UNIVERSITÉ DE TOURS

Santé, Sciences Biologiques et Chimie du Vivant (SSBCV)

EA 4245 – Transplantation, Immunologie et Inflammation (T2I)

THÈSE

présentée par

Arnaud BISSON

soutenue le

28 mars 2024

pour obtenir le grade de

Docteur de l'université de Tours

dans la spécialité

Cardiologie et Maladies Vasculaires

Machine Learning et Épidémiologie de la Fibrillation Atriale

THÈSE dirigée par :

Monsieur le Professeur ANGOULVANT Denis	PU-PH, université de Tours
Monsieur le Professeur FAUCHIER Laurent	PU-PH, université de Tours

RAPPORTEURS :

Madame la Docteure OLIÉ Valérie	PhD, HDR, Santé Publique France
Monsieur le Docteur ALGALARRONDO Vincent	MCU-PH, université Paris Diderot

---

JURY :

Madame la Professeure GRAMMATICO-GUILLON Leslie (Présidente du jury)	PU-PH, université de Tours
Monsieur le Professeur ANGOULVANT Denis	PU-PH, université de Tours
Monsieur le Professeur FAUCHIER Laurent	PU-PH, université de Tours
Monsieur le Professeur GARCIA Rodrigue	PU-PH, université de Poitiers
Madame la Docteure OLIÉ Valérie	PhD, HDR, Santé Publique France
Monsieur le Docteur ALGALARRONDO Vincent	MCU-PH, université Paris Diderot

REMERCIEMENTS

Monsieur le Professeur Denis ANGOULVANT,
Vous m'avez fait l'honneur de diriger ce travail et me faites le plaisir de le juger.
Je vous remercie pour votre soutien ainsi que votre disponibilité.
Vous avez toute ma reconnaissance.

Monsieur le Professeur Laurent FAUCHIER,
Merci d'avoir accepté de diriger et de juger ce travail.
La pertinence de tes conseils et la justesse de tes corrections sont toujours pour moi l'occasion d'apprendre.
Sois assuré de ma profonde estime et de ma gratitude.

Madame la Professeure Leslie GRAMMATICO-GUILLON,
Merci de me faire l'honneur de juger ce travail.
Ton expertise en épidémiologie et ton regard sont essentiels dans l'évaluation de ce travail.
Au plaisir de débuter de futurs projets collaboratifs.
Sois assurée de mon profond respect.

Monsieur le Professeur Rodrigue GARCIA,
Merci de me faire l'amitié de juger ce travail.
Échanger avec toi sur tous les sujets est un réel plaisir.
Sois assuré de ma sincère estime.

Madame la Docteure Valérie OLIÉ,
Vous me faites l'honneur de juger ce travail.
Merci pour l'intérêt que vous y accordez.
Soyez assurée de ma respectueuse considération.

Monsieur le Docteur Vincent ALGALARRONDO,
Tu me fais l'honneur de juger ce travail.
Merci pour l'intérêt que tu y accorde.
Sois assuré de ma respectueuse considération.

To Professor Gregory Lip, I really enjoyed my year in Liverpool. Thank you so much for your kindness and for welcoming me into the LCCS family.

A Monsieur le Professeur Dominique Babuty, vous qui m'avez inspiré et soutenu toutes ces années. Votre exemplarité marquera à jamais le médecin que vous avez formé.

A Nicolas, merci pour ces années passées à tes cotés pendant lesquelles j'ai beaucoup appris et beaucoup ri. Même si nos chemins se sont récemment séparés, c'est toujours un plaisir de se retrouver.

A Alexandre, merci pour ta sympathie et ta bonne humeur sur lesquelles je peux compter. Hâte de repartir sur de nouveaux projets ensemble !

A toute ma famille, soutien permanent et indéfectible qui me rappelle ce qui est le plus important dans la vie.

A mes quatre petites pupilles que je ne me lasserai jamais de voir grandir. N'allez pas trop vite s'il vous plait.

A ma chérie Audrey, continuons à tourner ensemble les pages d'une histoire qui est, grâce à toi, plus belle chaque jour…

SOMMAIRE

Travaux scientifiques conduits pendant la thèse ..... p 6

Résumé ..... p 7

Abstract ..... p 8

Liste des abréviations ..... p 9

INTRODUCTION

1 Fibrillation atriale ..... p 10

1.1 Définition .....p 10

1.2 Épidémiologie ..... p 10

1.3 Prédire la fibrillation atriale, son évolution et ces complications ..... p 11

1.4 Limite des modèles actuels..... p 13

2 Machine learning ..... p 16

2.1 Genèse et essor du machine learning ..... p 16

2.2 Terminologie et concepts ..... p 18

2.3 Application du machine learning en épidémiologie ..... p 20

2.3.1 Prétraitement des données ..... p 21

2.3.2 Entrainement ..... p 23

2.3.3 Évaluation ..... p 27

3 Fibrillation atriale et machine learning ..... p 27

4 Problématique ..... p 29

OBJECTIFS

..... p 30

TRAVAUX

1 Première partie : Modèles supervisés et prédiction de la fibrillation atriale ou d'événements cardiovasculaires indésirables associés à la fibrillation atriale ..... p 31

1.1 Étude 1 : Prédiction de la fibrillation atriale incidente après un accident vasculaire cérébral à l'aide d'une approche de type machine learning : Une étude de cohorte nationale française ..... p 32

1.2 Étude 2 : Prédiction de la mortalité précoce après un diagnostic de fibrillation atriale à l'aide d'une approche de type machine learning : Une étude de cohorte nationale française ..... p 42

1.3 Étude 3 : Modèle de machine learning de type multi-label gradient-boosting decision tree en chaine pour prédire les risques multiples chez les patients atteints de fibrillation atriale : Résultats issus de la phase II/III du registre GLORIA-AF ..... p 56

1.4 Discussion ..... p 80

2 Deuxième partie : Modèles non supervisés et analyse en cluster au sein d'une population de patients atteints de fibrillation atriale ..... p 83

2.1 Étude 4 : Phénotypes cliniques et récidives de fibrillation atriale après ablation par cathéter : Une analyse non supervisée en clusters ..... p 83

2.2 Étude 5 : Phénotypes et pronostic chez les patients non anti coagulés atteints de fibrillation atriale : Une analyse non supervisée en clusters ..... p 101

2.3 Discussion ..... p 112

CONCLUSION

..... p 113

PERSPECTIVES

..... p 114

ANNEXES

..... p 115

REFERENCES BIBLIOGRAPHIQUES

..... p 123

Travaux scientifiques conduits pendant la thèse

Publications :

***Bisson A**, Lemrini Y, El-Bouri W, Bodin A, Angoulvant D, Lip GYH, Fauchier L. **Prediction of incident atrial fibrillation in post-stroke patients using machine learning: a French nationwide study**. Clin Res Cardiol. 2023 Jun;112(6):815-823. doi: 10.1007/s00392-022-02140-w.*

***Bisson A**, Fawzy AM, El-Bouri W, Angoulvant D, Lip GYH, Fauchier L, Clementy N. **Clinical Phenotypes and Atrial Fibrillation Recurrences After Catheter Ablation: An Unsupervised Cluster Analysis.** Curr Probl Cardiol. 2023 Aug;48(8):101732. doi: 10.1016/j.cpcardiol.2023.101732.*

***Bisson A**, M Fawzy A, Romiti GF, Proietti M, Angoulvant D, El-Bouri W, Y H Lip G, Fauchier L. **Phenotypes and outcomes in non-anticoagulated patients with atrial fibrillation: An unsupervised cluster analysis.** Arch Cardiovasc Dis. 2023 Jun-Jul;116(6-7):342-351. doi: 10.1016/j.acvd.2023.06.001.*

***Bisson A**, Lemrini Y, Romiti GF, Proietti M, Angoulvant D, Bentounes S, El-Bouri W, Lip GYH, Fauchier L. **Prediction of early death after atrial fibrillation diagnosis using a machine learning approach: A French nationwide cohort study.** Am Heart J. 2023 Nov;265:191-202. doi: 10.1016/j.ahj.2023.08.006.*

Présentations orales :

***Bisson A.** Oral presentation. **Artificial intelligence: useful to predict and diagnose atrial fibrillation?** 22èmes Journées de Rythmologie. 6-8 octobre 2021, Paris, France.*

***Bisson A.** Short communication. **Prediction of incident atrial fibrillation in post stroke patients using machine learning: a French nationwide study.** 33èmes Journées Européennes de la Société Française de Cardiologie. 11-13 janvier 2023, Paris, France.*

***Bisson A.** Moderated Poster Presentation. **Prediction of early death after atrial fibrillation diagnosis: a French nationwide study.** EHRA Congress. 16 -18 April 2023, Barcelona, Spain.*

***Bisson A.** Oral abstract session. **Mortality prediction in patient with atrial fibrillation and 0 or 1 nongender-related CHA2DS2-Vasc risk factor using machine learning: a French nationwide study.** ESC Congress. 25-28 August 2023, Amsterdam, Netherlands.*

***Bisson A.** Moderated Poster Presentation. **Phenotypes and outcomes in non-anticoagulated patients with atrial fibrillation: an unsupervised cluster analysis.** 34èmes Journées Européennes de la Société Française de Cardiologie. 17-19 janvier 2024, Paris, France.*

***Bisson A.** Oral presentation. **The role of artificial intelligence in electrophysiology and cardiac pacing.** 34èmes Journées Européennes de la Société Française de Cardiologie. 17-19 janvier 2024, Paris, France.*

MARCHINE LEARNING
ET
EPIDEMIOLOGIE DE LA FIBRILLATION ATRIALE

Résumé

La fibrillation atriale est l'arythmie la plus fréquente et est associée à une morbi-mortalité importante. Les patients atteints de fibrillation atriale constituent une population hétérogène en termes de caractéristiques mais aussi en termes de pronostic. Les modèles statistiques conventionnels permettent d'explorer et de prédire la survenue d'évènements cardio-vasculaires chez les patients atteints de FA. Cependant, ces modèles sont limités et leurs performances ne sont pas optimales, particulièrement avec les données recueillies actuellement.

Les nouvelles approches de type machine learning pourraient rendre possible une meilleure compréhension de nos données et une amélioration de nos modèles prédictifs.

Après une revue de la littérature, nous présentons dans ce travail le développement, l'application et l'évaluation de modèles de machine learning supervisés dans la prédiction d'événements indésirables liés à la fibrillation atriale à travers 3 études réalisées partir de différents jeux de données de patients atteints de fibrillation atriale au niveau national et international. Nous rapportons ensuite l'application de modèles non supervisés de clustering dans la recherche de phénogroupes cliniquement pertinents à travers 2 études réalisées sur des bases de données locales.

Par ce travail, nous montrons l'intérêt des méthodes de machine learning, par la supériorité des modèles prédictifs développés par rapports aux modèles classiques, par la description de phénogroupes de patients cliniquement pertinents et par l'éclairage nouveau apporté par nos données permettant l'émergence de nouvelles questions et hypothèses.

Mots-clés : Fibrillation atriale ; Machine learning ; Épidémiologie.

MARCHINE LEARNING AND EPIDEMIOLOGIE OF ATRIAL FIBRILLATION

Abstract

Atrial fibrillation is the most common arrhythmia and is associated with significant morbidity and mortality. Patients suffering from atrial fibrillation constitute a heterogeneous population in terms of both characteristics and prognosis. Conventional statistical models can be used to explore and predict the occurrence of cardiovascular events in patients with AF. However, these models are limited and their performance is not optimal, particularly with the data currently being collected.

New machine learning approaches could provide a better understanding of our data and improve our predictive models.

After a review of the literature, we present in this work the development, application and evaluation of supervised machine learning models in the prediction of adverse events related to atrial fibrillation through 3 studies carried out using different datasets of patients with atrial fibrillation at national and international level. We then report on the application of unsupervised clustering models in the search for clinically relevant phenogroups from 2 studies based on local databases.

Through this work, we demonstrate the value of machine learning approach, through the superiority of the predictive models developed when compared with conventional models, through the description of clinically relevant phenogroups of patients and through the new light shed on our data, allowing new questions and hypotheses to emerge.

Key words: Atrial Fibrillation; Machine Learning; Epidemiology.

Liste des abréviations

ABC : Atrial fibrillation Better Care (ou Parcours ABC)

AUC : Aire sous la courbe ROC

AVCi : Accident Vasculaire Cérébral ischémique

ESC : European Society of Cardiology (Société Européenne de Cardiologie)

FA : Fibrillation Atriale

IA : Intelligence Artificielle

MeSH : Medical Subject Heading

ML : Machine Learning

NHS : National Health Service

PCA : Principal Componant Analysis

PMSI : Programme de Médicalisation des Systèmes d'Information

SMOTE : Synthetic minority oversampling technique

INTRODUCTION

1 Fibrillation atriale

1.1 Définitions

La fibrillation atriale (FA) est une arythmie d'origine supraventriculaire due à une activité électrique rapide (400–600/min) et anarchique des deux oreillettes ayant pour conséquences des contractions désynchronisées et hemodynamiquement inefficaces.

Elle se manifeste sur l'électrocardiogramme par une activation atriale irrégulière (trémulation de la ligne de base) sans onde P distincte et des intervalles RR irréguliers (quand la conduction atrioventriculaire est préservée). Sa réponse ventriculaire est sous la dépendance du nœud atrioventriculaire dont la capacité de filtration varie selon l'état du système nerveux autonome et de l'imprégnation en médicaments bradycardisants. Son diagnostic repose sur un enregistrement électrocardiographique 12 dérivations ou mono piste montrant ces éléments durant au moins 30 secondes [1].

Il existe plusieurs manières de classer la FA selon la présence de symptômes, de cause déclenchante sous-jacente (ex : FA post opératoire) ou encore de certaines conditions valvulaires (FA associée à un rétrécissement mitral moyen à sévère ou une prothèse valvulaire mécanique). Mais la classification la plus utilisée est celle dite en « P » qui distingue le mode de présentation en **P**remier épisode de FA, FA **P**aroxystique, **P**ersistante, **P**ersistante longue durée ou encore **P**ermanente [1].

1.2 Épidémiologie

La FA est l'arythmie la plus fréquente chez l'adulte. Sa prévalence au niveau mondiale est estimée entre 2 et 4% ce qui représentait 43.6 millions d'individus en 2016 [1,2]. Elle est en augmentation en raison de l'allongement de la durée de vie dans la population générale et de l'intensification de son dépistage, et devrait doubler au cours des prochaines décennies [3–6]. Chez les Européens de plus de 55 ans, le risque de développer une FA au cours de la vie est

d'environ un sur trois [7,8]. En France, elle concernerait entre 600 000 et 1 000 000 de patients [9].

La FA s'associe à tout un cortège de comorbidités qui constituent le terrain de celle-ci. Les cardiopathies associées les plus fréquentes sont la cardiopathie hypertensive (22% - 36%), les coronaropathies (14%-32%) et les cardiopathies valvulaires (12% - 26%) [10]. Tandis que les comorbidités les plus fréquentes sont l'hypertension (67% - 76%), l'insuffisance cardiaque (22% - 42%), le diabète (20% - 24%), l'obésité (20% - 35%), les maladies pulmonaires chroniques (10% - 18%), les dysthyroïdies (8% - 11%), l'insuffisance rénale (11% - 22%), les AVCi (9% - 16%), et les troubles neuropsychiatriques (19%) [10]. En particulier, un tiers des patients atteints de FA présentent au moins trois comorbidités associées [10].

De plus, la FA est grevée d'une morbi-mortalité importante [11,12]. En effet, la FA multiplie le risque d'un AVCi par 5 et est responsable de 20% d'entre eux [13–15]. Ces AVCi secondaires à la FA sont plus graves, sont associés à une plus grande mortalité et à un plus grand risque de récidive [2,16,17]. Un tiers des patients atteints de FA ont une insuffisance cardiaque qui, quelle en soit l'origine ou la conséquence, est la principale cause d'hospitalisation (14%) et de mortalité (14.5%) [18–20]. La FA induit aussi un déclin cognitif, de la dépression et une altération de la qualité de vie [21–23]. *In fine*, la FA est associée à une mortalité multipliée par 3.5 [7].

Enfin, la FA est aussi un enjeu économique car elle engendre des coûts très importants pour nos sociétés. Ces coûts sont majoritairement attribuables aux hospitalisations et aux complications comme l'AVCi [24,25]. En France, on estime que près de 2,6 % des dépenses annuelles en soins de santé sont associées aux hospitalisations liées à la FA pour un coût avoisinant les 2 milliards € [26]. Concernant le Royaume Uni, le coût total est estimé à 2 milliards £ soit 0.9 à 1.6% des dépenses du National Health Service (NHS) mais pourrait, à horizon 2040, s'élever entre 1.35 et 4.27% [24].

1.3 Prédire la fibrillation atriale, son évolution et ses complications

Pourquoi ?

Le constat est que, dans l'immense majorité des cas, on ne guérit pas de la FA. Les progrès réalisés ces dernières décennies nous permettent en partie de prévenir sa survenue, de mieux

la dépister, de ralentir son évolution, de réduire son retentissement en termes de symptômes et de qualité de vie ou encore de diminuer les complications et la surmortalité qui en résultent mais pas de guérir de cette maladie chronique [27–32]. Il est par conséquent primordial de pourvoir prédire en amont et détecter précocement les épisodes clés (bien souvent irréversibles) de l'histoire naturelle de cette maladie afin de pouvoir réduire leurs impacts en intervenant à temps par la prise en charge des facteurs de risque modifiables, l'introduction de traitements préventifs des complications et la mise en place de stratégies invasives efficaces [33].

Cette prise en charge globale et intégrative est citée dans les dernières recommandations de la Société Européenne de Cardiologie (ESC) sous le nom de 'Parcours ABC' (« The Atrial fibrillation Better Care (ABC) pathway for integrated management »; A: Anticoagulation/Avoid stroke; B: Better symptom management; C: Cardiovascular and Comorbidity optimization) [1,34]. Ce « parcours » holistique a démontré sa supériorité en termes de réduction de la mortalité tout cause (-58%), de réduction de la mortalité cardiovasculaire (-63%), de réduction des AVCi et saignements majeurs (-45% et -31%, respectivement) et de diminution des dépenses de santé [35–40]. Pour être pleinement efficace, il est néanmoins crucial que cette stratégie soit correctement appliquée au bon patient, au bon moment.

Cependant, diagnostiquer la FA n'est pas toujours chose aisée car elle est bien souvent asymptomatique et paroxystique. En effet, jusqu'à 90% des épisodes de FA sont asymptomatiques avec une faible corrélation entre les symptômes et les épisodes (17%), et un tiers des FA sont de forme paroxystique [41,42]. Quand bien même la FA est symptomatique, les signes fonctionnels peuvent être aspécifiques (palpitations, dyspnée, asthénie, douleurs thoraciques) [43]. De plus, le risque individuel au regard de la FA ou de ses complications est très hétérogène et variable d'un individu à l'autre [42,44–46].

C'est donc pour cela que le parcours ABC repose sur la classification et la sélection précise des patients afin de stratifier au mieux leur risque individuel et *in fine* de personnaliser au maximum leur prise en charge [1].

Comment ?

C'est classiquement à partir de registres ou cohortes que nous proviennent les données qui nous ont permis de construire les classifications et les modèles prédictifs essentiels à la prise en charge de la FA. On peut citer, parmi d'autres, les cohortes Euro Heart Suvey on AF

débutée en 2003 (n=5 333), ARIC débutée en 1987 (n=15 792) ou encore Framingham débutée en 1948 (n=5 209 pour la cohorte originelle) [45–50].

Ces cohortes ont permis le développement de modèles, utilisés en pratique clinique courante pour certains, prédisant le risque de survenue de FA (scores CHARGE-AF et C_2HEST), la progression de la FA de la forme paroxystique à persistante (score HATCH), le risque de récidive après ablation par cathéter (scores CAAP-AF, APPLE, ATLAS), le risque d'AVCi (scores $CHADS_2$, CHA_2DS_2-VASc, GARFIELD-AF, ATRIA, ABC-stroke) ou encore le risque de saignement (scores HASBLED, ORBIT et ATRIA bleeding) [45–47,51–61].

Ces scores se basent essentiellement sur des facteurs de risque cliniques (âge, hypertension artérielle, insuffisance cardiaque, diabète, …) dont l'impact sur l'histoire naturelle de la FA est majeur. Cela souligne l'importance du terrain qui fait le lit de la FA et de ses complications par ses conséquences sur l'oreillette gauche et le développement d'une véritable cardiomyopathie atriale [62–64]. Néanmoins, certains de ces scores sont multiparamétriques et intègrent des données électrocardiographiques, biologiques ou de l'imagerie reflétant directement cette atteinte atriale [53–55,57,58]. Par ailleurs, des pistes visant à intégrer d'autres variables provenant de nouvelles technologies comme la génomique, l'épigénomique, la transcriptomique, la protéomique et la métabolomique regroupées sous le terme de « omiques », ou encore de nouvelles voies comme celle de l'exposome incluant les influences environnementales, les conditions de vie et les aspects biopsychologiques, se développent et représentent une étape importante dans cette quête d'une médecine dite de précision [12].

« Ce qui est simple est toujours faux. Ce qui ne l'est pas est inutilisable » (Paul Valéry) et toute la difficulté reste pour ces scores de combiner simplicité et praticité d'utilisation au quotidien d'une part, et généralisation et précision d'autre part.

1.4 Les limites des modèles actuels

Les méthodologies statistiques dites « fréquentistes » (par opposition à l'approche « bayesienne ») sont les approches analytiques les plus couramment utilisées [65]. Les tests statistiques de base et les modèles de régression multivariés (logistique, linéaire et à taux proportionnel de Cox) sont les plus couramment utilisés pour tester les hypothèses afin de définir les associations entre les variables prédictives et les variables à prédire [66]. La plupart

des modèles traditionnels sont axées sur les données et font diverses hypothèses (linéarité, l'absence de multicolinéarité et la proportionnalité des risques dans le temps) sur les données utilisées dans le modèle. Ces hypothèses sont bien comprises par les épidémiologistes. Cependant, au fur et à mesure que le nombre de données disponibles pour l'analyse augmente, la « malédiction de la dimensionnalité » devient manifeste [67]. Dans ce contexte, les hypothèses de modélisation traditionnelles deviennent plus complexes, les relations sont fortement non linéaires et de nouvelles méthodes deviennent nécessaires.

Cela explique que malgré tous les efforts menés sur ce sujet ces dernières années et les centaines de publications scientifiques qui en résultent, les performances des modèles actuels restent relativement modestes.

Cela se traduit par une utilisation parfois réduite de ces scores. Par exemple, le score CHA_2DS_2-VASc ne prédit que modestement les patients à haut risque thromboembolique et son utilisation se limite dans les recommandations, malgré une échelle de valeurs de 0 à 9, qu'à la stratification des patients à bas risque (<1% d'AVCi/an) ne nécessitant pas d'anticoagulation (CHA_2DS_2-VASc = 0 pour les hommes ou 1 pour les femmes) (Classe I, Niveau de preuve A) [1,45]. Pour ce qui est des patients à risque intermédiaire (CHA_2DS_2-VASc = 1 pour les hommes ou 2 pour les femmes), le risque (estimé à 1-2%/an) n'est pas homogène, la décision d'introduire ou non une anticoagulation ne peut se baser que sur le seul score et il est demandé d'individualiser le traitement (Classe IIa, Niveau de preuve B), en particulier dans les populations asiatiques (**Annexe 1**) [1,68,69]. En effet, l'augmentation du risque d'AVCi chez les patients à risque intermédiaire par rapport aux patients à faible risque n'est pas le même selon le facteur de risque présent (âge de 65 à 74 ans : +224% ; diabète : +151% ; insuffisance cardiaque congestive : +98% ; hypertension artérielle : +74% ; maladie vasculaire : +56%) [70]. Pour ce qui est du score HASBLED, son impact clinique est limité car il catégorise mal les patients à bas risque hémorragique (score < 3) [46]. Et pour les patients à haut risque (score ≥ 3), son utilisation doit se limiter à les « marquer » (flag up) pour une surveillance rapprochée mais sans conduire à la modification ou interruption d'une anticoagulation [1,71]. Enfin, pour les autres scores cliniques cités précédemment, ils n'ont pas de place en pratique courante.

Plusieurs explications peuvent être avancées sur ce « plafond de verre statistique » semblant infranchissable. Tout d'abord, l'association de comorbidités, de poly médicamentation et de fragilité, ainsi que leurs intrications et interactions constituent ce qu'on appelle la « complexité clinique », bien qu'il n'y en ait pas de définition unique [72]. Cette complexité clinique est associée à un plus mauvais pronostic et une moindre efficacité du

parcours ABC chez les patients atteints de FA [40,73]. Modéliser la complexité clinique des patients atteints de FA ne se réduit pas à faire une simple addition, avec plus ou moins de pondération, des caractéristiques qui la composent comme le font la plupart des modèles actuels ne reflétant pas suffisamment la profondeur des interactions entre les différentes combinaisons de comorbidités et leur association avec le pronostic. De plus, l'impact de certaines variables sur le pronostic peut dépendre de la présence d'autres variables, en faisant un facteur modulant du risque (risk modifier) plus qu'un facteur de risque en soi comme c'est le cas pour le sexe féminin avec l'âge dans le score CHA_2DS_2-VASc [74–76].

Ensuite, à l'heure où les recommandations sont internationales et les populations multiethniques, il est difficile d'utiliser une approche unique de type « one size fits all » surtout si le modèle a été élaboré initialement à partir d'une population homogène et bien définie. En effet, il existe des disparités ethniques et géographiques concernant le risque hémorragique, l'efficacité et la sécurité des anticoagulants oraux directs, et la stratification du risque d'AVCi selon la population étudiée [77–79]. En conséquence et s'agissant des populations asiatiques, le risque thrombotique chez les patients à bas risque (évalué par un score CHA_2DS_2-VASc = 0 pour les hommes et 1 pour les femmes) et ayant entre 50 et 64 ans est plus élevé (1.78%/an) et dépasse le seuil tolérable de 1%/an avant l'introduction d'une anticoagulation [80]. Cette particularité a amené *Chao, Lip et al.* à modifier le critère d'âge du score CHA_2DS_2-VASc (1 point à partir de 50 ans au lieu de 65 ans) permettant une amélioration des performances prédictives et de l'index net de reclassification pour ces patients [79].

Enfin, l'évolution des populations, des méthodes diagnostiques et des thérapeutiques rendent rapidement obsolètes les scores souvent développés sur des cohortes constituées des années auparavant. Si nous reprenons l'exemple des scores CHA_2DS_2-VASc et HASBLED qui sont encore les références actuellement, ces scores ont été développés à partir d'une cohorte constituée il y a 20 ans et ont fait leur première apparition dans les recommandations en 2010 [1,81]. Depuis, les patients atteints de FA ont vu leur profil de comorbidités et leur pronostic changer au fil des années et on peut se poser la question de la précision de ces modèles au cours du temps [4,82]. De plus, les anticoagulants oraux directs, arrivés sur les marchés américains et européens peu de temps après (2010 et 2011, respectivement) suite à la publication de l'essai RE-LY ont bouleversé nos habitudes de prescription [83]. En effet, les prescriptions d'antivitamine K, seul anticoagulant oral disponible dans les années 2000, sont devenues minoritaires à partir de 2014 et ne représentaient que 12.9% des nouvelles prescriptions d'anticoagulants oraux en France en 2018 [84–86]. Or, la plupart des scores de

risque hémorragique ont été validés avec la warfarine [87]. Dans une vaste étude menée auprès de patients sous anticoagulants oraux directs, les scores HAS-BLED, ATRIA et ORBIT-AF avaient des performances prédictives modérées (air sous la courbe ROC (AUC) : 0,58-0,61). La sensibilité et la spécificité du score HAS-BLED dans cette population étaient respectivement de 62,8% et 53,5% quant à la valeur prédictive positive, elle n'était que de 3,0% [88].

2 Machine learning

2.1 Genèse et essor du machine learning

L'intelligence artificielle ou IA dont fait partie le machine learning (ML) est une discipline jeune dont l'essor au cours des 70 dernières années a été rythmée par « 3 vagues » successives entrecoupées de périodes appelées « hiver technologiques » (**Figure 1**).

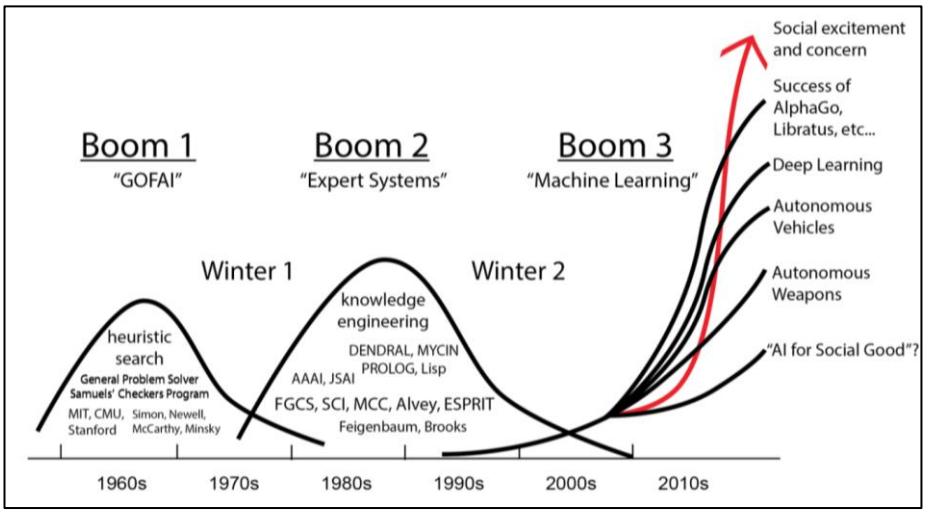


Figure 1. Évolutions de l'intelligence artificielle depuis les années 1960.

D'après Garvey CS. Technology's Stories. 2018. DOI : 10.15763/jou.ts.2018.03.16.02.

Tout commence dans les années 1940 avec la conjonction des développements technologiques de l'époque et l'avènement de la cybernétique, théorie dont l'objectif était

d'unifier les mathématiques, l'électronique et l'automatisation en tant que « théorie entière de la commande et de la communication, aussi bien chez l'animal que dans la machine » [89].

En 1943 est publié par Warren McCullough (neurophysiologiste) et Walter Pittsle (mathématicien) le premier modèle mathématique d'un réseau de neurone. Le SNARC (Stochastic Neural Analog Reinforcement Calculator), premier ordinateur à réseau de neurones (40 neurones artificiels) voit quant à lui le jours suite au travail de Marvin Minsky et Dean Edmonds au début des années 1950 [90]. C'est Alan Turing qui posera pour la première fois la question de l'éventuelle intelligence d'une machine dans son célèbre article de 1950 « Computing Machinery and Intelligence » où il décrira son fameux test [91]. Ce n'est qu'en 1956, que le terme d'Intelligence Artificielle sera prononcé pour la première fois, durant la conférence du « Dartmouth Summer Research Project on Artificial Intelligence » par John McCarthy. Lors de cet événement, les chercheurs ont présenté les ambitions de l'IA : répliquer ou simuler l'intelligence humaine par des machines.

Arthur Samuel, informaticien américain pionnier dans le secteur de l'intelligence artificielle, est le premier à faire usage de l'expression « machine learning » ou apprentissage automatique en 1959 à la suite de la création d'un programme de Jeu de Dames pour la société IBM en 1952 dont l'algorithme était capable de s'améliorer en jouant [92].

Si l'IA demeurait fascinante et remplie de promesses, l'engouement est retombé dans les années 1960. Les ordinateurs se sont assez vite heurtés aux limites technologiques de l'époque en termes de capacité de mémoire et de puissance de calcul. A ce premier boom appelé GOFAI (« Good Old-Fashioned Artificial Intelligence ») va succéder un premier « hiver technologique ».

C'est grâce à l'avènement des premiers microprocesseurs à la fin des années 1970 que l'IA connait un nouvel essor et entre dans l'âge d'or des « Systèmes Experts ». En effet, le développement de l'informatique va être important pendant plusieurs décennies suivant la loi de Moore. Ce modèle techno-économique prédit le doublement du nombre de transistors présent sur un microprocesseur tous les 18 à 24 mois expliquant la croissance exponentielle des performances des ordinateurs [93,94] (**Annexe 2**). Ces Systèmes Experts s'appuyaient sur un « moteur d'inférence », correspondant à un algorithme de simulation des raisonnements déductifs, programmé pour être un miroir logique d'un raisonnement humain. Le moteur était ainsi capable de conduire des raisonnements logiques d'un haut niveau d'expertise et de dériver des conclusions à partir d'une base de connaissances. Une fois encore, l'engouement retombera à nouveau fin des années 1980 du fait de la complexité de leur programmation et d'un effet « boite noire ». Le deuxième hiver technologique arrivera donc malgré le très

médiatique succès de Deep Blue (système expert d'IBM) au jeu d'échec contre le champion du monde Garry Kasparov en mai 1997.

Il faudra attendre les années 2010 pour que débute le troisième boom dans lequel nous nous trouvons encore. Cette nouvelle vague est liée à deux éléments : 1) L'explosion de la quantité de données numériques produites qui est devenue massive, appelée « Big data », permettant la constitution d'énormes bases de données ; 2) La poursuite de l'amélioration considérable de la puissance de calcul fournie par les processeurs. De plus, un changement complet de paradigme par rapport aux systèmes experts est initié sous l'impulsion de grandes sociétés comme IBM, Microsoft, et Google : l'approche inductive. Il ne s'agit plus de coder les règles comme pour les systèmes experts. L'algorithme n'est plus seulement sur une action itérative mais il va pouvoir être doté de la capacité d'apprentissage permettant l'élaboration de ses propres règles, et c'est le big data qui rendra possible l'entrainement de ces algorithmes. Cela permettra l'essor des techniques de machine learning, parmi lesquelles se trouve l'apprentissage profond (deep learning), techniques largement utilisées à l'heure actuelle en recherche mais aussi pour toutes sortes de taches (reconnaissance de texte, d'image, vocale, modélisation, analyse de données) dans de multiples domaines qui font notre quotidien (information, bourse, santé, sciences, sécurité).

2.2 Terminologie et concepts

Il semble important de rappeler quelques définitions nécessaires à la compréhension de la suite de cet exposé puisque certains concepts de machine learning n'ont pas de parallèle statistique ou épidémiologique, et que la terminologie entre ces deux disciplines diffère souvent. Nous passons ici brièvement en revue les principes de base du machine learning. Quelques exemples d'équivalence entre termes relatifs au machine learning et statistiques/épidémiologiques sont proposés en annexe (**Annexe 3**) [66,95,96].

Machine learning

Le machine learning est une branche de l'IA dont l'objectif général est de permettre aux ordinateurs « d'apprendre » à partir de la donnée sans avoir été programmés explicitement à cet effet [92]. C'est la différence avec la programmation conventionnelle, qui utilise des algorithmes « basés sur des règles » préétablies (rule-based methods). Cet apprentissage

automatique permet l'amélioration de leurs performances dans l'accomplissement de tâches grâce à leur expérience.

Le machine learning met généralement l'accent sur la précision prédictive plutôt que sur l'inférence fondée sur des hypothèses (*a priori*), et tire parti généralement d'ensembles de données de grande taille et de haute dimension (les big data) [97,98]. Bien que ces techniques existent depuis de nombreuses années, le machine learning n'a été accepté en tant que terme Medical Subject Heading (MeSH) de Medline qu'en 2016 (https://www.ncbi.nlm.nih.gov/mesh/2010029).

Big data

Le terme « big data » désigne les ensembles de données qui sont trop volumineux (en termes de quantité) ou trop complexes (en termes de structure ou dimensionnalité) pour être traités par les méthodes traditionnelles de traitement des données. Par « big », on entend généralement la taille des observations, par opposition aux « fat data », qui comportent moins d'observations par rapport au nombre de variables dont la plupart n'apporte aucune valeur informative et entrave même l'élaboration du processus de prise de décision (« infobesity leading to analysis paralysis ») [95]. Du point de vue de l'épidémiologie, les big data se caractérisent également par le fait que leur récupération et utilisation ultérieure à des fins de recherche n'est pas l'objectif initial de ces données : on peut citer comme exemple : les dossiers médicaux électroniques contenu dans le Programme de Médicalisation des Systèmes d'Information français (PMSI) qui est une base de données médico-administrative servant pour le remboursement des établissements de santé mais qui permet aussi l'élaboration de travaux scientifiques [62,99,100].

Le big data est l'essence du machine learning car ces algorithmes ont besoin d'une grande quantité de données pour apprendre et se développer. De plus, c'est la technologie qui permet d'exploiter le mieux le potentiel du big data.

Modèle supervisé ou non supervisé

Les modèles supervisés, aussi appelés estimateurs, sont généralement utilisés pour prédire une variable dépendante ou outcome comme le ferait un modèle statistique prédictif classique à l'aide d'une régression. L'algorithme de s'entrainer en connaissant a priori la bonne réponse

puisque la variable dépendante, ici appelée « label », va être pré-étiqueté (« labeled data ») afin lui permettre d'apprendre de ses erreurs et de s'améliorer.

Dans l'apprentissage non supervisé, il n'y a pas de labelling (outcome blinded analysis). L'algorithme fonctionne sans cet *a priori*. Une première méthode appelée Cluster analysis tente d'identifier les relations naturelles, tendances ou motifs (patterns) latents au sein des données pour définir des sous-groupes non spécifiés appelés clusters présentant des caractéristiques similaires. Cette approche sert d'analyse exploratoire d'un jeu de données pour ensuite permettre la production d'hypothèses. Une seconde méthode permet la réduction de dimensionnalité d'un jeu de données permettant une manipulation plus facile et aide à identifier les « composantes » influençant la variance des données. On peut citer par exemple la Principal Component Analysis (PCA) (**Annexe 4**).

Algorithme de classification versus régression

Concernant les algorithmes supervisés, la classification fait référence à la prédiction d'un résultats catégoriels, tandis que la régression fait référence à la prédiction de résultats continu.

Deep learning

L'expression deep learning ou « apprentissage profond » est un type spécifique de machine learning apparu pour décrire des réseaux de neurones artificiels avec de nombreuses couches cachées inspirés par l'architecture des réseaux de neurones biologiques. Ces modèles, bien que très complexes, sont extrêmement flexibles et permettent d'inclure un nombre presque illimité de variables pour des tâches de classification ou de régression, et sont très précis pour modéliser des relations hautement non linéaires. Une des limites de ces approches réside dans la sélection des hyper paramètres et l'entraînement qui peuvent être trop gourmands en calcul pour être exécutés sur des machines locales, ce qui réduit leur utilisation à grande échelle.

2.3 Application du machine learning en épidémiologie

L'utilisation d'un algorithme de machine learning (ML) nécessite une méthodologie particulière qui consiste en une succession d'étapes importantes avec toutes sortes d'outils statistiques et informatiques permettant *in fine* la création d'un modèle. La première partie de

ce chapitre est commune aux modèles supervisés et non supervisés, les suivantes sont spécifiques aux modèles supervisés.

2.3.1 Pré traitement des données (preprocessing)

Cette étape cruciale consiste à transformer la donnée brute (raw data) en donnée exploitable par ces algorithmes et par ce biais, d'améliorer leurs performances selon l'adage « better data often beats better algorithms ».

Mise à l'échelle des données continues (scaling)

L'objectif est d'homogénéiser les variables continues en termes d'ordre de grandeur tout en préservant les écarts de distance au sein d'une même variable. L'hypothèse sous-jacente d'un l'algorithme est que les nombres les plus élevés ont une certaine importance et vont donc à jouer un rôle plus décisif lors de la formation du modèle. Les techniques les plus courantes de mise à l'échelle sont la normalisation et la standardisation. La normalisation transforme les valeurs pour qu'elles soient toutes comprises entre un minimum et un maximum (généralement entre [0,1] ou [-1,1]) alors que la standardisation transforme les données pour qu'elles aient une moyenne nulle et une variance de 1. Les données sont alors sans unité.

Gestion des données manquantes (missing data handling)

La majorité des bases de données comportent des données manquantes par défaut de recueil ou corruption de données. La plupart des algorithmes de ML ne peuvent traiter ce genre de données et c'est pourquoi il est important de modifier ces valeurs. Pour cela, il est possible de les retirer (dropping) ou de les remplacer (imputation) par des valeurs arbitraires (0 ou moyenne/médiane des valeurs de la variable) ou par une valeur prédite en fonction des autres variables et observations (Multiple Imputation by Chain Equation (MICE), K Nearest Neighbors (KNN) imputation) (**Annexe 5**) [101].

Encodage des données catégorielles (encoding)

L'encodage consiste à convertir des données catégorielles ou textuelles en format numérique, de manière à ce qu'elles puissent être utilisées comme données d'entrée (input) par les algorithmes. Selon le type de donnée ordinale ou nominale, différentes techniques sont utilisées. En cas de variable ordinale, le label encoding est utilisé et consiste à transformer chaque catégorie en une valeur entière unique. Cette méthode est simple mais présente l'inconvénient d'induire une hiérarchie entre les catégories ce qui n'est pas le cas si la variable est nominale. Dans ce cas, on utilisera la technique dite du One-Hot-Encoding où une colonne binaire (0/1) est créée pour chaque catégorie unique dans la variable.

Suppression des données aberrantes (outliers)

Les données aberrantes ont plusieurs sources : erreur de recueil ou de mesure, valeurs extrêmes. Elles sont responsables d'un bruit statistique et nuisent à la performance de l'algorithme. Il est important de les repérer par des tests statistiques uni ou multivariés et de les traiter (suppression ou imputation).

Extraction / sélection / création des variables prédictives (feature engineering)

Un trop grand nombre de variables (grande dimensionnalité) rend leur manipulation complexe et induit un risque de sur ajustement (overfitting). Il peut donc être nécessaire d'appliquer des techniques de régularisation ou de réduction de la dimensionnalité (feature extraction). Cela permet une diminution du risque de sur-ajustement, une diminution du temps de calcul et une amélioration de la précision. Plusieurs techniques comme la Principal Componant Analysis (PCA) sont alors disponibles.

Le feature engineering correspond à la création de nouvelles variables à partir des données disponibles pour révéler des effets latents. Une simple feature engineering, telle que la création du logarithme d'une variable continue pour modifier sa distribution, la création d'une variable polynomiale, ou l'agrégation de deux variables pour tenir compte de la multicolinéarité, est parfois nécessaire pour améliorer les performances d'un modèle.

Enfin, la sélection des variables d'intérêt (feature selection) est une étape essentielle car les modèles de machine learning sont appliqués à des ensembles de données qui ont souvent été collectés pour des raisons autres qu'une hypothèse spécifique, comme c'est le cas des données des dossiers médicaux électroniques. Ces bases de données contiennent généralement un grand nombre de variables qui, pour beaucoup d'entre elles ne sont pas

pertinentes pour l'élaboration du modèle : « fat data » et loi de Pareto. En réduisant le nombre de variables redondantes ou qui n'affectent pas le résultat, l'objectif est multiple : réduire la dimensionnalité pour réduire l'overfitting ; diminuer le temps de calcul nécessaire à l'entrainement ; créer un modèle plus simple et donc plus facilement utilisable et généralisable (principe du « rasoir d'Ockham ») ; et améliorer la précision du modèle (principe du « garbage in – garbage out »). Cette sélection des caractéristiques peut se faire de nombreuses manières, notamment en sélectionnant manuellement des caractéristiques jugées cliniquement pertinentes, en explorant les corrélations statistiques entre variables prédictives (méthode non supervisée : élimine les variables redondantes) ou en étudiant la relation entre les variables prédictives et les variables à prédire (méthode supervisée : élimine les variables non informatives) (**Annexe 6**).

2.3.2 Entrainement (training)

Séparation des données et validation croisée (split, cross validation)

Les données ayant servi à l'entrainement du modèle ne doivent pas être utilisées pour l'évaluation au risque de surestimer les performances. Il est donc crucial de séparer données d'entrainement (training set) et données d'évaluation (testing set) : c'est le train-test split (aussi appelé hold out). Les données sont divisées en deux de manière aléatoire avec classiquement un ratio de 7:3 (train set = 70%, test set = 30%) (**Annexe 7**). Le test set ne servira qu'à la toute fin du processus une fois que les modèles entrainés sur le training set seront optimisés et que le meilleur sera sélectionné. Pour éviter toute fuite d'information du test set vers le train set et garder un test set le plus indépendant que possible, le split est réalisé avant certaines étapes comme le scaling ou le sampling.

Cependant, pour optimiser et sélectionner la meilleure version du meilleur modèle en vue d'une évaluation finale sur le test set, il faut pouvoir au préalable valider les prédictions de notre modèle sur le même training set. Une nouvelle séparation des données est donc réalisée au sein du training set avec création d'un set de validation (validation set). Mais cette nouvelle mise à l'écart de données réduit la quantité de données d'entrainement, entraine une perte de tendances statistiques et pose le problème d'élévation de la variance. C'est pourquoi une validation croisée avec généralement 5 plis (5-fold cross validation) est utilisée

permettant d'entrainer et de valider les modèles choisis sur l'entièreté du training set réduisant ainsi les biais et la variance.

Échantillonnage de la variable à prédire (sampling)

En épidémiologie, les classes de la variable à prédire sont souvent très déséquilibrées ce qui pénalise les algorithmes de machine learning qui sont meilleurs dans la prédiction de la classe majoritaire. Il est donc possible de rééquilibrer artificiellement ces proportions par des méthodes de sous ou sur échantillonnage en supprimant une partie de la classe majoritaire ou en ajoutant des observations synthétiques au sein de la classe minoritaire (Synthetic minority oversampling technique (SMOTE)) [66,102]. Cette étape n'est réalisée que sur le train set. Les deux approches peuvent être combinées. De cette façon, l'algorithme dispose d'un plus grand nombre d'exemples de la classe minoritaire à partir desquels il peut apprendre.

Choix du modèle

Des modèles de machine learning aux caractéristiques différentes existent. Parmi les modèles supervisés, on peut citer 5 principaux estimateurs couramment utilisés : la régression logistique/linéaire (L regression), les K voisins les plus proches (K-Nearest Neighbors), les machines à vecteurs de support (Support Vector Machines), les arbres de décision (Decision trees) et les réseaux de neurones (Neural Networks) (**Figure 2**).

Il n'existe pas de modèle idéal mais certains sont mieux adaptés à différents types de données et à différents problèmes (**Annexe 8**). Le choix va dépendre tout d'abord de la question posée : classification, régression, clustering ou réduction de dimensionnalité. Ensuite, la quantité de données est un élément critique. Le nombre idéal d'observations nécessaire va dépendre de la quantité de donnée disponible mais aussi de la quantité et de la qualité des variables. Là où certains algorithmes non supervisés peuvent être utilisés sur des populations très restreintes (<100) d'autres ne seront pas applicables [103]. Enfin le temps de calcul entre aussi en compte car il ne dépend pas seulement de la puissance de l'appareil utilisé ou de la quantité de données traitée mais aussi du type d'algorithme.

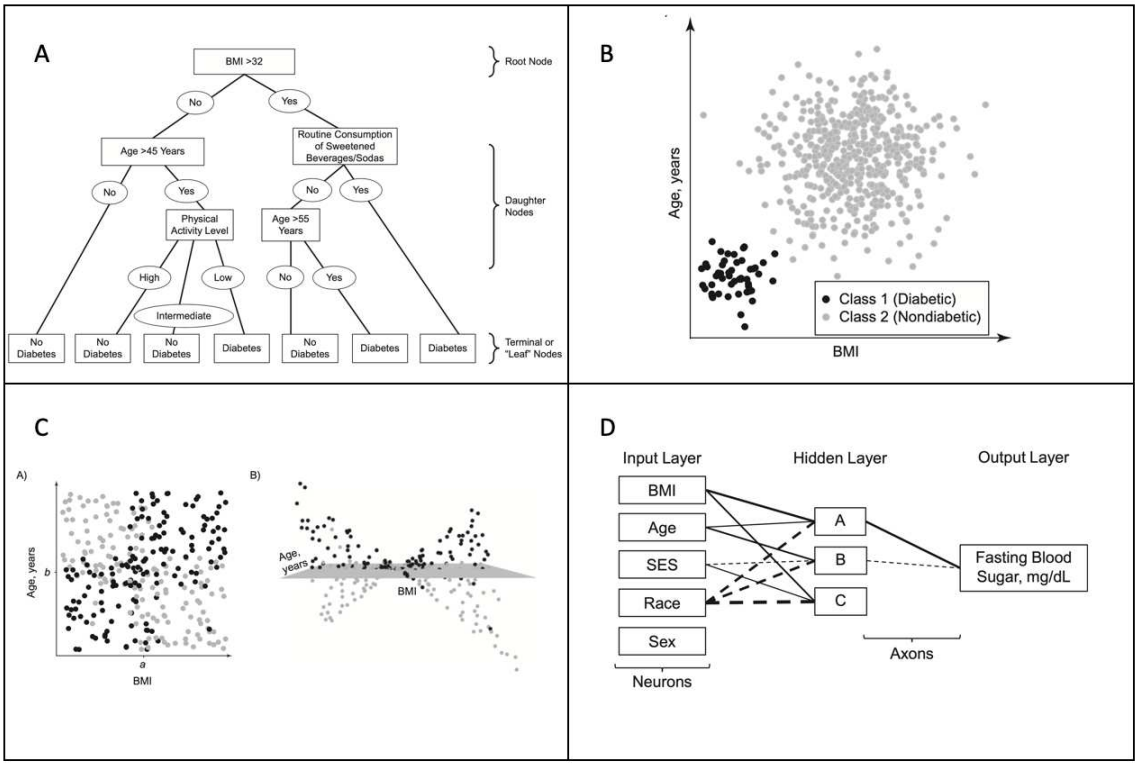


Figure 2. Représentation de 4 modèles supervisés de machine learning populaires : les arbres de décision (Decision tree) (**A**); les K voisins les plus proches (K-Nearest Neighbors) (**B**); les machines à vecteurs de support (Support Vector Machines) (**C**), les réseaux de neurones (Neural Networks) (**D**) [96].

Ensemble learning

Ces estimateurs précédemment décrits peuvent ensuite être assemblés. Ces méthodes d'ensemble permettent d'améliorer les performances des modèles en créant une population d'estimateurs soit en entrainant le même estimateur sur différentes versions d'un ensemble de données soit en entrainant des estimateurs qualitativement différents sur les mêmes données puis en combinant leurs résultats. C'est l'application du principe du « wisdow of the crowd » : l'opinion collective d'un groupe d'individus divers et indépendants est plus juste que celle d'un seul expert.

Cette approche peut être utilisée soit dans le cadre du même estimateur, soit dans le cadre d'approches imbriquées. Quatre méthodes sont principalement décrites : le Bagging utilisé dans les forêts d'arbres aléatoires (Random Forest); le Boosting utilisé par le gradient

boosting; le Bayesian model averaging (BMA) et le Stacking ou super learnner (**Annexe 9**) [96].

Optimisation du model (tuning)

Tous les algorithmes de machine learning possèdent divers hyper-paramètres (options des réglages) qui doivent être ajustés afin de fournir une prédiction valide et précise. On peut citer par exemple le taux d'apprentissage dans les réseaux neuronaux, C et sigma dans une machine à vecteurs de support, ou k dans l'algorithme des plus proches voisins (**Annexe 10**). Ce processus d'ajustement est appelé « tuning ». Cette étape est complexe car il s'agit de spécifier un ensemble de valeurs pour plusieurs hyper-paramètres puis les estimateurs doivent être construits pour tester toutes les combinaisons de valeurs.

L'automatisation de cette étape est donc nécessaire. Des approches de recherche en grille ou aléatoires (grid ou random search) permettent de pré-spécifier une multitude de valeurs possibles pour un grand nombre d'hyper-paramètres nécessaires requis par le modèle (**Annexe 11**) [104]. La limite de cette approche sera l'intensité de calcul requise.

Choix de la métrique

Enfin, beaucoup de méthodes de mesure peuvent être utilisées (l'erreur quadratique moyenne, l'accuracy, precision–recall, l'AUC et le score F1) selon le type de prédiction recherché (continue ou catégorielle). Ces méthodes fournissent des indications sur les performances globales du modèle et leur choix est une question complexe car elles peuvent être prises à défaut, et particulièrement en cas de déséquilibre entre classe minoritaire et majoritaire du label. Prenons l'exemple d'un algorithme qui ne prédirait que des malades dans une population où la prévalence de la pathologie en question ne serait que de 4%. L'accuracy (taux de prédictions correctes) serait de 96%.

La métrique la plus couramment utilisée est l'analyse de la courbe ROC [105]. Elle a l'avantage de fournir un score unique (l'AUC) qui peut être utilisé pour comparer d'autres modèles, contrairement aux autres métriques qui sont inconstamment rapportées. A partir de celle-ci, elle permet aussi d'établir une valeur seuil permettant de faire varier sensibilité/spécificité selon la problématique étudiée (diagnostique, dépistage).

2.3.3 Évaluation

C'est l'étape finale où le modèle préalablement entrainé et sélectionné va être évalué sur un jeu de données inédit pour celui-ci : le test set. Il pourra ensuite être comparé à d'autre modèles considérés comme gold standards. Il se posera ensuite la question de la pertinence et de la généralisation de ce nouveau modèle.

3 Fibrillation atriale et machine learning

Nous assistons depuis quelques années à une augmentation importante du nombre de publications sur le thème du ML en médecine et tout particulièrement dans le domaine de la FA (**Figure 3**). Le ML est principalement utilisé pour la détection de la FA, la prédiction de récurrences post ablation, la prédiction de complications ou l'analyse de données provenant d'objet connectés [106,107]. Cette application du ML dans le domaine de la FA est principalement due à deux phénomènes : de nouvelles sources de données disponibles et le développement d'outils facilitant l'application du ML.

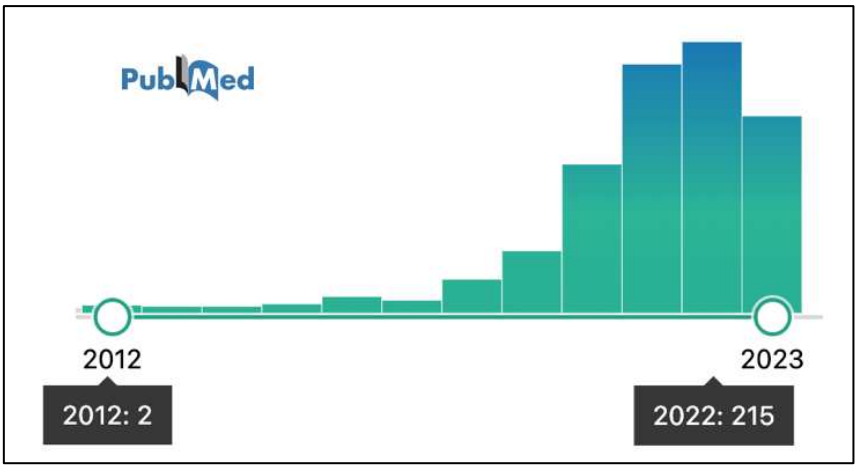


Figure 3. Évolution du nombre de publications depuis 10 ans sur le thème de la fibrillation atriale et du machine learning.
Recherche sur le site du NCBI (https://pubmed.ncbi.nlm.nih.gov/) avec les termes suivants : (atrial fibrillation[Title/Abstract]) AND ((machine learning[Title/Abstract]) OR (artificial intelligence[Title/Abstract]) OR (deep learning[Title/Abstract])).

De nouvelles données

Ces dernières années ont vu l'avènement de la eSanté (electronic health ou eHealth) et plus particulièrement des big data dans le domaine de la FA. On peut distinguer deux principales sources de big data particulièrement adaptées pour l'application du ML : les dossiers médicaux électroniques (electronic health records ou EHR) et la Santé connectée (mobile health ou mHealth).

Riches d'un grand nombre de variables, les dossiers médicaux électroniques permettent la constitution de bases données extrêmement complexes de plusieurs millions d'observations [108,109]. Ces bases ont pour avantage de reposer sur des données de vie réelle, uniformisées (Classification internationale des Maladies), continuellement mises à jour, exhaustives à l'échelle d'une population avec un long suivi dans le temps. Elles constituent un terrain d'application particulièrement intéressant pour le ML.

Par ailleurs, la Santé connectée par les téléphones mobiles, les montres connectées, les prothèses cardiaques implantables et autres capteurs physiologiques portatifs génèrent en continu un flux d'information massif sur la FA [12]. Ces appareils permettent aussi la constitution très rapide de volumineuses cohortes pour le screening de masse [110]. L'utilisation du ML sur ces données recueillies est un champ de recherche très actif [106].

De nouveaux outils

Il est souvent rare pour des épidémiologistes de formation d'avoir les compétences nécessaires à la fois en statistiques, en science de l'informatique (computing science) et en programmation pour tirer pleinement parti du ML. C'est en partie à cause de la popularité persistante de langages de programmation closed-source (SAS, Stata) répandus dans le domaine de l'épidémiologie mais peu adaptés pour le machine learning.

Cependant, les freins à la démocratisation du ML se sont considérablement réduits grâce à l'introduction de plateformes et logiciels open-source facilitant la construction, l'optimisation et l'application de modèles de ML. Ces outils extrêmement puissants sont largement disponibles au sein de librairies online. On peut citer en exemple TensorFlow dans R (R Foundation for Statistical Computing, Vienne, Autriche) ; Enterprise Miner Neural Network et AutoNeural dans SAS (SAS Institute, Inc., Cary, Caroline du Nord) ; et sklearn (https://scikit-learn.org/stable/) et TensorFlow dans Python (Python Software Foundation, Wilmington, Delaware).

4 Problématique

Les patients atteints de FA constituent une population très hétérogène en termes de caractéristiques et en termes de pronostic. Une meilleure caractérisation phénotypique à partir de données cliniques, biologiques et d'imagerie pourrait permettre l'essor d'une médecine personnalisée plus performante. Cependant les méthodes statistiques classiques permettant l'élaboration de modèles prédictifs, que ce soit pour la survenue de FA, de ses récidives ou de ses complications, sont limitées et ne permettent pas d'exploiter pleinement les données actuellement recueillies.

Dans cette quête d'une médecine personnalisée et de modèles plus précis et plus performants, le ML suscite depuis une décennie un engouement scientifique certain. Mais quand est-il réellement de la valeur ajoutée et de l'applicabilité de cette approche ?

OBJECTIFS

L'objectif de cette thèse est de préciser l'applicabilité et l'apport de nouvelles approches d'intelligence artificielle appelées ML dans le domaine de l'épidémiologie de la FA à travers différentes questions scientifiques et différentes sources de données.

Ce travail se divise en deux parties. La première comprenant 3 études explore le développement et l'application de modèles supervisés de ML dans la prédiction d'évènements cardiovasculaires majeurs liées à la FA à partir de bases de données nationales et internationales. La seconde comprenant 2 études explore quant à elle l'hétérogénéité et la complexité clinique des patients atteints de FA à l'aide d'algorithmes non supervisés de ML et évalue leur association avec le pronostic des patients à partir de bases de données locorégionales.

TRAVAUX

1 Première partie : Modèles supervisés et prédiction de la fibrillation atriale ou d'événements cardiovasculaires indésirables associés à la fibrillation atriale

1.1 Étude 1 : Prédiction de la fibrillation atriale incidente après un accident vasculaire cérébral à l'aide d'une approche de type machine learning : Une étude de cohorte nationale française

***Bisson A**, Lemrini Y, El-Bouri W, Bodin A, Angoulvant D, Lip GYH, Fauchier L. **Prediction of incident atrial fibrillation in post-stroke patients using machine learning: a French nationwide study**. Clin Res Cardiol. 2023 Jun;112(6):815-823. doi: 10.1007/s00392-022-02140-w. doi: 10.1007/s00392-022-02140-w.* [111]

1. Objectif de l'étude

L'objectif de cette première étude était de développer un algorithme de ML capable de prédire la survenue de FA dans les suites d'un AVCi et de comparer les performances de celui-ci aux scores cliniques déjà existants.

2. Méthode

Nous avons pour cela constitué rétrospectivement une cohorte nationale de patients victimes d'AVCi à partir des données du PMSI. Au total, 240,306 patients victimes d'AVCi et sans antécédent de FA au préalable ont été inclus. Après une étape de sélection des variables d'importance, 5 algorithmes supervisés de ML ont été entrainés et optimisés à partir d'une sous cohorte de dérivation pour prédire la survenue ou non de FA durant le suivi. Le meilleur des 5 modèles a ensuite été sélectionné pour être évalué sur une sous cohorte de validation et comparé aux scores cliniques préexistants.

3. Résultats

Au terme d'un suivi moyen de 7.5±11.1 mois, 14,095 patients (5.6%) (âge moyen : 77.6±10.6 ans, 50.3% de femmes) ont développé de la FA. Parmi les modèles de ML, celui présentant les meilleures performances était un réseau de neurones avec une AUC de 0.77 (95% CI 0.76-0.78) sur la cohorte de validation. Comparé aux scores cliniques traditionnels, le réseau de neurones était statistiquement supérieur au score CHA_2DS_2-VASc, au score de risque de Framingham, au score HAVOC et au score C_2HEST ($p < 0,0001$). L'augmentation nette de l'indice de reclassification ($p < 0,0001$) et l'analyse de la courbe de décision montraient une amélioration des taux de reclassification pour établir le risque de FA.

4. Conclusion

Le réseau de neurones a obtenu les meilleures performances dans la prédiction de la FA après un AVCi par rapport à d'autres algorithmes de ML et aux scores de risque clinique traditionnels. Cette méthode prometteuse peut aider à la stratification des risques pour la prise de décision en relation avec une stratégie de dépistage de la FA chez les patients après un AVCi.

Clinical Research in Cardiology
https://doi.org/10.1007/s00392-022-02140-w

ORIGINAL PAPER

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

Arnaud Bisson[1,2,3] · Yassine Lemrini[1] · Wahbi El-Bouri[4] · Alexandre Bodin[1] · Denis Angoulvant[1,2] · Gregory Y. H. Lip[4] · Laurent Fauchier[1]

Received: 6 November 2022 / Accepted: 7 December 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany 2022

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 CHA_2DS_2-VASc score, Framingham risk score, HAVOC score and C_2HEST 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.

Denis Angoulvant, Gregory Y. H. Lip and Laurent Fauchier are joint senior authors.

✉ Arnaud Bisson
arnaud.bisson@univ-tours.fr

1 Service de Cardiologie, Centre Hospitalier Régional Universitaire et Faculté de Médecine de Tours, 2 Boulevard Tonnellé, 37000 Tours, France

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

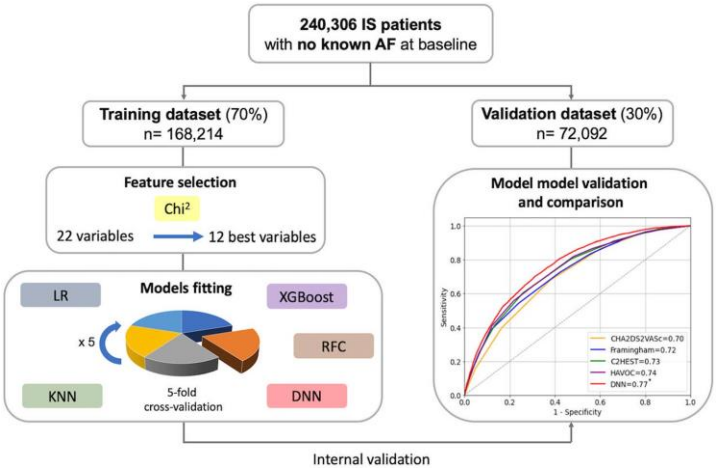
3 Service de Cardiologie, Centre Hospitalier Régional d'Orléans, Orléans, France

4 Liverpool Centre for Cardiovascular Science, University of Liverpool, Liverpool Heart and Chest Hospital, Liverpool, UK

Published online: 17 December 2022

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



Keywords Atrial fibrillation · Machine learning · Ischemic stroke · Prediction

Introduction

Atrial fibrillation (AF) is the most common arrhythmia affecting 2–4% of adults and is responsible for one in five 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 8–14 [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 as C_2HEST 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.

Method

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 1546 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 (i.e., the health problem that justified admission to hospital), the related diagnosis (i.e., potential chronic disease or health state during hospital stay), or the significantly associated diagnosis (i.e., 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 (Fig. 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).

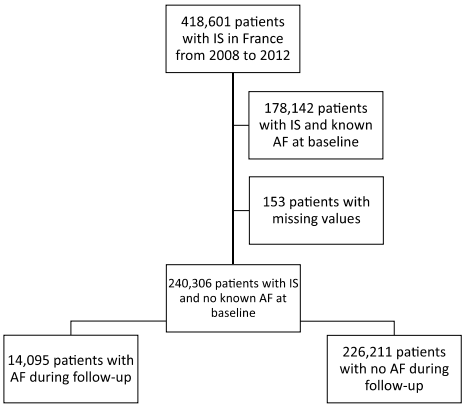


Fig. 1 Flowchart of the study

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 Chi^2 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 fivefold cross-validation: the training set was split into fivefold validation sets. For each validation set, the corresponding training set was given by the remaining fourfolds. Each candidate model was fitted and hyperparameters tuned in each training set. The performance measure was then performed on the validation sets (Fig. 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.

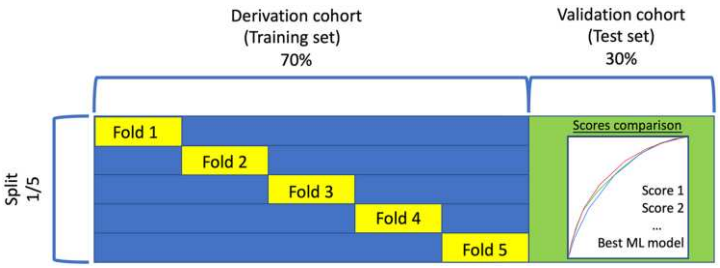


Fig. 2 Machine learning methods. Model training using five different algorithms and fivefold cross-validation

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 CHA_2DS_2-VASc score (congestive heart failure, hypertension, age ≥ 75 [*2 points*], diabetes, stroke [*2 points*], vascular disease, age 65–74, female), C_2HEST 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].

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 Chi^2 test was used to compare categorical data. Receiver operating characteristic curves were constructed, and Harrell C indexes (i.e., 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 with 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 $P < 0.05$.

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 (CHA_2DS_2-VASc score, C_2HEST 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

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	77.6 ± 10.6	70.8 ± 15.7	< 0.0001
Age ≥ 75 years, *n* (%)	9755 (69.2%)	110,782 (49%)	< 0.0001
Gender (female), *n* (%)	7082 (50.3%)	107,266 (47.4%)	< 0.0001
Underlying diseases, *n* (%)			
Hypertension	11,745 (83.3%)	141,045 (62.3%)	< 0.0001
Diabetes mellitus	4083 (29%)	50,977 (22.5%)	< 0.0001
Congestive heart failure	6261 (44.4%)	33,162 (14.7%)	< 0.0001
Vascular disease	6907 (49%)	70,636 (31.2%)	< 0.0001
CHA_2DS_2-VASc score, *n* (%)			
Score = 2	161 (1.1%)	16,992 (7.5%)	< 0.0001
Score = 3	523 (3.7%)	33,266 (14.7%)	< 0.0001
Score = 4	1192 (8.5%)	39,300 (17.4%)	< 0.0001
Score = 5	2596 (18.4%)	50,585 (22.4%)	< 0.0001
Score = 6	3903 (27.7%)	48,443 (21.4%)	< 0.0001
Score = 7	3441 (24.4%)	26,634 (11.8%)	< 0.0001
Score = 8	1881 (13.4%)	9214 (4.1%)	< 0.0001
Score = 9	398 (2.8%)	1777 (0.8%)	< 0.0001
Comorbidities			
Systemic embolism	807 (5.7%)	6261 (2.8%)	< 0.0001
Coronary artery disease	4969 (35.3%)	39,652 (17.5%)	< 0.0001
Obesity	2071 (14.7%)	22,901 (10.1%)	< 0.0001
Abnormal renal function	5393 (38.3%)	38,618 (17.1%)	< 0.0001
Liver disease	593 (4.2%)	6705 (3%)	< 0.0001
Anemia	3980 (28.2%)	31,165 (13.8%)	< 0.0001
Lung disease	3661 (26%)	35,320 (15.6%)	< 0.0001
Cancer within preceding 5 years	3056 (21.7%)	37,400 (16.5%)	< 0.0001
Inflammatory diseases	1470 (10.4%)	14,186 (6.3%)	< 0.0001
Alcohol-related diagnoses	954 (6.8%)	17,680 (7.8%)	< 0.0001
Thyroid disease	6040 (42.9%)	68,247 (30.2%)	< 0.0001
Dyslipidemia	5793 (41.1%)	69,428 (30.7%)	< 0.0001
PM-ICD	1643 (11.7%)	7201 (3.2%)	< 0.0001
Valvular disease	2780 (19.7%)	15,121 (6.7%)	< 0.0001
Tobacco smoking	1415 (10%)	28,840 (12.7%)	< 0.0001

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, Fig. 3).

When compared to the previously described clinical scores, DNN showed statistically significantly higher predictive ability: CHA_2DS_2-VASc score 0.702 (95% CI 0.694–0.709), C_2HEST 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, Fig. 4a).

The DNN had category free net reclassification improvement compared with the CHA_2DS_2-VASc score (64% [$P < 0.0001$]), C_2HEST score (52% [$P < 0.0001$]), HAVOC score (38% [$P < 0.0001$]) and modified Framingham risk score (48% [$P < 0.0001$]). Using decision curve analysis,

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

Scores	OR (95% CI)	AF prevalence per quartile (%)			
		Q1	*Q2*	*Q3*	*Q4*
C_2HEST*	2.32 (2.24–2.4)	1.7	3.4	6.9	16.4
CHA_2DS_2VASc*	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

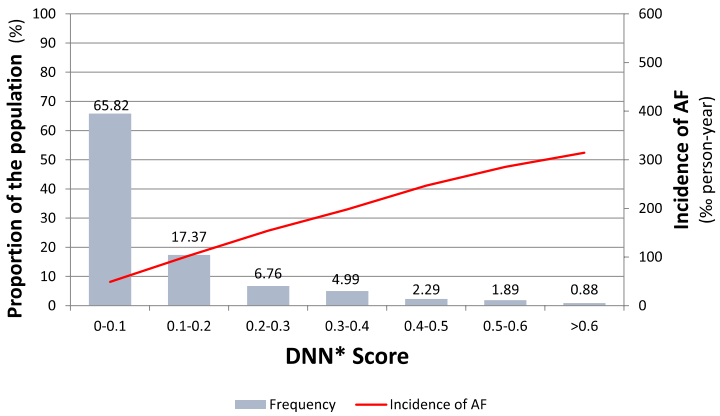


Fig. 3 Distribution of the population and incident rate of atrial fibrillation (AF) according to the deep neural network score. *DNN* deep neural network

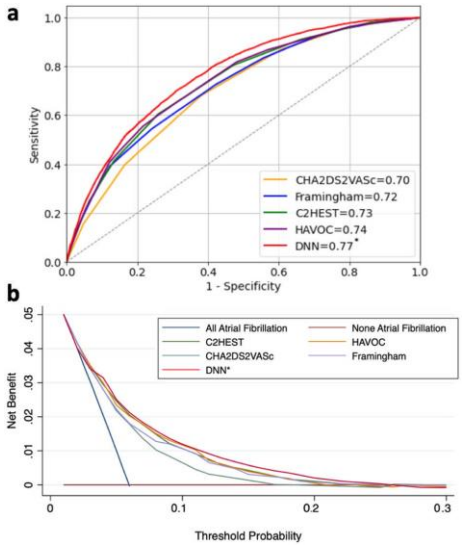


Fig. 4 **a** Receiver operating characteristic curves of incident atrial fibrillation after ischemic stroke. *DNN* deep neural network, $P<0.0001$ vs Framingham risk score, CHA_2DS_2VASc score, HAVOC score and C_2HEST score. **b** decision curve analyses for the deep neural network algorithm and C_2HEST, CHA_2DS_2-VASc, HAVOC and Framingham risk scores. *DNN* deep neural network

our DNN showed better clinical usefulness compared with the traditional scores (Fig. 4b).

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 held-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 with 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 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 (e.g., 0.578 for HAVOC score, 0.572 for CHA_2DS_2-VASc score), notwithstanding the caveat that some scores (CHA_2DS_2-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.

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. While 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.

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.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00392-022-02140-w.

Data availability Data will not be available according to the French legislation.

Declarations

Conflict of interest 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.

References

1. Benjamin EJ, Muntner P, Alonso A, Bittencourt MS, Callaway CW, Carson AP, Chamberlain AM 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. Guo Y, Tian Y, Wang H, Si Q, Wang Y, Lip G (2015) Prevalence, incidence, and lifetime risk of atrial fibrillation in China: new insights into the global burden of atrial fibrillation. Chest 147(1):109–119. https://doi.org/10.1378/chest.14-0321
3. Burdett P, Lip GYH (2022) Atrial fibrillation in the UK: predicting costs of an emerging epidemic recognizing and forecasting the cost drivers of atrial fibrillation-related costs. Eur Heart J 8(2):187–194. https://doi.org/10.1093/ehjqcco/qcaa093
4. Hart RG, Pearce LA, Aguilar MI (2007) Meta-analysis: antithrombotic therapy to prevent stroke in patients who have nonvalvular atrial fibrillation. Ann Intern Med 146(12):857–867. https://doi.org/10.7326/0003-4819-146-12-200706190-00007
5. Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomström-Lundqvist C, Boriani G 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. Eur Heart J 42(5):373–498. https://doi.org/10.1093/eurheartj/ehaa612
6. Chao T-F, Joung B, Takahashi Y, Lim TW, Choi E-K, Chan Y-H, Guo Y et al (2022) 2021 Focused update consensus guidelines of the Asia Pacific Heart rhythm society on stroke prevention in atrial fibrillation: executive summary. Thromb Haemost 122(1):20–47. https://doi.org/10.1055/s-0041-1739411
7. Kishore A, Vail A, Majid A, Dawson J, Lees KR, Tyrrell PJ, Smith CJ (2014) Detection of atrial fibrillation after ischemic stroke or transient ischemic attack: a systematic review and meta-analysis. Stroke 45(2):520–526. https://doi.org/10.1161/STROKEAHA.113.003433
8. Sposato LA, Cipriano LE, Saposnik G, Vargas ER, Riccio PM, Hachinski V (2015) Diagnosis of atrial fibrillation after stroke and transient ischaemic attack: a systematic review and meta-analysis. Lancet Neurol 14(4):377–387. https://doi.org/10.1016/S1474-4422(15)70027-X
9. Bisson A, Bodin A, Clementy N, Babuty D, Lip GYH, Fauchier L (2018) Prediction of incident atrial fibrillation according to gender in patients with ischemic stroke from a nationwide cohort. Am J Cardiol 121(4):437–444. https://doi.org/10.1016/j.amjcard.2017.11.016
10. Li Y-G, Bisson A, Bodin A, Herbert J, Grammatico-Guillon L, Joung B, Wang Y-T, Lip GYH, Fauchier L (2019) C2 HEST score and prediction of incident atrial fibrillation in poststroke patients: a french nationwide study. J Am Heart Assoc 8(13):e012546. https://doi.org/10.1161/JAHA.119.012546
11. Bishop CM (2006) Pattern recognition and machine learning, 3rd edn. Springer, New York
12. Hill NR, Ayoubkhani D, McEwan P, Sugrue DM, Farooqui U, Lister S, Lumley M 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 S, Sandler B, Johnston E, Pollock KG, Hill NR, Gordon J, Tsang C, Khan S, Ng FS, Farooqui U (2020) Detecting undiagnosed atrial fibrillation in UK primary care: validation of a machine learning prediction algorithm in a retrospective cohort study. Eur J Prevent Cardiol. https://doi.org/10.1177/2047487320942338
14. Lip GYH, Genaidy A, Tran G, Marroquin P, Estes C, Sloop S (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. Thromb Haemost 122(1):142–150. https://doi.org/10.1055/a-1467-2993
15. Couris CM, Polazzi S, Olive F, Remontet L, Bossard N, Gomez F, Schott A-M, Mitton N, Colonna M, Trombert B (2009) Breast cancer incidence using administrative data: correction with sensitivity and specificity. J Clin Epidemiol 62(6):660–666. https://doi.org/10.1016/j.jclinepi.2008.07.013
16. Fauchier L, Clementy N, Pelade C, Collignon C, Nicolle E, Lip GYH (2015) Patients with ischemic stroke and incident atrial fibrillation: a nationwide cohort study. Stroke 46(9):2432–2437. https://doi.org/10.1161/STROKEAHA.115.010270
17. Lip GYH, Nieuwlaat R, Pisters R, Lane DA, Crijns HJGM (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):263–272. https://doi.org/10.1378/chest.09-1584
18. Li Y-G, Pastori D, Farcomeni A, Yang P-S, Jang E, Joung B, Wang Y-T, Guo Y-T, Lip GYH (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):510–518. https://doi.org/10.1016/j.chest.2018.09.011
19. Kwong C, Ling AY, Crawford MH, Zhao SX, Shah NH (2017) A clinical score for predicting atrial fibrillation in patients with cryptogenic stroke or transient ischemic attack. Cardiology 138(3):133–140. https://doi.org/10.1159/000476030
20. Schnabel RB, Sullivan LM, Levy D, Pencina MJ, Massaro JM, D'Agostino RB, Newton-Cheh C et al (2009) Development of a risk score for atrial fibrillation (Framingham Heart Study): a community-based cohort study. Lancet (London, England) 373(9665):739–745. https://doi.org/10.1016/S0140-6736(09)60443-8
21. DeLong ER, DeLong DM, Clarke-Pearson DL (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 44(3):837–845
22. Pencina MJ, D'Agostino RB, D'Agostino RB, Vasan RS (2008) Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat Med 27(2):157–172. https://doi.org/10.1002/sim.2929. (**discussion 207–212**)
23. Vickers AJ, Cronin AM, Elkin EB, Gonen M (2008) Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. BMC Med Inform Decis Mak 8:53
24. Bisson A, Clementy N, Bodin A, Angoulvant D, Babuty D, Lip GYH, Fauchier L (2017) Relationship of preexisting cardiovascular comorbidities to newly diagnosed atrial fibrillation after ischemic stroke. Stroke 48(10):2878–2880. https://doi.org/10.1161/STROKEAHA.117.018251
25. Olier I, Ortega-Martorell S, Pieroni M, Lip GYH (2021) How machine learning is impacting research in atrial fibrillation: implications for risk prediction and future management. Cardiovasc Res 117(7):1700–1717. https://doi.org/10.1093/cvr/cvab169
26. Chua W, Purmah Y, Cardoso VR, Gkoutos GV, Tull SP, Neculau G, Thomas MR et al (2019) Data-driven discovery and validation of circulating blood-based biomarkers associated with prevalent atrial fibrillation. Eur Heart J 40(16):1268–1276. https://doi.org/10.1093/eurheartj/ehy815
27. Zheng X, Wang F, Zhang J, Cui X, Jiang F, Chen N, Zhou J, Chen J, Lin S, Zou J (2022) Using machine learning to predict atrial fibrillation diagnosed after ischemic stroke. Int J Cardiol 347(January):21–27. https://doi.org/10.1016/j.ijcard.2021.11.005
28. Lip GYH, Tran G, Genaidy A, Marroquin P, Estes C, Landsheft J (2022) Improving dynamic stroke risk prediction in non-anticoagulated patients with and without atrial fibrillation: comparing common clinical risk scores and machine learning algorithms. Eur Heart J 8(5):548–556. https://doi.org/10.1093/ehjqcco/qcab037
29. Lip GYH, Genaidy A, Tran G, Marroquin P, Estes C (2021) Incident atrial fibrillation and its risk prediction in patients developing COVID-19: a machine learning based algorithm approach. Eur J Intern Med 91(September):53–58. https://doi.org/10.1016/j.ejim.2021.04.023
30. Skrebelyte-Strøm L, Rønning OM, Dahl FA, Steine K, Kjekshus H (2022) Prediction of occult atrial fibrillation in patients after cryptogenic stroke and transient ischaemic attack: PROACTIA. Europace. https://doi.org/10.1093/europace/euac092
31. Nielsen JC, Lin Y-J, de Oliveira Figueiredo MJ, Shamloo AS, Alfie A, Boveda S, Dagres N 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 22(8):1147–1148. https://doi.org/10.1093/europace/euaa065
32. Hart RG, Sharma M, Mundl H, Kasner SE, Bangdiwala SI, Berkowitz SD, Swaminathan B et al (2018) Rivaroxaban for stroke prevention after embolic stroke of undetermined source. N Engl J Med 378(23):2191–2201. https://doi.org/10.1056/NEJMoa1802686
33. Diener H-C, Sacco RL, Donald Easton J, Granger CB, Bernstein RA, Uchiyama S, Kreuzer J et al (2019) Dabigatran for prevention of stroke after embolic stroke of undetermined source. N Engl J Med 380(20):1906–1917. https://doi.org/10.1056/NEJMoa1813959
34. Poli S, Meissner C, Baezner HJ, Kraft A, Hillenbrand F, Hobohm C, Liman J 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. Eur Heart J 42(Supplement_1):ehab724.2070. https://doi.org/10.1093/eurheartj/ehab724.2070
35. Sanna T, Diener H-C, Passman RS, Di Lazzaro V, Bernstein RA, Morillo CA, Rymer MM et al (2014) Cryptogenic stroke and underlying atrial fibrillation. N Engl J Med 370(26):2478–2486. https://doi.org/10.1056/NEJMoa1313600
36. Gladstone DJ, Spring M, Dorian P, Panzov V, Thorpe KE, Hall J, Vaid H et al (2014) Atrial fibrillation in patients with cryptogenic stroke. N Engl J Med 370(26):2467–2477. https://doi.org/10.1056/NEJMoa1311376
37. Rubiera M, Aires A, Antonenko K, Lémeret S, Nolte CH, Putaala J, Schnabel RB et al (2022) European Stroke Organisation (ESO) guideline on screening for subclinical atrial fibrillation after stroke or transient ischaemic attack of undetermined origin. Eur Stroke J. https://doi.org/10.1177/23969873221099478

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

1.2 Étude 2 : Prédiction de la mortalité précoce après un diagnostic de fibrillation atriale à l'aide d'une approche de type machine learning : Une étude de cohorte nationale française

***Bisson A**, Lemrini Y, Romiti GF, Proietti M, Angoulvant D, Bentounes S, El-Bouri W, Lip GYH, Fauchier L. **Prediction of early death after atrial fibrillation diagnosis using a machine learning approach: A French nationwide cohort study.** Am Heart J. 2023 Nov;265:191-202. doi: 10.1016/j.ahj.2023.08.006.* [112]

1. Objectif de l'étude

L'objectif de cette seconde étude était de développer un algorithme de ML capable de prédire la survenue de décès précoce suivant le diagnostic de FA et de comparer les performances de celui-ci aux scores cliniques déjà existants.

2. Méthode

C'est à partir d'une cohorte nationale de patients atteints de FA constituée rétrospectivement provenant des données du PMSI que nous avons développé notre modèle. Au total, 2,435,541 ayant une FA nouvellement diagnostiquée ont été inclus. Après une étape de sélection des variables d'importance, 3 algorithmes supervisés de ML ont été entrainés et optimisés à partir d'une sous cohorte de dérivation pour prédire la survenue de décès survenant dans l'année suivant le diagnostic de FA. Le meilleur des 3 modèles a ensuite été sélectionné pour être évalué sur une sous cohorte de validation et comparé aux scores cliniques préexistants.

3. Résultats

Au cours de la première année, 342,005 patients (14,4 %) sont décédés après une moyenne de 83±98 jours (médiane 37 [10-129]). Le meilleur modèle sélectionné était un réseau de neurones avec une AUC de 0,785 (IC 95 %, 0,781-0,789) sur la cohorte de validation. Comparé aux scores de risque cliniques, le modèle sélectionné était supérieur aux scores de CHA_2DS_2-VASc et HAS-BLED et supérieur aux scores dédiés tels que l'Indice de Comorbidité de Charlson et le Score de Risque de Fragilité Hospitalier pour prédire le décès durant l'année suivant le diagnostic de FA (AUC : 0,597 ; 0,562 ; 0,643 ; 0,626 respectivement, $p<0,0001$).

4. Conclusion

Les algorithmes de ML prédisent la mortalité chez les patients atteints de FA avec de meilleures performances que les scores de risque clinique traditionnels développés précédemment. Une approche fondée sur ce type de modèles permettrait de mieux stratifier les patients souffrant de FA et présentant un risque élevé de mortalité toute cause confondue, non cardiovasculaire ou cardiovasculaire, et pourrait *in fine* aider les cliniciens à prendre en charge ces patients de manière holistique et intégrée.

Prediction of early death after atrial fibrillation diagnosis using a machine learning approach: A French nationwide cohort study

Arnaud Bisson, MD [a,b,c,d], **Yassine Lemrini, MD** [a], **Giulio Francesco Romiti, MD** [d,e], **Marco Proietti, MD, PhD** [f,g], **Denis Angoulvant, MD, PhD** [a,b], **Sidahmed Bentounes, MSc** [a], **Wahbi El-Bouri, PhD** [d], **Gregory Y.H. Lip, MD** [d,h,#], **and Laurent Fauchier, MD, PhD** [a,#] ***Tours, France; Orléans, France; Liverpool, United Kingdom; Rome, Italy; Milan, Italy; Milano, Italy and Aalborg, Denmark***

Aims Atrial fibrillation is associated with important mortality but the usual clinical risk factor based scores only modestly predict mortality. This study aimed to develop machine learning models for the prediction of death occurrence within the year following atrial fibrillation diagnosis and compare predictive ability against usual clinical risk scores.

Methods and Results We used a nationwide cohort of 2,435,541 newly diagnosed atrial fibrillation patients seen in French hospitals from 2011 to 2019. Three machine learning models were trained to predict mortality within the first year using a training set (70% of the cohort). The best model was selected to be evaluated and compared with previously published scores on the validation set (30% of the cohort). Discrimination of the best model was evaluated using the C index. Within the first year following atrial fibrillation diagnosis, 342,005 patients (14.4%) died after a period of 83 (SD 98) days (median 37 [10-129]). The best machine learning model selected was a deep neural network with a C index of 0.785 (95% CI, 0.781-0.789) on the validation set. Compared to clinical risk scores, the selected model was superior to the CHA_2DS_2-VASc and HAS-BLED risk scores and superior to dedicated scores such as Charlson Comorbidity Index and Hospital Frailty Risk Score to predict death within the year following atrial fibrillation diagnosis (C indexes: 0.597; 0.562; 0.643; 0.626 respectively. $P < .0001$).

Conclusion Machine learning algorithms predict early death after atrial fibrillation diagnosis and may help clinicians to better risk stratify atrial fibrillation patients at high risk of mortality. (Am Heart J 2023;265:191–202.)

Translational Perspective

Atrial fibrillation is responsible for a substantial proportion of short-term mortality making futile, complex and expensive, cardiovascular procedures/devices or therapies that will not change overall prognosis due to competing risk between cardiovascular and non-cardiovascular death. Machine learning algorithms predict early mortality in atrial fibrillation patients with a better ability than previously developed traditional clinical risk scores. A Machine learning approach may help clinicians to better stratify atrial fibrillation patients at high risk of mortality and may assist physicians in decision-making when managing atrial fibrillation patients in a holistic and integrated care manner.

Background

Atrial fibrillation (AF) is the most common sustained cardiac arrhythmia with an estimated prevalence in adults between 2% and 4%, and is associated with a high burden of morbidity and mortality.[1,2] During the last decades, improvements in the care of patients with AF have been confirmed, including the development of stroke and bleeding risk calculators and the validation of the benefits of oral anticoagulation on stroke and survival.[3–6]

However, mortality is the most frequent major clinical event after the diagnosis of AF with a cumulative

From the [a]Service de Cardiologie, Centre Hospitalier Régional Universitaire et Faculté de Médecine de Tours, Tours, France, [b]EA4245, Transplantation Immunité Inflammation, Université de Tours, Tours, France, [c]Service de Cardiologie, Centre Hospitalier Régional Universitaire d'Orléans, Orléans, France, [d]Liverpool Centre for Cardiovascular Science at University of Liverpool, Liverpool John Moores University and Liverpool Heart & Chest Hospital, Liverpool, United Kingdom, [e]Department of Translational and Precision Medicine, Sapienza – University of Rome, Italy, [f]Department of Clinical Sciences and Community Health, University of Milan, Italy, [g]Division of Subacute Care, IRCCS Istituti Clinici Scientifici Maugeri, Milano, Italy, [h]Danish Center for Health Services Research, Department of Clinical Medicine, Aalborg University, Aalborg, Denmark

[#]Joint senior authors.

Submitted August 14, 2023; accepted August 14, 2023

Reprint requests: Arnaud Bisson, MD, Service de Cardiologie, Centre Hospitalier Regional Universitaire et Faculté de Médecine de Tours, 2 Boulevard Tonnellé 37000, Tours, France.

E-mail address: arnaud.bisson@univ-tours.fr.

0002-8703

© 2023 Elsevier Inc. All rights reserved.

https://doi.org/10.1016/j.ahj.2023.08.006

incidence of 19.5%, 28.2%, and 48.8% at 1, 2, and 5 years respectively and an overall 1-year all-cause mortality which has not improved from 2007 to 2015 despite a significant reduction in the rate of cardiovascular (CV) death.[7-9] This trend has been observed globally, with age-standardized AF mortality rates remaining stable between 1990 and 2017 but increasing in poorer countries.[10] Although stroke is the most feared complication of AF, stroke accounted for only 7.0% of deaths in AF, with sudden cardiac death (22.25%), progressive heart failure (15.1%), and non-CV death (35.8%) accounting for the majority of deaths in the RE-LY trial.[11] Indeed, a high residual risk of stroke and CV complications remains despite the use of oral anticoagulation.[12,13]

Recognizing AF patients particularly at risk of non-CV death is therefore crucial for the management of concomitant non-CV conditions and hence reducing mortality. Even though the existence of an association between contemporary clinical risk scores such as CHA_2DS_2-VASc score or HAS-BLED score and increased risk of all-cause death has been previously reported, these scores are not dedicated for the prediction of death and importantly, the predictive ability is modest.[14] More integrative tools such as the Charlson Comorbidity Index (CCI, as a measure of multimorbidity) or Hospital Frailty Risk Score (HFRS, as a measure of frailty) have been specifically developed to predict mortality but without wide validation in AF populations.[15,16]

Machine learning (ML) is emerging as a new method for improving the prediction of adverse outcomes in the field of CV disease, including AF.[17] Highly effective in large data sets with a multiplicity of variables, it has already been effective for prognostic prediction and decision making in several diseases.[18–20]

This study firstly aimed to train and evaluate ML models for the prediction of death occurrence within the year following AF diagnosis and second, to compare predictive ability again the usual clinical risk scores.

Methods

Study population

This longitudinal cohort study was based on he 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 health care 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.[21–23]

Data for all patients admitted with AF in France from January 2011 to December 2019 were collected from the PMSI. The study included adults (aged ≥18 years) with a diagnosis of AF (ICD-10 code I48 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, 2011, and December 31, 2019. Baseline characteristics corresponded to all the diagnoses collected at discharge of the first hospital stay were a diagnosis of AF was reported, as well as past clinical history in the year before the admission to establish history before AF.

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).

Follow-up and outcomes

Each patient was followed until death or December 31, 2019. All-cause mortality was defined as any death occurring during follow-up and cause of death (CV or non-CV) was identified based on the main diagnosis during hospitalization resulting in death based on ICD-10 codes (CV death: I00–I99 – Diseases of the heart and circulatory system; non-CV death: any other ICD-10 codes). The main outcome was the occurrence of all cause death within the year following the diagnosis of AF.

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 and make the model practical, only most critical variables (features) in the derivation cohort were subse-

quently used for model building after feature selection by a random forest classifier (RFC).

Best selected features then served as an input to train 3 well-accepted ML models, including logistic regression with L2 regularization (LR), RFC and deep neural network (DNN).[17] ML models were implemented in Python using open-source packages: Scikit-learn version 1.1.1.

A random train-test split methodology and cross-validation were then applied, as previously published, and is described as follows[24]: 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. Predictive ability of each ML models was assessed using C index and compared to each other. All models were trained, optimized and evaluated using the same 5-fold cross-validation and were therefore applied on the same whole training set. The model with the best performance was finally selected for analysis on the held-out validation cohort (test set) for final evaluation to evaluate the models' generalization performance and compared to clinical scores.

Model evaluation and comparison

Area under the ROC curve (AUC) of the best ML model selected was calculated on the validation cohort and C index was compared to predictive ability of clinical scores previously described such as the CHA_2DS_2-VASc score (Supplemental Table 1), the HAS-BLED score (Supplemental Table 2), the CCI (Supplemental Table 3), and the HFRS (Supplemental Table 4) for each patient.[3,4,15,16] Although CHA_2DS_2-VASc and HAS-BLED scores are used to evaluate the thromboembolic and bleeding risk in AF patients, they are also predictors of death in AF patients.[14] The CCI is dedicated tool to evaluate the burden of multimorbidity and is associated with an increased risk of all-cause death. CCI uses a system of relative weights to evaluate the impact of a list of chronic conditions on mortality risk, based on the evaluation of 17 chronic conditions. The HFRS was developed to evaluate frailty in hospitalized patients. The score was built according to the ICD-10 codes related to 109 clinical conditions. HFRS was established as an accurate tool to evaluate frailty in comparison to established frailty scales and validated in nationwide French population.[25]

As previously described in the literature, the category of highest thromboembolic risk patients was defined as a CHA_2DS_2-VASc score ≥ 2, the category of highest bleeding risk patients as a HAS-BLED ≥ 3, the category of patients with the most multimorbidity as a CCI ≥ 4 and the category of the frailest patients as a HFRS ≥ 15.

Statistical analysis

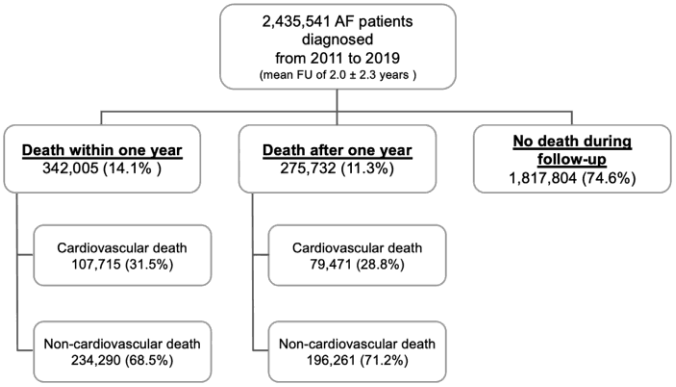
Qualitative variables were described using counts and percentages, and continuous quantitative variables were described as mean ± SD or median [interquartile range]. Comparisons were made using parametric or nonparametric tests, as appropriate: The Wilcoxon signed rank and Kruskal Wallis tests were used for comparing values between 2 independent groups, and the χ^2 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.[26] Net reclassification improvement was calculated according to the methods described by Pencina et al[27] to assess the reclassification performance of the scores. Clinical usefulness and net benefit of the best ML model in comparison to the best clinical score were estimated using decision curve analysis.[28] Multivariate logistic regression was used, and results were expressed as odds ratio (OR) and 95% confidence intervals (95%CI). Analyses were performed using Python version 3.09 and STATA version 16.0 (Stata Corp, College Station, TX). All statistical significance levels were 2-sided, and the significant differences were expressed as $P < 0.05$.

Results

Among 2,435,541 patients diagnosed with AF from 2011 to 2019 and included for analysis, 617,737 (25.4%) deaths were recorded during a mean follow-up of 2.0 ± 2.3 years (median 1.1 [0.07-3.39]) in whom were 187,186 (7.7%) CV death and 430,551 (17.7%) non-CV death (Figure 1). Among patient dead from CV cause, 80,913 (43.2%) died from heart failure and 43,429 (23.2%) died from stroke or systemic embolism. Overall, incidence rates (95% CI) for the study outcomes were accordingly 12.69 %/year (12.66-12.72) for all-cause death, 3.84 %/year (3.83-3.86) for CV death and 8.84 %/year (8.82-3.87) for non-CV death.

Early deaths, within the year following the first episode of AF, occurred in 342,005 patients after a mean time of 83 ± 98 days and represented 55.4% of all deaths and 14.04% of the whole population. These patients were older than patients still alive at 1 year with a mean age of 81.29 ± 10.28 years vs 76.53 ± 12.23 years respectively ($P < .0001$) and had more comorbidities as highlighted by clinical scores (mean CHA_2DS_2-VASc score 3.96 ± 1.53 vs 3.38 ± 1.62, HAS-BLED score 2.61 ± 1.12 vs 2.34 ± 1.19, CCI 4.56 ± 2.67 vs 3.28 ± 2.78 and HFRS 11.69 ± 8.57 vs 8.88 ± 9.19, $P < .0001$ for each) (Table 1). Patients still alive at 1 year had significantly more dyslipidemia (21.9% vs 19.2%, $P < .0001$), obesity (14.6% vs 12.3%, $P < .0001$), sleep apnea syndrome (5.3% vs4.4%, $P < .0001$) and gastroesophageal reflux (2.3% vs 2.0%, $P < .0001$).

Figure 1



Flowchart for the training and validation of models.

Among deaths recorded at 1 year, 107,715 were CV death (31.5%) and 234,290 non-CV death (68.5%). CV deaths occurred earlier than non-CV deaths (mean time 73 ± 95 days for CV death and 88 ± 99 days for non-CV death). Patients who died from non-CV death were younger than patients who died from CV death (mean age 80.66 ± 10.44 vs 82.66 ± 9.78 years, $P < .0001$) (Table 2), and had less prevalent CV comorbidities and more prevalent extra cardiac comorbidities than patients who died from CV death. Hence the CHA_2DS_2-VASc and HAS-BLED scores were lower in patient who died from non-CV death than CV death (3.74 ± 1.51 vs 4.45 ± 1.46, $P < .0001$ and 2.57 ± 1.14 vs 2.71 ± 1.07, $P < .0001$ respectively). Conversely, CCI and HFRS were higher in patients who died from non-CV death than CV death (4.75 ± 2.75 vs 4.14 ± 2.45, $P < .0001$ and 12.32 ± 8.72 vs 10.34 ± 8.05, $P < .0001$ respectively).

Among 110 variables available, only the 18 most important were selected, as adding any other variable did not improve prediction (Supplemental Figure 1). Most critical features selected for analysis were age, presence of cancer metastasis, resuscitated cardiac arrest, cancer, congestive heart failure, decubitus ulcers, renal failure, pneumonia, lung disease, difficulty in walking, malnutrition, anemia, impaired mobility, liver disease, acute renal failure, renal disease, blood transfusion and urinary tract infection (Supplemental Figure 2, Supplemental Table 5).

After hyperparameter tuning using the derivation cohort (training set), ML algorithms had good predictive performances based on C indexes: LR 0.780 (95% CI, 0.778-0.781), RF 0.785 (95% CI, 0.782-0.788), DNN 0.786 (95% CI, 0.783-0.789). The DNN model showed the best predictive performances and was therefore selected as the best model for further evaluation ($P = .0001$ vs LR, $P = .19$ vs RF). The chosen hyperparameters are displayed in Supplemental Table 6.

Once evaluated on the held-out validation cohort (test set), DNN had a good C index: 0.785 (95% CI, 0.781-0.789). The model was well calibrated to the validation cohorts (Supplemental Figure 3). The optimal cut-point was defined as the point closest to the point (0,1) on the ROC curve and was 0.16. Thus, we stratified AF patients into a low-risk group (<0.16) and high-risk group (≥0.16). At this decision threshold, sensitivity was 73.1%, specificity 69.1%, negative predicted value 94% and positive predicted value 28.1%. The incidence of all cause death at 1 year rises in a stepwise fashion from 13.8 per 1000 patients for the first quintile to 352.4 per 1,000 for the fifth quintile (Figure 2).

When compared to the previously described clinical scores, DNN showed significantly higher predictive ability: CHA_2DS_2-VASc score: 0.596 (95% CI, 0.592-0.602), HAS-BLED score: 0.562 (95% CI, 0.557-0.567), CCI: 0.643 (95% CI, 0.638-0.648) and HFRS: 0.626 (95% CI, 0.622-0.631) ($P < .0001$) (Figure 3, Panel A).

The DNN had a positive net reclassification improvement and integrated discriminatory improvement compared with the CHA_2DS_2-VASc score (79% [$P < .0001$] and 13% [$P < .0001$], respectively), HAS-BLED score (82% [$P < .0001$] and 14% [$P < .0001$], respectively), CCI (79% [$P < .0001$] and 12% [$P < .0001$], respectively) and modified HFRS (80% [$P < .0001$] and 13%

Table 1. Baseline characteristics of the AF population according to death before or after 1 year

	Total (n = 2,435,541)	Alive after 1 year (n = 2,093,536)	Dead within the first year (n = 342,005)	P*
Clinical characteristics				
Age, mean ± SD	77.2 ± 12.1	76.53 ± 12.23	81.29 ± 10.28	<.0001
CHA_2DS_2VASc score, mean ± SD	3.5 ± 1.6	3.38 ± 1.62	3.96 ± 1.53	<.0001
HASBLED score, mean ± SD	2.4 ± 1.2	2.34 ± 1.19	2.61 ± 1.12	<.0001
Charlson comorbidity index, mean ± SD	3.5 ± 2.8	3.28 ± 2.78	4.56 ± 2.67	<.0001
Frailty index, mean ± SD	9.3 ± 9.2	8.88 ± 9.19	11.69 ± 8.57	<.0001
Gender (male), n (%)	1,286,313 (52.8)	1,103,154(52.6)	183,159(53.6)	<.0001
Hypertension, n (%)	1,459,065 (59.9)	1,245,027 (59.5)	214,038 (62.6)	<.0001
Diabetes mellitus, n (%)	509,617 (20.9)	426,980 (20.4)	82,919 (24.2)	<.0001
Heart failure, n (%)	882,430 (36.2)	708,360 (33.8)	174,070 (50.9)	<.0001
Valve disease, n (%)	334,730 (13.7)	283,937 (13.6)	50,793 (14.9)	<.0001
Dilated cardiomyopathy, n (%)	169,108 (6.9)	142,331 (6.8)	26,777 (7.8)	<.0001
Coronary artery disease, n (%)	596,709 (24.5)	500,199 (23.9)	96,510 (28.2)	<.0001
Previous MI, n (%)	119,230 (4.9)	96,876 (4.6)	23,084 (6.7)	<.0001
Previous PCI, n (%)	97,241 (4.0)	83,207 (4)	14,034 (4.1)	.0004
Previous CABG, n (%)	79,822 (3.3)	67,744 (3.2)	12,078 (3.5)	<.0001
Vascular disease, n (%)	428,526 (17.6)	349,961 (16.7)	78,565 (23)	<.0001
Previous pacemaker or ICD, n (%)	94,691 (3.9)	77,622 (3.7)	17,069 (5)	<.0001
Ischemic stroke, n (%)	173,563 (7.1)	141,315 (6.8)	32,248 (9.4)	<.0001
Previous major bleeding, n (%)	187,806 (7.7)	144,263 (6.9)	43,543 (12.7)	<.0001
Smoker, n (%)	162,715 (6.7)	135,364 (6.5)	27,351 (8)	<.0001
Dyslipidemia, n (%)	524,489 (21.5)	458,680 (21.9)	65,809 (19.2)	<.0001
Obesity, n (%)	346,996 (14.2)	304,846 (14.6)	42,150 (12.3)	<.0001
Alcohol related diagnoses, n (%)	122,218 (5.0)	98,136 (4.7)	24,082 (7)	<.0001
Abnormal renal function, n (%)	166,930 (6.9)	127,027 (6.1)	39,903 (11.7)	<.0001
Lung disease, n (%)	403,259 (16.6)	310,776 (14.8)	92,483 (27)	<.0001
Sleep apnea syndrome, n (%)	127,100 (5.2)	111,998 (5.3)	15,102 (4.4)	<.0001
Liver disease, n (%)	88,742 (3.6)	63,429 (3)	25,313 (7.4)	<.0001
Gastroesophageal reflux, n (%)	55,201 (2.3)	48,477 (2.3)	6,724 (2)	<.0001
Thyroid diseases, n (%)	233,230 (9.6)	199,187 (9.5)	34,043 (10)	<.0001
Inflammatory disease, n (%)	145,531 (6.0)	119,309 (5.7)	262,22 (7.7)	<.0001
Anemia, n (%)	414,210 (17.0)	315,096 (15.1)	99,114 (29)	<.0001
Previous cancer, n (%)	428,273 (17.6)	311,607 (14.9)	116,666 (34.1)	<.0001
AF therapeutic strategy				
Medical treatment only, n (%)	2,181,323 (89.6)	1,844,118 (88.1)	337,205 (98.6)	<.0001
Electrical cardioversion, n (%)	206,153 (8.5)	201,911 (9.6)	4,242 (1.2)	<.0001
Atrial fibrillation ablation, n (%)	62,069 (2.6)	61,889 (3)	180 (0.1)	<.0001
AV node ablation, n (%)	13,784 (0.6)	13,280 (0.6)	504 (0.1)	<.0001
Left atrial appendage occlusion, n (%)	2,699 (0.1)	2,656 (0.1)	43 (<0.1)	<.0001

AF, atrial fibrillation; *CABG*, coronary artery bypass graft; *ICD*, implantable cardiac defibrillator; *MI*, myocardial infarction; *PCI*, percutaneous coronary intervention; *PM*, pacemaker.

* Patients alive vs dead at 1 year after AF diagnosis.

[P < .0001], respectively). Using decision curve analysis, our DNN showed better clinical usefulness compared with the traditional scores (Figure 3, Panel B).

Predictive abilities of our DNN model were broadly similar when applied to early non-CV death prediction (C index 0.786 [95%CI, 0.782-0.791]), while slightly lower when applied to early CV death prediction (C index 0.724 [95%CI, 0.716-0.731]) (Figure 4). In both cases, DNN model was significantly superior to all clinical scores described above (P < .0001) even after stratification on age (Supplemental Table 7).

Univariable and multivariable logistic regression analyses for association between clinical risk scores and outcomes occurrence are reported in Table 3. After adjustment on baseline clinical characteristics, the DNN was independently associated with increased risk of all cause death (odd ratio [OR] 1.56 [95% CI, 1.54-1.58] for each decile, OR 4.00 [95% CI, 3.78-4.22] DNN ≥ 0.16 vs <0.16, P < .0001), non-CV death (OR 1.57 [95% CI, 1.55-1.59] for each decile, OR 4.50 [95% CI, 4.21-4.80] DNN≥0.16 vs <0.16, P < .0001) and CV death (OR 1.34 [95% CI, 1.31-1.37] for each decile, OR 2.26 [95% CI, 2.05-2.47] DNN≥0.16 vs <0.16, P < .0001). DNN was also independently associated with increased risk of all cause death when using Cox regression analysis and increased risk of non-CV and CV death when using Fine-Gray competing analysis (Supplemental Table 8).

Table 2. Baseline characteristics of the AF patients dead within the first year following diagnosis of AF according to the cause of death

	Dead within the first year (n = 342,005)	Non-CV death (n = 234,290)	CV death (n = 107,715)	P*
Clinical characteristics				
Age	81.29 ± 10.28	80.66 ± 10.44	82.66 ± 9.78	<.0001
CHA_2DS_2VASc score, mean ± SD	3.96 ± 1.53	3.74 ± 1.51	4.45 ± 1.46	<.0001
HASBLED score, mean ± SD	2.61 ± 1.12	2.57 ± 1.14	2.71 ± 1.07	<.0001
Charlson comorbidity index, mean ± SD	4.56 ± 2.67	4.75 ± 2.75	4.14 ± 2.45	<.0001
Frailty index, mean ± SD	11.69 ± 8.57	12.32 ± 8.72	10.34 ± 8.05	<.0001
Gender (male), *n* (%)	183,159 (53.6)	131,264 (56)	51,895 (48.2)	<.0001
Hypertension, *n* (%)	214,038 (62.6)	144,197 (61.5)	69,841 (64.8)	<.0001
Diabetes mellitus, *n* (%)	82,919 (24.2)	56,164 (24)	26,755 (24.8)	<.0001
Heart failure, *n* (%)	174,070 (50.9)	105,237 (44.9)	68,833 (63.9)	<.0001
Valve disease, *n* (%)	50,793 (14.9)	27,286 (11.6)	23,507 (21.8)	<.0001
Dilated cardiomyopathy, *n* (%)	26,777 (7.8)	15,375 (6.6)	11,402 (10.6)	<.0001
Coronary artery disease, *n* (%)	96,510 (28.2)	57,702 (24.6)	38,808 (36)	<.0001
Previous MI, *n* (%)	23,084 (6.7)	10,805 (4.6)	12,279 (11.4)	<.0001
Previous PCI, *n* (%)	14,034 (4.1)	7,855 (3.4)	6,179 (5.7)	<.0001
Previous CABG, *n* (%)	12,078 (3.5)	7,081 (3)	4,997 (4.6)	<.0001
Vascular disease, *n* (%)	78,565 (23)	46,771 (20)	31,794 (29.5)	<.0001
Previous pacemaker or ICD, *n* (%)	17,069 (5)	10,676 (4.6)	6,393 (5.9)	<.0001
Ischemic stroke, *n* (%)	32,248 (9.4)	14,313 (6.1)	17,935 (16.7)	<.0001
Previous major bleeding, *n* (%)	43,543 (12.7)	32,989 (14.1)	10,554 (9.8)	<.0001
Smoker, *n* (%)	27,351 (8)	20,708 (8.8)	6,643 (6.2)	<.0001
Dyslipidemia, *n* (%)	65,809 (19.2)	43,344 (18.5)	22,465 (20.9)	<.0001
Obesity, *n* (%)	42,150 (12.3)	28,988 (12.4)	13,162 (12.2)	.2
Alcohol related diagnoses, *n* (%)	24,082 (7)	18,606 (7.9)	5,476 (5.1)	<.0001
Abnormal renal function, *n* (%)	39,903 (11.7)	26,662 (11.4)	13,241 (12.3)	<.0001
Lung disease, *n* (%)	92,483 (27)	66,907 (28.6)	25,576 (23.7)	<.0001
Sleep apnea syndrome, *n* (%)	15,102 (4.4)	10,659 (4.5)	4,443 (4.1)	<.0001
Liver disease, *n* (%)	25,313 (7.4)	18,965 (8.1)	6,348 (5.9)	<.0001
Gastroesophageal reflux, *n* (%)	6,724 (2)	4,989 (2.1)	1,735 (1.6)	<.0001
Thyroid diseases, *n* (%)	34,043 (10)	23,364 (10)	10,679 (9.9)	.6
Inflammatory disease, *n* (%)	26,222 (7.7)	19,143 (8.2)	7,079 (6.6)	<.0001
Anemia, *n* (%)	99,114 (29)	76,301 (32.6)	22,813 (21.2)	<.0001
Previous cancer, *n* (%)	116,666 (34.1)	98,778 (42.2)	17,888 (16.6)	<.0001
AF therapeutic strategy				
Medical treatment only, *n* (%)	337,205 (98.6)	231,659 (98.9)	105,546 (98)	<.0001
Electrical cardioversion, *n* (%)	4,242 (1.2)	2,341 (1)	1,901 (1.8)	<.0001
Atrial fibrillation ablation, *n* (%)	180 (0.1)	91 (<0.1)	89 (0.1)	<.0001
AV node ablation, *n* (%)	504 (0.1)	251 (0.1)	253 (0.2)	<.0001
Left atrial appendage occlusion, *n* (%)	43 (<0.1)	22 (<0.1)	21 (<0.1)	.01

AF, atrial fibrillation; *CABG*, coronary artery bypass graft; *CV*, cardiovascular; *ICD*, implantable cardiac defibrillator; *MI*, myocardial infarction; *PCI*, percutaneous coronary intervention; *PM*, pacemaker.

*CV vs non-CV death.

Discussion

In a large nationwide French hospital data base of 2,435,541 AF patients, ML scoring was an effective method for the prediction of 1-year all-cause mortality as well as non-CV and CV death. When compared to contemporary clinical risk scores, this approach had a significantly better predictive ability to target patients with a higher risk of mortality. These findings suggest that ML could have a crucial clinical role in assessing prognostic risk in patients with AF and improve holistic management of AF patients to reduce mortality.

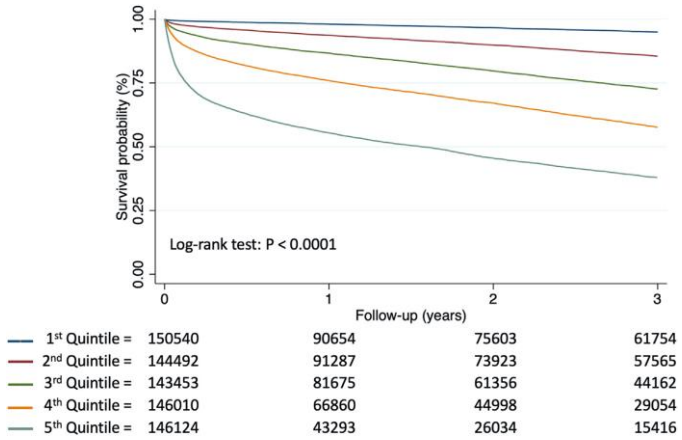
Mortality rates across studies

With an overall 1-year mortality rate of 14.1% (342,005 patients), including 68.5% (234,290 patients) of non-CV deaths, rates reported in our cohort are higher than results previously reported by Singh et al (8% at 1 year) with a similar ratio between CV and non-CV death.[9] However, these cohort included data from hospital-based care but also physician services and patients were therefore younger and had less comorbidities explaining a lower mortality. Our results are more consistent with Piccini et al who reported a 1-year mortality rate of 19.5% with slightly older and comorbid patients than in our cohort [8].

Morbidity, frailty, and mortality

It has been reported that comorbid long-term health conditions (LTC) and especially multimorbidity, defined as the presence of one more LTC in addition to an index condition, are frequent in AF patients when compared to

Figure 2



Kaplan-Meier curves of the prediction all cause death during follow-up.

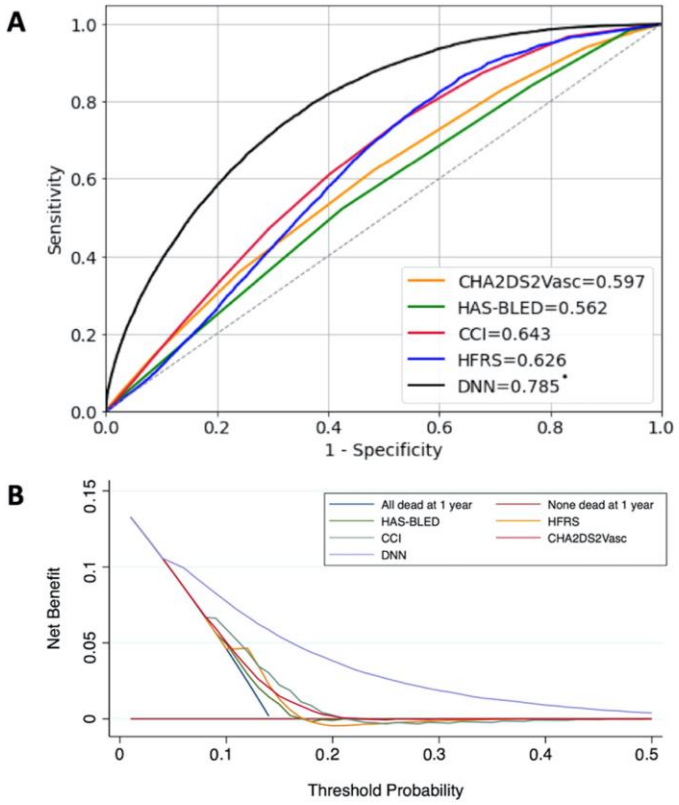
Table 3. Logistic regression analysis for clinical risk scores and occurrence of death within 1 year

	Univariate analysis		Multivariable analysis*	
	OR (95% CI)	*P*	OR (95% CI)	*P*
All-cause death				
DNN (for each decile)	1.52 (1.50-1.53)	<.0001	1.56 (1.54-1.58)	<.0001
DNN ≥ 0.16 (vs <0.16)	6.09 (5.84-6.35)	<.0001	4.00 (3.78-4.22)	<.0001
CCI (for each point)	1.16 (1.15-1.17)	<.0001	1.15 (1.14-1.16)	<.0001
CCI ≥4 (vs <4)	2.33 (2.24-2.42)	<.0001	1.88 (1.79-1.99)	<.0001
HFRS (for each point)	1.03 (1.03-1.03)	<.0001	1.00 (1.00-1.01)	.001
HFRS ≥15 (vs <15)	1.50 (1.44-1.56)	<.0001	0.94 (0.90-0.99)	.015
Non-CV death				
DNN (for each decile)	1.53 (1.51-1.55)	<.0001	1.57 (1.55-1.59)	<.0001
DNN ≥ 0.16 (vs <0.16)	6.23 (5.92-6.55)	<.0001	4.50 (4.21-4.80)	<.0001
CCI (for each point)	1.17 (1.16-1.18)	<.0001	1.19 (1.18-1.21)	<.0001
CCI ≥4 (vs <4)	2.49 (2.38-2.61)	<.0001	2.22 (2.09-2.35)	<.0001
HFRS (for each point)	1.03 (1.03-1.04)	<.0001	1.01 (1.01-1.02)	<.0001
HFRS ≥15 (vs <15)	1.70 (1.62-1.78)	<.0001	1.12 (1.06-1.18)	<.0001
CV death				
DNN (for each decile)	1.35 (1.34-1.37)	<.0001	1.34 (1.31-1.37)	<.0001
DNN ≥ 0.16 (vs <0.16)	3.99 (3.72-4.27)	<.0001	2.26 (2.05-2.47)	<.0001
CCI (for each point)	1.09 (1.08-1.10)	<.0001	0.99 (0 .97-1.01)	.412
CCI ≥4 (vs <4)	1.72 (1.61-1.83)	<.0001	1.10 (1.01-1.19)	.025
HFRS (for each point)	1.01 (1.01-1.02)	<.0001	0 .98 (0.98-0.99)	<.0001
HFRS ≥15 (vs <15)	1.03 (0.95-1.11)	.453	0.68 (0.62-0.74)	<.0001

CCI, Charlson Comorbidity Index; *CI*, confidence interval; *CV*, cardiovascular; *DNN*, deep neural network; *HFRS*, Hospital Frailty Risk Score; *OR*, odds ratio.
* Adjusted on all clinical characteristics displayed Table 1.

non-AF patients and is associated with mortality in AF patients.[29-33] This additional risk is added to AF itself which is already an independent risk factor for all-cause mortality.[30] In the same way, the high mean CHA_2DS_2-VASc score in our cohort reflects CV comorbidities and high CCI reflects non-CV comorbidities. These scores were even higher in AF patients who died at 1 year when compared to patients still alive. Both scores were independently associated to 1-year mortality underlying the burden of multimorbidity.

Figure 3



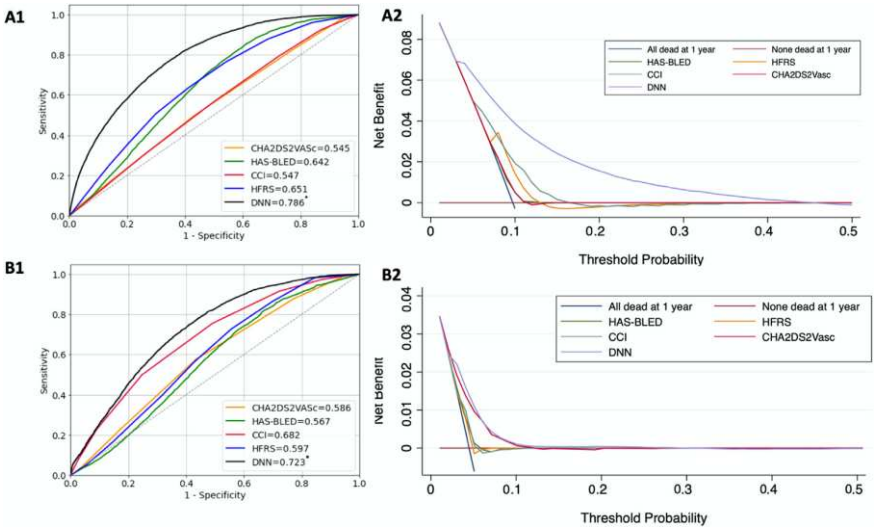
Panel A, Receiver operating characteristic curves of occurrence of all cause death during the year following atrial fibrillation diagnosis. **Panel B,** Decision curve analyses for the deep neural network algorithm and CHA_2DS_2-VASc score, HAS-BLED score, Charlson Comorbidity Index and Hospital Frailty Risk Score. * DNN (deep neural network) $P < .0001$ vs CHA_2DS_2VASc score, HAS-BLED score, CCI (Charlson Comorbidity Index) and HFRS (Hospital Frailty Risk Score).

Frailty is another concept developed recently. While multimorbidity and frailty represent 2 different ways of looking at the complexity of older persons and share some aspects, frailty differs from multimorbidity by integrating the concept of functional loss.[34,35] Moreover, multimorbidity is monodimensional and grounds its roots in the somewhat inadequate framework of "disease," whereas frailty implies a more exhaustive and comprehensive assessment of the individual, facilitating the implementation of multidimensional and tailored interventions. Indeed, AF patients are frequently frail with an estimated prevalence of 40% reported across the literature, and frailty is also strongly associated to mortality.[36] In our cohort, patients had an intermediate risk profile, with a mean ± SD HFRS of 9.3 ± 9.2 which increased to 11.69 ± 8.57 for patients who died at 1 year and HFRS was independently associated with death at 1 years.

Variable selection

Age plays a pivotal role in our algorithm and was the most important variable selected by Random Forest. More surprisingly, only 1 CV comorbidity (congestive heart failure) was included in our model. All other selected variables were non-CV underlying the burden of non-CV multimorbidity and frailty on survival. To some extent, this is consistent with previous observations. For

Figure 4



Panels A1 and **A2,** Receiver operating characteristic curves and Decision curve analyses for non-cardiovascular death within the year following AF diagnosis. **Panels B1 and B2,** Receiver operating characteristic curves and Decision curve analyses for cardiovascular death within the year following AF diagnosis. * DNN: deep neural network, $P < .0001$ vs CHA_2DS_2VASc score, HAS-BLED score, CCI (Charlson Comorbidity Index) and HFRS (Hospital Frailty Risk Score).

example, Singh et al reported that heart failure contributed to 3.8% of all deaths and was the second CV cause of death after ischemic heart disease, being responsible for 16% of all deaths.[9] Cancer and respiratory failure were the most common non-CV cause of death accounting for 30% and 10% respectively. Interestingly, ischemic heart disease was not retained by our model, and stroke deaths were infrequent (2.7%). Of note, the GARFIELD-AF and ROCKET-AF studies have previously shown that heart failure and sudden cardiac death are the major reasons for death of AF patients taking oral anticoagulant medication.[37,7]

Other selected variables related to non-CV conditions such as cancer, renal disease and lung disease are strongly associated with mortality in AF patients.[30,33] Once again, these conditions are found out with the common thromboembolic risk scores.

Another strength of random forest feature selection is the identification of variable un- or poorly described such as difficulty in walking, blood transfusion or urinary tract infection. Difficulty in walking or impaired mobility are not intuitively risk factors of death, but when considered as surrogates of falls, this is consistent with the REGARDS study which have shown association between the risk of falls and mortality particularly in AF population.[38]

Risk stratification of mortality

AF guidelines have mainly focused on identifying patients with different risks of stroke and major bleeding.[39–41] Prior prognostic models for stroke and bleeding in AF showed only modest predictive performance when evaluated to predict the outcome of death.[14] On the other hand, scores such as CCI and HFRS assessing mortality through multimorbidity and frailty have been developed from hospitalized patients but not especially on AF patient cohorts.[15,16] Additionally, HFRS was not fitted on patients under 75 years. Hence, assessing mortality remain a challenge, especially in patients with AF.

Few recent studies aimed to propose new dedicated scores developed with ML methods to predict the occurrence of all-cause death in AF patients.[42–45] Reported predictive ability ranged from 0.77 to 0.85 (c-index) which is comparable to our model. However, the GARFIELD-AF and BASIC-AF risk scores need to integrate biological and echocardiography variables unlike our only clinical factor based ML model.[42,43] Additionally, Lasso-Cox model and ABC risk score use variables not routinely available such as GDF-15, monoamine oxidase, cholinesterase, blood urea nitrogen in the clinical setting that might decrease their practicality in everyday clinical practice.[44,45] Moreover, clinical variables used as input for these models

were only CV items except cancer for Lasso-Cox model and dementia for GARFIELD-AF, and biomarker are non-specific, likely reflecting a sick patient or a sick heart.[46]

Yet we know that non-CV deaths represent the majority of mortality and that multimorbidity and frailty are strong predictors of mortality in AF patients, reflecting the clinical complexity of such patients.[47] Thus, including non-CV conditions multimorbidity and frailty to our model could explain the good performance of our model, similar to the others, even without biological or echography parameters. Unsurprisingly, the management of concurrent comorbidities (not limited to CV ones) is a pillar of the integrated and holistic management of AF patients, and how such a management approach can improve prognosis in clinically complex patients, characterized—among others—by frailty and multimorbidity.[48,49] Of note, our model was built using exhaustive real-life data at the country level, unlike other models derived from a single-center cohort or randomized control trial that might induce selection bias.[43–45] This would confer a good reproducibility of our results when applied for external validation but need to be confirmed in other populations (eg, North America, Asia).

Potential clinical practical implications and perspectives

Health care systems increasingly adopt electronic health records (EHR) for the management of patients. As AF is frequently associated with multimorbidity or frailty which are strong predictors of mortality, this obviates the current need for separate stroke and bleeding risk scores to aid mortality calculations for each individual. Indeed, the predictable aspect of survival outcomes in AF patients appear virtually all non-AF rhythm related and many are non-cardiac. More integrative and dedicated tools might allow the clinician to evaluate all these risks separately when deciding on whether to recommend management strategies. Indeed, our DNN model has the potential to be incorporated into online calculators, mobile applications, or routine electronic record systems and to automatically predict mortality (total, non-CV and CV) and would allow users to base treatment decisions on more precise measures of risk. Alerts can be implemented in case of a 'high risk' score to flag up those patients at high risk of early death.

Populations and medical practices change very rapidly and models are quickly outdated. This model has also the possibility to be implemented in nationwide administrative datasets and trained in real time to provide up to date versions of the model in order to fit perfectly the data over time. Moreover, using ICD-10 codes internationally enable comparisons and validation to be carried out in different populations across the world. Future attempts to modify the prognosis of AF patients in clinical trials may ultimately need to select out lower risk subsets using modeling approaches (eg, using ML), where the effects of unmodifiable relationships are much smaller and there is a better causal pathway between AF and AF-related adverse outcomes.

Limitations

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 were possibly underestimated. Echocardiographic, biological and imaging parameters were also lacking.

Conclusion

ML algorithms predict mortality in AF patients with a better ability than previously developed traditional clinical risk scores. A ML approach may help clinicians to better stratify AF patients at high risk of total, non-CV and CV mortality and may assist physicians in decision-making when managing AF patients in a holistic and integrated care manner.

Disclosures

Bisson has been a consultant or speaker for AstraZeneca, Bayer, BMS/Pfizer, Medtronic, Vitorpharma and Alnylam. Angoulvant has been a consultant or speaker for Amgen, Astra-Zeneca, Bayer, BMS/Pfizer, MSD, Novartis, Novo Nordisk, Sanofi, Servier. Lip has been a consultant or speaker for BMS/Pfizer, Boehringer Ingelheim, Daiichi-Sankyo, Anthem. No fees are received personally. He is co-principal investigator of the AFFIRMO project on multimorbidity in AF, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 899871. LF reports consulting fees for AstraZeneca, Bayer, BMS/Pfizer, Boehringer Ingelheim, Medtronic, Novo Nordisk and Novartis and lecture fees for AstraZeneca, Bayer, BMS/Pfizer, Boehringer Ingelheim and Zoll. All other authors have reported that they have no relationships relevant to the contents of this paper to disclose.

Data availability statement

Access to the PMSI (the French hospitalization database) is controlled by the CNIL, the independent national ethics committee protecting human rights in France. Due to the sensitive nature of the database, data sharing is not authorized according to the French legislation.

Funding

None.

Acknowledgments

None.

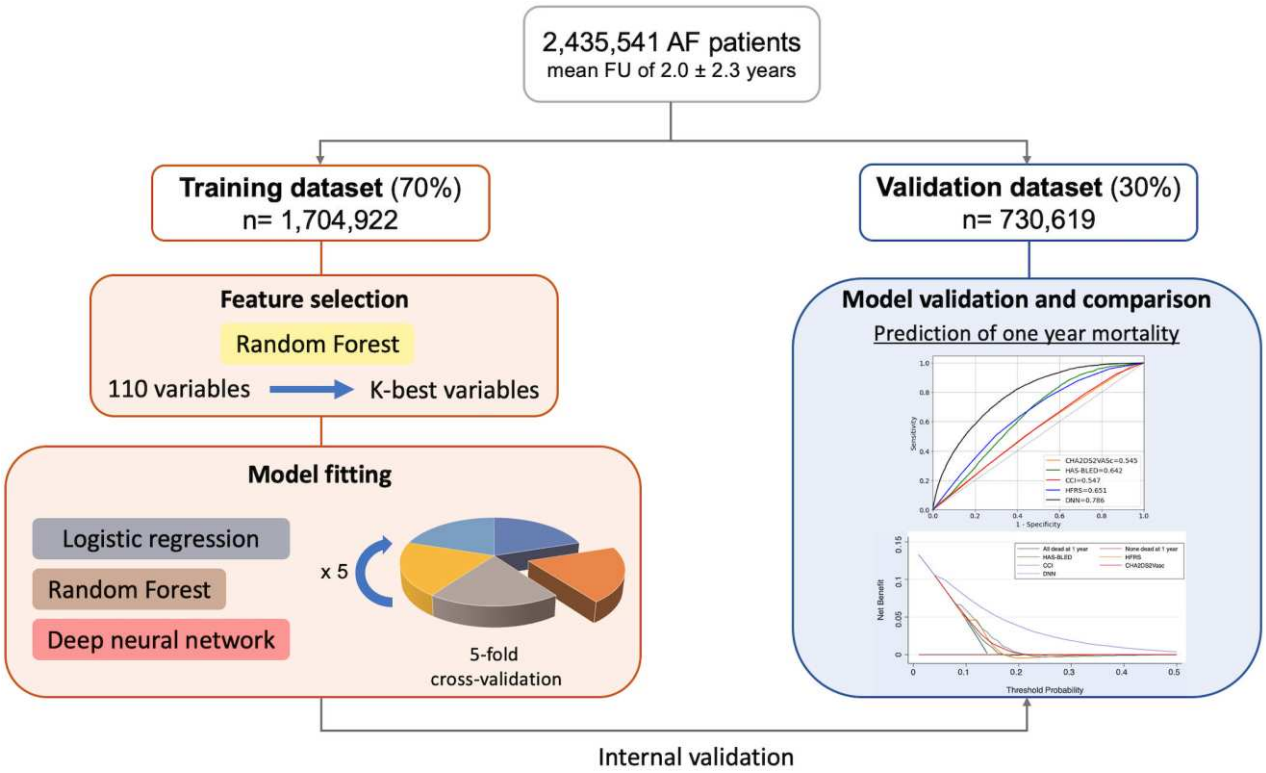
Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ahj.2023.08.006.

References

1. Benjamin EJ, Muntner P, Alonso A, et al. Heart disease and stroke statistics-2019 update: a report From the American Heart Association. Circulation 2019;139:e56–528.
2. Tsao CW, Aday AW, Almarzooq ZI, et al. Heart disease and stroke statistics-2022 update: a report from the American Heart Association. Circulation 2022;145:e153–639.
3. Lip GYH, Nieuwlaat R, Pisters R, et al. 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 2010;137:263–72.
4. Pisters R, Lane DA, Nieuwlaat R, et al. A novel user-friendly score (HAS-BLED) to assess 1-year risk of major bleeding in patients with atrial fibrillation: the Euro Heart Survey. Chest 2010;138:1093–100.
5. van Walraven C, Hart RG, Singer DE, et al. Oral anticoagulants vs aspirin in nonvalvular atrial fibrillation: an individual patient meta-analysis. JAMA 2002;288:2441–8.
6. Liew A, O'Donnell M, Douketis J. Comparing mortality in patients with atrial fibrillation who are receiving a direct-acting oral anticoagulant or warfarin: a meta-analysis of randomized trials. J Thromb Haemost 2014;12:1419–24.
7. Bassand JP, Accetta G, Camm AJ, et al. Two-year outcomes of patients with newly diagnosed atrial fibrillation: results from GARFIELD-AF. Eur Heart J 2016;37:2882–9.
8. Piccini JP, Hammill BG, Sinner MF, et al. Clinical course of atrial fibrillation in older adults: the importance of cardiovascular events beyond stroke. Eur Heart J 2014;35:250–6.
9. Singh SM, Abdel-Qadir H, Pang A, et al. Population trends in all-cause mortality and cause specific–death with incident atrial fibrillation. J Am Heart Assoc 2020;9:e016810.
10. Dai H, Zhang Q, Much AA, et al. Global, regional, and national prevalence, incidence, mortality, and risk factors for atrial fibrillation, 1990-2017: results from the Global Burden of Disease Study 2017. Eur Heart J Qual Care Clin Outcomes 2021;7:574–82.
11. Marijon E, Le Heuzey J-Y, Connolly S, et al. Causes of death and influencing factors in patients with atrial fibrillation. Circulation 2013;128:2192–201.
12. Fauchier L, Villejoubert O, Clementy N, et al. Causes of death and influencing factors in patients with atrial fibrillation. Am J Med 2016;129:1278–87.
13. Ding WY, Lane DA, Gupta D, et al. Incidence and risk factors for residual adverse events despite anticoagulation in atrial fibrillation: results from phase II/III of the GLORIA-AF registry. J Am Heart Assoc 2022;11:e026410.
14. Proietti M, Farcomeni A, Romiti GF, et al. Association between clinical risk scores and mortality in atrial fibrillation: systematic review and network meta-regression of 669,000 patients. Eur J Prev Cardiol 2020;27:633–44.
15. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40:373–383.
16. Gilbert T, Neuburger J, Kraindler J, et al. Development and validation of a Hospital Frailty Risk Score focusing on older people in acute care settings using electronic hospital records: an observational study. Lancet 2018;391:1775–82.
17. Olier I, Ortega-Martorell S, Pieroni M, Lip GYH. How machine learning is impacting research in atrial fibrillation: implications for risk prediction and future management. Cardiovasc Res 2021;117:1700–17.
18. Lip GYH, Genaidy A, Tran G, et al. improving stroke risk prediction in the general population: a comparative assessment of common clinical rules, a new multimorbid index, and machine-learning-based algorithms. Thromb Haemost 2022;122:142–50.
19. Zheng X, Wang F, Zhang J, et al. Using machine learning to predict atrial fibrillation diagnosed after ischemic stroke. Int J Cardiol 2022;347:21–7.
20. Mark E, Goldsman D, Gurbaxani B, et al. Using machine learning and an ensemble of methods to predict kidney transplant survival. PLoS One 2019;14:e0209068.
21. Bisson A, Bodin A, Clementy N, et al. Prediction of incident atrial fibrillation according to gender in patients with ischemic stroke from a nationwide cohort. Am J Cardiol 2018;121:437–44.
22. Li Y-G, Bisson A, Bodin A, et al. C2HEST score and prediction of incident atrial fibrillation in poststroke patients: a French nationwide study. J Am Heart Assoc 2019;8:e012546.
23. Couris CM, Polazzi S, Olive F, et al. Breast cancer incidence using administrative data: correction with sensitivity and specificity. J Clin Epidemiol 2009;62:660–6.
24. Bisson A, Lemrini Y, El-Bouri W, et al. Prediction of incident atrial fibrillation in post-stroke patients using machine learning: a French nationwide study. Clin Res Cardiol 2023;112:815–23.
25. Gilbert T, Cordier Q, Polazzi S, et al. External validation of the Hospital Frailty Risk Score in France. Age Ageing 2022;51:afab126.
26. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44:837–45.
27. Pencina MJ, D'Agostino Sr RB, D'Agostino Jr RB, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat Med 2008;27:157–72 discussion 207-112.
28. 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.
29. Radner H, Yoshida K, Smolen JS, Solomon DH. Multimorbidity and rheumatic conditions-enhancing the concept of comorbidity. Nat Rev Rheumatol 2014;10:252–6.
30. Andersson T, Magnuson A, Bryngelsson I-L, et al. All-cause mortality in 272,186 patients hospitalized with incident atrial fibrillation 1995-2008: a Swedish nationwide long-term case-control study. Eur Heart J 2013;34:1061–7.
31. Chamberlain AM, Alonso A, Gersh BJ, et al. Multimorbidity and the risk of hospitalization and death in atrial fibrillation: a population-based study. Am Heart J 2017;185:74–84.

32. Vanbeselaere V, Truyers C, Elli S, et al. Association between atrial fibrillation, anticoagulation, risk of cerebrovascular events and multimorbidity in general practice: a registry-based study. BMC Cardiovasc Disord 2016;16:61.
33. Jani BD, Nicholl BI, McQueenie R, et al. Multimorbidity and co-morbidity in atrial fibrillation and effects on survival: findings from UK Biobank cohort. Europace 2018;20(FI 3):f329–36.
34. Cesari M, Pérez-Zepeda MU, Marzetti E. Frailty and multimorbidity: different ways of thinking about geriatrics. J Am Med Dir Assoc 2017;18:361–4.
35. Fried LP, Tangen CM, Walston J, et al. Frailty in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci 2001;56:M146–56.
36. Proietti M, Romiti GF, Raparelli V, et al. Frailty prevalence and impact on outcomes in patients with atrial fibrillation: a systematic review and meta-analysis of 1,187,000 patients. Ageing Res Rev 2022;79:101652.
37. Pokorney SD, Piccini JP, Stevens SR, et al. Cause of death and predictors of all-cause mortality in anticoagulated patients with nonvalvular atrial fibrillation: data from ROCKET AF. J Am Heart Assoc 2016;5:e002197.
38. O'Neal WT, Qureshi WT, Judd SE, et al. Effect of falls on frequency of atrial fibrillation and mortality risk (from the REasons for Geographic And Racial Differences in Stroke Study). Am J Cardiol 2015;116:1213–18.
39. Hindricks G, Potpara T, Dagres N, et al. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): 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. Eur Heart J 2021;42:373–498.
40. Chao T-F, Joung B, Takahashi Y, et al. 2021 Focused Update Consensus Guidelines of the Asia Pacific Heart Rhythm Society on stroke prevention in atrial fibrillation: executive summary. Thromb Haemost 2022;122:20–47.
41. Gorog DA, Gue YX, Chao T-F, et al. Assessment and mitigation of bleeding risk in atrial fibrillation and venous thromboembolism: executive summary of a European and Asia-Pacific Expert Consensus Paper. Thromb Haemost 2022;122:1625–52.
42. Fox KAA, Lucas JE, Pieper KS, et al. Improved risk stratification of patients with atrial fibrillation: an integrated GARFIELD-AF tool for the prediction of mortality, stroke and bleed in patients with and without anticoagulation. BMJ Open 2017;7:e017157.
43. Samaras A, Kartas A, Akrivos E, et al. A novel prognostic tool to predict mortality in patients with atrial fibrillation: the BASIC-AF risk score. Hellenic J Cardiol 2021;62:339–48.
44. Hijazi Z, Oldgren J, Lindbäck J, et al. A biomarker-based risk score to predict death in patients with atrial fibrillation: the ABC (age, biomarkers, clinical history) death risk score. Eur Heart J 2018;39:477–85.
45. Chen Y, Wu S, Ye J, et al. Predicting all-cause mortality risk in atrial fibrillation patients: a novel LASSO-Cox model generated from a prospective dataset. Front Cardiovasc Med 2021;8:730453.
46. Camelo-Castillo A, Rivera-Caravaca JM, Marín F, et al. Predicting adverse events beyond stroke and bleeding with the ABC-Stroke and ABC-bleeding scores in patients with atrial fibrillation: the Murcia AF project. Thromb Haemost 2020;120:1200–7.
47. Romiti GF, Proietti M, Bonini N, et al. Clinical complexity domains, anticoagulation, and outcomes in patients with atrial fibrillation: a report from the GLORIA-AF Registry Phase II and III. Thromb Haemost 2022;122:2030–41.
48. Lip GYH. The ABC pathway: an integrated approach to improve AF management. Nat Rev Cardiol 2017;14:627–8.
49. Romiti GF, Proietti M, Vitolo M, et al. Clinical complexity and impact of the ABC (Atrial fibrillation Better Care) pathway in patients with atrial fibrillation: a report from the ESC-EHRA EURObservational Research Programme in AF General Long-Term Registry. BMC Med 2022;20:326.



Graphical abstract.

1.3 Étude 3 : Modèle de machine learning de type multi-label gradient-boosting decision tree en chaine pour la prédiction de risques multiples chez les patients atteints de fibrillation atriale : Résultats issus de la phase II/III du registre GLORIA-AF

Lu J, ***Bisson A****, Bennamoun M, Zheng Y, Sanfilippo FM, Hung J, Briffa T, McQuillan B, Stewart J, Figtree G, Huisman MV, Dwivedi G, Lip GYH.* ***Multi-label gradient-boosting machine learning model to predict multiple risks in patients with atrial fibrillation: results from Phase II/III of the GLORIA-AF registry.*** *Under review.*

1. Objectif de l'étude

L'objectif de ce troisième projet était de montrer l'applicabilité et l'utilité d'un modèle de ML de type multi-label gradient boosting decision tree (ML-GBDT) en chaine à partir d'une cohorte internationale de patients atteints de FA.

2. Méthode

C'est à partir du registre prospectif international GLORIA-AF que nous avons optimisé notre algorithme pour la prédiction de la survenue d'événements indésirables comme le décès, l'AVCi et les saignements majeurs durant l'année suivant le diagnostic de FA. Nous avons ensuite comparé notre modèle aux scores cliniques déjà existants et exploré ses performances au sein de différentes sous populations.

3. Résultats

Sur les 25,656 patients atteints de FA inclus, 215 patients (0,8%) ont été victime d'AVCi, 405 patients (1,6 %) d'hémorragie majeure et 897 patients (3,5%) sont décédés au cours de l'année suivant le diagnostic de FA. Notre modèle a obtenu une AUC de 0,785 (IC 95% : 0,757-0,813) pour la prédiction du risque de décès, de 0,691 (IC 95% : 0,626-0. 756) pour la prédiction sur risque d'AVCi et 0,698 (IC 95% : 0,651-0,745) pour la prédiction du risque de saignements majeurs. Notre modèle était supérieur à l'Indice de Comorbidité de Charlson (0,747, p=0,007) pour la prédiction du décès, supérieur au score CHA_2D_2S-VASc (0,613, p=0,028) pour la prédiction de l'AVCi et supérieur au score HAS-BLED (0,607, p=0,002) pour les saignements majeurs, avec une amélioration de l'indice net de reclassification (10,0 %, 12,5 % et 23,6 % respectivement). Les performances prédictives ainsi que l'ordre d'importance des variables variaient entre les différents sous-groupes.

4. Conclusion

Notre étude valide le fait que notre modèle de ML de type ML-GBDT en chaine surpasse les scores de risque clinique pour prédire le risque de décès toutes causes confondues, d'AVCi et d'hémorragie majeure chez les patients souffrant de FA. Des variations sont observables selon les sous populations. Cette approche pourrait être utilisée comme un outil holistique pour optimiser la compréhension et l'évaluation des risques d'événements indésirables et atténuer leur impact dans diverses sous populations de patients souffrant de FA.

Multi-label gradient-boosting machine learning model to predict multiple risks in patients with atrial fibrillation: results from Phase II/III of the GLORIA-AF registry

Juan Lu[1,2,3,4]*, Arnaud Bisson[1,5]*, Mohammed Bennamoun[3], Yalin Zheng[1,6], Frank M Sanfilippo[7], Joseph Hung[2], Tom Briffa[7], Brendan McQuillan[2,8], Jonathon Stewart[2], Gemma Figtree[9,10], Menno V Huisman[11], Girish Dwivedi[2,3,12]§, Gregory YH Lip[1,13]§

[1] Liverpool Centre for Cardiovascular Science at University of Liverpool, Liverpool John Moores University and Liverpool Heart & Chest Hospital, Liverpool, United Kingdom;
[2] Medical School, The University of Western Australia, Perth, Australia;
[3] Harry Perkins Institute of Medical Research, Perth, Australia;
[4] Department of Computer Science and Software Engineering, The University of Western Australia, Perth, Australia;
[5] Department of Cardiology, University Hospital and University of Tours, Tours, France;
[6] Department of Eye and Vision Sciences, University of Liverpool, Liverpool, United Kingdom;
[7] School of Population and Global Health, The University of Western Australia, Perth, Australia;
[8] Sir Charles Gairdner Hospital, Perth, Australia;
[9] Kolling Institute and Charles Perkins Centre, University of Sydney, Sydney, Australia;
[10] Department of Cardiology, Royal North Shore Hospital, Sydney, Australia;
[11] Department of Thrombosis and Hemostasis Leiden University Medical Center Leiden the Netherlands;
[12] Department of Cardiology, Fiona Stanley Hospital, Perth, Australia;
[13] Danish Center for Health Services Research, Department of Clinical Medicine, Aalborg University, Aalborg, Denmark

*Joint first authors
§ Joint senior authors
Drs Lip and Huisman are co-Chairs of the GLORIA-AF registry

Correspondence:
Prof Girish Dwivedi, MD, FRCPC, FACC, FASNC, FSCCT
Wesfarmers Chair in Cardiology, Harry Perkins Institute of Medical Research, The University of Western Australia
Consultant Cardiologist at Fiona Stanley Hospital, Murdoch, 6009, WA, Australia
Adjunct Professor of Cardiology, University of Ottawa, Canada
Adjunct Research Professor-School of Biomedical Sciences at Curtin University, WA, Australia
Email: girish.dwivedi@perkins.uwa.edu.au
T: +61 8 6488 8181

Prof Gregory Y H Lip, MD, FRCP, DGM, FACC, FESC, FEHRA
Liverpool Centre for Cardiovascular Science, University of Liverpool and Liverpool Heart & Chest Hospital, Liverpool, UK
Danish Center for Health Services Research, Department of Clinical Medicine, Aalborg University, Aalborg, Denmark
Email: gregory.lip@liverpool.ac.uk

Abstract

Background Patients with atrial fibrillation (AF) have a higher risk of ischaemic stroke and death. While anticoagulants are effective at reducing these risks, they increase the risk of major bleeding. Simple clinical risk stratification scores are used to assess risks and determine the appropriate treatment plan in AF. However, these risk scores only perform modestly in predicting adverse outcomes, especially for the outcome of death. We aimed to test the multi-label gradient boosting decision tree (ML-GBDT) model in predicting risks for death, stroke, and major bleeding in a prospective global AF registry.

Methods We studied patients from phase II/III of the GLORIA-AF registry. The outcomes were all-cause death, ischaemic stroke and major bleeding within one year following the diagnosis of AF. We trained the ML-GBDT model, which captures dependencies between input features and multiple output labels. We compared the discrimination of the ML-GBDT model and the clinical scores in predicting patient outcomes.

Findings A total of 25,656 patients were included (mean age 70.3 years (SD 10.3); 44.8% female). These patients were predominantly anticoagulated (86.0%). Within one-year after AF diagnosis, ischaemic stroke occurred in 215 (0.8%), major bleeding in 405 (1.6%), and death in 897 (3.5%) patients. Our model achieved an optimized area under the curve (AUC) in predicting risks of death (0.785, 95% CI: 0.757-0.813) compared to Charlson Comorbidity Index (0.747, p=0.007), ischaemic stroke (0.691, 0.626-0.756) comparing to CHA_2D_2S-VASc score (0.613, p=0.028), and major bleeding (0.698, 0.651-0.745) as opposed to HAS-BLED score (0.607, p=0.002), with improvement in net reclassification index (10.0%, 12.5% and 23.6% respectively).

Interpretation The ML-GBDT model outperformed clinical risk scores in predicting the risks in patients with AF. This approach could be used as a single multifaceted holistic tool to optimize patient risk assessment and mitigate adverse outcomes when managing AF.

Funding None
Abstract word count: 296

Research in Context

Evidence before this study

We did not do a formal systematic review as there is already established literature indicating patients with AF have a significantly elevated risk of stroke and mortality. While anticoagulation therapy reduces these risks, it increases the risk of bleeding. Most efforts have focused on using clinical risk scores in predicting risks for patients with AF, such as CHA2DS2-VASc, HAS-BELD, and the Charlson comorbidity index. While the Charlson comorbidity index is commonly used to predict all-cause mortality in general patients, and the CHA2DS2-VASc, HAS-BELD scores offer only modest performance in predicting ischaemic stroke and major bleeding in patients with AF. Machine learning models have been used but most of them have treated different outcomes independently. Incidentally, they also been instrumental in uncovering previously unidentified risk variables associated with adverse outcomes. We have previously shown that the deployment of multi-label optimized the risk prediction performance of machine learning models on an Australian retrospective cohort. This

serves as proof of concept that incorporating relationships between patient outcomes can enhance risk prediction.

Added value of this study

This is the first study to our knowledge to use the multi-label gradient boosting decision tree to predict the multifaceted risks for patients diagnosed with atrial fibrillation in a prospective registry. We used a large global prospective AF registry and identified ischaemic stroke, major bleeding, and all-cause death within the first year after diagnosis. The model was trained and internally validated on 70% of the data and tested in the remaining 30% of the data. The model incorporated patients' comorbidities and treatment information and also took into account the relationships between different outputs to predict patients' risks. We found that the multilabel machine learning model achieved a statistically superior performance compared to clinical risk scores by leveraging the relationships between different outcomes, as well as taking comorbidities and treatments into account. Importantly, the predictive performance varied among different patients' sub-groups featured by geographic regions, treatment and history of stroke. Lastly, we ranked the feature importance for the patient subsets in GLORIA-AF based on geographic regions, receipt of anticoagulation therapy, and history of stroke.

Implications of all the available evidence

Our study underscores the value of using multi-label machine learning models to predict multiple risks associated with patients diagnosed with AF. Our model benefits from including additional comorbidities and treatment information, and from incorporating the relationships between the different outcomes. The machine learning model also holds the potential for automation within the electronic health records system. Furthermore, the feature importance ranking on different sub-cohorts in the model could help elucidate the varied performance in predicting outcomes.

Introduction

Atrial fibrillation (AF) is the most common cardiac arrhythmia, increasingly affecting over 45 million people globally. Patients diagnosed with AF have significantly elevated risks of ischaemic stroke and death.[1] Oral anticoagulation therapy is used to mitigate the risks of ischaemic stroke and all-cause death, but they increase the risk of major bleeding.[2] It is also well recognized that patients may still suffer from residual risks of death and ischaemic stroke despite receiving anticoagulants[3].

The CHA_2DS_2-VASc and HAS-BLED scores are recommended by clinical guidelines for evaluating stroke and bleeding risks, respectively.[4,5] However, these scores have only modest discriminatory capacity for ischaemic stroke and major bleeding in patients with AF.[6,7] Apart from the risk factors employed in these simple clinical risk scores, ethnicity or geographic regions have been identified as potential additional factors that can influence the prognosis for patients after being diagnosed with AF.[8] For example, studies have shown that the modified risk stratification tool $mCHA_2DS_2$-VASc tend to provide more accurate risk predictions for Asian patients with AF.[9] Risk factors for stroke or bleeding are also risks for death, although the Charlson Comorbidity Index (CCI) is an established tool applied to predict 1-year death based on specific comorbidities.[10]

Machine learning algorithms have shown better risk predictive performance, after accounting for the presence of multimorbid conditions and dynamic changes in risk.[11] Additionally, multi-label machine learning can predict different patient outcomes simultaneously (e.g. stroke, major bleeding, and death) taking into account the interdependence between risks, therefore offering a more effective and multifaceted approach to predicting event risks faced by patients with AF. This would allow us to predict death, as well as ischaemic stroke and major bleeding risks for patients with AF by incorporating their diverse comorbidities, existing treatment strategies, and demographics including ethnicity and geographical region. This offers valuable insights for early intervention on the most significant risks, leading to improved clinical outcomes.

In our previous study in an Australian AF registry, we constructed and assessed different multi-label machine learning algorithms, among which the multi-label gradient boosting decision tree (ML-GBDT) model demonstrated superior performance for predicting stroke and bleeding risk in patients with non-valvular AF compared with clinical risk scores (CHA_2DS_2-VASc and HAS-BLED).[12] One major limitation of this previous study was the retrospective nature of the data set which was sourced exclusively from one geographical region with the predominant ethnic group being Caucasians. Moreover, non-vitamin K antagonist oral anticoagulants (NOAC) had not yet reached widespread use in our previous study. An external validation of our approach in global prospectively collected AF cohorts would be highly desirable.

Herein, we aimed to investigate our prior ML-GBDT algorithms to predict death, ischaemic stroke and major bleeding within the first year after the baseline visit for patients with newly diagnosed non-valvular AF (Figure 1). We used a large, contemporary global cohort of patients from the prospective GLORIA-AF (Global Registry on Long-Term Oral Anti-Thrombotic Treatment in Patients with Atrial Fibrillation) registry.[13] Our goal was to assess the applicability and utility of multi-label machine learning across diverse AF patient populations and determine its potential for widespread clinical adoption.

Methods
We used data from phase II/III of the prospective GLORIA-AF registry, which is a prospective international registry that gathered information about patients from 935 medical centres in 38 countries across Asia, Europe, North America, Latin America, and Africa/Middle East.[13] The registry comprised adults with newly diagnosed non-valvular AF (within 3 months prior to baseline visit) and an increased risk of stroke (CHA_2DS_2-VASc score ≥ 1).

We included patients from the GLORIA-AF registry enrolled between 2011 and 2020 who had at least one year of follow-up or died during the first year. We formed subsets based on age (age 65 years or more), anticoagulant treatment, residential region (Europe, North America, and Asia), and whether patients had primary prevention or secondary prevention of ischaemic stroke. This enabled an in-depth exploration into the performance of different risk prediction models within the subcohorts sharing certain characteristics, and thereby providing nuanced insights into their predictive capacities, identifying factors that may impact the risk prediction in patients with AF.

AF classification was as per to the European Society of Cardiology recommendations.[14] The severity of AF-related symptoms was determined based on the European Heart Rhythm Association classification. The CHA_2DS_2-VASc, HAS-BLED scores and CCI were calculated from data within the registry.

The study was governed by Good Clinical Practice and the Declaration of Helsinki. The protocol was approved by the European Medicines Agency and institutional review boards at each participating site. Patients provided written informed consent to be included in the GLORIA-AF registry.

Predictors
We included 110 variables (features) in the initial model, including patient demographics, vital signs, comorbidity history, medications, type of anticoagulant treatment, medical procedures, and laboratory results. The complete list of variables is available in Supplementary table S1.

We did not include variables with more than 25% missing values (n = 16) and imputed the remaining missing values using the multivariate feature imputation method. Categorical variables were presented as counts and percentage and tested for differences with the chi-squared test. Continuous variables are presented as mean ±standard deviation (SD) and tested for differences with the t-tests. A 2-sided $p < 0.05$ was considered statistically significant.

Sampling and processing
We split the data into three sets for different purposes. We randomly allocated 70% of the data for training and internal validation. 10% of the training data was used for internal validation and hyperparameter tuning. The remaining 30% of data were assigned for testing. Text data was encoded as binary or categorical variables. The continuous variables were standardized with a mean of 0 and a standard deviation of 1, ensuring that all continuous features are on a similar scale.

Outcome
Follow-up was 1 year from the baseline AF diagnosis. The primary outcome of interest was all-cause death, ischaemic stroke and major bleeding within the 1-year follow-up. Stroke was

defined as an acute onset of a localised neurological impairment which originated from a vascular source and was continuous for a minimum of 24 hours or led to death. Data from computed tomography scans, magnetic resonance imaging or autopsy were applied by the GLORIA-AF investigators to determine the presence of ischaemic stroke.[15] Major bleeding was classified by the GLORIA-AF investigators according to the definition from the International Society on Thrombosis and Hemostasis.[16]

Machine-Learning Model

A gradient boosting decision tree[19] based classifier chain[20], namely ML-GBDT, was the best performing algorithm (compared with support vector machine and multi-layer neural networks) in previous studies for predicting risks in AF patients.[13] The detailed model background is listed in the Supplementary file. The order of the classifier chain was established as major bleeding, ischaemic stroke and all-cause death based on our previous experiment results[12] and time-to-event for different outcomes. The predicted probability of major bleeding was used to predict ischaemic stroke, and both predictions were used to predict all-cause death. This effectively empowered the model to handle patients with multiple events. For comparison purposes, we also constructed single-label GBDT models to predict each outcome independently. Modelling and statistical analyses were performed using Python 3.9 and Stata (16.1, StataCorp LLC, USA).

Feature importance

We included all the relevant available variables in the initial model. We then used the permutation feature importance[19] and presented 25-30 top-ranked features using bar plots. The permutation feature importance was defined as the reduction in a model's performance metrics when the values of an individual feature were randomly shuffled. The feature importance ranking plot was presented as a horizontal bar chart, with each bar representing a variable in our data. The length of the bar indicated the level of impact on the ML-GBDT predictive performance.

Parameter tuning

Given the low event counts (significant skew) in the outcomes, we adopted a balanced class weights setting in the model during the training phase, thereby compensating for the uneven distribution. We optimised the ML-GBDT model using a Bayesian technique[20] with 5-fold cross-validation to identify the optimal hyperparameters from the single-output GBDT for each outcome, including number of estimators, learning rate, maximum leaves, minimum child weight, gamma, maximum bin, column sample by tree and subsample. We implemented the mean values for each hyperparameter in the ML-GBDT.

Development vs. Validation

Compared to our previous study, predictors were modified based on available variables[12]. We added medical treatments for AF as predictors but were unable to incorporate predictors such as haemoglobin and indigenous ethnicity, which were not recorded in the GLORIA-AF Registry. We utilized different data source to define outcomes. In the current study, we used clinical information. We also implemented different hyperparameter tuning strategy while doing the validation. The feature importance measurement strategy was updated to permutation features importance, which tends not to be biased on high cardinality features.

Performance measurement

Measures of prediction performance included the area under the receiver operating characteristic curve (AUC-ROC), sensitivity, specificity, and net reclassification index. A probability threshold of 0.5 was selected to determine the sensitivity and specificity for clinical risk scores. We then reserved the similar specificity for the ML-GBDT model while choosing the thresholds and calculated the sensitivity. A decision curve analysis was employed to show the net benefits of each model across a range of threshold probabilities, and Kaplan-Meier survival curves were presented to visualize the probability of an event occurring over the three-year follow-up time based on the predicted probability.

Table 1: Baseline characteristics of the study cohort identified from Phase II/III of the GLORIA-AF registry stratified by outcome.

	All-cause death			Ischaemic stroke			Major bleeding		
	No event (n = 24759)	Event (n = 897)	p-value	No event (n = 25441)	Event (n = 215)	p-value	No event (n = 25251)	Event (n = 405)	p-value
Demographics									
Age	70·1 ± 10·3	76·2 ± 9·2	<0·0001	70·2 ± 10·3	74·5 ± 9·2	<0·0001	70·2 ± 10·3	74·3 ± 8·9	<0·0001
Female	11114 (44·9)	369 (41·1)	0·03	11387 (44·8)	96 (44·7)	0·97	11299 (44·7)	184 (45·4)	0·78
BMI (kg/m2)	28·7 ± 6·2	27·1 ± 6·6	<0·0001	28·6 ± 6·3	27·2 ± 5·6	0·001	28·6 ± 6·3	28·6 ± 6·2	0·8
Region			<0·0001			0·066			<0·0001
Europe	12325 (49·8)	467 (52·1)		12684 (49·9)	108 (50·2)		12593 (49·9)	199 (49·1)	
North America	5593 (22·6)	201 (22·4)		5754 (22·6)	40 (18·6)		5651 (22·4)	143 (35·3)	
Asia	4661 (18·8)	108 (12·0)		4715 (18·5)	54 (25·1)		4726 (18·7)	43 (10·6)	
Latin America	1874 (7·6)	115 (12·8)		1977 (7·8)	12 (5·6)		1969 (7·8)	20 (4·9)	
Africa/Middle East	306 (1·2)	6 (0·7)		311 (1·2)	1 (0·5)		312 (1·24)	0 (0·0)	
Race			<0·0001			0·023			<0·0001
White	18840 (76·1)	716 (79·8)		19405 (76·3)	151 (70·2)		19221 (76·1)	335 (82·7)	
Asian	4333 (17·5)	112 (12·5)		4393 (17·3)	52 (24·2)		4399 (17·4)	46 (11·4)	
Black or Afro-Caribbean	424 (1·7)	22 (2·5)		440 (1·7)	6 (2·8)		431 (1·7)	15 (3·70)	
Arab or Middle East	324 (1·3)	5 (0·56)		329 (1·3)	0 (0·0)		327 (1·3)	2 (0·49)	
Other	838 (3·4)	42 (4·7)		874 (3·4)	6 (2·8)		873 (3·5)	7 (1·73	
Vital signs									
Heart rate (bpm)	80·0 ± 21·2	83·7 ± 20·6	<0·0001	80·1 ± 21·2	79·7 ± 18·7	0·77	80·1 ± 21·2	80·0 ± 20·3	0·96
Systolic BP (mmHg)	132·5 ± 18·5	128·4 ± 20·8	<0·0001	132·3 ± 18·6	135·7 ± 18·5	0·01	132·4 ± 18·6	130·9 ± 20·0	0·12
Diastolic BP (mmHg)	78·2 ± 11·9	74·6 ± 12·9	<0·0001	78·1 ± 11·9	77·3 ± 12·4	0·33	78·1 ± 11·9	75·7 ± 12·4	<0·0001
Scores									
CHA_2D_2S-VASc	3·15±1·5	4·1±1·6	<0·0001	3·18±1·5	3·95±1·7	<0·0001	3·17±1·5	3·83±1·5	<0·0001
HAS-BLED	1·32±0·9	1·7±0·9	<0·0001	1·33±0·9	1·68±0·9	<0·0001	1·33±0·9	1·69±0·9	<0·0001
CCI	3·84±1·9	4·53±1·8	<0·0001	3·84±1·9	4·53±1·8	<0·0001	3·83±1·9	4·86±2·2	<0·0001
Laboratory result									
Creatinine Clearance (mL/min)	82·2 ± 52·7	62·9 ± 32·6	<0·0001	81·6 ± 52·4	71·2 ± 32·7	0·004	81·7 ± 52·5	70·7 ± 30·5	<0·0001
Comorbidities									
Hypertension	18653 (75·3)	673 (75.0)	0·83	19149 (75·3)	177 (82·3)	0·02	19000 (75·2)	326 (80·5)	0·02
Hypercholesterolemia	9951 (40·2)	345 (38·5)	0·3	10198 (40·1)	98 (45·6)	0·1	10088 (40)	208 (51·4)	<0·0001
Diabetes mellitus	5670 (22·9)	266 (29·7)	<0·0001	5882 (23·1)	54 (25·1)	0·49	5822 (23·1)	114 (28·1)	0·02
Coronary artery disease	4628 (18·7)	269 (30)	<0·0001	4842 (19)	55 (25·6)	0·01	4792 (19)	105 (25·9)	0·0004
Congestive heart failure	5304 (21·4)	396 (44·1)	<0·0001	5653 (22·2)	47 (21·9)	0·9	5583 (22·1)	117 (28·9)	0·001

Left ventricular hypertrophy	4528 (18·3)	197 (22)	0·01	4683 (18·4)	42 (19·5)	0·67	4648 (18·4)	77 (19)	0·76
Prior thromboembolism	3615 (14·6)	205 (22·9)	<0·0001	3741 (14·7)	79 (36·7)	<0·0001	3739 (14·8)	81 (20)	0·004
Prior bleeding	1285 (5·2)	59 (6·6)	0·07	1335 (5·2)	9 (4·2)	0·49	1301 (5·2)	43 (10·6)	<0·0001
Peripheral artery disease	686 (2·8)	61 (6·8)	<0·0001	737 (2·9)	10 (4·7)	0·13	726 (2·9)	21 (5·2)	0·01
COPD	1409 (5·7)	124 (13·8)	<0·0001	1518 (6)	15 (7)	0·53	1485 (5·9)	48 (11·9)	<0·0001
Treatment									
OAC	21266 (85·9)	711 (79·3)	<0·0001	21812 (85·7)	165 (76·7)	0·0002	21614 (85·6)	363 (89·6)	0·02
Dabigatran	8326 (33·6)	213 (23·7)		8477 (33·3)	62 (28·8)		8428 (39·0)	111 (27·4)	
VKA	4547 (18·4)	226 (25·2)		4737 (18·6)	36 (16·7)		4656 (21·5)	117 (28·9)	
Apixaban	4274 (17·3)	134 (14·9)		4368 (17·2)	40 (18·6)		4347 (20·1)	61 (15·1)	
Rivaroxaban	3803 (15·4)	130 (14·5)		3908 (15·4)	25 (11·6)		3863 (17·9)	70 (17·3)	
Edoxaban	316 (1·3)	8 (0·9)		322 (1·3)	2 (0·9)		320 (1·5)	4 (1·0)	
AF ablation	438 (1·8)	4 (0·4)	0·003	441 (1·7)	1 (0·5)	0·15	439 (1·7)	3 (0·7)	0·13
Antiplatelet	5923 (23·9)	276 (30·8)	<0·0001	6139 (24·1)	60 (27·9)	0·2	6070 (24)	129 (31·9)	0·0003
Antiarrhythmic drug	6399 (25·8)	226 (25·2)	0·66	6592 (25·9)	33 (15·3)	0·0004	6530 (25·9)	95 (23·5)	0·27
ACE-i	7534 (30·4)	289 (32·2)	0·25	7765 (30·5)	58 (27)	0·26	7689 (30·5)	134 (33·1)	0·25
Angiotensin receptor blocker	6393 (25·8)	214 (23·9)	0·19	6554 (25·8)	53 (24·7)	0·71	6498 (25·7)	109 (26·9)	0·59
Beta blocker	15579 (62·9)	580 (64·7)	0·29	16025 (63)	134 (62·3)	0·84	15881 (62·9)	278 (68·6)	0·02
Digoxin	2001 (8·1)	119 (13·3)	<0·0001	2095 (8·2)	25 (11·6)	0·07	2075 (8·2)	45 (11·1)	0·04
Diuretic	9255 (37·4)	493 (55)	<0·0001	9673 (38)	75 (34·9)	0·35	9555 (37·8)	193 (47·7)	0·0001
Statin	11104 (44·8)	430 (47·9)	0·07	11406 (44·8)	128 (59·5)	<0·0001	11311 (44·8)	223 (55·1)	<0·0001
AF classification									
Type of AF			<0·0001			0·032			0·284
Paroxysmal	13807 (55·8)	429 (47·8)		14135 (55·6)	101 (47·0)		14027 (55·6)	209 (51·6)	
Persistent	8482 (34·3)	339 (37·8)		8730 (34·3)	91 (42·3)		8670 (34·3)	151 (37·3)	
Permanent	2470 (10·0)	129 (14·4)		2576 (10·1)	23 (10·7)		2554 (10·1)	45 (11·1)	
EHRA classification			<0·0001			0·007			0·003
I	8399 (33·9)	350 (39·0)		8653 (34·0)	96 (44·7)		8586 (34·0)	163 (40·3)	
II	10085 (40·7)	278 (31·0)		10297 (40·5)	66 (30·7)		10225 (40·5)	138 (34·1)	
III	4853 (19·6)	215 (24·0)		5027 (19·8)	41 (19·1)		4997 (19·8)	71 (17·5)	
IV	1422 (5·7)	54 (6·0)		1464 (5·8)	12 (5·6)		1443 (5·7)	33 (8·2)	

Note: Values are n (%) or mean ± SD. ACE-i = angiotensin-converting enzyme inhibitor, AF = atrial fibrillation, BMI = body mass index, BP = blood pressure, CCI = Charlson comorbidity index, COPD = chronic obstructive pulmonary disease, EHRA = European Heart Rhythm Association atrial fibrillation classification, OAC = oral anticoagulants, SD = standard deviation, VKA = vitamin K antagonist.

CHA_2DS_2-VASc calculated from congestive heart failure, hypertension, age, diabetes mellitus, stroke or transient ischaemic attack, vascular disease, sex (female).
HAS-BLED calculated from hypertension, abnormal renal or liver function, stroke, bleeding, international normalized ratio, elderly (age over 65 years), drugs or alcohol usage.
CCI calculated from age, myocardial infarction, congestive heart failure, peripheral vascular disease, cerebrovascular accident or transient ischaemic attack, dementia, COPD, peptic ulcer disease, liver disease, diabetes mellitus, chronic kidney disease, and cancer. However, it is noteworthy that certain factors such as connective tissue, hemiplegia, leukemic, lymphoma, and AIDS were not taken into consideration due to limitations in their availability in the data.

Results

The study cohort comprised 25,656 patients in total, which 10,961 patients excluded due to a lack of follow-up information. The patients were predominantly anticoagulated (86.0%) with a mean age of 70.3 years (SD 10.3), and 44.8% female, and with a mean CHA_2DS_2-VASc score of 3.2 (SD 1.5) (see Supplemental Table S2 for characteristics of the total cohort, Supplemental Table S3-4 for characteristics of the sub-cohorts).

During the first-year follow-up, there were 215 (0.8%) ischaemic stroke events with median time-to-event 139 days, 897 (3.5%) all-cause death with median time-to-event 158 days, and 405 (1.6%) major bleeding events with median time 138 days. Among patients who received anticoagulant therapy, the incidence of ischaemic stroke was 0.75 per 100 patient-years (PYs), the incidence of all-cause death was 3.2 per 100 PYs, and the incidence of major bleeding was 1.7 per 100 PYs.

The incidence of all-cause death varied across different regions, with incidence of 3.7 per 100 PYs among patients residing in Europe, 3.5 per 100 PYs for patients in North America, 2.3 per 100 PYs for patients residing in Asia, and 5.8 per 100 PYs for patients in Latin America. Table 1 presents a comparison of the baseline characteristics of patients with ischaemic stroke with those who had major bleeding or died during the same period. The use of oral anticoagulants was lower in patients who suffered an ischaemic stroke (76.7% vs 85.7%) or died (79.3% vs 85.9%), but higher in patients with major bleeding (89.6% vs 85.6%).

Performance

The predictive ability of the model for all-cause death was significantly better for ML-GBDT compared with CCI, with an AUC-ROC of 0.785 (95% CI 0.757-0.813) versus 0.747 (0.718-0.777), respectively ($p = 0.007$) (Table 2, Figure 2). We also evaluated the performance of CHA_2DS_2-VASc in predicting all-cause death, and it performed worse than the CCI (0.676 vs. 0.747, $p < 0.001$). The net reclassification index (NRI) for predicting all-cause death was 10.0% on the total cohort using our model. Correspondingly, the decision curve for ML-GBDT showed a higher net benefit than the CCI across the majority of the threshold probability range. Relative to CCI, the ML-GBDT yielded a higher sensitivity (0.726 vs. 0.677) while maintaining comparable specificity (0.707 vs. 0.704).

The AUC-ROC for predicting ischaemic stroke was significantly higher ($p = 0.028$) for ML-GBDT [0.691 (0.626-0.756)] compared to CHA_2DS_2-VASc [0.613 (0.538-0.689)]. The NRI showed an improvement of 12.5%. A similar specificity for both model and score were obtained while choosing the probability thresholds with ML-GBDT consistently demonstrating superior sensitivity.

An improvement was also observed in predicting major bleeding, with the AUC-ROC significantly ($p = 0.002$) higher for ML-GBDT than HAS-BLED: 0.698 (0.651-0.745) compared with 0.607 (0.559-0.655) (Table 2). The ML-GBDT model yielded a higher sensitivity (0.664 vs. 0.546) when predicting major bleeding. Out of the three outcomes, the NRI for major bleeding was the highest at 23.6%, in alignment with the decision curve analysis (Figure 2). The survival curve (Figure 2) for the "low risk" (first and second quartile) groups were observed to have a slower decline, indicating higher survival rates over time, compared to the "high risk" (third and fourth quartile) groups. It also outperformed single label GBDT in predicting the risks for patients independently (Supplementary Table S5).

Table 2: Comparison of the performance of risk scores and ML models to predict the outcomes in patients with AF.

		CCI (95% CI)	CHA_2DS_2-VASc (95% CI)	HAS-BLED (95% CI)	ML-GBDT (95% CI)
All cause death	**AUC**	0·747 (0·718, 0·777)	0·676 (0·643, 0·709)		**0·785 (0·757, 0·813)**
	SENS	0·677 (0·617, 0·733)			**0·726 (0·667, 0·778)**
	SPEC	0·704 (0·693, 0·714)			**0·707 (0·697, 0·717)**
Ischaemic stroke	**AUC**		0·613 (0·538, 0·689)		**0·691 (0·626, 0·756)**
	SENS		0·548 (0·425, 0·666)		**0·629 (0·505, 0·738)**
	SPEC		0·614 (0·603, 0·625)		**0·617 (0·606,0·628)**
Major bleeding	**AUC**			0·607 (0·559, 0·655)	**0·698 (0·651, 0·745)**
	SENS			0·546 (0·457, 0·633)	**0·664 (0·575, 0·742)**
	SPEC			0·634 (0·623, 0·644)	**0·634 (0·623, 0·645)**

AUC =area under the receiver operating characteristic curve, CI = confidence interval, SENS = sensitivity, SPEC = specificity, CCI = Charlson comorbidity index, ML-GBDT = multi-label gradient boosting decision tree.

Subgroup analyses
Compared to the CCI, the ML-GBDT model demonstrated superior discriminatory capacity ($p < 0.05$) on all sub-cohorts, except for patients for secondary stroke prevention. The CCI showed a reduced AUC-ROC while predicting all-cause death in patients who were older (age ≥ 65), anticoagulated, residing in European, North American or Asian regions, compared to the total cohort (Table 3). For patients who were not anticoagulated, those from Europe, or patients for primary or secondary stroke prevention, AUC-ROC of CCI was comparable or higher than the total cohort.

The AUC-ROC of the CHA_2DS_2-VASc score was higher among elderly, anticoagulated, European and North American sub-cohorts, as opposed to its AUC-ROC for the total cohort while estimating the risks of ischaemic stroke (Table 3). However, the performance of CHA_2DS_2-VASc score declined in patients from Asia or those for either primary or secondary stroke prevention, relative to the total cohort. The ML-GBDT model outperformed the CHA_2DS_2-VASc score in predicting ischaemic stroke across all the sub-groups, but none were statistically significant.

The AUC-ROC of HAS-BLED remained consistent or increased among several sub-groups of patients when compared its AUC-ROC for the total cohort. These subgroups included those who received anticoagulant therapy, those who were from European and North American regions, and patients for primary stroke prevention. However, it was lower for elderly patients, those not receiving anticoagulation therapy, patients from Asia, and those for secondary stroke prevention, in comparison to the total cohort. The ML-GBDT achieved higher AUC-ROC than the HAS-BLED, but not all observed predictions were statistically different.

Figure 3 and Figure 4 illustrate the top 25 or 30 most importance variables for total and sub-cohorts in the ML-GBDT model. Age, smoking, creatinine clearance, anticoagulant therapy, and prior history of stroke or TIA emerged as the most critical variables in predicting the risks for patients diagnosed with AF. In the sub-cohorts, the ranking of top features changed, including blood pressure, regions, BMI, previous thromboembolism, diabetes, respiratory disease, congestive heart failure and statin therapy. These additional variables, along with age, creatinine clearance and history of stroke/TIA, significantly impacted risk prediction.

Table 3: Comparison of the AUC-ROC of the ML-GBDT model and clinical risk scores across various study sub-groups.

	CCI (95% CI)	CHA_2DS_2-VASc (95% CI)	HAS-BLED (95% CI)	ML-GBDT (95% CI)	p-value (score vs. ML model)
age ≥ 65 years old					
All cause death	0·712 (0·679, 0·746)			0·752 (0·718, 0·785)	0·0283
Ischaemic stroke		0·653 (0·575, 0·730)		0·715 (0·652, 0·778)	0·0913
Major bleeding			0·548 (0·492, 0·605)	0·627 (0·572, 0·682)	0·0214
Anticoagulated					
All cause death	0·720 (0·687, 0·752)			0·759 (0·728, 0·790)	0·0027
Ischaemic stroke		0·698 (0·622, 0·774)		0·750 (0·677, 0·823)	0·0772
Major bleeding			0·601 (0·554, 0·647)	0·685 (0·637, 0·733)	0·0018
Not anticoagulated					
All cause death	0·759 (0·703, 0·816)			0·819 (0·766, 0·872)	0·0396
Ischaemic stroke		0·620 (0·475, 0·765)		0·694 (0·593, 0·796)	0·2394
Major bleeding			0·584 (0·416, 0·752)	0·678 (0·540, 0·815)	0·2838
Europe					
All cause death	0·747 (0·711, 0·783)			0·796 (0·763, 0·828)	0·0014
Ischaemic stroke		0·681 (0·600, 0·762)		0·737 (0·663, 0·810)	0·2102
Major bleeding			0·645 (0·581, 0·709)	0·690 (0·611, 0·753)	0·3232
North America					
All cause death	0·723 (0·663, 0·783)			0·804 (0·742, 0·865)	0·0050
Ischaemic stroke		0·696 (0·515, 0·878)		0·705 (0·530, 0·880)	0·8928
Major bleeding			0·653 (0·564, 0·742)	0·708 (0·626, 0·789)	0·3186
Asia					
All cause death	0·692 (0·584, 0·799)			0·840 (0·766, 0·914)	0·0123
Ischaemic stroke		0·532 (0·390, 0·675)		0·676 (0·515, 0·836)	0·1419
Major bleeding			0·582 (0·452, 0·713)	0·606 (0·467, 0·746)	0·7973
Primary prevention					
All cause death	0·744 (0·714, 0·773)			0·782 (0·750, 0·814)	0·0104
Ischaemic stroke		0·527 (0·455, 0·598)		0·598 (0·518, 0·677)	0·1343
Major bleeding			0·623 (0·569, 0·676)	0·656 (0·604, 0·709)	0·2380
Secondary prevention					

All cause death	0·743 (0·681, 0·805)			0·773 (0·707, 0·839)	0·4192
Ischaemic stroke		0·532 (0·440, 0·624)		0·627 (0·530, 0·724)	0·2319
Major bleeding			0·556 (0·414, 0·697)	0·630 (0·518, 0·743)	0·3848

The ML-GBDT was trained and tested on the sub cohort. AUC-ROC = area under the receiver operating characteristic curve, CI = confidence interval, CCI = Charlson comorbidity index, ML-GBDT = multi-label gradient boosting decision tree.

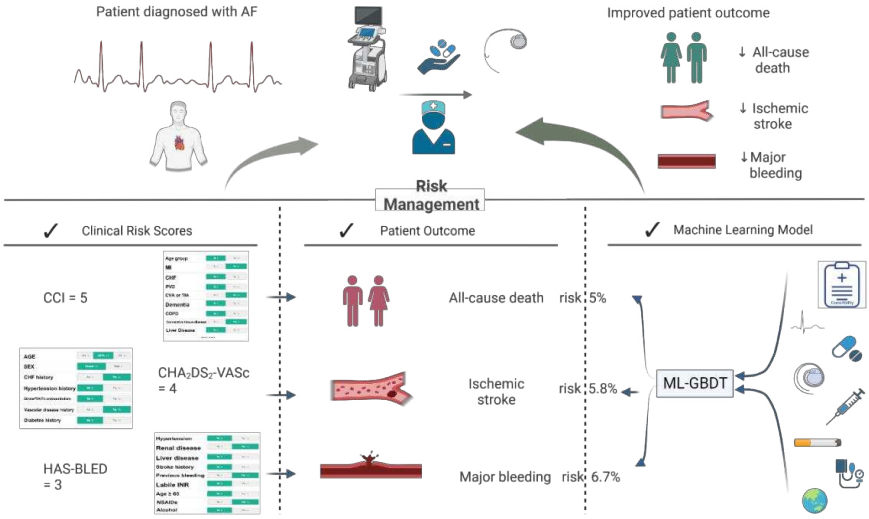


Figure 1. Graphic Abstract

AF=atrial fibrillation; CCI = Charlson Comorbidity Index; CHA_2DS_2-VASc includes congestive heart failure, hypertension, age, diabetes mellitus, stroke or transient ischaemic attack, vascular disease, sex (female); HAS-BLED includes hypertension, abnormal renal or liver function, stroke, bleeding, international normalized ratio, elderly (age over 65 years), drugs or alcohol usage; ML-GBDT = multi-label gradient boosting decision tree.

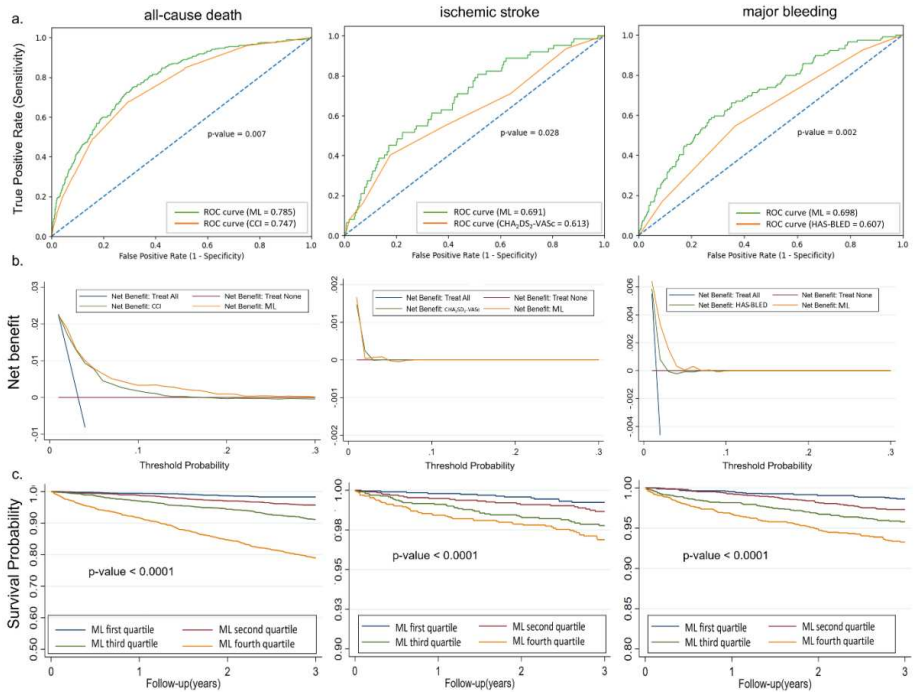


Figure 2. Performance of machine learning model compared to that of clinical scores in predicting risk of all-cause death, ischaemic stroke and major bleeding in patients with atrial fibrillation.

a) AUC-ROC curve. b) decision analysis curve. The x-axis represents the threshold probability, and the y-axis illustrates the net benefit derived from each model. The "treat all" and "treat none" strategies are also depicted on the decision curve. c) Kaplan Meier survival curves. The survival curves plot time in years on the x-axis, against the survival probability on the y-axis. The predicted probabilities were grouped into quartiles, with each represented by a distinct survival curve, corresponding to the respective predicted probability range.

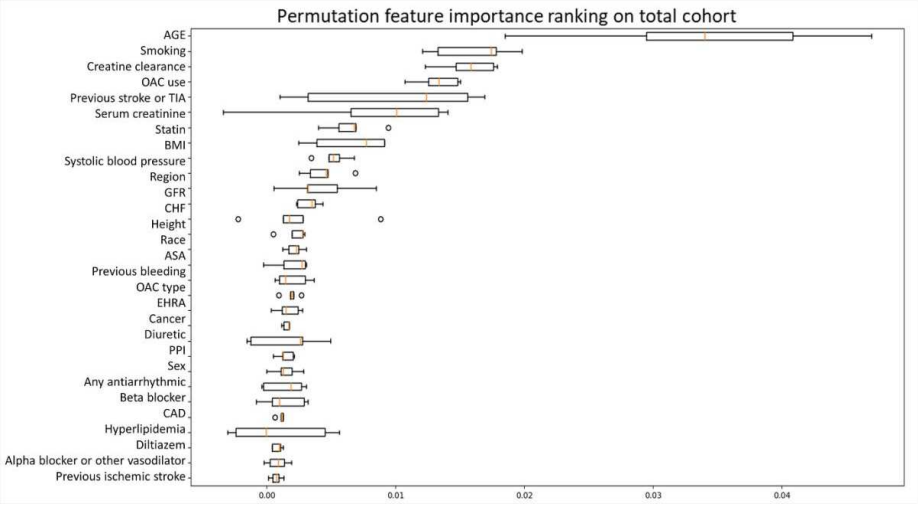


Figure 3. Permutation feature importance in the multi-label gradient boosting decision tree model on the total cohort while predicting outcomes of all-cause death, ischemic stroke and major bleeding.

ASA = aspirin, BMI = body mass index, CAD = coronary artery disease, CHF = congestive heart failure, EHRA = European Heart Rhythm Association atrial fibrillation classification, GFR = glomerular filtration rate, OAC = oral anticoagulant therapy, PPI = proton pump inhibitors, TIA = transient ischaemic attack.

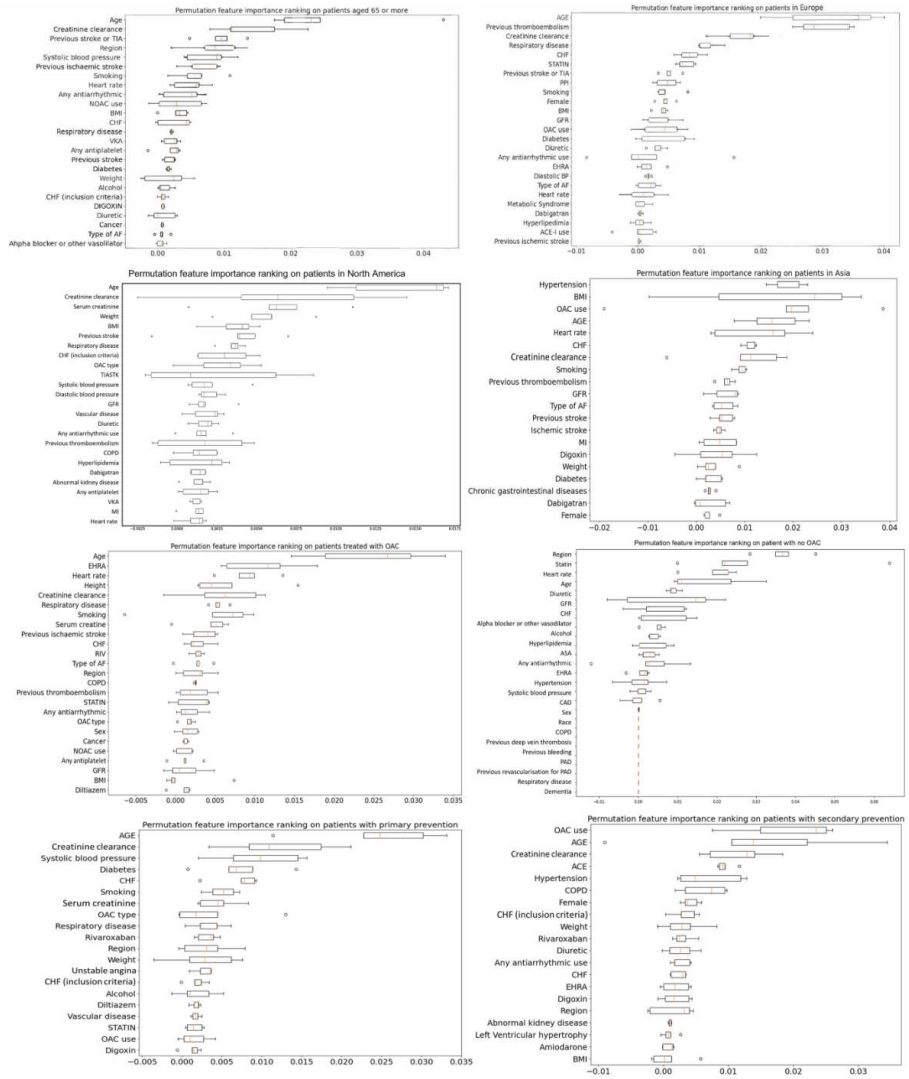


Figure 4. Permutation feature importance rankings in the multi-label gradient boosting decision tree model on the sub-cohort while predicting outcomes of all-cause death, ischemic stroke and major bleeding.

ACE = angiotensin-converting enzyme inhibitors, AF = atrial fibrillation, ASA = aspirin, BMI = body mass index, CAD = coronary artery disease, CHF = congestive heart failure, COPD = chronic obstructive pulmonary disease, EHRA = European Heart Rhythm Association atrial fibrillation classification, GFR = glomerular filtration rate, MI = myocardial infarction, NOAC = non-vitamin K antagonist oral anticoagulants, OAC = oral anticoagulants, PAD = peripheral artery disease, PPI = proton pump inhibitors, RIV = rivaroxaban, STK = stroke, TIA = transient ischemic attack, VKA = vitamin K antagonist.

Discussion

In this study, we have shown that a ML-GBDT model performed better than three clinical risk scores in predicting the risk of death, ischemic stroke and major bleeding, for patients diagnosed with AF. This attained statistically significant optimized performance in comparison to the CCI for predicting risk of all-cause death, the CHA_2DS_2-VASc score for ischaemic stroke risk, and the HAS-BLED score for major bleeding risk. Moreover, our study is the first to present the feature importance ranking for patient cohorts featured by geographical regions, receiving anticoagulant therapy, and history of stroke.

During the last decades, there has been growing recognition that death, especially cardiovascular death, associated with AF is increasing[21] although there is no dedicated score for this specific population. It is therefore important to implement appropriate measures to mitigate the risk of death for patients with AF. Our developed ML-GBDT model provides statistically significant improved performance in assessing the risk of all-cause death accounting for risk of ischaemic stroke and major bleeding.

Whilst oral anticoagulant therapy is used to reduce the risk of stroke, it simultaneously increases the likelihood of major bleeding events.[2] Patients may still experience an ischaemic stroke event despite receiving anticoagulant therapy.[3] Also, despite an uptake in the use of anticoagulants, one-third of high-risk patients with AF do not receive any anticoagulant treatment.[22] As a result, the risk of death, stroke and major bleeding should be monitored to identify patients at increased risk so that appropriate intervention or closer monitoring can be initiated. The low number of events, specifically ischaemic stroke, resulting from the use of anticoagulant therapy, is a challenge for prediction. This has indeed impacted the predictive performance for both clinical scores and our model.

Using clinical risk scores, anticoagulant therapy is recommended in patients with a CHA_2DS_2-VASc score ≥ 2 in males or ≥ 3 for females and should be considered when CHA_2DS_2-VASc score is 1 in males or 2 for females. Meanwhile, patients at high risk of bleeding (HAS-BLED score ≥ 3) should be scheduled for early and more frequent clinical review and follow-up.[4] However, scores which only include a limited number of risk factors report modest capacity in predicting patient risks. Our model which additionally incorporated treatments, interventions and an extensive range of comorbidities, yielded the best results. By leveraging the predicted risk probability and feature importance, clinicians could address modifiable risk factors and adjust treatments and interventions utilising our model.

AF necessitates a multifaceted, holistic approach to its management, and requires active patient involvement in partnership with clinicians. The Atrial Fibrillation Better Care (ABC) pathway was introduced as an approach to simplify the integrated care in patients with AF,[23] and adhering to it has been shown to improve the prognosis of AF patients.[24] Current guidelines stipulate that an holistic or integrated care approach to AF management should identify and manage risk factors and comorbidities.[4,5]

With insights from feature importance in the ML-GBDT model, it is additionally possible to characterize subclinical processes, and target key modifiable risk factors to modify patient risks, such as hypertension, diabetes and BMI. Our study shows the power of incorporating a wide range of comorbidities and existing treatments in enhancing the risk prediction performance. This could potentially enrich the precision of interventions and lead to optimized patient care. The ML-GBDT model and its feature importance could further be integrated with electronic

medical records or mobile apps in the future in daily clinical practice to facilitate better stratification of risk and improve the prognosis of patients through tailored management.

Anticoagulation

Despite receiving anticoagulation treatment, patients with AF may still face an enduring risk of stroke and mortality. Residual risks for death and ischaemic stroke were predicted by our model in patients who received anticoagulant therapy. It is worth highlighting that the use of anticoagulant therapy had a significant impact on the performance of risk prediction, as indicated in Table 3. The majority of patients in the GLORIA-AF registry received anticoagulant therapy due to their elevated risk of ischaemic stroke (CHA_2DS_2-VASc score $\geq$ 1), and the number of patients who were not anticoagulated was small, resulted in a less statistically significant finding while comparing the performance of model and scores in the latter group.

Patient location

Our results indicated that the location where patients reside was statistically significant for each outcome ($p < 0.0001$) and influenced the risk prediction performance. The AUC-ROCs for all the clinical risk scores were lower for patients from Asia. This could potentially be explained by the age difference across regions (Supplementary S3). Chao et al.[9] showed that a modified CHA_2DS_2-VASc optimized performance for stroke risk prediction in Asian patients, which reflected the age threshold in Asian patients with AF for an increased risk of ischaemic stroke could be 50 years. Our model possibly offers a more robust solution for patient risk assessment of death, ischaemic stroke, and major bleeding. The feature importance ranking of Asian patients showed different results compared to European and North American cohorts, explaining the performance variation in different populations and providing new avenues for future research.

Primary vs secondary prevention subgroups

Stroke history is a critical risk factor while assessing a patients' risk of ischaemic stroke. Both machine learning and clinical risk scores yielded sub-optimized performance in predicting the risks for patients who have not previously had a stroke. Lower event rates were observed in these patients (0.63%) compared to 2.1% for secondary prevention (Supplementary Table S4). This was lower than the observed event rate in the Darlington AF registry,[25] but this could be due to more than 85% of patients in the GLORIA-AF registry receiving anticoagulant therapy. The OAC use was also ranked as one of the top important factors for patients with secondary prevention of stroke in Figure 4.

Machine learning

Machine learning has been increasingly applied to predict risks for patients with AF using various data sources.[26–29] Ambale-Venkatesh et al. employed different data sources, including imaging, text, questionnaires, and biomarkers, to improve cardiovascular event predictions through machine learning algorithms.[27] Machine learning has also been used to uncover risk factors that were previously unknown and improve stroke prediction in patients with AF.[28] Sequential measures of international normalized ratio have also been used to predict clinical outcomes in newly diagnosed patients with using machine learning.[29] Compared to other applications of machine learning, our ML-GBDT model took into account the relationship between different outcomes (all-cause death, ischaemic stroke, and major bleeding) for patients with AF. We have also compared the feature importance ranking of different sub-cohorts, which could provide additional support in practice.

Limitations

A key limitation in our study cohort is the relatively low number of events, especially ischaemic stroke, due to the use of anticoagulant therapy. The current study does not account for possible changes in anticoagulant status during follow-up, nor the impact of clinical interventions that impact death, stroke, or bleeding, such as early rhythm control.[30] Another limitation to the predictive performance could be the structure of the classifier chains. The sequential nature of the model, where each subsequent classifier's input includes the output of the preceding one, implies that information propagates along the chain, including both performance gains and inaccuracies. Although inaccuracies may affect the prediction, the information propagation may have ultimately led to enhanced performance. One of the challenges associated with the application of machine learning models is the requirement for a much larger number of variables compared to risk scores because it accounts for an extensive range of patient comorbidities, current treatments and interventions. This might pose additional workload to collect these variables in already busy clinical settings. However, the machine learning process has the capability of full automation, by integrating the model directly into electronic health record systems. The whole process, from data extraction to risk prediction could occur in real-time. This could substantially mitigate the associated burden of required data collection, making the adoption of models more feasible and less disruptive in future.

Conclusions

Our study validates that the multi-label machine learning model outperforms the clinical risk scores in predicting the risk of all-cause death, ischaemic stroke and major bleeding in patients with AF from a contemporary global registry. We also provide feature importance ranking for various patient cohorts by geographical regions, receiving anticoagulation therapy, and history of stroke. Our approach could be used as a multifaceted and holistic tool to optimize patient risk assessments and mitigate adverse outcomes in diverse AF populations.

Contributors

GYL, GD, JL, and AB designed the study. JL and AB processed the data and did the statistical analysis. JL developed and tested the model. JL wrote the manuscript. JL, AB, MB, YZ, FMS, JH, TB, BM, JS, GF, MVH, GB and GYL critically reviewed the manuscript.

Declaration of interests

We declare no competing interests.

Reference

1. Kornej, J., Börschel, C. S., Benjamin, E. J. & Schnabel, R. B. Epidemiology of Atrial Fibrillation in the 21st Century. *Circ. Res.* **127**, 4–20 (2020).
2. Ruff, C. T. *et al.* Comparison of the efficacy and safety of new oral anticoagulants with warfarin in patients with atrial fibrillation: a meta-analysis of randomised trials. in *Database of Abstracts of Reviews of Effects (DARE): Quality-assessed Reviews [Internet]* (Centre for Reviews and Dissemination (UK), 2014).
3. Ding, W. Y. Residual Stroke Risk in Atrial Fibrillation. *Arrhythmia Electrophysiol. Rev.* **10**, 147–153 (2021).
4. Hindricks, G. *et al.* 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS). *Eur. Heart J.* ehaa612 (2020) doi:10.1093/eurheartj/ehaa612.
5. Chao, T.-F. *et al.* 2021 Focused Update Consensus Guidelines of the Asia Pacific Heart Rhythm Society on Stroke Prevention in Atrial Fibrillation: Executive Summary. *Thromb. Haemost.* **122**, 20 (2022).
6. Chen, J.-Y. *et al.* CHADS2 versus CHA2DS2-VASc score in assessing the stroke and thromboembolism risk stratification in patients with atrial fibrillation: a systematic review and meta-analysis. *J. Geriatr. Cardiol. JGC* **10**, 258–266 (2013).
7. Senoo, K., Proietti, M., Lane, D. A. & Lip, G. Y. H. Evaluation of the HAS-BLED, ATRIA, and ORBIT Bleeding Risk Scores in Patients with Atrial Fibrillation Taking Warfarin. *Am. J. Med.* **129**, 600–607 (2016).
8. van Doorn, S. *et al.* Predictive performance of the CHA2DS2-VASc rule in atrial fibrillation: a systematic review and meta-analysis. *J. Thromb. Haemost.* **15**, 1065–1077 (2017).
9. Chao, T.-F. *et al.* Validation of a Modified CHA2DS2-VASc Score for Stroke Risk Stratification in Asian Patients With Atrial Fibrillation. *Stroke* **47**, 2462–2469 (2016).
10. Charlson, M. E., Carrozzino, D., Guidi, J. & Patierno, C. Charlson Comorbidity Index: A Critical Review of Clinimetric Properties. *Psychother. Psychosom.* **91**, 8–35 (2022).
11. Lip, G. Y. H. *et al.* Improving Stroke Risk Prediction in the General Population: A Comparative Assessment of Common Clinical Rules, a New Multimorbid Index, and Machine-Learning-Based Algorithms. *Thromb. Haemost.* **122**, 142–150 (2022).
12. Lu, J. *et al.* Performance of multilabel machine learning models and risk stratification schemas for predicting stroke and bleeding risk in patients with non-valvular atrial fibrillation. *Comput. Biol. Med.* **150**, 106126 (2022).
13. Huisman, M. V. *et al.* Design and rationale of Global Registry on Long-Term Oral Antithrombotic Treatment in Patients with Atrial Fibrillation: A global registry program on long-term oral antithrombotic treatment in patients with atrial fibrillation. *Am. Heart J.* **167**, 329–334 (2014).
14. Kirchhof, P. *et al.* 2016 ESC Guidelines for the management of atrial fibrillation developed in collaboration with EACTS. *Eur. Heart J.* **37**, 2893–2962 (2016).
15. Ding, W. Y. *et al.* Incidence and Risk Factors for Residual Adverse Events Despite Anticoagulation in Atrial Fibrillation: Results From Phase II/III of the GLORIA-AF Registry. *J. Am. Heart Assoc.* **11**, (2022).
16. van der Wall, S. J. *et al.* Low bleeding and thromboembolic risk with continued dabigatran during cardiovascular interventions: the GLORIA-AF study. *Eur. J. Intern. Med.* **91**, 75–80 (2021).

17. Ke, G. *et al.* LightGBM: A Highly Efficient Gradient Boosting Decision Tree. in *Advances in Neural Information Processing Systems* vol. 30 (Curran Associates, Inc., 2017).
18. Read, J., Pfahringer, B., Holmes, G. & Frank, E. Classifier Chains for Multi-label Classification. in *Machine Learning and Knowledge Discovery in Databases* (eds. Buntine, W., Grobelnik, M., Mladenić, D. & Shawe-Taylor, J.) 254–269 (Springer, 2009). doi:10.1007/978-3-642-04174-7_17.
19. Breiman, L. Random Forests. *Mach. Learn.* **45**, 5–32 (2001).
20. Snoek, J., Larochelle, H. & Adams, R. P. Practical Bayesian Optimization of Machine Learning Algorithms. Preprint at https://doi.org/10.48550/arXiv.1206.2944 (2012).
21. Tanaka, Y. *et al.* Trends in Cardiovascular Mortality Related to Atrial Fibrillation in the United States, 2011 to 2018. *J. Am. Heart Assoc.* **10**, e020163 (2021).
22. Navar, A. M. *et al.* Trends in Oral Anticoagulant Use Among 436 864 Patients With Atrial Fibrillation in Community Practice, 2011 to 2020. *J. Am. Heart Assoc.* **11**, e026723 (2022).
23. Lip, G. Y. H. The ABC pathway: an integrated approach to improve AF management. *Nat. Rev. Cardiol.* **14**, 627–628 (2017).
24. Romiti, G. F. *et al.* Adherence to the Atrial Fibrillation Better Care (ABC) pathway and the risk of major outcomes in patients with atrial fibrillation: a post-hoc analysis from the prospective GLORIA-AF Registry. *eClinicalMedicine* **55**, (2023).
25. Mazurek, M. *et al.* Secondary Versus Primary Stroke Prevention in Atrial Fibrillation: Insights From the Darlington Atrial Fibrillation Registry. *Stroke* **48**, 2198–2205 (2017).
26. Olier, I., Ortega-Martorell, S., Pieroni, M. & Lip, G. Y. H. How machine learning is impacting research in atrial fibrillation: implications for risk prediction and future management. *Cardiovasc. Res.* **117**, 1700–1717 (2021).
27. Ambale-Venkatesh, B. *et al.* Cardiovascular Event Prediction by Machine Learning. *Circ. Res.* **121**, 1092–1101 (2017).
28. Li, X. *et al.* Integrated Machine Learning Approaches for Predicting Ischemic Stroke and Thromboembolism in Atrial Fibrillation. *AMIA. Annu. Symp. Proc.* **2016**, 799–807 (2017).
29. Goto, S. *et al.* New artificial intelligence prediction model using serial prothrombin time international normalized ratio measurements in atrial fibrillation patients on vitamin K antagonists: GARFIELD-AF. *Eur. Heart J. - Cardiovasc. Pharmacother.* **6**, 301–309 (2020).
30. Chao, T.-F. *et al.* Early Rhythm Control and the Risks of Ischemic Stroke, Heart Failure, Mortality, and Adverse Events When Performed Early (<3 Months): A Nationwide Cohort Study of Newly Diagnosed Patients with Atrial Fibrillation. *Thromb. Haemost.* **122**, 1899–1910 (2022).

1.4 Discussion

Dans le premier travail portant sur la prédiction de la FA survenant dans les suites d'un AVCi, nous avons pu rapporter une des premières applications du ML sur la base de données nationale Française du PMSI [111]. Parmi les cinq algorithmes de ML entraînés, le réseau de neurones a été sélectionné comme étant le modèle présentant les meilleures performances pour discriminer les patients à risque de développer une FA. Comparée aux scores de risque cliniques habituels, cette approche avait une capacité prédictive significativement supérieure.

La base du PMSI contient des millions de séjours et des milliers de variables constituant une mine d'informations et un terrain d'application privilégié pour le ML. Cependant, les systèmes de cotation essentiellement binaires (classification Internationale des Maladies (CIM-10), Classification Commune des Actes Médicaux (CCAM), Liste des Produits et Prestations (LPP)) réduisent considérablement la granularité des variables obtenues et par conséquent la complexité de ces données ce qui limite la valeur ajoutée du ML.

Malgré tout, les performances prédictives d'un réseau de neurones étaient supérieures aux modèles de régression de Cox préalablement développés sur les mêmes données et supérieures aux scores cliniques déjà existants avec un taux de reclassification intéressant [62,100,113].

Cette approche prometteuse pourra permettre à l'avenir une meilleure stratification des risques pour améliorer la prise de décision concernant le choix d'une stratégie de dépistage intensif de la FA chez les patients les plus à risque ayant subi un AVCi.

L'évolution des bases nationales de données de santé française au sein du Système de Données de Santé (SNDS), l'une des plus importante au monde en termes de volume et d'exhaustivité, et l'ajout de nouveaux indicateurs avec plus de granularité permettra certainement de tirer un encore meilleur parti du ML.

Le second travail a permis d'élaborer un modèle de prédiction de la mortalité chez les patients atteints de FA à partir d'un autre jeu de données provenant du PMSI [112]. Comparée aux scores de risque clinique actuels, cette approche avait une meilleure capacité prédictive pour cibler les patients présentant un risque de mortalité élevé.

Parmi les nombreuses données disponibles, nous avons sélectionné les variables d'importance en utilisant un estimateur de type Random Forest afin d'éliminer les variables statistiquement redondantes ou non prédictives. Cette méthode nous a permis de faire deux observations : 1) Certaines variables semblant, *a priori*, redondantes étaient pourtant sélectionnés par le modèle comme étant importantes si bien que leur exclusion nuisait aux performances finales ; 2) Cela nous a permis de retenir des variables pouvant apparaître comme intuitivement peu pertinentes ou jusqu'à lors peu décrites sur le sujet. Dans les deux cas, l'utilisation d'une méthode s'apparentant plus à une approche empirico-inductive (data-driven ou bottom-up) qu'hypothético-déductive (hypothesis-driven ou top-down), caractéristique du ML, nous a permis de tirer le plus d'information possible à partir de la donnée directement sans conceptions pré établies *a priori*.

Par ailleurs, dans un souci de simplification de notre algorithme et de puissance de calcul, nous avons réduit de nombre de variable à k=18 (meilleur ratio AUC/K) ce qui est assez limitant pour un usage généralisé en pratique courante, les scores de risque devant trouver un équilibre entre praticité d'une part et précision d'autre part. Cependant synthétiser les concepts de complexité et fragilité clinique en un nombre réduit de variables est illogique, comme le montrent les scores actuels intégrant un nombre d'items comparable voir supérieur [114,115].

Les modèles de ML peuvent aider les cliniciens à mieux stratifier le risque de mortalité des patients atteints de FA. La FA étant associée à une mortalité non cardiovasculaire importante à court terme, identifier les patients les plus à risque permet d'éviter la réalisation d'actes ou de thérapies cardiovasculaires futiles au profit d'une prise en charge plus globale. L'évolution vers une médecine de précision explorant la complexité clinique des patients va nécessiter l'utilisation de modèles intégrant de plus en plus variables provenant différentes sources. A l'aune des applications numériques et de la possibilité d'intégrer directement ces modèles aux logiciels des dossiers médicaux informatisés, la complexité de ces outils n'est plus un frein.

Enfin, le troisième travail montre l'applicabilité et l'utilité d'un modèle de type multi-label gradient boosting decision tree (ML-GBDT) en chaine dans la prédiction d'évènements indésirable à partir d'un registre prospectif international de patients atteints de FA (**Annexe 12**). Le ML-GBDT en chaine était plus performant que les scores cliniques usuels dédiés dans la prédiction des saignements majeurs, AVCi et décès.

Cette étude montre aussi que l'incorporation des relations entre différents labels (variables à prédire) améliorait les performances du modèle par rapport à un modèle à prédiction unique (single-label GBDT).

L'exploration de ce registre international souligne l'hétérogénéité de la population atteinte de FA. Les caractéristiques cliniques varient selon l'âge, la localisation géographique, les traitements, la présence d'un AVCi préalable et en conséquence les risques d'évènements indésirables. Du fait des variations des relations entre features (variables prédictives) et labels (variables à prédire), développer des modèles universels devient donc extrêmement difficile [79]. Ceci explique les variations en termes de performances des différents scores selon les sous-populations. Il est intéressant de constater que selon la sous-population choisie, le classement des variables par ordre d'importance selon l'algorithme n'était pas le même. Cette observation permet de générer des hypothèses afin d'explorer des pistes jusqu'à lors inconnues.

Le Parcours ABC a été introduit pour systématiser et simplifier la prise en charge holistique et intégrée des patients souffrant de FA et est actuellement promu dans les recommandations [1,34]. Il insiste sur l'identification des facteurs de risque modifiables. A la lumière du classement des variables selon leur importance fourni par l'algorithme, il semble désormais possible de cibler les facteurs de risque modifiables critiques et spécifiques à une population afin d'optimiser la prise en charge de la FA et ce de manière adaptée selon la population concernée. De plus, il serait tout aussi possible de personnaliser le traitement anticoagulant chez les populations à plus haut risque thromboembolique ou hémorragique. Enfin, prédire l'ensemble des risques liés à la FA grâce à un modèle à prédiction multiples (multi-label) s'inscrit parfaitement dans une démarche globale de prise en charge.

2 Deuxième partie : Modèles non supervisés et analyse en cluster au sein d'une population de patients atteints de fibrillation atriale

2.1 Étude 4 : Phénotypes cliniques et récidives de fibrillation atriale après ablation par cathéter : Une analyse non supervisée en clusters

Bisson A**, Fawzy AM, El-Bouri W, Angoulvant D, Lip GYH, Fauchier L, Clementy N.* ***Clinical Phenotypes and Atrial Fibrillation Recurrences After Catheter Ablation: An Unsupervised Cluster Analysis. *Curr Probl Cardiol. 2023 Aug;48(8):101732. doi: 10.1016/j.cpcardiol.2023.101732.* [116]

1. Objectif de l'étude

Dans ce travail, l'objectif était d'explorer l'hétérogénéité phénotypique des patients atteints de FA et de distinguer des clusters de patients associés au risque de récidive d'arythmie dans les suites d'une ablation par cathéter.

2. Méthode

Nous nous sommes basés pour cela sur une cohorte rétrospective de patients atteints de FA symptomatiques référés pour une première ablation par cathéter. Après une étape de preprocessing, 33 variables ont été analysées par un algorithme de machine learning non supervisé de type agglomerative hierarchical clustering. L'association en terme pronostique entre les phénogroupes et le risque de récidive d'arythmie a ensuite été évalué.

3. Résultats

Au total, 209 patients consécutifs ont été inclus dont l'âge moyen était de 62±10 ans parmi lesquels 103 (49%) étaient atteints de FA paroxystique. 3 clusters aux caractéristiques distinctes ont été identifiés. A 1 an, le taux de récidive d'arythmie était de 27.2% pour le Cluster 1, 43.2% pour le Cluster 2 et 60.9% pour le Cluster 3 ($p<0.0001$). Cette classification en clusters était associée aux récidives d'arythmie (HR 1,58, IC à 95 % 1,01-2,49, P = 0,046) et ce après ajustement sur l'âge, le score CHA_2DS_2-VASc, le volume de l'oreillette gauche, le type de FA et la fraction d'éjection ventriculaire gauche.

4. Conclusion

L'analyse non supervisée en clusters a permis d'identifier différents phénogroupes de patients atteints de FA et traités par ablation par cathéter. Ces résultats donnent un nouvel aperçu de la complexité clinique des patients traités par ablation de la FA et de son influence sur les résultats de cette procédure. D'autres études basées sur des bases de données de plus haute dimension avec des échantillons de grande taille sont nécessaires pour confirmer ces résultats et les implications futures potentielles dans la pratique clinique.

Clinical Phenotypes and Atrial Fibrillation Recurrences After Catheter Ablation: An Unsupervised Cluster Analysis

Arnaud Bisson, MD[a,b,c,d*], Ameenathul M. Fawzy, MD[d], Wahbi El-Bouri, PhD[d], Denis Angoulvant, MD, PhD[a,c,#], Gregory Y.H. LIP, MD[d,e,#], Laurent Fauchier, MD, PhD[a,#], and Nicolas Clementy, MD, PhD[f,#]

From the [a] Service de Cardiologie, Centre Hospitalier Régional Universitaire et Faculté de Médecine de Tours, Tours, France, [b] Service de Cardiologie, Centre Hospitalier Régional Universitaire d'Orléans, Orléans, France, [c] EA4245, Transplantation Immunité Inflammation, Université de Tours, Tours, France, [d] Liverpool Centre for Cardiovascular Science at University of Liverpool, Liverpool John Moores University and Liverpool Heart and Chest Hospital, Liverpool, UK, [e] Danish Center for Clinical Health Services Research, Aalborg University, Aalborg, Denmark and [f] Service de Cardiologie, Clinique du Millénaire, Montpellier, France.

Abstract: Catheter ablation (CA) is a well-established treatment of atrial fibrillation (AF). Data-driven cluster analysis is able to better distinguish prognostically-relevant phenotype clusters among patients with AF. We performed a hierarchical cluster analysis in a cohort of AF patients undergoing a first CA and evaluate associations between identified clusters and

[#] Joint senior authors.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Bisson has been a consultant or speaker for Astra-Zeneca, Bayer, BMS/Pfizer, Medtronic, Vitorpharma and Alnylam. Clementy has been a consultant or speaker for Abbott, Medtronic and Microport. Angoulvant has been a consultant or speaker for Amgen, Astra-Zeneca, Bayer, BMS/Pfizer, MSD, Novartis, Novo Nordisk, Sanofi, Servier. Lip has been a consultant or speaker for BMS/Pfizer, Boehringer Ingelheim, Daiichi-Sankyo, Anthos. No fees are received personally. LF reports consulting fees for AstraZeneca, Bayer, BMS/Pfizer, Boehringer Ingelheim, Medtronic, Novo Nordisk and Novartis and lecture fees for AstraZeneca, Bayer, BMS/Pfizer, Boehringer Ingelheim and Zoll. All other authors have reported that they have no relationships relevant to the contents of this paper to disclose.

*Correspondence: Arnaud BISSON, MD, Service de Cardiologie, Centre Hospitalier Regional Universitaire et Faculté de Médecine de Tours, Tours, France, 2 Boulevard Tonnellé, 37000 Tours, France, Tel: 33-247473659 E-mail: arnaud.bisson@univ-tours.fr

Curr Probl Cardiol 2023;48:101732
0146-2806/$ – see front matter
https://doi.org/10.1016/j.cpcardiol.2023.101732

recurrences of arrhythmia following ablation. The study included 209 AF patients treated with CA. A total of 3 clusters with distinct characteristics were identified. Recurrences at 1 year occurred in 27.2% in Cluster 1, 43.2% in Cluster 2 and 60.9% in Cluster 3 ($P < 0.0001$). Cluster classification was independently associated with arrhythmia recurrences (HR 1.58, 95% CI 1.01-2.49, $P = 0.046$) after adjustment for age, CHA_2DS_2-VASc score, left atrial volume, type of atrial fibrillation and ejection fraction. To concluded, cluster analysis identified 3 statistically-driven groups among AF patients treated with CA with different risks for arrhythmia recurrences. (Curr Probl Cardiol 2023;48:101732.)

Introduction

Over the last decades, catheter ablation has become a well-established therapy for rhythm control in patients with symptomatic atrial fibrillation (AF).[1-2] The latter has moved from a therapy of last resort for highly symptomatic patients with drug-refractory AF to an accepted first-line option for patients across the spectrum of AF type and severity. However, freedom from arrhythmia and maintenance of sinus rhythm at 1-year postablation is variable and the rate of recurrences ranges from 60 to almost 90% depending on many factors.[3-7] Identification of patient groups at higher risk for recurrent AF may help in developing preventive strategies and tailoring of rhythm control therapy postablation. A number of risk-prediction scores have been evaluated but with only modest predictive performance.[8-11] Of note, recurrence of AF after catheter ablation is driven by complex interactions of various factors and might not be optimally characterized with simplified risk scores. Rather than occurrence in isolation or a binary manner (eg. diabetes yes/no), clinical risk factors(s) often occur concurrently leading to potentially "clinically complex" phenotype clusters.

Unsupervised cluster analysis is a machine learning method that categorizes complex entities without investigators' supervision by segregating samples into homogenous groups based on each cluster's dissimilarities.[12] This data-driven approach helps unveil meaningful phenotypes within a disease that has been previously considered as homogenous. Thus, cluster analysis may be invaluable in phenotyping several cardiovascular diseases and exploring the clinical complexity of cardiovascular patients.

Hence, this study aimed to explore for the first time whether unsupervised cluster analysis can identify clinically relevant groups among AF patients treated with catheter ablation and second, to evaluate whether these cluster phenotypes are associated with different rates of arrhythmia recurrence.

Methods

Study Population

Consecutive patients with symptomatic AF referred to the Cardiology Department of the University Hospital of Tours (France) for AF ablation between February 2013 and May 2014 were included in this study. Patients with a prior ablation for AF were excluded. Collected clinical data included symptoms (EHRA classification) and history of arrhythmia, presence of thrombo-embolic risk factors (CHA2DS2-VASc score (Supplemental Table 1)), and past and current medications. Trans-thoracic echocardiography and cardiac magnetic resonance imaging (MRI) or computed tomography scan (CT-scan) were performed to assess left ventricular function and left atrial (LA) diameter and volume before ablation.

Procedures were performed under general anaesthesia. A 4-millimeter irrigated-tip catheter was used in all patients to deliver radiofrequency energy (Thermocool SF, Biosense Webster; Flexability, St Jude Medical). A circular lasso catheter (Biosense Webster, St Jude Medical) was used for mapping. After transseptal puncture, antral pulmonary vein isolation (PVI) was performed in all patients. A bidirectional block was systematically obtained in all veins. In patients with persistent AF and remaining arrhythmia at that stage a stepwise approach was performed: sequentially, anterior roof and mitral isthmus lines were obtained (endocardially, and epicardially through the coronary sinus, when necessary), complex fractionated atrial electrograms (CFAEs) within LA (LA appendage, inferior-posterior wall, interatrial septum), coronary sinus, and right atrium (crista terminalis, superior venacava, cavotricuspid isthmus), were mapped and defragmented when necessary. Stable atrial tachycardias were systematically mapped and ablated. When return to sinus rhythm was obtained, either with ablation, or with electrical cardioversion at the end of procedure, bidirectional block was confirmed on all performed lines.

Follow-Up and Outcomes

Recurrence was defined as $\geq$1 documented sustained episode ($\geq$30 seconds) of any atrial arrhythmia, symptomatic or not, on any ECG

or Holter monitoring strip (scheduled or additional), after a single ablation procedure, after a 3 months blanking period.

During the blanking period, antiarrhythmic drugs (AADs) were continued in most of the patients, and a cardioversion was performed in the event of persistent recurrence. At the end of the blanking period, AADs were systematically discontinued in all patients. All patients were rigorously followed for 12 months. A 24 hours Holter monitoring was performed at 3 and 6 months along with a clinical examination and a resting ECG. A 7 days Holter recording was systematically performed at 12 months (Spiderview, Sorin Group, Le Plessis-Robinson, France).

Cluster Analysis

The whole cohort was used for analysis. Variables with missing data rate for any variables above 30% were excluded and variables with missing rate under 30% were imputed using Multivariate Imputation By Chained Equations algorithm.[13] 33 Baseline clinical and procedural variables were therefore used for the analysis (Supplemental Table 2).

The hierarchical clustering method (using Ward's linkage criterion) was used to identify homogenous phenotypic subgroups of AF patients without prior knowledge of the outcomes. The algorithm begins with each element (ie, patient) as a separate cluster and then proceeds with a "bottom-up" approach grouping each cluster with the most similar 1 until all clusters become one.

A dendrogram is provided to display the distance obtained at each interaction of the clustering process (Fig 1). Small values of the distance indicate that the merged clusters were similar, and large values indicate the combination of 2 dissimilar (heterogeneous) clusters. The determination of the numbers of clusters was not prespecified. Examination of the dendrogram indicated that the groupings became more heterogeneous after being expanded to 3 clusters. Both 3-cluster and 4-cluster models were examined. The 3-cluster model formed much clearer patterns of patient groups than a 4-cluster model. Therefore the 3-cluster model was used in this study. Cluster models were implemented in Python using open-source packages: Scikit-learn version 1.1.1. Once clusters were identified, we assessed the association between clusters and clinical characteristics and outcomes.

Statistical Analysis

Qualitative variables were described using counts and percentages, and continuous quantitative variables were described as mean ± standard

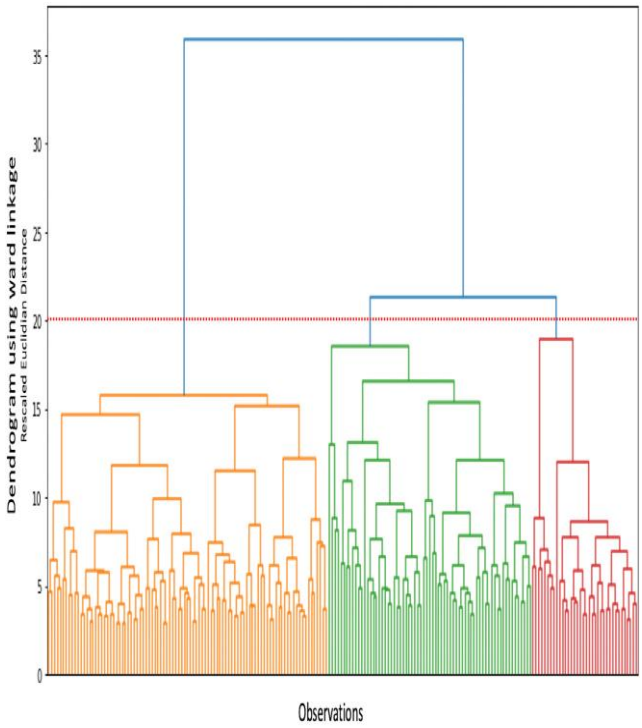


FIG 1. Dendrogram generated by hierarchical clustering showing 3 main clusters. The dendrogram represents the relative degree of similarity between individual patients. The greater the height of the branch points (distance), the greater the differences are between the branches and the more dissimilarity exist between clusters. The red line indicates the stopping location.

deviation (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 Chi square test was used to compare categorical data. Unadjusted and multivariate-adjusted Cox analyses were used and results were expressed as hazard ratio (HR) and 95% confidence intervals (95% CI). Analyses were performed using Python version 3.09 and STATA version 16.0 (Stata Corp, College Station, TX). All statistical significance levels were 2-sided, and significant differences were expressed as $P < 0.05$.

Ethics Approval

All methods, including the ablation procedure, were carried out in accordance with the guidelines effective at the time of the study. The ethics committee for human research of the University Hospital Centre of Tours (France) approved the study protocol. All patients signed informed consent before inclusion.

Results

A total of 209 patients were included in the study, of whom 103 had paroxysmal AF (49%). Patients were predominantly men (69%), with a mean age of 62 ± 10 years (median 63 [56-69]). Hypertension was found in 49% of patients, and diabetes in 19% and 13% had a CHA2DS2-VASc score >2. Mean ejection fraction was 53 ± 11%, LA volume 133 ± 41 mL and brain natriuretic peptide (BNP) 169 ± 284 ng/mL. The analysis identified 3 patient clusters as shown in the dendrogram (Fig 1). Figure 2 displays the projection of the classification according to the 3 phenogroups identified after reduction of dimension using principal component analysis. Baseline clinical characteristics and procedural characteristics according to the 3 clusters were compared and are reported in Tables 1 and 2.

Clusters

Cluster 1. Cluster 1 was composed of 103 patients, representing 49.3% of the cohort. These patients had a mean age of 61 ± 10 years, were mostly males (63%) and had few comorbidities as reflected by a mean CHA2DS2-VASc score of 1.7 ± 1.41. The phenotype of these patients was characterized by a higher prevalence of paroxysmal AF (90%),

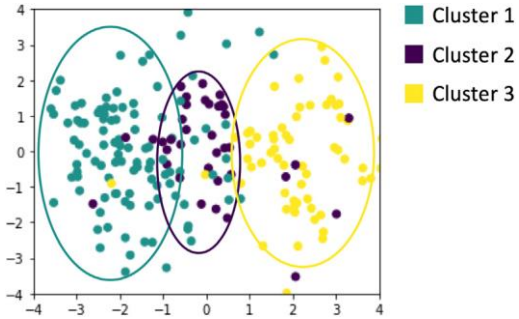


FIG 2. 2D projection of the classification according to the 3 clusters after reduction of dimension using principal component analysis.

higher rate of sinus rhythm at admission (95%) and lower LA volume (116 ± 32mL) than for other clusters (all *P*-values <0.05). Procedure, fluoroscopy and total ablation time were significantly shorter than for other clusters (all *P*-values <0.05) and PVI only was the ablation strategy performed in 76.5% of patients.

Cluster 2. Cluster 2 was composed of 37 (18%) patients. This intermediate cluster included patients of a similar age to Cluster 1 (60 ± 8 years) but with slightly lower rates of comorbidities and fewer symptoms but without reaching statistical significance. However, more patients had persistent AF than in Cluster 1, and had more dilated LA ($P < 0.005$). When considering procedural characteristics, anatomy comprised a lower rate of left common pulmonary vein (2.7%) ($P < 0.05$) than in other clusters and 51.6% of the patients received PVI associated with lines and CFAEs ablation ($P < 0.005$).

Cluster 3. Cluster 3 was composed of 69 (33%) patients and was characterized by older patients (mean age: 65.93 ± 9.96 years, $P < 0.005$) and highest mean CHA_2DS_2-VASc scores (2.81 ± 1.54, $P < 0.05$) reflecting a higher burden of cardiovascular comorbidities than the other clusters. These patients had more often persistent AF, higher weight, more comorbidities such as hypertension and heart failure, more dilated LA, higher level of BNP, galectine-3 and C reactive protein (CRP) and lower estimated glomerular filtration rate (eGFR) than in other clusters ($P < 0.05$). Class I AADs and betablockers were less prescribed ($P < 0.05$) and these patients underwent more complex procedures including PVI associated with lines and CFAEs ablation in 75.8% of cases ($P < 0.0001$).

TABLE 1. Baseline characteristics of patients stratified by clusters

	Cluster 1 (n = 103)	Cluster 2 (n = 37)	Cluster 3 (n = 69)	p 1 vs 2	p 2 vs 3	p 1 vs 3
Age, y	60.88 ± 9.85	60.16 ± 8.42	65.93 ± 9.96	0.69	0.003	0.001
Female sex	38 (36.9)	6 (16.2)	20 (29)	0.02	0.16	0.28
CHA_2DS_2-VASc score	1.7 ± 1.41	1.41 ± 1.14	2.81 ± 1.54	0.26	<0.0001	<0.0001
Paroxysmal AF	93 (90.3)	7 (18.9)	3 (4.4)	<0.0001	<0.0001	0.03
EHRA class	2.90 ± 0.65	2.68 ± 0.63	3.04 ± 0.65	0.07	0.006	0.17
Diagnosis to ablation time, d	2179 ± 2090	2028 ± 1959	1725 ± 2902	0.71	0.58	0.24
Weight, kg	81.6 ± 18.8	81.6 ± 14.1	91.8 ± 22.7	0.99	0.02	0.002
BMI, kg/m2	27.8 ± 5.5	27.1 ± 3.9	30.9 ± 6.6	0.53	0.003	0.001
Hypertension	49 (47.6)	13 (35.1)	41 (59.4)	0.19	0.02	0.13
Diabetes mellitus	14 (13.6)	5 (13.5)	21 (30.4)	0.99	0.06	0.01
Cardiomyopathy	17 (16.5)	2 (5.4)	38 (55.1)	0.09	<0.0001	<0.0001
Heart failure	9 (8.7)	7 (18.9)	39 (56.5)	0.1	0.0002	<0.0001
Mitral regurgitation	44 (42.7)	22 (59.5)	43 (62.3)	0.09	0.84	0.013
Vascular disease	9 (8.7)	2 (5.4)	8 (11.6)	0.52	0.49	0.54
Stroke	8 (7.8)	6 (16.2)	6 (8.7)	0.32	0.42	0.88
SR at admission	98 (95.2)	9 (24.3)	5 (7.3)	<0.0001	0.02	<0.0001
Ejection Fraction, %	57 ± 7	56 ± 7	44 ± 12	0.38	<0.0001	<0.0001
CT/MRI LA Vol., mL	116 ± 32	133 ± 36	159 ± 43	0.01	0.003	<0.0001
eGFR, mL/min/1.73 m^2	76.5 ± 15.7	78.8 ± 23	65.9 ± 24	0.51	0.01	0.001
BNP, ng/mL	81.4 ± 91.7	108.5 ± 87.5	328.6 ± 434.1	0.13	0.003	<0.0001
Galectin-3, ng/mL	13.3 ± 4.2	12.9 ± 3.2	19.9 ± 12.1	0.54	0.001	<0.0001
CRP, mg/L	3.5 ± 3.7	3.1 ± 3.31	6.8 ± 7.5	0.55	0.005	0.0002
Class I AADs	68 (66)	20 (54.1)	17 (24.6)	0.2	0.005	<0.0001
Class III AADs	60 (58.3)	31 (83.8)	58 (84.1)	0.005	1	0.0003
Betablockers	45 (43.7)	10 (27)	13 (18.8)	0.08	0.34	0.001
DOAC *versus* VKA	23 (22.3)	9 (24.3)	16 (23.2)	0.81	1	0.9
ACEi or ARBs	22 (21.4)	5 (13.5)	14 (20.3)	0.3	0.44	0.87

Values are n (%) or mean ± SD.

Abbreviations: AADs, antiarrhythmic drugs; ACEi, angiotensin-converting enzyme inhibitors; ARBs, angiotensin receptor blockers; BMI, body mass index; BNP, brain natriuretic peptide; CRP, C reactive protein; CT, computed tomography; DOAC, direct oral anticoagulant; eGFR, estimated glomerular filtration rate; EHRA, European Heart Rhythm association; LA Vol., left atrium volume; MRI, magnetic resonance imaging; SR, sinus rhythm; VKA, vitamin K antagonist.

TABLE 2. Characteristics of procedures stratified by clusters

	Cluster 1 (n = 103)	Cluster 2 (n = 37)	Cluster 3 (n = 69)	p 1 vs 2	p 2 vs 3	p 1 vs 3
Left common PV	27 (26.2)	1 (2.7)	14 (20.3)	0.001	0.02	0.47
Procedure duration, min	119.1 ± 45.9	178.1 ± 67.6	180.2 ± 47.9	<0.0001	0.87	<0.0001
Fluoroscopy time, min	20.9 ± 10.1	36.6 ± 25.8	33.4 ± 12.2	<0.0001	0.41	<0.0001
Total ablation duration, min	46.1 ± 24.1	64.3 ± 25.3	69.5 ± 19.2	0.003	0.34	<0.0001
Time to LPVs disconnection, min	31.8 ± 18.1	31.2 ± 15.9	26.8 ± 18.4	0.86	0.24	0.09
Time to RPVs disconnection, min	26 ± 15	23.8 ± 12.5	21.7 ± 12.6	0.45	0.44	0.06
Ablation strategy (n = 174)*	(n = 81)	(n = 31)	(n = 62)			
-PVI alone	62 (76.5)	5 (16.1)	2 (3.2)	<0.0001	0.04	<0.0001
-PVI + lines	16 (19.8)	10 (32.3)	13 (21)	0.21	0.31	1
-PVI + lines + CFAEs	3 (3.7)	16 (51.6)	47 (75.8)	<0.0001	0.03	<0.0001

Values are n (%) or mean ± SD.

Abbreviations: CFAEs, complex fractionated atrial electrograms; PVI, pulmonary vein isolation; L/R/PVs, left/right/pulmonary vein.

*Data available for 174 patients.

TABLE 3. Arrhythmia recurrence according to clusters

	Cluster 1 (n = 103)	Cluster 2 (n = 37)	Cluster 3 (n = 69)	p 1 vs 2	p 2 vs 3	p 1 vs 3
Recurrence at 12 months	28 (27.2)	16 (43.2)	42 (60.9)	0.09	0.1	<0.0001
-Paroxysmal AF*	22 (78.6)	3 (18.8)	2 (4.8)	<0.0001	0.12	<0.0001
-Persistent AF*	3 (10.7)	7 (43.8)	27 (64.3)	0.02	0.23	<0.0001
-Atrial flutter*	2 (7.1)	5 (31.3)	13 (31)	0.08	0.12	<0.0001
-Atrial tachycardia*	1 (3.6)	1 (6.3)	0 (0)	1	0.28	0.4
Time to recurrence, day	340 ± 211	241 ± 164	144 ± 139	0.11	0.03	<0.0001
Redo procedures	18 (64.3)	10 (62.5)	32 (76.2)	1	0.34	0.29

Values are n (%) or mean ± SD.
Abbreviation: AF, atrial fibrillation.
*(%) are among patients with recurrence.

Associations With Clinical Outcomes

Table 3 compares the crude rates of recurrence among clusters. Recurrence rate at 1 year was 27.2% for Cluster 1, 43.2% for Cluster 2, and 60.9% for Cluster 3 ($P = 0.09$ for Cluster 1 vs 2, $P = 0.1$ for Cluster 2 vs 3 and $P < 0.0001$ for Cluster 1 vs 3). The type of arrhythmia recurrence was also different among clusters with paroxysmal AF in 78.6% and persistent AF in 10.7% of patients with recurrences in Cluster 1 and persistent AF in 64.3% and atrial flutter in 64.3% for Cluster 3 ($P < 0.0001$). Time to first recurrence was the lowest for Cluster 3 and the longest for Cluster 1 ($P < 0.0001$). Patients among Cluster 3 experienced more redo procedures.

Figure 3 displays Kaplan-Meier curves of AF recurrences within the year following AF ablation.

In the unadjusted Cox regression analysis (Table 4), when compared with Cluster 1, Cluster 2, and Cluster 3 had a significantly higher risk of arrhythmia recurrence. After the adjustments on age, AF type, CHA_2DS_2-VASc score, ejection fraction and LA volume, patients included in Cluster 3 remained at higher risk of recurrence than patients included in Cluster 1 (aHR 0.29, 95% CI, 0.01-0.84, $P = 0.023$).

Discussion

To our knowledge, this is the first exploratory study using cluster analysis on a population of AF patients treated with catheter ablation. The major findings of this study are as follows: (i) cluster analysis led to the identification of 3 statistically-driven groups of AF patients with distinct phenotype characteristics; (ii) clusters significantly varied among the measures of demographic characteristics, blood testing, cardiac imaging,

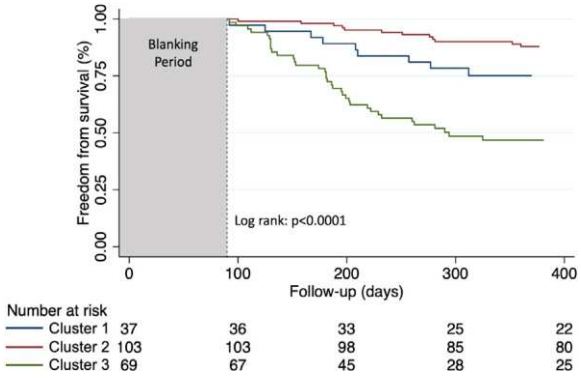


FIG 3. Kaplan-Meier curves for atrial fibrillation recurrences after catheter ablation.

concomitant treatment and procedural characteristics; and (iii) these distinct clusters were independently associated with different probability of arrhythmia recurrences.

Our analysis identified 3 main clinical phenotypes. The first cluster was characterized by young patients with mostly paroxysmal AF, low rate of associated comorbidities and nondilated LA. These patients were treated more frequently with class I AADs and experienced shorter AF ablation procedures mostly composed of PVI alone. This cluster had the best prognosis with the lower rate of arrhythmia recurrence. These recurrences occurred latter than in the other clusters and were mainly composed of paroxysmal AF. The second cluster was composed of patients similar to those in Cluster 1 in terms of comorbidities. Even though Cluster 2 appeared like a subset of borderline patients of Cluster 1 at first glance, these patients were distinguishable by a higher proportion of persistent AF, a higher proportion of men, more dilated LA, and a low rate of left common pulmonary vein anatomy. These patients had an intermediate risk profile for arrhythmia recurrences after AF ablation mainly

TABLE 4. Unadjusted and adjusted Cox regression analysis for arrhythmia recurrence

	Unadjusted analysis			Multivariate analysis		
	HR	95% CI	*P*	aHR*	95% CI	*P*
Arrhythmia recurrence						
Cluster 1 vs 2	0.48	0.26-0.90	0.022	0.78	0.34-1.81	0.57
Cluster 1 vs 3	0.29	0.18-0.47	<0.0001	0.29	0.01-0.84	0.023
Cluster 2 vs 3	0.55	0.31-0.98	0.043	0.67	0.30- 1.49	0.32

*Adjusted HR for age, paroxysmal atrial fibrillation, CHA_2DS_2-VASc score, ejection fraction and left atrial volume.

composed of persistent AF and atrial flutter. For the third cluster, patients were older, had more frequently persistent AF, more comorbidities and more dilated LA than the other clusters. Biomarkers (BNP, CRP, and galectine-3) associated with AF recurrences after ablation were also higher.[14-18] These patients were treated with longer and more complex procedures. The prognosis of patients included in this cluster was worse with more frequent and earlier recurrences mainly comprised of persistent AF and atrial flutter.

Although it is well known as a risk factor of recurrences after ablation, diagnosis to ablation time was not significantly different among the 3 clusters.[3,19,20] Moreover, there was a trend towards shorter duration in Cluster 2 and 3. This highlights the fact that, when considering the natural history of AF, the progression of the disease and symptoms may matter more than only the duration since its onset. This contradicts the rule that "the sooner the better" and underlines the importance of not recusing patients from ablation just because of a long history of AF without assessing the whole clinical complexity.

AF type (ie, paroxysmal, persistent or long standing) is also described as a strong predictor of postablation AF recurrences.[20-23] However, the risk of recurrence is the result of more complex and varied phenomena that cannot be summarized only by the type of AF. This simple classification often overlooks the variety of underlying conditions that are commonly associated with AF and that play an important role in recurrences. Comorbidities such as hypertension, diabetes, obesity, structural heart disease and mitral regurgitation are described as important risk factors contributing to the development of LA dilatation, atrial fibrosis and then atrial cardiomyopathy translating into poorer outcomes with rhythm control strategies.[23-29] Applying cluster analysis helps consider the great heterogeneity of AF ablated patient phenotypes (so-called "clinical complexity"), which is strongly involved in the progressive nature of atrial remodelling, and leads to a new relevant classification. This novel statistical approach defined phenogroups 1 and 3 which were independently associated with AF recurrence even after adjustment on strong and well described predictors such as age, AF type, CHA_2DS_2-VASc score, ejection fraction, and LA volume.[3,8,9,30-35]

Interestingly, cluster analysis can elucidate differences in practice patterns across a population. Indeed, patients included in Cluster 2 and 3 were more often treated with complex ablation procedure including lines and CFAEs ablation probably due to the higher rate of persistent AF. It has since been shown that the choice of such aggressive strategy is ineffective for the treatment of persistent AF when compared to PVI alone.[5]

Clinical Implications and Perspectives

These findings reveal a previously unrecognized heterogeneity of clinical phenotypes and outcomes among AF patients treated with catheter ablation, which may have important clinical implications. There is an important need for more sophisticated risk stratification tools using clinical factors, diverse imaging parameters, biomarkers, and procedural information to identify individuals at risk of recurrence. Different therapeutic strategies, optimal timing of intervention and surveillance might be needed for each cluster, as a step toward precision-medicine in AF. Specifically, a tailored and comprehensive management focusing on higher risk of recurrence clusters with more aggressive modifiable risk factor reduction could improve time to first recurrence, rather than aggressive procedural ablation (including additional lines and CFAEs) as it is often the case.[5,36] In a way, stop the progression of the substrate rather than make it worse.

Cluster analysis is a powerful hypothesis generating approach giving new insights on clinical complexity. This avenue of research should be investigated. Further studies clarifying the interaction between clusters and treatment, such as ablation strategy or lifestyle modification could substantially enhance the clinical implications of phenotype cluster analysis in daily practice and improve outcomes.

Limitations

First, clustering algorithm results are dependent on the underlying population and associated patterns of care in the community. Caution should be taken when generalizing these results from a single centre study. Second, as cluster analysis depends on available data, the incorporation of other variables such as LA substrate might yield different results. Since cluster analysis necessitates complete data on individual patients, we chose to drop variables with a greater than 30% missing data rate to insure the quality of the data. Another caveat is the limited sample size which is likely to result in a lack of statistical power, especially when comparing Cluster 2 to other clusters. This has also been a limitation to the incorporation of a larger number of clusters. Finally, patients were included from 2013 to 2014, and treatment strategies and clinical practice have changed over time (ie, new energy delivery tools such as cryoballoon or pulsed field ablation), partly limiting the generalization of the results to current clinical practice. Nonetheless, our aim was to provide first exploratory data on clusters analysis in patients treated with AF ablation.

Conclusion

Unsupervised cluster analysis led to the identification of statistically-driven phenogroups of AF patients treated with catheter ablation. These results provide new insight into the clinical complexity of AF ablated patients and its influence on outcomes. These data suggest a more holistic and tailored management approach to treat existing comorbidities alongside AF to improve outcomes. Further studies based on high-dimensional databases with large sample size are necessary to confirm these findings and potential future implications in clinical practice.

Data Statement

The data that support the findings of this study are available from the corresponding author, upon reasonable request, and after approval of all other coinvestigators.

Acknowledgment

Dr Arnaud BISSON would like to thank the Société Française de Cardiologie for providing this opportunity, through the Research and fellowship Grant 2020, to work in collaboration with the University of Liverpool.

Supplementary materials

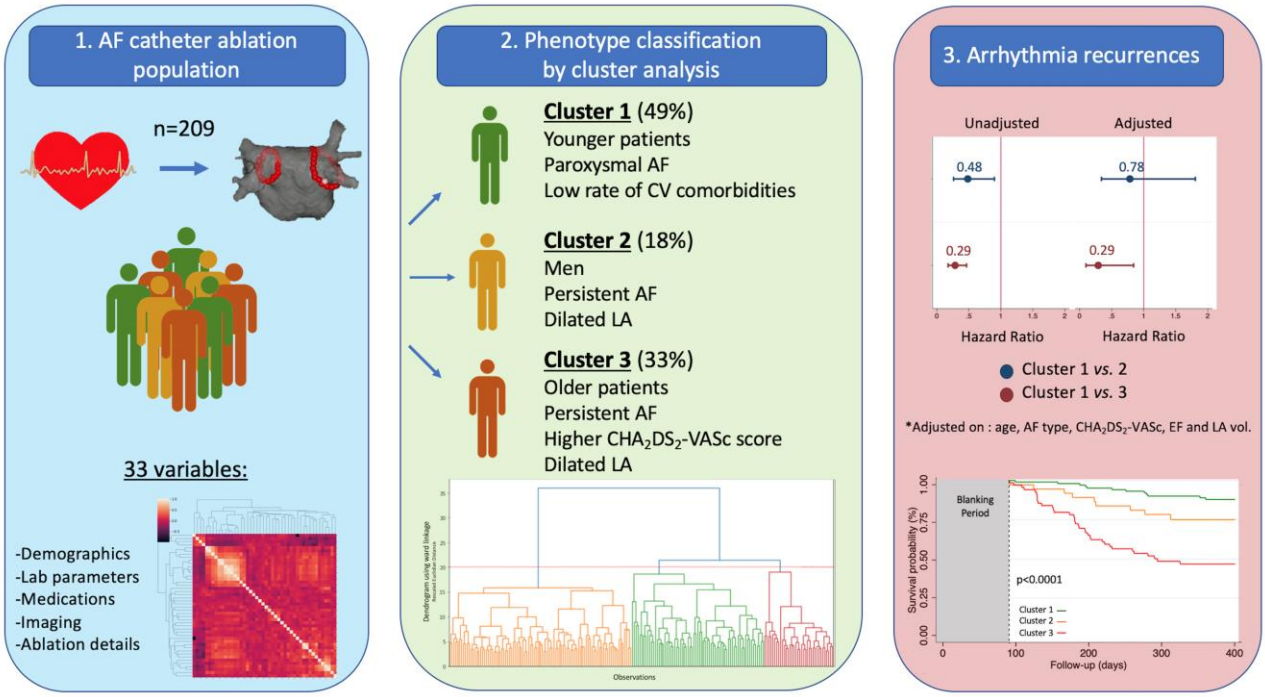
Supplementary material associated with this article can be found in the online version at doi:10.1016/j.cpcardiol.2023.101732.

REFERENCES

1. Hindricks G, Potpara T, Dagres N, et al. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): 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. *Eur Heart J* 2021;42(5):373–498.
2. Calkins H, Hindricks G, Cappato R, et al. 2017 HRS/EHRA/ECAS/APHRS/SOLAECE expert consensus statement on catheter and surgical ablation of atrial fibrillation. *Europace* 2018;20(1):e1–160.
3. Jaïs P, Cauchemez B, Macle L, et al. Catheter ablation versus antiarrhythmic drugs for atrial fibrillation. *Circulation* 2008;118(24):2498–505.

4. Mont L, Bisbal F, Hernández-Madrid A, et al. Catheter ablation vs. antiarrhythmic drug treatment of persistent atrial fibrillation: a multicentre, randomized, controlled trial (SARA study). *Eur Heart J* 2014;35(8):501–7.
5. Verma A, yang Jiang C, Betts TR, et al. Approaches to catheter ablation for persistent atrial fibrillation. *N Engl J Med* 2015;372(19):1812–22.
6. Kuck K-H, Brugada J, Fürnkranz A, et al. Cryoballoon or radiofrequency ablation for paroxysmal atrial fibrillation. *N Engl J Med* 2016;374:2235–45.
7. Poole JE, Bahnson TD, Monahan KH, et al. Recurrence of atrial fibrillation after catheter ablation or antiarrhythmic drug therapy in the CABANA trial. *J Am Coll Cardiol* 2020;75(25):3105–18.
8. Kornej J, Hindricks G, Shoemaker MB, et al. The APPLE score: a novel and simple score for the prediction of rhythm outcomes after catheter ablation of atrial fibrillation. *Clin Res Cardiol* 2015;104(10):871–6.
9. Winkle RA, Jarman JWE, Mead RH, et al. Predicting atrial fibrillation ablation outcome: the CAAP-AF score. *Heart Rhythm* 2016;13(11):2119–25.
10. Kornej J, Schumacher K, Dinov B. Prediction of electro-anatomical substrate and arrhythmia recurrences using APPLE, DR-FLASH and MB-LATER scores in patients with atrial fibrillation undergoing catheter ablation. *Sci Rep* 2018;8(1):12686.
11. Dretzke J, Chuchu N, Agarwal R, et al. Predicting recurrent atrial fibrillation after catheter ablation: a systematic review of prognostic models. *EP Europace* 2020;22(5):748–60.
12. Dey D, Slomka PJ, Leeson P, et al. Artificial intelligence in cardiovascular imaging: JACC state-of-the-art review. *J Am Coll Cardiol* 2019;73(11):1317–35.
13. van Buuren S, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained equations in R. *J Stat Softw* 2011;45(3):1–67.
14. Boyalla V, Harling L, Snell A, et al. Biomarkers as predictors of recurrence of atrial fibrillation post ablation: an updated and expanded systematic review and meta-analysis. *Clin Res Cardiol* 2022;111(6):680–91.
15. Malouf JF, Kanagala R, Al Atawi FO, et al. High sensitivity C-reactive protein: a novel predictor for recurrence of atrial fibrillation after successful cardioversion. *J Am Coll Cardiol* 2005;46(7):1284–7.
16. Clementy N, Benhenda N, Piver E, et al. Serum galectin-3 levels predict recurrences after ablation of atrial fibrillation. *Sci Rep* 2016;6:34357.
17. Clementy N, Garcia B, André C, et al. Galectin-3 level predicts response to ablation and outcomes in patients with persistent atrial fibrillation and systolic heart failure. *PLoS One* 2018;13(8):e0201517.
18. Clementy N, Piver E, Bisson A, et al. Galectin-3 in atrial fibrillation: mechanisms and therapeutic implications. *Int J Mol Sci* 2018;19(4):976.
19. Chew DS, Black-Maier E, Loring Z, et al. Diagnosis-to-ablation time and recurrence of atrial fibrillation following catheter ablation. *Circ Arrhythm Electrophysiol* 2020;13(4):e008128.
20. De Greef Y, Schwagten B, Chierchia GB, de Asmundis C, Stockman D, Buysschaert I. Diagnosis-to-ablation time as a predictor of success: early choice for pulmonary vein isolation and long-term outcome in atrial fibrillation: results from the Middelheim-PVI Registry. *Europace* 2018;20(4):589–95.

21. Ganesan AN, Shipp NJ, Brooks AG, et al. Long-term Outcomes of catheter ablation of atrial fibrillation: a systematic review and meta-analysis. *J Am Heart Assoc* 2013;2 (2):e004549.
22. Uemura T, Kondo H, Sato H, et al. Predictors of outcome after catheter ablation for atrial fibrillation: group analysis categorized by age and type of atrial fibrillation. *Ann Noninvasive Electrocardiol* 2023;28(2):e13020.
23. Wokhlu A, Hodge DO, Monahan KH, et al. Long-term outcome of atrial fibrillation ablation: impact and predictors of very late recurrence. *J Cardiovasc Electrophysiol* 2010;21(10):1071–8.
24. Santoro F, Di Biase L, Trivedi C, et al. Impact of uncontrolled hypertension on atrial fibrillation ablation outcome. *JACC Clin Electrophysiol* 2015;1(3):164–73.
25. Wong CX, Sullivan T, Sun MT, et al. Obesity and the Risk of incident, post-operative, and post-ablation atrial fibrillation: a meta-analysis of 626,603 individuals in 51 studies. *JACC Clin Electrophysiol* 2015;1(3):139–52.
26. Teh AW, Kistler PM, Lee G, et al. Electroanatomic remodeling of the left atrium in paroxysmal and persistent atrial fibrillation patients without structural heart disease. *J Cardiovasc Electrophysiol* 2012;23(3):232–8.
27. Marrouche NF, Wilber D, Hindricks G, et al. Association of atrial tissue fibrosis identified by delayed enhancement MRI and atrial fibrillation catheter ablation: the DECAAF study. *JAMA* 2014;311(5):498–506.
28. Guichard JB, Nattel S. Atrial cardiomyopathy: a useful notion in cardiac disease management or a passing fad? *J Am Coll Cardiol* 2017;70(6):756–65.
29. Kreimer F, Gotzmann M. Left atrial cardiomyopathy—a challenging diagnosis. *Front Cardiovasc Med* 2022;9:942385.
30. D'Ascenzo F, Corleto A, Biondi-Zoccai G, et al. Which are the most reliable predictors of recurrence of atrial fibrillation after transcatheter ablation?: a meta-analysis. *Int J Cardiol* 2013;167(5):1984–9.
31. Berruezo A, Tamborero D, Mont L, et al. Pre-procedural predictors of atrial fibrillation recurrence after circumferential pulmonary vein ablation. *Eur Heart J* 2007;28(7):836–41.
32. Nedios S, Kosiuk J, Koutalas E, et al. Comparison of left atrial dimensions in CT and echocardiography as predictors of long-term success after catheter ablation of atrial fibrillation. *J Interv Card Electrophysiol* 2015;43(3):237–44.
33. Kornej J, Hindricks G, Kosiuk J, et al. Comparison of CHADS2, R2CHADS2, and CHA2DS2-VASc scores for the prediction of rhythm outcomes after catheter ablation of atrial fibrillation. *Circ Arrhythm Electrophysiol* 2014;7(2):281–7.
34. Costa FM, Ferreira AM, Oliveira S, et al. Left atrial volume is more important than the type of atrial fibrillation in predicting the long-term success of catheter ablation. *Int J Cardiol* 2015;184:56–61.
35. Njoku A, Kannabhiran M, Arora R, et al. Left atrial volume predicts atrial fibrillation recurrence after radiofrequency ablation: a meta-analysis. *Europace* 2018;20(1):33–42.
36. Pathak RK, Middeldorp ME, Lau DH, et al. Aggressive risk factor reduction study for atrial fibrillation and implications for the outcome of ablation: the ARREST-AF cohort study. *J Am Coll Cardiol* 2014;64(21):2222–31.



Graphical abstract.

2.2 Étude 5 : Phénotypes et pronostic chez les patients non-anticoagulés atteints de fibrillation atriale : Une analyse non supervisée en clusters

Bisson A**, Fawzy AM, Romiti GF, Proietti M, Angoulvant D, El-Bouri W, Y H Lip G, Fauchier L.* ***Phenotypes and outcomes in non-anticoagulated patients with atrial fibrillation: An unsupervised cluster analysis. *Arch Cardiovasc Dis. 2023 Jun-Jul;116(6-7):342-351. doi: 10.1016/j.acvd.2023.06.001*. [117]

1. Objectif de l'étude

L'objectif de ce dernier travail était d'explorer et de décrire une nouvelle classification de la complexité clinique de patients atteints de FA à l'aide d'une analyse en clusters et d'évaluer l'association entre cette classification et le pronostic.

2. Méthode

A partir d'une cohorte régionale de patients atteints de FA (The Loire Valley Atrial Fibrillation Cohort), nous avons identifié 3,434 patients non-anticoagulés. Après une étape de preprocessing, 25 variables ont été analysées par un algorithme non supervisé de machine learning de type agglomerative hierarchical clustering. L'association entre la classification en clusters obtenue et la survenue d'événements majeurs tels que le décès, l'AVCi et les épisodes hémorragiques majeurs a ensuite été évaluée.

3. Résultats

Au sein de cette population dont l'âge moyen était de 70,3±17 ans dont 42,8 % étaient des femmes, trois clusters ont été identifiés : le Cluster 1 composé de patients plus jeunes, avec une faible prévalence de comorbidités ; le Cluster 2 comprenant des patients âgés souffrant de FA permanente, de pathologies cardiaques et d'un nombre élevé de comorbidités cardiovasculaires ; le Cluster 3 identifiant des patientes âgées avec un nombre élevé de comorbidités cardiovasculaires. Par rapport au Cluster 1, les Clusters 2 et 3 étaient indépendamment associés à un risque accru de survenue d'un critère composite comprenant AVCi/embole systémique/décès (HR 2,85, IC 95 % 1,32-6,16 et HR 1,52, IC 95 % 1,09-2,11, respectivement) et de décès toutes causes confondues (HR 3,54, IC 95 % 1,49-8,43 et HR 1,88, IC 95 % 1,26-2,79, respectivement). Le Cluster 3 était indépendamment associé à un risque accru d'hémorragie majeure (HR 1,72, IC 95 % 1,06-2,78).

4. Conclusion

Dans cette population de patients non anticoagulés atteints de FA, l'analyse non supervisée en clusters a permis d'identifier trois phénogroupes ayant des caractéristiques distinctes. Ces groupes étaient associés à des risques différents d'événements cliniques indésirables majeurs. Ils soulignent la nécessité d'une approche plus holistique des patients pour traiter les comorbidités existantes associées à la FA, afin de minimiser le risque d'évènements indésirables.

Archives of Cardiovascular Disease 116 (2023) 342–351

Disponible en ligne sur ScienceDirect www.sciencedirect.com

Elsevier Masson France EM|consulte www.em-consulte.com

Clinical Research

Phenotypes and outcomes in non-anticoagulated patients with atrial fibrillation: An unsupervised cluster analysis☆

Arnaud Bisson[a,b,c,d,*], Ameenathul M. Fawzy[d], Giulio Francesco Romiti[d,e], Marco Proietti[d,f,g], Denis Angoulvant[a,c,1], Wahbi El-Bouri[d], Gregory Y. H. Lip[d,h,1], Laurent Fauchier[a,1]

[a] Service de cardiologie, centre hospitalier régional universitaire et faculté de médecine de Tours, 37000 Tours, France
[b] Service de cardiologie, centre hospitalier régional universitaire d'Orléans, 45100 Orléans, France
[c] EA4245, transplantation immunité inflammation, université de Tours, 37032 Tours, France
[d] Liverpool Centre for Cardiovascular Science at University of Liverpool, Liverpool John Moores University and Liverpool Heart & Chest Hospital, L7 8TX Liverpool, United Kingdom
[e] Department of Translational and Precision Medicine, Sapienza – University of Rome, 00185 Rome, Italy
[f] Department of Clinical Sciences and Community Health, University of Milan, 20122 Milan, Italy
[g] IRCCS Istituti Clinici Scientifici Maugeri, 20138 Milan, Italy
[h] Department of Clinical Medicine, Aalborg University, 9000 Aalborg, Denmark

ARTICLE INFO

Article history:
Received 23 February 2023
Received in revised form 1st June 2023
Accepted 5 June 2023
Available online 24 June 2023

Keywords:
Atrial fibrillation
Cluster analysis
Outcomes
Machine learning

ABSTRACT

Background: Patients with atrial fibrillation are characterized by great clinical heterogeneity and complexity. The usual classifications may not adequately characterize this population. Data-driven cluster analysis reveals different possible patient classifications.
Aims: To identify different clusters of patients with atrial fibrillation who share similar clinical phenotypes, and to evaluate the association between identified clusters and clinical outcomes, using cluster analysis.
Methods: An agglomerative hierarchical cluster analysis was performed in non-anticoagulated patients from the Loire Valley Atrial Fibrillation cohort. Associations between clusters and a composite outcome comprising stroke/systemic embolism/death and all-cause death, stroke and major bleeding were evaluated using Cox regression analyses.
Results: The study included 3434 non-anticoagulated patients with atrial fibrillation (mean age 70.3 ± 17 years; 42.8% female). Three clusters were identified: cluster 1 was composed of younger patients, with a low prevalence of co-morbidities; cluster 2 included old patients with permanent atrial fibrillation, cardiac pathologies and a high burden of cardiovascular co-morbidities; cluster 3 identified old female patients with a high burden of cardiovascular co-morbidities. Compared with cluster 1, clusters 2 and 3 were independently associated with an increased risk of the composite outcome (hazard ratio 2.85, 95% confidence interval 1.32–6.16 and hazard ratio 1.52, 95% confidence interval 1.09–2.11, respectively) and all-cause death (hazard ratio 3.54, 95% confidence interval 1.49–8.43 and hazard ratio 1.88, 95% confidence interval 1.26–2.79, respectively). Cluster 3 was independently associated with an increased risk of major bleeding (hazard ratio 1.72, 95% confidence interval 1.06–2.78).
Conclusion: Cluster analysis identified three statistically driven groups of patients with atrial fibrillation, with distinct phenotype characteristics and associated with different risks for major clinical adverse events.

© 2023 Elsevier Masson SAS. All rights reserved.

☆ Tweet: Prognosis of patients with AF depends on burden of concomitant co-morbidities. Cluster analysis enables description of this so-called clinical complexity and identified three clinically distinct phenogroups associated with different risks for major clinical adverse events (Twitter handle: @rythmochutours).

* Corresponding author. Service de cardiologie, centre hospitalier regional universitaire et faculté de médecine de Tours, 2, boulevard Tonnellé, 37000 Tours, France.
E-mail address: arnaud.bisson@univ-tours.fr (A. Bisson).

[1] Joint senior authors.

https://doi.org/10.1016/j.acvd.2023.06.001
1875-2136/© 2023 Elsevier Masson SAS. All rights reserved.

1. Abbreviations

ABC	atrial fibrillation better care
AF	atrial fibrillation
aHR	adjusted hazard ratio
CI	confidence interval
HR	hazard ratio

2. Introduction

Atrial fibrillation (AF) is the most common cardiac arrhythmia worldwide and is still associated with high mortality and morbidity from stroke and thromboembolism [1]. The variety of underlying processes and disorders involved in AF and its prognosis might not be optimally characterized with simplified risk scores, which tend to use risk factors in a binary manner (yes/no), despite many risk factors occurring in combination with each other [2]. So-called clinical complexity, which reflects the burden of concomitant comorbidities and conditions that increase both thromboembolic and bleeding risks, and hence death, may influence treatment choices, and poses significant challenges in the management of AF [3].

Machine learning is emerging as a new method for improving the understanding of and the prediction of adverse outcomes in the field of cardiovascular disease, including AF [4]. In particular, unsupervised cluster analysis categorizes the complex entities without investigator supervision, by segregating samples into homogenous groups based on each cluster's dissimilarities [5]. This data-driven approach helps to unveil meaningful phenotypes within a disease that has been previously considered as homogenous. Thus, cluster analysis may be invaluable in phenotyping several cardiovascular diseases, and exploring the clinical complexity of patients with cardiovascular disease. A few cluster analyses have been performed in patients with AF [2,6–8], but these studies were derived from randomized controlled trials and registries, including a majority of patients with AF already taking oral anticoagulant therapy. We are unaware of any previous cluster analysis in a non-anticoagulated AF cohort, hence providing a clear picture of the "natural history" of clinical outcomes, uninfluenced by oral anticoagulation use.

This study therefore aimed to explore, for the first time, whether unsupervised cluster analysis could identify clinically relevant groups among non-anticoagulated patients with AF, and to evaluate whether these cluster phenotypes are associated with different treatment patterns and outcomes in a real-world population.

3. Methods

3.1. Study population

The detailed methods and design of the Loire Valley AF Project have been reported previously [9–12]. We included all patients with a diagnosis of AF seen in the Cardiology Department at the University Hospital of Tours between January 2000 and December 2010. Treatments at discharge and during follow-up were obtained by screening hospitalization reports. Patients were excluded from the study if their antithrombotic treatment was unknown. Patient characteristics were obtained from the institution's computerized system, based on the International Classification of Diseases, 10th revision. Using these characteristics, the CHA_2DS_2-VASc and HAS-BLED scores were calculated retrospectively [13,14] (Table A.1 and Table A.2).

In our analysis, we included only patients with non-valvular AF, not treated with anticoagulant therapy (vitamin K antagonists or heparin) at discharge and during follow-up (Fig. 1).

3.2. Follow-up and outcomes

During follow-up, deaths from all causes, strokes and major bleeding were recorded whenever they occurred in our institution, which includes four hospitals covering all medical and surgical specialties. Our hospital covers an area of 4000 km^2 and a population of 400,000 inhabitants, and is the only public institution in the area. In addition, mortality data were obtained using a search tool from the website of the main regional newspaper of the Région Centre, La Nouvelle République (http://nrco.lanouvellerepublique.fr/dossiers/necro/index.php). In this "real-world" registry, some deaths may have resulted from undiagnosed stroke, because cerebral imaging and post-mortem examinations were not mandated, unlike in a clinical trial. The primary endpoint for this study was the composite of "stroke/systemic embolism/death." Severe bleeding was defined as a decrease in haemoglobin concentration of > 5 g/dL, the need for transfusion with one or more units of blood, the need for corrective surgery, the occurrence of an intracranial or retroperitoneal haemorrhage, or any combination of these events.

3.3. Cluster analysis

The whole cohort was used for analysis. Variables with a missing data rate above 25% were excluded, and variables with a missing data rate $\leq$ 25% were imputed using the "Multivariate Imputation By Chained Equations" algorithm [15]. Twenty-five baseline clinical variables were therefore used for the analysis (Table A.3).

The hierarchical clustering method (using Ward's linkage criterion) was used to identify homogenous phenotypic subgroups of patients with AF without prior knowledge of the outcomes. The algorithm begins with each element (i.e. patient) as a separate cluster, and then proceeds with a "bottom-up" approach, grouping each cluster with the most similar one, until all clusters become one.

A dendrogram is provided to display the distance obtained at each interaction of the clustering process (Fig. 2). Small values of the distance indicate that the merged clusters are similar, and large values indicate the combination of two dissimilar (heterogeneous) clusters. The determination of the numbers of clusters was not prespecified. Examination of the dendrogram indicated that the groupings became more heterogeneous after being expanded to three clusters. Both three-cluster and four-cluster models were examined. The three-cluster model formed much clearer patterns of patient groups than the four-cluster model and was therefore used in this study. Cluster models were implemented in Python using open-source packages: Scikit-learn, version 1.1.1 [16]. Once clusters were identified, we assessed the association between clusters and clinical characteristics and outcomes.

3.4. Statistical analysis

Qualitative variables are described using counts and percentages, and continuous quantitative variables are described as mean ± standard deviation or median (interquartile range). The Kolmogorov–Smirnov test was used to assess the distribution of continuous variables. Comparisons were made using parametric or non-parametric tests, as appropriate: The Wilcoxon signed-rank and Kruskal-Wallis tests were used to compare continuous values between two independent groups, and the χ^2 test was used to compare categorical data. Kaplan-Meier curves were plotted with the duration from the enrolment to the last follow-up or death. Unadjusted and multivariable-adjusted Cox analyses were used to estimate the association between clusters and clinical outcomes, and results are expressed as hazard ratios (HRs) and 95% confidence intervals (CIs). The whole cohort was used for this analysis. Two multivariable models were built, adjusting for well-known confounders and clinical scores used as good integra-

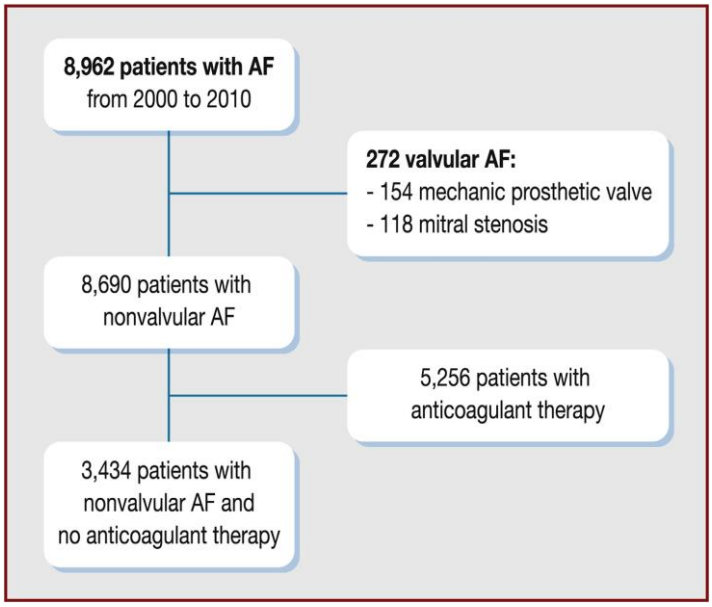


Fig. 1. Flow chart of study. AF: atrial fibrillation.

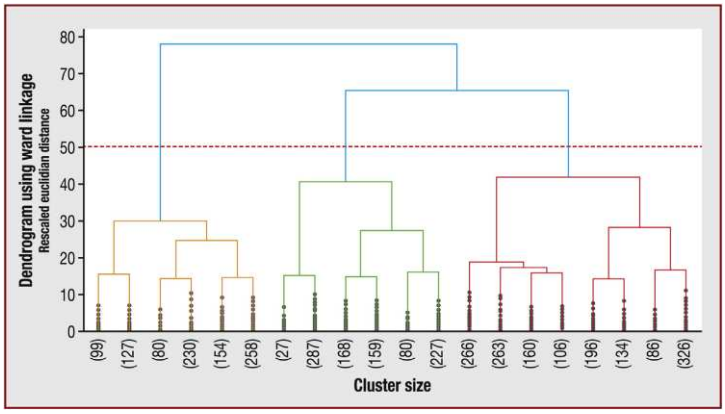


Fig. 2. Dendrogram generated by hierarchical clustering, showing three main clusters. The dendrogram represents the relative degree of similarity between individual patients. The greater the height of the branch points (distance), the greater the differences are between the branches, and the more dissimilarity exists between clusters. The red line indicates the stopping location.

tors of multimorbid state: for assessing the composite outcome, all-cause death and stroke, model 1 was adjusted for age, sex and type of AF; and model 2 was adjusted for CHA_2DS_2-VASc score. To evaluate the association between clusters and major bleeding events, model 2 was adjusted for the HAS-BLED score instead of the CHA_2DS_2-VASc score, whereas model 1 was adjusted for the same covariates as reported above (i.e. age, sex and type of AF).

Analyses were performed using Python, version 3.09, and STATA, version 16.0 (StataCorp, College Station, TX, USA). Statistical significance levels were two-sided, and significant differences were expressed as $P < 0.05$. When multiple comparisons where performed, Bonferroni's adjustment of α risk was applied.

3.5. Ethics approval

The local ethical committee of our institution was consulted, and approved this study on 07 December 2010, and it was registered as a clinical audit. The informed consent of patients was deemed unnecessary, as this was a retrospective analysis, and there was no impact on their care.

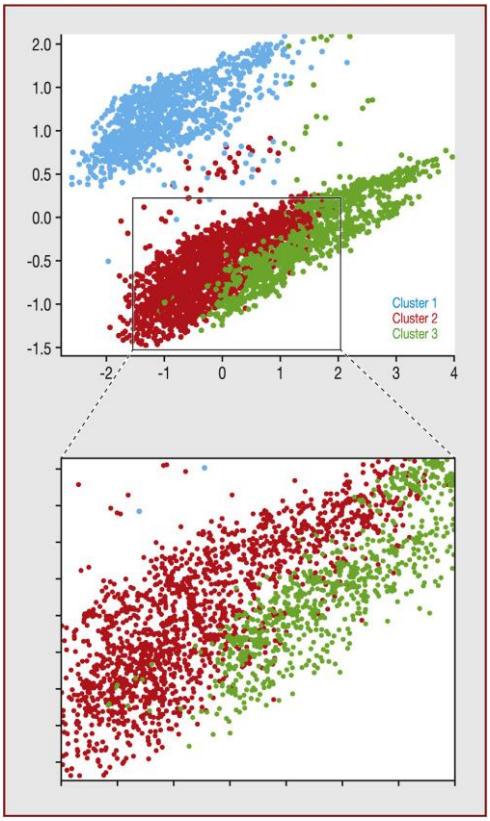


Fig. 3. Projection of the classification according to the three clusters after reduction of dimension using principal component analysis (from 25 to two dimensions).

4. Results

Among 8962 patients diagnosed with AF in our institution between 2000 and 2010, 8690 (97%) had non-valvular AF, of whom 3434 (38.3%) were not anticoagulated (Fig. 1). The analysis identified three patient clusters, as shown in the dendrogram (Fig. 2). Fig. 3 displays the projection of the classification, according to the three phenogroups identified after reduction of dimension using principal component analysis. Cluster 1 is clearly distinct from the two others and, despite their nearness, there is only a minor overlap between clusters 2 and 3. Baseline clinical characteristics according to the three clusters were compared, and are reported in Table 1.

4.1. Clusters

4.1.1. Cluster 1

Cluster 1 was composed of 949 patients, representing 27.6% of the cohort. The phenotype of the patients was characterized as being younger (mean age 51 ± 17 years), with male predominance (69%), low prevalence of permanent AF (2.5%), low CHA_2DS_2-VASc score (mean 1.47 ± 1.44), low prevalence of co-morbidities and higher rates of tobacco smoking (19.7%) ($P = 0.21$ vs. cluster 2 and $P = 0.001$ vs. cluster 3), dyslipidaemia (21.0%) ($P < 0.0001$ vs. cluster 2 and 3) and alcohol-related diagnoses (4.7%) ($P = 0.41$ vs. cluster 2 and $P = 0.05$ vs. cluster 3) compared with the other clusters. These patients were more often treated with antiplatelet agents (66.6%), beta-blockers (50.0%) and class I antiarrhythmic drugs (44.2%) ($P < 0.0001$ for all).

4.1.2. Cluster 2

Cluster 2 was composed of 948 (27.6%) patients. Patients in this cluster were older (mean age 78.52 ± 11.4 years) compared with the other two clusters ($P < 0.0001$ vs. cluster 1 and $P = 0.003$ vs. cluster 3), mainly male (61.7%) and almost exclusively in permanent AF (97.9%). These patients were characterized by a higher prevalence of cardiac pathologies, such as heart failure (63.4%), valve disease (20.1%) and pacemaker or implantable cardioverter-defibrillator implantations (19.9%) compared with the other two clusters ($P < 0.05$ for all). With regard to treatments, we observed a higher use of angiotensin-converting enzyme inhibitors (ACEi)/angiotensin receptor blockers (ARBs) (39.1%) and digoxin (35.3%) ($P = 0.001$ vs. cluster 1 and $P = 0.18$ vs. cluster 3 for ACEi/ARBs, and $P < 0.001$ vs. cluster 1 and 3 for digoxin).

4.1.3. Cluster 3

Cluster 3 was composed of 1537 (44.8%) patients, characterized by a mean age of 77.19 ± 10.3 years, a female majority (52.7%;

Table 1
Baseline characteristics of patients, stratified by cluster.

	Cluster 1	Cluster 2	Cluster 3	*P* values[a]		
	(*n* = 949; 27.6%)	(*n* = 948; 27.6%)	(*n* = 1537; 44.8%)	1 vs. 2	1 vs. 3	2 vs. 3
Age (years)	51 ± 16.89	78.52 ± 11.38	77.19 ± 10.25	< 0.0001	< 0.0001	0.003
Female sex	297 (31.3)	363 (38.3)	810 (52.7)	0.001	< 0.0001	< 0.0001
Permanent AF	24 (2.5)	928 (97.9)	0 (0)	< 0.0001	-	-
CHA_2DS_2-VASc score	1.47 ± 1.44	3.75 ± 1.61	3.85 ± 1.61	< 0.0001	< 0.0001	0.14
Low (0 for men, 1 for women)	461 (48.6)	30 (3.2)	63 (4.1)	< 0.0001	< 0.0001	0.28
Intermediate (1 for men, 2 for women)	167 (17.6)	60 (6.3)	110 (7.2)	< 0.0001	< 0.0001	0.17
High (> 1 for men, > 2 for women)	321 (33.8)	858 (90.5)	1364 (88.7)	< 0.0001	< 0.0001	0.17
HAS-BLED score	0.88 ± 1.16	1.89 ± 1.04	1.92 ± 1.07	< 0.0001	< 0.0001	0.46
Score ≥ 3	118 (12.4)	240 (25.3)	405 (26.4)	< 0.0001	< 0.0001	0.57
Co-morbidities						
Hypertension	179 (18.9)	442 (46.6)	804 (52.3)	< 0.0001	< 0.0001	0.01
Diabetes mellitus	94 (9.9)	164 (17.3)	239 (15.5)	< 0.0001	< 0.0001	0.25
Heart failure	198 (20.9)	601 (63.4)	849 (55.2)	0.003	< 0.0001	0.0001
Valve disease	58 (6.1)	191 (20.1)	258 (16.8)	< 0.0001	< 0.0001	0.03
Coronary artery disease	259 (27.3)	319 (33.6)	542 (35.3)	< 0.0001	< 0.0001	0.41
Vascular disease	275 (29.0)	371 (39.1)	601 (39.1)	0.01	0.003	0.99
Previous pacemaker/ICD	51 (5.4)	189 (19.9)	221 (14.4)	0.0001	0.001	0.0003
Ischaemic stroke	47 (5.0)	75 (7.9)	123 (8.0)	0.0001	< 0.0001	0.94
Previous major bleeding	32 (3.4)	72 (7.6)	99 (6.4)	0.01	0.84	0.27
Smoker	187 (19.7)	123 (13.0)	159 (10.3)	0.21	0.001	0.04
Dyslipidaemia	199 (21.0)	153 (16.1)	317 (20.6)	< 0.0001	< 0.0001	0.01
Alcohol-related diagnoses	45 (4.7)	34 (3.6)	35 (2.3)	0.41	0.005	0.05
Renal insufficiency	31 (3.3)	100 (10.5)	185 (12.0)	0.25	0.73	0.26
Dialysis	2 (0.2)	4 (0.4)	20 (1.3)	0.16	0.06	0.03
Liver disease	4 (0.4)	8 (0.8)	8 (0.5)	< 0.0001	< 0.0001	0.33
Thyroid disease	26 (2.7)	37 (3.9)	64 (4.2)	< 0.0001	0.0001	0.0001
Laboratory						
eGFR (mL/min/1.73 m^2)	80 ± 45.96	57.89 ± 19.64	56.86 ± 19.12	< 0.0001	< 0.0001	0.24
Medical therapies						
NSAID	1 (0.1)	0 (0)	5 (0.3)	0.32	0.28	0.08
Antiplatelet agent	624 (66.6)	508 (54.7)	755 (50.3)	< 0.0001	< 0.0001	0.03
$P2Y_{12}$ inhibitor	159 (17.1)	120 (13.0)	219 (14.7)	0.01	0.12	0.24
ACE inhibitor or ARB	98 (28.5)	197 (39.1)	273 (35.4)	0.001	0.02	0.18
Diuretic	73 (21.2)	294 (58.3)	315 (40.9)	< 0.0001	< 0.0001	< 0.0001
Beta-blocker	172 (50.0)	164 (32.5)	284 (36.8)	< 0.0001	< 0.0001	0.12
Digoxin	74 (21.5)	178 (35.3)	144 (18.7)	< 0.0001	0.27	< 0.0001
Class I AAD	92 (44.2)	11 (4.8)	70 (19.0)	< 0.0001	< 0.0001	< 0.0001
Class III AAD	165 (36.0)	126 (23.9)	394 (45.2)	< 0.0001	0.001	< 0.0001
Calcium channel blocker	25 (12.0)	55 (23.8)	83 (22.6)	0.001	0.002	0.72

Data are expressed as mean ± standard deviation or number (%). AAD: antiarrhythmic drug; ACE: angiotensin-converting enzyme; AF: atrial fibrillation; ARB: angiotensin receptor blocker; eGFR: estimated glomerular filtration rate; ICD: implantable cardioverter-defibrillator; NSAID: non-steroidal anti-inflammatory drug.
[a] $P < 0.0167$ was considered as significant after Bonferroni's adjustment.

$P < 0.0001$) and the highest mean CHA_2DS_2-VASc and HAS-BLED scores (3.85 ± 1.61 [$P < 0.0001$ vs. cluster 1 and $P = 0.14$ vs. cluster 2] and 1.92 ± 1.07 [$P < 0.0001$ vs. cluster 1 and $P = 0.46$ vs. cluster 2], respectively), reflecting a higher burden of cardiovascular co-morbidities than the other clusters. These patients were more often treated with class III antiarrhythmic drugs (45.2%; $P \leq 0.001$).

4.2. Associations with clinical outcomes

Table 2 compares the crude rates and annual risks of major adverse events between clusters. Over a mean follow-up of 812 ± 1034 days (median 299 days, interquartile range [7 – 1384] days), the occurrence of the composite outcome and all-cause death was higher in cluster 2 (27.3% and 23.8%) compared with cluster 1 (6.9% and 4.5%) and cluster 3 (22.8% and 18.4%) (all $P < 0.01$). The occurrence of stroke was higher in cluster 3 (7.9%) than in cluster 1 (3.0%) and cluster 2 (7.5%) ($P < 0.001$ and $P = 0.90$, respectively), whereas the occurrence of bleeding was higher in cluster 2 (8.0%) than in cluster 1 (3.9%) and cluster 3 (7.7%) ($P < 0.001$ and $P = 0.38$, respectively).

The occurrence of major adverse events across the different clusters is illustrated in the Kaplan-Meier curves in Fig. 4, and was significantly different across the three clusters for the primary composite outcome and all-cause death, and a significantly better prognosis was observed for cluster 1 patients in terms of stroke and major bleeding. Kaplan-Meier curves for cardiovascular and non-cardiovascular death are displayed in Fig. A.1.

In the unadjusted Cox regression analysis (Table 3), when compared with cluster 1, cluster 2 and cluster 3 had a significantly higher risk of the composite outcome, all-cause death, stroke and major bleeding. After the adjustments according to multivariable models (model 1 adjusted for the CHA_2DS_2-VASc score for the composite outcome, all-cause death and stroke, and the HAS-BLED score for major bleeding; model 2 adjusted for age, sex and type of AF), cluster 2 and cluster 3 were independently associated with an increased risk of the primary composite outcome (adjusted HRs [aHRs] between 1.52 and 2.85) and all-cause death (aHRs between 1.88 and 3.85) compared with cluster 1 (Table 3). When considering stroke, clusters 2 and 3 were not independently associated with an increased risk in model 1 or model 2, with only a trend seen for cluster 2 (aHR 1.63, 95% CI 0.99–2.66). Cluster 3 was independently associated with major bleeding in model 1, whereas both cluster 2 and cluster 3 were independently associated with major bleeding in model 2 (aHRs between 1.72 and 2.06).

5. Discussion

To our knowledge, this is the first exploratory study using cluster analysis on exclusively non-anticoagulated patients with AF. The principal findings of our study are as follows:

Table 2
Major adverse events, stratified by cluster.

	Cluster 1 (n = 949)		Cluster 2 (n = 948)		Cluster 3 (n = 1537)		P values[a]		
	n (%)	% pt-years (95% CI)	n (%)	% pt-years (95% CI)	n (%)	% pt-years (95% CI)	1 vs. 2	1 vs. 3	2 vs. 3
Composite outcome	65 (6.9)	2.86 (2.24–3.65)	259 (27.3)	14.21 (12.43–15.88)	350 (22.8)	11 (9.72–12.02)	< 0.001	< 0.001	0.002
All-cause death	43 (4.5)	1.82 (1.35–2.45)	226 (23.8)	11.81 (10.17–13.23)	282 (18.4)	8.38 (7.29–9.23)	< 0.001	< 0.001	0.001
Stroke	28 (3.0)	1.23 (0.85–1.78)	71 (7.5)	3.87 (3.07–4.88)	122 (7.9)	3.8 (3.18–4.54)	< 0.001	< 0.001	0.90
Major bleeding	37 (3.9)	1.61 (1.16–2.22)	76 (8.0)	4.21 (3.37–5.28)	118 (7.7)	3.71 (3.09–4.44)	< 0.001	< 0.001	0.38

CI: confidence interval; % pt-years: % patient-years.
[a] P < 0.0167 was considered as significant after Bonferroni's adjustment.

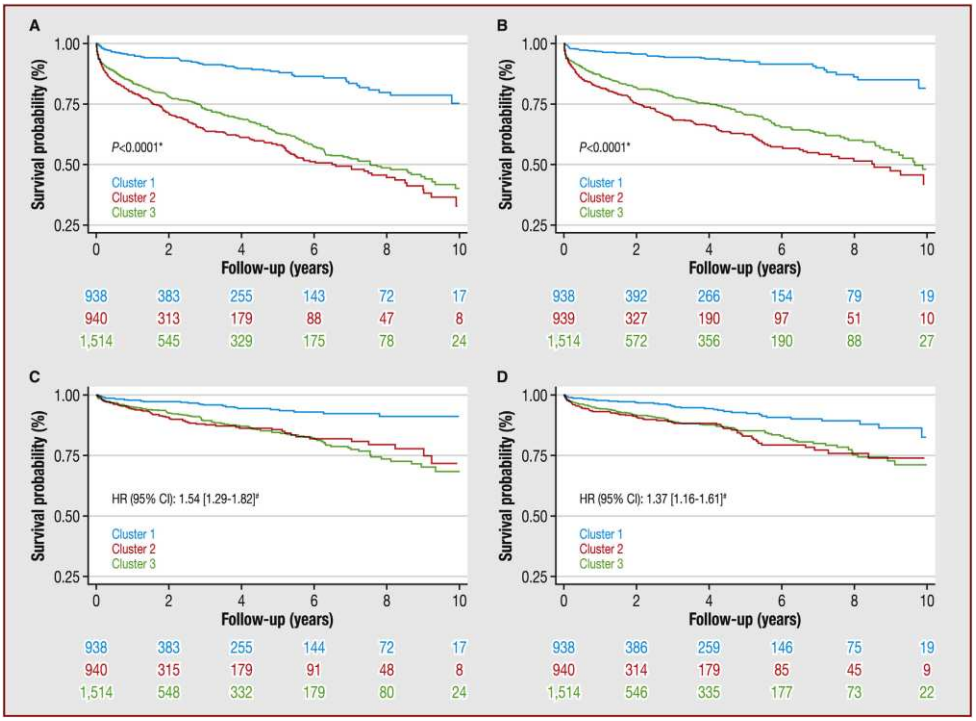


Fig. 4. A–D. Kaplan-Meier curves for the composite outcome of stroke, thromboembolism and all-cause death (A), all-cause death (B), stroke (C) and major bleeding (D). CI: confidence interval; HR: hazard ratio. [a] Log-rank and univariate Cox analysis P values. [b] Univariate Cox analysis.

- cluster analysis led to the identification of three statistically driven groups of patients with AF, with distinct phenotype characteristics
- the clusters varied significantly in terms of demographic characteristics, cardiovascular risk factors, co-morbidities and concomitant treatment;
- these distinct clusters were independently associated with different risks for all-cause death and major clinical adverse events.

Previous cluster analyses have been performed in patients with AF [2,6–8], but included a majority of patients already taking oral anticoagulant therapy, which may affect the clinical outcomes of interest, including stroke, death and major bleeding. The major challenge is to find a contemporary AF cohort that is non-anticoagulated, to provide a clear picture of the "natural history" of clinical outcomes, uninfluenced by anticoagulation use. However, prescription trends have evolved considerably since the 2000s. Patients with AF, during the vitamin K antagonist period, were much more undertreated than now, with only two thirds receiving anticoagulation [17,18]. This rate of anticoagulation was even lower in patients with AF and co-morbidities [19,20].

Using hierarchical cluster analysis applied to this large cohort of real-world non-anticoagulated patients with AF, we were able to find patterns of association, which allowed a novel categorization of patients with AF. Cluster 1 was characterized by younger patients with low co-morbidities, but a higher rate of behavioural disorders. Not surprisingly, this cluster had the lowest rates of adverse outcomes. Two thirds of these patients were prescribed antiplatelet agents, despite the low prevalence of coronary artery disease and vascular disease. This can be explained by the stroke risk stratification guidelines used during this period, which were based on the older $CHADS_2$ score until 2012, and not on the CHA_2DS_2-VASc score [21–23]. The mean $CHADS_2$ score in this cohort was actually

Table 3
Unadjusted and adjusted Cox regression analysis for major adverse events.

	Unadjusted analysis			Multivariable analysis: model 1[a]			Multivariable analysis: model 2[b]		
	HR	95% CI	*P*	HR	95% CI	*P*	HR	95% CI	*P*
Composite outcome									
Cluster 2 vs. 1	4.46	3.40–5.87	<0.0001	2.85	1.32–6.16	0.008	2.83	2.09–3.82	<0.0001
Cluster 3 vs. 1	3.57	2.74– 4.66	<0.0001	1.52	1.09–2.11	0.013	2.12	1.57–2.86	<0.0001
Cluster 2 vs. 3	1.26	1.07–1.48	0.005	2.28	1.13– 4.63	0.022	1.34	1.14–1.58	<0.0001
All-cause death									
Cluster 2 vs. 1	5.76	4.15–7.99	<0.0001	3.54	1.49–8.43	0.004	3.85	2.70–5.48	<0.0001
Cluster 3 vs. 1	4.24	3.08–5.85	<0.0001	1.88	1.26–2.79	0.002	2.90	2.03–4.15	<0.0001
Cluster 2 vs. 3	1.37	1.15–1.63	0.001	2.86	1.41–5.81	0.004	1.45	1.21–1.73	<0.0001
Stroke									
Cluster 2 vs. 1	2.92	1.89–4.54	<0.0001	0.61	0.09–3.91	0.60	1.63	0.99–2.66	0.05
Cluster 3 vs. 1	3.02	2.00–4.56	<0.0001	1.06	0 .63–1.79	0.82	1.43	0.90–2.28	0.13
Cluster 2 vs. 3	0.99	0.75–1.34	01.00	1.02	0.76–1.4	0.87	1.09	0.82–1.48	0.53
Major bleeding									
Cluster 2 vs. 1	2.47	1.66–3.66	<0.0001	1.88	0.51–6.93	0.34	2.06	1.37–3.10	<0.0001
Cluster 3 vs. 1	2.23	1.54–3.27	<0.0001	1.72	1.06–2.78	0.027	1.72	1.17–2.53	<0.0001
Cluster 2 vs. 3	1.11	0.83–1.48	0.50	0.67	0.09–4.83	0.69	1.14	0.86–1.53	0.37

CI: confidence interval; HR: hazard ratio.
[a] Adjusted for age, sex and type of atrial fibrillation, and adjusted for age and sex for stroke in the cluster 2 vs. 3 comparison, because of the small number issue.
[b] Adjusted for CHA_2DS_2-VASc score for the composite outcome, all-cause death and stroke, and adjusted for HAS-BLED score for major bleeding.

0.60 ± 0.91 in cluster 1. Aspirin 75–325 mg was still recommended during the 2000s for patients at low or intermediate risk of stroke (Class IB), explaining the high rate of antiplatelet agent prescription in this population. A low rate of anticoagulation and a high use of aspirin were also evident in other data from 2000–2010 [24,25].

Cluster 2 was characterized by older patients with permanent AF and a high burden of cardiovascular co-morbidities. The representative clinical features of these patients included a higher prevalence of cardiac pathologies. Hence, these patients were more likely to be treated with angiotensin-converting enzyme inhibitors or angiotensin receptor blockers, diuretics, digoxin and calcium channel blockers. Cluster 3 was also characterized by older patients with a high burden of cardiovascular co-morbidities, but comprised more women, with slightly higher CHA_2DS_2-VASc and HAS-BLED scores. In terms of medication, cluster 3 had the highest proportion of patients treated with class III antiarrhythmic drugs. Of note, cluster 2 seemed to be a subset of patients relatively similar to cluster 3 patients in terms of baseline characteristics (except for cardiac pathologies, sex and type of AF), but despite these similarities, cluster 2 patients had a worse prognosis with regard to the primary composite outcome and all-cause death than cluster 3 patients. This higher rate of adverse events in cluster 2 was mainly driven by the first year following the diagnosis of AF. Although the reasons are unclear, we speculated that a possible higher risk of stroke and major bleeding may have been masked by the high frequency of the competing event of all-cause death in cluster 2. Moreover, a proportion of sudden deaths may have been attributable to undiagnosed strokes.

Although cluster grouping is inconstant across previously published studies applied to mixed anticoagulated and non-anticoagulated cohorts, two phenotypes appear preserved. As in our study, a first cluster composed of younger patients, sometimes with behavioural disorders and associated with lower rates of adverse events, and a second cluster of atherosclerotic co-morbid patients, associated with a worse prognosis, have been described. Other clusters described previously vary according to the presence and prevalence of variables between datasets, practice pattern between countries and the origin of the population. For example, the type of AF served as the main driver to characterize cluster 2 in our cohort, and was not reported in a cluster analysis applied to a North American and a European anticoagulated population, but contributed to the formation of clusters in the Asian population of the KiCS-AF registry [2,6–8]. As the type of AF has been shown to be independently associated with an increase in death and thromboembolism, this may explain the excess of adverse events in cluster 2 [26].

5.1. Clinical implications and perspectives

In recent years, multimorbidity, polypharmacy and frailty, three components of so-called clinical complexity, have been described repeatedly in patients with AF, and several reports have underlined their detrimental impact on quality of care and prognosis [27–30]. This increasing awareness of clinical complexity in patients with AF has contributed to significant changes in the approach to their care. Recent guidelines on AF recommended appropriate characterization of patients with AF, followed by a more holistic or integrated care approach to AF management, based on the "Atrial Fibrillation Better Care" (ABC) pathway approach ("A" = Anticoagulation/Avoid stroke; "B" = Better symptom management; and "C" = Cardiovascular and Co-morbidity optimization) [23,31,32]. Compared with usual care, implementation of the ABC pathway has been significantly associated with a lower risk of adverse events and lower health-related costs, especially in clinically complex patients with AF [33–37]. However, applying the "C" component of the ABC pathway requires a good understanding and characterization, rather than classification, of the results of a combined effect of a complex continuum (often without specific threshold values) of multiple interacting factors, such as unhealthy lifestyle, risk factors and cardiovascular and non-cardiovascular disease.

Cluster analysis may be a good aid to further understand the complex nature of patients with AF. Hence, a cluster-driven characterization of clinically relevant phenotypes may help with a more tailored approach, facilitate AF management and decision-making, and improve outcomes. As some cluster phenotypes appear preserved across cohorts, identification of interventions specific to these groups of patients with AF is an important goal for future research. For example, the cluster of young AF patients with behavioural disorders, which can be assimilated to our cluster 1, has a non-negligible risk of adverse events, despite a low prevalence of co-morbidities. Focusing on lifestyle modifications in this specific population, and especially on diet and smoking cessation, could further reduce the impact of AF [38,39]. Cluster analysis also enables us to distinguish older patients with a high burden of cardiovascular disease (cluster 3) and a subset of patients (cluster 2) who, although apparently similar, do not have the same long-term

prognosis in terms of composite outcome of major adverse events, mainly driven by mortality, even after adjustment for age. This difference is apparent within the first year, highlighting the need for early and tailored management focused on cardiac pathologies and AF type.

5.2. Study limitations

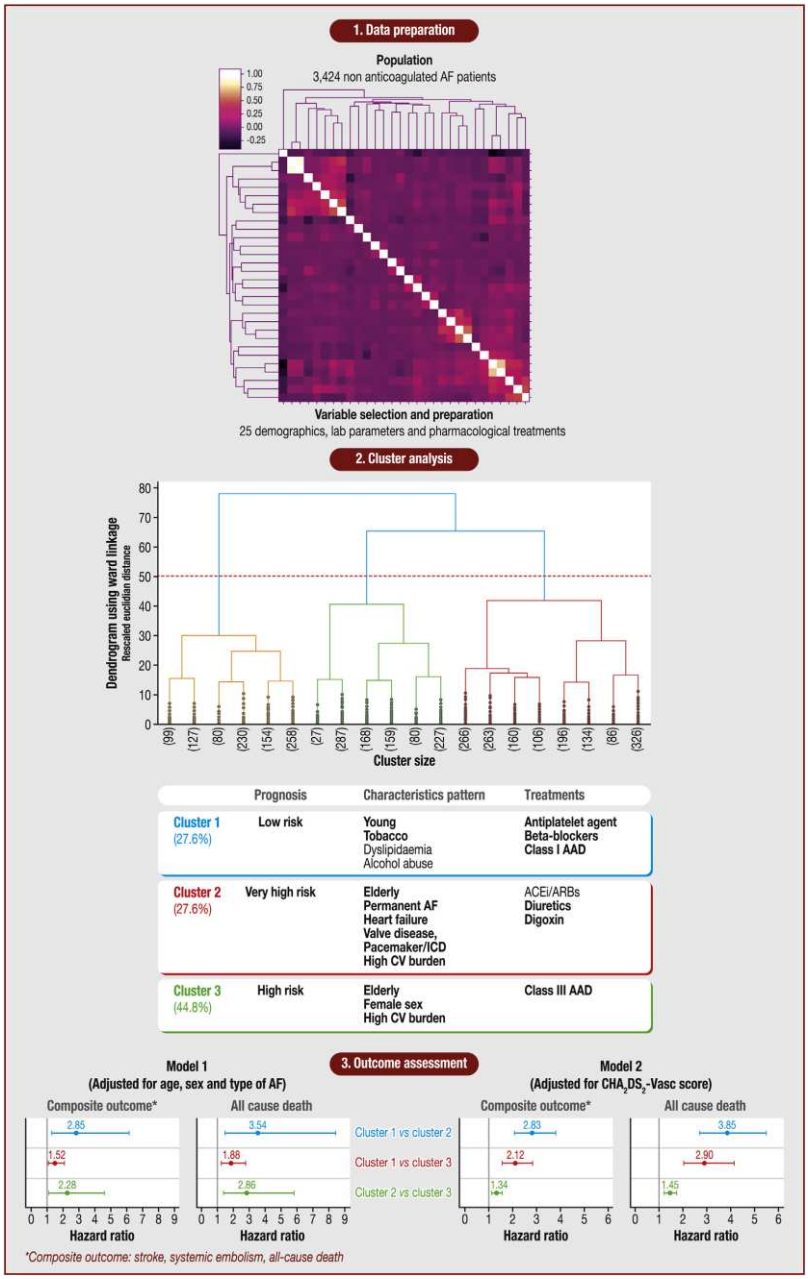
This study was based on a registry, and is subject to the limitations inherent to retrospective observational analyses. Knowing the exact reason for the absence of oral anticoagulation prescription is a challenge, especially in real-life data, where the reason(s) are not entirely clear. This can be explained by the combination of multiple factors related to the guidelines in application during the period of the study: low-risk patients not indicated for oral anticoagulation, patients taking antiplatelet agents instead of oral anticoagulation and the presence of misclassified patients as a result of the use of the $CHADS_2$ score. There were also patients with conditions such as dementia or with a possible so-called transient cause of AF, whom many physicians have been reluctant to treat with long-term vitamin K antagonists [19,40]. Another issue of using real-life data, particularly for outpatient management, is that a substantial part of oral anticoagulation may have been introduced and not recorded, despite the screening of hospital records, although the median follow-up was quite short: 0.82 (interquartile range 0.02–3.79) years.

Moreover, clustering algorithm results are dependent on the underlying population, available clinical variables and associated patterns of care in the community. The study population was hospital-based, and therefore may not be representative of all patients with AF. This study was not ethnically diverse, and our findings may not be generalizable to other populations. Then, as cluster analysis depends on available data, the incorporation of more clinical characteristics, such as ejection fraction and brain natriuretic peptide, might yield different results. However, as cluster analysis necessitates complete data on individual patients, we chose to drop variables with a missing data rate > 25%, to ensure the quality of the data. Another caveat is that we did not have access to data on events occurring outside of our area. Moreover, patients were included from 2000 to 2010, and treatment strategies and clinical practice have changed over time, partly limiting the generalization of the results to current clinical practice.

Although cluster analysis is not properly a predictive model, we aimed to demonstrate that this approach, which is an unsupervised method (outcome blinded), is able to distinguish relevant phenogroups with regard to prognosis in a non-anticoagulated AF cohort (which is difficult to find today). This complementary approach to usual clinical models may improve our understanding of clinical complexity, and provide us with new insights. This preliminary study provides novel data on clusters and their natural history in a fully non-anticoagulated AF cohort.

6. Conclusions

In this analysis of non-anticoagulated patients with AF, unsupervised cluster analysis led to the identification of three statistically driven groups of patients with AF with distinct phenotype characteristics. These distinct clusters were associated with different risks for all-cause death and major clinical adverse events. These results provide new insights into the natural history of non-anticoagulated clinically complex patients with AF. Such findings underlie the need for a more holistic management approach to treat existing comorbidities alongside AF, to minimize the risk of adverse outcomes (Central illustration).



Central Illustration. Overview of study design. Left: variable selection and data preparation among a population of 3424 non-anticoagulated patients with atrial fibrillation. Middle: identification of three statistically driven phenogroups. Right: assessment of major outcomes among each phenogroups. AAD: antiarrhythmic drug; ACEi: angiotensin-converting enzyme inhibitors; AF: atrial fibrillation; ARBs: angiotensin receptor blockers; CV: cardiovascular; ICD: implantable cardioverter-defibrillator.

Sources of funding

None.

Acknowledgments

Dr Arnaud Bisson would like to thank the Société Française de Cardiologie for providing this opportunity, through the Research and Fellowship Grant 2020, to work in collaboration with the University of Liverpool.

Disclosure of interest

A. B. has been a consultant or speaker for AstraZeneca, Bayer, BMS/Pfizer, Medtronic, Vitorpharma and Alnylam.

D. A. has been a consultant or speaker for Amgen, AstraZeneca, Bayer, BMS/Pfizer, MSD, Novartis, Novo Nordisk, Sanofi and Servier.

G. Y. H. L. has been a consultant or speaker for BMS/Pfizer, Boehringer Ingelheim, Daiichi-Sankyo and Anthos; no fees were received personally.

L. F. reports consulting fees for AstraZeneca, Bayer, BMS/Pfizer, Boehringer Ingelheim, Medtronic, Novo Nordisk and Novartis; and lecture fees for AstraZeneca, Bayer, BMS/Pfizer, Boehringer Ingelheim and Zoll.

The other authors declare that they have no competing interest.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.acvd.2023.06.001.

References

[1] Zhang J, Johnsen SP, Guo Y, Lip GYH. Epidemiology of atrial fibrillation: geographic/ecological risk factors, age, sex, genetics. Card Electrophysiol Clin 2021;13:1–23.
[2] Proietti M, Vitolo M, Harrison SL, et al. Impact of clinical phenotypes on management and outcomes in European atrial fibrillation patients: a report from the ESC-EHRA EURObservational Research Programme in AF (EORP-AF) general long-term registry. BMC Med 2021;19:256.
[3] Romiti GF, Proietti M, Bonini N, et al. Clinical complexity domains, anticoagulation, and outcomes in patients with atrial fibrillation: a report from the GLORIA-AF registry phase II and III. Thromb Haemost 2022;122:2030–41.
[4] Olier I, Ortega-Martorell S, Pieroni M, Lip GYH. How machine learning is impacting research in atrial fibrillation: implications for risk prediction and future management. Cardiovasc Res 2021;117:1700–17.
[5] Dey D, Slomka PJ, Leeson P, et al. Artificial intelligence in cardiovascular imaging: JACC state-of-the-art review. J Am Coll Cardiol 2019;73:1317–35.
[6] Inohara T, Piccini JP, Mahaffey KW, et al. A cluster analysis of the Japanese multicenter outpatient registry of patients with atrial fibrillation. Am J Cardiol 2019;124:871–8.
[7] Inohara T, Shrader P, Pieper K, et al. Association of of atrial fibrillation clinical phenotypes with treatment patterns and outcomes: a multicenter registry study. JAMA Cardiol 2018;3:54–63.
[8] Vitolo M, Proietti M, Shantsila A, Boriani G, Lip GYH. Clinical phenotype classification of atrial fibrillation patients using cluster analysis and associations with trial-adjudicated outcomes. Biomedicines 2021;9:843.
[9] Banerjee A, Fauchier L, Vourc'h P, et al. Renal impairment and ischemic stroke risk assessment in patients with atrial fibrillation: the Loire Valley atrial fibrillation project. J Am Coll Cardiol 2013;61:2079–87.
[10] Bisson A, Bodin A, Clementy N, et al. Stroke, thromboembolism and bleeding in patients with atrial fibrillation according to the EHRA valvular heart disease classification. Int J Cardiol 2018;260:93–8.
[11] Fauchier L, Lecoq C, Clementy N, et al. Oral anticoagulation and the risk of stroke or death in patients with atrial fibrillation and one additional stroke risk factor: the Loire Valley atrial fibrillation project. Chest 2016;149:960–8.
[12] Philippart R, Brunet-Bernard A, Clementy N, et al. Prognostic value of CHA2DS2-VASc score in patients with "non-valvular atrial fibrillation" and valvular heart disease: the Loire Valley atrial fibrillation project. Eur Heart J 2015;36:1822–30.
[13] Lip GY, Nieuwlaat R, Pisters R, Lane DA, Crijns HJ. 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 2010;137:263–72.
[14] Pisters R, Lane DA, Nieuwlaat R, de Vos CB, Crijns HJ, Lip GY. A novel user-friendly score (HAS-BLED) to assess 1-year risk of major bleeding in patients with atrial fibrillation: the Euro Heart Survey. Chest 2010;138:1093–100.
[15] van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. J Stat Softw 2011;45:1–67.
[16] Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in python. JMLR 2011;12:2825–30.
[17] Chen Q, Toorop MMA, Tops LF, Lijfering WM, Cannegieter SC. Time trends in patient characteristics, anticoagulation treatment, and prognosis of incident nonvalvular atrial fibrillation in the Netherlands. JAMA Netw Open 2023;6:e239973.
[18] Gorczyca-Glowacka I, Bielecka B, Walek P, et al. Temporal trends in oral anticoagulant prescription in atrial fibrillation patients between 2004 and 2019. Int J Environ Res Public Health 2022:19.
[19] Mongkhon P, Alwafi H, Fanning L, et al. Patterns and factors influencing oral anticoagulant prescription in people with atrial fibrillation and dementia: results from UK primary care. Br J Clin Pharmacol 2021;87:1056–68.
[20] Sur NB, Wang K, Di Tullio MR, et al. Disparities and temporal trends in the use of anticoagulation in patients with ischemic stroke and atrial fibrillation. Stroke 2019;50:1452–9.
[21] Camm AJ, Lip GY, De Caterina R, et al. 2012 focused update of the ESC Guidelines for the management of atrial fibrillation: an update of the 2010 ESC Guidelines for the management of atrial fibrillation. Developed with the special contribution of the European Heart Rhythm Association. Eur Heart J 2012;33:2719–47.
[22] European Heart Rhythm Association, European Association for Cardio-Thoracic Surgery, Camm AJ, et al. Guidelines for the management of atrial fibrillation: the task force for the management of atrial fibrillation of the European Society of Cardiology (ESC). Eur Heart J 2010;31:2369–429.
[23] Hindricks G, Potpara T, Dagres N, et al. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboation 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. Eur Heart J 2021;42:373–498.
[24] Nieuwlaat R, Capucci A, Lip GY, et al. Antithrombotic treatment in real-life atrial fibrillation patients: a report from the Euro Heart Survey on Atrial Fibrillation. Eur Heart J 2006;27:3018–26.
[25] Waldo AL, Becker RC, Tapson VF, Colgan KJ, NABOR Steering Committee. Hospitalized patients with atrial fibrillation and a high risk of stroke are not being provided with adequate anticoagulation. J Am Coll Cardiol 2005;46:1729–36.
[26] Ganesan AN, Chew DP, Hartshorne T, et al. The impact of atrial fibrillation type on the risk of thromboembolism, mortality, and bleeding: a systematic review and meta-analysis. Eur Heart J 2016;37:1591–602.
[27] Chen N, Alam AB, Lutsey PL, et al. Polypharmacy, adverse outcomes, and treatment effectiveness in patients ≥ 75 with atrial fibrillation. J Am Heart Assoc 2020;9:e015089.
[28] Jani BD, Nicholl BI, McQueenie R, et al. Multimorbidity and co-morbidity in atrial fibrillation and effects on survival: findings from UK Biobank cohort. Europace 2018;20:f329–36.
[29] Proietti M, Cesari M. Describing the relationship between atrial fibrillation and frailty: clinical implications and open research questions. Exp Gerontol 2021;152:111455.
[30] Woolford SJ, Aggarwal P, Sheikh CJ, Patel HP. Frailty, multimorbidity and polypharmacy. Medicine 2021;49:166–72.
[31] Lip GYH. The ABC pathway: an integrated approach to improve AF management. Nat Rev Cardiol 2017;14:627–8.
[32] Potpara TS, Lip GYH, Blomstrom-Lundqvist C, et al. The 4S-AF Scheme (stroke risk; symptoms; severity of burden; substrate): a novel approach to in-depth characterization (rather than classification) of atrial fibrillation. Thromb Haemost 2021;121:270–8.
[33] Pastori D, Farcomeni A, Pignatelli P, Violi F, Lip GY. ABC (atrial fibrillation better care) pathway and healthcare costs in atrial fibrillation: the ATHERO-AF study. Am J Med 2019;132:856–61.
[34] Pastori D, Pignatelli P, Menichelli D, Violi F, Lip GYH. Integrated care management of patients with atrial fibrillation and risk of cardiovascular events: The ABC (atrial fibrillation better care) pathway in the ATHERO-AF study cohort. Mayo Clin Proc 2019;94:1261–7.
[35] Proietti M, Romiti GF, Olshansky B, Lane DA, Lip GYH. Improved outcomes by integrated care of anticoagulated patients with atrial fibrillation using the simple ABC (atrial fibrillation better care) pathway. Am J Med 2018;131 [1359-1366 e1356].
[36] Proietti M, Romiti GF, Olshansky B, Lane DA, Lip GYH. Comprehensive management with the ABC (atrial fibrillation better care) pathway in clinically complex patients with atrial fibrillation: a post hoc ancillary analysis from the AFFIRM trial. J Am Heart Assoc 2020;9:e014932.
[37] Yoon M, Yang PS, Jang E, et al. Improved population-based clinical outcomes of patients with atrial fibrillation by compliance with the simple ABC (atrial fibrillation better care) pathway for integrated care management: a nationwide cohort study. Thromb Haemost 2019;119:1695–703.
[38] Angoulvant D, Villejoubert O, Bejan-Angoulvant T, et al. Effect of active smoking on comparative efficacy of antithrombotic therapy in patients with atrial fibrillation: the Loire Valley atrial fibrillation project. Chest 2015;148:491–8.
[39] Nalliah CJ, Sanders P, Kalman JM. The impact of diet and lifestyle on atrial fibrillation. Curr Cardiol Rep 2018;20:137.
[40] Fauchier L, Clementy N, Bisson A, et al. Prognosis in patients with atrial fibrillation and a presumed "temporary cause" in a community-based cohort study. Clin Res Cardiol 2017;106:202–10.

2.4 Discussion

A travers ces deux derniers travaux, nous avons appliqué une méthode non supervisée d'analyse en clusters afin de décrire des phénogroupes cliniquement pertinents à partir de patients atteints de FA non anticoagulés (Travail 1) ou traités par ablation par cathéter (Travail 2) [116,117]. Dans les deux cas, cette approche nous a permis d'identifier 3 phénogroupes présentant des caractéristiques significativement distinctes et associés de manière indépendante à différentes probabilités en terme d'événements indésirables.

Bien que les méthodes de ML soient réputées pour être 'data consuming', les modèles non supervisés ont l'avantage d'être applicables sur des cohortes locales avec de plus petits effectifs (n<1000) comme c'est le cas dans le premier travail. Ce qui est perdu en termes de nombre d'observations est compensé par des données souvent plus granuleuses (continues) et donc intéressantes à analyser avec les approches de type ML.

Distinguer, à partir d'une population hétérogène, des sous-groupes homogènes dont les caractéristiques ne diffèrent que légèrement (variations nuancées, continuum) est souvent complexe. Classiquement, les groupes sont définis à partir d'une variable pré-spécifiée dont la capacité à séparer est déjà connue [118,119]. La limite de cette méthode est que cela ne prend pas en compte la complexité des interactions entre variables. La particularité de l'analyse en cluster est de définir de nouveau phénogroupes sans *a priori* uniquement guidés par la donnée en aveugle de tout résultat. De plus, elle fournit des informations sur les relations entre les données, leur distance (degré de différences ou de similitudes), qui sont ensuite facilement visualisables (dendrogramme, nuage de points après réduction de dimensionnalité, heatmaps).

Le premier intérêt de cette approche est de découvrir de nouveaux patterns dans les données permettant, à partir de là, la formulation d'hypothèses afin d'ouvrir des pistes de recherche (hypothesis or knowledge generating approach). C'est le cas dans les deux travaux dès lors que l'on cherche à expliquer le lien entre phénogroupe et pronostic. Est-il lié au terrain, à l'histoire naturelle de la FA (type, ancienneté ou progression) ou aux thérapeutiques mises en œuvre ? Les questions sont nombreuses.

Le second intérêt est directement clinique. Décrire des profils types de patients dont le pronostic diffère permet de mieux adapter et d'optimiser la prise en charge. Dans l'optique d'une médecine personnalisée de plus en plus précise, les méthodes de clustering vont aider à la meilleure sélection des patients.

CONCLUSION

Ces dernières années, le ML est devenu une méthode populaire en épidémiologie où l'utilisation de méthodes statistiques classiques peut atteindre ses limites, suscitant un intérêt croissant auprès de la communauté scientifique.

Par ce travail, nous avons pu observer que cette approche est applicable sur une variété de sources de données de FA différentes disponibles actuellement et ce quel que soit les particularités celles-ci (registre international de populations hétérogènes (GLORIA-AF), données nationales constituées de variables essentiellement binaires (PMSI), données locales limitées nombre d'observations mais aux variables granuleuses (Loire Valley-AF cohorte)) [111,112,116,117].

De plus, nous avons montré l'intérêt de cette approche, par la supériorité des modèles prédictifs développés par rapports aux modèles classiques, par la description de phénogroupes de patients cliniquement pertinents et par l'éclairage nouveau apporté sur nos données permettant l'émergence de nouvelles questions et hypothèses.

D'un point de vue clinique, le ML semble donc être une technologie permettant l'essor d'une médecine intégrative et personnalisée performante.

PERSPECTIVES

Si cela vient de commencer, c'est surtout durant la décennie à venir que nous allons voir la véritable explosion du ML en santé et sa concrétisation.

La constitution d'énormes bases de données de santé nationales accessibles comme le Health Data Hub en France, la multiplicité des sources de données exploitables rendue possible par la eSanté, la facilité de l'accès aux ressources éducatives sur la méthode et la simplification d'utilisation des algorithmes via les environnements open-source participent à l'accélération du processus de développement de cette technique [12,66,96,120,121].

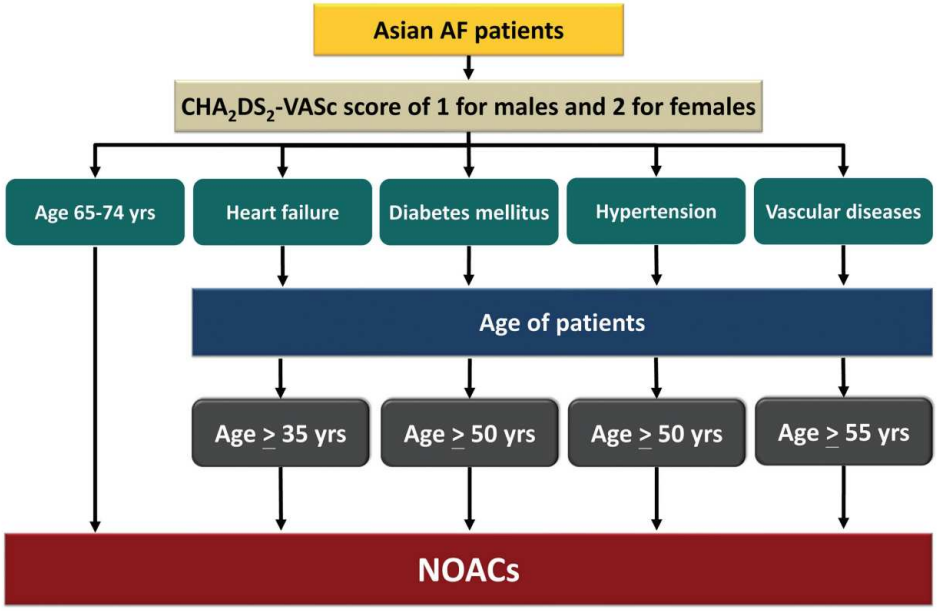
Les champs d'application sont considérables comme la stratification de risques, la personnalisation des prises en charge et le dépistage que ce soit via les dossiers médicaux informatisés, les prothèses cardiaques électroniques ou encore les objets connectées [121].

Cependant, il persiste quelques défis pour lesquels des solutions sont en cours d'élaboration [122]. En effet, si l'on reproche souvent l'opacité des mécanismes de décision des algorithmes de ML (effet « black box »), les experts estiment possible de trouver des solutions à ce problème au cours des prochaines années [123]. D'autres solutions sont en cours de développement comme de nouveau algorithmes pour les traitements de données non structurées (texte, image) ou la plus grande facilité d'accès à de grandes bases de données afin de permettre la validation externe de modèles.

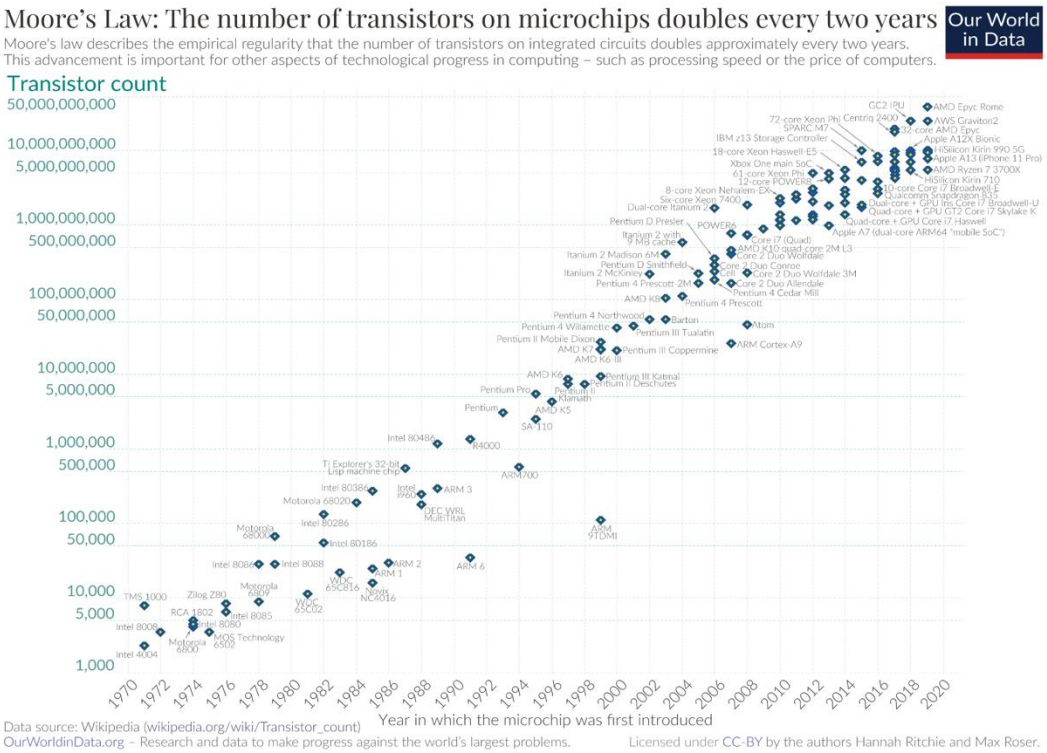
Bien que les diverses applications décrites soient prometteuses, des études prospectives d'envergure évaluant l'impact de ces technologies sur des critères d'évaluation cliniques durs restent nécessaires pour que cela se traduise dans les recommandations et dans la pratique clinique quotidienne.

A une échelle plus locale au sein de notre équipe, ces travaux ont permis acquérir la compréhension, les compétences et la maitrise nécessaire à l'utilisation de cette technique. Cela représente pour les prochaines années une formidable opportunité pour nos recherches à venir.

ANNEXES



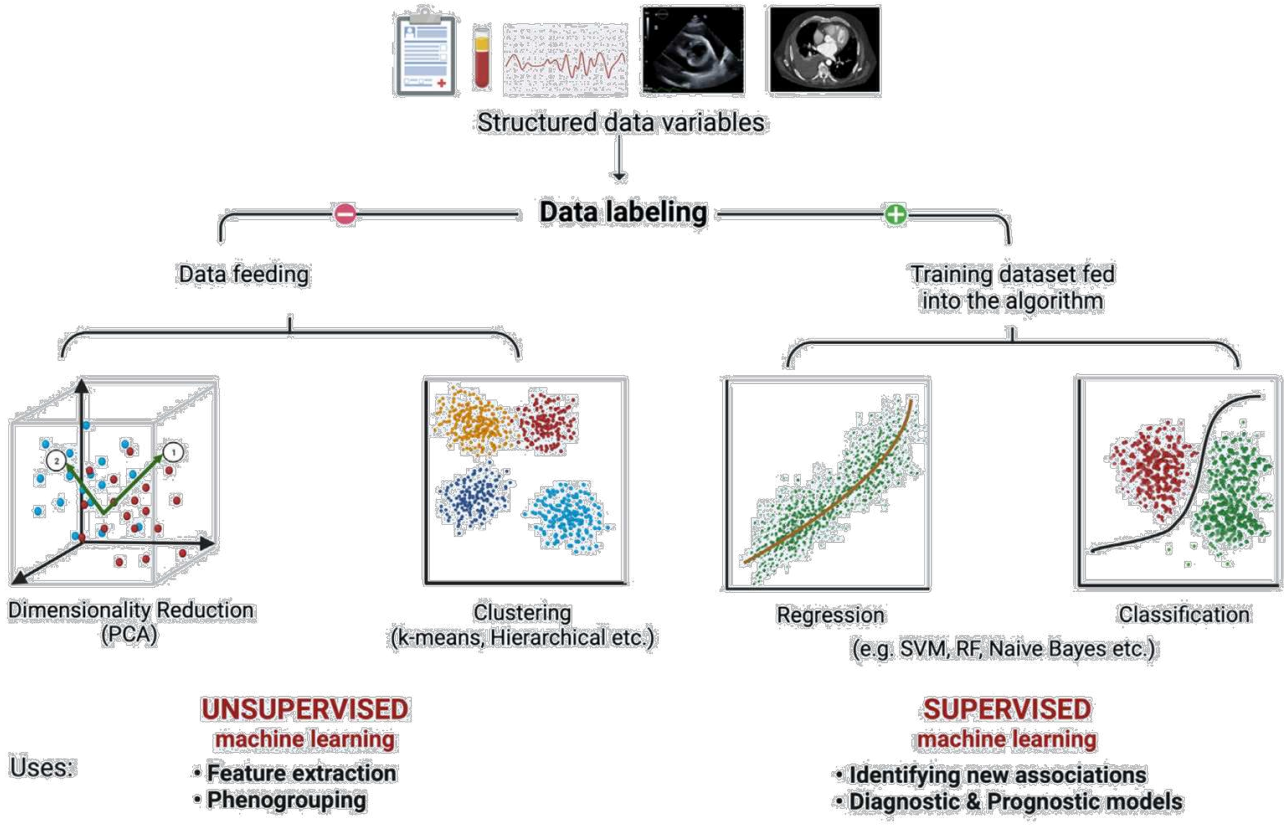
Annexe 1. Arbre de décision sur l'usage des anticoagulants oraux proposé chez les patients asiatiques à risque thromboembolique intermédiaire [69].



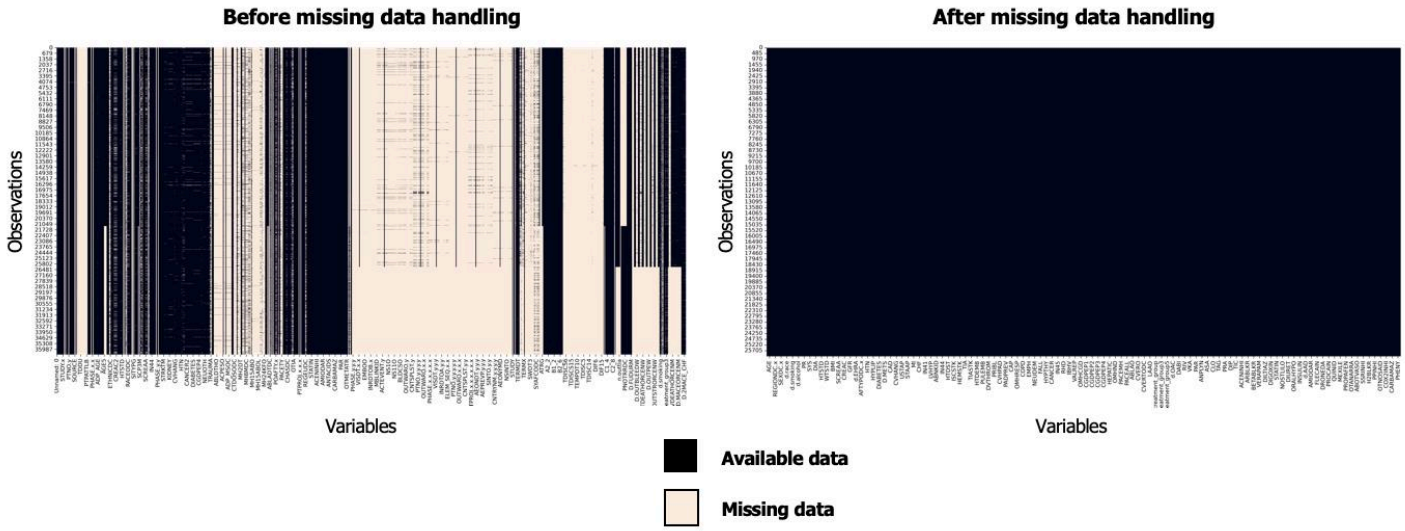
Annexe 2. Représentation de la loi de Moore au cours des 50 dernières années.

Machine Learning Term(s)	Epidemiology Term(s)	Definition and Notes
Attribute, feature, predictor, or field	Independent variable	Machine learning uses various terms to reference what epidemiologists would consider an "independent variable," including *attribute*, *feature*, *predictor*, and *field*.
Domain	Range of possible variable values	The domain is the set of possible values of an attribute. It can be continuous or categorical/binary.
Input and output	Independent (exposure) and dependent (outcome) variables	In machine learning, "input" refers to all of the predictors or independent variables that enter the model, and "output" generally refers to the predicted value (whether a number, classification, etc.) of the dependent variable or outcome.
Classifier, estimator	Model	"Classifiers" or "estimators" are used generally in the machine learning literature to refer to algorithms that perform a prediction or classification of interest. Their less common, though more technical, usage specifically refers to fully parameterized models that are used to predict or classify.
Learner	Model-fitting algorithm	A learner inputs a training set and outputs a classifier. Usually, but not always, *learner* refers to the fitting algorithm, while *classifier* refers to the fitted model.
Dimensionality	No. of covariates	No. of independent variables under consideration in a model.
Label	Value of dependent variables, outcomes	A variable's label is its value for each observation (e.g., 0 or 1). Although labels can technically describe any variable, common shorthand is that "labeled data" refers to data in which the dependent variable assumes a value for all observations.
Imbalanced data	Data set in which some cases or risk categories occur much less frequently than the others	In imbalanced machine learning data sets, the outcome or another risk category of interest occurs much less frequently, either because of the intrinsic nature of the problem (e.g., a rare disease in a database of medical records) or because of the sampling strategy (e.g., prevalence of cases in the study population is much lower than that in the target/source population). Heavily imbalanced data may pose challenges in some classification algorithms and require tuning parameters in order to correct for or otherwise address this imbalance. One method for addressing imbalanced data sets is to "balance" them artificially, either by oversampling instances of the minority class or undersampling instances of the majority class.
Loss function	Error measure	In machine learning, a loss function is generally considered a penalty for misclassification when assessing a model's predictive performance.

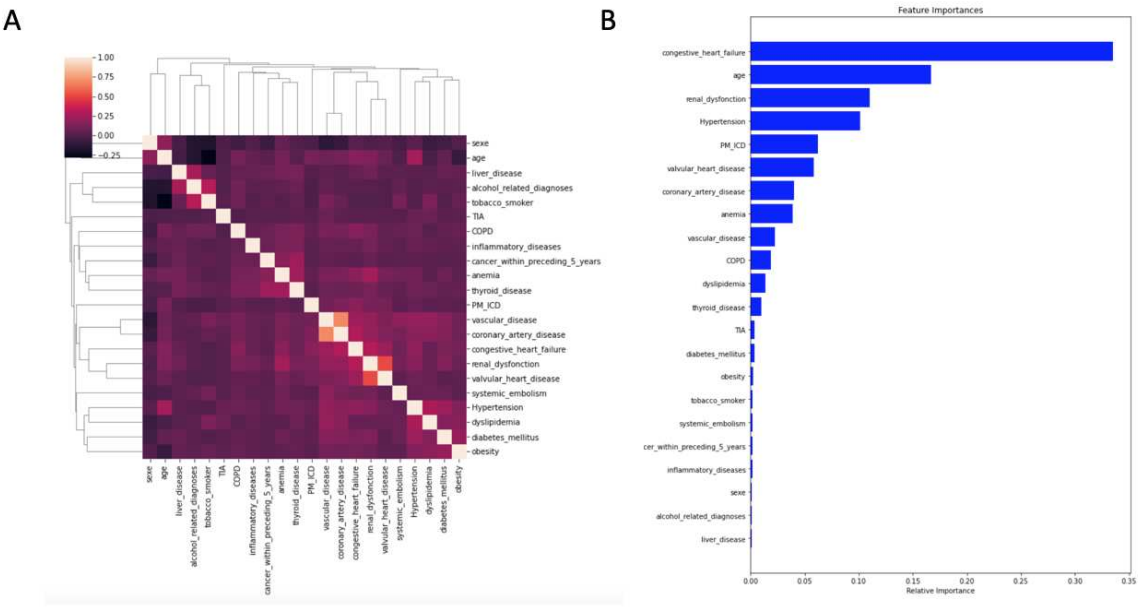
Annexe 3. Correspondances terminologiques entre épidémiologie et machine learning [96].



Annexe 4. Classification des algorithmes de machine learning [122].

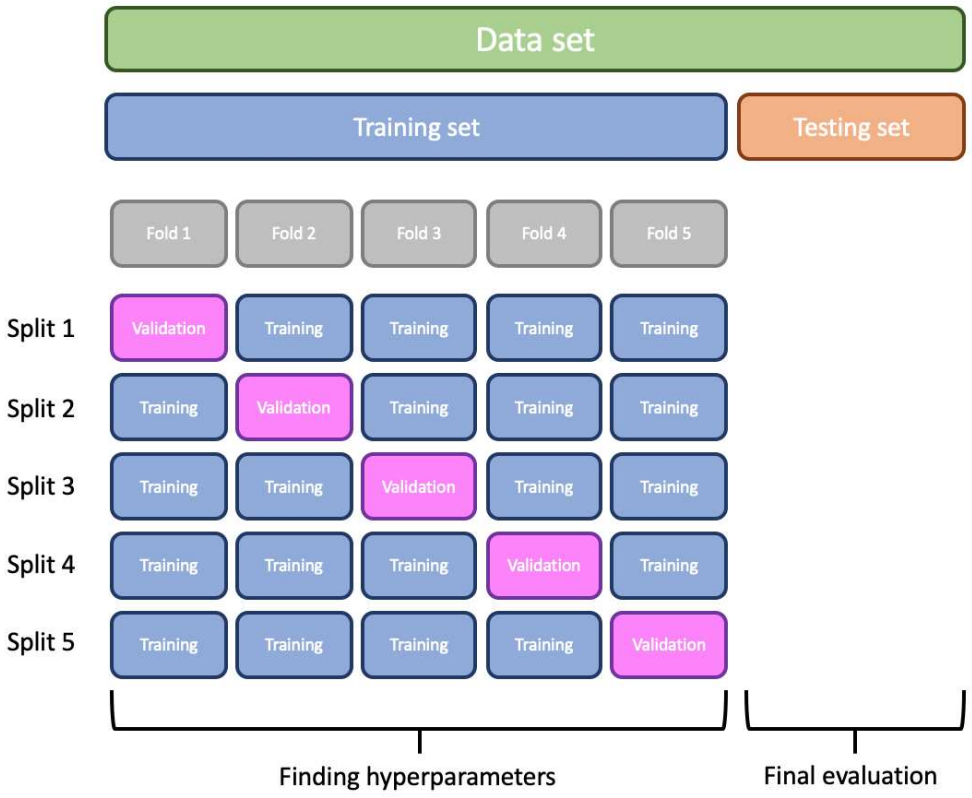


Annexe 5. Heatmaps représentant l'importance des données manquantes.
Exemple provenant du registre international GLORIA-AF.

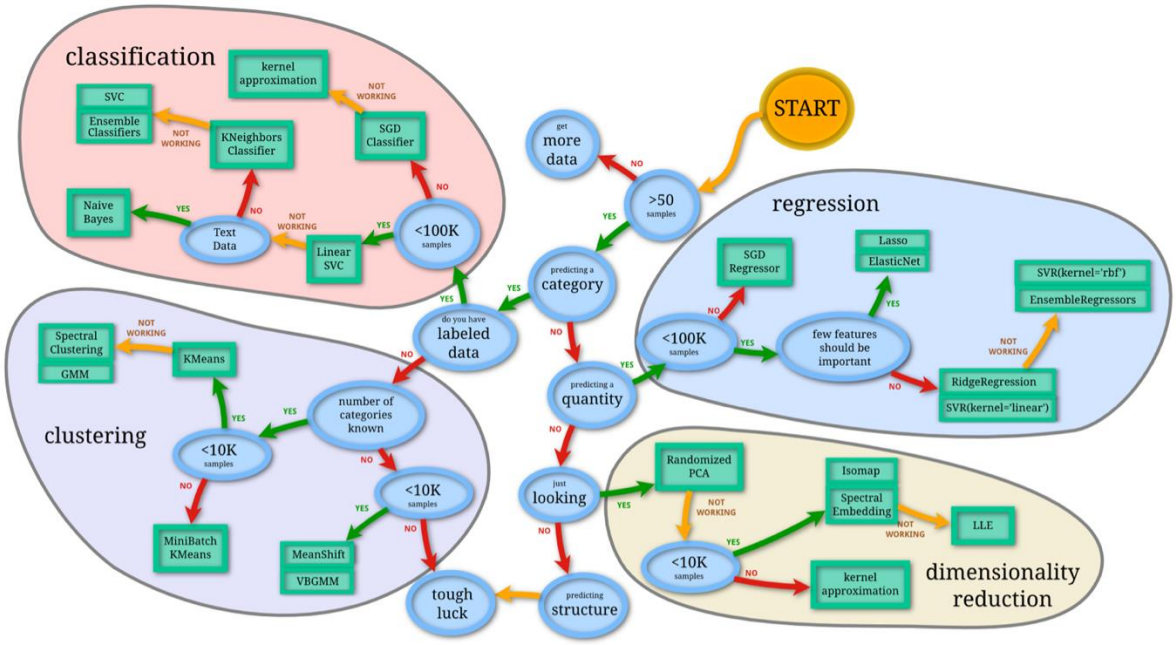


Annexe 6. Méthode de sélection des features.

(**A**) Heatmap représentant la corrélation des features entre elles afin de retirer les variables redondantes (méthode non supervisée).
(**B**) Classement par ordre d'importance de la participation de chaque feature dans la prédiction d'un estimateur pour un label défini (méthode supervisée).

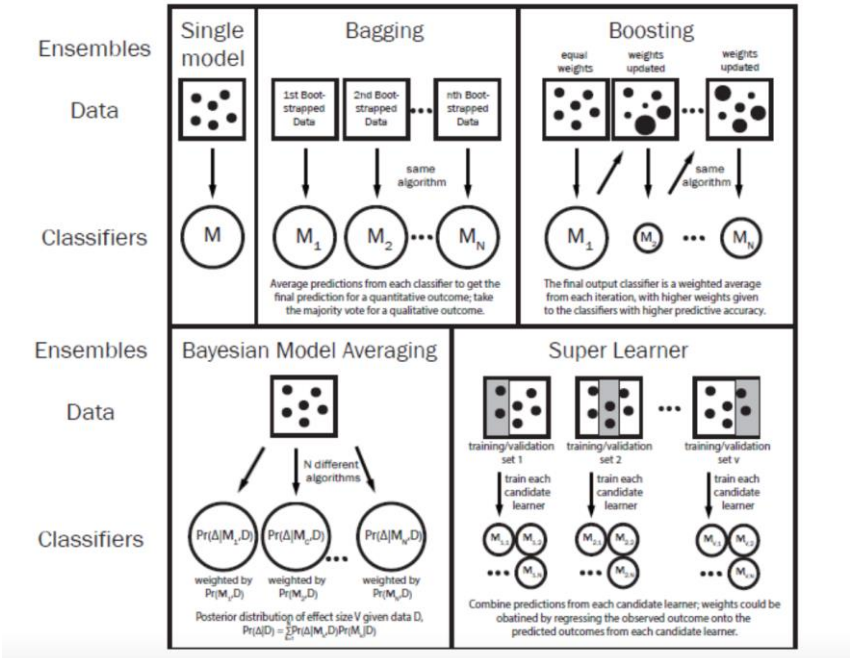


Annexe 7. Principe du train-test split combiné avec une validation croisée.



Annexe 8. Algorithme de choix des différents estimateurs selon les données et le problème proposé par Scikit-learn.

Disponible sur : https://scikit-learn.org/stable/tutorial/machine_learning_map/index.html



Annexe 9. Représentation des 4 modèles d'ensemble : bagging, boosting, Bayesian model averaging et Super Learner [96].

Les modèles bagging et boosting entraînent le même algorithme sur un sous-ensemble différent des données d'origine, de manière parallèle et séquentielle respectivement, alors que les modèles bayésiens et Super Learner entraînent un large éventail d'algorithmes sur le même ensemble de données.

Model	Hyper-parameters
LR	C
	Class weight
RFC	N estimators
	Max depth
	Min samples leaf
	Min samples split
	Class weight
SVM	C
	Gamma
	Kernel
	Class weight
XGBoost	Max depth
	Learning rate
	N estimators
	Min child weight
	Gamma
	Subsample
	Colsample bytree
	Scale pos weight
DNN	Hidden layers
	Learning rate
	Dropout
	Epochs
	Batch size
	Activation
	Loss functiuon
	Optimizer
	Class weight

Annexe 10. Exemples d'hyper-paramètres selon différents modèles supervisés d'apprentissage.

DNN: Deep neural network; LR: Logistic regression; RFC: RandomForest Classifier; SVM: Support Vector Machine; XGBoost: Extrem gradient boosting.

A

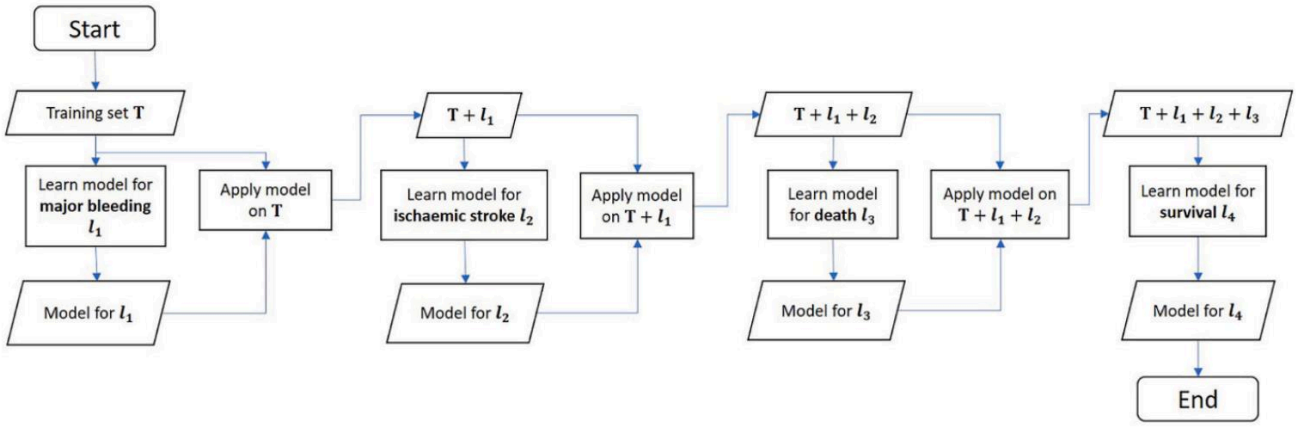
```
# Set the metric
metric = 'roc_auc'

# List estimators and hyperparameters in a dictionariy
model_params = {'logistic regression': {'model':LogisticRegression(solver='liblinear',multi_class='auto', random_state=42),
                                        'params' : {'C': [0.01,0.1,1,3,5,7,10],'class_weight':({},{0:1,1:7})}},
                'random_forest': {'model': RandomForestClassifier(random_state=42),
                                  'params': {'n_estimators':np.arange(5, 500, 20), 'max_depth':np.arange(1, 50, 5), 'class_weight':({},{0:1,1:7})}},
                'KNN': {'model': KNeighborsClassifier(),
                        'params': {'n_neighbors':np.arange(1, 15, 3)}},
                'MLPC': {'model': MLPClassifier(max_iter=500, random_state=42),
                         'params': {'hidden_layer_sizes': [(50,50,50), (50,100,50), (100,)], 'activation': ['tanh', 'relu'],
                                    'alpha': [0.0001, 0.05],'learning_rate': ['constant','adaptive']}}
               }

# Train estimators with a cross validation using GridSearchCV in a for loop
scores = []
for model_name, mp in model_params.items():
    grid = GridSearchCV(mp['model'], mp['params'], scoring=metric, n_iter=30, n_jobs=-1, cv=5)
    grid.fit(X_train, y_train)
    scores.append({'model': model_name,'best_score': grid.best_score_, 'best_params': grid.best_params_})

# Output the results
df = pd.DataFrame (scores, columns=['model', 'best_score', 'best_params'])
print(df)
```

B

	Model	Best_score	Best_params
0	LR	0.619	{'class_weight': {}, 'C': 7}
1	RFC	0.680	{'n_estimators': 185, 'max_depth': 1, 'class_weight': {}}
2	KNN	0.604	{'n_neighbors': 13}
3	MLPC	0.623	{'learning_rate': 'constant', 'hidden_layer_sizes': (100,), 'alpha': 0.0001, 'activation': 'tanh'}

Annexe 11. Exemple d'optimisation simultanée de plusieurs estimateurs par la méthode du GridSearch.

(**A)** Lignes de code en langage Python programmant l'entrainement de plusieurs combinaisons d'hyper-paramètres pour plusieurs estimateurs différents. (**B**) Tableau de résultats correspondant à l'AUC de la meilleure combinaison d'hyper-paramètre pour chaque estimateur.
LR: logistic regression; RFC: Random Forest Classifier; KNN: K nearest neiboughrs; MLPC: Multi Layer Perceptron Classifier.



Annexe 12. Architecture d'un modèle de machine learning multi label [124].

L'entrée (T) a été intégrée au premier modèle (M1) de la chaîne pour prédire les admissions pour hémorragie majeure (l1). Ensuite, l'entrée et le résultat prédit de l'hémorragie majeure (T + l1) ont été mis à jour en tant que nouvelle entrée à intégrer dans le deuxième modèle (M2) de la chaîne pour prédire les admissions pour AVCi (l2). Les résultats prédits ainsi que l'entrée du modèle (T + l1 + l2) ont servi d'entrée au modèle de prédiction du décès (l3) (M3). Enfin, la survie (l4) est prédite à l'aide de l'entrée composite (T + l1 + l2 + l3).

REFERENCES BIBLIOGRAPHIQUES

1. Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomström-Lundqvist C, et al. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): 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. 2021 Feb 1;42(5):373–498.
2. Benjamin EJ, Muntner P, Alonso A, Bittencourt MS, Callaway CW, Carson AP, et al. Heart Disease and Stroke Statistics-2019 Update: A Report From the American Heart Association. Circulation. 2019 Mar 5;139(10):e56–528.
3. Chugh SS, Havmoeller R, Narayanan K, Singh D, Rienstra M, Benjamin EJ, et al. Worldwide Epidemiology of Atrial Fibrillation: A Global Burden of Disease 2010 Study. Circulation. 2014 Feb 25;129(8):837–47.
4. Schnabel RB, Yin X, Gona P, Larson MG, Beiser AS, McManus DD, et al. 50 year trends in atrial fibrillation prevalence, incidence, risk factors, and mortality in the Framingham Heart Study: a cohort study. Lancet. 2015 Jul 11;386(9989):154–62.
5. Colilla S, Crow A, Petkun W, Singer DE, Simon T, Liu X. Estimates of current and future incidence and prevalence of atrial fibrillation in the U.S. adult population. Am J Cardiol. 2013 Oct 15;112(8):1142–7.
6. Krijthe BP, Kunst A, Benjamin EJ, Lip GYH, Franco OH, Hofman A, et al. Projections on the number of individuals with atrial fibrillation in the European Union, from 2000 to 2060. Eur Heart J. 2013 Sep;34(35):2746–51.
7. Magnussen C, Niiranen TJ, Ojeda FM, Gianfagna F, Blankenberg S, Njølstad I, et al. Sex Differences and Similarities in Atrial Fibrillation Epidemiology, Risk Factors, and Mortality in Community Cohorts: Results From the BiomarCaRE Consortium (Biomarker for Cardiovascular Risk Assessment in Europe). Circulation. 2017 Oct 24;136(17):1588–97.
8. Staerk L, Wang B, Preis SR, Larson MG, Lubitz SA, Ellinor PT, et al. Lifetime risk of atrial fibrillation according to optimal, borderline, or elevated levels of risk factors: cohort study based on longitudinal data from the Framingham Heart Study. BMJ. 2018 Apr 26;361:k1453.
9. Charlemagne A, Blacher J, Cohen A, Collet JP, Diévart F, de Groote P, et al. Epidemiology of atrial fibrillation in France: Extrapolation of international epidemiological data to France and analysis of French hospitalization data. Archives of Cardiovascular Diseases. 2011 Feb 1;104(2):115–24.
10. Zoni-Berisso M, Lercari F, Carazza T, Domenicucci S. Epidemiology of atrial fibrillation: European perspective. Clin Epidemiol. 2014;6:213–20.
11. Odutayo A, Wong CX, Hsiao AJ, Hopewell S, Altman DG, Emdin CA. Atrial fibrillation and risks of cardiovascular disease, renal disease, and death: systematic review and meta-analysis. BMJ. 2016 Sep 6;354:i4482.
12. Kornej J, Börschel CS, Benjamin EJ, Schnabel RB. Epidemiology of Atrial Fibrillation in the 21st Century. Circulation Research. 2020 Jun 19;127(1):4–20.
13. Kishore A, Vail A, Majid A, Dawson J, Lees KR, Tyrrell PJ, et al. Detection of atrial fibrillation after ischemic stroke or transient ischemic attack: a systematic review and meta-analysis. Stroke. 2014 Feb;45(2):520–6.
14. Henriksson KM, Farahmand B, Åsberg S, Edvardsson N, Terént A. Comparison of cardiovascular risk factors and survival in patients with ischemic or hemorrhagic stroke. Int J Stroke. 2012 Jun;7(4):276–81.
15. Grond M, Jauss M, Hamann G, Stark E, Veltkamp R, Nabavi D, et al. Improved

detection of silent atrial fibrillation using 72-hour Holter ECG in patients with ischemic stroke: a prospective multicenter cohort study. Stroke. 2013 Dec;44(12):3357–64.
16. Kolominsky-Rabas PL, Weber M, Gefeller O, Neundoerfer B, Heuschmann PU. Epidemiology of Ischemic Stroke Subtypes According to TOAST Criteria Incidence, Recurrence, and Long-Term Survival in Ischemic Stroke Subtypes: A Population-Based Study. Stroke. 2001 Jan 12;32(12):2735–40.
17. Pistoia F, Sacco S, Tiseo C, Degan D, Ornello R, Carolei A. The Epidemiology of Atrial Fibrillation and Stroke. Cardiol Clin. 2016 May;34(2):255–68.
18. Nieuwlaat R, Eurlings LW, Cleland JG, Cobbe SM, Vardas PE, Capucci A, et al. Atrial Fibrillation and Heart Failure in Cardiology Practice: Reciprocal Impact and Combined Management From the Perspective of Atrial Fibrillation: Results of the Euro Heart Survey on Atrial Fibrillation. Journal of the American College of Cardiology. 2009 May 5;53(18):1690–8.
19. König S, Ueberham L, Schuler E, Wiedemann M, Reithmann C, Seyfarth M, et al. In-hospital mortality of patients with atrial arrhythmias: insights from the German-wide Helios hospital network of 161 502 patients and 34 025 arrhythmia-related procedures. Eur Heart J. 2018 Nov 1;39(44):3947–57.
20. An Y, Ogawa H, Yamashita Y, Ishii M, Iguchi M, Masunaga N, et al. Causes of death in Japanese patients with atrial fibrillation: The Fushimi Atrial Fibrillation Registry. Eur Heart J Qual Care Clin Outcomes. 2019 Jan 1;5(1):35–42.
21. Dagres N, Chao TF, Fenelon G, Aguinaga L, Benhayon D, Benjamin EJ, et al. European Heart Rhythm Association (EHRA)/Heart Rhythm Society (HRS)/Asia Pacific Heart Rhythm Society (APHRS)/Latin American Heart Rhythm Society (LAHRS) expert consensus on arrhythmias and cognitive function: what is the best practice? Europace. 2018 Sep 1;20(9):1399–421.
22. Schnabel RB, Michal M, Wilde S, Wiltink J, Wild PS, Sinning CR, et al. Depression in atrial fibrillation in the general population. PLoS One. 2013;8(12):e79109.
23. Freeman JV, Simon DN, Go AS, Spertus J, Fonarow GC, Gersh BJ, et al. Association Between Atrial Fibrillation Symptoms, Quality of Life, and Patient Outcomes: Results From the Outcomes Registry for Better Informed Treatment of Atrial Fibrillation (ORBIT-AF). Circ Cardiovasc Qual Outcomes. 2015 Jul;8(4):393–402.
24. Burdett P, Lip GYH. Atrial fibrillation in the UK: predicting costs of an emerging epidemic recognizing and forecasting the cost drivers of atrial fibrillation-related costs. Eur Heart J Qual Care Clin Outcomes. 2022 Mar 2;8(2):187–94.
25. Le Heuzey JY, Paziaud O, Piot O, Said MA, Copie X, Lavergne T, et al. Cost of care distribution in atrial fibrillation patients: the COCAF study. Am Heart J. 2004 Jan;147(1):121–6.
26. Cotté FE, Chaize G, Gaudin AF, Samson A, Vainchtock A, Fauchier L. Burden of stroke and other cardiovascular complications in patients with atrial fibrillation hospitalized in France. Europace. 2016 Apr;18(4):501–7.
27. O'Keefe EL, Sturgess JE, O'Keefe JH, Gupta S, Lavie CJ. Prevention and Treatment of Atrial Fibrillation via Risk Factor Modification. American Journal of Cardiology. 2021 Dec 1;160:46–52.
28. Hobbs FDR, Fitzmaurice DA, Mant J, Murray E, Jowett S, Bryan S, et al. A randomised controlled trial and cost-effectiveness study of systematic screening (targeted and total population screening) versus routine practice for the detection of atrial fibrillation in people aged 65 and over. The SAFE study. Health Technol Assess. 2005 Oct;9(40):iii–iv, ix–x, 1–74.
29. Rienstra M, Hobbelt AH, Alings M, Tijssen JGP, Smit MD, Brügemann J, et al. Targeted therapy of underlying conditions improves sinus rhythm maintenance in patients

with persistent atrial fibrillation: results of the RACE 3 trial. Eur Heart J. 2018 Aug 21;39(32):2987–96.
30. Hart RG, Pearce LA, Aguilar MI. Meta-analysis: antithrombotic therapy to prevent stroke in patients who have nonvalvular atrial fibrillation. Ann Intern Med. 2007 Jun 19;146(12):857–67.
31. Secondary prevention in non-rheumatic atrial fibrillation after transient ischaemic attack or minor stroke. EAFT (European Atrial Fibrillation Trial) Study Group. Lancet. 1993 Nov 20;342(8882):1255–62.
32. Kirchhof P, Camm AJ, Goette A, Brandes A, Eckardt L, Elvan A, et al. Early Rhythm-Control Therapy in Patients with Atrial Fibrillation. New England Journal of Medicine. 2020 Oct 1;383(14):1305–16.
33. Gunawardene MA, Willems S. Atrial fibrillation progression and the importance of early treatment for improving clinical outcomes. EP Europace. 2022 Jun 6;24(Supplement_2):ii22–8.
34. Lip GYH. The ABC pathway: an integrated approach to improve AF management. Nat Rev Cardiol. 2017 Nov;14(11):627–8.
35. Proietti M, Romiti GF, Olshansky B, Lane DA, Lip GYH. Improved Outcomes by Integrated Care of Anticoagulated Patients with Atrial Fibrillation Using the Simple ABC (Atrial Fibrillation Better Care) Pathway. Am J Med. 2018 Nov;131(11):1359-1366.e6.
36. Yoon M, Yang PS, Jang E, Yu HT, Kim TH, Uhm JS, et al. Improved Population-Based Clinical Outcomes of Patients with Atrial Fibrillation by Compliance with the Simple ABC (Atrial Fibrillation Better Care) Pathway for Integrated Care Management: A Nationwide Cohort Study. Thrombosis and Haemostasis. 2019;19(10):1695–703.
37. Pastori D, Pignatelli P, Menichelli D, Violi F, Lip GYH. Integrated Care Management of Patients With Atrial Fibrillation and Risk of Cardiovascular Events: The ABC (Atrial fibrillation Better Care) Pathway in the ATHERO-AF Study Cohort. Mayo Clin Proc. 2019 Jul;94(7):1261–7.
38. Pastori D, Farcomeni A, Pignatelli P, Violi F, Lip GY. ABC (Atrial fibrillation Better Care) Pathway and Healthcare Costs in Atrial Fibrillation: The ATHERO-AF Study. Am J Med. 2019 Jul;132(7):856–61.
39. Guo Y, Lane DA, Wang L, Zhang H, Wang H, Zhang W, et al. Mobile Health Technology to Improve Care for Patients With Atrial Fibrillation. J Am Coll Cardiol. 2020 Apr 7;75(13):1523–34.
40. Romiti GF, Pastori D, Rivera-Caravaca JM, Ding WY, Gue YX, Menichelli D, et al. Adherence to the 'Atrial Fibrillation Better Care' Pathway in Patients with Atrial Fibrillation: Impact on Clinical Outcomes—A Systematic Review and Meta-Analysis of 285,000 Patients. Thromb Haemost. 2022 Mar;122(03):406–14.
41. Page RL, Wilkinson WE, Clair WK, McCarthy EA, Pritchett EL. Asymptomatic arrhythmias in patients with symptomatic paroxysmal atrial fibrillation and paroxysmal supraventricular tachycardia. Circulation. 1994 Jan;89(1):224–7.
42. Boriani G, Laroche C, Diemberger I, Fantecchi E, Popescu MI, Rasmussen LH, et al. Asymptomatic Atrial Fibrillation: Clinical Correlates, Management, and Outcomes in the EORP-AF Pilot General Registry. The American Journal of Medicine. 2015 May 1;128(5):509-518.e2.
43. Strickberger SA, Ip J, Saksena S, Curry K, Bahnson TD, Ziegler PD. Relationship between atrial tachyarrhythmias and symptoms. Heart Rhythm. 2005 Feb;2(2):125–31.
44. Benjamin EJ, Levy D, Vaziri SM, D'Agostino RB, Belanger AJ, Wolf PA. Independent risk factors for atrial fibrillation in a population-based cohort. The Framingham Heart Study. JAMA. 1994 Mar 16;271(11):840–4.
45. Lip GYH, Nieuwlaat R, Pisters R, Lane DA, Crijns HJGM. 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. 2010 Feb;137(2):263–72.
46. Pisters R, Lane DA, Nieuwlaat R, de Vos CB, Crijns HJGM, Lip GYH. A novel user-friendly score (HAS-BLED) to assess 1-year risk of major bleeding in patients with atrial fibrillation: the Euro Heart Survey. Chest. 2010 Nov;138(5):1093–100.
47. Alonso A, Krijthe BP, Aspelund T, Stepas KA, Pencina MJ, Moser CB, et al. Simple Risk Model Predicts Incidence of Atrial Fibrillation in a Racially and Geographically Diverse Population: the CHARGE-AF Consortium. Journal of the American Heart Association. 2(2):e000102.
48. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. Am J Epidemiol. 1989 Apr;129(4):687–702.
49. Mahmood SS, Levy D, Vasan RS, Wang TJ. The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. The Lancet. 2014 Mar 15;383(9921):999–1008.
50. Nieuwlaat R, Capucci A, Camm AJ, Olsson SB, Andresen D, Davies DW, et al. Atrial fibrillation management: a prospective survey in ESC member countries: the Euro Heart Survey on Atrial Fibrillation. Eur Heart J. 2005 Nov;26(22):2422–34.
51. Li YG, Pastori D, Farcomeni A, Yang PS, Jang E, Joung B, et al. 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. 2019 Mar;155(3):510–8.
52. de Vos CB, Pisters R, Nieuwlaat R, Prins MH, Tieleman RG, Coelen RJS, et al. Progression from paroxysmal to persistent atrial fibrillation clinical correlates and prognosis. J Am Coll Cardiol. 2010 Feb 23;55(8):725–31.
53. Winkle RA, Jarman JWE, Mead RH, Engel G, Kong MH, Fleming W, et al. Predicting atrial fibrillation ablation outcome: The CAAP-AF score. Heart Rhythm. 2016 Nov;13(11):2119–25.
54. Kornej J, Hindricks G, Shoemaker MB, Husser D, Arya A, Sommer P, et al. The APPLE score: a novel and simple score for the prediction of rhythm outcomes after catheter ablation of atrial fibrillation. Clin Res Cardiol. 2015 Oct;104(10):871–6.
55. Mesquita J, Ferreira AM, Cavaco D, Moscoso Costa F, Carmo P, Marques H, et al. Development and validation of a risk score for predicting atrial fibrillation recurrence after a first catheter ablation procedure - ATLAS score. Europace. 2018 Nov 1;20(FI_3):f428–35.
56. Fox KAA, Lucas JE, Pieper KS, Bassand JP, Camm AJ, Fitzmaurice DA, et al. Improved risk stratification of patients with atrial fibrillation: an integrated GARFIELD-AF tool for the prediction of mortality, stroke and bleed in patients with and without anticoagulation. BMJ Open. 2017 Dec 21;7(12):e017157.
57. Singer DE, Chang Y, Borowsky LH, Fang MC, Pomernacki NK, Udaltsova N, et al. A New Risk Scheme to Predict Ischemic Stroke and Other Thromboembolism in Atrial Fibrillation: The ATRIA Study Stroke Risk Score. Journal of the American Heart Association. 2(3):e000250.
58. Hijazi Z, Lindbäck J, Alexander JH, Hanna M, Held C, Hylek EM, et al. The ABC (age, biomarkers, clinical history) stroke risk score: a biomarker-based risk score for predicting stroke in atrial fibrillation. Eur Heart J. 2016 May 21;37(20):1582–90.
59. Gage BF, Waterman AD, Shannon W, Boechler M, Rich MW, Radford MJ. Validation of clinical classification schemes for predicting stroke: results from the National Registry of Atrial Fibrillation. JAMA. 2001 Jun 13;285(22):2864–70.
60. O'Brien EC, Simon DN, Thomas LE, Hylek EM, Gersh BJ, Ansell JE, et al. The ORBIT bleeding score: a simple bedside score to assess bleeding risk in atrial fibrillation. Eur

Heart J. 2015 Dec 7;36(46):3258–64.
61. Fang MC, Go AS, Chang Y, Borowsky LH, Pomernacki NK, Udaltsova N, et al. A new risk scheme to predict warfarin-associated hemorrhage: The ATRIA (Anticoagulation and Risk Factors in Atrial Fibrillation) Study. J Am Coll Cardiol. 2011 Jul 19;58(4):395–401.
62. Bisson Arnaud, Clementy Nicolas, Bodin Alexandre, Angoulvant Denis, Babuty Dominique, Lip Gregory Y.H., et al. Relationship of Preexisting Cardiovascular Comorbidities to Newly Diagnosed Atrial Fibrillation After Ischemic Stroke. Stroke. 2017 Oct 1;48(10):2878–80.
63. Guichard JB, Nattel S. Atrial Cardiomyopathy: A Useful Notion in Cardiac Disease Management or a Passing Fad? Journal of the American College of Cardiology. 2017 Aug 8;70(6):756–65.
64. Kreimer F, Gotzmann M. Left Atrial Cardiomyopathy – A Challenging Diagnosis. Front Cardiovasc Med. 2022 Jun 30;9:942385.
65. Greenland S, Poole C. Living with p values: resurrecting a Bayesian perspective on frequentist statistics. Epidemiology. 2013 Jan;24(1):62–8.
66. Wiemken TL, Kelley RR. Machine Learning in Epidemiology and Health Outcomes Research. Annu Rev Public Health. 2020 Apr 2;41:21–36.
67. Bellman R. Adaptive Control Processes: A Guided Tour. Princeton Univ Press. 2015;
68. Fauchier L, Clementy N, Bisson A, Ivanes F, Angoulvant D, Babuty D, et al. Should Atrial Fibrillation Patients With Only 1 Nongender-Related CHA2DS2-VASc Risk Factor Be Anticoagulated? Stroke. 2016;47(7):1831–6.
69. Chao TF, Joung B, Takahashi Y, Lim TW, Choi EK, Chan YH, et al. 2021 Focused Update Consensus Guidelines of the Asia Pacific Heart Rhythm Society on Stroke Prevention in Atrial Fibrillation: Executive Summary. Thromb Haemost. 2022 Jan;122(1):20–47.
70. Grymonprez M, Steurbaut S, De Sutter A, Lahousse L. Impact of a single non-sex-related stroke risk factor on atrial fibrillation and oral anticoagulant outcomes: a systematic review and meta-analysis. Open Heart. 2020 Dec 23;7(2):e001465.
71. Edmiston MK, Lewis WR. Bleeding Risk Scores in Atrial Fibrillation: Helpful or Harmful? Journal of the American Heart Association. 2018 Sep 18;7(18):e010582.
72. Turner BJ, Cuttler L. The Complexity of Measuring Clinical Complexity. Ann Intern Med. 2011 Dec 20;155(12):851.
73. Proietti M, Romiti GF, Olshansky B, Lane DA, Lip GYH. Comprehensive Management With the ABC (Atrial Fibrillation Better Care) Pathway in Clinically Complex Patients With Atrial Fibrillation: A Post Hoc Ancillary Analysis From the AFFIRM Trial. Journal of the American Heart Association. 2020 May 18;9(10):e014932.
74. Wu VCC, Wu M, Aboyans V, Chang SH, Chen SW, Chen MC, et al. Female sex as a risk factor for ischaemic stroke varies with age in patients with atrial fibrillation. Heart. 2020 Apr;106(7):534–40.
75. Tomasdottir M, Friberg L, Hijazi Z, Lindbäck J, Oldgren J. Risk of ischemic stroke and utility of CHA2DS2-VASc score in women and men with atrial fibrillation. Clinical Cardiology. 2019;42(10):1003–9.
76. Friberg L, Benson L, Rosenqvist M, Lip GYH. Assessment of female sex as a risk factor in atrial fibrillation in Sweden: nationwide retrospective cohort study. BMJ. 2012 May 31;344:e3522.
77. Shen AYJ, Yao JF, Brar SS, Jorgensen MB, Chen W. Racial/ethnic differences in the risk of intracranial hemorrhage among patients with atrial fibrillation. J Am Coll Cardiol. 2007 Jul 24;50(4):309–15.
78. Wang KL, Lip GYH, Lin SJ, Chiang CE. Non-Vitamin K Antagonist Oral Anticoagulants for Stroke Prevention in Asian Patients With Nonvalvular Atrial Fibrillation: Meta-Analysis. Stroke. 2015 Sep;46(9):2555–61.

79. Chao TF, Lip GYH, Liu CJ, Tuan TC, Chen SJ, Wang KL, et al. Validation of a Modified CHA2DS2-VASc Score for Stroke Risk Stratification in Asian Patients With Atrial Fibrillation. Stroke. 2016 Oct;47(10):2462–9.
80. Chao TF, Wang KL, Liu CJ, Lin YJ, Chang SL, Lo LW, et al. Age Threshold for Increased Stroke Risk Among Patients With Atrial Fibrillation: A Nationwide Cohort Study From Taiwan. J Am Coll Cardiol. 2015 Sep 22;66(12):1339–47.
81. 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, Camm AJ, Kirchhof P, Lip GYH, et al. 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. 2010 Oct 1;31(19):2369–429.
82. Chen Q, Toorop MMA, Tops LF, Lijfering WM, Cannegieter SC. Time Trends in Patient Characteristics, Anticoagulation Treatment, and Prognosis of Incident Nonvalvular Atrial Fibrillation in the Netherlands. JAMA Network Open. 2023 Apr 25;6(4):e239973.
83. Connolly SJ, Ezekowitz MD, Yusuf S, Eikelboom J, Oldgren J, Parekh A, et al. Dabigatran versus Warfarin in Patients with Atrial Fibrillation. New England Journal of Medicine. 2009 Sep 17;361(12):1139–51.
84. Bezabhe WM, Bereznicki LR, Radford J, Wimmer BC, Curtain C, Salahudeen MS, et al. Factors influencing oral anticoagulant use in patients newly diagnosed with atrial fibrillation. European Journal of Clinical Investigation. 2021;51(5):e13457.
85. Gorczyca-Głowacka I, Bielecka B, Wałek P, Chrapek M, Ciba-Stemplewska A, Jelonek O, et al. Temporal Trends in Oral Anticoagulant Prescription in Atrial Fibrillation Patients between 2004 and 2019. Int J Environ Res Public Health. 2022 May 4;19(9):5584.
86. Gabet A, Chatignoux E, Grave C, Béjot Y, Olié V. Épidémiologie de la fibrillation atriale en France. Incidence des patients nouvellement traités par un anticoagulant oral, évolutions temporelles et disparités territoriales. Bull Epidémiol Hebd. 2021;(5):72–81.
87. Apostolakis S, Lane DA, Guo Y, Buller H, Lip GYH. Performance of the HEMORR(2)HAGES, ATRIA, and HAS-BLED bleeding risk-prediction scores in patients with atrial fibrillation undergoing anticoagulation: the AMADEUS (evaluating the use of SR34006 compared to warfarin or acenocoumarol in patients with atrial fibrillation) study. J Am Coll Cardiol. 2012 Aug 28;60(9):861–7.
88. Lip GYH, Skjøth F, Nielsen PB, Kjældgaard JN, Larsen TB. The HAS-BLED, ATRIA, and ORBIT Bleeding Scores in Atrial Fibrillation Patients Using Non-Vitamin K Antagonist Oral Anticoagulants. Am J Med. 2018 May;131(5):574.e13-574.e27.
89. Wiener N. Cybernetics or control and communication in the animal and the machine. 2. ed., 10. print. Cambridge, Mass: MIT Press; 2000. 212 p.
90. McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics. 1943 Dec 1;5(4):115–33.
91. Turing AM. I.—COMPUTING MACHINERY AND INTELLIGENCE. Mind. 1950 Oct 1;LIX(236):433–60.
92. Samuel AL. Some Studies in Machine Learning Using the Game of Checkers. IBM Journal of Research and Development. 1959 Jul;3(3):210–29.
93. Moore GE. Cramming more components onto integrated circuits, Reprinted from Electronics, volume 38, number 8, April 19, 1965, pp.114 ff. IEEE Solid-State Circuits Society Newsletter. 2006 Sep;11(3):33–5.
94. Shalf J. The future of computing beyond Moore's Law. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2020 Jan 20;378(2166):20190061.
95. Sung J, Hopper JL. Co-evolution of epidemiology and artificial intelligence:

challenges and opportunities. International Journal of Epidemiology. 2023 Aug 1;52(4):969–73.
96. Bi Q, Goodman KE, Kaminsky J, Lessler J. What is Machine Learning? A Primer for the Epidemiologist. Am J Epidemiol. 2019 Dec 31;188(12):2222–39.
97. Rasmussen CE, Williams CKI. Gaussian Processes for Machine Learning. The MIT Press; 2005.
98. Breiman L. Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). Statistical Science. 2001 Aug;16(3):199–231.
99. Bisson A, Bodin A, Herbert J, Lacour T, Saint Etienne C, Pierre B, et al. Pacemaker Implantation After Balloon- or Self-Expandable Transcatheter Aortic Valve Replacement in Patients With Aortic Stenosis. J Am Heart Assoc. 2020 May 5;9(9):e015896.
100. Bisson A, Bodin A, Clementy N, Babuty D, Lip GYH, Fauchier L. Prediction of Incident Atrial Fibrillation According to Gender in Patients With Ischemic Stroke From a Nationwide Cohort. Am J Cardiol. 2018 Feb 15;121(4):437–44.
101. Buuren S van, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software. 2011 Dec 12;45:1–67.
102. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research. 2002 Jun 1;16:321–57.
103. Wiemken TL, Kelley RR, Fernandez-Botran R, Mattingly WA, Arnold FW, Furmanek SP, et al. Using cluster analysis of cytokines to identify patterns of inflammation in hospitalized patients with community-acquired pneumonia: a pilot study. Univ Louisville J Respir Infect. 2017;1(1):3–11.
104. Bergstra J, Bengio Y. Random Search for Hyper-Parameter Optimization. J Mach Learn Res. 2012;13:281–305.
105. Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognition. 1997 Jul 1;30(7):1145–59.
106. Olier I, Ortega-Martorell S, Pieroni M, Lip GYH. How machine learning is impacting research in atrial fibrillation: implications for risk prediction and future management. Cardiovascular Research. 2021 Jun 15;117(7):1700–17.
107. Wegner FK, Plagwitz L, Doldi F, Ellermann C, Willy K, Wolfes J, et al. Machine learning in the detection and management of atrial fibrillation. Clin Res Cardiol. 2022 Sep;111(9):1010–7.
108. Hemingway H, Asselbergs FW, Danesh J, Dobson R, Maniadakis N, Maggioni A, et al. Big data from electronic health records for early and late translational cardiovascular research: challenges and potential. Eur Heart J. 2018 Apr 21;39(16):1481–95.
109. Khurshid S, Kartoun U, Ashburner JM, Trinquart L, Philippakis A, Khera AV, et al. Performance of Atrial Fibrillation Risk Prediction Models in Over 4 Million Individuals. Circulation: Arrhythmia and Electrophysiology. 2021 Jan;14(1):e008997.
110. Perez MV, Mahaffey KW, Hedlin H, Rumsfeld JS, Garcia A, Ferris T, et al. Large-Scale Assessment of a Smartwatch to Identify Atrial Fibrillation. N Engl J Med. 2019 Nov 14;381(20):1909–17.
111. Bisson A, Lemrini Y, El-Bouri W, Bodin A, Angoulvant D, Lip GYH, et al. Prediction of incident atrial fibrillation in post-stroke patients using machine learning: a French nationwide study. Clin Res Cardiol. 2023 Jun;112(6):815–23.
112. Bisson A, Lemrini Y, Romiti GF, Proietti M, Angoulvant D, Bentounes S, et al. Prediction of early death after atrial fibrillation diagnosis using a machine learning approach: A French nationwide cohort study. Am Heart J. 2023 Aug 16;S0002-8703(23)00203-X.
113. Li YG, Bisson A, Bodin A, Herbert J, Grammatico-Guillon L, Joung B, et al. C2 HEST Score and Prediction of Incident Atrial Fibrillation in Poststroke Patients: A French Nationwide Study. J Am Heart Assoc. 2019 Jul 2;8(13):e012546.

114. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 1987;40(5):373–83.
115. Gilbert T, Neuburger J, Kraindler J, Keeble E, Smith P, Ariti C, et al. Development and validation of a Hospital Frailty Risk Score focusing on older people in acute care settings using electronic hospital records: an observational study. The Lancet. 2018 May 5;391(10132):1775–82.
116. Bisson A, Fawzy AM, El-Bouri W, Angoulvant D, Lip GYH, Fauchier L, et al. Clinical Phenotypes and Atrial Fibrillation Recurrences After Catheter Ablation: An Unsupervised Cluster Analysis. Curr Probl Cardiol. 2023 Mar 31;48(8):101732.
117. Bisson A, M Fawzy A, Romiti GF, Proietti M, Angoulvant D, El-Bouri W, et al. Phenotypes and outcomes in non-anticoagulated patients with atrial fibrillation: An unsupervised cluster analysis. Arch Cardiovasc Dis. 2023;116(6–7):342–51.
118. Bisson A, Bodin A, Clementy N, Bernard A, Babuty D, Lip GYH, et al. Stroke, thromboembolism and bleeding in patients with atrial fibrillation according to the EHRA valvular heart disease classification. Int J Cardiol. 2018 01;260:93–8.
119. Fauchier G, Bisson A, Bodin A, Herbert J, Semaan C, Angoulvant D, et al. Metabolically healthy obesity and cardiovascular events: A nationwide cohort study. Diabetes Obes Metab. 2021 Nov;23(11):2492–501.
120. Tseng AS, Noseworthy PA. Prediction of Atrial Fibrillation Using Machine Learning: A Review. Front Physiol. 2021;12:752317.
121. Siontis Konstantinos C., Yao Xiaoxi, Pirruccello James P., Philippakis Anthony A., Noseworthy Peter A. How Will Machine Learning Inform the Clinical Care of Atrial Fibrillation? Circulation Research. 2020 Jun 19;127(1):155–69.
122. Gautam N, Ghanta SN, Clausen A, Saluja P, Sivakumar K, Dhar G, et al. Contemporary Applications of Machine Learning for Device Therapy in Heart Failure. JACC: Heart Failure. 2022 Sep 1;10(9):603–22.
123. Savage N. Breaking into the black box of artificial intelligence. Nature [Internet]. 2022 Mar 29 [cited 2023 Aug 17]; Available from: https://www.nature.com/articles/d41586-022-00858-1
124. Lu J, Hutchens R, Hung J, Bennamoun M, McQuillan B, Briffa T, et al. Performance of multilabel machine learning models and risk stratification schemas for predicting stroke and bleeding risk in patients with non-valvular atrial fibrillation. Comput Biol Med. 2022 Sep 22;150:106126.

Arnaud BISSON

MACHINE LEARNING

ET

EPIDEMIOLOGIE DE LA FIBRILLATION ATRIALE

Résumé :
La fibrillation atriale est l'arythmie la plus fréquente et est associée à une morbi-mortalité importante. Les patients atteints de fibrillation atriale constituent une population hétérogène en termes de caractéristiques mais aussi en termes de pronostic. Les modèles statistiques conventionnels permettent d'explorer et de prédire la survenue d'évènements cardio-vasculaires chez les patients atteints de FA. Cependant, ces modèles sont limités et leurs performances ne sont pas optimales particulièrement avec les données recueillies actuellement.

Les nouvelles approches de type machine learning pourraient rendre possible une meilleure compréhension de nos données et une amélioration de nos modèles prédictifs.

Après une revue de la littérature, nous présentons dans ce travail le développement, l'application et l'évaluation de modèles de machine learning supervisés dans la prédiction d'événements indésirables liés à la fibrillation atriale à travers 3 études réalisées partir de différents jeux de données de patients atteints de fibrillation atriale au niveau national et international. Nous rapportons ensuite l'application de modèles non supervisés de clustering dans la recherche de phénogroupes cliniquement pertinents à travers 2 études réalisées sur des bases de données locales.

Par ce travail, nous montrons l'intérêt des méthodes de machine learning, par la supériorité des modèles prédictifs développés par rapports aux modèles classiques, par la description de phénogroupes de patients cliniquement pertinents et par l'éclairage nouveau apporté par nos données permettant l'émergence de nouvelles questions et hypothèses.
Mots clés : Fibrillation atriale ; Machine learning ; Épidémiologie.

Summary:
Atrial fibrillation is the most common arrhythmia and is associated with significant morbidity and mortality. Patients suffering from atrial fibrillation constitute a heterogeneous population in terms of both characteristics and prognosis. Conventional statistical models can be used to explore and predict the occurrence of cardiovascular events in patients with AF. However, these models are limited and their performance is not optimal, particularly with the data currently being collected.

New machine learning approaches could provide a better understanding of our data and improve our predictive models.

After a review of the literature, we present in this work the development, application and evaluation of supervised machine learning models in the prediction of adverse events related to atrial fibrillation through 3 studies carried out using different datasets of patients with atrial fibrillation at national and international level. We then report on the application of unsupervised clustering models in the search for clinically relevant phenogroups from 2 studies based on local databases.

Through this work, we demonstrate the value of machine learning approach, through the superiority of the predictive models developed when compared with conventional models, through the description of clinically relevant phenogroups of patients and through the new light shed on our data, allowing new questions and hypotheses to emerge.
Keywords: Atrial Fibrillation; Machine Learning; Epidemiology.