

QSAR Prediction of BBB Permeability Based on Machine Learning upon PETBD: A Novel Data Set of PET Tracers

Qing Su, Zhilong Ye, Chunyan Han, Ganyao Xiao, Jiazhi Chen, Haiyan Chen, Zigui Wang, Shun Huang, Lu Wang, Grace Peng, Wei Zhou,* and Zhiyun Du*



Cite This: <https://doi.org/10.1021/acs.jmedchem.5c01791>



Read Online

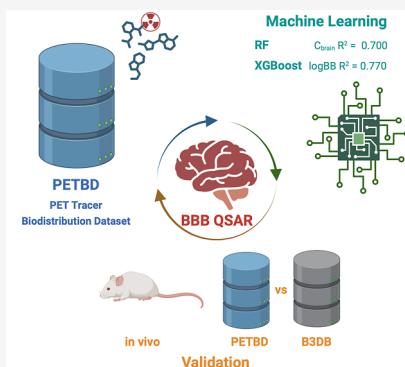
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: Assessing small-molecule blood–brain barrier permeability is laborious, yet critical in drug development. Quantitative prediction models are hindered by a lack of high-quality data set. To address this, we curated PETBD, a novel data set of drug concentrations for 1056 positron emission tomography tracers across 14 organs at 60 min post injection, as well as *in vivo* metadata. We developed machine learning models to predict the brain-to-blood concentration ratio (log BB), and for the first time, drug concentration in the brain. Extreme gradient boosting model reached the best performance in predicting C_{brain} ($R^2 = 0.700$) and also achieved state-of-the-art log BB prediction ($R^2 = 0.770$). Feature importance analysis was employed to explain the contributions of physicochemical-based features. The model's superior generalizability was validated against the B3DB benchmark and with unpublished PET tracers. The open-access PETBD data set (<https://github.com/GDUT-Computer-Medical-Science-Team/PETBD-QSAR/>) and models promise to accelerate CNS drug discovery by bridging computational predictions with empirical pharmacokinetic outcomes.



INTRODUCTION

Blood–brain barrier (BBB) permeability is the most important property for developing a therapeutic agent, no matter the case of central nervous system (CNS) or non-CNS diseases. Necessary brain exposure is prerequisite for CNS drugs to treat CNS disorders, while it may also cause a variety of undesirable side effects for non-CNS drugs.¹ Thus, it is critical to measure BBB penetration properties in a preclinical study. Many efforts have been made to develop *in vitro* and *in vivo* models for BBB permeability determination.² However, traditional drug discovery processes, reliant on these experimental methods, are often time-consuming, labor-intensive, and expensive. Thus, strategies that facilitate a more efficient determination of BBB permeability will lead to significant savings in time and cost during the process of drug discovery. While *in vivo* experiments are generally considered the most reliable for assessing BBB permeability, they remain inherently expensive and time-consuming. In contrast, computer-aided methods for the assessment of BBB permeability present compelling advantages, including high performance, low cost, and high efficiency. Therefore, the prediction of BBB permeability for small molecules represents a vital yet challenging endeavor, an enduring need particularly within the fields of CNS drug discovery and CNS radiopharmaceutical design.^{3–5}

To facilitate BBB permeability prediction, numerous approaches have been developed. Among these, Lipinski's rule of five (Ro5) has gained popularity as a general "rule of

thumb".⁶ However, the conventional Ro5, while useful for overall drug-likeness, is not specifically tailored for predicting the BBB penetration of CNS drugs. Building upon this concept, various optimized rules^{7,8} have been proposed that leverage physicochemical properties such as molecular weight (MW), polar surface area (PSA), $\log P$, and pK_a to estimate the probability of a compound crossing the BBB. Beyond these rule-based approaches, more sophisticated algorithms relying on structural or physicochemical descriptors have emerged to facilitate the design of CNS drug or PET tracers.^{5,9–11} These include tools like CNS MPO,^{12,13} CNS MPO PET score,^{3,4} BBB score,¹ large language model,¹⁴ and dedicated statistical algorithms,¹⁵ which collectively enable the classification of a drug candidate as either permeable (BBB+) or nonpermeable (BBB−) across the BBB.

With the recent advances in machine learning and other artificial intelligence techniques such as deep learning, numerous in silico models have been established to predict BBB permeability of chemical molecules.^{2,16–22} Currently, in silico BBB permeability prediction models can be categorized as either classification or regression models, which are

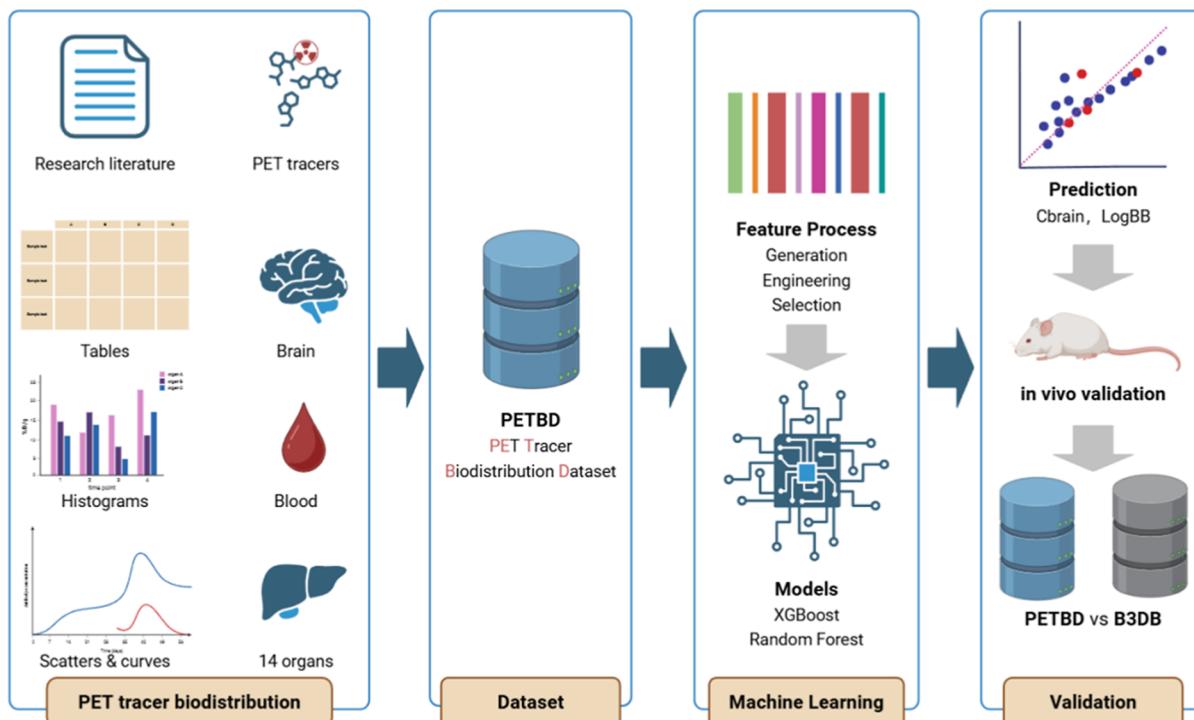
Received: June 30, 2025

Revised: December 16, 2025

Accepted: December 22, 2025

Table 1. Current and Recent Datasets and Algorithms for BBB Permeability Prediction

references	algorithm	data set name	categorical/numerical	size	notes
Meng et al. (2022) ²⁵	NA	B3DB	categorical	7807	4956 BBB+, 2851 BBB-, 1058 log BB
Shaker et al. (2021) ²⁴	LightGBM	LightBBB	numerical	1058	
Gupta et al. (2019) ¹	BBB score	NA	categorical	990	FDA approved drugs, 270 CNS+, 720 CNS-
Wang et al. (2019) ²³	machine learning resampling	NA	categorical	2358	1812 BBB+, 546 BBB-
Wang et al. (2023) ²⁸	RF, XGBoost, HGB, GCNN	NA	numerical	1958	Largest PAMPA-BBB data set
Wang et al. (2015) ²⁹	KNN, RF, SVM	NA	numerical	439	QSAR modeling, Log BB data set
Liu et al. (2021) ³⁰	SVM, RF, and XGBoost	NA	categorical	1593	1283 BBB+, 310 BBB-
Spielvogel et al. (2025) ⁹	KNN, RF, XGBoost, LR	NA	categorical	154	42 BBB-, 68 BBB+, 44 efflux transporter ligands

**Figure 1.** Schematic diagram of the workflow in this study. The illustration was created with BioRender.com.

predominantly based on categorical or numerical data sets, respectively. Within categorical data sets, drugs are labeled as BBB penetrating (BBB+) or nonpenetrating (BBB-) (e.g., data set BBBP at <https://moleculenet.org/datasets-1>), or labeled as CNS drugs (CNS+) or non-CNS drugs (CNS-) from a FDA-approved drug library.¹ For numerical data sets, parameters for indicating permeability of drugs can be log BB (a logarithmic ratio of the concentration of a drug in the brain to its concentration in the blood) or $K_{p,uu,brain}$ (unbound brain-to-plasma partition coefficient). However, the performance of these predictions is heavily reliant on the type, size, diversity, and reliability of data sets. Liu et al.²³ built a machine-learning model with resampling based on a binary data set (2358 entries). A lightBBB model was developed using the light gradient boosting machine (LightGBM) algorithm based on a database including 7162 entries.²⁴ Spielvogel et al.⁹ also recently developed a machine learning-based model for BBB penetration prediction. This model was trained on a data set comprising 24 molecular parameters (both calculated and experimentally derived) from 154 radiolabeled molecules and demonstrated robust classification performance. Quantitative models use logarithm of the permeable surface area product (log PS), of which mostly used data sets were light BBB (696

unique molecules) and B3DB which had 1058 molecules with numeric log BB values.²⁵

Although prediction results from models trained on numerical data sets are more informative for setting thresholds during virtual screening or determining dosages when designing animal experiments in the drug discovery process, there is a scarcity of numerical data sets related to experimental data on BBB permeability, such as log BB²⁶ or $K_{p,uu,brain}$ (Table 1). This scarcity stems from the time-consuming, labor-intensive, and financially costly nature of such data. Notably, one of the main challenges lies in the fact that various indices representing BBB permeability are acquired under different experimental conditions by researchers globally using diverse protocols. This heterogeneity makes it difficult to unify or normalize the data into a large-scale data set suitable for developing computer-aided models that facilitate BBB permeability prediction.

Inspired by the standard operating procedure for determining the biodistribution of positron emission tomography (PET) tracers in animal experiments,^{31–33} this study compiled PET tracer biodistribution data set (PETBD), a novel numerical data set curated from publicly available PET tracer research. This data set systematically integrates time-resolved,

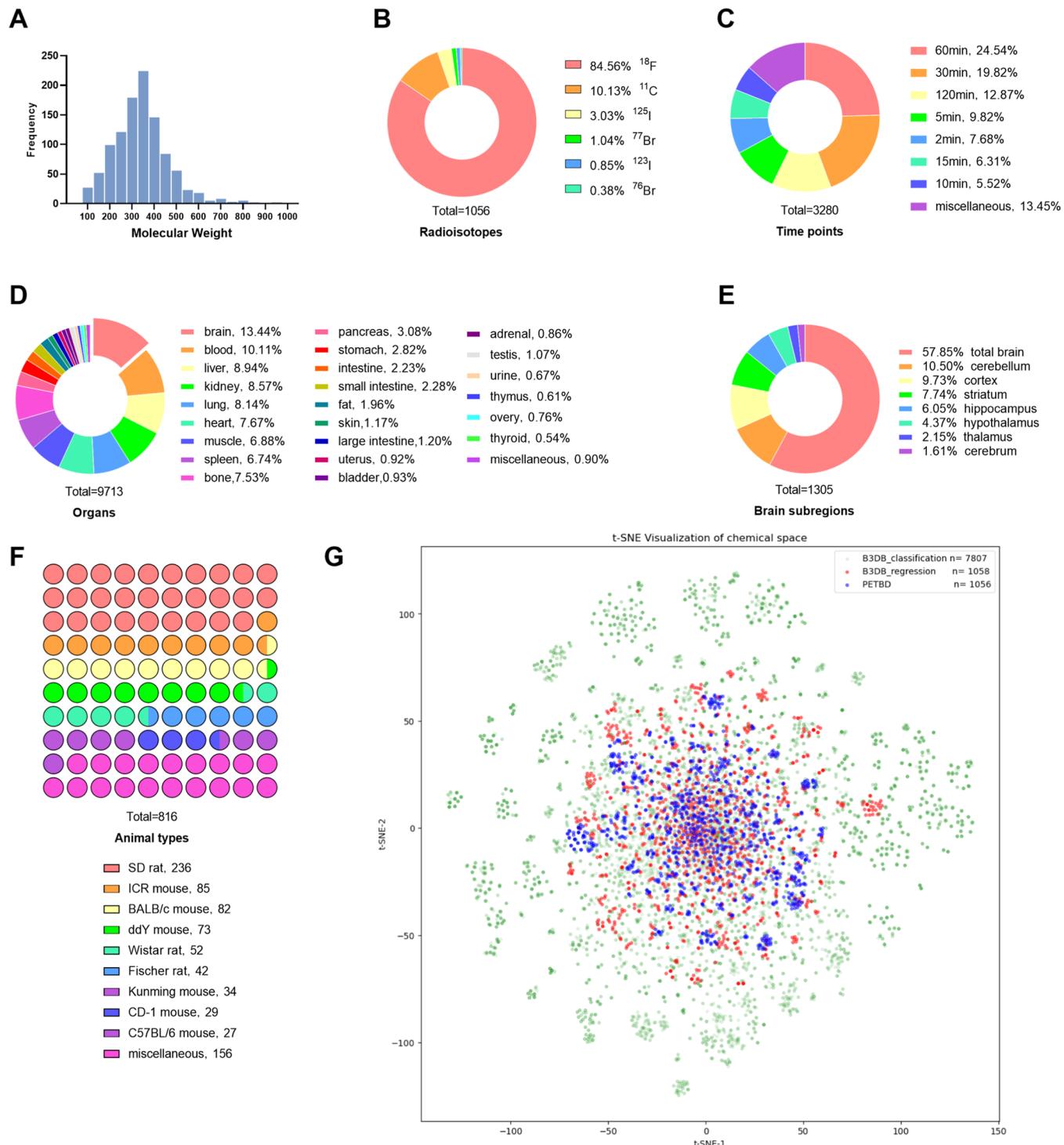


Figure 2. Summary of the PETBD data set. (A) Molecular weight distribution of the PET tracers. (B) Distribution of radioisotopes used in the PET tracers included in the data set. (C) Distribution of time points across the PET tracer studies. (D) Statistics on organs included in the data set, with brain subregions aggregated under the general term “brain”. (E) Distribution of brain subregions analyzed in the data set. (F) Animal types of PET tracers applied in animal experiments. (G) Chemical space visualization of the data set using t-SNE (t-distributed stochastic neighbor embedding) for dimensionality reduction, with entities from the B3DB data set included for comparison (blue: PETBD, red: regression set of B3DB, green: classification set of B3DB).

organ-specific biodistribution profiles of radiolabeled small molecules, normalized to ensure cross-study comparability. Leveraging PETBD, we developed machine learning-driven quantitative structure–activity relationship (QSAR) models to predict BBB permeability, using brain drug concentration (C_{brain}) and the brain-to-blood concentration ratio (log BB) as

key pharmacokinetic (PK) end points. The optimized models demonstrated robust predictive performance for both log BB and C_{brain} , validated through rigorous in silico cross-validation and prospective *in vivo* testing in murine models. These results underscore PETBD’s utility in bridging structural descriptors with empirical PK outcomes.

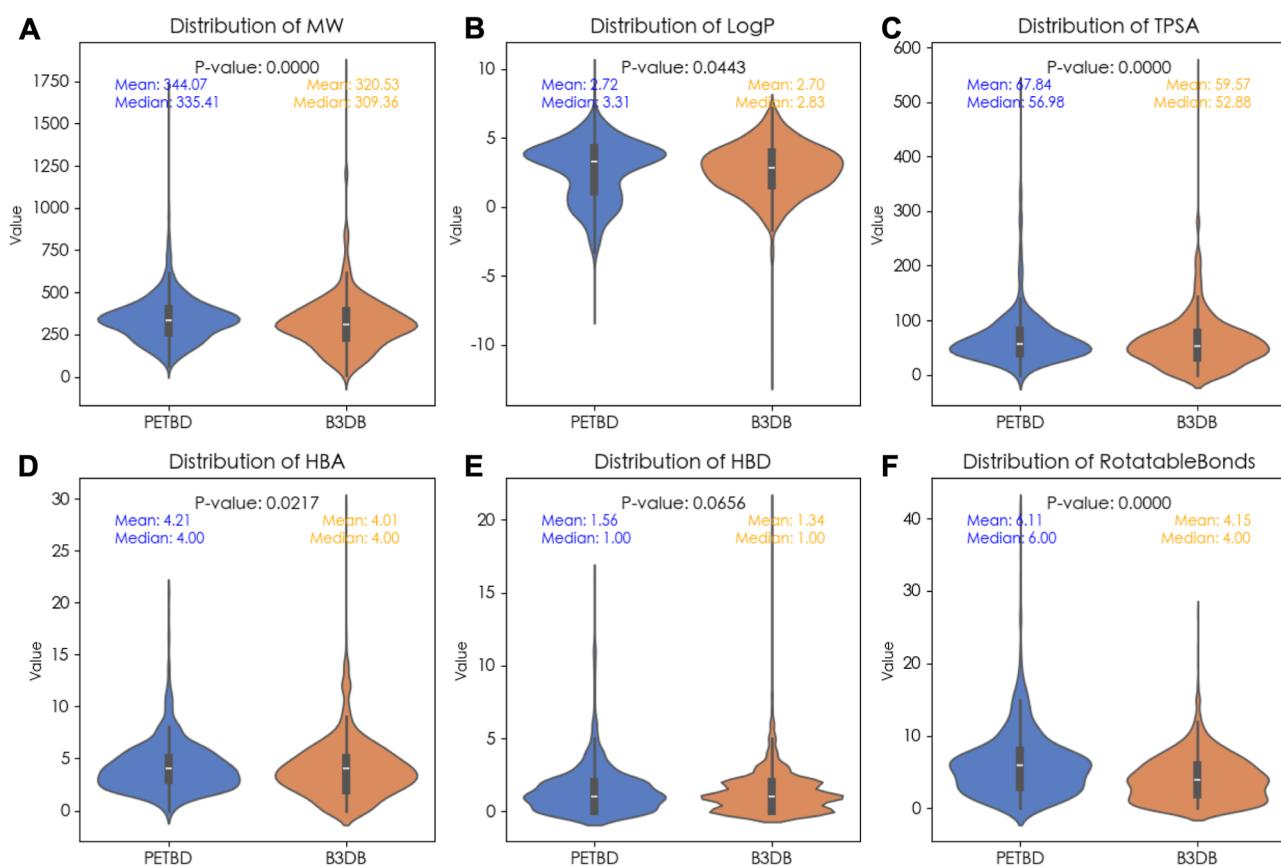


Figure 3. Comparative analysis of physicochemical properties for the PETBD and B3DB regression set. The distributions of six molecular descriptors critical for CNS permeability are visualized using violin plots for the PETBD ($n = 1056$) and B3DB regression set ($n = 1058$) data sets. Properties shown are (A) molecular weight, (B) log P , (C) topological polar surface area (TPSA), (D) hydrogen bond acceptors (HBA), (E) hydrogen bond donors (HBD), and (F) RotatableBonds. Each violin plot displays the kernel density of the data, with the median marked by a white dot and the interquartile range indicated by the thick black bar. Statistical significance between the data sets for each property was determined by a Mann–Whitney U test.

■ RESULTS AND DISCUSSION

To advance the development of predictive models for BBB permeability in CNS drug discovery, this study curated a novel data set derived from the biodistribution profiles of PET tracers in murine models. The PET tracer biodistribution data set (PETBD) comprises 1056 small-molecule PET tracers annotated with molecular representation SMILES, log BB values (logarithm of the brain-to-blood concentration ratio), and time-resolved drug concentrations (expressed as % ID/g; percentage of injected dose per gram of tissue) across 14 organs, including the brain and blood, as well the *in vivo* metadata that the biodistribution data originated from. To the best of our knowledge, PETBD represents the first comprehensive compilation of PET tracer biodistribution data and the first application of machine learning to predict organ-specific drug concentrations from structural features (Figure 1). This groundbreaking initiative bridges the gap between computational chemistry and empirical PKs, enabling data-driven prediction of both BBB permeability and multi-organ drug distribution—a critical capability for rational drug design and preclinical optimization.

Data Set PETBD

In this study, biodistribution data of PET tracers were systematically collected and organized into a comprehensive data set, PETBD. PET tracers publicly reported by 2022 were identified through systematic searches on PubMed and Web of

Science using the keywords “PET” and “biodistribution”. For PET tracers reported prior to 2013, primary references were drawn from the Molecular Imaging and Contrast Agent Database (MICAD, <https://www.ncbi.nlm.nih.gov/books/NBK5330/>) and the ^{18}F -Database of Imaging Radiolabeled Compounds (DIRAC, <http://www.iphc.cnrs.fr/dirac/>). Over 10,000 research articles were initially indexed and manually filtered, resulting in the refinement and data extraction from more than 2000 full-text articles.

The chemical structures and biodistribution data of each PET tracer were manually recorded in MOL file format and standardized Excel worksheets, with data conversion applied where necessary. All retrieved Excel datasheets and MOL files were integrated using a self-developed tool, *MedicalDatasets-Merger*, into a unified data set. Following the removal of outlier data, collation of chemical structures, and merging of duplicates, a novel data set was constructed.

Before finalizing the data set, a comprehensive analysis was conducted (Figure 2). Among all PET tracers included in this resource, the molecular weights of most of the tracers ranged between 100 and 600. ^{18}F is the most frequently used radioisotope in the collected PET tracers, likely due to its favorable decay half-life and ease of synthetic preparation. The incorporation of fluorine atoms is also a common feature in modern small-molecule drugs. We also analyzed the organs and time points reported in the collected research literature.

The brain and blood were the most frequently studied organs. Since some CNS-targeted PET tracers investigated specific brain subregions, these subregions were aggregated under the general term “brain”, and their data were calculated as total brain concentrations based on the weighted contributions of subregion concentrations and their mass or volume fractions relative to the whole brain.

Statistical analysis of collected biodistribution studies indicates that 60 min is the most frequently utilized time point. This prominence is significantly influenced by the PK properties of PET radiopharmaceuticals, as the most commonly used PET tracer labeling isotopes, such as fluorine-18 (half-life $T_{1/2} = 109.77$ min) and carbon-11 ($T_{1/2} = 20.34$ min), retain substantial concentrations 60 min postinjection for imaging, making this time point ideal for assessing initial uptake and distribution kinetics. Furthermore, a broader literature analysis reveals that many pharmaceuticals exhibit biological half-lives ranging from minutes to hours.³⁴ Consequently, the 60 min time point often serves as a critical benchmark for evaluating the metabolic disposition and PK characteristics of diverse drug candidates, which led to this time point being selected for further machine learning modeling.

We compared our PETBD data set (1056 entries, focused on log BB) with the largest publicly available BBB permeability data set B3DB (7807 classification/1058 regression entries). We found minimal molecular overlap with only 35 shared compounds (Table S7). To compare chemical space, we calculated six key physicochemical properties (MW, log P, TPSA, HBD, HBA, and RotatableBonds) using RDKit. Despite statistically significant differences in MW, TPSA, and RotatableBonds (Mann–Whitney U test, $p < 0.001$), the violin plots (Figure 3) show substantial distributional overlap with similar medians and interquartile ranges. We attribute the statistical significance to the large sample size of the B3DB rather than practical differences. We conclude that both data sets cover a comparable chemical space for these key properties.

Consequently, drug concentrations in 14 typical organs, including the brain and blood, at the 60 min time point for 1056 PET tracers, as well as their calculated log BB, canonical SMILES, metadata including animal type, gender, animal weight, and injection dosage (Table 2), were integrated into the final data set, PETBD. The data set is stored in comma-separated values (CSV) format and is openly accessible on GitHub (<https://github.com/GDUT-Computer-Medical-Science-Team/PETBD-QSAR/>).

The biodistribution data of PET tracers recorded in this data set are expressed as % ID/g, which represents the percentage of the injected dose of a PET tracer per gram of specific organs at certain time point. These values were measured using radioactive gamma counts after correcting for isotope decay and are typically presented in tabular form in the public literature. Biodistribution values reported as % ID/mL or SUV have been normalized and converted to % ID/g. Additionally, tabular data have been quantified into numerical values for the sake of consistency.

Since the data in this data set originated from typical PET tracer studies conducted following standard operating procedures with limited types of apparatus, they reflect the overall distribution of the drug across the BBB, regardless of *in vivo* metabolites. The uniform nature of the data allows for integration into a single data set, facilitating meaningful

Table 2. List of Information in the Dataset PETBD

column header	description	data type
compound index	compound identifier within data set PETBD	string
SMILES	canonical SMILES generated with RDKit	string
PMID	PubChem compound identifier	string
animal type	types of animals in biodistribution study	string
gender	gender of animals in biodistribution study	string
animal weight	weight of animals in biodistribution study, unit: g	float
injection dosage	injection dosage of PET tracers in biodistribution study, unit: μCi	float
log BB	log BB value calculated with concentrations of brain and blood	float
brain	drug concentration in brain 60 min post injection, unit: % ID/g	float
blood	drug concentration in blood 60 min post injection, unit: % ID/g	float
liver	drug concentration in liver 60 min post injection, unit: % ID/g	float
lung	drug concentration in lung 60 min post injection, unit: % ID/g	float
kidney	drug concentration in kidney 60 min post injection, unit: % ID/g	float
muscle	drug concentration in muscle 60 min post injection, unit: % ID/g	float
spleen	drug concentration in spleen 60 min post injection, unit: % ID/g	float
intestine	drug concentration in intestine 60 min post injection, unit: % ID/g	float
bone	drug concentration in bone 60 min post injection, unit: % ID/g	float
stomach	drug concentration in stomach 60 min post injection, unit: % ID/g	float
uterus	drug concentration in uterus 60 min post injection, unit: % ID/g	float
heart	drug concentration in heart 60 min post injection, unit: % ID/g	float
fat	drug concentration in fat 60 min post injection, unit: % ID/g	float
pancreas	drug concentration in pancreas 60 min post injection, unit: % ID/g	float
reference	DOI number of research reference	string

comparisons. As a result, these standardized data are highly suitable for use as training data for machine learning and further AI applications. Numerical inputs paired with numerical outputs are more informative than categorical inputs, offering greater utility for *in vivo* experimental design and drug discovery by providing detailed insights beyond simple BBB permeability classifications. Moreover, for PET tracers, biodistribution data in the brain and blood are included and can be converted into log BB values, providing a novel resource to complement currently available log BB data sets.

QSAR Models of Machine Learning

Based on our data set PETBD, we tuned various machine learning models, including SVM, LightGBM, RF, XGBoost, MLP, and CatBoost, with C_{brain} and log BB as the target variables, respectively, descriptors or fingerprints generated from SMILES strings, and metadata of PET animal experiments (animal type, gender, animal weight, and injection dosage) as variables. Performance and effectiveness of tuned models in predicting C_{brain} and log BB were systematically evaluated.

The overall modeling workflow began by partitioning the data set into a training set (81%), a validation set (9%), and a held-out test set (10%). All subsequent process was performed exclusively on the training set. A preprocessing pipeline was first applied, which involved oversampling, imputing missing values, and then recursive feature elimination (RFE) to select the top 50 most informative features. Using this processed data, a 10-fold cross-validation strategy coupled with Optuna's Bayesian optimization was employed to find the optimal model hyperparameters. The final model was then trained on the entire training set, and its generalization performance was assessed on the held-out test set.

Model Performance on C_{brain} Prediction

To evaluate the predictive performance for C_{brain} , six machine learning regression models were developed and assessed on an independent test set. The performance metrics, including R^2 , MSE, RMSE, MAE, adjusted R^2 , and MAPE, are summarized in Table 3. Among the evaluated models, RF, XGBoost, and

Table 3. Model Performance on C_{brain} Prediction

models ^a	R^2	MSE	RMSE	MAE	adjusted R^2	MAPE (%)
RF	0.662	0.568	0.754	0.488	0.288	15.16
XGBoost	0.700	0.505	0.710	0.440	0.367	14.75
LightGBM	0.473	0.452	0.672	0.526	0.460	17.12
CatBoost	0.449	0.473	0.688	0.468	0.435	18.65
SVM	0.389	0.525	0.725	0.563	0.373	20.13
MLP	0.260	0.635	0.797	0.648	0.240	22.92

^aAll models were trained using fused features derived from the training set ($n = 770$), underwent hyperparameter tuning via cross-validation and Optuna on the validation set ($n = 86$), and metrics reported were ultimately evaluated on a predefined held-out test set ($n = 96$).

LightGBM demonstrated markedly superior performance compared to CatBoost, SVM, and MLP. The XGBoost model yielded the highest coefficient of determination ($R^2 = 0.700$) and the lowest MAE (MAE = 0.440) and MAPE (MAPE = 14.75%), while the RF model achieved comparable R^2 but with highest adjusted R^2 (adjusted $R^2 = 0.288$), indication of its stability of predictive capability.

In our experiments, different models exhibited varying performance when using molecular descriptors and molecular fingerprints as input features (Table S3 and S4). Overall, RF demonstrated the best performance with molecular descriptors. As shown in Table S3, RF achieved the highest R^2 value of 0.447 among all models, highlighting its strong predictive capabilities. In contrast, XGBoost performed well using molecular fingerprints on other metrics such as RMSE and MAE, as indicated in Table S4.

Figure 4 provides a visual comparison of the predicted versus true values for the two top-performing models on both the training and test sets. Both RF (Figure 4A) and XGBoost (Figure 4C) achieved excellent fits on the training data, with R^2 values of 0.879 and 0.926, respectively. However, a clear drop in performance is visible for both models on the test set (Figure 4B,D), where the R^2 values decreased to 0.663 for RF and 0.700 for XGBoost. This discrepancy highlights the degree of overfitting in both models.

Our results identify tree-based ensemble methods, specifically RF and XGB, as the most effective models for this task. They significantly outperformed other popular methods,

including gradient-boosting machines (LightGBM, CatBoost), SVM, and a neural network (MLP). The close competition between RF and XGBoost is noteworthy. XGBoost showed a slight advantage in terms of absolute error (lowest RMSE and MAE), which is often a critical metric in quantitative prediction. In contrast, RF explained a slightly higher proportion of the variance (highest R^2). A key observation from the regression plots (Figure 4) is the performance gap between the training and test sets for both of the top models. The high R^2 values on the training data (0.879–0.926) compared with the test data are a classic sign of overfitting. This suggests that while the models are powerful, they may be capturing noise from the training data. Future work could focus on mitigating this overfitting through more aggressive regularization, feature selection, or by acquiring a larger and more diverse training data set. Despite this, an R^2 of 0.700 on a held-out test set represents a strong predictive model for a complex biological end point like C_{brain} . These findings underscore the importance of model selection in QSAR studies. Future work could explore hybrid ensemble approaches or physics-informed neural networks to further bridge the gap between empirical and mechanistic modeling.

Model Performance on Log BB Prediction

We next evaluated the models for the prediction of log BB. To address model bias arising from imbalanced sample distribution between ¹⁸F-labeled and non-¹⁸F compounds, we applied oversampling to balance the training data. This balancing strategy led to substantial performance improvements on the test set: R^2 increased from 0.484 to 0.632, RMSE decreased from 0.493 to 0.379 (a 23% reduction), and MAE decreased from 0.390 to 0.286 (a 27% reduction) during model optimization, collectively indicating enhanced model generalization (Supporting Information).

The comparative performance metrics for the independent test set are presented in Table 4. Overall, the models achieved a stronger predictive performance for log BB than for C_{brain} . XGBoost obtained the highest coefficient of determination ($R^2 = 0.770$), followed closely by RF ($R^2 = 0.759$) and CatBoost ($R^2 = 0.725$). However, a different trend emerged for the error-based metrics. The CatBoost model yielded substantially lower RMSE (0.485) and MAE (0.340) values than any other model. This indicates that its predictions were, on average, significantly more accurate than those of XGBoost (RMSE = 0.753) and RF (RMSE = 0.772). Furthermore, CatBoost also achieved the highest Adjusted R^2 (0.717) by a wide margin, suggesting it is a more robust and well-generalized model. Figure 5 illustrates the performance of the RF and XGBoost models. Similar to the C_{brain} task, both models showed signs of overfitting, with near-perfect performance on the training sets (Figure 5A,C, $R^2 > 0.97$) that dropped on the test sets (Figure 5B,D, $R^2 \approx 0.76$). Notably, MAPE remained largely unchanged, which aligns with its known sensitivity to small target values and its relative insensitivity to distribution balance. To ensure statistical reliability and prevent variance underestimation due to oversampling, we performed additional validation using stratified K-fold cross-validation with multiple random seeds. The consistent results across these validation schemes confirm the robustness of our findings and support the effectiveness of the balancing strategy.

In the log BB prediction task, the results provided a more nuanced picture. While XGBoost and RF again achieved the highest R^2 values, the CatBoost model was clearly the superior

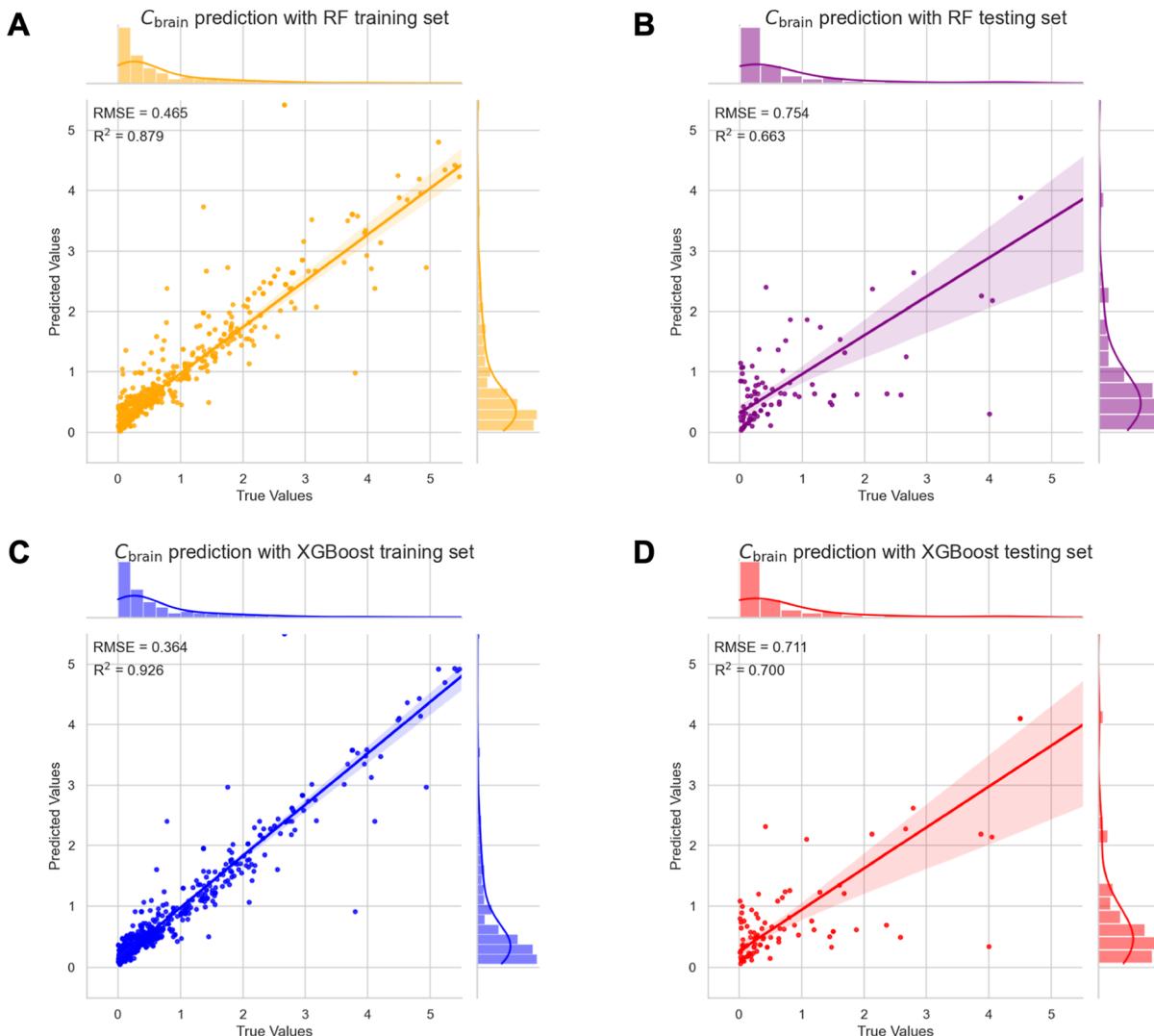


Figure 4. Models were trained on fused features (fingerprints, descriptors, and *in vivo* metadata) of training set ($n = 770$), tuned with CV and Optuna on validation set ($n = 86$), finally evaluated on held-out test set ($n = 96$) from a predefined data set split. Performance evaluation of final RF and XGBoost models in predicting C_{brain} using RF model, predicted values vs true values of training set (A) and test set; (B) using XGBoost model, predicted values vs true values of training set (C) and test set (D).

Table 4. Model Performance on Log BB Prediction

models ^a	R ²	MSE	RMSE	MAE	adjusted R ²	MAPE (%)
XGBoost	0.770	0.566	0.753	0.441	0.433	1.02
RF	0.759	0.596	0.772	0.459	0.403	1.05
CatBoost	0.725	0.235	0.485	0.340	0.717	1.05
LightGBM	0.644	0.306	0.553	0.418	0.635	1.32
MLP	0.369	0.542	0.736	0.593	0.350	2.44
SVM	0.342	0.565	0.752	0.610	0.322	2.53

^aAll models were trained using fused features derived from the training set ($n = 681$), underwent hyperparameter tuning via cross-validation and Optuna on the validation set ($n = 77$), and metrics reported were ultimately evaluated on a predefined held-out test set ($n = 86$).

performer in terms of predictive accuracy and model robustness. Conversely, the other models showed a large drop-off. This discrepancy suggests that while XGBoost and RF are effective at capturing the variance in the data (high R^2), they may be overly complex or penalized for feature

redundancy, making them less robust than CatBoost for this specific task.

Comparison with Prior Prediction Models

QSAR modeling of BBB permeability for small-molecule drugs presents a persistent challenge. While several predictive models have been developed using limited data sets of log BB values, their generalizability remains a concern. In this study, we trained models for both prediction log BB values (Table 5) and drug concentration (at 60 min post i.v.) of small molecular chemicals and rigorously evaluated its performance on an independent hold-out data set to ensure unbiased validation.

For those quantitative prediction models, comparative analysis revealed that our XGBoost model outperformed the recently published LogBB_Pred (LightGBM) model,²¹ demonstrating superior predictive accuracy with both a higher coefficient of determination ($R^2 = 0.77$ vs 0.61) and a lower mean squared error (MSE = 0.56 vs 0.36). Furthermore, our model exhibited improved performance compared to previously reported models,^{35,36} indicating enhanced reliability

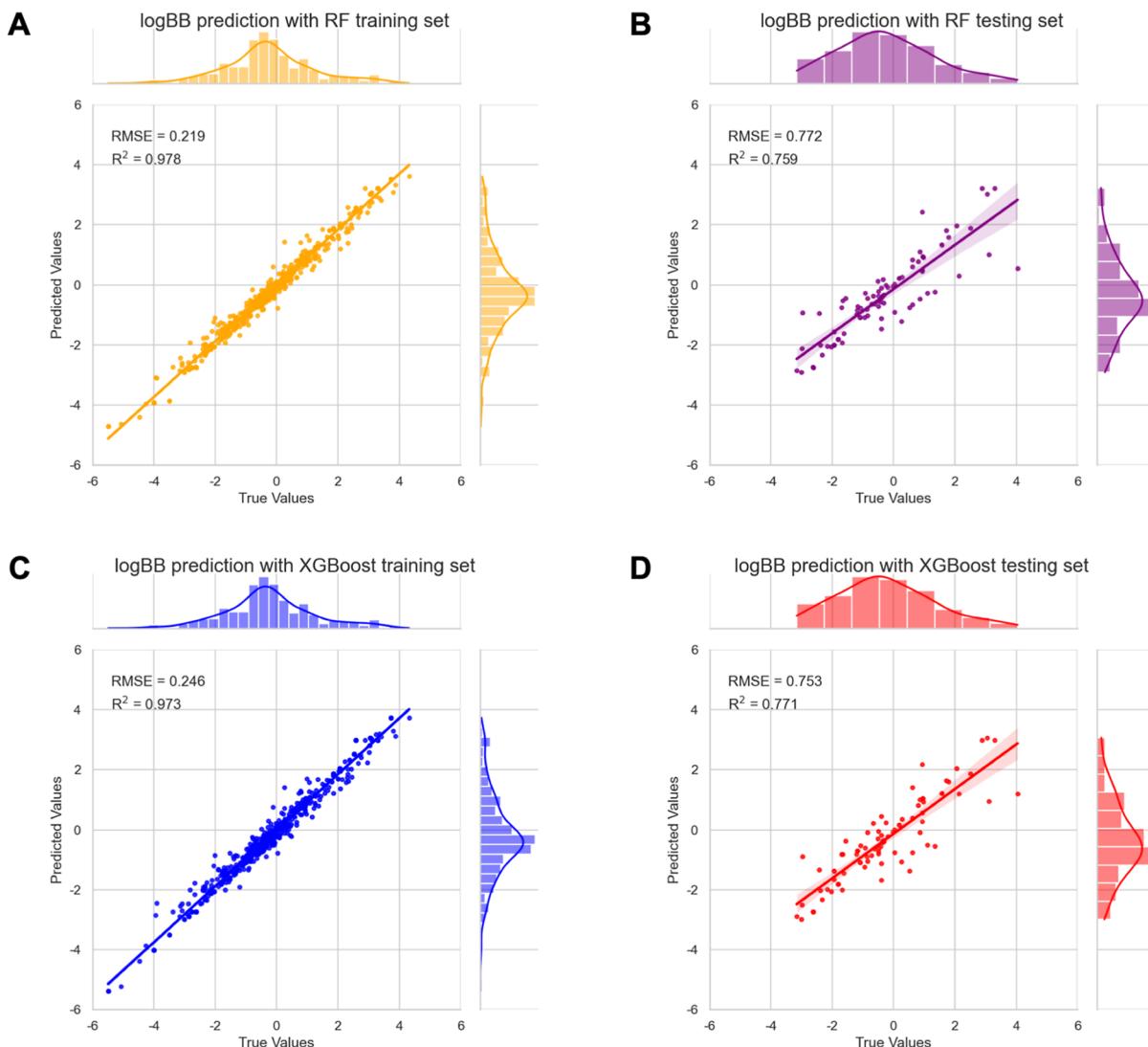


Figure 5. Models were trained on fused features (fingerprints, descriptors, and *in vivo* metadata) of training set ($n = 681$), tuned with CV and Optuna on validation set ($n = 77$), finally evaluated on held-out test set ($n = 86$) from a predefined data set split. Performance of predicting log BB using RF model, predicted values vs true values of training sets (A) and test set; (B) using XGBoost model, predicted values vs true values of training sets (C) and test set (D).

Table 5. Comparison of XGBoost Model Performance with Existing Quantitative Models

model	R ²	MSE	references
our model	0.77	0.56	this study
LogBB_Pred	0.61	0.36	Shaker et al. (2023) ²¹
ADMET prediction service	0.56	0.41	Dyabina et al. (2016) ³⁵
multiple linear regression	0.59	-	Muehlbacher et al. (2011) ³⁶

and accuracy of our model in predicting log BB values for novel compounds.

To facilitate comparison with more prevalent qualitative prediction models that classify compounds as BBB+/BBB– or CNS/non-CNS, our XGBoost model for log BB prediction was transformed into a binary classifier using a cutoff threshold on both our hold-out test set and the B3DB regression set. Currently, there is no universally accepted gold standard for this cutoff threshold. While the cutoff value can vary, a log BB value ≥ 0 (equal concentration in brain and blood) or ≥ -1 is often used to define a CNS + compound. A log BB > 0.3 is considered indicative of ready penetration, whereas a log BB $<$

-1 suggests poor diffusion into the brain. Since B3DB contains BBB+/BBB– labels, a cutoff threshold ($\log \text{BB} = -0.329$) was optimized (Figure 8D) to achieve the max MCC on B3DB when applying our model's log BB prediction results to this third-party data set. The optimized model exhibited satisfactory overall performance with a superior MCC (0.69) for classifying entries in PETBD.

The empirically derived threshold of -0.329 (maximizing MCC) represents a biologically functional boundary rather than an arbitrary statistical artifact. Correlating to a brain-to-blood ratio of $\approx 0.47:1$, this cutoff aligns with the free drug hypothesis, correctly classifying CNS-active compounds where high plasma protein binding depresses total K_p despite adequate free-drug equilibration. Furthermore, this threshold captures the “latent consensus” of clinically validated CNS drugs (e.g., morphine, risperidone), effectively distinguishing compounds with manageable efflux liability from those with prohibitive permeability issues better than standard integer cutoffs, such as $\log \text{BB} = 0$ or -1 .

However, the classification performance remained challenging to surpass that of state-of-the-art BBB ± classifiers (Table 6). This limitation may be attributed to the fact that our model

Table 6. Performance Comparison of Our XGBoost Model with Publicly Available Qualitative Models

model	MCC	accuracy	sensitivity	specificity	references
our model ^a	0.69	0.85	0.81	0.86	this study
our model ^b	0.34	0.69	0.70	0.66	this study
LogBB_Pred	0.60	0.85	0.42	0.99	Shaker et al. (2023) ²¹
random forest model	-	0.83	0.88	0.75	Spielvogel et al. (2025) ⁹
ML model	0.98	1.00	1.00	0.97	Banerjee et al. (2024) ¹¹
admetSAR	0.65	0.85	0.85	0.70	Tang et al. (2018) ³⁷ Tang et al. (2024) ³⁸
LightBBB	0.42	0.70	0.70	0.74	Shaker et al. (2021) ²⁴
SwissADME	0.29	0.70	0.57	0.75	Daina et al. (2016) ³⁹
BBB predictor	0.15	0.70	0.28	0.85	Liu et al. (2014) ⁴⁰

^aClassification performance with model XGBoost on hand-out test set of PETBD, classified as BBB+/BBB- with cutoff threshold -0.329.

^bClassification performance with model XGBoost on classification subset of B3DB, classified predicted log BB values as BBB+/BBB- with cutoff threshold -0.329.

was originally trained and optimized as a quantitative model intended to predict continuous variables, which may not be optimally suited for transformation into categorical classification tasks. Nevertheless, this broader comparative analysis provides valuable insights for the methodological evaluation.

Choosing a BBB prediction method depends on the project goals. Simple heuristics like MPO are fast for initial high-throughput triaging, but their static, binary (BBB+/BBB-) output offers limited guidance for *in vivo* experimental design. In contrast, our ML-driven approach provides a substantial leap in practical utility by predicting quantitative log BB values and, more significantly, dynamic drug concentration–time profiles within the brain. This actionable, high-resolution data is far more informative, enabling the optimization of dosing regimens, PK sampling, and target engagement estimates—tasks impossible with simple binary predictions. Therefore, while MPO is useful for initial screening, our model is the recommended starting point for lead optimization projects. Future efforts will be devoted to improving model accuracy and robustness through larger data sets and developing more sophisticated, user-friendly algorithms to guide molecular design.

Model Explanation

Explanation of Log BB Prediction Model. To systematically identify the most robust and consistently contributing features, we implemented a repeated pipeline encompassing feature selection, model training, and cross-fold validation. This procedure comprised 30 independent iterations with each execution employing a unique random seed. Subsequently, the

Table 7. Feature Importance Analysis of Top 20 Molecular Descriptors and Related Properties for the Model Predicting Log BB

feature	selection counts (30 runs)	total importance	ranking in permutation	ranking in SHAP	feature description (physio-chem property)
ATSC1are	25	1.39	1	1	centered Moreau-Broto autocorrelation of lag 1 weighted by Allred-Rocow EN
BCUT2D_CHGLO	30	0.36	2	4	incorporates atom charges in the burden matrix—returns the lowest eigenvalue
AATSC0c	15	0.21	3	3	averaged and centered Moreau-Broto autocorrelation of lag 0 weighted by Gasteiger charge
AATSC1Z	25	0.46	4	11	averaged and centered Moreau-Broto autocorrelation of lag 1 weighted by atomic number
ATSC1c	28	0.27	5	2	centered Moreau-Broto autocorrelation of lag 1 weighted by Gasteiger charge
SsssN	29	0.60	6	5	sum of sssN
ATSC1pe	19	0.32	7	7	centered Moreau-Broto autocorrelation of lag 1 weighted by Pauling EN
MATS1are	22	0.31	8	13	Moran coefficient of lag 1 weighted by Allred-Rocow EN
ATSC1m	28	0.44	9	9	centered Moreau-Broto autocorrelation of lag 1 weighted by mass
AATSC1pe	29	0.71	10	16	averaged and centered Moreau-Broto autocorrelation of lag 1 weighted by Pauling EN
MinEStateIndex	26	0.03	11	21	minimum EState index
GATS3c	13	0.06	12	14	Geary coefficient of lag 3 weighted by Gasteiger charge
BCUT2D_LOGPLOW	24	0.11	13	18	incorporates atom logarithms of the partition coefficient ($\log P$) in the Burden matrix—returns the lowest eigenvalue
BIC4	15	0.08	14	30	4-ordered bonding information content
MATS1c	29	0.33	15	12	Moran coefficient of lag 1 weighted by Gasteiger charge
GATS1pe	3	0.01	16	8	Geary coefficient of lag 1 weighted by Pauling EN
BCUT2D_MRHI	2	0.00	17	28	incorporates atom molar refractivity in the burden matrix—returns the highest eigenvalue
ATSC0p	29	0.45	18	10	centered Moreau-Broto autocorrelation of lag 0 weighted by polarizability
AATS1Z	15	0.06	19	26	averaged Moreau-Broto autocorrelation of lag 1 weighted by atomic number
BCUT2D_LOGPHI	29	0.07	20	23	incorporates atom logarithms of the partition coefficient ($\log P$) in the burden matrix—returns the highest eigenvalue

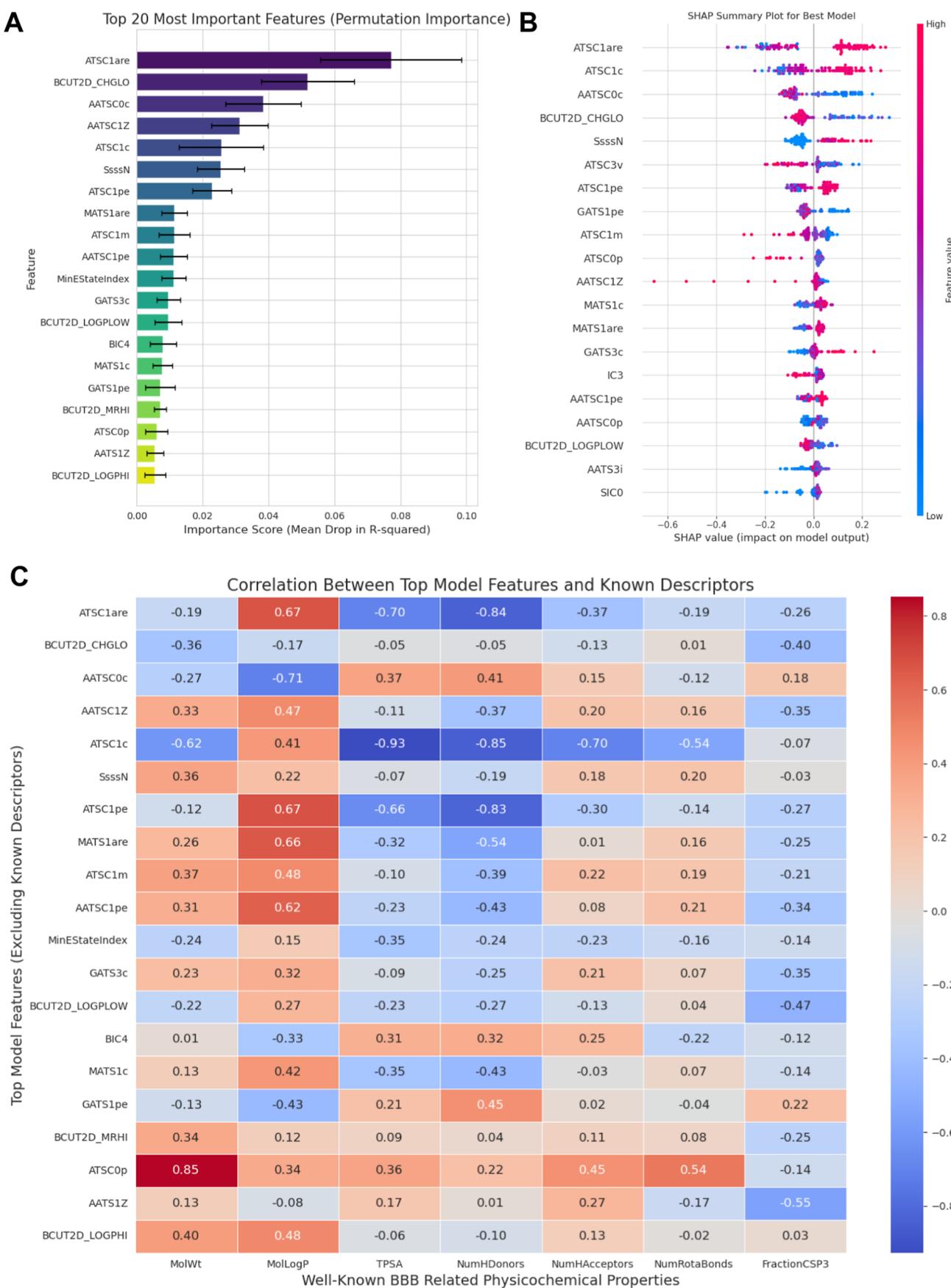


Figure 6. Feature analysis of the best log BB prediction model (XGBoost) presents a comprehensive analysis of feature importance and relationships. (A) Global permutation importance of features. (B) SHAP analysis of feature contributions. (C) Correlation heatmap illustrating relationships between model features and well-known BBB permeability-related descriptors and physicochemical properties.

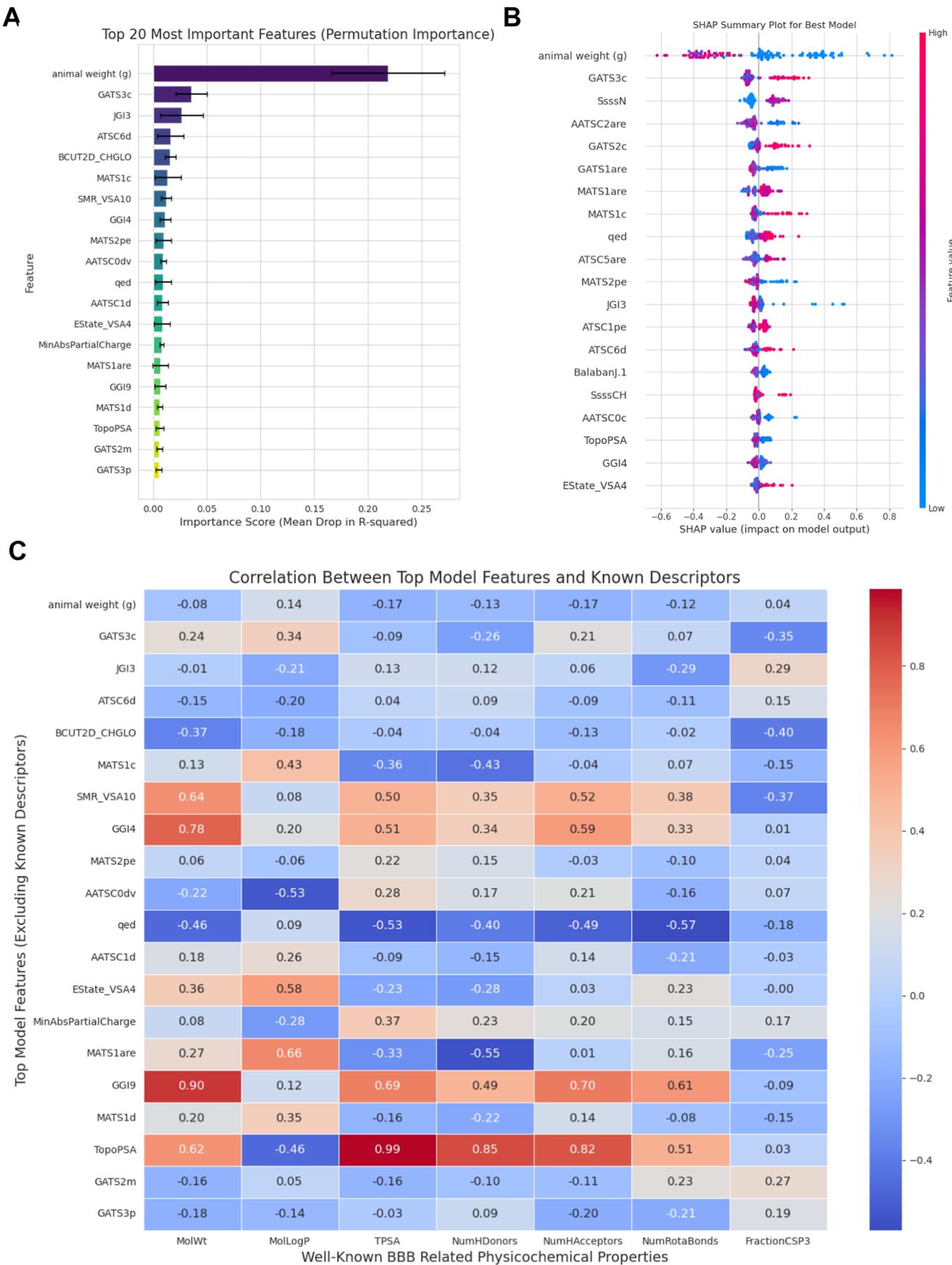


Figure 7. Feature analysis of the best C_{brain} prediction model (RF) presents a comprehensive analysis of feature importance and relationships. (A) Global permutation importance of features. (B) SHAP analysis of feature contributions. (C) Correlation heatmap illustrating relationships between model features and well-known BBB permeability-related descriptors and physicochemical properties.

Table 8. Feature Importance Analysis of Top 20 Molecular Descriptors and Related Properties for the Model Predicting C_{brain}

feature	selection counts (30 runs)	total importance	ranking in permutation	ranking in SHAP	feature description (physio-chem property)
animal weight (g)	30	0.27	1	1	weight of animals of metadata
GATS3c	20	0.25	2	2	Geary coefficient of lag 3 weighted by Gasteiger charge
JGI3	24	0.19	3	12	3-ordered mean topological charge
ATSC6d	10	0.11	4	14	centered Moreau-Broto autocorrelation of lag 6 weighted by sigma electrons
BCUT2D_CHGLO	28	0.15	5	30	incorporates atom charges in the Burden matrix—returns the lowest eigenvalue
MATS1c	24	0.21	6	8	Moran coefficient of lag 1 weighted by Gasteiger charge
SMR_VSA10	10	0.04	7	21	calculates the SMR (molar refractivity) VSA for a molecule by assigning atom contributions to predefined bins based on their Labute ASA and MR values
GGI4	29	0.38	8	19	4-ordered raw topological charge
MATS2pe	15	0.17	9	11	Moran coefficient of lag 2 weighted by Pauling EN
AATSC0dv	24	0.33	10	47	averaged and centered Moreau-Broto autocorrelation of lag 0 weighted by valence electrons
qed	20	0.08	11	9	quantitative estimation of drug-likeness
AATSC 1d	20	0.18	12	27	averaged and centered Moreau-Broto autocorrelation of lag 1 weighted by sigma electrons
EState_VSA4	13	0.11	13	20	EState VSA Descriptor 4 ($0.72 \leq x < 1.17$)
MinAbsPartialCharge	24	0.16	14	44	minimum absolute partial charge
MATS1are	26	0.58	15	7	Moran coefficient of lag 1 weighted by Allred-Rocow EN
GGI9	23	0.19	16	26	9-ordered raw topological charge
MATS 1d	13	0.10	17	45	Moran coefficient of lag 1 weighted by sigma electrons
TopoPSA	28	0.36	18	18	topological polar surface area
GATS2m	18	0.05	19	39	Geary coefficient of lag 2 weighted by mass
GATS 3p	12	0.06	20	35	Geary coefficient of lag 3 weighted by polarizability

features most frequently selected by the optimized pipeline across these iterations are detailed in Table 7. Features such as BCUT2D_CHGLO, SssssN, MATS1c, and ATSC0p were selected in over 95% of the runs, indicating their fundamental importance to building a predictive model for log BB, regardless of the specific training data.

On the final optimized model, permutation importance was used to assess the global impact of each feature on test set performance (Figure 6A). This analysis identifies the features the model relies on most heavily for its accuracy. This model-agnostic method measures the drop in model R^2 when a single feature's information is shuffled. Our analysis revealed that ATSC1are and BCUT2D_CHGLO were the most critical features; removing their signal significantly degraded the model's predictive accuracy.

To understand the directionality and magnitude of feature effects, we performed SHAP (SHapley Additive exPlanations) analysis. The SHAP summary plot (Figure 6B) corroborates the findings from permutation importance, showing that ATSC1are have the largest impact (mean absolute SHAP value). Furthermore, it provides mechanistic insight: high values of ATSC1are (red dots on the right) consistently push the prediction higher (positive SHAP value), confirming the well-established principle that higher lipophilicity increases the log BB. Conversely, the plot for AATSC0c shows that high values consistently lower the predicted log BB, aligning with chemical intuition.

Interestingly, while ATSC1se were selected frequently (100%) during model training, its permutation importance was lower than expected. The SHAP analysis revealed that its contributions were often smaller and more context-dependent than features such as SssssN (count of tertiary aliphatic amine groups). This suggests that ATSC1se is useful for fine-tuning predictions for specific molecules but is not a primary driver of overall performance. Notably, animal weight from the metadata in our data set, despite not being among the top

20 descriptors, made significant contributions to the model's performance (ranking 27th in permutation importance and 35th in SHAP analysis).

Notably, while the final model demonstrated a strong predictive performance, several descriptors traditionally associated with BBB permeability, such as molecular weight (MolWt) and TPSA, were not ranked among the most important features by permutation importance or SHAP analysis. We investigated the potential reasons for this observation. The primary reason for this appeared to be the feature redundancy. The correlation heatmap (Figure 6C) revealed that the information encoded by MolWt is strongly correlated with model-selected features. For instance, MolWt showed a high correlation with the descriptor ATSC0P positively ($r = 0.85$) and a moderate correlation with ATSC1c negatively ($r = -0.62$), both of which were ranked highly by the model. Similarly, TPSA shows a strong negative correlation ($r = -0.93$) with ATSC1c, while the number of hydrogen bond donors (NumHDonors) negatively correlates with ATSC1are, ATSC1c, and ATSC1pe simultaneously. This suggested that the model did not overlook the concepts of molecular size or polarity. Instead, it identified a more informative representation of these properties through a combination of other more nuanced descriptors, which ultimately led to a greater reduction in prediction error during training.

Explanation of C_{brain} Prediction Model. We also delved into the explanation of the C_{brain} prediction model. The features contributing to this model varied significantly from those in the model predicting log BB. Notably, animal weight from the research resource metadata was the most frequently selected feature by the model, indicating its consistent contribution. Furthermore, it consistently ranked highest in both permutation importance (Figure 7A) and SHAP analysis (Figure 7B), signifying its substantial influence, far exceeding that of other features (Table 8). This finding aligns with

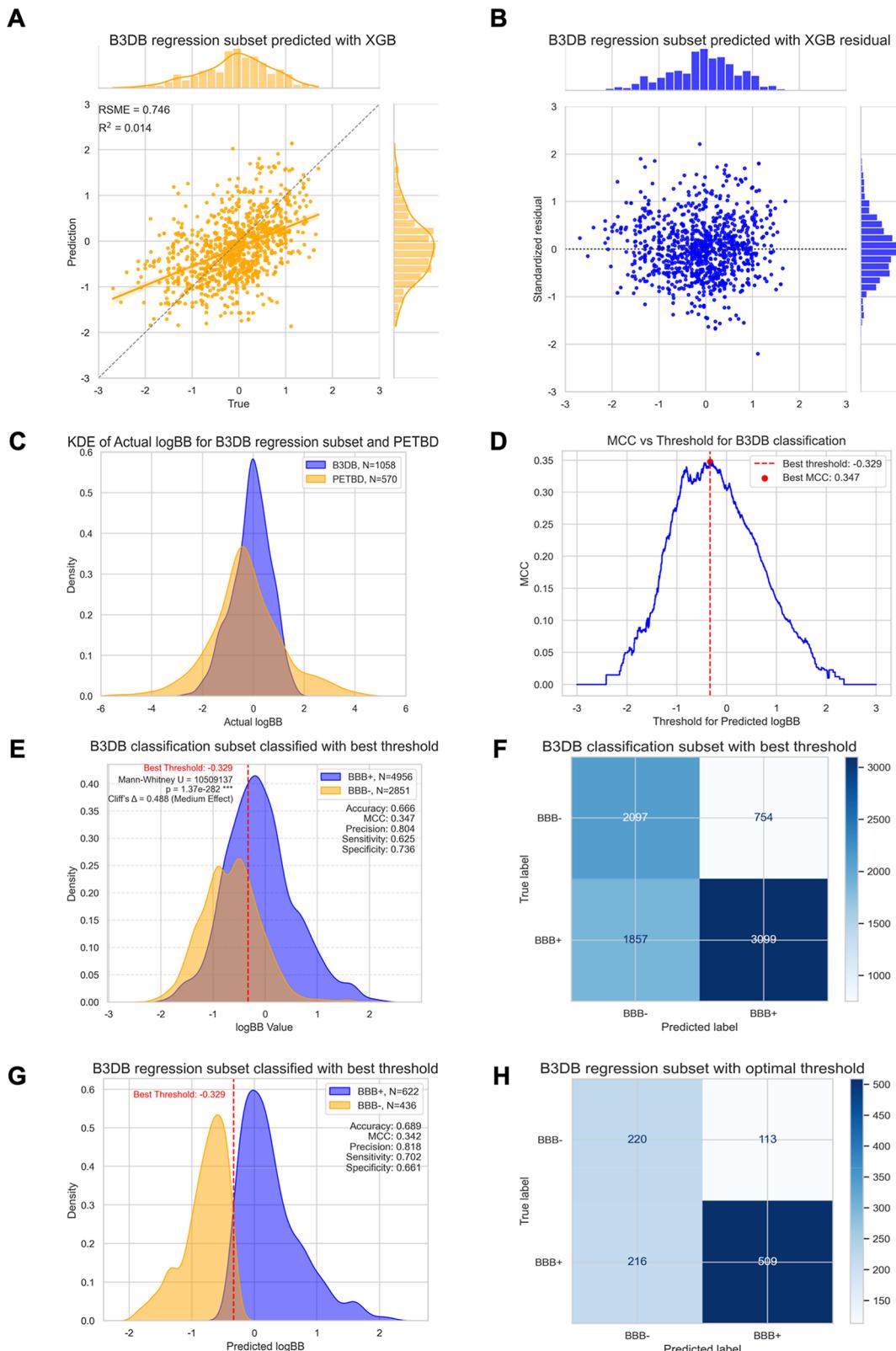


Figure 8. Validation of log BB prediction using the B3DB data set. (A) Predicted vs true log BB values for the regression subset ($n = 1058$). (B) Residual distribution between predicted and true log BB values. (C) Overlapping density distributions of true log BB values for the B3DB data set and the PETBD data set. (D) Threshold optimization based on MCC score maximization using the with the BBB+/BBB– label and predicted log BB of B3DB classification subset. The peak indicates the selected optimal threshold (red dot, -0.329), proposed as the cutoff log BB threshold for binary classification of BBB+/BBB–. (E) Predicted log BB values classified with cutoff threshold for the classification subset of B3DB (BBB +/BBB–). (F) Confusion matrix of the B3DB classification subset ($n = 7807$), with predicted log BB values classified using the cutoff threshold and actual labels. (G) Predicted log BB values classified with cutoff threshold for the B3DB regression subset. (H) Confusion matrix comparing classified labels (predicted log BB classified using the cutoff threshold) vs actual labels in the B3DB regression subset.

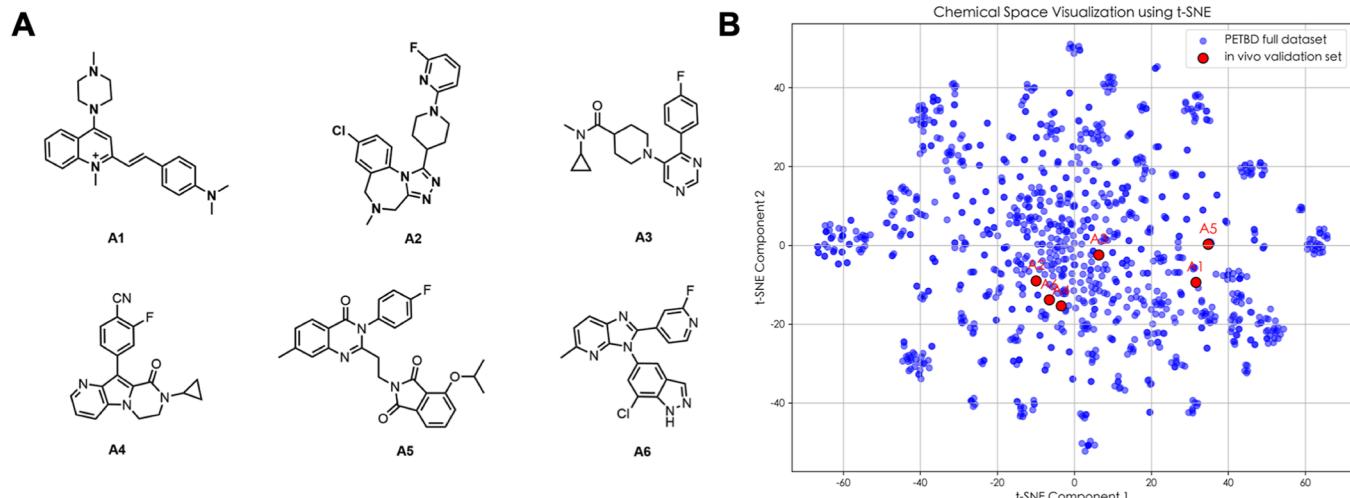


Figure 9. (A) Chemical structure of PET tracers from *in vivo* PET assays for external validation. (B) Chemical space of PET tracers of *in vivo* validation set and PETBD full data set.

existing knowledge and intuition that higher body weight often results in lower drug concentrations in the brain, as further illustrated by SHAP plot (Figure 7B). Additionally, typical polarity-related features, such as TopoPSA, JGI3, GGI4, GGI9, and GATS 3p, revealed that molecular polarity significantly affects the BBB permeability. Among these, TopoPSA also exhibited a positive correlation with well-known descriptors associated with BBB permeability, including MolWt, TPSA, NumHDonors, and NumHAcceptors (Figure 7C).

Predicting the dynamic concentrations of drugs in the brain is an exceptionally challenging endeavor, so much so that it has rarely been effectively explored. Data on drug tissue distribution is infrequently assembled into uniform data sets due to the heterogeneity of PK studies. Furthermore, the inherent data noise in medical research contributes to the scarcity of comprehensive data. While our study pioneered in addressing this complex challenge, it also demands further optimization of prediction performance since intricate mechanisms, including passive diffusion and active transport, may be involved in the process of drugs crossing the BBB.

Small molecules cross the BBB via passive diffusion if they are generally lipophilic (high log *P*), small (under 400–500 Da), neutrally charged at physiological pH, and possess low hydrogen bonding capacity (low PSA), allowing them to readily dissolve in and traverse the lipid membranes. Conversely, active transport is utilized for molecules that do not meet these criteria, such as essential nutrients such as glucose and amino acids, which are shuttled in by specific influx transporters (e.g., GLUT, LAT). This energy-dependent process also transports molecules with specific charges or structures via dedicated transporters and includes protective efflux systems (such as P-gp and BCRP) that actively pump various substances, including some lipophilic ones, out of the brain, thereby limiting their accumulation.

In future studies, we will consider incorporating data on the binding affinity of small molecules to GLUT, LAT, and P-gp, or their structural similarity to the substrates of these transporters, into our data set, or developing “Mixture of Experts” models. It is anticipated that by integrating factors related to passive diffusion and active transport we can more effectively evaluate the relationship between drug structure and BBB permeability.

Validation with B3DB Data Set. To assess the generalizability of our model, validation was performed on the external data set, B3DB, the largest publicly available resource for BBB permeability data set.²⁵ The regression subset ($n = 1058$ compounds with numerical log BB values) and classification subset ($n = 7807$ compounds with categorical BBB+/BBB− labels) were analyzed separately. Notably, its regression subset is also incorporated into the classification subset, with log BB = −1.0 serving as the threshold to discriminate between BBB+ and BBB− compounds.

The XGBoost model trained on SMILES-derived molecular descriptors demonstrated moderate predictive capability for log BB values in the B3DB regression subset, yielding an R^2 of 0.014 and a RMSE of 0.746 (Figure 8A,B). While the RMSE fell within an acceptable range for preliminary screening applications, the low R^2 suggests limited precision compared to its performance on the internal PETBD test set ($R^2 = 0.770$, and RMSE = 0.753). This discrepancy may reflect inherent variability in experimental protocols or chemical diversity across data sets, potentially limiting extrapolation to external data sets or experimental conditions. Consequently, further model optimization is warranted.

The log BB values in our PETBD data set exhibited wider distribution than those in the regression subset of B3DB (Figure 8C). To transform our regression model into a binary classifier, we applied it to this third-party data set (B3DB’s classification subset, 7807 entries with BBB+/BBB− labels). A new threshold for discriminating BBB+ from BBB was optimized to −0.329 (Figure 8D) by maximizing the MCC score, aiming to achieve the best discriminatory ability across both the regression and classification subsets. This optimization yielded an accuracy of 0.667 and an MCC score of 0.347 for the classification subset (Figure 8E,F). The Mann–Whitney *U* test revealed a highly significant difference between groups ($U = 10,509,137$, $p < 0.001$), with BBB + compounds exhibiting higher log BB values (Figure 8E). The effect size, quantified by Cliff’s Delta ($\Delta = 0.488$), exceeded the threshold for a “large” effect ($\Delta \geq 0.474$), suggesting meaningful discriminative power. However, substantial distribution overlap persists, underscoring limitations in binary classification based solely on log BB.

For the regression subset ($n = 1058$) of B3DB, our model exhibited substantial discriminatory ability when applying the proposed cutoff threshold of $\log \text{BB} = -0.329$ to the predicted $\log \text{BB}$ values. Analysis of the distribution plot of $\log \text{BB}$ values labeled as BBB+ or BBB- (Figure 8G) and the confusion matrix of predicted and actual labels (Figure 8H) revealed the following performance metrics: accuracy = 0.689, MCC = 0.342, sensitivity = 0.702, specificity = 0.661, and precision = 0.818. These results collectively suggested the considerable performance of our model, when combined with the proposed threshold, serving as a novel classification model. Notably, the proposed threshold ($\log \text{BB} = -0.329$) exceeds the conventional threshold for BBB- classification ($\log \text{BB} = -1.0$)²¹ suggesting a more conservative classification strategy driven by data-driven optimization.

While these metrics indicate moderate discriminative performance, caution is warranted when integrating the PETBD and B3DB data sets due to potential batch effects (e.g., experimental protocol variability) or domain shifts (e.g., divergent chemical space coverage). Systematic optimization, such as domain adaptation or ensemble modeling, is recommended to enhance cross-data set generalizability.

In Vivo Validation. In this study, *in vivo* validation played a pivotal role in evaluating the predictive power of the models in real-world applications. We obtained 6 unpublished PET tracers from multiple research institutes, and their biodistribution data were utilized as an independent test set (chemical structures and chemical space in Figure 9 and SMILES notation in Table 9), distinct from the PETBD training

Table 9. SMILES of PET Tracers from In Vivo Assays for External Validation

index	SMILES	max Tanimoto similarity ^a
A1	CN1CCN(c2 cm ³ (/C=C/c3ccc(N(C)C)cc3)[n]+](C)c3cccc23)CC1	0.3175
A2	CN1Cc2 cm3(Cl)ccc2-n2c(nnc2C2CCN(c3cccc(F)n3)CC2)C1	0.2222
A3	CN(C(=O)C1CCN(c2cncnc2-c2ccc(F)cc2)CC1)C1CC1	0.3171
A4	N#Cc1ccc(-c2c3n(c4cccnc24)CCN(C2CC2)C3=O)cc1F	0.2535
A5	Cc1ccc2c(=O)n(-c3ccc(F)cc3)c(CCN3C(=O)e4cccc(OC(C)C)c4C3=O)nc2c1	0.8382
A6	Cc1ccc2nc(-c3ccnc(F)c3)n(-c3 cm3(Cl)c4[nH]nc4c3)c2n1	0.2208

^aNote: Tanimoto similarity was calculated using 2048-bit Morgan fingerprints with RDKit. A value below 0.7 is generally considered as structurally distinct.

data set. The maximum Tanimoto similarity between each tracer and any entry in the PETBD full set was as stated in Table 9 calculated using Morgan fingerprints and dimension-

ally reduced with t-SNE, indicating significant structural dissimilarity between the molecules of the *in vivo* validation set and PETBD. A lower value indicates greater structural novelty relative to PETBD.

The models trained and tuned on the PETBD data set were applied on the *in vivo* validation set, with actual experimental concentration in brain (C_{brain}) are summarized in Table 10. Based on the results, XGBoost emerged as the most stable and accurate model for predicting C_{brain} values, demonstrating consistent predictive performance on the PETBD test set. CatBoost, LightGBM, and RF, although slightly less effective than XGBoost, consistently delivered reliable predictive results across most scenarios. In contrast, the remaining models, including MLP and SVM, exhibited moderate performance and did not demonstrate significant advantages. We concluded that the models trained on our data set demonstrate satisfactory reliability for C_{brain} prediction.

Model Output Interpretation. In the study, we developed machine-learning models for predicting C_{brain} and $\log \text{BB}$. The output of the prediction models C_{brain} corresponds to the dynamic concentration in the brain at a time point of 60 min post intravenous injection according to the intrinsic nature of the PETDB that was derived from PET tracer biodistribution studies. While C_{brain} is initially reported in % ID/g (percentage of injected dose per gram of tissue), it can be directly converted to PK concentration units (e.g., ng/mL or μM) using the injected dose (D_{inj}) via the equation below. This conversion holds regardless of whether D_{inj} is expressed in mass (e.g., milligrams per kilogram) or molar (e.g., millimolar units) units, enabling flexible application across experimental designs. To the best of our knowledge, this work represents the first predictive framework for estimating absolute brain drug concentrations from structural descriptors, bridging the gap between PET tracer biodistribution data and traditional PK modeling.

$$\% \text{ ID/g} = \frac{\frac{D_{\text{tissue}}}{D_{\text{inj}}} \times 100}{W_{\text{tissue}}} = \frac{\frac{D_{\text{tissue}}}{W_{\text{tissue}}}}{\frac{D_{\text{inj}}}{D_{\text{inj}}}} \times 100 = \frac{C_{\text{tissue}}}{D_{\text{inj}}} \times 100$$

$$C_{\text{tissue}} = \frac{\% \text{ ID/g} \times D_{\text{inj}}}{100}$$

whereas: % ID/g: percent injected dose per gram of tissue or organ, unit: %/g. D_{tissue} : dosage in tissue, typically measured as radioactivity with a gamma-counter or PET scanner typically in unit of Ci or Bq for PET tracers. D_{inj} : total injected dose. W_{tissue} : weight of tissue in units of gram. C_{tissue} : concentration in tissue, radioactivity per gram.

Quantifying the absolute brain drug concentration (C_{brain})—rather than relying solely on BBB permeability or discrete BBB ± classifications—is critical for optimizing therapeutic dosing regimens. This is because BBB permeability alone may not

Table 10. C_{brain} Prediction of In Vivo Validation Set by Models Tuned on PETBD

index	experimental	XGBoost	RF	CatBoost	LightGBM	MLP	SVM
A1	4.303	3.951	3.750	3.511	4.105	2.890	1.852
A2	0.196	0.356	0.651	0.512	0.433	1.152	1.031
A3	0.691	0.812	0.899	0.755	0.931	1.430	1.215
A4	0.991	1.150	1.101	1.354	1.218	0.955	1.350
A5	0.306	0.421	0.713	0.63	0.550	1.25	1.116
A6	0.076	0.298	0.590	0.488	0.379	0.875	0.988

accurately reflect drug concentration and efficacy, let alone the classification prediction of BBB \pm . For both CNS and non-CNS indications, achieving therapeutic success requires brain concentrations to exceed the minimum effective concentration (MEC) while remaining below the maximum safe concentration (MSC)—a balance that cannot be attained through permeability-based predictions alone.

Predicting C_{brain} directly offers greater utility in preclinical drug development, as it bypasses the limitations of log BB or $K_{\text{p},\text{uu,brain}}$ and eliminates the need for concurrent blood concentration data.^{41,42} This approach enables researchers to estimate therapeutically actionable doses while accounting for dynamic factors such as metabolism, protein binding, and efflux transporter activity. By bridging PK parameters with pharmacodynamic outcomes, C_{brain} prediction models advance translational drug discovery beyond binary permeability classifications.

Moreover, the prediction output C_{brain} can be contextualized within clinical PET imaging by linking it to SUV, a widely used metric in radiotracer studies. SUV is the ratio of the image-derived radioactivity concentration (C_{ROI}) and the whole-body concentration of the injected radioactivity (C_{inj}).⁴³ Predicted C_{brain} could also be interpreted as SUV to provide a reference for designing PET tracers. The conversion between the prediction readout (in units of % ID/g) and SUV can be performed using the equations below. This conversion enables the direct interpretation of model-predicted C_{brain} as SUV, bridging computational predictions with clinically actionable PET imaging data. By aligning preclinical PK modeling with clinical SUV thresholds (e.g., tumor detection thresholds of SUV > 2.5), this framework provides a critical reference for optimizing PET tracer design—ensuring sufficient brain uptake while minimizing off-target accumulation.

$$\begin{aligned} \text{SUV} &= \frac{C_{\text{ROI}}}{C_{\text{inj}}} \\ &= \frac{\frac{D_{\text{ROI}}}{W_{\text{ROI}}}}{\frac{D_{\text{inj}}}{W_{\text{body}}}} \\ &= \left(\frac{\frac{D_{\text{ROI}}}{D_{\text{inj}}}}{\frac{W_{\text{ROI}}}{W_{\text{body}}}} \times 100 \right) \times \frac{W_{\text{body}}}{100} \\ &= \% \text{ ID/g} \times \frac{W_{\text{body}}}{100} \\ \% \text{ ID/g} &= \text{SUV} \times \frac{100}{W_{\text{body}}} \end{aligned}$$

whereas: C_{ROI} : radioactive concentration of specific region of interest (ROI, e.g., area, tissue, or an organ). D_{ROI} : dose of radioactivity detected in ROI, typically in units of MBq or mCi. W_{ROI} : weight of ROI, occasionally estimated by its volume. C_{inj} : initial radioactive concentration of whole body. D_{inj} : total injected dose of radioactivity. W_{body} : body weight.

CONCLUSIONS

This study curated the PETBD, the first multiorgan PK resource for 1056 small-molecule PET tracers, and machine learning-driven QSAR models that predict both absolute brain drug concentration (C_{brain}) and the log BB with robust

accuracy. Cross-validation against the B3DB benchmark data set—the largest publicly available BBB permeability resource—demonstrated strong generalizability across its regression and classification subsets, affirming the model's utility in diverse chemical and pharmacological contexts. Prospective validation with unpublished murine data further confirmed its translational fidelity. While direct conversion of C_{brain} to SUV bridged preclinical predictions with clinical PET imaging thresholds. By integrating structural, temporal, and multiorgan distribution data, the model offers transformative utility in virtual screening, preclinical design, and therapeutic window estimation, addressing critical gaps in CNS drug development. Future work will expand PETBD size with multispecies PKs and mechanistic transport dynamics, refining hybrid models for enhanced precision. Open access to the data set, models, and integration scripts foster collaborative innovation, accelerating discovery and improving translational success rates in CNS therapeutics.

EXPERIMENTAL SECTION

Data Source

Resource data were collected from the public literature (PubMed, Web of Science, and Google Scholar et al.). Retrieved literature was filtered by PET tracers, small molecules, containing drug concentrations in brain and blood. Molecules containing heavy metal atoms are not included.

Data Collecting

For each research article, the chemical structure of every PET tracer was recorded in a separate MOL file. These structures were either manually drawn or identified using the software KingDraw. The biodistribution profiles of PET tracers across multiple organs at various time points were presented in parameters such as % ID/g, % ID/mL, or standard uptake value (SUV), typically in the form of tables, histograms, or scatter plots (sometimes with curves). Tabulated data were manually extracted and formatted into a standardized Excel worksheet. For histogram data, column heights were quantified into numerical values using the tool WebPlotDigitizer. The drug concentration of PET tracers was expressed as the radioactive concentration and recorded as % ID/g, which represents the percentage of injected dosage of a radiolabeled PET tracer per gram of a specific organ. This data originated from biodistribution studies of PET tracers, conducted mostly under standard operating procedures, and measured using a specific radio instrument (gamma counter). Data in units of % ID/g were extracted directly from tables. Data in % ID/mL were treated as equivalent to % ID/g, assuming an organ density of approximately 1 g/mL. Similarly, SUV, as a parameter of clinical use, represents ratio of radioactivity concentration within a region at a time and the injected dose of radioactivity per kilogram of the patient's body weight. SUV data were converted to % ID/g by using the equation below, assuming an organ density of approximately 1 g/mL.

$$\% \text{ ID/g} = \text{SUV} \times \frac{100}{W_{\text{body}}}$$

where: W_{body} is the average body weight of the experimental animals in the research.

Data Integration

Data from over 1600 PET tracers (each comprising a chemical structure MOL file and an Excel file) were systematically compiled into a database using the self-developed tool, *MedicalDatasetsMerger*. Each database entry includes the compound index, SMILES notation, chemical structure, and organ-specific concentrations (unit: % ID/g) at various time points. To ensure consistency, organ names expressed in different forms (e.g., lung/lungs, blood/plasma, bone/bones/femur) were unified (refer to the merge_constants_config.yml file of

MedicalDatasetsMerger on GitHub for details). For PET tracers with drug distribution data in specific brain regions (e.g., cortex and hippocampus), subregion concentrations were weighted by their relative mass or volume and unified as whole-brain concentrations according to the equation below and relative fractions in Table 11. Drug concentrations in less frequently studied organs were excluded during data integration.

Table 11. Relative Fraction of the Specific Subregion with Respect to the Whole Brain^{44–46}

brain subregion	relative fraction (f_i , %)
whole brain	100
cerebellum	9.1
cortex	44.5
striatum	8.4
hippocampus	6.5
hypothalamus	2.7
thalamus	2.2
cerebrum	9.1

The whole brain drug concentration (C_{brain}) can be calculated using the following equation

$$C_{\text{brain}} = \frac{\sum_i (C_i \times f_i)}{\sum_i f_i}$$

where: C_i : the drug concentration in a specific subregion if applicable in Table 11. f_i : the relative fraction of the specific subregion with respect to the whole brain.

Data Cleaning

Eliminate Anomalous Data. PET tracers that fell into the criteria below were considered as obviously abnormal data and thus were eliminated from the data set.

- Drug concentrations in brain (C_{brain}) larger than 10, which means more than 10% of injected drug distributed in per gram brain tissue.
- Molecular weight of PET tracers is larger than 1000, which may be nanoparticles, peptides, nucleic acids beyond the scope of small molecules.
- Entries detected by Elliptic Envelope algorithm⁴⁷ from *Scikit-learn* library⁴⁸ as outliers.

Merge Duplicate Tracers. For those tracers with identical chemical structures (including chirality) and distinct biodistribution data from different research, entries of duplicated tracers were merged following rules below.

- One identical chemical structure from 2 entries, if ratio of maximum and minimum larger than 5, both entries eliminated; otherwise, average values were kept.
- One identical chemical structure from 3 or more entries, if ratio of maximum and minimum larger than 5, all entries were eliminated; if ratio less than 2, average values were kept; otherwise, remove the entry with furthest values away from average values until the ratio less than 2.

Collate Chemical Structure. The chemical structures of PET tracers were obtained from the published research literature. Electronic representations of these structures were generated using KingDraw, primarily leveraging its image recognition functionality. In the chemical structures, radioactive isotopes (e.g., ¹⁸F, ¹¹C) were recorded in their nonradioactive forms. The MOL files were converted into isomeric SMILES (rather than canonical SMILES) using Python package RDKit. Isomeric SMILES specifically account for isotopic labeling and stereochemistry, ensuring molecular accuracy. A final manual examination was conducted to verify the correctness of the PET tracer structures, including atoms, bonds, electronic charges, and chirality.

Data Set Formulation. Considering the time point preference in PET tracer studies, drug concentrations at 60 min postintravenous injection across 14 organs—including the brain, blood, liver, lung, kidney, muscle, spleen, intestine, bone, stomach, uterus, heart, fat, and pancreas—were selected. Log BB, defined as the logarithm of the ratio of PET tracer concentration in the brain to its concentration in the blood, was calculated and incorporated into the data set. Alongside compound indices and SMILES representations, PMID identifier, animal type, gender, animal weight and injection dosage, a data set containing 1056 PET tracers were compiled. The data set is stored in comma-separated values (CSV) format and is openly accessible on GitHub (<https://github.com/GDUT-Computer-Medical-Science-Team/PETBD-QSAR/>).

Model Construction and Evaluation. This section provides a comprehensive overview of the processes involved in constructing and evaluating our predictive models. The steps include descriptor generation to create robust data sets, prepossessing of target data to ensure accuracy, and model training using multiple machine-learning techniques. We then assess the model performance through various qualitative metrics, ensuring the effectiveness and reliability of our predictive models. These steps are crucial in validating our data set and confirming the robustness of our machine-learning approaches. Furthermore, to strengthen our validation process, both *in silico* and *in vivo* validations are incorporated.

Feature Generation. To meet to the requirements of machine-learning input and identify features with higher predictive importance, the Modred descriptors (1613 features) and Morgan fingerprints (1024-bit) features were generated from SMILES strings using the *Modred*⁴⁹ and *RDKit* packages.⁵⁰ The *in vivo* metadata, consisting of animal type, gender, animal weight, and injection dosage, represent the most characteristic attributes of the PETBD data set. For categorical metadata encoding: animal type was encoded as 1 for mice (including substrains) and 0 for rats, with unknown values left as NaN; gender was encoded as 1 for female and 0 for male, with unknown values also retained as NaN. The computed molecular descriptors, fingerprints, and *in vivo* metadata collectively formed the fused feature set (2741 features in total) for subsequent feature selection and machine-learning model training.

Target Data Preprocess. In this study, drug concentration data of 14 organs at the time point of 60th minute were contained. Elliptic Envelope algorithm⁴⁷ from *Scikit-learn* was then applied to remove outliers from the concentration data. Outliers were removed and set to NaN. C_{brain} as well as the log BB, defined as the logarithm of the ratio of concentration of PET tracer in the brain to its concentration in the blood, were selected as targets, respectively.

Model Training. Support Vector Machine (SVM), CatBoost, Random Forest (RF), LightGBM, Extreme Gradient Boosting (XGB), and Multi-Layer Perceptron (MLP) were used for the prediction task of organ concentration at specific time points. These models were selected for their robustness and ability to handle complex data sets, thereby providing a comprehensive evaluation of our data.

To ensure a robust and unbiased evaluation, the data set was first partitioned into a training set (81%), a validation set (9%), and a held-out test set (10%) using stratified sampling to preserve the original isotope label distribution across all splits. To mitigate the bias introduced by the imbalanced distribution of ¹⁸F-labeled compounds, resampling strategies, either oversampling or undersampling, was applied exclusively to the training set. Specifically, the training set was first subdivided into a majority class (samples containing ¹⁸F) and a minority class (samples not containing ¹⁸F). In the oversampling approach, the minority class was randomly sampled with replacement, until its size matched that of the majority class. Conversely, for undersampling, the majority class was randomly downsampled without replacement to align with the size of the minority class. Following either resampling method, the augmented subset was combined with the remaining original subset and subjected to global shuffling, resulting in a newly balanced training set.

All model development, including hyperparameter optimization, was conducted exclusively on the 80% training set. Prior to model training, a preprocessing pipeline was established for feature

imputation and selection. For missing values (NaN) in molecular descriptors, fingerprints, and metadata, SimpleImputer was employed using median imputation for numerical features and most frequent imputation for categorical features. Subsequently, RFE was applied to select the 50 most representative features (with 4 metadata features forcibly retained) before proceeding to model training. Then we employed a 10-fold cross-validation strategy coupled with Optuna, a Bayesian optimization framework. For each fold, Optuna systematically explored the hyperparameter space to identify the configuration that maximized the model's predictive performance.

The best hyperparameter set, determined by the average performance across all 10-fold, was used to train a final model on the entire training set. This final model was then evaluated on the validation set, and its ultimate generalization performance was assessed on the held-out test set.

Model Performance Metrics. For the regression application, the performance of the developed models was evaluated using five metrics: root-mean-square error (RMSE), mean absolute error (MAE), MSE, mean absolute percentage error (MAPE), and coefficient of determination (R^2). RMSE refers to the standard deviation of the residuals or prediction errors. MAE assesses the average absolute discrepancy between the predicted and actual values. MSE calculates the average of the squares of the errors. MAPE measures the average absolute percentage error between predicted and actual values. R^2 ranges from 0 to 1 and represents the proportion of the variance in the dependent variable that is predictable from that of the independent variable. A well-performing model would typically exhibit low RMSE, MAE, MSE, and MAPE values while having an R^2 value approaching 1.

Feature Importance Analysis. To rigorously validate the robustness and stability of our models, the entire process, including feature selection, model training, and cross-fold validation, was independently executed for 30 iterations, each with a distinct random seed. Subsequently, the consistency of the selected features and their importance across these iterations were statistically analyzed.

Permutation importance and SHAP analysis are employed to explain the best trained final models. The TreeExplainer efficiently calculates the SHAP values for each feature in tree-based models, quantifying their positive or negative impact on the model's output. The sign of the SHAP value indicates whether the feature enhances or reduces the predicted value, and the absolute value reflects its influence.

Additionally, the correlation of the top 20 selected features with well-known features that are generally regarded as relevant to the BBB permeability was calculated to elucidate the significance of these features.

Validation with B3DB Data Set

Log BB Prediction of B3DB Data Set Regression Subset. The regression subset of the B3DB data set ($n = 1058$ entries) was subjected to log BB value prediction using our optimized model (XGBoost trained on fused feature set). As B3DB lacked specific metadata (i.e., animal type, gender, animal weight, and injection dosage), an imputation strategy was applied, consistent with the feature preprocessing used during model training. Specifically, a SimpleImputer was utilized: "median" was adopted for numerical features, and "most_frequent" for categorical features. During this prediction phase, the imputed values for the metadata were revealed as follows: "animal type" = "rat", "gender" = "male", "animal weight (g)" = 250.0, and "injection dosage (μ Ci)" = 100.0. The model's performance was evaluated on the external data set to assess its generalizability. Residuals between predicted and actual log BB values within the regression subset were calculated and visualized to assess the model's accuracy of this specific data set.

Log BB Prediction of B3DB Data Set Classification Subset. The classification subset of the B3DB data set ($n = 7807$ entries) was evaluated using our top-performing model, Random Forest. Predicted log BB values were statistically compared between BBB+ and BBB- labeled compounds. To address class imbalance (BBB+/BBB- = 4956:2851), the optimal classification threshold was determined by

maximizing the Matthews correlation coefficient (MCC) score. A confusion matrix was subsequently generated at this threshold to quantify the classification performance.

Statistical Analysis

The differences in predicted log BB values between BBB+ and BBB- compounds were analyzed using nonparametric methods due to the non-normal distribution of the data (assessed via Shapiro-Wilk tests). Homogeneity of variances was confirmed using the Levene's test ($p > 0.05$). A Mann-Whitney U test was employed to compare the central tendencies of the two groups. Effect size was quantified using Cliff's Delta (Δ), interpreted as follows: $\Delta < 0.147$ (negligible), $0.147 \leq \Delta < 0.33$ (small), $0.33 \leq \Delta < 0.474$ (medium), and $\Delta \geq 0.474$ (large). All analyses were performed by using Python (SciPy and Statsmodels libraries).

In Vivo Validation

To validate the model performance, we selected 6 additional unpublished PET tracers collected from PET research institutes. These compounds and their biodistribution data were used as an independent test data set for external validation.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jmedchem.Sc01791>.

Sampling strategies and performance; determination of feature number with RFE; prediction performance of C_{brain} and log BB with Mordred molecular descriptors and Morgan molecular fingerprints respectively; comparison prediction error of ^{18}F and non- ^{18}F subgroups in PETBD hand-out test set; and SMILES strings of identical entries in PETBD and B3DB ([PDF](#))

AUTHOR INFORMATION

Corresponding Authors

Wei Zhou – School of Biomedical and Pharmaceutical Sciences, Guangdong University of Technology, Guangzhou 510006, PR China;  orcid.org/0000-0003-4030-2022; Email: zhou_wei@gdut.edu.cn

Zhiyun Du – School of Biomedical and Pharmaceutical Sciences, Guangdong University of Technology, Guangzhou 510006, PR China; Email: zhiyundu@gdut.edu.cn

Authors

Qing Su – School of Computer Science, Guangdong University of Technology, Guangzhou 510006, PR China

Zhilong Ye – School of Computer Science, Guangdong University of Technology, Guangzhou 510006, PR China

Chunyan Han – School of Computer Science, Guangdong University of Technology, Guangzhou 510006, PR China

Ganyao Xiao – School of Computer Science, Guangdong University of Technology, Guangzhou 510006, PR China

Jiazhui Chen – School of Biomedical and Pharmaceutical Sciences, Guangdong University of Technology, Guangzhou 510006, PR China

Haiyan Chen – School of Biomedical and Pharmaceutical Sciences, Guangdong University of Technology, Guangzhou 510006, PR China

Zigui Wang – School of Biomedical and Pharmaceutical Sciences, Guangdong University of Technology, Guangzhou 510006, PR China

Shun Huang – Department of Nuclear Medicine, The Tenth Affiliated Hospital of Southern Medical University

(Dongguan People's Hospital), Dongguan 523059, PR China; orcid.org/0000-0002-0576-9308

Lu Wang – Center of Cyclotron and PET

Radiopharmaceuticals, Department of Nuclear Medicine and PET/CT-MRI Center, The First Affiliated Hospital of Jinan University, Guangzhou 510630, PR China; orcid.org/0000-0002-8049-1991

Grace Peng – Life Science Solution, PatSnap Information Technology (Suzhou) Co., Ltd., Suzhou 215009, PR China

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.jmedchem.5c01791>

Author Contributions

This manuscript was written through the collaborative efforts of all authors. W.Z. conceptualized the project, wrote and reviewed the manuscript. Z.D. acquired financial support. Q.S. designed the machine learning models and supervised the research. Z.Y., C.H., and G.X. optimized machine-learning models. J.C., H.C., and Z.W. collected and curated the data set. S.H and L.W. provided *in vivo* validation. G.P. coordinated computational resources.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (grant number 22178070). We sincerely thank Professor Yongmin Zhang from Sorbonne University for his valuable and constructive advice. The authors also wish to thank colleagues for their insightful discussions and encouragement. We thank PatSnap for providing data analysis capabilities and computational resources to support the research. Schematic figures were created with <https://BioRender.com>.

ABBREVIATIONS

BBB, blood–brain barrier; Bq, becquerel; CNS, central nervous system; CSV, comma-separated values; C_{brain} , brain drug concentration; % ID/g, percentage of injected dose per gram of tissue; GCNN, genetic convolutional neural network; KNN, K-nearest neighbors; $K_{\text{p,uu,brain}}$, unbound brain-to-plasma partition coefficient; LightGBM, light gradient boosting machine; log BB, logarithm of brain-to-blood concentration ratio; MAE, mean absolute error; MAPE, mean absolute percentage error; MCC, Matthews correlation coefficient; MBq, megabecquerel; MEC, minimum effective concentration; MICAD, molecular imaging and contrast agent database; MLP, multi-layer perceptron; MSC, maximum safe concentration; PET, positron emission tomography; PETBD, PET tracer biodistribution data set; PK, pharmacokinetics; QSAR, quantitative structure–activity relationship; RF, random forest; RFE, recursive feature elimination; RMSE, root mean square error; R^2 , coefficient of determination; SMILES, simplified molecular input line entry system; SUV, standardized uptake value; SVM, support vector machine; t-SNE, t-distributed stochastic neighbor embedding; XGBoost, extreme gradient boosting

REFERENCES

- (1) Gupta, M.; Lee, H. J.; Barden, C. J.; Weaver, D. F. The Blood–Brain Barrier (BBB) Score. *J. Med. Chem.* **2019**, *62* (21), 9824–9836.
- (2) Hanafy, A. S.; Dietrich, D.; Fricker, G.; Lamprecht, A. Blood–Brain Barrier Models: Rationale for Selection. *Adv. Drug Delivery Rev.* **2021**, *176*, 113859.
- (3) Zhang, L.; Villalobos, A. Strategies to Facilitate the Discovery of Novel CNS PET Ligands. *EJNMMI Radiopharm. Chem.* **2017**, *1* (1), 13.
- (4) Lindberg, A.; Chassé, M.; Varlow, C.; Pees, A.; Vasdev, N. Strategies for Designing Novel Positron Emission Tomography (PET) Radiotracers to Cross the Blood–Brain Barrier. *Labelled Comp Radiopharmac.* **2023**, *66* (9), 205–221.
- (5) Jackson, I. M.; Webb, E. W.; Scott, P. J. H.; James, M. L. In Silico Approaches for Addressing Challenges in CNS Radiopharmaceutical Design. *ACS Chem. Neurosci.* **2022**, *13* (12), 1675–1683.
- (6) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug Delivery Rev.* **1997**, *23* (1–3), 3–25.
- (7) Leeson, P. D.; Davis, A. M. Time-Related Differences in the Physical Property Profiles of Oral Drugs. *J. Med. Chem.* **2004**, *47* (25), 6338–6348.
- (8) Hitchcock, S. A.; Pennington, L. D. Structure–Brain Exposure Relationships. *J. Med. Chem.* **2006**, *49* (26), 7559–7583.
- (9) Spielvogel, C. P.; Schindler, N.; Schröder, C.; Stellnberger, S. L.; Wadsak, W.; Mitterhauser, M.; Papp, L.; Hacker, M.; Pichler, V.; Vraka, C. Enhancing Blood–Brain Barrier Penetration Prediction by Machine Learning-Based Integration of Novel and Existing, In Silico and Experimental Molecular Parameters from a Standardized Database. *J. Chem. Inf. Model.* **2025**, *65* (6), 2773–2784.
- (10) Steén, E. J. L.; Vugts, D. J.; Windhorst, A. D. The Application of In Silico Methods for Prediction of Blood–Brain Barrier Permeability of Small Molecule PET Tracers. *Front. Nucl. Med.* **2022**, *2*, 853475.
- (11) Dehnboestel, F. O.; Dixit, V. A.; Preissner, R.; Banerjee, P. Non-Animal Models for Blood–Brain Barrier Permeability Evaluation of Drug-like Compounds. *Sci. Rep.* **2024**, *14* (1), 8908.
- (12) Wager, T. T.; Hou, X.; Verhoest, P. R.; Villalobos, A. Moving beyond Rules: The Development of a Central Nervous System Multiparameter Optimization (CNS MPO) Approach To Enable Alignment of Druglike Properties. *ACS Chem. Neurosci.* **2010**, *1* (6), 435–449.
- (13) Wager, T. T.; Hou, X.; Verhoest, P. R.; Villalobos, A. Central Nervous System Multiparameter Optimization Desirability: Application in Drug Discovery. *ACS Chem. Neurosci.* **2016**, *7* (6), 767–775.
- (14) Huang, E. T. C.; Yang, J.-S.; Liao, K. Y. K.; Tseng, W. C. W.; Lee, C. K.; Gill, M.; Compas, C.; See, S.; Tsai, F.-J. Predicting Blood–Brain Barrier Permeability of Molecules with a Large Language Model and Machine Learning. *Sci. Rep.* **2024**, *14* (1), 15844.
- (15) Dichiara, M.; Amata, B.; Turnaturi, R.; Marrazzo, A.; Amata, E. Tuning Properties for Blood–Brain Barrier Permeation: A Statistics-Based Analysis. *ACS Chem. Neurosci.* **2020**, *11* (1), 34–44.
- (16) Cherian Parakkal, S.; Datta, R.; Das, D. DeepBBBP: High Accuracy Blood-brain-barrier Permeability Prediction with a Mixed Deep Learning Model. *Mol. Inf.* **2022**, *41* (10), 2100315.
- (17) Carracedo-Reboredo, P.; Liñares-Blanco, J.; Rodríguez-Fernández, N.; Cadrón, F.; Novoa, F. J.; Carballal, A.; Maojo, V.; Pazos, A.; Fernández-Lozano, C. A Review on Machine Learning Approaches and Trends in Drug Discovery. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 4538–4558.
- (18) Kumar, P.; Saini, V.; Gupta, D.; Chawla, P. A.; Kumar, A. BBBper: A Machine Learning-Based Online Tool for Blood–Brain Barrier (BBB) Permeability Prediction. *CNS Neurol. Disord.: Drug Targets* **2024**, *24*, 0118715273328174241007060331.
- (19) Fu, L.; Shi, S.; Yi, J.; Wang, N.; He, Y.; Wu, Z.; Peng, J.; Deng, Y.; Wang, W.; Wu, C.; Lyu, A.; Zeng, X.; Zhao, W.; Hou, T.; Cao, D. ADMETlab 3.0: An Updated Comprehensive Online ADMET Prediction Platform Enhanced with Broader Coverage, Improved Performance, API Functionality and Decision Support. *Nucleic Acids Res.* **2024**, *52* (W1), W422–W431.

- (20) Feng, T. W.; Edros, R.; Ab Ghani, N.; Umairah Mokhtar, S.; Dong, R.H. Prediction of Blood–Brain Barrier Permeability of Compounds by Machine Learning Algorithms. *ARASET* **2023**, *33* (2), 269–276.
- (21) Shaker, B.; Lee, J.; Lee, Y.; Yu, M.-S.; Lee, H.-M.; Lee, E.; Kang, H.-C.; Oh, K.-S.; Kim, H. W.; Na, D. A Machine Learning-Based Quantitative Model (LogBB_Pred) to Predict the Blood–Brain Barrier Permeability (logBB Value) of Drug Compounds. *Bioinformatics* **2023**, *39* (10), btad577.
- (22) Tong, X.; Wang, D.; Ding, X.; Tan, X.; Ren, Q.; Chen, G.; Rong, Y.; Xu, T.; Huang, J.; Jiang, H.; Zheng, M.; Li, X. Blood–Brain Barrier Penetration Prediction Enhanced by Uncertainty Estimation. *J. Cheminf.* **2022**, *14* (1), 44.
- (23) 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* (20), 2189–2201.
- (24) 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* (8), 1135–1139.
- (25) Meng, F.; Xi, Y.; Huang, J.; Ayers, P. W. A Curated Diverse Molecular Database of Blood–Brain Barrier Permeability with Chemical Descriptors. *Sci. Data* **2021**, *8* (1), 289.
- (26) Geldenhuys, W. J.; Mohammad, A. S.; Adkins, C. E.; Lockman, P. R. Molecular Determinants of Blood–Brain Barrier Permeation. *Ther. Delivery* **2015**, *6* (8), 961–971.
- (27) Loryan, I.; Reichel, A.; Feng, B.; Bundgaard, C.; Shaffer, C.; Kalvass, C.; Bednarczyk, D.; Morrison, D.; Lesuisse, D.; Hoppe, E.; Terstappen, G. C.; Fischer, H.; Di, L.; Colclough, N.; Summerfield, S.; Buckley, S. T.; Maurer, T. S.; Fridén, M. Unbound Brain-to-Plasma Partition Coefficient, $K_{p,\text{uu}}$, Brain—a Game Changing Parameter for CNS Drug Discovery and Development. *Pharm. Res.* **2022**, *39* (7), 1321–1341.
- (28) Kato, R.; Zeng, W.; Siramshetty, V. B.; Williams, J.; Kabir, M.; Hagen, N.; Padilha, E. C.; Wang, A. Q.; Mathé, E. A.; Xu, X.; Shah, P. Development and Validation of PAMPA-BBB QSAR Model to Predict Brain Penetration Potential of Novel Drug Candidates. *Front. Pharmacol.* **2023**, *14*, 1291246.
- (29) Wang, W.; 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* (9), 3055–3065.
- (30) Liu, L.; Zhang, L.; Feng, H.; Li, S.; Liu, M.; Zhao, J.; Liu, H. Prediction of the Blood–Brain Barrier (BBB) Permeability of Chemicals Based on Machine-Learning and Ensemble Methods. *Chem. Res. Toxicol.* **2021**, *34* (6), 1456–1467.
- (31) Rowland, D. J.; Cherry, S. R. Small-Animal Preclinical Nuclear Medicine Instrumentation and Methodology. *Semin. Nucl. Med.* **2008**, *38* (3), 209–222.
- (32) Kadrimas, D. J.; Hoffman, J. M. Methodology for Quantitative Rapid Multi-Tracer PET Tumor Characterizations. *Theranostics* **2013**, *3* (10), 757–773.
- (33) Cabral, E.; Montgomery, M. K.; Berg, M.; Manzuk, L. K.; Giddabasappa, A.; Jiang, Z. K. Preclinical Positron Emission Tomography with Body Conforming Animal Molds for Cloud-Based Automated Image Analysis in Mice. *JOVE* **2024**, *212*, 67370.
- (34) Relevance of Half-Life in Drug Design|Journal of Medicinal Chemistry. <https://pubs.acs.org/doi/abs/10.1021/acs.jmedchem.7b00969> (accessed June 28, 2025).
- (35) Dyabina, A. S.; Radchenko, E. V.; Palyulin, V. A.; Zefirov, N. S. Prediction of Blood–Brain Barrier Permeability of Organic Compounds. *Dokl. Biochem. Biophys.* **2016**, *470* (1), 371–374.
- (36) Muehlbacher, M.; Spitzer, G. M.; Liedl, K. R.; Kornhuber, J. Qualitative Prediction of Blood–Brain Barrier Permeability on a Large and Refined Dataset. *J. Comput. Aided Mol. Des.* **2011**, *25* (12), 1095–1106.
- (37) 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*.
- (38) Gu, Y.; Yu, Z.; Wang, Y.; Chen, L.; Lou, C.; Yang, C.; Li, W.; Liu, G.; Tang, Y. *admetSAR3.0: A Comprehensive Platform for Exploration, Prediction and Optimization of Chemical ADMET Properties*.
- (39) Daina, A.; Zoete, V. A BOILED-Egg To Predict Gastrointestinal Absorption and Brain Penetration of Small Molecules. *ChemMedChem* **2016**, *11* (11), 1117–1121.
- (40) Liu, H.; Wang, L.; Lv, M.; Pei, R.; Li, P.; Pei, Z.; Wang, Y.; Su, W.; Xie, X.-Q. AlzPlatform: An Alzheimer’s Disease Domain-Specific Chemogenomics Knowledgebase for Polypharmacology and Target Identification Research. *J. Chem. Inf. Model.* **2014**, *54*, 1050–1060.
- (41) Liu, H.; Dong, K.; Zhang, W.; Summerfield, S. G.; Terstappen, G. C. Prediction of Brain:Blood Unbound Concentration Ratios in CNS Drug Discovery Employing in Silico and in Vitro Model Systems. *Drug Discovery Today* **2018**, *23* (7), 1357–1372.
- (42) Lawrenz, M.; Svensson, M.; Kato, M.; Dingley, K. H.; Chief Elk, J.; Nie, Z.; Zou, Y.; Kaplan, Z.; Lagiakos, H. R.; Igawa, H.; Therrien, E. A Computational Physics-Based Approach to Predict Unbound Brain-to-Plasma Partition Coefficient, $K_{p,\text{uu}}$. *J. Chem. Inf. Model.* **2023**, *63* (12), 3786–3798.
- (43) Gu, F.; Wu, Q. Quantitation of Dynamic Total-Body PET Imaging: Recent Developments and Future Perspectives. *Eur. J. Nucl. Med. Mol. Imaging* **2023**, *50* (12), 3538–3557.
- (44) Welniak-Kaminska, M.; Fiedorowicz, M.; Orzel, J.; Bogorodzki, P.; Modlinska, K.; Stryjek, R.; Chrzanowska, A.; Pisula, W.; Grieb, P. Volumes of Brain Structures in Captive Wild-Type and Laboratory Rats: 7T Magnetic Resonance in Vivo Automatic Atlas-Based Study. *PLoS One* **2019**, *14* (4), No. e0215348.
- (45) Goerzen, D.; Fowler, C.; Devenyi, G. A.; Germann, J.; Madularu, D.; Chakravarty, M. M.; Near, J. An MRI-Derived Neuroanatomical Atlas of the Fischer 344 Rat Brain. *Sci. Rep.* **2020**, *10* (1), 6952.
- (46) Bolon, B.; Butt, M. T. Appendix 6. Comparative and Correlative Neurobiological Parameters for the Brain. In *Fundamental Neuropathology for Pathologists and Toxicologists*; John Wiley & Sons, Ltd, 2011; pp 551–554.
- (47) Rousseeuw, P. J.; Driessen, K. V. A Fast Algorithm for the Minimum Covariance Determinant Estimator. *Technometrics* **1999**, *41* (3), 212–223.
- (48) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
- (49) Moriwaki, H.; Tian, Y.-S.; Kawashita, N.; Takagi, T. Mordred: A Molecular Descriptor Calculator. *J. Cheminf.* **2018**, *10* (1), 4–14.
- (50) Yap, C. W. PaDEL-descriptor: An Open Source Software to Calculate Molecular Descriptors and Fingerprints. *J. Comput. Chem.* **2011**, *32* (7), 1466–1474.