

Artificial Intelligence for Detecting Preterm Uterine Activity in Gynecology and Obstetric Care

Ibrahim Olatunji Idowu, Paul Fergus, Abir Hussain, Chelsea Dobbins,
Mohammed Khalaf, Raul V. Casana Eslava, Robert Keight
School of Computing and Mathematical Sciences,
Liverpool John Moores University
Byrom Street, L3 3AF, UK.

I.O.Idowu@2009.ljmu.ac.uk, M.I.Khalaf@2014.ljmu.ac.uk, R.Keight@2015.ljmu.ac.uk, P.Fergus, A.Hussain, C.M.Dobbins,
R.V.CasanaEslava{@ljmu.ac.uk}

Abstract —Preterm birth brings considerable emotional and economic costs to families and society. However, despite extensive research into understanding the risk factors, the prediction of patient mechanisms and improvements to obstetrical practice, the UK National Health Service still annually spends more than £2.95 billion on this issue. Diagnosis of labour in normal pregnancies is important for minimizing unnecessary hospitalisations, interventions and expenses. Moreover, accurate identification of spontaneous preterm labour would also allow clinicians to start necessary treatments early in women with true labour and avert unnecessary treatment and hospitalisation for women who are simply having preterm contractions, but who are not in true labour. In this research, the Electrohysterography signals have been used to detect preterm births, because Electrohysterography signals provide a strong basis for objective prediction and diagnosis of preterm birth. This has been achieved using an open dataset, which contains 262 records for women who delivered at term and 38 who delivered prematurely. Three different machine learning algorithms were used to identify these records. The results illustrate that the Random Forest performed the best of sensitivity 97%, specificity of 85%, Area under the Receiver Operator curve (AUROC) of 94% and mean square error rate of 14%.

Keywords — *Electrohysterography(EHG); Random Forest;Preterm Delivery; Term Delivery, Classification, Rule-Base, AUROC and Features extraction, Wavelet Transform (WT), Fourier Transform (FT), Preterm Labour (PTL), Power Spectral Density (PSD), Leave-one-out Cross-validation (LOOCV, Synthetic Minority Over-Sampling Technique (SMOTE), Waveform Database (WFDB).*

I. INTRODUCTION

Premature birth usually occurs when a baby is born before 37 weeks of pregnancy, while term births occur when the baby is delivered after a gestational period of 37 weeks[1]. According to [2], preterm births can occur for three main reasons, including 1) deliveries are brought forward due to medical indication for the best interest of both the mother and the baby, 2) membrane raptures occur prior to labour (PPROM) and 3) the occurrence of spontaneous contractions (termed preterm labour or PTL). Moreover, other health and lifestyle factors, such as cervical and uterine abnormalities, recurrent antepartum haemorrhage, illnesses and infections, any invasive procedure or surgery, underweight or obese mothers, ethnicity, social deprivation, long working hours/late nights, alcohol and drug use, and folic acid deficiency could also facilitate this condition.

In England and Wales 7.3% of babies are born prematurely in 2009, 7.1% in 2010 and 7.2% in 2011.This costs the UK National Health Service (NHS) close to £2.95 billion per year [3]. In the United States one of the most common reasons for antenatal hospitalisations is the avoidance of preterm labour and delivery. Over 10% of the 4 million infants born each year are premature and this costs the medical hospitals \$1500 a day for neonatal intensive care and constitutes a national healthcare expenditure well over \$25.6 billion on women who actually went on to deliver during their stay [4].

In recent years, a number of researches have reported the use of a combination of machine learning classifiers, a better understanding of risk factors related to preterm birth and the use of Electrohysterography (EHG) signal processing has led to the introduction of several measures to improve the effectiveness of the essential treatments of pregnant women [5], [6]. EHG is used to measure electrical activity in the uterus, while machine learning algorithms are trained to distinguish between term and preterm EHG records through the detection of patterns in the data, while managing variance between subjects.

The current research direction of machine learning is moving towards the comparison of different algorithms and the implementation of ensemble algorithm techniques. It can improve the accuracy and robustness of a given classification or regression task, if each learner is better than a random guess and has certain diversity. The increased accuracy is attained by exploring a greater range of learner space, while the robustness is achieved by aggregating and averaging the results.

The contribution of this paper is to investigate the classification performance of different algorithms, such as Random forest classifier, Rule-based classifier and Penalized Logistic Regression classifier techniques, using EHG signals to detect the onset of preterm births. Various machine learning classifiers are compared using an open dataset containing the records of 300 records pregnant subjects. The dataset contains 262 records of women who delivered at term and 38 who delivered prematurely [7]. Using signal processing and oversampling techniques, pre-selected features have been extracted from raw EHG signals that are suited to classifying term and preterm records.

The structure of our paper is as follows. Section II presents the methodology. The simulation results are presented in section III. Finally the paper is concluded in section IV.

II. METHODOLOGY

Predicting preterm birth and diagnosing preterm labour have important consequences, for both the health of the mother and baby, as well as the economy. Most efforts have typically concentrated on mitigating the effects of preterm birth. Nevertheless, they have proven ineffective in the detection of preterm births and incorrect diagnoses of patients are having severe financial implications for the NHS every year [3]. However, one encouraging technique, which has gained recognition in monitoring uterine activity, is the use of advanced machine learning algorithms and Electrohysterography signal processing [8]. For term deliveries, true labour only starts within 24 hours and for preterm deliveries, it may start within 7 to 10 days prior to delivery. The change in EHG activity from non-labour to labour is dramatic; throughout the rest of pregnancy, any change in EHG is more gradual.

A raw EHG signal results from the propagation of electrical activity, between cells in the muscular wall of the uterus. This signal measures the potential difference between the electrodes, in a time domain. In most pregnancies, the connections between gap junctions are sparse, although gradually increasing, until the last few days before labour. A specific pacemaker site has not been conclusively identified, although, due to obvious physiological reasons, there may be a generalised propagation direction, from the top to the bottom of the uterus [9].

Fele-Zorz et al. in [10], conducted a comprehensive study that compared linear and non-linear signal processing techniques to separate EHG records of term and preterm delivery groups. The study was based on EHG records that were collected from a general population of pregnant patients at the Department of Obstetrics and Gynecology Medical Centre in Ljubljana, between 1997 and 2006. These records are publically available via Physionet Term and Preterm Electrohysterography (TPEHG) database [7], [10]. In the database, records were collected from the general population of pregnant patients, as well as those admitted to the hospital with diagnosed preterm labour. In the TPEHG database, there are 300 records/recordings - one record per pregnancy. Each recording was approximately 30 minutes long, had a sampling frequency (fs) of 20Hz, and had a 16-bit resolution with an amplitude range of ± 2.5 mV. Prior to sampling, the signals were passed through an analogue three-pole Butterworth filter, in the range of 0 to 5 Hz. Records were either recorded early, <26 weeks (at around 23 weeks of gestation) or later, ≥ 26 weeks (at around 31 weeks). In this experiment, linear and non-linear methods have been used and are, in both the time and frequency domains, used to improve the results obtained from classification algorithms. The extraction of features often forms part of the data pre-processing stage.

Zardoshti et al. in [11], evaluated a number of features commonly used when dealing with EHG signals. These included integrating absolute value, zero crossings and auto-regression coefficients. However, despite their good discriminant capabilities, a precise frequency threshold for accurate contraction distinction and delivery prediction, over different patients, could not be determined. Fergus et al. in [6],

conducted a broad study of techniques for analysing features of EHG signals, where features such as peak frequency, median frequency, root mean square and sample entropy performed particularly well when discriminating between term and preterm records. The experiment for this paper will be conducted using the following process:

A. Data Acquisition

The dataset used for this paper is the same as that used in [5], [9], [12], [13], with four features (root mean square, median frequency, peak frequency and sample entropy). The raw uterine EHG signal has been extracted from Physionet [7] using the Waveform Database (WFDB) toolbox. The retrieved EHG signals were recorded using four bipolar electrodes adhered to the abdominal surface and spaced at a horizontal, vertical, distance of 2.5cm to 7cm apart. This has been achieved by recording through channel three of the bipolar electrodes. Table 1 shows the details for the TPEHG dataset that has been used in this paper.

TABLE I. EHG TERM AND PRETERM DATASET

Recording Time	Number of Recordings	Time of Recordings	Deliveries Recording Weeks
Early Term	143	22.7	39.7
Term Later	119	30.8	39.6
Total Term Recording	262		
Early Preterm	19	23.0	34.2
Later Preterm	19	30.2	34.7
Total Preterm Records	38		
Total Number of Records	300		

B. Data Pre-processing/Feature Extraction

The collection of raw EHG signals is always temporal. However for analysis and feature extraction purposes, transformation into other domains is possible. These include a frequency representation via Wavelet Transform (WT) and Fourier Transform (FT) [13].

The advantage of frequency-related parameters is that they are less susceptible to signal quality variations, due to electrode placement or the physical characteristics of the subjects. In order to calculate these parameters, a transform from the time domain is required, i.e., using a Fourier transform of the signal. In several of the studies reviewed in [10], [14], Power Spectral Density (PSD) is used to obtain frequency parameters. Peak frequency is one of the features that will be considered within the datasets used within this paper. It describes the frequency of the highest peak in the PSD. Most studies focus on the peak frequency of the burst, in both human and animal studies, and it is confirmed to be one of the most reliable parameters for predicting true labour [9], [10].

Other studies have found medium frequency to be more helpful in determining whether delivery was going to be term or preterm[6], [15]. These along with amplitude based features, root mean squares, and sample entropy will be considered and extracted from the raw EHG signals. Table 2 described the mathematical equation of these four features used in the experiment. In this list, power spectrum is calculated using the

Fast Discrete Fourier Transform, where f_s and N describe the sample frequency and the number of samples, respectively. Conversely, median frequency is used to estimate the typical frequency present in the signal and is regarded in the literature as a useful feature in EHG signal. The median frequency is defined as the midpoint in the frequency power spectrum where the sum of the points on each side is equal. RMS is also used in this study as a signal strength estimator in EHG frequency bands. It provides a measure of the magnitude of the varying quantity of the EHG signal where a signal represented by a time-series $x(t)$ can be calculated as the root of the mean of the squares for all samples in the signal. Measuring the complexity of the signal is regarded as an important feature, which can be calculated using sample entropy. The signal parameters are set where m starts from {2,3,4} and r starts from {0.1,0.2,0.15,0.125,0.175}

TABLE II. FEATURE EXTRACTION TECHNIQUES USED IN EMG

Equation Name	Equation Abbreviations
Root Mean Square of EHG Signal	$RMS = \sqrt{1/N \sum_{n=1}^N x_n^2}$
Peak Frequency of EHG Signal	$F_{max} = \arg(\frac{f_s}{N} \max_{i=0}^{N-1} P(i))$
Median Frequency	$F_{med} = i_m \frac{f_s}{N}, \sum_{i=0}^{i=i_m} P(i) \doteq \sum_{i=i_m}^{N-1} P(i)$
Sample Entropy	$sampEn_{m,r}(y) = \begin{cases} -\log\left(\frac{c_m}{c_{(m-1)}}\right) & c_m \neq 0 \wedge c_{m-1} \neq 0 \\ -\log\left(\frac{N-m}{N-m-1}\right) & c_m = 0 \vee c_{m-1} = 0 \end{cases}$

The signals were filtered using an analogue three-pole Butterworth filter to obtain the 0.34 to 1Hz band, which contains most of the EHG information [16]. This is because Maner et al in [16], found that uterine electrical activity occurred within 1Hz and that the maternal heart-rate (ECG) was always higher than 1Hz. Furthermore, 95% of the patients measured had respiration rates of 0.33 Hz or less. Hence, this research considered that a 0.34-1 Hz bandpass filter removed most of these effects and also movement artefacts. In this paper the raw TPEHG signals were passed through the same Butterworth filter to focus on data between 0.34 and 1Hz.

C. Classification

This study employs the comparison of Random Forest classifier, Rule-based classifier and Penalized Logistic Regression classifier for predicting preterm and term EHG signals. The main reason of selecting these classifiers is because of the robustness of the learning algorithms with unbalanced data as well as the data with missing values.

Breiman et al. in [17], proposed Random Forests, which add an additional layer of randomness to bagging. In addition to constructing each tree using a different bootstrap sample of the data, random forests change how the classification or regression trees are constructed. In standard trees, each node is split using the best split among all variables. In a random forest, each node is split using the best among a subset of predictors randomly chosen at that node.

The Random Forest algorithm can be trained through the process of an ensemble of B trees $\{T_1(X), \dots, T_B(X)\}$, where, $X = x_1 \dots x_p$ p represents the dimensional vector of molecular

descriptors or properties associated with a molecule. The ensemble produces B outputs $\{Y_1 = T_1(X) \dots Y_B = T_B(X)\}$ where $Y_b, b = 1 \dots, B$, is the prediction for a molecule by the b^{th} tree. Outputs of all trees are aggregated to produce one final prediction Y . For classification problems, Y is the class predicted by the majority of trees. In regression, it is the average of the individual tree predictions. In theory, given data on a set of n molecules for training, $D = \{(X_1, Y_1) \dots (X_n, Y_n)\}$, where $X_i, i = 1 \dots, n$ is a vector of descriptors and Y_i is either the corresponding class label (e.g., active/inactive) or activity of interest (e.g., $logIC_{50}$) the training algorithm proceeds as follows. Firstly, from the training data of n molecules, draws a bootstrap sample (i.e., random sample, with replacement, n molecules). Secondly, for each bootstrap sample, grow a tree with the following modification: at each node, choose the best split among a randomly selected subset of m_{try} (rather than all) descriptors. Here, m_{try} is essentially the only tuning parameter in the algorithm. The tree is grown to the maximum size (i.e., until no further splits are possible) and not pruned back. Thirdly, repeat the above steps until (a sufficiently large number) B such trees are grown. When $m_{try} = p$ i.e., the best split at each node is selected among all descriptors, the random forest algorithm is the same as Bagging [17]–[22].

Random Forest algorithm can be very efficient, especially when the number of descriptors is very large. The efficiency of the algorithm, compared to that of growing a single decision tree, comes from two differences between the two algorithms. First, in the usual tree growing algorithm, all descriptors are tested for their splitting performance at each node, while random forest only tests m_{try} of the descriptors.

Since m_{try} is typically very small (the square root of the number of descriptors for classification), the search is very fast. In other words, at each node the Random Forest algorithm effectively only sees m_{try} , rather than p descriptors. Second, to get the right model complexity for optimal prediction strength, some pruning is usually needed for a single decision tree. This is typically done via cross-validation and can take up a significant portion of the computation. In regression, each tree is grown on the residuals of the previous trees. Prediction is done by weighted vote (in classification) or weighted average (in regression) of the ensemble outputs [19]. Recently, Ham et al. in [23], applied Random Forests to classification of hyperspectral remote sensing data. Their approach is implemented within a multi-classifier system arranged as a binary hierarchy. The obtained experimental results in [23] are good for a hyperspectral data set with limited training data.

Penalized Logistic Regression classifier [24] is another model used in our experiment. Huang et al. in [25], used this algorithm for feature selection and classification methods that are applicable to high-dimensional bioinformatics data their result were promising.

The classifier incorporate a penalty term to the log-likelihood function for the same purpose of achieving a stable, as well as accurate, classification model from higher-dimensional data and avoid overfitting by imposing a penalty

on large fluctuations on the estimated parameters [26]. The algorithm is illustrated below in (2).

$$l_p(\beta_0; \beta; \lambda) = -l(\beta_0; \beta) + \lambda J(\beta) \quad (2)$$

Where $l(\beta_0; \beta)$ denotes the unrestricted log-likelihood function phrased. λ represents the regularisation parameter controlling the amount of shrinkage and J represent a penalty function on the coefficient parameter β .

The last classifier used in the experiment is Rule Based classifier. Rules can be derived from decision trees by creating a separate rule for each path taken from the root to each leaf node [27], [28]. Rules can also be directly inferred from the data; algorithms using these rules are called ‘separate-and-conquer’ algorithms or ‘covering’ algorithms. As a minimum, prior knowledge of the domain is needed in order to specify the features (predicates) and the functions. The algorithm can then start by considering the most general rule and the ‘best one’ is chosen which describes the largest proportion of training instances. The algorithms are called covering, because as one rule is learnt, the algorithm eliminates the instances that the rule covers as it has solved them. The rules then become increasingly specific, with more conjunctions added. An example of a rule might be *IF red(x) AND round(x) then apple(x)*; this is an example of two conjunctions in a rule. After every stage of elimination, the ‘best’ rule is always chosen. The process is repeated until either there are no more training examples left, or when a threshold classification level has been reached. Rule quality measures such as the J-measure can be used to measure rule induction and classification processes (Smyth and Goodman, 1990) as quoted by [29]. In a classification situation, a decision rule is a logic statement with the following form below in (3).

$$conditional1 \wedge conditional2 \wedge \dots class, \quad (3)$$

Where a condition is usually an attribute-value pair indicating a certain value of certain attributes that is required to trigger the condition. If the training data records match all the conditions of the rules, then that mean the rule covers the data record; if the rule covers a data record and classifies the data record to the right class, then this means that the rule explains the data record effectively.

Rule-Based classifier are normally more comprehensible than decision trees for learning binary problems, since with rule-learners, only the rules for the positive class need to be learnt. The major disadvantage of rule-based systems is that they are not very suitable for handling reasoning with uncertainty. Rule-based systems are also not very adept at updating belief in the light of new evidence. In practice, certainty factors have been added to deal with such deficiencies, but in reality, they have led to wildly inaccurate estimates of beliefs [27], [28], [30], [31].

III. RESULT SIMULATION

The experiment was performed using the CARET Package in the R programming language. In order to train the selected classifiers, Leave-one-out Cross-validation (LOOCV) is selected. LOOCV is an estimate of the generalised performance of a model trained on $n - 1$ samples of data, which is generally a slightly pessimistic estimate of the

performance of a model trained on n samples. Rather than choosing one model in this experiment, we have fitted the model to all of the data, and used LOOCV to provide a slightly conservative estimate of the performance of that model. In the experiments, 30 iterations have been used to calculate the error rate for each of the classifiers.

Another validation technique used in this experiment is the K-fold Cross-validation technique, which estimates the accuracy of the classifiers. In the experiments, k-fold validation is performed using 5 folds and 1 repetition for the mean square errors to compare the true positives (sensitivity) and true negatives (specificity). The Receiver Operator Curve (ROC) has been used to select the optimal model using the largest value. Moreover, the overall performance of the model is calculated using the average of the area under curve (AUC) of the ROC curve for each fold. The benefit of using the Area under the Receiver Operator curve (AUROC) is that, for many tasks, the operational misclassification costs are unknown or variable, or the operational class frequencies are different to those in the training sample or are variable. In that case, the overall accuracy is often fairly meaningless and the AUROC is a better indicator of performance and ideally we want a classifier that outputs well-calibrated probabilities. This means that in our results, the AUROC, specificity and sensitivity values will be obtained for each fold, and then the mean and standard deviation are calculated to get the mean square error. The \pm in front of the figures in result table of both Table 3 and 4 below represents the mean square error of each result.

The TPEHG dataset is unbalanced (262 term and 38 preterm) subjects and the pre-process of the data was done through a 0.34-1 Hz bandpass filter. Table 3 shows the results of the selected classifiers and the mean averages that have been obtained over 30 simulations for the sensitivity, specificity, and AUROC.

TABLE III. ORIGINAL TPEHG SIGNAL (262 TERM AND 38 PRETERM)

Classifiers	Leave-one-out Cross Validation (LOOCV) 30 Repetitions		Cross Val, 5 Folds, 1 Repetitions
	Sensitivity	Specificity	AUROC
Random Forest	0.977 ±0.021	0.025 ±0.056	0.615 ±0.069
Penalized Logistic Regression	1.000 ±0.000	0.000 ±0.000	0.569 ±0.013
Rule-Based Classifier	0.992 ±0.010	0.0286 ±0.064	0.514 ±0.027

From the four selected classifiers used in this experiment, the results indicate that Random Forest classifier performed the best. This is because Random Forest has relatively good accuracy and robustness learning. They also provide two straightforward methods for feature selection; mean decrease impurity and mean decrease accuracy. The specificity in the initial test was low for all classifiers. The TPEHG dataset is unbalanced (262 term and 38 preterm), which has a significant impact on machine learning algorithms as classifiers are more prone to detect the majority class. Therefore, given that there are more term records, the probability of detecting a preterm

record is low. To address this issue, the minority class (preterm) has been oversampled using the Synthetic Minority Over-Sampling Technique (SMOTE). The technique is effective in solving class skew problems [32]–[34]. SMOTE is used to generate 262 preterm records, using the 38 records that are already available. The SMOTE algorithm allows a new dataset to be generated that contains an even split between term and preterm records (262 for each), which have been oversampled using the original preterm records. Table 4 illustrates the results from a k-fold cross validation technique and oversampling of the dataset used to improve the results obtained from the holdout method.

TABLE IV. SMOTE TPEHG SIGNAL (262 TERM AND 262 PRETERM)

Leave-one-out Cross Validation (LOOCV) 30 Repetitions			Cross Val, 5 Folds, 1 Repetitions
Classifiers	Sensitivity	Specificity	AUROC
Random Forest	0.973 ±0.017	0.858 ±0.009	0.942 ±0.014
Penalized Logistic Regression	0.901 ±0.053	0.876 ±0.028	0.919 ±0.030
Rule-Based Classifier	0.992±0.010	0.858±0.009	0.922±0.008

After oversampling our datasets (262/524, i.e. 50 %), the result indicate that Random Forest were more effective and yielded better outcomes in comparison to other classifications., Random Forest can ensemble the classifier using many decision tree models, thus making the parameters easy to set and prevent the problem of overfitting. The results acquired from this experiment have improved on those previously attained in [35]-[36]. These results are encouraging and suggest that the approach posited in our experiments shows a line of enquiry worth pursuing.

IV. CONCLUSION

Current techniques that are being utilized for diagnosing preterm labour are not sufficient enough to predict preterm labour. This paper has shown the experimental results of three different machine learning algorithms in classifying the TPEHG signal, which has been filtered between 0.34 and 1 Hz band, for predicting preterm births. The result indicates that Random Forest performed the best, with specificity of 86%, sensitivity of 97%, Area under the Receiver Operator curve (AUROC) of 94% and mean square error rate of 14%.As a result, the experiment illustrated that Random Forest classifier provides robust learning and can handle high dimensional data using a large number of trees in the ensemble during the classification process.

These results are encouraging and suggest that this would give practitioners the capability to better manage patients and to provide better health care for them and their unborn children. In turn, it could reduce resulting pregnancy complications. Perhaps one negative aspect of the work is the need to utilize oversampling to increase the number of preterm samples. A better way would be 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

- [1] WHO, "Born Too Soon: The Global Action Report on Preterm Birth," Geneva, 2012.
- [2] A. Greenough, "Long term respiratory outcomes of very premature birth (<32 weeks).," *Semin. Fetal Neonatal Med.*, vol. 17, no. 2, pp. 73–6, Apr. 2012.
- [3] S. Bulletin, "Statistical Bulletin Gestation-specific Infant Mortality in England and Wales," National Office for Statistics, 2011. [Online]. Available: http://www.ons.gov.uk/ons/dcp171778_237668.pdf. [Accessed: 05-Feb-2014].
- [4] W. K. Nicholson, K. D. Frick, and N. R. Powe, "Economic burden of hospitalizations for preterm labor in the United States.," *Obstet. Gynecol.*, vol. 96, no. 1, pp. 95–101, Jul. 2000.
- [5] I. O. Idowu, P. Fergus, A. Hussain, C. Dobbins, H. Al-askar, P. Fergus, A. Hussain, and C. M. Dobbins, "Advance Artificial Neural Network Classification Techniques Using EHG for Detecting Preterm Births," 8th Int. Conf. Complex, Intelligent, Softw. Intensive Syst., vol. 8, no. 1, 2014.
- [6] 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, 2013.
- [7] PhysioNet, "The Term -Preterm EHG Database (TPEHG- DB)," physionet.org, 2012. .
- [8] L. J. Muglia and M. Katz, "The enigma of spontaneous preterm birth.," *N. Engl. J. Med.*, vol. 362, no. 6, pp. 529–35, Feb. 2010.
- [9] M. Lucovnik, R. J. Kuon, L. R. Chambliss, W. L. Maner, S.-Q. Shi, L. Shi, J. Balducci, and R. E. Garfield, "Use of uterine electromyography to diagnose term and preterm labor.," *Acta Obstet. Gynecol. Scand.*, vol. 90, no. 2, pp. 150–7, Feb. 2011.
- [10] G. Fele-Žorž, G. Kavšek, Z. Novak-Antolič, and F. Jager, "A comparison of various linear and non-linear signal processing techniques to separate uterine EMG records of term and pre-term delivery groups.," *Med. Biol. Eng. Comput.*, vol. 46, no. 9, pp. 911–22, Sep. 2008.
- [11] R. H. M. Zardoshti, B. C. Wheeler, K. Badie, "Evaluation of EMG Features for Movement Control of Prostheses," in *proceedings.15th International. conference. IEEE EMBS*, 1993, pp. 1141–1142.
- [12] P. Steer, "The epidemiology of preterm labour.," *BJOG*, vol. 112 Suppl , no. March, pp. 1–3, Mar. 2005.
- [13] A. J. Hussain, P. Fergus, H. Al-Askar, D. Al-Jumeily, and F. Jager, "Dynamic neural network architecture inspired by the immune algorithm to predict preterm deliveries in pregnant women," *Neurocomputing*, vol. 151, pp. 963–974, 2015.
- [14] Y. Ye-Lin, J. Garcia-Casado, G. Prats-Boluda, J. Alberola-Rubio, and A. Perales, "Automatic Identification of Motion Artifacts in EHG Recording for Robust Analysis of Uterine Contractions.," *Comput. Math. Methods Med.*, vol. 2014, p. 470786, Jan. 2014.
- [15] P. Fergus, I. Idowu, and A. Hussain, *Intelligent Computing in Bioinformatics*, 10th Inter., vol. 8590. Taiyuan, China: Springer, 2014, p. pp 309–314.
- [16] W. Maner, "Predicting term and preterm delivery with transabdominal uterine electromyography," *Obstet. Gynecol.*, vol. 101, no. 6, pp. 1254–1260, Jun. 2003.

- [17] L. Breiman, "Random Forrest," *Mach. Learn.*, pp. 1–33, 2001.
- [18] D. R. Cutler, T. C. Edwards, K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler, "Random forests for classification in ecology," *Ecology*, vol. 88, no. 11, pp. 2783–2792, 2007.
- [19] V. Svetnik, A. Liaw, C. Tong, J. Christopher Culberson, R. P. Sheridan, and B. P. Feuston, "Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling," *J. Chem. Inf. Comput. Sci.*, vol. 43, pp. 1947–1958, 2003.
- [20] M. Pal, "Random forest classifier for remote sensing classification," *Int. J. Remote Sens.*, vol. 26, no. June 2015, pp. 217–222, 2005.
- [21] S. Timotheou, "The Random Neural Network□: A survey," vol. 00, no. 0, 2009.
- [22] G. Biau, "Analysis of a Random Forests model," *J. Mach. Learn. Res.*, vol. 13, pp. 1063–1095, 2012.
- [23] J. Ham, Y. Chen, M. M. Crawford, and J. Ghosh, "Investigation of the Random Forest Framework for Classification of Hyperspectral Data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 492–501, 2005.
- [24] C. Bielza, V. Robles, and P. Larrañaga, "Regularized logistic regression without a penalty term: An application to cancer classification with microarray data," *Expert Syst. Appl.*, vol. 38, no. 5, pp. 5110–5118, May 2011.
- [25] J. Huang and S. Ma, "Penalized feature selection and classification in bioinformatics," *Briefings in Bioinformatics*, vol. 9, pp. 392–403, 2008.
- [26] T. Universit and M. Deistler, "Using Penalized Logistic Regression Models for Predicting the Effects of Advertising Material ausgeführt am Institut für," 2009.
- [27] M. Srinivas, "Using Rule Based Classifiers for the Predictive Analysis of Breast Cancer Recurrence.," *J. Inf. Eng. ...*, vol. 2, no. 2, pp. 12–20, 2012.
- [28] L. W. Jinyan Li, "Rule-Based Data Mining Methods for Classification Problems in Biomedical Domains," Singapore, 2004.
- [29] S. B. Kotsiantis, "Supervised Machine Learning: A Review of Classification Techniques," *An Int. J. Comput. Informatics*, vol. 31, no. 3, p. 3, 2007.
- [30] D. Metz, S. Karadgi, U. Müller, and M. Grauer, "Self-Learning Monitoring and Control of Manufacturing Processes Based on Rule Induction and Event Processing," no. c, pp. 88–92, 2012.
- [31] W. W. Cohen, "Fast Effective Rule Induction," in *12th International Conference Machine Learning (ICML)*, 1995, pp. 115–123.
- [32] C. Bunkhumpornpat, K. Sinapiromsaran, and C. Lursinsap, "Safe-level-SMOTE: Safe-level-synthetic minority over-sampling technique for handling the class imbalanced problem," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2009, vol. 5476 LNAI, pp. 475–482.
- [33] L. M. Taft, R. S. Evans, C. R. Shyu, M. J. Egger, N. Chawla, J. A. Mitchell, S. N. Thornton, B. Bray, and M. Varner, "Countering imbalanced datasets to improve adverse drug event predictive models in labor and delivery.," *J. Biomed. Inform.*, vol. 42, no. 2, pp. 356–64, Apr. 2009.
- [34] N. Chawla and K. Bowyer, "SMOTE: synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, no. 1, pp. 321–357, 2002.
- [35] P. F. Al-askar, Haya, D. Al-Jumeily, A.J. Hussain, "The Application of Self-Organised Network Inspired by Immune Algorithm for Prediction of Preterm Deliveries from EHG Signals," in *PgNet*, 2013, pp. 1–6.
- [36] R. Romero, S. Mazaki-Tovi, E. Vaisbuch, J. P. Kusanovic, T. Chaiworapongsa, R. Gomez, J. K. Nien, B. H. Yoon, M. Mazor, J. Luo, D. Banks, J. Ryals, and C. Beecher, "Metabolomics in premature labor: a novel approach to identify patients at risk for preterm delivery.," *J. Matern. Fetal. Neonatal Med.*, vol. 23, no. December, pp. 1344–1359, 2010.