Annals of Tropical Research 47 (2):290-303(2025)
https://doi.org/10.32945/atr47218.2025

ANNALS OF TROPICAL RESEARCH

Review Article

Advancing pharmacovigilance through artificial intelligence: A review of applications and ethical considerations

Zhinya Kawa Othman[1,2*], Prisca Mirindi Nabashaho[3], Emmanuella Ojugbeli[4], David Olpengs[5], Mohamed Mustaf Ahmed[6] and Don Eliseo Lucero-Prisno III[7,8,9]

ABSTRACT

Artificial intelligence is increasingly applied in pharmacovigilance to identify, prioritize, and interpret adverse drug reactions across real-world data sources. This narrative review synthesizes recent peer-reviewed studies (2015–2024) and maps AI use across four domains: extraction of adverse drug reactions from social and clinical text, supervised and ensemble signal detection in spontaneous reporting systems and electronic health records, knowledge-graph-based discovery of drug–event associations, and prediction of outcome seriousness to support triage. Across domains, implementations most consistently enhance intake, coding, prioritization, and the timeliness of safety assessment, while graph-based methods surface plausible associations for follow-up and seriousness models aid risk stratification. Cross-cutting challenges include heterogeneous and shifting data, annotation burden, class imbalance (especially for rare events), and concerns around transparency, privacy, and fairness. Evidence remains predominantly retrospective, with uneven external validation, underscoring the need for prospective studies, standardized reporting and calibration, fairness audits, and closer alignment with regulatory signal-management workflows spanning detection, validation, analysis, prioritization, and assessment. By clarifying where AI is already dependable and where methodological and ethical gaps persist, this review offers practical directions for integrating AI into routine pharmacovigilance with auditable thresholds, monitoring, and human oversight.

Received: 21 September 2025
Revised: 26 October 2025
Accepted: 6 December 2025
Published: 12 December 2025

Keywords: Artificial intelligence, pharmacovigilance, adverse drug reaction, machine learning, applications, deep learning

[1]Department of Pharmacy, Kurdistan Technical Institute, Sulaymaniyah, Kurdistan Region, Iraq
[2]Faculty of Pharmaceutical Sciences, Chulalongkorn University, Bangkok, Thailand
[3]Department of Pharmacy, United States International University-Africa, Nairobi, Kenya
[4]Faculty of Pharmacy, University of Benin, Benin, Nigeria
[5]Health System Strengthening Department, Amref Health Africa. Nairobi, Kenya
[6]Faculty of Medicine and Health Sciences, SIMAD University, Mogadishu, Somalia
[7]Department of Global Health and Development, London School of Hygiene and Tropical Medicine, London, United Kingdom
[8]Research Office, Palompon Institute of Technology, Palompon, Leyte, Philippines
[9]Research Unit, Bukidnon State University, Malaybalay City, Bukidnon, Philippines

*Corresponding Author. Address: Department of Pharmacy, Kurdistan Technical Institute, Sulaymaniyah, Kurdistan Region, Iraq; Email: zhinya.kawa@kti.edu.iq

INTRODUCTION

Pharmacovigilance (PV) is the science and effort devoted to identifying, evaluating, comprehending, and avoiding side effects and other drug-related issues. Adverse drug reactions (ADR) are harmful and unintended responses to medicinal products that occur at doses normally used for prophylaxis, diagnosis, or therapy (Shukla et al., 2024). In 2023, the World Health Organization (WHO) reported that approximately 1 in 10 patients are harmed during care worldwide, contributing to over 3 million deaths each year. However, in low-middle-income countries (LMICs), up to 4% of people die from unsafe care. More than half of this harm, approximately 1 in 20 patients, is preventable, and about half of preventable harm is medication-related (Patient Safety, 2023). Medication errors alone are estimated to cost approximately $42 billion annually (Medication Without Harm, 2017).

As the pharmaceutical industry grows, traditional methods, such as clinical trials and manual reporting, are challenged by the increasing complexity of the data. Detecting and prioritizing true safety signals is especially difficult given the massive, heterogeneous, and noisy data streams (Shukla et al., 2024). In 2017, the WHO South-East Asia region urged action to prioritize and strengthen pharmacovigilance by building national ADR reporting systems that include diverse health facilities (The Pharmacovigilance System for Traditional Medicine in Thailand, 2017). Artificial intelligence (AI) and machine learning (ML) have been shown to be useful for multiple aspects of data ingestion and for assessment of reported causality in individual case safety reports (ICSRs) (Cherkas et al., 2022). Currently, the use of AI approaches is increasing in various areas of pharmacovigilance, including safe operations, signal management, and identification of target populations (Salas et al., 2022). These systems can automate the intake and analysis of safety information from electronic health records (EHRs), spontaneous reporting systems, and active surveillance, complementing statistical approaches such as penalized regression, Bayesian borrowing, and temporal scan statistics while enabling the discovery of syndromic patterns in complex data (Badary, 2025). Empirical studies have reported consistent gains for text-based intake and triage, with competitive performance in signal detection and seriousness prediction across multiple data sources (Li et al., 2018; Nikfarjam et al., 2015; Yang et al., 2019). Thus, leveraging AI technologies helps minimize preventable harm by incorporating digital approaches to improve patient safety (Salas et al., 2022).

However, significant gaps persist when extracting reliable safety signals from large, noisy, and heterogeneous data streams. Despite promising metrics, much of the AI–PV literature is retrospective, variably benchmarked, and unevenly externally validated across regions, languages, and data modalities (Kompa et al., 2022). Closing these gaps will align PV and regulatory practices with real-world needs by standardizing outcome/performance reporting across extraction, signal detection, triage, and seriousness prediction. In parallel, the field must address AI-related ethical issues, including privacy, bias, transparency, accountability, and deployment governance (Dimitsaki et al., 2024).

The primary objective of our narrative review is to synthesize recent peer-reviewed evidence on AI applications and measure benefits in PV while cataloging the ethical and practical challenges that remain, thereby delineating where the field is robust and where critical deficits still constrain the impact. In doing so, we aim to clarify how technical advances map to operational use, ethical safeguards, and regulatory signal management workflow. In assembling this narrative review, we synthesized peer-reviewed literature from 2015–2024 identified in PubMed/MEDLINE, Scopus,

and Web of Science, supplemented by Google Scholar. We included English-language original studies applying AI and ML to core PV tasks, including text-based adverse drug reaction (ADR) extraction; supervised and ensemble signal detection in spontaneous reporting systems or electronic health records (EHRs); knowledge-graph association discovery; and prediction of outcome seriousness. We excluded editorials, viewpoints, non-peer-reviewed preprints, rules-only automation, and studies outside of PV applications.

AI approaches, data sources, and performance metrics

Approaches include sequence labeling and natural language processing (NLP) for social media and clinical narratives, transformer fine-tuning for triage, supervised learning and meta-models for spontaneous reporting systems, gradient boosting and random forests for national databases, and knowledge-graph learning with external EHR validation (Table 1). Data sources covered Twitter, DailyStrength, PubMed abstracts, MADE 1.0 clinical notes, hospital EHRs, FDA Adverse Event Reporting System, Korea Adverse Event Reporting System, VigiBase, and benchmark sets such as CSIRO Adverse Drug Event Corpus (CADEC), Side Effect Resource (SIDER), and off-label Side Effects Database(Bae et al., 2021; Hussain et al., 2021; Jung et al., 2024; Lee et al., 2022; Li et al., 2018; Martin et al., 2022; Nikfarjam et al., 2015; Stanovsky et al., 2017; Yang et al., 2019; Zhao et al., 2023). Across text-mining tasks, studies consistently reported strong named entity recognitionand competitive relation extraction sufficient for automated intake, coding, and triage; detailed figures are summarized in Table 2 (Hussain et al., 2021; Li et al., 2018; Nikfarjam et al., 2015; Stanovsky et al., 2017). In signal detection, supervised and multimodal fusion approaches generally outperform disproportionality baselines and can improve timeliness, although the performance varies by database, time window, and event prevalence (Harpaz et al., 2017; Jeong et al., 2018; Lee et al., 2022; Martin et al., 2022). Beyond detection, knowledge-graph models prioritize plausible drug–event associations with supportive external validation, and seriousness-prediction models enable triage and prioritization; both are less reliable for rare classes and benefit from calibration and fairness checks before deployment (Bean et al., 2017; Jung et al., 2024; Zhao et al., 2023).

Applications and reported challenges

Applications cluster around automated ADR case findings from social media and clinical narratives, auto-coding and triage of patient reports for national pharmacovigilance portals, supervised and ensemble screening of spontaneous reporting systems and EHRs for earlier signal emergence, knowledge-graph-based association discovery to prioritize candidates for follow-up**,** and prioritization of serious outcomes to focus safety review (Lee et al., 2022; Li et al., 2018; Martin et al., 2022; Nikfarjam et al., 2015; Yang et al., 2019; Zhao et al., 2023). The challenges reported were informal language and annotation noise in social media, domain shift across sources, heterogeneous and delayed reference labels, coverage and bias in public data, single-center constraints, class imbalance, and low precision–recall for rare classes. Many studies have further highlighted the dependence on labeled references and limited external validation (Bae et al., 2021; Bean et al., 2017; Harpaz et al., 2017; Hussain et al., 2021; Jeong et al., 2018; Jung et al., 2024; Lee et al., 2022; Martin et al., 2022; Stanovsky et al., 2017; Zhao et al., 2023). Detailed metrics are listed in Table 1.

Table 1. Summary of water quality parameters and usage profiles of sampled wells

Type of AI intervention/ exposure	Type of outcome	Methodological approach	Data source	Key performance / outcome metric	Reported impact on pharmacovigilance	Challenges / limitations	Reference
Conditional Random Field (CRF) sequence labeler with word-embedding cluster features	ADR mention extraction (social media)	Supervised NER for ADR spans in tweets and patient forums	Twitter and DailyStrength	F-measure 0.82 on ADR extraction, surpassing baselines	Demonstrates reliable ADR extraction from lay text for PV triage streams	Informal language, annotation noise and domain drift across platforms	(Nikfarjam et al., 2015)
Multimodal signal fusion (FAERS + claims + MEDLINE + web search)	Signal detection performance & timeliness	Combine source-specific signals via predictive/weighted aggregation	FAERS and MEDLINE	AUC gain 0.04–0.09; lead-time 7–22 months vs label dates	Earlier and stronger signals by fusing heterogeneous sources	Data heterogeneity and benchmark/label timing constraints	(Harpaz et al., 2017)
Knowledge-infused Recurrent Neural Network (DBpedia embeddings)	ADR mention extraction (social media)	Recurrent Neural Network (RNN) transducer + knowledge-graph embeddings and active learning variant	CADEC corpus and social media posts	F1 93.4 (CADEC); 83.9 with limited expert labels	High-accuracy ADR NER with reduced annotation burden	New annotation tasks often have scarce labeled training data, and creating large, high-quality annotations is heavyweight and expensive	(Stanovsky et al., 2017)

Abbreviations: ADR – adverse drug reaction; ADE – adverse drug event; PV – pharmacovigilance; NER – named-entity recognition; EHR – electronic health record; SRS – spontaneous reporting system; FAERS – FDA Adverse Event Reporting System; KAERS – Korea Adverse Event Reporting System; VigiBase – WHO global ICSR database; KG – knowledge graph; GCAP – graph contrastive pretraining; GNN – graph neural network; CNN – convolutional neural network; AUC – area under the ROC curve (used when reported verbatim in the cited paper); AUROC – area under the ROC curve (used otherwise); AUPR – area under the precision–recall curve; PRR – proportional reporting ratio; ROR – reporting odds ratio; IC – information component; EBGM – empirical Bayes geometric mean; MADE 1.0 – Medical Annotation and Data Extraction challenge corpus; CADEC – CSIRO Adverse Drug Event Corpus; SIDER – Side Effect Resource; OFFSIDES – OFF-label Side Effects; CV – cross-validation.

Table 1. continued

Type of AI intervention/ exposure	Type of outcome	Methodological approach	Data source	Key performance / outcome metric	Reported impact on pharmacovigilance	Challenges / limitations	Reference
Knowledge-graph learning + EHR validation	Prediction of unknown ADRs	Graph features (drug/target/indication/ADR) with enrichment-based learner and validate in EHR	Public knowledge graph (KG) + South London and Maudsley EHR	Cross-validation AUC 0.92; predicts 68% held-out edges and EHR validation > baselines	Prioritizes plausible and previously unknown ADRs for follow-up	Coverage/bias of public data and validation limited to one health system	(Bean et al., 2017)
Meta-model (RF/SVM/NN) integrating CERT, CLEAR, PACE	Lab-event–related ADR signals	Meta-model stacking outputs of CERT, CLEAR, and PACE algorithms	21-year inpatient EHR (Korea)	AUROC 0.737–0.816, F1 0.629–0.709, exceeding each base method	Improves ADR signal detection from labs by combining methods	Single-centre data and mapping labs (ADR surrogates)	(Jeong et al., 2018)
End-to-end BiLSTM-CRF + relation extraction and multi-task learning	ADE info extraction from EHR notes	Joint NER + relation extraction and Multi-Task Learning (MTL) variants	MADE 1.0 clinical notes	Best single-task F1 65.9%; MTL (HardMTL) 66.7%	Automates ADE entities/relations to feed PV workflows	Dataset-specific tuning; external generalization uncertain	(Li et al., 2018)

Table 1. continued

Type of AI intervention/ exposure	Type of outcome	Methodological approach	Data source	Key performance / outcome metric	Reported impact on pharmacovigilance	Challenges / limitations	Reference
Deep learning pipeline (LSTM NER + ML relations), "MADEx"	Medication, ADE, relation extraction	NER with LSTM; SVM/RF for intra-/cross-sentence relations	Clinical notes (MADE 1.0)	NER F1 0.8233 (top-3 in challenge) and competitive relation extraction	Robust pipeline for ADE extraction to support PV case finding	Cross-sentence relations harder and manual review still needed	(Yang et al., 2019)
Transformer classifiers (BERT/RoBERTa)	ADR tweet/paper classification for PV intake/triage	Fine-tuned transformers and multi-corpus evaluation	Twitter and PubMed abstracts	F1 0.976 (PubMed), 0.896 (Twitter) for ADR classification	High-precision social/literature triage to augment PV intake	Domain shift across sources and reproducibility of noisy labels	(Hussain et al., 2021)
Ensemble ML (GBM/RF) vs disproportionality	Early signal detection (oncology)	Supervised models trained on known signals which apply to new data	FAERS (framework later applied to KAERS/FAERS)	Reported superior AUROC to disproportionality; FAERS AUROC around 0.75 in follow-up application	Identified earlier signals for cancer drugs and ML viable for PV screening	Performance varies by database/time window and simplicity	(Bae et al., 2021)

Table 1. continued

Type of AI intervention/ exposure	Type of outcome	Methodological approach	Data source	Key performance / outcome metric	Reported impact on pharmacovigilance	Challenges / limitations	Reference
Light Gradient Boosted Machine (LGBM), transformer models like Cross-lingual Language Model (XLM) and CamemBERT	General signal detection	Compare ML vs classical metrics across labeled sets	Patient reports from the French national pharmacovigilance web portal	AUC of 0.97	ML improves ranking of true signals over traditional stats	Requires labeled reference and risk of overfitting	(Martin et al., 2022)
GBM/ RF for early signal detection	Timeliness and accuracy of infliximab safety signals	Train on known label adverse events and year-stratified detection	KAERS and external test on FAERS	KAERS AUROC 0.82 (GBM), 0.79 (RF); FAERS 0.75/0.73; 4/5 adverse events detected in first year vs label	Earlier detection than PRR/IC and cross-database generalization	Lower AUROC on FAERS, class imbalance and lifecycle effects	(Lee et al., 2022)

Table 1. continued

Type of AI intervention/ exposure	Type of outcome	Methodological approach	Data source	Key performance / outcome metric	Reported impact on pharmacovigilance	Challenges / limitations	Reference
Multi-task deep model (GCAP; GNN+CNN+attention)	Seriousness classification for ADRs	Multimodal representation of drug and ADR semantics, multi-head attention	Benchmark set + SIDER and OFFSIDES tests	Avg AUC 0.917 (CV); 0.891 (SIDER) & 0.901 (OFFSIDES) avg AUC on tests	Predicts seriousness classes to prioritize PV risk	Lower AUPR in rare classes (CA/RI) and data sparsity	(Zhao et al., 2023)
Gradient boosting (XGBoost)	Safety signals for serious CV adverse events (tisagenlecleucel)	Supervised model using labeled positive/negative controls	WHO VigiBase (to Feb 2024)	Test AUROC 0.76; accuracy/sensitivity/specificity ≥0.98 and outperforms PRR/ROR/IC/ EBGM	Flags 6 previously "unknown" CV adverse events for monitoring	Dependent on labeling and VigiBase reporting biases	(Jung et al., 2024)

Critical Analysis and future directions

Across the included studies, AI systems were most consistently effective for pharmacovigilance tasks anchored in text processing and workflow triage, with several deployments showing robust real-world performance and others demonstrating competitive or improved signal detection compared with traditional disproportionality methods (Warner et al., 2025). These patterns are plausible given the maturation of clinical NLP and supervised learning, but they also reflect careful problem framing (e.g., case intake, coding, and prioritization), where ground truth is accessible and outcomes are proximal, conditions under which models tend to generalize better. Nonetheless, external performance can drift when data or practice patterns change, underscoring the need for transparent reporting and prospective evaluation (Wiens et al., 2019). Findings that multimodal or knowledge-graph-based approaches enrich detection breadth and relationship reasoning align with prior work showing that graphs can capture latent drug–drug–event structures, but utility depends on curation quality and harmonization across sources (Zitnik et al., 2018).

Applications that aimed to prioritize or de-duplicate spontaneous reports also mirrored earlier pharmacovigilance advances, showing tangible workflow gains, while reminding us that spontaneous systems carry endemic limitations that bound inference. Duplicates and missing denominators are well-recognized constraints that require method safeguards and cautious interpretations (Kiguba et al., 2024). The predominance of retrospective designs and class-imbalanced outcomes in the dataset likely contributed to both apparent gains in internal metrics and variability across tasks, which is consistent with the broader imbalanced-learning literature and argues for calibrated evaluation using metrics beyond accuracy as well as attention to minority-class fidelity (Johnson & Khoshgoftaar, 2019). Several studies have reported subgroup or product-specific differences in performance or outputs; given established sex-related and context-specific patterns in adverse drug reaction reporting, routine fairness checks and stratified analyses should be embedded into pharmacovigilance AI-PV to avoid masking heterogeneous errors (Hendriksen et al., 2021; Watson et al., 2019).

Methodologically, we interpret the evidence through three lenses, issues, ethics, and regulation, to make the analysis more actionable. Dataset shift (temporal, institutional, and linguistic), retrospective designs, and class imbalance are the primary threats to generalizability and reliability across tasks (Wiens et al., 2019). Signal detection systems are particularly sensitive to delayed or heterogeneous reference standards (Van Calster et al., 2019); knowledge-graph models depend on curation quality and coverage; and text-mining pipelines can be overfit to corpus-specific conventions (Nicholson & Greene, 2020). To mitigate these risks, studies should report calibration beyond accuracy (e.g., AUROC/AUPR, calibration curves), implement empirical calibration or control-based diagnostics where appropriate, and verify transportability via external validation (Collins et al., 2024; Liu et al., 2020; Rivera et al., 2020; Vasey et al., 2022).

Frameworks such as CONSORT-AI, SPIRIT-AI, DECIDE-AI, and TRIPOD+AI provide concrete checklists for pre-specifying endpoints, managing human–AI interaction, and documenting dataset shift and reproducibility (Collins et al., 2024; Liu et al., 2020; Rivera et al., 2020; Vasey et al., 2022). Bias, transparency, and privacy emerge as persistent concerns. Sources of bias include imbalanced

outcome prevalence, under-reporting of specific demographics, and context-specific reporting patterns that can yield subgroup performance gaps (Hendriksen et al., 2021; Watson et al., 2019). Transparency requires clear documentation of the model scope, data lineage, error modes, and human-in-the-loop decision points aligned with responsible-ML guidance (Wiens et al., 2019). Privacy risks span the secondary use of clinical notes and ingestion of public social media content; data minimization, robust de-identification, and governance for linkage/triage use cases are necessary safeguards (Dimitsaki et al., 2024). Routine subgroup reporting, fairness analyses (with prespecified thresholds for acceptable disparity), and post-deployment monitoring should be treated as first-class outcomes, not post hoc checks.

Operational integration should map model outputs to established signal-management steps (detection, validation, analysis, prioritization, and assessment) with auditable thresholds and documentation that align with pharmacovigilance practice (Warner et al., 2025). Prospective, real-world evaluations with external datasets can de-risk lifecycle effects and facilitate regulatory dialogue, whereas traceable data flows, versioning, and audit trails support inspections. Where duplicates and missing denominators constrain inference, methods should explicitly describe deduplication, exposure proxies, and uncertainty communication (Kiguba et al., 2024). Accordingly, we recommend prospective, externally validated evaluations with prespecified monitoring for dataset shift and human-factor impacts; the adoption of TRIPOD+AI to standardize reporting for prediction-oriented models, including fairness thresholds and calibration plans; the integration of empirical calibration and negative/positive control diagnostics in observational evaluations; and the explicit alignment of AI-enabled outputs with regulatory signal-management and benefit–risk assessment, supported by auditable thresholds, traceability, and oversight (Figure 1).

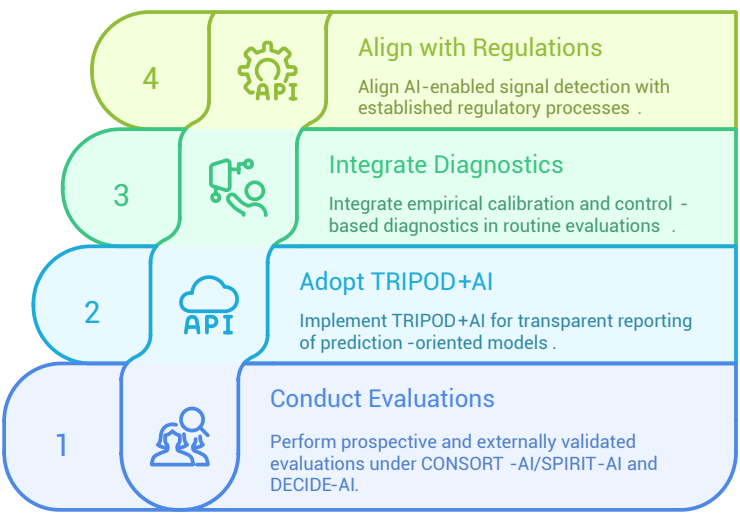

Figure 1. Recommendations to achieve AI integration in pharmacovigilance.

CONCLUSION

AI methods now reliably support PV tasks spanning ADR text mining, case intake/coding, workflow triage, and in many settings, more timely signal detection than disproportionality alone, while knowledge-graph and seriousness-prediction approaches extend discovery and prioritization. At the same time, the current evidence base is largely retrospective, variably externally validated, and sensitive to dataset shifts, class imbalances, and documentation gaps in transparency and privacy. Near-term priorities are prospective, externally validated evaluations, standardized reporting with calibration beyond accuracy, predefined fairness thresholds and subgroup monitoring, and explicit handling of PV constraints (deduplication, exposure proxies, and uncertainty communication). Operationally, model outputs should be mapped to established regulatory signal management steps with clear decision thresholds, traceability, and audit trails. With these safeguards, AI can be more safely and effectively embedded in routine PV to improve the timeliness, consistency, and overall quality of safety assessment.

Acknowledgment

None.

Author Contributions

ZKO: Conceptualization, Investigation, Writing—original draft. PMN, EO, DO, MMA: Writing—review & editing. DELP III supervised the study. All authors have read and approved the final manuscript.

Funding Source

Not applicable.

Availability of Data and Materials

No datasets were generated or analyzed during the current study.

Ethical Considerations

Not applicable.

Competing Interest

The authors declare no competing interests.

References

Badary, O. A. (2025). The potential of artificial intelligence and machine learning in pharmacovigilance: An update. *GenoMed Connect, 2*, 0010. https://doi.org/10.69709/GENOMC.2024.176699

Bae, J. H., Baek, Y. H., Lee, J. E., Song, I., Lee, J. H., & Shin, J. Y. (2021). Machine learning for detection of safety signals from spontaneous reporting system data: Example of nivolumab and docetaxel. *Frontiers in Pharmacology, 11*, 602365. https://doi.org/10.3389/fphar.2020.602365

Bean, D. M., Wu, H., Iqbal, E., Dzahini, O., Ibharim, Z. M., Broadbent, M., Stewart, R., & Dobson, R. J. B. (2017). Knowledge graph prediction of unknown adverse drug reactions and validation in electronic health records. *Scientific Reports, 7*, 16416. https://doi.org/10.1038/s41598-017-16674-x

Cherkas, Y., Ide, J., & van Stekelenborg, J. (2022). Leveraging machine learning to facilitate individual case causality assessment of adverse drug reactions. *Drug Safety, 45*(5), 571–582. https://doi.org/10.1007/s40264-022-01163-6

Collins, G. S., Moons, K. G. M., Dhiman, P., Riley, R. D., Beam, A. L., Van Calster, B., Ghassemi, M., Liu, X., Reitsma, J. B., Van Smeden, M., Boulesteix, A. L., Camaradou, J. C., Celi, L. A., Denaxas, S., Denniston, A. K., Glocker, B., Golub, R. M., Harvey, H., Heinze, G., Hoffman, M. M., Kengne, A. P., Lam, E., Lee, N., Loder, E. W., Maier-Hein, L., Mateen, B. A., McCradden, M. D., Oakden-Rayner, L., Ordish, J., Parnell, R., Rose, S., Singh, K., Wynants, L., & Logullo, P. (2024). TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ, 385*, e078378. https://doi.org/10.1136/bmj-2023-078378

Dimitsaki, S., Natsiavas, P., & Jaulent, M. C. (2024). Applying AI to structured real-world data for pharmacovigilance purposes: Scoping review. *Journal of Medical Internet Research, 26*(1), e57824. https://doi.org/10.2196/57824

Harpaz, R., DuMouchel, W., Schuemie, M., Bodenreider, O., Friedman, C., Horvitz, E., Ripple, A., Sorbello, A., White, R. W., Winnenburg, R., & Shah, N. H. (2017). Toward multimodal signal detection of adverse drug reactions. *Journal of Biomedical Informatics, 76*, 41–49. https://doi.org/10.1016/j.jbi.2017.10.013

Hendriksen, L. C., van der Linden, P. D., Lagro-Janssen, A. L. M., van den Bemt, P. M. L. A., Siiskonen, S. J., Teichert, M., Kuiper, J. G., Herings, R. M. C., Stricker, B. H., & Visser, L. E. (2021). Sex differences associated with adverse drug reactions resulting in hospital admissions. *Biology of Sex Differences, 12*(1), 34. https://doi.org/10.1186/s13293-021-00377-0

Hussain, S., Afzal, H., Saeed, R., Iltaf, N., & Umair, M. Y. (2021). Pharmacovigilance with transformers: A framework to detect adverse drug reactions using BERT fine-tuned with FARM. *Computational and Mathematical Methods in Medicine, 2021*, 5589829. https://doi.org/10.1155/2021/5589829

Jeong, E., Park, N., Choi, Y., Park, R. W., & Yoon, D. (2018). Machine learning model combining features from algorithms with different analytical methodologies to detect laboratory-event-related adverse drug reaction signals. *PLOS ONE, 13*(11), e0207749. https://doi.org/10.1371/journal.pone.0207749

Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data, 6*(1), 1–54. https://doi.org/10.1186/s40537-019-0192-5

Jung, J., Kim, J. H., Bae, J. H., Woo, S. S., Lee, H., & Shin, J. Y. (2024). A real-world pharmacovigilance study on cardiovascular adverse events of tisagenlecleucel using a machine learning approach. *Scientific Reports, 14*(1), 13641. https://doi.org/10.1038/s41598-024-64466-x

Kiguba, R., Isabirye, G., Mayengo, J., Owiny, J., Tregunno, P., Harrison, K., Pirmohamed, M., & Ndagije, H. B. (2024). Navigating duplication in pharmacovigilance databases: A scoping review. *BMJ Open, 14*(4), e081990. https://doi.org/10.1136/bmjopen-2023-081990

Kompa, B., Hakim, J. B., Palepu, A., Kompa, K. G., Smith, M., Bain, P. A., Woloszynek, S., Painter, J. L., Bate, A., & Beam, A. L. (2022). Artificial intelligence based on machine learning in pharmacovigilance: A scoping review. *Drug Safety, 45*(5), 477–491. https://doi.org/10.1007/s40264-022-01176-1

Lee, J. E., Kim, J. H., Bae, J. H., Song, I., & Shin, J. Y. (2022). Detecting early safety signals of infliximab using machine learning algorithms in the Korea adverse event reporting system. *Scientific Reports, 12*(1), 14869.https://doi.org/10.1038/s41598-022-18522-z

Li, F., Liu, W., & Yu, H. (2018). Extraction of information related to adverse drug events from electronic health record notes: Design of an end-to-end model based on deep learning. *JMIR Medical Informatics, 6*(4), e12159. https://doi.org/10.2196/12159

Liu, X., Rivera, S. C., Moher, D., Calvert, M. J., & Denniston, A. K. (2020). Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: The CONSORT-AI extension. *BMJ, 370*, m3164. https://doi.org/10.1136/bmj.m3164

Martin, G. L., Jouganous, J., Savidan, R., Bellec, A., Goehrs, C., Benkebil, M., Miremont, G., Micallef, J., Salvo, F., Pariente, A., & Létinier, L. (2022). Validation of artificial intelligence to support the automatic coding of patient adverse drug reaction reports using nationwide pharmacovigilance data. *Drug Safety, 45*(5), 535–548. https://doi.org/10.1007/s40264-022-01153-8

Medication Without Harm. (2017). *WHO*. https://www.who.int/initiatives/medication-without-harm

Nicholson, D. N., & Greene, C. S. (2020). Constructing knowledge graphs and their biomedical applications. *Computational and Structural Biotechnology Journal, 18*, 1414–1428. https://doi.org/10.1016/j.csbj.2020.05.017

Nikfarjam, A., Sarker, A., O'Connor, K., Ginn, R., & Gonzalez, G. (2015). Pharmacovigilance from social media: Mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. *Journal of the American Medical Informatics Association, 22*(3), 671–681. https://doi.org/10.1093/jamia/ocu041

Patient safety. (2023). *WHO*. https://www.who.int/news-room/fact-sheets/detail/patient-safety

Rivera, S. C., Liu, X., Chan, A. W., Denniston, A. K., & Calvert, M. J. (2020). Guidelines for clinical trial protocols for interventions involving artificial intelligence: The SPIRIT-AI extension. *BMJ, 370*. https://doi.org/10.1136/bmj.m3210

Salas, M., Petracek, J., Yalamanchili, P., Aimer, O., Kasthuril, D., Dhingra, S., Junaid, T., & Bostic, T. (2022). The use of artificial intelligence in pharmacovigilance: A systematic review of the literature. *Pharmaceutical Medicine, 36*(5), 295–306. https://doi.org/10.1007/s40290-022-00441-z

Shukla, D., Bhatt, S., Gupta, D., & Verma, S. (2024). Role of artificial intelligence in pharmacovigilance. *Journal of Drug Discovery and Health Sciences, 1*(4), 230–238. https://doi.org/10.21590/JDDHS.01.04.06

Stanovsky, G., Gruhl, D., & Mendes, P. N. (2017). Recognizing mentions of adverse drug reaction in social media using knowledge-infused recurrent models. *Proceedings of the Association for Computational Linguistics*, 1, 142–151. https://aclanthology.org/E17-1014/

The pharmacovigilance system for traditional medicine in Thailand. (2017). *WHO*. http://apps.who.int/bookorders

Van Calster, B., McLernon, D. J., Van Smeden, M., Wynants, L., & Steyerberg, E. W. (2019). Calibration: The Achilles heel of predictive analytics. *BMC Medicine, 17*(1), 230. https://doi.org/10.1186/s12916-019-1466-7

Vasey, B., Nagendran, M., Campbell, B., Clifton, D. A., Collins, G. S., Denaxas, S., Denniston, A. K., Faes, L., Geerts, B., Ibharim, M., Liu, X., Mateen, B. A., Mathur, P., McCradden, M. D., Morgan, L., Ordish, J., Rogers, C., Saria, S., Ting, D. S. W., Watkinson, P., Weber, W., Wheatstone, P., McCulloch, & the DECIDE-AI expert group. (2022). Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *Nature Medicine, 28*(5), 924–933. https://doi.org/10.1038/s41591-022-01772-9

Warner, J., Prada Jardim, A., & Albera, C. (2025). Artificial intelligence: Applications in pharmacovigilance signal management. *Pharmaceutical Medicine, 39*(3), 183–198. https://doi.org/10.1007/s40290-025-00561-2

Watson, S., Caster, O., Rochon, P. A., & den Ruijter, H. (2019). Reported adverse drug reactions in women and men: Aggregated evidence from globally collected individual case reports during half a century. *EClinicalMedicine, 17*, 100188. https://doi.org/10.1016/j.eclinm.2019.10.001

Wiens, J., Saria, S., Sendak, M., Ghassemi, M., Liu, V. X., Doshi-Velez, F., Jung, K., Heller, K., Kale, D., Saeed, M., Ossorio, P. N., Thadaney-Israni, S., & Goldenberg, A. (2019). Author correction: Do no harm: A roadmap for responsible machine learning for health care. *Nature Medicine, 25*(10), 1627.https://doi.org/10.1038/s41591-019-0609-x

Yang, X., Bian, J., Gong, Y., Hogan, W. R., & Wu, Y. (2019). MADEx: A system for detecting medications, adverse drug events, and their relations from clinical notes. *Drug Safety, 42*(1), 123–133. https://doi.org/10.1007/s40264-018-0761-0

Zhao, H., Ni, P., Zhao, Q., Liang, X., Ai, D., Erhardt, S., Wang, J., Li, Y., & Wang, J. (2023). Identifying the serious clinical outcomes of adverse reactions to drugs by a multi-task deep learning framework. *Communications Biology, 6*. 870. https://doi.org/10.1038/s42003-023-05243-w

Zitnik, M., Agrawal, M., & Leskovec, J. (2018). Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics, 34*(13), i457–i466. https://doi.org/10.1093/bioinformatics/bty294