

表 題 機械学習を用いた心房細動新規発症予測モデル：
個人リスクの定量化とその説明可能性

論 文 の 区 分 論文博士

著 者 名 鐘江 宏

所 属 医療法人社団こころとからだの元氣プラザ
健康づくり事業本部調査事業分析室

2024年8月15日申請の学位論文

紹 介 教 員 地域医療学系専攻 心血管病学専攻科
職名・氏名 教授・荻尾 七臣

目 次

第1章	はじめに	1
第2章	研究の方法	3
第3章	研究の成果	19
第4章	研究の考察	29
第5章	結論	33
第6章	謝辞	35
参考文献	36

第1章 はじめに

心房細動は脳卒中の重要なリスク要因であり[1]、一般集団における有症率は1-3%である[2, 3]。2030年に、ヨーロッパでは1400-1700万人、日本では108万人が心房細動を有すると推計されており[3, 4]、心房細動患者数は今後50年間で2.5倍に増加すると考えられている[1]。この心房細動患者の増加は人口の高齢化と循環器疾患における生存率の向上の影響が大きい[5, 6]。この傾向は世界で最も高齢化が進んでいる日本において特に顕著であると考えられる。

心房細動のリスク要因には、心不全、年齢、心臓弁膜症、高血圧、糖尿病がある[7]。疫学研究において高血圧、動脈硬化、脈動負荷、内皮機能不全が心房細動発症と関連していることが示されている[7-9]。動脈硬化については、収縮期血圧の増加や高血圧発症と強く関連しており、他の高血圧のリスク要因と独立している[10-13]。心房細動は脳卒中や全身性塞栓と関連しているが、高血圧を合併する心房細動では、高血圧と心房細動のいずれかを有する場合と比較し、脳卒中リスクがさらに高くなる[14, 15]。従って、高血圧や先行する動脈硬化を考慮したうえで、心房細動を早期に発見し、治療することは脳卒中や全身性塞栓症の発症を予防するために重要である。

個人の心房細動発症リスクを推定できることは医師にとって有用である。なぜなら、心房細動発症を予防し、または遅らせるために、個人の生活習慣への介

入や治療・処方が可能となるからである。心房細動の新規発症を予測するモデルは数多く存在するが、それらは欧米・アジア諸国で一般的な統計モデルを使用している[16-34]。近年、機械学習に代表されるデータ駆動型の手法の重要性が多くの領域で増しており[35]、機械学習を使用した心房細動の新規発症予測モデルも発表されている[36-44]。動脈硬化は高血圧発症に先行することが知られており[45]、心房細動のリスク要因でもあるが、高血圧と動脈硬化を考慮した機械学習を用いた心房細動の新規発症予測モデルはない。

本研究では、一般集団の健康診断データを利用し、高血圧と動脈硬化を考慮した心房細動の新規発症を予測する機械学習モデルを作成し、心房細動発症の個人リスクの定量化と関連する要因の評価可能性について検討した。

第2章 研究の方法

(1) 対象者

2005 年から 2015 年に一般財団法人日本健康増進財団で健康診断を受診した者のうち、4 年間連続して (Year 0~Year 3 とする) 健康診断を受診した者 13410 名を対象とした。なお、初回受診時 (Year 0) に心房細動の治療を受けておらず、標準 12 誘導心電図で心房細動所見がない者を対象とした。健康診断受診者 98467 名について本研究対象者 (13410 名) と非対象者 (85057 名) の背景に大きな差はみられなかった (参考資料 1)。本研究は自治医科大学倫理審査委員会で承認 (臨 A17-変 119) を得て実施した。

(2) 研究に用いた情報

アウトカム

アウトカムは心房細動の新規発症とし、Year 1 から Year 3 に医師の診断、標準 12 誘導心電図で心房細動所見あり (ミネソタコード 8-3-1, 8-3-3) に基づき判断した。

健康診断結果

健康診断結果には個人の治療歴、生活習慣、身体計測、臨床検査が含まれる。初回受診時 (Year 0) における健康診断結果を取得、整理した。

（３）対象集団の特性

カテゴリ変数は頻度およびパーセンテージを算出し、連続変数は平均値および標準偏差を算出した。群間の差の検定については、カテゴリ変数では Fisher 正確検定を、連続変数では Welch t 検定を用いた。

（４）心房細動新規発症に関連する要因

米国国立医学図書館の PubMed データベースを利用して、心房細動の新規発症予測モデルに関する医学文献を検索し、関連する文献から対象者数、イベント発生数、年齢、リスク要因および予測能に関するデータを抽出、整理した（表 1）。

治療歴、生活習慣、身体計測、臨床検査のうち、心房細動新規発症に関連すると考えられる 36 要因を選択し、予測モデルに含めることとした（表 2）。動脈硬化の指標としては cardio-ankle vascular index (CAVI) を使用した[46]。CAVI の測定はオシロメトリック法（フクダ電子株式会社製バセラシリーズ）を用いて、臥位にて数分間の安静後に測定した。左右の上腕と足首にカフを装着し、脈波センサーを用いて、50 mmHg の拍圧下で四肢の脈拍容積波形を同時に記録した。右側の測定の後、左側の測定を行い、被験者の負担を軽減した。CAVI の算出式[46]は $CAVI=2\rho/\Delta P \times \ln(Ps/Pd) \times PWV^2$ で、 ρ は血液密度、 ΔP は脈圧、 Ps は収縮期

血圧、Pdは拡張期血圧、PWVは心臓から足首の脈波伝搬速度(pulse wave velocity)である。

また、臨床検査 10 項目（アラニンアミノトランスフェラーゼ（ALT）、アスパラギン酸アミノトランスフェラーゼ（AST）、ガンマグルタミルトランスペプチターゼ（ γ -GTP）、尿酸、中性脂肪、高比重リポ蛋白質（HDL）コレステロール、低比重リポ蛋白質（LDL）コレステロール、推算糸球体ろ過量（eGFR）、尿蛋白、尿糖）を除外した 26 要因を用いた簡易予測モデルも作成し、臨床検査結果が不明な場合の予測可能性についても検討した。

（5）機械学習モデルの作成

機械学習モデルとして勾配ブースティング（eXtreme Gradient Boosting, XGBoost）を使用した[35]。

アウトカムおよび初回受診時の健康診断結果をモデル作成用データ 80%（n=10,728）とモデル検証用データ 20%（n=2,682）にランダムに分割した（図 1）。過学習を防ぎつつ適合度を向上するために、5 分割交差検証法（図 1）を用い、グリッドサーチおよびランダムサーチにより最適なハイパーパラメーターの組み合わせを選択した（参考資料 2）。なお、ハイパーパラメーターの評価には ROC 曲線とその曲線下面積（AUC）を使用した。モデル作成用データと選択し

た最適なハイパーパラメーターを用いて学習済み XGBoost モデルを作成した。

また、本データセットは心房細動新規発症者が非常に少ないデータ不均衡がある。このデータ不均衡は、少数群の識別が困難になり、多数群に対するモデルの偏りが生じることがあり、分類問題では大きな課題となる可能性がある。この問題に対処するために Synthetic Minority Over-sampling Technique (SMOTE) [47]を用いて、データ不均衡を改善した解析も行い、データ不均衡の影響を確認した。

(6) 機械学習モデルの評価

学習済み XGBoost モデルにモデル検証用データを適用し、ROC 曲線および AUC によりモデルの適合度を評価した。一般的に AUC が 0.7~0.8 で良好、0.8 以上で優秀と判断されている [48]。

(7) 機械学習モデルの解釈

学習済み XGBoost モデルにモデル検証用データを適用した場合、どの要因の寄与が大きいかを確認するためにゲーム理論に基づく Shapley Additive Explanation values (SHAP) を用いた [49]。

データ解析ソフトは SAS version 9.4 (SAS Institute Inc., Cary, NC) と

Python version 3.8 (Python Software Foundation, Wilmington DE) を用いた。

表 1(1) 心房細動の新規発症予測モデル（一般的な統計モデル）

著者, 発表 年, 国名	対象者数 発症者数(診断法)	年齢 (男性)	追跡 期間	説明変数の数および種類		統計モデル	訓練データ AUC	検証データ AUC
Schnabel et al. 2009[32] USA	Community-based cohort 4764 457 (ECG)	61 (45– 95)歳 (45%)	10 年	7	年齢, 性別, SBP, 高血圧, BMI, PR 間 隔, 心雑音, HF	Cox model	0.78	-
Chamberlain et al. 2011[20] USA	Community-based cohort 14546 515 (ECG)	45–64 歳 (45%)	10 年	12	年齢, 身長, 喫煙, SBP, 高血圧, 心雑 音, LVH, LAE, DM, CHD, HF	Cox model	0.78	0.77 (内部検証)
Everett et al. 2013[22] USA	Women HCPs 13061 404 (questionnaire, medical record)	53 歳 (0%)	10 年	6	年齢, 体重, 身長, SBP, 飲酒, 喫煙	Cox model	-	0.72 (内部検証)
Alonso et al. 2013[16] USA	3 community-based cohorts 18556 1186 (ICD9, ICD10, ECG)	65 歳 (57%)	5 年	13	年齢, 人種, 身長, 体重, SBP, DBP, 喫 煙, 高血圧, DM, MI, HF, LVH, PR 間隔	Cox model	0.77	0.67, 0.72 (外部検証)

Brunner et al. 2014[19] USA	Medical system database 97909 3431 (ICD9, medical record)	43 歳 (41%)	8.5 年	7	年齢, CAD, DM, 性別, HF, 高血圧, VD	Logistic model	0.81	-
Rienstra et al. 2016[30] Netherlands	2 community-based cohorts 8265 250 (ECG)	49 歳 (50%)	10 年	18	年齢, 性別, 家系, BMI, DBP, HR, 高血 圧, MI, HF, DM, 脳卒中, PAD, 喫煙, 飲酒, 脂質異常症, PR 間隔, GFR, UAE	Latent class model	0.84	0.73 (外部検証)
Alonso et al. 2016[17] USA	6 community-based cohorts 5477 296 (ICD9)	45–84 歳 (47%)	5 年	11	年齢, 人種, 身長, 体重, SBP, DBP, 喫 煙, 降圧薬, DM, NT-proBNP, hsCRP	Cox model	0.83	-
Saliba et al. 2016[31] Israel	EHR database (CHS) 1062073 23223 (ICD9)	66 歳 (45%)	2 年	5, 7	CHADS ₂ , CHA ₂ DS ₂ -VASc	Cox model	0.73 for CHADS ₂ , 0.74 for CHA ₂ DS ₂ - VASc	-
Suenari, et al. 2017[33] Taiwan	EHR database (NHIRD) 670804 9174 (ICD9)	42 歳 (51%)	9 年	5	年齢, 高血圧, 脳卒中 or TIA, COPD, HF	Cox model	0.72	-

Ding et al. 2017[21] China	Health check-up cohort 33186 134 (ECG)	57 (45– 85) 歳 (67%)	2 年	7	年齢, 性別, CHD, 心電図所見および波 形	Cox model	0.82	0.79 (内部検証)
Kokubo et al. 2017[27] Japan	Community-based cohort 6864 311 (ECG)	56 (30– 79) 歳 (47%)	10 年	10	年齢, 性別, 収縮期高血圧, BMI, 飲酒, 喫煙, non-HDL-C, 不整脈, CAD, 心雑 音	Cox model	0.75	-
Linker et al. 2018[29] USA	Community-based cohort 3790 509 (medical record)	59 歳 (47%)	9 年	14	年齢, 身長, 体重, 人種, うっ血性心不 全, CAD, MI, VD, COPD, CA, CAS, 脳 卒中, DM, 腎移植	Logistic model	0.80	0.79 (内部検証)
Aronson et al. 2018[18] Israel	EHR database (MHS) 96778 5660 (ECG)	62 歳 (46%)	10 年	10	年齢, 性別, BMI, 高血圧, SBP, CLD, MI, PAD, HF, ID	Cox model	0.74	0.75 (内部検証)
Li et al. 2019[28] China	EHR database 471446 921 (ECG, Holter)	47 歳 (53%)	11 年	6	CAD, COPD, 高血圧, 年齢, 収縮期 HF, 甲状腺機能亢進症	Cox model	0.75	0.75 (内部検 証) 0.65 (外部検 証)

Hamada et al. 2019[23] Japan	Medical check-up cohort 65984 349 (ECG, self- reported)	52 歳 (65%)	7 年	10	年齢, 性別, 腹囲, DBP, 飲酒, HR, LVH, AE, APC, VPC	Cox model	0.78	0.79 (内部検 証)
Hulme et al. 2019[25] USA	EHR database (RPDR) 206042 7216 (ICD9, ICD10, medical record)	61 歳 (42%)	5 年	16	年齢, 性別, 人種, 喫煙, 身長, 体重, DBP, 高血圧, 脂質異常症, HF, CHD, VD, 脳卒中 and TIA, PAD, CKD, 甲状 腺機能低下症	Cox model	0.77	0.78 (内部検 証)
Igarashi et al. 2021[26] Japan	Health check-up cohort 37526 135 (ECG)	52 歳 (67%)	3 年	7	年齢, 性別, 腹囲, DBP, LDL-C, GTP, ミ ネソタコード	Cox model	0.84	0.79 (内部検 証) 0.79 (外部検 証)
Hata et al. 2021[24] Japan	Population-based cohort 2442 230 (ECG)	59 歳 (43%)	10 年	8	年齢, 性別, 腹囲, SBP, eGFR, 心雑音, 心電図所見, 不整脈	Cox model	0.79	-
Yum et al. 2022[34] Korea	EHR database cohort 25584 281* (ICD9, ECG)	51 歳 (51%)	3 年	12	年齢, 性別, CKD, HF, MV 狭窄, VD, 脳 卒中, 心電図所見	Cox model	0.81	0.80 (内部+ 外部検証)

*発症率より算出した

AE, 心房肥大; APC, 心房期外収縮; AUC, 曲線下面積; BMI, body mass index; CA, 心停止; CAD, 虚血性心疾患; CAS, 冠動脈ステント; CHD, 冠状動脈

性心臓病; CKD, 慢性腎臓病; CLD, 慢性肺疾患; COPD, 慢性閉塞性肺疾患; DBP, 拡張期血圧; DM, 糖尿病; ECG, 心電図; EHR, 医療電子記録; HCPs, 医療専門家; HF, 心不全; HR, 心拍数; ID, 免疫不全症候群; LAE, 左房肥大; LVH, 左室肥大; MI, 心筋梗塞; MV, 僧帽弁; NT-proBNP, N 末端プロ脳性ナトリウム利尿ペプチド; PAD, 末梢動脈疾患; SBP, 収縮期血圧; TIA, 一過性脳虚血発作; UAE, 尿中アルブミン排泄量; VD, 弁膜症; VPC, 心室性期外収縮.

表 1(2) 心房細動の新規発症予測モデル（機械学習モデル）

著者, 発表 年, 国名	対象者数 発症者数(診断法)	年齢 (男性)	追跡 期間	説明変数の数および種類	統計モデル	訓練データ AUC	検証データ AUC
Hill et al. 2019[39] UK	Primary care data (CPRD) 2994837 95607 (Read code)	56 歳 (47%)	7.1 年	100 年齢, 性別, 人種, 喫煙, 身長, 体重, BMI, DBP, SBP, 高血圧, HF, CHD, MI, LVH, T1DM, T2DM など	Time- varying NN model	0.83	-
Hu et al. 2019[40] Taiwan	Health insurance database (LHID) 682237 14212 (ICD9)	41 歳 (51%)	12 年	19 年齢, 性別, 既往歴, CHA ₂ DS ₂ -VASc score, 追跡期間など	Random forest model	0.95	0.85(外部検 証)
Bundy et al. 2020[37] USA	Population-based cohort 3534 124 (ICD9)	61 歳 (48%)	5 年	23 年齢, 人種, 身長, 体重, SBP, DBP, 高血 圧, NT-proBNP, cardiac troponin-T など	LASSO, Cox model	0.81	-
Tiwari et al. 2020[43] USA	UCHealth Hospital system 2252219 28036 (ICD9, ICD10)	43 歳 (46%)	0.5 年	200 年齢, 性別, 診断, 手技, 服薬など	Logistic model (L2)	-	0.81(内部検 証)

Sekelj et al. 2021[42] UK	Primary care data (WSIC) 604135 17880 (Read code)	52 歳 (49%)	8 年	100	年齢, 性別, 人種, 喫煙, 身長, 体重, BMI, DBP, SBP, 高血圧, 高血圧治療など	Time- varying NN model	0.87	0.83 (内部検 証)
Ashburner et al. 2022[36] USA	Primary Care Practice-Based Research Network 32960 2839 (ICD10)	70 歳 (43%)	5 年	22	年齢, 性別, 人種, 肥満など	NLP model	0.74	0.74 (内部検 証) 0.75 (外部検 証)
Toprak et al. 2023[44] Europe	3 community-based cohorts 42280 1496 (ICD8, ICD9, ICD10)	52 歳 (48%)	5.7 年	23	年齢, 性別, BMI, SBP, 降圧薬, 喫煙, DM, HF, MI or 脳卒中, バイオマーカー など	LASSO, RSF model	-	-
He et al. 2023[38] China	Community-based cohort 13310 313 (ECG)	67 歳 (45%)	-	60	年齢, 性別, BMI, SBP, DBP, 診断, ECG など	Logistic model, SVM, RF	-	-
Nadarajah et al. 2023[41] UK	Primary care data (CPRD-GOLD) 1664911 5834 (Read code)	50 歳 (49%)	0.5 年	73	年齢, 性別, 合併症, 人種	RF model	-	0.82 (内部検 証)

AUC, 曲線下面積; BMI, body mass index; CAVI, 心臓血管足首指数; CHD, 虚血性心疾患; CKD, 慢性腎臓病; CVD, 循環器疾患; DBP, 拡張期血圧; DL, 脂質異常症; DM, 糖尿病; ECG, 心電図; EHR, 医療電子記録; HF, 心不全; LASSO, 正則化回帰; LVH, 左室肥大; MI, 心筋梗塞; NLP, 自然言語処理;

NN, ニューラルネットワーク; NT-proBNP, N 末端プロ脳性ナトリウム利尿ペプチド; RF, ランダムフォレスト; RSF, ランダムサバイバルフォレスト; SBP, 収縮期血圧; SVM, サポートベクターマシン.

表 2 心房細動発症に関連すると考えられる要因

No.	変数名	値
1	年齢	連続量、19-92 歳
2	性別	カテゴリ、0:女性 1:男性
3	Body mass index (BMI)	連続量、10-47 kg/m ²
4	収縮期血圧	連続量、84-212 mmHg
5	拡張期血圧	連続量、48-132 mmHg
6	脈圧	連続量、22-118 mmHg
7	ALT	連続量、4-532 IU/L
8	AST	連続量、7-463 IU/L
9	γ-GTP	連続量、6-1815 IU/L
10	尿酸	連続量、0.4-11.7 mg/dL
11	中性脂肪	連続量、14.0-4340.0 mg/dL
12	HDL コレステロール	連続量、23.0-185.0 mg/dL
13	LDL コレステロール	連続量、23.6-289.0 mg/dL
14	eGFR	連続量、4.2-117.9 mL/min/1.73m ²
15	尿蛋白	カテゴリ、0:陰性 1:陽性
16	尿糖	カテゴリ、0:陰性 1:陽性
17	高血圧	カテゴリ、0:なし 1:あり
18	不整脈	カテゴリ、0:なし 1:あり
19	左室肥大 1 ^{※1}	カテゴリ、0:なし 1:あり
20	左室肥大 2 ^{※2}	カテゴリ、0:なし 1:あり
21	糖尿病	カテゴリ、0:なし 1:あり

22	脂質異常症	カテゴリ、0:なし 1:あり
23	慢性腎臓病	カテゴリ、0:なし 1:あり
24	循環器疾患※ ³	カテゴリ、0:なし 1:あり
25	心疾患※ ⁴	カテゴリ、0:なし 1:あり
26	脳血管疾患※ ⁵	カテゴリ、0:なし 1:あり
27	喫煙	カテゴリ、0:なし 1:あり
28	飲酒頻度、0 日/週	カテゴリ、0:なし 1:あり
29	飲酒頻度、1-6 日/週	カテゴリ、0:なし 1:あり
30	飲酒頻度、7 日/週	カテゴリ、0:なし 1:あり
31	飲酒量、0 合/日	カテゴリ、0:なし 1:あり
32	飲酒量、1-2 合/日	カテゴリ、0:なし 1:あり
33	飲酒量、3 合/日以上	カテゴリ、0:なし 1:あり
34	Cardio-ankle vascular index (CAVI)	連続量、4.2-13.1 unit
35	CAVI 測定時の収縮期血圧	連続量、83-212 mmHg
36	CAVI 測定時の拡張期血圧	連続量、43-132 mmHg

※1 標準 12 誘導心電図の高振幅

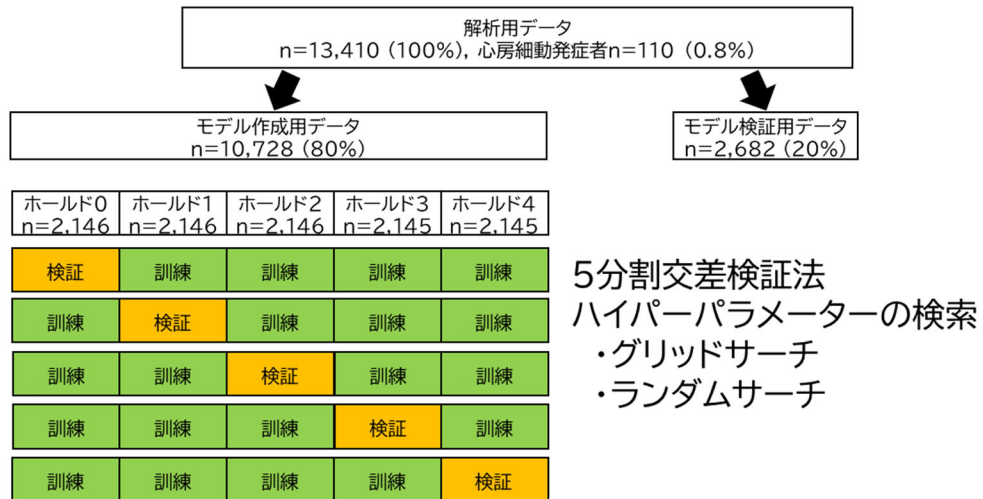
※2 標準 12 誘導心電図の高振幅+ST 異常

※3 心疾患+脳血管疾患

※4 狭心症、心筋梗塞、心臓病

※5 脳卒中、脳梗塞、脳内出血、くも膜下出血

図1 機械学習モデルの作成



第3章 研究の成果

(1) 対象者背景

対象者 13410 名のうち、3 年以内に心房細動を発症した者は 110 名 (0.82%) であった。モデル作成用データおよびモデル検証用データの対象者背景を表 3、表 4 に示す。心房細動を発症した者は高齢で、男性が多く、診療室血圧や CAVI が高く、高血圧、循環器疾患、不整脈や左室肥大が多かった。

表 3 対象者の背景 (モデル作成用データ)

No.	変数名	心房細動発症者 n=88	心房細動未発症者 n=10640	p 値
1	年齢, 歳	63.8±10.0	49.6±12.7	<0.001
2	性別, 男性	70 (79.6)	5164 (48.5)	<0.001
3	BMI, kg/m ²	24.0±2.9	22.8±3.5	0.003
4	収縮期血圧, mmHg	132.4±13.9	125.1±15.9	<0.001
5	拡張期血圧, mmHg	76.9±9.7	73.6±10.9	0.004
6	脈圧, mmHg	55.5±12.9	51.6±10.5	<0.001
7	ALT, IU/L	25.9±9.7	23.0±11.4	0.028
8	AST, IU/L	25.1±13.4	23.0±17.3	0.282
9	γ-GTP, IU/L	57.6±71.2	38.4±51.2	0.001
10	尿酸, mg/dL	5.9±1.2	5.1±1.4	<0.001
11	中性脂肪, mg/dL	120.7±57.3	107.3±81.4	0.145
12	LDL コレステロール, mg/dL	123.8±31.0	126.3±32.7	0.506
13	HDL コレステロール, mg/dL	59.6±17.2	67.8±19.2	<0.001
14	eGFR, ml/min/1.73m ²	73.2±10.2	84.3±10.8	<0.001
15	尿蛋白, 陽性	4 (4.6)	233 (2.2)	0.130
16	尿糖, 陽性	0 (0.0)	135 (1.3)	0.631
17	高血圧, あり	36 (40.9)	1273 (12.0)	<0.001
18	不整脈, あり	12 (13.6)	111 (1.0)	<0.001
19	左室肥大 ^{※1} , あり	12 (13.6)	601 (5.7)	0.004

20	左室肥大 2※ ² , あり	2 (2.3)	78 (0.7)	0.140
21	糖尿病, あり	5 (5.7)	291 (2.7)	0.096
22	脂質異常症, あり	5 (5.7)	801 (7.5)	0.684
23	慢性腎臓病, あり	1 (1.1)	195 (1.8)	1.000
24	循環器疾患※ ³ , あり	26 (29.6)	308 (2.9)	<0.001
25	心疾患※ ⁴ , あり	24 (27.3)	218 (2.1)	<0.001
26	脳血管疾患※ ⁵ , あり	2 (2.3)	95 (0.9)	0.189
27	喫煙, あり	15 (17.1)	1884 (17.7)	1.000
28	飲酒頻度, 0 日/週	37 (42.1)	5665 (53.2)	0.031
29	飲酒頻度, 1-6 日/週	30 (34.1)	3418 (32.1)	
30	飲酒頻度, 7 日/週	21 (23.9)	1557 (14.6)	
31	飲酒量, 0 合/日	37 (42.1)	5671 (53.3)	0.038
32	飲酒量, 1-2 合/日	38 (43.2)	4061 (38.2)	
33	飲酒量, 3 合/日以上	13 (14.8)	908 (8.5)	
34	CAVI, unit	8.7±0.9	7.7±1.0	<0.001
35	CAVI 測定時の収縮期血圧, mmHg	130.9±16.7	121.7±16.3	<0.001
36	CAVI 測定時の拡張期血圧, mmHg	80.2±11.1	75.8±11.2	<0.001

※1 標準 12 誘導心電図の高振幅

※2 標準 12 誘導心電図の高振幅+ST 異常

※3 心疾患+脳血管疾患

※4 狭心症、心筋梗塞、心臓病

※5 脳卒中、脳梗塞、脳内出血、くも膜下出血

表 4 対象者の背景（モデル検証用データ）

No.	変数名	心房細動発症者 n=22	心房細動未発症者 n=2660	p 値
1	年齢, 歳	61.8±8.4	49.7±12.8	<0.001
2	性別, 男性	18 (81.8)	1295 (48.7)	0.002
3	BMI, kg/m ²	23.7±4.6	22.8±3.3	0.211
4	収縮期血圧, mmHg	135.9±16.4	124.5±15.6	<0.001
5	拡張期血圧, mmHg	81.2±9.8	73.3±11.0	<0.001
6	脈圧, mmHg	54.7±11.5	51.2±10.3	0.116
7	ALT, IU/L	26.1±13.3	22.7±10.5	0.175
8	AST, IU/L	24.8±16.6	22.3±17.4	0.552
9	γ-GTP, IU/L	66.4±89.3	36.2±54.0	0.019
10	尿酸, mg/dL	5.9±1.4	5.1±1.4	0.008
11	中性脂肪, mg/dL	134.7±132.5	106.5±118.8	0.316
12	LDL コレステロール, mg/dL	124.2±25.6	126.7±32.8	0.748
13	HDL コレステロール, mg/dL	64.4±17.9	67.9±18.8	0.435
14	eGFR, ml/min/1.73m ²	76.5±10.8	84.1±10.8	0.003
15	尿蛋白, 陽性	0 (0.0)	64 (2.4)	1.000
16	尿糖, 陽性	1 (4.6)	41 (1.5)	0.294
17	高血圧, あり	5 (22.7)	310 (11.7)	0.170
18	不整脈, あり	1 (4.6)	32 (1.2)	0.239
19	左室肥大 1 ^{※1} , あり	4 (18.2)	147 (5.5)	0.032
20	左室肥大 2 ^{※2} , あり	0 (0.0)	20 (0.8)	1.000
21	糖尿病, あり	1 (4.6)	71 (2.7)	0.452
22	脂質異常症, あり	3 (13.6)	213 (8.0)	0.415
23	慢性腎臓病, あり	1 (4.6)	53 (2.0)	0.362
24	循環器疾患 ^{※3} , あり	4 (18.2)	67 (2.5)	0.002
25	心疾患 ^{※4} , あり	3 (13.6)	50 (1.9)	0.009
26	脳血管疾患 ^{※5} , あり	1 (4.6)	18 (0.7)	0.145
27	喫煙, あり	4 (18.2)	448 (16.8)	0.778
28	飲酒頻度, 0 日/週	11 (50.0)	1413 (53.1)	0.554
29	飲酒頻度, 1-6 日/週	6 (27.3)	849 (31.9)	
30	飲酒頻度, 7 日/週	5 (22.7)	398 (15.0)	
31	飲酒量, 0 合/日	11 (50.0)	1414 (53.2)	0.211

32	飲酒量, 1-2 合/日	7 (31.8)	1031 (38.8)	
33	飲酒量, 3 合/日以上	4 (18.2)	215 (8.1)	
34	CAVI, unit	8.5±1.0	7.7±1.0	<0.001
35	CAVI 測定時の収縮期血圧, mmHg	133.5±15.0	121.5±15.9	<0.001
36	CAVI 測定時の拡張期血圧, mmHg	81.7±9.0	75.7±10.9	0.010

※1 標準 12 誘導心電図の高振幅

※2 標準 12 誘導心電図の高振幅+ST 異常

※3 心疾患+脳血管疾患

※4 狭心症、心筋梗塞、心臓病

※5 脳卒中、脳梗塞、脳内出血、くも膜下出血

（２）予測モデルの評価

全 36 要因を考慮した予測モデルにおける AUC は、5 分割交差検証法の結果で 0.84~0.87、モデル作成用データ全体の結果で 0.99、モデル検証用データで 0.79 であった（図 2）。臨床検査 10 項目を除外した 26 要因を考慮した予測モデルにおける AUC は同様に 0.83~0.89、0.99、0.78 であった（図 3）。2 つの予測モデルの AUC に大きな差はみられなかった。また、SMOTE によりデータ不均衡を改善した場合でもモデルの向上はみられなかった（参考資料 3）。

モデル全体及び個々人における予測モデルの変数寄与度は図 4、図 5 に示す。全 36 要因を考慮した予測モデルにおける変数寄与度は、年齢が最も大きく、続いて CAVI、eGFR、男性、BMI、尿酸、 γ -GTP、中性脂肪、CAVI 測定時の収縮期血圧、ALT であった（図 4）。臨床検査 10 項目を除外した 26 要因を考慮した予測モデルにおける変数寄与度は、年齢が最も大きく、続いて CAVI、BMI、CAVI 測定時の収縮期血圧、CAVI 測定時の拡張期血圧、収縮期血圧、脈圧、拡張期血圧、男性、心疾患であった（図 5）。

図2 予測モデルの評価結果 (36 要因を考慮した予測モデル)

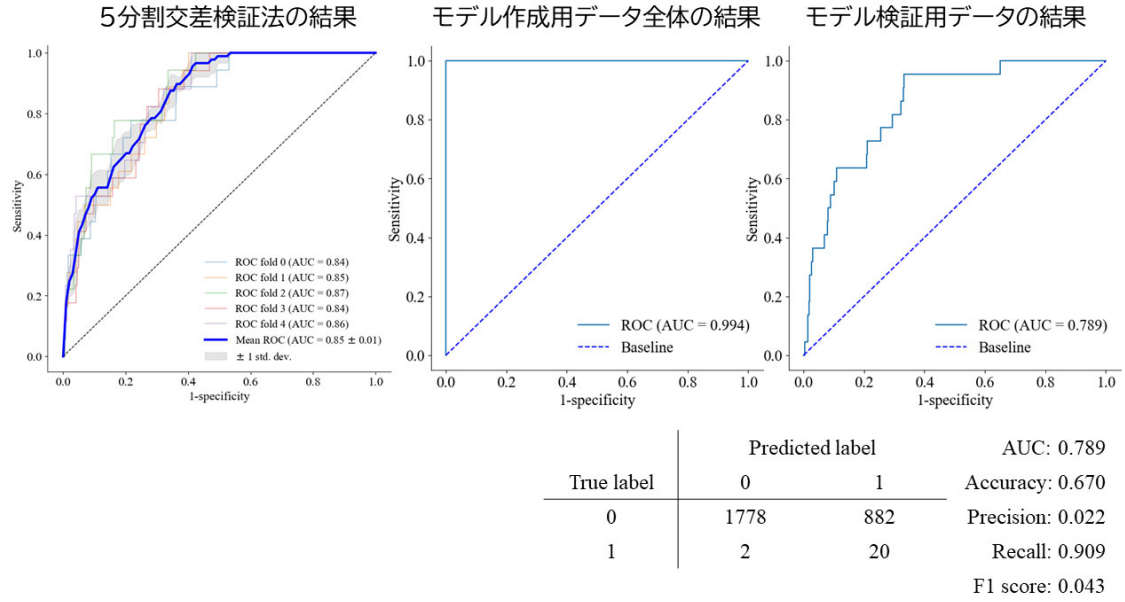


図3 予測モデルの評価結果 (26 要因を考慮した予測モデル)

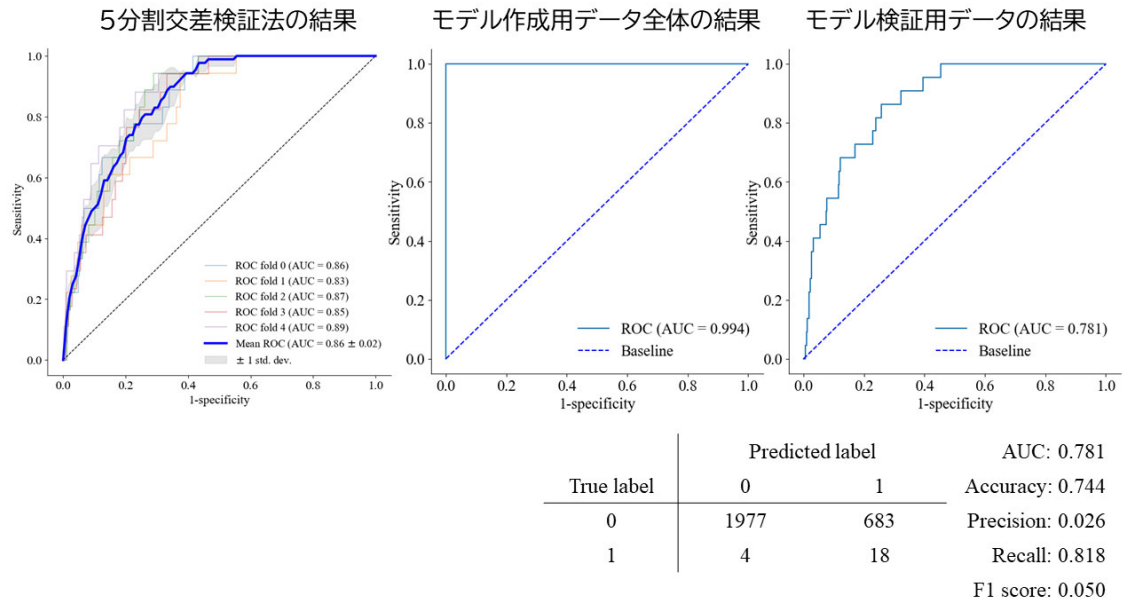


図4 変数重要度 (36 要因を考慮した予測モデル)

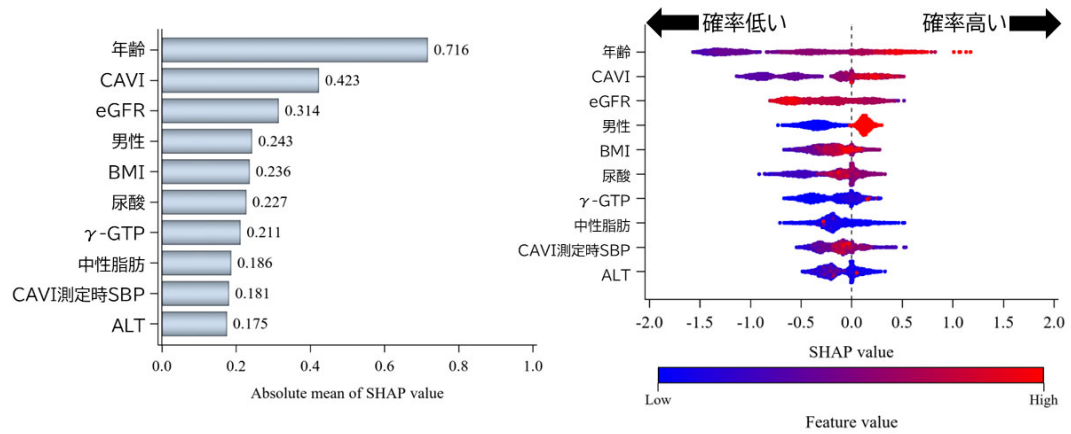
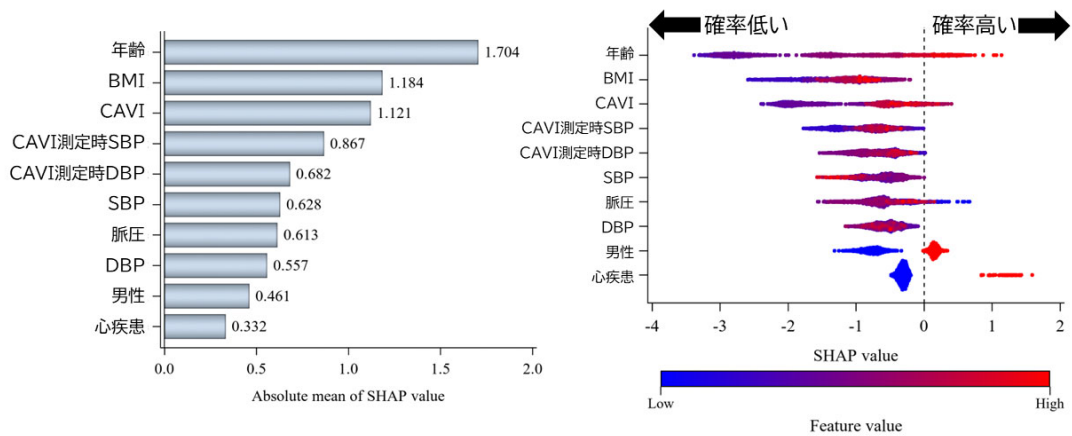


図5 変数重要度 (26 要因を考慮した予測モデル)



（３）予測モデルの解釈

症例２名における予測モデルの解釈結果は図６、図７に示す。症例１では、心疾患あり、循環器疾患あり、年齢（８０歳）の寄与が大きく、心房細動の発症確率が高くなる傾向がみられた（図６）。症例２では、年齢（７１歳）、収縮期血圧（１００ mmHg）、尿酸（６．８ mg/dL）、CAVI（１０．２）の寄与が大きく、心房細動の発症確率が高くなる傾向がみられた（図７）。

収縮期血圧に対する SHAP は図８に示す。収縮期血圧が １０６～１４６ mmHg の範囲では、収縮期血圧が増加するにしたがって SHAP も増加する傾向がみられた。収縮期血圧が １０６ mmHg 未満と １４８ mmHg 以上では対象者が非常に少なかった。収縮期血圧が １０６ mmHg 未満の者は、女性が多く、若く、比較的健康な者が多かった。一方、収縮期血圧が １４８ mmHg 以上の者は、男性が多く、高齢で、肥満など比較的不健康な者が多かった。収縮期血圧が １０６ mmHg 未満の者は、相対的に収縮期血圧に対する SHAP が大きく、その他の変数に対する SHAP は小さかった。収縮期血圧が １４８ mmHg 以上の者は、相対的に収縮期血圧に対する SHAP が小さく、その他の変数に対する SHAP は大きかった。

図6 予測モデルの解釈結果（症例1）

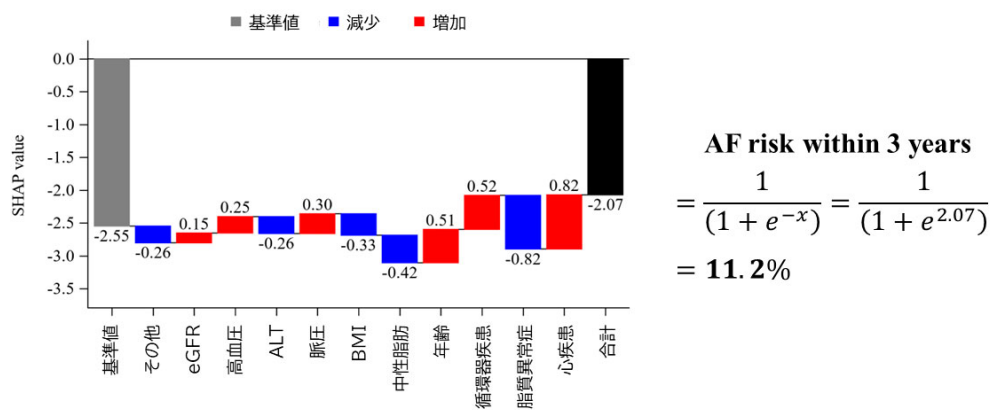


図7 予測モデルの解釈結果（症例2）

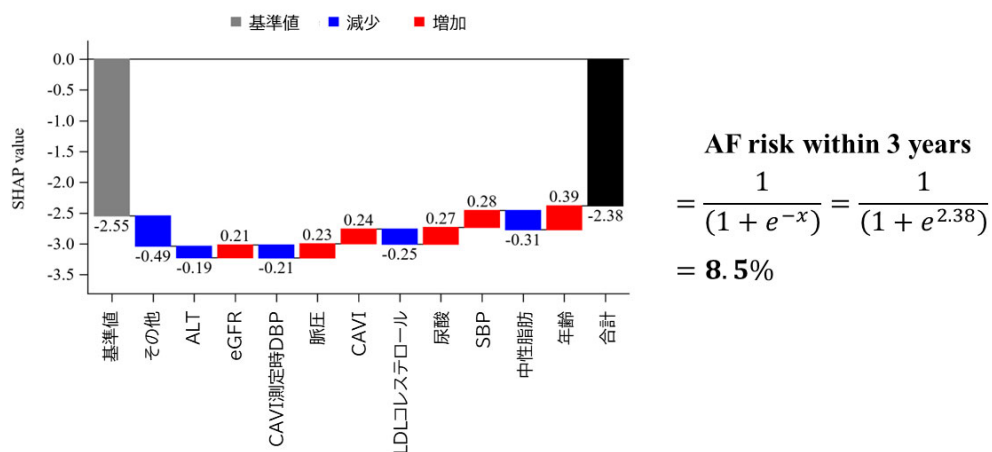
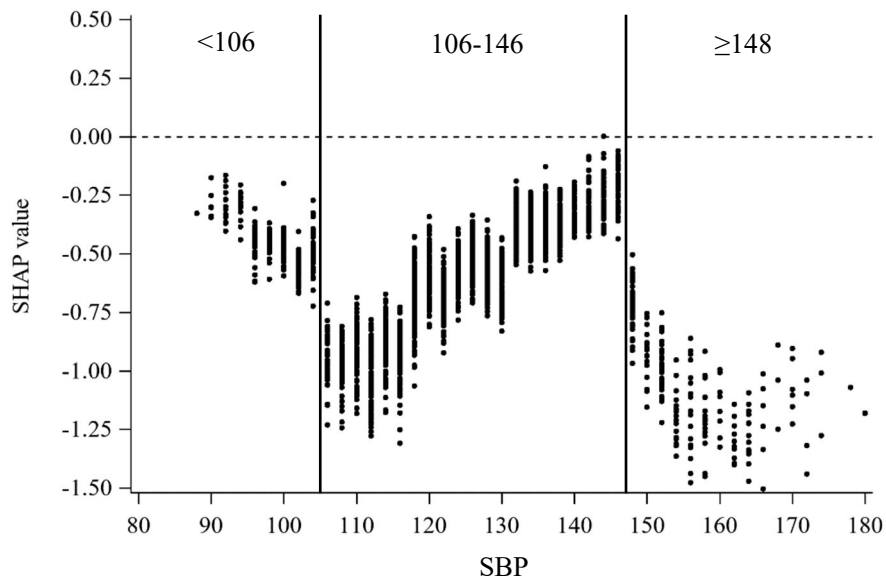


図8 SBP と SHAP の関係



収縮期血圧（座位）と対象者属性

SBP group (mmHg)	n	男性 (%)	年齢 (年)	BMI (kg/m ²)	CAVI (unit)	臥位 SBP (mmHg)
<106	286	29	42.9	20.7	7.2	104.0
106-146	2188	51	49.8	22.9	7.7	121.3
≥148	208	52	58.9	24.4	8.5	148.6

収縮期血圧（座位）と SHAP 値

SBP group (mmHg)	座位 SBP	男性	年齢	BMI	CAVI	臥位 SBP
<106	-0.470	-0.514	-2.172	-1.244	-1.557	-1.175
106-146	-0.609	-0.296	-1.637	-1.168	-1.103	-0.842
≥148	-1.051	-0.251	-0.845	-1.079	-0.582	-0.705

第4章 研究の考察

本研究では、健康診断受診データを使用して、心房細動新規発症を予測する機械学習モデルを作成した。本モデルの特徴は2つある。1つは高血圧、動脈硬化の指標である CAVI、5つの血圧指標（健診時における座位の収縮期血圧・拡張期血圧・脈圧および CAVI 測定時における臥位の収縮期血圧・拡張期血圧）を同時に加味したことである。もう1つは、SHAP を利用して、全体および個々人の心房細動発症に関連する要因を評価できることである。予測モデルには心房細動発症に関連すると考えられる 36 要因を投入したが、臨床検査 10 項目を除外した予測モデルも作成し、臨床検査結果が不明な場合でも対処できるようにした。

ロジスティック回帰モデルやコックス回帰モデルのような一般的な統計モデルではデータの独立性や変数間の多重共線性などいくつかの仮定が必要となるが、機械学習モデルでは仮定が少なくなると考えられている[35]。今回使用した勾配ブースティング (XGBoost) はデータサイエンス分野で広く利用されており、多くの機械学習コンペティションで良好な結果を記録している[35]。

年齢は心房細動発症に強く関連しており、多くの予測モデルが年齢を考慮している[20, 22, 32]。フラミンガム研究では、動脈硬化の指標である augmentation index、中心脈圧、血流依存性血管拡張反応、上腕動脈血管径が心房細動発症と

強く関連することを示した[9]。本研究においても、年齢と CAVI は心房細動発症と強く関連していた。

高血圧の新規発症と関連する CAVI[45]と CAVI 測定時の収縮期血圧は、36 要因を考慮した予測モデルにおいて心房細動新規発症の予測因子の上位 10 以内であった。また、臨床検査 10 項目を除外した予測モデルにおいても 5 つの血圧指標である CAVI 測定時の収縮期血圧、拡張期血圧および健診時の収縮期血圧、拡張期血圧、脈圧が心房細動新規発症の予測因子の上位 10 以内であった。これまでの疫学研究では、CAVI が脳卒中、冠動脈疾患、心不全など循環器疾患と関連していたとの報告は多い[50]。また、CAVI は発作性心房細動の独立したリスクであるとの報告もある[51]。CAVI が示す動脈硬化（ステッフネス）の増大は収縮期にかかる心臓後負荷の増大になる。また、高血圧は心房細動新規発症のリスクを 1.8 倍増加させ、持続性心房細動へ進展するリスクを 1.5 倍増加させることが示されている[7, 8]。一般的に座位測定時の血圧が高血圧の診断に使われているが、本研究では CAVI 測定時における臥位の収縮期血圧が心房細動新規発症と強く関連していた。我々が以前作成した高血圧新規発症を予測するモデルにおいても[52]、CAVI 測定時における臥位の収縮期血圧が高血圧新規発症と強く関連していた。臥位による心臓への静脈灌流量の増加は、心房と心室への過剰負荷を引き起こす[53]。この心前負荷の増加に加え、臥位における収縮期血圧増加

による心後負荷の増加が収縮期にさらに増大すると考えられる。24 時間自由行動下血圧測定と心不全の関連を検討した研究において、夜間における臥位血圧は日中における座位および立位血圧よりも心不全と強く関連していた[54]。これらの知見は様々な状況における血圧評価の重要性を示しており、臥位における血圧測定は座位における血圧測定よりも将来の心房細動発症の予測に有用であるかもしれない。

メタボリックシンドロームの構成要素である BMI、中性脂肪、収縮期血圧、尿酸と腎機能（eGFR）も心房細動新規発症の予測因子の上位 10 以内であった。Chamberlain et al. [20]は収縮期血圧と中性脂肪の増加が心房細動発症と関連していることを示した（ハザード比 2.63、収縮期血圧 160 mmHg 以上 vs. 120 mmHg 以下、ハザード比 2.61、中性脂肪 200 mg/dL 以上 vs. 150 mg/dL 以下）。また、日本の研究において[23]、収縮期血圧と尿酸は心房細動発症と有意に関連していた（1 標準偏差増加当たりのハザード比、収縮期血圧 1.19、尿酸 1.13）。

予測モデルの結果を解釈することは非常に重要である。ブラックボックスである機械学習のアルゴリズムを説明可能にし、なぜ予測モデルがそのように予測したのか理解するためにゲーム理論に基づく SHAP を使用した。心房細動発症の予測確率が高い場合、SHAP からどのリスク要因の影響が大きかったか特定することができた。

本研究は、経時的で大規模な標準化された健康診断データを使用した。対象集団の特性や計測方法に関して、いくつかの限界がある。連続した 4 年間の健康診断データを用いて予測モデルを作成したため、長期間における心房細動新規発症リスクを予測することはできない。心電図検査の判定結果を活用したが、波形データは保存されておらず、波形情報を活用することができなかった。日本人を対象とした研究であり、一般的に健康で若いと考えられる健康診断受診者を対象としたため、心房細動の新規発症率が小さく、心房細動新規発症者が少なかった。この点は SMOTE によりデータ不均衡を改善した場合でもモデルの向上はみられなかった原因であると考えられる。今後より心房細動有症率が高い集団について検証を行いたい。

第5章 結論

心房細動は脳卒中の重要なリスク要因であり、人口の高齢化と循環器疾患における生存率の向上により、心房細動患者数は今後 50 年間で 2.5 倍に増加すると考えられている。従って、心房細動の将来の新規発症リスクを予測し、予防することが重要である。

本研究では、健康診断受診者データを使用して、心房細動新規発症を予測する機械学習モデルを作成した。本モデルは心房細動発症の関連要因である高血圧、動脈硬化の指標である CAVI、5 つの血圧指標（健診時における座位の収縮期血圧・拡張期血圧・脈圧および CAVI 測定時における臥位の収縮期血圧・拡張期血圧）を同時に加味した唯一のモデルである。また、ゲーム理論に基づく SHAP 値を利用して、全体および個々人の心房細動発症に関連するリスク要因を評価できるモデルである。予測モデルには心房細動発症に関連すると考えられる 36 要因を投入したが、臨床検査結果が不明な場合でも対処できるように臨床検査 10 項目を除外した予測モデルも作成し、いずれの予測モデルでも予測能に大きな差はみられなかった。一般的に座位測定時の血圧が高血圧の診断に使われているが、本研究では CAVI 測定時における臥位の収縮期血圧が心房細動新規発症と強く関連していた。

本研究で開発されたモデルは、臨床医が患者の将来の心房細動新規発症リス

クを評価するために使用することができる。この結果に基づいて、臨床医は患者に合わせた生活習慣の改善を計画し、心房細動発症を予防または遅らせるような治療方針を決定することができる。今後は、本予測モデルを健康診断や実臨床で実装し、心房細動新規発症リスクを評価し、心房細動の発症予防の可能性について検証したい。また、臥位における血圧測定の有用性についても検討を進めたい。

第6章 謝辞

本研究を行うにあたり、研究全般にわたりご指導いただきました自治医科大学内科学講座 循環器内科学部門 苅尾七臣 教授に厚く御礼申し上げます。また、研究実施に協力いただきました自治医科大学循環器内科学教室の大倉綾子 様、松本祐里 様、濱寄春菜 様、諏訪裕美 様、医療法人社団こころとからだの元気プラザの中村哲也 様、及川孝光 様、研究データの使用に協力をいただいた一般財団法人日本健康増進財団の藤代健太郎 様、鈴木賢二 様、フクダ電子株式会社の深谷恭平 様、伊藤哲也 様、原田中裕 様に、心から感謝を申し上げます。

参考文献

1. 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;285:2370-5.
2. Friberg L, Bergfeldt L. Atrial fibrillation prevalence revisited. *J Intern Med*. 2013;274:461-8.
3. Zoni-Berisso M, Lercari F, Carazza T, Domenicucci S. Epidemiology of atrial fibrillation: European perspective. *Clin Epidemiol*. 2014;6:213-20.
4. Ohsawa M, Okayama A, Sakata K, Kato K, Itai K, Onoda T, et al. Rapid increase in estimated number of persons with atrial fibrillation in Japan: an analysis from national surveys on cardiovascular diseases in 1980, 1990 and 2000. *J Epidemiol*. 2005;15:194-6.
5. Berge T, Lyngbakken MN, Ihle-Hansen H, Brynildsen J, Pervez MO, Aagaard EN, et al. Prevalence of atrial fibrillation and cardiovascular risk factors in a 63-65 years old general population cohort: the Akershus Cardiac Examination (ACE) 1950 Study. *BMJ Open*. 2018;8:e021704.
6. World Health Organization. *World Population Prospects 2022*. Available from: https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/wpp2022_summary_of_results.pdf (Accessed 30 Apr 2024).
7. 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;271:840-4.
8. de Vos CB, Pisters R, Nieuwlaat R, Prins MH, Tieleman RG, Coelen RJ, et al. Progression from paroxysmal to persistent atrial fibrillation clinical correlates and prognosis. *J Am Coll Cardiol*. 2010;55:725-31.
9. Shaikh AY, Wang N, Yin X, Larson MG, Vasan RS, Hamburg NM, et al. Relations of Arterial Stiffness and Brachial Flow-Mediated Dilation With New-Onset Atrial Fibrillation: The Framingham Heart Study. *Hypertension*. 2016;68:590-6.
10. Dernellis J, Panaretou M. Aortic stiffness is an independent predictor of progression to hypertension in nonhypertensive subjects. *Hypertension*. 2005;45:426-31.
11. Liao D, Arnett DK, Tyroler HA, Riley WA, Chambless LE, Szklo M, et al. Arterial stiffness and the development of hypertension. The ARIC study. *Hypertension*. 1999;34:201-6.
12. Najjar SS, Scuteri A, Shetty V, Wright JG, Muller DC, Fleg JL, et al. Pulse wave

- velocity is an independent predictor of the longitudinal increase in systolic blood pressure and of incident hypertension in the Baltimore Longitudinal Study of Aging. *J Am Coll Cardiol*. 2008;51:1377-83.
13. Takase H, Dohi Y, Toriyama T, Okado T, Tanaka S, Sonoda H, et al. Brachial-ankle pulse wave velocity predicts increase in blood pressure and onset of hypertension. *Am J Hypertens*. 2011;24:667-73.
 14. Gage BF, Waterman AD, Shannon W, Boechler M, Rich MW, Radford MJ. Validation of clinical classification schemes for predicting stroke: results from the National Registry of Atrial Fibrillation. *JAMA*. 2001;285:2864-70.
 15. 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;137:263-72.
 16. 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.
 17. Alonso A, Roetker NS, Soliman EZ, Chen LY, Greenland P, Heckbert SR. Prediction of Atrial Fibrillation in a Racially Diverse Cohort: The Multi-Ethnic Study of Atherosclerosis (MESA). *J Am Heart Assoc*. 2016;5:e003077.
 18. Aronson D, Shalev V, Katz R, Chodick G, Mutlak D. Risk Score for Prediction of 10-Year Atrial Fibrillation: A Community-Based Study. *Thromb Haemost*. 2018;118:1556-63.
 19. Brunner KJ, Bunch TJ, Mullin CM, May HT, Bair TL, Elliot DW, et al. Clinical predictors of risk for atrial fibrillation: implications for diagnosis and monitoring. *Mayo Clin Proc*. 2014;89:1498-505.
 20. 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.
 21. Ding L, Li J, Wang C, Li X, Su Q, Zhang G, et al. Incidence of atrial fibrillation and its risk prediction model based on a prospective urban Han Chinese cohort. *J Hum Hypertens*. 2017;31:574-9.
 22. 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;34:2243-51.
 23. Hamada R, Muto S. Simple risk model and score for predicting of incident atrial

- fibrillation in Japanese. *J Cardiol*. 2019;73:65-72.
24. Hata J, Nagata T, Sakata S, Oishi E, Furuta Y, Hirakawa Y, et al. Risk Prediction Model for Incident Atrial Fibrillation in a General Japanese Population - The Hisayama Study. *Circ J*. 2021;85:1373-82.
 25. Hulme OL, Khurshid S, Weng LC, Anderson CD, Wang EY, Ashburner JM, et al. Development and Validation of a Prediction Model for Atrial Fibrillation Using Electronic Health Records. *JACC Clin Electrophysiol*. 2019;5:1331-41.
 26. Igarashi Y, Nochioka K, Sakata Y, Tamai T, Ohkouchi S, Irokawa T, et al. Risk prediction for new-onset atrial fibrillation using the Minnesota code electrocardiography classification system. *Int J Cardiol Heart Vasc*. 2021;34:100762.
 27. 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;81:1580-8.
 28. Li YG, Pastori D, Farcomeni A, Yang PS, Jang E, Joung B, et al. A Simple Clinical Risk Score (C(2)HEST) for Predicting Incident Atrial Fibrillation in Asian Subjects: Derivation in 471,446 Chinese Subjects, With Internal Validation and External Application in 451,199 Korean Subjects. *Chest*. 2019;155:510-8.
 29. Linker DT, Murphy TB, Mokdad AH. Selective screening for atrial fibrillation using multivariable risk models. *Heart*. 2018;104:1492-9.
 30. Rienstra M, Geelhoed B, Yin X, Siland JE, Vermond RA, Mulder BA, et al. Cluster Individuals Based on Phenotype and Determine the Risk for Atrial Fibrillation in the PREVEND and Framingham Heart Study Populations. *PLoS One*. 2016;11:e0165828.
 31. Saliba W, Gronich N, Barnett-Griness O, Rennert G. Usefulness of CHADS2 and CHA2DS2-VASc Scores in the Prediction of New-Onset Atrial Fibrillation: A Population-Based Study. *Am J Med*. 2016;129:843-9.
 32. 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;373:739-45.
 33. Suenari K, Chao TF, Liu CJ, Kihara Y, Chen TJ, Chen SA. Usefulness of HATCH score in the prediction of new-onset atrial fibrillation for Asians. *Medicine (Baltimore)*. 2017;96:e5597.
 34. Yum Y, Shin SY, Yoo H, Kim YH, Kim EJ, Lip GYH, et al. Development and Validation of 3-Year Atrial Fibrillation Prediction Models Using Electronic Health Record With or Without Standardized Electrocardiogram Diagnosis and a Performance Comparison Among Models. *J Am Heart Assoc*. 2022;11:e024045.

35. Chen T. XGBoost: A Scalable Tree Boosting System. *arXiv*. 2016:1603.02754.
36. Ashburner JM, Chang Y, Wang X, Khurshid S, Anderson CD, Dahal K, et al. Natural Language Processing to Improve Prediction of Incident Atrial Fibrillation Using Electronic Health Records. *J Am Heart Assoc*. 2022;11:e026014.
37. Bundy JD, Heckbert SR, Chen LY, Lloyd-Jones DM, Greenland P. Evaluation of Risk Prediction Models of Atrial Fibrillation (from the Multi-Ethnic Study of Atherosclerosis [MESA]). *Am J Cardiol*. 2020;125:55-62.
38. He J, Liu S, Yang C, Wei Y. Value of baseline characteristics in the risk prediction of atrial fibrillation. *Front Cardiovasc Med*. 2023;10:1068562.
39. 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.
40. Hu WS, Hsieh MH, Lin CL. A novel atrial fibrillation prediction model for Chinese subjects: a nationwide cohort investigation of 682 237 study participants with random forest model. *Europace*. 2019;21:1307-12.
41. Nadarajah R, Wu J, Hogg D, Raveendra K, Nakao YM, Nakao K, et al. Prediction of short-term atrial fibrillation risk using primary care electronic health records. *Heart*. 2023;Feb 9;heartjnl-2022-322076 (online ahead of print).
42. Sekelj S, Sandler B, Johnston E, Pollock KG, Hill NR, Gordon J, et al. Detecting undiagnosed atrial fibrillation in UK primary care: Validation of a machine learning prediction algorithm in a retrospective cohort study. *Eur J Prev Cardiol*. 2021;28:598-605.
43. Tiwari P, Colborn KL, Smith DE, Xing F, Ghosh D, Rosenberg MA. Assessment of a Machine Learning Model Applied to Harmonized Electronic Health Record Data for the Prediction of Incident Atrial Fibrillation. *JAMA Netw Open*. 2020;3:e1919396.
44. Toprak B, Brandt S, Brederbecke J, Gianfagna F, Vishram-Nielsen JKK, Ojeda FM, et al. Exploring the incremental utility of circulating biomarkers for robust risk prediction of incident atrial fibrillation in European cohorts using regressions and modern machine learning methods. *Europace*. 2023;25:812-9.
45. Kario K, Kanegae H, Oikawa T, Suzuki K. Hypertension Is Predicted by Both Large and Small Artery Disease. *Hypertension*. 2019;73:75-83.
46. Hayashi K, Yamamoto T, Takahara A, Shirai K. Clinical assessment of arterial stiffness with cardio-ankle vascular index: theory and applications. *J Hypertens*. 2015;33:1742-1757.
47. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res*. 2002;16:321-357

48. Ralph B. D'Agostino, Michael J. Pencina, Joseph M. Massaro, Sean Coady. Cardiovascular Disease Risk Assessment: Insights from Framingham. *GLOBAL HEART*. 2013;8:11-23.
49. Lundberg S, Lee SI. A Unified Approach to Interpreting Model Predictions. *arXiv*. 2017:1705.07874.
50. Miyoshi T, Ito H. Arterial stiffness in health and disease: The role of cardio-ankle vascular index. *J Cardiol*. 2021;78:493-501.
51. Miyoshi T, Doi M, Noda Y, Ohno Y, Sakane K, Kamikawa S, et al. Arterial stiffness determined according to the cardio-ankle vascular index is associated with paroxysmal atrial fibrillation: a cross-sectional study. *Heart Asia*. 2014;6:59-63.
52. Kanegae H, Suzuki K, Fukatani K, Ito T, Harada N, Kario K. Highly precise risk prediction model for new-onset hypertension using artificial intelligence techniques. *J Clin Hypertens (Greenwich)*. 2020;22:445-50.
53. Kario K, Williams B. Nocturnal Hypertension and Heart Failure: Mechanisms, Evidence, and New Treatments. *Hypertension*. 2021;78:564-77.
54. Kario K, Hoshida S, Mizuno H, Kabutoya T, Nishizawa M, Yoshida T, et al. Nighttime Blood Pressure Phenotype and Cardiovascular Prognosis: Practitioner-Based Nationwide JAMP Study. *Circulation*. 2020;142:1810-20.

参考資料

参考資料 1 2005 年から 2015 年に一般財団法人日本健康増進財団で健康診断を受診した者で、本研究対象者と非対象者の比較

No.	変数名	本研究対象者 n=13410	本研究非対象者 n=85057
1	年齢, 歳	49.8±12.8	47.7±13.6
2	性別, 男性	6547 (48.8)	43691 (51.4)
3	BMI, kg/m ²	22.8±3.4	22.7±3.5
4	収縮期血圧, mmHg	125.1±15.8	123.8±16.4
5	拡張期血圧, mmHg	73.5±10.9	73.4±11.6
6	脈圧, mmHg	51.5±10.5	50.4±10.7
7	ALT, IU/L	23.0±11.2	22.9±11.5
8	AST, IU/L	22.9±17.3	22.9±19.0
9	γ-GTP, IU/L	38.1±52.0	38.2±55.5
10	尿酸, mg/dL	5.1±1.4	5.1±1.4
11	中性脂肪, mg/dL	107.2±90.0	107.8±100.5
12	LDL コレステロール, mg/dL	126.3±32.7	123.6±33.8
13	HDL コレステロール, mg/dL	67.8±19.2	67.4±18.3
14	eGFR, ml/min/1.73m ²	84.2±10.8	86.0±11.6
15	尿蛋白, 陽性	301 (2.2)	4742 (5.7)
16	尿糖, 陽性	177 (1.3)	1752 (2.1)
17	高血圧, あり	1624 (12.1)	9248 (10.9)
18	不整脈, あり	156 (1.2)	1030 (1.3)
19	左室肥大 1 ^{※1} , あり	764 (5.7)	4488 (5.4)
20	左室肥大 2 ^{※2} , あり	100 (0.8)	667 (0.8)
21	糖尿病, あり	368 (2.7)	2438 (2.9)
22	脂質異常症, あり	1022 (7.6)	5327 (6.3)
23	慢性腎臓病, あり	250 (1.9)	1572 (1.9)
24	循環器疾患 ^{※3} , あり	405 (3.0)	2743 (3.2)
25	心疾患 ^{※4} , あり	295 (2.2)	2100 (2.5)
26	脳血管疾患 ^{※5} , あり	116 (0.9)	747 (0.9)
27	喫煙, あり	2351 (17.5)	13510 (15.9)
28	飲酒頻度, 0 日/週	7126 (53.1)	48022 (56.5)
29	飲酒頻度, 1-6 日/週	4303 (32.1)	26627 (31.3)

30	飲酒頻度, 7 日/週	1981 (14.8)	10408 (12.2)
31	飲酒量, 0 合/日	7133 (53.2)	48089 (56.5)
32	飲酒量, 1-2 合/日	5137 (38.3)	30051 (35.3)
33	飲酒量, 3 合/日以上	1140 (8.5)	6917 (8.1)
34	CAVI, unit	7.7±1.0	7.6±1.0
35	CAVI 測定時の収縮期血圧, mmHg	121.7±16.2	121.2±16.9
36	CAVI 測定時の拡張期血圧, mmHg	75.8±11.2	75.3±11.7

※1 標準 12 誘導心電図の高振幅

※2 標準 12 誘導心電図の高振幅+ST 異常

※3 心疾患+脳血管疾患

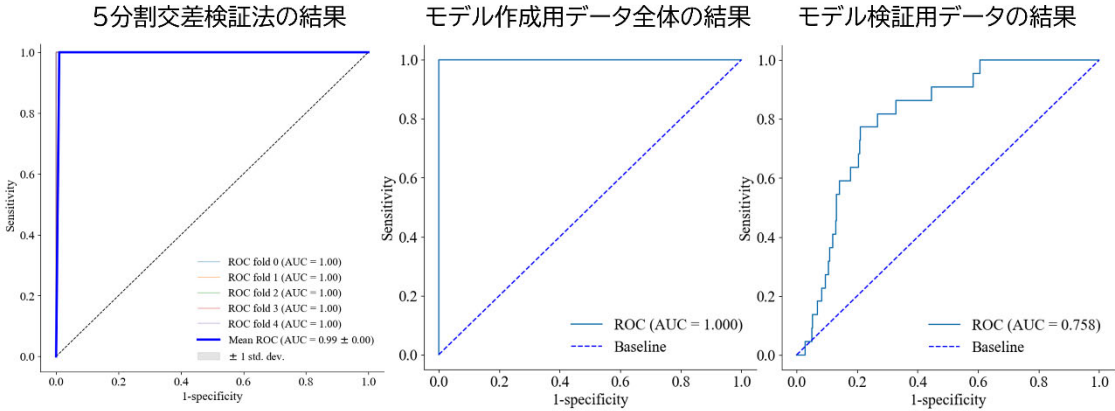
※4 狭心症、心筋梗塞、心臓病

※5 脳卒中、脳梗塞、脳内出血、くも膜下出血

参考資料 2 最適なハイパーパラメーターの組み合わせ

パラメーター	36 要因モデル	26 要因モデル
Colsample by tree	0.3	0.6
Gamma	0.0001	0.01
Max depth	15	15
Min child weight	7	5
Reg alpha	0.03	0.01
Reg lambda	0.0001	0.03
Subsample	1.0	0.3

参考資料3 Synthetic Minority Over-sampling Technique を用いた解析結果



True label	Predicted label		AUC: 0.758
	0	1	Accuracy: 0.789
0	2099	561	Precision: 0.028
1	6	16	Recall: 0.727
			F1 score: 0.053