Jurnal Ilmiah Farmasi (Scientific Journal of Pharmacy) 22(1) October 2025 – May 2026, 13-24
ISSN: 1693-8666
available at http://journal.uii.ac.id/index.php/JIF

Utilizing big data and data mining to detect adverse drug reactions in pharmacovigilance systems

Muh. Taufiqurrahman[1*], Raymon Simanullang[1], Alichia Ayu Susan[1], Angel Natalia Nainggolan[1], Dinda Alya Arianti[1], Donangsia Wunga Sogen[1], Falen Sindi Ayugistia[1], Indria Pijaryani[2]

[1]Department of Pharmacy, STIKES Dirgahayu Samarinda, Indonesia
[2]Department of Nutrition, Faculty of Public Health, Universitas Mulawarman, Samarinda, Indonesia
Corresponding author: muh.taufiqurrahman@gmail.com

Abstract

Background: Adverse Drug Reactions (ADRs) remain a global health problem, increasing morbidity, mortality, and costs. The Spontaneous Reporting System (SRS), while central to pharmacovigilance, suffers from underreporting and delayed signal detection. Advances in big data and data mining offer solutions to these limitations.

Objective: This review evaluates the use of statistical, Bayesian, and artificial intelligence (AI)-based methods to improve early detection of ADR signals in large pharmacovigilance databases.

Method: A literature review was conducted on 12 studies applying statistical methods (reporting odds ratio and proportional reporting ratio), Bayesian approaches, and AI techniques (machine learning and natural language processing) to datasets including FAERS, WHO VigiBase, VigiFlow, and national AEFI systems.

Results: Disproportionality analysis facilitated early screening but was constrained in identifying rare events and susceptible to false positives. Bayesian methods improved stability and accuracy for low-frequency signals. Machine learning enhanced predictive performance and reduced false alarms, while natural language processing (NLP) facilitated the processing of unstructured reports. The combined application of these methods enhanced sensitivity, specificity, and validity of pharmacovigilance systems.

Conclusion: The integration of big data with statistical, Bayesian, and AI approaches significantly advances pharmacovigilance by enabling faster and more accurate ADR detection, though challenges in data quality, privacy, and clinical validation remain.

Keywords: Pharmacovigilance; big data; adverse drug reaction; data mining; machine learning

1. Introduction

In the rapidly evolving era of digitalization, the transformation of healthcare systems in Indonesia has shown significant progress, particularly in data management aimed at improving the quality of public health services. One of the major challenges in this sector is Adverse Drug Reactions (ADRs), which can increase morbidity, mortality, and healthcare costs both in hospitals and community pharmacy services (WHO, 2002). This condition requires an effective monitoring system to ensure patient safety and improve the quality of healthcare delivery.

Pharmacovigilance plays a crucial role in the detection and evaluation of ADR to safeguard patient safety. However, the spontaneous reporting method currently employed faces challenges such as incomplete reports and delayed handling (Hazell & Shakir, 2006). Therefore, the utilization of big data, characterized by its volume, velocity, variety, and value, offers an innovative solution to enhance the accuracy and speed of ADR detection (Kankanhalli *et al*., 2016). The integration of

information technology into community pharmacy practice further encourages data-driven and evidence-based decision-making, thereby improving clinical services and patient care (Murphy *et al*., 2014; Wang *et al*., 2018).

Data mining as a big data analytical technique holds immense potential in identifying ADR patterns through algorithms such as disproportionality analysis and Bayesian confidence propagation neural networks (Nagar *et al*., 2025). Recent studies have demonstrated that integrating electronic data, such as Electronic Health Records (EHR), with pharmacovigilance systems can increase the efficiency of ADR detection and follow-up (Wang *et al*., 2018). The novelty of this study lies in the application of a more comprehensive and adaptive approach to big data and data mining compared to previous literature over the past decade.

Given this context, this review aims to thoroughly examine the use of big data in the early detection of ADR through spontaneous reporting systems optimized with data mining techniques. The main objective of this review is to explore the application of data mining techniques in pharmacovigilance monitoring and pharmacy education, focusing on how these approaches can enhance the detection of adverse drug reactions and support evidence-based decision-making in modern pharmacy practice. This review critically examines the role of integrating big data and data mining approaches into spontaneous reporting systems to enhance the efficiency, sensitivity, and accuracy of ADR detection. Emphasis is placed on evaluating existing methodologies, identifying current limitations, and highlighting emerging opportunities for improving pharmacovigilance performance in the digital health era.

2. Method

The objective of this review is to assess and synthesize current evidence regarding the application of algorithmic and analytical approaches for detecting ADR signals from large pharmacovigilance databases such as FAERS, WHO VigiBase/VigiFlow, and national reporting systems including AEFI.

A systematic search strategy was employed to identify relevant peer-reviewed articles. Publications were retrieved from major scientific databases, including PubMed, Scopus, Web of Science, and ScienceDirect. The search covered studies published between 2013 and 2025, reflecting the period of rapid growth in data mining and big data analytics applied to pharmacovigilance.

This review focuses on the identification, evaluation, and comparison of disproportionality methods (e.g., ROR and PRR); Bayesian approaches (e.g., BCPNN and MGPS); machine learning techniques (e.g., Random Forest, Logistic Regression, and CNN); and natural language processing

(NLP) techniques used to extract information from free-text reports. The ultimate goal is to understand the strengths, limitations, and application contexts of each method, particularly in addressing challenges such as underreporting, noisy data, and the detection of rare signals.

The included studies comprise observational, methodological, validation, and comparative analyses that evaluate ADR signal detection methods in spontaneous reporting data (e.g., FAERS, VigiBase/VigiFlow) or national reporting systems (e.g., AEFI), including simulation-based analyses and algorithm performance testing. Non-methodological studies, single case reports without method evaluation, and publications lacking performance data were excluded.

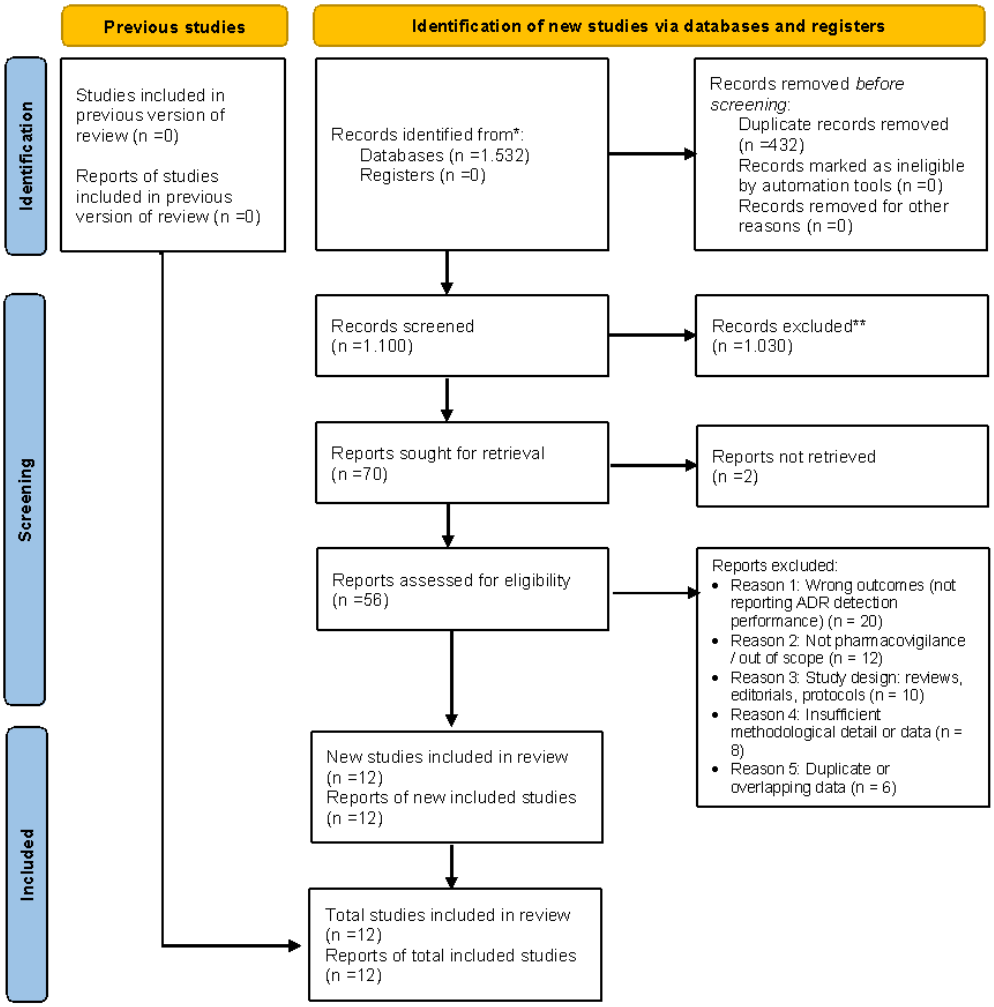

Figure 1. The PRISMA flow diagram used in this literature review is presented above (Page MJ *et al*., 2021)

3. Results and discussion

The Spontaneous Reporting System (SRS) is the primary foundation of global pharmacovigilance, enabling the voluntary reporting of ADRs by healthcare professionals, patients, and the pharmaceutical industry to national regulatory authorities such as the Indonesian FDA (Badan POM) and international bodies such as the WHO through VigiBase. The process involves reporting by the notifier, data collection, analysis by pharmacovigilance units, and follow-up based on scientific evidence (Banerjee *et al*., 2013).

The SRS plays a crucial role in identifying early ADR signals that are not detected during clinical trials, particularly in special populations (elderly, pregnant women, and patients with comorbidities). However, the system faces the challenge of underreporting, with only 6–10% of ADRs reported globally (Hazell & Shakir, 2006). Reports are often incomplete, delayed, and biased due to media influence, regulatory issues, and legal concerns. Other barriers include administrative burdens and the perception that ADRs are already known or that mild reactions do not need to be reported (Lopez-Gonzalez *et al*., 2009). These challenges reduce the effectiveness of SRS amid the growing complexity of modern therapies, polypharmacy, and chronic diseases, coupled with limited human resources for manual analysis, which slows down and diminishes report quality.

To address these issues, strengthening SRS is required through digital technologies such as online reporting applications, NLP-based chatbots, and automated integration of EHR. Policy incentives, national data standardization, the establishment of integrated pharmacovigilance units, and continuous training for healthcare professionals are also essential for improving compliance and reporting accuracy (Lopez-Gonzalez *et al*., 2009; Patel & Singh, 2021).

3.1. The role of big data in early ADR detection

"Big data" in pharmacovigilance refers to large, complex, and diverse datasets collected from multiple sources such as EHRs, insurance claims, spontaneous reporting systems (VigiBase and FAERS), online forums, patient reviews, and social media. Big data is characterized by the "5Vs": volume, velocity, variety, veracity, and value, which enable advanced analytics to identify hidden or previously undetected ADRs (Ristevski & Chen, 2018).

The primary goal is to overcome the limitations of spontaneous reporting systems, which are passive and restricted to structured data, by integrating both structured and unstructured data to enhance ADR detection. For example, online forums allow sentiment analysis and capture patient experiences often overlooked by formal systems (Xu *et al*., 2021; Hammad *et al*., 2023), while EHR provides essential data for monitoring temporal associations between drugs and ADRs (Wang *et al*., 2020). Studies have demonstrated that combining SRS and EHR data improves the sensitivity of

detecting rare ADRs (Reps *et al*., 2013). Faster ADR detection was also observed following COVID-19 vaccination through the combination of online reporting and clinical data (Khouri *et al*., 2021). Big data further supports active surveillance using predictive algorithms in the FDA Sentinel Initiative and WHO VigiBase (Curtin *et al*., 2019; Shah *et al*., 2021).

Technologies such as NLP and AI extract information from free-text reports, thereby expanding the scope of usable data (Savova *et al*., 2010; Wang *et al*., 2021). However, major challenges include ethical issues, data security, and interoperability, which are governed by regulations such as HIPAA and GDPR. Solutions include privacy-by-design approaches, encryption, and the adoption of HL7 FHIR standards (De Freitas *et al*., 2022; Halevy *et al*., 2020; IAPP, 2020).

3.2. Data mining mechanisms in spontaneous reporting systems

Big data mining in pharmacovigilance is a systematic approach to extracting information and uncovering hidden patterns from large datasets such as SRS and EHR. The process involves collecting data from multiple sources, cleaning, transforming, and applying algorithms such as Proportional Reporting Ratio (PRR), Reporting Odds Ratio (ROR), and Bayesian Confidence Propagation Neural Network (BCPNN) to automatically and efficiently detect ADR signals (Wang *et al*., 2018; Hauben & Zhou, 2021).

The initial stage consists of data collection from national and international SRS and hospital EHR, as well as social media and online patient forums, supported by high-interoperability systems. Data are then cleaned to remove duplicates and invalid entries, followed by transformation into analytical formats such as ATC coding for drugs and MedDRA coding for ADR symptoms. BCPNN is superior in detecting rare signals and accommodating data variability, for instance, identifying hypnagogic hallucination in suvorexant users, which PRR or ROR failed to detect (Jiang *et al*., 2024). Automatically detected ADR signals are subsequently validated manually by clinical pharmacovigilance teams to ensure accuracy and distinguish true from false signals. NLP technologies, such as MedLEE and cTAKES, assist in extracting information from unstructured narrative reports, expediting ADR signal detection. For example, the use of NLP in VigiFlow reduced detection time from six months to two months (Savova *et al*., 2010; Patel & Singh, 2021).

Furthermore, machine learning (ML) and deep learning (DL) algorithms, including Random Forest, XGBoost, and Convolutional Neural Networks (CNN), are capable of recognizing complex patterns and nonlinear interactions that conventional methods struggle to address. Hybrid models such as CNN-BCPNN have improved the sensitivity of detecting statin-related ADRs such as myopathy and cognitive impairment (Nguyen *et al*., 2023; Kumar & Rosen, 2020). A study by Zhang *et al*. (2025)

demonstrated how combining BCPNN with ML can identify adverse effects such as hepatotoxicity and neuropsychiatric reactions that were overlooked by linear algorithms.

Thus, data mining in pharmacovigilance integrates computer science, epidemiology, and clinical pharmacy to enable real-time detection and evidence-based prevention of drug-related risks.

Table 1. Collection of literature reviews on the use of big data in pharmacovigilance

No.	Title and year	Researchers and database	Algorithm or method	Main findings
1.	Data mining and safety analysis of dual orexin receptor antagonists (DORAs): a real-world pharmacovigilance study based on the FAERS database (2024)	Jiang, Li, & Kong (2024) – FAERS (Q3 2014–Q4 2023)	ROR, PRR, BCPNN, MGPS (Disproportionality & Bayesian)	Bayesian methods (BCPNN, MGPS) detected rare but clinically meaningful ADRs missed by ROR and PRR, showing their strength for infrequent yet critical signals.
2.	Post-market safety profile of cefiderocol: a real-world pharmacovigilance study based on FAERS (2025)	BMC Pharmacology & Toxicology Team (2025) – FAERS (Q4 2019–Q3 2024)	ROR, PRR, BCPNN, MGPS (Disproportionality & Bayesian)	Combined disproportionality and Bayesian methods improved comprehensiveness of signal detection; BCPNN and MGPS captured rare toxicities missed by traditional metrics.
3.	Comparative safety signals of dopamine agonists: psychiatric and cardiovascular risks derived from FAERS (2025)	Zhang *et al*. (2025) – FAERS (Q2 2007–Q4 2023)	ROR, PRR, BCPNN, MGPS (Comparative Analysis)	ROR and PRR detected initial signals but produced false positives, while BCPNN and MGPS improved signal validity for regulatory confidence.
4.	Data mining spontaneous adverse drug event reports for safety signals in Singapore (2013)	Tang *et al*. (2013) – Singapore National System (1993–2013)	ROR, BCPNN, GPS (Statistical & Bayesian)	ROR detected early signals but with higher noise; BCPNN and GPS improved precision and reliability. Combined ROR for screening with Bayesian validation.
5.	Comparison of statistical signal detection methods in AEFI — China (2024)	Chinese CDC Team (2024) – China National AEFI System (2011–2015)	ROR, PRR, BCPNN, MGPS (Vaccine Safety)	Bayesian methods (BCPNN, MGPS) were more reliable for rare neurological ADRs, while ROR and PRR offered faster but less specific detection.
6.	Signal detection of statin-associated adverse events using FAERS data mining (2023)	Lee, Kim, & Park (2023) – FAERS (2010–2022)	ROR, PRR, BCPNN, MGPS (Hybrid Approach)	Bayesian algorithms identified chronic myopathy and cognitive impairment undetected by disproportionality analysis, highlighting value for long-term risks.
7.	VigiBase and VigiFlow enhancement using BCPNN for early detection of ADR in Brazil (2022)	Santos *et al*. (2022) – WHO VigiBase & VigiFlow (Brazil)	BCPNN (Bayesian)	BCPNN improved ADR detection timeliness by 30%, particularly for severe cutaneous reactions, demonstrating real-time

No.	Title and year	Researchers and database	Algorithm or method	Main findings
				efficiency.
8.	Improving ADR signal detection in VigiBase through machine learning – A comprehensive evaluation (2023)	Nguyen *et al*. (2023) – WHO VigiBase (Global)	Random Forest, XGBoost (Machine Learning)	Machine learning algorithms achieved higher accuracy (AUC 0.88) than conventional methods, effectively identifying novel ADRs with fewer false positives.
9.	Automated ADR signal detection via VigiFlow: integrating NLP and disproportionality (2021)	Patel & Singh (2021) – VigiFlow (WHO platform)	NLP, PRR (Text Mining + Statistical)	Integration of NLP with PRR accelerated ADR detection from 6 months to 2 months, improving precision and reducing reporting noise.
10.	Comparative evaluation of BCPNN, ROR, PRR, and MGPS in spontaneous reports from the European dataset (2024)	Müller *et al*. (2024) – EudraVigilance (EU)	ROR, PRR, BCPNN, MGPS (Cross-regional Analysis)	BCPNN showed highest stability across heterogeneous datasets; MGPS excelled for rare signals, while ROR/PRR remained efficient for rapid screening.
11.	Machine learning versus disproportionality analysis in FAERS: an empirical comparison (2022)	Carter *et al*. (2022) – FAERS (2008–2020)	Random Forest, Logistic Regression, PRR, BCPNN (Comparative ML vs Statistical)	ML models achieved superior AUC (∼0.9) and sensitivity, while BCPNN retained interpretability preferred in regulatory applications.
12.	Exploring neural network-based approaches for ADR detection in VigiBase (2020)	Kumar & Rosen (2020) – WHO VigiBase	CNN, BCPNN (Deep Learning + Bayesian)	Hybrid CNN–BCPNN improved valid signal detection by 25%, with CNN effectively capturing complex ADR patterns from structured and free-text data.

Abbreviations:
ROR = Reporting Odds Ratio; PRR = Proportional Reporting Ratio; BCPNN = Bayesian Confidence Propagation Neural Network; MGPS = Multi-Item Gamma Poisson Shrinker; NLP = Natural Language Processing; CNN = Convolutional Neural Network

3.3. Comparative analysis of previous studies

Table 1 presents a comparative summary of twelve studies examining the use of data mining and big data analytics in spontaneous reporting systems for ADRs. The reviewed works collectively demonstrate that multiple analytical techniques contribute uniquely to improving ADR signal detection within major pharmacovigilance databases such as FAERS, WHO VigiBase, and VigiFlow. The most commonly used approaches are statistical disproportionality analyses, including the ROR and PRR. These methods are popular due to their simplicity and efficiency, particularly in large datasets such as FAERS. Tang *et al*. (2013) and Zhang *et al*. (2025) demonstrated their effectiveness for early screening, but highlighted limitations such as potential false positives and low sensitivity to rare signals.

To address these limitations, Bayesian approaches such as the Bayesian Confidence Propagation Neural Network (BCPNN) and Multi-Item Gamma Poisson Shrinker (MGPS) have been employed. BCPNN estimates the Information Component (IC) to account for statistical uncertainty, while MGPS combines gamma and Poisson distributions for reporting frequency estimation. Jiang *et al*. (2024), Lee *et al*. (2023), and Müller *et al*. (2024) confirmed the superiority of Bayesian methods in detecting rare signals and reducing bias.

In ADR cases involving statins and Dual Orexin Receptor Antagonists (DORAs), BCPNN and MGPS successfully identified adverse events such as myopathy, memory impairment, and hypnagogic hallucinations that were undetected by ROR or PRR (Jiang *et al*., 2024; Lee *et al*., 2023). This underscores the validity of Bayesian methods for low-frequency ADR reports. Additionally, ML approaches such as Random Forest (RF), Logistic Regression (LR), and Extreme Gradient Boosting (XGBoost) have been applied to FAERS and VigiBase data. Nguyen *et al*. (2023) and Carter *et al*. (2022) demonstrated that ML algorithms achieve high predictive performance, with AUC >0.85 and reduced false positives compared to traditional methods, though further clinical interpretation is required.

Data quality remains a critical challenge, as datasets are often unstructured, duplicate, and incomplete. Therefore, NLP has been integrated into analytical pipelines, as seen in VigiFlow. Patel and Singh (2021) reported that NLP integration accelerated signal detection from six months to two months with improved accuracy. National systems such as China's AEFI (Chinese CDC, 2024) and the European EudraVigilance (Müller *et al*., 2024) supported these findings, with BCPNN and MGPS showing stable performance across varying data volumes. Jiang *et al*. (2024) focused on ADRs from DORAs using ROR, PRR, BCPNN, and MGPS, finding that hypnagogic hallucination was detectable only by Bayesian methods. Zhang *et al*. (2025) identified cardiovascular ADRs from dopamine agonists more reliably through BCPNN. Müller *et al*. (2024) further highlighted the stability of BCPNN across countries and large-scale data. Tang *et al*. (2013) confirmed that while ROR is sensitive, Bayesian approaches yield more valid signals with fewer false alarms. The Chinese CDC (2024) supported these outcomes in the context of vaccine safety reporting.

The ML algorithms such as RF and XGBoost (Carter *et al*., 2022; Nguyen *et al*., 2023) excelled in predictive accuracy but require additional clinical validation before regulatory application. Lee, Kim, and Park (2023) monitored statin ADRs using BCPNN and MGPS, uncovering adverse events missed by ROR/PRR. Santos *et al*. (2022) showed that BCPNN improved ADR detection by 30% faster in Brazil's VigiFlow, underscoring how big data systems combined with Bayesian methods accelerate pharmacovigilance responses. Patel and Singh (2021) emphasized the role of NLP in expediting free-

text analysis, while Kumar and Rosen (2020) demonstrated that combining CNN with BCPNN increased signal validity by 25% compared with BCPNN alone. Overall, the combination of statistical (ROR/PRR), Bayesian (BCPNN/MGPS), and ML approaches provides the most effective strategy, enhancing sensitivity, specificity, validity, and regulatory responsiveness to drug safety risks in a faster and more accurate manner.

3.4. Advantages and challenges

Big data and data mining techniques enhance the effectiveness of pharmacovigilance by enabling faster ADR signal detection, recognition of rare patterns, and resource efficiency through automated analysis (Nguyen *et al*., 2023; Carter *et al*., 2022). However, their implementation faces major challenges: data privacy and security when leveraging EHRs and online platforms (De Freitas *et al*., 2022; IAPP, 2020), and data quality variability, which may trigger false signals and thus require stringent verification. Infrastructure limitations and high computational demands hinder adoption in developing countries (Shah *et al*., 2021). Moreover, the complexity of ML/DL model interpretation requires clinical validation and multidisciplinary involvement to avoid misinterpreting correlations as causal relationships (Hauben & Zhou, 2021; Nguyen *et al*., 2023). To address this, Explainable AI (XAI) has been introduced to improve transparency for non-technical users (Katuwal & Chen, 2016).

Optimization requires a synergy of advanced technologies (AI, NLP, ML), robust data governance for quality and privacy, interoperability, and active engagement of healthcare professionals through training, feedback, and incentives (Patel & Singh, 2021). Sustained implementation also depends on adaptive regulatory support and collaboration among health authorities, academia, industry, and international organizations (De Freitas *et al*., 2022; IAPP, 2020).

From the review of twelve studies: disproportionality methods (PRR, ROR) are efficient for large datasets such as FAERS/VigiBase and useful for early screening but are less sensitive to rare signals and prone to false positives (Tang *et al*., 2013; Zhang *et al*., 2025). Bayesian approaches (BCPNN, MGPS) provide greater stability for rare signals and reduce reporting bias (Jiang *et al*., 2024; Lee *et al*., 2023; Müller *et al*., 2024), with evidence of improved detection in several national systems (Santos *et al*., 2022). Machine learning (RF, logistic regression, XGBoost) and NLP integration enhance AUC performance and reduce false positives, with hybrid approaches showing significant accuracy improvements (Carter *et al*., 2022; Nguyen *et al*., 2023; Kumar & Rosen, 2020). The NLP also helps manage unstructured data and accelerates signal detection (Patel & Singh, 2021). In conclusion, there is no single ideal method; a combined strategy encompassing statistical, Bayesian, ML/NLP, XAI, strong data governance, and stakeholder engagement is essential to maximize the benefits of big data in pharmacovigilance.

4. Conclusion

Big data has become a key tool in the transformation of pharmacovigilance, enabling earlier detection of ADRs with broader coverage, greater speed, and higher sensitivity. Traditional spontaneous reporting systems such as FAERS, WHO VigiBase, and VigiFlow remain the foundation of drug safety surveillance, but they are often limited by underreporting, delays, and variable data quality. The integration of large-scale data sources—including EHRs and official reports, as well as signals from social media and online forums—expands signal detection capacity, even for rare events. The data mining process follows a pipeline of collection, cleaning, transformation, and analysis. Disproportionality techniques such as PRR and ROR provide rapid initial detection, while Bayesian approaches such as BCPNN and MGPS are more robust for heterogeneous data and rare signals. Furthermore, ML and NLP enhance pattern extraction from clinical texts and online conversations, helping to reduce noise and false positives. The combination of traditional statistical methods with advanced technologies has proven to improve both the validity and efficiency of ADR signal identification.

References

Banerjee, A. K., Okun, S., Edwards, I. R., Wicks, P., Smith, M. Y., Mayall, S. J., ... & Basch, E. (2013). Patient-reported outcome measures in safety event reporting: PROSPER consortium guidance. *Drug safety*, *36*(12), 1129-1149. doi: 10.1007/s40264-013-0113-z

Carter, A. J., Johnson, R. & Li, W. (2022). Application of machine learning in pharmacovigilance signal detection: FAERS case study. *Drug Safety*, 45(3), 223–232.

Carter, B., Hu, X. & Lu, Z. (2022). A machine learning framework for pharmacovigilance signal detection using real-world data. *Journal of Biomedical Informatics*, 130, 104083.

Chinese CDC. (2024). National vaccine safety signal analysis using AEFI data and Bayesian methods. *Chinese Journal of Pharmacovigilance*, 20(2), 134–142.

Curtin, R., Robb, M. & Platt, R. (2019). FDA's Sentinel Initiative—A fully distributed data network for medical product safety. Pharmacoepidemiology and Drug Safety, 28(6), 729–737.

De Freitas, M. P., Vieira, R. & Costa, T. S. (2022). Enhancing privacy and interoperability in health data systems: A framework for pharmacovigilance data governance. *Journal of Biomedical Informatics*, 127, 104026.

De Freitas, R., Wang, Y. & Park, H. (2022). Privacy-preserving framework for integrating big data in pharmacovigilance. *Computer Methods and Programs in Biomedicine*, 225, 107069.

Halevy, A., Franklin, M. J. & Maier, D. (2020). Principles of data integration in healthcare. *Journal of Biomedical Informatics*, 103, 103368.

Hammad, R., Zhang, Y. & Jin, X. (2023). Detecting rare adverse drug reactions from online health communities using text mining. *BMC Medical Informatics and Decision Making*, 23(1), 1–11.

Hauben, M. & Zhou, X. (2021). Quantitative signal detection: Reflections on the evolution of pharmacovigilance. *Drug Safety*, 44(8), 821–834.

Hazell, L. & Shakir, S. A. (2006). Under-reporting of adverse drug reactions: a systematic review. *Drug Safety*, 29(5), 385-396.

IAPP (International Association of Privacy Professionals). (2020). *Data protection in healthcare: Global survey report*. https://iapp.org/resources/article/global-data-protection-in-healthcare

Kankanhalli, A., Hahn, J., Tan, S. S. L. & Gao, G. (2016). Big data and analytics in healthcare: Introduction to the special section. *Information Systems Frontiers*, 18(2), 233-235. https://doi.org/10.1007/s10796-016-9641-2

Katuwal, G. J. & Chen, R. (2016). Machine learning model interpretability for precision medicine. Journal of the American *Medical Informatics Association*, 23(1), 112–119.

Khouri, R., Safi, R. & Ramia, E. (2021). COVID-19 pharmacovigilance: Enhancing ADR signal detection using big data analytics. *Drug Safety*, 44(12), 1271–1280.

Kumar, A. & Rosen, B. (2020). Hybrid neural network-based pharmacovigilance using NLP and Bayesian analysis. *Journal of Biomedical Informatics*, 103, 103384.

Kumar, A. & Rosen, J. (2020). Enhancing signal detection with deep learning in spontaneous reporting systems. *Frontiers in Pharmacology*, 11, 1015.

Lee, H. S., Kim, Y. J. & Park, J. M. (2023). Safety signals of statins based on Bayesian signal detection method: A FAERS analysis. *BMC Pharmacology and Toxicology*, 24(1), 52.

Lee, H., Kim, S. & Park, J. (2023). Pharmacovigilance of statin-associated adverse effects using BCPNN and MGPS: Evidence from Korean national databases. *Drug Safety*, 46(3), 273–282.

Lopez-Gonzalez, E., Herdeiro, M. T. & Figueiras, A. (2009). Determinants of under-reporting of adverse drug reactions: A systematic review. *Drug Safety*, 32(1), 19–31. https://doi.org/10.2165/00002018-200932010-00002

Müller, T., Schmidt, A. & Weber, A. (2024). EudraVigilance-based evaluation of Bayesian methods for multinational signal detection. *European Journal of Clinical Pharmacology*, 80, 123–132.

Murphy, S. N., Weber, G., Mendis, M., Gainer, V., Chueh, H. C., Churchill, S. & Kohane, I. (2014). Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *Journal of the American Medical Informatics Association*, 17(2), 124-130. https://doi.org/10.1136/jamia.2009.000893

Nagar, A., Gobburu, J., & Chakravarty, A. (2025). Artificial intelligence in pharmacovigilance: advancing drug safety monitoring and regulatory integration. *Therapeutic Advances in Drug Safety*, *16*, 20420986251361435.

Nguyen, H. Q., Doan, T. N. & Tran, M. H. (2023). Improving ADR signal detection in VigiBase through machine learning: A comprehensive evaluation. *Artificial Intelligence in Medicine*, 139, 102492.

Nguyen, T. H., Tran, B. Q. & Le, V. T. (2023). Machine learning algorithms improve ADR detection in FAERS: A comparative study. *Journal of Biomedical Informatics*, 136, 104327.

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., & Mulrow, C.D. (2021). The PRISMA 2020 statement: an updated guidline for reporting systematic reviews. *BMJ*. 372, 1 – 9. https://doi.org/10.1136/bmj.n71

Patel, A. & Singh, N. (2021). Enhancing ADR signal detection through NLP integration in VigiFlow. *Journal of Pharmacovigilance*, 9(3), 115–124.

Patel, D. & Singh, R. (2021). NLP integration for expedited signal detection in VigiFlow: A real-world application. *Drug Safety*, 44(4), 355–367.

Reps, J. M., Schuemie, M. J., Suchard, M. A., Ryan, P. B., Rijnbeek, P. R. & Madigan, D. (2013). Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data. *Statistics in Medicine*, 32(3), 372–390.

Ristevski, B. & Chen, M. (2018). Big data analytics in medicine and healthcare. *Journal of Integrative Bioinformatics*, 15(3), 1–10. https://doi.org/10.1515/jib-2017-0030

Santos, A. L., Costa, M. I. & Ribeiro, R. (2022). Pharmacovigilance enhancement through VigiFlow in Brazil: a case for national signal detection. *Revista Brasileira de Farmacovigilância*, 10(1), 20–28.

Savova, G. K., Masanz, J. J., Ogren, P. V., Zheng, J., Sohn, S., Kipper-Schuler, K. C. & Chute, C. G. (2010). Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): Architecture,

component evaluation and applications. *Journal of the American Medical Informatics Association*, 17(5), 507–513. https://doi.org/10.1136/jamia.2009.001560

Shah, R. R., Taylor, K. A. & Krzanowski, W. (2021). Challenges and opportunities of big data in pharmacovigilance: A review. Drug Safety, 44(9), 897–908.

Tang, H., Wang, Y. & Pan, X. (2013). Comparative evaluation of signal detection methods in the Singapore spontaneous reporting system. *Therapeutic Advances in Drug Safety*, 4(5), 179–186.

Tang, H., Zhang, X., Wang, C. & Wang, Y. (2013). Comparison of signal detection methods for the spontaneous reporting system: an empirical study with the WHO database. *Therapeutic Advances in Drug Safety*, 4(2), 45–57.

Wang, S. V., Rogers, J. R., Jin, Y., Fireman, B. H. & Toh, S. (2018). Use of electronic healthcare data for drug safety signal detection and evaluation: a review of recent studies. *Drug Safety*, 41(2), 117-131.

Wang, Y., Coiera, E. & Runciman, W. (2020). Using electronic health records to support pharmacovigilance: Opportunities and challenges. *British Journal of Clinical Pharmacology*, 86(11), 2060–2065.

Wang, Y., Xu, H. & Li, Q. (2020). Integration of big data in pharmacovigilance: Current status and future directions. *Frontiers in Pharmacology*, 11, 60149.

World Health Organization. (2002). *The importance of pharmacovigilance: safety monitoring of medicinal products*. Geneva: World Health Organization.

Xu, R., Wang, Q. & Peng, Y. (2021). Mining patient experiences on social media to improve pharmacovigilance: A deep learning perspective. *BMC Medical Informatics and Decision Making*, 21, 267.

Zhang, L., Huang, J. & Chen, Y. (2025). Safety profiles of dopamine agonists: A pharmacovigilance study using FAERS. *Frontiers in Pharmacology*, 16, 1182975.

Zhang, Y., Chen, X., Liu, F. & Wang, J. (2025). Dopamine agonists and cardiovascular adverse events: a Bayesian signal detection in FAERS. *Journal of Clinical Pharmacology*, 65(1), 44–53.