

## 再帰型ニューラルネットワークモデルによる健康診断データ用いた心筋梗塞発症予測確率と地域健康教育への展開

Application of a resident's predicted probability of an acute myocardial event derived from the data of an annual medical examination using a recurrent neural network model to community health education

○戸田園乃<sup>\*1</sup>, 戸田真子<sup>\*2</sup>, 岡田英作<sup>\*3</sup>, 山内加奈子<sup>\*4</sup>, 加藤匡宏<sup>\*5</sup>

TODA Sono<sup>\*1</sup>, TODA Mako<sup>\*2</sup>, OKADA Eisaku<sup>\*3</sup>, YAMAUCHI Kanako, KATO Tadahiro<sup>\*5</sup>

<sup>\*1</sup>松山大学経済学部, <sup>\*2</sup>愛媛大学大学院連合農学研究科 <sup>\*3</sup>高野山大学文学部密教学科 <sup>\*4</sup>福山市立大学教育学部 <sup>\*5</sup>愛媛大学大学院教育学研究科

<sup>\*1</sup>Faculty of Faculty of Economics, Matsuyama University, <sup>\*2</sup>The United Graduate School of Agricultural Sciences, Ehime University, <sup>\*3</sup>Department of Esoteric Buddhism, Koyasan University, <sup>\*4</sup>Faculty of Education, Fukuyama City University, <sup>\*5</sup>Graduate School of Education, Ehime University

**【要約】** われわれは、脳卒中および心筋梗塞（脳心事故）の危険因子を特定するために、愛媛県の中規模市である O 市において 2 つのコホート集団を設立した。(1) 第 1 期 O 市コホート I（1996 年から 1998 年：5161 人）(2) 第 2 期 O 市コホート II（2010 年から 2012 年：3600 人）である。(1)(2)のコホート集団は脳心事故発生をエンドポイントとし、1996 年より約 27 年にわたり O 市基幹病院において脳心事故の疾病登録を実施した。また、脳心事故死亡も把握してきた。心筋梗塞(既)発症者データを用いて再帰型ニューラルネットワークモデル(RNN)を構築し、健診受診者の将来の心筋梗塞発症を予測した。O 市コホート I および II 参加者が発症したかどうかについて 2023 年以降に採録した脳心事故(新)発症者を比較検討し、O コホート研究によって RNN モデルと学習済み重みデータが実際に心筋梗塞を予測できるかどうかを明らかにした。

**【キーワード】** ニューラルネットワークモデル, 心筋梗塞発症予測, 地域健康教育, ヘルスプロモーション

### I. 問題の所在

われわれは、脳卒中および心筋梗塞（以下脳心事故と称する）の危険因子を特定するために愛媛県の中規模市である O 市（人口約 5 万人規模）において以下(1)(2)の 2 つのコホート集団を設立した。最初に (1) 第 1 期 O 市コホート I（1996 年から 1998 年：5161 人）においては、老人保健法に基づく成人病基本健康診査受診結果データと食生活習慣を含む脳心事故に関連する危険因子に関する問診等の健康関連項目を 3 年間で蓄積した。次に (2) 第 2 期 O 市コホート II 集団（2010 年から 2012 年：3600 人）においてはメタボリック症候群予防の特定健康診断結果データ、飲酒、喫煙、日常身体活動量、睡眠など様々な生活習慣の状況、首尾一貫感覚(Sense of coherence 以下 SOC と略す)を含む詳細問診項目、自律神経系機能検査、各種バイオマーカー等の脳心事故危険因子に関するベースラインデータを 3 年間で蓄積した。(1)(2)のコホート集団は脳心事故発生をエンドポイントとする調査であるために 1996 年より約 27 年にわたり、O 市基幹病院において脳心事故の疾病登録を実施し、脳心事故発生および死

亡の動向を把握している。1996 年 4 月 1 日から 2022 年 12 月 31 日までの脳心事故疾病登録データを採録しており、2 つのコホート集団のベースラインデータをマッチング解析することにより脳心事故発生の危険因子を明らかにしてきた(斉藤, 2007)(斉藤, 2022)。

地域における疾病予防活動では、リスクの大小にかかわらず集団全員を対象とした「ポピュレーション戦略」に力を入れ、特に教育や啓発を中心とした一次予防の活動を行なってきた。また、疾病発症リスクの高い集団には「ハイリスク戦略」として早期発見・早期治療という二次予防策を取ってきた。しかし、ハイリスク戦略は、すでにリスクが上昇している集団を対象とするために、被介入者が治療対象者になってしまうことが多く、予防活動となりにくいことが多い。

われわれは機械学習のプラットフォーム基盤の確立することによって O 市で設立した 2 つのコホート集団(1)O 市コホート I および (2) O 市コホート II のベースラインデータおよび(3)脳心事故発生の疾病登録情報を再帰型ニューラルネットワークモデル(RNN)モデルによるディープラーニング(Deep Learning 以下 DL

と略す)手法によって健診受診者一人ひとりひとりに脳心事故の発症予測確率を算出する試みを開始している。そこで、ポピュレーション戦略の教育や啓発として、過去に集積された脳心事故疾病登録データと健診データから RNN モデルを持って、健診受診者各個人に対して脳心事故の発症確率を提示することができれば、健診受診後に実施される保健指導のエビデンスを提供することができ、健診受診者が脳心事故発症のハイリスク者になる前に早期の積極的な介入することができる。本研究においては、脳心事故発症のうち心筋梗塞発症予測確率の正解率(精度)と損失率を求めることを目的とした。

## II. 研究の方法

### 1. 対象と期間

#### (1)倫理的配慮

本研究は、愛媛大学教育学部研究倫理委員会での承認(R4-34-1)を経て実施している。

#### (2)研究デザイン

われわれの研究デザインを図1に示す。

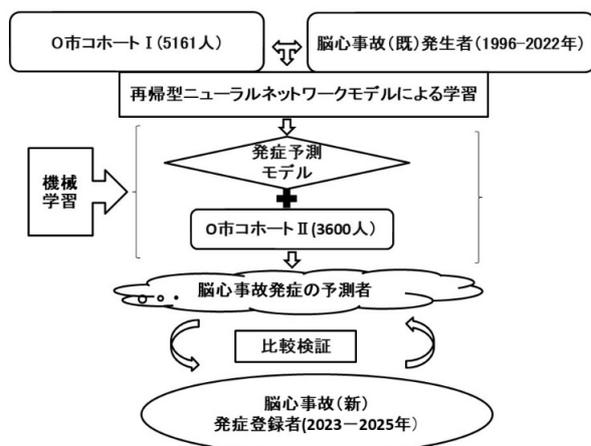


図1 研究デザイン

O市で設立した2つのコホート集団である(1)第1期O市コホートI(1996年から1998年:5161人)に蓄積された成人病基本健康診査データと循環器疾患に関連する危険因子等の健康関連データと(2)第2期O市コホートII集団(2010年から2012年:3600人)に蓄積されたメタボリック症候群予防のメタボリック症候群予防の特定健康診断結果データおよび喫煙、飲酒、日常身体活動量、睡眠など様々な生活習慣の状況やSOCを含む詳細な問診、TAS-9を用いた心電図のRR間隔スペクトラム解析から導出される自律神経系機能、各種バイオマーカーなど脳心事故発生に関連す

る健康関連データおよび(3)1996年からO市の主要病院で実施継続している脳心事故(既)発症者データを用いてRNNモデルを構築し、DLを用いて対象者の将来の脳心事故を予測した。O市コホートIおよびII参加者が発症したかどうかについて、2023年以降に採録した脳心事故(新)発症者を比較検討し、Oコホート研究によってRNNモデルと学習済み重みデータが実際に脳心事故を予測できるかどうかを明らかにする。

#### (3) 脳心事故発症の疾病登録

われわれは、1996年からO市においてポピュレーションベースで循環器疾患の疫学研究を推進してきた。O市の主要病院における脳心事故の発症登録を行い、WHO MONICA研究のプロトコールに基づいた医療情報について把握してきた。心筋梗塞に関しては、心筋梗塞の疾病登録は、発症までの臨床経過につき心筋梗塞の発症に合致するエピソードがあったかどうかを確認して採録した。すなわち、発症後の心電図、症状、血液データの値(白血球数、Aspartate aminotransferase (AST)、Lactate dehydrogenase (LDH)、Creatine kinase (CK)、Creatine kinase-MB (CK-MB)、トロポニンT)、心筋梗塞の既往歴を確認した。これらに加え、冠状動脈造影、超音波心臓検査の所見を採録から、心筋梗塞の新規発症者を採録した。加藤らそれら得られた医療情報から診断基準に基づき、心筋梗塞の確実・可能性例に再分類を行い、発症数を特定している(加藤, 2021)。

### 2. 分析方法

#### RNNモデル実装環境

上記(1)(2)(3)のデータを使用したRNNモデル作成のパイロットプログラムの実装環境を以下に示す。モデルの訓練と予測値の推定にWindows10、64ビット、CPU Core i7、メモリ16GBのスペックのコンピューターを用いた。モデルの実装はDL向けのライブラリであるTensorflow (Google)の仮想環境下において、Keras (Python2.6)、numpy、pandas、Matplotlib、Scikit-learningを用いた。anaconda2.4の環境下でJupyter notebookを用いてDL学習によるモデル構築と疾病発生予測を実施した。RNNのモデルは中間層1層のネットワーク(単層ニューラルネットワーク Single Layer Neural Networks:以下SLNNと略す)を用いた。入力層から隠れ層へのDenseは10、隠れ層から出力層へのDenseは2、エポック数は300、バッチサイズは20でモデルを構築し、訓練をおこなった。O市コホート集団

IとIIのデータから欠損値を除いた2088名のうち、2022年12月末までに心筋梗塞を発症した人33名、発症していない人2055名の健診データを用いて心筋梗塞予測モデルの作成を検討した。心筋梗塞発症の有無および健診データとしては、性別、年齢、身長、体重、BMI (Body Mass Index)、最大血圧 (収縮期血圧) 最小血圧 (拡張期血圧)、血液データのGOT, GPT,  $\gamma$  GTP, 総コレステロール, HDL コレステロール, 中性脂肪, , ヘモグロビン A1c, クレアチン, 尿酸,喫煙の有無の17項目を利用した。

### III. 結果

#### (1)SLNN モデル適正 (データ標準化なし)

第一作業として2022年12月末までに心筋梗塞を発症した人33名、発症していない人2055名の健診データを学習後に汎化性能を確かめるテストデータ(test)として切り分け、ニューラルネットワークの重みを更新する学習データ(train)とし、正解率(精度)と損失(図2と図3)からモデル適正を決定した。

正解率(精度)は正解率が0.98564, 損失率は0.0843であった。これら図2と図3は、データの標準化(前)の計算結果である。

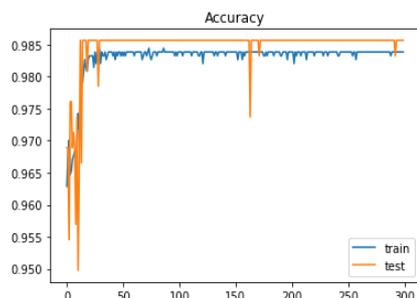


図2 標準化(前) 正解率(精度)

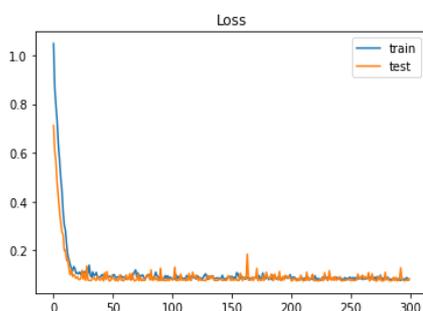


図3 標準化(前) 損失率

#### (2)SLNN モデル適正 (データ標準化あり)

機械学習においては、データ値の大小によって、

正解率と損失率が影響を受けることから、学習させるデータを0-100のパーセントに標準化させ、学習させることがなされる傾向にある。これら図3と図4は、データの標準化(後)の計算結果である

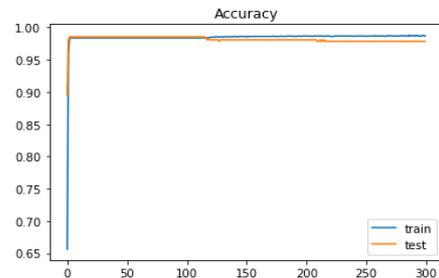


図3 標準化(後) 正解率(精度)

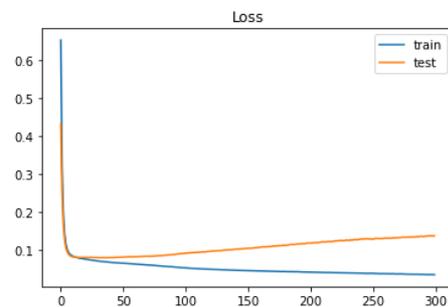


図4 標準化(後) 損失率

#### (3)心筋梗塞の発生の検証

第二作業は、第一作業のモデルと学習済み重みデータを用いて、観察集団(コホートII)において「実際に心筋梗塞を発症していない」住民の新規データ2421名について、predict関数で予測させたところ、発症する(yes)と予測した人は0人であり、発症しない(no)と予測した人は2421人(うち実際に発症しなかった人2415人)となり正解率は、 $2415 \div 2421$ で求められ0.99752となった。

#### (4)疾病登録の心筋梗塞による死亡者と厚生労働省の死亡小票との一致

1996年から2022年の27年間に、われわれの実施した0市コホートI(参加者5161人)および0市コホートII(参加者3600人)の心筋梗塞発症登録では、コホートIにおいて128人の発症、コホートIIでは、19人の発症を確認することができた。厚生労働省から入手した27年間の心臓疾患での死亡者数(ICDコードI20-25, I50, I00-02.0, I05-09, I30-52)は2899人であった。0市コホートIおよびコホートIIを併せた8161人の中で、厚生労働省の死亡票から確認した心筋梗塞による死亡者数

は50人であった。さらに、疾病登録において、心筋梗塞で死亡した人数は32人あり、心筋梗塞で死亡した参加者が厚生労働省の死亡票の病名として登録されている割合は64%であった。

#### IV. 考察

近年、機械学習・深層学習による電子カルテデータを用いた糖尿病の発症予測 (Thomas, 2018)や RNN を用いた複数年にわたる健康診断データを用いて健康診断で測定される測定値の予測が試みられている (Han-Gyu, 2017)。

健康診断が主な対象としている生活習慣病は、感染症とは異なり、発症までに時間がかかることから、健康診断データを用いて疾患を発症する前の段階である未病段階における疾患の進行について予測を行う機械学習による解釈可能な予測モデルの構築が必要となることが指摘されている (大場, 2020)。すなわち、O市コホート I と II のベースラインは脳心事故発生の危険因子を抽出するための詳細な健康診断であったとしても、研究参加者全員に脳心事故発症の危険因子が見つかるわけではない。O市コホート I と II 健康診断データセットは、健康な受診者と脳心事故発症の危険因子を持つ受診者の人数差が大きい不均衡なデータセットとなっている。

本研究においては、少しでも脳心事故発症の高い予測確率を得るために健康診断の間診項目のように脳心事故を発症したかどうかを聞き取る調査ではなく、主要基幹病院における脳心事故の疾病登録データを用いた信頼性の高い医療情報を用いてエンドポイントを決定している。

特に、われわれが使用した脳心事故発症の疾病登録結果をエンドポイントとして使用する手法は、従来の疫学指標として一般的に使用されている脳心事故の危険因子のトレンド検定、相対危険度と集団寄与割合を算出することができ、従来の疫学指標を用いて RNN による脳心事故の発症予測確率を検証可能な地域はごく限られており学術的にも有意義と考えられる。本研究の特徴は O市コホート I と II の研究参加者の脳心事故発症予測確率を算出するとともに、2023年以降に本当に脳心事故を発生したかどうかについては、脳心事故新発症者発症登録と突合せることによって脳心事故発症予測確率の確度を検証することができる。

一方、SLNN モデルから算出された脳心事故の発症予測確率を一般の健康診断受診者の健診結果に提示することは、疾病予防のポピュレーション戦略における健康の保持増進に説得力を持たせるとともに健診受診者の疾病予防の動機づけを高めることができると考えている。

O市コホート I および II 研究から得られる脳心事故発生の危険因子に関する従来の疫学指標と SLNN モデルによる脳心事故発症確率とを比較することによって、社会医学研究における AI の位置づけ、課題を明らかにし、今後社会医学における AI 研究の発展に寄与できると考える。

#### 疾病登録の限界

われわれの脳心事故の疾病登録に以下の限界がある。われわれは、脳心事故が原因となる病院到着時死亡や在宅死亡について詳細な疾病情報を得ることはできていない。また、救急搬送情報については、個人情報保護の観点から、救急隊の情報を入手することはできない。O市 Cohort I および II の参加者については脳心事故が原因となる死亡原因は、愛媛県 Y 保健所にある死亡票で死因の確認が可能である。しかしながら、O市 Cohort I および II の参加者以外の住民が脳卒中および心筋梗塞で死亡した場合については、個人情報保護の観点から、保健所にある死亡票で氏名の確認をすることができない。また、O市 Cohort I および II の参加者であっても、脳心事故が原因となり、寝たきり状態から肺炎によって死亡した場合は、死因が変更されている可能性がある。O市 Cohort I および II の参加者が脳心事故によって死亡した場合は、Y 保健所での死亡票と照合し、疾病登録の精度を管理してきた。しかしながら、訪問診療による在宅死亡もしくは警察官による検視による急性心不全や心臓内因死という死因の中に、急性心筋梗塞による不整脈死や心筋破裂が内在しており、即死する急性心筋梗塞における疾病登録の悉皆性を担保していることを確認できていない。また、救急搬送については、医師が同乗すれば O市以外に搬送することが可能であるため、O市以外に搬送された生存者数は不明であるため疾病登録の実際の精度は不明である。

#### V. おわりに

地域健康教育への展開

われわれの生活を脅かす脳心事故を予防するため

には、地域における健康づくりを推進していかねばならない。集団全体に働きかける方法をポピュレーションアプローチと呼ばれ、疾患発症の予防効果が期待される(厚生労働省, 総論)。そこで重要となってくるのが効果的な保健指導の実施である。大森(2012)は、保健指導が(1)病気や療養行動に関する知識の向上(2)セルフマネジメント行動の習得を可能となることを報告している。冠動脈の狭窄による別部位での心筋虚血による心筋梗塞の再発や陳旧性心筋梗塞が原因となる心不全の予防に貢献すると考えている。われわれが作成したエビデンスに基づいた脳心事故発生予測モデルを健診に用いることで、疾病予防のポピュレーションアプローチに新たな展開を提示することができる。今後、蓄積された健康診断データ、健康関連データから脳心事故発症について精度の高い予測モデルを開発するために必要な学習データ項目を明らかにすることによって地域のヘルスプロモーションに新しい切り口での疾病予防政策を提示することができ、公衆衛生学的にも重要な方略となる。

#### 付記・謝辞

本研究におけるベースライン調査ならびにコホート研究の基盤は、これまで以下の文部科学省科学研究費補助金により構築してきたを報告する。(基盤C(2008-2010) 研究代表者: 斉藤功「自律神経系機能と耐糖能異常に関する疫学研究」(課題番号 20590647); 基盤C(2010-2012) 研究代表者: 加藤匡宏「自律神経系機能と循環器疾患とのコホート研究」(課題番号 22590587); 基盤C(2014-2016) 研究代表者: 斉藤功「自律神経系機能の低下と糖尿病発症に関するコホート研究」(課題番号 26460767); 基盤C(2017-2021) 研究代表者: 加藤匡宏「首尾一貫感覚(SOC)と循環器疾患発症および総死亡・死因別死亡のコホート研究」(課題番号 17K09204)。研究代表者: 加藤匡宏「深層学習を用いた脳心事故発生予測モデルの構築と検証に関する研究」(課題番号 22K10557); 基盤C(2022-2027)からの研究助成を得ている。研究の一部はJSPS 科研費 JP12345678 の助成を受けている。

#### 文献

大場勇貴, 手塚太郎, 讃岐勝, ほか1名(2020) 健康診断データを用いた疾患予測における解釈可能なモデルの構築 DEIM2020 I2-2(day1) p58.

大森豊緑(2012) <https://mhlw-grants.niph.go.jp/project/18151> (2025年1月19日確認)

加藤匡宏, 丸山広達, 斉藤功(2021) 愛媛県0市における25年間の脳卒中および虚血性心疾患の疾病登録調査 愛媛大学教育実践センター紀要 39巻, 23-33.

厚生労働省, 総論

[https://www.mhlw.go.jp/www1/topics/kenko21\\_11/pdf/s0.pdf](https://www.mhlw.go.jp/www1/topics/kenko21_11/pdf/s0.pdf) (2025年1月19日確認)

Han-Gyu Kim, Gil-Jin Jang, Ho-Jin Choi, (et al.) (2017) Recurrent neural networks with missing information imputation for medical examination data prediction. 2017 IEEE International Conference on Big Data and Smart Computing (BigComp) Jeju, 2017-02-13/16, IEEE, 2017, 317-323.

斉藤功, 小西正光, 渡部和子, ほか3名(2007) 地域集団におけるメタボリックシンドロームの脳卒中罹患に及ぼす影響について 日本公衆衛生雑誌 54巻, 10号, 677-683.

斉藤功, 山内加奈子, 山泉雅光, ほか1名(2022) メタボリックシンドロームと脳卒中罹患: 18.6年間のコホート研究 日本公衆衛生雑誌 69巻, 5号, 394-402.

Thomas Garske(2018) Using deep learning on EHR data to predict diabetes. University of Colorado at Denver. 2018, Ph.D. thesis