

厚生労働科学研究費補助金 難治性疾患政策研究事業
難治性の肝・胆道疾患に関する調査研究
分担研究報告書

機械学習による原発性胆汁性胆管炎患者の予後および
予後規定因子予測モデルの作成

研究協力者 寺井 崇二 新潟大学大学院医歯学総合研究科消化器内科学分野 教授

研究要旨：原発性胆汁性胆管炎（PBC）は、肝硬変や肝不全に至る難治性疾患である。ウルソデオキシコール酸が第一選択であるがその治療反応性、掻痒感などの症状や肝細胞癌、消化管静脈瘤といった合併症によって予後が大きく左右される。本研究では、従来の統計学では困難だった PBC 患者の予後のみならず症状、合併症を機械学習（ML）の一つである transformers を用いて高精度に予測するモデルを開発したい。このモデルは、少ないデータで高精度な予測が可能と報告されており、さらに同一データから PBC 患者の生命予後、合併症を個別に予測可能でありテーラーメイド治療や不必要な検査の削減による医療費削減にも貢献できると期待される。

共同研究者

木村 成宏 新潟大学医歯学総合病院

A. 研究目的

原発性胆汁性胆管炎（PBC）は、自己免疫性肝疾患であり、肝内胆管が慢性・進行性に障害され、肝硬変や肝不全に至る可能性がある。ウルソデオキシコール酸（UDCA）は、1997年に米国食品医薬品局（FDA）により PBC の治療薬として承認され、胆汁鬱滞性疾患の第一選択薬となっている。しかし、20～40% 程度の患者さんが治療に反応せず、これらの症例では肝硬変や肝不全へと進行し予後不良となることが知られている。治療反応性が不良な症例では黄疸や掻痒感などの症状が出現する可能性が高くなり、有症状症例では無症状例と比較し予後不良であると言われている。また、門脈圧亢進症状が他の肝疾患よりも進行しやすいことも報告されており、消化管静脈瘤の発生および出血により時に致命的となる。さ

らに肝細胞癌も少ないながら発症することもある。一般には病状進行は緩徐であるが、症例によっては数年で急激に進行することもある。このように PBC ではさまざまな症状を呈することが判明しており、各症状によって予後不良となることが報告されている。そのため予後予測においては全体の傾向を掴むことも重要であるが、各患者がどのような症状で予後が規定されるかも非常に重要である。これまで疾患包括的な予後予測モデルは報告されているものの、各症状の出現を並行して予測するモデルは無かった。

これらの予後予測を行うことは、ひとつの症状を予測することであっても多数の因子が関与するため、従来の統計学では正確な予測は困難であった。しかし人工知能（AI）を発展させた機械学習（ML）は予測精度を高める一つの方法であることが明らかとなり、ML を用いた未来予測の有用性は、消化器病を含む様々な医学分野で報告

されている。MLではこれまでいわゆる深層学習であるディープニューラルネットワーク(DNN)などの大規模データを要するモデルが隆盛であった。しかしDNNでは10万を超えるデータが必要であり、さらに計算に時間を要することがデメリットであった。一方でTransformersを用いたモデルは並列処理が可能であることから大量のデータを高速に処理可能であり、さらにPre-trainingモデルを用いることでDNNと比べてモデル作成に必要なデータは少なくすむというメリットがある。それに加えてDNNと比べても正確性が高いことが報告されており、これらの報告に基づき、Transformersを用いたMLはPBC症例の予後予測に有用であると考へた。また、予測対象を予後だけでなく消化管静脈瘤や肝細胞癌、症状などへ変更することで同一データから予後や治療方針に影響する因子の予測も可能であると思われる。

本研究では、MLを用いてPBC症例の生命予後を予測するMLモデルを開発・検証したい。著者らは、新規のMLモデルを用いることで症例に応じたテーラーメイド的な治療方法を提案可能であり、総じて予後改善効果が期待できるものと考えている。さらに各症例における合併症のリスクに応じた検査を行うことで不必要な検査を回避可能であり、今後も増加する医療費削減にも貢献できる可能性がある。

B. 研究方法

本研究では二つの要因が重要であると考えられる。一つは精度が高いMLモデル、もう一つは十分なデータサンプル量である。TransformersをベースとしたMLモデルを作成する予定である。TransformersはGoogle AIが開発した自然言語処理モデルであり、従来のMLモデルと比較して①事前

学習済みであり、新規のデータに対してさらに多くのデータを必要としない、②汎用性が高く、さまざまなタスクに応用がきく、③精度が高い、といった特徴がある。このモデルは2017年に発表されたが、その後Chat-GPTなどへと応用され、2024年においても頻用されている。機械学習は非常に短いスパンでより良いモデルが報告されているものの、このtransformersベースのモデルは上記のように発表後7年経過しても第一線で活躍を続けている。このMLモデルであれば今後も十分に使用に値するモデルが作成できるものと思われる。

データに関して、申請者は厚生労働科学研究費補助金、難治性疾患政策研究事業難治性の「肝・胆道疾患に関する調査研究」田中班に所属しており、本邦で集積しているPBC症例のデータベースを利用可能となっている。おおよそ1万件の症例データを利用可能であり、予後のみならず合併症の経過も利用可能であり、本研究での目的に合致したデータベースである。DNNなどの深層学習ではその10倍以上のデータが必要になると言われているが、上記のようにtransformersの利用であれば必要十分なデータ数であると思われる。

このように本研究で必須の要件はすでに利用可能であり、十分に研究の遂行が可能であると思われる。

(倫理面への配慮)

ヘルシンキ宣言及び「人を対象とする医学系研究に関する倫理指針」を遵守して実施する。本研究により知り得た個人情報、個人情報保護法や関係各倫理指針に基づき、細心の注意を払い取り扱う。

C. 研究結果

研究報告書作成時点ではデータクリーニ

ング及びモデル作成を行なっているが、研究結果の発表ができる段階ではない。

D. 考察

特になし。

E. 結論

特になし。

F. 研究発表

1. 論文発表

特になし。

2. 学会発表

特になし。

G. 知的財産権の出願・登録状況

(予定を含む。)

1. 特許取得

特になし。

2. 実用新案登録

特になし。

3. その他

特になし。