

科学研究費助成事業 研究成果報告書

平成 29 年 9 月 1 日現在

機関番号：84404

研究種目：基盤研究(C) (一般)

研究期間：2014～2016

課題番号：26460790

研究課題名(和文)大規模コホートに基づく個別化されたリスク予測ツール、バイオマーカーの有用性の研究

研究課題名(英文)Chort based Individualized Risk Estimation By biomarkers and Machine Learning

研究代表者

西村 邦宏(Kunihiro, Nishimura)

国立研究開発法人国立循環器病研究センター・研究開発基盤センター・室長

研究者番号：70397834

交付決定額(研究期間全体)：(直接経費) 3,900,000円

研究成果の概要(和文)：冠動脈疾患発症の10年間発症危険度を予測する日本人を対象とした初めてのリスクスコアである吹田スコアを開発した。発症に関する前向き研究データから、古典的サポートベクターマシンおよびランダムフォレストの複数モデルの組み合わせにより、日常診療上よく使用される約40項目のルーチン検査から、M I D A S研究の総数1920症例に基づき、アンサンブル学習を用いた複数モデルの組み合わせによるMACE(Major Adverse Cardiac Event)予測モデルを構築し、約40モデルの識別距離の組み合わせから偽陽性5%程度、偽陰性5%程度で予測するモデルを構築し、既存予測モデルの成績より良好な成績を得た

研究成果の概要(英文)：We developed coronary heart disease (CHD) risk score, called Suita Score based on 10 years follow-up results from Suita study for primary prevention. We also developed the risk score for major cardiac disease(MACE) prediction score based on the follow-up data of 1920 prospective observational cases with cardiovascular diseases using machine learning. We adopted ensemble learning with support vectors and random forests to predict 1 year MACE in the secondary prevention using around 40 routinely used biomarkers in daily practices. The total accuracy was around 90% and it was reproducible in validation data sets, with more precise prediction ability than existing MACE prediction model, such as GRACE score.

研究分野：循環器疫学、医学統計学

キーワード：コホート研究 リスク予測 機械学習 虚血性心疾患

1. 研究開始当初の背景

世界に例をみない速度で進む超高齢社会の本邦において、脳卒中、心筋梗塞の医療の整備は喫緊の課題である。心脳血管疾患は日本人の死因の 25.5%を占め、国民医療費の約 20%を循環器病が占める(がん 13%)という現状を生み出している。一方で、近年、医療情報電子化の導入、推進において診療記録の電子化が大きな推進力となり、ビッグデータの利用が盛んにおこなわれるようになってきている。マッキンゼーの試算ではビッグデータの活用により、米国では年 30 兆円-40 兆円の医療費削減が見込まれている。AHA/ACC の脳卒中、心臓病に関する登録データベースである Get with the Guideline の登録件数が 200 万例を超えるなど、大規模臨床登録データベースに基づくビッグデータを活用した研究が盛んになっている。(Lancet.2015,JAMA2007)。これらのデータを用いた人工知能応用の進歩に伴い、従来の古典的単層ニューラルネットワークより、hidden layer を多層化させて、学習能を向上させることが可能な深層学習手法(いわゆる deep learning)が開発されつつあり、画像認識において、自動的に高い識別能を示すことが近年明らかになってきている。(Nature 2015) また予後に直接関連するアルゴリズムや (Sci Transl Med. 2011、拡張性心不全において、機械学習による “phenomapping” 解析により、新たな分類体系を確立し、予後予測を改善する可能性が報告された (Shara: Circ, 2015))

2. 研究の目的

循環器疾患、脳卒中高リスク群を、循環器学会 JROAD、厚労科研 J-ASPECT 研究などのレジストリーおよび国立循環器病研究センター、電子カルテ検査データに加え、MRI による不安定プラーク像などから、人工知能による機械学習により、1 次、2 次予防の革新的な臨床診断意思決定システムを開発する。これらのハイリスク群における診断治療の指摘戦略を、世界的に著名な Capewell らによる IMPACT モデルなど政策評価モデル、マイクロシミュレーションなどにより効率性、費用効果性を検討し、今後のわが国における循環器疾患診療、予防の重点を明らかにすることを目的とする。

3. 研究の方法

1. 一次予防に関するリスクスコア

吹田コホート研究は、都市部住民のランダムサンプルを対象としたわが国で数少ない前向きコホート研究であり 1989 年より 6485 人の健常者のフォローを行っており、健診で得られた結果と心血管疾患リスクとの関連を検討することが可能である。我々は 1989-2007 年までの 18 年間のフォローデータを基にフラミンガムリスクスコアと同様の手法を用いて 10 年間の虚血性心疾患発症リスクに関する予測スコアを開発した。(吹田スコア)

2. 二次予防に関するリスクスコア

循環器疾患発症に関する前向き研究 MIDAS 研究の 2273 症例のデータから、1 年後の MACE (Major Adverse Cardiac Event) の発症を人工知能により予測可能かを検討した。古典的サポートベクターマシンおよびランダムフォレストの複数モデルを組み合わせたアンサンブル学習により、日常診療上よく使用される約 40 項目のルーチン検査および既往歴を用いることで、MACE 予測モデルを構築した。学習用データセットでは偽陽性 5%程度、偽陰性立 5%程度で予測するモデルが生成可能であり、検証用データセットでも高リスク群、低リスク群に関して AUROC 0.88 で予後予測が可能であった。この成績は GRACE 研究などの既存の MACE 予測モデルの成績より感度、特異度ともに約 20%程度向上しており、現在特許出願中である。(図 1、図 2)

また野口らとともに、MRI を用いた不安定プラークの画像上の特徴に関する研究を世界に先駆けて行い、冠動脈高輝度プラーク(冠動脈 high-intensity plaque:HIP)の予測精度に関して検討を行った。

3. 人口知能応用による自然言語処理による予後予測改善の検討

また、同群の患者に関して、電子カルテ記事から人工知能応用による自然言語処理を用い、胸痛、浮腫など症状の有無(大項目 8 項目、小項目 25 項目)について自動抽出が可能かを検討した上で予測因子としての有用性を検討した。専門医 2 名、統計学者 2 名が辞書チューニングを行い、AI による教師あり学習をサポートした。(図 3) 入院中に自由記載されたカルテの SOAP 記事を使用した(看護記録含む)約 60 万行のカルテ記事を読み込み、症候の出現頻度を患者ごとに集積したところ、ほぼ医療職から見て問題ない精度の症候抽出は可能であった。

さらに抽出した症候を既存の予測モデルに加えた場合 10%程度の予測精度の向上を認めた。(図 4)

4. 研究成果

1. 一次予防に関するリスクスコア

冠動脈疾患発症の 10 年間発症危険度を予測する日本人を対象とした初めてのリスクスコアである吹田スコアを開発した。(表 1) 結果として、フラミンガムリスクスコアは、中国など他のアジア諸国において有効であった集団のリスク因子の平均値によるキャリブレーションなどの補正手段を用いても、発症確率が極めて低い日本人集団においてはリスクを過大評価する傾向があり、今回の吹田研究を用いた場合にはより正確に予測することが可能であった。本研究は、日本動脈硬化学会の次期ガイドラインでの利用が予定されており、また日本動脈硬化学会の最優秀論文賞を受賞した。

2.3 二次予防に関するリスクスコアおよび人工知能応用

予想される結果と意義

本研究におけるような人工知能応用によるリスク予測は、研究事例が少なく、また本研究におけるような全国規模のレジストリーおよび大規模専門病院の悉皆データによるビッグデータ分析の事例は本邦初となる。深層学習の精度は近年急速に進歩しており、今後さらに精度が向上することが見込まれる。特に画像認識については研究が蓄積されており、ツールも多い。ただし、今回の目的である将来のリスク予測のような、深層学習だけでは要因を特定しづらく、まだ事例はほとんどない。この研究のためには当然ながらビッグデータの蓄積が必要となるが、我々はすでに貴重なデータを大量に蓄積しており、競合に対して絶対的な優位性を持つ。本研究の成果は、医療保険者のデータベースから、JCN7 などの従来の高血圧ガイドラインでは治療対象にならないが、複合的なリスクを有する患者群に対して機械学習を行い、個人個人のリスクにより個別化された高血圧治療戦略を採用することにより、同一コストで、47%多く心筋梗塞を予防できるようになり、同一のシステムによりカイザーパーマネンテは加入者百万人あたり\$98M(約 95 億円)の医療費削減に成功している例(Ann Intern Med. 2011)など、医療費削減につながる個別化医療推進を行うことも可能と思われる。その他新規デバイス、新薬、難病希少疾患の case-finding など様々な応用が可能であると思われる。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計 件)

〔学会発表〕(計 件)

〔図書〕(計 0 件)

〔産業財産権〕

出願状況(計 0 件)

名称：
発明者：
権利者：
種類：
番号：
出願年月日：

国内外の別：

取得状況(計 0 件)

名称：
発明者：
権利者：
種類：
番号：
取得年月日：
国内外の別：

〔その他〕
特記事項なし

6. 研究組織

(1) 研究代表者

西村 邦宏 (NISHIMURA KUNIHIRO)
国立循環器病研究センター
循環器病統合情報センター・統計解析室長
研究者番号：70397834

(2) 研究分担者

(3) 連携研究者

(4) 研究協力者

危険因子	区分	配点	スコアの合計点	将来10年間に冠動脈疾患にかかる確率
年齢	35-44歳	30点	35点以下	1%未満
	45-54歳	38点		
	55-64歳	45点		
	65-69歳	51点		
	70歳以上	53点		
性別	女性	-7点	36-40点	1%
	喫煙	現在たばこを吸っている	5点	41-45点
糖尿病	ある	6点	46-50点	3%
	至適血圧	-7点		
血圧(mmHg)	正常血圧、正常高値血圧	0点	51-55点	5%
	Stage I 高血圧	4点		
	Stage II ~ IV 高血圧	6点	56-60点	9%
	~99	0点		
LDLコレステロール(mg/dl)	100~139	5点	61-65点	14%
	140~159	7点		
	160~179	10点	66-70点	22%
	180~	11点		
HDLコレステロール(mg/dl)	~39	0点	71点以上	28%以上
	40~59	-5点		
	60~	-6点		
CKDステージ	Stage 1または2	0点		
	Stage 3	3点		
	Stage 4または5	14点		

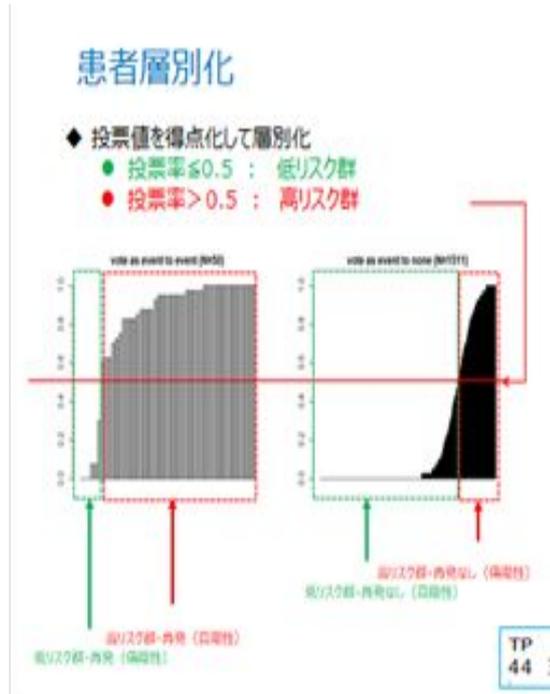


表1

図1

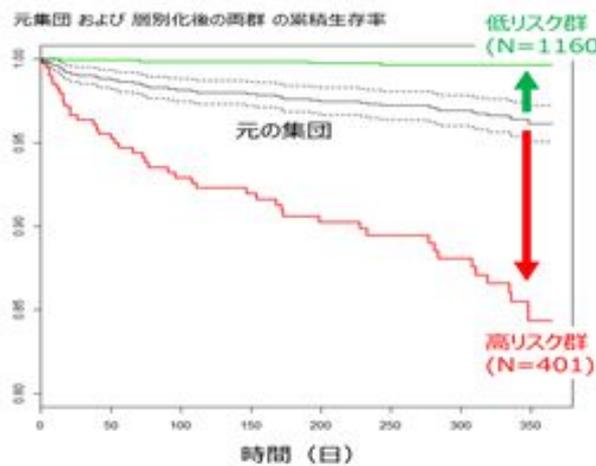


図2



図3

AMI患者への応用

17

- 今回のデータ取得対象者中AMI患者(N=225例)
- アウトカムは心臓死+MI再発+stroke
- Grace本来は心筋梗塞-院内死亡、一年後死亡予測

変数	HR(95%CI)	P-value
GRACE スコア	1.02(1.01,1.03)	<0.001
いびき (他覚的)	1.44(0.96,2.16)	0.078
胸痛 (放射痛)	0.34(0.10,1.16)	0.083
呼吸困難感	1.02(1.00,1.03)	0.019
浮腫	1.04(1.00,1.07)	0.042
無尿	1.15(1.00,1.32)	0.050
睡眠時無呼吸	1.04(1.01,1.07)	0.020

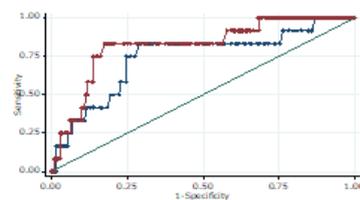


図4