Molecular Informatics

WILEY-VCH

molecular informatics

REVIEW OPEN ACCESS

Machine Learning in Drug Development for Neurological Diseases: A Review of Blood Brain Barrier Permeability Prediction Models

Aryon Eckleel Nabi[1] | Pedram Pouladvand[4] | Litian Liu[2] | Ning Hua[3] | Cyrus Ayubcha[1,3]

[1]Harvard Medical School, Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA USA | [2]Boonshoft School of Medicine, Wright State University, Dayton, OH USA | [3]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Boston, MA USA | [4]Department of Epidemiology, Harvard Chan School of Public Health, Boston, MA USA

Correspondence: Pedram Pouladvand (pedrampouladvand@gmail.com)

Received: 28 October 2024 | **Revised:** 25 January 2025 | **Accepted:** 5 February 2025

Keywords: artificial neural network (ANN) | blood brain barrier (BBB) | central nervous system (CNS) | deep learning (DL) | deep neural network (DNN) | extra-tree (EXT) | extreme gradient boosting (XGBoost) | fingerprints (FPs) | in silico model | machine learning (ML) | molecular descriptor(s) (MDs) | quantitative structure-activity relationship (QSAR) model | random forest (RF) | recurrent neural network (RNN) | simplified molecular line entry system (SMILES) | support vector machine (SVM)

ABSTRACT

The blood brain barrier (BBB) is an endothelial-derived structure which restricts the movement of certain molecules between the general somatic circulatory system to the central nervous system (CNS). While the BBB maintains homeostasis by regulating the molecular environment induced by cerebrovascular perfusion, it also presents significant challenges in developing therapeutics intended to act on CNS targets. Many drug development practices rely partly on extensive cell and animal models to predict, to an extent, whether prospective therapeutic molecules can cross the BBB. In interest to reduce costs and improve prediction accuracy, many propose using advanced computational modeling of BBB permeability profiles leveraging empirical data. Given the scale of growth in machine learning and deep learning, we review the most recent machine learning approaches in predicting BBB permeability.

Introduction

The Blood Brain Barrier (BBB) is a highly selective cellular film around the brain that inhibits certain compounds and molecules, on the basis of size and biochemical characteristics, from accessing the central nervous system (CNS). It is composed of brain capillary endothelial cells linked by tight junctions, which are surrounded by pericytes, astrocytes, and the basal lamina. Such structures are a major contributor to the high selectivity of the BBB, stopping the passage of 98% of all molecules present in the general circulatory system's blood pool [1, 2]. Figure 1.

Such functions of the BBB allow for CNS homeostasis. Specifically, this is achieved via highly selective ion and solute transport between the intravascular systemic plasma and the CNS vasculature through either passive transport or active transport [1]. Passive transport can either occur transcellularly across BBB endothelial cells or paracellularly, where these forms of passive transport are chiefly a function of molecular weight and lipid solubility [1]. Alternatively, active transport is mediated by influx and efflux transporters which are selective for their particular ion or molecules, such as hormones or proteins [1, 3]. Moreover, the functionality of the BBB has been noted to vary according to genetic disposition and environmental exposures; it is also known that the BBB is altered with

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2025 The Author(s). *Molecular Informatics* published by Wiley-VCH GmbH.

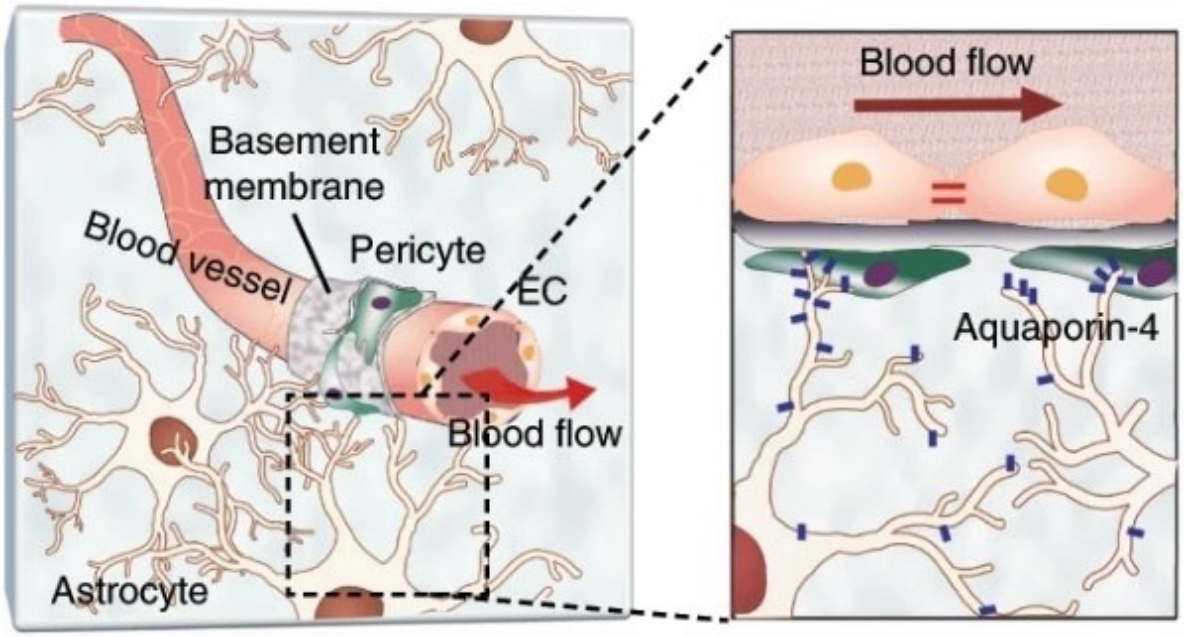


FIGURE 1 | The figure above illustrates the blood brain membrane surrounding cerebrovascular structures. Figure was reproduced with permission from Ahn et al. [102].

aging, infectious or inflammatory processes and neuropathologies such as neurodegenerative diseases [4, 5].

Considering these functions, the BBB prevents many unwanted toxins, antigens, pathogens or other unintended molecular components in the general somatic circulatory system from accessing the CNS. [1] Illustration of the diffusion mechanism of small solutes and drugs is provided in Figure 2. Nevertheless, the BBB often limits the concentration of most small molecule therapeutics leading to inadequate concentrations for therapeutic effects [6]. Unfortunately, BBB permeability is not a simple function [5]. While smaller and more lipophilic molecules often are the most restrictive predictive thresholds, predicting permeability in complex and dynamic settings above remains a significant challenge.

The earliest efforts to model BBB permeability (BBBp) used logP, a measure of lipophilicity, and molecular weight (MW). These models often predicted logBB which reflects the log concentration of drug in the brain/concentration in the blood [7]. Many times researchers classify logBB using a splitting criteria for BBB+ (permeable) and BBB- (permeable) [8].

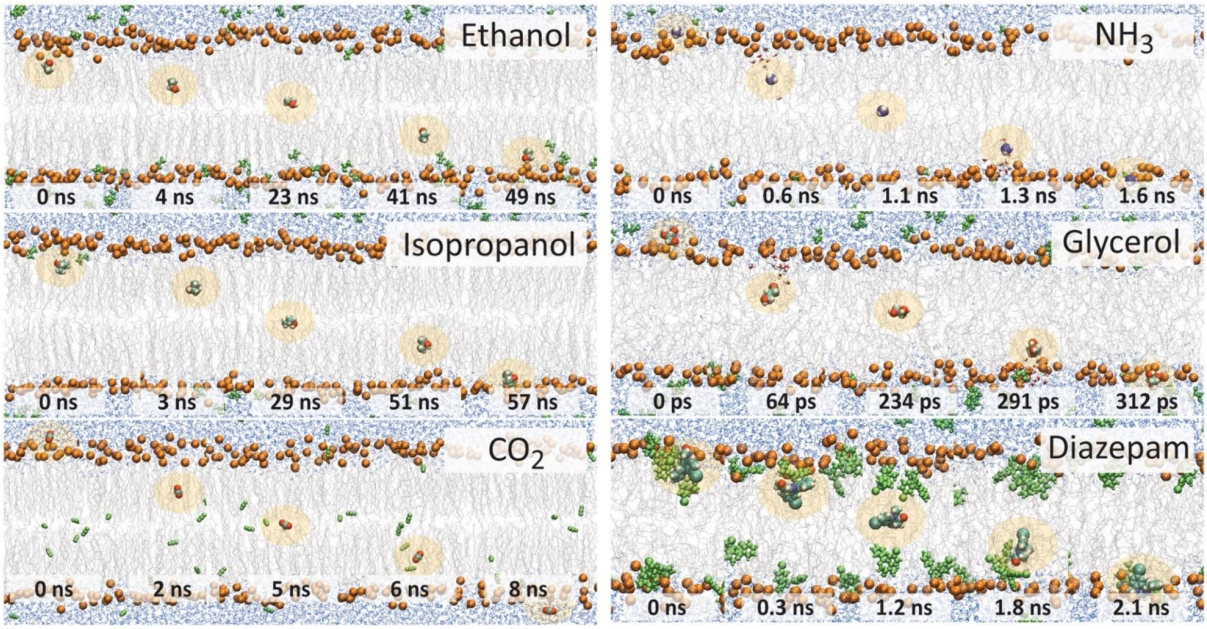


FIGURE 2 | Spontaneous diffusion of various small solutes and drugs through the BBB. Figure used with permission from Wang et al. [106].

However, hundreds of studies since then have attempted to derive superior models and understandings of the predictive factors that underlie BBBp [9].

In light of the extensive development of machine learning (ML) algorithms and computation power over the last decade, there is newfound hope that significant advances in predicting BBBp can be achieved. If sufficiently improved, these models may replace the need for expensive and ethically challenging cellular and animal testing. Moreover, these models may allow for better identification of BBBp molecules in early stages of drug development leading to a more efficient pipeline. In this paper, we conduct a literature review from 2019 to capture the recent developments in the field of predicting BBBp for molecules using ML methods.

Material and Methods

A comprehensive literature review was conducted using PubMed MEDLINE and Google Scholar intended to collect all original machine learning studies regarding BBBp prediction from Jan 01 2019 to December 1 2024. Keywords including "BBB," "ML," and "DL" were used in boolean search strings structured according to the PICO (Population of interest, Intervention, Control, Outcome) structured question of interest (Table 1). Our comprehensive PICO question allowed us to examine research studies using machine learning or deep learning methods to predict the BBBp of compounds or small molecules. Only full papers written in English were included, and only articles that met the PICO criteria were included. There were no specific number or type of compounds or performance/performance metrics needed for inclusion.

Results

14,596 and 7,870 total results were found from PubMed and Google Scholar. Upon reviewing the results 14,541were further excluded from our PubMed search and 7,862 were excluded from the Google Scholar search due to not being applicable to our field of study. After narrowing down the number of research articles to those applicable to our field of study, we further excluded articles due to not having a relation to our specific topic, having an abstract that was out of the scope of our research topic, and being a review article. This left a total of 27 papers for review (Table 2).

Following a comprehensive search of PubMed and Google Scholar, a Scopus search was also conducted. Scopus search yielded 824 search results and from these 819 were excluded due to not being of relevance to the topic or not meeting the inclusion criteria, four of the articles were already included in our compiled list, and one of the articles was a new addition.

Foundations of Machine Learning

ML is a branch of AI that specializes in algorithm and model creation [10]. While recent advancements in algorithms may contribute to improving "learning", algorithms are only as capable as the quality and amount of data available. Supervised ML leverages labeled data, where data and algorithm quality is further contingent on accurate labeling [10]. Unsupervised settings can use unlabeled data, where algorithms attempt to find patterns across each sample to derive some greater grouping structure. The general workflow of ML is demonstrated in Figure 3.

Support Vector Machine (SVM) is a supervised ML algorithm used for classification and regression tasks. It is particularly useful for data that becomes linearly separable when transformed into a higher-dimensional space through the use of kernel functions [11]. Tree methods are often trained to split the data into smaller and more specific subsets. A Decision Tree (DT) consists of nodes and edges. Each node can be one of three types: a root node, a decision node, and a leaf node. These nodes have different contributions to the overall function

TABLE 1 | Systematic Search Results.

Website	Search terms	Results
Google scholar	(BBB OR Blood Brain Barrier OR Brain Barrier) AND (DL OR Deep Learning OR AI OR Artificial Intelligence OR Transformers Or Convolutional Neural Network OR CNN OR Generative Models OR VAE OR AE OR Variational Auto-Encoder OR Autoencoder OR Machine Learning OR ML OR RNN OR Prediction OR Predictability OR Recurrent Neural Network OR LSTM OR Long Short Term Memory OR Convolution) AND (Perm*)	7,870 resulting studies
Pubmed	(BBB OR Blood Brain Barrier OR Perme*) AND (DL OR Learn* OR ML OR Machine Learning OR Random Forest OR RF OR Lapla* OR SVM OR Support Vector Machine OR Algor* OR AI OR Artificial Intelligence OR Transf* Or Convol* OR Neural Network OR CNN OR VAE OR AE OR Encode* OR RNN OR Predict* OR LSTM) NOT (Review[Publication Type])	14,596 resulting studies
Scopus	(BBB OR Blood Brain Barrier OR Brain Barrier) AND (DL OR Deep Learning OR AI OR Artificial Intelligence OR Prediction OR Permeability OR ODE	824 resulting studies

TABLE 2 | Reports Discussed.

Study	Methods used	Dataset	Model statistics	Type of model	Citation
Liu et al.	SVM RF RF XGB	1757 compounds. Validation: 213 Encoded using 9 FPs.	5-fold CV: Acc: 0.820 – 0.918 Validation: Acc (Best Ensemble Model): 0.930	Regression	Liu L. Zhang L. Feng H. Li S. Liu M. Zhao J. et al. Liu, H. (2021). Prediction of the blood–brain barrier (BBB) permeability of chemicals based on machine-learning and ensemble methods. Chemical Research in Toxicology, 34(6), 1456–1467.
Shaker et al.	LightGBM	7162 compounds Encoded using SMILES	CV: Acc: 89% Specificity: 0.77 Sensitivity: 0.93 Specificity: 0.94 (Achieved on external validation data) Sensitivity: 0.85 (Achieved on external validation data) Comment: seems only reporting classification statistics here	Classification/Regression	Shaker B. Yu M. S., Song, J. S., Ahn, S., Ryu, J. Y., Oh, K. S., et al. Na, D. (2021). LightBBB: computational prediction model of blood–brain-barrier penetration based on LightGBM. Bioinformatics, 37(8), 1135–1139.
Shaker et al.	LightGBM gradient-boost RF	913 log BB values (External test) 27 compounds	10-fold CV: MSE: 0.22 R^2: 0.59 External Test: MSE: 0.36 R^2: 0.61 Comment: seems only reporting regression statistics here	Classification/Regression	Shaker B. Lee J. Lee Y. Yu M. S., Lee, H. M., Lee, E., Kang, H. C., Oh, K. S., Kim, H. W., et al. Na, D. (2023). A machine learning-based quantitative model (LogBB Pred) to predict the blood-brain barrier permeability (logBB value) of drug compounds. Bioinformatics (Oxford, England), 39(10), btad577. doi:10.1093/bioinformatics/btad577.
Charoenkwan et al.	SCM Based Predictor	Kumar et al. [48]	Test: MCC: 0.716 Acc: 0.895	Classification	Charoenkwan, P., Chumnanpuen, P., Schaduangrat, N., Lio', P., Moni, M.

TABLE 2 | (Continued)

Study	Methods used	Dataset	Model statistics	Type of model	Citation
					A., & Shoombuatong, W. (2022). Improved prediction and characterization of blood-brain barrier penetrating peptides using estimated propensity scores of dipeptides. Journal of Computer-Aided Molecular Design, 36(11), 781–796.
Sakiyama et al.	RF Classifier	Training: 1565 molecules Validation: 196 molecules Test: 196 molecules	Test: ROC-AUC: 0.767	Classification	SAKIYAMA, H., MOTOKI, R., OKUNO, T., & LIU, J. Q. (2023). Improvement of Blood-Brain Barrier Permeability Prediction Using Cosine Similarity. Journal of Computer Chemistry, Japan-International Edition, 9.
Boulamaane et al.	RF	7,807 molecules FPs	5-fold CV: AUC: 0.97 External Set: Accuracy: 95% accuracy AUC: 0.92	Classification	Boulaamane, Y., & Maurady, A. (2023). EnsembleBBB: Enhanced accuracy in predicting drug blood-brain barrier permeability with a Machine Learning Ensemble model.
Kumar et al.	qRASAR	Training: 1,012 compounds Validation: 1,130,315 compounds split into two sets	R2: 0.634 Q2: 0.627 Rpred2: of 0.697	Classification/Regression	Kumar, V., Banerjee, A., & Roy, K. (2024). Innovative strategies for the quantitative modeling of blood–brain barrier (BBB) permeability: harnessing the power of machine learning-based q-RASAR approach. Molecular Systems Design & Engineering.
Wu et al.	NN-QSAR	260 compounds 70 MD′s	R:0.956	Classification/Regression	Wu, Z., Xian, Z., Ma, W., Liu, Q., Huang, X., Xiong, B., ... & Zhang, W. (2021). Artificial

TABLE 2 | (Continued)

Study	Methods used	Dataset	Model statistics	Type of model	Citation
					neural network approach for predicting blood brain barrier permeability based on a group contribution method. Computer methods and programs in biomedicine, 200, 105943.
Dehnbostel et al.	RF	479 active compounds 286 inactive compounds	RF: F1: 0.95 Specificity: 0.962	Classification	Dehnbostel, F. O., Dixit, V. A., Preissner, R., & Banerjee, P. (2024). Non-animal models for blood–brain barrier permeability evaluation of drug-like compounds. Scientific Reports, 14(1), 8908.
Cornelissen et al.	XGB	9316 compounds FPs	AUC: 0.78–0.92	Regression	Cornelissen, F. M., Markert, G., Deutsch, G., Antonara, M., Faaij, N., Bartelink, I., ... & Westerman, B. A. (2023). Explaining blood–brain barrier permeability of small molecules by integrated analysis of different transport mechanisms. Journal of Medicinal Chemistry, 66(11), 7253–7267.
Kim et al.	3D-QSAR NN	406 compounds	R^2: 0.809–0.963	Regression	Kim, T., You, B. H., Han, S., Shin, H. C., Chung, K. C., & Park, H. (2021). Quantum artificial neural network approach to derive a highly predictive 3D-QSAR model for blood–brain barrier passage. International journal of molecular sciences, 22(20), 10995.
Radchenko et al.	NN	529 compounds	Q2:0.815 (RMSEcv:0.318	Regression	Radchenko E. V., Dyabina, A. S., et al. Palyulin, V. A. (2020). Towards

TABLE 2 | (Continued)

Study	Methods used	Dataset	Model statistics	Type of model	Citation
					deep neural network models for the prediction of the blood–brain barrier permeability for diverse organic compounds. Molecules, 25(24), 5901.
Alsenan et al.	RNN	3606 compounds (1803 BBB + and 1803 BBB−) 3603 features represented by	10-fold CV: Acc: 96.53% AUC: 0.584	Classification	Alsenan, S., Al-Turaiki, I., & Hafez, A. (2020). A recurrent neural network model to predict blood–brain barrier permeability. Computational Biology and Chemistry, 89, 107377.
Atwereboannah et al.	FCNN CNN	1st Dataset: 2051 drugs 2nd Dataset: 213	FCNN: 1st dataset: Acc:99.5% F1: 0.989 AUROC: 0.995 2nd dataset: F1: 0.6544 Acc: 54.05% AUC: 0.584 CNN: 1st dataset: Acc: 89.6% F1: 0.896 AUROC: 0.992 2nd dataset: Acc: 66% F1: 0.667 AUC: 0.584	Classification	Achiaa Atwereboannah, A., Wu, W. P., & Nanor, E. (2021, May). Prediction of Drug Permeability to the Blood-Brain Barrier using Deep Learning. In 4th International Conference on Biometric Engineering and Applications (pp. 104–109).
Ma et al.	NN	Training: 269 BBBp + peptides 2069 BBBp- peptides External Validation: 119 BBBp + 119 BBBp-	Acc: 98.31% Sensitivity:98.15% Specificity:98.32%	Regression	Ma, C., & Wolfinger, R. (2023). A prediction model for blood-brain barrier penetrating peptides based on masked peptide transformers with dynamic routing. Briefings in Bioinformatics, 24(6), bbad399.
Yu et al.	GCN + SVM hybrid model	Training: 940 market drugs 315 BBB + 625 BBB- Validation: 117 drugs	Acc: 0.96 ROC AUC: 0.98 F1: 0.94	Classification/	Yu T. H., Su, B. H., Battalora, L. C., Liu, S., et al. Tseng, Y. J. (2022). Ensemble modeling with machine learning

TABLE 2 | (Continued)

Study	Methods used	Dataset	Model statistics	Type of model	Citation
		42 BBB+ 75 BBB- SMILES Notation			and deep learning to provide interpretable generalized rules for classifying CNS drugs with high prediction power. Briefings in bioinformatics, 23(1), bbab377.
Hamzic et al.	MT-GNN ST	1593 compounds 208 physicochemical descriptors and FPs	R^2: 0.42 MAE: 0.39 MCC: 0.66	Classification/Regression	Hamzic, S., Lewis, R., Desrayaud, S., Soylu, C., Fortunato, M., Gerebtzoff, G., & Rodríguez-Pérez, R. (2022). Predicting in vivo compound brain penetration using multi-task graph neural networks. Journal of chemical information and modeling, 62(13), 3180–3190.
Ajabani et al.	SVM, KNN, Logistic Regression, RF, Multi-Layer Perceptron, and Light Gradient Boosting	1593 compounds	78% to 94% accuracy RF (best performance): Acc: 90.36% AUC: 0.96 Sensitivity: 77.73% Specificity: 94.74%	Classification	Ajabani, D. (2023). A Computational Prediction Model of Blood-Brain Barrier Penetration Based on Machine Learning Approaches.
Mazumdar et al.	XGboost, RF, Extra Tree Classifiers, NN	8153 compounds 4894 BBBp+ 3259 BBBp 8 FPs 2 MDs	NN (best performance): Acc: 97.8% Sensitivity: 97.0%, Specificity: 98% BACC: 98% F1: 98% ROC-AUC: 98%	Classification	Mazumdar, B., Sarma, P. K. D., Mahanta, H. J., & Sastry, G. N. (2023). Machine learning based dynamic consensus model for predicting blood-brain barrier permeability. Computers in Biology and Medicine, 160, 106984.
Miao et al.	Four-layer NN	91 molecules 38 BBB+ 53 BBB- 210 molecules 136 BBB+ 74 BBB- 161 molecules 76 BBB+ 85 BBB-	Acc: 97% AUC: 0.98 F1 score: 0.92	Classification	Miao, R., Xia, L. Y., Chen, H. H., Huang, H. H., & Liang, Y. (2019). Improved classification of blood-brain-barrier drugs using deep learning. Scientific reports, 9(1), 8802.

TABLE 2 | (Continued)

Study	Methods used	Dataset	Model statistics	Type of model	Citation
Kumar et al.	DNN 1-D CNN Fined Tune CNN SVM kNN NB RF	2,607 BBB+ compounds 998 BBB- compounds FPs	DNN: Accuracy: 99.2% Sensitivity: 99.7% Specificity: 99.6% AUC: 98.7% HD: 4.048 1-D CNN: Accuracy: 96.9% Sensitivity: 95.6% Specificity: 97.5% AUC: 98.3% HD: 4.118 FNCNN: Accuracy: 97.2% Sensitivity: 98.3% Specificity: 98.3% AUC: 94.6% HD: 4.581 SVM: Accuracy: 91.6% Sensitivity: 96.9% Specificity: 93.8% AUC: 96.3% HD: 15.242 kNN: Acc: 92.7% Sensitivity: 97.4% Specificity: 94.9% AUC: 96.8% HD: 12.891 NB: Acc: 84.4% Sensitivity: 94.8% Specificity: 89.9% AUC: 93.5% HD: 14.543 RF: Acc: 81.5% Sensitivity: 94.3% Specificity: 88.7% AUC: 93.8% HD: 14.543	Classification	Kumar, R., Sharma, A., Alexiou, A., Bilgrami, A. L., Kamal, M. A., & Ashraf, G. M. (2022). DeePred-BBB: A blood brain barrier permeability prediction model with improved accuracy. Frontiers in neuroscience, 16, 858126.
	MMBART Transformer	3605 compounds 2607 BBB+ 998 BBB–SMILES	DeePred: AUC: 0.93 LightBBB: AUC: 0.96	Classification Comments: they use regression model but then convert the regression results to classification decision. All the results seem to be on classification without given intermediate regression results. I would suggest to remove the regres-	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. (2024). Predicting blood-brain barrier permeability of molecules with a large language model and machine learning. Scientific reports, 14(1),

TABLE 2 | (Continued)

Study	Methods used	Dataset	Model statistics	Type of model	Citation
				sion to keep things simple.	15844. https://doi.org/10.1038/s41598-024-66897-y
Rosa et al.	LIME + RF LIME + RF-ET LIME + DRN	2000 drugs, hormones, and neurotransmitters FPs	DRN: ROC-AUC: 0.894, 0.790, 0.882 (for three runs) Precision: 0.868 to 0.896 Recall: 0.868 to 0.896 F1: 0.868 to 0.896 Acc: 0.868 to 0.896 RF: ROC-AUC: 0.931, 0.943, 0.936 (for three runs) RF-ET slightly outperformed by RF. Similar performance.	Classification	Rosa, L. C. S., Argolo, C. O., Nascimento, C. M. C., & Pimentel, A. S. (2024). Identifying Substructures That Facilitate Compounds to Penetrate the Blood–Brain Barrier via Passive Transport Using Machine Learning Explainer Models. ACS Chemical Neuroscience.
	Deep-B^3	7224 compounds 5483 BBB + 1741 were BBB- SMILES	Sensitivity: 85% Specificity: 64% MCC: 0.49 Accuracy: 0.74 AUC: 0.80	Classification	Tang, Q., Nie, F., Zhao, Q., & Chen, W. (2022). A merged molecular representation deep learning method for blood-brain barrier permeability prediction. Briefings in bioinformatics, 23(5), bbac357. https://doi.org/10.1093/bib/bbac357
Aftab et al.	CNN-LSTM-ODE	LightBBB B3BD	Regression Results: LightBBB RMSE: 0.59; B3BD RMSE: 0.85 Classification Results: LightBBB AUC: 0.96 Sensitivity: 0.92 Specificity: 0.95 B3BD AUC: 0.94 Sensitivity: 0.92 Specificity: 0.91	Classification/Regression	Aftab, N., Masood, F., Ahmad, S., Rahim, S. S., Sanami, S., Shaker, B., & Wei, D. Q. (2024). An Optimized Deep Learning Approach for Blood-Brain Barrier Permeability Prediction with ODE Integration. Informatics in Medicine Unlocked, 101526.
Kato et al.	RF, XGBoost, Histogram	2000 compounds	RF (best performance):	Classification	Kato, R., Zeng, W., Siramshetty, V. B.,

TABLE 2 | (Continued)

Study	Methods used	Dataset	Model statistics	Type of model	Citation
	gradient boosting (HGB), Graph convolutional neural network (GCNN)		BACC: 0.698 ± 0.015 (5-fold CV) BACC: 0.708 (validation)		Williams, J., Kabir, M., Hagen, N., ... & Shah, P. (2023). Development and validation of PAMPA-BBB QSAR model to predict brain penetration potential of novel drug candidates. Frontiers in Pharmacology, 14, 1291246.
Mauri et al.	Classification models based on KNN and a Consensus Model	1915 compounds	Training set Acc: 0.819 Sensitivity: 0.917 Specificity: 0.645 CV BACC: 0.781 Test set sensitivity: 0.760 Test set specificity: 0.916 Acc: 0.827	Classification	Mauri, A., & Bertola, M. (2022). Alvascience: a new software suite for the QSAR workflow applied to the blood–brain barrier permeability. International Journal of Molecular Sciences, 23(21), 12882.
Kumar et al.	c-RASAR	7,807 compounds, 4956 (BBBp+) and 2851 and (BBBp−)	Acc: 0.781 Precision: 0.833 Recall: 0.819 F1: 0.826 MCC: 0.53 AUC: 0.84	Classification/Regression	Kumar, V., Banerjee, A., & Roy, K. (2024). Breaking the Barriers: Machine-Learning-Based c-RASAR Approach for Accurate Blood–Brain Barrier Permeability Prediction. Journal of Chemical Information and Modeling, 64(10), 4298–4309.

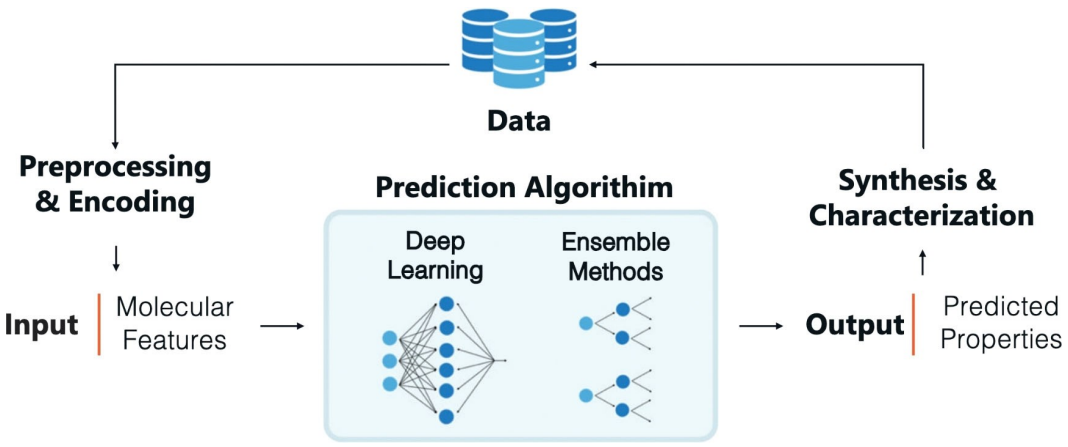


FIGURE 3 | General overview of Machine Learning workflow for the prediction of quality [107].

of a tree and are connected together by edges. Edges act as the branches of a DT, representing outcomes based on a particular feature. Often, individual models are used in ensembles. This describes a type of model that is a combination of multiple models. An ensemble method usually aims to combine individually weaker predictors into larger predictors.

Random Forest (RF) is one of the most widely-used ML methods. It serves as a learning method for both classification and regression. RF belongs to the ensemble approach and, by constructing multiple decision trees, works by averaging the predictions (for regression) or taking the majority vote (for classification) from multiple decision trees [12]. Another notable ML method is Extreme Gradient Boosting (XGBoost), an application of gradient boosting. Boosters are a family of algorithms that create a strong learner by combining multiple weaker learners; furthermore, they are helpful in fighting overfitting, handle missing values well, and support built-in cross-validation. These usually involve learning trees, which are formed sequentially, allowing each tree to correct the mistakes made by the last [13]. Another type of method is Extra Trees (EXT), an ML method that belongs to the ensemble family. It shares similarities with Random Forest (RF), being originally introduced as an extension of the latter. As an ensemble learner, it can build multiple decision trees during training and combine their predictions to improve generalization. Unlike RF, which selects the best split from a random subset of features, EXT randomly chooses splits for each node, which can sometimes lead to even further variance reduction at the expense of slightly increased bias [14].

Deep Learning (DL) is a branch of ML that specializes in learning representations of data via multi-layered neural networks (NNs) [15]. NNs consist of layers of interconnected neurons, with each connection having an associated weight. The input layer, which is composed of neurons or nodes, takes the inputted data and represents each attribute of the data with a neuron before sending it to the hidden layers. The hidden layers of an artificial neural network are a series of layers that process the data, allowing the network to learn complex patterns [15]. A Graph Neural Network (GNN) is a type of neural network designed to process data in the form of graphs, a series of nodes and edges that the model uses to pass information by aggregating it through the nodes. Each node in a GNN can represent an entity, and edges represent relationships between these entities [16]. On the other hand, a CNN is a type of neural network that consists of convolutional layers, pooling layers, and a dense NN component that produces an output[97]. CNNs specialize in processing spatially sorted data, a common example being two- or three-dimensional structures. The convolutional layers extract features from the input data and feed them into the pooling layer, which reduces spatial dimensions. This is then finally transferred into a fully connected layer that produces an output similar to a traditional dense NN [17].

There are other variants, including a Recurrent Neural Network (RNN), which is designed to handle sequential data by maintaining an internal state to process sequences of inputs [18]. More advanced RNNs, such as Long Short-Term Memory (LSTM) models, specialize in using gating mechanisms to retain longer sequences without depreciation, an issue that arises in most simple RNNs [19]. Currently state of the art for sequence data, transformers depend on “self-attention”, a mechanism that makes each word from the input sequence attend to all other positions to compute a representation of that token/position, allowing the model to capture the weight of each token in the context of the entire input [20]. Molecules can be characterized based on their 1D-, 2D-, 3D-structures and other molecular fingerprints as mentioned below. These characteristics can then be leveraged using methods that handle sequential data, such as RNNs and transformers, to predict the output of interest (e. g. BBBp).

Molecular Representations

To note, many models in this field are termed Quantitative Structure-Activity Relationship (QSAR) models. Such a model links chemical structure to its functional property. Oftentimes, they represent sophisticated regressions, but more recently, they have been entailing ML and DL techniques. There are many ways of representing the molecular structure (MS) of a compound in chemoinformatics, including but not limited to Molecular Descriptors (MDs), Fingerprints (FPs), and Simplified Molecular Input Line Entry System (SMILES) text notation. These three are the most prevalently used descriptors in the articles reviewed. These specific representations serve to represent certain molecular features of a compound for computational efficiency.

An MD is a numerical representation of a molecule′s chemical properties across multiple possible dimensions [21]. A 1D MD may represent simple characteristics, like a molecule′s weight and number of atoms. A 2D MD represents a molecule′s structural features and connectivity. A 3D MD represents the geometric and spatial features of a compound that considers the 3D aspects of its molecular structure. Resembling aspects of physics, a 4D MD includes aspects that change over time in a molecule [21].

An FP is a numerical representation of a molecule that shows the structural features of a molecule in a binary or bit-string format. FPs encode information about the presence or absence of certain structural fragments or molecular substructures within a molecule [22]. The main difference between FPs and generic MDs lies in the structure and method of description. Moreover, an FP provides qualitative information about a molecule′s structural features, unlike a lower-dimensional MD [22].

The Simplified Molecular Input Line Entry System (SMILES) is the most commonly used descriptor in the studies reviewed. SMILES, like the other descriptors, is a textual representation of a chemical structure [23]. It serves as a compact and standardized way to encode and communicate the structure of molecules in a machine-readable format, such as those used in BBBp prediction. SMILES provides additional information, as compared to MDs and FPs, including connectivity and stereochemistry [23]. This may be the reason why SMILES is being used more often than MDs or FPs. Of course, the use of MDs

or FPs may be chosen over SMILES depending on the nature of the specific task.

Discussion

Machine Learning Methods

In recent years, while deep learning (DL) methods have gained popularity, machine learning (ML) methods have typically demonstrated superior performance in tabular data contexts [24]. Their computational efficiency has also encouraged the exploration of multiple ML algorithms in various studies. For example, Liu et al utilized three models – SVM, RF, and RF XGBoost – to develop an ensemble algorithm [25]. They compiled a dataset of 1757 compounds (1276 BBB+, 481 BBB−) from Wang et al. [33]. and Zhao et al. [70]. and validated it using an external set of 213 compounds from Li et al. [26]. Their feature selection process involved removing features with low variance, such as carbon or nitrogen, and filtering out highly correlated features. In particular, they thresholded on Tanimoto correlation coefficient to identify compounds with highly correlated features. (e.g. optimal threshold was found to be 0.95). They built models with SVM, RF, and XGBoost, evaluated through 5-fold cross-validation with 100 repeats, showing accuracy ranging from 0.820 (SVM with APC2D FP) to 0.918 (RF with PubChem FP). Twenty-seven base classifiers were then created with the three algorithms and nine molecular fingerprints, forming ensemble models. The best ensemble model (Ensemble Top-9) achieved an AUC of 0.966 and accuracy of 0.930, significantly outperforming individual classifiers. This ensemble approach improved predictive performance and stability, with the model showing balanced sensitivity and specificity. However, overfitting and data diversity remained challenges, with lower accuracy in their validation set which included 213 compounds.

Shaker *et al* utilize a Light Gradient Boosting Model (LightGBM), termed LightBBB, to predict BBBp [27]. LightGBM is an implementation of an ensemble learning RF that grows trees leaf-wise. Their composite dataset consisted of 7162 compounds [28–35] encoded in SMILES. Overall, the model achieved an accuracy of 89% in cross-validation and 90% on the external test dataset. In 10-fold cross-validation, a specificity of 0.77 and sensitivity was 0.93 was achieved where in external validation data, a specificity and sensitivity of 0.94 and 0.85 was conferred. Compared to prior literature their model failed to achieve state of the art performance.

In a later study, Shaker *et al* also utilized gradient boosting RF with the LightGBM algorithm optimized using GridSearchCV [36]. This algorithm was tested off of a dataset of 913 logBB values that ranged from −2.69–1.7 [37–43]. There was also an external dataset of 27 compounds [44]. In the test set, the model achieved average mean squared error (MSE) of 0.22 and an R^2 score of 0.59 during a 10 fold cross-validation. This led to a final, optimized score of an R^2 of 0.61 and an MSE of 0.36, a better score when compared to other common models like ADMET Prediction Service [45] and PreADMET [46]. Note that while one instance of LightBBB used a dataset of over 7,000 compounds, the follow-up study was trained on only 913 compounds with logBB values. Charoenkwan *et al* developed a scoring-card method (SCM) based predictor to predict and characterize BBBp peptides (B3PPs) called SCMB3PP utilizing three datasets derived from Kumar et al. [47, 48]. Specifically, the model estimates amino acid and dipeptide tendencies to characterize permeability of the aforementioned B3PPs. The model works off of a scorecard algorithm calculating an initial propensity score by subtracting the score of all B3PPs from all non-B3PPs. These optimized scores are then used to create an SCM model (which in this case is SCMB3PP). Finally, the SCM is used to predict the permeability of unknown peptides. The average Matthew's Correlation Coefficient (MCC) for the independent test set across the three datasets was approximately 0.716, and the average accuracy was 0.895.

Sakiyama et al. examined the impact of employing training data that closely resembles the test data to improve the performance of machine learning models in predicting BBB permeability [49]. Their findings demonstrated that selecting training data with high cosine similarity (i.e., a vector-based method to quantify the similarity between data points) to the test data significantly enhanced their RF's prediction performance, even with a smaller training dataset [50]. The top-performing model in the study also exhibited superior performance on two external test sets, indicating robust generalization capabilities and surpassing existing leading models. This approach, leveraging cosine similarity, shows promise for accurately predicting the properties of compounds by balancing data from highly diverse and limited datasets. Boulaamane et al. analyzed a dataset of 7,807 from B3DB compounds [with various molecular binary FPs to find that their RF algorithm using MACCS FPs yielded the best results out of all used descriptors [51, 52]. This combination achieved a mean AUC of 0.94 in 5-fold cross-validation. The applicability domain assessment indicated that the MACCS dataset had the fewest outliers. The model also achieved a 95% accuracy and an AUC of 0.92 on an external benchmarking dataset, underscoring its efficacy for early-stage screening of compounds for BBB permeability.

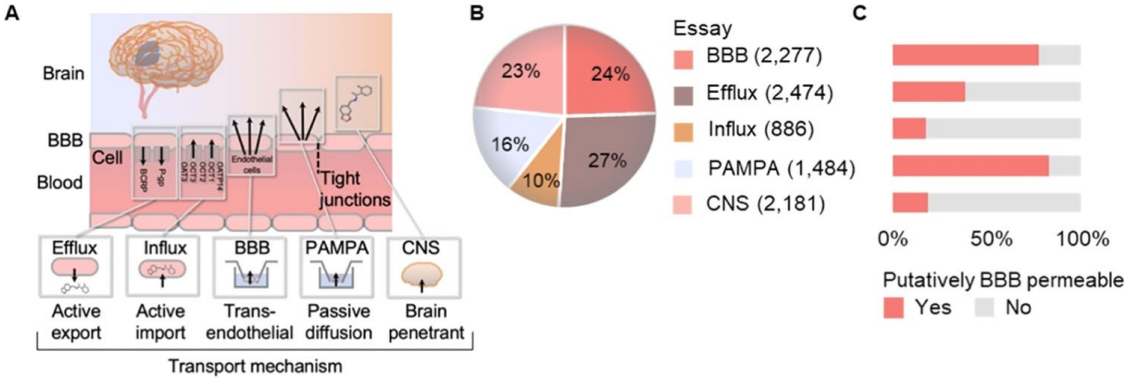
Kumar et al. introduced the quantitative read-across structure–activity relationship (q-RASAR) framework to enhance the precision of BBBp predictions [53]. Unlike traditional models, this study focused on improving the quantitative prediction of BBB permeability for organic compounds. Researchers developed a q-RASAR PLS regression model using a dataset of 1,012 diverse heterocyclic and aromatic compounds from the freely accessible B3DB database. The model's predictive capability was validated with two external sets, totaling 1,130,315 compounds, including synthetic compounds and natural products, and two additional external sets comprising 116 drug-like/drug compounds from the FDA and ChEMBL databases. The study highlighted the importance of hydrophobicity, electronic effects, degree of ionization, and steric factors in facilitating BBB traversal. Their q-RASAR achieved an R^2 of 0.634, a Q^2 of 0.627, and a R_{pred}^2 of 0.697 on their external sets; it was less accurate compared to Wu et al.'s neural network (NN) [54]. However, direct comparison requires caution because Wu et al.'s model was developed using a smaller, more limited dataset. This discrepancy reflects the limitations of descriptor-based QSAR models, which reduce a molecule's

three-dimensional information, such as electrostatics and conformation, into numerical parameters like TPSA, logP, and molecular weight. Although computationally efficient, these descriptors may overlook steric interactions, hydrogen-bond directionality, and non-linear relationships. In contrast, neural networks process structural features more comprehensively by leveraging three-dimensional data or graph-based representations. Descriptor-based approaches often rely on manually engineered features, which can exclude important molecular properties and result in lower performance compared to neural networks that extract features automatically.

In the Dehnbostel *et al* study, 63 distinct models were trained using SVM and RF classifiers, incorporating various combinations of datasets, descriptors, encoding methods, and kernel functions on data from both Muehlbacher et al. and an in-house dataset [55, 56]. Notably, the bbbPythoN-imb RF model achieved F1-scores around and above 0.95 in both internal and external validations. The bbbPythoN-bal RF model reached a specificity of 0.962 with 479 active and 286 inactive compounds in the test set highlighting the challenges of imbalanced datasets. Key descriptors like topological polar surface area only considering nitrogen and (oxygen TopoPSA(NO)) achieved the highest Mutual Information Differential Shannon Entropy (MI–DSE) score, crucial for distinguishing BBB-permeating compounds, although its MI–DSE score dropped from around 0.6–0.3 when positive samples were reduced. Additionally, descriptors like centered Moreau-Broto autocorrelation, with lag 1 and weighted by Sanderson electronegativity (ATSC1se) showed that active compounds conserved around 0.0, indicating low electronegativity disparity, essential for passive membrane crossing.

One of the chief advantages of machine learning (ML) over deep learning (DL) is its explainability, which is crucial in predicting blood-brain barrier permeability (BBBP) as demonstrated in Figure 4. Cornelissen et al. enhance this explainability by analyzing the predictive importance of various molecular features related to BBBP transport [57]. They use XGBoost models to construct multiple models on various sets of molecules that exclusively use one form of transport from a larger FP dataset [33] composed of 4 separate datasets [30, 56, 58, 59] and self-curated data, totalling 9316 compounds. Efflux models for ATP-binding cassette (ABC) transporters, influx models with solute carriers (SLCs) and PAMPA models for diffusion were developed.

The XGBoost models developed in this study scored 72% or higher, with an AUC ranging from 0.78–0.92 on the training dataset, and achieved an accuracy of 71% or higher on validation sets. Efflux models performed comparatively worse than influx models, with accuracies of 71% and 89%. Notably,



D

Dataset/Essay	Target(s)	Putative mechanism	Distribution outcome (measured transport effect)	Sources
BBB	BBB Membrane	BBB permeability	BBB permeable: 1,769 BBB impermeable: 508	Wang *et al*. 2018
Efflux	ABC Transporters	Efflux from brain	ABC substrate: 1,515 ABC non-substrate: 959	Metrabase (Mak *et al*., 2015), ChEMBL (Mendez *et al*., 2019)
Influx	SLC Transporters	Influx to brain	SLC substrate: 157 SLC non-substrate: 729	Metrabase (Mak *et al*., 2015), ChEMBL (Mendez *et al*., 2019)
PAMPA	Artificial Membranes	Passive membrane permeability	PAMPA permeable: 1,235 PAMPA impermeable: 249	ChEMBL (Mendez *et al*., 2019)
CNS (Benchmark)	CNS Drug	BBB permeability	CNS drug: 411 Non-CNS drug: 1,770	Drugbank (Wishart *et al*., 2018)

FIGURE 4 | The figure above illustrates the following: (A) A diagram depicting various active and passive transport mechanisms through which compounds can cross the blood-brain barrier (BBBp). (B) A pie chart comparing the sizes of different datasets by the number of compounds they include. (C) A bar chart displaying the distribution of compounds categorized as likely to be BBBp (in red) versus those not likely to be BBBp (in gray). (D) A table summarizing the datasets used for generating prediction models, including their sources and the specific transport mechanisms assessed through in vitro measurements (e.g., BBB, Efflux, Influx, PAMPA) or clinical data (e.g., CNS). Figure was reproduced with permission from Cornelissen et al. [57].

Topological Polar Surface Area (TPSA), a polarity descriptor, was found to be crucial for both influx and efflux, with lower TPSA thresholds significant for BBB permeation. The difference in influx and efflux model performance can be attributed to the inherent function of each models' transporters and the selectivity of each transporter. For example, the efflux models that performed worse utilize an ABC transporter that is less selective than SLC transporters used by influx models. Therefore, the less selective transporters of efflux models are more likely to have variability in terms of predicting BBB permeability which makes them perform worse when compared to influx models. Molecular weight also emerged as a key predictor for efflux transporters, while the number of hydrogen bond donors was important for influx transporters. The study also highlighted the importance of logD, another measure of lipophilicity, for the PAMPA datasets. Additionally, specific molecular substructures such as MACCS8 and MACCS43 were found to influence BBB permeability, with MACCS8 associated with BBB-impermeable compounds and MACCS43 significant for the influx dataset.

Deep Learning Methods

Kim et al developed a deep learning-based 3D-QSAR NN model, AlphaQ, which, unlike traditional 2D QSAR models, accounts for the three-dimensional intra-molecular spatial arrangements [60]. Models incorporating electrostatic potential (ESP) data represent an extension of traditional QSAR approaches by capturing three-dimensional charge distributions that influence hydrogen bonding and ionic interactions at the BBB. Unlike simpler descriptors, which may fail to account for localized charge gradients, ESP-based models can provide a more detailed representation of molecular interactions, improving predictions of transporter activity and permeability. Kim et al specifically utilize three-dimensional distribution of ESP. Their dataset was composed of 406 organic compounds that were derived from two sources [61, 62]. This dataset ranged from 200–600 atomic mass units (amu) and was split into 8 separate subsets based on their amu. Within these subsets, AlphaQ achieved an R^2 between 0.809–0.963 with no clear relationship between amu range and performance. This performance was validated using an external 5-fold cross validation and outperformed the performance of classic 2-D QSAR models. This study was however limited to using ESPs features and did not account for the presence of other descriptors (e.g. MDs, FPs, SMILES, etc.).

Radchenko *et al* created an NN using fragmental descriptors and to predict logBB [63]. The model uses a single layer architecture with 200 fragmental descriptors. Their dataset of 529 compounds, derived from over 100 source publications. The model was then externally validated using a set of 568 compounds [64]. The model scored a Q2 (compounds with error greater than 1–1.5 log units) of 0.815 and a Root Mean Square Error of Cross-Validation (RMSEcv) of 0.318. While the strength of this paper is in their use of primary literature and true external validation dataset as well as their use of logBB as opposed to a binary outcome, this study fails to compare well to other methods due to little overlap in databases. Similarly, Wu *et al* presented a NN QSAR approach[54]. Specifically, the QSAR was used to study the relationship between structural characteristics of compounds and their biological properties. Subsequently, the NN was used to build logBB prediction models from the predicted biological activity across 18 variables including TPSA and logP. Thus, the study used a total of 70 MDs, 52 being structure group descriptors, to create 18 physicochemical parameters. The model was trained off of a subset of 260 compounds, a further 40 being used for testing. The optimal NN model achieved an R score of 0.956 in the test set.

Alsenan *et al* were the first group to apply RNNs to the prediction of BBBp [65]. The group augments an imbalanced data set composed of 2350 compounds,(1803 BBB+, 547 BBB−) to 3606 compounds to reduce the imbalance (1803 BBB + and 1803 BBB−) [33]. Each molecule retained 6394 features composed of both FPs and MDs as well as SMILES representations. A kernel-PCA (non-linear dimensionality reduction) was used to reduce dimensionality from 6934–3603. The Synthetic Minority Oversampling Technique (SMOTE) resampling technique is used to further mitigate the class imbalance. While SMOTE introduces synthetic noise, this resampling method proved effective here. On 10-fold-cross-validation, the RNN model achieved an average accuracy of 96.53% along with a specificity of 98.09%.

Atwereboannah *et al* [66] developed a Fully Connected Neural Network (FCNN) and a CNN using SMILES notations from the MoleculeNet Repository [34] (2051 drugs) and Gao *et al*'s (213 drugs) [29]. In the MoleculeNet Repository, the FCNN achieved an accuracy of 99.5% and an F1 score of 0.989. The CNN achieved a lower score with an accuracy of 89.6% and a similar F1 score of 0.896. The FCNN and CNN achieved an AUROC of 0.995 and 0.992 respectively. In the second dataset, the FCNN scored an F1 of 0.6544, an accuracy of 54.05%, and an AUC of 0.584. The CNN achieved an F1 of 0.667, an accuracy of 66%, and an AUC of 0.584. Compared to SVM models from Gao *et al*, the CNN outperformed in only the second dataset [29]. Again, 529 training compounds and 568 external compounds remain relatively modest sample sizes compared to certain large-scale BBB databases.

Ma et al. develop a deep learning model to predict BBBp called DeepB3P3 [67]. Their hybrid model attempts to combine the strengths of transformer encoders, and capsule networks (CapsNet) layers to enhance performance. The authors use a transformer that attempts to better capture hierarchical and spatial relationships [68]. The model was trained on a dataset composed of compounds pulled from B3Pdb and the Swiss-Prot database [69, 70] that consisted of 269 BBBp+ peptides and 2069 BBBp- peptides. There was a smaller dataset of 119 BBB+ and 119 BBB- compounds pulled from Dai et al. that was also used as external validation [71]. The first two datasets were used for training and further evaluation, leading to the model scoring an accuracy of 98.31%, a sensitivity of 98.15%, and a specificity of 98.32% in the validation set [71]. The next highest performing model was He *et al*'s Mutual Information Maximization Meta-Learning (MIMML) [72], a meta-learning neural network, scoring an accuracy of 95.67%, a sensitivity of 64.08%, a specificity of 98.32%, AUC of 0.994, and an MCC of 0.997.

As illustrated above, appropriately developed and applied DL methods are highly accurate when it comes to predictability but remain challenging to analyze than most purely ML methods. Yu *et al* attempted to explore ML-level explainability in DL models [73]. The main dataset included 940 market drugs, of which 315 were BBB+ and 625 were BBB- [74]. Additionally, an external dataset of 117 market drugs was used for validation, with 42 BBB+ and 75 BBB- drugs [75]. A Graph Convolutional Network (GCN) was used to process molecular graphs and predict CNS activity probabilities (Prob_pred) from SMILES notations, leveraging its capability to handle complex data and automatically learn features. To enhance performance and interpretability, a Hybrid Ensemble Model was developed by integrating the GCN′s predictions with an SVM model. Molecular descriptors were calculated using PaDEL-Descriptor software, initially including 17,473 descriptors, which were then refined based on importance scores from various methods such as the Gini index for DT and RF, and coefficients for SVM, ensuring the most relevant features were selected for training.

The hybrid model achieved high performance metrics, including an accuracy of 0.96, ROC AUC of 0.98, and an F1 score of 0.94, and offered valuable insights for designing CNS drugs and improving predictive models. The authors further utilized these results to develop and propose a set of guidelines that can be used to distinguish between CNS- and non-CNS-active compounds. Furthermore, these results demonstrated that 75% of CNS drugs adhere to the authors new proposed guidelines while only 21% non-CNS drugs follow the guidelines. This ensemble model outperformed conventional ML and DL models with accuracy increases of 4% on training and validation sets and 8% on external testing. Using a fingerprint-splitting method for robust validation and effective hyperparameter tuning, the model achieved an accuracy of 0.85. They also identified eight key molecular descriptors for CNS drugs, including features such as nitrogen heterocycles and specific hydrogen bonding properties. The proposed classification rules, based on these descriptors, were superior to traditional methods like Lipinski′s Rule of Five [73].

Another hybrid model presented by Aftab et al. integrated three models of Ordinary Differential Equations (ODEs), RNNs, and CNNs. This model is capable of forecasting the logBB value of the compound in question for lightBBB and B3BD datasets. The output of the CNN model was further fed into Long Short-Term Memory (LSTM) network. The results demonstrated that the hybrid model of CNN-LSTM achieved an RMSE of 0.59 utilizing the LightBBB dataset and an RMSE of 0.85 with the B3DB dataset [76]. Kumar *et al.* created several models, mainly focusing on the classification read-across structure−activity relationship (c-RASAR) framework paired with an LDA model [77]. The study utilized a dataset of 7,807 compounds, including both 4956 (BBBp+) and 2851 and (BBBp−) pulled from the B3DB database. The LDA-RASAR model was analyzed in the test set, scoring an accuracy of 0.781, a precision of 0.833, a recall of 0.819, an F1 score of 0.826, and MCC of 0.53, and an AUC of 0.84.These results and others previously mentioned demonstrate the powerful ability of hybrid models when it comes to rapid assessment of compounds and informed decision making when it comes to the BBB permeability and drug development.

Kato *et al.* developed a QSAR model based off of PAMPA-BBB data. This was developed using a dataset of around 2,000 compounds which represented over 60 small molecule drug discovery projects in the National Center for Advancing Translational Sciences (NCATS). In the study there were four different predictive models applied, those being RF, XGBoost, Histogram Gradient Boosting (HGB), and Graph Convolutional Neural Networks (GCNN). Notably, the RF model, the highest 5-fold cross-validation balanced accuracy (BACC) of 0.698 ± 0.015 and validation set BACC of 0.708 were achieved using RDKit descriptors, outperforming MOE descriptors which achieved a 0.688 ± 0.019 5-fold cross-validation BACC and a validation set BACC of 0.659. In contrast, the GCNN model achieved a slightly lower 5-fold CV BACC of 0.683 ± 0.018 with RDKit descriptors but exhibited superior validation set performance (BACC = 0.723) [78]. These results further demonstrate the strengths and weaknesses of various models and how a combination of various models can be utilized to further develop more stringent models and methods. For example, Mauri *et al.* introduces and applies the Alvascience software suite, designed to facilitate all of the QSAR and Quantitative Structure-Property Relationship (QSPR) workflow, an essential component of predicting molecular endpoints for untested compounds. It does this by housing four different software, each covering a step of the QSAR process, that make up the suite but can also be used separately as their own software [79]. Models such as this one can be broken down and each software can be utilized to be implemented in other more models or serve as a source of validation for other models.

Hamzic et al. [80] developed learning models to predict in vivo compound logP from chemical structures using an internal Novartis dataset with in vivo and in vitro variables [81]. The group developed two types of models including multi-task graph neural network (MT-GNN) and single-task (ST) models. While both models utilized 208 physicochemical descriptors along with FPs. to predict logP, the MT-GNN was also trained to complete auxiliary tasks based on in vitro chemical data including LogD in octanol buffer, passive permeability, etc. MT-GNN models improved prediction accuracy compared to training solely on in vivo data. The best-performing MT-GNN model achieved a coefficient of determination of 0.42 and a mean absolute error of 0.39 on a prospective validation set, outperforming all tested ST models [52]. Additionally, the classification of compounds into brain-penetrant or non-penetrant categories achieved a MCC of 0.66. These were compared with K-Nearest Neighbors (KNN) and RF models which were largely outperformed (KNN: MCC = 0.02, RF: $R^2 = 0.39$, MAE = 0.44, Spearman′s Correlation Coefficient = 0.63).

Comparative Studies

Many studies have attempted to compare the relative performance of multiple models across the ML and DL spectrum on BBBp prediction tasks. For instance, Ajabani et al. [82] utilized a combination of two BBB datasets [83, 84] with an array of ML methods including SVM, KNN, Logistic Regression, RF,

Multi-Layer Perceptron, and Light Gradient Boosting with 10-fold cross-validation. Each approach ranged from 78% to 94% accuracy with RF models showing exceptional performance. The RF model attained an overall accuracy of 90.36%, AUC of 0.96, a sensitivity of 77.73%, and a specificity of 94.74% with similar performance in the external validation dataset.

Mazumdar *et al* also conducted their study comparing XGboost, Random Forest, Extra Tree Classifiers, and NN [85]. These models used datasets derived from the DrugBank database which contained a total of 8153 compounds (4894 BBB permeable, 3259 non-permeable), and were evaluated using a 5-fold cross-validation. The 8153 compounds were encoded using 10 separate molecular descriptors and totaled 4 different dataset, 3 balanced with one imbalanced. The descriptors were 8 FPs and 2 MDs, all being tested using the three methods and a resampling technique. The models utilized a dynamic consensus approach to achieve agreements between multiple models, which results in a more "agreed upon" conclusion between the models. On final evaluation, the NN significantly outperformed all other techniques with respect to by achieving an accuracy of 97.8%, a sensitivity of 97.0%, a specificity of 98%, a balance accuracy of 98%, an F1 score of 98%, and an ROC-AUC of 98% [85].

Miao *et al* created a four-layer NN and compared the proposed model with NN, SVM, KNN, and DT algorithms [86]. They used the clinical SIDER database and derived three separate datasets to train and evaluate the model. They were composed of 91 molecules (38 BBB+, 53 BBB−), 210 molecules (136 BBB+, 74 BBB−), and 161 molecules (76 BBB+, 85 BBB−) [83]. Their NN produced an average accuracy of 97% with an AUC of 0.98, and an F1 score of 0.92 across all datasets. When compared to other ML methods separately, there was a gain in accuracy from 14–45 percent. The DL method consistently outperformed the other methods, showing significant improvements in accuracy, AUC, and F1 score – up to 44% better than the next best methods. Despite its advantages, the DL method does not predict the mechanism of BBB penetration. We note that while these two studies contradict Ajabani et al, the NN used by Ajabani is limited in size and depth compared to the NN used in the latter two comparative studies.

Another study compared separate ML approaches composed of SVM, KNN, Naïve Bayes (NB), and RF to three DL models including a deep NN DeePred-BBB, a one dimension CNN (CNN-1D) and a fine-tuned CNN (i.e. CNN-VGG16) [88]. The FPs dataset included 2,607 BBB+ compounds and 998 BBB- compounds that were pulled from Zhao *et al* [70], Shen *et al* [58], and Roy *et al.*[89] [85]. The DeePred-BBB model achieved the highest performance, with 99.2% accuracy, 99.7% sensitivity, and 99.6% specificity, as well as an AUC of 98.7% and the lowest Hamming Distance (4.048). The CNN-1D model followed with 96.9% accuracy, 95.6% sensitivity, 97.5% specificity, an AUC of 98.3%, and an HD of 4.118. The fine-tuned CNN model also performed well, showing 97.2% accuracy, 98.3% sensitivity, and 98.3% specificity, with an AUC of 94.6% and an HD of 4.581. Among traditional models, the SVM with a linear kernel had 91.6% accuracy, 96.9% sensitivity, 93.8% specificity, an AUC of 96.3%, and an HD of 15.242. The kNN model (k = 3) achieved 92.7% accuracy, 97.4% sensitivity, 94.9% specificity, an AUC of 96.8%, and an HD of 12.891. The NB model had 84.4% accuracy, 94.8% sensitivity, 89.9% specificity, an AUC of 93.5%, and an HD of 14.543, while the RF model had 81.5% accuracy, 94.3% sensitivity, 88.7% specificity, an AUC of 93.8%, and an HD of 14.543.

Leveraging the transformer architecture, Huang et al. develop a MegaMolBART (MMBART) as a transformer encoder for SMILES [90]. The transformer encoded SMILES were then compared to traditional FPs in an XGBoost algorithm for final logBB prediction. The model was trained off of a dataset that derived compounds from LightBBB, DeePred-BBB, B3DB, and the Natural Products Research Laboratory (NPRL) [27,52,88,91][87,100,101]. MMBART outperformed traditional Morgan FPs, showing superior accuracy and AUC scores. Specifically, MMBART achieved an AUC of 0.93 on the LightBBB dataset and 0.96 on the DeePred dataset, comparable to or better than traditional models like LightGBM and NNs. In vitro validation using LC–MS/MS confirmed the model′s predictions, with BBB-permeable compounds accurately identified and BBB-impermeable compounds correctly excluded. The model′s ability to handle complex patterns from SMILES representations are perhaps more reliable than extensive physicochemical property calculations, however, the authors did note an overfitting as a possible reason for inflated performance. Figure 5

Rosa et al. utilized the local interpretable model-agnostic explanation (LIME) method to explain all individual predictions [92]. The LIME algorithm slightly alters data, getting predictions from the model then fitting a simpler regression to the results. The models, RF, RF-Extra Trees (RF-ET) and a Deep Residual Network (DRN), were trained from the Molecule Net dataset [34]. The dataset consisted of over 2000 drugs, hormones, and neurotransmitters which were encoded in FPs. DRN conferred ROC-AUC scores of 0.894, 0.790, and 0.882 for its three runs, with precision, recall, F1 score, and accuracy values ranging from 0.868–0.896. RF outperformed with ROC-AUC scores of 0.931, 0.943, and 0.936 with RF-ET performing slightly worse. All models exhibited high true positive values and relatively lower true negative values, indicating effective identification of BBB penetrating compounds, though they showed less accuracy in identifying nonpenetrating compounds. Additionally, LIME analysis identified key substructures, notably nitrogen-containing groups and aromatic rings, that contribute to BBB penetration, with consistent findings across DRN, RF, and RF-ET models.

In another attempt to leverage DL methods, Tang et al created the Deep B^3 algorithm [93]. This multi-model algorithm extracts molecular graph features using a modified ResNet-50 CNN, and SMILES strings features with a Long Short Term Memory (LSTM). These features are then concatenated with MDs and FPs, where a self-attention layer is applied to highlight key features related to BBB permeability within the final NN. The dataset used for training consists of 7224 compounds [25, 27–31]. Within these, there were 5483 being BBB+, while 1741 were BBB-. In the testing dataset, sourced from previous publications and databases, 2670 molecules (1258 BBB+ and 1412 BBB−) were included [2, 25]. To address dataset imbalance, augmentation was performed by renumber-

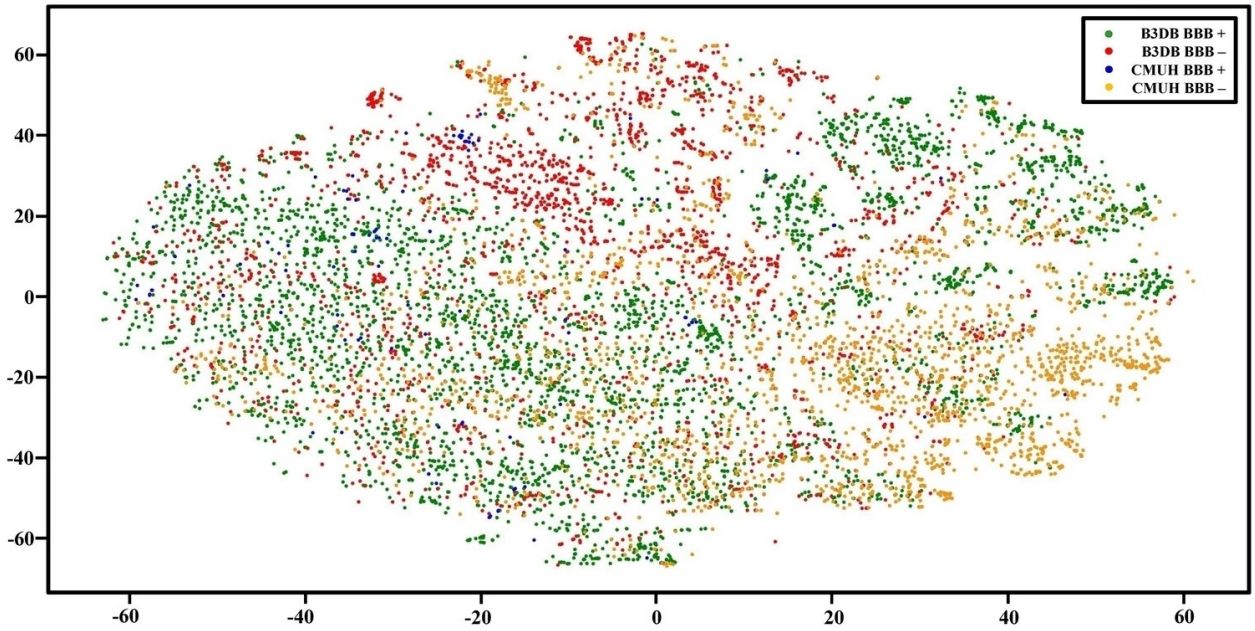


FIGURE 5 | The distribution of molecule embeddings using t-distributed stochastic neighbor embedding (T-SNE), with color coding indicating different datasets and BBB+ or BBB− classifications. Figure was reproduced with permission from Huang et al. [90].

ing atoms. The study's final results were a 85% sensitivity, 64% specificity, 0.49 MCC, 0.74 accuracy, and a 0.80 AUC for the post-SMILES-augmented model. The authors note superior performance compared to the Light Gradient Boosting-based LightBBB where there was increased MCC and ACC by 6% and 4% respectively [27]. LightBBB was also outperformed by Deep B^3 outperformed in sensitivity by 9%.

In reviewing BBB permeability prediction models, both ML and DL methods have shown substantial promise, with each approach offering different advantages that depend on data availability, descriptor choices, and target compound diversity. Several studies [25,27,51][94] highlight ensemble ML methods, attaining accuracy and AUC scores above 0.90 when using an array of molecular descriptors and fingerprints, underscoring the benefits of combining diverse weak predictors. Meanwhile, Shaker et al. [27,36] confirm the efficiency of gradient boosting algorithms for handling large, heterogeneous datasets, although overfitting and data imbalance remain common challenges. Classic descriptors like lipophilicity (logP/logD), molecular weight, TPSA, and hydrogen-bonding features continue to serve as robust predictors, yet more sophisticated representations – including 3D electrostatic potentials and SMILES-based encodings – are increasingly deployed to capture intricate structural and conformational subtleties. Indeed, Kim et al. [60] leverage electrostatic potential data to refine traditional QSAR approaches, while transformer models such as MegaMolBART [90] demonstrate strong feature extraction capabilities from SMILES. Deep learning architectures (e.g., CNN, RNN, GNN) can excel in large or specialized datasets, as illustrated by Ma et al.'s DeepB3P3 [67] for peptide modeling, though they often pose interpretability limitations. Methods like LIME or hybrid ML–DL ensembles from Yu et al. [73] aim to address these issues by balancing predictive power with transparency. Ensuring robust applicability domains and developing balanced, chemically diverse datasets are likewise crucial, as demonstrated by Sakiyama et al. [49] (using similarity-based data selection) and Alsenan et al. [65] (employing advanced augmentation). Across many studies, consensus or hybrid strategies show improved generalization, suggesting that combining ML and DL paradigms may deliver the most reliable solutions moving forward. Finally, user-friendly online servers such as ADMET Prediction Service [45] and PreADMET [46] provide additional avenues for benchmarking or cross-checking in-house models. This integration of computational tools with accessible platforms further fosters collaborative efforts to develop more accurate and widely applicable BBB permeability predictors.

Beyond improving classification metrics, these computational approaches also offer insights into molecular characteristics and mechanisms that influence CNS penetration. Models referencing lipophilicity (logP/logD), TPSA, hydrogen-bond donors, and specific structural motifs have outlined properties that favor passive diffusion or interact with efflux transporters. Such information can guide rational compound design, helping optimize BBB permeability while minimizing off-target effects and potential toxicity. Integrating these predictions into broader in silico ADMET frameworks facilitates earlier evaluation of pharmacokinetic and toxicity properties, reducing late-stage failures. One of the most notable limitations in most of these models are the limited data available. Oftentimes these models are trained on non-representative datasets that bias models leading them to fail in generalized drug development applications [107]. As datasets grow and feature-selection methods refine, ML and DL models may continue to improve, ultimately supporting more efficient screening and development strategies for CNS-targeted therapeutics.

Conclusion

The prediction of blood-brain barrier permeability has seen significant advancements through the application of both traditional ML and DL methods. Numerous studies have demonstrated the strengths and limitations of various models, highlighting the exceptional performance of ensemble approaches and modern DL techniques. For instance, Liu et al. achieved remarkable accuracy through combining various diverse data features of different models to develop their ensemble model using SVM, RF, and RF XGBoost. Other studies like Shaker et al. and Charoenkwan et al. showcased the efficacy of gradient boosting and scoring-card methods, respectively. Moreover, DL methods such as CNNs and GNNs have often performed at even higher levels comparatively. Nevertheless, challenges such as lack of diversity, dataset imbalance, and limited comprehensive molecular descriptors suggest that overfitting may be of concern.

To develop an ideal predictive model, a large, diverse, and balanced dataset is essential, encompassing various molecular descriptors like FPs and SMILES. Ensuring robust training through cross-validation techniques and enhancing model interpretability are crucial steps toward achieving reliable predictions. This can be done through combining methods of ML and DL models to enhance accuracy, interpretability, and generalizability. For instance, the ensemble model by Yu et al. is a hybrid model that utilizes both ML and DL methods and has shown remarkable accuracy and generalizability and provided insights on descriptors that distinguish CNS- and non-CNS active drugs. These hybrid models can further be used to develop drugs for targets outside of the CNS as well which potentially opens new doors for novel therapeutics in various fields of medicine. Moving forward, efforts should focus on deciphering DL pattern recognition to fully leverage their capabilities in understanding BBB permeability, ultimately paving the way for more effective drug development and therapeutic interventions.

Conflict of Interests

The authors declare no conflicts of interest.

Data Availability Statement

Data may be requested via the authors.

References

1. R. Gawdi, K. R. Shumway, and P. D. Emmady, *Physiology, Blood Brain Barrier* in *StatPearls*, (StatPearls Publishing, 2023). Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK557721/.

2. X. Zhang, T. Liu, X. Fan, and N. Ai, "In: Silico Modeling on ADME Properties of Natural Products: Classification Models for Blood-brain Barrier Permeability, its Application to Traditional Chinese Medicine and in Vitro Experimental Validation", *Journal of Molecular Graphics and Modelling* 75 (2017): 347–354, https://doi.org/10.1016/j.jmgm.2017.05.021.

3. D. Saxena, A. Sharma, M. H. Siddiqui, and R. Kumar, "Blood Brain Barrier Permeability Prediction using Machine Learning Techniques: An Update", *Current Pharmaceutical Biotechnology* 20, no. 14 (2019): 1163–1171, https://doi.org/10.2174/1389201020666190821145346.

4. X. Shao, Q. Shou, K. Felix, et al., "Age-related Decline in Blood-brain Barrier Function is More Pronounced in Males than Females in Parietal and Temporal Regions," *bioRxiv : The Preprint Server for Biology* 2024.01.12.575463 (2024), 10.1101/2024.01.12.575463.

5. A. Aragón-González, P. J. Shaw, and L. Ferraiuolo, "Blood-brain Barrier Disruption and its Involvement in Neurodevelopmental and Neurodegenerative Disorders", *International Journal of Molecular Sciences* 23, no. 23 (2022): 15271, https://doi.org/10.3390/ijms232315271.

6. A. Reichel, "The Role of Blood-brain Barrier Studies in the Pharmaceutical Industry", *Current Drug Metabolism* 7, no. 2 (2006): 183–203, https://doi.org/10.2174/138920006775541525.

7. D. E. Clark, "In: Silico Prediction Of Blood-brain Barrier Permeation", *Drug Discovery Today: Technologies* 1, no. 4 (2003): 441–448, 10.1016/S1740-6749(03)00163-6.

8. H. Pajouhesh, and G. R. Lenz, "Medicinal Chemical Properties of Successful Central Nervous System Drugs", *NeuroRx* 2 (2005): 541–553, https://doi.org/10.1602/neurorx.2.4.541.

9. D. Wu, Q. Chen, X. Chen, F. Han, Z. Chen, and Y. Wang, "The Blood-brain Barrier: Structure, Regulation, and Drug Delivery", *Signal Transduction and Targeted Therapy* 8, no. 1 (2023): 217, https://doi.org/10.1038/s41392-023-01481-w.

10. R. Y. Choi, A. S. Coyner, J. Kalpathy-Cramer, M. F. Chiang, and J. P. Campbell, "Introduction to Machine Learning, Neural Networks, and Deep Learning", *Translational Vision Science Technology* 9, no. 2 (2020): 14. https://doi.org/10.1167/tvst.9.2.14.

11. C. Cortes, "Support-vector Networks", *Machine Learning*, (1995). https://doi.org/10.1007/BF00994018.

12. What is Random Forest?. IBM. (2021, October 20). https://www.ibm.com/topics/random-forest.

13. T. Chen, and C. Guestrin, (2016, August). Xgboost: A Scalable Tree Boosting System. In: Proceedings of the 22nd acm Sigkdd International Conference on Knowledge Discovery and Data Mining (pp. 785–794).

14. A. Sharaff and H. Gupta, *Extra-tree Classifier with Metaheuristics Approach for Email Classification*: *Advances in Computer Communication and Computational Sciences: Proceedings of IC4S 2018* pp. 189–197. (Singapore: Springer, 2019).

15. N. Kriegeskorte, and T. Golan, "Neural Network Models and Deep Learning", *Current Biology* 29, no. 7 (2019): R231–R236, https://doi.org/10.1016/j.cub.2019.02.034.

16. T. N. Kipf, and M. Welling. 2016. "Semi-supervised Classification with Graph Convolutional Networks". arXiv preprint arXiv:1609.02907.

17. Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning", *Nature* 521, no. 7553 (2015): 436–444, https://doi.org/10.1038/nature14539.

18. Larry Hardesty | MIT News Office. (2014, April 4). Explained: Neural networks. MIT News | Massachusetts Institute of Technology.

19. A. Vaswani, N. Shazeer, N. Parmar, et al., "Attention is All you Need," *Advances in Neural Information Processing Systems 30*, 2017.

20. S. Hochreiter, and J. Schmidhuber, "Long short-term Memory", *Neural Computation* 9, no. 8 (1997): 1735–1780, https://doi.org/10.1162/neco.1997.9.8.1735.

21. B. Chandrasekaran, S. N. Abed, O. Al-Attraqchi, K. Kuche, and R. K. Tekade, *Computer-aided Prediction of Pharmacokinetic (ADMET) Properties : Dosage form Design Parameters*,(Academic Press, 2018), 731–755.https://www.sciencedirect.com/science/article/abs/pii/B978012814421300021X.

22. E. F. Gortari, C. R. Garcia-Jacas, K. Martinez-Mayorga, and J. L. Medina-Franco, "Database Fingerprint (DFP): An Approach to Represent Molecular Databases," *Journal of CheminformaticsBioMed Central*, (6, February) 2017, https://jcheminf.biomedcentral.com/articles/10.1186/s13321-017-0195-1#citeas.

23. D. Light, "SMILESTM Simplified Molecular Input Line Entry System," Daylight Cheminformatics, accesed May 2019, https://www.daylight.com/smiles/.

24. L. Breiman, "Statistical modeling: The Two Cultures (with comments and a rejoinder by the author)", *Statistical Science* 16, no. 3 (2001): 199–231, https://doi.org/10.1214/ss/1009213726.

25. L. Liu, L. Zhang, H. Feng, et al., "Prediction of the Blood-brain Barrier (BBB) Permeability of Chemicals based on Machine-learning and Ensemble Methods", *Chemical Research in Toxicology* 34, no. 6 (2021): 1456–1467, https://doi.org/10.1021/acs.chemrestox.0c00343.

26. H. Li, C. W. Yap, C. Y. Ung, Y. Xue, Z. W. Cao, and Y. Z. Chen, "Effect of Selection of Molecular Descriptors on the Prediction of Blood–brain Barrier Penetrating and Nonpenetrating Agents by Statistical Learning Methods", *Journal of Chemical Information and Modeling* 45, no. 5 (2005): 1376–1384, https://doi.org/10.1021/ci050135u.

27. B. Shaker, M. S. Yu, J. S. Song, et al., "LightBBB: Computational Prediction Model of Blood-brain-barrier Penetration based on LightGBM", *Bioinformatics* 37, no. 8 (2021): 1135–1139, https://doi.org/10.1093/bioinformatics/btaa918.

28. M. Adenot, and R. Lahana, "Blood-brain Barrier Permeation Models: Discriminating between Potential CNS and non-CNS Drugs Including P-glycoprotein Substrates", *Journal of Chemical Information and Computer Sciences* 44, no. 1 (2004): 239–248, https://doi.org/10.1021/ci034205d.

29. Z. Gao, Y. Chen, X. Cai, and R. Xu, "Predict Drug Permeability to Blood-brain-barrier from Clinical Phenotypes: Drug Side Effects and Drug Indications", *Bioinformatics* 33, no. 6 (2017): 901–908, https://doi.org/10.1093/bioinformatics/btw713.

30. I. F. Martins, A. L. Teixeira, L. Pinheiro, and A. O. Falcao, "A Bayesian Approach to in Silico Blood-brain Barrier Penetration Modeling", *Journal of Chemical Information and Modeling* 52, no. 6 (2012): 1686–1697, https://doi.org/10.1021/ci300124c.

31. F. Plisson, and A. M. Piggott, "Predicting Blood–brain Barrier Permeability of Marine-derived Kinase Inhibitors using Ensemble Classifiers Reveals Potential Hits for Neurodegenerative Disorders", *Marine Drugs* 17, no. 2 (2019): 81, https://doi.org/10.3390/md17020081.

32. M. Singh, R. Divakaran, L. S. K. Konda, R. Kristam, "A Classification Model for Blood Brain Barrier Penetration", *Journal of Molecular Graphics and Modelling* 96 (2020): 107516. https://doi.org/10.1016/j.jmgm.2019.107516.

33. Z. Wang, H. Yang, Z. Wu, et al., "In: Silico Prediction of Blood–brain Barrier Permeability of Compounds by Machine Learning and Resampling Methods", *ChemMedChem* 13, no. 20 (2018): 2189–2201, https://doi.org/10.1002/cmdc.201800533.

34. Z. Wu, B. Ramsundar, E. N. Feinberg, et al., "MoleculeNet: A Benchmark for Molecular Machine Learning", *Chemical Science* 9, no. 2 (2018): 513–530, https://doi.org/10.1039/C7SC02664A.

35. Y. Yuan, F. Zheng, and C. G. Zhan, "Improved Prediction of Blood–brain Barrier Permeability through Machine Learning with Combined use of Molecular Property-based Descriptors and Fingerprints", *AAPS Journal* 20 (2018): 1–10. https://doi.org/10.1208/s12248-018-0215-8.

36. B. Shaker, J. Lee, Y. Lee, et al., "A Machine Learning-based Quantitative Model (logBB Pred) to Predict the Blood-brain Barrier Permeability (logBB value) of Drug Compounds", *Bioinformatics (Oxford, England)* 39, no. 10 (2023): btad577, 10.1093/bioinformatics/btad577.

37. J. A. Platts, M. H. Abraham, Y. H. Zhao, A. Hersey, L. Ijaz, and D. Butina, "Correlation and Prediction of a Large Blood-brain Distribution Data Set–an LFER Study", *European Journal of Medicinal Chemistry* 36, no. 9 (2001): 719–730, https://doi.org/10.1016/S0223–5234(01)-01269–7.

38. X. C. Fu, Z. F. Song, C. Y. Fu, and W. Q. Liang, "A Simple Predictive Model for Blood-brain Barrier Penetration", *Die Pharmazie* 60, no. 5 (2005): 354–358.http://www.ingentaconnect.com/content/govi/pharmaz/2005/00000060/00000005/art00008?crawler=true.

39. Z. Bayat, J. Movaffagh, and S. Noruzi, "Development of a Computational Approach to Predict Blood-brain Permeability on Anti-viral Nucleoside Analogues", *Russian Journal of Physical Chemistry A* 85 (2011): 1923–1930, https://doi.org/10.1134/S0036024411110021.

40. M. Muehlbacher, G. M. Spitzer, K. R. Liedl, and J. Kornhuber, "Qualitative Prediction of Blood-brain Barrier Permeability on a Large and Refined Dataset", *Journal of Computer-aided Molecular Design* 25, no. 12 (2011): 1095–1106, https://doi.org/10.1007/s10822-011-9478-1.

41. T. S. Carpenter, D. A. Kirshner, E. Y. Lau, S. E. Wong, J. P. Nilmeier, and F. C. Lightstone, "A Method to Predict Blood-brain Barrier Permeability of Drug-like Compounds using Molecular Dynamics Simulations", *Biophysical Journal* 107, no. 3 (2014): 630–641, https://doi.org/10.1016/j.bpj.2014.06.024.

42. H. K. Shin, S. Lee, H. N. Oh, et al., "Development of Blood Brain Barrier Permeation Prediction Models for Organic and Inorganic Biocidal Active Substances", *Chemosphere* 277 (2021): 130330, https://doi.org/10.1016/j.chemosphere.2021.130330.

43. Q. Tang, F. Nie, Q. Zhao, and W. Chen, "A Merged Molecular Representation Deep Learning Method for Blood-brain Barrier Permeability Prediction", *Briefings in Bioinformatics* 23, no. 5 (2022): bbac357, https://doi.org/10.1093/bib/bbac357.

44. F. Grisoni, V. Consonni, and R. Todeschini, "Impact of Molecular Descriptors on Computational Models," *Computational Chemogenomics* (2018): 171–209. volume 1825, https://doi.org/10.1007/978-1-4939-8639-2-5.

45. A. S. Dyabina, E. V. Radchenko, V. A. Palyulin, "Prediction of Blood–brain Barrier Permeability of Organic Compounds", *Doklady Biochemistry and Biophysics* 470 (2016): 371–374, https://doi.org/10.1134/S1607672916050173.

46. S. Lee, G. S. Chang, I. H. Lee, "The PreADME: PC-based Program for Batch Prediction of ADME Properties", *EuroQSAR* 9 (2004): 5–10.https://preadmet.webservice.bmdrc.org/introduction/.

47. P. Charoenkwan, P. Chumnanpuen, N. Schaduangrat, P. Lio', M. A. Moni, and W. Shoombuatong, "Improved Prediction and Characterization of Blood-brain Barrier Penetrating Peptides using Estimated Propensity Scores of Dipeptides", *Journal of Computer-Aided Molecular Design* 36, no. 11 (2022): 781–796, https://doi.org/10.1007/s10822-022-00476-z.

48. V. Kumar, S. Patiyal, R. Kumar, et al., "B3Pdb: An Archive of Blood–brain Barrier-penetrating Peptides", *Brain Structure and Function* 226 (2021): 2489–2495, https://doi.org/10.1007/s00429-021-02341-5.

49. H. Sakiyama, R. Motoki, T. Okuno, and J. Q. Liu, "Improvement of Blood-brain Barrier Permeability Prediction using Cosine Similarity", *Journal of Computer Chemistry, Japan-International Edition* 9 (2023), https://doi.org/10.2477/jccjie.2023-0017.

50. Scikit-learn: Machine learning in Python-Scikit-learn 1.6.1 documentation. (n.d.). Retrieved January 1, 2025, from https://scikit-learn.org/stable/.

51. Y. Boulaamane, and A. Maurady, "EnsembleBBB: Enhanced Accuracy in Predicting Drug Blood-brain Barrier Permeability with a Machine Learning Ensemble Model," (2023). https://doi.org/10.21203/rs.3.rs-3752408/v1.

52. F. Meng, Y. Xi, J. Huang, and P. W. Ayers, "A Curated Diverse Molecular Database of Blood-brain Barrier Permeability with Chemical Descriptors", *Scientific Data* 8, no. 1 (2021): 289, https://doi.org/10.1038/s41597-021-01069-5.

53. V. Kumar, A. Banerjee, and K. Roy, "Innovative Strategies for the Quantitative Modeling of Blood–brain Barrier (BBB) Permeability: Harnessing the Power of Machine Learning-based q-RASAR Ap-

proach," *Molecular Systems Design Engineering*, (2024). https://doi.org/10.1039/D4ME00056K.

54. Z. Wu, Z. Xian, W. Ma, et al., "Artificial Neural Network Approach for Predicting Blood Brain Barrier Permeability based on a Group Contribution Method", *Computer Methods and Programs in Biomedicine* 200 (2021): 105943, https://doi.org/10.1016/j.cmpb.2021.105943.

55. F. O. Dehnbostel, V. A. Dixit, R. Preissner, and P. Banerjee, "Non-animal Models for Blood-brain Barrier Permeability Evaluation of Drug-like Compounds", *Scientific Reports* 14, no. 1 (2024): 8908, https://doi.org/10.1038/s41598-024-59734-9.

56. M. Muehlbacher, G. M. Spitzer, K. R. Liedl, and J. Kornhuber, "Qualitative Prediction Of Blood–brain Barrier Permeability on a Large and Refined Dataset", *Journal of Computer-Aided Molecular Design* 25 (2011): 1095–1106, https://doi.org/10.1007/s10822-011-9478-1.

57. F. M. Cornelissen, G. Markert, G. Deutsch, et al., "Explaining Blood–brain Barrier Permeability of Small Molecules by Integrated Analysis of Different Transport Mechanisms", *Journal of Medicinal Chemistry* 66, no. 11 (2023): 7253–7267. https://doi.org/10.1021/acs.jmedchem.2c01824.

58. J. Shen, F. Cheng, Y. Xu, W. Li, and Y. Tang, "Estimation of ADME Properties with Substructure Pattern Recognition", *Journal of Chemical Information and Modeling* 50, no. 6 (2010): 1034–1041, https://doi.org/10.1021/ci100104j.

59. W. Wang, M. T. Kim, A. Sedykh, and H. Zhu, "Developing Enhanced Blood–brain Barrier Permeability Models: Integrating External Bio-assay Data in QSAR Modeling", *Pharmaceutical Research* 32 (2015): 3055–3065, https://doi.org/10.1007/s11095-015-1687-1.

60. T. Kim, B. H. You, S. Han, H. C. Shin, K. C. Chung, and H. Park, "Quantum Artificial Neural Network Approach to Derive a Highly Predictive 3D-QSAR Model for Blood–brain Barrier Passage", *International Journal of Molecular Sciences* 22, no. 20 (2021): 10995, https://doi.org/10.3390/ijms222010995.

61. F. Ooms, P. Weber, P. A. Carrupt, and B. Testa, "A Simple Model to Predict Blood-Brain Barrier Permeation from 3D Molecular Fields", *Biochimica et Biophysica Acta* 1587, no. 2–3 (2002): 118–125, https://doi.org/10.1016/S0925-4439(02)00074-1.

62. Y. Brito-Sánchez, Y. Marrero-Ponce, S. J. Barigye, et al., "Towards Better BBB Passage Prediction using an Extensive and Curated Data Set", *Molecular Informatics* 34, no. 5 (2015): 308–330, 10.1002/minf.201400118.

63. E. V. Radchenko, A. S. Dyabina, and V. A. Palyulin, "Towards Deep Neural Network Models for the Prediction of the Blood–brain Barrier Permeability for Diverse Organic Compounds", *Molecules* 25, no. 24 (2020): 5901, https://doi.org/10.3390/molecules25245901.

64. Y. Brito-Sánchez, Y. Marrero-Ponce, S. J. Barigye, et al., "Towards Better BBB Passage Prediction using an Extensive and Curated Data Set", *Molecular Informatics* 34, no. 5 (2015): 308–330, https://doi.org/10.1002/minf.201400118.

65. S. Alsenan, I. Al-Turaiki, and A. Hafez, "A Recurrent Neural Network Model to Predict Blood–brain Barrier Permeability", *Computational Biology and Chemistry* 89 (2020): 107377, https://doi.org/10.1016/j.compbiolchem.2020.107377.

66. A. Achiaa Atwereboannah, W. P. Wu, E. Nanor, "Prediction of Drug Permeability to the Blood-Brain Barrier using Deep Learning," in *4th International Conference on Biometric Engineering and Applications* (2021, May). pp. 104–109.

67. C. Ma, R. Wolfinger, "A Prediction Model for Blood-brain Barrier Penetrating Peptides based on Masked Peptide Transformers with Dynamic Routing", *Briefings in Bioinformatics* 24, no. 6 (2023): bbad399. DOI:10.1093/bib/bbad399.

68. S. Sabour, N. Frosst, G. E. Hinton, Dynamic Routing between Capsules. Advances in Neural Information Processing Systems, 30 (2017).

69. V. Kumar, S. Patiyal, A. Dhall, N. Sharma, and G. P. S. Raghava, "B3Pred: A Random-Forest-Based Method for Predicting and Designing Blood-Brain Barrier Penetrating Peptides", *Pharmaceutics* 13, no. 8 (2021): 1237, https://doi.org/10.3390/pharmaceutics13081237.

70. Y. H. Zhao, M. H. Abraham, A. Ibrahim, et al., "Predicting Penetration Across the Blood-Brain Barrier from Simple Descriptors and Fragmentation Schemes", *Journal of Chemical Information and Modeling* 47, no. 1 (2007): 170–175. https://doi.org/10.1021/ci600312d.

71. R. Dai, W. Zhang, W. Tang, et al., "BBPpred: Sequence-Based Prediction of Blood-Brain Barrier Peptides with Feature Representation Learning and Logistic Regression", *Journal of Chemical Information and Modeling* 61, no. 1 (2021): 525–534, https://doi.org/10.1021/acs.jcim.0c01115.

72. W. He, Y. Jiang, J. Jin, et al., "Accelerating Bioactive Peptide Discovery via Mutual Information-Based Meta-Learning", *Briefings in Bioinformatics* 23, no. 1 (2022): bbab499, https://doi.org/10.1093/bib/bbab499.

73. T. H. Yu, B. H. Su, L. C. Battalora, S. Liu, and Y. J. Tseng, "Ensemble Modeling with Machine Learning and Deep Learning to Provide Interpretable Generalized Rules for Classifying CNS Drugs with High Prediction Power", *Briefings in Bioinformatics* 23, no. 1 (2022): bbab377, https://doi.org/10.1093/bib/bbab377.

74. C. W. Yap, "PaDEL-descriptor: An Open Source Software to Calculate Molecular Descriptors and Fingerprints", *Journal of Computational Chemistry* 32, no. 7 (2011): 1466–1474, https://doi.org/10.1002/jcc.21707.

75. H. van de Waterbeemd, G. Camenisch, G. Folkers, J. R. Chretien, and O. A. Raevsky, "Estimation of Blood-Brain Barrier Crossing of Drugs using Molecular Size and Shape, and H-bonding Descriptors", *Journal of Drug Targeting* 6, no. 2 (1998): 151–165, https://doi.org/10.3109/10611869808997889.

76. N. Aftab, F. Masood, S. Ahmad, "An Optimized Deep Learning Approach for Blood-Brain Barrier Permeability Prediction with ODE Integration," *Informatics in Medicine Unlocked* (2024): 101526, https://doi.org/10.1016/j.imu.2024.101526.

77. V. Kumar, A. Banerjee, and K. Roy, "Breaking the Barriers: Machine-Learning-Based c-RASAR Approach for Accurate Blood–Brain Barrier Permeability Prediction", *Journal of Chemical Information and Modeling* 64, no. 10 (2024): 4298–4309, https://doi.org/10.1021/acs.jcim.4c00433.

78. R. Kato, W. Zeng, V. B. Siramshetty, "Development and Validation of PAMPA-BBB QSAR Model to Predict Brain Penetration Potential of Novel Drug Candidates", *Frontiers in Pharmacology* 14 (2023): 1291246. https://doi.org/10.3389/fphar.2023.1291246.

79. A. Mauri, and M. Bertola, "Alvascience: A New Software Suite for the QSAR Workflow Applied to the Blood–brain Barrier Permeability", *International Journal of Molecular Sciences* 23, no. 21 (2022): 12882, https://doi.org/10.3390/ijms232112882.

80. S. Hamzic, R. Lewis, S. Desrayaud, "Predicting in Vivo Compound Brain Penetration using Multi-task Graph Neural Networks", *Journal of Chemical Information and Modeling* 62, no. 13 (2022): 3180–3190, https://doi.org/10.1021/acs.jcim.2c00412.

81. Novartis Open Source Science. Novartis. (n.d.). https://www.novartis.com/research-development/open-source-science.

82. D. Ajabani, "A Computational Prediction Model of Blood-Brain Barrier Penetration Based on Machine Learning Approaches" (2023).

83. Landrum greg. (n.d.). Open-source cheminformatics software. RDKit. https://www.rdkit.org/.

84. A. Mauri, V. Consonni, M. Pavan, and R. Todeschini, "Dragon Software: An Easy Approach to Molecular Descriptor Calculations", *Match* 56, no. 2 (2006): 237–248.https://match.pmf.kg.ac.rs/electronic_versions/Match56/n2/match56n2_237-248.pdf.

85. B. Mazumdar, P. K. D. Sarma, H. J. Mahanta, and G. N. Sastry, "Machine Learning based Dynamic Consensus Model for Predicting Blood-brain Barrier Permeability", *Computers in Biology and Medicine* 160 (2023): 106984, https://doi.org/10.1016/j.compbiomed.2023.106984.

86. R. Miao, L. Y. Xia, H. H. Chen, H. H. Huang, and Y. Liang, "Improved Classification of Blood-Brain-Barrier Drugs using Deep Learning", *Scientific Reports* 9, no. 1 (2019): 8802, https://doi.org/10.1038/s41598-019-44773-4.

87. I. Letunic, (n.d.). Sider 4.1 : Side Effect Resource. SIDER Side Effect Resource. http://sideeffects.embl.de/.

88. R. Kumar, A. Sharma, A. Alexiou, A. L. Bilgrami, M. A. Kamal, and G. M. Ashraf, "DeePred-BBB: A Blood Brain Barrier Permeability Prediction Model with Improved Accuracy", *Frontiers in Neuroscience* 16 (2022): 858126, https://doi.org/10.3389/fnins.2022.858126.

89. D. Roy, V. K. Hinge, and A. Kovalenko, "To Pass or Not to Pass: Predicting the Blood–brain Barrier Permeability with the 3D-RISM-KH Molecular Solvation Theory", *ACS Omega* 4, no. 16 (2019): 16774–16780.

90. E. T. Huang, J. S. Yang, K. Y. Liao, "Predicting Blood–brain Barrier Permeability of Molecules with a Large Language Model and Machine Learning", *Scientific Reports* 14, no. 1 (2024): 15844. https://doi.org/10.1038/s41598-024-66897-y.

91. Natural Products Research Laboratory NPRL. (n.d.). http://nprl.gov.np/.

92. L. C. S. Rosa, C. O. Argolo, C. M. C. Nascimento, and A. S. Pimentel, "Identifying Substructures that Facilitate Compounds to Penetrate the Blood–Brain Barrier via Passive Transport using Machine Learning Explainer Models", *ACS Chemical Neuroscience* (2024). https//ido.org/10.1021/acschemneuro.3c00840.

93. Q. Tang, F. Nie, Q. Zhao, and W. Chen, "A Merged Molecular Representation Deep Learning Method for Blood–Brain Barrier Permeability Prediction", *Briefings in Bioinformatics* 23, no. 5 (2022): bbac357, https://doi.org/10.1093/bib/bbac357.

94. A. K. Ghose, T. Herbertz, R. L. Hudkins, B. D. Dorsey, and J. P. Mallamo, "Knowledge-based, Central Nervous System (CNS) Lead Selection and Lead Optimization for CNS Drug Discovery", *ACS Chemical Neuroscience* 3, no. 1 (2012): 50–68, https://doi.org/10.1021/cn200100h.

95. W. A. Mackey, and G. B. Shaw, "Oral Hexamethonium Bromide in Essential Hypertension", *British Medical Journal* 2, no. 4726 (1951): 259, https://doi.org/10.1136/bmj.2.4726.259.

96. Z. Zhang, and Z. Zhang, et al., "Artificial neural network," *Multivariate Time Series Analysis in Climate and Environmental Research* ·XX· (2018): 1–35. https://doi.org/10.1007/978-3-319-67340-0_1.

97. L. Craig, and R. Awati, "What is a Convolutional Neural Network (CNN). Enterprise AI," accessed January 11, 2024, https://www.techtarget.com/searchenterpriseai/definition/convolutional-neural-network.

98. W. M. Pardridge, "Drug Transport Across the Blood-brain Barrier", *Journal of Cerebral Blood Flow and Metabolism* 32, no. 11 (2012): 1959–1972, https://doi.org/10.1038/jcbfm.2012.126.

99. Y. Yu, X. Si, C. Hu, and J. Zhang, "A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures", *Neural Computation* 31, no. 7 (2019): 1235–1270, https://doi.org/10.1162/neco a 01199.

100. Z. Zhang, L. Chen, F. Zhong, et al, "Graph Neural Network Approaches for Drug-target Interactions", *Current Opinion in Structural Biology* 73 (2022): 102327, https://doi.org/10.1016/j.sbi.2021.102327.

101. J. Pearl, M. Glymour, and N. P. Jewell, *Causal Inference in Statistics: A Primer*, (Wiley, 2016), https://www.wiley.com/en-us/Causal+Inference+in+Statistics%3A+A+Primer-p-9781119186847.

102. S. I. Ahn, Y. J. Sei, H. J. Park, et al., "Microengineered human blood–brain barrier platform for understanding nanoparticle transport mechanisms." *Nature Communications* 11 (2020): 175. https://doi.org/10.1038/s41467-019-13896-7.

103. V. Kumar, et al, "B3Pdb: An Archive of Blood–brain Barrier-penetrating Peptides", *Brain Structure Function* 226, no. 8 (2021): 2489–2495, https://doi.org/10.1007/s00429-021-02341-5.

104. Q. Tang, F. Nie, Q. Zhao, and W. Chen, "A Merged Molecular Representation Deep Learning Method for Blood-brain Barrier Permeability Prediction", *Briefings in Bioinformatics* 23, no. 5 (2022): bbac357, https://doi.org/10.1093/bib/bbac357.

105. S. M. McDonald, E. K. Augustine, Q. Lanners, C. Rudin, L. Catherine Brinson, and M. L. Becker, "Applied Machine Learning as a Driver for Polymeric Biomaterials Design", *Nature Communications* 14, no. 1 (2023): 4838, https://doi.org/10.1038/s41467-023-40459-8.

106. Y. Wang, E. Gallagher, C. Jorgensen, "An Experimentally Validated Approach to Calculate the Blood-brain Barrier Permeability of Small Molecules", *Scientific Reports* 9, no. 1 (2019): 6117, https://doi.org/10.1038/s41598-019-42272-0.

107. F. Kretschmer, J. Seipp, M. Ludwig, G. W. Klau, and S. Böcker, "Coverage Bias in Small Molecule Machine Learning", *Nature Communications* 16, no. 1 (2025): 554, https://doi.org/10.1038/s41467-024-55462-w.

108. Larry Hardesty | MIT News Office. (2014, April 4). Explained: Neural networks. MIT News | Massachusetts Institute of Technology. https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.