



UNIVERSITÀ
DEGLI STUDI
FIRENZE

CICLO XXXVII

COORDINATORE Prof. Gian Maria Rossolini

Machine Learning for Risk Stratification in Atrial Fibrillation

Settore Scientifico Disciplinare MED/09

Dottorando

Dott. Bernardini Andrea

Tutore

Prof.ssa Marcucci Rossella

Coordinatore

Prof. Gian Maria Rossolini

*«They said, get back, honky cat
Better get back to the woods
Well, I quit those days and my redneck ways
Oh, a change is gonna do me good»*

Honky Cat, Elton John, 1972

Melissa: «Why do they call you the Duck? »

*Rubber Duck: «Because it rhymes with "luck."
See, my daddy always told me to be just like a duck.
Stay smooth on the surface and paddle like the devil underneath! »*

Convoy, directed by Sam Peckinpah, 1978

INTRODUCTION	4
AIM OF THE STUDY.....	6
GRAPHICAL ABSTRACT	7
METHODS	8
RESULTS	12
DISCUSSION	21
CONCLUSIONS.....	28
REFERENCES	29

INTRODUCTION

Atrial fibrillation (AF) is the most common arrhythmia with a currently estimated prevalence in adults between 2% and 4%¹. Moreover, a 2.3-fold rise in prevalence is expected due to extended longevity in the general population and the intensifying search for undiagnosed AF^{1,2}. Anticoagulation is the therapy of choice for the prevention of cardioembolic events in patients with AF. However, the risk of on-therapy major bleeding is not negligible and represents a common and prognostically relevant complication³. Despite of different combinations of items incorporated in the available different risk scores, the accuracy of available prediction tools for bleeding in AF patients on anticoagulation therapy remains modest⁴. In addition to the bleeding risk, AF is associated with an increased risk of stroke and death. While anticoagulation reduces the relative risk of stroke by 60-70% in AF patients, a relevant residual stroke risk of 1-2% per year remains in randomized trials⁵. On the other hand, although the use of oral anticoagulants lowers all-cause mortality in patients with AF⁶, the mortality rate in these patients is higher than in the general population, with cardiovascular (CV) etiology being an important cause of death. Prediction functions for events can be derived from data using Machine Learning (ML) algorithms. ML systems have the potential to improve predictive performances especially when diagnoses are made from large volumes or complex patterns of data using non-linear, highly interactive combinations of variables to uncover novel patterns^{7,8}. Therefore, these algorithms have been used in various contexts in AF, focusing on detection of AF or risk prediction modelling, in particular clinical outcome prediction^{7,9-11}. However, the previously published ML models predicted clinical outcomes in a not anticoagulated population^{10,11}, in a partially anticoagulated population^{7,12}, or in patients exclusively

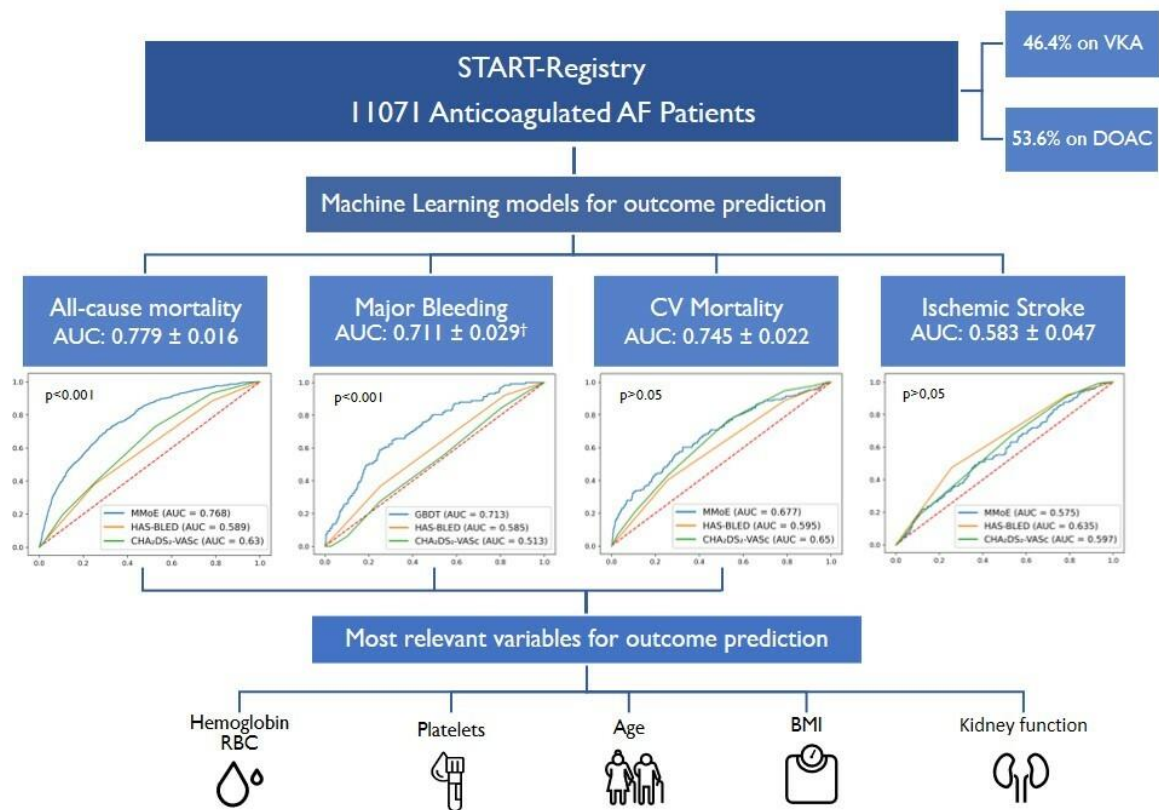
anticoagulated with vitamin K antagonists (VKA)¹³, thus limiting their applicability in the contemporary AF population, represented mainly by anticoagulated patients with DOACs¹⁴. Evidence is lacking for risk models to predict all-cause death, CV death, stroke, and major bleeding in an AF population entirely on anticoagulation therapy. Furthermore, when an anticoagulated population was considered in the aforementioned studies, the outcome prediction was carried out without taking into account the type of anticoagulation used (VKAs or DOACs).

The AF START-2 (Survey on anTicoagulated pAtients RegisTer, NCT02219984) is an Italian observational, multicenter, prospective cohort study designed to provide prescriptive and prognostic information on AF patients receiving long-term anticoagulation with VKAs or DOACs¹⁵.

AIM OF THE STUDY

The aim of this study is to use ML to provide an accurate tool to predict the risk of all-cause death, CV death, stroke, and major bleeding in a large cohort of anticoagulated patients with AF from the START-2 Register.

GRAPHICAL ABSTRACT



Graphical Abstract: Machine Learning approach for outcome prediction in AF patients.

[†]Major Bleeding AUC is referred to the DOAC population. ML: machine learning; CV: cardiovascular; AUC: area under the curve; RBC: red blood cell count; BMI: body mass index; VKA: vitamin K antagonists; DOAC: direct oral anticoagulants

METHODS

Study population

The data of the study were derived from the START-2 an independent, collaborative database that includes adult patients (>18 years) who started an anticoagulant therapy for non-valvular AF. Exclusion criteria were represented by life expectancy <6 months, and refusal of the patient to undergo follow-up visits. The START-Register procedures, ethical committee approval and data collecting have been previously described^{15,16}. The analysis was done on the whole cohort of patients reporting outcome events from June 2011 to January 2021. Patient's clinical features were recorded using a web-based CRF (case report forms). Baseline data, demographics, clinical characteristics, associated risk factors for stroke and bleeding, laboratory data, clinical indication for treatment, therapeutic range expected (in cases of treatment with VKA), and use of concomitant drugs were collected. Serum creatinine levels were measured by local hospital laboratories, with creatinine clearance (CrCl) calculated by the Cockcroft-Gault formula. Renal failure was defined as a creatinine clearance < 30 mL/min. Patients were stratified for stroke risk evaluation according to the CHA₂DS₂VASc¹⁷ score; baseline bleeding risk was evaluated with the HAS-BLED¹⁸ score.

Study Endpoints

The centers involved provided detailed clinical reports of any relevant clinical outcome by regular follow-up of the patients enrolled. Major bleeding and clinically relevant non-major bleeding (CRNMB) were defined according to the International

Society on Thrombosis and Haemostasis classification¹⁹. Stroke was defined as the onset of focal neurological symptoms lasting at least 24 hours and supported by congruent ischemic lesions at CT or MRI scans. The cause of death was considered cardiovascular if due to myocardial infarction, heart failure, stroke, CV hemorrhage, or an unwitnessed death occurring within 24 h after the onset of symptoms (e.g., sudden death)^{20,21}. The study endpoint is represented by the ML risk prediction of all-cause death, CV death, stroke, and major bleeding in a large cohort of anticoagulated patients with AF, considering the overall population and the two sub-groups (DOACs and VKA populations) in accordance to the specific anticoagulant strategy chosen.

Predictive Models

Different supervised ML models were used to assess the study endpoint. Stepwise logistic regression (SLR) and gradient-boosted decision trees (GBDT)²² were applied to predict all-cause death, CV death, stroke, and major bleeding individually. Different types of multi-task neural networks (MTNN) were trained to jointly predict the same outcomes. The first predictive technique that was used was SLR, which is a method of fitting regression models by iteratively testing the statistical significance of each independent variable and selecting the best ones ($p\text{-value} > 0.05$). The second technique used was the GBDT, which combines (by voting) multiple weak models to form a strong model. The number of estimators and the learning rate for GBDT were chosen to be 125 and 0.01, respectively, through a grid search. GBDT models can provide a list of the most important features used to make predictions. However, while the importance of a feature can indicate which features are generally most important for prediction, it does not reveal how the value of the feature influences it. SHAP

(SHapley Additive exPlanations), a game theory-based approach, was used to explain the model output²³. SHAP assigns each feature an importance value for a specific prediction and connects optimal credit allocation with local explanations using classical Shapley values from game theory. Therefore, it is possible to understand how the value of each feature influences the prediction. Some multi-task learning techniques using neural networks were also experimented, with either a simple shared-bottom architecture²⁴ or a Multi-Gate Mixture of Experts (MMoE)²⁵. The shared-bottom architecture had three dense layers with 128, 128 and 256 units, respectively, and RELU nonlinearities. Unlike SLR and GBDT, MMOE can learn both shared and task-specific features from the data. The MMOE approach had eight expert networks with 16 units each for the shared part and a layer of 32 specialized units for each task. In both cases, Adam (number of epochs=100) was used to fit the models²⁶. Regularization techniques such as L1/L2 regularization and dropout layers were used in neural networks to mitigate overfitting and create robust data representations. Baseline variables as reported on the registries' CRF were used as potential predictors. In total, 68 variables were considered for all predictive models (*Supplementary Material, Table 1*)

Metrics and statistical analysis

As a metric, the area under the ROC curve (AUC, Area Under the Curve) was used in a 5-fold cross-validation, in which the data were divided into five subsets and each subset was used once as a test set, while the remaining subsets were used as training

sets. To ensure a correct stratification of the data on the four targets (death, CV death, ischemic stroke and major bleeding) on the folds, the technique of iterative stratification²⁷ was used, which is a method that divides a multi-label dataset by considering each label separately, beginning with the one with the fewest positive examples and progressing to the best represented one. Therefore, each fold contains a balanced number of examples from each label combination. Pooled²⁸ ROC curves were compared using the DeLong test²⁹. Continuous variables were expressed as mean (M) and standard deviation (SD) while categorical variables as number (n) and percentage (%). Differences between continuous values were assessed using the unpaired t-test, categorical variables were compared by the Chi-square test or Fisher exact test, as appropriate. The incidence of death, ischemic events and bleeding were calculated by dividing the number of events by the person time at risk. The incidence rate ratio with the 95% confidence interval (95%CI) was calculated. P-values <0.05 were considered statistically significant. Statistical analysis was performed with SPSS Statistics version 22 (IBM Corporation) and Stata V.14 statistical software package (StataCorp, College Station, Texas, USA).

RESULTS

Study population

The study population consisted of 11078 patients, median age was 77 years (IQR 71-83) and 6029 (54.3%) were males. Patients were followed up for a median time of 1.5 years (IQR 1.0-2.6), with a maximum of 9.7 years. Detailed baseline characteristics of the study population are reported in **Table 1**. Regarding the anticoagulant treatment, 5135 (46.4%) were on VKA and 5943 (53.6%) were on DOACs, of whom 2562 (43.1%) were treated with a reduced dose. Median age was higher in VKA patients than in DOAC patients (79 vs. 76 years, $p < 0.001$), while sex distribution did not differ between the two groups. The rate of paroxysmal AF, hypertension, thrombocytopenia, history of previous bleeding, stroke, and dementia was higher in the DOAC population, while, coronary artery disease, heart failure and renal failure were more common in VKA patients. The mean HAS-BLED was 2.1 ± 0.8 in the overall population and did not differ between the two groups (VKA: 2.0 ± 0.9 vs. DOACs: 2.1 ± 0.8 ; $p = 0.4$); the mean CHA₂DS₂VASc score was 3.7 ± 1.5 in the total cohort, slightly higher in DOAC than VKA users (3.7 ± 1.6 vs 3.6 ± 1.5 ; $p = 0.03$).

	Overall population	Patients on VKAs	Patients on DOACs	p-value
Patients, n (%)	11078	5135 (46.4)	5943 (53.6)	
Males, n (%)	6029 (54.3)	2825 (55.0)	3191 (53.7)	0.2
Median age (range) [IQR]	77 (18-99) [71-83]	76 (18-98) [70-82]	79 (20-99) [72-84]	<0.001
Age > 75 years, n (%)	6896 (62.2)	2390 (57.1)	3966 (66.7)	<0.001
Median follow-up, years (range) [IQR]	1.5 (0.1-9.7) [1.0-2.6]	1.9 (0.1-9.7) [0.9-3.5]	1.3 (0.1-8.9) [1.0-2.0]	<0.001
Paroxysmal AF, n (%)	2340 (21.1)	1682 (32.8)	2340 (39.6)	<0.001
Hypertension, n (%)	8889 (80.2)	4002 (77.9)	4887 (82.2)	<0.001
Diabetes Mellitus, n (%)	2275 (20.5)	1048 (20.4)	1226 (20.6)	0.8
Heart failure, n (%)	2115 (19.1)	1046 (20.4)	1069 (18.0)	0.001

Coronary Artery Disease, n (%)	1869 (16.9)	1019 (19.9)	848 (14.3)	<0.001
Peripheral Arterial Disease, n (%)	648 (5.8)	332 (6.5)	316 (5.3)	0.004
COPD, n (%)	1355 (12.2)	623 (12.1)	732 (12.3)	0.03
Active Cancer, n (%)	308 (2.8)	140 (2.7)	168 (2.8)	0.5
Dementia, n (%)	341 (3.1)	113 (2.2)	228 (3.8)	<0.001
Anemia, n (%)	300 (2.7)	145 (2.8)	155 (2.6)	0.5
Frequent fall, n (%)	221 (2.0)	109 (2.1)	112 (1.9)	0.7
Previous bleeding, n (%)	412 (3.7)	122 (2.4)	290 (4.9)	<0.001
Previous Stroke, n (%)	1688 (15.2)	725 (14.1)	963 (16.2)	0.004
eGFR < 30 mL/min	560 (5.1)	369 (7.2)	191 (3.2)	<0.001
Antiplatelet drugs, n (%)	1306 (11.8)	707 (13.8)	599 (10.1)	<0.001
Mean CHA ₂ DS ₂ VASc score (SD)	3.7 (1.5)	3.6 (1.6)	3.7 (1.5)	0.03
CHA ₂ DS ₂ VASc ≤2, n (%)	2374 (21.4)	1255 (24.4)	1119 (18.8)	<0.001
Mean HAS-BLED score (SD)	2.1 (0.8)	2.0 (0.9)	2.1 (0.8)	0.4
HASBLED≥3, n (%)	2840 (25.6)	1335 (26.0)	1505 (25.3)	0.4
DOAC reduced dose, n (%)			2562 (43.1)	

Table 1. Characteristics of patients according to type of anticoagulant treatment. COPD: chronic obstructive pulmonary disease; GFR: glomerular filtration rate

Clinical outcomes

The incidence of adverse events recorded during follow-up is shown in **Table 2**. During follow-up, 785 (14.1%) patients died, of whom 169 (21.6%) due to CV causes; 542 bleeding events were recorded (2.44 per 100 pt-yrs), of which 240 were major (rate 1.08 x 100 pt-yrs). Major bleeding was fatal in 25 cases (10.4%). Stroke occurred in 52 (0.23 x 100 pt-yrs) patients during the follow-up. A detailed analysis of the study population and outcomes has been previously published¹⁶.

	Overall population N. (rateX100 pt-yrs)	Patients on AVKs N. (rateX100 pt-yrs)	Patients on DOACs N. (rateX100 pt-yrs)
<i>All Bleedings</i>	541 (2.44)	290 (2.40)	251 (2.48)
<i>Major Bleedings</i>	240 (1.08)	119 (0.99)	121 (1.19)
<i>CRNMB</i>	301 (1.34)	171 (1.41)	130 (1.28)

Median Time to Bleedings, years [range]	0.8 (0.1-7.0)	1.2 (0.1-6.7)	0.9 (0.1-5.8)
<i>Thrombotic events</i>	84 (0.37)	36 (0.30)	48 (0.47)
Stroke	52 (0.23)	21 (0.17)	31 (0.30)
Median Time to Stroke, years (range)	1.4 (0.1-7.0)	1.3 (0.1-6.5)	1.2 (0.1-5.4)
<i>Death</i>	781 (3.50)	550 (4.54)	231 (2.28)

Table 2. Adverse events recorded during follow-up. CRNMB: Clinically Relevant Non-Major Bleeding

ML models discrimination

The performance of ML models and logistic regression stepwise are shown in **Table 3**. The best performance models for each category and their comparison with clinical risk scores are reported in **Table 4**. In general, the AUC resulted higher for all-cause death and lower for stroke prediction. MMoE outperformed other predictors on all tasks, except for major bleeding in DOAC-treated patients, in which GBDT outperformed all other predictors. Naïve multi-task provided the best prediction regarding ischemic stroke. In all items, outcome prediction provided by ML models resulted slightly higher in the ML model than in logistic stepwise regression, however, without reaching a statistical difference.

Using MMoE, a cross-validated AUC of 0.779 ± 0.016 was obtained for the prediction of all-cause death in the overall population; when restricting to DOAC and VKA patients, the AUC resulted 0.733 ± 0.041 and 0.759 ± 0.026 , respectively. AUC for the prediction of CV death in the overall population resulted in 0.741 ± 0.029 , while the AUC obtained for DOAC and VKA patients were 0.733 ± 0.034 and 0.729 ± 0.020 , respectively. For major bleeding prediction, an AUC of 0.641 ± 0.020 was reached in the overall cohort, while an AUC of 0.711 ± 0.029 for DOAC patients and $0.630 \pm$

0.037 for VKA patients was obtained. The Naïve multi-task model predicted stroke occurrence with an AUC of 0.606 ± 0.117 , 0.566 ± 0.094 and 0.583 ± 0.047 , respectively, for the overall, VKA, and DOAC populations.

All-cause death	DOAC	0.709±0.030	0.691±0.035	0.733±0.041	0.723±0.032
All-cause death	VKA	0.74 ±0.021	0.704±0.046	0.759±0.026	0.707±0.028
CV death	Overall population	0.705±0.037	0.702±0.034	0.745±0.022	0.735±0.014
CV death	DOAC	0.649±0.072	0.658±0.041	0.733±0.034	0.661±0.050
CV death	VKA	0.677±0.074	0.671±0.043	0.729±0.020	0.695±0.051
Major bleeding	Overall population	0.619±0.031	0.592±0.026	0.641±0.020	0.628±0.013
Major bleeding	DOAC	0.711±0.029	0.672±0.045	0.689±0.069	0.685±0.041
Major bleeding	VKA	0.625±0.041	0.594±0.012	0.630±0.037	0.610±0.022
Stroke	Overall population	0.559±0.055	0.583±0.047	0.559±0.061	0.559±0.055
Stroke	DOAC	0.606±0.085	0.581±0.082	0.600±0.074	0.575±0.015
Stroke	VKA	0.513±0.056	0.566±0.094	0.557±0.083	0.512±0.131

Table 3. Predictive model results: the best predicting model is evidenced in bold. DOAC: direct oral anticoagulant; VKA: Vitamin K antagonists

Outcome	Population	Best ML Model	HASBLED	p-value	CHA ₂ DSVA ₂ SC	p-value
All-cause death	Overall population	0.779 ± 0.016	0.589 ± 0.028	< 0.001	0.629 ± 0.026	< 0.001
All-cause death	DOAC	0.733 ± 0.041	0.566 ± 0.062	< 0.001	0.641 ± 0.051	< 0.001
All-cause death	VKA	0.759 ± 0.026	0.608 ± 0.007	< 0.001	0.635 ± 0.020	< 0.001
CV death	Overall population	0.745 ± 0.022	0.596 ± 0.067	0.822	0.650 ± 0.019	0.787
CV death	DOAC	0.733 ± 0.034	0.557 ± 0.071	0.655	0.650 ± 0.020	0.298
CV death	VKA	0.729 ± 0.020	0.621 ± 0.050	0.153	0.654 ± 0.059	0.270
Major bleeding	Overall population	0.641 ± 0.020	0.576 ± 0.010	0.352	0.570 ± 0.030	0.243
Major bleeding	DOAC	0.711 ± 0.029	0.586 ± 0.054	< 0.001	0.605 ± 0.041	< 0.001
Major bleeding	VKA	0.630 ± 0.037	0.570 ± 0.034	0.831	0.540 ± 0.039	0.435
Stroke	Overall population	0.583 ± 0.047	0.592 ± 0.056	0.343	0.595 ± 0.041	0.282

Stroke	DOAC	0.606 ± 0.085	0.614 ± 0.089	0.525	0.623 ± 0.105	0.630
Stroke	VKA	0.566 ± 0.094	0.580 ± 0.093	0.414	0.536 ± 0.076	0.701

Table 4: Comparison between best ML predicting model and HAS-BLED or CHA₂DS₂VASC

Predictive risk model comparison

The comparison between the best ML-performing model and the clinical risk score HAS-BLED CHA₂DS₂VASC is shown in **Table 4** and **Figure 1**. ML outperformed HAS-BLED and CHA₂DS₂VASC in all-cause death prediction in all populations considered ($p < 0.001$, **Figure 1 A**). Moreover, ML improved major bleeding prediction in the DOAC population compared to HAS-BLED (0.711 Vs 0.586, $p < 0.001$, **Figure 1 C**). Despite numerically higher results, ML did not reach a statistically significant improvement for CV death or major bleeding prediction in all the populations considered or in the overall and VKA populations, respectively. No statistically significant differences were found in ischemic stroke prediction between ML and other risk scores.

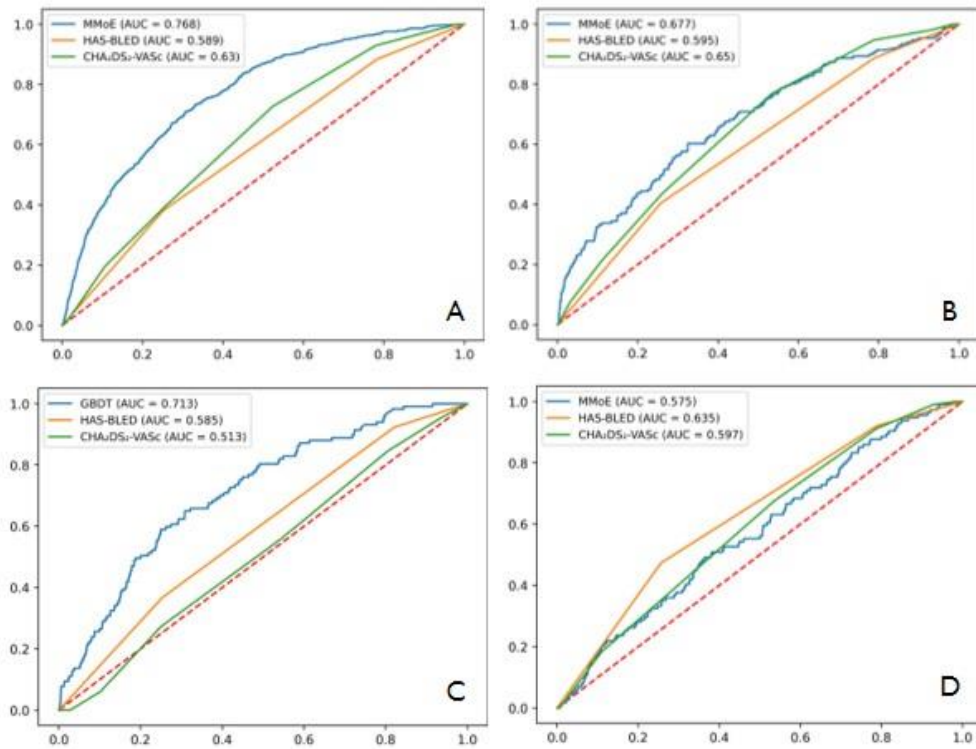


Figure 1: ROC comparison for all-cause death in the overall population (A), CV death in the overall population (B), major bleeding in DOAC patients (C) and ischemic stroke in the overall population (D) between the best-predicting ML model, CHA₂DS₂VASC and HAS-BLED scores.

Ranking Risk Predictors

The list of the most relevant features of the GBDT model in the prediction of major bleeding in the DOAC population is shown in **Figure 2 A**. The list of relevant features for the GBDT model for the other outcomes and populations is shown in **Figure 1 and 2, Supplementary Material**. The most important features for major bleeding in the DOAC population were hemoglobin levels, platelet count, BMI value, estimated GFR, age, red cell blood count, creatinine level and a history of previous major bleeding. With the exclusion of the previous major bleeding and apart from minor differences in

ranked feature importance, the risk predictors were similar for each outcome in all populations considered. SHAP algorithm (**Figure 2B**) selected the same aforementioned features as the more impactful in predicting major bleeding. In particular, the algorithm evidenced a linear inverse relationship (lower the value, higher the impact on the outcome prediction) between major bleeding and hemoglobin levels, platelet number, and BMI; contrarily, a linear direct relationship was evidenced with age and the history of previous major bleeding. On the other hand, the renal function, represented by GFR and creatinine levels, contributed in a non-linear relationship predicting major bleeding.

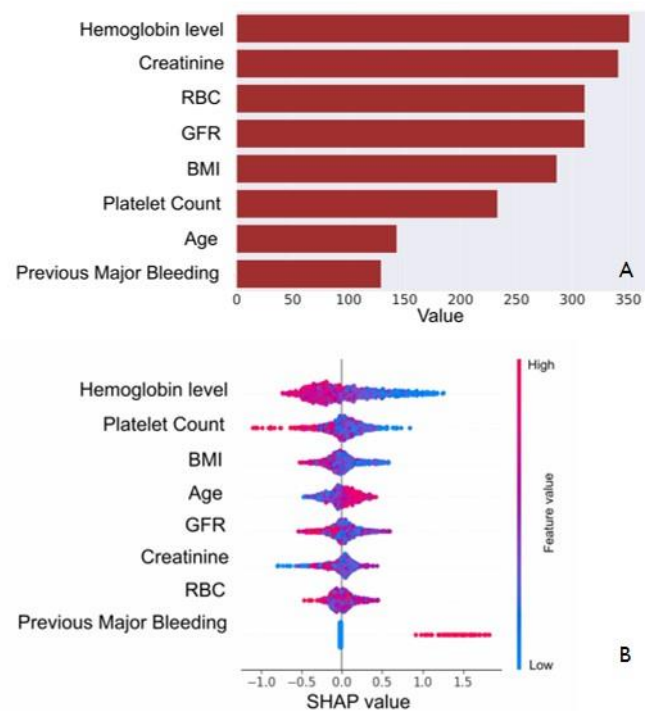


Figure 2: (A) Most relevant features for the prediction of major bleeding events in the DOAC population in the GBDT model as the best predicting ML model for this item. (B) SHAP values for GBDT model in the prediction of Major Bleeding in the DOAC population. RBC: Red Blood Cell Count; BMI: Body Mass Index; GFR: Glomerular Filtration Rate

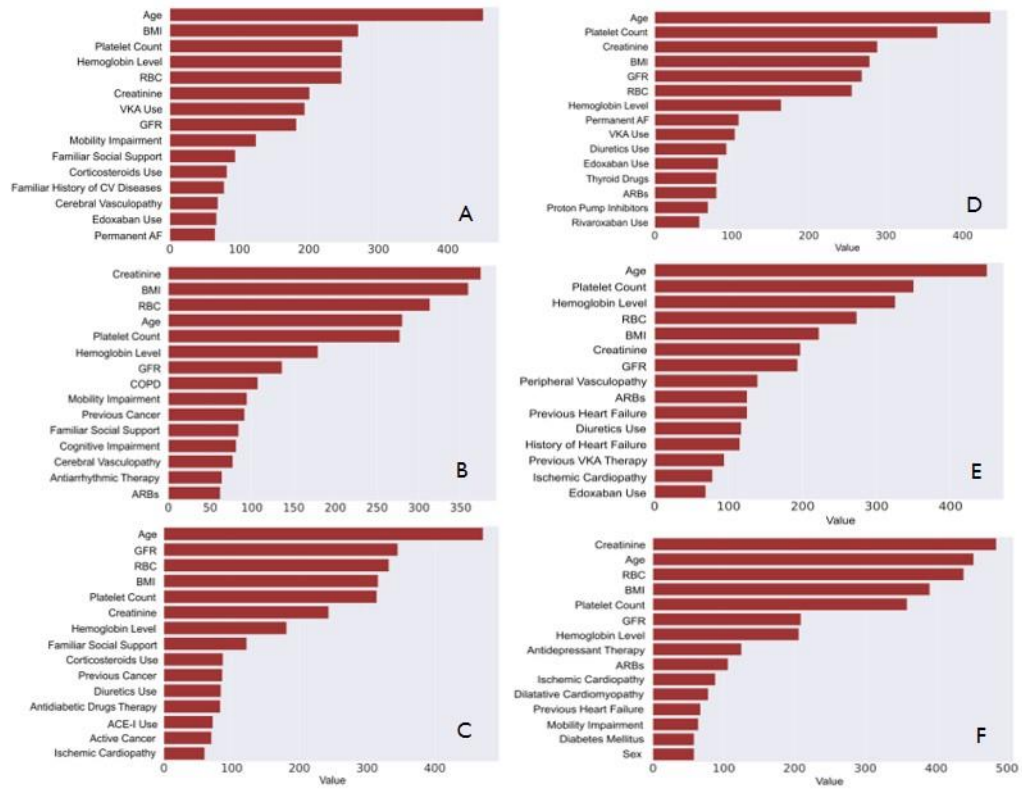


Figure 1 Supplemental: Most relevant features for the prediction of all-cause death and cardiovascular (CV) death in the GBDT model as the best predicting ML model. **A:** all-cause death in global population; **B:** all-cause death in DOAC population; **C:** all-cause death in VKA population; **D:** CV death in global population; **E:** CV death in DOAC population; **F:** CV death in VKA population

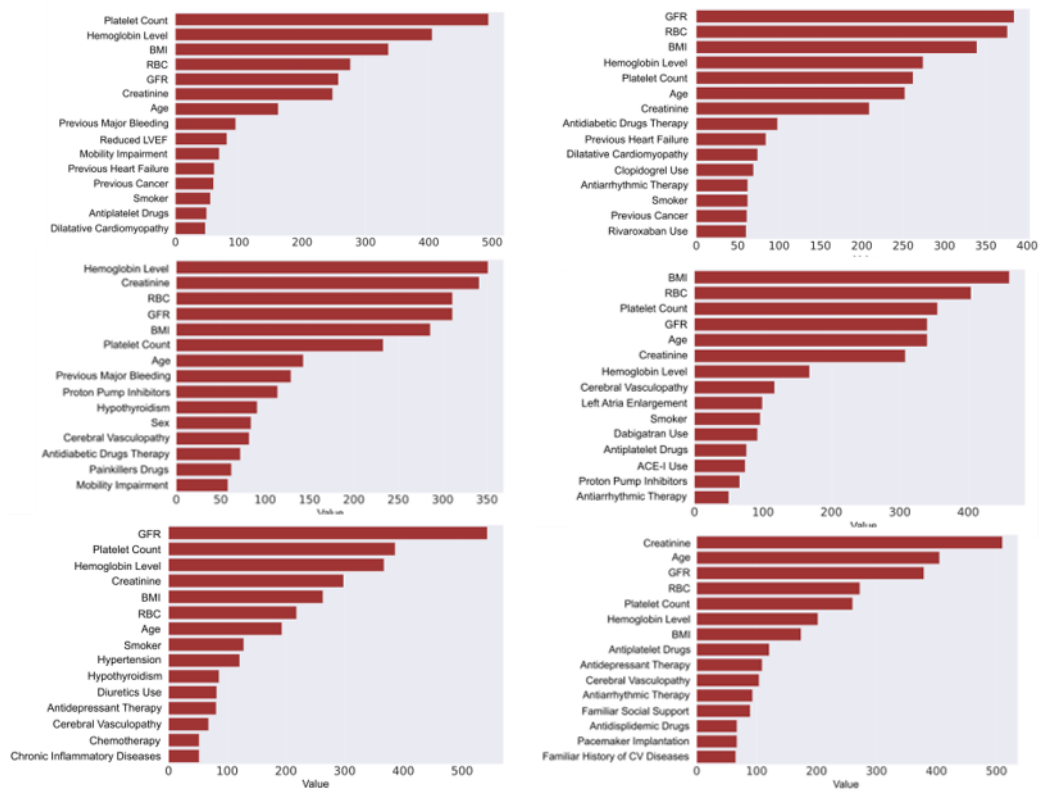


Figure 2 Supplemental: Most relevant features for the prediction of major bleeding and ischemic stroke in the GBDT model as the best predicting ML model. **A:** major bleeding in global population; **B:** major bleeding in DOAC population; **C:** major bleeding in VKA population; **D:** stroke in global population; **E:** stroke in DOAC population; **F:** stroke in VKA population. To be noted that when GBDT model was poorly predictive ($AUC < 0.7$), as for ischemic stroke prediction, the list of features is not very reliable because of the non-predictability of the model itself.

DISCUSSION

To our knowledge, this is the first study to use ML approach in an AF population entirely on oral anticoagulation to predict all-cause death, CV death, major bleeding and stroke. Data were obtained by a large, multicenter, real-world registry with a long-duration follow-up. Clinical outcomes were also evaluated in patients receiving different therapeutic approaches (DOACs and VKA), providing a different risk prediction in the two populations. The main findings can be recapitulated as follows:

- ML improved the prediction of all-cause mortality in all populations considered as compared to available risk scores and improved the prediction of major bleeding in the DOAC cohort as compared to HAS-BLED.
- The GBDT model selected important features for outcome prediction variables, such as hemoglobin, platelet count, and BMI, that are not included in available risk scores.

Anticoagulation represents the cornerstone of therapy in AF patients, with an overall anticoagulation rate that increased over the past years¹⁴, a decline in VKA prescription and a progressive increase in DOACs prescription. DOACs were prescribed in more than 95% of patients in our cohort during the last year of enrolment¹⁶. Our study, contrarily to the previous ones already published, provides results updated also to a contemporary population represented mainly by anticoagulated patients with DOACs. As anticoagulated AF patients represent the majority of AF patients¹⁴, a targeted ML analysis for event prediction in this setting is crucial to improve the ability to identify patients with a higher risk, allowing a better clinical stratification.

AF is associated with an increased risk of mortality³⁰. In recent studies, ML has provided an effective tool to predict this outcome with consistent results in AF patients, with a c-index ranging from 0.61 at 1 year derived from three measurements of PT-INR taken over the first 30 days after starting VKA in patients from the GARFIELD-AF registry¹³, to 0.762 in two analyses of the GARFIELD-AF and ORBIT-AF registries, 0.765 in an Australian study, and up to 0.785 in a French cohort of newly diagnosed AF patients admitted to Hospitals³¹. In our study, the discriminative ability was good with an AUC for all-cause death prediction of 0.779, a result that is consistent with the aforementioned studies, with the differences that are extrapolated from a different setting (patients not necessarily admitted to Hospitals) and clinical situation (patients completely on oral anticoagulation). Despite differences in prevalence between studies, ranging from 64% of all deaths in a metanalysis of randomized trials³² to nearly 38% in real-world registries³³, CV death represents an important cause of mortality in AF patients. For these reasons, we provided a risk prediction for CV death with an AUC of 0.741, which is consistent with a recent finding of a ML prediction of 0.724 in a large cohort derived from a French database that, however, did not provide any information regarding the use of anticoagulation therapy. The lower rate of CV death (19% vs 31.5%) compared to the French study, could explain the slightly lower prediction performance in both ML systems in comparison with all-cause death. START-Register enrollment centers are anticoagulation clinics that closely monitored AF patients and although a relevant percentage of patients were enrolled after a hospital discharge, the more critical patients could not have been referred to the anticoagulation clinics: these aspects,

added to the fact that oral anticoagulation *per se* reduces the rate of CV-death, may explain the lower rate of events in our patients.

Anticoagulation reduces the risk of thromboembolic events in AF patients⁶: as a consequence, the number of thromboembolic events and, in particular, of ischemic stroke resulted very low in the START-AF registry, with a rate $\times 100$ pt-yrs of 0.37 and 0.23, respectively. This aspect mainly contributed to the suboptimal stroke prediction in our study, that was not different from that obtained with the CHA₂DS₂VASC score. Of note, our performance was comparable with previous ML-derived results, that also reported a C statistic ≤ 0.70 for stroke prediction^{7,9,13}. As in previous reports, our model showed a better performance in predicting all-cause mortality than stroke or bleeding. A possible explanation of this finding could be that since the risk factors for stroke, hemorrhage, and mortality overlap significantly, many of the items recorded in registries are linked to a higher risk of all three outcomes. Nevertheless, models of stroke and hemorrhage should consider the competing risk of death from other causes, whereas models of all-cause mortality do not⁷.

Regarding the different ML models used, MMoE outperformed the other methods in predicting all-cause and CV death among all the populations considered. Unlike other naïve multi-task learning models, MMoE introduces greater diversity among experts, creating representations that are better suited to highly unbalanced and heterogeneous multi-task learning. This allows the model to better capture shared patterns that might be ignored by independent training, as MMoE explicitly learns to model task relationships from data, having to share a part of the architecture for the prediction of different targets. Despite in a previous work stepwise logistic regression models performed as well or better than ML models for outcome prediction⁷, in our study ML

techniques provided slightly better results (although without significant differences) than stepwise logistic regression. The differences in the number and type of variables included in the datasets could have influenced the capability of more complex models to better capture non-linear relationships and interactions between variables.

The most commonly used scores in clinic practice in AF patients are represented by CHA₂DS₂-VASc and HAS-BLED. Although they were created respectively to predict the thromboembolic risk before the introduction of antithrombotic therapy and the major bleeding risk amongst patients on anticoagulation, these scores have been widely used beyond their original purpose³⁴. Due to the overlapping of items considered, both these scores resulted in reasonably good predictors even of stroke, major bleeding, and mortality in AF patients³⁵. Anyway, these scores are not dedicated to the prediction of death and their predictive ability remains modest³⁶. Our ML approach allowed us to provide better accuracy of prediction regarding all-cause death in all populations considered and of major bleeding in the DOAC population. Also if recommended by the current guidelines, our study confirmed the HAS-BLED score poor major bleeding predictive ability in the DOAC population (AUC 0.586 ± 0.054), providing a better tool to stratify the hemorrhagic risk in a contemporary AF cohort of patients. Moreover, it is conceivable that all cause-death is related to factors not included in the current stratification risk scores, which ML is more likely to reveal. The better prediction of death and major bleeding could allow the individuation of high-risk patients, filling an unmet need for an individualized patient management strategy³⁷.

ML approach could help to uncover additional new potential risk factors of clinical outcomes in addition to those utilized in the classic clinical risk score, derived from

standard statistical methods. In addition to age and renal function, which are items already considered in CHA₂DS₂VASC and HAS-BLED scores, our study identified other risk factors for worse outcomes in AF patients, such as anemia (represented by hemoglobin level and red blood count), platelet count and BMI.

Anemia is a common finding in AF patients and a well-known risk factor for bleeding, cardiac events and all-cause mortality³⁸. Anemia could contribute to the development of new-onset AF³⁹, or to increase the recurrences of paroxysmal AF⁴⁰. Moreover, anemia could represent a marker of frailty, another known risk factor for a worse outcome in AF patients⁴¹.

Thrombocytopenia is also associated with an increased risk of mortality and bleeding in AF patients⁴². This aspect has to be considered in the case of AF patients who require anticoagulation, as starting anticoagulation increases the number of major bleeding events in patients with low platelet counts compared to patients without thrombocytopenia⁴³. Moreover, a low platelet count represents an established risk factor for bleeding in patients with AF undergoing PCI, regardless of clinical presentation⁴⁴. The SHAP algorithm (**Figure 2B**) in our study showed a reverse-linear association between the platelet count and the bleeding risk in DOAC patients, with a directional impact of the feature towards the prediction of an increased bleeding risk together with a lowering of the platelet count (as observed for hemoglobin level).

BMI is the most frequently used anthropometric measure of adiposity. Although obesity has a key role in the incidence and progression of AF, its impact on prognosis is still controversial, with many studies stressing that obesity is not associated with an increase in adverse events, in particular regarding stroke, death, or CV death⁴⁵, providing higher protection in those with established CV disease⁴⁶. This phenomenon

is also known as “obesity paradox” and seems to be applicable also to major bleedings⁴⁷. On the other hand, being underweight could be associated with an increased bleeding risk⁴⁸, systemic embolism, cardiovascular death and all-cause death⁴⁵. A recent paper from the START registry found that DOACs seem to be safe for treatment in extremely overweight or underweight AF patients⁴⁹, however this study included an heterogeneous cohort of AF and venous thromboembolism patients, younger than those included in our study, moreover considering only the extreme sides of the BMI in the population. Our ML analysis demonstrates that BMI is one of the most important variables for feature importance in all populations, with a trend for an increase in the risk of major bleeding associated with a decrease in BMI, in particular in the DOAC population.

Therefore, our analysis suggests that adjunctive risk factors, which are currently not used for the prediction of death, bleeding and thromboembolic risk in AF, should be taken into account in the management of these AF patients. Actual clinical scores may not have captured all relevant risk predictors: adding different variables beyond the overlapping items of the currently available scores could improve prediction accuracy. When applied to semi-structured data, machine learning techniques have produced remarkable predictive performance⁵⁰. The incorporation of unstructured data into existing data repositories, allowing ML to reveal some additional non-linear relationships, would enhance the predictive accuracy of future models⁷.

Limitations

We acknowledge that the study has some limitations. Outcome events were adjudicated by the local investigators. Despite being a large, multicenter registry, all

centers of AF START-2 are from a single European country; therefore, our results should be confirmed in other populations. Our results lack for an external validation cohort, so extrapolation to general practice should be done cautiously. However, the consistency of our results with other recent ML studies could indirectly confirm their validity. Although it has numerous items, the database may not have captured all relevant risk predictors: the dataset is in prevalence composed of categorical features and did not include more information regarding other biomarkers (such as Troponin level, B natriuretic peptide) or imaging features. Nevertheless, if on one hand the increased number of more unstructured data could improve algorithm performance, on the other hands the addition of not routinely performed exams could limit algorithm's reproducibility and increase its complexity. The study provided a long follow-up of patients (up to 9.4 years): the prediction of future clinical events based on a single time point remains challenging.

CONCLUSIONS

In a large cohort of patients with AF treated with oral anticoagulation, ML models demonstrated good discriminative ability to predict all-cause death and CV death in the overall population, and major bleeding in DOAC patients, outperforming the CHA₂DS₂VASC and HAS-BLED scores. Hemoglobin values, platelet count, and BMI emerged as new potential risk predictors in anticoagulated AF patients. The applications of ML prediction models in clinical practice could improve the risk assessment and subsequent management of anticoagulated patients with AF.

REFERENCES

1. Hindricks, G. *et al.* 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS). *Eur. Heart J.* **42**, 373–498 (2021).
2. Benjamin, E. J. *et al.* Heart Disease and Stroke Statistics-2019 Update: A Report From the American Heart Association. *Circulation* **139**, e56–e528 (2019).
3. Ageno, W. & Donadini, M. Breadth of complications of long-term oral anticoagulant care. *Hematol. Am. Soc. Hematol. Educ. Program* **2018**, 432–438 (2018).
4. Borre, E. D. *et al.* Predicting Thromboembolic and Bleeding Event Risk in Patients with Non-Valvular Atrial Fibrillation: A Systematic Review. *Thromb. Haemost.* **118**, 2171–2187 (2018).
5. Garcia, C. *et al.* Admission Severity of Atrial-Fibrillation-Related Acute Ischemic Stroke in Patients under Anticoagulation Treatment: A Systematic Review and Meta-Analysis. *J. Clin. Med.* **11**, 3563 (2022).
6. Hart, R. G., Pearce, L. A. & Aguilar, M. I. Meta-analysis: Antithrombotic Therapy to Prevent Stroke in Patients Who Have Nonvalvular Atrial Fibrillation. *Ann. Intern. Med.* **146**, 857–867 (2007).
7. Loring, Z. *et al.* Machine learning does not improve upon traditional regression in predicting outcomes in atrial fibrillation: an analysis of the ORBIT-AF and GARFIELD-AF registries. *EP Eur.* **22**, 1635–1644 (2020).
8. Olier, I., Ortega-Martorell, S., Pieroni, M. & Lip, G. Y. H. How machine learning is impacting research in atrial fibrillation: implications for risk prediction and future management. *Cardiovasc. Res.* **117**, 1700–1717 (2021).
9. Lu, J. *et al.* Performance of multilabel machine learning models and risk stratification schemas for predicting stroke and bleeding risk in patients with non-valvular atrial fibrillation. *Comput. Biol. Med.* **150**, 106126 (2022).
10. Lip, G. Y. H. *et al.* Improving dynamic stroke risk prediction in non-anticoagulated patients with and without atrial fibrillation: comparing common clinical risk scores and machine learning algorithms. *Eur. Heart J. Qual. Care Clin. Outcomes* **8**, 548–556 (2022).
11. Li, X. *et al.* Integrated Machine Learning Approaches for Predicting Ischemic Stroke and Thromboembolism in Atrial Fibrillation. *AMIA Annu. Symp. Proc. AMIA Symp.* **2016**, 799–807 (2016).
12. Steinberg, B. A. *et al.* International trends in clinical characteristics and oral anticoagulation treatment for patients with atrial fibrillation: Results from the

- GARFIELD-AF, ORBIT-AF I, and ORBIT-AF II registries. *Am. Heart J.* **194**, 132–140 (2017).
13. Goto, S. *et al.* New artificial intelligence prediction model using serial prothrombin time international normalized ratio measurements in atrial fibrillation patients on vitamin K antagonists: GARFIELD-AF. *Eur. Heart J. - Cardiovasc. Pharmacother.* **6**, 301–309 (2020).
 14. Navar, A. M. *et al.* Trends in Oral Anticoagulant Use Among 436 864 Patients With Atrial Fibrillation in Community Practice, 2011 to 2020. *J. Am. Heart Assoc.* **11**, e026723 (2022).
 15. Antonucci, E. *et al.* The Italian START-Register on Anticoagulation with Focus on Atrial Fibrillation. *PloS One* **10**, e0124719 (2015).
 16. Poli, D. *et al.* Management of anticoagulation in atrial fibrillation patients in Italy: insight from the *Atrial Fibrillation-Survey on Anticoagulated Patients Register* (AF-START). *Bleeding Thromb. Vasc. Biol.* **2**, (2023).
 17. Lip, G. Y. H., Nieuwlaat, R., Pisters, R., Lane, D. A. & Crijns, H. J. G. M. Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: the euro heart survey on atrial fibrillation. *Chest* **137**, 263–272 (2010).
 18. Pisters, R. *et al.* A novel user-friendly score (HAS-BLED) to assess 1-year risk of major bleeding in patients with atrial fibrillation: the Euro Heart Survey. *Chest* **138**, 1093–1100 (2010).
 19. Kaatz, S., Ahmad, D., Spyropoulos, A. C., Schulman, S., & Subcommittee on Control of Anticoagulation. Definition of clinically relevant non-major bleeding in studies of anticoagulants in atrial fibrillation and venous thromboembolic disease in non-surgical patients: communication from the SSC of the ISTH. *J. Thromb. Haemost. JTH* **13**, 2119–2126 (2015).
 20. Hicks, K. A. *et al.* 2017 Cardiovascular and Stroke Endpoint Definitions for Clinical Trials. *Circulation* **137**, 961–972 (2018).
 21. Sudden cardiac death. Report of a WHO Scientific Group. *World Health Organ. Tech. Rep. Ser.* **726**, 5–25 (1985).
 22. Ke, G. *et al.* LightGBM: A Highly Efficient Gradient Boosting Decision Tree. in (2017).
 23. Lundberg, S. M. & Lee, S.-I. A Unified Approach to Interpreting Model Predictions. in *Advances in Neural Information Processing Systems* vol. 30 (Curran Associates, Inc., 2017).

24. Caruana, R. Learning Many Related Tasks at the Same Time with Backpropagation. in *Advances in Neural Information Processing Systems* vol. 7 (MIT Press, 1994).
25. Ma, J. *et al.* Modeling Task Relationships in Multi-task Learning with Multi-gate Mixture-of-Experts. in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining 1930–1939* (Association for Computing Machinery, New York, NY, USA, 2018). doi:10.1145/3219819.3220007.
26. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. Preprint at <https://doi.org/10.48550/arXiv.1412.6980> (2017).
27. Sechidis, K., Tsoumakas, G. & Vlahavas, I. On the Stratification of Multi-label Data. in *Machine Learning and Knowledge Discovery in Databases* (eds. Gunopulos, D., Hofmann, T., Malerba, D. & Vazirgiannis, M.) vol. 6913 145–158 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2011).
28. Hogan, J. & Adams, N. M. On Averaging ROC Curves. *Trans. Mach. Learn. Res.* (2023).
29. DeLong, E. R., DeLong, D. M. & Clarke-Pearson, D. L. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* **44**, 837–845 (1988).
30. Odutayo, A. *et al.* Atrial fibrillation and risks of cardiovascular disease, renal disease, and death: systematic review and meta-analysis. *BMJ* **354**, i4482 (2016).
31. Bisson, A. *et al.* Prediction of early death after atrial fibrillation diagnosis using a machine learning approach: A French nationwide cohort study. *Am. Heart J.* S0002-8703(23)00203-X (2023) doi:10.1016/j.ahj.2023.08.006.
32. Gómez-Outes, A. *et al.* Causes of Death in Anticoagulated Patients With Atrial Fibrillation. *J. Am. Coll. Cardiol.* **68**, 2508–2521 (2016).
33. Romiti, G. F. *et al.* Clinical Complexity Domains, Anticoagulation, and Outcomes in Patients with Atrial Fibrillation: A Report from the GLORIA-AF Registry Phase II and III. *Thromb. Haemost.* **122**, 2030–2041 (2022).
34. Melgaard, L. *et al.* Assessment of the CHA2DS2-VASc Score in Predicting Ischemic Stroke, Thromboembolism, and Death in Patients With Heart Failure With and Without Atrial Fibrillation. *JAMA* **314**, 1030–1038 (2015).
35. Poli, D., Antonucci, E., Pengo, V., Testa, S. & Palareti, G. Comparison of HAS-BLED and HAS-BED Versus CHADS2 and CHA2DS2VASc Stroke and Bleeding Scores in Patients With Atrial Fibrillation. *Am. J. Cardiol.* **119**, 1012–1016 (2017).
36. Proietti, M. *et al.* Association between clinical risk scores and mortality in atrial fibrillation: Systematic review and network meta-regression of 669,000 patients. *Eur. J. Prev. Cardiol.* **27**, 633–644 (2020).

37. D'Ascenzo, F. *et al.* Machine learning-based prediction of adverse events following an acute coronary syndrome (PRAISE): a modelling study of pooled datasets. *The Lancet* **397**, 199–207 (2021).
38. Bonde, A. N. *et al.* Oral anticoagulation among atrial fibrillation patients with anaemia: an observational cohort study. *Eur. Heart J.* **40**, 3782–3790 (2019).
39. Xu, D. Z. *et al.* Reduced kidney function and anemia as risk factors for new onset AF. *Eur. Heart J.* **34**, P4271 (2013).
40. Kim, M. *et al.* Clinical relationship between anemia and atrial fibrillation recurrence after catheter ablation without genetic background. *Int. J. Cardiol. Heart Vasc.* **27**, 100507 (2020).
41. Proietti, M. *et al.* Frailty prevalence and impact on outcomes in patients with atrial fibrillation: A systematic review and meta-analysis of 1,187,000 patients. *Ageing Res. Rev.* **79**, 101652 (2022).
42. Pastori, D. *et al.* Thrombocytopenia and Mortality Risk in Patients With Atrial Fibrillation: An Analysis From the START Registry. *J. Am. Heart Assoc.* **8**, e012596 (2019).
43. Iyengar, V. *et al.* Bleeding Risk in Atrial Fibrillation and Thrombocytopenia: A Propensity Matched Cohort Study. *Blood* **140**, 345–346 (2022).
44. Gragnano, F. *et al.* Impact of clinical presentation on bleeding risk after percutaneous coronary intervention and implications for the ARC-HBR definition. *EuroIntervention J. Eur. Collab. Work. Group Interv. Cardiol. Eur. Soc. Cardiol.* **17**, e898–e909 (2021).
45. Zhu, W. *et al.* Relation of Body Mass Index With Adverse Outcomes Among Patients With Atrial Fibrillation: A Meta-Analysis and Systematic Review. *J. Am. Heart Assoc.* **5**, e004006 (2016).
46. Horwich, T. B., Fonarow, G. C. & Clark, A. L. Obesity and the Obesity Paradox in Heart Failure. *Prog. Cardiovasc. Dis.* **61**, 151–156 (2018).
47. Proietti, M., Guiducci, E., Cheli, P. & Lip, G. Y. H. Is There an Obesity Paradox for Outcomes in Atrial Fibrillation? A Systematic Review and Meta-Analysis of Non-Vitamin K Antagonist Oral Anticoagulant Trials. *Stroke* **48**, 857–866 (2017).
48. Bertomeu-Gonzalez, V. *et al.* Association of Body Mass Index With Clinical Outcomes in Patients With Atrial Fibrillation: A Report From the FANTASIA Registry. *J. Am. Heart Assoc.* **9**, e013789 (2020).
49. Guarascio, M. *et al.* DOACs use in extreme body-weighted patients: results from the prospective START-register. *Intern. Emerg. Med.* (2023) doi:10.1007/s11739-023-03334-4.

50. Rajpurkar, P. & Lungren, M. P. The Current and Future State of AI Interpretation of Medical Images. *N. Engl. J. Med.* **388**, 1981–1990 (2023).