Assessing Intrapartum Risk of Hypoxic Ischemic Encephalopathy Using Fetal Heart Rate with Long Short-Term Memory Networks

Derek Kweku DEGBEDZUI[1], Michael Kuzniewicz[2], Cornet Marie-Coralie[3], Yvonne Wu[3]
Heather Forquer[2], Lawrence Gerstley[2], Emily Hamilton[4], Doina Precup[2]
Philip Warrick [2], Robert Kearney [1]

[1] McGill University, Montreal, Canada
[2] Kaiser Permanente, Northern California, USA
[3] University of California, San Francisco, USA
[4] PeriGen Inc., Montreal, Canada

Abstract

This study investigated the prediction of the risk of hypoxic ischemic encephalopathy using intrapartum cardiotocography (CTG) records with a long short-term memory recurrent neural network. Across the 12 hours of labour, HIE sensitivity rose from 0.25 to 0.56 as delivery approached while the specificity remained approximately constant. The results show that classification improves as delivery approaches but that performance needs improvement. Future work will address the limitations of this preliminary studies by investigating input signal transformations and the use of other network architectures to improve the model performance.

1. Introduction

Hypoxic ischemic encephalopathy (HIE) is a brain injury caused by the impaired delivery of oxygen to the brain[1]. The estimated incidence of neonatal HIE is around 1.5 per 1000 live births in developed countries and 10-20 per 1000 live births in low- and middle-income countries [1, 2].

Neonatal HIE, a syndrome of disturbed neurologic function in the earliest days of life, is characterized by difficulty with initiating and maintaining respiration, depression of tone and reflexes, sub-normal levels of consciousness, and often seizures [3]. Generally, over half of newborns diagnosed with neonatal HIE may die or develop major impairment (such as cerebral palsy, hearing loss, visual impairment, cognitive and behavioral problems) by age 3 years [4].

Clinicians monitor both fetal and maternal well-being during labour using cardiotocography (CTG) which measures fetal heart rate (FHR), uterine pressure (UP) and maternal heart rate (MHR). Clinicians use visual interpretation of these CTG signals to assess fetal condition and to determine when to intervene – by performing a Cesarian section (CS) – to prevent neonatal death or HIE [5]. Unfortunately, the visual assessment of CTG signals has significant inter- and intra-observer variability [5, 6]. The inaccurate interpretation of CTG signals has also contributed to an increased rate of CS and assisted deliveries [6, 7]. Furthermore, clinical guidelines give no specific management recommendations for the great majority of FHR signals (over 80%) that are categorized as “indeterminate” [8, 9].

Machine learning (ML) and deep learning (DL) methods may have the potential to improve the interpretation of CTG signals. The computation of engineered FHR features for use with classical ML methods is a difficult task that has yet to produce features sufficiently informative for FHR classification [10]. Some publications have reported approaches to fetal state detection using DL with EFM data [11–14]. However, the majority of these have used datasets that were too small to fully support the development of DL models. Additionally, these datasets did not include HIE cases. Thus, there is a need to conduct DL experiments with a large cohort of births. Using the 250,000 EFM records in our EFM database, the objective of our project is to investigate DL models for the early assessment of the risk of developing HIE using raw and transformed EFM records. This paper presents early results using FHR input and a long short-term memory (LSTM) to assess the intrapartum risk of HIE.

2. Method

This section briefly discusses the clinical data and the data preprocessing carried out. Next, it describes the classification experiments conducted using these data.

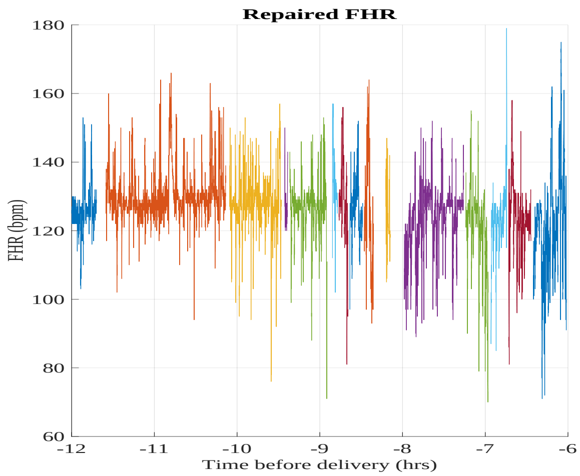

Figure 1. A typical repaired FHR signal. Each CTG signal, with a maximum length of 12 hours, was repaired using PeriCALM Patterns [15]. Each color corresponds to a different segment due to the presence of a gap. This displayed repaired FHR had 14 gaps of varying length within the 12^{th} and 6^{th} hour before delivery.

2.1. Clinical data

The clinical data comprised 12 hours of de-identified CTG signals obtained from 246,973 births at the 15 hospitals of Kaiser Permanente Northern California between 2010 and 2019. We included singleton live births, with gestational age $\geq$ 35 weeks, without congenital or chromosomal abnormalities and with electronic fetal monitoring. Analysis was limited to vaginal births since CS preempts observation of the labour time course that would have occurred without intervention.

HIE was defined as the presence of both acidosis and encephalopathy, where acidosis was defined as pH $<$ 7 or base deficit $\geq$ 10 mmol/L from the umbilical cord blood gas measurements shortly after birth. The healthy group comprised newborns who exhibited no encephalopathy or seizures, had Apgar at 5 minutes $\geq$ 7, had no chest compression or intubation and were discharged home alive.

The resulting data set comprised CTG signals from 173 HIE, 2003 Acidosis and 24620 healthy (without acidosis) cases. The number of HIE cases reflects its expected low incidence. This study focused on EFM signals from 145 HIE and 170 randomly under-sampled healthy fetuses.

2.2. Data preprocessing

We used PeriCALM Patterns, a proprietary software of PeriGen Inc. [15], to identify artifacts, remove noise and repair the FHR signals. Figure 1 shows a repaired FHR signal. Each colored segment of FHR is separated by a gap, which occurs when sensors detach during acquisition. The repaired signals were decimated from 4 Hz to 1 Hz to reduce training time. The factor of 4 was selected after determining that this retained 96.9% of the FHR power for both HIE and healthy cases. The decimated FHR retained the low frequency (LF, 30–150 mHz) and movement frequency (MF, 150–500 mHz) bands [16, 17]. Finally, we divided the recording into 20-minute non-overlapping segments, which generated 36 segments for the longest recordings of 12 hrs.

2.3. Classification

Train, test, and validation indexes were randomly permuted for each fold, generating 10 non-overlapping test sets for 10-fold cross validation. In each fold during training, the train and validation data for all 36 labour epochs were concatenated and used to train and validate a 3-layer LSTM model, with three hidden layers of 128, 256 and 128 cells. An early stopping value of 30 training epochs ensured that the trained model did not overfit the training dataset. For each fold, the LSTM model was trained 30 times and the model with the lowest validation loss retained. This model was used to independently evaluate the fold's test data for each labour epoch (i.e., we evaluated the classification performance for each of the 36 labour epochs separately) in terms of the average, across folds, of the sensitivity, specificity, and area under the receiver operating characteristic (AUROC) as functions of labour time.

3. Results

Figure 2 shows the AUROC, sensitivity, and specificity for the test data as functions of time. The AUROC rose steadily as delivery approached (from 0.51 to 0.67) as did HIE sensitivity (from 0.25 to 0.56); specificity remained approximately constant over the 12 hours with a mean of 0.71 and standard deviation of 0.04.

4. Discussion

The current performance shows some promising results but there is room to greatly improve the prediction performance. Our approach is limited by class imbalance mitigation (through random under-sampling) that did not tap the potential of our available data size. In future works, we will explore other class imbalance approaches such as cost sensitive learning.

The authors in [12] reported that convolutional neural networks (CNN) models outperformed LSTM models, hence we will also assess these DL structures. The use of a single-channel FHR input also results in a reliance on

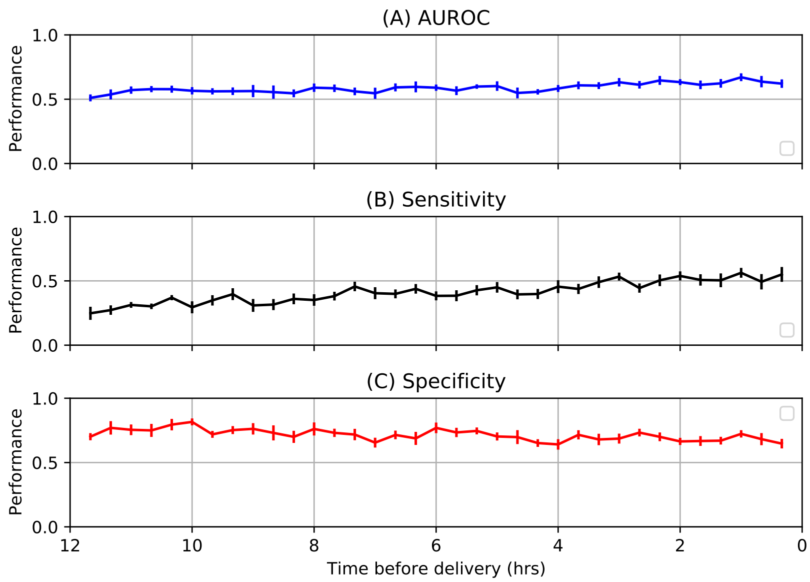

Figure 2. Classification performance metrics as a function of time before birth. (A) Area under the receiver operating characteristic (AUROC), (B) Sensitivity and (C) Specificity. The lines and error bars indicate the mean $\pm$ standard deviation over the 10-folds of cross-validation.

an LSTM-alone representation. We will explore the added benefits of adding MHR and UP inputs to the DL model, which may also be beneficial. Finally, we will assess transformation of the raw preprocessed EFM using methods such as the spectrogram and scalogram.

The class label of each 12-hour FHR is determined after the time of birth and our current reference point for analysis is the time of delivery. This is a major of a clinical problems as we have no objective measure during labour about when the fetal compromise commences. Additionally, our current approach of separately classifying each of the 36 20-minute non-overlapping labour epoch is not directly applicable in a clinical setting. Despite these limitations, our results demonstrate that as labour progresses, the model is better able to predict the fetal condition at birth. In future works, we will shift our analysis from time of delivery to labour onset and explore other methods to resolve the limitations of the study.

5. Conclusion

The aim of this project is to improve the early detection of intrapartum risk of developing HIE using a large cohort of births. This current study focused on single channel FHR input to train and test an LSTM network. The preliminary studies show some promising results, but the model performance needs improvement and the limitation of the study needs to be addressed. Future work will work on improving the model performance and addressing the current limitations by exploring transformation methods for the input signals and other DL network architectures.

Acknowledgments

The research is being supported by the Bill Melinda Gates Foundation and the National Institute of Health.

References

[1] Greco P, Nencini G, Piva I, Scioscia M, Volta C, Spadaro S, Neri M, Bonaccorsi G, Greco F, Cocco I, et al. Pathophysiology of hypoxic–ischemic encephalopathy: a review of the past and a view on the future. Acta Neurologica Belgica 2020;120(2):277–288.

[2] Kurinczuk JJ, White-Koning M, Badawi N. Epidemiology of neonatal encephalopathy and hypoxic–ischaemic encephalopathy. Early human development 2010;86(6):329–338.

[3] of Pediatrics AA, et al. Neonatal encephalopathy and neurologic outcome, second edition report of the american college of obstetricians and gynecologists' task force on neonatal encephalopathy. Pediatrics 2014;133(5):e1482–e1488.

[4] Adstamongkonkul D, Hess DC. Ischemic conditioning and neonatal hypoxic ischemic encephalopathy: a literature review. Conditioning medicine 2017;1(1):9.

[5] Chauhan SP, Klauser CK, Woodring TC, Sanderson M, Magann EF, Morrison JC. Intrapartum nonreassuring fetal heart rate tracing and prediction of adverse outcomes: in-

terobserver variability. American journal of obstetrics and gynecology 2008;199(6):623–e1.

[6] Ayres-de Campos D, Spong CY, Chandraharan E. Figo consensus guidelines on intrapartum fetal monitoring: Cardiotocography. International Journal of Gynecology Obstetrics 2015;131(1):13–24.

[7] Michaeli J, Srebnik N, Zilberstein Z, Rotem R, Bin-Nun A, Grisaru-Granovsky S. Intrapartum fetal monitoring and perinatal risk factors of neonatal hypoxic–ischemic encephalopathy. Archives of Gynecology and Obstetrics 2020;1–9.

[8] Clark SL, Nageotte MP, Garite TJ, Freeman RK, Miller DA, Simpson KR, Belfort MA, Dildy GA, Parer JT, Berkowitz RL, et al. Intrapartum management of category ii fetal heart rate tracings: towards standardization of care. American journal of obstetrics and gynecology 2013;209(2):89–97.

[9] Macones GA, Hankins GD, Spong CY, Hauth J, Moore T. 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; 37(5):510–515.

[10] Zhao Z, Zhang Y, Comert Z, Deng Y. Computer-aided diagnosis system of fetal hypoxia incorporating recurrence plot with convolutional neural network. Frontiers in physiology 2019;10:255.

[11] Warrick PA, Hamilton EF. Intrapartum fetal-state classification using long short-term memory neural networks. In 2017 Computing in Cardiology (CinC). 2017; 1–4.

[12] Petrozziello A, Jordanov I, Papageorghiou TA, Redman WC, Georgieva A. Deep learning for continuous electronic fetal monitoring in labor. In 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2018; 5866–5869.

[13] Cömert Z, Kocamaz AF. Fetal hypoxia detection based on deep convolutional neural network with transfer learning approach. In Computer Science On-line Conference. Springer, 2018; 239–248.

[14] Gao W, Lu Y. Fetal heart baseline extraction and classification based on deep learning. In 2019 International Conference on Information Technology and Computer Application (ITCA). IEEE, 2019; 211–216.

[15] Warrick P, Hamilton E, Macieszczak M. Neural network based detection of fetal heart rate patterns. In Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005., volume 4. IEEE, 2005; 2400–2405.

[16] Cerutti S, Civardi S, Bianchi A, Signorini MG, Ferrazzi E, Pardi G. Spectral analysis of antepartum heart rate variability. Clinical Physics and Physiological Measurement 1989; 10(4B):27.

[17] Warrick PA, Hamilton EF, Precup D, Kearney RE. Classification of normal and hypoxic fetuses from systems modeling of intrapartum cardiotocography. IEEE Transactions on Biomedical Engineering 2010;57(4):771–779.

Address for correspondence:

Derek Kweku DEGBEDZUI

Department of Biomedical Engineering, McGill University

derek.degbedzui@mail.mcgill.ca