Preparation of the Database for Machine Learning for Atrial Fibrillation Ablation

Yijun Liu, Hung-I Hsia, Mustapha Oloko-Oba, Kathryn Wood, Joyce Ho, and Vicki Hertzberg

Nell Hodgson Woodruff School of Nursing of Emory University

EMORY COLLEGE OF ARTS AND SCIENCES | Undergraduate Research Programs

INTRODUCTION

Affecting over 6 million people in the U.S., atrial fibrillation (AF), the most common cardiac arrhythmia, is a major public health concern.[1-4] AF is costly to the health care system and leads to significant health consequences (e.g., stroke, heart failure, dementia, decreased quality of life).[5] With time, AF patients experience increased frequency and duration of AF episodes. Many AF patients seek out atrial fibrillation ablation (AFA) in order to improve quality of life and decrease AF episodes. AFA, cauterization of areas of the left atrium, is the most effective treatment for persistent / paroxysmal AF.[1-3]

AFA success rates vary, but many patients will not be AF-free following AFA. At leading AFA centers, AF-free rates at one and two years after initial AFA were 40% and 37%, respectively.[3] Given the modest success rates of AFA, patient selection for this procedure should receive more attention. Sociodemographic and clinical phenotype data have been used to predict AFA response, but collectively they have poor predictive ability.[6]

The widespread adoption of electronic health record (EHR) systems presents a ripe opportunity for a paradigm shift for predicting AFA outcomes. A better understanding of patient specific factors predicting AFA outcome will inform patient selection for this procedure. To this end we propose to use machine learning techniques to develop predictive models for outcomes of primary AFA procedures.

METHODS – IBM® MarketScan®

In the study, we use the IBM® MarketScan® Research Databases. It is a comprehensive collection of deidentified patient-level healthcare data that integrate various sources, including electronic medical records that contain more than 245 million unique patients since 1995.[7]

In this study, we focus on utilizing the **Inpatient Admissions, Outpatient Services, and Inpatient Services** datasets from the IBM® MarketScan® Research Databases, specifically for Medicare patients due to their age. The three datasets contains the diagnosis and procedure codes (ICD-9/10) for AF and AFA. AFA has started since 2013, and the time range of our data is **2013-2021**.

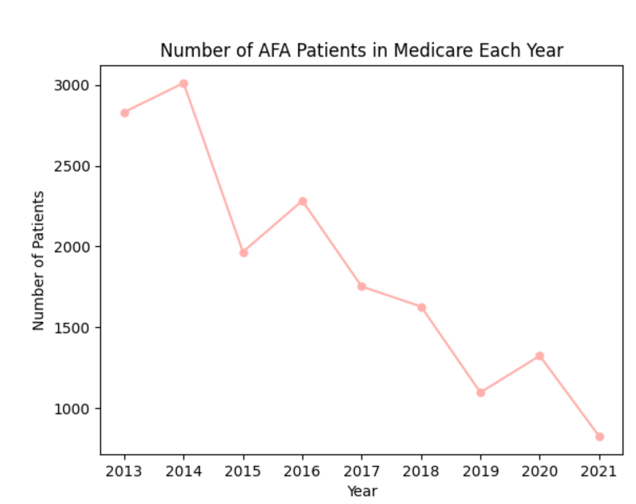
METHODS – Data Processing

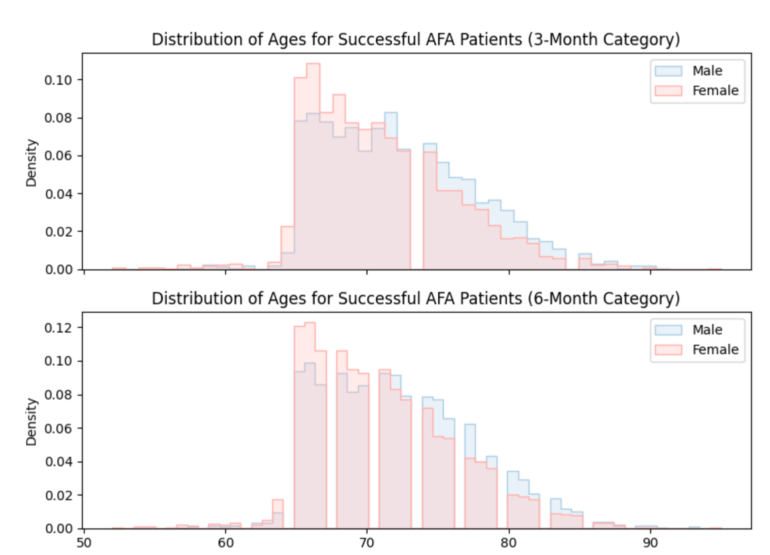
Failure: AF/AFA recurrences after 3 (or 6) months to 1 year of the procedure (AFA)
Success: Every AFA patient except the failure

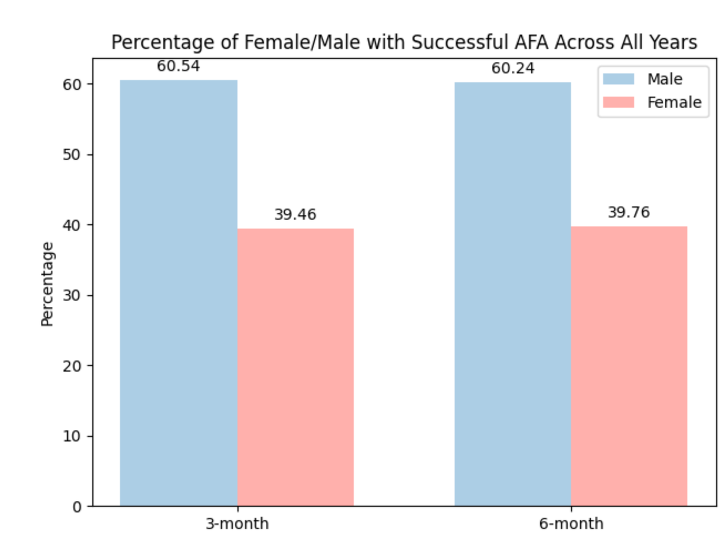
- Integrate and process Inpatient Admissions, Outpatient Services, and Inpatient Services datasets collectively.
- Modify Operation Date: Use discharge date minus one day as the operation date for AFA
- Verify AFA patients' prior AF diagnoses to address potential coding discrepancies.
- Categorize AFA patients' outcomes into success and failure datasets.
- Retrieve AFA patients' all medical records from 2011 to 2021.
- Identify the earliest diagnosis date for each AFA patient on the operation day, tracing records from 2011 to 2021.

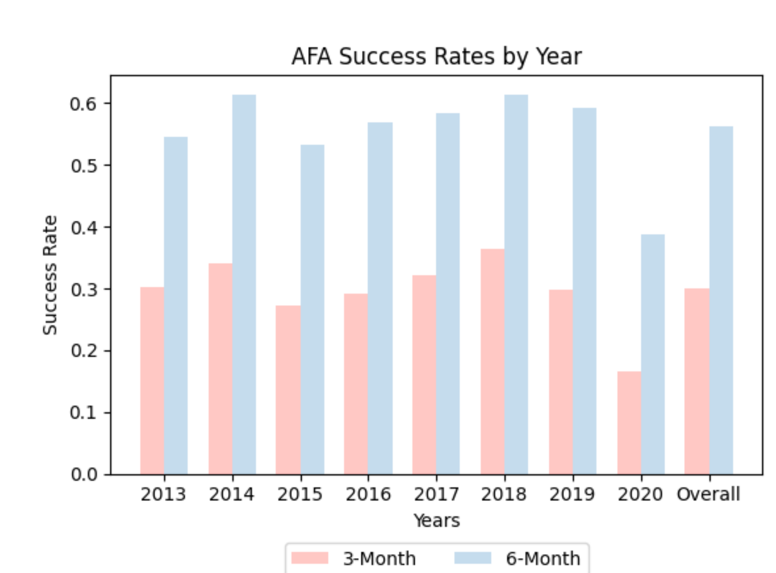
PRELIMINARY RESULTS

Total AFA Patients: 16724









FUTURE DIRECTIONS

Machine Learning (ML) Approaches

- Exploit EHR data for outcome prediction
- Conventional, supervised **ML** (logistic regression, support vector machines, decision trees, random forests, and gradient boosting decision trees)
  - Feature selection, dimensionality reduction, and varied observation window
- Explore deep learning models (CNNs, RNNs)

Evaluation Plan

- Discrimination metrics: AUC and AUPRC
- Calibration metrics: Brier score, scaled Brier score, Hosmer-Lemeshow test, etc.
- Monte Carlo cross-validation (70-30 train-test split)

Potential Limitations and Alternative Strategies

- Small sample size and overfitting
  - Regularization techniques
- Imbalanced classes
  - Sampling strategies (SMOTE) or cost-sensitive objective functions
- Missing data
  - Missing data mechanisms (dummy variables, imputation)

Expected Outcomes

- Impact of using existing ML algorithms to predict AFA outcomes
- Identify patient characteristics as predictors of AFA outcomes
- Improved risk prediction model for better decision-making

ACKNOWLEDGEMENT

This research is supported by R21HL156184.

REFERENCE

[1] H. Calkins, G. Hindricks, R. Cappato, Y.-H. Kim, E. B. Saad, L. Aguinaga, J. G. Akar, V. Badhwar, J. Brugada, J. Camm, P.-S. Chen, S.-A. Chen, M. K. Chung, J. C. Nielsen, A. B. Curtis, et al. 2017 HRS/EHRA/ECAS/APHRS/SOLAECE expert consensus statement on catheter and surgical ablation of atrial fibrillation. Heart Rhythm, 14(10):e275–e444, 2017.
[2] D. Mozaffarian, E. J. Benjamin, A. S. Go, D. K. Arnett, M. J. Blaha, M. Cushman, S. R. Das, S. De Ferranti, J. P. Després, H. J. Fullerton, V. J. Howard, M. D. Huffman, C. R. Isasi, M. C. Jiménez, S. E. Judd, et al. Executive summary: heart disease and stroke statistics-2016 update: a report from the American Heart Association. Circulation, 133(4):447–454, 2016.
[3] R. Cappato, H. Calkins, S.-A. Chen, W. Davies, Y. Iesaka, J. Kalman, Y.-H. Kim, G. Klein, A. Natale, D. Packer, and A. Skanes. Updated worldwide survey on the methods, efficacy, and safety of catheter ablation for human atrial fibrillation. Circulation: Arrhythmia and Electrophysiology, 3(1):32–38, 2010.
[4] C. T. January, L. S. Wann, J. S. Alpert, H. Calkins, J. E. Cigarroa, J. C. Cleveland Jr, J. B. Conti, P. T. Ellinor, M. D. Ezekowitz, M. E. Field, K. T. Murray, R. L. Sacco, W. G. Stevenson, P. J. Tchou, C. M. Tracy, and C. W. Yancy. 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation: executive summary: a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines and the Heart Rhythm Society. Journal of the American College of Cardiology, 64(21):2246–2280, 2014.
[5] S. S. Chugh, G. A. Roth, R. F. Gillum, and G. A. Mensah. Global burden of atrial fibrillation in developed and developing nations. Global Heart, 9(1):113–119, 2014.
[6] L. O'Neill, J. Harrison, M. O'Neill, and S. E. Williams. Clinical, electrophysiological and imaging predictors of atrial fibrillation ablation outcome. Expert Review of Cardiovascular Therapy, 15(4):289–305, 2017.
[7] IBM Watson Health. (n.d.). IBM MarketScan Research Databases for life sciences researchers. [White paper]. Retrieved from https://www.ibm.com/downloads/cas/0NKLE57Y