Evaluation of Advanced Artificial Neural Network Classification and Feature Extraction Techniques for Detecting Preterm Births Using EHG Records

Paul Fergus, Ibrahim Olatunji Idowu, Abir Hussain, Chelsea Dobbins, Haya Al-Askar

Liverpool John Moores University, Byrom Street, Liverpool, L3 3AF, UK.
I.O.Idowu@2009.ljmu.ac.uk, H.Alaskar@2011.ljmu.ac.uk,
{P.Fergus, A.Hussain, C.M.Dobbins}@ljmu.ac.uk

Abstract. Globally, the rate of preterm births is increasing and this is resulting in significant health, development and economic problems. Current methods for the early detection of such births are inadequate. However, there has been some evidence to suggest that the analysis of uterine electrical signals, collected from the abdominal surface, could provide an independent and easier way to diagnose true labour and detect when preterm delivery is about to occur. Using advanced machine learning algorithms, in conjunction with electrohysterography signal processing, numerous studies have focused on detecting true labour several days prior to the event. In this paper however, the electrohysterography signals have been used to detect preterm births. This has been achieved using an open dataset that contains 262 records for women who delivered at term and 38 who delivered prematurely. Several new features from Electromyography studies have been utilized, as well as feature-ranking techniques to determine their discriminative capabilities in detecting *term* and *preterm* records. Seven artificial neural network algorithms are considered with the results showing that the *Radial Basis Function Neural Network* classifier performs the best, with 85% *sensitivity*, 80% *specificity*, 90% area under the curve and a 17% mean error rate.

Keywords: Electrohysterography (EHG); Preterm Delivery; Term Delivery, Artificial Neural Networks

1. Introduction

The World Health Organisation (WHO) defines preterm birth as the delivery of any baby born alive before 37 weeks of gestation. In other words, births that occur before 259 days of pregnancy are defined as *preterm* and births that occur between 259 and 294 days, as *term* (WHO, 2012). Preterm births have a significant adverse impact on the newborn, including an increased risk of death and other health related defects. In 2009, preterm births accounted for approximately 7% of live births, in England and Wales (Bulletin, 2011).

During pregnancy, the monitoring of uterine contractions is vital in order to differentiate between those that are normal and those that may lead to premature birth. The early onset of such contractions can be caused by a number of conditions, including abnormalities in the cervix and uterus, recurrent antepartum hemorrhage and infection (Lucovnik et al., 2011). In the USA, the cost of treatment is reportedly $25.6 billion, whilst in England and Wales, it is estimated to be £2.95 billion, annually (Bulletin, 2011). Consequently, in the last twenty years, a great deal of research has been undertaken to detect and prevent the threat of preterm birth.

One promising technique, which has gained recognition in monitoring uterine activity, is the use of advanced machine learning algorithms and Electrohysterography (*EHG*). This method records signals from the abdominal surface of pregnant women. These readings are then used to study the electrical activity produced by the uterus. The results are convincing and suggest that it is an interesting line of enquiry to pursue.

In conjunction with *EHG* signal processing, the research carried out by Lucovnik *et al.* (Lucovnik et al., 2011) and Hassan *et al.* (Hassan, Muszynski, Alexandersson, & Marque, 2013) illustrates that extracting features from *EHG* signals is key to finding particular spectral information specific to *term* and *preterm* deliveries. The aim of this paper is to evaluate features and their use with several advanced artificial neural network classification algorithms and their ability to distinguish between *term* and *preterm* births. An open dataset has been used, which contains 300 records of pregnant subjects (262 *term* and 38 *preterm*). The results indicate that the selected classifiers, in conjunction with selected features, outperform a number of previous approaches.

The remainder of the paper is structured as follows. Section 2 discusses related studies. Section 3 describes the experimental methodology and the selected extracted feature sets, including the design of the experiment. The results are presented in section 4 before the paper is concluded in Section 5.

2. Related Studies

Over the past 20 years, research has focused on the use of pattern recognition techniques to extract features from *EHG* signals. These include *linear* and *nonlinear* methods, in both the *time* and *frequency* domains, to improve the results obtained from classification algorithms. The extraction of features often forms part of the data pre-processing stage. In our previous work (Fergus et al., 2013), features such as *peak frequency*, *median frequency*, *root mean square* and *sample entropy*, performed particularly well when discriminating between *term* and *preterm* records.

However, it is in the Electromyography (*EMG*) field that we find some new and interesting works. In one study, Lucovnik *et al.* (Lucovnik et al., 2011) investigated whether uterine *EMG* could be used to evaluate propagation velocity (*PV*). In this study, the electrical signals of the uterus were measured both in labour and non-labour patients who delivered at *term* and *prematurely*. The results indicate that, the combination of power spectrum (*PS*) and *PV* peak frequency parameters yielded the best predictive results in identifying true *preterm* labour. However, only one dimension of propagation is considered at a time, which is based on the estimation of time delays between spikes. In comparison, Lange *et al*. (Lange et al., 2014) estimate the *PV* of the entire *EHG* bursts that occur during a contraction by calculating the bursts corresponding to a full contraction event. The results illustrate that the estimated average propagation velocity is 2.18 (60.68) cm/s. No single preferred direction of propagation was found.

Meanwhile, Alamedine *et al.* (Alamedine, Khalil, & Marque, 2013) presented three techniques to identify the most useful features relevant for contraction classification. These included *linear* features, such as *peak frequency, mean frequency* and *root mean square*, and *nonlinear* features, such as the *Lyapunov exponent* and *sample entropy*. In order to choose the most suitable features that represent contractions, feature selection algorithms have been used. This process involved using a binary particle swarm-optimization (*BPSO*) algorithm and calculating the Jeffrey Divergence (*JD*) distance. This is a sequential forward selection (*SFS*) algorithm. The results show that the *BPSO* and *SFS* algorithms could select features with the greatest discriminant capabilities. In this case, out of the six features considered, *sample entropy* produced the best results.

Vasak *et al.* (Vasak et al., 2013) studied whether uterine *EMG* can identify inefficient contractions. This can lead to first-stage labour and caesarean delivery in *term* nulliparous women, with the unplanned onset of labour. In this study, *EMG* was recorded during spontaneous labour in 119 such cases, with singleton *term* pregnancies in the cephalic position. Electrical activity of the myometrium, during contractions, is characterized by its power density spectrum (*PDS*). The diagnosis of labour was made if the patient was in active labour, with no increase in dilation, for at least two hours. The data was analysed to calculate the Intra-class correlation coefficients. This was achieved by comparing the variance of contraction characteristics, within subjects, to the variance between subjects.

3. Methodology

The TPEHG dataset contains the raw *EHG* signals that have been used in our study (PhysioNet, 2012). This data has been pre-processed using data segmentation, feature extraction and classification. The study in (Leman H, Marque C, 1999) illustrates how *EHG* signals can be pre-processed using various frequency related parameters. The study uses several *linear* and *non-linear* signal pre-processing techniques, via three different channels, to discern *term* and *preterm* deliveries. The pre-processing technique used in (Leman H, Marque C, 1999) passed the *EHG* signal through a Butterworth filter configured to filter 0.8-4 Hz, 0.3-4 Hz, and 0.3-3Hz frequencies. However, (Maner, 2003) found that uterine electrical activity occurred within 1Hz and that the maternal heart-rate was always higher than 1Hz. Furthermore, 95% of the patients measured had respiration rates of 0.33 Hz or less. Based on these findings, in this paper, the raw *TPEHG* signals have been passed through the same Butterworth filter to focus on data between 0.34 and 1Hz.

3.1 Raw Data Collection

The raw *EHG* signals, obtained from the Physionet database (PhysioNet, 2012), have been recorded using four bipolar electrodes. These have been adhered to the abdominal surface and spaced at a horizontal and vertical distance between 2.5 and 7cm apart. The total number of records in the EHG dataset is 300 (38 *preterm* records and 262 *term* records). Each of the signals were either recorded early, <26 weeks (at around 23 weeks of gestation) or later, =>26 weeks (at around 31 weeks). Within the dataset, three signals have been obtained simultaneously, 'per record'. This has been achieved by recording through three different channels.

3.2 Feature Selection

The literature reports that *peak frequency*, *median frequency*, *sample entropy* and *root mean squares* have the most potential to discriminate between *term* and *preterm* records. Furthermore, the literature also reports that in *EMG* studies, features such as *waveform length*, *log detector* and *variance* are equally as good at discriminating between different muscle activities. To validate these findings, the above mentioned features have been ranked using *statistical significance*, *linear discriminant analysis using independent search (LDAi)*, *linear discriminant analysis using forward search (LDAf)*, *linear discriminant analysis using backward search (LDAb)* and *gram-schmidt (GS)* analysis.

The *Radial Basis Function Neural Network* (*RBNC*), using the *Linear Discriminant Analysis Forward Search* feature ranking technique showed that, *sample entropy*, *waveform length*, *log detector*, and *variance* provide the best discriminant capabilities and are therefore used to evaluate the classifiers used in this paper.

3.3 Classifiers

This study evaluates the use of seven advanced artificial neural network classifiers. These are the Back-Propagation Trained Feed-Forward Neural Network Classifier (*BPXNC*), Levenberg-Marquardt Trained Feed-Forward Neural Network Classifier (*LMNC*), Perceptron Linear Classifier (*PERLC*), Radial Basis Function Neural Network Classifier (*RBNC*), Random Neural Network Classifier (*RNNC*), Voted Perceptron Classifier (*VPC*) and the Discriminative Restricted Boltzmann Classifier (*DRBMC*) (37steps, 2013).

4. Results

This section presents the classification results for *term* and *preterm* delivery records. This has been achieved using the extracted feature set from the 0.34-1 Hz filter on Channel 3. Using the 80% holdout technique, the initial validation results have been presented. This provides a baseline for comparison against all subsequent evaluations that have been performed, using the oversampled dataset.

4.1 Original Results for 0.34-1 Hz Filter on Channel 3

The performance of each classifier has been evaluated using the *sensitivity*, *specificity*, *errors*, and *AUC* values. In this trial, the experiments have been repeated 30 times. Randomly selected training and test sets have been used in each iteration.

Classifier Performance

The first evaluation uses the original *TPEHG* dataset, which contains 38 *preterm* and 262 *term* records. Table 1, below, illustrates the mean averages obtained over 30 simulations for the *sensitivity*, *specificity*, and *AUC* values. As it can be seen, the *sensitivities* (i.e. the ability to classify a *preterm* record), in this initial test, are low for all classifiers. This is expected because there are a limited number of *preterm* records from which the classifiers can learn. Consequently, *specificities* are higher than *sensitivities*.

Table 1. Original TPEHG Signal (262 *Term* And 38 *Preterm*)

Classifiers	Sensitivity	Specificity	AUC
BPXNC	0	0.9987	54%
LMNC	0.0667	0.9519	58%
PERLC	0.1619	0.8647	57%
RBNC	0.1286	0.9622	56%
RNNC	0.0667	0.9474	56%
VPC	0	1.0000	50%
DRBMC	0	0.9981	58%

4.2 Results for 0.34-1 Hz TPEHG filter on Channel 3 – Oversampled using *SMOTE*

In order to solve the class skew problem, the *preterm* records have been oversampled using the Synthetic Minority Oversampling Technique (*SMOTE*) (Richman & Moorman, 2000). This algorithm oversamples the minority class (38 *preterm* records) to 262, which equals the 262 *term* samples already provided by the *TPEHG* database. A new dataset now contains an even split between *term* and *preterm* records. Using this dataset, the experiment has been repeated a further 30 times.

Classifier Performance

Table 2, illustrates the mean averages obtained over 30 simulations for the *sensitivity*, *specificity*, and *AUC* values. As it can be seen, the *sensitivities*, for all of the algorithms, have significantly improved, while *specificities* have decreased. In addition, the *AUC* results also show a significant improvement in accuracy for all of the classifiers. In particular, the *RBNC* has dramatically improved with an accuracy of 90%.

Table 2. SMOTE TPEHG signal (262 *Term* and 262 *Preterm*)

Classifiers	Sensitivity	Specificity	AUC
BPXNC	79%	58%	72%
LMNC	82%	69%	82%
PERLC	46%	67%	63%
RBNC	85%	80%	90%
RNNC	86%	72%	83%
VPC	98%	2%	50%
DRBMC	59%	55%	56%

5. Conclusion

The development of medical information systems has played an important role in the biomedical domain. This has led to the extensive use of Artificial Intelligence (AI) techniques for extracting biological patterns in data. Furthermore, data pre-processing and validation techniques have also been used extensively to analyze such datasets for classification problems. In this paper, seven classifiers have been used to classify *term* and *preterm* records from the *TPEHG* dataset, filtered between 0.34 and 1 Hz. The results demonstrate that the best performing classifier was the *RBNC* with 85% *sensitivity*, 80% *specificity*, 90% *AUC* and a 17% mean error rate. These results are encouraging and suggest that the approach posited in this paper is a line of enquiry worth pursuing.

Perhaps one negative aspect of the work is the need to utilize oversampling to increase the number of *preterm* samples. A better way would have been to balance the dataset using actual recordings obtained from pregnant women who delivered prematurely. This will be the focus of future research, alongside a more extensive investigation into different machine learning algorithms and techniques.

References

37steps. (2013). Pattern Recognition Tools. *Version 5*.

Alamedine, D., Khalil, M., & Marque, C. (2013). Comparison of different EHG feature selection methods for the detection of preterm labor. *Computational and mathematical methods in medicine*, *2013*, 485684. doi:10.1155/2013/485684

Bulletin, S. (2011). Statistical Bulletin Gestation-specific Infant Mortality in England and Wales. *National Office for Statistics*.

Fergus, P., Cheung, P., Hussain, A., Al-Jumeily, D., Dobbins, C., & Iram, S. (2013). Prediction of Preterm Deliveries from EHG Signals Using Machine Learning. *PloS one*, *8*(10), e77154. doi:10.1371/journal.pone.0077154

Hassan, M., Muszynski, C., Alexandersson, A., & Marque, C. (2013). Nonlinear Correlation Analysis of External Uterine Electromyography. *IEEE Transactions on BioMedical Engineering*, *60*(4), 1160–1166.

Lange, L., Vaeggemose, A., Kidmose, P., Mikkelsen, E., Uldbjerg, N., & Johansen, P. (2014). Velocity and directionality of the electrohysterographic signal propagation. *PloS one*, *9*(1), e86775. doi:10.1371/journal.pone.0086775

Leman H, Marque C, G. J. (1999). Use of the electrohyster- ogram signal for characterization of contractions during pregnancy. *IEEE Trans Biomed Eng*, *46*(10), 1222–1229.

Lucovnik, M., Maner, W. L., Chambliss, L. R., Blumrick, R., Balducci, J., Novak-Antolic, Z., & Garfield, R. E. (2011). Noninvasive uterine electromyography for prediction of preterm delivery. *American journal of obstetrics and gynecology*, *204*(3), 228.e1–10. doi:10.1016/j.ajog.2010.09.024

Maner, W. (2003). Predicting term and preterm delivery with transabdominal uterine electromyography. *Obstetrics & Gynecology*, *101*(6), 1254–1260. doi:10.1016/S0029-7844(03)00341-7

PhysioNet. (2012). The Term -Preterm EHG Database (TPEHG- DB). *physionet.org*.

Richman, J., & Moorman, J. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology- …*, *278*(6), H2039–49.

Vasak, B., Graatsma, E. M., Hekman-Drost, E., Eijkemans, M. J., van Leeuwen, J. H. S., Visser, G. H., & Jacod, B. C. (2013). Uterine electromyography for identification of first-stage labor arrest in term nulliparous women with spontaneous onset of labor. *American journal of obstetrics and gynecology*, *209*(3), 232.e1–8. doi:10.1016/j.ajog.2013.05.056

WHO. (2012). *Born too soon: The Global Action Report on Preterm Birth* (p. 126). Retrieved from http://www.who.int/maternal_child_adolescent/documents/born_too_soon/en/index.html