

● 研究演題 6

特定健康診査データを用いた高血圧症発症を予測する機械学習モデルの構築

群馬大学大学院医学系研究科 博士3年 大川貴史

【目的】

高血圧症を5年後に発症しているか予測する、新たなモデルを構築すること。

【方法】

本研究では、群馬県の国民健康保険の2011-2021年の特定健康診査データを使用した。特徴量として、特定健康診査データから得られる個人情報、生体検査値、既往歴・服薬状況・生活習慣の問診情報など32の項目を用いた。予測する生活習慣病の対象を高血圧症の新規発症とし、高血圧症該当の有無は、収縮期血圧 $\geq 140$  mmHg or 拡張期血圧 $\geq 90$  mmHg or 降圧剤の服薬により定義した。機械学習は、識別器にLightGBMを用いるモデルを採用した。また、不均衡データを有効的に学習するために、クラス間のデータ数のバランスを調整する「アンダーサンプリング」とアンサンブル学習の1つである「バギング」を組み合わせた学習を行なった。評価指標は再現率、適合率及びROC-AUCを用いて行なった。

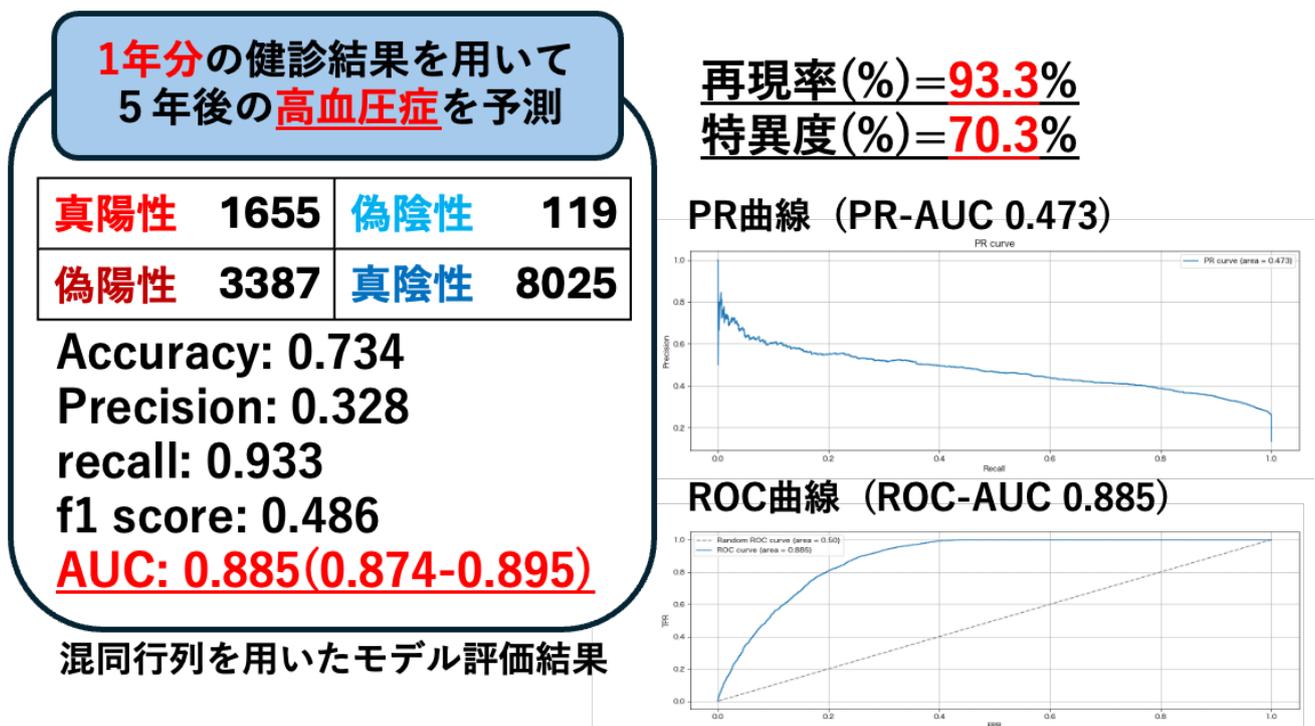
【結果】

2011年から2016年に健診を受診した人(n=924,932)のうち、5年後に再度受診していた131,857人を高血圧発症予測モデル構築に用いた。高血圧症の再現率は0.991、適合率は0.28、ROC-AUCは0.885であった。ROC-AUCは既存の統計学的手法と比較しても高い結果を示した。

【考察】

単年のデータにもかかわらず再現率、ROC-AUCが高い値を示したため、本モデルは生活習慣病のスクリーニングに有用であることが示唆された。

図1 機械学習を用いた高血圧症発症予測モデル結果



混同行列を用いたモデル評価結果