Title: Predicting Atrial Fibrillation Ablation Outcomes: A Machine Learning Approach Leveraging a Large Administrative Claims Database

Authors: Yijun Liu,[a,b] Mustapha Oloko-Oba PhD,[c] Kathryn Wood RN PhD,[c] Michael S. Lloyd MD,[d] Joyce C. Ho PhD,[a*] Vicki Stover Hertzberg PhD[c,a]

Author Affiliations:

[a]Department of Computer Science, Emory University

400 Dowman Drive

Atlanta, GA 30322

[b]Department of Quantitative Theory and Methods, Emory University

36 Eagle Row

Atlanta, GA 30322

[c]Nell Hodgson Woodruff School of Nursing, Emory University

1520 Clifton Rd. NE

Atlanta, GA 30322

[d]Division of Cardiology, Emory University

100 Woodruff Circle

Atlanta, GA 30322

Funding: This work was supported by grant R21HL156184 (PI: Vicki Stover Hertzberg) from the National Institutes of Health to Emory University, Atlanta, Georgia.

Abstract

Background: Atrial fibrillation (AF) ablation is an effective treatment for reducing episodes and improving quality of life in patients with AF. However, in some patients there are only modest long-term AF-free rates after AF ablation. There is a need to address the limited benefits some patients experience by developing predictive algorithms to improve AF ablation outcomes.

Objective: The authors aim to utilize machine learning models on claims data to explore if innovative coding models may lead to better patient outcomes than use of traditional stroke risk score prediction.

Methods: The Merative MarketScan® Research Medicare data was used to examine claims for AF ablation. To predict 1-year AF-free outcomes after AF ablation, logistic regression and XGBoost models were used. Model predictions were compared with established risk scores $CHADS_2$ and CHA_2DS_2-VASC. These models were also assessed on subgroups of patients with paroxysmal AF, persistent AF, and both AF and atrial flutter from October 2015 onwards.

Results: The sample included 14,521 patients with claims for AF ablation. XGBoost achieved an area under the receiver operating characteristic curve (AUC) of 0.525, 0.521, and 0.527 for the entire AF ablation population, female, and male, respectively. Within the subgroups, machine learning models performed the best for the paroxysmal AF subgroup using ICD codes, demographic information, and comorbidity indexes, achieving an AUC of 0.546.

Conclusion: Machine learning models outperformed $CHADS_2$ and CHA_2DS_2-VASC in all AF ablation patient groups (whole population, female, and male). Using patient data for those who had their AF ablation on or after October 2015, machine learning models performed best in all

subgroups and the population, indicating that including ICD codes in machine learning models may improve performance.

Keywords: logistic regression, XGBoost, $CHADS_2$, CHA_2DS_2-VASC, atrial fibrillation, atrial fibrillation ablation

Abbreviations

Abbreviation	**Definition**
AF	Atrial fibrillation
AUC	Area under the curve
CCI	Charlson comorbidity index
Com	Comorbidity Indices
Demo	Demographic characteristic(s)
ECG	Electrocardiogram
ECI	Elixhauser comorbidity index
EHR	Electronic health record(s)
I Table	Inpatient Admission Table
ICD	International Classification of Disease
ML	Machine learning
O Table	Outpatient Services Table
ROC	Receiver operating characteristic
S Table	Inpatient Services Table
SD	Standard Deviation
US	United States
XGB	XGBoost
XGBoost	Extreme Gradient Boosting

Introduction

Affecting over 6 million people in the U.S., atrial fibrillation (AF), the most common cardiac arrhythmia, is a major public health concern that will continue to increase with the aging United States (US) population.[1-4] AF is costly to health care systems and leads to significant health consequences of stroke, heart failure, dementia, and decreased quality of life.[5] Although there is no cure for AF currently, AF ablation is the most effective treatment to restore normal sinus rhythm and decrease symptoms in episodes of paroxysmal or persistent AF, thereby reducing AF burden and improving quality of life.[1-3] AF is associated with increased risk of cardiovascular events that may affect treatment outcomes, yet there limited knowledge of personalized risk stratification for patients undergoing AF ablation. Existing risk scores, like $CHADS_2$ and CHA_2DS_2-VASc, have traditionally been applied to predict stroke risk and are now also utilized in risk assessment following COVID-19, heart surgery, and AF ablation.[6-13]

Success rates for AF ablation from the literature vary based on individual clinical variables such as type of AF, left atrial size or volume index,[1,2,4] yet these variables are frequently difficult to access in large electronic health record datasets (EHR). Patients continue to experience episodes of AF following an initial AF ablation with long-term AF-free rates after de novo AF ablation reported as 50% to 75%.[1,2,14] Additionally, the chances of developing any complications after AF ablation range around 6%, with 0.1% to 0.9% of patients experiencing complications that could result in death.[15-17] Given the modest success rates of AF ablation, prediction of outcomes could be personalized to more easily identify those who would be most likely to benefit from an AF ablation.

Machine learning (ML) has emerged as a powerful approach that leverages increased computational power with large datasets to help achieve complex decisions to guide clinical practice.[18] Artificial intelligence and ML have been used in the field of electrophysiology since the 1970s for automated ECG interpretation.[18,19] More recently, innovations in algorithms, development and labeling of large databases, and improvements in hardware and software have rapidly increased the role of ML in cardiac electrophysiology and cardiovascular imaging to identify predictors of patient outcomes.[20] ML has already been used to improve prediction of AF ablation outcomes, primarily via electronic health records (EHRs). Nevertheless, health systems are not widely interoperable,[21] making it both costly and challenging to extend these prognostic tools across multiple health systems. Studies utilizing EHR data have often been limited to datasets from one to two hospitals, limiting the generalizability of the models and hindering broad adoption.[22,23] Claims data, on the other hand, is commonly collected, more readily available, and usually collected on a large national scale.[21] Although EHR data, which can include medications, laboratory data, and radiology reports, is more granular than claims data and can offer more accurate predictions, claims-based prognostic models can offer better scalability across health systems and populations.[24]

In this study, we propose to develop ML-based predictive models for outcomes of de novo AF ablation procedures using national level claims data in the United States. Our goal is to evaluate a ML derived risk prediction model for AF ablation patient outcomes. We hypothesize that the ML models will be comparable to or exceed existing AF risk scores with respect to predictive power. Existing risk scores including $CHADS_2$ and CHA_2DS_2-VASc have achieved non-trivial improvements in predicting the outcomes of AF procedures, achieving area under the receiver operating characteristic curve (AUROC) of 0.785 and 0.830 respectively in a dataset consists of

565 patients.[25] Thus, in this study, we utilize $CHADS_2$ and CHA_2DS_2-VASc as a baseline to compare with our ML approaches. We will also characterize outcomes by subgroups as well as examine differences in AF ablation outcomes by sex.

Methods

This research leveraged de-identified claims data sourced from the Merative MarketScan® Research Medicare Databases (MarketScan, Inc., Ann Arbor, MI) to investigate AF ablation (CPT code '93656') claims spanning January 1, 2013, to December 31, 2020. MarketScan contains claims made for individuals with Medicare Supplemental and Medicare Advantage plans.

Inclusion criteria comprised patients across inpatient admission (I), inpatient services (S), and outpatient services (O) tables in the database. Medical history and postoperative outcomes were scrutinized using claims data from January 1, 2011, to December 31, 2021, with a focus on patients possessing valid patient IDs for traceability. Each patient's medical history included all ICD codes from visits within the two years before the initial occurrence of AF ablation. The final cohort comprised 14,521 patients after excluding those without valid patient IDs. The research was conducted at Emory University, Atlanta, GA.

Patient Population

We included patients undergoing AF ablation (CPT code '93656') between January 1, 2013, and December 31, 2020, using the I, S, and O tables from Medicare claims. Only the I, S, and O tables have ICD codes which are the primary factors we used for our predictions. These patients

were validated to avoid CPT coding errors by verifying the co-presence of AF (ICD-9 code of '427.31' or ICD-10 code of 'I48.X') and AF ablation.

The decision to include exclusively Medicare patients stemmed from the MarketScan database limitations preventing integration of commercial and Medicare claims, as the unique patient IDs differ between the Medicare and commercial claims datasets. Additionally, mean age for first AF ablation usually happens around 55-62.[2-4] Medicare therefore covered a greater range of aged patients in the database. Moreover, the absence of postoperative outcomes for numerous patients in the commercial database rendered it unsuitable for this study.

Definition

Our study's objective is to forecast the binary outcome—success or failure—of AF ablation prior to the procedure based on patient past medical history and demographics. Success entailed the absence of recurring AF or AF ablation within the 6-month interval from the end of the sixth to the end of the twelfth month following the initial AF ablation procedure. Conversely, failure was defined by any occurrence of recurring AF or AF ablation during the same 6-month period. The patient population was further validated by confirming that they had revisited the clinics within a year after ablation to ensure the success population was a true success.

While the O datasets comprehensively documented the operation date for AF ablation, the I and S datasets exclusively provided admission and discharge dates. In navigating this constraint, we pragmatically designated the admission date from the I and S datasets as a surrogate for the AF ablation operation date in our analysis. This procedural adjustment was necessitated by the dataset structure, allowing us to maintain temporal coherence in our predictive modeling.

The study employed subgroup analysis, where we identified groups with paroxysmal AF, persistent AF, and AF patients with flutter. ICD-10 codes clearly distinguish between paroxysmal AF, persistent AF, and chronic AF, in comparison to ICD-9 codes that do not make these distinctions. Within the Medicare datasets, all patient records after October 1, 2015, adopted ICD-10 codes instead of ICD-9 codes. To reflect current changes in terminology of types of AF, we combined persistent AF and chronic AF in the database as persistent AF. We defined paroxysmal AF as patients with ICD-10 code of I48.0, I48.20, and free of persistent AF code; persistent AF as patients with I48.1, I48.11, I48.19, I48.2, and I48.21. Patients with atrial flutter had any AF code with I48.3 (ICD-10) or I48.4 (ICD-10) or 427.32 (ICD-9).

Data Processing

Each patient was distinctly identified by a unique patient ID that facilitates cross-referencing across datasets, namely the I, S, and O tables, housing medical claims for individual medical visits. Our predictive focus necessitated a retrospective analysis, limited to the patient's history preceding the date of AF ablation.

To present a comprehensive historical snapshot, we constructed a table encompassing patient ID, demographic details (sex, region, age, and industry), the earliest AF ablation date, the failure date (if applicable), and an exhaustive list of ICD and CPT codes from past medical visits. This retrospective tableau extended back over a 2-year period.

In adapting to Medicare's transition from ICD-9 to ICD-10 post-October 2015, we integrated the conversion from ICD-10 to ICD-9 codes using the ICD-10 Lookup tool.[26] To simplify our dataset and maintain a manageable feature set, we retained only the first three digits of the converted ICD codes.

Within the claims data, we calculated two established indices—the Charlson Comorbidity Index (CCI) and the Elixhauser Comorbidity Index (ECI)—to capture patients' comorbid conditions.[27,28] These indices used a weighted system based on specific conditions to provide a score, with higher values indicating more severe comorbidities.

Modeling

We used two popular and efficient supervised ML classifiers: Logistic Regression and XGBoost (eXtreme Gradient Boosting) (XGB).[29] Logistic Regression computes the probability of a binary outcome by employing a logistic function (sigmoid curve) to transform the linear combination of input features into probabilities. This model is particularly advantageous due to its simplicity and interpretability, especially in scenarios where the relationship between input variables and the outcome is expected to be linear. On the other hand, XGB represents a more sophisticated approach. As an ensemble technique, it constructs multiple decision trees in a sequential manner, with each subsequent tree focusing on addressing the errors made by its predecessors. This method does not presuppose a linear relationship between input and output variables, offering greater flexibility and efficacy in dealing with larger and more intricate datasets. Despite its computational intensity, XGB is celebrated for its high efficiency and versatility, making it a potent tool in predictive modeling, especially in situations where the complexity of the data surpasses the capabilities of simpler models like Logistic Regression.[24]

The $CHADS_2$ and CHA_2DS_2-VASC risk scores have been widely used to predict stroke risk in patients with AF.[12,13,25] These risk scores more recently have been used to predict outcomes in patients with AF, heart failure, coronary artery disease, and postoperative AF undergoing cardiovascular surgical procedures.[6,11] Still other investigators have reported that the $CHADS_2$

and CHA_2DS_2-VASC risk scores were useful predictors of adverse events after AF ablation.[7] Therefore, we chose to use $CHADS_2$ and CHA_2DS_2-VASC risk score as our baseline comparison.

Statistical Analysis

For the continuous variable age, we used a t-test to obtain the p-value. For categorical variables, we used Chi-square tests. Continuous variables were reported as the mean ± standard deviation, while categorical variables were expressed as percentages. The area under receiver operating characteristic curve (AUC) was used to assess the performance of each model.

For two-point statistics, bootstrap resampling was used to generate a distribution and then perform a t-test to obtain the p-values. Specifically, bootstrap resampling was only used to obtain the p-values in table 2.

Ethical Considerations

This study used commercially available data that have been deidentified. As such, the study was deemed exempt by Emory University Institutional Review Board.

Results

The demographic and clinical profiles of patients with AF are detailed in Tables 1 and 2. Our study cohort consisted of 14,521 patients, with an average age of 71.5 years (SD = 5.31). A successful outcome from AF ablation procedures was observed in 54.01% (n=7,843) of patients. Females constituted 39.94% of the study population. Clinically, 24.73% (n=3,591) of the patients were diagnosed with concomitant atrial flutter. The precise identification of patients with

paroxysmal and persistent AF was limited, relative to the total cohort, due to the use of ICD-9 instead of ICD-10 prior to October 2015. A total of 7,646 patients were identified using ICD-10 codes for AF ablation, demonstrating a slightly reduced AF ablation success rate of 53.28% in comparison to the broader patient population. A subset of 6,983 patients was accurately categorized as having paroxysmal or persistent AF. Within this subset, 37.63% (n=2,877) were diagnosed with paroxysmal AF, while 53.70% (n=4,106) had persistent AF. The AF ablation success rates for paroxysmal and persistent AF were 52.55% and 53.90%, respectively.

Table 3 showed the comparative performance, as measured by the Area Under the Receiver Operating Characteristic Curve (AUC ROC), of the XGB machine learning model and the risk scores, which were $CHADS_2$ and CHA_2DS_2VASc, in predicting the outcomes of AF ablation. $CHADS_2$ and CHA_2DS_2VASc are stroke prevention risk scores, with $CHADS_2$ involving congestive heart failure, hypertension, age ≥75, diabetes, stroke (doubled) and CHA_2DS_2VASc involving congestive heart failure, hypertension, age ≥75 (doubled), diabetes, stroke (doubled), vascular disease, age 65 to 74 and sex category (female). Notably, both $CHADS_2$ and CHA_2DS_2VASc scores yielded predictions worse than random chance (AUC ROC < 0.5) except for CHA_2DS_2VASc in female group (which ROC = 0.500, equaled to the chance of random guessing). The XGB model exhibited modest predictive accuracy across the entire study population with an AUC ROC of 0.525. This accuracy was slightly improved (0.527) for males but marginally decreased (0.521) for females. Nevertheless, the XGB model surpassed the predictive capabilities of the $CHADS_2$ and CHA_2DS_2VASc scores across all patient cohorts.

Table 4 presented type of atrial arrhythmia subgroup analyses focusing on patients with ICD-10 codes (patients' AF ablation dates were on or after October 2015). ICD-10 codes are more specific and enables patients to be identified as having paroxysmal AF or persistent AF. In table

4, we presented results for patients with paroxysmal AF, persistent AF, and those who had both AF and atrial flutter (in ICD-10 space), utilizing three distinct datasets: 1) comprising converted, simplified 3-digit ICD codes, demographic information, and comorbidity scales (Charlson Comorbidity Index and Elixhauser Comorbidity Scale); 2) demographic data and comorbidity scales; and 3) solely demographic information. Across all subgroups and the entire ICD-10 population, models based on the first dataset type demonstrated superior predictive performance. The highest AUC ROC was observed in the entire ICD-10 population, achieving 0.550. Among subgroups, the highest AUC ROC was found in patients with paroxysmal AF (0.546) when ICD codes were included. For the first dataset type, the AUC ROC values for paroxysmal AF, persistent AF, AF patients with atrial flutter, and the entire population were 0.546, 0.526, 0.521, and 0.550, respectively. The lowest AUC ROC in each subgroup occurred for AF patients with atrial flutter using only demographic data (0.511). Similarly, for paroxysmal AF and the entire ICD-10 population, the lowest AUC ROC (both 0.534) was observed using demographic data alone. In contrast, the lowest AUC-ROC for persistent AF (0.516) was observed when using demographics combined with comorbidity scales.

Discussion

In this study, we have developed ML learning models that predict outcomes of de novo AF ablation procedures. Our XGB model demonstrated improved performance and predictive power as compared to $CHADS_2$ and CHA_2DS_2VASc in all patient and sex subgroups. However, XGB's predictive ability of outcomes after AF ablation was found to be lower in females than it was in males or in the entire population. There was no difference in predictive power when comparing $CHADS_2$ to CHA2DS2-VASc risk scores for outcomes after AF ablation except for females where CHA_2DS_2-VASc performs better than $CHADS_2$. Additionally, in comparing

outcomes in patient subgroups of types of AF (paroxysmal, persistent, or AF with atrial flutter) the ICD codes model demonstrated better prediction power than using either only demographic/clinical variables or those variables plus comorbidity scales, the Charlson Comorbidity Index (CCI) and the Elixhauser Comorbidity Index (ECI). Use of these machine learning models may be useful in clinical practice in patient selection for AF ablation in the future.

Previous clinical models for predicting AF ablation success have reported an AUC ranging from 0.55 to 0.65, with only three models achieving an AUC of 0.75.[8-10] In other studies, $CHADS_2$ and CHA_2DS_2-VASc can usually achieve an AUC of 0.785 and 0.830 respectively in predicting AF ablation.[25] However, in our study, $CHADS_2$ and CHA_2DS_2-VASc only achieved an AUC of 0.498-0.5, performing mostly worse than random guessing (AUC = 0.5). This finding highlights the difficulty in predicting the success and failure of AF ablation for the dataset used in this study. In contrast, our machine learning (ML) models achieved an AUC of 0.521-0.527, outperforming both $CHADS_2$ and CHA_2DS_2-VASc with statistical significance (p-value = 0.000) in all patient groups (whole population, female, and male) for this dataset.

In addition to demonstrating that ML models outperform these risk score predictions, we have also conducted an analysis to understand what types of features should be included in the ML models. We identified three sets of features and three clinically meaningful subgroups in the population (and the whole population). The three sets of features respectively were: 1) comprised with ICD codes, demographic information, and comorbidity scales; 2) comprised with demographic data and comorbidity scales; and 3) restricted to demographic information; and the three clinically meaningful subgroups are: paroxysmal AF, persistent AF, as well as AF patients with atrial flutter. These subgroups could only be identified in the ICD-10 space, and thereby we

also compared these subgroup results with the ICD-10 population. This resulted in 16 unique ML models for each combination of feature sets and subgroups. Across all subgroups, ML models performed best when including ICD codes as features, indicating that the incorporation of ICD codes improved model performance. Among the three subgroups (paroxysmal AF, persistent AF, and patients with atrial flutter), the ML models performed best for paroxysmal AF patients and AF patients with atrial flutter have the least success. The entire ICD-10 population achieved the highest overall AUC ROC compared to other subgroups, which was likely due to the larger sample size.

Our findings demonstrate that ML models using ICD codes to estimate AF ablation procedural outcomes are robust and valid across all populations. Improvement of outcome predictions for AF ablation using ML has the potential for widespread use in research and clinical practice to determine optimal patient selection for AF ablation and AF patient management. Advances in artificial intelligence and ML technology have an ability to rapidly analyze and synthesize innumerable variables to predict outcomes of AF ablation and discover new patterns of clinical variables that greatly surpass prior conventional methods of gauging success. These findings will be important to consider as healthcare policymakers struggle to allocate limited resources to as many patients as possible and search for ways to improve patient outcomes. ML technologies will play increasingly more important roles in medicine with future advances and as we better learn how to incorporate ML for improvements in clinical practice and patient outcomes.

Limitations

This data used in this study comprised health insurance claims data filed for people holding Medicare Advantage or Medicare Supplemental plans. As with all administrative data, some

human errors may be present. We addressed these issues by reporting clear descriptions of our data source, the inclusion and exclusion criteria, the billing codes used, and any potential confounding factors included in the modeling. We also verified the patients with AF ablation by identifying at least one past occurrence of AF in their past medical history.

Acknowledgment

This research is supported by R21HL156184.

No generative AI models (including Large Language Models such as ChatGPT) were used in the conduct of this research nor in the writing of this manuscript. We use python scikit-learn module to perform logistic regression and XGB which formed the body of this research.[30]

Supplemental Materials

The code used for this study is publicly released in the GitHub Repository:

https://github.com/isSherrrrry/AFA-Claims-CodeRelease

References

1. Andrade JG, Champagne J, Dubuc M, et al. Cryoballoon or radiofrequency ablation for atrial fibrillation assessed by continuous monitoring: a randomized clinical trial. *Circulation*. 2019;140(22):1779-1788.

2. Joglar JA, Chung MK, Armbruster AL, et al. 2023 ACC/AHA/ACCP/HRS guideline for the diagnosis and management of atrial fibrillation: a report of the american college of cardiology/american heart association joint committee on clinical practice guidelines. *Circulation*. 2024;149(1):e1-e156.

3. Martin SS, Aday AW, Almarzooq ZI, et al. 2024 heart disease and stroke statistics: a report of US and global data from the American Heart Association. *Circulation*. 2024;149(3):e347-e913.

4. Van Gelder IC, Rienstra M, Bunting KV, et al. 2024 ESC guidelines for the management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS). *European Heart Journal*. 2024;45(23):3314-3414.

5. Chugh SS, Roth GA, Gillum RF, Mensah GA. Global burden of atrial fibrillation in developed and developing nations. *Global Heart*. 2014;9(1):113-119.

6. Zheng JY, Li DT, Qiu YG, et al. CHA2DS2-VASc score as a prognostic indicator in patients with atrial fibrillation undergoing coronary stenting. *Turkish Journal of Medical Sciences*. 2022;52(4):1103-1110.

7. Mulder MJ, Kemme MJB, Hopman LHGA, et al. Comparison of the predictive value of ten risk scores for outcomes of atrial fibrillation patients undergoing radiofrequency pulmonary vein isolation. *International Journal of Cardiology*. 2021;334:103-110.

8. Jacobs V, May HT, Bair TL, et al. The impact of risk score (CHADS2 versus CHA2DS2-VASc) on long-term outcomes after atrial fibrillation ablation. *Heart Rhythm*. 2015;12(4):681-686.

9. Kornej J, Hindricks G, Arya A, Sommer P, Husser D, Bollmann A. The APPLE score: a novel score for the prediction of rhythm outcomes after repeat catheter ablation of atrial fibrillation. *PLoS ONE*. 2017;12(1):e0169933.

10. Kosich F, Schumacher K, Potpara TS, Lip GYH, Hindricks G, Kornej J. Clinical scores used for the prediction of negative events in patients undergoing catheter ablation for atrial fibrillation. *Clinical Cardiology*. 2019;42(2):320-329.

11. Shuvy M, Zwas DR, Keren A, Gotsman I. Value of the CHA2DS2-VASc score for predicting outcome in patients with heart failure. *ESC Heart Failure*. 2020;7(5):2553-2560.

12. 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;137(2):263-272.

13. Chen J, Zhang AD, Lu HY, Guo J, Wang F, Li ZZ. CHADS2 versus CHA2DS2-VASc score in assessing the stroke and thromboembolism risk stratification in patients with atrial fibrillation: a systematic review and meta-analysis. *Journal of Geriatric Cardiology*. 2013;10(3):258-266.

14. Clarnette J, Brooks AG, Mahajan R, et al. Outcomes of persistent and long-standing persistent atrial fibrillation ablation: a systematic review and meta-analysis. *Europace*. 2018;20(FI_3):f366-f376.

15. Ngo L, Ali A, Ganesan A, Woodman R, Adams R, Ranasinghe I. Ten-year trends in mortality and complications following catheter ablation of atrial fibrillation. *European Heart Journal Quality Care and Clinical Outcomes*. 2022;8(4):398-408.

16. Ngo L, Ali A, Ganesan A, Woodman R, Adams R, Ranasinghe I. Gender differences in complications following catheter ablation of atrial fibrillation. *European Heart Journal Quality Care and Clinical Outcomes*. 2021;7(5):458-467.

17. Rehman KA, Wazni OM, Barakat AF, et al. Life-threatening complications of atrial fibrillation ablation: 16-year experience in a large prospective tertiary care cohort. *JACC: Clinical Electrophysiology*. 2019;5(3):284-291.

18. Feeny AK, Chung MK, Madabhushi A, et al. Artificial intelligence and machine learning in arrhythmias and cardiac electrophysiology. *Circulation: Arrhythmia and Electrophysiology*. 2020;13(8):e007952.

19. Nygårds ME, Hulting J. An automated system for ECG monitoring. *Computers and Biomedical Research*. 1979;12(2):181-202.

20. Dey D, Slomka PJ, Leeson P, et al. Artificial intelligence in cardiovascular imaging: JACC state-of-the-art review. *Journal of the American College of Cardiology*. 2019;73(11):1317-1335.

21. MacKay EJ, Stubna MD, Chivers C, et al. Application of machine learning approaches to administrative claims data to predict clinical outcomes in medical and surgical patient populations. *PLoS ONE*. 2021;16(6):e0252585.

22. Deng H, Shantsila A, Xue YM, et al. Using the MB-LATER score for predicting arrhythmia outcome after catheter ablation for atrial fibrillation: The Guangzhou atrial fibrillation project. *International Journal of Clinical Practice*. 2018;72(11):e13247.

23. Yuan N, Duffy G, Dhruva SS, et al. Deep learning of electrocardiograms in sinus rhythm from US veterans to predict atrial fibrillation. *JAMA Cardiology*. 2023;8(12):1131-1139.

24. Tseng AS, Noseworthy PA. Prediction of atrial fibrillation using machine learning: a review. *Frontiers in Physiology*. 2021;12:752317.

25. Chao TF, Lin YJ, Tsao HM, et al. CHADS(2) and CHA(2)DS(2)-VASc scores in the prediction of clinical outcomes in patients with atrial fibrillation after catheter ablation. *Journal of the American College of Cardiology*. 2011;58(23):2380-2385.

26. Lookup IC. ICD Code Lookup website. https://icdcodelookup.com

27. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of Chronic Diseases*. 1987;40(5):373-383.

28. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Medical Care*. 1998;36(1):8-27.

29. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016:785-794.

30. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *Journal of Machine Learning Research*. 2011;12(85):2825-2830.

Table 1. Baseline Demographic Characteristics of Patients undergoing Atrial Fibrillation

Demographic Variable	Overall (N=14521)	AF ablation Success (N=7843)	AF ablation Failure (N=6678)	p-value
Age (mean ± standard deviation)	71.5 ± 5.31	71.5 ± 5.27	71.6 ± 5.34	0.628
Female	5800 (39.94%)	3118 (39.76%)	2682 (40.17%)	0.630
Region				**0.000**
Northeast	2790 (19.21%)	1544 (19.69%)	1246 (18.66%)	0.122
North Central	4467 (30.76%)	2263 (28.85%)	2204 (33.00%)	0.000
South	4733 (32.59%)	2599 (33.14%)	2134 (31.96%)	0.134
West	2393 (16.48%)	1360 (17.34%)	1033 (15.47%)	0.003
Unknown	138 (0.95%)	77 (0.98%)	61 (0.91%)	0.736
Industry				**0.000**
Oil & Gas Extraction, Mining	6 (0.04%)	5 (0.06%)	1 (0.01%)	0.302
Manufacturing, Nondurable Goods	3013 (20.75%)	1486 (18.94%)	1527 (22.87%)	0.000
Manufacturing, Durable Goods	467 (3.21%)	254 (3.24%)	213 (3.19%)	0.905
Transportation, Communication, Utilities	1768 (12.18%)	1007 (12.84%)	761 (11.40%)	0.009
Retail Trade	42 (0.29%)	22 (0.28%)	20 (0.30%)	0.954
Finance, Insurance, Real Estate	661 (4.55%)	371 (4.73%)	290 (4.34%)	0.281
Services	2866 (19.74%)	1479 (18.86%)	1387 (20.77%)	0.004
Agriculture, Forestry, Fishing	4 (0.03%)	2 (0.03%)	2 (0.03%)	1.000
Construction	33 (0.23%)	20 (0.26%)	13 (0.19%)	0.558
Wholesale	54 (0.37%)	37 (0.47%)	17 (0.25%)	0.045
Unknown	5607 (38.61%)	3160 (40.29%)	2447 (36.64%)	0.000

Ablation.

The number in parentheses represents the percentage within the overall, success, or failure groups. Industry is categorized based on the employer responsible for the claim payment, and regions follow the Census Bureau's regional definitions.

Clinical Variable	Overall (N=14521)	AF ablation Success (N=7843)	AF ablation Failure (N=6678)	p-value
Charlson Comorbidity Index				0.292
0	4371 (30.10%)	2375 (30.28%)	1996 (29.89%)	0.607
1	4295 (29.58%)	2277 (29.03%)	2018 (30.22%)	0.872
>=2	5855 (40.32%)	3191 (40.68%)	2664 (39.89%)	0.611
Elixhauser Comorbidity Index				0.447
0	44 (0.30%)	24 (0.31%)	20 (0.30%)	1.000
1	989 (6.81%)	515 (6.57%)	474 (7.10%)	1.000
>=2	13488 (92.89%)	7304 (93.13%)	6184 (92.60%)	1.000
Both Atrial Flutter and AF (ICD-9 and ICD-10)	3591 (24.73%)	1963 (25.03%)	1628 (24.38%)	0.376
Patients with ICD-10	N=7646	N=4074	N=3572	0.511
Paroxysmal AF (ICD-10 only)	2877 (37.63%)	1512 (37.11%)	1365 (38.21%)	0.333
Persistent AF (ICD-10 only)	4106 (53.70%)	2213 (54.32%)	1893 (53.00%)	0.256
Unspecified AF	663 (8.67%)	349 (8.57%)	314 (8.79%)	0.759

Table 2. Baseline clinical characteristics of patients in sample undergoing AF ablation

The number in the brackets represents the percentage within the overall, success, or failure groups. The paroxysmal and persistent AF only exists in the ICD-10 space, which the overall ICD-10 population is 7646, success population is 4074, and failure population is 3572.

Table 3. Performance Comparison between XGB and $CHADS_2$ and CHA_2DS_2VASc Risk Scores Stratified by Sex).

	XGB (mean ± SD)	**$CHADS_2$ Scores**	**CHA_2DS_2VASc Scores**	**p-value (XGB C2)**	**p-value (XGB C2V)**	**p-value (C2 C2V)**
Population (n=14,521)	0.525 (0.522-0.527)	0.498	0.498	**0.000**	**0.000**	0.507
Female (n=5,800)	0.521 (0.517-0.525)	0.498	0.50	**0.000**	**0.000**	0.510
Male (n=8,721)	0.527 (0.524-0.531)	0.498	0.498	**0.000**	**0.000**	0.498

All performance values are reported in terms of AUC ROC. Cell values for XGB report average ± standard deviation (SD).

Table 4. ROC Performance by Clinical and Demographic Predictors Across Atrial Arrhythmia Subgroups.

	ICD+Demo+ (CCI+ECI)	**Demo+ (CCI+ECI)**	**Demo**
Paroxysmal AF (n=2,877)	0.546 (0.542-0.551)	0.543 (0.538-0.549)	0.534 (0.527-0.540)
Persistent AF (n=4,106)	0.526 (0.520-0.532)	0.516 (0.512-0.520)	0.525 (0.520-0.530)
ICD-10, With Atrial Flutter (n=1,503)	0.521 (0.514-0.530)	0.514 (0.504-0.525)	0.511 (0.504-0.519)
ICD-10 Population (n=7,646)	0.550 (0.545-0.554)	0.536 (0.533-0.541)	0.534 (0.531-0.537)

This population only includes patients who had their first AF ablation in or after October 2015. “ICD” means ICD codes of patients past medical history; “Demo” represents demographic variables (region, sex, age, and industry); the numbers in the brackets represent the confidence interval after 30 Monte Carlo runs.