2nd INTERNATIONAL CONFERENCE ON
Artificial Intelligence
31 May - 1 June 2026

IICAI
Second International Interdisciplinary Conference on Artificial Intelligence

Predicting Plasma Protein Binding (Fraction Unbound) with Machine Learning Using Molecular Descriptors

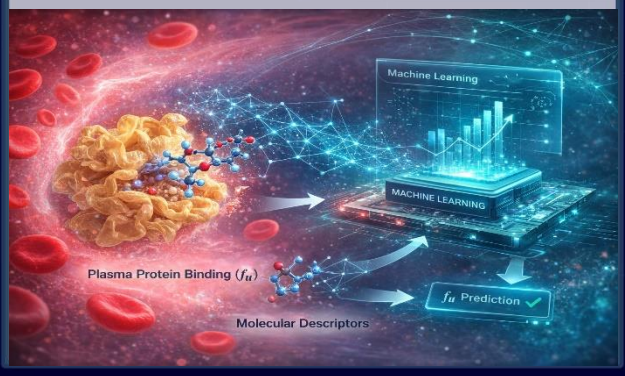
Code: iai-1046

Arash Maghsoudlou[1], Fatemeh Ghorbani-Bidkorpeh*[2], Madjid Soltani[3]

1-PharmD Student, School of Pharmacy, Shahid Beheshti University of Medical Sciences, Tehran, Iran, arash_maghsoudlou@sbmu.ac.ir
2-Associate Professor, PhD of Pharmaceutics, School of Pharmacy, Shahid Beheshti University of Medical Sciences, Tehran, Iran, f.ghorbani@sbmu.ac.ir
3-Adjunct Assistant Professor, Department of Electrical & Computer Engineering, University of Waterloo, Waterloo, Ontario, Canada, msoltani@uwaterloo.ca

Abstract

fraction unbound is a key pharmacokinetic parameter affecting drug distribution and efficacy. Because experimental measurement is costly and time-consuming, this study applied machine learning to predict fu using physicochemical descriptors and molecular fingerprints from 647 compounds. Gradient Boosting showed the best performance (R^2 = 0.61, RMSE = 0.19). Feature analysis identified hydrogen bonding capacity, lipophilicity, and polarity as major determinants. These findings highlight the value of ML-based approaches in early drug discovery.



Research method

Experimental fu data were obtained from the Krumpholz dataset. After removing duplicates, 647 unique compounds were used for analysis. Physicochemical descriptors and molecular fingerprints (Morgan, RDK, MACCS) were generated, and dimensionality reduction was applied using Truncated SVD. Multiple machine learning algorithms, including Gradient Boosting and Random Forest, were trained using an 80/20 train–test split with 5-fold cross-validation. Model performance was evaluated using R^2, RMSE, and MAE.

conclusion

This study demonstrated that the fraction unbound in plasma (fu) can be predicted with reasonable accuracy using machine learning models based on physicochemical descriptors and molecular fingerprints. Among the evaluated algorithms, Gradient Boosting showed the most consistent and reliable performance. The analysis confirmed that hydrogen bonding capacity, lipophilicity, and polarity-related properties play a major role in determining plasma protein binding. These findings align with established pharmacokinetic principles and highlight the capability of machine learning to model complex structure–property relationships. Overall, the proposed approach offers a cost-effective and efficient in-silico tool that can support early-stage drug discovery and pharmacokinetic screening.

The purpose of the research

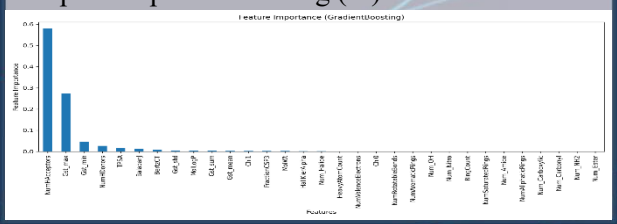
The purpose of this study was to develop a reliable and interpretable machine learning model for predicting the fraction unbound in plasma (fu), a key pharmacokinetic parameter that influences drug distribution, clearance, and therapeutic efficacy. Because experimental measurement of plasma protein binding is costly and time-consuming, there is a strong need for efficient computational prediction tools in early drug discovery.This research evaluated multiple machine learning algorithms using physicochemical descriptors and truncated molecular fingerprints to model the nonlinear relationship between molecular structure and fu. Additionally, the study aimed to identify the most influential molecular properties—such as hydrogen bonding capacity, lipophilicity, and polarity—that govern plasma protein binding.

Practical or simulation results

The predictive performance of the developed machine learning models was evaluated using standard regression metrics. Gradient Boosting achieved the best results in estimating plasma fraction unbound (fu), outperforming other algorithms. The main performance metrics are summarized below:

Model	R^2 Test	RMSE Test	MAE Test
GB	0.613052	0.187547	0.146494
RF	0.530220	0.206648	0.159843
SVR	0.297104	0.252772	0.197067
MLP	-0.17971	0.327470	0.250847
Ridge Regression	-0.38748	0.355137	0.282893

Feature importance analysis revealed that hydrogen bonding capacity and polarity-related descriptors are the primary determinants of plasma protein binding (fu).



References

[1]A. Talevi and C. L. Bellera, "Drug Binding to Plasma Proteins," in The ADME Encyclopedia: A Comprehensive Guide on Biopharmacy and Pharmacokinetics, A. Talevi Ed. Cham: Springer International Publishing, 2022, pp. 315–326.

[2]R. Watanabe et al., "Predicting Fraction Unbound in Human Plasma from Chemical Structure: Improved Accuracy in the Low Value Ranges," Molecular Pharmaceutics, vol. 15, no. 11, pp. 5302–5311, 2018/11/05/ 2018, doi: https://doi.org/10.1021/acs.molpharmaceut.8b00785.

[3]C. W. Lindsley, "Lipophilicity," in Encyclopedia of Psychopharmacology, I. P. Stolerman and L. H. Price Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 1–6.

[4]S. Pore and K. Roy, "Insights into pharmacokinetic properties for exposure chemicals: predictive modelling of human plasma fraction unbound (fu) and hepatocyte intrinsic clearance (Clint) data using machine learning," Digital Discovery, 10.1039/D4DD00082J vol. 3, no. 9, pp. 1852–1877, 2024, doi: 10.1039/D4DD00082J.

[5]L. Sun et al., "In Silico Prediction of Compounds Binding to Human Plasma Proteins by QSAR Models," ChemMedChem, vol. 13, no. 6, pp. 572–581, 2018/03/20 2018, doi: https://doi.org/10.1002/cmdc.201700582.

[6]L. Krumpholz, A. Klimczyk, W. Bieniek, S. Polak, and B. Wiśniowska, "Data set of fraction unbound values in the in vitro incubations for metabolic studies for better prediction of human clearance," Database, vol. 2024, p. baae063, 2024, doi: 10.1093/database/baae063.