



Artificial Intelligence and Machine Learning for Safe Medicines

Andrew Bate^{1,2,3} · Yuan Luo⁴

Accepted: 16 March 2022

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2022

This special issue on machine learning (ML) in *Drug Safety* illustrates the extent to which the excitement around ML in broader society is now pervading pharmacovigilance (PV). Pharmacovigilance is in a time of great change [1] and there is much discussion of the role that newer technologies including ML can and will play in driving this necessary change. This excitement is not of course unique to PV and we see widespread scientific research and discussion of ML and artificial intelligence (AI) in the broader healthcare field. Machine learning used routinely in many applications, for example, voice recognition for automated clinical visit scribing, and visual pattern recognition for medical imaging such as in retinopathy [2]. Given the complexity of medicine and healthcare delivery, rule-based systems that typically involve human-curated rule sets are necessarily limited in capability both for recognizing large numbers of complex patterns and for automation in data ingestion, pre-processing, and dissemination. One would theoretically anticipate expert-crafted rule-based systems will be upper bounded by human capacity, where with sufficient rich well-labeled training data and generalizability the potential for ML seems much more promising.

While the use of ML is not new in safety, see for example [3] nor even in application to safety reports, for example [4], there has been limited routine use in PV and there are many reasons for this [5]. However, there are signs this is changing and some of these barriers are beginning to be overcome, particularly in the area of natural language processing, which is finding extensive use in the extraction of information from free-text clinical notes in electronic health records [6]. In this issue, there is a scoping review [7] showing the breadth of research from data ingestion to signal detection. There

clearly remains much confusion and lack of clarity around the scope of ML and AI and the usage as discussed in a systematic review [8] shows there is clearly a huge increase in published research on AI-based ML [7]. This is further illustrated by the wide range of examples of original research covering applications as diverse as predicting drug approvals [9], automated patient-reported adverse event [10] and drug coding [11], and adverse event report causality assessment [12], and disease prediction [13] and its role in supporting decision making by safety experts during signal validation [14]. This issue also contains perspectives from different stakeholders and data networks [15–17], insights and challenges into how ML can help facilitate identifying the completely unexpected ‘black swan events’ [18] and insights into how ML is making inroads into causal inference and telehealth and in resource-limited settings [19–21].

Despite the range of articles, it would however be a mistake to believe that all the challenges for effective, trusted, routine production ML have been resolved—we are still some way from ubiquitous ML yet! Many thorny issues remain for the use of ML in PV. Consider a few examples. How important is contemporaneous explainability in the broadest sense of the term? Clearly, the ability to explain so another understands the reasoning behind an output boosts trust in the system, but is it essential? Does this depend on the application or even the choice of the algorithm? For example, should we prefer deterministic over non-deterministic algorithms? A requirement on contemporaneous explainability may limit performance especially if we require the ML to only do what a human can do or at least comprehend the value in the immediacy of a ML-based suggestion. The example of the board game ‘Go’ and move 37 during the second game of a series defeat of the human champion by an AI system springs to mind, which at the time of the move was not readily appreciated, for example, “that’s a very strange move” and “I thought it was a mistake” [22], and was since seen as brilliant through the lens of retrospect. Yet if one was always to require some sort of retrospective comprehensibility to a human as a condition for trust, how would one define the length that a retrospective evaluation

✉ Andrew Bate
andrew.x.bate@gsk.com

¹ GSK, Brentford, UK

² LSHTM, London, UK

³ New York University, New York, NY, USA

⁴ Northwestern University, Evanston, IL, USA

would apply? If this was a necessary cause, this could clearly preclude the timely use of some ML outputs for decision making in certain circumstances. Would AlphaGo have lost that second game of Go had a human-required approval been needed for proposed move 37 and denied?

Similarly, performance of ML is a contested issue, in safety, we have a responsibility to show we are doing all that we can to ensure the safe use of medicines and continuously strive to improve. How do we show improvement with new technologies? Much has been said that strong performance of ML in a subtask is promising for future PV, but it must also improve performance across the overall PV lifecycle [18], rather than just creating work or inefficiencies or delays elsewhere that overwhelm improvement at the particular steps or tasks. A compelling evaluation of safety system performance on a holistic level is notoriously difficult, however, with much discussion about reference sets for method evaluation [23]: progress in this area will be needed to show the future value of ML.

Much data of relevance to PV cannot be shared readily because of the broader ethical and privacy concerns around sharing healthcare data. Arguably, sharing code has limited value unless it is simple to follow and therefore can easily be accurately implemented and adapted with minimal effort to run on a data source or run on public domain data. In other parts of healthcare, many articles rely on data sets available in the public domain [24]. Data sources that have been used to train ML models more generally are increasingly being released so that results can be fully reproduced by other researchers. For example, the use of ML on real-world data for PV might accelerate if we see more publications with open source code that runs on public domain real-world data sources such as for example, MIMIC-III [25] or the US Centers for Disease Control and Prevention NHANES data set, and this could similarly foster reproducibility as well as address confidence in performance.

So what to the future? We expect to see more efforts to integrate ML holistically across the entire PV lifecycle, and as the need for rapid and effective learning from emerging data for decision making has become even more evident during the COVID-19 pandemic, we anticipate also evolutions in workflows [26]. For example, performance of disproportionality analyses of individual case safety reports is more limited in identifying high-order dependences such as drug–drug interactions because of the increased impact on quantitative scores of coincidentally similar reports, which may be artefactual. Machine learning to cluster and quantify similarity of reports [27] has been used to downweigh likely duplicative reports to make such higher order dependency signal detection more effective [28], therefore linking ML in data ingestion to a subsequent data analysis.

In addition to more clear links between ML for data ingestion and analysis, we expect to see more evidence of

the implementation and practical impact of routine use of ML, not as sole ML solutions but embedded in overall production systems including together with rule-based systems. We hope this special issue provides the reader with a clear perspective of ML in evolution in PV, and we expect it to herald ever more interesting, informative, and important articles on the use of ML in PV. Please see video 1 (online only) for authors' views on this theme issue of *Drug Safety*.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40264-022-01177-0>.

Declarations

Funding No funding was received to assist with the preparation of this editorial.

Conflict of interest Andrew Bate is an employee of GSK and holds stock and stock options. Yuan Luo has received grants from the National Institutes of Health (U01TR003528 and R01LM013337) and has no other disclosures to report.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Data availability Not applicable.

Code availability Not applicable.

Author contributions All authors read and approved the final version.

References

1. Bate A, Stegmann JU. Safety of medicines and vaccines: building next generation capability. *Trends Pharmacol Sci*. 2021;42(12):1051–63.
2. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;25(1):44–56.
3. Alvager T, Smith TJ, Vijai F. Neural-network applications for analysis of adverse drug reactions. *Biomed Instrum Technol*. 1993;27(5):408–11.
4. Orre R, Lansner A, Bate A, Lindquist M. Bayesian neural networks with confidence estimations applied to data mining. *Comput Stat Data Anal*. 2000;34(8):473–93.
5. Bate A, Hobbiger SF. Artificial intelligence, real-world automation and the safety of medicines. *Drug Saf*. 2021;44(2):125–32.
6. Luo Y, Thompson WK, Herr TM, Zeng Z, Berendsen MA, Jonnalagadda SR, et al. Natural language processing for EHR-based pharmacovigilance: a structured review. *Drug Saf*. 2017;40(11):1075–89.
7. Kompa B, Hakim JB, Palepu A, Kompa KG, Smith M, Pain PA, et al. Artificial intelligence based on machine learning in pharmacovigilance: a scoping review. *Drug Saf* (**in press**).
8. Aronson JK. Artificial intelligence in pharmacovigilance: terms, concepts, and limitations. *Drug Saf* (**in press**).

9. Lo A. Identifying and mitigating potential biases in predicting drug approvals. *Drug Saf* (**in press**).
10. Letinier L. Validation of an artificial intelligence pipeline to support the automatic coding of patient adverse drug reaction reports, using nationwide pharmacovigilance data. *Drug Saf* (**in press**).
11. Gattepaille LM. Automated drug coding using AI: an evaluation of WHODrug Koda on adverse event reports. *Drug Saf* (**in press**).
12. Cherkas Y. Leveraging machine learning to facilitate individual causality: assessment of adverse drug reactions. *Drug Saf* (**in press**).
13. Williams RD. Using iterative pairwise external validation to contextualize prediction model performance: a use case predicting 1-year heart-failure risk in diabetes patients across five data sources. *Drug Saf* (**in press**).
14. Manlik K. Supervised machine learning-based decision support for signal validation classification. *Drug Saf* (**in press**).
15. Ball R. "Artificial intelligence" for pharmacovigilance: ready for prime time?. *Drug Saf* (**in press**).
16. Grabowski N. Industry perspective on artificial intelligence/machine learning in pharmacovigilance. *Drug Saf* (**in press**).
17. Toh D. Successes and challenges of applying machine learning in distributed networks like sentinel. *Drug Saf* (**in press**).
18. Kjoersvik O, Bate A. Black swan events and intelligent automation for routine safety surveillance. *Drug Saf* (**in press**).
19. Bates D. Intelligent telehealth in pharmacovigilance: a future perspective. *Drug Saf* (**in press**).
20. Lianga L, Hub J, Sunc G, Hongd N, Lia Y, Haoa T, et al. AI/ML-based pharmacovigilance in the setting of limited resources. *Drug Saf* (**in press**).
21. Luo Y. Machine learning in causal inference: application in pharmacovigilance. *Drug Saf* (**in press**).
22. Metz C. In two moves, alphago and lee sedol redefined the future. *Wired*. 16-Mar-2016. Accessed 21 Feb 2022.
23. Bate A, Edwards IR. Data mining techniques in pharmacovigilance. In: Hartzema AG, Tilson HH, Chan KA, editors. *Pharmacoepidemiology and therapeutic risk management*. 3rd ed. Cincinnati: Harvey Whitney; 2008: p. 239–72.
24. Beam AL, Manrai AK, Ghassemi M. Challenges to the reproducibility of machine learning models in health care. *JAMA*. 2020;323(4):305–6.
25. Johnson AE, Pollard TJ, Shen L, Lehman LW, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. *Sci Data*. 2016;3: 160035.
26. Luo Y, Wunderink RG, Lloyd-Jones D. Proactive vs reactive machine learning in health care: lessons from the COVID-19 pandemic. *JAMA*. 2022;327(7):623–4.
27. Norén GN, Orre R, Bate A, Edwards IR. Duplicate detection in adverse drug reaction surveillance. *Data Min Knowl Discov*. 2007;14(3):305–28.
28. Noren GN, Sundberg R, Bate A, Edwards IR. A statistical methodology for drug–drug interaction surveillance. *Stat Med*. 2008;27(16):3057–70.