**Computational Prediction of Adverse Drug Reactions and Toxicity Using**

**AI and ML**

**ABSTRACT**

Adverse drug reactions (ADRs) and drug-induced toxicity pose significant challenges in healthcare, jeopardizing patient safety and escalating healthcare costs. Despite their subtlety compared to infectious diseases, their ramifications are profound. Early detection of ADRs and toxicity is crucial for assessing a drug's safety and efficacy. The integration of artificial intelligence (AI) and machine learning (ML) techniques has revolutionized early detection practices, enabling swift and precise prediction of potential ADRs and toxicity even before drug synthesis, preclinical, and clinical trials. This review amalgamates the roles of AI and ML in early detection, encompassing diverse methodologies from data mining to deep learning. It provides comprehensive insights into databases, modeling algorithms, and software pertinent to ADR and toxicity prediction. By elucidating the capabilities of these technologies, it underscores their potential to transform drug discovery and enhance patient safety.

The abstract also examines AI's transformative impact on healthcare, particularly in advanced drug monitoring. AI facilitates real-time data analysis and continual surveillance, improving drug efficacy and reducing adverse reactions. Leveraging sophisticated algorithms, AI analyzes intricate datasets, enabling personalized treatment plans and precision medicine. Moreover, AI-driven systems mitigate risks, reduce errors, and optimize patient care, ensuring improved health outcomes. The abstract anticipates AI's future trajectory in drug monitoring, addressing ethical and regulatory considerations. It emphasizes AI's role in shaping a more efficient, personalized, and patient-centric healthcare paradigm, promising to redefine healthcare delivery and elevate patient outcomes.

**Keyword:** **Machine Learning, Adverse Drug Reaction, Drug Monitoring, Patient Safety, Precision Medicine**

**Introduction**

The process of developing pharmaceuticals is arduous and expensive, often taking over a decade and costing billions of dollars. Despite rigorous clinical trials and regulatory oversight, adverse drug reactions (ADRs) and drug-induced toxicity persist as significant challenges in healthcare, leading to hospitalizations and substantial costs.[1,2,3,4]Traditional detection methods, reliant on post-marketing surveillance, are reactive and may fail to identify harmful effects until they affect a large number of patients.[5,6,7]

ADRs, unintended reactions to medications, contribute significantly to global hospital admissions, underscoring the ongoing prevalence of this issue. Additionally, drug-induced toxicity, an extreme form of ADR, can result in organ damage and severe health consequences. These challenges pose clinical and economic burdens, emphasizing the need for early detection and mitigation strategies.[8,9]

Improving early detection could mitigate patient harm, enhance safety, reduce costs, and streamline drug development. Quantitative Structure-Activity Relationship (QSAR) models offer a promising approach, using mathematical equations to predict drug toxicity based on chemical structure. Integrating bioinformatics and genomics enhances our ability to predict toxicity by analyzing diverse datasets.[10,11,12]

Artificial Intelligence (AI) and Machine Learning (ML) are revolutionizing healthcare, offering efficient processing of large datasets and predictive capabilities. They show promise in predicting ADRs by analyzing electronic health records and genomic data. ML algorithms can identify genetic factors contributing to drug responses, improving our understanding of toxicity.[13,14]

The integration of AI and ML into QSAR models has significantly enhanced predictive toxicology. Deep learning, a subset of ML, yields more accurate predictions and handles complex chemical data, aiding in early ADR identification. This approach holds promise for safer drug development and improved patient outcomes.[15]

This review emphasizes the importance of early detection in drug design, discusses available toxicity prediction tools and databases, and explores the innovative applications of AI and ML in ADR and toxicity modeling across various categories.[16]

1. **Medication recalls: the rationale for early identification of adverse drug reactions (ADRs) and drug-induced toxicity**

Thalidomide, introduced in the late 1950s as a sedative and remedy for morning sickness, gained widespread use across Europe, Australia, and Japan [17]. However, within a few years, it resulted in the birth of nearly 10,000 children with phocomelia, leading most countries to ban its use by 1961, although some continued distribution for a while [18]. Thalidomide's adverse effects extended beyond limb malformations to include congenital heart issues, ear deformities, and visual impairments [19]. This tragic event underscored the importance of rigorous pharmaceutical testing before a drug's release to the public [20].

Introduced by Merck in 1999, Rofecoxib (Vioxx) was marketed as a safer alternative to non-steroidal anti-inflammatory drugs, specifically targeting osteoarthritis-related pain management [21]. The APPROVe clinical trial (Adenomatous Polyp Prevention of Vioxx), a three-year study evaluating Rofecoxib's efficacy in preventing colorectal polyp recurrence in patients with a history of colorectal adenomas, was halted prematurely in late September, two months ahead of schedule. The trial revealed a significantly increased risk of severe cardiovascular events, such as strokes and heart attacks, emerging after 18 months of Rofecoxib treatment. This risk was approximately double that of patients receiving a placebo [22]. However, during the initial 18-month period of the study, there were no indications of increased risk. Table 1 provides examples of significant drug withdrawals resulting from adverse drug reactions (ADRs) and drug-induced toxicities.

**TABLE 1. List of withdrawal drugs over the years due to the ADRs and toxicities.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Drug name**  | **Brand**  | **Drug intend to use for**  | **Reason of withdrawal**  | **Year of Approval**  | **Year of withdrawal**  |
| **Umbralisib****[23]** | **Ukoniq** | **Treatment of adult patients with Marginal Zone Lymphoma and Follicular Lymphoma that are relapsed or refractory** | **The overall survival (OS)****results from the UNITY -CLL Phase 3 study revealed a growing OS imbalance**  | **2021**  | **2022** |
| **Belantamab mafodotinblmf****[24]** | **Blenrep**  | Intended for the treatment of adults with multiple myeloma that is relapsed or refractory and who have received at least four prior therapies, including an anti CD-38 monoclonal antibody, a proteasome inhibitor, and an immunomodulatory agent. | **The outcome of the DREAMM 3 Phase II Confirmatory trial did not meet the requirements of the FDA accelerated approval regulations**  | **2020** | **2023** |
| **Troglitazone[ 25]**  | **Rezulin**  | **Type 2 Diabetes Mellitus**  | **Liver Toxicity**  | **1997** | **2000** |
| **Sibutranine [26]**  | **Meridia**  | **Weight Loss and Maintenance**  | **Increased Risk of heart attack and stroke**  | **1997** | **2010** |

1. **ADRs data Resources**

ADRs refer to any adverse, undesired, or unforeseen reactions to a medication, whether intended for preventing, diagnosing, or treating diseases or altering physiological functions [36]. This aspect of drug response undergoes thorough scrutiny at various stages, including pre-clinical assessments, clinical trials, and post-marketing surveillance [27].

In clinical settings, healthcare professionals play a crucial role in monitoring and reporting ADRs. These reactions can be identified through patient assessments, lab tests, and monitoring medication responses. Healthcare providers often notice and report ADRs, especially with newly approved drugs. Patients are encouraged to inform their healthcare providers or national regulatory authorities about any unexpected effects. Pharmacovigilance, particularly post-marketing surveillance, is vital for ongoing drug safety monitoring, encompassing ADRs and toxicity. In the United States, the FDA manages the Adverse Event Reporting System (FAERS) to track reported ADRs, serving as a valuable resource for real-world data and enhancing understanding of drug safety profiles [28].

Official resources provide information on ADRs associated with specific drugs. The FDA offers a comprehensive database of approved drugs on its website, each with links to prescribing information detailing known ADRs, available at: https://www.accessdata.fda.gov/scripts/cder/daf/. Additionally, the U.S. National Library of Medicine's DailyMed service furnishes reliable data on marketed drugs, including FDA-approved labels listing known ADRs. Users can simply search for a specific drug on the DailyMed website to access the 'Adverse Reactions' section of its labeling for a detailed list of known ADRs, accessible at https://dailymed.nlm.nih.gov/dailymed/.[29,30]

1. **Drug redesigning and restructuring**

Another facet of pharmaceutical advancement involves the redesign and restructuring of drugs, aiming to enhance therapeutic effectiveness while minimizing adverse effects [31]. This procedure entails modifying a drug's molecular composition to enhance its pharmacological characteristics, including bioavailability, potency, selectivity, and stability, among other factors [32].

1. **Types of ADRs and toxicity to be checked for drug development**

**5.1 Mechanism of drug induced toxicity**

Prior to delving into the particular mechanisms of drug-induced toxicity, it's crucial to establish a shared comprehension of how this toxicity is delineated and measured. One characterization of drug toxicity involves the concentration at which a drug or its metabolite inflicts harm upon an organism. This harm can manifest through various mechanisms, ranging from direct cellular harm to unintentional off-target effects.[33]

**5.1.1 Direct cellular damage**

Medications have the potential to induce toxicity by directly harming cells, a process in which the drug or its metabolites disrupt cellular structures or biochemical functions. This disruption can result in cell damage or death, ultimately eliciting an adverse reaction.[34]

**5.1.2 Immune mediated-Response**

The immune system is essential for safeguarding our well-being by shielding us from harmful agents. Nevertheless, at times, it may erroneously recognize drugs or their metabolites as foreign entities, prompting an immune reaction known as drug-induced immune-mediated responses.[35]

**5.1.3 Metabolite induced- toxicity**

Toxicity induced by metabolites, commonly referred to as bioactivation, represents a significant form of drug toxicity. In this process, a drug undergoes metabolic conversion, typically within the liver, resulting in the production of a harmful metabolite.[36]

**5.1.4 off target effects**

Another plausible origin of ADRs and toxicity arises from off-target effects. These instances occur when a drug interacts with unintended biological targets, resulting in unforeseen and occasionally adverse side effects.[37]

**5.2 Measurements of toxicity**

Assessing the safety profile of new drugs relies heavily on toxicity testing, which encompasses various methodologies offering distinct insights into potential harm to biological organisms. One traditional method is the LD50, determining the dose causing mortality in 50% of test animals, pivotal in toxicology since the 1920s, offering a quantitative toxicity measure for substance comparisons [38]. Nonetheless, LD50 is just one of many toxicity endpoints; modern assessments encompass diverse measures. These include acute toxicity measures like the median effective dose (ED50), indicating the dose eliciting a specified effect in 50% of the population. Sub-chronic and chronic toxicity studies investigate repeated exposure effects over prolonged periods, focusing on organ damage, cancer, or long-term health implications. The no observed adverse effect level (NOAEL), defining the highest dose with no adverse effects, aids risk assessment to establish safe human exposure levels [39]. Further tests target specific harms like genotoxicity (DNA damage), teratogenicity (birth defects), and neurotoxicity (nervous system harm). Advancements in high-throughput screening and computational toxicology promise accelerated and comprehensive toxicity testing, minimizing animal usage and offering pertinent data for human health risk assessment[40]

**5.3 Classification of ADRs**

ADRs pose a significant concern in healthcare, contributing to patient morbidity, mortality, and escalated healthcare expenses. Throughout the years, various classification systems have emerged to enhance comprehension, prediction, and prevention of these occurrences. Two extensively recognized systems are the Rawlins and Thompson classification and the DoTS classification.

Introduced in 1977, the Rawlins and Thompson classification divides ADRs into Type A and Type B [41].

Type A reactions, predominant among ADRs, are dose-dependent and foreseeable based on a drug's known pharmacological actions. In contrast,

Type B reactions are less frequent, unpredictable, and not dose-dependent, often necessitating drug discontinuation due to their severity [42]. This classification system's advantage lies in its simplicity and effectiveness in identifying prevalent and predictable reactions (Type A), which are frequently preventable.

**6.4 Major types of drugs induced toxicities**

**6.4.1 Hepatotoxicity**

Hepatotoxicity denotes liver damage induced by chemical substances, particularly certain medications [43]. Medications are a leading contributor to liver injury, with over 1000 drugs identified as hepatotoxic [44.]

**6.4.2 Nephrotoxicity**

Nephrotoxicity denotes the damaging impact on the kidneys caused by specific substances, such as medications and environmental toxins. Various drugs, including certain antibiotics, non-steroidal anti-inflammatory drugs (NSAIDs), chemotherapy drugs, and contrast agents utilized in radiological exams, have been linked to nephrotoxicity.[45.46]

**6.4.3 Cardiotoxicity**

Cardiotoxicity is characterized by heart dysfunction or muscle impairment due to exposure to specific drugs or toxins[47]. It can present in various forms, including myocardial infarction, heart failure, arrhythmias, and alterations in blood pressure. Anthracyclines, such as doxorubicin, utilized in cancer chemotherapy, are among the most infamous drug classes associated with cardiotoxic effects[48].

**6.4.4 Neurotoxicity**

Neurotoxicity pertains to harm or negative impacts on the structure or operation of either the peripheral or central nervous system (CNS) resulting from exposure to toxic substances, including specific medications [49].

**6.4.5 Carcinogenicity , geno-toxicity, and mutagenicity**

Assessing carcinogenicity, genotoxicity, and mutagenicity is pivotal in the process of drug development and assessment, as they address the potential of a drug to trigger cancer or genetic harm, which can profoundly impact long-term patient well-being. Carcinogenicity specifically concerns a substance's capability to initiate cancer or stimulate its progression. Drugs demonstrating carcinogenic traits warrant serious attention, particularly those intended for prolonged usage, as they could elevate the likelihood of cancer development over time [50].

**6.4.6 Skin Sensitization**

Skin sensitization is a critical safety issue during the development of novel drugs, as it encompasses an allergic reaction upon skin contact with specific substances, resulting in dermatitis or skin inflammation. This reaction is intricate and entails a two-step process of immune system activation: induction, where the immune system adapts to respond, and elicitation, where the response is provoked upon subsequent exposure [51].

1. **AI and ML methods for early detection of ADRs**

AI and ML have shown tremendous potential in transforming drug development procedures, opening up new avenues for more effective, quicker, and cost-efficient approaches to discovering and developing innovative therapeutics [52]. AI, a broader concept, mimics human intelligence processes through machines, particularly computer systems, and can execute tasks like learning, reasoning, problem-solving, perception, and language comprehension [53]. Machine learning, a subset of AI, is a technique for data analysis that automates the construction of analytical models. It operates on the principle that systems can learn from data, recognize patterns, and make decisions with minimal human intervention [54]. These computational technologies find application across various stages of drug development, including target identification, drug discovery, preclinical testing, and clinical trials [55]. During target identification and drug discovery, AI/ML models can forecast protein structures and potential drug-target interactions, thereby diminishing the necessity for extensive laboratory work [50]. In preclinical testing, predictive models are formulated using machine learning algorithms to evaluate the potential toxicity of new compounds, further refining safety profiles and reducing failure rates in subsequent phases [1]. In clinical trials, AI/ML algorithms can expedite patient recruitment, forecast patient outcomes, and monitor adverse events, streamlining the trial process and prioritizing patient needs [43]. Nevertheless, despite the considerable promise AI and ML hold for drug development, their implementation also poses challenges. These encompass issues such as data quality and availability, model transparency and interpretability, and regulatory and ethical considerations, which require meticulous attention for successful integration [22].

**7.1 Data mining for ADR and toxicity detection**

Data mining, a pivotal aspect of AI and ML, serves as a significant tool for promptly identifying adverse drug reactions and toxicity. This procedure entails identifying, extracting, and analyzing patterns and valuable insights from extensive datasets, such as electronic health records (EHRs), clinical trial data, post-marketing surveillance data, and even social media posts [60]. One primary method utilized for adverse drug reaction (ADR) detection in data mining is signal detection. This approach relies on statistical algorithms to identify 'signals', patterns suggesting a higher-than-expected correlation between drug use and reported adverse events [32]. By detecting potential ADRs before clinical confirmation, this approach offers an early warning system for clinicians and regulatory bodies. Machine learning techniques like neural networks, decision trees, and support vector machines have been effectively employed in data mining to develop predictive models for ADRs and toxicity [34]. These models are trained on existing data, learning the patterns associated with confirmed ADRs, and then applied to new data to forecast potential ADRs. Data mining also greatly enhances toxicity detection. For example, the use of ML in predicting drug-induced liver injury (DILI) has led to algorithms capable of analyzing extensive databases of drug compounds and predicting potential hepatotoxic effects [13]. This capability to identify toxicity early in the drug development process diminishes the risk of costly failures in later stages and contributes to ensuring patient safety [44]. Despite its potential, data mining for ADR and toxicity detection faces challenges. Foremost among these are issues such as data quality and privacy, the interpretability of machine learning models, and the necessity for careful validation to prevent false positives [25].

**7.2 Machine Learning for Predicting ADRs and Drug toxicity**

**7.2.1 Traditional Machine Learning methods**

Machine learning (ML) plays a vital role in advancing drug toxicity prediction within the realm of artificial intelligence. By employing algorithms to analyze and derive insights from existing datasets, ML facilitates the anticipation of outcomes on novel data sets. Within drug toxicity prediction, ML models excel at detecting correlations between toxic effects and various molecular and clinical data attributes. A variety of ML models, including decision trees, neural networks, support vector machines, and deep learning techniques, are utilized for this purpose. For example, decision trees have been effectively employed to forecast hepatotoxicity based on compound structural characteristics. Deep learning, a subset of ML inspired by the human brain's architecture, demonstrates remarkable accuracy in predicting drug toxicity, particularly when handling vast and intricate biological data sets. The applicability of ML extends beyond preclinical stages, proving valuable in clinical trial settings and post-marketing surveillance. ML models enable the prediction of individual patient risk by considering their unique genomic and phenotypic traits, thereby facilitating a more personalized approach to managing toxicity risk. Nonetheless, ML-based toxicity prediction encounters challenges, notably concerning the quality and representativeness of training data. Biases within data sets can skew predictions, while issues surrounding data privacy and sharing can restrict the availability of suitable training data for ML applications. Moreover, the complexity and lack of interpretability inherent in many ML models pose obstacles to comprehending and validating their predictions.[2,15,28].

**7.2.2 Advanced mathematics-based machine learning**

Recent years have witnessed a growing trend at the intersection of mathematics, biology, and machine learning, particularly in the field of molecular property prediction, specifically toxicity. Advanced machine learning models leveraging sophisticated mathematics-based featurization techniques, incorporating principles of topology, geometry, and graph theory, have emerged. For example, Wu et al. introduced the element-specific persistent homology (ESPH) model, which employs topology-based multitask deep neural networks for quantitative toxicity prediction. This approach preserves critical chemical information and outperforms other methods in representing small molecules. Additionally, the algebraic graph-assisted bidirectional transformer (AGBT) framework combines algebraic graphs, bidirectional transformers, and various machine learning algorithms to achieve state-of-the-art performance in molecular property prediction tasks, including toxicity. Similarly, the geometric graph learning toxicity (GGL-Tox) model proposed by Jiang et al. integrates multiscale weighted colored graph (MWCG) features with the gradient boosting decision trees (GBDT) algorithm, demonstrating accuracy and efficiency in toxicity prediction. Furthermore, advanced mathematics-based learning models have been applied in drug repurposing efforts, such as Feng et al.'s machine learning screening of DrugBank compounds for opioid use disorder treatment. These models systematically analyze compound binding affinities on opioid receptors, considering absorption, distribution, metabolism, excretion, and toxicity properties. Overall, the integration of mathematics, biology, and machine learning is pushing the boundaries of molecular property prediction, including toxicity, with anticipated increasing significance in drug development.[19,40,].

**7.2.3 QSAR**

Quantitative Structure-Activity Relationships (QSAR) have become crucial in computationally predicting drug toxicity, employing rigorous validation criteria and applicability domain (AD) identification for reliable predictions. QSAR models correlate chemical biological activity, including toxicity, with physicochemical properties or theoretical molecular descriptors. These models utilize various machine learning algorithms, ranging from linear regression to neural networks, to predict toxicity endpoints like mutagenicity, carcinogenicity, and hepatotoxicity. QSAR models serve as initial filters in drug development by swiftly screening compound libraries for potential candidates with favorable toxicity profiles. However, constructing and validating robust QSAR models require large datasets, quality molecular descriptors, and careful selection of machine learning methods. Despite challenges, QSAR models have significantly improved early detection of drug toxicity, enhancing the safety and efficiency of drug development processes.[23]

**7.3 Deep learning**

Deep learning, a subset of AI, mimics human brain processing to make data-driven decisions and has become increasingly utilized in drug discovery, notably in predicting adverse drug reactions (ADRs). Its ability to handle large, complex datasets, including genomic and proteomic data, sets it apart from other machine learning models. Deep learning excels in detecting intricate structures within data and has shown superior performance in tasks such as predicting protein structures. Researchers leverage algorithms like Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs) to predict ADRs by processing sequential and image data, respectively. For instance, Zeng et al. achieved excellent results in ADR prediction using deep neural networks to extract features from chemical structures and biological systems. Despite its advantages, challenges remain, including the need for large data sets and computational resources, as well as interpretability issues in medical applications[46]

**7.4 Al in Pharmacovigilance**

Pharmacovigilance focuses on detecting, assessing, understanding, and preventing adverse drug effects, prioritizing safe and rational medication use. AI is increasingly used in pharmacovigilance to speed up and improve adverse drug reaction (ADR) detection. Machine learning (ML) algorithms automate the review of large pharmacovigilance databases like the FDA Adverse Event Reporting System. AI-driven text mining techniques efficiently detect ADRs from various sources, including electronic health records and social media. AI models, particularly those based on ML, enhance signal detection by identifying complex patterns in data, complementing traditional methods. Despite its transformative potential, challenges such as data privacy and model interpretability need addressing to fully exploit AI's benefits in pharmacovigilance.[19,40,45].

**7.5 AI and ML in drug redesign and restructuring**

AI and ML have become essential in drug redesign and restructuring, assisting in the development of new drug candidates with enhanced safety and efficacy profiles. These technologies allow researchers to digitally generate and test thousands of potential drug structures, expediting drug development significantly. ML algorithms within AI systems can learn from existing drug structures and their pharmacological attributes, enabling the prediction of therapeutic potential for novel structures. For example, ML algorithms can analyze critical parameters like pharmacokinetics, drug-target interactions, and toxicity profiles to create improved analogs of existing drugs. Additionally, AI aids in predicting potential off-target effects, offering a thorough safety analysis of drug candidates before costly clinical trials. It also contributes to understanding structure-activity relationships, identifying modifications that boost potency and reduce toxicity during drug optimization. Moreover, AI and ML support drug repurposing efforts by restructuring and retesting existing drugs for alternative therapeutic uses, leveraging existing toxicity and safety data to accelerate the development process significantly.[12,53,44]

1. **Al and ML tools and software for modeling and predicting of Drugs ADR and toxicity**

AI and ML-driven computational methods have transformed the evaluation of drug safety, emerging as crucial facilitators in this field. These tools accelerate the drug development process while ensuring the safety and effectiveness of new treatments through modeling and predicting adverse drug reactions (ADR) and toxicity. They mitigate risks associated with medication development, such as unforeseen side effects, costly late-stage clinical trial failures, and post-market withdrawals. The subsequent examination in Table 2 will assess the capabilities and applications of available AI and ML tools, including those accessible via open access, web-based platforms, and commercial sources.

Table 2. Al and ML - based tools and software for the prediction and modeling of drugs ADR and toxicity.

|  |  |  |  |
| --- | --- | --- | --- |
| Tools | Type | Usage | Description  |
| ProTox-II  | Free access to web server  | Toxicity Prediction  | A total 33 models for toxicity endpoints prediction  |
| pkCSM  | Free access to web server  | ADMET Prediction  | 14 quantitative regression models and 16 predictive classification models for predicting  |
| ADMETSAR 2.0  | Free access to web server  | ADMET Prediction | 47 models available for drug discovery and environmental risk assessment  |
| cypREACT | open access tools  | metabolism prediction  | A set of in silico metabolic prediction tools for precisely forecasting the production of human metabolism  |
| Pred-Skin 3.0  | Free access to web server  | Skin sensitization prediction  | A QSAR model base on Skin sensitization prediction tool  |
| autoQSAR  | commercial  | QSAR modeling  | QSAR modeling will generated by build in featurization for molecular systems  |
| Double cross validation  | free download tool  | QSAR modeling  | The tool involved internal external cross validation loops for modeling validation  |

1. **Database for modeling and prediction of ADRs and toxicity**

Various databases, such as the FDA's Adverse Event Reporting System (FAERS), the SIDER database, and ToxNet, offer valuable information for predicting adverse drug reactions (ADRs) and drug toxicity, enhancing pharmacovigilance efforts. FAERS contains over 14 million reports of adverse events, while SIDER provides data on marketed medicines and their recorded adverse reactions. ToxNet, managed by the US National Library of Medicine, offers toxicology data for predicting drug toxicity. Integration of these databases into AI and ML models allows for rapid and accurate analysis of vast data sets, aiding in the early detection of potential ADRs and toxicities during drug development. Continual updates to these databases enable real-time monitoring and prediction, supporting the development of safer and more effective therapeutics.[85,86,87]. For a detailed overview of these databases and others, including their primary features and data types, please refer to Table 3.

Table 3. Detailed overview of selected databases

|  |  |  |
| --- | --- | --- |
| **Database** Biochemical Genetic and Genomic  | Data typeMetabolomics database  | **Description** Using the Constraint Based Reconstruction and Analysis (COBRA) framework, Biochemical Genetic and Genomic is a knowledge base of extensive genome-scale metabolic reconstructions that are biochemically, genetically, and gnomically structured. These reconstructions help assess the metabolic capabilities of organisms and for interpreting experimental data. |
| BioCyc | Metabolomic database  | A collection of 20,055 Pathway/Genome Databases (PGDBs) for model eukaryotes and thousands of microorganisms, BioCyc includes curated data from 130,000 literatures. |
| Human Metabolome database  | metabolomics database | The Human Metabolome Database is a database that contains data on small-molecule metabolites that can be discovered in the human body and that can be used in clinical chemistry, metabolomics, and the development of biomarkers. |
| Kyoto Encyclopedia of Gene and Genomes  | Metabolomics database | The Kyoto Encyclopedia of Genes and Genomes is an excellent source for working on the high-level functions and applications of biological systems, as well as the genome, diseases, biological pathways, medications, and chemicals. |
| Small molecules pathway database  | Metabolomic database  | SMPDB is an interactive and visually appealing database with more than 30.000 small molecule pathways that are exclusive to humans. |
| Pubchem | Toxicity database | PubChem is a database covering chemicals and drugs, diseases and the environment, environmental health, occupational safety and health, poisoning, risk assessment and regulations, and toxicology, among others. |
| Genetic Toxicology Data Bank  | Toxicology Database | For more than 3000 chemicals from the United States Environmental Protection Agency (EPA), GENE-TOX from the National Library of Medicine at the NIH provides genetic toxicology test data from expert peer evaluation of available scientific literature. |
| CompTop chemical Dashboard  | Toxicology database  | The CompTox Chemicals Dashboard is a freely accessible online database provided by the U.S. Environmental Protection Agency (EPA). It provides information on over 700,000 chemicals, including toxicity data. |
| TOXLINE ( Toxicology Information Online) | Toxicology database  | The comprehensive online bibliographic database TOXLINE is maintained by the National Library of Medicine and provides information on the pharmacological, biochemical, physiological, and toxicological effects of medications and other compounds. |
| ChemIDplus | Toxicology database  | This is a database provided by the U.S. National Library of Medicine, containing over 400,000 chemicals. Toxicity information can be found on each chemical's page, often with links to more detailed resources. |
| Hazardous Substances Data Bank | Toxicology database  | A toxicology database with a focus on the toxicity of potentially dangerous substances is called the Hazardous Substances Data Bank (HSDB). It offers details on human exposure, occupational hygiene, emergency response techniques, environmental fate, legal specifications, nanomaterials, and related topics. |
| Integrated Risk Information Systems  | Toxicology database  | IRIS, which is managed by the EPA, houses data on numerous compounds' effects and health risk information. It focuses on the potential negative consequences on human health brought on by exposure to various environmental toxins. |
| LiverTox | Toxicity database  | A free access web-based database providing liver-related injury including medication, nutritional supplements, and herbal chemicals. |

1. **Role of AI and ML in ADR and toxicity modeling**

**10.1 Drug- Induced Liver Injury**

Jaganathan et al. utilized a Support Vector Machine (SVM) to predict Drug-induced liver injury (DILI) using 1253 compounds gathered from various sources. Molecular descriptors from different tools were integrated into a dataset containing 2648 descriptors. After cleaning and applying learning algorithms, the SVM model yielded accuracy, sensitivity, and specificity scores of 0.811, 0.840, and 0.783 respectively, indicating good classification accuracy with fewer descriptors required. In 2023, Rao et al. developed another DILI AI/ML model, reducing drug discovery time. They employed six machine learning models including SVM, achieving notable scores for receiver operating characteristic area under the curve, sensitivity, and specificity. Additionally, they identified key off-target genes for further model development.[18,19]

**10.2 Nephrotoxicity and kidney Injury**

Kandasamy et al. showcased the applicability of machine learning in predicting drug-induced nephrotoxicity, focusing on the renal proximal tubule as the target. Using a dataset of 30 compounds derived from experimental data, they employed the Random Forest algorithm for model development. Through 10-fold cross-validation, they achieved training accuracy of 99.8% and test accuracy of 87.0%.

Acute kidney injury (AKI) poses significant healthcare challenges and risks, including increased expenditures, mortality, cardiovascular events, and progression to chronic kidney disease. Cheng et al. developed a machine learning model for AKI using electronic medical records from 60,534 patients. Their random forest prediction model, validated through cross-validation, achieved an impressive area under the curve of 0.765, predicting AKI events a day in advance.[50,51,52]

**10.3 Cardiotoxicity**

Kar and Roy conducted a significant study on hERG prediction using AI and ML, focusing on identifying potential hERG channel blockers crucial for predicting cardiotoxicity. They developed QSAR-based models using a diverse set of 242 compounds with varying structures and therapeutic actions. These models provided quantitative insights into the molecular structures and properties influencing hERG binding affinity. By adhering to OECD principles, they rigorously validated the models internally and externally, and introduced Pharmacological Distribution Diagrams (PDDs) for better classification visualization. The study concluded with the application of these models to screen compounds from the DrugBank database for hERG channel-blocking properties.

In another work by Cai et al., a deep learning algorithm named deephERG was proposed in 2019 for predicting small molecule hERG blockers. Using a dataset of 7889 compounds and a deep neural network (DNN), they achieved an impressive AUROC of 0.967 on validation sets. DeephERG identified 29.6% of FDA-approved drugs as potentially having hERG inhibitory activities, underscoring the importance of hERG risk assessment in early drug discovery.[53,54]

**10.4 Neurotoxicity**

Zhao et al. delved into drug-induced neurotoxicity (DINeurot) by leveraging ML-based models, compiling a substantial dataset of neurotoxic and non-neurotoxic pharmacological structures. Their study led to the development of the top-performing MACCS\_SVM model, exhibiting remarkable prediction accuracy rates, potentially offering a cost-effective alternative to traditional experimental methods for identifying DINeurot. The findings also highlighted significant differences in physicochemical characteristics between neurotoxic and non-neurotoxic substances, unveiling potential indicators of neurotoxicity. The identification of 18 structural alerts (SAs) associated with neurotoxicity enhances DINeurot prediction and furthers understanding of chemical neurotoxicity mechanisms. These SAs have been incorporated into the previously developed SApredictor web server, enhancing its utility for toxicity prediction.

Jiang et al. developed a QSAR model for neurotoxicity utilizing LD50 data of 442 organic compounds and molecular descriptors computed through PyBioMed software. They employed various algorithms to build QSAR models and reported a Q2 value of 0.784 for predictive effect. MATSe3, MATSv3, and Smin32 were identified as key descriptors for neurotoxicity prediction. Based on accuracy assessments, the authors recommended the extra-trees regressor model for further modeling and predictions.[55,46]

**10.5 Carcinogenicity, genotoxicity, and mutagenicity**

Kar and Roy devised a robust QSAR model to establish correlations between chemical structure and specific activities or properties, using a dataset of 1464 compounds. Their aim was to reduce reliance on expensive and time-consuming animal testing. Their findings identified molecular features such as lipophilicity, conjugated ring systems, keto, and nitro groups as potential contributors to carcinogenicity, while tertiary and secondary nitrogens, phenolic, enolic, and carboxylic OH fragments, and the presence of three-membered rings were associated with decreased carcinogenic potential. This work highlights the promise of AI and ML in toxicological modeling, facilitating efficient in-silico screening of drug candidate molecules for their carcinogenic potential.

Li et al. introduced DeepCarc, utilizing the National Center for Toxicological Research liver cancer database (NCTRlcdb) with 692 compounds in the training set and 171 in the test set for its development. DeepCarc achieved a Matthews correlation coefficient (MCC) of 0.432 for the test set, surpassing four sophisticated deep learning (DL)-powered QSAR models by an average of 37%. Furthermore, DeepCarc assessed the carcinogenic potential of compounds from DrugBank and Tox21, positioning itself as an early detection tool for carcinogenicity prediction.[47,48]

**10.6 Skin sensitization**

Kar and Roy developed multiple QSAR models to forecast the skin sensitization potential of 51 organic compounds, uncovering insights into the chemical characteristics governing skin sensitivity. They found aromatic compounds generally have higher sensitizing potential than aliphatic ones, except for carbonyl compounds where aliphatic compounds take precedence. Electrophilic and hydrophilic properties emerged as crucial in determining sensitization propensity, along with features like nitrogen atoms, triple bonds, and the Al-C(=X)-Al fragment.

Nandy et al. employed Linear Discriminant Analysis (LDA) to predict skin sensitization potential, aiming to modernize toxicological modeling and circumvent costly traditional assays. Their LDA model, built and validated with 147 chemicals, demonstrated exceptional ability in distinguishing skin-sensitizing from non-skin-sensitizing substances. Key findings revealed that rotatable bonds and molecular flexibility negatively impact skin sensitivity, while the Dragon branching index has a positive effect. Descriptors like the quadric index and nitrogen atoms showed comparable contributions to both categories. Applying the model to the DrugBank database showcased its potential in identifying potential skin-sensitizing substances, advancing AI/ML's role in reliable toxicological profiling.[39,50]

1. **Future direction and overview**

The advent of AI and ML represents a transformative phase in pharmacovigilance, offering promising prospects for early detection of adverse drug reactions (ADRs) and drug-induced toxicities. These advanced technologies have already revolutionized how we analyze and mitigate toxic effects and adverse medication reactions. Real-time monitoring and reporting hold potential for further exploration, as AI/ML algorithms could swiftly identify probable ADRs, facilitating prompt treatments and reducing harm to patients, especially with the increasing adoption of remote patient monitoring technologies. Despite advancements, there's a pressing need for more accurate and reliable models capable of handling complex datasets, with deep learning offering potential for enhanced prediction power by discerning intricate patterns in high-dimensional data. Addressing the opacity of 'black box' models in healthcare is essential to foster trust among healthcare professionals, necessitating the development of clearer and more understandable AI models. Integrating AI and ML with electronic health records presents an intriguing opportunity to efficiently sift through vast amounts of unstructured data to identify potential ADRs, significantly bolstering patient safety and pharmacovigilance efforts. Furthermore, the convergence of AI and ML with genomic data in personalized medicine holds promise for tailored ADR predictions, enabling safer and more effective treatments tailored to individual genetic profiles. Though still nascent, the application of AI and ML in early ADR and toxicity detection holds immense potential to enhance drug safety, reduce healthcare costs, and save lives, shaping the future landscape of pharmacovigilance.

In summary, while still in its early stages, the field of AI and ML in early ADR and toxicity detection holds great promise for the future. The progress made so far is encouraging, and further advancements have the potential to enhance drug safety, reduce healthcare expenses, and ultimately save lives. Moving forward, strategic development and utilization of these technologies can truly revolutionize pharmacovigilance and drug safety practices.

**COMPETING INTERESTS DISCLAIMER:**

Authors have declared that they have no known competing financial interests OR non-financial interests OR personal relationships that could have appeared to influence the work reported in this paper.

REFERENCE

1. J.A. DiMasi, H.G. Grabowski, R.W. Hansen Innovation in the pharmaceutical industry: new estimates of R&D costs J. Health Econ., 47 (2016), pp. 20-33
2. R.K. Harrison Phase II and phase III failures: 2013–2015 Nat. Rev. Drug Discov., 15 (2016), pp. 817-818
3. World Health Organization Quality Assurance and Safety of Medicines Team, Safety of medicines: a guide to detecting and reporting adverse drug reactions: why health professionals need to take action., (2002).
4. C. Kongkaew, P.R. Noyce, D.M. Ashcroft Hospital admissions associated with adverse drug reactions: a systematic review of prospective observational studies Ann. Pharmacother., 42 (2008), pp. 1017-1025
5. E.S. Björnsson Drug-induced liver injury: an overview over the most critical compounds

 Arch. Toxicol., 89 (2015), pp. 327-334

1. J. Sultana, P. Cutroneo, G. Trifirò Clinical and economic burden of adverse drug reactions J. Pharm. Pharm., 4 (2013), pp. S73-S77
2. T. Hartung Toxicology for the twenty-first century Nature, 460 (2009), pp. 208-212
3. S. Kar, K. Roy QSAR of phytochemicals for the design of better drugs Expert Opin. Drug Discov., 7 (2012), pp. 877-902
4. P. Zhang, F. Wang, J. Hu, R. Sorrentino Towards personalized medicine: leveraging patient similarity and drug similarity analytics AMIA Jt Summits Transl. Sci. Proc., 2014 (2014), pp. 132-136
5. E.J. Topol High-performance medicine: the convergence of human and artificial intelligence Nat. Med., 25 (2019), pp. 44-56 2023 June 13
6. P.B. Jensen, L.J. Jensen, S. Brunak Mining electronic health records: towards better research applications and clinical careNat. Rev. Genet., 13 (2012), pp. 395-405
7. M. Liu, Y. Wu, Y. Chen, J. Sun, Z. Zhao, X.-W. Chen, M.E. Matheny, H. Xu Large-scale prediction of adverse drug reactions using chemical, biological, and phenotypic properties of drugs J. Am. Med Inf. Assoc., 19 (2012), pp. e28-e35
8. D.W. Bates, R.S. Evans, H. Murff, P.D. Stetson, L. Pizziferri, G. Hripcsak Detecting adverse events using information technology J. Am. Med Inf. Assoc., 10 (2003), pp. 115-128
9. A. Cherkasov, E.N. Muratov, D. Fourches, A. Varnek, I.I. Baskin, M. Cronin, J. Dearden, P. Gramatica, Y.C. Martin, R. Todeschini, V. Consonni, V.E. Kuz’min, R. Cramer, R. Benigni, C. Yang, J. Rathman, L. Terfloth, J. Gasteiger, A. Richard, A. Tropsha QSAR modeling: where have you been? Where are you going to? J. Med. Chem., 57 (2014), pp. 4977-5010
10. A. Mayr, G. Klambauer, T. Unterthiner, S. Hochreiter DeepTox: toxicity prediction using deep learning Front. Environ. Sci., 3 (2016)16
11. J. Vamathevan, D. Clark, P. Czodrowski, I. Dunham, E. Ferran, G. Lee, B. Li, A. Madabhushi, P. Shah, M. Spitzer, S. Zhao Applications of machine learning in drug discovery and development Nat. Rev. Drug Discov., 18 (2019), pp. 463-477 View at publisherCrossrefView in ScopusGoogle Scholar
12. 17J.H. Kim, A.R. Scialli Thalidomide: the tragedy of birth defects and the effective treatment of disease Toxicol. Sci., 122 (2011), pp. 1-6 Google Scholar
13. W. Lenz A short history of thalidomide embryopathy Teratology, 38 (1988), pp. 203-215 View at publisherCrossrefView in ScopusGoogle Scholar
14. M.T. Miller, K. Strömland Teratogen update: thalidomide: a review, with a focus on ocular findings and new potential uses Teratology, 60 (1999), pp. 306-321 View in ScopusGoogle Scholar
15. F.O. Kelsey Thalidomide update: regulatory aspects Teratology, 38 (1988), pp. 221-226 View at publisherCrossrefView in ScopusGoogle Scholar
16. H.M. Krumholz, J.S. Ross, A.H. Presler, D.S. Egilman What have we learnt from Vioxx? Bmj, 334 (2007), pp. 120-123 View at publisherCrossrefView in ScopusGoogle Scholar22
17. B. Sibbald Rofecoxib (Vioxx) voluntarily withdrawn from market Cmaj, 171 (2004), pp. 1027-1028 View in Scopus Google Scholar
18. FDA approval of lymphoma medicine Ukoniq (umbralisib) is withdrawn due to safety concerns; [cited 2023 June 13]. Available from: 〈https://www.fda.gov/drugs/drug-safety-and-availability/fda-approval-lymphoma-medicine-ukoniq-umbralisib-withdrawn-due-safety-concerns〉.Google Scholar
19. FDA granted accelerated approval to belantamab mafodotin-blmf for multiple myeloma; [cited 2023 June 13].Available from: 〈https://www.fda.gov/drugs/resources-information-approved-drugs/fda-granted-accelerated-approval-belantamab-mafodotin-blmf-multiple-myeloma〉,Google Scholar
20. FDA approves atezolizumab for PD-L1 positive unresectable locally advanced or metastatic triple-negative breast cancer; [cited 2023 June 13]. Available from: 〈https://www.fda.gov/drugs/drug-approvals-and-databases/fda-approves-atezolizumab-pd-l1-positive-unresectable-locally-advanced-or-metastatic-triple-negative〉,Google Scholar
21. Analysis and recommendations for Agency action regarding non- steroidal anti-inflammatory drugs and cardiovascular risk; [cited 2023 June 13]. Available from: 〈https://www.fda.gov/media/74279/download〉,Google Scholar
22. FDA requests the withdrawal of the weight-loss drug Belviq, Belviq XR (lorcaserin) from the market; [cited 2023 June 13]. Available from: 〈https://www.fda.gov/drugs/fda-drug-safety-podcasts/fda-requests-withdrawal-weight-loss-drug-belviq-belviq-xr-lorcaserin-market#:∼:text=On%20February%2013%2C%202020%20FDA,an%20increased%20occurrence%20of%20cancer〉.,Google Scholar
23. D. Halegoua-De Marzio, V.J. Navarro Chapter 29 - hepatotoxicity of cardiovascular and antidiabetic drugs N. Kaplowitz, L.D. DeLeve (Eds.), Drug-Induced Liver Disease (Third Edition), Academic Press (2013), pp. 519-540, 10.1016/B978-0-12-387817-5.00029-7 View PDFView article View in Scopu sGoogle Scholar
24. H.P. Parkman Chapter 24 - Prokinetic agents for gastroparesis R.W. McCallum, H.P. Parkman (Eds.), Gastroparesis, Academic Press (2021), pp. 323-339, 10.1016/B978-0-12-818586-5.00024-7 View PDF View articleGoogle Scholar
25. S. Czernichow, G.D. Batty Withdrawal of sibutramine for weight loss: where does this leave clinicians? Obes. Facts, 3 (2010), pp. 155-156 View in ScopusGoogle Scholar31
26. Withdrawal of sibutramine for weight loss: where does this leave clinicians? Obes. Facts, 3 (2010), pp. 155-15632
27. S. Gottlieb Antihistamine drug withdrawn by manufacturer Bmj, 319 (1999), p. 7
28. P.C. Hébert, D.A. Fergusson, B. Hutton, C.D. Mazer, S. Fremes, M.Blajchman, C. MacAdams, G. Wells, J. Robblee, J. Bussières, K. Teoh

Regulatory decisions pertaining to aprotinin may be putting patients at risk Cmaj, 186 (2014), pp. 1379-1386

1. W.J. Wight, P.M. Wright Pharmacokinetics and pharmacodynamics of rapacuronium bromide Clin. Pharm., 41 (2002), pp. 1059-1076
2. M.B. Zazzara, K. Palmer, D.L. Vetrano, A. Carfì, G. Onder Adverse drug reactions in older adults: a narrative review of the literature Eur. Geriatr. Med., 12 (2021), pp. 463-473
3. I. Kola, J. Landis Can the pharmaceutical industry reduce attrition rates? Nat. Rev. Drug Discov., 3 (2004), pp. 711-716
4. O. World Health. (2002). The importance of pharmacovigilance. In. Geneva: World Health Organization.
5. G. Patrick Getting the drug to market In An Introduction to Medicinal Chemistry, Oxford University Press,(2017), pp. 274-305
6. R.S. Obach, J.G. Baxter, T.E. Liston, B.M. Silber, B.C. Jones, F. MacIntyre, D.J. Rance, P. Wastall The prediction of human pharmacokinetic parameters from preclinical and in vitro metabolism data J. Pharm. Exp. Ther., 283 (1997), pp. 46-58
7. Z. Zhang, W. Tang Drug metabolism in drug discovery and development Acta Pharm. Sin. B, 8 (2018), pp. 721-732
8. J. Brian Houston, D.J. Carlile Prediction of hepatic clearance from microsomes, hepatocytes, and liver slices Drug Metab.Rev., 29 (1997), pp. 891-922
9. J.C. Kalvass, S.M. Tristan, M.P. Gary Use of plasma and brain unbound fractions to assess the extent of brain distribution of 34 drugs: comparison of unbound concentration ratios to in vivo p-glycoprotein efflux ratios Drug Metab. Dispos., 35 (2007), p. 660
10. M. Rowland, & T.N. Tozer. (2010). Clinical Pharmacokinetics and Pharmacodynamics: Concepts and Applications Fourth Edition. LWW.
11. Y. Qiu, Y. Chen, G. Zhang, L. Yu, & R.V. Mantri. (2016). Developing Solid Oral Dosage Forms-Pharmaceutical Theory and Practice, 2nd Edition. Elsevier.
12. R. Collier Rapidly rising clinical trial costs worry researchers Cmaj, 180 (2009), pp. 277-278
13. U.S.Fa.D. Administration Drug Dev. Process (2018)
14. S.-M. Huang, J. Lertora, P. Vicini, & J. Arthur Atkinson (2021). Atkinson's Principles of Clinical Pharmacology, 4th Edition. Elsevier.

48. K.I. Kaitin, J.A. DiMasi Pharmaceutical innovation in the 21st century: new drug approvals in the first decade, 2000-2009 Clin. Pharm. Ther., 89 (2011)

 49.E.L. Eisenstein, R. Collins, B.S. Cracknell, O. Podesta, E.D. Reid, P. Sandercock, Y. Shakhov, M.L. Terrin, M.A. Sellers, R.M. Califf, C.B. Granger, R. Diaz

Sensible approaches for reducing clinical trial costs Clin. Trials, 5 (2008), pp. 75-84

50.C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. FeeneyExperimental and computational approaches to estimate solubility and permeability in drug discovery and development settings Adv. Drug Deliv. Rev., 46 (2001), pp. 3-26

51.A. Gabizon, R. Catane, B. Uziely, B. Kaufman, T. Safra, R. Cohen, F. Martin, A. Huang, Y. Barenholz Prolonged circulation time and enhanced accumulation in malignant exudates of doxorubicin encapsulated in polyethylene-glycol coated liposomes Cancer Res., 54 (1994), pp. 987-992

52.Braeuning, R.A. Budinsky, B. Burkhardt, N.R. Cameron, G. Camussi, C.-S. Cho, Y.-J. Choi, J. Craig Rowlands, U. Dahmen, J.G. Hengstler Recent advances in 2D and 3D in vitro systems using primary hepatocytes, alternative hepatocyte sources and non-parenchymal liver cells and their use in investigating mechanisms of hepatotoxicity, cell signaling and ADME Arch. Toxicol., 87 (2013), pp. 1315-1530

53.W.J. Pichler Delayed drug hypersensitivity reactions Ann. Intern Med, 139 (2003), pp. 683-693

54.E. Tiligada, M. Ishii, C. Riccardi, M. Spedding, H.U. Simon, M.M. Teixeira, M.L. Cuervo, S.T. Holgate, F. Levi-Schaffer The expanding role of immunopharmacology: IUPHAR Review 16 Br. J. Pharm., 172 (2015), pp. 4217-4227

55. L.B. Schwartz Effector cells of anaphylaxis: mast cells and basophils Novartis Found. Symp., 257 (2004), pp. 65-74 discussion 74-69, 98-100, 276-185