DTEK0086 Biosignal Analytics

Preterm Birth Detection using electrohysteography

Background
Preterm birth is the birth that occurs before weeks 37 of pregnancy. Approximately 15 million infants are born preterm every year, and the rate is rising. Complications related to preterm birth are the most frequent cause of neonatal deaths and cause 1 million deaths yearly among children under five years. In addition, prematurely born babies are at risk of numerous different medical conditions (e.g., hearing problems, sight problems). Effective prediction of preterm births at the early stages of the pregnancy could be useful. There are studies to prevent preterm birth with early medical interventions. Electrohysteogram (EHG) is a method used for preterm prediction. EHG is a form of electromyography (EMG), where the electrical signals are collected from the surface of the uterus [1, 2].

Objective
The objective of this project is to explore preterm birth prediction using EHG signals collected from mothers before week 26 of pregnancy. Using machine learning, you need to differentiate the EHG signals into two classes: i.e., 1 for the mothers that gave preterm birth and 0 for the mothers with normal pregnancy. The analysis should be done in Python (more details in the Instruction Section).

For this course project, you need to:

1. Submit your Python script and your report of the observations, graphs, and conclusions made upon analyzing the given signals. It is suggested to submit a Jupyter Notebook file, including your code and report.
2. Give a 20-minute presentation about your work. Your presentation should include a description of
    a. The problem and the biosignal
    b. The steps in your analysis: e.g., what pre-processing methods you use, which features you extract, which machine-learning algorithms you use
    c. The results that you obtain: e.g., the accuracy of two machine learning methods
    d. Your evaluation and conclusion on the findings and methods

Data collection setup
The EHG signals have been recorded with 4 electrodes (E1-E4) placed around the navel:

- E1 placed 3.5 cm to the left and 3.5 cm above the navel
- E2 placed 3.5 cm to the right and 3.5 cm above the navel
- E3 placed 3.5 cm to the right and 3.5 cm below the navel
- E4 placed 3.5 cm to the left and 3.5 cm below the navel

In total, three different channels (potential differences) have been collected from these electrodes: i.e., S1=E2-E1, S2=E2-E3, S3=E4-E3. Each EHG record is about 30 minutes long. Records have been digitized at 20 Hz with a 16-bit resolution over a range of ± 2.5 mV. The data is extracted from the Physionets Term-Preterm EHG Database (https://physionet.org/content/tpehgdb/1.0.1/).

Structure of the data

The project includes two folders as "normal" and "preterm." The "normal" folder contains 20 records (i.e., CSV files) from the mothers whose pregnancy had normal length, and the "preterm" folder contains 19 records from the mothers that gave preterm birth. Each record was collected from a mother, including 3 channels (i.e., S1, S2, and S3) of 30-minute EHG signals.

Instruction

For the analysis, you should:

1. Use pre-processing techniques (such as filtering) if necessary.
2. Extract relevant time-domain and frequency-domain features from the EHG signals (e.g., summary statistics, RMS value, and resonance frequency).
3. Select two supervised machine learning algorithms and train two classifiers. Each classifier should predict 1 for preterm or 0 otherwise.
4. Evaluate the two classifiers using leave-one-out cross-validation:
    a. In leave-one-out cross-validation, each mother acts once as test data, and the remaining mothers act as training data.
    b. Remember to standardize the features inside the cross-validation to prevent information leakage. i.e., use the mean and standard deviation of the training data to standardize the training data and the test data
    c. Pool the results of the leave-one-out cross-validation and obtain the confusion matrix, accuracy, precision, recall, and f1-score.

Hint: You can utilize packages such as scipy, tsfresh, and tsfel for the pre-processing and feature-extraction steps, and packages such as scikit-learn for the machine-learning step.

[1] FERGUS, Paul, et al. Prediction of preterm deliveries from EHG signals using machine learning. PloS one, 2013, 8.10: e77154. (https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0077154)

[2] Despotović, D., Zec, A., Mladenović, K., Radin, N., & Turukalo, T. L. (2018, September). A machine learning approach for an early prediction of preterm delivery. In 2018 IEEE 16th International Symposium on Intelligent Systems and Informatics (SISY) (pp. 000265-000270). IEEE. (https://ieeexplore.ieee.org/abstract/document/8524818/)