2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 979-8-3503-2447-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/EMBC40787.2023.10340095

Timely Detection of Infants at Risk of Intrapartum Acidosis and Hypoxic-Ischemic Encephalopathy Using Cardiotocography*

Johann Vargas-Calixto, *Student Member*, *IEEE*, Yvonne Wu, Michael Kuzniewicz, Marie-Coralie Cornet, Heather Forquer, Lawrence Gerstley, Emily Hamilton, Philip A. Warrick, *Member*, *IEEE*, and Robert E. Kearney, *Life Fellow*, *IEEE*

***Abstract*— This work aims to improve the intrapartum detection of fetuses with an increased risk of developing fetal acidosis or hypoxic-ischemic encephalopathy (HIE) using fetal heart rate (FHR) and uterine pressure (UP) signals. Our study population comprised 40,831 term births divided into 3 classes based on umbilical cord or early neonatal blood gas assessments: 374 with verified HIE, 3,047 with acidosis but no encephalopathy and 37,410 healthy babies with normal gases. We developed an intervention recommendation system based on a random forest classifier. The classifier was trained using classical and novel features extracted electronically from 20-minute epochs of FHR and UP. Then, using the predictions of the classifier on each epoch, we designed a decision rule to determine when to recommended intervention. Compared to the Caesarean rates in each study group, our system identified an additional 5.68% of babies who developed HIE (54.55% vs 60.23%, $p < 0.01$) with a specific alert threshold. Importantly, about 75% of these recommendations were made more than 200 minutes before birth. In the acidosis group, the system identified an additional 17.44% (37.15% vs 54.59%, $p < 0.01$) and about 2/3 of these recommendations were made more than 200 minutes before birth. Compared to the Caesarean rate in the healthy group, the associated false positive rate was increased by 1.07% (38.80% vs 39.87%, $p<0.01$).**

***Clinical Relevance*— This method recommended intervention in more babies affected by acidosis or HIE, than the intervention rate observed in practice and most often did so 200 minutes before delivery. This was early enough to expect that interventions would have clinical benefit and reduce the rate of HIE. Given the high burden associated with HIE, this would justify the marginal increase in the normal Cesarean rate.**

I. INTRODUCTION

Neonatal hypoxic-ischemic encephalopathy (HIE) is a serious brain dysfunction caused by hypoxia during labor. It occurs in about 1.3 - 1.7 cases per 1,000 births in developed countries [1]. However, in low- and middle-income countries (LMIC) the incidence is much higher ranging from 6.7 to 26.5 per 1,000 births [2]. HIE has catastrophic consequences in LMIC, where 12 – 31% of affected infants die during the neonatal period [2].

During labor, fetal mechanisms control the fetal heart rate (FHR) to assure proper delivery of blood and oxygen to the brain in periods of hypoxia. The failure of these mechanisms can lead to fetal metabolic acidosis and HIE. As such, clinicians use cardiotocography (CTG) during labor to monitor the FHR and uterine pressure (UP) and to infer fetal condition. Through visual assessment, clinicians categorize the signals in three levels of abnormality according to clinical guidelines [3]. These levels reflect low, indeterminate, or high risk of developing acidosis and HIE. Unfortunately, about 80% of tracings are assigned to the indeterminate category and few tracings from babies with HIE reach the high-risk category. Moreover, inter-clinician agreement on tracing assessment is low [4]. When a fetus is identified as being at high risk of developing HIE, clinicians perform an emergency Caesarean delivery (CD) in the attempt to prevent progression to injury. Thus, it is crucial to identify those fetuses at substantial risk of HIE, and to do so as early as possible.

Computerized FHR features have been shown to be associated with an increased risk of acidosis and HIE [5, 6]. Automated tracing analysis provides more objective feature detection than visual interpretation of FHR and UP. Furthermore, digital signal processing facilitates finding new features or combinations and trends over time that can be used by machine learning and deep learning techniques to predict the risks of acidosis and HIE.

Digital records have been in use for many years so that adequate data can be collected to study rare events such as HIE. In this study we have applied machine learning techniques to develop a method for the intrapartum detection of fetuses with an increased risk of developing fetal acidosis or HIE. Our method used data from the electronic medical record and the digital CTG recording from births in a group of hospitals from the United States.

II. CLINICAL DATA

CTG signals and clinical data were acquired in 15 Kaiser Permanente Northern California hospitals between 2011 and 2019. The resulting dataset includes data from 246,968 singleton births with a gestational age of at least 35 weeks and no congenital defects. The study was performed on de-identified data and was approved by the Research Ethics Board of Kaiser Permanente and McGill University. We limited this study to a subgroup of 40,831 infants that had cord blood gas or neonatal blood gas available within the first 120

* Research supported by the Bill & Melinda Gates Foundation and the National Institutes of Health.

J. Vargas-Calixto and R. E. Kearney are with the Department of Biomedical Engineering, McGill University, Montreal, QC H3A 2B4, Canada. (e-mail: carlos.vargascalixto@mail.mcgill.ca).

Y. Wu and M. Cornet are with the University of California, San Francisco, CA 94158, USA.

M. Kuzniewicz, H. Forquer, and L. Gerstley are with Kaiser Permanente, Oakland, CA 94612, USA.

E. Hamilton and P. Warrick are with McGill University, Montreal, QC H3A 2B4, Canada, and with PeriGen Inc., Montreal, QC H4Z1E8, Canada.

minutes after birth. We used the blood gas results to define three groups:

a) A healthy group, comprising 37,410 infants with blood pH > 7, base deficit < 10 mmol/L, and no admission to the neonatal intensive care unit (NICU),

b) An acidosis group, comprising 3,047 infants with pH < 7 or base deficit > 10 mmol/L and no signs of encephalopathy. Fetal acidosis is a transitory state that could eventually lead to HIE if the conditions are maintained for long periods without time for recovery.

c) An HIE group, comprising 374 infants with acidosis and clinical evidence of encephalopathy present at 1 to 6 hours of age, accompanied by electrographic seizures or therapeutic hypothermia.

FHR was sampled at 4 Hz; UP was sampled at 1 Hz and upsampled to 4 Hz. No attempt was made to account for differences in medications administered during labor.

III. Methods

A. Preprocessing

The FHR signals were divided into nonoverlapping 20-minute-long epochs. We will refer to Epoch-N as the epoch starting N minutes before delivery and ending N+20 minutes before delivery. We processed the FHR using PeriCALM Patterns, a software system from PeriGen Inc. that identified classical FHR and UP patterns. It first identified gaps and periods where the signal was uninterpretable due to noise. Gaps shorter than 60 samples (15 s) were filled by linear interpolation. Then, the signals were preprocessed using low-pass, high-pass, median and FHR Karhunen-Loève filters. A long short-term memory (LSTM) network then identified FHR patterns and logistic regression was used to identify UP contractions [7, 8]. We only considered for analysis those epochs that had at least 80% valid samples.

B. FHR Patterns and Features

The FHR patterns identified by PeriCALM Patterns were:

a) Baseline (BAS): relatively flat segments of the FHR with a typical range between 110 – 160 bpm, and a typical peak-to-peak variability between 5 – 15 bpm.

b) Acceleration (ACC): segments with an increase in the FHR of more than 15 bpm and lasting for more than 15 seconds before returning to baseline.

c) Deceleration (DEC): segments, with a decrease in the FHR, usually more than 15 bpm that lasted for more than 15 seconds before returning to baseline.

PeriCALM Patterns also identified the location and duration of uterine contractions (CON). Once the FHR was preprocessed, we extracted relevant features from the FHR signals for classification. These features were estimated separately for each available epoch for each subject.

1) Features of Fetal Heart Rate Variability

Features that have been proposed in the literature to reflect fetal status during the intrapartum were included in our classification approach [6]. For each epoch, features were estimated twice: once for the whole epoch, and once for only the baseline periods within that epoch.

a) Frequency domain features: were estimated from the power spectral density (PSD), computed using the Lomb-Scargle periodogram to handle gaps [9]. We estimated the power in three bands: low frequency (LF, 30 – 150 mHz); movement frequency (MF, 150 – 500 mHz); and high frequency (HF, 0.5 – 1 Hz) bands [10]. The band ratio LF/(MF+HF) was also computed.

b) Nonlinear and nonparametric features: included the approximate entropy (ApEn) which quantifies signal irregularity and has been shown to reflect fetal compromise [10]. We also estimated features using phase rectified signal averaging (PRSA) including the deceleration capacity (DC), acceleration capacity (AC), and the deceleration reserve (DR) [11].

2) Features of Center and Variability

We estimated the mean (μFHR) and standard deviation (σFHR) for the whole epoch and the BAS within each epoch.

3) Features Associated with FHR Patterns

This group of features were estimated from the FHR patterns identified by PeriCALM Patterns.

a) Missing beats (MB): were estimated from the area of the DEC in the FHR. The expected number of beats in an epoch was estimated from baseline sections. However, during DEC, the FHR decreases. Thus, the MB was equal to the expected number of beats in the epoch minus the actual number of beats in that epoch.

b) Extra beats (EB): were the number beats in excess of the expected number, caused by ACC events.

c) Number of decelerations and contractions prior to the beginning of the epoch (DEC60 and CON60): The number of DEC and CON in the 60 minutes prior to the beginning of each epoch. These values were divided by six and rounded to obtain the average CON and DEC rate in a 10-minute period. This led to CON60 having values between 0 and 5, and DEC60 between 0 and 4.

C. Classification Framework

Our system had two steps. First, we classified each epoch independently using FHR features. Then, we aggregated the predictions of the epoch classifier to recommend intervention for each infant.

1) Classification of FHR epochs

This step performed binary classification using a random-forest (RF) classifier. The pathological class contained the infants from the acidosis and HIE groups, whereas the healthy class comprised infants from the healthy group.

The RF classifier used an ensemble of decision trees trained on different subsets of the data. To reduce the effect of class imbalance, we subsampled the majority class – healthy – such that each tree was trained on a balanced subset of infants. The RF classifier was trained on the FHR features of all epochs up to nine hours before delivery.

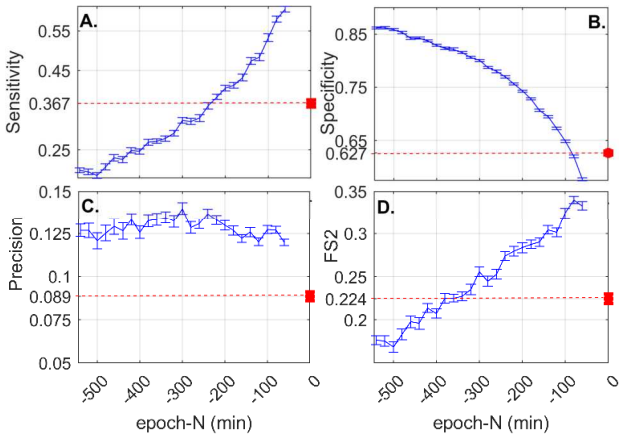

Figure 1. Performance metrics for the random-forest classifier (blue). The equivalent observed intervention rates are indicated in red. These metrics are the (A) sensitivity, the (B) specificity, the (C) precision, and the (D) F_2 score for the predictions of the classifiers on the test set. The lines show the median of these estimates, and the error bars the 95% confidence interval of the median. Notice that these metrics were estimated for each epoch and only reflect those snapshots in time.

After training, we used the out-of-bag (OOB) predictions for validation, and the test set predictions for performance evaluation. We only tested Epoch-60 and earlier. It takes about 30 minutes from the moment that the decision is made to the end of a Caesarean delivery (CD). Thus, any decision made at Epoch-40 or Epoch-20 would not result in interventions that make any clinical difference.

The RF classifier provided the posterior probability that an epoch belonged to the pathological class. We selected a classification threshold that gave the maximum sensitivity, or true positive rate, while keeping the false positive rate below the Caesarian rate of the healthy class.

2) Recommending Intervention

This step analyzed the cumulative predictions on the epochs of each infant to generate a recommendation. Each epoch predicted to be pathological was considered an alert. For each infant, we counted the number of alerts as a function of the time before delivery. Using the predictions on the validation set, we varied the number of alerts to compute a receiver-operating characteristics (ROC) curve. We used this curve to define the decision rule that defines the minimum number of alerts required for the system to recommend intervention. Thus, when an infant had more or equal alerts than the defined threshold, the system recommended intervention. Using this decision rule, we compared the recommended intervention rates for each group and as a function of the time before delivery in the test sets.

D. Performance Metrics

We randomly divided the subjects into training and testing sets. A random sample of 90% of the individuals was used for training and the remaining 10% for testing. This random sampling was repeated 100 times to obtain a distribution of performance metrics.

The predictions of the epoch-based classifier were assessed using the following metrics:

a) Sensitivity: the true positive rate of the predictions.

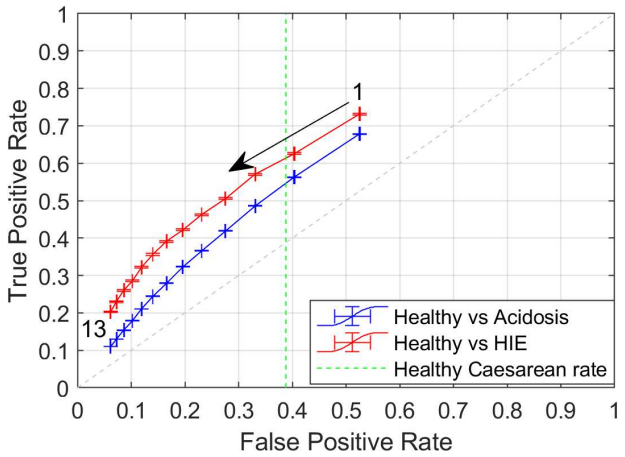

Figure 2. Receiver-operating characteristic curve when the minimum number of alerts varies from 1 to 13 (as indicated in the plot). This plot was generated using the out-of-bag predictions of the random-forest classifier. Curves were plotted for the healthy vs acidosis (blue) and healthy vs HIE (red) comparisons. The identity line was plotted as a dashed grey line and the healthy Caesarean delivery rate was plotted as a dashed green line.

$$Sensitivity = \frac{TP}{TP+FN}, \quad (1)$$

where TP is the number of correctly identified pathological epochs, and FN is the number of pathological epochs predicted to be healthy.

b) Specificity: the true negative rate of the predictions.

$$Specificity = \frac{TN}{TN+FP}, \quad (2)$$

where TN is the number of correctly identified healthy epochs, and FP is the number of healthy epochs predicted to be pathological.

c) Precision: the positive predictive value of the predictions.

$$Precision = \frac{TP}{TP+FP}. \quad (3)$$

d) F_2 score: a version of the F-score that weighs the sensitivity higher than the precision.

$$F_2 = 5 * \frac{TP}{5*TP+4*FN+FP}. \quad (4)$$

IV. Results

A. Classification of FHR Epochs

Fig. 1 shows the performance of the RF classifier as a function of Epoch-N. Fig. 1.A shows that the sensitivity increased as the time of delivery approached and became higher than the pathological intervention rate at 220 min before birth. Fig. 1.B shows that the specificity was initially high and decreased as time of delivery approached but was better than the healthy CD rate until Epoch-80. Thus, the sensitivity increased, and specificity decreased as birth approached. Fig. 1.C shows that the classifier precision was always higher than the clinical intervention rate. Finally, Fig. 1.D shows that the classifier had a better F_2 score than the intervention rate during the last 300 min of labor.

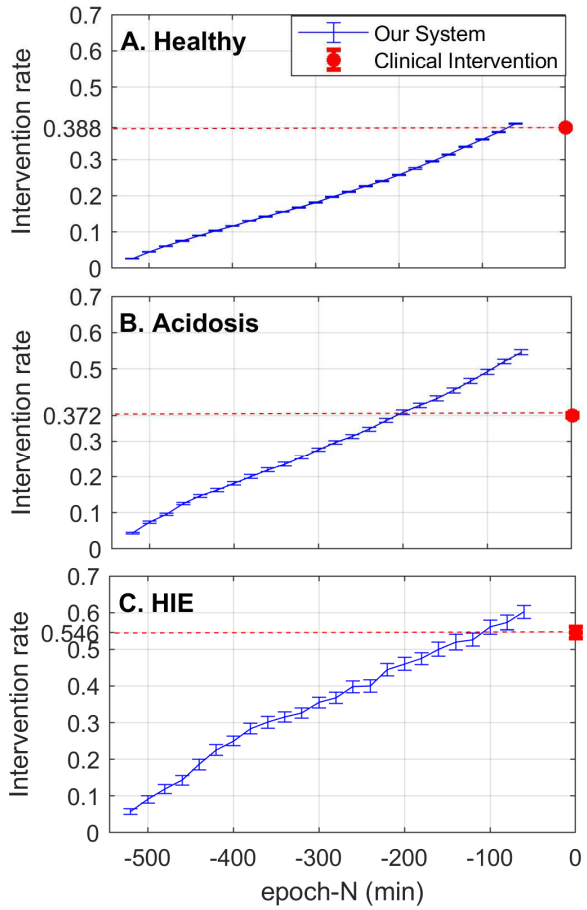

Figure 3. Rate of intervention recommendation based on the cumulative predictions in the (A) healthy, (B) acidosis, and (C) HIE groups. The current intervention rate for each group (red) is shown for comparison. The plot shows the median and its 95% confidence interval.

B. *Recommending Intervention*

Fig. 2 shows the ROC curve of the intervention recommendations as a function of the minimum number of alerts needed to recommend intervention. As the criteria became stricter (i.e. more alerts were needed), the false positive rate and the true positive rate decreased. We selected a threshold of two alerts since this yielded a false positive rate close to the healthy CD rate while keeping a high sensitivity.

Fig. 3 shows that the rate of recommendation with this threshold increased monotonically with time. For the healthy population the recommendation rate was lower than the CD rate until Epoch-60. For the acidosis and HIE populations, the recommendation rate surpassed the CD rate at Epoch-200 and Epoch-100 respectively.

Table I summarizes the information of Fig. 3 at two times: Epoch-200 and Epoch-60. When compared to the clinical intervention rates at Epoch-60, our system increased the detection of HIE by 5.68% (54.55% vs 60.23%, $p<0.01$) and acidosis by 17.44% (37.15% vs 54.59%, $p < 0.01$). The Caesarean rate in the healthy population increased by only 1.07% (38.80% vs 39.87%, $p<0.01$). Importantly, 3/4 of the recommendations in the HIE group, and 2/3 in the acidosis group were made by Epoch-200.

TABLE I. COMPARISON OF THE INTERVENTION RATES OBSERVED IN THE TEST SET VERSUS THOSE OBTAINED WITH OUR RECOMMENDATION SYSTEM AT EPOCH-200 AND EPOCH-60

Group	Intervention Rate	Recommendation System	
		Epoch-200	***Epoch-60***
Healthy	38.80% (38.66 – 38.94)	25.81% (25.66 – 25.97)	39.87% (39.70 – 40.04)
Acidosis	37.15% (36.58 – 37.72)	38.03% (37.35 – 38.71)	54.59% (53.93 – 55.25)
HIE	54.55% (53.06 – 56.03)	46.05% (44.30 – 47.80)	60.23% (58.47 – 61.99)

V. DISCUSSION

A. *Key Findings*

There are two critical performance objectives for a CTG-based decision-support system for HIE. First, the system must identify more infants at high risk than current standards, without causing too many unnecessary interventions. Second, the identification must occur as early as possible to allow enough time for a clinical intervention to plausibly prevent injury and development of HIE. Our proposed system did both: (1) it recommended intervention in more infants from the pathological groups than the clinical intervention rate, and (2) ~3/4 of the recommendations in the HIE group and ~2/3 in the acidosis group were made at least three hours before delivery. Furthermore, these improvements were made with a negligible increase in the healthy group CD rate. Thus, use of our system would result in early CD for HIE infants. This could help reduce the incidence and severity of HIE since it is reasonable to expect that earlier CD will improve outcome.

B. *Prediction of the Risk of Acidosis and HIE Using FHR*

Our system was trained on FHR features from 20-minute-long epochs but had no knowledge of the time before delivery. As such, it could be used during pregnancy when the time before delivery is unknown.

FHR epochs are not independent, and there will be many epochs for most infants. Thus, rather than using single epoch predictions, our system recommended intervention based on an alert threshold computed across epochs. This better accounted for FHR patterns associated with pathological outcomes that persisted over multiple epochs.

Other FHR classification approaches have reported lower false positive rates than our approach, with similar levels of sensitivity. However, these focused on separating subjects at the extremes of the pH and base deficit spectra. Thus, Warrick et. al defined their healthy class to have a base deficit < 8 mmol/L, while the pathological class had a base deficit > 12 mmol/L and evidence of death or HIE [12]. In comparison, Petrozziello et al. defined their healthy class as those with arterial cord pH ≥ 7.15, and the pathological class as those with arterial cord pH < 7.05 and presence of stillbirth, neonatal death, neonatal encephalopathy, intubation or cardiac massage and admission to the NICU [13]. As such, it is reasonable to expect more misclassifications with our system which included many more intermediate cases. Furthermore, these previous studies generated most of their alerts during the last 90 to 60 minutes before birth. In contrast,

our system has a sensitivity of ~46% for HIE and ~38% for acidosis three hours before delivery.

C. Strengths and Limitations

A major study strength was the use of a comprehensive dataset with automated extraction of clinical data from the electronic medical and CTG records from 15 hospitals where umbilical cord gases are obtained frequently. In addition, all HIE records, including the associated neuroimaging findings, were reviewed by a study pediatric neurologist to verify the accuracy of the diagnosis. All 40,831 CTG records were analyzed electronically, avoiding inconsistency and bias that is well documented to occur with visual analysis by clinicians.

One limitation was the requirement for subjects to have umbilical or very early neonatal blood gases and CTG recording. Clinicians generally perform a blood gas when concerned about fetal acid base status. Thus, the individuals in our healthy population might have had more risk factors or concerning fetal heart rate patterns compared to individuals where blood gases were not acquired. Patients with no CTG recordings and no blood gas measurements are usually those with few risk factors and short uncomplicated labors. Thus, these selection criteria do not diminish the significance of our findings; in contrast, they strengthen them since they make the discrimination task more challenging. The effect of selection requiring blood gas measurement will be assessed in the future by relaxing the inclusion criteria. That said, the requirement for blood gases does eliminate speculation that the "healthy" control group could have included babies with unrecognized acidosis.

D. Future Work

Future classifiers could also include clinical risk factors [5], probabilistic models of FHR and UP patterns [14], and models of the relationship between FHR and UP signals [12], all of which have shown to carry discriminatory information about the risk of developing HIE. Other alternatives to the work presented here include using the time from the beginning of labor instead of the number of decelerations and contractions to inform the classifier of the progression of labor. Also, instead of counting the number of alerts, we could look at the cumulative posterior probability of the predictions in consecutive epochs before recommending intervention.

VI. Conclusion

The methods described in this retrospective study recommended intervention in more babies affected by acidosis or HIE than the intervention rate observed in actual practice. Moreover, they did so early enough before delivery. This interval was long enough to expect that interventions would have clinical benefit. Earlier interventions in both outcome classes are meaningful because the clinical objective is to avoid injury and HIE and they are likely to contribute to reducing the rate of HIE. The increase in Caesarean deliveries in the healthy group was minimal making the risk/benefit ratio very advantageous.

References

[1] J. J. Kurinczuk, M. White-Koning, and N. Badawi, "Epidemiology of neonatal encephalopathy and hypoxic-ischaemic encephalopathy," *Early Hum Dev*, 2010, vol. 86, no. 6, pp. 329-38.

[2] J. E. Lawn, A. CC. Lee, M. Kinney, L. Sibley, W. A. Carlo, V. K. Paul, R. Pattinson, and G. L. Darmstadt, "Two million intrapartum-related stillbirths and neonatal deaths: where, why, and what can be done?," *Int J Gynaecol Obstet*, 2009, vol. 107 Suppl 1, pp. S5-19.

[3] G. A. Macones, G. D. V. Hankins, C. Y. Spong, J. Hauth, and T. Moore, "The 2008 National Institute of Child Health and Human Development Workshop Report on Electronic Fetal Monitoring: Update on Definitions, Interpretation, and Research Guidelines," *Journal of Obstetric*, Gynecologic & Neonatal Nursing, 2008, vol. 37, no. 5, pp. 510-515.

[4] C. M. Farquhar, S. Armstrong, V. Masson, J. M. D. Thompson, and L. Sadler, "Clinician Identification of Birth Asphyxia Using Intrapartum Cardiotocography Among Neonates With and Without Encephalopathy in New Zealand," *JAMA Network Open*, 2020, vol. 3, no. 2, pp. e1921363-e1921363.

[5] A. Georgieva, C. W. G. Redman, and A. T. Papageorghiou, "Computerized data-driven interpretation of the intrapartum cardiotocogram: a cohort study," *Acta Obstetricia et Gynecologica Scandinavica*, 2017, vol. 96, no. 7, pp. 883-891.

[6] J. Vargas-Calixto, Y. Wu, M. Kuzniewicz, M. C. Cornet, H. Forquer, L. Gerstley, E. Hamilton, P. Warrick, R. Kearney, "Temporal Evolution of Intrapartum Fetal Heart Rate Features," in *2021 Computing in Cardiology (CinC)*, 2021, pp. 1-4.

[7] P. Warrick, E. Hamilton, and M. Macieszczak, "Neural network based detection of fetal heart rate patterns," in *Proceedings 2005 IEEE International Joint Conference on Neural Networks*, 2005, vol. 4, pp. 2400-2405.

[8] P. A. Warrick and E. F. Hamilton, "Antenatal fetal heart rate acceleration detection," in *2016 Computing in Cardiology Conference (CinC)*, 2016, pp. 893-896.

[9] G. D. Clifford and L. Tarassenko, "Quantifying errors in spectral estimates of HRV due to beat replacement and resampling," *IEEE Trans Biomed Eng*, 2005, vol. 52, no. 4, pp. 630-8.

[10] M. G. Signorini, G. Magenes, S. Cerutti, and D. Arduini, "Linear and nonlinear parameters for the analysis of fetal heart rate signal from cardiotocographic recordings," *IEEE Trans Biomed Eng*, 2003, vol. 50, no. 3, pp. 365-74.

[11] M. W. Rivolta, T. Stampalija, M. G. Frasch, and R. Sassi, "Theoretical Value of Deceleration Capacity Points to Deceleration Reserve of Fetal Heart Rate," *IEEE Trans Biomed Eng*, 2020, vol. 67, no.4, pp. 1176-85.

[12] P. A. Warrick, E. F. Hamilton, D. Precup, and R. E. Kearney, "Classification of Normal and Hypoxic Fetuses From Systems Modeling of Intrapartum Cardiotocography," *IEEE Transactions on Biomedical Engineering*, 2010, vol. 57, no. 4, pp. 771-779.

[13] A. Petrozziello, C. W. G. Redman, A. T. Papageorghiou, I. Jordanov, and A. Georgieva, "Multimodal Convolutional Neural Networks to Detect Fetal Compromise During Labor and Delivery," *IEEE Access*, 2019, vol. 7, pp. 112026-112036.

[14] J. Vargas-Calixto, Y. Wu, M. Kuzniewicz, M. C. Cornet, H. Forquer, L. Gerstley, E. Hamilton, P. Warrick, R. Kearney, "Multi-Chain Semi-Markov Analysis of Intrapartum Cardiotocography," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2022, pp. 1948-1952.