

令和 6 年 6 月 25 日現在

機関番号：17701

研究種目：基盤研究(C)（一般）

研究期間：2021～2023

課題番号：21K08034

研究課題名（和文）70万人、43年の大規模健診データによる心血管病発症AI予測モデルの構築と実装

研究課題名（英文）Construction and implementation of an AI prediction model for the onset of cardiovascular disease based on 700,000 people and 43 years of large-scale health checkup data

研究代表者

川添 晋（KAWASOE, SHIN）

鹿児島大学・医歯学総合研究科・特任講師

研究者番号：00810201

交付決定額（研究期間全体）：（直接経費） 3,200,000円

研究成果の概要（和文）：健診データの可視化とAI適用への前処理を行い、そのデータにLAMPおよび機械学習アルゴリズムを適用してモデルの作成を行った。ハイパーパラメータのチューニングを行い、いくつかの機械学習モデル（ランダムフォレスト、XGBoosting、ロジスティック回帰、ニューラルネットワーク、サポートベクトルマシン、他）について最適化を行った。アウトカムとして高血圧、慢性腎臓病、メタボリックシンドローム、動脈硬化（baPWV高値）についてモデルの作成を行った。研究成果について国内外の複数の学会で成果の発表を行い、成果の論文化も順次進めている。

研究成果の学術的意義や社会的意義

超少子高齢化社会への突入と医療の高度化・高額化に伴い、わが国の医療財政は逼迫しており、疾患の予防および早期発見による医療費の抑制が急務である。我々は43年間にわたる70万人の健診データをもとにして、個人単位での疾患の早期発見・早期治療に役立つ心血管疾患発症予測の人工知能（AI）モデルを構築・実装した。さらにそのアルゴリズムを保健指導の場で実際に指導に役立てるべく、アプリ化を現在進めている。健康寿命の延伸と医療費の抑制に寄与する、世界に先駆けた新たな医療モデルであるものと考えている。

研究成果の概要（英文）：Visualization of health checkup data and preprocessing to AI application were performed, and LAMP and machine learning algorithms were applied to the data to create models. Hyperparameters were tuned and optimized for several machine learning models (random forest, XGBoosting, logistic regression, neural network, support vector machine, and others). Models were created for hypertension, chronic kidney disease, metabolic syndrome, and atherosclerosis (high baPWV) as outcomes. The results of the research were presented at several domestic and international conferences, and the results are being published in a series of papers.

研究分野：循環器内科学

キーワード：AI データベース研究 健康診断 心血管疾患

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

1. 研究開始当初の背景

わが国は他国に先がけて超少子高齢化社会に突入し、医療費は 43 兆円を超え（平成 29 年厚労省調べ）医療財政の健全化は喫緊の課題である。死因や要介護の原因をみると、心臓病や脳卒中などの心血管疾患が第一位で全体の 19.8%をしめている（平成 30 年内閣府調べ）。国内外からの報告で、高血圧、糖尿病、喫煙などが心血管疾患の強いリスク因子として同定され、その治療薬も数多く開発されているが、疾患発症は減少していない。心血管疾患は、生活習慣がその発症に大きく影響している。個人の健康意識をさらに向上させる新たなアプローチが必要である。近年のビッグデータ活用や AI 技術の発展に伴い、疾患リスク分析の研究は新たな局面を迎えている。健診データから個人の将来の疾患発症率を予測する疾患リスクモデルを開発し、医療者と個人が協同して生活習慣病の予防に取り組んでいくことが今後重要になっていく。

2. 研究の目的

健診データに AI 技術を用い、リスク因子を総合して個人の将来の疾患発症を予測するモデルを開発することを目的とする。医学的見地からの AI モデルの評価を行い、最終的には独立コホートをを用いた社会実装を行う。個人単位での疾患発症リスクの把握や、より詳細な予防指導につなげることができ、予防医療の向上に寄与すると思われる。AI 解析を行う上で、利用可能なデータ数は重要である。我々は、鹿児島県内の 3 カ所の大規模健診センターと契約を結び、約 70 万人の重複のない個人の、1977 年からの 43 年間にわたる経年の健診データを利用可能である。これは過去の健診データにもとづく研究と比較して突出したデータ数である。全受診者の 46%において 10 年以上の長期のデータがあり、追跡性も高い。20～80 歳代まで幅広い世代をカバーしており、年代別のモデル構築も可能である。鹿児島市のみならず、郡部から離島まで幅広い生活圏を含んでいることも特徴である。対象者数、年齢層、年代、生活圏などにおいて非常に広範囲のデータベースを利用可能であることは、AI の予測精度の根幹に関わるという点で非常に大きな強みである。本研究では、AI を用いた研究についての「ブラックボックス化」の問題点が解決され、精度の高さだけでなく、人智と AI が協調した心血管疾患の予測モデルの開発が期待できる。また、43 年にわたる 70 万人規模の健診結果での発症リスク予測は先例が無く、本研究が実施できれば、世界に先駆けた研究となる。幅広い生活圏の対象者を含んでいることから、ここで開発した AI モデルは、広く全国に展開することができると考えられる。

本研究で作成された AI モデルの利用により、受診者自身の健康意識の改革、生活習慣の改善につながり、健康増進に役立てることができる。予防医療の発展に貢献し、医療・介護費用の抑制による財政の健全化に資することができる。また、当該データは 1977 年からの記録であるが、40 年前と今では食事や生活スタイルも大きく変わり、体格も異なる。心血管疾患リスクに関する「時代効果」への検証も可能となると思われる。

3. 研究の方法

健診データベースに AI を導入し、より個別化指導につながる予測モデルを構築する。作成されたモデルを医学的な立場から検証し、臨床的に実装可能な状態にする。独立したコホートに対してモデルを適用し、予測精度や医療費削減効果を評価する。

健診データの可視化と AI 適用への前処理（初年度）

- ・データエラーやノイズ、欠損値に対するクレンジング
- ・既存の統計解析手法によるデータ傾向の調査
- ・データの変換（正規化、汎化など）を実施し、AI 解析の実施に耐えうるデータに整形
- ・リスク因子の年次変動を加味した変数の合成

LAMP および他の機械学習モデルの作成（次年度以降）

- ・統計的に有意な相互因子を発見可能な LAMP を適用
- ・パラメータチューニング
- ・他の機械学習モデルを適用（ランダムフォレスト、SVM、ニューラルネットワーク等）
- ・上記結果の比較による予測精度の検証
- ・医学的な見地からの解釈を検討し、臨床的に実装可能な状態にする

- 実装に向けての独立コホートをを用いた検証（次年度以降）
- ・垂水研究に予測アルゴリズムを適用
- ・疾患予測精度の評価
- ・実装により疾患予防が実現した場合の医療費削減効果を調査

4．研究成果

健診データの可視化と AI 適用への前処理を行い、そのデータに LAMP および機械学習アルゴリズムを適用してモデルの作成を行った。ハイパーパラメータのチューニングを行い、いくつかの機械学習モデル（ランダムフォレスト、XGBoosting、ロジスティック回帰、ニューラルネットワーク、サポートベクトルマシン、他）について最適化を行った。アウトカムとして高血圧、慢性腎臓病、メタボリックシンドローム、動脈硬化（baPWV 高値）についてモデルの作成を行った。研究成果について国内外の複数の学会で成果の発表を行い、成果の論文も順次進めている。

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計4件（うち招待講演 0件 / うち国際学会 2件）

1. 発表者名 川添晋
2. 発表標題 Development of a risk prediction equation for chronic kidney disease using Japanese health checkup data
3. 学会等名 日本循環器学会総会
4. 発表年 2022年

1. 発表者名 川添晋
2. 発表標題 健診データを用いた機械学習アルゴリズムによる高血圧の発症予測
3. 学会等名 日本心臓病学会
4. 発表年 2022年

1. 発表者名 Kawasoe Shin
2. 発表標題 Ability of machine learning algorithms to predict chronic kidney disease using health examination data from the Japanese general population
3. 学会等名 American Heart Association (国際学会)
4. 発表年 2022年

1. 発表者名 Kawasoe Shin
2. 発表標題 Ability of machine learning algorithms to predict Hypertension using health examination data from the Japanese general population
3. 学会等名 European Society of Cardiology (国際学会)
4. 発表年 2022年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分担者	窪菌 琢郎 (Kubozono Takuro) (00598013)	鹿児島大学・医歯学域医学系・講師 (17701)	
研究 分担者	大石 充 (Ohishi Mitsuru) (50335345)	鹿児島大学・医歯学域医学系・教授 (17701)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------