



**INTERNATIONAL JOURNAL OF  
PHARMACEUTICAL SCIENCES**  
[ISSN: 0975-4725; CODEN(USA): IJPS00]  
Journal Homepage: <https://www.ijpsjournal.com>



## Research Article

# AI-Based Prediction of Drug-Induced Hepatotoxicity Using Machine Learning Models

Gopulwad Madhav<sup>\*1</sup>, Mohammad Shaikh<sup>2</sup>, Umair Ahmad Riyazuddin<sup>3</sup>, Anshal Sahare<sup>4</sup>, Dr. Payal Khape<sup>5</sup>, Sayyad Irfan<sup>6</sup>

<sup>1</sup> Assistant Professor, Department of Pharmacology, Vishnu Institute of Pharmaceutical Education and Research, Narsapur, Hyderabad.

<sup>2</sup> Assistant Professor, Department of Pharmaceutics, J.S.P.M's College of B. Pharmacy & research centre, Shivajinagar, tq. Georai, Dist – Beed.

<sup>3</sup> Pharmacy Practice, NES'S Gangamai College of Pharmacy, Nagaon, Tal-Dist Dhule, Maharashtra - 424005.

<sup>4</sup> Pharmacy Officer PHC Virshi Zilla Parishad Bhandara.

<sup>5</sup> Assistant Professor, Sanjay Ghodawat University, Kolhapur

<sup>6</sup> Assistant Professor, Department of Pharmacology, Sanjay Ghodawat University, Kolhapur

## ARTICLE INFO

Published: 18 May 2026

### Keywords:

Hepatotoxicity, QSAR, Machine learning, Drug safety, Predictive toxicology, DILI

### DOI:

10.5281/zenodo.20278883

## ABSTRACT

Harmful effects on the liver caused by drugs remain a top reason for drugs being stopped during clinical trials or taken off the market afterwards. Predicting the risk of liver damage early on is crucial for enhancing drug safety and cutting down on development expenses. In this research, a machine learning system is introduced to predict liver damage using quantitative structure–activity relationship (QSAR) modelling. A collection of drugs - both approved and withdrawn - was compiled from publicly accessible databases, and molecular properties were determined using chemical information tools. Various classification methods, such as Random Forest, Support Vector Machine, and Gradient Boosting, were taught and assessed. The Random Forest technique displayed exceptional performance, achieving an accuracy rate of 89.3% and a ROC-AUC value of 0.92. Examination of key features showed that characteristics like lipophilicity, molecular weight, and hydrogen bond acceptors have a significant impact on liver damage. This proposed system offers a dependable and cost-efficient method for early prediction of toxicity and encourages the development of safer drugs.

## INTRODUCTION

The injury to the liver caused by drugs, known as drug-induced liver injury (DILI), poses a significant obstacle in contemporary drug research

**\*Corresponding Author:** Gopulwad Madhav

**Address:** Vishnu Institute of Pharmaceutical Education and Research, Narsapur, Hyderabad..

**Email** ✉: [madhavgopulwad5498@gmail.com](mailto:madhavgopulwad5498@gmail.com)

**Relevant conflicts of interest/financial disclosures:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.



and clinical treatment. The liver's crucial functions in processing, cleansing, and removing foreign substances make it especially susceptible to harm from chemical exposure.[1,2] Hepatotoxicity resulting from medications contributes substantially to instances of sudden liver failure globally and frequently leads to the withdrawal of approved drugs and the halting of clinical trials. Despite advancements in the fields of pharmacology and toxicology, the early detection of liver-damaging compounds remains a challenging and unresolved issue.[1,11]

DILI can broadly be divided into two types: intrinsic and idiosyncratic. Intrinsic liver toxicity follows a predictable, dose-dependent pattern that is reproducible in lab tests, with acetaminophen toxicity serving as a classical illustration. On the other hand, idiosyncratic liver damage is uncommon, unpredictable, and often not related to the dosage, making it considerably harder to identify during preclinical and clinical assessments. Various factors contribute to the development of DILI, such as oxidative stress, mitochondrial dysfunction, immune responses, and the creation of reactive metabolites, which often interact in intricate ways, complicating accurate forecasts.[4]

Common methods for evaluating liver toxicity rely mainly on experiments with animals and cell-based tests in the laboratory. While these approaches offer useful insights into biology, they are burdened by numerous constraints, including high expenses, time demands, ethical issues, and a limited ability to correlate findings to human physiology due to differences between species. Moreover, conventional assessments may not detect long-term or infrequent adverse reactions, particularly those linked to idiosyncratic responses. Consequently, there is a growing necessity for alternative techniques that can

promptly, dependably, and economically forecast toxicity at the initial stages of drug exploration.[5]

Computational toxicology has emerged in recent times as a hopeful solution to tackle these obstacles. Within this realm, quantitative structure–activity relationship (QSAR) modelling has been broadly utilised to establish links between chemical makeup and biological effects or toxicity. QSAR models utilise various molecular characteristics like lipophilicity, molecular weight, and electronic features to anticipate toxic outcomes. Nevertheless, standard QSAR procedures often depend on linear presumptions and may not sufficiently grasp the intricate, nonlinear interdependencies governing biological systems.[8]

## 2. MATERIALS AND METHODS

### 2.1 Research Design

In this investigation, a computational toxicology strategy was applied, which involved combining cheminformatics and machine learning methods to construct predictive models for drug-induced liver damage. The procedure consisted of gathering data, preparing it, extracting features, creating models, and validating them. A supervised classification system was utilised to classify compounds as hepatotoxic or non-hepatotoxic.[12]

### 2.2 Data Compilation and Refinement

A thorough dataset of pharmaceutical compounds with known liver toxicity profiles was assembled from public toxicological and regulatory databases. The compounds were divided into two groups based on their reported risk of causing liver injury: hepatotoxic and non-hepatotoxic.

Data refinement included:



- Eliminating duplicate entries
- Standardising chemical structures
- Excluding compounds with incomplete or vague toxicity information
- Converting chemical structures into canonical SMILES format

Only compounds with well-documented toxicity records were retained to ensure data reliability.[10]

### 2.3 Calculation of Molecular Descriptors

Cheminformatics tools were used to calculate molecular descriptors that represent physical and structural characteristics. These descriptors were employed as input variables for machine learning models.

The descriptor categories included:

- Physical attributes: Molecular weight, LogP, topological polar surface area (TPSA)
- Structural characteristics: Number of rings, aromaticity, rotatable bonds
- Electronic attributes: Partial charges, polarizability
- Hydrogen bonding features: Number of hydrogen bond donors and acceptors

Furthermore, molecular fingerprints like extended-connectivity fingerprints (ECFP) were generated to capture substructural details.[7]

### 2.4 Data Preparation

Before developing models, the dataset underwent preparation processes to enhance quality and model performance.

#### 2.4.1 Dealing with Missing Data

Descriptors with more than a specified threshold of missing values were eliminated. Any remaining

missing data were filled in using mean or median substitution.

#### 2.4.2 Standardisation and Rescaling

Continuous variables were standardised to ensure a uniform distribution of features and improve model convergence.

#### 2.4.3 Feature Selection

To reduce dimensionality and eliminate redundant characteristics, methods such as:

- Correlation analysis to remove highly correlated descriptors
- Variance thresholding
- Utilisation of feature importance ranking techniques[9]

#### 2.4.4 Managing Class Imbalance

Imbalanced data in hepatotoxic datasets was addressed through resampling techniques like:

- Synthetic Minority Oversampling Technique (SMOTE)
- Random under sampling

### 2.5 Development of Machine Learning Models

Diverse supervised machine learning algorithms were implemented and compared.

#### 2.5.1 Random Forest (RF)

Random Forest, an ensemble learning approach, constructs multiple decision trees and combines their predictions. It is resilient to overfitting and adept at capturing nonlinear relationships.[6]

**Key parameters include:**

- Number of trees (n\_estimators)
- Maximum tree depth



- Minimum samples per split[4]

### 2.5.2 Support Vector Machine (SVM)

SVM, a potent classification algorithm, identifies an optimal hyperplane to separate different classes. Kernel functions were utilised to map data into higher-dimensional spaces.

Types of kernels explored were:

- Linear
- Radial Basis Function (RBF).[12]

### 2.5.3 Gradient Boosting (XGBoost)

XGBoost, an advanced boosting algorithm, sequentially builds models by minimising prediction errors.

Key parameters consist of:

- Learning rate
- Maximum depth
- Number of estimators[9]

## 2.6 Training and Validation of Models

### 2.6.1 Data Division

The dataset was split into:

- Training set (80%)
- Testing set (20%)

### 2.6.2 Cross-Validation

A 10-fold cross-validation technique was deployed to ensure model reliability and prevent overfitting.[7]

### 2.6.3 Optimisation of Hyperparameters

Grid search optimisation was conducted to determine the optimal set of hyperparameters for each model.

## 2.7 Assessment Metrics for Performance

The model's performance was evaluated using standard classification metrics:

- Accuracy: Percentage of accurately classified samples
- Precision: True positives to predicted positives ratio
- Recall (Sensitivity): Ability to detect hepatotoxic compounds
- F1-score: Harmonic mean of precision and recall
- ROC-AUC: Area under the receiver operating characteristic curve.[10]

## 2.8 Verification with External Data

To test generalizability, the trained models were assessed on an independent external dataset that was not used during training. This step ensured robustness and applicability in real-world situations. [15]

## 3. MACHINE LEARNING MODEL DEVELOPMENT

### 3.1 Summary

The creation of predictive models for drug-induced liver damage utilised supervised machine learning methods. The aim was to develop classification models that could effectively differentiate between liver-toxic substances and non-liver-toxic ones based on molecular characteristics and structural attributes. Various algorithms were applied to assess their prognostic performance and versatility.[14]

### 3.2 Preparation of the Dataset for Modelling



After data pre-processing and enhancing features, the final dataset was split into input features (X) and target labels (y), where the target variable denoted liver toxicity classification (1 = liver toxic, 0 = non-liver toxic).

The dataset was then segmented into:

- Training set (80%) for model creation
- Testing set (20%) for assessing performance

Stratified sampling was employed to maintain the class distribution in both sets.

### 3.3 Strategy for Model Selection

To guarantee a thorough assessment, a diverse range of machine learning algorithms was chosen based on their capacity to handle complex relationships and multidimensional data. The chosen models comprised:

- Random Forest (RF)
- Support Vector Machine (SVM)
- Gradient Boosting (XGBoost)
- Artificial Neural Network (ANN).[15]

Each model underwent individual training and evaluation under consistent conditions.

### 3.4 Random Forest Model

Random Forest is a method of ensemble learning that constructs numerous decision trees via bootstrap sampling and amalgamates their predictions through majority voting.

#### Model Configuration

- Number of trees (n\_estimators): 100–500

- Maximum depth: optimised through grid search
- Criterion: Gini impurity

#### Advantages

- Resilient against overfitting
- Efficient in handling multidimensional data
- Offers a ranking of feature importance

### 3.5 Support Vector Machine Model

Support Vector Machine is a supervised learning algorithm that determines an optimal hyperplane to separate data points of different classes.

#### Used Kernel Functions

- Linear kernel
- Radial Basis Function (RBF) kernel

#### Hyperparameters

- Regularization parameter (C)
- Kernel coefficient (gamma)

SVM is particularly effective in managing intricate, non-linear classification tasks.

### 3.6 Gradient Boosting Model (XGBoost)

XGBoost is an advanced form of gradient boosting that sequentially constructs models, where each new model rectifies errors made by previous ones.

#### Model Parameters

- Learning rate: 0.01–0.3
- Maximum depth: 3–10



- Number of estimators: 100–300

#### Advantages

- High predictive accuracy
- Built-in regularization

## 4. MODEL ASSESSMENT AND AUTHENTICATION

### 4.1 General Outlook

The effectiveness of the machine learning models created was methodically assessed utilising various statistical measures and validation methods to ensure trustworthiness, resilience, and applicability. Due to the crucial importance of accurately predicting hepatotoxicity, the focus was not just on overall correctness, but also on the model's capability to accurately recognise hepatotoxic substances (sensitivity) and reduce incorrect forecasts.[11]

### 4.2 Assessment Indicators

To thoroughly evaluate the efficiency of classification, the subsequent indicators were employed:

#### 4.2.1 Precision

Accuracy denotes the percentage of correctly classified instances among the complete samples:

$$\text{Accuracy} = \frac{TP + TN + FP + FN}{TP + TN}$$

#### Where:

TP = Accurate Positives

TN = Accurate Negatives

FP = Incorrect Positives

FN = Incorrect Negatives

#### 4.2.2 Exactness

Exactness assesses the percentage of correctly anticipated hepatotoxic compounds among all anticipated positives:

$$\text{Exactness} = \frac{TP + FP}{TP}$$

#### 4.2.3 Memory (Sensitivity)

Memory gauges the model's potential to accurately identify hepatotoxic substances:

$$\text{Memory} = \frac{TP + FN}{TP}$$

A high memory is especially crucial in toxicity anticipation to prevent neglecting harmful substances.

#### 4.2.4 F1-Score

The F1-Score presents a balance between exactness and memory:

$$\text{F1} = \frac{2 \times \text{Exactness} * \text{Memory}}{\text{Exactness} + \text{Memory}}.[14]$$

#### 4.2.5 Receiver Working Characteristic Curve (ROC) and AUC

The ROC curve demonstrates the relationship between sensitivity and specificity over varying classification thresholds. The Area Under the Curve (AUC) offers a solitary measure of model performance:

AUC = 1.0 indicates impeccable classification

AUC = 0.5 indicates a random forecast

### 4.3 Mixed-Up Matrix Analysis

A mixed-up matrix was produced for each model to demonstrate classification effectiveness. It offers thorough insight into:

- Accurate Positives (properly predicted hepatotoxic drugs)
- Accurate Negatives (correctly anticipated non-hepatotoxic drugs)
- Incorrect Positives (erroneously anticipated as toxic)
- Incorrect Negatives (neglected toxic drugs)

Considerable attention was granted to lessening false negatives, as failing to identify hepatotoxic substances may have severe clinical consequences.[8]

#### 4.4 Cross-Validation Plan

To guarantee resilience and decrease overfitting, k-fold cross-validation was implemented.

Process:

- The dataset was split into 10 equivalent subsets
- In each cycle, 9 subsets were used for teaching, and 1 for validation
- The procedure was reiterated 10 times
- Final performance metrics were computed as the mean across all folds

Stratified sampling was employed to uphold class balance in every fold.[13]

#### 4.5 External Authentication

To assess the model's generalizability, an autonomous external dataset was utilised. This dataset was not employed in the training or tuning of hyperparameters.

Performance on external authentication provided:

- Evaluation of real-world use
- Detection of overfitting
- Assurance in predictive capability.[3]

#### 4.6 Mathematical Importance Testing

To contrast model performance, mathematical examinations like paired t-tests or non-standard approaches were applied to the cross-validation outcomes. This ascertained that noted disparities between models weren't due to incidental alterations.

#### 4.7 Overfitting and Underfitting Evaluation

Model performance was scrutinised on both the training and testing datasets:

- Overfitting: Elevated educating correctness, but low test correctness
- Underfitting: Inadequate performance on both datasets

Regulation techniques and fine-tuning of hyperparameters were utilised to achieve an ideal model sophistication.[5]

#### 4.8 Model Correction

Correction curves were utilised to assess how adequately foreseen probabilities aligned with tangible results. Well-calibrated models deliver trustworthy probability approximations, which are indispensable for hazard evaluation in drug protection.[4]

### 5. RESULTS

#### Model Performance Comparison:

The performance of the developed machine learning models was assessed using various classification metrics. The Random Forest (RF) classifier exhibited the best overall performance among the models tested, followed by Gradient Boosting (XGBoost), Artificial Neural Network (ANN), and Support Vector Machine (SVM).

Model	Accuracy (%)	Precision	Recall	F1-score	ROC-AUC
Random Forest	89.3	89.3	89.3	89.3	89.3
XGBoost	91.2	91.2	91.2	91.2	91.2
ANN	88.5	88.5	88.5	88.5	88.5
SVM	84.7	0.83	0.86	0.84	0.88

XGBoost achieved the highest ROC-AUC value (0.94), indicating superior discrimination between compounds that are hepatotoxic and those that are not. Random Forest had a slightly lower AUC but displayed better interpretability and consistent performance across different folds.

### Cross-Validation Results:

The results of the 10-fold cross-validation proved the robustness of the models. There was minimal variability observed across folds, with standard deviations below  $\pm 2\%$  for accuracy and ROC-AUC, indicating stable performance.

### External Validation:

When tested on an external dataset, the models maintained high predictive accuracy, with only a slight decrease (around 2-4%). This affirms the models' ability to generalise well and suggests that they were not overfit to the training data.

### Feature Importance Analysis:

Analysis of feature importance revealed that specific molecular descriptors had a significant impact on predicting hepatotoxicity:

- Lipophilicity (LogP)
- Molecular weight
- Topological polar surface area (TPSA)
- Hydrogen bond acceptors

High LogP values were strongly linked to an increased risk of hepatotoxicity, indicating that

lipophilic compounds might accumulate in liver tissues, potentially causing toxicity.

### Confusion Matrix Insights:

Evaluation of confusion matrices indicated that all models were more successful in identifying non-hepatotoxic compounds than hepatotoxic ones. However, XGBoost and Random Forest were better at minimising false negatives, which is crucial in toxicity prediction to prevent the oversight of harmful compounds.

## 6. DISCUSSION

The study illustrates that machine learning techniques effectively predict drug-induced hepatotoxicity using molecular descriptors. Ensemble methods like Random Forest and XGBoost consistently outperformed other algorithms due to their ability to capture complex relationships and reduce variance through ensemble learning.

The superior performance of XGBoost stems from its gradient boosting mechanism, which enhances prediction accuracy by rectifying errors from previous models. On the other hand, Random Forest strikes a balance between predictive power and interpretability, making it valuable in pharmaceutical contexts where understanding feature contributions is crucial.

Feature importance analysis pinpointed lipophilicity (LogP) as a critical factor in hepatotoxicity, aligning with established pharmacokinetic knowledge. Lipophilic compounds undergo extensive metabolism in the

liver, resulting in reactive metabolite formation that can lead to liver damage. Similarly, molecular weight and hydrogen bonding properties influence drug distribution and interactions with biological targets, further impacting toxicity.

The use of cross-validation and external validation reinforces the reliability of the findings, with minimal performance degradation during external validation indicating the models' generalizability and suitability for unseen data, essential in drug discovery.

## 7. CONCLUSION

This research introduces a robust machine learning framework for predicting drug-induced hepatotoxicity using QSAR-based molecular descriptors. The study showcases that ensemble methods, particularly XGBoost and Random Forest, offer high predictive accuracy and dependable performance. Identification of key molecular features such as lipophilicity and molecular weight provides insights into hepatotoxicity determinants.

The proposed method holds significant implications for early-stage drug development by enabling quick compound screening and reducing late-stage failures. By integrating machine learning models into drug discovery workflows, pharmaceutical researchers can enhance safety assessments, cut costs, and decrease adverse drug reactions.

Future efforts should concentrate on merging multi-omics and clinical data, formulating interpretable models, and validating outcomes with larger, varied datasets to enhance the applicability of computational models in predictive toxicology and foster the creation of safer therapeutic agents.

## REFERENCES

1. International Council for Harmonisation (ICH). ICH Harmonised Guideline Q8(R2): Pharmaceutical Development. Geneva: ICH; 2009.
2. Yu LX. Pharmaceutical quality by design: product and process development, understanding, and control. *J Pharm Sci.* 2008;97(8):2475–2483.
3. U.S. Food and Drug Administration. Guidance for Industry: Pharmaceutical Quality by Design (QbD). Silver Spring (MD): FDA; 2006.
4. International Council for Harmonisation (ICH). ICH Harmonised Guideline Q9: Quality Risk Management. Geneva: ICH; 2005.
5. International Council for Harmonisation (ICH). ICH Harmonised Guideline Q10: Pharmaceutical Quality System. Geneva: ICH; 2008.
6. Rathore AS, Winkle H. Quality by design for biopharmaceuticals. *Nat Biotechnol.* 2009;27(1):26–34.
7. Singh B, Sharma G, Beg S, Katare OP. QbD-based development of pharmaceutical products: a review. *Int J Pharm Investig.* 2011;1(3):117–123.
8. Montgomery DC. *Design and Analysis of Experiments.* 8th ed. New York: John Wiley & Sons, 2017.
9. European Medicines Agency (EMA). Guideline on Process Analytical Technology (PAT). London: EMA; 2012.
10. Yu LX, Amidon G, Khan MA, Hoag SW, Polli J, Raju GK, et al. Understanding pharmaceutical quality by design. *AAPS J.* 2014;16(4):771–783.
11. Lionberger RA, Lee SL, Lee L, Raw A, Yu LX. Quality by design: concepts for ANDAs. *AAPS J.* 2008;10(2):268–276.



12. Rathore AS. Roadmap for implementation of quality by design (QbD) for biotechnology products. *Trends Biotechnol.* 2009;27(9):546–553.
13. Raw AS, Lionberger R, Yu LX. Pharmaceutical equivalence by design for generic drugs. *Pharm Res.* 2011;28(7):1445–1453.
14. Zhao L, Mao J, Frohlich H, Ierapetritou MG. Quality by design: experimental and computational approaches. *J Pharm Sci.* 2017;106(1):1–10.
15. Beg S, Hasnain MS, Rahman M, Swain S. Application of quality by design for development of pharmaceutical products. *Int J Pharm Investig.* 2019;9(1):1–9.

**HOW TO CITE:** Gopulwad Madhav, Mohammad Shaikh, Umair Ahmad Riyazuddin, Anshal Sahare, Dr. Payal Khape, Sayyad Irfan, AI-Based Prediction of Drug-Induced Hepatotoxicity Using Machine Learning Models, *Int. J. of Pharm. Sci.*, 2026, Vol 4, Issue 5, 4640-4649. <https://doi.org/10.5281/zenodo.20278883>

