

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

Derek Kweku DEGBEDZUI¹, Michael Kuzniewicz³, Cornet Marie-Coralie⁴, Yvonne Wu⁴, Heather Forquer³, Lawrence Gerstley³, Emily Hamilton^{1,2}, Doina Precup¹, Philip Warrick^{1,2}, Robert Kearney¹

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

Aims: Our goal is to improve the analysis of intrapartum electronic fetal monitor signals for the early detection of fetuses at risk of hypoxic-ischemic encephalopathy (HIE) during labor. This study examines the performance of a long short-term memory (LSTM) model with fetal heart rate (FHR) input to classify fetuses whose postnatal diagnosis was either healthy or HIE.

Methods: We selected de-identified records from 145 HIE and 170 healthy neonates born at hospitals of Kaiser Permanente Northern California between 2010 and 2019. Recordings >12 hours before delivery were excluded. FHR was downsampled from 4 Hz to 1 Hz and divided into 20-minute non-overlapping segments. Approximately 15% of the fetuses (26 HIE and 25 healthy cases) served as a hold-out test dataset, generating 507 HIE and 801 healthy randomly selected segments. The complement set of fetuses served as the train-validation dataset. We achieved class balance by randomly under sampling healthy cases, generating a total of 3659 HIE and 3631 healthy segments. We then trained 3-layer LSTM models using 5-fold cross-validation, retaining the model for each fold with the lowest loss over 5 trials. We evaluated performance with the independent hold-out dataset.

Results: Figure 1 shows test sensitivity, specificity, and area under the receiver operating characteristic (AUROC) as functions of time. The AUROC rose steadily as delivery approached (from 0.42 to 0.71) as did HIE sensitivity (from 0.21 to 0.49); specificity remained approximately constant over the 12 hours with a mean of 0.74 and standard deviation of 0.10.

Conclusion: This study demonstrates that better prediction occurs as labor progresses, consistent with clinical expectations. In future work, we will investigate preprocessing methods and other network architectures.

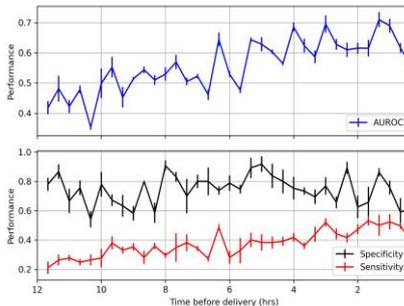


Figure 1: Test performance metrics over time. The lines and error bars indicate the mean \pm standard deviation over 5-folds of cross-validation.