臨床薬学領域における AI の活用によるビックデータ解析の可能性

桐 生 嘉 浩

Potential for Big Data Analysis Using AI in the Field of Clinical Pharmacy

Yoshihiro Kiryu

M&B Collaboration Medical Corporation Hokuetsu Hospital; 2-20-19 Midori-cho, Shibata, Niigata 957-0018, Japan.

(Received August 26, 2020)

Industrial reforms utilizing artificial intelligence (AI) have advanced remarkably in recent years. The application of AI to big data analysis in the medical information field has also been advancing and is expected to be used to find drug adverse effects that cannot be predicted by conventional methods. We have developed an adverse drug reactions analysis system that uses machine learning and data from the Japanese Adverse Drug Event Report (JADER) database. The system was developed using the C# programming language and incorporates the open source machine learning library Accord.Net. Potential analytical capabilities of the system include discovering unknown drug adverse effects and evaluating drug-induced adverse events in pharmaceutical management. However, to apply the system to pharmaceutical management, it is important to examine the characteristics and suitability of the level of AI used in the system and to select statistical methods or machine learning when appropriate. If these points are addressed, there is potential for pharmaceutical management to be individualized and optimized in the clinical setting by using the developed system to analyze big data. The system also has the potential to allow individual healthcare facilities such as hospitals and pharmacies to contribute to drug repositioning, including the discovery of new efficacies, interactions, and drug adverse events.

Key words——adverse effect; database; machine learning; clinical pharmacy information system; medical informatics computing

はじめに

近年、世界中で人工知能 (artificial intelligence; AI) を活用した産業改革が、目覚ましく進んでいる。¹⁾ 医療分野における AI は、IBM Watson Health[®]などに代表されるような診断支援を目的とした医療 AI サービスとしても利用されている。²⁻⁴⁾ また、医療情報領域では AI を利用したビックデータ解析の応用も進められており、従来の手法からは予測できない副作用や効果を見出す手法の 1 つとしても期待されている。⁵⁻⁷⁾

これまで、筆者らは、独立行政法人医薬品医療機器総合機構 (Pharmaceuticals and Medical Devices Agency; PMDA) が 2012 年から一般公開している有害事象自発報告データベース (Japanese Adverse Drug Event Report database; JADER) を利用し、

AI を実現する手段として機械学習を組み入れた副作用解析システムの開発を進めている。^{8,9)} 本稿では、臨床現場での副作用解析を進める上で、機械学習を用いた副作用の評価、支援の方法について紹介するとともに、医薬品副作用データベースを利用した機械学習システムの援用の可能性について叙述する。

1. 副作用解析システム

独自開発は、開発用マシンとして Microsoft[®] Windows 10[®] 64 bit (Microsoft Corporation, Redmond), プロセッサ Intel[®] Core™ i7-4790 (Intel Corporation, Santa Clara), CPU 3.20 GHz, 実装メモリ (RAM) 8.00 GB を用いた。統合開発環境は、Visual Studio 2013[®] (Microsoft Corporation) とコンパイラ言語のプログラミング言語 C# (Microsoft Corporation) を使用し、主要なコントロール以外は、サードパーティ製の外部コンポーネント DLL を取込む仕様で構築した。ロードする機械学習ライブラリは、オープンソースライブラリの Accord.Net Machine Learning Framework [Version

医療法人社団 M&B コラボレーション北越病院 (〒957-0018 新潟県新発田市緑町 2-20-19)

e-mail: hpharma@mbc.or.jp

本総説は、日本薬学会第 140 年会シンポジウム S45 で発表した内容を中心に記述したものである。

	Target adverse events	Non-target adverse events	Total
Target drug	a	b	a+b
Non-target drug	c	d	c+d
Total	a+c	b+d	a+b+c+d

$$ROR = \frac{a/c}{b/d} = \frac{ac}{bd} \quad p = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{(a+b+c+d)!a!b!c!d!}$$

$$95\% CI = \exp \left[\ln(ROR) \pm 1.96 \times \sqrt{\left(\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d} \right)} \right]$$

Fig. 1. Two-by-Two Cross-tabulation Table Used to Calculate the Reporting Odds Ratio (ROR), 95% Confidence Interval (95% CI), and *p*-Value by Fisher's Exact Test

3.8.0；ライセンス：GNU Library or Lesser General Public License (LGPL) version 3.0] を使用し，グラフや散布図はオープンソースのグラフコントロールライブラリである ZedGraph [Version 5.1.7；ライセンス：LGPL version 2.1] を用いた。^{10,11)} 独自開発の非同期的高速多重並列処理アルゴリズムを搭載した解析用サーバーは，逐次動的にデータ解析しミリ秒単位 (ms) でクライアントとデータ共有する仕様であり，解析用データベースのすべての前処理は，データソース更新時のみ全自動処理され，所

要時間としては 30 分以内に完了する。

2. 解析用データベース

JADER データベースは年々データ登録件数が増加しているが，本稿で例示する解析データは 2004 年 4 月–2017 年 11 月までの有害事象自発報告データを使用している。対応する International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) 国際医薬用語集日本語版 (Medical Dictionary for Regulatory Activities/Japanese version; MedDRA/J) は，Version 20.1 である。また，ダウンロードした csv ファイル (症例一覧テーブル，原疾患テーブル，副作用テーブル，原疾患テーブル) の各データは，重複データの削除等のクリーニング処理を施して整形し，これを解析用データベースとして動的にシステムにインポートしている。

JADER データを用いた解析における特性については，副作用が発現しなかった患者情報は含まれていないことが重要な基礎情報の 1 つとして挙げられる。ゆえに，JADER データを用いた場合では，ある特定の薬物が誘発する特定の副作用の発現率は算出できない。そこで，薬物と副作用・有害事象の相関を吟味する手法として，2×2 のクロス集計表を用いた 2 変量解析により，報告オッズ比 (report-

Fig. 2. Sample Form of the Adverse Drug Reactions Analysis System for Acute Liver Failure (Color figure can be accessed in the online version.)

Fig. 3. Optimal Number of Clusters Using X-means Analysis of Volcano Plot Data Associated with Acute Liver Failure

Fig. 4. Volcano Plot with Cluster Analysis Based on Gaussian Mixture Models
(Color figure can be accessed in the online version.)

ing odds ratio; ROR) や 95%信頼区間 (95% confidence interval; 95% CI) などのシグナル指標を用いて評価する方法が広く応用されている (Fig. 1).^{12,13)} これに基づいて、筆者らは先行研究として毒性学の領域で利用されている volcano プロットに機械学習を組み込み、瞬時に任意の医薬品と副作用の相関を視覚的に捉えられるアプリケーションツール

を作製した (Fig. 2).¹⁴⁾ 散布図は、クロス集計表の各報告数に 0.5 を加えて補正した ROR を基に、縦軸は 5 未満の報告数でも有効なフィッシャーの正確検定から得られる危険率 (p 値) の常用対数 $[-\log(p \text{ 値})]$ 、横軸は ROR の 95% 信頼区間下側 (lower limit of 95% CI; LCI) の自然対数 $[\ln(\text{LCI})]$ を表している。¹⁵⁾ ここで、volcano プロットの右上

Table 1. Volcano Plot Data of Medicines Associated with Acute Liver Failure: Distribution Group, $\ln(\text{LCI})$, $-\log(p\text{-value})$, Number of Reports, ROR, and 95% CI

Drug Name	$\ln(\text{LCI})^a$	$-\log(p\text{-value})$	Number of reports	ROR ^b	95% CI ^c [Upper–Lower]
Distribution group 1					
Isoniazid	2.394	49.702	15	18.301	[10.96–30.55]
Acetaminophen	2.064	40.602	19	12.474	[7.87–19.75]
Pyrazinamide	2.418	32.978	8	22.253	[11.22–44.13]
Rifampicin	2.139	32.145	12	14.991	[8.49–26.44]
Amiodarone Hydrochloride	1.899	23.105	11	12.052	[6.67–21.75]
Distribution group 2					
Pyridoxal Phosphate Hydrate	2.611	21.224	3	39.376	[13.61–113.88]
Clobazam	2.235	17.325	4	23.813	[9.34–60.67]
Estazolam	2.098	15.002	4	20.742	[8.14–52.81]
Mosapride Citrate Hydrate	1.891	13.818	5	15.441	[6.62–35.97]
Peramivir Hydrate	2.128	12.959	3	24.206	[8.39–69.78]
Disopyramide	2.086	12.412	3	23.212	[8.05–66.90]
Chlorpromazine/Promethazine	1.816	11.121	4	15.638	[6.14–39.78]
Distribution group 3					
Tegafur/Uracil	1.426	10.840	8	8.240	[4.16–16.31]
Ethambutol Hydrochloride	1.504	10.002	6	9.796	[4.49–21.33]
Omeprazole Sodium	1.675	9.547	4	13.578	[5.33–34.52]
Blonanserin	1.519	8.040	4	11.615	[4.56–29.52]
Abiraterone Acetate	1.396	7.942	5	9.400	[4.03–21.88]
Antipyretic Analgesic Anti-inflammatory Agent (OTC drug)	1.482	7.711	4	11.186	[4.40–28.43]
Tolvaptan	1.182	6.814	6	7.100	[3.26–15.45]
Biperiden Hydrochloride	1.221	5.724	4	8.610	[3.38–21.87]
Sorafenib Tosilate	0.806	5.686	11	4.052	[2.24–7.30]
Zonisamide	1.084	5.471	5	6.876	[2.95–15.99]
Pilsicainide Hydrochloride Hydrate	1.128	5.134	4	7.848	[3.08–19.93]
Edaravone	1.101	4.972	4	7.639	[3.00–19.40]
Lansoprazole	0.698	4.117	8	3.984	[2.01–7.88]
Distribution group 4					
Asunaprevir	1.128	4.419	3	8.881	[3.09–25.52]
⋮	⋮	⋮	⋮	⋮	⋮
Fludarabine Phosphate	0.058	1.064	3	3.048	[1.06–8.74]

^a Lower limit of 95% confidence interval, ^b reporting odds ratio, ^c 95% confidence interval.

方のデータ群は対象有害事象を惹起し易い可能性が示唆されるが、その数値の大小は統計的有意差を表しており、可能性の尺度を示すわけではない。そこで、個々のデータの類似性を群分類に分けてビジュアル化することで、データ群としての傾向を視覚的に把握できるように機械学習を用いて改修した。組込んだ機械学習の種類は、教師あり学習、教師なし学習、及び強化学習に分類される。¹⁶⁾ Volcano プロットでは教師なし学習のクラスタ分析 (k-means) とガウス混合モデルを複合させた。^{17,18)} さらに、クラスタ分析は、初期値に依存して最適クラスタ数に変化してしまう問題点を改善するために、ベイズ情

報量基準 (Bayesian Information Criterion; BIC) によるクラスタ数の最適化分析 (x-means) を応用し、クラスタ群をより客観的に分割する手法を適用した (Fig. 3)。^{19,20)}

3. 副作用解析

医薬品による潜在的リスクは、医薬品リスク管理計画書 (risk management plan; RMP)、医薬品添付文書、インタビューフォーム (interview form; IF) などの医薬品情報を参照して把握されている。しかしながら、市販後調査を経た医薬品再評価時に改めて認められるような、後に重大なリスクとして認識される副作用や有害事象は数多くある。このような

Fig. 5. Flow Chart of Pharmaceutical Management Model with the Adverse Drug Reactions Analysis System

副作用リスクをあらかじめ視野に入れた副作用評価を研究機関に限らず、臨床現場においても進められるのであれば、より網羅的に副作用管理を実践することが可能になると筆者は考えている。例えば、バソプレシン V₂ 受容体拮抗薬であるトルバプタン（サムスカ®）の添付文書に記載されている副作用の項の急性肝不全は、2018 年 3 月に改訂第 17 版で追記された。^{21,22)} この改訂が公開される 4 ヶ月前の 2017 年 11 月時の JADER データに基づいたシステム解析結果では、 $-\log(p \text{ 値})$ は、 $6.814 [> -\log(0.05)]$ であり、機械学習によるクラスタ分類は、重心点の縦軸値が 3 番目に高い群（D3 群）に含まれ、かつ同群内の各データはすべて統計的に有意であった（ $p < 0.05$ ）（Fig. 4）。また、MedDRA 標準検索式（standardized MedDRA queries; SMQ）の肝障害（SMQ: 20000005）に属する基本語（preferred term; PT）の急性肝不全（PT: 10000804）とその類縁疾患である劇症肝炎（PT: 10019772）の添付文書記載内容を調査した。²³⁾ その結果、クラスタ群内のデータがすべて統計的に有意であった 1-3 群の同群内医薬品の急性肝不全、又は劇症肝炎の記載状況は、1 群 100.0%（5/5 件）、2 群 28.6%（3/7 件）、及び

3 群 54.5%（6/11 件）であった（Table 1）。加えて、企業の作成する RMP の特定されたリスク、及び潜在的リスクの項において、この時点での該当の疾患名の記載はなかった。²⁴⁾ しかしながら、総合的なリスクマネジメントの観点から鑑みると、重要な副作用の項に肝機能障害が記載されていることから、状態によっては急性増悪を回避するよう注意深く観察し管理することが望ましいと評価することができる。^{21,25)} ただし、添付文書に追記される副作用は、それ自体の発現率に限って記載されるわけではないことに注意が必要である。さらに、特に自発報告データの特性や機械学習の仕組みを吟味せず、報告バイアスや意味づけを考慮しない場合、それは大変誤った「予測結果」につながる可能性がある。ゆえに、これらのことに留意して、副作用評価やリスク管理を実践しなければならない。つまり、このようなアプローチで、かならずしも添付文書改訂にまでつながるような結果が早々に判定できるというわけではない。²⁶⁾ また、一般的な機械学習が導き出す様々な結果は、あくまでもヒトが与えたデータに対する「予測結果」であり、因果関係を強調するような統計学的解析により算出された結果ではないという

ことである。もちろん、広義的には適切なバリデーションにより得られた機械学習手法であれば、解析結果を高く評価できるが、それは因果や意味の解釈を含んだものではない点を再認識しなければならない。

4. 副作用管理と運用

現在、臨床現場での副作用管理は、あらゆる情報を総動員してリスクマネジメントを実践していくことが大前提である。当然、当システムを利用した結果は、そのうちの情報の1つであって、この情報をより吟味する必要がある検討材料として採用するか却下するか、あるいはリスク管理のために適用するかしないかはヒトが判断する。加えて、機械学習による結果は、データベースの質や前処理とバリデーションに大きく依存するため、結果を解釈する基準の1つとして考慮しなければならない。このため、当システムはヒトに依存した医薬品情報支援ツールであり、それがなくてはならない基幹システムレベルでは使用していない。また、個別の最適な医薬品のリスク管理は、リスクとなる複雑な要因を個々の患者の状態に照らして総合的に判断する必要がある。機械学習を用いた場合であっても、判定方法の詳細を含めてすべて自動化した画一的な手法は現在のところ極めて困難、かつ危険であると言える。一方、更なる機能を追加した上で、支援ツールとして適切に配慮されることによって、より強力に個別の医薬品リスク管理に資するモノとして扱える。したがって、統計学的解析手法を適宜用いたリスク管理に限らず、機械学習手法を用いる場合も同様に、常に解析された結果の検証、調査、評価を必須として運用しなければならない (Fig. 5)。

おわりに

副作用解析は、臨床現場での薬剤師が実践できる業務の1つではあるが、対象の副作用の種類は膨大であり、複雑な要因により発現することが少なくない。そのため、副作用以外の可能性も含めて、可及的速やかに対象の副作用の起因薬物を推論することは決して容易ではない。しかしながら、様々な薬剤師業務支援システムの独自開発が臨床現場で進められていることから、フリーのオープンソースやソフトウェアを利用することで、ビックデータと機械学習を利用した有益性が高い副作用解析システムも、現場主導で発展させることができると考える。²⁷⁻²⁹⁾

また、解析技術の向上や機械学習手法の知見が蓄積されることで、そのようなツールの創出の可能性は高まっていくであろう。今後、臨床現場の薬剤師の副作用解析を支援するためにも、機械学習と統計学を交えた副作用評価手法やリスク管理を支援するツールの開発、あるいはそれらの臨床応用が分野横断的に多方面からさらに進むことを期待したい。

利益相反 開示すべき利益相反はない。

REFERENCES

- 1) Columbus L., "Roundup of Machine Learning Forecasts and Market Estimates for 2019," Forbes Media LLC, 27 March, 2019: <https://www.forbes.com/sites/louiscolumbus/2019/03/27/roundup-of-machine-learning-forecasts-and-market-estimates-2019/>, cited 23 January, 2020.
- 2) IBM Watson Health. "Healthcare provider solutions for the modern health system.": <https://www.ibm.com/watson-health/solutions/healthcare-provider>, cited 3 July, 2020.
- 3) Liu C., Liu X., Wu F., Xie M., Feng Y., Hu C., *J. Medical Internet Res.*, **20**, e11087 (2018).
- 4) Jiang F., Jiang Y., Zhi H., Dong Y., Li H., Ma S., Wang Y., Dong Q., Shen H., Wang Y., *Stroke Vasc. Neurol.*, **2**, 230-243 (2017).
- 5) Ahmed Z., Mohamed K., Zeeshan S., Dong X. Q., *Database*, baaa010 (2020).
- 6) Harrer S., Shah P., Antony B., Hu J., *Trends Pharmacol. Sci.*, **40**, 577-591 (2019).
- 7) Kozawa S., Sagawa F., Endo S., De Almeida G. M., Mitsuishi Y., Sato T. N., *iScience*, **23**, 100791 (2020).
- 8) Pharmaceuticals and Medical Devices Agency. "Drug Side Effects Database Terms of Use.": <https://www.pmda.go.jp/safety/info-services/drugs/adr-info/suspected-adr/0003.html>, cited 3 July, 2020.
- 9) Kiryu Y., Yoshino T., Murai Y., Sato Y., Abstract of papers, the 21st Annual Meeting of the Japanese Society of Drug informatics, Suzuka, June 2018, O1-5.
- 10) Accord.NET Framework: <http://accord-framework.net/index.html>, cited 7 April, 2020.

- 11) ZedGraph: <https://sourceforge.net/projects/zedgraph/>, cited 7 April, 2020.
- 12) Niinomi I., Hosohata K., Oyama S., Inada A., Wakabayashi T., Iwanaga K., *Int. J. Toxicol.*, **38**, 487–492 (2019).
- 13) Inaba I., Kondo Y., Iwasaki S., Tsuruhashi S., Akaishi A., Morita K., Oniki K., Saruwatari J., Ishitsuka Y., Irie T., *Front. Pharmacol.*, **10**, 874 (2019).
- 14) Uesawa Y., *Yakugaku Zasshi*, **138**, 185–190 (2018).
- 15) Sugawara H., Uchida M., Suzuki S., Suga Y., Uesawa Y., Nakagawa T., Takase H., *Biol. Pharm. Bull.*, **42**, 1185–1191 (2019).
- 16) Akiba S., Sugiyama A., Terada M., “Kikaigakushu Zukan,” Shoeisha Co., Ltd., Tokyo, 2019.
- 17) Hartigan J., Wong M., *Appl. Stat.*, **28**, 100–108 (1979).
- 18) Akiba S., Sugiyama A., Terada M., “Kikaigakushu Zukan,” Chap. 15, Shoeisha Co., Ltd., Tokyo, 2019, pp. 130–135.
- 19) Schwarz G., *Ann. Stat.*, **6**, 461–464 (1978).
- 20) Pelleg D., Moore A., Proceedings of the 17th International Conference on Machine Learning, Stanford, June–July 2000, pp. 727–734.
- 21) Otsuka Pharmaceutical Co., Ltd., Samusca® (torvaptan) package insert, 1st ed., June 2020: https://www.pmda.go.jp/PmdaSearch/iyakuDetail/ResultDataSetPDF/180078_2139011D1022_1_08, Pharmaceuticals and Medical Devices Agency Web, cited 5 July, 2020.
- 22) Otsuka Pharmaceutical Co., Ltd., Samusca® (torvaptan) interview form, 20th ed., June 2020: https://www.info.pmda.go.jp/go/interview/1/180078_2139011D1022_1_023_1F.pdf, Pharmaceuticals and Medical Devices Agency Web, cited 5 July, 2020.
- 23) Mochida S., *Nihon Shokakibyō Gakkai Zasshi*, **112**, 813–821 (2015).
- 24) Otsuka Pharmaceutical Co., Ltd., Samusca® (torvaptan) risk management plan, 10 December 2019: https://www.pmda.go.jp/RMP/www/180078/bdcf0385-dca9-4b6d-a2df-8c28c7632717/180078_2139011D1022_004RMP.pdf, Pharmaceuticals and Medical Devices Agency Web, cited 5 July, 2020.
- 25) National Institute of Diabetes and Digestive and Kidney Diseases. “LiverTox: Clinical and Research Information on Drug-Induced Liver Injury.”: https://www.ncbi.nlm.nih.gov/books/NBK548503/pdf/Bookshelf_NBK548503.pdf, National Center for Biotechnology Information Web, cited 5 July, 2020.
- 26) Matsui K., Iimura Y., *Farumashia*, **50**, 543–547 (2014).
- 27) Kiryu Y., Honma N., Yoshino T., Murai Y., Sobue T., Sato Y., *Jpn. J. Drug Inform.*, **19**, 8–16 (2017).
- 28) Sato A., Sakata H., Suzuki A., Kato N., Kuroda N., Katayama H., Amano T., Matsumoto K., Sakakibara Y., Itakura Y., *Jpn. J. Med. Inf.*, **34**, 75–80 (2014).
- 29) Igarashi T., Konno A., Tsukamoto H., Yano R., Watanabe K., Nakamura T., Masada M., Goto N., *Jpn. J. Drug Inform.*, **20**, 66–71 (2018).