

氏 名	かねがえ ひろし 鐘江 宏
学 位 の 種 類	博士 (医学)
学 位 記 番 号	乙第 870 号
学位授与年月日	令和 6 年 12 月 23 日
学位授与の要件	自治医科大学学位規定第 4 条第 3 項該当
学 位 論 文 名	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 審 査 委 員	(委員長) 教 授 藤 田 英 雄 (委 員) 教 授 今 井 靖 准教授 桑 原 政 成 (学外委員) 准教授 野 村 章 洋

## 論文内容の要旨

### 1 研究目的

心房細動は脳卒中の重要な危険因子である。高血圧を合併する心房細動では脳卒中リスクがさらに高くなる。心房細動の新規発症を予測するモデルは数多く存在し、機械学習を用いたモデルも発表されている。動脈硬化は高血圧発症に先行することが知られており、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。

### 2 研究方法

本研究は、2005 年から 2015 年の間に少なくとも 4 年連続して健康診断を受診した 13,410 人(新規発症心房細動:110 名、非心房細動:13,300 名)を対象とした。機械学習モデルは eXtreme Gradient Boosting を使用し、Shapley Additive Explanation によりモデルを解釈した。データは、トレーニングセット (80%) とテストセット (20%) にランダムに分割し、5 分割交差検証法によりハイパーパラメータのチューニングを行った。モデルには既往歴、生活習慣、身体計測や生化学検査値など心房細動と関連すると考えられる 36 変数を投入した。

### 3 研究成果

対象集団は平均年齢 50 歳、男性 49%であった。心房細動新規発症者は高齢者で、男性が多く、診療室血圧や CAVI が高く、高血圧、循環器疾患、不整脈や左室肥大を有する者が多かった。モデル作成用データの AUC は 0.85 で、モデル検証用データの AUC は 0.79 であった。予測モデルにおける重要度の高い予測因子は年齢であり、次いで CAVI、eGFR、性別、BMI、尿酸、 $\gamma$ -GTP、中性脂肪、CAVI 測定時の収縮期血圧、ALT であった。個人の SHAP 値を整理することにより、心房細動の発症確率の高低と合わせて、どのような要因の影響が大きかったかを把握することが可能であった。

#### 4 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。

#### 5 結論

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。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、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。

### 論文審査の結果の要旨

脳卒中の重要な危険因子としての心房細動は臨床的に重要である。心房細動を発症予測するモデルは数多く発表されているが、高血圧と動脈硬化からの心房細動を発症予測モデルはこれまでにない。本研究は、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ている。2005 年から 2015 年の間に少なくとも4年連続して健康診断を受診した13,410人（新規発症心房細動:110名、非心房細動:13,300名）を対象とした。機械学習モデルは eXtreme Gradient Boosting を使用し、Shapley Additive Explanation によりモデルを解釈した。動脈硬化指標として CAVI を採用し、血圧は通常の座位血圧に加えて CAVI 測定時の臥位血圧を含め、交絡因子の考察への必要性が乏しい XGBoost を採用して心房細動発症予測モデルを構築したことにまず新規性と独創性がある。更に、包括的に36変数という多数のパラメータの要因分析が可能となり、その結果モデル作成用データの AUC は 0.85 で、モデル検証用データの AUC は 0.79 と高い予測能をもつモデルの構築に成功している。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の評価が可能となった。

審査会では、細部の指摘事項とともに改訂を要する事項として①対象期間中に健診を受けたが研究対象にならなかった9万例と、対象集団の背景が異なっていないかの検証が必要、②

AF 発症が 0.8%という不均衡データの扱いについて適正性を強化する必要があるとの指摘、③XGBoost の実態を明らかにする、があった。申請者はこれらに適切に対応し論文の改訂を行った結果、博士論文としての水準を満たしたものと認め合格判定を行った。

## 試問の結果の要旨

申請者は、以下の研究成果について発表を行った。2005 年から 2015 年の間に少なくとも 4 年連続して健康診断を受診した 13,410 人（新規発症心房細動：110 名、非心房細動：13,300 名）を対象とした。機械学習モデルは eXtreme Gradient Boosting を使用し、Shapley Additive Explanation によりモデルを解釈した。データは、トレーニングセット（80%）とテストセット（20%）にランダムに分割し、5 分割交差検証法によりハイパーパラメータのチューニングを行った。モデルには既往歴、生活習慣、身体計測や生化学検査値など心房細動と関連すると考えられる 36 変数を投入した。対象集団は平均年齢 50 歳、男性 49%であった。心房細動新規発症者は高齢者で、男性が多く、診療室血圧や CAVI が高く、高血圧、循環器疾患、不整脈や左室肥大を有する者が多かった。モデル作成用データの AUC は 0.85 で、モデル検証用データの AUC は 0.79 であった。予測モデルにおける重要度の高い予測因子は年齢であり、次いで CAVI、eGFR、性別、BMI、尿酸、 $\gamma$ -GTP、中性脂肪、CAVI 測定時の収縮期血圧、ALT であった。個人の SHAP 値を整理することにより、心房細動の発症確率の高低と合わせて、どのような要因の影響が大きかったかを把握することが可能であった。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。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。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、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。

この発表に対し、質疑応答が行われた。

今井委員より、心電図波形を予測パラメータとして採用していないことについて指摘があり、研究の限界として追記を行った。また、AF を発症している高リスク集団の分析や比較につい

での指摘に関しては、定期健康診断における心房細動の有症率は 0.3%程度と報告されており、本研究でも 0.8%であった。より心房細動を多く含むような集団での検証について今後の検討課題とした。

桑原委員より、対象期間中に健診を受けたが研究対象にならなかった 9 万例と、対象集団の背景が異なっていないかの検証が必要との指摘があり、健康診断受診者全体と本研究対象者について対象者背景を比較した表を追加した改訂を行った。また検査項目として HDL、AST、尿糖など従来報告されていなかった多数の項目が入っていることや血圧指標として SBP、DBP、CAVI 時臥位 SBP の複数が変数として採用されていることに関しては、本研究で用いた XG Boost は交絡因子について考察の必要性が乏しい手法となっていることが回答され、本論文での妥当性を説明し新規性としての確認も行った。

野村委員より、AF 発症が 0.8%という不均衡データの扱いについて適正性を強化する必要があるとの指摘に対しては、Over sampling の問題は重要であり、Synthetic Minority Over-sampling Technique (SMOTE)を用いた結果について追記を行った。

藤田委員長より XGBoost の実態をもう少し明らかにするためアンサンブルにおいて二分木の数などの表示を求めるとの指摘に対して、5 分割交差検証を用いてハイパーパラメータの探索を行った 7 パラメータについて追記を行った。また、経時的変化の寄与は分析対象となり得るかの質問に対して参考論文としている高血圧新規発症を予測する機械学習モデルを作成した際には経時的変化についても考慮した結果、モデルへの寄与としては発症前年が大きい傾向を示していたため今回の研究ではモデルや結果解釈が複雑とならないようにベースラインの情報のみを採用したとの説明を行った。

以上の指摘事項に対する回答および論文改訂により、本論文は新規性・独創性を有する研究論文として纏まった結果、委員全員が博士論文としての水準を満たしたものと認め、合格との判定を行った。