Cardiac arrhythmias classification in Kardiovize population study

Martin Pesl1,2,3*, Jakub Hejc4*, Tomas Kulik1,2, Tomas Vicar4, Petra Novotna4, Marina Ronzhina4, Juan Pablo Gonzalez Rivas2,5, Juraj Jakubik2, Zdenek Starek1,2

1 1st Department of Internal Medicine, Cardio-Angiology, Faculty of Medicine, Masaryk University, Brno, Czech Republic
2 ICRC, St. Anne's University Hospital, Brno, Czech Republic
3 Department of Biology, Faculty of Medicine, Masaryk University, Brno, Czech Republic
4 Department of Biomedical Engineering, Brno University of Technology 5 Department of Global Health and Population, Harvard T.H. Chan School of Public Health, Boston, US.
* These authors contributed equally

Aims: Automatic detection of arrhythmias becomes essential in population studies. Besides available commercial solutions, there are new options for population data handling, such as deep-learning models. In this study, we compare two different approaches and evaluate them on data from atrial fibrillation patients.

Methods: 12-lead ECGs recorded in 397 volunteers of the Kardiovize study by ELI™ 350 ECG Mortara were used. The Mortara system automatically assigned each record to one or more of 88 various categories. Additionally, the ECGs were analyzed by a self-developed deep-learning arrhythmia detector. The detection approach consisted of data preprocessing (downsampling, random amplification, time shift, and random noise alteration) and the ResNet model training. The training database was created using publicly available datasets (total 43,000 records) containing variable data. The detector assigned each record to one or more of 24 categories. The evaluation results were compared. The mismatches were visually inspected and revised. The evaluation process was focused on atrial fibrillation (AF) as one of the most common arrhythmias in the Czech population.

Results: On the training set, the F1-score of the model reached 0.86 and 0.87 for normal sinus rhythm and AF, respectively. In both categories, false positives occur. One of the reasons for the model misclassification was incorrect expert evaluation used as a predicted model output. In the test phase, no records were assigned to the AF category. On the contrary, the Mortara system classified six records as AF. Visual verification confirmed the correctness of the model output.

Conclusion:
From the present pilot study, the deep-learning classificator output contradicts the commercially available ECG evaluator. Notably, no false atrial fibrillation detections were indicated in the model output. According to the expert, ECG evaluation using a deep-learning model seems to be more appropriate for handling population data than the recent solution (Mortara).

Acknowledgement: This study was funded by the European Union's Horizon 2020 Future and Emerging Technologies Programme (Grant number 732170).