Introduction of Artificial Intelligence in Cardiotocography Interpretation for the Clinical Use

A. R. Soltani, J. V. Zaen, C. Aguet, H. Sigurthorsdottir, A. Radan •, M. Lemay

Cardiotocography is among the most important medical surveillance methods, particularly for the assessment of the fetal state during labor. Yet, its conventional subjective interpretation mainly depends on the observer's experience which can lead to different interpretations of a same situation. In this article, we present how we have overcome this well-known issue at CSEM by a close collaboration with the Inselspital Bern. We have taken the advantage of Artificial Intelligence to devise a smart system capable of interpreting the cardiotocograms and provide accurate estimation of the pH level of the fetus (as a strong biomarker of hypoxia) as well as the detection of fetal hypoxia.

Cardiotocography (CTG) is the main medical surveillance method used to assess the fetal state during labor. As fetal hypoxia remains an important cause of neonatal morbidity and mortality, the question of suboptimal intrapartum surveillance is still omnipresent in obstetrics. Whereas the main goal of CTG is to prevent poor neonatal outcomes like stillbirth, neonatal acidosis or hypoxia, its impact has been widely discussed ever since its introduction in the late 1960s. CTG interpretation strongly depends on the observer's experience, thus, intra- and interindividual differences in interpreting the CTG are a long-known issue. This offers a suitable background for the use of artificial intelligence (AI) in CTG interpretation, following the resolutions that have occurred in fields of cardiology and radiology. In the literature, there are a few studies such as Petrozziello-2019 [1] and Cömert-2019 [2] who tried to predict fetal hypoxia using AI-based methods on CTGs. Such studies focused only on the prediction of hypoxia and have not provided any estimation of pH values.

In this work, we have designed a hybrid method based on i) digital signal processing and conventional machine learning (DSP), and ii) deep learning models (DL) to estimate the pH level of the fetus and the predication of fetal hypoxia using CTGs. The hybrid method takes the advantage of both DPS and DL methods by providing the pH values as the weighted average of the results of DSP and DL methods. For the DSP method, we have developed several physiological features to represent CTGs and various models (e.g., Bayes, support-vector machine, decision tree). In addition, we have devised and optimized a convolutional neural network (CNN) in the DL method.

The proposed method has been evaluated on a large-scale clinical data consisted of around 14000 CTGs recorded by Inselspital. After a careful data cleaning and screening process, around 6000 CTGs with pH values ranging from 6.7–7.6 have been used to evaluate the method. However, this work is still ongoing, and more data inclusion is expected. As the validation strategy, the data has been split into 60-20-20% portions as the training, validation, and test sets. Table 1 compares the performance of the methods based on DSP, DL, and hybrid for the detection of hypoxia as well as the estimation of pH values.

The results (Table 1a) show that the hybrid method is able to detect hypoxia with an accuracy of 76%. The specificity of the hybrid method is slightly higher than the sensitivity which shows that the method is a bit biased toward the negative class (not hypoxia). One reason for this might be the imbalanced nature of the training dataset. Comparing the results of DSP, DL, and hybrid methods, the hybrid method is able to keep a proper balance between the sensitivity and specificity and provide a more reliable method. Another important point is the poor sensitivity of the DL method probably due to having limited number of samples, especially for the positive class (samples with hypoxia), in the training set. Compared to the literature, the proposed hybrid model has achieved comparable results where slight differences can be observed, probably since these works have used different and more limited datasets for their evaluations. The results of Table 1b suggest that all three models have almost similar performance in the estimation of pH with MSE, RMSE, and MAE of 0.01, 0.07, and 0.05, respectively.

As future prospective, it has been planned to add more data (around 3500 samples) to the database to make it more suitable for the training of the models, especially for the DL-based models. A bigger dataset would allow better optimization and training of the models and would provide further improvement of the results.

Table 1: Performance of the proposed methods for the detection of hypoxia and estimation of pH level in comparison with the literature.

a) Hypoxia detection

Reference	Method	Model	Acc.	Sen.	Spe.
[1]	DL	CNN	-	58	85
[2]	DSP	kNN, decision tree, SVM	-	77	93
CSEM	DSP	Bayes	72	69	73
CSEM	DL	CNN	77	31	82
CSEM	Hybrid	Bayes - CNN	76	63	78

Acc: Accuracy, Sen: Sensitivity, Spe: Specificity, kNN: k-Nearest Neighboir.

b) pH estimation

Reference	Method	Model	MSE	RMSE	MAE
CSEM	DSP	Support vector regression	0.01	0.07	0.05
CSEM	DL	RNN	0.01	0.07	0.06
CSEM	Hybrid	Bayes - CNN	0.01	0.07	0.05

MSE: Mean Square Error, RMSE: Root Mean Square Error, MAE: Mean Absolut Error

This work was partly funded by Canton of Bern. CSEM is thankful for their support.

• Fachärztin FMH Gynäkologie und Geburtshilfe Oberärztin, Inselspital Bern, Switzerland

[1] A. Petrozziello, C. W. Redman, A. T. Papageorghiou, I. Jordanov, A. Georgieva (2019). Multimodal convolutional neural networks to detect fetal compromise during labor and delivery. IEEE Access, 7, 112026-112036.

[2] Z. Cömert, A. Şengür, Ü. Budak, A. F. Kocamaz (2019). Prediction of intrapartum fetal hypoxia considering feature selection algorithms and machine learning models. Health information science and systems, 7(1), 1-9.