

AI（人工知能）による3大生活習慣病(糖尿病・脂質異常症・高血圧)の3年以内発症確率の予測

特定健診の結果は分かりますか？

「特定健診を受けてください」と毎年封筒が市役所から送られてきます。それではと特定健診を受けると1~2ヶ月くらい経って特定健診の結果が送られてきます。いろんな数字が書かれていて、医者コメントもついています。身長・体重・血圧くらいまではわかるけど、並んでいる数字の意味するところがわからない。医者が推奨するコメントではとにかく医者に行けと言っている。行けば良いのはわかるけど数字が理解できないし、面倒なので放っておく。こんな感じの生活を送っている人が多いのではないのでしょうか。

私たちは特定健診の結果が個人にとっては非常に重要な情報であることをお知らせしたくて、AI（人工知能）を作ることになりました。特定健診の目的は、生活習慣病の予防であり、起こりうる病気を出来るだけ発症させないようにするためです。そこで分かりにくいあなたの特定健診の結果から、近い将来どれくらいの確率で生活習慣病になるのかをAIに計算してもらおうことにしました。AIはあなたの特定健診の数字を読み解いて、生活習慣病の3年以内の発症確率を算出します。あなたの発症確率が、天気予報の降水確率のように表現されます。これで難しい数字を専門知識を使って読み解くのではなくて、生活習慣病の発症確率という尺度で特定健診の結果を見ることができます。

次に、生活習慣病の発症確率が高いと判定された際には、自分と同じ特定健診の結果であるが、生活習慣だけが異なる場合における生活習慣病の発症確率をAIが計算します。生活習慣が異なる場合の確率と自分の確率を比較することによって、生活習慣の重要性をAIに教えてもらいます。喫煙はがん以外にも、肥満、高血圧、糖尿病などの様々な生活習慣病のリスクです。飲酒は適度であることが推奨されています。体重は飲食に大きく依存します。これらの生活習慣が異なる場合における発症確率がどの程度異なるかを自分で調べることができます。どんどん使ってください。そしてあなた自身でみずからの健康を守ってください。

ここから先は、ちょっと硬い内容になります。何を使って、どのようにして機械学習によりAIを作り上げたのかを知っていただければ嬉しく思います。

背景

代表的な生活習慣病である糖尿病・脂質異常症・高血圧は、心筋梗塞や脳梗塞、末期腎不全や認知症などの重篤な病気を引き起こす危険因子であることが知られている。一方で生活習慣病は、個人の生活習慣の改善により発症を抑制することが可能である。

アスマイルでは、大阪府の国保加入者の健診データ等を活用して健康予測モデルを構築し、アスマイルユーザー個人の現在の健

診データから近い未来に起こり得る生活習慣病の発症確率を予測するAIを搭載する。発症確率の算出を行うのは、3つの代表的な生活習慣病である糖尿病、脂質異常症、高血圧とする。これらの生活習慣病はそれぞれに診断基準が設定されている。今回のAIではそれぞれの病気で対象となる健康因子の数値が3年以内の健診結果において、診断基準に達することで病気発症とする。また発症確率は天気予報における降水確率と同じ指標であり、ユーザーにとってはわかりやすいのではないかと考えている。本稿では、アスマイルに搭載したAIの開発手順と評価方法について説明する。

今回は現在から3年以内の発症確率の予測を行うAIを導入する。そのために現在の体重や喫煙状況などを追加で入力頂く。そのほかの情報については直近の特定健診結果を用いる。さらには体重や生活習慣だけが異なる場合の発症確率と自分の発症確率を比較することで、生活習慣の改善がどれくらい重要であるかを実感することができます。

大阪府の国民健康保険データと対象者

現在大阪府では年間で約200万人が国民健康保険に加入している。そのうちの約30%の54万人程度が特定健診を受診している。この特定健診データなどの詳細な分析を行い、3年以内の生活習慣病の発症確率の予測を行う。

AIの学習に用いる対象者の選択は次のように行う。特定健診データは個人が特定できないように加工された2013年度から2017年度までの5年分の大阪府国保データを用いる。特定健診受診日から3年間の健診データを追跡するために、2013年4月から2014年12月の期間に特定健診を受けた人を対象とする。このうち、対象期間の特定健診日から3年間で定期的な特定健診の受診がな

い人を除くと、約25万人となった。そこから、初回特定健診時点で既に各疾病の診断基準を満たしている人や各疾病の治療薬を服用している人は解析対象者から除いた。なお、生活習慣病の診断基準は、各疾病のガイドラインに基づいて表1に示す基準を採用した。

表 1. 診断基準

生活習慣病	健診項目	診断基準数値
糖尿病	空腹時血糖	126mg/dL 以上
	ヘモグロビン A1c(HbA1c)	6.5%以上
脂質異常症	LDL コレステロール	140mg/dL 以上
	HDL コレステロール	40mg/dL 未満
	中性脂肪	150mg/dL 以上
高血圧	収縮期血圧	140mmHg 以上
	拡張期血圧	90mmHg 以上

初回特定健診時点で予測する疾病を発症している人(診断基準を満たす者、または服薬者)の割合は、糖尿病では約11%、脂質異常症では約59%、高血圧では約45%となった。アスマイルユーザーのうち、直近の特定健診データにおいて、予測する疾病を発症していた場合には、「発症予測の対象外です」と表示する。

次に、AIに入力する特定健診項目の選択を行う。連続変数として、年齢、BMI、収縮期血圧、中性脂肪、HDL コレステロール、LDL コレステロール、ALT(GPT)、HbA1cを用いた。今回は生活習慣を重要視するので、喫煙、飲酒の頻度、朝食摂取の有無、内服薬や既往歴の有無などを用いる。これらの質問項目の解答は必ずしも電子化されていないので、アスマイルでAIを使用する際には、現在の生活習慣として入力して頂くことにした。

機械学習の方法として勾配ブースティング決定木を用いた。計算速度と精度の点で優れているPython言語で書かれたLightGBMと呼ばれる公開されているソフトウェアを使うことにした。計算結果の信頼度向上には今回のようにビッグデータを使うのが重要である。

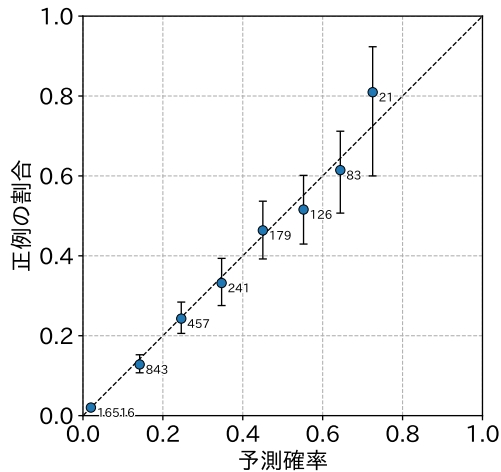


図 1. 糖尿病の信頼度曲線

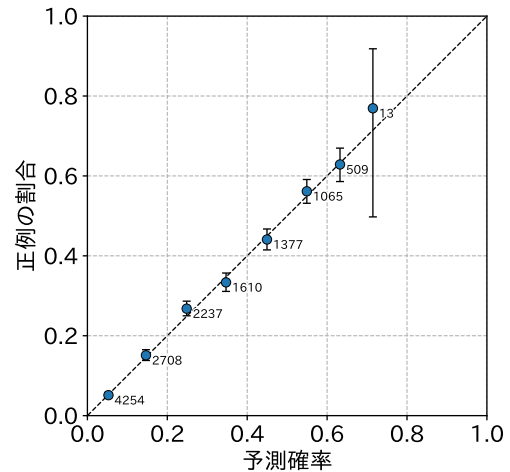


図 3. 高血圧の信頼度曲線

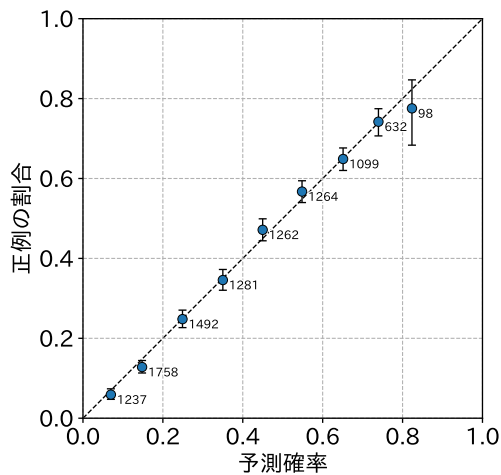


図 2. 脂質異常症の信頼度曲線

計算結果

予測確率の評価には信頼度曲線を用いる。信頼度曲線の横軸は、AIが予測した確率であり、縦軸はデータから計算した確率である。予測確率の評価には学習で用いたデータとは異なる検証用のデータを用いる。予測確率とデータから求めた確率が一致するというのは、45度の直線に点が乗ることに対応する。図中のそれぞれの点の横に書かれている数字は、データの数を表している。データの数が大きい場合には、それぞれの点の誤差棒が小さくなる。

図 1 では糖尿病の計算結果がプロットさ

れている。結果は黒点線で示してある45度の直線によくのっている。病気の発症確率を予測する良いモデルになっていることがわかる。図2には脂質異常症、図3には高血圧の信頼度曲線がプロットされている。AIによる予測とデータが非常に良く一致している。

AIは発症確率の計算結果に、どの健康因子が寄与しているかを示す重要度についても教えてくれる。協力ゲーム理論のシャープレイ値を機械学習に応用したSHAP値を全ての検証用データで計算し、その絶対値の平均値から重要度を計算する。糖尿病の場合に一番重要な健康因子はHbA1cである。二番目から四番目には降圧薬の内服、性別、LDLコレステロールが重要な要因になっている。脂質異常症の場合には、一番目はLDLコレステロールであり、二番目から四番目は中性脂肪、HDLコレステロール、BMIになっている。高血圧の場合には、一番目は収縮期血圧であり、二番目から四番目は年齢、BMI、飲酒の頻度になっている。

今回の機械学習の精度は素晴らしいものだが、それはひとえに大阪府全体のビッグデータを使うことで達成することができた。

アスマイルでの予測確率の表示

過去の大阪府国保加入者の大規模な特定健診データを使って学習させたAIに、アスマイルユーザーの健診データを入力することで、3年以内の生活習慣病の発症確率を予測する。結果は42%などの1%刻みの確率で表示する。確率を3つの領域(20%未満、20%以上50%未満、50%以上)に分けて、それぞれの生活習慣病のガイドラインに沿って確率に応じたコメントを行う。コメントについては糖尿病では日本糖尿病学会の糖尿病診療ガイドライン、高血圧では日本高血圧学会の高血圧治療ガイドライン、脂質異常症では日本動脈硬化学会の動脈硬化性疾患予防ガイドラインの推奨する生活改善の方法についての内容を表示する。

さらには、アスマイルユーザーごとにそれぞれの生活習慣病において、SHAP 値から算出された重要度が最も高い健康因子の分布とユーザーの値を表示する。そのことにより、自らの健康因子の数字は同じ性別・年齢層の人たちの中でどの位置にいるかを分かるようにしている。糖尿病についてはHbA1c、高血圧については収縮期血圧値、脂質異常症についてはLDL コレステロールを使う。

生活習慣病の発症リスクシミュレーション
生活習慣病の発症確率が20%以上の際には、特定健診の結果は同じであるが、体重や生活習慣だけが異なる場合の発症確率を、AIがシミュレーションすることによって、生活習慣の改善がどれくらい重要であるかを容易に知ることができる「発症リスクシミュレーション」機能を実装した。体重が3%低い場合、喫煙していない場合、ほとんど飲酒しない場合における発症確率をAIがシミュレーションし、自分の発症確率と比較することで、生活習慣の改善の重要性を実感することができる。それぞれのシミュレー

ション結果を比較することによって、3%の体重減少と生活習慣の改善のうち、発症確率の低下を最も期待できる方法を知ることができる。

生活習慣病の発症予防には、家で簡単に測定できる体重の管理が重要である。今回のAIではBMIが25 kg/m²以上の方が3%以内の体重減を目指す上で役立つ機能も付け加えた。

まとめ

アスマイルでは大阪府国保加入者の健康促進のために機械学習プロジェクトを立ち上げました。今回のアスマイルAIを十分に活用して下さることで、自らの健康を自らで知ることを経験していただき、行動変容につながれば嬉しい限りです。

(文責：大阪大学キャンパス健康支援・相談センター)