

# Machine Learning Approaches for Medication Safety Analysis

## Introduction

Ensuring the safe use of medicines is a core component of healthcare. Adverse drug events (ADEs), adverse drug reactions (ADRs), medication errors, and harmful drug–drug interactions (DDIs) remain major causes of preventable morbidity, mortality, and costs worldwide. The World Health Organization identifies medication-related harm as a leading safety concern, estimating that medication errors alone cost health systems billions annually (World Health Organization, 2024). In hospitalised patients, ADEs are associated with longer stays, increased readmissions, and higher mortality (Patel *et al.*, 2023).

Reducing medication-related harm has therefore become a global priority. Traditional pharmacovigilance systems (e.g. spontaneous reporting schemes, manual chart reviews, and audits) have enabled detection of important and yet undetected safety concerns. Nonetheless, these systems face enduring limitations: they are passive, resource-intensive, and retrospective, and underreporting is common (Bate and Evans, 2009; Kim *et al.*, 2022). The expansion of digital health infrastructures has created opportunities to address these shortcomings. Electronic health records (EHRs), claims databases, registries, and social media now provide large-scale, real-world data (RWD), but these data are heterogeneous, high-dimensional, and often unstructured, posing challenges for traditional approaches (Wong *et al.*, 2022).

Machine learning (ML), a subset of artificial intelligence (AI), offers new opportunities. By learning patterns from data, ML methods can enhance detection, prediction, and prevention of medication-related harms (Jordan and Mitchell, 2015). They differ from hypothesis-driven, rule-based approaches by enabling discovery of novel associations, accommodating complex data, and providing predictive outputs in real-time (Deimazar and Sheikhtaheri, 2023). This essay reviews the current state of ML in medication safety analysis. It considers methodological foundations, explores applications, and critically evaluates opportunities and challenges, highlighting the promise of ML to complement traditional pharmacovigilance while recognising the need for robust validation, interpretability, and regulatory oversight.

## Background

Pharmacovigilance systems have historically relied on spontaneous reporting systems (SRS), such as the FDA's Adverse Event Reporting System, and WHO's VigiBase. These systems have successfully identified safety concerns that were not apparent in clinical trials, such as rhabdomyolysis risks associated with cerivastatin (Bate and Evans, 2009). However, it is estimated that fewer than 10% of ADRs are reported (Alomar *et al.*, 2020). Moreover, SRS are prone to reporting biases, documentation inconsistencies, and delayed signal recognition (Noguchi, Tachi and Teramachi, 2021; Kim *et al.*, 2022). In turn, manual chart reviews and institutional reporting mechanisms, while valuable, are labour-intensive, retrospective, and have limited scope (Deimazar and Sheikhtaheri, 2023).

The digital healthcare transformation has expanded the data available for safety monitoring. EHRs capture longitudinal data on prescribing, diagnoses, laboratory results, and outcomes, while unstructured clinical narratives contain nuanced information not coded in structured fields (Henry *et al.*, 2019; Wong *et al.*, 2022). Administrative claims and registries provide population-level perspectives, while patient-generated data from social media and online communities contribute insights into real-world experiences (Xie, Zeng and Marcum, 2017; Brown *et al.*, 2022; Golder *et al.*, 2024). These developments promise to improve the timeliness and breadth of safety surveillance, but they also create analytic challenges due to data scale, heterogeneity, and complexity.

Regulatory agencies have begun to adapt. For example, the FDA's Sentinel Initiative has enabled active surveillance of medical products since 2016 through distributed observational data networks (Brown *et al.*, 2022). In Europe, the DARWIN EU project is being developed to integrate RWD into regulatory decision-making (Raventos and Prieto-Alhambra, 2025). Both industry and regulatory agencies are also increasingly considering ML, with the FDA and EMA emphasising the importance of transparency, reproducibility, and validation in their adoption (Kassekert *et al.*, 2022; Ball *et al.*, 2024; European Medicines Agency, 2024a, 2024b). This evolving environment provides fertile ground for the integration of ML into medication safety frameworks.

## Machine learning approaches

ML refers to algorithms that learn from data to make predictions or identify patterns without being explicitly programmed (Jordan and Mitchell, 2015; Russell and Norvig, 2021). Unlike traditional statistical methods, which rely on predefined hypotheses and linear associations, ML adapts to complex, non-linear relationships and uncovers hidden structures in large, heterogeneous datasets. In medication safety, this enables integration of structured (e.g. laboratory values, prescriptions), unstructured (e.g. clinical notes), and patient-generated data to detect otherwise hidden patterns and risks (Henry *et al.*, 2019; Wong *et al.*, 2022).

A key departure from traditional pharmacovigilance approaches is that ML models are designed not only to describe existing data, but also predict future outcomes. For instance, where disproportionality analyses in SRS suggest associations between drugs and events, supervised ML models can classify whether a patient is at high risk of an ADE based on their medical history (Kim *et al.*, 2022; Deimazar and Sheikhtaheri, 2023). Similarly, while manual chart reviews retrospectively identify errors, predictive ML models can generate real-time alerts to prevent harm (Yalçın *et al.*, 2023; Alqaraleh, Almagharbeh and Ahmad, 2025).

Several categories of ML methods have been applied in medication safety:

1. **Supervised learning** (e.g. logistic regression, decision trees, random forests [RF], support vector machines [SVM], and neural networks) uses labelled data for classification or prediction. RF and SVM have been applied to SRS data to improve ADR detection over disproportionality metrics (Kim *et al.*, 2022). Neural networks, including deep learning architectures, have shown strong performance in processing complex inputs such as free-text notes (Henry *et al.*, 2019; Kompa *et al.*, 2022).
2. **Unsupervised learning** identifies hidden patterns in unlabelled data. Clustering can reveal patient subgroups with elevated risk, while anomaly detection highlights unusual prescribing behaviours. Association rule mining has been used in SRS data to detect

previously unrecognised drug–event co-occurrences (Noguchi, Tachi and Teramachi, 2021; Barbieri *et al.*, 2025).

3. **Natural language processing** (NLP) extracts information from unstructured text or social media. Named entity recognition (NER) can identify drugs and symptoms, while classification algorithms assess whether text indicates an ADE. Henry *et al.* (2019) used recurrent neural networks (RNNs) to extract ADEs from EHR text, and Xie, Zeng and Marcum (2017) highlighted the value of deep learning for noisy patient-generated data.
4. Finally, **advanced methods** include ensemble learning, which combines multiple models to improve accuracy, and graph neural networks (GNNs), which model drug–patient relationships to predict complex interactions (Al-Rabeah and Lakizadeh, 2022; Lee *et al.*, 2022; Zhong *et al.*, 2023). Time-series approaches are also applied to longitudinal safety monitoring, providing early warnings of emerging risks (Li, 2025).

These approaches expand medication safety beyond retrospective detection towards proactive, personalised, and scalable risk assessment. Nonetheless, challenges remain: supervised methods depend on high-quality labelled data, unsupervised methods risk spurious associations, and deep learning models often lack interpretability (Kassekert *et al.*, 2022; Ball *et al.*, 2024; Simpson and Qasim, 2025; Toni *et al.*, 2024).

## Current applications and case studies

### Identification of safety events

Detecting ADEs and ADRs remains a primary focus of ML applications. In hospital settings, ML models trained on EHRs, including RNNs and gradient boosting, consistently outperformed rule-based methods in detecting ADEs (Deimazar and Sheikhtaheri, 2023). Henry *et al.* (2019) demonstrated that deep learning NLP models could extract ADE mentions from unstructured clinical notes with high precision, broadening the scope of surveillance beyond structured data. In SRSs, supervised models such as RF and SVM improved ADR signal detection compared with disproportionality analysis (Kim *et al.*, 2022). Bayesian networks applied to SRS data allowed causal inference, reducing false positives by modelling conditional dependencies (Cherkas, Ide and van Stekelenborg, 2022; Kim *et al.*, 2022; Yiqing Zhao *et al.*, 2022). Social media has also been explored: Xie, Zeng and Marcum (2017) argued that deep learning could filter complex patient posts, while Lee *et al.* (2022) demonstrated a hybrid ML system that combined text mining with classification models to detect potential side effects in online drug reviews.

### Identification of potential medication errors

Preventing medication errors is another domain where ML has shown tangible benefits. In neonatal intensive care units, predictive ML models that combine patient and provider factors to classify error risk yielded high accuracy (area-under-the-curve [AUC] 0.82-0.92) (Yalçın *et al.*, 2023; Henry Basil *et al.*, 2024). Alqaraleh, Almagharbeh and Ahmad (2025) reviewed studies on ML-enhanced clinical decision support systems, finding reductions of up to 95% in operating room medication errors and 80% in intravenous medication errors. These findings illustrate how ML can augment human vigilance by providing real-time, context-specific alerts.

## Prediction of drug–drug interactions

ML has advanced the prediction of DDIs by moving beyond known pharmacological mechanisms. Ibrahim *et al.* (2021) developed SMDIP, a similarity-based supervised ML model, achieving strong predictive performance (AUC 0.76, accuracy 0.79) and identifying novel antiviral DDIs later supported by pharmacological evidence. Barbieri *et al.* (2025) applied network analysis and ML to EHRs to identify drug-induced acute myocardial infarction, showing how relational modelling can capture higher-order interactions. In pre-clinical research, Bai *et al.*, 2025 reviewed ML-enabled toxicity prediction across multiple toxicity classes, highlighting applications of deep learning and GNNs for predicting organ-specific toxicities. These approaches extend pharmacovigilance into predictive domains, supporting both safer prescribing and earlier identification of risks in drug development.

## Limitations and future directions

Despite promising advances, the use of ML in medication safety analysis faces important limitations. A paramount concern is data quality and bias. The effectiveness of any model is inherently constrained by the quality of its inputs, and this remains a challenge across healthcare sources. EHRs frequently contain incomplete, inconsistent, or erroneous entries owing to variation in documentation practices (Cohen *et al.*, 2019). SRSs are similarly limited by underreporting and selective reporting bias, yielding samples that are rarely representative of wider populations (Kim *et al.*, 2022). Social media, while rich in valuable patient perspectives, is noisy and lacks clinical validation (Xie, Zeng and Marcum, 2017). Training models on such biased data risks perpetuating existing inequities; underrepresentation of older adults or minority populations, for example, can result in systematic under-detection of risks precisely in the groups most in need of protection.

Concerns around generalisability and reproducibility add further complexity. Many ML models are developed and validated within single institutions or on narrow datasets (Deimazar and Sheikhtaheri, 2023). Without rigorous external validation, their performance in other healthcare settings remains highly uncertain. Differences in coding standards, population demographics, and clinical practices can significantly degrade accuracy when models are transferred across sites (Rockenschaub *et al.*, 2024). These issues limit reproducibility and slow the transition from research prototypes to dependable tools for global pharmacovigilance.

Interpretability is another critical challenge. Deep learning architectures, such as recurrent and convolutional neural networks, often outperform simpler models but at the expense of transparency (Teng *et al.*, 2022). For clinicians and regulators, understanding why a model produces a particular output is fundamental to trust and accountability. The “black-box” nature of complex models may therefore hinder clinical adoption and regulatory approval (Kassekert *et al.*, 2022). Conversely, simpler, interpretable models offer clarity but usually at the cost of predictive performance, leaving an unresolved trade-off between accuracy and transparency.

Integration into clinical workflows also presents obstacles. Even highly accurate models may have little practical impact if not embedded seamlessly into care processes. ML-based alerts that are overly frequent or insufficiently specific can contribute to “alert fatigue,” causing clinicians to override or ignore them (Alqaraleh, Almagharbeh and Ahmad, 2025). Outputs must therefore be presented in actionable, context-sensitive formats to be clinically useful. Without careful user-centred design, ML systems risk adding burden rather than enhancing safety.

Finally, regulatory and ethical challenges continue to shape the trajectory of ML adoption. Regulators remain cautious, emphasising the need for transparency, rigorous validation, and ongoing monitoring. The FDA's Emerging Drug Safety Technology Program signals openness to innovation, but agencies stress that models must be explainable, robust, and accountable (Ball *et al.*, 2024). Ethical considerations also arise, including responsibility for missed or erroneous predictions and the imperative to mitigate bias. These concerns highlight that advances in governance and ethics are as critical as technical progress.

Although substantial limitations persist, ML offers considerable promise. Many studies report high predictive accuracy and reduced error rates in experimental or pilot deployments (Ibrahim *et al.*, 2021; Yalçın *et al.*, 2023). Yet scepticism remains among regulators and clinicians, who caution that methodological rigour, interpretability, and reproducibility are prerequisites before such systems can be trusted in high-stakes safety contexts (Kassekert *et al.*, 2022; Ball *et al.*, 2024). This tension between innovation and caution underscores the transitional stage of the field. Addressing these limitations will require both methodological advances and strong governance. Federated learning may improve generalisability by enabling training across distributed datasets without compromising privacy. Explainable AI techniques such as SHapley Additive exPlanations (SHAP) values and counterfactual reasoning can enhance transparency, while integration of genomic and multi-omics data offers opportunities for personalised pharmacovigilance. Crucially, collaboration between technical experts, clinicians, and regulators will be essential to ensure that ML systems are safe, equitable, and clinically useful.

## Conclusion

Machine learning is reshaping medication safety analysis through enhanced scalability, adaptability, and predictive capacity beyond traditional approaches. By complementing traditional pharmacovigilance frameworks, ML enhances ADE detection, supports prediction of medication errors, and improves DDI identification, with examples from EHR-based ADE detection, medication error prediction, and network-based DDI modelling illustrating its potential value. Nevertheless, challenges remain, namely suboptimal data quality, a need for improved interpretability generalisability, and slow regulatory endorsement. The most promising avenues for increased adoption lie in integrated, explainable, validated ML systems embedded within existing frameworks. Such approaches could advance the central goal of medication safety by preventing avoidable harm while ensuring therapeutic benefit.

## References

- Alomar, M. *et al.* (2020) 'Post marketing surveillance of suspected adverse drug reactions through spontaneous reporting: current status, challenges and the future', *Therapeutic Advances in Drug Safety*, 11, p. 2042098620938595. Available at: <https://doi.org/10.1177/2042098620938595>.
- Alqaraleh, M., Almagharbeh, W.T. and Ahmad, M.W. (2025) 'Exploring the impact of artificial intelligence integration on medication error reduction: A nursing perspective', *Nurse Education in Practice*, 86, p. 104438. Available at: <https://doi.org/10.1016/j.nepr.2025.104438>.
- Al-Rabeah, M.H. and Lakizadeh, A. (2022) 'Prediction of drug-drug interaction events using graph neural networks based feature extraction', *Scientific Reports*, 12, p. 15590. Available at: <https://doi.org/10.1038/s41598-022-19999-4>.
- Bai, C. *et al.* (2025) 'Machine Learning-Enabled Drug-Induced Toxicity Prediction', *Advanced Science*, 12(16), p. 2413405. Available at: <https://doi.org/10.1002/adv.202413405>.
- Ball, R. *et al.* (2024) 'Trust but Verify: Lessons Learned for the Application of AI to Case-Based Clinical Decision-Making From Postmarketing Drug Safety Assessment at the US Food and Drug Administration'. Available at: <https://www.jmir.org/2024/1/e50274>.
- Barbieri, M.A. *et al.* (2025) 'Network Analysis and Machine Learning for Signal Detection and Prioritization Using Electronic Healthcare Records and Administrative Databases: A Proof of Concept in Drug-Induced Acute Myocardial Infarction', *Drug Safety*, 48(5), pp. 513–526. Available at: <https://doi.org/10.1007/s40264-025-01515-y>.
- Bate, A. and Evans, S.J.W. (2009) 'Quantitative signal detection using spontaneous ADR reporting', *Pharmacoepidemiology and Drug Safety*, 18(6), pp. 427–436. Available at: <https://doi.org/10.1002/pds.1742>.
- Brown, J.S. *et al.* (2022) 'The US Food and Drug Administration Sentinel System: a national resource for a learning health system', *Journal of the American Medical Informatics Association*, 29(12), pp. 2191–2200. Available at: <https://doi.org/10.1093/jamia/ocac153>.
- Cherkas, Y., Ide, J. and van Stekelenborg, J. (2022) 'Leveraging Machine Learning to Facilitate Individual Case Causality Assessment of Adverse Drug Reactions', *Drug Safety*, 45(5), pp. 571–582. Available at: <https://doi.org/10.1007/s40264-022-01163-6>.
- Cohen, G.R. *et al.* (2019) 'Variation in Physicians' Electronic Health Record Documentation and Potential Patient Harm from That Variation', *Journal of General Internal Medicine*, 34(11), pp. 2355–2367. Available at: <https://doi.org/10.1007/s11606-019-05025-3>.
- Deimazar, G. and Sheikhtaheri, A. (2023) 'Machine learning models to detect and predict patient safety events using electronic health records: A systematic review', *International Journal of Medical Informatics*, 180, p. 105246. Available at: <https://doi.org/10.1016/j.ijmedinf.2023.105246>.
- European Medicines Agency (2024a) *Reflection paper on the use of Artificial Intelligence (AI) in the medicinal product lifecycle\_240903*. Available at: [https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-use-artificial-intelligence-ai-medicinal-product-lifecycle\\_en.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-use-artificial-intelligence-ai-medicinal-product-lifecycle_en.pdf).

European Medicines Agency (2024b) *Review of AI/ML applications in medicines lifecycle (2024)*. Available at: [https://www.ema.europa.eu/en/documents/report/review-artificial-intelligence-machine-learning-applications-medicines-lifecycle-2024-horizon-scanning-short-report\\_en.pdf](https://www.ema.europa.eu/en/documents/report/review-artificial-intelligence-machine-learning-applications-medicines-lifecycle-2024-horizon-scanning-short-report_en.pdf).

Golder, S. *et al.* (2024) 'The Value of Social Media Analysis for Adverse Events Detection and Pharmacovigilance: Scoping Review', *JMIR Public Health and Surveillance*, 10(1), p. e59167. Available at: <https://doi.org/10.2196/59167>.

Henry Basil, J. *et al.* (2024) 'Machine learning-based risk prediction model for medication administration errors in neonatal intensive care units: A prospective direct observational study', *Digital Health*, 10, p. 20552076241286434. Available at: <https://doi.org/10.1177/20552076241286434>.

Henry, S. *et al.* (2019) '2018 n2c2 shared task on adverse drug events and medication extraction in electronic health records', *Journal of the American Medical Informatics Association : JAMIA*, 27(1), pp. 3–12. Available at: <https://doi.org/10.1093/jamia/ocz166>.

Ibrahim, H. *et al.* (2021) 'Similarity-based machine learning framework for predicting safety signals of adverse drug–drug interactions', *Informatics in Medicine Unlocked*, 26, p. 100699. Available at: <https://doi.org/10.1016/j.imu.2021.100699>.

Jordan, M.I. and Mitchell, T.M. (2015) 'Machine learning: Trends, perspectives, and prospects', *Science*, 349(6245), pp. 255–260. Available at: <https://doi.org/10.1126/science.aaa8415>.

Kassekert, R. *et al.* (2022) 'Industry Perspective on Artificial Intelligence/Machine Learning in Pharmacovigilance', *Drug Safety*, 45(5), pp. 439–448. Available at: <https://doi.org/10.1007/s40264-022-01164-5>.

Kim, H.R. *et al.* (2022) 'Analyzing adverse drug reaction using statistical and machine learning methods'. Available at: [https://journals.lww.com/md-journal/fulltext/2022/06240/analyzing\\_adverse\\_drug\\_reaction\\_using\\_statistical.30.aspx](https://journals.lww.com/md-journal/fulltext/2022/06240/analyzing_adverse_drug_reaction_using_statistical.30.aspx).

Kompa, B. *et al.* (2022) 'Artificial Intelligence Based on Machine Learning in Pharmacovigilance: A Scoping Review', *Drug Safety*, 45(5), pp. 477–491. Available at: <https://doi.org/10.1007/s40264-022-01176-1>.

Lee, E. *et al.* (2022) 'Drug Usage Safety from Drug Reviews with Hybrid Machine Learning Approach', *Computer Systems Science and Engineering*, 45(3), pp. 3053–3077. Available at: <https://doi.org/10.32604/csse.2023.029059>.

Li, T. (2025) 'Enhancing Adverse Event Monitoring and Management in Phase IV Chronic Disease Drug Trials: Applications of Machine Learning', in *Proceedings of the 4th International Conference on Biomedical and Intelligent Systems*. New York, NY, USA: Association for Computing Machinery (IC-BIS '25), pp. 69–73. Available at: <https://doi.org/10.1145/3745034.3745046>.

Noguchi, Y., Tachi, T. and Teramachi, H. (2021) 'Detection algorithms and attentive points of safety signal using spontaneous reporting systems as a clinical data source', *Briefings in Bioinformatics*, 22(6), p. bbab347. Available at: <https://doi.org/10.1093/bib/bbab347>.

Patel, T.K. *et al.* (2023) 'Impact of suspected adverse drug reactions on mortality and length of hospital stay in the hospitalised patients: a meta-analysis', *European Journal of Clinical Pharmacology*, 79(1), pp. 99–116. Available at: <https://doi.org/10.1007/s00228-022-03419-7>.

Raventos, B. and Prieto-Alhambra, D. (2025) 'Real-world evidence for regulatory purposes: The example of DARWIN EU®', *Farmacia Hospitalaria: Organo Oficial De Expresion Cientifica De La Sociedad Espanola De Farmacia Hospitalaria*, 49(2), pp. 62–64. Available at: <https://doi.org/10.1016/j.farma.2025.02.011>.

Rockenschaub, P., Ansell, D. and Shallcross, L. (2018) 'Linking individual-level data on diagnoses and dispensing for research on antibiotic use: Evaluation of a novel data source from English secondary care', *Pharmacoepidemiology and Drug Safety*, 27(2), pp. 206–212. Available at: <https://doi.org/10.1002/pds.4367>.

Russel, S. and Norvig, P. (2021) *Artificial intelligence: a modern approach*. 4th edn. Upper Saddle River, NJ : Prentice Hall: Pearson. Available at: <https://doi.org/10.1109/MSP.2017.2765202>.

Simpson, M.D. and Qasim, H.S. (2025) 'Clinical and Operational Applications of Artificial Intelligence and Machine Learning in Pharmacy: A Narrative Review of Real-World Applications', *Pharmacy*, 13(2), p. 41. Available at: <https://doi.org/10.3390/pharmacy13020041>.

Teng, Q. *et al.* (2022) 'A survey on the interpretability of deep learning in medical diagnosis', *Multimedia Systems*, 28(6), pp. 2335–2355. Available at: <https://doi.org/10.1007/s00530-022-00960-4>.

Toni, E. *et al.* (2024) 'Machine Learning Techniques for Predicting Drug-Related Side Effects: A Scoping Review', *Pharmaceuticals*, 17(6), p. 795. Available at: <https://doi.org/10.3390/ph17060795>.

Wong, J. *et al.* (2022) 'Applying Machine Learning in Distributed Data Networks for Pharmacoepidemiologic and Pharmacovigilance Studies: Opportunities, Challenges, and Considerations', *Drug Safety*, 45(5), pp. 493–510. Available at: <https://doi.org/10.1007/s40264-022-01158-3>.

World Health Organization (2024) *Medication without harm: Policy brief*. Available at: <https://www.who.int/publications/i/item/9789240062764> (Accessed: 7 September 2025).

Xie, J., Zeng, D.D. and Marcum, Z.A. (2017) 'Using deep learning to improve medication safety: the untapped potential of social media', *Therapeutic Advances in Drug Safety*, 8(12), pp. 375–377. Available at: <https://doi.org/10.1177/2042098617729318>.

Yalçın, N. *et al.* (2023) 'Development and validation of a machine learning-based detection system to improve precision screening for medication errors in the neonatal intensive care unit', *Frontiers in Pharmacology*, 14. Available at: <https://doi.org/10.3389/fphar.2023.1151560>.

Yiqing Zhao *et al.* (2022) 'Machine Learning in Causal Inference: Application in Pharmacovigilance', *Drug Safety*, 45(5), pp. 459–476. Available at: <https://doi.org/10.1007/s40264-022-01155-6>.

Zhong, Y. *et al.* (2023) 'DDI-GCN: Drug-drug interaction prediction via explainable graph convolutional networks', *Artificial Intelligence in Medicine*, 144, p. 102640. Available at: <https://doi.org/10.1016/j.artmed.2023.102640>.