Automatic Diagnosis of Cardiac Disease from Twelve-lead and Reduced-lead ECGs using Multi-label Classification

Saman Parvaneh, Prathic Sundararajan

Edwards Lifesciences, Irvine, USA

Background:

ECG is an essential tool for the clinical diagnosis of multiple cardiac diseases. The 2021 PhysioNet/CinC Challenge aims to develop algorithms to classify cardiac abnormalities from Twelve-lead and Reduced-lead ECGs. In this article, a multi-label classification using XGBoost was used for this challenge.

Method:

About 55% of the training dataset, consisting of 43,101 ECG records, had more than one diagnosis. Therefore, multi-label classification was selected for ECG classification. The training data was split randomly into in-house training (80%) and test sets (20%). Basic features (RMS, age, and gender) were extracted from all available leads for each record. The label powerset approach was used to transform a multi-label problem into a multi-class problem—one multi-class classifier XGBoost trained on all unique label combinations found in the training data. Python was used for feature extraction and model development.

Results:

Challenge scores on the challenge test set for 12-leads, 6-leads, 3-leads, and 2-leads were -0.14, -0.16, -0.18, and -0.18, respectively (Team: LINC). On average, the challenge score on a test set was 0.02 lower than the score that we got on the held-out in-house test set (e.g., subset of training set) for different lead combinations.

Discussion:

We proposed a multi-label XGBoost for the classification of cardiac abnormalities from twelve-lead and reduced-lead ECGs. We faced technical challenges in integrating extracted features (e.g., signal quality index, morphological features, and geometric heart rate variability) in Matlab with model development in Python. Therefore, our goal for the official phase is to further improve our models' performance by incorporating a richer feature set extracted in Python.