



Nagoya City University Academic Repository

学位の種類	博士（薬学）
報告番号	甲第2037号
学位記番号	第420号
氏名	渡邊 崇
授与年月日	令和6年3月22日
学位論文の題名	副作用報告データベースを用いた添付文書の改訂を支援する機械学習モデルの開発
論文審査担当者	主査： 牧野 利明 副査： 頭金 正博, 桑 和彦, 館 知也

名古屋市立大学 学位論文

副作用報告データベースを用いた
添付文書の改訂を支援する機械学習モデルの開発

令和5年度（2024年 3月）

所属： 名古屋市立大学大学院薬学研究科
レギュラトリーサイエンス分野

氏名： 渡邊 崇

本論文は、令和 6 年 3 月名古屋市立大学大学院薬学研究科において審査されたものである。

主査 牧野 利明 教授
副査 頭金 正博 教授
副査 条 和彦 教授
副査 舘 知也 教授

本論文は、学術情報雑誌に掲載された次の報文を基礎とするものである。

1. T.Watanabe, K.Ambe, M.Tohkin

Predicting the addition of information regarding clinically significant adverse drug reactions to Japanese drug package inserts using a machine-learning model

Ther Innov Regul Sci. (in press)

<https://doi.org/10.1007/s43441-023-00603-4>

2. T.Watanabe, K.Ambe, M.Tohkin

Streamlining considerations for safety measure: A predictive model for addition of clinically significant adverse reactions to Japanese drug package inserts

Biol. Pharm. Bull. (accepted)

本論文の基礎となる研究は、頭金 正博 教授の指導の下に名古屋市立大学大学院薬学研究科において行われた。

目次

略語一覧.....	1
要旨.....	2
序論.....	8
本論.....	11
第 1 章：国内副作用症例の集積を根拠とした添付文書の「重大な副作用」欄への 追記を予測する機械学習モデルの構築	11
1-1 背景	11
1-2 方法	12
1-3 結果	18
1-4 考察	25
1-5 小括	28
第 2 章：安全対策措置の検討を効率的に行うための汎用的予測モデルの構築	30
2-1 背景	30
2-2 方法	30
2-3 結果	38
2-4 考察	43
2-5 小括	45
総括.....	46
謝辞.....	49
引用文献.....	50
Supplemental.....	53

略語一覧

ADR	Adverse drug reaction 副作用
AUC	Area under the curve
CCDS	Company core data sheet 企業中核データシート
CSAR	Clinically significant adverse reaction 重大な副作用
EU	European Union 欧州連合
EFS	Exhaustive feature selection
FAERS	U.S. Food and Drug Administration Adverse Event Reporting System FDA有害事象報告システム
FDA	U.S. Food and Drug Administration アメリカ食品医薬品局
JADER	Japanese Adverse Drug Event Report 日本の医薬品副作用データベース
LGB	Light gradient boosting machine
MCC	Matthews correlation coefficient マッシュューズ相関係数
MedDRA	Medical Dictionary for Regulatory Activities 国際医薬用語集
NN	Neural network
PMDA	Pharmaceuticals and Medical Devices Agency 医薬品医療機器総合機構
PRR	Proportional reporting ratio
PT	Preferred term 基本語
RBF-SVM	Support vector machine with radial basis function
ROR	Reporting odds ratio
SD	Standard deviation 標準偏差
SMQ	Standardized MedDRA Queries MedDRA標準検索式
XGB	Extreme gradient boosting

要旨

【序論】

独立行政法人医薬品医療機器総合機構（PMDA）は、医薬品や医療機器の製造販売後の安全性に関する情報として、製薬企業、医療機関及び患者から報告された副作用に関する報告や外国で行われた安全対策措置報告、論文や学会報告等の研究報告等から得られた安全性情報を幅広く収集している。PMDAはこれらの収集した情報から既知あるいは未知の副作用が発生している兆候（リスクシグナル）を検知すると、収集した安全性情報を評価及びスクリーニングし、製薬企業や専門家との議論を経て最終的に添付文書改訂等の安全対策措置の要否を決定している。このような一連のワークフローには多くの人員と時間を要しており、リスクシグナルを検知してから最終的な安全対策措置を実施するまでに一般的には数か月程度を要している。このリスクシグナルを効率的に検出するための技術として、現在統計的シグナル検出手法が用いられている。これは、薬剤と副作用の組み合わせ件数を計算し、注目する医薬品の注目する副作用が、データ全体と比較して高頻度に認められるか否かを検討しており、不均衡分析とも呼ばれている。代表的な不均衡分析手法としてproportional reporting ratio（PRR）やreporting odds ratio（ROR）が知られており、これらの不均衡分析手法はPMDA、製薬企業等でシグナルの検出のために汎用されている。しかし、不均衡分析で検出されたシグナルのうち、より詳細な検討へ進めるべきと判断されるものは2.1%という報告がされており、不均衡分析で検出されたシグナルには多くの偽陽性が含まれていることが報告されている¹⁾。よって幅広くリスク候補であるシグナルを検出する従来の不均衡分析に加えて、正確性の高い、安全対策措置要否の意思決定を効率化させるツール開発が求められている。

そこで本研究では、安全対策措置として実施されることが多い添付文書への「重大な副作用」欄への追記に注目し、「重大な副作用」欄の改訂を早期に検出可能な予測モデルの構築に取り組んだ。第一章では、「重大な副作用」が添付文書へ追記される理由の中で、根拠とされる頻度の最も高い国内副作用症例の集積に基づく医薬品と副作用の組み合わせ情報に注目し、国内副作用症例の集積に基づく「重大な副作用」欄の改訂を高い精度で早期に予測可能な機械学習モデルの構築を検討した。第二章では、第一章で構築した予測モデルでは考慮していなかった海外の副作用情報や、各国の添付文書を作成する際の基準となる文書である企業中核データシート（CCDS）の改訂といった「重大な副作用」欄の追記理由にも注目し、第一章で構築した予測モデルと相補的な関係にある、国内副作用症例の集積に限定しない汎用的な機械学習予測モデルの構築を検討した。これら予測モデルを構築することで、収集した安全性情報から直接添付文書の重大な副作用の項に追記される可能性の高い医薬品と副作用の組み合わせ情報を

早期に予測することが可能となり、迅速かつ効率的に安全性対策措置の要否の意思決定を支援することが可能となる。また、予測モデルにおいて重要と判断された機械学習の特徴量を明らかにすることで、これまで医薬品と副作用の関連性を評価してきた指標や、安全対策措置の要否を決定する際に重要と考えられてきた情報が予測に寄与しているのかを評価した。

【本論】

第1章：国内副作用症例の集積を根拠とした添付文書の「重大な副作用」欄への追記を予測する機械学習モデルの構築

我が国の添付文書での「使用上の注意」欄の改訂の81%が「重大な副作用」欄での追記であり、うち63%が国内副作用症例の集積を理由としていたと報告がされている²⁾。そこで第一章では、主要な安全対策措置である国内症例の情報を根拠とする添付文書の「重大な副作用」欄の追記を予測対象とし、安全対策措置要否の意思決定を支援する効果的な機械学習ベースの予測モデルの構築を行った。

予測対象の正解データは、2011年8月から2020年3月の間に国内症例の集積を理由として「重大な副作用」欄へ追記された医薬品と副作用の組み合わせのうち、添付文書の「重大な副作用」の項に記載された名称と国際医薬用語集（MedDRA）の基本語（PT）が一致したものとした。不正解データは、2020年3月時点で本研究の対象となった重大な副作用が添付文書に記

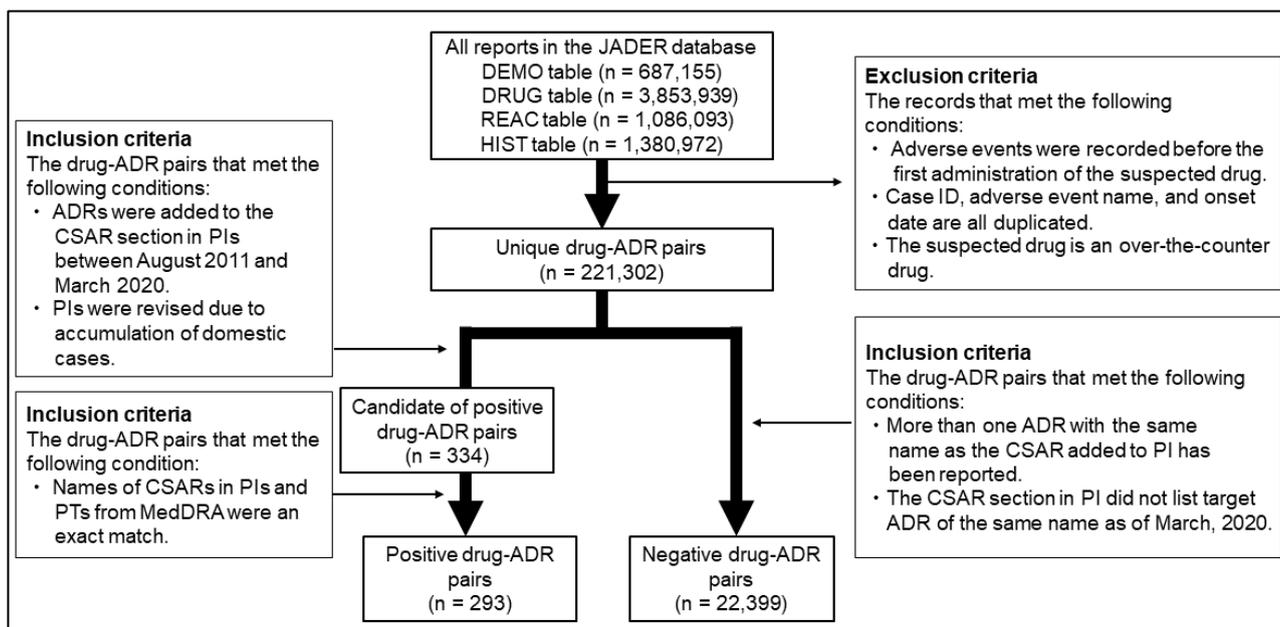


図1 国内副作用症例の集積に基づく予測モデルの構築に使用したデータセットの選択・除外基準

JADER, Japanese Adverse Drug Event Report; ADR, adverse drug reaction; CSAR, clinically significant adverse drug reaction; PI, package insert; PT, preferred term; MedDRA, medical dictionary for regulatory activities.

載されておらず、かつ該当副作用の報告が1件以上されている医薬品と副作用の組み合わせとした。予測モデルの構築に用いるデータセットは、PMDAが公開している国内副作用症例が集積されているJapanese Adverse Drug Event Report (JADER) データベースを採用した。JADERには被疑薬の初回投与以前に副作用が発生した報告や、重複、医療用医薬品以外の情報が含まれているため、これらのデータは事前に除外した(図1)。

早期に「重大な副作用」欄の追記を予測するために、実際に「重大な副作用」が追記された時点より2四半期前(6か月前)の情報を用いて特徴量を作成した。特徴量として、医薬品と副作用のペアごとに医薬品と副作用間の関連性を示唆すると考えられる累積副作用症例数や、累積死亡者数、累積再投与再発症例数、不均衡分析の相対的指標等を計34個作成した。機械学習モデルはeXtreme gradient boosting (XGB)、light gradient boosting machine (LGB)、support vector machine with radial basis function kernel (RBF-SVM)の3種類を使用し、各モデル構築時に最適な特徴量の組み合わせ探索と、ハイパーパラメータの探索を実施した。学習とテストは、作成したデータセットを7:3に分割し、70%を学習に使用して層化8分割交差検証にて評価を行った。モデルの性能評価は不均衡データの二値分類モデルの指標であるマシューズ相関係数(MCC)を採用した。その結果、正解データ293例、不正解データ221,302例を対象とした予測モデルの構築において、RBF-SVMを用いた予測モデルが最も高い性能を示し、テストデータにてMCCが0.941を示した。選択された特徴量は「1四半期あたりの平均副作用報告件数」、「注目する副作用に関して、全医薬品での報告件数のうち注目する医薬品における報告の割合(index B)」、「1四半期前から新規に報告された副作用報告件数」、「ROR」であった。うち前者3つの特徴量は実験に用いた全ての機械学習アルゴリズムにおいて共通な重要な特徴量に選ばれた。

以上より、JADERと機械学習モデルを使用することで、国内症例の集積を根拠とした添付文書の「重大な副作用」欄への追記を高い精度で予測可能なことが示された。また、最適な特徴量の組み合わせを探索した結果、機械学習アルゴリズムに依存せず共通した重要な特徴量として「1四半期あたりの平均副作用報告件数」、「Index B」、「1四半期前から新規に報告された副作用報告件数」が選択された。この結果より、医薬品と副作用との関係性の強さを示唆すると従来から考えられてきた再投与再発症例数や投与からイベント発生までの時間といった情報は、添付文書へ重大な副作用が追記されることを予測する上で重要性が低いことが示唆された。

第2章：安全対策措置の検討を効率的に行うための汎用的予測モデルの構築

「重大な副作用」欄への追記の63%が国内症例の集積を根拠としたものである一方で、37%

の大半が海外で集積された副作用情報が改訂理由となっている。添付文書の改訂理由は様々なパターンが存在し、事前にどのパターンで改訂を行うかを予測することは難しい。そこで、より汎用的に使用可能な、国内・海外の副作用症例情報の集積を根拠とした「重大な副作用」欄への追記を予測するモデルの構築を行った。

本研究では国内で報告された副作用症例の集積に限定せず、国内外の症例集積を理由として「重大な副作用」が追記された医薬品と副作用の組み合わせを予測することを目指した。よって、JADERに加えて世界中の副作用症例を対象にしている米国の副作用自発報告データベースであるthe United States (U.S.) Food and Drug Administration Adverse Event Reporting System (FAERS) を新たに特徴量作成用のデータセットとして採用した。

予測対象の正解データは2011年8月から2022年3月の間に国内副作用症例の集積または国内外の副作用症例集積、海外の副作用症例集積、CCDSの改訂を理由として我が国の添付文書の「重大な副作用」欄へ追記された649件のうち、添付文書の「重大な副作用」の項に記載された副作用名称と、MedDRA PTが一致した414件とした。不正解データは、2022年3月時点で、本研究の対象となった重大な副作用が添付文書に記載されておらず、かつ該当副作用の報告がJADERまたはFAERSにおいて1件以上されている医薬品と副作用の組み合わせ94,391件とした(図2)。

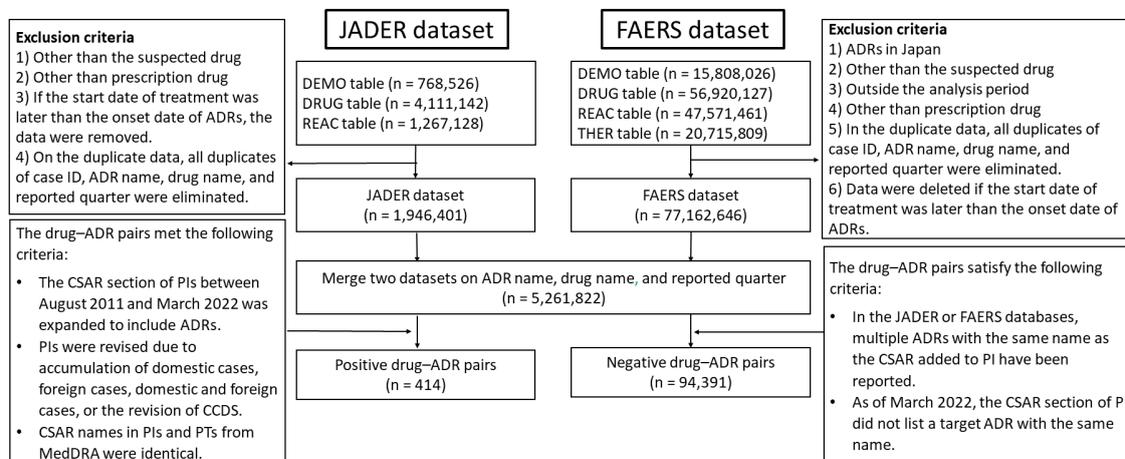


図2 国内外の副作用症例集積に基づく予測モデルの構築に使用したデータセットの選択・除外基準
The JADER and FAERS databases extracted 414 positive and 94,391 negative drug-ADR pairs.

JADER, Japanese Adverse Drug Event Report; FAERS, the U.S. Food and Drug Administration Adverse Event Reporting System; ADR, adverse drug reaction; CSAR, clinically significant adverse reaction; PI, package insert; PT, preferred term; MedDRA, medical dictionary for regulatory activities.

特徴量は、第一章の研究で作成した特徴量に加えて、FAERSの特性を考慮した報告者別の特徴量や交互作用特徴量、時系列性を考慮したラグ特徴量を用いた。これらの特徴量は実際に重大な副作用が追記された時点より2四半期以前の情報を用いて計1,701個作成した。予測モデル

はXGB、LGB、RBF-SVM、neural network (NN) といった単一の機械学習アルゴリズムに加えて、それらのモデルをアンサンブルしたブレンダーモデルを採用し、2つの特徴量データセット (informative features、top 100 features) にて予測モデルを構築した。

予測対象に含まれた重大な副作用の上位10種類について、改訂理由の内訳を確認したところ、国内副作用症例の集積ならびにCCDSの改訂が、重大な副作用の添付文書への追記の主たる理由であることが明らかとなった (図3)。また、副作用ごとに改訂の理由に偏りがあり、間質性肺疾患の90%以上が国内副作用症例の集積を改訂の理由であった一方で、急性汎発性発疹性膿疱症や薬剤性過敏症はCCDSの改訂を理由とした追記が約50%を占めていた。

1,511個の特徴量で構成されたinformative features を使用したgeneralized linear model blenderが平均MCCが0.868という非常に高い予測性能を示した。予測が外れ偽陰性と評価された割合は、国内症例の集積を理由としたものが7件 (11.7%)、国内外の症例集積を理由としたものが1件 (33.3%)、海外症例の集積を理由としたものが3件 (75%)、CCDSの改訂を理由としたものが1件 (6.3%) であった。

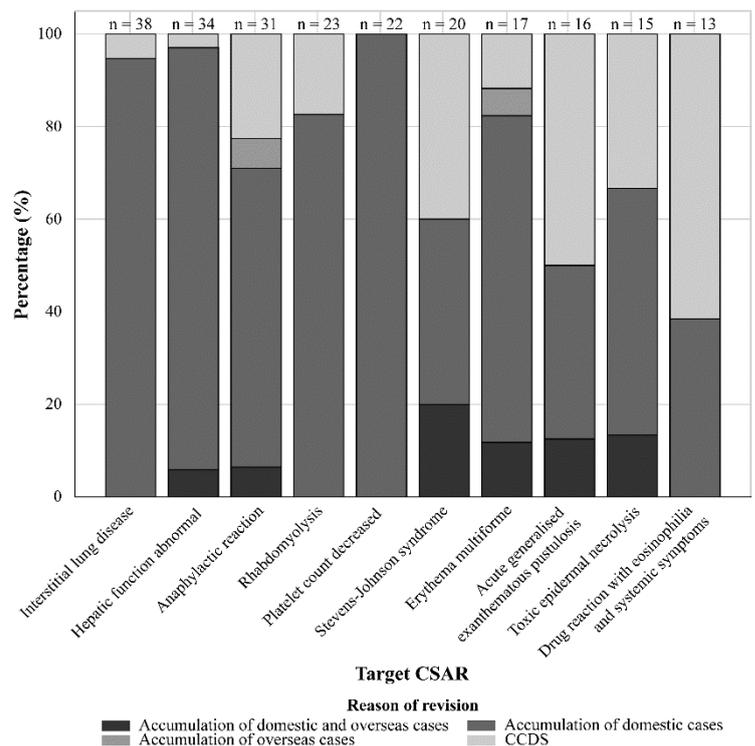


図3 重大な副作用別の重大な副作用追記理由
CCDS, company core data sheet

最も良い予測性能を示したモデルにおける上位5件の重要な特徴量は「イベント2四半期前時点の国内外副作用症例由来のindex B」、「2四半期前時点の国内+欧州副作用症例由来のPRR」、「3四半期前時点の国内+欧州副作用症例由来のPRR」、「3四半期前時点の国内+欧州副作用症例由来のindex B」、「国内症例副作用初回報告から経過した四半期数」であった。

以上より、JADERとFAERS、機械学習モデルを使用することで、国内外に限らず副作用症例の集積を理由とする重大な副作用の追記を高い性能で早期に予測することが可能であることが示唆された。予測に寄与した特徴量として、index BやPRRといったシグナル値が選ばれており、これらは国内症例の集積に限定した第一章の予測モデルにおいても選択されていることか

ら、国内外に関わらず、副作用症例の集積を理由とした重大な副作用の追記を予測する上で一貫して重要であることが示唆された。また、片方のデータベース由来の特徴量が重要な特徴量の上位を占めていることはなく、両データベース共通の特徴量が多く選出されたことから、予測モデルが特定の改訂理由に特化しておらず、汎用的に使用可能な予測モデルとなっていることが示された。

【結論】

第一章では、JADERと機械学習を用いることで、国内副作用症例情報の集積を理由とする添付文書の「重大な副作用」欄への追記を、高い性能で予測できる機械学習モデルが構築できた。また第二章では、予測対象を広げるために、国内・海外に限らず副作用症例の集積を理由とする「重大な副作用」欄への追記を予測可能な、汎用性の高い予測モデルを構築した。以上の研究結果より、副作用報告データベースと機械学習を用いることで、高い精度で「重大な副作用」欄への追記を予測できること、又、我が国の添付文書へ重大な副作用が追記されることを予測する上で不均衡分析から得られたシグナル値やindex Bといった他剤や他の副作用との相対指標が重要であることも確認できた。

両モデルは相補的な関係にあり、第一章で作成したモデルは日本でのみ販売されている医薬品や、間質性肺疾患のような国内副作用症例の集積を理由とした重大な副作用の追記が起りやすい副作用に、第二章で作成したモデルはグローバルで販売されている医薬品に使用することで、それぞれのモデルの性能が発揮しやすいシーンで安全対策措置の要否検討の意思決定支援に貢献できると考える。

序論

医薬品は承認前に臨床試験を通して、有効性及び安全性が科学的に検証された上で承認される。しかし、承認前に臨床試験を通して収集された安全性情報は、限られた患者数や患者背景が限定された、一般的な日常診療とは異なる特殊な条件下で収集されたものである。よって、市販後に医薬品の使用患者数が増加し、患者背景が多様化することで承認時に判明できなかった副作用が顕在化することがある。この課題に対して、医薬品の安全性情報、特に有害事象及び副作用情報を収集し、検出、評価、監視、予防する活動がファーマコヴィジランスと定義され、基本的に全ての医薬品について実施されており、収集された情報を元に規制当局と製薬企業が協力して様々な安全対策措置を講じている³⁾。

安全対策措置には追加の注意喚起や添付文書の「使用上の注意」の改訂、承認事項の変更、承認の取消し等があるが、最も頻度が高く、現場の医療関係者に必要な情報を適時・適切に提供する安全対策措置が添付文書の「使用上の注意」の改訂である。添付文書の「使用上の注意」の改訂は、厚生労働省が発出する「使用上の注意の改訂指示通知」に記載されており、医薬品医療機器総合機構（PMDA）のホームページで確認が可能である⁴⁾。PMDAは、製薬企業や医療関係者、患者等からの有害事象報告や、海外の規制当局が実施した安全対策措置に関する情報など、市販後の安全性情報を幅広く収集・評価し、安全対策措置の必要性を検討する。これらの検討に基づいて、製薬企業は添付文書の改訂などの安全対策措置を実施する⁴⁾。

添付文書の「使用上の注意」の改訂は、以下のステップを含む：（1）安全性情報の収集、（2）有害事象のシグナル検出、（3）シグナル検証、（4）シグナル評価、（5）リスク分類に基づく安全対策措置の検討、（6）専門家による協議、（7）安全対策措置の実施。最初のステップは、製薬企業や医療関係者等からの副作用報告を収集し、PMDAが保有する安全性データベースに蓄積することである。次に、収集されたデータに対してシグナル検出が行われる。シグナルとは、世界保健機構の定義で「それまで知られなかったか、もしくは不完全にしか立証されていなかった薬剤と有害事象との因果関係の可能性に関する、更に検討が必要な情報」とされている³⁾。その後、添付文書の記載内容や関連す

る報告、シグナル検出で得られた評価指標の値、海外の規制当局の対応状況などに基づいてシグナル評価の必要性が検討される。シグナル評価が必要と判断されたものは、新たなリスクであるかを評価するために二次スクリーニングが行われ、その後、製薬企業や専門家と議論を重ねてリスクが新たなものか否か、安全対策措置が必要であるかを判断する。安全対策措置が必要となった場合、PMDAは安全対策の草案を作成するとともに、シグナル評価結果とリスク分類に基づく安全対策措置の妥当性を専門家と議論し、最終的に安全対策措置として添付文書の「使用上の注意」の改訂を製薬企業に指示する^{6,7)}。

シグナル検出及びシグナル検証のステップにおいて、リスク候補であるシグナルを効率的にスクリーニングするシグナル検出手法として、不均衡分析が使われている^{8,9)}。具体的には、proportional reporting ratio (PRR) やreporting odds ratio (ROR)、multi-item gamma Poisson shrinker等の不均衡分析の手法が各国の規制当局において使用されている¹⁰⁻¹³⁾。また、近年、傾向スコアや機械学習を用いた手法も提案されており、有害事象報告の質や内容を考慮することで、検出されるシグナルの解釈性を向上させるだけでなく、予測性能の改善が報告されている¹⁴⁻¹⁷⁾。しかし、これらのシグナル検出手法は直接安全対策措置の要否を予測するものではなく、あくまで大量の情報から効率的にある薬物と有害事象との特異的な関連の可能性をスクリーニングする方法にすぎない。実際に、欧州医薬品庁のファーマコヴィジランス・リスク評価委員会の報告によると、2020年に検出された1,888件の安全性シグナルのうち、39件(2.1%)のみがファーマコヴィジランス・リスク評価委員会によって検証すべきシグナルと判断されており、検出されたシグナルの中に偽陽性が多く含まれていることが明らかとなった¹⁾。現在、安全対策措置の要否を直接予測する手法は報告されておらず、安全対策措置の要否を検討する上で多くのステップを経る必要があり、多大な時間と人的コストを必要としている。従って、収集された安全性情報から直接かつ迅速に、安全対策措置の必要性を判断する際の支援となる根拠を提示する意思決定支援ツールの開発が求められている。

本研究はシグナル検出手法だけで解決できていない、安全対策措置の要否検討を効率化する新しい手法を開発することを目的とし、第1章では添付文書の「使用上の注意」の改訂の

中で最も頻度の高い「国内の副作用症例の集積を根拠とする添付文書の『重大な副作用』欄へ新しい副作用が追記される」イベントを対象とし、このイベントを高い精度で早期に予測可能な機械学習モデルの開発を検討した。機械学習とは、データを通じてパターンを学習し、その学習結果を基に新たなデータに対して予測や決定を行う技術である。機械学習モデルは与えられたデータセットからパターンを抽出し、未知のデータに対する予測や意思決定を支援する。このプロセスは、人間の介入を最小限に抑え、データ駆動型のアプローチを可能にする。本研究が統計モデルではなく、機械学習を選択した理由は、機械学習は統計モデルに対して大量のデータから複雑なパターンを識別し、予測する能力に優れているためである。特に、非線形関係や高次元のデータ構造を扱う場合、機械学習アルゴリズムは統計モデルよりも精度が高い^{18, 19)}。本研究の予測対象のイベントは、少数の線形性を示す情報のみで予測することは困難であると考えられ、またより高い精度で予測可能なモデルの開発を目指したことから機械学習によるモデル開発を実施した。第2章では予測対象を国内の副作用症例の集積を根拠とするものだけではなく、海外の副作用症例の集積等を根拠とする添付文書の「重大な副作用」欄への新しい副作用の追記に拡大し、国内・海外に関わらず副作用症例の集積を根拠とする重大な副作用の追記を予測可能な、より汎用的に使用可能な機械学習モデルの開発を検討した。

本論

第1章：国内副作用症例の集積を根拠とした添付文書の「重大な副作用」欄への追記を予測する機械学習モデルの構築

1-1 背景

本邦の医薬品添付文書の「使用上の注意」欄の改訂の81%が「重大な副作用」欄への追記であり、うち63%が国内副作用症例の集積を改訂のための根拠としていたと報告がされている²⁾。重大な副作用は、副作用の転帰や重篤性を考慮し、「特に注意を要するもの」が記載される。「特に注意を要するもの」とは、致死性や、後遺症の残存が考慮される。法令上は、日米欧医薬品規制調和国際会議の副作用の重篤の定義に基づいて①死亡、②障害、③死亡又は障害につながるおそれのある症例、④治療のために病院又は診療所への入院又は入院期間の延長が必要とされる症例、⑤死亡又は①～④までに掲げる症例に準じて重篤である症例、⑥後世代における先天性の疾病または異常、である²⁰⁾。これらの基準に該当する重大な副作用は、臨床的に大きな影響を与えるため、新規の重大な副作用を添付文書へ追記する場合は、副作用を検知してから添付文書へ追記するまでの時間を短縮し、迅速な安全対策措置を実施する必要がある。しかしながら、PMDAが公表する「安全対策措置の検討に関する標準的な作業の流れのフロー」によると、副作用等報告データベースの中に蓄積された大量の副作用報告の中から、シグナルを検出、検証、評価し、最終的に添付文書の改訂が行われるまでには一般的に数か月程度の時間を要している^{4,5)}。安全対策措置の要否検討を効率化する手法として、大量の医薬品と副作用の組み合わせの中からシグナルを検出する統計学的なシグナル検出手法をPMDAも採用しているものの、シグナルの検出から安全対策措置の実施までにかかる時間に与える効果は限定的である。

そこで第1章では、最も頻度の高い安全対策措置である「国内の副作用症例の集積を根拠とする添付文書の『重大な副作用』欄への追記」を予測対象とし、副作用報告データベースに蓄積された情報から直接予測可能な機械学習モデルを構築することで、安全対策措置要否の意思決定を支援する効果的な手法の開発を目指した。

1-2 方法

1-2-(1) データソースと対象集団

2011年8月から2020年3月に実施された添付文書の改訂のうち、国内症例の副作用集積情報を根拠とする添付文書の「重大な副作用」欄の追記を予測対象とした。予測モデルの構築には、PMDAが公開している日本の医薬品副作用データベース（JADER）を使用した。JADERには国内の副作用報告データが集積されており、（1）DEMOテーブル（性別、年齢、体重等の患者背景情報）、（2）DRUGテーブル（医薬品名、投与開始日、投与終了日等の医薬品情報）、（3）REACテーブル（有害事象名、転帰、発現日等の副作用情報）、（4）HISTテーブル（原疾患名等の疾患情報）の4つから構成されている。本研究ではHISTテーブルを除く3つのテーブルを用いて特徴量作成を行った。DRUGテーブルには被疑薬の他、併用薬、相互作用が含まれる。本研究は国内症例の副作用集積情報を根拠とする添付文書の「重大な副作用」欄の追記を予測することを目的としているため、被疑薬のみを対象とした。また、副作用発症日以降に被疑薬の初回投与日が記録されているものや、重複症例、被疑薬が一般用医薬品の場合は予測においてノイズを発生させると考えたため、本研究では解析対象集団から除外した。

1-2-(2) アウトカム

対象期間中に国内症例の副作用集積情報を根拠として添付文書の「重大な副作用」欄に新たな副作用が追記され、かつ追加された副作用の添付文書記載名と国際医薬用語集（MedDRA）の基本語（PT）とが一致した75種類の副作用をアウトカム（正例）とした。対照群（負例）は、（1）対象期間中に解析対象の75種類の副作用のいずれかが1件以上副作用報告されている、（2）2020年3月時点で、解析対象の副作用が添付文書の「重大な副作用」欄に記載されていない、以上2つの条件を満たすものとした。負例の条件のうち、報告件数を1件以上と定義した理由は、対象期間中に厚生労働省が発出した「使用上の注意の改訂指示」に関する通知の中に、直近3年度の国内副作用症例の集積が1件であっても「重大な副作用」が追記した事例が存在したためである^{21, 22)}。

各副作用はMedDRA version 23.0のPTで集計した。アウトカムの75種類の副作用は、MedDRAのPTの他、特定の医学的状态を表すStandardized MedDRA Queries (SMQ) を参考に定義した。各アウトカムの定義は表1に示す。

表1 予測対象の副作用名を構成するMedDRA基本語一覧

予測対象の副作用名	基本語 (PT)
Interstitial lung disease	Interstitial lung disease, eosinophilic pneumonia, lung disorder, pneumonitis, organizing pneumonia, pulmonary fibrosis, idiopathic pulmonary fibrosis
Hepatic function abnormal	Hepatic function abnormal, liver disorder, drug-induced liver injury
Anaphylactic reaction	Anaphylactic shock, anaphylactic reaction, shock, blood pressure decreased, urticaria
Platelet count decreased	Thrombocytopenia, platelet count decreased
Rhabdomyolysis	Rhabdomyolysis
Erythema multiforme	Erythema multiforme
Toxic epidermal necrolysis	Toxic skin eruption, toxic epidermal necrolysis
Agranulocytosis	Agranulocytosis, granulocytopenia, and granulocyte count decreased
Hepatitis fulminant	Hepatitis fulminant
Ileus	Ileus, intestinal obstruction, ileus paralytic, mechanical ileus
Stevens–Johnson syndrome	Stevens–Johnson syndrome, oculomucocutaneous syndrome
Pancytopenia	Pancytopenia
Cardiac failure congestive	Cardiac failure congestive, cardiac failure
Acute generalized exanthematous pustulosis	Acute generalized exanthematous pustulosis
Acute kidney injury	Acute kidney injury
Pemphigoid	Pemphigoid
Hepatitis B reactivation	Hepatitis B, hepatitis B DNA increased, hepatitis B DNA assay positive, hepatitis B virus test positive, hepatitis B reactivation
Pancreatitis acute	Pancreatitis acute
Drug hypersensitivity	Drug reaction with eosinophilia and systemic symptoms, drug hypersensitivity
Nephrotic syndrome	Nephrotic syndrome
Hypoglycemia	Blood glucose decreased, hypoglycemia, hypoglycemic seizure
Neuroleptic malignant syndrome	Neuroleptic malignant syndrome
Deep vein thrombosis	Deep vein thrombosis, pulmonary embolism
Hemophagocytic lymphohistiocytosis	Hemophagocytic lymphohistiocytosis

Shock symptom	Anaphylactoid shock, shock symptom
Hyponatremia	Blood sodium decreased, hyponatremia
Hypotension	Hypotension
Neutropenia	Neutrophil count decreased, neutropenia, autoimmune neutropenia, idiopathic neutropenia, febrile neutropenia
Tubulointerstitial nephritis	Tubulointerstitial nephritis
Ventricular tachycardia	Ventricular tachycardia, ventricular tachyarrhythmia
Depressed level of consciousness	Depressed level of consciousness, loss of consciousness, altered state of consciousness, hallucination, hallucination, visual
Cholangitis sclerosing	Cholangitis sclerosing
Tuberculosis	Tuberculosis
Intestinal perforation	Ileal perforation, large intestine perforation, intestinal perforation, rectal perforation
Colitis ischemic	Colitis ischemic
Guillain–Barre syndrome	Guillain–Barre syndrome
Ketoacidosis	Diabetic ketoacidosis, ketoacidosis
Metabolic acidosis	Metabolic acidosis
Hypomagnesemia	Hypomagnesemia
Ovarian hyperstimulation syndrome	Ovarian hyperstimulation syndrome
Stomatitis	Stomatitis
Laryngospasm	Laryngospasm
Enteritis	Enteritis
Acute respiratory distress syndrome	Acute respiratory distress syndrome
Acute disseminated encephalomyelitis	Acute disseminated encephalomyelitis
Cholecystitis acute	Cholecystitis acute
Inappropriate antidiuretic hormone secretion	Inappropriate antidiuretic hormone secretion
Disseminated intravascular coagulation	Disseminated intravascular coagulation
Sepsis	Bacterial sepsis, sepsis, septic shock
Neonatal cardiac failure	Cardiac failure high output, neonatal cardiac failure
Pneumothorax	Pneumothorax
Gastrointestinal perforation	Gastrointestinal perforation, upper gastrointestinal perforation, lower gastrointestinal perforation
Hemolytic anemia	Autoimmune hemolytic anemia, hemolytic anemia
Meningitis aseptic	Meningitis aseptic
Hyperthyroidism	Hyperthyroidism
Hypothyroidism	Thyroiditis, hypothyroidism
Leukoencephalopathy	Posterior reversible encephalopathy syndrome, progressive multifocal leukoencephalopathy, leukoencephalopathy
Hepatic failure	Hepatic failure
Gastric antral vascular ectasia	Gastric antral vascular ectasia

Gastric ulcer	Gastroduodenal ulcer, gastric ulcer, duodenal ulcer, gastric ulcer hemorrhage, duodenal ulcer hemorrhage
Cerebral infarction	Thrombotic cerebral infarction, cerebellar infarction, cerebral infarction, hemorrhagic cerebral infarction, embolic cerebral infarction, ischemic cerebral infarction
Nephrogenic diabetes insipidus	Nephrogenic diabetes insipidus
Renal impairment	Renal impairment
Pyelonephritis	Pyelonephritis acute, pyelonephritis
Tumor lysis syndrome	Tumor lysis syndrome
Tendon rupture	Tendon rupture
Thrombotic microangiopathy	Thrombotic microangiopathy
Optic neuritis	Optic neuritis
Myasthenia gravis	Myasthenia gravis, myasthenia gravis crisis
Copper deficiency	Blood copper decreased, copper deficiency
Deafness	Mixed deafness, deafness neurosensory, deafness
Osteomalacia	Osteomalacia
Hyperkalemia	Blood potassium increased, hyperkalemia
Hyperglycemia	Hyperglycemia
Type 1 diabetes mellitus	Type 1 diabetes mellitus, fulminant type 1 diabetes mellitus

1-2-(3) 特徴量作成と前処理

添付文書の改訂はシグナル検出後、数週から数か月以上後に実施される。早期に添付文書の「重大な副作用」欄に追記されることを予測するために、正例は実際に添付文書の「重大な副作用」欄に新たな副作用が追記された時点から2四半期（6か月）前時点のJADERの情報を使用して特徴量を作成した。一方、負例に関しては、解析対象期間の最も新しい時点（2020年3月）のJADERの情報を使用して特徴量を作成した。

特徴量は、CIOMS Working Group VIII並びにGVP Module VIIIで医薬品と有害事象との因果関係を示唆すると言及された情報^{8,23)}や、統計学的シグナル検出手法の改善を目的とした研究においてシグナル検出の改善に貢献した「薬物と副作用について十分な情報を含む報告数」、「過去3年間の副作用報告数」、「副作用報告の偏り」¹⁷⁾、統計学的シグナル値の性質と添付文書改訂の関連性の研究で報告された重要な因子「医薬品の発売からの経過年数」「イベントの重篤性」²⁴⁾を参考に、副作用報告件数や被疑薬再投与再発件数、被疑薬の初回投与から副作用発症までの期間の中央値、統計学的シグナル値、一報告当たり

の欠損値数といった特徴量を計34個作成した（表2）。統計学的シグナル並びに他剤や他の副作用に対する相対的指標は、以下の計算式で作成した。

$$\text{ROR} = \frac{(a/c)}{(b/d)} = \frac{ad}{bc}$$

$$\text{PRR} = \frac{a / (a + c)}{b / (b + d)} = \frac{a(b + d)}{b(a + c)}$$

$$\text{Pearson's chi-squared test with Yates' continuity correction} = \frac{(a + b + c + d)(|ad - bc| - (a + b + c + d)/2)^2}{(a + b)(c + d)(a + c)(b + d)}$$

$$\text{Index A} = \frac{100 \times a}{a + c}$$

$$\text{Index B} = \frac{100 \times a}{a + b}$$

a: 特定の薬剤に関して、注目する副作用が報告された件数、 b: 特定の薬剤以外に関して、注目する副作用が報告された件数、 c: 特定の薬剤に関して、注目する副作用以外の副作用が報告された件数、 d: 特定の薬剤以外に関して、注目する副作用以外の副作用が報告された件数

使用したデータ中の「医薬品投与日」と「発現日」には、4,920件（21.7%）の欠損値が存在した。これらを使用して作成する特徴量「初回投与から発現日までの経過日数の中央値」「初回投与から発現日までの経過日数の平均値」に関して、欠損値は他の薬剤で報告された同じ副作用における「初回投与から発現日までの経過日数の中央値」を代入し対応した。各特徴量は、モデルに応じて標準化または分位点情報を使用した変換を前処理として適用した。

1-2-(4) モデル構築と評価指標

添付文書の「重大な副作用」欄に追記される医薬品と副作用の組み合わせを予測するために、本研究ではeXtreme gradient boosting（XGB）²⁵⁾、light gradient boosting machine（LGB）²⁶⁾、radial basis function kernelを使用したsupport vector machine（RBF-SVM）²⁷⁻³²⁾

の計3つの機械学習アルゴリズムを使用した。

作成した特徴量全てが予測に寄与するとは限らないため、我々は全ての特徴量を用いて機械学習モデルを学習させたものと、特徴量選択手法の一つである **exhaustive feature selection** (EFS) で選択された特徴量のみを学習させたものの2通りのモデリングを実施した³¹⁾。データセットは訓練データとテストデータに7:3で分割し、訓練データは層化8分割交差検証で評価した。ハイパーパラメータ最適化は、**Optuna**を使用した³⁴⁾。モデルの評価指標として、マシューズ相関係数 (MCC) を使用した。MCCは、アウトカムが不均衡なときに使用できるバランスの取れた評価指標であり、二値分類モデルの精度を測定するものである。MCCは-1から+1の範囲で、+1は完全な予測を、0は平均的なランダム予測を表す。添付文書の「重大な副作用」欄に新しい副作用が追記されることは稀であり、データセットは不均衡であるため、機械学習モデルのパフォーマンスはMCCを用いて評価した。

$$\text{MCC} = \frac{(\text{TP} \times \text{TN}) - (\text{FP} \times \text{FN})}{\sqrt{(\text{TP} + \text{FP}) \times (\text{TP} + \text{FN}) \times (\text{TN} + \text{FP}) \times (\text{TN} + \text{FN})}}$$

TP, true positive; TN, true negative; FP, false positive; FN, false negative.

データの加工と機械学習モデルの作成は、Python 3.9.3、pandas 1.2.3、numpy 1.24.4、scikit-learn 1.0.2、mlxtend 0.20.0、lightgbm 3.2.1、xgboost 1.4.2、optuna 2.10.0、matplotlib 3.4.1、seaborn 0.11.2を使用した。

1-3 結果

1-3-(1) 収集結果

解析対象期間中に実施された添付文書の「重大な副作用」欄への追記のうち、54%（334例）が国内の副作用症例集積を根拠としていた。334例中、293例が添付文書の「重大な副作用」欄に記載された副作用名がMedDRAのPTと完全一致したため、これらを本研究の予測対象とした（図1）。

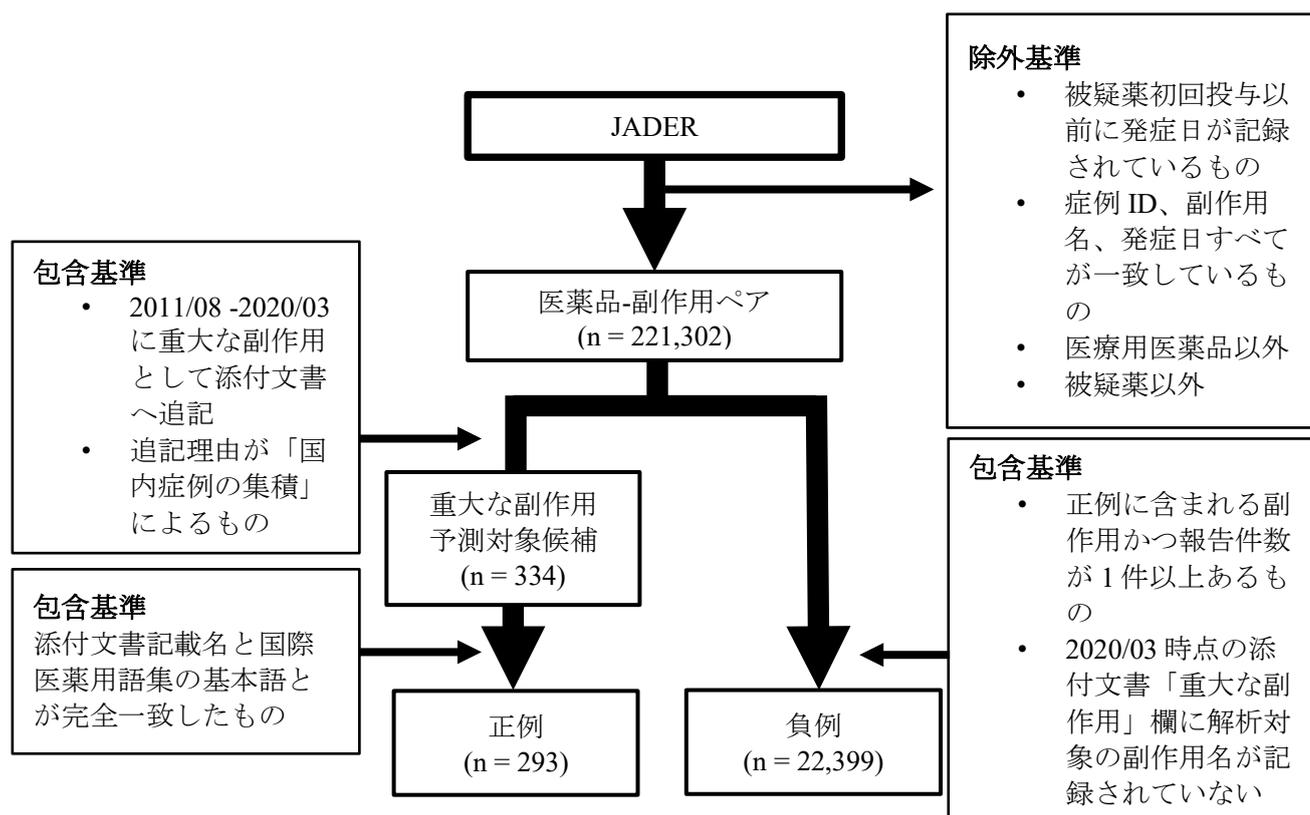


図1 予測対象の重大な副作用追記事例と対照群の抽出フロー図

JADERを使用し、医療用医薬品の被疑薬を対象を限定し、データの重複やデータの不整合を除き、最終的に添付文書の「重大な副作用」欄へ新たな副作用が追記された正例293件、正例に含まれる副作用が報告されているものの、添付文書の「重大な副作用」欄には記載されていない負例22,399件が抽出された。

図1に示した通り、221,302件の医薬品と副作用の組み合わせが、正例ならびに負例として抽出され、最終的に正例293例、負例22,399例を機械学習モデルの構築と評価に使用した。正例は負例に対して一症例あたりの欠損データ数が少なく、一症例当たりのPMDAへの平均報告回数が多い傾向が見られた。また、両群において被疑薬の再投与例の報告は非常に

稀であった（表2）。

表2 JADERを用いて作成した全34特徴量

特徴量名	正例	負例
	Mean (SD)	Mean (SD)
Number of patients	12.4 (19.6)	7.5 (44.4)
Number of deaths	0.3 (1.3)	0.2 (1.8)
Number of patients with drug re-administration	0.0 (0.1)	0.0 (0.3)
Number of patients who discontinued the suspected drug	6.6 (13.4)	2.6 (22.3)
The average number of days between administration of the suspected drug and the onset of the adverse drug reaction	117.5 (254.0)	123.1 (331.0)
The median number of days between administration of the suspected drug and the onset of the adverse drug reaction	87.9 (267.8)	105.3 (325.0)
Number of patients within 15 days from the administration of the suspected drug to the onset of the adverse drug reaction	3.3 (10.7)	1.4 (23.6)
Number of patients within 30 days from the administration of the suspected drug to the onset of the adverse drug reaction	4.1 (11.2)	1.8 (26.0)
Number of patients within 90 days from the administration of the suspected drug to the onset of the adverse drug reaction	5.5 (12.0)	2.5 (29.7)
The average number of missing values per case	3.9 (1.5)	5.5 (3.0)
The average number of reports to the regulatory authority per case	3.1 (1.5)	2.4 (1.2)
Index A	3.9 (7.3)	2.5 (6.8)

Index B	1.6 (5.9)	0.2 (0.9)
Number of quarters that have elapsed since the first adverse event was reported	17.8 (13.5)	36.7 (20.4)
Number of newly reported patients from a quarter ago	1.2 (1.9)	0.0 (0.0)
Number of new patients reported since before 2 quarters	2.1 (2.9)	0.2 (1.5)
Number of new patients reported since before 3 quarters	3.0 (4.0)	0.3 (2.2)
Number of new patients reported since before 4 quarters	3.7 (4.9)	0.4 (2.8)
Number of newly reported deaths from a quarter ago	0.1 (1.0)	0.0 (0.0)
Number of newly reported patients with drug re-administration from a quarter ago	0.0 (0.2)	0.0 (0.0)
Number of newly reported patients who discontinued the suspected drug a quarter ago	0.6 (4.3)	0.0 (5.0)
Number of newly reported patients within 15 days from the administration of the suspected drug to the onset of the adverse drug reaction from a quarter ago	-0.5 (8.2)	0.0 (0.2)
Number of newly reported patients within 30 days from the administration of the suspected drug to the onset of the adverse drug reaction from a quarter ago	-0.5 (8.2)	0.0 (0.2)
Number of newly reported patients within 90 days from the administration of the suspected drug to the onset of the adverse drug reaction from a quarter ago	-0.6 (8.7)	0.0 (0.3)
The average number of patients reported per quarter	11.8 (16.5)	14.4 (24.4)
PRR	49.0 (423.7)	6.3 (63.9)

Log PRR	0.8 (1.8)	0.2 (1.5)
The lower limit of the confidence interval for PRR	21.0 (178.1)	2.1 (40.7)
The upper limit of the confidence interval for PRR	124.5 (1036.1)	26.7 (197.8)
YatesChisq	248.8 (1645.3)	65.8 (2518.4)
ROR	57.9 (530.8)	8.1 (109.8)
Log ROR	0.9 (1.8)	0.3 (1.6)
The lower limit of the confidence interval for ROR	22.2 (191.6)	2.2 (60.2)
The upper limit of the confidence interval for ROR	163.7 (1494.8)	51.7 (706.4)

PRR, proportional reporting ratio; ROR, reporting odds ratio; YatesChisq, Pearson's chi-squared test with Yates' continuity correction; Index A, the ratio of the target adverse event number of specific drug to all adverse event numbers that were reported for the same drug; index B, ratio of adverse events covered by a particular drug to those covered by all drugs; SD, standard deviation.

1-3-(2) 機械学習モデルの予測性能の比較

各モデルの予測性能は、図2、Supplemental 表1（表S1）に示した。RBF-SVMモデルは、偽陽性と偽陰性が最も少なく、MCCが最も高かった。特徴量選択を行ったRBF-SVMモデルは、交差検証でのMCCが0.938、テストデータでのMCCが0.922と、最も良好な予測性能を示した。最も予測性能が優れたRBF-SVMモデルのハイパーパラメータを最適化（Cが0.01–3000の範囲、 γ が0.001–1000の範囲で5000回の反復探索）した結果、MCCは0.941（C = 2388、 $\gamma = 4.718$ ）であった（表3）。

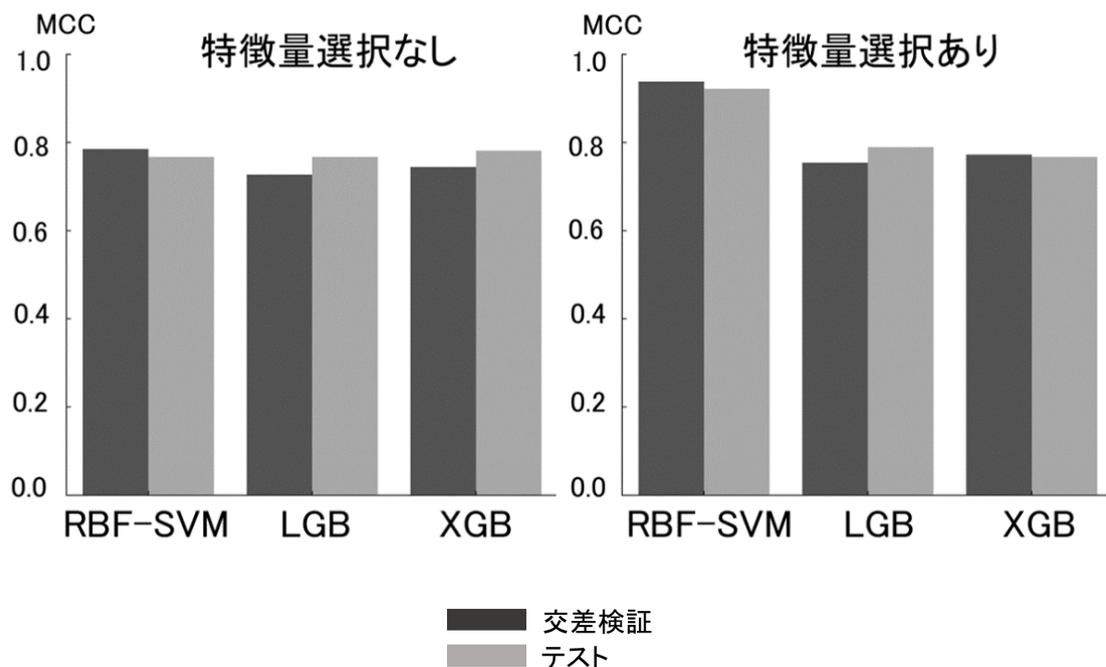


図2 特徴量選択有無別の各機械学習モデルの予測性能

MCC, matthews correlation coefficient; RBF-SVM, support vector machine with radial basis function kernel; LGB, light gradient boosting machine; XGB, extreme gradient boosting.

表3 RBF-SVMモデルの評価指標一覧

特徴量選択有無	訓練データ				テストデータ			
	MCC (SD)	Precision (SD)	Recall (SD)	AUC (SD)	MCC (SD)	Precision (SD)	Recall (SD)	AUC (SD)
無し	0.785 (0.044)	0.899 (0.066)	0.693 (0.068)	0.966 (0.024)	0.767	1.0	0.591	0.983
有り	0.938 (0.032)	0.989 (0.019)	0.892 (0.047)	0.989 (0.012)	0.922	0.987	0.864	0.998
有り+最適化	-	-	-	-	0.941	0.988	0.898	0.999

AUC, area under the curve; MCC, matthews correlation coefficient; RBF-SVM, support vector machine with radial basis function kernel; SD, standard deviation.

最適化されたRBF-SVMモデルにおける特徴量重要度の結果は表4に示した。特徴量重要度はpermutation importanceを使用した³⁵⁾。Permutation importanceは、特徴量の値をランダムに交換することでモデルの誤差が増加した場合、その特徴は重要と考えられ、予測はこの特徴に依存していることを示す。モデルで最も重要度が高かった特徴は、1四半期あたりの平均副作用報告件数を意味する「The average number of patients reported per quarter」であった。

表4 ハイパーパラメータ最適化済みRBF-SVMモデルにおける特徴量重要度

特徴量名	相対重要度
The average number of patients reported per quarter	1.0000
ROR	0.9769
Index B	0.9646
Number of newly reported patients from a quarter ago	0.0191

ROR, reporting odds ratio; index B, ratio of adverse events covered by a particular drug to those covered by all drugs.

1-3-(3) 統計学的シグナル検出手法と機械学習モデルの比較

機械学習モデルの予測性能と統計学的シグナル検出手法を比較するために、従来の統計学的シグナル検出手法の代表例であるPRRと添付文書の「重大な副作用」欄への副作用の追記との関係性を分析した。シグナル検出に一般的に適用されるPRRの閾値である報告数が3件以上かつPearson's chi-squared test with Yates' continuity correctionが4.0以上、PRRが2.0以上を示す医薬品と副作用の組み合わせをシグナル陽性とした。表5にテストデータにおける検出されたシグナルと、実際に添付文書の「重大な副作用」欄に副作用が追記された事例との関係性を示した。PRRが多数の偽陽性を検出し、検出されたシグナルのわずか4.1% (35/859) が添付文書の「重大な副作用」欄への副作用追記に繋がっていることが明らかとなった。一方、ハイパーパラメータ最適化後のRBF-SVMのテストデータにおける予測では、添付文書の「重大な副作用」欄へ副作用が追記されると予測されたもののうち、98.8% (84/85) が実際に添付文書の「重大な副作用」欄へ副作用が追記されていた (表6)。

表5 添付文書の「重大な副作用」欄への追記とPRRの関係性

		添付文書の「重大な副作用」欄への追記件数	
		有り	無し
シグナル検出結果	陽性	35	824
	陰性	53	5,896

表6 ハイパーパラメータ最適化後のRBF-SVMのテストデータにおける予測結果

		添付文書の「重大な副作用」欄への追記件数	
		有り	無し
モデルの予測結果	陽性	84	1
	陰性	4	6,719

1-4 考察

日本国内の副作用症例の集積を根拠とした添付文書の「重大な副作用」欄への追記は、従来の統計学的シグナル検出手法と比較して、RBF-SVMを使用することで高い予測性能で早期に予測可能であることが示された。検討した3つの機械学習アルゴリズムの中で、最も予測性能が高かったRBF-SVMにおいて選ばれた4個の特徴量のうち、1四半期あたりの平均副作用報告件数を意味する「The average number of patients reported per quarter」、注目する副作用に関して全医薬品での報告件数のうち注目する医薬品が占める報告の割合を意味する「Index B」、前四半期から新たに追加された副作用報告件数を意味する「Number of newly reported patients from a quarter ago」の3個は、RBF-SVM以外の機械学習アルゴリズムにおいても共通して予測に必要な特徴量として選ばれた（表S2）。このことから、国内の副作用症例の集積を根拠とした添付文書の「重大な副作用」欄への追記を予測する上で、機械学習アルゴリズムに依存せず、一貫してこれら3つの特徴量が重要であると言える。また、予測に重要と考えていた医薬品と副作用の因果関係を示唆する被疑薬の再投与情報や、被疑薬の投与から副作用発現までの期間の短さは、予測に必要な特徴量として選ばれなかった³⁶⁾。更に、副作用報告件数や死亡件数は、PMDAや製薬企業が安全対策措置の要否を検討する際の指標として使われてきたが、これらも予測に必要な特徴量として選ばれなかった。シグナル検出手法の改善を目的とした先行研究において、被疑薬の再投与情報や、被疑薬の投与から副作用発現までの期間、副作用報告件数はシグナル検出の改善への貢献度は低く、直近の副作用報告件数や副作用報告の不均衡性の貢献度が高かったと報告されている¹⁶⁾。シグナル検出が安全対策措置の要否検討の最初のステップであり、検出されたシグナルの一部が最終的に安全対策措置の要否に繋がることを考慮すると、シグナル検出と本研究で作成した予測モデルの目的は異なるものの、類似の特徴量が重要である点は自然であると考えられる。

日本国内の副作用症例の集積を根拠とした添付文書の「重大な副作用」欄への追記は、従来の統計学的シグナル検出手法と比較して、RBF-SVMを使用することでMCC 0.941という高い予測性能で早期に予測可能であることが示された。RBF-SVMはカーネル法を使用

することで、非線形な特徴量をより高次元の空間に写像することで線形分離を行う方法である³¹⁾。RBF-SVMが高い予測性能を示した理由は、EFSによって選択された特徴量の数と、RBF-SVMのモデルの特徴から推察が可能である。RBF-SVMを含むSVMは、学習に特徴量すべてを使用して2つのクラスを分ける境界線を探査する。よって、予測に貢献しない特徴量が含まれている場合、その特徴量も考慮した上で2つのクラスを分ける最適な境界線が作成されるため予測性能が低下してしまう。一方で、今回実験したXGBやLGBは学習の過程で予測に貢献しない特徴量を除外しながら2つのクラスを分ける最適な境界線を探査するため、EFSによる特徴量選択の影響がRBF-SVMに比べて小さい。最終的にEFSで選ばれた特徴量数が4個と少ないこと、すべての特徴量を学習に使用するRBF-SVMのモデルの特徴を考慮すると、国内の副作用症例の集積を根拠とした添付文書の「重大な副作用」欄への追記を予測する問題は、限られた特徴量で高次元空間において線形分離可能な単純な問題であったため、RBF-SVMが高い予測性能を発揮したと考えられる。

高い予測性能を示した一方で、表6が示す通りRBF-SVMにおいて偽陽性が1件、偽陰性が4件検出された。これら予測が外れた医薬品と副作用の組み合わせにおける特徴的な傾向は認められなかった。ただし、選択された重要な特徴量のindex BやRORは他の医薬品や副作用に対する相対的指標であり、例えば他の医薬品で全く報告されていない、あるいは非常に報告頻度が低い副作用の場合、相対的指標が注目した医薬品において計算できない、または非常に高い異常値を示してしまう。このような非常に稀なケースは、本予測モデルで正しい予測をすることが難しいため、機械学習モデルではなく、各症例報告の内容を個別に精査する個別症例評価による安全対策措置の要否検討が望ましいと考えられる。

本研究結果より、機械学習を使用した本予測モデルは、添付文書の「重大な副作用」欄への追記を検討する安全対策措置要否を判定する時間を短縮し、より効率的な意思決定を実現することを支援できる可能性がある。従来の統計学的シグナル検出手法は、大量の医薬品と副作用の組み合わせの中からリスクの高い組み合わせを検出することで、安全対策措置の要否を検討するワークフローの効率性を高めている。一方で、シグナル検出はあくまでリスクの高い組み合わせを検出し、優先順位付けするための役割しか持っておらず、

安全対策措置が必要な真のリスクを検出する役割は有していない^{9, 11, 37)}。欧州医薬品庁のファーマコヴィジランス・リスク評価委員会の報告によると、2020年に検出された1,888件の安全性シグナルのうち、39件（2.1%）のみがファーマコヴィジランス・リスク評価委員会によって検証すべきシグナルと判断された¹⁾。このことから、統計学的シグナル検出手法により検出されたシグナルが、安全対策措置に直接繋がることが稀であることがわかる。本研究においても、同様の傾向が認められるか確認するために、従来の統計学的シグナル検出手法の代表としてPRRと、安全対策措置として添付文書の「重大な副作用」欄への追記との関係性を分析した（表5）。結果、本研究においても検出された大多数のシグナルが、添付文書の「重大な副作用」欄への追記に影響を及ぼしていないことが明らかとなった。この結果は、本研究で使用したデータが特殊ではないこと、そして従来の統計学的シグナル検出手法と比較して、本予測モデルは添付文書の「重大な副作用」欄への追記を直接予測する性能が高く、実社会への外挿が可能であることを支持していると考えられる。

本研究では、他の目的で集積したデータを利用しているため、結果の解釈上の限界が存在する。その限界を以下に述べる。

- (1) 本研究で作成した機械学習モデルは、副作用自発報告データベースのJADERを使用している。自発報告には医師、薬剤師を含む医療専門家の他、弁護士、消費者からの報告も含まれており、質に差があることが考えられる。本研究では報告者の違いが機械学習モデルに与える影響を検討していない。しかし、JADERは消費者からの自発報告が多数含まれる米国の副作用自発報告データベースであるFAERSと異なり、報告の大半が医療専門家を占めている³⁸⁾。よって、国内の副作用症例の集積を根拠とした添付文書の「重大な副作用」の欄への追記を予測する本研究においては、報告者による自発報告の質が結果に与える影響は小さいと考えられる。
- (2) 予測対象は添付文書の重大な副作用欄に記載された副作用名と、MedDRAのPTが完全一致したものを対象としており、添付文書記載名とPTが一致しない場合は本機械学習モデル

ルを使用することができない。このようなケースにおいては、医学専門家や類薬における副作用集積状況を考慮した上で、集計する副作用のグルーピングを実施する必要がある。本研究において、添付文書記載名とPTが一致せず、解析対象に含まれなかったケースが41件あったものの、293件（87.7%）が解析対象に含まれている。これら大半の副作用において高い性能で予測できている点、副作用の発現部位などの副作用の属性情報を特徴量として学習していない点から、国内の副作用症例の集積を根拠とした添付文書の「重大な副作用」の欄への追記を予測する目的においては、汎用的に使用可能な機械学習モデルとなっていると考えられる。

- (3) 本機械学習モデルは、添付文書の「重大な副作用」の欄への追記が発生する2四半期前時点の情報を利用することで予測が可能であるという仮説に基づいて開発された。しかし、医薬品の上市直後に致命的な副作用が頻発した場合は、2四半期分の情報が集積される前に重大な副作用が追記されることがある。このようなケースにおいては、本機械学習モデルは適応できない。

以上の限界はあるものの、本研究は国内の副作用症例の集積を根拠とした添付文書の「重大な副作用」欄へ新たな副作用が追記されることを早期に、高い精度で、直接予測可能な機械学習モデルを開発した初めての研究である。本機械学習モデルを安全対策措置の要否検討の標準的なワークフローに組み込むことで、重大な副作用として添付文書へ記載すべきより高いリスクの医薬品と副作用の組み合わせを検出し、時間と人的コストを削減したより効率的な安全対策措置の要否検討の実現が期待される。

1-5 小括

JADERと機械学習を用いることで、高い精度で2四半期後に国内の副作用症例の集積を根拠とした添付文書の「重大な副作用」欄へ新たな副作用が追記されることを予測することが可能である。また、医薬品と副作用との関係性の強さを示唆すると従来から考えられてきた再投与再発症例数や投与からイベント発生までの時間といった情報は、「重大な副

作用」欄へ新たな副作用が追記されることへ与える影響が小さいことが示唆される。

第2章：安全対策措置の検討を効率的に行うための汎用的予測モデルの構築

2-1 背景

第1章では、国内の副作用症例の集積を根拠とした添付文書の「重大な副作用」欄へ新たな副作用が追記されることを予測する機械学習モデルを作成することで、安全対策措置の要否検討を効率化させる検討を行った。しかしながら、添付文書の「重大な副作用」欄への追記の63%が国内症例の集積を根拠としたものである一方で、残り37%は各国の添付文書を作成する際に基準となる文書であるCCDSの改訂や海外の副作用症例の集積、国内および海外の副作用用例の集積、海外の臨床試験結果、公表文献・ガイドライン・学会における報告状況等が改訂根拠となっている²⁾。このように、様々な改訂根拠を元に「重大な副作用」欄へ新たな副作用が追記されるため、安全対策措置の要否検討を効率化する目的において、改訂根拠にできるだけ依存しない、適用範囲の広い予測モデルを構築することが実用性の観点で重要である。そこで本研究は、国内の副作用症例の集積に限定せず、国内・海外に関わらず副作用症例の集積が関わる改訂根拠に基づく添付文書の「重大な副作用」欄への新たな副作用の追記を早期に予測可能な実用性の高い機械学習モデルを構築することを目的とした。

2-2 方法

2-2-(1) データソースと対象集団

2011年8月から2022年3月までに新しい重大な副作用が添付文書に追記された事例のうち、国内の副作用症例の集積、海外の副作用症例の集積、国内及び海外の副作用症例の集積、CCDSの改訂を根拠とした添付文書の「重大な副作用」欄への新たな副作用の追記を予測対象とした。CCDSの改訂は、海外の副作用症例の集積に基づき実施されるため、副作用症例の集積を根拠とする事例と判断し、解析対象に含めた。予測モデルの構築に必要な特徴量の作成は、国内・海外それぞれの副作用集積情報等を考慮する必要があるため、PMDAが公開している日本の医薬品副作用データベースであるJADERと、米国食品医薬品局であるThe United States (U.S.) Food and Drug Administration (FDA) が公開している米国を中心とした

海外の医薬品副作用データベースであるthe U.S. Food and Drug Administration Adverse Event Reporting System (FAERS) を使用した³⁸⁾。なお、FAERSはFDAによって無償公開されているが、本研究はFAERSに含まれる重複症例の削除、報告医薬品名の正規化等のクレンジング処理が事前にされたデータであるJAPIC AERSを一般財団法人日本医薬情報センターから購入し、予測モデルの構築に使用した。

FAERSは (1) DEMOテーブル (患者の年齢、性別等)、 (2) DRUGテーブル (使用薬剤等)、 (3) REACテーブル (有害事象名等)、 (4) OUTCテーブル (症例転帰情報)、 (5) RPSRテーブル (情報源等)、 (6) INDIテーブル (原疾患等)、 (7) THERテーブル (治療期間等)、以上7つから構成されている。本研究では、2022年3月までの情報を含むDEMOテーブル、DRUGテーブル、REACテーブル、THERテーブルの4つを使用して特徴量を作成した。また、FAERSのデータ収集開始時期がJADERとは異なるため、JADERのデータ収集開始時期と揃えるために、2004年以降のデータを使用した。

FAERSには日本の副作用症例が含まれており、JADERと症例が重複する可能性が考えられる。そのため、特徴量を作成する前にFAERSから日本の副作用症例を削除した。第一章の研究と同様に、副作用症例の集積に関わる改訂根拠に基づく添付文書の「重大な副作用」欄への新たな副作用の追記を予測するために、医療用医薬品の被疑薬以外の薬剤、重複症例、副作用発症日以降に被疑薬の初回投与日が記録されているものを解析対象集団から事前に除外した。

2-2-(2) アウトカム

対象期間中に国内の副作用症例集積、海外の副作用症例集積、国内及び海外の副作用症例集積、CCDSの改訂を根拠として添付文書の「重大な副作用」欄に新たな副作用が追記され、かつ追加された副作用の添付文書記載名とMedDRAの基本語 (PT) と一致した副作用をアウトカム (正例) とした (表7)。対照群 (負例) は、 (1) 対象期間中に正例と同じ副作用がJADERまたはFAERSで報告されている、 (2) 2022年3月時点で、正例と同じ副作用が添付文書の「重大な副作用」欄に記載されていない、以上2つの条件を

満たすものとした。各副作用はMedDRA version 25.0のPTで集計した。

表7 予測対象の副作用名と解析対象件数

基本語 (PT)	解析対象件数
Acute generalised exanthematous pustulosis	16
Acute hepatic failure	1
Acute kidney injury	7
Acute respiratory distress syndrome	1
Agranulocytosis	9
Anaphylactic reaction	31
Angioedema	1
Autoimmune hepatitis	1
Bradycardia	3
Cardiac failure	6
Cerebral infarction	1
Cholangitis sclerosing	3
Cholecystitis acute	2
Colitis	1
Colitis ischaemic	2
Contrast encephalopathy	2
Deafness	1
Deep vein thrombosis	4
Dermatitis exfoliative	1
Disseminated intravascular coagulation	1
Drug reaction with eosinophilia and systemic symptoms	13
Enteritis	1
Erythema multiforme	17
Fanconi syndrome	1
Gastric antral vascular ectasia	1
Gastric ulcer	1
Gastrointestinal perforation	2
Guillain-Barre syndrome	1
Haemolytic anaemia	1
Haemophagocytic lymphohistiocytosis	4
Hepatic failure	2
Hepatic function abnormal	34
Hepatitis fulminant	10
Hydrocephalus	1

Hyperglycaemia	1
Hyperkalaemia	2
Hyperparathyroidism	1
Hyperthyroidism	3
Hypocalcaemia	2
Hypoglycaemia	3
Hypoglycaemia neonatal	1
Hypomagnesaemia	2
Hyponatraemia	2
Hypotension	1
Hypothyroidism	2
Ileus	1
Ileus paralytic	2
Inappropriate antidiuretic hormone secretion	2
Interstitial lung disease	38
Intestinal obstruction	5
Intestinal perforation	1
Intestinal ulcer	1
Ketoacidosis	1
Laryngospasm	1
Leukoencephalopathy	3
Metabolic acidosis	1
Myasthenia gravis	1
Myocarditis	2
Neonatal cardiac failure	1
Nephrogenic diabetes insipidus	1
Nephrotic syndrome	4
Neuroleptic malignant syndrome	3
Neuropathy peripheral	1
Neutropenia	3
Osteomalacia	1
Ovarian hyperstimulation syndrome	1
Pancreatitis	1
Pancreatitis acute	4
Pancytopenia	5
Pemphigoid	5
Pemphigus	1
Peripheral arterial occlusive disease	1
Platelet count decreased	22

Pleural effusion	1
Pleurisy	1
Pneumothorax	1
Progressive multifocal leukoencephalopathy	4
Pseudomembranous colitis	1
Pulmonary embolism	1
Pulmonary hypertension	2
Pyelonephritis	1
Renal impairment	1
Retinal detachment	1
Rhabdomyolysis	23
Sepsis	1
Stevens-Johnson syndrome	20
Stomatitis	1
Thrombotic microangiopathy	1
Thrombotic thrombocytopenic purpura	2
Toxic epidermal necrolysis	15
Tuberculosis	3
Tubulointerstitial nephritis	6
Tumour lysis syndrome	4
Type 1 diabetes mellitus	3
Vasculitis	1
Ventricular tachycardia	2

2-2-(3) 特徴量作成

正例は実際に添付文書の「重大な副作用」欄に新たな副作用が追記された時点から2四半期前時点、3四半期前時点、4四半期前時点のJADER並びにFAERSに含まれる情報を使用して、負例に関しては解析対象期間の最も新しい時点（2022年3月）のJADER並びにFAERSに含まれる情報を使用して計1,701個の特徴量を作成した。これら特徴量は第一章の研究で重要な特徴量として選ばれた「統計学的シグナル値」や「1四半期あたりの平均副作用報告件数」、全医薬品での報告件数のうち注目する医薬品が占める報告の割合を意味する「Index B」、「前四半期から新たに追加された副作用報告件数」に加えて、医薬品と有害事象との因果関係を示唆する情報、製薬企業がファーマコヴィジランス活動で注視する情報等を参考に作成した^{8, 17, 23, 24}。更に作成した特徴量を用いて、各

特徴量作成時点間で交互作用を作成した。

JADERとFAERSは、報告される副作用の情報源に異なる偏りが見られる。具体的には、FAERSの副作用報告の半数程度が消費者からのものであるのに対し、JADERの副作用報告の9割程度が医療専門家からのものである³⁸⁾。この情報源の偏りは、副作用報告の質並びに予測性能に影響を与えると考えたため、情報源別に特徴量を作成した。また、FAERSと異なり、JADERには「第一被疑薬」「第二被疑薬」といった被疑薬の順位を表す情報が含まれていない。被疑薬が一つに限定できる場合と、複数存在する場合とでは医薬品と副作用との因果関係の確からしさが異なると考えた。そこで、JADERを使用して作成した特徴量は、被疑薬と副作用の関係が1:1のものに限定して集計したもの、1:2に限定して集計したもの、2:1に限定して集計したものを追加で作成した。以上の特徴量作成に関して、表8にまとめた。

表8 各データベース由来の特徴量一覧

	JADER 由来の特徴量	FAERS 由来の特徴量	両データベース由来の特徴量
Basic features (基本特徴量)	Number of reporters, number of deaths, number of recurrence patients of re-administration, number of discontinued patients, number of days from administration to event onset (average, median), number of patients with event onset within N days from administration (N = 15, 30, 90), average number of missing values per report, average number of reports per case, number of quarters since the first AE was reported,	Number of reporters, number of reporters from physicians, number of reporters from pharmacists, number of reporters from other healthcare professionals, number of reporters from consumers, total number of reporters from healthcare professionals, number of reporters from the United States (U.S.), number of reporters from the European Union (EU)'s 5 major countries, number of primary suspect drugs, number of	Number of reporters from Japan and the U.S., number of reporters from Japan and the EU's 5 major countries, difference in the number of quarters since the first report in Japan and the first report overseas, ratio of reporters from Japan to U.S., ratio of reporters from Japan to the EU's 5 major countries

	average number of reports per quarter	quarters since the first AE was reported, average number of reports per quarter	
	These are all segmented by reporter type and within the groups where the ADR to suspected drug relationship is limited to 1:1, 1:2, or 2:1.		
Disproportionality analysis (統計学的シグナル値)	PRR, chi-square test, signals determined by PRR	PRR, chi-square test, signals determined by PRR	Signal values based on the number of reports from JADER and FAERS (PRR, chi-square test), signal values based on the number of reports from JADER and the EU's 5 major countries in FAERS (PRR, chi-squared test), signals determined by PRR
	These features were calculated for each report count according to the conditions computed in basic features.	These features were calculated for each report count according to the conditions computed in the basic features.	
Relative features (相対的指標)	Index A, index B		
	These features were computed for each report count based on the conditions computed in Basic features.		
Lag features (ラグ特徴量)	Using the point two quarters prior to the addition of CSAR to the PI as a reference generates the basic features, disproportionality analysis, relative features, and interaction features for one and two quarters prior to that reference point.		
Interaction features (交互作用)	At various time points, subtraction and multiplication were performed between the same features.		

Index A, percentage of a specific type of ADR cases due to the suspected drug out of all ADR cases that occurred after using the same drug; index B, percentage of ADR cases attributed to the suspected drug out of the total ADR cases from all drugs; PRR, proportional reporting ratio; chi-squared test, Pearson's chi-squared test with Yates continuity correction; CSAR, clinically significant adverse reaction; PI, package insert; JADER, Japanese Adverse Drug Event Report; FAERS, the U.S. Food and Drug Administration Adverse Event Reporting System.

2-2-(4) モデル構築と評価指標

予測モデルの構築は、AutoMLツールのDataRobotを使用した。DataRobotは3,000以上のデータ前処理、特徴量エンジニアリング、機械学習アルゴリズムを有しており、これらを自

動的に組み合わせて最適な予測モデルを作成する自動化ツールである。データセットは訓練データとテストデータに8:2で分割し、訓練データは層化5分割交差検証で評価した。評価指標は、不均衡データの評価に頑強なMCCを採用した。また、データセットの分割方法による予測性能の誤差を考慮するために、5つのランダムシードを使用してデータセットを分割し、それぞれのランダムシードにおける予測性能の平均で最終的な予測モデルの性能を評価した。

モデルの構築には2つの特徴量セットを使用した。1つは作成した1,701個の特徴量の中から、DataRobotによって自動的に予測に寄与しない特徴量が除外された結果残った1,511個の特徴量からなる「有用な特徴量」、他方が有用な特徴量で予測モデルを一度作成した後に、その予測モデルにおける重要な特徴量上位100個を選抜した「重要な特徴量上位100個」である。両特徴量セットを使用して予測モデルを構築し、最も性能が高かったモデルを最終的に予測モデルとして選択した。予測モデルにおいて重要な特徴量は、permutation importanceを用いて評価した。予測モデル作成を除く分析、可視化はpython 3.9.3、Python 3.9.3、pandas 1.2.3、numpy 1.24.4、matplotlib 3.4.1、seaborn 0.11.2を使用した。

2-3 結果

2-3-(1) 収集結果

解析対象期間中に、701件の副作用が添付文書の「重大な副作用」欄へ新たに追記された。その内、648件（92.3%）が国内・海外に関わらず副作用症例の集積を根拠としたものであった。組み入れ基準、除外基準に従い414件の正例（国内の副作用症例集積を根拠とした追記：302件、国内及び海外の副作用症例集積を根拠とした追記：22件、海外の副作用症例主席を根拠とした追記：12件、CCDSの改訂を根拠とした追記：78件）と、94,391件の負例が抽出された（図3）。

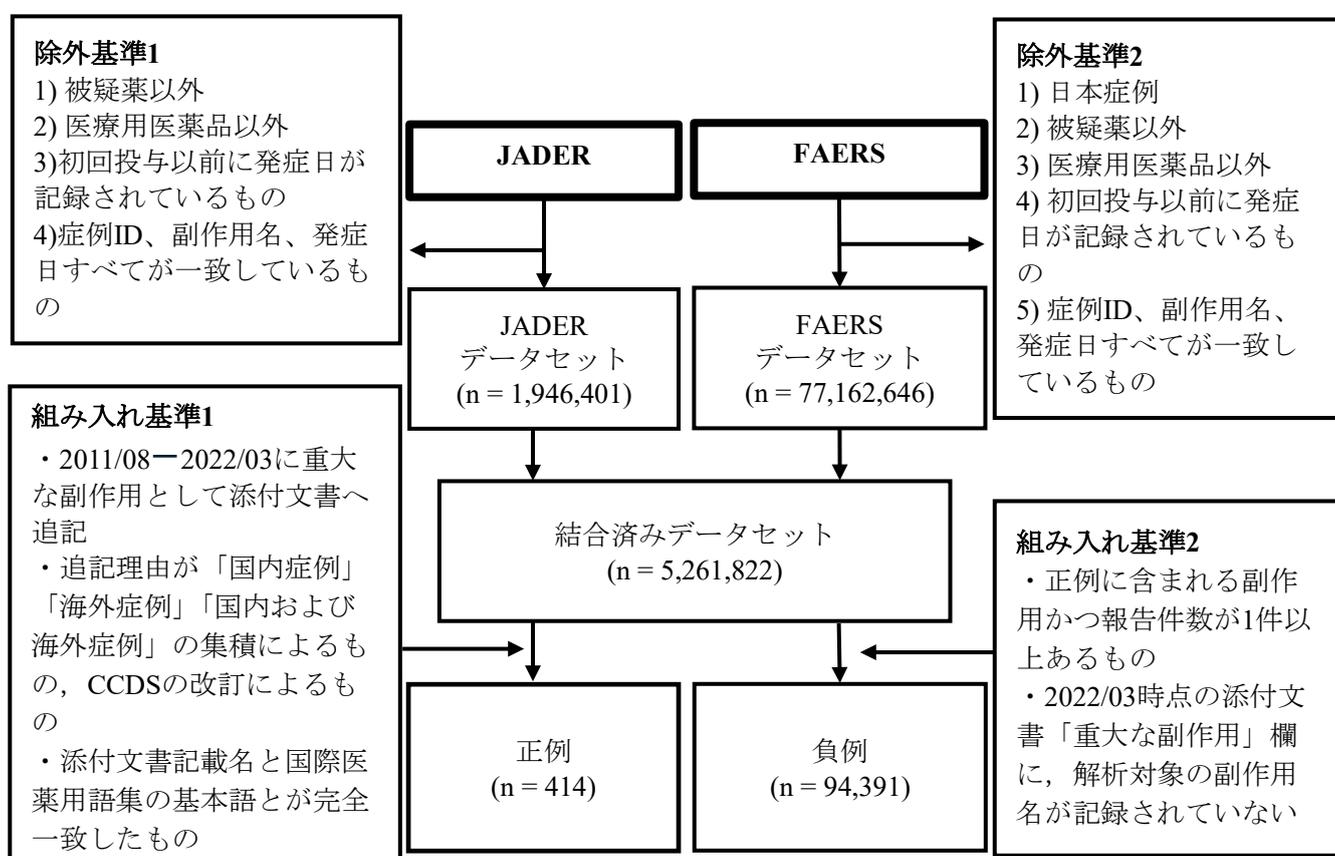


図3 解析対象集団の抽出するフロー

JADER並びにFAERSを使用し、最終的に添付文書の「重大な副作用」欄へ新たな副作用が追記された正例414件、正例に含まれる副作用が報告されているものの、添付文書の「重大な副作用」欄には記載されていない負例94,391件が抽出された。

JADER, Japanese Adverse Drug Event Report; FAERS, U.S. Food and Drug Administration Adverse Event Reporting System

414件の正例中、「重大な副作用」欄へ追記された頻度の高い副作用は、間質性肺疾患（38件）、肝機能異常（34件）、アナフィラキシー反応（31件）であった。また、副作用別に改訂根拠の内訳を確認した結果、間質性肺疾患は36件（94.7%）が国内の副作用症例の集積を根拠としており、一方でアナフィラキシー反応は11件（35.4%）をCCDSの改訂を含む海外症例の集積を根拠としており、副作用ごとに改訂の根拠に大きな偏りがあることが明らかとなった（図4）。

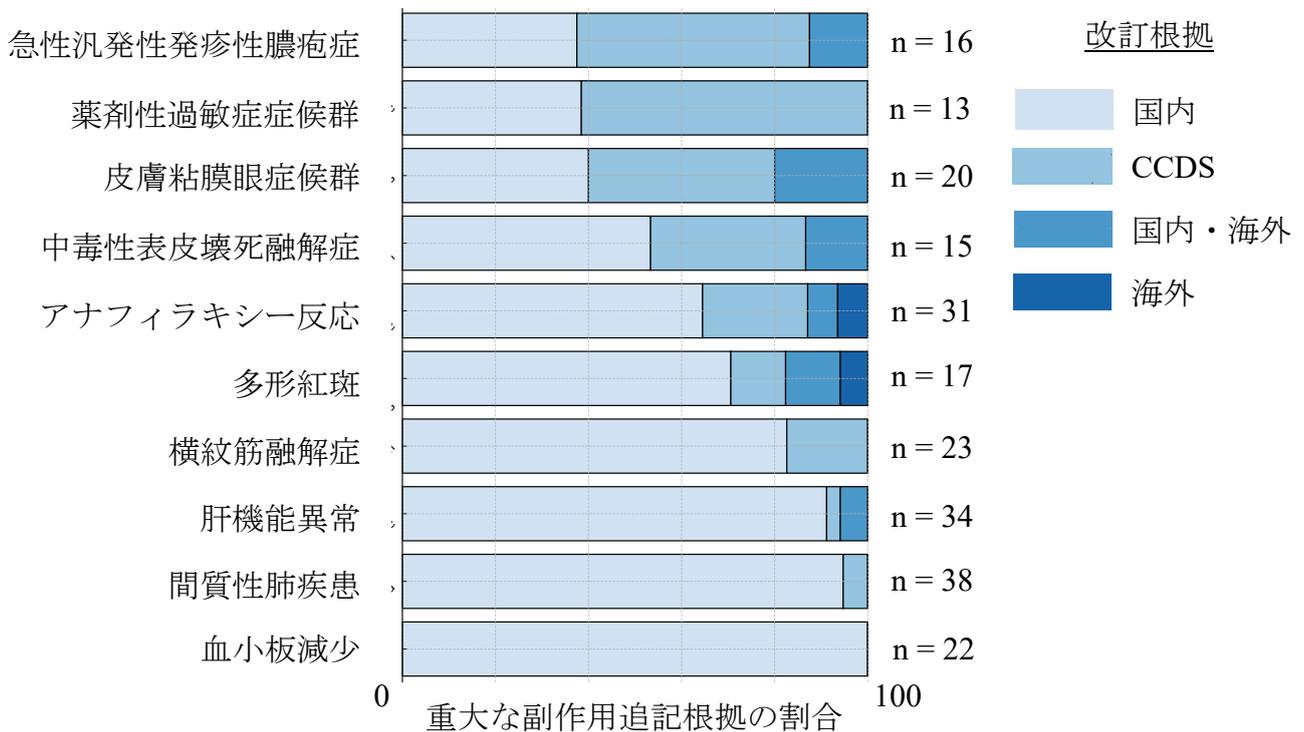


図4 追記頻度の高い副作用別の改訂根拠内訳

2-3-(2) 予測モデルの比較と重要な特徴量

全てのシード値において、有用な特徴量を使用したニューラルネットワークモデルとXGBを一般化線形モデルでアンサンブルしたgeneralized linear model (GLM) blenderが最も高い予測性能を示し、交差検証では平均MCC 0.8771、テストデータでは平均MCC 0.8677であった（表9、表S3）。

表9 各シード値別の最も予測性能が高かったモデル一覧

シード番号	モデル	特徴量セット	MCC (訓練データ)	MCC (テストデータ)
1	GLM blender	有用な特徴量	0.8754	0.8995
2	GLM blender	有用な特徴量	0.8827	0.8639
3	GLM blender	有用な特徴量	0.8891	0.8556
4	GLM blender	有用な特徴量	0.8546	0.8402
5	GLM blender	有用な特徴量	0.8548	0.8792
Mean (SD)			0.8713 (0.0159)	0.8677 (0.0227)

SD, standard deviation; GLM, generalized linear model; MCC, matthews correlation coefficient

最も高い予測性能を示したシード番号1において、テストデータには国内の副作用症例の集積を根拠とする重大な副作用の追記が60件、国内及び海外の副作用症例の集積を根拠とするものが3件、海外の副作用症例の集積を根拠とするものが4件、CCDSの改訂を根拠とするものが16件含まれていた（表10）。これらのテストデータのうち、国内の副作用症例の集積を根拠とするものは7件、国内及び海外の副作用症例の集積を根拠とするものは1件、海外の副作用症例の集積を根拠とするものは3件、CCDSの改訂を根拠とするものは1件が予測の結果偽陰性であった。

表10 重大な副作用の根拠別テストデータにおける予測結果

重大な副作用の追記根拠	真陽性/データ数 (%)
国内の副作用症例の集積	53/60 (88.3)
国内及び海外の副作用症例の集積	2/3 (66.7)
海外の副作用症例の集積	1/4 (25.0)
CCDSの改訂	15/16 (93.8)

最も高い予測性能を示したGLM blenderにおいて、重要な特徴量と選ばれた上位5件は「JADERとFAERS両者を使用して計算した、イベント発生2四半期前時点の全医薬品での報告件数のうち注目する医薬品が占める報告の割合を意味する『Index B derived from JADER + FAERS case reports 2 quarters pre-event』」、 「JADERとFAERSに含まれる欧州

主要5か国からの報告を使用して計算した、イベント発生2四半期前時点のシグナル値を意味する『PRR from JADER case reports + reports from major 5 EU countries 2 quarters pre-event』」、「JADERとFAERSに含まれる欧州主要5か国からの報告を使用して計算した、イベント発生3四半期前時点のシグナル値を意味する『PRR from JADER case reports + reports from major 5 EU countries 3 quarters pre-event』」、「JADERとFAERSに含まれる欧州主要5か国からの報告を使用して計算した、イベント発生3四半期前時点の全医薬品での報告件数のうち注目する医薬品が占める報告の割合を意味する『Index B derived from JADER + reports from major 5 EU countries 3 quarters pre-event』」、「JADERを使用して計算した、イベント発生2四半期前時点の初回報告から経過した四半期数を意味する『Number of quarters elapsed since the first report of the ADR derived from JADER 2 quarters pre-event』」であった。これらの重要な特徴量の中で、「Index B derived from JADER + FAERS case reports 2 quarters pre-event」が最も重度が高かった（表11）。また、各シード値における重要な特徴量上位5件を確認した結果、「PRR from JADER case reports + reports from major 5 EU countries 3 quarters pre-event」がすべてのシード値で共通して重要な特徴量として選抜されており、「Index B derived from JADER + FAERS case reports 2 quarters pre-event」は4つのシード値において重要な特徴量として選抜されていた（表12）。

表11 GLM blenderにおける重要な特徴量上位5件

特徴量名	相対重要度
Index B derived from JADER + FAERS care reports 6 months pre-event	1.000
PRR from JADER case reports + reports from major 5 EU countries 6 months pre-event	0.781
PRR from JADER case reports + reports from major 5 EU countries 9 months pre-event	0.586
Index B derived from reports + reports from major 5 EU countries 9 months pre-event	0.519
Number of quarters elapsed since the first report of the AE derived from JADER, 6 months pre-event	0.363

JADER, Japanese Adverse Drug Event Report; FAERS, U.S. Food and Drug Administration Adverse Event Reporting System; EU, European Union; PRR, proportional reporting ratio; index B, ratio of adverse events associated with a specific drug to those associated with all drugs.

表12 各シード値における最も予測性能の高いモデルの重要な特徴量上位5件

特徴量名	出現回数
PRR from JADER case reports + reports from five main EU countries 3 quarters pre-event	5
Index B derived from JADER + FAERS case reports 2 quarters pre-event	4
Comparison of reports from other healthcare professionals from FAERS 2 and 4 quarters pre-event	3
Increase in reports from other healthcare professionals from FAERS between 2 and 4 quarters pre-event	3
PRR from JADER case reports + reports from major 5 EU countries 2 quarters pre-event	3
Index B derived from JADER + FAERS case reports 3 quarters pre-event	2
Index B derived from JADER + reports from major 5 EU countries 3 quarters pre-event	1
Number of quarters elapsed since the first report of the AE derived from JADER, 2 quarters pre-event	1
PRR derived from JADER and FAERS case reports 3 quarters pre-event	1
PRR derived from the multiplication of JADER and FAERS case reports 2 quarters pre-event	1

Index B, percentage of ADR cases attributed to the suspected drug out of the total ADR cases from all drugs; PRR, proportional reporting ratio; JADER, Japanese Adverse Drug Event Report; FAERS, the U.S. Food and Drug Administration Adverse Event Reporting System ; EU, European Union

2-4 考察

本研究で最も性能が優れた予測モデルはMCCがテストデータにおいて0.8995であり、添付文書の改訂根拠別の正答率は、国内の副作用症例集積によるものが88.3% (53/60)、海外の副作用症例集積が関連するものが78.3% (18/23)であった。海外の副作用症例集積が関連する添付文書の改訂根拠別の正答率は、国内及び海外の副作用症例集積によるものが66.7% (2/3)、海外の副作用症例集積によるものが25% (1/4)、CCDSの改訂によるものが93.8% (15/16)であった。更に、異なるシード値ごとのモデルの性能を平均するとMCCは0.8677であり、シード値によらず安定して高い予測性能を示していることが明らかとなった。この結果より、国内・海外によらず副作用症例の集積を根拠とした添付文書の「重大な副作用」欄への副作用の追記は、副作用報告データベースのJADER、FAERSそして機械学習を用いることで高い精度で予測が可能であることが示唆された。また、この高い予測性能に寄与した重要な特徴量上位10件のうち、3件はJADERのみから作成された特徴量であり、残り7件はJADER並びにFAERSの両データベースの情報を組み合わせて作成された特徴量であった。上位の重要な特徴量の中に、両データベースを組み合わせた特徴量が多数含まれていることは、両データベースを組み合わせた情報が添付文書の「重大な副作用」欄への新たな副作用の追記を予測する上で必須であり、構築した予測モデルが特定の改訂根拠、例えば国内の副作用症例集積や海外の副作用症例集積のみに注目して学習されたものではないことを示している。よって、本研究で開発した予測モデルは、高い予測性能を持ち、特定の改訂根拠に依存しない汎用的な予測モデルであると考えられる。

各シード値の最良なモデルにおける重要な特徴量上位5件を確認した結果、PRR並びにindex Bが共通して重要な特徴量に選ばれていた。選ばれた特徴量の傾向は、第1章の研究と同様であった。また、ワクチン製剤の添付文書改訂について研究した事例においても同様の傾向が報告されており、添付文書の改訂16.5ヶ月前に統計学的シグナル値に有意な差が見られ、早期の統計学的シグナル値がワクチン製剤の添付文書改訂に影響を与えることが示唆された³⁹⁾。これらの一貫した傾向を踏まえると、早期の統計学的シグナル値やindex B

のような他剤や他の副作用に対する相対的指標が添付文書の「重大な副作用」欄へ新たな副作用が追記されることを予測する上で重要であると考えられる。第1章の研究で作成した予測モデルは、国内の副作用症例集積に限定したことで、添付文書の「重大な副作用」欄への副作用の追記を高い精度で予測することを可能にした。図5で示した通り、副作用ごとに添付文書の「重大な副作用」欄へ追記される改訂根拠に偏りが見られる。よって、第1章で作成した予測モデルは、間質性肺疾患や肝機能異常、血小板減少といった改訂根拠の大半が国内の副作用症例の集積であるものを予測する際に特に力を発揮すると考えられる。一方で、本研究で作成した予測モデルは、第1章で作成した予測モデルとは異なり多様な改訂根拠に対応する汎用性の高い予測モデルとなっており、かつ高い予測精度を示している。本予測モデルは注目する副作用の過去の「重大な副作用」欄へ追記された改訂根拠に偏りが認められないケースや、2四半期後に「重大な副作用」として追記されやすい医薬品と副作用の組み合わせを抽出する最初のスクリーニングとして使用することで、予測モデルの特性を生かして安全対策措置の検討の効率化に貢献できると考えられる。

本研究の結果の解釈上の限界を以下に述べる。

- (1) 第1章と同様、予測対象は添付文書の重大な副作用欄に記載された副作用名と、MedDRAのPTが完全一致したものを対象としており、添付文書記載名とPTが一致しない場合は本機械学習モデルを使用することができない。
- (2) 実際の添付文書改訂検討時には、SMQや企業が独自に作成した副作用別のPTリスト等を参考にして、リストに含まれる副作用報告を合算し、添付文書の改訂要否を検討することが一般的である。しかし、SMQは一部の副作用のみにしか対応しておらず、また企業独自のPTリストは入手が困難である。よって、本研究では恣意的な副作用選択を避け、公平な予測モデルを作成するために添付文書記載名とPTが完全一致した組み合わせのみで予測モデルを構築した。今後、副作用ごとに添付文書改訂の要否を検討する際に使用可能な国内・海外で共通のPTリストを作成することで、よりバイアスが少なく、かつ高い予測精度を有する機械学習モデルの構築が可能になると考えられる。

(3) 第1章と同様、本研究で作成した予測モデルも2四半期後のイベントを予測するものであり、上市直後に重大な副作用として追記すべきと判断される副作用には適応できない。

以上の限界はあるものの、本研究はJADERとFAERS、機械学習を使用することで副作用症例の集積を根拠とする添付文書の「重大な副作用」欄への追記を予測する汎用的な予測モデルを開発した初めての研究である。

2-5 小括

JADERとFAERS、機械学習を用いることで、高い精度で2四半期後に国内、海外に関わらず副作用症例の集積を根拠とした添付文書の「重大な副作用」欄へ、新たな副作用が追記されることを予測することが可能であることが示唆された。また、本研究においても医薬品と副作用との関係性の強さを示唆すると従来から考えられてきた再投与再発症例数や投与からイベント発生までの時間といった情報は、「重大な副作用」欄へ新たな副作用が追記されることへ与える影響が小さいことが示唆された。以上の結果から、副作用症例の集積を根拠とした添付文書の「重大な副作用」欄への追記は予測が可能であり、予測する上で早期の統計学的シグナル値や他の薬剤や副作用との相対的指標が重要であると考えられる。

総括

第1章ではJADERと機械学習を用いることで、国内の副作用症例集積を根拠とする添付文書の「重大な副作用」欄への追記を、高い性能で予測可能な予測モデルを構築できた。国内の副作用症例集積を根拠とする添付文書の「重大な副作用」欄への追記は、添付文書の改訂の中で最も頻度の高い安全対策措置であり、この安全対策措置をMCC 0.9を超える高い性能で予測できたことは、安全対策措置の要否検討の効率化に貢献できると考えられる。しかし、安全対策措置の添付文書の「重大な副作用」欄への追記は、約4割を海外の副作用症例集積や、CCDSの改訂等のその他の根拠を理由に実施されている。そのため、安全対策措置の要否検討をより効率化するためにはこれらその他の根拠に適応可能な汎用性の高い予測モデルの構築が重要であると考えた。

そこで第2章では、第1章で構築した予測モデルの限界を補う「国内、海外の副作用症例集積に限らず、副作用症例の集積を根拠とする添付文書の『重大な副作用』欄への追記」を予測可能な、より汎用性の高い予測モデルの構築を検討した。重大な副作用の追記の9割以上が、副作用症例の集積を根拠としており、これらを高い精度で予測できるモデルを構築することで、安全対策措置の要否検討を更に効率化できると考えた。結果、JADERとFAERS、機械学習を用いることでMCC 0.85を超える高い予測性能を示す予測モデルを構築することができた。構築した予測モデルにおける重要な特徴量や、複数のシード値によるモデリングの結果から、一貫して高い予測性能と、両データベース由来の特徴量が上位を占めた点より特定の改訂根拠に依存しない汎用性の高い予測モデルとなっていることが示唆された。

両研究の結果から、副作用症例を根拠とした添付文書の「重大な副作用」欄への追記を予測する上で、従来PMDAや製薬企業が安全対策措置の要否検討で注目していた副作用報告件数や死亡件数、そして医薬品と副作用の因果関係の強さを示唆する被疑薬の再投与情報や、被疑薬の投与から副作用発現までの期間の短さは、重要性が低く、安全対策措置の要否検討の効率化に貢献していないことが一貫して示唆された。むしろ、早期の統計学的シグナル値やindex Bといった他の医薬品、副作用に対する相対的指標が重要であることが示唆され、

この結果はワクチン製剤の添付文書の改訂に影響を与える因子を探索した先行研究の結果とも一致していた。

本研究の成果は、副作用報告データベースと機械学習を組み合わせることで、副作用症例の集積を根拠とする添付文書の「重大な副作用」欄への追記を早期に予測可能なことを示唆し、また安全対策措置の要否検討の際に従来重要と考えられてきた情報は、重要性が高くないことを初めて見出した研究である。さらに、従来の統計学的シグナル検出手法の性能と対比したことで、それぞれの手法の役割を明確化することができた。

本研究から得られた知見は、PMDAにおける安全対策措置の検討に関する標準的な作業の流れのフローを改善するだけでなく、製薬企業においても、従来より効率的に添付文書へ重大な副作用を追記すべきか否かを判断できるようになると考えられる。例えば、膨大な医薬品と副作用の組み合わせの中からリスク候補であるシグナルを検出する目的においては従来通り統計学的シグナル検出手法を使用し、特に重要性の高い添付文書の改訂を伴う重大な副作用の検出には本研究で開発した機械学習モデルを使用することで、双方の強みを活かした安全対策措置の要否検討が可能となると考えられる。また、本研究で開発した2つの機械学習モデルは相補的な関係にあり、第2章で開発した予測モデルはその汎用性の高さを活かし、収集された膨大な副作用報告全体を対象にスクリーニングし、重大な副作用として追記される可能性が高い医薬品と副作用の組み合わせを抽出する役割が適していると考えられる。一方、第1章で開発した予測対象が限定されるものの、より高い予測性能を有していることから、国内の副作用症例の集積を根拠として重大な副作用が追記されやすい間質性肺疾患や肝機能異常、血小板減少など特定の副作用、または海外より先に日本で承認された医薬品、日本でのみ販売されている医薬品における安全対策措置の要否を検討するスクリーニングで特にその性能の高さを活かせると考えられる。

本研究で開発された機械学習モデルが、安全対策措置の要否検討に必要なプロセスすべてを代替するものではない点に注意が必要である。本モデルは、あくまで従来の統計学的シグナル検出手法の弱点である偽陽性の検出を減らし、安全対策措置の一つである重大な副作用の追記を直接かつ早期に、高精度で予測可能な点に強みを持つ。よって、本モデルによって検出され

た医薬品と副作用の組み合わせに対して、専門家による精査は引き続き必要である。また、重大な副作用以外のリスクの検出は本モデルの対象外のため、このようなリスクの検出には従来の安全対策措置の要否検討プロセスが引き続き必要である。

本モデルは、定期的なシグナル評価のタイミングで実行することで、重大な副作用の追記が発生する可能性が高い医薬品と副作用の組み合わせを検出し、専門家が精査すべき医薬品と副作用の組み合わせを減らすことで安全対策措置の要否検討の効率化を実現している。今後、本モデルをPMDAや製薬企業が安全対策措置の要否検討に使用し、定量的な評価を行うことが期待される。

謝辞

本研究の遂行と博士論文の作成にあたり、終始、懇切なる御指導と御鞭撻を賜り、親身に励まし頂きました、指導教官の名古屋市立大学大学院薬学研究科レギュラトリーサイエンス分野 頭金 正博 教授に厚く御礼申し上げます。

本研究を遂行する環境を整えて頂き、研究上の有用な御助言と心温まる励ましを賜りました名古屋市立大学大学院薬学研究科レギュラトリーサイエンス分野 安部 賀央里 講師に深く御礼申し上げます。

本研究の議論を通じて様々な視点から有用な御助言を頂きました名古屋市立大学大学院薬学研究科レギュラトリーサイエンス分野の皆様にご心より感謝致します。

大変お忙しい中、本論文を審査頂き、多角的な視点から本論文への御指導を頂きました名古屋市立大学大学院薬学研究科 牧野 利明 教授、糸 和彦 教授、舘 知也 教授に深く御礼申し上げます。

引用文献

1. European Medicines Agency. Annual report on EudraVigilance for the European parliament, the council, and the commission. [Internet]. Available from: https://www.ema.europa.eu/en/documents/report/2020-annual-report-eudravigilance-european-parliament-council-commission_en.pdf; 2020
2. Suzuki Y, Kishi T, Nakamura M, et al. Evaluation of factors influencing addition of clinically significant adverse reactions section in drug package inserts. *Jpn J Drug Inform.* 2017;19:17–23.
3. The importance of pharmacovigilance: Safety monitoring of medicinal products. WHO Lib Catalog. 2002. [Internet]. Available from: <https://www.who.int/publications/i/item/10665-42493>
4. Pharmaceuticals and Medical Devices Agency. Revisions of PRECAUTIONS [Internet]. Available from: <https://www.pmda.go.jp/safety/info-services/drugs/calling-attention/revision-of-precautions/0001.html>
5. Pharmaceuticals and Medical Devices Agency. Outline of post-marketing safety measures. [Internet]. Available from: <https://www.pmda.go.jp/english/safety/outline/0001.html>
6. Pharmaceuticals and Medical Devices Agency. Standard workflow for consideration of safety measures such as revision of electronic drug product package inserts. [Internet]. Available from: <https://www.pmda.go.jp/files/000243072.pdf>.
7. Pharmaceuticals and Medical Devices Agency. Reference: Standard workflow for consideration of safety measures. [Internet]. Available from: <https://www.pmda.go.jp/files/000243073.pdf>.
8. CIOMS practical aspects of signal detection in pharmacovigilance. Report of CIOMS Working Group VIII. [Internet]. Available from: https://cioms.ch/working_groups/working-group-viii/
9. European Medicines Agency. Guideline on Good Pharmacovigilance Practices (GVP) Module IX (Rev. 1). [Internet]. Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-good-pharmacovigilance-practices-gvp-module-ix-signal-management-rev-1_en.pdf.
10. Evans SJ, Waller PC, Davis S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiol Drug Saf.* 2001;10:483–486.
11. Hauben M, Madigan D, Gerrits CM, et al. The role of data mining in pharmacovigilance. *Expert Opin Drug Saf.* 2005;4:929–948.
12. Stephenson WP, Hauben M. Data mining for signals in spontaneous reporting databases: proceed with caution. *Pharmacoepidemiol Drug Saf.* 2007;16:359–365.

13. Szarfman A, Machado SG, O'Neill RT. Use of screening algorithms and computer systems to efficiently signal higher-than-expected combinations of drugs and events in the US FDA's spontaneous reports database. *Drug Saf.* 2002;25:381–392.
14. Courtois É, Tubert-Bitter P, Ahmed I. New adaptive lasso approaches for variable selection in automated pharmacovigilance signal detection. *BMC Med Res Methodol.* 2021;21:271.
15. Courtois É, Pariente A, Salvo F, et al. Propensity score-based approaches in high dimension for pharmacovigilance signal detection: an empirical comparison on the French spontaneous reporting database. *Front Pharmacol.* 2018;9:1010.
16. Bae JH, Baek YH, Lee JE, et al. Machine learning for detection of safety signals from spontaneous reporting system data: example of nivolumab and docetaxel. *Front Pharmacol.* 2020;11:602365.
17. Caster O, Juhlin K, Watson S, et al. Improved statistical signal detection in pharmacovigilance by combining multiple strength-of-evidence aspects in vigiRank. *Drug Saf.* 2014;37:617–628.
18. Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. *BMC Medical Research Methodology.* 2019;19:64.
19. Vogenberg FR. Predictive and prognostic models: Implications for healthcare decision-making in a modern recession. *Am Health Drug Benefits.* 2009;2:218-222.
20. International conference on harmonisation of technical requirements for registration of pharmaceuticals for human use. E2D Post-approval safety data management: Note for guidance on definitions and standards for expedited reporting. ICH. 2003. [Internet]. Available from: <https://www.ich.org/page/efficacy-guidelines>.
21. Summary of investigation results “Amantadine hydrochloride”. [Internet]. Available from: <https://www.pmda.go.jp/files/000207392.pdf>
22. Summary of investigation results “Panitumumab (genetical recombination)”. [Internet]. Available from: <https://www.pmda.go.jp/files/000206607.pdf>
23. European Medicines Agency. Guideline on Good Pharmacovigilance Practices (GVP) Module VIII (Rev. 3). [Internet]. Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-good-pharmacovigilance-practices-gvp-module-viii-post-authorisation-safety-studies-rev-3_en.pdf
24. Insani WN, Pacurariu AC, Mantel-Teeuwisse AK, et al. Characteristics of drugs safety signals that predict safety related product information update. *Ther Innov Regul Sci.* 2018;27:789–796.
25. Chen T, Guestrin C. Proceedings of the 22nd ACM SIGKDD International conference on knowledge discovery and data mining, New York, NY: ACM; 2016. p. 785–794.
26. Ke G, Meng Q, Finley T, et al. Lightgbm: A highly efficient gradient boosting decision tree. *Adv Neural Inform Process Syst.* 2017;30:3146–3154.

27. Vapnik V, Lerner A. Pattern recognition using generalized portrait method. *Autom Remote Control*. 1963;24:774–780.
28. Aruna S, Rajagopalan SP. A novel SVM based CSSFFS feature selection algorithm for detecting breast cancer. *Int J Comput Appl*. 2011;31:14–20.
29. Noble WS. What is a support vector machine? *Nat Biotechnol*. 2006;24:1565–1567.
30. Boser BE, Guyon IM, Vapnik VN. A training algorithm for optimal margin classifiers (COLT'92), Proceedings of the 5th annual workshop on computational learning theory, Pittsburgh; 1992. p. 27–29.
31. Amari S, Wu S. Improving support vector machine classifiers by modifying kernel functions. *Neural Netw*. 1999;12:783–789.
32. Hofmann T, Schölkopf B, Smola AJ. Kernel methods in machine learning. *Ann Statist*. 2008;36:1171–1220.
33. Raschka S. MLxtend: providing machine learning and data science utilities and extensions to Python's scientific computing stack. *J Open Source Softw*. 2018;3:638.
34. Akiba T, Sano S, Yanase T, et al., Optuna: A next-generation hyperparameter optimization framework. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining; Anchorage, AK, USA; 2019. p. 2623–2631.
35. Breiman L. Random forests. *Mach Learn*. 2001;45:5–32.
36. CIOMS VI working group; 2005. Management of safety information from clinical trials, Appendix. [Internet]. Available from: https://cioms.ch/wp-content/uploads/2017/01/Mgment_Safety_Info.pdf.
37. Bate A, Evans SJ. Quantitative signal detection using spontaneous ADR reporting. *Pharmacoepidemiol Drug Saf*. 2009;18:427–436.
38. Nomura K, Takahashi K, Hinomura Y, et al. Effect of database profile variation on drug safety assessment: an analysis of spontaneous adverse event reports of Japanese cases. *Drug Des Devel Ther*. 2015;9:3031–41.
39. Suzuki S, Imai S, Mitsuboshi S, et al. Detection of vaccine adverse events before package insert revisions using a Japanese spontaneous reporting system. *J Clin Pharmacol*. 2023;63:903–908.

Supplemental

表S1. 各機械学習モデルの予測性能一覧

モデル名	特徴量選 択有無	訓練データ				テストデータ			
		MCC	Precision	Recall	AUC	MCC	Precision	Recall	AUC
SVM - RBF	無し	0.785	0.899	0.693	0.966	0.767	1.000	0.591	0.983
Kernel		(0.044)	(0.066)	(0.068)	(0.024)				
Extreme	無し	0.744	0.981	0.575	0.787	0.781	1.000	0.614	0.807
Gradient		(0.093)	(0.033)	(0.128)	(0.064)				
Boosting									
Light	無し	0.727	0.960	0.560	0.780	0.767	0.981	0.602	0.801
Gradient		(0.099)	(0.042)	(0.125)	(0.063)				
Boosting									
SVM - RBF	有り	0.938	0.989	0.892	0.989	0.922	0.987	0.864	0.998
Kernel		(0.032)	(0.019)	(0.047)	(0.012)				
Extreme	有り	0.772	0.973	0.619	0.809	0.767	0.981	0.602	0.801
Gradient		(0.078)	(0.035)	(0.101)	(0.051)				
Boosting									
Light	有り	0.754	0.954	0.604	0.802	0.789	1.000	0.625	0.812
Gradient		(0.078)	(0.042)	(0.113)	(0.056)				
Boosting									

RBF-SVM, support vector machine with the radial basis function kernel; EFS, exhaustive feature selection; MCC, matthews correlation coefficient; AUC, area under the curve.

表S2. 各機械学習モデルにおける重要な特徴量

モデル	選択された重要な特徴量
RBF-SVM	<ul style="list-style-type: none"> • Index B • Number of newly reported patients from a quarter ago • Average number of patients reported per quarter • ROR
Light Gradient Boosting	<ul style="list-style-type: none"> • Index B • Number of quarters that have elapsed since the first adverse event was reported • Number of newly reported patients from a quarter ago • Average number of patients reported per quarter • Upper limit of the confidence interval of ROR
Extreme Gradient Boosting	<ul style="list-style-type: none"> • Number of patients • Index B • Number of newly reported patients from a quarter ago • Average number of patients reported per quarter • Lower limit of the confidence interval of PRR • Upper limit of the confidence interval of PRR

PRR, proportional reporting ratio; ROR, reporting odds ratio; index A, the ratio of the target adverse event number of specific drug to all adverse events numbers that reported for the same drug; index B, ratio of adverse events covered by a particular drug to those covered by all drugs.

表S3. 重要な特徴量上位100個を使用したモデリング結果

シード番号	モデル	MCC (訓練データ)	MCC (テストデータ)
1	GLM blender	0.8651	0.8799
2	AVG blender	0.7785	0.7813
3	GLM blender	0.8811	0.8104
4	GLM blender	0.8205	0.7769
5	GLM blender	0.7026	0.7446
Mean (SD)		0.8096 (0.0720)	0.7986 (0.0457)

SD, standard deviation; GLM, generalized linear model; AVG, average; MCC, matthews correlation coefficient.