

令和 6 年 6 月 3 日現在

機関番号：32665

研究種目：若手研究

研究期間：2022～2023

課題番号：22K15347

研究課題名（和文）機械学習アプローチによる薬物性肝障害の発症に寄与する未知の薬物間相互作用の検出

研究課題名（英文）Detection of potential drug-drug interactions for drug-induced liver injury using machine-learning algorithms

研究代表者

秋元 勇人（AKIMOTO, Hayato）

日本大学・医学部・助教

研究者番号：80847658

交付決定額（研究期間全体）：（直接経費） 400,000円

研究成果の概要（和文）：高齢化により多剤処方を受ける高齢患者数の増加が予想されている。薬物の代表的な副作用の一つとして肝障害があり、多剤処方を受ける患者では肝障害のリスクを高める複数の薬物が意図せず処方されているかもしれない。そこで本研究では医療機関で集積される電子カルテ情報を機械学習により解析することで、肝障害のリスクを高める未知の薬物の組み合わせ（薬物間相互作用）を探索した。その結果、機械学習アルゴリズムは従来の手法である線形回帰分析よりも優れた予測性能を示し、このアルゴリズムは臨床において高頻度で使用されているジクロフェナクとファモチジンの組み合わせが肝障害の発症リスクを高めることを示唆した。

研究成果の学術的意義や社会的意義

本研究では電子カルテ情報を機械学習により解析することで、薬物誘発性肝障害発症リスクに関連する潜在的な薬物間相互作用の検出を試みた。機械学習アルゴリズムはOTC医薬品としても使用されているジクロフェナクやファモチジンの組み合わせが肝障害のリスクを高めることを示唆し、医療機関に受診している患者のみならず、セルフメディケーションを実施している国民へも一定の注意を払う必要があると思われる。また従来の薬剤疫学的薬物間相互作用は線形回帰分析や不均衡分析により検出されてきたが、機械学習を用いた検出はこれらの手法よりも高い汎化性を示しつつ相互作用の検出が可能な手法として今後活用が期待できる。

研究成果の概要（英文）：The number of elderly patients receiving polypharmacy tends to increase by the aging of Japan's population. One of the most common adverse drug events is liver injury, and these patients receiving polypharmacy may be unintentionally prescribed drugs that increase the risk of liver injury. In this study, potential drug-drug interactions for increased risk of drug-induced liver injury were identified from electronic medical records using machine learning models. The machine learning models such as the logistic least absolute shrinkage and selection operator (LASSO) model and the gradient boosting decision tree model showed better predictive performance than multiple logistic regression model, one of the traditional statistical models, and suggested that the combination of diclofenac and famotidine significantly has been associated with an increased risk of drug-induced liver injury.

研究分野：薬剤疫学

キーワード：薬物間相互作用 薬物性肝障害 機械学習 人工知能 リアルワールドデータ

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

日本社会が直面している高齢化は今後も進行し、2045年に高齢者数は最多となることが予測されている。加えて、後期高齢者の約65%は何らかの併存疾患を複数有しているとされている。したがって、本邦においては今後、高齢者数の増加に伴い複数の併存疾患に罹患する高齢者数も増加し、これらの併存疾患を治療するために多剤処方を受ける高齢者もまた増加するであろうことは想像に難くない。薬物性肝障害(DILI)は広範な薬物で見られる薬物性副作用の1つであり、DILIのリスクを有する薬物を2剤以上併用するとそのリスクは6倍増加することが報告されている。以上のことから、高齢化が進行する本邦では複数の併存疾患を有する高齢患者への多剤処方において、意図せずDILIリスクが高まるような薬物の組み合わせが処方されるかもしれない。このようなDILIリスクを相乗的に高める未知の薬物間相互作用を探索することは薬物の処方設計において重要である。

2. 研究の目的

薬剤疫学的な薬物間相互作用を検出するために、米国食品医薬品局が公開している有害事象自発報告システム FAERS や世界保健機関が公開している副作用自発報告システム Vigibase™ 等がしばしば活用されている。しかしながら、これらのデータは患者自身の主観的な報告が多数を占めており、報告バイアスが存在する。また、薬物間相互作用を検出するために用いられる代表的な統計手法には、分割表を用いた不均衡分析や多重ロジスティック回帰分析(MLR)のような線形回帰モデルがある。しかしながら、これらの手法は訓練データに対して過学習を引き起こしやすく、テストデータ(未知のデータ)に対して脆弱になる傾向にある。近年、コンピュータの処理能力の向上と人工知能の基盤技術の1つである機械学習アルゴリズムの発達により、個人のコンピュータでも機械学習モデルを構築することが容易になりつつある。機械学習は訓練データへの過学習を抑制しつつ、未知データの予測能力を高めることができ、さらには構築された機械学習モデルを解釈することでどの因子がモデルの予測に寄与しているかを評価することも可能である。したがって、訓練データへの過学習を抑制したDILIの発症を予測する機械学習モデルを構築し、そのモデルを解釈することにより未知の薬物間相互作用を検出することが可能となる。そこで本研究の目的は、患者による主観に依らないデータである電子カルテ情報を機械学習で解析することによってDILI発症予測モデルを構築し、そのモデルからDILI発症リスクを増大させる未知の薬物間相互作用を検出することを目的とする。

3. 研究の方法

本研究ではデータソースとして日本大学医学部において蓄積された電子カルテ情報データベース(NUSM CDW)を利用する。NUSM CDWは匿名加工された患者基本情報、診断情報、投薬情報、検査値情報の4つの情報源からなるデータベースであり、日本大学系列の3病院で収集された約250万人分の患者情報を有する大規模なリアルワールドデータである。本研究ではまず、NUSM CDWから3か月以内に3回以上肝機能検査を受け、初回及び2回目の血清アラニンアミノ基転移酵素(ALT)値が基準値[4.0-44.0 U/L]範囲内の患者122,285名分の患者情報を抽出した。抽出された患者のうち、3回目あるいはそれ以降に測定されたALT値が基準値上限の3倍を超えた患者を症例群(Case group, N=14,634)、いずれの測定時点においても基準値範囲内であった患者を対照群(Control group, N=80,484)とし、これをアウトカムとした。次に、以下の除外基準に適合した患者を除外した計54,261名分(Cases=4,152; Controls=54,261)の患者情報を得た。54,261名分の患者の基本情報(年齢、性別、3病院)、5つの既往歴(高血圧、糖尿病、脂質異常症、心不全、敗血症)の有無、180種類の薬物の使用状況(2値変数)、100人以上の患者において併用されていた2種薬物の組み合わせ1080パターン(積項; 2値変数)、観察期間中に併用されていた薬物の数(併用薬物数)の計1,375変数を特徴量とした。この54,261名分のデータセットを機械学習で解析するために訓練データ(70%; N=40,889)とテストデータ(30%; 17,524)に分割した。

次に未知の薬物間相互作用を検出するために、ALTの異常な上昇を予測する機械学習モデルを構築した。本研究では、従来手法であるMLRモデルに加え、MLRの損失関数中にL1正則化項を加えたLogistic Least Absolute Shrinkage and Selection Operator (LASSO) regressionモデル、決定木ベースのアルゴリズムである勾配ブースティング決定木(GBDT)モデルの3つの機械学習モデルを構築した。MLRモデルではアウトカムの対数オッズを従属変数とし、1,375特徴量を説明変数とした。Logistic LASSO regressionモデルの構築には統計解析ソフトRのglmnetパッケージを使用し、MLRと同様の従属変数及び説明変数を設定後、10分割交差検証法により、訓練データにおける誤分類率が最小となるような値(最適化値)を決定した。GBDTモデルの構築にはRのxgboostパッケージを使用した(以下、GBDTモデルをxgboostモデルと呼ぶ)。xgboostモデルには調整すべきパラメータ(ハイパーパラメータ)が複数あり、本研究ではこれらのハイパーパラメータを10分割交差検証法及びグリッドサーチによりチューニングした。チューニング手順及びサーチ空間を以下のようにした。

Steps in parameter tuning are as follows:

- 1) Learning rate (eta, default=0.3) was fixed at 0.1.
- 2) Ten-fold cross-validation for the training set was performed using the eta and other default parameters to determine the best iterations ('nrounds' , set at 343).
- 3) Grid search was performed with fixed learning rate and the best iterations to determine the rest of the hyperparameters minimizing misclassification error rate for the training set.

Search space is as follows:

- 'max_depth' : {2, 4, 6, 8, 10}
- 'min_child_weight' : {ranged from 1 to 20 by 1}
- 'gamma' : {ranged from 0 to 20 by 1}
- 'colsample_bytree' : {ranged from 0.5 to 1.0 by 0.1}
- 'subsample' : {ranged from 0.5 to 1.0 by 0.1}

構築された 3 つの機械学習モデルの判別性能と頑健性を評価するために受信者動作特性曲線下面積(AUROC)及び Precision-Recall 曲線下面積(AUPR)を、モデルの較正を評価するために較正曲線の傾き及び切片を、モデルが出力する予測確率の正確さの指標として Brier score をそれぞれ算出した。

本研究の目的である未知の薬物間相互作用の候補を探索するために、Logistic LASSO regression モデルにおいては 180 種類の個々の薬物の偏回帰係数と興味ある積項の偏回帰係数を使用して Relative Excess Risk due to Interaction (RERI)を算出した。さらに、偏回帰係数及び RERI の 95%信頼区間を算出するために、2,000 回のブートストラップサンプリングを実施し、95%ブートストラップパーセンタイル信頼区間(95%BootCI)を算出した。xgboost モデルにおいて薬物間相互作用を検出するために全特徴量の特徴量重要度(feature importance)を算出した。

4. 研究成果

本研究では線形回帰手法である MLR モデル及び Logistic LASSO regression モデル、非線形手法である xgboost モデルの 3 つの ALT 値異常上昇予測モデルを構築した。これら 3 つの予測モデルの判別性能、頑健性及びモデルの較正を図 1 に示す。図 1A,B には訓練データ及びテストデータに対する 3 つのモデルの AUROC 及び AUPR を示した。テストデータにおける判別性能について、Logistic LASSO regression モデルと xgboost モデルは MLR よりも優れていた。また図 1C,D に示したように、Logistic LASSO regression モデルと xgboost モデルは訓練データのサンプルサイズが少ない条件下でもテストデータに対して優れた判別能力を保有していた。一方で MLR モデルは少ないサンプル数ではテストデータをうまく判別することが困難であった。モデル較正について、較正曲線の傾きは 1.0 に近いほど、切片については 0.0 に近いほど良好であり、Brier score はより低値である方が正確に予測確率を算出していることになる。図 1E-G において MLR モデルは傾きが小さく、切片は大きい。また Brier score は他の 2 つのモデルよりも高値であった。以上のことから、Logistic LASSO regression モデルや xgboost モデルの ALT の異常な上昇の予測性能は MLR モデルよりも優れていることが明らかとなった。

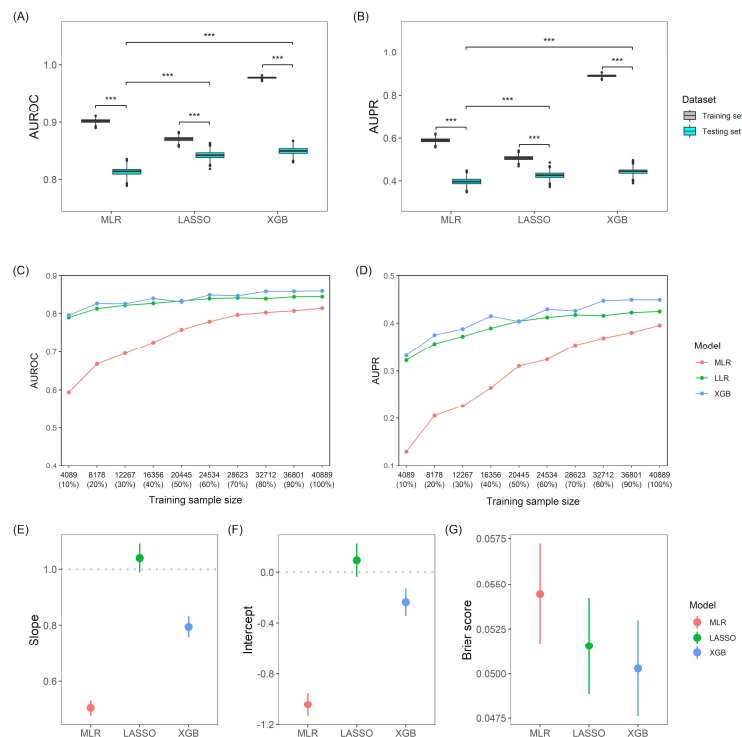


図 1. 3 つの機械学習モデルの判別性能、頑健性、モデル較正の比較。

次に、予測性能が MLR よりも優れており、線形回帰アルゴリズムである Logistic LASSO regression モデルにおける薬物間相互作用の候補を図 2A に示す。図 2A の Product term 欄には Logistic LASSO regression モデルに投入された積項のうち、95%BootCI 下限値が 1.0 よりも大

きかった3つの積項の調整済みオッズ比(Adjusted Odds Ratio; Adjusted OR)を示しており、Individual drug 欄にはこれら3つの積項を構成する個々の薬物6種の Adjusted OR をそれぞれ示している。これら3つの積項(組み合わせ)のうち、アスピリンとシロスタゾールの組み合わせは積項の Adjusted OR の下限値が1.0よりも大きいと算出されたが(Adjusted OR 1.232; lower limit 1.143), アスピリン及びシロスタゾールの個々の Adjusted OR は1.0を下回っていた。一般に、個々の効果が1.0を下回ると RERI を用いた相互作用の検出はできない。一方で、ジクロフェナクとファモチジン、アセトアミノフェンとアンブロキシソールの組み合わせでは、積項の Adjusted OR が1.0を上回っていることに加え、個々の薬物の Adjusted OR もまた1.0を上回っている。とりわけジクロフェナクとファモチジンは個々の Adjusted OR の下限値も1.0を上回っていた。この2つの組み合わせ(ジクロフェナク+ファモチジン、アセトアミノフェン+アンブロキシソール)について、Adjusted OR から算出された RERI を

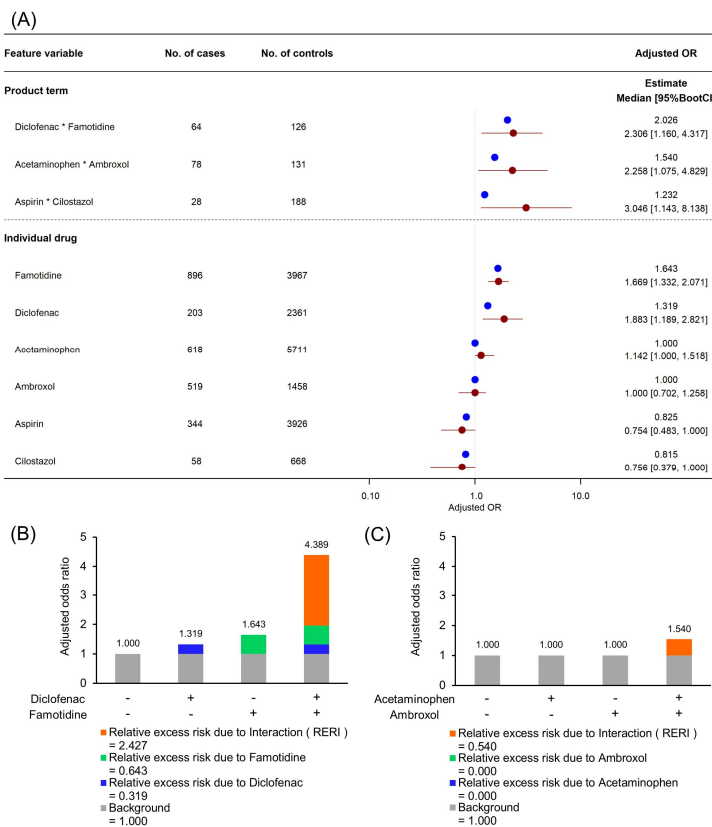


図2. Logistic LASSO regression モデルから算出された RERI に基づく薬物間相互作用の候補。

図2B,Cにそれぞれ示す。図2Bはジクロフェナクとファモチジンの個々の ALT 異常上昇リスク(青及び緑)と相互作用による上昇リスクのさらなる上昇(橙色)を示している。例えば、この2種薬物を両方とも使用していないときの ALT 異常上昇リスクを1.0倍とすると、ジクロフェナク単独では1.3倍、ファモチジン単独では1.6倍であり、この2つを併用すると ALT 異常上昇リスクは4.4倍になる。図2Cにはアセトアミノフェンとアンブロキシソールの個々の ALT 異常上昇リスク(両方ともに1.0倍)と相互作用による上昇リスクのさらなる上昇(橙色)が示されている。この組み合わせでは、薬物単独ではALTの異常な上昇リスクに影響はないものの、併用したときにリスクが上昇することを示している(ただし、その程度は比較的小さいものであった)。

xgboost モデルは本研究で構築された3つの機械学習モデルの中で、テストデータに対して最も高い予測性能や頑健性を有していた。このモデルにおける ALT の異常な上昇イベントの予測に寄与する特徴量重要度(feature importance)の上位50変数を図3に示す。ALTの異常な上昇に寄与する特徴量は特徴量の重要度が高い順に、併用薬剤数、年齢、フェンタニルの使用、性別、カルバペネム系抗菌薬の使用であった。この上位50変数中には、赤色で示したように、ヘパリンナトリウムとフロセミド間、ジクロフェナクとファモチジン間の積項が含まれており、これら4つの個々の薬物の特徴量重要度を緑色で示した。これら2つの積項及び個々の4つの薬物は1,375変数で構築されたxgboostモデルにおける特徴量重要度の上位50位以内に入っており、ALTの異常な上昇イベントの予測に一定程度寄与していることがわかった。この2つの積項のうち、ジクロフェナクとファモチジンの組み合わせは Logistic LASSO regression モデルのような線形アルゴリズムだけでなく、xgboost モデルのような非線形なアルゴリズムにおいても ALT の異常な上昇に寄与することが確認され、個々の薬物においても線形アルゴリズムと同様、ALTの異常な上昇の発症

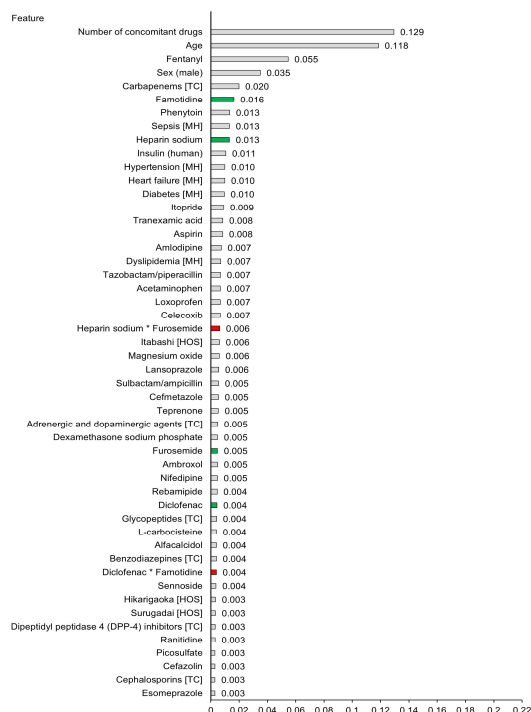


図3. xgboost モデルにおける ALT の異常上昇イベントの予測に寄与する相対的に重要な特徴量。

に寄与していることがわかった。

以上のことから、機械学習アルゴリズムのリアルワールドデータへの適用は伝統的な統計解析手法である線形回帰分析よりも未知データに対して、より良好にフィッティングすることがわかる。さらには構築された機械学習モデルがどのようにして興味あるイベントの発生予測をしているか、すなわち発生予測にどの因子が寄与しているかを評価可能である。本研究ではこの技術を活用し、リアルワールドにおいてジクロフェナクとファモチジン間に ALT の異常な上昇リスクに寄与する相互作用があることを見出した。機械学習のリアルワールドデータへの適用は薬物性肝障害の潜在的な相互作用の検出に限定されるものではなく、リアルワールドにおける薬物の有効性や安全性の評価に幅広く応用可能であることを確信する。

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 2件 / うち国際共著 0件 / うちオープンアクセス 2件）

1. 著者名 Akimoto Hayato, Nagashima Takuya, Minagawa Kimino, Hayakawa Takashi, Takahashi Yasuo, Asai Satoshi	4. 巻 13
2. 論文標題 Detection of synergistic interaction on an additive scale between two drugs on abnormal elevation of serum alanine aminotransferase using machine-learning algorithms	5. 発行年 2022年
3. 雑誌名 Frontiers in Pharmacology	6. 最初と最後の頁 910205
掲載論文のDOI（デジタルオブジェクト識別子） 10.3389/fphar.2022.910205	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Akimoto Hayato, Hayakawa Takashi, Nagashima Takuya, Minagawa Kimino, Takahashi Yasuo, Asai Satoshi	4. 巻 14
2. 論文標題 Detection of potential drug-drug interactions for risk of acute kidney injury: a population-based case-control study using interpretable machine-learning models	5. 発行年 2023年
3. 雑誌名 Frontiers in Pharmacology	6. 最初と最後の頁 1176096
掲載論文のDOI（デジタルオブジェクト識別子） 10.3389/fphar.2023.1176096	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計1件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 秋元勇人、長島卓也、早川 隆、皆川貴美乃、高橋泰夫、浅井 聡
2. 発表標題 肝障害のリスクに相乗的に影響を及ぼす薬物間相互作用の探索：機械学習アルゴリズムを用いたシグナル検出
3. 学会等名 日本薬学会第143年会
4. 発表年 2023年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

6. 研究組織

氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------