ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - XIII^th SISMEC Congress - Vol. 20 Suppl. 1

# Prediction of Atrial Fibrillation 10–Year Risk with Optimal Survival Tree Models

*Lofaro Danilo[1], Armentaro Giuseppe[2], Vizza Patrizia[3], Veltri Pierangelo[3], Sciacqua Angela[2], Conforti Domenico[4]*

*(1) Department of Mathematics and Computer Science, University of Calabria, Rende (CS), Italy*
*(2) Department of Medical and Surgical Sciences, Magna Graecia University, Catanzaro, Italy*
*(3) BioHER Lab, Department of Computer Engineering, Modeling, Electronics and Systems, University of Calabria, Rende (CS), Italy*
*(4) deHealth Lab, Department of Mechanical, Energetic and Management Engineering, University of Calabria, Rende (CS), Italy*

CORRESPONDING AUTHOR: *Lofaro Danilo, danilo.lofaro@unical.it*

## BACKGROUND

Atrial fibrillation (AF) is the most common cardiac rhythm disorder in adults and old subjects with an estimated global prevalence of 35 million cases worldwide and increasing incidence in the next decades [1]. Traditional AF risk scores – Framingham, ARIC, CHARGE AF, CHA2DS2-VASc and SAAFE [2–6] – reach C-statistics around 0.75-0.80. Recently, there has been a growing interest in applying machine learning (ML) techniques to develop predictive models for AF. Many of these models have pushed discrimination performance a little higher, sacrificing interpretability, since the "black-box" nature of the employed algorithms [7].

## OBJECTIVE

To build and internally validate an interpretable model that predicts the 10–year probability of AF-free survival, using the recently proposed Optimal Survival Tree (OST) algorithm [8].

## METHODS

Data analyzed came from the Catanzaro Atrial Fibrillation project [9], an observational prospective cohort study that included outpatients enrolled from January 1998 to December 2018, referred to the University Hospital of Catanzaro - Italy, for cardiac clinical evaluation.

Patients with end-stage renal disease, active malignancy, thyroid dysfunction, cardiomyopathy, rheumatic and non-rheumatic valvular heart disease, or prosthetic valves, were excluded as well as those with previous acute myocardial infarction or stroke.

Predictors included in the analyses were: i) Demographic and anthropometric measures: age, sex, BMI, waist circumference; ii) Medical history: hypertension, diabetes, heart failure, vascular disease, COPD, previous TIA, CHA2DS2 VASc components; iii) Laboratory measures: fasting glucose, total/HDL/LDL cholesterol, triglycerides, eGFR; Imaging derived variables: Left Atrial Volume index (LAVi), Left Ventricular Mass index, Carotid Intima-Media thickness. Time-to-first AF diagnosis was right-censored at 10 years.

The OST model was benchmarked against three established tree-based algorithms: survival CART, survival conditional-inference trees (cTree) and random survival forests (RF).

The models were trained on a randomly selected subset of patients (70%) and their predictive performances were subsequently evaluated and compared on the remaining 30%. A 5-fold cross-validation based grid search was used to tune the models' hyper-parameters. Discrimination (time-dependent AUC), accuracy (Brier score, integrated Brier score, Index of Prediction Accuracy – IPA) and calibration (Integrated Calibration Index - ICI, E50, E90) were assessed .

## RESULTS

A total of 4114 patients were selected (mean age 59.06 ± 11.73, 48.1% Females). During a mean follow-up of 59 ± 19 months, AF occurred in 533 patients (13%). At baseline, AF patients showed on average a worse clinical profile in terms of anthropometric measures (BMI and Waist circumference), renal function (eGFR), cardiovascular risk factors (Diabetes, Hypertension, Heart failure, previous TIA), CHA2DS2-VASc score and echocardiographic parameters.

The final OST model (Figure 1) relied on only four variables - LAVi, Glucose, Age, and CHA2DS2-VASc - creating six leaves that collapsed into four clinically meaningful risk profiles: i) Very-low risk: Either LAVi ≤ 34 mL/m2, glucose ≤ 97 mg/dL, CHA2DS2-VASc ≤ 2, or same LAVi, Glucose > 97, and age ≤ 71 y (n = 2082, expected AF-free

DOI: 10.54103/2282-0930/29402

survival 115-118 mo); ii) Low risk: Same LAVi/glucose but CHA2DS2-VASc > 2 (n = 213, 106 mo); iii) Moderate risk: Either LAVi ≤ 34 with higher glucose and age > 71 y or LAVi 34-39 (n = 399, 87-89 mo); iv) High risk: LAVi ≥ 40 (n = 186, 56 mo).

On the test cohort OST achieved AUCs of 0.856 and 0.794 and Brier scores of 0.086 and 0.134 at 5 and 10 years, respectively, slightly outperforming CART (5/10-y AUC 0.849/0.764; Brier 0.096/0.137) and cTree (0.846/0.766; 0.096/0.156), and trailing RF in the 5-year (0.894, 0.083), but not the 10–year prediction (0.804, 0.131). Calibration metrics favored OST over RF at both horizons.

## CONCLUSIONS

A parsimonious, easily explainable four-variable OST predicted 10–year AF risk almost as accurately as RF yet with superior calibration and bedside transparency. Adding a single echocardiographic measure (LAVi) to routine clinical data may enable personalized AF screening and targeted prevention. External validation in independent, multicentre cohorts is required to confirm the model's generalisability and to support its adoption in routine clinical practice.

## REFERENCES

1. Lippi G, Sanchis-Gomar F, Cervellin G. Global epidemiology of atrial fibrillation: An increasing epidemic and public health challenge. Int J Stroke. 2021 Feb;16(2):217–21.
2. Schnabel RB, Sullivan LM, Levy D, et al. Development of a risk score for atrial fibrillation (Framingham Heart Study): a community-based cohort study. The Lancet. 2009;373(9665).
3. Chamberlain AM, Agarwal SK, Folsom AR, et al. A clinical risk score for atrial fibrillation in a biracial prospective cohort (from the Atherosclerosis Risk in Communities [ARIC] study). Am J Cardiol. 2011 Jan;107(1):85–91.
4. Alonso A, Krijthe BP, Aspelund T, et al. Simple risk model predicts incidence of atrial fibrillation in a racially and geographically diverse population: the CHARGE-AF consortium. J Am Heart Assoc. 2013;2(2).
5. Saliba W, Gronich N, Barnett-Griness O, Rennert G. Usefulness of CHADS2 and CHA2DS2-VASc Scores in the Prediction of New-Onset Atrial Fibrillation: A Population-Based Study. American Journal of Medicine. 2016;129(8).
6. Linker DT, Murphy TB, Mokdad AH. Selective screening for atrial fibrillation using multivariable risk models. Heart. 2018 Sep;104(18):1492–9.
7. Tseng AS, Noseworthy PA. Prediction of Atrial Fibrillation Using Machine Learning: A Review. Vol. 12, Frontiers in Physiology. 2021.
8. Bertsimas D, Dunn J, Gibson E, Orfanoudaki A. Optimal survival trees. Mach Learn [Internet]. 2022 Aug;111(8):2951–3023. Available from: https://doi.org/10.1007/s10994-021-06117-0
9. Sciacqua A, Perticone M, Tripepi G, et al. Renal disease and left atrial remodeling predict atrial fibrillation in patients with cardiovascular risk factors. Int J Cardiol. 2014 Jul;175(1):90–5.

DOI: 10.54103/2282-0930/29402

ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - XIII$^{th}$ SISMEC Congress - Vol. 20. Suppl. 1

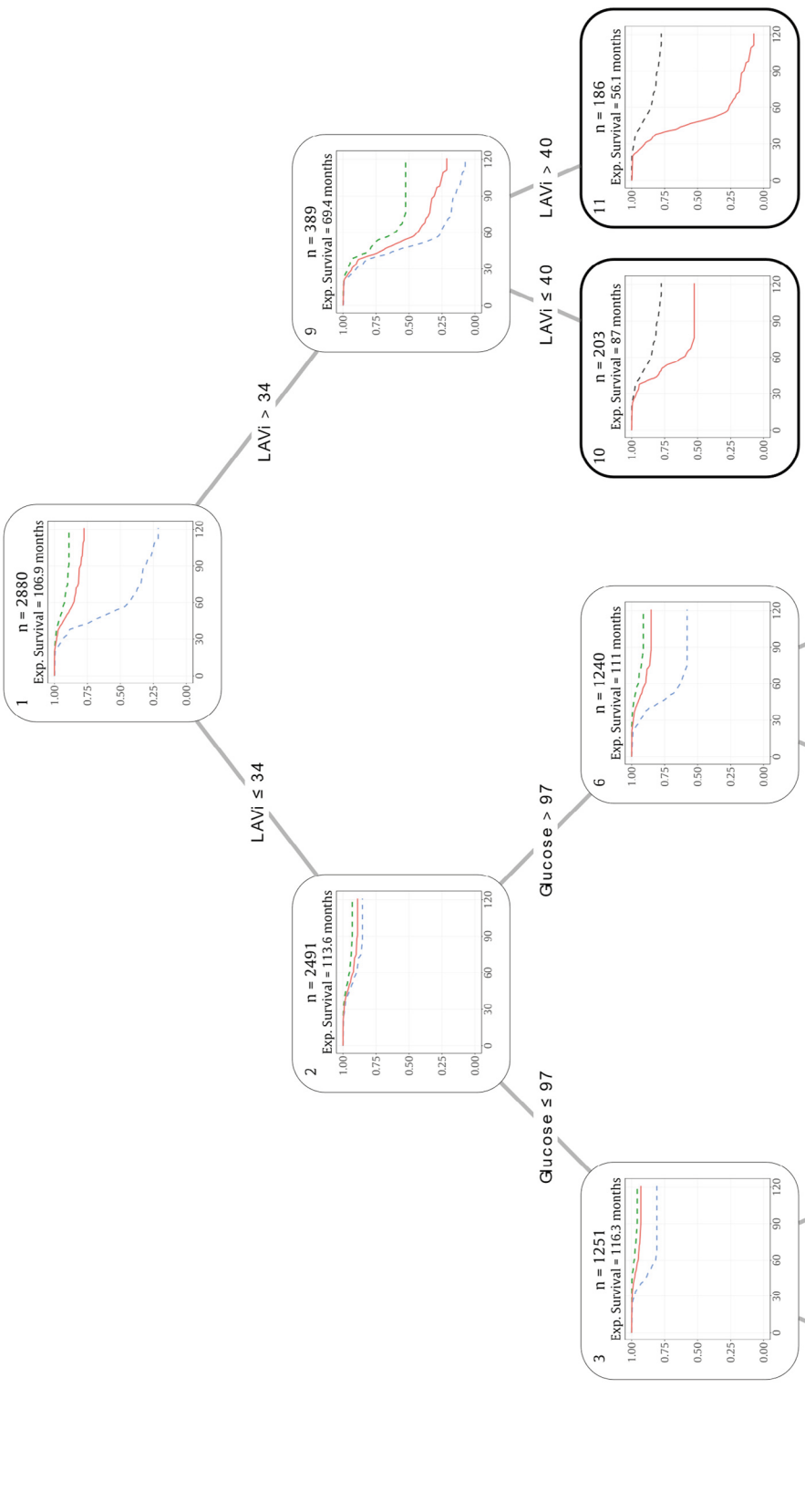


*Figure 1. Optimal Survival Tree for 10–years AF risk. The red survival curve shows the survival probability of the subgroup of patients included in the tree node, the dashed green curve shows the survival of the subgroup in the left child, and the dashed blue line shows that of the subgroup in the right child. In leaf nodes (ticker border) the dashed grey curve shows the survival in the entire cohort used to develop the tree.*

DOI: 10.54103/2282-0930/29402