Determination of Window Size for Baseline Estimation of Fetal Heart Rate Using CTG

Abstract—One of the major features of fetal heart rate (FHR) is its baseline, the accurate classification of which is of utmost importance as all the other parameters of cardiotocography (CTG) rely on it. Though guidelines for interpretation had been put forward by different organizations, they lack uniformity, hence, difficult to implement in automated systems. We aim to propose a standardized baseline estimation algorithm addressing. Baseline was calculated within continuous sliding window of duration 10 minutes and varying slide-lengths. The best slide-length is chosen and the corresponding baseline value is compared with the one obtained with discreet window as suggested by NICHD [1] and the estimate given by three experienced physicians. The result is analyzed using standard statistical methods such as confidence level, error margin, standard error etc. at a Confidence Level of 95%. The results of the algorithm and mathematical calculations incorporated in this work is found encouraging. Thus the proposed algorithm can not only be used to easily estimate the baseline using overlapping window but can also be adopted in an automated system for the interpretation of CTG.

Keywords—cardiotocography; fetal heart rate; baseline; confidence level; confidence interval; error margin; standard error.

I. INTRODUCTION

Medical decision making nowadays depend largely on the automated analysis of medical data. This considerably helps improve medical diagnosis and treatment. Thus, though the traditional mode of interpretation of CTG by visual analysis is not time consuming, the accuracy of it is largely dependent on the knowledge and experience of the clinician and many a time varies from clinician to clinician. An automated system for decision support may be able to eliminate some of the shortcomings of visual interpretation.

The features of FHR are Baseline, Variability, Acceleration and Deceleration. Each of these can be classified as Reassuring, Non-Reassuring and Abnormal. In order for proper interpretation of CTG, each of these features needs to be separately classified. Recognizing normal or grossly pathological CTG is easy, but problem arises in interpretation of suspicious or pathological CTG that has one or two abnormal features. Thus to identify such cases all four parameters need to be carefully evaluated [2].

Baseline is the most fundamental feature of FHR as all the other features are directly dependent on it. It is a common belief that robustness of the algorithm for the estimation and classification of FHR baseline is largely responsible for the efficiency of the automated fetal monitoring system. It is thought that if correctly interpreted, FHR evaluation may lead to a better prediction of fetal status than the measurement of fetal blood pH [3].

Last few decades had seen several attempts at automation of the CTG analysis and especially the estimation of baseline. These range from classical approaches that employ non-linear filtering of FHR signals to estimate baseline to intelligent approaches based on Artificial Neural Network. System 8000 was developed in 1991 for antenatal analysis based on the work of Dawes. This system was later upgraded to SonicAid FetalCare, which is used commercially for FHR analysis [4]. 2CTG2 was developed based on the work of Magenes [5]. NST-EXPERT, which was developed in 1995 by Alonso-Betanzos is an expert system that is capable of making a diagnosis and proposing a treatment [6]. CAFE is an extension of NST-EXPERT and was developed in 2002 by Guijarro-Berdinas and Alonso-Betanzos. It uses neuro-fuzzy approach to automate all the tasks associated with the analysis of CTG [7]. SisPorto was developed over a span of 14 years at the University of Porto, Portugal. It uses an expert system to estimate the individual aspects of CTG. The current version of SisPorto that is in use is Omniview SisPorto 3.5 [8]. So far only Sonicaid, which is based on traditional crisp method of CTG analysis, is used commercially worldwide, but the others failed to get any wide acceptance due to lack of specificity and poor practicality in every day use.

A. Cardiotocography

Fetal heart rate is controlled mainly by autonomic nervous system whose activity basically depends on the oxygen content of the fetal blood. Thus, analysis of FHR signals can predict asphyxia [9] which is the third most common cause of fetal death. The Cardiotocograph (CTG) is being used by the obstetricians since 1960s as a means for recording (graph) the heart beat (cardio) and the uterine contraction pressure (toco) of the mother, to evaluate the well being of the fetus. CTG is
non-invasive, cost-effective tool for checking the fetal status. Its induction to the clinical practice limited the occurrence of fetal problems leading to a decline in child morbidity and mortality [10].

Baseline is the resting level of fetal heart rate. Acceleration of FHR is a temporary increase of fetal heart rate 15 bpm above the baseline and lasting for 15 secs or longer. This occurs in response to fetal movements and signifies fetal central nervous system alertness and fetal well-being. The temporary decreases of FHR at least 15 bpm below the baseline and lasting for 15 secs or longer is called Deceleration. It usually suggests such hazardous events as compression of the umbilical cord.

The baseline of FHR is established by approximating the average FHR rounded to increments of 5 beats per minute (bpm) during a 10-minute observation. Periods of acceleration, deceleration, and marked fetal heart rate variability (> 25 bpm) are to be excluded to define the baseline. There must be at least 2 minutes of recognizable baseline segments which need not be contiguous in any 10-minute window, or the baseline for that period is undetermined [10].

Baselines can be classified as Reassuring(R): 110–160 bpm, Non-Reassuring (NR): 100 – 109 bpm or 161 – 180 bpm, Abnormal (Ab) : less than 100 bpm or more than 180 bpm

The 3-tier classification of CTG trace is as follows: A CTG trace is said to be Normal (N) if all the four features of FHR fall into reassuring category. Suspicious (S) if one of the features of FHR fall into non-reassuring category, while the others fall into reassuring category. Pathological (P) if two or more features fall into non-reassuring category or more than one feature fall into abnormal category [10].

B. Guidelines of FHR Interpretations

Even though CTG was in vogue from 1960s there was a lack of agreement regarding the definition, terminology and clinical interpretation of FHR patterns. The first standardized guidelines and definitions that could be quantitated, for the interpretation of FHR signal was proposed by FIGO in 1986 which were later modified by RCOG in 2001 and by NICE in 2007 [1]. NICHD took up this issue in 1997 and provided its own set of guidelines [1]. The guidelines were further reviewed in 2008 by NICHD and ACOG and a set of standardized guidelines was established. Though the definitions were mainly for visual analysis, it was also possible to adapt these definitions for computerized interpretation [1]. NICHD at this time introduced its 3-tier classification system which we followed in this paper. Despite the various guidelines the interpretation of CTG is still biased and suffers from high inter and intra observer variation. Even though these guidelines are thorough, they lack the definition to identify the transition from “normal” to “suspicious” or “suspicious” to “pathological” [Schifrin]. Also, the comparison between the guidelines showed that they are consistent for “normal” patterns but exhibited wide variations in “suspicious” and “pathological” patterns [11].

Among all the guidelines provided by different organizations, the ones provided by the NICHD is considered most standard one as it is more detailed and takes into consideration different factors to determine the baseline value. We thus checked whether our proposed method matches with the guideline of NICHD.

II. CHALLENGES IN CTG INTERPRETATION

The accurate interpretations of CTG are impeded by several factors; FHR waveform has a complex shape which doesn’t render well to visual interpretation. Though FHR waveform is a source of a lot of information, only a small part of it can be extracted by visual analysis. To analyze a CTG, each of its features is identified and then classified using rules derived from the guidelines of FIGO or NICHD. But since these guidelines are solely based on experimental observations, they lack precision, giving rise to difference in opinion among the medical practitioners [12]. Also, great care needs to be taken to interpret the data at the boundaries. These factors may lead to misdiagnosis which can cause lifetime affliction such as cerebral palsy that occurs due to hypoxia, or even death. Many of these incidences can be prevented by precise identification of the abnormal FHR pattern [13]. Enquiry conducted in UK, between 1997 and 2000, found out that many stillbirths and deaths in infancy occurred due to the poor interpretation of CTG [14].

III. METHODOLOGY

A. Baseline Estimation

For our work we have used CTU-UHB [15] database of Czech Technical University, Department of Cybernetics. We have considered the first 30 minutes of the 90 minutes trace as rest of the traces in most cases were not of adequate quality to perform an assessment. We used 25 samples to test our algorithm.

The algorithm is based on iterative calculation of real baseline based on initial virtual baseline. Though the proposed method is based on the initial virtual baseline, its effect on the signal will be annulled. This is due to the iterative calculation of the modified baseline which eventually approaches the original baseline of the FHR signal. Iterative approach removes accelerations and decelerations. This algorithm is applied within 10 mins. window. If within each window identifiable baseline is for more than 2 minutes, not necessarily contiguous, then that value is accepted, otherwise the baseline value is rejected. The window is slid by various time intervals ranging from 1 min. – 10 mins. The algorithm is depicted using a flowchart shown in Fig 1.

$B_{\text{vir}}$ is the virtual baseline and $B_{\text{mod}}$ is the modified baseline. $B_{\text{mod}}$ is obtained by computing the weighted average of the $B_{\text{mod}}$ values. Once $B_{\text{mod}}$ is calculated for each 10 minutes interval, the average baseline $B_{\text{avg}}$, which is the baseline B for that particular sample, is obtained.

The standard deviation $\sigma$ is estimated by calculating the difference between the consecutive signal values within each window. This process is repeated with different slide lengths. Average value of standard deviation $\sigma_{\text{avg}}$ for each slide length j is estimated by taking the average standard deviation values of all the samples:
\[ \sigma_{avg-j} = \frac{1}{N} \sum_{j=1}^{N} \sigma_i \]  

where, \( i = 1, \ldots, N \), the N samples under consideration and \( j = 1, \ldots, 10 \), the 10 slide-lengths from 1min – 10 mins. \( \sigma_i \) is the standard deviation value for each sample.

The minimum average standard deviation is calculated as follows:

\[ \sigma_{avg-min} = \min(\{\sigma_{avg-j}\}) \]  

The B value of the slide-length with the minimum average standard deviation is compared with the visual estimate provided by three expert obstetricians using different statistical methods. We have also compared this baseline value with the value obtained for slide length of 10 minutes. i.e. the discreet window as suggested by the NICHD.

B. Statistical Analysis

Along with a visual analysis by obstetricians, baseline estimation by the proposed method were compared using mean, standard deviation, error margin, confidence interval at a Confidence Level of 95%. A comparison between our result and the estimates given by the obstetricians is shown in Table I. Statistical estimations like mean, standard deviation, error margin, confidence interval (CI) are used to compare the estimated values of baseline for the set of data and the observed values of baseline by the physician for the same set of data. Calculations are done according to the following equations.

Mean :

\[ \bar{x} = \frac{\sum x_i}{n} \]  

Standard Deviation:

\[ \sigma = \sqrt{\frac{1}{n-1} \sum (x_i - \mu)^2} \]  

Margin of Error:

\[ m = 1.96 \times \frac{\sigma}{\sqrt{n}} \]  

where, the value 1.96 is called the confidence coefficient.

Confidence Interval :

\[ CI = \bar{x} \pm m \]  

Standard Error:

\[ SE = \frac{\sigma}{\sqrt{n}} \]  

The width of the CI gives an idea about the certainty of a given parameter. A wide interval means more data should be used in order to say anything definite about the parameter.

Error margin (m) is the amount of random sampling error in the result of a survey. If the margin of error is large then one should have less confidence in the result. Increase in sample size reduces the margin of error.

Standard error (SE) is an estimate of standard deviation, derived from a given sample. SE is smaller for a large sample.

The lowest average value of standard deviation is obtained for the window size 10 minutes and slide-length of 3 minutes. Using statistical methods we then compared this estimate with the estimate using 10 minutes discreet window (slide-length of 10 minutes) and average of the estimates provided by the physicians as shown in Table II.

IV. RESULTS AND DISCUSSION

In contrast to the recommendations provided by NICHD which takes the discrete window into consideration, this algorithm assumed the continuous window. We believe this is more logical and can detect any anomaly in baseline at the earliest, leading to early diagnosis of fetal distress which is justified by our experimental evaluation.

From Table II it can be seen that the statistical measurements for discreet window are only slightly better than the continuous window, in the range of around 7%. Standard error, error margin and CI are also within acceptable limit, indicating the accuracy of the proposed method.

Though in both cases all these parameters are slightly higher than the visual interpretation we can conclude that the proposed method of the use of overlapping window can be used for the automated estimation of baseline of FHR.

But in both cases all these values are slightly higher than the visual interpretation.

<table>
<thead>
<tr>
<th>FHR</th>
<th>Doctor1</th>
<th>Doctor2</th>
<th>Doctor3</th>
<th>Average</th>
<th>Result (3 mins interval)</th>
<th>Result with discreet window</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>142</td>
<td>144</td>
<td>144</td>
<td>143.3333</td>
<td>143.34</td>
<td>143.67</td>
</tr>
<tr>
<td>2</td>
<td>145</td>
<td>157</td>
<td>144</td>
<td>148.6667</td>
<td>153.82</td>
<td>152.28</td>
</tr>
<tr>
<td>3</td>
<td>150</td>
<td>150</td>
<td>154</td>
<td>151.3333</td>
<td>151.52</td>
<td>147.81</td>
</tr>
<tr>
<td>4</td>
<td>140</td>
<td>140</td>
<td>140</td>
<td>140</td>
<td>130.47</td>
<td>136.73</td>
</tr>
<tr>
<td>5</td>
<td>135</td>
<td>136</td>
<td>138</td>
<td>136.3333</td>
<td>132.99</td>
<td>133.59</td>
</tr>
<tr>
<td>6</td>
<td>160</td>
<td>162</td>
<td>160</td>
<td>160.6667</td>
<td>160.21</td>
<td>158.91</td>
</tr>
<tr>
<td>7</td>
<td>145</td>
<td>145</td>
<td>148</td>
<td>146</td>
<td>148.42</td>
<td>145.8</td>
</tr>
<tr>
<td>8</td>
<td>152</td>
<td>155</td>
<td>150</td>
<td>152.3333</td>
<td>151.25</td>
<td>151.28</td>
</tr>
<tr>
<td>9</td>
<td>138</td>
<td>142</td>
<td>138</td>
<td>139.3333</td>
<td>133.39</td>
<td>134.32</td>
</tr>
<tr>
<td>10</td>
<td>150</td>
<td>154</td>
<td>148</td>
<td>150.6667</td>
<td>146.66</td>
<td>146.61</td>
</tr>
<tr>
<td>TABLE II. THE COMPARISON USING STATISTICAL METHOD BETWEEN OBSERVED AND ESTIMATED VALUES OF BASELINE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>----------------</td>
<td>----------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Observed</td>
<td>Estimated (with 3 minutes overlap)</td>
<td>Estimated (with discreet 10 minutes window)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>144.89</td>
<td>143.34</td>
<td>142.39</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Std. dev.</td>
<td>10.63</td>
<td>11.83</td>
<td>10.99</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error Margin</td>
<td>4.39</td>
<td>4.88</td>
<td>4.54</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Std. Error</td>
<td>2.12</td>
<td>2.37</td>
<td>2.20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CI upper bound</td>
<td>149.28</td>
<td>148.22</td>
<td>146.93</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CI lower bound</td>
<td>140.51</td>
<td>138.46</td>
<td>137.86</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

V. CONCLUSION AND FUTURE SCOPE

A baseline calculation algorithm is proposed here within continuous sliding window of duration 10 minutes and varying slide-lengths. Even though the average standard deviation is the lowest for the 10 minutes discreet window, it can be concluded that overlapping window will provide a more clinically acceptable estimation of the baseline value. From Table I it can be seen that the estimates provided by the physicians tally with the estimates provided by our method.

NICHD guideline claims that baseline should be estimated within 10 minutes discreet window. It can thus be safely concluded that our method conforms well to NICHD guideline for visual interpretation. This method can also be used for automated estimation of baseline.

Width of the Confidence Interval suggests that more data needs to be included in the sample. In our future works we aim to use not only more data for parameter estimation, but also include the opinion of more physicians of varied level of experience.

Fig. 1. Flowchart for the estimation of baseline in 10 mins window with a slide of t mins

Start
  ↓
Calculate the initial virtual baseline $B_{vir}$ by averaging the FHR values
  ↓
FHR → $FHR_{i}$
  ↓
Calculate acceleration and deceleration and remove them
  ↓
Calculate modified baseline $B_{mod}$ by averaging FHR,
  ↓
$D = |B_{vir} - B_{mod}|$
  ↓
No
  ↓
$D > \varepsilon$ ?
  ↓

Yes
  ↓
$B_{vir} \leftarrow B_{mod}$
  ↓
Slide the window by t mins
  ↓
Duration of $B_{mod} ~ > 2$ mins?
  ↓

No
  ↓
Yes
  ↓
Calculate $B_{mod}'$ using weighted avg.
  ↓
Calculate $B_{avg}$ the avg. baseline
  ↓
Baseline $B$
  ↓
Reject the $B_{mod}$ value
  ↓
Length of the trace is reached?
  ↓

No
  ↓
Yes
  ↓
Calculate $B_{mod}$ using weighted avg.
  ↓
Calculate $B_{mod}$ the avg. baseline
  ↓
Baseline $B$
  ↓
End
REFERENCES


