www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 04, Issue 12, December 2024, pp : 292-295

e-ISSN : 2583-1062
Impact Factor : 7.001

MACHINE LEARNING IN CAUSAL INFERENCE: APPLICATION IN PHARMACOVIGILANCE

Datir Sanket Dilip[1], Prof. Poonam P Khade[2], Dr. Megha T Salve[3]

[1,2,3]Department of Pharmacology, Shivajirao Pawar College of Pharmacy Pachegaon Dist – Ahmednagar, Maharashtra 422602.

Email Id: sanketdatir2003@gmail.com

DOI: https://www.doi.org/10.58257/IJPREMS37472

ABSTRACT

Pharmacovigilance is crucial for ensuring the safety of medicines. Machine learning and causal inference paradigms have been increasingly applied to pharmacovigilance to improve the detection and prediction of adverse drug events. This review aims to summarize the current state of machine learning in causal inference for pharmacovigilance. We discuss the data sources used in pharmacovigilance, including spontaneous reporting systems, real-world data, social media, and biomedical literature. We also review machine learning techniques, such as association rule mining, clustering, and neural networks, and causal inference paradigms, including propensity score matching, instrumental variable analysis, and regression discontinuity design. Finally, we highlight challenges and future directions in this field, including data quality and standardization, integration of multiple data sources, and development of more advanced machine learning algorithms. The majority of current data sources and tasks related to pharmacovigilance were not originally intended for causal inference. Pharmacovigilance has been slow to embrace integrated models that combine machine learning with causal inference. The implementation of causal frameworks has the potential to address recognized challenges associated with machine learning models, thereby improving the application of machine learning in pharmacovigilance activities

1. INTRODUCTION

Machine Learning in Causal Inference : Application in Pharmacovigilanc**e**

The World Health Organization has been advocating for pharmacovigilance initiatives to ensure the safety of pharmaceuticals through prompt and dependable information sharing concerning drug safety matters, such as adverse drug events (ADEs) [1]. An ADE refers to an unintended and harmful reaction resulting from a medication [2]. Among hospitalized patients, 16.9% experienced ADEs, with 6.7% classified as serious and 0.3% as fatal [2, 3]. Although medication errors (including incorrect or omitted doses, improper administration techniques, and equipment malfunctions) and the prescription of multiple medications are recognized as significant risk factors for ADEs [4, 5], numerous instances of ADEs still arise from undetected signals during clinical trials [3]. This issue may stem from limited sample sizes and strict patient eligibility criteria in pre-approval studies [3]. Consequently, pharmacovigilance plays a crucial role in the safe administration of medications. This review emphasizes the importance of ADE detection and monitoring tasks (including pre-clinical prediction) within the pharmacovigilance program lifecycle, as these tasks are most likely to be effectively addressed through machine learning and causal inference.

Currently, significant data sources utilized in pharmacovigilance encompass spontaneous reporting systems (SRS), real-world data (RWD) such as electronic health records (EHRs), social media platforms, biomedical literature, and various knowledge bases. Each of these data sources presents distinct advantages and inherent biases, which will be elaborated upon in the subsequent sections. Although data mining techniques have been employed to improve the efficiency of pharmacovigilance, the strength of evidence derived from identified signals is largely contingent upon the selected data source and the design of the study. In summary, we have identified three primary tasks within the domain of pharmacovigilance.

1. Extraction of drug-event pairs. This task typically involves the utilization of either structured data from EHRs or machine learning/deep learning methods based on natural language processing (NLP) to derive drug-event co-occurrence pairs from unstructured text. It is important to note that these pairs merely suggest a potential associative relationship between the drug and the event and should not be regarded as a confirmed adverse drug event (ADE). The symptoms reported may arise from various medical conditions unrelated to the ADE, necessitating further validation through additional statistical analyses or alternative data sources.

2. Detection of adverse drug events. In traditional pharmacovigilance, the primary objective is the timely detection of ADEs associated with post-marketing drugs. The goal of ADE detection is to identify and confirm these events based on real-world medication usage data as early as possible. We view ADE detection as a task that establishes a more robust associative relationship compared to disproportionality or NLP-based extraction of drug-event co-occurrence pairs.

www.ijprems.com
editor@ijprems.com
INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 04, Issue 12, December 2024, pp : 292-295
e-ISSN : 2583-1062
Impact Factor : 7.001

However, it is crucial to recognize that ADE detection remains associative without further confirmation when relying on SRS due to the limitations of this data source, which lacks a matched control group and does not allow for causality assessment. Conversely, detecting adverse drug events using an RWD database can provide more comprehensive evaluations

2. DATA SOURCES FOR PHARMACOVIGILANCE

2.1 Spontaneous Reporting System

The SRS database, including the FDA Adverse Event Reporting System (FAERS) and WHO's VigiBase, represents the most conventional dataset for the detection of adverse drug events (ADEs). Historically, the analysis of SRS data has relied on statistically based techniques, such as disproportionality measures and multivariate analyses.

In recent times, machine learning approaches, including association rule mining, clustering, graph mining, and neural networks, have also been employed to analyze SRS data. Nevertheless, these methodologies have primarily been effective in identifying 'signals of suspected causality.

2.2Real-World Data

Real-world data, which encompasses both structured and unstructured formats such as insurance claims, electronic health records (EHRs), and registry databases, presents significant opportunities for pharmacovigilance.

This data type allows for extended follow-up periods, improved assessment of exposure and outcomes, and a more comprehensive gathering of confounding variables, including comorbidities and co-prescribed medications. Additionally, it is possible to identify comparison groups within real-world data databases through matching techniques. Nonetheless, the timeliness of data collection from claims or registry databases has posed challenges. In contrast, electronic health records are regarded as a more favorable option regarding the promptness of data availability.

Traditional Causal Inference Paradigm and Integration with Machine Learning :

Pharmacovigilance studies predominantly consist of observational studies due to the characteristics of the data utilized for analysis. Nonetheless, these observational studies possess a limited capacity to establish causality, particularly regarding probabilities associated with altered conditions (adverse events) resulting from treatments or external interventions. To conduct causal inference in observational studies, either randomization or a robust study design is essential. In many instances of long-term pharmacovigilance, randomized trials are impractical, leading to a preference for observational studies in this context. However, significant challenges arise during both the design and analysis phases when attempting to derive causal conclusions from retrospective observational studies.

The foremost challenge lies in differentiating between causal and associative relationships within observational data, especially in the presence of confounders (factors that influence both the exposure and the outcome) and colliders (factors affected by both the exposure and the outcome). Although multivariable regression analysis is frequently employed to adjust for potential confounders, direct estimation of causal effects remains elusive. Additionally, it is crucial to capture and evaluate temporal relationships in observational studies prior to establishing causal relationships. Hill's criteria—comprising Strength, Consistency, Specificity, Temporality, Biological gradient, Plausibility, Coherence, Experiment, and Analogy between exposures and outcomes—are commonly cited as the foundational definition of causality in epidemiology. These criteria have informed the development of numerous causal inference models, statistical tests, and machine learning applications aimed at assessing causality.

Challenges in Machine Learning and the Importance of Causality:

Machine learning and deep learning algorithms excel at recognizing correlations; however, they fall short in establishing causation. In numerous applications, correlation may be adequate. Nonetheless, this is not applicable in the realm of pharmacovigilance and, more broadly, in healthcare.

The absence of causality assessment leads to various challenges for ML and DL algorithms, including issues related to generalizability, explainability, and fairness. In recent years, the research community in ML and DL has increasingly focused on enhancing these aspects. As previously mentioned, the integration of ML and DL with established causal inference frameworks has been pursued to improve the efficacy of traditional methods. Conversely, incorporating ML and DL within a causal inference framework can bolster the generalizability, explainability, and fairness of these models. Tackling these challenges is essential for ensuring the provision of robust evidence in pharmacovigilance, particularly if machine learning is to be utilized for signal detection.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
www.ijprems.com
editor@ijprems.com
Vol. 04, Issue 12, December 2024, pp : 292-295
e-ISSN : 2583-1062
Impact Factor : 7.001

- **Generalizability**
- **Fairnes**

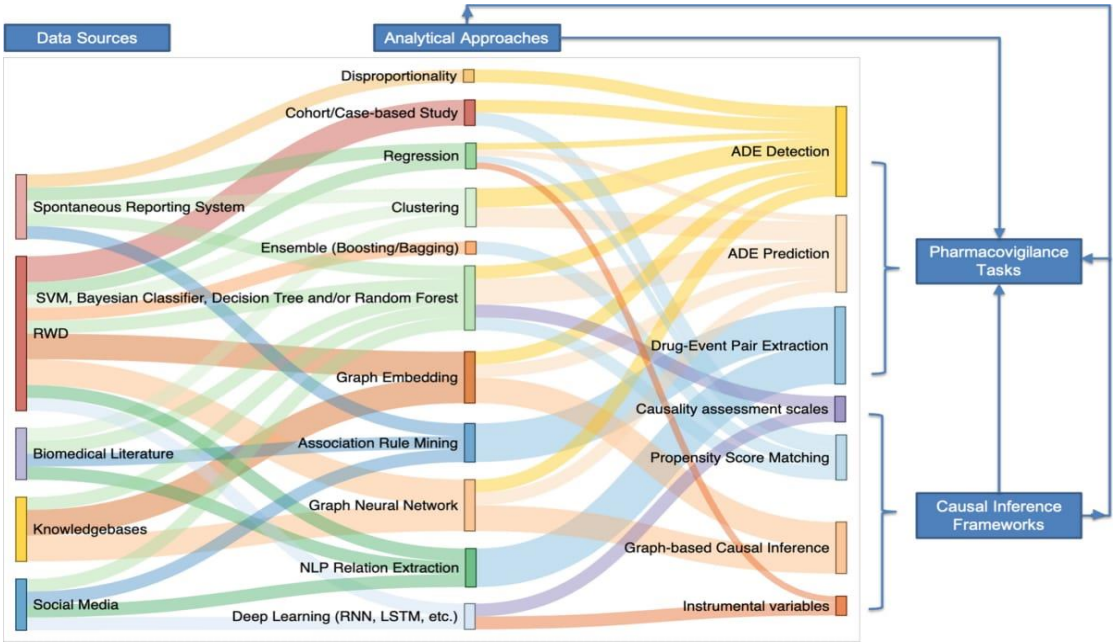

Fig.1

Fig.1 The interconnections among pharmacovigilance data sources, analytical methodologies, pharmacovigilance activities, and causal inference frameworks are significant. Each data source is typically examined using particular analytical methodologies that align with the inherent characteristics of the data it contains. Furthermore, each pharmacovigilance activity is linked to specific analytical methodologies. Causal inference frameworks are incorporated with various analytical methodologies and utilized in pharmacovigilance activities. ADE refers to adverse drug events, LSTM denotes long short-term memory, NLP stands for natural language processing, RNN indicates recurrent neural networks, RWD represents real-world data, and SVM signifies support vector machines.

3. FUTURE DIRECTION

Researchers should explore the application of machine learning to rare adverse events in pharmacovigilance.

Development of more advanced machine learning algorithms: Researchers should develop more advanced machine learning algorithms that can handle complex data structures and relationships.

Increased transparency and explainability_: Future research should focus on increasing transparency and explainability of machine learning models in pharmacovigilance.

Real-world validation of machine learning models: Researchers should conduct real-world validation of machine learning models in pharmacovigilance to ensure their accuracy and reliability.

Development of machine learning-based methods for signal detection: Future research should focus on developing machine learning-based methods for signal detection in pharmacovigilance.

Integration of machine learning with pharmacovigilance databases: Researchers should explore the integration of machine learning with pharmacovigilance databases, such as the FDA Adverse Event Reporting System.

4. CONCLUSION

This paper presents a comprehensive review of (1) data sources and tasks pertinent to pharmacovigilance, (2) traditional causal inference frameworks alongside the incorporation of machine learning into these frameworks, and (3) challenges associated with machine learning, as well as potential solutions through causal designs. Initially, it was determined that the majority of existing data sources and tasks related to pharmacovigilance were not specifically tailored for causal inference.

Additionally, the presence of low data quality significantly hindered the assessment of causal relationships. Given the critical nature of establishing causal links in pharmacovigilance, advancing data quality and representation is essential for conducting high-caliber studies in this field. Furthermore, it was noted that pharmacovigilance has been slow to embrace integrated models that combine machine learning with causal inference, indicating several missed opportunities. For instance, there is potential for further development and refinement of machine learning-based propensity score matching (PSM) and instrumental variable (IV) learning for pharmacovigilance applications. Lastly,

we acknowledged ongoing efforts to tackle the prevalent issues associated with correlation-based machine learning and deep learning models, particularly through the integration of causal paradigms. Consequently, we foresee that the pharmacovigilance sector could greatly benefit from advancements in the machine learning and deep learning domains, particularly through the synthesis of machine learning techniques with causal inference methodologies.

5. REFERANCE

[1] World Health Organization. The importance of pharmacovigi-lance. Geneva: World Health Organization; 2002.
[2] Bailey C, Peddie D, Wickham ME, Badke K, Small SS, Doyle-Waters MM, et al. Adverse drug event reporting systems: a sys-tematic review. Br J Clin Pharmacol. 2016;82(1):17–29.
[3] Lee CY, Chen Y. Machine learning on adverse drug reactions for pharmacovigilance. Drug Discov Today. 2019;24(7):1332–43
[4] Pirmohamed M, James S, Meakin S, Green C, Scott AK, Wal-ley TJ, et al. Adverse drug reactions as cause of admission to hospital: prospective analysis of 18 820 patients. BMJ.
[5] 2004;329(7456):15–9.
[6] Rothschild JM, Churchill W, Erickson A, Munz K, Schuur JD, Salzberg CA, et al. Medication errors recovered by emergency department pharmacists. Ann Emerg Med. 2010;55(6):513–21.
[7] Schachterle SE, Hurley S, Liu Q, Petronis KR, Bate A. An imple-mentation and visualization of the tree-based scan statistic for safety event monitoring in longitudinal electronic health data. Drug Saf. 2019;42(6):727–41.
[8] Kulldorf M, Dashevsky I, Avery TR, Chan AK, Davis RL, Gra-ham D, et al. Drug safety data mining with a tree-based scan statistic. Pharmacoepidemiol Drug Saf. 2013;22(5):517–23.
[9] Harpaz R, Callahan A, Tamang S, Low Y, Odgers D, Finlayson S, et al. Text mining for adverse drug events: the promise, chal-lenges, and state of the art. Drug Saf. 2014;37(10):777–90.
[10] Zeng Z, Deng Y, Li X, Naumann T, Luo Y. Natural language processing for EHR-based computational phenotyping. IEEE/ACM Transact Comput Biol Bioinform. 2018;16(1):139–53.
[11] 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
[12] US FDA. FDA Adverse Event Reporting System (FAERS);2021. https://www.fda.gov/drugs/questions-and-answers-fdas-adverse-event-reporting-systemfaers/fda-adverse-event-reporting-system-faers-public-dashboard. Accessed 20 Feb 2022.
[13] Lindquist M. VigiBase, the WHO global ICSR database system: basic facts. Drug Inf J. 2008;42(5):409–19.
[14] Harpaz R, DuMouchel W, Shah NH, Madigan D, Ryan P, Friedman C. Novel data-mining methodologies for adverse drug event discovery and analysis. Clin Pharmacol Ther. 2012;91(6):1010–21.
[15] Rouane-Hacene M, Toussaint Y, Valtchev P. Mining safety signals in spontaneous reports database using concept analysis. p. 285–94.
[16] Harpaz R, Chase HS, Friedman C. Mining multi-item drug adverse efect associations in spontaneous reporting systems. p. 1–8.
[17] Bate A, Hornbuckle K, Juhaeri J, Motsko SP, Reynolds RF. Hypothesis-free signal detection in healthcare databases: fnding its value for pharmacovigilance. London: SAGE Publications; 2019.
[18] Bate A, Hornbuckle K, Juhaeri J, Motsko SP, Reynolds RF. Hypothesis-free signal detection in healthcare databases: fnd-ing its value for pharmacovigilance. London: SAGE Publications; 2019.
[19] Chen, Y., et al. (2019). Causal inference for pharmacovigilance using machine learning and propensity scores. Proceedings of the 2019 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics,
[20] Chen, Y., et al. (2020). Machine learning-based methods for causal inference in pharmacovigilance. Journal of Biomedical Informatics