

Title	リアルワールドデータに基づくBERTを用いた新しい医薬品安全性シグナル検出手法の提案
Sub Title	
Author	八重樫, 昇吾(Andō, Takayuki) 石川, 智貴(Hara, Azusa) 安藤, 崇之(Urushihara, Hisashi) 原, 梓 大石, 尚央 漆原, 尚巳
Publisher	慶應義塾大学AI・高度プログラミングコンソーシアム
Publication year	2023
Jtitle	AICカンファレンス予稿集 (2023. ) ,p.6- 12
JaLC DOI	
Abstract	従来の医薬品安全性評価に関わるシグナル検出において、複数の医薬品と複数の有害事象間、つまり多対多の複雑な関連を考慮した安全性情報の評価はなされていない。そこで、本研究ではレセプトデータベースにBERTを適用することで、患者の背景・過去病歴の情報に基づいて各有害事象の分散表現を作成し、その特有な共起パターンをクラスターとして検出することでシグナル検出が可能か検討した。有害事象をクラスタリングした結果を比較し、経口糖尿病薬のSGLT-2阻害薬、DPP-4阻害薬、及びビグアナイド薬に特有な既知の副作用と関連するクラスターを検出し、シグナル検出に活用できる可能性が示唆された。さらなる改良のために今後は他の薬剤に適用し、本技法の特性を明確にすることが重要である。
Notes	会議名：AICカンファレンス2023 開催地：慶應義塾大学日吉キャンパス 日時：2023年3月4日 第1章研究論文 論文-1
Genre	Conference Paper
URL	<a href="https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO11003001-20230304-0006">https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO11003001-20230304-0006</a>

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the Keio Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

# リアルワールドデータに基づくBERTを用いた新しい医薬品安全性シグナル検出手法の提案

八重樫昇吾<sup>1</sup>, 石川智貴<sup>2</sup>, 安藤崇之<sup>3</sup>, 原梓<sup>1</sup>, 大石尚央<sup>2</sup>, 漆原尚巳<sup>1</sup>

<sup>1</sup>慶應義塾大学大学院薬学研究科 医薬品開発規制化学講座

<sup>2</sup>慶應義塾大学大学院理工学研究科総合デザイン工学専攻

<sup>3</sup>慶應義塾大学医学部 総合診療教育センター

## Abstract:

従来の医薬品安全性評価に関わるシグナル検出において、複数の医薬品と複数の有害事象間、つまり多対多の複雑な関連を考慮した安全性情報の評価はなされていない。そこで、本研究ではレセプトデータベースにBERTを適用することで、患者の背景・過去病歴の情報に基づいて各有害事象の分散表現を作成し、その特有な共起パターンをクラスターとして検出することでシグナル検出が可能か検討した。有害事象をクラスタリングした結果を比較し、経口糖尿病薬のSGLT-2阻害薬、DPP-4阻害薬、及びビグアナイド薬に特有な既知の副作用と関連するクラスターを検出し、シグナル検出に活用できる可能性が示唆された。さらなる改良のために今後は他の薬剤に適用し、本技法の特性を明確にすることが重要である。

**Keywords:** Signal detection, Real World Data, BERT, Claims data

## 1. 研究背景・目的

医薬品安全性評価を検討してきたCIOMSワーキンググループVIでは、副作用検出において有害事象間の複雑な関連を考慮して安全性情報を評価することが統計学上の重大な課題としている<sup>1</sup>。これまでの安全性シグナル検出は専ら、薬剤と事象間の関連を1対1で評価し、並行してみられる類縁事象を考慮していない。しかし、薬剤が与える影響は複数器官に及ぶ可能性があるため、複数の要因を考慮の上、1対多、もしくは多対多の関連を考慮し、安全性を評価することが望ましい。さらに、発現数の少ない有害事象を個別に評価する場合、検出力が低くなり、また、他の関連する有害事象から得られる有用な情報が無視されることになる。そこで、Berry & Berryはランダム無作為化試験(RCT)における有害事象と医薬品との関連性の特定に、同一器官内における類縁症状・徴候を考慮した階層ベイズモデルを提唱した<sup>2</sup>。この手法では、同一器官の有害事象は異なる器官の有害事象よりも同時性をもって類似した事象が発生するという仮定を置いている。Crooksらはこの階層ベイズモデルを自発報告データベースのVigiBaseに適用し、有害事象をWHO Adverse Reaction Terminology (WHO-ART)の辞書構造に基づき消化器系、または肝胆膵系でグルーピングすることで有害事象間の類似性を仮定したシグナル検出の方法を提唱した<sup>3</sup>。これらの手法は、医薬品と有害事象の関連の評価に他の関連する有害事象の情報も考慮することで既存手法より精度の高いシグナル検出が可能一方で、有害事象間のグルーピングの定義は人手により予め恣意的に設定されたものであり、全ての有害事象についての類縁症状・徴候の人手によらないグルーピングは困難であった。

市販後の医薬品安全性監視において重要な情報源の1つとされている自発報告データベースは、曝露、分母情報の欠如、及び一般的に過小報告などによるバイアスのため因果関係の判定には大きな限界があることが知られている<sup>4</sup>。そこで、BrownらはHMO Research NetworkのCenter for Education and Research on Therapeutics (CERT)に参加し

ている9つのヘルスプランのレセプトデータ<sup>5</sup>にMaximum Sequential Probability Ratio Testing法(MaxSPART法)を適用することで、医薬品の有害事象に関するリアルタイムのサーベイランスにおけるレセプトデータの有用性評価、及び安全性サーベイランスにレセプトデータを使用する際の主要な問題点を明らかにした。また、Choiらは韓国のHealth Insurance Review & Assessment Service (HIRA)のレセプトデータベース<sup>6</sup>に不均衡分析を適用してシグナル検出を試みた<sup>7</sup>。いずれの研究においても既知の副作用を含むシグナルを検出し、レセプトデータによるリアルタイムでのシグナル検出への活用可能性を示した。しかし、これらの研究では前述のような一つの医薬品に対し、複数の有害事象を並列で考慮する、もしくは複数の要因に対し、複数の有害事象を同時に考慮することはできていない。

近年のコンピュータ技術の発展に伴い利活用が進んでいる機械学習は、大規模データにおける複雑な共起パターンを扱うことに長けているため、医薬品と有害事象との間の1対多、もしくは多対多の関連についての評価に利用できる可能性がある。副作用予測の関連研究<sup>8</sup>では、医薬品の有害事象の予測においてDeep Neural Network(以下、DNN)で既存手法よりも優位な結果を示している。しかし、これらの研究では薬剤の成分や疾患の分類情報などの既存の情報を活用したにすぎない。そこで本研究では、患者固有の背景情報を考慮しつつ医薬品が生体内で引き起こす特有の生理・病理メカニズムに起因する同時並行の反応の結果として観測される複数の関連する潜在的な有害事象の共起パターンを検出する手法を提案する。具体的には、診療報酬請求データベースに機械学習を適用することで、患者の背景・過去病歴の情報を用いて、発現した医薬品の有害事象を分散表現とし、その特有な共起パターンをクラスターとして検出することでシグナル検出が可能か検討した。安全性プロファイルがよく知られている糖尿病治療薬を対象に機械学習モデルを適用し、形成された有害事象のクラスターについて添付文書や医薬品安全性監視計画書の調査、及び医学レ

ビューを行うことでモデルの有用性について記述的に評価した。

## 2. 方法

### 2.1. 研究の構想

薬剤の投与により体内の様々な器官に引き起こされる様々な応答が、結果として薬剤の有効性・安全性として表れる。そのうち副作用は、薬剤投与により引き起こされる意図しない患者にとって有害なものである。薬剤が引き起こす一連の生体内の生理学的な応答が、ある患者特有の容態や要因の影響を受けた結果、別の患者で発現した症状・徴候と同一症状・徴候として表れる場合や、類似する症状・徴候の全部もしくは一部を示す場合、または異なる症状・徴候を示す場合と様々な表現形を取ることで、患者間では同一の、もしくは異なる副作用が認められる。この仮定によれば、副作用として発現する症状・徴候は、生体内におけるある共通の生理的・病理的メカニズムによる応答として、類似したかつ内容の一部異なる症状・徴候のグループが、類似した容態や要因を有する患者で頻繁に認められると考えられる。

本研究では、医薬品が生体内で引き起こす特有の生理・病理メカニズムに起因する同時並行の反応の結果として観測される複数の関連する潜在的な有害事象の共起パターンを検出する手法を提案する。リアルワールドデータであるレセプトデータから既知もしくは未知の有害事象発現リスクを推定するため、医薬品の特性や前提条件によらない機械学習を用いた新たなシグナル検出手法について検討を行うことが目的である (Fig 1 参照)。本技法では、レセプトデータで観測された繰り返し認められる医薬品の有害事象の共起パターンを検出するために、関連すると考えられる有害事象のクラスターをクラスタリング手法に基づき形成し、パターン認識に基づき潜在的な副作用クラスターを特定する。結果を薬効群間で比較することで、薬剤クラスに特有の副作用クラスターを特定する。

薬効群レベル

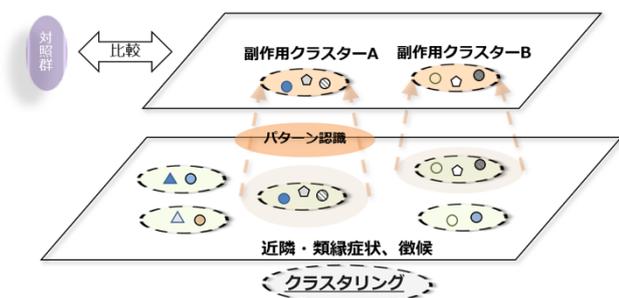


Fig. 1 研究の構想

### 2.2. 研究のデザイン

株式会社JMDC及びDeSCヘルスケア株式会社が保有するレセプトデータベース(計2270万人)を用いた。調査対象患者は、2005年1月から2021年8月の間に調査対象薬の処方特定され、かつ、糖尿病のICD-10コード(E10-E14)を調査対象薬の処方日の1年後以前に持ち、かつ、有害事象の記録がある患者とした。調査対象薬の初回処方日前に1回以

上レセプトがない患者は除外した。有害事象は調査対象薬の初回処方後初めて記録された傷病データとし、リスク期間は処方日1日後から最終処方日までとした。有害事象の定義はTreatment-emergent adverse event (以下、TEAE)を用いた<sup>9</sup>。TEAEは治療下で発現した有害事象であり、本研究では「調査対象薬処方後に存在する傷病データのうち、調査対象薬処方前には存在しなかった傷病データ」と定義した。ただし、TEAEの定義に含まれる併存疾患の増悪はレセプト上で特定することが困難であるため考慮していない。リスク期間は、処方日1日後から一連の処方が終了した日に最後の処方日数を加えた最終処方日までとした。併存疾患は対象薬の初回処方日以前に診断を受けた疾患、もしくは初回処方日後のリスク期間以外で診断された有害事象以外の疾患とした。対象薬以外の糖尿病治療薬による併用・既治療歴は対象薬の最終処方日までに処方された糖尿病治療薬とした。

患者情報に基づき予測するタスクを通じて有害事象の分散表現を獲得し、得られた有害事象のベクトルをWard法でクラスタリングすることで、背景が類似する患者の有害事象、もしくは同一患者で並行して発症した類縁有害事象がクラスターを形成することを意図した (Fig 2 参照)。

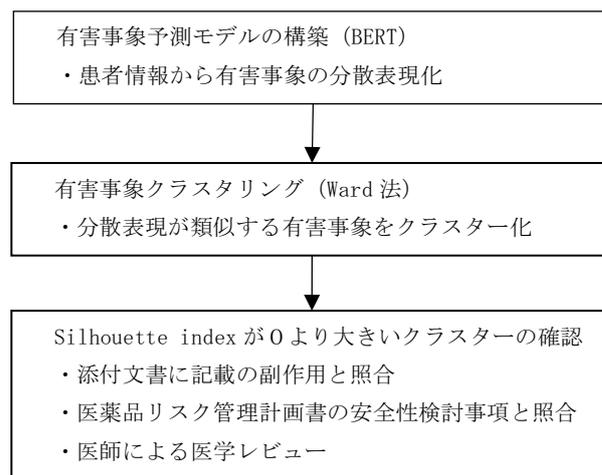


Fig. 2 研究のフローチャート

関連研究で提案されたBidirectional Encoder Representations from Transformers (BERT)を基にしたモデルで、薬剤クラスごとに患者の年齢、性別、病歴、対象薬以外の糖尿病治療薬による併用・既治療歴、有害事象を入力として、有害事象の分散表現を出力としたモデルの構築を行った。ここで、入力した変数の意味関係を表さないone hotベクトルから高次元ベクトル空間への埋め込みを行なった。文章の学習に活用されるNext Sentence Prediction (NSP)は本研究では適用できないため、事前学習はMasked Language Model (MLM)で行った。関連研究<sup>10</sup>で提案されていたように、BERTのInputに各診療時点を区別する情報 (Segment), 診療時点の年齢 (Age), 性別の埋め込み層を導入した。Segmentには各診療単位でAとBを交互に入力し、Positionの順序情報と組み合わせることで別時点の診療の区別と同一診療内の診断順序を順不同として扱うことが可能である。入力した患者情報を基にMLMで有害事象の分散表

現を獲得した。関連研究<sup>11,12</sup>を参考に患者の有害事象の12%をマスクすることとした。変数間での置換は実臨床のデータと異質な学習データとなるため実装しなかった。BERTのパラメータはTable 1に示す通りとした。入力データ数が多すぎるとBERTが構築できないため、入力データは有害事象全てに加えてIndex dateに近い病歴、対象薬剤以外の糖尿病治療薬による併用・既治療歴の合計入力数を変更して学習が進む設定を探索した。また、発症した患者数が極端に少ない有害事象は学習結果に大きな影響を与えるため解析対象患者内で発症数が20人未満の有害事象を除外、または除外しない両方の設定で構築した。その他のBERTで調整するパラメータはTable 1の通りである。BERTの予測精度の評価指標にはPrecision microを用いた。クラスタリングでは教師なし学習であるWardを用い、各クラスターの要素数は平均5個とした。形成されたクラスターをSilhouette Index(SI)とDavies-Bouldin Index(DBI)で評価した。

Table 1 BERTの設定

最適化アルゴリズム	Adam
損失関数	Binary crossentropy
学習係数	0.0001
バッチサイズ	16
隠れ層のDropout率	0.01
隠れ層の数	9
Attention head数	12
Feed-forward層のサイズ	256
Transformerのencoderとpoolerの活性化関数	GeLu
embedding, Transformerのencoder, attention probabilities, poolerの全結合層のドロップアウト率	0.1
学習回数	[100, 150]
解析対象患者内で発症数がx未満の有害事象を除外	[0, 20]
分散表現の次元数	[144, 256]
全有害事象、病歴、対象薬剤以外の糖尿病治療薬による併用・既治療歴の総入力数	[50, 100, 150]

患者背景は記述統計で要約した。機械学習アルゴリズムに適応させるために抽出した有害事象を2値表現に変換した。BERTへ入力するため、病歴、対象薬剤以外の糖尿病治療薬による併用・既治療歴、有害事象を時系列順に並べた。疾患はICD-10コードの医学的な概念を構成する最小限の単位である小分類でコード化した。SIが0以下の場合、クラスターが誤っている可能性があるとして判断される<sup>13</sup>。したがって、SIが0より大きいクラスターについて医学レビューを行い、合併症(糖尿病性網膜症、糖尿病性腎症、糖尿病性神経障害、糖尿病性足病変、及び糖尿病性大血管症)を含むクラスターまたは、感染症、精神疾患、及び悪性腫瘍の糖尿病の関連疾患を含むクラスターに分類した。解析アプリケーションはPython3.7を使用し、OSはUbuntu 18.04 LTS Server, CPUとGPUはそれぞれAMD EPYC 7702P, NVIDIA RTX A6000 2機及びメインメモリ512GBを実装したHPC3000-ERMGPU4TS(HPCシステムズ株式会社)で解析を行った。本研究は慶應義塾大学薬学部人を対象とする研究倫理委員会での審査を受け、承認後に実施した(承認番号:221005-2)。

### 3. 結果

調査対象患者はインスリン製剤で98,667人(平均年齢70.4歳, 男性59%, リスク期間の中央値210日), DPP-4阻害薬で64,183人(66.4歳, 61%, 761日), SGLT-2阻害薬で23,866人(61.7歳, 63%, 670日), ビグアナイド薬で40,167人(61.6歳, 63%, 792日), 及びGLP-1受容体作動薬で20,070人(66歳, 57%, 480日)であった(Fig 3, Table 2参照)。併存疾患と有害事象は各薬剤クラスで上位3個に入った疾患全てを示しており、いずれの薬剤クラスにおいても同じ疾患で糖尿病の合併症として考えられる疾患が上がっていた。糖尿病治療薬の併用・既治療歴はインスリン製剤, SGLT-2阻害薬及びビグアナイド薬でDPP-4阻害薬の処方が高くそれぞれ78.2%, 77.4%, 及び81%であった。DPP-4阻害薬ではビグアナイド薬の処方が52%, GLP-1受容体作動薬ではインスリン製剤の処方が63%と最も高かった。BERTのパラメータ調整の結果、学習回数が150回、解析対象患者内で発症数が20未満の疾患を除外し、分散表現の次元数を240、患者一人あたりの総入力数は100個となった。最大precision microとSIはインスリン製剤で0.33, 0.062, DPP-4阻害薬で0.67, 0.06, SGLT-2阻害薬で0.46, 0.056, ビグアナイド薬で0.43, 0.058, 及びGLP-1受容体作動薬で0.67, 0.056となり、DBIはいずれも2であった。

Ward法によって形成したクラスターに対して医学レビューを行い、分類した結果をTable 3に示す。SIが0より大きいクラスターのうち、全体では添付文書に記載の副作用を含むクラスターが109個(16.2%)形成され、薬剤クラス別ではSGLT-2阻害薬で最も多く形成がみられた(30.6%)。糖尿病の合併症のクラスターはインスリン製剤で最も多く形成がみられた。糖尿病の関連疾患のうち、感染症クラスターはSGLT-2阻害薬以外で10個以上形成されており、精神疾患はSGLT-2阻害薬で最も多く形成がみられた。悪性腫瘍はDPP-4阻害薬とGLP-1受容体作動薬で多く形成がみられ、SIが0以下のクラスターはDPP-4阻害薬で最も多く形成がみられた。

添付文書に記載の副作用を含むクラスターをFig 4に示す。SGLT-2阻害薬を除く薬剤クラスで低血糖クラスターの形成がみられ、全クラスターのうちSI値の上位順でインスリン製剤(Cluster 1), DPP-4阻害薬(Cluster 2), ビグアナイド薬(Cluster 3), GLP-1受容体作動薬(Cluster 4)でそれぞれ8番目, 137番目, 61番目, 90番目であり、他の薬剤クラスと比較してインスリン製剤のクラスターはSI値が高かった。DPP-4阻害薬のCluster 5とGLP-1受容体作動薬のCluster 6, 及びDPP-4阻害薬のCluster 7で類天疱瘡を含むクラスターが形成されていた。DPP-4阻害薬のCluster 7に関節炎のクラスター, Cluster 8に剥脱性皮膚炎を含む皮膚疾患のクラスター, 及びCluster 9に間質性肺炎・喘息を含む肺疾患のクラスターが形成されていた。SGLT-2阻害薬のCluster 10で、性器感染症・頻尿クラスターがみられ、ビグアナイド薬のCluster 11には乳酸アシドーシスを含むクラスターがみられた。

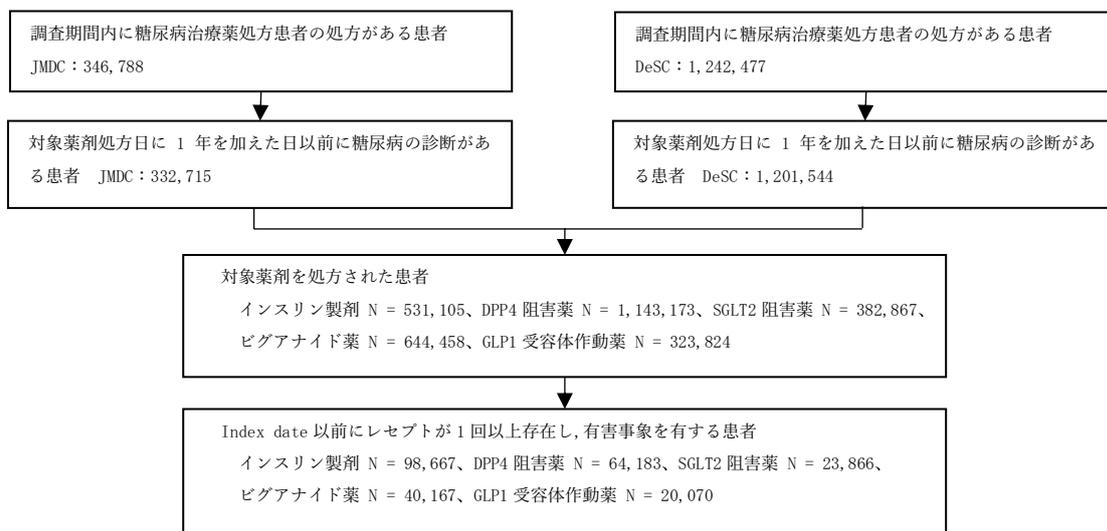


Fig. 3 調査対象患者の内訳

Table 2 調査対象患者の患者背景

	インスリン製剤 N=94,395	DPP-4 阻害薬 N=64,183	SGLT-2 阻害薬 N=23,866	ビグアナイド薬 N=40,167	GLP-1 受容体作動薬 N=20,070
年齢 平均(SD)	70.4 (14.1)	66.4 (14.4)	61.7 (14.6)	61.6 (14.6)	66 (14.4)
男性 (%)	55,709 (59)	39,188 (61)	14,948 (63)	25,370 (63)	11,361 (57)
リスク期間 (日), Median, [IQR]	210 [91, 570]	761 [397, 1219]	670 [366, 1037]	792 [426, 1279]	480 [271, 751]
<b>併存疾患</b>					
I10:本態性(原発性<一性性>)高血圧(症)	63,388 (64.2)	33,778 (52.6)	15,387 (64.5)	21,048 (52.4)	14,143 (70.5)
E78:リポタンパク<蛋白>代謝障害及びその他の脂血症	55,832 (56.6)	32,422 (50.5)	16,546 (69.3)	22,642 (56.4)	14,389 (71.7)
H52:屈折及び調節の障害	42,345 (42.9)	23,860 (37.2)	10,262 (43.0)	15,241 (37.9)	10,715 (53.4)
<b>有害事象</b>					
K59:その他の腸の機能障害	23,021 (23.3)	15,111 (23.5)	3,508 (14.7)	7,605 (18.9)	3,261 (16.2)
K21:胃食道逆流症	22,616 (22.9)	15,379 (24.0)	3,704 (15.5)	8,263 (20.6)	2,997 (14.9)
I50:心不全	20,465 (20.7)	14,918 (23.2)	3,594 (15.1)	7,285 (18.1)	3,011 (15.0)
H52:屈折及び調節の障害	17,504 (17.7)	14,353 (22.4)	3,817 (16)	9,406 (23.4)	3,202 (16.0)
E78:リポタンパク<蛋白>代謝障害及びその他の脂血症	15,774 (16.0)	16,366 (25.5)	3,154 (13.2)	9,463 (23.6)	2,486 (12.4)
L30:その他の皮膚炎	15,395 (15.6)	12,884 (20.1)	3,646 (15.3)	7,240 (18.0)	2,740 (13.7)
<b>糖尿病治療薬による併用・既治療歴</b>					
インスリン製剤	-	21,491 (44.1)	6,876 (31)	11,744 (31.7)	12,498 (63.7)
DPP-4 阻害薬	65,545 (78.2)	-	17,169 (77.4)	29,957 (81.0)	11,926 (60.8)
SGLT-2 阻害薬	16,537 (19.7)	13,459 (27.6)	-	12,316 (33.3)	7,751 (39.5)
ビグアナイド薬	36,800 (43.9)	25,347 (52)	14,430 (65)	-	11,397 (58.1)
GLP-1 受容体作動薬	9,484 (11.3)	3,110 (6.4)	2,176 (9.8)	2,425 (6.6)	-
チアゾリジン薬	11,705 (14)	6,971 (14.3)	4,214 (19)	6,759 (18.3)	3,017 (15.4)
グリニド薬	18,674 (22.3)	8,247 (16.9)	3,217 (14.5)	5,010 (13.5)	4,781 (24.4)
スルホニル尿素薬	31,061 (37)	16,904 (34.7)	8,030 (36.2)	13,645 (36.9)	6,843 (34.9)
αグルコシダーゼ阻害薬	24,268 (28.9)	12,276 (25.2)	4,856 (21.9)	8,486 (22.9)	5,421 (27.6)

Table 3 医学レビューによるクラスター分類(SI>0)

	全体 N=672	インスリン製剤 N=160	DPP-4 阻害薬 N=162	SGLT-2 阻害薬 N=111	ビグアナイド薬 N=132	GLP-1 受容体作動薬 N=107
添付文書に記載の副作用を含有	109 (16.2%)	14 (6.8%)	24 (14.8%)	34 (30.6%)	26 (19.7%)	27 (25.2%)
糖尿病の合併症	58	13	11	11	12	11
<b>糖尿病の関連疾患</b>						
感染症	57	17	15	7	13	5
悪性腫瘍	40	13	10	7	3	7
精神疾患	15	2	2	6	1	4

Cluster1, 低血糖 インスリン製剤, SI=0.16	Cluster2, 低血糖 DPP-4 阻害薬, SI=0.057	Cluster3, 低血糖 ビッグアナイド薬, SI=0.0002	Cluster4, 低血糖 GLP-1 受容体作動薬, SI=0.016
D64; その他の貧血 D65; 播種性血管内凝固症候群 [脱線維素症候群] D68; その他の凝固障害 D69; 紫斑病及びその他の出血性病態 E03; その他の甲状腺機能低下症 E05; 甲状腺中毒症 [甲状腺機能亢進症] E07; その他の甲状腺障害 E16; その他の膵内分泌障害 (添付文書に記載の副作用: 低血糖)	E16; その他の膵内分泌障害 (添付文書に記載の副作用: 低血糖) E88; その他の代謝障害 G51; 顔面神経障害	D33; 脳及び中枢神経系のその他の部位の良性新生物<腫瘍> E16; その他の膵内分泌障害 (添付文書に記載の副作用: 低血糖) E88; その他の代謝障害	C73; 甲状腺の悪性新生物<腫瘍> D32; 膵臓の良性新生物<腫瘍> E16; その他の膵内分泌障害 (添付文書に記載の副作用: 低血糖) E28; 卵巣機能障害
Cluster5, 皮膚疾患(類天疱瘡) DPP-4 阻害薬, SI=0.003	Cluster6, 皮膚疾患(類天疱瘡) GLP-1 受容体作動薬, SI=0.042	Cluster7, 関節炎、横紋筋融解症 DPP-4 阻害薬, SI=0.13	Cluster8, 皮膚炎(剥脱性皮膚炎) DPP-4 阻害薬, SI=0.13
B81; その他の腸ぜんく蠕虫症, 他に分類されないもの L08; 皮膚及び皮下組織のその他の局所感染症 L10; 天疱瘡 L12; 類天疱瘡 (添付文書に記載の副作用: 類天疱瘡) L21; 脂漏性皮膚炎	B59; ニューモシスチス症 L08; 皮膚及び皮下組織のその他の局所感染症 L12; 類天疱瘡 L21; 脂漏性皮膚炎 L50; じんまき蕁麻疹 L51; 多形紅斑 (添付文書に記載の副作用: 紅斑)	M11; その他の結晶性関節障害 M16; 股関節症 [股関節部の関節症] M18; 第1手根中手関節の関節症 M24; その他の明示された関節内障害 M53; その他の脊柱障害, 他に分類されないもの M62; その他の筋障害 (添付文書に記載の副作用: 横紋筋融解症) M72; 線維芽細胞性障害 M77; その他の腱(靭帯)付着部症 R23; その他の皮膚変化	K93; 他に分類される疾患におけるその他の消化器の障害 L01; 膿か<痂>疹 L02; 皮膚膿瘍, せつ<フルンケル>及びよく<カルブンケル> L04; 急性リンパ節炎 L13; その他の水疱症 L20; アトピー性皮膚炎 L22; おむつ<ナプキン>皮膚炎 L24; 刺激性接触皮膚炎 L26; 剥脱性皮膚炎 (添付文書に記載の副作用: 剥脱性皮膚炎) L27; 摂取物質による皮膚炎 L28; 慢性単純性苔せん<癬>及び痒疹 L43; 扁平苔せん<癬>
Cluster9, 間質性肺炎、喘息 DPP-4 阻害薬, SI=0.053	Cluster10, 性器感染症、頻尿 SGLT-2 阻害薬, SI=0.081	Cluster11, 乳酸アシドーシス、精神疾患 ビッグアナイド薬, SI=0.071	
J45; 喘息 J46; 喘息発作重積状態 J70; その他の外的因子による呼吸器病態 J84; その他の間質性肺炎 (添付文書に記載の副作用: 間質性肺炎) J85; 肺及び縦隔の膿瘍 J90; 胸水, 他に分類されないもの J92; 胸膜斑<ブランク> J93; 気胸 J94; その他の胸膜病態 J96; 呼吸不全, 他に分類されないもの	N34; 尿道炎及び尿道症候群 N41; 前立腺の炎症性病態 N48; 陰茎のその他の障害 (添付文書に記載の副作用: 亀頭炎、亀頭包皮炎) R31; 詳細不明の血尿 R33; 尿閉 R35; 多尿 (添付文書に記載の副作用: 頻尿、尿量増加) R39; 尿路系に関するその他の症状及び徴候 (添付文書に記載の副作用: 排尿困難、尿意逼迫)	E63; その他の栄養欠乏症 E79; プリン及びピリミジン代謝障害 (添付文書に記載の副作用: 血中尿酸増加) E87; その他の体液、電解質及び酸塩基平衡障害 (添付文書に記載の副作用: 乳酸アシドーシス、乳酸上昇、血中カリウム上昇) F31; 双極性感情障害<躁うつ病> F40; 恐怖症性不安障害 F44; 解離性 [転換性] 障害 F45; 身体表現性障害	

SI; Silhouette Index

Fig. 4 添付文書に記載の副作用を含むクラスター

## 4. 考察

本研究では機械学習技術を活用することで、患者情報に基づいた糖尿病治療薬の有害事象の分散表現ベクトルを作成し、クラスタリングでベクトルの近似度に基づいた有害事象のクラスターを検出し、シグナル検出に応用可能か検討した。添付文書に記載されている副作用を含むクラスターを医学レビューにより確認したところ、関連する有害事象同士が同じクラスターに所属していたことから患者情報に基づく類縁有害事象のクラスター化が可能であることを示した。既知の副作用として知られている SGLT-2 阻害薬、ビッグアナイド薬、及び DPP-4 阻害薬において薬剤クラスに特有のクラスターが検出されたことから、本技法で薬剤クラス特有の副作用を検出できる可能性があることを示した。

SGLT-2 阻害薬とビッグアナイド薬の平均年齢が他の薬剤クラスと比較して低かったことは、これらの薬剤に対して日本糖尿病学会は安全性の観点から高齢者へは慎重投与としている<sup>14,15</sup> ためと考えられる。リスク期間はインスリン製剤が最も短かった。インスリン製剤は血糖コントロール不良

になった2型糖尿病患者に対して短期インスリン強化療法を行い、β細胞の機能及びインスリン感受性が回復した後にインスリン投与を中止し、経口糖尿病治療薬にて以降は血糖コントロールを行う<sup>16</sup> ためと考えられる。インスリン製剤、SGLT-2 阻害薬及びビッグアナイド薬で DPP-4 阻害薬の処方者が最も多く見られた。DPP-4 阻害薬においてはビッグアナイド薬、GLP-1 受容体作動薬ではインスリン製剤の処方者が最も処方されていた。これは、JMDC DB 及び MDV provider を使用した日本における糖尿治療薬の処方実態調査の結果と概ね一致している<sup>17</sup>。

糖尿病の合併症はインスリン製剤で最も多く形成がみられた。これはインスリン製剤が血糖コントロール不良の2型糖尿病患者に処方されるためであると考えられる。感染症のクラスターはいずれの薬剤クラスにおいても糖尿病の関連疾患の中で最も多く形成がみられた。1型、2型に関わらずリスクが上昇し、予後に関しても糖尿病を基礎疾患として有する場合感染症が重症化しやすい<sup>18</sup> ことが反映された可能性がある。精神疾患のクラスターがいくつか形成され

たが、糖尿病に精神疾患を合併する例は少なくなく、特にうつ病の合併率が高い<sup>19</sup>。また、第二世代抗精神病薬では、副作用として体重増加が報告されており<sup>20</sup>、機序が明確になっていないが抗精神病薬が血糖上昇を引き起こすことが示唆されている<sup>21</sup>。したがって、これらの体重減少や血糖上昇に対して、体重増加がみられないDPP-4阻害薬、もしくは体重減少が報告されているGLP-1受容体作動薬が抗精神病薬処方患者に処方された可能性がある<sup>22</sup>。悪性腫瘍に関してもいずれの薬剤クラスについてもクラスターが形成されていた。先行研究においてインスリングルゲン使用者における癌発症リスク上昇が示唆されたが<sup>23</sup>、メタ解析においてはリスク上昇と示唆されなかったため一貫した見解が得られていない<sup>24</sup>。メトホルミンについては先行研究で癌の発症リスク低下が示唆されている<sup>25</sup>。しかし、糖尿病それ自身が悪性腫瘍のリスクファクターであり、インスリン投与患者では悪性腫瘍の副作用報告が多数なされているが、日本の糖尿病診療ガイドライン2019によると現時点で糖尿病治療薬と癌罹患リスクとの関連は十分に解明されていないと結論付けられている<sup>26</sup>。

SGLT-2阻害薬以外の薬剤クラスで低血糖クラスターが形成され、そのうちインスリン製剤のSI値が上位のクラスターとなった。先行研究では重症低血糖の原因薬としてインスリン製剤が最も高いことが報告されている<sup>27</sup>。したがって、インスリン製剤が他の糖尿病治療薬と比較して低血糖を引き起こしやすく、そして低血糖と同時に同じクラスターに含まれる疾患が共起する回数が多かったことでSI値が高くなったことが考えられる。

SGLT-2阻害薬投与患者のみで認められた尿路感染症クラスターと性器感染症と頻尿のクラスターは、RMPで重要な特定されたリスクとして記載されている<sup>28</sup>ことからSGLT-2阻害薬の副作用と関連するクラスターと考えられる。

DPP-4阻害薬投与患者で横紋筋融解症と関節炎、剥離性皮膚炎を含む皮膚疾患、及び間質性肺炎と喘息を含む肺疾患クラスターが認められた。DPP-4阻害薬による関節炎として、多発性関節炎、多関節症、関節リウマチ、重症滑膜炎が報告<sup>29,30</sup>されている。皮膚疾患に関しては、サキサグリブチン、リナグリブチン、アログリブチン、ビルダグリブチンでは剥離性皮膚症を含む重度皮膚有害反応が報告されている<sup>31</sup>。肺疾患に関しては、本邦では2013年に添付文書の重要な副作用として間質性肺炎が追加されている<sup>32</sup>。喘息も同じクラスターに属していたが、Singhらは糖尿病患者の高血糖状態と高インスリン血症が肺機能低下と喘息の高い発生に寄与している可能性を示している<sup>33</sup>。台湾のレセプトデータベースを用いて2型糖尿病におけるDPP-4阻害薬の使用と喘息の発症リスクと重症度の関連を調査した研究では、DPP-4阻害薬の使用が喘息の発症リスクと重症度の低下に関与していた<sup>34</sup>。以上のことから、間質性肺炎・喘息のクラスターは糖尿病の影響も考えられるが、これらの3つのクラスターはDPP-4阻害薬の影響によって形成された可能性が考えられる。

DPP-4阻害薬の副作用として知られており、RMPにも重要な潜在的リスクとして記載されている類天疱瘡のクラスター

がDPP-4阻害薬、SGLT-2阻害薬、及びGLP-1受容体作動薬で認められた。SGLT-2阻害薬とGLP-1受容体作動薬の投与患者ではDPP-4阻害薬を処方された割合が高かったことから、対象薬剤処方前、もしくは併用薬として処方されたDPP-4阻害薬の影響の可能性が考えられる。

ビグアナイド薬で乳酸アシドーシスを含むクラスターが形成されていた。本邦ではメトホルミンによる重篤な乳酸アシドーシスが起る危険があるため、全身状態が悪い患者への投与を禁忌とし、投与する際の注意事項が発出されており<sup>14</sup>、重要な特定されたリスクとして記載している<sup>35</sup>。精神疾患がノイズとして含まれているが、ビグアナイド薬の副作用と関連するクラスターの可能性がある。

本研究ではBERTとWard法の評価指標がいずれも低い値を取ったため、手法の妥当性については改良の余地があり、更なる手法の検討が必要である。特に、インスリン製剤では誤った予測に基づき作成された分散表現によるクラスターが形成された可能性がある。BERTを用いてCPRDの患者を対象に病歴を入力としてMLMを行った関連研究<sup>12</sup>ではprecision microが0.66と本研究のDPP-4阻害薬とGLP-1受容体作動薬の投与患者の結果と同程度であることから分散表現の妥当性はある程度確保されたと考えられる。薬剤クラスごとにprecision microの値がばらついており、これは対象薬剤クラスごとの投与患者のデータの性質に依存している可能性がある。また、自然言語処理技術分野の研究<sup>36</sup>において、BERTで文書の分散表現を獲得し、k-means法や凝集型クラスタリング手法を複数のデータセットに適応して文書のクラスタリングに対する有用性を検証した報告では、SI値が最高で0.46を示すデータセットが存在した一方で、0.086と本研究と同程度のSI値を示すデータセットも存在した。しかし、いずれのデータセットにおいてもBERTを適用しない場合と比較してSI値が上昇していたことからBERTによる変数の分散表現化は有用とされていた。

本研究の強みとして下記の点が挙げられる。まず、本技法では対象薬剤と有害事象間の関連性について、年齢、性別、対象薬剤以外の糖尿病治療薬、併存疾患、及び他の有害事象の情報を考慮しており、多対多の関連について検出を行っている。本技法では患者個人の疾患情報を活用することで、患者の多様性を考慮した有害事象の評価が可能である。次に、本技法は教師なし学習のため、既知情報に左右されず、有害事象辞書のコード体系を超えた関連性を特定可能な点である。つまり、本研究ではデータドリブンな関連性の考慮が可能である。次に、レセプトデータのみで構築できる点である。従来、安全性シグナル検出に用いられている自発報告データベースのように医療従事者からの報告の労務を必要としない。レセプトデータでは診療報酬請求対象の診療情報が全て含まれるため、未知の潜在的な有害事象の特定が可能なが自発報告より優れている可能性がある。次に、疾患と薬剤に依存せずに構築できる点である。本研究では糖尿病治療薬である5つの薬剤クラスを対象に本技法を適用したが、十分な患者数が見込める場合に薬剤クラスを問わずに本技法を適用可能である。

本研究にはいくつかの限界が存在する。本研究では有害事象の定義に TEAE を用い、対象薬剤処方後に認められる疾患全てを有害事象として定義した。したがって、原疾患に伴う合併症が多く検出される。クラスターの解釈の際には合併症のクラスターを除外するための医学的な解釈の必要性があるため、本研究では検出したクラスターに対して医学レビューを行った。また、有害事象を発症したケースのみを対象患者としているため、有害事象発症のリスクファクターが特定できない。次に、BERT が深層学習を基にしたモデルのため解釈性が低い。したがって、疾患の分散表現の作成に寄与した患者集団を特定することができない。次に、Ward 法によって形成したクラスターは、併用薬剤もしくは対象薬剤の処方より前の他の糖尿病治療薬の影響を受けている可能性がある。したがって、本研究では各薬剤クラスで形成したクラスターを比較することで特定の薬剤クラスで形成した特有なクラスターを検出した。次に、解析対象集団内で発症人数が 20 人未満の有害事象は除外した。このため、稀な有害事象について考慮ができていない。次に、本技法では SI が発現回数に依存する可能性があるため、発現回数の多い糖尿病の合併症クラスターの SI は高値をとりやすく、発現回数の少ない有害事象の SI が低くなる。次に、学習のために大きなサンプルサイズと膨大なデータサイズを処理可能なコンピューテーションが必要である。

## 5. 結論（と今後の展望）

本研究では分散表現とクラスタリング手法をレセプトデータベースに程起用し、新たな着想に基づく安全性シグナル検出手法提案した。糖尿病治療薬を対象に本技法で有害事象をクラスタリングした結果を比較し、SGLT-2 阻害薬、DPP-4 阻害薬、及びビグアナイド薬に特有な既知の副作用と関連するクラスターを検出した。以上より、本技法がシグナル検出に活用できる可能性が示唆された。さらなる改良のために今後は他の薬剤に適用し、本技法の特性を明確にすることが重要である。

## 参考文献

- Council for International Organizations of Medical Sciences WG. Management of Safety Information from Clinical Trials –Report of CIOMS Working Group VI. 2005; [https://cioms.ch/wp-content/uploads/2017/01/Mgmt\\_Safety\\_Info.pdf](https://cioms.ch/wp-content/uploads/2017/01/Mgmt_Safety_Info.pdf).
- Berry SM, et al. Accounting for multiplicities in assessing drug safety: a three-level hierarchical mixture model. *Biometrics*. 2004;60(2):418-426.
- Crooks CJ, et al. Prieto-Merino D, Evans SJ. Identifying adverse events of vaccines using a Bayesian method of medically guided information sharing. *Drug Saf*. 2012;35(1):61-78.
- 漆原尚巳. 自発報告に基づくシグナル検出ができること、できないこと—その理論と実践—. *医薬品情報学*. 2020;21(4):135-141.
- Platt R, et al. Multicenter epidemiologic and health services research on therapeutics in the HMO Research Network Center for Education and Research on Therapeutics. *Pharmacoepidemiol Drug Saf*. 2001;10(5):373-377.
- Kim JA, et al. Towards Actualizing the Value Potential of Korea Health Insurance Review and Assessment (HIRA) Data as a Resource for Health Research: Strengths, Limitations, Applications, and Strategies for Optimal Use of HIRA Data. *J Korean Med Sci*. 2017;32(5):718-728.
- Choi NK, et al. Signal detection of rosuvastatin compared to other statins: data-mining study using national health insurance claims database. *Pharmacoepidemiol Drug Saf*. 2010;19(3):238-246.
- Wang CS, et al. Detecting Potential Adverse Drug Reactions Using a Deep Neural Network Model. *J Med Internet Res*. 2019;21(2):e11016.
- バイエル薬品株式会社. 2.7.4 臨床的安全性. [https://www.pmda.go.jp/drugs/2014/P201400097/630004000\\_22000AMX00014\\_K101\\_L.pdf](https://www.pmda.go.jp/drugs/2014/P201400097/630004000_22000AMX00014_K101_L.pdf). Accessed 2023-01-04.
- Meng Y, et al. Bidirectional representation learning from transformers using multimodal electronic health record data to predict depression. *IEEE Journal of Biomedical and Health Informatics*. 2021;25(8):3121-3129.
- Devlin J, et al. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*. 2018.
- Li Y, et al. BEHRT: Transformer for Electronic Health Records. *Sci Rep*. 2020;10(1):7155.
- Vendramin L, et al. Relative clustering validity criteria: A comparative overview. *Statistical Analysis and Data Mining: The ASA Data Science Journal*. 2010;3.
- 公益社団法人 日本糖尿病協会. メトホルミンの適正使用に関する Recommendation. [https://www.nittokyo.or.jp/modules/information/index.php?content\\_id=23](https://www.nittokyo.or.jp/modules/information/index.php?content_id=23). Accessed 2022-12-30.
- 公益社団法人 日本糖尿病協会. 糖尿病治療における SGLT2 阻害薬の適正使用に関する Recommendation. [http://www.fa.kyorin.co.jp/jds/uploads/recommendation\\_SGLT2.pdf](http://www.fa.kyorin.co.jp/jds/uploads/recommendation_SGLT2.pdf). Accessed 2023-01-05.
- 岩城正輝, et al. 短期強化インスリン療法と  $\beta$  細胞保護療法を併用した 2 型糖尿病治療の有用性. *糖尿病*. 2008;51(4):309.
- Nishimura R, et al. Treatment patterns, persistence and adherence rates in patients with type 2 diabetes mellitus in Japan: a claims-based cohort study. *BMJ Open*. 2019;9(3):e025806.
- Muller LM, et al. Increased risk of common infections in patients with type 1 and type 2 diabetes mellitus. *Clin Infect Dis*. 2005;41(3):281-288.
- Anderson RJ, et al. The prevalence of comorbid depression in adults with diabetes: a meta-analysis. *Diabetes Care*. 2001;24(6):1069-1078.
- Doane MJ, et al. Antipsychotic Treatment Experiences of People with Schizophrenia: Patient Perspectives from an Online Survey. *Patient Prefer Adherence*. 2020;14:2043-2054.
- Sawagashira R, et al. Subthreshold Change in Glycated Hemoglobin and Body Mass Index After the Initiation of Second-Generation Antipsychotics Among Patients With Schizophrenia or Bipolar Disorder: A Nationwide Prospective Cohort Study in Japan. *J Clin Psychiatry*. 2022;83(3).
- Yamada Y, et al. Dose-response, efficacy, and safety of oral semaglutide monotherapy in Japanese patients with type 2 diabetes (PIONEER 9): a 52-week, phase 2/3a, randomised, controlled trial. *Lancet Diabetes Endocrinol*. 2020;8(5):377-391.
- Jonasson JM, et al. Insulin glargine use and short-term incidence of malignancies—a population-based follow-up study in Sweden. *Diabetologia*. 2009;52(9):1745-1754.
- Johnson JA, et al. Intensive glycaemic control and cancer risk in type 2 diabetes: a meta-analysis of major trials. *Diabetologia*. 2011;54(1):25-31.
- Evans JM, et al. Metformin and reduced risk of cancer in diabetic patients. *Bmj*. 2005;330(7503):1304-1305.
- 日本糖尿病学会 社. 糖尿病診療ガイドライン 2019. [http://www.jds.or.jp/modules/publication/index.php?content\\_id=4](http://www.jds.or.jp/modules/publication/index.php?content_id=4). Accessed 2022-01-04.
- 難波光義, et al. 日本糖尿病学会糖尿病治療に関連した重症低血糖の調査委員会. 糖尿病治療に関連した重症低血糖の調査委員会報告. *糖尿病*. 2017; 60:826-842.
- アステラス製薬株式会社. スーグラ錠 25mg、50mg に係る 医薬品リスク管理計画書. 2022; [https://www.pmda.go.jp/RMP/www/800126/99381dc5-3a82-4336-aeba-912ccd0305a8/800126\\_3969018F2029\\_016RMP.pdf](https://www.pmda.go.jp/RMP/www/800126/99381dc5-3a82-4336-aeba-912ccd0305a8/800126_3969018F2029_016RMP.pdf). Accessed 2023-01-06.
- Yokota K, et al. Sitagliptin (DPP-4 inhibitor)-induced rheumatoid arthritis in type 2 diabetes mellitus: a case report. *Intern Med*. 2012;51(15):2041-2044.
- Sasaki T, et al. Acute onset of rheumatoid arthritis associated with administration of a dipeptidyl peptidase-4 (DPP-4) inhibitor to patients with diabetes mellitus. *Diabetology International*. 2010;1(2):90-92.
- Huang J, et al. Adverse event profiles of dipeptidyl peptidase-4 inhibitors: data mining of the public version of the FDA adverse event reporting system. *BMC Pharmacology and Toxicology*. 2020;21(1):68.
- 厚生労働省. 医薬品・医療機器等安全性情報 No. 305. 2013; [https://www.mhlw.go.jp/www1/kinkyu/iyaku\\_j/iyaku\\_j/anzenseijyouhou/305.pdf](https://www.mhlw.go.jp/www1/kinkyu/iyaku_j/iyaku_j/anzenseijyouhou/305.pdf). Accessed 2023-1-17.
- Singh S, et al. Insulin and the lung: connecting asthma and metabolic syndrome. *J Allergy (Cairo)*. 2013;2013:627384.
- Li PF, et al. Association of dipeptidyl peptidase-4 inhibitor use and the risk of asthma development among type 2 diabetes patients. *Ther Adv Respir Dis*. 2022;16:17534666221135320.
- 株式会社三和化学研究所. メトホルミン塩酸塩錠 250mgMT 「三和」/メトホルミン塩酸塩錠 500mgMT 「三和」に係る医薬品リスク管理計画書. 2022; [https://med.skk-net.com/supplies/item/202212/300297\\_3962002F2078\\_003RMP.pdf](https://med.skk-net.com/supplies/item/202212/300297_3962002F2078_003RMP.pdf). Accessed 2023-01-06.
- Mehta V, et al. WEClustering: word embeddings based text clustering technique for large datasets. *Complex & Intelligent Systems*. 2021;7(6):3211-3224.