



Machine Learning Approach to Predict Risk of 90-Day Hospital Readmissions in Patients With Atrial Fibrillation: Implications for Quality Improvement in Healthcare

Man Hung^{1,2,3,4} , Eric S. Hon⁵, Evelyn Lauren², Julie Xu⁶, Gary Judd¹, and Weicong Su⁷

Abstract

Background: Atrial fibrillation (AF) in the elderly population is projected to increase over the next several decades. Catheter ablation shows promise as a treatment option and is becoming increasingly available. We examined 90-day hospital readmission for AF patients undergoing catheter ablation and utilized machine learning methods to explore the risk factors associated with these readmission trends.

Methods: Data from the 2013 Nationwide Readmissions Database on AF cases were used to predict 90-day readmissions for AF with catheter ablation. Multiple machine learning methods such as k-Nearest Neighbors, Decision Tree, and Support Vector Machine were employed to determine variable importance and build risk prediction models. Accuracy, precision, sensitivity, specificity, and area under the curve were compared for each model.

Results: The 90-day hospital readmission rate was 17.6%; the average age of the patients was 64.9 years; 62.9% of patients were male. Important variables in predicting 90-day hospital readmissions in patients with AF undergoing catheter ablation included the age of the patient, number of diagnoses on the patient's record, and the total number of discharges from a hospital. The k-Nearest Neighbor had the best performance with a prediction accuracy of 85%. This was closely followed by Decision Tree, but Support Vector Machine was less ideal.

Conclusions: Machine learning methods can produce accurate models in predicting hospital readmissions for patients with AF. The likelihood of readmission to the hospital increases as the patient age, total number of hospital discharges, and total number of patient diagnoses increase. Findings from this study can inform quality improvement in healthcare and in achieving patient-centered care.

Keywords

machine learning, atrial fibrillation, hospital readmission, quality improvement, patient-centered care

Introduction

Atrial fibrillation (AF) is growing in prevalence and is costly. Being the most common cardiac rhythm disorder,¹ it is estimated to impact 33.5 million individuals across the world,² with the number of cases projected to increase exponentially over the next several decades.³ In the United States, the number of AF cases is expected to double by 2050.⁴ The rapid increase is likely attributed to the growing elderly population in the world³ as AF is often associated with the aging process.⁵ The rapid rise in AF cases also lead to increased medical cost and resulted in a public health crisis. The annual cost for AF

¹ Roseman University of Health Sciences College of Dental Medicine, South Jordan, UT, USA

² University of Utah School of Medicine, Salt Lake City, UT, USA

³ Utah Center for Clinical and Translational Sciences, Salt Lake City, UT, USA

⁴ Huntsman Cancer Institute, Salt Lake City, UT, USA

⁵ University of Chicago Department of Economics, Chicago, IL, USA

⁶ University of Utah College of Nursing, Salt Lake City, UT, USA

⁷ University of Utah Department of Mathematics, Salt Lake City, UT, USA

Corresponding Author:

Man Hung, College of Dental Medicine, Roseman University of Health Sciences, 10894 South River Front Parkway, South Jordan, UT 84095, USA.

Email: mhung@roseman.edu



treatments in the United States was estimated at 6.65 billion in 2006⁶ and is expected to increase quickly over the next decades.

AF is a common cardiac arrhythmia with chaotic electrical activity showing in the atria, causing symptoms such as palpitations, shortness of breath, effort intolerance and fatigue,² and is associated with an increase in morbidity and mortality from heart failure, stroke, cognitive impairment⁷ and other thromboembolic conditions.⁸ Such conditions have contributed to lower quality of life in AF patients compared to the general population and other patients with coronary heart diseases.^{1,9} Catheter ablation is an increasingly widespread method of treatment for atrial fibrillation and has shown good outcomes.¹⁰ Using radio frequency or cryotherapy to electrically isolate the pulmonary veins and ablate arrhythmia foci¹¹ during catheter ablation have demonstrated improvement in atrial fibrillation-related symptoms and enhancement in health-related quality of life (HQoL).² Additionally, ablation also has measurable positive effects on risk of death, stroke, and dementia⁸ and is more effective in than anti-arrhythmic medications.¹²

To improve healthcare quality while also reducing healthcare costs, the Centers for Medicaid and Medicare Services (CMS) have developed the Hospital Readmission Reduction Program (HRRP), which penalizes healthcare providers that have high hospital readmission rates.^{13,14} Since the implementation of HRRP, readmission rates have been reduced by approximately 1%¹⁵; however, this may not be sufficient to prove that the HRRP program has caused a decline in hospital readmissions. More studies are needed. Additionally, with 2,592 out of 5,627 hospitals penalized in 2015 in the United States, the overall hospital readmission rate remains high.¹³ Comorbid AF conditions, such as heart attack and heart failure, are among the predominant hospitalization diagnoses being penalized by HRRP. Reducing HRRP penalties relies strongly upon understanding the reasons behind hospital readmissions of AF patients. This understanding is also critical for minimizing the rising healthcare costs incurred from raising AF cases.

Hospital readmission rate for patients with AF undergoing catheter ablation was reported to be approximately 10%.¹¹ Age, sex, primary payer, heart failure, hypertension, chronic renal disease, lung disease, and the number of AF hospitalizations during the prior years were significant predictors for 30-day hospital readmission.¹¹ While readmission rates for AF patients (10%)¹⁶ are comparatively lower than those with other conditions affected by HRRP penalties such as acute myocardial infarction (20%), heart failure (25%), and pneumonia (18%),¹⁷ the readmissions rates for AF patients undergoing catheter ablation was quite high at 16.5%.¹⁸ Even though CMS currently only tracks 30-day readmissions, it does not mean that 90-day readmissions are not important. It may mean that CMS wants to see studies evaluating the 90-day readmission rates in order to make an informed decision. Since past research has demonstrated that hospital readmissions during the first 90-days are actually very common,¹⁹ but 90-day hospital readmission predictors were less widely studied, there is a knowledge gap in the literature. Thus, there is a strong need for

carrying out this study to examine 90-day hospital readmission in AF.

Compared to the general population, AF patients are 3 times more likely to undergo multiple hospitalizations and they spend 73% more annually in direct medical costs, including Medicare payments.^{1,20} As expected AF cases will rise within the next few decades, so does the urgency to understand AF's risk factors and to create accurate models to predict AF. A clearer comprehension of AF will alleviate the impending economic and public health burden.

Past research regarding hospital readmissions have typically utilized traditional hypothesis-driven statistical techniques to identify the causal factors, which rely heavily on assumptions and have many limitations when the data are large.^{21,22} For example, traditional linear regression assumes homoscedasticity, independence of observations, normally distributed errors for each dependent variable's value and linear relationship between dependent and independent variables, etc. These assumptions are often very difficult to meet for traditional hypothesis-testing methods, especially when there exists a large number of variables and cases. Hospital readmission data typically consist of a large number of variables and cases and are susceptible to the limitations imposed by traditional hypothesis-driven techniques. However, machine learning is both an innovative and efficient method that allows a large amount of data to be processed efficiently without relying on traditional assumptions. This study aimed to use machine learning methods to develop prediction models of 90-day hospital readmissions for AF patients undergoing catheter ablation.

Methods

Data

The 2013 cycle of the Nationwide Readmissions Database (NRD) provided the raw data used in this study. The NRD was developed for the Healthcare Cost and Utilization Project (HCUP) to address the lack of nationally representative data on hospital readmissions for patients of all ages. It used HCUP State Inpatient Databases (SID) and the corresponding verified patient numbers to track patients within participating states a while following strict privacy guidelines. It included inpatients treated and discharged at community hospitals that were not rehabilitation or long-term acute care facilities. The 2013 NRD was created from 21 SID that contained geographically diverse information and contained 49.3% of the total population and 49.1% of all hospitalizations in the United States. Detailed information on NRD can be found at the HCUP website.²³

Outcome

The main outcome for this study was the 90-day hospital readmissions status. The NRD defined an index event as the starting point for analyzing repeat hospital visits, while hospital readmission was a subsequent inpatient admission within a

specified period of time. Thus, 90-day hospital readmission was defined as the index admission that had at least one readmission within 90 days after hospital discharge. It was a dichotomous variable with 1 representing a patient had one or more readmissions within 90 days after discharge and 0 otherwise. Further, as defined by CMS, to be considered as readmission, patients had to be readmitted to the same hospital or another applicable acute care hospital.

Demographics

Age, diagnosis, number of unique chronic conditions, patients' length of hospital stay, procedures reported for a patient on their discharge, gender, income, and primary payer were the demographic variables included in this study. Both weighted and unweighted prevalence estimates were calculated for the demographic characteristics. To compute the weighted demographic descriptive, the R 4.0.0 software Survey package was utilized. Clusters, stratum, and weights were incorporated into the data to produce nationally representative results.

Data Processing

Patients were identified via the primary diagnosis code of AF (427.31) using the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM),²⁴ and the primary or secondary procedure code of catheter ablation as 37.34. Patients who died during hospitalization, or were under the age of 18 years old, or had missing data on the length of hospital stay were excluded. For the 90-day hospital readmission status, patients discharged after September were excluded in order to allow for 90-day follow-up before the year ended in 2013. Patients having the following secondary diagnoses were excluded: atrial flutter, paroxysmal supraventricular tachycardia, atrioventricular nodal tachycardia, Wolff-Parkinson-White syndrome, paroxysmal ventricular tachycardia and ventricular premature beats.^{11,18} Additional exclusion criteria were patients with diagnosis or procedural codes showing prior or current implantation of pacemaker or implantable cardioverter-defibrillator and patients with open surgical ablations.^{11,18}

In order to prepare the data for variable selection, additional data processing was performed. Irrelevant variables such as patient IDs, key identifiers, and weighting variables were excluded. Variables with all cases missing were deleted. To ensure that the scales were consistent with all other variables, age and total hospital discharges were standardized. To prepare the data for machine learning, resampling methods were applied to the readmitted cases in order to adjust imbalanced data. Categorical data such as hospital bed size and discharge quarter were dummy coded to prevent the classifiers from incorrectly interpreting the variables as continuous data.

Variable Selection

With nearly 2,000 variables contained in the NRD, conducting variable selection to select a subset of top predictors was

needed and it could provide reduce computer storage requirements, machine learning model training times, and data dimensionality, which can lead to improved model performance.²⁵ Top predictor variables were chosen based on relative variable importance computed using random forest. Random forest is a well-established tree method for variable selection. It works by identifying a small number of relevant predictors that can produce a more parsimonious model but has a similar predictive performance of a logistic model.²⁶ Using random forest, the top 30 features (i.e., variables) were identified and ordered by their predictive performance. The 30 features were subsequently narrowed down into a simpler model with the top 6 features that had relatively high variable importance, which were then input into the machine learning to produce risk prediction models.

Machine Learning Algorithms

In conventional statistical approaches, a model is built and then input into a machine (e.g., computer).²⁷ This conventional approach is model-driven and heavily relies on assumptions about the shape of the data and these assumptions may be prone to bias and error. On the other hand, machine learning provides a data-driven approach in analyzing data. Machine learning inputs the data directly into the machine instead of imposing assumptions on the model itself. The goal of the machine is then to perform pattern recognition in order to "learn" and output a model observed in the data.²⁷ Such a data-driven approach is particularly efficient for analyzing large complex data such as those of hospital readmissions data, genomic data, imaging data or stock market data where patterns can be difficult to discern. Machine learning has great potential and implication in the public health for identification of healthcare needs as well as for crisis prediction and prevention.²⁸

We used supervised machine learning approaches for model development which included k-nearest neighbors (k-NN), support vector machine (SVM), and decision tree classifier. Supervised machine learning was chosen because the outcome of interest had already been identified (e.g., hospital readmission status of the patients).²⁹ K-NN, SVM, and decision tree are some of the most well-known and well-used machine learning methods to apply classification algorithms. Decision tree provides advantages of efficiency and flexibility that can lead to performance improvements, and it is used in a wide array of areas such as medical diagnosis, remote sensing, and speech recognition.³⁰ K-NN is widely used for pattern classification, and is very effective when the probability distribution of the input variables are unknown since it does not make probability assumption of the variables.³¹ Because it is well-matched for binary classification,³² SVM has been shown to work well with high dimensional data.³³ All of these methods have excellent ability to model non-linear relationship of the data observed in real-life situations. They are also easy to implement in clinical settings. The Python software version 3.8.3 scikit-learn package was used for machine learning modeling in this study.

Table 1. Demographic Characteristics of 90-Day Readmissions (Numbers Outside of the Parentheses Are Weighted; Numbers Inside the Parentheses Are Unweighted).

Variables	Mean	SD	Median	n	%
Age (year)	64.9(65.0)	11.5(11.3)	66(66)	10,547(4,922)	100
Number of Chronic Conditions	5.2(5.1)	2.7(2.7)	5(5)	10,547(4,922)	100
Number of Diagnosis	8.1(8.0)	4.8(4.7)	7(7)	10,547(4,922)	100
Number of Procedures	3.6(3.6)	1.6(1.6)	3(3)	10,547(4,922)	100
Length of Stay (days)	2.4(2.4)	3.0(2.9)	1(1)	10,547(4,922)	100
Gender					
Male				6,630(3,075)	62.9(62.5)
Female				3,917(1,847)	37.1(37.5)
Income					
0-25th percentile				2,042(956)	19.6(19.7)
26th to 50th percentile				2,558(1,179)	24.6(24.3)
51st to 75th percentile				2,771(1,256)	26.7(25.9)
76th to 100th percentile				3,021(1,451)	29.1(30.0)
Expected Primary Payer					
Medicare				5,943(2,789)	56.4(56.7)
Medicaid				333(158)	3.2(3.2)
Private Insurance				3,929(1,815)	37.3(36.9)
Self-pay				59(31)	0.6(0.6)
No charge				22(11)	0.2(0.2)
Other				258(117)	2.4(2.4)

To account for overfitting, the data were randomly split into a 60% training set and a 40% test set. Models were then applied to both the training and test sets, and their accuracies were recorded. We aimed to keep the difference of the accuracies between the training and the test sets to be no greater than 7%, to avoid overfitting of the data. When the data were overfitted, adjustments were made to the model parameters. Specifically, we applied L2 regularization to the model to overcome overfitting issues commonly occurred with k-NN methods. Per United States federal regulations (45 CFR 46, category 4), this is a secondary data analysis and the study does not require ethnic review since the data were deidentified and publicly available.

Results

For AF patients undergoing catheter ablation, there was a total of 9,468 (weighted N = 20,612) cases for the 90-day readmissions. After applying exclusion criteria and accounting for index admissions and death, there were 4,922 cases (weighted N = 10,547) remaining. The 90-day hospital readmission rate was 17.6%. The average age of the patients was 64.9 years old and there were 62.9% males (Table 1).

Figure 1 displays the relative variable importance score for the top 30 features. The higher the importance score, the more useful a feature in predicting the outcome. Patient's age was the most important feature among the top 30 features selected by random forest for determination of the likelihood for being readmitted (Figure 1). The patient's age, the total number of discharges from a hospital, the number of diagnoses a patient had at discharge, the number of chronic conditions a patient had at discharge, the number of procedures a patient had at

discharge, length of initial hospital stay, and gender were the top predictor variables identified for the 90-day readmissions.

With a predictive accuracy of approximately 85%, k-NN performed the best among the machine learning methods. It was followed by decision tree at 72.3% (Figure 2). SVM showed an accuracy of 62.6%. K-NN also had the highest positive predictive value (i.e., precision) at 0.875 (Figure 2). Overall, accuracy, precision sensitivity, specificity, and AUC were similar between k-NN and decision tree (Figure 3). K-NN and decision tree performed better than SVM.

Discussion

The purpose of this study was to predict 90-day hospital readmissions status for AF patients undergoing catheter ablation. Results demonstrated that machine learning methods were able to predict the occurrence of hospital readmissions at approximately 85% accuracy. The top predictors were: age, total discharges from hospital, number of diagnoses a patient had upon discharge, the number of chronic conditions a patient had upon discharge, the number of procedures on patient's record, length of hospital stay, and gender.

The cross-sectional nature of the NRD data is a limitation of the study. Using data from multiple years would allow development of potentially more accurate predictive models. Future studies may consider collecting longitudinal data to model prediction and confirm the results. Also, updates to ICD manuals and other healthcare references and tools have occurred after the data collection. While hospital characteristics are influential factors in predicting hospital readmissions for different conditions such as heart failure,³⁴ this does not seem to be as conclusive for atrial fibrillation readmission predictions.

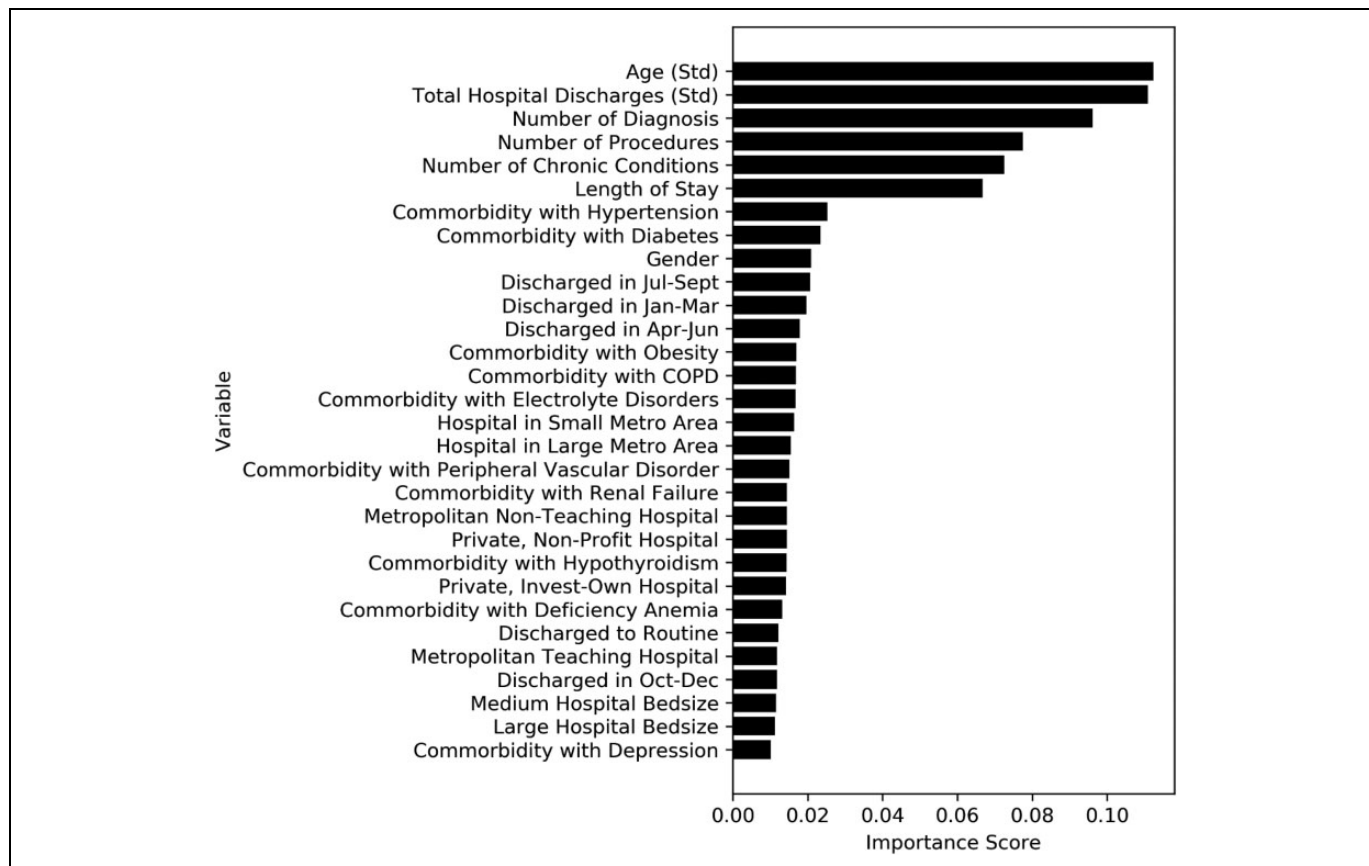


Figure 1. Relative variable importance of the top 30 features in predicting 90-day hospital readmissions in atrial fibrillation patients undergoing catheter ablation.

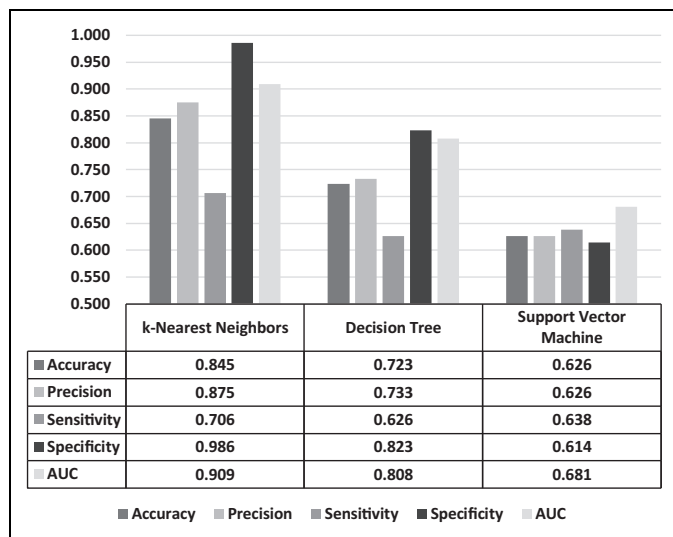


Figure 2. Performance metrics of machine learning models using the top 6 features (90-day readmissions).

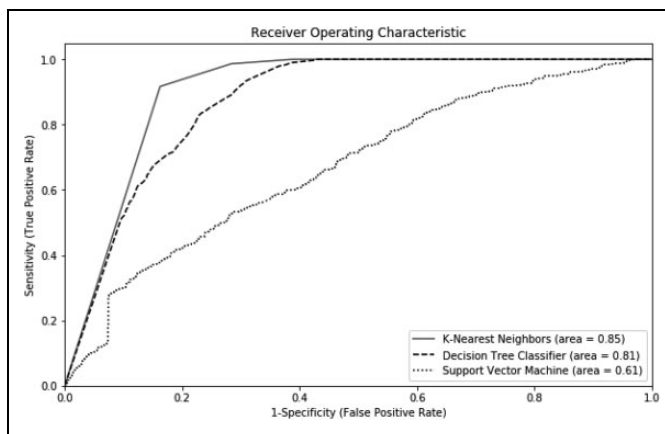


Figure 3. Receiver operating characteristic curves for the various machine learning methods (90-day readmissions).

Furthermore, translating findings into institutional policies can be difficult for hospitals without the adequate budget. For larger hospitals that have many beds, have academic affiliation, adequate staffing, and greater proportion of Medicare and

privately insured patients, readmissions prevention measures are normally more feasible. Outpatient management of AF may be relatively easy when compared with other cardiac conditions such as heart failure.

Another limitation is that large national databases such as NRD have a long latency period between the time of data collection and the time that data are available for the public to use and analyze. The results could have changed if there are

more recent data. Nonetheless, there are still insights gained from the data that can help design prospective investigations in the future.

Not all studies can be benefited by machine learning techniques. In general, large national or international data, or big data collected from any organizations have the potential to reap the benefits from machine learning to discover new insights, but small data do not. Traditional statistical approaches are regarded as model-driven, meaning a model is predetermined by preformulated hypotheses or existing theories and then it is being tested to determine whether the null hypotheses can be accepted or rejected. However, machine learning approaches are data-driven, meaning there is not a predetermined model. Thus, hypothesis testing or theory testing is not applicable. In fact, the goal of the machine learning is to discover new insights, new knowledge and new theories. Machine learning is relevant for organizations that want to gain more insights from their data to innovate and not do business as usual, and is useful for handling large and complex data when the relationships between the variables are not apparent.³⁵

Previous study demonstrated that older age and various comorbidities of patients who underwent AF ablation are factors independently associated with increased likelihood of 90-day readmissions,³⁶ which matched with our findings. Specifically, patients having 5 or more comorbidities were 2 times more likely to be readmitted within 90 days of initial hospital discharge. The literature had also reported gender, length of initial hospital stay, disposition to facility¹⁸ as well as number of chronic conditions³⁶ as the top predictors, which was consistent with our study findings. Furthermore, using machine learning our study was able to discover additional top predictors which were missed by prior studies that used traditional statistical approaches. These additional predictors include the total number of hospital discharges, the total number of diagnoses a patient had at discharge and the total number of procedures a patient had at discharge. The differences in analytical methods likely attribute to the discrepancies between our research and prior research. Prior research had utilized mainly traditional statistical methods for analysis. Using machine learning to conduct analyses can lead to an improved understanding of the data and an innovative opportunity for new frontiers of discovery.

Our models were able to reach a high predictive accuracy of 85% while using a supervised machine learning approach. Such models can be valuable for both policymakers and healthcare providers. Healthcare providers might find it helpful to look closely into a patient's record and provide patients with more personalized medical treatments to improve healthcare quality and minimize hospital readmissions. Applying these predictive models to assess hospital readmission risks can contribute to effective preventative treatments, lowering of medical costs, improvement in patient care, and having fewer mortalities.²⁸

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The authors sincerely thank the Clinical Outcome Research and Education at Roseman University College of Dental Medicine for supporting this study.

ORCID iD

Man Hung  <https://orcid.org/0000-0003-2827-3740>

References

1. Thrall G, Lane D, Carroll D, Lip GY. Quality of life in patients with atrial fibrillation: a systematic review. *Am J Med.* 2006; 119(5):448. e1-19.
2. Allan KS, Henry S, Aves T, et al. Comparison of health-related quality of life in patients with atrial fibrillation treated with catheter ablation or antiarrhythmic drug therapy: a systematic review and meta-analysis protocol. *BMJ Open.* 2017;7(8):e017577.
3. Ball J, Carrington MJ, McMurray JJ, Stewart S. Atrial fibrillation: profile and burden of an evolving epidemic in the 21st century. *Int J Cardiol.* 2013;167(5):1807-1824.
4. Go AS, Hylek EM, Phillips KA, et al. Prevalence of diagnosed atrial fibrillation in adults: national implications for rhythm management and stroke prevention: the AnTicoagulation and Risk Factors in Atrial Fibrillation (ATRIA) Study. *J Am Med Assoc.* 2001;285(18):2370-2375.
5. Tsang TS, Gersh BJ. Atrial fibrillation: an old disease, a new epidemic. *Am J Med.* 2002;113(5):432-435.
6. Coyne KS, Paramore C, Grandy S, Mercader M, Reynolds M, Zimetbaum P. Assessing the direct costs of treating nonvalvular atrial fibrillation in the United States. *Value Health.* 2006;9(5): 348-356.
7. Piccini JP, Sinner MF, Greiner MA, et al. Outcomes of Medicare beneficiaries undergoing catheter ablation for atrial fibrillation. *Circulation.* 2012;126(18):2200-2207.
8. Bunch TJ, Crandall BG, Weiss JP, et al. Patients treated with catheter ablation for atrial fibrillation have long-term rates of death, stroke, and dementia similar to patients without atrial fibrillation. *J Cardiovasc Electrophysiol.* 2011;22(8):839-845.
9. Thrall G, Lip GY, Carroll D, Lane D. Depression, anxiety, and quality of life in patients with atrial fibrillation. *Chest.* 2007; 132(4):1259-1264.
10. Weerasooriya R, Khairy P, Litalien J, et al. Catheter ablation for atrial fibrillation: are results maintained at 5 years of follow-up? *J Am Coll Cardiol.* 2011;57(2):160-166.
11. Shah RU, Freeman JV, Shilane D, Wang PJ, Go AS, Hlatky MA. Procedural complications, rehospitalizations, and repeat procedures after catheter ablation for atrial fibrillation. *J Am Coll Cardiol.* 2012;59(2):143-149.
12. Zhuang J, Lu Y, Tang K, Peng W, Xu Y. Influence of body mass index on recurrence and quality of life in atrial fibrillation patients after catheter ablation: a meta-analysis and systematic review. *Clin Cardiol.* 2013;36(5):269-275.
13. Shameer K, Johnson KW, Yahi A, et al. Predictive modeling of hospital readmission rates using electronic medical record-wide machine learning: a case-study using Mount Sinai Heart Failure

- Cohort. Paper presented at: Pacific Symposium on Biocomputing 2017; 2017. Fiaromont Orchid, Big Island of Hawaii.
14. Boccuti C, Casillas G. Aiming for fewer hospital U-turns: the Medicare hospital readmission reduction program. Policy Brief. 2015. Published March 2017, Henry J. Kaiser Family Foundation.
 15. McIlvannan CK, Eapen ZJ, Allen LA. Hospital readmissions reduction program. *Circulation*. 2015;131(20):1796-1803.
 16. Opolski G, Januskiewicz L, Szczerba E, et al. Readmissions and repeat procedures after catheter ablation for atrial fibrillation. *Cardiol J*. 2015;22(6):630-636.
 17. Krumholz HM, Lin Z, Keenan PS, et al. Relationship between hospital readmission and mortality rates for patients hospitalized with acute myocardial infarction, heart failure, or pneumonia. *J Am Med Assoc*. 2013;309(6):587-593.
 18. Arora S, Lahewala S, Tripathi B, et al. Causes and predictors of readmission in patients with atrial fibrillation undergoing catheter ablation: a national population-based cohort study. *J Am Heart Assoc*. 2018;7(12):e009294.
 19. Clarnette JA, 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.
 20. Kim MH, Johnston SS, Chu BC, Dalal MR, Schulman KL. Estimation of total incremental health care costs in patients with atrial fibrillation in the United States. *Circ Cardiovasc Qual Outcomes*. 2011;4(3):313-320.
 21. Hosseinzadeh A, Izadi MT, Verma A, Precup D, Buckeridge DL. Assessing the predictability of hospital readmission using machine learning. Paper presented at: IAAI, 2013, Bellvue, Washington, USA.
 22. Kuo CY, Yu LC, Chen HC, Chan CL. Comparison of models for the prediction of medical costs of spinal fusion in Taiwan diagnosis-related groups by machine learning algorithms. *Health Inform Res*. 2018;24(1):29-37.
 23. Project HCUP. Overview of the Nationwide Readmissions Database (NRD). 2017. Accessed August 27, 2020. www.hcup-us.ahrq.gov/nrdoverview.jsp.HCUP
 24. Centers for Medicare & Medicaid Services. *ICD-9-CM Diagnosis and Procedure Codes: Abbreviated and Full Code Title*. 2014. Accessed August 27, 2020. <https://www.cms.gov/Medicare/Coding/ICD9ProviderDiagnosticCodes/codes>
 25. Guyon I, Elisseeff A. An introduction to variable and feature selection. *J Mach Learn Res*. 2003;3(7-8):1157-1182.
 26. Genuer R, Poggi J-M, Tuleau-Malot C. Variable selection using random forests. *Pattern Recognit Lett*. 2010;31(14):2225-2236.
 27. Wiens J, Shenoy ES. Machine learning for healthcare: on the verge of a major shift in healthcare epidemiology. *Clin Infect Dis*. 2017;66(1):149-153.
 28. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. *Health Inform Sci Syst*. 2014;2(1):3.
 29. Kotsiantis SB, Zaharakis I, Pintelas P. Supervised machine learning: a review of classification techniques. *Emerg Artif Intellig Appl Comput Eng*. 2007;160:3-24.
 30. Safavian SR, Landgrebe D. A survey of decision tree classifier methodology. *IEEE Trans Syst Man Cybern*. 1991;21(3):660-674.
 31. Wu Y, Ianakiev K, Govindaraju V. Improved k-nearest neighbor classification. *Pattern Recognit*. 2002;35(10):2311-2318.
 32. Chapelle O, Haffner P, Vapnik VN. Support vector machines for histogram-based image classification. *IEEE Trans Neural Netw*. 1999;10(5):1055-1064.
 33. Furey TS, Cristianini N, Duffy N, Bednarski DW, Schummer M, Haussler D. Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics* 2000;16(10):906-914.
 34. Ziaiean B, Fonarow GC. The prevention of hospital readmissions in heart failure. *Prog Cardiovasc Dis*. 2016;58(4):379-385.
 35. Hung M, Voss MW, Rosales MN, et al. Application of machine learning for diagnostic prediction of root caries. *Gerodontology*. 2019;36(4):395-404. doi:10.1111/ger.12432
 36. Noseworthy PA, Kapa S, Haas LR, et al. Trends and predictors of readmission after catheter ablation for atrial fibrillation, 2009-2013. *Am Heart J*. 2015;170(3):483-489.