Influence of Feature Selection on Naïve Bayes Classifier for Recognizing Patterns in Cardiotocograms

Mohamed El Bachir Menai, Fatimah J. Mohder, and Fayha Al-mutairi
Department of Computer Science, College of Computer and Information Sciences, King Saud University
Riyad 11543, Saudi Arabia
Email: menai@ksu.edu.sa

***Abstract*— Cardiotocography is a technical procedure that consists in recording the fetal heart rate (FHR) and uterine activity (UA) during the last months of a pregnancy. Cardiotocogram (CTG) analysis consists in identifying some patterns associated to fetal activity in order to detect potential fetal pathologies. Several automatic classification methods have been already tested on CTG data sets, while a few feature selection (FS) methods have been considered. The aim of this paper is to investigate the influence of FS on the performance of a naïve Bayes classifier for FHR patterns and fetal states. We empirically compare the performance of several models using four different FS methods (Correlation-based, ReliefF, Information Gain, and Mutual Information). We find that ReliefF yields to a better performance for fetal state classification, while no FS method worth the effort for FHR pattern classification.**

***Index Terms*— cardiotocography, fetal heart rate (FHR), fetal states, feature selection, naïve bayes classifier**

I. Introduction

Complicated childbirth might cause death or brain injury of baby (approximately 1-7 in 1000 babies) [1]. Nowadays, Cardiotocography procedure is used for electronic fetal monitoring in order to measure and record uterine activity (UA), baseline fetal heart rate (FHR), baseline of FHR variability, presence of accelerations, periodic or episodic decelerations, and changes of FHR patterns over time.

According to some reviews, incorrect analysis of FHR signal, such as recognizing the acceleration and deceleration patterns in FHR signal, might cause around 50% of birth-related brain injuries [2], [3]. Such medical errors could be prevented by using automated analysis of FHR signal to assist the clinicians in the assessment of the fetal state and improve the intra-partum care, especially as more than 90% of FHR parameters are automatically monitored and detected. Automated analysis of FHR signals would be more accurate to detect FHR patterns than the visual analysis achieved by the clinician.

Several machine learning techniques have been already considered for automated FHR classification, such as Artificial Neural Networks (ANNs) [4], [5, [6], support vector machines (SVM) [7], [8], [9], [10], decision trees [9], naïve Bayes classifiers [9], and other hybrid methods [5]. However, only a few FS methods have been investigated for Cardiotocogram analysis.

In this work, we demonstrate the influence of FS on the performance of naïve Bayes classifier for the automated analysis of FHR signals in order to obtain information on the FHR patterns and fetal state. We compare the performance of several classifiers based on a naïve Bayes classifier and four different feature selection methods (Correlation-based, ReliefF, Information Gain, and Mutual Information) using UCI Cardiotocography dataset [11]. We demonstrate the positive impact of ReliefF on fetal state classification, and show that no FS method worth the effort for FHR pattern classification.

The remainder of this paper is organized as follows. Section 2 outlines the Cardiotocography procedure and related concepts. Section 3 presents some work related to Cardiotocogram analysis using automatic classification methods. Section 4 describes the FS methods to compare and outlines the fundamentals of a simple naïve Bayes classifier. Section 5 presents the empirical comparison and discusses the results obtained. Section 6 concludes the paper.

II. Cardiotocography

Since the 1960's, obstetricians are using the Cardiotocography, an electronic method for recording (graphy) the fetal heartbeat (cardio) and uterine contractions (toco) during pregnancy, by means of a Cardiotocograph or an electronic fetal monitor (EFM). Fig. 1 illustrates a typical Cardiotocogram (CTG).

The continuous monitoring by using CTG requires qualitative and quantitative interpretations of several parameters described as follows [12]:

- Uterine activity (contractions):
  - Frequency: Number of contraction in a standard interval.

Manuscript received Jan 15, 2013; revised March 1, 2013.

doi: 10.12720/jomb.2.1.66-70

– Duration: The amount of time from the start of a contraction to the end of the same contraction.
– Intensity: A measure of how strong a contraction is.
– Resting tone: A measure of how relaxed the uterus is between contractions.
– Interval: The amount of time between the end of one contraction to the beginning of the next contraction.

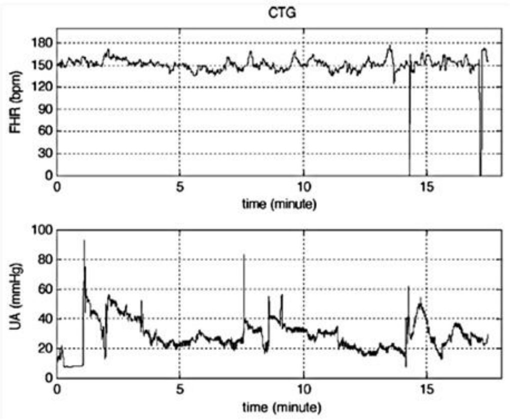

Figure 1. A typical CTG [9]

Uterine activity may be defined as:
– Normal- less than or equal to 5 contractions in 10 minutes, averaged over a 30-minute window.
– Tachysystole- more than 5 contractions in 10 minutes, averaged over a 30-minute window.

- Baseline fetal heart rate (FHR), which is determined by approximating the mean FHR rounded to increments of five beats per minute during a 10-minute window, excluding accelerations and decelerations and periods of marked FHR variability.
  – Baseline FHR less than 110 beats per minute and symptoms are termed as Bradycardia.
  – Baseline FHR greater than 160 beats per minute and symptoms are termed as tachycardia.
- Baseline FHR variability, which is determined in a 10-minute window, excluding accelerations and decelerations. Baseline FHR variability is defined as fluctuations in the baseline FHR that are irregular in amplitude and frequency. The fluctuations are visually quantified as the amplitude of the peak-to-trough in bpm (beat per minute).
  – Absent
  – Minimal
  – Moderate
  – Marked
- Presence of accelerations: Visually apparent abrupt increase in FHR. An abrupt increase is an increase from an onset of acceleration to the peak in less than or equal to 30 seconds (to be considered as acceleration, the peak must be greater than or equal to 15 bpm).
- Periodic or episodic decelerations
  – Periodic: Refers to decelerations that are associated with contractions
  – Episodic: Refers to those not associated with contractions

  There are four types of decelerations:
  – Early deceleration: It is related to a gradual decrease in the FHR with an onset of deceleration to a nadir (more than 30 seconds) where the nadir occurs with the peak of a contraction.
  – Late deceleration: It is related to a gradual decrease in the FHR with an onset of deceleration to a nadir (more than 30 seconds).
  – Variable deceleration: It is related to an abrupt decrease in the FHR (more than 15 bpm) that was measured from the most recently baseline where from the deceleration's onset to nadir is less than 30 seconds and the deceleration lasts (more than 15 seconds).
  – Prolonged deceleration: It is present when there is a visually apparent decrease in FHR from the baseline that is greater than or equal to 15 bpm, lasting greater than or equal to 2 minutes, but less than 10 minutes. A deceleration that lasts greater than or equal to 10 minutes is a baseline change.
- Changes or trends of FHR patterns over time.
  – Category I (Normal): Baseline rate 110-160 bpm, Moderate variability, Absence of late, or variable decelerations, and early decelerations and accelerations may or may not be present.
  – Category II (Indeterminate): Tracing is not predictive of abnormal fetal acid-base status, but evaluation and continued surveillance and reevaluations are indicated.
  – Category III (Abnormal): Absence of baseline variability with recurrent late or variable decelerations or Bradycardia; or sinusoidal fetal heart rate.

III. RELATED WORK

The International Federation of Obstetrics and Gynaecology (FIGO) guidelines [13] were introduced as an attempt to standardize the use of electronic monitoring of FHR. The first work of automatic CTG analysis following FIGO guidelines consists in describing and extracting the CTG morphological features [14]. Bernades [15] developed SisPorto, a system for automatic analysis of CTG tracings, based on an improvement of the morphological feature extraction introduced in [14].

Artificial Neural Networks (ANNs) were used as a classifier [1] to detect FHR acceleration and deceleration patterns and to estimate the FHR baseline and variability. ANNs were used to classify deceleration patterns into episodic and periodic decelerations [6] according to FIGO guidelines and based on the relationship between the parameters of the deceleration and the associated uterine contraction. ANNs with Radial Basis Functions (RBF) and MultiLayer Perceptrons (MLP) were the best performing classifiers.

Support Vector Machines (SVM) have been used for FHR signal analysis. Georgoulas et al. [7] used discrete wavelet transformation to extract scale-dependent features of the FHR signal and SVM for their classification. Georgoulas et al. [9] used SVM with RBF and polynomial kernels to identify fetal and neonatal compromise, namely metabolic acidosis [16]. The RBF kernel machines outperformed the polynomial machines and both of them outperformed the conventional methods of k-nearest neighbor (k-NN), linear and quadratic discriminant classifiers.

Chudáček *et al*. [8] used SVM, naïve Bayes, and a decision tree (C4.5 algorithm) with a polynomial kernel to analyze FHR signals based on linear features (e.g. Description of the FHR baseline using mean) and non-linear features (e.g. Fractal dimension of waveform). They used three FS methods: Principal Component Analysis, Information Gain, and Group of Adaptive Models Evolution (GAME).

Krupa *et al.* [10] proposed a new method for FHR signal analysis based on Empirical Mode Decomposition (EMD) for feature extraction and SVM with RBF for classification of FHR recordings.

Georgoulas *et al*. [17] proposed a FS method based on bPSO (binary Particle Swarm Optimization) for FHR signal analysis using SVM and k-NN.

Hybrid methods have been also considered for automated FHR signal analysis. Fontenla-Romero et al. [5] proposed several approaches for the recognition of acceleration and deceleration patterns in FHR signals, including rule-based approach, ANNs, and a neuro-fuzzy approach.

IV. Proposed Approach

In this Section, we describe the FS methods and naïve Bayes classifier used in our experiments.

FS is the process that outputs a subset of the input feature set by retaining the main characteristics necessary for the classification process. FS can reduce the dimensionality of the measurement space and lead to more efficient classification.

Four basic phases represent a typical FS procedure: Subset generation, subset evaluation, stopping criterion, and results' validation. The process starts with the generation of subset by employing a particular search strategy in order to give rise to candidate feature subsets. This is followed by an evaluation of each candidate subset according to a specific evaluation criterion and then compared with the best feature subset. If it is better, the previous best one is replaced by it. These processes of generation and evaluation of a subset are repeated until a certain stopping criterion is attained. Finally, the selected best feature subset is carefully validated by the test data set.

We select four FS methods among the most useful ones in medical data classification: Correlation-based FS, ReliefF, Information Gain, and Mutual Information. These methods are based on assigning a score to each feature.

A. Correlation-based Feature Selection (CFS)

CFS is a filter FS method. It is based on measuring the correlation between a feature and a class as well as the inter correlation between features. CFS evaluates subsets of features on the basis of the hypothesis "Good feature subsets hold features highly interrelated with the class, yet uncorrelated to each other" [18]. This hypothesis gives rise to two concepts. One is the feature-classification correlation and another is the feature-feature correlation. Equation (1) gives the merit of a feature subset *S* of *k* features:

$$M_s = \frac{k\ r_{fc}}{\sqrt{k + k(k-1)r_{ff}}} \tag{1}$$

M_S is the merit of a feature subset *S*, r_{fc} is the average of the correlations between the features and the class, r_{ff} is the average of the inter-correlation between features in the subset

B. ReliefF

One of the most used algorithms for FS is ReliefF. It is considered as a fast, easy to implement and accurate algorithm. In addition, it is robust to noise and feature redundancy. This algorithm assigns relevance weight to each feature by using an instance based learning method. A weight refers to the ability of the feature to distinguish between the class values. The features are then ranked according to their weights and the best ones are selected based on a specific threshold.

ReliefF algorithm starts by selecting random instances from the training data. Then, for each instance it finds the nearest *k* instances to the same class (nearest hits) and to the classes other than the class of the selected instance (nearest misses). The feature's weight is updated according to how well its values distinguish the sampled instance from its nearest hits and nearest misses. A feature gets a high weight if it distinguishes between instances from different classes and has the same value for instances of the same class.

C. Information Gain (IG)

Information entropy describes the amount of impurity in a set of features. IG is defined as the entropy of the whole set minus the entropy when a particular feature is chosen. Initially, we compute the IG for each feature in the dataset and then we remove the features whose information gain is less than a fixed threshold.

D. Mutual Information (MI)

MI is a measure of interdependence between random features [19]. The idea behind the MI method is to reduce the amount of uncertainty a feature (or more than one feature) possesses in relation to the class. It determines the level of uncertainty for each feature by looking at dependencies between them. The features which have a high level of uncertainty are highly dependent on other features and are thus discarded. Therefore, the features which are only dependent on the target variable are left. In this way, the features selected should be highly relevant to the target variable.

E. Naïve Bayes Classifier

Naïve Bayes classifiers have been found to perform surprisingly well in medical diagnosis. Indeed, they have been used successfully for several medical classification problems, including breast cancer, thyroid disease, and identification of neonatal hearing impairment [20].

Let $X = \{x_1, \dots, x_n\}$ represents the set of features of a CTG (with/without FS) and $C = \{c_1, \dots, c_m\}$ represents the set of classes of FHR patterns/fetal state. Given a new CTG instance X, the classification problem asks to assign a class c_k and a class c_l to FHR patterns and fetal state, respectively.

The simple Naïve Bayes classifier is a probabilistic classifier based on Bayesian probability theory. It classifies a new CTG to the most probable FHR pattern c_k (resp. fetal state c_l) according to (3).

$$c_k = argmax_{c_k \in C} P(c_k) \prod_{i=1}^{n} P(x_i \mid c_k) \qquad (2)$$

$P(c_k)$ is the prior probability of each class, and the probability $P(x_i|c_k)$ can be estimated using a Gaussian distribution or Laplacean prior.

V. Experimental Results

A. Data Sets and Protocol

UCI cardiotocography data set [11], [21] is used in our experiments. It contains 2126 fetal CTGs represented by 21 diagnostic features related to FHR and UA. The classification of CTGs is both with respect to FHR patterns and fetal states. The features represent UA, FHR, baseline FHR variability, presence of accelerations, periodic or episodic decelerations, and characteristics of histogram.

FHR pattern class code (1 to 10) for classes (A to SUSP):

- A: Calm sleep,
- B: REM sleep,
- C: Calm vigilance,
- D: Active vigilance,
- SH: Shift pattern (A or SUSP with shifts),
- AD: Accelerative/decelerative pattern (stress situation),
- DE: Decelerative pattern,
- LD: Largely decelerative pattern,
- FS: Flat-sinusoidal pattern (pathological state),
- SUSP: Suspect pattern.

Fetal state class code:

- Normal (N),
- Suspect (S),
- Pathologic (P).

A cardiotocogram is classified as (Pathological) in one of the following cases:

- FHR baseline greater than 170 or baseline less than l00.
- Reduction in LTV longer than 40 minutes.
- Severe or prolonged and repetitive decelerations.
- Bradycardia longer than 10 min.
- Sinusoidal pattern.

The four FS methods are applied on the data set. For CFS, ReliefF and IG, three threshold values are tested: 10, 15, and 20. A naïve Bayes classifier is performed with randomly selected train/test splits for both FHR pattern and fetal state data sets with a 30% test data. Results are reported as average classification accuracy, sensitivity, and specificity across 100 trials. They are defined as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (3)$$

$$Sensitivity = \frac{TP}{TP + FN} \qquad (4)$$

$$Specificity = \frac{TN}{TN + FP} \qquad (5)$$

TP: True Positives; *TN*: True Negatives; *FP*: False Positives; *FN*: False Negatives.

B. Results

Table 1 and Table 2 show the average best results obtained with the naïve Bayes classifier based on the FS methods. “FS-threshold” denotes the “threshold” (number of attributes) that leads to the best results obtained using “FS”.

TABLE I. Results of Naïve Bayes for the Classification of FHR Patterns with/without FS

	Accuracy%	Sensitivity%	Specificity%
CFS-10	86,790	82,557	98,440
ReliefF-10	79,300	78,134	97,491
ReliefF-15	82,031	80,440	97,800
IG-10	81,498	81,841	97,737
IG-15	88,344	87,155	98,585
MI (20/21)	89,467	87,290	**98,722**
Without FS	**89,499**	**98,722**	86,959

TABLE II. Results of Naïve Bayes for the Classification of Fetal States with/without FS

	Accuracy%	Sensitivity%	Specificity%
CFS-10	93,509	89,984	94,139
ReliefF-10	92,642	88,408	94,357
ReliefF-15	**93,979**	**91,582**	**95,790**
IG-10	92,078	87,411	94,573
IG-15	91,221	85,728	93,394
MI (20/21)	92,078	87,417	94,573
Without FS	92,076	87,241	94,625

Table I shows the results on the FHR pattern data set. The naïve Bayes classifier reaches the maximum accuracy (89.499%) and sensitivity (98.722%) when no FS method is used. However, the maximum specificity (98.722%) is reached when a subset of features is selected by MI (20/21): The feature DS (number of severe decelerations per second) is not selected. The second best accuracy (89.467%) is obtained when this feature DS is not selected. It appears that the specificity can be increased by removing DS at the expense of decreasing the sensitivity.

Table II shows the results of the fetal state data set. The best average results in terms of accuracy (93.979%), sensitivity (91.582%), and specificity (95.790%) are reached when a subset of 15 features is selected by ReliefF. The features that have not been retained are:

- FM (number of fetal movements per second),
- DS (number of severe decelerations per second),
- DP (number of prolonged decelerations per second),
- Mode (histogram mode),
- Variance (histogram variance),
- Tendency (histogram tendency).

These results suggest that the six features that have not been selected by ReliefF could not be correlated to the target. This result requires an assessment and validation by obstetric clinicians.

VI. Conclusion and Future Work

This paper has compared four FS methods for the classification of FHR patterns and fetal state. The methods considered are among the most popular one, especially in medical data classification. A naïve Bayes classifier has been trained and tested on UCI cardiotocography data set. Based on overall empirical results, we find that the tested FS methods have no significant impact on the classification of FHR patterns. However, ReliefF method has a significant impact on the performance of fetal state classification. Indeed, the best results are obtained when only 15 features are selected.

In future work, we plan to collaborate with obstetric clinicians and physicians in order to assess the computational results. We will also investigate other FS methods, looking for correspondence between attributes and classification performance.

References

[1] P. A. Warrick, E. F. Hamilton, and M. Macieszczak, "Neural network based detection of fetal heart rate patterns," in *Proc. IEEE Int. Joint Conf. on Neural Netw. IJCNN'05*, Aug. 2005, vol. 5, pp. 2400- 2405.

[2] E. Draper, J. Kurinczuk, C. Lamming, M. Clarke, D. James, and D. Field, "A confidential enquiry into cases of neonatal encephalopathy," *Arch Dis Child Fetal Neonatal*, 87(3): F176–F180, doi: 10.1136/fn.87.3.F176, Nov. 2002.

[3] S. B. Ransom, D. M. Studdert, M. P. Dombrowski, M. M. Mello, and T. A. Brennan, "Reduced medicolegal risk by compliance with obstetric clinical pathways: A case-control study," *Obstet. & Gynecol.*, vol. 101, Issue 4, pp. 751-755, Apr, 2003.

[4] C. Ulbricht, G. Dorffner, and A. Lee, "Neural networks for recognizing patterns in cardiotocograms," *Artif. Intell. Med.*, vol. 12, pp. 271–284, 1998.

[5] O. Fontenla-Romero, A. Alonso-Betanzos, and B. Guijarro-Berdinas, "Adaptive pattern recognition in the analysis of cardiotocographic records," *IEEE Trans. Neural Netw.*, vol. 12, no. 5, pp. 1188–1195, Sep. 2001.

[6] M. Jezewski, P. Labaj, J. Wrobel, A. Matonia, J. Jezewski, and D. Cholewa, "Automated classification of deceleration patterns in fetal heart rate signal using neural networks," in *Proc. IV Latin American Cong. on Biomed. Eng. 2007, Bioeng. Sol. for Latin America Health IFMBE Proc.*, 2008, vol. 18, pp. 5-8.

[7] G. Georgoulas, D. Stylios, and P. P. Groumpos, "Classification of fetal heart rate using scale dependent features and support vector machines," in *Proc of 16th IFAC World Cong.*, July 2005.

[8] V. Chudáček, J. Spilka, M. Huptych, G. Georgoulas, P. Janků, M. Koucký, C. Stylios, and L. Lhotská, "Comparison of linear and non-linear features for intrapartum cardiotocography evaluation – Clinical usability vs. Contribution to classification," *Analysis Biomed. Signals Images, BIOSIGNAL*, pp. 369-372, 2010.

[9] G. Georgoulas, D. Chrysostomos, and P. P. Groumpos, "Predicting the risk of metabolic acidosis for newborns based on fetal heart rate signal classification using support vector machines," *IEEE Trans. Biomed. Eng.*, pp. 875- 884, Apr. 2006.

[10] N. Krupa , M. A. MA , E. Zahedi, S. Ahmed, and F. M. Hassan, "Antepartum fetal heart rate feature extraction and classification using empirical mode decomposition and support vector machine," *BioMed. Eng. OnLine*, 10:6, 2011.

[11] A. Frank and A. Asuncion, UCI Machine Learning Repository [http://archive.ics.uci.edu/ml] Irvine, CA: University of California, School of Information and Computer Science.

[12] D. Robertsa, B. Kumara, D.G. Tincellob, and S. A. Walkinshawa, "Computerised antenatal fetal heart rate recordings between 24," *British J. Obstet. Gynaecol.*, pp. 858-862, Aug. 2001.

[13] G. Rooth, A. Huch, and R. Huch, "Guidelines for the use of fetal monitoring," *Int. J. Gynaecol. Obstet.*, vol. 25, pp. 159-167, 1987.

[14] R. Mantel, H. P. van Geijn, F. J. Caron, J. M. Swartjes, E. E. van Woerden, and H. W. Jongsma, "Computer analysis of antepartum fetal heart rate: 2. Detection of accelerations and decelerations," *Int. J. Biomed. Comput.*, 25(4), pp. 273-286, May 1990.

[15] J. Bernardes, C. Moura, J. P. de Sa, and L. P. Leite, "The Porto system for automated cardiotocographic signal analysis," *J. Perinat. Med.*, pp. 61–65, 1991.

[16] H. P. van Geijn, "Developments in CTG analysis," *Bailliere's Clin. Obstet. Gynaecol.*, vol. 10, no. 2, pp. 185–209, June 1996.

[17] G. Georgoulas, C. Stylios, V. Chudáček, M. Macas, J. Bernardes, and L. Lhotska, "Classification of fetal heart rate signals based on features selected using the binary particle swarm optimization algorithm," *World Cong. Med. Phy. Biomed.l Eng. 2006, IMBE Proc.*, 2007, pp. 1156-1159.

[18] M. A. Hall, "Correlation-based feature selection for machine learning," University of Waikato, 1999.

[19] G. D. Tourassi, E. D. Frederick, M. K. Markey, C. E. Floyd Jr, "Application of the mutual information criterion for feature selection in computer-aided diagnosis," *Med. Phys.*, 28(12), pp. 2394-402, Dec. 2001.

[20] R. Abraham, J.B. Simha, and S. S. Iyengar, "Effective discretization and hybrid feature selection using naïve Bayesian classifier for medical data mining," *Int. J. Comp. Intel. Res.*, pp. 116–129, 2009.

[21] D. Ayres-de Campos, J. Bernardes, A. Garrido, J. Marques-de-Sá, and L. Pereira-Leite, "SisPorto 2.0: A program for automated analysis of cardiotocograms," *J. Matern. Fetal Med.*, 9(5), pp.311-8, Sep.-Oct. 2000.

Mohamed El Bachir Menai received a Ph.D. degree in computer science from Mentouri University of Constantine, Algeria, and University of Paris VIII, France, in 2005. He received also a "Habilitation universitaire" in computer science from Mentouri University of Constantine in 2007 (it is the highest academic qualification in Algeria and France). He is currently associate professor in the department of computer science at King Saud University. His main interests include satisfiability problems, evolutionary computing, machine learning, and natural language processing.

Fatimah J. Mohder received an MSc degree in computer science from King Saud University, Riyad, in 2012. Her research interests include machine learning and medical informatics.

Fayha Al-mutairi received an MSc degree in computer science from King Saud University, Riyad, in 2013. Her areas of interest include health informatics and medical decision support systems.