Detection rate of fetal distress using contraction-dependent fetal heart rate variability analysis

Document Version:
Accepted manuscript including changes made at the peer-review stage

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.
Link to publication

Detection rate of fetal distress using contraction-dependent fetal heart rate variability analysis

G.J.J. Warmerdam[1]**, R. Vullings**[1]**, J.O.E.H. Van Laar**[2]**, M.B. Van der Hout-Van der Jagt**[1,2]**, J.W.M. Bergmans**[1]**, L. Schmitt**[3]**, and S.G. Oei**[1,2]

[1] Faculty of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands
[2] Máxima Medical Center, Veldhoven, The Netherlands.
[3] Philips Research, Eindhoven, The Netherlands.

E-mail: `g.j.j.warmerdam@tue.nl`

objective, approach, main results, significance,

Abstract. *Objective*: Monitoring of the fetal condition during labor is currently performed by cardiotocography (CTG). Despite the use of CTG in clinical practice, CTG interpretation suffers from a high inter- and intra-observer variability and a low specificity. In addition to CTG, analysis of fetal heart rate variability (HRV) has been shown to provide information on fetal distress. However, fetal HRV can be strongly influenced by uterine contractions, particularly during the second stage of labor. Therefore, the aim of this study is to examine if distinguishing contractions from rest periods can improve the detection rate of HRV features for fetal distress during the second stage of labor. *Approach*: We used a dataset of 100 recordings, containing 20 cases of fetuses with adverse outcome. The most informative HRV features were selected by a Genetic Algorithm and classification performance was evaluated using Support Vector Machines. *Main results*: Classification performance of fetal heart rate segments closest to birth improved from a geometric mean of 70% to 79%. If the classifier was used to indicate fetal distress over time, the geometric mean at 15 minutes before birth improved from 60% to 72%. *Significance*: Our results show that combining contraction-dependent HRV features with HRV features calculated over the entire fetal heart rate signal improves the detection rate of fetal distress.

1. Introduction

The introduction of cardiotocography (CTG) in the 1960s, has enabled continuous monitoring of the fetal heart rate (FHR) and uterine contractions. In current obstetric units, CTG has become the worldwide standard for detection of fetal distress during labor. However, the CTG is interpreted visually and the inter- and intra-observer variability is high (Blix et al.,

2003). Furthermore, despite a high sensitivity of CTG interpretation, the specificity is poor (Ayres-de Campos et al., 2015). As a result of the low specificity, the rate of unnecessary operative interventions has increased since the introduction of CTG in the clinic (Alfirevic et al., 2006). To detect if the fetus is in distress, additional information is therefore often required when the CTG is abnormal.

In recent studies, analysis of fetal heart rate variability (HRV) has been shown to provide information on fetal distress (Gonçalves et al., 2006; van Laar et al., 2010, 2011; Spilka et al., 2012; Georgieva et al., 2013; Abry et al., 2013; Xu et al., 2014; Warmerdam et al., 2016). Since the autonomic nervous system (ANS) regulates the heart rate, variations in the heart rate reveal autonomic regulation and might indirectly provide information on fetal distress. To quantify changes in HRV, several features have been developed in the literature that describe different aspects of HRV. For overviews of the most commonly used HRV features the reader is referred to (Voss et al., 2009) or (Bravi et al., 2011).

Most HRV features have been developed and validated for adults in controlled experiments. However, unlike for adults, changes in the exterior of the fetus cannot be controlled during labor. Uterine contractions can lead to a temporary reduction of oxygen supply to the fetus (e.g. due to umbilical cord occlusion or reduced maternal placental blood flow). Besides, uterine contractions can cause large fluctuations in the intrauterine pressure, directly influencing the fetal cardiovascular system. As labor progresses into a stage of active pushing, the strength and influence of contractions increases.

In Romano et al. (2006) and Cesarelli et al. (2010), it was shown for healthy fetuses that HRV was significantly higher during contractions as compared to rest periods (i.e. the period in between contractions). Moreover, Warrick et al. (2010) used an input-output model to describe the relation between the uterine activity (UA) and FHR signal, and showed that features extracted from this model provided information about fetal distress. The change in FHR during contractions could indicate a healthy response of the fetal ANS to stabilize the cardiovascular system. In a recent study we showed that the differences between HRV features during contractions and rest might be used to improve the detection of fetal distress (Warmerdam et al., 2016).

In (Spilka et al., 2012; Georgoulas et al., 2006; Georgieva et al., 2013; Xu et al., 2014) it was shown that the detection of fetal distress can be improved by classifying combinations of multiple HRV features. Therefore, in this paper we examine whether the detection of fetal distress can be improved by combining HRV features that were calculated over the entire heart rate with HRV features that were calculated separately during contractions and during rest periods. To determine the most informative subset of HRV features, both with and without the additional contraction-dependent HRV features, we used a Genetic Algorithm (GA) (Mitchell, 1998). Furthermore, the classification performance of these subsets for detection of fetal distress was evaluated over time, using Support Vector Machines (SVMs) (Shawe-Taylor and Cristianini, 2000).

The rest of this paper is organized as follows; Section 2.1 discusses the data acquisition and pre-processing. In Sections 2.2 and 2.3 we summarize the implementation of the used HRV features. Feature selection, classification, and validation are explained in Sections 2.4-
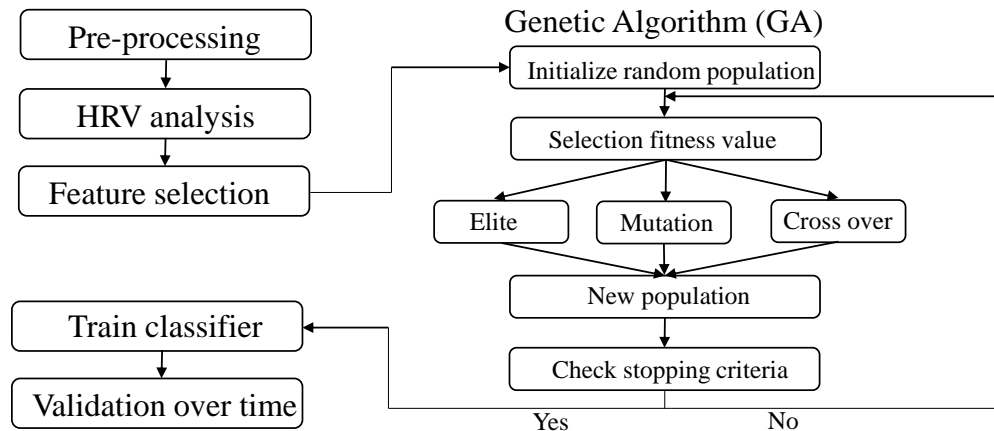
Figure 1. Flowchart of signal processing.

2.6. Results and discussion are presented in Sections 3 and 4.

2. Methods

This section describes the different HRV features, feature selection, and the classification that was used in this paper. The overall classification procedure is presented in Fig. 1.

2.1. Data acquisition and pre-processing

We used FHR and UA signals that were recorded on a Neoventa STAN® (Mölndal, Sweden) in the Máxima Medical Center in Veldhoven, the University Medical Center in Utrecht, and registrations collected in Sweden in the context of STAN evaluation over the period 2001-2008. Registrations of at least 36 weeks of gestation were included. Pregnancies complicated by intra-uterine growth restriction, fever, fetal congenital anomalies, or use of ritodrine were excluded. We used beat-to-beat FHR signals that were recorded by a scalp electrode. UA was recorded with either a tocodynamometer or intra-uterine pressure catheter. Contractions were annotated by a clinical expert that had no knowledge of the FHR signal or fetal outcome.

Because the beat-to-beat R-R intervals are determined when a heartbeat occurs, the R-R intervals are not equidistantly sampled. We used linear interpolation to obtain an equidistantly sampled FHR signal, sampled at a rate of 4 Hz. If an R-R interval changed by more than 25 BPM or more than 200 ms with respect to the adjacent R-R interval, the R-R interval was considered an artifact. Artifacts were corrected by linear interpolation.

Each registration was divided into segments of 10 minutes, with five minute overlap. If a segment contained less than 20% loss of FHR signal and at least annotation of two contractions and two rest periods, the segment was used for further analysis. We only included FHR signal during the second stage of labor (stage of active pushing), since during this stage contractions have a strong effect on the fetus. Besides, because we expect that the effects of oxygen deficiency increase as labor progresses, we only considered FHR segments up to maximally 45 minutes before birth. Since we are also interested in the classification

Table 1. Clinical characteristics of included fetuses.

Clinical features	healthy fetuses (cases n=80)	fetuses with adverse outcome (cases n=20)
Gestational age (days)	280 ± 10	283 ± 8
Birth weight (g)	3551 ± 471	3560 ± 467
1-minute Apgar score	8.6 ± 1.1	6.3 ± 2.0
5-minute Apgar score	9.5 ± 0.5	7.8 ± 2.0
Cord arterial pH	7.25 ± 0.03	7.00 ± 0.04
Cord arterial base excess (mmol/l)	-5.8 ± 2.3	-16.0 ± 2.2
Length second stage of labor (min)	58 ± 24	62 ± 24

Values are expressed as mean and standard deviation.

performance over time, only registrations were included with at least 3 usable 10 minute segments within the final 45 minutes before birth.

The fetal outcome was based on the umbilical arterial acid-base state. Adverse outcome was defined as a pH below 7.05 and base excess below -12 mmol/l, and healthy as a pH above 7.20 (Sundström et al., 2000). Out of 1232 registrations, our dataset contained 44 cases with adverse outcome. Of these 44 cases, 6 cases were excluded due to either intra-uterine growth restriction, fever, fetal congenital anomalies, or used of ritodrine. Another 6 had to be excluded because no information was available on the onset of the second stage of labor. Of the remaining 32 cases, 20 cases had sufficient quality FHR and UA signal, and were used for further analysis. The 20 cases with adverse outcome were matched with 80 healthy cases (1-to-4 ratio between the size of both groups). We have attempted to match the prevalence of certain medications that can cross the placenta and influence fetal HRV (e.g. anti-hypertensives or pain killers) in accordance with the difference in size of the two groups. The main clinical characteristics of the included fetuses is shown in Table 1.

2.2. HRV analysis

For each 10 minute FHR segment, conventional HRV features were calculated over the resampled FHR signal. The HRV features used in this paper have been described in detail in our previous study (Warmerdam et al., 2016) and only the implementation is discussed here. The used HRV features are listed in Table 2. In the rest of this paper we will denote the resampled FHR signal by $x[n]$, for samples $n = 1,...,N$, with N the length of the signal.

As shown by Peters et al. (2011), artifact correction by interpolation influences the HRV analysis. Excluding these periods for the computation of HRV features reduces the error made in the HRV analysis (Peters et al., 2011). This section also describes how we reduced the influence of artifacts for each HRV feature.

Statistical features   The standard deviation (*SD*) and the root mean square of successive differences (*RMSSD*) were calculated as statistical features. For the calculation of *SD* and *RMSSD* all artifact corrected samples were excluded.

Table 2. HRV features.

HRV features	abbreviation
Standard deviation	SD
Root mean square of successive differences	$RMSSD$
Low frequency power	LF
High frequency power	HF
Total power	TP
Normalized low frequency power	LF_n
Normalized high frequency power	HF_n
Sample entropy	$SampEn$
scaling exponent	α
Deceleration capacity	DC

Frequency domain features   Spectral analysis was performed to examine the heart rate at specific frequency bands that are related to sympathetic and parasympathetic activity (Van Laar et al., 2008). Frequency bands for analysis of fetal HRV differ from frequency bands for analysis of adult HRV (van Laar et al., 2009). Power in the low frequency band (LF, 0.04-0.15 Hz) is related to both sympathetic and parasympathetic activity, and power in the high frequency band (HF, 0.4-1.5 Hz) is related to parasympathetic activity only. Besides absolute LF and HF, also normalized frequency powers were calculated (LF_n and HF_n), by dividing LF and HF by the total power (TP, 0.04-1.5 Hz).

Spectral analysis was performed by the Continuous Wavelet Transform (CWT) with a fifth order symlet wavelet (Peters et al., 2011; Warmerdam et al., 2016). In this study, we used the complex CWT (Addison, 2002). Because the complex CWT also provides phase information, it is possible to calculate the spectral power at each sample, leading to more reliable calculation of LF_n and HF_n. Note that resampling of the beat-to-beat FHR signal to an equidistant signal may cause some distortion of the spectral analysis (Clifford and Tarassenko, 2005). Yet, we expect that the influence of this distortion on the outcome is limited since resampling was performed for all registrations.

CWT compares $x[n]$ to a family of analyzing wavelet functions $\psi[\frac{n-\tau}{s}]$, where s is a scaling parameter and τ the position parameter of the wavelet. The CWT coefficients $W[s, \tau]$ can be obtained by varying the scale parameter s and the position parameter τ. The power at each time instant and scale is proportional to the square of the CWT coefficients. The total power in a frequency band is then calculated by integrating the power over the scales within the frequency band of interest and averaging over time.

Each wavelet $\psi[\frac{n-\tau}{s}]$ has a support width that is related to the scaling parameter s. Similar to Peters et al. (2011), we considered one-third of support width of $\psi[\frac{n-\tau}{s}]$ as the effective support width. A CWT coefficient was excluded for the calculation of HF if one or more samples within the effective support width was artifact corrected. To calculate LF, we excluded CWT coefficients when more than five seconds of consecutive FHR signal was artifact corrected within the effective support width.

Complexity features   The presence of irregularities in the heart rate is seen as healthy blood pressure control by the ANS. Complexity in short time signals can be quantified by Sample Entropy (*SampEn*) (Richman and Moorman, 2000). To calculate *SampEn*, the FHR is divided into vectors $\mathbf{u}[i] = \{x[i], x[i+1], ..., x[i+m-1]\}$. *SampEn* estimates the conditional probability that if patterns are similar for length $m-1$, they will also be similar for length m. A tolerance parameter r is used to define a threshold for similarity between two patterns. We excluded vectors $\mathbf{u}[i]$ from the calculation of *SampEn* if any of the samples $\{i, i+1, ..., i+m-1\}$ was artifact corrected. Similar to Gonçalves et al. (2006), Spilka et al. (2012) and (Warmerdam et al., 2016), the length of the vectors was set to $m = 2$ and the tolerance to $r = 0.2SD$.

Fractal features   To describe the scaling properties of HRV over different time scales, we used scaling exponent α, which is obtained from Detrended Fluctuation Analysis (DFA) (Ihlen, 2012). To calculate α, first the cumulative sum $y[n]$ of the FHR signal is obtained. Next, $y[n]$ is divided into equally spaced segments of length a and the linear trend of each segment ($y_a[n]$) is determined. The characteristic fluctuation F at each scale a is computed as the root mean square (RMS) deviation between $y[n]$ and its trend $y_a[n]$.

     The scaling exponent α is given by the slope of the relation between $\log(F[a])$ and $\log(a)$. Similar to Abry et al. (2013), α was computed over the scales ranging from 1 to 60 seconds, which corresponds to the physiological range of the baroreflex.

     Artifact corrected samples are excluded from the calculation of the linear trend $y_a[n]$ and the characteristic fluctuation F. Moreover, if the percentage of artifact correction within a segment of length a is more than 20%, it is excluded entirely.

Phase rectified signal averaging   In addition to the HRV features that were included in our previous study (Warmerdam et al., 2016) and have been described above, we also used Phase Rectified Signal Averaging (PRSA). PRSA has shown promising results for the detection of fetal distress (Georgieva et al., 2014) and (Xu et al., 2014). The basic idea of PRSA is to account for non-stationarities in the heart rate by aligning heart rate segments based on predefined events, called anchor points (ν). For analysis of the heart rate, anchor points are typically defined as decelerations and accelerations in the heart rate, and PRSA quantifies the average response to a deceleration (the deceleration capacity, *DC*) or acceleration (the acceleration capacity, *AC*). It was shown by Georgieva et al. (2014) that *DC* and *AC* are highly correlated and that their performance for detection of fetal distress is similar. Therefore, we only included *DC* in our analysis.

     PRSA depends on two parameters: a filter parameter T and a window parameter L (Kantelhardt et al., 2007). Similar to Georgieva et al. (2014) we used $T = 5$ and $L = 45$. Artifact corrected samples were not used as anchor points. Moreover, when there was more than 20% artifact correction within a window $\nu \pm L$, the anchor point ν was also excluded.
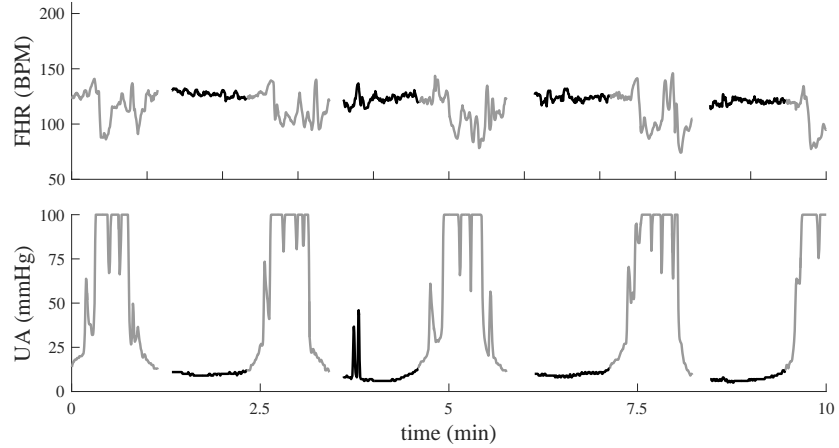
Figure 2. Example of an FHR (upper graph) and UA signal (bottom graph) of healthy fetus. Black lines are rest periods and gray lines contractions.

2.3. Contraction-dependent HRV features

In addition to calculating HRV features over the entire FHR signal, we also calculated HRV features separately during contractions and rest periods. Using the expert annotations, the rest period was defined as the period in between two contractions. No contraction or rest period was defined if the length of a contraction or the time between contractions was less than 20 seconds. Note that the fetus typically requires some recovery time directly after a contraction, but has enough time to recover before a new contraction. Therefore, if the time between two contractions was more than one minute, we only considered the final minute preceding the second contraction as rest period. In Fig. 2 an example is shown of an FHR signal that is recorded during contractions and during rest periods.

HRV features during contractions were based on R-R intervals that were recorded during contractions (denoted as $feature_{uc}$). Similarly, HRV features during rest periods were based on R-R intervals that were recorded during rest periods (denoted as $feature_{rest}$). Also the ratio between HRV features calculated during contractions and during rest was included (denoted as $feature_{ratio} = feature_{uc}/feature_{rest}$), as the ratios have shown promising results in our previous studies (Warmerdam et al., 2016). Because the time between contractions or the length of the contractions was often less than one minute, we could not calculate LF, LF_n, HF_n, α and DC separately during contractions and during rest periods. Our analysis of contraction-dependent HRV features was therefore limited to SD, $RMSSD$, HF, and $SampEn$.

In total, 22 HRV features were used: 10 calculated over the entire FHR, 4 during rest, 4 during contractions, and 4 ratio features, as presented in Table 3. We defined HRV feature set $S_1 \in \mathbb{R}^{10}$ as the set of HRV features that are calculated over the entire FHR, without distinguishing contractions and rest periods. The set of HRV features that contain the contraction-dependent HRV features is defined as $S_2 \in \mathbb{R}^{12}$. The set of HRV features that contains both HRV features that are calculated over the entire FHR signal and contraction-dependent HRV features is defined as $S_3 = S_1 \cup S_2$.

Table 3. HRV feature sets.

feature set	HRV features
$S_1 \in \mathbb{R}^{10}$	$SD, RMSSD, LF, HF, TP, LF_n, HF_n, SampEn, \alpha, DC$
$S_2 \in \mathbb{R}^{12}$	$SD_{uc}, RMSSD_{uc}, HF_{uc}, SampEn_{uc}$
	$SD_{rest}, RMSSD_{rest}, HF_{rest}, SampEn_{rest}$
	$SD_{ratio}, RMSSD_{ratio}, HF_{ratio}, SampEn_{ratio}$
$S_3 \in \mathbb{R}^{22}$	$S_1 \cup S_2$

2.4. Feature selection

Fulcher et al. (2012) showed that many HRV features that are described in Section 2.2 are correlated and might therefore contain redundant information. Besides, our dataset is relatively small, with only 20 cases of fetuses with adverse outcome. As a consequence, there is a risk of overfitting when too many HRV features are used for classification. Therefore, we used a Genetic Algorithm (GA) to select the best combination of HRV features. GA has the advantage that it can effectively explore the entire feature space, without running through all possible combinations of features and it has successfully been used for detection of fetal distress by Xu et al. (2014).

A schematic overview of GA is presented on the right side of Fig. 1. GA iteratively creates populations of candidate feature subsets (genotypes) by randomly combining and mutating genotypes of the population in a previous iteration, until certain convergence criteria are reached. In each iteration, the genotypes of the current population are scored by a fitness function. Genotypes are then selected at random to generate a new population with a probability that is proportional to their fitness value. Each new population consists of elite children (best performing genotypes that are guaranteed to survive to the next iteration), children that are produced by random mutations of a single genotype, and children that are produced by random crossover between pairs of genotypes.

For the implementation of GA, we used *ga* from the standard Global Optimization Toolbox of Matlab® (The Mathworks, Inc. Natick, MA). As fitness function, we used the performance of an SVM classifier, which is explained in Section 2.5. Based on the ranking in fitness value, genotypes are selected using a roulette system (Mitchell, 1998). Genotypes with equal scores were given equal rank and only the best 50% was used to generate the next population. Similar to Xu et al. (2014), the population size was set to 100, two elite children were used, and the crossover fraction was set to 80%. The remaining 20% for the new population was obtained by single point mutation. If two equal genotypes were selected for crossover, single point mutation is used instead to generate the child. To prevent GA from reaching a local minimum with a relatively large number of selected features, the number of selected features of a genotype was restricted to a maximum of five features. The GA was terminated if the best fitness value of the population did not improve for 10 iterations or in case 200 iterations were done. The settings for GA are summarized in Table 4.

Table 4. GA settings.

Parameter	Value
Population size	100
Number of elite children	2
Crossover fraction	80%
Mutation function	single point
fitness value	SVM performance
fitness function	rank
Genotype selection	roulette system
Maximum number of features in genotype	5
Maximum iterations with same fitness value	10
Maximum iterations	200

Table 5. SVM settings.

Parameter	Value
Kernel function	rbf
Kernel width (σ)	1
Penalty factor (C)	1
Optimization metric	g
Misclassification cost majority class	0.25
Misclassification cost minority class	1

2.5. Classification

We used SVMs to classify healthy fetuses and fetuses with adverse outcome. SVMs are supervised learning models that construct a set of hyperplanes that minimize the classification error while maximizing the distance between the hyperplanes and its nearest data points (Shawe-Taylor and Cristianini, 2000). We used a Gaussian radial base function (rbf) to allow for non-linear decision boundaries. To implement the SVMs, we used the standard implementation *fitcsvm* in the Statistics and machine learning toolbox of Matlab®. We used the default settings of the SVM for the width of the Gaussian kernel function ($\sigma = 1$) and the penalty parameter for misclassification ($C = 1$).

The imbalance in the distribution of the two classes in our dataset causes the SVM to be accurate for classification of the majority class (healthy fetuses) but to perform poorly for classification of the minority class (fetuses with adverse outcome) (He and Garcia, 2009). To prevent poor classification of the minority class, we defined a cost function based on the imbalance in the class distribution: the cost for misclassification of the minority class was thus four times higher than misclassification for the majority class. The SVM settings are summarized in Table 5.

2.6. Validation

In case of an imbalanced dataset, the accuracy (*Ac*) is not the best metric to evaluate the classification performance (He and Garcia, 2009) and (Georgoulas et al., 2006). Instead, we used the geometric mean (*g*) as metric to train the classifier and as fitness function for the feature selection in GA. The metric *g* provides a balance between the classification accuracy of the majority class (specificity, *Sp*) and the classification accuracy of the minority class (sensitivity, *Se*), and is defined as $g = \sqrt{Sp \cdot Se}$ (He and Garcia, 2009). The SVM rbf is a soft-type classifier, meaning that the output of the classifier can be a continuous value. Changing the threshold settings of the classifier will vary *Se* and *Sp*. Receiver operating characteristic (ROC) curves can be used to evaluate how the classifier performs for various classification threshold settings. The area under the ROC curve (*AUC*) can be used as a metric to evaluate the classification performance for various classification threshold settings.

For feature selection, the FHR segments of all registrations were used that were closest to birth. We used the 10-fold cross validation performance of the SVM as fitness value in GA. Note that for feature selection we only used the 10 minute FHR segment of each fetus that was closest to birth, meaning that from each fetus only a single FHR segment was used. The 10 minute FHR segment was either included in the training or in the test partition, but never in both. Using 10-fold cross validation ensures that the fitness value for genotypes is not overestimated. Due to our relatively small dataset, containing only 20 cases of fetuses with adverse outcome, feature selection and classification could be influenced by the location where the dataset was split into ten fold. To obtain a more objective result, feature selection was repeated 50 times, starting from different data splits. Hence, we generated 50 subsets of HRV features with 50 outcome measures for sets S_1, S_2, and S_3.

Besides classification performance for the FHR segments closest to birth, we are also interested in the classification performance over time. To this end, a classifier was trained for each generated feature subset on the FHR segments closest to birth. Then, the trained classifiers were used to classify if a fetus was in distress or not for all FHR segments within a registration, starting from 45 minutes before birth up to the time of birth. Since the purpose of the classifier would be to initiate an intervention in case an FHR segment is classified as fetal distress, all remaining FHR segments of that registration up to the time of birth were also indicated as fetal distress.

At each time instant the *g*, *Se*, and *Sp* for a classifier were calculated. Note that since not all registrations have 45 minutes of active pushing, the outcome measures were calculated with respect to the total number of registrations with at least one FHR segment up to that time (as shown in Fig. 3). Furthermore, for some recordings we were unable to analyse the final minutes before birth, either due to loss of FHR signal or insufficient quality UA signal. For these recordings, we retained the classification value of the final available segment up to the time of birth.
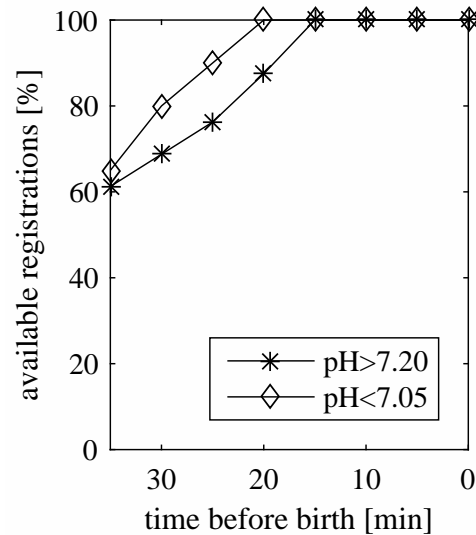
Figure 3. Percentage of registrations with at least one FHR segment up to that time.
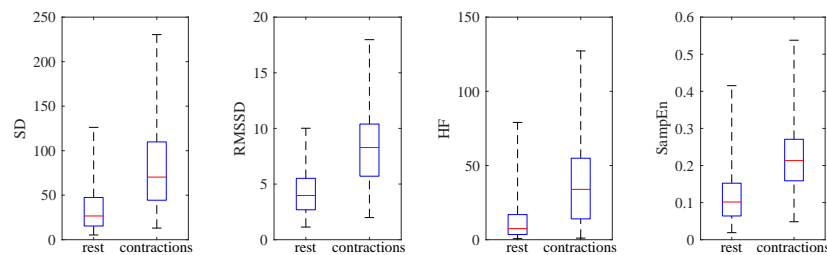

Figure 4. Comparison of HRV features measured during rest periods and contractions. Results are shown as boxplots where the central marker indicates the median and the edges of the box the interquartile range. The whiskers correspond to the full range of the data.

3. Results

Figure 4 show boxplots of the HRV features that were calculated separately during contractions and rest periods (*SD*, *RMSSD*, *HF*, and *SampleEn*) for the entire dataset. The correlation matrix of the HRV features used in this study is presented in Figure 5. Figure 6 shows the HRV features that were selected in the 50 runs by GA. From S_1 (only HRV features that are calculated over the entire FHR segments), TP, LF_n, HF_n, and DC were most frequently selected. In case of S_2 (only contraction-dependent HRV features) SD_{ratio} and $RMSSD_{uc}$ were most frequently selection. In case of S_3 (combined set of HRV features in S_1 and S_2), SD_{ratio} and HF were most frequently selected. The average size of selected subsets from S_1 was 3.0, from S_2 2.6, and from S_2 3.5 HRV features.

The average cross validation performance for classification of the FHR segments closest to birth is shown in Table 6. The average g for S_1, S_2, and S_2 were $g = 70\%$, $g = 76\%$, and $g = 79\%$, respectively. Finally, Fig. 7 shows the classification performance over time.
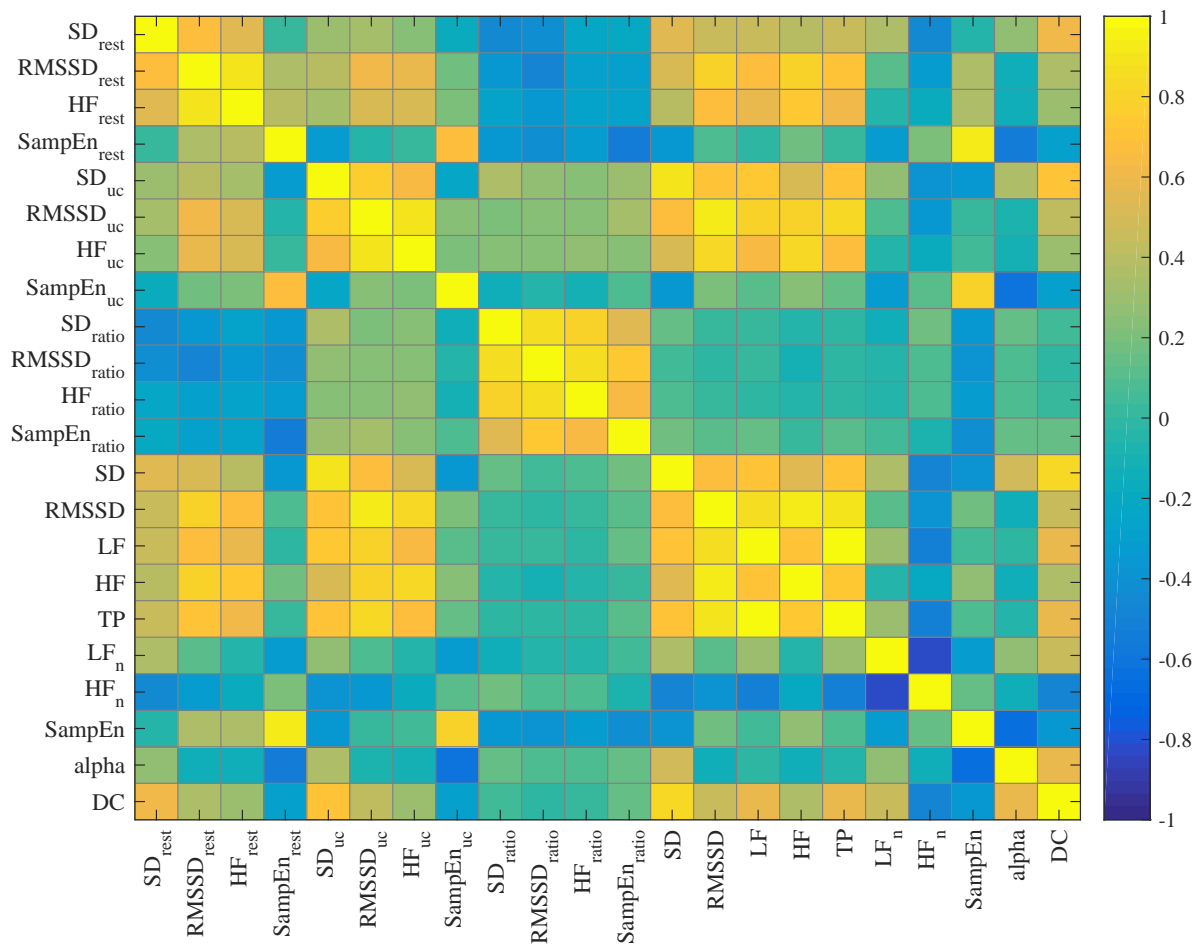
Figure 5. Correlation matrix of the HRV features in S_3. The color indicates the strength of the correlation.

**Table 6.** Average 10-fold cross validation performance for classification of the FHR segments closest to birth.

Feature set	g (%)	Se (%)	Sp (%)	Ac (%)	AUC
S_1	70	66	76	74	69
S_2	76	75	76	76	78
S_3	**79**	**81**	**77**	**78**	**80**

4. Discussion

Analysis of fetal HRV provides information on fetal distress. Most HRV features have been developed and validated for adults in controlled experiments. However, during labor the fetal cardiovascular system is strongly influenced by contractions. As can be seen in Fig. 4, HRV features are higher during contractions compared to HRV features during rest periods. In this study we showed that separating contractions from rest periods improves HRV analysis for the detection of fetal distress during labor.
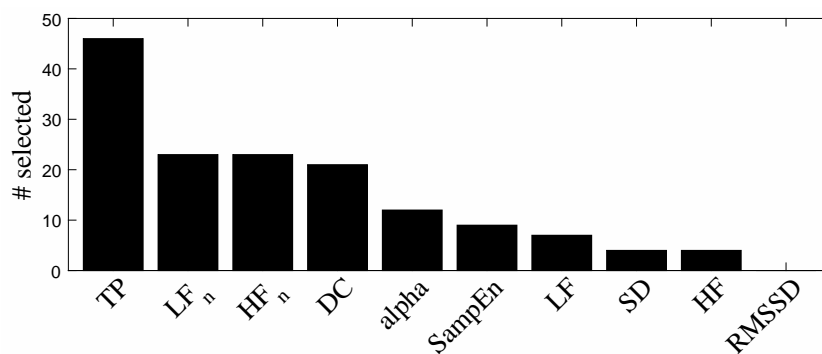
(a) Selected HRV features from S_1.
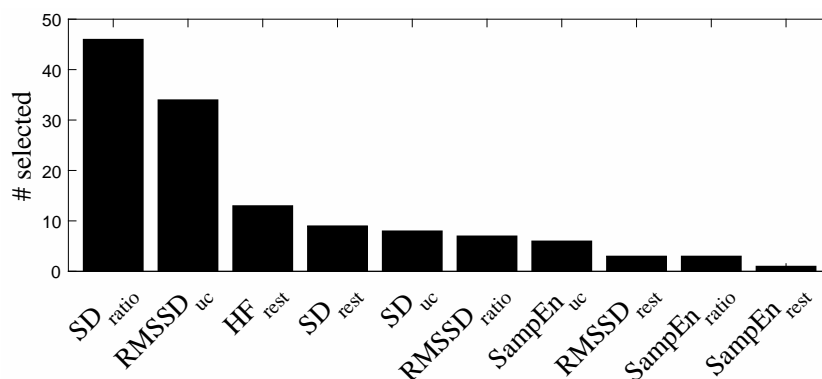
(b) Selected HRV features from S_2.
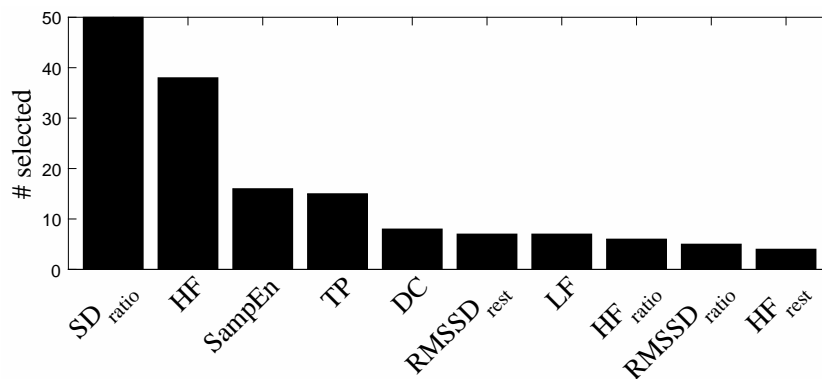
(c) Selected HRV features from S_3.

Figure 6. Results feature selection by GA.

4.1. Data acquisition

Out of the relatively large number of available registrations, we could only include 20 cases of fetuses with adverse outcome. This is partly due to the low incidence of fetuses with pH below 7.05 and partly due to our strict inclusion criteria. Moreover, we required a good quality UA signal to be able to clearly identify contractions. Because in clinical practice the UA signal is often of poor quality, this requirement limits the application of contraction-dependent HRV features. A potential solution to improve the quality of the UA signal might be to use electrohysterography to record the UA instead of a tocodynamometer (Euliano et al.,
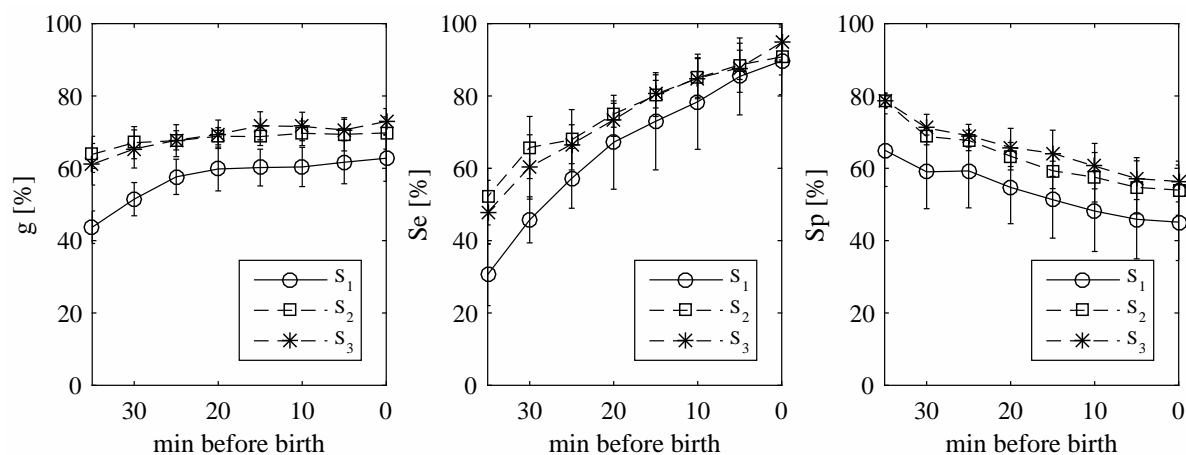
Figure 7. Average performance classification over time.

2013) and (Vlemminx et al., 2017). In addition to good quality UA signal, less than 20% loss of FHR signal was allowed. To minimize the influence of artifact correction of the FHR signal, periods with artifact corrected FHR were excluded from the calculation of HRV features.

In this study we only focussed on FHR segments during the second stage of labor. During the second stage, contractions generally have a strong effect on the fetus. In contrast, during the first stage of labor (stage of cervical dilation), the effect of contractions is less pronounced. Therefore, a classifier that is trained on FHR from the second stage of labor might be less predictive during the first stage of labor. For future work it would be interesting to also examine the classification performance during the first stage of labor.

Because the goal of this study was to examine whether separating contractions from rest periods can improve the HRV analysis, we limited our dataset to two groups that are well separated in terms of fetal outcome and did not consider registrations with pH between 7.05 and 7.20. However, this approach simplifies the classification task and does not represent true clinical conditions. Before a classifier can be used in clinical practice, future study should use a dataset that contains registration with all pH values.

4.2. Feature selection

In Fig. 5 it can been seen that several of the HRV features that were used in this study were correlated. In particular, HRV features calculated on the entire FHR signal were correlated to their counterpart that was calculated during contractions or during rest periods (e.g. *SD* was correlated to SD_{rest} and S_{uc}). To select the best combination of HRV features for detection of fetal distress we used GA. The average size of selected subsets from S_1, S_2, and S_3 (3.0, 2.6, and 3.5, respectively) was smaller than the maximum number of five features that was allowed for GA, indicating that larger feature subsets would not have resulted in better classification.

Because GA uses the classification performance to score the candidate subsets, selected HRV features could perform differently for other classifiers. Furthermore, 10-fold cross validation was used for feature selection. Since feature selection could be influenced by

splitting the dataset, GA was repeated fifty times using different data splits. The influence of splitting the dataset on feature selection can be seen from Fig. 6 by the number of HRV features that were selected multiple times.

The most frequently selected HRV feature for S_1 was TP, and the most frequently selected HRV feature for S_2 and S_3 was SD_{ratio}. Although TP was selected most of the time from S_1, TP was only selected a few times for S_3, as shown in Fig. 6. Since both TP and SD_{ratio} are related to the presence of decelerations, these HRV features contain similar information. SD_{ratio} also contains information about how well the fetus is able to recover from contractions: a high SD_{ratio} indicates that the fetus quickly recovers from a contraction and is able to stabilize its cardiovascular system during the rest periods. On the other hand, a low ratio indicates that the variability during contractions and rest periods is similar, meaning that the fetus is unable to recover during the rest periods. In case SD_{ratio} is selected, the information of TP becomes redundant.

Other HRV features that were often selected for S_1 were LF_n, HF_n, and DC. van Laar et al. (2010) also found that normalized spectral powers were predictive of fetal distress. Since changes in normalized spectral powers are not masked by changes in the total power, normalized spectral powers can give better insight in cardiovascular control than absolute spectral powers.

Interestingly, although HF_n was frequently selected for S_1 it was not often selected for S_3. Instead, for S_3 HF was often selected in combination with SD_{ratio}. Since SD_{ratio} is related to the presence of decelerations and the total power, it could be that the combination of SD_{ratio} and HF also contains the information in HF_n, since SD_{ratio} is related to TP. In line with the selection of HF for S_3 is the selection of $RMSSD_{uc}$ and HF_{rest} for S_2, as the combined information of $RMSSD_{uc}$ and HF_{rest} is similar to the information of HF.

To gain more insight in cardiovascular control, it would be interesting to examine HF_n and LF_n during contractions and rest periods. However, the length of contractions or rest periods was often less than the required length to calculate LF_n and HF_n, and we were unable to calculate HF_n and LF_n separately during contractions and rest periods. Besides LF_n and HF_n, in this study DC was also frequently selected from S_1, similar to findings in (Xu et al., 2014). Since DC contains information about decelerations, it contains similar information as SD_{ratio} and was not frequently selected from S_3. In our results, scaling exponent α was selected less often. Recently, Chudacek et al. (2014) obtained promising results for detection of fetal distress by calculating scaling exponents using a wavelet based scattering transform. Abry et al. (2013) showed that the influence of decelerations on the scattering transform is limited and using the scattering transform might improve the fractal analysis.

4.3. Classification

Many factors can influence the response of the fetus to contractions. There are clinical parameters that can influence the FHR, such as gestational age (Laar et al., 2014), behavioral states (van Laar et al., 2009), or medication (Verdurmen et al., 2013). Furthermore, uterine contractions can directly influence the fetal cardiovascular system through a rise in

intrauterine pressure, or indirectly by blocking the oxygen supply to the fetus (e.g. umbilical cord occlusion) (van der Hout-van der Jagt et al., 2013). As labor progresses, the effect of contractions on the fetal cardiovascular system increases. As a consequence, in the literature many different FHR patterns have been described with different characteristics, several of which have been related to fetal distress (e.g. late decelerations, bradycardia, sinusoidal patterns, or saltatory patterns) (Sundström et al., 2000). The relatively low number of cases with adverse fetal outcome in combination with the large variation in FHR patterns makes training of a classifier challenging.

We did not compare our classification performance to visual CTG interpretation, because the inter- and intra-observer variability of CTG interpretation is generally high (Blix et al., 2003). Besides, in the final minutes before birth the CTG is typically abnormal in most registrations. Hence, results obtained from visual CTG interpretation might in this case not be representative.

The average 10-fold cross validation performance for classification of the FHR segments closest to birth improved from $g = 70\%$ for S_1 to $g = 76\%$ for S_2, and $g = 79\%$ for S_3. The difference is less pronounced in terms of Ac ($Ac = 74\%$ for S_1, $Ac = 76\%$ for S_3, and $Ac = 78\%$ for S_3), because Ac is mostly determined by correct classification of the majority class. Note that the classifier was trained to optimize g, which is a balance between Se and Sp. Depending on the use of the classifier in clinical practice, it would be more appropriate to train a classifier that is either more sensitive to misclassification of the minority group (and increase Se) or misclassification of the majority group (and increase Sp). This can be achieved by changing the penalty parameter for misclassification or changing the threshold settings of the SVM. As AUC is higher for S_2 and S_3 compared to S_1 ($AUC = 69\%$ for S_1, $AUC = 78\%$ for S_2, and $AUC = 80\%$ for S_3), using S_2 and S_3 will perform better for varying settings for the penalty parameter or classification threshold.

The earlier a prediction, the more useful the information for clinical intervention. For classification over time, the performance of S_2 and S_3 were similar, although performance of S_3 was slightly better. The average g over time for S_3 increased by 12% with respect to the average g of S_1 (57% for S_1, 68% for S_2, and 69% for S_3). At 15 minutes before birth, the classification performance for S_1 was $g = 60\%$, for S_2 $g = 69\%$, and for S_3 $g = 72\%$. For all sets S_1, S_2, and S_3, classification performance over time was lower than the classification performance for the FHR segments closest to birth. It should be noted that fetal distress develops gradually over time (Fleischer et al., 1982). Since it is unclear at which point the fetus is no longer capable of handling the stress, the relatively low Se at 35 minutes before birth ($Se = 31\%$ for S_1, $Se = 52\%$ for S_2, and $Se = 48\%$ for S_3) could be because at that time some of the fetuses with adverse outcome were still relatively healthy. Further study using a larger dataset is required to gain more insight into which combination of HRV features is the most informative and to improve the classification.

5. Conclusion

Fetal HRV analysis provides information on fetal distress. Combining HRV features calculated over the entire fetal heart rate with contraction-dependent HRV features improves the classification performance during the second stage of labor. Further studies are required to gain more insight in the which combination of HRV features is most informative.

Acknowledgments

References

Abry, P., Roux, S., Chudácek, V., Borgnat, P., Goncalves, P. and Doret, M. (2013). Hurst exponent and intrapartum fetal heart rate: Impact of decelerations, *Computer-Based Medical Systems (CBMS), 2013 IEEE 26th International Symposium on*, IEEE, pp. 131–136.

Addison, P. (2002). *The illustrated wavelet transform handbook: introductory theory and applications in science, engineering, medicine and finance*, CRC press.

Alfirevic, Z., Devane, D. and Gyte, G. (2006). Continuous cardiotocography (CTG) as a form of electronic fetal monitoring (EFM) for fetal assessment during labour., *Cochrane Database Syst Rev* **3**: CD006066.

Ayres-de Campos, D., Spong, C. Y. and Chandraharan, E. (2015). Figo consensus guidelines on intrapartum fetal monitoring: Cardiotocography, *International Journal of Gynecology & Obstetrics* **131**(1): 13–24.

Blix, E., Sviggum, O., Koss, K. and Øian, P. (2003). Inter-observer variation in assessment of 845 labour admission tests: comparison between midwives and obstetricians in the clinical setting and two experts, *BJOG: An International Journal of Obstetrics & Gynaecology* **110**(1): 1–5.

Bravi, A., Longtin, A. and Seely, A. (2011). Review and classification of variability analysis techniques with clinical applications, *Biomed Eng Online* **10**(1): 90.

Cesarelli, M., Romano, M., Ruffo, M., Bifulco, P. and Pasquariello, G. (2010). Foetal heart rate variability frequency characteristics with respect to uterine contractions, *JBiSE* **03**(10): 10141021.

Chudacek, V., Andén, J., Mallat, S., Abry, P. and Doret, M. (2014). Scattering transform for intrapartum fetal heart rate variability fractal analysis: A case-control study, *Biomedical Engineering, IEEE Transactions on* **61**(4): 1100–1108.

Clifford, G. D. and Tarassenko, L. (2005). Quantifying errors in spectral estimates of hrv due to beat replacement and resampling, *IEEE transactions on biomedical engineering* **52**(4): 630–638.

Euliano, T., Nguyen, M., Darmanjian, S., McGorray, S., Euliano, N., Onkala, A. and Gregg, A. (2013). Monitoring uterine activity during labor: a comparison of 3 methods, *American journal of obstetrics and gynecology* **208**(1): 66.e1–6.

Fleischer, A., Schulman, H., Jagani, N., Mitchell, J. and Randolph, G. (1982). The development of fetal acidosis in the presence of an abnormal fetal heart rate tracing. i. the average for gestational age fetus., *American journal of obstetrics and gynecology* **144**(1): 55–60.

Fulcher, B. D., Georgieva, A. E., Redman, C. W. G. and Jones, N. S. (2012). Highly comparative fetal heart rate analysis., *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference* **2012**: 3135–3138.

Georgieva, A., Papageorghiou, A., Payne, S., Moulden, M. and Redman, C. (2014). Phase-rectified signal averaging for intrapartum electronic fetal heart rate monitoring is related to acidaemia at birth, *BJOG: An International Journal of Obstetrics & Gynaecology* **121**(7): 889–894.

Georgieva, A., Payne, S. J., Moulden, M. and Redman, C. (2013). Artificial neural networks applied to fetal monitoring in labour, *Neural Computing and Applications* **22**(1): 85–93.

Georgoulas, G., Stylios, C. and Groumpos, P. (2006). Classification of fetal heart rate using scale dependent features and support vector machines.

Gonçalves, H., Rocha, A., Ayres-de Campos, D. and Bernardes, J. (2006). Linear and nonlinear fetal heart rate analysis of normal and acidemic fetuses in the minutes preceding delivery, *Medical and Biological Engineering and Computing* **44**(10): 847–855.

He, H. and Garcia, E. A. (2009). Learning from imbalanced data.

Ihlen, E. (2012). Introduction to multifractal detrended fluctuation analysis in matlab, *Front. Physio.* **3**.

Kantelhardt, J. W., Bauer, A., Schumann, A. Y., Barthel, P., Schneider, R., Malik, M. and Schmidt, G. (2007). Phase-rectified signal averaging for the detection of quasi-periodicities and the prediction of cardiovascular risk, *Chaos: An Interdisciplinary Journal of Nonlinear Science* **17**(1): 015112.

Laar, J. O., Warmerdam, G. J., Verdurmen, K. M., Vullings, R., Peters, C. H., Houterman, S., Wijn, P. F., Andriessen, P., Pul, C. and Guid Oei, S. (2014). Fetal heart rate variability during pregnancy, obtained from non-invasive electrocardiogram recordings, *Acta obstetricia et gynecologica Scandinavica* **93**(1): 93–101.

Mitchell, M. (1998). *An introduction to genetic algorithms*, MIT press.

Peters, C., Vullings, R., Rooijakkers, M., Bergmans, J., Oei, S. and Wijn, P. (2011). A continuous wavelet transform-based method for time-frequency analysis of artefact-corrected heart rate variability data., *Physiol Meas* **32**(10): 1517–1527.

Richman, J. and Moorman, J. (2000). Physiological time-series analysis using approximate entropy and sample entropy, *American Journal of Physiology-Heart and Circulatory Physiology* **278**(6): H2039–H2049.

Romano, M., Bifulco, P., Cesarelli, M., Sansone, M. and Bracale, M. (2006). Foetal heart rate power spectrum response to uterine contraction, *Medical and Biological Engineering and Computing* **44**(3): 188–201.

Shawe-Taylor, J. and Cristianini, N. (2000). Support vector machines, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods* pp. 93–112.

Spilka, J., Chudáček, V., Kouckỳ, M., Lhotská, L., Huptych, M., Janků, P., Georgoulas, G. and Stylios, C. (2012). Using nonlinear features for fetal heart rate classification, *Biomedical Signal Processing and Control* **7**(4): 350–357.

Sundström, A., Rosén, D. and Rosén, K. (2000). Fetal surveillance, *Technical report*, Gothenburg: Neoventa Medical AB.

van der Hout-van der Jagt, M. B., Jongen, G. J., Bovendeerd, P. H., Oei, S. G. et al. (2013). Insight into variable fetal heart rate decelerations from a mathematical model, *Early Human Development* **89**(6): 361–369.

van Laar, J., Peters, C., Houterman, S., Wijn, P., Kwee, A. and Oei, S. (2011). Normalized spectral power of fetal heart rate variability is associated with fetal scalp blood ph., *Early Hum Dev* **87**(4): 259–263.

van Laar, J., Peters, C., Vullings, R., Houterman, S., Bergmans, J. and Oei, S. G. (2010). Fetal autonomic response to severe acidaemia during labour., *BJOG* **117**(4): 429–437.

van Laar, J., Peters, C., Vullings, R., Houterman, S. and Oei, S. (2009). Power spectrum analysis of fetal heart rate variability at near term and post term gestation during active sleep and quiet sleep., *Early Hum Dev* **85**(12): 795–798.

Van Laar, J., Porath, M., Peters, C. and Oei, S. (2008). Spectral analysis of fetal heart rate variability for fetal surveillance: review of the literature., *Acta Obstet Gynecol Scand* **87**(3): 300–306.

Verdurmen, K. M., Renckens, J., van Laar, J. O. and Oei, S. G. (2013). The influence of corticosteroids on fetal heart rate variability: a systematic review of the literature, *Obstetrical & gynecological survey* **68**(12): 811–824.

Vlemminx, M. W., Thijssen, K. M., Bajlekov, G. I., Dieleman, J. P., Van Der Hout-Van, M. B., Jagt, D. and OEI, S. G. (2017). Electrohysterography for uterine monitoring during term labour compared to external tocodynamometry and intra-uterine pressure catheter, *European Journal of Obstetrics & Gynecology and Reproductive Biology* .

Voss, A., Schulz, S., Schroeder, R., Baumert, M. and Caminal, P. (2009). Methods derived from nonlinear dynamics for analysing heart rate variability, *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **367**(1887): 277–296.

Warmerdam, G., Vullings, R., Van Laar, J., Van der Hout-Van der Jagt, M., Bergmans, J.,

Schmitt, L. and Oei, S. (2016). Using uterine activity to improve fetal heart rate variability analysis for detection of asphyxia during labor, *Physiological measurement* **37**(3): 387.

Warrick, P., Hamilton, E., Precup, D. and Kearney, R. (2010). Classification of normal and hypoxic fetuses from systems modeling of intrapartum cardiotocography, *Biomedical Engineering, IEEE Transactions on,* **57**(4): 771–779.

Xu, L., Redman, C., Payne, S. and Georgieva, A. (2014). Feature selection using genetic algorithms for fetal heart rate analysis, *Physiological Measurement* **35**(7): 13571371.