An Effective Ensemble Deep Learning Framework for Blood-brain Barrier Permeability Prediction

Thanh-Hoang Nguyen-Vo*,†, Trang T. T. Do‡, and Binh P. Nguyen*
*School of Mathematics and Statistics, Victoria University of Wellington, Wellington 6012, New Zealand
Email: hoang.nguyen@weltec.ac.nz, binh.p.nguyen@vuw.ac.nz
†School of Innovation, Design and Technology, Wellington Institute of Technology, Lower Hutt 5012, New Zealand
‡Ministry of Business, Innovation and Employment, Wellington 6011, New Zealand
Email: trang.do@mbie.govt.nz

***Abstract*—As a highly protective biological structure, the blood-brain barrier prevents the uncontrolled passage of molecules to keep the central nervous system free from chemical toxification and maintain brain homeostasis. Since most substances are not allowed to freely penetrate the blood-brain barrier, examination of the blood-brain barrier permeability (BBBP) of drug candidates is highly essential in drug discovery. To screen the BBBP of molecules, several computational methods were developed with satisfactory outcomes. These methods, however, have shortcomings that need to be addressed to improve prediction performance. In our study, we propose iBBBP-Ensemble, an ensemble deep learning model that combines two types of neural networks: a convolutional neural network and multilayer perceptrons, and three types of molecular representations: the Extended-Connectivity Fingerprint, RDKit molecular descriptors, and Mol2vec-embedded features. Experimental results confirmed the effectiveness and stability of our proposed model. The benchmarking analysis also indicated that iBBBP-Ensemble outperformed all machine learning and deep learning baseline models.**

***Index Terms*—blood-brain barrier permeability, BBBP, deep learning, ensemble learning, ADME, drug discovery**

I. Introduction

Optimizing the pharmacokinetic and pharmacodynamic properties of drug candidates is an essential stage in drug discovery. Among these properties, the ability or inability to penetrate the blood-brain barrier is crucial for investigating drugs targeting receptors in the brain. The blood-brain barrier (BBB) is a dynamic biological layer that strictly controls the non-selective penetration of compounds to protect the central nervous system (CSN) and maintain brain homeostasis [1], [2]. The molecular structures of drugs are designed to either enhance or suppress their ability to pass the BBB. To effectively treat CNS disorders, drug candidates are driven to have enhanced BBB penetration abilities. While treating other diseases, drug candidates are designed to have suppressed BBB penetration abilities or even BBB penetration inabilities to reduce undesirable CNS-related side effects [1], [2]. Numerous prospective drugs have been discontinued because they can penetrate the BBB and exert their full effectiveness. One hundred percent of large-molecule drugs and more than ninety percent of small-molecule drugs are unable to penetrate the BBB [3]. Therefore, it is essential to develop a framework that can predict the BBB permeability (BBBP) of drug candidates.

Conventionally, the BBBP of drug molecules has been determined via *in vitro* [4], [5] and *in vivo* [6], [7] methods. Those approaches, however, are expensive, time-consuming, and labor-intensive [8], [9]. As computer-assisted approaches for analyzing BBBP offer significant benefits, including high throughput, low cost, and great efficiency [10], many computational models for BBBP prediction have been developed using diverse techniques and features [11]. According to the diversity of their underlying methodologies, these models are divided into four categories: mathematical learning, statistical learning, machine learning, and deep learning. The MPO [12], MPO_V2 [13], and BBB Score [14] are typical mathematical learning models designed with specific probability functions based on the physicochemical properties of molecules. Several statistical learning models were constructed using different techniques, such as multiple linear regression, partial least squares, and linear discriminant analysis [15], [16]. Besides, machine learning models were developed using distinct algorithms, encompassing Logistic Regression [17], Random Forest [18]–[20], Support Vector Machines [20], [21], Naïve Bayes [20], Light Gradient Boosting Machine [22], and eXtreme Gradient Boosting [23], [24]. Multiple deep learning architectures, comprising artificial neural networks (ANN) [25]–[27], recurrent neural networks [28], and convolutional neural networks [29], were utilized to develop models for BBBP prediction. Besides, hybrid models combining several techniques were released with promising outcomes [10], [27], [29]. Although the mathematical and statistical learning models are explainable, their prediction efficiency is limited due to the fact that their probability functions were derived using insufficient data. Since these machine learning models were developed using larger datasets, they were more generalizable and accurate at making predictions. However, the model creators did not examine the class imbalance issue and how it affected model precision. While the existing deep learning models are either outdated [25], [26] or inappropriately formulated [27]–[29]. Wu *et al.* [27] created their model using a very small dataset

The work of BPN was supported by the Faculty Strategic Research Grant (FSRG) numbers 410132 & 411494 at Victoria University of Wellington (VUW) and the Endeavour Fund (Smart Ideas) from the New Zealand Ministry of Business, Innovation and Employment (MBIE) under contract VUW RTVU2301. The work of THNV was partly supported by the Whitireia and WelTec Contestable Fund.

979-8-3503-5409-6/24/$31.00 ©2024 IEEE
DOI 10.1109/CAI59869.2024.00038

of 260 training samples and 40 test samples. Alsenan *et al.*'s model was overfitted due to data leakage [28]. They pre-processed all the data before splitting it into training and test sets [28]. Thus, an effective and robust computational model for BBBP prediction is highly essential. The model needs to be trained with an up-to-date, well-curated dataset under an appropriate modeling strategy and evaluated by reasonable metrics.

In our study, we propose an effective deep learning framework called iBBBP-Ensemble to predict the BBBP of compounds. Three types of features: Extended connectivity fingerprints (ECFP) [30], RDKit molecular descriptors (RDKitMD) [31], and Mol2vec-embedded features (Mol2vec) [32] are used as independent inputs of three separate neural networks. The outputs of these three neural networks are then averaged to obtain the final outputs. Our proposed model is also benchmarked with several conventional machine learning and deep learning models to fairly assess its performance.

II. Materials and Methods

A. Dataset

The chemical data used for model development and evaluation were collected from the B3DB database, which contains curated BBBP compounds [33]. The B3DB database has 4956 compounds that are able to pass the blood-brain barrier (BBB+) and 2851 compounds that are unable to pass the blood-brain barrier (BBB-). This database is currently the largest BBB dataset, whose chemical records were collected from numerous peer-reviewed publications and open-access datasets until now.

Prior to constructing the model, a data curation process was carried out to qualify the input data for both the development and evaluation stages. We used the data curation procedure developed by Fourches *et al.* [34] with slight adjustments [35], [36]. The procedure consists of four main stages: (1) validation, (2) cleaning, (3) normalization, and (4) final verification. Stage 1 eliminates any molecules that are classified as inorganic, mixtures, or organometallics. Stage 2 discards salts and resolves two kinds of charged molecules: metal-containing and organic ones. While metal-containing charged molecules are rejected due to possessing metal atoms with positive charges, organic charged molecules are kept unchanged because precisely defining the experimental settings under which charged organic molecules are active is infeasible. Hence, charged organic molecules are neutralized. Stage 3 transforms all tautomers (detautomerization) and stereoisomers (destereoisomerization) of the same molecules into their unique, standardized forms to ensure no interconvertible isomers are available. Any duplicates detected in these three stages are removed at the end of each stage. Stage 4 removes any chemical data (molecules) holding conflicting labels, or the so-called '*conflicting record*'. All chemical data were formatted in canonical SMILES form before entering any processing stages. All chemical data were formatted in canonical SMILES form before entering any processing stages. It is worth noting that BBB+ and BBB- samples were separately processed. Conflicting records may appear in one of the following three cases: (a) two or more identical molecules carrying different labels; (b) duplicate molecules carrying the same label; and (c) two or more distinct molecules whose CAS registry numbers are the same. While any records that match cases (a) or (c) were discarded, records that match case (b) were merged into one. Furthermore, all molecules were cross-referenced in the PubChem and ChEMBL databases, the two largest and most reliable chemical databases, to ascertain their structural identities and validities.

TABLE I: Datasets for model development and evaluation

Dataset	Number of molecules		
	BBB+	BBB-	Total
Training	1764	1000	2764
Validation	311	177	488
Test	366	208	574

After the chemical data were curated, we obtained 3826 curated chemical samples. Then, we used stratified random sampling to select 15% of the total curated data to build an independent test set. The remaining chemical data were used as a training set (for developing machine learning models) or continued to be divided into a training set and a validation set with a ratio of 85:15 (for developing deep learning models). Table I provides the numbers of curated chemical data for model development and evaluation.

B. Molecular Representations

1) Extended Connectivity Fingerprints: Extended-connectivity fingerprints (ECFPs) are a subtype of topological fingerprints, also known as circular fingerprints or Morgan fingerprints [30]. They were initially created for substructural similarity searching before being upgraded to support structure-activity modeling. The ECFP scheme allows users to generate various types of circular fingerprints (ECFP$\underline{D}$ with D as the diameter) by adjusting the radius and bit counts. The BBB molecules were converted into 2048-dimensional binary vectors using a radius of 2 so-called '*ECFP4-2048*' or '*ECFP4*' vectors. Since the ECFP4 vectors are sparse vectors, they were transformed into their corresponding index vectors used for deep learning models. The details of transforming ECFP4 vectors into corresponding index vectors are clearly explained in other studies [37], [38].

2) RDKit Molecular Descriptors: The RDKit molecular descriptor (RDKit MD) set, which consists of 196 physicochemical properties, was calculated using RDKit, an open-source cheminformatics library [31]. The feature set includes 4 molecular property descriptors, 58 MOE-type descriptors, 7 topological descriptors, 1 charged partial surface area descriptor, 12 connectivity descriptors, and 106 constitutional descriptors. The molecular descriptors of BBB molecules were computed as vectors of size 1×196.

3) Mol2vec-embedded Features: Mol2vec [32] is a pre-trained molecular encoder for converting molecular structures into numerical vectors, also known as molecular embeddings. These Mol2vec-embedded vectors can be used as input for

machine learning models or other types of data analysis [39]. The concept of Mol2vec is based on the idea of a molecular fingerprint. Learning to map molecular fingerprints to continuous vectors helps it to capture more complex structural information than traditional fingerprints. This mapping enables the use of molecular embeddings in various downstream applications, such as virtual screening, drug discovery, and quantitative structure-activity relationship modeling. The BBB molecules were converted into Mol2vec-embedded vectors with a dimension of 1×300.

C. Model Architecture

Our proposed deep learning model is designed with three independent neural networks that accept three different types of molecular representation vectors: ECFP-embedded vectors, RDKitMD vectors, and Mol2vec-embedded vectors (Figure 1). The neural networks for ECFP-embedded vectors, RDKitMD vectors, and Mol2vec-embedded vectors are termed NN1, NN2, and NN3, respectively. NN1 is a convolutional neural network (CNN), while NN2 and NN3 are multilayer perceptrons (MLP). The NN1 branch has one embedding layer, one 2D-convolutional (Conv2D) layer, one batch normalization (BatchNorm) layer, one 2D max-pooling (MaxPool2D) layer, and two fully connected (FC) layers. NN2 and NN3 are similarly designed with three fully connected layers. The first fully connected (FC1) layers of NN2 and NN3 accept RDKitMD and Mol2vec-embedded vectors with dimensions of 1×196 and 1×300, respectively. A dropout rate of 0.1 is applied after the first and second fully connected layers. The outputs of NN1, NN2, and NN3 are probabilities that are then ensembled to form the final predicted probability. The Leaky Rectified Linear Unit (LeakyReLU) is the only activation function used in all layers of these neural networks, except for the final layers, which are activated by the sigmoid function.

Our model was optimized using the Adam optimizer [40] with a learning rate of 1×10^{-4} to iteratively tune the network parameters. Since the problem is binary classification, we employed binary cross-entropy as the loss function. All deep learning models in our study were implemented using PyTorch 1.10.2 and trained on an i7-9700 CPU equipped with 32GB of RAM and one NVIDIA Titan Xp GPU. Our models were trained over a period of 20 epochs. The final model selection was based on the epoch corresponding to the lowest validation loss. The duration of training for each epoch was approximately 2 seconds, while the duration of testing for each epoch was approximately 0.25 seconds.

III. Results and Discussion

A. Model Evaluation

To examine the effectiveness of the ensemble learning strategy, we compared iBBBP-Ensemble to its component neural network models. All models were developed and evaluated using the same training, validation, and test sets. We assessed the performance of all models with four evaluation metrics, including balanced accuracy (BA), F1 score (F1), Matthews's correlation coefficient (MCC), the area under the receiver operating characteristic curve (AUCROC), and the area under the precision-recall curve (AUCPR). Table II summarizes the results of model evaluation on the independent test set. The results indicate that ensemble learning can improve the model's performance. Our proposed model outperforms its component neural network models, with AUCROC and AUCPR values of 0.9031 and 0.9426, respectively. Our model also achieves an F1 score of 0.8730 and an MCC value of 0.6316. The second-best model is the MLP developed with RDKitMD features, followed by the MLP developed with Mol2vec-embedded features, and the CNN developed with ECFP4-embedded features.

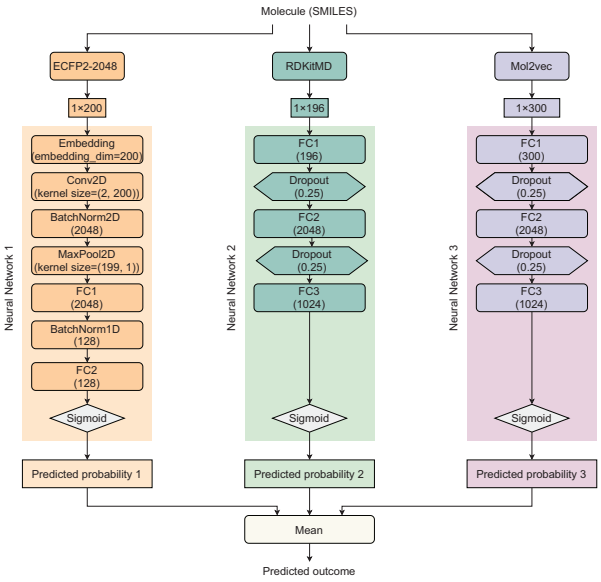

Fig. 1: Model architecture of iBBBP-Ensemble.

B. Model Benchmarking

1) Benchmarking with Machine Learning Models: To examine the performance of iBBBP-Ensemble, we also compared it with 20 conventional machine learning models developed by combining five learning algorithms: AdaBoost Classifier [41] (ABC), Extremely Randomized Tree (ERT), Gradient Boosting [42] (GB), Random Forest (RF) [43], and eXtreme Gradient Boosting (XGB) [44], and three types of molecular representations: ECFP4, RDKitMD, and Mol2vec-embedded features. Since the construction of machine learning models often requires hyper-parameter tuning via k-fold cross-validation, we merged the training data and validation data (used for training iBBBP-Ensemble) to create a new training set. All machine learning models were tuned and trained on the new training set. The independent test set remains unchanged.

Table III compares the performance of iBBBP-Ensemble and that of 20 machine learning models on the independent test set. It is unsurprising that the machine learning models trained

TABLE II: The performance of iBBBP-Ensemble and other deep learning models

Model	Feature			Metric				
	ECFP4	RDKitMD	Mol2vec	AUCROC	AUCPR	BA	F1	MCC
CNN	+	-	-	0.8701	0.9198	0.7776	0.8388	0.5551
MLP	-	+	-	0.8817	0.9168	0.7894	0.8632	0.5997
MLP	-	-	+	0.8802	0.9231	0.7623	0.8531	0.5582
Ours	+	+	+	**0.9031**	**0.9426**	**0.8066**	**0.8730**	**0.6316**

TABLE III: The performance of iBBBP-Ensemble and other conventional machine learning models

Model	Feature			Metric				
	ECFP4	RDKitMD	Mol2vec	AUCROC	AUCPR	BA	F1	MCC
ABC	+	-	-	0.7584	0.8240	0.5677	0.7987	0.2684
ERT	+	-	-	0.8047	0.8726	0.6381	0.8230	0.4058
GB	+	-	-	0.8125	0.8755	0.6467	0.8272	0.4258
RF	+	-	-	0.8177	0.8795	0.6302	0.8223	0.4032
XGB	+	-	-	0.8134	0.8854	0.6828	0.8303	0.4482
ABC	-	+	-	0.8163	0.8601	0.6959	0.8350	0.4690
ERT	-	+	-	0.8767	0.9172	0.7330	0.8501	0.5309
GB	-	+	-	0.8766	0.9144	0.7413	0.8529	0.5427
RF	-	+	-	0.8744	0.9170	0.7214	0.8408	0.5005
XGB	-	+	-	0.8777	0.9184	0.7375	0.8445	0.5209
ABC	-	-	+	0.8171	0.8764	0.6632	0.8264	0.4259
ERT	-	-	+	0.8756	0.9261	0.6811	0.8329	0.4555
GB	-	-	+	0.8652	0.9093	0.6911	0.8329	0.4605
RF	-	-	+	0.8720	0.9203	0.6914	0.8347	0.4660
XGB	-	-	+	0.8727	0.9107	0.7375	0.8504	0.5343
ABC	+	+	+	0.8347	0.8866	0.6577	0.8267	0.4248
ERT	+	+	+	0.8799	0.9261	0.6862	0.8367	0.4702
GB	+	+	+	0.8807	0.9234	0.6996	0.8431	0.4964
RF	+	+	+	0.8805	0.9251	0.7031	0.8437	0.4995
XGB	+	+	+	0.8864	0.9306	0.7443	0.8575	0.5567
Ours	+	+	+	**0.9031**	**0.9426**	**0.8066**	**0.8730**	**0.6316**

with three feature types show better performance compared to their corresponding models trained with a single feature type. Also, results indicate that among these three feature types, RDKitMD is more effective than Mol2vec and ECPF4. The machine learning models trained with ECFP4 features work the least effectively, with all AUCROC values around 0.85. All the conventional machine learning models work less effectively than ours, while their computational costs required for model development are higher. Overall, the use of different feature types can elevate the distinctly extracted features of input molecules.

2) Benchmarking with Deep Learning Models: The performance of iBBBP-Ensemble was also examined by comparing it with three advanced deep learning architectures using molecular graphs as input features. These graph-based deep learning models are Graph Convolutional Network (GCN), TrimNet [45], and Graph Multiset Transformer [46] (GMT). The GCN model consists of three graph convolutional layers developed by Kipf *et al.* [47], followed by three fully connected layers. The main idea of the TrimNet model is the utilization of the multi-head triplet attention mechanism in learning molecular graphs [45], while the GMT model is specified by multi-head attention-based global pooling layers capturing the node-node interaction based on their structural dependencies. All deep learning models, including ours, were trained and evaluated under the same training, validation, and test sets. Table IV compares the performance of iBBBP-Ensemble and that of three deep learning models on the independent test set. The generation of node features used for developing the GCN, TrimNet and GMT models was exactly based on their original source code. Results reveal that among these three advanced graph-based deep learning models, the GCN model has the highest performance, followed by the TrimNet, and GMT models. However, these models' performances do not exceed our model's. The volume of data is very likely one of the main reasons that affects the efficiency of these advanced deep learning models. Under this circumstance, our model proves its effectiveness with less complex neural network architectures.

TABLE IV: The performance of iBBBP-Ensemble and other advanced deep learning models

Model	Metric				
	AUCROC	AUCPR	BA	F1	MCC
GCN	0.8708	0.9165	0.7827	0.8495	0.5730
TrimNet	0.8669	0.9137	0.7606	0.8436	0.5409
GMT	0.8629	0.9103	0.7682	0.8489	0.5565
Ours	**0.9031**	**0.9426**	**0.8066**	**0.8730**	**0.6316**

C. Leave-One-Compound-Cluster-Out Validation

To assess the model's robustness and stability, we conducted Leave-One-Compound-Cluster-Out (LOCCO) validation [48] to explore the practical situation where the models predict the BBBP of novel molecules. The Self-Organizing Map (SOM) algorithm [49] was employed to create a two-dimensional map of size 3×3 of all the compounds featured by ECFP4 (using 1024 bits) in the curated dataset. The SOM algorithm can

TABLE V: The performance of iBBBP-Ensemble and other models under the LOCCO validation

Model	Feature			Metric				
	ECFP4	RDKitMD	Mol2Vec	AUCROC	AUCPR	BA	F1	MCC
ABC	+	+	+	0.7054	0.7900	0.5948	0.7889	0.2788
ERT	+	+	+	0.7857	0.8446	0.6209	**0.8085**	0.3568
GB	+	+	+	0.7670	0.8318	0.5286	0.7737	0.1584
RF	+	+	+	0.7924	0.8508	0.6417	0.8077	0.3821
XGB	+	+	+	0.7892	0.8523	0.6660	0.8040	0.3841
GCN	n/a	n/a	n/a	0.7795	0.8293	0.7021	0.7836	0.4267
TrimNet	n/a	n/a	n/a	0.7396	0.8072	0.6795	0.7535	0.3646
GMT	n/a	n/a	n/a	0.7742	0.8149	0.6762	0.7869	0.3983
Ours	+	+	+	**0.8070**	**0.8539**	**0.6991**	0.7921	**0.4317**

be utilized for clustering as it can maintain the topological structure of the chemical data [50]. With the two-dimensional map of size 3×3, we obtained 9 clusters of all the molecules. Each of the 9 clusters was iteratively chosen as the test set, while the 8 remaining clusters were combined and then randomly divided into a training set and a validation set with a ratio of 85:15.

Table V shows the average test performance of the iBBBP-Ensemble and eight models (five conventional machine learning and three advanced deep learning models) under the LOCCO validation. For each test cluster, we calculated the evaluation metrics and multiplied these values by the weight of the test cluster's size. Generally, the performances of all models, including iBBBP-Ensemble, under the LOCCO validation are lower than those under the random sampling strategy. This situation reflects a practical approach in which the chemical space of the test cluster is far different from the chemical space of the training cluster. This experiment was designed to examine the lower limit of the model's performance when the model has to predict structurally distant molecules. Although the performance of all models slightly declines, iBBBP-Ensemble is still dominant over other models in terms of both AUCROC and AUCPR values. The results of this experiment demonstrate the robustness and stability of our model when predicting novel molecules.

IV. Conclusion

The determination of the blood-brain barrier permeability (BBBP) of molecules is crucial in drug discovery, particularly in the development of central nervous system drugs. In our study, we developed iBBBP-Ensemble, an effective and robust computational framework for predicting the BBBP of molecules. This framework uses an ensemble deep learning architecture combined with three molecular representations, a combination that has been shown to enhance the predictive power for screening BBB-permeable molecules. Benchmarking results confirmed that our method outperformed other deep learning and conventional machine learning models. Moreover, the iBBBP-Ensemble model has a minimal computational expense, allowing users to effortlessly make adjustments to the model by incorporating supplementary data in the future or to adapt it for alternative objectives.

References

[1] M. W. Bradbury, "The blood-brain barrier," *Experimental Physiology: Translation and Integration*, vol. 78, no. 4, pp. 453–472, 1993.

[2] M. Segarra, M. R. Aburto, and A. Acker-Palmer, "Blood-brain barrier dynamics to maintain brain homeostasis," *Trends in Neurosciences*, vol. 44, no. 5, pp. 393–405, 2021.

[3] W. M. Pardridge, "Blood-brain barrier drug targeting: the future of brain drug development," *Molecular Interventions*, vol. 3, no. 2, p. 90, 2003.

[4] S. Bagchi, T. Chhibber, B. Lahooti, A. Verma, V. Borse, and R. D. Jayant, "In-vitro blood-brain barrier models for drug screening and permeation studies: an overview," *Drug Design, Development and Therapy*, vol. 13, p. 3591, 2019.

[5] F. Sivandzade and L. Cucullo, "In-vitro blood-brain barrier modeling: A review of modern and fast-advancing technologies," *Journal of Cerebral Blood Flow & Metabolism*, vol. 38, no. 10, pp. 1667–1681, 2018.

[6] A. M. Palmer and M. S. Alavijeh, "Overview of experimental models of the blood-brain barrier in CNS drug discovery," *Current Protocols in Pharmacology*, vol. 62, no. 1, pp. 7–15, 2013.

[7] J. Bicker, G. Alves, A. Fortuna, and A. Falcão, "Blood-brain barrier models and their relevance for a successful development of CNS drug delivery systems: a review," *European Journal of Pharmaceutics and Biopharmaceutics*, vol. 87, no. 3, pp. 409–432, 2014.

[8] C. Åberg, "Quantitative analysis of nanoparticle transport through in vitro blood-brain barrier models," *Tissue Barriers*, vol. 4, no. 1, p. e1143545, 2016.

[9] M. Singh, R. Divakaran, L. S. K. Konda, and R. Kristam, "A classification model for blood brain barrier penetration," *Journal of Molecular Graphics and Modelling*, vol. 96, p. 107516, 2020.

[10] L. Liu, L. Zhang, H. Feng, S. Li, M. Liu, J. Zhao, and H. Liu, "Prediction of the blood-brain barrier (BBB) permeability of chemicals based on machine-learning and ensemble methods," *Chemical Research in Toxicology*, vol. 34, no. 6, pp. 1456–1467, 2021.

[11] Y. Ding, X. Jiang, and Y. Kim, "Relational graph convolutional networks for predicting blood-brain barrier penetration of drug molecules," *Bioinformatics*, vol. 38, no. 10, pp. 2826–2831, 2022.

[12] T. T. Wager, X. Hou, P. R. Verhoest, and A. Villalobos, "Moving beyond rules: the development of a central nervous system multiparameter optimization (CNS MPO) approach to enable alignment of druglike properties," *ACS Chemical Neuroscience*, vol. 1, no. 6, pp. 435–449, 2010.

[13] A. K. Ghose, G. R. Ott, and R. L. Hudkins, "Technically extended multiparameter optimization (TEMPO): an advanced robust scoring scheme to calculate central nervous system druggability and monitor lead optimization," *ACS Chemical Neuroscience*, vol. 8, no. 1, pp. 147–154, 2017.

[14] M. Gupta, H. J. Lee, C. J. Barden, and D. F. Weaver, "The blood-brain barrier (BBB) score," *Journal of Medicinal Chemistry*, vol. 62, no. 21, pp. 9824–9836, 2019.

[15] U. Norinder, P. Sjöberg, and T. Österberg, "Theoretical calculation and prediction of brain–blood partitioning of organic solutes using MolSurf parametrization and PLS statistics," *Journal of Pharmaceutical Sciences*, vol. 87, no. 8, pp. 952–959, 1998.

[16] 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*, vol. 36, no. 9, pp. 719–730, 2001.

[17] 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*, vol. 17, no. 2, p. 81, 2019.

[18] V. Svetnik, A. Liaw, C. Tong, J. C. Culberson, R. P. Sheridan, and B. P. Feuston, "Random forest: a classification and regression tool for compound classification and QSAR modeling," *Journal of Chemical Information and Computer Sciences*, vol. 43, no. 6, pp. 1947–1958, 2003.

[19] 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*, vol. 52, no. 6, pp. 1686–1697, 2012.

[20] 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*, vol. 75, pp. 347–354, 2017.

[21] 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," *The AAPS Journal*, vol. 20, no. 3, pp. 1–10, 2018.

[22] B. Shaker, M.-S. Yu, J. S. Song, S. Ahn, J. Y. Ryu, K.-S. Oh, and D. Na, "LightBBB: computational prediction model of blood-brain-barrier penetration based on LightGBM," *Bioinformatics*, vol. 37, no. 8, pp. 1135–1139, 2021.

[23] Z. Shi, Y. Chu, Y. Zhang, Y. Wang, and D.-Q. Wei, "Prediction of blood-brain barrier permeability of compounds by fusing resampling strategies and extreme gradient boosting," *IEEE Access*, vol. 9, pp. 9557–9566, 2020.

[24] M. S. Ramakrishnan and N. Ganapathy, "Extreme gradient boosting based improved classification of blood-brain-barrier drugs," *Studies in Health Technology and Informatics*, vol. 294, pp. 872–873, 2022.

[25] P. Garg and J. Verma, "*In silico* prediction of blood brain barrier permeability: an artificial neural network model," *Journal of Chemical Information and Modeling*, vol. 46, no. 1, pp. 289–297, 2006.

[26] A. Guerra, J. A. Páez, and N. E. Campillo, "Artificial neural networks in ADMET modeling: prediction of blood-brain barrier permeation," *QSAR & Combinatorial Science*, vol. 27, no. 5, pp. 586–594, 2008.

[27] Z. Wu, Z. Xian, W. Ma, Q. Liu, X. Huang, B. Xiong, S. He, and W. Zhang, "Artificial neural network approach for predicting blood brain barrier permeability based on a group contribution method," *Computer Methods and Programs in Biomedicine*, vol. 200, p. 105943, 2021.

[28] S. Alsenan, I. Al-Turaiki, and A. Hafez, "A recurrent neural network model to predict blood-brain barrier permeability," *Computational Biology and Chemistry*, vol. 89, p. 107377, 2020.

[29] S. C. Parakkal, R. Datta, and D. Das, "DeepBBBP: High accuracy blood-brain-barrier permeability prediction with a mixed deep learning model," *Molecular Informatics*, p. 2100315, 2022.

[30] D. Rogers and M. Hahn, "Extended-connectivity fingerprints," *Journal of Chemical Information and Modeling*, vol. 50, no. 5, pp. 742–754, 2010.

[31] G. Landrum *et al.*, "RDKit: Open-source cheminformatics software (release 2022.03.2)," 2022. [Online]. Available: http://www.rdkit.org

[32] S. Jaeger, S. Fulle, and S. Turk, "Mol2vec: unsupervised machine learning approach with chemical intuition," *Journal of Chemical Information and Modeling*, vol. 58, no. 1, pp. 27–35, 2018.

[33] 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*, vol. 8, no. 1, pp. 1–11, 2021.

[34] D. Fourches, E. Muratov, and A. Tropsha, "Trust, but verify: on the importance of chemical structure curation in cheminformatics and QSAR modeling research," *Journal of Chemical Information and Modeling*, vol. 50, no. 7, p. 1189, 2010.

[35] L. Nguyen, T.-H. Nguyen-Vo, Q. H. Trinh, B. H. Nguyen, P.-U. Nguyen-Hoang, L. Le, and B. P. Nguyen, "iANP-EC: identifying anticancer natural products using ensemble learning incorporated with evolutionary computation," *Journal of Chemical Information and Modeling*, vol. 62, no. 21, pp. 5080–5089, 2022.

[36] T. Vinh, L. Nguyen, Q. H. Trinh, T.-H. Nguyen-Vo, and B. P. Nguyen, "Predicting cardiotoxicity of molecules using attention-based graph neural networks," *Journal of Chemical Information and Modeling*, vol. 64, no. 6, pp. 1816–1827, 2024.

[37] T.-H. Nguyen-Vo, L. Nguyen, N. Do, P. H. Le, T.-N. Nguyen, B. P. Nguyen, and L. Le, "Predicting drug-induced liver injury using convolutional neural network and molecular fingerprint-embedded features," *ACS Omega*, vol. 5, no. 39, pp. 25 432–25 439, 2020.

[38] T.-H. Nguyen-Vo, Q. H. Trinh, L. Nguyen, P.-U. Nguyen-Hoang, T.-N. Nguyen, D. T. Nguyen, B. P. Nguyen, and L. Le, "iCYP-MFE: identifying human cytochrome P450 inhibitors using multitask learning and molecular fingerprint-embedded encoding," *Journal of Chemical Information and Modeling*, vol. 62, no. 21, pp. 5059–5068, 2021.

[39] S. Shibayama, G. Marcou, D. Horvath, I. I. Baskin, K. Funatsu, and A. Varnek, "Application of the mol2vec technology to large-size data visualization and analysis," *Molecular Informatics*, vol. 39, no. 6, p. 1900170, 2020.

[40] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," *arXiv*, 2014.

[41] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, 1997.

[42] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of Statistics*, pp. 1189–1232, 2001.

[43] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[44] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794.

[45] P. Li, Y. Li, C.-Y. Hsieh, S. Zhang, X. Liu, H. Liu, S. Song, and X. Yao, "TrimNet: learning molecular representation from triplet messages for biomedicine," *Briefings in Bioinformatics*, vol. 22, no. 4, p. bbaa266, 2021.

[46] J. Baek, M. Kang, and S. J. Hwang, "Accurate learning of graph representations with graph multiset pooling," *arXiv*, 2021.

[47] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv*, 2016.

[48] I. Cortés-Ciriano, G. J. P. V. Westen, G. Bouvier, M. Nilges, J. P. Overington, A. Bender, and T. E. Malliavin, "Improved large-scale prediction of growth inhibition patterns using the NCI60 cancer cell line panel," *Bioinformatics*, vol. 32, no. 1, pp. 85–95, 2016.

[49] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biological Cybernetics*, vol. 43, no. 1, pp. 59–69, 1982.

[50] L. Cai, B. P. Nguyen, C.-K. Chui, and S.-H. Ong, "A two-level clustering approach for multidimensional transfer function specification in volume visualization," *The Visual Computer*, vol. 33, no. 1, pp. 163–167, 2015.