



Predictive Classification of Clinical Drug Toxicity Using Molecular and Pharmacological Characteristics

Anamika Chauhan

Chandigarh University, Punjab, Email: pro.anamika@gmail.com

***Corresponding author:** Anamika Chauhan, Email: pro.anamika@gmail.com

Article History:

Received: 2026-03-17

Revised: 2026-06-08

Accepted: 2026-07-11

Published: 2026-07-26

Abstract

Drug-induced liver injury is still a significant problem in drug development as clinically important toxicity can occur following exposure even when preclinical results are satisfactory. In this study, the predictive classification of clinical drug toxicity was examined using the drug-induced liver injury severity and toxicity (DILIST) dataset from the U.S. Food and Drug Administration (FDA). A dataset of 1,279 distinct drugs that were DILIST-positive or DILIST-negative, along with route-of-administration data, was provided. Descriptive analyses described prevalence of toxicity and patterns of pharmacological routes and chi-square testing assessed the association between categories of routes and toxicity classification. The route-based logistic regression model was tested by stratified 5-fold cross validation and an independent 20% test set. Of 1,279 drugs, 768 (60.0%) were DILIST-positive and 511 (40.0%) were DILIST-negative. Route of administration was reported for 1013 compounds and most were administered orally. A significant difference in the proportion of DILIST-positive compounds was detected between the route categories, but the ability to discriminate with route alone was limited. The analysis confirms the utility of curated clinical toxicity labels, and reveals that pharmacological route data alone is not sufficient to predict toxicity well. For future studies and validation, integration of validated molecular descriptors, structural fingerprints, exposure variables and mechanistic pharmacology is therefore needed to improve and interpret clinical drug-toxicity classification models.

Keywords: Drug Toxicity, Drug-Induced Liver Injury, Predictive Classification, Pharmacological Characteristics, Molecular Descriptors.

1. Introduction

Historically, drug toxicity has been a major reason for attrition in pharmaceutical development and continues to be a concern post marketing due to the interactions between chemical structure, pharmacokinetics, metabolism, dose, target engagement and patient susceptibility. As a result, the focus of computational toxicology has shifted from rules-based alerts to those that are driven by data and can incorporate heterogeneous chemical and biological data (Wang et al., 2020). Drug-induced liver injury (DILI) has a special significance due to the fact that it can occur as transient biochemical perturbations up to the severe hepatic failure, and can result in clinical restrictions and withdrawals. However, machine-learning studies on curated DILI resources have shown that clinically relevant hepatotoxicity can be tackled as a classification problem, but it relies heavily on the definition of the endpoint, quality of the features, and the design of the validation strategy (Ancuceanu et al., 2020). Chierici et al. (2020) also demonstrated that there are limitations to the predictability of DILI, cautioning interpretation of a DILI score as biological causality.

Molecular and pharmacological characteristics complement each other in providing information about toxicity. Physicochemical characteristics affect the passage of the compound through the membrane, its distribution in tissues, the metabolic burden, the occurrence of reactive metabolites and the exposure, while pharmacological characteristics influence the mode and site of administration and action. Oral DILI studies have demonstrated that the combination of multiple features and endpoints could be beneficial for classification, as a single descriptor family was used previously (Liu et al., 2020). However, medicinal chemistry evidence suggests that structural and physicochemical liabilities can also lead to a hepatotoxic risk via covalent binding, mitochondrial stress, transporter inhibition and metabolic activation (Norman, 2020). But, clinically observed DILI cannot be simply explained by one molecular property. Aleo et al. (2020) suggested to shift from binary notions of risk to a relative-risk characterization, and provide examples of how the output from models should be understood in the context of exposure and pharmacology.

Recently, several artificial-intelligence methods have led to an increase in the number of feature spaces that can be used for toxicity modeling. DeepDILI was able to enhance DILI prediction by leveraging learned representations, and provided proof-of-concept for representation learning for complicated toxicity endpoints (Li et al., 2021). Descriptor-based support vector machines have also demonstrated that carefully selected molecular characteristic can still be able to contain significant predictive information (Jaganathan et al., 2021). Concurrently, the overall promise of AI for predicting DILI also comes with challenges such as the size of the available datasets, the definition of the classes, the applicability of the classes to the specific domains, and external validation (Vall et al., 2021). These problems are in line with the broader research in predictive toxicology, which indicates that no single algorithm is found to be generally applicable (Wang et al., 2021).

Clinically anchored datasets are thus useful as they provide a link between the predicted and the human safety data. Models should be clear in terms of evidence being captured in their predictors and early screening for DILI can help prioritize compounds prior to expensive late-stage development. To evaluate the feasibility of the binary classification models, large sets of approved drugs have been used as the basis for model development and demonstrated to be practical (Zhang et al., 2022). In other reviews of toxicity endpoints, the importance of preprocessing, feature selection and independent testing are also reiterated as necessary steps that must be taken to obtain useful prediction (Cavasotto & Scardino, 2022). Interpretability is also essential; knowledge-guided approaches may be able to offer a clearer understanding of what biological or chemical features are important for predictions, as opposed to just opaque labels (Hao et al., 2022). Recent clinical-toxicity modelling has integrated neural networks with molecular explanations to further enhance prediction and interpretation (Sharma et al.,

2023). In this context, we sought to define the clinical classification resource, DIList, to describe the toxicity patterns and evaluate the predictive information in the available pharmacological data, and to identify the additional molecular information needed for a fully integrated model, using the DIList clinical classification resource.

An important additional feature of the clinically useful classifier is that it should maintain a clear relationship between the definition of the predictors and the intended decision context. Molecular descriptors can capture information on size, polarity, lipophilicity, flexibility, aromaticity and hydrogen-bonding capacity, while structural fingerprints will enable the encoding of local chemical environment, which may show liabilities not represented by global physical-chemical measures. Pharmacological variables can provide information about pathways of exposure and therapeutic situations. But these features must be created from real structures and handled during the training procedure so as not to leak. It is therefore the discrimination, sensitivity to toxic compounds, specificity, calibration and reproducibility of the model, and not only the accuracy, that should be emphasized in model comparisons. The endpoint DIList is especially useful for this purpose as it is clinically oriented and at the drug level. A baseline can be used to set contribution of pharmacological information prior to molecular representations.

Research Objectives

- To characterize the distribution of DIList-positive and DIList-negative drugs and their available pharmacological route profiles.
- To evaluate whether route-of-administration characteristics provide reproducible predictive information for clinical drug-toxicity classification.
- To establish an interpretable analytical framework for subsequent integration of validated molecular descriptors with clinical toxicity labels.

2. Methodology

2.1 Study Design and Data Source

A retrospective computational classification study was performed with the Drug-Induced Liver Injury Severity and Toxicity (DIList) dataset from the U.S. Food and Drug Administration (U.S. Food and Drug Administration, 2023). A total of 1,279 drug records (including a unique DIList identifier, compound name, binary DIList classification and route of administration) was supplied in the provided workbook. The coded DIList positive compounds were 1 and DIList negative compounds were 0. Duplicate compound names were checked before analysis.

2.2 Data Preprocessing and Pharmacological Feature Engineering

Entries for the route of administration were harmonized with respect to capitalization and obvious spelling variations. Routes were aggregated into clinically meaningful categories: oral, intravenous/vascular, subcutaneous, intramuscular, topical/transdermal, inhalation, other and not reported. The names of the compounds were kept for future matching of structures. No molecular structures or calculated descriptors for molecules were included in the supplied workbook and thus no molecular values were inferred or fabricated.

2.3 Descriptive and Inferential Analysis

Summary of toxicity classes and route categories in counts and percentages. Within each route group, proportions of DIList positiveness were derived. Association of route category and toxicity classification was tested with a Pearson chi-square test. P-values < 0.05 were considered statistically significant.

2.4 Predictive Classification

Logistic regression was performed to build a logistic regression classifier with the single pharmacologic predictor, route category. An 80:20 stratified train-test split with

fixed random seed was used to split data. The categorical route values were one-hot encoded and logistic regression was used to reduce the bias caused by the unequal class sizes. A 5-fold stratified cross-validation was only carried out in the training data.

2.5 Model Evaluation

Cross-validated ROC-AUC scores and independent test-set accuracy, sensitivity, specificity, precision, F1-score and ROC-AUC scores were used to measure performance. Sensitivity and specificity were obtained from a confusion matrix. As the source workbook did not contain molecular descriptors the route-only model was used as a baseline rather than a full molecular toxicity model. All analyses were carried out in Python with packages pandas, SciPy, scikit-learn and Matplotlib.

3. Results

3.1 Clinical Toxicity Profile

The composition of the analytical data is presented in table 1. A binary toxicity classification was obtained for all 1,279 compounds (768 DILIST-positive drugs (60.0%) and 511 DILIST-negative drugs (40.0%)). For 1013 compounds (79.2%) route information was available; 266 records did not include a reported route. The moderate class imbalance towards DILIST-positive compounds is shown in Figure 1.

Table 1. Clinical and Pharmacological Characteristics of the Analytical Cohort

Characteristic	n	%
Total drugs	1279	100.0
DILIST-positive	768	60.0
DILIST-negative	511	40.0
Route reported	1013	79.2
Route not reported	266	20.8

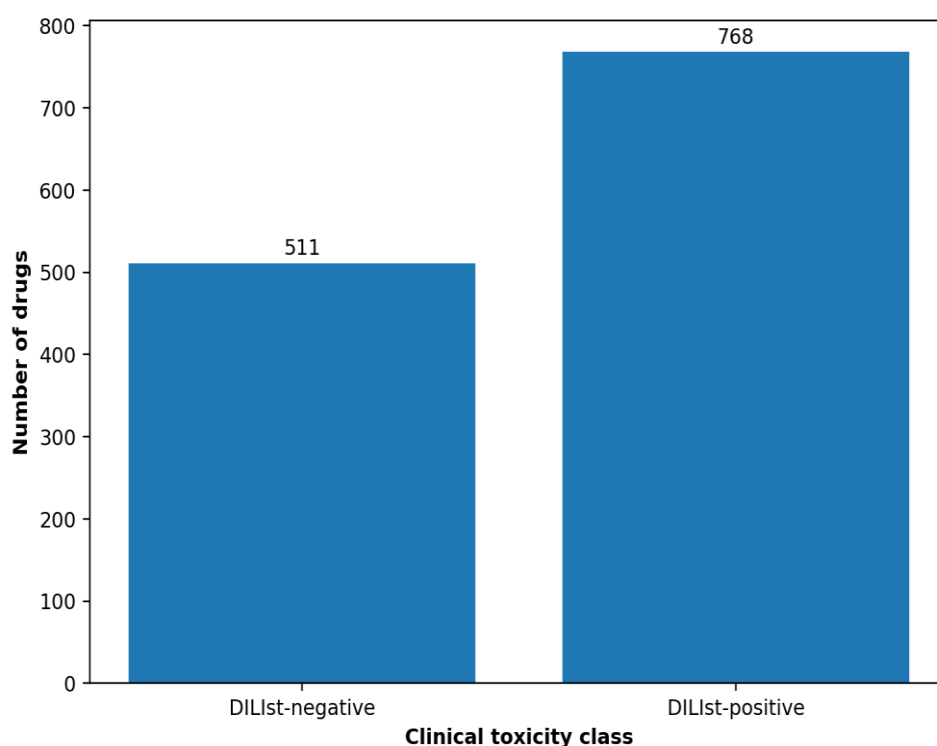


Figure 1. Distribution of Clinical Drug Toxicity Classes

3.2 Route-of-Administration Patterns

The route distribution and toxicity proportions are given as standardized values in Table 2. The most commonly recorded administration route was oral, with the second being intravenous/vascular. As can be seen in Figure 2, there is a substantial focus on these two route groups as opposed to the smaller subcutaneous, intramuscular, topical/transdermal, inhalation and other route groups. DIList-positive was more or less present in different proportions according to the route category as shown in Figure 3.

Table 2. Route-of-Administration Profile According to DIList Classification

Route_group	Total	Positive	Negative	Positive (%)
Oral	715	427	288	59.7
Not reported	266	248	18	93.2
Intravenous/vascular	213	62	151	29.1
Subcutaneous	42	14	28	33.3
Topical/transdermal	16	4	12	25.0
Intramuscular	16	7	9	43.8
Other	7	3	4	42.9
Inhalation	4	3	1	75.0

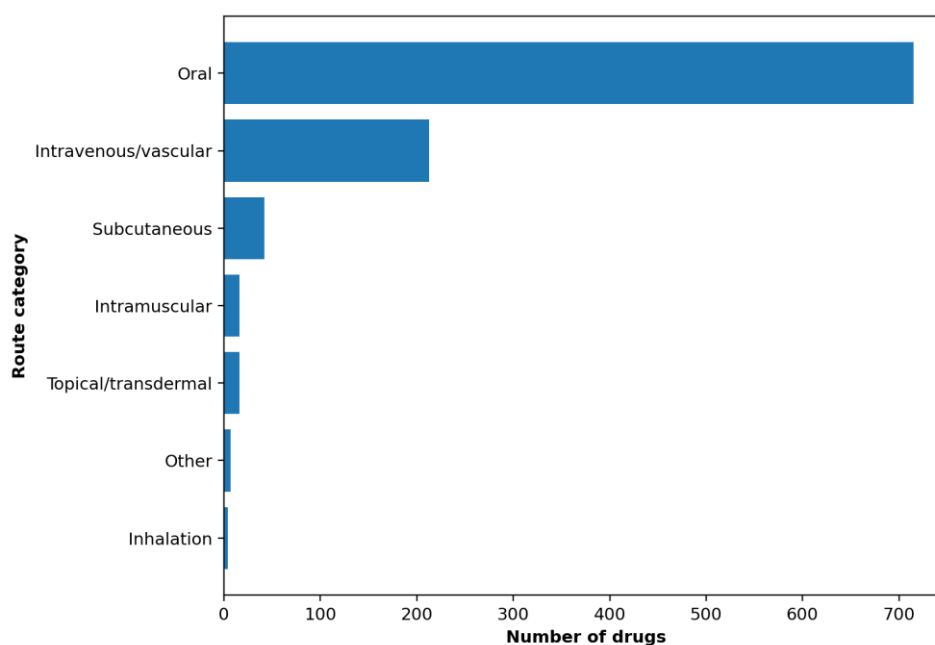


Figure 2. Distribution of Standardized Route-of-Administration Categories

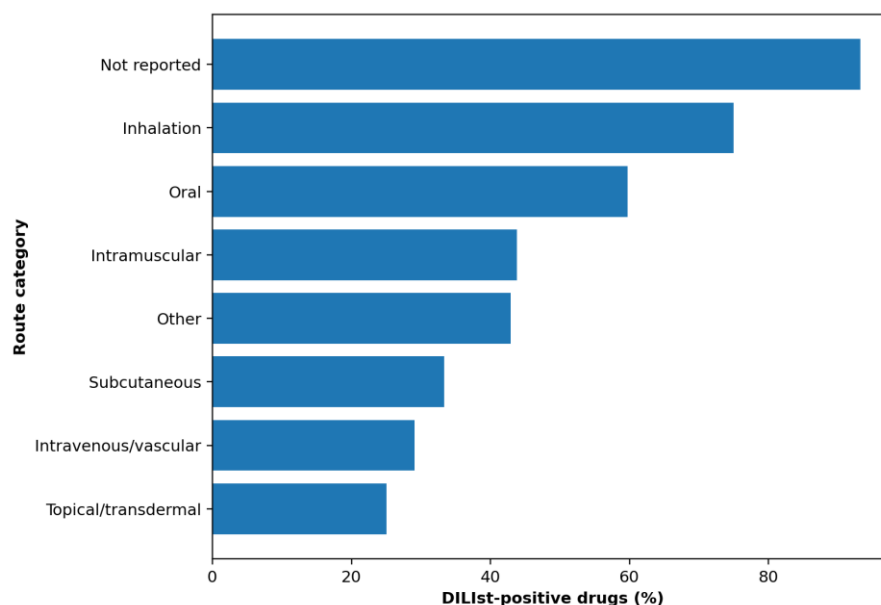


Figure 3. DILIst-Positive Proportion Across Route Categories

3.3 Association Between Pharmacological Route and Toxicity

Table 3 shows the inferential comparison between route category and DILIst classification. The chi-square statistic was 230.82 with 7 degrees of freedom ($p < 0.001$). This result indicated that toxicity-class proportions were not uniformly distributed across route categories. Because several route groups were small and route was not a molecular determinant, the association was interpreted as descriptive pharmacological structure rather than evidence of causation.

Table 3. Association Between Route of Administration and DILIst Classification

Comparison	Chi-square	df	p-value
Route of administration	230.82	7	<0.001

3.4 Baseline Predictive Classification

Table 4 summarizes the route-only logistic regression model. Mean five-fold cross-validated ROC-AUC was 0.722 (SD 0.032). On the independent test set, accuracy was 0.668, sensitivity 0.857, specificity 0.382, precision 0.677, F1-score 0.756, and ROC-AUC 0.704. Figure 4 displays the test-set ROC curve. The modest discrimination demonstrated that route information alone did not provide sufficient information for clinically useful toxicity classification and supported the need for molecular descriptors, structural fingerprints, dose/exposure variables, and mechanistic pharmacology.

Table 4. Performance of the Route-Only Baseline Toxicity Classifier

Model	CV ROC-AUC	CV SD	Accuracy	Sensitivity	Specificity	Precision	F1-score	Test ROC-AUC
Route-only logistic regression	0.722	0.032	0.668	0.857	0.382	0.677	0.756	0.704

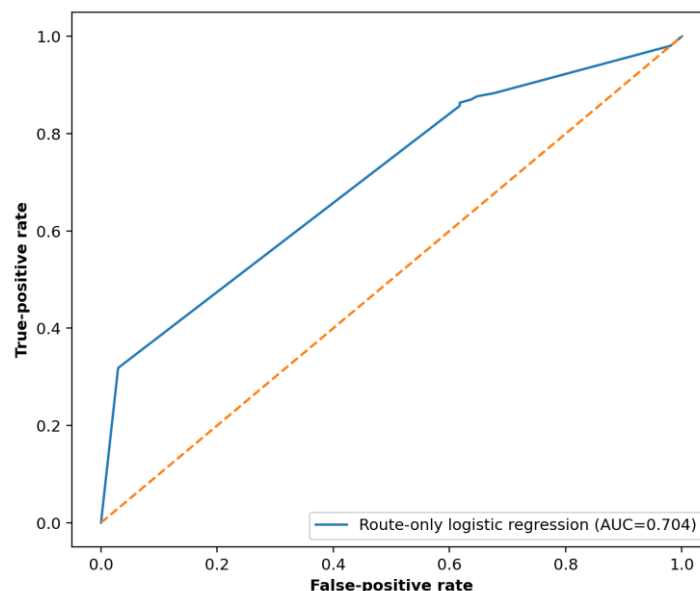


Figure 4. ROC Curve for the Route-Only Logistic Regression Model

4. Discussion

The results of the present analysis illustrate both the promise and the shortcomings of clinically based toxicity data for prediction of classification. The DILIST resource was clearly binary with 60.0% of the drugs being DILIST-positive. This class distribution was adequate for conventional supervised learning, and justifies the stratification and class-aware modeling. As more and more data becomes available that enables the prediction of toxicity, there is an increasing trend of merging chemical and biological data, since the toxicity seen clinically is more than just one data domain (Liu et al., 2023). The results of this study confirm this: with regard to the class of toxicity, the route of administration was found to vary across the classes, but no stand alone predictive discrimination was obtained. Comparisons of machine-learning and deep-learning models for toxicity also show that biologically relevant features are the key to success in making a good model, not just the use of a machine-learning algorithm (Guo et al., 2023).

The correlation of route category to DILIST classification should not be considered a reflection of hepatotoxicity associated with a route. Route can alter bioavailability, first pass exposure, peak concentration, and tissue distribution, but also intrinsic chemical reactivity and metabolism, dose, length of treatment, immune response and host susceptibility influence DILI. Thus, models aimed at estimating the severity of DILI have been built that include a wider array of molecular and pharmacological data instead of a single administration variable (Rao et al., 2023). In a broader sense, the artificial-intelligence (AI) methods for toxicity prediction benefit from the use of combined descriptors, fingerprints, biological assays, and a pharmacological context and ensure keeping training and evaluation data separated (Tran et al., 2023). Clinical reviews also highlight the variety and individuality of DILI, which make it difficult to predict simply and deterministically (Allison et al., 2023).

The relatively poor performance of the baseline (route only) is therefore not surprising. Modern molecular models are able to represent the structures, which are not shown in the route label. Graph-based methods, for instance, represent atom and bond relationships and are able to capture substructures relevant for toxicity prediction that are not possible with the classical categorical predictors. When combined with clinically relevant labels, machine-learning DILI prediction is also found to be more useful when compared to failed drug candidates (Mostafa et al., 2024). Further improvement in hepatotoxicity classification has been achieved using ensemble algorithms that combine complementary algorithms and molecular representation, however, the performance of ensemble relies on the composition of the dataset and

applicability domain (Khan et al., 2024). Based on these observations, I suggest to extend the current DILIST framework by adding molecular weight, lipophilicity, polar surface area, hydrogen-bonding, rotatable-bond, aromaticity, and structural fingerprint features which can be reproducibly generated.

Those molecular variables will need to be incorporated and interpretability will be important. However, if the classifier makes a prediction that is not associated with a plausible chemical and/or pharmacological liability, then it cannot be translated to a high-performing classifier. A recent study on prediction of clinical adverse events demonstrates the value of combining molecular or target-level properties with machine learning techniques to find patterns that relate to adverse events (Cao et al., 2024). Effective molecular representations have also been used to stack various strategies and boost DILI prediction performance while also harnessing the complementary model strengths (Guan et al., 2025). Future feature-rich analysis of the presented data set should be performed on the transparent logistic models and nonlinear models (random forest, support vector machine, gradient boosting and graph based) along with the reporting of sensitivity and specificity in addition to ROC-AUC.

A number of restrictions should be noted. First, molecular structures were not included in the DILIST workbook provided and nor were the molecular descriptors precomputed, so the present reproducible analysis could not realistically be said to have molecular predictor effects. Second, for a significant subpopulation, no route information was available and multi-route entries (which include pharmacological solutions) had to be consolidated, which might have reduced pharmacological resolution. Third, DILIST labels are clinically informed toxicity classification which do not reflect patient level susceptibility, dose, exposure, duration, comorbidity or interacting medications. Fourth, random train-test partitioning can provide a good estimate of internal generalization, but cannot replace external validation of chemically different compounds. Future research efforts are thus warranted to expand the same compounds to include verified structures and molecular descriptors, to include exposure and mechanism information, and to test scaffold-aware validation and calibration in an independent clinical toxicity dataset. The extensions would create an integrated molecular-pharmacological classifier as suggested by the title of the study and maintain the clinical relevance of the DILIST endpoint.

The next step of the analysis from a translational point of view should be focused on the replication of the model rather than on superfluous model complexities. If the structure matching could be verified and the salts, mixtures etc., could be dealt with explicitly and the calculation of the descriptors pre-specified and the exclusion criteria transparent, the confidence in molecular associations would be greater. By providing an indication of the calibration, applicability domains and errors by chemical class, a level of clarity is provided as to where predictions can be trusted and assessment is necessary.

5. Conclusion

The present study assessed the clinical relevance of drug toxicity data and the prediction of classification of drug toxicity data with the information available in the provided data set and FDA DILIST resource. A binary clinically anchored endpoint was created by identifying 768 DILIST-positive and 511 DILIST-negative drugs from a list of 1279 unique drugs. Most compounds had some variability in proportions of toxicity by route of administration, and route of administration was reported for the majority of compounds. However, the route-only logistic regression model had only modest discrimination, showing that administration route alone is not enough to have adequate information to be a good predictor for drug-induced liver injury. This discovery is crucial because it represents a pharmaco-logical correlation that can be reproduced, but not a molecular predictability that cannot. Molecular structures and/or calculated descriptors were not provided with the supplied work book and molecular effects were

not fabricated or inferred. The proposed framework should be more robustly implemented to include verified chemical structure and molecular weight, lipophilicity, polar surface area, hydrogen-bonding, flexibility, aromaticity, structural fingerprint, or other exposure or mechanism variables. Thus, these features can be tested with interpretable and nonlinear classifiers which are tested under rigorous cross validation and independent testing. The DIList dataset thus serves as an excellent starting point for clinical outcome data, and the present analysis as a clear baseline and guide for specific data enrichment to enable defensible prediction of molecular and pharmacological toxicity. The step-wise sequence enables quick and easy comparison of simple and advanced algorithms without over-stating the source data. Before considering the use of prediction models for drug-development or clinical safety decision support, external validation of models on chemically distinct drugs and evaluation of false-negative predictions should be emphasized.

References

1. Aleo, M. D., et al. (2020). Moving beyond binary predictions of human drug-induced liver injury (DILI) toward contrasting relative risk potential. *Chemical Research in Toxicology*, 33(1), 223-238. <https://doi.org/10.1021/acs.chemrestox.9b00262>
2. Allison, R., et al. (2023). Drug induced liver injury-A 2023 update. *Journal of Toxicology and Environmental Health, Part B*, 26(8), 442-467. <https://doi.org/10.1080/10937404.2023.2261848>
3. Ancuceanu, R., et al. (2020). Computational models using multiple machine learning algorithms for predicting drug hepatotoxicity with the DILIRank dataset. *International Journal of Molecular Sciences*, 21(6), 2114. <https://doi.org/10.3390/ijms21062114>
4. Cao, A., Zhang, L., Bu, Y., & Sun, D. (2024). Machine learning prediction of on/off target-driven clinical adverse events. *Pharmaceutical Research*, 41(8), 1649-1658. <https://doi.org/10.1007/s11095-024-03742-x>
5. Cavasotto, C. N., & Scardino, V. (2022). Machine learning toxicity prediction: Latest advances by toxicity end point. *ACS Omega*, 7(51), 47536-47546. <https://doi.org/10.1021/acsomega.2c05693>
6. Chierici, M., et al. (2020). Predictability of drug-induced liver injury by machine learning. *Biology Direct*, 15(1), 3. <https://doi.org/10.1186/s13062-020-0259-4>
7. Guan, J., et al. (2025). StackDILI: Enhancing drug-induced liver injury prediction through stacking strategy with effective molecular representations. *Journal of Chemical Information and Modeling*, 65(2), 1027-1039. <https://doi.org/10.1021/acs.jcim.4c02079>
8. Guo, W., et al. (2023). Review of machine learning and deep learning models for toxicity prediction. *Experimental Biology and Medicine*, 248(21), 1952-1973. <https://doi.org/10.1177/15353702231209421>
9. Hao, Y., Romano, J. D., & Moore, J. H. (2022). Knowledge-guided deep learning models of drug toxicity improve interpretation. *Patterns*, 3(9), 100565. <https://doi.org/10.1016/j.patter.2022.100565>
10. Jaganathan, K., Tayara, H., & Chong, K. T. (2021). Prediction of drug-induced liver toxicity using SVM and optimal descriptor sets. *International Journal of Molecular Sciences*, 22(15), 8073. <https://doi.org/10.3390/ijms22158073>
11. Khan, M. Z. I., et al. (2024). Comprehensive hepatotoxicity prediction: Ensemble model integrating machine learning and deep learning. *Frontiers in Pharmacology*, 15, 1441587. <https://doi.org/10.3389/fphar.2024.1441587>
12. Li, T., Tong, W., Roberts, R., Liu, Z., & Thakkar, S. (2021). DeepDILI: Deep learning-powered drug-induced liver injury prediction using model-level representation. *Chemical Research in Toxicology*, 34(2), 550-565. <https://doi.org/10.1021/acs.chemrestox.0c00374>

13. Liu, A., Seal, S., Yang, H., & Bender, A. (2023). Using chemical and biological data to predict drug toxicity. *SLAS Discovery*, 28(3), 53-64. <https://doi.org/10.1016/j.slasd.2022.12.003>
14. Liu, X., et al. (2020). Machine-learning prediction of oral drug-induced liver injury (DILI) via multiple features and endpoints. *BioMed Research International*, 2020, 4795140. <https://doi.org/10.1155/2020/4795140>
15. Martin, M. T., et al. (2022). Early drug-induced liver injury risk screening: “Free,” as good as it gets. *Toxicological Sciences*, 188(2), 208-218. <https://doi.org/10.1093/toxsci/kfac054>
16. Mostafa, F., Howle, V., & Chen, M. (2024). Machine learning to predict drug-induced liver injury and its validation on failed drug candidates in development. *Toxics*, 12(6), 385. <https://doi.org/10.3390/toxics12060385>
17. Norman, B. H. (2020). Drug induced liver injury (DILI): Mechanisms and medicinal chemistry avoidance/mitigation strategies. *Journal of Medicinal Chemistry*, 63(20), 11397-11419. <https://doi.org/10.1021/acs.jmedchem.0c00524>
18. Rao, M., et al. (2023). AI/ML models to predict the severity of drug-induced liver injury for small molecules. *Chemical Research in Toxicology*, 36(7), 1129-1139. <https://doi.org/10.1021/acs.chemrestox.3c00098>
19. Sharma, B., et al. (2023). Accurate clinical toxicity prediction using multi-task deep neural nets and contrastive molecular explanations. *Scientific Reports*, 13, 4908. <https://doi.org/10.1038/s41598-023-31169-8>
20. Teng, S., et al. (2023). MolFPG: Multi-level fingerprint-based graph Transformer for accurate and robust drug toxicity prediction. *Computers in Biology and Medicine*, 164, 106904. <https://doi.org/10.1016/j.compbimed.2023.106904>
21. Tran, T. T. V., Wibowo, A. S., Tayara, H., & Chong, K. T. (2023). Artificial intelligence in drug toxicity prediction: Recent advances, challenges, and future perspectives. *Journal of Chemical Information and Modeling*, 63(9), 2628-2643. <https://doi.org/10.1021/acs.jcim.3c00200>
22. U.S. Food and Drug Administration. (2023). Drug-Induced Liver Injury Severity and Toxicity (DILIST) Dataset. FDA Liver Toxicity Knowledge Base. <https://www.fda.gov/science-research/liver-toxicity-knowledge-base-ltkb/drug-induced-liver-injury-severity-and-toxicity-dilist-dataset>
23. Vall, A., et al. (2021). The promise of AI for DILI prediction. *Frontiers in Artificial Intelligence*, 4, 638410. <https://doi.org/10.3389/frai.2021.638410>
24. Wang, L., et al. (2020). An overview of machine learning and big data for drug toxicity evaluation. *Chemical Research in Toxicology*, 33(1), 20-37. <https://doi.org/10.1021/acs.chemrestox.9b00227>
25. Wang, M. W. H., Goodman, J. M., & Allen, T. E. H. (2021). Machine learning in predictive toxicology: Recent applications and future directions for classification models. *Chemical Research in Toxicology*, 34(2), 217-239. <https://doi.org/10.1021/acs.chemrestox.0c00316>
26. Zhang, H., Zhang, H.-R., Hu, M.-L., & Qi, H.-Z. (2022). Development of binary classification models for assessment of drug-induced liver injury in humans using a large set of FDA-approved drugs. *Journal of Pharmacological and Toxicological Methods*, 116, 107185. <https://doi.org/10.1016/j.vascn.2022.107185>