Dedicated entropy measures for early assessment of pregnancy progression from single-channel electrohysterography

Massimo Mischi, Chuan Chen, Tanya Ignatenko, Hinke de Lau, Beijing Ding, Guid S. Oei and Chiara Rabotti

Abstract—Objective: Preterm birth is a large-scale clinical problem involving over 10% infants. Diagnostic means for timely risk assessment are lacking and the underlying physiological mechanisms unclear. To improve the evaluation of pregnancy before term, we introduce dedicated entropy measures derived from a single-channel electrohysterogram (EHG). Methods: The estimation of Approximate Entropy (ApEn) and Sample Entropy (SampEn) is adjusted to monitor variations in the regularity of single-channel EHG recordings, reflecting myoelectrical changes due to pregnancy progression. In particular, modifications in the tolerance metrics are introduced for improving robustness to EHG amplitude fluctuations. An extensive database of 58 EHG recordings with 4 monopolar channels in women presenting with preterm contractions was manually annotated and used for validation. The methods were tested for their ability to recognize the onset of labor and the risk of preterm birth. Comparison with the best single-channel methods according to the literature was performed. Results: The reference methods were only performed. SampEn and ApEn produced the best prediction of delivery, although only one channel showed a significant difference (p < 0.04) between labor and non-labor. The modified ApEn produced the best prediction of preterm delivery, showing statistical significance (p < 0.02) in 3 channels. These results were also confirmed by the area under the receiver operating characteristic curve and 5-fold cross-validation. Conclusion: The use of dedicated entropy estimators improves the diagnostic value of EHG analysis earlier in pregnancy. Significance: Our results suggest that changes in the EHG might manifest early in pregnancy, providing relevant prognostic opportunities for pregnancy monitoring by a practical single-channel solution.

Index Terms—Electrohysterogram, pregnancy, preterm delivery, entropy, uterus

I. INTRODUCTION

Birth occurring before completing the 37th week of gestation (preterm), accounts for 75% of perinatal mortality and more than half of birth-related long-term morbidity [1]. With the exception of births induced for medical reasons before the 37th week, preterm birth is spontaneous and the most common cause. Accurate distinction between these two types of uterine activity is crucial because, while threatening contractions need to be promptly suppressed by medical intervention, rest is often sufficient to sedate physiological, harmless uterine activity [3]. Eventually, about half of the women hospitalized due to premature contractions deliver at term, even without treatment [4]; this evidences the limits of the available diagnostic tools, often causing overdiagnosis and overtreatment.

Only few options are nowadays available for characterizing the uterine activity. Among these options, intra-uterine pressure measurement is the most reliable; however, due to the need for catheterization, it cannot be used for pregnancy monitoring before labor [5]. External tocography, being non-invasive, is extensively used, but it provides only quantification of contraction frequency and duration. Furthermore, tocography has been demonstrated not to be able to predict delivery. Even cervical changes may not be an accurate indicator of labor alone, as a large percentage of women with established cervical changes do not deliver preterm, even without treatment [4].

Impending preterm delivery is likely associated with a change in uterine cell excitability, favoring conduction of electrical activity across the uterine muscles, the myometrium. Analysis of the electrohysterogram (EHG), which is non-invasive measurement of the uterine electrical activity, has therefore been investigated as a potential method for pregnancy monitoring and early diagnosis of preterm birth [6]–[22].

Previous research on EHG-signal analysis has mainly addressed the diagnosis of preterm birth, with focus on either one of two distinct objectives. The first objective is to predict the delivery time irrespective of its occurrence before (preterm) or after (at term) the 37th week of gestation (prediction of delivery). This translates into the recognition of contractions related to imminent delivery [18], e.g., by evaluating the correlation between the evaluated features and the measurement-to-delivery interval. The second objective focuses on early differentiation between those pregnancies that will eventually reach the 37th week of gestation and those that will end with a preterm delivery, irrespective of the measurement-to-delivery time interval (diagnosis of preterm birth).

While the first objective allows inclusion of measurements at any gestational age, the latter necessarily requires the availability of EHG measurements early in pregnancy, i.e., from nate, at the onset of the first symptoms, between contractions that represent a real threat of imminent preterm birth and physiological uterine activity triggered e.g. by environmental causes. Accurate distinction between these two types of uterine activity is crucial because, while threatening contractions need to be promptly suppressed by medical intervention, rest is often sufficient to sedate physiological, harmless uterine activity [3]. Eventually, about half of the women hospitalized due to premature contractions deliver at term, even without treatment [4]; this evidences the limits of the available diagnostic tools, often causing overdiagnosis and overtreatment.

Impending preterm delivery is likely associated with a change in uterine cell excitability, favoring conduction of electrical activity across the uterine muscles, the myometrium. Analysis of the electrohysterogram (EHG), which is non-invasive measurement of the uterine electrical activity, has therefore been investigated as a potential method for pregnancy monitoring and early diagnosis of preterm birth [6]–[22].

Previous research on EHG-signal analysis has mainly addressed the diagnosis of preterm birth, with focus on either one of two distinct objectives. The first objective is to predict the delivery time irrespective of its occurrence before (preterm) or after (at term) the 37th week of gestation (prediction of delivery). This translates into the recognition of contractions related to imminent delivery [18], e.g., by evaluating the correlation between the evaluated features and the measurement-to-delivery interval. The second objective focuses on early differentiation between those pregnancies that will eventually reach the 37th week of gestation and those that will end with a preterm delivery, irrespective of the measurement-to-delivery time interval (diagnosis of preterm birth).

While the first objective allows inclusion of measurements at any gestational age, the latter necessarily requires the availability of EHG measurements early in pregnancy, i.e., from...
patients with preterm contractions. The challenge here is that not only are these patients more difficult to recruit, but preterm measurements are also difficult to annotate and process, due to the poor signal-to-noise ratio (SNR) and the lack of a golden standard for contraction assessment. Furthermore, extending the results obtained in term patients for the prediction of the delivery time to the diagnosis of preterm birth underlies the hypothesis that the pathogenesis of preterm delivery coincides with an earlier evolution of the physiological process leading to a delivery at term. Despite its viability, this hypothesis has not been proven yet.

Motivated by the lack of insight into the physiology underlying the onset of labor and the exact mechanism underpinning spontaneous preterm birth, a number of different features extracted from the EHG have been proposed for (preterm) delivery prediction, both on single- and multi-channel signals.

Promising single-channel features are related to the frequency content of the EHG signal. In particular, an increase in the peak frequency extracted from the signal power spectral density (PSD) seems to be an indicator of approaching labor both at term and preterm, when measured within 1-7 days to delivery [8], [10], [23]–[29]. A significant distinction between term and preterm EHG measurements has been shown in [10] by estimation of the median frequency.

Studies on parameters related to EHG signal amplitude and energy show more controversial results [10], [17], [28], [30]. A number of studies report an increase in the power spectrum amplitude just prior to delivery both by analyzing the whole signal as well as each individual burst separately, and a significant difference in root mean squares between recordings performed before and after the 26th week of gestation [10], [24]. However, a number of studies failed to show a significant correlation of amplitude-related parameters with gestational age or with the measurement-to-delivery time [18], [28], [30].

The use of nonlinear features such as entropy, time reversibility, and maximal Lyapunov exponent [10], [16], has also been tested for prediction of labor. Among the methods evaluated for the estimation of EHG nonlinear features, the use of time reversibility accurately separated pregnancy and labor contractions on a dataset of 12 women, processing manually-selected contraction segments [17]. Time reversibility can therefore be considered as a potential method for the prediction of delivery [17].

The initiation of labor is likely related to altered levels of myometrial cell connectivity that induce changes in the regularity of the measured EHG signal [31]. In line with this hypothesis, Sample Entropy (SampEn) measured early in pregnancy provided the best separation between term and preterm delivery in a large cohort of preterm women when compared to a number of other linear and nonlinear features [10]. Noteworthy, these results have been obtained by analyzing the whole EHG recording without distinction between contractile and quiescent segments. While very advantageous for the clinical implementation of the method, this approach does not discriminate between the contribution of quiescent and contractile signal segments to the final results.

More recently, estimation of the EHG conduction velocity has been proposed to predict (preterm) delivery [14], [32]. Physiologically, this approach hypothesizes a relation between the increased amount of gap-junctions observed with approaching delivery and changes in the conduction properties of the myometrial cells, which can be reflected in variations of the EHG conduction velocity [14], [31], [33], [34]. To this end, multichannel EHG recordings have been used and a number of algorithms proposed for the analysis. Promising results have been shown by visual analysis and selection of EHG signal collected in a large population with preterm contractions [18].

Beyond the approximation of the EHG as resulting from propagating plane waves, additional connectivity measures, such as nonlinear correlation, have been tested and evaluated for their ability to reflect the underlying uterine activity [17].

The use of graph theory is also being investigated to assess connectivity (spatial dependency) and detect labor [35].

In this paper, we propose new dedicated methods for preterm birth diagnosis from single-channel EHG signals that are based on entropy features. The use of a single channel facilitates the clinical uptake of the methods, requiring the application of two or three (reference plus ground) electrodes only. With the aim of detecting variations in signal regularity, the proposed methods are based on a modification of the traditional calculation of Approximate Entropy (ApEn) and Sample Entropy (SampEn).

ApEn [36] and SampEn [37] are signal features derived from information theory. They reflect the regularity of a time series and, therefore, have been extensively used for analysis and characterization of physiological signals. The available methods for entropy estimation are more accurate for time series with moderate fluctuations of their mean and standard deviation [38], [39]. Due to the occurrence of uterine action potentials in bursts and to the progressive recruitment often observed at the beginning and at the end of each burst [40], the EHG signal typically shows large amplitude fluctuations that may weaken the hypothesis of stationarity. This may explain the poor results reported for the use of ApEn in order to estimate labor in a small cohort of 7 patients [17].

Here we revisited and adjusted the derivation of ApEn and SampEn in order to propose dedicated methods for EHG signal analysis that aim at suppressing the influence of amplitude fluctuations on the estimation of the signal entropy. A dataset containing EHG measurements on 81 patients with preterm contractions was available for validation. These patients were recruited at the Máxima Medical Center (MMC) in Veldhoven (the Netherlands) and included in a clinical study referred to as PoPE [41]. Based on this dataset, the proposed methods were evaluated on both the possible objectives: 1) the ability of classifying each patient as in labor or not and 2) the accuracy in the diagnosis of preterm birth. Only contraction periods were considered for the analysis. For comparison, previously proposed statistical methods based on single channel EHG analysis were also tested on our dataset. In particular, we selected those two methods that have shown, according to the literature, the best classification performance in each of the two classification objectives; these are Time Reversibility and SampEn, respectively.
II. MATERIALS AND METHODS

A. Data acquisition

Data were collected at the MMC within the PoPE study, approved by the ethical medical committee of the hospital. After signing an informed consent, 120 women admitted to the hospital for preterm contractions underwent multichannel EHG recording. An electrode patch containing five sensors (Nemo Healthcare, Veldhoven, the Netherlands), including a ground electrode, was placed in the middle line of the abdomen [41], as shown Fig. 1. An additional reference electrode was placed on the right hip to obtain monopolar derivations.

A multichannel Porti system (Twente Medical Systems International, Enschede, the Netherlands) was used to amplify and digitize the data at a sampling frequency of 500 Hz. Out of the 120 included women, 39 were originally excluded from the PoPE study because of technical failure with the acquisition. The duration of the remaining recordings ranged from 20 minutes to 60 minutes. Of the included 81 patients, 18 had a synchronized registration with a tocodynamometer.

Two independent clinical experts annotated manually the contraction periods on the full dataset after training on six patients with synchronized tocodynamometer measurement. The contraction periods were identified as uterine electrical bursts showing a rise in the intra-uterine pressure estimated by EHG analysis ([11] and [42]) above twice the baseline for a minimum duration of 20 s. In addition, an accelerometer was placed next to the electrode patch and the recorded signal analyzed in order to exclude false contractions due to motion artifacts. Low concordance between observers necessitated a consensus meeting in order to reduce dissimilarity in segmentation. Eventually, only segments selected by both observers were used for the analysis.

In order to reinforce our choice to limit the analysis to contractile activity, only contraction segments recognized by all experts, and confirmed by the automatic selection method, were retained for further analysis. Of the remaining dataset, to ensure sufficient data for the following analysis, only recordings with at least 5 min of annotated contraction time were considered, resulting in a final validation dataset comprising the recordings of 58 patients.

B. Pre-processing

For each recording, the contraction segments were concatenated to form a new sequence. Similar to previous studies [14] [20] [25], the recorded signals were band-pass filtered within the frequency range from 0.3 to 0.8 Hz in order to suppress the main noise components due to respiration, motion artifacts, and maternal ECG. Figure 2 shows an example of EHG signal containing two contractions before and after band-pass filtering. After filtering, down-sampling to 20 Hz was applied.

The literature on the choice of unipolar or bipolar derivations for EHG signal analysis is rather controversial. Although bipolar derivations offer a better SNR due to the suppression of common mode components, unipolar derivations should be preferred when the spatial properties of the signal in terms of origin and direction of propagation are a priori unknown and may impact bipolarization in an unpredictable way [40]. Therefore, we performed our analysis independently on the 4 unipolar signals of the electrode patch.

C. Entropy features

1) Approximate Entropy: ApEn, originally formalized in [36], is a family of statistics that provides a measure of regularity, closely related to the Kolmogorov entropy; it can be applied to typically short and noisy clinical data [36] [43] [44]. When this statistic is used to compare time series for similar epochs, more frequent and more similar epochs lead to lower values of ApEn [37]. Therefore, we consider ApEn to be particularly suitable for revealing EHG changes in relation to pregnancy progression and labor.

For a description of ApEn, we refer to a time series, \( u(j) \), with \( j = 1, \ldots, N \). In order to estimate its ApEn, we subdivide \( u(j) \) in epochs of length \( m \), forming \( N - (m+1)\tau \) vectors \( x_m(i) \), with \( x_m(i) = (u(i+k\tau) : 0 \leq k \leq m-1) \), \( \{i | 1 \leq i \leq N - \tau(m-1)\} \), and \( \tau \) an integer time delay. A tolerance, \( r \), is defined for accepting matches among the epochs. This tolerance is typically defined based on the...
vectors’ Euclidean distance or on the maximum difference
of their scalar components, namely, \( \| \mathbf{x}_m(i) - \mathbf{x}_m(j) \|_2 \leq r \)
or \( \max_k |u(i+k\tau) - u(j+k\tau)| \leq r \), respectively [43]. In this
work, as the signals are already pre-filtered and down-sampled
based on the EHG bandwidth, no additional down-sampling is
applied for the definition of the epochs, i.e., \( \tau = 1 \). Moreover,
due to the large amplitude fluctuations of the EHG, the
Euclidian’s distance is used to define the tolerance \( r \). This
results in a definition of tolerance that is less sensitive to the
presence of spikes and large fluctuations in EHG recordings.

Given tolerance \( r \), let \( B_i \) be the number of vectors \( \mathbf{x}_m(j) \)
within \( r \) of \( \mathbf{x}_m(i) \). The empirical probability \( C^m(r) \) that a vector
\( \mathbf{x}_m(j) \) is within \( r \) of \( \mathbf{x}_m(i) \) can be estimated by

\[
C^m(r) = \frac{B_i}{N - m + 1}.
\]

(1)

After introducing \( \Phi^N_m(r) \), given as

\[
\Phi^N_m(r) = \frac{\sum_{i=1}^{N-m+1} \ln[C^m(r)]}{N - m + 1},
\]

(2)

ApEn can be defined [38] as

\[
\text{ApEn}(m, r, N) = \lim_{N \to \infty} | \Phi^N_m(r) - \Phi^N_{m+1}(r) |,
\]

which, for a finite time series, is estimated as

\[
\text{ApEn}(m, r, N) = \Phi^N_m(r) - \Phi^N_{m+1}(r).
\]

(4)

This estimated ApEn\((m, r, N)\), for fixed parameters \( m \), \( N \),
and \( r \), provides a measure of regularity and complexity [38].
Theoretically, a perfectly regular time series would show ApEn
close to 0, while a perfectly irregular time series would show
ApEn close to 2 [45].

2) Sample Entropy: SampEn is a family of statistics that
has been specifically proposed to reduce the bias caused by
self-matching counts introduced in the calculation of ApEn by
the occurrence of \( \ln(0) \).

Based on the same definitions used for ApEn, and making
the same choices for \( \tau \) and \( r \), \( C^m_a(r) \) is defined as the template-
wise probability that \( \mathbf{x}_m(j) \) is located within the tolerance
\( r \) of \( \mathbf{x}_m(i) \), with \( j \neq i \), to exclude self-matches in the epoch
comparison. The average of \( C^m_a(r) \), given as

\[
C^m_a(r) = \frac{1}{(N - m + 1)} \sum_{i=1}^{N-m+1} C^m_a(r),
\]

(5)
is then used as the overall probability to approximate the
entropy as

\[
\text{SampEn}(m, r, N) = -\ln \left[ \frac{C^{m+1}_a(r)}{C^m_a(r)} \right], C^m_a(r), C^{m+1}_a(r) \neq 0.
\]

(6)

Notice that, although the definitions of ApEn and SampEn
look similar, Eq. 6 represents a conditional entropy, different
from the entropy rate represented by Eq. 4.

3) Modified distance metrics: In general, more accurate
entropy estimation is obtained when the mean and standard
deviation of the analyzed signal shows limited variation over
time. Therefore, the presence of large amplitude fluctuations
and spikes, typical of EHG signals [40], may affect the
estimated ApEn and SampEn more than signal regularity. This
makes direct application of entropy measures unsuitable for
most EHG applications, hampering the possibility to perform
accurate classification [17], [38], [46]. To overcome this problem,
we propose a modification of the original distance metrics
aiming at limiting the tolerance dependency on large amplitude
fluctuations and spikes.

In particular, we modify the tolerance and define it by
angular metrics, i.e., the tolerance \( \rho \), with \(-1 < \rho < 1\), is
defined based on the angle between the vectors \( \mathbf{x}_m(i) \) and
\( \mathbf{x}_m(j) \), \( \theta(\mathbf{x}_m(i), \mathbf{x}_m(j)) \) expressed by:

\[
\cos(\theta(\mathbf{x}_m(i), \mathbf{x}_m(j))) = \frac{\mathbf{x}_m(i) \cdot \mathbf{x}_m(j)}{\| \mathbf{x}_m(i) \|_2 \| \mathbf{x}_m(j) \|_2} \geq \rho,
\]

(7)

where \( \cdot \) denotes the inner product of two vectors. We can
then use the estimated entropy measures (ApEn and SampEn)
based on this new tolerance to compare epochs that are scaled
in amplitude.

We can observe that the proposed entropy analysis with
modified tolerance is equivalent to the analysis of the original
entropy applied to normalized data. In fact, for the vectors
\( \mathbf{x}_m(i) = \frac{\mathbf{x}_m(i)}{\| \mathbf{x}_m(i) \|_2} \) and \( \mathbf{x}_m'(j) = \frac{\mathbf{x}_m(j)}{\| \mathbf{x}_m(j) \|_2} \) with a unit length, it
holds that

\[
\| \mathbf{x}_m'(i) - \mathbf{x}_m'(j) \|^2 = (1 - \cos(\theta(\mathbf{x}_m'(i), \mathbf{x}_m'(j)))).
\]

(8)

Therefore, defining \( B_i \) for this normalized data, we obtain

\[
B_i = \# j : \| \mathbf{x}_m'(i) - \mathbf{x}_m'(j) \|_2 \leq r = \# j : \| \mathbf{x}_m'(i) - \mathbf{x}_m'(j) \|^2 \leq r^2
= \# j : (1 - \cos(\theta(\mathbf{x}_m'(i), \mathbf{x}_m'(j)))) \leq r^2
= \# j : \cos(\theta(\mathbf{x}_m'(i), \mathbf{x}_m'(j))) \geq 1 - r^2 / 2 \geq \rho,
\]

where \# denotes number of elements.

Because of the similarity in their definition, SampEn and
ApEn are often interchanged and compared for analysis of
physiological signals [45], [46]. Therefore, the proposed
modified tolerance metrics based on angular distance was applied
to both entropy features.

D. Validation strategy

1) Classification objectives: All the features, namely,
ApEn, SampEn, Modified ApEn, and Modified SampEn, were
evaluated for their classification ability in the two identified,
relevant scenarios.

1) Prediction of delivery within 1 week from the measure-
ment, irrespectively of being at term or preterm. For
preterm cases, 1 week is a common cut-off to separate
pregnancy from labor.
2) Diagnosis of preterm birth, i.e., for distinguishing be-
tween real and false threat among women who were
prospectively classified at risk of preterm delivery.

According to the literature, SampEn provides the best results
for the prediction of labor (first scenario) [10]. However, the
best results for the second scenario are provided by the assess-
ment of Time Reversibility [17]. Therefore, for completeness,
this method was also implemented and compared.
2) Time Reversibility: A stochastic process is defined as time-reversible if it is invariant under the reversal of the time scale [47]. An example of time-reversible process is a linear Gaussian random process. Since Time Reversibility can be a strong signature of nonlinearity, in [22] it has been proposed as feature to distinguish between pregnancy and labor contractions.

A common measure of Time Reversibility is the third-order quantity

\[ Tr(\tau) = \frac{1}{N - \tau} \sum_{n=\tau+1}^{N} (S_n - S_{n-\tau})^3, \]  

where \( S \) is a signal with \( N \) samples, and \( \tau = 1 \) [47], [48]. According to the original study, 99 surrogate data are created by iterative amplitude-adjusted Fourier transform in order to match the same PSD of the original signal [48]. The value of Time Reversibility \( Tr(\tau) \) of the original time series is tested in order assess whether it is likely to be drawn, within a confidence level, from the distribution of values of the surrogates. Comparatively different values for the original series lead to the rejection of the null hypothesis and the original series is considered to be nonlinear. To this end, the Rank test is adopted, and the discriminative statistics of the original signal is used as feature for classification.

To accurately reproduce the original study, different from ApEn and Sample Entropy, assessment of Time Reversibility was based on separate contractions rather than on sliding windows. Since decimation to 20 Hz has been demonstrated not to degrade the performance of Time Reversibility for EHG signal analysis, 20-Hz sampling frequency was used also in this study [49].

Although all features were estimated on monopolar derivations, for the estimation of Time Reversibility only, some additional tests on bipolar derivations were also performed to fully reproduce the original study [50].

3) Statistics: The classification ability of each feature was assessed on the average values extracted from each woman data sequence in each electrode. The average of the results for the 4 electrodes was also considered.

The Wilcoxon rank sum test was adopted to evaluate the discrimination ability of the selected features in both scenarios. This statistics was chosen to account for the non-Gaussian distribution of the data and unbalanced classes [51]. The feature ability to provide correct classification was also assessed by the area under the receiver operating characteristic (ROC) curve, derived over the full dataset of 58 patients.

In addition, 5-fold cross-validation was performed in order to evaluate sensitivity, specificity, and accuracy of all the features for classification in both scenarios. To this end, the data were subdivided in 5 patient groups. The optimal threshold was then determined by ROC curve analysis (point closest to the upper-left corner) on 4 groups and applied to the remaining group to evaluate the classification performance, rounding in a way that all groups would undergo classification. This procedure was repeated for 50 random subdivisions in 5 groups.

4) Adopted parametrization: Prior to the proposed statistical analysis, the parametrization for the implemented methods was determined. In line with the literature, ApEn and SampEn were estimated using \( m = 3 \) and \( r \) equal to 0.15 the standard deviation of the signal [10], [38], [45], [46].

For the modified metrics, these parameters were chosen to optimize the classification performance on the subset of annotated EHG signals for which the external tocogram was also available. Classification aimed at differentiating between contraction and quiescent periods as annotated by the experts with the support of the tocogram and additional automatic detection algorithms based on [11], [42]. For each recording, quiescent and contraction periods were first concatenated to produce two separate signals representing the two different conditions. The differentiation was evaluated based on the p-value estimated by Student t-test for class separation. The estimation of modified ApEn and SampEn was therefore performed with \( m = 3 \) and \( \rho = \cos(\pi/18) \approx 0.98 \).

Based on the same optimization, all entropy features were estimated employing an 80-s sliding windows with 40-s overlap. For each patient, the average over all windows was considered in our statistical analysis presented in Section II-D3.

All signals were analyzed after preprocessing (Section II-B), including the concatenation of the contraction segments. For completeness, SampEn was also calculated on the whole signal, without contraction selection, in line with the original study described in [10].

All the analysis was implemented in Matlab® (The MathWorks, Natick, MA) running on a windows system with Intel(R) Core(TM) i7-4710MQ and 8 GB RAM. With this implementation, the computational time for the measurement of entropy using the Euclidian distance metrics was 10-s per window, and 16-s per window using the modified distance metrics.

III. RESULTS

In total, 58 women with preterm contractions were analyzed, of which 34 delivered preterm. Nine of the women who delivered preterm were measured at less than one week to delivery (labour). The gestational age at delivery versus the measurement-to-delivery interval is shown in Fig. 3. We analyzed a total number of 804 contractions which were distributed across the included 58 subjects as shown in Fig. 4.

A. Prediction of delivery

Focussing on the classification of each women as being in labor or not, the left side of Table I shows the average values of all the features together with the statistical significance (p-value) of the difference between non-labor and labor condition. The results are reported for each monopolar electrode and for the average values of all the electrodes, estimated for each woman. The best results are obtained by SampEn and ApEn, with minor differences between the two; women in labor showed lower values (average SampEn=0.312 and average ApEn=0.364) compared to those who were not in labor (average SampEn=0.403 and average ApEn=0.442). Significant differences, however, are obtained only with electrode 1 (E1 in Fig. 1). These results are also confirmed by the ROC curve areas, also reported in Table I. The full ROC curves are shown
Fig. 3. Measurement-to-delivery time interval vs. gestational age at delivery in the study population.

Fig. 4. Distribution of the measured contractions over the population included in the study.

in Fig. 5, where all the electrodes are considered independently to derive an average value of each feature per woman.

The assessment of the classification performance by 5-fold cross-validation is reported in Table II (left side). The best classification results are achieved with SampEn and ApEn, with differences depending on the classification objective: higher sensitivity for SampEn (63% vs. 54%) and higher specificity for ApEn (77% vs. 57%). The classification results for the modified ApEn are also very close.

When the statistical analysis was performed on Time Reversibility extracted by bipolar derivations, as presented in the original work, no statistical difference could be found.

B. Diagnosis of preterm birth

Following the same statistical analysis as for the prediction of labor, the right side of Table I shows the average values of all the features together with the statistical significance (p-value) of the difference between term and preterm deliveries.

The results show a significant improvement in the classification performance when the modified distance metrics is adopted. In particular, the best results are obtained by the modified ApEn (average p-value of 0.004); women who eventually delivered at term showed higher values of modified ApEn (0.251) compared to those who delivered preterm (0.248). This result is also confirmed by the ROC curve areas, reported in Table I. The full ROC curves are shown in Fig. 6, where all the electrodes are considered independently to derive an average value of each feature per woman.

The assessment of the classification performance by 5-fold cross-validation is reported in Table II (right side). Again, a clear advantage by Modified ApEn is evidenced by the results, especially in terms of specificity and accuracy.

For completeness, SampEn was also calculated on the whole signal, without contraction selection, following the same procedure reported in [10]. This solution led to a deterioration of the classification performance by SampEn in both classifi-
TABLE I

| Feature Difference Between Preterm/Term and Between Non-labor/Labor Groups |
|-----------------------------|-----------------------------|-----------------------------|
| Non-labor/Labor             | Preterm/Term                |
| Method                      | Electrode                   | Non-labor                   | Labor                        | p-value          | ROC area         | Preterm         | Term             | p-value          | ROC area         |
| ApEn                        | E1 0.451 ± 0.097            | 353 ± 0.150                | 0.039                       | 0.710                        | 0.438 ± 0.106   | 0.429 ± 0.125   | 0.641           | 0.277           |
|                             | E2 0.437 ± 0.105            | 345 ± 0.157                | 0.086                       | 0.675                        | 0.422 ± 0.112   | 0.421 ± 0.131   | 0.543           | 0.590           |
|                             | E3 0.437 ± 0.100            | 373 ± 0.147                | 0.198                       | 0.631                        | 0.423 ± 0.108   | 0.430 ± 0.116   | 0.482           | 0.573           |
|                             | E4 0.441 ± 0.093            | 385 ± 0.122                | 0.093                       | 0.671                        | 0.421 ± 0.091   | 0.447 ± 0.112   | 0.087           | 0.652           |
|                             | average 0.442 ± 0.092       | 364 ± 0.139                | 0.054                       | 0.696                        | 0.426 ± 0.098   | 0.431 ± 0.115   | 0.372           | 0.593           |
| Modified ApEn               | E1 0.250 ± 0.008            | 0.259 ± 0.006              | 0.943                       | 0.508                        | 0.249 ± 0.008   | 0.251 ± 0.005   | 0.193           | 0.602           |
|                             | E2 0.250 ± 0.007            | 0.570 ± 0.006              | 0.339                       | 0.398                        | 0.247 ± 0.006   | 0.252 ± 0.005   | 0.001           | 0.754           |
|                             | E3 0.250 ± 0.008            | 0.249 ± 0.004              | 0.198                       | 0.631                        | 0.248 ± 0.007   | 0.252 ± 0.007   | 0.000           | 0.705           |
|                             | E4 0.250 ± 0.008            | 0.247 ± 0.006              | 0.141                       | 0.650                        | 0.248 ± 0.008   | 0.252 ± 0.005   | 0.015           | 0.690           |
|                             | average 0.250 ± 0.007       | 0.248 ± 0.005              | 0.237                       | 0.621                        | 0.248 ± 0.007   | 0.251 ± 0.005   | 0.004           | 0.728           |
| SampEn                      | E1 0.413 ± 0.120            | 0.301 ± 0.156              | 0.025                       | 0.731                        | 0.391 ± 0.121   | 0.398 ± 0.151   | 0.455           | 0.359           |
|                             | E2 0.397 ± 0.127            | 0.294 ± 0.168              | 0.072                       | 0.683                        | 0.374 ± 0.126   | 0.387 ± 0.159   | 0.435           | 0.361           |
|                             | E3 0.397 ± 0.127            | 0.315 ± 0.162              | 0.153                       | 0.646                        | 0.376 ± 0.129   | 0.392 ± 0.147   | 0.502           | 0.533           |
|                             | E4 0.405 ± 0.121            | 0.340 ± 0.146              | 0.191                       | 0.633                        | 0.377 ± 0.115   | 0.417 ± 0.142   | 0.120           | 0.621           |
|                             | average 0.403 ± 0.117       | 0.312 ± 0.150              | 0.054                       | 0.696                        | 0.380 ± 0.115   | 0.398 ± 0.143   | 0.356           | 0.572           |
| Modified SampEn             | E1 0.244 ± 0.009            | 0.241 ± 0.006              | 0.371                       | 0.592                        | 0.242 ± 0.010   | 0.246 ± 0.005   | 0.026           | 0.674           |
|                             | E2 0.245 ± 0.009            | 0.241 ± 0.008              | 0.299                       | 0.606                        | 0.242 ± 0.009   | 0.247 ± 0.007   | 0.015           | 0.689           |
|                             | E3 0.245 ± 0.010            | 0.243 ± 0.007              | 0.382                       | 0.590                        | 0.243 ± 0.010   | 0.247 ± 0.008   | 0.051           | 0.652           |
|                             | E4 0.245 ± 0.009            | 0.240 ± 0.009              | 0.106                       | 0.665                        | 0.242 ± 0.010   | 0.247 ± 0.007   | 0.021           | 0.680           |
|                             | average 0.245 ± 0.008       | 0.241 ± 0.007              | 0.184                       | 0.635                        | 0.242 ± 0.009   | 0.247 ± 0.006   | 0.020           | 0.681           |
| Time                        | E1 9.463 ± 0.308            | 10.792 ± 0.354             | 0.560                       | 0.585                        | 9.865 ± 0.017   | 9.446 ± 0.631   | 0.390           | 0.394           |
|                             | E2 8.269 ± 5.324            | 8.161 ± 5.253              | 0.894                       | 0.500                        | 8.270 ± 5.355   | 8.223 ± 5.251   | 1.000           | 0.515           |
|                             | average 9.009 ± 4.089       | 10.152 ± 5.336             | 0.673                       | 0.540                        | 9.441 ± 4.430   | 8.873 ± 4.174   | 0.608           | 0.344           |

The results on the prediction of delivery, based on the differentiation between labor and non-labor condition, show SampEn and ApEn as providing the best classification performance. This result is confirmed by all figures, namely, p-value, area under the ROC curve and, although to a lesser extent, the results of the 5-fold cross validation. Instead, the modified metrics, as well as the reference method from the literature, Time Reversibility, resulted in lower performance. The ApEn improvement compared to the results reported in the literature [17] may relate to the use of the Euclidian’s metrics, which is more robust than the maximum distance in the case of low SNR.

The results on the prediction of preterm delivery show the modified metrics to outperform the standard definitions of ApEn and SampEn. This is confirmed by all statistical figures. In particular, the modified ApEn results in the most accurate prediction and best differentiation between the two groups. This result suggests that while approaching labor is reflected in amplitude changes in the signal, signs related to the risk of preterm labor are independent of the signal amplitude and are mainly related to the regularity of the normalized EHG time series. This would support the hypothesis that preterm delivery cannot be simply regarded as early idiopathic activation of the normal labor process. In general, although the mechanisms underlying term and pre-term delivery are poorly understood, the EHG seems to offer an important tool for early assessment of preterm risk, as well as for revealing relevant changes in the uterine activity during pregnancy.

Our study reveals a decrease in signal entropy, both ApEn and SampEn, for women who are in labor compared to women who are not in labor. This result is also in line with the results obtained with the modified measures of entropy for the assessment of the risk of preterm birth. Women that are at risk...
show a lower signal value of modified entropy, both ApEn and SampEn. Physiologically, a decrease in signal entropy may be expected as a result of increased coordination among cells at the myometrial level. However, different underlying mechanisms may be reflected by the different distance metrics employed for the entropy estimation.

In general, when single electrodes are evaluated, each method shows the best results for different positions. This result, in line with previous studies [54], may be due to the complex spatial distribution of the EHG characteristics, which may affect differently each feature depending on the electrode position. In our measurements, E1 (upper right electrode in Fig. 1) provided the best results for labor detection, while it provided the worst results for prediction of preterm delivery. This result may again relate to the different physiological mechanisms in pregnancies leading to term or preterm delivery.

In the pregnant uterus, contractions occur in bursts of action potentials. Recently, the approach towards the analysis of the EHG has polarized toward processing the whole signal regardless of the alternation between active bursts and quiescent periods. Indeed, the detection of contractions is a critical preprocessing step, particularly preterm, and the lack of annotations on the publicly available databases does not allow for validation. Sample Entropy has been previously evaluated on the whole signal and Time Reversibility on manually segmented contraction segments. In this study, we focussed on an in-house dataset, where only validated contractions segments were used for the analysis. The aim was to focus on the contribution of the uterine activity only, while neglecting factors such as contraction frequency or duration. For completeness, to fully reproduce the results reported in the literature, SampEmp was additionally applied to the whole signal, including the quiescent periods. However, these results showed lower classification performance. We may therefore hypothesize that most entropy information for the characterization of pregnancy relates to the contraction phase, while the remaining signal introduces noise in the measurement.

Time Reversibility had been previously presented as a measure of signal nonlinearity that shows significant increase with approaching delivery. Our results did not evidence the same performance. The reason for this discrepancy may relate to the different strategy adopted in the original study [50], averaging the results based on the contractions in the entire groups, without distinction between subjects. This way, the results are influenced by the number of contractions recorded per woman, which are highly variable and dependent, among other factors, on the time to delivery. Therefore, we opted to evaluate the average value per woman. Importantly, the use of monopolar derivations may have also affected the estimation of the Time Reversibility due to the bias introduced by the common mode. However, additional tests using bipolar derivations did not show any significant improvement.

In general, although the use of a single EHG channel is desirable for practical use and the obtained results are promising, the obtained classification accuracy shows room for improvement. In order improve the obtained results on single channel recordings, multiple features reflecting different aspects of the underlying physiological phenomena may be combined in the future. Alternative entropy measures, such as correlation and variance entropy, could therefore be investigated for their ability to provide additional, complementary information. Machine-learning techniques could then be employed to achieve ranking and optimal combination of all the available features and information [55]. In addition, multichannel analysis, possibly coupled with advanced multiscale
modeling of the uterine electrical activation [56]–[58], may provide additional features that are valuable to elucidate the role of cell activity and connectivity in relation to the onset of labor and preterm labor.

V. CONCLUSION

Entropy measures have been revisited and adjusted with the objective of achieving noninvasive prediction of delivery and diagnosis of preterm birth by analysis of single-channel EHG signals. On an extended database of in-house preterm EHG measurements, the estimated ApEn and SampEn have shown the best prediction of delivery, while the modified ApEn has shown a clear advantage for early diagnosis of delivery. These results prove the value of EHG entropy analysis as a tool for early, prognostic evaluation of pregnancy.

REFERENCES


