Learning From Alarms: A Novel Robust Learning Approach to Learn an Accurate Photoplethysmography-Based Atrial Fibrillation Detector using Eight Million Samples Labeled with Imprecise Arrhythmia Alarms

Cheng Ding[1], ZhiCheng Guo[2], Cynthia Rudin[3], Ran Xiao[4], Amit Shah[5], Duc H. Do[6], Randall J Lee[7], Gari Clifford[1,8], Fadi B Nahab[9], Xiao Hu[1,4,8]

[1]Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA, USA
[2]Department of Electrical and Computer Engineering, Duke University, Durham, NC, USA
[3]Department of Computer Science, Duke University, Durham, NC, USA
[4]Nell Hodgson Woodruff School of Nursing, Emory University, Atlanta, GA, USA
[5]Department of Medicine, Emory University School of Medicine, Atlanta, GA, USA
[6]School of Medicine, University of California at Los Angeles, Los Angeles, CA, USA
[7]School of Medicine, University of California, San Francisco, CA, USA
[8]Department of Biomedical Informatics, Emory University School of Medicine, Atlanta, GA, USA
[9]Department of Neurology, Emory University School of Medicine, Atlanta, GA, USA

Atrial fibrillation (AF) is a common irregular heart rhythm indicating serious health conditions. Detecting AF in ambulatory settings is crucial, and Photoplethysmography (PPG) is a suitable method for passive monitoring due to its availability in wearable devices. However, existing AF detection algorithms using PPG often produce false positives, leading to unnecessary resource utilization. Deep neural networks (DNNs) offer accurate AF detection potential but lack sufficient training data and struggle with other arrhythmias. To overcome these challenges, this study introduces two innovations: a large-scale dataset of 8 million PPG strips from 24,100 patients and a novel approach using cluster membership consistency (CMC) loss, which introduces a cluster label from a pre-trained autoencoder, to alleviate the impact of label noise.

The performance of the AF detection on the Stanford dataset is summarized in table 1, comparing seven baseline algorithms with the proposed CMC method using 2 and 6 clusters. The comparison across the two signal quality subgroups shows consistently better performance on the good quality subgroup compared to the bad quality subgroup. With all three groups, CMC produces either the best or the second-best performance among testing scenarios. Notably, CMC-6 demonstrates a strong performance on good quality signals and ranks second to DivideMix on bad quality signals.

Table 1. Performance comparison between the proposed method with state-of-the-art algorithms on the test set.

Methods	whole dataset	bad quality	good quality
CE	0.5853	0.5277	0.9051
SCE	0.5859	0.5359	0.7429
Co-teaching	0.5448	0.4626	0.8217
INCV	0.6082	0.5472	0.9245
DivideMix	**0.7437**	**0.6983**	0.9631
ELR	0.503	0.4807	0.8266
SOP	0.6633	0.5752	0.9614
CMC-2	0.7277	0.6779	0.9614
CMC-6	0.7416	0.6847	**0.9724**

The contributions of this study include 1) a new approach to generate large labeled datasets using the concurrent alarm data; 2) the introduction of the CMC loss for robust learning, and 3) the demonstration of the proposed approach's effectiveness in handling label noise and poor-quality signals. This work provides valuable insights for PPG-based AF detection and has the potential for broader applications in deep learning.