

Pharmacovigilance in the Age of Digital Transformation: Leveraging Big Data Analytics and Machine Learning for Enhanced Adverse Drug Reaction Surveillance

Dr. R. Pushpalakshmi¹, Nivethitha T², M. Vengateshwaran³, Lathika B A⁴, Dr. D. Satheesh Kumar⁵, D. Sujeetha⁶

¹Associate Professor Grade-1, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India. Email: pushpalakshmi.r@vit.ac.in

²Assistant Professor, Department of ECE, Sri Eshwar College of Engineering, India. Email: nivethitha1704@gmail.com

³Assistant Professor, Department of Computer Science and Engineering, R.M.K. Engineering College (Autonomous), Chennai, India. Email: mkvengatesh@gmail.com

⁴Assistant Professor, Department of CSE, KGiSL Institute of Technology, India. ORCID: 0000-0001-8059-0917. Email: lathika.b.a@kgkite.ac.in

⁵Associate Professor, Department of Artificial Intelligence and Machine Learning, Hindusthan College of Engineering and Technology, Coimbatore, India. Email: dsatheeshme@gmail.com

⁶Assistant Professor, Department of CSE, Nehru Institute of Engineering and Technology, Coimbatore, India. Email: sujeetha.venkatachalam@gmail.com

Received: 03.03.2026, Reviewed: 11.03.2026 Accepted: 24.03.2026, Published: 05.04.2026

ABSTRACT

Digital transformation has contributed to the emergence of pharmacovigilance that is highly advanced through the use of big data analytics and machine learning algorithms. Conventional adverse drug reaction (ADR) surveillance systems which are based on spontaneous reporting are usually characterised by underreporting, delays and limited data integration. The increase in the number of electronic health records, social media, and real-life data sources has allowed the creation of sophisticated pharmacovigilance systems that can carry out real-time surveillance and prompt signal isolation. Natural language processing and deep learning machine learning algorithms enable the derivation of meaningful insights on structured and unstructured data to enhance the accuracy and efficacy of ADR detection. Although this has been achieved, there are also issues of data quality, interoperability, ethics, and algorithmic biasing. This paper is a critical review of the role of digital technologies in pharmacovigilance and has shown the necessity of adopting hybrid solutions that can involve computational intelligence and regulatory control to guarantee the safety of drugs in today's healthcare systems.

Keywords: Pharmacovigilance, Adverse Drug Reactions, Big Data Analytics, Machine Learning, Artificial Intelligence, Signal Detection, Digital Health

How to cite this article: Pushpalakshmi R, Nivethitha T, Vengateshwaran M, Lathika BA, Satheesh Kumar D, Sujeetha D. Pharmacovigilance in the Age of Digital Transformation: Leveraging Big Data Analytics and Machine Learning for Enhanced Adverse Drug Reaction Surveillance. *Int J Drug Deliv Technol.* 2026;16(20s): 72-75. DOI: 10.25258/ijddt.16.20s.9

Source of support: Nil.

Conflict of interest: None

1. Introduction

Pharmacovigilance, science and the activities involved with the identification, evaluation, and comprehension as well as the prevention of adverse drug reactions (ADRs) are very important in patient security. Conventional pharmacovigilance systems are mostly dependent on spontaneous reporting systems (SRS) that are constrained in most cases by underreporting, delayed response to signals, and incomplete information (Hussain, 2021). Such restrictions

minimise post-marketing drug surveillance effectiveness and delay regulatory interventions.

Pharmacovigilance in the Age of Digital Transformation: Leveraging Big Data Analytics and Machine Learning for Enhanced Adverse Drug Reaction Surveillance

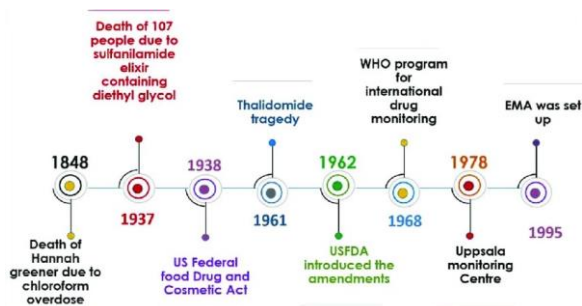


Figure 1: Evolution of Pharmacovigilance from Traditional Reporting Systems to Digital Data-Driven Models

(Source: Pilipiec *et al.*, 2022)

Digitalisation of healthcare has brought new possibilities to improve pharmacovigilance based on the concept of big data analytics and machine learning. The growing accessibility of real-life data in the form of electronic health records (EHRs), insurance claims, and social media sites are becoming an excellent source of information in identifying ADRs (Sarker *et al.*, 2020). These large and complex datasets can be processed by machine learning algorithms to find patterns and associations that could potentially help to determine safety signals (Pilipiec *et al.*, 2022). Such a paradigm shift from traditional to data-driven pharmacovigilance is the benefit of the paradigm and the possibility of continuous surveillance of the safety of drugs proactively. Nevertheless, it also creates some additional problems with the integration of data, the model reliability, and ethics.

2. Foundations of Digital Pharmacovigilance

The convergence of the big data technologies with sophisticated computing methods inherently predetermines the emergence of digital pharmacovigilance that will allow the transfer of the passive surveillance practices to the actively controlled and information-orientated approach to the monitoring of drug safety. The four dimensions of big data in healthcare, though, are commonly described as the volume, velocity, variety, and veracity dimensions, which combine to form the complexity and size of the modern health-related data. In this case, such data cannot be adequately analysed with conventional analytical methods, and, thus, it is mandatory to embrace advanced computing tools and infrastructures (Sharif *et al.*, 2023). Pharmacovigilance systems should have the capacity then to combine and examine both structured information, including electronic health records and laboratory findings, and unstructured information, including clinical stories, physician notes, and patient posts on social media networks.

Machine learning gives the fundamental conceptual framework for deriving significant understanding of these heterogeneous data. Supervised learning methods are often used to categorise and forecast adverse drug reactions using labelled training examples, whereas unsupervised learning methods help to reveal concealed patterns and anomalies that could manifest novel safety issues (Harpaz *et al.*, 2021). One of the most important subfields of artificial intelligence, natural language processing (NLP), is especially important in digital pharmacovigilance because it allows an organised extracting and understanding of the information derived out of the unstructured textual sources of information. NLP makes it possible to find adverse events that were not reported before or discussed in social media before and were underreported (Xu and Wang, 2020).

Moreover, data mining methods also play an important role in signal identification by pointing out statistically important links among drug exposures and adverse events. These methods make pharmacovigilance systems more sensitive, and it is possible to detect rare, delayed, or more complicated ADRs that are not necessarily present in other traditional spontaneous reporting tools (Bate and Edwards, 2020).

3. Data Sources in Modern Pharmacovigilance

Digital pharmacovigilance effectiveness is determined by the quality and different nature of data sources. Electronic health records constitute one of the primary sources of clinical data that is organised, such as patient demographics, diagnoses and treatment histories. Such datasets can be used to do longitudinal analysis of patient outcomes and identify ADR patterns (Zhao *et al.*, 2021). Social media sites have been found to be useful in the collection of real-time patient-reported data. The experiences that users have on the use of medication are usually shared, which gives information about ADRs that might not be communicated through other channels. This unstructured data may be analysed by machine learning models to identify emerging safety signals (Sarker *et al.*, 2020).

Empirical evidence, which comes through observational data not controlled in the presence of clinical trials, has become a significant factor in regulatory decision-making. Such information gives a better picture of drug safety among various categories of patients (Schneeweiss, 2020). Nevertheless, these sources of data also have their own problems, such as problems in data quality, standardisation, and interoperability. The crucial thing in the provision of

Pharmacovigilance in the Age of Digital Transformation: Leveraging Big Data Analytics and Machine Learning for Enhanced Adverse Drug Reaction Surveillance

reliable pharmacovigilance is data accuracy and consistency.

4. Machine Learning Techniques in ADR Detection

The use of machine learning has facilitated the efficiency and accuracy with which ADR is detected. Supervised learning models are widely applied in classifying and predicting adverse events, depending upon labelled data. Such models are able to detect complicated interactions between drug exposure and adverse outcomes (Pilipiec *et al.*, 2022). Deep learning methods such as neural networks have proved to perform better with large and unstructured databases. Such models have the ability to learn feature representations automatically without the need to feature engineer them manually (Zhao *et al.*, 2021). Also, deep learning can be successfully used to process medical text and imaging information.

It uses the unsupervised learning techniques which include clustering and anomaly detection to detect unexpected patterns in data which could signify possible ADRs. These techniques are useful in identifying uncommon or unfamiliar adverse events (Harpaz *et al.*, 2021). Irrespective of these benefits, machine learning models have several requirements, such as high-quality training data and strict validation in order to be reliable. Bias in the data may influence the performance of the model and the need to have strong evaluation frameworks.

5. Applications and Impact on Drug Safety

Drug safety monitoring has been massively boosted by the application of machine learning and big data analytics in the field of pharmacovigilance. With automated signal-detecting systems, ADRs can be identified in real time and it saves more time to take a regulatory action. The systems enhance the efficiency of the pharmacovigilance procedures and decrease the use of manual reporting (Sharif *et al.*, 2023).

Risk stratification with the help of machine learning models also selects patient groups that have a higher vulnerability to adverse reactions. This will allow specific intervention and unique treatment plans, which enhance patient outcomes. In addition, potential safety problems can be predicted with the help of predictive analytics and prevented before they expand (Khan *et al.*, 2022). Accurate incorporation of real-life data into the pharmacovigilance systems has enhanced generalisability of the results as it presents the real-life experiences of patients. In this way, the relevance of safety assessments will be improved, and the evidence-based decision-making process will be achieved (Schneeweiss, 2020).

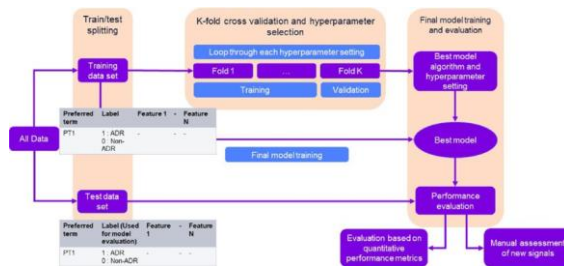


Figure 2: Machine Learning Workflow for ADR Signal Detection and Risk Prediction

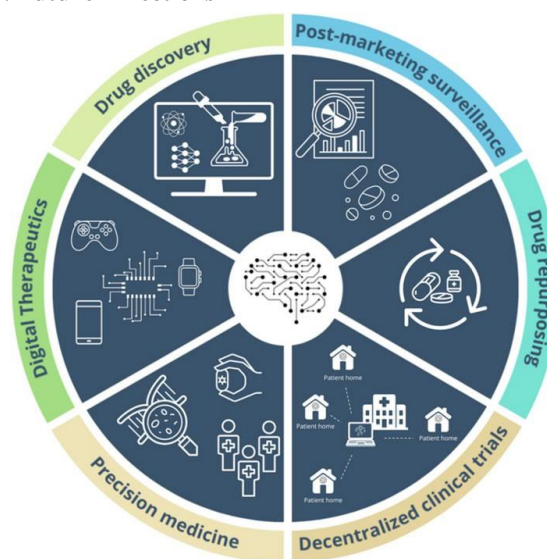
(Source: Schneeweiss, 2020)

6. Challenges and Ethical Considerations

Even with the development in digital pharmacovigilance, there are still a number of issues. The issue of data quality is a key problem because incomplete or inaccurate data could translate to wrong conclusions. The fact that the data sources and formats vary also makes data integration and analysis more complicated (Hussain, 2021). Another vital problem is algorithmic bias. Machine learning can be used to make poor or discriminatory predictions based on biased data, which can have an impact on patient safety. Bias should be dealt with through a meticulous data selection and validation (Khan *et al.*, 2022).

Data security and privacy are also important issues, especially when accessing sensitive patients' data. To ensure that the people trust their pharmacovigilance systems, it is crucial to ensure that they comply with the data protection regulations (Alomar, 2021). Also, complex machine learning models may not be easily accepted in a regulatory environment due to the lack of interpretability. More models are required that are transparent and explainable to facilitate decision-making.

7. Future Directions



Pharmacovigilance in the Age of Digital Transformation: Leveraging Big Data Analytics and Machine Learning for Enhanced Adverse Drug Reaction Surveillance

Figure 3: Future Framework of AI-Driven Pharmacovigilance Systems

(Source: Sharif *et al.*, 2023)

The future of pharmacovigilance is associated with the creation of better-developed and more interconnected analytic systems. The concept of explainable artificial intelligence (XAI) is bound to become highly valuable in enhancing the interpretability and transparency of machine learning models. It will increase the confidence of medical workers and authorities. Multi-source data (genomic and wearable) will also be integrated, which will allow performing more comprehensive safety checks. The data interoperability and standardisation will enable the exchange of information across healthcare systems (Sharif *et al.*, 2023). It will be necessary to use collaborative strategies in promoting digital pharmacovigilance, involving regulatory agencies, healthcare providers, and technology companies. These alliances can facilitate the workings of standard frameworks and best practices.

8. Conclusion

The digital transformation has radically transformed pharmacovigilance, as it has allowed the more efficient and correct detection of adverse drug reactions. Integration of big data analytics and machine learning has helped to improve the capability of processing large and complicated datasets, helping to detect the early signs of a signal and prevent risks proactively. Such innovations can transform patient safety and outcomes in healthcare to a considerable degree. Nevertheless, issues of data quality, algorithm bias, and ethical issues need to be dealt with so as to realise the full potential of digital pharmacovigilance. Greater regulation and technology innovation that is balanced should be employed to allow safe and effective utilisation of these tools. The future of pharmacovigilance will be greatly dependent on the cooperation of stakeholders as the field is still developing.

References

Alomar, M.J., 2021. Factors affecting the reporting of adverse drug reactions. *Saudi Pharmaceutical Journal*, 29(4), pp.334–339. <https://doi.org/10.1016/j.jsps.2021.01.004>

Bate, A. and Edwards, I.R., 2020. Data mining in pharmacovigilance: A practical guide. *Therapeutic Advances in Drug Safety*, 11, pp.1–13. <https://doi.org/10.1177/2042098620930651>

Harpaz, R., DuMouchel, W., Shah, N.H., Madigan, D., Ryan, P. and Friedman, C., 2021. Computational

methods for pharmacovigilance. *Clinical Pharmacology & Therapeutics*, 110(2), pp.313–324. <https://doi.org/10.1002/cpt.2129>

Hussain, R., 2021. Big data, medicines safety and pharmacovigilance. *Journal of Pharmaceutical Policy and Practice*, 14(1), pp.1–9. <https://doi.org/10.1186/s40545-021-00329-4>

Khan, L.M., Al-Harhi, S.E., Saad, A.A. and AlAmoudi, A.A., 2022. Artificial intelligence in pharmacovigilance systems: A review. *Frontiers in Artificial Intelligence*, 5, p.879655. <https://doi.org/10.3389/frai.2022.879655>

Pilipiec, P., Sieradzki, R. and Jurczak, W., 2022. Use of machine learning for pharmacovigilance: A systematic review. *Pharmaceutics*, 14(2), p.266. <https://doi.org/10.3390/pharmaceutics14020266>

Sarker, A., Ginn, R., Nikfarjam, A., O'Connor, K., Smith, K., Jayaraman, S., Upadhaya, T. and Gonzalez, G., 2020. Utilizing social media data for pharmacovigilance: A review. *Briefings in Bioinformatics*, 21(4), pp.1309–1325. <https://doi.org/10.1093/bib/bbz070>

Schneeweiss, S., 2020. Real-world evidence in pharmacovigilance and regulatory decision making. *Nature Reviews Drug Discovery*, 19(11), pp.727–728. <https://doi.org/10.1038/s41573-020-00064-7>

Sharif, K.S., Dhar, S., Saha, S. and Paul, S., 2023. Big data analytics and machine learning in pharmacovigilance: A review. *Frontiers in Pharmacology*, 14, p.1271309. <https://doi.org/10.3389/fphar.2023.1271309>

Xu, R. and Wang, Q., 2020. Large-scale biomedical text mining for pharmacovigilance. *BMC Bioinformatics*, 21(1), p.354. <https://doi.org/10.1186/s12859-020-03684-7>

Yang, C., Delcher, C., Shenkman, E. and Ranka, S., 2020. Machine learning approaches for detecting adverse drug reactions using social media data. *Journal of Biomedical Informatics*, 103, p.103374. <https://doi.org/10.1016/j.jbi.2019.103374>

Zhao, S., Jin, Z., Zhang, Y. and Cao, J., 2021. Deep learning for adverse drug reaction detection using electronic health records. *BMC Medical Informatics and Decision Making*, 21(1), p.137. <https://doi.org/10.1186/s12911-021-01437-6>