Cardiovascular Toxicology (2024) 24:365–374
https://doi.org/10.1007/s12012-024-09843-8

RESEARCH

Development and Validation of Machine Learning Algorithms to Predict 1-Year Ischemic Stroke and Bleeding Events in Patients with Atrial Fibrillation and Cancer

Bang Truong[1] · Jingyi Zheng[2] · Lori Hornsby[3] · Brent Fox[1] · Chiahung Chou[1] · Jingjing Qian[1]

Received: 15 January 2024 / Accepted: 21 February 2024 / Published online: 18 March 2024

Abstract

In this study, we leveraged machine learning (ML) approach to develop and validate new assessment tools for predicting stroke and bleeding among patients with atrial fibrillation (AFib) and cancer. We conducted a retrospective cohort study including patients who were newly diagnosed with AFib with a record of cancer from the 2012–2018 Surveillance, Epidemiology, and End Results (SEER)-Medicare database. The ML algorithms were developed and validated separately for each outcome by fitting elastic net, random forest (RF), extreme gradient boosting (XGBoost), support vector machine (SVM), and neural network models with tenfold cross-validation (train:test = 7:3). We obtained area under the curve (AUC), sensitivity, specificity, and F2 score as performance metrics. Model calibration was assessed using Brier score. In sensitivity analysis, we resampled data using Synthetic Minority Oversampling Technique (SMOTE). Among 18,388 patients with AFib and cancer, 523 (2.84%) had ischemic stroke and 221 (1.20%) had major bleeding within one year after AFib diagnosis. In prediction of ischemic stroke, RF significantly outperformed other ML models [AUC (0.916, 95% CI 0.887–0.945), sensitivity 0.868, specificity 0.801, F2 score 0.375, Brier score = 0.035]. However, the performance of ML algorithms in prediction of major bleeding was low with highest AUC achieved by RF (0.623, 95% CI 0.554–0.692). RF models performed better than CHA$_2$DS$_2$-VASc and HAS-BLED scores. SMOTE did not improve the performance of the ML algorithms. Our study demonstrated a promising application of ML in stroke prediction among patients with AFib and cancer. This tool may be leveraged in assisting clinicians to identify patients at high risk of stroke and optimize treatment decisions.

Keywords AFib · Cancer · Stroke · Bleeding · Prediction · Machine learning

Introduction

In the United States (US), atrial fibrillation (AFib) is projected to affect 12 million people by 2030 [1]. AFib has been recorded as the primary diagnosis for more than 454,000 hospitalizations and contributed to more than 158,000 deaths annually [2–4]. The coexistence of cancer among patients with AFib increases incidence of adverse events such as ischemic stroke, venous thromboembolism (VTE), bleeding, and death compared with AFib patients without cancer [5–8]. Current management of patients with AFib and cancer with oral anticoagulants (OACs) remains

Handling Editor: Vittorio Fineschi.

✉ Jingjing Qian
jzq0004@auburn.edu

Bang Truong
bct0022@auburn.edu

Jingyi Zheng
jzz0121@auburn.edu

Lori Hornsby
hornslb@auburn.edu

Brent Fox
foxbren@auburn.edu

Chiahung Chou
czc0109@auburn.edu

[1] Department of Health Outcomes Research and Policy, Auburn University Harrison College of Pharmacy, 4306d Walker Building, Auburn, AL 36849, USA

[2] Department of Mathematics and Statistics, Auburn University College of Sciences and Mathematics, Auburn, AL, USA

[3] Department of Pharmacy Practice, Auburn University Harrison College of Pharmacy, Auburn, AL, USA

suboptimal due to insufficient evidence regarding risk assessment and treatment optimization from clinical practice guidelines [9].

CHA_2DS_2-VASc score, a composite score of congestive heart failure (1 point), hypertension (1), age ≥ 75 (2), diabetes mellitus (1), prior stroke, TIA, or thromboembolism (2), vascular disease (e.g. peripheral artery disease, myocardial infarction, aortic plaque) (1), age 65–74 years, and sex category (1), has been used to evaluate of risk of stroke in patients with AFib [10, 11]. The clinical guidelines recommend OACs for patients with CHA_2DS_2-VASc scores ≥ 2 [11, 12]. However, CHA_2DS_2-VASc score is not highly predictive in patients with AFib and cancer [13, 14]. HAS-BLED score has been widely used for risk of bleeding stratification. The HAS-BLED is calculated by the presence of hypertension (1), abnormal renal/liver function (1 + 1), stroke (1), bleeding tendency or predisposition (1), labile INR for patients taking warfarin (1), age ≥ 65, drugs (concomitant aspirin or NSAIDs) or excess alcohol use (1 + 1) [15]. The 2020 European Society of Cardiology (ESC) guideline suggests a score of ≥ 3 indicates "high risk" [12]. However, it is not recommend against the use of anticoagulants, but caution and regular monitoring after treatment initiation are needed [12]. Nonetheless, the usefulness of HAS-BLED in cancer patients are inconclusive because cancer is an independent risk factor of bleeding among patients with AFib [16]. Pastori et al. compared the performances of multiple bleeding risk scores among cancer patients and found that HAS-BLED was not highly predictive of major and gastrointestinal bleeding [17].

Therefore, it is an urgent need to develop new risk assessment tools for stroke and bleeding in patients with AFib and cancer. Traditional risk assessment tools such as CHA_2DS_2-VASc and HAS-BLED are simple and easy for implementation among clinicians because they are linear combinations of patients' diseases and conditions. However, when the relationships between patients' characteristics and outcomes become more complicated, these tools may not perform well in patients with AFib and cancer. Recently, machine learning (ML) algorithms have been increasingly used to support clinical decision-making such as or identifying patients with dementia in primary care, anticoagulation monitoring, and measuring pretreatment quality of care before treatment in patients with hepatitis C [18–20]. Compared with conventional regression-based methods, ML models are able to learn from the data when the association between predictors and outcome variables is not linear. ML models have overperformed parametric regressions in handling high-dimensional data and interactions between variables in a complex data structure [20–22].

In this study, we developed and validated ML algorithms to predict risk of stroke and bleeding events among patients with AFib and cancer, using US cancer registry and administrative claims linked datasets.

Materials and Methods

Study Design and Data Source

We followed the Transparent Reporting of a multivariable prediction model for individual Prognosis Or Diagnosis (TRIPOD) guideline to develop and validate ML algorithms to separately predict risk of stroke and risk of bleeding in patients with AFib and cancer [23]. We conducted a retrospective cohort study using the 2011–2019 Surveillance, Epidemiology, and End Results (SEER) registry linked to Medicare database. SEER registry contains demographics, cancer characteristics, treatment, and follow-up of cancer patients across the US, [24] while Medicare data capture health care services utilization (medical claims, procedures, and prescriptions) of beneficiaries [25]. The study design and approach are illustrated in Figures S1, S2.

Participants

We included individuals aged ≥ 66, newly diagnosed non-valvular atrial fibrillation (NVAF) from 1/1/2012 to 12/31/2018. AFib was defined as any International Classification of Disease-9th Revision-Clinical Modification (ICD-9-CM) codes 427.31 or 427.32 or any International Classification of Disease-10th Revision-Clinical Modification (ICD-10-CM) codes I48.xx in any position on one Medicare inpatient claim or on two outpatient claims at least 7 days but < 1 year apart [26]. We removed patients with valvular diseases, repair or replacement, venous thromboembolism, or joint replacement during the 12 months baseline period because OACs are also indicated for these conditions and their clinical management are different from AFib [27, 28]. Eligible records were then linked to SEER files to identify patients with breast, lung, or prostate cancer—the most common cancer types with AFib—from at any time before the initial AFib diagnosis (ICD-O-3 codes C50.0–C50.9 for breast; C34.0, C34.1, C34.2, C34.3, C34.8, C34.9, C33.9 for lung; C61.9 for prostate cancer). Patients were required to continuously enroll in Medicare part A, B, D, and without Medicare Advantage or Health Maintenance Organization (HMO) for at least 12 months before and 12 months after NVAF diagnosis. Since OAC initiation during follow-up may modify the risk of stroke and bleeding, we excluded patients who initiated warfarin or direct anticoagulants (DOACs) within 12 months before or after NVAF diagnosis. All ICD codes to identify these conditions can be found in Table S1, Supplementary materials.

Outcomes

The outcomes of interest were ischemic stroke and major bleeding events identified within 12 months after AFib diagnosis. We defined major bleeding and ischemic stroke using validated algorithms defined by ICD-9-CM and ICD-10-CM codes in the primary diagnosis from Medicare medical claims files [29–31].

Predictors

We selected potential predictors from literature review and based on availability in SEER-Medicare data [29, 32, 33]. The following predictors were included: *demographics* (index age, sex, race/ethnicity, calendar year, geographical region, urbanicity), *socioeconomic factors* (household median income, percentage of household with education level below high school, and Medicaid eligibility), *comorbidities* (hypertension, congestive heart failure, diabetes, prior stroke, vascular diseases, prior bleeding, renal diseases, liver diseases, alcohol use disorders, asthma/chronic obstructive pulmonary disease, hematological disorders, dementia, depression, thrombocytopenia, acute kidney disease, peptic ulcer disease), *cancer characteristics* (time from cancer diagnosis to the onset of AFib, cancer type, cancer stage, tumor grade, active cancer status [29, 32]), *cancer treatment* (radiation, and cancer-directed surgery, and potentially interacting antineoplastic agents), and *medication history* (antiplatelet/non-steroidal anti-inflammatory drugs, angiotensin-converting enzyme (ACE) inhibitors/angiotensin II receptor blockers (ARBs), calcium channel blockers, beta blockers, antiarrhythmic medications, diuretics, statin, pump proton inhibitors, and serotonin reuptake inhibitors). Features were obtained during 12 months before the index date. All diagnosis codes and procedure codes for covariate ascertainment are described in Table S1, Supplementary materials.

Algorithms, Model Training and Validation

Descriptive statistics was used to compare the characteristics of the full cohort and between patients with and without the outcomes. MissForest was used to impute missing values for predictors [34, 35]. The original dataset was then randomly split into two datasets: training (70%) and testing datasets (30%) [36, 37], with similar distribution of the outcomes in both datasets. In the algorithm training process, ML models (elastic net logistic regression, random forest (RF), support vector machine (SVM), extreme gradient boosting (XGBoost), and neural network) were fitted with ten-fold cross-validation (CV) [38]. The fitted models were then tested on the rest of the data. Since stroke and bleeding occurred in less than 10% within one year among AFib patients [29, 39], our classification is severely imbalanced due to the prediction of minority class (stroke and bleeding) [40]. Therefore, we shifted the decision threshold to the true event probability rather than using the default threshold of 0.50 [40, 41]. The description of the models can be found in Technical Appendix.

Model Performance, Calibration, and Evaluation

To assess algorithm discrimination, we calculated the area under the receiver operating characteristic curve (AUROC or AUC) as the main metrics and compared the AUC across algorithms using DeLong's test [42]. Other performance metrics were also extracted, including sensitivity, specificity, and F2 score. Since true positive (patients actually having stroke/bleeding) is more important and false negative cases (patients at high risk of the event were not identified) are more costly, we selected F2 score over F1 score [41, 43]. In addition, we generated feature importance plots to identify the contribution of each variable in predicting the outcomes [44]. Since feature importance using Gini index in tree-based algorithms (i.e., RF and XGBoost) are subject to bias [45], we computed out-of-bag impurity reduction feature importance as an alternative [46]. Model calibration was performed to compare the true probability of the outcome versus a model's prediction with Brier score [47]. We also compared the performances of ML algorithms with CHA_2DS_2-VASc score or HAS-BLED score in predicting ischemic stroke and major bleeding, respectively. In predicting ischemic stroke, we fitted logistic regressions with CHA_2DS_2-VASc score as the predictor using the training data and validated the model on the testing data. Likewise, we predicted the risk of major bleeding on HAS-BLED score. Sensitivity, specificity, and AUC were obtained for these models. ML algorithms were developed using RStudio (version 3.6.2; Boston, MA, USA) and data analysis was conducted using SAS (version 9.4, SAS Institute, Inc., Cary, NC, USA).

Sensitivity Analyses

Since our classification problem was severely imbalanced, we used Synthetic Minority Oversampling Technique (SMOTE) to account for imbalance distribution of the outcome variables [48]. Model development and validation were conducted on the resampling dataset.

Results

Study Sample and Characteristics

The final cohort consisted of 18,388 patients, of whom 523 (2.84%) had ischemic stroke and 221 (1.20%) had major bleeding within one year after AFib diagnosis (Fig. 1). The characteristics of study sample are described in Table S2. Overall, the mean (standard deviation) age was 76.59 (7.13), 8483 (46.13%) were women, and the majority were White (85.11%) and residing in the Northeast (39.13%), or West (34.40%) region. The median (interquartile range) duration from cancer diagnosis to AFib onset was 17 (2–40) months. Compared with non-stroke patients, patients who had stroke were more likely to have breast cancer [(227 (43.40%) vs. 5416 (30.32%)], use potential interaction agents [139 (26.58%) vs. 825 (21.41%)], diabetes [188 (35.95%) vs. 5433 (30.41)], history of stroke [96 (18.36) vs. 1446 (8.09)], and vascular diseases [136 (26.00) vs. 4249 (23.78)] but less likely to have lung cancer [106 (20.27%) vs 6059 (33.92%)]. Compared with non-bleeding patients, patients who had bleeding were more likely to have breast cancer [84 (38.01%) vs. 5559 (30.60%)], history stroke [53 (23.98%) vs. 1489 (8.20%)], vascular diseases [63 (28.51%) vs. 4322 (23.79%)], and history of bleeding [72 (32.58%) vs. 3771 (20.76%)] (Table S3).

Algorithm Performance and Comparison

Ischemic Stroke Prediction

The performances of ML models in the original sample are described in Table 1. The AUCs of elastic net, RF, XGBoost, SVM, and neural network were 0.684 (95% CI 0.641–0.727), 0.916 (95% CI 0.887–0.945), 0.737 (95% CI 0.698–0.777), 0.545 (95% CI 0.502–0.588), and 0.625 (95% CI 0.579–0.672), respectively. RF outperformed other ML models in AUC (0.916, 95% CI 0.887–0.945), sensitivity (0.868), specificity (0.801), and F2 score (0.375). The best calibration was achieved in RF algorithm (Brier score = 0.035) (Table 1). Although CHA$_2$DS$_2$-VASc score showed a higher sensitivity (0.829) compared to other ML

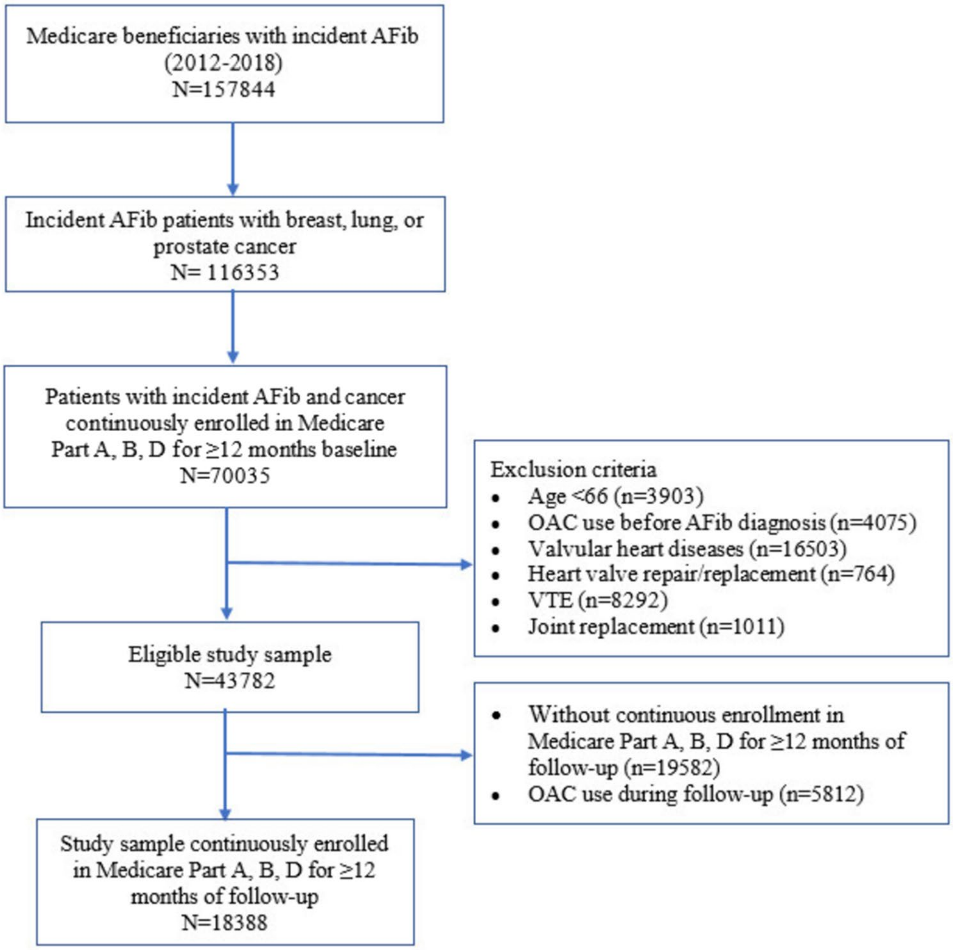
Medicare beneficiaries with incident AFib (2012-2018) N=157844
Incident AFib patients with breast, lung, or prostate cancer N= 116353
Patients with incident AFib and cancer continuously enrolled in Medicare Part A, B, D for ≥12 months baseline N=70035
Exclusion criteria
• Age <66 (n=3903)
• OAC use before AFib diagnosis (n=4075)
• Valvular heart diseases (n=16503)
• Heart valve repair/replacement (n=764)
• VTE (n=8292)
• Joint replacement (n=1011)
Eligible study sample N=43782
• Without continuous enrollment in Medicare Part A, B, D for ≥12 months of follow-up (n=19582)
• OAC use during follow-up (n=5812)
Study sample continuously enrolled in Medicare Part A, B, D for ≥12 months of follow-up N=18388

Fig. 1 Flowchart diagram for study sample. *VTE* Venous thromboembolism, *AFib* Atrial Fibrillation, *OAC* Oral Anticoagulant

Table 1 Model performance of machine learning models for ischemic stroke prediction

	Sensitivity	Specificity	AUROC	*p*-value*	F2 score	Brier score
Original data						
Elastic net	0.698	0.574	0.684 (0.641–0.727)	Reference	0.183	0.055
RF	0.868	0.801	0.916 (0.887–0.945)	<0.001	0.375	0.035
XGBoost	0.723	0.608	0.737 (0.698–0.777)	0.005	0.202	0.054
SVM	0.434	0.589	0.545 (0.502–0.588)	<0.001	0.121	0.055
NN	0.692	0.511	0.625 (0.579–0.672)	0.023	0.161	0.056
CHA_2DS_2-VASc	0.829	0.268	0.580 (0.534–0.623)	–	–	–
SMOTE resampling						
Elastic net	0.577	0.620	0.648 (0.603–0.693)	Reference	0.164	0.446
RF	0.801	0.334	0.633 (0.587–0.675)	0.0352	0.213	0.442
XGBoost	0.667	0.529	0.633 (0.588–0.678)	0.0408	0.160	0.440
SVM	0.560	0.633	0.650 (0.604–0.695)	0.1027	0.172	0.446
NN	0.372	0.752	0.580 (0.534–0.626)	<0.001	0.152	0.309

AUROC area under receiver operating characteristic curve, RF random forest, XGBoost extreme gradient boosting, SVM support vector machine, NN neural network, SMOTE synthetic minority oversampling technique

*DeLong's test

–: not calculated

models (except for RF), its specificity was low (0.268). Top five important features of RF algorithm were socioeconomic factors (proportion of household with no high school education level and median household median income), time from cancer diagnosis to AFib onset, history of stroke, and concomitant use of ACE inhibitors or ARBs. History of stroke and time from cancer diagnosis to AFib onset were the most important features in all ML models (Figs. 2, S3–S6).

Major Bleeding Prediction

For bleeding prediction, performances of ML models were poor (all AUCs < 0.7 in original sample) (Table 2).

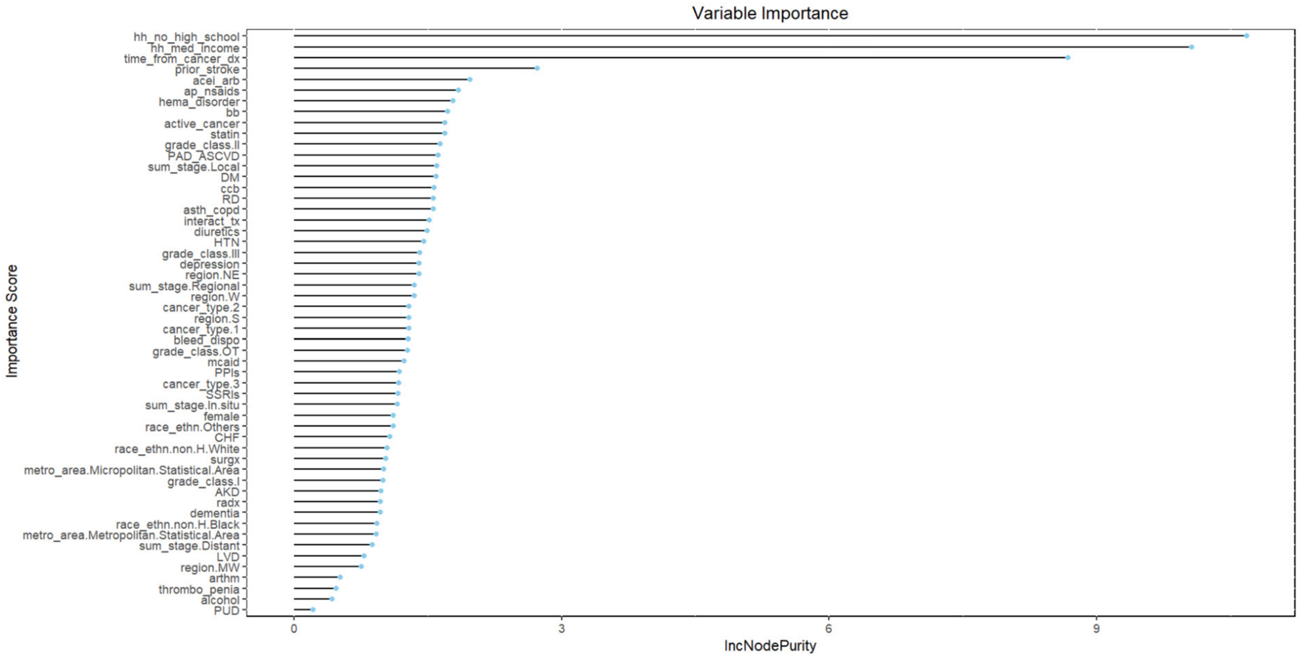

Fig. 2 Feature importance plot of random forest algorithm for ischemic stroke prediction (original data)

Table 2 Model performance of machine learning models for major bleeding prediction

	Sensitivity	Specificity	AUC	p-value	F2	Brier score
Original data						
Elastic net	0.424	0.689	0.575 (0.503–0.649)	Reference	0.070	0.023
RF	0.515	0.671	0.623 (0.554–0.692)	0.0003	0.081	0.024
XGBoost	0.439	0.641	0.578 (0.510–0.646)	0.7210	0.064	0.024
SVM	0.652	0.357	0.546 (0.472–0.619)	0.0726	0.056	0.024
NN	0.470	0.497	0.504 (0.432–0.575)	0.0122	0.051	0.024
HAS-BLED	0.052	0.960	0.574 (0.506–0.637)	–	–	–
SMOTE resampling						
Elastic net	0.348	0.722	0.564 (0.492–0.635)	Reference	0.064	0.378
RF	0.863	0.182	0.551 (0.478–0.625)	0.2752	0.057	0.048
XGBoost	0.515	0.517	0.553 (0.477–0.630)	0.2813	0.057	0.050
SVM	0.348	0.714	0.562 (0.490–0.634)	0.4052	0.062	0.375
NN	0.136	0.849	0.520 (0.449–0.590)	0.2752	0.041	0.200

AUROC area under receiver operating characteristic curve, RF random forest, XGBoost extreme gradient boosting, SVM support vector machine, NN neural network, SMOTE synthetic minority oversampling technique

*DeLong's test

–: not calculated

The AUCs of elastic net, RF, XGBoost, SVM, and neural network were 0.575 (95% CI 0.503–0.649), 0.623 (95% CI 0.554–0.692), 0.578 (95% CI 0.510–0.646), 0.546 (95% CI 0.472–0.619), 0.504 (95% CI 0.432–0.575), respectively. RF outperformed other models in AUC (0.623 (95% CI 0.554–0.692). However, sensitivity was highest for SVM algorithm (0.652), and best specificity was achieved in elastic net algorithm (0.689). There was no difference in calibration of five algorithms. HAS-BLED score failed to identify patients with major bleeding (sensitivity = 0.052). Proportion of household with no high school education level, median household median income, time from cancer

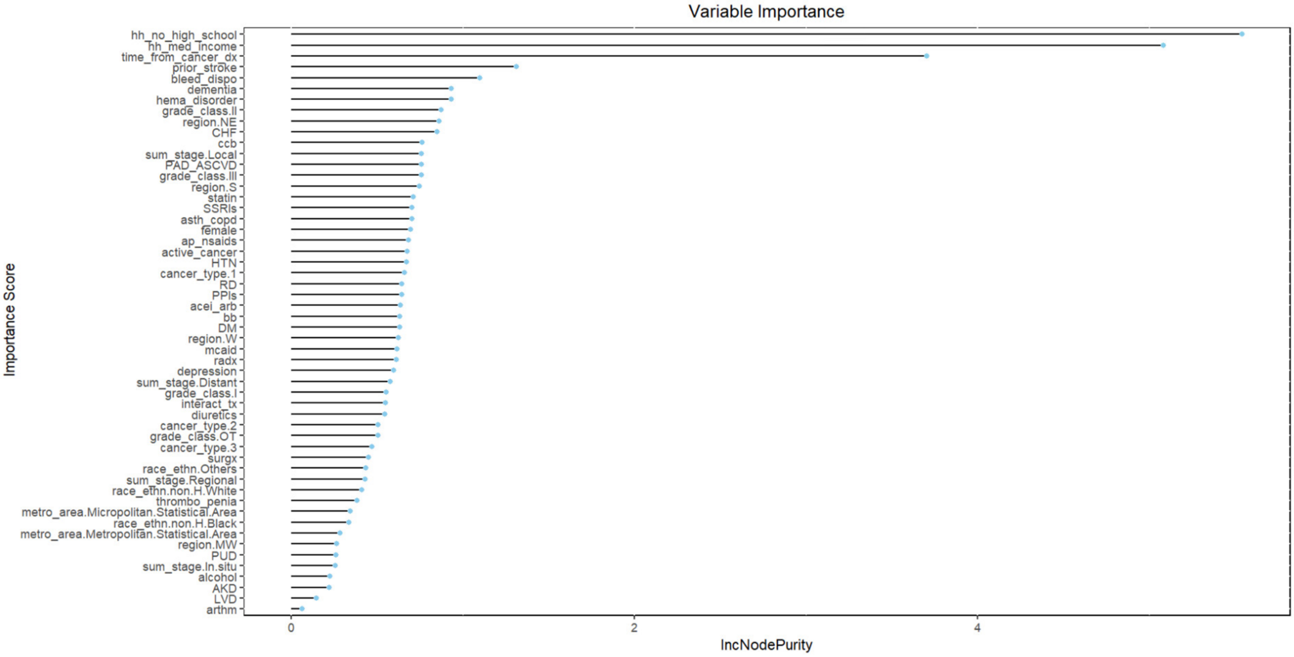

Fig. 3 Feature importance plot of random forest algorithm for major bleeding prediction (original data)

diagnosis to AFib onset, history of stroke, history of bleeding were top five important features of RF algorithm. History of bleeding was among top 5 important features in elastic net, RF, XGBoost, and neural network algorithms (Figs. 3**, S7-S10**).

Sensitivity Analysis

There was no major improvement in the performance metrics using SMOTE resampling for ischemic stroke and major bleeding prediction across five ML algorithms. SMOTE resampling worsened the calibration of the algorithms with larger Brier score compared with original sample (Table 1 and 2). Feature importance of ML algorithms in SMOTE samples are described in Figures S11-S15 (ischemic stroke) and Figures S16-20 (major bleeding).

Discussion

Our study is among the first studies that developed and validated ML algorithms to predict adverse outcomes exclusively for patients with AFib and cancer. In this cohort study, we demonstrated that incorporating ML algorithms into SEER-Medicare data can be a promising tool to predict short-term (1 year) risk of stroke among patients with AFib and cancer. Among older adults with cancer who were newly diagnosed with AFib, clinicians can collect patients' demographics, socioeconomic status, medical history, and medication history from routine medical records and/or patient survey, then leverage this tool to predict patients' risk of stroke. Our ML algorithms help clinicians identify high-risk patients and facilitate treatment decision (i.e., medication or non-pharmacological intervention) among older adults with AFib and cancer across the US.

RF outperformed other ML models in all metrics (AUC, sensitivity, specificity, and F2 score) for ischemic stroke. Although widely accepted as a risk assessment tool for stroke among patients with AFib, CHA_2DS_2-VASc score failed to achieve high performance in patients with AFib and cancer, especially in new onset AFib [9, 14, 49]. In this study, CHA_2DS_2-VASc score performed better than ML models, except for RF in identifying patients with ischemic stroke, however, CHA_2DS_2-VASc score could not differentiate those with lower risk (low specificity). In fact, 91.9% of patients in this study have $CHA_2DS_2\text{-VASc} \geq 2$ and would have been recommended for OACs according to current guidelines [11, 12]. The major limitation of CHA_2DS_2-VASc is the absence of cancer indicator, which has been suggested as an independent risk factor of stroke [50, 51]. A recently published study suggested the incorporation of cancer to CHA_2DS_2-VASc score to improve predictability of the original score [52]. Indeed, CHA_2DS_2-VASc score is the linear combination of conditions in prediction of stroke [10]. In the presence of cancer, the relationship between patient characteristics and ischemic stroke may become more complicated (i.e., non-linear), it is not surprising that CHA_2DS_2-VASc score failed to achieve high performance. In our study, linear models such as elastic net and SVM had lower performance metrics compared with non-linear models such as RF and XGBoost. Similar to CHA_2DS_2-VASc, we found prior stroke was among most important features in all ML algorithms. However, our approach incorporated a comprehensive set of patients' characteristics. For example, patients' socioeconomic status (household median income and education level) and cancer characteristics (cancer type, active cancer status) were ranked among top features in RF and XGBoost. The importance of these features highlighted contributions of health disparities and cancer characteristics in stroke prediction. The inclusion of these variables may be useful in identifying high-risk patients [53]. However, it is also noticed that tree-based models may inflate the impact of continuous features in their prediction [45]. Clinicians may consider initiating OACs for those who are at high risk of stroke identified by our RF algorithm.

Traditional tools such as HAS-BLED or $HEMORR_2HAGES$ showed poor predictability in patients with cancer [16, 17, 54]. Our ML algorithms also failed to obtain high performance metrics in prediction of major bleeding. Such poor performance suggested complex interactions between patients' characteristics and outcomes in the presence of cancer. First, although we obtained additional cancer characteristics compared with traditional risk scores, the performance was not improved [55]. This may suggest that our models failed to capture important features in prediction of major bleeding. In fact, genetic factors and disease severity were not available in SEER-Medicare data and dynamic features (i.e., cancer progression, new diagnosis of diseases) were not included in the models due to complexities. Similar to previous risk scores, we found that bleeding history was an important factor in prediction of subsequent major bleeding [17, 55]. Second, we excluded patients who have already initiated OACs before AFib diagnosis and those who initiated AFib during follow-up because OACs may increase risk of bleeding. As a result, only 1.2% patients in our cohort experienced bleeding events during follow-up and this created a severe imbalance classification problem for our ML algorithms and may lead to poor predictability [56]. Future studies may expand the outcomes to other types of bleeding (i.e., intracranial bleeding, gastrointestinal bleeding, or other non-critical site bleeding) to improve the performance and the clinical utility of the algorithms.

In our study, SMOTE resampling approach did not improve the performance of the model. In the training set, SMOTE created new synthetic 'stroke' individuals from interpolations of the original, real 'stroke' cases [48]. Studies have shown that SMOTE-like methods could improve the performance of weak classifiers such as SVM, decision tree [57]. In our study, SMOTE improved AUCs in SVM only. Another limitation of SMOTE is that it resulted in poorly calibrated models where the probability of the minority class (stroke) was strongly inflated demonstrated by Brier score.

Our study is subject to some limitations. We were unable to capture some important variables in the ML models (i.e., BMI, genetic factors, frailty, and health behaviors—not available in SEER-Medicare). Socioeconomic factors such as household income and education level are available on the aggregate area level (Census tract) but not individual level. In addition, our algorithms did not incorporate the impact of some post-baseline predictors (i.e., treatment dosage, adherence, recent CHA_2DS_2-VASc and HAS-BLED scores, recent use of NSAIDs, and other time-varying variables such as interactions between oral anticoagulants between OACs and antineoplastic agents) [58]. Our study is applicable to the study period 2011–2019. From 2020, the presence Covid-19 has worsened outcomes of patients with AFib or cancer patients and has negatively impacted health services, delayed and reduced cancer screening and diagnosis in the United States [59–63]. Therefore, the model should be updated and validated incorporating Covid-related factors during and after the pandemic. In addition, our ML algorithms could not further stratify the risk of stroke and major bleeding (i.e., low, moderate, high, or very high). Future study may leverage advanced ML algorithms such as survival ML in predicting the probability of adverse events after 1 year or extended follow-up time. Last, the generalizability of our ML models to other populations may be limited (i.e., commercial insurance, anticoagulated patients, or patients with other cancer types).

Conclusion

Our study demonstrated a promising application of ML in stroke prediction among older adults with cancer who are newly diagnosed with AFib in the US. This tool may be leveraged in assisting clinicians in identification of patients at high risk of stroke and improving treatment decisions.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12012-024-09843-8.

Acknowledgements This study used the linked SEER-Medicare database. The interpretation and reporting of these data are the sole responsibility of the authors. The authors acknowledge the efforts of the National Cancer Institute; Information Management Services (IMS), Inc.; and the Surveillance, Epidemiology, and End Results (SEER) Program tumor registries in the creation of the SEER-Medicare database. The collection of cancer incidence data used in this study was supported by the California Department of Public Health pursuant to California Health and Safety Code Section 103885; Centers for Disease Control and Prevention's (CDC) National Program of Cancer Registries, under cooperative agreement 1NU58DP007156; the National Cancer Institute's Surveillance, Epidemiology and End Results Program under contract HHSN261201800032I awarded to the University of California, San Francisco, contract HHSN261201800015I awarded to the University of Southern California, and contract HHSN261201800009I awarded to the Public Health Institute. The ideas and opinions expressed herein are those of the author(s) and do not necessarily reflect the opinions of the State of California, Department of Public Health, the National Cancer Institute, and the Centers for Disease Control and Prevention or their Contractors and Subcontractors.

Author Contribution BT: Conceptualization; Data curation; Formal analysis; Methodology; Writing—original draft. BF: Methodology; Validation; Writing—review and editing. EC: Methodology; Validation; Writing—review and editing. JZ: Methodology; Validation; Writing—review and editing. LH: Methodology; Validation; Writing—review and editing. JQ: Conceptualization; Methodology; Supervision; Validation; Writing—review and editing.

Funding This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Declarations

Conflict of interest No conflicts of interest to report for all authors.

Ethical Approval The study was granted exemption by the Auburn University institutional review board (IRB) and the study was performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards.

References

1. Patel, N. J., Deshmukh, A., Pant, S., et al. (2014). Contemporary trends of hospitalization for atrial fibrillation in the United States, 2000 through 2010: Implications for healthcare planning. *Circulation, 129*(23), 2371–2379.
2. Benjamin, E. J., Muntner, P., Alonso, A., et al. (2019). Heart disease and stroke statistics-2019 update: A report from the American Heart Association. *Circulation, 139*(10), e56–e528.

3. Centers for Disease Control and Prevention - National Center for Health Statistics. About Multiple Cause of Death, 1999–2019. https://wonder.cdc.gov/mcd-icd10.html. Published 2019. Retrieved October 14, 2021.
4. Chung, M. K., Eckhardt, L. L., Chen, L. Y., et al. (2020). Lifestyle and risk factor modification for reduction of atrial fibrillation: A Scientific statement from the American Heart Association. *Circulation, 141*(16), e750–e772.
5. Timp, J. F., Braekkan, S. K., Versteeg, H. H., & Cannegieter, S. C. (2013). Epidemiology of cancer-associated venous thrombosis. *Blood, 122*(10), 1712–1723.
6. Prandoni, P., Lensing, A. W. A., Piccioli, A., et al. (2002). Recurrent venous thromboembolism and bleeding complications during anticoagulant treatment in patients with cancer and venous thrombosis. *Blood, 100*(10), 3484–3488.
7. Melloni, C., Shrader, P., Carver, J., et al. (2017). Management and outcomes of patients with atrial fibrillation and a history of cancer: The ORBIT-AF registry. *European Heart Journal - Quality of Care and Clinical Outcomes, 3*(3), 192–197.
8. Fanola, C. L., Ruff, C. T., Murphy, S. A., et al. (2018). Efficacy and safety of Edoxaban in patients with active malignancy and atrial fibrillation: Analysis of the ENGAGE AF-TIMI 48 trial. *Journal of the American Heart Association., 7*(16), e008987.
9. Sorigue, M., & Miljkovic, M. D. (2019). Atrial fibrillation and stroke risk in patients with cancer: A primer for oncologists. *Journal of Oncology Practice., 15*(12), 641–650.
10. Lip, G. Y., Nieuwlaat, R., Pisters, R., Lane, D. A., & Crijns, H. J. (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.
11. January, C. T., Wann, L. S., Calkins, H., et al. (2019). 2019 AHA/ACC/HRS focused update of the 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation: A report of the American College of Cardiology/American Heart Association task force on clinical practice guidelines and the heart rhythm society in collaboration with the society of thoracic surgeons. *Circulation, 140*(2), e125–e151.
12. Hindricks, G., Potpara, T., Dagres, N., et al. (2020). 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): The Task Force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) Developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. *European Heart Journal., 42*(5), 373–498.
13. D'Souza, M., Carlson, N., Fosbøl, E., et al. (2018). CHA(2)DS(2)-VASc score and risk of thromboembolism and bleeding in patients with atrial fibrillation and recent cancer. *European Journal of Preventive Cardiology, 25*(6), 651–658.
14. Patell, R., Gutierrez, A., Rybicki, L., & Khorana, A. A. (2017). Usefulness of CHADS2 and CHA2DS2-VASc scores for stroke prediction in patients with cancer and atrial fibrillation. *American Journal of Cardiology, 120*(12), 2182–2186.
15. Pisters, R., Lane, D. A., Nieuwlaat, R., de Vos, C. B., Crijns, H. J., & Lip, G. Y. (2010). 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, 138*(5), 1093–1100.
16. Brown, J. D., Goodin, A. J., Lip, G. Y. H., & Adams, V. R. (2018). Risk stratification for bleeding complications in patients with venous thromboembolism: Application of the HAS-BLED bleeding score during the first 6 months of anticoagulant treatment. *Journal of the American Heart Association*. https://doi.org/10.1161/JAHA.117.007901
17. Pastori, D., Marang, A., Bisson, A., Herbert, J., Lip, G. Y. H., & Fauchier, L. (2021). Comparison of the HAS-BLED, ORBIT and ATRIA bleeding risk scores in 399,344 patients with atrial fibrillation and cancer. *European Heart Journal*. https://doi.org/10.1093/eurheartj/ehab724.0438
18. Chirikov, V. V., Shaya, F. T., Onukwugha, E., Mullins, C. D., dosReis, S., & Howell, C. D. (2017). Tree-based claims algorithm for measuring pretreatment quality of care in medicare disabled hepatitis C patients. *Medical Care, 55*(12), e104.
19. Gordon, J., Norman, M., Hurst, M., et al. (2021). Using machine learning to predict anticoagulation control in atrial fibrillation: A UK clinical practice research datalink study. *Informatics in Medicine Unlocked., 25*, 100688.
20. Spooner, A., Chen, E., Sowmya, A., et al. (2020). A comparison of machine learning methods for survival analysis of high-dimensional clinical data for dementia prediction. *Scientific Reports., 10*(1), 20410.
21. Thottakkara, P., Ozrazgat-Baslanti, T., Hupf, B. B., et al. (2016). Application of machine learning techniques to high-dimensional clinical data to forecast postoperative complications. *PLoS ONE, 11*(5), e0155705.
22. Ryo, M., & Rillig, M. C. (2017). Statistically reinforced machine learning for nonlinear patterns and variable interactions. *Ecosphere., 8*(11), e01976.
23. Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. M. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *BMC Medicine., 13*(1), 1.
24. National Cancer Institute. Overview of the Surveillance, Epidemiology, and End Results (SEER) Program. Retrieved December 27, 2021. from https://seer.cancer.gov/about/overview.html
25. Warren, J. L., Klabunde, C. N., Schrag, D., Bach, P. B., & Riley, G. F. (2002). Overview of the SEER-Medicare data: content, research applications, and generalizability to the United States elderly population. *Medical Care*. https://doi.org/10.1097/00005650-200208001-00002
26. Jensen, P. N., Johnson, K., Floyd, J., Heckbert, S. R., Carnahan, R., & Dublin, S. (2012). A systematic review of validated methods for identifying atrial fibrillation using administrative data. *Pharmacoepidemiology and Drug Safety*. https://doi.org/10.1002/pds.2317
27. Lyman, G. H., Carrier, M., Ay, C., et al. (2021). American Society of Hematology 2021 guidelines for management of venous thromboembolism: Prevention and treatment in patients with cancer. *Blood Advances., 5*(4), 927–974.
28. Otto, C. M., Nishimura, R. A., Bonow, R. O., et al. (2021). 2020 ACC/AHA guideline for the management of patients with valvular heart disease: A report of the American College of Cardiology/American Heart Association Joint Committee on clinical practice guidelines. *Circulation, 143*(5), e72–e227.
29. Deitelzweig, S., Keshishian, A. V., Zhang, Y., et al. (2021). Effectiveness and Safety of oral anticoagulants among nonvalvular atrial fibrillation patients with active cancer. *JACC: CardioOncology., 3*(3), 411–424.
30. Thigpen, J. L., Dillon, C., Forster, K. B., et al. (2015). Validity of international classification of disease codes to identify ischemic stroke and intracranial hemorrhage among individuals with associated diagnosis of atrial fibrillation. *Circulation Cardiovascular Quality and Outcomes, 8*(1), 8–14.
31. Cunningham, A., Stein, C. M., Chung, C. P., Daugherty, J. R., Smalley, W. E., & Ray, W. A. (2011). An automated database case definition for serious bleeding related to oral anticoagulant use. *Pharmacoepidemiology and drug safety., 20*(6), 560–566.

32. Shah, S., Norby, F. L., Datta, Y. H., et al. (2018). Comparative effectiveness of direct oral anticoagulants and warfarin in patients with cancer and atrial fibrillation. *Blood Advances, 2*(3), 200–209.
33. Connolly, S. J., Ezekowitz, M. D., Yusuf, S., et al. (2009). Dabigatran versus Warfarin in patients with atrial fibrillation. *New England Journal of Medicine., 361*(12), 1139–1151.
34. Waljee, A. K., Mukherjee, A., Singal, A. G., et al. (2013). Comparison of imputation methods for missing laboratory data in medicine. *British Medical Journal Open, 3*(8), e002847.
35. Stekhoven, D. J., & Bühlmann, P. (2012). MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics, 28*(1), 112–118.
36. Xu, Y., & Goodacre, R. (2018). On splitting training and validation set: A comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning. *Journal of Analysis and Testing., 2*(3), 249–262.
37. Liu, H., & Cocea, M. (2017). Semi-random partitioning of data into training and test sets in granular computing context. *Granular Computing., 2*(4), 357–386.
38. Cawley, G. C., & Talbot, N. L. C. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. *Journal of Machine Learning Research, 11*, 2079–2107.
39. Claxton, J. S., MacLehose, R. F., Lutsey, P. L., et al. (2019). A new model to predict ischemic stroke in patients with atrial fibrillation using warfarin or direct oral anticoagulants. *Heart Rhythm, 16*(6), 820–826.
40. Brownlee, J. (2020). *Imbalanced Classification with Python: Better Metrics, Balance Skewed Classes, Cost-Sensitive Learning*. Machine Learning Mastery.
41. van den Goorbergh, R., van Smeden, M., Timmerman, D., & Van Calster, B. (2022). The harm of class imbalance corrections for risk prediction models: Illustration and simulation using logistic regression. *Journal of the American Medical Informatics Association, 29*(9), 1525–1534.
42. DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics, 44*(3), 837–845.
43. Devarriya, D., Gulati, C., Mansharamani, V., Sakalle, A., & Bhardwaj, A. (2020). Unbalanced breast cancer data classification using novel fitness functions in genetic programming. *Expert Systems with Applications., 140*, 112866.
44. Saarela, M., & Jauhiainen, S. (2021). Comparison of feature importance measures as explanations for classification models. *SN Applied Sciences., 3*(2), 272.
45. Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics, 8*(1), 25.
46. Loecher, M. (2022). Unbiased variable importance for random forests. *Communications in Statistics - Theory and Methods., 51*(5), 1413–1425.
47. Huang, Y., Li, W., Macheret, F., Gabriel, R. A., & Ohno-Machado, L. (2020). A tutorial on calibration measurements and calibration models for clinical prediction models. *Journal of the American Medical Informatics Association., 27*(4), 621–633.
48. Chawla, N., Bowyer, K., Hall, L., & Kegelmeyer, W. (2002). SMOTE: Synthetic minority over-sampling technique. *J Artif Intell Res (JAIR)., 16*, 321–357.
49. D'Souza, M., Carlson, N., Fosbøl, E., et al. (2018). CHA2DS2-VASc score and risk of thromboembolism and bleeding in patients with atrial fibrillation and recent cancer. *European Journal of Preventive Cardiology., 25*(6), 651–658.
50. Navi, B. B., Reiner, A. S., Kamel, H., et al. (2015). Association between incident cancer and subsequent stroke. *Annals of Neurology, 77*(2), 291–300.
51. Bang, O. Y., Chung, J. W., Lee, M. J., Seo, W. K., Kim, G. M., & Ahn, M. J. (2020). Cancer-related stroke: An emerging subtype of ischemic stroke with unique pathomechanisms. *J Stroke., 22*(1), 1–10.
52. Bungo, B., Chaudhury, P., Arustamyan, M., et al. (2022). Better prediction of stroke in atrial fibrillation with incorporation of cancer in CHA2DS2VASC score: CCHA2DS2VASC score. *IJC Heart & Vasculature., 41*, 101072.
53. Lindmark, A., Eriksson, M., & Darehed, D. (2022). Socioeconomic status and stroke severity: Understanding indirect effects via risk factors and stroke prevention using innovative statistical methods for mediation analysis. *PLoS ONE, 17*(6), e0270533.
54. Raposeiras Roubín, S., Abu Assi, E., Muñoz Pousa, I., et al. (2022). Incidence and predictors of bleeding in patients with cancer and atrial fibrillation. *American Journal of Cardiology, 167*, 139–146.
55. Trinks-Roerdink, E. M., Geersing, G. J., Hemels, M. E. W., et al. (2023). External validation and updating of prediction models of bleeding risk in patients with cancer receiving anticoagulants. *Open Heart., 10*(1), e002273.
56. Wang, S., Dai, Y., Shen, J., & Xuan, J. (2021). Research on expansion and classification of imbalanced data based on SMOTE algorithm. *Scientific Reports., 11*(1), 24039.
57. Elor Y, Averbuch-Elor H. To SMOTE, or not to SMOTE? *arXiv preprint arXiv:220108528*. 2022.
58. Truong, B., Hornsby, L., Fox, B. I., Chou, C., Zheng, J., & Qian, J. (2023). Screening for clinically relevant drug-drug interactions between direct oral anticoagulants and antineoplastic agents: A pharmacovigilance approach. *Journal of Thrombosis and Thrombolysis., 56*(4), 555–567.
59. Lee, L. Y., Cazier, J. B., Angelis, V., et al. (2020). COVID-19 mortality in patients with cancer on chemotherapy or other anticancer treatments: A prospective cohort study. *Lancet, 395*(10241), 1919–1926.
60. Pardo Sanz, A., Salido Tahoces, L., Ortega Pérez, R., González Ferrer, E., Sánchez Recalde, Á., & Zamorano Gómez, J. L. (2021). New-onset atrial fibrillation during COVID-19 infection predicts poor prognosis. *Cardiology Journal, 28*(1), 34–40.
61. Rosenblatt, A. G., Ayers, C. R., Rao, A., et al. (2022). New-onset atrial fibrillation in patients hospitalized with COVID-19: Results from the american heart association COVID-19 cardiovascular registry. *Circulation: Arrhythmia and Electrophysiology., 15*(5), e010666.
62. Mariotto, A. B., Feuer, E. J., Howlader, N., Chen, H.-S., Negoita, S., & Cronin, K. A. (2023). Interpreting cancer incidence trends: challenges due to the COVID-19 pandemic. *JNCI: Journal of the National Cancer Institute., 115*(9), 1109–1111.
63. National Cancer Institute. Impact of COVID on 2020 SEER Cancer Incidence Data. Retrieved February 12, 2024 from https://seer.cancer.gov/data/covid-impact.html

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.