

科学研究費助成事業 研究成果報告書

令和 5 年 6 月 14 日現在

機関番号：31305

研究種目：基盤研究(C)（一般）

研究期間：2019～2022

課題番号：19K12206

研究課題名（和文）基本的検査の時系列データを活用した甲状腺機能異常症予測モデルの構築

研究課題名（英文）Constructing a prediction model of thyroid dysfunction making use of a set of time-series routine tests data

研究代表者

青木 空真（Aoki, Sorama）

東北医科薬科大学・薬学部・助教

研究者番号：40584462

交付決定額（研究期間全体）：（直接経費） 3,300,000円

研究成果の概要（和文）：甲状腺機能異常症の見逃しを防ぐため、専門検査である甲状腺関連ホルモンを測定する前段階として、測定頻度が高い血液検査項目を機械学習手法によって組み合わせて学習させて高精度に甲状腺中毒症と機能低下症を予測するモデルを構築、これをクラウド上で稼働させることで実際の病院において導入可能であることを示した。

また、検査値が前回受診時からどれくらい変動したのかを年間変動速度として表した特徴量を追加、時系列モデルとすることで、中毒症予測モデルの偽陽性率を半分にまで抑えることに成功した。一方、機能低下症に関しては時系列モデルにしても精度は向上せず、モデルの学習手法を各種試行しても改善は認められなかった。

研究成果の学術的意義や社会的意義

甲状腺機能異常症見逃し防止のための機械学習モデルについて、クラウド化により実際の病院における3健診施設にて稼働させることができたことで、人工知能の応用化が多忙な健診施設においても可能であることを示せたと考えている。また、本モデルはこのうち対象者数が最大であった施設（東北公済病院）にて見逃されていた患者25例の発見にも成功しており、患者QOLの向上にも繋がることを実証した。さらに検査値の時系列情報について、速度として取り扱うと有用である疾患と有用でない疾患の差異を限定的ながら明らかにできたことで、臨床検査値の取り扱いに関する知見を深めることができた。

研究成果の概要（英文）：In order to prevent thyroid dysfunctions from being overlooked, we constructed a machine-learning-based prediction model for thyrotoxicosis and hypothyroidisms by learning a combination of frequently measured blood test items, as a preliminary step to measuring thyroid-related hormones. We have also demonstrated that it is possible to introduce that model into actual hospitals by running it on the cloud.

Additionally, by adding a feature that expresses the rate of annual change from the values at the last examination, and turning it into a time-series model, we have successfully reduced the false-positive rate of the hyperthyroidism prediction model by more than half. On the other hand, even by turning it into a time-series model, the accuracy of the hypothyroidism prediction did not improve, and no improvement was observed even after trying various learning methods for the model.

研究分野：医療情報学

キーワード：甲状腺機能異常症 機械学習 人工知能 医療統計 基本的検査 時系列解析 スクリーニング

1. 研究開始当初の背景

甲状腺はヒトの咽頭隆起下、気管の前方に位置している体内で最も大きな内分泌器官であり、甲状腺ホルモンの分泌によってエネルギー産生を調整している。このため、自己抗体の出現や感染などの原因によってこの調整のバランスが崩れた場合は甲状腺機能異常症となり、全身性の多岐にわたる症状が現れる。こうした症状は疾患特異性に乏しく、各症状を個別に見ると他の病気と間違われやすいため、本来は甲状腺機能異常症であるにも関わらず循環器科や精神科を受診した結果、更年期障害や認知症、うつ病などと診断されて誤った治療が行われている症例も存在しており、見逃されやすい。一方でこうした見逃しを防止するためには、少なくとも健診受診者全例の甲状腺刺激ホルモン(TSH)を測定しなければ難しいという報告が甲状腺専門医らのグループにより複数なされているものの、TSHを測定すると1例あたり1,000円前後は必要であるため、費用対効果の観点から議論が続いているのが現状である。

そこで我々は、こうしたホルモンバランスの乱れが全身の基本的検査項目(総コレステロール(TC)や赤血球(RBC)など)にも影響を及ぼしていることに注目した。これら基本的検査項目は追加料金も発生しない、健診などで測定済みの項目であることから、こうした項目を組み合わせることで甲状腺機能異常症患者を発見可能になるならば安価で手軽なスクリーニングの実現が期待できる。本研究課題開始時点において、我々は健常対照群と甲状腺機能異常症患者群の検査値データを用意した上で、人工知能の一種であるパターン認識手法を3種類併用して検査値の変動パターンを学習させて予測・分類可能か検討を続けてきた。その結果、多数の検査項目より絞り込みを行うことで、機能亢進症ではALP上昇、S-Cr低下、TC低下、心拍数上昇の4項目[Sato W, et al. Biomed Pharmacother 2010]、機能低下症ではLDH上昇、S-Cr低下、TC低下、RBC低下の4項目[Aoki S, et al. Biomed Pharmacother 2012]を用いると高精度に分類可能なモデルが構築できることを明らかにしてきた。さらにこうして構築した予測モデルを組み込んだ、臨床現場において容易に計算することができる簡易スクリーニングツールを作成した上で実際の間ドック施設(JR 仙台病院ならびに東北公済病院)における受診者を対象としてスクリーニングを実施したところ、2008年から2013年の5年間で計13,186名について計算し、その中から25名の新規症例を発見することに成功してきた[青木 空真, ほか. 人間ドック 2012] [Aoki S, et al. Ningen Dock International 2014]。これら発見症例は1例を除いて人間ドック担当医にも疑われておらず、本スクリーニングを経なければ見逃されていた症例と考えられた。

一方、本スクリーニングの問題点としては偽陽性の多さが挙げられ、例えば前述の期間において予測モデルにより疾患が疑われたものの、甲状腺ホルモンを測定すると基準範囲内であった偽陽性例が225名存在した。これは偽陰性を防ぐため、研究段階ということで予測のカットオフ値を低く設定したことにもよるが、他にも健常対照群と疾患群を用いた比較のみでは明らかにならなかった、健診における受診者の多様性(基礎疾患の存在や生活習慣など)に起因する検査値の個人差によってももたらされていることが明らかとなった。今後さらなる実用段階を目指すため、より精度を向上させるにはこうした検査値の個人差を補正できるモデルの構築が必須であると考えられた。また、あわせてスクリーニングツールも簡易なものではなく、実際に臨床現場において使用可能な枠組みを作成し、運用が可能でなければ本研究の目的は達成されないものと考えられた。

2. 研究の目的

そこで本研究では基本的検査値の時系列データに着目し、前回受診時から今回受診時にかけての検査値変化を予測モデルに組み込むことで、こうした検査値に存在する個人差を補正可能かどうか検討した。もともと本手法は、例えばコレステロールの1項目が低いというだけでは甲状腺機能が亢進しているのか、栄養状態が悪いのか、慢性の消耗性疾患があるのかなど様々な要因による変動が考えられるところ、人工知能によって複数の項目の変動を同時に考慮することで、各検査値の変動の度合いから総合的に何パーセントその疾患が疑われるのか、という判断を行わせることが可能となったものである。ここに前回受診時から各項目がどのように変動したのかという時系列の情報を組み込むならば、例えば前述の慢性疾患であれば前回受診時も今回受診時も低値のまま推移しているところ、甲状腺機能亢進症であれば大きく低下している動きが期待され、両者を判別するための大きな特徴を追加できるものと考えられる。こうして精度を高めた上で本スクリーニングを全国に展開できれば、人間ドック受診者は年間約760万人と推計されていることから、見逃されている患者を15,000人前後発見できると期待でき、これら患者のQOL向上へ大きく寄与することが可能である。

3. 研究の方法

時系列データを保持している健常対照群、甲状腺中毒症患者群、甲状腺機能低下症患者群のデータを対象に、各検査値の時系列変動速度を算出した。この時、人間ドックは受診間隔が概ね1年前後になることが多いため、時系列変動速度は次式に基づき1年間あたりの変動速度へと変換した。

$$v_i = \frac{(x_1 - x_2) \cdot 365}{x_1(t_2 - t_1)}$$

ここで、 v_i は検査項目*i*(ALP, S-Cr, TC, ChE, LDH, RBC, 心拍数)の年間変動速度、 x_1 は検査項目*i*の今回測定値、 t_1 は x_1 の測定日、 x_2 は検査項目*i*の前回測定値、 t_2 は x_2 の測定日を表す。

こうして得られた各検査項目とそれぞれの変動速度を変数として、ベイズ正則型ニューラルネットワーク(BRNN)およびサポートベクトルマシン(SVM)に入力、学習させて予測モデルを構築した。構築の際はBRNNとSVMにおける予測・分類傾向の性質を踏まえてアンサンブル学習モデルとして、BRNNあるいはSVMのうち予測率がより高い方を出力結果として採用した。このモデルに対し、従来の1時点予測モデルでは偽陽性となったケースを主としたテストデータを予測させ、特異度が向上するか(偽陽性を減少させるか)検討した。併せて感度も維持しているかどうか確認するため、並行して学習データについてLeave-One-Out法を用いて確認した。また、結果に応じて時系列データの取り扱い方法について改良を加え、年間変動速度に変換するのではなく、リカレントニューラルネットワーク(RNN)を用いて時系列情報を直接学習させるモデルの構築も試みた。まずは甲状腺中毒症の予測モデルについて、健常対照群45名、甲状腺中毒症患者群21名を学習データ、1時点モデルによる中毒症予測率が60%以上で、甲状腺ホルモンの測定を行ったが基準範囲内だった偽陽性例122名をテストデータとして解析を実施した。このとき、従来有用性を明らかにしてきた4項目(ALP, S-Cr, TC, 心拍数)に加えて、甲状腺ホルモンの生化学的作用を考慮し、測定頻度は少ないものの関係性が深いと推察されるコリンエステラーゼ(ChE)も加えた時系列モデルを別個に作成・評価することでよりさらなる精度の向上が確認できるかも検討した。次に甲状腺機能低下症の予測モデルについて、同様に健常対照群42名、甲状腺機能低下症患者群14名を学習データ、テストデータは偽陽性例に限らず甲状腺機能異常者を除いた人間ドック受診者4,571名のデータを用いて検討した。

4. 研究成果

(1) 甲状腺中毒症の時系列予測モデル

使用した各群のデータの概要を表 1 に示す。また、本データを用いて予測計算を行った結果を横軸に 1 時点予測モデル、2 時点予測モデル、ChE 追加 2 時点予測モデルを並べ、縦軸を各々のモデルにより計算された甲状腺中毒症予測率とし、患者群と偽陽性群における各サンプルの予測率推移を図 1, および図 2 に示す。図 1 より、患者群予測率の推移に着目すると予測率 85%以上の予測例は 1 時点モデル、2 時点モデル、ChE 追加の 2 時点モデルでそれぞれ 21 名中 15 名、17 名、20 名となり、時系列

表 1. 使用データ各群における検査値と算出した変動速度の平均値 ± 標準偏差

	健常群 (n=45)	患者群 (n=21)	偽陽性群 (n=122)
ALP [IU/L]	186.69±53.54	277.62±110.73	311.63±110.38
TC [mg/dL]	197.20±22.85	144.19±21.61	170.24±28.66
S-Cr [mg/dL]	0.57±0.08	0.45±0.07	0.46±0.08
HR [/min]	66.71±9.57	78.48±9.94	75.55±14.01
ChE [IU/L]	320.16±60.17	423.20±70.35	338.5±83.69
V _{ALP} [%/year]	3.60±13.70	44.52±49.26	6.86±30.22
V _{TC} [%/year]	-0.13±10.27	-20.04±15.84	-5.36±15.04
V _{S-Cr} [%/year]	-0.81±5.97	-21.97±15.61	-2.51±8.68
V _{HR} [%/year]	-0.49±11.84	28.86±19.13	6.6±14.78
V _{ChE} [%/year]	-1.11±7.45	26.05±22.42	-3.1±11.00

情報の追加と ChE の追加で感度の上昇が確認された。特に ChE 追加の 2 時点モデルは 1 例の未鑑別軽症例を除いて、軽症例もほぼ全て含む発見 20 症例を閾値以上で算出することができた。偽陽性群予測率の推移を同様に図 2 より確認すると、こちらは 1 時点予測モデルから 2 時点予測モデルで偽陽性数が 45 名 24 名とほぼ半減していたが、ChE 追加 2 時点予測モデルはさらにこれが 6 名となり、1 時点予測モデルから比べると 9 割弱偽陽性を減少させた。このことから、時系列情報は精度の向上に大きく寄与するとともに、さらに ChE の追加は極めて有用性が高いことが示唆された。これは、特に従来モデルでは低栄養状態による TC 低下例や肝機能障害による ALP 高値例が偽陽性となりやすかったが、ChE はこれらの例では低値を示す一方、甲状腺中毒症患者では高値を示すという逆の傾向を持つことが要因になったと考えられ、こうした生化学的な機序や臨床背景に着目しながら予測モデルを構築することの重要性が示された。ただし、ChE は測定されていない施設も存在するため、各施設において測定されている項目に応じた最適なモデルで計算されるようなシステムに調整する必要性は認められた。

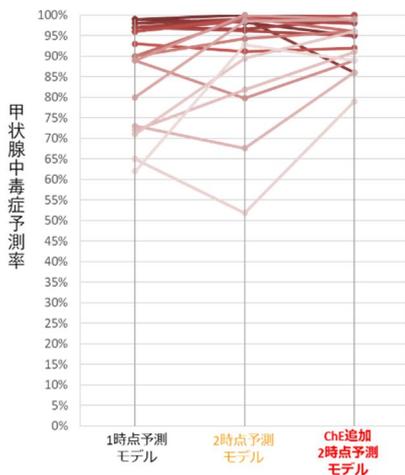


図 1. 甲状腺中毒症患者群 21 名の予測率推移

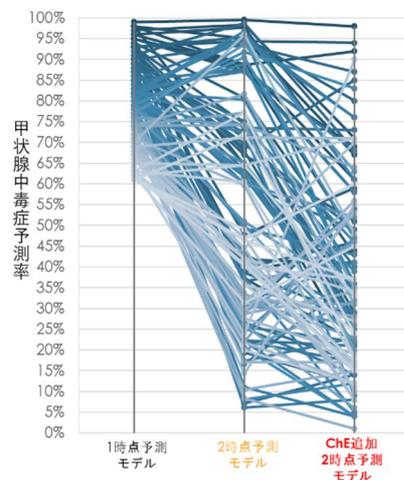


図 2. 偽陽性群 21 名の予測率推移

(2) 甲状腺機能低下症の時系列予測モデル

同様に、本モデルの構築に当たって用いたデータの概要を表1に示す。また、本データを用いて1時点予測モデルおよび2時点予測モデルを構築し、感度と特異度を算出してプロットしたROC曲線を図3に示す。結果を比較すると、曲線下面積(AUC)では1時点予測モデルが0.990、2時点予測モデルが0.987であり、時系列情報を追加した予測モデルでも精度にほとんど変化が見られなかった。予測結果の内訳を確認するため、縦軸に1時点予測モデルおよび2時点予測モデルにて計算された甲状腺機能低下症予測率をプロットした、各群の予測率分布を箱ひげ図にて図4、および図5に示す。予測率の分布を比較すると、2時点予測モデルの方が予測率の分散が僅かに増加していたため、これは計算された時系列速度に個人差が大きく、明確なパターンを持たない可能性が示唆された。この傾向は、時系列速度の取り扱い方を変えて、RNNにより時系列情報を直接学習させたモデルによる計算を実施しても変わらなかった(図6, AUC=0.980)。以上の結果より、甲状腺中毒症において本時系列モデルは精度向上に大きく寄与したが、甲状腺機能低下症では中毒症と比べて検査値の変動がより緩やかで、生活習慣などに起因する個人差の影響の方が疾患による影響よりも大きいこと、このため1年間の変動を単純に追跡するだけでは情報が限られてしまっている可能性が高く、今後は複数年のデータを収集し、まとめて取り扱うモデルの必要性が示唆された。

表2. 使用データ各群における検査値と算出した変動速度の平均値 ± 標準偏差

	健常群 (n=42)	疾患群 (n=14)	テストデータ群 (n=4571)
LDH [IU/mL]	171.45 (22.91)	257.14 (66.01)	179.18 (30.38)
RBC [10 ⁶ /μL]	4.43 (0.29)	3.85 (0.33)	4.26 (0.36)
TC [mg/dL]	197.26 (21.9)	279.7 (42.35)	208.31 (33.94)
S-Cr [mg/dL]	0.57 (0.07)	0.8 (0.16)	0.65 (0.2)
V _{LDH} [%/year]	-4.28 (9.89)	26.34 (30.2)	0.94 (12.07)
V _{RBC} [%/year]	-0.64 (4.21)	-0.21 (5.27)	-0.18 (4.48)
V _{TC} [%/year]	-0.48 (10.69)	26.58 (25.89)	0.04 (10.8)
V _{S-cr} [%/year]	-0.95 (6.13)	18.58 (24.31)	0.57 (7.93)

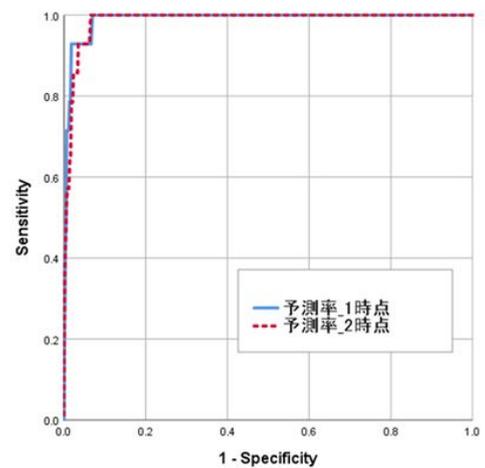


図3. 健常対照+テストデータ群を陰性群、機能低下症患者群を陽性群としたROC曲線

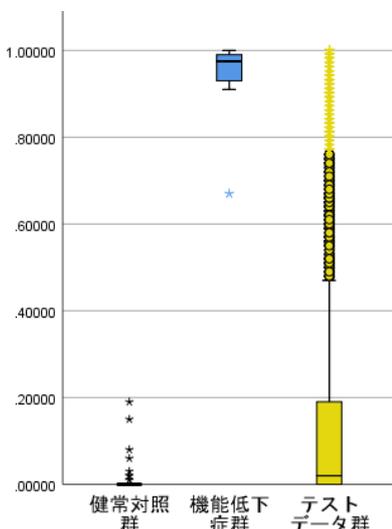


図4. 3群の予測率分布
(1時点予測モデル)

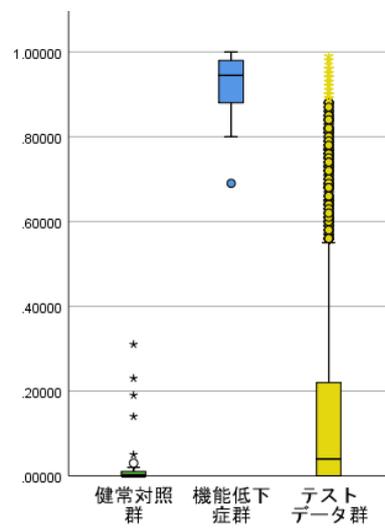


図5. 3群の予測率分布
(2時点予測モデル)

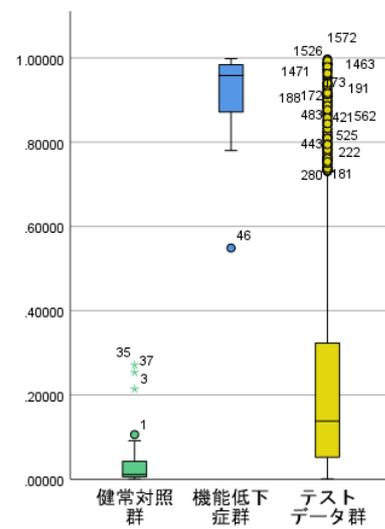


図6. 3群の予測率分布
(RNN2時点予測モデル)

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件/うち国際共著 0件/うちオープンアクセス 1件）

1. 著者名 佐藤 憲一, 青木 空眞, 阿部 杏奈, 小沢 晃世, 星 憲司, 川上 準子, 中川 吉則, 森 弘毅, 飛田 渉, 吉岡 明美, 佐藤 譲, 藤井 豊, 吉田 克己	4. 巻 34
2. 論文標題 人間ドック受診時の基本的検査データを用いた 甲状腺機能異常症の診断支援 - クラウドスクリーニングサービスの展開と有用性 -	5. 発行年 2020年
3. 雑誌名 人間ドック	6. 最初と最後の頁 718-730
掲載論文のDOI (デジタルオブジェクト識別子) 10.11320/ningendock.34.718	査読の有無 有
オープンアクセス オープンアクセスとしている(また、その予定である)	国際共著 -

〔学会発表〕 計6件（うち招待講演 0件/うち国際学会 0件）

1. 発表者名 吉田 克己, 飛田 渉, 佐藤 憲一, 青木 空眞, 星 憲司, 川上 準子
2. 発表標題 一般検査による甲状腺機能異常症のクラウドスクリーニング - 東北公済病院での5年間の経験とホルモン過剰症の真陽性と偽陽性の鑑別法 -
3. 学会等名 第34回東北甲状腺談話会
4. 発表年 2023年

1. 発表者名 青木空眞, 星憲司, 川上準子, 佐藤憲一, 中川吉則, 森弘毅, 深澤洋, 飛田渉, 吉田克己
2. 発表標題 血液の基本的検査項目を組み合わせた甲状腺機能異常症の診断支援～クラウドスクリーニングの現況と甲状腺機能低下症時系列予測へ向けた取り組み
3. 学会等名 第33回東北甲状腺談話会
4. 発表年 2022年

1. 発表者名 青木空眞, 星憲司, 川上準子, 佐藤憲一, 中川吉則, 森弘毅, 飛田渉, 吉田克己
2. 発表標題 複数の基本的検査項目を組み合わせた甲状腺機能低下症の診断支援: 人工知能(AI)による2時点の検査値を用いた時系列予測モデル構築の試み
3. 学会等名 日本薬学会第142年会
4. 発表年 2022年

1. 発表者名 吉田 克己、飛田 渉、青木 空眞、星 憲司、川上 準子、佐藤 憲一
2. 発表標題 一般検査による甲状腺機能異常症のスクリーニングー東北公済病院での9年半の経験ー
3. 学会等名 第32回東北甲状腺談話会
4. 発表年 2021年

1. 発表者名 青木 空眞、齋藤 咲、星 憲司、川上 準子、佐藤 憲一、渡部 輝明、中川 吉則、森 弘毅、深澤 洋、飛田 渉、吉田 克己
2. 発表標題 複数の基本的検査項目を組み合わせた甲状腺中毒症の診断支援：コリンエステラーゼを新規に追加した2時点予測モデルの構築と評価
3. 学会等名 第62回日本甲状腺学会学術集会
4. 発表年 2019年

1. 発表者名 佐藤 憲一、青木 空眞、阿部 杏奈、小沢 晃世、星 憲司、川上 準子、中川 吉則、森 弘毅、飛田 渉、吉岡 明美、佐藤 謙、藤井 豊、吉田 克己
2. 発表標題 人間ドック受診時の基本的検査データを用いた甲状腺機能異常症の診断支援：クラウドスクリーニングサービスの展開と有用性評価
3. 学会等名 第36回甲状腺病態生理研究会
4. 発表年 2020年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究協力者	吉田 克己 (Yoshida Katsumi)		

6. 研究組織（つづき）

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究協力者	飛田 渉 (Hida Wataru)		
研究協力者	川上 準子 (Kawakami Junko)		
研究協力者	星 憲司 (Hoshi Kenji)		
研究協力者	佐藤 憲一 (Sato Kenichi)		

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関