



Review

Recent Studies of Artificial Intelligence on In Silico Drug Distribution Prediction

Thi Tuyet Van Tran ^{1,2,3} , Hilal Tayara ^{4,*} and Kil To Chong ^{5,*}

¹ Department of Electronics and Information Engineering, Jeonbuk National University, Jeonju 54896, Republic of Korea

² Department of Information Technology, An Giang University, Long Xuyen 880000, Vietnam

³ Vietnam National University–Ho Chi Minh City, Ho Chi Minh 700000, Vietnam

⁴ School of International Engineering and Science, Jeonbuk National University, Jeonju 54896, Republic of Korea

⁵ Advances Electronics and Information Research Center, Jeonbuk National University, Jeonju 54896, Republic of Korea

* Correspondence: hilaltayara@jbnu.ac.kr (H.T.); kitchong@jbnu.ac.kr (K.T.C.)

Abstract: Drug distribution is an important process in pharmacokinetics because it has the potential to influence both the amount of medicine reaching the active sites and the effectiveness as well as safety of the drug. The main causes of 90% of drug failures in clinical development are lack of efficacy and uncontrolled toxicity. In recent years, several advances and promising developments in drug distribution property prediction have been achieved, especially in silico, which helped to drastically reduce the time and expense of screening undesired drug candidates. In this study, we provide comprehensive knowledge of drug distribution background, influencing factors, and artificial intelligence-based distribution property prediction models from 2019 to the present. Additionally, we gathered and analyzed public databases and datasets commonly utilized by the scientific community for distribution prediction. The distribution property prediction performance of five large ADMET prediction tools is mentioned as a benchmark for future research. On this basis, we also offer future challenges in drug distribution prediction and research directions. We hope that this review will provide researchers with helpful insight into distribution prediction, thus facilitating the development of innovative approaches for drug discovery.



Citation: Tran, T.T.V.; Tayara, H.; Chong, K.T. Recent Studies of Artificial Intelligence on In Silico Drug Distribution Prediction. *Int. J. Mol. Sci.* **2023**, *24*, 1815. <https://doi.org/10.3390/ijms24031815>

Academic Editors: Luca Pinzi and Giulio Rastelli

Received: 14 December 2022

Revised: 11 January 2023

Accepted: 13 January 2023

Published: 17 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: ADMET; distribution prediction; drug discovery; artificial intelligence; machine learning; deep learning

1. Introduction

Pharmacokinetics, the study of how pharmaceuticals are handled in the body, consists of four stages: absorption, distribution, metabolism, and excretion (ADME) (Figure 1A). It plays a very important role in drug research and development (R&D) because any drug candidate must be checked for pharmacokinetics and toxicity (ADMET) properties to ensure efficacy and safety. The average capitalized investment in R&D to bring a new medicine to the market is estimated at USD 1.1417 billion, after considering the cost of unsuccessful studies [1]. A key problem in drug R&D is the failure of compound candidates in clinical trials. Increasing success rates in clinical trials is believed to be the most profound factor in overall cost reductions and to outweigh savings in other phases [2]. If research improves the prediction of a drug's failure by 10% before clinical trials, it could save about USD 100 million in development expenses for each drug [3]. Therefore, identifying chemical candidates with higher efficacy and no toxic or otherwise unfavorable side effects is a major challenge.

Early ADMET property assessment research may considerably increase the drug's success rate, decrease the drug's R&D costs, minimize the incidence of side effects and toxicities, and provide a direct therapeutic rationale for drug usage. Drug distribution is a

critical step in the ADMET process because it has the potential to influence both the amount of drug that reaches the active sites as well as the effectiveness and toxicity of the drug. Lack of efficacy and uncontrolled toxicity are the main causes of 90% of medication failures during clinical development [4]. Drug distribution can cause unwanted reactions and side effects. Moreover, optimizing the distribution property affects other properties of ADMET because drug distribution is an important mediating process. Therefore, predicting drug distribution properties is essential during the early phases of drug research.

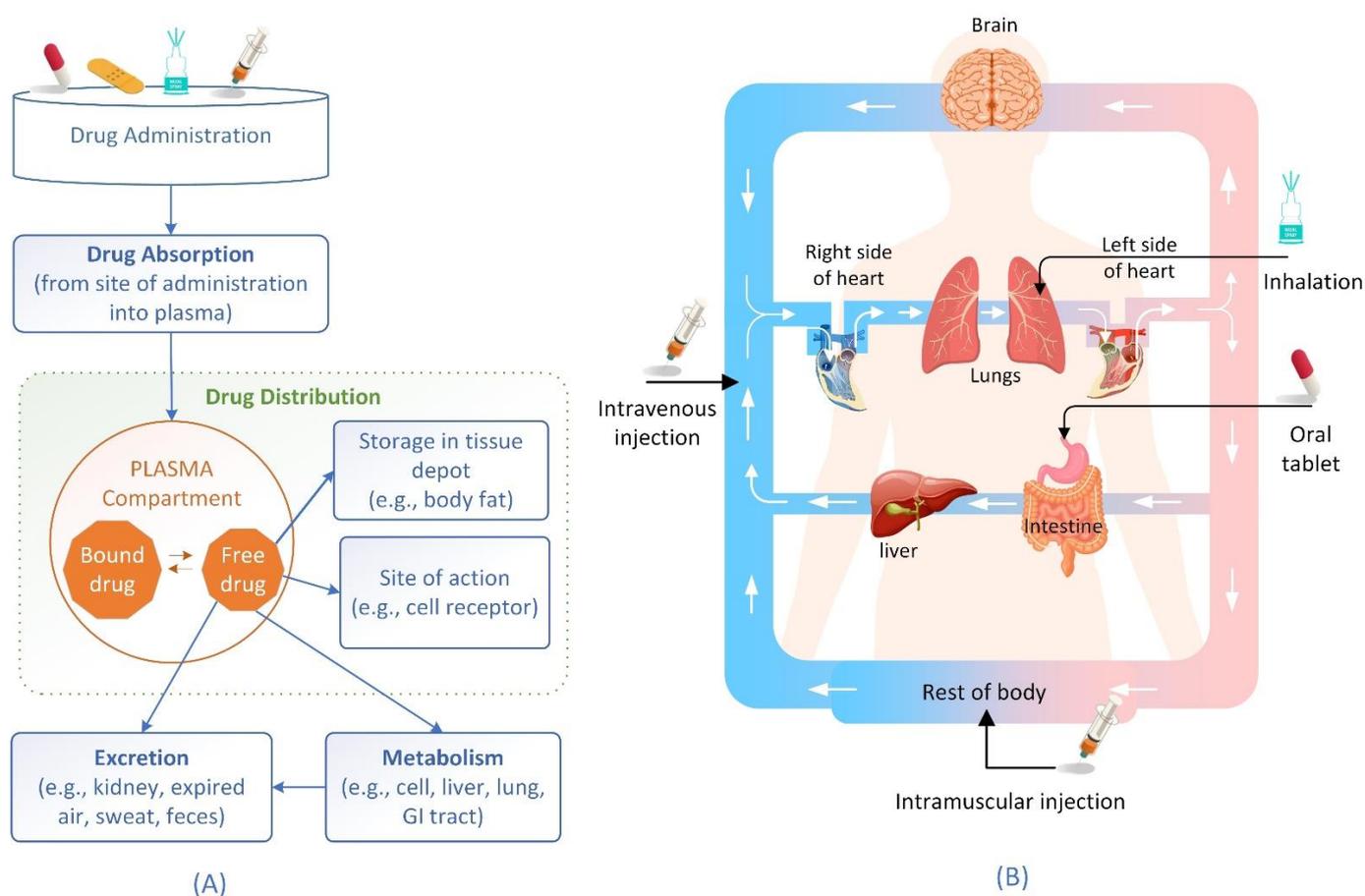


Figure 1. (A) Schematic description of the pharmacokinetic (ADME) process. (B) Drug administration and drug distribution process in the body [5].

In traditional drug R&D, predicting distribution properties relies heavily on *in vitro* and *in vivo* studies. Despite progress in technological innovation, conventional experimental evaluations of distribution properties are often costly and time consuming. For example, assessing the blood–brain barrier penetration of every chemical takes one week and costs nearly USD 10,000 in a non-good laboratory practice facility [3]. Moreover, *in vitro* screening of compounds is typically limited to a few properties, with emphasis placed on only a few of the most promising chemical candidates. As a result, *in silico* distribution-related models are widely employed for quick and early screening of the drug distribution properties before they are further explored *in vitro* [6]. Because of the recent enormous success of artificial intelligence (AI) in many fields, AI-based drug R&D is ready to become a large force in the field of pharmaceuticals and is projected to make significant improvements in preclinical research. AI systems can efficiently and cheaply screen thousands or millions of candidate molecules rather than limiting the examination of distribution features to a select few. To construct, optimize, and improve the model performance, it is crucial to have a clear understanding of the distribution properties and the latest advances in AI-based distribution prediction models. Therefore, in this study,

we comprehensively reviewed recent studies that used AI to predict drug distribution properties. Additionally, we collected available databases and datasets that the scientific community often uses for distribution prediction. We provided a list of free tools that support ADMET property prediction, along with the distributional property prediction performance of five recent tools. Finally, challenges and future directions for researchers working on AI-based distribution prediction are discussed. We believe that it will be helpful for researchers to work on improving the distribution property prediction model and other properties of the ADMET. It is important to note that the examination of drug distribution should be evaluated within the context of a particular drug delivery strategy. However, this manuscript primarily focuses on the application of AI-based models for predicting drug distribution.

2. Drug Distribution Process and Factors Affecting the Process

Once absorbed into the bloodstream, the medication circulates rapidly throughout the body. Drugs are transported from the bloodstream to body tissues during the recirculation process, which is the drug distribution stage. Drug distribution is the process by which an unmetabolized drug is distributed throughout the bloodstream and tissues of an organism (Figure 1B). The effectiveness or toxicity of medicine is dependent on its distribution in certain tissues, which explains in part why there is a lack of relationship between plasma levels and observed effects [7]. Drugs have varying distributions in various tissues, including fat, muscle, lungs, and brain, depending on their molecular structure and administration. The pharmacological effect of a drug depends on its concentration at the action site. This means that the distribution is a key factor in determining when, how strong, and sometimes how long the drug will work.

There are two stages in the transport of medications from the bloodstream to the tissues outside the blood vessels: one, the rapid passage of free or unbound medication from the blood through the capillary wall and into the interstitial/extracellular fluid (ECF), and two, the passage of drug from the ECF across cell membranes to the intracellular fluid [8]. Various tissues absorb the drug from the plasma at various speeds and to varying degrees, leading to a non-uniform distribution of the drug throughout the body. Drug distribution is a passive process influenced by many factors, including drug permeability across tissues, organ/tissue size, perfusion rate, drug binding to tissue components, and other factors, as depicted in Figure 2. In preclinical *in silico* studies, the endpoints or properties of distribution were determined based on these factors. Based on large ADMET prediction systems and recent studies, we synthesized some properties (endpoints) of the distribution, such as physicochemical properties (molecular weight, heavy atoms, log P, log D, log S, pKa, etc.), plasma protein binding, blood–brain barrier, human placenta barrier, volume of distribution, fraction unbound in human plasma, and fraction unbound in the brain.

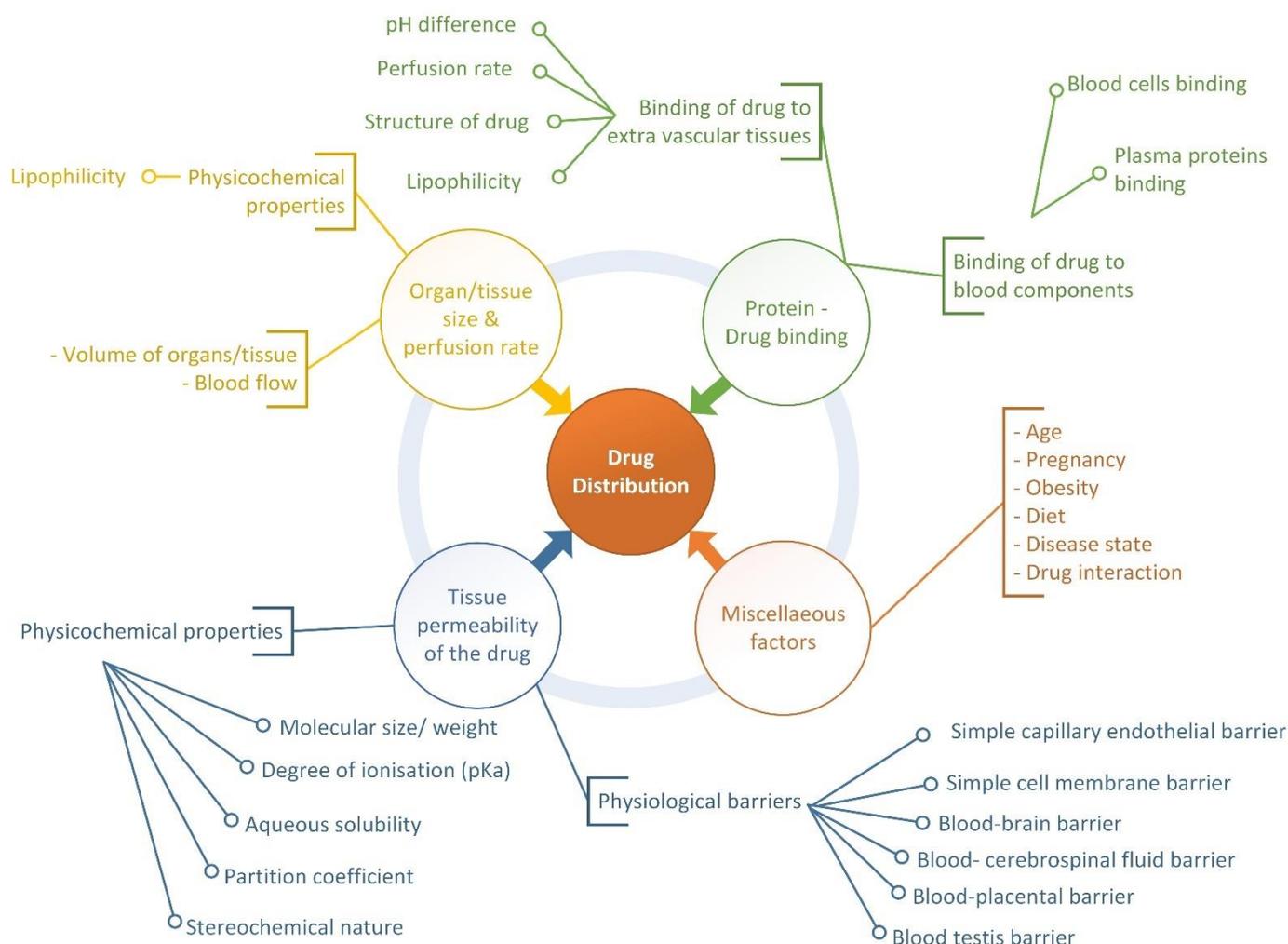


Figure 2. Factors influencing drug distribution.

3. Performance Metrics to Evaluate and Compare AI-Based Distribution Prediction Methods

Evaluating the performance of AI methods is important to measure how effective a method is and to compare the performance of different methods fairly. In this review, we present the following evaluation metrics: accuracy (ACC), precision, recall, F1 score, area under the receiver operating characteristic curve (AUC), mean absolute error (MAE), root mean squared error (RMSE), coefficient of determination (R^2), predictive relevance (Q^2), and geometric mean fold error (GMFE). The formulas are as follows:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \text{Range}[0,1] \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \text{Range}[0,1] \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \text{Range}[0,1] \quad (3)$$

$$F1score = \frac{2 \times Precision \times Recall}{Precision + Recall} \text{Range}[0,1] \quad (4)$$

$$AUC = \text{Area under the receiver operating characteristic curve} \text{Range}[0,1] \quad (5)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}| \quad (6)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2} \quad (7)$$

$$R^2 = 1 - \frac{\sum (y_i - \hat{y})^2}{\sum (y_i - \bar{y})^2} \text{Range}[0, 1] \quad (8)$$

$$Q^2 = 1 - \frac{\sum (y_i - \hat{y})^2}{\sum (y_i - \bar{y})^2} \text{Range}[0, 1] \quad (9)$$

$$GMFE = \frac{1}{N} \sum_{i=1}^N \left| \log_{10} \left(\frac{\hat{y}}{y_i} \right) \right| \quad (10)$$

In the Equations (1)–(3), TN, FN, TP, and FP represent the number of true negatives, false negatives, true positives, and false positives, respectively. In the Equations (6)–(10), N , y_i , \hat{y} , and \bar{y} represent total numbers of observation, the actual value for the i th observation, the predicted value of y , and mean value of y , respectively.

ACC is a measure of how many predictions made by a model are correct. It is calculated by dividing the number of correct predictions by the total number of predictions made. When the positive and negative classes are uneven, evaluating system performance based on accuracy may result in excessive bias [9]. In virtual screening data, the number of negative samples is frequently much greater than the number of positive ones. The accuracy score would exaggerate the performance of a failing prediction model that labels all cases as negative (i.e., inactive, or non-interacting). Precision is a measure of a model's positive predictions' accuracy. It is computed by dividing the number of accurate positive predictions by the total number of positive predictions generated by the model. Recall quantifies the proportion of actual positive cases that the model accurately identified. It is determined as the number of accurate positive predictions divided by the total number of actual positive cases. The F1 score quantifies the balance between recall and precision. It is computed based on the harmonic mean of precision and recall. AUC is a measure of the performance of a binary classification model. It is calculated by plotting the true positive rate against the false positive rate at various classification thresholds. The AUC ranges from 0 to 1, with higher values indicating better performance. MAE is a measure of the difference between predicted and actual values that is calculated by taking the mean of the absolute differences between the predicted and actual values. It is easy to understand and interpret, but it is not as sensitive to outliers as other measures. RMSE is calculated by taking the square root of the mean of the squared differences between the predicted and actual values. It is one of the most widely used measures of model performance, as it is easy to interpret and sensitive to both the mean and the variance of the error. R^2 is a measure of how well a model fits the data. It is calculated as the ratio of the variance of the predicted values to the variance of the actual values. An R^2 value of 1 indicates that the model perfectly fits the data, while a value of 0 indicates that the model does not fit the data at all. Q^2 is a measure of the performance of a model in a predictive modeling context. It is calculated as the ratio of the variance of the predicted values to the total variance of the observed data. It ranges from -1 to 1 , with a higher value indicating a better performing model. The main difference between R^2 and Q^2 is that R^2 is a measure of model fit, while Q^2 is a measure of model predictive ability [10]. R^2 is calculated using the same data that was used to fit the model, while Q^2 is calculated using a different set of data (i.e., a "holdout" set). This means that Q^2 is a more conservative estimate of a model's performance, as it reflects the model's ability to generalize to new data. Another difference between R^2 and Q^2 is that R^2 is only defined for regression models, while Q^2 can be used for both regression and classification models. GMFE is calculated as the geometric mean of the fold errors, where the fold error

is defined as the absolute value of the difference between the predicted and actual values. A lower GMFE indicates a better performing model.

It is worth noting that the results presented in the following sections are based on a specific dataset that may not have enough chemical diversity to accurately predict properties for new, first-in-class therapeutics.

4. AI-Based Distribution Property Prediction

AI is an area of computer science that aims to develop machines exhibiting human-like intelligence in problem solving, task performance, and learning [11]. Machine learning (ML) is a subset of AI involving the use of algorithms and statistical models to enable ML and improve its performance without explicit programming, which requires feeding the machine with large amounts of data and allowing it to learn patterns and relationships in the data. Deep learning (DL) is a specific type of ML that involves using neural networks designed to mimic the structure and function of the human brain. Deep learning algorithms are particularly effective on complex tasks and can process and analyze vast amounts of data and make decisions based on the patterns and relationships in the data. AI techniques can be used to analyze and predict potential ADMET properties of molecules, based on their structural and chemical properties. These techniques can be used to identify and prioritize molecules likely to exhibit good ADMET properties, which are important considerations during drug development. These techniques can also help reduce the time and cost of traditional testing methods by identifying molecules that are likely to have poor ADMET properties and removing them before clinical testing.

Drug distribution property prediction, notably AI-based drug distribution property prediction, has made great strides in recent years, allowing researchers to drastically reduce the time and money spent on eliminating unsuitable drug candidates. An overview of the general structure of an AI-based drug distribution property prediction model using the ML and DL approaches is presented in Figure 3. In this section, we focus on some properties such as plasma protein binding (PPB), fraction unbound in plasma (F_u), blood–brain barrier (BBB), and volume of distribution (V_d). For every distribution property, we discuss why it should be predicted and the evolution of AI-based studies that have predicted it in recent years.

4.1. Blood–Brain Barrier Permeability Prediction

BBB refers to the unique features of the central nervous system's (CNS's) microvasculature [12]. Drugs that target the CNS must penetrate the BBB to reach their molecular target. In contrast, drugs with peripheral targets may require minimal or no BBB penetration to avoid adverse CNS effects. Furthermore, the BBB plays an important role in protecting the brain parenchyma from blood-borne pathogens and significantly interferes with the entry of drug and other exogenous compounds into the CNS [13]. The logarithmic ratio of drug concentrations in the brain and blood, log BB, is the most commonly used quantitative measure of a molecule's capability [14]. Experiments in vivo to determine log BB are complex and time consuming; the log BB data obtained are usually small with no high reliability. To overcome these disadvantages, recent AI-based BBB permeability prediction methods have focused on classifying whether a given compound is BBB permeable (BBB+) or not (BBB-) rather than on the log BB data.

Recently, AI-based approaches have increasingly been used to predict of BBB permeability (Table 1). Some commonly used ML algorithms, such as random forest (RF), support vector machine (SVM), k-nearest neighbor (k-NN), decision trees (DT), gradient boosting (GB), and extreme boosting system (XGB), achieve a prediction accuracy higher than 80% [15–19]. Notably, the LightBBB server [16] was developed by Shaker et al. They used the light gradient boost machine (LightGBM) algorithm, a large dataset of 7162 compounds collected from previous studies and 2432 1D and 2D descriptors. Using 10-fold cross-validation, they achieved an ACC of 89% and an AUC of 93%. This performance was lower than that of the mixed DL model on the same dataset, with an ACC of 92% and

an AUC of 96% [20]. However, they built a useful BBB prediction tool, that is available at <http://ssbio.cau.ac.kr/software/bbb> (accessed on 15 October 2022). Two ML-based ADMET predictors, admetSAR 2.0 [21] and FP-ADMET [22], also achieved very high BBB prediction results with an ACC of 90.7% and 81%, respectively.

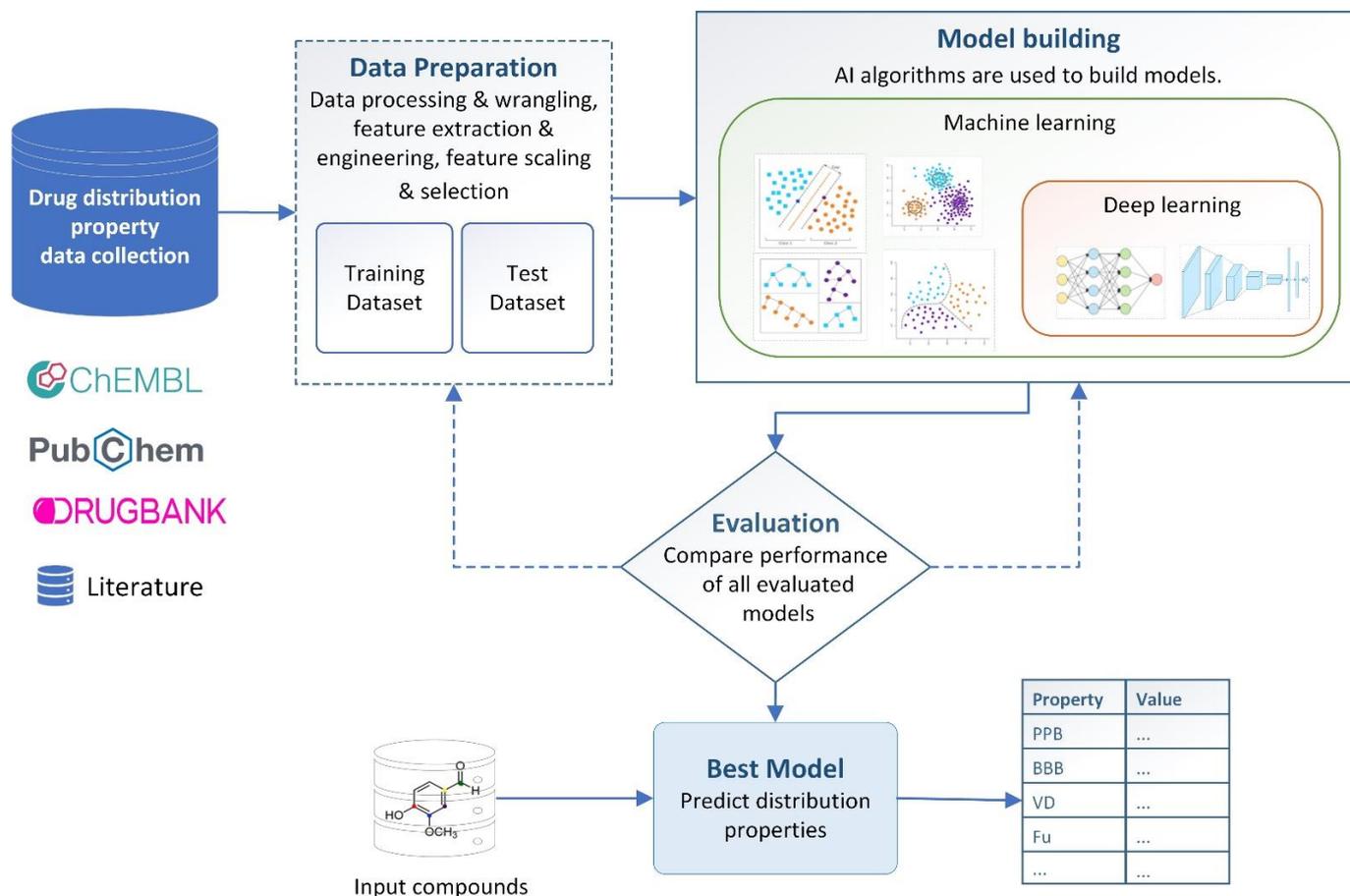


Figure 3. General structure of a drug distribution property prediction model using AI.

In the last few years, several researchers have presented various DL algorithms for BBB permeability prediction with excellent results, such as artificial neural networks (ANN), deep neural networks (DNN), convolutional neural networks (CNN), recurrent neural networks (RNN), and graph convolutional neural networks (GCNN) [23–31]. Alsenan et al. used an RNN model to improve the accuracy of BBB permeability prediction [29]. The dataset for their experiment contains 2350 compounds collected from Wang et al. [32] and 6394 descriptors and fingerprints for each compound. With an ACC of 96.53% and an AUC of 98.6%, the obtained results demonstrated that their RNN model solved the three identified issues of the previous BBB prediction model: imbalanced datasets, high dimensionality, and enhanced classifier performance. Their DL model also achieved the same performance using the same dataset [30].

Despite their outstanding performance, these models have a problem in common with other AI-based models: lack of interpretability [27]. This nature of the “black box” does not help researchers learn how to better design CNS drugs. To overcome this drawback, Yu’s research team developed a method to combine the strengths of ML and DL to produce a set of simple rules that are simple to understand and make predictions with better accuracy [27]. This is a hybrid method between the SVM and GCNN algorithms using a dataset of 940 drugs on the market and eight optimum descriptors with the highest essential scores. As a result, the novel hybrid ensemble model performed better than other traditional constitutive quantitative structure–activity relationship (QSAR) models, with an

ACC of 96% and an AUC of 98%. This hybrid model is not limited to CNS drug prediction but can also be used for other ADMET property predictions.

AI-based BBB permeability prediction models use smaller datasets and fewer features, which are often less reliable because they do not cover sufficient chemical diversity [16]. Therefore, reliable datasets that are sufficiently large to build models should be chosen.

Table 1. Summary of recent AI-based studies predicting BBB property.

Method	Data Sources	No. of Compounds	Performance	Ref.
SVM, RF, XGB	[32–35]	1970	AUC = 0.957, ACC = 0.910	[15]
LightGBM	[17,18,32,36–40]	7162	AUC = 0.94, ACC = 0.89	[16]
Mixed DL: Multilayer Perceptron (MLP), CNN	[16]	7162	AUC = 0.96, ACC = 0.92	[20]
RF, MLP, Sequential Minimal Optimization	[33,36,41]	2313	ACC = 0.88	[17]
Logistic Regression, DT, RF, GB	[42]	968	AUC = 0.78, ACC = 0.817	[18]
SVM, k-NN, DT, DNN	SIDER [40,43]	1000	ACC = 0.97, AUC = 0.98	[19]
Multichannel Substructure-Graph Gated Recurrent Unit Architecture	[37]	2053	AUC = 0.753	[23]
CNN	[37]	2039	AUC = 0.694	[24]
CNN	[35,44]	2254	ACC = 0.755, AUC = 0.784	[25]
CNN	[15,16,18,36,37,40]	7224	ACC = 0.74, AUC = 0.83	[45]
ANN	[46,47]	300	RMSE = 0.171	[26]
SVM and GCNN	[48]	940	ACC = 0.96, F1 score = 0.95	[27]
Fully Connected Neural Network, CNN	[37,40]	2264	AUC = 0.995	[28]
RNN	[32]	2350	ACC = 0.965, AUC = 0.98	[29]
DNN	[32]	2350	ACC = 0.962, AUC = 0.968	[30]
XGraphBoost	[38,49]	2039	AUC = 0.932	[31]

4.2. Plasma Protein Binding Prediction

One of the main mechanisms of drug absorption and distribution is through PPB. Therefore, drug binding to plasma proteins has a strong effect on the pharmacodynamic activity of drugs. PPB can directly affect oral bioavailability because free drug concentrations are at stake when the drug binds to serum proteins [50]. Figure 4 depicts the bi-dimensional interaction between drug–protein binding in the plasma, drug distribution, and drug elimination [51]. Many *in silico* predictive PPB models using ML and DL have been constructed using various datasets and evaluation units (Table 2). In this sense, *in silico* approaches can be cost effective, quick, and potent for screening large amounts of molecules, even without the need to synthesize the substance, as its structure suffices [52].

In an *in silico* strategy, Yuan et al. developed a QSAR model for predicting human PPB on a large dataset that was collected and curated from multiple studies over the past 15 years with 6741 compounds [53]. The QSAR model for different levels was constructed for the three corresponding descriptor sets: ADMET, Dragon, and the PaDEL, using five ML algorithms: RF, support vector regression (SVR), k-NN, boost tree (BT), and gradient-enhanced regression (GER). The best performance of their model is much higher than that of the previous models, with an MAE of 0.076 on the test set and an MAE falling to 0.041 at high binding (PPB > 0.8), 0.127 at moderate binding (PPB = 0.4–0.8), and 0.156 at low binding (PPB < 0.4). Their models performed well in the external evaluation set, which included 99 compounds from traditional Chinese medicine with an MAE value of 0.149.

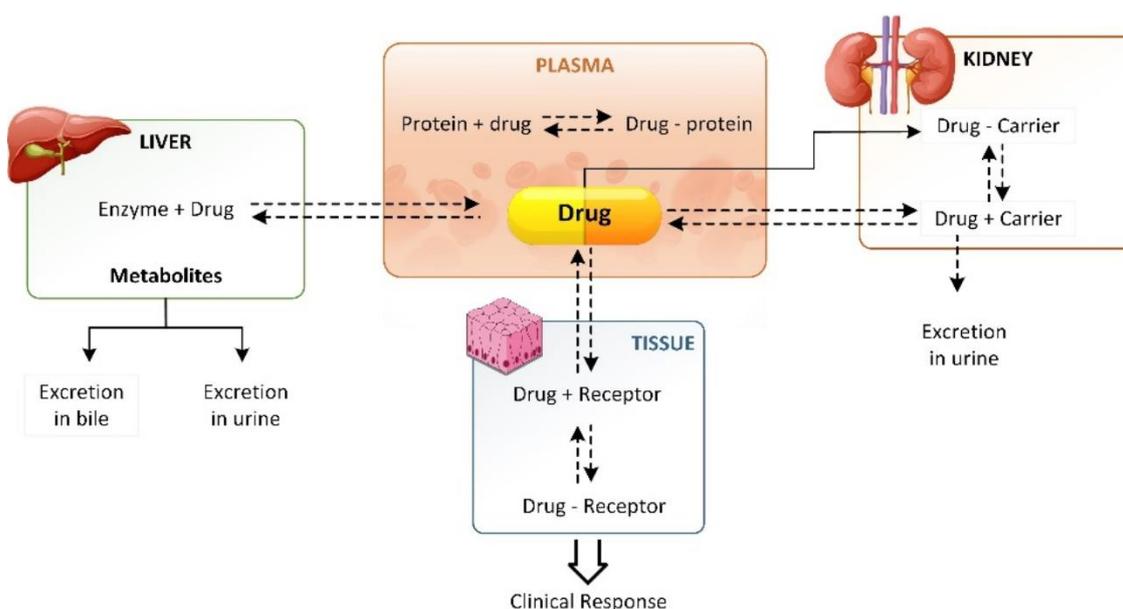


Figure 4. Relationship of reversible drug–protein binding in the plasma, drug distribution, and elimination.

Recently, Venkatraman et al. used the RF algorithm based on the fingerprint to predict PPB for 8103 compounds, achieving a balanced ACC of 84% and an AUC of 92% [22]. Xiong and co-authors developed an ADMET prediction-free web tool using the multitask graph attention framework [50]. They used 4712 compounds to predict PPB, and their model yielded a high-accuracy prediction with an R^2 of 0.733 and an RMSE of 0.135. An enhanced Graph Isomorphism Network (MolGIN) was proposed by Peng et al., utilizing the bond features and distinctions in the impact of atom neighbors for predicting ADMET properties [54]. The PPB dataset was collected by Wang et al. [55] contained 1830 compounds. The test results for an R^2 of 0.738 show that MolGIN is significantly superior to other baseline models (RF, graph neural network (GNN), DNN) in terms of efficiency measurement and achieves performance comparable to or superior to modern models (admetSAR 2.0 [21], ADMETLab 1.0 [6]) on the same dataset.

In 2022, Lou et al. proposed a new strategy for predicting and optimizing the human BBB for substances using an interpretable DL approach [56]. They used the attentive fingerprint algorithm, 3921 compounds, and Morgan fingerprints to develop an interpretable DL model. With an RMSE of 0.112 on the test set, their model showed promising predictive ability. Moreover, it could offer lead compounds with particular structural change plans to improve the PPB properties, unlike conventional QSAR models. Interpretable DL approaches allow us to understand why the model generates such predictions, helping us comprehend chemical pathways and rationally construct a structural change scheme.

Other PPB prediction studies are summarized in Table 2. In general, most models are built on different datasets; therefore, their performances cannot be compared with each other. Nevertheless, they will be helpful tools for assessing PPB during the process of drug design or structural modification.

Table 2. Summary of recent AI-based studies predicting PPB property.

Method	Data Sources	No. of Compounds	Performance	Ref.
RF	[57]	670	$R^2 = 0.74$, RMSE = 0.12	[52]
RF	[53,58]	8103	ACC = 0.84, AUC = 0.92	[22]
SVM	AstraZeneca in-house	100,550	RMSE = 0.444, $R^2 = 0.721$	[59]
k-NN, SVR, RF, BT, and GER	[55,60–69], ChEMBL and DrugBank	6741	MAE = 0.076	[53]
GCNN	[62]	1209	$R^2 = 0.668$, RMSE = 0.191	[21]
Multitask graph attention framework	ChEMBL, PubChem, OCHEM, Literature	4712	$R^2 = 0.733$, RMSE = 0.135	[50]
GNN	[61,62]	1744	$R^2 = 0.747$	[70]
MolGIN method	[55]	1830	$R^2 = 0.738$	[54]
GCNN, GAT	ChEMBL, PubChem, DrugBank, Literature	1830	$R^2 = 0.563$, RMSE = 0.211	[71]
Attentive fingerprint algorithm (GNN)	[56]	3921	$R^2 = 0.841$, RMSE = 0.112	[56]

4.3. Fraction Unbound in Plasma Prediction

In pharmacodynamic and pharmacokinetic studies, the F_u is a critical determinant of therapeutic efficacy. Most drugs in plasma are in an equilibrium state between unbound and bound to serum proteins [50]. The unbound fraction of the drug diffuses into tissues and is metabolized, or eliminated from the body [72]. In other words, only this fraction can be transferred to the sites of action across the membranes, whereas the bound fraction acts as a reservoir for the free drug concentration and prolongs the duration of action [73]. Distinct pharmacokinetic effects were observed as this fraction varied. The degree to which a drug attaches to proteins in the bloodstream may impair its efficacy; the more bound it is, the less efficiently it may pass cellular membranes or diffuse. F_u influences the renal glomerular filtration rate and hepatic metabolism. As a result, it affects the drug's volume of distribution and total clearance, both of which are critical elements in determining its pharmacokinetics [74]. Consequently, it is critical to make an accurate estimate of the F_u of drug candidates, particularly in low-value regions, throughout the drug development process. This section summarizes the progress of AI-based F_u prediction studies since 2019 (see Table 3).

Venkatraman investigated the efficiency of fingerprint-based RF models for predicting many ADMET-related properties [22]. In particular, his F_u prediction method achieved comparable or better performance compared with traditional 2D and 3D molecular descriptors on 2391 compounds with an R^2 of 0.63 and an RMSE of 0.44. The QSAR model was constructed by Wang et al. using several highly efficient, powerful, and widely used ML methods such as RF, SVM, GB, and XGB [75]. They used a dataset of 1352 drugs from a previous study [76], and the results once again proved that the RF model has a superior predictive power to the other methods in predicting F_u , with an R^2 of 0.818 and an RMSE of 0.291. Recently, Mulpuru and Mishra used a freely available automated ML framework, including AutoKeras, PyCaret, Auto-Sklearn, and TPOT, with chemical fingerprints to build F_u predictions [77]. Their best prediction model on a large dataset of 5471 compounds from ChEMBL was impressive, with an R^2 of 0.85, giving their model a significant contribution to ADMET modeling.

In addition to the ML techniques, DL techniques have also been exploited by many authors with the goal of improving prediction efficiency on large datasets. Zhou et al. built a QSAR model using DNNs and chemical fingerprints on 24 industrial ADME datasets from Lilly's in-house ADME assay with 9730 molecules to train the F_u prediction model [78].

However, the comparison results showed that their DNN model was not better than the SVM model on the same dataset, with an RMSE of 0.086 (SVM: RMSE = 0.083). In another study, Feinberg et al. built an ADMET prediction model with multitask deep featurization using a GCNN on a large dataset of 13,388 training compounds and 4462 testing compounds [79]. Their prediction model achieved significantly higher prediction accuracy than with the RF model, which had an R^2 of 0.919 (RF: R^2 = 0.582).

Table 3. Summary of recent AI-based studies predicting F_u property.

Method	Data Sources	No. of Compounds	Performance	Ref.
SVM, RF, GB, XGB	[76]	1352	R^2 = 0.82, RMSE = 0.291	[75]
AutoML Framework	ChEMBL v.27	5471	R^2 = 0.85, RMSE = 8.44	[77]
QSAR/Partial Least Squares (PLS) model	[69,80]	599	Q^2 = 0.69	[81]
DNNs	ADMET assays	9730	RMSE = 0.086	[78]
PotentialNet GCNNs	ADMET assays	17,850	R^2 = 0.919	[79]

4.4. Volume of Distribution Prediction

The volume of distribution (V_d) is a pharmacokinetic measure that indicates how long a drug will remain in the plasma or whether it will redistribute to other tissue compartments [82]. In other words, V_d is a theoretical concept of the dose used with actual initial concentrations in circulation, and it is a critical property for describing drug distribution in the human body [50]. V_d affects the half-life and duration of the activity of the compound at a steady state [83]. When two drugs have the same daily dose, the one with a lower V_d (shorter half-life) may require more frequent dosing (at lower individual doses) to attain a pharmacodynamic profile comparable to that with a higher V_d at a steady state. V_d is also a critical pharmacokinetic metric for determining the plasma concentration–time profile and half-life of drugs [84]. Figure 5 illustrates how to calculate V_d when we use three different drugs (A, B, and C) at a dose of 500 mg in an intuitive manner [85]. In addition, the table in Figure 5 shows the V_d values of some commonly used drugs. Several AI-based models have been successful in predicting V_d (Table 4).

Based on the ML technology, three authors used RF algorithms to develop V_d prediction models and made significant contributions in the past year [22,59,86]. Especially, the FP-ADMET prediction software was developed by Venkatraman [22]. This is a powerful tool for ADMET prediction based on fingerprints. The efficiency of the V_d prediction model for 1951 compounds were R^2 = 0.45 and RMSE = 0.51. AstraZeneca has nearly 20 years of development experience in AI-based ADME models [59]. AstraZeneca’s in-house data and models are updated regularly, and their accuracy increased over time. Their V_d prediction model achieved high accuracy with an R^2 of 0.67 and an RMSE of 0.371. Simeon et al. constructed a QSAR model for predicting V_d in humans, rats, dogs, mice, and monkeys using RF, PLS, and ANN algorithms [87]. Their models were built using physicochemical descriptors, electronic state descriptors, fingerprint descriptors, or a combination of physicochemical descriptors and one of the other two descriptors. Using the V_d human dataset of 1442 compounds, the RF model had a highly accurate prediction on the test set with an R^2 of 0.61 and an RMSE of 0.41.

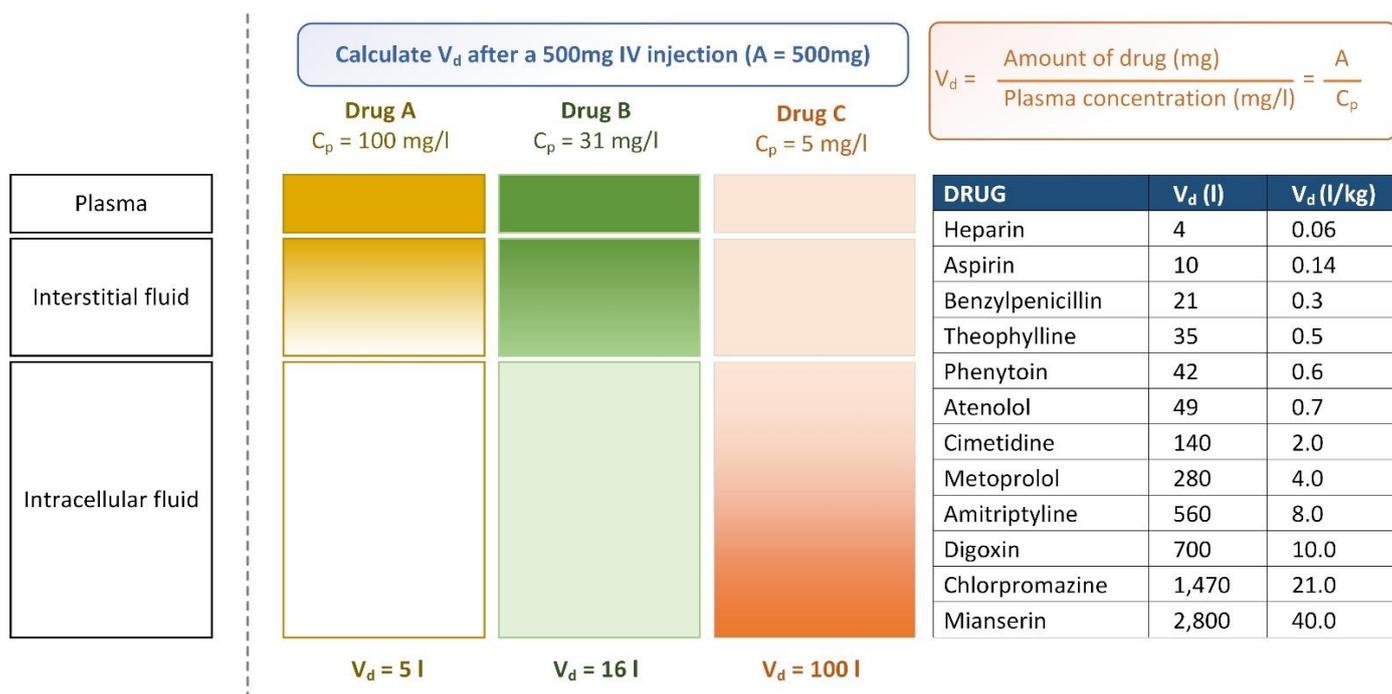


Figure 5. Examples of volume of distribution. Adapted from Maxwell [85].

In another study, Wang et al. used four ML algorithms, RF, SVM, GBM, and XGB, to develop a quantitative property–structure relationship (QSPR) model to predict V_d [75]. Their models were assessed by 10-fold cross-validation on a dataset containing 1352 drugs from Lomabardo et al. [76] using 209 selected features. The best-performing model was the SVM model, with an R^2 of 0.870 and an RMSE of 0.208. A new model called DeepPharm, using integrated transfer learning and multitask learning approaches, was developed by Ye et al. [88]. DeepPharm is more efficient than conventional ML methods such as PLS regression, SVM, ANN, RF, and k-NN on 412 molecules from the FDA, with an accuracy of 63.33% and a MAE of 0.175. To improve ADMET prediction, Feinberg and co-authors proposed a multitask deep featurization method applying GCNN using a large dataset containing 45,229 compounds for training and 15,076 compounds for testing [79]. However, compared with the RF method on the same dataset, this method did not improve significantly, with an R^2 of 0.525 (RF: $R^2 = 0.520$).

Table 4. Summary of recent AI-based studies predicting V_d property.

Method	Data Sources	No. of Compounds	Performance	Ref.
RF	[89]	1303	GMFE = 2.15 % < 2-fold = 54 % < 3-fold = 73	[86]
RF	AstraZeneca in-house	1440	RMSE = 0.371, $R^2 = 0.67$	[59]
SVM, RF, GB machine, XGB	[76]	1352	$R^2 = 0.87$, RMSE = 0.208	[75]
PLSANN, RF	ChEMBL [76,90]	1442	$R^2 = 0.61$, RMSE = 0.41	[87]
PLS regression, SVM, ANN, RF, k-NN, multitask learning feed-forward neural network, DeepPharm	Drugbank	412	ACC = 0.63, MAE = 0.174	[88]
PotentialNet GCNNs	ADMET assays	63,305	$R^2 = 0.525$	[79]

5. Public AI-Based ADMET Prediction Tools

With the continual collection of experimental ADMET data in recent years, many AI-based prediction tools for diverse endpoints have been developed to efficiently facilitate ADMET evaluation. More specifically, they can help researchers evaluate the ADMET properties in a time and money-saving manner, screen for undesirable compounds, and gather timely feedback on ADMET information for lead optimization. Moreover, they are also good support for distribution prediction researchers in developing and improving models. In Table 5, we list popular AI-based ADMET prediction tools that have been newly developed or updated in the last few years. These tools have built-in drug distribution property predictions and are available for free.

Table 5. Public AI-based ADMET prediction tools.

Name	No. of ADMET Prediction Models	Methods	Website *	Ref.
OECD QSAR Toolbox	902	QSAR	https://qsartoolbox.org/	[91]
iDrug ADMET prediction	60	AI	https://drug.ai.tencent.com/console/en/admet	
AdmetSAR 2.0	52	RF, SVM, k-NN	http://lmmd.ecust.edu.cn/admetSar2/	[21]
ADMETlab 2.0	67	GNN	https://admetmesh.scbdd.com/	[50]
Interpretable-ADMET	59	GCNN GAT	http://cadd.pharmacy.nankai.edu.cn/interpretableadmet/	[71]
HelixADMET	52	RF, GNN	https://paddlehelix.baidu.com/app/drug/admet/train	[70]
FP-ADMET	50	RF	https://gitlab.com/vishsoft/fpadmet	[22]
SwissADME	35	MLR, SVM, RNN, etc.	http://www.swissadme.ch/	[92]
vNN-ADMET	15	k-NN	https://vnnadmet.bhsai.org/	[93]
ICDrug ADMET	14	RF	www.icdrug.com/ICDrug/ADMET	[94]
Virtual Rat	12	RF, C5.0, DT	https://virtualrat.cmdm.tw/	[3]
LightBBB	1 (BBB)	Light GBM	http://ssbio.cau.ac.kr/software/bbb	[16]
Deep B ³	1 (BBB)	CNN	http://cpcb.cdutcm.edu.cn/deepb3/	[45]

* The websites were accessed on 10 October 2022.

In particular, we are interested in five publicly available ADMET predictors developed from 2019 to 2022, namely AdmetSAR 2.0 [21], ADMETLab 2.0 [50], FP-ADMET [22], Interpretable-ADMET [71], and HelixADMET [70]. They predicted most of the main ADMET-related properties (from 50 to 67 endpoints) and demonstrated good predictive performance. In Table 6, we analyze the predictive performance across the four distribution properties. These free and user-friendly tools can help ADMET researchers quickly and easily identify ADMET profiles for a wide range of drug candidates. Furthermore, these tools can serve as benchmarks for future ADMET studies. More interestingly, the Interpretable-ADMET tool helps optimize drug candidates with undesirable ADMET properties by automatically creating a new set of virtual candidates based on matching molecular pair rules.

Table 6. Performance of five ADMET prediction tools on distribution property prediction.

Property	Tool	Methods	No. of Compounds	Performance	
				AUC	R ²
BBB	AdmetSAR 2.0	SVM	1839	0.944	
	ADMETLab 2.0	GNN	1601	0.908	
	FP-ADMET	RF	7236	0.92	
	Interpretable-ADMET	GCNN & GAT	1830	0.897	
	HelixADMET	GNN	1791	0.944	
PPB	AdmetSAR 2.0	GCNN	1209		0.668
	ADMETLab 2.0	GNN	1573		0.733
	FP-ADMET	RF	8103	0.92	
	Interpretable-ADMET	GCNN & GAT	2044		0.563
	HelixADMET	GNN	1744		0.747
F _u	AdmetSAR 2.0	-	-	-	-
	ADMETLab 2.0	GNN	1494		0.763
	FP-ADMET	RF	2319		0.63
	Interpretable-ADMET	-	-	-	-
	HelixADMET	-	-	-	-
V _d	AdmetSAR 2.0	-	-	-	-
	ADMETLab 2.0	GNN	1399		0.782
	FP-ADMET	RF	1951		0.45
	Interpretable-ADMET	-	-	-	-
	HelixADMET	-	-	-	-

Overall, practical applications demonstrate that the tool is limited to qualitative analysis of chemicals and cannot accurately anticipate the quantitative values of certain properties [95]. Moreover, most reports indicate that these tools achieve very high or acceptable prediction accuracies. However, most predictive data have considerable uncertainty, and the decision is sensitive to a particular property [96]. We should choose tools with a larger amount of training data, higher accuracy, and higher citations, and use various tools to analyze data to make more accurate decisions.

6. Data Sources for Distribution Prediction Research

The success of an AI-based predictive model is highly dependent on the data and the modeling approach. The availability of an increasing number of public datasets on human pharmacokinetics facilitates the collection of a large number of structural substances and their associated experimental values for modeling purposes. A thorough understanding of the origin and reliability of the data is essential. Table 7 summarizes the most popular data sources for ADMET prediction research, specifically AI-based distribution prediction. In addition, researchers in the field of distribution prediction often use multiple datasets aggregated from different studies to develop and test their models. Interested readers may refer to the additional data sources from the literature provided in Section 3. Although the data sources for the study predicting the distribution were numerous, their quality was insufficient. Therefore, when using any source of test data to build a model, experts must carefully evaluate the certainty and reliability of the test. We should select reliable data sources, aggregate them, and use datasets that are sufficiently large for model training. Sharing more experimental data from pharmaceutical companies would be helpful to the

scientific community. We hope that further development of big data will bring promising prospects for future drug distribution research.

Table 7. Public data sources for AI-based distribution prediction research.

Name	Data Size (Compounds) *	Website *	Ref.
ZINC20	>750 million	https://zinc20.docking.org/	[97]
ChemSpider	115 million	http://www.chemspider.com/	[98]
PubChem	>111 million	https://pubchem.ncbi.nlm.nih.gov/	[99]
Therapeutics Data Commons	4,264,939	https://tdcommons.ai/	[100]
OCHEM 4.2	3,791,680	https://ochem.eu/home/show.do	[58]
openFDA	>3 million	https://open.fda.gov/	[101]
ChEMBL	>2.2 million	www.ebi.ac.uk/chembl/	[102]
GOSTAR	1.76 million	https://www.gostardb.com/	
BindingDB	>1 million	https://www.bindingdb.org/	[103]
Supernatural II	325,508	http://bioinformatics.charite.de/supernatural	[104]
NIST Chemistry WebBook	>70,000	http://webbook.nist.gov/	[105]
SIDER 4.1	55,730	http://sideeffects.embl.de/	[43]
ContaminantDB	>54,000	https://contaminantdb.ca/	
DrugBank 5.1.9	14,665	http://www.drugbank.ca/	[106]
IMPPAT 2.0	17,967	https://cb.imsc.res.in/imppat	[107]
KEGG	12,000	https://www.kegg.jp/	

* Data size and websites were accessed on 10 October 2022.

7. Challenges for AI-Based Distribution Prediction Researcher

The increasingly powerful AI technology in drug R&D presents not only many opportunities but also many challenges for researchers in predicting drug distribution and ADMET properties.

The first challenge is the lack of data quality [95,108–113]. Public data sources for drug R&D are undeniably increasing significantly; however, AI algorithms need not only the quantity of data but also the quality of data that are high enough to make accurate models. The chemicals tested should be sufficiently diverse to allow generation methods to cover the entire chemical search space [114]. Therefore, to solve this problem, it is important to collect high-quality data. Experts argue that more empirical data are required to create higher quality models and maximize the potential of AI-based applications [109]. However, in vivo and in vitro data collection is complex and limited [115]. Other problems related to the variability of the experiment, such as errors occurring in the process of data collection, management, and manipulation, also affect the quality of data. The statistical challenges and combination of diverse data with varying noise and bias are significant. Therefore, extracting and collecting high-quality data to train computational models is a laborious and challenging task that must be performed by experts. Recently, data sources from fields such as biology, chemistry, pharmacology, and clinical trials have been collected to build “big data” for drug R&D [116]; however, many obstacles still exist. Technical challenges, such as missing data, dimensional inaccuracies, and bias control, make big data analytics complex [117]. More time is needed to build a complete big data system for drug discovery and development. Meanwhile, there are many useful sources of data from proprietary pharmaceutical companies that are yet to be shared publicly with the research community [118]. Therefore, security and reasonable sharing policies are essential and contribute to solving data difficulties during this period.

The second challenge is model quality. In addition to data quality, a suitable learning model is required to harness the power of AI to predict distributional properties. First, it

requires researchers to have extensive knowledge of building AI-based models such as ML and DL algorithms. Without the expertise needed to build an effective data-mining project, researchers sometimes rely on incorrect methods that can lead to common errors or overly optimistic results [119]. AI-based models are evolving rapidly, and their complexity is increasing exponentially, requiring researchers to grasp new techniques quickly. Additionally, the current data landscape also necessitates the creation of powerful novel computational approaches capable of accurately predicting outcomes with diverse, large, multidimensional, and sparse data.

The third challenge faced by researchers is the difficulty in understanding the nature of AI models. Although the performance of AI-based distributed property prediction models is impressive, the mechanistic interpretation is still lacking. Therefore, it is difficult for scientists to assess the novelty or reliability of the hypothesis generated by AI because of its black-box nature, which hinders the improvement of the model and the optimization of compounds with undesirable distribution properties.

As in drug discovery, drug distribution is multidisciplinary. To study and build predictive models of distribution characteristics, researchers need to equip themselves with relevant knowledge in areas such as biology, bioinformatics, pharmacology, chemistry, and chemical informatics [120]. This is a big challenge for independent researchers. In fact, programmers and modelers who analyze huge datasets and build AI models often have a theoretical background and are ignorant of data-generating experiments and their flaws. AI experts are rarely chemists or biologists, especially structural representation specialists. Therefore, identifying potential mistakes in large datasets and interpreting the results of AI models remains difficult. Researchers must understand drug properties, endpoint roles, effects, metrics and assessments, structure–exposure–activity relationships, drug interactions, and other related knowledge aspects. Not all distribution properties are detrimental to all medicines. For instance, medications that target disorders of the central nervous system must typically be able to cross the BBB, although this trait is generally absent in other diseases. Therefore, collaboration between scientists is essential to ensure the correctness, effectiveness, and usability of drug property prediction models.

8. Conclusions and Future Perspectives

The application of AI to improve the drug R&D process is still in its infancy at this point [121]. For AI to reach the pinnacle of drug R&D, time and effort are required from multidisciplinary researchers. This is both an opportunity and a challenge for researchers. An accurate prediction of the distribution property is an important part of determining the ADMET profile of a drug candidate. Optimization of the distribution properties has a direct influence on the effectiveness and toxicity of the drug. Despite data challenges, the current efforts of AI-based, distributed property prediction models have made positive contributions, such as reducing costs and time in drug R&D.

One approach to overcome the challenges in AI-based drug distribution prediction is to focus on improving the quality and quantity of the data used for training and testing the models. This can be achieved by incorporating a wider range of data sources, including clinical trials, electronic health records, and real-world evidence. Additionally, implementing advanced data-cleaning and preprocessing techniques can help reduce the noise and bias in the data, leading to more accurate and reliable predictions. Another direction for future research in this field is the development of more sophisticated and robust AI algorithms that can handle complex and dynamic data. This includes the use of DL techniques, which have shown promising results in various medical and health applications. Additionally, it is crucial to ensure the interpretability of AI-based drug distribution predictions, which can be achieved by developing methods to visualize and interpret the underlying mechanisms and processes behind the model's predictions, thus, increasing trust and confidence in the results and enabling good decision making in drug development and personalized medicine. Furthermore, it is essential to incorporate human expertise and domain knowledge into the development and evaluation of AI-based drug

distribution prediction models. This can be achieved through collaboration between AI researchers and medical experts, such as pharmacologists and clinicians, to ensure that the models align with existing knowledge and practices in the field.

The core problem of AI systems is the process of “learning”. Good learning requires high-quality data and a high-quality learning model. “Quality” requires expert supervision. In the future, drug R&D data for learning will increase rapidly in quantity and complexity, requiring powerful AI-based predictive models such as DL. Ensuring the correctness and effectiveness of an AI-based predictive model requires close collaboration between multidisciplinary scientists. The contribution and sharing of data from pharmaceutical companies and academic researchers will accelerate the development of big data. In the next decade, data, computation, and multidisciplinary scientists will become highly connected to AI-based drug R&D. There is a continuous feedback loop between interpretable AI and experimental biology. Through incremental improvements to workflow and comprehensible insights, researchers can track, evaluate, and construct better prediction models.

In this study, the recent development of AI-based distribution property prediction models were analyzed and synthesized. Although each model has its own limitations, the models show a remarkable effort by researchers. The basics of distribution and the role of endpoints, along with related resources such as public data sources and free prediction tools, are provided. We hope that this is a useful document for researchers who develop and improve distribution property prediction models and other properties of ADMET based on AI.

Author Contributions: Conceptualization, H.T. and T.T.V.T.; Methodology, T.T.V.T.; Validation, T.T.V.T.; H.T. and K.T.C.; Resources, H.T. and K.T.C.; Writing—Original Draft, T.T.V.T.; Writing—Review & Editing, H.T. and K.T.C.; Supervision, H.T. and K.T.C.; Project Administration, H.T. and K.T.C.; Funding Acquisition, H.T. and K.T.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by a National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. 2020R1A2C2005612) and (No. 2022R1G1A1004613).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wouters, O.J.; McKee, M.; Luyten, J. Estimated research and development investment needed to bring a new medicine to market, 2009–2018. *JAMA* **2020**, *323*, 844–853. [[CrossRef](#)] [[PubMed](#)]
2. Luo, Y.; Peng, J.; Ma, J. Next Decade’s AI-Based Drug Development Features Tight Integration of Data and Computation. *Health Data Sci.* **2022**, *2022*, 9816939. [[CrossRef](#)]
3. Hsiao, Y.; Su, B.H.; Tseng, Y.J. Current development of integrated web servers for preclinical safety and pharmacokinetics assessments in drug development. *Brief. Bioinform.* **2021**, *22*, bbaa160. [[CrossRef](#)] [[PubMed](#)]
4. Sun, D.; Gao, W.; Hu, H.; Zhou, S. Why 90% of clinical drug development fails and how to improve it? *Acta Pharm. Sin. B* **2022**, *12*, 3049–3062. [[CrossRef](#)] [[PubMed](#)]
5. Ksir, C.J.; Carl, L.; Hart, D. *Drugs, Society, and Human Behavior*; McGraw-Hill Education: New York, NY, USA, 2017.
6. Dong, J.; Wang, N.N.; Yao, Z.J.; Zhang, L.; Cheng, Y.; Ouyang, D.; Lu, A.P.; Cao, D.S. ADMETlab: A platform for systematic ADMET evaluation based on a comprehensively collected ADMET database. *J. Cheminformatics* **2018**, *10*, 29. [[CrossRef](#)]
7. Onetto, A.J.; Sharif, S. *Drug Distribution*; StatPearls Publishing: Treasure Island, FL, USA, 2022.
8. Brahmanekar, D.M.; Jaiswal, S.B. Distribution of drugs. In *Biopharmaceutics and Pharmacokinetics: A Treatise*, 3rd ed.; Vallabh Prakashan: Delhi, India, 2019.
9. Rifaioglu, A.S.; Atas, H.; Martin, M.J.; Cetin-Atalay, R.; Atalay, V.; Doğan, T. Recent applications of deep learning and machine intelligence on in silico drug discovery: Methods, tools and databases. *Brief. Bioinform.* **2019**, *20*, 1878–1912. [[CrossRef](#)]
10. Worley, B.; Powers, R. PCA as a practical indicator of OPLS-DA model reliability. *Curr. Metab.* **2016**, *4*, 97–103. [[CrossRef](#)]

11. Avci, O.; Abdeljaber, O.; Kiranyaz, S.; Hussein, M.; Gabbouj, M.; Inman, D.J. A review of vibration-based damage detection in civil structures: From traditional methods to Machine Learning and Deep Learning applications. *Mech. Syst. Signal Process.* **2021**, *147*, 107077. [[CrossRef](#)]
12. Daneman, R.; Prat, A. The blood-brain barrier. *Cold Spring Harb. Perspect. Biol.* **2015**, *7*, a020412. [[CrossRef](#)]
13. Kadry, H.; Noorani, B.; Cucullo, L. A blood-brain barrier overview on structure, function, impairment, and biomarkers of integrity. *Fluids Barriers CNS* **2020**, *17*, 69. [[CrossRef](#)]
14. Bickel, U. How to measure drug transport across the blood-brain barrier. *Neurotherapeutics* **2005**, *2*, 15–26. [[CrossRef](#)] [[PubMed](#)]
15. Liu, L.L.; Zhang, L.; Feng, H.W.; Li, S.M.; Liu, M.; Zhao, J.; Liu, H.S. Prediction of the Blood-Brain Barrier (BBB) Permeability of Chemicals Based on Machine-Learning and Ensemble Methods. *Chem. Res. Toxicol.* **2021**, *34*, 1456–1467. [[CrossRef](#)] [[PubMed](#)]
16. Shaker, B.; Yu, M.S.; Song, J.S.; Ahn, S.; Ryu, J.Y.; Oh, K.S.; Na, D. LightBBB: Computational prediction model of blood-brain-barrier penetration based on LightGBM. *Bioinformatics* **2021**, *37*, 1135–1139. [[CrossRef](#)] [[PubMed](#)]
17. Singh, M.; Divakaran, R.; Konda, L.S.K.; Kristam, R. A classification model for blood brain barrier penetration. *J. Mol. Graph. Model.* **2020**, *96*, 107516. [[CrossRef](#)] [[PubMed](#)]
18. Plisson, F.; Piggott, A.M. Predicting blood–brain barrier permeability of marine-derived kinase inhibitors using ensemble classifiers reveals potential hits for neurodegenerative disorders. *Mar. Drugs* **2019**, *17*, 81. [[CrossRef](#)]
19. Miao, R.; Xia, L.Y.; Chen, H.H.; Huang, H.H.; Liang, Y. Improved Classification of Blood-Brain-Barrier Drugs Using Deep Learning. *Sci. Rep.* **2019**, *9*, 1–11. [[CrossRef](#)]
20. Cherian Parakkal, S.; Datta, R.; Das, D. DeepBBBP: High Accuracy Blood-Brain-Barrier Permeability Prediction with a Mixed Deep Learning Model. *Mol. Inform.* **2022**, *41*, 2100315. [[CrossRef](#)]
21. Yang, H.; Lou, C.; Sun, L.; Li, J.; Cai, Y.; Wang, Z.; Li, W.; Liu, G.; Tang, Y. admetSAR 2.0: Web-service for prediction and optimization of chemical ADMET properties. *Bioinformatics* **2019**, *35*, 1067–1069. [[CrossRef](#)]
22. Venkatraman, V. FP-ADMET: A compendium of fingerprint-based ADMET prediction models. *J. Cheminformatics* **2021**, *13*, 75. [[CrossRef](#)]
23. Wang, S.; Li, Z.; Zhang, S.G.; Jiang, M.J.; Wang, X.F.; Wei, Z.Q. Molecular Property Prediction Based on a Multichannel Substructure Graph (vol 8, pg 18601, 2020). *IEEE Access* **2020**, *8*, 127968. [[CrossRef](#)]
24. Chen, J.-H.; Tseng, Y.J. A general optimization protocol for molecular property prediction using a deep learning network. *Brief. Bioinform.* **2022**, *23*, bbab367. [[CrossRef](#)] [[PubMed](#)]
25. Shi, T.T.; Yang, Y.W.; Huang, S.H.; Chen, L.X.; Kuang, Z.Y.; Heng, Y.; Mei, H. Molecular image-based convolutional neural network for the prediction of ADMET properties. *Chemom. Intell. Lab. Syst.* **2019**, *194*, 103853. [[CrossRef](#)]
26. Wu, Z.; Xian, Z.; Ma, W.; Liu, Q.; Huang, X.; Xiong, B.; He, S.; Zhang, W. Artificial neural network approach for predicting blood brain barrier permeability based on a group contribution method. *Comput. Methods Programs Biomed.* **2021**, *200*, 105943. [[CrossRef](#)] [[PubMed](#)]
27. Yu, T.H.; Su, B.H.; Battalora, L.C.; Liu, S.; Tseng, Y.J. Ensemble modeling with machine learning and deep learning to provide interpretable generalized rules for classifying CNS drugs with high prediction power. *Brief. Bioinform.* **2022**, *23*, bbab377. [[CrossRef](#)] [[PubMed](#)]
28. Achiaa Atwereboannah, A.; Wu, W.-P.; Nanor, E. Prediction of Drug Permeability to the Blood-Brain Barrier using Deep Learning. In Proceedings of the 4th International Conference on Biometric Engineering and Applications, Taiyuan, China, 25–27 May 2021; pp. 104–109.
29. Alsenan, S.; Al-Turaiki, I.; Hafez, A. A Recurrent Neural Network model to predict blood-brain barrier permeability. *Comput. Biol. Chem.* **2020**, *89*, 107377. [[CrossRef](#)]
30. Alsenan, S.A.; Al-Turaiki, I.M.; Hafez, A.M. Feature extraction methods in quantitative structure–activity relationship modeling: A comparative study. *IEEE Access* **2020**, *8*, 78737–78752. [[CrossRef](#)]
31. Deng, D.G.; Chen, X.W.; Zhang, R.C.; Lei, Z.R.; Wang, X.J.; Zhou, F.F. XGraphBoost: Extracting Graph Neural Network-Based Features for a Better Prediction of Molecular Properties (vol 61, pg 2697, 2021). *J. Chem. Inf. Model.* **2021**, *61*, 4820–4822. [[CrossRef](#)]
32. Wang, Z.; Yang, H.; Wu, Z.; Wang, T.; Li, W.; Tang, Y.; Liu, G. In silico prediction of blood–brain barrier permeability of compounds by machine learning and resampling methods. *ChemMedChem* **2018**, *13*, 2189–2201. [[CrossRef](#)]
33. Wang, W.Y.; Kim, M.T.; Sedykh, A.; Zhu, H. Developing Enhanced Blood-Brain Barrier Permeability Models: Integrating External Bio-Assay Data in QSAR Modeling. *Pharm. Res.* **2015**, *32*, 3055–3065. [[CrossRef](#)]
34. Zhao, Y.H.; Abraham, M.H.; Ibrahim, A.; Fish, P.V.; Cole, S.; Lewis, M.L.; de Groot, M.J.; Reynolds, D.P. Predicting penetration across the blood-brain barrier from simple descriptors and fragmentation schemes. *J. Chem. Inf. Model.* **2007**, *47*, 170–175. [[CrossRef](#)]
35. Li, H.; Yap, C.W.; Ung, C.Y.; Xue, Y.; Cao, Z.W.; Chen, Y.Z. Effect of selection of molecular descriptors on the prediction of blood-brain barrier penetrating and nonpenetrating agents by statistical learning methods. *J. Chem. Inf. Model.* **2005**, *45*, 1376–1384. [[CrossRef](#)] [[PubMed](#)]
36. Adenot, M.; Lahana, R. Blood-brain barrier permeation models: Discriminating between potential CNS and non-CNS drugs including P-glycoprotein substrates. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 239–248. [[CrossRef](#)] [[PubMed](#)]
37. Martins, I.F.; Teixeira, A.L.; Pinheiro, L.; Falcao, A.O. A Bayesian approach to in silico blood-brain barrier penetration modeling. *J. Chem. Inf. Model.* **2012**, *52*, 1686–1697. [[CrossRef](#)] [[PubMed](#)]
38. Wu, Z.; Ramsundar, B.; Feinberg, E.N.; Gomes, J.; Geniesse, C.; Pappu, A.S.; Leswing, K.; Pande, V. MoleculeNet: A benchmark for molecular machine learning. *Chem. Sci.* **2018**, *9*, 513–530. [[CrossRef](#)] [[PubMed](#)]
39. Yuan, Y.; Zheng, F.; Zhan, C.-G. Improved prediction of blood–brain barrier permeability through machine learning with combined use of molecular property-based descriptors and fingerprints. *AAPS J.* **2018**, *20*, 54. [[CrossRef](#)] [[PubMed](#)]

40. Gao, Z.; Chen, Y.; Cai, X.; Xu, R. Predict drug permeability to blood-brain-barrier from clinical phenotypes: Drug side effects and drug indications. *Bioinformatics* **2017**, *33*, 901–908. [[CrossRef](#)]
41. Brito-Sanchez, Y.; Marrero-Ponce, Y.; Barigye, S.J.; Yaber-Goenaga, I.; Perez, C.M.; Huong, L.T.T.; Cherkasov, A. Towards Better BBB Passage Prediction Using an Extensive and Curated Data Set. *Mol. Inform.* **2015**, *34*, 308–330. [[CrossRef](#)]
42. Chico, L.K.; Van Eldik, L.J.; Watterson, D.M. Targeting protein kinases in central nervous system disorders. *Nat. Rev. Drug Discov.* **2009**, *8*, 892–909. [[CrossRef](#)]
43. Kuhn, M.; Letunic, I.; Jensen, L.J.; Bork, P. The SIDER database of drugs and side effects. *Nucleic Acids Res.* **2016**, *44*, D1075–D1079. [[CrossRef](#)]
44. Shen, J.; Cheng, F.; Xu, Y.; Li, W.; Tang, Y. Estimation of ADME properties with substructure pattern recognition. *J. Chem. Inf. Model.* **2010**, *50*, 1034–1041. [[CrossRef](#)]
45. Tang, Q.; Nie, F.; Zhao, Q.; Chen, W. A merged molecular representation deep learning method for blood–brain barrier permeability prediction. *Brief. Bioinform.* **2022**, *23*, bbac357. [[CrossRef](#)] [[PubMed](#)]
46. Kouskoura, M.G.; Piteni, A.I.; Markopoulou, C.K. A new descriptor via bio-mimetic chromatography and modeling for the blood brain barrier (Part II). *J. Pharm. Biomed. Anal.* **2019**, *164*, 808–817. [[CrossRef](#)] [[PubMed](#)]
47. Toropov, A.A.; Toropova, A.P.; Beeg, M.; Gobbi, M.; Salmona, M. QSAR model for blood-brain barrier permeation. *J. Pharmacol. Toxicol. Methods* **2017**, *88*, 7–18. [[CrossRef](#)] [[PubMed](#)]
48. Ghose, A.K.; Herbertz, T.; Hudkins, R.L.; Dorsey, B.D.; Mallamo, J.P. Knowledge-based, central nervous system (CNS) lead selection and lead optimization for CNS drug discovery. *ACS Chem. Neurosci.* **2012**, *3*, 50–68. [[CrossRef](#)]
49. Mayr, A.; Klambauer, G.; Unterthiner, T.; Steijaert, M.; Wegner, J.K.; Ceulemans, H.; Clevert, D.-A.; Hochreiter, S. Large-scale comparison of machine learning methods for drug target prediction on ChEMBL. *Chem. Sci.* **2018**, *9*, 5441–5451. [[CrossRef](#)]
50. Xiong, G.; Wu, Z.; Yi, J.; Fu, L.; Yang, Z.; Hsieh, C.; Yin, M.; Zeng, X.; Wu, C.; Lu, A.; et al. ADMETlab 2.0: An integrated online platform for accurate and comprehensive predictions of ADMET properties. *Nucleic Acids Res.* **2021**, *49*, W5–W14. [[CrossRef](#)]
51. Venkateswarlu, V. Distribution of Drugs. In *Biopharmaceutics and Pharmacokinetics*; PharmaMed Press: Watford, UK, 2008; p. 77.
52. Toma, C.; Gadaleta, D.; Roncaglioni, A.; Toropov, A.; Toropova, A.; Marzo, M.; Benfenati, E. QSAR Development for Plasma Protein Binding: Influence of the Ionization State. *Pharm. Res.* **2019**, *36*, 28. [[CrossRef](#)]
53. Yuan, Y.W.; Chang, S.; Zhang, Z.; Li, Z.G.; Li, S.Z.; Xie, P.; Yau, W.P.; Lin, H.S.; Cai, W.M.; Zhang, Y.C.; et al. A novel strategy for prediction of human plasma protein binding using machine learning techniques. *Chemom. Intell. Lab. Syst.* **2020**, *199*, 103962. [[CrossRef](#)]
54. Peng, Y.Z.; Lin, Y.M.; Jing, X.Y.; Zhang, H.; Huang, Y.R.; Luo, G.S. Enhanced Graph Isomorphism Network for Molecular ADMET Properties Prediction. *IEEE Access* **2020**, *8*, 168344–168360. [[CrossRef](#)]
55. Wang, N.-N.; Deng, Z.-K.; Huang, C.; Dong, J.; Zhu, M.-F.; Yao, Z.-J.; Chen, A.F.; Lu, A.-P.; Mi, Q.; Cao, D.-S. ADME properties evaluation in drug discovery: Prediction of plasma protein binding using NSGA-II combining PLS and consensus modeling. *Chemom. Intell. Lab. Syst.* **2017**, *170*, 84–95. [[CrossRef](#)]
56. Lou, C.; Yang, H.; Wang, J.; Huang, M.; Li, W.; Liu, G.; Lee, P.W.; Tang, Y. IDL-PPBopt: A Strategy for Prediction and Optimization of Human Plasma Protein Binding of Compounds via an Interpretable Deep Learning Method. *J. Chem. Inf. Model.* **2022**, *62*, 2788–2799. [[CrossRef](#)] [[PubMed](#)]
57. Obach, R.S.; Lombardo, F.; Waters, N.J. Trend analysis of a database of intravenous pharmacokinetic parameters in humans for 670 drug compounds. *Drug Metab. Dispos.* **2008**, *36*, 1385–1405. [[CrossRef](#)]
58. Sushko, I.; Novotarskyi, S.; Körner, R.; Pandey, A.K.; Rupp, M.; Teetz, W.; Brandmaier, S.; Abdelaziz, A.; Prokopenko, V.V.; Tanchuk, V.Y. Online chemical modeling environment (OCHEM): Web platform for data storage, model development and publishing of chemical information. *J. Comput.-Aided Mol. Des.* **2011**, *25*, 533–554. [[CrossRef](#)] [[PubMed](#)]
59. Oprisiu, I.; Winiwarter, S. In *Silico ADME Modeling*; Academic Press: Cambridge, MA, USA, 2021; pp. 208–222. [[CrossRef](#)]
60. Zhivkova, Z. Quantitative Structure—Pharmacokinetics Relationships for Plasma Protein Binding of Basic Drugs. *J. Pharm. Pharm. Sci.* **2017**, *20*, 349–359. [[CrossRef](#)] [[PubMed](#)]
61. Votano, J.R.; Parham, M.; Hall, L.M.; Hall, L.H.; Kier, L.B.; Oloff, S.; Tropsha, A. QSAR modeling of human serum protein binding with several modeling techniques utilizing structure— information representation. *J. Med. Chem.* **2006**, *49*, 7169–7181. [[CrossRef](#)] [[PubMed](#)]
62. Sun, L.; Yang, H.; Li, J.; Wang, T.; Li, W.; Liu, G.; Tang, Y. In silico prediction of compounds binding to human plasma proteins by QSAR models. *ChemMedChem* **2018**, *13*, 572–581. [[CrossRef](#)]
63. Zhu, X.-W.; Sedykh, A.; Zhu, H.; Liu, S.-S.; Tropsha, A. The use of pseudo-equilibrium constant affords improved QSAR models of human plasma protein binding. *Pharm. Res.* **2013**, *30*, 1790–1798. [[CrossRef](#)]
64. Watanabe, R.; Esaki, T.; Kawashima, H.; Natsume-Kitatani, Y.; Nagao, C.; Ohashi, R.; Mizuguchi, K. Predicting fraction unbound in human plasma from chemical structure: Improved accuracy in the low value ranges. *Mol. Pharm.* **2018**, *15*, 5302–5311. [[CrossRef](#)]
65. Douguet, D. Data sets representative of the structures and experimental properties of FDA-approved drugs. *ACS Med. Chem. Lett.* **2018**, *9*, 204–209. [[CrossRef](#)]
66. Tajimi, T.; Wakui, N.; Yanagisawa, K.; Yoshikawa, Y.; Ohue, M.; Akiyama, Y. Computational prediction of plasma protein binding of cyclic peptides from small molecule experimental data using sparse modeling techniques. *BMC Bioinform.* **2018**, *19*, 157–170. [[CrossRef](#)]
67. Ingle, B.L.; Veber, B.C.; Nichols, J.W.; Tornero-Velez, R. Informing the human plasma protein binding of environmental chemicals by machine learning in the pharmaceutical space: Applicability domain and limits of predictability. *J. Chem. Inf. Model.* **2016**, *56*, 2243–2252. [[CrossRef](#)]
68. Li, H.; Chen, Z.; Xu, X.; Sui, X.; Guo, T.; Liu, W.; Zhang, J. Predicting human plasma protein binding of drugs using plasma protein interaction QSAR analysis (PPI-QSAR). *Biopharm. Drug Dispos.* **2011**, *32*, 333–342. [[CrossRef](#)]

69. Zhivkova, Z.; Doytchinova, I. Quantitative structure—Plasma protein binding relationships of acidic drugs. *J. Pharm. Sci.* **2012**, *101*, 4627–4641. [CrossRef]
70. Zhang, S.; Yan, Z.; Huang, Y.; Liu, L.; He, D.; Wang, W.; Fang, X.; Zhang, X.; Wang, F.; Wu, H.; et al. HelixADMET: A robust and endpoint extensible ADMET system incorporating self-supervised knowledge transfer. *Bioinformatics* **2022**. [CrossRef]
71. 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* **2022**, *38*, 2863–2871. [CrossRef]
72. Roberts, J.A.; Pea, F.; Lipman, J. The Clinical Relevance of Plasma Protein Binding Changes. *Clin. Pharmacokinet.* **2013**, *52*, 1–8. [CrossRef]
73. Seyfinejad, B.; Ozkan, S.A.; Jouyban, A. Recent advances in the determination of unbound concentration and plasma protein binding of drugs: Analytical methods. *Talanta* **2021**, *225*, 122052. [CrossRef]
74. Bohnert, T.; Gan, L.-S. Plasma protein binding: From discovery to development. *J. Pharm. Sci.* **2013**, *102*, 2953–2994. [CrossRef]
75. Wang, Y.C.; Liu, H.C.; Fan, Y.R.; Chen, X.Y.; Yang, Y.; Zhu, L.; Zhao, J.N.; Chen, Y.D.; Zhang, Y.M. In Silico Prediction of Human Intravenous Pharmacokinetic Parameters with Improved Accuracy. *J. Chem. Inf. Model.* **2019**, *59*, 3968–3980. [CrossRef]
76. Lombardo, F.; Berellini, G.; Obach, R.S. Trend analysis of a database of intravenous pharmacokinetic parameters in humans for 1352 drug compounds. *Drug Metab. Dispos.* **2018**, *46*, 1466–1477. [CrossRef]
77. Mulpuru, V.; Mishra, N. In Silico Prediction of Fraction Unbound in Human Plasma from Chemical Fingerprint Using Automated Machine Learning. *Acs Omega* **2021**, *6*, 6791–6797. [CrossRef]
78. Zhou, Y.D.; Cahya, S.; Combs, S.A.; Nicolaou, C.A.; Wang, J.B.; Desai, P.V.; Shen, J. Exploring Tunable Hyperparameters for Deep Neural Networks with Industrial ADME Data Sets. *J. Chem. Inf. Model.* **2019**, *59*, 1005–1016. [CrossRef] [PubMed]
79. Feinberg, E.N.; Joshi, E.; Pande, V.S.; Cheng, A.C. Improvement in ADMET Prediction with Multitask Deep Featurization. *J. Med. Chem.* **2020**, *63*, 8835–8848. [CrossRef]
80. Yamagata, T.; Zanelli, U.; Gallemann, D.; Perrin, D.; Dolgos, H.; Petersson, C. Comparison of methods for the prediction of human clearance from hepatocyte intrinsic clearance for a set of reference compounds and an external evaluation set. *Xenobiotica* **2017**, *47*, 741–751. [CrossRef]
81. Fagerholm, U.; Spjuth, O.; Hellberg, S. Comparison between lab variability and in silico prediction errors for the unbound fraction of drugs in human plasma. *Xenobiotica* **2021**, *51*, 1095–1100. [CrossRef]
82. Mansoor, A.; Mahabadi, N. *Volume of Distribution*; StatPearls Publishing: Treasure Island, FL, USA, 2022.
83. Smith, D.A.; Beaumont, K.; Maurer, T.S.; Di, L. Volume of Distribution in Drug Design. *J. Med. Chem.* **2015**, *58*, 5691–5698. [CrossRef]
84. Hsu, F.; Chen, Y.C.; Broccatelli, F. Evaluation of Tissue Binding in Three Tissues across Five Species and Prediction of Volume of Distribution from Plasma Protein and Tissue Binding with an Existing Model. *Drug Metab. Dispos.* **2021**, *49*, 330–336. [CrossRef]
85. Maxwell, S. How Are Drugs Distributed around the Body. Available online: <https://vimeo.com/469366240> (accessed on 11 November 2022).
86. Lombardo, F.; Bentzien, J.; Berellini, G.; Muegge, I. In Silico Models of Human PK Parameters. Prediction of Volume of Distribution Using an Extensive Data Set and a Reduced Number of Parameters. *J. Pharm. Sci.* **2021**, *110*, 500–509. [CrossRef]
87. Simeon, S.; Montanari, D.; Gleeson, M.P. Investigation of Factors Affecting the Performance of in silico Volume Distribution QSAR Models for Human, Rat, Mouse, Dog & Monkey. *Mol. Inform.* **2019**, *38*, e1900059. [CrossRef]
88. Ye, Z.; Yang, Y.; Li, X.; Cao, D.; Ouyang, D. An Integrated Transfer Learning and Multitask Learning Approach for Pharmacokinetic Parameter Prediction. *Mol. Pharm.* **2019**, *16*, 533–541. [CrossRef]
89. Lombardo, F.; Jing, Y.K. In Silico Prediction of Volume of Distribution in Humans. Extensive Data Set and the Exploration of Linear and Nonlinear Methods Coupled with Molecular Interaction Fields Descriptors. *J. Chem. Inf. Model.* **2016**, *56*, 2042–2052. [CrossRef]
90. Bento, A.P.; Gaulton, A.; Hersey, A.; Bellis, L.J.; Chambers, J.; Davies, M.; Krüger, F.A.; Light, Y.; Mak, L.; McGlinchey, S. The ChEMBL bioactivity database: An update. *Nucleic Acids Res.* **2014**, *42*, D1083–D1090. [CrossRef] [PubMed]
91. Schultz, T.W.; Diderich, R.; Kuseva, C.D.; Mekenyan, O.G. The OECD QSAR toolbox starts its second decade. In *Computational Toxicology*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 55–77.
92. Daina, A.; Michielin, O.; Zoete, V. SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* **2017**, *7*, 42717. [CrossRef]
93. Schyman, P.; Liu, R.; Desai, V.; Wallqvist, A. vNN web server for ADMET predictions. *Front. Pharmacol.* **2017**, *8*, 889. [CrossRef]
94. Wei, M.; Zhang, X.; Pan, X.; Wang, B.; Ji, C.; Qi, Y.; Zhang, J.Z. HobPre: Accurate prediction of human oral bioavailability for small molecules. *J. Cheminformatics* **2022**, *14*, 1. [CrossRef]
95. Wu, F.; Zhou, Y.; Li, L.; Shen, X.; Chen, G.; Wang, X.; Liang, X.; Tan, M.; Huang, Z. Computational Approaches in Preclinical Studies on Drug Discovery and Development. *Front. Chem.* **2020**, *8*, 726. [CrossRef]
96. Jia, C.Y.; Li, J.Y.; Hao, G.F.; Yang, G.F. A drug-likeness toolbox facilitates ADMET study in drug discovery. *Drug Discov. Today* **2020**, *25*, 248–258. [CrossRef]
97. Irwin, J.J.; Tang, K.G.; Young, J.; Dandarchuluun, C.; Wong, B.R.; Khurelbaatar, M.; Moroz, Y.S.; Mayfield, J.; Sayle, R.A. ZINC20—A free ultralarge-scale chemical database for ligand discovery. *J. Chem. Inf. Model.* **2020**, *60*, 6065–6073. [CrossRef]
98. Pence, H.E.; Williams, A. ChemSpider: An Online Chemical Information Resource. *J. Chem. Educ.* **2010**, *87*, 1123–1124. [CrossRef]
99. Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B.A.; Thiessen, P.A.; Yu, B. PubChem 2019 update: Improved access to chemical data. *Nucleic Acids Res.* **2019**, *47*, D1102–D1109. [CrossRef]
100. Huang, K.; Fu, T.; Gao, W.; Zhao, Y.; Roohani, Y.; Leskovec, J.; Coley, C.W.; Xiao, C.; Sun, J.; Zitnik, M. Artificial intelligence foundation for therapeutic science. *Nat. Chem. Biol.* **2022**, *18*, 1033–1036. [CrossRef] [PubMed]

101. Kass-Hout, T.A.; Xu, Z.; Mohebbi, M.; Nelsen, H.; Baker, A.; Levine, J.; Johanson, E.; Bright, R.A. OpenFDA: An innovative platform providing access to a wealth of FDA's publicly available data. *J. Am. Med. Inform. Assoc.* **2016**, *23*, 596–600. [[CrossRef](#)]
102. Mendez, D.; Gaulton, A.; Bento, A.P.; Chambers, J.; De Veij, M.; Félix, E.; Magariños, M.P.; Mosquera, J.F.; Mutowo, P.; Nowotka, M. ChEMBL: Towards direct deposition of bioassay data. *Nucleic Acids Res.* **2019**, *47*, D930–D940. [[CrossRef](#)]
103. Gilson, M.K.; Liu, T.; Baitaluk, M.; Nicola, G.; Hwang, L.; Chong, J. BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res.* **2016**, *44*, D1045–D1053. [[CrossRef](#)] [[PubMed](#)]
104. Banerjee, P.; Erehman, J.; Gohlke, B.-O.; Wilhelm, T.; Preissner, R.; Dunkel, M. Super Natural II—A database of natural products. *Nucleic Acids Res.* **2015**, *43*, D935–D939. [[CrossRef](#)] [[PubMed](#)]
105. Linstrom, P.J.; Mallard, W.G. The NIST Chemistry WebBook: A chemical data resource on the internet. *J. Chem. Eng. Data* **2001**, *46*, 1059–1063. [[CrossRef](#)]
106. Wishart, D.S.; Feunang, Y.D.; Guo, A.C.; Lo, E.J.; Marcu, A.; Grant, J.R.; Sajed, T.; Johnson, D.; Li, C.; Sayeeda, Z. DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Res.* **2018**, *46*, D1074–D1082. [[CrossRef](#)]
107. Mohanraj, K.; Karthikeyan, B.S.; Vivek-Ananth, R.; Chand, R.; Aparna, S.; Mangalapandi, P.; Samal, A. IMPPAT: A curated database of Indian medicinal plants, phytochemistry and therapeutics. *Sci. Rep.* **2018**, *8*, 4329. [[CrossRef](#)]
108. Danishuddin; Kumar, V.; Faheem, M.; Woo Lee, K. A decade of machine learning-based predictive models for human pharmacokinetics: Advances and challenges. *Drug Discov. Today* **2022**, *27*, 529–537. [[CrossRef](#)]
109. Kumar, A.; Kini, S.G.; Rathi, E. A Recent Appraisal of Artificial Intelligence and In Silico ADMET Prediction in the Early Stages of Drug Discovery. *Mini Rev. Med. Chem.* **2021**, *21*, 2788–2800. [[CrossRef](#)]
110. Ferreira, L.L.G.; Andricopulo, A.D. ADMET modeling approaches in drug discovery. *Drug Discov. Today* **2019**, *24*, 1157–1165. [[CrossRef](#)] [[PubMed](#)]
111. Pantaleao, S.Q.; Fernandes, P.O.; Goncalves, J.E.; Maltarollo, V.G.; Honorio, K.M. Recent Advances in the Prediction of Pharmacokinetics Properties in Drug Design Studies: A Review. *ChemMedChem* **2022**, *17*, e202100542. [[CrossRef](#)] [[PubMed](#)]
112. Pasrija, P.; Jha, P.; Upadhyaya, P.; Khan, M.S.; Chopra, M. Machine Learning and Artificial Intelligence: A Paradigm Shift in Big Data-Driven Drug Design and Discovery. *Curr. Top. Med. Chem.* **2022**, *22*, 1692–1727. [[CrossRef](#)] [[PubMed](#)]
113. Huang, D.Z.; Baber, J.C.; Bahmanyar, S.S. The challenges of generalizability in artificial intelligence for ADME/Tox endpoint and activity prediction. *Expert Opin. Drug Discov.* **2021**, *16*, 1045–1056. [[CrossRef](#)]
114. Kantify. AI in Drug Discovery: ADMET Property Prediction. Available online: <https://kantify.com/use-cases/ai-in-drug-discovery-admet-property-prediction> (accessed on 14 October 2022).
115. Bhatarai, B.; Walters, W.P.; Hop, C.E.C.A.; Lanza, G.; Ekins, S. Opportunities and challenges using artificial intelligence in ADME/Tox. *Nat. Mater.* **2019**, *18*, 418–422. [[CrossRef](#)]
116. Brown, N.; Cambuzzi, J.; Cox, P.J.; Davies, M.; Dunbar, J.; Plumbley, D.; Sellwood, M.A.; Sim, A.; Williams-Jones, B.I.; Zwierzyna, M. Big data in drug discovery. *Prog. Med. Chem.* **2018**, *57*, 277–356. [[CrossRef](#)]
117. Lee, C.H.; Yoon, H.-J. Medical big data: Promise and challenges. *Kidney Res. Clin. Pract.* **2017**, *36*, 3. [[CrossRef](#)]
118. Schneider, P.; Walters, W.P.; Plowright, A.T.; Sieroka, N.; Listgarten, J.; Goodnow, R.A.; Fisher, J.; Jansen, J.M.; Duca, J.S.; Rush, T.S. Rethinking drug design in the artificial intelligence era. *Nat. Rev. Drug Discov.* **2020**, *19*, 353–364. [[CrossRef](#)]
119. Chicco, D. Ten quick tips for machine learning in computational biology. *BioData Min.* **2017**, *10*, 35. [[CrossRef](#)]
120. Fourches, D.; Williams, A.J.; Patlewicz, G.; Shah, I.; Grulke, C.; Wambaugh, J.; Richard, A.; Tropsha, A. Computational Tools for ADMET Profiling. *Comput. Toxicol. Risk Assess. Chem.* **2018**, 211–244. [[CrossRef](#)]
121. Healthcare, G. It Will Take Years for AI Use to Peak in Drug Discovery and Development Process. Available online: <https://www.pharmaceutical-technology.com/comment/ai-peak-drug-discovery-development/> (accessed on 1 November 2022).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.