**Prediction of incident atrial fibrillation in post stroke patients using machine learning: A French nationwide study**

Arnaud BISSON1,2,3, MD, Yassine LEMRINI1, MD, Wahbi EL-BOURI4, Alexandre BODIN1, MD, Denis Angoulvant1,2, MD,PhD\*, Gregory Y. H. LIP4, MD\*, Laurent FAUCHIER1, MD,PhD\*

1 - Service de Cardiologie, Centre Hospitalier Régional Universitaire et Faculté de Médecine de Tours, Tours, France

2 - EA4245, Transplantation Immunité Inflammation, Université de Tours, Tours, France

3 - Service de Cardiologie, Centre Hospitalier Régional d’Orléans, Orléans, France

4 - Liverpool Centre for Cardiovascular Science, University of Liverpool, Liverpool Heart & Chest Hospital, Liverpool, United Kingdom

[\*joint senior authors]

**Correspondence:**

Arnaud BISSON, MD

Service de Cardiologie, Centre Hospitalier Regional Universitaire et Faculté de Médecine de Tours, Tours, France

2 Boulevard Tonnellé, 37000 Tours

Tel : 33-247473659

Email: [arnaud.bisson@univ-tours.fr](mailto:arnaud.bisson@univ-tours.fr)

ORCID : 0000-0002-3449-1800

**Word count:** 3055

**Key words:** Atrial fibrillation, Machine learning, Ischemic stroke, Prediction

**Statements and Declarations:** A.Bisson has been a consultant or speaker for Astra-Zeneca, Bayer, BMS/Pfizer, Medtronic, Vitorpharma and Alnylam. D.Angoulvant has been a consultant or speaker for Amgen, Astra-Zeneca, Bayer, BMS/Pfizer, MSD, Novartis, Novo Nordisk, Sanofi, Servier. GYH.Lip has been a consultant for Bayer/Janssen, BMS/Pfizer, Medtronic, Boehringer Ingelheim, Novartis, Verseon, and Daiichi-Sankyo and a speaker for Bayer, BMS/Pfizer, Medtronic, Boehringer Ingelheim, and Daiichi-Sankyo; no fees are directly received personally. L.Fauchier has been a consultant or speaker for Bayer, BMS/Pfizer, Boehringer Ingelheim, Medtronic, and Novartis.

**Abstract:**

*Background:* Targeting ischemic strokes patients at risk of incident atrial fibrillation (AF) for prolonged cardiac monitoring and oral anticoagulation remains a challenge. Clinical risk scores have been developed to predict post stroke AF with suboptimal performances. Machine learning (ML) models are developing in the field of AF prediction and may be used to discriminate post-stroke patients at risk of new onset AF. This study aimed to evaluate ML models for the prediction of AF and to compare predictive ability to usual clinical scores.

*Methods:* Based on a French nationwide cohort of 240,459 ischemic stroke patients without AF at baseline from 2009 to 2012, ML models were trained on a train set and the best model was selected to be evaluate on the test set. Discrimination of the best model was evaluated using the C index. We finally compared our best model with previously described clinical scores.

Results: During a mean follow-up of 7.9±11.5 months, 14,095 patients (mean age 77.6±10.6; 50.3% female) developed incident AF. After training, the best ML model selected was a deep neural network with a C index of 0.77 (95% CI, 0.76-0.78) on the test set. Compared to traditional clinical scores, the selected model was statistically significantly superior to the CHA2DS2-VASc score, Framingham risk score, HAVOC score and C2HEST score (P<0.0001). The ability to predict AF was improved as shown by net reclassification index increase (P<0.0001) and decision curve analysis.

Conclusions: ML algorithms predict incident AF post stroke with a better ability than previously developed clinical scores.

**1. Introduction**

Atrial fibrillation (AF) is the most common arrhythmia effecting 2-4% of adults and is responsible for one in 5 of all ischemic strokes (IS) with a high risk of recurrent strokes [1,2] and increasing healthcare costs [3]. When compared to placebo or control, the risk of stroke in AF is reduced by 64% and all-cause mortality by 26% by oral anticoagulant (OAC) therapy [4].

In current international guidelines, OAC initiation for secondary prevention after IS is recommended only after AF documentation [5,6]. Conversely, if AF is not documented, antiplatelet agents remain recommended despite their inefficacy for stroke prevention in AF [4]. Hence, the screening for unknown AF in post stroke patients is recommended.

Several RCTs have established the effectiveness of ECG monitoring for post-stroke AF detection, with a number needed to screen of 8 to 14 [7,8]. Looking harder and longer and using more sophisticated monitoring may generally improve AF detection. Additional ECG monitoring using long-term non-invasive ECG monitors or insertable cardiac monitors should be considered to detect AF in selected post stroke patients without previously known AF. Indeed, this is a class II recommendation in the absence of strong evidence of the effect of prolonged ECG monitoring and subsequent prescription of OAC on stroke or mortality in patients with detected AF due to underpowered studies [5,6].

To improve cost effectiveness, resolve logistic difficulties and enhance efficiency, targeting best candidates for prolonged cardiac monitoring post IS seems to be necessary. However, despite the use of clinical scores such C2HEST scores to identify patients at higher risk for incident AF after IS, it still remains challenging in clinical practice [9,10]. Indeed, the efficacy of traditional clinical scores simply calculated from the sum of the points of some variables is modest at best. A limited number of covariate and discretized weights of models result in deterioration in model performance.

Machine learning (ML) is a branch of artificial intelligence (AI) that deals with the development of algorithms that use data to make predictions and to improve their accuracy without being explicitly programmed to do so [11]. Compared to traditional risk scores, ML algorithms can easily incorporate a large number of variables and then identify nonlinear and complex interactions. Accordingly, ML may improve prediction performance and better describe complex and unpredictable associations [12-13], even in stroke risk prediction [14].

The aim of this study was to evaluate and compare ML predictive models versus traditional clinical risk scores to predict AF after IS from a nationwide database.

**2. Method**

2.1. Study population

This longitudinal cohort study was based on the French hospitalization database, the PMSI (Programme de Médicalisation des Systèmes d’Information), covering hospital care across the entire population. In France, each hospital discharge from one of the 1,546 French healthcare facilities, whether public or private hospital, must be registered in the National Hospital Discharge Database. A standardized discharge summary is collected for every hospital stay and categorized into a single medical or surgical diagnosis-related group based on the International Classification of Diseases, Tenth Revision (ICD-10). A unique patient identification number make it possible to link multiple hospital stays across time corresponding to a single patient without revealing his or her identity. The reliability of PMSI data has already been assessed and used previously to study patients with stroke and AF [9, 10, 15-16].

Data for all patients admitted with ischemic stroke in France from January 2008 to December 2012 were collected from the PMSI. The study included adults (aged ≥18 years) with a diagnosis of acute ischemic stroke (ICD-10 code I63 and its subsections) coded as the primary diagnosis (ie, the health problem that justified admission to hospital), the related diagnosis (ie, potential chronic disease or health state during hospital stay), or the significantly associated diagnosis (ie, comorbidity or associated complication) who were hospitalized between January 1, 2008, and December 31, 2012. We performed an analysis restricted to the patients seen after 2009, meaning that all patients had at least 1 year in which previous events were recorded to establish history of previous AF and comorbidities. Patients with no diagnosis of AF (ICD-10 code I48 and its subsections) were considered to have sinus rhythm (**Figure** **1**).

The medical information contained in the database is anonymous and protected by professional confidentiality. Consequently, ethics review was not required and patient consent was not sought. The study was conducted retrospectively, patients were not involved in its conduct, and there was no impact on their care. This type of study was approved by the institutional review board of the Pole Coeur Thorax Vaisseaux from the Trousseau University Hospital (Tours, France) on December 1, 2015, and registered as a clinical audit. Procedures for data collection and management were approved by the Conseil National de l’Informatique et des Libertés (CNIL), the independent national ethics committee protecting human rights in France, which ensures that all information is kept confidential and anonymous (authorization no. 1749007).

2.2. Feature selection and model development

From the imputed dataset, continuous variables were centralized to the mean and scaled to the standard deviation, whereas categorical variables were coded into binary numbers (0 and 1). Then, the whole dataset was randomly split into derivation (training set) and validation cohort (test set) (7:3). To increase the prediction accuracy, only most critical features in the derivation cohort were selected for analysis. A Chi2 test was used since it is a suitable method for categorical input and output. The 12 best variables were introduced into five well-accepted ML models to assess post stroke AF risk. There includes logistic regression with L2 regularization (LR), random forest classifier (RFC), K-nearest neighbors (KNN), extreme gradient boosting (XGBoost), and deep neural network (DNN). ML models were implemented in Python using open-source packages: Scikit-learn version 1.1.1 and XGboost version 1.6.1.

ML models were trained and optimized on the derivation cohort (training set) using a random search algorithm with 5-fold cross-validation: the training set was split into 5-fold validation sets. For each validation set, the corresponding training set was given by the remaining 4 folds. Each candidate model was fitted and hyper-parameters tuned in each training set. The performance measure was then performed on the validation sets (**Figure 2**). Predictive ability of each ML models was assessed using C index and compared to each other.

The model with the best performance was then selected for analysis on the held-out validation cohort for final evaluation (test set) to evaluate the models' generalization performance and compared to clinical scores.

2.3. Model evaluation and comparison

Area under the ROC curve (AUC) of the best ML model selected was finally calculated on the validation cohort and C index was compared to predictive ability of clinical scores previously described such as CHA2DS2-VASc score (congestive heart failure, hypertension, age ≥ 75 [*2 points*], diabetes, stroke [*2 points*], vascular disease, age 65-74, female), C2HEST score (coronary artery disease / chronic obstructive pulmonary disease [*1 point each*], hypertension [*1 point*], elderly [age ≥ 75, *2 points*], systolic HF [*2 points*], thyroid disease [*1 point*]), HAVOC score (hypertension [*2 points*], age ≥75 years [*2 points*], valvular heart disease [*2 points*], peripheral vascular disease [*1 point*], obesity [*1 point*], congestive heart failure [*4 points*] and coronary artery disease [*2 points*])

and modified Framingham risk score in the validation cohort [10, 17-20].

2.4. Statistical analysis

Qualitative variables were described using counts and percentages, and continuous quantitative variables were described as mean±SD or median [interquartile range]. Comparisons were made using parametric or nonparametric tests, as appropriate: The Wilcoxon signed rank and Kruskal Wallis tests were used for comparing values between 2 independent groups, and the Chi2 test was used to compare categorical data. Receiver operating characteristic curves were constructed, and Harrell C indexes (ie, AUC) were calculated as a measure of model performance and compared using the DeLong test [21]. Net reclassification improvement was calculated according to the methods described by Pencina et al to assess the reclassification performance of the scores [22]. Clinical usefulness and net benefit of the best ML model in comparison to the best clinical score were estimated using decision curve analysis [23].

Analyses were performed using Python version 3.09 and STATA v16.0 (StataCorp). All statistical significance levels were two-sided, and the significant differences were expressed as *p* < 0.05.

**3. Results**

Among 240,306 patients included for analysis, 14,095 patients (5.9%) (mean age 77.6±10.6; 50.3% female) developed incident AF after a mean follow-up of 7.5 ±11.1 months. Baseline characteristics are presented in **Table 1**. Patients with AF were significantly older than those without AF and more frequently female (P<0.0001 both). The prevalence of each comorbidity was higher in AF patients, including hypertension, diabetes mellitus, coronary arterial disease, valve disease, hyperlipidemia, vascular disease, chronic obstructive pulmonary disease, renal dysfunction, thyroid disease, and HF (P<0.0001, respectively). Consequently, patients who developed AF had higher clinical scores (CHA2DS2-VASc score, C2HEST scores, HAVOC score and modified Framingham risk score) at baseline than those who did not (P<0.0001).

Most critical features selected for analysis were age, hypertension, congestive heart failure, renal dysfunction, anemia, lung disease, thyroid disease, dyslipidemia, vascular disease, coronary artery disease, pacemaker/implantable cardiac defibrillator (PM-ICD) and valvular heart disease.

After hyperparameter tuning using the derivation cohort composed of 168,214 randomly assigned patients (training set), ML algorithms had good predictive performances based on C indexes: LR 0.761 (95% CI, 0.757- 0.769), KNN 0.741 (95% CI, 0.736 - 0.752), RF 0.763 (95% CI, 0.760 - 0.769), XGBoost 0.758 (95% CI, 0.750 – 0.766), DNN 0.766 (95% CI, 0.763 - 0.771).

The DNN model showed the best predictive performances and was therefore selected as the best model for further evaluation (P=0.002 vs LR, P<0.0001 vs KNN, P=0.02 vs RF, P=0.002 vs XGBoost). Hyperparameters are displayed in **Supplemental table** **1**.

Once evaluated on the held-out validation cohort composed of the 72,092 remaining patients (test set), DNN had a good C index: 0.766 (95% CI, 0.759 - 0.773). The incidence of AF roses in a stepwise fashion from 19.9 per 1000 patient-years for the first quartile to 175.8 per 1000 for the fourth quartile (**Table 2**, **Figure 3**).

When compared to the previously described clinical scores, DNN showed statistically significantly higher predictive ability: CHA2DS2-VASc score 0.702 (95% CI, 0.694 - 0.709), C2HEST scores 0.735 (95% CI, 0.727 - 0.742), HAVOC score 0.738 (95% CI, 0.730 - 0.745) and modified Framingham risk score 0.720 (95% CI, 0.712 - 0.727) (P<0.0001) (**Supplemental** **table 2**, **Figure 4a**).

The DNN had category free net reclassification improvement compared with the CHA2DS2-VASc score (64% [P<0.0001]), C2HEST score (52 % [P<0.0001]), HAVOC score (38% [P<0.0001]) and modified Framingham risk score (48% [P<0.0001]). Using decision curve analysis, our DNN showed better clinical usefulness compared with the traditional scores (**Figure 4b**).

**4. Discussion**

This study is the first application of ML on a nationwide hospital data base to derive and validate the usefulness of detecting post-stroke AF. Among the five ML algorithms trained, the DNN model was selected as the best model with good performance for discriminating patients at risk of developing AF after stroke as well as in the hold out validation cohort. When compared to usual clinical risk scores, this approach had a significantly better predictive ability to target these patients at risk of arrhythmia.

Using clinical scores based on cardiac or extra cardiac comorbidities to predict AF occurrence after IS is a well-established concept [24]. Preexisting comorbidities may lead to left atrium overload, enlargement and fibrosis and in fine to a true atrial cardiomyopathy responsible for AF and IS. This atrial substrate supports the fact that patients with so-called new onset AF more often had a history of transient ischemic attack suggests and that AF was previously unknown rather than being true new-onset AF.

The model developed in the present study was based on these comorbidities as features. Some comorbidities used were intuitively associated to AF development and were previously described and integrated in clinical scores such as age, hypertension, congestive heart failure, renal dysfunction, vascular disease, coronary artery disease or valvular heart disease. However, more unusual extra cardiac comorbidities were able to improve predictive ability of our models (anemia, lung disease, thyroid disease, dyslipidemia), as well as dynamic changes in risk with ageing and incident comorbidities, may need further investigations.

Over the last decade, ML has captured the interest of the medical and healthcare community with an explosion of publications using ML in the cardiovascular field and especially in AF research [25]. This approach is currently improving AF understanding and prediction and is already outperforming traditional predictive risk scores in a broad range of clinical settings [12,13,26-29]. Due to the complex etiology of AF and the highlight of ML algorithms in handling complex relationships between large numbers of variables, models constructed using ML methods may show better predictive performance in AF risk assessing than those models using classical statistical methods. One particular benefit of ML approaches is the ability to extract higher level features from the input data to learn from. The algorithm is able to learn how to classify imaging data. As such, ML approaches may have superior performance when the inputs are complex, the features of the input data cannot be readily discerned, or the relationships between input data are complex and nonlinear. Moreover, finding the ultimate combination of variables using a method often derived from multivariate analysis is not always easy. The advantage of the approach used in this study, where several algorithms based on different mathematic concepts (neural network, ensemble learning method, linear models…) were trained, is to optimize chances to find the most informative model.

The predictive ability of traditional scores in our held-out validation cohort (test set) used of for external validation was consistent with values described in other studies and our best ML model was able to outperform all of them [9,10]. Our best model was a DNN with a C index of 0.766 (95% CI, 0.759 - 0.773). Indeed, such an algorithm is the most widely used ML algorithm for research in the field of AF with an excellent efficacy compared to other algorithms [25]. Moreover, predictive ability of our model based on clinical characteristics was similar than other models based on echocardiography, baseline holter ECG or biomarkers [30].

This was also reported by Zheng *et al* who developed a ML algorithm for AF prediction with better predictive performances than traditional risk scores (C index of 0.922 for DNN) [27]. The excellent predictive performance of their model when compared to our model is also due to key features such as NIHSS score (National Institutes of Health Stroke Scale) for stroke severity and stroke etiology which were missing on our data base. Moreover, the gap between ML model and traditional clinical scores described by X. Zheng *et al* was wider when compared to our study. Such improvement of prediction can also be explained by the integration of continuous features where traditional risk scores dichotomized continuous variables, which lost information. Except for age, our database did not contain any continuous variable reducing the improvement of prediction. Moreover, C index of traditional clinical risk scores were very low in this study (eg. 0.578 for HAVOC score, 0.572 for CHA2DS2-VASc score), notwithstanding the caveat that some scores (CHA2DS2-VASc) were design for stroke risk stratification and not the prediction of incident AF. Indeed, a recent EHRA position paper appealed for the appropriate use of the score for the condition it was designed for [31].

External performance of our DNN model in the held-out validation cohort (test set) was similar to that described in the derivation cohort (train set) (0.766 (95% CI, 0.763 - 0.771) versus 0.766 (95% CI, 0.759 - 0.773)) demonstrating the absence of overfitting to the data thanks to the split and K fold cross validation methodology.

**Potential Clinical practical implication and perspectives**

The question of OAC in post stroke patients is still debated. However, guidelines recommend to diagnose AF before introducing OAC for secondary prevention. Indeed, two RCTs (NAVIGATE and RESPECT ESUS trials) aiming to demonstrate the efficacy of OAC after stroke without previously documented AF were negative and therefore did not support this strategy [32-33]. More recently, ATTICUS randomized trial, which enrolled ESUS patients with supposed risk profile for cardiac thromboembolism, was prematurely stopped for futility [34]. Another ongoing trial may be able to provide new insights on post stroke atrial cardiopathy (ARCADIA (NCT03192215)). Diagnosing AF previously to introduce OAC seems therefore mandatory. It is well known since CRYSTAL AF and EMBRACE that the longer is the monitoring for searching AF after stroke, the more AF is detected [8,35,36]. However, longer monitoring incurs burden in terms of healthcare work and costs. Therefore, the identification of biomarkers that can potentially increase the diagnostic yield (and thus cost-effectiveness) of AF screening in patients with stroke is intensively researched but still challenging.

Although blood, cardiac and neuroimaging biomarkers have been identified as promising independent predictors of AF detection in patients with stroke, they have insufficient negative predictive value to avoid missing sub-clinical AF in patients with a lower but still clinically important chance of detecting it and their use in clinical practice is therefore not retained in the current guidelines [37]. Despite their moderate performance, clinical risk scores can be used for ‘high risk’ patient identification. Based on the same easily available patient characteristics, ML models could further improve in the future risk-stratification and decision making about AF screening strategy after IS.

**5. Strengths and Limitations**

This is the first time that a large nationwide database is used for an analysis by ML algorithms.

Despite a large amount of data, ideal to feed and train ML algorithms, the nature of PMSI data base derived from ICD-10 codes and therefore composed almost exclusively by categorical features minimize algorithms performances because of a lack of granularity.

Moreover, diagnoses during outpatient visits were not included in our analysis and this possibly underestimated the true incidence of AF. Echocardiographic, biological and brain imaging parameters were also lacking. Whilst the ML models were statistically superior, the implementation of such models into clinical practice would require Health solutions or backend solutions with explainable AI integrated into electronic health records.

**6. Conclusion**

In a nationwide cohort of IS, using a ML approach demonstrates effective risk prediction for post stroke AF. DNN model achieved best prediction performance compared with other ML algorithms and traditional clinical risk scores. This promising method may help risk-stratification for decision making in relation to a screening strategy for AF in post stroke patients.

**7. References**

1. Benjamin, Emelia J., Paul Muntner, Alvaro Alonso, Marcio S. Bittencourt, Clifton W. Callaway, April P. Carson, Alanna M. Chamberlain, et al. 2019. “Heart Disease and Stroke Statistics-2019 Update: A Report From the American Heart Association.” *Circulation* 139 (10): e56–528. <https://doi.org/10.1161/CIR.0000000000000659>.

2. Y, Guo, Tian Y, Wang H, Si Q, Wang Y, and Lip Gyh. 2015. “Prevalence, Incidence, and Lifetime Risk of Atrial Fibrillation in China: New Insights into the Global Burden of Atrial Fibrillation.” *Chest* 147 (1): 109–19. <https://doi.org/10.1378/chest.14-0321>.

3. Burdett, Paul, and Gregory Y. H. Lip. 2022. “Atrial Fibrillation in the UK: Predicting Costs of an Emerging Epidemic Recognizing and Forecasting the Cost Drivers of Atrial Fibrillation-Related Costs.” *European Heart Journal. Quality of Care & Clinical Outcomes* 8 (2): 187–94. <https://doi.org/10.1093/ehjqcco/qcaa093>.

4. Hart, Robert G., Lesly A. Pearce, and Maria I. Aguilar. 2007. “Meta-Analysis: Antithrombotic Therapy to Prevent Stroke in Patients Who Have Nonvalvular Atrial Fibrillation.” *Annals of Internal Medicine* 146 (12): 857–67. <https://doi.org/10.7326/0003-4819-146-12-200706190-00007>.

5. Hindricks, Gerhard, Tatjana Potpara, Nikolaos Dagres, Elena Arbelo, Jeroen J. Bax, Carina Blomström-Lundqvist, Giuseppe Boriani, 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): 373–498. <https://doi.org/10.1093/eurheartj/ehaa612>.

6. Chao, Tze-Fan, Boyoung Joung, Yoshihide Takahashi, Toon Wei Lim, Eue-Keun Choi, Yi-Hsin Chan, Yutao Guo, et al. 2022. “2021 Focused Update Consensus Guidelines of the Asia Pacific Heart Rhythm Society on Stroke Prevention in Atrial Fibrillation: Executive Summary.” *Thrombosis and Haemostasis* 122 (1): 20–47. <https://doi.org/10.1055/s-0041-1739411>.

7. Kishore, Amit, Andy Vail, Arshad Majid, Jesse Dawson, Kennedy R. Lees, Pippa J. Tyrrell, and Craig J. Smith. 2014. “Detection of Atrial Fibrillation after Ischemic Stroke or Transient Ischemic Attack: A Systematic Review and Meta-Analysis.” *Stroke* 45 (2): 520–26. <https://doi.org/10.1161/STROKEAHA.113.003433>.

8. Sposato, Luciano A., Lauren E. Cipriano, Gustavo Saposnik, Estefanía Ruíz Vargas, Patricia M. Riccio, and Vladimir Hachinski. 2015. “Diagnosis of Atrial Fibrillation after Stroke and Transient Ischaemic Attack: A Systematic Review and Meta-Analysis.” *The Lancet. Neurology* 14 (4): 377–87. <https://doi.org/10.1016/S1474-4422(15)70027-X>.

9. Bisson, Arnaud, Alexandre Bodin, Nicolas Clementy, Dominique Babuty, Gregory Y. H. Lip, and Laurent Fauchier. 2018. “Prediction of Incident Atrial Fibrillation According to Gender in Patients With Ischemic Stroke From a Nationwide Cohort.” *The American Journal of Cardiology* 121 (4): 437–44. <https://doi.org/10.1016/j.amjcard.2017.11.016>.

10. Li, Yan-Guang, Arnaud Bisson, Alexandre Bodin, Julien Herbert, Leslie Grammatico-Guillon, Boyoung Joung, Yu-Tang Wang, Gregory Y. H. Lip, and Laurent Fauchier. 2019. “C2 HEST Score and Prediction of Incident Atrial Fibrillation in Poststroke Patients: A French Nationwide Study.” *Journal of the American Heart Association* 8 (13): e012546. <https://doi.org/10.1161/JAHA.119.012546>.

11. Bishop CM. Pattern Recognition and Machine Learning, 3rd edn. New York, NY: Springer; 2006.

12. Hill, Nathan R., Daniel Ayoubkhani, Phil McEwan, Daniel M. Sugrue, Usman Farooqui, Steven Lister, Matthew Lumley, et al. 2019. “Predicting Atrial Fibrillation in Primary Care Using Machine Learning.” *PLoS ONE* 14 (11): e0224582. <https://doi.org/10.1371/journal.pone.0224582>.

13. Sekelj, Sara, Belinda Sandler, Ellie Johnston, Kevin G Pollock, Nathan R Hill, Jason Gordon, Carmen Tsang, Sadia Khan, Fu Siong Ng, and Usman Farooqui. 2020. “Detecting Undiagnosed Atrial Fibrillation in UK Primary Care: Validation of a Machine Learning Prediction Algorithm in a Retrospective Cohort Study.” *European Journal of Preventive Cardiology*, August, 2047487320942338. <https://doi.org/10.1177/2047487320942338>.

14. Lip, Gregory Y. H., Ash Genaidy, George Tran, Patricia Marroquin, Cara Estes, and Sue Sloop. 2022. “Improving Stroke Risk Prediction in the General Population: A Comparative Assessment of Common Clinical Rules, a New Multimorbid Index, and Machine-Learning-Based Algorithms.” *Thrombosis and Haemostasis* 122 (1): 142–50. <https://doi.org/10.1055/a-1467-2993>.

15. Couris, Chantal Marie, Stephanie Polazzi, Frederic Olive, Laurent Remontet, Nadine Bossard, Frederic Gomez, Anne-Marie Schott, Nicolas Mitton, Marc Colonna, and Beatrice Trombert. 2009. “Breast Cancer Incidence Using Administrative Data: Correction with Sensitivity and Specificity.” *Journal of Clinical Epidemiology* 62 (6): 660–66. <https://doi.org/10.1016/j.jclinepi.2008.07.013>.

16. Fauchier, Laurent, Nicolas Clementy, Christele Pelade, Cecile Collignon, Emmanuelle Nicolle, and Gregory Y. H. Lip. 2015. “Patients With Ischemic Stroke and Incident Atrial Fibrillation: A Nationwide Cohort Study.” *Stroke; a Journal of Cerebral Circulation* 46 (9): 2432–37. <https://doi.org/10.1161/STROKEAHA.115.010270>.

17. Lip, Gregory Y. H., Robby Nieuwlaat, Ron Pisters, Deirdre A. Lane, and Harry J. G. M. 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): 263–72. <https://doi.org/10.1378/chest.09-1584>.

18. Li, Yan-Guang, Daniele Pastori, Alessio Farcomeni, Pil-Sung Yang, Eunsun Jang, Boyoung Joung, Yu-Tang Wang, Yu-Tao Guo, and Gregory Y. H. Lip. 2019. “A Simple Clinical Risk Score (C2HEST) for Predicting Incident Atrial Fibrillation in Asian Subjects: Derivation in 471,446 Chinese Subjects, With Internal Validation and External Application in 451,199 Korean Subjects.” *Chest* 155 (3): 510–18. <https://doi.org/10.1016/j.chest.2018.09.011>.

19. Kwong, Calvin, Albee Y. Ling, Michael H. Crawford, Susan X. Zhao, and Nigam H. Shah. 2017. “A Clinical Score for Predicting Atrial Fibrillation in Patients with Cryptogenic Stroke or Transient Ischemic Attack.” *Cardiology* 138 (3): 133–40. <https://doi.org/10.1159/000476030>.

20. Schnabel, Renate B., Lisa M. Sullivan, Daniel Levy, Michael J. Pencina, Joseph M. Massaro, Ralph B. D’Agostino, Christopher Newton-Cheh, et al. 2009. “Development of a Risk Score for Atrial Fibrillation (Framingham Heart Study): A Community-Based Cohort Study.” *Lancet (London, England)* 373 (9665): 739–45. <https://doi.org/10.1016/S0140-6736(09)60443-8>.

21. DeLong, E. R., D. M. DeLong, and D. L. Clarke-Pearson. 1988. “Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach.” *Biometrics* 44 (3): 837–45.

22. Pencina, Michael J., Ralph B. D’Agostino, Ralph B. D’Agostino, and Ramachandran S. Vasan. 2008. “Evaluating the Added Predictive Ability of a New Marker: From Area under the ROC Curve to Reclassification and Beyond.” *Statistics in Medicine* 27 (2): 157–72; discussion 207-212. <https://doi.org/10.1002/sim.2929>.

23. Vickers AJ, Cronin AM, Elkin EB, Gonen M. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. BMC Med Inform Decis Mak. 2008;8:53.

24. Bisson, Arnaud, Nicolas Clementy, Alexandre Bodin, Denis Angoulvant, Dominique Babuty, Gregory Y. H. Lip, and Laurent Fauchier. 2017. “Relationship of Preexisting Cardiovascular Comorbidities to Newly Diagnosed Atrial Fibrillation After Ischemic Stroke.” *Stroke* 48 (10): 2878–80. <https://doi.org/10.1161/STROKEAHA.117.018251>.

25. Olier, Ivan, Sandra Ortega-Martorell, Mark Pieroni, and Gregory Y H Lip. 2021. “How Machine Learning Is Impacting Research in Atrial Fibrillation: Implications for Risk Prediction and Future Management.” *Cardiovascular Research* 117 (7): 1700–1717. <https://doi.org/10.1093/cvr/cvab169>.

26. Chua, Winnie, Yanish Purmah, Victor R. Cardoso, Georgios V. Gkoutos, Samantha P. Tull, Georgiana Neculau, Mark R. Thomas, et al. 2019. “Data-Driven Discovery and Validation of Circulating Blood-Based Biomarkers Associated with Prevalent Atrial Fibrillation.” *European Heart Journal* 40 (16): 1268–76. <https://doi.org/10.1093/eurheartj/ehy815>.

27. Zheng, Xiaohan, Fusang Wang, Juan Zhang, Xiaoli Cui, Fuping Jiang, Nihong Chen, Junshan Zhou, Jinsong Chen, Song Lin, and Jianjun Zou. 2022. “Using Machine Learning to Predict Atrial Fibrillation Diagnosed after Ischemic Stroke.” *International Journal of Cardiology* 347 (January): 21–27. <https://doi.org/10.1016/j.ijcard.2021.11.005>.

28. Lip, Gregory Y. H., George Tran, Ash Genaidy, Patricia Marroquin, Cara Estes, and Jeremy Landsheft. 2022. “Improving Dynamic Stroke Risk Prediction in Non-Anticoagulated Patients with and without Atrial Fibrillation: Comparing Common Clinical Risk Scores and Machine Learning Algorithms.” *European Heart Journal. Quality of Care & Clinical Outcomes* 8 (5): 548–56. <https://doi.org/10.1093/ehjqcco/qcab037>.

29. Lip, Gregory Y. H., Ash Genaidy, George Tran, Patricia Marroquin, and Cara Estes. 2021. “Incident Atrial Fibrillation and Its Risk Prediction in Patients Developing COVID-19: A Machine Learning Based Algorithm Approach.” *European Journal of Internal Medicine* 91 (September): 53–58. <https://doi.org/10.1016/j.ejim.2021.04.023>.

30. Skrebelyte-Strøm, Loreta, Ole Morten Rønning, Fredrik A. Dahl, Kjetil Steine, and Harald Kjekshus. 2022. “Prediction of Occult Atrial Fibrillation in Patients after Cryptogenic Stroke and Transient Ischaemic Attack: PROACTIA.” *Europace: European Pacing, Arrhythmias, and Cardiac Electrophysiology: Journal of the Working Groups on Cardiac Pacing, Arrhythmias, and Cardiac Cellular Electrophysiology of the European Society of Cardiology*, July, euac092. <https://doi.org/10.1093/europace/euac092>.

31. Nielsen, Jens Cosedis, Yenn-Jiang Lin, Marcio Jansen de Oliveira Figueiredo, Alireza Sepehri Shamloo, Alberto Alfie, Serge Boveda, Nikolaos Dagres, et al. 2020. “European Heart Rhythm Association (EHRA)/Heart Rhythm Society (HRS)/Asia Pacific Heart Rhythm Society (APHRS)/Latin American Heart Rhythm Society (LAHRS) Expert Consensus on Risk Assessment in Cardiac Arrhythmias: Use the Right Tool for the Right Outcome, in the Right Population.” *Europace: European Pacing, Arrhythmias, and Cardiac Electrophysiology: Journal of the Working Groups on Cardiac Pacing, Arrhythmias, and Cardiac Cellular Electrophysiology of the European Society of Cardiology* 22 (8): 1147–48. <https://doi.org/10.1093/europace/euaa065>.

32. Hart, Robert G., Mukul Sharma, Hardi Mundl, Scott E. Kasner, Shrikant I. Bangdiwala, Scott D. Berkowitz, Balakumar Swaminathan, et al. 2018. “Rivaroxaban for Stroke Prevention after Embolic Stroke of Undetermined Source.” *New England Journal of Medicine* 378 (23): 2191–2201. <https://doi.org/10.1056/NEJMoa1802686>.

33. Diener, Hans-Christoph, Ralph L. Sacco, J. Donald Easton, Christopher B. Granger, Richard A. Bernstein, Shinichiro Uchiyama, Jörg Kreuzer, et al. 2019. “Dabigatran for Prevention of Stroke after Embolic Stroke of Undetermined Source.” *New England Journal of Medicine* 380 (20): 1906–17. <https://doi.org/10.1056/NEJMoa1813959>.

34. Poli, S, C Meissner, H J Baezner, A Kraft, F Hillenbrand, C Hobohm, J Liman, et al. 2021. “Apixaban for Treatment of Embolic Stroke of Undetermined Source (ATTICUS) Randomized Trial – Update of Patient Characteristics and Study Timeline after Interim Analysis.” *European Heart Journal* 42 (Supplement\_1): ehab724.2070. <https://doi.org/10.1093/eurheartj/ehab724.2070>.

35. Sanna, Tommaso, Hans-Christoph Diener, Rod S. Passman, Vincenzo Di Lazzaro, Richard A. Bernstein, Carlos A. Morillo, Marilyn Mollman Rymer, et al. 2014. “Cryptogenic Stroke and Underlying Atrial Fibrillation.” *New England Journal of Medicine* 370 (26): 2478–86. <https://doi.org/10.1056/NEJMoa1313600>.

36. Gladstone, David J., Melanie Spring, Paul Dorian, Val Panzov, Kevin E. Thorpe, Judith Hall, Haris Vaid, et al. 2014. “Atrial Fibrillation in Patients with Cryptogenic Stroke.” *New England Journal of Medicine* 370 (26): 2467–77. <https://doi.org/10.1056/NEJMoa1311376>.

37. Rubiera, Marta, Ana Aires, Kateryna Antonenko, Sabrina Lémeret, Christian H Nolte, Jukka Putaala, Renate B Schnabel, et al. 2022. “European Stroke Organisation (ESO) Guideline on Screening for Subclinical Atrial Fibrillation after Stroke or Transient Ischaemic Attack of Undetermined Origin.” *European Stroke Journal*, May, 23969873221099480. <https://doi.org/10.1177/23969873221099478>.

**Table 1.** Baseline characteristics of the patients with ischemic stroke and no known AF at baseline according to AF occurrence during follow-up.

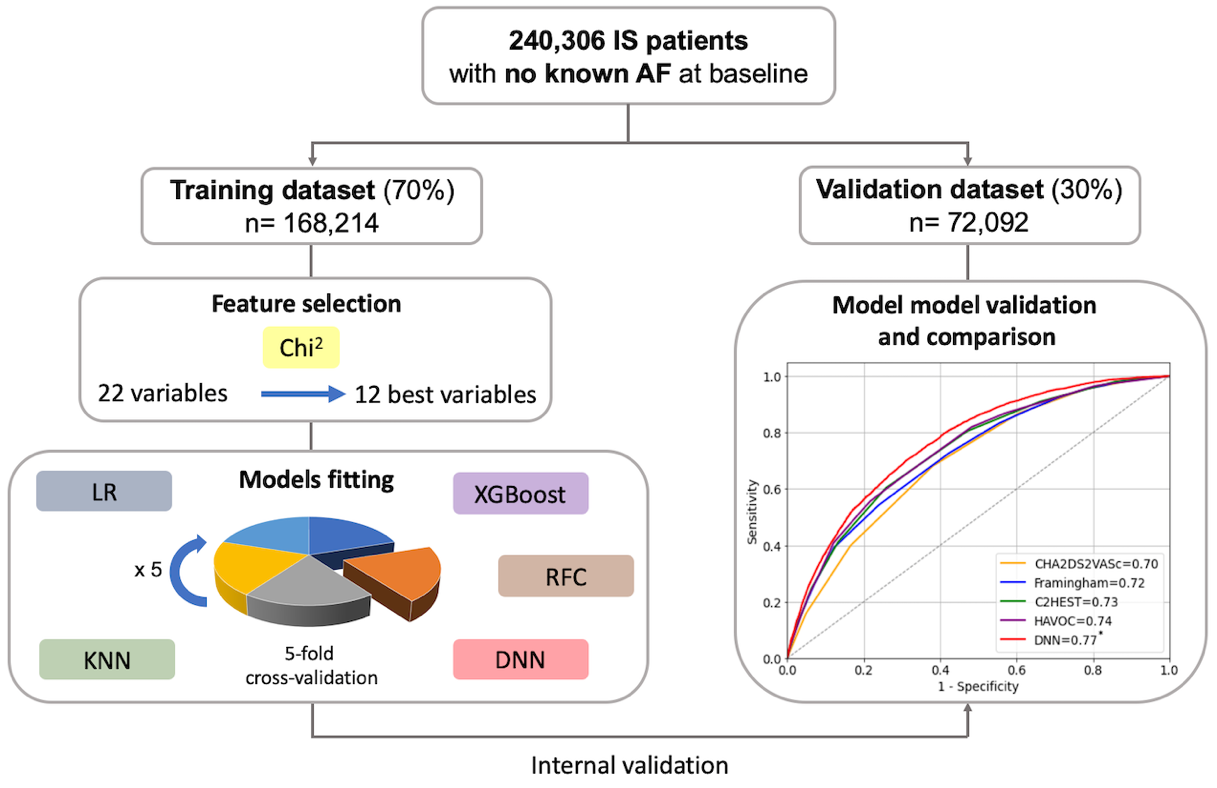
|  |  |  |  |
| --- | --- | --- | --- |
| Variables | AF during follow-up  n=14,095 | No AF during follow-up  n=226,364 | p value |
| Age, years  Age ≥75 years old, n (%)  Gender (female), n (%)  Underlying diseases, n (%)  Hypertension  Diabetes mellitus  Congestive heart failure  Vascular disease  CHA2DS2-VASc score, n (%)  Score=2  Score=3  Score=4  Score=5  Score=6  Score=7  Score=8  Score=9 | 77.6±10.6  9,755(69,2%)  7,082(50.3%)  11,745(83.3%)  4,083(29%)  6,261(44.4%)  6,907(49%)  161(1.1%)  523(3.7%)  1,192(8.5%)  2,596(18.4%)  3,903(27.7%)  3,441(24.4%)  1,881(13.4%)  398(2.8%) | 70.8±15.7  110,782(49%)  107,266(47.4%)  141,045(62.3%)  50,977(22.5%)  33,162(14.7%)  70,636(31.2%)  16,992(7.5%)  33,266(14.7%)  39,300(17.4%)  50,585(22.4%)  48,443(21.4%)  26,634(11.8%)  9,214(4.1%)  1,777(0.8%) | <0.0001  <0.0001  <0.0001  <0.0001  <0.0001  <0.0001  <0.0001  <0.0001  <0.0001  <0.0001  <0.0001  <0.0001  <0.0001  <0.0001  <0.0001 |
| Comorbidities  Systemic embolism  Coronary artery disease | 807(5.7%)  4,969(35.3%) | 6,261(2.8%)  39,652(17.5%) | <0.0001  <0.0001 |
| Obesity | 2,071(14.7%) | 22,901(10.1%) | <0.0001 |
| Abnormal renal function | 5,393(38.3%) | 38,618(17.1%) | <0.0001 |
| Liver disease | 593(4.2%) | 6,705(3%) | <0.0001 |
| Anaemia | 3,980(28.2%) | 31,165(13.8%) | <0.0001 |
| Lung disease | 3,661(26%) | 35,320(15.6%) | <0.0001 |
| Cancer within preceding 5 years | 3,056(21.7%) | 37,400(16.5%) | <0.0001 |
| Inflammatory diseases | 1,470(10.4%) | 14,186(6.3%) | <0.0001 |
| Alcohol-related diagnoses | 954(6.8%) | 17,680(7.8%) | <0.0001 |
| Thyroid disease  Dyslipidaemia | 6,040(42.9%)  5,793(41.1%) | 68,247(30.2%)  69,428(30.7%) | <0.0001  <0.0001 |
| PM-ICD | 1,643(11.7%) | 7,201(3.2%) | <0.0001 |
| Valvular disease | 2,780(19.7%) | 15,121(6.7%) | <0.0001 |
| Tobacco smoking | 1,415(10%) | 28,840(12.7%) | <0.0001 |

**Table 2.** Odds ratios and post stroke atrial fibrillation prevalence according scores.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Scores | OR (95% CI) | AF prevalence per quartile (%) | | | |
| Q1 | Q2 | Q3 | Q4 |
| C2HEST\* | 2.32 (2.24- 2.4) | 1.7 | 3.4 | 6.9 | 16.4 |
| CHA2DS2VASc\* | 1.87 (1.82 - 1.92) | 2.1 | 5 | 7.7 | 13.3 |
| HAVOC\* | 1.94 (1.89 – 2) | 1.9 | 5.1 | 6.1 | 14 |
| Framingham\* | 2.24 (2.16 - 2.32) | 1.6 | 4.2 | 6.6 | 17 |
| DNN | 2.54 (2.45 - 2.63) | 1 | 2.7 | 5.9 | 14.3 |

\* P<0.0001 vs DNN. AF: Atrial Fibrillation; CI: Confidence interval; DNN: Deep Neural Network OR: Odds ratio;

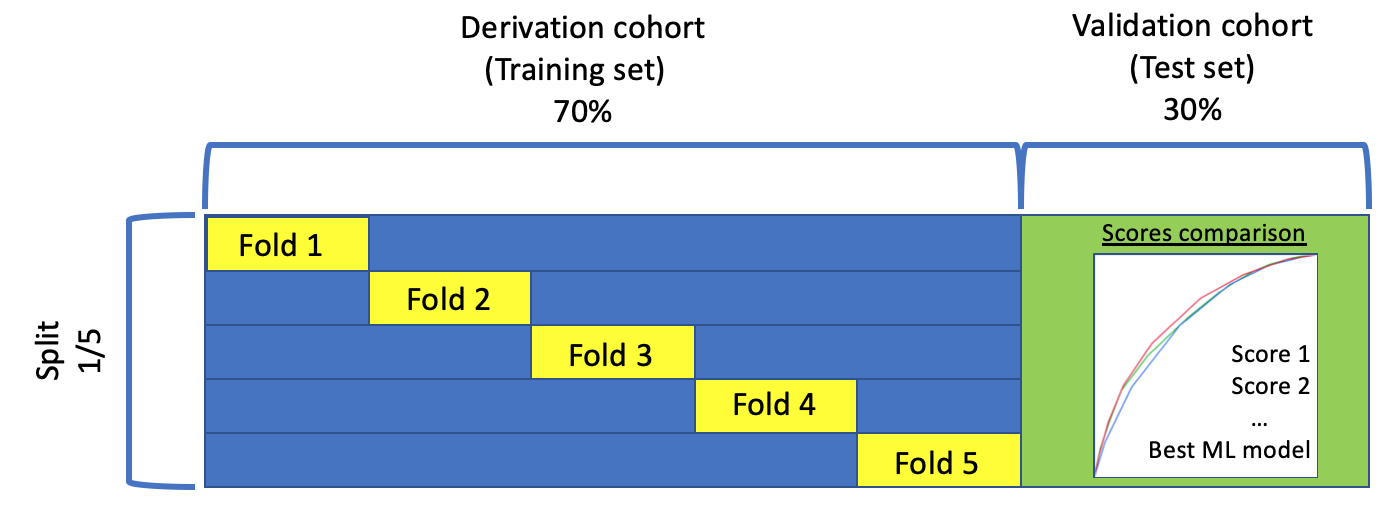
`



**Graphic abstract.**

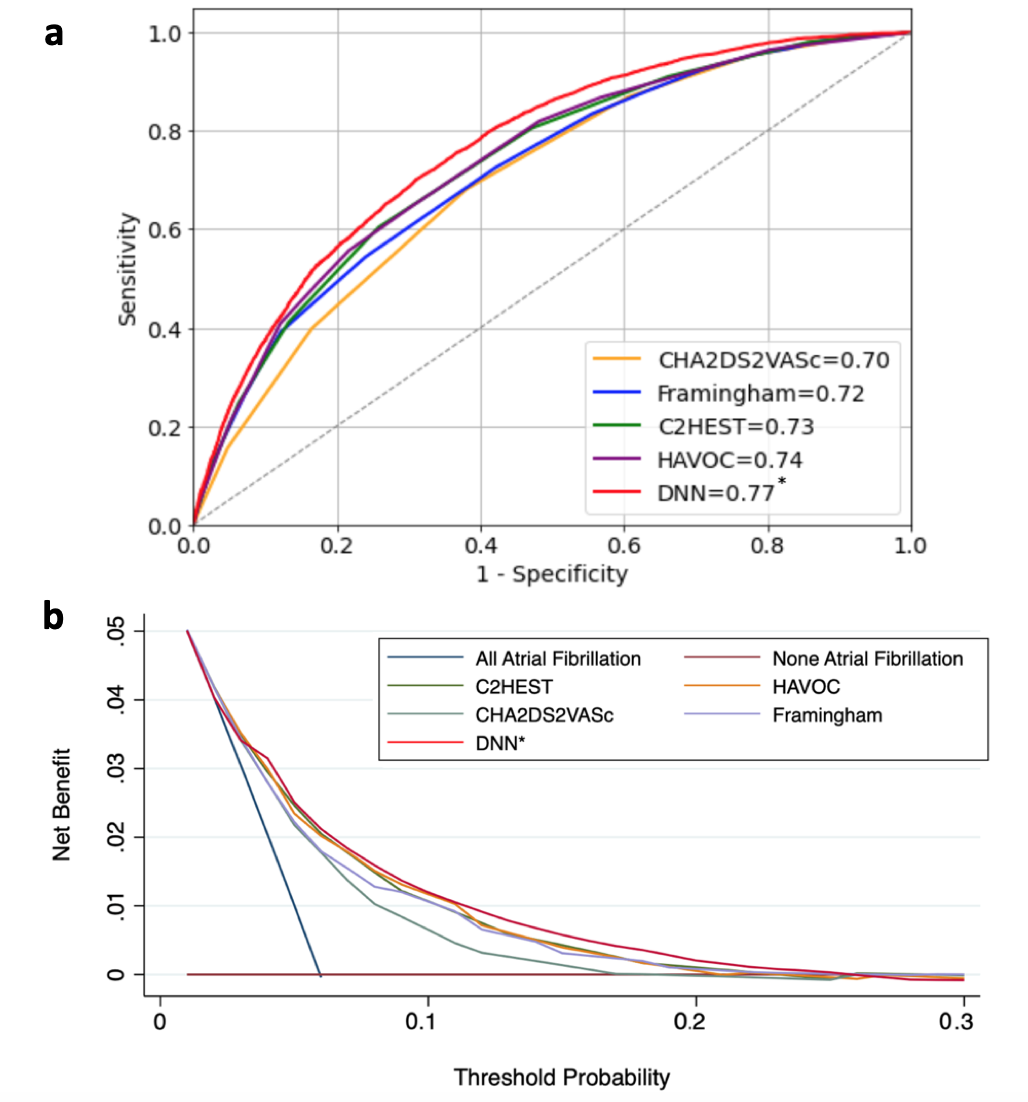
AF: Atrial Fibrillation; DNN: Deep Neural Network; IS: Ischemic Stroke; KNN: K-Nearest Neighbors; LR: Logistic Regression; RFC: Random Forest Classifier; XGBoost: Extreme Gradient Boosting

**Figure 1.** Flow chart of the study



**Figure 2.** Machine learning methods. Model training using five different algorithms and five-fold cross-validation.

**Figure 3.** Distribution of the population and incident rate of atrial fibrillation (AF) according to the deep neural network score. \* DNN: Deep Neural Network



**Figure 4. a,** Receiver operating characteristic curves of incident atrial fibrillation after ischemic stroke.

\* DNN: Deep Neural Network, P<0.0001 vs Framingham risk score, CHA2DS2VASc score, HAVOC score and C2HEST score. **b,** Decision curve analyses for the Deep Neural network algorithm and C2HEST, CHA2DS2-VASc, HAVOC and Framingham risk scores. \* DNN: Deep Neural Network

**SUPPLEMENTARY MATERIAL**

**Supplemental table 1.** Hyperparameters of machine learning models.

| Model | Hyper-parameters | Value |
| --- | --- | --- |
| LR | C | 0.01 |
|  | Class weight | 0:1; 1:16 |
| RFC | N estimators | 245 |
|  | Max depth | 6 |
| KNN | N\_neighbors | 171 |
|  | Leaf\_size | 49 |
| XGBoost | Learning rate | 0.8 |
|  | Min child weight | 10 |
|  | Colsample bytree | 0.04 |
| DNN | Hidden layers | 50-100-50 |
|  | Learning rate  Activation  Alpha | adaptive  relu  0.0001 |

Abbreviations: ML, machine learning; LR, logistic regression with L2 regularization; RFC, random forest classifer; KNN, K nearest neighbors; XGBoost, extreme gradient boosting; DNN, deep neural network. All options were left as default other than Hyper-parameters in table.

**Supplemental table 2.** Predictive Ability of the ML model and clinical Scores for atrial fibrillation occurrence after stroke.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **C-Index (95% CI)** |  | **C-Index Comparison** | | |  |
|  |  | **C2HEST** | **CHA2DS2-VASc** | **HAVOC** | **Framingham** | **DNN** |
| C2HEST | 0.735 (0.727 - 0.742) | - | <0.0001 | 0.1537 | <0.0001 | <0.0001 |
| CHA2DS2VASc | 0.714 (0.713 - 0.714) | <0.0001 | - | <0.0001 | <0.0001 | <0.0001 |
| HAVOC | 0.738 (0.730 - 0.745) | 0.1537 | <0.0001 | - | <0.0001 | <0.0001 |
| Framingham | 0.720 (0.712 - 0.727) | <0.0001 | <0.0001 | <0.0001 | - | <0.0001 |
| DNN | 0.766 (0.759 - 0.773) | <0.0001 | <0.0001 | <0.0001 | <0.0001 | - |
|  |  |  |  |  |  |  |

CI: Confidence interval; DNN: Deep Neural Network.