

再帰型ニューラルネットワークモデルを用いた脳卒中発症予測の検証

Validation of prediction probability of an occurrence of stroke using recurrent neural network model

○戸田真子^{*1}, 戸田園乃^{*2}, 丸山広達^{*3}, 加藤匡宏^{*4}

TODA Mako^{*1}, TODA Sono^{*2}, MARUYAMA Kotatsu^{*3}, KATO Tadahiro^{*4}

^{*1}愛媛大学大学院連合農学研究科, ^{*2}愛媛大学大学院医農融合公衆衛生学環, ^{*3}愛媛大学大学院農学研究科, ^{*4}愛媛大学大学院教育学研究科

^{*1}The United Graduate School of Agricultural Sciences, Ehime University, ^{*2}Integrated Medical and Agricultural School of Public Health, Ehime University, ^{*3}Graduate School of Agriculture, Ehime University,

^{*4}Graduate School of Education, Ehime University

[要約] われわれは、愛媛県 O 市における脳心事故発症調査を実施するために、1996 年～1998 年に健康診断に参加し同意が得られた 5161 人をベースライン集団とする O 市コホート I を設立した。ベースライン集団については、毎年、脳心事故発症および異動状況を追跡している。さらに、O 市の基幹病院における脳心事故発症者を採録することで、脳卒中の発症を把握した。脳心事故採録調査は、O 市の基幹病院で実施されているため、脳卒中の発症漏れが少なく、極めて信頼性の高いデータであると考えられる。本稿では、健診項目に欠損値のない 5021 人を対象として、機械学習を用いて脳卒中発症者と健常者の予測を行った。さらに、脳心事故採録に基づく実際の脳卒中発症者と突合させることによって、機械学習による脳卒中発症予測の感度および特異度を検証することを目的とした。

[キーワード] ニューラルネットワークモデル, 脳卒中発症予測, 健康診断, 脳心事故疾病登録

I. 問題の所在

近年、様々な分野で AI が導入され、機械学習への期待が高まっている。疫学分野においても、機械学習を活用した新たな取り組みが進められている。高齢化が進行する日本では、疾患の早期発見・重症化予防による健康寿命の増進が求められている。特に、脳卒中や心筋梗塞等の疾患は、生活習慣から起因していると考えられ、その予防に向けた画策が立てられている。

現在、日本では生活習慣病の予防を目的として、特定健康診査および特定保健指導が実施されている。包國・麻原 (2013) は、生活習慣改善の方法を共に創るためには、生活習慣病予防を自分のこととして考えられるようにすることが重要なプロセスの 1 つであることを報告している。そのため、生活習慣病予防の重要性の理解を得るために、疾病の発症リスクを高い精度で予測し、その結果を基に保健指導を実施することが必要であると考えられる。

われわれは、愛媛県 O 市における脳心事故発症調査を目的として、1996 年～1998 年に健康診断を受診した 5161 人をベースライン集団とする「O 市コホート I」を設立した。5161 人のベースライン集団に関しては、同意を得ているため、毎年、脳心事故発症および

異動状況を追跡することが可能である。また、脳心事故発症を確定するために、O 市の基幹病院において採録調査を実施している。基幹病院での採録調査によって、脳心事故の発症漏れが少なく、極めて信頼性の高いデータであると考えられる。

本稿では、ベースライン集団 5161 人のうち、健診項目に欠損値のない 5021 人のデータを用いて、脳卒中の発症の有無を予測する。そして、予測結果と脳心事故発症採録に基づく実際の発症者を突合し、機械学習による脳卒中発症予測の感度と特異度を調べることを目的とした。

II. 研究の方法

1. 倫理的配慮

O 市における脳心事故発症登録は、愛媛大学医学部研究倫理委員会 (承認番号 14-1, 20-1) および愛媛大学教育学部研究倫理委員会 (承認番号 R4-34-1) の承認を得て実施した。また、疾病登録を実施する各病院については疾病登録データの利用に関するオプトアウトを実施している。研究参加者については、脳心事故発症に関する疾病登録データおよび転出・死亡に関する異動情報の利用について、書面による同意を得てい

る。

2. 研究デザイン

1996年～1998年にO市コホートI調査に参加し、老人保健法に基づく成人病基本健康診査を受診した5161人において、性別、年齢、身長、体重、血圧、血清総コレステロール、HDLコレステロール、中性脂肪（随時）、血糖値（随時）、ウエスト周囲長（臍周囲）、飲酒状況、喫煙状況を確認することが可能であった。

また、われわれは住所地在O市にあり、外傷性脳血管障害を除く入院患者を対象に所定の調査票(山海他, 1992)を用いて、1996年から2023年までの27年間にわたり脳心事故発症登録を実施してきた。そのため、O市コホートIの研究参加者5161人における脳卒中の発症を確認することができる。

3. 機械学習による脳卒中発症者の予測

深層学習は、層の数が多い階層的なニューラルネットワークによってデータから抽象度の高い内部表現を獲得させる方法として知られている(麻生, 2013)。ニューラルネットワークは、生体の神経細胞を模した人工ニューロンを基本構成要素とし、生体の神経回路網と同じように人工ニューロンがシナプスで結合したネットワーク構造を作った計算アルゴリズムである(神田, 2018)。ニューラルネットワークにおける「学習」とは、重み係数(シナプス結合の強さ)やバイアス(活動電位の発火閾値)を最適化することを示す(神田, 2018)。

本稿では、単層型ニューラルネットワークを用いて、脳卒中発症者の予測確率を算出した。研究の流れを示した模式図・手順を以下に示す。



図1 研究の流れ

a. 環境設定

モデルの訓練と予測値の推定にWindows10, 64ビット, CPUCorei7, メモリ16GBのスペックのコンピュータを使用した。モデルの実装はDeep Learning (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。Anaconda 2.4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。再帰型ニューラルネットワーク (Recurrent Neural Network; RNN) モデルとしては、中間層1層からなる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, Batch Sizeは20, Epoch数は300としてモデルを構築し、学習を行った。

b. データセットの準備

本稿では5161人のデータセットから欠損値を除いた5021人を対象とした。27年間にわたる疾病登録の結果、5021人のうち、脳卒中発症者は461人（くも膜下出血33人、脳梗塞341人、脳出血87人）であり、発症割合は9.2%であった。5021人の性年齢別階級別の分布を表1に示す。

	参加者(N=5021)	
	男(N=1723)	女(N=3298)
30～39	109 (6.3)	364 (11.0)
40～49	195 (11.3)	442 (13.4)
50～59	209 (12.1)	628 (19.0)
60～69	667 (38.7)	1098 (33.3)
70～79	442 (25.7)	651 (19.7)
80～89	97 (5.6)	114 (3.5)
90～	4 (0.2)	1 (0.0)

表1 5021人の性年齢階級

5021人のデータセットから、ランダムに抽出した1000人の健診データを用いて予測モデルの構築を行った。さらに、1000人のうち、脳卒中発症者は92人(9.2%)となるように設定した。

健診データは、性別、年齢、BMI (Body Mass Index), 最大血圧, 最小血圧, 血液データの総コレステロール, HDLコレステロール, 中性脂肪, 血糖, 飲酒, 喫煙である。それぞれの項目に関しては、特定保健指導の区分(異常なし: 0, 要指導 a: 1, 要

指導 b : 2, 要医療 : 3) に準じてダミー変数を作成した。

Python code では, `train_size=0.8` と設定し, 1000 人の健診データをランダムに選択したうえで, 800 人 (80%) の健診データは学習用データ (train), 200 人 (20%) をテスト用データ (test) に分割した。すなわち, 研究参加者のデータをランダムに並べて最初の 800 人を用いて SLNN の重みを更新し, 残りの 200 人を用いて汎化性能を評価した。train と test の一致率に基づき, モデルの適合度を決定した。

c. 単層ニューラルネットワークを用いたアルゴリズムの決定

Batch Size を 20 と設定したため, 1 回に処理するサンプル数は 20 である。Epochs は 300 と設定したため, 学習の繰り返し回数は 300 回である。Batch Size 20 とは, 機械学習におけるミニバッチ学習を設定し, 1 回の SLNN 更新に使われるデータ数が 20 であることを意味する。すなわち, SLNN の訓練に使用する 800 人の健診データすべてを一度に処理するのではなく, 800 人の健診データセットからランダムに抽出された 20 人分の健診データを使って SLNN を訓練し, その後パラメーターを更新する。

このプロセスを繰り返しながら, 最終的にデータ全体 (ランダム選択による train 用健診データ 800 人, test 用健診データ 200 人) を活用した SLNN において, Accuracy は 0.90, loss function は 0.34 であった。

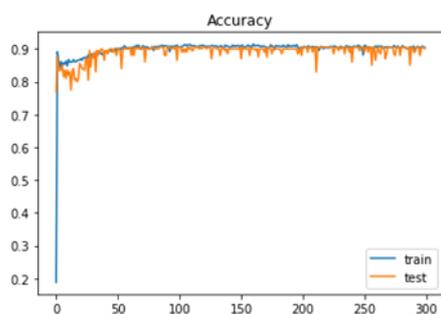


図2 正確率

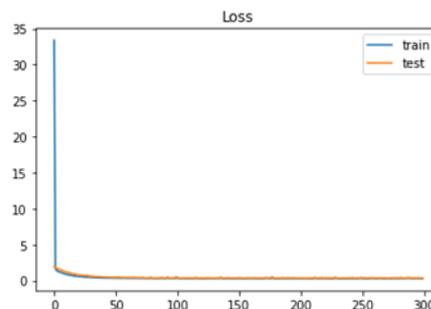


図3 損失率

d. 実測値と比較することで得られる予測値の適合度の決定

c. で構築したアルゴリズムを用いて, 5021 人から 1000 人を除いた 4021 人に対して脳卒中発症予測を行った。SLNN から得られた脳卒中発症予測者と脳心事故疾病採録調査に基づく実際の脳卒中発症者を比較した。

III. 結果

本稿において, 5021 人のうちランダムに抽出した 1000 人を用いて SLNN のパラメーターを学習させ, その後, 残りの 4021 人に学習済み SLNN を適用し, 個人の脳卒中発症予測確率 (発症の有無) を計算した。

学習済み SLNN による予測では, 4021 人のうち 61 人が脳卒中を発症すると予測された。予測された 61 人について脳心事故発症の疾病登録情報と照合したところ, 8 人が実際に脳卒中を発症していた。これより, SLNN の感度は 13.1% であった。

一方, 学習済み SLNN は 4021 人のうち 3960 人を「脳卒中を発症しない」と予測した。疾病登録情報と比較した結果, 3960 人のうち 3599 人が実際に脳卒中を発症していなかった。これより, 特異度は 90.9% であった。

なお, 4021 人における実際の脳卒中発症者は 369 人であり, SLNN はそのうち 361 人の脳卒中は予測できなかった。

IV. 考察

現代の日本では, 健康寿命の延伸が重要課題として掲げられ, 疾病予防・重症化予防の取り組みが強化されている (厚生労働省, 2020)。その中で, 生活習慣から起因すると考えられる脳卒中の発症予防の構築は極めて重要である。近年, 健康診断データをもとに, 機

械学習を用いた疾病の発症予測に関する研究が報告されており (Tsunekawa M, et al. 2019 ; 大場他, 2020 ; 道庭他, 2022), 発症予防に着目した取り組みに期待されていることが明白である。

われわれは, 脳卒中発症を予測し, その結果を特定保健指導に活用することで地域の健康づくり活動の推進に寄与できる可能性を検討してきた (戸田他, 2025). 本稿の SLNN の感度は 13.1%, 特異度は 90.9% であり, 感度が低い結果となった. しかし, 今後, パラメーター決定に用いる変数の追加やモデル構造の改善を行うことで, 感度向上の余地があると考えられる. 一方で, SLNN による脳卒中発症の予測は脳卒中のリスク因子と脳卒中発症との因果関係を考慮せず, パラメーターとノード数のみで判定する手法である. そのため, エビデンスを基盤とした疾病予測モデルではない. 因果関係の解明には, 従来のコホート研究 (追跡調査) を実施し, Cox 比例ハザードモデルを用いた解析が不可欠である.

日本には, 生活習慣病における危険因子を明らかにし, 日本人の生活習慣病予防と健康寿命の延伸に資することを目的とした国立がん研究センターによる多目的コホート研究 (JPHC 研究) が存在する. Yatsuya et al. (2013) は, JPHC 研究の追跡調査結果を用いて, リスクスコアによる脳卒中発症確率を算出し, 発症予測モデルを開発している. 本稿では, Python を用いた SLNN によって脳卒中発症を予測したが, Yatsuya et al. のように, コホート研究でのリスク評価を取り入れたモデルこそが, 医学的エビデンスの高い発症確率が算出できると考えられる.

脳卒中発症を予防するためには, 日常的に健康意識を高め, 規則正しい生活をするのが重要である. 脳卒中発症予測モデルによって算出された脳卒中発症確率を特定保健指導の場で提示することは, 健康意識向上の動機付けとして有効である可能性がある.

今後は, 蓄積された健康診断データや健康関連データを活用し, より精度の高い脳卒中発症予測モデルを開発することで, 地域のヘルスプロモーションに新たな視点から疾病予防政策を提示できると考える.

V. おわりに

O 市コホート I 参加者 5021 人のうち脳卒中発症者 92 人を含む 1000 人を用いて SLNN の学習を行い, 最適化したパラメーターを実装した学習済み SLNN

により, 残りの 4021 人の脳卒中発症予測を実行した. その結果, SLNN の感度は 13.1%, 特異度は 90.9% であった.

付記・謝辞

本稿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, これまでに文部科学省科学研究費補助金により構築してきた ((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)). さらに, 研究の一部は JSPS 科研費 JP12345678 の助成を受けている.

文献

- 麻生 英樹 (2013). 多層ニューラルネットワークによる深層表現の学習, 人工知能学会誌, 28 (4), 649–659.
- 道庭 賢一・田中 孝浩・春木 耕祐 (2022). 健康診断データに基づく機械学習による疾病リスク予測 AI 及び生活習慣改善 AI, 東芝レビュー, 77 (1), 48–51.
- 神田 寛行 (2018). Deep learning (深層学習) の基礎知識と, 医療応用に向けた課題, 視覚の科学, 39 (3), 75–79.
- 包國 幸代・麻原 きよみ (2013). 対象者中心の保健指導を実践する保健師の技術, 日本看護科学会誌, 33 (1), 71–80.
- 厚生労働省 (2020). 予防・健康づくりについて https://www.mhlw.go.jp/content/12401000/00061286_2.pdf (2026 年 3 月 3 日閲覧)
- 大場 勇貴・手塚 太郎・讃岐 勝也 (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, 第 12 回データ工学と情報マネジメ

ントに関するフォーラム(DEIM2020)予稿集.

山海 知子・磯 博康・嶋本 喬他 (1992). CT 所見を中心とした脳卒中の疫学的研究--コホート内症例対照研究による脳出血ならびに脳梗塞の病型別発生要因の検討, 日本公衆衛生雑誌, 7, 410-420.

戸田 園乃・戸田 真子・岡田 英作他 (2025). 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を用いた脳心事故発症予測確率と地域健康教育への展開, 科学教育研究センター紀要, 4, 39-43.

Tsunekawa M, Oka N, Araki M, et al. (2019). Prediction of the Onset of Lifestyle-related Diseases Using Regular Health Checkup Data, The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019, 1-4.

Yatsuya H, Iso H, Yamagishi K, et al. (2013). Development of a Point-based Prediction Model for the Incidence of Total Stroke Japan Public Health Center Study, Stroke, 44(5), 1295-1302.