Comparison of Fetal Heart Rate Baseline Estimation by the Cardiotocograph Network and Clinicians: A Multidatabase Retrospective Assessment Study

Xiuyu Pan  
Guangzhou Women and Children's Medical Center

Yaosheng Lu  
Jinan University

Mei Zhong  
Southern Medical University Nanfang Hospital

Huijin Wang  
Jinan University

Zheng Zheng  
Guangzhou Women and Children's Medical Center

Jieyun Bai (✉️ bai_jieyun@126.com)  
Jinan University  https://orcid.org/0000-0002-2847-350X

Research Article

Keywords: Artificial Intelligence, Deep Learning, Fetal Heart Rate, Baseline, Computer Analysis

Posted Date: August 3rd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1793588/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Objective

This study aims to compare the fetal heart rate (FHR) baseline predicted by the cardiotocograph network (CTGNet) with that estimated by clinicians.

Results

A total of 1267 FHR recordings acquired with different electrical fetal monitors (EFM) were collected from five datasets. Each FHR baseline was estimated by clinicians and CTGNet, respectively. Agreement between CTGNet and clinicians was evaluated using the kappa statistics, intra-class correlation coefficient, and the limits of agreement. The number of differences < 3 beats per minute (bpm), 3–5 bpm, 5–10 bpm and ≥ 10 bpm, is 64.88%, 15.94%, 14.44% and 4.74%, respectively. Kappa statistics and intra-class correlation coefficient are 0.873 and 0.969, respectively. Limits of agreement are −6.81 and 7.48 (mean difference: 0.36 and standard deviation: 3.64). An excellent agreement was found between CTGNet and clinicians in the baseline estimation from FHR recordings with different signal loss rates.

1. Introduction

Although deaths in children have declined substantially in the past 30 years, more than 5 million still die every year[1]. Electronic fetal heart rate (FHR) monitoring was introduced to detect fetuses' pathological states as early as possible in the obstetrics practice in the late 1950s. However, the misinterpretation and ambiguity of FHR patterns may increase unnecessary interventions, such as operative deliveries and cesarean sections[2–4]. Different guidelines over the past decades have recommended some modifications for interpreting FHR tracings, but beliefs in the etiology of basic FHR patterns (including the baseline, the variability, accelerations, decelerations, and sinusoidal patterns) have remained essentially unchanged[5, 6]. In these FHR patterns, the baseline is a pre-requisite for evaluating the other patterns[7]. Gynecologists and obstetricians usually estimated the baseline by visual analysis, but the unreliability of visual interpretation with a high degree of inter-and intra-observer variability is found[7–12]. Therefore, computer-assisted analysis has been sought to mitigate the variability of visual explanation [13–17].

The Lian-Med Technology Co. Ltd. developed an artificial intelligence-based program (cardiotocograph network, CTGNet) for automatic FHR analysis [18]. Using the CTGNet, the baseline is firstly calculated by detecting unstable segments with a convolutional neural network and then excluding unreliable points with a progressive trimming approach with long-term and short-term sliding windows. Its performance has been compared with 12 traditional methods [10, 14, 15, 19–27] on a single-center dataset with high-quality FHR recordings. Although the trained CTGNet achieved the best estimation[18], its performance on FHR tracings acquired with different devices in multi-centers is unclear.
To evaluate the clinical usability of the CTGNet, we compare the FHR baseline predicted by the CTGNet with that estimated by clinicians using a large dataset with 1267 FHR recordings acquired with fetal monitors of five device manufacturers.

2. Main Text

2.1 Material and Methods

2.1.1 Data Source

This prospective assessment study used 1267 FHR recordings from 5 datasets covering the years 2011 to 2021: (1) an FHR dataset with 84 recordings collected from the Guangzhou Women and Children's Medical Center of Guangzhou Medical University (GMU_DB; May 2021 to July 2021); (2) a dataset with 331 CTG records of the First Affiliated Hospital of Jinan University (JNU_DB; January 2015 to December 2020); (3) a dataset with 234 CTG records collected from the NanFang Hospital of Southern Medical University (SMU_DB; January 2012 to December 2020); (4) the open-access database with 66 FHR recordings collected at Saint Vincent de Paul Maternity Hospital of Lille Catholic University (LCU_DB; February 1st, 2011, and December 31st, 2016)\[13, 28, 29\] and the open-access database with 552 CTG recordings collected at the obstetrics ward of the University Hospital in Brno (UHB_DB; April 2010 and August 2012)\[30–32\]. In these datasets, the signal loss rate of FHR recordings from GMU_DB, JNU_DB and SMU_DB is < 10% per 10 mins, whereas those from GMU_DB and UHB_DB are < 7% and < 50%, respectively (Fig. 1).

The Medical Ethics Committees of the Guangzhou Women and Children's Medical Center (273A01), the Jinan University (JNUKY-2022-018) and the NanFang Hospital of Southern Medical University (NFEC-2019-024) approved this retrospective study.

2.1.2 Devices

Fetal monitors used for acquiring FHR recordings include: F15 EFM (Edan, Shenzhen, China), SRF618B5 EFM (Sanrui, Guangzhou, China), F3 EFM (Lian-Med, Guangzhou, China), STAN S21 and S31 (Neoventa Medical, Mölndal, Sweden), and Avalon FM40 and FM50 (Philips Healthcare, Amsterdam, The Netherlands).

2.1.3 Methods

According to the baseline definition of the FIGO consensus guideline\[5\]: (1) the baseline is estimated as the mean level of the most horizontal and less oscillatory FHR segments of 10 minutes; (2) it is necessary to review previous and subsequent 10-minute sections to estimate the baseline in recordings with unstable FHR signals. Clinicians (Z.Z. and X.P.) with more than seven years of experience in CTG analysis independently assessed the baselines of FHR tracings. In order to obtain a consistent baseline, FHR recordings were re-evaluated when the differences between clinicians exceeded three bpm, and the baseline was determined as an average of these clinicians’ estimations when their difference was less
than three bpm. A difference between the baseline estimated by clinicians and that predicted with the CTGNet were then computed to evaluate their agreement.

2.1.4 Statistical analysis

For each FHR recording, baseline values estimated by CTGNet and the consensus of clinicians were attributed to 5 bpm classes (such as class 0: ≤100 bpm, class 1: 100 < baseline ≤ 105, and class 2: 105 < baseline ≤ 110) in the following manner [7]: (1) when the baseline difference does not exceed five bpm, CTGNet’s and clinicians’ baselines are assigned to the same class according to their mean (e.g., if CTGNet’s baseline value is 109 and clinicians’ estimation is 113, both values are assigned to the class 110–115); (2) when the baseline difference exceeds five bpm, baseline values are assigned to their respective classes. Kappa and intra-class correlation (ICC) coefficient values (i.e., excellent agreement: >0.75, good agreement: 0.4–0.75 and poor agreement: <0.4) were calculated to evaluate agreement in the baseline estimation and 95% confidence intervals (95% CI) were computed for all results.

2.2 Results

Table 1 summarizes the comparisons of the baselines predicted by the CTGNet and those estimated by clinicians. In 99% of FHR recordings from GMU_DB, JNU_DB and SMU_DB, differences do not exceed 5 bpm, whereas the percentages of the number of differences less than 5 bpm dropped to 58–85% in low-quality FHR recordings from LCU_DB and UHB_DB. For the baseline difference ≥ 5 bpm, proportions are <0.5% for datasets from China (i.e., GMU_DB, JNU_DB and SMU_DB), 13.6% for LCU_DB and 42.2% for UHB_DB, respectively. In general, the ratio of differences < 5 bpm and ≥ 5 bpm is 4:1 among the 1267 FHR recordings.
Table 1
Comparisons of the baselines predicted by the CTGNet and those estimated by clinicians.

<table>
<thead>
<tr>
<th></th>
<th>GMU_DB (n = 84)</th>
<th>JNU_DB (n = 331)</th>
<th>SMU_DB (n = 234)</th>
<th>LCU_DB (n = 66)</th>
<th>UHB_DB (n = 552)</th>
<th>All (n = 1267)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean difference (bpm) (95% IC)</td>
<td>0.09 (-0.05 to 0.22)</td>
<td>-0.20 (-0.27 to -0.14)</td>
<td>0.88 (0.71 to 1.03)</td>
<td>0.72 (-0.99 to 2.41)</td>
<td>0.42 (0.02 to 0.83)</td>
<td>0.36 (0.14 to 0.54)</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.63</td>
<td>0.58</td>
<td>1.28</td>
<td>7.05</td>
<td>4.83</td>
<td>3.64</td>
</tr>
<tr>
<td>Maximum differences (bpm)</td>
<td>2.81</td>
<td>3.43</td>
<td>16.52</td>
<td>48.88</td>
<td>18.86</td>
<td>48.88</td>
</tr>
<tr>
<td>Number of differences &lt; 3 bpm</td>
<td>83 (98.81%)</td>
<td>327 (98.79%)</td>
<td>227 (97.01%)</td>
<td>44 (66.67%)</td>
<td>141 (25.54%)</td>
<td>822 (64.88%)</td>
</tr>
<tr>
<td>Number of differences 3–5 bpm</td>
<td>1 (1.19%)</td>
<td>4 (1.21%)</td>
<td>6 (2.56%)</td>
<td>13 (19.70%)</td>
<td>178 (32.25%)</td>
<td>202 (15.94%)</td>
</tr>
<tr>
<td>Number of differences 5–10 bpm</td>
<td>0 (0.00%)</td>
<td>0 (0.00%)</td>
<td>0 (0.00%)</td>
<td>6 (9.09%)</td>
<td>177 (32.07%)</td>
<td>183 (14.44%)</td>
</tr>
<tr>
<td>Number of differences ≥ 10 bpm</td>
<td>0 (0.00%)</td>
<td>0 (0.00%)</td>
<td>1 (0.43%)</td>
<td>3 (4.55%)</td>
<td>56 (10.14%)</td>
<td>60 (4.74%)</td>
</tr>
<tr>
<td>Lower limit (95% IC)</td>
<td>-0.82 (-0.98 to -0.64)</td>
<td>-1.34 (-1.45 to -1.24)</td>
<td>-1.63 (-1.91 to -1.35)</td>
<td>-13.10 (-10.12 to -16.07)</td>
<td>-9.05 (-9.74 to -8.36)</td>
<td>-6.81 (-7.15 to -6.46)</td>
</tr>
<tr>
<td>Upper limit (95% IC)</td>
<td>0.93 (0.77 to 1.10)</td>
<td>0.93 (0.83 to 1.04)</td>
<td>3.39 (3.11 to 3.67)</td>
<td>14.52 (11.55 to 17.50)</td>
<td>9.90 (9.21 to 10.59)</td>
<td>7.48 (7.14 to 7.82)</td>
</tr>
<tr>
<td>Kappa (95% IC)</td>
<td>0.991 (0.974 to 1.008)</td>
<td>0.997 (0.992 to 1.001)</td>
<td>0.983 (0.966 to 1.000)</td>
<td>0.885 (0.808 to 0.963)</td>
<td>0.771 (0.745 to 0.797)</td>
<td>0.873 (0.857 to 0.889)</td>
</tr>
<tr>
<td>ICC (95% IC)</td>
<td>0.998 (0.997 ~ 0.999)</td>
<td>0.999 (0.999 ~ 1.000)</td>
<td>0.995 (0.993 ~ 0.996)</td>
<td>0.956 (0.928 ~ 0.973)</td>
<td>0.955 (0.947 ~ 0.962)</td>
<td>0.969 (0.966 ~ 0.972)</td>
</tr>
</tbody>
</table>

ICC: intra-class correlation coefficient, IC: confidence intervals.
Table 2
The performance of different computer-assisted methods.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Type</th>
<th>Dataset</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dawes et al. (1982)[38, 39]</td>
<td>Signal Processing</td>
<td>Private dataset $(n = 41)[33]$</td>
<td>SI :45.84%, ADIR: 30.70%</td>
</tr>
<tr>
<td>Dalton et al. (1984)[40]</td>
<td>Signal Processing</td>
<td>Private dataset $(n = 41)[33]$</td>
<td>SI :50.08%, ADIR: 36.65%</td>
</tr>
<tr>
<td>Searle et al. (1988)[41]</td>
<td>Signal Processing</td>
<td>Private dataset $(n = 41)[33]$</td>
<td>SI : 40.84%, ADIR: 30.05%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Private dataset $(n = 234)[42]$</td>
<td>BL:1.79 bpm, SI : 55.13%, MADI: 2.98%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Private dataset $(n = 234)[42]$</td>
<td>BL:3.97 bpm, SI : 83.13%, MADI: 9.80%</td>
</tr>
<tr>
<td>Aeyels et al. (1992)[43]</td>
<td>Signal Processing</td>
<td>Private dataset $(n = 41)[33]$</td>
<td>SI : 47.37%, ADIR: 35.07%</td>
</tr>
<tr>
<td>Bernardes et al. (1993) [44, 45]</td>
<td>Signal Processing</td>
<td>Private dataset $(n = 41)[33]$</td>
<td>SI : 56.18%, ADIR: 36.33%</td>
</tr>
<tr>
<td>Arduini et al. (1993)[9, 34]</td>
<td>Signal Processing</td>
<td>Private dataset $(n = 41)[33]$</td>
<td>SI : 36.97%, ADIR: 26.33%</td>
</tr>
<tr>
<td>van Alphen et al. (1994) [48]</td>
<td>Signal Processing</td>
<td>Private dataset $(n = 41)[33]$</td>
<td>SI : 48.57%, ADIR: 35.28%</td>
</tr>
<tr>
<td>Mantel et al. (1994)[49]</td>
<td>Signal Processing</td>
<td>Private dataset $(n = 41)[33]$</td>
<td>SI : 44.03%, ADIR: 30.88%</td>
</tr>
<tr>
<td>Mongelli et al. (1997)[10]</td>
<td>Signal Processing</td>
<td>Private dataset $(n = 41)[33]$</td>
<td>SI : 45.45%, ADIR: 34.26%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Private dataset $(n = 90)[13]$</td>
<td>BL:4.7 bpm, SI : 64%, MADI: 12%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Private dataset $(n = 234)[42]$</td>
<td>BL:2.48 bpm, SI : 55.09%, MADI: 6.03%</td>
</tr>
</tbody>
</table>

* was reprogrammed and evaluated on the private dataset $(n = 90)[13]$. BL: the root mean-squared difference between baselines; SI: the synthetic inconsistency coefficient; MADI: the morphological analysis discriminant index; ADIR: the inconsistency between two baselines calculated using differences in the number, location and area of the matching episodes (i.e., acceleration and deceleration)
<table>
<thead>
<tr>
<th>Methods</th>
<th>Type</th>
<th>Dataset</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Private dataset (n = 234)[42]</td>
<td>BL: 3.82 bpm, SI: 76.51%, MADI: 9.34%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Private dataset (n = 234)[42]</td>
<td>BL: 3.52 bpm, SI: 89.13%, MADI: 10.23%</td>
</tr>
<tr>
<td>Ayres-de Campos et al. (2002)[23]</td>
<td>Signal Processing</td>
<td>Private dataset (n = 90)[13]</td>
<td>BL: 5.6 bpm, SI: 71%, MADI: 12%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Private dataset (n = 234)[42]</td>
<td>BL: 3.08 bpm, SI: 61.43%, MADI: 7.52%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Private dataset (n = 234)[42]</td>
<td>BL: 3.35 bpm, SI: 63.29%, MADI: 9.18%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Private dataset (n = 234)[42]</td>
<td>BL: 4.50 bpm, SI: 86.79%, MADI: 12.88%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Private dataset (n = 234)[42]</td>
<td>BL: 2.76 bpm, SI: 57.40%, MADI: 6.32%</td>
</tr>
<tr>
<td>Maeda et al. (2012)[26]</td>
<td>Signal Processing</td>
<td>Private dataset (n = 90)[13]</td>
<td>BL: 5.7 bpm, SI: 71%, MADI: 15%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Private dataset (n = 234)[42]</td>
<td>BL: 3.64 bpm, SI: 66.75%, MADI: 11.38%</td>
</tr>
<tr>
<td>Wrobel et al. (2013)[27]</td>
<td>Signal Processing</td>
<td>Private dataset (n = 41)[33]</td>
<td>SI: 43.70%, ADIR: 33.00%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Private dataset (n = 90)[13]</td>
<td>BL: 5.0 bpm, SI: 65%, MADI: 12%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Private dataset (n = 234)[42]</td>
<td>BL: 3.30 bpm, SI: 64.94%, MADI: 10.60%</td>
</tr>
</tbody>
</table>

* was reprogrammed and evaluated on the private dataset (n = 90)[13]. BL: the root mean-squared difference between baselines; SI: the synthetic inconsistency coefficient; MADI: the morphological analysis discriminant index; ADIR: the inconsistency between two baselines calculated using differences in the number, location and area of the matching episodes (i.e., acceleration and deceleration)
### Methods

<table>
<thead>
<tr>
<th>Methods</th>
<th>Type</th>
<th>Dataset</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Private dataset (n = 234)[42]</td>
<td>BL: 2.29 bpm, SI : 51.94%, MADI: 4.67%</td>
</tr>
<tr>
<td>Lu et al.(2020)[16]*</td>
<td>Signal Processing</td>
<td>Private dataset (n = 90)[13]</td>
<td>BL: 4.59 bpm, SI : 76.5%, MADI: 15.6%</td>
</tr>
</tbody>
</table>

* was reprogrammed and evaluated on the private dataset (n = 90)[13]. BL: the root mean-squared difference between baselines; SI: the synthetic inconsistency coefficient; MADI: the morphological analysis discriminant index; ADIR: the inconsistency between two baselines calculated using differences in the number, location and area of the matching episodes (i.e., acceleration and deceleration).

According to Kappa values (i.e., excellent agreement: Kappa > 0.75, good agreement: 0.4 ≤ Kappa ≤ 0.75 and poor agreement: Kappa < 0.4), agreement in the baseline estimation is excellent for different datasets. However, Kappa values for datasets from China (i.e., GMU_DB, JNU_DB and SMU_DB) are > 0.98, whereas those for LCU_DB and UHB_DB are 0.885 and 0.771, respectively. For 1267 FHR recordings, the Kappa value and the correlation coefficient between manual measurement and CTGNet calculation are 0.873 and 0.969, respectively.

Figure S1(A-E) are Bland-Altman plots demonstrating the interchangeability of the clinical measurement and the CTGNet for the baseline estimation. The mean differences are < 1. Limits of agreement for GMU_DB, JNU_DB, SMU_DB, LUC_DB and UHB_DB, respectively, are −0.82/0.93, -1.34/0.93, -0.82/0.93, -13.10/14.52 and −9.05/9.90. Maximum differences for GMU_DB, JNU_DB, SMU_DB, LUC_DB and UHB_DB, are 2.81 bpm, 3.43 bpm, 16.52 bpm, 48.88 bpm and 18.86 bpm, respectively.

These most divergent examples in each database are shown in Fig. S2(A-F). It can be observed that the difference is large when the signal loss rate increases. Anyway, on the whole dataset including 1267 FHR recordings, the mean difference, deviation and limits of agreement are 0.36, 3.64 and −6.81/7.48, respectively.

### 2.3 Discussion

Reliable FHR interpretation is the base of the fetal state assessment. The poor recognition performance of FHR patterns can propagate the error to subsequent steps, thereby decreasing classification accuracy. In all these FHR patterns, the baseline is a precondition for evaluating of the other patterns. Visual estimation of the FHR baseline is subject to inter-and intra-observer variability. Computer-assisted baseline estimation has been proposed as a promising way to reduce this variability. In order to evaluate the performance of our computer-assisted method (i.e., CTGNet), a comparison of FHR baseline estimation by the CTGNet and a consensus of clinicians presents in this study.
Several studies have evaluated the performance of different computer-assisted methods. For example, in 2016, Jezewski et al. evaluated 11 different algorithms using two inconsistency coefficients based on three properties (i.e., number, location and area) of accelerations/accelerations[33]. They found that the algorithm of Arduini et al. [9, 34] outperforms other methods by achieving the lowest mean inconsistency coefficients on a private dataset with 41 FHR signals. This nonlinear filtering method is similar to the algorithm proposed by Mantel et al. [19]. The difference is that Arduini’s baseline is computed in 10-mins windows with 5-mins shift, whereas Mantel’s baseline is calculated for the whole FHR tracing. Considering Mantel’s method, Houzé de l’Aulnoit et al. further evaluated 11 newer algorithms by comparing the computed baselines with that estimated by clinicians on a dataset with 90 FHR signals [13]. This study found that Lu and Wei’s algorithm [14] achieves better results than other methods by achieving a new morphological analysis discriminant index (MADI) of 7.3%. Recently, a weighted median filter was proposed by Boudet et al. to compute the FHR baseline, and more agreement (represented by a MADI of 4.0%) with clinicians’ consensus than Lu and Wei’s method was shown on this dataset with 90 FHR recordings [15]. Similar to Lu and Wei’s method, an algorithm for the baseline estimation based on singular spectrum analysis and empirical mode decomposition was also proposed by Lu et al. [16] and evaluated on another public dataset with 552 FHR recordings. This method also was objectively evaluated on the dataset with 90 FHR recordings [13] by achieving a MADI of 15.6%. Unlike signal processing methods, the CTGNet based on deep learning was proposed in our previous study [18] and evaluated on a larger dataset with 234 FHR recordings. This method was compared with 12 signal processing methods and the lowest metrics (including the root-mean-squared difference between baselines and MADI) were obtained. These methodological studies illustrate the excellent performance of the CTGNet. However, its clinical application still requires a comparative study with large-scale multicenter data.

Baselines assigned by computer-assisted methods had been compared with those estimated by clinicians in several studies [7–12]. In the studies of Arduini et al. and Ayres-de-Campo et al., limits of agreement (LoA) were no more than −6.45 and 7.07 in ≤150 FHR tracings. In addition, Kappa value and ICC coefficient also were used to evaluate agreement in baseline determination between a computer-assisted method and several experts in previous studies. Obtained Kappa values varied from 0.18 to 0.97, while the ICC coefficient was within the range of 0.83–0.98 (Table S1). All these results were obtained on several small datasets with ≤150 FHR tracings acquired with EFM from ≤3 different manufacturers. In the present study, a larger dataset with 1267 FHR tracings acquired with EFM from 5 different manufacturers was used to evaluate agreement in baseline determination between the CTGNet and clinicians. This dataset included ~50% high-quality tracings with a signal loss rate of <10% per 10 mins and ~50% low-quality recordings with a signal loss rate of <50% per 30 mins. Kappa values were >0.98 for these high-quality tracings from GMU_DB, JNU_DB and SMU_DB, while the Kappa value was 0.771 for the lowest-quality FHR recordings (n = 552) from UHB_DB. Regardless, an excellent agreement was obtained on the whole dataset (n = 1276). These results indicate possibilities for the clinical application of CTGNet in FHR baseline estimation.
The high loss rate of FHR tracings severely affected the performance of CTGNet. In the present study, the Kappa value was reduced from 1.0 for the FHR recordings with a low signal loss rate to 0.885 for tracings with a medium loss rate and 0.771 for the low-quality recordings with a high signal loss rate. Although high-quality recordings with a signal loss of $\leq 20\%$ are recommended by the FIGO guidelines to assess the FHR patterns\cite{5}, low-quality tracings with a mean signal loss of 28–55\% are shown in clinical practice\cite{35, 36}. For example, the mean signal loss of 13\% and 30\% were found during the first and the second stage of labor, respectively\cite{37}. Therefore, the CTGNet can be further trained on datasets with low-quality tracings to improve its robustness.

2.4 Conclusions

The CTGNet for the FHR baseline estimation provided an excellent agreement with clinicians. However, this occurs in FHR recordings with low and medium signal loss rates. In the future, the CTGNet can be further improved by training it with more low-quality tracings.

3. Limitations

In the study, these baselines were evaluated by two clinicians from the same hospital, and the evaluations should be conducted by more clinicians from several centers in different countries in the future.

Declarations

4.1 Acknowledgements

The authors would like to thank Mujun Liu, Hao Yi, and Xue Kang for their assistance in calculating the baseline with the cardiotocograph network.

4.2 Authors’ contributions

Conceptualization, J.B., X.P., Y.L., Z.Z. and H.W.; writing-original draft preparation, J.B. and X.P.; writing-review and editing, J.B., Z.Z. and X.P.; visualization, J.B.; funding acquisition, H.W., J.B., and Y.L; Data collection, Z.Z., Y.L., H.W. and J.B; Data annotation, X.P. and Z.Z. All authors contributed to editorial changes in the manuscript. All authors read and approved the final manuscript.

4.3 Funding

This research was funded by the National Natural Science Foundation of China (61901192) (J.B.), the Science and Technology Program of Guangzhou (202201010544) (J.B.), Guangdong Provincial Key Laboratory of Traditional Chinese Medicine Informatization (2021B1212040007) (Y.L.) and National Key Research and Development Project (2019YFC0120100, and 2019YFC0121907) (H.W., J.B., and Y.L.).

4.4 Availability of data and materials
The data that support the findings of this study are available from the corresponding authors upon reasonable request.

4.5 Ethics approval and consent to participate

The Medical Ethics Committees of the Guangzhou Women and Children's Medical Center (273A01), the Jinan University (JNUKY-2022-018) and the NanFang Hospital of Southen Medical University (NFEC-2019-024) approved this retrospective study. The institutional review board approved the present study and allowed for a waiver of informed consent owing to the minimal risk posed to the patients. The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

4.6 Consent for publication

Not applicable.

4.7 Conflict of Interest

The authors declare no conflict of interest.

References


47. Jezewski J, Wrobel J. *Global baseline determination in fetal heart rate records with the identification of local prominent rates*. in Proc. of the 1st Int. Conf. on Medical Physics and Biomedical Engineering. 1994.


**Figures**

**Figure 1**

Flow chart of the FHR baseline estimation.

**Supplementary Files**

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterial.docx