

[研究助成 (C)]

機械学習を用いた低侵襲かつ汎用的な腎臓病診断支援モデルの構築

2237009

研究代表者
(助成金受領者)

共同研究者

聖マリアンナ医科大学
腎臓・高血圧内科聖マリアンナ医科大学
腎臓・高血圧内科聖マリアンナ医科大学
腎臓・高血圧内科大学院生/
診療助手

准教授

主任教授

野田 竜之介

市川 大介

柴垣 有吾

[研究の目的]

IgA 腎症は、世界で最も頻度の高い原発性糸球体腎炎の一つであり、診断後 20 年以内に 30-40% の患者が末期腎不全に至る。早期発見と治療が重要であるが、確定診断には侵襲的な腎生検が必要であり、合併症のリスクもある。本研究の目的は、以下の 2 点である。

1. 腎生検に頼らず、血液検査や尿検査などの一般的な臨床データを用いて IgA 腎症を非侵襲的に診断予測する機械学習モデルを開発し、検証すること。
2. 従来の統計モデル（ロジスティック回帰）と比較して、機械学習モデルが IgA 腎症の診断において高い精度と信頼性を持つことを示すこと。

[研究の内容, 成果]

本研究は、聖マリアンナ医科大学病院で腎生検を受けた成人患者のレトロスペクティブなデータを用いたコホート研究である。2006 年 1 月から 2022 年 9 月までに腎生検を受けた患者を対象とし、診断が不確かな患者、複数の主要な診断がある患者は除外した。最終的に 1268 名の患者が解析対象となり、2006 年 1 月から 2019 年 12 月までのデータを開発コホート (n=1027)、2020 年 1 月から 2022 年 9 月まで

のデータを検証コホート (n=241) とした。

潜在的な予測変数として、人口統計学的特徴（年齢、性別、身長、体重、BMI）、血液検査項目（白血球数、ヘモグロビン、総蛋白、アルブミン、BUN、クレアチニン、尿酸、AST、ALT、ALP、LDH、CK、総コレステロール、グルコース、HbA1c、CRP、IgG、IgA、IgM、C3、C4、IgA/C3 比、抗核抗体）、尿検査項目（尿蛋白/クレアチニン比、尿中赤血球）を収集した。欠損値は k-最近傍法で補完し、4 つの変数選択手法（LASSO, Random Forest-Recursive Feature Elimination, Random Forest-Filtering, SelectFromModel with Extra Trees）を組み合わせ、最終的に 14 個の予測変数（年齢、ヘモグロビン、総蛋白、アルブミン、LDH、CK、CRP、IgG、IgA、C3、C4、IgA/C3 比、尿中赤血球、尿蛋白/クレアチニン比）を選択し、機械学習モデルに組み入れた。

モデル開発には、5 つの機械学習法（XGBoost, LightGBM, Random Forest, Artificial Neural Network, 1D-CNN）とロジスティック回帰を適用した。開発コホートで 5 回反復 5 分割交差検証とベイズ最適化を用いてハイパーパラメータを調整し、AUROC（Area Under the Receiver Operating Characteristic curve）と AUPRC（Area Under the Precision-Recall Curve）で識別能を評価した。また、95% 信頼区間は 1000 回のブートストラップ反

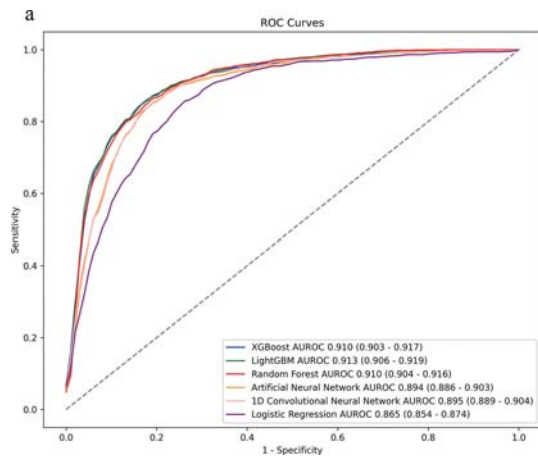
復で推定し、Deep ROC 解析で識別能を詳細に評価、キャリブレーションプロットと Brier スコアでモデルの較正能と総合性能を評価した。

結果、開発コホートでは LightGBM は 0.913 (95% CI 0.906-0.919) という最も高い AUROC を達成した。これは、ロジスティック回帰、Artificial Neural Network, 1D-CNN よりも有意に高く、XGBoost および Random Forest とは有意差はなかった (図 1a)。検証コホートでは、XGBoost が 0.894 (95% CI 0.850-0.935) という最も高い AUROC を示したが、他のモデルと有意差は認めなかった (図 1b)。

AUPRC も同様の傾向を示し、開発コホートにおいて、XGBoost の AUPRC は 0.779 (95%

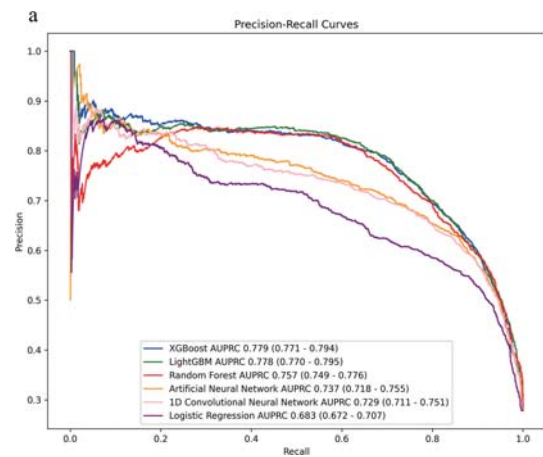
CI 0.771-0.794) であり、ロジスティック回帰、Artificial Neural Network, 1D-CNN よりも有意に高く、LightGBM および Random Forest とは有意差はなかった (図 2a)。検証コホートでは、XGBoost が同様に 0.748 (95% CI 0.630-0.846) という最も高い AUPRC を記録したが、他のモデルとの間に有意差は見られなかった (図 2b)。

キャリブレーションプロットは良好な較正能を示し、Brier スコアは 0.107 から 0.131 の範囲であった。Deep ROC 解析を用いて、開発コホートおよび検証コホートにおける各機械学習モデルのグループ正規化 AUROC を確認したところ、XGBoost と LightGBM は、偽陽性率



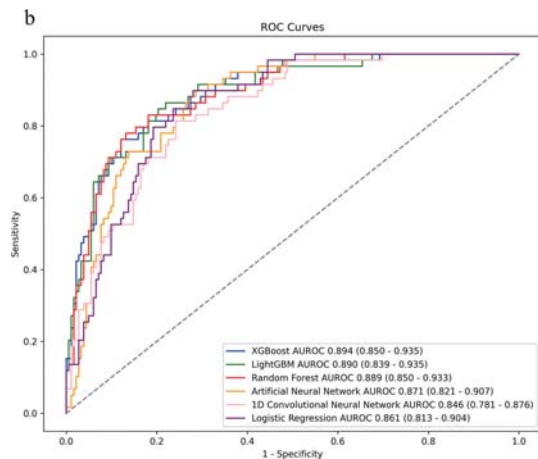
(文献 1 より引用)

図 1a 開発コホートにおける予測モデルの ROC 曲線



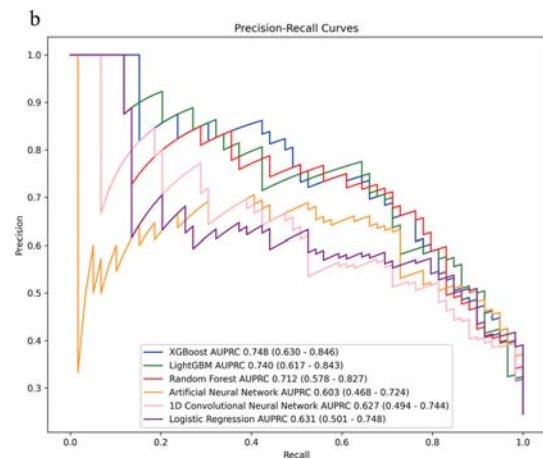
(文献 1 より引用)

図 2a 開発コホートにおける予測モデルの PR 曲線



(文献 1 より引用)

図 1b 検証コホートにおける予測モデルの ROC 曲線



(文献 1 より引用)

図 2b 検証コホートにおける予測モデルの PR 曲線

で分割された3つのグループすべてにおいて、良好なグループ正規化 AUROC を示した。

さらに、モデルの解釈のために SHAP (SHapley Additive exPlanations) 法を用い、予測に重要な変数を特定した。その結果、年齢、アルブミン、IgA/C3 比、尿中赤血球が、予測に重要な変数であることが明らかになった。感度分析として、データを分割せずに 10 分割交差検証を行った結果でも、LightGBM が最も高い AUROC を示し、XGBoost と Random Forest とは有意差がないことを確認した。

本研究では、IgA 腎症の診断における機械学習モデルの有用性を検証した。複数の機械学習モデルを評価した結果、XGBoost, LightGBM, Random Forest が従来のロジスティック回帰モデルよりも高い予測性能を示すことが明らかになった。これらのモデルは、年齢、血清アルブミン値、血清 IgA/C3 比、尿中赤血球といった既知の重要な予測因子を捉え、さらに予測因子と IgA 腎症の関連が非線形である可能性を示唆した。これは、機械学習が IgA 腎症の非侵襲的診断に有用である可能性を示唆している。

先行研究と比較して、本研究はより大規模なデータで複数のモデルを統計的に比較評価した点に強みがある。特に、ツリーベースの機械学習モデルである XGBoost, LightGBM, Random Forest の優れた性能は、他の疾患予測研究とも一致しており、これらのモデルが IgA 腎症の非侵襲的診断に有用である可能性を裏付けている。SHAP 分析によってモデルの予測根拠を解釈した結果、年齢、アルブミン、IgA/C3 比、尿中赤血球が重要な予測因子であ

ることが改めて確認され、これらの因子と IgA 腎症の複雑な関係性が示唆された。

本研究の臨床的意義として、簡便で正確な非侵襲的診断モデルが開発可能であること、日常臨床で得られるデータのみでモデルが利用できるため追加コストがかからないこと、重要な変数の特定と関係性の可視化が新たな臨床的視点を提供できることが挙げられる。一方で、単一施設データであること、サンプルサイズに限界があること、対象患者が広範であることなどが研究の限界として指摘され、今後の多施設共同研究による検証が求められる。結論として、機械学習モデルは IgA 腎症の診断予測に有用であり、特に XGBoost, LightGBM, Random Forest は、従来の統計モデルよりも高性能で、複雑なデータ構造も扱えることから、非侵襲的かつ信頼性の高い診断法として期待される。

本原稿で使用した図は全て文献 1 Sci Rep. 2024 May 30; 14(1): 12426. から引用した。

[成果の発表、論文など]

1. Noda R, Ichikawa D, Shibagaki Y. Machine learning-based diagnostic prediction of IgA nephropathy: model development and validation study. Sci Rep. 2024 May 30; 14(1): 12426.
2. Ryunosuke Noda, Daisuke Ichikawa, Yugo Shibagaki. Machine Learning Models for IgA Nephropathy Diagnosis: A Retrospective Study on Predictive Performance and Influential Variables. ASN Kidney Week. Nov 2023.
3. 野田竜之介, 市川大介, 柴垣有吾. 機械学習による血液・尿検査を用いた IgA 腎症の診断予測. 第 53 回日本腎臓学会東部学術大会. 2023 年 9 月.