Supplemental Information: Geometric Multi-color Message Passing Graph Neural Networks for Blood-brain Barrier Permeability Prediction

Trung Nguyen[1*], Md Masud Rana[2*], Farjana Tasnim Mukta[2], Chang-Guo Zhan[3], and Duc Duy Nguyen[4†]
[1]The Bredesen Center, University of Tennessee, Knoxville, TN 37996, USA
[2]Department of Mathematics, Kennesaw State University, Kennesaw, GA 30144, USA
[3]Department of Pharmaceutical Sciences, University of Kentucky, Lexington, KY 40506, USA
[4]Department of Mathematics, University of Tennessee, Knoxville, TN 37996, USA

This document provides supplemental information for the main manuscript, including details on model hyperparameters, data preprocessing pipelines, and comprehensive experimental results.

1 Methodological Details

This section provides additional mathematical details for the construction of the Weighted Colored Subgraphs (WCS) discussed in the main text.

In the BBBP datasets, we consider 12 atom types: C, H, O, N, P, Cl, F, Br, S, Si, I, and X (where X represents all other atom types). Thus, $C_1 = \mathrm{C}$, $C_2 = \mathrm{H}$, and so on.

In all subgraph constructions, we exclude covalent interactions where two atoms are too close to each other. To enforce this constraint, we set a cutoff such that if the distance between two atoms satisfies $\|r_i - r_j\| < r_{\mathrm{vdw}_i} + r_{\mathrm{vdw}_j} + \sigma$, the interaction is ignored. Here, σ is the standard deviation of all van der Waals radii in the given dataset.

The pairwise interaction of all the atoms in a subgraph $\mathcal{G}_{kk'}$ can be represented in the weighted adjacency matrix as follows:

$$\left(A_{kk'}(\Phi, \eta_{kk'})\right)_{ij} = \begin{cases} \Phi\left(\|\mathbf{r}_i - \mathbf{r}_j\|; \eta_{kk'}\right), & \text{if } (i,j) \in \mathcal{E}_{kk'} \\ 0, & \text{otherwise} \end{cases} \tag{S1}$$

The degree matrix $D_{kk'}$ is a diagonal matrix where each entry $(D_{kk'})_{ii}$ represents the sum of the weights of all edges connected to atom i. This value is used directly in the final feature construction and is computed as:

$$\left(D_{kk'}\right)_{ii} = \sum_{j \in \mathcal{V}_{kk'}} \left(A_{kk'}\right)_{ij} \tag{S2}$$

*These authors contributed equally to this work.
†Address correspondence to Duc Duy Nguyen. E-mail: ducnguyen@utk.edu

Finally, the Laplacian matrix $L_{kk'}$ is computed as:

$$L_{kk'} = D_{kk'} - A_{kk'} \tag{S3}$$

The adjacency and Laplacian matrices capture the structural properties of the subgraph and are useful for further spectral and geometric analysis of the graph.

2 Model Parametrization and Hyperparameter Optimization

Table S1: Hyperparameter search space for GMC-MPNN.

Hyperparameter	**Search Space**
Activation	RELU, LEAKYRELU, PRELU, TANH, SELU, ELU
Aggregation	MeanAggregation, SumAggregation, NormAggregation
Aggregation Norm	Uniform distribution: [1, 200] (step size = 1)
Batch Size	Discrete choices: 16, 32, 64, 128, 256
Depth	Random integer: [2, 6] (step size = 1)
Dropout	0.0, 0.05, 0.1, 0.15, ..., 0.45
FFN Hidden Dim	Random integer: [300, 2400] (step size = 100)
FFN Num Layers	Random integer: [1, 3]
Final LR Ratio	Log-uniform distribution: [1e-2, 1]
Message Hidden Dim	Random integer: [300, 2400] (step size = 100)
Init LR Ratio	Log-uniform distribution: [1e-2, 1]
Max LR	Log-uniform distribution: [1e-4, 1e-2]
Kernel type	Exponential, Lorentz
Scaling factor (τ)	0.5, 1.0, ..., 10 (Step size 0.5)
Power (κ)	0.5, 1.0, ..., 20 (Step size 0.5)

We systematically optimize hyperparameters using Optuna over the search space defined in Table S1. The search focuses on parameters such as activation functions, aggregation methods, model depth, dropout rates, feedforward network dimensions, learning rate schedules, and batch sizes. The search was performed on a cluster with a single NVIDIA H100 GPU (80GB VRAM). A complete search for one dataset required about 3-5 minutes.

Additionally, we also optimize WCS kernel parameters (β, κ, τ) to achieve robust atom-level descriptors. Here, β specifies the kernel type (generalized exponential, $\beta = E$, or Lorentz, $\beta = L$), τ acts as a scaling factor determining the characteristic distance $\eta_{kk'} = \tau(r_k + r_{k'})$, and κ is the power parameter of the kernel function [1].

We generated 1600 WCS feature variants, each corresponding to a unique combination of kernel parameters (β, κ, τ) selected from the grid $\tau \in [0.5, 10]$ and $\kappa \in [0.5, 20]$ (increments of 0.5). For each GMC-MPNN model variant (defined by the feature aggregation method, e.g., Max Features), we evaluated performance using all 1600 kernel configurations, selecting the kernel that yielded the best validation set score. This optimal configuration was then used for final benchmarking on the test set. Table S2 shows the best hyperparameters identified for the primary model variants on the

three datasets. The generated phase was performed on a cluster with a 40-core CPU, and each of the WCS features variants cost about 8 minutes per dataset.

Table S2: Best overall hyperparameter configurations found for the BBBP classification datasets ($BBBP_{\text{cls}}^{\text{MolNet}}$, $BBBP_{\text{cls}}^{\text{B3DB}}$) and B3DB regression dataset ($BBBP_{\text{reg}}^{\text{B3DB}}$).

Hyperparameter	$BBBP_{\mathbf{cls}}^{\mathbf{MolNet}}$	$BBBP_{\mathbf{cls}}^{\mathbf{B3DB}}$	$BBBP_{\mathbf{reg}}^{\mathbf{B3DB}}$
Activation	LEAKYRELU	RELU	PRELU
Aggregation	SumAggregation	MeanAggregation	MeanAggregation
Aggregation Norm	57	4	41
Batch Size	32	32	16
Depth	5	3	3
Dropout	0.0	0.0	0.0
FFN Hidden Dim	900	700	2200
FFN Num Layers	2	2	1
Final LR Ratio	0.104	0.024	0.033
Message Hidden Dim	2200	1300	1400
Init LR Ratio	0.011	0.041	1.0
Max LR	0.00521	0.00385	6.38e-04
Kernel Type	Exponential	Lorentz	Lorentz
Scaling Factor (τ)	0.5	0.5	7.5
Power (κ)	17.0	7.0	16.0

3 Dataset Preparation

The dataset $BBBP_{\text{cls}}^{\text{MolNet}}$ is the MoleculeNet [2] classification dataset comprising a total of 2,039 compounds, with 1,560 classified as positive (BBB+) and 479 as negative (BBB–). We clean the SMILES strings by removing isolated ions such as [H^+] or [Cl^-], along with their associated prefixes. The 3D structures for this dataset are provided by WeiLab [3].

Both the second classification dataset, $BBBP_{\text{cls}}^{\text{B3DB}}$, and the regression dataset, $BBBP_{\text{reg}}^{\text{B3DB}}$, are derived from the B3DB benchmark introduced by Meng *et al.* [4]. The classification dataset initially contained 7,807 compounds, while the regression dataset included 1,051 compounds annotated with SMILES strings and logBB values.

Since neither the B3DB dataset provides 3D molecular structures, we applied a consistent preprocessing pipeline to both. First, we cleaned the SMILES strings by removing isolated ions and then validated them using RDKit to ensure compatibility with graph-based models. Molecules that could not be parsed by RDKit were excluded (2 from the classification set, none from the regression set).

Next, to generate a single, low-energy 3D conformer for each molecule, we implemented a hierarchical pipeline. Our primary method utilized the OpenEye OMEGA Toolkit [5] to perform a knowledge-based conformational search and write the single, lowest-energy structure to a MOL2 file. For molecules failing with OMEGA, we employed an open-source fallback: RDKit generated an initial 3D structure (ETKDGv2), saved as PDB, which OpenBabel [6] converted to MOL2.

Despite this, 65 molecules in the classification set and 4 in the regression set failed conversion. Consequently, we obtained 7,740 valid molecules for $BBBP_{\text{cls}}^{\text{B3DB}}$ and 1,047 for $BBBP_{\text{reg}}$, used in

downstream experiments.

A summary of the datasets is provided in Table S3.

Table S3: Summary of datasets used in this study.

Dataset	Total Compounds	# of BBB+	# of BBB−	Dataset Type
$BBBP_{\text{cls}}^{\text{MolNet}}$	2,039	1,560	479	Classification
$BBBP_{\text{cls}}^{\text{B3DB}}$	7,740	4,905	2,835	Classification
$BBBP_{\text{reg}}^{\text{B3DB}}$	1,047	N/A	N/A	Regression

4 Sensitive Analysis

This section provides a comprehensive summary of the sensitivity analysis, illustrating the impact of varying τ and κ values on the final prediction performance. The following figures (Figures S1, Figures S2, and Figures S3) present a detailed analysis of the kernel when running different τ and κ values on the validation set for each dataset. Each figure provides a pair of plots for exponential kernels and Lorentz kernels, measuring by AUC-ROC for classification and RMSE for the regression task.

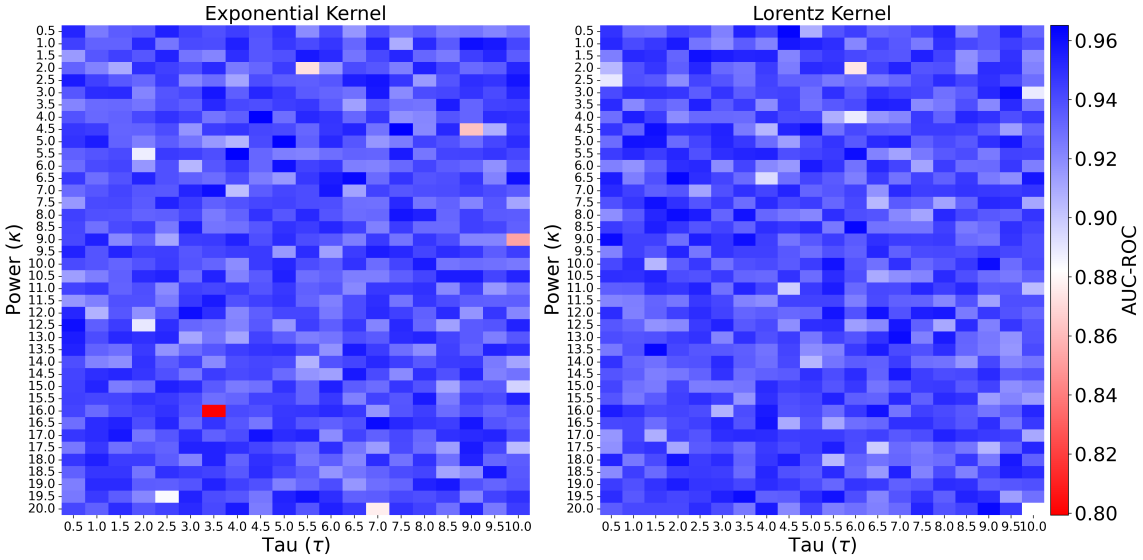


Figure S1: Sensitivity analysis of kernel hyperparameters on the validation set of the $BBBP_{\text{cls}}^{\text{MolNet}}$ dataset.

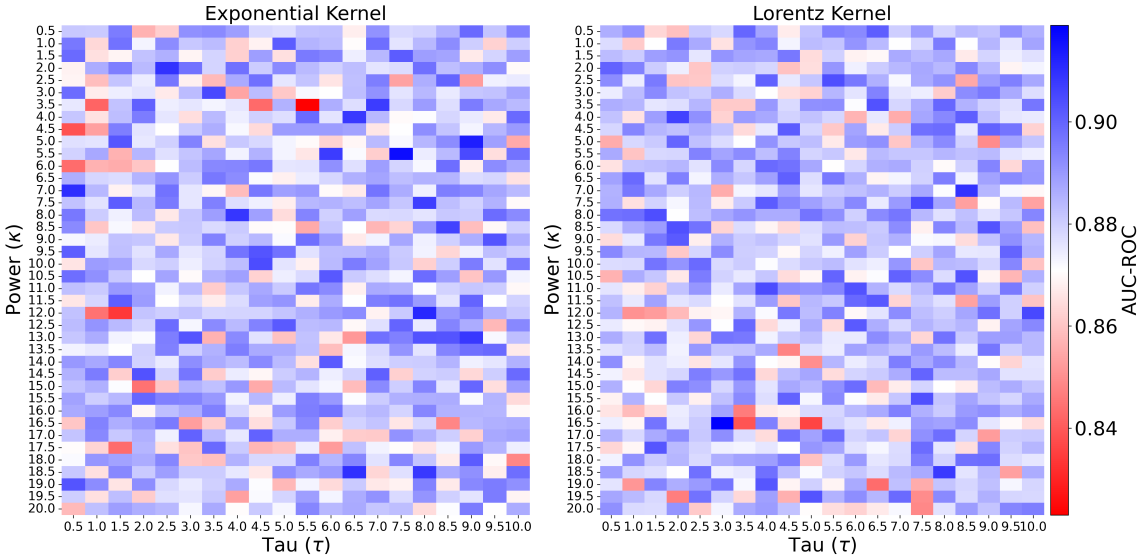


Figure S2: Sensitivity analysis of kernel hyperparameters on the validation set of the $BBBP^{\text{B3DB}}_{\text{cls}}$ dataset.

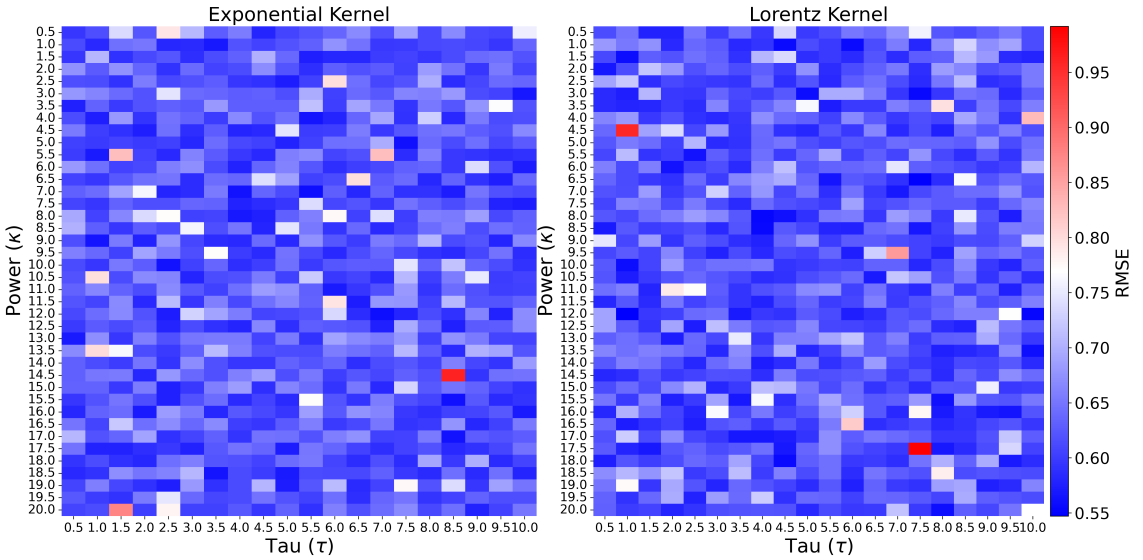


Figure S3: Sensitivity analysis of kernel hyperparameters on the validation set of the $BBBP^{\text{B3DB}}_{\text{regression}}$ dataset.

5 Ablation Study Results

Table S4 presents the statistical significance of the top 20 most impactful atom-pair interactions identified through our ablation study. A one-sample t-test was performed on the change in AUC-ROC over 20 runs against a null hypothesis of zero change, formally quantifying the contribution of each interaction type.

Table S4: Statistical significance of top 20 impactful atom-pair interactions from the ablation study.

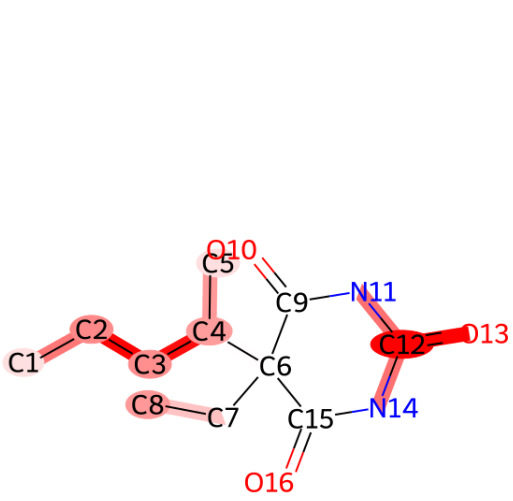
Ablated Atom Pair	**Mean Change in AUC**	**t-statistic**	**p-value**
N.ar-O.2	9.035	8.809	2.76×10^{-26}
B-C.3	9.033	9.367	8.62×10^{-27}
N.pl3-O.2	9.029	9.358	8.79×10^{-27}
N.3-O.3	9.023	9.913	2.95×10^{-27}
C.2-O.3	9.022	9.956	2.72×10^{-27}
F-O.3	9.019	9.159	1.32×10^{-26}
B-N.3	9.014	1.0668	7.33×10^{-28}
N.3-N.ar	9.014	1.0177	1.79×10^{-27}
N.3-N.am	9.013	9.362	8.72×10^{-27}
C.ar-N.2	9.012	1.2047	7.30×10^{-29}
N.2-N.2	9.009	9.726	4.23×10^{-27}
F-N.ar	9.006	1.0012	2.44×10^{-27}
N.am-O.2	9.005	1.0011	2.45×10^{-27}
N.3-N.pl3	9.002	9.530	6.22×10^{-27}
O.2-O.2	9.002	8.998	1.85×10^{-26}
N.2-N.am	8.997	1.0292	1.45×10^{-27}
B-F	8.997	1.0168	1.82×10^{-27}
C.3-S.o2	8.996	1.0493	1.00×10^{-27}
C.2-F	8.995	9.462	7.13×10^{-27}
Cl-N.3	8.995	9.252	1.09×10^{-26}

6 Model Interpretability

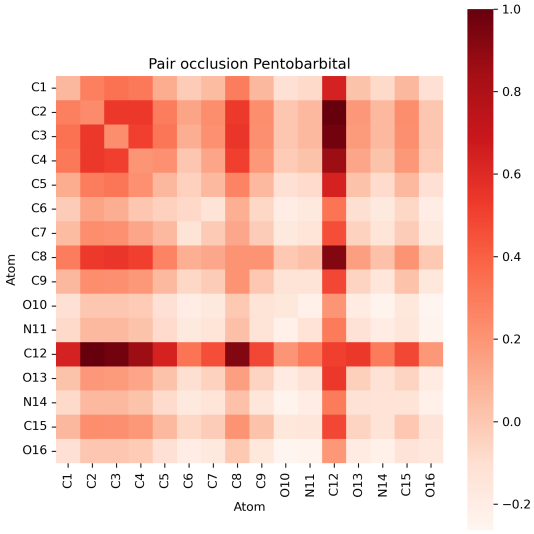
To provide insight into our model's predictions, we used an occlusion-based attribution method [7] to see how the model's prediction changes when specific input features are removed. In our implementation, we systematically zero out the WCS feature vectors associated with individual atoms and pairs of atoms. The importance score is calculated as the difference between the original prediction and the prediction after occlusion, measuring the model's reliance on the geometric information encoded in those WCS features.

We present 3 case studies, including Pentobarbital (positive, BBB+), M2L-663581 (negative, BBB-), and 2-Methylpentane (positive, BBB+) from Figures S4 to S6 to demonstrate chemically meaningful interpretations. Each figure includes a Saliency Overlay, a Pairwise Heatmap showing the importance score for each atom pair interaction, and a Atom Importance Bar Plot that summary of the total influence per atom.

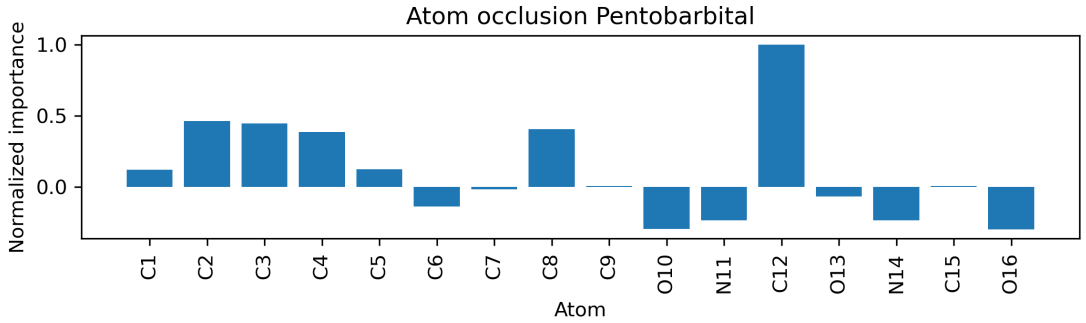
For Pentobarbital, the model identifies the central carbon of the isopropyl group and the amine nitrogen (N) and ether oxygen (O) atoms, along with adjacent carbons, as having the highest positive influence on the BBB+ prediction (Figures S4a and Figures S4c). The heatmap also highlights interactions involving these heteroatoms (Figures S4b).



(a) Saliency overlay



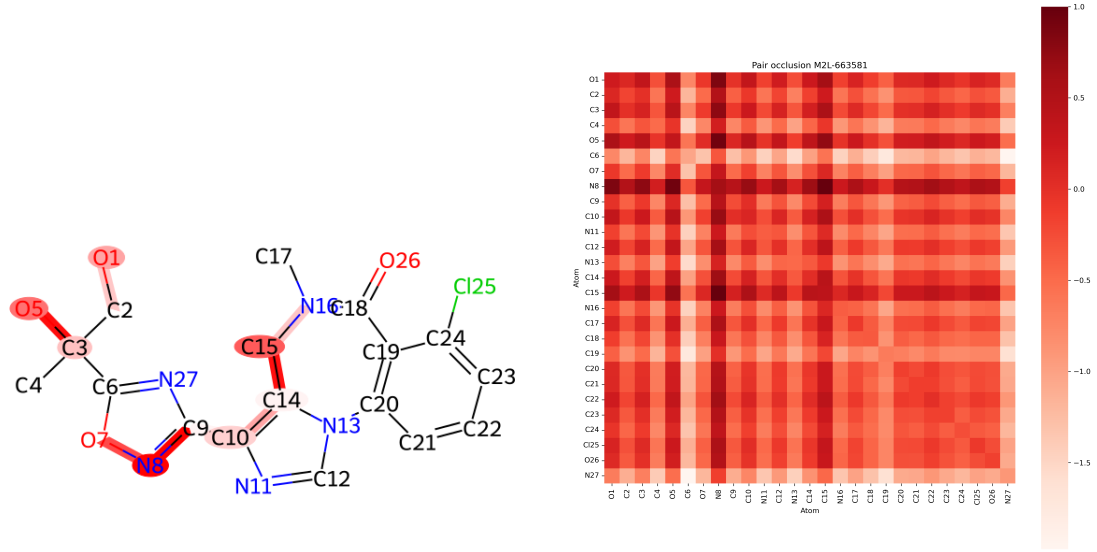
(b) Pairwise importance heatmap



(c) Atom-level importance

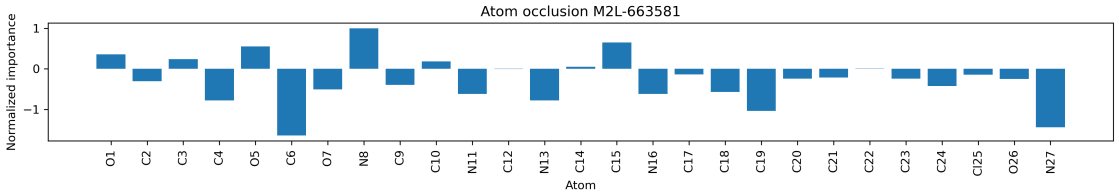
Figure S4: Model interpretability results for Pentobarbital (Predicted: BBB+).

For the BBB- compound M2L-663581, the model assigns high positive and negative importance to various oxygen and nitrogen atoms within the complex heterocyclic core and polar side chain (Figures S5a and Figures S5c), indicating these regions contribute significantly to the prediction of low permeability.



(a) Saliency overlay

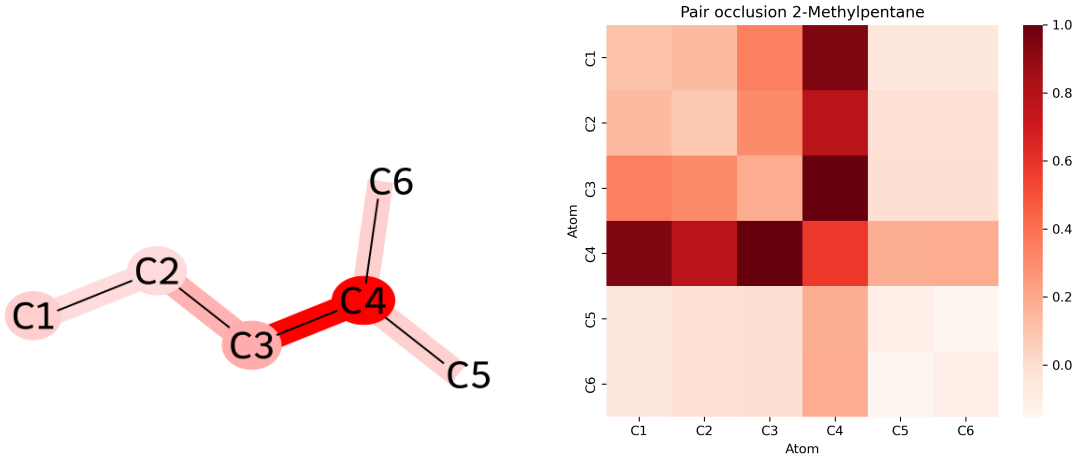
(b) Pairwise importance heatmap



(c) Atom-level importance

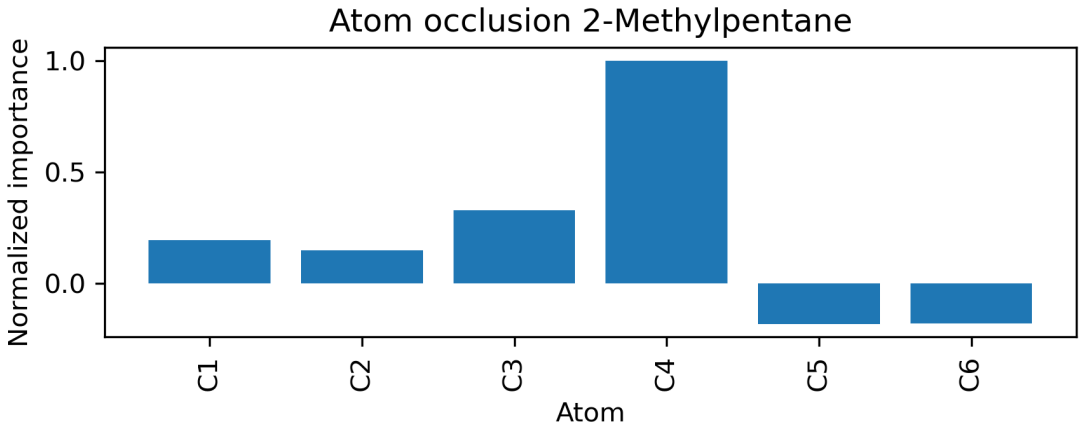
Figure S5: Model interpretability results for M2L-663581 (Predicted: BBB-).

For 2-Methylpentane, a simple lipophilic alkane, the model focuses importance on the branched carbon center (Figures S6a and Figures S6c), suggesting that the model correctly identifies the molecule as belonging to the class of small, lipophilic compounds known to permeate the BBB effectively.



(a) Saliency overlay

(b) Pairwise importance heatmap



(c) Atom-level importance

Figure S6: Model interpretability results for 2-Methylpentane (Predicted: BBB+).

The occlusion analysis confirms our model identifies chemically relevant features influencing BBB permeability. For the BBB+ compounds, Pentobarbital (Figure S4) and 2-Methylpentane (Figure S6), the model highlights features associated with successful permeation. In Pentobarbital, the geometric context of the isopropyl carbon, amine nitrogen, and ether oxygen show high positive importance. For 2-Methylpentane, the branched carbon center, linked to lipophilicity, is correctly identified as crucial. Conversely, for the BBB- compound M2L-663581 (Figure S5), significant importance is attributed to heteroatoms in the complex core and polar side chain, reflecting features known to hinder permeability.

References

[1] Duc D Nguyen, Tian Xiao, Menglun Wang, and Guo-Wei Wei. Rigidity strengthening: A mechanism for protein–ligand binding. Journal of chemical information and modeling, 57(7):1715–1721, 2017.

[2] Zhenqin Wu, Bharath Ramsundar, Evan N Feinberg, Joseph Gomes, Caleb Geniesse, Aneesh S Pappu, Karl Leswing, and Vijay Pande. Moleculenet: A benchmark for molecular machine learning. arXiv preprint arXiv:1703.00564, 2017.

[3] D. Chen, K. Gao, D. D. Nguyen, X. Chen, Y. Jiang, G. W. Wei, and F. Pan. Algebraic graph-assisted bidirectional transformers for molecular property prediction. Nature Communications, 12(1):3521, 2021.

[4] F. Meng, Y. Xi, J. Huang, et al. A curated diverse molecular database of blood-brain barrier permeability with chemical descriptors. Scientific Data, 8(1):289, 2021.

[5] Paul C. D. Hawkins, A. G. Skillman, G. L. Warren, B. A. Ellingson, and M. T. Stahl. Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and the Cambridge Structural Database. Journal of Chemical Information and Modeling, 50(4):572–584, 2010.

[6] N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison. Open babel: An open chemical toolbox. Journal of Cheminformatics, 3(1):33, 2011.

[7] Matthew D. Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In European Conference on Computer Vision (ECCV), pages 818–833. Springer, 2014.