IOP Publishing Physiol. Meas. 43 (2022) 025008 https://doi.org/10.1088/1361-6579/ac4e6d

Physiological Measurement

IPEM Institute of Physics and Engineering in Medicine

PAPER

Detection of fetal arrhythmias in non-invasive fetal ECG recordings using data-driven entropy profiling

RECEIVED
14 October 2021

REVISED
9 January 2022

ACCEPTED FOR PUBLICATION
24 January 2022

PUBLISHED
21 March 2022

Emerson Keenan[1,2], **Chandan Karmakar**[3], **Radhagayathri K Udhayakumar**[3], **Fiona C Brownfoot**[2], **Igor Lakhno**[4], **Vyacheslav Shulgin**[5], **Joachim A Behar**[6] and **Marimuthu Palaniswami**[1]

[1] Department of Electrical and Electronic Engineering, The University of Melbourne, Melbourne, VIC 3010, Australia
[2] Department of Obstetrics and Gynaecology, The University of Melbourne, Heidelberg, VIC 3084, Australia
[3] School of Information Technology, Deakin University, Geelong, VIC 3220, Australia
[4] Obstetrics and Gynecology Department, Kharkiv Medical Academy of Postgraduate Education, Kharkiv, Ukraine
[5] Aerospace Radio-Electronic Systems Department, National Aerospace University, Kharkiv Aviation Institute, Kharkiv, Ukraine
[6] Biomedical Engineering Faculty, Technion-Israel Institute of Technology, Haifa, Israel

E-mail: e.keenan@ieee.org

Keywords: arrhythmia, entropy, fetal ECG, fetal monitoring, signal quality

Abstract

Objective. Fetal arrhythmias are a life-threatening disorder occurring in up to 2% of pregnancies. If identified, many fetal arrhythmias can be effectively treated using anti-arrhythmic therapies. In this paper, we present a novel method of detecting fetal arrhythmias in short length non-invasive fetal electrocardiography (NI-FECG) recordings. *Approach.* Our method consists of extracting a fetal heart rate time series from each NI-FECG recording and computing an entropy profile using a data-driven range of the entropy tolerance parameter *r*. To validate our approach, we apply our entropy profiling method to a large clinical data set of 318 NI-FECG recordings. *Main Results.* We demonstrate that our method (*TotalSampEn*) provides strong performance for classifying arrhythmic fetuses (AUC of 0.83) and outperforms entropy measures such as *SampEn* (AUC of 0.68) and *FuzzyEn* (AUC of 0.72). We also find that NI-FECG recordings incorrectly classified using the investigated entropy measures have significantly lower signal quality, and that excluding recordings of low signal quality (13.5% of recordings) increases the classification performance of *TotalSampEn* (AUC of 0.90). *Significance.* The superior performance of our approach enables automated detection of fetal arrhythmias and warrants further investigation in a prospective clinical trial.

1. Introduction

Fetal arrhythmias are a disorder in the rhythm of the fetal heart rate (FHR), occurring in up to 2% of pregnancies (Southall *et al* 1977). Although many are benign, up to 10% of arrhythmias may result in fetal morbidity and mortality (Krapp *et al* 2003). These arrhythmias can typically be categorized as bradycardias (FHR $<$ 100 bpm), tachycardias (FHR $>$ 180 bpm) or other irregularities in FHR rhythm (Strasburger 2000).

Currently, the primary technique used to diagnose fetal arrhythmias is fetal echocardiography (Allan *et al* 1983). This technique allows for simultaneous recording of both ventricular and atrial activity using pulsed wave Doppler, motion mode (M-mode) or tissue velocity imaging ultrasound (Hornberger 2007). However, the correct use of these techniques requires a trained sonographer to accurately position the ultrasound probe over the fetal heart and a perinatal cardiologist to interpret the ultrasound, limiting their utility to the clinical setting.

The detection of fetal arrhythmias is an important area of research, as if identified, many types of arrhythmia can be effectively treated with anti-arrhythmic agents (Wacker-Gussmann *et al* 2014). Unfortunately due to the transient nature of many arrhythmias, existing strategies are not ideally suited for their detection with undiagnosed fetal arrhythmias thought to be responsible for up to 10% of unexplained fetal deaths (Crotti *et al* 2013).

To improve available techniques for identifying fetal arrhythmias, several works have investigated the application of non-invasive fetal electrocardiography (NI-FECG) for this purpose (Komáromy *et al* 1977, Behar

et al 2019). Recent works have shown that processed NI-FECG recordings can be utilized by a perinatal cardiologist to accurately diagnose fetal arrhythmia (Behar *et al* 2019). However, cardiologist interpretation is time consuming and cannot be performed widely. As such, the ultimate aim is to develop an automated method for identifying fetal arrhythmias to maximize the sensitivity and speed of detection.

In the field of adult ECG analysis, the application of nonlinear entropy measures for automated arrhythmia detection has been well studied, beginning with the application of Approximate Entropy (*ApEn*) (Caswell Schuckers 1998). *ApEn* was first introduced to quantify system complexity by analyzing the self-similarity of a time series at an embedding dimension m (the segment size used for comparison), and a tolerance r (the threshold used to classify segments as similar). An increase in entropy thus corresponds to an increase in system complexity or irregularity (Pincus 1991, Mayer *et al* 2014). Nonlinear measures are preferred to linear measures for analyzing heart rate time series as it has been shown that the dynamics of these signals are inherently nonlinear in nature (Rajendra Acharya *et al* 2004).

Since the introduction of *ApEn*, several improvements have been suggested, including the introduction of Sample Entropy (*SampEn*) (Richman and Moorman 2000) which removed the self-matching bias of *ApEn* and Fuzzy Entropy (*FuzzyEn*) (Chen *et al* 2009), which reduced dependence on the tolerance parameter r by introducing a fuzzy similarity metric. However, both of these modifications still rely on the manual selection of an appropriate r value for each application, leading to unknown reliability for novel uses such as fetal arrhythmia detection.

To address this limitation, we recently introduced the method of entropy profiling (Udhayakumar *et al* 2017, 2018, 2019) which computes an entropy estimate by using a non-parametric, data-driven set of r values. This work expands on our previous conference paper (Keenan *et al* 2020), to assess the performance of our entropy profiling approach against existing entropy measures such as *SampEn* and *FuzzyEn* for detecting fetal arrhythmias in a large clinical data set of NI-FECG recordings.

2. Materials and methods

2.1. Data set

For assessing each entropy measure we will utilize 22 NI-FECG recordings from the open-access NIFEA data set (Behar *et al* 2019) acquired via Physionet (Goldberger *et al* 2000), and a private set of 296 NI-FECG recordings obtained at the Ukrainian Children's Cardiac Center in Kyiv. All recordings were captured between 2013 and 2017 with approval by the Bioethics Committee of the Kharkiv Medical Academy of Postgraduate Education and the Ukrainian Children's Cardiac Center and registered under the ID 0116U002865. The combined data set contains 19 recordings from fetuses with a diagnosed arrhythmia and 299 recordings from fetuses with no diagnosed arrhythmia (hence referred to as normal). Several participants provided multiple NI-FECG recordings at different time points with a total of 15 unique arryhthmic participants and 245 unique normal participants in the combined data set. Within the arrhythmic fetuses, the following types of arrhythmia as confirmed by echocardiography were present: four with irregular atrial rhythm, three with premature atrial contractions, three with premature junctional contractions, two with atrial tachycardia, one with paroxysmal junctional tachycardia, one with intermittent second degree atrioventricular block, and one case with both atrial tachycardia and premature atrial contractions. In this work, only recordings from singleton pregnancies were assessed to avoid uncertainty between twin fetuses FHR time series.

2.2. Preprocessing and FHR time series extraction

Each NI-FECG recording was acquired at 16 bit resolution and 500 Hz or 1000 Hz sampling rate using an electrode configuration in one of the two placement schemes shown in figure 1. The alteration in sampling rate and electrode configuration comes from technical improvements made over the duration of the data collection period (Behar *et al* 2019).

The typical recording length and gestational age for the normal and arrhythmic participants are shown in table 1 and figure 2. For 18 of the normal recordings, information on gestational age was not available and as such, have been excluded from the gestational age statistics.

Using each NI-FECG recording, we extract a FHR time series using the fetal QRS (*fQRS*) extraction technique developed by Varanini *et al* (2014). This technique is implemented using all abdominal sensors excluding the maternal chest sensor, beginning with a preprocessing stage which reduces baseline wander, impulsive artifacts and power-line interference at the fundamental frequency (50 Hz) and its subsequent three harmonics.

Following preprocessing, the fetal and maternal ECG signals are separated using independent component analysis (ICA) followed by singular value decomposition for cancelling the maternal ECG. The residual fetal

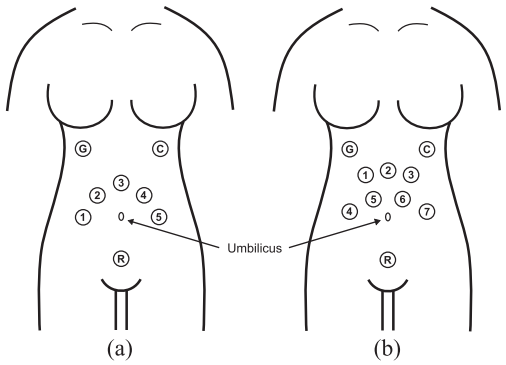

Figure 1. Electrode placement schemes in the utilized clinical data set where some NI-FECG recordings used electrodes in configuration (a), while others used electrodes in configuration (b). All leads were measured with respect to the reference sensor ®. © represents the ground electrode while © represents the maternal chest electrode (not utilized in this study).

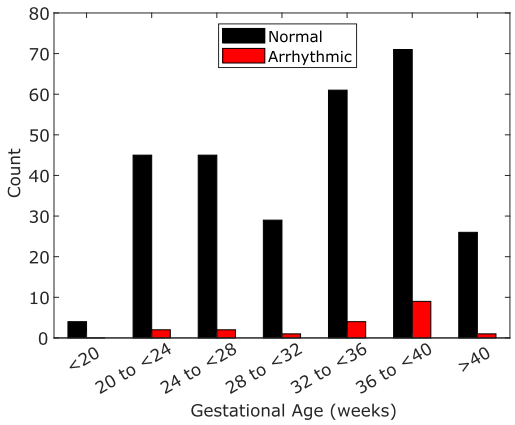

Figure 2. Distribution of gestational age for normal and arrhythmic recordings. *18 normal recordings were missing gestational age and have been excluded from the gestational age statistics.

Table 1. Characteristics of the normal and arrhythmic recordings.

	Normal	Arrhythmic
Number of recordings	299	19
Average recording length (*mm:ss*)	9:46	13:16
Gestational age in weeks ($\mu \pm \sigma$)	$31.3 \pm 6.5^{*}$	33.2 ± 5.4

ECG signals obtained from each sensor channel are then passed to a second ICA stage and a QRS detection process applied to all independent components to determine the potential *fQRS* locations.

The final *fQRS* locations are then selected from a single independent component on the basis of *a priori* knowledge of the typical FHR pattern by minimizing the first and second derivatives between adjacent *fQRS* intervals and the ratio of detected *fQRS* locations to matching maternal QRS complexes. For further details on the technique used for *fQRS* extraction, please refer to Varanini *et al* (2014).

The final *fQRS* locations are then transformed into a FHR time series, expressed in beats per minute (bpm), as follows:

$$FHR(i) = \frac{60}{fQRS(i+1) - fQRS(i)}, \quad i = 1, 2 \ldots N - 1,$$

where N is the number of *fQRS* detections and each FHR value is rounded to the nearest whole number. As this work focuses on the application of entropy measures in short length recordings, we consider up to the first 1000 FHR values extracted from each recording for analysis. The number of recordings which met each data length requirement (number of extracted FHR values) are shown in table 2.

Table 2. Distribution of NI-FECG recordings which met the required number of extracted FHR values.

# of FHR values	Normal	Arrhythmic	% of all recordings
10	299	19	100.0
25	299	19	100.0
50	299	19	100.0
100	295	18	98.4
250	280	18	93.7
500	213	18	72.6
1000	198	16	67.3

2.3. Entropy features

Using the extracted FHR values, a set of entropy features are computed for time series of length $N = 10,\ 25,\ 50,\ 100,\ 250,\ 500$ and 1000, considering the first N FHR values in each recording. Recordings which do not contain the minimum number of FHR values at a particular value of N are excluded from analysis. In previous studies, the required data length to produce reliable *SampEn* values for neonatal heart rate variability analysis is reported to be between 100 and 5000, with no defined limit documented for *FuzzyEn* (Lake *et al* 2002). However, in this work we will demonstrate that this requirement is an artificial limitation and that reliable entropy estimates can be determined for extremely short signal lengths. The entropy features compared in this work are as follows:

2.3.1. Sample entropy

Sample Entropy (*SampEn*), originally formulated by Richman and Moorman (2000), is a nonlinear measure of signal regularity used to assess diagnostic information from physiological signals such as heart rate time series. *SampEn* can be defined for a time series of length N as the conditional probability of two signal segments that are similar for m data points being similar at the next data point. The segment similarity is determined by the threshold r, commonly set to a multiple of the signal's standard deviation (SD). Thus, *SampEn* can be computed as follows, where m refers to the length of each signal segment:

(i) Form $(N - m)$ vectors, each of length m as follows:

$$\{X_i^m\colon 1 \leqslant i \leqslant (N - m)\},$$

where $X_i^m = \{x(i + k)\colon 0 \leqslant k \leqslant m - 1\}$

(ii) Take each X_i^m vector and find its distance from each vector X_j^m, where the inter-vector distance is given by:

$$d_{ij}^m = \{\max|X_i^m - X_j^m|\colon 1 \leqslant j \leqslant (N - m),\ j \neq i\}$$

(iii) Then we define:

$$\Phi^m(r) = \frac{1}{N - m}\sum_{i=1}^{N-m} C_i^m(r),$$

where $C_i^m(r)$ is the probability that a vector X_j^m lies within a distance r of the vector X_i^m for $j \neq i$.

(iv) Finally, *SampEn* is calculated as:

$$SampEn(N,\ m,\ r) = \ln\frac{\Phi^m(r)}{\Phi^{m+1}(r)}.$$

In this work, we represent *SampEn* as H_S.

2.3.2. Fuzzy entropy

Fuzzy Entropy (*FuzzyEn*) was subsequently proposed as an improvement to *SampEn* by Chen *et al* (2009). This refinement was achieved by using a fuzzy similarity metric in place of the fixed r threshold to determine segment similarity, thus reducing the impact of a specific choice of r on the entropy estimate. *FuzzyEn* can be computed as follows:

(i) For each inter-vector distance d_{ij}^m compute the similarity degree D_{ij}^m as:

$$D_{ij}^m = \exp(-\left(\frac{d_{ij}^m}{r}\right)^2).$$

(ii) Compute the probability of a vector X_j^m to be similar to a vector X_i^m by degree D_{ij}^m as:

$$C_i^m(r) = \frac{1}{N-m-1}\sum_{i=1,i\neq j}^{N-m} D_{ij}^m.$$

(iii) Then we define:

$$\Phi^m(r) = \frac{1}{N-m}\sum_{i=1}^{N-m} C_i^m(r).$$

(iv) Finally, *FuzzyEn* is calculated as:

$$FuzzyEn(N, m, r) = \ln\frac{\Phi^m(r)}{\Phi^{m+1}(r)}.$$

In this work, we represent *FuzzyEn* as H_F.

2.3.3. Entropy profiling

In this work, we also compute an entropy profile of each FHR time series using the recently introduced technique of cumulative histogram based entropy profiling, originally described in Udhayakumar *et al*(2018). For a given time series, if we represent D as the matrix containing all elements of d^m and d^{m+1}, *range* as the set of unique elements of D sorted in ascending order, and n_{bin} as the length of *range*, its entropy profile can be generated as follows:

(i) Calculate the cumulative distribution function:

$$cdf_{iq}^m = p(d_i^m \leqslant range(q)): 1 \leqslant q \leqslant n_{bin},$$

where each value of q is a distinct r value in the entropy profile and p is the probability of occurrence.

(ii) For $1 \leqslant q \leqslant n_{bin}$ compute:

$$\Phi^m(q) = \frac{1}{N-m}\sum_{i=1}^{N-m} cdf_{iq}^m.$$

(iii) Calculate *SampEn* for each q as:

$$SampEn(q) = \ln\frac{\Phi^m(q)}{\Phi^{m+1}(q)}.$$

(iv) Across all values of q, this process produces a profile of *SampEn* values. From this profile, we can derive secondary measures of signal regularity. In this work, we compute the total sample entropy (*TotalSampEn*) as the sum of all individual *SampEn* values in the profile:

$$TotalSampEn(N, m) = \sum_{q=1}^{n_{bin}} SampEn(q).$$

In this work, we represent *TotalSampEn* as H_{TS}.

2.3.4. Selection of m *and* r

For this work, both *SampEn* and *FuzzyEn* are evaluated at $r = 0.15 \times SD$ of the FHR time series and $m = 2$, typically suggested values for physiological signals (Yentes *et al* 2013), while *TotalSampEn* is evaluated at $m = 2$.

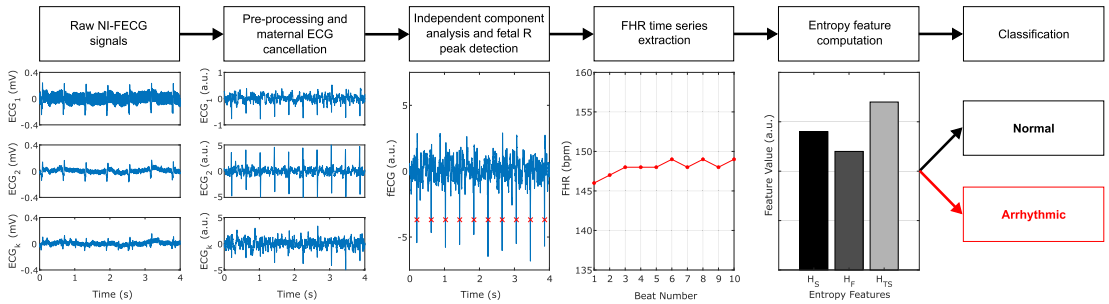

Figure 3. Overview of NI-FECG processing, FHR time series extraction, entropy feature computation and signal classification process for each recording showing from left to right: (1) raw NI-FECG signals, (2) NI-FECG signals after pre-processing and maternal ECG cancellation, (3) selected fetal ECG signal component after independent component analysis and detected *fQRS* locations (×), (4) fetal heart rate (FHR) time series computed from *fQRS* intervals, (5) entropy features *SampEn* (H_S), *FuzzyEn* (H_F) and *TotalSampEn* (H_{TS}) computed for the selected FHR time series and (6) decision step to classify recording as normal or arrhythmic using each entropy feature. Note: feature values shown are for illustrative purposes only, a.u. = arbitrary units.

It is important to note that the data-driven range of r used for computing *TotalSampEn* is calculated using only the single recording being analyzed and does not share information across recordings. Therefore, all assessed entropy methods utilize standard parameters that have not been specifically tuned to our data set.

An overview of the data analysis process used in this work showing the NI-FECG processing, FHR time series extraction, entropy feature computation and signal classification is shown in figure 3.

2.4. Clinical baseline

To compare the presented entropy methods against a clinical baseline, we also present the classification performance of detecting fetal arrhythmias using the clinical method as defined by Strasburger (2000). For this approach, all FHR time series with a mean FHR either below 100 bpm or above 180 bpm are classified as arrhythmic. This simple assessment approximates the method through which many fetal arrhythmias are initially detected in clinical practice and subsequently referred for further cardiac evaluation.

2.5. Surrogate analysis

To confirm that all entropy features are measuring nonlinear behavior in the FHR time series, we perform a surrogate analysis on all normal recordings in our data set. A surrogate analysis works by shuffling the temporal structure of the time series and comparing feature values before and after surrogation. If the distribution of feature values significantly changes after surrogation, the hypothesis that they represent purely linear behavior must be rejected (Schreiber and Schmitz 2000). To achieve this, we perform amplitude adjusted Fourier transform (AAFT) surrogation on all normal recordings with $N = 1000$, $r = 0.15 \times SD$ and $m = 2$ (Theiler *et al* 1992).

2.6. Signal quality

An important consideration when analyzing NI-FECG recordings is the wide range of signal quality present due to power line interference, muscle artifacts, sensor noise and inter-individual anatomic variations (Andreotti *et al* 2017, Kahankova *et al* 2019, Keenan *et al* 2021).

To determine the impact of changes in signal quality on arrhythmia classification performance, we compute the *bSQI* signal quality index for each NI-FECG recording which consists of running two different QRS detectors over the ECG signal and comparing their output (Li *et al* 2008, Behar *et al* 2013). The final *bSQI* value is computed as the ratio of beats detected by both detectors within an agreement interval compared to all beats identified by either detector. As specific QRS detection methods are sensitive to different types of noise, this method provides an estimate of signal quality of the underlying ECG. The *bSQI* signal quality measure is used over other signal quality measures as it does not rely on long term signal dynamics which may be altered in arrhythmic recordings.

For this work, we utilize the *jqrs* (Behar *et al* 2014) and Pan–Tompkins QRS detectors (Pan and Tompkins 1985) with an agreement interval of 50 ms, applied to the selected fetal ECG signal component used for FHR computation (see panel 3, figure 3). This combination of QRS detectors has been chosen as it was identified as the most relevant *bSQI* metric when compared to expert annotations on clinical NI-FECG recordings in a previous study by Andreotti *et al* (2017).

Table 3. Mean and standard deviation of each feature in normal recordings for original and surrogate data at $N = 1000$. p value indicates Mann–Whitney U test between original and surrogate feature values.

Feature	Original	Surrogate	p value
$SampEn(H_S)$	0.95 ± 0.55	1.62 ± 0.52	1.2E-27
$FuzzyEn(H_F)$	0.64 ± 0.42	1.12 ± 0.40	1.1E-30
$TotalSampEn(H_{TS})$	5.66 ± 7.54	8.69 ± 8.25	1.2E-14

2.7. Performance metrics

The metrics used to compare the performance of each entropy feature for classifying FHR time series as arrhythmic (positive) or normal (negative) are as follows: area under the receiver operating characteristic curve (AUC), sensitivity, specificity and F_1 score. Sensitivity (SE), specificity (SP) and F_1 score were calculated as follows:

$$SE = \frac{TP}{TP + FN}$$

$$SP = \frac{TN}{TN + FP}$$

$$F_1 = \frac{2 \cdot TP}{2 \cdot TP + FN + FP},$$

where TP, FP, TN and FN respectively are the number of true positives (correctly identified arrhythmia), false positives (falsely identified arrhythmia), true negatives (correctly identified normal) and false negatives (falsely identified normal). The classification threshold for each feature is determined by traversing the ROC curve and calculating the sensitivity, specificity and F_1 score at each threshold value. Detection is performed by classifying all recordings with feature values greater than the threshold to optimise the F_1 score or achieve a fixed specificity as desired. p values were calculated using the non-parametric Mann–Whitney U test and MATLAB R2019b (The Mathworks Inc.) was used to perform all computations.

3. Results

3.1. Surrogate analysis

The results of the surrogate analysis are given in table 3 where all entropy features show a statistically significant difference before and after surrogation with H_S ($p = 1.2\text{E} - 27$), H_F ($p = 1.1\text{E} - 30$) and H_{TS} ($p = 1.2\text{E} - 14$), indicating that all entropy measures represent nonlinear behavior, confirming their validity for this application.

3.2. Arrhythmia classification

The mean and standard error of entropy features H_S, H_F and H_{TS} for normal and arrhythmic FHR time series are shown in figure 4. It can be seen from table 2 that most recordings (93.7%) had at least $N = 250$ FHR values available for analysis, while for $N = 500$ and $N = 1000$, 72.6% and 67.3% of all recordings respectively met the data length requirement.

From figure 4, it can be observed that H_S is not defined in normal and arrhythmic fetuses for $N < 250$. This indicates that the standard choice of $r = 0.15 \times SD$ for computing H_S may be ill-suited for the short data lengths implemented in this work. However, this is an artificial limitation as H_F and H_{TS} are defined at all data lengths, with H_{TS} demonstrating greater differentiation between normal and arrhythmic recordings for all values of N.

This difference between entropy features is reflected in table 4 which shows that H_{TS} demonstrates the strongest performance for detecting fetal arrhythmias with a maximum AUC of 0.83 ($N = 250$ and $N = 1000$), compared to H_S which achieves a maximum AUC of 0.68 ($N = 1000$), and H_F which achieves a maximum AUC of 0.72 ($N = 1000$). Furthermore, it can be seen that even the worst performing data length for H_{TS}, with an AUC of 0.72 at $N = 25$, is equal to or greater than the best performance achieved by either H_S and H_F, indicating its reliable performance with smaller dependence on data length.

In terms of F_1 score, the highest performance for all entropy features is achieved at $N = 1000$ where H_{TS} achieves an F_1 score of 50.0% compared to 33.3% for H_S and 35.9% for H_F. These overall modest F_1 scores reflect the extreme class imbalance in our data set.

Comparing these results against the clinical baseline method, it can be seen that for $N > 50$, all investigated entropy features provide superior performance for arrhythmia detection in terms of AUC and F_1 score. Interestingly for $N \leqslant 50$, the performance of the clinical baseline method is equivalent or better compared to H_F, while H_{TS} provides the highest performance in all comparisons.

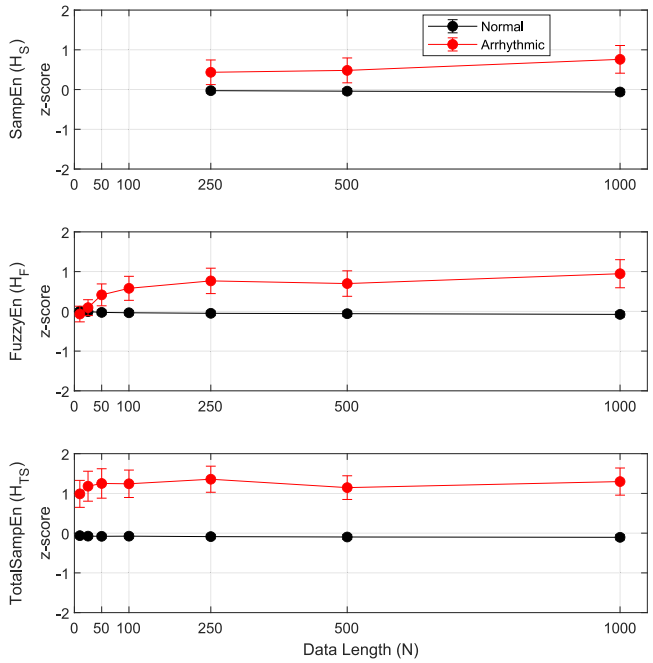

Figure 4. Mean and standard error of entropy features expressed as *z*-scores for the first *N* FHR values in each time series. *SampEn* values could not be computed for all recordings for $N < 250$. Number of recordings which met each data length requirement shown in table 2.

Table 4. Performance of clinical baseline method and entropy features in terms of AUC and F_1 score for classifying normal versus arrhythmic recordings at varying data lengths. *SampEn* values could not be computed for all recordings for $N < 250$.

N	Clinical baseline		SampEn (H_S)		FuzzyEn (H_F)		TotalSampEn (H_{TS})	
	AUC	F_1(%)	AUC	F_1(%)	AUC	F_1(%)	AUC	F_1(%)
10	0.57	23.1	NA	NA	0.48	12.3	0.74	30.0
25	0.57	24.0	NA	NA	0.56	13.7	0.72	37.3
50	0.60	30.8	NA	NA	0.60	20.8	0.77	39.1
100	0.58	25.0	NA	NA	0.65	29.4	0.80	39.2
250	0.55	17.4	0.60	28.6	0.69	26.2	0.83	43.6
500	0.55	17.4	0.61	28.6	0.68	28.6	0.82	47.6
1000	0.55	19.0	0.68	33.3	0.72	35.9	0.83	50.0

Table 5. Sensitivity of *TotalSampEn* (H_{TS}) at specificities of 95.0%, 90.0% and 85.0% and determined H_{TS} threshold for arrhythmia detection for each data length *N*. Best result per column in bold. Number of recordings which met each data length requirement shown in table 2.

	Sensitivity at selected specificity			Threshold at selected specificity		
N	95.0%	90.0%	85.0%	95.0%	90.0%	85.0%
100	22.2%	55.6%	66.7%	21.7	13.5	8.2
250	27.8%	**66.7%**	**83.3%**	20.3	**12.0**	**7.3**
500	27.8%	55.6%	61.1%	22.7	13.5	10.8
1000	**43.8%**	62.5%	68.8%	**22.9**	14.4	8.7

To understand the performance of H_{TS} in terms of sensitivity and specificity, figure 5 shows the receiver operating characteristic (ROC) curves for all data lengths with an AUC ≥ 0.8, while table 5 gives the sensitivity at specificities of 85.0%, 90.0% and 95.0% and the determined H_{TS} thresholds for arrhythmia detection. At a specificity of 95.0%, a data length of $N = 1000$ provides the highest sensitivity of 43.8%, while at specificities of 90.0% and 85.0%, $N = 250$ provides the highest sensitivities of 66.7% and 83.3% respectively.

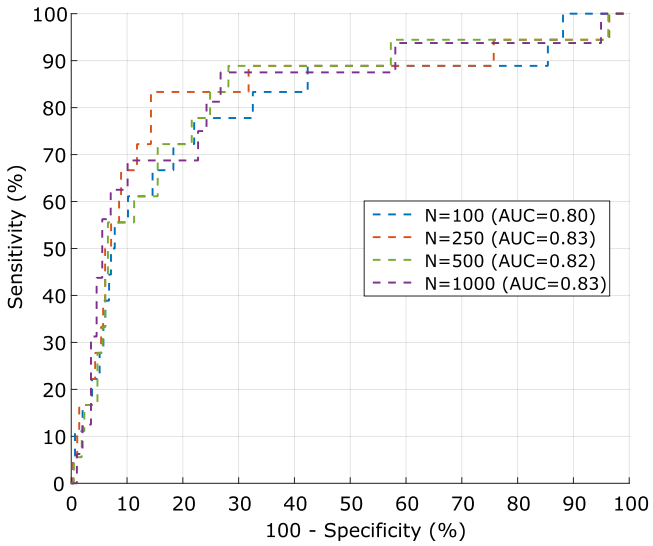

Figure 5. Receiver operating characteristic (ROC) curves for *TotalSampEn* (H_{TS}) at varying data length N for AUC $\geqslant$ 0.8. Number of recordings which met each data length requirement shown in table 2.

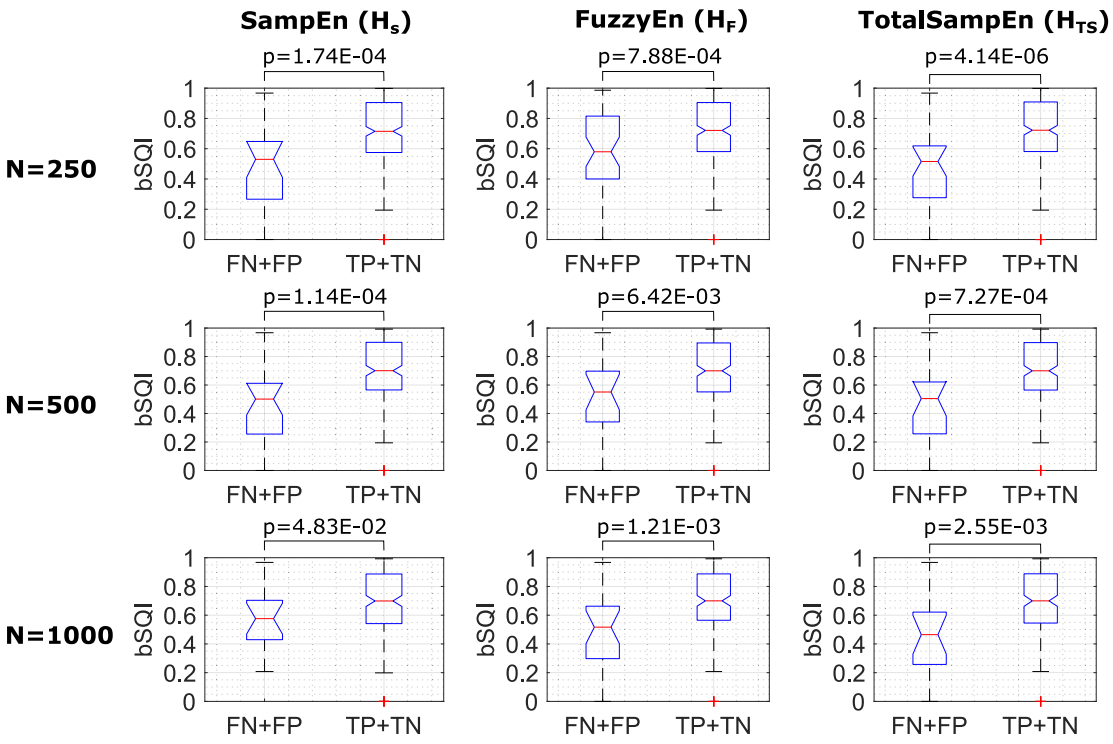

Figure 6. Boxplots of differences in signal quality (*bSQI*) between incorrectly classified (FP + FN) and correctly classified (TP + TN) NI-FECG recordings at the optimal F_1 decision point for each entropy feature at data lengths N = 250, 500 and 1000. Each boxplot shows the median (center line), inter-quartile range (box bounds), largest (top whisker) and smallest (bottom whisker) values within 1.5 × inter-quartile range, and outliers. p value represents statistical significance of the difference between median *bSQI* values in each group using the Mann–Whitney U test. Number of recordings which met each data length requirement shown in table 2.

3.3. Signal quality

The impact of signal quality on the classification performance of each method is given in figure 6 which shows the difference in *bSQI* between correctly classified recordings (TP + TN) and incorrectly classified recordings (FP + FN) at the optimal F_1 point for N = 250, 500 and 1000 for all entropy features. As can be seen, there is a robust statistical association ($p < 0.05$) between *bSQI* and the correctness of detection for all entropy features, where incorrect classifications are significantly associated with lower signal quality.

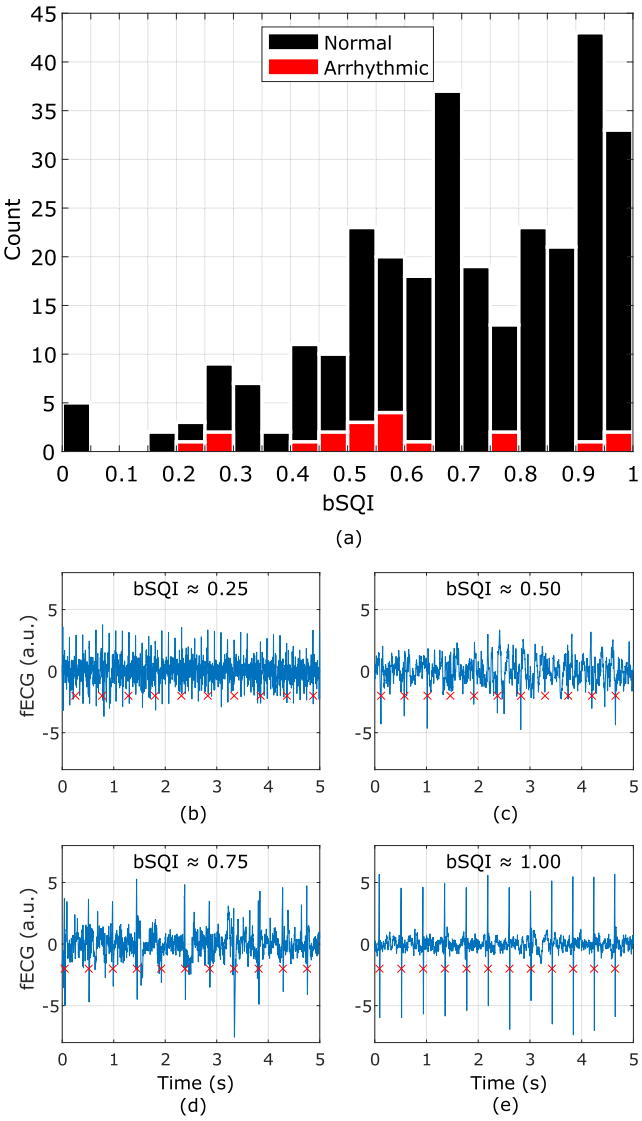

Figure 7. (a) Signal quality distribution of *bSQI* across normal and arrhythmic NI-FECG recordings, and example signal segments of normal recordings with *bSQI* approximately equal to (b) 0.25, (c) 0.50, (d) 0.75 and (e) 1.00. Each panel shows the extracted fetal ECG signal component and detected *fQRS* locations (×) after maternal ECG cancellation and independent component analysis. a.u. = arbitrary units.

To investigate the impact of excluding low signal quality recordings on classification performance, figure 7 shows the signal quality distribution of *bSQI* across all recordings in our data set, as well as example signal segments of normal recordings with *bSQI* approximately equal to 0.25, 0.50, 0.75 and 1.00.

Based on the signal quality distribution in figure 7 we calculate the AUC and F_1 score for each entropy feature, excluding NI-FECG recordings below *bSQI* thresholds from 0 to 0.5 in 0.05 increments. The results of this analysis are shown in figure 8, where it can be seen that excluding low signal quality recordings improves the classification performance of H_{TS} to a maximum AUC of 0.90 and maximum $F_1 = 60.9\%$ at a *bSQI* threshold of 0.45 for $N = 1000$. H_F achieves a smaller benefit, reaching a maximum AUC of 0.73 at a *bSQI* threshold of 0.2 and maximum $F_1 = 40.0\%$ at a *bSQI* threshold of 0.25 for $N = 1000$. Finally, H_S achieves no change in maximum AUC and reaches a maximum $F_1 = 35.7\%$ at a *bSQI* threshold of 0.25 for $N = 1000$. For comparison purposes, the amounts of overall data loss at *bSQI* thresholds of 0.2, 0.25 and 0.45 are 2.2%, 3.5% and 13.5% respectively.

4. Discussion

As shown in table 4, our entropy profiling approach *TotalSampEn* achieves the highest performance as an automated method for detecting fetal arrhythmias, outperforming entropy measures such as *SampEn* and

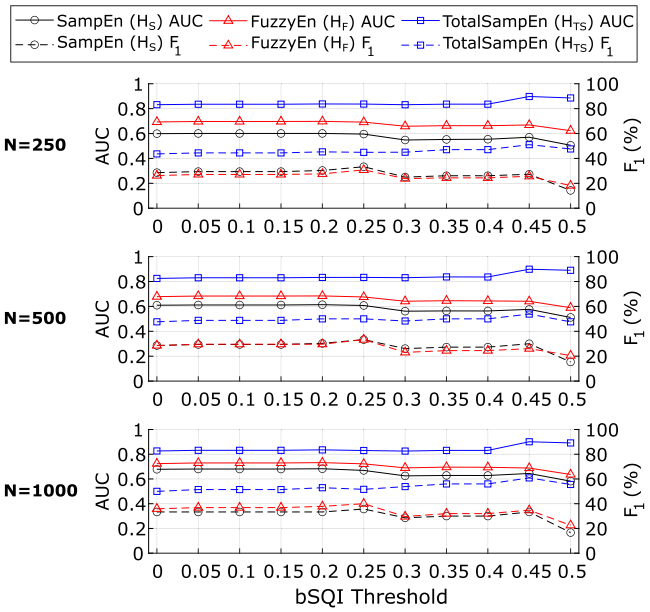

Figure 8. Classification performance of each entropy feature in terms of AUC and F_1 as recordings below the indicated *bSQI* threshold are excluded.

FuzzyEn in terms of AUC and F_1 at all data lengths. The increased performance of our approach can be attributed to its ability to retrieve a data-driven entropy profile for each signal, rather than relying on a fixed choice of the tolerance parameter r. This advantage also corresponds to an ability to retrieve reliable entropy estimates for extremely small data lengths.

Furthermore, based on the signal quality analysis in figure 6, it can be seen that recordings of poor signal quality are more likely to be misclassified for all investigated entropy features. However, figure 8 demonstrates that *TotalSampEn* benefits more strongly from excluding low signal quality recordings, which may provide a beneficial tradeoff in improving the accuracy of classification.

Due to the low incidence of fetal arrhythmias in the general population (Southall *et al* 1977), it is important that the presented method performs adequately at high specificity to ensure that the clinical burden of false positives is minimal. For this purpose, our method demonstrates excellent suitability with 83.3% sensitivity at 85.0% specificity for $N = 250$ as shown in table 5. It is interesting to observe that at 95.0% specificity, H_{TS} computed for $N = 1000$ strongly outperforms all other data lengths, contributing to its higher F_1 score. However, it is important to note that more recordings met the data length requirement for $N = 250$ compared to $N = 1000$ (93.7% versus 67.3%).

These differences in available data length are due to the retrospective nature of our study in which NI-FECG recordings were collected during routine clinical practice in the pursuit of optimizing sensor placement and signal processing techniques. Based on these results, we recommend future studies collect at least $N = 1000$ FHR values to provide the opportunity for extremely high specificity diagnosis.

Based on the signal quality analysis, we further recommend utilizing a signal quality metric such as *bSQI* when assessing FHR values for arrhythmia analysis. In this work we utilized a *bSQI* metric applied to the entire NI-FECG recording, however it is possible to compute *bSQI* for smaller windows and exclude only subsections of a recording which fall below the desired quality threshold.

One limitation of the present work is that we have not used a test-train methodology to validate the optimal N and *bSQI* values for maximising classification performance. This is due to the small number of arrhythmic participants in our cohort and as such we recommend that future studies validate the decision points found in this work. Future studies may also investigate the potential for constructing multi-feature classifiers utilizing time and frequency domain features similar to adult arrhythmia analysis, as well the potential for automated identification of the type of arrhythmia (Alonso-Atienza *et al* 2014).

Finally, previous work has shown that NI-FECG recordings are suitable for identifying the mechanism of fetal arrhythmia when interpreted by a trained clinician (Behar *et al* 2019). The results of this study demonstrate that entropy profiling analysis can be used as a first line tool for identifying fetal arrhythmias in NI-FECG recordings to improve detection performance in the wider population. If a suspected arrhythmia is detected, the extracted NI-FECG waveforms may be transferred to a perinatal cardiologist for examination to determine the type of arrhythmia and course of treatment.

5. Conclusion

In summary, this work presents a novel method of detecting fetal arrhythmias in short length non-invasive fetal electrocardiography (NI-FECG) recordings. Our method (*TotalSampEn*) consists of computing an entropy profile from the FHR time series extracted from NI-FECG recordings using a data-driven range of the entropy tolerance parameter *r*. In our clinical data set of 318 NI-FECG recordings, *TotalSampEn* demonstrates excellent performance for detecting fetal arrhythmias and strongly outperforms entropy measures such as *SampEn* and *FuzzyEn*. The superior performance of our approach enables automated detection of fetal arrhythmias and warrants further investigation in a prospective clinical trial.

Acknowledgments

Fiona Brownfoot is supported by a NHMRC Early Career Fellowship (NHMRC #1142636) and a Norman Beischer Clinical Research Fellowship. Emerson Keenan is supported by a University of Melbourne Department of Obstetrics and Gynaecology Early Career Fellowship. This work was supported by the Australian Research Council (ARC) Discovery Project under Grant DP190101248.

ORCID iDs

Emerson Keenan https://orcid.org/0000-0003-1966-2293
Chandan Karmakar https://orcid.org/0000-0003-1814-0856
Fiona C Brownfoot https://orcid.org/0000-0002-1439-7016
Igor Lakhno https://orcid.org/0000-0002-7914-7296
Joachim A Behar https://orcid.org/0000-0001-5956-7034
Marimuthu Palaniswami https://orcid.org/0000-0002-3635-4252

References

Allan L D, Anderson R H, Sullivan I D, Campbell S, Holt D W and Tynan M 1983 Evaluation of fetal arrhythmias by echocardiography *Br. Heart J.* **50** 240–5

Alonso-Atienza F, Morgado E, Fernández-Martínez L, García-Alberola A and Rojo-Álvarez J L 2014 Detection of life-threatening arrhythmias using feature selection and support vector machines *IEEE Trans. Biomed. Eng.* **61** 832–40

Andreotti F, Grasser F, Malberg H and Zaunseder S 2017 Non-invasive fetal ECG signal quality assessment for multichannel heart rate estimation *IEEE Trans. Biomed. Eng.* **64** 2793–802

Behar J, Oster J and Clifford G D 2014 Combining and benchmarking methods of foetal ECG extraction without maternal or scalp electrode data *Physiol. Meas.* **35** 1569–1589

Behar J, Oster J, Li Q and Clifford G D 2013 ECG signal quality during arrhythmia and its application to false alarm reduction *IEEE Trans. Biomed. Eng.* **60** 1660–6

Behar J A, Bonnemains L, Shulgin V, Oster J, Ostras O and Lakhno I 2019 Noninvasive fetal electrocardiography for the detection of fetal arrhythmias *Prenatal Diagn.* **39** 178–87

Caswell Schuckers S A 1998 Use of approximate entropy measurements to classify ventricular tachycardia and fibrillation *J. Electrocardiol.* **31** 101–5

Chen W, Zhuang J, Yu W and Wang Z 2009 Measuring complexity using FuzzyEn, ApEn, and SampEn *Med. Eng. Phys.* **31** 61–8

Crotti L *et al* 2013 Long QT syndrome-associated mutations in intrauterine fetal death *JAMA* **309** 1473–82

Goldberger A L, Amaral L A N, Glass L, Hausdorff J M, Ivanov P C, Mark R G, Mietus J E, Moody G B, Peng C and Stanley H E 2000 Physiobank, physiotoolkit, and physionet *Circulation* **101** e215–20

Hornberger L K 2007 Echocardiographic assessment of fetal arrhythmias *Heart* **93** 1331–3

Kahankova R, Martinek R, Jaros R, Behbehani K, Matonia A, Jezewski M and Behar J A 2019 A review of signal processing techniques for non-invasive fetal electrocardiography *IEEE Rev. Biomed. Eng.* **13** 51–73

Keenan E, Karmakar C, Brownfoot F C and Palaniswami M 2021 Personalized anatomic modeling for noninvasive fetal ECG: methodology and applications *IEEE Trans. Instrum. Meas.* **70** 1–12

Keenan E, Udhayakumar R K, Karmakar C K, Brownfoot F C and Palaniswami M 2020 Entropy profiling for detection of fetal arrhythmias in short length fetal heart rate recordings *2020 42nd Annual Int. Conf. of the IEEE Engineering in Medicine Biology Society (EMBC)* pp 621–4

Komáromy B, Gaál J and Lampé L 1977 Fetal arrhythmia during pregnancy and labour *BJOG: An Int. J. Obstetrics Gynaecol.* **84** 492–6

Krapp M, Kohl T, Simpson J M, Sharland G K, Katalinic A and Gembruch U 2003 Review of diagnosis, treatment, and outcome of fetal atrial flutter compared with supraventricular tachycardia *Heart* **89** 913–7

Lake D E, Richman J S, Griffin M P and Moorman J R 2002 Sample entropy analysis of neonatal heart rate variability *Am. J. Physiol.-Regulatory, Integr. Comparative Physiol.* **283** R789–97

Li Q, Mark R G and Clifford G D 2008 Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter *Physiol. Meas.* **29** 15–32

Mayer C C, Bachler M, Hörtenhuber M, Stocker C, Holzinger A and Wassertheurer S 2014 Selection of entropy-measure parameters for knowledge discovery in heart rate variability data *BMC Bioinf.* **15** S2

Pan J and Tompkins W J 1985 A real-time QRS detection algorithm *IEEE Trans. Biomed. Eng.* **BME-32** 230–6

Pincus S M 1991 Approximate entropy as a measure of system complexity *Proc. Natl Acad. Sci.* **88** 2297–301

Rajendra Acharya U, Kannathal N, Sing O W, Ping L Y and Chua T 2004 Heart rate analysis in normal subjects of various age groups *Bio. Med. Eng. Online* **3** 24

Richman J S and Moorman J R 2000 Physiological time-series analysis using approximate entropy and sample entropy *Am. J. Physiol.-Heart Circ. Physiol.* **278** H2039–49

Schreiber T and Schmitz A 2000 Surrogate time series *Physica* D **142** 346–82

Southall D P, Orrell M J, Talbot J F, Brinton R J, Vulliamy D G, Johnson A M, Keeton B R, Anderson R H and Shinebourne E A 1977 Study of cardiac arrhythmias and other forms of conduction abnormality in newborn infants *Br. Med. J.* **2** 597–9

Strasburger J F 2000 Fetal arrhythmias *Prog. Pediatric Cardiol.* **11** 1–17

Theiler J, Eubank S, Longtin A, Galdrikian B and Doyne Farmer J 1992 Testing for nonlinearity in time series: the method of surrogate data *Physica* D **58** 77–94

Udhayakumar R K, Karmakar C and Palaniswami M 2017 Approximate entropy profile: a novel approach to comprehend irregularity of short-term HRV signal *Nonlinear Dyn.* **88** 823–37

Udhayakumar R K, Karmakar C and Palaniswami M 2018 Understanding irregularity characteristics of short-term HRV signals using sample entropy profile *IEEE Trans. Biomed. Eng.* **65** 2569–79

Udhayakumar R K, Karmakar C and Palaniswami M 2019 Entropy profiling to detect ST change in heart rate variability signals *2019 41st Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC)* pp 4588–91

Varanini M, Tartarisco G, Billeci L, Macerata A, Pioggia G and Balocchi R 2014 An efficient unsupervised fetal QRS complex detection from abdominal maternal ECG *Physiol. Meas.* **35** 1607–1619

Wacker-Gussmann A, Strasburger J F, Cuneo B F and Wakai R T 2014 Diagnosis and treatment of fetal arrhythmia *Am. J. Perinatol.* **31** 617–28

Yentes J M, Hunt N, Schmid K K, Kaipust J P, McGrath D and Stergiou N 2013 The appropriate use of approximate entropy and sample entropy with short data sets *Ann. Biomed. Eng.* **41** 349–65