

厚生労働科学研究費
厚生労働行政推進調査事業費
 補助金総合研究報告書

令和 5 年 5 月 30 日

厚生労働大臣
~~(国立医薬品食品衛生研究所長)~~ 殿
~~(国立保健医療科学院長)~~

(研究代表者)

研究者の住所	〒606-8507 京都府京都市左京区聖護院川原町 54
所属機関名	京都大学
部署・職名	医学研究科 特定准教授
氏名	岡本 里香

補助事業名 : 令和 4 年度 厚生労働科学研究費
厚生労働行政推進調査事業費 補助金
 (政策科学総合 (臨床研究等 ICT 基盤構築・人工知能実装研究事業) 研究事業)

研究課題名 (課題番号) : 薬局ヒヤリハット事例に対する安全管理対策評価に関する AI 開発 (20AC1005)
 研究実施期間 : 令和 2 年 6 月 1 日から令和 5 年 3 月 31 日まで
 国庫補助金精算所要額 : 金 58,900,000 円也 (※研究期間の総額を記載すること)
 (うち間接経費 13,591,000 円)

上記補助事業について、厚生労働科学研究費補助金等取扱規程 (平成 10 年 4 月 9 日厚生省告示第 130 号) 第 16 条第 3 項の規定に基づき下記のとおり研究成果を報告します。

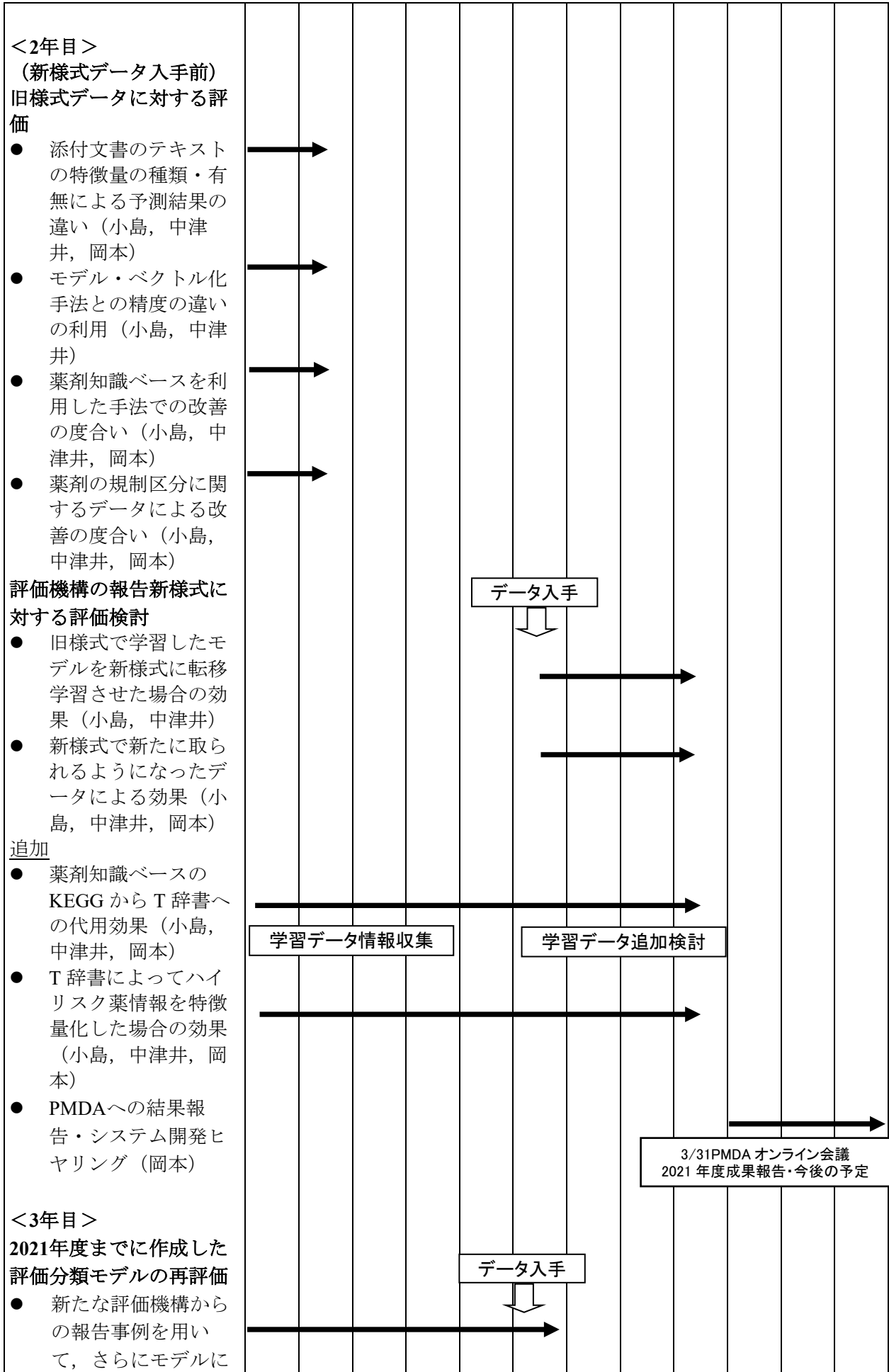
記

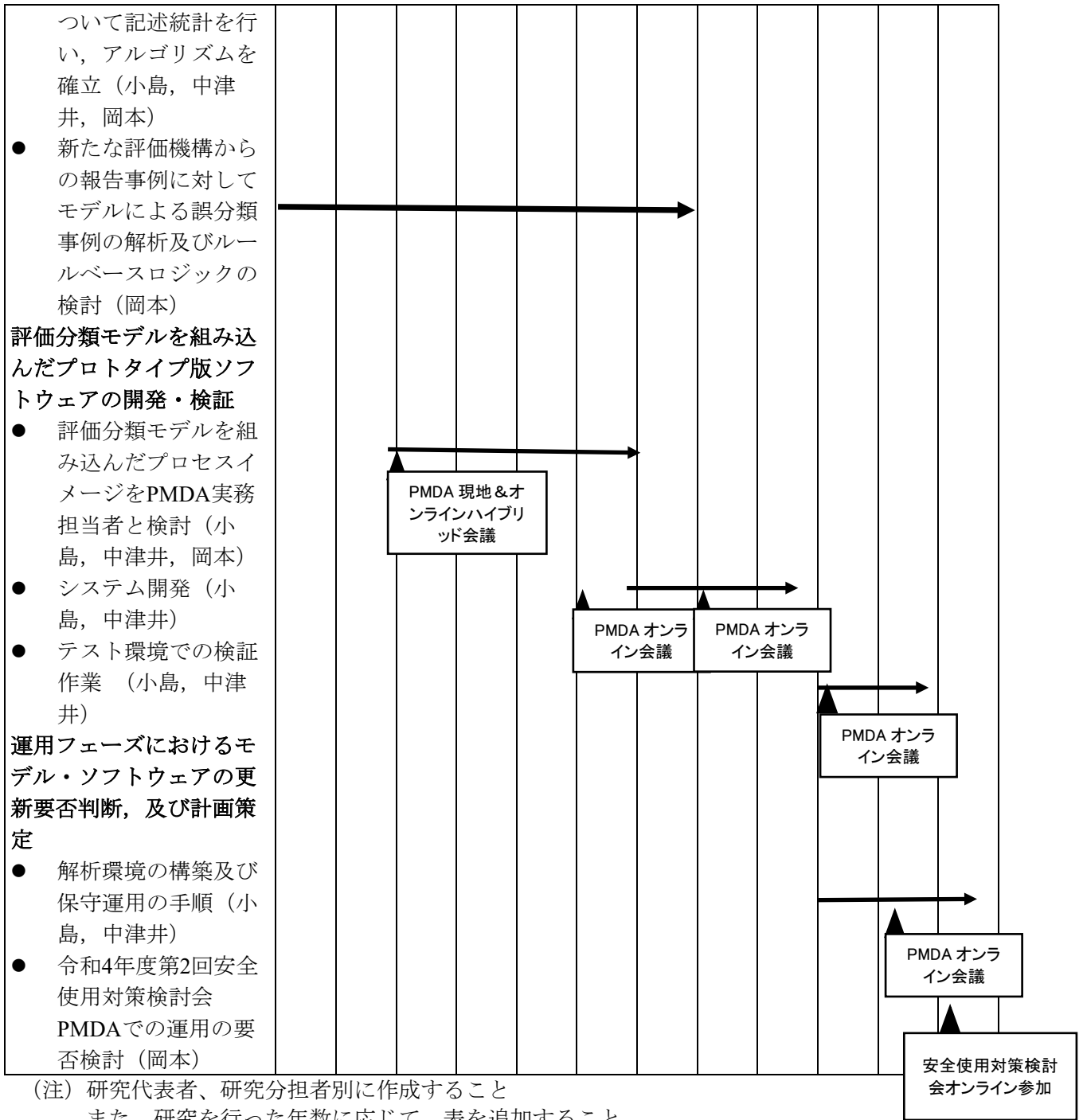
1. 研究概要の説明
 (1) 研究者別の概要

所属機関・部署・職名	氏名	分担した研究項目及び研究成果の概要	研究実施期間	配分を受けた研究費	間接経費
京都大学 大学院医学研究科 人間健康科学系専攻 ビッグデータ医科学分野・特定准教授	岡本里香	研究期間を通じて、PMDA や外部委託先との連携を行い、研究の進捗管理、方針検討等の研究全体の統括を実施した。 2020 年度 <分担項目> 薬剤に関する情報 (添付文書、複数名称、規制区分など) の仕様ノウハウ <研究成果概要> 【探索的研究で構築したモデルの精度向上】、【モデルの精度検証及び精度向上検討】において、精度向上の過程で新たに追加すべき薬剤情報を検討し、複数名称や規制区分として「劇薬」を特徴量として追加した。	令和 2 年 6 月 1 日から令和 5 年 3 月 31 日	40,120,725	13,591,000

		<p>2021 年度 <分担項目> 薬剤に関する情報（添付文書、複数名称、規制区分など）の仕様ノウハウ <研究成果概要> 【評価機構の報告新様式に対する評価検討】において、これまで開発した評価分類モデルの適用可能性について、薬剤特徴量側から検討した。また PMDA 実装化に向け、商用の観点での辞書変更を検討した。</p> <p>2022 年度 <分担項目> ・研究全体の統括 ・薬剤に関する情報（添付文書、複数名称、規制区分など）の仕様ノウハウ <研究成果概要> 評価分類モデルの再評価、評価分類モデルを組み込んだプロトタイプ版ソフトウェアの開発・検証、運用フェーズにおけるモデル・ソフトウェアの更新要否判断、及び計画策定において、医薬品視点で検討した。</p>			
<p>山口大学大学院医学系研究科・医学部附属病院 AI システム医学・医療研究教育センター 特命教授</p>	<p>中津井雅彦</p>	<p>2020 年度 <分担研究項目> 機械学習の手法検討とその精度結果の分析 <研究成果概要> 【探索的研究で構築したモデルの精度向上】、【モデルの精度検証及び精度向上検討】において、精度向上の過程で、機械学習の手法として 2 種類を使用し、その比較の解析結果を評価した。また分類精度に影響している要因を検討するため、分類閾値ごとの差分検討やサンプリング方法についても検討結果について評価した。</p> <p>2021 年度 <分担項目> 機械学習の手法検討とその精度結果の分析 <研究成果の概要> 【評価機構の報告新様式に対する評価検討】において、これまで開発した評価分類モデルの適用可能性について、解析結果について評価した。</p> <p>2022 年度 <分担項目> 機械学習の手法検討とその精度結果の分析</p>	<p>令和 2 年 6 月 1 日から 令和 5 年 3 月 31 日</p>	<p>3,844,000</p>	

		<p><研究成果概要> 評価分類モデルの再評価，評価分類モデルを組み込んだプロトタイプ版ソフトウェアの開発・検証，運用フェーズにおけるモデル・ソフトウェアの更新要否判断，及び計画策定において、解析・システム開発視点で検討した。</p>			
京都大学 大学院医学 研究科 人間健康科 学系専攻 ビッグデー タ医科学分 野・特定講 師	小島諒介	2020 年度 <分担項目> 機械学習の手法検討とその精度結果の分析 <研究成果概要> 【探索的研究で構築したモデルの精度向上】，【モデルの精度検証及び精度向上検討】において，精度向上の過程で，機械学習の手法として2種類を使用し，その比較を実施した．また分類精度に影響している要因を検討するため，分類閾値ごとの差分検討やサンプリング方法についても検討した． 2021 年度 <分担項目> 機械学習の手法検討とその精度結果の分析 <研究成果概要> 【評価機構の報告新様式に対する評価検討】において，これまで開発した評価分類モデルの適用可能性について，これまでの機械学習手法に対して，新様式で適正な手法を評価検討し，同定した． 2022 年度 <分担項目> 機械学習の手法検討とその精度結果の分析 <研究成果概要> 評価分類モデルの再評価，評価分類モデルを組み込んだプロトタイプ版ソフトウェアの開発・検証，運用フェーズにおけるモデル・ソフトウェアの更新要否判断，及び計画策定において、解析・システム開発視点で検討した。	令和 2 年 6 月 1 日から 令和 5 年 3 月 31 日	1,344,275	研究代表 者 一括計上





(3) . 研究成果の説明

研究の目的：

独立行政法人医薬品医療機器総合機構（以下、PMDA）では、公益財団法人日本医療機能評価機構（以下、評価機構）が薬局ヒヤリ・ハット事例収集・分析事業（以下、本事業）に基づき収集・分析・公表した「ヒヤリ・ハット事例」に対し、医薬品の名称・包装等の観点から安全対策を講じる必要がないか検討を行っている。本事業は、全国の薬局からの事例を収集・分析し、薬局における医療安全対策に有用な情報を共有するなど、医療安全対策の一層の推進を図ることを目的として行われている。収集される事例は、薬局で発生した調剤や疑義照会等に関するヒヤリ・ハット事例であるが、例えば、調剤に関する事例のうち、薬剤の名称の視覚的、音韻的な類似に起因したことで薬剤を取り違えた等の事例の場合、PMDA では製造販売会社に対し、薬剤の取違えを防ぐための注意喚起の必要性等について指導する、といった医薬品の物的要因に対する安全管理対策の評価・検討を実施している。当該評価・検討の過程で、PMDA では、評価機構が公表する事例に対して、人による目で、以下の評価1～5に分類することにより、安全対策の必要な事例を選定している。近年、評価機構の公表事例は約10万近い数におよび、PMDA での人による評価分類の負担が増加していることを踏まえ、本研究では、この評価分類を人工知能（以下、AI）が行えるようにすることを目的とし、PMDA での評価分類の負担を軽減し、また事例数が増加している中でも、一貫性のある評価を行えるようにすることを目指す。

<5 評価分類>

評価1：医薬品の安全使用に関して製造販売業者等による対策が必要又は可能と考えられた事例

評価2：製造販売業者等により既に対策がとられているもの、もしくは対策を既に検討中の事例

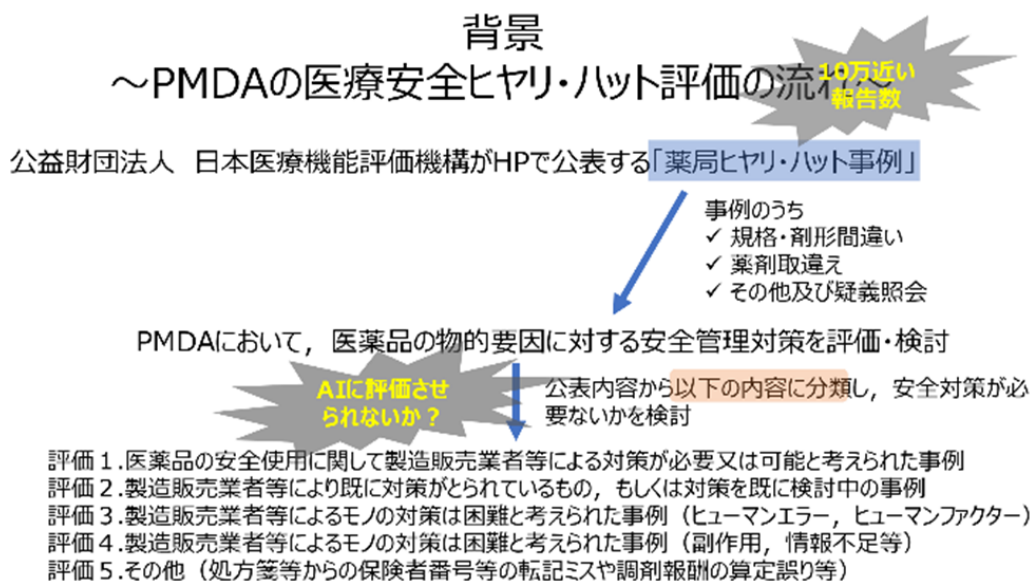
評価3：製造販売業者等によるモノの対策は困難と考えられた事例（ヒューマンエラー、ヒューマンファクター）

評価4：製造販売業者等によるモノの対策は困難と考えられた事例（副作用、情報不足等）

評価5：その他（処方箋等からの保険者番号等の転記ミスや調剤報酬の算定誤り等）

研究結果の概要：

本研究で、交付申請時点では、PMDA の評価1～5の5分類をPMDA の人による評価に置き換わるAI開発を計画した。



しかしながら、評価分類モデルの精度向上を検討する中で、例えば、PMDA は評価時点で、製造販売業者による安全対策の実施あるいは予定を確認したうえで、評価結果に反映していることや、PMDA の人による評価において判断の揺れがあること等が挙げられ、AI では解決できない問題があ

り、PMDA 評価と評価分類モデルによる評価を完全に一致させることは難しいことが判明した。そこで、1次スクリーニングの位置づけで、評価分類モデル（AI）にルールベースロジックを組み合わせることにより、安全対策措置を検討する必要のある事例が低レベルのリスク分類に誤分類することを防止しつつ、2次スクリーニングとして調査すべき事例に絞る（調査不要の事例をソートすることにより、PMDA の人による評価の業務負担を軽減することを目的し、ルールベースロジックの検討を PMDA と共に実施した。

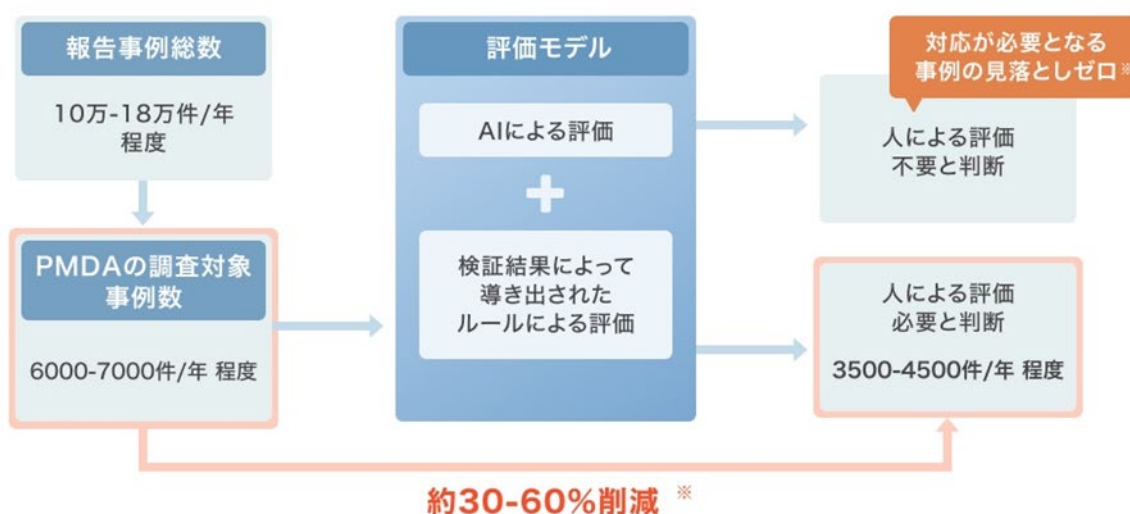
AI部分



ルールベース部分

その結果、誤分類を0件とし、約30~60%の事例について PMDA での人による評価が不要と判定された。

本施策における評価モデルによるPMDA調査件数の削減効果



*検証時において、2020年3月-2022年6月のPMDA評価対象件数に対して、評価モデルが人による評価が不要と判断したものの割合。PMDAでは数か月単位ごとに報告事例を評価しており、評価タイミングごとの事例種類の分布に応じて削減割合が変動。

我々は、AI とルールベースロジックを組み合わせたモデルを最終モデルとし、PMDA の人による評価を併用し、事例に対する評価を実施する運用ルールを決定した。当該運用については、2023年3月1日開催 PMDA の令和4年度第2回安全使用対策検討会において、2023年度4月以降のPMDAの本業務において事例のスクリーニングに試行的に導入することが了承され、本研究期間内に実装化に繋げることができた。

なお、PMDA の評価分類は、5分類から類型Ⅰ～Ⅲの3分類に変更となったが、以下の通り、これまでの評価1及び評価2は類型Ⅰ及び類型Ⅱに対応しており、評価3~5が類型Ⅲとして扱われるようになっており、分類内容は同じであるため、本開発の評価分類モデルへの影響はなかった。

<現行分類>

類型Ⅰ：製造販売業者等により、速やかに新たな対策をとる必要がある事例

類型Ⅱ：製造販売業者等により、既に対策がとられている事例、または既に対策が検討されている事例

類型Ⅲ：上記以外の事例（事例の集積が必要な事例、ヒューマンエラーによる事例など）

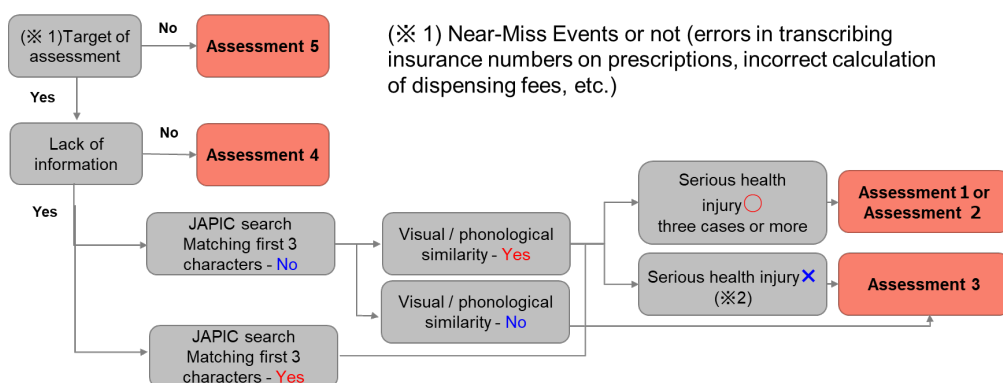
本研究で開発したAIは、当初の人の評価に100%置き換えることはできなかったが、当初の目的の一つであったPMDAにおける業務負担の軽減については達成することができた。また、医療安全におけるAI活用に関する課題を解決すべく、ルールベースロジックを組み合わせることによる効果を示すことができ、今後の医療安全領域での評価支援に関するAI開発の先駆けとなる開発事例として本研究を報告する。

研究の実施経過

【開発する評価分類モデルと精度】

本研究開始時点では、PMDA が当時行っていた評価 1～5 の 5 分類を行う AI 開発を目指すこととしたが、本研究の先行研究として実施した探索的研究において、評価機構の報告数に対して、PMDA が安全対策を検討した事例（評価 1 及び評価 2 が該当）は非常に少ないため、過去の対策検討事例は学習データとしては数が不十分であり、評価 1 か評価 2 かを分類することはできないという課題が挙げられた。そこで、開発するモデルの分類では「評価 1」「評価 2」を区別せず、「評価 1-2」として分類し、他、「評価 3」「評価 4」「評価 5」の計 4 分類とした。また、評価 1 及び評価 2 は安全管理対策が必要な事例であるため、モデルによる分類が誤ってこれらを評価 3、4 あるいは 5 の低リスクに分類してしまうと、安全対策が必要な事例を見逃すことになる。そこで、PMDA の評価 3～5 の事例は評価分類モデルで「評価 1-2」に分類されることを許容することとした。

以上を踏まえ、次に示す評価スキームに準じた評価分類モデルを作成した。



(※2) As a general rule, "If there is no similarity, classify as human factor (Assessment 3)"

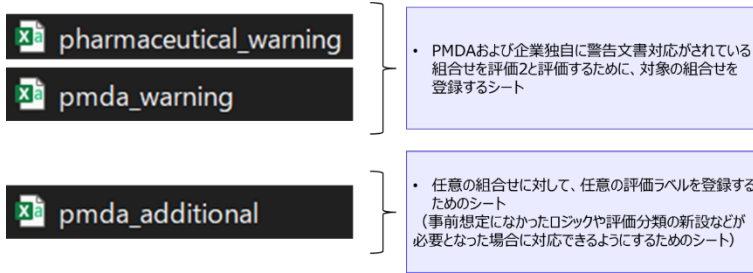
【評価分類モデルの分類精度の向上】

評価分類モデルの精度は、precision, recall, f1-score を評価指標とし、特徴量の検討、自然言語処理技術方法（Universal Sentence Encoder と word2vec の比較）、サンプリング方法（Random Over Sampling と Random Under Sampling の比較）、評価スキーム（no-step 法と 2-step 法の比較）を行った。分類精度向上の検討は、本研究期間の 2 年目までは、評価機構の旧様式の報告形式を用いて検討したが、新様式の報告形式に対しても、同様の分類精度を検出することができ、3 年目時点で新様式に最適な評価分類モデルの条件を決定した。

特徴量	販売名のみ類似性	編集距離など複数の手法を利用
	一般的名称まで含めた類似性	T 辞書を利用して特徴量化
	T 辞書の同義語を利用した類似性	T 辞書を利用して特徴量化
	テキスト情報（Universal Sentence Encoder 使用）	事例の詳細 背景要因 改善策 発生要因 を encode して連結
	薬効情報	T 辞書を利用して特徴量化
	ハイリスク薬情報	「特定薬剤管理指導加算等の算定対象となる薬剤一覧」を利用して特徴量化
分類評価スキーム	no-step 法	
サンプリング方法	Random Over Sampling	

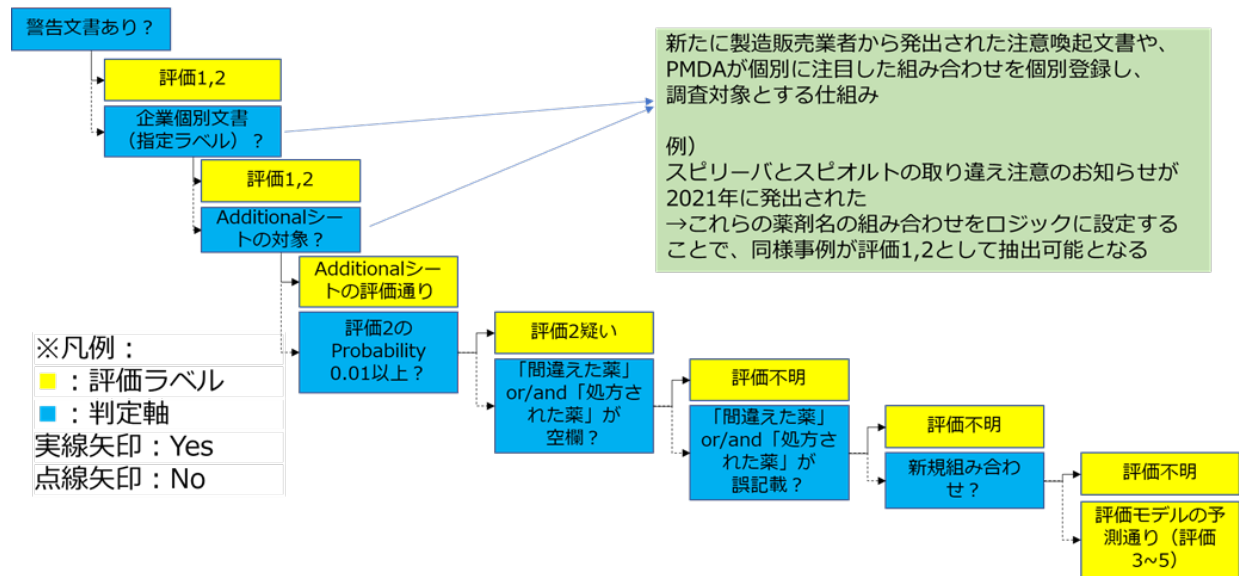
【評価分類モデルの実装化検討】

最適な評価分類モデルの条件で分類した場合においても、評価 1-2 が評価 3 に分類される事例があった。この誤分類の事例を分析し、ルールベースとして、以下のシートを PMDA と協力して作成した。



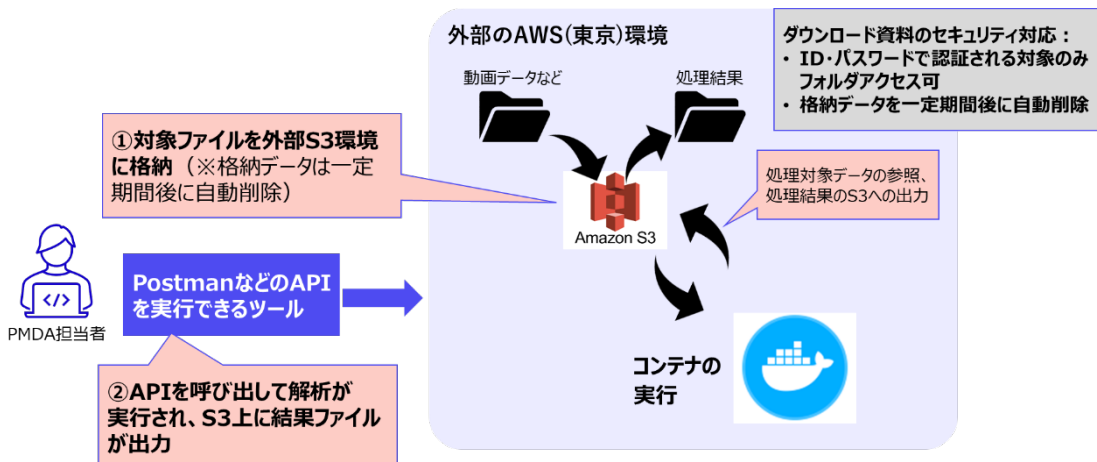
上記シートを以下のスキームに適用させることにより、評価分類モデルとルールベースロジックを組み合わせた。その結果、評価 1-2 の事例が評価 3 に分類される誤分類を 0 事例とすることができた。

AI分類結果+ルールベースロジック→最終判定結果



【PMDA 実装環境の検討】

PMDA システム担当も含め、PMDA で本開発システムをどのように実装できるのか、PMDA 内でのセキュリティーポリシーなども踏まえて検討し、以下を構築した。



本研究期間において、当該研究交付申請時に以下を計画したが、実装に繋げることができた。

これまでの探索的研究で認められた課題

〈誤分類のパターン〉

- 薬局・PMDA側に起因するもの(判断揺れ・データ欠損など)
- 薬剤の名称が一意でないことに起因するもの
 - 一般名やジェネリック名など複数名称があるもの
 - 既に名称が変更されてしまっているもの

2020年度

探索的研究で構築したモデルの精度向上

- 複数名称に関しては薬剤知識ベースを利用した手法で改善検討
- 薬剤の規制区分に関するデータでの検討
- 令和元年(2019年)度後期データでモデルの精度検証及び精度向上検討

2021年度

評価機構の報告新様式に対する評価検討

- PMDAの新様式に対する評価ルールの変更点を踏まえ、2020年度で改良したモデルで新様式報告に対する評価分類の検討と課題抽出

2022年度

評価機構の報告新様式に対する評価のアルゴリズム確立

2021年度の課題を改善し、実用化までのアルゴリズムを確立

研究成果の刊行に関する一覧表：

1. 岡本里香, 中津井雅彦, 小島諒介 薬局ヒヤリ・ハット事例に対する安全管理対策評価に関するAI開発 第7回日本医療安全学会学術総会(オンライン) 2021年5月25日
2. 岡本里香, 中津井雅彦, 小島諒介 薬局ヒヤリ・ハット事例に対する安全管理対策評価に関するAI開発 第11回レギュラトリーサイエンス学会学術大会 2021年9月17日
3. Rika Okamoto, Ryosuke Kojima, Masahiko Nakatsui. Toward AI-supported Evaluation for Safety Control Measures against Near-Miss Events in Pharmaceutical Products. Safety Science に 2023.1.submitted, 現在 reviewer comments に対して投稿論文修正中

研究成果による知的財産権の出願・取得状況：

特記事項なし

研究により得られた成果の今後の活用・提供：

2023年3月1日開催の令和4年度第2回安全使用対策検討会において、2023年度4月以降のPMDAの本業務で開発したAIが、事例のスクリーニングに試行的に導入されることが了承された。

厚生労働科学研究費補助金

政策科学総合研究事業（臨床研究等ICT基盤構築・人工知能実装研究事業）

薬局ヒヤリハット事例に対する安全管理対策評価に関するAI開発

令和2年度～4年度 総合研究報告書

研究代表者 岡本 里香

令和 5 (2023) 年 5月

目 次

I. 総合研究報告
薬局ヒヤリハット事例に対する安全管理対策評価に関するAI開発----- 1
岡本 里香

総合研究報告書

薬局ヒヤリハット事例に対する安全管理対策評価に関するAI研究

研究代表者 岡本 里香
京都大学大学院医学研究科 人間健康科学系専攻 ビッグデータ医科学系分野
特定准教授

研究要旨

独立行政法人医薬品医療機器総合機構（以下、PMDA）では、公益財団法人日本医療機能評価機構（以下、評価機構）が薬局ヒヤリ・ハット事例収集・分析事業（以下、本事業）に基づき収集・分析・公表した「ヒヤリ・ハット事例」に対し、医薬品の名称・包装等の観点から安全対策を講じる必要があるか、PMDAの調査員によって、評価を5段階に分類し、安全対策の必要な事例を抽出し、検討を行っている。しかしながら、評価機構の薬局ヒヤリ・ハット事例報告数は増加しており、ここ数年では、1年で10数万件が報告されている。この中から対策の必要性を検討しなければならない事例を抽出するだけでも、かなりの労力であるため、本研究では、この分類を人工知能（以下、AI）が行えるようにし、PMDAの人による評価の業務負担を軽減し、一貫性のある評価を行うことを目的とした。

本開発過程で、分類精度の高い評価分類モデルを作成することができたが、例えば、PMDAは評価時点で、製造販売業者による安全対策の実施あるいは予定を確認したうえで、評価結果に反映していることや、PMDAの人による評価において判断の揺れがあること等が挙げられ、AIでは解決できない問題があり、過去に評価されたPMDA評価と評価分類モデルによる評価を完全に一致させることは難しいことが判明した。そこで、1次スクリーニングの位置づけで、評価分類モデル（AI）にルールベースロジックを組み合わせることで、安全対策措置を検討する必要のある事例が低レベルのリスク分類に誤分類することを防止しつつ、2次スクリーニングとして調査すべき事例に絞る（調査不要の事例をソートする）ことにより、PMDAの人による評価の業務負担を軽減することを目的とした成果物を目標とした。その結果、誤分類を0件とした上で、約30～60%の事例についてPMDAでの人による評価が不要と判定された。

開発した「評価分類モデル（AI）とルールベースロジック」をPMDAで実装するため、実装環境を構築するとともに、2023年3月1日開催の令和4年度第2回安全使用対策検討会において、2023年度4月以降のPMDAの本業務において、まずは事例のスクリーニングに試行的に導入することが了承された。

なお、研究期間内にPMDAの評価は5分類（評価1～5）から3分類（類型Ⅰ～Ⅲ）に変更されたが、分類のリスクは対応しており、開発した評価分類モデルへの影響はなかった。

（研究分担者）

中津井 雅彦

山口大学大学院医学系研究科・医学部附属病院AIシステム医学・医療研究教育センター・特命教授
小島 諒介

京都大学大学院医学研究科・人間健康科学系専攻・ビッグデータ医科学分野・講師

A. 研究目的

独立行政法人医薬品医療機器総合機構（以下、PMDA）では、公益財団法人日本医療機能評価機構（以下、評価機構）が薬局ヒヤリ・ハット事例収集・分析事業に基づき収集・分析・公表したヒヤリ・ハット事例に対し、医薬品の名称・包装等の観点から安全対策を講じる必要がないか検討を行っている。例えば、薬剤の名称の視覚的、音韻的な類似に起因したことで薬剤を取り違えた等の事例の場合、PMDAでは製造販売会社に対し、薬剤の取違えを防ぐための注意喚起の必要性等について指導するといった医薬品の物的要因に対する安全管理対策の評価・検討を実施しており、現状、人による目で評価分類し、安全対策が必要な事例を抽出している。評価機構が公表するヒヤリ・ハット事例は年々増加傾向にあり、PMDAにおいてこれらに対する安全対策要否の検討・評価は負担が大きくなっている。（図1）

<図1>薬局ヒヤリ・ハット報告集と評価対象事例数の推移

薬局ヒヤリ・ハット報告数と評価対象事例数の推移

PMDA安全使用対策検討会	2013	2014	2015	2016	2017	2018		2019		2020		2021
						前期	後期	前期	後期	前期	後期	前期
評価機構の報告数	6,497	5,387	4,923	4,865	5,151	3,457	20,684 (*1)	34,373 (*2)	92,741 (*3)	97,707 (*4)	23,575	40,186 (*5)
PMDA評価対象	<ul style="list-style-type: none"> * 事例のうち ✓ 規格・剤形間違い ✓ 薬剤取違え ✓ その他及び疑義照会 						評価対象事例の抽出方法の検討のため評価事例なし				新様式	
	3,670	3,019	2,959	2,893	3,356	2,424		1,255	2,369	2,416	575	2,959
	調査の結果、対策を講じたものの事例数は非常に少なく、ほとんどが情報不足、ヒューマンエラーの事例である											
対策検討事例数	0	6	12	12	20	11	-	57	122	161	33	148
評価対象事例に占める割合	0.2%	0.2%	0.4%	0.4%	0.6%	0.5%	-	4.5%	5.1%	6.7%	5.7%	5.0%

(*1) 増加背景として、H27年10月公表「患者のための薬局ビジョン」、H29年3月公表「『患者のための薬局ビジョン』実現のためのアクションプラン検討委員会報告書」の影響、さらに、平成30年度診療報酬改定における地域支援体制加算の新設が本事業への参加と報告の増加に影響していると考えられる。

(*2) これまで6か月分の報告を対象としてきたが、H30年度後期のPMDA評価をスキップしたため、9か月間分の報告対象（評価機構報告H30年1～9月分）

(*3) 7か月間分の報告対象（評価機構報告2018年10～2019年4月分）

(*4) 8か月間分の報告対象（評価機構報告2020年5～12月分）

(*5) 評価機構報告 2020年3月17日～2020年7月31日

PMDAにおける評価は、評価機構が公表する事例のうち、「規格・剤形間違い」「薬剤取違え」「その他及び疑義照会」として報告された事例を抽出し、各事例の内容を確認・評価し、次の評価1～5に分類している。

評価1：医薬品の安全使用に関して製造販売業者等による対策が必要又は可能と考えられた事例

評価2：製造販売業者等により既に対策がとられているもの、もしくは対策を既に検討中の事例

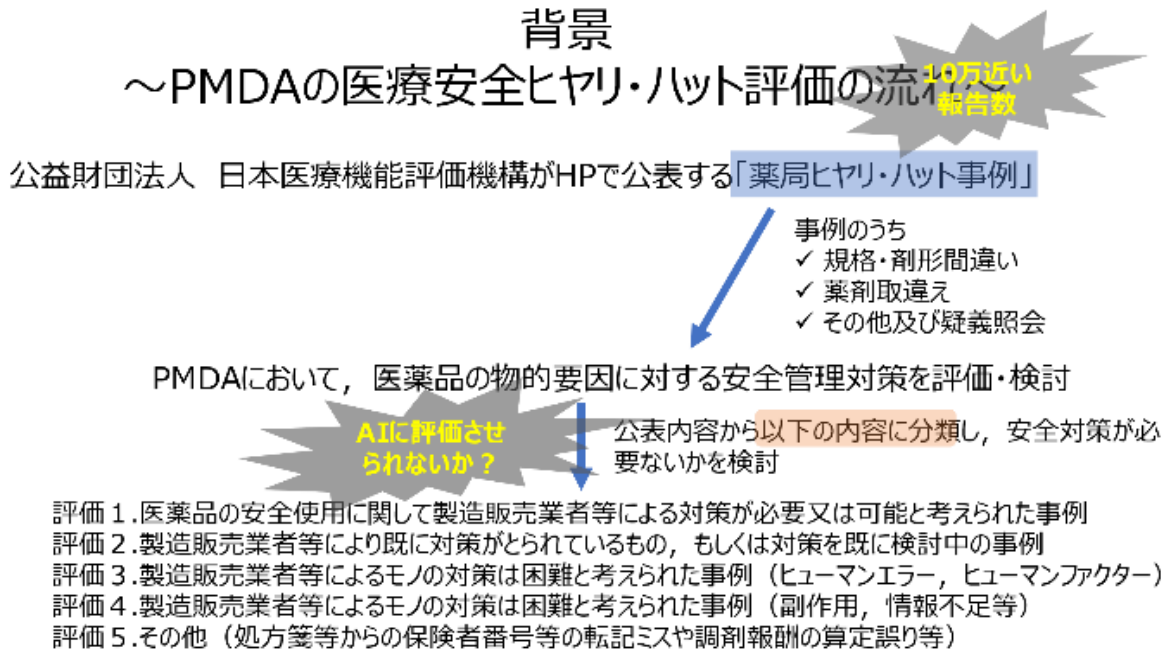
評価3：製造販売業者等によるモノの対策は困難と考えられた事例（ヒューマンエラー、ヒューマンファクター）

評価4：製造販売業者等によるモノの対策は困難と考えられた事例（副作用、情報不足等）

評価5：その他（処方箋等からの保険者番号等の転記ミスや調剤報酬の算定誤り等）

本研究では、PMDAにおける評価1～5の分類をAIに実施させる（図2）ことにより、PMDAにおける本業務の負担を軽減し、ヒヤリ・ハット事例の増加に対しても一貫性のある評価を行うことを目的とする。

<図 2>本研究の概念



B. 研究方法

初年度（2020年度）は、評価機構の旧様式による報告事例を用いて、評価分類モデルを検討した。

ベースとなる初期の評価分類モデルとして、Light GBM (light gradient boosting machine) を適用し、特徴量：表層類似度*、テキストのベクトル化：word 2 vec、サンプリング方法：Random Under Sampling とし、特徴量の追加による評価指標 (precision, recall, f1-score, macro-f1) の向上させることで、精度の高い評価分類モデルを決定することとした。

さらに、分類評価スキームとして、1回で4分類をする no-step 法に対して、評価 1-2 が評価 3 に誤分類されることを低減させるために、まずは、評価 1-3、評価 4、評価 5 の3分類に分類し、その後、評価 1-3 を評価 1-2、評価 3 に分類することにより、最終的に4分類とする 2-step 法についても検討した。

また、テキストのベクトル化では Universal Stentence Encoder (USE) を検討し、サンプリング方法**は Random Over Sampling も検討し、最適な評価分類モデルの条件を決定した。

表層類似度*

表層類似度

- 先頭3文字の一致の有無
- 先頭5文字の一致の有無
- 完全一致の有無
- 薬剤名称の平均長
- 最長一致文字数
- 最長一致文字数 / 名称の平均長
- 編集距離※1 / 最大長
- ゲシュタルトパターンマッチング※2

※1: 文字列Aを何回編集(追加, 削除, 変更)すると文字列Bになるかによって計算される
 ※2: 文字列同士の連続する共通部分を抜き出し、抜き出した文字列の前後に対しても同じ処理を繰り返すことで計算される

例えば「アテック」と「アレック」に対して計算される特徴量

	数値
先頭3文字の一致の有無	0
先頭5文字の一致の有無	0
完全一致の有無	0
薬剤名称の平均長	6
最長一致文字数 / 名称の平均長	0.5
文字種的一致数 / 最大長	0.714
編集距離 / 最大長	0.333
ゲシュタルトパターンマッチング	0.833

サンプリング方法*

Random Under Sampling	Random Over Sampling
<ul style="list-style-type: none"> • データが少ないラベルにデータ数を合わせる手法 →ここでは、一番少ない評価1及び2のデータ数と同じになるように、他の評価データをランダムに使わないこと 	<ul style="list-style-type: none"> • データが多いラベルにデータ数を合わせる手法 →ここでは、一番少ない評価1及び2のデータをたくさんコピーし、あたかもデータ数が他と同じくらいあるようにすること

2020年度の評価分類モデルの精度向上に関する検討の結果を以下にまとめる。

- (1) 特徴量：KEGG (Kyoto Encyclopedia of Genes and Genomes) を用いた複数名称と薬効の特徴量化により精度向上が認められた。劇薬以外の重篤な健康被害に関する情報を特徴量に含める必要性が示唆された。
- (2) 分類評価スキーム：2step 法において、両段階にテキスト情報を特徴量と含めた場合に精度向上に最も寄与し、no-step 法よりも有効であった。
- (3) サンプリング方法：Random Under Sampling より Random Over Sampling が有効と考えられた。
- (4) 自然言語処理技術：テキスト情報に対して word2vec よりも Universal Sentence Encoder が有用であった。

2021年度は、評価機構の旧様式から新様式への変更に伴い、新様式による報告を入手できたので、これまでに作成した評価分類モデルが新様式データに対しても同様に分類できるかを検討した。

- ① 新様式データの PMDA 評価結果の分類パターンを検出した。
→評価機構が報告した 2020 年 3 月 17 日から 2020 年 7 月 31 日までの新様式事例の内訳が、様式が変更したことに伴う、評価機構の報告事例からの PMDA による抽出の結果として、PMDA の評価 1 から 5 の事例配分に旧様式に比して不均衡がないかを確認する。
- ② 上記①より、ラベル 4 及び 5 の事例が極端に少ないことから、no-step 法でも高い分類精度の可能性が考えられたため、no-step 法と 2step 法の比較検討 (閾値 0.5) を実施した。評価指標は precision, recall, f1-score 及び AUC precision recall curve を用いた。

さらに、精度向上等に向けた検討として、以下の (1), (2) の学習データについて追加検討を、新様式データに対して no-step 法 (4 値分類として出力し最も良い指標値を採用), Random Over Sampling, Universal Sentence Encoder の使用条件 (図 3) で、学習データ KEGG と T 辞書の比較及び特徴量としてハイリスク薬を含めた場合を検討した。評価指標は precision, recall, f1-score 及び AUC precision recall curve を用いた。

- (1) T 辞書：商用利用の観点で、KEGG から有限会社ティ辞書企画が販売する医薬品情報データに関する辞書 (以下「T 辞書」) の変更を検討する。収録語に、医薬品名 (一般名, 商品名, 略語等), 薬効名, 疾患名 (病名・症状, 副作用用語等), 独自で整理された同義語等を含み、同義語については KEGG にない情報。
- (2) ハイリスク薬情報：劇薬以外の重篤な健康被害に関する情報として、PMDA では、厚生労働省保険局による「特定薬剤管理指導加算等の算定対象となる薬剤一覧」を活用しているため、当該薬剤一覧にある薬剤を特徴量に追加検討する。

<表 1>新様式データに対する特徴量等の条件

特徴量	販売名のみ類似性	編集距離など複数の手法を利用
	一般的名称まで含めた類似性	T 辞書を利用して特徴量化
	T 辞書の同義語を利用した類似性	T 辞書を利用して特徴量化
	テキスト情報 (Universal Sentence Encoder 使用)	事例の詳細 背景要因 改善策 発生要因 を encode して連結
	薬効情報	T 辞書を利用して特徴量化
	ハイリスク薬情報	「特定薬剤管理指導加算等の算定対象となる薬剤一覧」を利用して特徴量化
分類評価スキーム	no-step 法	
サンプリング方法	Random Over Sampling	

最終年度の 2022 年度は、2021 年度に検討した評価分類モデルに対して、新たに評価機構からの報告事例を用いて、さらにモデルについて記述統計を行い、アルゴリズムを確立した。(評価分類モデルの確定)

また、これまでの検討において得られた課題として、PMDA は評価時点で、製造販売業者による安全対策の実施あるいは予定を確認したうえで、評価結果に反映していることや、PMDA の人による評価において判断の揺れがあること等が挙げられ、PMDA 評価と評価分類モデルによる評価を完全に一致させることは難しい。そこで、評価分類モデルを PMDA 評価過程の 1 次スクリーニングとして使用する場合の許容できる分類精度を事例分析から検討し、事例に対するルールベースを設計した。(運用ルールの決定)

PMDA の業務フローの中での評価分類モデルを組み込んだプロセスイメージを PMDA 実務担当者と検討し、プロセスイメージの決定後、システム開発を行い、開発したシステムを費用、メンテナンス性、運用の継続性等の観点で検証を行った。

PMDA での運用の要否については、令和4年度第2回安全使用対策検討会において、評価委員の先生方に対して、本研究で開発した AI システムについて説明を行い、2023 年度 4 月からヒヤリハット評価支援として事例のスクリーニングに試行的に導入することとなった。

(倫理面への配慮)

本研究で使用するデータは個人情報を含まず、公表済のものを使用しているため、倫理面での配慮は特にない。

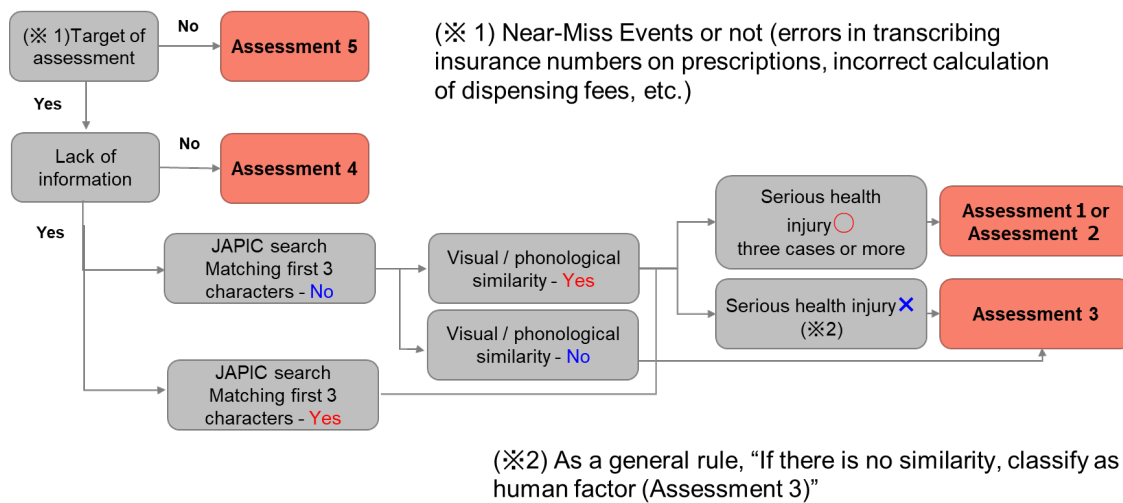
C. 研究結果

【開発する評価分類モデルの分類パターンと精度】

本研究開始時点では、PMDA が当時行っていた評価 1~5 の 5 分類を行う AI 開発を目指すこととしたが、本研究の先行研究として実施した探索的研究において、評価機構の報告数に対して、PMDA が安全対策を検討した事例（評価 1 及び評価 2 が該当）は非常に少ないため、過去の対策検討事例は学習データとしては数が不十分であり、評価 1 か評価 2 かを分類することはできないという課題が挙げられた。そこで、開発するモデルの分類では「評価 1」「評価 2」を区別せず、「評価 1-2」として分類し、他、「評価 3」「評価 4」「評価 5」の計 4 分類とした。また、評価 1 及び評価 2 は安全管理対策が必要な事例であるため、モデルによる分類が誤ってこれらを評価 3、4 あるいは 5 の低リスクに分類してしまうと、安全対策が必要な事例を見逃すことになる。そこで、PMDA の評価 3~5 の事例は評価分類モデルで「評価 1-2」に分類されることを許容することとした。

以上を踏まえ、図 3 に示す評価スキームに準じた評価分類モデルを作成した。

<図 3>評価分類モデルのスキーム



【評価分類モデルの分類精度の向上】

- サンプルング方法：Random Over Sampling と Random Under Sampling の比較
 特徴量として、①表層類似度*、②一般的名称（KEGG を使用し、一般名と記載）、③テキスト情報のベクトル化（word 2 vec 使用し、テキスト情報と記載）、④薬剤情報（KEGG 使用し、KEGG 情報と記載）、⑤劇薬情報（KEGG 使用）を追加し、2つのサンプルング方法を比較した。

Random Over Sampling の場合と Random Under Sampling の場合とで検出したところ、図 4 で示すように、全体として、Random Over Sampling の方が Random Under Sampling よりも精度がよい結果が得られた。また、Random Over Sampling の場合に、評価 1 及び 2 の分類について、precision は学習データセットを追加することにより 0.16→0.27→0.66→0.71 と高まる結果が得られた。しかし、同じく評価 1 及び 2 の分類における recall については、「テキスト情報」を学習させると、「表層類似度+一般名」→「表層類似度+一般名+テキスト情報」は 0.86→0.17 となり、低くなる結果となった。一方、評価 4 については、「テキス

ト情報」を追加することにより， recall は「表層類似度+一般名」→「表層類似度+一般名+テキスト情報」は 0.20→0.91 と高まった。

<図 4> Random Over Sampling と Random Under Sampling の比較（劇薬なし）

組み合わせ	評価	Over sampling			Under sampling		
		precision	recall	F1-score	precision	recall	F1-score
表層類似度	1 及び 2	0.16	0.83	0.27	0.04	0.85	0.07
	3	1.00	0.78	0.80	1.00	0.66	0.80
	4	0.76	0.20	0.29	0.62	0.19	0.29
	5	0.23	1.00	0.37	0.23	1.00	0.37
表層類似度+一般名	1 及び 2	0.27	0.86	0.41	0.06	0.92	0.11
	3	1.00	0.80	0.89	1.00	0.70	0.82
	4	0.77	0.20	0.32	0.61	0.19	0.28
	5	0.23	1.00	0.37	0.23	1.00	0.37
表層類似度+一般名+テキスト情報	1 及び 2	0.66	0.17	0.27	0.03	0.91	0.06
	3	0.98	0.97	0.98	0.99	0.70	0.82
	4	0.78	0.91	0.84	0.58	0.85	0.69
	5	0.89	0.88	0.89	0.52	0.88	0.65
表層類似度+一般名+テキスト情報+KEGG情報	1 及び 2	0.71	0.19	0.30	0.03	0.89	0.06
	3	0.99	0.98	0.98	0.99	0.72	0.83
	4	0.86	0.93	0.89	0.59	0.85	0.70
	5	0.91	0.88	0.89	0.53	0.88	0.65

*ここでの一般名は，薬剤知識ベース（KEGG）を用いた別名称のことであり，製品名だけでなく，一般的名称に対しても対応し，当該対応により，後発医薬品の名称にも対応可能とする。

*テキスト情報は，背景情報である。

さらに，評価 1 及び 2 の分類精度を高めるために，規制区分情報として「劇薬」か否かの情報を学習データに追加した（図 5）．その結果，図 6 との結果と比較すると，Random Under Sampling では「劇薬情報」の追加による精度の変化は認められなかったが，Random Over Sampling での評価 1 及び 2 の precision は，「表層類似度」→「表層類似度+劇薬情報」が，0.16→0.22，「表層類似度+一般名」→「表層類似度+一般名+劇薬情報」0.27→0.32，「表層類似度+一般名+テキスト情報」→「表層類似度+一般名+テキスト情報+劇薬情報」0.66→0.69，「表層類似度+一般名+テキスト情報+KEGG 情報」→「表層類似度+一般名+テキスト情報+KEGG 情報+劇薬情報」0.71→0.74 と高まる結果が得られた。

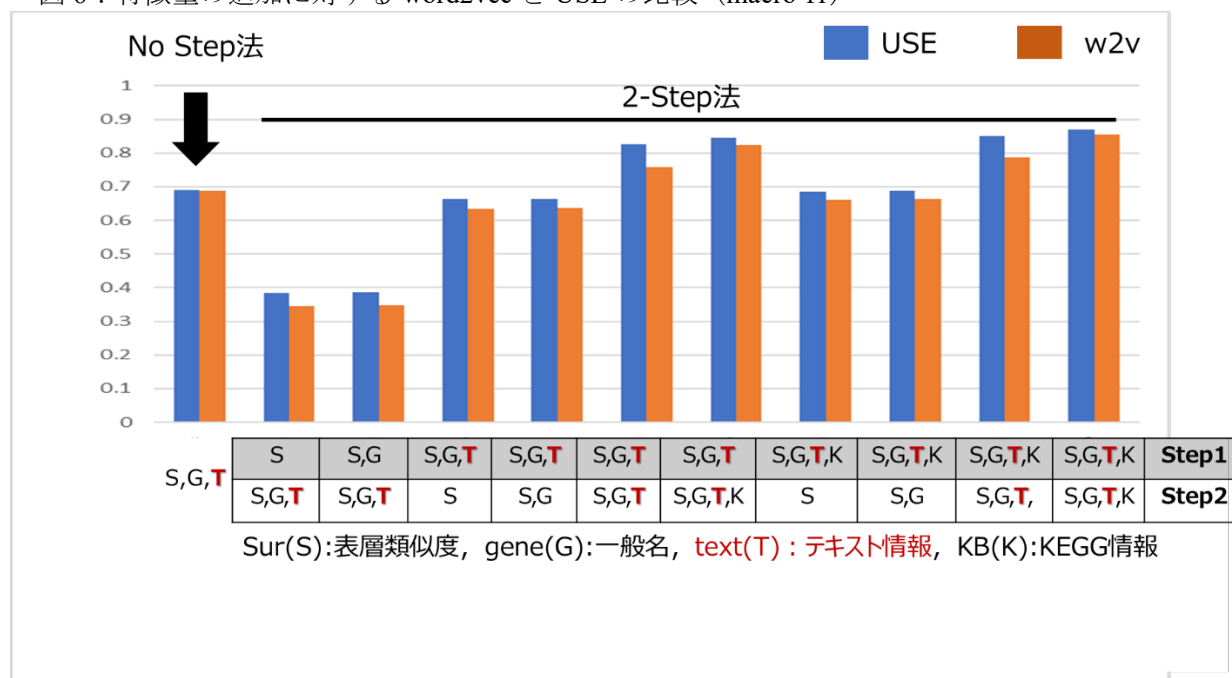
<図 5> Random Over Sampling と Random Under Sampling の比較（劇薬あり）

組み合わせ	評価	Over sampling			Under sampling		
		precision	recall	F1-score	precision	recall	F1-score
表層類似度+劇薬情報	1及び2	0.22	0.83	0.34	0.04	0.87	0.08
	3	0.99	0.83	0.90	0.99	0.69	0.82
	4	0.76	0.20	0.32	0.65	0.19	0.29
	5	0.25	0.96	0.40	0.24	0.98	0.39
表層類似度+一般名+劇薬情報	1及び2	0.32	0.85	0.46	0.07	0.92	0.13
	3	0.99	0.84	0.91	1.00	0.74	0.85
	4	0.77	0.20	0.31	0.64	0.19	0.29
	5	0.25	0.96	0.40	0.24	0.98	0.39
表層類似度+一般名+テキスト情報+劇薬情報	1及び2	0.69	0.15	0.25	0.03	0.91	0.06
	3	0.98	0.97	0.98	0.99	0.70	0.82
	4	0.79	0.92	0.85	0.58	0.85	0.69
	5	0.89	0.89	0.89	0.52	0.88	0.65
表層類似度+一般名+テキスト情報+KEGG情報+劇薬情報	1及び2	0.74	0.19	0.30	0.03	0.91	0.06
	3	0.99	0.98	0.98	0.99	0.71	0.83
	4	0.86	0.93	0.89	0.59	0.85	0.70
	5	0.91	0.88	0.89	0.52	0.87	0.65

● word2vec と USE の比較及び no-step 法と 2-step 法の比較

macro-f1 の比較として、図 6 はテキスト情報ありの場合のみを示すが、テキスト情報が含まれる場合に、2-step 法で差が認められた。no-step 法では、USE と w2v で差異はないが、2-step 法では、Step1 あるいは Step2 に「テキスト情報」を追加した場合に、USE が word2vec よりも精度が高い傾向が見られた。これは、USE の特性が「テキスト情報」に対して有効であると考えられた。

図 6：特徴量の追加に対する word2vec と USE の比較 (macro-f1)



初年度（2020年度）の評価分類モデルの精度向上に関する検討の結果を以下にまとめる。

- 特徴量：KEGG を用いた一般的名称を対象とすることによる複数名称と薬効情報の特徴量化により精度向上が認められた
- 分類評価スキーム：2step 法において、テキスト情報を特徴量と含めた場合に精度向上に最も寄与し、no-step 法よりも有効であった。
- サンプリング方法：Random Under Sampling より Random Over Sampling が有効と考えられた。
- 自然言語処理技術：テキスト情報に対して word2vec よりも Universal Sentence Encoder が有用であった。

2021年度は、これまでに作成した評価分類モデルが新様式データに対して再現できるかを検討した。

- ① 新様式データの PMDA 評価結果の分類パターンを検出した。
→評価機構が報告した 2020 年 3 月 17 日から 2020 年 7 月 31 日までの新様式事例の内訳が、様式が変更したことに伴う、評価機構の報告事例からの PMDA による抽出の結果として、PMDA の評価 1 から 5 の事例配分に旧様式に比して不均衡がないかを確認する。
- ② 上記①より、ラベル 4 及び 5 の事例が極端に少ないことから、no-step 法でも高い分類精度の可能性が考えられたため、no-step 法と 2step 法の比較検討（閾値 0.5）を実施した。評価指標は precision, recall, f1-score 及び AUC precision recall curve を用いた。

さらに、精度向上等に向けた検討として、以下の (1), (2) の学習データについて追加検討を、新様式データに対して no-step 法（4 値分類として出力し最も良い指標値を採用）、Random Over Sampling, Universal Sentence Encoder の使用条件（図 7）で、学習データ KEGG と T 辞書の比較及び特徴量としてハイリスク薬を含めた場合を検討した。評価指標は precision, recall, f1-score 及び AUC precision recall curve を用いた。

実施した。

- (1) T 辞書：商用利用の観点で、KEGG から有限会社ティ辞書企画が販売する医薬品情報データに関する辞書（以下「T 辞書」）の変更を検討する。収録語に、医薬品名（一般名、商品名、略語等）、薬効名、疾患名（病名・症状、副作用用語等）、独自で整理された同義語等を含み、同義語については KEGG にない情報。
- (2) ハイリスク薬情報：劇薬以外の重篤な健康被害に関する情報として、PMDA では、厚生労働省保険局による「特定薬剤管理指導加算等の算定対象となる薬剤一覧」を活用しているため、当該薬剤一覧にある薬剤を特徴量に追加検討する。

<図 7>新様式データに対する特徴量等の条件

特徴量	2020年度の検討	2021年度
販売名のみ類似性	編集距離など複数の手法を利用	編集距離など複数の手法を利用
一般名まで含めた類似性	KEGGを利用	T辞書を利用して特徴量化
T辞書の同義語を利用した類似性	-	T辞書を利用して特徴量化
テキストの埋め込み表現 (Universal Sentence Encoder)	事例の詳細 背景要因 改善策 発生要因 をencodeして連結	事例の詳細 背景要因 改善策 発生要因 をencodeして連結
薬効情報	KEGGを利用	T辞書を利用して特徴量化
ハイリスク薬情報	-	「特定薬剤管理指導加算等の算定対象となる薬剤一覧」を利用して特徴量化

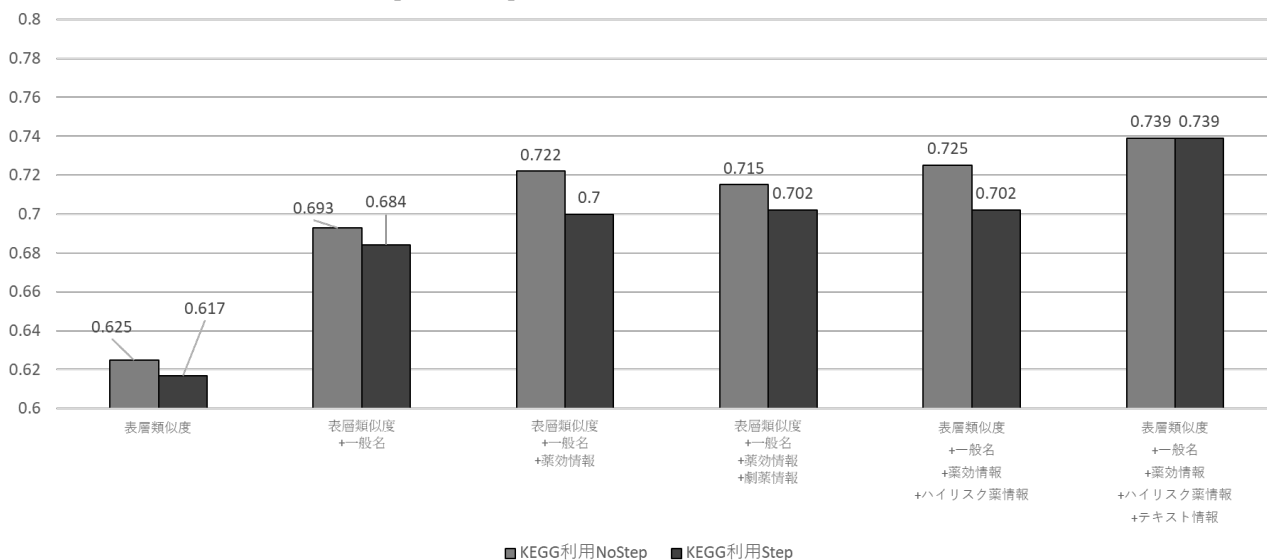
新様式データに対して、no-step 法と 2step 法の比較検討（閾値 0.5）を実施した結果を、ハイリスク薬情報の特徴量追加検討も含めて示す（図 8）。

2 step 法による recall, f1-score の精度向上がラベル 3 に認められた。しかし、この結果は、あくまで閾値 0.5 での結果であり、AUC の結果（図 9）では、no-step 法の場合に精度向上が認められ、全体の性能としては no-step 法で表層類似度＋一般名＋ハイリスク薬＋薬効情報＋テキスト情報の特徴量の組合せの場合に AUC は最も高く、実用化に向けては、実務上許容できる recall を参考に no-step 法か 2step 法かを選択する必要があると考える。

<図 8>新様式データに対する no-step 法と 2step 法の比較 (各評価指標)

feature	label	No-Step			2-Step		
		precision	recall	f1score	precision	recall	f1score
表層類似度	2	0.58	0.64	0.61	0.58	0.65	0.61
	3	0.98	0.80	0.88	0.97	0.97	0.97
	4	0.00	0.00	0.00	0.00	0.00	0.00
	5	0.02	0.77	0.04	0.00	0.00	0.00
			0.97	0.64	0.77	0.97	0.64
表層類似度+一般名	2	0.98	0.82	0.89	0.97	0.99	0.98
	3	0.00	0.00	0.00	0.00	0.00	0.00
	4	0.02	0.77	0.04	0.00	0.00	0.00
	5	0.99	0.66	0.79	0.99	0.66	0.79
			0.98	0.82	0.89	0.97	0.99
表層類似度+一般名+薬効情報	2	0.98	0.82	0.89	0.97	0.99	0.98
	3	0.00	0.00	0.00	0.00	0.00	0.00
	4	0.02	0.77	0.04	0.00	0.00	0.00
	5	0.99	0.66	0.79	0.99	0.66	0.79
			0.98	0.82	0.89	0.97	0.99
表層類似度+一般名+薬効情報+劇薬情報	2	0.98	0.82	0.89	0.97	0.99	0.98
	3	0.00	0.00	0.00	0.00	0.00	0.00
	4	0.02	0.77	0.04	0.00	0.00	0.00
	5	0.99	0.66	0.79	0.99	0.66	0.79
			0.98	0.82	0.89	0.97	0.99
表層類似度+一般名+薬効情報+ハイリスク薬情報	2	0.95	0.66	0.78	0.97	0.62	0.76
	3	0.98	0.99	0.98	0.97	0.99	0.98
	4	0.00	0.00	0.00	0.00	0.00	0.00
	5	0.11	0.31	0.16	0.00	0.00	0.00
			0.98	0.99	0.98	0.97	0.99
表層類似度+一般名+薬効情報+ハイリスク薬情報+テキスト情報	2	0.95	0.66	0.78	0.97	0.62	0.76
	3	0.98	0.99	0.98	0.97	0.99	0.98
	4	0.00	0.00	0.00	0.00	0.00	0.00
	5	0.11	0.31	0.16	0.00	0.00	0.00
			0.98	0.99	0.98	0.97	0.99

<図 9>新様式データに対する no-step 法と 2step 法の比較 (AUC PrecisionRecallCurve)

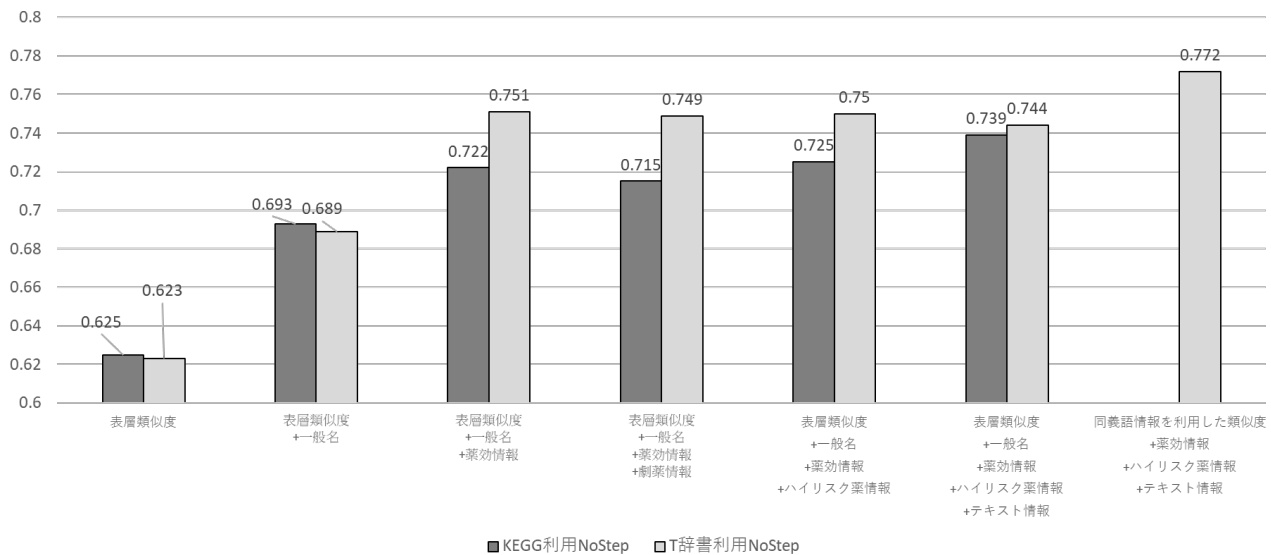


次に、上記の結果で最適と考えられた分類スキームの no-step 法、そして、これまでに最適であることを検討したサンプリング方法として Random Over Sampling 法を組合せ条件として、KEGG と T 辞書の比較を上述の図 3 の内容で行った。その結果、precision, recall, f1-score では既存手法と T 辞書利用で差はほとんどないことから、T 辞書利用で分類精度の低下はないことが示され (図 10)、AUC では、T 辞書利用は既存手法よりも高く、T 辞書に含まれる同義語を特徴量として含めた場合に、最も高い AUC が得られた (図 7)。

<図 10>新様式に対する KEGG (既存手法) と T 辞書との比較 (各評価指標)

feature	label	既存手法			T辞書利用		
		precision	recall	f1score	precision	recall	f1score
表層類似度	2	0.58	0.64	0.61	0.58	0.64	0.61
	3	0.98	0.80	0.88	0.98	0.80	0.88
	4	0.00	0.00	0.00	0.00	0.00	0.00
	5	0.02	0.77	0.04	0.02	0.77	0.04
表層類似度+一般名	2	0.97	0.64	0.77	0.92	0.66	0.76
	3	0.98	0.82	0.89	0.98	0.82	0.89
	4	0.00	0.00	0.00	0.00	0.00	0.00
	5	0.02	0.77	0.04	0.02	0.77	0.04
表層類似度+一般名+薬効情報	2	0.99	0.66	0.79	0.95	0.68	0.80
	3	0.98	0.82	0.89	0.98	0.82	0.89
	4	0.00	0.00	0.00	0.01	0.60	0.02
	5	0.02	0.77	0.04	0.00	0.00	0.00
表層類似度+一般名+薬効情報+劇薬情報	2	0.98	0.66	0.79	0.97	0.68	0.80
	3	0.98	0.82	0.89	0.98	0.82	0.89
	4	0.00	0.00	0.00	0.01	0.60	0.02
	5	0.02	0.77	0.04	0.00	0.00	0.00
表層類似度+一般名+薬効情報+ハイリスク薬情報	2	0.99	0.66	0.79	0.95	0.68	0.80
	3	0.98	0.82	0.89	0.98	0.82	0.89
	4	0.00	0.00	0.00	0.01	0.60	0.02
	5	0.02	0.77	0.04	0.00	0.00	0.00
表層類似度+一般名+薬効情報+ハイリスク薬情報+テキスト情報	2	0.95	0.66	0.78	0.93	0.64	0.76
	3	0.98	0.99	0.98	0.97	0.99	0.98
	4	0.00	0.00	0.00	0.00	0.00	0.00
	5	0.11	0.31	0.16	0.16	0.23	0.19
T辞書同義語+薬効情報+ハイリスク薬情報+テキスト情報	2				0.94	0.68	0.79
	3				0.98	0.99	0.98
	4				0.00	0.00	0.00
	5				0.15	0.23	0.18


<図 11>新様式に対する KEGG と T 辞書との比較 (AUC PrecisionRecallCurve)



また、KEGG と T 辞書でそれぞれ最も分類精度が良い特徴量組合せの分類事例数を比較したところ、KEGG と、T 辞書利用で正しく分類されたラベル 1,2 の事例数は、KEGG では 95 事例に対し、T 辞書では 99 事例であり、T 辞書を利用することによりラベル 1,2 への分類精度が高まった (図 12)。KEGG でラベル 1,2 と分類されるべき事例はラベル 3 に誤分類されていることから、この 4 事例を精査したところ、KEGG には含まれていない情報として、T 辞書に含まれている同義語が、名称変更前後の販売名に関する取違え事例の分類に寄与していることが分かった。

<図 12> 最も分類精度が良い特徴量組合せの KEGG と T 辞書での分類事例数

		評価分類モデル			
		1-2	3	4	5
PMDA	1-2	97	51	0	0
	3	5	2748	4	31
	4	0	8	0	2
	5	0	7	2	4



		評価分類モデル			
		1-2	3	4	5
PMDA	1-2	101	47	0	0
	3	7	2760	6	15
	4	0	8	0	2
	5	0	8	2	3

<図 13> T 辞書利用で正しく分類されたラベル 2 の事例

薬1	薬2	考察
アレロック	シルニジピン	シルニジピンの旧称がアテレックであり、T辞書に同義語として登録されていた
シルニジピン	オロパタジン	シルニジピンの旧称がアテレックであり、オロパタジンの同義語にアレロックが登録されていた
セルシン	プロマゼパム	プロマゼパムの旧称がセニランであり、T辞書に同義語として登録されていた
イソコナールR	一硝酸イソソルビド	2020年6月にイソコナールRは名称変更で硝酸イソソルビドとなっているが、クローラしたKEGGに未反映だった

しかし、T 辞書をした場合でも、図 12 のとおり、ラベル 1,2 をラベル 3 に誤分類する事例が 49 事例ある。49 事例の内容を表 2 のようにケース分けを行ったところ、ケース 1 では、情報提供文書の発出により、過去にラベル 3 として判断されていたものでも、その後取り違いに関する情報提供文書が発出された場合、ある時点から評価がラベル 3 からラベル 1,2 に変更になったケースであり、AI ではカバーできない要因による誤分類が存在することが分かった。

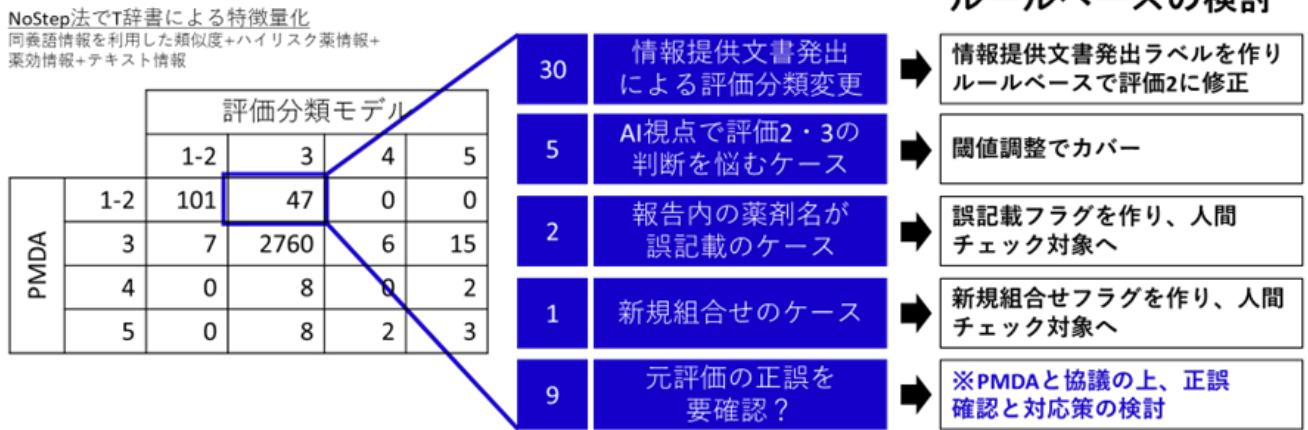
<表 2>

ケース No.	ケース内容	事例数
1	情報提供文書発出により評価分類が変更になっているケース	30
2	AI 視点で評価 2・3 の判断を悩むケース	5
3	報告内の薬剤名が誤記載のケース	2
4	新規薬剤組合せのケース	1
5	過去 PMDA 評価結果について確認が必要なケース	9

2021 年度の成果として、新様式となった評価機構報告に対して PMDA の 2021 年度前期検討会で評価されたデータ 2959 事例（2020 年度前期データ）を評価分類モデルで分類した結果から、誤分類の分析を行い、分析結果から、評価分類モデルの精度に関係しない原因（例えば、過去にラベル 3 として判断されていたものでも、その後取り違いに関する情報提供文書が発出された場合、ある時点から評価 3→評価 2 に変更になる場合がある）で誤分類が生じていることが分かった。誤分類事例を分析したところ、AI の精度以外の問題がある事例であった。そこで、評価分類モデルとしては、表 1 で示す最適な評価分類モデルの条件を確定し、評価分類モデルによる分類に加えて、ルールベースロジックを設定することを検討した。

以下は、2021 年度に実施した 2020 年度前期データを評価分類モデルで分類した結果である。誤分類 47 件を分析し、それぞれに対するルールベース①～④を検討することとした。

<図 14>



<ルールベースの検討>

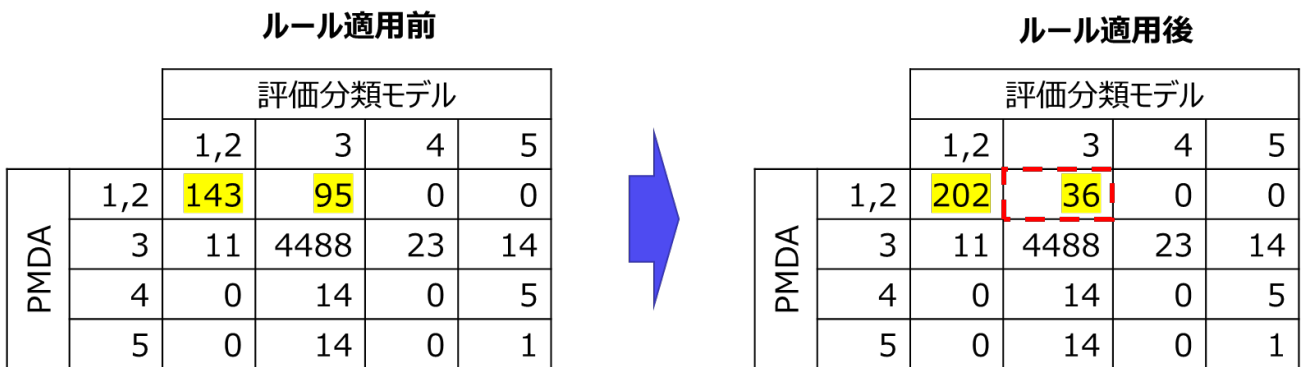
- ルールベース①：情報提供文書発出のラベル作成
- ルールベース②：閾値調整（評価 1,2 の確率が 0.01 以上）
- ルールベース③：存在しない薬剤名のラベル作成
- ルールベース④：新規薬剤組合せのラベル作成

評価分類モデルとルールベースを組み合わせた場合の分類を検討するため、新たに入手した PMDA の 2021 年度後期検討会で評価されたデータ 4808 件（2021 年度後期データ）について、評価分類モデルとルールベース①～④を適用した。

その結果、PMDA 評価 1 あるいは 2 を評価分類モデルでは評価 3 に分類する事例が 95 件あったが、ルールベースを適用すると、36 件に誤分類が減少し、59 件（62%）はルールベースで修正できた。59 件の内訳（件同一データで複数該当するものあり）は、情報提供文書に該当するもの 47 件、誤記載と考えられるもの 6 件、新規薬剤組合せ 1 件、評価 1,2 の確率 0.01 以上 22 件であった。

<図 15>

NoStep法でT辞書による特徴量化
同義語情報を利用した類似度+ハイリスク薬情報+薬効情報+テキスト情報



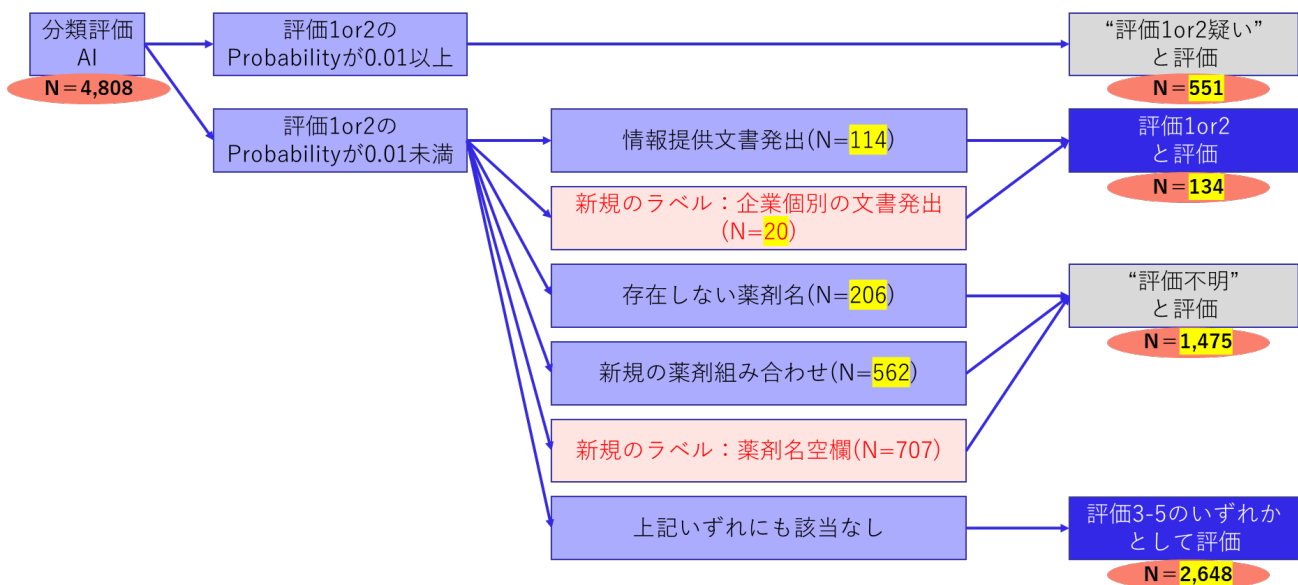
ルール適用後の 36 件について PMDA に詳細確認を実施して頂いたところ、36 件中 25 件は、企業独自で安全対策を講じていた事例であり、11 事例は薬剤名が空欄の事例であった。

<表 3：企業独自で安全対策を講じていた 25 事例>

処方された薬	間違えた医薬品	ラベル2の確率	ラベル3の確率	PMDAコメント	
9	ベポタスチンベシル酸塩錠	ベタヒスチンメシル酸塩錠	0.001755	0.996998	ベポタスチンとベタヒスチンの取り違えに係る注意喚起文書については、企業側が自主的に発出していることが明らかとなったため。
890	バルデスローション	バンデルローション	0.00063	0.998767	バルデスローション及びバルデス軟膏の販売名は、既に企業により一般名に変更済みのため。
891	バルデス軟膏	バンデル軟膏	0.000371	0.998766	同上
1888	アストマリ錠	アストミン錠	0.009832	0.985542	アストマリ錠の販売名は、既に企業により一般名に変更済みのため。
2775	スチプロン軟膏	スチブロン軟膏	0.001661	0.997321	スチブロン軟膏の販売名は、既に企業により一般名に変更済みのため。
3405	バルデス軟膏	バンデル軟膏	0.001793	0.995177	A列No.890と同じ理由
3696	ベギンクリーム	ベギンクリーム	0.000145	0.998835	ベギンクリームの販売名は、既に企業により一般名に変更済みのため。
3697	ベギンクリーム	ベギンクリーム	0.00046	0.997735	同上
3719	ベポタスチンメシル酸塩錠	ベポタスチンベシル酸塩錠	0.000623	0.998291	A列No.9と同じ理由
3720	ベポタスチンメシル酸塩錠	ベポタスチンベシル酸塩錠	0.001373	0.997763	同上
3722	ベポタスチンメシル酸塩錠	ベポタスチンベシル酸塩錠	0.006452	0.991418	同上
3725	ベポタスチンメシル酸塩錠	ベポタスチンベシル酸塩錠	0.008457	0.986695	同上
3726	ベポタスチンメシル酸塩錠	ベポタスチンベシル酸塩錠	0.005847	0.986852	同上
3761	ベポタスチンベシル酸塩OD錠	ベタヒスチンメシル酸塩錠	0.001995	0.995874	同上
3763	ベポタスチンベシル酸塩OD錠	ベタヒスチンメシル酸塩錠	0.000878	0.997363	同上
3766	ベポタスチンベシル酸塩錠	ベタヒスチンメシル酸塩錠	0.002152	0.996499	同上
3768	ベポタスチンベシル酸塩錠	ベタヒスチンメシル酸塩錠	0.004331	0.994417	同上
3769	ベポタスチンベシル酸塩錠	ベタヒスチンメシル酸塩錠	0.005871	0.991438	同上
3770	ベポタスチンベシル酸塩錠	ベタヒスチンメシル酸塩錠	0.00042	0.998581	同上
3771	ベポタスチンベシル酸塩錠	ベタヒスチンメシル酸塩錠	0.001785	0.996938	同上
3773	ベポタスチンベシル酸塩錠	ベタヒスチンメシル酸塩錠	0.004902	0.990022	同上
3774	ベポタスチンベシル酸塩錠	ベタヒスチンメシル酸塩錠	0.006595	0.981055	同上
3775	ベポタスチンベシル酸塩錠	ベタヒスチンメシル酸塩錠	0.001492	0.997645	同上
3776	ベポタスチンベシル酸塩錠	ベタヒスチンメシル酸塩錠	0.001562	0.996503	同上

そこで、上記で作成したルールベース①～④に対するラベルに対して、新規のラベルとして企業独自で安全対策を講じていた事例及び薬剤名空欄のルールベースを追加した。以上のルールベースを組み合わせることにより、2021年度後期データ 4808 件のうち、人の目で再評価が必要な「評価 1, 2 疑い」及び「評価不明」計 2026 件であり、誤分類させることなく（誤分類 0）、分類後の人による評価対象事例は 4808 件から 2026 件と、約 60%減少させることができた。

< 図 16 >

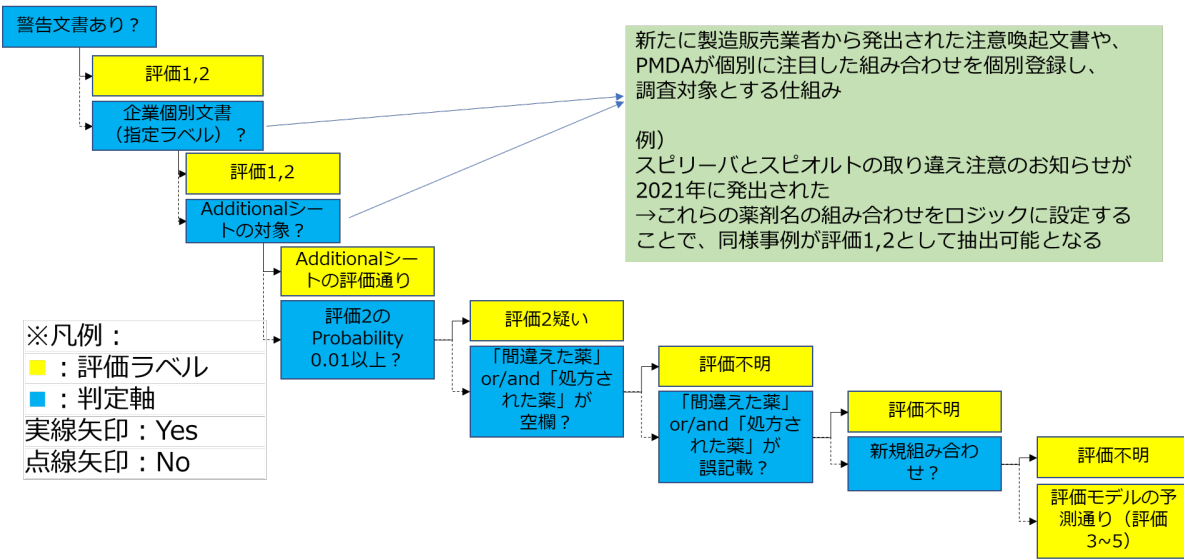


① 運用ルールの決定

以上を踏まえ、これまで検討してきた評価分類モデルの結果とルールベースロジックを組合せて分類した結果を最終判定結果とし、この結果を PMDA で 2 次スクリーニングとして評価する運用を提案する。

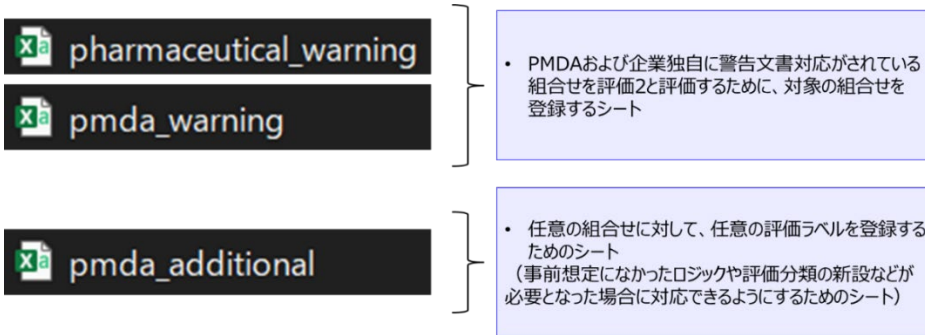
< 図 17 >

AI分類結果+ルールベースロジック→最終判定結果



ルールベースロジックの内容は、以下の通りである。

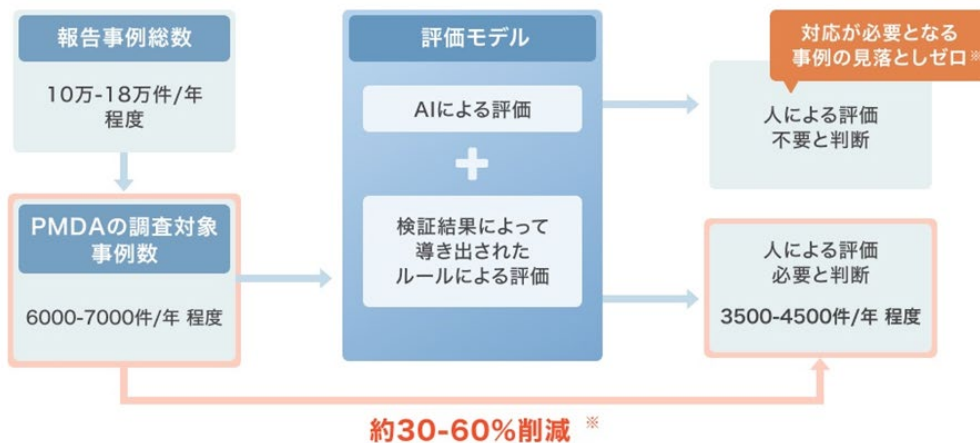
< 図 18 >



2020年3月～2022年6月のPMDA評価対象事例に対しても、評価分類モデルとルールベースロジックを併せた評価モデルを適用した結果、誤分類0件で、かつ、約30～60%の事例についてPMDAでの人による評価が不要と判定された。

< 図 19 >

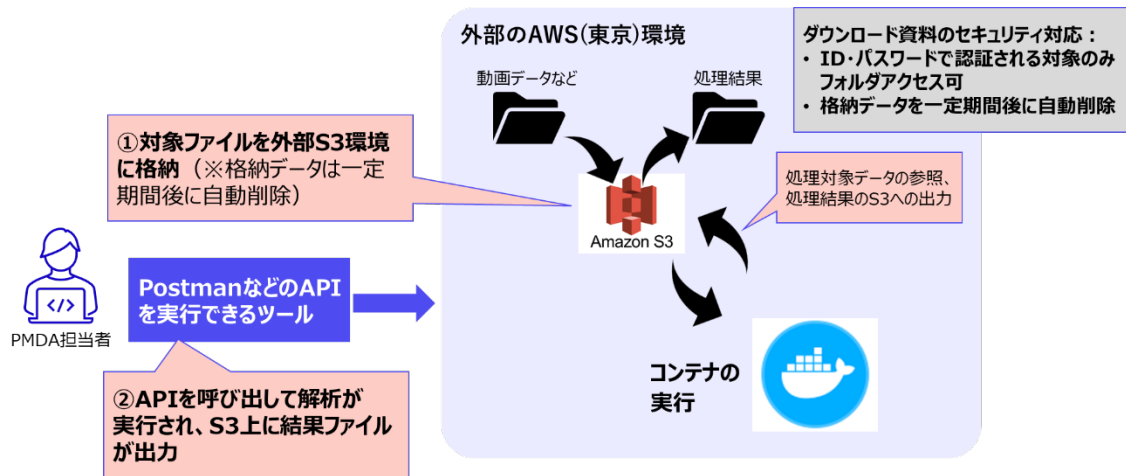
本施策における評価モデルによるPMDA調査件数の削減効果



※検証時において、2020年3月-2022年6月のPMDA評価対象件数に対して、評価モデルが人による評価が不要と判断したものの割合。PMDAでは数か月単位ごとに報告事例を評価しており、評価タイミングごとの事例種類の分布に応じて削減割合が変動。

② 実装環境の構築

<図 20>



外注先のエクサウィザーズ（EXA）のAWS環境上に予測モデルを配置し、APIを通じてPMDAのPC上から解析の実行が可能な実装環境を構築した。

以下のマニュアルを作成し、EXA以外のAWS環境でも実装環境を設置できるようにした。

D. 考察

2022年度までに検討した評価分類モデル構築は新様式データに対しても適用可能であり、これまでテキストを多く含んでいた旧様式に対しては、分類スキームを2-step法にすることで、評価1,2が評価3に誤分類するのを軽減できることを示してきたが、新様式データになったことで、シンプルなno-step法の分類でむしろAUCが向上する結果を示した。no-step法を適用した評価分類モデルで分類し、誤分類について分析したところ、これまでの誤分類分析と同様にやはりAIではカバーできない要因による誤分類であった。実用化の課題として、誤分類をいかに削減して、かぎりなく0件とするか、に対して、誤分類の分析、PMDAによる精査により、ルールベースロジックを確立した。評価分類モデル+ルールベースロジックによる1次スクリーニングとして分類の結果、誤分類を0とすることができ、次に続くスクリーニングにおける、人による評価が必要な事例数を約30~60%削減できる見込みであり、評価分類モデルの開発として実用化レベルとすることができた。また、実装環境の構築については、外注先を中心として検討・構築が実施されたが、PMDAで将来実装される場合に、今般の外注先以外の企業でも、環境維持ができるようマニュアル作成を指導するなどの対応を実施した。

E. 結論

本開発当初は、PMDAの業務負担軽減と人による評価のばらつきをなくすことを目的として、PMDAの人による評価分類に代わり、PMDA評価と100%一致する評価分類モデルとしてのAIを開発する計画で開始した。2020年度~2021年度で評価分類モデルの精度向上を目指し、テキスト情報に対する特徴量化、分類スキーム、サンプリング方法などの検討を実施したが、結果的に、AI以外の要因でPMDA評価と一致することは難しいことが分かった。そこで、評価分類モデルは1次スクリーニングとして実装し、できる限りPMDAの人による評価を行う2次スクリーニングにおける評価対象件数を減らし、PMDAの業務負担軽減に寄与することを目指し、2022年度はこれまで検討した評価分類モデルの条件を確立し、ルールベースロジックを併せることとした。ルールベースロジックは、PMDAにおける蓄積しているフラグ等を踏まえて作成された。結果として、1次スクリーニングの位置づけとして求められる、誤分類0件を評価分類モデルにルールベースロジックを併せることで実現できた。また、PMDAの業務負担軽減に繋がる2次スクリーニング段階での評価対象件数を削減することができた。以上、開発当初の評価分類モデル（AI）は完成しなかったが、実装可能なAIを開発することができた。本開発は、将来的な医療安全に関連するAI開発を計画する際の事例になると考えており、現行の人による評価において、AIが人に置き換わることは難しいが、より迅速に、人による精査が必要な事例に絞り、効率的な医療安全評価にAIが寄与することを本開発で示せたと考える。

F. 研究発表

1. 論文発表

Rika Okamoto, Ryosuke Kojima, Masahiko Nakatsui. Toward AI-supported Evaluation for Safety Control Measures against Near-Miss Events in Pharmaceutical Products. *Safety Science* (2023.1.submitted, 2023.5 revised)

2. 学会発表

- 岡本里香, 中津井雅彦, 小島諒介 薬局ヒヤリ・ハット事例に対する安全管理対策評価に関する AI 開発 第7回日本医療安全学会学術総会 (オンライン) 2021年5月25日
- 岡本里香, 中津井雅彦, 小島諒介 薬局ヒヤリ・ハット事例に対する安全管理対策評価に関する AI 開発 第11回レギュラトリーサイエンス学会学術大会 2021年9月17日

3. その他

2023年3月1日開催の令和4年度第2回安全使用対策検討会において, これまでの開発について説明をし, 2023年度4月以降のPMDAの本業務において, 本研究で開発したAIを事例のスクリーニングに試行的に導入することが了承された.

G. 知的所有権の取得状況

なし