



# Explainable Machine Learning for Atrial Fibrillation in the General Population Using a Generalized Additive Model

## — A Cross-Sectional Study —

Masaki Kawakami; Shigehiro Karashima, MD, PhD; Kento Morita;  
Hayato Tada, MD, PhD; Hirofumi Okada, MD, PhD; Daisuke Aono, MD, PhD;  
Mitsuhiro Kometani, MD, PhD; Akihiro Nomura, MD, PhD; Masashi Demura, MD, PhD;  
Kenji Furukawa, MD, PhD; Takashi Yoneda, MD, PhD;  
Hidetaka Nambo; Masa-aki Kawashiri, MD, PhD

**Background:** Atrial fibrillation (AF) is the most common arrhythmia and is associated with increased thromboembolic stroke risk and heart failure. Although various prediction models for AF risk have been developed using machine learning, their output cannot be accurately explained to doctors and patients. Therefore, we developed an explainable model with high interpretability and accuracy accounting for the non-linear effects of clinical characteristics on AF incidence.

**Methods and Results:** Of the 489,073 residents who underwent specific health checkups between 2009 and 2018 and were registered in the Kanazawa Medical Association database, data were used for 5,378 subjects with AF and 167,950 subjects with normal electrocardiogram readings. Forty-seven clinical parameters were combined using a generalized additive model algorithm. We validated the model and found that the area under the curve, sensitivity, and specificity were 0.964, 0.879, and 0.920, respectively. The 9 most important variables were the physical examination of arrhythmia, a medical history of coronary artery disease, age, hematocrit,  $\gamma$ -glutamyl transpeptidase, creatinine, hemoglobin, systolic blood pressure, and HbA1c. Further, non-linear relationships of clinical variables to the probability of AF diagnosis were visualized.

**Conclusions:** We established a novel AF risk explanation model with high interpretability and accuracy accounting for non-linear information obtained at general health checkups. This model contributes not only to more accurate AF risk prediction, but also to a greater understanding of the effects of each characteristic.

**Key Words:** Atrial fibrillation; General population; Generalized additive model; Machine learning; Prediction

Atrial fibrillation (AF) is the most common arrhythmia, with a global prevalence estimated at high as 46.3 million, with annual increases of 3.8 million cases in recent years.<sup>1–4</sup> AF is associated with an increased risk of thromboembolic stroke and heart failure, which is not only life threatening, but can also result in severe sequelae, leaving patients with severe mobility impairments or requiring nursing care.<sup>5,6</sup> Individuals with

AF have an approximately 5-fold higher risk of stroke than those without AF.<sup>7</sup> Therefore, the development of AF should be predicted at the early onset.

Artificial intelligence (AI) is gaining attention in preventive medicine, and there is an urgent need for the detection and prediction of AF. AI methods can automatically identify obscure but significant patterns from individual clinical data to predict the clinical outcomes of a patient.

Received December 2, 2021; accepted December 2, 2021; J-STAGE Advance Publication released online December 28, 2021 Time for primary review: 1 day

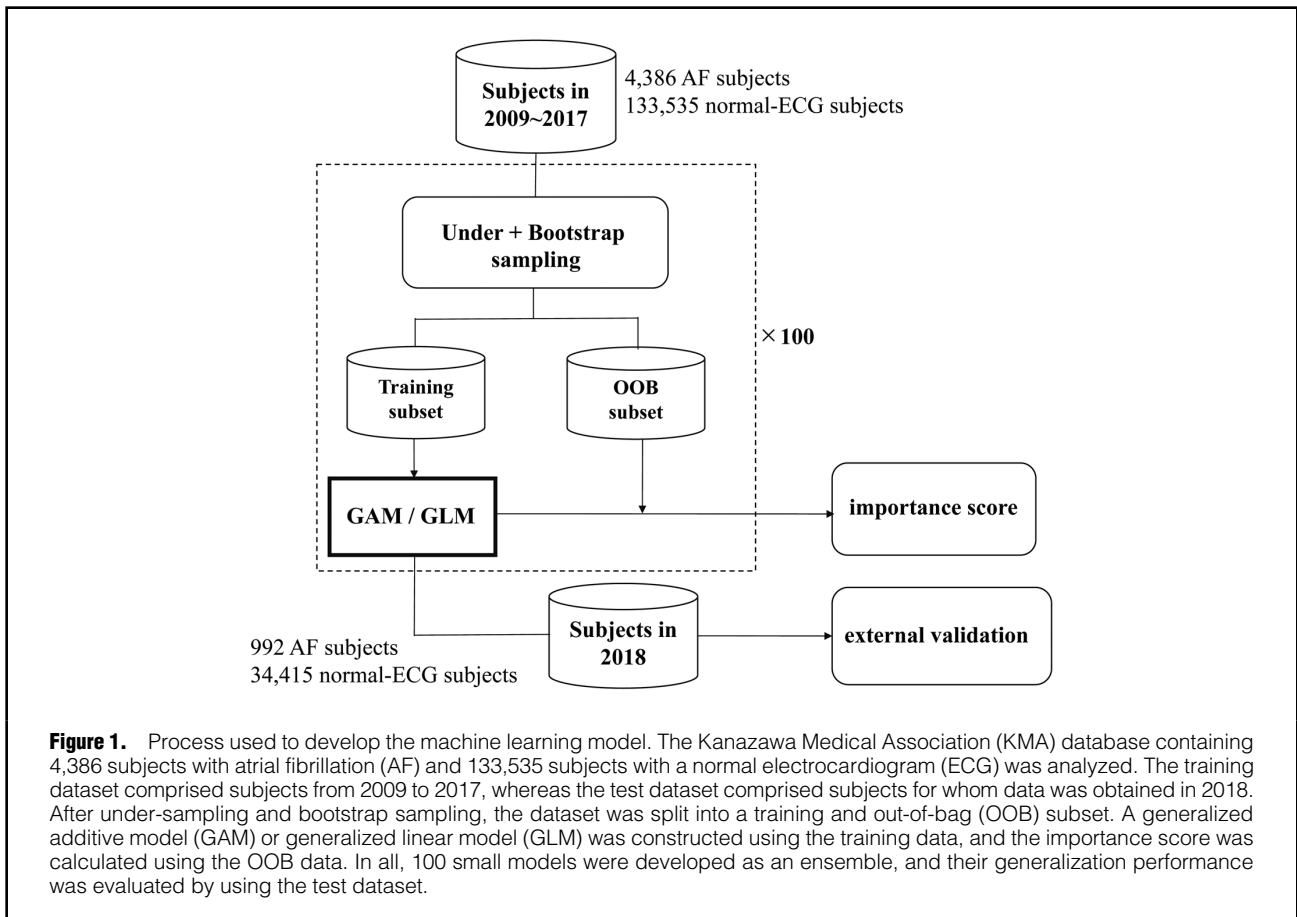
School of Electrical Information Communication Engineering, College of Science and Engineering (M. Kawakami, K.M., H.N.), Institute of Liberal Arts and Science (S.K.), Institute of Transdisciplinary Sciences (T.Y.), Kanazawa University, Kanazawa; Department of Cardiovascular Medicine (H.T., H.O., A.N., M. Kawashiri), Department of Endocrinology and Metabolism (D.A., M. Kometani, T.Y.), Departments of Hygiene (M.D.), Department of Health Promotion and Medicine of the Future (T.Y.), Graduate School of Medical Science, Kanazawa University, Kanazawa; and Health Care Center, Japan Advanced Institute of Science and Technology, Nomi (K.F.), Japan

The first two authors contributed equally to this study (M. Kawakami, S.K.).

Mailing addresses: Hidetaka Nambo, School of Electrical Information Communication Engineering, College of Science and Engineering, Kanazawa University, Kakuma-machi, Kanazawa 920-1192, Japan. E-mail: nambo@blitz.ec.t.kanazawa-u.ac.jp and Shigehiro Karashima, MD, PhD, Institute of Liberal Arts and Science, Kanazawa University, Kakuma-machi, Kanazawa 920-1192, Japan. E-mail: skarashima@staff.kanazawa-u.ac.jp

All rights are reserved to the Japanese Circulation Society. For permissions, please e-mail: cr@j-circ.or.jp  
ISSN-2434-0790





Some studies have demonstrated the high accuracy of AI models in the field of AF.<sup>8,9</sup> However, the internal workings or the basis for decisions on inference results are not known by most users of such systems. The difficulty in explaining the results of and reasoning behind AI analysis is commonly referred to as a “black box”. Consequently, most models cannot explain their predictions and evidence to doctors and patients in clinical settings. However, doctors must use models that have both high accuracy and high interpretability when they explain diagnosis or treatment plans to patients.

Several algorithms have been proposed to explain the black box model.<sup>10–12</sup> The most simple and easy to understand algorithm is the generalized linear model (GLM). The GLM extends the linear model to the exponential family of distribution and is expressed as a linear form of the relationship between independent variables and the outcome. Meanwhile, the generalized additive model (GAM) is an extension of the GLM that allows non-parametric forms of independent variables.<sup>13</sup> The GAM is trained on linear predictors, such as the sum of each variable’s non-linear functions.

Some AF prediction models have been proposed in prior studies, but few were developed with interpretable methodology. In addition, in Japan, specific health checkups started from 2008, and these are mainly conducted at clinics.<sup>14–18</sup> However, an electrocardiogram (ECG) is considered an additional rather than mandatory item on these health checkups. An interpretable algorithm for the

prediction of AF using the information obtained from such checkups would help the general population to think about lifestyle or other factors related to the development of AF. The aim of this study was to develop and validate an interpretable model for the detection of AF, using data from specific health checkups, and to examine how clinical characteristics affect the risk of AF.

## Methods

### Study Participants

The study was a cross-sectional observational study that included 489,073 subjects aged  $\geq 40$  years who underwent community-based medical checkups between 2009 and 2018 in Kanazawa City. All clinicians were sent a manual that followed the guidelines of each academic society, and the checkups were performed accordingly. During the checkups, clinicians conduct standard medical examinations, recording data such as height, weight, waist circumference, blood pressure, results of biochemical examinations, urinalysis, and resting 12-lead ECG. AF was diagnosed on the basis of the baseline ECG, according to Minnesota Codes.<sup>19</sup> Specifically, subjects diagnosed with Minnesota Code 8-3 were classified as having AF, and those diagnosed with Minnesota Code 1-0-0 were classified as having a normal ECG. All ECG findings were recorded and stored by the Kanazawa Medical Association (KMA). Subjects with partially missing data and Minnesota Codes other than 8-3 or 1-0-0 (**Supplementary Figure 1**) were excluded from the

study.

The KMA collected and anonymized the data. This study was approved by the Ethics Committee of KMA (No. 16000003) and Kanazawa University (No. 2019-080), and was conducted in accordance with the Declaration of Helsinki and the ethical guidelines for human medical research. The data were exempt from informed consent because they were secondary data. Hence, an opt-out-style announcement regarding this study was issued on the KMA website ([http://www.kma.jp/kenkyu/kenkyu\\_index.html](http://www.kma.jp/kenkyu/kenkyu_index.html)).

### Dataset and Features

The KMA database contains information on several clinical parameters, including physical observations and medical histories. Forty-seven variables, including age, sex, body mass index (BMI), waist circumference, systolic blood pressure (SBP), and diastolic blood pressure (DBP), were collected in the dataset. Moreover, the following were

measured in blood samples within 24 h of collection using an automated clinical chemical analyzer: serum creatinine (s-Cr), estimated glomerular filtration rate (eGFR), serum uric acid, HbA1c, total cholesterol, triglycerides, low-density lipoprotein cholesterol, high-density lipoprotein cholesterol (HDL-C), red blood cell (RBC) count, white blood cell (WBC) count, hemoglobin, hematocrit, platelet count, aspartate aminotransferase, alanine transaminase (ALT),  $\gamma$ -glutamyl transpeptidase ( $\gamma$ -GTP), and plasma glucose. Although the methods used for blood analyses were not calibrated between laboratories, the analyses were performed according to the methods for laboratory tests recommended by the Japan Society of Clinical Chemistry, which have been widely adopted by laboratories across Japan. Urinary protein, glucose, and occult blood were semiquantitatively examined and graded using 5 levels: “negative”, “trace”, “1+”, “2+”, and “3+”. In this study, urinary protein was analyzed as 3 levels (“negative”, “trace”,

Table 1. Baseline Characteristics of Subjects in the Atrial Fibrillation (AF) and Normal Electrocardiogram (ECG) Groups in the Training and Test Datasets				
	Training (n=137,921)		Test (n=35,407)	
	AF (n=4,386)	Normal ECG (n=133,535)	AF (n=992)	Normal ECG (n=34,415)
<b>Age (years)</b>	80±8**	70±12	79±8**	71±11
<b>Male sex (%)</b>	59.9	31.2	63.0	32.1
<b>BMI (kg/m<sup>2</sup>)</b>	23.1±3.6**	22.6±3.4	23.6±3.6**	22.8±3.4
<b>WC (cm)</b>	84.7±9.8**	82.8±9.7	86.1±10.3**	83.4±9.6
<b>SBP (mmHg)</b>	126±16**	127±16	125±16**	127±16
<b>DBP (mmHg)</b>	72±11**	73±10	73±11	73±10
<b>Laboratory tests</b>				
WBC (/μL)	5,506±2,027**	5,384±1,727	5,290±1,429	5,264±1,449
RBC (×10 <sup>4</sup> /μL)	430±57	431±46	437±52	439±44
Hemoglobin (g/dL)	13.3±1.9**	13.1±1.5	13.5±1.7**	13.4±1.4
Hematocrit (%)	40.3±5.1**	39.7±4.0	40.9±4.7**	40.5±3.8
PLT (×10 <sup>4</sup> /μL)	18.9±5.5**	22.5±5.9	19.8±5.4**	23.3±5.7
Cr (mg/dL)	1.0±0.5**	0.7±0.3	1.0±0.4**	0.8±0.2
eGFR (mL/min/1.73 m <sup>2</sup> )	58.3±17.6**	71.5±17.2	56.3±15.7**	67.5±15.3
UA (mg/dL)	5.8±1.5**	4.9±1.3	5.7±1.5**	5.0±1.3
AST (U/L)	26.3±12.4**	23.8±12.7	25.5±10.1**	23.7±10.8
ALT (U/L)	20.5±12.5*	20.4±14.5	20.1±11.0	20.7±13.0
$\gamma$ -GTP (U/L)	52.7±82.5**	32.6±50.8	45.9±81.6**	31.9±46.1
HbA1c (NGSP; %)	5.7±0.7**	5.5±0.6	6.0±0.7**	5.8±0.6
PG (mg/dL)	112±40**	102±30	112±37**	101±27
TC (mg/dL)	180±35**	201±34	180±32**	202±34
LDL-C (mg/dL)	104±29**	118±30	104±28**	117±29
HDL-C (mg/dL)	54±16**	60±15	56±14**	62±16
TG (mg/dL)	114±73**	119±74	107±60**	119±76
<b>Urine tests</b>				
Protein (%)				
(-)/(±)	62.6/16.3	80.5/11.9	65.4/16.9	81.8/11.3
(+)/(2+)/(3+)	12.6/6.5/0.5	5.1/1.9/0.1	11.2/3.8/1.6	4.7/1.4/0.4
Glucose (%)				
(-)/(±)	86.7/4.1	93.3/1.9	87.5/2.9	93.6/1.6
(+)/(2+)/(3+)	3.2/4.3/0.4	1.8/2.1/0.3	3.7/2.0/2.8	1.7/1.1/1.6
Occult blood (%)				
(-)/(±)	60.0/18.9	66.4/16.7	57.7/20.4	65.0/17.1
(+)/(2+)/(3+)	11.6/7.7/0.4	9.5/6.4/0.4	12.1/6.0/2.6	10.1/5.6/1.8

(Table 1 continued the next page.)

	Training (n=137,921)		Test (n=35,407)	
	AF (n=4,386)	Normal ECG (n=133,535)	AF (n=992)	Normal ECG (n=34,415)
<b>Past</b>				
Hypertension (%)	59.3	38.8	63.6	41.4
Diabetes (%)	16.9	10.7	18.4	11.2
Dyslipidemia (%)	19.5	24.6	26.9	29.9
Stroke (%)	23.4	6.6	19.8	5.5
CAD (%)	63.0	8.7	63.9	7.6
CKD (%)	2.9	0.8	3.1	1.1
Anemia (%)	15.9	17.6	15.2	15.6
<b>Exam</b>				
Anemia (%)	1.1	0.6	1.2	0.3
Jaundice (%)	0.1	0.0	0.0	0.0
Arrhythmia (%)	57.3	0.3	57.6	0.3
Heart murmur (%)	6.2	1.0	3.5	0.8
Crackles (%)	0.8	0.5	1.2	0.3
Hepatomegaly (%)	0.3	0.1	0.3	0.1
Edema (%)	9.4	2.3	8.2	2.0
Cervical tumors (%)	0.4	1.0	1.0	1.0
Neuropathy (%)	2.9	1.4	1.7	0.8
Malnutrition (%)	0.6	0.3	0.5	0.1
Other (%)	8.7	4.7	5.8	3.0

Data are given as the mean  $\pm$  SD or as percentages. \* $P < 0.05$ , \*\* $P < 0.001$  compared with the normal ECG group. ALT, alanine aminotransferase; AST, aspartate aminotransferase; BMI, body mass index; CAD, coronary artery disease; CKD, chronic kidney disease; Cr, creatinine; DBP, diastolic blood pressure; eGFR, estimated glomerular filtration rate; Exam, physical examination;  $\gamma$ -GTP,  $\gamma$ -glutamyl transpeptidase; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; NGSP, National Glycohemoglobin Standardization Program; Past, past history; PG, plasma glucose; PLT, platelet count; RBC, red blood cell count; SBP, systolic blood pressure; TC, total cholesterol; TG, triglycerides; UA, uric acid; WBC, white blood cell count; WC, waist circumference.

and “others”), whereas urinary glucose and urine occult blood were analyzed as 2 levels (“negative” and “others”). A medical history of stroke, coronary artery disease (CAD), chronic kidney disease (CKD), and anemia was analyzed as past history based on self-reported questionnaire. A history of hypertension, diabetes, and dyslipidemia “under treatment” or “untreated” were assigned a value of 1 and others were assigned a value of 0. Physical findings such as jaundice, arrhythmia, heart murmur, crackles, hepatomegaly, edema, cervical tumors, neuropathy, malnutrition, and anemia were also included in the model.

### Statistical Analysis for Clinical Background

Data are expressed as the mean  $\pm$  SD or as percentages. Parameters were compared between the AF and normal ECG groups. Normality was evaluated using the Shapiro-Wilk test. Data that were not normally distributed were compared using the Mann-Whitney U test. The variance of formally distributed data was evaluated using the Bartlett test; normally distributed data with equal variance were compared using Student's t-test, whereas data without equal variance were compared using Welch's t-test. Two-sided  $P < 0.05$  was considered significant. Statistical analyses were performed using Python 3.8.3 programming language (Python Software Foundation, Wilmington, DE, USA) and SciPy 1.5.2.

### General Process of Model Construction and Evaluation

This study adopted the ideas of the GAMensPlus algorithm combining the GAM and ensemble learning.<sup>20</sup> Ensemble learning is a method of estimating the final prediction result by building several small models and averaging the

class probability output. We used the GAM or the GLM as the small model to compare them in terms of the difference in non-linear and linear coefficients (**Figure 1**).

The database was divided into 2 different datasets: a training set, comprising subject data collected from 2009 to 2017, and a validation set, comprising data collected in 2018. Data for the 2018 study participants were removed from the training dataset to avoid using information from the same subjects twice (**Supplementary Figure 1**).

The model was constructed by repeating the following 4 steps for the number of ensemble members: (1) under-sampling and bootstrap sampling; (2) training the small model; (3) calculating the generalized importance score with an out-of-bag (OOB) dataset; and (4) external validation with the validation dataset. First, a sub-dataset was constructed by using the under-sampling method to address class imbalance. After under-sampling, the training and OOB datasets were separated by repeat sampling with replacement named bootstrap sampling. Second, a small prediction model was constructed with the training dataset. Third, permutation importance scores were obtained with the OOB dataset to evaluate how much each feature contributed to the predictions of the model. Finally, the overall prediction and generalized importance scores were obtained by averaging the AF probability and the permutation importance score for every GAM, respectively. The generalization performance of the ensemble model can be evaluated in terms of the area under the receiver operating characteristic curve (AUC), sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). Detailed methods are provided in the **Supplementary Material**.

**Table 2. Feature Importance Ranking and Scores Calculated on the Basis of Area Under the Curve (AUC), Sensitivity, and Specificity**

Ranking	AUC	Score	Sensitivity	Score	Specificity	Score
1	Exam arrhythmia	0.429	Exam arrhythmia	0.149	Exam arrhythmia	0.454
2	Past CAD	0.177	Past CAD	0.121	Past CAD	0.144
3	Age	0.073	Age	0.088	Hematocrit	0.051
4	Hematocrit	0.054	Hematocrit	0.041	Age	0.034
5	$\gamma$ -GTP	0.030	$\gamma$ -GTP	0.035	Hemoglobin	0.029
6	Cr	0.029	Cr	0.032	ALT	0.024
7	Hemoglobin	0.028	Hemoglobin	0.024	SBP	0.023
8	SBP	0.021	HbA1c	0.023	Cr	0.022
9	ALT	0.018	UA	0.020	$\gamma$ -GTP	0.021
10	HbA1c	0.012	TC	0.019	Exam no symptoms	0.016
11	TG	0.011	SBP	0.018	TG	0.014
12	TC	0.010	Past stroke	0.018	DBP	0.011
13	UA	0.010	AST	0.016	eGFR	0.010
14	AST	0.009	PLT	0.016	TC	0.010
15	DBP	0.008	HDL-C	0.016	AST	0.009
16	eGFR	0.008	eGFR	0.015	UA	0.008
17	Past stroke	0.007	UP (1+, 2+, 3+)	0.015	Sex	0.008
18	RBC	0.006	ALT	0.015	WC	0.008
19	PLT	0.006	RBC	0.014	LDL-C	0.008
20	UP (1+, 2+, 3+)	0.006	BMI	0.014	HbA1c	0.007
21	Exam no symptoms	0.005	Past hypertension	0.014	RBC	0.006
22	HDL-C	0.005	DBP	0.013	PLT	0.006
23	BMI	0.004	UG	0.013	UP (1+, 2+, 3+)	0.006
24	UP(-)	0.004	UP( $\pm$ )	0.013	PG	0.006
25	LDL-C	0.004	TG	0.013	UP(-)	0.006
26	Past dyslipidemia	0.004	Past dyslipidemia	0.012	Past stroke	0.005
27	WC	0.003	UP(-)	0.012	Past diabetes	0.005
28	UP( $\pm$ )	0.003	UOB	0.012	Past dyslipidemia	0.005
29	Sex	0.003	LDL-C	0.011	UOB	0.004
30	PG	0.002	WBC	0.011	UP( $\pm$ )	0.004
31	UOB	0.002	Past anemia	0.011	BMI	0.004
32	Past diabetes	0.002	Exam anemia	0.011	HDL-C	0.004
33	UG	0.001	Exam cervical	0.011	Exam edema	0.003
34	Past hypertension	0.001	Exam crackles	0.011	Exam heart murmur	0.002
35	Exam edema	0.001	Exam malnutrition	0.011	WBC	0.002
36	Exam heart murmur	<0.001	Exam jaundice	0.011	Exam others	0.002
37	Exam anemia	<0.001	Exam neuropathy	0.011	Exam crackles	0.002
38	Exam crackles	<0.001	Exam hepatomegaly	0.011	Exam anemia	0.002
39	WBC	<0.001	Exam heart murmur	0.011	Exam jaundice	0.002
40	Past anemia	<0.001	Past CKD	0.010	Exam neuropathy	0.002
41	Exam jaundice	<0.001	Past diabetes	0.010	Exam malnutrition	0.002
42	Exam neuropathy	<0.001	Exam edema	0.010	Past anemia	0.002
43	Exam others	<0.001	Exam others	0.010	Exam cervical tumors	0.002
44	Exam malnutrition	<0.001	Sex	0.009	Past CKD	0.002
45	Exam cervical tumors	<0.001	WC	0.009	Exam hepatomegaly	0.002
46	Past CKD	<0.001	PG	0.009	UG	0.002
47	Exam hepatomegaly	<0.001	Exam no symptoms	<0.001	Past hypertension	<0.001

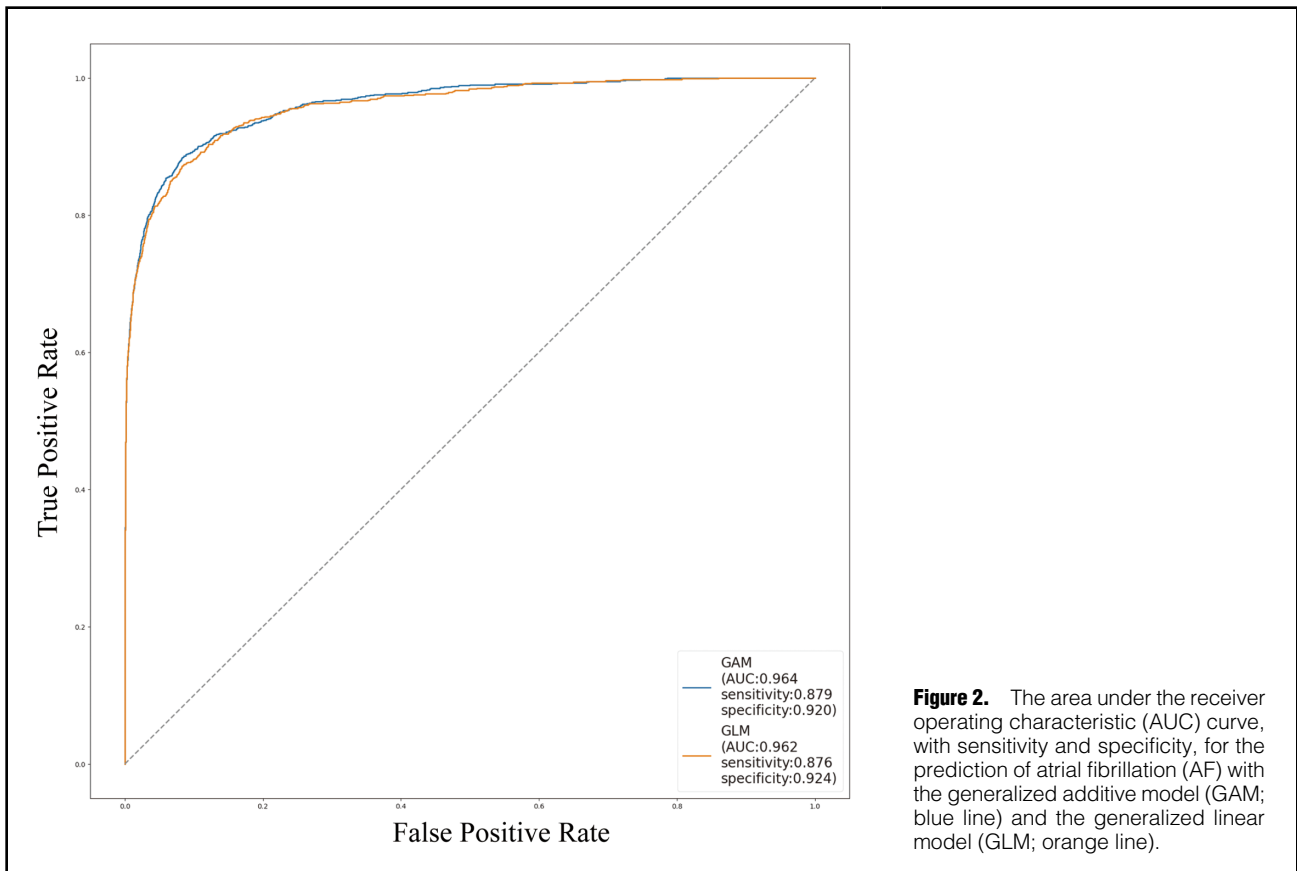
UG, urinary glucose; UOB, urine occult blood; UP, urinary protein. Other abbreviations as in Table 1.

## Results

### Subject Characteristics

Residents who underwent health checkups (n=489,073) were registered in the KMA database (**Supplementary Figure 1**). After applying the inclusion and exclusion criteria, the study subjects were classified into AF (n=5,378) and

normal ECG (n=167,950) groups. **Table 1** shows the baseline characteristics of subjects with AF and normal ECGs, divided into the training and validation subsets. Significant differences were observed between the AF and normal ECG groups for all variables, except for the RBC count in the training subset and DBP, WBC count, RBC count, and ALT in the validation subset.



**Figure 2.** The area under the receiver operating characteristic (AUC) curve, with sensitivity and specificity, for the prediction of atrial fibrillation (AF) with the generalized additive model (GAM; blue line) and the generalized linear model (GLM; orange line).

### Feature Importance Ranking

**Table 2** shows the generalized feature importance ranking and scores calculated by the GAM based on AUC, sensitivity, and specificity. Physical examination of arrhythmia, a history of CAD, age, and hematocrit were the top 4 items in all evaluation metrics, with blood  $\gamma$ -GTP, s-Cr, and hemoglobin concentrations ranked among the top 10 items.

### Predictive Performance

In this study we developed an explainable prediction model that set the number of ensemble members at 100. **Figure 2** shows the receiver operating characteristic curves for the final prediction in the validation dataset. The GAM (AUC, 0.964; sensitivity, 0.879; specificity, 0.920; PPV, 0.242; NPV, 0.996) was as accurate as the GLM (AUC, 0.962; sensitivity, 0.876; specificity, 0.924; PPV, 0.249; NPV, 0.996).

### Interpretation of Features

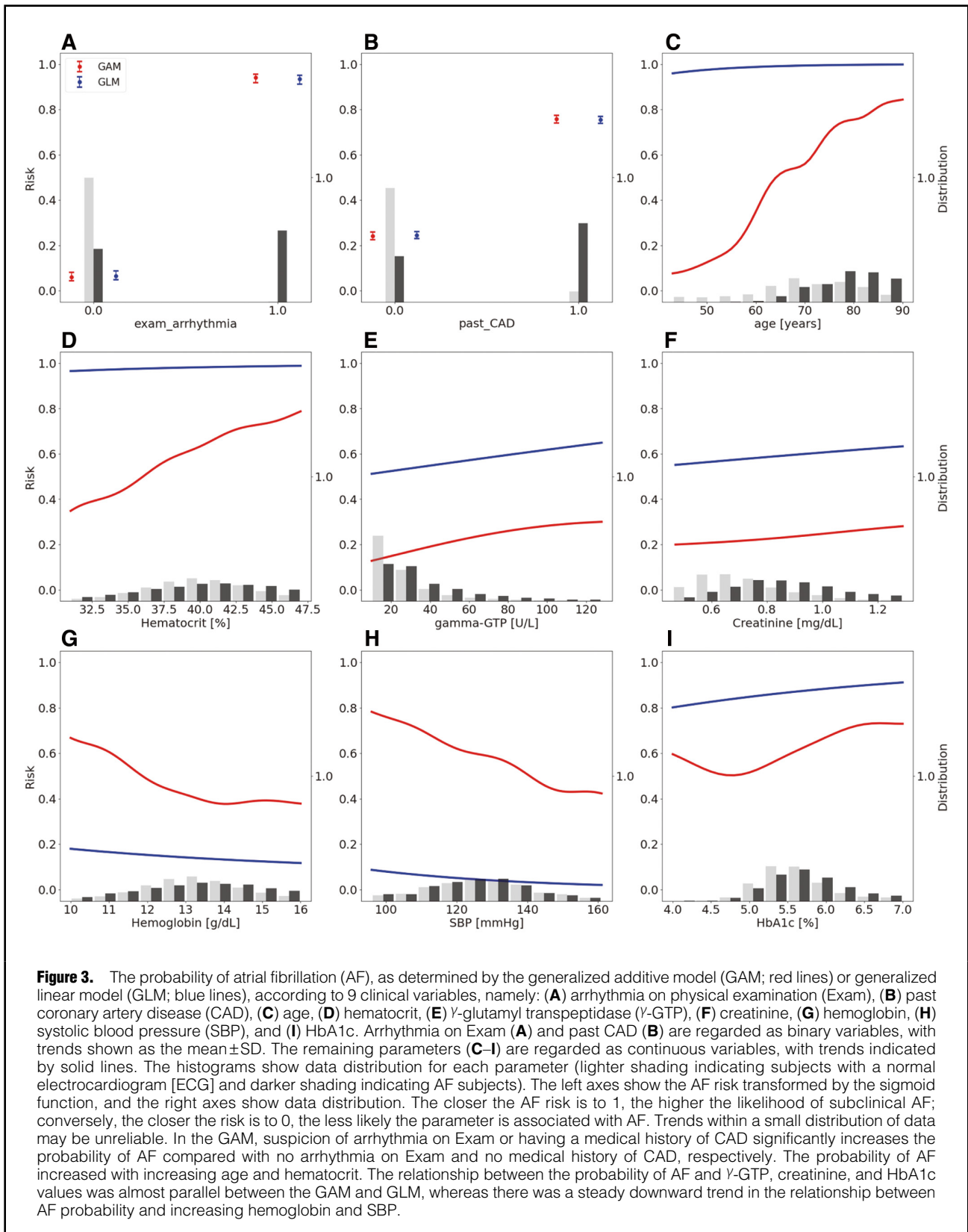
**Figure 3** shows the relationship between 9 variables selected on the basis of the generalized importance score and the risk of having AF. **Figure 3** shows average trends for the GAM and GLM, data distribution, and mean trends. Physical examination of arrhythmia, history of CAD, age, hematocrit,  $\gamma$ -GTP, s-Cr, hemoglobin, SBP, and HbA1c were selected as the 9 features with a high contribution to AUC, sensitivity, and specificity (**Table 2**). The relationship between the remaining 38 variables and the risk of AF is shown in **Supplementary Figure 2**. Both a physical finding of arrhythmia and a past history of CAD exhibited strong relationships with AF incidence compared with no arrhythmia on physical examination and no medical history

of CAD, respectively. In addition, the trends for age and hematocrit revealed an overall positive relationship between the variable and AF incidence; in particular, the magnitude of the increase in risk with increasing age was high. The relationship between HbA1c and AF showed a J-shaped waveform, with a minimum around 5%, with HbA1c values >5% positively associated with the development of AF. In contrast, the relationship between AF and both hemoglobin and SBP exhibited a negative trend. Finally, the relationship between  $\gamma$ -GTP or s-Cr and the probability of AF was almost parallel between the GAM and GLM.

### Discussion

We have developed and validated an explainable machine learning algorithm to predict AF, using clinical parameters obtained during health checkups. Using our model, we were able to detect incidents of AF with both high interpretability and high accuracy based on the easily obtained clinical findings. Based on AUC, sensitivity, and specificity, the 9 most essential elements were found to be a physical examination of arrhythmia, a medical history of CAD, age, hematocrit,  $\gamma$ -GTP, s-Cr, hemoglobin, SBP, and HbA1c. Notably, we revealed non-linear relationships between these clinical parameters and the probability of AF.

The developed model has high detection ability, with correct diagnoses in 87.9% of AF subjects and 92.0% of subjects with normal ECGs. Because the NPV was 99.6%, ECGs may no longer need to be performed in cases that the model predicts to be “normal”. In addition, compared with complex methods, such as deep learning models or



random forest classifiers, our model has the advantage of explaining the non-linear association between the occurrence of AF and clinical parameters. Therefore, the highly accurate predictions made by our model may be explained in detail, using clinical parameters, to the general population who participated in the medical examinations. Several AF risk score prediction models have been reported previously.<sup>21–26</sup> Those studies identified the following clinical variables as risk factors: age,<sup>21–24,26</sup> BMI (height or weight),<sup>21–24,26</sup> blood pressure (SBP, DBP, or the use of antihypertensive medication),<sup>21–24,26</sup> lifestyle habits (smoking, drinking status),<sup>22–24,26</sup> lipids (non-HDL-C),<sup>26</sup> on medication for the treatment of diabetes,<sup>22,24</sup> eGFR,<sup>25</sup> hemoglobin,<sup>25</sup> heart disease (myocardial infarction and heart failure), CAD, cardiac murmur,<sup>21,22,24,26</sup> and ECG findings (PR interval, left atrial enlargement, and left ventricular hypertrophy).<sup>21,22,24</sup> Our model also extracted several variables (i.e., age, SBP,  $\gamma$ -GTP [liver function], HbA1c, a past history of CAD, kidney function [creatinine], and anemia [hematocrit, hemoglobin]) that are similar to the factors derived from previous studies. These previous studies used Cox proportional hazard regression models, in which the significance of the influence different factors have is determined on the basis of the  $\beta$  coefficients, hazard ratios, and confidence intervals. Expressing the results in terms of coefficients is easier to understand as a linear function, but how the impact will change with changes in the clinical parameters remains unclear. However, in our model, the contribution of each of the 9 important features could be better understood by considering each feature as a non-linear function.

Arrhythmia on physical examination, a medical history of CAD, and age were recognized as significant risk factors, as reported previously. Lower hemoglobin levels and anemia were previously reported to contribute to the prediction of AF, but hematocrit also seems to be important. Although hematocrit is an index of the same RBC properties, a high hematocrit may be caused by relative polycythemia due to dehydration, which can also be assessed by s-Cr. The present study has shown that the risk of AF increased with increasing s-Cr. Both hematocrit and s-Cr may be explained as predictors of AF via dehydration. In our database, we did not include whether the subjects drank alcohol as a variable. However,  $\gamma$ -GTP could be used as a surrogate of alcohol consumption, which was detected as a risk factor for AF in the present study.

HbA1c ranging from 5% to 7% appeared to increase the risk of AF, and the effect of diabetes as a risk factor of AF was significant. Sandhu et al examined the differential associations between risk factors and the development of paroxysmal vs. non-paroxysmal AF.<sup>27</sup> In that study, HbA1c quartiles ( $\leq 4.84\%$ , 4.84 to  $\leq 5.00\%$ , 5.00 to  $\leq 5.19\%$ ,  $> 5.19\%$ ) and the occurrence of AF were analyzed, with higher HbA1c levels found to be preferentially associated with the early development of non-paroxysmal AF.<sup>27</sup> Iguchi et al also investigated whether the prevalence of AF was associated with HbA1c levels in Japanese adults.<sup>28</sup> In that study, the presence of AF was associated with the HbA1c level, especially in subjects with HbA1c  $< 6.5\%$ . The results of the present study show that high HbA1c increases the risk of AF with respect to glycemic control.

Hypertension has been reported as a risk for AF in previous reports. A cross-sectional study enrolling elderly subjects suggested a negative relationship between SBP and AF prevalence.<sup>29</sup> A similar trend has been observed in

other studies that enrolled and followed older subjects who started hemodialysis.<sup>30</sup> In subjects receiving antihypertensive treatment, a J-shaped relationship between SBP and AF was observed.<sup>31,32</sup> In the present study, SBP in the range 100–120 mmHg was more likely to complicate AF than SBP above that range, and there was no J-shaped relationship between SBP and AF. We present 3 hypotheses regarding the negative correlation between SBP and AF: (1) low SBP may, indeed, increase the risk of AF via chronic coronary ischemia, myocardial proliferation, and fibrosis induced by inadequate coronary perfusion; (2) low SBP may be the result of AF or AF-related cardiac structural and functional abnormalities; and (3) subjects with higher SBP are more likely to receive better clinical care and more antihypertensive medications, which play an important role in the mutually affected relationship between hypertension and AF.

In terms of interpretability, the GAM captured more detailed trends for the clinical parameters than the GLM. For example, the slope of the relationship between age and AF risk was steep for those subjects in their 60s, but the trend changed in subjects their 70s and 80s. BMI was associated with a higher probability of AF starting from a BMI around 25 kg/m<sup>2</sup>, which was defined as “overweight” in the Suita City survey.<sup>26</sup> HbA1c also exhibited a J-shaped relationship with AF risk, with a stronger positive relationship in for HbA1c 5–6.5%. The non-linear model for each clinical parameter developed using the GAM revealed part of the “black box” and helped understand the results of the prediction algorithm.

This study has several limitations. First, subjects with Minnesota Codes other than 8-3 and 1-0-0 were excluded (i.e., some subjects were classified as neither AF [Minnesota Codes 8-3] nor normal ECG [Minnesota Codes 1-0-0]). Indeed, approximately 150,000 subjects were excluded. The algorithm constructed by excluding that group may be less accurate when examined using the actual general population. It may be desirable to re-evaluate the true accuracy of our model in specific health checkups without an ECG test in the future as part of a prospective study.

Second, we achieved high accuracy in terms of AUC, sensitivity, and specificity. However, the PPV, representing the percentage of cases diagnosed as positive that are actually positive, was assessed as 0.242 in the GAM. This indicates that approximately three-quarters of the predicted AF subjects were normal ECG subjects who were misclassified. Hill et al developed machine learning models for risk prediction of AF using routinely collected data.<sup>33</sup> In that study, the final neural network model achieved a PPV of 0.295, 0.183, and 0.115 to identify 25%, 50%, and 75% of diagnosed AF cases, respectively.<sup>33</sup> In that study, there was a trade-off between sensitivity and PPV. However, we recommend using our model as screening test. We would strongly recommend that eligible subjects who test positive using our algorithm undergo a thorough an ECG. This is because approximately 25% of subjects may already have AF and could have a stroke at any time. This explainable model is expected to make physicians aware of AF complications even in the absence of ECG findings and prevent subjects from experiencing stroke and heart failure. Our model could detect persistent or chronic AF that appeared during medical examinations, but paroxysmal AF may be difficult to detect. However, high-risk subjects who are predicted to fall in the AF category in our model could undergo secondary examinations such as ECG and Holter



ECG to confirm the presence of AF.

We constructed an explainable AI system to detect current AF or normal ECG status and visualize the impact of clinical parameters, but the model cannot predict the future. The model is only designed to predict subjects with current AF complications in the general population. A useful method in terms of prevention could be, for example, a risk scoring system to predict the onset of AF within 3 years. An explainable AF future prediction and diagnosis system would be even better if it could visualize, in an understandable way, which clinical parameters could be improved in high-risk subjects to prevent AF and by how much.

In summary, we established a novel AF risk prediction model by using data from physical observations, past histories, and general laboratory test results obtained during health checkups. Further, we expressed how clinical characteristics affect the incidence of AF with non-linear trends. Our model is expected to contribute not only to more accurate AF risk prediction, but also understanding of the effects of each parameter, potentially leading to personalized medicine.

### Acknowledgments

The authors thank all those who agreed to participate in this study, as well as all general practitioners who helped collect the data. The authors express special thanks to Mr. Yoshitaka Sakikawa (KMA) for his assistance with data management. The authors also thank Editage ([www.editage.jp](http://www.editage.jp)) for English language editing a draft of this manuscript.

### Sources of Funding

This study did not receive any specific funding.

### Disclosures

K. Furukawa has received lecture fees from Sanofi K.K., Eli Lilly Japan K.K., and Ono Pharmaceutical Co. Ltd. A. Nomura has received consulting fees from CureApp, Inc. The remaining authors have no conflicts of interest to declare.

### IRB Information

This study was approved by the Ethics Committee of KMA (No. 16000003) and Kanazawa University (No. 2019-080).

### Data Availability

The deidentified participant data will be shared on a request basis. Please contact Shigehiro Karashima ([skarashima@staff.kanazawa-u.ac.jp](mailto:skarashima@staff.kanazawa-u.ac.jp)) or the Kanazawa University IRB ([crc.irb-knz@esct.jp](mailto:crc.irb-knz@esct.jp)). No additional, related documents will be available. Anyone who wishes to can access all data related to the study for 3 years after publication. For any purpose, the data will be shared as Excel files via E-mail, but the data used in this study has been restricted by the Kanazawa University IRB.

### References

- Chugh SS, Havmoeller R, Narayanan K, Singh D, Rienstra M, Benjamin EJ, et al. Worldwide epidemiology of atrial fibrillation: A Global Burden of Disease 2010 Study. *Circulation* 2014; **129**: 837–847.
- Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, et al. Heart disease and stroke statistics – 2016 update: A report from the American Heart Association. *Circulation* 2016; **133**: e38–e360.
- Alonso A, Bengtson LG. A rising tide: The global epidemic of atrial fibrillation. *Circulation* 2014; **129**: 829–830.
- Vos T, Abajobir AA, Abate KH, Abbafati C, Abbas KM, Abd-Allah F, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016. *Lancet* 2017; **390**: 1211–1259.
- Benjamin EJ, Levy D, Vaziri SM, D'Agostino RB, Belanger AJ, Wolf PA. Independent risk factors for atrial fibrillation in a population-based cohort: The Framingham Heart Study. *JAMA* 1994; **271**: 840–844.
- 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**: 946–952.
- Wolf PA, Abbott RD, Kannel WB. Atrial fibrillation as an independent risk factor for stroke: The Framingham Study. *Stroke* 1991; **22**: 983–988.
- Hu WS, Hsieh MS, Lin CL. A novel atrial fibrillation prediction model for Chinese subjects: A nationwide cohort investigation of 682237 study participants with random forest model. *EuroPace* 2019; **21**: 1307–1312.
- Attia ZI, Noseworthy PA, Lopez-Jimenez F, Asirvatham SJ, Deshmukh AJ, Gersh BJ, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: A retrospective analysis of outcome prediction. *Lancet* 2019; **394**: 861–867.
- Caruana R, Lou Y, Gehrke J, Koch P, Sturm M, Elhadad N. Intelligent models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In: *KDD '15. Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2015; 1721–1730. <https://dl.acm.org/doi/10.1145/2783258.2788613> (accessed December 15, 2021).
- Ribeiro MT, Singh S, Guestrin C. Why should I trust you? Explaining the predictions of any classifier. In: *KDD '16. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016; 1135–1144. <https://dl.acm.org/doi/10.1145/2939672.2939778> (accessed December 15, 2021).
- Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: *NIPS '17. Proceedings of the 31st International Conference on Neural Information Processing Systems*. 2017; 4765–4774. <https://dl.acm.org/doi/10.5555/3295222.3295230> (accessed December 15, 2021).
- Hastie T, Tibshirani R. Generalized additive models. *Stat Sci* 1986; **1**: 297–310.
- Tada H, Kawashiri MA, Nohara A, Inazu A, Kobayashi J, Yasuda K, et al. Lipid management in a Japanese community: Attainment rate of target set by the Japan Atherosclerosis Society Guidelines for the Prevention of Atherosclerotic Cardiovascular Diseases 2012. *J Atheroscler Thromb* 2017; **24**: 338–345.
- Tada H, Kawashiri MA, Sakata K, Yoneda T, Yasuda K, Yamagishi M, et al. Renal glucosuria is not associated with atherosclerotic cardiovascular disease outcome in a general Japanese community. *Atherosclerosis* 2017; **261**: 111–116.
- Tada H, Kawashiri MA, Yasuda K, Yamagishi M. Associations between questionnaires on lifestyle and atherosclerotic cardiovascular disease in a Japanese general population: A cross-sectional study. *PLoS One* 2018; **13**: e0208135.
- Tada H, Shibayama J, Nishikawa T, Okada H, Nomura A, Usui S, et al. Prevalence, self-awareness, and LDL cholesterol levels among patients highly suspected as familial hypercholesterolemia in a Japanese community. *Pract Lab Med* 2020; **22**: e00181.
- Tanaka Y, Tada H, Hara S, Hayashi K, Patel RB, Nishikawa T, et al. Association of proteinuria with incident atrial fibrillation in the general Japanese population. *J Cardiol* 2021; **77**: 100–105.
- Prineas R, Crow R, Blackburn H. The Minnesota code manual of electrocardiographic findings. Littleton, MA: John Wright-PSG, 1982.
- De Bock KW, Van den Poel D. Reconciling performance and interpretability in customer churn prediction using ensemble learning based on generalized additive models. *Expert Sys Appl* 2012; **39**: 6816–6826.
- Schnabel RB, Sullivan LM, Levy D, Pencina MJ, Massaro JM, D'Agostino RB, et al. Development of a risk score for atrial fibrillation (Framingham Heart Study): A community-based cohort study. *Lancet* 2009; **373**: 739–745.
- Chamberlain AM, Agarwal SK, Folsom AR, Soliman EZ, Chambless LE, Crow R, et al. A clinical risk score for atrial fibrillation in a biracial prospective cohort (from the Atherosclerosis Risk In Communities [ARIC] Study). *Am J Cardiol* 2011; **107**: 85–91.
- Everett BM, Cook NR, Conen D, Chasman DI, Ridker PM, Albert CM. Novel genetic markers improve measures of atrial fibrillation risk prediction. *Eur Heart J* 2013; **34**: 2243–2251.

24. Alonso A, Krijthe BP, Aspelund T, Stepas KA, Pencina MJ, Moser CB, et al. Simple risk model predicts incidence of atrial fibrillation in a racially and geographically diverse population: The CHARGE-AF consortium. *J Am Heart Assoc* 2013; **2**: e000102.
25. Xu D, Murakoshi N, Sairenchi T, Irie F, Igarashi M, Nogami A, et al. Anemia and reduced kidney function as risk factors for new onset of atrial fibrillation (from the Ibaraki Prefectural Health Study). *Am J Cardiol* 2015; **115**: 328–333.
26. Kokubo Y, Watanabe M, Higashiyama A, Nakao YM, Kusano K, Miyamoto Y. Development of a basic risk score for incident atrial fibrillation in a Japanese general population: The Suita Study. *Circ J* 2017; **81**: 1580–1588.
27. Sandhu RK, Conen D, Tedrow UB, Fitzgerald KC, Pradhan AD, Ridker PM, et al. Predisposing factors associated with development of persistent compared with paroxysmal atrial fibrillation. *J Am Heart Assoc* 2014; **3**: e000916.
28. Iguchi Y, Kimura K, Shibasaki K, Aoki J, Sakai K, Sakamoto Y, et al. HbA1c and atrial fibrillation: A cross-sectional study in Japan. *Int J Cardiol* 2012; **156**: 156–159.
29. Chen Y, Huang QF, Sheng CS, Lei L, Xu SK, Zhang W, et al. Cross-sectional association between blood pressure status and atrial fibrillation in an elderly Chinese population. *Am J Hypertens* 2019; **32**: 777–785.
30. Chang TI, Liu S, Airy M, Niu J, Turakhia MP, Flythe JE, et al. Blood pressure and incident atrial fibrillation in older patients initiating hemodialysis. *Clin J Am Soc Nephrol* 2019; **14**: 1029–1038.
31. Thomas MC, Dublin S, Kaplan RC, Glazer NL, Lumley T, Longstreth WT Jr, et al. Blood pressure control and risk of incident atrial fibrillation. *Am J Hypertens* 2008; **21**: 1111–1116.
32. Chen C, Liu L, Yu Y, Shen G, Huang J, Huang Y, et al. Association of systolic blood pressure with atrial fibrillation among treated hypertensive patients. *Ann Palliat Med* 2020; **9**: 1752–1763.
33. Hill NR, Ayoubkhani D, McEwan P, Sugrue DM, Farooqui U, Lister S, et al. Predicting atrial fibrillation in primary care using machine learning. *PLoS One* 2019; **14**: e0224582.

### Supplementary Files

Please find supplementary file(s);  
<http://dx.doi.org/10.1253/circrep.CR-21-0151>