J. Fac. Pharm. Ankara / Ankara Ecz. Fak. Derg., 50(2): 457-474, 2026
Doi: 10.33483/jfpau.1655343
Review Article/ Derleme Makale

MACHINE LEARNING APPLICATIONS IN TOXICOLOGY: CURRENT APPROACHES AND FUTURE PERSPECTIVES

TOKSİKOLOJİDE MAKİNE ÖĞRENMESİ UYGULAMALARI: GÜNCEL YAKLAŞIMLAR VE GELECEK PERSPEKTİFLERİ

Selinay YAŞAR[1], Rumeysa ÇETİN TÜRKER[1,2], Merve BECİT KIZILKAYA[3]*

[1]Afyonkarahisar Health Sciences University, Faculty of Pharmacy, Department of Pharmaceutical Toxicology, Afyonkarahisar, Türkiye
[2]Gazi University, Faculty of Pharmacy, Department of Pharmaceutical Toxicology, Ankara, Türkiye
[3]University of Health Sciences, Gülhane Faculty of Pharmacy, Department of Pharmaceutical Toxicology, Ankara, Türkiye

ABSTRACT

Objective: *Toxicology faces challenges with increasing chemicals and complex exposure scenarios. This review examines the historical development and current status of artificial intelligence (AI) and machine learning (ML) applications in toxicology, from QSAR models of the 1960s to today's deep learning algorithms. Advances achieved through AI integration in fields such as drug discovery, toxicokinetics, nanotoxicology, and environmental toxicology, along with toxicological endpoints including cardiovascular toxicity, hepatotoxicity, carcinogenesis, genotoxicity, and neurotoxicity, have been evaluated.*

Result and Discussion: *AI technologies offer significant advantages such as reduction in animal experimentation, rapid pattern detection, and toxicity prediction with minimal experimental data. Integrated analysis of genomic, proteomic, and chemical structure data elucidates toxicity mechanisms at the molecular level, while ML models developed for multiple organs contribute to reducing drug attrition rates. Data quality, model validation, and interpretability remain primary challenges to be overcome. In the future, integration of traditional toxicological methods with modern computational approaches will provide more reliable and efficient results in risk assessments, while regulatory authorities need to develop standards for AI-supported models. Multi-omic data integration, personalized toxicology approaches, and development of new ML algorithms that can illustrate toxicity mechanisms in humans represent promising directions for advancing the toxicology sciences.*

Keywords: *AOP, machine learning, PBTK, QSAR, toxicology*

ÖZ

Amaç: *Toksikoloji, artan kimyasallar ve karmaşık maruziyet senaryolarıyla zorluklarla karşı karşıyadır. Bu derleme, 1960'lardaki QSAR modellerinden günümüzün derin öğrenme algoritmalarına kadar toksikolojide yapay zeka (AI) ve makine öğrenmesi (ML) uygulamalarının tarihsel gelişimini ve mevcut durumunu incelemektedir. İlaç keşfi, toksikokinetik, nanotoksikoloji ve çevresel toksikoloji gibi alanlarda AI entegrasyonu ile kardiyovasküler toksisite, hepatotoksisite,*

* **Corresponding author:** Merve Becit Kızılkaya
E-mail: mervebecit@hotmail.com; merve.becit@afsu.edu.tr
Submitted date: 11.03.2025 **Accepted date:** 02.03.2026 **Published date:** 19.05.2026
Citation: Yaşar, S., Çetin-Türker, R., Becit Kızılkaya, M. (2026). Machine learning applications in toxicology: Current approaches and future perspectives. Journal of Faculty of Pharmacy of Ankara University, 50(2), 457-474.

karsinogenez, genotoksisite ve nörotoksisite gibi toksikolojik sonlanım noktalarında sağlanan ilerlemeler değerlendirilmiştir.

Sonuç ve Tartışma: *AI teknolojileri, hayvan deneylerinin azaltılması, hızlı örüntü tespiti ve minimum deneysel veriyle toksisite tahmini gibi önemli avantajlar sunmaktadır. Genomik, proteomik ve kimyasal yapı verilerinin entegre analizi, moleküler düzeyde toksisite mekanizmalarını aydınlatırken, çoklu organlar için geliştirilen ML modelleri ilaç kayıp oranlarının azaltılmasına katkıda bulunmaktadır. Veri kalitesi, model validasyonu ve yorumlanabilirlik, aşılması gereken başlıca zorluklar olarak kalmaktadır. Gelecekte, geleneksel toksikolojik yöntemlerin modern hesaplamalı yaklaşımlarla entegrasyonu, risk değerlendirmelerinde daha güvenilir ve verimli sonuçlar sağlarken, düzenleyici otoritelerin AI destekli modeller için standartlar geliştirmesi gerekmektedir. Multi-omik veri entegrasyonu, kişiselleştirilmiş toksikoloji yaklaşımları ve insanlarda toksisite mekanizmalarını gösterebilen yeni ML algoritmalarının geliştirilmesi, toksikoloji bilimlerinin ilerlemesi için umut verici yönleri temsil etmektedir.*

Anahtar Kelimeler: *AOP, makine öğrenmesi, PBTK, QSAR, toksikoloji*

INTRODUCTION

Toxicology is a scientific discipline that studies the adverse effects of xenobiotics on organisms and the environment, as well as the mechanisms that cause this toxicity. Toxicology also encompasses the application of toxicological knowledge for the mitigation and/or prevention of this toxicity, and for the risk assessment of xenobiotics [1]. Toxicology faces thousands of new chemical substances introduced to the market each year, increasing regulatory requirements, and complex exposure scenarios. According to Chemical Abstracts Service records, there are currently more than 200 million chemical substances, and this number is increasing every day. Chemical risk analyses and safety tests are critically important for the timely identification of toxic chemicals in different industrial fields. In drug development, for example, toxicity assessments conducted in the early stages of the development process significantly reduce the financial burden of new drug development by lowering failure rates that might be encountered in later phases [2]. Traditional toxicology testing methods are generally time-consuming, costly, and largely dependent on animal models. This creates significant limitations, including ethical issues and the inability to precisely represent human biology. The time-consuming and costly nature of traditional toxicity tests has encouraged the development of artificial intelligence (AI)-based solutions as an alternative strategy [3,4].

The foundations of AI applications in toxicology were laid in the 1960s with the development of Quantitative Structure-Activity Relationship (QSAR) models. Early QSAR models were based on simple mathematical relationships. With advancements in computer technologies, more complex models began to emerge in the 1980s. In the 1990s, the application of artificial neural networks in toxicology significantly increased the accuracy of prediction models. In the early 2000s, toxicity prediction models became more sophisticated with the development of data mining techniques and machine learning (ML) algorithms [5]. Since the 2010s, the proliferation of deep learning (DL) technologies, coupled with the development of new experimental protocols such as high-throughput screening (HTS) analyses, has generated toxicity data for millions of compounds. This large accumulation of data enables faster and more accurate toxicity predictions. Some DL models provide more successful results than traditional methods [2,6].

ML and DL technologies can significantly reduce the use of test animals by offering alternative methods to animal experiments. These technologies enable rapid pattern detection in complex toxicological data sets, prediction of toxic effects with minimal experimental data, and estimation of multiple organ toxicity using data obtained from a single *in-vitro* experiment. Integrated analysis of genomic, proteomic, and chemical structure data becomes possible, thereby elucidating toxicity mechanisms at the molecular level. These technologies play a complementary role in determining the toxicokinetic properties of xenobiotics, establishing safety profiles, and conducting risk assessments [4,7-9].

In recent years, the application areas of ML technologies in toxicology have been continuously expanding. Successful applications are seen in various fields such as drug discovery and development [10], toxicokinetics [11], oncological drug efficacy and toxicity [12], nanotoxicology [13], human health

hazard assessment of industrial chemicals [14], and histopathology [15]. Significant developments have also been recorded in specific areas among toxicology's comprehensive endpoints, including hepatotoxicity [16], cardiotoxicity [17], carcinogenesis [18], genotoxicity [19], cytotoxicity [20], dermal toxicity [21]. *In silico* drug toxicity prediction offers significant advantages, particularly in terms of reducing drug attrition rates. Similar to organ-on-a-chip technologies, ML prediction models for multiple organs are being successfully implemented [3,22].

The most critical issue in AI-based toxicology models is the "black box" nature of DL models, particularly those with neural network structures. This situation makes it difficult for toxicologists to trust these predictions, as it leads to a lack of understanding of how the models make decisions. As a solution to this problem, interpretable ML approaches are being developed. These approaches enable explanation of how models make decisions and render toxicity mechanisms comprehensible. Reducing the lack of interpretability underlying a trained model is an effective approach to reveal underlying toxicity mechanisms and strengthen decision-making processes [2]. Although some technical challenges exist, such as data quality and model validation, the development of alternative testing methods in line with technological advancements and industrial needs presents significant opportunities.

The present review aims to comprehensively address the current state of ML applications in toxicology, which has gained increasing importance. Specifically, it focuses on the integration of traditional toxicological methods with modern computational approaches, aiming to examine the methodological developments, application areas, and technical challenges encountered in this field within a systematic framework. Additionally, it aims to discuss the potential contributions of this integration to the future of toxicological science and new research opportunities.

Machine Learning Frameworks Advancing Toxicological Science

AI is a field of science focused on developing computer systems capable of performing tasks that require human intelligence. ML is a technique for creating prediction models through learning from input feature data using computational algorithms (Figure 1) [2,23].

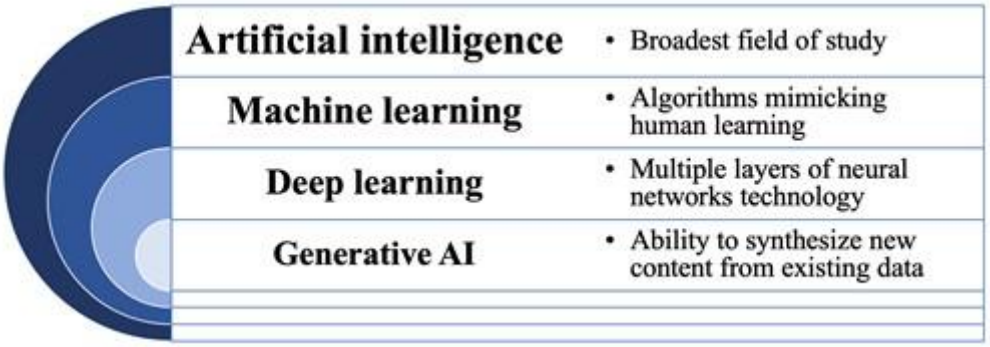

Figure 1. The hierarchical structure of artificial intelligence systems

ML is examined in three fundamental categories: supervised learning, unsupervised learning, and reinforcement learning. Supervised learning conducts evaluations using algorithms such as linear regression, logistic regression, and decision trees based on input and output data. Unsupervised learning groups data or extracts structural relationships using clustering and principal component analysis algorithms based solely on input data. Reinforcement learning, on the other hand, learns by interacting with an environment based on a reward-cost mechanism [23]. DL is a subfield of ML and learns more complex relationships using multi-layer neural networks (Figure 1). With this characteristic, it provides superiority over traditional ML methods [24].

In ML algorithms used in toxicology, regression models based on quantitative toxicity values and classification models using categorical toxicity values stand out [7]. Each of these algorithms has been adapted for specific applications in toxicology, serving different learning objectives and can be used for common purposes (Table 1). The fundamental difference lies in the variability of each algorithm's

prediction capability and accuracy rate. Therefore, selecting the most appropriate algorithm for specific toxicological problems plays a significant role in the success of the study [9,12].

Table 1. Machine learning algorithms commonly used in toxicological research

Algorithm	Brief Description
Support vector machines	Classification method that separates data points in multi-dimensional space using an optimal hyperplane (such as separating toxic and non-toxic chemicals)
Random forest	Prediction model that combines multiple decision trees through random sampling (such as predicting mutagens in test sets consisting of numerous compounds)
Ensemble modeling	Stronger and more balanced toxicity prediction through combination of different ML models (such as model designs incorporating combinations of support vector machines, k-nearest neighbors, and Naïve Bayes)
k-nearest neighbors	Toxicity prediction based on the properties of k nearest neighboring chemicals (such as cardiotoxicity prediction)
Naïve Bayes	Classification method based on probability theory, operating under the assumption of independence between variables
Multilayer perceptron	Artificial neural network model with a multi-layered structure that processes input data through intermediate layers to produce output (such as predicting toxicity from chemical structural properties)
Decision tree	Prediction model that progresses through branching with hierarchical decision rules (such as classification of carcinogens)
Bayesian	Statistical learning model based on probability theory that makes new predictions using prior data (such as probabilistic prediction of toxicity classes of chemicals)
Extreme gradient boosting	High-performance and optimized ensemble learning model based on gradient boosting principle (such as toxicity classification in large chemical datasets)
Adaptive boosting	Ensemble learning algorithm that creates a strong prediction model by sequentially combining weak classifiers (such as correcting misclassified toxic compounds through repetitive learning)
Convolutional neural network	Deep learning model that extracts features through layered convolution operations (such as toxicity prediction from molecular structure images)

Reorganized based on the conceptual framework in [25] and the mini-review by [7]. The algorithms are ranked considering the distribution percentage of toxicity prediction models for machine learning and deep learning algorithms [7]

Model Development Procedure

A typical prediction model development procedure includes data collection and organization, model creation, and model validation. Different types of data provide researchers with varying levels of understanding and interpretation. Data used in toxicity studies significantly differ from raw datasets; as the vast majority of these data represent specific properties and interactions of chemical substances, they enable scientific interpretation both during the modeling phase and after the model is completed [2].

The systematic approach in ML studies begins with problem definition. Subsequently, 1K-100K data points need to be collected. The collected data should be cleaned and transformed. In the next stage, model selection should be made and training should be conducted. Finally, testing and validation procedures should be performed (Figure 2) [9].

The model's performance is evaluated with various metrics such as confusion matrix and Receiver Operating Characteristic (ROC) curves [26]. Criteria such as accuracy, precision, and mean square error are used to determine model success [3]. In a successful ML model, it is important to avoid superficial analyses and focus on mechanistic information. The overfitting problem and ethical implications should be considered [9,27]. In toxicity modeling, it is of critical importance for the scientific value and reliability that the model not only makes accurate predictions but can also explain the underlying toxicological mechanisms behind these predictions.

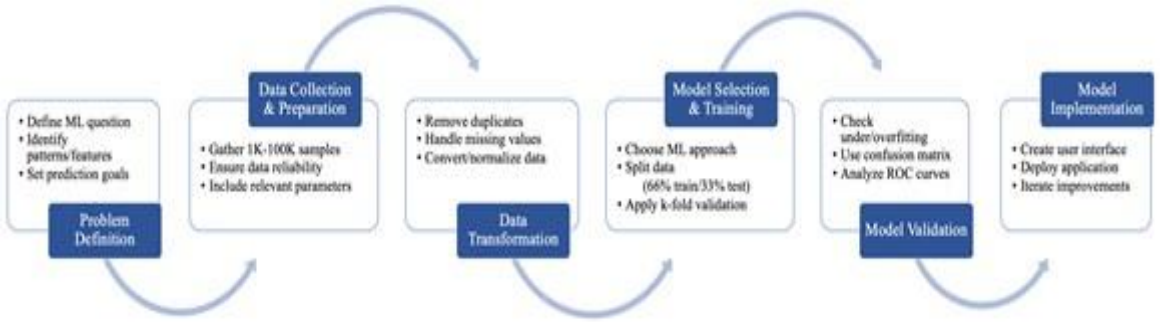

Figure 2. Machine learning workflow in toxicological applications
The systematic process starts with defining specific ML questions and prediction goals, followed by collecting reliable data samples (1K-100K). Data transformation ensures quality through duplicate removal and normalization. Model selection and training involves splitting data (66% training/33% testing) and k-fold validation. Model validation examines under/overfitting using confusion matrix and Receiver Operating Characteristic (ROC) curves. Finally, model implementation focuses on deployment and continuous improvement (Adapted from [9])

Artificial Intelligence-Based Toxicological Modeling Approaches

Quantitative Structure-Activity Relationship Models

The most intuitive data for toxicity modeling is chemical structure information. QSAR, traditionally used for chemical toxicity modeling, is a powerful computational method that analyzes the relationships between structural properties of chemicals and their biological activities (toxicokinetics and toxicity). Molecular structure information, which can be obtained from open access digital data repositories such as PubChem and ChEMBL, is used in generating descriptors required for toxicological modeling. These raw structural data can be comprehensively analyzed and transformed through advanced chemoinformatic software like RDKit, converting them into numerical parameters that can be used in toxicity prediction models [2].

Chemical structure similarity-based modeling methods (like QSAR and read-across approach) enable the prediction of biological activities of new chemicals without conducting experimental assay [25]. The QSAR model development process consists of four main stages [25,28]. First, a training set consisting of experimental data is prepared, then the properties of each molecule are encoded with molecular descriptors, subsequently the model is trained using mathematical algorithms, and finally the prediction performance of the model is evaluated using a validation data set. Some studies using QSAR modeling with ML and AI approaches [18,21,28-39] are summarized in Table 2. In 2014, the Tox21 Data Challenge initiated by the National Center for Advancing Translational Sciences (NCATS) to develop and compare different computational models for toxicity prediction based on chemical structure data, along with studies conducted on 11.764 chemicals and the development of the DeepTox system, presented new methodological approaches in toxicity prediction [39]. Another significant development is the toxicity modeling of polyfluorinated alkyl substances (PFAS), which are widespread environmental contaminants and pose a significant global public health concern. Using five different ML algorithms, this provides a valuable model for constraining the bioactivity of numerous PFAS based on chemical structures (Table 2) [28]. In recent years, the acute toxicity modeling study by the U.S. Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM), with a database covering 11.992 chemicals and 139 prediction models, has provided a robust alternative to *in-vivo* tests [40].

Moreover, QSAR models have shown significant developments in carcinogenicity tests. The DeepCarc model, developed to predict carcinogenicity for small molecules using DL-based model-level representations, achieved high accuracy (0.754) and sensitivity (0.910) values using a training set of 692 chemicals and a test set of 171 chemicals (Table 2) [18]. This model offers a rapid and effective screening tool as an alternative to traditional two-year rodent tests. However, the main challenge computational models face in carcinogenicity assessment is the limited applicability domains of individual models and restricted chemical diversity that can be evaluated.

Table 2. Representative studies integrating machine learning approaches with QSAR modeling

Best Algorithm	Data Size	Target Analysis	Reference
Support vector machines	650+ pesticides	Acute toxicity (bee)	[29]
Ensemble modeling	7000+ chemicals	Acute toxicity (rat, oral)	[30]
Meta Ensembling of Multitask, Deep Learning (QuantitativeTox)	1000+ chemicals	Acute toxicity (LD_{50} and LC_{50})	[31]
Multitask neural network	1000+ polyfluorinated alkyl substances	Bioassay	[28]
Ensemble XGBoost, Deep Learning (CapsCarcino, DeepCarc), Random Forest	600+ chemicals	Carcinogenicity	[18,32-34]
Random Forest	866,000+ chemicals	Multiple organ toxicity	[35]
Extra trees	1000+ chemicals	Various toxicities	[36]
Extra trees, Random Forest, k-nearest neighbors	18,000+ drug-gut microbiome interactions	Microbiome	[37]
Bernoulli naïve Bayes, k-nearest neighbors, Random Forest, Support vector machine	1200+ chemicals	Prenatal developmental toxicity	[38]
Deep Learning	31 chemicals	Skin toxicity	[21]

In [34], researchers developed an innovative weight of evidence model to overcome current challenges in carcinogenicity assessment. The model, integrating QSAR models and *in silico* toxicogenomic models, was applied to a training set of 597 chemicals and a test set of 198 chemicals obtained from the International Agency for Research on Cancer (IARC). Using the random forest algorithm, the model achieved an 8% improvement in Area Under the Receiver Operating Characteristic Curve (AUROC) value and a 19.7% improvement in chemical coverage. The model identified 44 chemicals with high carcinogenicity risk among approximately 1600 food contact chemicals, and 34 of these showed consistency with data in the IARC and European Chemicals Agency (ECHA) databases. These results demonstrate the model's effectiveness in identifying chemicals with carcinogenic potential and its ability to provide efficient alternatives.

QSAR models, also accepted by regulatory authorities, reduce the number of *in-vitro* and *in-vivo* toxicity tests and increase resource efficiency. These models, developed and validated in accordance with OECD principles, have become priority tools in chemical safety assessments. AI technologies facilitate regulatory acceptance processes by enhancing the transparency and interpretability of QSAR models [41].

However, all these structural modeling approaches have certain limitations. Especially in the assessment of complex toxicity endpoints such as carcinogenicity and hepatotoxicity, models based solely on structural information and chemical properties (including QSAR, read-across) may be inadequate. This becomes more evident in cases where compounds with similar chemical structures or properties exhibit different toxicity profiles, significantly reducing the prediction accuracy of the models [2]. Therefore, in evaluating complex toxicity endpoints, including other biological and toxicological data in addition to structural information is of critical importance.

Physiologically Based Toxicokinetic Models

ML-based physiologically based pharmacokinetic (PBTK) modeling and ADMET (absorption, distribution, metabolism, excretion, toxicity) predictions form the cornerstones of modern toxicology. These approaches demonstrate a synergistic effect in the comprehensive evaluation of safety profiles of chemical compounds [42]. PBTK modeling is a computational process that simulates the ADME (absorption, distribution, metabolism, excretion) processes of xenobiotics in the body using mathematical equations, involving experimental measurements of parameters such as tissue-plasma distribution coefficients and metabolic rates, and the prediction of parameters without experimental values using *in-vivo* pharmacokinetic data [43]. The LightGBM-based model developed by Kamiya et

al. [11] predicted PBTK parameters such as absorption rate, volume of distribution, and hepatic clearance of 246 different compounds with high accuracy (r=0.83). Similarly, Pradeep et al. [44] achieved significant results in predicting toxicokinetic parameters. These models play a critical role especially in dose-response analysis and *in-vitro* to *in-vivo* extrapolation (IVIVE), and in cross-species extrapolation and consequently in next-generation risk assessment that integrates next-generation approach methodologies [8,25,45]. Some studies from the last 5 years [46-49], summarized in Table 3, address the prediction of important *in-vivo* toxicity endpoints such as neurotoxicity, developmental toxicity, hepatotoxicity, and endocrine system toxicity based on *in-vitro* toxicity data.

Table 3. *In-vivo* toxicity prediction with PBTK model-based IVIVE

Model	Toxicity	Chemicals	*In-Vitro* Tests	Parameters	Target Species	Reference
Five-compartment PBTK	Neurotoxicity	Saxitoxin	Various *in-vitro* tests	BMDL10, BMD50, NOAEL, LD50	Rat, mouse, human	[46]
Four TK or PBTK models	Developmental toxicity	Valproic acid and 9 structural analogs	DevTOXquickPredict	LEL, clinical doses, EAD	Human, rat	[47]
PBTK model from GastroPlus 9.0 (Simulations Plus, Inc., Lancaster, CA, USA)	Hepatotoxicity	Acetaminophen	2D/3D HepaRG cell culture	LOEC, toxic dose	Human	[48]
Six-compartment PBTK model for maternal	Endocrine disrupting effect	Pesticides	Androgen receptor-transfection (OECD TG 458), H295R steroidogenesis assay	IC20, NOAEL, LOAEL	Rat	[49]

*Adapted from [8]

In recent years, the integration of PBTK and ML has shown new developments. The studies by Lu et al. [50] have revealed that the neural-ordinary differential equations approach, which modernizes traditional PBTK models, demonstrates superior performance especially in new dosing regimens. The development of user-friendly tools such as ADMET Predictor® provides time and cost savings for researchers and industry [9]. Parallel to these developments, the application area of AI-supported PBTK models has not been limited to conventional chemicals, but significant progress has also been made in nanotechnology. The superiority of DL approaches in predicting tumor delivery efficiency of nanoparticles has been reported in a meta-analysis study [51]. These integrated approaches enable faster, more efficient, and reliable results in drug discovery and toxicity assessment.

Adverse Outcome Pathway Analysis

Adverse Outcome Pathway (AOP) is a conceptual framework that links chemical-induced molecular, cellular, and organ-level responses to organism-level adverse outcomes. Mechanism-based test results can be used systematically within an AOP pathway to evaluate the likelihood of a compound triggering the target adverse outcome [2,52]. In the last 10 years, numerous AOPs have been characterized and documented in the AOP Knowledge Base [53]. Comprehensive studies such as ToxCast and Tox21 have evaluated thousands of chemicals using over 70 HTS and examined more than 125 biological processes. This program is conducted in collaboration with leading institutions such as the Environmental Protection Agency (EPA), Food and Drug Administration (FDA), NCATS, and National Toxicology Program (NTP), generating 120 million data points and making significant contributions to toxicology [25].

AI algorithms contribute to elucidating complex toxicity mechanisms by identifying interactions and intersections between different AOPs. This information is used in evaluating safety profiles of new

drug candidates and in risk assessments. Nuclear estrogen receptors (α and β) stand out as one of the most studied AOPs in toxicology, with these receptors playing critical roles in fundamental biological processes such as cell differentiation and reproduction [54,55]. As an alternative to the labor and resource-intensive nature of traditional methods, Ciallella et al [54] developed an innovative deep neural network model. This model includes 1024 functional connectivity chemical fingerprints and 3 estrogen receptor α/β toxicophores in the input layer, 5 hidden layers with 57 neurons arranged according to the AOP framework, and *in-vivo* rodent uterotrophic bioactivity in the output layer. Following a path from chemical structures to biological effects, this AI-based model can successfully predict compounds with potential estrogenic activity and is used as an effective tool in risk assessment of toxic compounds.

While traditional AOP approaches are useful in chemical risk assessment, they cannot fully reveal quantitative relationships [25]. Therefore, when sufficient data is available, quantitative AOPs (qAOPs) are being developed to mathematically model the relationships between chemical exposure and key events. Zgheib et al. [56] used three different approaches to quantitatively define the oxidative stress-induced chronic kidney disease AOP: empirical dose-response modeling, Bayesian network calibration, and systems biology modeling. The researchers demonstrated that the Bayesian network approach was more balanced and effective compared to other methods. Due to potential regulatory applications in chemical risk assessment, the number of qAOP models as computational toxicity prediction tools is gradually increasing [57,58].

The molecular initiating event has special importance among the key events of an AOP because it links the structural properties of chemicals to biological targets. Allen et al. [59,60] developed an effective tool that predicts molecular initiating events using the 2D structural properties of chemicals. This tool is based on the rich dataset in the ChEMBL database: more than one million compounds, over 12 million bioactivities, and more than 10.000 biological targets. The developed model contains 4810 structural alerts for 39 pharmacological targets and demonstrates impressive performance: achieving 82% sensitivity, 93% specificity, and 93% overall quality rates [60].

Data Science in Modern Toxicology

Big-Data

The concept of big data in toxicology refers to large-volume data sets that require high-performance computers and advanced computational approaches [61,62]. HTS data (e.g., ToxCast/Tox21 data), omics technologies data, toxicity data in large public databases (such as ACToR, ChEMBL, CTD, DrugMatrix, eNanoMapper, PubChem), epidemiological data, and environmental monitoring data are among the examples of big data [25]. For example, the ACToR database contains more than 800.000 compounds and over 500.000 tests, housing *in-vitro* and *in-vivo* toxicity data [63]. ChEMBL includes binding, function, and toxicity data of drugs and drug-like chemicals with 1.1 million bioassays and 1.8 million compounds [64]. While the CTD database houses relationships between more than 14.000 compounds, over 42.000 genes, and more than 6.000 diseases [65]. DrugMatrix contains gene expression data for approximately 600 drug molecules and 10.000 genes [66]. eNanoMapper houses physicochemical properties and safety data for more than 700 types of nanomaterials [67]. PubChem is one of the most comprehensive databases with over 111 million compounds, 1.39 million bioassays, and 293 million bioactivity data points [68]. The REACH database contains information on 21.405 unique substances from 89.905 dossiers submitted under European Union chemical legislation [69]. RepDose houses repeated dose study data where 364 compounds were examined in 1017 studies in dogs, mice, and rats [70]. While ToxicoDB contains toxicogenomic data for 231 chemicals [71], ToxNET (which was retired in 2019 and partially integrated into other NIH resources) houses *in-vitro* and *in-vivo* toxicity data for more than 50.000 environmental chemicals from 16 sources [72]. The existence of various big data sources in toxicology enables the application of AI approaches to predictive toxicology.

Moreover, Luechtefeld et al. [35] developed Read-Across Structure Activity Relationship (RASAR) models based on a combination of unsupervised k-nearest neighbor algorithm and logistic regression, as well as Data fusion RASAR models using similarity-based features and random forest, utilizing a database containing more than 866.000 chemical properties. The results of this study revealed

that big data analysis and ML-based QSAR/RASAR models demonstrated success exceeding the reproducibility rates of traditional animal experiments. In nanomaterials, Yan et al. [73] developed a comprehensive database containing 705 nanomaterials and 2142 nano-descriptors, while the eNanoMapper project [67,74] provides an extensive infrastructure based on FAIR (Findable, Accessible, Interoperable and Reusable) principles, including physicochemical and toxicological parameters. These developments are shaping the future of big data analytics and AI applications in toxicology, enabling more reliable, rapid, and ethical toxicity assessments.

High-Content Image-Based Screening Data

High-content image-based screening is one of the pioneering technologies for AI applications in toxicology. This system can perform comprehensive phenotypic analysis from microscopic images using convolutional neural networks and DL algorithms. The technology can detect toxic effects at the subcellular level with high-resolution imaging by simultaneously evaluating multiple parameters such as cell morphology, organelle structure, and protein localization. This approach is used as an effective tool especially in early-stage toxicity assessments in drug development processes [25].

The detection of DNA damage is particularly critical in toxicology, and high-content image-based methods are creating a major transformation in this area. A notable example in this context relates to the comet assay, one of the sensitive and reliable methods widely used in genotoxicity assessment. Despite being used for many years, the lack of automation constitutes a significant disadvantage. To solve this problem, Atila et al. [75] developed an innovative convolutional neural network model based on comet assay image data. Developed using a dataset containing 796 original and 9995 augmented images, the model consists of an input layer, an output layer, and 9 hidden layers. This model was able to classify comet images into four classes (healthy, slightly damaged, damaged, and severely damaged) with an accuracy rate of 96.1%. This study demonstrates the potential of AI applications in high-content imaging.

Omics/Toxicogenomics Data

Omics technologies (genomics, proteomics, metabolomics) are used to examine changes in biological systems in detail. For example, data was generated by combining proteomics technology with convolutional neural network methods to visualize histopathological images in kidney cancer [76]. In another study, genomics technology was combined with the Restricted Boltzmann Machine method for the diagnosis of ovarian cancer [77]. AI and omics technologies have critical importance in their contributions to toxicology science. They are used for purposes such as conducting big data analysis, discovering various biomarkers [78], and elucidating toxicological metabolic pathways [79].

Toxicogenomics, as a sub-branch of toxicology, is a field that uses genomic technologies to examine the adverse effects of xenobiotics at the gene and protein level. This field has become an important tool as a biomarker for predictive toxicology and for determining molecular toxicity mechanisms in organisms resulting from exposure to environmental chemicals [80,81]. In a recent study by Rahman et al. [82], *in-vitro* toxicogenomic data was collected regarding the expression of 38 key proteins exposed to 20 different chemicals. Optimal biomarkers were identified to predict phenotypic toxicity endpoints using ML-based feature selection and Support Vector Machine classification method. The researchers demonstrated that a small number of molecular biomarkers associated with DNA damage and repair pathways could predict *in-vivo* carcinogenicity and Ames genotoxicity endpoints with 70% accuracy.

Chen et al. [83] developed an AI-based Tox-GAN framework that can generate *in-vivo* gene activities and expression profiles in rats at multiple doses and treatment durations based on chemical structures. This model was trained using a large-scale toxicogenomic database containing transcriptomic data for 170 compounds [84]. Toxicogenomic data obtained from Tox-GAN showed more than 87% concordance with experimental gene expression data in terms of Gene Ontology. This framework offers a promising alternative tool for generating high-quality *in-vivo* toxicogenomic data without animal experiments.

Predictive Toxicity

The use of ML and DL algorithms in toxicity prediction has brought a new dimension to toxicology research by enabling the evaluation of different toxicity endpoints for numerous xenobiotics [85].

Hepatotoxicity

Drug-Induced Liver Injury (DILI) is one of the most unpredictable adverse reactions in clinical practice and is the main reason for the withdrawal of approved drugs after their market release. With symptoms ranging from changes in enzyme levels to liver necrosis, DILI seriously threatens patient safety. Ongoing inconsistencies between laboratory tests and real-world data have increased the importance of AI-based computational models in early prediction of DILI risk [41]. Significant progress has been made in DILI prediction using databases such as SIDER, TG-GATEs, and DrugBank [86]. Models in this field follow the pioneering triple classification system developed by Chen et al. [16], "no DILI", "less DILI", and "most DILI". Bayesian analysis, Support Vector Machine, ensemble modeling, random forest, multilayer perceptron within deep neural networks, convolutional neural network techniques, and k-nearest neighbors algorithms stand out [7]. Models developed using large databases (such as Assay Central, ptoxra, DNN-liverTox) can predict liver damage with over 80% accuracy [7, 41,87].

Cardiotoxicity

Regarding cardiotoxicity, blockage of the human ether-à-go-go related gene (hERG) potassium channel is the main known adverse drug reaction. Consequently, various AI-based computer models such as CardPred and hERGdb have been developed to test *in-vivo* hERG blockage [41]. Databases such as BindingDB, PubChem Bioassay, ChEMBL, and DrugBank are used for evaluating hERG inhibition potential, which is critical in the early stages of drug development. Bayesian models [88] provided 0.91 balanced accuracy, while graph convolutional network models [89] achieved 0.863 balanced accuracy. Another study [17] obtained 0.873 average balanced accuracy on 7889 compounds. Research using random forest and support vector machine models [90] achieved average balanced accuracies of 0.865 and 0.86 respectively. Additionally, a study [91] obtained 0.912 average balanced accuracy on 1668 compounds.

Carcinogenicity and Genotoxicity

Chemical carcinogenesis is becoming an increasingly critical issue in drug discovery processes. The main reason for this importance is the serious and long-term effects of carcinogenic substances on human health. Carcinogens are compounds that cause damage to the genome or trigger tumor formation by disrupting cellular metabolic processes, increase the frequency of existing tumors, or accelerate their emergence process. Carcinogenicity is a critical research area in toxicology science due to its direct effects on DNA and cellular metabolism. Traditionally, carcinogenicity assessment is conducted through *in-vivo* tests on laboratory animals. However, this approach presents a methodology that is costly, time-consuming, and ethically problematic. These limitations have directed researchers to seek alternative solutions. In recent years, prediction models developed in computer environments (*in silico*) have emerged as an effective alternative for carcinogenicity assessment. Among these computer-based approaches, QSAR models are the most comprehensively researched and widely accepted methods (Table 2). QSAR models can predict the carcinogenic potential of untested compounds based on mathematical relationships between molecular structure and biological activity [41]. ML approaches developed using the Carcinogenic Potency Database (CPDB) have provided significant advancements in this field with 147 different models such as CarcinoPred-EL and CapsCarcino. Similarly, in mutagenicity studies, various ML and DL classification models have been developed using databases based on the Ames test and *in-vitro* experiments. These models are successfully used to predict disease risks that could be caused by changes in DNA and abnormal genetic mutations [7].

Computational models can make comprehensive predictions by also correlating physicochemical and structural properties of chemicals with *in-vitro* genotoxicity endpoints [7]. ML and DL models are

successfully used particularly in genotoxicity assessments such as toxicity in mammalian cells [92], *in-vivo* micronucleus assay [93], Comet assay [19], and Ames test [94].

Cytotoxicity

Cytotoxicity is one of the fundamental types of toxicity in terms of examining toxic effects at the cellular level. Cytotoxicity assessments, which play a critical role in the early stages of the drug development process, are supported by ML and DL algorithms in the modern approach. Computational models developed using algorithms such as random forest, Support vector machine, and Multilayer perceptron (like ProTox 3.0., admetSAR1, and admetSAR2) show high success rates in cytotoxicity prediction [7]. Webel et al. [20] demonstrated that *in silico* methods save time and cost by reducing animal experiments and guide the early design process. The researchers used an in-house dataset containing more than 34.000 compounds for cytotoxicity prediction, and their developed DL model achieved over 70% balanced accuracy rate with less than 5% cytotoxic molecule ratio. In the study, identification of toxicophores was enabled using the Deep Taylor Decomposition method, and visual analysis of the relationship between cytotoxicity maps and substructures with toxicity was performed.

Skin Toxicity

Skin sensitization is one of the most common forms of human immune toxicity and therefore is an important endpoint in drug discovery and cosmetics. Models developed using compounds from OECD's eChemPortal database and databases such as NICEATM have shown toxic substructures that likely cause skin sensitization [41]. AI applications offer effective solutions in evaluating effects such as irritation, rash, and burns caused by chemicals on the skin. ML and DL models are successfully used as alternatives to the traditional mouse lymph node assay, particularly in predicting skin sensitization and irritation, which are important in terms of allergic contact dermatitis [7,21].

Reproductive Toxicity

Reproductive toxicity is an important type of toxicity covering developmental and reproductive toxicity endpoints. While developmental toxicity can lead to the death of the organism or structural-functional abnormalities, reproductive toxicity can cause problems affecting the fetus such as teratogenicity and growth retardation [95]. While traditional approaches use *in-vitro* tests in rats and rabbits, computational approaches such as ML and DL models are now successfully used as alternative methods for evaluating various endpoints of reproductive toxicity [7]. Zhang et al. [96] compared seven different ML methods for predicting chemical-induced developmental toxicity in the early stages of drug development. The researchers, using Naïve Bayes, support vector machine, k-nearest neighbor, decision tree, random forest, and AdaBoost methods, achieved the highest performance with the Naïve Bayes classifier. The model showed 91.11% overall prediction accuracy and 91.50% balanced accuracy in the training set, while providing 83.93% concordance and 81.85% balanced accuracy in the test set. The study also identified 10 important molecular descriptors associated with developmental toxicity using genetic algorithm.

Other Specific Toxicity Prediction Methods

AI applications in acute toxicity prediction provide a powerful tool for evaluating oral, dermal, and inhalation toxicity categories determined by the EPA. The acute oral LD50 data repository of 12.000 chemicals, created through collaboration between the NTP Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) and the National Center for Computational Toxicology (NCCT), makes significant contributions to the development of ML and DL classification models [7].

As an example of AI applications in medical toxicology, the Tak model can successfully distinguish six different toxidromes (anticholinergic, cholinergic, opioid, sedative-hypnotic, serotonin toxicity, and sympathomimetic). The model has shown higher performance than both physicians and traditional decision tree models in challenging clinical cases, with a transparent decision-making process and minimal maintenance requirements [97].

Additionally, Hidden Markov Models stand out as an effective AI tool in detecting hidden states in toxicological and clinical studies. These models provide high accuracy in evaluating ADMET

parameters in toxicogenomic, toxicoproteomic, and toxicometabolomic analyses [98]. Successful results have been obtained in clinical applications such as schizophrenia prediction [99], depression monitoring [100], and opioid use disorder treatment evaluation [101].

In environmental toxicology, AI applications offer effective solutions in pollutant modeling and ecosystem risk assessment. DL techniques are used for predicting water quality parameters, while artificial neural networks are used for detecting pollutant sources [102,103]. Random forest technique has been used to predict the possible timing of harmful algal blooms that threaten environmental and human health, and $PM_{2.5}$ pollution levels in the atmosphere over the last 40 years [104-106]. The EPA's ToxCast program evaluates the health effects of environmental pollutants quickly and cost-effectively by performing toxicity predictions of chemical components using *in-vitro* and *in-silico*-based methods [107,108].

Regulatory Aspects of AI-Based Tools in Toxicology

Regulatory agencies worldwide are increasingly evaluating and implementing AI-based tools in toxicological assessments. Agencies like US EPA and FDA are adopting defined approaches for various compounds including pesticides, industrial chemicals, and pharmaceuticals, while European implementation remains more limited to specific endpoints such as skin sensitization (OECD Guideline No. 497) and eye irritation (OECD Guideline No. 467) [109]. Regulatory bodies are developing frameworks for standardized reporting of "New Approach Methodologies" data, including the OECD's OHT 201 template for mechanistic toxicity information and the OECD Omics Reporting Framework (OORF) for omics data types [110]. Traditional validation processes are being reconsidered, with movements toward performance-based test guidelines, reference substances, and case studies to accelerate acceptance while maintaining scientific rigor [111]. Initiatives like the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) are working to define regulatory needs and opportunities for new approach methodologies both within the US and globally [109,112]. There is growing recognition that legislation must shift from animal-centric paradigms to mechanism-based approaches that can accommodate new approach methodologies data, with projects like the European Partnership for the Assessment of Risks from Chemicals (PARC) addressing these critical points [113].

Future Perspectives and Limitations

AI and ML applications in toxicology sciences are rapidly developing with the concept of ToxAIcology [114]. They play a complementary role in the transition from traditional animal-based risk assessment to the modern paradigm supported by *in-vitro* and *in-silico* modeling [41]. The study by Sharma et al. [115] showed that *in-vitro* data (based on AUC) provides high accuracy in clinical toxicity predictions. This finding emphasizes the importance of chemical parallels in Tox21 (*in-vitro*) and ClinTox (clinical) databases and reveals the necessity of a rigorous approach in data analysis [9].

The heterogeneous structure of toxicological data and the requirements of different algorithms [25,28] create significant limitations regarding data quality and standardization. The "black box" structure of DL systems [116] and challenges in mechanistic explainability [54] are issues that require attention in terms of model interpretability. To address this limitation, the AOP framework is used to provide a theoretical basis for IVIVE and to improve the prediction capabilities of ML algorithms [8]. Personalized toxicology approaches and pharmacogenomic research, which center on individuals' genetic profiles [117], play an important role in understanding the complexity of biological systems and person-specific risk assessment.

Additionally, in line with the fundamental principle of toxicology that "The dose makes the poison," computational toxicology models such as read-across and QSAR, and dose-response relationship predictions shape computational resource and technical infrastructure requirements [31]. Green toxicology and modern testing platforms [118] integrate with AI systems to provide more ethical and reliable assessments.

The future of ML in toxicology is built on data quality and accessibility [61]. The 3R principle for personalized toxicology and animal welfare [119] demonstrates the advantages provided by ML. Model interpretability [111,120] and regulatory oversight mechanisms [121,122] are critical

development points in toxicology sciences.

RESULT AND DISCUSSION

AI applications in toxicology have become a complementary factor shaping the future of the field by modernizing traditional toxicity assessment methods. The integration of methods such as PBTK models, QSAR analyses, AOP approaches, and toxicogenomic with AI has accelerated drug development processes, optimized costs, and increased the accuracy of toxicity predictions. Personalized toxicology approaches, thanks to AI technologies, enable the development of safer drugs by taking into account genetic variations and metabolic differences, while also allowing comprehensive risk assessments in environmental toxicology. In the future, AI-supported toxicology applications are expected to become more widespread and standardized, with technological advancements and methodological innovations overcoming current limitations. The increasing acceptance and inclusion of AI-supported toxicity assessment systems by regulatory authorities in legal frameworks will contribute to the introduction of safer products to the market, while the continuous development of these technologies will continue to play a critical role in the evolution of toxicology science.

AUTHOR CONTRIBUTIONS

Concept: S.Y., M.B.K.; Design: M.B.K.; Control: M.B.K.; Sources: S.Y., R.Ç.T., M.B.K.; Materials: - ; Data Collection and/or Processing: S.Y., R.Ç.T., M.B.K.; Analysis and/or Interpretation: S.Y., R.Ç.T., M.B.K.; Literature Review: S.Y., R.Ç.T., M.B.K.; Manuscript Writing: S.Y., R.Ç.T., M.B.K.; Critical Review: S.Y., R.Ç.T., M.B.K.; Other: -

CONFLICT OF INTEREST

The authors declare that there is no real, potential, or perceived conflict of interest for this article.

REFERENCES

1. Klaassen, C.D. (2018). Casarett & Doull's Toxicology: The Basic Science of Poisons, 9th ed. McGraw Hill, New York, 1-1648.
2. Jia, X., Wang, T., Zhu, H. (2023). Advancing computational toxicology by interpretable machine learning. Environmental Science and Technology, 57(46), 17690-17706. [CrossRef]
3. Sinha, K., Ghosh, N., Sil, P.C. (2023). A review on the recent applications of deep learning in predictive drug toxicological studies. Chemical Research in Toxicology, 36(8), 1174-1205. [CrossRef]
4. Shaki, F., Amirkhanloo, M., Chahardori, M. (2024). The future and application of artificial intelligence in toxicology. Asia Pacific Journal of Medical Toxicology, 13(1), 21-28. [CrossRef]
5. Mullowney, M.W., Duncan, K.R., Elsayed, S.S., et al. (2023). Artificial intelligence for natural product drug discovery. Nature Reviews Drug Discovery, 22(11), 895-916. [CrossRef]
6. Kleinstreuer, N., Tetko, I., Tong, W. (2021). Introduction to special issue: Computational toxicology. Chemical Research in Toxicology, 34(2), 171-175. [CrossRef]
7. Guo, W., Liu, J., Dong, F., et al. (2023). Review of machine learning and deep learning models for toxicity prediction. Experimental Biology and Medicine, 248(21), 1952-1973. [CrossRef]
8. Han, P., Li, X., Yang, J., et al. (2024). Advancing toxicity predictions: A review on *in vitro* to *in vivo* extrapolation in next-generation risk assessment. Environment and Health, 2(7), 499-513. [CrossRef]
9. Tonoyan, L., Siraki, A.G. (2024). Machine learning in toxicological sciences: Opportunities for assessing drug toxicity. Frontiers in Drug Discovery, 4, 1336025. [CrossRef]
10. Gupta, R., Srivastava, D., Sahu, M., et al. (2021). Artificial intelligence to deep learning: Machine intelligence approach for drug discovery. Molecular Diversity, 25(3), 1315-1360. [CrossRef]
11. Kamiya, Y., Handa, K., Miura, T., et al. (2021). In silico prediction of input parameters for simplified physiologically based pharmacokinetic models for estimating plasma, liver, and kidney exposures in rats after oral doses of 246 disparate chemicals. Chemical Research in Toxicology, 34(2), 507-513. [CrossRef]
12. Badwan, B.A., Liaropoulos, G., Kyrodimos, E., et al. (2023). Machine learning approaches to predict drug efficacy and toxicity in oncology. Cell Reports Methods, 3(2), 100413. [CrossRef]
13. Singh, A.V., Ansari, M.H.D., Rosenkranz, D., et al. (2020). Artificial intelligence and machine learning in computational nanotoxicology: Unlocking and empowering nanomedicine. Advanced Healthcare

Materials, 9(17), 1901862. [CrossRef]
14. Keshava, C., Nicolai, S., Vulimiri, S.V., et al. (2023). Application of systematic evidence mapping to identify available data on the potential human health hazards of selected market-relevant azo dyes. Environment International, 176, 107952. [CrossRef]
15. Mehrvar, S., Himmel, L.E., Babburi, P., et al. (2021). Deep learning approaches and applications in toxicologic histopathology: Current status and future perspectives. Journal of Pathology Informatics, 12(1), 42. [CrossRef]
16. Chen, M., Suzuki, A., Thakkar, S., et al. (2016). DILIrank: The largest reference drug list ranked by the risk for developing drug-induced liver injury in humans. Drug Discovery Today, 21(4), 648-653. [CrossRef]
17. Cai, C., Guo, P., Zhou, Y., et al. (2019). Deep learning-based prediction of drug-induced cardiotoxicity. Journal of Chemical Information and Modeling, 59(3), 1073-1084. [CrossRef]
18. Li, T., Tong, W., Roberts, R., et al. (2021). DeepCarc: deep learning-powered carcinogenicity prediction using model-level representation. Frontiers in Artificial Intelligence, 4, 757780. [CrossRef]
19. Sizochenko, N., Syzochenko, M., Fjodorova, N., et al. (2019). Evaluating genotoxicity of metal oxide nanoparticles: Application of advanced supervised and unsupervised machine learning techniques. Ecotoxicology and Environmental Safety, 185, 109733. [CrossRef]
20. Webel, H.E., Kimber, T.B., Radetzki, S., et al. (2020). Revealing cytotoxic substructures in molecules using deep learning. Journal of Computer-Aided Molecular Design, 34(7), 731-746. [CrossRef]
21. Hu, F., Santagostino, S.F., Danilenko, D.M., et al. (2022). Assessment of skin toxicity in an *in vitro* reconstituted human epidermis model using deep learning. The American Journal of Pathology, 192(4), 687-700. [CrossRef]
22. Hu, Y., Ren, Q., Liu, X., et al. (2023). *In silico* prediction of human organ toxicity via artificial intelligence methods. Chemical Research in Toxicology, 36(7), 1044-1054. [CrossRef]
23. Janiesch, C., Zschech, P., Heinrich, K. (2021). Machine learning and deep learning. Electronic Markets, 31(3), 685-695. [CrossRef]
24. Sharma, N., Sharma, R., Jindal, N. (2021). Machine learning and deep learning applications-a vision. Global Transitions Proceedings, 2(1), 24-28. [CrossRef]
25. Lin, Z., Chou, W.C. (2022). Machine learning and artificial intelligence in toxicological sciences. Toxicological Sciences, 189(1), 7-19. [CrossRef]
26. Rashidi, H.H., Albahra, S., Robertson, S., et al. (2023). Common statistical concepts in the supervised machine learning arena. Frontiers in Oncology, 13, 1130229. [CrossRef]
27. Duffull, S., Isbister, G. (2022). Challenges faced when modeling clinical toxicology and toxinology events. CPT: Pharmacometrics & Systems Pharmacology, 11(5), 532. [CrossRef]
28. Cheng, W., Ng, C.A. (2019). Using machine learning to classify bioactivity for 3486 per- and polyfluoroalkyl substances (PFASs) from the OECD list. Environmental Science and Technology, 53(23), 13970-13980. [CrossRef]
29. Xu, X., Zhao, P., Wang, Z., et al. (2021). *In silico* prediction of chemical acute contact toxicity on honey bees via machine learning methods. Toxicology in Vitro, 72, 105089. [CrossRef]
30. Russo, D.P., Strickland, J., Karmaus, A.L., et al. (2019). Nonanimal models for acute toxicity evaluations: Applying data-driven profiling and read-across. Environmental Health Perspectives, 127(4), 047001. [CrossRef]
31. Karim, A., Riahi, V., Mishra, A., et al. (2021). Quantitative toxicity prediction via meta ensembling of multitask deep learning models. ACS Omega, 6(18), 12306-12317. [CrossRef]
32. Zhang, L., Ai, H., Chen, W., et al. (2017). CarcinoPred-EL: Novel models for predicting the carcinogenicity of chemicals using molecular fingerprints and ensemble learning methods. Scientific Reports, 7(1), 1-14. [CrossRef]
33. Wang, Y.W., Huang, L., Jiang, S.W., et al. (2020). CapsCarcino: A novel sparse data deep learning tool for predicting carcinogens. Food and Chemical Toxicology, 135, 110921. [CrossRef]
34. Wang, C.C., Liang, Y.C., Wang, S.S., et al. (2022). A machine learning-driven approach for prioritizing food contact chemicals of carcinogenic concern based on complementary *in silico* methods. Food and Chemical Toxicology, 160, 112802. [CrossRef]
35. Luechtefeld, T., Marsh, D., Rowlands, C., et al. (2018). Machine learning of toxicological big data enables read-across structure activity relationships (RASAR) outperforming animal test reproducibility. Toxicological Sciences, 165(1), 198-212. [CrossRef]
36. Pu, L., Naderi, M., Liu, T., et al. (2019). EToxPred: A machine learning-based approach to estimate the toxicity of drug candidates. BMC Pharmacology and Toxicology, 20(1), 1-15. [CrossRef]
37. McCoubrey, L.E., Elbadawi, M., Orlu, M., et al. (2021). Machine learning uncovers adverse drug effects

on intestinal bacteria. Pharmaceutics, 13(7), 1026. [CrossRef]
38. Ciallella, H.L., Russo, D.P., Sharma, S., et al. (2022). Predicting prenatal developmental toxicity based on the combination of chemical structures and biological data. Environmental Science and Technology, 56(9), 5984-5998. [CrossRef]
39. Mayr, A., Klambauer, G., Unterthiner, T., et al. (2016). DeepTox: Toxicity prediction using deep learning. Frontiers in Environmental Science, 3, 167215. [CrossRef]
40. Mansouri, K., Karmaus, A.L., Fitzpatrick, J., et al. (2021). CATMoS: Collaborative acute toxicity modeling suite. Environmental Health Perspectives, 129(4), 14. [CrossRef]
41. Pérez Santín, E., Rodríguez Solana, R., González García, M., et al. (2021). Toxicity prediction based on artificial intelligence: A multidisciplinary overview. Wiley Interdisciplinary Reviews: Computational Molecular Science, 11(5), e1516. [CrossRef]
42. Maltarollo, V.G., Gertrudes, J.C., Oliveira, P.R., et al. (2015). Applying machine learning techniques for ADME-Tox prediction: A review. Expert Opinion on Drug Metabolism & Toxicology, 11(2), 259-271. [CrossRef]
43. Chou, W.C., Lin, Z. (2023). Machine learning and artificial intelligence in physiologically based pharmacokinetic modeling. Toxicological Sciences, 191(1), 1-14. [CrossRef]
44. Pradeep, P., Patlewicz, G., Pearce, R., et al. (2020). Using chemical structure information to develop predictive models for *in vitro* toxicokinetic parameters to inform high-throughput risk-assessment. Computational Toxicology, 16, 100136. [CrossRef]
45. Tan, Y.M., Worley, R.R., Leonard, J.A., et al. (2018). Challenges associated with applying physiologically based pharmacokinetic modeling for public health decision-making. Toxicological Sciences, 162(2), 341-348. [CrossRef]
46. Chen, J., Noorlander, A., Wesseling, S., et al. (2023). Integrating *in vitro* data and physiologically based kinetic modeling to predict and compare acute neurotoxic doses of saxitoxin in rats, mice, and humans. Environmental Science and Technology, 57(30), 10974-10984. [CrossRef]
47. Chang, X., Palmer, J., Lumen, A., et al. (2022). Quantitative *in vitro* to *in vivo* extrapolation for developmental toxicity potency of valproic acid analogues. Birth Defects Research, 114(16), 1037-1055. [CrossRef]
48. Zhang, C., Zhang, Q., Li, J., et al. (2020). Integration of *in vitro* data from three dimensionally cultured HepaRG cells and physiologically based pharmacokinetic modeling for assessment of acetaminophen hepatotoxicity. Regulatory Toxicology and Pharmacology, 114, 104661. [CrossRef]
49. Scholze, M., Taxvig, C., Kortenkamp, A., et al. (2020). Quantitative *in vitro* to *in vivo* extrapolation (QIVIVE) for predicting reduced anogenital distance produced by anti-androgenic pesticides in a rodent model for male reproductive disorders. Environmental Health Perspectives, 128(11), 1-17. [CrossRef]
50. Lu, J., Deng, K., Zhang, X., et al. (2021). Neural-ODE for pharmacokinetics modeling and its advantage to alternative machine learning models in predicting new dosing regimens. IScience, 24(7), 102804. [CrossRef]
51. Cheng, Y.H., He, C., Riviere, J.E., et al. (2020). Meta-analysis of nanoparticle delivery to tumors using a physiologically based pharmacokinetic modeling and simulation approach. ACS Nano, 14(3), 3075-3095. [CrossRef]
52. Ankley, G.T., Bennett, R.S., Erickson, R.J., et al. (2010). Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment. Environmental Toxicology and Chemistry, 29(3), 730-741. [CrossRef]
53. OECD. (2025). Adverse outcome pathways, molecular screening and toxicogenomics. The Organisation for Economic Co-operation and Development (OECD). https://www.oecd.org/en/. Access date: 02.02.2025.
54. Ciallella, H.L., Russo, D.P., Aleksunes, L.M., et al. (2021). Revealing adverse outcome pathways from public high-throughput screening data to evaluate new toxicants by a knowledge-based deep neural network approach. Environmental Science and Technology, 55(15), 10875-10887. [CrossRef]
55. Huang, R., Sakamuru, S., Martin, M.T., et al. (2014). Profiling of the Tox21 10K compound library for agonists and antagonists of the estrogen receptor alpha signaling pathway. Scientific Reports, 4(1), 1-9. [CrossRef]
56. Zgheib, E., Gao, W., Limonciel, A., et al. (2019). Application of three approaches for quantitative AOP development to renal toxicity. Computational Toxicology, 11, 1-13. [CrossRef]
57. Sinitsyn, D., Garcia-Reyero, N., Watanabe, K.H. (2022). From qualitative to quantitative AOP: A case study of neurodegeneration. Frontiers in Toxicology, 4, 838729. [CrossRef]
58. Spinu, N., Cronin, M.T.D., Enoch, S.J., et al. (2020). Quantitative adverse outcome pathway (qAOP) models for toxicity prediction. Archives of Toxicology, 94(5), 1497-1510. [CrossRef]

59. Allen, T.E.H., Liggi, S., Goodman, J.M., et al. (2016). Using molecular initiating events to generate 2D structure-activity relationships for toxicity screening. Chemical Research in Toxicology, 29(10), 1611-1627. [CrossRef]
60. Allen, T.E.H., Goodman, J.M., Gutsell, S., et al. (2018). Using 2D structural alerts to define chemical categories for molecular initiating events. Toxicological Sciences, 165(1), 213-223. [CrossRef]
61. Richarz, A.N. (2019). Big data in predictive toxicology: Challenges, opportunities and perspectives. Issues in Toxicology, 41, 1-37. [CrossRef]
62. Ciallella, H.L., Zhu, H. (2019). Advancing computational toxicology in the big data era by artificial intelligence: Data-driven and mechanism-driven modeling for chemical toxicity. Chemical Research in Toxicology, 32(4), 536-547. [CrossRef]
63. Judson, R., Richard, A., Dix, D., et al. (2008). ACToR - Aggregated computational toxicology resource. Toxicology and Applied Pharmacology, 233(1), 7-13. [CrossRef]
64. Gaulton, A., Bellis, L.J., Bento, A.P., et al. (2012). ChEMBL: A large-scale bioactivity database for drug discovery. Nucleic Acids Research, 40(D1), D1100-D1107. [CrossRef]
65. Davis, A.P., Grondin, C.J., Johnson, R.J., et al. (2021). Comparative toxicogenomics database (CTD): Update 2021. Nucleic Acids Research, 49(D1), D1138-D1143. [CrossRef]
66. Ganter, B., Tugendreich, S., Pearson, C.I., et al. (2005). Development of a large-scale chemogenomics database to improve drug candidate selection and to understand mechanisms of chemical toxicity and action. Journal of Biotechnology, 119(3), 219-244. [CrossRef]
67. Jeliazkova, N., Chomenidis, C., Doganis, P., et al. (2015). The eNanoMapper database for nanomaterial safety information. Beilstein Journal of Nanotechnology, 6, 1609-1634. [CrossRef]
68. Kim, S., Chen, J., Cheng, T., et al. (2021). PubChem in 2021: New data content and improved web interfaces. Nucleic Acids Research, 49(D1), D1388-D1395. [CrossRef]
69. Luechtefeld, T., Maertens, A., Russo, D.P., et al. (2016). Global analysis of publicly available safety data for 9,801 substances registered under reach from 2008-2014. ALTEX, 33(2), 95. [CrossRef]
70. Bitsch, A., Jacobi, S., Melber, C., et al. (2006). REPDOSE: A database on repeated dose toxicity studies of commercial chemicals-A multifunctional tool. Regulatory Toxicology and Pharmacology, 46(3), 202-210. [CrossRef]
71. Nair, S.K., Eeles, C., Ho, C., et al. (2020). ToxicoDB: An integrated database to mine and visualize large-scale toxicogenomic datasets. Nucleic Acids Research, 48(W1), W455-W462. [CrossRef]
72. Fonger, G.C., Stroup, D., Thomas, P.L., et al. (2000). Toxnet: A computerized collection of toxicological and environmental health information. Toxicology and Industrial Health, 16(1), 4-6. [CrossRef]
73. Yan, X., Sedykh, A., Wang, W., et al. (2020). Construction of a web-based nanomaterial database by big data curation and modeling friendly nanostructure annotations. Nature Communications, 11(1), 1-10. [CrossRef]
74. Jeliazkova, N., Apostolova, M.D., Andreoli, C., et al. (2021). Towards FAIR nanosafety data. Nature Nanotechnology, 16(6), 644-654. [CrossRef]
75. Atila, Ü., Baydilli, Y.Y., Sehirli, E., et al. (2020). Classification of DNA damages on segmented comet assay images using convolutional neural network. Computer Methods and Programs in Biomedicine, 186, 105192. [CrossRef]
76. Zhang, S., Xu, Y., Hui, X., et al. (2017). Improvement in prediction of prostate cancer prognosis with somatic mutational signatures. Journal of Cancer, 8(16), 3261. [CrossRef]
77. Lu, T.P., Kuo, K.T., Chen, C.H., et al. (2019). Developing a prognostic gene panel of epithelial ovarian cancer patients by a machine learning model. Cancers, 11(2), 270. [CrossRef]
78. Patti, G.J., Yanes, O., Siuzdak, G. (2012). Metabolomics: The apogee of the omics trilogy. Nature Reviews Molecular Cell Biology, 13(4), 263-269. [CrossRef]
79. Ideker, T., Thorsson, V., Ranish, J.A., et al. (2001). Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. Science, 292(5518), 929-934. [CrossRef]
80. Chen, Q., Chou, W.C., Lin, Z. (2022). Integration of toxicogenomics and physiologically based pharmacokinetic modeling in human health risk assessment of perfluorooctane sulfonate. Environmental Science and Technology, 56(6), 3623-3633. [CrossRef]
81. Liu, Z., Huang, R., Roberts, R., et al. (2019). Toxicogenomics: A 2020 vision. Trends in Pharmacological Sciences, 40(2), 92-103. [CrossRef]
82. Rahman, S.M., Lan, J., Kaeli, D., et al. (2022). Machine learning-based biomarkers identification from toxicogenomics - Bridging to regulatory relevant phenotypic endpoints. Journal of Hazardous Materials, 423, 127141. [CrossRef]
83. Chen, X., Roberts, R., Tong, W., (2022). Tox-GAN: An artificial intelligence approach alternative to animal studies-a case study with toxicogenomics. Toxicological Sciences, 186(2), 242-259. [CrossRef]

84. Igarashi, Y., Nakatsu, N., Yamashita, T., et al. (2015). Open TG-GATEs: A large-scale toxicogenomics database. Nucleic Acids Research, 43(D1), D921-D927. [CrossRef]
85. Cavasotto, C.N., Scardino, V. (2022). Machine learning toxicity prediction: Latest advances by toxicity end point. ACS Omega, 7(51), 47536-47546. [CrossRef]
86. Zhang, F., Sun, B., Diao, X., et al. (2021). Prediction of adverse drug reactions based on knowledge graph embedding. BMC Medical Informatics and Decision Making, 21(1), 1-11. [CrossRef]
87. Moein, M., Heinonen, M., Mesens, N., et al. (2023). Chemistry-based modeling on phenotype-based drug-induced liver injury annotation: From public to proprietary data. Chemical Research in Toxicology, 36(8), 1238–1247. [CrossRef]
88. Liu, L.L., Lu, J., Lu, Y., et al. (2014). Novel Bayesian classification models for predicting compounds blocking hERG potassium channels. Acta Pharmacologica Sinica, 35(8), 1093-1102. [CrossRef]
89. Chen, Y., Yu, X., Li, W., et al. (2023). In silico prediction of hERG blockers using machine learning and deep learning approaches. Journal of Applied Toxicology, 43(10), 1462-1475. [CrossRef]
90. Siramshetty, V.B., Chen, Q., Devarakonda, P., et al. (2018). The Catch-22 of predicting hERG blockade using publicly accessible bioactivity data. Journal of Chemical Information and Modeling, 58(6), 1224-1233. [CrossRef]
91. Shen, M.Y., Su, B.H., Esposito, E.X., et al. (2011). A comprehensive support vector machine binary hERG classification model based on extensive but biased end point hERG data sets. Chemical Research in Toxicology, 24(6), 934-949. [CrossRef]
92. Fujita, Y., Honda, H., Yamane, M., et al. (2019). A decision tree-based integrated testing strategy for tailor-made carcinogenicity evaluation of test substances using genotoxicity test results and chemical spaces. Mutagenesis, 34(1), 101-109. [CrossRef]
93. Fan, D., Yang, H., Li, F., et al. (2018). *In silico* prediction of chemical genotoxicity using machine learning methods and structural alerts. Toxicology Research, 7(2), 211-220. [CrossRef]
94. Zhang, H., Kang, Y.L., Zhu, Y.Y., et al. (2017). Novel naive Bayes classification models for predicting the chemical Ames mutagenicity. Toxicology In Vitro, 41, 56-63. [CrossRef]
95. OECD. (2018). Test No. 414: Prenatal Developmental Toxicity Study. https://www.oecd.org/en. Access date: 02.02.2025.
96. Zhang, H., Mao, J., Qi, H.Z., (2020). *In silico* prediction of drug-induced developmental toxicity by using machine learning approaches. Molecular Diversity, 24(4), 1281-1290. [CrossRef]
97. Chary, M., Boyer, E.W., Burns, M.M. (2021). Diagnosis of acute poisoning using explainable artificial intelligence. Computers in Biology and Medicine, 134, 104469. [CrossRef]
98. Martinelli, D.D. (2023). Machine learning for metabolomics research in drug discovery. Intelligence-Based Medicine, 8, 100101. [CrossRef]
99. Boeker, M., Hammer, H.L., Riegler, M.A., et al. (2023). Prediction of schizophrenia from activity data using hidden Markov model parameters. Neural Computing and Applications, 35(8), 5619-5630. [CrossRef]
100. Liu, Q., Cole, D., Tran, T., et al. (2024). Intraindividual phenotyping of depression in high-risk youth: An application of a multilevel hidden Markov model. Development and Psychopathology, 36(3), 1262-1271. [CrossRef]
101. King, C.A., Englander, H., Todd Korthuis, P., et al. (2021). Designing and validating a Markov model for hospital-based addiction consult service impact on 12-month drug and non-drug related mortality. PLOS ONE, 16(9), e0256793. [CrossRef]
102. Wang, N., Dong, G., Qiao, R., et al. (2024) Bringing artificial intelligence (AI) into environmental toxicology studies: A perspective of ai-enabled zebrafish high-throughput screening. Environmental Science and Technology, 58(22), 9487-9499. [CrossRef]
103. Asha, P., Natrayan, L., Geetha, B.T., et al. (2022). IoT enabled environmental toxicology for air pollution monitoring using AI techniques. Environmental Research, 205, 112574. [CrossRef]
104. Wu, X., Zhou, Q., Mu, L., et al. (2022). Machine learning in the identification, prediction and exploration of environmental toxicology: Challenges and perspectives. Journal of Hazardous Materials, 438, 129487. [CrossRef]
105. Nelson, N.G., Munoz-Carpena, R., Phlips, E.J., et al. (2018). Revealing biotic and abiotic controls of harmful algal blooms in a shallow subtropical lake through statistical machine learning. Environmental Science and Technology, 52(6), 3527-3535. [CrossRef]
106. Li, H., Yang, Y., Wang, H., et al. (2021). Constructing a spatiotemporally coherent long-term PM2.5 concentration dataset over China during 1980-2019 using a machine learning approach. Science of the Total Environment, 765, 144263. [CrossRef]
107. Wang, M., Wang, X., Huang, K., et al. (2024). Human biomonitoring of environmental chemicals among

elderly in wuhan, china: Prioritizing risks using EPA's ToxCast database. Environmenta Science and Technology, 58(23), 10001-10014. [CrossRef]
108. Punt, A., Firman, J., Boobis, A., et al. (2020). Potential of ToxCast data in the safety assessment of food chemicals. Toxicological Sciences, 174(2), 326-340. [CrossRef]
109. Schmeisser, S., Miccoli, A., von Bergen, M., et al. (2023). New approach methodologies in human regulatory toxicology-Not if, but how and when! Environment International, 178, 108082. [CrossRef]
110. Harrill, J.A., Viant, M.R., Yauk, C.L., et al. (2021). Progress towards an OECD reporting framework for transcriptomics and metabolomics in regulatory toxicology. Regulatory Toxicology and Pharmacology, 125, 105020. [CrossRef]
111. Burgdorf, T., Piersma, A.H., Landsiedel, R., et al. (2019). Workshop on the validation and regulatory acceptance of innovative 3R approaches in regulatory toxicology- Evolution versus revolution. Toxicology In Vitro, 59, 1-11. [CrossRef]
112. Ajisafe, O.M., Adekunle, Y.A., Egbon, E., et al. (2025). The role of machine learning in predictive toxicology: A review of current trends and future perspectives. Life Sciences, 1(378), 123821. [CrossRef]
113. Marx-Stoelting, P., Rivière, G., Luijten, M., et al. (2023). A walk in the PARC: Developing and implementing 21st century chemical risk assessment in Europe. Archives of Toxicology, 97(3), 893-908. [CrossRef]
114. Hartung, T. (2023). ToxAIcology-The evolving role of artificial intelligence in advancing toxicology and modernizing regulatory science. ALTEX, 40(4), 559-570. [CrossRef]
115. Sharma, B., Chenthamarakshan, V., Dhurandhar, A., et al. (2023). Accurate clinical toxicity prediction using multi-task deep neural nets and contrastive molecular explanations. Scientific Reports, 13(1), 1-16. [CrossRef]
116. Petch, J., Di, S., Nelson, W. (2022). Opening the black box: The promise and limitations of explainable machine learning in cardiology. Canadian Journal of Cardiology, 38(2), 204-213. [CrossRef]
117. Capone, P., Chiarella, P., Sisto, R. (2024). Advanced technologies in genomic toxicology: Current trend and future directions. Current Opinion in Toxicology, 37, 100444. [CrossRef]
118. Maertens, A., Luechtefeld, T., Knight, J., et al. (2024). Alternative methods go green! green toxicology as a sustainable approach for assessing chemical safety and designing safer chemicals. Altex, 41(1), 3-19. [CrossRef]
119. Hubrecht, R.C., Carter, E. (2019). The 3Rs and humane experimental technique: Implementing change. Animals, 9(10), 754. [CrossRef]
120. Vellido, A. (2020). The importance of interpretability and visualization in machine learning for applications in medicine and health care. Neural Computing and Applications, 32(24), 18069-18083. [CrossRef]
121. Guan, H., Dong, L., Zhao, A. (2022). Ethical risk factors and mechanisms in artificial intelligence decision making. Behavioral Sciences, 12(9), 343. [CrossRef]
122. Meskó, B., Topol, E.J. (2023). The imperative for regulatory oversight of large language models (or generative AI) in healthcare. Npj Digital Medicine, 6(1), 1-6. [CrossRef]