

機械学習による2型糖尿病の新たなサブタイプ分類手法の開発 ～日常臨床データだけで高精度な分類・予測が可能に～

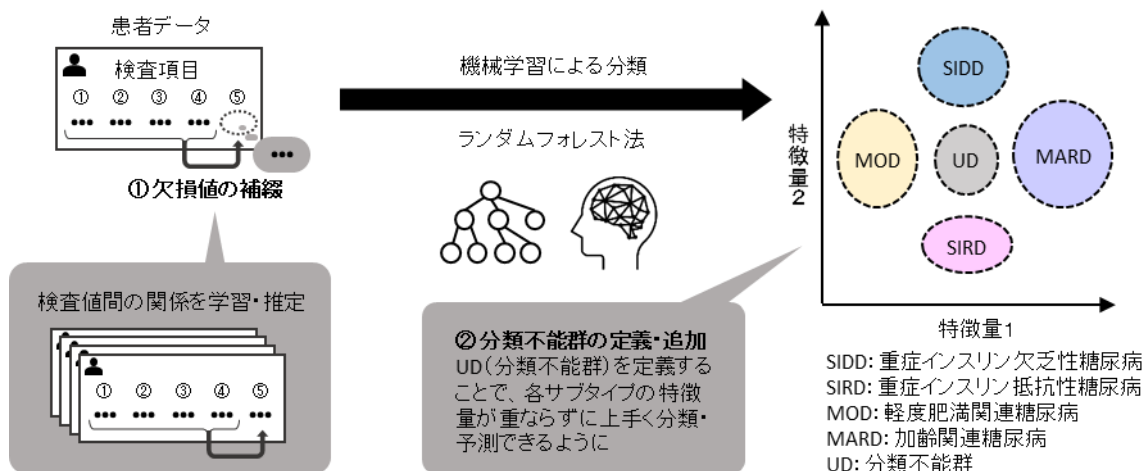
千葉大学大学院医学研究院の川上英良教授、福島県立医科大学糖尿病内分泌代謝内科学講座の田辺隼人助教、島袋充生教授らは、ランダムフォレスト^{注1)}と呼ばれる機械学習の手法に欠損データの補綴（ほてい）と、予測確率の低い分類カテゴリー（分類不能群）の定義・追加を組み合わせることにより、日常的に得られる臨床データのみを用いて、2型糖尿病のサブタイプを高精度に予測できる機械学習モデルを開発しました。本研究により患者さん一人ひとりの合併症リスクや治療反応性を、本モデルと手持ちのデータから高精度に予測できるようになるため、臨床現場での糖尿病のサブタイプの予測、糖尿病の個別化医療実現への大きな寄与が期待されます。

本研究成果は、2024年8月21日に欧州糖尿病学会雑誌 Diabetologia に掲載されました。

■ 研究の背景

糖尿病は、インスリンが十分に働かないために、血液中のブドウ糖（血糖）の量が多い状態が続く病気です。日本における糖尿病の大半を占めるのが2型糖尿病で、その患者数は予備軍を含めると1,000万人をこえると推定されています。膵臓から分泌されるインスリンによって、血糖値は一定の範囲にコントロールされますが、2型糖尿病ではインスリン分泌の減少や、インスリンへの反応の低下により、血糖値が高くなります。血糖値が高い状態が長期間続くと、血管がダメージを受け、失明、神経障害、脳卒中、心筋梗塞、腎不全などの重篤な合併症のリスクにさらされます。

Ahlqvist による先行研究^{参考文献1)}では、2型糖尿病は重症インスリン欠乏性糖尿病、重症インスリン抵抗性糖尿病、軽度肥満関連糖尿病、加齢関連糖尿病の4種類のサブタイプに分類されますが、病気の状態や原因、合併症の有無・種類、治療反応が個人で異なるため、患者さんの糖尿病のサブタイプを知った上で、それに合わせた治療を進めることが重要です。2型糖尿病の分類には個々の患者さんのインスリンの分泌・感受性に関連するデータが必要ですが、日常臨床データには含まれない項目のため、臨床現場で活用しにくいという課題がありました。また理論上は糖尿病のサブタイプは長期にわたって安定的と考えられていますが、従来法での分類では経時的なサブタイプの変化がしばしば生じ、サブタイプ分類は長期的な治療計画には使いにくいと考えられてきました。そこで本研究では、日常臨床データだけで2型糖尿病のサブタイプを高精度かつ一貫性をもって予測できるモデルの開発を試みました。



【図】機械学習による2型糖尿病の新たなサブタイプ分類手法の開発。

ランダムフォレストと呼ばれる機械学習の手法に、①欠損値の補綴（ほてい）②分類不能群の定義・追加を組み合わせることで、日常的に得られる臨床データのみを用いて、2型糖尿病のサブタイプを高精度に予測できる機械学習モデルを開発した。

■研究の成果

研究グループはランダムフォレストと呼ばれる教師あり機械学習の手法を用いた予測モデルを開発しました。その際、検査値間の関係を学習することで、日常臨床データの値からインスリンに関するデータの値を推定可能にしました(図-①欠損データの補綴)。また先行研究では4つだった2型糖尿病のカテゴリーに、分類予測確率の低いカテゴリー(分類不能群: UD)を定義・追加することで(図-②)、各サブタイプの特徴量の分布が重ならずによく分類・予測できるようになりました。得られたモデルによるサブタイプ予測は、教師データである従来法での分類結果に対し、82.9-94.0%という高い精度を示しました。またサブタイプ分類の一貫性の検討のため、観察期間前後のサブタイプ予測の結果を比較したところ、前後でサブタイプが一致していた割合は重症インスリン欠乏性糖尿病(SIDD): 100%、重症インスリン抵抗性糖尿病(SIRD): 68.6%、軽度肥満関連糖尿病(MOD): 94.4%、加齢関連糖尿病(MARD): 97.9%となり、従来法よりも高い一貫性を示しました。

今回得られた予測モデルによって、従来必須であった項目が欠損していても日常臨床データを用いたサブタイプ分類が可能となり、患者さん一人ひとりの合併症リスク(失明、神経障害、脳卒中、心筋梗塞、腎不全など)や治療反応性を、手持ちのデータから高精度に予測できるようになりました。

■今後の発展・展望

本成果により、日常診療データからの糖尿病のサブタイプ予測が可能となり、臨床現場におけるサブタイプ分類、糖尿病の個別化医療実現への大きな寄与が期待されます。また一貫性の高いサブタイプ分類が可能となったことで、個々の患者さんの合併症リスクや治療反応性に基づいた長期的な治療計画や予後の推定にも本モデルが貢献できると考えられます。

■用語解説

注1) ランダムフォレスト: 教師あり機械学習の代表的な手法の一つ。教師データ(正解がわかっているデータ)を用いて学習し、分類や回帰を行う。外れ値に強く、変数の依存関係を考慮できる、特定の変数のみに影響を受けすぎないといった長所を持つ。

■論文情報

タイトル: Machine learning-based reproducible prediction of type 2 diabetes subtypes

著者: Hayato Tanabe, Masahiro Sato, Akimitsu Miyake, Yoshinori Shimajiri, Takafumi Ojima, Akira Narita, Haruka Saito, Kenichi Tanaka, Hiroaki Masuzaki, Junichiro J. Kazama, Hideki Katagiri, Gen Tamiya, Eiryo Kawakami, Michio Shimabukuro

雑誌名: Diabetologia

DOI: 10.1007/s00125-024-06248-8

■参考文献

1. Ahlqvist E, Storm P, Käräjämäki A et al. (2018) Novel subgroups of adult-onset diabetes and their association with outcomes: a datadriven cluster analysis of six variables. *Lancet Diabetes Endocrinol* 6(5):361–369.

<本研究に関するお問い合わせ>

千葉大学大学院医学研究院 / 国際高等研究基幹 教授 川上 英良 (カワカミ エイリョウ)

TEL : 043-226-2826 E-mail : eiryo.kawakami@chiba-u.jp

<広報に関するお問い合わせ>

千葉大学広報室

TEL : 043-290-2018 E-mail : koho-press@chiba-u.jp