

Development of a predictive model for new onset of atrial fibrillation in a hospital cohort

Ilaria Gandin^a, Arjuna Scagnetto^b, Andrea Di Lenarda^b, Giulia Barbati^a

Introduction

Atrial fibrillation (AF) is a type of cardiac arrhythmia associated with major adverse events, like hospitalizations and death. Tools and techniques for the prediction of AF are required to identify high-risk individuals. Currently, the most used tool to this aim is the CHARGE-AF risk model [1], a standard survival model based on a small number of variable readily available in primary care setting.

The present work has two main objectives:

- 1) the development of a predictive risk score based on machine learning (ML) algorithms
- 2) performing an extensive comparison between standard and ML methods

Data and Methods

We analysed data of 16887 patients examined at the Cardiovascular Observatory of Trieste [2] between 2009 and 2014.

96 features were considered from Electronic Health Records (EHRs): demographics, clinical parameters measured during the cardiological visit, ECG parameters, prescriptions, previous diagnosis, comorbidities.

We implemented two algorithms:

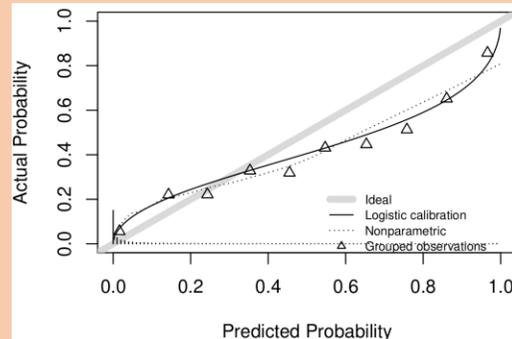
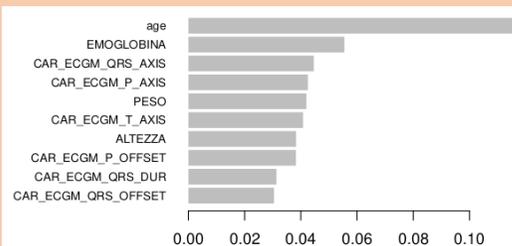
- the penalised logistic regression (LR)
- the gradient boosting (XGB), an ensemble method based on tree weak learners.



Results

In our preliminary analysis we observe a modest increase in the performance of machine learning approaches.

Task	Model	AUC	95% CI
1-year n=32795 AF=3%	CHARGE	0.70	[0.67,0.75]
	LR	0.74	[0.68,0.78]
	XGB	0.72	[0.68,0.76]
5-years n=28611 AF=11%	CHARGE	0.70	[0.70,0.71]
	LR	0.80	[0.78,0.82]
	XGB	0.84	[0.82,0.86]



Discussion

Further analyses are required to understand the benefit in using ML algorithms in this context. Next developments of the work are:

- implementation of more advanced tuning techniques for model's hyper-parameters (instead of grid search)
- recalibration techniques, in particular isotonic regression fitting and assessment of recalibrates predictions using Brier score and reliability plots
- investigation of the role of new emerging high-impact variables