



Predicting Toxicity of Herbal and Synthetic Organic Compounds Using Machine Learning-Based QSAR Models

Pratik Khanal^{1*}, Sanjiv Regmi²

¹ Crimson College of Technology, Butwal, Nepal

² Institute of Engineering and IT, Lumbini Technological University, Nepalgunj, Nepal

*Corresponding Email: PratikKhanal18@gmail.com

Received: 7 December 2025 / Accepted: 1 February 2026 / Published: 1 March 2026

Abstract

This study focuses on the development of a machine learning-based Quantitative Structure-Activity Relationship (QSAR) model to predict the toxicity of organic compounds, including both traditional herbal remedies and synthetic compounds. The study employs Logistic Regression, Random Forest, and Support Vector Machines (SVM) to predict potential toxicity based on molecular descriptors calculated using RDKit, achieving over 90% accuracy across models. Feature importance analysis reveals that molecular descriptors such as lipophilicity (logP), hydrogen bond donors, and specific molecular fingerprints (e.g., FP_375, FP_243, FP_417) significantly correlate with toxicity. A Random Forest-based model highlighted these fingerprint bits as key contributors to toxicity prediction, showing strong correlations with known toxicological properties. The top 20 fingerprint features were analyzed, with their importance ranking depicted in a bar chart. The model demonstrates promising results in predicting hepatotoxicity and neurotoxicity, offering an early-stage toxicity screening tool for drug discovery. Validated on external datasets, the model generalizes well to unseen herbal and synthetic compounds, making it a valuable tool for pharmaceutical and herbal compound safety evaluation. This research underscores the potential of integrating traditional medicinal knowledge with advanced computational methods to enhance safety profiling of diverse organic compounds.

Keywords: Machine learning, toxicity, herbal compounds, synthetic compounds, random forest, support vector machine, logistic regression

1. Introduction

Traditional herbal remedies have played a significant role in healthcare systems worldwide, including Ayurveda, Traditional Chinese Medicine (TCM), and other ethnopharmacological practices (Sen et al., 2011). While herbal compounds are often assumed to be safe due to their natural origins, many lack extensive toxicity studies, particularly through modern scientific methods (Woo et al., 2012). This can lead to safety concerns, especially in long-term use or high doses (Moreira et al., 2014).

As drug discovery and toxicology studies progress, computational approaches have gained prominence (Cherkasov et al., 2014). Specifically, Quantitative Structure-Activity Relationship (QSAR) models have become a popular tool for predicting the biological activity of chemical compounds based on their molecular structures (Varsou et al., 2024). QSAR models can predict potential toxic effects without the need for large-scale clinical testing or experimental setups, making them efficient for early-stage screening (EBSCOhost, 2023).

While QSAR models have been applied extensively in the pharmaceutical and industrial sectors, little attention has been given to herbal medicines (Xu et al., 2024). The lack of data about traditional medicines creates a barrier to ensuring the safe use of those compounds in modern healthcare.

There is a growing need for computational models that can predict the toxicity of herbal compounds, thus bridging the gap between traditional medicine and modern toxicological evaluation (Machhar et al., 2019).

This study aims to develop a machine learning – based QSAR model to predict toxicity of herbal and synthetic organic compounds, using RDKit-calculated molecular descriptors. The model will be trained and validated with known toxicity data from herbal and plant derived compounds.

The outcomes could benefit both the scientific community and the herbal medicine industry by offering a tool for early-stage screening, potentially reducing the need for extensive in vivo and in vitro testing (Krewski et al., 2010).

2. Methodology

2.1 Dataset Collection

This study gathered data on the molecular structures and toxicity profiles of herbal compounds from publicly available databases such as:

- a. FDA: FDA approved drugs from fda.gov (2024)
- b. HMDB: HMDB (Human Metabolome Database) for non-toxic metabolites (D. S. Wishart et al., 2022)

- c. T3DB: Data on toxic substances was obtained from T3DB. (D. Wishart et al., 2015)
- d. CPDB: The CPDB is a single standardized resource of the results of 45 years of chronic, long-term carcinogenesis bioassays.
- e. ProTox-3.0: The external validation set as SDfile.(Banerjee et al., 2024)

The dataset included toxicity data for various endpoints, including hepatotoxicity, neurotoxicity, and general cytotoxicity(Yang et al., 2019). Compounds without toxicity data were excluded from the study(Fourches et al., 2016)

2.2 Molecular Descriptor Calculation

To represent the molecular structure of each herbal compound numerically, molecular descriptors were calculated using RDKit, a powerful tool for cheminformatics(Ekins et al., 2014). The descriptors included:

0D Descriptors: 0D (zero-dimensional) descriptors in cheminformatics are typically scalar values that provide information about the overall properties of a molecule without considering its spatial arrangement. E.g., molecular weight, number of atoms, number of heavy atoms, number of hydrogen bond donors and others.

1D Descriptors: 1D (one-dimensional) descriptors in cheminformatics represent counts or specific attributes of molecular features, focusing on individual atom types, bond types, or functional groups without considering the spatial arrangement of the molecule. E.g., total hydrogen atoms, total number of C atoms, number of aromatic amino groups and others.

2D Descriptors: 2D (two-dimensional) descriptors in cheminformatics provide information about the molecular structure and properties based on the arrangement of atoms and bonds, without taking into account three-dimensional (3D) conformations. E.g., Maximum absolute Estate values, Minimum absolute partial charges, charge components and various shape descriptors.

These descriptors were used as input features for the machine learning models.

Molecular Fingerprints: Fingerprints are a specific type of descriptor that encodes molecular structures as bit strings. This encoding consists of a sequence of binary digits (bits), which indicate the presence (1) or absence (0) of particular substructures within the molecule. The resulting numeric array is of length nnn , where nnn is determined by the specific fingerprint algorithm employed. In this study, molecular fingerprints were calculated for the input data using the RDKit library in Python(Rogers & Hahn, 2010). The types of fingerprints calculated include:

- a. Morgan Fingerprint
- b. Atom Pair Fingerprint

- c. Topological Torsion Fingerprint
- d. RDKit Fingerprint

2.3 Machine Learning Algorithms

Three machine learning algorithms were selected for the toxicity prediction task(Lo et al., 2018):

Logistic Regression used 11 molecular descriptors (e.g., LogP, TPSA) selected for toxicity relevance, with features scaled via RobustScaler and missing values imputed with zeros. Hyperparameters (regularization strength C, solvers: liblinear, saga) were optimized using GridSearchCV with 5-fold cross-validation, and the model was evaluated for accuracy and classification metrics, saved with joblib for reproducibility.

Support Vector Machines (SVM) utilized the same 11 descriptors, standardized with StandardScaler, and employed an RBF kernel (C=1, gamma='scale'). The model was assessed for accuracy, precision, recall, and F1-score, and serialized using joblib.

Random Forest used 1024-bit Morgan fingerprints (radius=2) from RDKit as input, with the dataset split into 80% training and 20% testing sets. The model, configured with 100 decision trees (random_state=42), was evaluated on validation and test sets and saved with joblib for external validation.

2.4 Model Training and Validation

The complete dataset was divided into training (80%) and testing (20%) subsets to train and evaluate the machine learning models. Feature engineering involved the calculation of molecular descriptors and fingerprints using RDKit, which served as the input features, while toxicity classification served as the target variable.

For Logistic Regression, hyperparameter tuning was conducted using GridSearchCV with 5-fold cross-validation. The model's regularization strength (C) and solver type (liblinear, saga) were optimized based on validation accuracy.

The Random Forest and Support Vector Machine (SVM) models were trained using fixed hyperparameters. Random Forest was implemented with 100 estimators and a fixed random state to ensure reproducibility, while the SVM used an RBF kernel with C=1 and gamma='scale' as default settings.

After training, all models were evaluated using the test dataset. Performance metrics such as accuracy, precision, recall, and F1-score were calculated to assess the predictive ability of each model(Ballabio et al., 2018). Additionally, confusion matrices and feature importance analyses were employed to understand model decisions and interpretability.

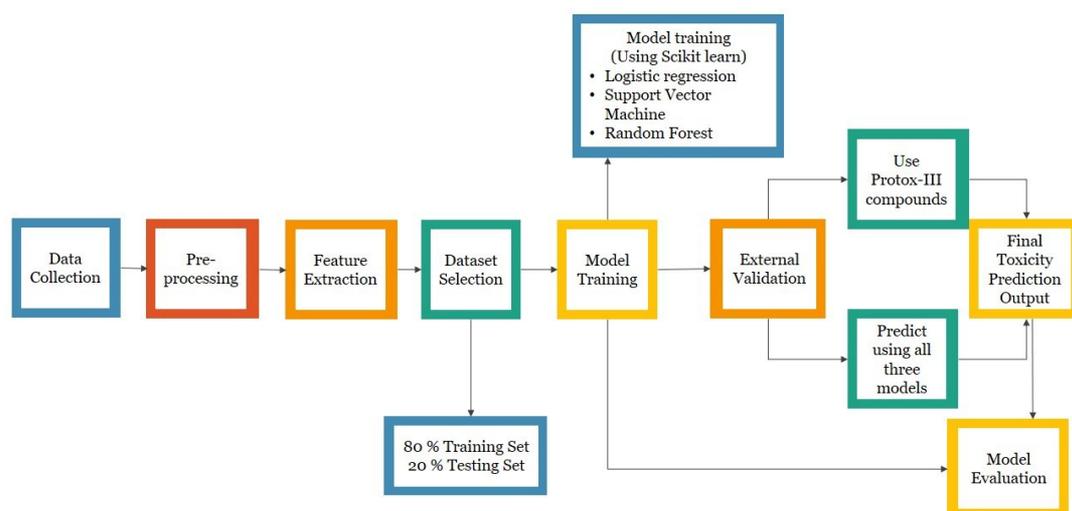


Figure 1: Methodology for Toxicity Prediction using ML-based QSAR Model

2.5 External validation

To assess the generalization ability of the trained Random Forest, Logistic Regression, and Support Vector Machine (SVM) models, an external dataset of 506 structurally diverse compounds from the ProTox-III (Khaouane et al., 2023) validation set (<https://tox.charite.de/protox3/index.php?site=links>) was used. SMILES notations of the compounds were converted into 1024-bit Morgan fingerprints (radius = 2) for the Random Forest model and into 11 molecular descriptors (e.g., molecular weight, LogP, TPSA, rotatable bonds, ring count) using RDKit for the Logistic Regression and SVM models, consistent with their training methodologies. Infinite or missing descriptor values were replaced with zeros. Each model predicted toxicity based solely on chemical structure, without using experimental LD50 values or pre-assigned toxicity classes. Post-prediction, LD50 values were used to classify compounds per the Globally Harmonized System (GHS) as toxic (Classes I–IV, LD50 ≤ 2000 mg/kg) or non-toxic (Classes V–VI, LD50 > 2000 mg/kg) for performance evaluation. Predictions were exported for further analysis.

3. Results and Discussion

3.1 Model Performance

The models were evaluated based on several key performance metrics, including accuracy, precision, recall, and F1 score. The performance of the Logistic Regression, Random Forest, and Support Vector Machine (SVM) models on the test set for toxicity prediction is summarized in the table below:

Table 1: Table Shows Test Set for Toxicity Prediction

| Model | Accuracy | Precision | Recall | F1 Score |
|------------------------|-----------------|------------------|---------------|-----------------|
| Logistic regression | 92.22% | 98.73% | 85.56% | 91.56% |
| Random forest | 92.78% | 98.73% | 86.67% | 92.31% |
| Support vector machine | 92.22% | 87.00% | 100% | 93.00% |

Both Logistic Regression and SVM models achieved an accuracy of 92.22%, while the Random Forest model performed slightly better with an accuracy of 92.78%. All models exhibited high precision (98.73% for Logistic Regression and Random Forest, 87% for SVM), indicating a low rate of false positives.

- a. Logistic Regression showed high precision and good recall (85.56%), resulting in an F1 score of 91.56%.
- b. Random Forest outperformed the others slightly in terms of recall (86.67%) and F1 score (92.31%), identifying slightly more toxic compounds while maintaining a good balance between precision and recall.
- c. Support Vector Machine (SVM) demonstrated a perfect recall (100%), meaning it correctly identified all toxic compounds in the test set, though its precision (87%) was lower than both Logistic Regression and Random Forest. The SVM model achieved an F1 score of 93%, the highest among the three models.

3.2 Analysis of Molecular Descriptors

Feature Importance analysis of Random Forest Model: Feature importance analysis using the Random Forest model revealed that not only classical molecular descriptors such as logP and the number of hydrogen bond donors were correlated with compound toxicity, but also specific molecular fingerprints contributed significantly. This finding aligns with established toxicological principles, where lipophilicity and hydrogen bonding influence membrane permeability and biological activity.

To understand the structural patterns influencing toxicity, the top 10 most important fingerprint bits (FP_375, FP_243, FP_417, FP_595, FP_887, FP_540, FP_591, FP_118, FP_695, and FP_69) were visualized using RDKit. Each fingerprint bit corresponds to a specific molecular substructure that frequently appeared in toxic compounds within the dataset. Representative SMILES structures were identified for each bit:

Table 2: Top 10 Fingerprint Bits

| Fingerprint Bit | Matched SMILES | Structural Insight |
|-----------------|---|---|
| FP_375 | <chem>Cc1ccc(-c2cc(C(F)(F)F)nn2-c2ccc(S(N)(=O)=O)cc2)cc1</chem> | Aromatic ring with trifluoromethyl & sulfonamide groups – both associated with membrane interaction and enzyme binding. |
| FP_243 | <chem>CCCC(C)C1(CC)C(=O)NC(=O)NC1=O</chem> | Branched alkyl chain with cyclic urea – linked to hydrophobicity and metabolic stability. |
| FP_417 | <chem>CC(=O)NC[C@H]1CN(c2ccc(N3CCOCC3)c(F)c2)C(=O)O1</chem> | Piperazine and fluorinated aryl groups – common in CNS-active drugs with potential neurotoxicity. |
| FP_595 | Long chain phospholipid-like ester | Highly lipophilic, mimicking biological membranes – relevant in cytotoxicity. |
| FP_887 | <chem>OCCN1CCN(CCCN2c3ccccc3Sc3cc(Cl)cc32)CC1</chem> | Tertiary amines and sulfur-containing heterocycles – often associated with hepatotoxicity. |
| FP_540 | <chem>CC(=O)Nc1cccc2c1-c1cccc1C2</chem> | Fused aromatic rings with acetamide – planar structures affecting DNA intercalation. |
| FP_591 | <chem>c1ccc2c(c1)[nH]c1ncccc12</chem> | Indole-pyridine fused ring – found in many bioactive compounds, potentially toxic at high doses. |
| FP_118 | <chem>CC(C)CON=O</chem> | N-nitroso group – a classic structural alert for mutagenicity and carcinogenicity. |
| FP_695 | <chem>Cc1ccc2c(c1[N+](=O)[O-])C(=O)c1cccc1C2=O</chem> | Nitroaromatic ketone – known for redox cycling and liver toxicity. |
| FP_69 | <chem>COC12C(COC(N)=O)C3=C(C(=O)C(C)=C(N)C3=O)N1CC1NC12</chem> | Complex fused heterocycles – structurally rich motifs often flagged in lead optimization for off-target effects. |

Visual representations of each substructure (highlighted in red) are provided in the supplementary materials (Figure S1–S10).

These fingerprint-based substructures, derived from the Morgan algorithm, do not directly convey semantic chemical features, but they consistently match recurring toxic motifs in chemical space. Their high importance values in the model emphasize their predictive power and their likely involvement in pharmacokinetic and toxicodynamic pathways.

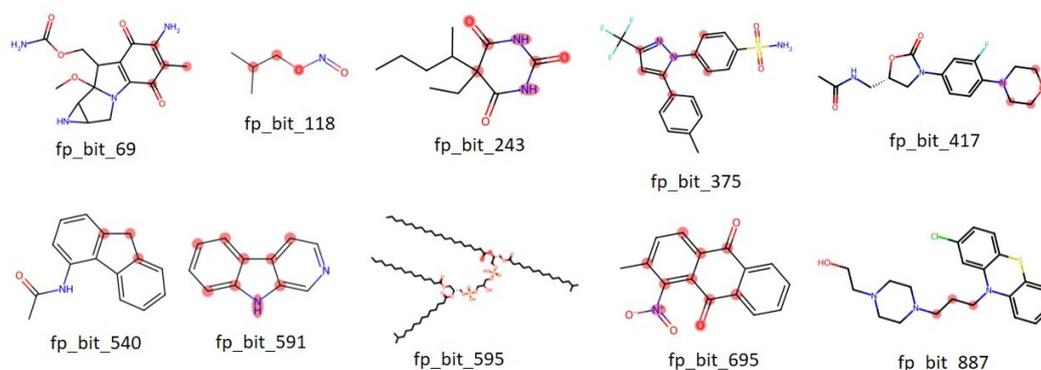


Figure 2: Top 10 Important Fingerprint Hits Identified as Relevant to Toxicity Prediction

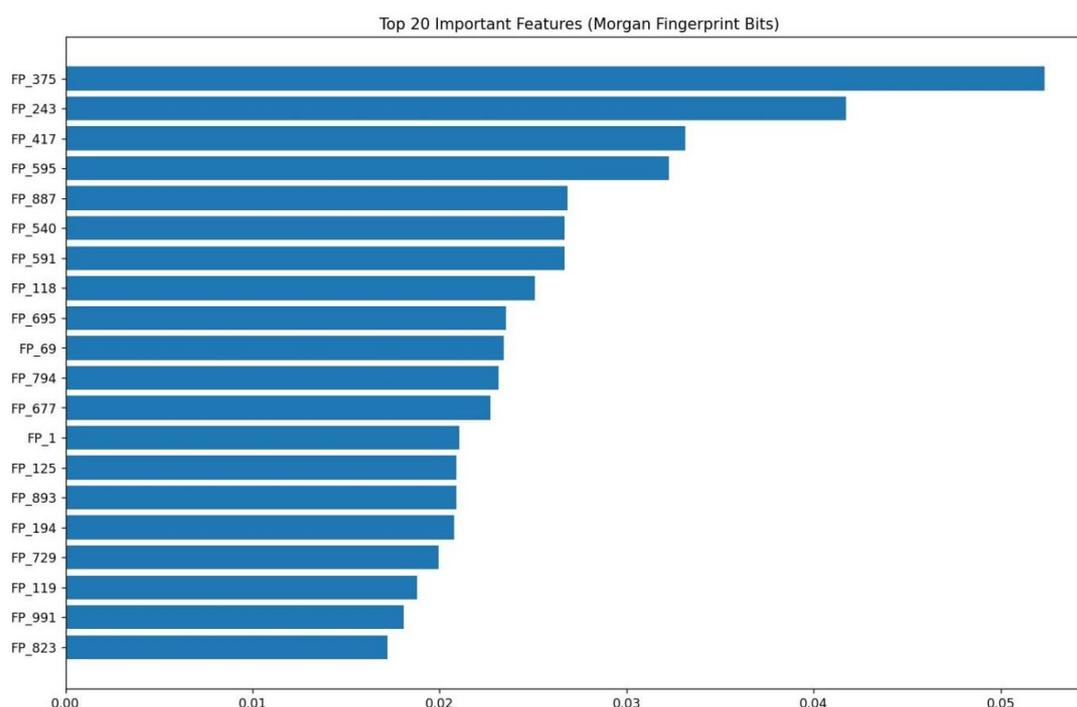


Figure 3: Top 20 Important Features (Morgan Fingerprint Bits)

Figure 3. This bar chart shows the relative importance of the top 20 Morgan fingerprint bits that contributed to the toxicity prediction model. Higher bars indicate more important features for the Random Forest model, suggesting these fingerprint bits are linked to key molecular characteristics that influence toxicity.

Feature Importance Analysis of Logistic Regression Model: - The feature importance analysis for the logistic regression model was performed by examining the absolute values of the model's coefficients. This provides insights into which features most significantly influence the toxicity predictions of organic compounds. The feature importance values were visualized in a bar chart, highlighting the

molecular descriptors that played key roles in distinguishing between toxic and non-toxic compounds.

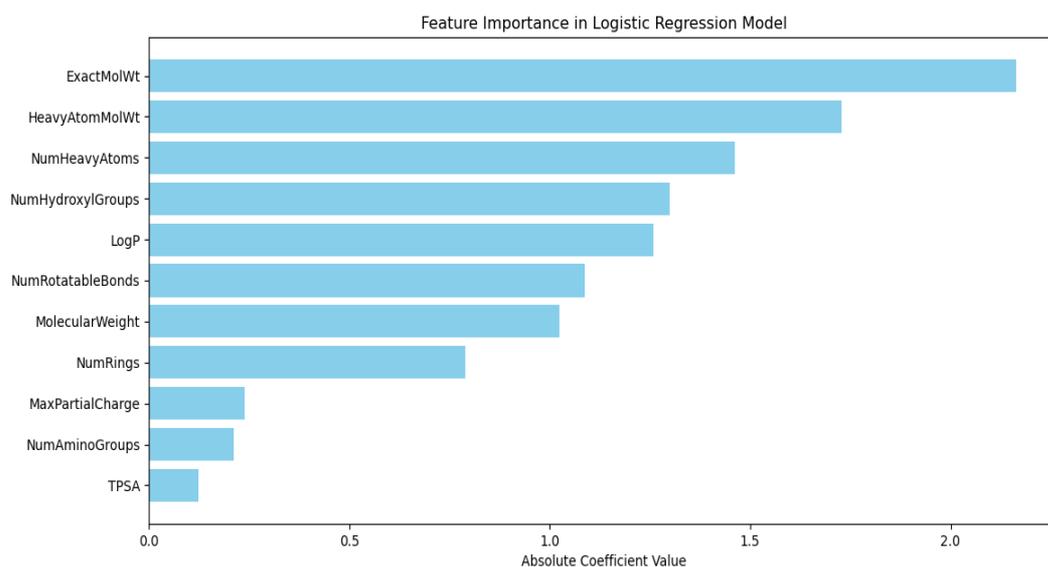


Figure 4: Feature Importance Analysis of Logistic Regression Model.

Figure 4 presents the feature importance analysis of the logistic regression model. As observed, the most significant features for toxicity prediction were ExactMolWt and HeavyAtomMolWt, which showed the highest coefficient values. Other important features include NumHeavyAtoms, NumHydroxylGroups, and LogP, suggesting that molecular weight and functional groups related to hydrophobicity and hydrogen bonding are crucial for toxicity prediction.

The bar chart in Figure 4 provides a clear visual representation of these features, where higher bars indicate more important features for the model's prediction of toxicity.

Feature Importance Analysis of SVM Model: - Feature importance analysis was conducted to better understand the role of each molecular descriptor in predicting toxicity using the Support Vector Machine (SVM) model. The analysis was performed using permutation importance, which evaluates the impact of each feature by measuring the decrease in model accuracy when a feature's values are shuffled.

The most important features for the SVM model, based on the mean decrease in accuracy, were found to be ExactMolWt and HeavyAtomMolWt, which had the highest importance values, followed by NumRotatableBonds, LogP, and NumHydroxylGroups. These results align with known toxicological principles, where molecular weight and lipophilicity (LogP) are key factors in predicting the bioactivity and toxicity of compounds.

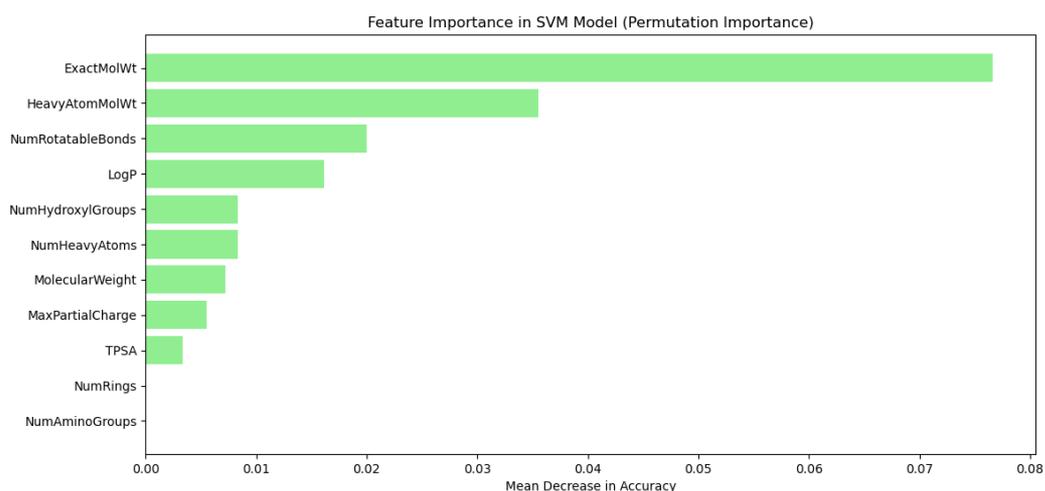


Figure 5: Feature Importance Ranking of SVM Model

The feature importance ranking is presented in Figure 5, which visually represents the relative importance of each feature in the model. As shown, ExactMolWt and HeavyAtomMolWt have a considerable influence on the toxicity predictions, suggesting that molecular weight-related descriptors are crucial in understanding compound toxicity. LogP and NumHydroxylGroups also play important roles, highlighting the relevance of hydrophobicity and functional group presence in predicting toxicity.

3.3 Validation with External Data

External Validation of Random Forest Model: - The Random Forest model was externally validated using 506 structurally diverse compounds from the ProTox-III validation set. SMILES strings were converted into 1024-bit Morgan fingerprints, consistent with the training pipeline. Experimental LD50 values were used post-prediction to evaluate performance against ProTox-defined toxicity classes.

The model achieved an accuracy of 77.1%, with particularly strong performance for identifying toxic compounds (Classes I–IV), achieving 96.2% recall and 79.0% precision. However, its precision for non-toxic compounds (Classes V–VI) was lower (37.5%), indicating a conservative bias toward predicting toxicity. These results suggest the Random Forest model effectively recognizes harmful structural motifs but may overpredict toxicity in less harmful or benign compounds.

External Validation of Logistic Regression Model: - The Logistic Regression model achieved a 100% recall for toxic compounds (Classes I–IV), correctly identifying all 396 toxic entries. However, it misclassified all 110 non-toxic compounds (Classes V–VI) as toxic, resulting in 0% precision for the non-toxic class. The overall accuracy was 78.3%, driven by the model's conservative bias toward predicting

toxicity. This high sensitivity may be advantageous in early hazard screening but underscores the need for improved specificity and balanced training to reduce false positives.

External Validation of Support Vector Machine (SVM) Model: - The SVM model achieved an overall accuracy of 78.3% in external validation. It successfully identified 100% of the 396 toxic compounds (Classes I–IV) with a recall of 1.00. However, it failed to identify any of the 110 non-toxic compounds (Classes V–VI), leading to 0% precision and recall for the non-toxic class. These results suggest the model has a high sensitivity for toxic compounds, favoring toxicity prediction across structurally varied chemical space. While this conservatism helps minimize false negatives, it also results in false positives among non-toxic chemicals. Future work may focus on improving class balance and calibration to enhance non-toxic compound identification.

3.4 Limitations

The study faced limitations related to the availability and quality of toxicity data for herbal compounds. Many herbal compounds lack comprehensive toxicity profiles, which restricted the size of the dataset. Additionally, the QSAR models were limited to predicting toxicity for individual compounds and did not account for the potential synergistic effects of multiple compounds in a herbal mixture.

4. Discussion and Conclusion

The development of machine learning-based Quantitative Structure-Activity Relationship (QSAR) models for predicting the toxicity of herbal and synthetic organic compounds marks a significant advancement in computational toxicology. Our study demonstrates that Logistic Regression, Random Forest, and Support Vector Machines (SVM) can achieve high accuracy (>90%) in predicting hepatotoxicity, neurotoxicity, and general acute toxicity. Random Forest slightly outperformed the others, with an accuracy of 92.78%, precision of 98.73%, recall of 86.67%, and F1 score of 92.31%. SVM exhibited perfect recall (100%) but lower precision (87%), making it particularly suitable for applications where missing toxic compounds is critical, even at the cost of some false positives.

Feature importance analysis provided valuable insights into the structural determinants of toxicity. For Random Forest, key features included molecular fingerprints corresponding to substructures such as aromatic rings, nitro groups, and tertiary amines, which are well-known for their association with toxic effects. Logistic Regression and SVM highlighted the significance of molecular weight and lipophilicity (logP), consistent with established toxicological principles. These findings not only validate the models' predictive capabilities but also offer

actionable insights for designing safer compounds by identifying structural features that contribute to toxicity.

When compared to recent advancements in the field, our models' performance is commendable. For instance, a study by (Romano et al., 2022) utilized graph neural networks (GNNs) with publicly aggregated semantic graph data, achieving a mean area under the receiver operating characteristic curve (AUROC) of 0.883 for toxicity prediction across 52 assays from the Tox21 dataset (Improving QSAR Modeling). While AUROC is a different metric from accuracy, the high performance of GNNs suggests that incorporating relational data between chemicals, genes, and assays can enhance prediction accuracy. Similarly, (Sharma et al., 2023) employed multi-task deep neural networks (MTDNN) with pre-trained SMILES embeddings, achieving an AUC-ROC of 0.991 for clinical toxicity prediction, demonstrating the potential of deep learning for complex endpoints (Accurate Clinical Toxicity). Our study's focus on both herbal and synthetic compounds addresses a gap in the literature, where herbal compounds are often underrepresented, thus bridging traditional herbal medicine with modern toxicology.

The high sensitivity of our models, particularly SVM, ensures that potentially toxic compounds, including those derived from herbal sources, are identified, enhancing public health safety. This is particularly relevant given the increasing use of herbal supplements and the need for robust safety profiling. However, our models exhibited a conservative bias in external validation, over-predicting toxicity, which could lead to false positives. This is a common challenge in toxicity prediction, potentially due to class imbalance in the datasets, where toxic compounds are less prevalent than non-toxic ones. The lower accuracy (77–78%) on the external ProTox-III dataset further highlights the need for more diverse training data to improve generalizability.

The practical implications of our findings are substantial. By providing early-stage toxicity screening, our models can reduce reliance on extensive animal testing, aligning with ethical and regulatory trends toward alternative methods. In pharmaceutical development, these models can prioritize compounds for further testing, optimizing resources and accelerating the identification of safe candidates. The inclusion of herbal compounds also supports the integration of traditional medicine into modern safety assessment frameworks, potentially informing regulatory guidelines for herbal products.

Despite these strengths, our study has limitations. The conservative bias in external validation suggests that class imbalance and dataset specificity may affect model performance. Additionally, the reliance on specific molecular descriptors and fingerprints may limit the models' applicability to novel chemical spaces. Future research should explore strategies to mitigate class imbalance, such as oversampling

minority classes or employing cost-sensitive learning techniques. Incorporating more diverse data sources, such as those aggregated in platforms like ComptoxAI, could enhance model robustness. Moreover, adopting advanced machine learning architectures, such as GNNs or deep neural networks(Mayr et al., 2016), may improve prediction accuracy and enable the modeling of more complex toxicity endpoints, including clinical toxicity.

In conclusion, this study underscores the efficacy of machine learning-based QSAR models in predicting the toxicity of a diverse range of compounds, including those from traditional herbal medicine. While opportunities for refinement remain, particularly in addressing class imbalance and enhancing generalizability, the current models offer a powerful tool for early-stage toxicity screening, supporting both scientific research and public health initiatives.

Future Directions

- a. Expansion to mixtures: Investigating toxicity in herbal mixtures rather than isolated compounds.
- b. Larger datasets: Collecting additional toxicity data for a wider range of herbal compounds.
- c. Model refinement: Experimenting with advanced machine learning techniques, such as neural networks(Wei et al., 2024), for more complex toxicological endpoints.

Acknowledgement:

The list of FDA-approved drugs used in this research was obtained from the U.S. Food and Drug Administration (FDA) website. The Human Metabolome Database (HMDB) was used for non-toxic metabolite data in this research. The database is freely available and must be cited as per the following reference: Wishart DS, Guo AC, Oler E, et al., HMDB 5.0: the Human Metabolome Database for 2022. *Nucleic Acids Res.* 2022. We thank the Toxic Exposome Database (T3DB) for providing the comprehensive data on toxic substances, which was essential for this study. We acknowledge the National Library of Medicine (NLM) for providing access to Carcinogenic Potency Database (CPDB) used in this study. This work utilized data from the NLM, 'Courtesy of the U.S. National Library of Medicine'.

The author gratefully acknowledge the use of the publicly available external validation dataset provided by ProTox-III (<https://tox.charite.de/prottox3/index.php?site=home#>). The validation set, described by the developers as a diverse subset of compounds spanning multiple toxicity classes, was used in this study to perform structure-based external validation of

predictive toxicity models. The resource significantly contributed to the evaluation of our models against real-world chemical diversity.

References:

- i. Ballabio, D., Grisoni, F., & Todeschini, R. (2018). Multivariate comparison of classification performance measures. *Chemometrics and Intelligent Laboratory Systems*, 174, 33–44. <https://doi.org/10.1016/j.chemolab.2017.12.004>
- ii. Banerjee, P., Kemmler, E., Dunkel, M., & Preissner, R. (2024). ProTox 3.0: A webserver for the prediction of toxicity of chemicals. *Nucleic Acids Research*, 52(W1), W513–W520. <https://doi.org/10.1093/nar/gkae303>
- iii. Cherkasov, A., Muratov, E. N., Fourches, D., Varnek, A., Baskin, I. I., Cronin, M., Dearden, J., Gramatica, P., Martin, Y. C., Todeschini, R., Consonni, V., Kuz'min, V. E., Cramer, R., Benigni, R., Yang, C., Rathman, J., Terfloth, L., Gasteiger, J., Richard, A., & Tropsha, A. (2014). QSAR Modeling: Where Have You Been? Where Are You Going To? *Journal of Medicinal Chemistry*, 57(12), 4977–5010. <https://doi.org/10.1021/jm4004285>
- iv. EBSCOhost (Ed.). (2023). *Qsar In Safety Evaluation And Risk Assessment*. Elsevier Academic Press.
- v. Ekins, S., Freundlich, J. S., Hobrath, J. V., Lucile White, E., & Reynolds, R. C. (2014). Combining Computational Methods for Hit to Lead Optimization in Mycobacterium Tuberculosis Drug Discovery. *Pharmaceutical Research*, 31(2), 414–435. <https://doi.org/10.1007/s11095-013-1172-7>
- vi. Fourches, D., Muratov, E., & Tropsha, A. (2016). Trust, but Verify II: A Practical Guide to Chemogenomics Data Curation. *Journal of Chemical Information and Modeling*, 56(7), 1243–1252. <https://doi.org/10.1021/acs.jcim.6b00129>
- vii. Khaouane, A., Ferhat, S., & Hanini, S. (2023). A Quantitative Structure-Activity Relationship for Human Plasma Protein Binding: Prediction, Validation and Applicability Domain. *Advanced Pharmaceutical Bulletin*, 13(4), 784–791. <https://doi.org/10.34172/apb.2023.078>
- viii. Krewski, D., Acosta, D., Andersen, M., Anderson, H., Bailar, J. C., Boekelheide, K., Brent, R., Charnley, G., Cheung, V. G., Green, S., Kelsey, K. T., Kerkvliet, N. I., Li, A. A., McCray, L., Meyer, O., Patterson, R. D., Pennie, W., Scala, R. A., Solomon, G. M., ... Staff Of Committee On Toxicity Test. (2010). Toxicity Testing in the 21st Century: A Vision and a

- Strategy. *Journal of Toxicology and Environmental Health, Part B*, 13(2–4), 51–138. <https://doi.org/10.1080/10937404.2010.483176>
- ix. Lo, Y.-C., Rensi, S. E., Torng, W., & Altman, R. B. (2018). Machine learning in chemoinformatics and drug discovery. *Drug Discovery Today*, 23(8), 1538–1546. <https://doi.org/10.1016/j.drudis.2018.05.010>
- x. Machhar, J., Mittal, A., Agrawal, S., Pethe, A. M., & Kharkar, P. S. (2019). Computational prediction of toxicity of small organic molecules: State-of-the-art. *Physical Sciences Reviews*, 4(10). <https://doi.org/10.1515/psr-2019-0009>
- xi. Mayr, A., Klambauer, G., Unterthiner, T., & Hochreiter, S. (2016). DeepTox: Toxicity Prediction using Deep Learning. *Frontiers in Environmental Science*, 3. <https://doi.org/10.3389/fenvs.2015.00080>
- xii. Moreira, D. D. L., Teixeira, S. S., Monteiro, M. H. D., De-Oliveira, A. C. A. X., & Paumgarten, F. J. R. (2014). Traditional use and safety of herbal medicines1. *Revista Brasileira de Farmacognosia*, 24(2), 248–257. <https://doi.org/10.1016/j.bjp.2014.03.006>
- xiii. Rogers, D., & Hahn, M. (2010). Extended-Connectivity Fingerprints. *Journal of Chemical Information and Modeling*, 50(5), 742–754. <https://doi.org/10.1021/ci100050t>
- xiv. Romano, J. D., Hao, Y., & Moore, J. H. (2022). Improving QSAR modeling for predictive toxicology using publicly aggregated semantic graph data and graph neural networks. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 27, 187.
- xv. Sen, S., Chakraborty, R., & De, B. (2011). Challenges and opportunities in the advancement of herbal medicine: India's position and role in a global context. *Journal of Herbal Medicine*, 1(3–4), 67–75. <https://doi.org/10.1016/j.hermed.2011.11.001>
- xvi. Sharma, B., Chenthamarakshan, V., Dhurandhar, A., Pereira, S., Hendler, J. A., Dordick, J. S., & Das, P. (2023). Accurate clinical toxicity prediction using multi-task deep neural nets and contrastive molecular explanations. *Scientific Reports*, 13(1), 4908. <https://doi.org/10.1038/s41598-023-31169-8>
- xvii. U.S. Food and Drug Administration. (2024). *Drugs@FDA: FDA-approved drugs*. <https://www.fda.gov/drugs/drug-approvals-and-databases/drugsfda-data-files>
- xviii. Varsou, D.-D., Kolokathis, P. D., Antoniou, M., Sidiropoulos, N. K., Tsoumanis, A., Papadiamantis, A. G., Melagraki, G., Lynch, I., & Afantitis,

- A. (2024). In silico assessment of nanoparticle toxicity powered by the Enalos Cloud Platform: Integrating automated machine learning and synthetic data for enhanced nanosafety evaluation. *Computational and Structural Biotechnology Journal*, 25, 47–60. <https://doi.org/10.1016/j.csbj.2024.03.020>
- xix. Wei, J., Tian, L., Nie, F., Shao, Z., Wang, Z., Xu, Y., & He, M. (2024). Quantitative structure-activity relationship model development for estimating the predicted No-effect concentration of petroleum hydrocarbon and derivatives in the ecological risk assessment. *Heliyon*, 10(5), e26808. <https://doi.org/10.1016/j.heliyon.2024.e26808>
- xx. Wishart, D., Arndt, D., Pon, A., Sajed, T., Guo, A. C., Djoumbou, Y., Knox, C., Wilson, M., Liang, Y., Grant, J., Liu, Y., Goldansaz, S. A., & Rappaport, S. M. (2015). T3DB: The toxic exposome database. *Nucleic Acids Research*, 43(Database issue), D928-934. <https://doi.org/10.1093/nar/gku1004>
- xxi. Wishart, D. S., Guo, A., Oler, E., Wang, F., Anjum, A., Peters, H., Dizon, R., Sayeeda, Z., Tian, S., Lee, B. L., Berjanskii, M., Mah, R., Yamamoto, M., Jovel, J., Torres-Calzada, C., Hiebert-Giesbrecht, M., Lui, V. W., Varshavi, D., Varshavi, D., ... Gautam, V. (2022). HMDB 5.0: The Human Metabolome Database for 2022. *Nucleic Acids Research*, 50(D1), D622–D631. <https://doi.org/10.1093/nar/gkab1062>
- xxii. Woo, C. S. J., Lau, J. S. H., & El-Nezami, H. (2012). Herbal Medicine. In *Advances in Botanical Research* (Vol. 62, pp. 365–384). Elsevier. <https://doi.org/10.1016/B978-0-12-394591-4.00009-X>
- xxiii. Xu, Y.-Q., Huang, P., Li, X.-W., Liu, S.-S., & Lu, B.-Q. (2024). Derivation of water quality criteria for paraquat, bisphenol A and carbamazepine using quantitative structure-activity relationship and species sensitivity distribution (QSAR-SSD). *Science of The Total Environment*, 948, 174739. <https://doi.org/10.1016/j.scitotenv.2024.174739>
- xxiv. Yang, H., Lou, C., Sun, L., Li, J., Cai, Y., Wang, Z., Li, W., Liu, G., & Tang, Y. (2019). admetSAR 2.0: Web-service for prediction and optimization of chemical ADMET properties. *Bioinformatics*, 35(6), 1067–1069. <https://doi.org/10.1093/bioinformatics/bty707>