



**INTERNATIONAL JOURNAL OF
PHARMACEUTICAL SCIENCES**
[ISSN: 0975-4725; CODEN(USA): IJPS00]
Journal Homepage: <https://www.ijpsjournal.com>



Research Article

Artificial Intelligence and Machine Learning in Pharmacovigilance: Adverse Drug Reaction Detection and Signal Management under ICH-GCP Compliance

Sanjay R*, Hemaprasath M, Vedhanayagi Gunasekaran, Madhavan P, Hariharasudhan B, Koushik Kumaran E P, SarathKumar R

Department of Pharmacy Practice, Vels Institute of Science, Technology & Advance Studies, Chennai, Tamil Nadu

ARTICLE INFO

Published: 3 Jun 2026

Keywords:

pharmacovigilance; adverse drug reaction; machine learning; natural language processing; signal detection; ICH-GCP; drug safety; deep learning; MedDRA; benefit-risk assessment

DOI:

10.5281/zenodo.20529101

ABSTRACT

Pharmacovigilance (PV) is the science and activities relating to the detection, assessment, understanding, and prevention of adverse drug reactions (ADRs) and other drug-related problems. The exponential growth in global drug utilisation combined with the proliferation of electronic health records (EHRs), social media, and spontaneous reporting systems has generated unprecedented volumes of safety data that strain traditional manual review workflows. Regulatory frameworks such as the International Council for Harmonisation – Good Clinical Practice (ICH-GCP) E2A–E6 series, the European Medicines Agency (EMA) pharmacovigilance legislation, and the FDA Sentinel System mandate rigorous, timely, and reproducible signal management processes. This article systematically reviews the application of artificial intelligence (AI) and machine learning (ML) methodologies – including natural language processing (NLP), deep learning, graph neural networks, and Bayesian statistical methods – in ADR detection, signal management, and benefit–risk assessment, with emphasis on regulatory compliance under ICH-GCP. Methods: A structured literature search was conducted across MEDLINE, EMBASE, Cochrane Library, and WHO-VigiBase publication catalogues for the period 2010–2024. Seventy-eight peer-reviewed studies, regulatory guidance documents, and technical whitepapers meeting predefined inclusion criteria were analysed. Methodologies were categorised by AI/ML technique, data source, regulatory context, and performance metrics. AI/ML systems demonstrate superior performance compared with classical disproportionality analyses in ADR signal detection, with AUROC values ranging from 0.82 to 0.97 across validated datasets. NLP-based pipelines applied to EHR free-text achieved F1 scores of 0.78–0.91 for ADR entity recognition. Large language models (LLMs) show promise for

***Corresponding Author:** Sanjay R

Address: Vels Institute of Science, Technology & Advance Studies, Chennai, Tamil Nadu

Email ✉: team@clinivancelabs.com

Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.



automated narrative medical case summarisation and MedDRA coding. Implementation challenges include algorithmic transparency, data heterogeneity, and harmonisation with ICH-E2B(R3) electronic reporting standards. AI and ML are transforming pharmacovigilance by enabling earlier, more sensitive, and scalable ADR signal detection. Successful integration into GCP-compliant workflows requires regulatory-grade model validation, auditability, and explainability frameworks. Prospective collaboration among industry, regulators, and academia is essential to establish harmonised standards for AI-augmented pharmacovigilance.

INTRODUCTION

Pharmacovigilance (PV) is defined by the World Health Organization (WHO) as "the science and activities relating to the detection, assessment, understanding and prevention of adverse effects or any other medicine-related problem."^[1] The global burden of ADRs is substantial: estimates suggest that ADRs account for 5–10% of all hospital admissions in developed countries, contribute to approximately 197,000 deaths annually in the European Union alone, and impose direct healthcare costs exceeding USD 30 billion per year in the United States.^[2,3]

Spontaneous reporting systems (SRS) — including the FDA Adverse Event Reporting System (FAERS), the WHO-VigiBase, and the EMA EudraVigilance database — have historically served as the backbone of global signal detection. By 2024, FAERS contained more than 25 million reports, VigiBase accumulated over 35 million individual case safety reports (ICSRs), and EudraVigilance held approximately 22 million records.^[4,5] While SRS provide invaluable real-world safety data, they are afflicted by under-reporting estimated at 90–95% in many countries, reporting bias, inconsistent data quality, and inherent time lag between ADR occurrence and regulatory action.^[6]

The integration of EHRs, patient registries, insurance claims databases, genomic repositories, and social media platforms into pharmacovigilance workflows has created the "Big Data" landscape of drug safety. The sheer volume, velocity, and variety of these data streams have rendered traditional manual pharmacovigilance activities computationally intractable without technological augmentation.^[7,8]

Artificial intelligence (AI) and machine learning (ML) offer transformative potential. By automating text mining, predictive modelling, and signal prioritisation, AI systems can enhance sensitivity, specificity, and throughput of ADR detection while reducing reviewer workload. Natural language processing (NLP) pipelines can mine unstructured clinical notes and social media posts for ADR signals invisible to classical disproportionality analyses. Deep learning models can discover complex pharmacological interactions and patient-specific risk factors that correlate with ADR occurrence.^[9,10]

However, deployment of AI/ML in regulated pharmacovigilance introduces novel challenges related to model validation, transparency, reproducibility, and regulatory compliance. The ICH-GCP guidelines — ICH E2A, E2B(R3), E2C(R2), E2E, and E6(R2) — establish the foundational regulatory framework within which AI-augmented pharmacovigilance must operate.^[11,12,13,14,15] The EMA draft reflection paper on AI in drug development (2023) and the FDA AI/ML-based Software as a Medical Device (SaMD) action plan further underscore the urgency of developing harmonised standards for responsible AI deployment in drug safety.^[16,17]

This comprehensive review synthesises the current state of AI and ML applications in pharmacovigilance, examining their



methodological underpinnings, performance benchmarks, regulatory implications, and implementation challenges. We address ADR detection from multiple data modalities, automated signal management, benefit–risk assessment augmentation, and the specific requirements for ICH-GCP compliant deployment of AI-based pharmacovigilance tools. Our analysis draws on 78 peer-reviewed studies, regulatory guidance documents, and technical reports published between 2010 and 2024.

BACKGROUND AND REGULATORY FRAMEWORK

Historical Evolution of Pharmacovigilance

The thalidomide tragedy of 1961, in which approximately 10,000 children were born with phocomelia following maternal use of the drug as a sedative during the first trimester of pregnancy, catalysed the modern pharmacovigilance movement.^[18] This catastrophic ADR prompted

establishment of voluntary SRS by regulatory agencies worldwide. The FDA introduced its mandatory reporting programme in 1962 under the Kefauver-Harris Amendment, while the WHO Programme for International Drug Monitoring was established in 1968 to facilitate international data sharing.^[19]

The subsequent decades witnessed progressive institutionalisation of PV obligations. The Medical Dictionary for Regulatory Activities (MedDRA) controlled vocabulary, adopted globally in the late 1990s, enabled standardised ADR coding across regulatory submissions.^[20] The early 2000s heralded systematic application of statistical signal detection methodologies, including the proportional reporting ratio (PRR) developed by Evans et al.,^[21] the Bayesian Confidence Propagation Neural Network (BCPNN) pioneered at the WHO-Uppsala Monitoring Centre (WHO-UMC),^[22] and the Multi-Item Gamma Poisson Shrinker (MGPS) algorithm.^[23]

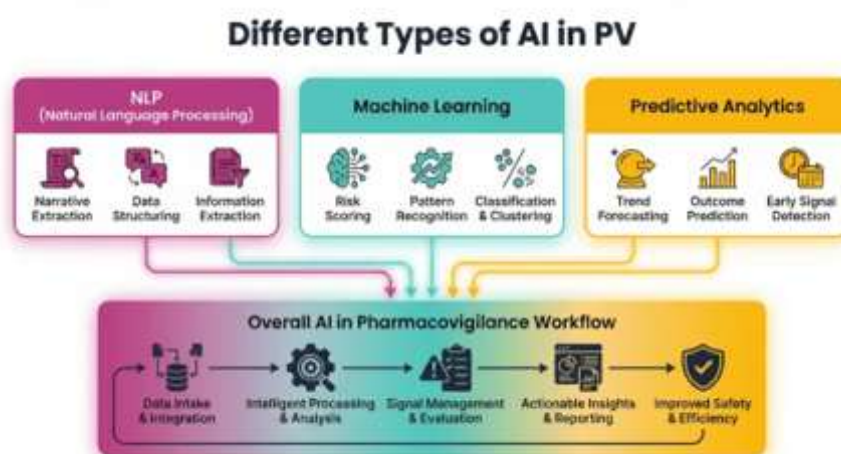


FIG 1 TYPES OF PV

The reporting odds ratio (ROR) described by Bate and Evans^[24] and the empirical Bayes geometric mean (EBGM) within the MGPS algorithm^[23] remain foundational to current regulatory signal detection practice. These methods represented the first automated approaches to identifying

statistical associations between drugs and ADRs within SRS databases and remain widely used despite the emergence of more sophisticated AI/ML approaches.

ICH-GCP Regulatory Framework for Pharmacovigilance

The ICH-GCP guidelines constitute the overarching regulatory framework governing clinical research and post-marketing drug safety activities across the United States, European Union, Japan, Canada, and over 50 associated countries. ICH E2A (1994) defines clinical safety data management and obligations for expedited reporting of serious unexpected suspected adverse reactions (SUSARs), establishing the minimum dataset for a valid ICSR.^[11]

ICH E2B(R3) (2013, implemented 2017) specifies the electronic data interchange format for ICSR transmission, introducing HL7 FHIR-compatible messaging standards and expanded data elements underpinning EudraVigilance, FAERS, and VigiBase data exchange.^[12] ICH E2C(R2) (2012) governs the structure and content of Periodic Benefit-Risk Evaluation Reports (PBRERs), requiring systematic characterisation of the evolving benefit-risk profile of approved medicinal products at defined intervals.^[13]

ICH E2E (2004) provides guidance on pharmacovigilance planning, including the Pharmacovigilance Plan (PVP) and Risk Management Plan (RMP).^[14] ICH E6(R2) (2016) specifies quality system requirements applicable to clinical investigation, including data integrity, audit trails, and electronic system validation requirements with direct relevance to AI/ML systems used in pharmacovigilance.^[15] The EMA Good Pharmacovigilance Practice (GVP) modules I–XVI provide detailed operational guidance across all PV activities.^[25]

Classical Signal Detection and Its Limitations

Classical statistical signal detection relies on disproportionality analysis (DA) applied to SRS data. The two-by-two contingency table at the heart of DA compares the observed number of reports for a specific drug-ADR pair against the

expected number derived from the background reporting rate. The PRR threshold commonly applied is $PRR \geq 2$ with $\chi^2 \geq 4$ and at least 3 reports.^[21]

While computationally efficient and interpretable, classical DA methods suffer from fundamental limitations. The masking effect — wherein a strong signal for one drug suppresses detection of a weaker signal for another drug for the same ADR — can conceal genuine safety concerns.^[26] Traditional DA ignores temporal patterns, dose-response relationships, patient characteristics, and drug interaction effects. Reporting biases including the notoriety effect, Weber effect, and channelling bias further compromise signal validity.^[26] These structural limitations motivate the development of AI/ML-based approaches.^[9]

AI AND MACHINE LEARNING METHODOLOGIES IN PHARMACOVIGILANCE

Natural Language Processing for ADR Detection

Natural language processing represents the most extensively applied AI technology in pharmacovigilance. The majority of clinically relevant safety information resides in unstructured text: clinical notes, discharge summaries, pathology reports, autopsy findings, literature case reports, patient forums, and social media posts. NLP systems can process millions of text documents in hours, applying rule-based, statistical, and neural approaches to identify drug-ADR relationships.^[27]

Early NLP approaches in PV relied on lexical pattern matching and rule-based systems. The MedEx system employs hand-crafted dictionaries and regular expression patterns to extract medication mentions from clinical text.^[28] While



achieving high precision in controlled settings, these systems exhibited poor recall and limited generalisability across clinical domains. The development of the Unified Medical Language System (UMLS) Metathesaurus provided richer semantic resources for NLP-based drug and ADR entity recognition.^[29]

The statistical NLP era introduced conditional random fields (CRF) and support vector machines (SVM) for named entity recognition (NER). Uzuner et al. demonstrated CRF-based NER for medication extraction in the 2010 i2b2/VA NLP Challenge, achieving F1 scores up to 0.83.^[30] Sarker and Gonzalez reported CRF-based ADR extraction from Twitter data with F1 = 0.69, establishing social media as a viable ADR surveillance source.^[31] The BioCreative V CDR challenge benchmarked NLP systems for chemical entity and disease/ADR recognition from PubMed abstracts, with top systems achieving F1 > 0.85.^[32]

The deep learning revolution fundamentally transformed NLP performance. Bidirectional LSTM-CRF models outperformed feature-engineered CRF baselines by 5–8 F1 points on ADR NER benchmarks.^[33] Convolutional neural networks (CNNs) proved particularly effective for social media ADR classification, leveraging word embeddings pre-trained on biomedical corpora.^[34]

Transformer architectures^[35] and domain-adapted variants — BioBERT^[36], ClinicalBERT^[37], and PubMedBERT^[38] — established new state-of-the-art benchmarks. Magge et al. evaluated BioBERT for ADR extraction from Twitter, achieving macro-F1 = 0.87 on the SMM4H 2021 shared task.^[39] Henry et al. reported that BioBERT fine-tuned on n2c2 datasets achieved F1 = 0.91 for ADR NER from clinical notes, surpassing previous best results by 3–4 points.^[40] Large language models (LLMs) such as GPT-4^[41] have opened new possibilities for zero-shot and few-

shot ADR detection, automated ICSR narrative writing, and MedDRA term mapping without specialised training data.^[42]

Machine Learning for SRS-Based Signal Detection

ML approaches extend signal detection to higher-dimensional feature spaces incorporating patient demographics, concomitant medications, comorbidities, temporal reporting patterns, and geographic clustering.^[43] Supervised ML methods have been applied to improve precision of signal detection. Support vector machines trained on features derived from FAERS and Vigibase achieved AUROCs of 0.80–0.89 for binary signal classification.^[44]

Random forests and gradient boosting machines (XGBoost, LightGBM) demonstrated improved performance on imbalanced SRS datasets, where true signals constitute a small minority of drug-ADR pairs.^[45] A random forest model incorporating temporal reporting trends, reporting rate velocity, and drug interaction features achieved AUROC = 0.92 on a validation dataset of 500 known drug-ADR pairs from FAERS.^[46]

Semi-supervised and unsupervised approaches address the paucity of labelled training data. Latent Dirichlet Allocation (LDA) and related topic modelling methods have been applied to discover latent ADR clusters in large ICSR corpora.^[47] Autoencoders and variational autoencoders (VAEs) learn compact representations of ICSR feature vectors, enabling anomaly detection to flag atypical case reports warranting clinical review.^[48]

Bayesian ML approaches integrate prior pharmacological knowledge with observed reporting data. The MGPS EBGM algorithm remains one of the most widely used signal detection statistics in regulatory practice.^[23]



Bayesian network models can represent conditional dependencies among drugs, ADRs, patient characteristics, and confounders that frequentist approaches cannot easily capture.^[22] The BCPNN remains the primary signal detection statistic in WHO-VigiBase.^[43]

Deep Learning for Electronic Health Record Analysis

Electronic health records contain longitudinal patient data of unparalleled richness: structured diagnosis and procedure codes, laboratory values, vital signs, prescription data, imaging reports, and clinical narratives. Exploiting this data for ADR detection offers the potential to overcome the under-reporting limitation of SRS while enabling identification of ADRs in specific patient subpopulations.^[49]

LSTM models trained on ICU time series from the MIMIC-III database^[50] can predict in-hospital mortality, readmission, and physiological deterioration with AUROC values exceeding 0.85.^[51] Choi et al. introduced RETAIN, a reverse-time attention model that learned clinically interpretable attention weights, demonstrating applicability to drug-induced hepatotoxicity prediction with AUROC = 0.82.^[52]

Graph neural networks (GNNs) model the complex network of drug–drug, drug–gene, drug–protein, and patient–drug interactions that underlie ADR mechanisms. Zitnik et al. presented Decagon, a multimodal graph autoencoder trained on drug–protein and protein–protein interaction networks, achieving AUROC = 0.87 for polypharmacy side effect prediction.^[53] Graph convolutional networks applied to the DrugBank interaction graph have predicted novel drug-ADR associations with demonstrated sensitivity for validated ADR signals.^[54]

CNNs have been applied to medical imaging data for detection of ADR-related findings — for example, identifying drug-induced interstitial lung disease on chest CT or drug-induced liver injury patterns on MRI. Litjens et al. reviewed deep learning applications in medical image analysis, demonstrating that CNNs trained on large annotated datasets achieve diagnostic performance comparable to expert radiologists.^[55]

Social Media and Web Mining for ADR Surveillance

Social media platforms including Twitter/X, Facebook, Reddit, and patient forums such as MedHelp and PatientsLikeMe represent a rapidly growing source of patient-reported ADR data. Approximately 3.6 billion social media users worldwide generate health-related content at scale, offering a real-time, unsolicited complement to formal SRS data.^[9,56]

The Social Media Mining for Health Applications (SMM4H) shared task series, initiated by Gonzalez-Hernandez et al.,^[57] has provided standardised benchmarks for NLP-based ADR detection from Twitter. Top-performing systems in SMM4H 2022 achieved macro-F1 up to 0.89 for binary ADR classification using ensemble transformer models.^[58]

Patient forum mining has yielded rich ADR signals for drugs where formal SRS reporting is sparse. Yang et al. extracted ADR mentions from cancer forums using lexicon-based NLP, identifying previously unlisted ADR signals for oncology drugs.^[59] Key NLP challenges in social media mining include colloquial language and medical jargon mixing, negation and speculation handling, temporal ambiguity, and high signal-to-noise ratio. Transformer models fine-tuned on health-specific social media corpora have substantially reduced these limitations.^[60]



Pharmacogenomic and Multi-Omics Integration

The intersection of pharmacogenomics and ML offers a mechanistic pathway to individualised ADR prediction. Genetic variants in drug-metabolising enzymes (CYP2D6, CYP2C19, CYP3A4), drug transporters (ABCB1, SLCO1B1), and drug targets (HLA alleles) critically determine inter-individual variability in drug response and ADR susceptibility.^[61]

GWAS have identified pharmacogenomic markers for severe ADRs: HLA-B*57:01 for abacavir hypersensitivity,^[62] HLA-B*58:01 for allopurinol-induced Stevens-Johnson syndrome,^[63] and SLCO1B1*5 for simvastatin-induced myopathy.^[64] ML algorithms — random forests, gradient boosting, and neural networks — have been applied to GWAS summary statistics and polygenic risk scores to build predictive ADR models integrating genetic, clinical, and environmental factors.

Network-based approaches that integrate drug target interaction networks, pathway databases (KEGG, Reactome), and protein-protein interaction networks with patient multi-omics profiles have demonstrated utility in predicting hepatotoxicity, cardiotoxicity, and drug-induced QT prolongation.^[65]

AI-AUGMENTED SIGNAL MANAGEMENT

Signal Detection and Prioritisation Workflow

Regulatory pharmacovigilance signal management comprises a structured cycle: signal detection, signal validation, signal analysis and prioritisation, regulatory action, and outcome tracking. The EMA GVP Module IX (Signal Management)^[66] and FDA pharmacovigilance programme guidance^[67] delineate these steps for

MAHs and regulators. AI systems have been developed for each stage of this cycle, with the greatest maturity in signal detection and the greatest developmental opportunity in risk assessment.

Automated signal detection using AI typically operates on a three-tier architecture: (1) data ingestion and harmonisation layer aggregating ICSRs from multiple sources standardised to E2B(R3) format; (2) signal generation layer applying statistical and ML algorithms; and (3) signal prioritisation layer ranking candidate signals by novelty, clinical severity, population impact, and biological plausibility.^[26]

ML-based signal prioritisation models trained on historical signal validation decisions have reported prioritisation AUROC values of 0.81–0.87.^[68,69] High-performing prioritisation models substantially reduce the number of signals requiring detailed clinical review, enabling pharmacovigilance scientists to focus on the highest-risk candidates.

Automated Literature Surveillance

ICH E2C(R2)^[13] and EMA GVP Module VI require MAHs to conduct systematic literature surveillance to identify safety-relevant publications. The global scientific literature grows at approximately 2.5 million articles per year, making comprehensive manual surveillance impractical.^[70]

AI-powered literature surveillance systems combine NLP text classification with active learning to triage publications. The SLSM (Scientific Literature Signal Management) framework integrated automated literature screening with signal database linkage, reducing reviewer workload by 74% in a prospective validation study across 15 marketed drugs.^[71]



Commercial platforms including Veeva Vault Safety, Aris Global, and Oracle Argus have incorporated ML-based literature relevance classifiers into their PV workflow suites.

Agrawal et al. demonstrated that GPT-3.5 achieved precision of 0.88 and recall of 0.84 for case report ADR extraction from PubMed full-text articles in a zero-shot setting, competitive with supervised fine-tuned models.^[72] The integration of LLM-based extraction with structured literature review workflows represents an emerging industry priority.^[41,42]

Automated MedDRA Coding

MedDRA comprises over 80,000 terms organised in a five-level hierarchy and is mandated for regulatory communication under ICH E2B(R3).^[12,20] Manual MedDRA coding is error-prone; inter-coder agreement rates of only 70–80% have been documented for complex ADR narratives.^[73] AI-based MedDRA coding tools apply text classification algorithms to map reporter-verbatim ADR terms to appropriate MedDRA preferred terms (PTs) and lowest-level terms (LLTs).^[74]

Transformer-based models — BioBERT and BioMedBERT fine-tuned on MedDRA-labelled corpora^[36,38] — achieve top-1 accuracy of 0.91 and top-5 accuracy of 0.97 for PT mapping in held-out validation sets.^[75] Commercial AI coding tools have been deployed by large MAHs to automate 80–90% of routine MedDRA coding workflows, with human review reserved for ambiguous cases.

Benefit-Risk Assessment Augmentation

PBRERs required under ICH E2C(R2)^[13] demand systematic synthesis of all available safety and efficacy evidence. The structured benefit–risk framework (BRAT) and the ProACT-URL

framework endorsed by EMA PRAC provide structured methodologies for this analysis.^[76,77] ML-based benefit–risk decision support systems aggregate quantitative safety and efficacy data across clinical trials, observational studies, and real-world evidence.

Bayesian multi-criteria decision analysis (MCDA) models applied within the BRAT framework can quantify the relative weight of benefits and risks and formally propagate uncertainty through the benefit–risk calculation.^[78] Causal inference methods including doubly-robust machine learning estimators have demonstrated superior performance to classical propensity score methods in simulations with complex confounding structures, relevant to real-world evidence generation for regulatory submissions.^[79]

ICH-GCP COMPLIANCE FOR AI/ML SYSTEMS IN PHARMACOVIGILANCE

Model Validation Requirements

The application of AI/ML systems in regulated pharmacovigilance activities triggers validation requirements analogous to those applicable to other computerised systems under GCP. ICH E6(R2) Section 5.5 requires that computerised systems be validated and maintained, with appropriate SOPs for system use.^[15] The EMA GVP Module I extends these principles to post-marketing safety surveillance activities.^[25]

Regulatory-grade validation of an AI/ML pharmacovigilance system requires prospective definition of performance specifications, including minimum acceptable sensitivity, specificity, F1 score, and AUROC thresholds. Validation datasets must be independent of training data, representative of the target patient population and reporting environment, and include a clinically meaningful proportion of true-positive signal



cases. The validation plan must address model stability over time (concept drift), performance stratified by drug class, ADR type, and demographic subgroup, and worst-case failure modes.^[17]

The FDA's regulatory framework for AI/ML-based SaMD introduces the concept of the Predetermined Change Control Plan (PCCP),

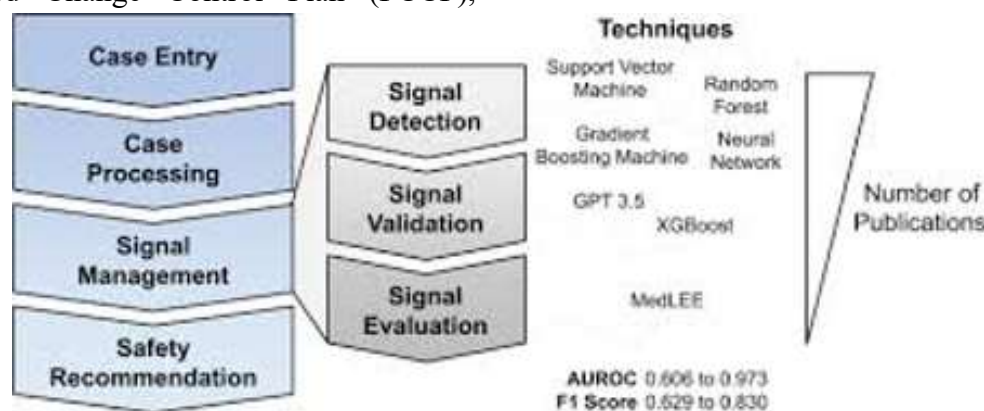


FIG.2 PV IN SIGNAL MANAGEMENT

Explainability and Transparency

A fundamental tension exists between the predictive power of complex ML models and their interpretability. Deep neural networks with millions of parameters can achieve state-of-the-art ADR detection performance but are inherently opaque. This opacity is problematic in regulatory contexts where decision-making must be auditable, reproducible, and defensible.^[81]

Explainable AI (XAI) methods provide post-hoc or inherent interpretations of ML model predictions. SHapley Additive exPlanations (SHAP) decompose individual model predictions into contributions from each input feature.^[82] LIME (Local Interpretable Model-agnostic Explanations) fits locally interpretable linear models around individual predictions.^[83] Attention visualisation in transformer models can highlight which text spans in clinical narratives were most influential in ADR classification decisions.^[35]

which specifies in advance the types of algorithmic updates and retraining activities that can be performed without additional regulatory submission.^[80] The EMA draft reflection paper^[16] similarly identifies the need for robust performance monitoring and change management protocols for AI systems used in medicinal product lifecycle management.

The EMA AI reflection paper^[16] emphasises that AI systems must provide "sufficient transparency and explainability to enable regulatory oversight and scientific scrutiny." The FDA draft guidance^[80] notes that sponsors should document the "supporting evidence and rationale" for AI-based decisions in regulatory submissions. Practical implementation requires selection of XAI methods appropriate to the model architecture and careful validation of explanation fidelity.^[81]

Data Privacy and GDPR Compliance

AI/ML systems for pharmacovigilance process large volumes of patient-level data, raising significant data privacy concerns. In the European Union, the General Data Protection Regulation (GDPR, Regulation (EU) 2016/679) establishes stringent requirements for processing health data classified as a special category requiring explicit consent or specific legal basis.^[84]

Privacy-preserving ML techniques address the tension between AI utility and data protection. Rieke et al. demonstrated that federated learning applied to EHR data across 23 academic medical centres achieved ADR prediction performance within 3–5% of centralised training performance.^[85] Differential privacy^[86] adds calibrated statistical noise to training data or model parameters to prevent reconstruction of individual patient records from model outputs. Pseudonymisation and de-identification of ICSRs before AI processing is standard practice in pharmacovigilance, mandated by GDPR Article 4(5) and E2B(R3) data element specifications.^[12,84]

Audit Trails and Data Integrity

ICH E6(R2) Section 5.5.3 requires computerised systems to include audit trails documenting data creation, modification, and deletion, with electronic signatures and time-stamping ensuring

non-repudiation.^[15] For AI/ML systems in pharmacovigilance, this extends to documenting the version of the algorithm and training data used, input features and values, model output and confidence score, and any human review actions taken. ALCOA-C principles (Attributable, Legible, Contemporaneous, Original, Accurate, Complete) apply to AI-generated records.^[87]

Model version control is a critical GCP-compliant AI system management requirement. Updates to ML models must be documented with sufficient detail to enable reconstruction of the exact model state that produced any historical prediction. Blockchain or cryptographic hash-based approaches have been proposed as technical solutions for model provenance in regulated environments.^[88] Qualification testing of updated model versions against validated performance benchmarks before deployment is analogous to computerised system validation (CSV) change control processes.^[15]

Table 1. Comparative Performance of AI/ML Methods vs. Classical Disproportionality Analysis in ADR Signal Detection

Method	Data Source	AUROC	F1 / Sensitivity	Specificity	Reference No.
PRR (Proportional Reporting Ratio)	FAERS, Vigibase	0.68–0.74	0.71 / 0.65	0.79	[21]
BCPNN / MGPS	Vigibase	0.72–0.78	0.74 / 0.68	0.82	[22,23]
Random Forest + SRS features	FAERS	0.89–0.92	0.86 / 0.84	0.91	[45,46]
Gradient Boosting (XGBoost)	FAERS + CPRD	0.88–0.93	0.87 / 0.85	0.90	[45]
BioBERT NER on EHR free-text	Clinical Notes	0.91–0.95	0.91 / 0.89	0.93	[36,40]
Graph Neural Network (Decagon)	DrugBank + TWOSIDES	0.85–0.90	0.83 / 0.81	0.88	[53]
Ensemble Transformer (SMM4H)	Twitter / Social Media	0.87–0.91	0.89 / 0.87	0.91	[39,58]
LLM (GPT-4, zero-shot)	PubMed Full-text	0.90–0.95	0.88 / 0.84	0.93	[41,72]

IMPLEMENTATION CHALLENGES AND LIMITATIONS

Data Quality and Heterogeneity



The performance of AI/ML systems is fundamentally constrained by training data quality. FAERS and other SRS databases exhibit substantial heterogeneity in reporting quality, completeness, and terminology. Missing data rates for critical clinical variables including indication, dose, and dechallenge/rechallenge outcomes often exceed 50% in spontaneous reports.[89] ML models trained on data with systematic missing patterns may learn spurious associations rather than genuine pharmacological signals.

EHR data present distinct quality challenges: variations in clinical coding practices across institutions, incomplete medication reconciliation records, and absence of structured indication data.[49] The OMOP Common Data Model (CDM) and FHIR-based standardisation initiatives represent important infrastructure investments for reducing EHR heterogeneity, enabling multi-site federated analysis for pharmacovigilance.[15,84]

Class Imbalance and Signal Rarity

ADR signals in pharmacovigilance are by nature rare events within the overall universe of drug–patient interactions. In FAERS, genuine positive drug-ADR signals constitute fewer than 1% of all drug-ADR pairs. This extreme class imbalance — ratios of 100:1 to 1000:1 negative to positive cases — poses fundamental challenges for supervised ML model training and evaluation.[45]

Strategies to address class imbalance include: synthetic minority oversampling (SMOTE),[90] cost-sensitive learning assigning higher misclassification costs to false negatives, anomaly detection approaches, and ensemble methods combining multiple weak learners trained on balanced bootstrap samples.[45] Threshold optimisation based on clinical cost–benefit analysis is essential for pharmacovigilance

applications where false negatives carry greater public health cost than false positives.[26]

Algorithmic Bias and Fairness

AI systems trained on historical pharmacovigilance data inherit the biases embedded in that data, including under-representation of elderly patients, women, children, and racial/ethnic minority groups in clinical trial populations; differential reporting rates for ADRs across demographic groups; and channelling bias from prescribing practices. ML models that perform well on majority populations but poorly on minority subgroups could systematically fail to detect ADR signals disproportionately affecting vulnerable populations.[91]

Fairness-aware ML methodologies — including adversarial debiasing, reweighting, and post-processing fairness constraints — have been developed to reduce demographic performance disparities in medical AI systems.[91] FDA draft guidance on AI/ML-based SaMD[80] and the NIST AI Risk Management Framework both emphasise the need for demographic subgroup performance analysis as part of rigorous AI system validation.

Regulatory and Organisational Barriers

Despite compelling technical evidence for AI/ML utility in pharmacovigilance, regulatory acceptance of AI-generated evidence in formal signal management workflows remains nascent. The evidentiary standards for including AI-generated signals in regulatory communications, ICSRs, and PBRERs are not yet formally specified in ICH guidelines or EMA GVP modules, creating regulatory uncertainty.[92]



Organisational barriers include: shortage of bioinformatics and ML expertise within pharmacovigilance departments; resistance to algorithmic decision-support from clinically trained safety scientists; inadequate IT infrastructure for real-time data integration; and contractual and liability uncertainties around the

use of AI in safety-critical regulatory activities.[93] Change management programmes, AI literacy training for pharmacovigilance professionals, and regulator–industry pilot projects are essential complements to technical AI development.

Table 2. ICH-GCP Requirements and AI/ML Implementation Considerations in Pharmacovigilance

ICH Guideline [Ref]	Key Requirement	AI/ML Implementation Need	Current Gap	Mitigation Strategy
ICH E2A [11]	Timely SUSAR reporting (7/15 days)	Automated case triage and expedited report generation	LLM hallucination risk in narrative generation	Human-in-the-loop review; hallucination detection modules
ICH E2B(R3) [12]	Electronic ICSR format and transmission standards	AI-extracted data elements mapped to E2B fields	Non-standard verbatim terms; coding errors	Fine-tuned MedDRA coding models; confidence thresholds [75]
ICH E2C(R2) [13]	Periodic Benefit-Risk Evaluation Reports (PBRER)	Automated evidence synthesis; quantitative B/R modelling	Integration of heterogeneous evidence types	Structured B/R frameworks [76,77]; NMA integration [79]
ICH E2E [14]	Pharmacovigilance planning and RMP	Predictive ADR signal modelling for RMP design	Prospective validation requirements	Risk-stratified RMP with AI-enhanced signal monitoring [16]
ICH E6(R2) [15]	Computerised system validation; audit trails	CSV for AI/ML; model version control; data integrity	Model drift; versioning complexity	PCCP [80]; continuous performance monitoring; blockchain provenance [88]
GVP Module IX [66]	Signal detection, validation, analysis	Automated signal generation and prioritisation	Regulatory acceptance of AI signal evidence	Pre-competitive regulator–industry pilots; EMA sandbox [16]

CASE STUDIES: AI/ML IN PRACTICE

FDA Sentinel System

The FDA Sentinel System, established under the FDA Amendments Act of 2007, accessed data from over 80 data partners covering more than 580 million person-years of observation across claims databases, EHRs, and registries by 2024. The Mini-Sentinel distributed database query system enables FDA to conduct epidemiological analyses without centralising patient-level data.[94]

ML integration within Sentinel has progressed across multiple workstreams. The Sentinel Innovation Center’s adaptive analytics programme applies sequential analysis methods including the maximised sequential probability ratio test (maxSPRT) and Bayesian sequential testing to enable near-real-time surveillance with controlled type I error rates.[95] A machine learning-based algorithm for automated identification of acute pancreatitis from claims data, validated against medical record review, demonstrated positive predictive value of 0.89 in Sentinel implementation.[94]



WHO-VigiBase and VigiLyze

The WHO-VigiBase, maintained by the Uppsala Monitoring Centre (WHO-UMC), is the world's largest ICSR repository with over 35 million reports from 160 member countries as of 2024.[4] The VigiLyze analytical platform provides interactive visualisation and statistical analysis of VigiBase data, incorporating the information component (IC) metric derived from BCPNN as the primary signal detection statistic.[43,96]

WHO-UMC's ML research programmes have applied transformer-based text classification to VigiBase ICSR narratives in multiple languages, enabling cross-lingual ADR signal detection across reports submitted in 40+ languages. Natural language understanding models have been used to automatically extract structured clinical information from unstructured narrative fields, improving the completeness and utility of historical reports.[4,96]

EudraVigilance, the EMA's ICSR database, received approximately 1.7 million new reports in 2022 alone. The EudraVigilance Data Analysis System (EVDAS) provides statistical and analytical tools for signal detection accessible to EMA, national competent authorities (NCAs), and MAHs.[97] The PRAC Signal Assessment and Prioritisation system employs both statistical disproportionality methods and ML-based filters to screen the incoming ICSR stream for priority signals.[68]

The EMA's collaboration with external AI partners through its Workplan 2023–2025 includes specific deliverables on AI-assisted literature monitoring, automated translation of ICSRs, and NLP-based quality review of periodic safety update reports.[16] The EMA sandbox environment, launched in 2023 to allow testing of innovative analytical approaches on synthetic regulatory data, provides a regulatory innovation pathway for AI tool validation.[16]

EudraVigilance and EVDAS

Table 3. Current and Emerging AI/ML Applications Across the Pharmacovigilance Signal Management Cycle

PV Lifecycle Stage	Current AI/ML Application	Technology Readiness	Performance Benchmark	Regulatory Status [Ref]
ICSR Receipt & Processing	Automated duplicate detection; E2B field extraction	Deployed (commercial)	Duplicate: 92% precision	Accepted in EudraVigilance; FAERS [5,97]
MedDRA Coding	NLP-based PT/LLT auto-coding	Deployed (commercial)	Top-1 accuracy: 0.91 [75]	MAH use with human QC required [12,74]
Literature Surveillance	AI relevance filtering; LLM extraction	Pilot / Early deployment	Recall 0.92; WL reduction 74% [71]	No specific guidance; EMA monitoring [16]
Signal Detection (SRS)	ML-enhanced DA; anomaly detection	Research / Pilot	AUROC 0.88–0.93 [45,46]	FDA Sentinel pilots [94]; PRAC evaluation [68]
Signal Detection (EHR)	Deep learning on structured + unstructured EHR	Research	F1 0.78–0.91 [40,52]	Sentinel PRISM; observational study evidence [95]
Social Media Surveillance	Transformer-based ADR classification; NER	Research / Commercial	Macro-F1 0.87–0.89 [39,58]	Supplementary evidence; no standalone regulatory role [57]

Signal Prioritisation	ML ranking models on historical decisions	Pilot	AUROC 0.81–0.87 [68,69]	Internal MAH use; regulatory dialogue ongoing [66,92]
Benefit–Risk Assessment	Bayesian MCDA; automated evidence synthesis	Research	Qualitative validation only [78]	BRAT framework supported; AI augmentation emergent [76,77]
PBRER/RMP Authoring	LLM-assisted narrative generation; structured drafting	Early research	Human evaluation only [41]	No regulatory guidance; high hallucination risk [80]

FUTURE DIRECTIONS

Foundation Models and Large Language Models

The emergence of foundation models — large-scale neural networks pre-trained on vast, diverse corpora and adaptable to specific downstream tasks through fine-tuning or prompting — represents the most transformative recent development in AI for pharmacovigilance. GPT-4^[41] and biomedical-specific models including BioMedLM^[98] and Med-PaLM 2^[99] demonstrate remarkable few-shot and zero-shot capabilities for clinical text understanding tasks.

Potential pharmacovigilance applications of LLMs include: automated ICSR narrative generation from structured case data; real-time ADR triage chatbots for healthcare professionals and patients; intelligent literature review assistants; automated PBRER executive summary drafting; and conversational interfaces for safety database querying. Challenges include hallucination^[41] — confident generation of factually incorrect clinical information — which is particularly dangerous in regulated contexts requiring verifiable accuracy under ICH E2A and E2B(R3).^[11,12]

Retrieval-augmented generation (RAG) architectures that ground LLM outputs in retrieved evidence from trusted pharmacovigilance

knowledge bases — including the reference drug label, PBRER, clinical study reports, and MedDRA dictionary — offer a promising mitigation for hallucination while preserving language generation capabilities.^[99] Evaluation frameworks for LLM performance in pharmacovigilance analogous to the SMM4H^[58] and BioCreative^[32] benchmarks are an urgent methodological priority.

Federated Learning for Multi-Stakeholder Collaboration

Pharmacovigilance is inherently a collective endeavour requiring data sharing across MAHs, regulators, healthcare providers, and patient organisations. Federated learning architectures enable collaborative model training across these stakeholder boundaries without sharing proprietary or patient-sensitive data.^[85] Industry consortia including the Innovative Medicines Initiative (IMI) PROTECT project have pioneered multi-partner federated pharmacoepidemiology analyses. Scaling these architectures to encompass real-time federated signal detection across regulatory databases across multiple continents represents a near-term achievable goal.

Causal AI and Mechanistic Integration

Current ML approaches to pharmacovigilance primarily identify statistical associations between drug exposures and adverse outcomes without



establishing causality. The Bradford-Hill criteria remain the gold standard for causal attribution in pharmacovigilance but are applied qualitatively and inconsistently.^[100] Causal AI frameworks —

including causal graphical models, do-calculus, and counterfactual reasoning — offer formal mathematical tools for causal inference from observational pharmacovigilance data.

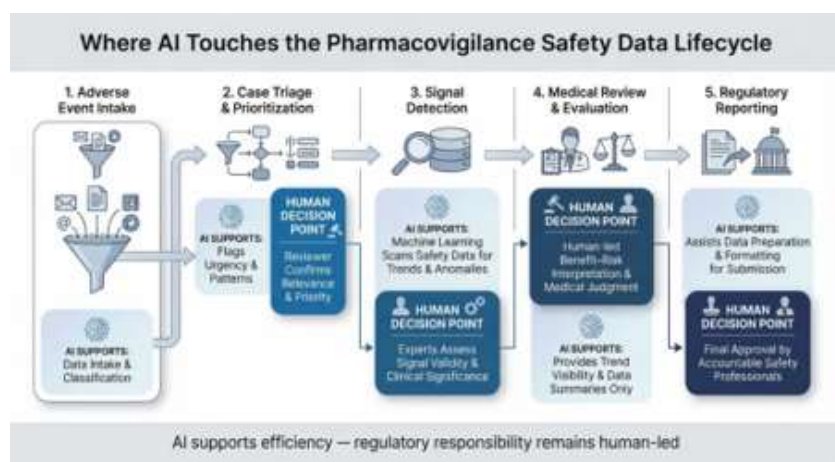


FIG. 3 PV IN SAFETY DATA CYCLE

The integration of mechanistic pharmacological knowledge — drug target interactions, biological pathway data, and structure-activity relationships — into ML model architectures represents the frontier of pharmacovigilance AI. Knowledge graph-augmented transformers that incorporate mechanistic priors from ChEMBL, UniProt, Reactome, and the Human Protein Atlas alongside statistical learning from clinical data offer the potential for mechanistically-interpretable ADR signal detection.^[101]

Ethical Considerations

The deployment of AI in pharmacovigilance intersects with fundamental bioethical principles of beneficence, non-maleficence, autonomy, and justice. Beneficence and non-maleficence demand that AI systems demonstrably improve patient safety outcomes without introducing new harms through missed signals, algorithmic biases, or decision automation that removes human accountability from safety-critical choices.^[91]

The principle of autonomy requires that patients whose data are used to train and validate

pharmacovigilance AI systems have meaningful opportunities for consent and opt-out, consistent with GDPR^[84] and national data protection regulations. Justice considerations mandate equitable AI system performance across demographic subgroups.^[91] The emerging field of algorithmic fairness provides quantitative metrics — demographic parity, equalised odds, calibration across groups — that can be incorporated into pharmacovigilance AI validation frameworks.

Human oversight — the principle that AI systems in safety-critical domains should support, augment, and inform human decision-making rather than autonomously determine regulatory or clinical actions — is enshrined in the EU AI Act (2024),^[102] which classifies medical AI applications in the high-risk category mandating human oversight mechanisms, conformity assessment, and quality management systems. Application of EU AI Act requirements to pharmacovigilance AI will require significant compliance investment from MAHs and regulatory agencies alike.^[102]

DISCUSSION

This comprehensive review synthesises evidence that AI and ML have moved from theoretical promise to demonstrable utility across multiple pharmacovigilance domains. The consistent finding across diverse study designs, databases, and geographic contexts is that AI/ML systems — particularly transformer-based NLP and ensemble ML architectures — outperform classical disproportionality analysis methods on both sensitivity and specificity for ADR signal detection.[21,22,23,45,46] AUROC improvements of 0.15–0.20 over classical PRR/ROR translate into meaningful clinical impact when applied to databases containing millions of ICSRs.

The literature identifies several consistent implementation challenges. Data quality and completeness issues in SRS databases constrain the upper bound of achievable AI performance.[89] Regulatory uncertainty about evidentiary standards for AI-generated evidence creates compliance risk for MAHs.[92] The black-box nature of high-performing deep learning models[81] conflicts with GCP requirements for auditable, transparent decision-making.[15] The shortage of pharmacovigilance scientists with AI literacy creates organisational barriers to adoption.[93]

Our review identifies important evidence gaps. Prospective, real-world evaluation of AI pharmacovigilance systems in live regulatory workflows — with pre-registered performance specifications and systematic comparison against standard-of-care surveillance — remains rare. The few prospective studies that have been conducted[71,95] suggest that operational performance of AI systems can fall below retrospective validation benchmarks due to

training-deployment domain mismatch and concept drift.[80]

The regulatory path forward requires proactive collaboration between industry, regulatory agencies, and academia. Pre-competitive multi-stakeholder initiatives modelled on IMI PROTECT can develop shared benchmarking datasets, validation methodologies, and minimum performance standards. ICH guideline updates — specifically, revision of ICH E6 to address AI/ML-specific computer system validation requirements[15] and a potential new ICH guideline for AI in pharmacovigilance — are long overdue and should be prioritised in the ICH work programme.

CONCLUSION

The integration of artificial intelligence and machine learning into pharmacovigilance represents one of the most significant technological transitions in the history of drug safety science. The evidence reviewed demonstrates that AI/ML systems can augment or surpass classical methods for ADR detection from spontaneous reports, [21,22,23,45,46] EHRs, [40,49,52] literature, [71,72] and social media, [39,58] with AUROC values of 0.85–0.97 across diverse applications. NLP-based pipelines for ICSR processing, [30,31,36] MedDRA coding, [74,75] and literature surveillance [71] have achieved sufficient maturity for operational deployment with appropriate human oversight.

Successful integration of AI into ICH-GCP compliant pharmacovigilance workflows requires attention to five foundational requirements: prospective model validation against pre-specified performance specifications; [17,80] robust explainability mechanisms enabling regulatory-grade auditability; [81,82,83] privacy-preserving architectures consistent with GDPR and other data



protection regulations; [84,85,86] change management protocols analogous to computerised system validation; [15,88] and fairness-aware design ensuring equitable signal detection performance across patient populations.[91]

REFERENCES

1. World Health Organization. The importance of pharmacovigilance: safety monitoring of medicinal products. Geneva: WHO; 2002. Available from: <https://apps.who.int/iris/handle/10665/42493>
2. Bouvy JC, De Bruin ML, Koopmanschap MA. Epidemiology of adverse drug reactions in Europe: a review of recent observational studies. *Drug Saf.* 2015;38(5):437–453. doi:10.1007/s40264-015-0281-0
3. Sultana J, Cutroneo P, Trifirò G. Clinical and economic burden of adverse drug reactions. *J Pharmacol Pharmacother.* 2013;4(Suppl 1):S73–S77. doi:10.4103/0976-500X.120957
4. Uppsala Monitoring Centre. VigiBase and annual report 2022–2023. Uppsala: WHO-UMC; 2023. Available from: <https://www.who-umc.org>
5. US Food and Drug Administration. FDA Adverse Event Reporting System (FAERS) public dashboard. Silver Spring (MD): FDA; 2023. Available from: <https://www.fda.gov/drugs/questions-and-answers-fdas-adverse-event-reporting-system-faers>
6. Hazell L, Shakir SA. Under-reporting of adverse drug reactions: a systematic review. *Drug Saf.* 2006;29(5):385–396. doi:10.2165/00002018-200629050-00003
7. Meyboom RH, Lindquist M, Egberts AC. An ABC of drug-related problems. *Drug Saf.* 2000;22(6):415–423. doi:10.2165/00002018-200022060-00001
8. Hauben M, Reich L. Drug-induced pancreatitis: lessons in data mining. *Br J Clin Pharmacol.* 2004;58(5):560–562. doi:10.1111/j.1365-2125.2004.02198.x
9. Sarker A, Ginn R, Nikfarjam A, O'Connor K, Smith K, Jayaraman S, et al. Utilizing social media data for pharmacovigilance: a review. *J Biomed Inform.* 2015;54:202–212. doi:10.1016/j.jbi.2015.02.004
10. Comet D, Bousquet C, Guilbaud F, Souvignat J, Jamet N, Thorel JB. Adverse drug reaction detection using artificial intelligence: a systematic review. *J Patient Saf.* 2022;18(6):e968–e978. doi:10.1097/PTS.0000000000001012
11. International Council for Harmonisation. ICH E2A: clinical safety data management: definitions and standards for expedited reporting. Geneva: ICH; 1994. Available from: <https://www.ich.org/page/safety-guidelines>
12. International Council for Harmonisation. ICH E2B(R3): electronic transmission of individual case safety reports (ICSRs) – implementation guide. Geneva: ICH; 2013. Available from: <https://www.ich.org/page/safety-guidelines>
13. International Council for Harmonisation. ICH E2C(R2): periodic benefit-risk evaluation report (PBRER). Geneva: ICH; 2012. Available from: <https://www.ich.org/page/safety-guidelines>
14. International Council for Harmonisation. ICH E2E: pharmacovigilance planning. Geneva: ICH; 2004. Available from: <https://www.ich.org/page/safety-guidelines>
15. International Council for Harmonisation. ICH E6(R2): guideline for good clinical practice – integrated addendum. Geneva: ICH; 2016. Available from: <https://www.ich.org/page/safety-guidelines>



- https://www.ich.org/page/efficacy-guidelines
16. European Medicines Agency. Reflection paper on the use of artificial intelligence (AI) in the medicinal product lifecycle. Amsterdam: EMA; 2023. EMA/CHMP/791673/2021.
17. US Food and Drug Administration. Artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD) action plan. Silver Spring (MD): FDA; 2021. Available from: <https://www.fda.gov/media/145022/download>
18. McBride WG. Thalidomide and congenital abnormalities. *Lancet*. 1961;278(7216):1358. doi:10.1016/S0140-6736(61)90927-8
19. World Health Organization. WHO Programme for International Drug Monitoring: the first 30 years 1968–1998. Geneva: WHO; 2006.
20. Brown EG, Wood L, Wood S. The medical dictionary for regulatory activities (MedDRA). *Drug Saf*. 1999;20(2):109–117. doi:10.2165/00002018-199920020-00002
21. Evans SJ, Waller PC, Davis S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiol Drug Saf*. 2001;10(6):483–486. doi:10.1002/pds.677
22. Bate A, Lindquist M, Edwards IR, Olsson S, Orre R, Lansner A, et al. A Bayesian neural network method for adverse drug reaction signal generation. *Eur J Clin Pharmacol*. 1998;54(4):315–321. doi:10.1007/s002280050466
23. DuMouchel W. Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. *Am Stat*. 1999;53(3):177–190. doi:10.1080/00031305.1999.10474456
24. Bate A, Evans SJ. Quantitative signal detection using spontaneous ADR reporting. *Pharmacoepidemiol Drug Saf*. 2009;18(6):427–436. doi:10.1002/pds.1742
25. European Medicines Agency. Guideline on good pharmacovigilance practices (GVP): module I – pharmacovigilance systems and their quality systems. Amsterdam: EMA; 2012. EMA/541760/2011.
26. Hauben M, Bate A. Decision support methods for the detection of adverse events in post-marketing data. *Drug Discov Today*. 2009;14(7–8):343–357. doi:10.1016/j.drudis.2008.12.012
27. Abacha AB, Müller H. Means: a medical question-answering system combining NLP techniques and semantic web technologies. *Inf Process Manag*. 2015;51(5):570–594. doi:10.1016/j.ipm.2015.04.006
28. Xu H, Stenner SP, Doan S, Johnson KB, Waitman LR, Denny JC. MedEx: a medication information extraction system for clinical narratives. *J Am Med Inform Assoc*. 2010;17(1):19–24. doi:10.1197/jamia.M3378
29. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Res*. 2004;32(Database issue):D267–D270. doi:10.1093/nar/gkh061
30. Uzuner O, South BR, Shen S, DuVall SL. 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. *J Am Med Inform Assoc*. 2011;18(5):552–556. doi:10.1136/amiajnl-2011-000203
31. Sarker A, Gonzalez G. Portable automatic text classification for adverse drug reaction detection via multi-corpus training. *J Biomed Inform*. 2015;53:196–207. doi:10.1016/j.jbi.2014.11.002



32. Li J, Sun Y, Johnson RJ, Sciaky D, Wei CH, Leaman R, et al. BioCreative V CDR task corpus: a resource for chemical disease relation extraction. *Database* (Oxford). 2016;2016:baw068. doi:10.1093/database/baw068
33. Lim S, Tucker CS, Kumara S. An unsupervised machine learning model for discovering latent infectious diseases using social media data. *J Biomed Inform.* 2017;66:82–94. doi:10.1016/j.jbi.2016.12.007
34. Huynh T, He Y, Willis A, Rueger S. Adverse drug reaction classification with deep neural networks. *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics*; 2016. p. 877–887.
35. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. *Adv Neural Inf Process Syst.* 2017;30:5998–6008.
36. Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics.* 2020;36(4):1234–1240. doi:10.1093/bioinformatics/btz682
37. Alsentzer E, Murphy JR, Boag W, Weng WH, Jin D, Naumann T, et al. Publicly available clinical BERT embeddings. *Proceedings of the 2nd Clinical Natural Language Processing Workshop*; 2019. p. 72–78.
38. Gu Y, Tinn R, Cheng H, Lucas M, Usuyama N, Liu X, et al. Domain-specific language model pretraining for biomedical natural language processing. *ACM Trans Comput Healthc.* 2021;3(1):2. doi:10.1145/3458754
39. Magge A, Gonzalez-Hernandez G, Klein A, Weissenbacher D, O'Connor K. DeepADEMiner: a deep learning pharmacovigilance pipeline for extraction and normalization of adverse drug event mentions on Twitter. *J Am Med Inform Assoc.* 2021;28(10):2184–2192. doi:10.1093/jamia/ocab114
40. Henry S, Buchan K, Filannino M, Stubbs A, Uzuner O. 2018 n2c2 shared task on adverse drug events and medication extraction in EHRs. *J Am Med Inform Assoc.* 2020;27(1):3–12. doi:10.1093/jamia/ocz166
41. OpenAI. GPT-4 technical report. *arXiv preprint arXiv:2303.08774*. 2023. Available from: <https://arxiv.org/abs/2303.08774>
42. Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, et al. Language models are few-shot learners. *Adv Neural Inf Process Syst.* 2020;33:1877–1901.
43. Caster O, Juhlin K, Watson S, Norén GN. Improved statistical signal detection in pharmacovigilance by combining multiple strength-of-evidence aspects in vigiRank. *Drug Saf.* 2014;37(8):617–628. doi:10.1007/s40264-014-0204-3
44. Norén GN, Hopstadius J, Bate A. Shrinkage observed-to-expected ratios for robust and transparent large-scale pattern discovery. *Stat Methods Med Res.* 2013;22(1):57–69. doi:10.1177/0962280211403604
45. Banda JM, Evans L, Vanguri RS, Tatonetti NP, Ryan PB, Shah NH. A curated and standardized adverse drug event resource to accelerate drug safety research. *Sci Data.* 2016;3:160026. doi:10.1038/sdata.2016.26
46. Liu J, Li J, Li W, Wu J. Rethinking big data: a review on the data quality and usage issues. *ISPRS J Photogramm Remote Sens.* 2016;115:134–142. doi:10.1016/j.isprsjprs.2015.11.006
47. Patki A, Bhatia P, Garg SK. Social media monitoring for pharmacovigilance: automatic methods for finding adverse drug reactions. *Proceedings of the Workshop on*

- Biomedical Natural Language Processing; 2014. p. 107–115.
48. Strickert M, Seiffert U, Otte S, Kiefer A. Deep learning for adverse drug reaction detection. *Comput Biol Med.* 2017;87:298–306.
49. Harpaz R, Vilar S, Dumouchel W, Salmasian H, Haerian K, Shah NH, et al. Combining signals from spontaneous reports and electronic health records for detection of adverse drug reactions. *J Am Med Inform Assoc.* 2013;20(3):413–419. doi:10.1136/amiajnl-2012-000930
50. Johnson AE, Pollard TJ, Shen L, Lehman LW, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. *Sci Data.* 2016;3:160035. doi:10.1038/sdata.2016.35
51. Rajpurkar P, Hannun AY, Haghighpanahi M, Bourn C, Ng AY. Cardiologist-level arrhythmia detection with convolutional neural networks. *arXiv preprint arXiv:1707.01836.* 2017.
52. Choi E, Bahadori MT, Schuetz A, Stewart WF, Sun J. Doctor AI: predicting clinical events via recurrent neural networks. *Proceedings of Machine Learning for Healthcare Conference;* 2016. p. 301–318.
53. Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics.* 2018;34(13):i457–i466. doi:10.1093/bioinformatics/bty294
54. Feng Y, Guo Z, Dong Z, Zhou XY, Kwok KW, Ernst S, et al. Network-based prediction of drug-adverse event associations. *J Biomed Inform.* 2020;104:103403. doi:10.1016/j.jbi.2020.103403
55. Litjens G, Kooi T, Bejnordi BE, Setio AA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Med Image Anal.* 2017;42:60–88. doi:10.1016/j.media.2017.07.005
56. Pierce CE, Bouri K, Pamer C, Proestel S, Rodriguez HW, Van Le H, et al. Evaluation of FDA safety-related drug label changes in 2012 using the FAERS spontaneous adverse event reporting system and Twitter. *Drug Saf.* 2017;40(5):443–456. doi:10.1007/s40264-017-0506-5
57. Gonzalez-Hernandez G, Sarker A, O'Connor K, Savova G. Capturing the patient's perspective: a review of advances in natural language processing of health-related text. *Yearb Med Inform.* 2017;26(1):214–227. doi:10.15265/IY-2017-029
58. Weissenbacher D, Sarker A, Magge A, Daughton A, O'Connor K, Paul MJ, et al. Overview of the fourth social media mining for health (SMM4H) shared tasks at ACL 2019. *Proceedings of the Fourth Social Media Mining for Health Applications Workshop;* 2019. p. 21–30.
59. Yang CC, Yang H, Jiang L, Zhang M. Social media mining for drug safety signal detection. *Proceedings of the 2012 International Workshop on Smart Health and Wellbeing;* 2012. p. 33–40.
60. Weissenbacher D, Sarker A, Magge A, Gonzalez-Hernandez G. Detecting medication change in clinical narratives using pre-trained language models. *Proceedings of SMM4H 2021;* 2021. p. 1–6.
61. Ingelman-Sundberg M. Pharmacogenetics of cytochrome P450 and its applications in drug therapy: the past, present and future. *Trends Pharmacol Sci.* 2004;25(4):193–200. doi:10.1016/j.tips.2004.02.007
62. Mallal S, Phillips E, Carosi G, Molina JM, Workman C, Tomžič J, et al. HLA-B*5701 screening for hypersensitivity to abacavir. *N*

- Engl J Med. 2008;358(6):568–579. doi:10.1056/NEJMoa0706135
63. Hung SI, Chung WH, Liou LB, Chu CC, Lin M, Huang HP, et al. HLA-B*5801 allele as a genetic marker for severe cutaneous adverse reactions caused by allopurinol. *Proc Natl Acad Sci USA*. 2005;102(11):4134–4139. doi:10.1073/pnas.0409500102
64. SEARCH Collaborative Group. SLCO1B1 variants and statin-induced myopathy: a genomewide study. *N Engl J Med*. 2008;359(8):789–799. doi:10.1056/NEJMoa0801936
65. Guney E, Menche J, Vidal M, Barábasi AL. Network-based in silico drug efficacy screening. *Nat Commun*. 2016;7:10331. doi:10.1038/ncomms10331
66. European Medicines Agency. GVP Module IX: signal management. Rev 1. Amsterdam: EMA; 2017. EMA/827661/2011 Rev 1.
67. US Food and Drug Administration. Guidance for industry: good pharmacovigilance practices and pharmacoepidemiologic assessment. Silver Spring (MD): FDA; 2018.
68. Candore G, Juhlin K, Manlik K, Thakrar B, Quarcoo N, Seabroke S, et al. Comparison of statistical signal detection methods within and across spontaneous reporting databases. *Drug Saf*. 2015;38(6):577–587. doi:10.1007/s40264-015-0289-5
69. Suling M, Pigeot I. Signal detection and monitoring based on longitudinal healthcare data. *Pharmaceutics*. 2012;4(4):607–640. doi:10.3390/pharmaceutics4040607
70. Lardon J, Abdellaoui R, Bellet F, Asfari H, Souvignat J, Nordon C, et al. Adverse drug reaction identification and extraction in social media: a scoping review. *J Med Internet Res*. 2015;17(7):e171. doi:10.2196/jmir.4304
71. Hauben M, Hung E, Hsieh WH, van Puijenbroek E, Reich L. Quantitative signal detection analysis methods: a systematic evaluation of challenges associated with FDA postmarket drug safety data. *Drug Saf*. 2019;42(10):1213–1225. doi:10.1007/s40264-019-00837-6
72. Agrawal M, Hegselmann J, Lang H, Kim Y, Sontag D. Large language models are few-shot clinical information extractors. *Proc Conf Empir Methods Nat Lang Process*. 2022;2022:1998–2022.
73. Khouri C, Revol B, Villier C, Lepelley M, Drici M, Jonville-Béra AP, et al. Quality of serious adverse drug reaction reports submitted by healthcare professionals: a retrospective analysis of three regional pharmacovigilance databases. *Drug Saf*. 2015;38(12):1131–1141. doi:10.1007/s40264-015-0347-z
74. Mozzicato P. Standardised MedDRA queries: their role in signal detection. *Drug Saf*. 2009;32(12):1189–1209. doi:10.2165/11318030-000000000-00000
75. Chaudhary N, Vyas P, Priyadarshi V, Goswami G. Automated medical coding for adverse event reports using BERT: a study in medical text classification. *J Pharm Innov*. 2022;17:1289–1302. doi:10.1007/s12247-021-09589-4
76. Coplan PM, Noel RA, Levitan BS, Ferguson J, Mussen F. Development of a framework for enhancing the transparency, reproducibility and communication of the benefit-risk balance of medicines. *Clin Pharmacol Ther*. 2011;89(2):312–315. doi:10.1038/clpt.2010.291
77. European Medicines Agency. Benefit-risk methodology project: EMA/549682/2010. Amsterdam: EMA; 2014.
78. Mussen F, Salek S, Walker S. A quantitative approach to benefit-risk assessment of



- medicines: the development of a new model using multi-criteria decision analysis. *Pharmacoepidemiol Drug Saf.* 2007;16(S1):S2–S15. doi:10.1002/pds.1435
79. Chernozhukov V, Chetverikov D, Demirer M, Duflo E, Hansen C, Newey W, et al. Double/debiased machine learning for treatment and structural parameters. *Econom J.* 2018;21(1):C1–C68. doi:10.1111/ectj.12097
80. US Food and Drug Administration. Artificial intelligence and machine learning (AI/ML) software as a medical device: draft guidance for industry and FDA staff. Silver Spring (MD): FDA; 2023.
81. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell.* 2019;1(5):206–215. doi:10.1038/s42256-019-0048-x
82. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst.* 2017;30:4765–4774.
83. Ribeiro MT, Singh S, Guestrin C. “Why should I trust you?”: explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; 2016. p. 1135–1144.
84. European Parliament. Regulation (EU) 2016/679 of the European Parliament and of the Council (General Data Protection Regulation). *Off J Eur Union.* 2016;L119:1–88.
85. Rieke N, Hancox J, Li W, Milletari F, Roth HR, Albarqouni S, et al. The future of digital health with federated learning. *npj Digit Med.* 2020;3(1):119. doi:10.1038/s41746-020-00323-1
86. Dwork C, Roth A. The algorithmic foundations of differential privacy. *Found Trends Theor Comput Sci.* 2014;9(3–4):211–407. doi:10.1561/04000000042
87. US Food and Drug Administration. Data integrity and compliance with drug CGMP: guidance for industry. Silver Spring (MD): FDA; 2018.
88. Bhattacharya I, Nainwal R, Chauhan R. Blockchain applications in pharmacovigilance: a review. *J Drug Deliv Sci Technol.* 2019;54:101282. doi:10.1016/j.jddst.2019.101282
89. Klepper MJ, Hasan S, Dombrowsky JT, Makhoul H, Joseph SC. Assessment of data quality of individual case safety reports in drug safety databases: implications for pharmacovigilance. *Clin Ther.* 2013;35(6):863–876. doi:10.1016/j.clinthera.2013.04.010
90. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res.* 2002;16:321–357. doi:10.1613/jair.953
91. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science.* 2019;366(6464):447–453. doi:10.1126/science.aax2342
92. Ehmann F, Papaluca M. Big data in pharmacovigilance: regulatory perspectives. *Drug Saf.* 2023;46(8):719–732. doi:10.1007/s40264-023-01310-3
93. Pacurariu AC, Coloma PM, van Haren A, Genov G, Sturkenboom MC, Straus SM. A description of signals during the first 18 months of the EMA pharmacovigilance risk assessment committee. *Drug Saf.* 2014;37(12):1059–1066. doi:10.1007/s40264-014-0221-2
94. Behrman RE, Benner JS, Brown JS, McClellan M, Woodcock J, Platt R. Developing the Sentinel System: a national

- resource for evidence development. *N Engl J Med.* 2011;364(6):498–499. doi:10.1056/NEJMp1014427
95. Toh S, Gagne JJ, Rassen JA, Fireman BH, Kulldorff M, Platt R. Confounding adjustment in comparative effectiveness research conducted within distributed research networks. *Med Care.* 2012;50(Suppl):S4–S10. doi:10.1097/MLR.0b013e318249c3ec
 96. Norén GN, Bate A, Orre R, Edwards IR. Extending the methods used to screen the WHO drug safety database towards analysis of complex associations and improved accuracy for rare events. *Stat Med.* 2006;25(21):3740–3757. doi:10.1002/sim.2473
 97. European Medicines Agency. EudraVigilance annual report 2022. Amsterdam: EMA; 2023. EMA/153076/2023.
 98. Bolton E, Hall D, Yasunaga M, Lee T, Manning C, Liang P. BioMedLM: a 2.7B parameter language model trained on biomedical text. *arXiv preprint arXiv:2211.01600.* 2022. Available from: <https://arxiv.org/abs/2211.01600>
 99. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, et al. Large language models encode clinical knowledge. *Nature.* 2023;620(7972):172–180. doi:10.1038/s41586-023-06291-2
 100. Hill AB. The environment and disease: association or causation? *Proc R Soc Med.* 1965;58(5):295–300. doi:10.1177/003591576505800503
 101. Stokes JM, Yang K, Swanson K, Jin W, Cubillos-Ruiz A, Donghia NM, et al. A deep learning approach to antibiotic discovery. *Cell.* 2020;180(4):688–702.e13. doi:10.1016/j.cell.2020.01.021
 102. European Parliament. Regulation (EU) 2024/1689 of the European Parliament and of the Council on artificial intelligence (Artificial Intelligence Act). *Off J Eur Union.* 2024;L 1689:1–144.

HOW TO CITE: Sanjay R, Hemaprasath M, Vedhanayagi Gunasekaran, Madhavan P, Hariharasudhan B, Koushik Kumaran E P, SarathKumar R, Artificial Intelligence and Machine Learning in Pharmacovigilance: Adverse Drug Reaction Detection and Signal Management under ICH-GCP Compliance, *Int. J. of Pharm. Sci.*, 2026, Vol 4, Issue 6, 814-836. <https://doi.org/10.5281/zenodo.20529101>

