Preterm-Term Birth Classification Using EMD-Based Time-Domain Features of Single-Channel Electrohysterogram Data

Suparerk Janjarasjitt (✉ suparerk.j@ubu.ac.th)
Ubon Ratchathani University  https://orcid.org/0000-0002-0252-8795

Research Article

Keywords: time-domain features, empirical mode decomposition, support vector machine, electrohysterogram, preterm birth

DOI: https://doi.org/10.21203/rs.3.rs-570938/v1

License: © This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Preterm-Term Birth Classification Using EMD-Based Time-Domain Features of Single-Channel Electrohysterogram Data

Suparerk Janjarasjitt

Abstract The preterm birth anticipation is a crucial task that can reduce the rate of preterm birth and also the complications of preterm birth. Electrohysterogram (EHG) or uterine electromyogram (EMG) data have been evidenced that they can provide an information useful for preterm birth anticipation. Four distinct time-domain features, i.e., mean absolute value, average amplitude change, difference absolute standard deviation value, and log detector, commonly applied to EMG signal processing are applied and investigated in this study. A single-channel of EHG data is decomposed into its constituent components, i.e., intrinsic mode functions, using empirical mode decomposition (EMD) before their time-domain features are extracted. The time-domain features of intrinsic mode functions of EHG data associated with preterm and term births are applied for preterm-term
birth classification using support vector machine (SVM) with a radial basis function. The preterm-term classifications are validated using 10-fold cross validation. From the computational results, it is shown that the excellent preterm-term birth classification can be achieved using a single-channel of EHG data. The computational results further suggest that the best overall performance on preterm-term birth classification is obtained when thirteen (out of sixteen) EMD-based time-domain features are applied. The best accuracy, sensitivity, specificity, and $F_1$-score achieved are, respectively, 0.9382, 0.9130, 0.9634, and 0.9366.

**Keywords** time-domain features · empirical mode decomposition · support vector machine · electrohysterogram · preterm birth

1 Introduction

Preterm birth or premature birth is defined by the World Health Organization (WHO) [1] as all births before 37 completed weeks of gestation [2]. Preterm birth is one of major health concerns as a preterm birth gives the baby less time to develop in the womb [3] and also a major cause of death [2]. Preterm birth can cause short-term complications and also long-term complications [3]. Solutions for reducing preterm birth include improving health care before, between and during pregnancies, providing effective treatments and interventions, and identifying women at risk for preterm delivery [4,5]. A preterm birth anticipation is therefore one of crucial tasks that useful in reducing preterm birth and its complications.

Electrohysterogram (EHG) or uterine electromyogram (EMG) measuring the electrical activity of uterine muscles [6] becomes a potential noninvasive tool for assessing and monitoring uterine activity [6–8]. There have been a number of com-
putational tools and techniques applied for classifying and predicting preterm birth as some of those summarized in Table 1. From the previous studies, the root mean square, the peak frequency, the median frequency, and the sample entropy are the most common quantitative features of EHG signals extracted and applied [6,9,10]. Time-domain features of EHG signals as well as the root mean square [11] are also investigated. Furthermore, Hemthanon et al. [12] examine the correlation between the time-domain features of EHG signals and gestational age. A pre-processing stage such as wavelet decomposition [10] and empirical mode decomposition [13], may be applied to EHG signals before quantitative features are extracted in a number of studies. Various classifiers including random forest [10,13], support vector machine (SVM) [13], and artificial neural network [13], are applied. In addition to machine learning techniques, basic statistical analyses such as t-test [6] and Wilcoxon rank sum test [14] are applied to validate the performance on preterm and term birth discrimination. Improved classification and discrimination results are obtained when oversampling techniques are also applied to enable a balanced number of preterm and term birth [10,13,9].

In this study, the capability of time-domain features of a single-channel EHG data on preterm-term birth classification is further examined as time-domain features have been common and useful quantitative features applied for various applications of EMG signal processing and analysis. Four time-domain features including the mean absolute value, the average amplitude change, the difference absolute standard deviation value, and the log detector are determined. Empirical mode decomposition (EMD) is applied to decompose a single-channel EHG data into a finite set of oscillatory components, called intrinsic mode functions (IMFs), associating with sequentially decreasing frequencies [15–17].
2 Methodology

2.1 Dataset

The electrohysterogram (EHG) dataset examined in this study is obtained from EHG recordings collected at the Department of Obstetrics and Gynecology, Medical Centre Ljubljana Ljubljana [6,18] from 1997 until 2006. Three hundred EHG recordings are publicly available at https://www.physionet.org/content/tpehgdb/ on the PhysioNet [18]. The EHG recordings were recorded from a general population of pregnant women during regular check-ups either around the 22nd week of gestation or around the 32nd week of gestation [6,18]. Three channels of EHG recordings obtained from four electrodes placed around the navel [6,18] are referred to as $s_1$, $s_2$, and $s_3$. The EHG data were acquired using the sampling frequency of 20 Hz.

The EHG dataset is classified into four classes, referred to as PE, PL, TE, and TL, corresponding to two subject groups and two periods of recordings summarized in Table 2. Those two subject groups are a group of preterm birth (P) which is associated with the delivery before or on the 37th week of gestation and a group of term birth (T) which is associated with the delivery was after the 37th of gestation [6,18]. Those two periods of recordings are an early period (E) which is associated with the recordings acquired before the 26th week of gestation and a late period (L) which is associated with the recordings acquired during or after the 26th week of gestation [6,18]. There are 19, 19, 143, and 119 of EHG recordings associated with the PE, PL, TE, and TL classes, respectively.
2.2 Data Processing

EHG signals corresponding to each channel of original EHG recordings are segmented with a length of 14400 samples, equivalently 720 seconds, from the end of EHG recordings. The EHG segments are further divided into 3600-sample epochs with 50% overlap. Accordingly, there are twelve EHG epochs for each EHG segment. EHG epochs are dissected and decomposed into four intrinsic mode functions (IMFs) using the empirical mode decomposition (EMD) [15,17,19]. Time-domain features which are commonly applied in EMG signal processing and analysis are subsequently extracted from the intrinsic mode functions of EHG epochs. The mean absolute value (MAV), the average amplitude change (AAC), the difference absolute standard deviation value (DASDV), and the log detector are the four distinct time-domain features applied for preterm-term birth classification.

Denote a sequence of intrinsic mode functions $x[n]$ with a length of $N$ applied for time-domain feature extraction. The time-domain features of each EHG epoch, respectively referred to as $F_1$, $F_2$, $F_3$, and $F_4$, are given as follows

- Mean absolute value (MAV):
  \[
  F_1 = \frac{1}{N} \sum_{n=0}^{N-1} |x[n]|. 
  \]  

- Average amplitude change (AAC):
  \[
  F_2 = \frac{1}{N} \sum_{n=0}^{N-1} |x[n+1] - x[n]|. 
  \]  

- Difference absolute standard deviation value (DASDV):
  \[
  F_3 = \sqrt{\frac{1}{N-1} \sum_{n=0}^{N-1} (x[n+1] - x[n])^2}. 
  \]  

- Log detector:
  \[
  F_4 = e^{\frac{1}{N} \sum_{n=0}^{N-1} \log(|x[n]|)}. 
  \]
2.3 Classification and Evaluation

A feature vector applied for preterm-term birth classification is composed of time-domain features of intrinsic mode functions of EHG epochs. A number of time-domain features used to form the feature vector ranges from 2 to 16, i.e., all of the time-domain features. The time-domain features of intrinsic mode functions of EHG epochs selected and applied correspond to their ranking scores obtained from the \( p \)-value of \( \chi^2 \) test, i.e., \( \log(p) \). There are two preterm-term birth classifications examined in this study: 1) PE-TE classification and 2) PL-TL classification. Support vector machines (SVMs) are applied for the preterm-term birth classifications. In the training of SVM models, the kernel function used is the radial basis function (RBF) with the standardization applied.

For the evaluation of preterm-term birth classifications, the 10-fold cross-validation is applied. The performance of preterm birth classifications is evaluated using four conventional classification performance measures including accuracy (Ac), sensitivity (Se), specificity (Sp), precision, recall, and \( F_1 \)-score, which are, respectively, given by

\[
Ac = \frac{TP + TN}{TP + TN + FP + FN}, \tag{5}
\]

\[
Se = \frac{TP}{TP + FN}, \tag{6}
\]

\[
Sp = \frac{TN}{TN + FP}, \tag{7}
\]

\[
F_1\text{-score} = \frac{TP}{TP + \frac{1}{2}(FP + FN)}, \tag{8}
\]

where \( TP, TN, FP, \) and \( FN \) denote a number of true positives, a number of true negatives, a number of false positives, and a number of false negatives, respectively.
3 Results

3.1 Preterm-Term Birth Classifications Using Early Period EHG Data

The accuracy, sensitivity, specificity, and $F_1$-score of the preterm-term birth classification obtained using the channels, $s_1$, $s_2$, and $s_3$, of early period EHG data are shown in Fig. 1(a)–(d), Fig. 2(a)–(d), and Fig. 3(a)–(d), respectively. The accuracy, sensitivity, specificity, and $F_1$-score of the preterm-birth classification range, respectively, from 0.7002 to 0.9382, 0.5515 to 0.9153, 0.6774 to 0.9634, and 0.6479 to 0.93662. The performance of the preterm-term birth classification corresponding to all metrics tend to be boosted as the number of time-domain features of intrinsic mode functions of EHG epochs applied increases.

The best performances obtained from the preterm-term birth classifications using the time-domain features of intrinsic mode functions of early period EHG data corresponding to each performance metrics are summarized in Table 3. All of performance metrics, i.e., the accuracy, sensitivity, specificity, and $F_1$-score, of the preterm-term birth classification obtained using the channel $s_1$ of early period EHG data are completely higher than those of the preterm-term birth classification using either channel $s_2$ or $s_3$. The best performance of the preterm-term birth classifications is generally achieved using thirteen time-domain features of intrinsic mode functions of channel $s_1$ of early period EHG data in which the best accuracy, specificity, and $F_1$-score are obtained. The accuracy, sensitivity, specificity, and $F_1$-score of the preterm-term birth classification using thirteen time-domain features of intrinsic mode functions of channel $s_1$ of early period EHG data are 0.9382, 0.9130, 0.9634, and 0.9366, respectively.
3.2 Preterm-Term Birth Classifications Using Late Period EHG Data

The accuracy, sensitivity, specificity, and $F_1$-score of the preterm-term birth classification obtained using the channels, $s_1$, $s_2$, and $s_3$, of late period EHG data are, respectively, shown in Fig. 4(a)–(d), Fig. 5(a)–(d), and Fig. 6(a)–(d). The accuracy, sensitivity, specificity, and $F_1$-score of the preterm-birth classification range, respectively, from 0.6842 to 0.8970, 0.5858 to 0.9176, 0.6728 to 0.8993, and 0.6498 to 0.8980. The trend of increasing accuracy, sensitivity, specificity, and $F_1$-score of the preterm-term birth classifications as the number of time-domain features of intrinsic mode functions of EHG epochs applied increases is obviously evidenced in the channels $s_1$ and $s_3$.

Table 4 summarizes the best performances obtained from the preterm-term birth classifications using the time-domain features of intrinsic mode functions of late period of EHG data corresponding to each performance metrics. Likewise, the higher values of accuracy, sensitivity, specificity, and $F_1$-score, of the preterm-term birth classifications using late period EHG data are obtained in the channel $s_1$ compared to the channels $s_2$ and $s_3$. The best accuracy and $F_1$-score of the preterm-term birth classifications are obtained using thirteen time-domain features of intrinsic mode functions of channel $s_1$ of late period EHG data. The accuracy, sensitivity, specificity, and $F_1$-score of the preterm-term birth classification using thirteen time-domain features of intrinsic mode functions of channel $s_1$ of late period EHG data are, respectively, 0.8970, 0.9062, 0.8879, and 0.8980.
4 Discussion

From the computational results, it is shown that the performance on preterm-term birth classification using time-domain features is significantly improved when the empirical mode decomposition (EMD) is applied to decompose a single-channel of EHG data into intrinsic mode functions. Only four time-domain features, i.e., the mean absolute value, the average amplitude change, the difference absolute standard deviation value, and the log detector, are applied compared to seventeen time-domain features of original (full-bandwidth) and filtered EHG data examined in the previous study [11]. The product of sensitivity and specificity of preterm-term birth classification is increased from 0.5134 to 0.8796.

The better performance on preterm-term birth classification is achieved using the channel $s_1$ of EHG data which records the difference of electrical activity of uterine muscles above the navel [6]. The number of time-domain features of intrinsic mode functions of single-channel EHG data providing the best performance on preterm-term birth classifications is 13. The accuracy, sensitivity, specificity, and $F_1$-score of preterm-term birth classification using thirteen time-domain features of intrinsic mode functions of single-channel EHG data acquired in the early period are, respectively, 0.9382, 0.9130, 0.9634, and 0.9366 while the accuracy, sensitivity, specificity, and $F_1$-score of preterm-term birth classification using thirteen time-domain features of intrinsic mode functions of single-channel EHG data acquired in the late period are 0.8970, 0.9062, 0.8879, and 0.8980, respectively. Interestingly, the better performance can be obtained from the preterm-term birth classification using single-channel EHG data acquired in the early period, i.e., before the 26th
week of gestation compared those acquired in the late period, i.e., during or after the 26th week of gestation.

The performance on preterm-term birth classification and discrimination of notable previous studies summarized in Table 1 are compared in Table 5. Regarding the previous studies [9,10] examining and reporting comparable performance metrics, the performance on preterm-term birth classification obtained using the time-domain features of intrinsic mode functions of single-channel of EHG data is shown to be generally superior to the best accuracy, sensitivity, and specificity of 0.92 [10], 0.92 [9], and 0.96 [10]. All four time-domain features applied in this study therefore reveal distinguishing characteristics of EHG data which are useful for preterm-term birth classification.

5 Conclusion

Four distinct time-domain features, i.e., the mean absolute value, the average amplitude change, the difference absolute standard deviation value, and the log detector, commonly applied for EMG signal processing and analysis are extracted from the first four intrinsic mode functions of single-channel of EHG data. The computational results suggest that the number of time-domain features of intrinsic mode functions of EHG data required for preterm-term birth classification is 13 (out of 16). The accuracy, sensitivity, specificity, and $F_1$-score of preterm-term birth classification using thirteen time-domain features of intrinsic mode functions of channel $s_1$ of EHG data acquired in the early period are 0.9382, 0.9130, 0.9634, and 0.9366, respectively.
Compliance with Ethical Standards

Funding: This study was funded by a TRF Research Career Development Grant, jointly funded by Thailand Research Fund (TRF) and Ubon Ratchathani University (contract number RSA6180041).

Conflict of Interest: The author declares that he has no conflict of interest.

References

1. WHO. WHO: Recommended definitions, terminology and format for statistical tables related to the perinatal period and use of a new certificate for cause of perinatal deaths, volume = 56, pages = 247-253, year = 1976, journal = Acta Obstetricia et Gynecologica Scandinavica;.


Table 1 Previous studies on preterm birth classification and discrimination using EHG data

<table>
<thead>
<tr>
<th>Ref.</th>
<th>First author surname</th>
<th>Description</th>
</tr>
</thead>
</table>
| [6]  | Fele-Zorž et al.     | Features: root mean square; peak frequency; median frequency; autocorrelation zero-crossing; Lyapunov exponent; correlation dimension; sample entropy  
Statistical analysis: t-test |
| [9]  | Fergus et al.        | Features: root mean square; peak frequency; median frequency; sample entropy  
Dimensionality reduction: principal component analysis  
Classifier: linear discriminant classifier; quadratic discriminant classifier; uncorrelated normal density based classifier; polynomial classifier; logistic classifier; 3-nearest neighbor; decision tree; parzen classifier; support vector classifier  
Oversampling technique: SMOTE |
| [10] | Peng et al.          | Features: root mean square; autocorrelation zero-crossing; peak frequency; median frequency; wavelet-based features; time reversability; Lyapunov exponent; sample entropy; correlation dimension  
Classifier: random forest  
Oversampling technique: adaptive synthetic sampling |
Features: Shannon entropy  
Classifier: support vector machine; random forest; artificial neural network; AdaBoost; Bayesian network  
Oversampling technique: SMOTE |
| [14] | Mas-Cabo et al.      | Features: root mean square; dominant frequency; decile of power spectrum; spectral moment ratio  
Statistical analysis: Wilcoxon rank sum test |
| [11] | Hemthanon et al.     | Features: integrated EMG; mean absolute value; modified mean absolute value; simple square integral; absolute value of temporal moments; root mean square; ν-order; log detector; waveform length; average amplitude change; difference absolute standard deviation value; zero crossing; slope sign change  
Feature selection: Pearson’s correlation coefficient; t-test  
Classifier: support vector machine |
<table>
<thead>
<tr>
<th>Class</th>
<th>Subject group</th>
<th>Period of recordings</th>
<th>Number of recordings</th>
</tr>
</thead>
<tbody>
<tr>
<td>PE</td>
<td>Preterm birth (P)</td>
<td>Early period (E)</td>
<td>19</td>
</tr>
<tr>
<td>PL</td>
<td>Preterm birth (P)</td>
<td>Late period (L)</td>
<td>19</td>
</tr>
<tr>
<td>TE</td>
<td>Term birth (T)</td>
<td>Early period (E)</td>
<td>143</td>
</tr>
<tr>
<td>TL</td>
<td>Term birth (T)</td>
<td>Late period (L)</td>
<td>119</td>
</tr>
</tbody>
</table>
### Table 3  Best Performance on Preterm-Term Birth Classification Using Early Period EHG Data

<table>
<thead>
<tr>
<th>Channel</th>
<th>Metric</th>
<th>Best performance</th>
<th>No. of features</th>
</tr>
</thead>
<tbody>
<tr>
<td>s₁</td>
<td>Ac</td>
<td>0.9382</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Se</td>
<td>0.9153</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Sp</td>
<td>0.9634</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>F₁-Score</td>
<td>0.9366</td>
<td>13</td>
</tr>
<tr>
<td>s₂</td>
<td>Ac</td>
<td>0.8936</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Se</td>
<td>0.8673</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Sp</td>
<td>0.9245</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>F₁-Score</td>
<td>0.8907</td>
<td>16</td>
</tr>
<tr>
<td>s₃</td>
<td>Ac</td>
<td>0.8513</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Se</td>
<td>0.8101</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Sp</td>
<td>0.9199</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>F₁-Score</td>
<td>0.8449</td>
<td>10</td>
</tr>
</tbody>
</table>
### Table 4  Best Performance on Preterm-Term Birth Classification Using Late Period EHG Data

<table>
<thead>
<tr>
<th>Channel</th>
<th>Metric</th>
<th>Best performance</th>
<th>No. of features</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_1$</td>
<td>Ac</td>
<td>0.8970</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Se</td>
<td>0.9176</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Sp</td>
<td>0.8993</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>$F_1$-Score</td>
<td>0.8980</td>
<td>13</td>
</tr>
<tr>
<td>$s_2$</td>
<td>Ac</td>
<td>0.8387</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Se</td>
<td>0.8673</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Sp</td>
<td>0.8650</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>$F_1$-Score</td>
<td>0.8366</td>
<td>15</td>
</tr>
<tr>
<td>$s_3$</td>
<td>Ac</td>
<td>0.8387</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Se</td>
<td>0.8787</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Sp</td>
<td>0.8192</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>$F_1$-Score</td>
<td>0.8418</td>
<td>10</td>
</tr>
<tr>
<td>Ref.</td>
<td>First author surname</td>
<td>Best performance</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>----------------------</td>
<td>------------------</td>
<td></td>
</tr>
<tr>
<td>[9]</td>
<td>Fergus et al.</td>
<td>Area under curve: 0.89</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sensitivity: 0.92</td>
<td></td>
</tr>
<tr>
<td>[10]</td>
<td>Peng et al.</td>
<td>Accuracy: 0.92</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sensitivity: 0.88</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Specificity: 0.96</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Area under curve: 0.88</td>
<td></td>
</tr>
<tr>
<td>[13]</td>
<td>Ren et al.</td>
<td>Area under curve: 0.986</td>
<td></td>
</tr>
<tr>
<td>[14]</td>
<td>Mas-Cabo et al.</td>
<td>$p$-value: &lt; 0.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Current study</td>
<td>Accuracy: 0.9382</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Janjarasjitt</td>
<td>Sensitivity: 0.9130</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Specificity: 0.9634</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$F_1$-score: 0.9366</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Product of sensitivity and specificity: 0.8796</td>
<td></td>
</tr>
</tbody>
</table>
Fig. 1 Performance of the preterm-term birth classification obtained using the channel $s_1$ of early period EHG data.
Fig. 2 Performance of the preterm-term birth classification obtained using the channel $s_2$ of early period EHG data.
Fig. 3 Performance of the preterm-term birth classification obtained using the channel s3 of early period EHG data.
Fig. 4 Performance of the preterm-term birth classification obtained using the channel $s_1$ of late period EHG data.
Preterm-Term Birth Classification Using EMD-Based Time-Domain Features

![Graphs showing performance metrics](image)

(a) Accuracy  
(b) Sensitivity  
(c) Specificity  
(d) $F_1$-score

**Fig. 5** Performance of the preterm-term birth classification obtained using the channel $s_2$ of late period EHG data.
Fig. 6 Performance of the preterm-term birth classification obtained using the channel s3 of late period EHG data.
Figures

Figure 1

Performance of the preterm-term birth classification obtained using the channel s1 of early period EHG data.
Figure 2

Performance of the preterm-term birth classification obtained using the channel s2 of early period EHG data.
Figure 3

Performance of the preterm-term birth classification obtained using the channel s3 of early period EHG data.
Figure 4

Performance of the preterm-term birth classification obtained using the channel s1 of late period EHG data.
Figure 5

Performance of the preterm-term birth classification obtained using the channel s2 of late period EHG data.
Figure 6

Performance of the preterm-term birth classification obtained using the channel s3 of late period EHG data.