A Study on Classifying Fetal Distress from Large-Scale Cardiotocographic (CTG) Data Using Different Machine Learning Approaches

Obstetricians use Cardiotocography (CTG) during labor to look for vital information affecting the health and safety of mother and fetus. However, CTG faces various challenges such as high noise, interpretation inconsistency, and the need for continuous expert presence. This study aims to provide different artificial intelligence approaches where each one acts as supportive guidance to help overcome the present challenges of CTG and predict fetal with potential complications. In the first part of the study, we denoised CTG signals and used an algorithm for extracting important features using Japan Society of Obstetricians and Gynecology (JSOG) guidelines and applied four machine learning methods: SVM, RF, DT, and ANN on extracted features to test the performance of the algorithm in detecting high-risk birth on large data gathered under clinical conditions. The process was tested on pH alone, then on Apgar 1 and 5 only, and finally on pH, Apgar scores 1 and 5. The best result achieved was of Apgar 1 and 5 only using RF with an area under the curve (AUC) of 0.89. The second part discusses our multi-input convolutional neural network (CNN) model which bypasses the need for CTG guidelines and extracts features directly from CTG images and gestational age. The model predicts infants with potentially low Apgar scores and achieved an AUC of 0.958 when classifying infants with Apgar score 5 minutes < 7 and an AUC of 0.955 when Apgar score 1 or 5 minutes < 6. In the third part of the study, we used anomaly detection generative adversarial networks (ANOGAN) to analyze and rate normal and abnormal CTG images based on anomalies found in them. This study overcomes data imbalance, a major challenge in artificial intelligence, by training only on the majority class which is the patients with a pH of 7.1 or higher. Moreover, the original ANOGAN uses DCGAN architecture so we also implemented ANOGAN on WGAN and WGANGP architectures. The trained model is tested on a dataset that includes an identical number of normal and abnormal CTG where it generates each image in the test dataset and provides a score based on the similarity between real and generated images. The model is trained solely on negative class. We found that CTG images with pH < 7.1 tend to have a higher anomaly score than CTG with pH >= 7.1. Based on the results of the aforementioned studies, we conclude that our approaches could support each other in guiding medical teams in mitigating risks they encounter during childbirth.