HIR
Healthcare Informatics Research

Association Rules to Identify Complications of Cerebral Infarction in Patients with Atrial Fibrillation

Sun-Ju Jung, MS[1], Chang-Sik Son, PhD[2], Min-Soo Kim, PhD[2], Dae-Joon Kim, MD[3], Hyoung-Seob Park, MD[4], Yoon-Nyun Kim, MD, PhD[2,4]

[1]Departments of Medical Informatics and [2]Biomedical Informatics Technology Center, Keimyung University School of Medicine, Daegu; [3]Department of Thoracic and Cardiovascular Surgery, Yonsei University College of Medicine, Seoul; [4]Department of Internal Medicine, Keimyung University School of Medicine, Daegu, Korea

Objectives: The purpose of this study was to find risk factors that are associated with complications of cerebral infarction in patients with atrial fibrillation (AF) and to discover useful association rules among these factors. **Methods:** The risk factors with respect to cerebral infarction were selected using logistic regression analysis with the Wald's forward selection approach. The rules to identify the complications of cerebral infarction were obtained by using the association rule mining (ARM) approach. **Results:** We observed that 4 independent factors, namely, age, hypertension, initial electrocardiographic rhythm, and initial echocardiographic left atrial dimension (LAD), were strong predictors of cerebral infarction in patients with AF. After the application of ARM, we obtained 4 useful rules to identify complications of cerebral infarction: age (>63 years) and hypertension (Yes) and initial ECG rhythm (AF) and initial Echo LAD (>4.06 cm); age (>63 years) and hypertension (Yes) and initial Echo LAD (>4.06 cm); hypertension (Yes) and initial ECG rhythm (AF) and initial Echo LAD (>4.06 cm); age (>63 years) and hypertension (Yes) and initial ECG rhythm (AF). **Conclusions:** Among the induced rules, 3 factors (the initial ECG rhythm [i.e., AF], initial Echo LAD, and age) were strongly associated with each other.

Keywords: Atrial Fibrillation, Cerebral Infarction, Risk Factors, Association Learning, Data Mining

I. Introduction

Atrial fibrillation (AF), the most common type of arrhythmia, decreases cardiac function by halting cardiac activity and causing irregular ventricular activity, and it is the main cause of 85% of systemic thromboembolic events and 33% of strokes resulting from blood clots in the atria [1]. Among the AF-induced cardiac diseases, cerebral infarction results in a large number of lesions with a high recurrence rate, and it is therefore associated with high mortality and morbidity rates. Therefore, it is of utmost importance to prevent strokes, for which it is essential to precisely identify the risk factors involved [2]. For AF, the annual incidence rate of thromboembolism is approximately 4%–6%, and its risk factors include congestive heart failure, coronary artery disease, hypertension, age (>65 years), diabetes mellitus, mitral valve

disease, and history of thromboembolism. Vigorous research is being conducted to identify the risk factors of strokes that are complications resulting from AF [3].

With the recent introduction of the Electronic Medical Record (EMR) and Electronic Health Record (EHR), collecting various types of clinical data has become easier. However, the clinical data within a hospital contain various types of unclear and incomplete data that are difficult to comprehend. As such, it is difficult to identify important factors related to the diagnosis or prognosis of certain diseases or to obtain meaningful knowledge from these types of clinical data [4,5]. In clinical medicine, the independent risk factors for the corresponding disease are generally extracted using multivariate statistical analysis utilizing logistic regression analysis to identify the risk factors associated with the diagnosis or the prognosis of certain diseases and to design a diagnosis and prediction model based on this. This method may be most effective for identifying independent risk factors of certain diseases; however, this method has a limitation, that is, the extraction of the relationships among the risk factors may be difficult. To compensate for this limitation, many previous studies have proposed methods using various data mining techniques [5-10]. In particular, some studies [5,10] have proposed a hybrid decision model (e.g., "multivariate statistical analysis and decision tree" and "rough set and decision tree"), which combines the advantages of the statistical analysis technique and machine learning techniques, and to verify its effectiveness, this model was applied to uncover useful information for acute appendicitis and heart failure. As with the method described in the these studies [5,10], we combined multivariate statistical analysis with the association rule mining technique to analyze the risk factors for cerebral infarction, a complication of AF, and we discuss the methods of obtaining useful information for decision making.

II. Methods

1. Subjects
The study included 1,134 patients with AF who were among the patients who visited the outpatient clinic of the Dongsan Medical Center in Daegu, Korea, between September 1983 and September 2010. Medical records were collected on demographic characteristics, medical history, initial electrocardiographic (ECG) findings, and initial echocardiographic (Echo) findings; samples with missing values were excluded. To obtain information related to the risk factors associated with cerebral infarction in patients with AF, 227 patients with cerebral infarction complications were selected from among those with AF for the study group. 907 patients with-

out cerebral infarction complications were recruited for the control group.

2. Statistical Analysis
To compare the differences between the study group and the control group, the chi-square test or Fisher's exact test was performed for category variables. For continuous variables, the student's t-test was carried out if they satisfied the normality after the Kolmogorov-Smirnov test; otherwise, the Mann-Whitney test was used if they did not. Here, the distribution of the category variables were given as percentages (%), and the continuous variables were given as mean ± standard deviation. In addition, the level of statistical significance was defined as $p < 0.05$. To identify the independent risk factors for cerebral infarction, multivariate statistical analysis was conducted on the significant factors with the p-value of 0.05 and 0.10 for entry and removal, respectively. All statistical analysis were carried out using the SPSS ver. 12.0 (SPSS Inc., Chicago, IL, USA).

3. Association Rule
The present study used an a priori algorithm [11] to obtain the information related to cerebral infarction. The a priori algorithm is one of the typical association rule mining methods with the generated rule expressed as R: IF A THEN B (or A => B) to extract association rules based on 2 variable

$$Supp(R) = \frac{|A \cap B|}{|N|}, \; Supp(R) \in [0,1]$$

$$Conf(R) = \frac{|A \cap B|}{|A|}, \; Conf(R) \in [0,1]$$

$$Lift(R) = \frac{|A \cap B|}{|A| \times |B|} = \frac{P(B|A)}{P(B)}, \; Lift(R) \in [0,\infty]$$

parameters, i.e., support and confidence. In addition, the generated rules are judged based on the lift or improvement scale [12].
Here, |.| represents the cardinality of the set, i.e., the number of the element, and N represents the number of samples in the entire data.

In general, support represents the ratio of samples that simultaneously satisfy both condition A in the antecedent part and condition B in the consequent part in the entire set of samples, while confidence represents the ratio of samples that satisfy both conditions A and B, among those that satisfy condition A in the antecedent part. A confidence value of 1 for a certain rule means that the possibility of obtaining outcome B when A is a given condition (A → B) is 100% (i.e., certain rule); if not, the possibility of A → B is defined

as a value (possible rule) between 0 and 1. In addition, the lift or improvement value of ≥1 represents a positive correlation, while a value of <1 represents a negative correlation. Therefore, more general or useful information would have a higher confidence level, with a lift or improvement value of ≥1. However, as discussed above, it is difficult to determine appropriate values for the two free parameters in association rule mining, because information must be obtained based on the minimum threshold for support and for confidence. As such, in this study, the minimum confidence level was variably adjusted to 10%–50%, when the minimum support was defined as 10%; cardiology specialists were consulted for the association rules generated here, while the final confidence level was determined by a clinical specialist's opinion. Furthermore, the a priori component of a commercial data mining program, Clementine ver. 12.0 (SPSS Inc., Chicago, IL, USA), was used for these experiments, and default values were used for the experimental parameters.

III. Results

1. Characteristics of Study Patients

Upon comparing the general statistical characteristics of the study group and the control group, it was found that the mean age was higher in the study group (69.31 ± 9.04 years) than in the control group (66.17 ± 11.22 years), with statistical significance ($p < 0.001$). As to medical history, 116 patients (51.1%) in the study group had hypertension compared to 342 patients (37.7%) in the control group ($p < 0.001$), while there were 43 patients (18.9%) in the study group who had coronary artery disease compared with 117 patients (12.9%) in the control group, also showing statistical significance ($p < 0.05$). With regard to the initial ECG findings, statistically significant ($p < 0.01$) differences were observed in ECG rhythm for AF with 205 patients in the study group (90.3%) and 738 patients in the control group (81.4%), for AF, with 4 patients in the study group (1.8%) and 15 patients in the control group (1.7%), and for normal sinus rhythm with 18 patients in the study group (7.9%) and 154 patients in the control group (17.0%). In addition, with regard to the type of AF, 58 patients from the study group (25.6%) and 325 patients from the control group (35.8%) showed paroxysmal AF, 17 patients from the study group (7.5%) and 73 patients from the control group (8.0%) showed persistent AF, and 152 patients from the study group (67.0%) and 509 patients from the control group (56.1%) showed permanent AF, with both the study and the control groups showing a relatively high frequency of permanence compared to the other 2 types, both with statistically significant difference ($p < 0.01$) (Table 1).

2. Clinical Reference for Continuous Variables

In the present study, the area under the receiver operating characteristic curve (AUC) was used to determine the clinical reference or criteria for independent risk factors with continuous attribute values. The data point value that provided the best AUC value in the ROC curve was selected as the cutoff value for each independent factor. Table 2 shows the independent variables having continuous attribute values; i.e., the clinical references for age, initial ECG heart rate, and for initial Echo ejection fraction (EF, %), diastolic left ventricular dimension (LVDd, cm), systolic left ventricular dimension (LVDs, cm), and left atrial dimension (LAD, cm), as well as the standard error and the AUC. As shown by the results in Table 2, findings include age older than 63 years, heart rate higher than 78 beats/min, EF ≤63%, LVDd ≤4.87 cm, LVDs >3.04 cm, and LAD >4.06 cm. In addition, the results of comparing the statistical characteristics between the study group and the control group after differentiating these 6 variables are shown in Table 3, with a statistically significant difference ($p < 0.01$) for LAD.

3. Risk Factors Associated with Cerebral Infarction in Patients with Atrial Fibrillation

Binary logistic regression analysis was used to extract the risk factors associated with complications of cerebral infarction from the 6 factors (age, hypertension, coronary artery disease, initial ECG rhythm, AF type, and initial Echo LAD) that were selected after univariate statistical analysis. The results showed that age, hypertension, initial ECG rhythm, and initial Echo LAD are the independent risk factors associated with cerebral infarction, with the risk increasing by 1.949 times for patients older than 63 years, 1.587 times for patients with a history of hypertension, 2.026 times when AF is the initial ECG rhythm, and 1.482 times when the LAD on initial Echo findings exceeded 4.06 cm. Furthermore, the Hosmer-Lemeshow goodness-of-fit test results showed that the model was appropriate, because the significance level for the chi-square value was 0.538 (Table 4).

4. Rules Associated with Cerebral Infarction

To obtain clinically reliable rules to determine the complications of cerebral infarction, the association rules were generated for the case when the minimum confidence level was adjusted to 10%–50%. As a result, the most reliable rule set could be obtained when the minimum confidence level was defined at 30%, yielding 44 rules (40 rules related to AF and 4 related to cerebral infarction) (Table 5). As shown in Table 5, the first rule showed the highest confidence level and improvement scale, which could be interpreted as follows.

Table 1. Characteristics of study patients between study and control groups (n = 1,134)

	Variable		Non-CI	CI	p-value
Demographic	Age (yr)		66.17 ± 11.22	69.31 ± 9.04	0.000
	Sex	Male	516 (56.9)	135 (59.5)	0.482
		Female	391 (43.1)	92 (40.5)	
History	Smoking	No	581 (64.1)	141 (62.1)	0.751
		Yes	177 (19.5)	44 (19.4)	
		Ex	149 (16.4)	42 (18.5)	
	Alcohol use	No	590 (65.0)	137 (60.4)	0.159
		Yes	278 (30.7)	74 (32.6)	
		Ex	39 (4.3)	16 (7.0)	
	Hypertension	No	565 (62.3)	111 (48.9)	0.000
		Yes	342 (37.7)	116 (51.1)	
	Diabetes mellitus	No	792 (87.3)	191 (84.1)	0.207
		Yes	115 (12.7)	36 (15.9)	
	Coronary artery disease	No	790 (87.1)	184 (81.1)	0.019
		Yes	117 (12.9)	43 (18.9)	
	Heart valve disease	No	774 (85.3)	200 (88.1)	0.284
		Yes	133 (14.7)	27 (11.9)	
	Congestive heart failure	No	729 (80.4)	177 (78.0)	0.419
		Yes	178 (19.6)	50 (22.0)	
	Thyroid disease	No	843 (92.9)	209 (92.1)	0.650
		Yes	64 (7.1)	18 (7.9)	
	Cerebral hemorrhage	No	903 (99.6)	225 (99.1)	0.414
		Yes	4 (0.4)	2 (0.9)	
Initial ECG	Heart rate (beats/min)		86.30 ± 26.78	87.50 ± 26.13	0.363
	Rhythm	AF	738 (81.4)	205 (90.3)	0.003
		A.Flutter	15 (1.7)	4 (1.8)	
		NSR	154 (17.0)	18 (7.9)	
	Hypertrophy	No	543 (59.9)	143 (63.0)	0.389
		Yes	364 (40.1)	84 (37.0)	
	Bundle branch block BB	No	874 (96.4)	215 (94.7)	0.519
		Right	28 (3.1)	10 (4.4)	
		Left	5 (0.6)	2 (0.9)	
	ST-T wave changes	No	795 (87.7)	193 (85.0)	0.290
		Yes	112 (12.3)	34 (15.0)	
	Type	Paroxysmal	325 (35.8)	58 (25.6)	0.009
		Persistent	73 (8.0)	17 (7.5)	
		Permanent	509 (56.1)	152 (67.0)	
Initial Echo	EF (%)		59.71 ± 12.35	59.48 ± 12.39	0.655
	LVDd (cm)		5.14 ± 0.64	5.13 ± 0.62	0.803
	LVDs (cm)		3.49 ± 0.76	3.49 ± 0.73	0.550
	LAD (cm)		4.50 ± 0.76	4.63 ± 0.69	0.016[a]
	LV dysfunction	No	827 (91.2)	198 (87.2)	0.071
		Yes	80 (8.8)	29 (12.8)	
	LA thrombus	No	906 (99.9)	227 (100)	0.617
		Yes	1 (0.1)	0 (0)	

Values are presented as number (%) or mean ± standard deviation.

ECG: electrocardiogram, Non-CI: non cerebral infarction, CI: cerebral infarction, DM: diabetes mellitus, CHF: congestive heart failure, AF: atrial fibrillation, A.Flutter: atrial flutter, NSR: normal sinus rhythm, BBB: bundle branch block, RBBB: right bundle branch block, LBBB: left bundle branch block, Echo: echocardiogram, EF: ejection fraction, LVDd: diastolic left ventricular dimension, LVDs: systolic left ventricular dimension, LAD: left atrial dimension, LV: left ventricular, LA: left atrial.

[a]Mann-Whitney U test.

Table 2. Clinical reference for continuous variables

	Variable	Cutoff	AUC	SE	95% CI	*p*-value
Demographic	Age (yr)	>63	0.583	0.020	0.553−0.613	0.000
Initial ECG	Heart rate (beats/min)	>78	0.525	0.020	0.497−0.552	0.210
Initial Echo	EF (%)	≤63	0.504	0.020	0.476−0.532	0.848
	LVDd (cm)	≤4.87	0.505	0.020	0.477−0.533	0.815
	LVDs (cm)	>3.04	0.504	0.020	0.476−0.533	0.829
	LAD (cm)	>4.06	0.547	0.019	0.518−0.575	0.017

AUC: area under curve, SE: standard error, CI: confidence interval, ECG: electrocardiogram, Echo: echocardiogram, EF: ejection fraction, LVDd: diastolic left ventricular dimension, LVDs: systolic left ventricular dimension, LAD: left atrial dimension.

Table 3. Statistical analysis after dichotomization of continuous variables

	Variable	Reference	Non-CI (control group)	CI (study group)	*p*-value
Demographic	Age (yr)	≤63	351 (38.7)	50 (22.0)	0.000
		>63	556 (61.3)	177 (78.0)	
Initial ECG	Heart rate (beats/min)	≤78	407 (44.9)	86 (37.9)	0.058
		>78	500 (55.1)	141 (62.1)	
Initial Echo	EF (%)	>63	398 (43.9)	89 (39.2)	0.203
		≤63	509 (78.7)	138 (60.8)	
	LVDd (cm)	>4.87	626 (69.0)	147 (64.8)	0.218
		≤4.87	281 (31.0)	80 (35.2)	
	LVDs (cm)	≤3.04	268 (29.5)	54 (23.8)	0.085
		>3.04	639 (70.5)	173 (76.2)	
	LAD (cm)	≤4.06	266 (29.3)	41 (18.1)	0.001
		>4.06	641 (70.7)	186 (81.9)	

Values are presented as number (%).

CI: cerebral infarction, ECG: electrocardiogram, Echo: echocardiogram, EF: ejection fraction, LVDd: diastolic left ventricular dimension, LVDs: systolic left ventricular dimension, LAD: left atrial dimension.

Those who were older than 63 years, had a history of hypertension, showed AF on the initial ECG Rhythm, and had initial Echo LAD of over 4.06 cm comprised 7% of the total 1,134 patients, with a confidence level of approximately 33% and an improvement scale of the rule of 1.7 times showing a positive correlation.

As previously mentioned, Table 6 shows the change in the number of corresponding rules for each group when the minimum confidence level was adjusted to 10%−50%, and the rules associated with cerebral infarction could not be obtained when the minimum level of confidence was adjusted to 0.35−0.50 (35%−50%). In addition, the results of Web node analysis conducted to examine the relationship between the fields showed that the factors most closely associated with cerebral infarction in patients with AF include the following in decreasing order of association: AF as the rhythm on initial ECG, LAD of greater than 4.06 cm on ini-

tial Echo, age older than 63 years in terms of demographic characteristics, and hypertension in terms of medical history (Figure 1).

IV. Discussion

Atrial fibrillation, the most common supraventricular arrhythmia, in which irregular atrial muscle contractions produce an irregular pulse, is known as a major risk factor of thromboembolism. Furthermore, it leads to strokes and causes hemodynamic instability, deterioration of renal function, and systemic embolic events [13,14].

In the present study, multivariate statistical analysis using logistic regression analysis was conducted to extract the risk factors for cerebral infarction complications in patients with AF, and the relationship between these factors was analyzed by applying the association rule mining technique. As a re-

Table 4. Multivariate analysis of predictors of cerebral infarction (entry and removal criteria of 0.05 and 0.10)

Variable			Coefficient	SE	OR (95% CI)	p-value	H^a
Demographic	Age (yr)	>63	0.667	0.178	1.949 (1.376−2.760)	0.000	0.538
History	Hypertension	Yes	0.462	0.153	1.587 (1.176−2.140)	0.002	-
Initial ECG	Rhythm	-	-	-	-	0.032	-
		AF	0.706	0.268	2.026 (1.197−3.428)	0.009	-
		A.Flutter	0.616	0.629	1.852 (0.540−6.350)	0.327	-
Initial Echo	LAD (cm)	>4.06	0.393	0.194	1.482 (1.013−2.168)	0.043	-
Intercept		-	-2.980	0.304	0.051	0.000	

R^2 = 0.042, n = 1,134. Reference categories of 4 independent predictors: Age ≤63 years, Hypertension (No), ECG rhythm (NSR), and Echo LAD ≤4.06 cm.

SE: standard error, OR: odds ratio, CI: confidence interval, ECG: electrocardiogram, AF: atrial fibrillation, A.Flutter: atrial flutter, Echo: echocardiogram, LAD: left atrial dimension, NSR: normal sinus rhythm.

[a]Hosmer-Lemeshow goodness-of-fit (H) statistic.

Table 5. Rules related to cerebral infarction

Rule		No.[a]	Support (%)	Confidence (%)	Lift
Age (>63 yr) Hypertension (Yes) Initial ECG rhythm (AF) Initial Echo LAD (>4.06 cm)	→ Cerebral infarction	233	6.88	33.48	1.67
Age (>63 yr) Hypertension (Yes) Initial Echo LAD (>4.06 cm)	→ Cerebral infarction	267	7.50	31.84	1.59
Hypertension (Yes) Initial ECG rhythm (AF) Initial Echo LAD (>4.06 cm)	→ Cerebral infarction	308	8.29	30.52	1.53
Age (>63 yr) Hypertension (Yes) Initial ECG rhythm (AF)	→ Cerebral infarction	279	7.50	30.47	1.52

ECG: electrocardiogram, AF: atrial fibrillation, Echo: echocardiogram, LAD: left atrial dimension.

[a]Number of samples (or cases) that satisfy the antecedent part of generated rules.

sult, the independent risk factors associated with cerebral infarction complications were found to include age, hypertension, initial ECG rhythm, and initial Echo LAD, and the following information associated with cerebral infarction could be obtained: 1) age >63 years, hypertension is present, initial ECG rhythm is AF, initial Echo LAD >4.06 cm => cerebral infarction (support, 6.88%; confidence, 33.48%); 2) age >63 years, hypertension is present, initial Echo LAD >4.06 cm => cerebral infarction (support, 7.50%; confidence, 31.84%); 3) hypertension is present, initial ECG rhythm is AF, initial Echo LAD >4.06 cm => cerebral infarction (support, 8.29%; confidence, 30.52%); and 4) age >63 years, hy-

pertension is present, initial ECG rhythm is AF => cerebral infarction (support, 7.50%; confidence, 30.47%). In addition, the analysis results using web node revealed AF as the initial ECG rhythm to be the factor most closely associated with cerebral infarction in patients with AF, followed by initial Echo LAD >4.06 cm, age >63 years, and hypertension in decreasing order of association.

An existing numeric tool that can be used to estimate the risk of stroke in patients with atrial fibrillation is CHADS$_2$ (congestive heart failure, hypertension, age, diabetes mellitus, prior stroke or TIA or thromboembolism [double]) score [15]. In this tool, 1 point is given for congestive heart

Table 6. Number of changed rules according to confidence levels

Confidence	No. of rules		
	Total	CI	Non-CI
0.10	79	39	40
0.15	69	29	40
0.20	59	19	40
0.25	51	11	40
0.30	44	4	40
0.35	40	-	40
0.40	40	-	40
0.45	40	-	40
0.50	40	-	40

CI: cerebral infarction.

Figure 1. Web node. ECG: electrocardiogram, AF: atrial fibrillation, Echo: echocardiogram, LAD: left atrial dimension.

failure, hypertension, age 75 years or older, and medical history of diabetes mellitus, and 2 points are given for history of stroke or transient cerebral ischemic attack, to categorize a low-risk group (0 point), a moderate-risk group (1 point), and a high-risk group (more than 2 points). In addition, the CHA_2DS_2-VAS_c (congestive heart failure/left ventricular dysfunction, hypertension, age ≥75 [doubled], diabetes, stroke [doubled], vascular disease, age 65–74, and sex category [female]) score (Birmingham 2009 scheme), based on the new guidelines reported at the European Society of Cardiology in 2010, is a more detailed stroke risk assessment tool than the previous $CHADS_2$ score, because it includes the risk factors (female, 65–75 years of age, left ventricular dysfunction, vascular diseases) that affect thromboembolism in patients whose $CHADS_2$ score is between 0 and 1 [16]. In the two stroke assessment tools mentioned above, the factors used to assess the risk of stroke in patients with atrial fibrillation include age, hypertension, diabetes mellitus, stroke or history of transient ischemic attack, sex, left ventricular dysfunction, and vascular diseases. These factors show a consistent trend with the 4 factors that are suggested in this study as risk factors (age, hypertension, AF as the initial ECG rhythm, and initial Echo LAD).

The results of the present study suggest the risk factors for complications of cerebral infarction in atrial fibrillation patients and the association rules between these factors, based on medical record data collected retrospectively. However, the effectiveness and reliability of these risk factors and the association rules suggested in this study have yet to be verified for clinical application; further research is required for such verification in addition to comparative studies with existing stroke assessment tools.
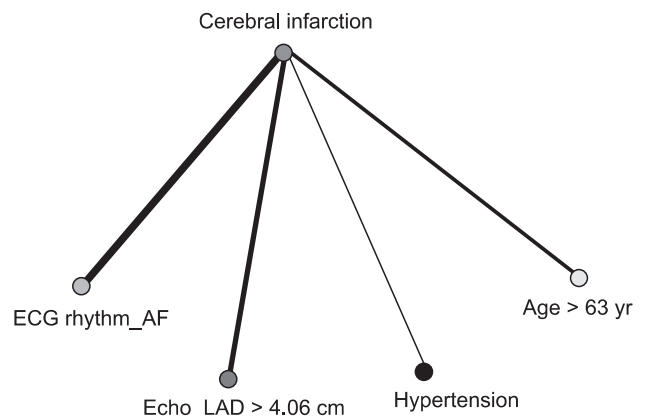
Conflict of Interest

No potential conflict of interest relevant to this article was reported.

Acknowledgments

References

1. Marini C, De Santis F, Sacco S, Russo T, Olivieri L, Totaro R, et al. Contribution of atrial fibrillation to incidence and outcome of ischemic stroke: results from a population-based study. Stroke 2005;36(6):1115-9.

2. Sherman DG, Goldman L, Whiting RB, Jurgensen K, Kaste M, Easton JD. Thromboembolism in patients with atrial fibrillation. Arch Neurol 1984;41(7):708-10.

3. The Stroke Prevention in Atrial Fibrillation Investigators. Predictors of thromboembolism in atrial fibrillation: I. clinical features of patients at risk. Ann Intern Med 1992;116(1):1-5.

4. Lee SM, Park RW. Basic concepts and principles of data mining in clinical practice. J Korean Soc Med Inform 2009;15(2):175-89.

5. Son CS, Jang BK, Seo ST, Kim MS, Kim YN. A hybrid decision support model to discover informative knowledge in diagnosing acute appendicitis. BMC Med Inform Decis Mak 2012;12:17.

6. Prather JC, Lobach DF, Goodwin LK, Hales JW, Hage ML, Hammond WE. Medical data mining: knowledge discovery in a clinical data warehouse. Proc AMIA Annu Fall Symp 1997:101-5.

7. Richards G, Rayward-Smith VJ, Sonksen PH, Carey S, Weng C. Data mining for indicators of early mortality in a database of clinical records. Artif Intell Med 2001;22(3):215-31.

8. Doddi S, Marathe A, Ravi SS, Torney DC. Discovery of association rules in medical data. Med Inform Internet Med 2001;26(1):25-33.

9. Mullins IM, Siadaty MS, Lyman J, Scully K, Garrett CT, Miller WG, et al. Data mining and clinical data repositories: insights from a 667,000 patient data set. Comput Biol Med 2006;36(12):1351-77.

10. Son CS, Kim YN, Kim HS, Park HS, Kim MS. Decision-making model for early diagnosis of congestive heart failure using rough set and decision tree approaches. J Biomed Inform 2012;45(5):999-1008.

11. Agrawal R, Srikant R. Fast algorithms for mining association rules in large databases. In: Bocca J, Jarke M, Zaniolo C, editors. Proceedings of the 20th International Conference on Very Large Data Bases; 1994 Sep 12-15; Santiago de Chile, Chile. San Francisco, CA: Morgan Kaufmann; 1994. p. 487-99.

12. Shin AM. System design and implementation of disease association rule mining using large electronic medical record data. Daegu, Korea: Keimyung University; 2011.

13. Go AS, Hylek EM, Phillips KA, Chang Y, Henault LE, Selby JV, 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. JAMA 2001;285(18):2370-5.

14. Olsson SB, Halperin JL. Prevention of stroke in patients with atrial fibrillation. Semin Vasc Med 2005;5(3):285-92.

15. Fuster V, Ryden LE, Cannom DS, Crijns HJ, Curtis AB, Ellenbogen KA, et al. ACC/AHA/ESC 2006 Guidelines for the management of patients with atrial fibrillation: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the European Society of Cardiology Committee for Practice Guidelines (writing committee to revise the 2001 guidelines for the management of patients with atrial fibrillation): developed in collaboration with the European Heart Rhythm Association and the Heart Rhythm Society. Circulation 2006;114(7):e257-354.

16. 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):263-72.