

DrugXplorer: An AI – Assisted Open-Source Framework for Drug Analysis and Screening

Rishi Senthil Kumar^{1,2†}, Jagadesshwar Muthukumaran^{1,2†},
Ashwin Sivakumar^{1,2†}, Sheena Christabel Pravin^{1,2},
Reena Monica P^{1,2*}

^{1*}School of Electronics Engineering, Vellore Institute of Technology,
Chennai, India.

²MedxAI Innovations, Vellore Institute of Technology, Chennai, India.

*Corresponding author(s). E-mail(s): reenamonica@vit.ac.in;
Contributing authors: rishi.senthilkumar2021@vitstudent.ac.in;
jagadesshwar.muthukumaran2021@vitstudent.ac.in;
ashwin.sivakumar2021@vitstudent.ac.in;
sheenachristabel.p@vit.ac.in.com;

†These authors contributed equally to this work.

Abstract

Researchers rely on in-silico analysis to screen potential drug candidates for various diseases. While multiple software and web applications exist for tasks such as Absorption, Distribution, Metabolism, Excretion (ADME) profiling, molecular docking, and molecular dynamics, using various tools can be time-consuming and challenging. Additionally, there is no easily accessible in-silico method for predicting drug synergy across different cell lines. To address these challenges, we developed a web application that integrates essential in-silico analyses, making drug discovery more efficient and accessible. A majority of the ADME properties were predicted using the RDKit package, while cytochrome P450 (CYP450) inhibition was predicted using five deep-learning classification models trained to predict inhibitory activity towards CYP1A2, CYP2C9, CYP2C19, CYP2D6, and CYP3A4. The accuracy of the DrugXplorer model was on par with benchmarked models such as MumCYP and others and outperforms in the prediction of CYP2C19 enzyme with an accuracy of 0.84. We developed a regression model using Convolutional Neural Networks (CNN) and trained it on 24 different protein datasets to predict binding affinity and inhibition constant (Kd). On average, the model achieved an R - squared (R^2) value of 0.68 across the 24 datasets. We also built another regression model to predict bliss synergy score across eight

cancer cell lines, achieving an average R^2 value of 0.95. Finally, we use an Agentic - AI framework to provide useful inferences and context to the predictive results with the help of Large Language Models and CrewAI packages. All of these processes are combined in a web application with an easy-to-use user interface. This would help any drug discovery researcher to seamlessly perform preliminary drug analysis and screening of the target of their choice.

Keywords: In-silico analysis, Cytochrome P450, Convolutional Neural Networks, Binding Affinity, Bliss Synergy Score, Large Language Models

1 Introduction

The accessibility of data and user-friendly software for drug discovery and analysis remains limited. Despite significant advances in research and development, progress is hindered by high costs and the absence of predictive models [1]. Machine learning (ML) has emerged as a promising solution to address these challenges by enabling faster and more accurate predictive modeling [2].

There is also a noticeable lack of accessibility to current ML models. The high learning curve required to use and edit such models is to be considered. There is a dire requirement for a user-friendly user interface (UI) for drug discovery and predictive analysis applications [3]. The need for comprehensive predictive models comes not only from the goal of experimentally reducing expenses, but also from providing a promising alternative to conventional chemical simulation software, which is expensive in computational resources and time [4].

1.1 Previous Works

Conventional simulation software, such as AutoDock Vina [5, 6], PyMOL [7], consists of a python wrapper to complement code-friendly usage. However, a researcher may have to use multiple tools for a comprehensive single drug analysis. No - Code or Low - Code interfaces are helpful in such cases to reduce user workload. DOCKSTRING [8] and PandaDock[9] are such python packages which provides a reliable ligand and target preparation protocol, enabling non-experts to generate meaningful docking scores. On general chemical analysis, packages such as RDKit [10] and Deepchem [11] have critical tools for simulations as well as machine learning models. Analysis pipelines and generative Artificial Intelligence (AI) models have been developed using these packages, as demonstrated by [Bento et al.](#), [Cao and Kipf](#), [Chithrananda et al.](#), and others.

The availability of credible and useful experimental data is of importance in order to create ML models with more relevant predictions. A study by [Landrum and Riniker](#) emphasizes the importance of high-quality datasets by analyzing noise in combined datasets. The study finds that when the same compound is tested across multiple assays for the same target, minimal curation of IC50 assays leads to poor agreement, underscoring the need for rigorous data curation.

A useful guide by [Özçelik and Grisoni](#), explored various strategies for chemical language processing: Embedding Techniques, Encoding Methods, ML Algorithms and Benchmark Datasets. This paper uses CNN, Recurrent Neural Network (RNN), Transformers with input from different encoding strategies for predictive modelling. Transformers offer a promising avenue of future work for molecular representation learning and property prediction. ChemBERTa [14] follows this approach for molecular property predictions with Pre - Trained Large Language Models (LLMs) using transformer encoding methods. Other studies have also implemented Generative Pre - Trained Transformers (GPTs) such as ChatGPT, Llama focusing on molecular property predictions, analysis of drug - drug interactions, material characteristics and reaction yields [17, 18]

Web applications with easy to use UI have been created for multiple drug discovery applications. DrugComb [19] is one such application used for drug synergy analysis as well as predictions. ADMET 3.0 [20] and SwissADME [21] is widely used for ADME analysis. FPocketWeb [22] is a newly created web application derived from the tools of FPocket[23] which is essential for optimal pocket detection sites in a protein for a efficient molecular docking.

1.2 Our Approach

With the increasing complexity of in-silico methods and their high computational demands, machine learning models present a promising alternative. However, researchers often spend valuable time re-running previously executed code, emphasizing the need for a no-code solution to streamline molecular property predictions. The goal is to develop an application for a researcher in the field of drug discovery. We intend to create a framework to order to perform the following tasks:

1. Extract preliminary details using Chemical Language Processing: e.g., Drug-likeness, Components of Extracts, etc.
2. Automate the process for running predictive models for a required molecular property prediction based on the prompt input.
3. Provide an AI-assisted inference paragraph to provide context and validation to the predicted property.

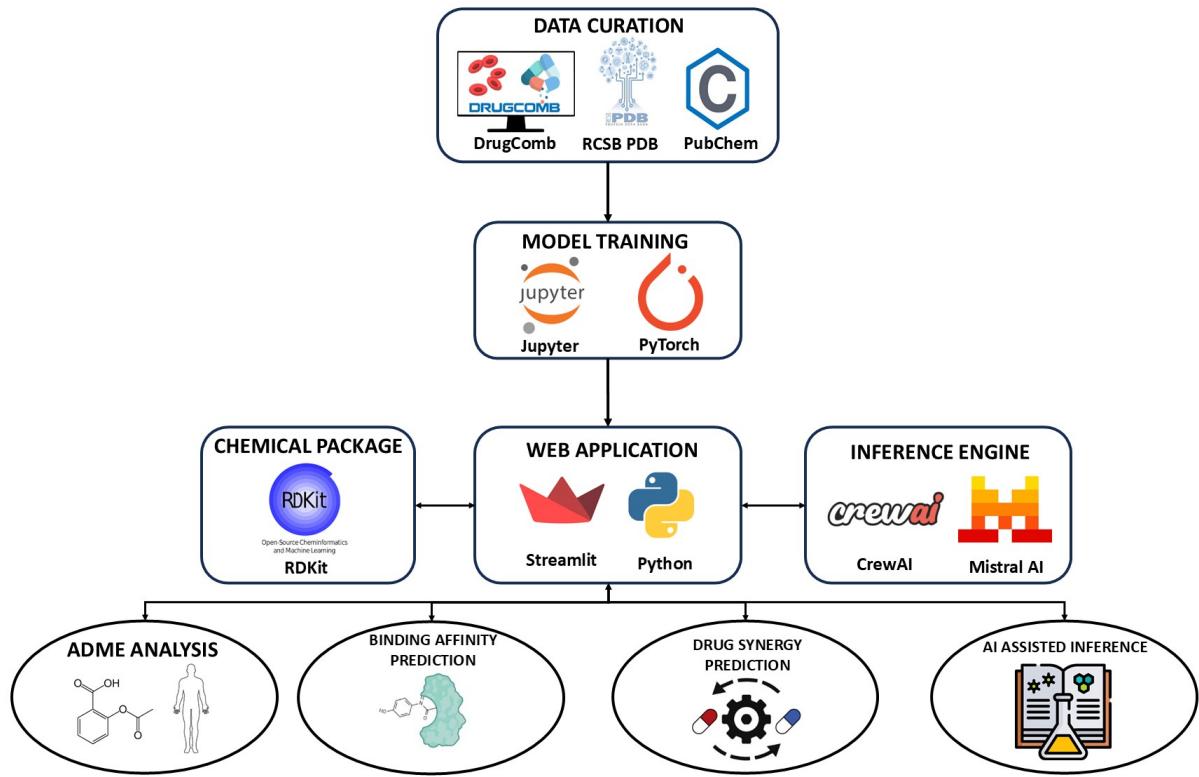


Fig. 1 Architecture of DrugXpler framework and web application

Figure 1 is an architecture diagram of the application’s framework. DrugXpler focuses on three main drug properties: ADME analysis, Binding Affinity Prediction and Drug Synergy Prediction. In the upcoming sections of the paper, we describe the methodology used for our predictive models as well as the features of the DrugXpler web application.

2 Methodology

2.1 Selections of Target Proteins and Cell Lines

To train our deep learning model to predict binding affinity using SMILES representation, we used data from the Docstring database [8]. The Docstring database contains data for over 50 protein receptors, from which we selected 24 to improve computational efficiency while maintaining a diverse set of receptors to cover important protein groups.

Nuclear receptors are prominent therapeutic targets for drug researchers, so we selected five nuclear receptors: PPARD, PPARG, AR, ESR1, and NR3C1 [24]. Kinases play a crucial role in cell signaling pathways, making them key targets for inflammatory

disease treatments. To capture these pathways, we included six kinase proteins: ABL1, JAK2, AKT1, MAPK1, PLK1, and EGFR1 [25–28]. Metabolic enzymes are essential in research involving drug metabolism and disease progression, so we incorporated PTGS2, HMGCR, CYP3A4, and DPP4 [29–32].

Neurotransmitter receptors and neurological targets are widely studied in neurodegenerative diseases and psychiatric disorders. To cover this domain, we selected ADRB1, ADORA2A, DRD2, ACHE, and BACE1, which play key roles in cardiovascular function, dopamine signaling, cholinergic pathways, and amyloid plaque formation [33–37]. Lastly, cancer remains a major focus in drug discovery, so we included CASP3, PARP1, ROCK1, and KDR, which are critical for apoptosis regulation, DNA repair, and angiogenesis [38–41]. By selecting these receptors, we ensured a broad representation of key therapeutic targets across multiple biological processes. The summary of graphical visualizations of our curated proteins for binding affinity predictions is given in figure 2

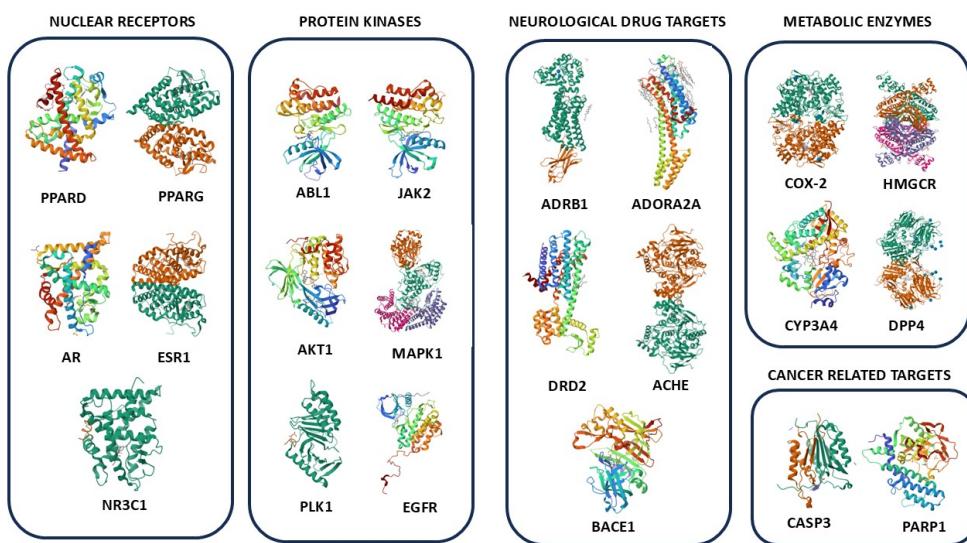


Fig. 2 Graphical Summary of 24 protein targets curated for binding affinity predictions

Drug synergy analysis is crucial in identifying effective combination therapies, particularly in oncology, where resistance to monotherapies often necessitates combination strategies. We wanted to train multiple models on various cell lines for our synergy prediction model. We utilized data from the Drugcomb database to train our deep learning models [19].

We selected eight well-characterized cancer cell lines, each representing different tissue origins and cancer types. These include MCF7 (breast cancer), 786-O (renal cell carcinoma), A549 (lung cancer), DU145 (prostate cancer), HCT116 (colorectal cancer), K562 (chronic myelogenous leukemia), OVCAR3 (ovarian cancer), and SNB75

(glioblastoma). Each of these cell lines has been widely used in drug screening studies due to their distinct genetic profiles and relevance in preclinical research. All selected cell lines are part of the NCI-60 cancer cell line panel, a widely recognized set of human cancer cell lines used by the National Cancer Institute (NCI) for anticancer drug screening. The NCI-60 panel represents a diverse range of cancer types and has been instrumental in evaluating drug responses, identifying biomarkers, and advancing precision oncology research [42].

2.2 Data Preparation

2.2.1 Simplified Molecular Input Line Entry System (SMILES)

SMILES is a linear notation used to represent molecular structures as text strings. Due to its machine-readable format, SMILES [43] has become a standard input representation in cheminformatics, for tasks like molecular property prediction, drug database searches, and ADME analysis. The basic components of SMILES are atoms (which are represented by chemical symbols), bonds, rings, branches and stereochemical combinations, that are involved in the molecule.

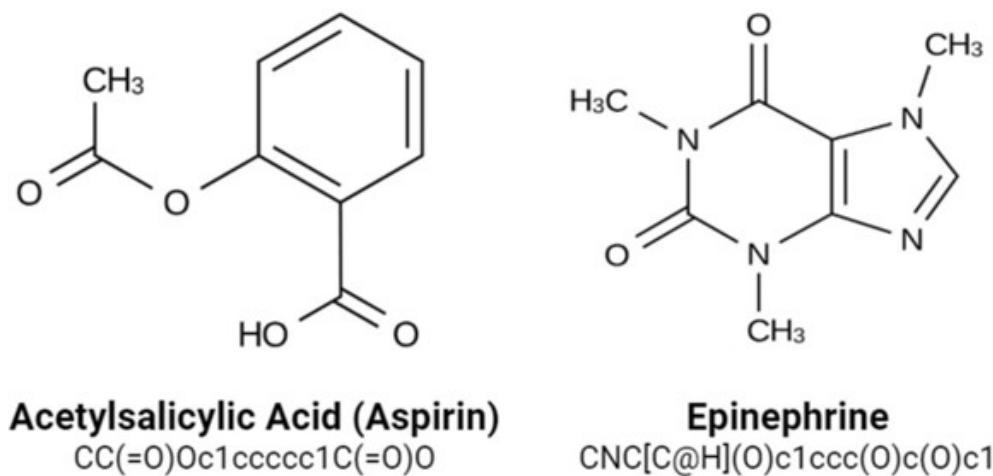


Fig. 3 Chemical representation of Aspirin and Epinephrine along with their SMILES representation.

2.2.2 Morgan Fingerprinting

Morgan Fingerprinting, [44] is one of the most popular algorithms which generate fingerprints of variable length to engineer data from drug molecules. The libraries used to implement this algorithm are Pandas, NumPy and RDKit [10]. The SMILES notation of a drug molecule is passed into the Morgan Fingerprinting built-in function, which is imported from the Chem module of the RDKit Library [10]. The function proceeds to assign an initial identifier to all the atoms of the molecule, based on its

atomic properties. It is transformed to a RDKit molecular object, which is passed through the fingerprint algorithm. The algorithm takes in parameters that include the radius, to define the degree of the neighbourhood bond expansion, and the number of bits to be generated, which are only binary digits of 0s and 1s. For the binding affinity dataset, we have transformed the SMILES notation into 150-bit length fingerprint with radius specified as 6. The 1 in a particular bit position represent a particular substructure which extends to 6 atoms in length and 0 represents the absence of that particular substructure. For every drug molecule's Morgan fingerprint [44], every bit is split into a feature vector. We have created 150 columns with the fingerprints of each drug molecule and concatenated the created fingerprint data into the existing dataset. Similarly for drug synergy, we have performed the same steps, where two 150-bit fingerprints were generated from the SMILES of drug1 and drug2 and then, the concatenated vectors of 300 bits were added into the dataset for 8 different cell-lines.

2.2.3 Data Cleaning

The target feature of the CYP datasets, comprises of three different classes, which are namely Active, Inactive and Inconclusive. We have removed the data points that have Inconclusive as the target value, to further improve the quality of the dataset.

We have applied Median Absolute Deviation (MAD) on the target column of both the binding affinity and the synergy datasets, and the threshold value had to be modified according to the spread of the target column in each dataset. MAD is a statistical measure of dispersion that quantifies the spread of the dataset. Due to its robustness, MAD is less sensitive to outliers, and provide a better estimate of scale for skewed and heavy-tailed distributions. The MAD function removes the outliers of the data, according to the threshold value provided.

2.3 Predictive Modeling using Machine Learning

2.3.1 Classification Model for CYP using Neural Networks

We have mapped the target classes from Active and Inactive to 0 and 1, respectively. The dataset was then, fit into feed-forward neural network, implemented using PyTorch [45]. The model takes in 150 columns as the input dimension and number of classes as 2. We have used 4 Dense hidden layers of 1024, 512, 256, 128 nodes each for the model. The ReLU activation function was used after each hidden layer to introduce non-linearity [46]. We ran the model on 100 epochs, with learning rate as 0.001 using Adam optimizer and CrossEntropyLoss acting as the loss function.

2.3.2 Prediction of Binding Affinity And Drug Synergy using Neural Networks

The dataset for binding affinity was fed into the CNN-based regression model, where the input dimension default value was set to 150. The model consisted of 2 convolution layers with kernel size 3, and a pooling layer with kernel size 2, to preserve important features while reducing size. The Dropout layer was included to drop 30% of the

neurons, to avoid overfitting during the training phase. This was followed by the fully-connected layers of 128,64 and an output layer of 1 neuron, with no activation function (to produce a raw continuous value). We trained the model for 1000 epochs, with an early stopping parameter active to stop the training phase at a minimum loss function value. The same model was used to train the synergy dataset; however the input dimension had to be changed to 300, since drug 1 and drug 2 each requires 150 bits.

2.3.3 Modelling Dose Response Curves

After retrieving the binding affinity score from the predictive model. We aim to derive the approximate IC₅₀ value, which is the approximate concentration at which the ligand inhibits the activity of a target pathway by 50%. We assume the binding affinity score to be the Gibbs free energy ΔG . Using the Gibbs Free equation (eqn 1), we derive the inhibition constant K_d . The IC₅₀ value is approximated from the Cheng - Prusoff Equation[47] where we equate the Michelis Menten constant with the IC₅₀ value. We then use the 4 parameter logistic equation (eqn 2) to graph the rate in inhibition of a pathway depending on the inputs given in the application. In this context, Y_{\min} and Y_{\max} represent the percentage of inhibition, where Y_{\max} is 100% and Y_{\min} is 0%. The variable x denotes the range of concentrations, which we perform a sweep test from 0.01 μM to 100 μM .

$$\Delta G = RT \ln \left(\frac{K_i}{c} \right) \quad (1)$$

$$y = y_{\min} + \frac{y_{\max} - y_{\min}}{1 + 10^{\lambda(\log_{10} IC_{50} - x')}} \quad (2)$$

2.4 ADME Profiling

We used the RDKit package [10] to generate the 2D structure of a molecule from its SMILES representation and calculate several ADME properties. These include molecular weight, logP (hydrophobicity), topological polar surface area (TPSA), hydrogen bond donors and acceptors, rotatable bonds, and Lipinski's rule of five compliance. Additionally, we assessed water solubility (LogS), synthetic accessibility, bioavailability score, and blood-brain barrier (BBB) permeability. By analyzing these properties early in the drug discovery process, researchers can prioritize promising compounds, reducing time and cost before experimental validation.

During drug discovery, one of the most critical factors to consider is CYP enzyme inhibition, as it plays a key role in drug metabolism. Inhibition of these enzymes can lead to drug-drug interactions, toxicity, and altered drug efficacy. To address this, we trained five deep learning classification models to predict whether a molecule will inhibit the following CYP enzymes: CYP1A2, CYP2C9, CYP2C19, CYP2D6, and CYP3A4. These enzymes are responsible for metabolizing a wide range of drugs in the human body. We obtained the training data from PubChem to train these models [48–52]. The summary of graphical visualizations of the CYP enzymes is given in figure 4

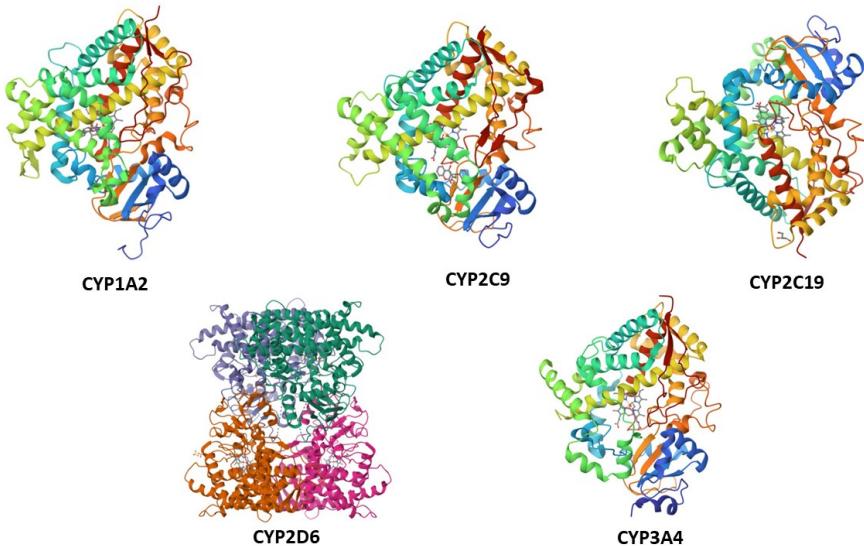


Fig. 4 Graphical Summary of CYP Models

2.5 Agentic AI Framework for Inference Assistance

In addition to the predicted values from the machine learning models. We have added an additional step to provide context to such results. For this, we have implemented an Agentic AI model with the help of CrewAI. At the helm, we create a main agent called a "Drug Discovery scientist", we initialize a backstory for this agent as a molecular modelling expert who interprets docking scores and analyses supporting literature. This agent is provided multiple tasks, particularly in inferring and providing context to results from the predictive models. Listing 1 provides one such example where the role, goal and backstory of the drug discovery scientist agent is defined as well as the task to be done, which in this case is to provide an inference and context to a predictive binding affinity result.

```

1 inferencer = Agent(
2     role="Drug Discovery Scientist",
3     goal="Analyze molecular docking results and provide insights based on
        supporting studies.",
4     backstory="A molecular modeling expert who interprets docking scores and
        identifies supporting literature.",
5     verbose=True,
6     llm=llm,
7 )
8
9 give_insight = Task(
10     description=(
11         "Write a concise, fact-based paragraph interpreting the following result: A
            molecule with SMILES: {smiles1} docked with protein {protein} is predicted to
            have a binding affinity of {value} kcal/mol and an IC50 value of {valueic50}
            \mu M."
12     ),

```

```

13     expected_output="A scientific analysis discussing the binding affinity and IC50
14     value, potential implications, and references to credible studies if
15     applicable.",
) agent=inferencer,

```

Listing 1 Sample Agent and task initialization for drug analysis

2.6 Web Application

We used the Python package Streamlit to develop the DrugXplorer web application that performs multiple in-silico drug discovery analyses. Our platform provides researchers with ADME profile analysis, binding affinity prediction, and drug synergy prediction along with AI inference in a single, easy-to-use interface. Many existing software and web applications for in-silico drug discovery are expensive and have steep learning curves. Additionally, different types of analysis such as ADME prediction, molecular docking, and molecular dynamics often require separate platforms, making the workflow inefficient and time-consuming. Our goal is to create a free, user-friendly web application that provides these essential analyses into one accessible platform. By streamlining the drug discovery process, we aim to help researchers make faster and more informed decisions, ultimately accelerating the development of new therapeutics [53]

3 Results

3.1 Predictive Models

3.1.1 ADME Profiling and CYP Predictions

Table 1 provides a comprehensive overview of the datasets for the 5 key CYP enzymes, which are important in drug metabolism and pharmacokinetics. Table 2 compares the Accuracy, Area Under the Curve (AUC), and Matthews Correlation Coefficient (MCC) of our model with other published models. Our model's performance is comparable to existing models, as reflected in the performance metrics. The AUC values range from 0.83 to 0.90, aligning with those of other models. Accuracy scores were consistent across all datasets, with our model achieving the highest accuracy on the CYP2C19 dataset. The MCC values ranged from 0.51 to 0.63, further demonstrating the model's reliability. Figure 5 presents a comparison of the average AUC, accuracy, and MCC across all five datasets against the benchmark models, highlighting our model's strong predictive capability for CYP inhibition.

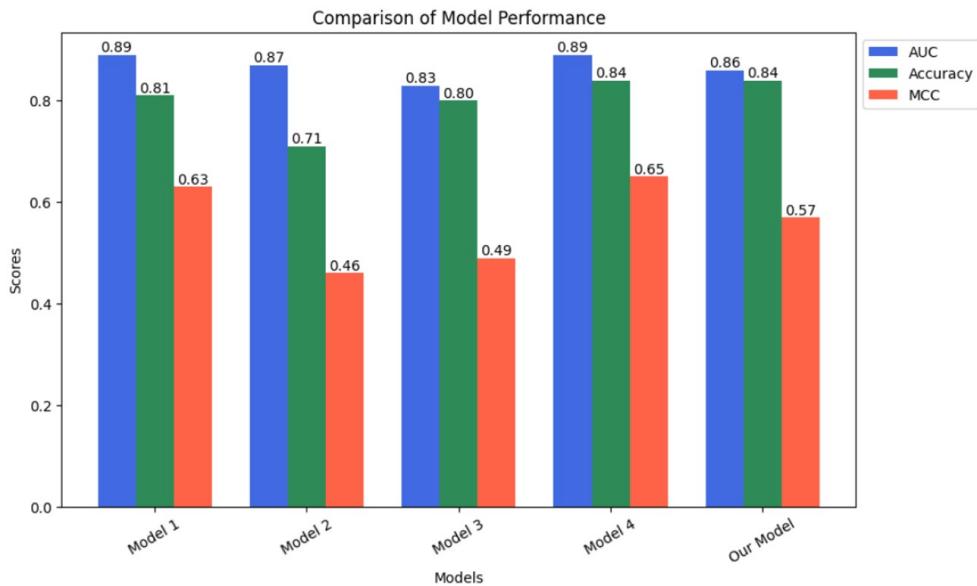
3.1.2 Drug synergy Predictions

Table 3 presents the performance analysis of our model for drug synergy prediction across eight different cell lines. The dataset sizes range from approximately 4,700 to 7,300 rows. R² value measures the proportion of variance in the dependent variable explained by the independent variable, ranging from 0 (poor fit) to 1 (excellent fit). Our model's R² values range from 0.89 to 0.98, indicating a strong fit across all cell lines. The Mean Squared Error (MSE) quantifies the average squared difference

Table 1 CYP Inhibitors and Non-inhibitors Summary

CYPs	No. inhibitors	No. non-inhibitors	Total No. rows
1A2	4439	4005	8444
2C9	1315	7577	8892
2C19	1954	7030	8984
2D6	1773	6809	8582
3A4	3569	7626	11195

between actual and predicted values, with lower values indicating better performance. For most datasets, the MSE is below 1, reflecting strong predictive capability. The Mean Absolute Error (MAE) represents the absolute difference between actual and predicted values, with lower values signifying better accuracy. Our model's MAE values range from 0.34 to 1.38, demonstrating robust performance across most cell lines. The Root Mean Squared Error (RMSE), the square root of MSE, represents the standard deviation of prediction errors, with lower values indicating better performance. Our model's RMSE values remain close to 1 for most datasets, further confirming its reliability.

**Fig. 5** Average AUC, MCC and accuracy of our model against benchmarked models for all CYP datasets

3.1.3 Binding Affinity Prediction

Table 4 presents the performance analysis of our model for binding affinity prediction across 24 different protein receptors. The dataset contains approximately 257,000 rows.

Table 2 Performance comparison of different models on CYP datasets

Models	CYP1A2			CYP2C9			CYP2C19			CYP2D6			CYP3A4		
	AUC	Accuracy	MCC	AUC	Accuracy	MCC	AUC	Accuracy	MCC	AUC	Accuracy	MCC	AUC	Accuracy	MCC
[54]	0.93	0.85	0.70	0.89	0.80	0.59	0.59	0.59	0.88	0.80	0.64	0.91	0.82	0.64	0.64
[55]	0.83	0.69	0.46	0.87	0.61	0.38	0.90	0.64	0.40	0.88	0.81	0.53	0.88	0.81	0.53
[56]	0.81	0.71	0.60	0.86	0.87	0.46	0.84	0.81	0.51	0.88	0.88	0.43	0.78	0.75	0.47
[57]	0.90	0.82	0.64	0.90	0.85	0.63	0.86	0.82	0.64	0.90	0.90	0.68	0.92	0.85	0.66
Our Model	0.84	0.78	0.57	0.83	0.85	0.53	0.85	0.84	0.51	0.85	0.86	0.58	0.90	0.84	0.63

Table 3 Performance metrics for different cell lines

Cell Line	No. Rows	R Square	MSE	MAE	RMSE
MCF7	5091	0.9856	0.3409	0.3446	0.5839
786-0	4862	0.9311	2.3441	1.1777	1.5311
A549	4930	0.9692	0.8985	0.6692	0.9479
DU145	5033	0.9756	0.8534	0.6940	0.9238
HCT116	7369	0.8992	4.5641	1.3275	2.1364
K562	4941	0.9931	0.3809	0.4084	0.6172
OVCAR3	7137	0.9088	4.0807	1.3840	2.0201
SNB75	4719	0.9771	0.4327	0.4691	0.6578

The R^2 value across all 24 datasets is 0.68, indicating that the model provides a reliable estimate of binding affinity, aiding researchers in identifying potential inhibitors for target proteins. The MSE values range from 0.21 to 0.56, reflecting strong predictive performance. Similarly, the MAE values, ranging from 0.37 to 0.59, highlight the model's accuracy in predicting binding affinity for most protein receptors. The RMSE values range from 0.43 to 0.76, demonstrating a strong overall performance. Notably, the model performs exceptionally well for protein receptors such as AKT1, ROCK1, and BACE1.

Table 4 Performance metrics for different proteins

Protein	No. Rows	R Square	MSE	MAE	RMSE
ABL1	258279	0.5832	0.4792	0.5538	0.6922
ACHE	258272	0.5891	0.5476	0.5879	0.7400
ADORA2A	256589	0.6871	0.3471	0.4611	0.5891
ADRB1	257842	0.7022	0.4037	0.4996	0.6354
AKT1	258696	0.8252	0.2248	0.3723	0.4741
AR	257436	0.4673	0.4322	0.5183	0.6574
BACE1	256676	0.7275	0.1891	0.3434	0.4349
CASP3	258395	0.7181	0.2138	0.3643	0.4624
CYP3A4	258192	0.8045	0.2868	0.4215	0.5355
DPP4	256154	0.7335	0.2622	0.4050	0.5121
DRD2	257888	0.7568	0.3389	0.4615	0.5822
EGFR	257288	0.6992	0.3340	0.4537	0.5780
ESR1	258057	0.5486	0.3719	0.4836	0.6098
HMGCR	257232	0.6382	0.2314	0.3802	0.2810
JAK2	255831	0.6497	0.2886	0.4216	0.5373
KDR	259469	0.6278	0.5565	0.5974	0.7460
MAPK1	257209	0.6607	0.3054	0.4333	0.5527
NR3C1	257344	0.6382	0.4623	0.5248	0.6799
PARP1	257895	0.7579	0.3174	0.4449	0.5633
PLK1	257803	0.6840	0.3199	0.4462	0.6840
PPARD	258082	0.7156	0.3386	0.4608	0.5819
PPARG	258822	0.7132	0.2854	0.4229	0.5343
PTGS2	258038	0.4591	0.4988	0.5594	0.7063
ROCK1	256552	0.7730	0.2194	0.3702	0.4684

3.2 Web Application Features

Figure 6 showcases the Home page of DrugXplorer which provides a brief introduction to the three main sections of our application: ADME analysis, binding affinity prediction, and drug synergy prediction.

For ADME analysis, the application presents key molecular properties of a compound in a tabular format, which includes molecular weight, hydrophobicity, TPSA, number of hydrogen donors and acceptors, number of rotatable bonds, number of Lipinski violations, Lipinski Rule of Five validators, water solubility, synthesizability, bioavailability, and BBB permeability. These properties are calculated using RDKit. These properties are used to assess the compound's potential for oral bioavailability, BBB permeability, and drug-drug interactions. The classification models predict whether the compound inhibits CYP1A2, CYP2C9, CYP2C19, CYP2D6, and CYP3A4 enzymes as shown in Figure 7.

Following this, the application generates an inference by analyzing the ADME properties, offering insights into the compound's absorption, distribution, metabolism, and excretion (ADME) characteristics as shown in Figure 8.

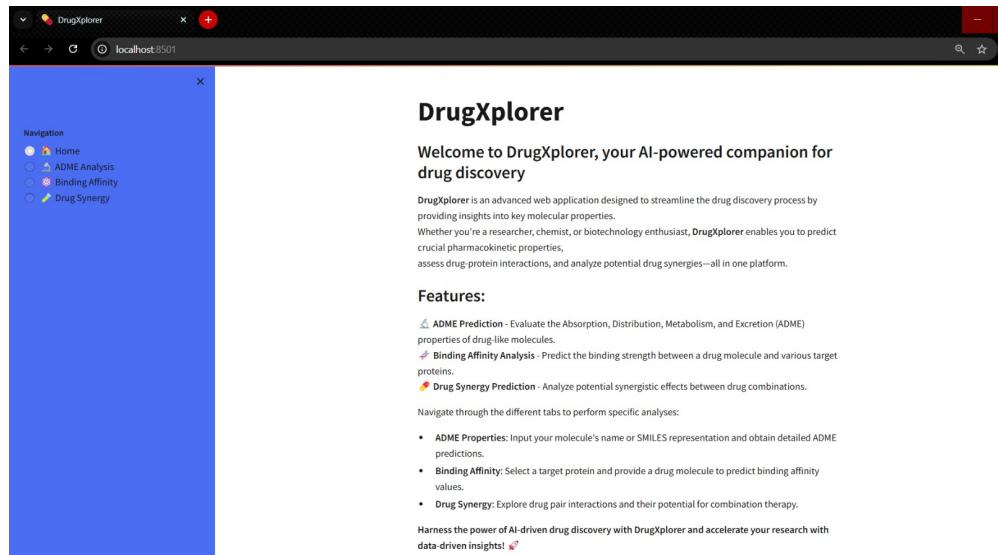


Fig. 6 DrugXplorer Homepage

ADME Analysis

• 📈 Molecular Properties:

- **Molecular Weight (MW):** Measures the size of the molecule.
- **$\log P$ (Hydrophobicity):** Indicates lipid solubility; affects absorption.
- **Topological Polar Surface Area (TPSA):** Predicts permeability & solubility.
- **H-bond Donors:** Number of hydrogen bond donors in the molecule.
- **H-bond Acceptors:** Number of hydrogen bond acceptors in the molecule.
- **Rotatable Bonds:** Determines molecule flexibility; impacts oral bioavailability.
- **Lipinski Violations:** Rules for drug-likeness (≤ 1 violation is preferred).
- **Lipinski Rule of Five Pass:** Whether the molecule meets Lipinski's criteria.
- **Water Solubility (LogS):** Predicts solubility; lower LogS = better solubility.
- **Synthetic Accessibility Score:** Estimates ease of synthesis (lower is better).
- **Bioavailability Score:** Probability of good oral bioavailability.
- **Blood-Brain Barrier (BBB) Permeability:** Predicts CNS drug potential.

• 🌱 Drug-Likeness & CYP Inhibition:

- **Lipinski Rule of Five Pass:** Evaluates overall drug-likeness.
- **Bioavailability Score:** Assesses potential for oral absorption.
- **Blood-Brain Barrier (BBB) Permeability:** Predicts CNS drug capability.
- **CYP Inhibition:** Predicts whether molecule will inhibit CYP1A2, CYP2C9, CYP2C19, CYP2D6, and CYP3A4.

ADME Properties

Property	Value	Interpretation
0 Molecular Weight	368.3850	Should be < 500 for good permeability
1 $\log P$	3.3699	Measures hydrophobicity; affects absorption
2 TPSA	93.0600	Below 140 \AA^2 favors permeability
3 H-bond Donors	2.0000	Should be ≤ 5 for drug-likeness
4 H-bond Acceptors	6.0000	Should be ≤ 10 for permeability
5 Rotatable Bonds	8.0000	Flexibility affects oral bioavailability
6 Lipinski Violations	0.0000	≤ 1 violation preferred
7 Lipinski Rule of Five Pass	1.0000	Indicates drug-likeness
8 Water Solubility (LogS)	-3.7949	Lower LogS = better solubility
9 Synthetic Accessibility	0.4519	Lower value = easier synthesis
10 Bioavailability Score	1.0000	1 indicates good oral bioavailability
11 BBB Permeability (Heuristic)	1.0000	Predicts CNS drug potential

☒ This molecule passes Lipinski's Rule of Five (drug-like).

☒ This molecule meets Veber's bioavailability criteria.

☒ This molecule has good potential for Blood-Brain Barrier (BBB) permeability.

• ⚡ How to Use This App

1. Enter drug name or its SMILES representation
2. Click **Analyze** to get the ADME properties.

Do you want to enter Name or SMILES representation

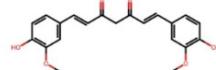
Name

Enter molecule's name

curcumin

Analyze

Molecular Structure



⚠ This molecule is an inhibitor of CYP1A2.

⚠ This molecule is an inhibitor of CYP2C9.

⚠ This molecule is an inhibitor of CYP2C19.

✓ This molecule is not an inhibitor of CYP2D6.

✓ This molecule is not an inhibitor of CYP3A4.

Fig. 7 ADME Analysis and prediction of inhibitory activity of CYP by RDKit and classification models

Inference	Metabolism
<p>Absorption</p> <ul style="list-style-type: none"> Molecular Weight: The molecular weight of 368.385 g/mol is within the recommended range for oral bioavailability (less than 500 g/mol). This suggests that the compound may have good absorption properties. Hydrophobicity (LogP): The hydrophobicity value of 3.37 indicates that the compound is lipophilic. This property can facilitate passive diffusion across biological membranes, enhancing absorption. Solubility (Log S): The solubility value of -3.79486 indicates that the compound has low water solubility. This might pose a challenge for absorption, as poorly soluble compounds may have limited dissolution in the gastrointestinal tract. Lipinski's Rule of Five: The compound has 0 violations of Lipinski's rule, suggesting it has potential for oral absorption. Bioavailability: The prediction of high bioavailability (true) is encouraging for absorption, as the compound is likely to reach systemic circulation in significant amounts. 	<ul style="list-style-type: none"> CYP Inhibitors: The compound is predicted to be an inhibitor of CYP enzymes other than CYP2D6 or CYP3A4. This suggests potential for drug-drug interactions with substrates of these enzymes. Careful consideration should be given to potential interactions. Number of Rotatable Bonds: The compound has 8 rotatable bonds, which typically allows for good oral bioavailability and metabolism. However, bonds can sometimes lead to increased metabolic liability.
<p>Distribution</p> <ul style="list-style-type: none"> Topological Polar Surface Area (TPSA): The TPSA of 93.06 Å² is within the range that allows for good oral bioavailability and cell membrane permeability, suggesting favorable distribution properties. Blood-Brain Barrier Permeability: The prediction of true for blood-brain barrier permeability indicates that the compound may be able to cross the blood-brain barrier, which is crucial for targeting central nervous system diseases. Hydrophobicity (LogP): The moderate hydrophobicity (LogP of 3.37) can facilitate the controlled distribution into tissues and across membranes. 	<p>Excretion</p> <ul style="list-style-type: none"> Molecular Weight and TPSA: The molecular weight and TPSA values suggest the compound will be excreted via both renal and biliary routes. The moderate lipophilicity of the compound allows for potential excretion through urine and bile. Metabolic Stability: The inhibition of multiple CYP enzymes suggests the compound may undergo extensive metabolism, which can influence its excretion process and contribute to the overall excretion process. Synthetic Accessibility: The synthetic accessibility score of 0.45188 indicates the compound is relatively easy to synthesize, which is beneficial for further studies on pharmacokinetic properties.

Fig. 8 Inference generated from Agentic AI Framework for ADME Analysis using Mistral and CrewAI

As Figure 9 displays, once the compound and target protein have been entered, the binding affinity is predicted and displayed along with the amount of dosage required to modulate the selected protein. An inference has been developed, which analyses molecular properties of the given drug compound, such as binding affinity, equilibrium constant and structural context, all of which are derived from the predictions of the back-end model. These properties indicate the strength of the compound-target interaction, the stability of the complex formed, and the potential for favourable interactions within the target's binding pocket. Also, a dose-response curve has been visualized, which is a graphical representation of the relationship between the drug concentration and the biological response or activity of the compound. This curve is used to show the inhibition or activation of the target protein with respect to the increasing drug concentration. Additionally, an inference into the potential therapeutic implications was generated, which may help researchers prioritize the drug compound for further development and testing.

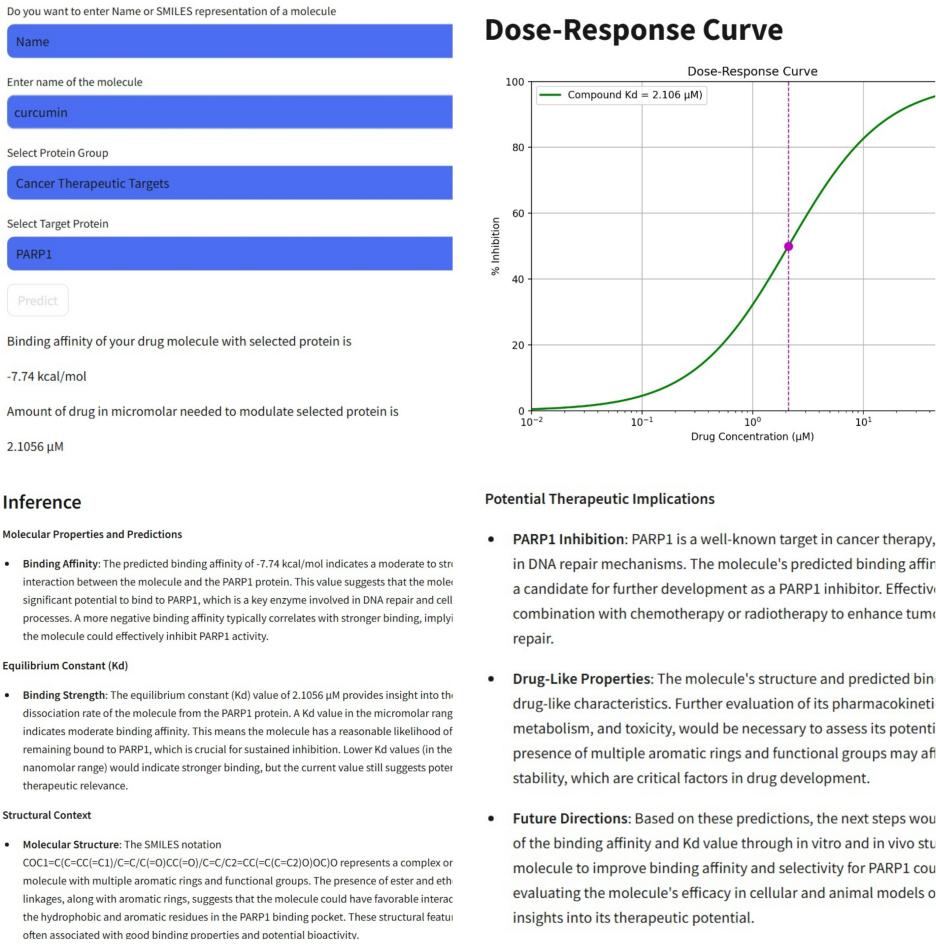


Fig. 9 Binding Affinity prediction and inference is demonstrated in DrugXplorer

As Figure 10 illustrates, once two compounds and a target cell line have been given as input, the bliss synergy score is predicted. The bliss synergy score is quantified as ‘synergistic’ when the score is greater than 1 (increasing effectiveness when combined in the input cell line), ‘additive’ if it falls between 1 and -1 (effective without the drug-drug interaction), ‘antagonistic’ if it is less than -1 (the drug combination hindering the effectiveness). The inference generated based on the synergy bliss score has offered insights into the potential efficacy of the drug combination on the selected cell line, as well as some background information on the cell line, where it was derived from, and its applicability to various cancer treatments. Because it emphasizes the potential advantages of combination therapy in cancer treatments, this inference has also raised possible implications for clinical applications and drug development.

Drug Synergy Prediction

• Cell Lines

Cell Lines & Cancer Types

- MCF7 – Breast cancer (ER⁺, Luminal A).
- A549 – Lung adenocarcinoma (NSCLC).
- HCT116 – Colorectal carcinoma.
- DU145 – Prostate cancer (androgen-independent).
- K562 – Chronic myelogenous leukemia (CML).
- OVCAR3 – Ovarian adenocarcinoma.
- SNB75 – Glioblastoma (brain tumor).
- 786-O – Renal cell carcinoma (RCC, kidney cancer).

Do you want to enter the Name or SMILES representation

Name

Enter name of the first molecule

Curcumin

Enter name of the second molecule

Piperine

Choose your desired cell line

A549

Predict Synergy

Bliss score of the two molecules with desired cell line is

3.55

How to Interpret Bliss Synergy Score

The Bliss Synergy Score quantifies the interaction between two drugs compared to their independent effects.

Bliss Synergy Score	Interpretation
> 1	Synergistic – Drugs work significantly better together than independently.
1 to -1	Additive – Drugs work as expected without interacting.
< -1	Antagonistic – Drug combination reduces effectiveness of individual drugs.

How to Use This App

1. Select two drugs from the input list or enter SMILES.
2. Choose a cancer cell line from the dropdown menu.
3. Click Predict to calculate the Bliss Synergy Score.

Inference

- The Bliss Drug Synergy score of 3.55 for the combination of Curcumin and Piperine is synergistic. The Bliss score of 3.55 indicates that the combined effect is 3.55 times greater than the sum of their individual effects. This synergistic interaction implies that the drugs may target different pathways within the cancer cells, leading to a more comprehensive inhibition of cancer growth.
- A Bliss score higher than 1 is considered synergistic, meaning the combined effect is greater than the sum of their individual effects. This synergistic effect observed in the combination could be particularly effective in treating lung cancer treatments. The synergistic effect observed in the combination could be particularly effective in treating lung cancer treatments.
- The A549 cell line is derived from human lung carcinoma, making it a common model system for studying lung cancer treatments. The synergistic effect observed in the combination could be particularly effective in treating lung cancer treatments.
- The synergistic interaction implies that the drugs may target different pathways within the cancer cells, leading to a more comprehensive inhibition of cancer growth.
- This finding could have significant implications for drug development, highlighting the potential benefits of combination therapy in cancer treatment.

Fig. 10 Drug synergy prediction and inference is demonstrated in DrugXplorer

3.3 Setup

DrugXplorer can directly be setup in a local system or server. The user has the option to customize the choice of LLM backend (Mistral or GPT4) and must activate the API key before deploying the application. The code for the same is available in [Supplementary Information](#).

4 Conclusion

DrugXplorer provides an AI-driven platform for in-silico drug discovery, addressing key challenges researchers face when using multiple tools for different analyses. By

combining deep-learning models, in-built Python packages, and an Agentic-AI framework powered by large language models, we have created an efficient, accessible, and comprehensive platform for preliminary drug screening by predicting ADME properties, CYP450 inhibition, binding affinity, and drug synergy. Our CNN-based regression model for binding affinity prediction, trained on 24 protein datasets, achieved an average R^2 value of 0.68 with an MSE of 0.34 and MAE of 0.45. The drug synergy prediction model, which estimates the Bliss synergy score across eight cancer cell lines, demonstrated strong performance with an average R^2 value of 0.95, MSE of 1.7, and MAE of 0.8.

Additionally, our deep-learning classification models for CYP450 inhibition prediction showed accuracy, AUC, and MCC values on par with benchmarked models such as MumCYP and others and outperforms in the prediction of CYP2C19 enzyme with an accuracy of 0.84. Furthermore, our Agentic-AI framework provides contextual inferences, making predictive results more interpretable. With a user-friendly interface built using Streamlit, our platform simplifies the drug discovery workflow, making it more accessible, efficient, and cost-effective. This unified approach reduces the need for multiple tools, saving time and effort while maintaining high predictive accuracy. By streamlining these critical analyses into a single, free tool, we aim to accelerate the development of new therapeutics and support researchers in making data-driven decisions with greater confidence.

For our future work, we plan to enhance the accuracy of our models further by incorporating more advanced architectures, such as Quantum Convolutional Neural Networks. Additionally, expanding the number of proteins used for binding affinity prediction will improve the application's generalizability and robustness, making it an even more powerful tool for computational drug discovery.

Supplementary Information. Code for DrugXplorer application is available in the following link: <https://github.com/AshwinSivakumar/Project2-DrugAnalysis>

Acknowledgements. RSK, JM, AS, SCP and RMP thanks the School of Electronics Engineering, VIT Chennai for their support towards the completion of this project.

Declarations

Conflict of interest/Competing interests. The authors declare no conflict of interest.

Code availability. Code is available in following link: <https://github.com/AshwinSivakumar/Project2-DrugAnalysis>

References

- [1] Scannell, J.W., Bosley, J.: When quality beats quantity: Decision theory, drug discovery, and the reproducibility crisis. PLOS ONE **11**(2), 1–21 (2016) <https://doi.org/10.1371/journal.pone.0147215>

- [2] Zhang, K., Yang, X., Wang, Y., Yu, Y., Huang, N., Li, G., Li, X., Wu, J.C., Yang, S.: Artificial intelligence in drug development. *Nature Medicine* **31**(1), 45–59 (2025) <https://doi.org/10.1038/s41591-024-03434-4>
- [3] Jamkhande, P.G., Ghante, M.H., Ajgunde, B.R.: Software based approaches for drug designing and development: A systematic review on commonly used software and its applications. *Bulletin of Faculty of Pharmacy, Cairo University* **55**(2), 203–210 (2017) <https://doi.org/10.1016/j.bfopcu.2017.10.001>
- [4] Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., Li, B., Madabhushi, A., Shah, P., Spitzer, M., Zhao, S.: Applications of machine learning in drug discovery and development. *Nature Reviews Drug Discovery* **18**(6), 463–477 (2019) <https://doi.org/10.1038/s41573-019-0024-5>
- [5] Eberhardt, J., Santos-Martins, D., Tillack, A.F., Forli, S.: Autodock vina 1.2.0: New docking methods, expanded force field, and python bindings. *Journal of Chemical Information and Modeling* **61**(8), 3891–3898 (2021) <https://doi.org/10.1021/acs.jcim.1c00203>. PMID: 34278794
- [6] Trott, O., Olson, A.J.: Autodock vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry* **31**(2), 455–461 (2010) <https://doi.org/10.1002/jcc.21334> <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.21334>
- [7] Trott, O., Olson, A.J.: Autodock vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry* **31**(2), 455–461 (2010) <https://doi.org/10.1002/jcc.21334> <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.21334>
- [8] García-Ortegón, M., Simm, G.N.C., Tripp, A.J., Hernández-Lobato, J.M., Bender, A., Bacallado, S.: Dockstring: Easy molecular docking yields better benchmarks for ligand design. *Journal of Chemical Information and Modeling* **62**(15), 3486–3502 (2022) <https://doi.org/10.1021/acs.jcim.1c01334> <https://doi.org/10.1021/acs.jcim.1c01334>. PMID: 35849793
- [9] Panda, P.K.: PandaDock: Python Molecular Docking. <https://github.com/pritampanda15/PandaDock>. GitHub repository (2025)
- [10] Landrum, G.: RDKit: Open-Source Cheminformatics Software (2023). <https://doi.org/10.5281/zenodo.591637> . <https://www.rdkit.org>
- [11] Ramsundar, B., Eastman, P., Walters, P., Pande, V., Leswing, K., Wu, Z.: Deep Learning for the Life Sciences. O'Reilly Media, ??? (2019)
- [12] Bento, A.P., Hersey, A., Félix, E., Landrum, G., Gaulton, A., Atkinson, F.,

- Bellis, L.J., De Veij, M., Leach, A.R.: An open source chemical structure curation pipeline using rdkit. *Journal of Cheminformatics* **12**(1), 51 (2020) <https://doi.org/10.1186/s13321-020-00456-1>
- [13] Cao, N.D., Kipf, T.: MolGAN: An implicit generative model for small molecular graphs (2022). <https://arxiv.org/abs/1805.11973>
- [14] Chithrananda, S., Grand, G., Ramsundar, B.: ChemBERTa: Large-Scale Self-Supervised Pretraining for Molecular Property Prediction (2020). <https://arxiv.org/abs/2010.09885>
- [15] Landrum, G.A., Riniker, S.: Combining *ic50* or *ki* values from different sources is a source of significant noise. *Journal of Chemical Information and Modeling* **64**(5), 1560–1567 (2024) <https://doi.org/10.1021/acs.jcim.4c00049> <https://doi.org/10.1021/acs.jcim.4c00049>. PMID: 38394344
- [16] Özçelik, R., Grisoni, F.: A hitchhiker’s guide to deep chemical language processing for bioactivity prediction. *Digital Discovery* **4**, 316–325 (2025) <https://doi.org/10.1039/D4DD00311J>
- [17] Jablonka, K.M., Schwaller, P., Ortega-Guerrero, A., Smit, B.: Leveraging large language models for predictive chemistry. *Nature Machine Intelligence* **6**(2), 161–169 (2024) <https://doi.org/10.1038/s42256-023-00788-1>
- [18] Sadeghi, S., Bui, A., Forooghi, A., Lu, J., Ngom, A.: Can large language models understand molecules? *BMC Bioinformatics* **25**(1), 225 (2024) <https://doi.org/10.1186/s12859-024-05847-x>
- [19] Zagidullin, B., Aldahdooh, J., Zheng, S., Wang, W., Wang, Y., Saad, J., Malyutina, A., Jafari, M., Tanoli, Z., Pessia, A., Tang, J.: Drug-comb: an integrative cancer drug combination data portal. *Nucleic Acids Research* **47**(W1), 43–51 (2019) <https://doi.org/10.1093/nar/gkz337> <https://academic.oup.com/nar/article-pdf/47/W1/W43/28880110/gkz337.pdf>
- [20] Wei, Y., Li, S., Li, Z., Wan, Z., Lin, J.: Interpretable-admet: a web service for admet prediction and optimization based on deep neural representation. *Bioinformatics* **38**(10), 2863–2871 (2022) <https://doi.org/10.1093/bioinformatics/btac192> <https://academic.oup.com/bioinformatics/article-pdf/38/10/2863/49009491/btac192.pdf>
- [21] Daina, A., Michelin, O., Zoete, V.: Swissadme: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific Reports* **7**(1), 42717 (2017) <https://doi.org/10.1038/srep42717>
- [22] Kochnev, Y., Durrant, J.D.: Fpocketweb: protein pocket hunting in a web browser. *Journal of Cheminformatics* **14**(1), 58 (2022) <https://doi.org/10.1186/>

- [23] Schmidtke, P., Le Guilloux, V., Maupetit, J., Tuffi^éry, P.: fpocket: online tools for protein ensemble pocket detection and tracking. *Nucleic Acids Research* **38**(suppl_2), 582–589 (2010) <https://doi.org/10.1093/nar/gkq383>
- [24] Frigo, D.E., Bondesson, M., Williams, C.: Nuclear receptors: from molecular mechanisms to therapeutics. *Essays in Biochemistry* **65**(6), 847–856 (2021) <https://doi.org/10.1042/EBC20210020>
- [25] Hantschel, O.: Structure, regulation, signaling, and targeting of abl kinases in cancer. *Genes & Cancer* **3**(5-6), 436–446 (2012) <https://doi.org/10.1177/1947601912458584>
- [26] Yamaoka, K., Saharinen, P., Pesu, M., Holt, V.E., Silvennoinen, O., O’Shea, J.J.: The janus kinases (jaks). *Genome Biology* **5**(12), 253 (2004) <https://doi.org/10.1186/gb-2004-5-12-253>
- [27] Chiwoneso, T.C., Luo, Y., Xu, Y., Chen, X., Chen, L., Sun, J.: Kinases and their derived inhibitors from natural products. *Bioorganic Chemistry* **156**, 108196 (2025) <https://doi.org/10.1016/j.bioorg.2025.108196>
- [28] Wee, P., Wang, Z.: Epidermal growth factor receptor cell proliferation signaling pathways. *Cancers* **9**(5) (2017) <https://doi.org/10.3390/cancers9050052>
- [29] Sharpe, L.J., Brown, A.J.: Controlling cholesterol synthesis beyond 3-hydroxy-3-methylglutaryl-coa reductase (hmgcr). *Journal of Biological Chemistry* **288**(26), 18707–18715 (2013) <https://doi.org/10.1074/jbc.R113.479808>
- [30] Zhang, Y., Wang, Z., Wang, Y., Jin, W., Zhang, Z., Jin, L., Qian, J., Zheng, L.: Cyp3a4 and cyp3a5: the crucial roles in clinical drug metabolism and the significant implications of genetic polymorphisms. *PeerJ* **12**, 18636 (2024) <https://doi.org/10.7717/peerj.18636>
- [31] Lim, S.W., Jin, J.Z., Jin, L., Jin, J., Li, C.: Role of dipeptidyl peptidase-4 inhibitors in new-onset diabetes after transplantation. *Korean Journal of Internal Medicine* **30**(6), 759–770 (2015) <https://doi.org/10.3904/kjim.2015.30.6.759>
- [32] Martín-Vázquez, E., Cobo-Vuilleumier, N., López-Noriega, L., Lorenzo, P.I., Gauthier, B.R.: The ptgs2/cox2-pge2 signaling cascade in inflammation: Pro or anti? a case study with type 1 diabetes mellitus. *International Journal of Biological Sciences* **19**(13), 4157–4165 (2023) <https://doi.org/10.7150/ijbs.86492>
- [33] Hampel, H., Vassar, R., De Strooper, B., Hardy, J., Willem, M., Singh, N., Zhou, J., Yan, R., Vanmechelen, E., De Vos, A., Nisticò, R., Corbo, M., Imbimbo, B.P., Streffer, J., Voytyuk, I., Timmers, M., Tahami Monfared, A.A., Irizarry, M., Albala, B., Koyama, A., Watanabe, N., Kimura, T., Yarenis, L., Lista, S.,

Kramer, L., Vergallo, A.: The beta-secretase bace1 in alzheimer's disease. Biological Psychiatry **89**(8), 745–756 (2021) <https://doi.org/10.1016/j.biopsych.2020.02.001>

- [34] Yi, B., Jahangir, A., Evans, A.K., Briggs, D., Ravina, K., Ernest, J., Farimani, A.B., Sun, W., Rajadas, J., Green, M., Feinberg, E.N., Pande, V.S., Shamloo, M.: Discovery of novel brain permeable and g protein-biased beta-1 adrenergic receptor partial agonists for the treatment of neurocognitive disorders. PLOS ONE **12**(7), 1–30 (2017) <https://doi.org/10.1371/journal.pone.0180319>
- [35] Hohoff, C., Kroll, T., Zhao, B., *et al.*: Adora2a variation and adenosine a1 receptor availability in the human brain with a focus on anxiety-related brain regions: modulation by adora1 variation. Translational Psychiatry **10**, 406 (2020) <https://doi.org/10.1038/s41398-020-01085-w>
- [36] Noble, E.P.: The drd2 gene in psychiatric and neurological disorders and its phenotypes. Pharmacogenomics **1**(3), 309–333 (2000) <https://doi.org/10.1517/14622416.1.3.309>
- [37] Colović, M.B., Krstić, D.Z., Lazarević-Pašti, T.D., Bondžić, A.M., Vasić, V.M.: Acetylcholinesterase inhibitors: pharmacology and toxicology. Current Neuropharmacology **11**(3), 315–335 (2013) <https://doi.org/10.2174/1570159X11311030006>
- [38] Kim, C., Wang, X.-D., Jang, S., Yu, Y.: Parp1 inhibitors induce pyroptosis via caspase 3-mediated gasdermin e cleavage. Biochemical and Biophysical Research Communications **646**, 78–85 (2023) <https://doi.org/10.1016/j.bbrc.2023.01.055>
- [39] Eskandari, E., Eaves, C.J.: Paradoxical roles of caspase-3 in regulating cell survival, proliferation, and tumorigenesis. Journal of Cell Biology **221**(6), 202201159 (2022) <https://doi.org/10.1083/jcb.202201159>
- [40] Loirand, G.: Rho kinases in health and disease: From basic science to translational research. Pharmacological Reviews **67**(4), 1074–1095 (2015) <https://doi.org/10.1124/pr.115.010595>
- [41] Modi, S.J., Kulkarni, V.M.: Vascular endothelial growth factor receptor (vegfr-2)/kdr inhibitors: Medicinal chemistry perspective. Medicine in Drug Discovery **2**, 100009 (2019) <https://doi.org/10.1016/j.medidd.2019.100009>
- [42] Workman, P.: The nci-60 human tumor cell line screen: A catalyst for progressive evolution of models for discovery and development of cancer drugs. Cancer Research **83**(19), 3170–3173 (2023) <https://doi.org/10.1158/0008-5472.CAN-23-2612>
- [43] Weininger, D.: Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules. Journal of Chemical Information and

- Computer Sciences **28**(1), 31–36 (1988) <https://doi.org/10.1021/ci00057a005>
- [44] Rogers, D., Hahn, M.: Extended-connectivity fingerprints. Journal of Chemical Information and Modeling **50**(5), 742–754 (2010) <https://doi.org/10.1021/ci100050t>
 - [45] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S.: PyTorch: An Imperative Style, High-Performance Deep Learning Library (2019). <https://arxiv.org/abs/1912.01703>
 - [46] Agarap, A.F.: Deep learning using rectified linear units (relu). CoRR **abs/1803.08375** (2018) [arXiv:1803.08375](https://arxiv.org/abs/1803.08375)
 - [47] Yung-Chi, C., Prsoff, W.H.: Relationship between the inhibition constant (k_i) and the concentration of inhibitor which causes 50 per cent inhibition (i_{50}) of an enzymatic reaction. Biochemical Pharmacology **22**(23), 3099–3108 (1973)
 - [48] National Center for Biotechnology Information: PubChem Bioassay Record for AID 410, p450-cyp1a2. Source: National Center for Advancing Translational Sciences (NCATS) (2025)
 - [49] National Center for Biotechnology Information: PubChem Bioassay Record for AID 883, qHTS Assay for Inhibitors and Substrates of Cytochrome P450 2C9. Source: National Center for Advancing Translational Sciences (NCATS) (2025)
 - [50] National Center for Biotechnology Information: PubChem Bioassay Record for AID 899, qHTS Assay for Inhibitors and Substrates of Cytochrome P450 2C19. Source: National Center for Advancing Translational Sciences (NCATS) (2025)
 - [51] National Center for Biotechnology Information: PubChem Bioassay Record for AID 891, qHTS Assay for Inhibitors and Substrates of Cytochrome P450 2D6. Source: National Center for Advancing Translational Sciences (NCATS) (2025)
 - [52] National Center for Biotechnology Information: PubChem Bioassay Record for AID 884, qHTS Assay for Inhibitors and Substrates of Cytochrome P450 3A4. Source: National Center for Advancing Translational Sciences (NCATS) (2025)
 - [53] Treuille, A., Nelson, J., Dobbins, Z.: Streamlit: A faster way to build and share data apps (2019)
 - [54] Ai, D., Cai, H., Wei, J., Zhao, D., Chen, Y., Wang, L.: Deepcyps: A deep learning platform for enhanced cytochrome p450 activity prediction. Frontiers in Pharmacology **14** (2023) <https://doi.org/10.3389/fphar.2023.1099093>

- [55] Banerjee, P., Dunkel, M., Kemmler, E., Preissner, R.: Supercyprspred—a web server for the prediction of cytochrome activity. Nucleic Acids Research **48**(W1), 580–585 (2020) <https://doi.org/10.1093/nar/gkaa166> <https://academic.oup.com/nar/article-pdf/48/W1/W580/33432872/gkaa166.pdf>
- [56] Cheng, F., Yu, Y., Shen, J., Yang, L., Li, W., Liu, G., Lee, P.W., Tang, Y.: Classification of cytochrome p450 inhibitors and noninhibitors using combined classifiers. Journal of Chemical Information and Modeling **51**(5), 996–1011 (2011) <https://doi.org/10.1021/ci200028n>. PMID: 21491913
- [57] Ngnamsie Njimbouom, S., Kim, J.-D.: Mumcyp.net: A multimodal neural network for the prediction of cyp450 inhibition. Expert Systems with Applications **255**, 124703 (2024) <https://doi.org/10.1016/j.eswa.2024.124703>