

Supplementary Table 1: Baseline characteristics of the derivation cohort stratified by stroke outcome within 1 year after atrial fibrillation diagnosis

	Stroke group (N=1,448)	Non-stroke group (N=8,063)	p-value
Age	75 (66, 82)	69 (60, 78)	<0.001
Sex, female	45.7	44.1	0.252
Clinical history (%)			
Any malignancy	8.0	8.8	0.348
Arterial embolism	2.6	1.9	0.108
Chronic lung disease	6.7	7.7	0.191
Congestive heart failure	14.9	17.3	0.031
Diabetes mellitus	28.9	18.0	<0.001
Hyperlipidemia	23.8	17.0	<0.001
Hypertension	60.5	44.6	<0.001
Prior ischemic stroke	78.6	3.4	<0.001
Mild liver disease	3.5	4.6	0.091
Myocardial infarction	1.9	2.1	0.618
Peptic ulcer disease excluding bleeding	4.6	3.4	0.024
Peptic ulcer disease	5.0	3.6	0.013
Peripheral vascular disease	3.9	2.0	<0.001
Renal failure	5.6	5.8	0.842
Valvular heart disease	4.2	5.2	0.122
Medication history (%)			
OHA	15.7	8.7	<0.001
Warfarin	15.4	7.2	<0.001
APA	32.2	18.7	<0.001
Digoxin	8.2	7.0	0.105
AAD	10.5	13.0	0.009
Vasodilators	12.8	8.2	<0.001
Diuretics	21.1	18.9	0.053
BBA	25.3	22.9	0.045
CCB	26.5	20.3	<0.001
ACEI	3.6	3.1	0.341
ARB	24.0	18.0	<0.001
Lipid modifying agents	15.3	10.6	<0.001

AAD indicates antiarrhythmic drugs; ACEI, angiotensin-converting enzyme inhibitor; APA, antiplatelet agents; ARB, angiotensin II receptor blocker; BBA, beta-blocking agents; CCB, calcium channel blocker; OHA, oral hypoglycemic agents.

Supplementary Table 2: Final hyperparameters used for model training

Model	Parameter	Value
Logistic regression	Solver	liblinear
	C	0.1
	Class Weight	balanced
Extreme gradient boosting	max_depth	3
	learning_rate	0.01
	n_estimators	250
	scale_pos_weight	5.6
	subsample	0.8
	colsample_bytree	0.7

Supplementary Table 3: Predictive performance of logistic regression (LR), extreme gradient boosting (XGB), and the CHA₂DS₂-VASc score across the internal and external validation cohorts

Metric	Internal validation cohort			External validation cohort 1			External validation cohort 2		
	LR	XGB	CHA ₂ DS ₂ -VASc	LR	XGB	CHA ₂ DS ₂ -VASc	LR	XGB	CHA ₂ DS ₂ -VASc
AUROC (95%CI)	0.91 (0.89-0.93)	0.91 (0.89-0.93)	0.67 (0.65-0.69)	0.88 (0.85-0.91)	0.88 (0.86-0.91)	0.61 (0.59-0.64)	0.88 (0.85-0.91)	0.88 (0.85-0.91)	0.62 (0.60-0.64)
AUPRC (95%CI)	0.74 (0.68-0.79)	0.75 (0.70-0.80)	0.22 (0.19-0.24)	0.77 (0.70-0.82)	0.77 (0.71-0.82)	0.23 (0.21-0.26)	0.79 (0.74-0.84)	0.80 (0.74-0.84)	0.25 (0.23-0.28)
Accuracy (%)	93.9	93.9	51.5	93.7	93.7	41.7	93.4	93.4	44.9
Sensitivity (Recall) (%)	79.3	79.3	90.3	73.5	73.5	93.1	75.1	75.1	91.3
Specificity (%)	96.5	96.5	44.5	98.4	98.4	29.8	98.1	98.1	33.0
Precision (%)	80.1	80.1	22.6	91.4	91.4	23.5	90.9	90.9	25.8
F1 Score	0.80	0.80	0.36	0.81	0.81	0.38	0.82	0.82	0.40
DeLong <i>p</i> -value* (vs. CHA ₂ DS ₂ -VASc)	<0.001	<0.001	-	<0.001	<0.001	-	<0.001	<0.001	-

**p*-values were calculated using DeLong's test to compare AUROC values among the three models. Bonferroni correction was applied for multiple pairwise comparisons ($\alpha = 0.05 / 3 = 0.0167$).

Abbreviations:

LR, logistic regression; XGB, extreme gradient boosting; AUROC, area under the receiver operating characteristic curve; AUPRC, area under the precision–recall curve; CI, confidence interval.

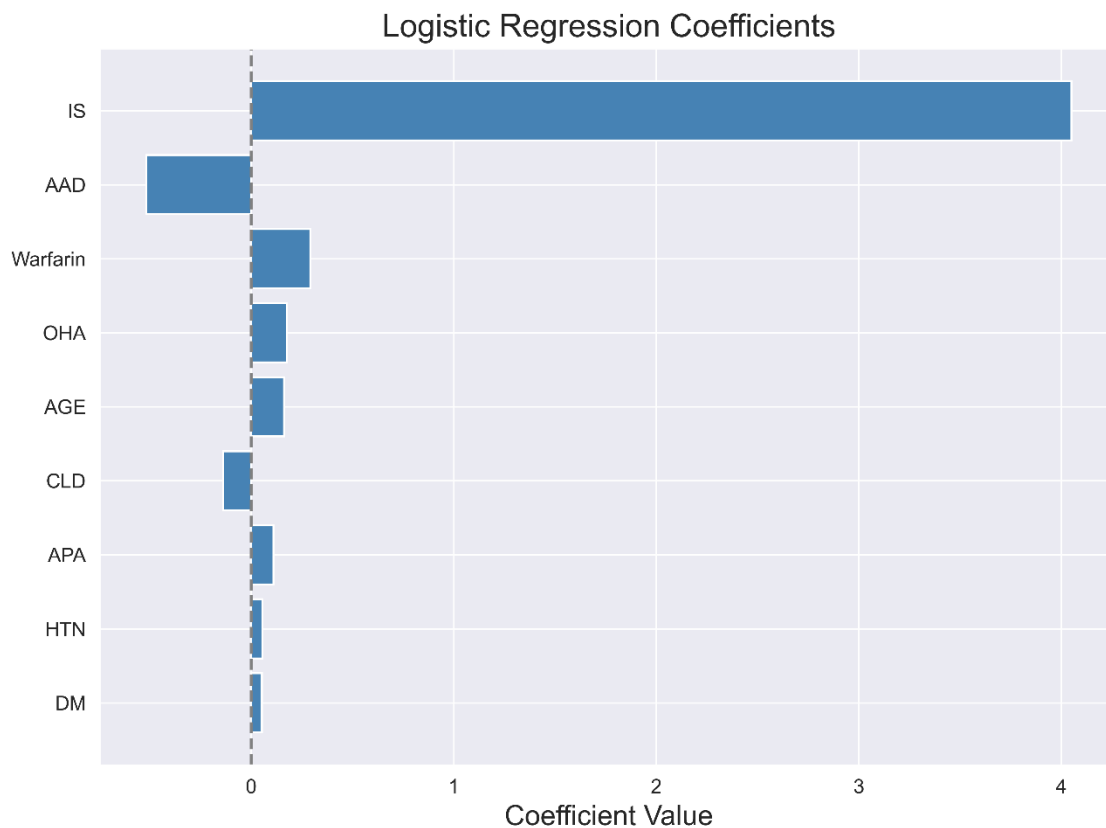
Supplementary Table 4: Definition and scoring criteria of CHA₂DS₂-VASc score for stroke risk stratification in atrial fibrillation

Items		Points
C	Congestive Heart Failure	1
H	Hypertension	1
A₂	Age ≥ 75 years	2
D	Diabetes Mellitus	1
S₂	Ischemic stroke/TIA	2
V	Vascular Disease	1
A	Age 65 - 74 years	1
Sc	Sex category (female)	1
	Maximum Score	9
	High risk of stroke events for men	≥ 2
	High risk of stroke events for women	≥ 3

Supplementary Table 5. Net Reclassification Improvement (NRI) of Logistic Regression (LR) vs. CHA₂DS₂-VASc Score at Different Decision Thresholds

LR vs CHA₂DS₂-VASc (External Validation Cohort 1)		
Threshold	Net Reclassification Improvement (Total)	<i>p</i>-value
0.1	0.4894 (95% CI [0.4098, 0.5690])	< 0.001
0.2	0.4903 (95% CI [0.4107, 0.5700])	< 0.001
0.3	0.4903 (95% CI [0.4107, 0.5700])	< 0.001
0.4	0.4903 (95% CI [0.4107, 0.5700])	< 0.001
0.5	0.4903 (95% CI [0.4107, 0.5700])	< 0.001
0.6	0.4903 (95% CI [0.4107, 0.5700])	< 0.001
0.7	0.4740 (95% CI [0.3944, 0.5537])	< 0.001
0.8	-0.1331 (95% CI [-0.2580, -0.0082])	0.037
LR vs CHA₂DS₂-VASc (External Validation Cohort 2)		
Threshold	Net Reclassification Improvement (Total)	<i>p</i>-value
0.1	0.4891 (95% CI [0.4143, 0.5639])	< 0.001
0.2	0.4891 (95% CI [0.4143, 0.5639])	< 0.001
0.3	0.4891 (95% CI [0.4143, 0.5639])	< 0.001
0.4	0.4891 (95% CI [0.4143, 0.5639])	< 0.001
0.5	0.4891 (95% CI [0.4143, 0.5639])	< 0.001
0.6	0.4891 (95% CI [0.4143, 0.5639])	< 0.001
0.7	0.4377 (95% CI [0.3586, 0.5168])	< 0.001
0.8	-0.1399 (95% CI [-0.2620, -0.0178])	0.025

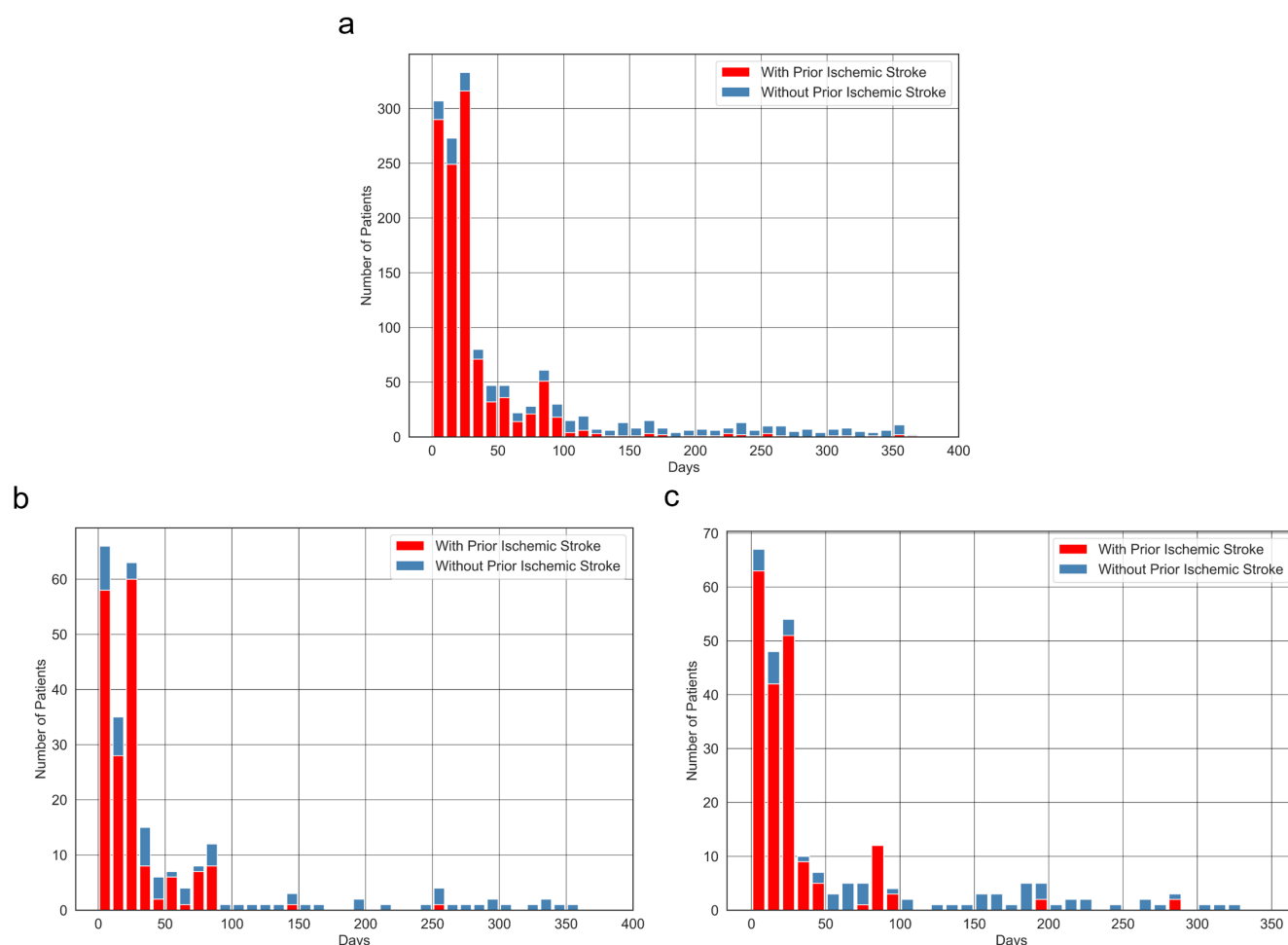
Supplementary Figure 1: Logistic regression coefficients for 1-year stroke risk prediction.



Bar plot showing the logistic regression coefficients of the 9 predictors selected via Bayesian Information Criterion (BIC)-based feature selection. Positive values indicate an increased log-odds of stroke within 1 year, while negative values reflect protective effects. Prior ischemic stroke (IS) had the strongest positive association with stroke risk, followed by use of warfarin, oral hypoglycemic agents (OHA), age, and other comorbidities. Anti-arrhythmic drug (AAD) use and chronic lung disease (CLD) were associated with reduced stroke risk.

IS indicates prior ischemic stroke; AAD, anti-arrhythmic drugs; Warfarin, warfarin use; OHA, oral hypoglycemic agents; AGE, age at diagnosis; CLD, chronic lung disease; APA, antiplatelet agents; HTN, hypertension; DM, diabetes mellitus.

Supplementary Figure 2. Distribution of stroke onset timing within 1 year after AF diagnosis, stratified by prior ischemic stroke history.



Bar plots depict the number of post-AF ischemic stroke events occurring at different time intervals (in days) following the initial AF diagnosis in:

- (a) the derivation cohort (N = 9,511),
- (b) external validation cohort 1 (N = 1,300), and
- (c) external validation cohort 2 (N = 1,242).

Events are stratified by the presence (red bars) or absence (blue bars) of prior ischemic stroke. Across all cohorts, a substantial proportion of stroke events occurred within the first 30 to 60 days, particularly among patients with a prior stroke history. These findings underscore the elevated early risk and potential diagnostic overlap in this population, reinforcing the need for careful temporal separation in model development.

Section/Topic	Item	Development / evaluation ¹	Checklist item	Reported on page
TITLE				
<i>Title</i>	1	D;E	Identify the study as developing or evaluating the performance of a multivariable prediction model, the target population, and the outcome to be predicted	Title page
ABSTRACT				
<i>Abstract</i>	2	D;E	See TRIPOD+AI for Abstracts checklist	P.2
INTRODUCTION				
<i>Background</i>	3a	D;E	Explain the healthcare context (including whether diagnostic or prognostic) and rationale for developing or evaluating the prediction model, including references to existing models	P.3
	3b	D;E	Describe the target population and the intended purpose of the prediction model in the context of the care pathway, including its intended users (e.g., healthcare professionals, patients, public)	P.3
	3c	D;E	Describe any known health inequalities between sociodemographic groups	P.3
<i>Objectives</i>	4	D;E	Specify the study objectives, including whether the study describes the development or validation of a prediction model (or both)	P.4
METHODS				
<i>Data</i>	5a	D;E	Describe the sources of data separately for the development and evaluation datasets (e.g., randomised trial, cohort, routine care or registry data), the rationale for using these data, and representativeness of the data	P.13
	5b	D;E	Specify the dates of the collected participant data, including start and end of participant accrual; and, if applicable, end of follow-up	P.13-14
<i>Participants</i>	6a	D;E	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including the number and location of centres	P.13
	6b	D;E	Describe the eligibility criteria for study participants	P.13
	6c	D;E	Give details of any treatments received, and how they were handled during model development or evaluation, if relevant	P.14
<i>Data preparation</i>	7	D;E	Describe any data pre-processing and quality checking, including whether this was similar across relevant sociodemographic groups	P.14
<i>Outcome</i>	8a	D;E	Clearly define the outcome that is being predicted and the time horizon, including how and when assessed, the rationale for choosing this outcome, and whether the method of outcome assessment is consistent across sociodemographic groups	P.14
	8b	D;E	If outcome assessment requires subjective interpretation, describe the qualifications and demographic characteristics of the outcome assessors	N.A.
	8c	D;E	Report any actions to blind assessment of the outcome to be predicted	N.A.
<i>Predictors</i>	9a	D	Describe the choice of initial predictors (e.g., literature, previous models, all available predictors) and any pre-selection of predictors before model building	P.14
	9b	D;E	Clearly define all predictors, including how and when they were measured (and any actions to blind assessment of predictors for the outcome and other predictors)	P.14
	9c	D;E	If predictor measurement requires subjective interpretation, describe the qualifications and demographic characteristics of the predictor assessors	N.A.
<i>Sample size</i>	10	D;E	Explain how the study size was arrived at (separately for development and evaluation), and justify that the study size was sufficient to answer the research question. Include details of any sample size calculation	P.13,16
<i>Missing data</i>	11	D;E	Describe how missing data were handled. Provide reasons for omitting any data	N.A.
<i>Analytical methods</i>	12a	D	Describe how the data were used (e.g., for development and evaluation of model performance) in the analysis, including whether the data were partitioned, considering any sample size requirements	P.15-16
	12b	D	Depending on the type of model, describe how predictors were handled in the analyses (functional form, rescaling, transformation, or any standardisation).	P.15-16
	12c	D	Specify the type of model, rationale ² , all model-building steps, including any hyperparameter tuning, and method for internal validation	P.15-16
	12d	D;E	Describe if and how any heterogeneity in estimates of model parameter values and model performance was handled and quantified across clusters (e.g., hospitals, countries). See TRIPOD-Cluster for additional considerations ³	P.15-17
	12e	D;E	Specify all measures and plots used (and their rationale) to evaluate model performance (e.g., discrimination, calibration, clinical utility) and, if relevant, to compare multiple models	P.16-18
	12f	E	Describe any model updating (e.g., recalibration) arising from the model evaluation, either overall or for particular sociodemographic groups or settings	N.A.
	12g	E	For model evaluation, describe how the model predictions were calculated (e.g., formula, code, object, application programming interface)	P.16
<i>Class imbalance</i>	13	D;E	If class imbalance methods were used, state why and how this was done, and any subsequent methods to recalibrate the model or the model predictions	P.16
<i>Fairness</i>	14	D;E	Describe any approaches that were used to address model fairness and their rationale	P.17
<i>Model output</i>	15	D	Specify the output of the prediction model (e.g., probabilities, classification). Provide details and rationale for any classification and how the thresholds were identified	P.16-18

¹ D=items relevant only to the development of a prediction model; E=items relating solely to the evaluation of a prediction model; D;E=items applicable to both the development and evaluation of a prediction model

² Separately for all model building approaches.

³ TRIPOD-Cluster is a checklist of reporting recommendations for studies developing or validating models that explicitly account for clustering or explore heterogeneity in model performance (eg, at different hospitals or centres). Debray et al, BMJ 2023; 380: e071018 [DOI: 10.1136/bmj-2022-071018]

<i>Training versus evaluation</i>	16	D;E	Identify any differences between the development and evaluation data in healthcare setting, eligibility criteria, outcome, and predictors	P.13-14
<i>Ethical approval</i>	17	D;E	Name the institutional research board or ethics committee that approved the study and describe the participant-informed consent or the ethics committee waiver of informed consent	P.13
OPEN SCIENCE				
<i>Funding</i>	18a	D;E	Give the source of funding and the role of the funders for the present study	P.19
<i>Conflicts of interest</i>	18b	D;E	Declare any conflicts of interest and financial disclosures for all authors	P.19
<i>Protocol</i>	18c	D;E	Indicate where the study protocol can be accessed or state that a protocol was not prepared	N.A.
<i>Registration</i>	18d	D;E	Provide registration information for the study, including register name and registration number, or state that the study was not registered	N.A.
<i>Data sharing</i>	18e	D;E	Provide details of the availability of the study data	P.18
<i>Code sharing</i>	18f	D;E	Provide details of the availability of the analytical code ⁴	P.18
PATIENT & PUBLIC INVOLVEMENT				
<i>Patient & Public Involvement</i>	19	D;E	Provide details of any patient and public involvement during the design, conduct, reporting, interpretation, or dissemination of the study or state no involvement.	P.13
RESULTS				
<i>Participants</i>	20a	D;E	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	P.4 and Fig.1
	20b	D;E	Report the characteristics overall and, where applicable, for each data source or setting, including the key dates, key predictors (including demographics), treatments received, sample size, number of outcome events, follow-up time, and amount of missing data. A table may be helpful. Report any differences across key demographic groups.	P.4, Table 1 and Table S1
	20c	E	For model evaluation, show a comparison with the development data of the distribution of important predictors (demographics, predictors, and outcome).	P.4 and Table 1
<i>Model development</i>	21	D;E	Specify the number of participants and outcome events in each analysis (e.g., for model development, hyperparameter tuning, model evaluation)	P.4-6
<i>Model specification</i>	22	D	Provide details of the full prediction model (e.g., formula, code, object, application programming interface) to allow predictions in new individuals and to enable third-party evaluation and implementation, including any restrictions to access or re-use (e.g., freely available, proprietary) ⁵	P.4-5, P.18
<i>Model performance</i>	23a	D;E	Report model performance estimates with confidence intervals, including for any key subgroups (e.g., sociodemographic). Consider plots to aid presentation.	P.5-8, and Fig. 3-5
	23b	D;E	If examined, report results of any heterogeneity in model performance across clusters. See TRIPOD Cluster for additional details ³ .	P.6-7, and Fig.3-4
<i>Model updating</i>	24	E	Report the results from any model updating, including the updated model and subsequent performance	N.A.
DISCUSSION				
<i>Interpretation</i>	25	D;E	Give an overall interpretation of the main results, including issues of fairness in the context of the objectives and previous studies	P.8-10
<i>Limitations</i>	26	D;E	Discuss any limitations of the study (such as a non-representative sample, sample size, overfitting, missing data) and their effects on any biases, statistical uncertainty, and generalizability	P.11-12
<i>Usability of the model in the context of current care</i>	27a	D	Describe how poor quality or unavailable input data (e.g., predictor values) should be assessed and handled when implementing the prediction model	P.11
	27b	D	Specify whether users will be required to interact in the handling of the input data or use of the model, and what level of expertise is required of users	P.11
	27c	D;E	Discuss any next steps for future research, with a specific view to applicability and generalizability of the model	P.11-12

From: Collins GS, Moons KGM, Dhiman P, et al. *BMJ* 2024;385:e078378. doi:10.1136/bmj-2023-078378

⁴ This relates to the analysis code, for example, any data cleaning, feature engineering, model building, evaluation.

⁵ This relates to the code to implement the model to get estimates of risk for a new individual.