



A merged molecular representation deep learning method for blood–brain barrier permeability prediction

Qiang Tang, Fulei Nie, Qi Zhao  and Wei Chen 

Corresponding authors: Qi Zhao, School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan 114051, China. E-mail: zhaoshi@lnu.edu.cn; Wei Chen, Innovative Institute of Chinese Medicine and Pharmacy, Chengdu University of Traditional Chinese Medicine, Chengdu 611137, China. E-mail: greatchen@ncst.edu.cn

Abstract

The ability of a compound to permeate across the blood–brain barrier (BBB) is a significant factor for central nervous system drug development. Thus, for speeding up the drug discovery process, it is crucial to perform high-throughput screenings to predict the BBB permeability of the candidate compounds. Although experimental methods are capable of determining BBB permeability, they are still cost-ineffective and time-consuming. To complement the shortcomings of existing methods, we present a deep learning-based multi-model framework model, called Deep-B³, to predict the BBB permeability of candidate compounds. In Deep-B³, the samples are encoded in three kinds of features, namely molecular descriptors and fingerprints, molecular graph and simplified molecular input line entry system (SMILES) text notation. The pre-trained models were built to extract latent features from the molecular graph and SMILES. These features depicted the compounds in terms of tabular data, image and text, respectively. The validation results yielded from the independent dataset demonstrated that the performance of Deep-B³ is superior to that of the state-of-the-art models. Hence, Deep-B³ holds the potential to become a useful tool for drug development. A freely available online web-server for Deep-B³ was established at <http://cbcb.cdutcm.edu.cn/deepb3/>, and the source code and dataset of Deep-B³ are available at <https://github.com/GreatChenLab/Deep-B3>.

Keywords: blood–brain barrier, molecular descriptors and fingerprints, molecular graph, SMILES, deep learning

Introduction

The blood–brain barrier (BBB) is a multicellular vascular structure that separates the central nervous system (CNS) from the peripheral blood circulation [1]. By tightly controlling the transport of molecules and ions [2], BBB not only maintains brain homeostasis but also protects the brain from intrusive xenobiotic chemical compounds [2, 3]. Hence, BBB plays important roles in maintaining the stability of the physiological environment of brain tissues and protecting the CNS from infraction by harmful agents or microorganisms in the blood. On the other hand, the ineffective transportation of molecules across the BBB is also a big challenge to drug design for CNS diseases [4–6]. It was reported that 98% of the small-molecule drugs and almost all macromolecular drugs could be blocked by BBB [7, 8]. Thus, the assessment of BBB permeability is a prerequisite of drug discovery and development for CNS diseases.

Over the past decade, numerous *in vivo* and *in vitro* assay methods have been proposed to evaluate the BBB permeability of chemicals [9–13]. For the *in vivo* methods, logBB and logPS are two gold-standard measures for analyzing the BBB penetration of drugs [10–13]. However, the *in vivo* methods are animal-based, labor-intensive, low throughput and expensive. The *in vitro* methods

can be divided into cell-based and non-cell-based assays [11, 13]. Functional expression and metabolic capacity of BBB transporters are the main advantages of the cell-based approaches [14]. By using this method, not only passive diffusion can be measured but also the active transport of the molecule through BBB can be observed. However, there are also some limitations, such as high cost and the differences of BBB between animal species and humans. The representative technology of non-cell-based permeability assays is the parallel artificial membrane permeability assay (PAMPA) [15]. By modifying the lipid composition of the artificial membranes, PAMPA has the ability to predict BBB permeability [15, 16]. However, the non-cell-based permeability method have not been proved suitable for the medium-to-high-throughput operation [14–16].

Given that the above-mentioned experimental techniques are time-consuming and cost-ineffective, a number of computational methods have been proposed to characterize the BBB permeability as well [17–21]. These methods can be mainly grouped into two categories, namely quantitative methods and qualitative methods. The quantitative methods aim to predict the properties describing the BBB permeability of compounds [22, 23], such as logBB and logPS [24, 25]. The qualitative methods are machine

Qiang Tang is a PhD candidate at the School of Basic Medical Sciences, Chengdu University of Traditional Chinese Medicine. His research interests include bioinformatics and machine learning.

Fulei Nie is a PhD candidate at the School of Public Health, North China University of Science and Technology. Her research interests include bioinformatics and machine learning.

Qi Zhao is a professor at the School of Computer Science and Software Engineering, University of Science and Technology Liaoning. His research interests include bioinformatics, complex network and machine learning.

Wei Chen is a professor at the Innovative Institute of Chinese Medicine and Pharmacy, Chengdu University of Traditional Chinese Medicine. His research interests include bioinformatics and machine learning.

Received: June 4, 2022. Revised: July 27, 2022. Accepted: July 30, 2022

© The Author(s) 2022. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com

learning based methods that classify whether a given compound is BBB permeable (BBB+) or not (BBB-). However, the performance of qualitative models for characterizing the BBB permeability of compounds depends on the selected representations of the molecule.

The most commonly used molecule representations were molecular descriptors and fingerprints [26]. Molecular descriptors describe molecules based on experimental or theoretical properties. According to the level of molecular representation required for calculating the descriptor, the molecular descriptors can be classified into three categories, namely one-dimensional (1D), two-dimensional (2D) or three-dimensional (3D) [26]. Molecular fingerprints are fixed-length bit strings that encode the structural, topological or functional features of a molecule. Different types of fingerprints differed in the form that a molecule is decomposed, such as the size of the representation and the hashing algorithm [27, 28].

In the past decade, a series of machine learning methods have been proposed to qualitatively predict BBB permeability based on molecular descriptors and fingerprints representations. In 2005, Li *et al.* trained an SVM model based on 415 molecules and reported an accuracy (ACC) of 0.837 [17]. Later on, Martins *et al.* established a series of random forest and SVM models based on a dataset of 1970 compounds and obtained the best ACC of 0.885 in 5-fold cross-validation test and 0.947 in the external validation test [18]. To improve predictive ACC, Wang *et al.* constructed an ensemble model with 1437 2D molecular descriptors and six types of molecular fingerprints [19]. Their model yielded an ACC of 0.975 and 0.907 in the training and test sets, respectively [19]. In 2013, Yan *et al.* trained several models by using multilinear regression, SVM and artificial neural network methods to predict the logBB value of 320 compounds [20]. By encoding the compounds using the dragon software [29], Na *et al.* proposed a lightGBM-based model and obtained an ACC of 0.90 for qualitatively predicting 74 independent compounds [30]. Recently, Liu *et al.* proposed an ensemble-learning model for predicting BBB permeability in which the samples were encoded by using nine molecular fingerprints and obtained an ACC of 0.78 in external validation [21]. It was found that the above-mentioned studies were only built based on molecular descriptors or fingerprints to predict the BBB permeability. None of them combined the features from different modalities which cover the chemical diversities of compounds.

Simplified molecular input line entry system (SMILES) has been widely used as a standard representation of compounds and can represent the functional substructures and express the structural differences in a richer feature space [31]. The success of deep neural networks in natural language processing (NLP) makes the SMILES-based representations become a research focus of cheminformatics [32, 33]. Since the SMILES is sequential and composed of text, the pre-trained models, such as Transformer and BERT, have been used to extract the latent features from SMILES [34–36].

Inspired by the remarkable achievements of deep learning, the pre-training strategy was also used to deduce the graph representations of molecules [37–39]. Based on the molecular graphs, convolutional neural network (CNN) and transfer learning method were proposed for virtual screening [40]. It was found that CNN could automatically obtain features from the graphs and can avoid the extraction of hand-crafted features.

Hence, in the present study, a deep learning-based multi-model framework, called Deep-B³, was proposed to predict BBB permeability. In Deep-B³, the compounds were encoded by using molecular descriptors and fingerprints, molecular graph and SMILES text

notation. For balancing the number of samples between the positive and negative datasets, each compound in the negative dataset was amplified by enumerating its different SMILES formats. When tested on the independent dataset, it was found that the Deep-B³ outperforms the state-of-the-art models for predicting BBB permeability.

Materials and Methods

Data collection and preparation

In this study, the training dataset, including compounds with experimentally confirmed BBB permeability, was built by integrating the data from recent studies and online repositories [18, 21, 30, 41–43]. After representing them by SMILES, the collected compounds were curated to eliminate duplicates, inorganic material and mixtures. Accordingly, a dataset containing 7224 compounds was obtained in which 5483 were positive samples (BBB+) and 1741 were negative ones (BBB-). Both lightBBB [30] and Liu *et al.*'s method [21] were built on the subset of this dataset.

The independent dataset used for testing the proposed method was retrieved from previous publications [21, 44–46]. Among them, 213 molecules (154 BBB+ and 59 BBB-) were constructed by Liu *et al.* [21], 31 traditional Chinese medicine molecules (13 BBB- and 18 BBB+) were experimentally verified by Ai *et al.* [44], 1 BBB- molecule (hexamethonium bromide) was from Shaw *et al.*'s work [45] and the rest 7807 molecules were collected from the B3DB [46].

In most cases, one drug molecule can be represented with multiple SMILES strings. Accordingly, many duplicated drugs were represented by different SMILES, which will lead to the redundancy of the training and testing dataset. Canonicalization is a method to ensure that each molecule corresponds to a unique canonical SMILES. Hence, all SMILES strings were converted into canonical SMILES to remove duplicated compounds in the training and testing datasets, respectively. Finally, we obtained a non-redundant training dataset, including 4364 molecules with 3125 BBB+ and 1239 BBB-, and a testing dataset, including 2670 molecules with 1258 BBB+ and 1412 BBB-.

Multiple SMILES-based augmentation

If a machine learning model is trained on an imbalanced dataset, its performance will be negatively affected. Therefore, an augmentation approach was used to balance the positive and negative training samples. The augmentation was implemented by renumbering atoms using 'MolToSmiles' in RDKit (<https://www.rdkit.org/>) with the parameter 'canonical=False'. For a compound in BBB- dataset with multiple SMILES formats, two more SMILES strings were randomly extracted and selected as the negative samples. After performing this procedure, a balanced training dataset containing 3125 positive and 3692 negative samples was built. The training dataset was randomly divided into two parts with a ratio of 8:2, which were used to fit the model and validate its performance while tuning model hyperparameters, respectively. The detail information about the training and testing dataset was presented in Table 1. The datasets are available at <http://cbcb.cdutcm.edu.cn/deepb3Static/data/data.tar.gz>.

Feature representations for molecules

The samples in the dataset were encoded by using three kinds of features, namely tabular data, text and image. By using RDKit, the 208 1D/2D molecular descriptors, 167 bits MACCS fingerprints [47] and 1024 bits Morgan fingerprints [48] were obtained for each molecule, which were then transferred into tabular data. The

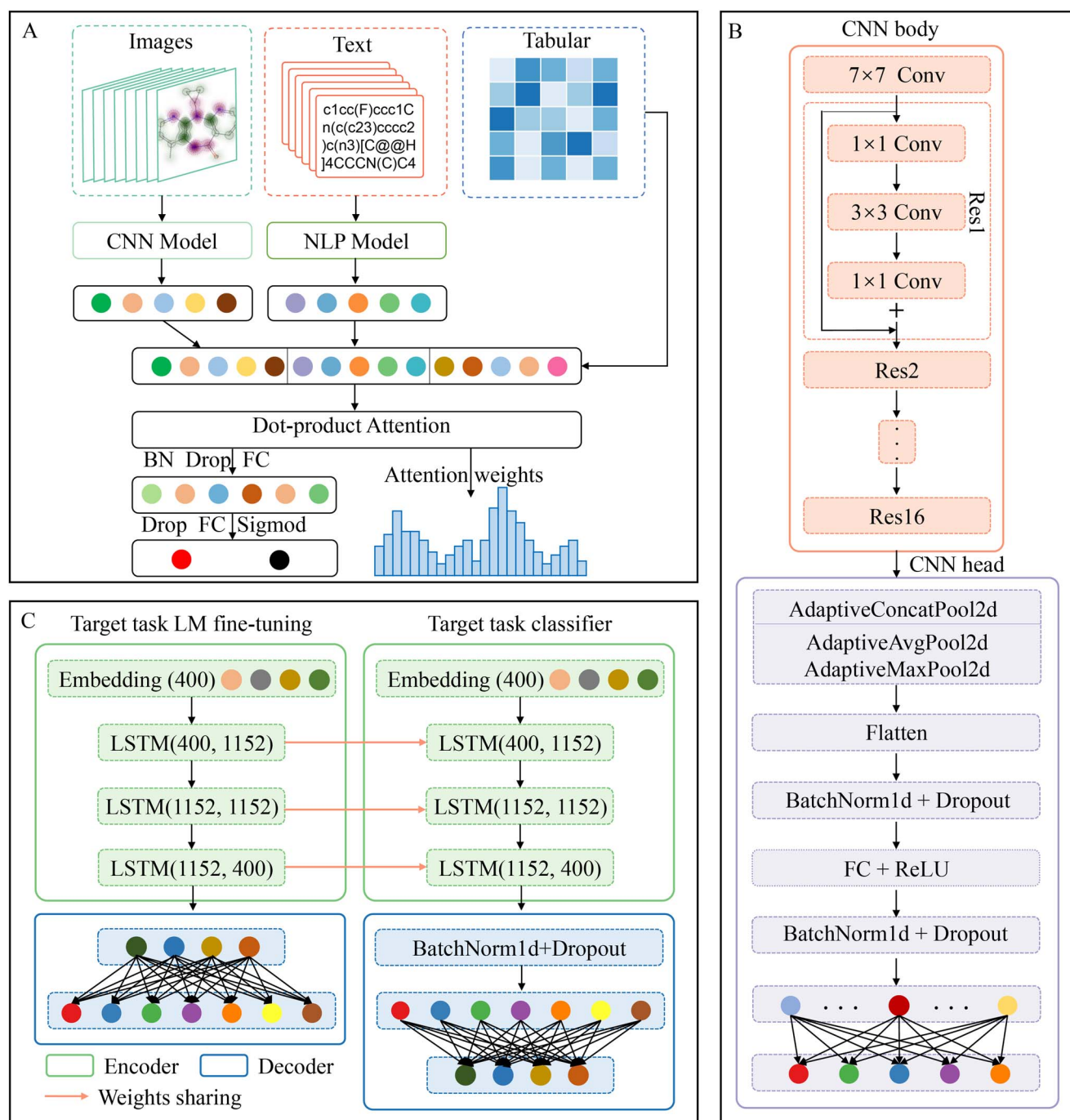


Figure 1. An illustration for the framework of Deep-B³. (A) The workflow of the Deep-B³ model. (B) The diagram of the CNN model. (C) The diagram of the NLP model.

canonical SMILES string of each molecule was used as the text feature. Based on the SMILES string, by running the rdkit.Chem.Draw package in RDKit, the molecular graph of each molecule was obtained and regarded as the image feature. The images with smaller or bigger sizes will result in low ACC or more memory usage [49, 50]. In order to standardize the distance scale, the images were resized to 224 × 224 pixels which has been widely used in the field of image analysis [51–53].

Performance evaluation strategies

The proposed model was evaluated by using both threshold-dependent and -independent metrics. The four threshold

dependent metrics [54–56], namely sensitivity (SN), specificity (SP), ACC and Matthews correlation coefficient (MCC) are defined in Eq. (1).

$$\left\{ \begin{array}{l} \text{SN} = \frac{\text{TP}}{\text{TP} + \text{FN}} \\ \text{SP} = \frac{\text{TN}}{\text{TN} + \text{FP}} \\ \text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \\ \text{MCC} = \frac{\text{TP} * \text{TN} - \text{FP} * \text{FN}}{\sqrt{(\text{TN} + \text{FP})(\text{TN} + \text{FN})(\text{TP} + \text{FP})(\text{TP} + \text{FN})}} \end{array} \right. \quad (1)$$

where TP, TN, FP and FN represent the number of true positive, true negative, false positive and false negative, respectively. The threshold independent metric is the area under the receiver operating characteristic curve (AUC) [57]. The value of AUC ranges

from 0.5 to 1. The higher the value of the AUC, the better performance is implied.

Result and Discussion

Architecture of the Deep-B³

By integrating the three kinds of features, namely molecular descriptors and fingerprints, molecule graph and SMILES string, a deep learning-based multi-model framework, called Deep-B³, was developed to predict the BBB permeability. The schematic diagram of Deep-B³ is depicted in Figure 1A. The Fast.AI 1.0 (<https://fastai1.fast.ai/>) was applied to build the frameworks. The pre-trained CNN model was used to extract features from the molecule graphs. A pre-trained NLP model was used to extract features from SMILES. Subsequently, the features obtained from the CNN and NLP were concatenated with the tabular features generated based on molecular descriptors and fingerprints. A self-attention layer monitored the key features for predicting BBB permeability and fed the results into a series of fully connected (FC) layer to make final predictions.

The ResNet-50 CNN [58] with 50 layers deep was used to build up the framework to automatically extract the features from molecule graph data. The architecture of the modified ResNet-50 model is shown in Figure 1B. The model begins with a convolution (Conv) with a kernel size of 7*7, followed by 16 ResNet blocks (Res). In each Res, there are three sequential Conv with the kernel size of 1*1, 3*3 and 1*1, respectively. Then, the output was processed by using the AdaptiveConcatPool2d, followed by the batch normalization layer (BN), dropout and FC layers. Finally, the latent features were obtained from the molecule graph data.

The language model (LM) Average Stochastic Gradient Descent Weight-Dropped Long Short-Term Memory (AWD-LSTM) [59] was used to extract features from the SMILES text notations. As shown in Figure 1C, the model contains two blocks, namely target task LM fine-tuning classifier and target task classifier. Each block consists of an encoder layer and a decoder layer. The encoder layer includes three LSTM that can learn a compact representation from the input. The three LSTM in decoder layer share the same weights with that of encoder layer. To achieve the latent features

Table 1. Detail information about the dataset used in this study

Sample	Raw training	Training	Testing
BBB+	3125	3125	1258
BBB-	1239	3692	1412

from SMILES, the decoder layer in the target task classifier block was replaced by a BN layer followed by a FC layer.

Model training and parameter optimization

For training a deep learning model, the most important hyperparameter is the learning rate. The layer-specific learning rate method was applied to overcome the slow learning. First, the proposed model was divided into seven layers (Table S1 available online at <http://bib.oxfordjournals.org/>), and then the best learning rate of each layer was determined by using the lr_find method. With the gradual unfreezing, the specific learning rate was used to tune the weights for each layer. To prevent overtraining, the early stopping technique was used to stop training when the valid loss does not decrease within five epochs.

The feature dimension of the CNN and NLP model also affects the performance of Deep-B³. In order to obtain the best combination, we optimized the dimensions of the output features of CNN and NLP. At first, we set the dimension of the output features from the NLP to 64 and then tested the models when the dimension of the output features from CNN were 64, 128, 256, 512 and 1024, respectively. The performances of those models on the testing dataset are shown in Figure 2 and Table S2 available online at <http://bib.oxfordjournals.org/>. Although the best AUC were observed for the models with the CNN output feature dimension of 512 and 128, the values of the metrics defined in Eq. (1) of the former model were higher. Hence, we set the dimension of the output features from the CNN model to 512 and tested the models when the dimension of the output features from NLP were 1024, 512, 256 and 128, respectively. However, the overall performances were not improved at all (Figure 2 and Table S2 available online at <http://bib.oxfordjournals.org/>). Accordingly, the best model was built based on the 512 and 64 output features from CNN and NLP models, respectively.

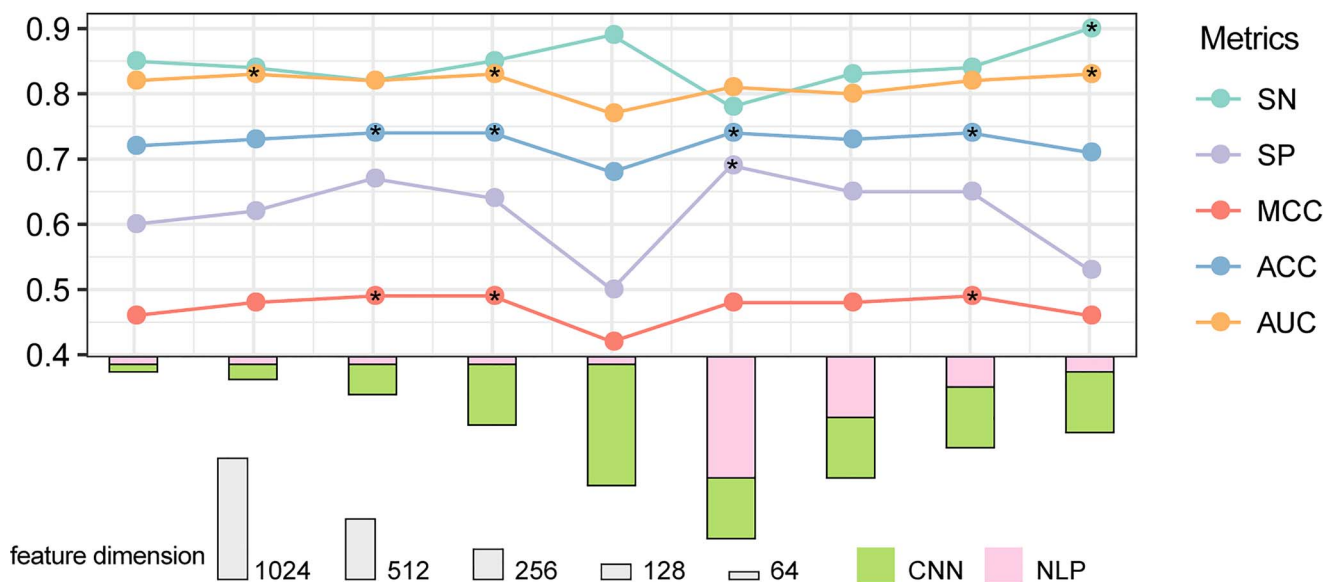


Figure 2. The performance of the models by combining the different dimensions of features output from CNN and NLP. The symbol * indicated the highest value.

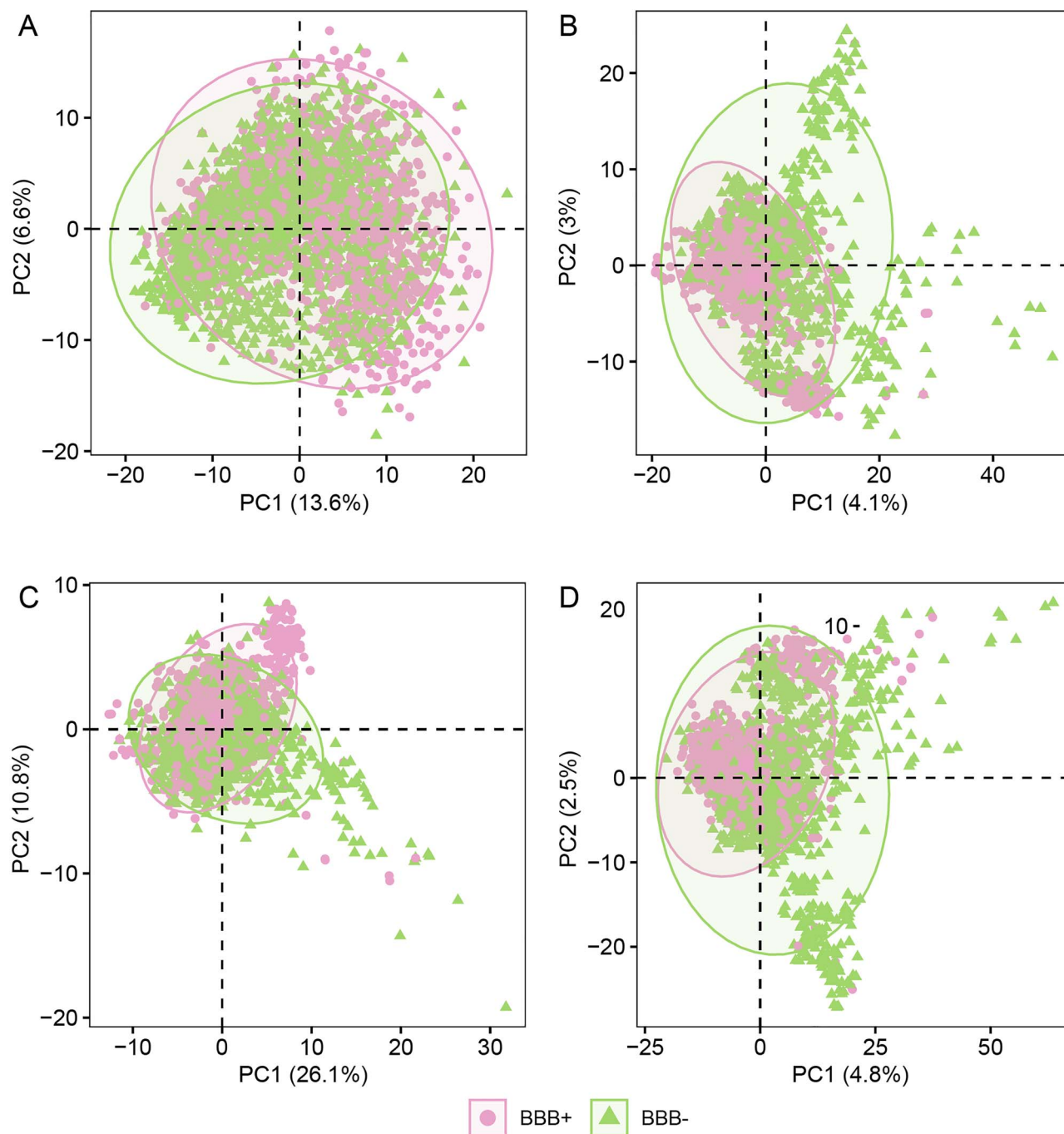


Figure 3. The PCA score plot of the first two PCs. (A) The PCA of the image features generated by the CNN model. (B) The PCA of the tabular features. (C) The PCA of the SMILES features extracted by the NLP model. (D) The PCA of the combined features.

Model evaluation and comparison with baseline methods

The proposed Deep-B³ method was then tested on the independent dataset, and we obtained an AUC of 0.83 with the SN of 85%, SP of 64% and ACC of 0.74 (Table 2). To further demonstrate its superiority, we also compared the performance of Deep-B³ with the baseline method lightBBB [30]. Their performances for predicting the BBB permeability are listed in Table 2. It was found that Deep-B³ outperforms lightBBB in terms of most of the metrics defined in Eq. (1). The MCC and ACC are 6% and 4% higher than that of lightBBB, respectively. Although the SN of Deep-B³ is 1% lower than that of lightBBB, its SN was 9% higher. This result

Table 2. Comparison of Deep-B³ with baseline methods for predicting BBB permeability

Model	SN (%)	SP (%)	MCC	ACC	AUC
lightBBB	86	55	0.43	0.70	–
Deep-B ³	85	64	0.49	0.74	0.83
Deep-B ³ (Raw)	95	41	0.41	0.66	0.80

Note: Raw means by using the raw SMILES without augmentations.

indicates that Deep-B³ is an excellent tool for predicting BBB permeability and holds the potential to become a useful tool of drug development.

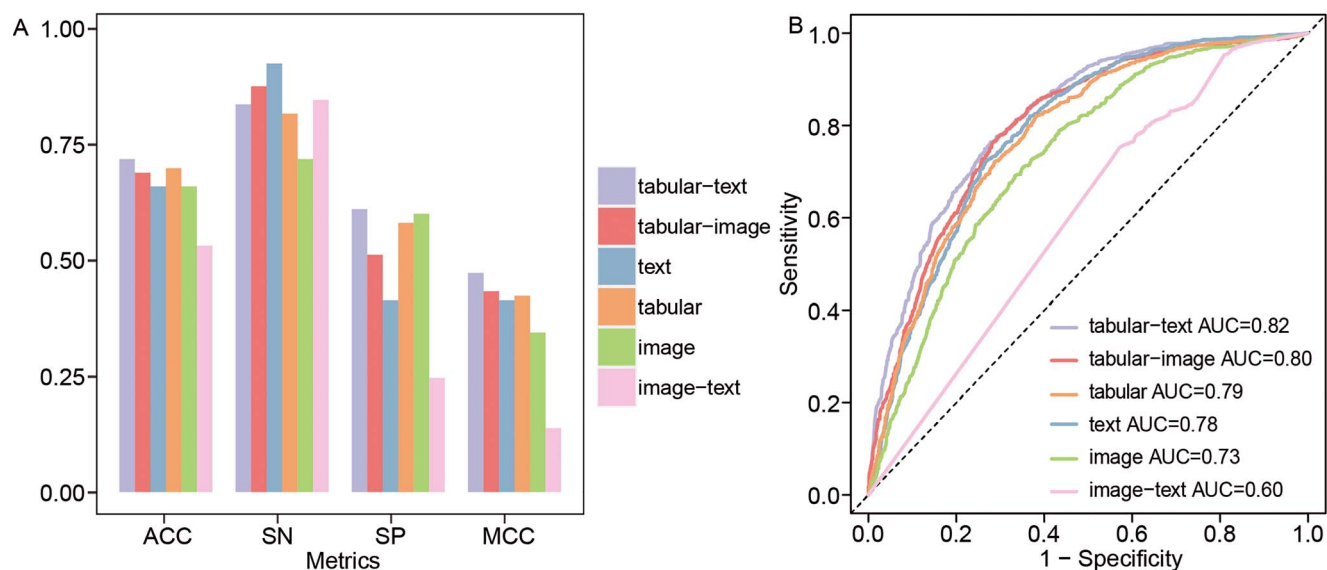


Figure 4. The performance of the models built based on different kinds of features for predicting the BBB permeability. (A) The values of the threshold dependent and independent metrics obtained from different models. (B) The ROC curves obtained from different models.

The multiple SMILES-based augmentation strategy has been used to improve the performance of molecular property prediction [36, 60]. To validate whether this strategy can also improve the performance of BBB permeability prediction, we compared the performance of Deep-B³ built based on the augmentation dataset and raw dataset. As indicated in Table 2, the model based on the augmented dataset outperforms that on the raw dataset. Compared with the model built based on raw dataset, the 8% and 3% improvements in terms of ACC and AUC were observed for the model based on the augmented dataset. Although the augmented SMILES share the same molecular descriptors and fingerprints with the raw SMILES, they differed in terms of molecule graph and SMILES string (Figure S1 available online at <http://bib.oxfordjournals.org/>). These results demonstrated that the SMILES-based augmentation strategy can improve the performance of Deep-B³ for BBB permeability prediction.

Feature contribution analysis

In order to interpret the latent features generated by the CNN and NLP model, the principal component analysis (PCA) method was used to reduce the feature dimensionality in an interpretable way. In the PCA plot, features contributing similar information or with high correlations will be clustered together. The PCA results are shown in Figure 3.

As shown in Figure 3A, the first two principal components (PCs) of the image features output from the CNN model explains 20.2% of the variation between BBB+ and BBB-. Therefore, if only using the molecular graph, it could not satisfactorily identify the BBB+ residing in the overlapping region. The first two PCs of the tabular, text and concatenating features explained 7.1%, 36.9% and 8.3% of the overall variation of between BBB+ and BBB- (Figure 3B-D), respectively. Although some BBB+ reside in the clusters covered by BBB-, a specific cluster of BBB+ is still noticeable in the PCA results. The PCA results indicate that the three kinds of features may contribute differently for distinguishing BBB+ from the BBB- (Figure 3).

In order to further demonstrate the contribution of tabular, image and text features to the Deep-B³, we compared the performances of the models based on different combinations of

these features in the independent dataset. The related results are shown in Figure 4. Among the models based on the single feature, although the model based on the image obtained poor performance than others, it has the highest SP value (Figure 4A, Table S3 available online at <http://bib.oxfordjournals.org/>). For the other two models, the model constructed based on text features had the highest SN, and the model on tabular features obtained the best ACC and AUC. For the models based on the image feature combined with either tabular or text yielded the higher SN than that based on the tabular-text feature. The model based on tabular-text feature exhibits the superior performance in terms of SP, MCC, ACC and AUC. However, none of these models is superior to Deep-B³. This result indicates that tabular, image and text features all contributed positively to the prediction of BBB permeability and that the tabular feature is most essential.

Model interpretation based on attention weights

For each molecule, the feature importance was determined according to its attention weight, which was calculated by taking the average of the attention values in a given column of the attention matrix. The attention weights were then visualized to investigate the impact of different kinds of features on predicting the BBB permeability. For example, the attention weight for recognin is shown in Figure 5. It was found that the bit-397 substructure from the Morgan fingerprints exhibits the highest weight in all features. In addition, the high attention weights were also observed for the image and text features, indicating that they were important for BBB permeability prediction as well.

The attention weights were also used to monitor the key features for BBB permeability prediction. The feature importance was defined by the average of the attention weight of all molecules. In Deep-B³, 1975 features were fed into the attention layer, including 512 features from the CNN model, 64 features from the NLP mode and 1399 tabular features.

The top 20% (395) features are tabular features, including 338-bits Morgan fingerprints and 14-bits MACCS fingerprints. Their distribution differences between BBB+ and BBB- were assessed

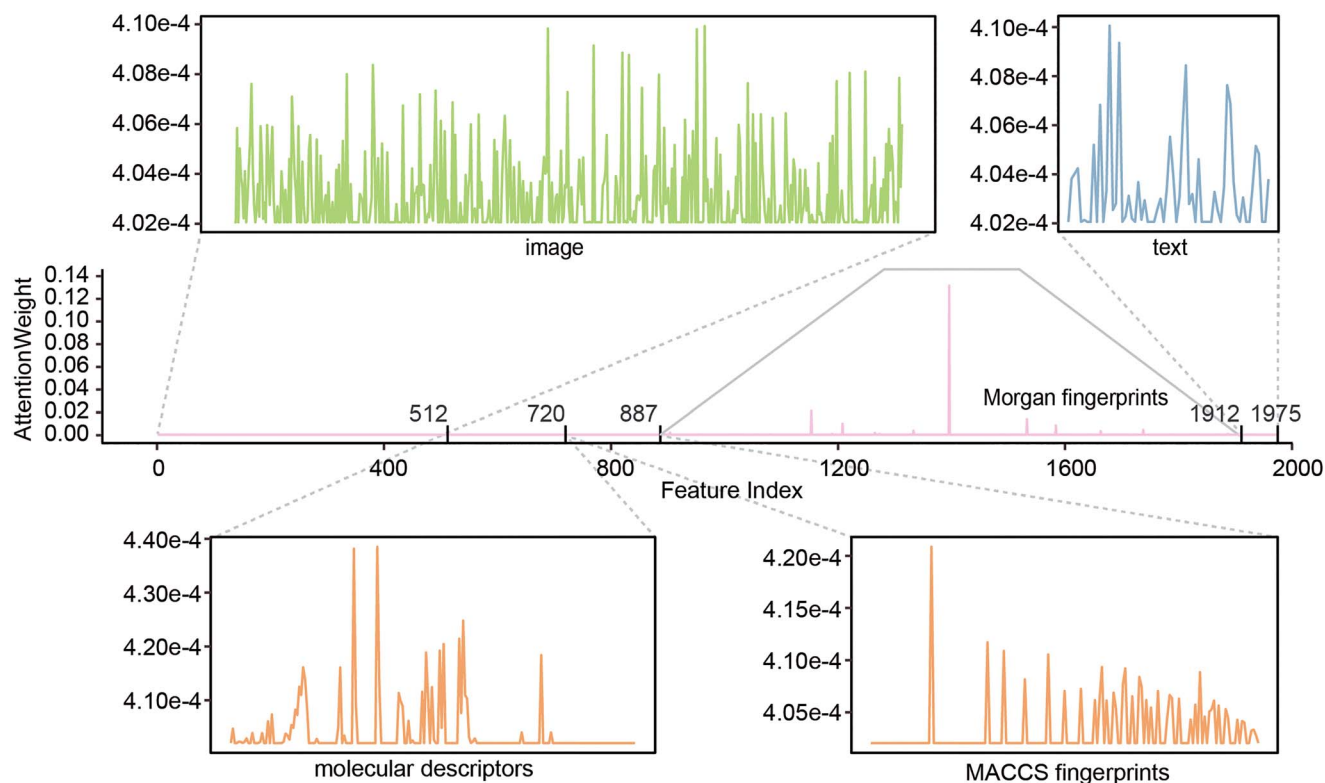


Figure 5. Visualization of attention weights. An example of attention weight visualization for the molecule recognin.

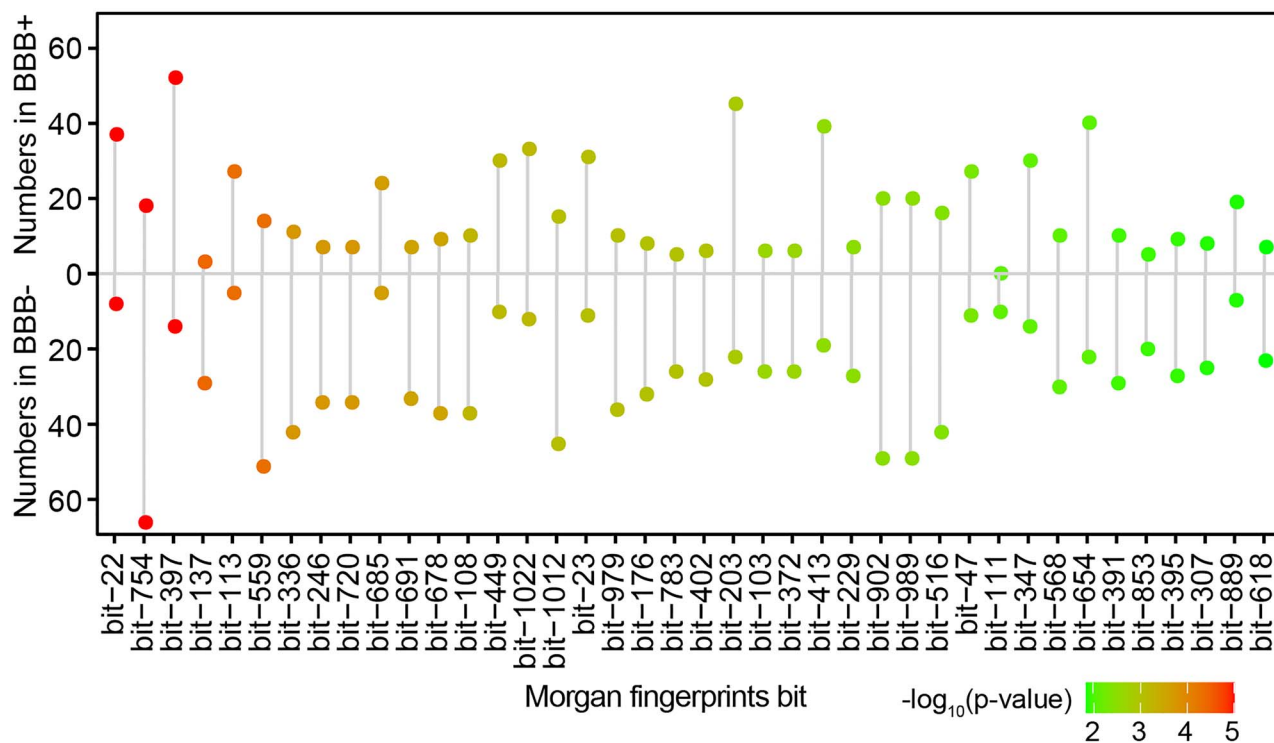


Figure 6. Distribution of Morgan fingerprints substructures in BBB+ and BBB- samples. The x-axis is the substructural bit, while the y-axis is the count numbers of the bit in the independent testing dataset. The color dot indicates the $-\log_{10}P$ -value.

by using chi-square test ($P < 0.05$). The top 40 substructures with significantly distinct distributions are shown in Figure 6, and their appearances are provided in Table S4 available online at <http://bib.oxfordjournals.org/>. For example, the bit-22,

bit-754 and above-mentioned bit-397 substructures differed most significantly between the BBB+ and BBB-. The bit-22 substructure is present in over 2.94% of all BBB+ samples and only in 0.57% of BBB- samples. On the contrary, about 4.67% of BBB- have

Input query molecules

Input molecular SMILES text

Paste your sequences with raw text format below (click [here](#) for example)

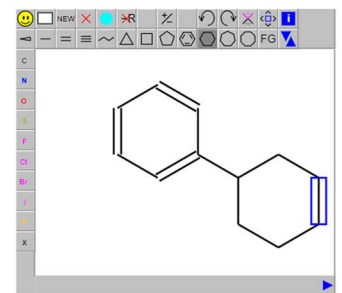
```
CN1CCC[C@@H](c2nc3ccccc3n2Cc2ccc(F)cc2)C1
C1CCCI
CC12NC(Cc3ccccc31)c1ccccc12
c1ccccc1
CCCCCCCC
```

Upload a File

Browse no file selected

Graph to SMILES

Draw a molecule using JMSE editor, then you can get the smiles of the molecule



Molecular SMILES generated by JMSE

Canonical SMILES

SMILES:

```
C2(C1CCCC1)=CC=CC=C2
```

Prediction results

Prediction Result

SMILES	Permeable or Not	probability
CN1CCC[C@@H](c2nc3ccccc3n2Cc2ccc(F)cc2)C1	Permeable	0.7311
C1CCCI	Permeable	0.7301
CC12NC(Cc3ccccc31)c1ccccc12	Permeable	0.7311
c1ccccc1	Permeable	0.7309
CCCCCCCC	Permeable	0.7311
C1=CC(=C(C=C1CC(C(=O)O)OC(=O)C=CC2=C3C(C(OC3=C(C=C2)O)C4=CC(=C(C=C4)O)O)C(=O)OC(CC5=CC(=C(C=C5)O)O)C(=O)O)O	Non-Permeable	0.2689

Figure 7. Screenshots for the interface and result page of the Deep-B³ web server.

the bit-754 substructure, while this proportion is only 1.43% for BBB+. These results provide a descriptive summary to indicate the contribution of substructures to BBB permeability prediction.

Web server implementation

For facilitating researchers to implement the proposed Deep-B³ model, an easy-to-use web server was established at <http://cbcb.cdutcm.edu.cn/deepb3/>. The server interface is shown in Figure 7. To obtain the desired results, users only need to paste or upload the SMILES strings of the compounds of interested. If the SMILES string is unknown, it could be obtained by drawing the molecule in the built-in molecule editor tool JSME [61]. Finally, by clicking the Submit button, the results will be shown in a new page.

Conclusion

In this study, we proposed a deep learning-based model, called Deep-B³, to predict the BBB permeability of compounds. The modified ResNet50 and AWD-LSTM model were used to extract latent features from molecules graph and SMILES, respectively. By integrating the latent features with the tabular features and feeding them to an attention layer followed by a series of FCs, the Deep-B³ will determine whether a compound is BBB+ or BBB-.

Experimental results from the independent dataset demonstrated that Deep-B³ outperformed existing baseline models in terms of both threshold-dependent and -independent metrics, indicating its superiority and robustness for predicting BBB permeability. The attention weight was used to quantify the importance of features. The substructural importance analysis revealed the difference of key molecular substructures between BBB+ and

BBB-. The observed substructural differences might contribute to the accurate prediction of BBB permeability.

For the convince of scientific community, the source code of Deep-B³ is provided at <https://github.com/GreatChenLab/Deep-B3>. As the implementation of the proposed Deep-B³, a freely accessible web-server has also been provided at <https://cbcb.cdutcm.edu.cn/deepb3>. We believe that Deep-B³ holds the potential to become a useful tool for drug development.

Key Points

- A multi-model framework, called Deep-B³, was proposed to predict the BBB permeability of compounds.
- Pre-trained models were used to learn the latent features from the molecule graph and SMILES text notations.
- Comparative results demonstrate that Deep-B³ significantly outperforms the baseline methods for BBB permeability prediction.

Data availability

The data and code that support the findings of this study are available at <https://github.com/GreatChenLab/Deep-B3>.

Authors' contributions

W.C. and Q.T. conceived and designed the work. Q.T. performed the data collection and analysis and visualized the results. Q.T. and

F.N. collected the data. Q.T., Q.Z. and W.C. wrote the manuscript. All authors read and approved the final manuscript.

Supplementary data

Supplementary data are available online at <https://academic.oup.com/bib>.

Funding

Natural Science Foundation of Sichuan Province (No. 2022NSFSC 1770), Innovation Team and Talents Cultivation Program of National Administration of Traditional Chinese Medicine (No. ZYYCXTD-D-202209), Foundation of Education Department of Liaoning Province (No. LJKZ0280).

References

- Daneman R, Prat A. The blood-brain barrier. *Cold Spring Harb Perspect Biol* 2015;**7**:a020412.
- Abbott NJ, Patabendige AA, Dolman DE, et al. Structure and function of the blood-brain barrier. *Neurobiol Dis* 2010;**37**:13–25.
- Obermeier B, Daneman R, Ransohoff RM. Development, maintenance and disruption of the blood-brain barrier. *Nat Med* 2013;**19**:1584–96.
- Vilella A, Ruozi B, Belletti D, et al. Endocytosis of nanomedicines: the case of glycopeptide engineered PLGA nanoparticles. *Pharmaceutics* 2015;**7**:74–89.
- Gao H, Pang Z, Jiang X. Targeted delivery of nano-therapeutics for major disorders of the central nervous system. *Pharm Res* 2013;**30**:2485–98.
- Dong X. Current strategies for brain drug delivery. *Theranostics* 2018;**8**:1481–93.
- Chen Y, Liu L. Modern methods for delivery of drugs across the blood-brain barrier. *Adv Drug Deliv Rev* 2012;**64**:640–65.
- Baratta MG. Getting to the brain. *Nat Nanotechnol* 2018;**13**:536.
- Di L, Kerns EH, Bezar IF, et al. Comparison of blood-brain barrier permeability assays: in situ brain perfusion, MDR1-MDCKII and PAMPA-BBB. *J Pharm Sci* 2009;**98**:1980–91.
- Carpenter TS, Kirshner DA, Lau EY, et al. A method to predict blood-brain barrier permeability of drug-like compounds using molecular dynamics simulations. *Biophys J* 2014;**107**:630–41.
- Abbott NJ. Prediction of blood-brain barrier permeation in drug discovery from in vivo, in vitro and in silico models. *Drug Discov Today Technol* 2004;**1**:407–16.
- Ciura K, Dziomba S. Application of separation methods for in vitro prediction of blood-brain barrier permeability—the state of the art. *J Pharm Biomed Anal* 2020;**177**:112891.
- Mensch J, Oyarzabal J, Mackie C, et al. In vivo, in vitro and in silico methods for small molecule transfer across the BBB. *J Pharm Sci* 2009;**98**:4429–68.
- Reichel A, Begley DJ, Abbott NJ. An overview of in vitro techniques for blood-brain barrier studies. *Methods Mol Med* 2003;**89**:307–24.
- Carrara S, Reali V, Misiano P, et al. Evaluation of in vitro brain penetration: optimized PAMPA and MDCKII-MDR1 assay comparison. *Int J Pharm* 2007;**345**:125–33.
- Mensch J, Melis A, Mackie C, et al. Evaluation of various PAMPA models to identify the most discriminating method for the prediction of BBB permeability. *Eur J Pharm Biopharm* 2010;**74**:495–502.
- Li H, Yap CW, Ung CY, et al. Effect of selection of molecular descriptors on the prediction of blood-brain barrier penetrating and nonpenetrating agents by statistical learning methods. *J Chem Inf Model* 2005;**45**:1376–84.
- Martins IF, Teixeira AL, Pinheiro L, et al. A Bayesian approach to in silico blood–brain barrier penetration modeling. *J Chem Inf Model* 2012;**52**:1686–97.
- Wang Z, Yang H, Wu Z, et al. In silico prediction of blood-brain barrier permeability of compounds by machine learning and resampling methods. *ChemMedChem* 2018;**13**:2189–201.
- Yan A, Liang H, Chong Y, et al. In-silico prediction of blood-brain barrier permeability. *SAR QSAR Environ Res* 2013;**24**:61–74.
- Liu L, Zhang L, Feng H, et al. Prediction of the blood-brain barrier (BBB) permeability of chemicals based on machine-learning and ensemble methods. *Chem Res Toxicol* 2021;**34**:1456–67.
- Konovalov DA, Coomans D, Deconinck E, et al. Benchmarking of QSAR models for blood-brain barrier permeation. *J Chem Inf Model* 2007;**47**:1648–56.
- Kim T, You BH, Han S, et al. Quantum artificial neural network approach to derive a highly predictive 3D-QSAR model for blood-brain barrier passage. *Int J Mol Sci* 2021;**22**:10995.
- Pardridge WM. Blood–brain barrier delivery. *Drug Discov Today* 2007;**12**:54–61.
- Allen DD, Smith QR. Characterization of the blood–brain barrier choline transporter using the in situ rat brain perfusion technique. *J Neurochem* 2001;**76**:1032–41.
- Xue L, Bajorath J. Molecular descriptors in chemoinformatics, computational combinatorial chemistry, and virtual screening. *Comb Chem High Throughput Screen* 2000;**3**:363–72.
- Capecchi A, Probst D, Reymond JL. One molecular fingerprint to rule them all: drugs, biomolecules, and the metabolome. *J Chem* 2020;**12**:43.
- Nisius B, Bajorath J. Molecular fingerprint recombination: generating hybrid fingerprints for similarity searching from different fingerprint types. *ChemMedChem* 2009;**4**:1859–63.
- Mauri A, Consonni V, Pavan M, et al. DRAGON software: an easy approach to molecular descriptor calculations. *MATCH / Communications In Mathematical & In Computer Chemistry* 2006;**56**:237–48.
- Shaker B, Yu MS, Song JS, et al. LightBBB: computational prediction model of blood-brain-barrier penetration based on Light-GBM. *Bioinformatics* 2021;**37**:1135–9.
- Weininger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J Chem Inf Comput Sci* 1988;**28**:31–5.
- Ozturk H, Ozgur A, Schwaller P, et al. Exploring chemical space using natural language processing methodologies for drug discovery. *Drug Discov Today* 2020;**25**:689–705.
- Li X, Fourches D. SMILES pair encoding: a data-driven substructure tokenization algorithm for deep learning. *J Chem Inf Model* 2021;**61**:1560–9.
- Kim Y, Zheng S, Tang J, et al. Anticancer drug synergy prediction in understudied tissues using transfer learning. *J Am Med Inform Assoc* 2021;**28**:42–51.
- Honda S, Shi S, Ueda HR. SMILES transformer: pre-trained molecular fingerprint for low data drug discovery. *arXiv* 2019;**1911**:04738.
- Wang, S., Y. Guo, Y. Wang, et al. SMILES-BERT: large scale unsupervised pre-training for molecular property prediction. In: *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*. Niagara Falls, NY, USA: ACM, 2019, 429–36.
- Kearnes S, McCloskey K, Berndl M, et al. Molecular graph convolutions: moving beyond fingerprints. *J Comput Aided Mol Des* 2016;**30**:595–608.

38. Chen D, Gao K, Nguyen DD, et al. Algebraic graph-assisted bidirectional transformers for molecular property prediction. *Nat Commun* 2021;**12**:3521.
39. Liu, S., M.F. Demirel, and Y.J.A.I.N.I.P.S. Liang. N-gram graph: Simple unsupervised representation for graphs, with applications to molecules. In: *Advances in Neural Information Processing Systems*. Vancouver, BC, Canada: Neural Information Processing Systems Foundation, 2019, 8466–78.
40. Imrie F, Bradley AR, van der Schaar M, et al. Protein family-specific models using deep neural networks and transfer learning improve virtual screening and highlight the need for more data. *J Chem Inf Model* 2018;**58**:2319–30.
41. Adenot M, Lahana R. Blood-brain barrier permeation models: discriminating between potential CNS and non-CNS drugs including P-glycoprotein substrates. *J Chem Inf Comput Sci* 2004;**44**:239–48.
42. Gao Z, Chen Y, Cai X, et al. Predict drug permeability to blood-brain-barrier from clinical phenotypes: drug side effects and drug indications. *Bioinformatics* 2017;**33**:901–8.
43. Plisson F, Piggott AM. Predicting blood(–)brain barrier permeability of marine-derived kinase inhibitors using ensemble classifiers reveals potential hits for neurodegenerative disorders. *Mar Drugs* 2019;**17**:81.
44. Zhang X, Liu T, Fan X, et al. 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. *J Mol Graph Model* 2017;**75**: 347–54.
45. Mackey WA, Shaw GB. Oral hexamethonium bromide in essential hypertension. *Br Med J* 1951;**2**:259–65.
46. Meng F, Xi Y, Huang J, et al. A curated diverse molecular database of blood-brain barrier permeability with chemical descriptors. *Sci Data* 2021;**8**:289.
47. Durant JL, Leland BA, Henry DR, et al. Reoptimization of MDL keys for use in drug discovery. *J Chem Inf Comput Sci* 2002;**42**: 1273–80.
48. Rogers D, Hahn M. Extended-connectivity fingerprints. *J Chem Inf Model* 2010;**50**:742–54.
49. Sabottke CF, Spieler BM. The effect of image resolution on deep learning in radiography. *Radiol Artif Intell* 2020;**2**:e190015.
50. Ikuzawa T, Ino F, Hagihara K. Reducing memory usage by the lifting-based discrete wavelet transform with a unified buffer on a GPU. *J Parallel Dist Com* 2019;**6**:70–80.
51. Hira S, Bai A, Hira S. An automatic approach based on CNN architecture to detect Covid-19 disease from chest X-ray images. *Appl Intell (Dordr)* 2021;**51**:2864–89.
52. Hoefling H, Sing T, Hossain I, et al. HistoNet: a deep learning-based model of normal histology. *Toxicol Pathol* 2021;**49**:784–97.
53. Tao T, Wei X. A hybrid CNN-SVM classifier for weed recognition in winter rape field. *Plant Methods* 2022;**18**:29.
54. Tang Q, Nie F, Kang J, et al. mRNALocator: enhance the prediction accuracy of eukaryotic mRNA subcellular localization by using model fusion strategy. *Mol Ther* 2021;**29**:2617–23.
55. Wang CC, Zhao Y, Chen X. Drug-pathway association prediction: from experimental results to computational models. *Brief Bioinform* 2021;**22**:bbaa061.
56. Chen X, Guan NN, Sun YZ, et al. MicroRNA-small molecule association identification: from experimental results to computational models. *Brief Bioinform* 2018;**21**:47–61.
57. Chen X, Yan CC, Zhang X, et al. Drug-target interaction prediction: databases, web servers and computational models. *Brief Bioinform* 2016;**17**:696–712.
58. He, K., X. Zhang, S. Ren, et al. Deep residual learning for image recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA: IEEE, 2016. p.770–8.
59. Merity S, Keskar NS, Socher R. Regularizing and optimizing LSTM language models. *arXiv* 2017;**1708**:02182.
60. Li C, Feng J, Liu S, et al. A novel molecular representation learning for molecular property prediction with a multiple SMILES-based augmentation. *Comput Intell Neurosci* 2022;**2022**:8464452.
61. Bienfait B, Ertl P. JSME: a free molecule editor in JavaScript. *J Chem* 2013;**5**:24.