Journal of Comprehensive Pharmaceutical Sciences
Online peer reviewed Journal | Published by COGNIX PRESS
Content available at www.cognixpress.in Online ISSN: 3139-5163

REVIEW ARTICLE
OPEN ACCESS

MACHINE LEARNING-BASED PHARMACOVIGILANCE FOR EARLY DETECTION OF ADVERSE DRUG REACTION SIGNALS

M.MANOORYA*, R.KOUSHIK, Y.SAI PRIYA, B.YAMINI, K. NAVYA, PURVI NETAM

Department of Pharmacy Practice, Avanthi Institute of Pharmaceutical Sciences, Vizianagaram.

Corresponding Author
M.Manoorya

DOI: https://doi.org/10.67396/jcps.v1i2.41

Article History: 25.06.2026 Received: 02.07.2026 Revised: Accepted: 21.08.2026

ABSTRACT

Pharmacovigilance is an essential component of healthcare systems that focuses on the detection, assessment, understanding, and prevention of adverse drug reactions (ADRs) and other medicine-related problems.[1] Conventional pharmacovigilance primarily relies on spontaneous reporting systems and statistical disproportionality methods, which may be affected by under-reporting, incomplete information, reporting bias, and the increasing volume of safety data.[2] The rapid development of artificial intelligence (AI), machine learning (ML), natural language processing (NLP), and real-world data analytics provides new opportunities for improving the efficiency and sensitivity of pharmacovigilance.[3] Machine learning algorithms can analyze large and heterogeneous datasets to identify complex relationships between medicines and adverse events that may not be readily recognized using conventional approaches. Algorithms such as random forest, support vector machine, gradient boosting, logistic regression, neural networks, and deep learning have been investigated for adverse event detection and safety signal identification.[4,5] Recent evidence indicates that ML-based approaches may complement traditional disproportionality methods and facilitate earlier identification and prioritization of potential safety signals.[6] However, challenges involving data quality, model interpretability, external validation, bias, privacy, and regulatory acceptance remain important. This article discusses the principles, workflow, applications, advantages, limitations, and future potential of ML-based pharmacovigilance for early ADR signal detection.

Keywords: *Pharmacovigilance; adverse drug reactions; machine learning; artificial intelligence; safety signal; pharmacovigilance databases; natural language processing; adverse drug events.*

1. INTRODUCTION

Medicines provide substantial benefits in the prevention and treatment of disease; however, their use may also result in unintended and harmful reactions. Pharmacovigilance is defined by the World Health Organization (WHO) as the science and activities associated with the detection, assessment, understanding, and prevention of adverse drug effects or other drug-related problems [1]. Effective pharmacovigilance is therefore essential for maintaining a favorable benefit–risk profile throughout the life cycle of a medicinal product.

Traditional pharmacovigilance relies substantially on spontaneous adverse event reporting systems. These systems provide an important source of post-marketing safety information but have several limitations, including under-reporting, duplicate reports, incomplete clinical information, stimulated reporting, and reporting biases [2]. As the number of medicines and adverse-event reports increases, manual evaluation of individual reports and safety signals becomes increasingly challenging.

Data-mining techniques have consequently become an important component of modern pharmacovigilance. Conventional approaches such as the **Reporting Odds Ratio (ROR)** and **Proportional Reporting Ratio (PRR)** can identify disproportionate reporting of specific drug–event combinations [7]. However, these methods may have limitations when dealing with high-dimensional, heterogeneous, and temporally complex data.

Machine learning provides an additional analytical approach. ML models can learn patterns from historical datasets and subsequently classify, predict, rank, or prioritize potential safety signals. A scoping review of 393 publications found that safety-signal detection represented a major application of ML in pharmacovigilance, although adoption of advanced ML methods remained limited [4]. More recent evidence suggests growing interest in random forest, gradient boosting, neural networks, and NLP-based approaches [6].

2. CONCEPT OF MACHINE LEARNING-BASED PHARMACOVIGILANCE

Machine learning is a branch of AI in which computational algorithms learn patterns from data without requiring every decision rule to be explicitly programmed.

In pharmacovigilance, an ML system can be designed to learn associations between:

Drug → Patient characteristics → Exposure → Adverse event → Outcome

The overall process can be represented as:

Pharmacovigilance Data
↓
Data Cleaning & Pre-processing
↓
Feature Extraction
↓
Training Dataset
↓
Machine Learning Model
↓
Prediction / Classification
↓
Potential Drug–ADR Association
↓
Signal Prioritization
↓
Clinical &Pharmacovigilance Review
↓
Signal Validation
↓
Regulatory / Safety Action

Importantly, an ML-generated signal should generally be considered a **hypothesis requiring validation**, rather than definitive proof that a drug caused an ADR. Signal management includes detection, validation, evaluation, and final assessment of potential safety signals [6].

3. DATA SOURCES FOR ML-BASED PHARMACOVIGILANCE

ML models can integrate information from multiple sources.

3.1 Spontaneous Reporting Systems

Large pharmacovigilance databases contain reports submitted by healthcare professionals, patients, and other sources. Such datasets can contain information regarding:

- Suspected medicine
- Adverse event
- Patient demographics
- Concomitant medicines
- Indication
- Dose
- Outcome
- Seriousness
- Reporter information

Spontaneous-reporting data remain one of the most important sources for post-marketing signal detection [2].

3.2 Electronic Health Records

Electronic health records (EHRs) provide longitudinal clinical information and can potentially identify adverse events that are not captured through spontaneous reporting. NLP and ML methods can be applied to clinical narratives to identify drug exposures and adverse events [8].

3.3 Biomedical Literature

Published scientific literature represents another valuable source of pharmacovigilance information. NLP systems can automatically identify relationships among drugs, diseases, symptoms, and adverse events from large volumes of literature.

3.4 Social Media and Patient-Generated Data

Patient-generated content can provide early information regarding symptoms and treatment experiences. NLP-based pharmacovigilance approaches have been investigated for extracting potential ADR information from user-generated content [8].

4. MACHINE LEARNING ALGORITHMS USED IN PHARMACOVIGILANCE

Several ML algorithms have been investigated for ADR detection and signal prioritization.

4.1 Logistic Regression

Logistic regression can classify whether a drug–event pair is likely to represent a potential safety signal. Its relative simplicity and interpretability make it useful as a baseline model [4].

4.2 Random Forest

Random forest is an ensemble learning technique that combines multiple decision trees. It can accommodate nonlinear relationships and interactions among variables. Recent literature identifies random forest as one of the frequently investigated ML approaches for pharmacovigilance signal detection [6].

4.3 Support Vector Machine

Support vector machines classify observations by identifying an optimal decision boundary between classes.

They can be useful for high-dimensional pharmacovigilance datasets.

4.4 Gradient Boosting

Gradient boosting algorithms construct sequential predictive models to improve classification performance. Recent signal-management literature has highlighted gradient boosting as a promising approach for pharmacovigilance applications [6].

4.5 Neural Networks and Deep Learning

Deep learning models can process complex and high-dimensional information. They are particularly useful when pharmacovigilance data contain unstructured text or complex relationships.

4.6 Natural Language Processing

NLP allows computers to analyze human language. In pharmacovigilance, NLP can identify:

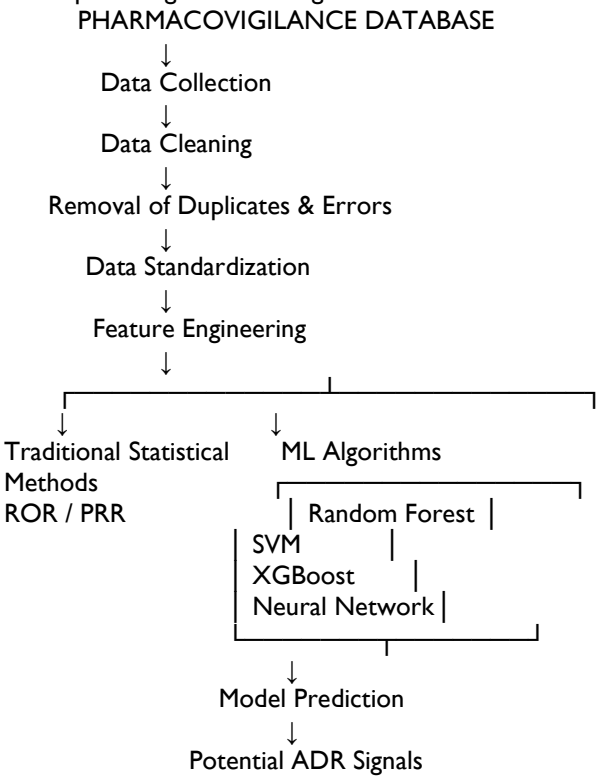
Drug names + adverse events + patient characteristics + temporal relationships

from clinical notes, scientific publications, social media, and safety reports [5,8].

A 2024 review reported that deep learning approaches generally performed better than traditional ML approaches for certain adverse-event extraction tasks, while models such as BERT showed strong performance in end-to-end adverse-event extraction [5].

5. PROPOSED WORKFLOW FOR EARLY ADR SIGNAL DETECTION

A practical ML-based pharmacovigilance system can be developed using the following workflow:



The most effective approach may therefore be a **hybrid model**, in which conventional disproportionality measures are combined with ML predictions rather than completely replacing established methods.

6. ADVANTAGES OF MACHINE LEARNING IN PHARMACOVIGILANCE

ML-based pharmacovigilance offers several potential advantages.

First, scalability: ML can process very large datasets more rapidly than manual review [3].

Second, pattern recognition: ML can identify nonlinear and complex relationships among multiple variables.

Third, early signal identification: Continuous analysis of accumulating data may help prioritize emerging drug–ADR associations for further investigation.

Fourth, integration of heterogeneous data: ML can potentially combine spontaneous reports, EHRs, literature, and patient-generated information [8,9].

Fifth, reduction in manual workload: Automated screening can prioritize reports or drug–event pairs requiring expert assessment.

A 2025 review concluded that ML and NLP are increasingly being applied to signal management and that several ML approaches demonstrated promising performance compared with conventional measures in particular study settings [6].

7. LIMITATIONS AND CHALLENGES

Despite considerable potential, ML-based pharmacovigilance has important limitations.

Data Quality

Pharmacovigilance datasets frequently contain missing, inconsistent, duplicated, or incorrectly coded information. ML models trained on poor-quality data may produce unreliable results.

Reporting Bias

Spontaneous reports are not generated randomly. Media attention, regulatory warnings, publicity, and professional awareness may influence reporting frequency.

Class Imbalance

True ADR signals may represent a very small proportion of the total observations. Consequently, models may become biased toward the majority class.

Interpretability

Complex models, particularly deep neural networks, may function as "black boxes." Pharmacovigilance professionals need to understand why a model has prioritized a particular signal.

Generalizability

A model trained on one database may not perform similarly on another database or in another therapeutic area. Systematic reviews have highlighted limitations in external validation and heterogeneous performance assessment [2,10].

Regulatory and Ethical Considerations

Automated systems should complement rather than replace qualified pharmacovigilance professionals. Patient privacy, data governance, algorithmic bias, transparency, validation, and accountability must be addressed before implementation in routine safety decision-making.[6]

8. FUTURE PERSPECTIVES

The future of ML-based pharmacovigilance is likely to involve **multimodal and hybrid analytical systems** rather than reliance on a single algorithm. Combining disproportionality analysis with ML, NLP, temporal analysis, and real-world clinical data could provide more comprehensive safety surveillance.

Explainable AI may become particularly important because pharmacovigilance experts need to understand the factors responsible for a model-generated signal. Federated learning could also facilitate analysis across institutions while reducing the need to centralize sensitive patient data.

Recent literature demonstrates that although conventional statistical approaches continue to dominate pharmacovigilance signal detection, ML and deep learning are increasingly being investigated. A 2026 scoping review found that traditional disproportionality methods remained predominant, while only a smaller proportion of studies used ML or deep learning, highlighting both the opportunity and the current implementation gap [2,6].

9. CONCLUSION

Machine learning represents a promising advancement in pharmacovigilance because it can analyze large and heterogeneous datasets, recognize complex patterns, and prioritize potential ADR signals for expert evaluation. ML approaches including random forest, gradient boosting, support vector machines, neural networks, and NLP can complement conventional pharmacovigilance methods.[4–6] However, ML should not be considered a substitute for clinical and pharmacovigilance expertise. Data quality, bias, interpretability, validation, privacy, and regulatory acceptance remain major challenges. The future of pharmacovigilance will likely depend on **transparent, validated, explainable, and human-supervised ML systems** capable of integrating multiple sources of real-world safety information. Such systems have the potential to improve the speed and efficiency of ADR signal detection and ultimately contribute to safer medicine use.

10. REFERENCES

1. World Health Organization. WHO Pharmacovigilance Indicators: A Practical Manual for the Assessment of Pharmacovigilance Systems. Geneva: World Health Organization; 2015. (Iris)
2. Coste A, et al. Methods for drug safety signal detection using routinely collected observational electronic health care data: a systematic review. *Pharmacoepidemiol Drug Saf*. 2022. doi:10.1002/pds.5548. (PubMed)
3. Salas M, Petracek J, Yalamanchili P, et al. The use of artificial intelligence in pharmacovigilance: a systematic review of the literature. *Pharm Med*. 2022;36(5):295-306. doi:10.1007/s40290-022-00441-z. (PubMed)
4. Kompa B, Hakim JB, Palepu A, et al. Artificial intelligence based on machine learning in pharmacovigilance: a scoping review. *Drug Saf*. 2022;45(5):477-491. doi:10.1007/s40264-022-01176-1. (PubMed)
5. Li Y, Tao W, Li Z, et al. Artificial intelligence-powered pharmacovigilance: a review of machine and deep learning in clinical text-based adverse drug event detection for benchmark datasets. *J Biomed Inform*. 2024;104621. doi:10.1016/j.jbi.2024.104621. (PubMed)
6. Warner J, Prada Jardim A, Albera C. Artificial intelligence: applications in pharmacovigilance signal management. *Pharm Med*. 2025;39(3):183-198. doi:10.1007/s40290-025-00561-2. (PubMed)
7. Bate A, Evans SJW. Quantitative signal detection using spontaneous ADR reporting. *Pharmacoepidemiol Drug Saf*. 2009;18(6):427-436.
8. Harpaz R, DuMouchel W, Shah NH, et al. Novel data-mining methodologies for adverse drug event discovery and analysis. *ClinPharmacolTher*. 2012;91(6):1010-1021.
9. Pappa D, Stergioulas LK. Harnessing social media data for pharmacovigilance: a review of current approaches. *Expert Opin Drug Saf*. 2019.
10. Jacoby AC, Matsuda MA, Blatt CR, Cazella SC. Data mining methods, tasks, and algorithms for adverse drug reaction analysis in pharmacovigilance: a scoping review. *Int J Med Inform*. 2026;219:106554. doi:10.1016/j.ijmedinf.2026.106554. (PubMed)