Machine Learning for Rodent Liver Toxicity Prediction: Leveraging Drug Substructure Properties

Akash Parekh, Omkar Nanekar, Nilanjan Chattopadhyay, Nitin Singhal
AIRAMATRIX, Mumbai, India

EUROTOX 2024 COPENHAGEN | DENMARK 8-11 SEPTEMBER 2024

Objective

To develop robust predictive models to assess liver toxicity in rodents (rats and mice) during the drug development process. This objective emphasizes the importance of accurate models for:

- **Safety Assessment**: Early identification of potential liver toxicity risks
- **Regulatory Compliance**: Meeting regulatory requirements for drug development
- **Cost and Time Savings**: Avoiding unnecessary animal testing and delays in drug development

Background

Drug-induced liver toxicity poses significant hurdle in the process of drug development, given its complex nature and potential consequences on safety and regulatory compliance. Traditional methods are often resource-intensive and may lack the accuracy for early-stage evaluations.

Predictive models utilizing machine learning algorithms, which are built using **structural and chemical** data, offer enhanced accuracy and efficiency in assessing toxicity. This has the potential to reduce costs and faster time to market of novel pharmaceuticals.

Experimental Design

Interaction with the target, metabolism, and other related elements. Therefore, it is necessary to employ a combination of experimental assays, computational approaches, and expert knowledge in order to accurately determine the hazardous substructures of a compound. The following describes our current strategy:

Structure-Activity Relationship (SAR) Analysis:

Analyse the substructures shared by toxic molecules that may not exist in non-toxic molecules.Examine toxic molecules for patterns in which specific functional groups, aromatic rings, or molecular fragments are consistently present.

Fragment-Based Methods:

Assess the toxic potential of each fragment by dividing the molecule into fragments or substructures. Fragment-based toxicity prediction using predefined toxic fragments.

Machine Learning Interpretation Techniques:

Feature importance analysis to identify features (substructures).

PredTox, Automated Machine Learning (AutoML) platform:

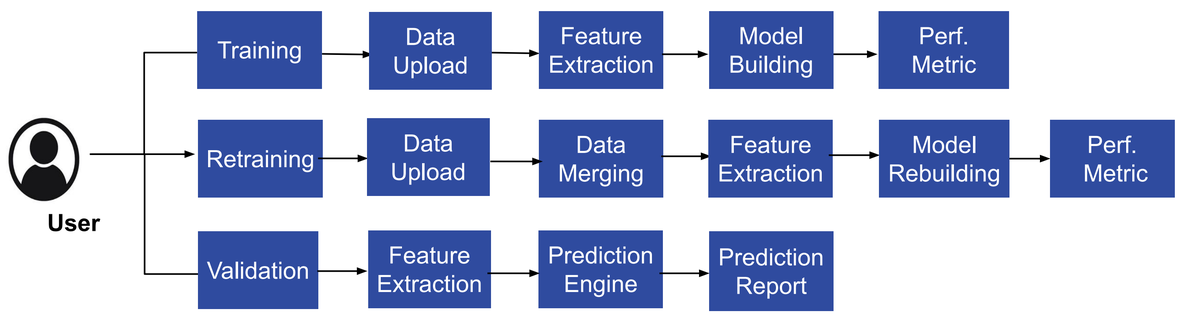
An AutoML application has been used for doing all the above analysis and it automates the construction and optimization of predictive models. By outsourcing tasks like model selection, training, and optimization, PredTox enhances efficiency and accuracy in toxicity prediction.

Experimental Setup

- Data preparation for model training is done using Open Tg-Gates [1.] and LiverTox data [2.] Five-fold cross-validation is conducted to ensure model robustness
- External validation done on and compounds from a global pharmaceutical company (GPC). and all toxic compounds from open source NTP (National Toxicology Program) [3.] data.

Dataset Name	Type	Compound	Non-Toxic	Toxic	Event rate
Training	Model Training	520	273	247	47.50%
Testing (Blind)	Internal Validation	130	68	62	47.69%
GPC	External Validation	27	68	62	47.69%
NTP	External Validation	181	0	181	100.00%

PredTox Architecture

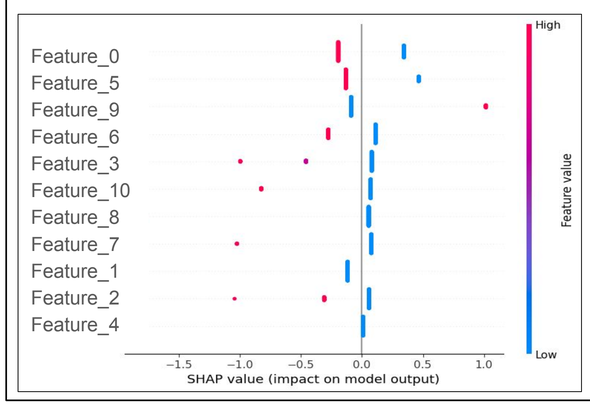


Results

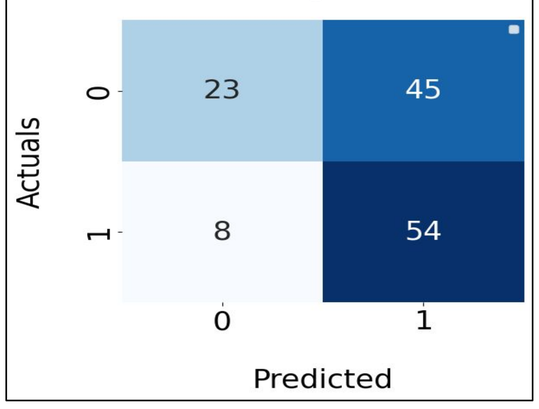
Dataset	AUC Mean	AUC Standard Deviation
Training	0.7272	+/- 0.0143
Validation	0.7081	+/- 0.0221
Testing	0.7155	+/- 0.8015
GPC	0.7056	-
NTP	-	-

Dataset	Precision	Recall	F1-Score	Accuracy
Testing	54.55%	87.10%	67.08%	59.23%
GPC	75.00%	80.00%	77.42%	74.07%
NTP (All Toxic)	100.00%	81.22%	89.63%	81.22%

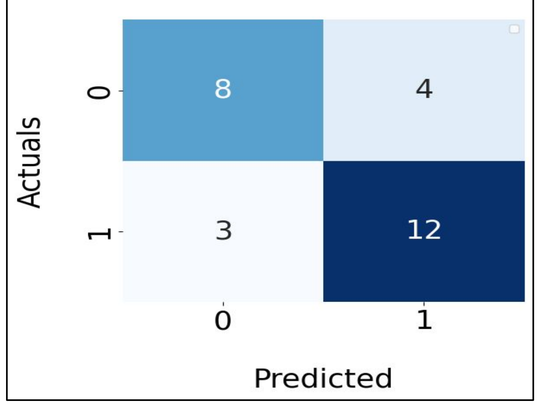
Feature Importance & Target Relation



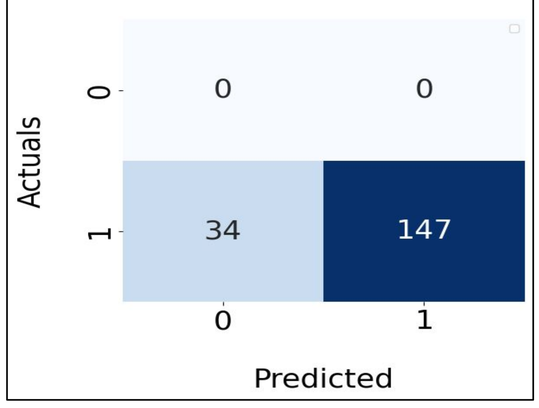
Testing



GPC



NTP



Conclusion

- The machine learning models developed in this study show promising potential for predicting rodent liver toxicity, achieving an AUC of approximately 0.72 across training, validation, and testing datasets, highlighting the effectiveness of leveraging substructure properties.
- The integration of Structure-Activity Relationship (SAR) analysis and fragment-based methods within the PredTox AutoML platform enables efficient identification of hazardous substructures, providing a reliable approach to early-stage toxicity assessment.
- Strong performance in the GPC and NTP datasets indicates potential to reduce traditional animal testing, lowering costs and speeding up drug development.

References

[1.] Yoshinobu Igarashi, Noriyuki Nakatsu, Tomoya Yamashita, Atsushi Ono, Yasuo Ohno, Tetsuro Urushidani and Hiroshi Yamada (2014) Open TG-GATEs: a large-scale toxicogenomics database. Nucleic Acids Research [online]. 43, D921-D927

[2.] Hoofnagle, J.H.; Serrano, J.; Knoben, J.E.; Navarro, V.J. LiverTox: A website on Drug-Induced Liver Injury; Wiley Online Library: Hoboken, NJ, USA, 2013.

[3.] Krewski D, Acosta D Jr, Andersen M, Anderson H, Bailar JC 3rd, Boekelheide K, Brent R, Charnley G, Cheung VG, Green S Jr, Kelsey KT, Kerkvliet NI, Li AA, McCray L, Meyer O, Patterson RD, Pennie W, Scala RA, Solomon GM, Stephens M, Yager J, Zeise L. Toxicity testing in the 21st century: a vision and a strategy. J Toxicol Environ Health B Crit Rev. 2010 Feb;13(2-4):51-138. doi: 10.1080/10937404.2010.483176. PMID: 20574894; PMCID: PMC4410863.

[4.] RDKit: Open-source cheminformatics. https://www.rdkit.org

solutions@airamatrix.com
www.airamatrix.com