Analysis of uterine EMG signals in term and preterm conditions using generalised Hurst exponent features

N. Punitha✉ and S. Ramakrishnan

An attempt has been made in this Letter to analyse term (week of gestation (WOG) >37) and preterm (WOG ≤ 37) conditions using uterine electromyography (uEMG) signals and generalised Hurst exponent (GHE) features. For this analysis, public database signals recorded from the surface of abdomen are considered. Multifractal detrended fluctuation analysis is performed on the signals and the GHE is calculated. From the exponent, seven features are extracted and data-balancing based on synthetic minority over-sampling technique is used to retain a balanced feature contribution by the term and preterm records. Two classification algorithms namely, Naive Bayes and logistic regression (LR) are employed to classify the signals. Ten-fold cross validation approach is executed and the performance is validated using accuracy, precision and recall. The results show the uEMG signals exhibit multifractal characteristics and five GHE features are significant in distinguishing the term and preterm uEMG signals. The LR classifier gives the highest accuracy of 97.8%. Therefore, it appears that the multifractal Hurst exponent features in combination with LR classifier can be used as biomarkers for predicting the preterm or term delivery during the early stage of gestation.

Introduction: Preterm delivery, referred to as the birth of a baby before the 37th week of gestation (WOG), is a major problem during pregnancy resulting in a global incidence of 15 million each year. Neonates born preterm have a higher threat of mortality and may suffer from poor neurodevelopment. At present, there is no efficient technique to detect preterm labour. However, early diagnosis is necessary to prevent preterm birth [1].

Uterine electromyography (uEMG) is a method that records the uterine electrical activity non-invasively from the surface of the abdomen. The uEMG signals are associated with the physiological process that causes contractions. Hence, investigation of these signals provides a better understanding of pregnancy and labour processes that allow prediction of preterm threat at an early stage of pregnancy (28th WOG) [2].

Researchers emphasise on applying signal processing techniques on uEMG signals to differentiate term and preterm records. Non-linear techniques are proven to provide better results to investigate uEMG signals [3]. One such method is the fractal analysis where, a non-integer parameter called fractal dimension is estimated to define the scale invariant characteristics of the signal. Fractals are structures that repeat themselves across different scales and give a measure of complexity [4].

A signal that possesses different fractal characteristics is said to be multifractal in nature. Multifractal analyses are found to be advantageous as a predictive and diagnostic tool in the investigation of physiological signals [5]. The generalised Hurst exponent (GHE) calculated using multifractal analysis represents the presence of long-range self-dependencies in a signal. GHE have been used to analyse electrocardiogram recordings to differentiate different levels of atrial fibrillation, as it provides information on the complexity of the atrial activation patterns [6]. Hence, in this work, an attempt has been made to illustrate the application of multifractal GHE features for differentiating uEMG signals in term and preterm groups before 26 WOG.

Methodology: The uEMG signals considered in this study are acquired from the Term–Preterm ElectroHysteroGram (TPEHG) database [3]. This dataset contains 143 term and 19 preterm delivery records, which are recorded before 26th WOG at a sampling frequency of 20 Hz. The signals at the frequency range of 0.3–3 Hz are reported to be efficient in differentiating term and preterm conditions [3]. Hence, data from this band of frequencies are considered for analysis.

The uEMG signals are subjected to multifractal detrended fluctuation analysis (MFDFA). It is a generalisation of detrended fluctuation analysis that is used to analyse non-stationary multifractal time series [7]. For this, the uEMG series is denoted as a series of cumulative sums Y and divided into N_l non-overlapping segments of same length l. A polynomial w_l is fitted for every segment u to identify the local trend and the corresponding root-mean-square gives the fluctuation function

$$F^2(l,u) = \frac{1}{l}\sum_{i=1}^{l}\left\{Y[(u-1)l+i] - w_l(i)\right\}^2 \quad (1)$$

Mean of all segments gives the q-order fluctuation function $F_q(s)$. The power-law function is calculated by changing the scale, and GHE is computed for each order which is given by $h(q)$ as follows:

$$F_q(l) = \left\{\frac{1}{N_l}\sum_{u=1}^{N_l}\left[F^2(l,u)\right]^{q/2}\right\}^{1/q} \approx l^{h(q)} \quad (2)$$

GHE indicates the presence of long-range dependencies in the q-order moments. A suitable choice of q for biological signals is from −5 to +5 [5] and hence, the same is used in this study.

For all the signals, GHE is calculated and seven features namely, maximum (HE_{MAX}), minimum (HE_{MIN}), zero-order (HE_0) and Hurst exponents (HE_2), degree of multifractality (DOM), degree of small (DSF) and large (DLF) fluctuations are extracted. HE_{MAX} and HE_{MIN} represent the maximum and minimum values of GHE for negative and positive orders indicating small and large fluctuations, respectively. HE_0 and HE_2 correspond to the values of GHE at $q = 0$ and $q = 2$ respectively. DOM is an indicator of multifractality, given by the difference in GHE between the extremes. DSF and DLF are calculated using (3) and (4).

$$\mathrm{DSF} = h(q=-5) - h(q=0) \quad (3)$$

$$\mathrm{DLF} = h(q=0) - h(q=5) \quad (4)$$

In the TPEHG database, the number of term records is found to be more than that of preterm records. As a consequence, the classifiers are profound to detect the majority class than the minority class, leading to bias [1]. Hence, there is a need for data-balancing, which is performed bythe synthetic minority over-sampling technique (SMOTE). SMOTE is an over-sampling technique which generates new features from the existing ones, resulting in a balanced training set. This algorithm works in the feature space instead of data space [8].

Further, the balanced feature set is checked for normality using the Quantile–Quantile plot and statistically analysed using the t-test. Two classifiers namely, Naive Bayes (NB) and logistic regression (LR) are trained with these features and the performances are analysed using metrics namely, accuracy, precision and recall.

Results and discussion: The representative uEMG signals recorded before 26th WOG that lead to term and preterm delivery are shown in Fig. 1. It is seen that the amplitude of the preterm signal is comparatively higher than that in term condition. This may be due to an increase in the strength of uterine muscle contraction towards labour.

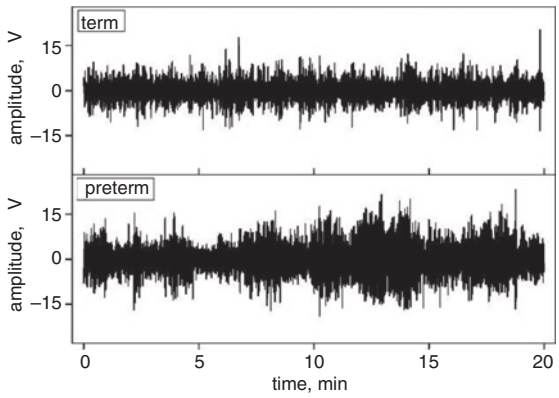

Fig. 1 *Representative uEMG signals*

The average variation of GHE for increasing orders of the fluctuation is presented in Fig. 2. It is seen that GHE in term and preterm delivery signals vary non-linearly with an increase in the orders of fluctuation and this indicate that uEMG signals are multifractal in nature. Also, the values of GHE in preterm are higher than that of a term group which can be due to the existence of more fluctuations in the amplitude of the signal. The presence of fluctuations is related to frequency and amplitude changes in a time series.

The statistical measures of the extracted features are shown in Table 1 and the scatterplots of representative significant features are shown in Fig. 3. It is inferred that five features namely, HE_{MAX}, HE_0, DLF, DOM and HE_2 show statistical significance in differentiating the signals in term and preterm groups. Conversely, there is no statistical difference in HE_{MIN} and DSF features. It can be seen that there is an increase in the average values of HE_{MAX} and HE_0 in the preterm group. This can be due to the increase in smaller fluctuations in the signal which influence the low-frequency components in the signals. HE_2 denotes the average fluctuation function of a time series where a higher value corresponds to a more regular signal [6]. It is witnessed that the value of HE_2 in the preterm condition is comparatively higher than in term group. This increase in regularity could be due to increased coordination of the uterus muscles. DOM is increased in preterm condition, signifying an increase in the multifractality. This shows that there is an increase in multifractal characteristics as labour approaches. DLF is observed to be 0.17 and 0.20 in term and preterm groups, respectively. It is an indicator of large fluctuations that may occur due to a significant change in both frequency and amplitude.

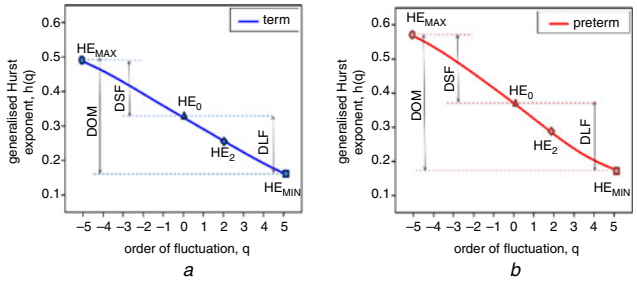

Fig. 2 *Representative GHEs with features*

a Generalised Hurst exponent, $h(q)$ – term
b Generalised Hurst exponent, $h(q)$ – preterm

Table 1: Statistics of GHE features

Features	Min	Max	Mean(SD)	Min	Max	Mean(SD)
HE_{MAX}*	0.30	0.62	0.49(0.07)	0.40	0.67	0.55(0.06)
HE_0*	0.16	0.46	0.33(0.05)	0.30	0.42	0.37(0.03)
HE_2*	0.10	0.36	0.27(0.04)	0.22	0.33	0.29(0.03)
HE_{MIN}	0.08	0.29	0.16(0.08)	0.07	0.24	0.17(0.07)
DSF	0.01	0.27	0.16(0.04)	0.10	0.26	0.18(0.03)
DLF*	0.07	0.46	0.17(0.08)	0.09	0.48	0.20(0.07)
DOM*	0.13	0.69	0.33(0.10)	0.19	0.67	0.37(0.09)

*Statistically significant ($p < 0.05$).

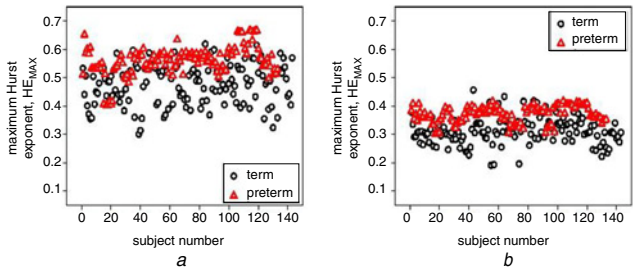

Fig. 3 *Scattergram of representative significant features*

a Maximum Hurst exponent, HE_{MAX}
b Zero-order Hurst exponent, HE_0

Fig. 4*a* displays the performance measures of the classifiers trained with all seven extracted features. It is observed that LR performs better than NB with a classification accuracy of 96%. Fig. 4*b* shows the performance measures of the classifiers trained with five significant features. It is seen that when only significant features are considered, there is an increase in the classifiers' performance. Highest recall of 97% and a precision of 98.5% is achieved using LR classifier.

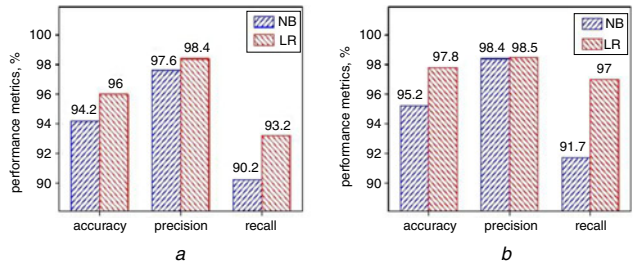

Fig. 4 *Performance metrics of the considered NB and LB classifiers*

a With all the seven extracted GHE features
b With only five significant GHE features

Conclusion: In this study, multifractal characteristics of uEMG signals are analysed to differentiate term and preterm conditions in the early stage of pregnancy (second trimester). MFDFA algorithm is implemented on the uEMG signals to calculate GHE. The results of GHE prove the multifractal characteristics of uEMG signals. In order to differentiate the signals, features are extracted from GHE and investigated. It is found that that there is a rise in multifractality in the preterm condition which can be due to the synchronisation of action potentials generated from the uterine cells towards labour. Five features namely, HE_{MAX}, HE_0, HE_2, DLF and DOM are able to distinguish signals in term and preterm groups and can aid in classification. LR classifier achieved the highest classification accuracy when trained with significant GHE features. Hence, it appears that the proposed approach of using multifractal Hurst exponent features extracted from MFDFA method, along with the LR classifier, is efficient to classify the uEMG signals and help in predicting the term or preterm delivery of pregnant women.

Submitted: *06 March 2019* E-first: *15 May 2019*
doi: 10.1049/el.2019.0803
One or more of the Figures in this Letter are available in colour online.

N. Punitha and S. Ramakrishnan (*Biomedical Engineering Group, Department of Applied Mechanics, Indian Institute of Technology Madras, Chennai, India*)

✉ E-mail: npunitha92@gmail.com

References

1 Fergus, P., Cheung, P., Hussain, A., *et al.*: 'Prediction of preterm deliveries from EHG signals using machine learning', *Plos One*, 2013, **8**, (10), p. e77154, doi: 10.1371/journal.pone.0077154
2 Punitha, N., and Ramakrishnan, S.: 'Analysis of uterine electromyography signals in preterm condition using multifractal algorithm'. Conf. Proc. IEEE Engineering Medicine Biology Society, Honolulu, Hawaii, July 2018, pp. 1–4, doi: 10.1109/EMBC.2018.8512891
3 Fele-Žorž, G., Kavšek, G., Novak-Antolič, Ž., *et al.*: 'A comparison of various linear and non-linear signal processing techniques to separate uterine EMG records of term and pre-term delivery groups', *Med. Biol. Eng. Comput.*, 2008, **46**, (9), pp. 911–922, doi: 10.1007/s11517-008-0350-y
4 West, B.J.: 'Fractal physiology and chaos in medicine' (World Scientific, Singapore, 2012, 2nd edn.), p. 332
5 Ihlen, E.A.F.E.: 'Introduction to multifractal detrended fluctuation analysis in Matlab', *Front. Physiol.*, 2012, **3**, (141), pp. 1–18, doi: 10.3389/fphys.2012.00141
6 Julián, M., Alcaraz, R., and Rieta, J.J.: 'Study on the optimal use of generalized Hurst exponents for noninvasive estimation of atrial fibrillation organization'. IEEE Conf. Computing Cardiology, Zaragoza, Spain, September 2013, pp. 1039–1042
7 Kantelhardt, J.W., Zschiegner, S.A., Koscielny-Bunde, E., *et al.*: 'Multifractal detrended fluctuation analysis of nonstationary time series', *Phys. A, Stat. Mech. Appl.*, 2002, **316**, (1–4), pp. 87–114, doi: 10.1016/S0378-4371(02)01383-3
8 Chawla, N.V., Bowyer, K.W., Hall, L.O., *et al.*: 'SMOTE: synthetic minority over-sampling technique', *J. Artif. Intell. Res.*, 2002, **16**, pp. 321–357, doi: 10.1613/jair.953