短 報

学習型 AI を用いた労働者の健康診断結果の予測：ある企業の労働者におけるパイロット研究

Application of artificial intelligence for the prediction of health examination results of workers: A pilot study

内田 満夫¹, 筒井 保博², 野下 浩司³, 小山 洋¹¹群馬大学大学院医学系研究科公衆衛生学分野²福岡産業保健総合支援センター³九州大学理学研究院生物科学部門数理生物科学研究室Mitsuo UCHIDA¹, Yasuhiro TSUTSUI²,
Koji NOSHITA³ and Hiroshi KOYAMA¹¹Department of Public Health, Graduate School of Medicine,
Gunma University²Fukuoka Occupational Health Support Center³Mathematical Biology Laboratory, Department of Biology, Kyushu
University(産衛誌 2021; 63(3): 95–98)
doi: 10.1539/sangyoeisei.2020-031-Cキーワード: Artificial intelligence, Deep learning, Recurrent
neural network, Long- and short-term memory

目 的

わが国は、平均寿命の延伸と共に出生数が減少し、どの国も経験したことのない少子高齢社会となった¹⁾。総務省の報告では、わが国の総人口は2008年に1億2,800万人とピークに達したが2050年には1億200万人程度まで減

少し、また65歳以上人口の割合は現在の28.9%から37.7%に増加すると予想されている²⁾。この人口構造の変化は産業現場にも影響を及ぼし、わが国では定年退職の延長³⁾などにより高齢者の雇用を促進して労働力の損失や社会保障費の増加を抑制する取り組みが進められている⁴⁾。しかし、わが国の労働者の健康診断所見割合は10年連続で50%を超え、また脳・心臓疾患による労災請求も増加し続けている⁵⁾。高齢労働者が増加すると、治療と仕事の両立⁶⁾や、心身の衰えによる労災を予防するための健康管理がさらに重要となってくる。以上より、産業保健領域においては、労働者の疾病予防策を検討し、長く健康に働く方法を提案することが求められている。

予防医学の観点からは、労働者の健康管理は1次予防から3次予防まで段階的に取り組むことが可能である。まず、リスクの大小にかかわらず集団全員を対象とした疾病予防策である「ポピュレーション戦略」があり、教育や啓発活動を中心とした1次予防がおこなわれる⁷⁾。しかし、1次予防が達成できず疾病リスクが上昇した場合、健康診断でそのリスクを早期に発見し、保健指導や治療などの対策、つまり2次予防に相当する「ハイリスク戦略」がおこなわれる⁷⁾。これらの方法で予防しきれなかった場合は疾病発症に至り、その後の社会復帰や再発予防という3次予防が行われる。したがって、職域で高齢者が著しく増加する近い将来、これらの予防活動はますます重要になると考えられる。もし健康状態の将来予測が可能となれば、労働者が早い段階でこれらの予防策を講じる動機づけの手段に活用できる可能性がある。

筆者らは、産業保健の領域において、学習型 AI の主要構成技術である“ディープラーニング (DL: deep learning)”の技術を展開し、職場巡視⁸⁾や情報機器作業者の作業管理のサポート⁹⁾に活用できるか、その可能性を示してきた。本研究では、蓄積された健康診断データから再帰型ニューラルネットワークのモデルを構築し、DLと教師あり学習に基づく学習型 AI を用いて対象の将来の健康診断結果を予測する手法、つまり上記で言うところの2次予防の方法を提案することを目的とした。このたびある企業1社を対象にパイロット研究をおこない、その可能性について検討した。

方 法

健康データの収集

本研究では、群馬県内にある運輸業のA社（従業員数490人）より協力を得た。A社からは、2014年から2018年までの5年分の健康診断データを得た。データを受け取る際には、企業側の担当者により、解析に用いない氏名などの個人情報とは全て削除してもらい、データと個人が連結できない状態にしてから研究者に送付してもらった。

2020年6月29日受付；2020年8月7日受理

J-STAGE 早期公開日：2020年8月26日

連絡先：内田満夫 〒371-8511 群馬県前橋市昭和町3-39-22
群馬大学大学院医学系研究科公衆衛生学分野Correspondence to: Mitsuo Uchida, Department of Public Health,
Graduate School of Medicine, Gunma University, 3-39-22, Showamachi,
Maebashi City, Gunma Prefecture 371-8511, Japan
(e-mail: uchida01@gunma-u.ac.jp)

健康診断項目は、対象企業で管理している項目のうち、数値データとして5年分を取り扱うことのできるBMI、腹囲、収縮期血圧、拡張期血圧、HDL コレステロール、LDL コレステロール、中性脂肪、GOT、GPT、 γ GTP、ヘモグロビン、赤血球の12項目を用いた。今回の研究では、調査期間中に同じ健康診断業者により検査を行っている対象を抽出したため、健診機関や検査機器の相違による影響はないと見なした。

DL の方法

今回の研究では、蓄積された過去情報を学習して将来データを予測するRNN (Recurrent Neural Network) モデルのひとつであるLSTM (Long-Short Term Memory) を採用した。このモデルは時系列データでニューラルネットワークを構築し、データの並び方を学習しながら過去データにもフィードバックし、RNNの弱点であった過去データの重みづけも含めた手法であり、時系列分析や自然言語処理に用いられる¹⁰⁾。本研究では、入手したデータから欠損値を除き、12項目全てのデータを利用できる対象を抽出した結果、合計174人を研究対象とした。まず初めに、作成するモデルの精度を後から評価するために、最初の10名を“テストデータ”として取り除いた。残りの164人は“訓練データ”とし、モデルを構築するために用いた。時系列データを用いて、過去4年分の健康診断結果を学習し、5年目を教師ありデータとして予測するモデルを作成し、予測値と観測値との一致割合を調べた。本研究では、モデルの訓練と予測値の推定にWindows 10、64ビット、CPUはIntel 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、Keras、Scikit learnを用いた。LSTMはモデルの精度の高くなるパラメータの値をトライアンドエラーで探索した。今回のモデルでは、データに活性化関数をあてはめて順番に次の層に渡していく中間層は256、その中間層の1層あたりのデータを分割するノード数は128、モデルの精度を上げるために計算を繰り返すエポック数は300、計算を高速化させるためグループ分けを行うバッチサイズは32という条件で訓練をおこなった。健康診断結果は、12項目すべての数値データの予測値を算出した。はじめに、訓練データで作成したモデルで訓練データ自身の将来予測をおこない、その精度を評価した。次にこのモデルで、予め取り除いておいた10人のテストデータの将来予測をおこない、モデルの一般化性能を評価した。数値データを完全に一致させることは困難であるため、観測値に対して予測値が $\pm 10\%$ 範囲内 (10%誤差)、および $\pm 20\%$ 範囲内 (20%誤差)の値を得た場合の正答割合を評価した。全てのモデルにおいて、予測精度が1.0 (全ての的中) に近いほど精度の高いモデルとあるとみなした。また20%誤差の場合、

正当割合がかなり高くなると予想されたため、最終的な評価値は10%誤差の方を採用した。

次に、作成モデルによるテストデータの予測精度の検証のためk-fold cross-validation法を用いた¹¹⁾。これは、サンプルからテストデータを順番に抽出してモデルを複数回作成し、それぞれ作成したモデルの予測精度を平均して、そのモデルの性能を確認する方法である。DLの研究では、モデル精度を評価する方法としてしばしば用いられている。今回は174人から10人をテストデータとして予め抽出し、残り164人で訓練して予測をおこなう方法を採用したため、10名ずつをテストデータとして切り分ける (モデル1はNo. 1からNo. 10をテストデータに、モデル2はNo. 11からNo. 20をテストデータに、以降繰り返す) ことを合計16回おこない、最終的には16回分のモデル精度の平均値 $\pm$ 標準誤差で表した。

研究倫理

本研究では、情報提供元の企業が個人情報の削除をおこなってから数値データのみを提供したため、個人を特定することができず、研究対象個別に対するインフォームドコンセントは取得しなかった。また、研究開始後のオプトアウトの処理ができない状態であったため、調査研究にデータを使用してほしくない対象がいた場合、研究開始時に公開する文書においてアナウンスを行い、申し出のあった人のデータを企業側担当者が除外することとした。本研究内容は群馬大学の医学研究倫理審査委員会において審査され承認された (HS2019-092)。

結 果

LSTMで健康診断項目の将来予測を行った結果、収縮期血圧を例にとると、まず訓練データのモデルの予測精度は10%誤差レベルで1.0 (20%誤差レベルも1.0)、つまり全て予測的中させた。続いて予め取り除いておいたテストデータを用いて予測精度を調べると10%誤差レベルで0.6 (20%誤差レベルで0.9) だった (図1)。さらにモデルの性能を検証するため、16回分のcross-validationをおこなった結果、10%誤差レベルで 0.64 ± 0.03 (平均 $\pm$ 標準誤差) (20%誤差レベルで 0.93 ± 0.03) だった (表1)。同様に健康診断の12項目すべての検証をおこなった結果、10%誤差レベルでの予測精度は、最も低いものは 0.10 ± 0.02 (γ GTP)、最も高いものは 0.94 ± 0.02 (赤血球) だった (表2)。

考察と結論

本研究では、164人の健康診断の情報 (BMI、腹囲、収縮期血圧、拡張期血圧、HDL コレステロール、LDL コレ

図 1. テストデータの観測値と予測値

収縮期血圧を例に、あらかじめ抽出したテストデータ10人の観測値（左バー）と予測値（右バー）を示した。1つ目のモデルは、10%誤差レベルの予測精度は0.60、20%誤差レベルの予測精度は0.90だった。

表 1. 収縮期血圧の k-fold cross-validation による精度の検証
対象174人から10人ずつテストデータに抽出して合計16回モデルを作成した結果、予測精度の平均値は10%誤差レベルで 0.64 ± 0.03 、20%誤差レベルで 0.93 ± 0.03 だった。

モデル	10% 誤差	20% 誤差
1	0.60	0.90
2	0.50	0.90
3	0.80	1.00
4	0.70	0.90
5	0.50	0.90
6	0.60	0.90
7	0.70	1.00
8	0.70	1.00
9	0.70	0.90
10	0.50	0.60
11	0.60	1.00
12	0.50	1.00
13	0.50	0.90
14	0.90	1.00
15	0.60	0.90
16	0.80	1.00
平均	0.64	0.93
標準誤差	0.03	0.03

ステロール、中性脂肪、GOT、GPT、 γ GTP、ヘモグロビン、赤血球）を用いて LSTM モデルを作成し、テストデータである10人の健康診断結果の将来予測をおこなった。その結果、10%誤差レベルでは約1割から9割の精度で予測することができた。この結果から、変動の大きい脂質や肝機能の血液データは予測が難しいが、一方で変動が少ないと考えられる BMI や赤血球のデータは9割以上の精度で予測できていた。今後、サンプルやデータ

表 2. 健診12項目の予測精度

健康診断の各項目の予測精度を検証した（平均値 $\pm$ 標準誤差）。精度は10%誤差レベルで0.10から0.94と幅を認めた。

項目	10% 誤差レベル	20% 誤差レベル
BMI	0.92 ± 0.02	0.99 ± 0.01
腹囲	0.89 ± 0.02	1.00 ± 0.00
収縮期血圧	0.64 ± 0.03	0.93 ± 0.03
拡張期血圧	0.56 ± 0.04	0.90 ± 0.02
HDL コレステロール	0.48 ± 0.05	0.82 ± 0.03
LDL コレステロール	0.44 ± 0.02	0.78 ± 0.02
中性脂肪	0.17 ± 0.03	0.33 ± 0.04
GOT	0.33 ± 0.04	0.59 ± 0.05
GPT	0.19 ± 0.03	0.33 ± 0.04
γ GTP	0.10 ± 0.02	0.28 ± 0.03
ヘモグロビン	0.93 ± 0.02	0.99 ± 0.01
赤血球	0.94 ± 0.02	0.99 ± 0.01

ポイントなどのデータ数を増加させることで予測精度の向上が期待され、学習型 AI を用いた健康診断結果の予測研究はさらに進展する可能性がある。また、本手法が確立されると、労働者が自ら健康管理をおこなうための支援ツールとなるかもしれない。

これまでの時系列データを取り扱う医学研究は、大気中有害物質と呼吸器系疾患との関連を評価する研究等が知られている¹²⁾。これらの研究では、有害物質の変動が時系列データとして利用され、主に回帰分析を用いて疾病発症との関連が評価されている。ただし生体データを時系列データとして取り扱っているのではなく、本研究が目的とした生体データの変動については議論されていない。ヒトの生体データを時系列情報として評価したものでは、データマイニングの手法を用いて生体データの

過去の変動の説明を試みた研究もある¹³⁾。この研究では生体データを時系列データとして取り扱っている点は共通するが、生体データ自身の将来予測を目的とはしていない。本研究では生体データの変動をDLで評価し、その将来予測を目的とした。つまり本研究では生体データ自身を予測することを目的としており、産業保健現場における実務への応用を検討している点に新規性があると考えられる。

将来予測を行う研究においてはしばしば回帰分析が用いられる。しかし予測する変数は1種類に限定されたり、また時系列データの場合はその途中経過も含めて予測精度の判定に用いられるため変動の大きいデータには適していないなどの限界がある。一方、学習型AIで利用するニューラルネットワークの場合、学習に使用する変数(本研究の場合は健康診断項目)を全てアウトプットデータとして予測することが可能である。また途中のデータポイントは予測精度の評価に用いず、最終年のデータの一致割合が重要となる。したがって、学習型AIは健康診断結果のような複数項目を予測したい場合に有用であると考えられる。この手法は周期性や経年変化を示すデータに利用できるため、産業保健現場の様々な情報に応用できると考えられ(例えば季節性変化を示す労働時間、労働災害、欠勤など)、今後の発展が期待される。

本研究にはいくつかの限界が存在する。1点目は、本研究がパイロット研究のため、訓練に用いたデータ数が164例と少なく、訓練データに依存する過学習が起きていた。研究を継続してより多くのサンプルを収集し、過学習を防ぐことが重要である。2点目は、DLは使用したデータセットに応じたモデルを作成するため、今回の研究では運輸業の労働者に限定されたモデルが作成され、他の業種の労働者には応用できない可能性がある。対象とする企業数および労働者を増やすことで一般化できる可能性がある。3点目は、健康診断データを用いて学習型AIを利用することを目的としたために、健康診断以外のデータを使用しなかった。したがって、論理的にはそのデータが影響を受ける可能性のある因子(例えば飲酒、喫煙、食習慣など)をモデルに組み込む検討の余地があるが、今回は使用しなかった。将来的にAIを労働者の行動変容を促すためのツールとして利用するのであれば、生活習慣の因子なども用いることが必要かもしれない。

本研究の結果、LSTMを利用した学習型AIは健康診断結果の将来予測に利用できる可能性が示唆された。本研究では近い将来の健診結果しか予測しなかったが、将来的に過去数年分の情報から10年後や20年後の予測を行い、疾病の2次予防に役立てたいと考えている。そのためには大量のデータを用いてモデルを作成することが必要である。本研究が進むと、地域や業種のバリエーションを包含した予測モデルを作成できることが期待され、また

将来的には労働者が自ら健康管理を行うための補助的ツールとなるかもしれない。今後も引き続き、産業保健分野でAI研究を進めることが必要である。

謝辞: 本研究の一部は大和証券ヘルス財団の研究助成を受けて実施した。またAI研究の議論の機会を提供してくれた産業保健AI研究会のメンバーに謝意を表する。

利益相反自己申告: 申告すべきものなし

文 献

- 1) 厚生労働省: 平成28年版厚生労働白書—人口高齢化を乗り越える社会モデルを考える—。[Online]. 2016 [cited 2020 May 24]; Available from: URL: <https://www.mhlw.go.jp/wp/hakusyo/kousei/16/>
- 2) 総務省: 世界の統計。[Online]. 2020 [cited 2020 May 24]; Available from: URL: <https://www.stat.go.jp/data/sekai/pdf/2020a1.pdf#page=15>
- 3) 厚生労働省: 高齢者の雇用。[Online]. 2012 [cited 2020 May 24]; Available from: URL: https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/koyou_roudou/koyou/jigyounushi/page09.html
- 4) 厚生労働省: 高齢者雇用対策。[Online]. 2018 [cited 2020 May 24]; Available from: URL: http://www.mhlw.go.jp/stf/seisakunitsuite/bunya/koyou_roudou/koyou/koureisha/index.html
- 5) 厚生労働省: 平成30年度「過労死等の労災補償状況」を公表します。[Online]. 2019 [cited 2020 May 24]; Available from: URL: https://www.mhlw.go.jp/stf/newpage_05400.html
- 6) 厚生労働省: 治療と仕事の両立について。[Online]. 2020 [cited 2020 May 24]; Available from: URL: <https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000115267.html>
- 7) 曾田研二, 田中平三. 予防医学のストラテジー. 東京: 医学書院, 1998:31-110.
- 8) 内田満夫, 野下浩司, 筒井保博, 小山 洋. 産業保健におけるDeep Learningの活用: 物流企業の職場巡視画像を用いたパイロット研究. 産業衛生学雑誌 2018;60:191-5.
- 9) 内田満夫, 野下浩司, 小山 洋. VDT作業者の安全衛生評価におけるDeep Learningの活用. 産業衛生学雑誌 2019;61:256-60.
- 10) 斎藤康毅. ゼロから作るDeep Learning 自然言語処理篇. 東京: オライリー・ジャパン 2018:223-75.
- 11) Hastie H, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. New York: Springer, 2017:241-2.
- 12) Nhung NTT, Amini H, Schindler C, et al. Short-term association between ambient air pollution and pneumonia in children: A systematic review and meta-analysis of time-series and case-crossover studies. Environ Pollut 2017;230:1000-8.
- 13) 竹内裕之, 児玉直樹, 高橋真悟. 健康データマイニングの評価(Ⅱ) 長期間の時系列データに基づく事例. 高崎健康福祉大学紀要 2015;14:1-12.