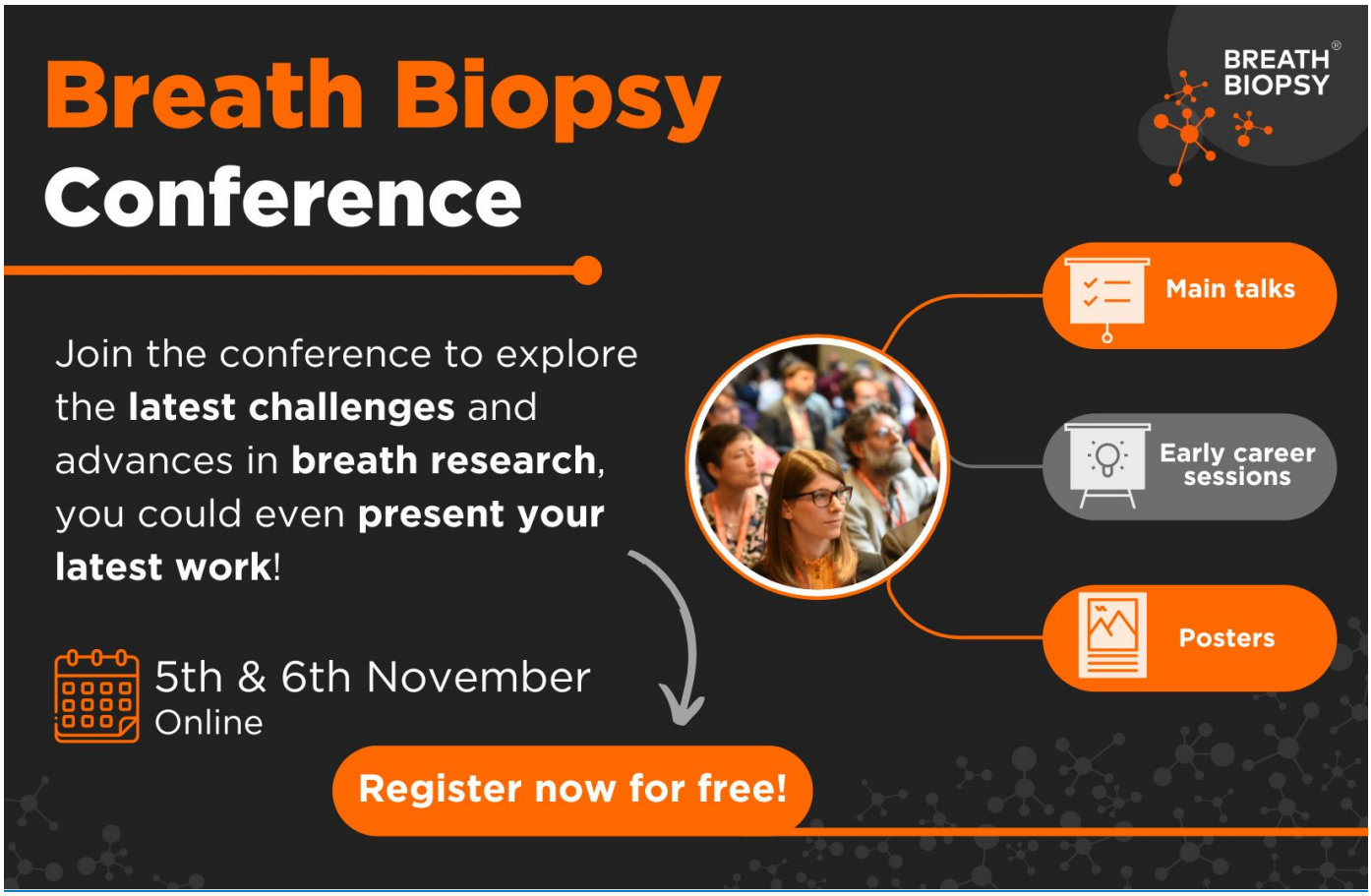
Physiological Measurement

Power-MF: robust fetal QRS detection from non-invasive fetal electrocardiogram recordings



IOP Publishing Physiol. Meas. 45 (2024) 055009 https://doi.org/10.1088/1361-6579/ad4952

Physiological Measurement

IPEM Institute of Physics and Engineering in Medicine

OPEN ACCESS

RECEIVED
20 July 2023

REVISED
18 March 2024

ACCEPTED FOR PUBLICATION
9 May 2024

PUBLISHED
21 May 2024

PAPER

Power-MF: robust fetal QRS detection from non-invasive fetal electrocardiogram recordings

Katharina M Jaeger[1,*], **Michael Nissen**[1], **Simone Rahm**[1], **Adriana Titzmann**[2], **Peter A Fasching**[2], **Janina Beilner**[1], **Bjoern M Eskofier**[1,3] and **Heike Leutheuser**[1]

[1] Friedrich-Alexander-Universitat Erlangen-Nürnberg, Machine Learning and Data Analytics Lab, Carl-Thiersch-Straße 2b, 91052 Erlangen, Germany
[2] Department of Gynecology and Obstetrics, Erlangen University Hospital, Universitätsstraße 21–23, 91054 Erlangen, Germany
[3] Translational Digital Health Group, Institute of AI for Health, Helmholtz Zentrum München—German Research Center for Environmental Health, Neuherberg, Germany
[*] Author to whom any correspondence should be addressed.

E-mail: katharina.jaeger@fau.de

Keywords: digital health, fetal electrocardiography, prenatal care, biomedical signal analysis, wearable sensing, fetal QRS detection

Supplementary material for this article is available online

Abstract

Objective. Perinatal asphyxia poses a significant risk to neonatal health, necessitating accurate fetal heart rate monitoring for effective detection and management. The current gold standard, cardiotocography, has inherent limitations, highlighting the need for alternative approaches. The emerging technology of non-invasive fetal electrocardiography shows promise as a new sensing technology for fetal cardiac activity, offering potential advancements in the detection and management of perinatal asphyxia. Although algorithms for fetal QRS detection have been developed in the past, only a few of them demonstrate accurate performance in the presence of noise and artifacts. *Approach.* In this work, we propose *Power-MF*, a new algorithm for fetal QRS detection combining power spectral density and matched filter techniques. We benchmark *Power-MF* against three open-source algorithms on two recently published datasets (Abdominal and Direct Fetal ECG Database: ADFECG, subsets B1 Pregnancy and B2 Labour; Non-invasive Multimodal Foetal ECG-Doppler Dataset for Antenatal Cardiology Research: NInFEA). *Main results.* Our results show that *Power-MF* outperforms state-of-the-art algorithms on ADFECG (B1 Pregnancy: 99.5% $\pm$ 0.5% F1-score, B2 Labour: 98.0% $\pm$ 3.0% F1-score) and on NInFEA in three of six electrode configurations by being more robust against noise. *Significance.* Through this work, we contribute to improving the accuracy and reliability of fetal cardiac monitoring, an essential step toward early detection of perinatal asphyxia with the long-term goal of reducing costs and making prenatal care more accessible.

1. Introduction

Perinatal asphyxia, characterized by insufficient oxygen supply to the fetus, is a critical issue that can result in severe neurologic and developmental impairments (Wang *et al* 2021). Immediate identification and intervention are crucial for mitigating the adverse effects of perinatal asphyxia. The current standard method for assessing fetal well-being during labour is cardiotocography (CTG), which measures fetal heart rate and uterine contractions (Chen *et al* 2011). Diagnosis of perinatal asphyxia is guided by the FIGO guidelines (Ayres-de Campos *et al* 2015). However, concerns have been raised regarding the accuracy of CTG due to subjective interpretation and potential errors, leading to false positives and false negatives (Ojala *et al* 2006, Benton *et al* 2020). Moreover, their bulky design requires stationary operation and demands expert knowledge for transducer placement. These limitations necessitate the exploration of novel technologies that can provide more precise and reliable information regarding fetal well-being. Non-invasive fetal electrocardiography (NI-fECG) is a promising sensing technology for fetal heart monitoring that is being

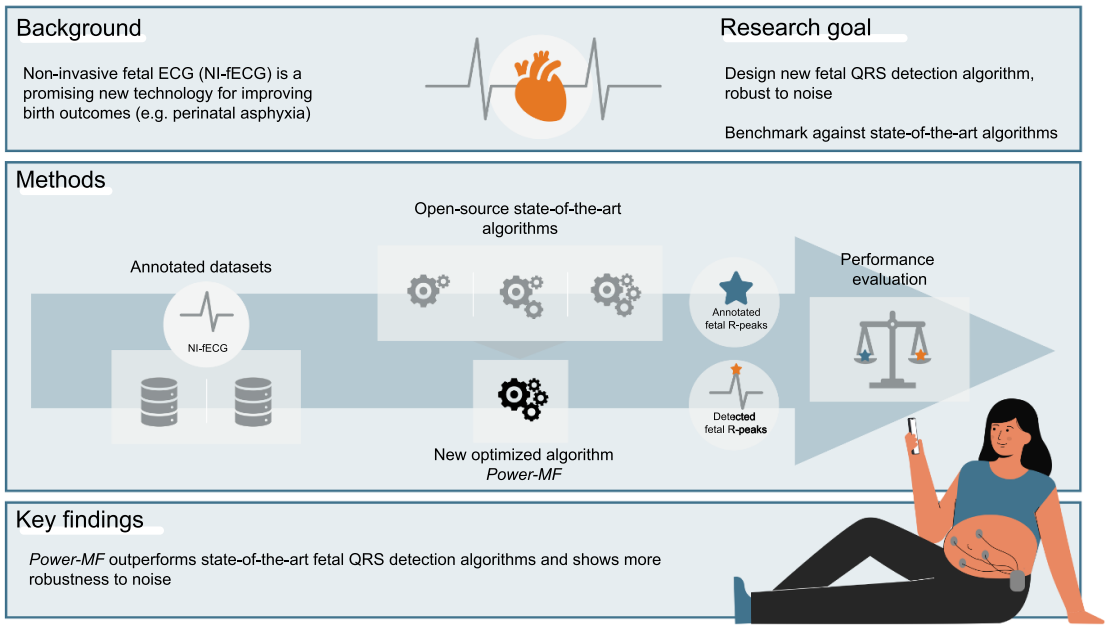


Figure 1. Graphical abstract of this paper, showing the benchmark process of *Power-MF* against state-of-the-art fetal QRS detection algorithms. Further, the background, research goal, methods, and key findings are described.

developed to refine the determination of perinatal asphyxia (Oudijk *et al* 2004, Jezewski *et al* 2017). It records fetal cardiac activity unobtrusively and non-invasively via electrodes from the maternal abdomen. Different electrode positions allow the signal to be acquired at different angles, which are displayed as channels. There is a huge amount of possible electrode configurations, yet standardized positioning remains undefined. Minimizing the number of electrodes offers increased comfort for the pregnant woman. Compared to CTG, NI-fECG offers high levels of fetal heart rate accuracy, which is less influenced by fetal movements and more accurate for women with high BMI (Sänger *et al* 2012, Hayes-Gill *et al* 2020, Liu *et al* 2023). NI-fECG provides an unobtrusive solution for long-term fetal monitoring as a self-applied wearable device, addressing the challenges of limited access and time-consuming prenatal care.

Although NI-fECG technology is gaining acceptance, and its potential beyond fetal heart rate monitoring is being explored (Jaeger *et al* 2022), its implementation in clinical practice is still limited (Wakefield *et al* 2022). The proof of reliable heart rate extraction under real conditions is yet to be established, and there is currently no definition of normative values (Smith *et al* 2018). Signal processing challenges arise due to increased noise levels in wearable NI-fECG and the low signal-to-noise ratio resulting from the mixed signal containing fetal and maternal ECG, uterine muscle signals, and other confounders, making fECG extraction challenging. Existing algorithms for fetal QRS detection are available. However, previous comparisons of those algorithms lacked standardization (Hasan *et al* 2009, Clifford *et al* 2014, Andreotti *et al* 2016, Li *et al* 2017, Kahankova *et al* 2020, Vaidya and Chaitra 2020). Additionally, there is a need for an algorithm that exhibits stable and reliable performance in the presence of noisy recordings.

In this paper, we introduce *Power-MF*, a new fetal QRS detection algorithm designed to be robust against noise. *Power-MF* is based on a combination of power spectral density (PSD) and matched filter techniques. Further, we objectively benchmark *Power-MF* against the state-of-the-art for fetal QRS detection on the two most recently published *Non-invasive Multimodal Foetal ECG-Doppler Dataset for Antenatal Cardiology Research* (NInFEA) (Sulas *et al* 2021) and *Abdominal and Direct Fetal ECG Database* (ADFECG) (Matonia *et al* 2020) datasets. For benchmarking, we selected three relevant open-source algorithms. We analyze the algorithms' performances concerning electrode configurations suitable for wearable NI-fECG devices.

A graphical abstract of this paper is shown in figure 1.

2. State-of-the-art algorithms

We conducted literature research in central databases to identify the currently most relevant fetal QRS detection algorithms. We also took algorithms published with open-source datasets and algorithms mentioned in review papers into account. From the collection of algorithms, we decided to focus on three algorithms by Behar *et al* (2014), Varanini *et al* (2014), and Sulas *et al* (2021). These algorithms were published after 2013, and their source code is publicly available. Further, they have been evaluated on a

Table 1. Self-stated performance of state-of-the-art fetal QRS detection algorithms. The algorithms are outlined along with their respective publications and the acronyms we defined for this work. The performance metrics include Se, PPV, and F1-score. Descendants are reported as stated in the corresponding publication.

Publication	Algorithm acronym	Evaluation dataset	Se in %	PPV in %	F1-score in %
Varanini *et al* (2014)	*Varanini*	CinC2013	99.1	98.9	98.9[a]
Behar *et al* (2014)	*Behar*	CinC2013	95.9	96.0	96.0
Sulas *et al* (2021)	*Sulas*	NInFEA	97	81	88[a]

[a] Retrospective calculation.

complete public dataset with the most commonly used metrics Sensitivity (Se), positive predictive value (PPV), and F1-score, and achieved good results. Table 1 shows the selected algorithms, including their self-stated accuracy and the respective evaluation datasets. In the following, the algorithms are described in detail.

2.1. Sulas (2021)

The algorithm by Sulas *et al* (2021) (*Sulas*) has a self-reported Se of 97% and PPV of 81% and was evaluated on the NInFEA dataset (Sulas *et al* 2021). The first step of *Sulas* consists of a preprocessing step using a bandpass filter between 0.05 Hz and 250 Hz to suppress low and high-frequency components. Then, the maternal QRS complexes are detected from the thoracic reference channels, and the maternal ECG is suppressed using a deflation algorithm based on periodic component analysis (πCA). The deflation algorithm comprises a decomposition step followed by wavelet denoising and reconstruction of the signal to remove the maternal component (Sameni and Clifford 2010).

To enhance the fetal signal, an independent component analysis (ICA) step is performed on the residual signals by means of the joint approximate diagonalization of eigenmatrices (JADE) algorithm based on the work of Cardoso and Souloumiac (1993). A channel selection based on a matched filter is performed to find the channel that contains the strongest fetal signal after ICA. A generic template of a fetal QRS complex is used for this purpose. To find the fetal QRS complexes, peak detection is performed on the matched filter output of each channel. The RR intervals of each channel are computed, and the channel with the slightest standard deviation of RR intervals is selected.

In order to correct the detected fetal QRS, another πCA step is performed with the temporary fetal QRS as an input. A new template is created based on their positions by averaging multiple fetal QRS. Using the new template, a matched filter is applied. The peak detection results applied to the output signal indicate the final fetal QRS.

2.2. Behar (2014)

The algorithm by Behar *et al* (2014) (*Behar*) was published during the Physionet / Computing in Cardiology Challenge 2013 (CinC2013) (Goldberger *et al* 2000). According to the self-reported performance it achieved a Se of 95.9%, PPV of 96.0%, and an F1-score of 96.0% on CinC2013. In this approach, template subtraction (TS) and ICA are performed in various sequences (TS, TS-ICA, ICA, ICA-TS), and from those, the channel with the smoothest heart rate variability is selected.

First, a notch filter followed by a Butterworth low-pass is applied to remove powerline interference and high-frequency components above 200 Hz. Furthermore, baseline wanders below 3 Hz are canceled using a Butterworth high-pass filter.

On each preprocessed channel, maternal QRS detection is performed with a QRS detector similar to Pan and Tompkins (1985) with a refractory period of 250 ms. Further, a channel selection approach is applied to select the channel with the most plausible maternal QRS sequence as a reference.

Maternal ECG suppression is applied in four different ways, and the best fetal QRS sequence out of all four approaches is selected. The first approach, referred to as 'TS', consists of a simple TS, where several maternal R-peaks are averaged with a fixed window width around the maternal R-peak. In the second approach, 'TS-ICA', an ICA step, is applied after TS, using the JADE algorithm based on the work of Cardoso and Souloumiac (1993). In the third approach, 'ICA', only an ICA step is performed. In the fourth approach, 'ICA-TS', ICA is followed by a TS. On the residuals of all four methods, peak detection is performed using a QRS detector with an adjusted refractory period of 150 ms to account for the higher fetal heart rate. Out of all fetal QRS sequences, the best one is selected using a beat comparison measure. The final step consists of a smoothing process to remove extra detected fetal QRS and insert missed fetal QRS based on physiologically reasonable heart rates.

2.3. Varanini (2014)

The algorithm by Varanini *et al* (2014) (*Varanini*) has a high self-reported Se of 99.1% and PPV of 98.9% on the CinC2013 dataset, resulting in an F1-score of 98.9%. The algorithm uses ICA and singular value decomposition (SVD) for detecting and suppressing maternal ECG. Another ICA step is performed on the residual signal to enhance the fECG.

First, the signals are preprocessed to remove impulsive artifacts using a median filter. In the second step, baseline wandering is removed with a low pass Butterworth filter in forward and backward directions to achieve a zero-phase shift in the baseline estimate. The difference between the baseline and original signals is then used as the detrended signal. Finally, the power line interference is removed using notch filters.

After preprocessing, a FastICA algorithm according to Hyvarinen (1999) with deflationary orthogonalization and the hyperbolic cosine as contrast function is performed to separate the maternal ECG from the signal. Then, the channel with the best maternal ECG signal is selected, and the maternal QRS are detected using an adaptive threshold on the absolute derivative of the selected maternal channel.

To suppress the maternal ECG, the maternal R-peaks are approximated by an SVD-based method. A template of the maternal ECG is created by weighting the maternal R-peaks with a trapezoidal window and decomposing it with SVD. Afterward, the maternal beats are reconstructed using eigenvectors corresponding to the three largest eigenvalues, which are subtracted from the signal. Maternal ECG cancellation is performed on all channels.

Another ICA step is performed to further enhance and separate the weak fetal signal after maternal ECG suppression. For fetal QRS detection, the signal is filtered with a derivative filter consisting of a comb filter followed by a moving average. After filtering, the QRS detector is applied, similar to the maternal QRS detector but with an adapted RR-interval size. From the detected fetal QRS, the RR intervals are calculated, and a segment is identified in which a good SNR is assumed. This segment is characterized by constant RR intervals. From the beginning/end of this segment, a second QRS detector is initiated in the forward/backward direction. This detector searches for maximum points in the weighted derivative signal. The weights depend on the predicted RR intervals, i.e. the weight is higher in the area where the next QRS is expected. The lengths of the predicted RR intervals are initialized from the values in the start segment and adjusted from beat to beat using a least mean square algorithm. The best RR sequence is selected from each channel based on a-priori knowledge of typical fetal RR values to obtain plausible R-peaks.

3. Methods

3.1. Power-MF fetal QRS detection

In this section, we present *Power-MF*, a fetal QRS detection algorithm that utilizes PSD and matched filter techniques. The main goal of *Power-MF* is to improve robustness against noisy signal segments, a common issue in fetal QRS detection. The algorithm continues the work presented by Varanini *et al* (2014).

Previous state-of-the-art algorithms, such as *Varanini*, have shown good performance in clean signal segments, see table 1. However, their performance degrades when dealing with noisy segments. By inspecting the false detections of *Varanini*, we observed inaccuracies in noisy segments where fetal peaks are not clearly visible, see figure 2. *Power-MF* addresses this issue by incorporating a fetal QRS detection method based on a matched filter. The use of matched filters has been previously shown to be robust to noise in adult QRS detection, and we hypothesize that it will also provide good performance in the presence of noise in fECG signals (Eskofier *et al* 2008, Smigiel and Marciniak 2017, Jamshidian-Tehrani and Sameni 2018). Additionally, *Power-MF* focuses on identifying the channel with the strongest fECG component for QRS detection. We use the PSD of the fECG to distinguish it from the maternal ECG and other background noise. By focusing on the channel with the strongest fECG component, *Power-MF* aims to improve the SNR and increase the accuracy of QRS detection. *Power-MF* employs *Varanini's* steps for preprocessing and maternal ECG cancellation, see section 2.3. Figure 3 shows the algorithms steps of *Power-MF* and *Varanini*.

3.1.1. Channel selection based on PSD

We assume that in the channel containing the fECG, the fetal QRS complexes occur at a specific frequency corresponding to the fetal heart rate. Therefore, the channel with the highest PSD in the range of the expected fetal heart rate is selected.

First, the signal is preprocessed to highlight the fetal peaks. Therefore, the filtered absolute derivative of the signal is calculated for each channel. A comb filter with 8 ms delay is used as the derivative filter, followed by a moving average with a 5 ms window length. The absolute values of the derivative are calculated and filtered with a Butterworth bandpass between 0.7 Hz and 8.0 Hz. The derivative highlights the high-frequency components caused by the high slopes of the QRS. The bandpass is then used to smooth the

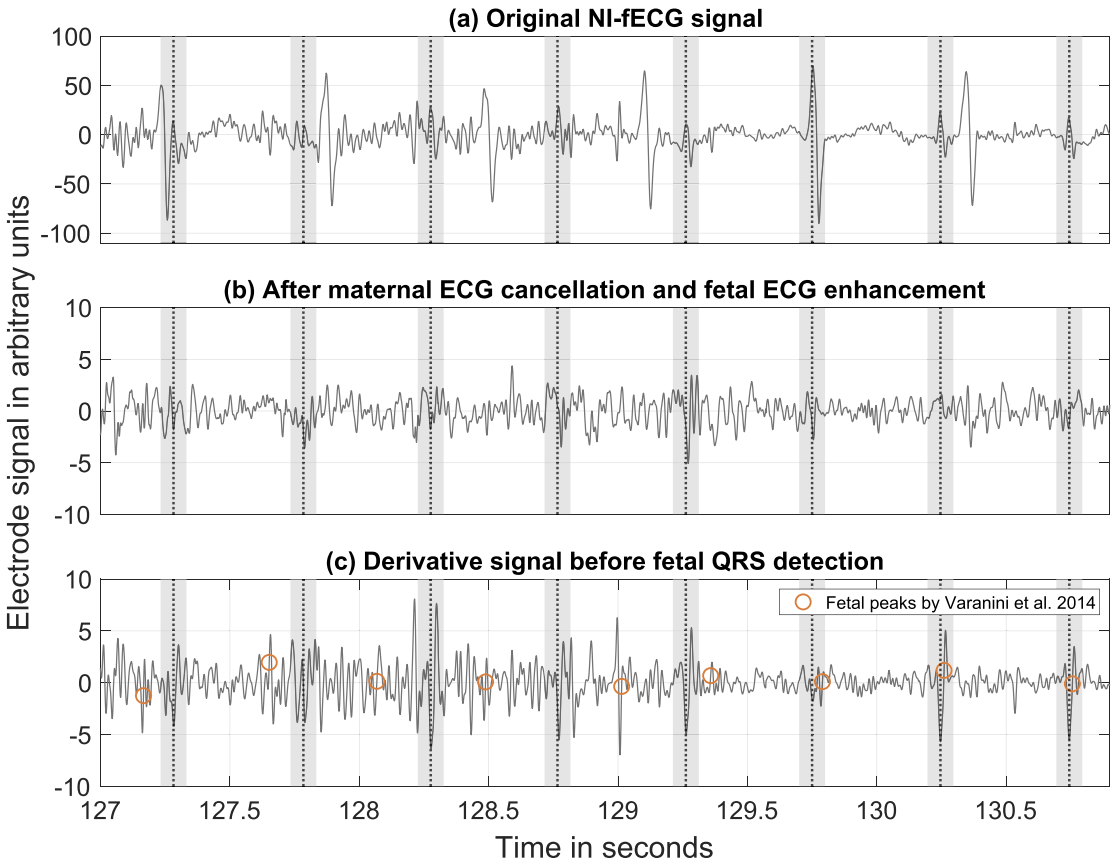


Figure 2. Processing steps of the fetal QRS detection algorithm by Varanini *et al* (2014). Varanini *et al* (2014) falsely detects the fetal peaks in the noisy parts of the signal (127.0–129.5 s). The dashed vertical lines indicate the ground truth fetal peak annotations. The circles in the bottom plot indicate the fetal peaks detected by Varanini *et al* (2014). The grey areas mark the acceptance interval of 50 ms. The signal is the fourth channel of recording 6 from the dataset ADFECG B2 Labour. The electrode signal is displayed in arbitrary units.

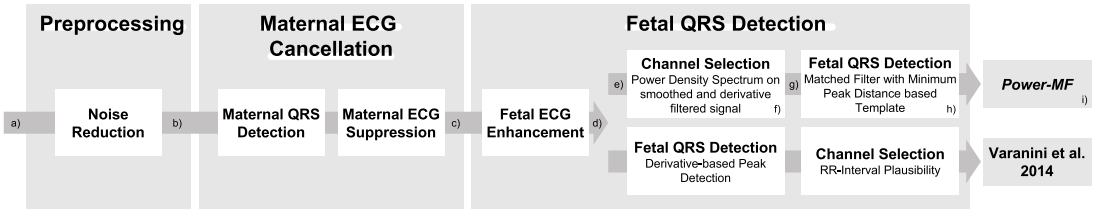


Figure 3. Overview of the algorithm steps of *Power-MF* and Varanini *et al* (2014). *Power-MF* and Varanini *et al* (2014) share the preprocessing and maternal ECG cancellation. Their fetal QRS detection methodology differs in channel selection and fetal QRS detection. The labels (a)–(i) correspond to different time points during the algorithm. Graphical visualizations of (a)–(i) based on an example signal are included in the supplementary material.

signal. The filter parameters were determined through practical observation and experience exclusively with the CinC2013 dataset.

The PSD was obtained by calculating Welch's PSD estimate (Welch 1967) with a Gaussian window and a 50% overlap. The window size *ws* was set to $ws = 15 \cdot \frac{60\,\mathrm{s}}{110\,\mathrm{bpm}} \cdot f_s$, with sampling frequency f_s, to capture at least 15 cardiac cycles, assuming a lower fetal heart rate boundary of 110 bpm. The PSD of the channel with the clearest fetal signal is assumed to contain a peak at a specific frequency corresponding to the fetal heart rate. The channel with the most prominent peak between 1.8 Hz and 3.0 Hz in the PSD is selected, corresponding to a fetal heart rate between 108 bpm and 180 bpm. The healthy fetal heart rate is between 110 bpm and 160 bpm. The upper limit was chosen slightly higher to include boundary cases. The lower limit was shifted only slightly to avoid reaching the range of the maternal heart rate in case the maternal ECG was not completely suppressed.

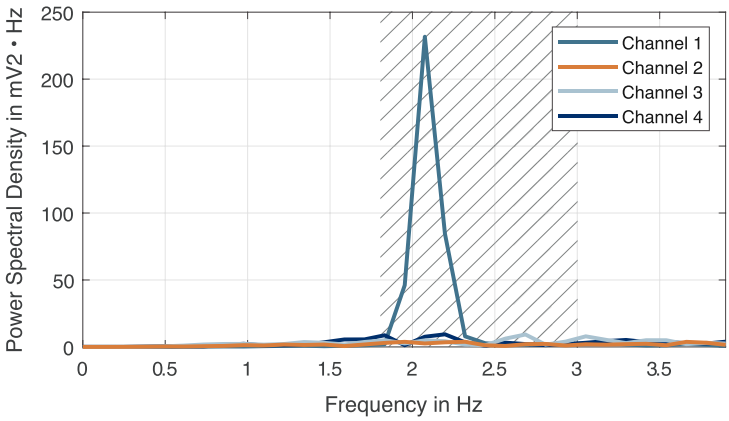


Figure 4. Channel selection of *Power-MF* based on PSD. PSD was estimated using Welch's method. A distinct peak occurs for Channel 1 at about 2.1 Hz. The graphs of the other three channels do not show a clear peak. The highlighted dashed area between 1.8 Hz and 3.0 Hz corresponds to a fetal heart rate between 108 bpm and 180 bpm and indicates the range in which the fetal peak is expected. This is exemplarily shown for a signal with four channels.

Figure 4 shows the PSD for all channels of a signal. In the highlighted dashed area between 1.8 Hz and 3.0 Hz only one channel has a clear peak at 2.1 Hz, corresponding to the channel with the best fECG representation.

3.1.2. Fetal QRS detection based on a matched filter

A matched filter maximizes the detection of a target signal by correlating the input signal with a known reference waveform called a template. In the first step, a fetal QRS template is generated. The template is generated for each processed signal separately and uses only the selected channel. For this purpose, all local maxima in the absolute derivative of the signal are determined, which have a certain minimum peak distance to their neighbor peak. Around each detected maximum, the signal is cropped to the size of the median RR interval in the recording, resulting in an array of waveforms centered around preliminary QRS peaks. The template is then calculated as the median of these waveforms. Then, the fetal ECG-enhanced signal of the selected channel is filtered with the time-reversed template. The resulting signal has peaks at locations where there is a high correlation between the signal and the template, i.e. where a fetal QRS waveform is expected.

3.1.3. Parameter optimization of Power-MF

The creation of a fetal QRS template involves the optimization of the minimum peak distance in *Power-MF*. The minimum peak distance is treated as a hyperparameter and is optimized in a training step on an independent dataset, adjusting the minimum peak distance to achieve the highest F1-score. Specific physiologically meaningful values are systematically tested, and the chosen value is set for further procedure.

3.1.3.1. Data

For parameter optimization of *Power-MF*, we used the CinC2013 dataset. It contains signals with a length of one minute and comprises four channels. They were collected with different devices, resolutions, and configurations. All recordings have a sampling frequency of 1000 Hz. The dataset contains three subsets, but only subset A is publicly available and contains reference R-peak annotations. In this work subset A was used to optimize the parameters of *Power-MF*. Due to inaccurate reference annotations, recordings a33, a38, a52, a54, and a71 were excluded, as suggested in previous publications (Behar *et al* 2014, Varanini *et al* 2014).

3.1.3.2. Evaluation and Results

For the minimum peak distance optimization, physiologically meaningful values from 290 ms to 360 ms in 10 ms steps were set. For each value, the local maxima were computed on all recordings of CinC2013 and compared with the ground truth fetal R-peak annotations. The minimum peak distance is set to the value leading to the highest F1-score.

The algorithm achieved the highest mean F1-score of 94.25% with a minimum peak distance of 340 ms on the training dataset (CinC2013). Consequently, this value was set for further procedure. Table 2 shows all results of the parameter optimization.

Table 2. Results of the parameter optimization of *Power-MF* on the CinC2013 dataset. The optimal minimum distance between two peaks was determined for the peak detection of the fetal QRS detection step. Mean F1-scores (Min, Max) in %. The highest mean F1-score was achieved at 340 ms (highlighted in bold).

Minimum peak distance	Mean F1-score in %	Min/max F1-score in %
290 ms	93.63	28.98/100.0
300 ms	93.89	29.39/100.0
310 ms	94.04	28.78/100.0
320 ms	94.09	28.46/100.0
330 ms	94.24	27.38/100.0
340 ms	**94.25**	**29.01**/100.0
350 ms	93.88	25.00/100.0
360 ms	92.76	27.34/100.0

Table 3. Characteristics of the datasets *Non-invasive Multimodal Foetal ECG-Doppler Dataset for Antenatal Cardiology Research* (NInFEA) and *Abdominal and Direct Fetal ECG Database* (ADFECG), that are used in this study.

	Year	Number of signals	Number of channels	Pregnancy week	Signal length	Sampling frequency
NInFEA (Sulas *et al* 2021)	2021	60	27 channels (22 abdominal, 3 thoracic, 2 back)	21st–27th	varies between 7.5 s and 2 min	2048 Hz
ADFECG B1 Pregnancy (Matonia *et al* 2020)	2020	10	4 abdominal channels	32nd–42nd	20 min	500 Hz
ADFECG B2 Labour (Matonia *et al* 2020)	2020	12	4 abdominal channels	32nd–42nd during labour	5 min	500 Hz

3.2. Benchmarking Power-MF against the state-of-the-art

3.2.1. Data

The most recently published open-source datasets NInFEA (Sulas *et al* 2021) and the ADFECG (Matonia *et al* 2020), consisting of the two subsets B1 Pregnancy and B2 Labour, were selected as test data for the algorithm benchmarking. An overview of the datasets is given in table 3. All datasets contain ground truth annotations. To the best of our knowledge, further publicly available datasets do not have reference annotations or are simulated data (Moor *et al* 1997, Andreotti *et al* 2016, Behar *et al* 2019).

For ADFECG the labels were acquired through automated fetal QRS detection and were subsequently verified by clinical experts (Matonia *et al* 2020). We used all leads as they were published.

For NInFEA, a synchronized pulse-wave Doppler sonography (PWD) of the fetal heart was acquired simultaneously and clinically annotated for heartbeat references. The PWD signal annotations are V-peaks, characteristic points representing the blood flow through the aortic valve. From the physiological perspective, blood flow through the aorta is immediately preceded by the depolarization and contraction of the ventricles. It can, therefore, be assumed that a V-wave directly follows an R-peak in the PWD signal. The V-peaks in the PWD signal have been annotated by expert clinicians during signal acquisition. We use all 22 abdominal channels of the NInFEA dataset for our study. Additionally, we included five electrode configurations based on market-available wearable devices, as proposed by Sulas *et al* (2021) proposed, see figure 5. We excluded the 9th channel of recording 34 in NInFEA due to corrupted signal values.

3.2.2. Algorithm preparation

The proposed algorithm *Power-MF* covers all steps of the fetal QRS detection pipeline. Its preprocessing and maternal ECG suppression overlap with *Varanini*, as these have shown to be robust. *Power-MF* differs from *Varanini* in the channel selection and fetal QRS detection steps. A complete overview of the algorithm steps of *Power-MF* and *Varanini* is given in figure 3.

The algorithms we employed for benchmarking required some adjustments to function properly on the datasets we selected. The source codes of *Behar* and *Varanini* are publicly available on the CinC2013 website[4]. For *Sulas*, we used the authors' publicly available repository on Github[5], which depends on the OSET toolbox[6]. The test datasets have sampling frequencies of 2048 Hz (NInFEA) and 500 Hz (ADFECG dataset). Since both algorithms, *Behar* and *Varanini*, were developed for CinC2013, they were published with a

[4] https://archive.physionet.org/challenge/2013/sources/.
[5] https://github.com/rsameni/NInFEADataset.
[6] https://sameni.org/OSET/.

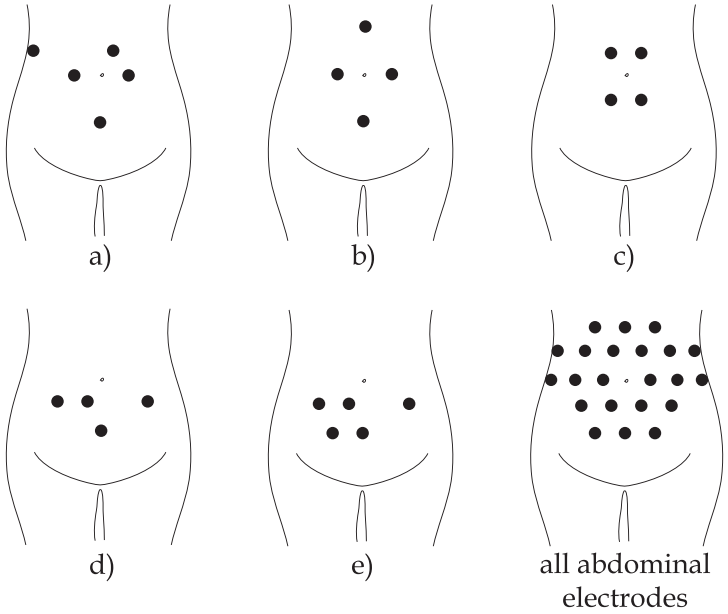


Figure 5. Electrode configurations of the NInFEA dataset. This overview shows the electrode configurations of the NInFEA dataset, as proposed by Sulas *et al* (2021). Configurations (a)–(e) model market-available wearable NI-fECG devices.

sampling frequency set to 1000 Hz. *Sulas*, on the other hand, was published with a sampling frequency set to 2048 Hz. All algorithms take the sampling frequency of the signals as an input parameter. Depending on the dataset we evaluated the algorithms on, this input parameter was set to the respective sampling frequency of the dataset. This step was necessary to ensure conformity between datasets and algorithms. Although we did not modify the core methodology behind the algorithm or the data directly, we cannot rule out the possibility that the algorithms behave differently, as they may depend on the sampling frequency. It was out of the scope of this work to test if additional modifications of this algorithm could improve the performance of the unseen dataset.

In the algorithm *Behar*, the authors defined specific parameters to improve their performance on the CinC2013 dataset. More precisely, the identified time series are flagged as implausible if less than 85 fetal QRS or more than 200 fetal QRS in the 60 s long recordings are detected or if the standard deviation of the detected RR intervals is above 17 ms. For our study, these parameters were removed, ensuring no signals longer than 60 s are rejected.

Sulas was adapted to detect maternal QRS from abdominal channels, as not all datasets contain thoracic reference channels. All algorithms were implemented in Matlab.

3.2.3. Performance metrics

For performance evaluation, we computed Se, PPV, and F1-score as follows:

$$\mathrm{Se} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{1}$$

$$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{2}$$

$$\mathrm{F1} = 2 \cdot \frac{\mathrm{PPV} \cdot \mathrm{Se}}{\mathrm{PPV} + \mathrm{Se}} = \frac{2 \cdot \mathrm{TP}}{2 \cdot \mathrm{TP} + \mathrm{FP} + \mathrm{FN}}. \tag{3}$$

The available reference fetal heartbeat annotations serve as ground truth. True positives (TP) are the number of correctly detected QRS, false positives (FP) are the number of wrongly detected QRS, and false negatives (FN) are the number of missed QRS.

For adults, a detected QRS is considered to be a true positive if it is within 150 ms of the reference annotation (Di Marco and Chiari 2011, Heryan *et al* 2021). For ADFECG, we used an interval of 50 ms due to the higher fetal heart rate, as suggested by Behar *et al* (2016). For the evaluation of the NInFEA dataset, though, a different evaluation method was chosen: In the original publication by Sulas *et al* (2021), a true positive is defined as a detected QRS that has a distance less than 200 ms to the annotation. As described before, from a physiological perspective, the QRS is expected to occur shortly before the V-peak. Therefore, we consider a QRS to be a true positive if it is within 200 ms before the annotation.

Se (1) can be interpreted as the percentage of correctly detected true fetal QRS out of the total number of true fetal QRS, i.e. it indicates how successful an algorithm is at finding the true fetal QRS. PPV (2) is the percentage of correctly detected true fetal QRS out of all detected fetal QRS. This value indicates how well the algorithm can detect true fetal QRS out of all detections. F1-score (3) is the harmonic mean of Se and PPV and is used to summarize the overall performance of a detector.

4. Results

4.1. Performance on NInFEA dataset

Table 4 shows detailed results for the five wearable electrode configurations and for all 22 abdominal electrodes, as introduced in figure 5. Our proposed algorithm *Power-MF* outperforms state-of-the-art algorithms on three of six electrode configurations (see figure 5) of the NInFEA dataset. *Power-MF* achieves an F1-score of 84.5% ± 17.7%, 89.3% ± 14.8%, and 90.5% ± 13.4% for the electrode configurations (b), (d) and (e). For electrode configurations (a) and (c), *Varanini* achieves the highest F1-score with 85.8% ± 21.0% and 89.5% ± 18.4%, respectively. On all abdominal electrodes, the algorithm *Sulas* performed best with a mean F1-score of 93.0% ± 12.4%.

Figure 6 presents boxplots of the F1-scores over the individual recordings to illustrate the distribution of scores for each algorithm. It shows that *Power-MF* and *Varanini* have a similar performance. *Sulas* performs best when all abdominal electrodes are considered. However, for configurations (a)–(e), the results are relatively low and spread over a wide range. *Behar* has a consistently mean score around 50% on all configurations.

Most algorithms achieve the best performance when all abdominal electrodes are included. The scores among the wearable electrode configurations are similar.

4.2. Performance on ADFECG dataset

Power-MF achieves the best performance on both subsets of the ADFECG dataset with a mean F1-score of 99.5% ± 0.5% on subset B1 Pregnancy and 98.0% ± 3.0% on subset B2 Labour. In table 5, detailed results of all state-of-the-art algorithms on the two subsets of the ADFECG dataset are shown. Figure 7 shows the distribution of F1-scores of the individual recordings. It is noticeable that the results among individual recordings are widely spread for *Behar* and *Sulas*, while *Power-MF* achieves the lowest standard deviation across all algorithms.

Power-MF, *Varanini* and *Behar* had the best performance in the ADFECG dataset with higher F1-scores and a smaller range of values. The number TP and FP for all algorithms are shown in tables 6 and 7. The supplementary material includes the performances on individual recordings, including F1, Se, PPV, TP, FP, and FN values.

4.3. Fetal heart rate analysis

A residual plot in figure 8 shows the relationship between the extracted heart rate of the algorithms *Power-MF* and *Varanini* and the ground truth heart rate for each recording of all datasets. For *Power-MF* the bias is 0.13 bpm and the limits of agreement (LoA) (±1.96 SD) are 6.83 bpm and −6.60 bpm. For *Varanini* the bias is 0.83 bpm and the LoA are 10.64 bpm and −9.00 bpm. Based on the analysis, there is a consistent performance of both algorithms across the entire range of heart rantes. It appears that *Power-MF* algorithm underestimated some individual recordings at higher heart rates. However, the overall limit of agreements across all heart rates is substantially lower than for *Varanini*.

5. Discussion

In this work, we propose *Power-MF*, an algorithm for robust fetal QRS detection from noisy NI-fECG recordings. Benchmarking against three relevant open-source state-of-the-art fetal QRS detection algorithms (Behar *et al* 2014, Varanini *et al* 2014, Sulas *et al* 2021) followed a standardized protocol using ADFECG (Matonia *et al* 2020) and NInFEA Sulas *et al* (2021) datasets. We chose ADFECG and NInFEA as test datasets since they are relatively new and, hence, few publications have been published using these datasets compared to the CinC2013 dataset.

In our benchmark, we identified *Power-MF* as the best-performing algorithm, as it achieves the highest F1 score and the least standard deviation on both subsets of the ADFECG datasets (table 5). Further, it outperforms the other algorithms on three electrode configurations of NInFEA (table 4). *Varanini* also performs convincingly on individual electrode configurations of NInFEA and achieves comparable results to *Power-MF* on the ADFECG dataset. However, manual spot checks revealed that *Power-MF* is more robust for

Table 4. Mean F1-scores ± standard deviation (min, max) in % for *Power-MF* and state-of-the-art fetal QRS detection algorithms on the NInFEA dataset for five different electrode configurations (see figure 5) and all abdominal electrodes, as proposed by (Sulas *et al* 2021). The results are averaged over all recordings. Minimum and maximum F1-scores are given in brackets. The best result for each configuration is highlighted in bold.

	NInFEA electrode configuration					
	(a)	(b)	(c)	(d)	(e)	All abdominal electrodes
Power-MF	85.3 ± 18.1 (45.7, 100.0)	**84.5 ± 17.7 (40.4, 100.0)**	86.6 ± 18.4 (35.2, 100.0)	**89.3 ± 14.8 (45.1, 100.0)**	**90.5 ± 13.4 (50.5, 100.0)**	91.7 ± 14.5 44.9, 100.0)
Varanini *et al* (2014)	**85.5 ± 21.0 (0.0, 100.0)**	84.4 ± 21.4 (0.0, 100.0)	**88.9 ± 18.4 (16.4, 100.0)**	88.8 ± 16.9 (48.0, 100.0)	88.4 ± 17.9 (33.7, 100.0)	92.3 ± 13.8 (46.9, 100.0)
Behar *et al* (2014)	50.8 ± 11.3 (0.0, 79.6)	50.8 ± 9.2 (0.0, 70.0)	52.4 ± 8.3 (32.0, 74.8)	49.7 ± 15.2 (0.0, 76.9)	52.2 ± 13.0 (0.0, 79.8)	52.9 ± 13.1 (0.0, 88.9)
Sulas *et al* (2021)	73.3 ± 20.5 (30.9, 100.0)	69.3 ± 21.1 (40.0, 100.0)	69.5 ± 19.9 (38.5, 100.0)	73.5 ± 19.9 (45.3, 100.0)	73.4 ± 20.2 (44.5, 100.0)	**93.0 ± 12.4 (54.0, 100.0)**

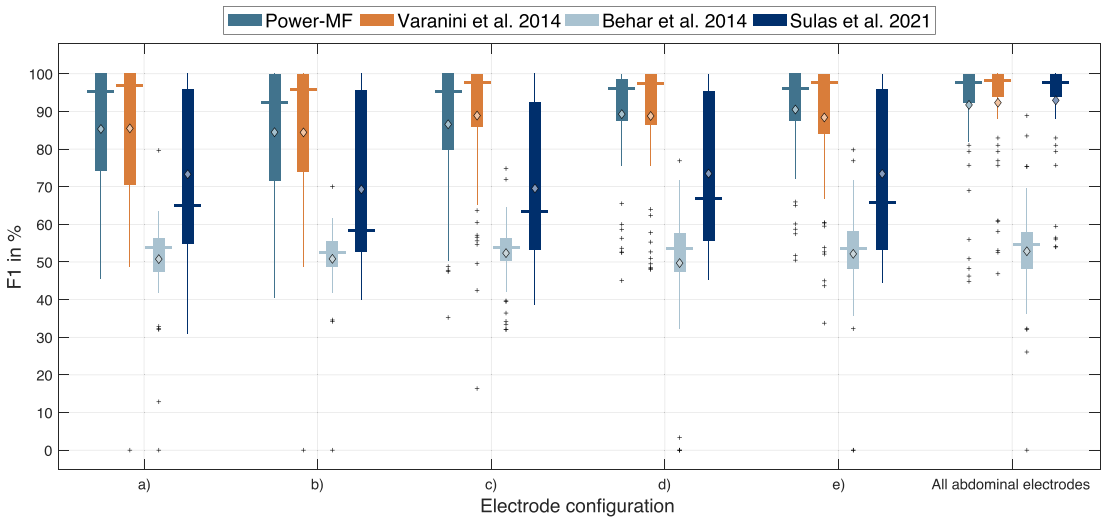


Figure 6. Boxplots of the F1-scores in % of the individual recordings of the NInFEA dataset for each electrode configuration. The horizontal bars represent the median F1-score, plus signs (+) indicate outliers. The boxes represent the interquartile range, and the vertical lines extend to the furthest observation that is not considered an outlier, i.e. that is at most 1.5 times the interquartile range from the top or bottom of the box. Diamonds (◇) indicate the mean F1-score over all recordings.

Table 5. Mean F1-scores ± standard deviation (min, max) in % for state-of-the-art fetal QRS detection algorithms on both subsets of the ADFECG dataset. The results are averaged over all recordings. Minimum and maximum F1-scores are given in brackets. The best result for each subset is highlighted in bold.

	B1 Pregnancy	B2 Labour
Power-MF	**99.5 ± 0.5 (98.4, 99.9)**	**98.0 ± 3.0 (89.1, 100.0)**
Varanini *et al* (2014)	99.4 ± 0.7 (98.1, 100.0)	97.9 ± 3.8 (86.3, 100.0)
Behar *et al* (2014)	90.4 ± 5.5 (83.2, 98.7)	87.7 ± 9.4 (64.9, 96.9)
Sulas *et al* (2021)	63.0 ± 22.0 (31.0, 99.5)	65.4 ± 27.9 (27.6, 99.7)

noisy data, see figure 9. We further showed that *Power-MF* can reliably extract fetal heart rate across all tested fetal heart rate ranges and achieves a lower bias across all datasets compared to *Varanini* (see figure 8).

We introduce *Power-MF*, a new algorithm for fetal QRS detection that builds on the *Varanini* algorithm and incorporates its effectiveness in preprocessing and maternal ECG cancellation. A major difference from

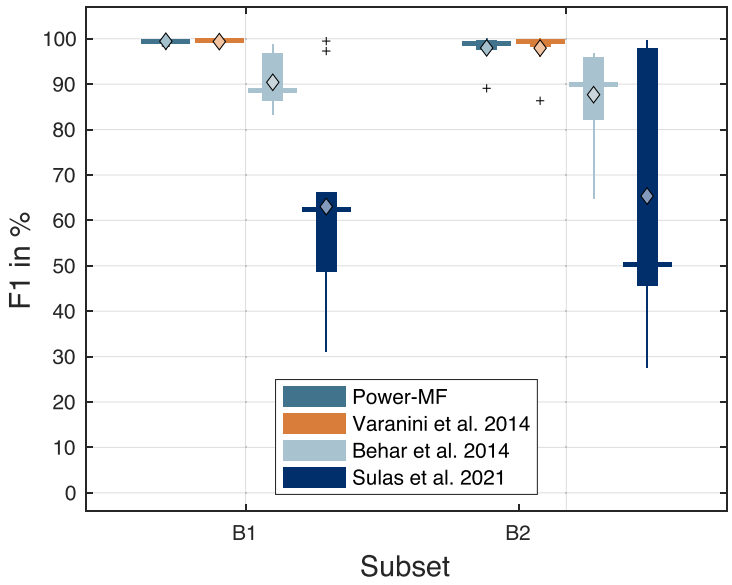


Figure 7. Boxplots of the F1-scores in % of the individual recordings of subsets B1 Pregnancy and B2 Labour of the ADFECG dataset. The horizontal bars represent the median F1-score, plus signs (+) indicate outliers. The boxes represent the interquartile range, and the vertical lines extend to the furthest observation that is not considered an outlier, i.e. that is at most 1.5 times the interquartile range from the top or bottom of the box. Diamonds (◇) indicate the mean F1-score over all recordings.

Table 6. Number of correctly detected fetal QRS peaks (true positives) for *Power-MF* and state-of-the-art QRS detection algorithms on the datasets ADFECG B1 Pregnancy and B2 Labour and NINFEA with electrode configurations (a)–(e) and all abdominal electrodes. The number of ground truth fetal QRS peaks is in bold.

	ADFECG		NInFEA					
	B1 Pregnancy	B2 Labour	(a)	(b)	(c)	(d)	(e)	All
Power-MF	28 189	7762	3677	3613	3703	3866	3939	3977
Varanini *et al* (2014)	28 264	7759	3738	3683	3868	3902	3803	4004
Behar *et al* (2014)	26 912	7507	2053	1983	2049	2020	2075	2078
Sulas *et al* (2021)	17 931	5197	3056	2783	2854	3057	3034	4067
Ground truth	**28 405**	**7903**			**4220**			

Table 7. Number of wrongly detected fetal QRS peaks (false positives) for *Power-MF* and state-of-the-art QRS detection algorithms on the datasets ADFECG B1 Pregnancy and B2 Labour and NINFEA with electrode configurations (a)–(e) and all abdominal electrodes.

	ADFECG		NInFEA					
	B1 Pregnancy	B2 Labour	(a)	(b)	(c)	(d)	(e)	All
Power-MF	165	190	573	614	544	408	339	300
Varanini *et al* (2014)	137	200	586	684	469	430	538	323
Behar *et al* (2014)	4238	1804	2248	2267	2255	2247	2285	2208
Sulas *et al* (2021)	10 495	2853	1317	1496	1454	1427	1393	248

Varanini is our introduction of a PSD-based approach for channel selection. While PSD has been used for noise removal in ECG signals (Biswas *et al* 2014) or to analyze the frequency spectrum of the fetal heart rate (Ferrazzi *et al* 1989), our novelty lies in its application specifically for selecting the channel with the most prominent fECG signal. To the best of our knowledge, this unique combination has not been explored before. As a fetal QRS detection approach, we applied a matched filter. Matched filters have been proven robust for adult QRS detection in various studies (Eskofier *et al* 2008, Smigiel and Marciniak 2017, Jamshidian-Tehrani and Sameni 2018). Additionally, (Matonia *et al* 2020) have shown its effectiveness for fetal QRS detection. Other works use e.g. derivative-based approaches, adaptive thresholds, or neural-networks for detecting the fetal QRS peaks in the signal (Varanini *et al* 2014, Mollakazemi *et al* 2021, Jebastine 2023).

The strengths of *Power-MF* lie in its robustness against noise, particularly during labor (see table 5), ensuring stable performance across a wide spectrum of fetal heart rates. The algorithm's adaptability to diverse heart rate ranges, as shown in figure 8, underscores its versatility. However, limitations arise from the optimization of the *minimum peak distance* hyperparameter for a specific heart rate range, potentially

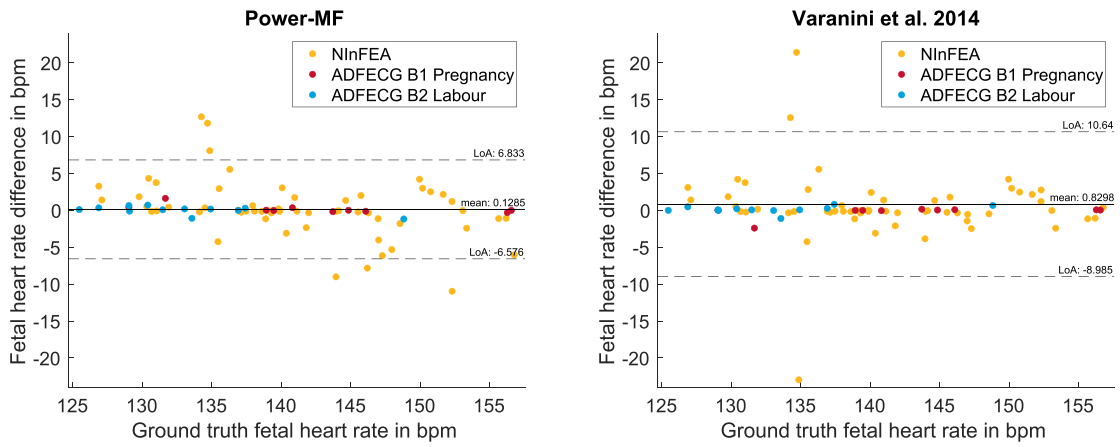


Figure 8. Residual plot for fetal heart rate differences for individual recordings in ADFECG B1 Pregnancy, B2 Labour and NInFEA datasets. Differences are plotted between *Power-MF* algorithm and ground truth (left) and between Varanini *et al* (2014) algorithm and ground truth (right). The differences are mean fetal heart rate values and each dot represents a single recording. The solid line represents the mean difference across all recordings, the dashed line represents the limits of agreement (LoA).

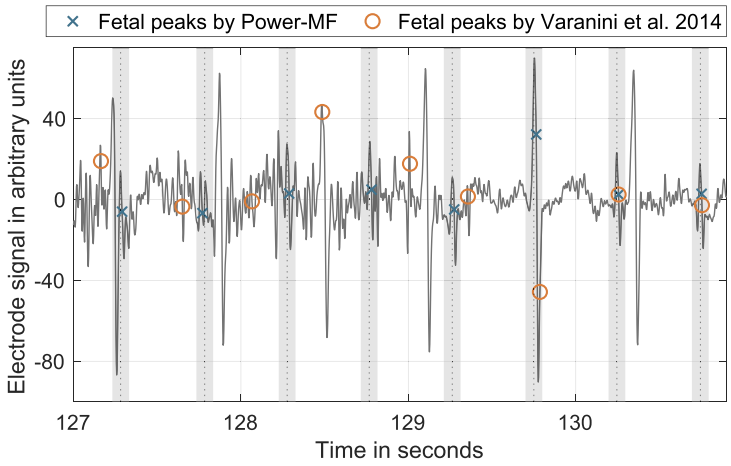


Figure 9. Fetal peak detection in a noisy signal segment during labour. This signal exemplifies that in a noisy signal segment (127.0–129.5 s) *Power-MF* detects the fetal peaks correctly while the algorithm of (Varanini *et al* 2014) misses them. In the less-noisy segment (from 129.5 s), both algorithms detect the fetal peaks correctly. The dashed vertical lines indicate the ground truth fetal peak annotations. The crosses and circles indicate the fetal peaks detected by *Power-MF* and Varanini *et al* (2014), respectively. The grey areas mark the acceptance interval of 50 ms. The signal is the fourth channel of recording 6 from the dataset ADFECG B2 Labour. The electrode signal is displayed in arbitrary units.

impacting performance at pathologically low fetal heart rates, as we aimed to avoid maternal heart rate interference. Another potential limitation of the algorithm lies in the fixed nature of the matched filter template, determined as the median of temporarily detected fetal QRS. Enhancing adaptability to situations where the QRS pattern is not stationary, such as during fetal movements, could be considered for future improvements. In order to integrate our algorithm into a new decision support system, it is necessary to test it in real-life situations that include pathological heart rate patterns and a wide range of gestational ages. This will ensure that the algorithm functions reliably for fetal heart rate extraction in routine clinical practice.

We evaluated the algorithms on different electrode configurations to get more generalizable results, including the potential performance on wearable devices. A connection between configuration and algorithm performance could not be identified, as the algorithms performed similarly for all five electrode configurations of NInFEA. It appears that more electrodes lead to better results (Aggarwal and Wei 2021). For the ADFECG dataset, we investigated only the originally published configuration, which is not comparable to those of NInFEA due to a different reference electrode location.

The algorithms exhibit lower performance on the NInFEA dataset (table 4, figure 6), possibly due to a lower data quality from recordings at an early gestational age, resulting in a lower SNR (Kahankova *et al* 2017). In about half of the NInFEA signals, fetal peaks are not visually identifiable across the 22 abdominal channels, contrasting with ADFECG, where fetal peaks are constantly visible. Performance varied among

signals, with deficiencies in signals lacking visible fetal peaks and high performance in those with visible peaks. This variation of signal quality might explain the high standard deviation in NInFEA results. Further, NInFEA does not contain reference R-peak annotations. Instead, the V-peaks in the PWD signal were used as ground truth labels. This required an increase in confidence interval due to the physiological time difference between R and V-peak, potentially leading to more falsely detected peaks being identified as correct.

Sulas shows low performance, particularly with fewer electrodes on NInFEA, suggesting reduced effectiveness with fewer channels, possibly due to the πCA method's dependency on multiple channels. Overfitting to the dataset due to the high number of channels might be conceivable. *Sulas* has been published together with the NInFEA dataset. The algorithm works well on the total dataset but might not generalize well. Our re-implementation achieves considerably higher performances than stated in the original publication (tables 1 and 4). As our work is based on the implementation of *Sulas*, we cannot rule out contradictions to the description by Sulas *et al* (2021).

A limitation of this work relates to the literature research. Although new algorithms are constantly being published, few provide the corresponding source code. These include mainly algorithms published in CinC2013 and algorithms based on the two toolboxes FECGSYN and OSET. This restriction has limited the variety of algorithms compared. To allow a comprehensive and representative comparison of state-of-the-art algorithms, we encourage authors to make their source code publicly available.

As a performance metric, we reported the F1-score as it describes the harmonic mean between Se and PPV. However, for other applications, e.g. morphological analysis, other metrics such as true and false positives might be more appropriate since the focus is more on the individual heartbeats than on the completeness of the detected heartbeats.

Another limitation concerns the algorithm's ability to adapt to different sampling frequencies. Open-source algorithms were typically published with a specific sampling frequency linked to the dataset for which they were initially developed. However, the datasets used for the algorithm comparison in this work had deviating sampling frequencies, necessitating adjustments in the source code. Although we did not modify the algorithm or data directly, we cannot rule out the possibility that the algorithms behave differently, as they may depend on the sampling frequency. It is important to point out that there might be an effect of different sampling frequencies on the performance of the algorithms and should, therefore, be investigated in future studies to evaluate the algorithms under different sampling conditions.

The CinC2013 dataset served as training set and was, therefore, not the primary evaluation dataset in our study. Comprehensive analysis results for this dataset are available in the supplemental material. Reproducing exact results from original algorithm publications was challenging, potentially due to differing experimental setups and software environments. Notably, *Behar* and *Varanini* were tailored for the CinC2013 dataset, which may have contributed to their slightly better performance compared to *Power-MF* on the CinC2013 dataset. However, *Power-MF* demonstrates promising generalization capabilities, maintaining a stable performance across all datasets, including labor conditions, despite comparatively lower performance on CinC2013.

We performed a parameter optimization of *Power-MF* on the CinC2013 dataset. *Varanini* and *Behar* have been developed for CinC2013. *Sulas* has been presented with the NInFEA dataset. We applied state-of-the-art algorithms as proposed in their original publications without further parameter optimization. A parameter optimization on CinC2013 might have improved *Sulas*' (Sulas *et al* 2021) performance on the ADFECG dataset.

Our work addresses methodological limitations, as highlighted by Kahankova *et al* (2020), ensuring an objective comparison by excluding incomplete, non-public, or synthetic datasets (Clifford *et al* 2014, Andreotti *et al* 2016, Li *et al* 2017, Vaidya and Chaitra 2020). Similar to Li *et al* (2017), we investigated the performance among different electrode configurations relevant to wearable devices and are using the maximum available electrodes in the datasets for thorough evaluation. Unlike previous studies employing algorithms from a single toolbox, we established a standardized evaluation using re-implemented algorithms, similar to the approach by Clifford *et al* (2014) in the CinC2013 evaluation. This approach enabled us to account for variations in data subsets and evaluation metrics, resulting in a more robust evaluation. In order to gain more detailed insights into the algorithm's capacity to preserve the morphological structures of the extracted fECG, a morphological analysis becomes essential. However, such analysis requires the availability of morphologically annotated or simulated data, as done by Andreotti *et al* (2016). In this study, we focused our analysis on real-world data. While we chose not to use simulated data, acknowledging its potential for a more detailed evaluation under varying noise levels, we plan to explore this in our future research.

We achieved acceptable performance of state-of-the-art algorithms for late pregnancy data. However, there is room for improvement in extracting data from early pregnancy, which necessitates further research and evaluation. Our focus centered on evaluating fetal QRS detection using a limited number of channels, as wearable NI-fECG devices often have a constrained electrode count. Despite *Power-MF* exhibiting only a

marginal performance gain compared to *Varanini*, we are convinced that we are heading in the right direction. Notably, our algorithm shows promising potential for stable fetal heart rate extraction from wearable devices, even with sparse channels and noisy recordings in early pregnancy.

Specific fetal heart rate patterns during labour provide insights into the presence of perinatal asphyxia (Ayres-de Campos *et al* 2015). *Power-MF* exhibits consistent performance on labour recordings (ADFECG B2 Labour: mean F1 $= 98.0 \pm 3.0$) and outperforms state-of-the-art algorithms (see table 5), making it the most suitable algorithm for additional investigations concerning the early detection of perinatal asphyxia. Further analyses will determine whether the extraction of these specific fetal heart rate patterns is feasible using *Power-MF*.

In our upcoming research, we plan to tackle these issues by collecting a new NI-fECG dataset. This dataset will serve as a foundation to evaluate *Power-MF's* capability to reconstruct specific fetal heart rate patterns indicative of perinatal asphyxia based on the detected fetal R-peaks. Further, this dataset will be used to further develop *Power-MF* with an adaptive template and validate its robustness to extract moving heart rate patterns and pathologic fetal heart rates.

6. Conclusion

This paper presents *Power-MF*, a new approach for fetal QRS detection. Our experimental results demonstrate that *Power-MF* outperforms three state-of-the-art algorithms on two recently published NI-fECG datasets. Furthermore, *Power-MF* is stable for different electrode configurations relevant to wearable devices and is robust against noisy recordings, especially during labour.

The development of *Power-MF* contributes to the advancement of fetal QRS detection algorithms. The stability and reliability of such algorithms play a crucial role in the early detection and diagnosis of birth complications, e.g. perinatal asphyxia. Our long-term goal of reducing costs and making prenatal care more accessible underscores the significance of these efforts in enhancing fetal QRS detection algorithms.

Our plans for a newly collected dataset will also help to answer other research questions that could not be answered with the existing datasets. For instance, morphological analyses of the NI-fECG have the potential to improve the early detection of perinatal asphyxia (Oudijk *et al* 2004), as well as cardiac malfunctions (Velayo *et al* 2011, Verdurmen *et al* 2016, Lakhno *et al* 2017). Additionally, as we are strong advocates of open and reproducible science, we plan to make our dataset available to the public, thus, also allowing and encouraging other researchers to develop and evaluate new approaches for fECG extraction to address the important issue of prenatal complications.

Data availability statement

The source code of this analysis is available on GitHub (https://github.com/mad-lab-fau/fecg-benchmarking). The datasets analyzed during this study are available in online repositories: CinC2013 (https://physionet.org/content/challenge-2013/1.0.0/), NInFEA (https://doi.org/10.13026/c4n5-3b04), ADFECG (https://doi.org/10.6084/m9.figshare.c.4740794.v1).

Acknowledgment

We want to thank the authors of the applied algorithms for providing the source code and their willingness to answer questions.

This work was supported by the Federal Ministry of Health on the basis of a decision by the German Bundestag (Grant Number ZMVI1-2519DAT400). Bjoern M Eskofier gratefully acknowledges the support of the German Research Foundation (DFG) within the framework of the Heisenberg professorship programme (Grant Number ES 434/8-1).

ORCID iD

Katharina M Jaeger https://orcid.org/0000-0002-2478-4079

References

Aggarwal G and Wei Y 2021 Non-invasive fetal electrocardiogram monitoring techniques: potential and future research opportunities in smart textiles *Signals* **2** 392–412

Andreotti F, Behar J, Zaunseder S, Oster J and Clifford G D 2016 An open-source framework for stress-testing non-invasive foetal ECG extraction algorithms *Physiol. Meas.* **37** 627–48

Ayres-de Campos D and Arulkumaran S 2015 FIGO consensus guidelines on intrapartum fetal monitoring: physiology of fetal oxygenation and the main goals of intrapartum fetal monitoring *Int. J. Gynecol. Obstet.* **131** 5–8

Behar J A, Bonnemains L, Shulgin V, Oster J, Ostras O and Lakhno I 2019 Noninvasive fetal electrocardiography for the detection of fetal arrhythmias *Prenat. Diagn.* **39** 178–87
Behar J, Andreotti F, Zaunseder S, Oster J and Clifford G D 2016 A practical guide to non-invasive foetal electrocardiogram extraction and analysis *Physiol. Meas.* **37** R1–R35
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** 1569–89
Benton M, Salter A, Simpson B, Wilkinson C and Turnbull D 2020 A qualitative study of a sample of women participating in an Australian randomised controlled trial of intrapartum fetal surveillance *Midwifery* **83** 102655
Biswas U, Das A, Debnath S and Oishee I 2014 ECG signal denoising by using least-mean-square and normalised-least-mean-square algorithm based adaptive filter *2014 Int. Conf. on Informatics, Electronics & Vision* (*ICIEV*) pp 1–6
Cardoso J and Souloumiac A 1993 Blind beamforming for non-gaussian signals *IEEE Proc. F* **140** 362
Chen H-Y, Chauhan S P, Ananth C V, Vintzileos A M and Abuhamad A Z 2011 Electronic fetal heart rate monitoring and its relationship to neonatal and infant mortality in the United States *Am. J. Obstet. Gynecol.* **204** 491.e1-10
Clifford G D, Silva I, Behar J and Moody G B 2014 Non-invasive fetal ECG analysis *Physiol. Meas.* **35** 1521–36
Di Marco L Y and Chiari L 2011 A wavelet-based ECG delineation algorithm for 32-bit integer online processing *Biomed. Eng.* **10** 23
Eskofier B M, Kornhuber J and Hornegger J 2008 Embedded QRS detection for noisy ECG sensor data using a matched filter and directed graph search *4th Russian-Bavarian Conf. on Bio-Medical Engineering* (*MIET, Moscow*)
Ferrazzi E, Pardi G, Setti P L, Rodolfi M, Civardi S and Cerutti S 1989 Power spectral analysis of the heart rate of the human fetus at 26 and 36 weeks of gestation *Clin. Phys. Physiol. Meas.* **10** 57–60
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-K and Stanley H E 2000 PhysioBank, PhysioToolkit and PhysioNet: components of a new research resource for complex physiologic signals *Circulation* **101** e215-20
Hasan M A, Reaz M B I, Ibrahimy M I, Hussain M S and Uddin J 2009 Detection and processing techniques of FECG signal for fetal monitoring *Biol. Proced.* **11** 263–95
Hayes-Gill B R, Martin T R P, Liu C and Cohen W R 2020 Relative accuracy of computerized intrapartum fetal heart rate pattern recognition by ultrasound and abdominal electrocardiogram detection *Acta Obstet. Gynecol. Scand.* **99** 413–22
Heryan K, Reklewski W, Szaflarski A, Ordowski M, Augustyniak P and Miskowicz M 2021 *Sensitivity of QRS Detection Accuracy to Detector Temporal Resolution* (*2021 Computing in Cardiology* (*CinC*) (*Brno, Czech Republic*) (IEEE) pp 31–4
Hyvarinen A 1999 Fast and robust fixed-point algorithms for independent component analysis *IEEE Trans. Neural Netw.* **10** 626–34
Jaeger K M, Nissen M, Richer R, Rahm S, Titzmann A, Fasching P A, Eskofier B M and Leutheuser H 2022 Machine learning-based detection of in-utero fetal presentation from non-invasive fetal ECG *2022 IEEE-EMBS Int. Conf. on Biomedical and Health Informatics (BHI)* pp 01–04
Jamshidian-Tehrani F and Sameni R 2018 Fetal ECG extraction from time-varying and low-rank noninvasive maternal abdominal recordings *Physiol. Meas.* **39** 125008
Jebastine J 2023 Fetal ECG extraction and QRS detection using advanced adaptive filtering-based signal decomposition and peak threshold technique from abdominal ECG signals *Circuits Syst. Signal Process.* **42** 6058–88
Jezewski J, Wrobel J, Matonia A, Horoba K, Martinek R, Kupka T and Jezewski M 2017 Is abdominal fetal electrocardiography an alternative to doppler ultrasound for FHR variability evaluation? *Front. Physiol.* **8** 305
Kahankova R, Jezewski J, Nedoma J, Jezewski M, Fajkus M, Kawala-Janik A, Wen H and Martinek R 2017 Influence of gestation age on the performance of adaptive systems for fetal ECG extraction *Adv. Electr. Electron. Eng.* **15** 491–501
Kahankova R, Martinek R, Jaros R, Behbehani K, Matonia A, Jezewski M and Behar J A 2020 A review of signal processing techniques for non-invasive fetal electrocardiography *IEEE Rev. Biomed. Eng.* **13** 51–73
Lakhno I, Behar J A, Oster J, Shulgin V, Ostras O and Andreotti F 2017 The use of non-invasive fetal electrocardiography in diagnosing second-degree fetal atrioventricular block *Matern. Health Neonatol. Perinatol.* **3** 14
Li R, Frasch M G and Wu H-T 2017 Efficient fetal-maternal ECG signal separation from two channel maternal abdominal ECG via diffusion-based channel selection *Front. Physiol.* **8** 277
Liu B, Thilaganathan B and Bhide A 2023 Effectiveness of ambulatory non-invasive fetal electrocardiography: impact of maternal and fetal characteristics *Acta Obstet. Gynecol. Scand.* **102** 577–84
Matonia A, Jezewski J, Kupka T, Jezewski M, Horoba K, Wrobel J, Czabanski R and Kahankowa R 2020 Fetal electrocardiograms, direct and abdominal with reference heartbeat annotations *Sci. Data* **7** 200
Mollakazemi M J, Asadi F, Tajnesaei M and Ghaffari A 2021 Fetal QRS detection in noninvasive abdominal electrocardiograms using principal component analysis and discrete wavelet transforms with signal quality estimation *J. Biomed. Phys. Eng.* **11** 197
Moor B, Gersem P D, Schutter B and Favoreel W 1997 DAISY: a database for identification of systems *J. A* **38** 4–5
Ojala K, Vääräsmäki M, Mäkikallio K, Valkama M and Tekay A 2006 A comparison of intrapartum automated fetal electrocardiography and conventional cardiotocography-a randomised controlled study *BJOG: Int. J. Obstet. Gynaecol.* **113** 419–23
Oudijk M A, Kwee A, Visser G H A, Blad S, Meijboom E J and Rosén K G 2004 The effects of intrapartum hypoxia on the fetal QT interval *BJOG* **111** 656–60
Pan J and Tompkins W J 1985 A real-time QRS detection algorithm *IEEE Trans. Biomed. Eng.* **32** 230–6
Sameni R and Clifford G D 2010 A review of fetal ECG signal processing; issues and promising directions *Open Pacing, Electrophysiol. Ther. J.* **3** 4–20
Sänger N, Hayes-Gill B, Schiermeier S, Hatzmann W, Yuan J, Herrmann E, Louwen F and Reinhard J 2012 Prenatal foetal non-invasive ECG instead of Doppler CTG - a better alternative? *Geburtshilfe Frauenheilkunde* **72** 630–3
Smigiel S and Marciniak T 2017 Detection of QRS complex with the use of matched filtering *Innovations in Biomedical Engineering* (*Advances in Intelligent Systems and Computing*) ed M Gzik, E Tkacz, Z Paszenda and E Pietka (Springer) pp 310–22
Smith V, Arunthavanathan S, Nair A, Ansermet D, da Silva Costa F and Wallace E M 2018 A systematic review of cardiac time intervals utilising non-invasive fetal electrocardiogram in normal fetuses *BMC Pregnancy Childbirth* **18** 370
Sulas E, Urru M, Tumbarello R, Raffo L, Sameni R and Pani D 2021 A non-invasive multimodal foetal ECG–Doppler dataset for antenatal cardiology research *Sci. Data* **8** 30
Vaidya R and Chaitra N 2020 Comparison of adaptive filters in extraction of fetal ECG *Proc. - Int. Conf. on Smart Electronics and Communication* (*ICOSEC 2020*) (https://doi.org/10.1109/ICOSEC49089.2020.9215256)
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** 1607–19

Velayo C, Sato N, Ito T, Chisaka H, Yaegashi N, Okamura K and Kimura Y 2011 Understanding congenital heart defects through abdominal fetal electrocardiography: case reports and clinical implications *J. Obstet. Gynaecol. Res.* **37** 428–35

Verdurmen K M J, Eijsvoogel N B, Lempersz C, Vullings R, Schroer C, Laar J O E H v and Oei S G 2016 A systematic review of prenatal screening for congenital heart disease by fetal electrocardiography *Int. J. Gynecol. Obstet.* **135** 129–34

Wakefield C, Yao L, Self S and Frasch M G 2022 Wearable technology for health monitoring during pregnancy: an observational cross-sectional survey study *Arch. Gynecol. Obstet.* **308** 73–78

Wang B, Zeng H, Liu J and Sun M 2021 Effects of prenatal hypoxia on nervous system development and related diseases *Front. Neurosci.* **15** 755554

Welch P 1967 The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms *IEEE Trans. Audio Electroacoust.* **15** 70–73