

主論文の要約

**Artificial intelligence in a prediction model for  
postendoscopic retrograde cholangiopancreatography  
pancreatitis**

〔人工知能を用いたERCP後膵炎リスク予測モデルの構築〕

名古屋大学大学院医学系研究科 総合医学専攻  
病態内科学講座 消化器内科学分野

(指導：川嶋 啓揮 教授)

高橋 秀和

## 【緒言】

内視鏡的逆行性胆道膵管造影 (endoscopic retrograde cholangiopancreatography; ERCP) は、胆膵疾患の診断および治療のために広く行われている検査である。しかし、ERCP には合併症のリスクが伴い、ERCP 後膵炎が最も高頻度で問題となっている(発生率 3-10%, 死亡率 0.7%)。これまで女性・膵炎既往・膵管造影など多様なリスク因子が報告されているものの、実臨床の患者は複数因子を同時に有しているため統合的に判断する必要があった。従来スコアリングシステムについては、ほとんどが単施設の検討に留まり、また因子間の非線形な関係性までは評価していなかった。患者個別のリスクを把握することは、適切な予防的処置や重症化予防だけでなく、限られた医療資源の適正配分に寄与する。そこで近年さまざまな分野で注目されている AI 技術を用いて ERCP 後膵炎リスクの予測モデルを構築し、その有用性を検討することを目的とした。

## 【対象及び方法】

2018 年 1 月から 2020 年 12 月までに、当院およびトヨタ記念病院で ERCP を受けた 20 歳以上の患者を後方視的に検討した。十二指腸乳頭処置後・Billroth-I 法を除く胃再建術後・カニューレション不可・検査前に膵炎を発症していた症例は除外した。ERCP 後膵炎の診断は改訂アトランタ分類に基づいて実施した。既報のリスク因子を含むように患者側・手技側の合計 40 因子を収集し、統計学的検討を行った。さらにランダムフォレストアルゴリズム (RF model) と多重ロジスティック回帰分析 (LR model) を用いて ERCP 後膵炎発症予測モデルを構築した。ランダムフォレストは機械学習手法の一種で、多数の決定木を組み合わせて動作するアンサンブル学習アルゴリズムである。特徴量選択については、RF model は wrapper method を、LR model は forward stepwise selection method を用いた。当院コホートは 3 分割交差検証で Training と Validation を行い、トヨタ記念病院コホートで Test した。さらに RF model のリスクスコアから low-・medium-・high-risk 群を抽出し、各群の ERCP 後膵炎発症率を比較した。モデルの精度は area under the receiver operating characteristic curve (AUROC) で評価した。

## 【結果】

当院コホート (n=615)、トヨタ記念病院コホート (n=554) について検討した (Figure 1)。当院では悪性疾患をはじめ多様な疾患に対する検査が実施されており、トヨタ記念病院では 70%以上が総胆管結石と common disease が多くを占めていた。一方、施設間の ERCP 後膵炎発症率に有意な差は認めなかった (当院 10.2%, トヨタ記念病院 9.6%) (Table 1)。当院コホートの単変量解析結果では、ERCP 後膵炎群において血小板値・胆道癌患者の割合・膵管ガイドワイヤー法や胆管生検実施例が多い傾向にあった (Table 2)。RF model の特徴量は Training set の AUROC を最大化するように 21 因子が選択されたが、out-of-bag classification error の増加も見られたため過学習状態と考えられた。そこでエラー率が最も低い 10 因子までの特徴量を選択した (血清アルブミン値,

血清クレアチニン値, 胆道癌, 膵癌, 総胆管結石, 検査時間>30 分, 膵管造影, 膵管ガイドワイヤー法, IDUS, 胆管生検) (Figure 2, 3)。LR model では単変量解析で  $P<0.10$  の 8 因子からさらに 4 つの特徴量を選択した(胆道癌, 膵管ガイドワイヤー法, EST, EPBD) (Table 3)。Validation・Test ともに、RF model は LR model より有意に精度が高かった(平均 AUROC 0.821 vs. 0.660; AUROC 0.770 vs. 0.663,  $P=0.002$ ) (Figure 4)。RF model から算出されたリスクスコアを用い、ERCP 後膵炎発症率に応じた層別化が可能であった (Validation: low-risk 0.0%, medium-risk 4.5%, high-risk 28.9%; Test: 2.9%, 10.0%, 23.9%)。Test set では、各 risk 群に統計学的な有意差を認めた (low-risk vs. medium-risk:  $P=0.011$ ; medium-risk vs. high-risk:  $P=0.007$ ) (Figure 5)。

### 【考察】

本研究では、機械学習アルゴリズムを用いて ERCP 後膵炎リスク予測モデルを構築した。3 分割交差検証および外部検証において RF model は LR model より優れた予測性能を示した。また患者を low-risk・medium-risk・high-risk 群に分類することが可能であった。膵炎発症には、複数因子が複雑に関連しており、非線形な因子の関係性まで評価できるアルゴリズムが適していた可能性がある。また out-of-bag classification error を最小化するように特徴量を選択したことが汎化性の高いモデル構築に貢献したと考えられる。本研究は、患者背景の異なる施設で外部検証がされていることに強みがある。患者個別のリスクを予測することによって、high-risk 群に対しては、膵管ステント留置や NSAIDs 坐剤などの予防的処置を積極的に検討し、注意深くモニタリングすることで膵炎の早期発見や重症化予防に役立てることができる。一方で low-risk 群には、早期食事再開やリハビリテーションを検討し、不要な入院期間を減らすことが可能となる。ERCP 症例には多様な患者背景や手技が含まれているため、それぞれのリスクに応じて治療方針を選択することが個別化医療に必要である。Limitation として、本検討には除外基準が設けられており、すべての患者を対象にしたものではない。またランダムフォレスト以外にも数多くの機械学習アルゴリズムが存在しており検討の余地がある。今後、多施設からさらに症例を収集することでより高性能なモデル構築が望まれる。また医療者はその機械学習モデルの特性を理解することで、臨床のパフォーマンスを最大化するように利用すべきであると考ええる。

### 【結論】

機械学習を用いた高精度の ERCP 後膵炎リスク予測モデルを構築した。このモデルはリスクの層別化においても有用であった。個別のリスクを正確に把握することで適切な治療方針の選択および医療資源の配分が期待される。