Improving Anticoagulant Treatment Strategies of Atrial Fibrillation Using Reinforcement Learning

Lei Zuo, MS[1], Xin Du, MD[2,3], Wei Zhao, PhD[1], Chao Jiang, MD[2], Shijun Xia, MD[2], Liu He, PhD[2], Rong Liu, MPH[3], Ribo Tang, MD[2], Rong Bai, MD[2], Jianzeng Dong, MD[2,4], Xingzhi Sun, PhD[1], Gang Hu, PhD[1], Guotong Xie, PhD[1], Changsheng Ma, MD[2]
[1]Ping An Health Technology, Beijing, China; [2]Beijing Anzhen Hospital, Capital Medical University; National Clinical Research Center for Cardiovascular Diseases; Beijing Advanced Innovation Center for Big Data-Based Precision Medicine for Cardiovascular Diseases, Beijing, China; [3]Heart Health Research Center, Beijing, China; [4]The First Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan Province, China

Abstract

In this paper, we developed a personalized anticoagulant treatment recommendation model for atrial fibrillation (AF) patients based on reinforcement learning (RL) and evaluated the effectiveness of the model in terms of short-term and long-term outcomes. The data used in our work were baseline and follow-up data of 8,540 AF patients with high risk of stroke, enrolled in the Chinese Atrial Fibrillation Registry (CAFR) study during 2011 to 2018. We found that in 64.98% of patient visits, the anticoagulant treatment recommended by the RL model were concordant with the actual prescriptions of the clinicians. Model-concordant treatments were associated with less ischemic stroke and systemic embolism (SSE) event compared with non-concordant ones, but no significant difference on the occurrence rate of major bleeding. We also found that higher proportion of model-concordant treatments were associated with lower risk of death. Our approach identified several high-confidence rules, which were interpreted by clinical experts.

Introduction

Atrial fibrillation (AF) is a common cardiac arrhythmia in adults, affecting up to approximately 10 million in China[1]. AF is a risk factor for stroke/thromboembolism and death[2,3], with an estimated 5-fold higher risk[4]. Warfarin is effective in preventing ischemic stroke (IS) in AF patients. However, major bleeding is not uncommon in patients treated with warfarin[5]. The efficacy and safety of non-vitamin K antagonist oral anticoagulants (NOACs) have been reported in clinical trials[6], but NOACs are expensive. Therefore, treating patients precisely by identifying the right patients to be treated and choosing the right oral anticoagulation (OAC) agent is important to achieve the largest net benefit.

The CHA_2DS_2-VA (CV) score (congestive heart failure, hypertension, age ≥75 years, diabetes mellitus, prior stroke or thromboembolism, vascular disease and age 65-74 years) ranging from 0 to 8 has been widely recommended to identify the risk level of IS for AF patients[7]. The larger the CV score is, the higher risk of IS the patient has. For AF patients with CV score of 2 or greater, OAC is recommended by several clinical guidelines[8,9]. However, only 6%-8% of AF patients died from stroke[10] but warfarin results in an annual major bleeding risk of 2%-5%[11]. It is still skeptical that the CV score may not precisely capture the risk of particular AF patients, and it might be unnecessary to treat all those patients. Thus, it is urgent to achieve personalized medication recommendation. Liu et al. proposed an outcome-driven approach to identify a precise group of patients with low risk of IS and described their unique characteristics[12].

In complementary to the traditional machine learning (ML) models, reinforcement learning (RL) has the distinctive advantage of optimizing sequences of decisions by learning the best policy. With the explosive increase of electronic medical records (EMR), RL approach has been successfully applied in healthcare domain, specifically in treatment recommendation[13-16]. For example, Komorowski et al.[14] aimed to reduce septic patients' mortality by recommending personalized optimal dosage of intravenous fluids and vasopressors. They found that patients who received doses similar to the model recommendation had the lowest 90-day mortality. Phuong D. Ngo et al.[15] proposed a RL model for optimal insulin injection policy in patients with type-1 diabetes. The result showed that the proposed methodology significantly reduced and successfully regulated the fluctuation of the blood glucose. Besides, there is an application to optimize the radiation therapy for cancer patients using Q-learning algorithm[16]. Inspired by these successful applications in treatment recommendation, we developed a personalized RL-based model to recommend anticoagulant therapy for AF patients and evaluated the effectiveness of the model in terms of short-term and long-term outcomes.

The RL treatment model learned the policy from real-world data collected in the Registry study. The Chinese Atrial Fibrillation Registry (CAFR) study was initiated at 2011, and has enrolled more than 25,000 AF patients from 32 hospitals in Beijing. Data of patients' demographics, symptoms and signs, results of physical examination and laboratory test, medical history, details of precious and current treatments at baseline, and follow-up every 6 months were collected. At every follow-up visit, symptoms and signs, the clinical events such as stroke and major bleeding, treatment condition and results of physical examination and laboratory test if any were collected.

In this study, we proposed a method to recommend the personalized anticoagulant treatment for AF patients. First, we defined ischemic stroke and systemic embolism (SSE) and major bleeding events as short-term outcomes and death as long-term outcome for AF patients. Then we developed the RL model by designing the reward function and loss function based on AF expertise. The effectiveness of the model was demonstrated by lower incidence rate of SSE in patient visits with model-concordant prescriptions than those non-concordant. There was no significant difference on the occurrence rate of major bleeding. We also found that higher proportion of model-concordant treatments were associated with lower risk of death. Furthermore, to better interpret the RL model and assist clinicians in making treatment decisions, we built a decision tree to simulate the RL model recommendations.

Method

Figure 1 shows the pipeline of building anticoagulant treatment model for AF patients. We first selected samples from CAFR data that meet the inclusion criteria. At the model training stage, we formulated the reinforcement learning (RL) problem for AF treatment by carefully designing state, action, reward function and loss function based on the expertise of AF. Then we trained a RL model, which is a deep neural network[17], from the selected data. The state was represented by the values of demographics, lab tests, vital signs, medical history and previous drugs of AF patients. The action was defined to simulate the actual prescription. The reward function was defined to assess the action in a given state according to CHA_2DS_2-VA score, treatment and clinical outcomes. The novelty also lied in the design of the loss function for model training, in which the loss consisted of the temporal difference (TD) loss, a regularization term and a supervised large margin classification loss. Specifically, the supervised loss enabled the algorithm to learn to imitate the clinician. After that, model-concordant was defined as the consistency between clinician's actual prescription and model recommendation, and we evaluated the performance of the RL model in terms of short-term outcomes, i.e., SSE and major bleeding events and long-term outcome, i.e., death. To evaluate the short-term outcomes at patient-visit level, the occurrence rates of SSE and major bleeding events were compared between the model-concordant treatments and the model-non-concordant treatments. To evaluate the long-term outcome at patient level, the relationship between patient's model-concordant rate and the occurrence rate of the death was depicted. Finally, we interpreted the RL model by building a decision tree to classify visits labeled model actions. The decision tree can simulate the RL model approximatively and derive rules according to the tree structure. The rules with enough coverage and high confidence could be used to interpret the RL model and assist clinicians to make clinical decision.

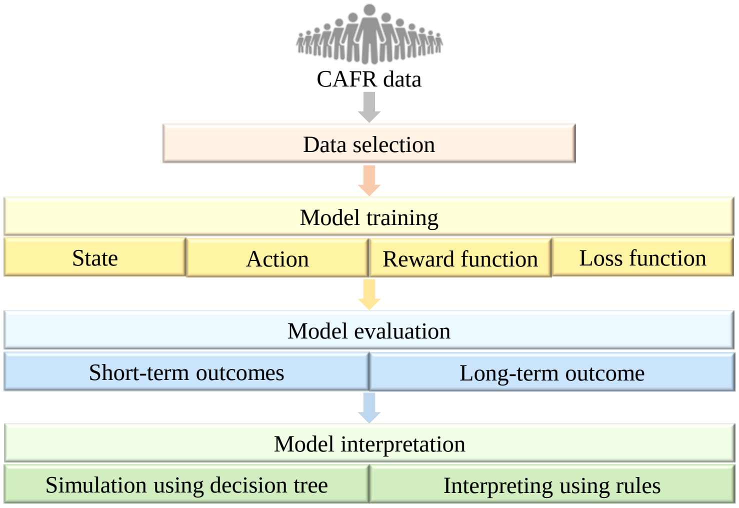

Figure 1. Pipeline of building anticoagulant treatment model for AF patients

Data selection and clinical outcomes

The purpose of this study is to assist clinicians to decide whether the AF patient truly need to been treated with oral anticoagulation. Both short-term and long-term clinical outcomes were taken into account. Short-term outcomes were evaluated at patient-visit level, including the occurrence rate of ischemic stroke and systemic embolism (SSE) event and major bleeding event after 6 months' therapy. Long-term outcome was evaluated at patient level, that is the occurrence rate of death in up to 8 years.

In this study, the patient-visits of interest are those which had not been treated with radiofrequency ablation (RFA) and CHA$_2$DS$_2$-VA score is 2 or greater. This is because RFA is associated with the reduction of the risk of stroke and mortality in AF patients[18]. To build and evaluate the RL treatment model, we deleted the visits without short-term outcomes. Finally, from the CAFR data, we identified 8,540 AF patients and 52,172 patient-visits that meet the selection criteria shown in Figure 2.

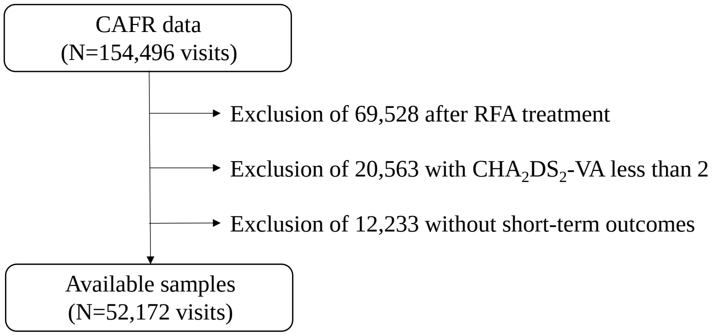
CAFR data
(N=154,496 visits)
Exclusion of 69,528 after RFA treatment
Exclusion of 20,563 with CHA$_2$DS$_2$-VA less than 2
Exclusion of 12,233 without short-term outcomes
Available samples
(N=52,172 visits)

Figure 2. Patient-visit selection criteria for anticoagulant treatment recommendation model

Reinforcement Learning model

Applying RL to the oral anticoagulant treatment problem of AF patients involves several elements. The first is to define the input of the treatment model (i.e., state) and the treatment options (i.e., action). Next we describe how we quantify the effectiveness of applying an action for a given state (i.e., reward). Finally, we describe how to design our model architecture and loss function for training the model.

1) State

For each patient visit, the clinical conditions included demographics, lab values, vital signs, medical history and previous drugs. We applied Z-score standardization for continuous variables, performed one-hot transformation for categorical variables, and kept binary variables unchanged. After the standardization and transformation, all values were rescaled into [-1,1], and we had a state of a 31 × 1 feature vector for each patient-visit, denoted as s_t. Note that the last visit of a patient was not used for model training because it did not associated with the outcomes.

2) Action

We focused on two types of drugs: warfarin and non-vitamin K oral anticoagulants (NOACs). Thus, we defined three actions: 0 represented no drug, 1 represented warfarin and 2 represented NOAC.

3) Reward

The reward function was clinically oriented and defined based on the expertise of AF. We considered the CHA$_2$DS$_2$-VA (CV) score, treatment, and the outcomes that indicate a patient's health status. The CV score not only identified the risk of IS, but also reflected the risk of bleeding to some extent. According to expert advice and related work[19], the CV score was stratified as shown in Table 1. For the last follow-up of a patient, if death is reported, a negative reward (i.e. penalty, -20) is given; Otherwise, a positive reward (i.e., 20) is given. For other available visits, Table 1 showed the reward values in detail. The reward function is defined based on the following principles. First, when SSE occurs, anticoagulant therapy is considered insufficient. Therefore, action of anticoagulation is given a positive reward, while no anticoagulation is given a negative reward; When major bleeding occurs, anticoagulant therapy is considered beyond the patient's tolerance. So, anticoagulation is given a negative reward while no anticoagulation is given a positive reward; when no SSE or major bleeding occurs, a positive reward is given no matter what the treatment is. Second, the higher the CV score is, the more necessary anticoagulant therapy is considered. For example, in Table 1, when the treatment is "anticoagulation" and the outcome is "SSE occurs", the value of the reward increases with the CV score. Third, the occurrence of either SSE or major bleeding is the severe event for AF patients, so the absolute value of their reward is higher than that of no occurrence. After defining the relative relationship between reward values according to the principles above, we determined the values in Table 1 by experiments. Specifically, we set the minimum value as 1, adjusted other values, and selected the value combination that lead to good model convergence.

Table 1. Reward of non-last follow-up

Treatment	**Outcome**	**CHA$_2$DS$_2$-VA score**		
		=2	**3-4**	**≥5**
Anticoagulation	SSE occurs	4	5	6
	Major bleeding occurs	-6	-6	-6
	No occurs	3	4	5
No anticoagulation	SSE occurs	-7	-6	-6
	Major bleeding occurs	4	3	2
	No occurs	3	2	1

4) Model architecture

To learn treatment policies, we used a reinforcement learning model which is a variant of Deep Q Networks (DQN)[17], called Prioritized Dueling Double DQN (PDD DQN)[20-22]. DQN gets a state vector as input and a Q vector of action values as output. The Q vector is the evaluation of all 3 actions' effects on patients' status and is calculated by a deep neural network (DNN) for a given state and all actions. However, as discussed in related work[20], using the same values both to select and evaluate an action can lead to overestimate the Q-values. To mitigate this problem, Double DQN (DDQN) uses the main network to determine the max Q values and the corresponding action, and then uses the target network to estimate the target Q-values[20]. When evaluating the effectiveness of a treatment, we model the influence from state and action on Q-values separately. So we use a Dueling Q Network where the function Q(s, a) is split into two streams, *value* and *advantage*[21]. The *value* represents the quality of a patient-visit's underlying state and the *advantage* represents the quality of the action being taken at that time-step. In addition, we use Prioritized Experience Replay (PER) to accelerate learning and improve the final policy quality by sampling the input data according to the samples' weight, which measured by their temporal-difference errors[22]. Our final PDD DQN network architecture has two hidden layers of 128 units with the batch normalization for main network and target network. The learning rate is 0.001 and the batch size is 128.

5) Loss function

The loss function is given in Equation (1), which consists of three terms. The first term is the temporal difference (TD) loss between the output Q of the network and the desired target Q. The traditional DQN optimizes the network parameters to minimize this TD loss. In addition to TD loss, we added a regularization term and a supervised large margin classification loss. The regularization term penalizes the output Q-values that differ significantly from the predefine threshold (Q_{thre} = 20), in order to learn a more appropriate Q-function. The supervised loss forces the values of clinician's actual action higher than the value of the other actions by a predefined margin[23]. The novelty is that we apply the value of the clinician's action and adjust the value of other actions. The supervised loss enables the algorithm to learn the model that imitates the clinician, while the TD loss ensures that the network satisfies the Bellman equation[17] (Equation (2)) and the reinforcement learning framework can be leveraged.

$$L(\theta) = \left[\left(Q_{target} - Q(s,a;\theta)\right)^2\right] + [\lambda \cdot max(|Q(s,a;\theta)| - Q_{thre}, 0)] + \left[\max\left(Q(s,a;\theta) + l(a_c,a)\right) - Q(s,a_c;\theta)\right] \tag{1}$$

$$Q_{target} = r + \gamma Q_{target}(s', \arg max_{a'} Q(s',a';\theta); \theta') \tag{2}$$

where θ are the parameters of the main network and θ' are the parameters of the target network, a_c is the action the clinician took in state s and $l(a_c, a)$ is a margin function that is 0 when $a = a_c$ and positive value otherwise.

Statistical analyses for model evaluation

To evaluate the short-term outcomes, model-concordant was regarded as the exposure factor[24], which is determined by whether the clinician's actual prescription is concordant to the RL model recommended medication at patient-visit level. The occurrence rates of short-term outcomes were compared between the model-concordant treatments and the model-non-concordant treatments via chi-square test. To adjust the key confounder, we stratified the CHA$_2$DS$_2$-VA score into: 2, 3-4 and ≥5[19], since both medical history and age are considered into the score. Furthermore, we stratified the patient visits according to the model action to evaluate the model performance.

To evaluate the long-term outcome, we performed it at patient level. First, the patient's model-concordant rate was calculated as the ratio between the number of model-concordant visits and the total number of visits. Then the patients were divided into different groups by applying the bin with size 0.2 on the patient's model-concordant rate, and the occurrence rate of death was computed in each group. Finally, the relationship between patient's model-concordant rate and the occurrence rate of the long-term outcome can be illustrated. A test for linear trend was conducted using Cochran-Armitage test. For all tests, a P value < 0.05 was considered statistically significant.

Group characterization for model interpretation

After obtaining the recommended treatment, our interest is to identify the characteristics of each group stratified by the action recommended by the RL model and understand the difference between groups. We addressed this by using CART to build a decision tree, which classify visits into model actions. To compare with the CHA_2DS_2-VA score, the candidate features just included the risk factors of CV score and previous anticoagulant drugs. To avoid overfitting of the decision tree, we used a parameter representing the minimum number of samples at a leaf node (min_samples_leaf) to constrain the tree construction. To interpret the RL model, we further derived the explicit rules from the decision tree. Finally, the coverage and confidence of the rule were assessed quantitatively, where the coverage of a rule is the number of samples in the corresponding leaf node and the confidence of the rule is the accuracy at that leaf node. The rules with enough coverage and high confidence would assist clinicians to make informed clinical decision.

Result

Data Set

We used baseline and follow-up data enrolled in CAFR study during 2011 to 2018 to build our RL model. In total, there were 52,172 visits with the diagnosis as AF, corresponding to 8,540 unique patients. Baseline characteristics of these patients were listed in Table 2. The average age of the patients was 71.57±8.89 with 46.4% female. The median CHA_2DS_2-VA score was 3 (IQR, 2-4). Specifically, patients were followed up every 6 months consecutively, and data of symptoms and signs, physical examination and laboratory test results, treatments and the clinical events were collected. The mean follow-up time was 2.72±1.84 years and the median number of follow-up visits was 5 (IQR, 2-8), in spite of different enrolled time. Table 3 shows characteristics for all patients' visits and three groups according to clinician's prescription. It was found that the occurrence rate of SSE for NOAC group was higher than the other two groups, probably because the clinical status of the patients were worse in these visits. As we discovered, the median age of NOAC group was more than 75 years old, while the median age of warfarin group was less than 75. Moreover, there were more visits with CV score of ≥ 5 in NOAC group than in warfarin group (24.69% vs 21.80%).

Table 2. Baseline characteristics of all patients

Characteristics	**Overall (N=8,540)**	**Characteristics**	**Overall (N=8,540)**
Age,years (Mean±SD)	71.57±8.89	Heart Rate, bmp (Mean±SD)	80.74±20.29
Age≥75	3661 (42.9%)	BMI (Mean±SD)	25.27±3.69
Age 65-74	3383 (39.6%)	Heart Failure	2155 (25.2%)
Age<65	1496 (17.5%)	SSE	1993 (23.3%)
Female Gender	3961 (46.4%)	Major bleeding	236 (2.8%)
CHA_2DS_2-VA score		Vascular Disease	2095 (24.5%)
(Median [IQR])	3 [2,4]	Hypertension	6742 (78.9%)
2	2944 (34.5%)	Diabetes	2746 (32.2%)
3-4	3871 (45.3%)	Beta Blocker	4737 (55.5%)
≥5	1725 (20.2%)	ACEI/ARB	153 (1.8%)
AST>50U/L	289 (3.4%)	Statin	166 (1.9%)
ALT>40U/L	668 (7.8%)	**Follow-up**, year (Mean±SD)	2.72±1.84
eGFR<60 mL/min/1.73m^2	1617 (18.9%)	number (Median [IQR])	5 [2,8]

* Values for continuous variables given as mean ± standard deviation or median [interquartile range]; for categorical variables, as count (percentage). Abbreviations and definitions: AST, aspartate aminotransferase; ALT, alanine aminotransferase; eGFR, estimated glomerular filtration rate; BMI, body mass index; ACEI, angiotensin converting enzyme inhibitors; ARB, angiotensin receptor blocker.

Table 3. Characteristics for all patients' visits and three groups according to clinician's prescription

Clinician's prescription	Visit number	Age, years (Median [IQR])	CHA$_2$DS$_2$-VA score $\geq$ 5 (n(%))	Count of SSE	Occurrence rate of SSE	Count of major bleeding	Occurrence rate of major bleeding
No drug	32,524	75 [68,80]	7,669 (23.58%)	463	1.42%	90	0.28%
Warfarin	17,424	74 [67,78]	3,798 (21.80%)	197	1.13%	178	1.02%
NOAC	2,224	76 [70,80]	549 (24.69%)	45	**2.02%**	20	0.90%
Overall	52,172	74 [68,79]	12,016 (23.03%)	705	1.35%	288	0.55%

Medication patterns

We found that in 64.98% of patient visits, the anticoagulant treatments recommended by the developed RL model were concordant with the actual prescriptions of the clinicians. As depicted in Figure 3, the medication patterns of clinicians' prescriptions and model recommendations were visualized by a 2-D histogram, in which the x axis represents the medication patterns prescribed by clinicians and y axis represents the medication patterns recommended by RL model (value 0 indicates no OAC, value 1 indicates warfarin, and value 2 indicates NOAC). The color indicates the usage number of corresponding medication patterns. The number on the diagonal indicates the number of model-concordant visits in which the actual prescription of the clinician was the same as model-recommended medication. There were 11,123 visits with warfarin and 2,565 visits with NOAC recommended by the RL model, but actually in these visits the patients were not treated with any OAC by clinicians. Similarly, there were 2,812 visits with warfarin and 658 visits with NOAC prescribed by clinicians, but our model recommended no anticoagulation. Table 4 showed the distribution of medication patterns between clinicians' prescriptions and model recommendations. We found that in most of patient visits (62.34%) clinicians did not prescribe any OAC, while in 42.75% of visits our model did not recommend any OAC. The model recommendations preferred to use warfarin, and the percentage of warfarin recommended by our RL model was higher than that prescribed by clinicians.

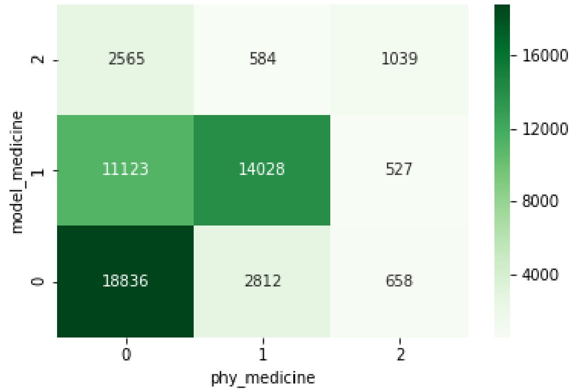

Figure 3. Medication pattern comparison between clinicians' prescriptions and model recommendations.

Table 4. The distribution of medication patterns between clinicians' prescriptions and model recommendations.

Action	Description	Model recommendation		Clinician prescription	
		Number	Ratio	Number	Ratio
0	No drug	22,306	42.75%	32,524	62.34%
1	Warfarin	25,678	49.22%	17,424	33.40%
2	NOAC	4,188	8.03%	2,224	4.26%
Overall		52,172	100%	52,172	100.00%

Since the number of using NOAC is quite small in both clinician's prescription and model recommendation, we combined it with warfarin as "Drug" group. As shown in Table 5, the percentage of using anticoagulation drug were similar among the three groups stratified by CV score in clinicians' prescriptions, while the percentage of using anticoagulation drug increased with the CV score obviously in model recommendations. Specifically, in the groups of visits with CV score ≥ 5, the OAC-using rate (88.02%) in model recommendations was much higher than that (36.18%) in clinicians' prescriptions. As is known that more anticoagulation treatments for the patients with CV score ≥ 5 may lead to good clinical outcome. In other words, the higher the CV score, the more anticoagulant treatment recommended by model, which is consistent with clinical experience.

Table 5. Comparison between model recommendations and clinicians' prescriptions stratified by CHA_2DS_2-VA

CHA_2DS_2-VA	**Total number**	**Model recommendations**		**Clinicians' prescriptions**	
		No drug	**Drug**	**No drug**	**Drug**
2	14,761	10,248 (69.43%)	4,513(30.57%)	9,477 (64.20%)	5,284(35.80%)
3-4	25,395	10,618(41.81%)	14,777(58.19%)	15,378(60.56%)	10,017(39.44%)
≥5	12,016	1,440(11.98%)	10,576(88.02%)	7,669(63.82%)	4,347(36.18%)
Overall	52,172	22,306(42.75%)	29,866(57.25%)	32,524(62.34%)	19,648(37.66%)

Short-term outcomes at patient-visit level

We evaluated the short-term outcomes at patient-visit level. First, we partitioned the patient visits into model-concordant group and model-non-concordant group. Then the short-term clinical outcomes were compared between the two groups in terms of SSE event and major bleeding event, namely the percentages of patient visits with the events. The evaluation results of short-term outcomes were shown in Table 6. Among all the samples, 33,903 visits (64.98%) were model-concordant and 18,269 (35.02%) were non-concordant, and the model-concordant treatments were associated with less SSE event compared with non-concordant ones. We further stratified the patient visits by CHA_2DS_2-VA score, which was the confounder that was most strongly correlated to the clinical outcome. For visit group with CV score of 2, the model-concordant treatments were associated with improved occurrence rate of both SSE and major bleeding events. For visits with CV score of 3 to 4, the result was similar with all visits. For visits with CV score of 5 or greater, the model-concordant treatments were associated with reduced occurrence rate of SSE event but with increased occurrence rate of major bleeding event. However, the amplitude of reduced rate (1.04%) was greater than the increased rate (0.43%).

Table 6. Short-term clinical outcomes comparison between model-concordant and model-non-concordant groups stratified by CHA_2DS_2-VA

CHA_2DS_2-VA	**Model-concordant**	**Total**	**SSE event**				**Major bleeding event**			
			Count of occurrence	**Occurrence rate**	**Diff of rate**	**P-value**	**Count of occurrence**	**Occurrence rate**	**Diff of rate**	**P-value**
2	Yes	11,187	60	0.54%	-0.52%	**<0.01 ****	42	0.38%	-0.52%	**<0.001 *****
	NO	3,574	38	1.06%			32	0.90%		
3-4	Yes	17,436	193	1.11%	-0.61%	**<0.001 *****	110	0.63%	0.19%	0.0742
	NO	7,959	137	1.72%			35	0.44%		
≥5	Yes	5,280	91	1.72%	-1.04%	**<0.001 *****	43	0.81%	0.42%	<0.01 **
	NO	6,736	186	2.76%			26	0.39%		
Overall	Yes	33,903	344	1.01%	-0.97%	**<0.001 *****	195	0.58%	0.07%	0.3627
	NO	18,269	361	1.98%			93	0.51%		

To further evaluate the effectiveness of the RL model, we stratified the patient visits by model actions. Similarly, model-concordant was regarded as the exposure factor in each group. As shown in Table 7, in each group of the model action, the occurrence rate of SSE event for model-concordant treatments was lower than that for model-non-concordant treatments, and the reduction in "1:warfarin" group is significant. Specifically, in "1:warfarin" group, the amplitude of reduced rate in SSE event is greater than the increased rate in major bleeding event. In "0:noDrug" group, the model-concordant treatments were associated with reduced occurrence rate of major bleeding event compared with the model-non-concordant ones. It was noticeable that there was no significant difference in "2:NOAC" group since the sample number was quite small.

Table 7. Short-term clinical outcomes comparison between model-concordant and model-non-concordant groups stratified by model action

Model action	**Model-concordant**	**Total**	**SSE event**				**Major bleeding event**			
			Count of occurrence	**Occurrence rate**	**Diff of rate**	**P-value**	**Count of occurrence**	**Occurrence rate**	**Diff of rate**	**P-value**
0: noDrug	Yes	18,836	169	0.90%	-0.22%	0.2377	52	0.28%	-0.87%	**<0.001 *****
	NO	3,470	39	1.12%			40	1.15%		
1: warfarin	Yes	14,028	156	1.11%	-1.06%	**<0.001 *****	135	0.96%	0.61%	<0.001 ***
	NO	11,650	253	2.17%			41	0.35%		
2: NOAC	Yes	1,039	19	1.83%	-0.36%	0.5608	8	0.77%	0.39%	0.1878
	NO	3,149	69	2.19%			12	0.38%		
Overall	Yes	33,903	344	1.01%	-0.97%	**<0.001 *****	195	0.58%	0.07%	0.3627
	NO	18,269	361	1.98%			93	0.51%		

Long-term outcome at patient level

We evaluated the long-term outcome of death at patient level. For each patient, we first computed the patient's model-concordant rate by dividing the number of model concordant visits by the total number of visits. In order to ensure the rationality of the result, we selected the patients with follow-up time ≥1 year. Then, the patients were divided into different groups according to the patient's model-concordant rate (e.g. every 20% as a group), and the occurrence rate of death in each group was computed. Figure 4 illustrated the relationship between patient's model-concordant rate and the occurrence rate of death outcome. The number next to the red dot indicated the total number of patients with corresponding model-concordant range. It can be found that the curves were downwards trending (p for trend < 0.001). In other words, the higher the patient's model-concordant rate was, the lower the occurrence rate of death was. For model-concordant rate between 0 and 0.2, there were 1360 patients, among which 322 patients died. In the 322 died patients, 262 (81.4%) were not treated with OAC by clinicians, but only 6 (1.9%) were not recommended with anticoagulation by the RL model in all their visits. On the contrary, for model-concordant rate between 0.8 and 1.0, there were 2984 patients, among which 163 patients died. In the 163 died patients, 67 patients (41.1%) were not treated with OAC by clinicians and 50 (30.7%) were not recommended with anticoagulation by the RL model.

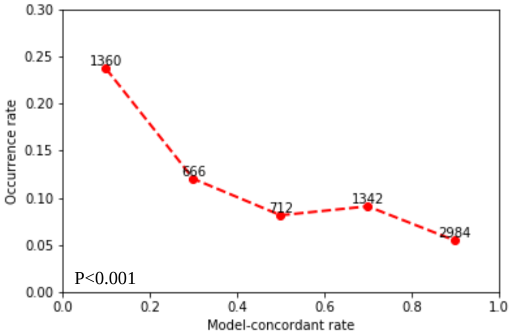

Figure 4. The relationship between patient's model-concordant rate and the occurrence rate of death

Group characterization results

As mentioned above, the RL model preferred to recommend anticoagulation in visits with CHA_2DS_2-VA score of 5 or greater, but it is relatively ambiguous for other visits. In order to suggest clinically interpretable treatment strategies learned by our model, we built a decision tree model to generate rules for visits with CV score of 2 to 4. The decision tree achieved a classification accuracy of 80%. As listed in Table 8, several rules with enough coverage and high confidence (all above 90%) were derived, which could be used to interpret when the RL model recommended anticoagulation or when not. In summary, these rules cover 13,876 visits (35%) out of a total sample size of 40,156, and 4,923 visits (12%) were not recommended anticoagulant treatment. Another rule with confidence of 70% was to be noticeable: if OAC not used, age≥75 and with HF, then anticoagulation was recommended by RL model. This represented that the RL model might have learned when to change the prescription.

Table 8. Rules to characterize the visits with CHA$_2$DS$_2$-VA score of 2 to 4

index	Conditions	Model recommendation	Coverage	Confidence
1	OAC not used, age≥65, no HF, no prior stroke/TIA/TE, no hypertension, no VD	Not on anticoagulation	2,368	93%
2	OAC not used, age<65, no HF, no prior stroke/TIA/TE		2,257	93%
3	OAC not used, age65-74, no HF, no prior stroke/TIA/TE, no hypertension, with VD, no DM		298	92%
4	OAC used, age≥75, no prior stroke/TIA/TE	On anticoagulation	5,467	92%
5	OAC used, age≥65, with prior stroke/TIA/TE		1,740	99%
6	OAC used, age65-74, with HF, no prior stroke/TIA/TE		844	96%
7	OAC used, age65-74, no HF, no prior stroke/TIA/TE, with hypertension, with VD, no DM		302	92%
8	OAC used, age65-74, no HF, no prior stroke/TIA/TE, with VD, with DM		250	97%
9	OAC used, age<65, with HF, with prior stroke/TIA/TE		214	93%
10	OAC used, age<65, with HF, no prior stroke/TIA/TE, with VD		136	95%
Total			13,876	93%

*"OAC not used" means that OAC was not in clinician's prescription in last visit. Abbreviations: HF, heart failure; TIA, transient ischemic attack; TE, thromboembolism; VD, vascular disease; DM, diabetes mellitus.

Discussion

In this study, we proposed a reinforcement learning model to recommend the personalized anticoagulant treatment for AF patients. The strengths of this study are twofold. First, the CAFR data used for model training and evaluation is of good quality. It covers a long follow-up time span of 8 years and consists of AF patients' demographics, symptoms and signs, physical examination and laboratory test, medical history, treatments and clinical events. Second, the novelties of method include RL model training in terms of state, action, reward function and especially loss function, model evaluation at patient-visit level and patient level respectively, and finally the group characterization to interpret the model.

The limitation of the current work is that the number of patients received NOACs therapy and the time of follow up after the initiation of NOACs are limited as these agents are expensive and not covered by medical insurance until recently. Therefore, we combined NOACs as a category rather than individual agents. When more data are accumulated with longer time span, we can update the treatment model and form rules interpreting when to prescribe NOAC, even which type of NOAC. Another limitation is that we evaluated long-term outcome by the relationship between patient's model-concordant rate and the occurrence rate of the long-term outcome without adjusting confounding factors. The cox model can be applied with more comprehensive evaluation.

Our approach was tested with AF patients' data. The method can be applied to treatment of other chronic diseases by designing state, action and reward function.

Conclusion

In this paper, we proposed a reinforcement learning model to recommend the personalized anticoagulant treatment for AF patients and attempted to interpret the model by group characterization. We demonstrated that the proposed RL model can lead to better expected short-term and long-term outcomes and identified several high-confidence rules, which were interpreted by clinical experts. The data used in our work were baseline and follow up data of 8,540 AF patients, enrolled in the CAFR study during 2011 to 2018. While much further study is required to truly mine rules, the proposed method represents a novel approach to take advantage of the strengths of different treatment policies.

Acknowledgments

We thank all the participants in this work for their support. This work was supported by the National Key Research and Development Program of China (2016YFC0900901, 2016YFC1301002, 2017YFC0908803, 2018YFC1312501), and grant from the National Science Foundation of China (81530016).

References

1. Du X, Ma C, Wu J, et al. Rationale and design of the Chinese Atrial Fibrillation Registry study. BMC Cardiovascular Disorders. 2016;16(1):130. doi: 10.1186/s12872-016-0308-1.
2. Benjamin EJ, Wolf PA, D'Agostino RB, Silbershatz H, Kannel WB, Levy D. Impact of atrial fibrillation on the risk of death: the Framingham Heart Study. Circulation. 1998;98(10):946-952.
3. Krahn AD, Manfreda J, Tate RB, Mathewson FA, Cuddy TE. The natural history of atrial fibrillation: incidence, risk factors, and prognosis in the Manitoba Follow-Up Study. The American journal of medicine. 1995;98(5):476-484.
4. Wolf PA, Abbott RD, Kannel WB. Atrial fibrillation as an independent risk factor for stroke: the Framingham Study. Stroke. 1991;22(8):983-988.
5. Verheugt FWA, Christopher BG. Oral anticoagulants for stroke prevention in atrial fibrillation: current status, special situations, and unmet needs. The Lancet. 2015;386(9990):303-310.
6. Coleman CI, Peacock WF, Bunz TJ, Alberts MJ. Effectiveness and safety of apixaban, dabigatran, and rivaroxaban versus warfarin in patients with nonvalvular atrial fibrillation and previous stroke or transient ischemic attack. Stroke. 2017;48(8):2142-2149.
7. Lip GY, Nieuwlaat R, Pisters R, Lane DA, Crijns HJ. Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: the euro heart survey on atrial fibrillation. Chest. 2010;137(2):263-272.
8. Kirchhof P, Benussi S, Kotecha D, et al. 2016 ESC Guidelines for the management of atrial fibrillation developed in collaboration with EACTS. European journal of cardio-thoracic surgery. 2016;50(5):e1-e88.
9. January CT, Wann LS, Calkins H, et al. 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. Journal of the American College of Cardiology. 2019;74(1):104-132.
10. Healey JS, Oldgren J, Ezekowitz M, et al. Occurrence of death and stroke in patients in 47 countries 1 year after presenting with atrial fibrillation: a cohort study. The Lancet. 2016;388(10050):1161-1169.
11. Lip GYH, Andreotti F, Fauchier L, et al. Bleeding risk assessment and management in atrial fibrillation patients. Thrombosis and haemostasis. 2011;106(12):997-1011.
12. Liu H, Li X, Xie G, et al. Precision cohort finding with outcome-driven similarity analytics: a case study of patients with atrial fibrillation. MedInfo. 2017;491-495.
13. Ghassemi MM, AlHanai T, Westover MB, Mark RG, Nemati S. Personalized medication dosing using volatile data streams. The Workshops of the Thirty-Second AAAI Conference on Artificial Intelligence. 2018.
14. Komorowski M, Celi LA, Badawi O, Gordon AC, Faisal AA. The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. Nature medicine. 2018;24(11):1716-1720.
15. Ngo PD, Wei S, Holubová A, Muzik J, Godtliebsen F. Reinforcement-learning optimal control for type-1 diabetes. 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI). 2018;333-336.
16. Jalalimanesh A, Haghighi HS, Ahmadi A, Soltani M. Simulation-based optimization of radiotherapy: Agent-based modelingand reinforcement learning. Mathematics and Computers in Simulation. 2017;133:235-248.
17. Mnih V, Kavukcuoglu K, Silver D, et al. Playing atari with deep reinforcement learning. NIPS. 2013.
18. Friberg L, Tabrizi F, Englund A. Catheter ablation for atrial fibrillation is associated with lower incidence of stroke and death: data from Swedish health registries. European heart journal. 2016;37(31):2478-2487.
19. Kaplan RM, Koehler J, Ziegler PD, Sarkar S, Zweibel S, Passman RS. Stroke risk as a function of atrial fibrillation duration and CHA_2DS_2-VASc score. Circulation. 2019;140(20):1639-1646.
20. Van Hasselt H, Guez A, Silver D. Deep reinforcement learning with double q-learning. Thirtieth AAAI conference on artificial intelligence. 2016.
21. Wang Z, Schaul T, Hessel M, van Hasselt H, Lanctot M, De Freitas N. Dueling network architectures for deep reinforcement learning. In International Conference on Machine Learning (ICML). 2016.
22. Schaul T, Quan J, Antonoglou I, Silver D. Prioritized experience replay. In Proceedings of the International Conference on Learning Representations. 2016. volume abs/1511.05952.
23. Piot B, Geist M, Pietquin O. Boosted bellman residual minimization handling expert demonstrations. In European Conference on Machine Learning (ECML). 2014.
24. Chen SY, Lee YC, Alas V, Greene M, Brixner D. Outcomes associated with nonconcordance to national kidney foundation guideline recommendations for oral antidiabetic drug treatments in patients with concomitant type 2 diabetes and chronic kidney disease. Endocrine Practice. 2014;20(3):221-231.