

Reproducibility of machine learning models for paroxysmal atrial fibrillation onset prediction

Cédric Gilon¹, Jean-Marie Grégoire^{1,2}, Jérôme Hellinckx¹, Stéphane Carlier², Hugues Bersini¹

1 IRIDIA, Université Libre de Bruxelles

2 Université de Mons

Background

Atrial fibrillation (AF) is the most common heart arrhythmia. Paroxysmal AF onset prediction is a much more complex task than screening AF. Published methods using the Physionet AFPDB database show excellent results, suggesting that forecasting AF episodes is possible by implementing machine learning (ML) models using heart rate variability (HRV) parameters.

Aims

Reproduce previously obtained results by published studies using the Physionet database and a larger database of unselected real-life patients.

Methods

We searched the literature for all articles on the forecasting of paroxysmal AF episodes. We analysed in depth the methodology of 3 recent studies using ML methods, to replicate their results. We screened our ECG Holter monitoring database of 11833 Holters to find those with paroxysmal AF episodes recorded. A total of 214 Holters with paroxysmal AF were labelled. We developed two ML models (deep neural network and a random forest model) for AF forecast using 13 HRV parameters. We compared performances of published models and our models using the Physionet database and our real-life database of patients.

Results

21 papers dedicated to the prediction of the onset of AF episodes have been published so far, showing exciting results culminating in sensitivities of 98%, specificity of 95% and accuracy of 98%. Using each model description available in the publications, we could not reach the published performances on the Physionet database. In addition, our models obtained a lower sensitivity of 84% for a specificity of 49% on the Physionet database, similar to the sensitivity of 80.1% for a specificity of 52.8% on our larger database.

Conclusion

ML models need to be more detailed if the reported results must be reproducible. Progress must still be made before the clinical use of algorithms that can anticipate paroxysmal AF. The use of larger databases is mandatory for this type of prediction.

300 words

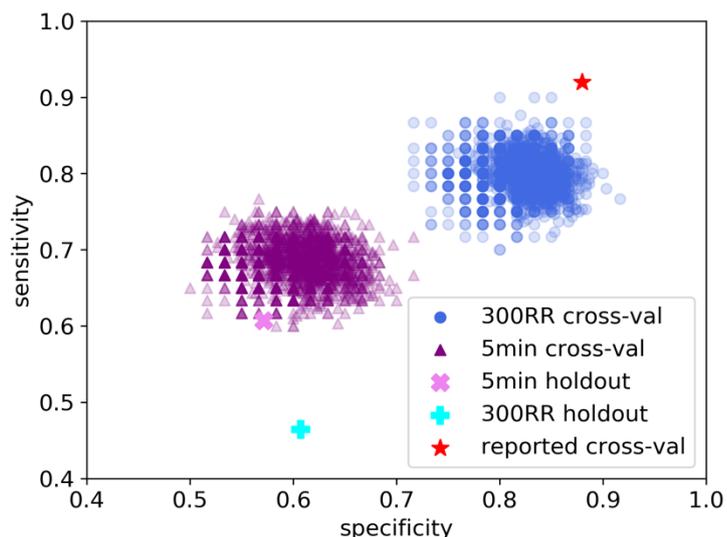


Figure 1: Reproduced results from [1] on the Physionet database. The reported sensitivity and specificity in the published work are shown as the red star. In total, 1000 10-fold cross-validation were evaluated with either 300 RR intervals window or 5 minutes window (blue dot and purple triangle). Our results using described methods on the Physionet database are poorer on the test sets (pink and light blue crosses).

[1] Ali Narin et al. "Early prediction of paroxysmal atrial fibrillation based on short-term heart rate variability". In: *Physica A: Statistical Mechanics and its Applications* 509 (Nov. 2018), pp. 56–65. issn: 03784371. doi: 10.1016/j.physa.2018.06.022.