



Leveraging quantum chemical properties in transfer learning for predicting blood-brain barrier permeability of drugs

Megan A. Lim¹ · Marybeth G. Yonk² · Kimberly B. Hoang² · Annette M. Molinaro³ · Monika Raj⁴ · Yuhong Du⁵ · Nicholas M. Boulis^{2,6} · Wael Hassaneen^{1,7} · Kecheng Lei²

Received: 3 August 2025 / Accepted: 22 October 2025
© Controlled Release Society 2025

Abstract

The blood-brain barrier (BBB), crucial for central nervous system (CNS) homeostasis, poses challenges for drug delivery in CNS diseases due to selective permeability. Because of this difficulty, there are limited treatments developed for CNS diseases. As a solution, computational models can be implemented in treatment development to enable rapid screening of drug permeability, saving time and resources. This study explores machine learning, deep learning, and transfer learning models to predict the BBB permeability of drug molecules, validated through an in vitro assay known as Parallel Artificial Membrane Permeability Assay-BBB (PAMPA-BBB). Using the Blood-Brain Barrier Database (B3DB) of ~8,000 compounds of known BBB permeability, classification models including support vector machines (SVMs), deep neural networks (DNNs), direct message passing neural networks (D-MPNNs), and transfer learning with quantum chemical properties were developed. Experimental validation with 18 compounds from the Emory Enriched Bioactive Library (EEBL), a library containing 1,018 FDA-approved pharmacologically active compounds of known activity, highlighted PAMPA-BBB as a robust validation method. The SVM model with combined 2D RDKit and Morgan fingerprint molecular representation achieved high performance (accuracy: 89.08%) on the B3DB test set. The best-performing models for the 18 EEBL compounds were transfer learning models. In particular, the model trained on the QM9-extended polarizability property correctly classified 17 out of 18 compounds, while the model trained on the QM9-extended dipole moment property achieved correct classification across all 18 experimental compounds. Additional analyses demonstrated that QC-based transfer learning provides complementary predictive value beyond traditional molecular descriptors such as LogP and molecular weight. QC-pretrained models achieved higher accuracy and ROC-AUC on both the B3DB and external PAMPA test sets, with performance remaining robust even after descriptor ablation. Moreover, QC-pretrained models outperformed the baseline of P-glycoprotein (P-gp) inhibition, underscoring the unique contribution of quantum-derived representations to BBB permeability prediction. Therefore, this study motivates the synergy of computational and experimental methods in enabling faster, more cost-effective, and targeted identification of CNS-active or CNS-sparing drugs.

Keywords Machine learning · Deep learning · Transfer learning · Quantum chemical property · PAMPA

✉ Wael Hassaneen
wael.mostafa@carle.com

✉ Kecheng Lei
kecheng.lei@emory.edu

¹ Carle Illinois College of Medicine, University of Illinois Urbana-Champaign, Champaign, IL, USA

² Department of Neurosurgery, Emory University School of Medicine, Atlanta, GA, USA

³ Department of Neurological Surgery, University of California, San Francisco, San Francisco, CA, USA

⁴ Department of Chemistry, Emory University, Atlanta, GA, USA

⁵ Department of Pharmacology and Chemical Biology, Emory Chemical Biology Discovery Center, Emory University School of Medicine, Atlanta, GA, USA

⁶ Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA, USA

⁷ Department of Neurosurgery, Carle Foundation Hospital, Urbana, IL, USA

Introduction

The blood-brain barrier (BBB) is a highly selective, protective boundary of the central nervous system (CNS) that plays a critical role in maintaining the brain's microenvironment [1]. It functions to shield the brain from harmful substances, pathogens, and toxins present in the bloodstream while allowing essential nutrients and molecules to pass through. Its protective mechanisms include tight junctions, active transport systems such as efflux pumps, and enzymatic barriers [2]. While the BBB is highly effective, it also poses challenges for drug delivery, as it restricts the entry of most therapeutic agents [3–5]. The inability of drugs to cross the BBB creates a major hurdle in developing treatments for neurological conditions, such as Alzheimer's disease, Parkinson's disease, brain and spinal cord tumors, and epilepsy, where effective drug delivery to the brain is essential for therapeutic success [6]. Understanding and overcoming these challenges is a critical focus of modern drug development.

Laboratory experiments based on cell-based methods are the current standard for accurately determining whether drugs can effectively cross the BBB, but they are also time consuming and labor intensive [7]. For reference, cell culture supplies and equipments cost thousands of dollars and require delicate and time intensive experimental procedures [8]. In recent years, machine learning (ML) and deep learning (DL) have become increasingly effective tools for predicting BBB permeability at virtually limited cost [9, 10]. This approach has the potential to significantly accelerate the drug discovery process by allowing researchers to screen thousands of compounds and prioritize promising candidates early on, saving both resources and time. In 2020, Singh et al. developed random forest, multilayer perceptron, and sequential minimal optimization models using a small dataset of 605 compounds and achieved an accuracy of 86.5% on an external set of 1,566 compounds [11]. The DeePred-BBB study applied ML (support vector machines, k-Nearest Neighbor, Random Forest, naïve Bayes) and DL (deep neural network, convolutional neural network 1-dimension, convolutional neural network by transfer learning) algorithms to a dataset of 3,605 diverse compounds [12]. They discovered that a DNN model with three layers (depth) having 200, 100, and 2 nodes each and integrated calculated features from the open-source Padel tool was most accurate.

In 2021, Blood-Brain Barrier Database (B3DB), the largest benchmark dataset for BBB permeability to date, was introduced. Compiled from 50 published sources and consisting of 7,807 compounds, it was designed to address the limitations of previous studies, which were constrained by small datasets and limited chemical diversity [13]. While

the literature had shown significant progress in applying ML and DL models to predict BBB permeability, there was now the opportunity to improve upon prior approaches and explore new methodologies that leverage the expanded chemical diversity of B3DB. Following its release, an ML-based classification read-across structure-activity relationship linear discriminant analysis model using the B3DB dataset emerged, highlighting the role of lipophilicity, electronic effects, and steric factors in facilitating BBB prediction [14]. Using a validation set of compounds from the DrugBank dataset, which had undergone experimental assessments for BBB penetration, their model achieved a predictive accuracy of 0.673, precision of 0.928, F-measure of 0.757, and balanced accuracy of 0.723, results comparable to the performance of the LightBBB online server [15]. Also utilizing the B3DB dataset, the transformer-based model MegaMolBART combined with an XGBoost classifier demonstrated improved results over traditional machine learning approaches achieving an accuracy of around 0.83 on the B3DB test set [16]. These studies have demonstrated the utility of the B3DB dataset; however, there is still potential to further improve model accuracy and ensure robust generalizability to novel compounds.

Transfer Learning is a powerful ML technique that has been highly successful in various domains, including natural language processing [17], computer vision [18], and speech recognition [19]. By leveraging knowledge from a pre-trained model on one task, it enhances performance on a different but related task. In drug discovery, transfer learning has been used to predict molecular properties and activities, including physiological, biophysical, and physicochemical characteristics [20]. However, its application to BBB permeability prediction remains largely unexplored. Quantum chemical properties have been shown to augment DL models for predicting molecular properties, including absorption, distribution, metabolism, and excretion (ADME), which are critical in modern drug discovery [21]. But no prior work has integrated quantum chemical properties with BBB permeability prediction using transfer learning.

In our study, we developed a novel transfer learning approach that leverages quantum chemical (QC) property as the source domain to predict the binary classification of BBB permeability, validating its performance on an external dataset derived from *in vitro* experiments. QC properties including electronic, topological, and geometric descriptors are anticipated to be useful in BBB prediction due to their relation to the fundamental factors influencing BBB transport including logD, pKa, rate of transport, binding affinity with p-glycoproteins [22]. Moreover, transfer learning is particularly powerful when large, well-characterized datasets (such as QC properties) are leveraged to generate predictions in a lower-data domain (BBB permeability)

[23]. We pretrained deep neural network-based models on QC properties using the QM9 and QM9-extended datasets, then fine-tuned them for BBB permeability prediction using the B3DB dataset. To further evaluate our approach, we compared our transfer learning models against traditional machine learning and deep learning methods, including support vector machines (SVM), deep neural networks (DNN), and direct message passing neural networks (D-MPNN), using the B3DB dataset. To assess the ability of these models to generalize beyond the B3DB chemical space, we curated an external test set of compounds. We selected 18 compounds from the Emory Enriched Bioactive Library (EEBL) with limited documentation of their BBB permeabilities in the literature and performed in vitro experiments with Parallel Artificial Permeability Assay-BBB (PAMPA-BBB) designed to simulate the passage of drugs across BBB membranes using dried lipids in a system of 96-well plates. The PAMPA assay serves as a more efficient alternative to traditional, often expensive and time-consuming cell-based methods, and is used as the ground truth in this study. Ultimately, the synergy of computation and experimentation creates a self-improving system in which the continuous collection of new data allows for refining and improving model accuracy over time. This iterative process accelerates drug discovery, enabling faster, more cost-effective, and targeted identification of CNS-active or CNS-sparing drugs.

Results

The B3DB dataset was used for model development and divided into a 75/25 train/test split. These splits were used to build a D-MPNN model as implemented in Chemprop (<https://github.com/chemprop>)³. For the other models, molecular representations including “feature set 1”, 2D RDKit descriptors, Morgan fingerprints, combined 2D RDKit descriptors and Morgan fingerprints, and Mol2Vec were computed. The descriptor set, which we refer to as “feature set 1” consists of molecular weight (MW), partition coefficient (LogP), topological surface area (TPSA), and number of hydrogen bond donors (HBD). These were selected based on the CNS drug space as defined by the 6 physicochemical properties: calculated log partition

coefficient (ClogP), calculated log distribution coefficient (ClogD), MW, TPSA, HBD, acid dissociation constant (pKa). For transfer learning, the source domain models were built first with the QM9 and QM9-extended datasets prior to the fine-tuning of the second model with the B3DB data [21, 24]. Details on the datasets used in this study are provided in Table 1. Model performances were assessed using the following metrics: accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic (ROC) curve (AUC). The overall workflow applied in this study is illustrated in Fig. 1.

Evaluation of model performances on the B3DB test set

Molecular representation plays a critical role in model performance [25, 26]. Different representations encode different aspects of chemical information, capturing local to global properties, and directly impacting the model’s ability to learn meaningful relationships. To gain insight into which representations worked best depending on the complexity of the associated model, we evaluated several molecular representations, including molecular descriptors, Morgan fingerprints, combination of molecular descriptors and Morgan fingerprints, and learned vector representations of molecular substructures (Mol2Vec), across several machine learning algorithms. As depicted in Fig. 2, SVMs achieved the highest overall performance on the B3DB test set with the combined RDKit and Morgan fingerprints (accuracy: 89.08%, F1: 0.92, AUC: 0.877). SVMs also achieved high performance with the 208 RDKit computed 2D physicochemical descriptors (accuracy: 87.96%, F1: 0.91, AUC: 0.866) and Mol2Vec vector representations (accuracy: 87.7%, F1: 0.91, AUC: 0.861), but lost performance with Morgan fingerprints (accuracy: 83.18%, F1: 0.88, AUC: 0.771) and the simpler initial descriptor set, feature set 1 (accuracy: 82.47%, F1: 0.87, AUC: 0.794). For DNNs, model performance was consistent across all representations involving 2D physicochemical properties, Morgan fingerprints, and their combination, with accuracies of 86.99%, 86.13%, and 87.35% respectively, but had lower performance with Mol2Vec (42%). The D-MPNNs achieved better performance with hyperparameter optimization (accuracy: 88.01%) and did

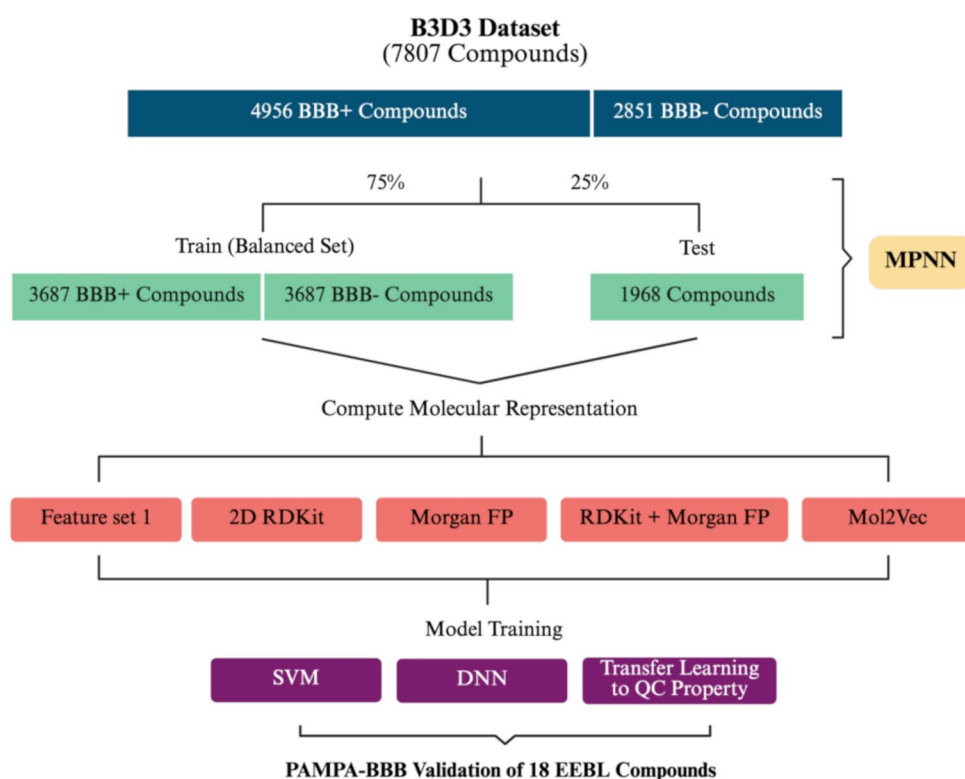
Table 1 Datasets for BBB permeability and QC property prediction

Name	Task	Used for training or testing	Data set size	Elements covered	No. Heavy atoms
B3DB	BBB+/BBB-	Both	7,807	CNOFSCIBr	1-136
QM9	QC property	Training	133,885	CNOF	1-9
QM9-extended	QC property	Training	153,716	CNOFSCI	1-9
Emory Test	BBB+/BBB-	Testing	18	CNOFSCIBr	6-45

B3DB was used in training and testing for model development for BBB permeability prediction. QM9 and QM9-extended datasets were used for source domain model development in the first half of transfer learning. Emory Test consists of the 18 compounds selected to run through PAMPA and validate the computational models

Fig. 1 Illustration of the workflow applied in this study.

The B3DB dataset was split into a random train/test 75/25 split. To achieve a more balanced dataset, we applied the oversampling strategy to the training split. A D-MPNN model was built as implemented in Chemprop. For the other models, molecular representations including “feature set 1”, 2D RDKit, Morgan fingerprints, combined 2D RDKit and Morgan fingerprints, and Mol2Vec were computed. For the transfer learning models, the source domain models were built first with the QM9 and QM9-extended datasets prior to the fine-tuning of the second model with the B3DB data. The QC properties used were dipole moment, polarizability, HOMO, LUMO, gap, electronic spatial extent, ZPVE, and heat capacity. Created with [Biorender.com](https://biorender.com)



Best Model Performance on the B3DB Test Set

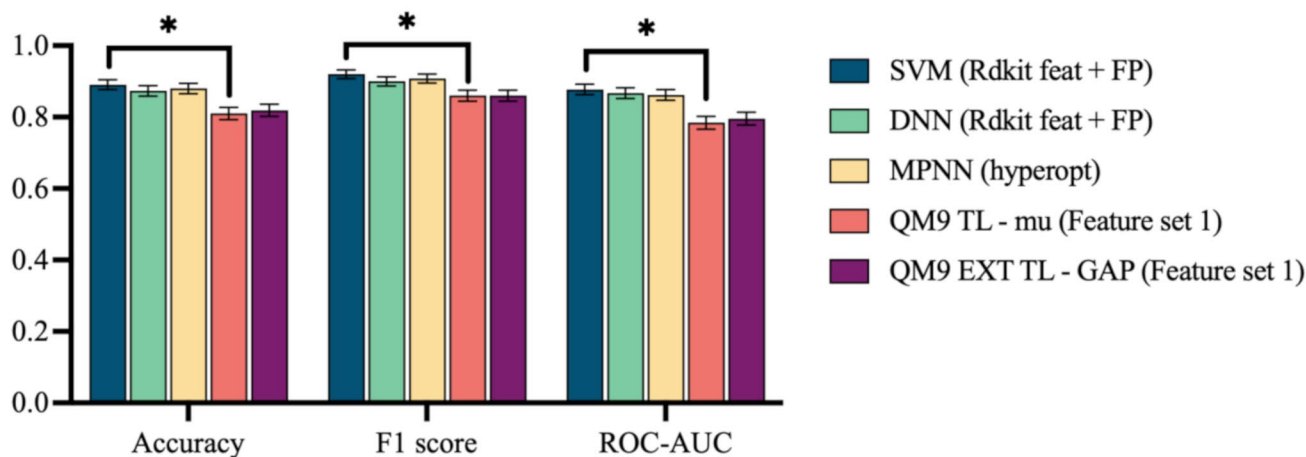


Fig. 2 Model performances on the B3DB test set.

The models shown here represent the top-performing ones on the B3DB test set, evaluated using three key metrics: accuracy, F1 score, and ROC-AUC. SVMs achieved the highest overall performance with

not demonstrate any significant improvement from RDKit feature concatenation to their learned molecular representation (accuracy: 87.75%).

For QM9-based transfer learning, the source QC task of polarizability with the initial descriptor set (feature set 1) as representation, achieved predictive performance of 80.74% accuracy and dipole moment of 81.00% on the B3DB test set (Table 2). This descriptor set (feature set 1) achieved

the combined RDKit and Morgan fingerprints. Significant difference was found between SVM (Rdkit feat + FP) and QM9 TL - mu (Feature set 1) (* $p < 0.05$). Error bars express binomial confidence interval at 95% confidence level.

consistent performance across all QC properties. For transfer learning models based on the QM9-extended source domain, the best QC properties to offer transfer learning capability with the initial descriptor set (feature set 1) as representation were polarizability (79.88% accuracy on the B3DB test set), dipole moment (81.10%), HOMO (79.98%), LUMO (80.23%), gap (81.91%), and ZPVE (81.40%). Overall, the transfer learning models had improved performance

Table 2 Model performances on the B3DB test set. This table outlines the performances of several molecular representation and model combinations (feature set 1, 2D RDKit, Morgan fingerprint, Rdkit feat+FP, and Mol2Vec) on the B3DB test set. Metrics include accuracy, F1 score, and ROC-AUC

Model	Molecular Representation	Accuracy	F1 Score	ROC-AUC
SVM (grid search hyper opt)	Feature set 1	0.8247	0.87	0.7943
	2D RDKit	0.8796	0.91	0.8661
	Morgan fingerprint	0.8318	0.88	0.7716
	Rdkit feat+FP	0.8908	0.92	0.8774
	Mol2Vec	0.8770	0.91	0.8619
DNN [4]	Feature set 1	0.8232	0.86	0.805
	2D RDKit	0.8699	0.9	0.8683
	Morgan fingerprint	0.8613	0.89	0.8503
	Rdkit feat+FP	0.8735	0.9	0.8669
	Mol2Vec	0.4200	0.89	0.8413
MPNN, hyperopt	Learned	0.8801	0.908	0.8620
MPNN	Learned	0.8780	0.905	0.8669
MPNN+RdKit features	Learned	0.8775	0.906	0.8619
QM9 TL - alpha	Feature set 1	0.8074	0.86	0.7752
	2D RDKit	0.6611	0.73	0.6479
	Morgan fingerprint	0.7246	0.81	0.6419
	Rdkit feat+FP	0.6662	0.79	0.5458
	Mol2Vec	0.6936	0.78	0.6349
QM9 TL - mu	Feature set 1	0.8100	0.86	0.7842
	2D RDKit	0.7678	0.82	0.7563
	Morgan fingerprint	0.7119	0.77	0.6889
	Rdkit feat+FP	0.7348	0.81	0.6851
	Mol2Vec	0.7774	0.83	0.742
QM9 TL - HOMO	Feature set 1	0.7200	0.78	0.701
	2D RDKit	0.7256	0.79	0.6986
	Morgan fingerprint	0.4705	0.41	0.5448
	Rdkit feat+FP	0.6662	0.71	0.6778
	Mol2Vec	0.6387	0.7	0.6328
QM9 TL - LUMO	Feature set 1	0.7759	0.82	0.7646
	2D RDKit	0.7561	0.81	0.728
	Morgan fingerprint	0.7300	0.73	0.6212
	Rdkit feat+FP	0.7444	0.82	0.6849
	Mol2Vec	0.7434	0.81	0.6976
QM9 TL - gap	Feature set 1	0.7815	0.83	0.7589
	2D RDKit	0.7332	0.80	0.6865
	Morgan fingerprint	0.6946	0.78	0.6353
	Rdkit feat+FP	0.6911	0.76	0.6223
	Mol2Vec	0.6768		0.6296
QM9 TL - ZPVE	Feature set 1	0.7825	0.83	0.7594
	2D RDKit	0.7215	0.78	0.6954
	Morgan fingerprint	0.6748	0.75	0.6367
	Rdkit feat+FP	0.6723	0.76	0.6132
	Mol2Vec	0.6540	0.74	0.6167
QM9 TL - cv	Feature set 1	0.7998	0.85	0.775
	2D RDKit	0.7307	0.79	0.7089
	Morgan fingerprint	0.7109	0.80	0.6251
	Rdkit feat+FP	0.7038	0.80	0.618
	Mol2Vec	0.6479	0.72	0.6315
QM9 TL - r2	Feature set 1	0.7576	0.83	0.6806
	2D RDKit	0.6621	0.71	0.6708
	Morgan fingerprint	0.3918	0.14	0.5203
	Rdkit feat+FP	0.6108	0.63	0.6465

Table 2 (continued)

Model	Molecular Representation	Accuracy	F1 Score	ROC-AUC
QM9 ext TL - alpha	Mol2Vec	0.6692	0.72	0.6693
	Feature set 1	0.7988	0.85	0.7611
	2D RDKit	0.6936	0.78	0.6426
	Morgan fingerprint	0.6916	0.79	0.6137
	Rdkit feat+FP	0.3552	0	0.5
QM9 ext TL - mu	Mol2Vec	0.6443	0.72	0.6227
	Feature set 1	0.8110	0.86	0.7808
	2D RDKit	0.7256	0.78	0.7191
	Morgan fingerprint	0.7053	0.78	0.6661
	Rdkit feat+FP	0.7571	0.82	0.7239
QM9 ext TL - HOMO	Mol2Vec	0.7749	0.84	0.7284
	Feature set 1	0.7998	0.85	0.7712
	2D RDKit	0.6580	0.71	0.6558
	Morgan fingerprint	0.4837	0.45	0.5473
	Rdkit feat+FP	0.6585	0.75	0.607
QM9 ext TL - LUMO	Mol2Vec	0.6580	0.73	0.6372
	Feature set 1	0.8023	0.85	0.7844
	2D RDKit	0.7754	0.83	0.7481
	Morgan fingerprint	0.5767	0.61	0.604
	Rdkit feat+FP	0.6540	0.72	0.6436
QM9 ext TL - gap	Mol2Vec	0.7327	0.80	0.689
	Feature set 1	0.8191	0.86	0.7958
	2D RDKit	0.7393	0.80	0.7172
	Morgan fingerprint	0.6606	0.73	0.6356
	Rdkit feat+FP	0.6997	0.77	0.6582
QM9 ext TL - zpve	Mol2Vec	0.7149	0.79	0.6643
	Feature set 1	0.8140	0.86	0.7893
	2D RDKit	0.7231	0.79	0.6808
	Morgan fingerprint	0.7114	0.78	0.6689
	Rdkit feat+FP	0.6575	0.74	0.6223
QM9 ext TL - cv	Mol2Vec	0.7058	0.80	0.6302
	Feature set 1	0.7154	0.80	0.6377
	2D RDKit	0.6982	0.76	0.6808
	Morgan fingerprint	0.6895	0.78	0.6153
	Rdkit feat+FP	0.7790	0.83	0.757
QM9 ext TL - r2	Mol2Vec	0.6601	0.73	0.6371
	Feature set 1	0.3552	0	0.5
	2D RDKit	0.6773	0.74	0.6563
	Morgan fingerprint	0.3948	0.14	0.5237
	Rdkit feat+FP	0.6570	0.73	0.6267
	Mol2Vec	0.7129	0.78	0.6775

with a simpler molecular representation (feature set 1) and expanded chemical space that included Sulfur (S) and Chlorine (Cl) atoms.

Validation of model performances on the PAMPA-BBB assay derived test set

To validate the effectiveness of the computational models, 18 of the 2,036 active EEBL compounds were chosen to complete in vitro validation [27]. These 18 compounds were chosen after filtering compounds found in the B3D3 training

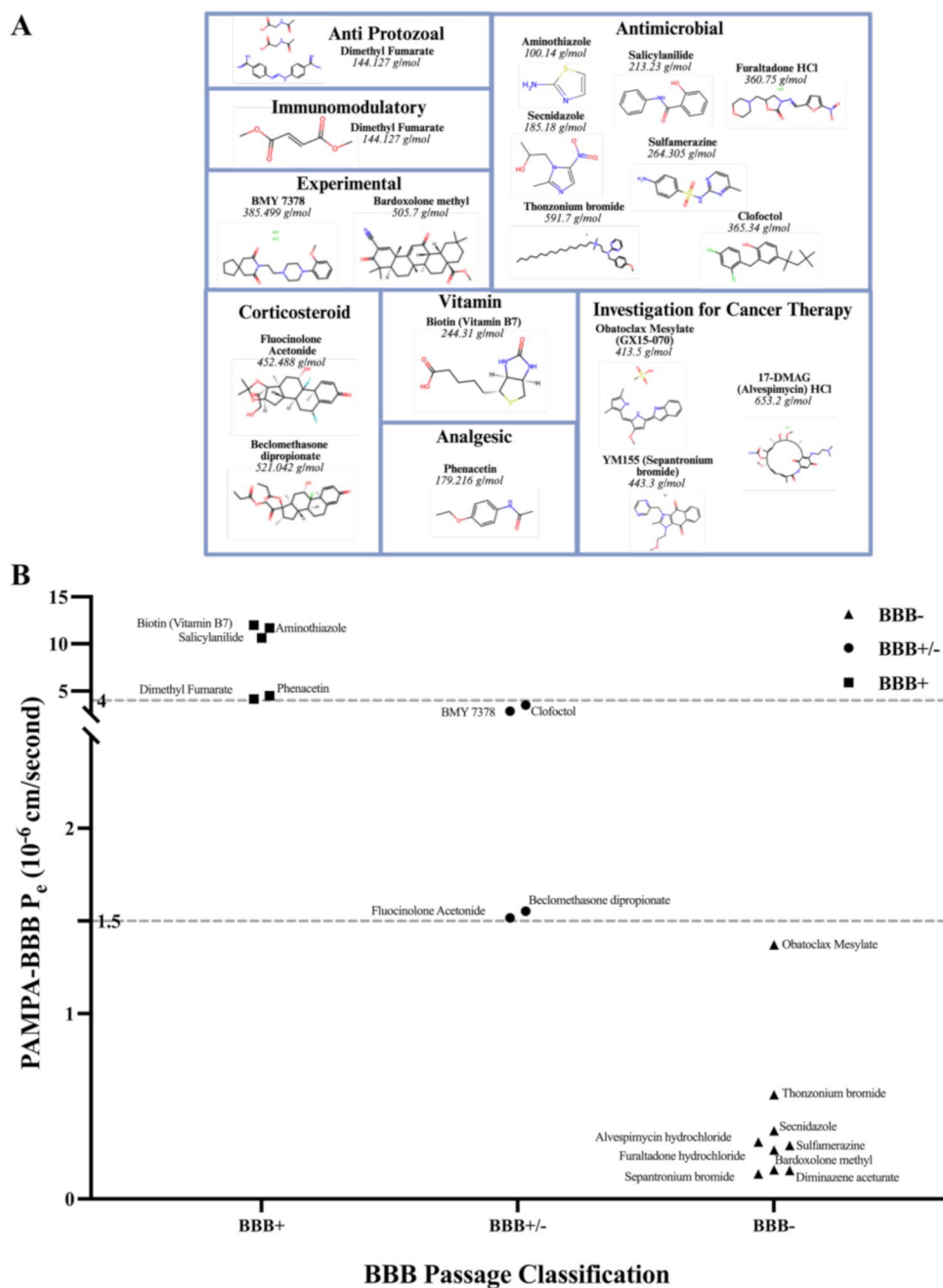
set. The goal was to form an external validation set approximating equal representation of permeable and impermeable compounds. The permeation for each of the 18 compounds as determined by the PAMPA-BBB Assay are shown in Supplementary Table 1. The compounds can be further classified based on their drug class: experimental (2), corticosteroid (2), immunomodulatory (1), vitamin (1), analgesic (1), antimicrobial (7), investigation for cancer therapy (3), and antiprotozoal (1). These drugs vary significantly in terms of class, mechanism of action, and therapeutic uses. Some, like beclomethasone dipropionate [28] and dimethyl fumarate

[29] are well-established medications for inflammatory and autoimmune diseases, whereas others, like 17-DMAG [30] and Obatoclast Mesylate [31], are experimental and under investigation for cancer and other conditions. Figure 3A shows the compound classes and Fig. 3B outlines the overall permeability trends in form of P_e value for each of the tested drugs.

The heat map in Fig. 4 shows the performance of selected models on this test set of 18 compounds. These models were chosen based on their performance on a 25% held-out test split from the B3DB dataset. The D-MPNN achieved an accuracy of 77.78%, SVM of 72.22%, and DNN of 83.33%.

Fig. 3 PAMPA-BBB permeability value versus BBB passage grouping for each drug.

(A) This image outlines the various classifications of each of the 18 compounds whose predicted permeability through the BBB was validated with the PAMPA-BBB assay. Images created with RDKit and figure created with BioRender.com. (B) This figure shows the relative P_e or permeability values of each drug as it pertains to its ability to pass the BBB. Boundaries set by low and high permeability controls fall at 1.5×10^{-6} cm/second and 4×10^{-6} cm/second, respectively. These boundaries are indicated by color change and dotted lines on the graph. Each dot represents one of the drugs tested in the PAMPA-BBB Assay and their respective P_e values



Overall, the best performing models were transfer learning models. Specifically, the model trained on the QM9-extended polarizability property correctly classified 17 out of 18 compounds, while the model trained on the QM9-extended dipole moment property achieved correct classification across all 18 experimental compounds.

Additional analyses to assess the contribution of QC transfer learning

To clarify the contribution of QC-based transfer learning, we compared a descriptor-only DNN baseline to models

Model Validation on the PAMPA Curated EEBL Test Set

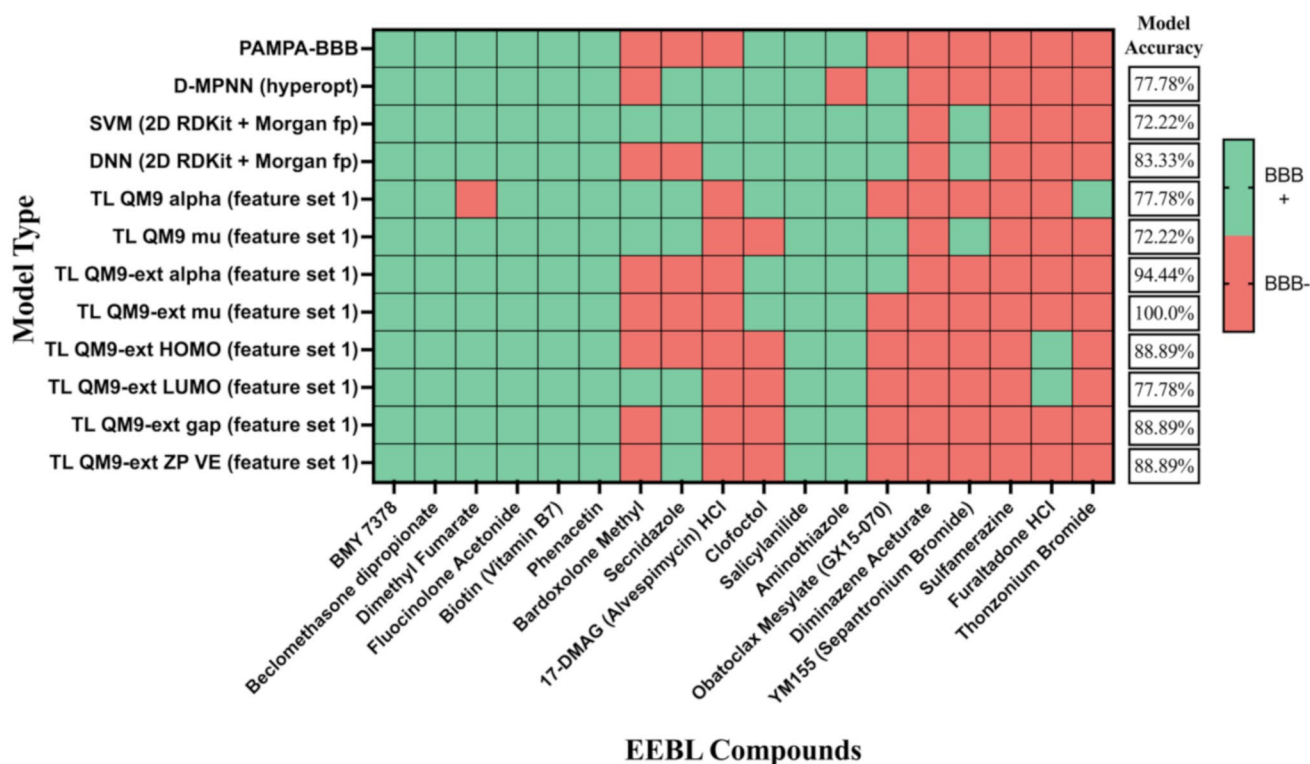


Fig. 4 Classification of PAMPA-validated compounds and heat map of model predictions on the PAMPA external test set. Model performances on the external dataset curated from the results of

the PAMPA experiments of EEBL compounds. Models listed here are the models that performed best on the B3DB test set

initialized with QC-pretrained embeddings across the two independent test sets (Fig. 5). On the B3DB held out test set, QC-transfer learning models achieved comparable or modestly higher performance relative to the descriptor baseline (Accuracy 0.8059, ROC-AUC 0.8010), with several QC-derived embeddings (e.g., gap, zpve, mu) showing improvements in accuracy and F1. In contrast, on the PAMPA curated test set, the benefit of QC pretraining was more pronounced: while the descriptor-only baseline achieved Accuracy/ROC-AUC 0.78, with QM9-ext TL dipole moment and polarizability embeddings achieving near-perfect accuracy and ROC-AUC.

Feature ablation

To evaluate whether QC-transfer learning provides predictive value beyond lipophilicity and molecular weight, we repeated the analysis after removing LogP and MW from the descriptor set (Fig. 6). On the B3DB test set, performance decreased across all models when these descriptors were excluded, yet QC-transfer learning models maintained competitive accuracy and ROC-AUC (e.g., QM9-ext TL

gap, Accuracy 0.819 $\rightarrow$ 0.753). On the PAMPA test set, QC-pretrained models continued to achieve strong performance despite the removal of LogP and MW, including QM9-ext TL alpha (Accuracy 0.944 $\rightarrow$ 0.889) and QM9-ext TL gap/zpve (Accuracy $\sim$ 0.833).

Comparison with Pgp inhibition

To further evaluate the specificity and generalizability of QC-based transfer learning, we compared models pretrained on QC properties (QM9 and QM9-ext) to models pretrained on P-glycoprotein (P-gp) inhibition, a task-specific biological property relevant to drug absorption, metabolism, and brain penetration (Fig. 7). On the B3DB test set, QC-transfer learning models achieved higher predictive performance than P-gp pretrained models. Accuracy ranged from 0.7988 to 0.8191 for QC-based embeddings compared to 0.7739–0.7769 for P-gp embeddings, with similar trends observed for F1 score (0.85–0.86 vs. 0.82–0.83) and ROC-AUC (0.7611–0.7958 vs. 0.7521–0.7657). On the PAMPA curated test set, the differences were more pronounced. QC-transfer learning models showed substantially higher performance,

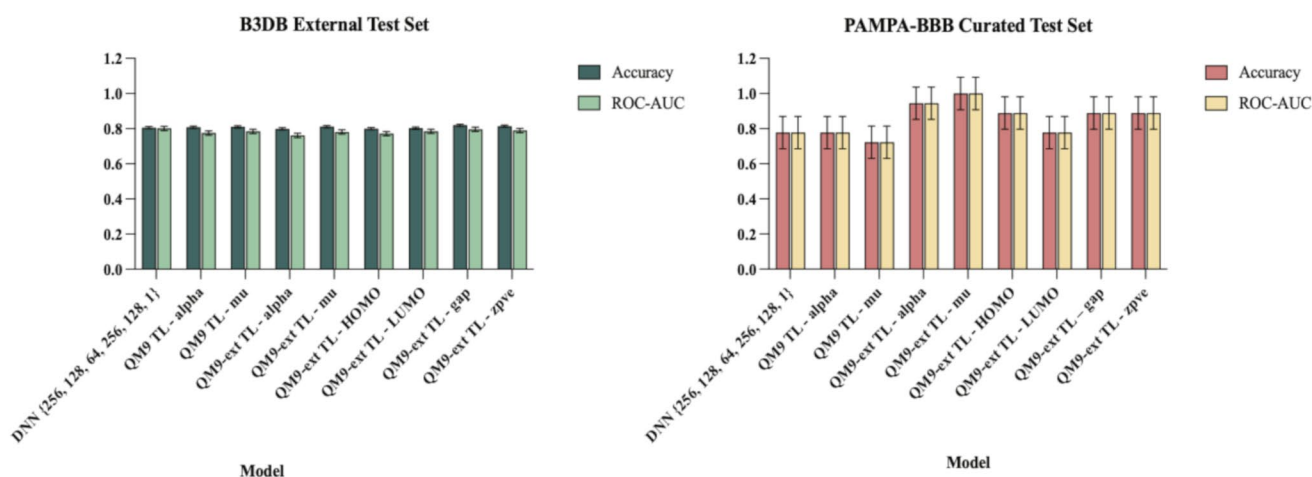


Fig. 5 Performance comparison of descriptor-only and QC-transfer learning models for BBB permeability prediction.

Accuracy and ROC-AUC are shown for nine models evaluated on the B3DB external test set (left) and the PAMPA curated test set (right). The descriptor-only DNN serves as the baseline. QC-transfer learn-

ing models achieved comparable or modest improvements on B3DB, while showing substantial performance gains on the PAMPA set, with QM9-ext TL μ and α embeddings achieving near-perfect predictive performance. Error bars express binomial confidence interval at 95% confidence level

with some embeddings achieving correct prediction for all 18 compounds, whereas P-gp pretrained models performed poorly (Accuracy/F1/ROC-AUC $\approx$ 0.61).

Scaffold split

To evaluate model generalizability to novel chemical scaffolds, we constructed models using a scaffold-based split to ensure test molecules from the B3DB set are structurally distinct from those used in training. Overall, models exhibited a modest decline in performance on the held out B3DB and PAMPA curated test sets compared to random splits, consistent with the more stringent nature of scaffold-based evaluation (Table 3).

On the B3DB test set, the MPNN model achieved 0.8683 / 0.8971 / 0.9262 (Accuracy / F1 / ROC-AUC) under scaffold split compared to 0.8801 / 0.908 / 0.8620 under random split. The DNN model showed a similar trend (0.7982 / 0.84 / 0.7820 vs. 0.8232 / 0.86 / 0.805). For the SVM baseline, performance decreased slightly from 0.8247 / 0.87 / 0.7943 to 0.8140 / 0.86 / 0.7872, while the QC-based transfer learning models (QM9-ext pretraining) generally maintained accuracies between 0.77 and 0.82 on B3DB.

Evaluation on the PAMPA dataset reflected a comparable pattern. Accuracies across models ranged from 0.72 to 0.83 under scaffold split, compared to 0.78–1.00 under random split, indicating reduced but still meaningful predictive performance on unseen molecular scaffolds.

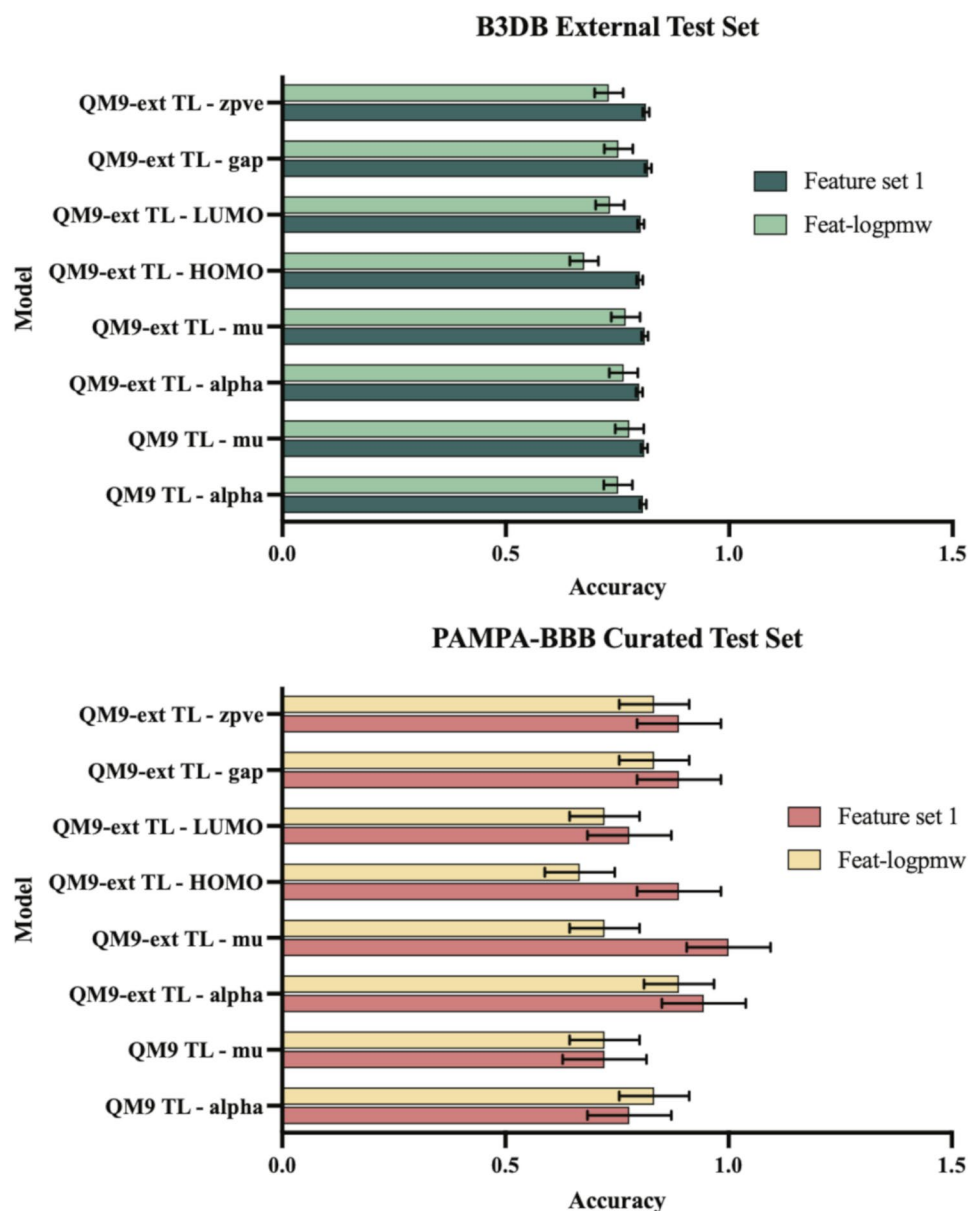
Discussion and conclusion

Leveraging the B3DB dataset, we investigated several molecular representations and machine learning models, evaluating their effectiveness in predicting BBB permeability. Additionally, we examined the potential for transfer learning techniques, particularly when trained on a source dataset of QC properties, to transfer to the domain task of BBB penetration. To assess the generalizability of these approaches, we curated an external dataset comprising 18 compounds, each with corresponding in vitro data obtained through PAMPA experiments.

Evaluation of model performances with various molecular representations on the B3DB test set demonstrated that SVMs (combined RDKit and Morgan fingerprint representation) attained best overall performance with accuracy of 89.08%, AUC of 0.877, and an F1 score of 0.92. RDKit computed physicochemical properties reflect the drug's passive diffusion capabilities, while Morgan fingerprints capture specific interactions like uptake, efflux, and protein binding. The integration of these complementary descriptors provided a more complete and comprehensive representation of the molecules.

Given the potential for transfer learning to enhance model performance when data from related domains is available, we investigated its applicability to the B3DB dataset. Transfer learning methods are particularly useful in situations where labeled data is limited or expensive to obtain, as they allow a model to leverage knowledge learned from a source domain (QC properties) and apply it to a target domain (BBB permeability) [32]. We selected QC properties to be the source domain in our study because

Fig. 6 Impact of removing LogP and molecular weight (MW) descriptors on transfer learning model performance. Bar plots compare accuracy for models trained with the full feature set versus an ablated feature set excluding LogP and MW. Results are shown for (upper) the B3DB test set and (lower) the PAMPA test set. Across most models, removal of LogP and MW led to a reduction in predictive performance, with the largest drops observed in QM9-ext TL models using HOMO, LUMO, and gap descriptors. Notably, QM9-ext TL models with dipole moment (μ) retained relatively strong accuracy even without LogP and MW, particularly in the B3DB test set, suggesting that dipole-derived quantum properties capture complementary information to standard physicochemical descriptors. These findings highlight the importance of LogP and MW as influential features in BBB permeability prediction, while also underscoring the added predictive value of incorporating quantum chemical descriptors. Error bars express binomial confidence interval at 95% confidence level



they provide detailed, fundamental insights into the electronic structure and reactivity of molecules that influence the fundamental factors influencing BBB transport including logD, pKa, rate of transport, binding affinity with p-glycoproteins. Both the QM9 and QM9-extended datasets were used for source domain model development to investigate whether training on an expanded chemical space including the additional heavy atoms of S and Cl had any impact on model performance. Notably, these datasets consist primarily of small molecules (≤ 250 Da), whereas typical BBB-permeable drugs are larger (~ 300 – 400 Da). Performing high-quality QC calculations on many larger molecules is computationally intensive and impractical, so pretraining on smaller molecules allows the model to efficiently learn fundamental QC relationships, which can then be fine-tuned

on BBB permeability data spanning a broader, drug-like chemical space.

For transfer learning models based on the QM9 source domain, polarizability and dipole moment emerged as the most effective QC properties for enhancing predictive performance (Table 2). To provide chemical context for their influence on BBB permeability, we examined representative compounds from both the training and test sets. From the training set, diazepam, a benzodiazepine with a fused aromatic ring system, exhibits high polarizability, moderate dipole moment, and high lipophilicity, features that align with its known BBB permeability. In contrast, furosemide, a sulfonamide diuretic with a polar sulfonamide group and multiple hydrogen bond donors, displays reduced polarizability and higher dipole moment, contributing to its

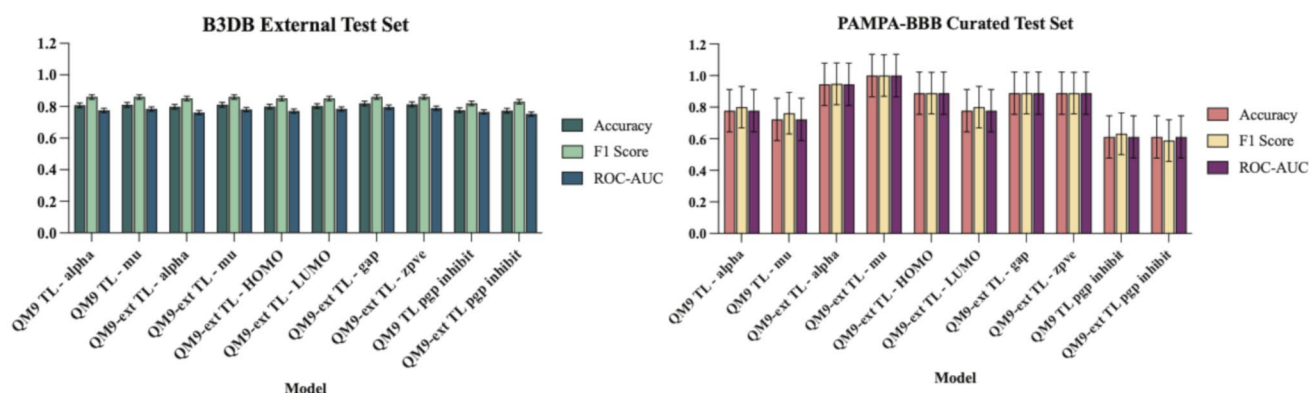


Fig. 7 Comparison of QC-based transfer learning vs. Pgp inhibition pretrained embeddings for BBB permeability prediction. Grouped bar plots show Accuracy, F1 Score, and ROC-AUC for models pretrained on general QC properties (QM9 and QM9-ext) versus models pretrained on P-gp inhibition. Results are shown separately for the B3DB test set (left) and PAMPA curated test set (right). QC-trans-

fer learning models consistently outperform P-gp TL models, particularly on the PAMPA set, demonstrating that QC-derived embeddings capture generalizable molecular features relevant for BBB permeability, whereas task-specific embeddings (P-gp inhibition) are less transferable. Error bars express binomial confidence interval at 95% confidence level

Table 3 Performance of different machine learning models and transfer learning approaches on the B3DB and PAMPA test sets

Model	Molecular Representation	B3DB (Acc / F1 / ROC-AUC)	PAMPA (Acc / F1 / ROC-AUC)
DNN	Feature set 1	0.7982 / 0.84 / 0.7820	0.8333 / 0.8421 / 0.8333
MPNN	Learned	0.8683 / 0.8971 / 0.9262	0.8333 / 0.8571 / 0.8333
SVM	Feature set 1	0.8140 / 0.86 / 0.7872	0.7222 / 0.7826 / 0.7222
QM9-ext TL - alpha	Feature set 1	0.7772 / 0.85 / 0.7058	0.7222 / 0.7826 / 0.7222
QM9-ext TL - mu	Feature set 1	0.8181 / 0.87 / 0.7810	0.7222 / 0.7619 / 0.7222
QM9-ext TL - HOMO	Feature set 1	0.8053 / 0.85 / 0.7859	0.7222 / 0.7059 / 0.7222
QM9-ext TL - LUMO	Feature set 1	0.8145 / 0.86 / 0.7852	0.7222 / 0.7619 / 0.7222
QM9-ext TL - gap	Feature set 1	0.8145 / 0.86 / 0.7855	0.7778 / 0.8000 / 0.7778
QM9-ext TL - zpve	Feature set 1	0.8130 / 0.86 / 0.7849	0.6667 / 0.7273 / 0.6667
QM9-ext TL - r2	Feature set 1	0.7367 / 0.80 / 0.7016	0.6111 / 0.6957 / 0.6111
QM9-ext TL - cv	Feature set 1	0.7725 / 0.83 / 0.7240	0.6111 / 0.6957 / 0.6111

Accuracy (Acc), F1 score (F1), and ROC-AUC are reported for each model. “Feature set 1” refers to molecular descriptors (MW, LogP, TPSA, and HBD) used as input, while “Learned” indicates representations learned directly by the model (e.g., MPNN). QM9-ext TL denotes models pretrained on QM9-extended QC properties (α , μ , HOMO, LUMO, gap, ZPVE, r^2 , CV) and subsequently fine-tuned on the BBB permeability task

non-permeable classification. From the test set, fluradoline contains aromatic systems imparting high polarizability, aligning with its BBB permeability. Mezlocillin, a BBB-impermeable β -lactam antibiotic with a penicillin core, shares structural and QC features with furosemide including high dipole moment from multiple polar moieties.

For transfer learning models based on the QM9-extended source domain, the best QC properties to offer transfer learning capability with the initial descriptor set (feature set 1) as representation were polarizability, dipole moment, HOMO, LUMO, gap, and ZPVE. The four descriptors of feature set 1 (MW, LogP, TPSA, and HBD), capture core molecular properties and are consistent across chemical spaces, making them ideal for transfer learning. A simpler representation is preferred for transfer learning as it reduces noise and redundancy, allowing the model to focus on generalizable patterns learned during pretraining. By expanding

the chemical space to include the S and Cl atoms of QM9-extended, several QC properties including the electronic properties (HOMO, LUMO, gap) had improved transfer learning performance. This supports the chemical intuition that S and Cl can alter orbital energies due to their higher electronegativity and larger atomic radius, respectively.

Evaluating the generalizability of ML models on external datasets is crucial to ensure robustness and reliability in real-world applications. To curate our external dataset, we selected 18 molecules from the EEBL to form an external validation set with the goal of approximating equal representation of permeable and impermeable compounds. Special attention was given to selecting compounds with limited or poorly documented permeability data in the literature to challenge the models.

All model predictions for the six compounds, BMY 7378, beclomethasone dipropionate, fluocinolone acetonide,

Table 4 12 ground state QC properties of the QM9 and QM9-extended Datasets

Molecular Property	Description
mu	Dipole moment
alpha	Norm of static polarizability
HOMO	Highest unoccupied molecular orbital
LUMO	Lowest unoccupied molecular orbital
gap	Difference between HOMO and LUMO
<R ² >	Electronic spatial extent
ZPVE	Zero-point vibrational energy
cv	Heat capacity
U0	Internal energy at 0 K
U298	Internal energy at 298.15 K
G298	Free energy at 298.15 K
H298	Enthalpy at 298.15 K

The QM9 and QM9-extended dataset provides QC properties of small organic molecules, which are essential for Understanding molecular structure, behavior, and interactions. Electronic properties include dipole moment, polarizability, HOMO, LUMO, gap. Spatial and structural properties include <R²> and ZPVE. Thermodynamic properties include heat capacity, internal energies, Gibbs free energy, and enthalpy

biotin (vitamin B7), phenacetin, and salicylanilide, aligned with the results of the in vitro PAMPA assay for positive BBB permeability. These compounds share the common characteristics of lipophilicity and small molecular size. BMY 7378, a selective α 1D-adrenoceptor antagonist and partial agonist/antagonist of the 5-HT_{1A} receptor, is an experimental drug studied for its potential use for treating hypertension as well as dual angiotensin-converting enzyme (ACE) inhibition [33]. There is currently limited knowledge and data detailing its BBB permeability in widely accessible sources or databases. Fluocinolone acetonide shares structural traits with diazepam, a compound from the training set, including a steroid backbone and fused rings, resulting in high polarizability and moderate lipophilicity. These features support its correct classification as BBB-permeable. Biotin (vitamin B7) is known to function as a cofactor for several carboxylation reactions in the brain and its mechanism of entry via saturable transport systems into and from the CNS has been studied [34, 35]. Phenacetin, a non-opioid analgesic, has been shown to have some degree of CNS effects including relaxation, drowsiness, euphoria supporting its ability to cross the BBB [35]. Salicylanilides are a class of compounds that have demonstrated antimicrobial properties but are also being investigated for their anticancer properties. The application of salicylanilide derivatives for treatment of glioblastoma, a brain tumor in which drug penetration of the BBB is needed for therapeutic effect, is an area of active medicinal chemistry research [36].

All model predictions for sulfamerazine and diminazene aceturate were consistent with the PAMPA assay results of BBB impermeability [37]. Sulfamerazine, a sulfonamide

antibiotic used to treat bacterial infections, is not typically used as a first-line agent for CNS infections, as other antibiotics with better BBB penetration are usually preferred [38]. Diminazene aceturate is a drug commonly used to treat trypanosomiasis (African sleeping sickness) and babesiosis, acting mainly on the bloodstream and tissues with limited central effect [39].

Transfer learning models performed well in aligning with in vitro PAMPA assay results for the compounds being investigated for cancer research and treatment. Transfer learning models are the only models shown in Fig. 4 that accurately predicted the impermeability of 17 DMAG (Alvespimycin) HCl, a heat shock protein 90 (HSP90) inhibitor [40, 41] with potential antineoplastic activity via the mechanism of reactive oxygen species (ROS) generation. This compound has known challenges in crossing barriers such as the BBB, likely due to its large polar surface area and high molecular weight [41]. All transfer learning models, excluding the QM9 dipole moment property, also do well to predict the impermeability of YM155 (sepantronium bromide). The ability of the QM9-extended dipole moment model to correctly predict sepantronium bromide's impermeability highlights the importance of the expanded chemical space of S and Cl atoms in the training dataset. Sepantronium bromide contains a quaternary ammonium group, imparting a separation of positive and negative charges and high dipole moment, properties that align with its non-permeability. This compound shows promise for antineoplastic activity, particularly against non-small cell lung cancer (NSCLC), prostate cancer, and melanoma. For obatox mesylate, a compound being explored for its Bcl-2 antagonism, transfer learning models to QM9-ext properties did well to predict its impermeability. The BBB permeabilities of both sepantronium bromide and obatox mesylate are not yet extensively documented in the literature.

Clofoctol is an antibiotic that has been investigated for its potential to treat gliomas, however, specific quantitative data on its BBB permeability are limited [42]. The D-MPNN, DNN, SVM outperform the transfer learning models to support the PAMPA assay in clofoctol's permeability, which is likely due to its small molecular weight and hydrophobicity. Amino-thiazole [43] is another agent that has served as a scaffold in medicinal chemistry and shown diverse clinical applications including infection, cancer, inflammation, and Alzheimer's, a neurological disorder in which BBB penetration is essential for therapeutic efficacy [44]. Except for the D-MPNN, all models predicted amino-thiazole's permeability. Secnidazole is a nitroimidazole antimicrobial agent primarily used to treat parasitic infections like amebiasis, giardiasis, and bacterial vaginosis [45]. Secnidazole's impermeability was accurately predicted by the transfer learning models to QM9-extended properties

of polarizability, dipole moment, and HOMO. Furaltadone HCl was predicted to be impermeable across all models.

Excluding the transfer learning model to QM9 polarizability, all models predicted BBB permeability of dimethyl fumarate, a drug whose ability to exert effects within the CNS is essential for treating the neurological condition, multiple sclerosis. For bardoxolone methyl, the deep learning models (D-MPNN, DNN, and 5 out of the 6 transfer learning models to QM9-extended properties) were consistent with the PAMPA assay result of impermeability. Bardoxolone is a novel drug initially developed as a modulator of inflammation-associated carcinogenesis by inducing the Nrf2 pathway, inhibiting NF- κ B, leading to antioxidant and anti-inflammatory effects [46].

To evaluate the contribution of QC-derived embeddings, we performed analyses for contribution beyond traditional descriptors, descriptor ablation experiments, and comparison for alternate pretraining tasks. The results of direct comparison between descriptor-only models and QC-based transfer learning suggests that while LogP and MW remain strong contributors to BBB prediction, QC-derived embeddings may provide complementary molecular-level information that is not fully captured by traditional descriptors. The gains were modest on the B3DB test set but more pronounced on the external PAMPA curated test set, indicating that QC-based transfer learning has the potential to improve generalizability.

Feature ablation analyses further confirmed that the predictive signal from QC-derived embeddings is not solely attributable to lipophilicity or molecular weight. When LogP and MW were removed from the descriptor set, QC-transfer learning models retained strong performance across both the B3DB and PAMPA test sets, with only modest decreases in Accuracy and ROC-AUC. This finding demonstrates that QC embeddings contribute unique information, likely reflecting electronic and quantum mechanical properties that influence passive diffusion and membrane interactions, which are not captured by traditional descriptors.

To further isolate the effect of QC-based pretraining, we compared models pretrained on QC properties (QM9 and QM9-extended) to models pretrained on P-glycoprotein (P-gp) inhibition, a biologically relevant property affecting drug absorption and brain penetration. P-gp was chosen as a comparison because it represents a task-specific endpoint directly related to CNS drug disposition. QC-transfer learning models showed modestly higher performance on the B3DB test set and substantially better performance on the external PAMPA curated test set, with some embeddings achieving near-perfect predictions, whereas P-gp pretrained models performed relatively poorly. These results suggest that QC-derived embeddings may capture molecular

features that are broadly informative for BBB permeability and not limited to task-specific signals.

We acknowledge that the experimental dataset used for model validation includes only 18 compounds; however, the objective was to experimentally determine the BBB permeability of select compounds using PAMPA, particularly those for which permeability data were previously unavailable. To further evaluate the robustness and generalizability of the developed models beyond this small validation set, we additionally assessed model performance using a scaffold split, which partitions molecules by structural scaffolds to simulate prediction on unseen ones. As expected, model performance modestly decreased relative to random splits, reflecting the more stringent and realistic nature of scaffold-based evaluation. Nonetheless, the models maintained consistent predictive ability across both evaluation schemes, suggesting that they capture transferable molecular features rather than relying on scaffold-specific correlations. Future studies will aim to expand this experimentally derived test set and complement PAMPA-BBB assays with cell-based and in vivo models to account for transporter-mediated effects and in situ permeability dynamics.

A limitation of this study is that both the computational models and the in vitro experiments assume an intact BBB. While this assumption holds true in many physiological and pathological contexts, there are scenarios where the BBB is compromised. For instance, in gliomas and other neurological conditions such as stroke, multiple sclerosis, or traumatic brain injury, the BBB integrity is disrupted, allowing for altered transport dynamics and permeability. These disturbances may significantly impact the applicability and accuracy of the models in predicting drug delivery or other interactions in such cases. Future work will focus on developing models and curating datasets that account for BBB disruption to better reflect the complexities of these pathological conditions.

Ultimately, our approach has promising potential to extrapolate effectively to novel compounds with no existing BBB permeability data, enabling the identification of candidates without the need for extensive testing and allowing the design of compounds with an increased probability of success. As we continue to refine our models and expand data collection, this synergy of computation and experimentation could accelerate drug discovery, enabling faster, more cost-effective, and targeted identification of CNS-active or CNS-sparing drugs.

Experimental section

Datasets

B3DB. The dataset used for model development is B3DB, one of the largest public BBB benchmark datasets consisting of 7,807 compounds compiled from 50 published resources with accompanying precomputed 2D chemical descriptors from the Mordred package (Table 1) [13]. The set of 7,807 compounds has categorical values of BBB permeability (4956 BBB-permeating and 2851 BBB-non-permeating) and as stated in their publication, compounds were cleaned of salts and solvents, and those with heavy metal atoms were removed. It consists of a range of heavy atoms from 1 to 136 and the heavy atoms C, N, O, F, S, Cl, Br. We divided the B3DB dataset into a random train/test 75/25 split and applied a data balancing technique to the training set. Using a balanced dataset is essential in preventing bias of the model towards the more dominant class of compounds. To achieve a more balanced 1:1 ratio of BBB+:BBB-, we applied the random oversampling technique to the training split in which instances within the BBB- class were duplicated.

To evaluate model generalizability across diverse chemical structures, we performed a scaffold-based split of the dataset using Bemis–Murcko scaffolds. Molecules sharing the same scaffold were assigned together to either the training or test set, ensuring that structurally similar compounds were not divided across sets. Approximately 75% of compounds were allocated to training and 25% to testing, resulting in 5,853 molecules in the training set and 1,952 in the test set. Two molecules with invalid SMILES strings were excluded.

QM9. The QM9 dataset is a widely used benchmark dataset in computational chemistry and machine learning consisting of 133,885 molecules [24]. The molecules are derived from the GDB-17 chemical space and contain up to 9 heavy atoms (carbon, oxygen, nitrogen, fluorine). Each molecule was optimized at the B3LYP/6-31G(2df, p) level of theory and twelve ground-state quantum chemical properties are provided: dipole moment, isotropic polarizability, energy of HOMO, energy of LUMO, energy of gap, electronic spatial extent, zero point vibrational energy (ZPVE), heat capacity, and atomization energies (U0, U298, G298, H298 (Table 4). This dataset was utilized for source domain model development in transfer learning and was randomly split into train/test 80/20.

QM9-Extended. The QM9-extended dataset expands beyond the QM9 chemical space and enables greater application to drug discovery by covering an additional ~20,000 molecules containing sulfur and chlorine atoms [21]. The GDB-17 space was filtered for compounds with S or Cl and

no more than nine heavy atoms. Each compound underwent a low-mode conformational search with the MMFF94x and subsequent geometry optimization at the same level of theory as the QM9 data set, B3LYP/6-31G(2df, p). This dataset was utilized for source domain model development in transfer learning and was randomly split into train/test 80/20.

P-gp (P-glycoprotein) Inhibition. We used the P-glycoprotein (Pgp) inhibition dataset curated by Broccatelli et al., comprising 1,212 compounds with experimentally determined activity. Pgp is an ABC transporter that influences drug absorption, metabolism, and brain penetration, and its inhibition can affect bioavailability and multidrug resistance. The task is binary classification: given a compound's SMILES string, predict whether it inhibits Pgp. A D-MPNN was trained on this dataset to generate predictions for QM9 and QM9-extended compounds, which were then used in a transfer learning framework: first pretraining on Pgp inhibition as a source task and subsequently fine-tuning for BBB permeability prediction [47].

Molecular representation

In this study, we explored a range of molecular representations with different machine learning models to determine which representations performed best, depending on the complexity of the model used. Below, we describe the representations used in our experiments: molecular descriptors, Morgan fingerprints, combined descriptors and Morgan fingerprints, and learned embeddings.

Molecular Descriptors. The first descriptor set, which we refer to as “feature set 1”, serving as inputs to the models consists of MW, LogP, TPSA, and HBD. These were selected based on the CNS drug space as defined by the 6 physicochemical properties: ClogP, ClogD, MW, TPSA, HBD, pKa [48]. These features were normalized for consistency and to improve model performance. Another descriptor set included 208 two-dimensional physicochemical descriptors computed using RDKit. Interpolation was applied to address any missing values (Na) for descriptors that could not be calculated.

Morgan Fingerprints. Morgan fingerprints encode which substructures are within a molecule within a certain radius of organic molecule bonds [49]. Morgan fingerprints were generated using the RDKit library and computed with a radius of 2 and fingerprint size of 2048 bits. The radius defines the number of bonds to include around each atom, capturing increasingly larger substructures as the radius increases. A radius of 2 was chosen to balance computational efficiency and the level of structural detail captured. A fingerprint size of 2048 bits was selected to provide a sufficiently large bit vector for distinguishing between different

molecular structures while maintaining a reasonable computational cost.

Molecular Descriptors combined with Morgan fingerprints. We adopted the approach demonstrated by Yuan et al., combining molecular property-based descriptors and molecular fingerprints to improve the prediction accuracy of BBB permeability [50]. As described above, the 208 2D RDKit descriptors served as the molecular property-based descriptors. The molecular descriptors and Morgan fingerprints were concatenated into a unified feature set for each compound. The values of property-based descriptors were normalized, and the Boolean values of fragment-based descriptors were not.

Learned Embeddings. Mol2Vec is an unsupervised machine learning approach inspired by the Word2Vec model that learns vector representations of molecular substructures. It encodes molecular structures into continuous vector representations, which can be used as input features for various predictive tasks in cheminformatics [51]. The pre-trained Mol2Vec model, which was trained on 20 million compounds with radius of 1 for Morgan fingerprints, was used to generate embeddings to serve as inputs to the models in this study.

Model development

Support vector machine (SVM). Hyperparameter optimization was conducted using grid search to identify the best combination of the kernel, regularization parameter (C), kernel coefficient (gamma), and degree (for the polynomial kernel). The grid search explored a range of values for C (e.g., 1, 10, 100, 1000), gamma (e.g., 0.01, 0.1, 0.02, 0.2, 0.03, 0.3, 0.04, 0.4, 0.05, 0.5, 0.6, 0.7, 0.8, 0.9), and degree (e.g., 2, 3, 4). The optimized SVM model was trained using the best hyperparameters identified in the grid search (Table 5).

DNN. We conducted manual hyperparameter optimization using a series of trial runs. Key hyperparameters including learning rate, number of hidden layers, number of neurons per layer, batch size, and dropout rate were varied individually and in combination based on prior domain

Table 5 Grid search optimized SVM hyperparameters for each molecular Representation

Molecular Representation	kernel	C	gamma	degree
Feature set 1	rbf	1000	0.9	Na
2D RDKit	poly	1	0.02	4
Morgan fingerprint	poly	100	0.03	4
2D RDKit+Morgan fingerprint	poly	1	0.01	4
Mol2Vec	poly	1	0.03	4

Hyperparameter optimization was conducted for each of the molecular representations using grid search to identify the best combination of the kernel, regularization parameter (C), kernel coefficient (gamma), and degree (for the polynomial kernel)

knowledge and empirical performance. A DNN with a feed-forward architecture consisting of four layers with 64, 32, 16, and 1 neurons, respectively, was constructed. ReLU (Rectified Linear Unit) activation functions were applied to all hidden layers and dropout regularization with a rate of 0.1 was applied to prevent overfitting. The model was trained using the Adam optimizer with an initial learning rate of 0.001 and a batch size of 32. Training was conducted for 1000 epochs, with the learning rate dynamically adjusted using a learning rate step scheduler to ensure convergence. Binary cross-entropy loss was employed as the loss function.

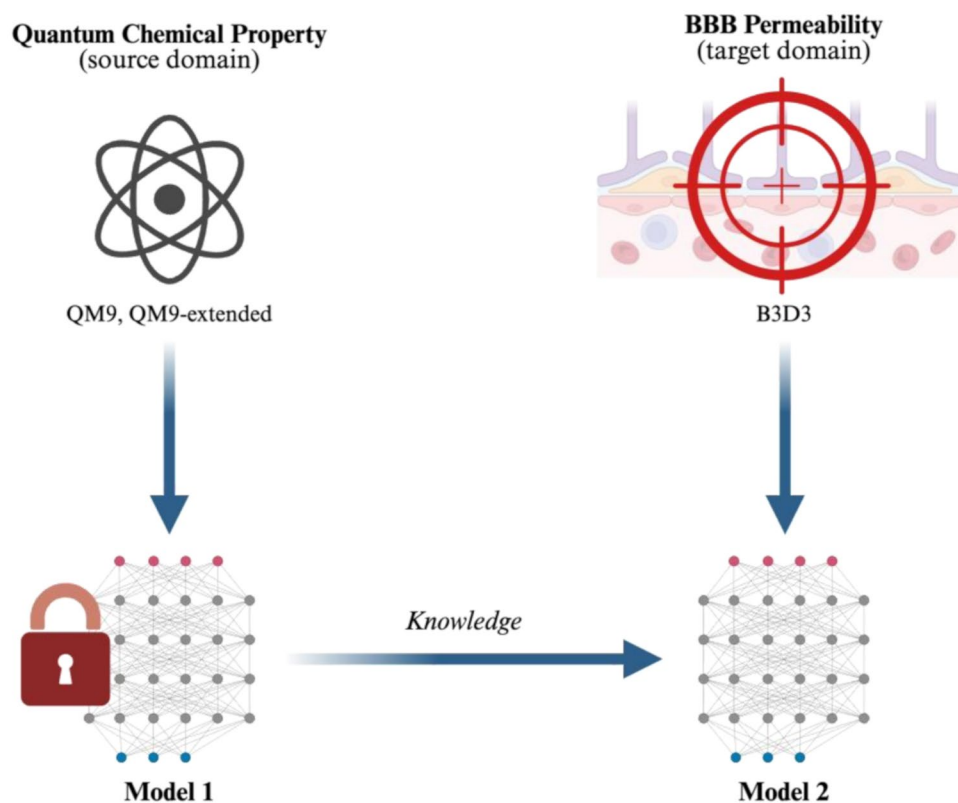
Baseline DNN. To enable direct comparison with the transfer learning architecture, we trained a baseline DNN on the BBB permeability dataset using the same layer structure as the transfer learning model. The network consisted of six fully connected layers: three initial layers for feature encoding ([256, 128, 64]) followed by a classification head ([256, 128, 1]). ReLU activations were applied after each hidden layer, and a dropout rate of 0.2 was used to mitigate overfitting. The final layer employed a sigmoid activation to output the probability of BBB permeability. The model was trained using the Adam optimizer (initial learning rate=0.0001) with a step learning rate scheduler (step size=250) and a batch size of 32 for 800 epochs. This architecture served as a baseline control to evaluate the performance gains attributable to transfer learning.

D-MPNN. To build GCNN models, we used D-MPNN as implemented in Chemprop (<https://github.com/chemprop>) [52]. The D-MPNN network architecture parameters were hyperparameter optimized for the B3DB data set using Bayesian optimization, and the optimal hyperparameters (depth: 6, hidden size: 1000, number of fully connected layers: 3, dropout: 0.25) were then used to train the model. Additional featurization with RDKit features as provided with Chemprop was also explored.

Transfer Learning. The transfer learning approach applied in this study is outlined in Fig. 8 and involves two DNN models trained sequentially on different but related tasks. Model 1 was trained on the source task, the regression of QC properties using either the QM9 or QM9-extended dataset. This model was a feedforward DNN consisting of four fully connected layers with hidden dimensions of [256, 128, 64] and a single output neuron (for the predicted QC property). The network was trained using the Adam optimizer with an initial learning rate of 0.0001 and a batch size of 32.

After convergence, the learned representations from this source model were transferred to Model 2, which was designed for a target task: binary classification of BBB permeability. To enable transfer learning, all layers in Model 1 except the output layer were preserved and their weights

Fig. 8 Transfer learning from QM property to BBB permeability. Illustration of the transfer learning methodology applied in this study. Knowledge is learned from a source domain (quantum chemical properties) and transferred to a target domain (BBB permeability). The lock icon represents the pretrained model's frozen layers that are not updated during training to the new task. Created with [BioRender.com](https://www.biorender.com/)



frozen, meaning they were excluded from gradient updates during further training. The original output layer, used for the regression task, was removed and replaced with a new classification head consisting of three layers [256, 128, 1] with ReLU activations in the hidden layers and a sigmoid activation in the final layer for binary output. Model 2 was trained on the BBB dataset using the Adam optimizer (initial learning rate of 0.0001), a step learning rate scheduler (step size = 250), and a batch size of 32 for 800 epochs.

Learning was transferred by reusing the feature representations learned from the QC property prediction task, with the assumption that these representations capture molecular features relevant to both source and target tasks. Fine-tuning was then applied only to the new layers, adapting the previously learned molecular embeddings to the BBB classification task. This transfer was achieved not merely by copying weights, but by repurposing the frozen layers as a molecular feature extractor, while allowing the new classification layers to specialize to the BBB prediction objective.

In terms of architectural flexibility, while the backbone architecture of the frozen layers remained constant between Models 1 and 2, the classification head (new output layers) was designed independently. Therefore, it is not necessary for the entire architecture, particularly the output layers, to match between the source and target models. Only the encoder portion (shared layers) needs to remain unchanged

to enable proper weight reuse and compatibility during transfer.

Data analysis

All data was characterized as a continuous or ordinal variable, and then, further analyzed with GraphPad Prism (GraphPad Software, Boston, MA, USA). For model performance, binomial 95% confidence intervals of the mean of F1-score, ROC-AUC, and accuracy data were analyzed for each model [53].

Curation of external validation dataset with PAMPA-BBB

The EEBL comprises 2,036 diverse small molecules with confirmed biological and pharmacological activities, including 1,018 FDA-approved compounds. Datapoints included in the EEBL are these specific small molecules. These molecules target over 200 distinct proteins involved in more than 20 signaling pathways, including those related to survival and apoptosis [54].

The eighteen commercial drugs from the curated test set were obtained from MedChemExpress (Monmouth Junction, NJ, USA). These drugs were chosen after stratification of drugs based off predicted BBB permeability and formed an external validation set with the goal of approximating

equal representation of permeable and impermeable compounds. The Parallel Artificial Membrane Permeability Assay-BBB Kit (PMBBB-096) was obtained from BioAssay Systems (Hayward, CA, USA). The PAMPA kit allows for the quantitative determination of BBB permeability through the usage of an artificial lipid membrane. This kit contains a donor plate, acceptor plate, working tray, dodecane, dried brain lipids, high permeability control, and low permeability control. The PAMPA was conducted as described in the protocol and past literature and absorbance was measured with the BioTek Cytation 5 Cell Imaging Multimode Reader (Agilent Technologies, Inc., Santa Clara, CA, USA) [55]. After absorbance reading, the permeability equation provided in the PAMPA protocol, the high permeability control, and the low permeability control were used to assign “high permeability” (BBB+), “average permeability” (BBB+/-), or “poor permeability” (BBB-) to each test compound (Supplementary Table 1):

‘BBB+’ (high BBB permeation predicted); $Pe (10^{-6} \text{ cm s}^{-1}) > 4.0$.

‘BBB-’ (low BBB permeation predicted); $Pe (10^{-6} \text{ cm s}^{-1}) < 1.5$.

‘BBB+/-’ (BBB permeation average); $Pe (10^{-6} \text{ cm s}^{-1})$ from 4.0 to 1.5.

Abbreviations

BBB	Blood-brain barrier
CNS	Central nervous system
QC	Quantum chemical
PAMPA-BBB	Parallel Artificial Membrane Permeability Assay-BBB
B3DB	Blood-Brain Barrier Database
SVMs	Support vector machines
DNNs	Deep neural networks
D-MPNNs	Direct message passing neural networks
EEBL	Emory Enriched Bioactive Library
ML	Machine learning
DL	Deep learning
Mol2Vec	Learned vector representations of molecular substructures
ROC-AUC	Area under the receiver operating characteristic curve
ACE	Angiotensin-converting enzyme
HSP90	Heat shock protein 90
ROS	Reactive oxygen species
ZPVE	Zero point vibrational energy
P-gp	P-glycoprotein
M155	Sepantronium bromide
NSCLC	Non-small cell lung cancer
ReLU	Rectified Linear Unit
MW	Molecular weight

TPSA
HBD

Topological surface area
Hydrogen bond donors

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s13346-025-02005-5>.

Author contributions The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. W.H.M., M.A.L., and K.L. conceived the project. M.A.L. and M.G.Y. designed and performed the experiments, analyzed the data, manuscript drafting and made figures. K.B.H., A.M.M., M.R., Y.D., N.M.B., and K.L. critically read the manuscript.

Funding This research is financially supported by the National Cancer Institute (NCI), National Institutes of Health (NIH), under grant R01CA251393. The publication also acknowledges support from the National Institute of General Medical Sciences, National Institutes of Health, through grant number R25GM143298, Georgia CTSA UL1 (UL1TR002378) and KL2 (KL2TR002381).

Data availability The 18 compounds tested in the PAMPA assay are available in the Supplementary Information. Upon publication, the scripts, B3DB dataset, and computed descriptors (“features1”, 2D RDKit, Mol2Vec, and Morgan FP) will be made publicly accessible via <https://github.com/meganlim>. Software used to train the GCNN models is freely available from Yang et al. at <https://github.com/chemprop>. Mol2Vec is freely available from Jaeger et al. at <https://github.com/samoturk/mol2vec>.

Data availability (Required): The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare that they have no known competing financial interests that could have appeared to influence the work reported in this paper.

References

- Kadry H, Noorani B, Cucullo L. A blood-brain barrier overview on structure, function, impairment, and biomarkers of integrity. *Fluids Barriers CNS*. 2020;17:69.
- Zhao Y, Gan L, Ren L, Lin Y, Ma C, Lin X. Factors influencing the blood-brain barrier permeability. *Brain Res*. 2022;1788:147937.
- Harilal S, Jose J, Parambi DGT, Kumar R, Unnikrishnan MK, Uddin MS, Mathew GE, Pratap R, Marathakam A, Mathew B. Revisiting the blood-brain barrier: A hard nut to crack in the transportation of drug molecules. *Brain Res Bull*. 2020;160:121–40.
- Zhang S, Gan L, Cao F, Wang H, Gong P, Ma C, Ren L, Lin Y, Lin X. The barrier and interface mechanisms of the brain barrier, and brain drug delivery. *Brain Res Bull*. 2022;190:69–83.
- Wu D, Chen Q, Chen X, Han F, Chen Z, Wang Y. The blood-brain barrier: structure, regulation, and drug delivery. *Signal Transduct Target Ther*. 2023;8:217.

6. Han L, Jiang C. Evolution of blood-brain barrier in brain diseases and related systemic nanoscale brain-targeting drug delivery strategies. *Acta Pharm Sin B*. 2021;11:2306–25.
7. Bayir E, Sendemir A. Vitro human Blood-Brain barrier model for drug permeability testing. *Methods Mol Biol*. 2021;2367:73–85.
8. Canfield SG, Stebbins MJ, Morales BS, Asai SW, Vatine GD, Svendsen CN, Palecek SP, Shusta EV. An isogenic blood-brain barrier model comprising brain endothelial cells, astrocytes, and neurons derived from human induced pluripotent stem cells. *J Neurochem*. 2017;140:874–88.
9. Vatansever S, Schlessinger A, Wacker D, Kaniskan HU, Jin J, Zhou MM, Zhang B. Artificial intelligence and machine learning-aided drug discovery in central nervous system diseases: State-of-the-arts and future directions. *Med Res Rev*. 2021;41:1427–73.
10. Clegg LE, Mac F, Gabhann. Molecular mechanism matters: benefits of mechanistic computational models for drug development. *Pharmacol Res*. 2015;99:149–54.
11. Singh M, Divakaran R, Konda LSK, Kristam R. A classification model for blood brain barrier penetration. *J Mol Graph Model*. 2020;96:107516.
12. Kumar R, Sharma A, Alexiou A, Bilgrami AL, Kamal MA, Ashraf GM. DeePred-BBB: A blood brain barrier permeability prediction model with improved accuracy. *Front Neurosci*. 2022;16:858126.
13. Meng F, Xi Y, Huang J, Ayers PW. A curated diverse molecular database of blood-brain barrier permeability with chemical descriptors. *Sci Data*. 2021;8:289.
14. Kumar V, Banerjee A, Roy K. Breaking the barriers: Machine-Learning-Based c-RASAR approach for accurate Blood-Brain barrier permeability prediction. *J Chem Inf Model*. 2024;64:4298–309.
15. Shaker B, Yu MS, Song JS, Ahn S, Ryu JY, Oh KS, Na D. Light-BBB: computational prediction model of blood-brain-barrier penetration based on LightGBM. *Bioinformatics*. 2021;37:1135–9.
16. Huang ETC, Yang JS, Liao KYK, Tseng WCW, Lee CK, Gill M, Compas C, See S, Tsai FJ. Predicting blood-brain barrier permeability of molecules with a large Language model and machine learning. *Sci Rep*. 2024;14:15844.
17. Zhu X, Chen Y, Gu Y, Xiao Z. SentiMedQAer: A transfer Learning-Based Sentiment-Aware model for biomedical question answering. *Front Neurorobot*. 2022;16:773329.
18. Varde AS, Karthikeyan D, Wang W. Facilitating COVID recognition from X-rays with computer vision models and transfer learning. *Multimed Tools Appl*. (2023) 1–32.
19. Tao H, Wang Y, Zhuang Z, Fu H, Guo X, Zou S. Cross-Corpus Speech Emotion Recognition Based on Transfer Learning and Multi-Loss Dynamic Adjustment. *Comput Intell Neurosci*. 2022;2022:5019384.
20. Cai C, Wang S, Xu Y, Zhang W, Tang K, Ouyang Q, Lai L, Pei J. Transfer learning for drug discovery. *J Med Chem*. 2020;63:8683–94.
21. Lim MA, Yang S, Mai H, Cheng AC. Exploring deep learning of quantum chemical properties for Absorption, Distribution, Metabolism, and excretion predictions. *J Chem Inf Model*. 2022;62:6336–41.
22. Oguma T, Uehara S, Nakahara K, Okuyama Y, Fuchino K, Suzuki N, Kan Y, Kanegawa N, Ogata Y, Kusakabe KI. A quantum Mechanics-Based method to predict intramolecular hydrogen bond formation reflecting P-glycoprotein recognition. *ACS Med Chem Lett*. 2023;14:223–8.
23. Zehui Z, Laith A, Jinglan Z, Ye D, Yuantong G. A comparison review of transfer learning and self-supervised learning: Definitions, applications, advantages and limitations. *Expert Syst Appl*. 2024;242:122807.
24. Ramakrishnan R, Dral PO, Rupp M, von Lilienfeld OA. Quantum chemistry structures and properties of 134 Kilo molecules. *Sci Data*. 2014;1:140022.
25. Chuang KV, Gunsalus LM, Keiser MJ. Learning molecular representations for medicinal chemistry. *J Med Chem*. 2020;63:8705–22.
26. Stepisnik T, Skrlj B, Wicker J, Kocev D. A comprehensive comparison of molecular feature representations for use in predictive modeling. *Comput Biol Med*. 2021;130:104197.
27. Li J, Lei K, Wu Z, Li W, Liu G, Liu J, Cheng F, Tang Y. Network-based identification of MicroRNAs as potential Pharmacogenomic biomarkers for anticancer drugs. *Oncotarget*. 2016;7:45584–96.
28. Calzetta L, Matera MG, Facciolo F, Cazzola M, Rogliani P. Beclomethasone dipropionate and formoterol fumarate synergistically interact in hyperresponsive medium bronchi and small airways. *Respir Res*. 2018;19:65.
29. Narapureddy B, Dubey D. Clinical evaluation of dimethyl fumarate for the treatment of relapsing-remitting multiple sclerosis: efficacy, safety, patient experience and adherence. *Patient Prefer Adherence*. 2019;13:1655–66.
30. Mellatyar H, Talaei S, Pilehvar-Soltanahmadi Y, Barzegar A, Akbarzadeh A, Shahabi A, Barekati-Mowahed M, Zarghami N. Targeted cancer therapy through 17-DMAG as an Hsp90 inhibitor: overview and current state of the Art. *Biomed Pharmacother*. 2018;102:608–17.
31. Goard CA, Schimmer AD. An evidence-based review of Obactox mesylate in the treatment of hematological malignancies. *Core Evid*. 2013;8:15–26.
32. Hosna A, Merry E, Gyalmo J, Alom Z, Aung Z, Azim MA. Transfer learning: a friendly introduction. *J Big Data*. 2022;9:102.
33. Rodriguez JE, Andrade-Jorge E, Barquet-Nieto A, Estrada-Soto SE, Gallardo-Ortiz IA, Villalobos-Molina R. BMY 7378, a selective $\alpha(1D)$ -adrenoceptor antagonist, is a new angiotensin converting enzyme inhibitor: in silico, in vitro and in vivo approach. *Biochim Biophys Acta Gen Subj*. 2025;1869:130732.
34. Spector R, Mock D. Biotin transport through the blood-brain barrier. *J Neurochem*. 1987;48:400–4.
35. Piper BJ, Alinea AA, Wroblewski JR, Graham SM, Chung DY, McCutcheon LRM, Birkett MA, Kheloussi SS, Shah VM, Szarek JL, Zalim QK, Arnott JA, McLaughlin WA, Lucchesi PA, Miller KA, Waite GN, Bordonaro M. A Quantitative and Narrative Evaluation of Goodman and Gilman's Pharmacological Basis of Therapeutics, Pharmacy (Basel). 2019;8.
36. Horvath L, Biri-Kovacs B, Baranyai Z, Stipsicz B, Mehes E, Jezso B, Kratky M, Vinsova J, Bosze S. New Salicylanilide derivatives and their peptide conjugates as anticancer compounds: Synthesis, Characterization, and in vitro effect on glioblastoma. *ACS Omega*. 2024;9:16927–48.
37. Kato R, Zeng W, Siramshetty VB, Williams J, Kabir M, Hagen N, Padilha EC, Wang AQ, Mathe EA, Xu X, Shah P. Development and validation of PAMPA-BBB QSAR model to predict brain penetration potential of novel drug candidates. *Front Pharmacol*. 2023;14:1291246.
38. Vardanyan RS, Hruby VJ. 33 - Antimicrob Drugs, (2006) 499–523.
39. de Brito MG, Mengarda AC, Oliveira GL, Cirino ME, Silva TC, de Oliveira RN, Allegretti SM, de Moraes J. Therapeutic effect of diminazene aceturate on parasitic blood fluke schistosoma mansoni infection. *Antimicrob Agents Chemother*. 64 (2020).
40. Kim JG, Lee SC, Kim OH, Kim KH, Song KY, Lee SK, Choi BJ, Jeong W, Kim SJ. HSP90 inhibitor 17-DMAG exerts anticancer effects against gastric cancer cells principally by altering oxidant-antioxidant balance. *Oncotarget*. 2017;8:56473–89.
41. Hu ZY, Lu J, Zhao Y. A physiologically based Pharmacokinetic model of alvespimycin in mice and extrapolation to rats and humans. *Br J Pharmacol*. 2014;171:2778–89.
42. Hu Y, Zhang M, Tian N, Li D, Wu F, Hu P, Wang Z, Wang L, Hao W, Kang J, Yin B, Zheng Z, Jiang T, Yuan J, Qiang B, Han W,

- Peng X. The antibiotic clofexol suppresses glioma stem cell proliferation by activating KLF13. *J Clin Invest*. 2019;129:3072–85.
43. Farouk Elsadek M, Mohamed Ahmed B, Fawzi Farahat M. An overview on synthetic 2-Aminothiazole-Based compounds associated with four biological activities. *Molecules*; 2021. p. 26.
 44. Soni A, Kumar A, Kumar V, Rawat R, Eyupoglu V. Design, synthesis and evaluation of aminothiazole derivatives as potential anti-Alzheimer's candidates. *Future Med Chem*. 2024;16:513–29.
 45. Muzny CA, Van Gerwen OT, Legendre D. Secnidazole: a treatment for trichomoniasis in adolescents and adults. *Expert Rev Anti Infect Ther*. 2022;20:1067–76.
 46. Avula U.M.R., Harris L, Hassanein M. Bardoxolone for CKD: the paradox of confusion and dogma. *Kidney360*. 2022;3:1955–60.
 47. Broccatelli F, Carosati E, Neri A, Frosini M, Goracci L, Oprea TI, Cruciani G. A novel approach for predicting P-glycoprotein (ABCB1) Inhibition using molecular interaction fields. *J Med Chem*. 2011;54:1740–51.
 48. Wager TT, Chandrasekaran RY, Hou X, Troutman MD, Verhoest PR, Villalobos A, Will Y. Defining desirable central nervous system drug space through the alignment of molecular properties, in vitro ADME, and safety attributes. *ACS Chem Neurosci*. 2010;1:420–34.
 49. Liu X, Cheng Q, Sun D, Wei W, Zheng Z. Anonymous pattern molecular fingerprint and its applications on property identification. *IEEE/ACM Trans Comput Biol Bioinform*. 2023;20:3759–71.
 50. Yuan Y, Zheng F, Zhan CG. Improved prediction of Blood-Brain barrier permeability through machine learning with combined use of molecular Property-Based descriptors and fingerprints. *AAPS J*. 2018;20:54.
 51. Jaeger S, Fulle S, Turk S. Mol2vec: unsupervised machine learning approach with chemical intuition. *J Chem Inf Model*. 2018;58:27–35.
 52. Heid E, Greenman KP, Chung Y, Li SC, Graff DE, Vermeire FH, Wu H, Green WH, McGill CJ. Chemprop: A machine learning package for chemical property prediction. *J Chem Inf Model*. 2024;64:9–17.
 53. Smith AD, Genz A, Freiburger DM, Belenky G, Van Dongen HP. Efficient computation of confidence intervals for bayesian model predictions based on multidimensional parameter space. *Methods Enzymol*. 2009;454:213–31.
 54. Mo X, Tang C, Niu Q, Ma T, Du Y, Fu H. HTiP: High-Throughput Immunomodulator phenotypic screening platform to reveal IAP antagonists as Anti-cancer immune enhancers. *Cell Chem Biol*. 2019;26:331–e339333.
 55. Di L, Kerns EH, Fan K, McConnell OJ, Carter GT. High throughput artificial membrane permeability assay for blood-brain barrier. *Eur J Med Chem*. 2003;38:223–32.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.