

Research Article

Prediction of Labor Induction Success from the Uterine Electrohysterogram

Carlos Benalcazar-Parra,¹ Yiyao Ye-Lin,¹ Javier Garcia-Casado ,¹ Rogelio Monfort-Ortiz,² Jose Alberola-Rubio,² Alfredo Perales,^{2,3} and Gema Prats-Boluda ¹

¹Centro de Investigación e Innovación en Bioingeniería, Universitat Politècnica de València, Edif. 8B, Camino de Vera SN, 46022 Valencia, Spain

²Servicio de Obstetricia y Ginecología, Hospital Universitario y Politécnico La Fe de Valencia, Av. Fernando Abril Martorell 106, Edificio F, 3ª Planta, Valencia, Spain

³Departamento de Pediatría, Obstetricia y Ginecología Universidad Valencia, Av Blasco Ibañez 15, 46010 Valencia, Spain

Correspondence should be addressed to Gema Prats-Boluda; gprats@ci2b.upv.es

Received 27 March 2019; Revised 19 July 2019; Accepted 8 October 2019; Published 15 November 2019

Guest Editor: Lourdes Martínez-Villaseñor

Copyright © 2019 Carlos Benalcazar-Parra et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Pharmacological agents are often used to induce labor. Failed inductions are associated with unnecessarily long waits and greater maternal-fetal risks, as well as higher costs. No reliable models are currently able to predict the induction outcome from common obstetric data (area under the ROC curve (AUC) between 0.6 and 0.7). The aim of this study was to design an early success-predictor system by extracting temporal, spectral, and complexity parameters from the uterine electromyogram (electrohysterogram (EHG)). Different types of feature sets were used to design and train artificial neural networks: Set_1: obstetrical features, Set_2: EHG features, and Set_3: EHG+obstetrical features. Predictor systems were built to classify three scenarios: (1) induced women who reached active phase of labor (APL) vs. women who did not achieve APL (non-APL), (2) APL and vaginal delivery vs. APL and cesarean section delivery, and (3) vaginal vs. cesarean delivery. For Scenario 3, we also proposed 2-step predictor systems consisting of the cascading predictor systems from Scenarios 1 and 2. EHG features outperformed traditional obstetrical features in all the scenarios. Little improvement was obtained by combining them (Set_3). The results show that the EHG can potentially be used to predict successful labor induction and outperforms the traditional obstetric features. Clinical use of this prediction system would help to improve maternal-fetal well-being and optimize hospital resources.

1. Introduction

The induction of labor consists of promoting uterine contractions and cervical ripening before the onset of spontaneous labor. This common procedure is indicated when continuing pregnancy increases maternal and/or fetal risks. In the United States, 22.8% of all births were induced in 2012 [1]. Pharmacological labor induction is mainly obtained by prostaglandins [2] but can take up to 20 hours [3] and has been known to take more than 36 hours, with no guarantee of success. It has also been associated with maternal and fetal risks such as abnormal uterine activity, fetal distress, and higher cesarean rates [4]. Failed inductions lead to unneces-

sary waits, greater maternal-fetal exhaustion and suffering, and the need for additional resources, thus increasing medical care costs. Predicting successful induction is an important aspect in improving maternal and fetal well-being, reducing healthcare costs and improving labor management.

Obstetric variables have been considered for this purpose and are usually based on cervix assessment by the Bishop score [5, 6], although cervical length, maternal age, height, weight, parity, and birth weight [7–9] have also been used. The predictive capacity values given by the area under the curve (AUC) of the receiver operating characteristic (ROC) curves are 0.69 for cervical length [7], 0.72 for cervical dilatation [7], 0.52 for Bishop score [6], and 0.60 for fetal weight

[8], showing that obstetrical data cannot at present be used to reliably predict induction of labor.

The electrohysterogram (EHG), i.e., uterine myoelectric activity recorded on the abdominal surface, is an alternative method of monitoring uterine dynamics and consists of intermittent bursts of action potentials derived from the simultaneous activation of multiple uterine muscle cells. Uterine myoelectric activity evolves throughout gestation, being scarce and uncoordinated in the early stages, and becomes intense and synchronized as delivery approaches [10]. Previous studies have shown that EHG signals can discriminate effective contractions associated with imminence of labor [11] or whether delivery will be term or preterm [12]. EHG records have also been used to characterize the uterine myoelectrical response to labor induction drugs [13–16]. Aviram et al. found that uterine electrical activity significantly increases 2 hours after prostaglandin E2 (PGE2) vaginal application and up to 8 hours after PGE2 application [13]. However, their aim was not to predict labor induction success or to compare the responses between successful and failed groups. Toth studied the possibility of predicting induction success using local prostaglandin [14]. They assessed uterine activity by means of an index that takes the intrinsic characteristics of EHG bursts into account (number of impulses, amplitudes, series, and shape) and found a statistically significant difference in the uterine activity index between successful (vaginally completed) and unsuccessful inductions between the 210th and 270th minutes. Benalcazar-Parra et al. also studied the differences between failed and successful (reaching the active phase of labor (APL)) inductions by comparing the evolution of different EHG parameters. They found different responses, mainly in amplitude and spectral parameters after 60'–120' from labor induction onset [15, 16]. However, to date, no work has been done on predicting successful induction from EHG records, while EHG-based neural networks have been applied to the prediction of term and preterm labor [12, 17–19]. In this context, the aim of the present study was to design a system capable of reliably predicting successful labor induction, based on EHG features and obstetrical data in the first 4 hours after labor induction onset.

Vaginal delivery can be considered a 2-step process. First, the woman has to reach the APL, i.e., regular uterine dynamic with 3–5 contractions every 10 minutes, 4 cm of cervical dilatation, and cervical effacement [20]. This is a necessary condition to be able to expel the fetus outside the uterus via the vaginal route (Step 2). It should be noted that although there is some controversy as regards establishing the value of the cervical dilatation and cervical effacement associated with APL, in the present work, we considered 4 cm, being the most widely extended definition [21]. A cesarean is needed if the APL cannot be reached. However, even if APL has been reached, various conditions may prevent vaginal delivery, such as labor arrest, pelvic-fetal disproportion, or loss of maternal-fetal well-being [22]. In the labor induction context and from the pharmacologic point of view, induction can be considered successful if drug action helps the patients achieve APL [15, 16, 23]. From the medical point of view, only vaginal deliveries are commonly considered

successful [24, 25]. Taking this into account, we considered three different scenarios in designing and validating prediction systems for labor induction success (see Figure 1).

2. Materials and Methods

2.1. Signal Acquisition. The study was conducted on 115 healthy pregnant women with gestational ages of between 40 and 41 weeks and singleton pregnancies who were determined to undergo labor induction by medical prescription. The distribution of the labor outcome population is shown in Figure 1 according to the different scenarios:

- (i) Scenario 1: women achieving active phase of labor (successful group; $N = 98$) vs. women nonachieving active phase of labor (failed group; $N = 17$)
- (ii) Scenario 2: from women who achieved active phase of labor, those achieving vaginal delivery (successful group; $N = 82$) vs. cesarean section (failed group; $N = 16$)
- (iii) Scenario 3: women achieving vaginal delivery (successful group; $N = 82$) vs. cesarean deliveries (failed group; $N = 33$)

The recordings were performed at the *Hospital Universitario y Politécnico La Fe de Valencia* (Spain), and the study was approved by the Hospital Ethics Committee (2015/0455, 12/01/2016). The women were previously informed of the nature of the study and gave their written consent. Labor induction was by vaginal administration of two different types of drugs commonly used in obstetrics: either a vaginal insert of 25 μg of misoprostol tablets (Misofar, Bial S.A., Portugal) with repeated doses every 4 hours up to a maximum of 3 doses or 10 mg of vaginal dinoprostone insert (Propess, Ferring, Germany). The women were kept under constant observation until the end of labor. The women's obstetrical characteristics and labor induction outcomes are shown in Table 1.

TOCO and EHG signals were simultaneously acquired by tocodynamometer and four monopolar disposable Ag/AgCl electrodes (3 M red dot 2560), respectively, in the recording sessions, which comprised 30 minutes of basal activity (before drug administration) and 4 hours of recording after drug administration. The abdominal surface was first exfoliated (Nuprep, Weaver and Company, USA) to reduce skin-electrode impedance. The monopolar electrodes (M1 and M2) were placed over the navel at each side of the median axis at a distance of 8 cm from each other, which has been found to be the optimal electrode placement in the literature [26]. A reference electrode was placed on the right hip and a ground electrode on the left hip (Figure 2). Monopolar EHG signals were amplified and filtered between 0.1 and 30 Hz by a commercial biosignal amplifier (Grass 15LT+4 Grass 15A94; Grass Instruments, West Warwick, RI) and digitalized at a sampling frequency of 1000 Hz. Since EHG signal energy principally ranges from 0.1 to 4 Hz, the signal was digitally filtered between 0.2 and 4 Hz to eliminate undesired components and then downsampled at 20 Hz to reduce the

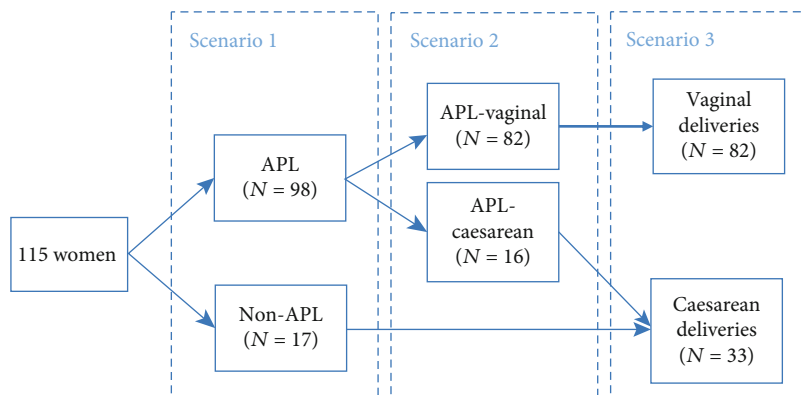


FIGURE 1: Study population and group distribution in each scenario.

TABLE 1: Women's obstetrical parameters and labor induction outcome; mean \pm std.

Obstetric variables	Mean \pm std
Maternal age	32.5 \pm 4.7
BMI (kg/m ²)	26.0 \pm 9.4
Gestations	1.4 \pm 0.6
Parity	
0	106/115 (92%)
1	8/115 (7%)
2	1/115 (1%)
Abortions	
0	94/115 (82%)
1	18/115 (16%)
2	3/115 (2%)
Bishop	2.0 \pm 1.2
Fetal weight (g)	3411.9 \pm 381.2
Active phase of labor	98/115 (85%)
Vaginal delivery	82/115 (71%)
Cesareans	33/115 (29%)

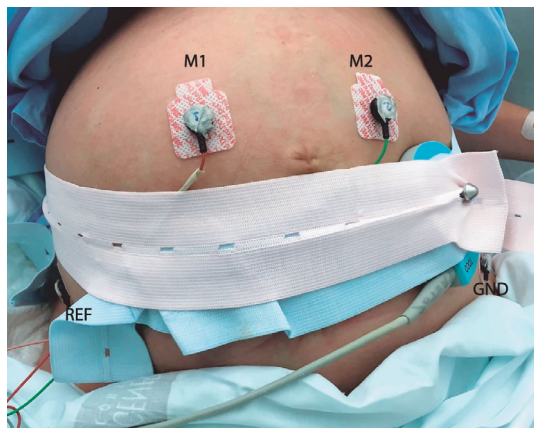


FIGURE 2: Photograph that illustrates TOCO probe and the surface electrode arrangement.

amount of data and the computational cost, obtaining the preprocessed M1P and M2P signals. One bipolar EHG signal was then obtained (M1P-M2P) to further reduce common-mode interference. The TOCO signal was recorded by a Corometrics 250cx (General Electric Healthcare, US) commercial maternal monitor at a sampling rate of 4 Hz. All EHG bursts associated with uterine contractions were identified by visual inspection of the bipolar EHG signal using the same criteria as in Benalcazar-Parra et al. [15].

2.2. EHG Signal Characterization. Several studies have shown that the temporal and spectral parameters obtained from EHG recordings change between pregnancy and labor onset [11]. It has been reported that temporal parameters such as amplitude, duration, and number of contractions (EHG bursts) change during pregnancy [27, 28]. As with spectral features, parameters such as peak frequency, mean frequency, and deciles, among others, have been extracted from the power spectral density to characterize EHG burst frequency components [27, 29–31]. In this regard, it is worth mentioning that EHG bursts are mainly composed of two distinct frequency components: fast wave low (FWL), a low frequency component associated with EHG propagation, and fast wave high (FWH), a high frequency component related to uterine cell excitability [32]. It is well known that both components are mainly distributed between 0.2 and 1 Hz [32], although some authors consider that it can extend up to 4 Hz [33]. However, some studies focus only on the FWH, restricting the bandwidth between 0.34 and 1 Hz to minimize breathing and cardiac interference [30]. It has also been shown that EHG burst spectral content shifts to higher frequencies, in the range of 0.34 to 1 Hz as labor approaches [34]. Furthermore, considering the nonlinear nature of the underlying mechanisms of the biological systems, parameters such as sample entropy, spectral entropy, and Lempel-Ziv have also been proposed to characterize EHG signals [33, 35].

Therefore, in the present work, 21 temporal, spectral, and complexity parameters were computed from each EHG burst (see Table 2). Peak-to-peak amplitude was computed from the temporal series associated with uterine contractions. The following parameters were extracted from the power spectral density distribution estimated by the periodogram method: dominant frequency in the range of 0.2-1 Hz (DF),

TABLE 2: Summary of the extracted parameters that will be used to design the classifiers.

EHG temporal parameters	EHG spectral parameters	EHG complexity parameters	Obstetric parameters
Peak to peak amplitude	DF, H/L ratio, Deciles [D1-D9], Teager energy	SampEn, LZ, SpEn, FuzzEn, SD1, SD2, SDRR, SD1/SD2	Maternal age, body mass index (BMI), gestations, parity, abortions, Bishop, fetal weight

ratio between the energy contents in high (0.34-1 Hz) and low (0.2-0.34 Hz) frequency bands (H/L ratio), and deciles (D1, D2, ..., D9), which correspond to frequencies below in which 10, 20, ..., 90%, respectively, of the total energy in the range 0.2-1 Hz are contained [36]. The Teager energy operator was computed to measure the energy of the EHG burst. This measure takes into account not only the amplitude but also the frequency of the signal [37].

As previously mentioned, due to the nonlinear nature of the underlying physiological mechanism of the biological systems, a set of 8 nonlinear parameters was computed for each EHG burst, where some of them were already used to characterize EHG signals: sample entropy (SampEn) has been used to discriminate between preterm and term labor and to assess the progress of labor [33], and the Lempel-Ziv (LZ) parameter has been used to distinguish between patients who give birth in less/more than 7 days [38]. We also computed some complexity parameters that have been used in other applications. Fuzzy entropy (FuzzEn) has been shown to be efficient at measuring the regularity of time series in surface EMG signals [39]. Spectral entropy (SpEn) has also given good results in monitoring the depth of anesthesia [40] and predicting epileptic seizures [41]. Poincare parameters (SD1, SD2, SDRR, and SD1/SD2) have been widely used for heart rate variability analysis [42] and have been claimed to be valuable for their ability to extract the nonlinear characteristics of time series [43].

In a previous work, to analyze the evolution of the EHG burst parameters in response to labor induction drugs, we first computed the median values of each parameter associated with the EHG bursts present in nonoverlapping intervals of 30 minutes [15, 16]. Results showed that for successful inductions, statistically significant and sustained increases with respect to the basal period were obtained after 60 minutes and 120 minutes in patients induced with misoprostol and dinoprostone, respectively [15, 16]. This is the reason why, in the present work, in order to use only the significant intervals for both drugs, for each parameter, we analyzed 5 intervals of 30 minutes (basal period—before drug administration: 120', 150', 180', and 210'), giving rise to a total of $21 \times 5 = 105$ EHG features.

Additionally, we considered the following obstetric parameters that have been used in the literature [5–9]: maternal age, body mass index (BMI), number of gestations, parity, number of abortions, Bishop before drug administration, and fetal weight.

Then, for the inputs to the different labor induction success predictor systems developed, the parameters were grouped into three sets: Set_1—containing only obstetrical features, Set_2—containing only EHG features, and Set_3—containing both EHG and obstetrical features.

2.3. Data Balancing. The disadvantage of imbalanced datasets is that classification learning algorithms are often biased towards the majority class, so that there is a higher misclassification rate for the minority class instances. The synthetic minority oversampling technique (SMOTE) was used in this study to deal with the unbalanced data problem. SMOTE is an oversampling approach proposed by Chawla et al. [44] and consists of increasing the number of observations of the minority class in the original dataset by creating new synthetic observations. SMOTE is an accepted technique for dealing with the unbalanced problem and has been used in several studies (e.g. [12, 45]).

Nine databases (3 scenarios \times 3 feature sets) were generated (see Table 3) using SMOTE to balance the number of observations of each class in every database.

2.4. Feature Selection. In order to use only relevant data and avoid redundant information, particle swarm optimization (PSO) was used for feature selection. PSO is a population-based stochastic optimization technique that is based on the social behavior of flocking birds or schooling fish developed by Eberhart and Kennedy [46]. PSO is an iterative algorithm that consists of a number of particles (the swarm) moving around in the search space in order to achieve the best solution. A particle representing a candidate solution moves to the optimal position by updating its position and velocity.

PSO was adapted for feature selection as shown in Figure 3. The algorithm starts from a training set to select a subset of relevant features with PSO (the winning particle). A reduced training set and a reduced validation set are obtained by removing the features that are not selected. An artificial neural network for classification is trained with the reduced training set and then applied to the reduced validation set to obtain the final validation classification accuracy. The algorithm is run iteratively k times from $k = 1$ to $k =$ number of original features (7 for Set_1, 21 for Set_2, and 28 for Set_3). Then, the subset of k features with the lowest accuracy error is chosen. The algorithm was computed for each database to reduce the dimensionality.

2.5. Classifiers. Artificial neural networks (ANN) have been used to classify term and preterm deliveries [12, 17]. In the present study, we used the multilayer perceptron network which is a unidirectional network with one input layer, one output layer, and a certain number of hidden layers. The hyperbolic tangent function was used as the transfer function of each neuron. After selecting the optimal structure, for each scenario and set of features, we obtained a total of nine predictor systems (PS) based on ANN ($PS_{\text{SCENARIO-SET}}$: PS1_1, PS1_2, ..., PS3_3). For each $PS_{\text{SCENARIO-SET}}$, the corresponding $DB_{\text{SCENARIO-SET}}$ database was used for training

TABLE 3: Databases used to build the different predictor systems for each scenario (SMOTE balanced) and each feature set.

	Set_1 (7 Obst. features)	Set_2 (21EHG features)	Set_3 (7 Obst. 21 EHG features)
Scenario 1 (164 observations)	DB1_1	DB1_2	DB1_3
Scenario 2 (196 observations)	DB2_1	DB2_2	DB2_3
Scenario 3 (164 observations)	DB3_1	DB3_2	DB3_3

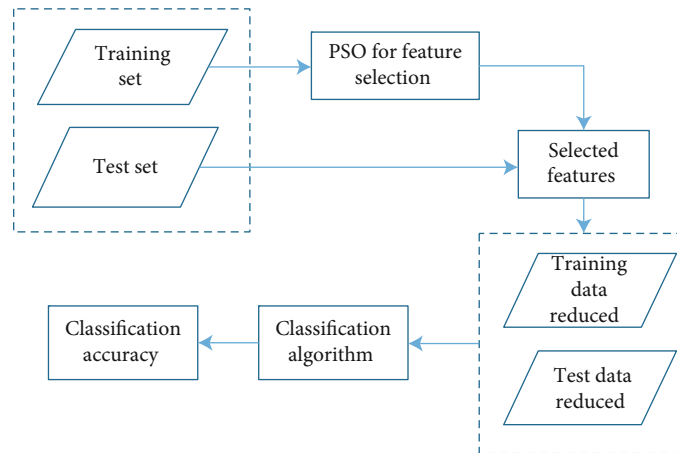


FIGURE 3: Diagram of the particle swarm optimization method for feature selection.

and validation (five-fold cross-validation). Figure 4 shows the scheme of each of the predictor systems.

In order to choose the optimal structure for each predictor system, we performed a grid search to select the number of hidden layers and hidden neurons. The rules in the grid search were as follows: maximum 2 hidden layers and maximum 10 hidden neurons in the first hidden layer. In addition, the number of neurons in the second hidden layer must not exceed the number of neurons of the first hidden layer, thus yielding a pyramidal structure with 2 hidden layers, which ensures optimal learning for multilayer networks [47]. In each scenario, we trained 165 ANN (55 ANN \times 3 feature sets). The best structure was selected from the 55 ANN of each case, measuring the average performance of each ANN from the validation set in a five-fold cross-validation. The implementation of the proposed algorithms to obtain the nine optimal predictor systems is shown in Figure 5.

Considering that vaginal delivery (Scenario 3) is a 2-step process, a fourth classifier was generated by cascading the predictor systems of Scenario 1 and Scenario 2 (PS1_{SET}-PS2_{SET}). The first system (PS1_{SET}) separates patients who achieve APL from those who fail to do so (non-APL) when using a particular set of features. Women classified as non-APL are directly classified as cesarean deliveries, while those who achieve APL are subclassified by a second system trained with the same set of features (PS2_{SET}). To evaluate this 2-step predictor system, the same validation partitions of the corresponding one-step predictor systems (DB3_{SET}) were used to compare the results between both approaches;

i.e., validation partitions from DB3_1 were used to evaluate PS1_1-PS2_1, from DB3_2 to evaluate PS1_2-PS2_2, and from DB3_3 to evaluate PS1_3-PS2_3.

2.6. Performance Measures. We validated the performance of each classifier by five-fold cross-validation. The following measures were calculated to evaluate classification performance:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \quad (1)$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (2)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}, \quad (3)$$

where TP represents the true positives, TN represents the true negatives, FP represents the false positives, and FN represents the false negatives. The area under the ROC curves (AUC) was computed for each PS_{SCENARIO-SET}.

3. Results

A total of 115 women with singleton pregnancies took part in the study. Their obstetric characteristics and labor induction outcome are summarized in Table 1. 98 women reached the active phase of labor, and 82 reached vaginal delivery. 33 ended up with a C-section: those who did not reach APL

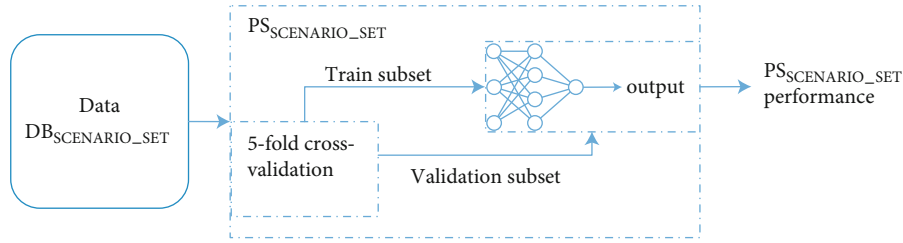


FIGURE 4: Scheme of the predictor systems ($PS_{SCENARIO_SET}$) obtained for each database ($DB_{SCENARIO_SET}$).

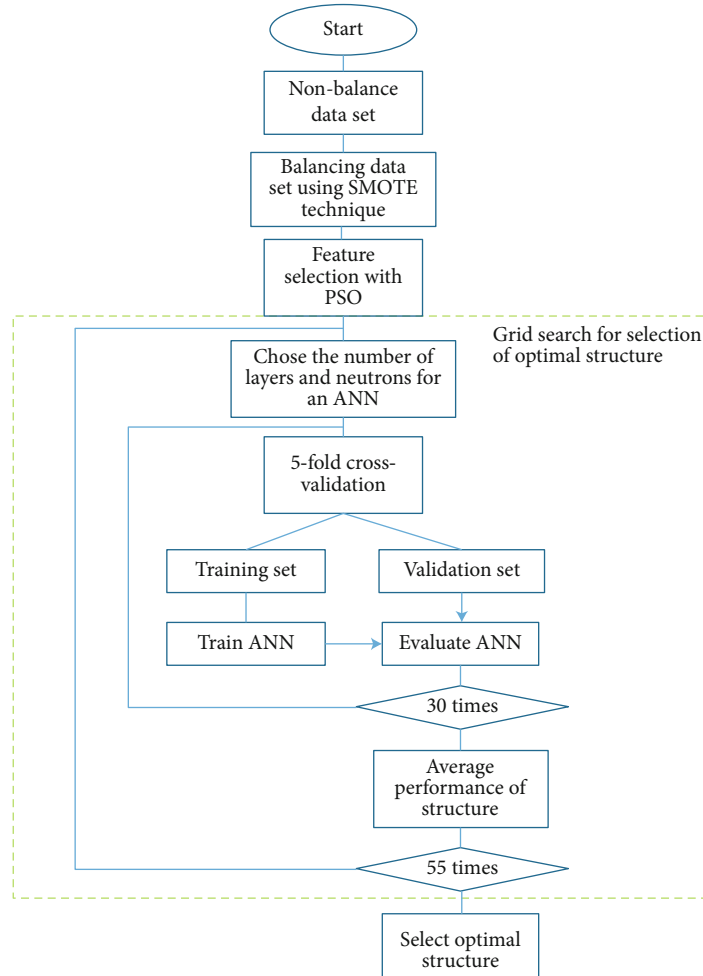


FIGURE 5: Diagram of the proposed algorithm to obtain the optimal predictor systems.

and some who did but were given a caesarian due to labor progression complications.

The mean and 95% confidence interval (CI) of the performance measures of the training and validation subsets when predicting APL (Scenario 1) are shown in Table 4. The predictor system using EHG features (PS1_2) outperformed that of obstetrical features (PS1_1). The highest performance measures were obtained when combining obstetrical and EHG features (PS1_3). The accuracy achieved in PS1_3 was 93.5% (CI 92.6-95.6%) for training subsets and 84.6% (CI 83.4-86.6%) for validation subsets. ROC curves of the three systems in Scenario 1 are depicted in Figure 6(a). The AUC

was greater for PS1_3 with an AUC of 0.96, while PS1_2 and PS1_1 yielded an AUC of 0.94 and 0.89, respectively.

The performance of the predictor systems in Scenario 2, which is aimed at distinguishing between APL-vaginal and APL-caesarian, is shown in Table 5. The best performance measures were reached for PS2_3, yielding an accuracy value of 95.2% (CI 94.4-96.1%) in the training subset and 86.5% (CI 85.3-87.8%) in the validation subset. The performance measures of this scenario were slightly better than those in Scenario 1 in Set_2 and Set_3. The ROC curves of the three classifiers in Scenario 2 are depicted in Figure 6(b). The AUC was 0.98 for PS2_3, 0.95 for PS2_2, and 0.84 for PS2_1.

TABLE 4: Mean and 95% confidence interval of performance measures of predictor systems of Scenario 1 (PS1, APL vs. non-APL). T: train; V: validation. PS1_1 uses DB1_1, PS1_2 uses DB1_2, and PS1_3 uses DB1_3.

		PS1_1	PS1_2	PS1_3
Accuracy (%)	T	84.5 (83.4-85.6)	91.4 (89.8-92.9)	93.5 (92.6-95.6)
	V	75.9 (74.5-77.3)	81.4 (79.9-82.8)	84.6 (83.4-86.6)
Sensitivity (%)	T	83.1 (82.0-84.4)	90.0 (88.2-91.9)	91.8 (90.7-94.6)
	V	73.1 (71.2-75.1)	76.5 (74.4-78.5)	78.9 (78.2-82.6)
Specificity (%)	T	85.8 (84.5-87.2)	92.7 (91.3-94.1)	95.1 (94.4-96.8)
	V	78.7 (76.8-80.6)	86.3 (84.7-88.0)	90.4 (88.0-91.5)

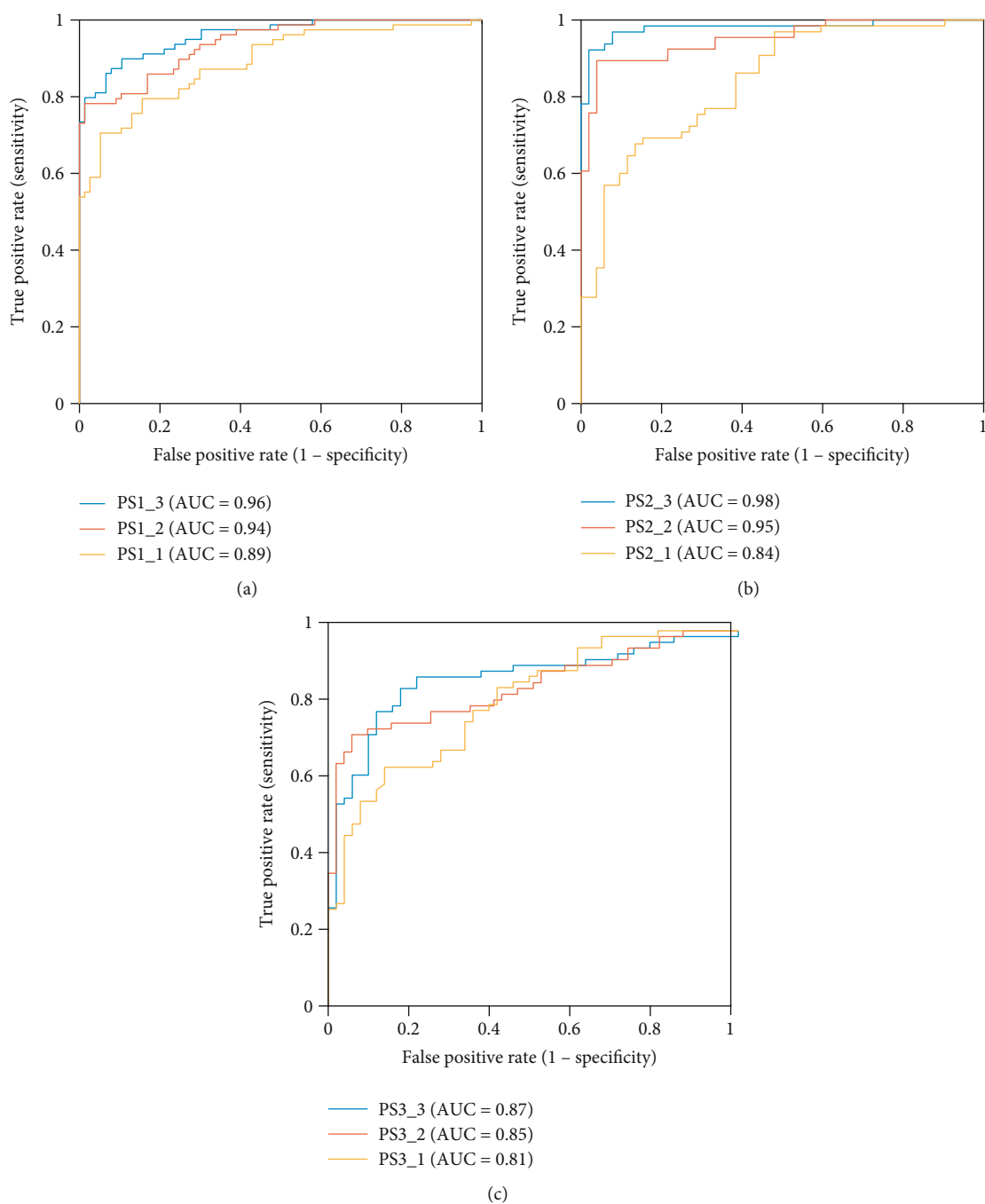


FIGURE 6: ROC curves of predictor systems for each scenario: (a) Scenario 1, (b) Scenario 2, and (c) Scenario 3 (1-step).

TABLE 5: Mean and 95% confidence interval of performance measures of predictor systems of Scenario 2 (APL-vaginal vs. APL-cesarean). T: train; V: validation. PS2_1 uses DB2_1, PS2_2 uses DB2_2, and PS2_3 uses DB2_3.

		PS2_1	PS2_2	PS2_3
Accuracy (%)	T	79.1 (77.7-80.5)	92.1 (91.2-93.1)	95.2 (94.4-96.1)
	V	72.2 (70.6-73.9)	82.7 (81.4-84.1)	86.5 (85.3-87.8)
Sensitivity (%)	T	78.7 (77.1-80.2)	91.1 (90.0-92.3)	94.4 (93.1-95.3)
	V	70.9 (68.8-73.0)	79.4 (77.4-81.4)	83.8 (81.1-84.6)
Specificity (%)	T	79.6 (77.9-81.3)	93.4 (92.4-94.4)	96.3 (95.7-97.3)
	V	73.9 (71.6-76.3)	87.0 (85.2-88.8)	89.9 (87.4-91.0)

The results of the 1-step predictor systems which are aimed at distinguishing between vaginal and cesarean deliveries (Scenario 3) are shown in Table 6. Accuracy values are around 80% for the training subset and 70% for the validation. The table shows that the best performance measures in the training and validation subsets were obtained for PS3_3 but were quite close to those of PS3_2. PS3_3 gave an accuracy of 70.4% (CI 67.7–70.5%), a sensitivity of 67.4% (CI 65.3–69.3%), and a specificity of 74.2% (CI 71.2–75.7%) in the validation subset. However, these figures are only slightly higher (around 2% in training, around 0.5% in validation) than using only EHG features (PS3_2). The ROC curves of the three systems are depicted in Figure 6(c). The highest AUC was found for system PS3_3 (AUC = 0.87). A slightly lower AUC was found for PS3_2 (AUC = 0.85), while the lowest AUC was found for PS3_1 (AUC = 0.81).

The results of the vaginal vs. cesarean predictor system with a 2-step approach are shown in Table 7. Performance values were calculated for the same validation partitions of the database used in the 1-step predictor system in Scenario 3. The best performance measures were obtained by the two-step system, which combines obstetrical and EHG features (PS1_3-PS2_3). The accuracy reached for the 2-step prediction system using Set_1 (PS1_1-PS2_1) was 71.9 (CI: 70.8-73.0%). A great improvement was noted when cascading PS1_2-PS2_2 for Set_2, with an accuracy of 79.9% (CI 78.8-81.0), and slightly higher for PS1_3-PS2_3 for Set_3, with an accuracy 81.4% (CI 80.3-82.5). This latter also achieved a better balance between sensitivity and specificity: 80.3% (CI 78.8–81.8) and 82.8% (CI 81.2–84.8), respectively. The best 2-step predictor system (PS1_3-PS2_3) also gave a much better performance than the best 1-step predictor system—PS3_3: average accuracy 81.4% vs 70.4%, sensitivity 80.3% vs 67.4%, and specificity 82.8% vs 74.2%.

4. Discussion

Predicting the success of labor induction has always been a challenge for obstetricians, and a reliable technique would be an invaluable aid that would help to minimize long waits, maternal-fetal exhaustion and suffering, and the medical costs. Although several attempts have already been made to predict labor induction success from obstetrical information [6–9], these studies have shown poor predictive performance.

In this study, we therefore opted to assess the potential role of EHG for this task.

In the active phase of labor, a necessary step before delivery, the electrical properties of the uterine myocytes undergo changes that generate increased uterine activity. The aim of pharmacologically induced labor is to promote uterine contractions and cervical ripening to achieve vaginal delivery. The reliable prediction of whether an induction agent could trigger APL or not would help clinicians to reduce unnecessary waits and decide whether or not to perform a cesarean section. Benalcazar et al. found a significantly different response between the EHG characteristics of patients that succeeded in achieving APL and those that did not [15, 16]. In the present work, we performed APL predictor systems (Scenario 1) with different sets of features: obstetrical (PS1_1), EHG (PS1_2), and a combination of both (PS1_3). The best performance measures were obtained in PS1_3, which yielded an accuracy of 84.6% in the validation subset and 0.96 for the predictor system AUC.

Vaginal delivery is not always guaranteed even after reaching APL, e.g., in conditions of labor arrest, pelvic-fetal disproportion or loss of maternal-fetal well-being. Knowing that it will definitely happen would help to reduce unnecessary waits. We designed PS2_1, PS2_2, and PS2_3 to discriminate between APL-vaginal and APL-cesarean (Scenario 2). However, as it is necessary to wait until the APL is reached (rarely in the first 4 hours from the onset of labor induction), its clinical significance is lower. In this scenario, combining obstetrical and EHG features also provided the best performance. However, this combination did not significantly improve the predictive performance with EHG features only (3.2% more accuracy in Scenario 1 and 3.8% in Scenario 2), and the EHG feature sets outperformed the results of the obstetrical features in both scenarios, indicating that EHG features provide more accurate information for classifying labor induction success.

As induction success after drug administration is usually defined as vaginal delivery, we developed vaginal delivery predictor systems (Scenario 3) which are potentially of the greatest clinical interest. Our first approach was a 1-step predictor system (PS3_1, PS3_2, and PS3_3). The average accuracy with obstetrical data only (PS3_1) was 68.9%, slightly lower than that in Sievert et al., in which 73.9% of the subjects were correctly classified in the validation cohort using obstetrical data only: gestational age, Bishop score, suspected growth restriction, chronic hypertension, and body mass

TABLE 6: Mean and 95% confidence interval of performance measures of 1-step predictor systems of Scenario 3 (vaginal deliveries vs. cesarean). T: train; V: validation. PS3_1 uses DB3_1, PS3_2 uses DB3_2, and PS3_3 uses DB3_3.

		PS3_1	PS3_2	PS3_3
Accuracy (%)	T	77.1 (75.8-78.5)	80.0 (78.9-81.2)	82.5 (80.3-83.0)
	V	68.9 (67.4-70.4)	69.9 (68.6-71.3)	70.4 (67.7-70.5)
Sensitivity (%)	T	77.0 (75.3-78.7)	77.4 (76.0-78.8)	80.2 (78.2-81.4)
	V	69.2 (66.7-71.6)	67.1 (65.0-69.3)	67.4 (65.3-69.3)
Specificity (%)	T	77.3 (75.7-78.9)	83.4 (82.0-84.7)	85.3 (83.4-86.6)
	V	68.7 (66.5-71.0)	73.6 (71.0-76.1)	74.2 (71.2-75.7)

TABLE 7: Mean and 95% confidence interval of performance measures of cascade predictor system of Scenario 3. V: validation partitions. DB3_Set.

		PS2_1-PS3_1	PS2_2-PS3_2	PS2_3-PS3_3
Accuracy (%)	V	71.9 (70.8-73.0)	79.9 (78.8-81.0)	81.4 (80.3-82.5)
Sensitivity (%)	V	54.9 (53.1-56.7)	75.2 (73.5-76.8)	80.3 (78.8-81.8)
Specificity (%)	V	93.8 (92.8-94.7)	85.9 (84.6-87.3)	82.8 (81.2-84.4)

index [9], and the area under the receiver-operating curve was 75%, which is lower than the 81% obtained in the present work. The larger AUC could be due to the different methods used to design the systems. In our case, we used neural networks, while Sievert et al. used multivariate logistic regression. Our results were also quite close to those obtained by Pitarello et al [7], in which transvaginal sonographic cervical measurements were carried out on 190 pregnant women to predict success (defined as vaginal deliveries). The AUC of all the prediction ultrasound cervical parameters were 68.9% for cervical length, 71.6% for fetal head stage, and 72.0% for cervical dilatation.

Using alternative or additional EHG features slightly improved the accuracy of the validation sets (<71%), in contrast to the enhanced EHG prediction achieved in the previous scenarios. This could have been due to the heterogeneous myoelectrical response to induction drugs in the cesarean delivery cohort, composed of subjects that succeeded in achieving regular and intense contractile activity and APL but could not deliver vaginally for other reasons, plus those who did not reach the necessary contractile activity. This situation would have given rise to bad training, poor generalization capacity, and system performance. We thus turned to a second two-step approach for predicting successful APL and vaginal delivery. The accuracy improved insignificantly when using only obstetrical data, but remarkably when using the EHG parameters (79.9% average accuracy in validation), confirming that two-phase assessment of uterine muscle response to the induction drug reduces class heterogeneity, makes it easier to extract information from the EHG, and gives more accurate predictions. It can also be seen that adding obstetrical information to EHG features does not significantly improve accuracy, but does help to balance sensitivity-specificity.

To the best of our knowledge, this is the first time that EHG has been used to predict successful labor induction. The results obtained show that EHG can play an important

role in labor management decisions and would help clinicians to avoid or reduce unnecessarily long inductions, decrease maternal-fetal risk and suffering, and reduce hospitalization costs.

The study has certain methodological limitations; firstly, it was composed of subjects administered with two different drugs (prostaglandin E1 and prostaglandin E2), which could have given rise to different electrophysiological responses. However, in a clinical context, the ability to predict the success of labor induction with an overall accuracy of 80%, regardless of the drug used, would be a huge advantage. Furthermore, the results of a randomized study would have had greater impact, especially if it compared the effects of various drugs. However, our aim here was to predict pharmacological induction outcomes using EHG and obstetrical information. In this regard, a previous study revealed no statistically significant differences between women who received prostaglandin E1 and prostaglandin E2 in the obstetrical parameters related to labor progress or outcomes, such as the number of women who delivered vaginally before or after 24h of induction, the number of women who achieved active labor period and time to reach labor, and the number of women who underwent cesarean section, arterial pH, and vein pH [15]. In our case, we observed the results of the pharmacological induction and its predictive capacity. Secondly, the unbalanced database of success and failure records in the different scenarios could have caused a bias in favor of the majority class, as was found in [12]. For this reason, the SMOTE data oversampling technique was used, which adds synthetic data to alleviate the problem of class imbalance. Other techniques such as ADASYN have been explored to deal with the problem of imbalance and have given similar results. The use of classification methods that take into account unbalanced data such as the weighted extreme learning machine [48] or weighted decision trees [49] could also be explored. In the same context, we should like to point out that we applied SMOTE before splitting up

the data subsets (training/validation) as has been done in several studies [50–52]. It was seen that when performing cross-validation after simple oversampling, the same samples can be included to build the prediction model and evaluate its performance [53]. Although this is not exactly the case when oversampling with the SMOTE technique, the samples in the training subset can be correlated with samples in the validation subset. It is thus advisable to oversample after data splitting. However, our limited database, mainly the small samples of the minority class in Scenarios 1 and 2, would yield non-extrapolatable validation performance results, since the validation subset would contain very few samples of the minority class in each iteration of the k -fold cross validation (3 samples in Scenarios 1 and 2). On the other hand, applying SMOTE to such a low minority class would yield samples similar to the original ones and would not solve this limitation. We thus opted to perform SMOTE on the entire database, as has been done in numerous other studies [50–52]. We hope to address this limitation in a future work with a larger database. Finally, PSO is a type of wrapped approach for feature selection that uses a learning/classification algorithm to evaluate the quality of a particular feature subset and so is computationally expensive [54]. In a future work, we plan to evaluate other methods with similar performance but computationally less expensive, such as the embedded or hybrid approaches [55].

5. Conclusions

In this work, the use of uterine electromyography for the prediction of the success of labor induction was evaluated for the first time. The predictor system of three labor induction scenarios was designed using a different set of features: obstetrical, EHG, and both. The EHG features outperformed traditional obstetric features in all the scenarios of labor induction outcome prediction. The combination of the obstetrical and the EHG features resulted in greater performance measures but close to those when using only EHG features. Average accuracies of about 85% were obtained when classifying APL vs. non-APL (scenario 1) and APL-vaginal vs. APL-cesarean (scenario 2). Two approaches were assessed and compared for the classification of vaginal vs. cesarean deliveries (scenario 3). One-step predictor systems resulted in a low predictive capacity (accuracy < 71%) The 2-step predictor system, cascade of the classifiers of Scenario 1 and Scenario 2, yielded accuracy values greater than 80% when EHG features were used. These results indicate that EHG parameters can be used to predict labor induction success in the early stages of labor induction. Therefore, an EHG-based labor induction success predictor system could be implemented to assist obstetricians in the task of labor management, improving maternal-fetal well-being, and reducing hospitalization times and costs.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

In accordance with my ethical obligation as a researcher, I declare that this research project received funding from Bial S.A., which could be affected by the results reported in the enclosed paper. I declare that none of the authors have a conflict of interest.

Acknowledgments

This work received financial support from the Spanish Ministry of Economy and Competitiveness, the European Regional Development Fund (DPI2015-68397-R and RTI2018-094449-A-I00), Universitat Politècnica de València VLC/Campus (UPV-FE-2018-B02), Generalitat Valenciana (GV/2018/104), and Bial S.A. The authors are grateful to the Obstetrics Unit of the Hospital Universitario y Politécnico La Fe de Valencia, where recording sessions were carried out.

References

- [1] J. A. Martin, B. E. Hamilton, M. J. K. Osterman, S. C. Curtin, and T. J. Matthews, “Births: final data for 2012,” *National Vital Statistics Reports*, vol. 62, no. 9, pp. 1–68, 2013.
- [2] M. Romanelli, E. Ribiani, L. Burnelli, G. Luzi, G. Affronti, and G. C. Di Renzo, “Pharmacological induction of labour: benefits and risks,” *Minerva Ginecologica*, vol. 59, no. 4, pp. 347–355, 2007.
- [3] O. B. M. Filho, R. M. Albuquerque, and J. G. Cecatti, “A randomized controlled trial comparing vaginal misoprostol versus Foley catheter plus oxytocin for labor induction,” *Acta Obstetrica et Gynecologica Scandinavica*, vol. 89, no. 8, pp. 1045–1052, 2010.
- [4] S. T. Seyb, R. J. Berka, M. L. Socol, and S. L. Dooley, “Risk of cesarean delivery with elective induction of labor at term in nulliparous women,” *Obstetrics and Gynecology*, vol. 94, no. 4, pp. 600–607, 1999.
- [5] E. H. Bishop, “Pelvic scoring for elective induction,” *Obstetrics and Gynecology*, vol. 24, pp. 266–268, 1964.
- [6] L. Hou, Y. Zhu, X. Ma, J. Li, and W. Zhang, “Clinical parameters for prediction of successful labor induction after application of intravaginal dinoprostone in nulliparous Chinese women,” *Medical Science Monitor*, vol. 18, no. 8, pp. CR518–CR522, 2012.
- [7] P. da Rocha Pennachiotti Pitarello, C. T. Yoshizaki, R. Ruano, and M. Zugaib, “Prediction of successful labor induction using transvaginal sonographic cervical measurements,” *Journal of Clinical Ultrasound*, vol. 41, no. 2, pp. 76–83, 2013.
- [8] C. A. de Campos Prado, E. Araujo Júnior, G. Duarte et al., “Predicting success of labor induction in singleton term pregnancies by combining maternal and ultrasound variables,” *The Journal of Maternal-Fetal & Neonatal Medicine*, vol. 29, no. 21, pp. 3511–3518, 2016.
- [9] R. A. Sievert, S. G. Kuper, V. C. Jauk, M. Parrish, J. R. Biggio, and L. M. Harper, “Predictors of vaginal delivery in medically indicated early preterm induction of labor,” *American Journal of Obstetrics and Gynecology*, vol. 217, no. 3, pp. 375.e1–375.e7, 2017.
- [10] R. E. Garfield, W. L. Maner, H. Maul, and G. R. Saade, “Use of uterine EMG and cervical LIF in monitoring pregnant

- patients,” *BJOG: An International Journal of Obstetrics & Gynaecology*, vol. 112, pp. 103–108, 2005.
- [11] Y. Ye-Lin, G. Prats-Boluda, J. Alberola-Rubio, J.-M. Bueno Barrachina, A. Perales, and J. Garcia-Casado, “Prediction of labor using non-invasive laplacian EHG recordings,” in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, vol. 2013, pp. 7428–7431, Osaka, Japan, July 2013.
 - [12] P. Fergus, P. Cheung, A. Hussain, D. Al-Jumeily, C. Dobbins, and S. Iram, “Prediction of preterm deliveries from EHG signals using machine learning,” *PLoS One*, vol. 8, no. 10, article e77154, 2013.
 - [13] A. Aviram, N. Melamed, E. Hadar, O. Raban, L. Hiersch, and Y. Yogev, “Effect of prostaglandin E2 on myometrial electrical activity in women undergoing induction of labor,” *Journal of Perinatology*, vol. 31, no. 5, pp. 413–418, 2014.
 - [14] T. Toth, “Transcutaneous electromyography of uterus in prediction of labor outcome induced by oxytocine and prostaglandine shapes,” *Gynaecologia et Perinatologia: Journal for Gynaecology, Perinatology, Reproductive Medicine and Ultrasonic Diagnostics*, vol. 14, no. 2, pp. 75–76, 2005.
 - [15] C. Benalcazar-Parra, Y. Ye-Lin, J. Garcia-Casado et al., “Electrohysterographic characterization of the uterine myoelectrical response to labor induction drugs,” *Medical Engineering & Physics*, vol. 56, pp. 27–35, 2018.
 - [16] C. Benalcazar-Parra, R. Monfort-Orti, Y. Ye-Lin et al., “Comparison of labour induction with misoprostol and dinoprostone and characterization of uterine response based on electrohysterogram,” *The Journal of Maternal-Fetal & Neonatal Medicine*, vol. 32, no. 10, pp. 1586–1594, 2019.
 - [17] W. L. Maner and R. E. Garfield, “Identification of human term and preterm labor using artificial neural networks on uterine electromyography data,” *Annals of Biomedical Engineering*, vol. 35, no. 3, pp. 465–473, 2007.
 - [18] M. O. Diab, C. Marque, and M. Khalil, “An unsupervised classification method of uterine electromyography signals: classification for detection of preterm deliveries,” *The Journal of Obstetrics and Gynaecology Research*, vol. 35, no. 1, pp. 9–19, 2009.
 - [19] S.-Q. Shi, W. L. Maner, L. B. Mackay, and R. E. Garfield, “Identification of term and preterm labor in rats using artificial neural networks on uterine electromyography signals,” *American Journal of Obstetrics and Gynecology*, vol. 198, no. 2, pp. 235.e1–235.e4, 2008.
 - [20] T. B. Østborg, P. R. Romundstad, and T. M. Eggebø, “Duration of the active phase of labor in spontaneous and induced labors,” *Acta Obstetrica et Gynecologica Scandinavica*, vol. 96, no. 1, pp. 120–127, 2017.
 - [21] WHO, “WHO recommendation on definitions of the latent and active first stages of labour,” 2018, <https://extranet.who.int/rhl/topics/preconception-pregnancy-childbirth-and-postpartum-care/care-during-childbirth/care-during-labour-1st-stage/who-recommendation-definitions-latent-and-active-first-stages-labour-0>.
 - [22] L. Flores Padilla, G. J. González Pérez, J. Trejo Franco et al., “Risk factors in cesarean section,” *Ginecología y Obstetricia de México*, vol. 76, no. 7, pp. 392–397, 2008.
 - [23] N. Baños, F. Migliorelli, E. Posadas, J. Ferreri, and M. Palacio, “Definition of failed induction of labor and its predictive factors: two unsolved issues of an everyday clinical situation,” *Fetal Diagnosis and Therapy*, vol. 38, no. 3, pp. 161–169, 2015.
 - [24] B. Bueno, L. San-Frutos, F. Salazar et al., “Variables that predict the success of labor induction,” *Acta Obstetrica et Gynecologica Scandinavica*, vol. 84, no. 11, pp. 1093–1097, 2005.
 - [25] V. Ware and B. D. Raynor, “Transvaginal ultrasonographic cervical measurement as a predictor of successful labor induction,” *American Journal of Obstetrics and Gynecology*, vol. 182, no. 5, pp. 1030–1032, 2000.
 - [26] M. J. Rooijakkers, S. Song, C. Rabotti et al., “Influence of electrode placement on signal quality for ambulatory pregnancy monitoring,” *Computational and Mathematical Methods in Medicine*, vol. 2014, Article ID 960980, 12 pages, 2014.
 - [27] R. E. Garfield and W. L. Maner, “Physiology and electrical activity of uterine contractions,” *Seminars in Cell & Developmental Biology*, vol. 18, no. 3, pp. 289–295, 2007.
 - [28] H. Leman, C. Marque, and J. Gondry, “Use of the electrohysterogram signal for characterization of contractions during pregnancy,” *IEEE Transactions on Biomedical Engineering*, vol. 46, no. 10, pp. 1222–1229, 1999.
 - [29] C. Buhimschi, M. B. Boyle, and R. E. Garfield, “Electrical activity of the human uterus during pregnancy as recorded from the abdominal surface,” *Obstetrics and Gynecology*, vol. 90, no. 1, pp. 102–111, 1997.
 - [30] D. Schlembach, W. L. Maner, R. E. Garfield, and H. Maul, “Monitoring the progress of pregnancy and labor using electromyography,” *European Journal of Obstetrics, Gynecology, and Reproductive Biology*, vol. 144, pp. S33–S39, 2009.
 - [31] D. Alamedine, A. Diab, C. Muszynski, B. Karlsson, M. Khalil, and C. Marque, “Selection algorithm for parameters to characterize uterine EHG signals for the detection of preterm labor,” *Signal, Image Video Process.*, vol. 8, no. 6, pp. 1169–1178, 2014.
 - [32] J. Terrien, C. Marque, and B. Karlsson, “Spectral characterization of human EHG frequency components based on the extraction and reconstruction of the ridges in the scalogram,” in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1872–1875, Lyon, France, August 2007.
 - [33] G. Fele-Zorz, 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 preterm delivery groups,” *Medical & Biological Engineering & Computing*, vol. 46, no. 9, pp. 911–922, 2008.
 - [34] J. Terrien, C. Marque, J. Gondry, T. Steingrimsdottir, and B. Karlsson, “Uterine electromyogram database and processing function interface: an open standard analysis platform for electrohysterogram signals,” *Computers in Biology and Medicine*, vol. 40, no. 2, pp. 223–230, 2010.
 - [35] M. Hassan, J. Terrien, C. Marque, and B. Karlsson, “Comparison between approximate entropy, correntropy and time reversibility: application to uterine electromyogram signals,” *Medical Engineering & Physics*, vol. 33, no. 8, pp. 980–986, 2011.
 - [36] C. Benalcazar-Parra, A. I. Tendero, Y. Ye-Lin et al., “Feasibility of labor induction success prediction based on uterine myoelectric activity spectral analysis,” in *Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies*, pp. 70–77, Prague, Czech Republic, 2018.
 - [37] J. F. Kaiser, “On a simple algorithm to calculate the ‘energy’ of a signal,” in *International Conference on Acoustics, Speech, and Signal Processing*, pp. 381–384, Albuquerque, NM, USA, 1990.
 - [38] A. Lemancewicz, M. Borowska, P. Kuć et al., “Early diagnosis of threatened premature labor by electrohysterographic

- recordings - The use of digital signal processing,” *Biocybernetics and Biomedical Engineering*, vol. 36, no. 1, pp. 302–307, 2016.
- [39] W. Chen, Z. Wang, H. Xie, and W. Yu, “Characterization of surface EMG signal based on fuzzy entropy,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 15, no. 2, pp. 266–272, 2007.
- [40] X.-S. Zhang, R. J. Roy, and E. W. Jensen, “EEG complexity as a measure of depth of anesthesia for patients,” *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 12, pp. 1424–1433, 2001.
- [41] S. Blanco, A. Garay, and D. Coulombie, “Comparison of frequency bands using spectral entropy for epileptic seizure prediction,” *ISRN Neurology*, vol. 2013, Article ID 287327, 5 pages, 2013.
- [42] M. Nardelli, A. Greco, G. Valenza, A. Lanata, R. Bailon, and E. P. Scilingo, “A novel heart rate variability analysis using lagged Poincaré plot: a study on hedonic visual elicitation,” in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2300–2303, Seogwipo, South Korea, July 2017.
- [43] M. Brennan, M. Palaniswami, and P. Kamen, “Do existing measures of Poincaré plot geometry reflect nonlinear features of heart rate variability?,” *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 11, pp. 1342–1347, 2001.
- [44] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: synthetic minority over-sampling technique,” *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.
- [45] B. Makond, K.-J. Wang, and K.-M. Wang, “Probabilistic modeling of short survivability in patients with brain metastasis from lung cancer,” *Computer Methods and Programs in Biomedicine*, vol. 119, no. 3, pp. 142–162, 2015.
- [46] R. Eberhart and J. Kennedy, “A new optimizer using particle swarm theory,” in *Proceedings of the 6th International Symposium on MicroMachine and Human Science (MHS '95)*, pp. 39–43, Nagoya, Japan, 1995.
- [47] M. Gori and A. Tesi, “On the problem of local minima in back-propagation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 1, pp. 76–86, 1992.
- [48] W. Zong, G.-B. Huang, and Y. Chen, “Weighted extreme learning machine for imbalance learning,” *Neurocomputing*, vol. 101, pp. 229–242, 2013.
- [49] S. Debray, S. Kannan, and M. Paithane, “Weighted Decision Trees,” in *Joint International Conference and Symposium on Logic Programming*, pp. 654–668, Washington DC, 1992.
- [50] U. R. Acharya, V. K. Sudarshan, S. Q. Rong et al., “Automated detection of premature delivery using empirical mode and wavelet packet decomposition techniques with uterine electromyogram signals,” *Computers in Biology and Medicine*, vol. 85, pp. 33–42, 2017.
- [51] L. M. Taft, R. S. Evans, C. R. Shyu et al., “Countering imbalanced datasets to improve adverse drug event predictive models in labor and delivery,” *Journal of Biomedical Informatics*, vol. 42, no. 2, pp. 356–364, 2009.
- [52] A. Smrdel and F. Jager, “Separating sets of term and pre-term uterine EMG records,” *Physiological Measurement*, vol. 36, no. 2, pp. 341–355, 2015.
- [53] R. Blagus and L. Lusa, “Joint use of over- and under-sampling techniques and cross-validation for the development and assessment of prediction models,” *BMC Bioinformatics*, vol. 16, no. 1, p. 363, 2015.
- [54] J. Loughrey and P. Cunningham, “Overfitting in wrapper-based feature subset selection: the harder you try the worse it gets,” in *Research and Development in Intelligent Systems XXI*, pp. 33–43, Springer London, London, 2004.
- [55] T. Butler-Yeoman, B. Xue, and M. Zhang, “Particle swarm optimisation for feature selection: a hybrid filter-wrapper approach,” in *2015 IEEE congress on evolutionary computation (CEC)*, pp. 2428–2435, Sendai, Japan, May 2015.



Hindawi

Submit your manuscripts at
www.hindawi.com

