SUPPLEMENTAL MATERIALS

Predicting cerebral infarction in patients with atrial fibrillation

using machine learning: the Fushimi AF study

Table I. Hyperparameters of the machine learning model

Algorithm	Hyperparameter	Searched value or category
CatBoost	Tree depth	2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15
	Random strength	1e-9, 1e-8, 1e-7, 1e-6, 1e-5, 1e-4, 1e-3, 1e-2, 1e-1, 1, 10
	Bagging temperature	0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0
	Border count	1−255
	Scale pos weight	0.01, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50, 0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95, 1.00
	Iterations	100, 200, 300, 400, 500, 600, 700, 800, 900, 1000

Table II. The baseline variables after data preprocessing and its missing data

Variables (n=68)	Number of missing data	% of missing data
Age	0	0.0
Sex	0	0.0
Hight	84	8.4
Weight	56	5.6
Body mass index	90	9.0
Body surface area	90	9.0
Systolic blood pressure	3	3.0
Diastolic blood pressure	3	3.0
Pulse rate	2	0.2
Baseline cardiac rhythm on electrocardiogram	0	0.0
Type of atrial fibrillation	0	0.0
Palpitation	0	0.0
Shortness of breath	0	0.0
Fatigue	0	0.0
Chest pain	15	1.5
Dizziness	15	1.5
Previous stroke (cerebral infarction / transient ischemic attack)	0	0.0
Previous intracranial hemorrhage	0	0.0
Previous systemic thromboembolism	0	0.0
Location of systemic thromboembolism	0	0.0
Previous major bleeding	0	0.0
Congestive heart failure	0	0.0
The New York Heart	0	0.0

Association functional classification		
Mitral regurgitation	0	0.0
Atrial regurgitation	0	0.0
Mitral valve stenosis	0	0.0
Atrial valve stenosis	0	0.0
Mitral valve prolapse	0	0.0
Surgical history of valvular heart disease	0	0.0
Hypertrophic cardiomyopathy	0	0.0
Dilated cardiomyopathy	0	0.0
Hypertension	0	0.0
Diabetes mellitus	0	0.0
Dyslipidemia	0	0.0
Coronary artery disease	0	0.0
Previous myocardial infarction	0	0.0
Previous percutaneous coronary intervention	0	0.0
Previous coronary artery bypass grafting	0	0.0
Peripheral artery disease	0	0.0
Chronic kidney disease	0	0.0
Hemodialysis	0	0.0
Chronic obstructive pulmonary disease	0	0.0
Pacemaker implantation	0	0.0
Implantable cardioverter defibrillator	0	0.0
Cardiac resynchronization therapy	0	0.0
History of cardiac ablation	0	0.0
History of direct cardioversion	0	0.0

Smoking history	259	25.8
Intake of warfarin	0	0.0
Intake of direct oral anticoagulant	0	0.0
Intake of antiplatelet drug	0	0.0
Intake of antiarrhythmic drug	0	0.0
Creatine phosphokinase	235	23.4
Creatinine	40	4.0
Blood urea nitrogen	56	5.6
Creatinine clearance	92	9.2
Triglyceride	160	15.9
High-density lipoprotein cholesterol	264	26.3
Low-density lipoprotein cholesterol	258	25.7
Blood sugar	114	11.3
Cardio-thoracic ratio on chest X-ray	124	12.3
Ejection fraction	215	21.4
Left atrial dimension	222	22.1
Relative wall thickness	226	22.5
Left ventricular mass	226	22.5
Left ventricular mass index	273	27.2
Wall motion asynergy	199	19.8

Figure I. the important variables for each algorithm in the variable selection step