

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

Disease Risk Prediction AI and Lifestyle Improvement AI Solutions Based on Machine Learning Using Data Collected in Health Examinations

道庭 賢一 DONIWA Kenichi 田中 孝浩 TANAKA Takahiro 春木 耕祐 HARUKI Kosuke

我が国では、1970年頃から高齢化率が急激に上昇し、生活習慣病の患者数が増加している。それに伴い、医療費が年々増加し、社会問題になっている。更に、人々の健康状態は、経済や企業の活動にも大きな影響を及ぼすことから予防医療への意識が高まり、健康経営に積極的に取り組む企業も増えてきた。

東芝は、医療費増加と予防医療に対応するため、人々の健康な生活を支える精密医療の領域に取り組んでいる。今回、定期健康診断データなどを基に、生活習慣病の発症リスクを予測する疾病リスク予測 AI と、生活習慣改善を提案する生活習慣改善 AI を開発した。前者で発症リスクを高い精度で予測し、それを基に後者で客観的な保健指導を行って行動変容を促すことで、生活習慣病の予防に貢献できる。

Continually rising healthcare costs have become a social issue in Japan accompanying the increase in the number of patients suffering from lifestyle-related diseases due to the rapid growth of the elderly population since the 1970s. Various companies are consequently focusing on health promotion activities based on preventive medical care with the aim of securing the health of their employees, which affects economic and business activities.

In order to suppress the growth in healthcare costs and accelerate the dissemination of preventive medical care, Toshiba Corporation is promoting advances in the area of precision medicine aimed at supporting people's healthy lives. As part of these efforts, we have developed the following two artificial intelligence (AI) solutions based on machine learning using information collected in annual health examinations: (1) a disease risk prediction AI solution that can predict the risk of developing lifestyle-related diseases with high accuracy, and (2) a lifestyle improvement AI solution that can facilitate behavioral changes by providing objective health guidance based on predicted risks.

1. まえがき

我が国の大きな社会問題の一つに、生活習慣病罹患（りかん）者の増加による医療費の増加がある。代表的な生活習慣病である糖尿病は、自覚症状がないままに網膜症・腎症・神経障害の三大合併症を伴い、患者の生活の質（QOL）を著しく低下させる。更に、心臓病や脳卒中のリスクも上昇する。

厚生労働省が公表した2016年「国民健康・栄養調査」⁽¹⁾によると、糖尿病罹患者の数は、強く疑われる人が1,000万人、可能性を否定できない人も含めると2,000万人と推計され、年齢の上昇とともに、糖尿病罹患者と強く疑われる人の割合が増加する。生活習慣病罹患者の増加は、医療費の増加にもつながる。厚生労働省が公表した2018年度「国民医療費の概況」⁽²⁾によると、生活習慣病に関する医療費の割合は、国民医療費の約30%を占める。

生活習慣病の予防のためには、発症前の段階で、個人の特徴に応じて保健指導を行うことが重要である。これにより、医療費の増加とQOL低下を抑制できる。

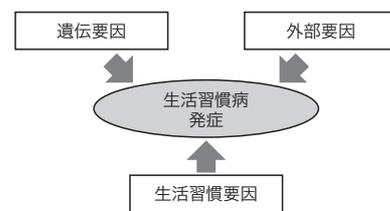


図1. 生活習慣病の要因

生活習慣病は、遺伝要因、外部要因、及び生活習慣要因が、発症の主要因と考えられる。

Factors affecting lifestyle-related diseases

生活習慣病の発症要因は、個人が持つ体質などによる遺伝要因と、病原体や有害物質などの外部要因、運動習慣・食習慣・休養・喫煙・飲酒などの生活習慣要因が考えられている（図1）。遺伝要因と外部要因は改善することが困難であるが、生活習慣要因は個人の努力により改善が可能である。保健指導と個人の努力により生活習慣要因を改善することで、発症リスクを低減できる。

個人ごとに保健指導を行うためには、

- (1) 保健指導の対象にする生活習慣病高リスク者の抽出
- (2) 個人に応じた生活習慣改善の目標設定が必要である。

そこで東芝は、疾病リスク予測AIと生活習慣改善AIを開発した。ここでは、二つのAIの概要と、性能評価結果について述べる。

2. 疾病リスク予測AI

生活習慣病の発症リスクを高精度に予測するために、疾病リスク予測AIを開発した。機械学習では、膨大な学習データが必要で、そのデータの質が予測精度に大きく影響する。当社は、良質なデータを保有しており、それを機械学習に用いることで、高精度のAIモデルを構築した。

2.1 データ

疾病リスク予測AIモデルの構築には、東芝健康保険組合の健康診断データとレセプトデータ（医療報酬明細書）を使用した。これらは、ガイドライン⁽³⁾に従って、匿名化、及び個人が特定できる特異値の削除を実施済みである。

健康診断データには、受診者情報や、生体検査値情報、既往歴・運動習慣・飲酒の習慣などの問診情報が含まれる（図2）。このうち、性別・年齢・身長・体重と、生活習慣病5疾病の発症の判定に使用するHbA1c（ヘモグロビンエーワンシー）や、収縮期血圧、拡張期血圧、中性脂肪（トリグリセライド）、HDL（高比重リポタンパク質）コレステロール、LDL（低比重リポタンパク質）コレステロール、GOT（グルタミン酸オキサロ酢酸トランスアミナーゼ）、GPT（グルタミン酸ピルビン酸トランスアミナーゼ）、 γ -GTP（ガンマグルタミルトランスペプチダーゼ）、BMI（体格指数）などの検査値を必須項目とし、必須項目が全てそろっているデータだけを使用した。使用するデータに必須項目以外の欠損値がある場合は、平均値で補間した。

レセプトデータには、病院などで診察を受けたときの診療

健康診断データ

	性別	年齢 (歳)	身長 (cm)	体重 (kg)	GOT (U/L)*	HbA1c (%)	...	糖尿病 既往歴	運動 習慣	...
Aさん	男性	43	172	68	25	5.7	...	なし	あり	...
Bさん	女性	52	147	45	20	5.3	...	なし	なし	...
...

*JSCC（日本臨床化学会）標準化対応法における酵素活性の国際単位

図2. 健康診断データに含まれる項目

健康診断データには、性別や年齢などの受診者情報、GOTやHbA1cなどの生体検査値情報、及び既往歴・運動習慣・飲酒の習慣などの問診情報が含まれる。

Information included in physical examination data

報酬明細書や、薬局で薬を処方されたときの調剤報酬明細書などがあり、調剤報酬明細書を使用して、予測対象とする生活習慣病の服薬者を抽出した。

疾病リスク予測AIの対象は、糖尿病、高血圧症、脂質異常症、肝機能障害、及び肥満の五つの生活習慣病とした。モデルの学習には、健康診断データと、予測対象とする疾病発症の有無が必要である。発症の有無の判定条件は、日本人間ドック学会の「検査表の見方（2018年4月改訂版）」⁽⁴⁾、及び日本高血圧学会の「高血圧治療ガイドライン」⁽⁵⁾を基に決定した。また、予測対象生活習慣病の医療用医薬品の服薬者は、発症者と判定した。例えば、収縮期血圧が正常値であっても、血圧降下剤を投与されている人は、発症者と判定する。調剤報酬明細書を使用することで、服薬者を抽出し、健康な人と発症者を正確に判別できた。

2.2 アルゴリズム

疾病リスク予測AIのアルゴリズムには、ランダムサバイバルフォレストを使用した。発症の有無と発症するまでの期間に焦点を当てた生存時間分析と、複数の決定木の結果を平均するランダムフォレストを組み合わせたアルゴリズムである。生存時間分析は、発症や死亡など対象となる事象をイベント、イベント発生までの期間を生存時間と定義し、図3のように、観察者数と発症していない健康者数を期間ごとに集計して、累積生存率と発症リスク率を計算する手法である。

ランダムサバイバルフォレストで学習し、モデルを作成することで、図4のように、個人の健康診断データを入力して、疾病リスクを予測できた。予測精度を高めるために、学習時に、当社独自の最適化技術を適用した。

期間	健康者数/観察者数	生存率	累積生存率	発症リスク値
1	9/10	0.90	0.90	0.10
2	8/9	0.89	0.80	0.20
3	7/8	0.88	0.70	0.30
4	6/7	0.86	0.60	0.40
5	5/6	0.83	0.50	0.50
6	5/5	1.00	0.50	0.50

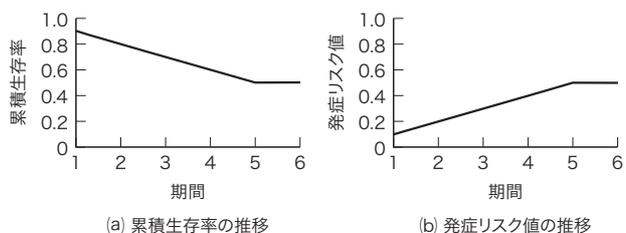


図3. ランダムサバイバルフォレストによる生存時間分析結果

観察者数と健康者数から、累積生存率と発症リスク率を計算する。

Example of results of survival time analysis using random survival forest

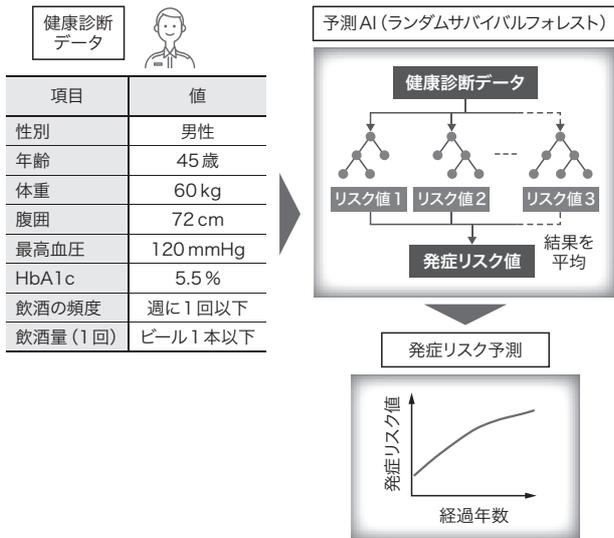


図4. 疾病リスク予測AIの概要

学習済みモデルに健康診断データを入力することで、疾病リスクを予測する。
Outline of disease risk prediction AI solution

2.3 性能

東芝健康保険組合のデータを使用して、3交差検証で評価した。3年以内に糖尿病を発症するリスクを予測し、検証データの平均AUC (Area Under the Curve)^(注1)が0.96という、高い予測性能を得た。東芝健康保険組合とは別の集団で実証実験を行ったところ、AUC 0.94を達成し、未知のデータに対する汎化性能も高いことが確認できた^{(6), (7)}。

3. 生活習慣改善AI

疾病リスク予測AIで発症リスクが高いと予測された人に、健康診断データに基づく客観的な保健指導を行い、目標を設定して行動変容を促すための生活習慣改善AIを開発した。

先行研究では、生活習慣の改善により体重が減少することが報告されている。3年間の生活習慣変化と体重増減について分析を行い、運動習慣、身体活動、就寝前の2時間以内の夕食、及び夕食後の間食が改善した群では、悪い生活習慣を持続した群と比べ、有意な体重減少が認められている⁽⁸⁾。また、体重減少により生体検査値が改善することも報告されている。レシピ情報・特定健診等情報データベース(NDB)のデータを使用し、積極的支援終了者について1年後の健康診断データを分析し、体重減少率と検査値の改善について調査を行っている。1～3%減量群では、体重変化なし群(±1%未満)と比較し、中性脂肪、LDLコレステロール、HDLコレステロール、HbA1c、GOT、

(注1) 予測の性能を表す指標。0～1.0の値で表し、1.0に近いほど予測性能が高く、0.5はランダムな判定と同等なことを意味する。



図5. 生活習慣改善フロー

生活習慣改善による体重減少と生体検査値改善により、発症リスクが低下する。

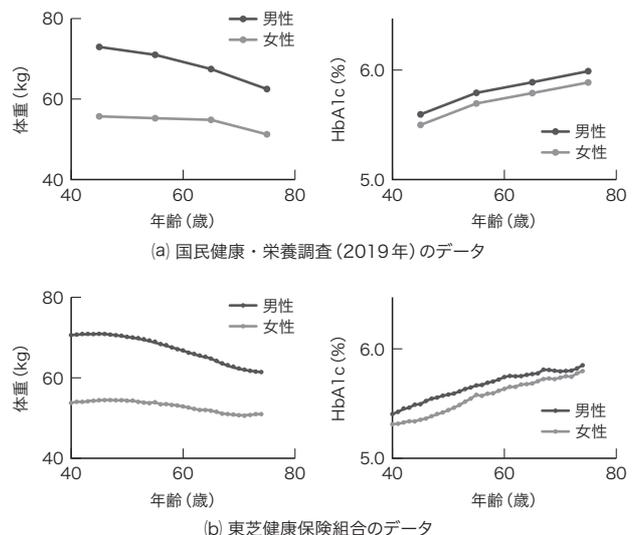
Flow of lifestyle improvement

GPT、及びγ-GTPに有意な改善が認められている⁽⁹⁾。

図5で示すように、生活習慣を改善することで体重が減少し、体重減少により生体検査値も改善する。生活習慣改善AIは、これらの変化を考慮して生活習慣改善によるリスク低下を予測し、個人ごとに最適な生活習慣改善を提案する。

3.1 体重減量率、生体検査値変化率

我が国の国民の健康状態については、厚生労働省の2019年「国民健康・栄養調査」に報告されている⁽¹⁰⁾。2019年に、約5,000人を対象に行われた調査結果の一部を、図6(a)に示す。10歳ごとの体重とHbA1cの平均値を示している。一方、東芝健康保険組合のデータを基に、同様の値を1歳ごとに記載したものを、図6(b)に示す。体重は45歳がピークで年齢上昇とともに減少し、HbA1cは上昇し続けることが読み取れる。東芝健康保険組合の健康診断データからは、国民健康・栄養調査より詳しい情報を読み取ることができる。東芝健康保険組合の健康診断データから、生活習慣改善による体重変化率と、体重減少による生体検査値の



*厚生労働省「令和元年 国民健康・栄養調査報告」⁽¹⁰⁾に基づいて作成

図6. 体重及びHbA1c平均値と年齢との関係の調査結果

東芝健康保険組合のデータを使用することで、国民健康・栄養調査のデータより、詳細な分析が可能になる。

Results of investigation on dependence of body weight and hemoglobin A1c (HbA1c) on age

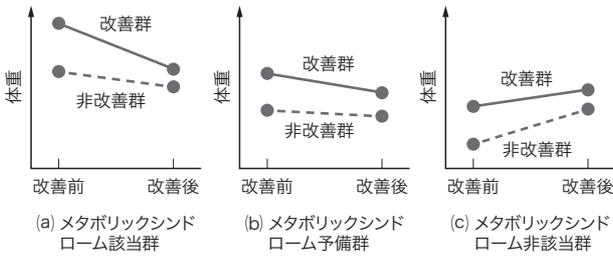


図7. 生活習慣改善による体重変化の例

生活習慣改善による体重変化では、メタボリックシンドロームの3分類で体重変化の傾向が異なる。

Examples of changes in body weight due to lifestyle improvement

変化率を解析して求めた。

ある生活習慣の改善群と非改善群の平均体重変化の例を、図7に示す。生活習慣改善による体重変化は、メタボリックシンドロームによる3分類(該当群, 予備群, 非該当群)で異なっているため、分類ごとに体重変化率を計算した。

健康診断データには、生活習慣に関する問診が複数ある。喫煙、運動、日常の歩行、歩行速度、食べる速度、遅い夕食、夜食・間食、朝食抜き、飲酒頻度、飲酒量、及び睡眠・休養を生活習慣改善の対象にして、体重変化を計算した。

同様に、体重減少による生体検査値変化率も、メタボリックシンドロームによる3分類で層別化して計算した。計算する生体検査値変化は、HbA1c, 収縮期血圧, 拡張期血圧, 中性脂肪, HDLコレステロール, LDLコレステロール, GOT, GPT, 及びγ-GTPである。

3.2 最適パターン探索

個人ごとに最適な生活習慣改善を提案するためには、複数ある生活習慣から、最適な改善パターンを探索しなければならない。これは、複数の選択肢の中から最適なパターンを選択する、ロス最小化問題として定義できる。

疾病リスクを目標以下に抑え、生活習慣改善の努力が少なくなるようなパターンを探索する。ロス、は、目標未達ロスと生活習慣改善ロスの合計として定義する。目標未達ロスは探索候補のリスク値と目標リスク値との差を示し、生活習慣改善ロスは生活習慣を改善する努力の大きさを示す。

このようにロスを定義することにより、今後生活習慣改善の対象が増加し、探索するパターン数が膨大になったとしても、ベイズ探索などの手法で効率良く求めることができる。

4. あとがき

生活習慣病の発症リスクを低減するために、疾病リスク予測AIと、生活習慣改善AIを開発した。

今後、ゲノムデータも使用し、個人の遺伝要因も考慮したAIを開発することで、医療費の増加とQOL低下を抑制し、社会問題の解決に貢献していく。

文献

- (1) 厚生労働省. “平成28年「国民健康・栄養調査」の結果”. <<https://www.mhlw.go.jp/stf/houdou/0000177189.html>>, (参照 2021-09-03).
- (2) 厚生労働省. “平成30年度 国民医療費の概況”. <<https://www.mhlw.go.jp/toukei/saikin/hw/k-iryohi/18/index.html>>, (参照 2021-09-03).
- (3) 個人情報保護委員会. 個人情報の保護に関する法律についてのガイドライン(匿名加工情報編). 2016, 25p. <<https://www.ppc.go.jp/files/pdf/guidelines04.pdf>>, (参照 2021-09-03).
- (4) 日本人間ドック学会. 検査表の見方(2018年4月改訂版). 2018, 10p. <<https://www.ningen-dock.jp/wp/wp-content/uploads/2013/09/dock-kensa-mikata.pdf>>, (参照 2021-09-03).
- (5) 日本高血圧学会 高血圧治療ガイドライン作成委員会編. 高血圧治療ガイドライン2019. ライフサイエンス出版, 2019, 302p. <https://www.jpnh.jp/data/jsh2019/JSH2019_hp.pdf>, (参照 2021-08-06).
- (6) 田口淳一, ほか. “健康診断データの AI 解析による糖尿病発症予測アルゴリズム開発”. 第60回日本人間ドック学会学術大会抄録集, 岡山, 2019-07, 日本人間ドック学会. 2019, p.260.
- (7) Taguchi, J. et al. "Prediction of Diabetes, Dyslipidemia, Hypertension, Liver function and Renal function using Artificial Intelligence (AI)". Joint Meeting of the 27th International Health Evaluation and Promotion Association and the 4th World Congress on Ningen Dock. Web congress, 2020-11. Japan Society of Ningen Dock. 2020, p.261.
- (8) 上村精一郎, ほか. 3年間の生活習慣の変化が体重増減に及ぼす影響. 総合健診. 2019, 46, 2, p.259-265.
- (9) 津下一代. 生活習慣介入のエビデンスと実際. 日本内科学会雑誌. 2016, 105, 9, p.1654-1661.
- (10) 厚生労働省. 令和元年 国民健康・栄養調査報告. 2020, 242p. <<https://www.mhlw.go.jp/content/000710991.pdf>>, (参照 2021-10-13).



道庭 賢一 DONIWA Kenichi
 研究開発センター
 知能化システム技術センター AI 応用推進部
 AI Application Dept.



田中 孝浩 TANAKA Takahiro
 研究開発センター
 知能化システム技術センター AI 応用推進部
 AI Application Dept.



春木 耕祐 HARUKI Kosuke
 研究開発センター
 知能化システム技術センター AI 応用推進部
 人工知能学会会員
 AI Application Dept.