Title: Validation of Deep Learning-Based Arrhythmia Classification Using 12-Leads ECGs: A Multi-Database Approach

Authors:
Long Dang, PhD. Pedro A. Moreno-Sánchez, Prof. Mark van Gils

Keywords:
deep learning, electrocardiogram, arrhythmia, bioengineering and biotechnology, bioinformatics and computational biology,

Abstract
Current practices in detecting of arrhythmia based on the electrocardiogram (ECG) demand extensive expertise. Deep Learning (DL) approaches have revealed high accuracy results in detecting arrhythmia. However, due to the fact that unique datasets are used for different studies, noticeable performance variations exist when methods applied to unseen data. This paper examines the reliability of a DL 12-leads ECG arrhythmia classifier by evaluating its performance variation in several dataset combinations.

We used a model from a prior study[1], which leveraged ResNet34 in analyzing annotated recordings (CPSC2018). We developed extra models with public datasets from the PhysioNet-Challenge-2021[2][3]: CPSC-2018-extra, PTB-XL, Chapman-Shaoxing. Nine categories were examined, all contained in the databases, except one class missing in the Chapman. Two approaches were used: a 08-class classification, using CPSC2018, CPSC2018-extra, PTB-XL for development with 80/20 hold-out, an independent test from the Chapman; and a 09-class case, with PTB-XL for training with the same hold-out ratio, the combination of CPSC2018, CPSC2018-extra for testing.

When compared to single-source models, our study reveals that while the classifier for the 08-class problem shows a slight F1-score decrease of 2.35% and for the 09-class cases experiences a significant drop of 39.52%, 08-class model's performance metrics demonstrates superior performance in 05 out of 08 classes.

The findings indicate that whereas the model performs marginally better when trained on a combined dataset, suggesting data diversification enhances the model's generalization ability, however its performance is notably reduced in imbalanced training data observed in the 9-class case, highlighting the model's sensitivity to class distribution.

References:

[1] Chukwu, Emmanuel C., and Pedro A. Moreno-Sánchez. «Enhancing Arrhythmia Diagnosis with Data-Driven Methods: A 12-Lead ECG-Based Explainable AI Model». https://doi.org/10.1007/978-3-031-59091-7_16.
[2] Reyna MA, Sadr N, Perez Alday EA, Gu A, Shah AJ, Robichaux C, Rad AB, Elola A, Seyedi S, Ansari S, Ghanbari H, Li Q, Sharma A, Clifford GD. Will Two Do? Varying Dimensions in Electrocardiography: The PhysioNet/Computing in Cardiology Challenge 2021. Computing in Cardiology 2021; 48: 1-4,
[3] Reyna MA, Sadr N, Perez Alday EA, Gu Annie, Shah AJ, Robichaux C, Rad AB, Elola A, Seyedi S, Ansari S, Ghanbari H, Li Q, Sharma A, Clifford GD. Issues in the automated classification of multilead ECGs using heterogeneous labels and populations. Physiol. Meas. 2022.