European Heart Journal (2024) 45 (Suppl 1)
Preventive Cardiology – Risk Factors and Prevention, Cardiovascular Risk Assessment

Predicting atrial fibrillation within 7 years follow-up using machine learning models

K.H. Tsarapatsani[1], D. Biros[2], V. Tsakanikas[1], A. Sakellarios[1], H.J. Trampisch[3], H. Rudolf[4], G.K. Matsopoulos[5], D.I. Fotiadis[1]

[1]Foundation for Research and Technology-Hellas, Ioannina, Greece
[2]University Hospital of Ioannina, First department of internal medicine, Ioannina, Greece
[3]Ruhr University Bochum, Department of Medical Informatics, Biometry and Epidemiology, Bochum, Germany
[4]Rostock University Medical Centre Part of the Rostock University, Rostock, Germany
[5]National Technical University of Athens, School of Electrical and Computer Engineering, Athens, Greece

Funding Acknowledgements: Type of funding sources: Public grant(s) – EU funding. Main funding source(s): This work has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 101017424, as part of the TIMELY project

Background: The prevalence of atrial fibrillation (AF), the most common type of cardiac arrhythmia, has increased globally. AF is also the leading cardiac factor for several cardiovascular and cerebrovascular events. Early prediction and treatment of AF prevents major cardiovascular and cerebrovascular events.

Purpose: Our objective is to predict AF risk within 7-years follow-up using only twelve basic factors and applying machine learning (ML) models.

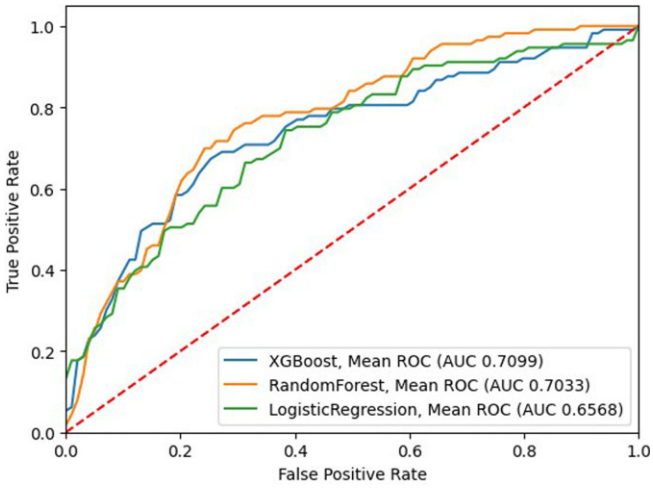
Methods: Data from the German epidemiological trial on ankle brachial index (getABI) were utilized in this work. GetABI includes 6,457 patients with mean age of 72.53 years. 3,985 individuals were assessed for AF of which 393 patients suffered from AF at 7-years follow-up. The feature selection was initially done manually, ending up in 45 features. The 10 best features were extracted from this dataset according to the SHAP (SHapley Additive exPlanations) values. The final features were body mass index, sensitive CRP, cholesterol, LDL-cholesterol, troponin I, NT-pro-BNP, glomerular filtration rate, ratio of triglycerides and HDL-cholesterol, gamma-GT, heart rate, including age and sex as major factors. The handling of missing values was achieved by SimpleImputer, estimating the mean values. XGBoost and Random Forest predicted AF. Logistic Regression was also used as a baseline model. The data were split into training and test set and ML models were applied for 100 iterative runs, extracting results with mean values.

Results: XGBoost was the most accurate model, predicting the AF with 71.19 % mean accuracy, 66.37 % mean sensitivity and 75.61 % mean specificity. Figure 1 depicts the performance of all models. Additionally, Receiver Operating Characteristic Curve (ROC) and Area Under Curve (AUC) for each model are presented in Figure 2. XGBoost also outperformed all the others models, achieving the highest mean AUC value, equal to 70.99 %. Contrariwise, logistic regression had the lowest overall performance.

Conclusion: ML models, specifically XGBoost, achieved accurate AF prediction in 3,985 individuals utilizing only twelve clinical-routine data.

ML models\Metrics	Mean Accuracy (%)	Mean Sensitivity (%)	Mean Specificity (%)
XGBoost	**71.19**	**66.37**	**75.61**
Random forest	70.76	60.18	80.49
Logistic regression	66.10	55.75	75.60

Evaluation results of each used model.



ROC curve and AUC value of each model.