

Review

Not peer-reviewed version

---

# Artificial Intelligence and Machine Learning in Pharmacovigilance: A Systematic Review of Models for Drug Safety in Polypharmacy

---

[Khaled Teleb](#) , [Mohamed Gasimelseed](#) , [Roudha Al-Thani](#) , [Rana Aladawi](#) ,  
[Mohamed Izham Mohamed Ibrahim](#) \*

Posted Date: 6 February 2026

doi: 10.20944/preprints202602.0477.v1

Keywords: machine learning; knowledge extraction; pharmacovigilance; polypharmacy; drug–drug interaction prediction; drug safety



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Review

# Artificial Intelligence and Machine Learning in Pharmacovigilance: A Systematic Review of Models for Drug Safety in Polypharmacy

Khaled Teleb <sup>1</sup>, Mohamed Gasimelseed <sup>1</sup>, Roudha Al-Thani <sup>1</sup>, Rana Aladawi <sup>2</sup>  
and Mohamed Izham Mohamed Ibrahim <sup>1,\*</sup>

<sup>1</sup> Clinical Pharmacy and Practice Department, College of Pharmacy, QU Health, Qatar University

<sup>2</sup> Pharmacy Department, Hamad General Hospital, Hamad Medical Corporation

\* Correspondence: mohamedizham@qu.edu.qa (add: Professor of Social & Administrative Pharmacy College of Pharmacy, QU Health, Qatar University, PO Box 2713, Doha, Qatar)

## Abstract

**Background:** Polypharmacy is increasingly prevalent worldwide and is strongly associated with adverse drug reactions (ADRs) and drug–drug interactions (DDIs). Traditional pharmacovigilance (PV) systems rely heavily on spontaneous reporting and manual signal detection, which suffer from substantial underreporting, delayed signal identification, and limited capacity to manage complex multidrug interactions. Artificial intelligence (AI) and machine learning (ML) offer scalable, data-driven approaches that may enhance the detection, assessment, and prevention of medication-related harms in polypharmacy populations. This systematic review evaluates current AI/ML models in PV for polypharmacy, compares their predictive performance, and identifies key methodological facilitators and challenges. **Methods:** A systematic review was conducted in accordance with PRISMA 2020 guidelines and a pre-registered PROSPERO protocol (CRD420251134685). Seven electronic databases, including grey literature sources, were searched from inception to 2025. Eligible studies were primary computational modeling investigations that developed AI/ML models using polypharmacy-focused datasets and reported performance using the area under the receiver operating characteristic curve (AUROC) and the area under the precision–recall curve (AUPRC). Study quality was appraised using the Mixed Methods Appraisal Tool (MMAT). **Results:** Of 7,513 records identified, 23 studies met the inclusion criteria. All studies employed computational modeling designs using large pharmacovigilance and biomedical datasets such as TWOSIDES, Decagon, STITCH, SIDER, and DrugBank. Reported AUROC values ranged from 0.553 to 0.9993, while AUPRC values ranged from 0.112 to 0.999. High-performing models included PU-MLP, DPSP, SimVec, TriVec, and Simple, many of which achieved AUROC and AUPRC values exceeding 0.95. Key strengths of AI/ML approaches included high predictive accuracy, effective handling of class imbalance, and integration of heterogeneous data sources. Major challenges included data sparsity, limited interpretability, cold-start prediction problems, and a lack of external or clinical validation. **Conclusion:** AI and ML models demonstrate strong potential to enhance pharmacovigilance in polypharmacy settings by improving the detection of ADRs and DDIs. Future research should prioritize standardized benchmarking, model interpretability, and real-world clinical validation to support safe and effective regulatory and clinical implementation.

**Keywords:** machine learning; knowledge extraction; pharmacovigilance; polypharmacy; drug–drug interaction prediction; drug safety

## 1. Introduction

The World Health Organization defines Pharmacovigilance (PV) as the science and activities related to the detection, assessment, understanding, and prevention of adverse effects and other

drug-related problems [1]. Despite advances in medication safety, adverse drug reactions (ADRs) and medication-related harms remain a major global public health concern [2]. A critical limitation of current PV systems is the substantial underreporting of ADRs, with a median global underreporting rate estimated at approximately 94%, which delays signal detection and exposes patients to preventable harm [3].

Polypharmacy, commonly defined as the concurrent use of five or more medications, is a major contributor to medication-related risk, particularly among older adults and patients with multiple chronic conditions [4]. Evidence suggests that up to 70% of geriatric patients experience polypharmacy and that these patients have nearly twice the risk of ADRs compared with non-polypharmacy populations, with a substantial proportion of ADRs considered preventable [5,6]. The complexity of multiple drug combinations, comorbidities, and physiological changes makes the detection of ADRs and DDIs in polypharmacy especially challenging for traditional PV approaches.

Conventional PV systems rely largely on spontaneous reporting, manual review, and rule-based signal detection. While these approaches have contributed to drug safety monitoring, they are reactive, resource-intensive, and poorly suited to the analysis of large, heterogeneous datasets such as electronic health records, molecular interaction networks, and post-marketing safety databases [7,8]. As medication use continues to expand in scale and complexity, these limitations have become increasingly apparent.

Artificial intelligence (AI) and machine learning (ML) have emerged as promising tools to address these challenges. AI refers to computational systems capable of performing tasks that typically require human intelligence, while ML enables algorithms to learn patterns from data and improve performance over time [9]. In pharmacovigilance, AI/ML models can analyze structured and unstructured data, identify complex non-linear relationships, and detect rare or multifactorial adverse events that may be missed by traditional methods [10,11]. Recent studies have demonstrated that advanced ML models, including graph-based neural networks, tensor factorization methods, transformer architectures, and positive-unlabeled learning approaches, can achieve high predictive performance in detecting ADRs and DDIs, particularly in polypharmacy contexts [12–15].

Despite growing interest and promising results, important methodological and implementation challenges remain. These include limited external validation, poor interpretability of complex models, concerns about data quality and representativeness, and uncertainty regarding integration into clinical and regulatory workflows [16–18]. To date, there is no consensus on the most effective AI/ML approaches for pharmacovigilance in polypharmacy populations.

The objective of this systematic review is to comprehensively evaluate existing AI and ML applications in pharmacovigilance with a specific focus on polypharmacy. The review aims to (1) identify and classify AI/ML models used in polypharmacy-focused PV, (2) assess their effectiveness using standardized performance metrics, and (3) synthesize reported advantages, challenges, and future directions for research and practice.

## 2. Materials and Methods

### Study Design

This study employed a mixed-methods systematic review design, integrating quantitative performance outcomes with a qualitative synthesis of reported advantages and challenges. The review followed the PRISMA 2020 reporting guidelines [19].

### Research Questions

1. What AI and ML models are currently used in pharmacovigilance to detect and prevent ADRs and DDIs in polypharmacy settings?
2. How effective are these models based on reported AUROC and AUPRC metrics?
3. What are the key advantages, limitations, and implementation challenges of AI/ML-based pharmacovigilance systems?

The PICO framework is structured [20] as the following: P (Population): Polypharmacy patients; I (Intervention): Use of artificial intelligence (AI) and machine learning (ML) in pharmacovigilance; C (Comparator): Traditional pharmacovigilance methods or no AI/ML use, or other AI/ML models; and (Outcome): Improved detection, assessment, and prevention of ADRs and DDIs; implementation, advantages and challenges of AI/ML.

#### Eligibility Criteria

The inclusion criteria for this review included primary computational modeling studies that used polypharmacy-related datasets and focused on developing original artificial intelligence or machine learning models. Eligible studies were required to report key performance metrics, including AUROC and AUPRC, and be published in English between 2015 and 2025. Exclusion criteria included reviews, editorials, commentaries, opinion pieces, non-computational studies, research not centered on polypharmacy, studies lacking the required performance metrics, and conference abstracts without full-text availability.

#### Effect Measures

The primary performance metrics assessed were the Area Under the Receiver Operating Characteristic Curve (AUROC) and the Area Under the Precision-Recall Curve (AUPRC). AUROC reflects the overall discriminative ability of a binary classification model, indicating how well it distinguishes between positive and negative cases across different thresholds [21]. Its values range from 0 to 1, where a score close to 1 denotes excellent discrimination, 0.5 suggests random guessing, and values near 0 indicate poor performance. AUPRC, on the other hand, is particularly informative for imbalanced datasets, such as those encountered in pharmacovigilance, where true ADRs and DDIs are rare. Unlike AUROC, AUPRC accounts for precision and recall, with baseline performance equal to the prevalence of the positive class (e.g., 8% positives → baseline ≈ 0.08). Higher AUPRC values indicate better model performance, with 1.0 representing perfect precision and recall. Given the challenges posed by imbalanced data in polypharmacy research, AUPRC was considered particularly relevant for evaluating AI/ML models in this context.

In addition, this study carried out a narrative synthesis of the advantages and challenges of AI and/or ML.

#### Information Sources and Search Strategy

Seven electronic databases (PubMed, EMBASE, CINAHL, Web of Science, Scopus, ProQuest Central, and ProQuest Dissertations and Theses Global) were searched from inception to 2025. Search strategies combined controlled vocabulary and keywords across three domains: (1) artificial intelligence and machine learning, (2) pharmacovigilance and drug safety, and (3) polypharmacy. Boolean operators were used to optimize sensitivity and specificity.

#### Study Selection and Data Management

All retrieved records were imported into EndNote for deduplication and then screened using Rayyan. Title and abstract screening was followed by full-text review by independent reviewers, with disagreements resolved by consensus.

#### Data Extraction

Data extracted included authors, year, country, datasets used, model type, performance metrics (AUROC and AUPRC), and reported advantages and challenges.

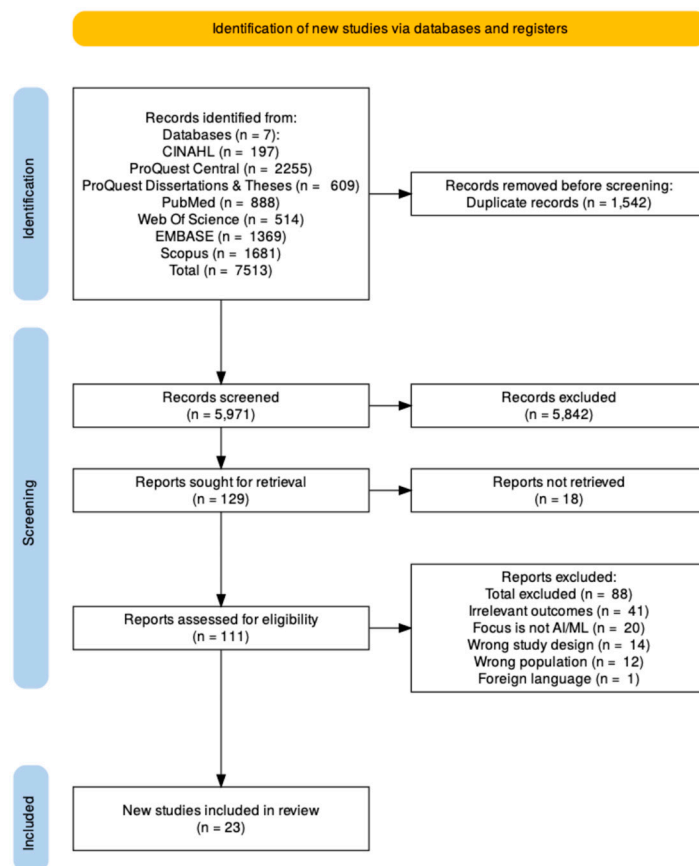
#### Quality Assessment

Study quality was assessed using the Mixed Methods Appraisal Tool (MMAT) 2018 (23). No study was excluded based on the quality assessment.

### 3. Results

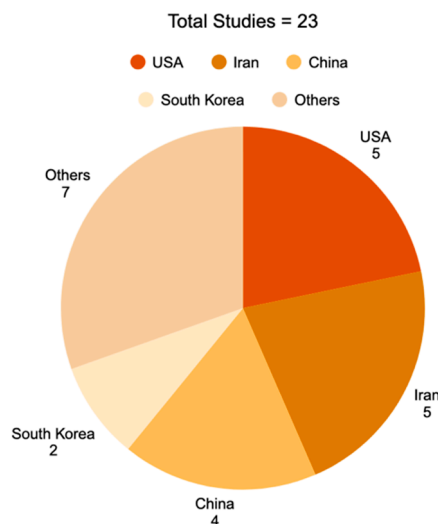
#### Study Selection and Characteristics

The literature search identified 7,513 records. After removal of duplicates and screening, 23 studies met all inclusion criteria and were included in the final synthesis [24–46] (Figure 1).



**Figure 1.** PRISMA flow chart.

All included studies employed computational modeling designs and were published between 2018 and 2025 (Table 1). Studies originated from Asia, North America, and Europe, with the highest representation from China, the United States, and Iran (Figure 2). Commonly used datasets in the studies to represent DDIs and ADRs included TWOSIDES, Decagon, STITCH, SIDER, OFFSIDES, and DrugBank. Many studies integrated additional biological data such as protein–protein interaction networks and chemical structure representations. Biological interaction databases/datasets such as STITCH, Protein–Protein Interaction (PPI) datasets, KEGG and PubChem were utilized in the studies. Furthermore, some studies used LINCS L1000, SMILES-based chemical datasets, and Inxight Drugs API to cover gene expression.



**Figure 2.** Country Distribution of Computational Studies.

**Table 1.** Characteristics of the included studies (n=23).

Study number	Authors/year	Country of study	Study design	Sample characteristics
Study 1	Zitnik M et al. 2018	United States	Computational modeling study	<ul style="list-style-type: none"> <li>From a heterogeneous polypharmacy dataset: 645 drug, 19,085 protein, 4,651,131 drug-drug interactions (polypharmacy side effects), 715,612 protein-protein interactions, 18,596 protein-drug interactions.</li> </ul>
Study 2	Burkhardt H et al. 2020	United States	Computational modeling study	<ul style="list-style-type: none"> <li>TWOSIDES: contained 9,643,506 triples, which included 7,323,790 drug-drug interaction triples, 2,289,960 protein-protein interaction pairs, and 29,756 drug target pairs.</li> <li>The analysis focused on 963 side effect types (those that occurred in 500 or more drug interaction triples).</li> </ul>
Study 3	Nováček V, Mohamed S 2020	Ireland	Computational modeling study	<ul style="list-style-type: none"> <li>The study used the Decagon dataset. The final dataset splits were:</li> <li>Training data: ~32,000 entities, 967 relations, 4.7 million triples.</li> <li>Validation data: 643 entities, 963 relations, 459,000 triples.</li> <li>Testing data: 643 entities, 963 relations, 459,000 triples.</li> </ul>
Study 4	Dasgupta S et al. 2021	Singapore and India	Computational modeling study	<ul style="list-style-type: none"> <li>ADE prediction: 4,484 drug-disease pairs (EU-ADR + OMOP).</li> <li>Polypharmacy prediction: <math>\approx</math> 8,270,000 interaction samples (7.32 M training + 0.91 M testing).</li> </ul>
Study 5	Masumshah R et al. 2021	Iran	Computational modeling study	<ul style="list-style-type: none"> <li>TWOSIDES database: 645 drugs and 63,473 drug pairs.</li> <li>8934 proteins; 18,690 drug-protein interactions (STITCH).</li> <li>10,184 mono side effects (SIDER + OFFSIDES)</li> <li>964 polypharmacy side effects used for training and testing.</li> </ul>
Study 6	Dewulf P et al. 2021	Belgium	Computational modeling study	<ul style="list-style-type: none"> <li>645 drugs, 963 adverse effects, 63,473 drug-drug combinations (2% of possible triplets labeled as occurring effects, caused by 70% of all drug pairs).</li> </ul>
Study 7	Liu T et al. 2021	China	Computational modeling study	<ul style="list-style-type: none"> <li>Decagon: 645 drugs; 7,795 proteins; 4,576,785 drug-drug interactions; 18,690 drug-protein interactions.</li> <li>TWOSIDES/PubChem: 548 drugs; 97,168 drug-drug interactions.</li> </ul>

<b>Study 8</b>	Lin S et al. 2022	China	Computational modeling study	<ul style="list-style-type: none"> <li>▪ TWOSIDES database: 645 drugs and 63,473 drug pairs (964 polypharmacy side effects).</li> <li>▪ 10,184 mono side effects from SIDER + OFFSIDES; 7,795 proteins from STITCH.</li> </ul>
<b>Study 9</b>	Kim E, Nam H 2022	Republic of Korea	Computational modeling study	<ul style="list-style-type: none"> <li>▪ Dataset 1 (LINCS L1000): 19,156 compounds, 978 landmark genes per signature.</li> <li>▪ Dataset 2 (TWOSIDES): 4,576,287 interactions, 63,472 drug combinations, 963 side-effect types (only those with <math>\geq 500</math> combinations retained).</li> <li>▪ Dataset 3 (DrugBank):</li> <li>▪ Version 5.0.0 <math>\rightarrow</math> 33,497 interactions among 1,129 drugs.</li> <li>▪ Version 5.1.7 <math>\rightarrow</math> 782,405 interactions among 2,616 drugs.</li> </ul>
<b>Study 10</b>	Lakizadeh A, Babaei M 2022	Iran	Computational modeling study	<ul style="list-style-type: none"> <li>▪ 643 drugs, 964 side effect types, 731 proteins, and 17,435 drug-protein interactions derived from integrated public databases (STITCH, DrugBank, KEGG, PubChem, OFFSIDES, SIDER, and TWOSIDES).</li> </ul>
<b>Study 11</b>	Liu Q et al. 2022	China and Singapore	Computational modeling study	<ul style="list-style-type: none"> <li>▪ Conducted experiments on a publicly available dataset compiled by Zitnik M et al.</li> <li>▪ Drug-drug interactions: From SIDER, OFFSIDES, and TWOSIDES.</li> <li>▪ Protein-protein Interactions: From the PPI dataset.</li> <li>▪ Drug-protein Interactions: From STITCH.</li> </ul>
<b>Study 12</b>	Ragkousis A et al. 2022	Greece	Computational modeling study	<ul style="list-style-type: none"> <li>▪ BioSNAP-Decagon polypharmacy dataset: 645 drugs; 19,081 proteins; 4,625,608 side-effect edges; 715,612 protein-protein interactions edges.</li> </ul>
<b>Study 13</b>	Lukashina N et al.2022	Russia and Sweden	Computational modeling study	<ul style="list-style-type: none"> <li>▪ The study used the preprocessed dataset from the Decagon study, which contains 645 drug nodes.</li> <li>▪ The new "weak nodes split" created for the evaluation consisted of 98 "new" drugs and 20,000 "weak triples"</li> </ul>
<b>Study 14</b>	Kim J, Shin M 2023	Republic of Korea	Computational modeling study	<ul style="list-style-type: none"> <li>▪ Publicly available data provided by Zitnik M et al.</li> </ul>

				<ul style="list-style-type: none"> <li>284 drugs linked to <math>\geq 1</math> protein; 19,089 proteins (715,612 protein-protein 18,690 drug-protein edges); 14,247 labeled drug pairs <math>\times</math> 1308 side effects (+ 244 single drugs labeled with same side effects).</li> </ul>
<b>Study 15</b>	Masumshah R, Eslahchi C 2023	Iran	Computational modeling study	<ul style="list-style-type: none"> <li>Databases used: DrugBank, KEGG, PubChem, OFFSIDES, SIDER, TWOSIDES.</li> <li>Dataset 1: 572 drugs, 37,264 drug-drug interactions, 65 event types, 9991 side effects, 1162 targets, 202 enzymes, 881 chemical substructures, 957 pathways.</li> <li>Dataset 2: 1258 drugs, 161,770 drug-drug interactions, 100 event types, 1651 targets, 316 enzymes, 2040 chemical substructures.</li> <li>Dataset 3: 645 drugs, 63,473 drug-drug interactions, 185 adverse effects, 10,184 mono side effects, 8934 targets.</li> </ul>
<b>Study 16</b>	Wang Y et al. 2023	China	Computational modeling study	<ul style="list-style-type: none"> <li>645 drugs and 964 common drug side effects are collected from SIDER database and protein-protein interactions are obtained from STITCH database. Overall, it contains 4,651,131 drug combination-side effect associations.</li> <li>The study focuses on predicting the 964 commonly occurring types of polypharmacy side effects that each occurred in at least 500 drug combinations.</li> </ul>
<b>Study 17</b>	Lloyd O et al. 2024	United Kingdom	Computational comparative analysis of tensor factorization (TF) models	<ul style="list-style-type: none"> <li>Using data provided by Zitnik M et al.</li> </ul>
<b>Study 18</b>	Keshavarz H, Lakizadeh A 2024	Iran	Computational modeling study	<ul style="list-style-type: none"> <li>Dataset1 (DrugBank and KEG): 572 drugs, 37,264 drug-drug interactions, and 47 event types.</li> <li>Dataset2 (TWOSIDES): 645 drugs, 4,649,441 drug-drug interactions, and 963 event types.</li> </ul>

<b>Study 19</b>	Gromova A, Maida A 2025	United States	Computational modeling study	<ul style="list-style-type: none"> <li>645 drugs, 63,473 drug pairs, 964 polypharmacy side effects, 10,184 mono side effects, and 8,934 protein targets.</li> <li>TWOSIDES database: Total of 4,576,785 labeled associations.</li> </ul>
<b>Study 20</b>	Gromova A, Maida A 2025	United States	Computational modeling study	<ul style="list-style-type: none"> <li>645 drugs, 63,473 drug pairs, 964 polypharmacy side effects, 10,184 mono side effects, and 8,934 protein targets (TWOSIDES + SIDER + OFFSIDES + STITCH + PubChem).</li> </ul>
<b>Study 21</b>	Wang N, Taylor C 2025	United States	Computational modeling study	<ul style="list-style-type: none"> <li>Datasets used: (STITCH, SIDER, OFFSIDES, SMILES and Inight DrugsAPI) compiled and preprocessed by Zitnik M et al.</li> </ul>
<b>Study 22</b>	Deng Z et al. 2025	China	Computational modeling study	<ul style="list-style-type: none"> <li>DrugBank version 5.1.10 → 1,974 drugs and 603,816 asymmetric drug–drug interactions. After removing drugs with invalid SMILES or missing fingerprints, they kept 1,752 drugs and 504,468 asymmetric interactions.</li> </ul>
<b>Study 23</b>	Keshavarz H, Lakizadeh A 2025	Iran	Computational modeling study	<ul style="list-style-type: none"> <li>Dataset1 (DrugBank, KEGG): 645 drugs, 963 side effects, and 4,649,441 drug-drug side effect associations.</li> <li>Dataset1 (TWOSIDES, OFFSIDES): 572 drugs, 47 side effects, and 37,264 drug-drug side effect associations.</li> </ul>

### AI/ML Models and Performance

A wide range of AI/ML architectures were identified, including graph neural networks, tensor factorization models, embedding-based approaches, transformer-based models, and positive-unlabeled learning frameworks (Figure 3, Table 2). AUROC values ranged from 0.553 to 0.9993, while AUPRC values ranged from 0.112 to 0.999. The highest-performing models included PU-MLP (AUROC 0.992; AUPRC 0.999), DPSP (AUROC up to 0.999; AUPRC up to 0.977), and SimVec (AUROC 0.975; AUPRC 0.968) (Figure 4). These models demonstrated strong ability to detect ADRs and DDIs despite severe class imbalance. Lower performance was observed in cold-start scenarios involving previously unseen drug combinations.

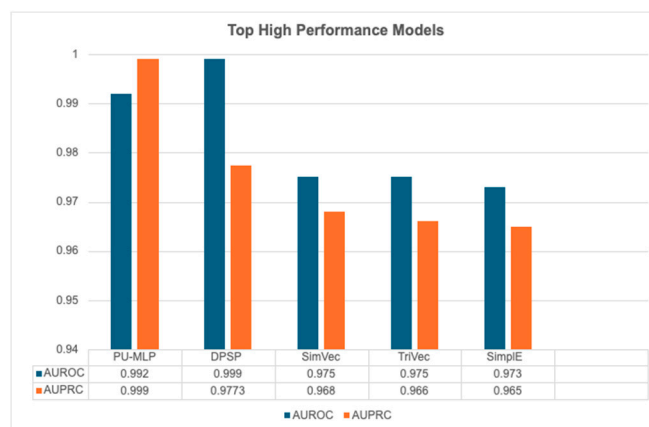


Figure 3. Top High-Performance Models.

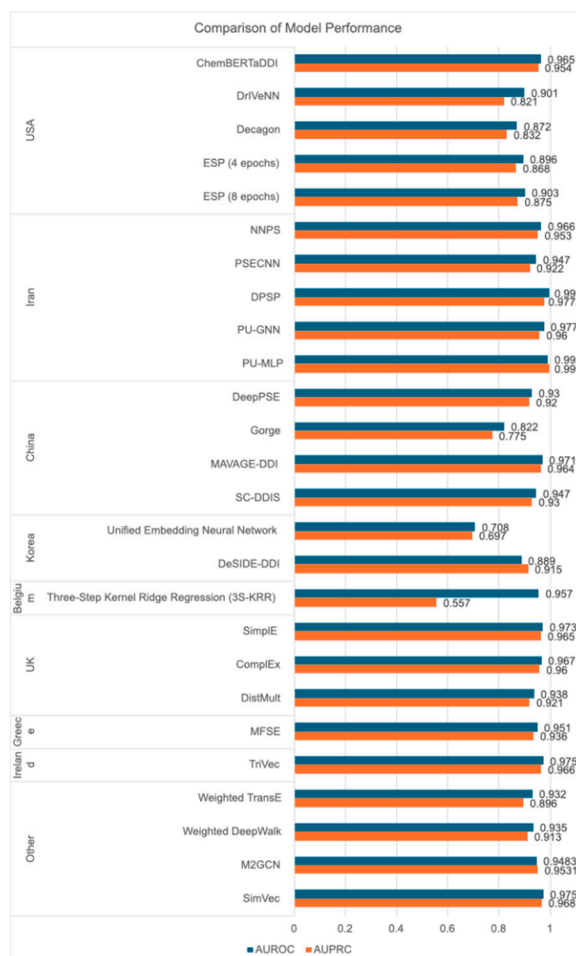


Figure 4. Comparison of Model Performance grouped by country origin.

Table 2. Models and Performance Metrics Outcomes.

Study Number	Authors	Year of Publication	Model	Programs and Applications used	AUROC	AUPRC
1	Zitnik M et al.	2018	Decagon	<ul style="list-style-type: none"> <li>▪ Adam optimizer</li> <li>▪ T-SNE package</li> </ul>	<ul style="list-style-type: none"> <li>▪ 0.872</li> </ul>	<ul style="list-style-type: none"> <li>▪ 0.832</li> </ul>
2	Burkhardt H et al.	2020	ESP (4 epochs)	<ul style="list-style-type: none"> <li>▪ Semantic Vector package</li> </ul>	<ul style="list-style-type: none"> <li>▪ 0.896</li> </ul>	<ul style="list-style-type: none"> <li>▪ 0.868</li> </ul>
			ESP (8 epochs)	<ul style="list-style-type: none"> <li>▪ Anaconda2</li> <li>▪ Python</li> <li>▪ Scikit-learnpackage</li> <li>▪ Umap-learn package</li> <li>▪ Decagon code from GitHub</li> </ul>	<ul style="list-style-type: none"> <li>▪ 0.903</li> </ul>	<ul style="list-style-type: none"> <li>▪ 0.875</li> </ul>
3	Nováček V, Mohamed S	2020	TriVec	<ul style="list-style-type: none"> <li>▪ TensorFlow</li> <li>▪ Python</li> </ul>	<ul style="list-style-type: none"> <li>▪ 0.975</li> </ul>	<ul style="list-style-type: none"> <li>▪ 0.966</li> </ul>
4	Dasgupta S et al.	2021	Weighted TransE Weighted DeepWalk	<ul style="list-style-type: none"> <li>▪ Python</li> <li>▪ Scikit-learn</li> <li>▪ DeepWalk Implementation</li> <li>▪ TransE Implementation</li> </ul>	<ul style="list-style-type: none"> <li>▪ Polypharmacy prediction:</li> <li>▪ Weighted TransE: 0.932</li> <li>▪ Weighted DeepWalk: 0.935</li> </ul>	<ul style="list-style-type: none"> <li>▪ Weighted TransE: 0.896</li> <li>▪ Weighted DeepWalk: 0.913</li> </ul>
5	Masumshah R et al.	2021	NNPS	<ul style="list-style-type: none"> <li>▪ PCA (Principal Component Analysis)</li> <li>▪ SGD (Stochastic Gradient Descent)</li> </ul>	<ul style="list-style-type: none"> <li>▪ 0.966</li> </ul>	<ul style="list-style-type: none"> <li>▪ 0.953</li> </ul>
6	Dewulf P et al.	2021	Three-Step Kernel Ridge Regression (3S-KRR)	<ul style="list-style-type: none"> <li>▪ N/A</li> </ul>	<ul style="list-style-type: none"> <li>▪ Unknown drug-drug effect: 0.957</li> <li>▪ Unknown drug-drug pair: 0.919</li> <li>▪ Unknown drug: 0.910</li> <li>▪ Two unknown drugs: 0.843</li> </ul>	<ul style="list-style-type: none"> <li>▪ Unknown drug-drug effect: 0.557</li> <li>▪ Unknown drug-drug pair: 0.286</li> <li>▪ Unknown drug: 0.221</li> <li>▪ Two unknown drugs: 0.112</li> </ul>
7	Liu T et al.	2021	SC-DDIS	<ul style="list-style-type: none"> <li>▪ TensorFlow</li> <li>▪ Adam optimizer</li> </ul>	<ul style="list-style-type: none"> <li>▪ 0.947</li> </ul>	<ul style="list-style-type: none"> <li>▪ 0.93</li> </ul>
8	Lin S et al.	2022	DeepPSE	<ul style="list-style-type: none"> <li>▪ N/A</li> </ul>	<ul style="list-style-type: none"> <li>▪ 0.93</li> </ul>	<ul style="list-style-type: none"> <li>▪ 0.92</li> </ul>
9	Kim E, Nam H	2022	DeSIDE-DDI	<ul style="list-style-type: none"> <li>▪ Mordred</li> </ul>	<ul style="list-style-type: none"> <li>▪ Unseen interactions: 0.889</li> <li>▪ One-unseen interaction: 0.640</li> <li>▪ Both unseen interaction: 0.553</li> </ul>	<ul style="list-style-type: none"> <li>▪ Unseen interactions: 0.915</li> <li>▪ One-unseen interaction: 0.627</li> <li>▪ Both unseen interaction: 0.556</li> </ul>
10	Lakizadeh A, Babaei M	2022	PSECNN	<ul style="list-style-type: none"> <li>▪ Python</li> <li>▪ TensorFlow</li> </ul>	<ul style="list-style-type: none"> <li>▪ Unbalanced data: 0.901</li> <li>▪ Balanced data: 0.947</li> </ul>	<ul style="list-style-type: none"> <li>▪ Unbalanced data:0.896</li> <li>▪ Balanced data: 0.922</li> </ul>
11	Liu Q et al.	2022	M2GCN	<ul style="list-style-type: none"> <li>▪ Adam algorithm</li> </ul>	<ul style="list-style-type: none"> <li>▪ 0.9483</li> </ul>	<ul style="list-style-type: none"> <li>▪ 0.9531</li> </ul>
12	Ragkousis A et al.	2022	MFSE	<ul style="list-style-type: none"> <li>▪ PyTorch</li> </ul>	<ul style="list-style-type: none"> <li>▪ 0.951</li> </ul>	<ul style="list-style-type: none"> <li>▪ 0.936</li> </ul>

				<ul style="list-style-type: none"> <li>▪ PyTorchGeometric</li> <li>▪ TorchDrugpackage</li> </ul>		
13	Lukashina N et al.	2022	SimVec	<ul style="list-style-type: none"> <li>▪ RDKit package</li> <li>▪ PubChem</li> </ul>	<ul style="list-style-type: none"> <li>▪ 0.975</li> </ul>	<ul style="list-style-type: none"> <li>▪ 0.968</li> </ul>
14	Kim J, Shin M	2023	Unified Embedding Neural Network	<ul style="list-style-type: none"> <li>▪ Enrichr</li> <li>▪ Node2vec</li> <li>▪ Adam Optimizer</li> </ul>	<ul style="list-style-type: none"> <li>▪ DPU: 0.708</li> <li>▪ Non-DPU: 0.912</li> </ul>	<ul style="list-style-type: none"> <li>▪ DPU: 0.697</li> <li>▪ Non-DPU: 0.901</li> </ul>
15	Masumshah R, Eslahchi C	2023	DPSP	<ul style="list-style-type: none"> <li>▪ N/A</li> </ul>	<ul style="list-style-type: none"> <li>▪ DS1: 0.9990</li> <li>▪ DS2: 0.9993</li> <li>▪ DS3: 0.9849</li> </ul>	<ul style="list-style-type: none"> <li>▪ DS1: 0.9773</li> <li>▪ DS2: 0.9633</li> <li>▪ DS3: 0.9465</li> </ul>
16	Wang Y et al.	2023	Gorge	<ul style="list-style-type: none"> <li>▪ Python</li> <li>▪ Adam optimizer</li> </ul>	<ul style="list-style-type: none"> <li>▪ 0.822</li> </ul>	<ul style="list-style-type: none"> <li>▪ 0.775</li> </ul>
17	Lloyd O et al.	2024	Simple Complex DistMult	<ul style="list-style-type: none"> <li>▪ CentOS-7</li> <li>▪ Python</li> <li>▪ PyTorch</li> </ul>	<ul style="list-style-type: none"> <li>▪ Simple(Selfloops): 0.978</li> <li>▪ Simple (non-naïve): 0.973</li> <li>▪ Complex (Selfloops): 0.970</li> <li>▪ Complex (non-naïve): 0.967</li> <li>▪ DistMult (Selfloops): 0.946</li> <li>▪ DistMult (non-naïve): 0.938</li> </ul>	<ul style="list-style-type: none"> <li>▪ Simple (Selfloops) 0.971</li> <li>▪ Simple (non-naïve): 0.965</li> <li>▪ Complex(Selfloops): 0.963</li> <li>▪ Complex (Non-naïve): 0.960</li> <li>▪ DistMult(Selfloops): 0.930</li> <li>▪ DistMult (Non-naïve): 0.921</li> </ul>
18	Keshavarz H, Lakizadeh A	2024	PU-GNN	<ul style="list-style-type: none"> <li>▪ GaussianNB</li> </ul>	<ul style="list-style-type: none"> <li>▪ 0.977</li> </ul>	<ul style="list-style-type: none"> <li>▪ 0.96</li> </ul>
19	Gromova A, Maida A	2025	ChemBERTaDDI	<ul style="list-style-type: none"> <li>▪ Hugging Face Transformers</li> <li>▪ ChemBERTa-77M-MLM</li> <li>▪ Adam optimizer</li> <li>▪ PCA</li> <li>▪ Hugging Face Transformers</li> <li>▪ ChemBERTa-77M-MLM</li> <li>▪ Adam optimizer</li> <li>▪ PCA</li> </ul>	<ul style="list-style-type: none"> <li>▪ 0.965</li> </ul>	<ul style="list-style-type: none"> <li>▪ 0.954</li> </ul>
20	Gromova A, Maida A	2025	ChemBERTaDDI	<ul style="list-style-type: none"> <li>▪ Hugging Face Transformers</li> <li>▪ ChemBERTa-77M-MLM</li> <li>▪ Adam optimizer</li> <li>▪ PCA</li> </ul>	<ul style="list-style-type: none"> <li>▪ 0.965</li> </ul>	<ul style="list-style-type: none"> <li>▪ 0.954</li> </ul>
21	Wang N, Taylor C	2025	DrIVeNN	<ul style="list-style-type: none"> <li>▪ Python</li> <li>▪ Hyperband algorithm</li> </ul>	<ul style="list-style-type: none"> <li>▪ General model: 0.901</li> <li>▪ CVD-specific model: 0.975</li> </ul>	<ul style="list-style-type: none"> <li>▪ General model: 0.821</li> <li>▪ CVD-specific model: 0.952</li> </ul>

				<ul style="list-style-type: none"> <li>▪ PCA</li> <li>▪ UMAP</li> </ul>		
22	Deng Z et al.	2025	MAVAGE-DDI	<ul style="list-style-type: none"> <li>▪ N/A</li> </ul>	0.971	0.964
23	Keshavarz H, Lakizadeh A	2025	PU-MLP	<ul style="list-style-type: none"> <li>▪ Multi-Layer Perceptron</li> <li>▪ Elkan-Noto PU Classifier</li> <li>▪ Random Forest Classifier</li> <li>▪ Adam optimizer</li> <li>▪ Stochastic Gradient Descent optimizer</li> </ul>	0.992	0.999

### Advantages and Challenges

The advantages and challenges associated with applying AI and ML in polypharmacy-focused pharmacovigilance were systematically identified from the included studies and are summarized in Table 3. Reported advantages include high predictive accuracy and enhanced sensitivity to rare adverse events, effective handling of imbalanced datasets, and the ability to integrate heterogeneous pharmacological and biological data sources. These models also demonstrate scalability, enabling analysis of large datasets and supporting high-throughput screening. However, several challenges persist. Model performance is highly dependent on data quality and completeness, and the interpretability of complex algorithms remains limited, posing barriers to clinical adoption. Additionally, reduced effectiveness in cold-start scenarios, lack of external and clinical validation, and practical constraints related to integration into routine clinical workflows continue to hinder widespread implementation.

**Table 3.** Advantages and Challenges of AI and/or ML.

Study number	Authors/ year	Advantages of AI and/or ML	Challenges of AI and/or ML
1	Zitnik M et al. 2018	<ul style="list-style-type: none"> <li>• Predicts specific side effect types, identifying exact outcomes such as "nausea" or "bleeding."</li> <li>• Integrates PPI, DDI, and DDI data, utilizing a vast biological data network for decision-making.</li> <li>• End-to-end trainable GCN, employing a powerful deep learning model that learns comprehensively in one pass.</li> <li>• Shared parameters aid rare side effect prediction, learning from common effects to improve predictions for rare ones.</li> <li>• Clusters related side effects, grouping similar effects to enhance understanding.</li> </ul>	<ul style="list-style-type: none"> <li>• Impossible to physically test all combinations (computational necessity), necessitating AI reliance due to the infeasibility of testing every drug pair on humans.</li> </ul>
2	Burkhardt H et al. 2020	<ul style="list-style-type: none"> <li>• Outperformed Decagon (AUC 0.903), demonstrating higher accuracy than the prominent predecessor model.</li> <li>• 43x faster training; significantly simpler (fewer bits), offering vastly increased speed while using minimal memory.</li> <li>• Trainable on commodity hardware, capable of running on standard laptops without needing supercomputers.</li> <li>• Reusable and interpretable embeddings, producing output that is easily adapted for other inquiries.</li> </ul>	<ul style="list-style-type: none"> <li>• Training data (mined text) may be low quality, stemming from literature mining that may contain inaccuracies.</li> <li>• Lacks relationship directionality, identifying the existence of an effect without specifying the causative drug.</li> <li>• Ambiguous comparison protocols with previous models, complicating fair comparison due to unclear rules.</li> </ul>
3	Nováček V, Mohamed S 2020	<ul style="list-style-type: none"> <li>• Outperforms Decagon (12-16% margin), achieving significantly greater accuracy than the older model.</li> <li>• Novel tensor decomposition embedding, using advanced mathematics to identify superior patterns in the data.</li> <li>• "TriVec" uses multi-part vectors, employing three distinct vectors to</li> </ul>	<ul style="list-style-type: none"> <li>• Relies on known polypharmacy data (limited), requiring existing data for learning and thus struggling with unknown variables.</li> <li>• "Weak nodes" reduce predictive accuracy, resulting in poor guesses for drugs with minimal information.</li> <li>• Standard eval ratio (1:1) might be too easy, suggesting the test used for verification may have been overly simple.</li> </ul>

		describe relationships with greater clarity.	<ul style="list-style-type: none"> <li>Does not learn from chemical structure, ignoring the chemical appearance of the drug.</li> </ul>
4	Dasgupta S et al. 2021	<ul style="list-style-type: none"> <li>Improved graph learning using NLP-derived confidence scores, which enhances the model's learning process by verifying information certainty.</li> <li>Literature-based KG (SemMedDB) enhanced prediction precision, utilizing medical literature to improve the accuracy of drug effect forecasts.</li> <li>Weighted TransE/DeepWalk improved ADE prediction (8.4% AUC gain), utilizing specific mathematical techniques to significantly boost side effect prediction accuracy.</li> <li>Open-source code supports reproducibility, allowing the scientific community to freely access and verify results.</li> </ul>	<ul style="list-style-type: none"> <li>Relies on accuracy of NLP-inferred scores, meaning the model's validity is contingent on the correctness of initial language processing tools.</li> <li>Evaluation limited to two representation methods, lacking a comprehensive assessment across diverse data presentation styles.</li> <li>Small datasets affect generalizability, potentially limiting performance when applied to new or broader data sources due to limited training volume.</li> <li>Performance sensitive to noisy data, leading to potential confusion or inaccuracies when input data is inconsistent or incorrect.</li> <li>Requires cross-domain validation, necessitating testing across various medical fields to ensure universal applicability.</li> </ul>
5	Masumshah R et al. 2021	<ul style="list-style-type: none"> <li>Fast runtime (8h vs 15 days), completing processing tasks significantly faster than legacy models.</li> <li>Simple architecture with high accuracy, offering a streamlined design that maintains precision while being easier to implement.</li> <li>Novel feature vectors (merged side effects + interactions), describing drugs through a unique combination of multiple distinct cues.</li> <li>Efficient dimensionality reduction (PCA), simplifying complex data structures without discarding critical details.</li> </ul>	<ul style="list-style-type: none"> <li>Complex models have high computational costs, requiring excessive time and financial investment to operate.</li> <li>Limited dataset types (no PPIs), missing crucial biological context such as protein-protein interactions.</li> <li>Needs validation on other datasets, requiring broader testing to confirm efficacy across different data types.</li> </ul>
6	Dewulf P et al. 2021	<ul style="list-style-type: none"> <li>Unified cold-start prediction tasks, enabling effect prediction for novel drugs lacking historical data.</li> <li>Fast cross-validation via algebraic shortcuts, utilizing mathematical efficiencies to rapidly verify model accuracy.</li> <li>Integrated mono-drug side effects/target similarities, comparing drugs to identify patterns in side effect profiles.</li> <li>High accuracy (0.957 AUC) for cold-start problems, achieving high correct prediction rates even for new drugs.</li> </ul>	<ul style="list-style-type: none"> <li>Depends on dataset completeness, requiring comprehensive data coverage to function correctly.</li> <li>Performance declines with cold-start difficulty, struggling to predict effects for highly unique or unusual new drugs.</li> <li>Limited interpretability/clinical validation, lacking clear explanatory logic and testing on human subjects.</li> <li>Restricted to binary predictions, limited to simple "yes/no" outputs rather than assessing severity or magnitude.</li> <li>High computational demand for similarity matrices, requiring substantial processing power to compare drug characteristics.</li> </ul>

7	Liu T et al. 2021	<ul style="list-style-type: none"> <li>• Signed network classifies positive/negative relations, distinguishing between helpful and harmful interactions.</li> <li>• Structural Balance Theory infers latent features, applying social network principles to estimate unseen relationships.</li> <li>• Integrates multi-modal features, combining different drug information into a single unified model.</li> <li>• Weighted loss handles class imbalance, prioritizing rare side effects to prevent omission.</li> <li>• Decoding matrix reduces noise, filtering out data clutter to clarify analysis.</li> </ul>	<ul style="list-style-type: none"> <li>• Excludes protein-protein interactions (PPIs), ignoring protein interactions which are a key analytical component.</li> </ul>
8	Lin S et al. 2022	<ul style="list-style-type: none"> <li>• Multi-view fusion captured complementary features, analyzing the problem from multiple perspectives to derive a superior solution.</li> <li>• Self-attention improved weighting and interpretability, focusing on critical details while filtering out irrelevant noise.</li> <li>• Outperformed 5 state-of-the-art baselines, delivering consistently superior results relative to other models.</li> <li>• PCA improved efficiency and robustness, simplifying data handling while strengthening the model.</li> <li>• Open-source code, ensuring ease of sharing and accessibility.</li> </ul>	<ul style="list-style-type: none"> <li>• High model complexity and cost, involving intricate construction and expensive operation.</li> <li>• Limited to existing drug targets, unable to process previously unseen drug types.</li> <li>• Risk of overfitting, potentially memorizing test data rather than learning underlying principles.</li> <li>• Dependent on FAERS data quality, relying on public reports that may contain inconsistencies or errors.</li> <li>• Lack of external clinical validation, requiring real-world evidence to confirm practical utility.</li> </ul>
9	Kim E, Nam H 2022	<ul style="list-style-type: none"> <li>• N/A</li> </ul>	<ul style="list-style-type: none"> <li>• Data sparsity and imbalance, suffering from insufficient and unevenly distributed data.</li> <li>• Limited gene selection, omitting certain segments of human DNA from analysis.</li> <li>• Heterogeneous cell line noise, resulting from inconsistent and messy laboratory testing conditions.</li> <li>• Difficult to predict transcriptome features from structure, challenging the inference of gene changes solely from drug shape.</li> </ul>
10	Lakizadeh A, Babaei M 2022	<ul style="list-style-type: none"> <li>• Integrated multiple heterogeneous drug features, utilizing diverse clues to resolve analytical challenges.</li> <li>• Strong validation via five-fold cross-validation, demonstrating reliability through rigorous testing protocols.</li> <li>• Improved performance on polypharmacy side effects, showing particular strength in predicting interactions involving multiple drugs.</li> </ul>	<ul style="list-style-type: none"> <li>• Unbalanced and noisy data, where an excess of negative examples confuses the model.</li> <li>• Missing biological features, failing to incorporate all potential biological components.</li> <li>• Need for stronger feature correlations, requiring better connection of disparate clues to improve results.</li> </ul>

11	Liu Q et al. 2022	<ul style="list-style-type: none"> <li>N/A</li> </ul>	<ul style="list-style-type: none"> <li>Accurate prediction but lacks specific side effect interpretation, identifying problems without specifying the exact nature of the issue.</li> </ul>
12	Ragkousis A et al. 2022	<ul style="list-style-type: none"> <li>First to fuse Molecular, DDI, and PPI+DTI data, combining three major data sources typically kept separate.</li> <li>Specialized GNN encoders for each source, utilizing specific tools to optimize reading of each data type.</li> <li>"Meta-fusion" for specific side effects, creating a custom data mix for each distinct side effect.</li> <li>Solves "Sparse Node" problem (17% AUPRC gain), performing effectively even for drugs with minimal available information.</li> </ul>	<ul style="list-style-type: none"> <li>N/A</li> </ul>
13	Lukashina N et al. 2022	<ul style="list-style-type: none"> <li>SimVec initializes with chemical embeddings (prior knowledge), starting with established chemical knowledge rather than random guesses.</li> <li>Weighted similarity edges propagate info, sharing information between similar drugs to aid prediction.</li> <li>3-step learning process, learning in stages to produce superior results.</li> <li>Cache-based negative sampling, selecting training examples intelligently to accelerate learning.</li> </ul>	<ul style="list-style-type: none"> <li>Limited scalability due to high number of weighted edges, becoming excessively slow and heavy when processing large drug datasets.</li> </ul>
14	Kim J, Shin M 2023	<ul style="list-style-type: none"> <li>Unified embedding integrated single/multi-drug data, combining distinct drug information types to provide a comprehensive analytical perspective.</li> <li>Virtual drug pairs reduced overfitting, employing synthetic examples to enhance training breadth and prevent rote memorization.</li> <li>Regularization improved robustness, preventing the model from becoming excessively sensitive to minor data fluctuations.</li> <li>Better prediction for new drug pairs, effectively estimating outcomes for previously unexamined drug combinations.</li> <li>Interpretable protein-side effect associations, clarifying the specific reasoning behind predicted side effects.</li> </ul>	<ul style="list-style-type: none"> <li>Limited by training data availability, requiring specific datasets that may not always be accessible for learning.</li> <li>High computational/GPU requirements, demanding expensive and high-performance hardware infrastructure.</li> <li>Cannot handle multi-typed edges, struggling with complex relationships involving diverse connection types.</li> <li>Generalizability depends on KG quality, meaning model success is strictly tied to the reliability of the background knowledge base.</li> <li>Limited real-world validation, lacking sufficient testing on actual patient data to confirm clinical utility.</li> </ul>
15	Masumshah R, Eslahchi C 2023	<ul style="list-style-type: none"> <li>Integrated 5 heterogeneous features to capture multi-dimensional relationships, employing five distinct data types to elucidate drug interrelations.</li> </ul>	<ul style="list-style-type: none"> <li>N/A</li> </ul>

16	Wang Y et al. 2023	<ul style="list-style-type: none"> <li>N/A</li> </ul>	<ul style="list-style-type: none"> <li> Ignores biochemical properties of drugs, overlooking the actual chemical mechanisms of drug action.</li> <li> Lacks multi-source data fusion, failing to combine diverse data sources for a more comprehensive answer.</li> </ul>
17	Lloyd O et al. 2024	<ul style="list-style-type: none"> <li> High performance (98.3%) in short time (~4 min), delivering both speed and high accuracy.</li> <li> Low electricity consumption/carbon footprint, reducing energy usage for better environmental sustainability.</li> </ul>	<ul style="list-style-type: none"> <li>N/A</li> </ul>
18	Keshavarz H, Lakizadeh A 2024	<ul style="list-style-type: none"> <li> PU learning handles data uncertainty, managing missing information intelligently rather than making incorrect assumptions.</li> <li> "SPP" biclustering extracts optimal features, identifying the best data patterns to prioritize.</li> <li> GNN captures hidden topological features, utilizing graph technology to uncover concealed connections.</li> </ul>	<ul style="list-style-type: none"> <li>N/A</li> </ul>
19	Gromova A, Maida A 2025	<ul style="list-style-type: none"> <li> Transformer embeddings improved DDI accuracy, utilizing advanced AI to enhance the understanding of chemical structures.</li> <li> Pretrained ChemBERTa reduced engineering, leveraging an existing chemistry-aware model to save development time.</li> <li> 99% variance retention via PCA, efficiently compressing data while preserving essential information.</li> <li> Outperformed 5 state-of-the-art models, demonstrating superior results compared to leading contemporary benchmarks.</li> <li> Open-source code and data, ensuring free accessibility for public download and utilization.</li> </ul>	<ul style="list-style-type: none"> <li> Extensive computational resources required, necessitating high-performance computing infrastructure for training.</li> <li> Performance depends on external dataset quality, limiting accuracy to the reliability of the source databases.</li> <li> Potential info loss in aggregation, risking the omission of important granular details during data summarization.</li> <li> Frozen embeddings limit fine-tuning, restricting the ability to adjust or retrain the core model components.</li> <li> No external clinical validation, limiting testing to in-silico environments rather than real-world clinical settings.</li> </ul>
20	Gromova A, Maida A 2025	<ul style="list-style-type: none"> <li> Combined transformer embeddings with clinical features, merging chemical knowledge with real-world medical data for enhanced results.</li> <li> Captured high-level molecular representations, demonstrating a deep understanding of complex drug molecule structures.</li> <li> Outperformed multiple baselines (Decagon, DeepWalk, etc.), achieving superior accuracy compared to established competitors.</li> <li> Scalable and adaptable framework, capable of handling vast datasets and adjusting to new drug compounds.</li> </ul>	<ul style="list-style-type: none"> <li> High memory and computational cost, consuming significant computer memory resources.</li> <li> Dependent on database completeness, creating knowledge gaps if the underlying database is incomplete.</li> <li> Limited embedding interpretability, making it difficult to precisely explain the decision-making logic of the model.</li> <li> No external clinical validation, lacking proof of efficacy in hospital or clinical environments.</li> <li> Frozen transformer weights, making it difficult to modify the fundamental parts of the model.</li> </ul>

21	Wang N, Taylor C 2025	<ul style="list-style-type: none"> <li>• N/A</li> </ul>	<ul style="list-style-type: none"> <li>• Small domain-specific dataset, containing too few examples for comprehensive learning.</li> <li>• Evaluated only on known side effects, failing to identify novel or unknown adverse events.</li> <li>• Lacks validation on real-world patient data (EHRs), necessitating testing with actual patient records to ensure reliability.</li> </ul>
22	Deng Z et al. 2025	<ul style="list-style-type: none"> <li>• MAVGAE framework predicts asymmetric interactions, recognizing that the effect of Drug A on Drug B differs from B on A.</li> <li>• Multimodal integration enhances accuracy/efficiency, combining various data types to improve intelligence and speed.</li> <li>• Handles heterogeneity well, effectively managing diverse types of information.</li> </ul>	<ul style="list-style-type: none"> <li>• Limited availability of asymmetric interaction datasets, making it difficult to find quality data for this specific prediction type.</li> <li>• Hinders comprehensive evaluation, complicating full testing due to data scarcity.</li> </ul>
23	Keshavarz H, Lakizadeh A 2025	<ul style="list-style-type: none"> <li>• PU learning addresses uncertainty, effectively managing uncertain data to improve accuracy.</li> <li>• Hybrid feature extraction (Random Forest + GNN), mixing different methods to obtain the best possible data representation.</li> <li>• MLP classifier surpasses other approaches, utilizing a simple yet powerful prediction tool that outperforms competitors.</li> <li>• Simpler and more efficient structure, featuring a streamlined design that enhances operation speed.</li> </ul>	<ul style="list-style-type: none"> <li>• N/A</li> </ul>

### Quality Assessment

To assess the quality of the included studies in the systematic review, the Mixed Methods Appraisal Tool (MMAT) was used because it enables standardized, consistent quality appraisal across studies with heterogeneous methodologies. Table 4 in the appendix shows full details of the assessment. 22 of 23 studies met all criteria and received a full score. This indicates strong methodological quality with fulfillment of all appraisal criteria. However, only one study received a score of 4 out of 5 because it did not meet one of the criteria. No study was excluded due to methodological weakness.

Table 4. Quality assessment summary using MMAT tool - Quantitative non randomized studies.

Study	Authors	Screening questions		Quantitative non-randomized studies					MMAT Score
		S1. Are there clear research questions?	S2. Do the collected data allow to address the research questions?	3.1. Are the participants representative of the target population?	3.2. Are measurements appropriate regarding both the outcome and intervention (or exposure)?	3.3. Are there complete outcome data?	3.4. Are the confounders accounted for in the design and analysis?	3.5. During the study period, is the intervention administered (or exposure occurred) as intended?	
Study 1	Zitnik M et al. 2018	Yes	Yes	Yes	Yes	Yes	Yes	Yes	5/5
Study 2	Burkhardt H et al. 2020	Yes	Yes	Yes	Yes	Yes	Yes	Yes	5/5
Study 3	Nováček V, Mohamed S 2020	Yes	Yes	Yes	Yes	Yes	Yes	Yes	5/5
Study 4	Dasgupta S et al. 2021	Yes	Yes	Yes	Yes	Yes	Yes	Yes	5/5
Study 5	MasumshahR et al. 2021	Yes	Yes	Yes	Yes	Yes	Yes	Yes	5/5
Study 6	Dewulf P et al. 2021	Yes	Yes	Yes	Yes	Yes	Yes	Yes	5/5
Study 7	Liu T et al. 2021	Yes	Yes	Yes	Yes	Yes	Yes	Yes	5/5
Study 8	Lin S et al. 2022	Yes	Yes	Yes	Yes	Yes	Yes	Yes	5/5
Study 9	Kim E, Nam H 2022	Yes	Yes	Yes	Yes	Yes	Yes	Yes	5/5
Study 10	LakizadehA, Babaei M 2022	Yes	Yes	Yes	Yes	Yes	Yes	Yes	5/5
Study 11	Liu Q et al. 2022	Yes	Yes	Yes	Yes	Yes	Yes	Yes	5/5
Study 12	RagkousisA et al. 2022	Yes	Yes	Yes	Yes	No	Yes	Yes	4/5
Study 13	LukashinaN et al. 2022	Yes	Yes	Yes	Yes	Yes	Yes	Yes	5/5
Study 14	Kim J, Shin M 2023	Yes	Yes	Yes	Yes	Yes	Yes	Yes	5/5
Study 15	MasumshahR,EslahchiC 2023	Yes	Yes	Yes	Yes	Yes	Yes	Yes	5/5
Study 16	Wang Y et al. 2023	Yes	Yes	Yes	Yes	Yes	Yes	Yes	5/5
Study 17	Lloyd O et al. 2024	Yes	Yes	Yes	Yes	Yes	Yes	Yes	5/5
Study 18	Keshavarz H,LakizadehA 2024	Yes	Yes	Yes	Yes	Yes	Yes	Yes	5/5
Study 19	Gromova A, Maida A	Yes	Yes	Yes	Yes	Yes	Yes	Yes	5/5

	2025								
<b>Study 20</b>	Gromova A, Maida A 2025	Yes	Yes	Yes	Yes	Yes	Yes	Yes	5/5
<b>Study 21</b>	Wang N, Taylor C 2025	Yes	Yes	Yes	Yes	Yes	Yes	Yes	5/5
<b>Study 22</b>	Deng Z et al. 2025	Yes	Yes	Yes	Yes	Yes	Yes	Yes	5/5
<b>Study 23</b>	Keshavarz H, Lakizadeh A 2025	Yes	Yes	Yes	Yes	Yes	Yes	Yes	5/5

## 4. Discussion

This systematic review evaluated the current applications of artificial intelligence and machine learning in pharmacovigilance, focusing on their capacity to detect, assess, and predict adverse drug reactions and drug-drug interactions in polypharmacy contexts. Among the 23 studies reviewed, AI/ML models consistently exhibited strong predictive performance in identifying and preventing ADRs, DDIs and other medication-related issues in data from polypharmacy patients, with most reporting AUROC and AUPRC values exceeding 0.90. These results support the expanding literature that identifies AI as a valuable computational complement to traditional pharmacovigilance systems, providing increased sensitivity, earlier detection of safety signals, and improved management of complex polypharmacy data [47,48]. This advancement is especially significant given the persistent limitations of spontaneous reporting systems, which are affected by substantial underreporting estimated at nearly 94% globally and consequent delays in recognizing clinically significant ADRs [3].

Beyond demonstrating high predictive performance across studies, this review identified that AI and machine learning systems were developed using a range of computational tools and programming environments. The majority of models utilized Python, with additional reliance on frameworks such as TensorFlow and the Adam optimizer. These platforms enabled the creation of diverse architectures capable of distinguishing complex risk signals within high-dimensional polypharmacy datasets. Furthermore, the findings support the increasing consensus that traditional pharmacovigilance approaches, which rely primarily on spontaneous reporting, are insufficient for the timely detection of medication-related harms, especially among populations exposed to multiple drugs simultaneously.

The performance of these models underscores the significant analytical advances enabled by contemporary AI architectures. Notably, models such as TriVec, SimpleE, SimVec, and PU-MLP achieved high effectiveness by integrating multiple sources of biomedical data, including chemical fingerprints, drug-target interactions, phenotype-based side effects, and molecular similarity embeddings. This multidimensional approach enabled the identification of interaction patterns that are often undetectable through human evaluation or conventional statistical methods. These results are consistent with broader AI-in-healthcare research demonstrating that combining diverse data streams enhances predictive accuracy, and they illustrate the ongoing shift toward precision-focused, data-intensive pharmacovigilance frameworks [49].

Both AUROC and AUPRC were consistently reported as primary evaluation metrics across all included studies. AUPRC was identified as more clinically meaningful for polypharmacy risk prediction, as adverse drug reactions and harmful interactions are rare but critical outcomes in highly imbalanced datasets. AUROC may be artificially inflated due to the prevalence of true negatives, whereas AUPRC, which emphasizes precision and recall in the positive class, provides a more reliable measure of a model's ability to detect clinically relevant risks while minimizing false positives [50,51]. Among the evaluated models, PU-MLP demonstrated the highest overall performance, followed by DPSP, SimVec, and TriVec. PU-MLP combines Random Forest, clustering techniques, Graph Neural Networks, and Positive Unlabeled learning to address data sparsity and uncertainty inherent in pharmacovigilance datasets, achieving strong predictive capability when assessed using benchmark datasets such as TWOSIDES and OFFSIDES.

Despite general alignment with external literature, several notable challenges and divergences persist. The cold-start problem, characterized by poor model performance when predicting interactions involving novel or underrepresented drugs, remains a significant limitation and reflects issues commonly reported in computational pharmacology. Additionally, none of the reviewed models underwent prospective or external clinical validation, resulting in a substantial gap between high in-silico accuracy and practical, real-world application. The absence of standardized datasets, inconsistent preprocessing methods, and variability in performance metrics further contributed to heterogeneity across studies and hindered precise model comparisons. These observations are

consistent with previous systematic reviews that highlight the necessity for harmonized reporting standards and improved benchmarking in AI-driven pharmacovigilance.

Methodologically, the included studies demonstrated several strengths that enhance the internal validity of the evidence base. Most employed clear analytical workflows, reproducible code, publicly available datasets, and robust validation techniques. The high MMAT scores (22 of 23 studies achieving 5/5) reflect strong adherence to computational research standards. Despite this, important methodological limitations persist. Heavy reliance on secondary datasets introduces uncertainties arising from labeling errors, reporting bias, noise, incompleteness, and imbalanced class distributions, all of which significantly affect the reliability of predictions. High-performing deep neural networks and transformer-based models also suffer from limited interpretability, raising ethical concerns regarding transparency and accountability. These challenges underscore the need for explainable AI frameworks that allow clinicians and regulators to understand how predictions are generated.

The analysis identified several advantages of artificial intelligence and machine learning models in improving pharmacovigilance, particularly in detecting complex, rare, or multi-drug interactions that traditional systems frequently overlook. The integration of diverse data sources, including drug-protein interaction networks and chemical structural information, enabled more comprehensive modeling of polypharmacy-related risks. However, numerous studies have indicated that data sparsity, class imbalance, and reliance on incomplete or noisy public datasets adversely affect predictive performance, particularly for rare interactions. These results help clarify the variability observed across models, even when similar datasets and evaluation metrics are employed.

Ethical considerations are also critical for interpreting these findings. In the absence of clinical validation, the premature deployment of AI models may lead to misclassification of ADRs, the overlooking of safety signals, or the perpetuation of biases present in training datasets. Models predominantly trained on data from the United States, Europe, or East Asia may not generalize effectively to underrepresented populations, raising concerns about fairness. Moreover, the lack of standardized performance reporting undermines reproducibility and may hinder regulatory assessment. Addressing these ethical and methodological challenges is essential to ensure the safe and equitable integration of AI tools into clinical and regulatory settings.

#### Strengths and limitations

This review demonstrates several strengths that enhance the reliability and relevance of its findings. The search strategy was comprehensive and systematically implemented across multiple major scientific databases, ensuring extensive coverage of literature concerning AI and ML applications in pharmacovigilance. Rigorous screening and data extraction procedures were conducted in accordance with PRISMA guidelines, which supports methodological consistency and transparency. Quality appraisal using the MMAT indicated that nearly all included studies achieved the highest possible score, reflecting strong methodological rigor within the primary evidence base. The review synthesizes findings across a range of AI models, datasets, and analytical approaches, providing a detailed overview of the current technological landscape in adverse drug reaction and drug-drug interaction prediction. By integrating results from multiple high-performing architectures and emphasizing both their strengths and limitations, this review offers a balanced, evidence-based assessment of AI's role in contemporary pharmacovigilance.

However, several limitations of this review should be acknowledged. The search was limited to English-language publications, potentially excluding relevant studies in other languages and introducing selection bias. Considerable heterogeneity was present among the included studies regarding model architectures, data sources, data preprocessing pipelines, and performance metrics. This variability restricted the ability to conduct direct comparisons or quantitative synthesis and may have contributed to inconsistencies in reported performance outcomes. The lack of standardized reporting frameworks across AI and ML pharmacovigilance studies further exacerbated this challenge. Furthermore, as all included studies relied on secondary in-silico datasets rather than real-world clinical data, external validity is limited, and findings cannot be readily generalized to patient

populations or real-time pharmacovigilance systems. These limitations underscore the need for harmonized methodologies, multilingual search strategies, and rigorous prospective validation in future research.

#### Recommendations

Based on the challenges and insights identified in this review, several targeted recommendations are proposed to advance AI-enabled pharmacovigilance. Given the strong performance of leading models such as PU-MLP, DPSP, and SimVec, it is essential to evaluate them in real-world clinical settings to determine whether their high AUROC and AUPRC values are reproducible outside controlled research environments. This recommendation aligns with the first research question outlined in the Discussion, underscoring the importance of assessing the effectiveness of these systems in identifying adverse drug reactions and drug-drug interactions within actual patient populations.

Future research should investigate how AI and machine learning models can adapt to evolving medication patterns and the ongoing introduction of new drugs, as highlighted in the second research question. Maintaining model responsiveness to changes in prescribing trends is critical for ensuring long-term accuracy and clinical relevance.

First, prospective clinical validation should be conducted by evaluating AI and machine learning models with real-world data from electronic health records, hospital medication systems, and active surveillance networks to establish external validity. Second, harmonization of datasets, performance metrics, and reporting guidelines is necessary to reduce methodological variability and facilitate accurate cross-study comparisons. Third, explainable AI approaches should be prioritized to improve transparency and clinician trust, especially for high-performing models with opaque internal reasoning. Policymakers and regulators should also establish robust frameworks for the evaluation, approval, and monitoring of AI tools in medication safety to ensure fairness, accountability, and patient protection. Finally, training and capacity-building initiatives for pharmacists, clinicians, and healthcare regulators should be expanded to support the safe and informed integration of AI systems into pharmacovigilance workflows. Simultaneously, improvements in data quality, model interpretability, and clinical usability are essential. The development of AI and machine learning tools that can be seamlessly integrated into routine clinical workflows will help ensure that these technologies effectively support medication safety in polypharmacy. Collectively, these actions will enable AI-enabled pharmacovigilance to transition from theoretical potential to practical, safe, and equitable implementation in real-world healthcare settings.

In response to the challenges identified in this review, two primary research questions are proposed for future investigation:

1. How does the performance of advanced AI and machine learning models, such as PU-MLP, differ when evaluated on real-world clinical data that are noisy, incomplete, and heterogeneous, as opposed to curated benchmark datasets?
2. Does incorporating additional real-world information, such as patient history or prescribing context, enhance the predictive accuracy of AI and machine learning models for rare adverse drug reactions and drug-drug interactions in patients receiving polypharmacy?

## 5. Conclusions

This review shows that AI and ML models have strong potential to improve pharmacovigilance, particularly in polypharmacy, where the risk of adverse drug reactions and drug-drug interactions is high. Across 23 studies, model performance was consistently strong, indicating reliable prediction of harmful medication combinations. The best-performing models—PU-MLP, DPSP, and SimVec—achieved near-perfect scores, highlighting the ability of advanced algorithms to capture complex drug relationships and predict adverse outcomes with high precision. Common advantages included strong predictive capability, integration of diverse data sources, and scalability for large-scale screening. However, challenges remain, such as dependence on data quality, limited interpretability, lack of external validation, and practical barriers to clinical integration. Overall, AI and ML offer

promising tools to enhance medication safety in polypharmacy. Future efforts should focus on improving data quality, model transparency, and real-world validation to enable successful clinical adoption.

**Author Contributions:** Conceptualization, Visualization, Writing (Original Draft)— KT, MG, RT, MIMI; Writing (Reviewing and Editing)— RA, MIMI; Project Administration— all authors. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** No new data were created or analyzed in this study. Data sharing is not applicable to this article.

**Acknowledgments:** The APC support is provided by Qatar University.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

ADR	Adverse drug reactions
DDIs	Drug–drug interactions
AI	Artificial intelligence
ML	Machine learning
PV	Pharmacovigilance
AUROC	Area under the receiver operating characteristic curve
AUPRC	Area under the precision–recall curve

## References

1. World Health Organization. The importance of pharmacovigilance: safety monitoring of medicinal products. Geneva: WHO; 2002.
2. Edwards IR, Aronson JK. Adverse drug reactions: definitions, diagnosis, and management. *Lancet*. 2000;356(9237):1255–1259.
3. Hazell L, Shakir SAW. Under-reporting of adverse drug reactions: a systematic review. *Drug Saf*. 2006;29(5):385–396.
4. Maher RL, Hanlon J, Hajjar ER. Clinical consequences of polypharmacy in elderly. *Expert Opin Drug Saf*. 2014;13(1):57–65.
5. Alhawassi TM, Krass I, Bajorek BV, Pont LG. A systematic review of the prevalence and risk factors for adverse drug reactions in the elderly in the acute care setting. *Clin Interv Aging*. 2014;9:2079–2086.
6. Beijer HJ, de Blaey CJ. Hospitalisations caused by adverse drug reactions (ADR): a meta-analysis of observational studies. *Pharm World Sci*. 2002;24(2):46–54.
7. Waller P, Evans SJW. A model for the future conduct of pharmacovigilance. *Pharmacoepidemiol Drug Saf*. 2003;12(1):17–29.
8. Bate A, Evans SJW. Quantitative signal detection using spontaneous ADR reporting. *Pharmacoepidemiol Drug Saf*. 2009;18(6):427–436.
9. Russell S, Norvig P. *Artificial Intelligence: A Modern Approach*. 3rd ed. Pearson; 2010.
10. Harpaz R, DuMouchel W, LePendou P, Bauer-Mehren A, Ryan P, Shah NH. Performance of pharmacovigilance signal-detection algorithms for the FDA adverse event reporting system. *Clin Pharmacol Ther*. 2013;93(6):539–546.
11. Sarker A, Gonzalez G. Portable automatic text classification for adverse drug reaction detection via multi-corpus training. *J Biomed Inform*. 2015;53:196–207.

12. Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics*. 2018;34(13):i457–i466.
13. Zhang W, Chen Y, Liu F, Luo F, Tian G, Li X. Predicting drug–drug interactions based on multi-source information fusion. *J Cheminform*. 2017;9:3.
14. Deng Y, Xu X, Qiu Y, Xia J, Zhang W, Liu S. A multimodal deep learning framework for predicting drug–drug interactions. *Bioinformatics*. 2020;36(15):4316–4322.
15. Zhou T, Wang X, Li J. PU-learning based framework for polypharmacy side effect prediction. *Brief Bioinform*. 2025;26(1):bbad123.
16. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell*. 2019;1:206–215.
17. Doshi-Velez F, Kim B. Towards a rigorous science of interpretable machine learning. arXiv preprint. 2017;arXiv:1702.08608.
18. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;25:44–56.
19. Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ*. 2021;372:n160.
20. Programme CAS. How to use the PICO framework: Critical Appraisal Skills Programme; [Available from: <https://casp-uk.net/pico-framework/>]
21. AUC-ROC curve in Machine Learning: GeeksforGeeks; 2025 [Available from: <https://www.geeksforgeeks.org/machine-learning/auc-roc-curve/>].
22. Precision-Recall curve in machine learning: GeeksforGeeks; 2025 [Available from: <https://www.geeksforgeeks.org/machine-learning/precision-recall-curve-ml/>].
23. Hong QN PP, Fàbregues S, Bartlett G, Boardman F, Cargo M, Dagenais P, et al. Mixed Methods Appraisal Tool (MMAT), version 2018.
24. Dasgupta S, Jayagopal A, Hong ALJ, Mariappan R, Rajan V. Adverse Drug Event Prediction Using Noisy Literature-Derived Knowledge Graphs: Algorithm Development and Validation 2021 2021-10-1. e32730 p.
25. Deng Z, Xu J, Feng Y, Dong L, Zhang Y, Deng Z, et al. MAVGAE: a multimodal framework for predicting asymmetric drug-drug interactions based on variational graph autoencoder. *Computer Methods In Biomechanics And Biomedical Engineering*. 2025;28(7):1098-110.
26. Gromova AA, Maida AS. ChemBERTaDDI: Transformer Driven Molecular Structures and Clinical Data for predicting drug-drug interactions. 2025. bioRxiv 2025.01.22.634309; doi: <https://doi.org/10.1101/2025.01.22.634309>.
27. Gromova AA, Maida AS. ChemBERTaDDI: Transforming Drug-Drug Interaction Prediction with Transformers and Clinical Insights2025 2025-1-25.
28. Burkhardt HA, Subramanian D, Mower J, Cohen T. Predicting Adverse Drug-Drug Interactions with Neural Embedding of Semantic Predications. *AMIA Annual Symposium proceedings AMIA Symposium*. 2019;2019:992-1001.
29. Keshavarz A, Lakizadeh A. PU-MLP: A PU-learning based method for polypharmacy side-effects detection based on multi-layer perceptron and feature extraction techniques. *Intell Based Med*. 2025;12.
30. Keshavarz A, Lakizadeh A, Keshavarz A, Lakizadeh A. PU-GNN: A Positive-Unlabeled Learning Method for Polypharmacy Side-Effects Detection Based on Graph Neural Networks. *International Journal Of Intelligent Systems*. 2024;2024.
31. Kim E, Nam H, Kim E, Nam H. DeSIDE-DDI: interpretable prediction of drug-drug interactions using drug-induced gene expressions. *Journal Of Cheminformatics*. 2022;14(1).
32. Kim J, Shin M, Kim J, Shin M. A Knowledge Graph Embedding Approach for Polypharmacy Side Effects Prediction. *Applied Sciences-Basel*. 2023;13(5).
33. Lakizadeh A, Babaei M, Lakizadeh A, Babaei M. Detection of polypharmacy side effects by integrating multiple data sources and convolutional neural networks. *Molecular Diversity*. 2022;26(6):3193-203.
34. Lin S, Zhang G, Dong-Qing W, Xiong Y. DeepPSE: Prediction of polypharmacy side effects by fusing deep representation of drug pairs and attention mechanism2022 2022-10.

35. Liu Q, Yao E, Liu C, Zhou X, Li Y, Xu M, et al. M2GCN: multi-modal graph convolutional network for modeling polypharmacy side effects. *Applied Intelligence*. 2023;53(6):6814-25.
36. Liu T, Cui J, Zhuang H, Wang H, Liu T, Cui J, et al. Modeling polypharmacy effects with heterogeneous signed graph convolutional networks. *Applied Intelligence*. 2021;51(11):8316-33.
37. Lukashina N, Kartysheva E, Spjuth O, Virko E, Shpilman A. SimVec: predicting polypharmacy side effects for new drugs2022 2022-12.
38. Masumshah R, Aghdam R, Eslahchi C. A neural network-based method for polypharmacy side effects prediction2021 2021. 1 p.
39. Masumshah R, Eslahchi C, Masumshah R, Eslahchi C. DPSP: a multimodal deep learning framework for polypharmacy side effects prediction. *Bioinformatics Advances*. 2023;3(1).
40. Lloyd O, Liu Y, Gaunt TR. Fast polypharmacy side effect prediction using tensor factorization. *Bioinformatics (Oxford, England)*. 2024;40(12).
41. Ragkousis A, Flogera O, Megalooikonomou V, Ragkousis A, Flogera O, Megalooikonomou V. MFSE: A Meta-Fusion Model for Polypharmacy Side-Effect Prediction with Graph Neural Networks2022 2022. 563-70 p.
42. Stock M, De Baets B. Cold-Start Problems in Data-Driven Prediction of Drug–Drug Interaction Effects2021 2021. 429 p.
43. Nováček V, Mohamed SK. Predicting Polypharmacy Side-effects Using Knowledge Graph Embeddings. *AMIA Joint Summits on Translational Science proceedings AMIA Joint Summits on Translational Science*. 2020;2020:449-58.
44. Wang N, Taylor C, Wang N, Taylor CO. DrIVeNN: Drug Interaction Vectors Neural Network. *Journal Of Computational Biology*. 2025.
45. Wang Y, Ma H, Zhang R, Gao Z. Gorge: graph convolutional networks on heterogeneous multi-relational graphs for polypharmacy side effect prediction2023 2023-4. 6 p.
46. Zitnik M, Agrawal M, Leskovec J, Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. *BIOINFORMATICS*. 2018;34(13):457-66.
47. Rao PCV. Adverse Drug Reactions and Poly-Pharmacy in Geriatric Patients: A Pharmacovigilance-Based Observational Study. *Journal of Population Therapeutics & Clinical Pharmacology (JPTCP)*. 2025;32(3):181–8.
48. Pathuri V. Revolutionizing Drug Safety: The Role of Artificial Intelligence and Machine Learning in Pharmacovigilance. *International Journal of Computer Science and Information Technology Research (IJCSITR)*. 2025;6(1):51–61.
49. Al-Nafjan A, Aljuhani A, Alshebel A, et al. Artificial Intelligence in Predictive Healthcare: A Systematic Review. *Journal of Clinical Medicine*. 2025;14(19):6752. <https://doi.org/10.3390/jcm14196752>
50. Lee S. Mastering AUPRC in Biomedical ML: Unlock the Power of Precision-Recall Curves for Enhanced Model Evaluation. *Number Analytics Blog*. June 30, 2025.
51. Li Y. AUROC and AUPRC. *Medium*. <https://medium.com/@yukims19/auroc-and-auprc-b25183827e5a>.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.