Prediction of Atrial Fibrillation Risks at Primary Care Level using Longitudinal Learning Stances

Henrique Anjos
henriquemdanjos@tecnico.ulisboa.com
Instituto Superior Técnico, Lisboa, Portugal

Abstract

Atrial fibrillation (AF) is the most prevalent cardiac arrhythmia worldwide and is strongly associated with increased risks of stroke, heart failure, and mortality. Traditional methods to detect and prognostic AF and its associated risks often fail to capture the full complexity of AF patterns, limiting their predictive accuracy. In spite of the improvements achieved by machine learning (ML) techniques, state-of-the-art AF-focused predictors do not generally incorporate longitudinal data, reducing their capacity to model the dynamic and evolving nature of individual behaviors and cardiophysiological indicators over time. The absence of a longitudinal perspective restricts understanding of how AF risk develops and changes across different prediction horizons. This study addresses these limitations by developing superior ML models tailored to predict adverse events within a longitudinal cohort of individuals with AF, while also laying the groundwork for future models that predict AF onset. Our work focuses on six critical clinical endpoints: stroke, all-cause death, cardiovascular death, heart failure hospitalizations, inpatient visits, and acute coronary syndrome. The ML models yielded an AUC of 0.65 for 1-year stroke prediction, outperforming CHA_2DS_2-VASc (0.59) and GARFIELD-AF (0.63). For all-cause mortality prediction, the models achieved an AUC of 0.78 against the 0.72 reference of GARFIELD-AF. In addition to predictive advances, the study identifies determinants of AF-related risks and introduces a prototype decision-support tool for clinical use. The work was conducted in collaboration with ULS Matosinhos, which reviewed and validated the findings.

Keywords

Atrial Fibrillation, Machine Learning, Electronic Health Records, Longitudinal Clinical Data, Medical Interface, Decision Support.

1 Introduction

AF is the most common cardiac arrhythmia worldwide, and its occurrence associated with a significant risk increase of stroke, heart failure, and mortality [5]. Studies estimate that around 9 million individuals in the European Union (EU) were affected by AF in 2010, and the number is projected to double by 2060 [18]. In Portugal, 4070 deaths were attributable to AF in 2010, corresponding to nearly 4% of all deaths [9]. Despite its prevalence and severe consequences, AF often goes undiagnosed until complications arise due to its episodic nature and the lack of consistent early-warning signs [4]. This diagnostic gap highlights the need for tools that can detect AF and prognosticate its complications effectively.

A plethora of traditional methods have been proposed for predicting AF and its complications that rely on point-based systems or traditional statistical models [14]. While these methods are functional, they often fail to capture the complexity of AF patterns [16]. In the last years, ML models have demonstrated superior performance compared to these traditional approaches [24]. However, the existing state-of-the-art ML methods generally suffer from two major drawbacks:

(1) failing to incorporate longitudinal data, which limits their ability to analyze the progression of biometric and cardiophysiological indicators across varying prediction horizons.
(2) underestimating the significance of integrating specific data modalities. Many approaches do not fully leverage the predictive value embedded in electrophysiological signals and neglect other critical factors such as risk behaviors, comorbidities, and drug regimen.

To address these challenges, this thesis lays the groundwork for developing a clinical decision-support tool designed to detect and prognosticate AF, as well as its associated clinical endpoints, using machine learning models specifically tailored for these tasks. This work is part of a broader initiative in collaboration with the Unidade Local de Saúde de Matosinhos (ULSM), which is expected to proceed in three phases: (1) prognostication of complications in the AF patient population; (2) Prediction of AF onset in a case-control population; (3) Incorporation of advanced modalities such as cardiac imaging and electrophysiological data. Given the data available at the time, this thesis focuses on phase 1, while also laying the groundwork for the following phases. It leverages a longitudinal cohort of AF patients, using electronic health records from ULSM collected over the past decades to derive population-specific insights. The work involves developing and optimizing machine learning algorithms to accurately predict AF-associated clinical endpoints, including stroke/systemic embolism (SE), all-cause mortality, cardiovascular death, heart failure hospitalizations, inpatient visits, and acute coronary syndrome. In parallel, a prototype clinical decision-support tool is designed, consisting of a backend API that provides programmatic access to the predictive models and a user-friendly graphical interface that allows healthcare professionals to input patient data and visualize predictions intuitively. The thesis also evaluates classical risk calculators, providing a comparative context, and establishes a foundation for future development and clinical application.

The main contributions of this thesis include the development of machine learning models for AF-related outcomes, the exploration of time-aware approaches from longitudinal data, systematic evaluation against classical risk scores, and the design of a prototype clinical decision-support tool.

2 Related Work

Predictive scores for stroke risk in AF patients have been developed since 2009 [22]. This section provides a brief overview of traditional

and machine learning approaches to predict complications among patients with pre-existing AF.

The $CHADS_2$ score is the primary stroke risk stratification scheme in patients with pre-existing AF and made part of the 2006 American College of Cardiology/American Heart Association/European Society of Cardiology (ACC/AHA/ESC) guidelines for nonvalvular AF [6]. This score assigns points based on the following risk factors: C - congestive heart failure (1 point), H - hypertension (1 point), A - age (1 point), D - diabetes mellitus (1 point), and S - prior stroke or transient ischemic attack (2 points). This score's simplicity made it widely adopted; however, it did not account for certain risk factors, which led to the development of more refined scores.

The Birmingham 2009 schema, also known as CHA_2DS_2-VASc score [19] emerged as an improvement over $CHADS_2$ and has been widely used since 2010 [25]. It provides a more comprehensive assessment of stroke risk in AF patients by incorporating additional risk factors, such as: V - prior vascular disease, A - age between 65 and 74 years, and Sc - sex category. This refinement significantly improved risk stratification, especially for patients at lower or intermediate risk [17].

In 2010, user-friendly HAS-BLED score was developed to assess 1-year risk of major bleeding in patients with AF, using data from the Euro Heart Survey on AF. The risk score gives points based on the following comorbidities: Hypertension, Abnormal renal/liver function, Stroke, Bleeding history or predisposition, Labile international normalized ratio, Elderly (> 65 years), Drugs/alcohol concomitantly. The bleeding score achieved an AUC of 0.72 in the derivation cohort of AF patients with extensive use of oral anticoagulation and was consistent when applied in several subgroups.

In 2012, HAS-BLED bleeding risk score was experimented in predicting cardiovascular events and mortality in anticoagulated patients with AF, and was shown to be useful. However, a multivariate analysis was slightly better at predicting cardiovascular events and mortality, showing that HAS-BLED is not the most accurate score to predict these events[7].

In 2016, ABC (age, biomarkers, clinical history) stroke risk score was developed to predict 6-year stroke risk in AF patients with a Cox regression model. The ABC-Stroke score achieved higher AUCs than CHA_2DS_2-VASc in both the derivation cohort (0.68 vs. 0.62) and the external validation cohort (0.66 vs. 0.58). The score used biomarkers such as Troponin I and NT-proBNP[11]. Later in 2016, the ABC score was also developed to predict bleeding risk score in patients with AF, and also achieved similar results[12]. Moreover, in 2017, ABC was also developed to predict death in anticoagulated patients with AF, including both clinical information and biomarkers. The model yielded higher AUC than a model based on all clinical variables, both the derivation (0.74 vs. 0.68) and validation cohorts (0.74 vs. 0.67)[13]. Later in 2021, ABC-AF was also validated in patients not receiving oral anticoagulation and yielded similar positive results[2].

In 2020, GARFIELD-AF risk model was developed to predict stroke, major bleeding or mortality in AF patients. The model is described as developed using sophisticated statistical modeling techniques, and can be found on https://af.garfieldregistry.org/garfield-af-risk-calculator. It achieved a higher score in stroke and mortality prediction compared with CHA_2DS_2-VASc and a higher score than HAS-BLED score in predicting major bleed events across all risk groups [1]. The model was then externally validated in ORBIT-AF cohort, and the discriminatory value was still superior than CHA_2DS_2-VASc score.

In 2022, the Fushumi AF registry — a cohort from Fushumi, Japan — was used to learn predictors of ischemic events in patients with AF. Apart from biological and past medical and treatment history, the model also used medication, blood test, and echocardiogram data. The model was able to have an increased performance compared to CHA_2DS_2-VASc with an AUC of 0.72 vs 0.62 [20].

Also in 2022, using the same Fushimi AF registry, another machine learning risk model was developed to predict incident heart failure in patients with AF. The model outperformed Framingham risk score with an AUC of 0.75 vs 0.67 [10]. However, it is to note that the Framingham risk score is not specifically tailored to predict heart failure events in an AF specific population.

In 2024, another model was developed to predict 1-year stroke risk in a South Asian, Indian population, and achieved an AUC of 0.82 compared with CHA_2DS_2-VASc 0.67. However, when externally validated in another Asian cohort, the model only achieved an AUC of 0.67 while CHA_2DS_2-VASc achieved 0.62 [3].

In 2024, a meta-analysis evaluated 13 studies on stroke prediction in patients with atrial fibrillation using machine learning methods. The mean AUC across studies was 0.73, with models such as XGBoost and logistic regression achieving higher performance, while neural networks showed comparatively lower AUC values [8].

Within the prediction of AF-related outcomes, stroke is the most extensively studied, given the substantial increase in stroke risk associated with AF. Other predictive models have focused on outcomes such as bleeding or all-cause mortality, while less attention has been given to events like acute coronary syndrome, cardiovascular death, and hospitalizations.

3 Dataset

This section provides a description of the dataset acquired for this work. The data comes from anonymised electronic health records (EHRs) of patients followed at the Unidade Local de Saúde de Matosinhos (ULSM). The dataset was specifically designed within the scope of preventive assessment initiatives led by ULSM, with a focus on the prevention of AF at the primary care level. Importantly, the cohort study was purposefully built by ULSM to address the specific objectives of this work. The section also provides an overview of the monitored variables and an exploratory data analysis, highlighting descriptive statistics and cohort-level patterns.

3.1 Data Description

The available dataset spans approximately 25 years and includes patients over 40 years old who were diagnosed with AF between 1 January 2012 and 31 December 2021. The dataset comprises a population of 7,203, and 167 features encompassing demographics, clinical records (pertaining to laboratory, pharmaceutical, and surgical acts), and the temporal context of the previous electronic registry.

This study was approved by the Ethical Committee and Data Protection Officer of ULSM. The original data was de-identified according to the HIPAA Safe Harbour Method with noise added to all variables.

In Table 1, we present some key features of the dataset. In addition to the features shown in the table, there are several binary features representing comorbidities that are not listed. These include: chronic obstructive pulmonary disease, myocardial infarction or unstable angina, type 1 diabetes mellitus, type 2 diabetes mellitus, and valvular heart disease. Additionally, there are several outcomes: acute coronary syndrome, arterial embolism, stroke, inpatient visit, heart failure hospitalization, cardiovascular death, and all-cause death.

Table 1: Some selected variables of the USLM dataset. Binary features representing comorbidities and dispensed medications are not listed for simplicity's sake.

Feature Name	Description	Data Type	Units
index_date	index date (random)	integer	N/A
age_hipaa	age (HIPAA compliant)	categorical	years
female	female	binary	N/A
wgt	weight	float	kg
hgt	height	integer	cm
bmi	body mass index	float	kg/m^2
sbp	systolic blood pressure	float	mmHg
dbp	diastolic blood pressure	float	mmHg
tc	total cholesterol	float	mg/dL
ldl	LDL cholesterol	float	mg/dL
hdl	HDL cholesterol	float	mg/dL
glc	glycemia	float	mg/dL
tg	triglycerides	float	mg/dL
crt	creatinine	float	mg/dL
egfr	estimated glomerular filtration rate	float	$mL/min/1.73m^2$
tsh	thyroid stimulating hormone	float	mIU/L
uacr	urine albumin-to-creatinine ratio	float	mg/g
a1c	Hba1c (glycated hemoglobin)	float	mmol/mol
cancer	cancer	binary	N/A
carotd	carotid disease	binary	N/A
cd	coronary disease	binary	N/A
dlp	dyslipidemia	binary	N/A
esrd	end stage renal disease	binary	N/A
flt	flutter	binary	N/A
hf	heart failure	binary	N/A
hta	hypertension	binary	N/A
slap	sleep apnea	binary	N/A
stk	stroke	binary	N/A
tyrd	thyroid disease	binary	N/A

The dataset also includes binary variables indicating whether a patient is currently on any of the following medications: antiarrhythmics, anticoagulants, angiotensin-aonverting anzyme anhibitor (ACEi), angiotensin receptor brlockers (ARBs), angiotensin receptor-neprilysin inhibitor (ARNi), antiplatelets, beta blockers, calcium channel blockers, digoxin, dipeptidyl peptidase-4 inhibitor (DPP4i), GLP-1 agonists, insulin, ivabradine, loop diuretics, other diuretics, metformin, mineralocorticoid receptor antagonist (MRA), nitrates, sodium-glucose cotransporter-2 inhibitors (SGLT2i), statins, and sulfonylurea. Moreover, the data contains additional binary variables that represent whether a patient had the following interventions before: cardiac surgery, cardiac device, coronary surgery, and percutaneous coronary intervention. To represent smoking, the dataset has a set of binary features: current smoker, former smoker, never smoked, and no information smoker. Some laboratory tests also include an additional feature representing the patient's initial measurement for that specific exam. These tests are: HbA1c, creatinine, estimated glomerular filtration rate, glycemia, HDL cholesterol, LDL cholesterol, total cholesterol, triglycerides, thyroid simulating hormone, urine albumin-to-creatinine ratio, and international normalized ratio (INR).

Every feature—except for index date, age (HIPAA), and sex (female)—has an associated integer day count indicating when the event occurred relative to the AF diagnosis. For comorbidities, the index date defines the diagnosis data. For laboratory tests, it indicates the date of the measurement. For biometric features, it corresponds to the date the information was extracted, while those for medications refer to the last prescription.

3.2 Exploratory Data Analysis (EDA) of the AF Cohort

To provide a clearer understanding of the dataset and its key variables, descriptive statistics for selected numerical variables—including mean, standard deviation, and proportion of missing values—are presented in Table 2. As shown, LDL cholesterol, HbA1c, TSH, UACR, and eGFR have a relatively high proportion of missing values (>20%), which may impact subsequent analyses. In contrast, variables age, weight, systolic, and diastolic blood pressure have negligible missingness (<1%). Overall, the mean and standard deviation values indicate moderate variability for most anthropometric and biochemical measures. Triglycerides, glycemia, and UACR display relatively large standard deviations, suggesting substantial variability across participants.

Table 2: Descriptive statistics of selected numerical variables: mean and missing rate (after outlier removal)

Variable	Value (mean ± SD)	Missing Values (%)
Age*	77.1 ± 11.2	0.0%
Weight	74.4 ± 15.4	15.0%
Height	162.1 ± 9.6	16.5%
BMI	28.3 ± 5.5	17.0%
SBP	136.6 ± 18.9	12.2%
DBP	77.4 ± 12.0	12.2%
Total cholesterol	176.7 ± 41.9	14.3%
LDL cholesterol	108.7 ± 35.6	31.2%
HDL cholesterol	45.5 ± 13.6	15.0%
Triglycerides	116.9 ± 59.1	14.6%
Creatinine	1.1 ± 0.6	4.8%
Glycemia	127.2 ± 55.9	5.0%
Hba1c	6.3 ± 1.2	45.6%
TSH	1.9 ± 1.4	33.4%
UACR	33.0 ± 57.1	55.7%
eGFR	63.7 ± 21.1	35.3%

* Age is estimated from categorical age groups using midpoints.

Table 3 presents the distribution of the most relevant variables, including demographic and clinical characteristics. For each variable, the table shows the number of subjects in each category, the corresponding percentage of the total population, and the number of cardiovascular death cases observed within that category. The majority of the population is between 70 and 89 years old (64.4%), with a slightly higher proportion of women than men (53.3% women). BMI and weight are higher than reference values from healthy populations. Hypertension is highly prevalent, affecting 81% of the population; however, systolic and diastolic blood pressure values appear to be within normal ranges, likely due to the widespread use of antihypertensive medication (91.5%). Total cholesterol and triglycerides are elevated in only a subset of subjects (24% and 17.9%, respectively). Nearly 40% of the population is diabetic (37.4%), and

Table 3: Descriptive statistics of the study cohort

Risk Factor	N	Percentage	CV Death
Age (years)			
40-49	130	1.8%	3
50-59	435	6.0%	24
60-69	1,178	16.4%	107
70-79	2,184	30.3%	322
80-89	2,453	34.1%	620
≥ 90	774	10.7%	204
Gender			
Male	3,366	46.7%	582
Female	3,837	53.3%	698
BMI (kg/m^2)			
<20	261	3.6%	55
20–<25	1513	21.0%	298
25–<30	2296	31.9%	386
≥ 30	2087	29.0%	332
Height (cm)			
<156	1653	22.9%	322
156–<164	1816	25.2%	332
164–<173	1773	24.6%	291
≥ 170	947	13.1%	138
Weight (kg)			
<55	567	7.9%	122
55–<70	1919	26.6%	368
70–<85	2361	32.8%	394
85–<100	1079	15.0%	169
≥ 100	361	5.0%	41
SBP (mmHg)			
<100	152	2.1%	25
100–<120	935	13.0%	186
120–<140	2682	37.2%	416
140–<160	2023	28.1%	355
≥ 160	662	9.2%	149
DBP (mmHg)			
<70	1612	22.4%	359
70–<80	2047	28.4%	342
80–<90	1909	26.5%	300
90–<100	693	9.6%	100
≥ 100	194	2.7%	31
Total cholesterol (mg/dL)			
<200	4580	63.6%	852
200–<240	1292	17.9%	208
≥ 240	441	6.1%	70
Triglycerides (mg/dL)			
<150	5008	69.5%	913
150–<200	770	10.7%	125
≥ 200	520	7.2%	89
Diabetes mellitus (mg/dL)			
Type 1	323	4.5%	72
Type 2	2368	32.9%	466
Smoking status			
Never	5456	75.8%	969
Former	426	5.9%	62
Current	428	5.9%	59
Hypertension medication use	6593	91.5%	1202
Coronary heart disease	777	10.8%	192
Cancer	1622	22.5%	324
Carotid disease	454	6.3%	110
Dyslipidemia	4635	64.4%	821
Chronic obstructive pulmonary disease	894	12.4%	204
Flutter	602	8.4%	125
Heart failure	2184	30.3%	529
Myocardial infarction or unstable angina	1234	17.1%	276
Peripheral artery disease	651	9.0%	159
Stroke	197	2.73%	29
Thyroid disease	815	11.3%	151
Valvular heart disease	1885	26.2%	409
Hypertension	5859	81.3%	1085
Cardiac device	157	2.2%	28
Anticoagulants	3192	44.3%	473

most have never smoked (75.8%). Several comorbidities are common: dyslipidemia (64.4%), heart failure (30.3%), valvular heart disease (26.2%), and cancer (22.5%). Other conditions with notable prevalence include myocardial infarction or unstable angina (17.1%), chronic obstructive pulmonary disease (12.4%), thyroid disease (11.3%), coronary heart disease (10.8%), peripheral artery disease (9%), and carotid disease (6.3%). The distribution of cardiovascular deaths across population groups is consistent with the underlying group sizes.

4 Solution

This section outlines the methodology employed in this study. It begins with a description of the data preprocessing steps, including cleaning, transformation, and preparation for analysis. We then detail the implementation of the predictive models.

4.1 Data Preprocessing

The study followed a structured data preprocessing and modeling workflow. Data cleaning involved recalculating all time-related variables relative to each patient's atrial fibrillation (AF) diagnosis and excluding individuals under 40 years of age. Outliers were treated through capping and log transformation, and implausible values were corrected. Missing data were handled through targeted removal of incomplete records and mean imputation, after evaluating each variable's relevance. Constant and redundant features were removed, and new clinically meaningful features were engineered, including aggregated indicators for vascular disease, diabetes, kidney function, and medication use. Temporal variables were used to derive longitudinal and delta-based representations to capture within-patient variation over time. Three dataset versions were produced: static, slope-based, and longitudinal. Class imbalance was addressed through resampling methods such as random sampling and SMOTE, depending on the outcome and model. Finally, standardization using the `StandardScaler` was applied to scale-sensitive algorithms (e.g., logistic regression, Naïve Bayes, multilayer perceptron), ensuring that the training and test sets remained properly separated.

4.2 Classical Risk Calculators

To evaluate the performance of classical risk calculators within the AF cohort, three representative models were selected: CHA2DS2-VASc, which follows a point-based system, and GARFIELD-AF, which is based on Cox regression. CHA2DS2-VASc and GARFIELD-AF were selected for their ability to predict stroke, with GARFIELD-AF also capable of predicting mortality in AF cohorts. These scores serve as a baseline for comparison with subsequently developed machine learning models, as they are specifically tailored for these comorbidities in AF cohorts. To ensure that the calculators provide true predictions of future AF risk, they were applied using only data available prior to AF diagnosis, without relying on any information obtained after the event.

To access stroke risk in the cohort CHA_2DS_2-VASc score is further calculated, estimating stroke risk over a 1-year period. CHA_2DS_2-VASc is a risk-stratification score ranging from 0 to 9, depending on the number and weight of the score's risk components. Table 4 shows each risk factor and the corresponding score.

Table 4: CHA_2DS_2-VASc risk factors and correspondent score [19]

Risk Factor	Score
Congestive heart failure (C)	1
Hypertension (H)	1
Age 75+ years (A)	2
Diabetes mellitus (D)	1
Stroke (S)	2
Vascular disease (V)	1
Age 65-74 years (A)	1
Sex category - female (Sc)	1

* If a woman is assigned only one point due to sex category, her score should be adjusted to zero.

The dataset contains all necessary information to compute CHA_2DS_2--VASc score. Data on congestive heart failure (equivalent to heart failure), hypertension, diabetes mellitus, stroke, and sex are directly available. Vascular disease was derived in the data preprocessing. Age was calculated using the middle value of the category interval: patients aged 60–69 were assigned a score of 1 (using 65 as the representative age), while those aged 70–79 or older were assigned a score of 2.

The GARFIELD-AF score was calculated for all patients in the cohort to evaluate their risk of stroke and all-cause mortality. Calculation of the GARFIELD-AF score requires the following variables: age, sex, ethnicity, diastolic blood pressure, pulse, history of heart failure, vascular disease, prior stroke, history of bleeding, diabetes, moderate-to-severe chronic kidney disease, dementia, current smoking, and use of vitamin K antagonist (VKA) or non–vitamin K antagonist oral anticoagulants (NOACs). Among these, sex, diastolic blood pressure, history of heart failure, prior stroke, current smoking, and chronic kidney disease were directly available from the dataset. Vascular disease and diabetes were derived during data preprocessing. Age was estimated as the midpoint of each age category, similarly to the other classical calculators. Variables not captured in the dataset were handled with assumptions: all patients were considered Caucasian, pulse was set at 80 bpm, and patients were assumed to have no history of bleeding or dementia. Regarding anticoagulant use, VKA treatment was assumed absent, and NOAC treatment was considered present for patients recorded as taking any anticoagulant. The score for each outcome was then computed using the original regression coefficients, summarized in Table 5, together with the Cox proportional hazards formulas (Equation 1),

$$\text{GARFIELD-AF all-cause mortality} = 1 - 0.987921904^{\exp(\sum \beta X)}, \tag{1}$$

for all-cause mortality, and Equation 2,

$$\text{GARFIELD-AF ischemic stroke/SE} = 1 - 0.9925445321^{\exp(\sum \beta X)}, \tag{2}$$

1. β is the regression coefficient associated with each risk factor.
2. X is the value or presence of each corresponding risk factor, with the respective adjustments.
3. All-cause mortality equation is for estimating risk at 6 months, and the ischemic stroke/SE formula is for estimating risk at 1 year. 4. The original GARFIELD-AF equations are expressed as percentages and are therefore multiplied by 100.

for ischemic stroke or systemic embolism. Other GARFIELD-AF formulas can be found at https://af.garfieldregistry.org/garfield-af-risk-calculator.

Table 5: GARFIELD-AF model coefficients for 6-month all-cause mortality and 1-year ischemic stroke/systemic embolism

Variable	All-cause death β	Stroke/SE β
Female sex	-0.306202287	-
Heart failure	0.693789082	0.233182644
Vascular disease	0.306120964	0.197919709
Prior stroke	0.265852980	0.800863063
History of bleeding	0.385407386	0.298839670
Diabetes	0.280133213	0.211995445
Moderate-to-severe CKD	0.377903886	0.349516938
Dementia	0.489453313	0.513221391
Current smoking	0.345481149	0.478831506
OAC treatment: NOAC	-0.414591263	-0.572199357
OAC treatment: VKA	-0.185935610	-0.352373263
Ethnicity: Hispanic/Latino	0.157023564	-
Ethnicity: Asian	-0.609609055	-
Ethnicity: Black/Other	0.375675102	-
(Age-65) (if $\leq$ 65)	0.031050027	0.039138147
(Age-65) (if $>$ 65)	0.064594824	-
(Weight-75) (if $\leq$ 75)	-0.021535182	-
(Pulse-120) (if $\leq$ 120)	0.007678035	-
(Diastolic BP-80) (if $\leq$ 80)	-0.019304333	-
(Diastolic BP-80) (if $>$ 80)	-	0.015900160

4.3 Machine Learning Predictors

To predict the outcomes, a set of well-established machine learning models was implemented. These included Naïve Bayes, Logistic Regression, Decision Tree, Random Forest, XGBoost, and a Multi-Layer Perceptron (MLP). This selection encompasses both classical statistical methods and modern ensemble and neural network approaches, allowing for a comprehensive evaluation of predictive performance across different modeling strategies.

Each model was trained for every outcome variable at the 6-month time horizon, except for stroke combined with arterial embolism, which was trained using data from the full 10-year cohort due to class imbalance. Models were trained separately for each dataset type—static, slope-based, and longitudinal—and both independently for each outcome to avoid multi-target bias, as well as jointly in a multi-target framework.

Robustness was ensured through 5-fold cross-validation, with model hyperparameters optimized via Bayesian optimization using the F_2 score (Equation 3),

$$F_2 = (1 + 2^2) \cdot \frac{\text{Precision} \cdot \text{Recall}}{(2^2 \cdot \text{Precision}) + \text{Recall}}, \tag{3}$$

as the objective, and final selection was also guided by the F_2 score during sampling. Using the F-measure with $\beta = 2$ gives greater weight to recall, ensuring higher sensitivity to false negatives compared to false positives. For each outcome, the best-performing predictor was identified through rank fusion of F_1, F_2, and AUC (Equation 4),

$$\text{AUC} = \int_0^1 \text{TPR}(\text{FPR})\, d(\text{FPR}). \tag{4}$$

1. TPR: True positive rate
2. FPR: False positive rate

The best-performing model was subsequently used to assess performance across all time horizons and subjected to explainability analysis using SHAP-based feature importance.

5 Results and Evaluation

In this section, we evaluate several machine learning models for predicting stroke/systemic embolism and all-cause death outcomes under three data settings: (i) static, (ii) slope-based, and (iii) longitudinal. Model performance is compared against traditional methods CHA_2DS_2-VASc and GARFIELD-AF. Multi-label classification was also experimented with, but results were substantially poorer and are therefore not reported. Each model was tuned using Bayesian optimization with the F_2 score as the objective, and also selected based on the F_2 score during sampling. The best-performing models were then selected using a rank fusion of F_1, F_2, and AUC, and were carried forward for further analyses, including feature importance assessment and outcome prediction across additional time horizons. Sometimes, the assessment of feature importance is unclear for longitudinal predictors, so a static model is used to obtain more traditional insights.

5.1 Stroke and Systemic Embolism

The CHA_2DS_2-VASc score was developed to estimate 1-year stroke or systemic embolism risk in patients with atrial fibrillation, with traditional risk categories defined as 0 for low risk, 1 for intermediate risk, and ≥2 for high risk [19]. According to the 2020 ESC guidelines for the diagnosis and management of AF, risk stratification should be interpreted in a sex-specific manner: low risk (score = 0 in men, or 1 in women), for whom antithrombotic therapy is generally not recommended; intermediate risk (score = 1 in men, or 2 in women), for whom oral anticoagulation (OAC) may be considered; and high risk (score ≥2 in men, or ≥3 in women), for whom OAC is recommended [15].

In our cohort, the distribution of CHA_2DS_2-VASc scores is shown in Figure 1 (left), with a mean of 4.09 and a standard deviation of 1.44. The score achieved an area under the curve (AUC) of 0.588. Most patients had elevated scores, with approximately 94% classified as high risk, 5% as intermediate risk, and 1% as low risk according to the ESC 2020 thresholds. These results are consistent with the presence of a high-risk, frail population and indicate that our dataset contains strong predictors of stroke and systemic embolism.

The ROC curve results were also favorable, with the CHA_2DS_2-VASc score achieving an AUC comparable to that of its original derivation cohort (0.588 vs. 0.606). However, it is important to note that the original derivation cohort [19] primarily included patients not receiving anticoagulants, although antiplatelet therapy was allowed. In contrast, a substantial proportion of our cohort (44%) was on anticoagulants, in addition to other cardiovascular medications, which can significantly alter individual stroke risk profiles.

Figure 2 shows the distribution of GARFIELD-AF risk estimates for 1-year ischemic stroke or systemic embolism across the cohort. Although GARFIELD-AF has not defined formal risk categories, higher deciles of predicted risk correspond to increased observed stroke incidence [1]. GARFIELD-AF also demonstrates superior predictive performance compared with CHA_2DS_2-VASc, achieving a higher AUC both in the derivation cohort (0.65 vs. 0.59) and in our treated cohort (0.633 vs. 0.588). In this task, an AUC of 0.63 in a validation cohort represents a strong performance, demonstrating the efficacy of GARFIELD-AF in predicting stroke and systemic embolism. It is also noteworthy that several variables required to calculate GARFIELD-AF—such as pulse, ethnicity, bleeding history, dementia, and type of anticoagulant (NOAC vs. VKA)—had to be approximated or inferred from our available dataset, which may have affected the predictive accuracy of the score. Nevertheless, the good AUCs observed for both CHA_2DS_2-VASc and GARFIELD-AF demonstrate the strength of these predictors, and indicate that the dataset provides reliable indicators of stroke and arterial embolism risk.

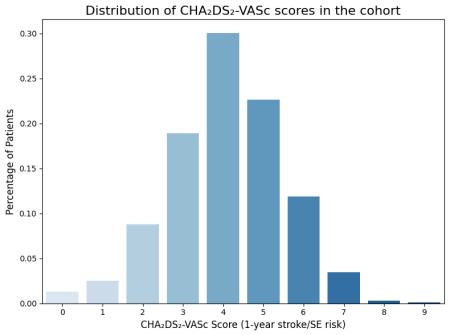

Figure 1: Distribution of CHA_2DS_2-VASc scores in the cohort for predicting stroke or systemic embolism at 1 year

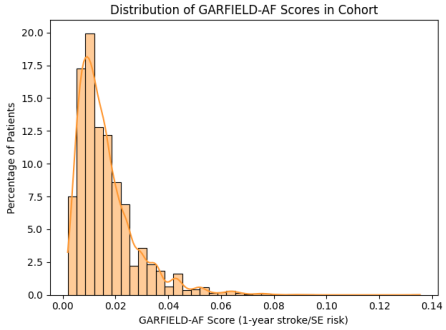

Figure 2: Distribution of GARFIELD-AF scores in the cohort for predicting stroke or systemic embolism at 1 year

Table 6 summarizes the performance of the machine learning models in predicting stroke or arterial embolism over a 10-year period. The best-performing model was Logistic Regression (LR), with the Multi-Layer Perceptron (MLP) and XGBoost trailing behind. Among the models, the static XGBoost achieved the highest precision, while the slope-based LR attained the highest F_1 and F_2 scores, reflecting a strong balance between precision and sensitivity. Additionally, LR achieved the highest AUC (0.634) and the lowest NNS (29.24), highlighting its superior overall performance.

The Naive Bayes demonstrated high sensitivity, while the remaining models showed comparatively limited performance, generally characterized by low precision and sensitivity. Incorporating slope

and other longitudinal features enhanced model performance, highlighting the added value of temporal variables in predicting stroke or systemic embolism.

Figure 3 presents the SHAP feature importance for the slope-based Logistic Regression model, which was the best-performing model. The analysis suggests that shorter height is associated with an increased risk of stroke or systemic embolism. Although height is not traditionally recognized as a stroke/SE risk factor, prior studies have linked shorter stature to cardiovascular disease or stroke, proposing that it may serve as a marker of early-life social and physical conditions [21].

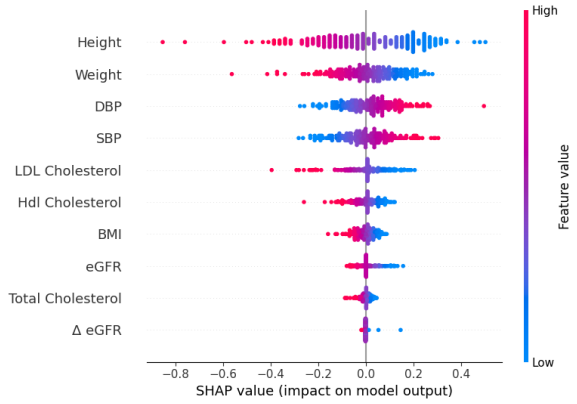

Figure 3: SHAP value of slope-based Logistic Regression model on predicting stroke and arterial embolism

Low body weight and BMI also emerged as predictors of increased risk, whereas higher values seemed protective—a pattern consistent with the so-called "obesity paradox." This paradox is thought to arise from factors such as malnutrition, frailty, and the limitations of BMI and weight in distinguishing fat from lean mass.

Blood pressure was another strong predictor, with both systolic and diastolic elevations associated with higher risk. Elevated systolic pressure is well established as a stroke risk factor, while diastolic pressure is generally less predictive; however, both contributed meaningfully in this model.

Interestingly, lower LDL, HDL, and total cholesterol levels were also linked to higher risk. While elevated LDL is classically considered harmful, the inverse association observed here may reflect the widespread use of statins in this population, which lowers LDL and therefore total cholesterol. Similar "cholesterol paradox" findings have been reported in older, frail, and atrial fibrillation cohorts [23].

Additionally, lower eGFR values were associated with a higher risk, consistent with existing evidence that impaired kidney function contributes to stroke and embolism. Changes in eGFR over time also influenced model performance, with declining eGFR linked to additional risk, in line with baseline eGFR findings. Although most slope features did not rank among the most important predictors in the model, they nonetheless made a meaningful contribution to risk prediction.

Age, despite being a well-established predictor, did not rank among the strongest features in this model. This may reflect overlap with other clinical variables such as blood pressure, renal function, and other measures, which may have captured much of the age-related risk.

The model was subsequently trained across additional time horizons, including 1 year, to assess its performance and compare it with CHA_2DS_2-VASc and GARFIELD-AF. The results for these time ranges are presented in Table 7. Model performance increases from 6 months to 1 year and is similar at 1 year and 2 years; however, the highest AUC is observed at 1 year. The 1-month and 3-month horizons were excluded, as the degree of class imbalance prevented the generation of reliable predictions.

Table 7: Performance of slope-based Logistic Regression model in predicting stroke and artery embolism, reported as mean ± standard deviation for F_1 score, F_2 score, and AUC.

Interval	F_1 score	F_2 score	AUC
6 months	0.020 ± 0.007	0.046 ± 0.017	0.582 ± 0.078
1 year	0.034 ± 0.014	0.076 ± 0.031	**0.651 ± 0.098**
2 years	0.034 ± 0.004	0.076 ± 0.010	0.562 ± 0.055

The static Logistic Regression model achieved only modest predictive performance, with very low precision and sensitivity. Nevertheless, its AUC for 1-year stroke risk (0.651) exceeded that of CHA_2DS_2-VASc (0.588) and GARFIELD-AF (0.633), outperforming the classical risk calculators. However, the weaker results at the 6-month and 2-year horizons indicate that the predictions are not yet reliable, and that addressing the severe data imbalance will likely require either a specifically tailored algorithm or a larger dataset to achieve both robust performance and greater reliability.

5.2 All-Cause Death

Figure 4 shows the distribution of GARFIELD-AF predictions at 6 months in the cohort (left panel) and the corresponding ROC curve (right panel). The distribution indicates that several patients are at high risk (≥ 10%) of mortality within the next six months, with some exhibiting even higher predicted probabilities. An AUC of 0.72 demonstrates good discriminatory ability, indicating that the model can reliably differentiate between patients who did and did not experience all-cause mortality within six months.

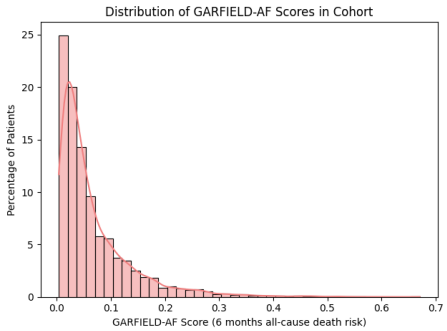

Figure 4: Distribution of GARFIELD-AF scores in the cohort for predicting death at 6 months

Table 8 presents the performance of the ML models in predicting all-cause death at 6 months. XGBoost, Random Forest, and Logistic

Table 6: Top performance machine learning models in predicting stroke/SE, reported as mean ± standard deviation for accuracy, precision, sensitivity, F_1 score, F_2 score, AUC, and NNS. The best models in each metric are highlighted in bold.

Model	Accuracy	Precision	Sensitivity	F_1 score	F_2 score	AUC	NNS	Rank
Static LR	0.522 ± 0.017	0.030 ± 0.001	**0.640 ± 0.028**	0.056 ± 0.003	0.125 ± 0.006	0.605 ± 0.022	33.93 ± 1.76	3
Static MLP	**0.859 ± 0.018**	**0.036 ± 0.012**	0.205 ± 0.076	0.061 ± 0.021	0.106 ± 0.037	0.574 ± 0.043	32.59 ± 15.23	4
Slope-based LR	0.612 ± 0.031	0.034 ± 0.002	0.603 ± 0.059	**0.065 ± 0.004**	**0.140 ± 0.008**	**0.634 ± 0.022**	**29.24 ± 1.70**	**1**
Slope-based MLP	0.856 ± 0.014	0.035 ± 0.008	0.205 ± 0.047	0.060 ± 0.013	0.104 ± 0.023	0.575 ± 0.033	29.91 ± 6.45	5
Longitudinal LR	0.774 ± 0.008	0.034 ± 0.008	0.338 ± 0.071	0.063 ± 0.014	0.122 ± 0.027	0.611 ± 0.032	30.65 ± 7.45	2

Regression achieved the best overall performance, combining high sensitivity with strong precision. Other models performed less well: Naive Bayes showed higher sensitivity but at the expense of precision, while the MLP and Decision Tree exhibited generally poor performance. Incorporating slope features did not improve predictive ability, whereas including the full set of longitudinal features enhanced model performance, indicating that these longitudinal features contain valuable predictive information.

The best-performing model was the longitudinal XGBoost, and its AUC for 6-month all-cause death risk (0.779) exceeded that of GARFIELD-AF (0.715), showing once again the efficacy of machine learning.

The longitudinal XGBoost has its feature importance presented in Figure 5.

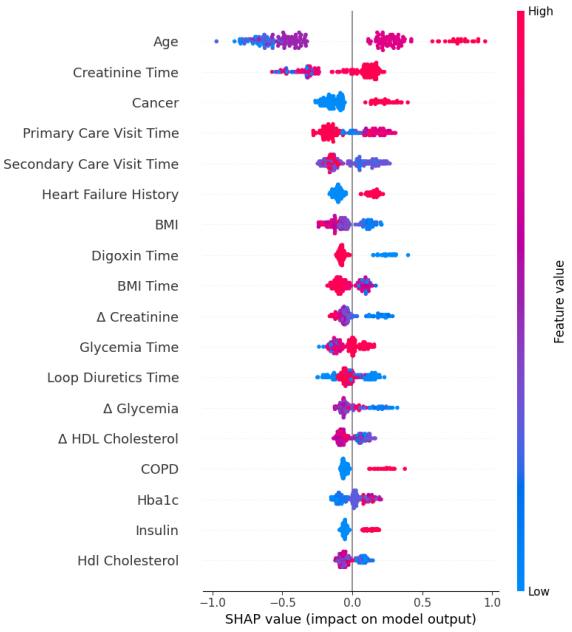

Figure 5: SHAP value of longitudinal XGBoost model on predicting all-cause death at 6 months

Age emerged as the strongest predictor, consistent with its role as a major risk factor, with patients aged 80–89—and especially those over 90—strongly related with high risk of all-cause mortality. A clinical history of cancer, heart failure, and COPD also contributed to higher risk, in line with established medical knowledge.

The obesity paradox was observed once again, with lower BMI values associated with a higher risk. Similarly, lower HDL cholesterol levels were also linked to a higher risk, aligning with established medical knowledge. Medication use also contributed to model performance; for instance, insulin use and elevated HbA1c levels, indicating prolonged hyperglycemia over the preceding 2–3 months and consistent with the diagnostic criteria for diabetes, were both related to higher risk. These findings align well with existing clinical understanding.

Among the slope-based features, decreases in creatinine, glycemia, and HDL cholesterol appeared to increase risk. Although elevated creatinine is typically indicative of impaired renal function, a decrease in creatinine may instead reflect reduced muscle mass or underlying liver disease—both recognized markers of frailty. A decline in HDL cholesterol increasing risk is consistent with established knowledge, whereas a decrease in glycemia increasing risk is contrary to expectations. This counterintuitive finding may occur because substantial declines in glycemia often follow initially elevated baseline levels, suggesting an underlying diagnosis of diabetes.

The longitudinal (temporal) features also showed high relevance within the model. The time-related variables that are associated with binary indicators, enable the model to capture both the occurrence and timing of clinical events. Typically, lower time values indicate a positive event, while higher values correspond to its absence.

Creatinine time had a notable influence, with more recent measurements associated with higher risk—likely reflecting its connection to current renal function status. Primary and secondary care visit times were also informative: individuals without recent primary care visits tended to have lower risk, possibly reflecting overall good health and reduced healthcare utilization, while recent visits were associated with increased mortality risk. Conversely, shorter intervals since a secondary care visit were associated with higher risk, suggesting that more frequent specialist monitoring likely reflects greater patient frailty.

Medication- and test-related time variables also provided insights. Recent digoxin use was associated with increased risk, as expected given its prescription in patients with cardiac dysfunction. Similarly, recent BMI and glycemia measurements correlated with higher risk, potentially reflecting closer monitoring of patients with existing health concerns. However, the relevance of Loop Diuretics time was less clear.

Overall, these temporal features demonstrate that their interpretation is complex and not always directly aligned with medical

Table 8: Performance of top performance machine learning models in predicting all-cause death, reported as mean ± standard deviation for accuracy, precision, sensitivity, F_1 score, F_2 score, AUC, and NNS.

Model	Accuracy	Precision	Sensitivity	F_1 score	F_2 score	AUC	NNS	Rank
Static LR	0.615 ± 0.032	0.096 ± 0.004	**0.815 ± 0.063**	0.171 ± 0.006	0.325 ± 0.011	0.774 ± 0.020	10.45 ± 0.42	4
Static RF	0.654 ± 0.043	0.102 ± 0.009	0.771 ± 0.051	0.180 ± 0.015	0.332 ± 0.023	0.769 ± 0.024	9.91 ± 0.98	5
Longitudinal LR	0.655 ± 0.020	0.102 ± 0.005	0.778 ± 0.035	0.180 ± 0.007	0.334 ± 0.011	0.774 ± 0.011	9.82 ± 0.43	2
Longitudinal RF	0.659 ± 0.040	0.105 ± 0.017	0.778 ± 0.068	0.184 ± 0.029	0.339 ± 0.047	0.753 ± 0.068	9.85 ± 1.72	3
Longitudinal XGB	**0.695 ± 0.021**	**0.111 ± 0.012**	0.741 ± 0.051	**0.192 ± 0.020**	**0.346 ± 0.033**	**0.779 ± 0.031**	**9.16 ± 1.04**	**1**

knowledge. Many of these associations likely reflect patterns of healthcare utilization, where sicker or frailer individuals undergo more frequent testing and medication changes. It is not clear if the models are creating relationships between the variable and its timing.

Some medications showed that their use is associated with an elevated risk. However, these associations are unlikely to represent direct causal effects of the drugs themselves. Rather, they likely capture the higher baseline risk of patients with complex comorbidities for which these medications are prescribed. Accordingly, their predictive value reflects underlying disease burden rather than harmful effects of the treatments.

To identify the most influential predictors in a static model, static Logistic Regression was selected, and its feature importance is illustrated in Figure 6.

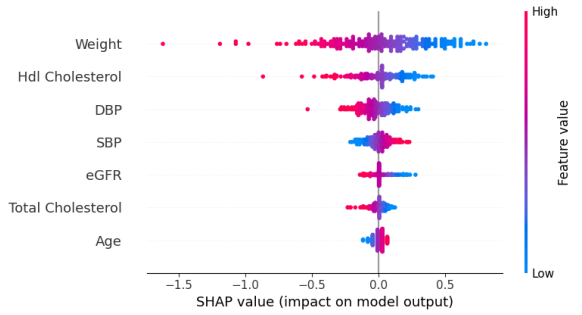

Figure 6: SHAP value of static Logistic Regression model on predicting all-cause death at 6 months

The results closely mirror those of the slope-based Logistic Regression model for stroke/SE prediction. Similar features, including weight, HDL cholesterol, systolic blood pressure (SBP), estimated glomerular filtration rate (eGFR), and total cholesterol, consistently emerged as important predictors. Weight reflected the obesity paradox, with higher values associated with lower risk. Low HDL cholesterol increased risk, as expected, while reduced eGFR indicated kidney dysfunction. Interestingly, lower total cholesterol was associated with a higher risk, contrary to expectations but potentially explained by statin use. Elevated SBP was linked to greater risk of death, consistent with established medical knowledge, whereas low diastolic blood pressure appeared to increase risk—a surprising finding, as it is not typically considered a strong risk factor in these populations. Age also contributed modestly, with higher values associated with increased risk, as expected.

The longitudinal XGBoost model was then further evaluated across different time intervals. Table 9 reports the results: model performance generally improves as the time horizon increases, reflecting the larger number of positive cases; however, the highest AUC is observed at 3 months.

Table 9: Performance of longitudinal XGBoost model in predicting stroke and artery embolism, reported as mean ± standard deviation for F_1 score, F_2 score, and AUC.

Interval	F_1 score	F_2 score	AUC
1 month	0.061 ± 0.019	0.095 ± 0.031	0.722 ± 0.049
3 months	0.165 ± 0.025	0.279 ± 0.047	**0.803 ± 0.030**
6 months	0.192 ± 0.020	0.346 ± 0.033	0.779 ± 0.031
1 year	0.243 ± 0.022	0.412 ± 0.033	0.768 ± 0.038
2 years	**0.312 ± 0.014**	**0.478 ± 0.022**	0.753 ± 0.013

6 Clinical Decision Support Tool

The developed clinical decision support tool comprises two core components: a backend API and a graphical user interface (GUI). Together, they enable seamless interaction between predictive models and end users. The source code is available at https://github.com/ulsmatosinhos/tre-nova-af, with access provided by ULSM upon request.

6.1 Application Programming Interface (API)

The API enables integration with existing clinical systems and workflows, offering standardized access to the predictive models and front-end interface. It was implemented using `FastAPI`, a modern Python web framework chosen for performance and scalability.

Key endpoints such as `/predict` accept patient data in JSON format, process it through the trained machine learning models, and return predictions with associated probabilities. This design decouples the computational backend from the user interface, allowing multiple clients (e.g., the `Dash` app, hospital systems, or third-party services) to securely access the prediction service.

Responses are returned in JSON for interoperability, and the system runs on an ASGI server (`uvicorn`) that can be deployed standalone or containerized. This modular architecture enhances maintainability, enabling model updates without modifying the interface or external integrations.

6.2 Graphical User Interface (GUI)

The GUI serves as the front end of the tool, allowing clinicians to input patient data and visualize predictions. Built with `Dash`, it

provides interactive forms for variables such as age, weight, blood pressure, and laboratory values. Future versions will support dynamic form customization to handle missing or optional inputs.

Predictions are displayed through clear visualizations that position the patient's data within the training cohort, improving interpretability. When data are submitted, the GUI communicates with the backend API, retrieves predictions, and displays the results in real time.

By separating the interface from the API, the system maintains flexibility—visual elements and layouts can evolve independently of the underlying models. Although the current version supports predictions for cardiovascular death, the framework is designed to extend to additional outcomes, including AF onset and related complications.

Overall, the prototype demonstrates that predictive modeling can be effectively integrated into a practical, user-friendly clinical tool. Its modular, extensible design ensures that future iterations can expand both functionality and usability while remaining compatible with clinical workflows.

Figure 7 presents the graphical user interface of the tool, showing how patient data is entered and processed through the API.

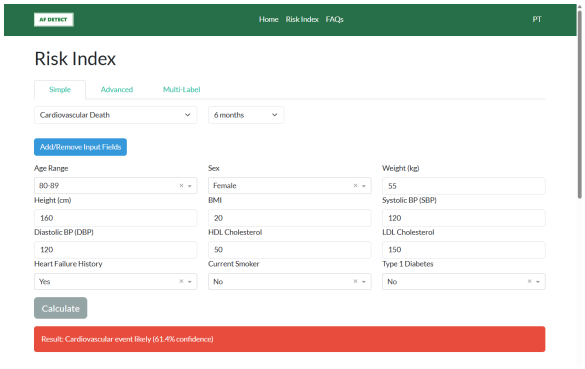

Figure 7: Prototype interface of the medical prediction tool predicting cardiovascular death at 6 months

7 Conclusion

This chapter summarizes the key results and insights of the study, linking model findings with clinical knowledge while identifying unexpected patterns shaped by comorbidities and medication use. It also discusses limitations and outlines directions for future research.

7.1 Concluding Remarks

This thesis addressed the challenge of prognosticating atrial fibrillation (AF)–related outcomes by pursuing two main objectives: (1) developing predictive models for clinically relevant endpoints, and (2) designing a prototype clinical decision-support tool. Together, these efforts aim to improve AF management in primary care.

The predictive models achieved strong performance, often surpassing classical risk calculators for stroke, systemic embolism, and all-cause mortality, and performed robustly across different outcomes and time horizons. Integrating diverse diagnostic, clinical, and behavioral data—along with temporal information—enabled a more holistic view of patient risk. Including timing-related features, such as diagnosis and medication events, consistently improved model performance by capturing progression patterns.

Across outcomes, several recurring associations emerged. Height showed a consistent inverse relationship with risk, possibly reflecting early-life social or physical factors influencing cardiovascular health. The "obesity paradox" was also observed: patients with lower BMI tended to have higher risk, likely due to frailty or confounding by illness. Lipid patterns showed that lower LDL and total cholesterol were often linked to higher risk, likely reflecting widespread statin use in the cohort. Blood pressure variables provided further insight—high systolic pressure increased risk as expected, but lower diastolic pressure was also associated with poorer outcomes, potentially due to medication effects. Notably, hypertension itself rarely appeared as important, likely because it was nearly universal in the population and thus lost predictive value.

Certain comorbidities, such as diabetes, did not consistently appear as predictors, while medication use (e.g., insulin, digoxin, loop diuretics) often did, highlighting how treatment practices shape model interpretation. Clinical variables like eGFR and age contributed meaningfully to predictions, though sex was significant only for heart failure hospitalization. These findings emphasize that in a heavily medicated AF cohort, traditional risk factors often behave differently, providing valuable insight into the complex interplay between treatment and risk.

While temporal models dominated, interpreting temporal features remains challenging. Some variables, such as "nitrates time," likely represent medication use rather than true temporal dynamics. Others, linked to clinical activity, may reflect healthcare utilization patterns rather than underlying pathology—an important consideration when interpreting predictive importance.

The second objective—developing a prototype clinical tool—demonstrated the feasibility of translating complex predictive models into a usable interface for clinicians. The tool integrated the trained models via a backend API, allowing real-time data entry and visualization of predictions. Though preliminary, it provides a foundation for future clinical decision-support systems.

Overall, this work highlights the potential of machine learning and temporal modeling for AF risk prediction, uncovers meaningful population-specific patterns, and lays groundwork for clinical implementation. With further validation, these contributions could enhance risk stratification, improve decision-making, and support better outcomes for patients with atrial fibrillation.

7.2 Limitations and Future Work

While this study demonstrates promising results, several limitations should be acknowledged. The dataset represented only an initial iteration of clinical records, limiting variable completeness and model generalizability. Important information such as AF type, pulse, dementia status, anticoagulant class (NOAC vs. VKA), cardiac murmurs, and ethnicity was missing. Additionally, data on bleeding history, liver health, and alcohol use—key for predicting bleeding risk—were unavailable.

Future work should include richer Electronic Health Record (EHR) integration and expand to include both AF and non-AF cohorts to enable AF incidence prediction. Access to electrocardiograms (ECGs), whether as raw signals or extracted features, would provide valuable complementary data, enhancing model robustness and interpretability.

Methodologically, models were designed for longitudinal data but not fully optimized for its structure. Future studies could explore deep learning architectures that capture variable–timestamp relationships and irregular sampling patterns. Improved feature engineering—such as clinically informed thresholds for lab values like eGFR or creatinine—could yield more interpretable and accurate predictions. Multi-output learning also warrants further exploration, especially for related outcomes such as cardiovascular and all-cause death. Threshold tuning to optimize specific clinical metrics (F_1, F_2, or cost-sensitive loss functions) could further enhance applicability.

Finally, the prototype interface should be refined through iterative design with clinicians to improve usability and integration into clinical workflows. With continued development, this system could serve as a practical tool supporting evidence-based, data-driven AF management in real-world settings.

References

[1] Jean-Pierre Bassand, Patricia N Apenteng, Dan Atar, A John Camm, Frank Cools, Ramon Corbalan, David A Fitzmaurice, Keith AA Fox, Shinya Goto, Sylvia Haas, et al. 2021. GARFIELD-AF: a worldwide prospective registry of patients with atrial fibrillation at risk of stroke. *Future cardiology* 17, 1 (2021), 19–38.

[2] Alexander P Benz, Ziad Hijazi, Johan Lindbäck, Stuart J Connolly, John W Eikelboom, Jonas Oldgren, Agneta Siegbahn, and Lars Wallentin. 2021. Biomarker-based risk prediction with the ABC-AF scores in patients with atrial fibrillation not receiving oral anticoagulation. *Circulation* 143, 19 (2021), 1863–1873.

[3] Yang Chen, Ying Gue, Peter Calvert, Dhiraj Gupta, Garry McDowell, Jinbert Lordson Azariah, Narayanan Namboodiri, Tommaso Bucci, A Jabir, Hung Fat Tse, et al. 2024. Predicting stroke in Asian patients with atrial fibrillation using machine learning: a report from the KERALA-AF registry, with external validation in the APHRS-AF registry. *Current problems in cardiology* 49, 4 (2024), 102456.

[4] Ehud Davidson, Zvi Rotenberg, Itzhak Weinberger, Jacob Fuchs, and Jacob Agmon. 1989. Diagnosis and characteristics of lone atrial fibrillation. *Chest* 95, 5 (1989), 1048–1050.

[5] Valentin Fuster, Lars E Rydén, Davis S Cannom, Harry J Crijns, Anne B Curtis, Kenneth A Ellenbogen, Jonathan L Halperin, G Neal Kay, Jean-Yves Le Huezey, James E Lowe, et al. 2011. 2011 ACCF/AHA/HRS focused updates incorporated into the ACC/AHA/ESC 2006 Guidelines for the management of patients with atrial fibrillation: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines developed in partnership with the European Society of Cardiology and in collaboration with the European Heart Rhythm Association and the Heart Rhythm Society. *Journal of the American College of Cardiology* 57, 11 (2011), e101–e198.

[6] Valentin Fuster, Lars E Rydén, David S Cannom, Harry J Crijns, Anne B Curtis, Kenneth A Ellenbogen, Jonathan L Halperin, Jean-Yves Le Heuzey, G Neal Kay, James E Lowe, et al. 2006. ACC/AHA/ESC 2006 guidelines for the management of patients with atrial fibrillation—executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the European Society of Cardiology Committee for Practice Guidelines (writing committee to revise the 2001 guidelines for the management of patients with atrial fibrillation) Developed in Collaboration with the European Heart Rhythm Association and the Heart Rhythm Society. *Journal of the American College of Cardiology* 48, 4 (2006), 854–906.

[7] Pilar Gallego, Vanessa Roldán, José Miguel Torregrosa, Josefa Gálvez, Mariano Valdés, Vicente Vicente, Francisco Marín, and Gregory YH Lip. 2012. Relation of the HAS-BLED bleeding risk score to major bleeding, cardiovascular events, and mortality in anticoagulated patients with atrial fibrillation. *Circulation: Arrhythmia and Electrophysiology* 5, 2 (2012), 312–318.

[8] Bill Goh and Sonu MM Bhaskar. 2024. Evaluating machine learning models for stroke prognosis and prediction in atrial fibrillation patients: a comprehensive meta-analysis. *Diagnostics* 14, 21 (2024), 2391.

[9] Miguel Gouveia, João Costa, Joana Alarcão, Margarida Augusto, Daniel Caldeira, Luís Pinheiro, António Vaz Carneiro, and Margarida Borges. 2015. Burden of disease and cost of illness of atrial fibrillation in Portugal. *Revista Portuguesa de Cardiologia (English Edition)* 34, 1 (2015), 1–11.

[10] Yasuhiro Hamatani, Hidehisa Nishi, Moritake Iguchi, Masahiro Esato, Hikari Tsuji, Hiromichi Wada, Koji Hasegawa, Hisashi Ogawa, Mitsuru Abe, Shunichi Fukuda, et al. 2022. Machine learning risk prediction for incident heart failure in patients with atrial fibrillation. *JACC: Asia* 2, 6 (2022), 706–716.

[11] Ziad Hijazi, Johan Lindbäck, John H Alexander, Michael Hanna, Claes Held, Elaine M Hylek, Renato D Lopes, Jonas Oldgren, Agneta Siegbahn, Ralph AH Stewart, et al. 2016. The ABC (age, biomarkers, clinical history) stroke risk score: a biomarker-based risk score for predicting stroke in atrial fibrillation. *European heart journal* 37, 20 (2016), 1582–1590.

[12] Ziad Hijazi, Jonas Oldgren, Johan Lindbäck, John H Alexander, Stuart J Connolly, John W Eikelboom, Michael D Ezekowitz, Claes Held, Elaine M Hylek, Renato D Lopes, et al. 2016. The novel biomarker-based ABC (age, biomarkers, clinical history)-bleeding risk score for patients with atrial fibrillation: a derivation and validation study. *The Lancet* 387, 10035 (2016), 2302–2311.

[13] Ziad Hijazi, Jonas Oldgren, Johan Lindbäck, John H Alexander, Stuart J Connolly, John W Eikelboom, Michael D Ezekowitz, Claes Held, Elaine M Hylek, Renato D Lopes, et al. 2018. A biomarker-based risk score to predict death in patients with atrial fibrillation: the ABC (age, biomarkers, clinical history) death risk score. *European heart journal* 39, 6 (2018), 477–485.

[14] Jelle CL Himmelreich, Lieke Veelers, Wim AM Lucassen, Renate B Schnabel, Michiel Rienstra, Henk CPM van Weert, and Ralf E Harskamp. 2020. Prediction models for atrial fibrillation applicable in the community: a systematic review and meta-analysis. *EP Europace* 22, 5 (2020), 684–694.

[15] Gerhard Hindricks, Tatjana Potpara, Nikolaos Dagres, Elena Arbelo, Jeroen J Bax, Carina Blomström-Lundqvist, Giuseppe Boriani, Manuel Castella, Gheorghe-Andrei Dan, Polychronis E Dilaveris, et al. 2021. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS) The Task Force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) Developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. *European heart journal* 42, 5 (2021), 373–498.

[16] Pooja S Jagadish and Rajesh Kabra. 2019. Stroke risk in atrial fibrillation: Beyond the CHA 2 DS 2-VASc score. *Current cardiology reports* 21 (2019), 1–9.

[17] David F Katz, Thomas M Maddox, Mintu Turakhia, Anil Gehi, Emily C O'Brien, Steven A Lubitz, Alexander Turchin, Gheorghe Doros, Lanyu Lei, Paul Varosy, et al. 2017. Contemporary trends in oral anticoagulant prescription in atrial fibrillation patients at low to moderate risk of stroke after guideline-recommended change in use of the CHADS2 to the CHA2DS2-VASc score for thromboembolic risk assessment: analysis from the National Cardiovascular Data Registry's Outpatient Practice Innovation and Clinical Excellence Atrial Fibrillation Registry. *Circulation: Cardiovascular Quality and Outcomes* 10, 5 (2017), e003476.

[18] Bouwe P Krijthe, Anton Kunst, Emelia J Benjamin, Gregory YH Lip, Oscar H Franco, Albert Hofman, Jacqueline CM Witteman, Bruno H Stricker, and Jan Heeringa. 2013. Projections on the number of individuals with atrial fibrillation in the European Union, from 2000 to 2060. *European heart journal* 34, 35 (2013), 2746–2751.

[19] Gregory YH Lip, Robby Nieuwlaat, Ron Pisters, Deirdre A Lane, and Harry JGM Crijns. 2010. 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, 2 (2010), 263–272.

[20] Hidehisa Nishi, Naoya Oishi, Hisashi Ogawa, Kishida Natsue, Kento Doi, Osamu Kawakami, Tomokazu Aoki, Shunichi Fukuda, Masaharu Akao, Tetsuya Tsukahara, et al. 2022. Predicting cerebral infarction in patients with atrial fibrillation using machine learning: The Fushimi AF registry. *Journal of Cerebral Blood Flow & Metabolism* 42, 5 (2022), 746–756.

[21] Inger Njølstad, Egil Arnesen, and Per G Lund-Larsen. 1996. Body height, cardiovascular risk factors, and risk of stroke in middle-aged men and women: a 14-year follow-up of the Finnmark Study. *Circulation* 94, 11 (1996), 2877–2882.

[22] Renate B Schnabel, Lisa M Sullivan, Daniel Levy, Michael J Pencina, Joseph M Massaro, Ralph B D'Agostino, Christopher Newton-Cheh, Jennifer F Yamamoto, Jared W Magnani, Thomas M Tadros, et al. 2009. Development of a risk score for atrial fibrillation (Framingham Heart Study): a community-based cohort study. *The Lancet* 373, 9665 (2009), 739–745.

[23] Shinya Suzuki. 2011. " Cholesterol paradox" in atrial fibrillation. *Circulation Journal* 75, 12 (2011), 2749–2750.

[24] Andrew S Tseng and Peter A Noseworthy. 2021. Prediction of atrial fibrillation using machine learning: a review. *Frontiers in Physiology* 12 (2021), 752317.

[25] Developed with the Special Contribution of the European Heart Rhythm Association (EHRA), Endorsed by the European Association for Cardio-Thoracic Surgery (EACTS), Authors/Task Force Members, A John Camm, Paulus Kirchhof, Gregory YH Lip, Ulrich Schotten, Irene Savelieva, Sabine Ernst, Isabelle C Van Gelder, et al. 2010. Guidelines for the management of atrial fibrillation: the Task Force for the Management of Atrial Fibrillation of the European Society of Cardiology (ESC). *European heart journal* 31, 19 (2010), 2369–2429.