REVIEW ARTICLE

UPI JOURNAL OF CHEMICAL AND LIFE SCIENCES (UPI-JCLS)

ISSN: 2581-4648 (An International online Peer Reviewed Open Access Journal)

www.uniquepubinternational.com Published by Unique Pub International (UPI)

AUTOMATION IN PHARMACOVIGILANCE – A REVIEW

Boddepalli Rakesh*[1], VC Randeep Raj[2]

*[1]Student ,Avanthi Institute Pharmaceutical Sciences, vizianagaram
[2]Associate Professor, Avanthi Institute of Pharmaceutical Sciences, vizianagaram[2]

***Corresponding Author**
Boddepalli Rakesh

Received: 15 July 2025 Revised: 24 Aug 2025 Accepted: 03 September 2025

Abstract

The pharmacovigilance (PV) function is undergoing a rapid transformation as automated technologies-ranging from robotic process automation (RPA) and natural language processing (NLP) to machine learning (ML) and large language models (LLMs)-are adopted across the safety lifecycle. Drivers include rising case volumes, expanding real-world data (RWD) sources, evolving global data standards, and increasing regulatory expectations for quality, timeliness, and traceability. This review synthesizes how automation is reshaping case intake and processing, coding and deduplication, literature surveillance, signal detection and evaluation, and regulatory reporting. It also outlines governance and validation practices for compliant deployment, and highlights emerging opportunities and limitations.

Keywords: automation, pharmacovigilance, ADR.

INTRODUCTION

Pharmacovigilance ensures the detection, assessment, understanding, and prevention of adverse effects or other drug-related problems across a product's life cycle. Global PV systems now manage millions of individual case safety reports (ICSRs) transmitted through standardized electronic formats into national and international repositories such as FDA's FAERS and the WHO Uppsala Monitoring Centre's VigiBase® [1–4]. Concurrent growth in healthcare digitization and advanced analytics has catalyzed automation of PV workflows, promising efficiency gains and earlier detection of safety signals with robust quality controls [5,6].

REGULATORY AND DATA-STANDARDS FOUNDATIONS FOR AUTOMATION

Automation in PV is enabled by harmonized data standards and guidance:

- **ICH E2B(R3) defines the XML** message structure and data elements for electronic ICSR exchange, facilitating straight-through processing from intake systems to regulators and global databases. [1,2,7,8].
- **FAERS E2B (R3) implementation** resources and testing environments support industry preparation for electronic submissions at scale, a prerequisite for reliable automation in U.S. reporting [3,9,10].
- **EMA Good Pharmacovigilance Practices (GVP) Module IX** provides a structured framework for signal management-detection, validation, confirmation, analysis, prioritization, and recommendation-that technology solutions must align with when automating tasks [11–13].
- **WHO Programme for International Drug Monitoring and VigiBase** exemplify standardized, quality-checked global aggregation; automated screening, controlled vocabularies, and completeness checks enhance downstream analytics [4,14].

These frameworks reduce ambiguity, allowing RPA/NLP tools and ML models to operate on consistent, machine-readable safety data and to automate routing, validation, and exchange.

DRIVERS AND BENEFITS OF AUTOMATION

Automation targets three persistent PV challenges. First, **volume and velocity:** case counts rise due to broader reporting channels (EHRs, patient portals), expanded indications, and global market penetration [5,6].Second, **data heterogeneity:** ICSRs arrive as unstructured narratives, scanned PDFs, emails, and call-center transcripts. Third, **compliance pressure:** strict timelines and auditability expectations require reproducible, traceable processes.

Implemented properly, automation can reduce manual touchpoints, shorten cycle times, improve coding consistency, surface earlier signals, and reallocate expert time to high-value medical assessment [5,6,15].

AUTOMATION ACROSS THE PV WORKFLOW

Case Intake, Triage, and De-duplication

RPA can capture cases from emails, portals, and partner gateways, perform completeness checks (e.g., minimum criteria), and convert inputs into E2B(R3) structures for validation and routing [1–3]. **NLP** classifies reportability, extracts key entities (suspect drug, event, seriousness, dates), and assigns priority. De-duplication leverages rules and **probabilistic/ML matchers** trained on identifiers, narrative similarity, and temporal proximity; VigiBase's long-standing automated signal and matching approaches illustrate the feasibility and value of statistical automation at scale.(4,14,16)

Medical Coding and Data Standardization

Automated coders use NLP and terminology services to map narratives to **MedDRA®** preferred terms and concomitant products to WHO Drug; ML models learn common phrasings and synonyms to improve recall and precision. Consistency in coding strengthens disproportionate-analysis signal detection later in the pipeline [5,6].

Literature and External Source Monitoring

Automated literature surveillance systems continuously harvest citations from bibliographic databases and apply **topic models** and **classification** to flag potentially reportable articles. Models fine-tuned on PV relevance decrease manual screening volume, enabling safety scientists to focus on borderline cases and clinical assessment. Similar techniques support monitoring of regulatory websites and safety communications, with audit trails for inspection readiness [5,6,15].

Case Processing and Medical Review Augmentation

Structured data population, case narrative summarization, and seriousness/expectedness checks can be partially automated. **Decision-support** dashboards surface inconsistencies (e.g., chronology conflicts) and prompt targeted human review. Early studies show LLMs can assist with drafting medically-relevant summaries from case narratives, provided robust human oversight, versioning, and validation guardrails are in place [5,6,20].

Signal Detection and Evaluation

Automation has perhaps the richest history in signal detection. **Disproportionality methods**—such as Proportional Reporting Ratios (PRR), Empirical Bayes Geometric Mean (EBGM), and the Bayesian Confidence Propagation Neural Network (BCPNN)-are routinely automated across global databases to prioritize drug–event pairs for review [14,16]. Newer **machine-learning pipelines** (gradient boosting, random forests, deep learning) incorporate covariates (age, sex, reporter type), temporal patterns, or multiple data sources (spontaneous, literature, EHR) to refine prioritization and reduce false positives. Automation also supports **workflow orchestration** across GVP Module IX steps-tracking detection, validation rationales, and benefit–risk impact-while preserving full traceability [11,12].

Regulatory Reporting and Submissions

With **E2B(R3)** ubiquitous, automated gateways validate, transform, and transmit ICSRs to FAERS/EudraVigilance, manage acknowledgments, and reconcile errors with minimal manual intervention. FDA's **pre-production testing** and regional implementation guides help sponsors industrialize their automated pipelines safely [3,9,10,21].

ARCHITECTURES AND TECHNICAL BUILDING BLOCKS

Modern PV automation solutions combine:

- **Ingestion & RPA:** email/API connectors, OCR for scanned content, format normalization, de-duplication queues [1–3].
- **NLP/LLM Services:** entity extraction (drug, indication, event, dose, dates), negation/temporality detection, narrative summarization, and MedDRA mapping [5,6,20].
- **Analytics & ML:** disproportionality engines, supervised classifiers for case relevance, anomaly detection for spikes or outliers, and causal-inference inspired approaches for structured RWD [5,6,15].
- **Standards & Terminologies:** E2B(R3) for exchange, MedDRA/WHO-Drug dictionaries, and controlled vocabularies for product and reporter attributes [1,2,4].
- **Governance & Audit:** model versioning, validation packs, monitoring, explainabilityartifacts, and role-based access controls integrated with quality management systems (QMS). [11,12,20].

VALIDATION, COMPLIANCE, AND HUMAN OVERSIGHT

Regulators emphasize that automation must not erode scientific integrity or traceability. Best practices include:

- **Intended Use & Risk Assessment:** define the automation's role (assistive vs. autonomous) and criticality of decisions it influences [11,12, 20].
- **Computer System Validation (CSV) / Computer Software Assurance (CSA):** qualify data pipelines, NLP/ML components, and RPA scripts; test against representative edge cases and maintain change control.

- **Performance Monitoring:** track precision/recall for case relevance classifiers, coding concordance vs. human gold standards, and stability of signal-ranking outputs over time; implement drift detectionand periodic re-validation [5,6,20].
- **Human-in-the-Loop:** retain medical review for causality, expectedness, and labeling impact; automation provides recommendations and triage, not conclusions, unless explicitly validated and permitted [11,12,20].
- **Bias and Explainability:** document training data, feature importance, and known limitations; provide rationale snippets (e.g., narrative spans) for flagged outputs to support reviewer trust [5,6,20].

IMPLEMENTATION ROADMAP AND CHANGE MANAGEMENT

Organizations moving from manual or semi-manual operations to advanced automation typically follow a stepped approach:

1. **Stabilize Standards & Gateways:** ensure robust E2B(R3) compliance, acknowledgments handling, dictionary management, and error reconciliation for straight-through electronic reporting. [1–3,9].
2. **Automate High-Volume, Low-Risk Tasks:** intake, deduplication, and coding with strong back-up by human QC to build confidence and measure ROI.
3. **Augment Medical Review and Literature Screening:** deploy NLP ranking and LLM summarization with strict SOPs and sampling-based QC, documenting concordance and intervention rates [5,6,20].
4. **Industrialize Signal Workflows:** embed automated detection with GVP IX-aligned workflow states and audit trails, linking signals to aggregate reports and labeling governance [11–13].
5. **Continuous Improvement:** monitor KPIs (cycle time, first-pass yield, coding deviation rate, signal time-to-decision); institute model lifecycle management and retraining cadences [5,6,20)].

OPPORTUNITIES AND FRONTIERS

- **Foundation models for PV:** domain-adapted LLMs can draft narratives, rationalize case classifications, and harmonize multi-lingual reports; guardrails like retrieval-augmented generation (RAG) can ground outputs in source evidence [5,6,20].
- **Integrated RWD analytics:** linking spontaneous reports with EHR claims and registries can improve signal specificity; semi-automated causal frameworks may help differentiate confounding and background incidence [15].
- **Proactive risk management:** near-real-time monitoring, automated change detection, and scenario simulation can surface safety issues earlier and support targeted pharmacovigilance activities [5,6].

LIMITATIONS AND RISKS

Automation is not a panacea. **Data quality** issues (incomplete narratives, misclassification), **class imbalance** (rare events), and **reporting biases** persist. Over-reliance can introduce automation bias, where users defer to model outputs, and concept drift can degrade performance as products and reporters change. Finally, interoperability and governance overheads—terminology updates, software validation, SOP changes—require sustained investment [5,6,20].

CONCLUSION

Automation is now foundational to efficient, compliant pharmacovigilance. Anchored in standards like E2B(R3) and aligned with GVP Module IX, modern PV organizations can safely automate intake, coding, literature screening, and signal detection, while preserving expert judgment for clinical interpretation and benefit–risk decisions. The highest-value programs pair robust engineering with transparent validation and continuous monitoring, ensuring that automation enhances, rather than replaces, pharmacovigilance science.

REFERENCES

1. ICH E2B(R3) Individual Case Safety Report (ICSR) specification and related files. International Council for Harmonisation (ICH).
2. ICH guideline E2B(R3) on electronic transmission of ICSRs—Implementation Guide. European Medicines Agency (EMA).
3. FDA. FAERS Electronic Submissions – E2B(R3) Standards (premarket and postmarket resources).
4. Uppsala Monitoring Centre (WHO). About VigiBase—quality assurance and automated checks.
5. Artificial intelligence in pharmacovigilance: a narrative review and practical applications. (Recent review).
6. Artificial intelligence in pharmacovigilance: advancing drug safety operations and ethics. (Recent review).
7. FDA. E2B(R3) Electronic Transmission of ICSRs—Implementation Guide & data elements (overview page).
8. EMA. ICH E2B(R3) electronic transmission of ICSRs—scientific guideline page.
9. FDA. FAERS Electronic Submissions (main page for electronic submissions to FAERS).
10. FDA. Preparing for the electronic exchange of safety reports—FAERS E2B(R3) pre-production testing (PDF).
11. EMA. Guideline on good pharmacovigilance practices (GVP) Module IX—Signal management (Rev. 1).

12. EMA. Good pharmacovigilance practices (GVP) overview page (Module IX details and links).
13. TriNetX. GVP Module IX for Signal Management: The Complete Guide (overview of process alignment).
14. WHO Programme for International Drug Monitoring (WHO-PIDM) / VigiBase information page (WHO site).
15. Artificial intelligence and big data for pharmacovigilance and patient safety. (ScienceDirect overview of efficiency gains).
16. Paediatric safety signals identified in VigiBase—methods and results (discussion of BCPNN and automated screening at UMC).
17. CIOMS Working Group XIV (Draft report). Artificial Intelligence in Pharmacovigilance—governance, automation bias, and validation considerations.
18. ProPharma/industry commentary on FDA FAERS E2B(R3) harmonization timelines and preparations.