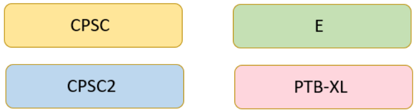
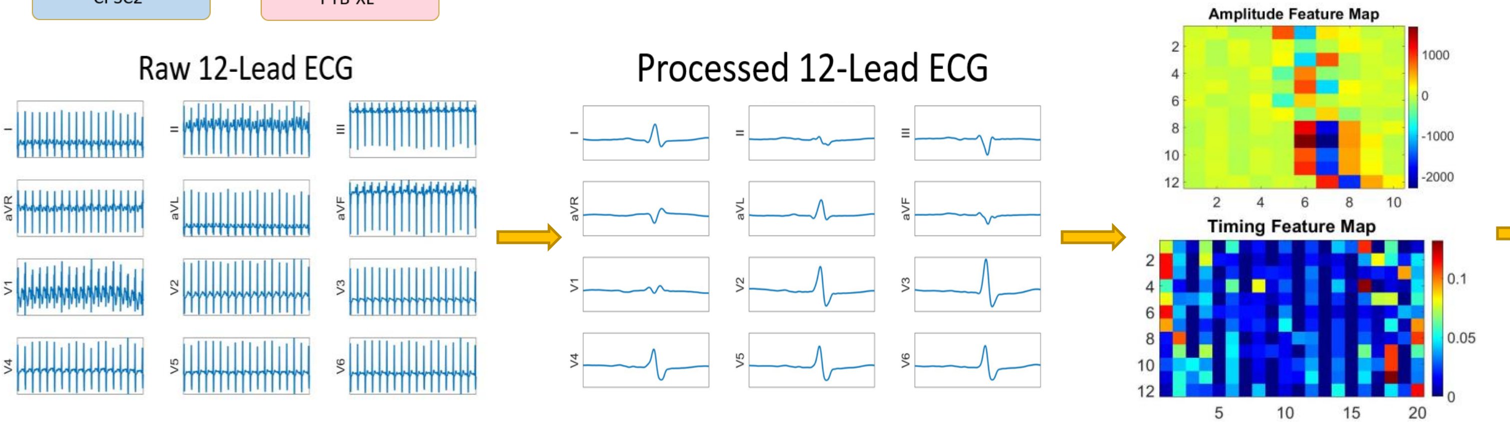
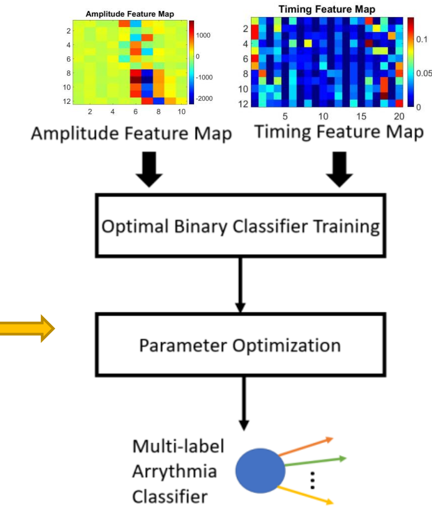
Interpretable Classifiers for Multi-label Arrhythmia with 12-Lead Electrocardiograms



Po-Ya Hsu*[1], Po-Han Hsu [1], Tsung-Han Lee [1], & Hsin-Li Liu [2]

[1]UC San Diego, [2]Central Taiwan University of Science and Technology





Abstract

In participation of the PhysioNet Arrhythmia Classification Challenge, our team, JuJuRock, devised a novel algorithm to classify multi-label arrhythmias. We developed a 12-lead ECG signal processing method that can handle ECG data of different lengths and from variant datasets. We formulated a novel 12-lead ECG amplitude and timing feature map generation technique. We also came up with an arrhythmia classification method that is adaptable to ECG data of different number of heartbeats. We conclude that our algorithm is not only efficient but also physiologically interpretable.

Methods

Our arrhythmia classification method is composed of five steps: **1) Dataset Selection:** We select CPSC, CPSC2, E, and PTB-XL datasets to build the classifier based on the signal quality and portions of evaluated classes. **2) Signal Processing:** We pass the raw signal to a low-pass filter and smooth the data of each lead with a 10ms window. Next, we search the R-peaks with the Pan-Tompkins algorithm and partition the data into 1-second long signal centered at the R-peak. **3) Feature Map Generation:** We devise a saliency-oriented feature map generation technique. The concept is to quantify the amplitude and timing of the P wave, QRS-complex, and the T wave. We select the representative ECG beat by using the structural similarity index and compute the amplitudes and timings of the salient waveforms. **4) Classifier Selection:** We experiment with convolutional neural network, recurrent neural network, boosting, support vector machine, random forest, and logistic regression classifiers. **5) Parameter Optimization:** We fine tune the number of heartbeats and the length of the data window required from a patient to correctly identify the arrhythmias in our algorithm.

Results & Conclusion

We considered boosting as the best classifier based on the sensitivity and specificity of the classification. Also, we found out that 10-second data length works the best in our algorithm. Below is the current result of our classifier:

Data	Time	Score
Training	6:00:00	0.402
Official	1:55:00	0.405

On top of its efficiency and satisfying score, our algorithm has successfully selected the physiologically reasonable feature maps for each type of arrhythmia.