

# **Cardiac Abnormalities Prediction using Ensemble of Diverse Sequence Labeling Models**

Yale Chang, Sara Mariani, Annamalai Natarajan, Asif Rahman,  
Gregory Boverman, Shruti Vij and Jonathan Rubin  
Philips Research North America, Cambridge, USA

## **Background:**

Cardiac abnormalities are leading causes of deaths and early diagnosis is important for providing timely interventions. The goal of 2020 PhysioNet/CinC Challenge is to develop algorithms to diagnose multiple cardiac abnormalities using 12-lead ECG data. In this work, we develop an ensemble of sequence labeling models to predict 9 classes of cardiac abnormalities.

## **Method:**

The training dataset was split into 10 folds, where 8 folds were used for model training and the remaining 2 folds were used for model validation and testing respectively. Z-normalization was applied to each of the 12 ECG channels for each patient. We hypothesized that different cardiac abnormalities could be better predicted by applying different models to different subsets of input channels. Therefore, we trained an ensemble of sequence labeling models. A collection of temporal convolutional neural networks were trained on different combinations of 12 input ECG channels. The average binary cross entropy over 9 classes was used as the training loss. Class-balance sample weighting was applied for each class. The average AUROC was used as the validation metric. We selected the optimal cutoffs of 9 classes by maximizing the geometric mean of F-beta and G-beta on the validation set.

## **Results:**

Geometric mean of F-beta and G-beta on the in-house test set was 0.648. The result on the official test set was 0.626. In comparison, the geometric mean of the best individual model evaluated on the official test set was 0.600, giving evidence that the ensemble model would outperform any individual model.

## **Discussion:**

We proposed an ensemble of diverse sequence labeling models to predict 9 classes of cardiac abnormalities using 12-lead ECG data. Applying multiple techniques, including 1) random combinations of input channels; and 2) adding alternative sequence labeling models to the ensemble should increase the diversity of the ensemble model and further improve performance.