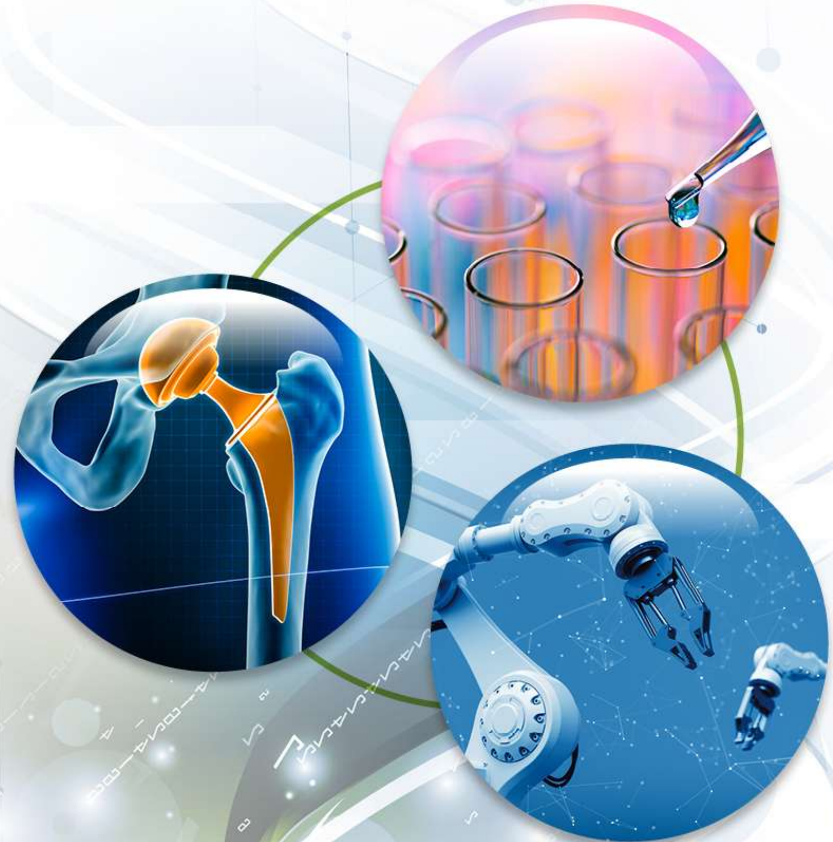
BMES

BIOMEDICAL ENGINEERING SOCIETY



De Novo CNS Drug Generation via Multimodal AI and SHAP-Guided Reinforcement Learning

Lars Arntsen • Researcher • PhD Applicant

Deep Imaging Analytics Lab

Rensselaer Polytechnic Institute

Troy, NY

Introduction & Motivation

Lab's Interests

Reinforcement Learning

Lab's Ongoing Projects

Blood-Brain Barrier Permeability (BBBP) Classifier

Reading!

Central Nervous System (CNS) Drug Development

Model Interpretability

Generative AI

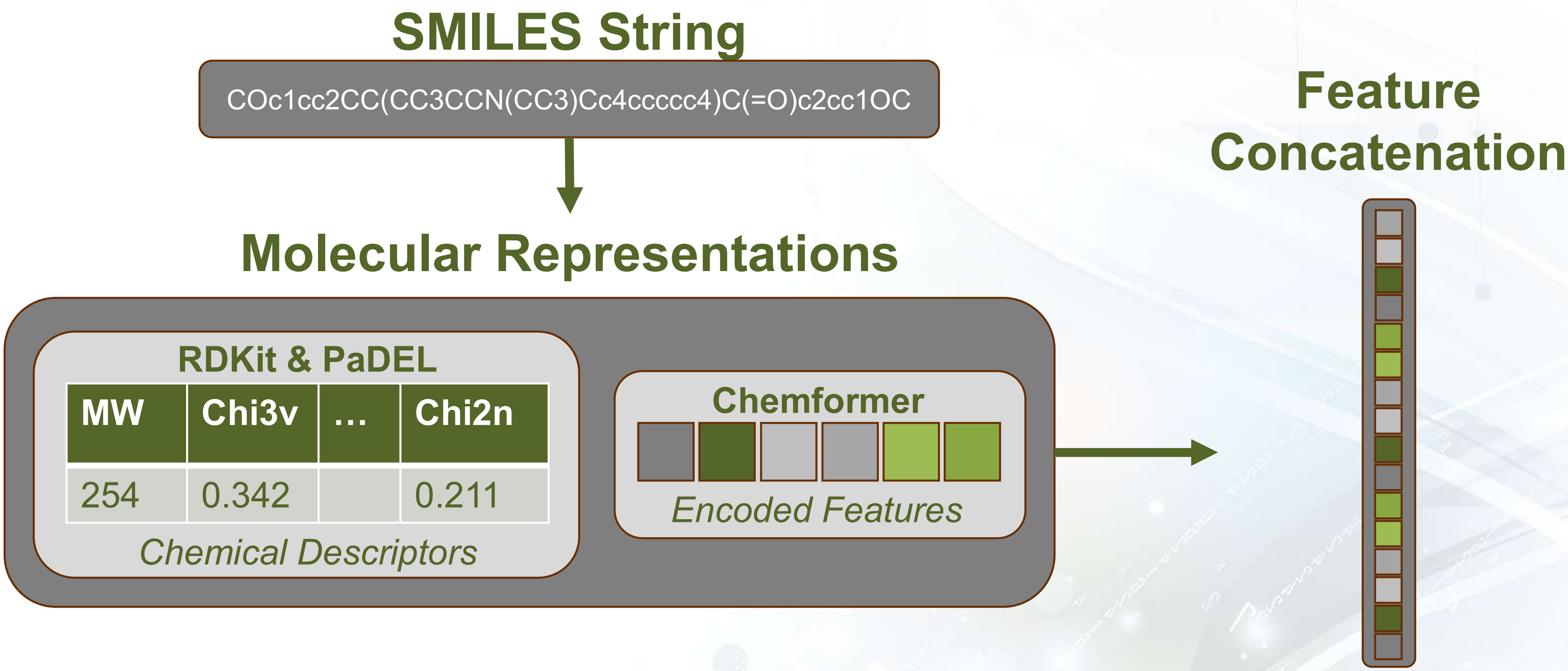
Objectives

- **Generate CNS Optimized SMILES Strings**
  1. Develop Multimodal Classifiers for Key CNS Properties
     - Conduct SHAP Analysis
  2. Design a RL Curriculum
  3. Finetune SMILES Generator

Model Setup

- **Classifiers:** LightGBM
  - **Modals:** PaDEL, RDKit, and Chemformer
- **Generator:** MolGPT (GPT-2 on ZINC-15)
- **Fine Tuning:** Policy Gradient RL
- **Goal:** Valid, Novel, and CNS-Optimized SMILES

I. Feature Extraction

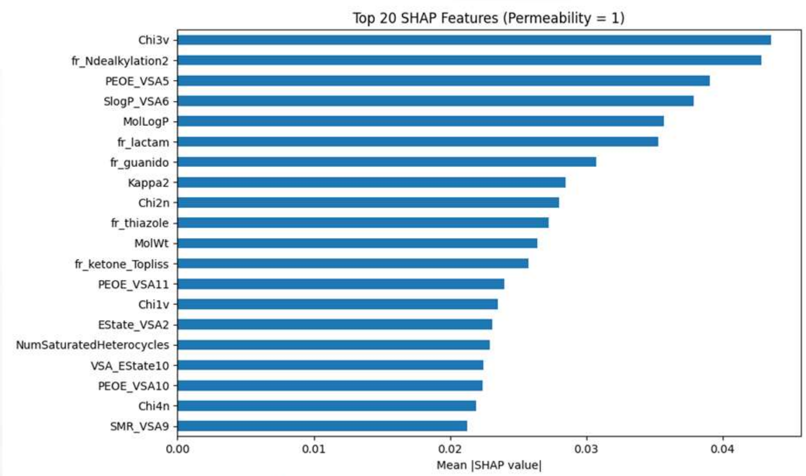


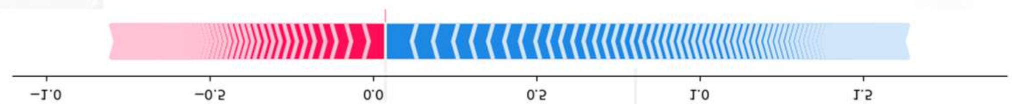
II. Feature Analysis

LightGBM Classifiers

1. Permeability
2. Neuronal cytotoxicity
3. Microelectrode Array–Based Neural Activity
4. Mammalian neurotoxicity

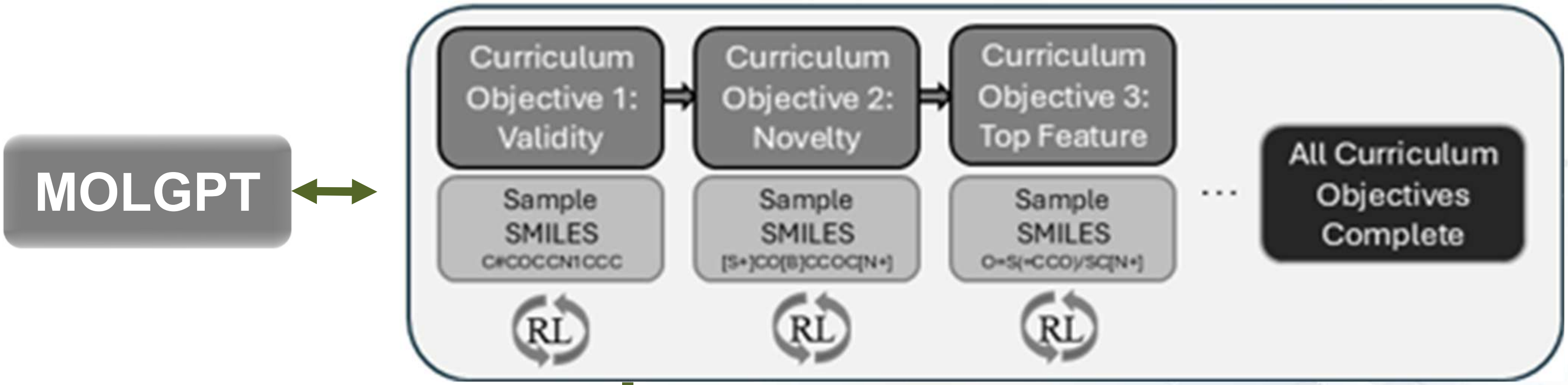
SHAP Analysis





III. Reward Design

SMILES String Generation & RL Curriculum



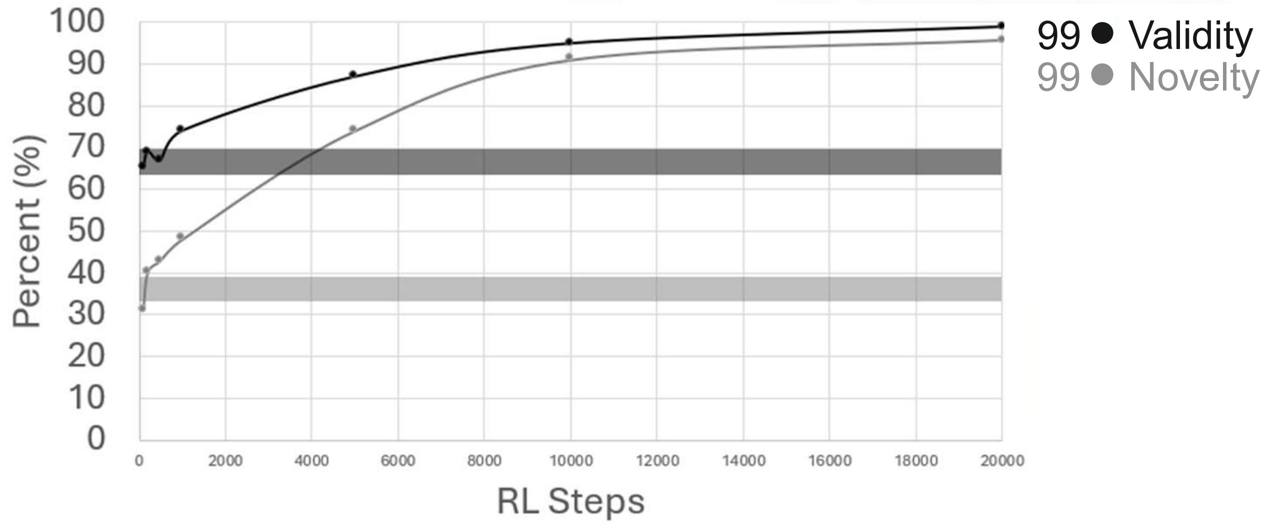
CNS Optimized SMILES Strings

C[P+](CNCNN)NC(=CCCCC)[B]CSC[N]CCOC=CCS1=CCCC

BMES
BIOMEDICAL ENGINEERING SOCIETY

Validation

Generated SMILES Strings Reward During Training



Validation RL → Classifiers → SHAP → Reward Function → Evaluation

Classifiers Performance

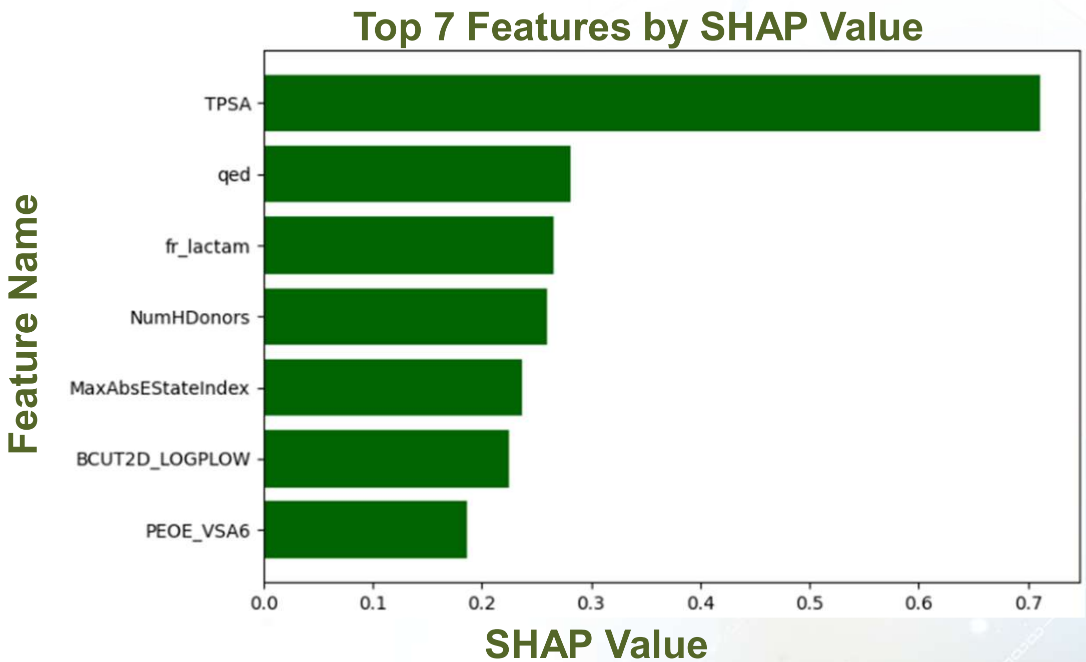
Permeability	NA	NT	NC
AUROC: 0.9771	AUROC: 0.7777	AUROC: 0.8178	AUROC: 0.9070
Comparison 0.9745	*Comparison 0.8509*	*Comparison 0.7945*	*Comparison 0.9637*

Pang et al., "NeuTox 2.0," *Environ. Int.*, 2025

Validate RL → **Classifiers** → SHAP → Reward Function → Evaluation

BMES
BIOMEDICAL ENGINEERING SOCIETY

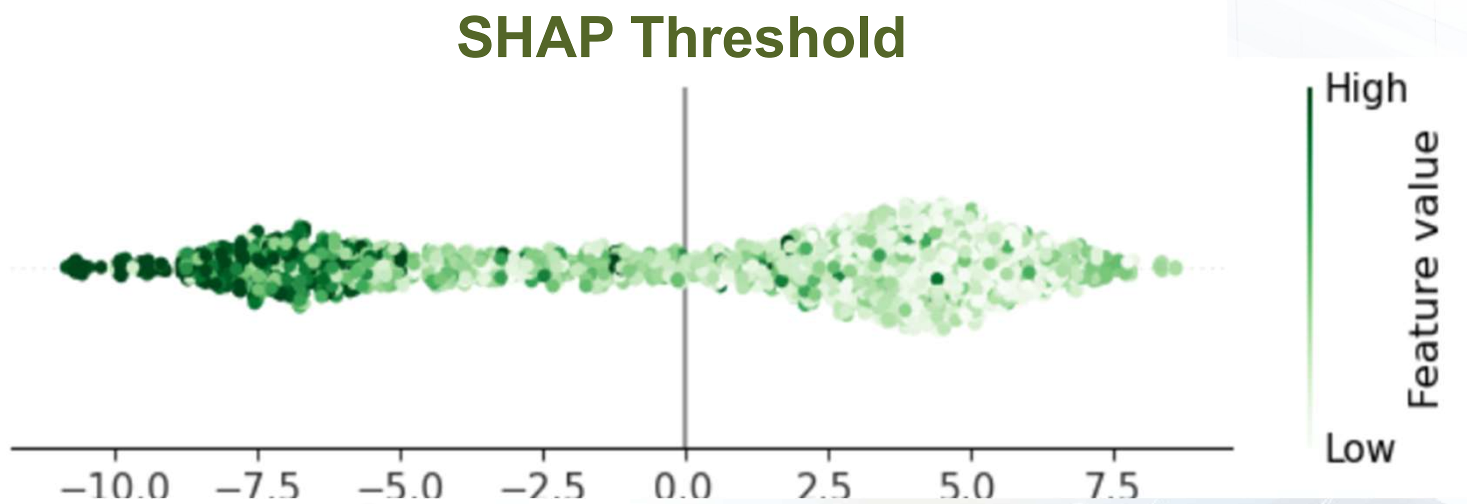
SHAP Analysis



Validate RL → Classifiers → **SHAP** → Reward Function → Evaluation

BMES
BIOMEDICAL ENGINEERING SOCIETY

SHAP Analysis



Validate RL → Classifiers → **SHAP** → Reward Function → Evaluation

BMES
BIOMEDICAL ENGINEERING SOCIETY

SHAP Analysis

Feature	Threshold	SHAP Value
fr_Ndealkylation2	0.379	0.0512
Chi3v	-0.170	0.0325
PEOE_VSA5	1.593	0.0321
MolLogP	-0.100	0.0317
SlogP_VSA6	-0.141	0.0309
…continues for all 722 features…		

Validate RL → Classifiers → **SHAP** → Reward Function → Evaluation

BMES
BIOMEDICAL ENGINEERING SOCIETY

Reward Function Design

Feature	Property	Abs. SHAP Value	Directionality	Threshold
TPSA	Permeability	0.9933	1	60.41
BertzCT	NC	0.6912	1	942.9
MolgLogP	NA	0.6723	1	3.848
…	…	…	…	…

Validate RL → Classifiers → SHAP → **Reward Function** → Evaluation

BMES
BIOMEDICAL ENGINEERING SOCIETY

Reward Function Design

- **97 Features with Abs. SHAP Value >0.10**
  - Introduce a new feature every 900 steps ~89000 steps

Validity & Novelty | TPSA | BertzCT | MolLogP

⊢ 900 Training Steps ⊣

Validate RL → Classifiers → SHAP → **Reward Function** → Evaluation

BMES
BIOMEDICAL ENGINEERING SOCIETY

Reward Function Design

Reward Function Components

- **Validity:** +0.1 if SMILES is chemically valid
- **Novelty:** Reward = 1 – max similarity
- **Summary:**

• Starts with validity and novelty

• Gradually adds property-based rewards

Validate RL → Classifiers → SHAP → **Reward Function** → Evaluation

Training Evaluation

Validate RL → Classifiers → SHAP → Reward Function → **Evaluation**

BMES
BIOMEDICAL ENGINEERING SOCIETY

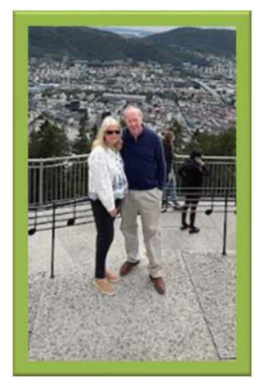
Future Work

- **Experiment with Curriculum Design**
- **Adapt to a Binding Site**
- **SMILE String Evaluation**

Acknowledgements

- Dr. Pingkun Yan
- Nadine Grant
- Dr. Diego Machado Reyes
- Zefan Yang
- Tom and Susan Arntsen
- Avery Caggiano




Personal Portfolio

SCAN ME

BMES
BIOMEDICAL ENGINEERING SOCIETY

References

[1] W. M. Pardridge, “The blood-brain barrier: Bottleneck in brain drug development,” NeuroRX, vol. 2, no. 1, pp. 3–14, Jan. 2005, doi: 10.1602/neurorx.2.1.3.

[2] K.-K. Mak and M. R. Pichika, “Artificial intelligence in drug development: present status and future prospects,” Drug Discov. Today, vol. 24, no. 3, pp. 773–780, Mar. 2019, doi: 10.1016/j.drudis.2018.11.014.

[3] B. Shaker et al., “LightBBB: computational prediction model of blood–brain-barrier penetration based on LightGBM,” Bioinformatics, vol. 37, no. 8, pp. 1135–1139, May 2021, doi: 10.1093/bioinformatics/btaa918.

[4] D. Bajusz, A. Rácz, and K. Héberger, “Chemical Data Formats, Fingerprints, and Other Molecular Descriptions for Database Analysis and Searching,” in Comprehensive Medicinal Chemistry III, Elsevier, 2017, pp. 329–378. doi: 10.1016/B978-0-12-409547-2.12345-5.

[5] M. Withnall, E. Lindelöf, O. Engkvist, and H. Chen, “Building attention and edge message passing neural networks for bioactivity and physical–chemical property prediction,” J. Cheminformatics, vol. 12, no. 1, p. 1, Jan. 2020, doi: 10.1186/s13321-019-0407-y.

[6] R. Irwin, S. Dimitriadis, J. He, and E. J. Bjerrum, “Chemformer: a pre-trained transformer for computational chemistry,” Mach. Learn. Sci. Technol., vol. 3, no. 1, p. 015022, Jan. 2022, doi: 10.1088/2632-2153/ac3ffb.

[7] Z. A. Rollins, A. C. Cheng, and E. Metwally, “MolPROP: Molecular Property prediction with multimodal language and graph fusion,” J. Cheminformatics, vol. 16, no. 1, p. 56, May 2024, doi: 10.1186/s13321-024-00846-9.

[8] D. Deng, Z. Lei, X. Hong, R. Zhang, and F. Zhou, “Describe Molecules by a Heterogeneous Graph Neural Network with Transformer-like Attention for Supervised Property Predictions,” ACS Omega, vol. 7, no. 4, pp. 3713–3721, Feb. 2022, doi: 10.1021/acsomega.1c06389.

[9] X. Pang et al., “NeuTox 2.0: A hybrid deep learning architecture for screening potential neurotoxicity of chemicals based on multimodal feature fusion,” Environ. Int., vol. 195, p. 109244, Jan. 2025, doi: 10.1016/j.envint.2024.109244.

[10] B. Yu, F. N. Baker, Z. Chen, X. Ning, and H. Sun, “LlaSMol: Advancing Large Language Models for Chemistry with a Large-Scale, Comprehensive, High-Quality Instruction Tuning Dataset,” Aug. 10, 2024, arXiv: arXiv:2402.09391. doi: 10.48550/arXiv.2402.09391.

[11] E. Wang et al., “TxGemma: Efficient and Agentic LLMs for Therapeutics”.

[12] Z. Wu et al., “MoleculeNet: a benchmark for molecular machine learning,” Chem. Sci., vol. 9, no. 2, pp. 513–530, Jan. 2018, doi: 10.1039/C7SC02664A.

BMES
BIOMEDICAL ENGINEERING SOCIETY

References

[13] I. F. Martins, A. L. Teixeira, L. Pinheiro, and A. O. Falcao, “A Bayesian Approach to in Silico Blood-Brain Barrier Penetration Modeling,” J. Chem. Inf. Model., vol. 52, no. 6, pp. 1686–1697, Jun. 2012, doi: 10.1021/ci300124c.

[14] F. Meng, Y. Xi, J. Huang, and P. W. Ayers, “A curated diverse molecular database of blood-brain barrier permeability with chemical descriptors,” Sci. Data, vol. 8, no. 1, p. 289, Oct. 2021, doi: 10.1038/s41597-021-01069-5.

[15] L. Liu et al., “Prediction of the Blood–Brain Barrier (BBB) Permeability of Chemicals Based on Machine-Learning and Ensemble Methods,” Chem. Res. Toxicol., vol. 34, no. 6, pp. 1456–1467, Jun. 2021, doi: 10.1021/acs.chemrestox.0c00343.

[16] B. Mazumdar, P. K. Deva Sarma, H. J. Mahanta, and G. N. Sastry, “Machine learning based dynamic consensus model for predicting blood-brain barrier permeability,” Comput. Biol. Med., vol. 160, p. 106984, Jun. 2023, doi: 10.1016/j.compbiomed.2023.106984.

[17] Q. Tang, F. Nie, Q. Zhao, and W. Chen, “A merged molecular representation deep learning method for blood–brain barrier permeability prediction,” Brief. Bioinform., vol. 23, no. 5, p. bbac357, Sep. 2022, doi: 10.1093/bib/bbac357.

[18] L. Storchi, G. Cruciani, and S. Cross, “DeepGRID: Deep Learning Using GRID Descriptors for BBB Prediction,” J. Chem. Inf. Model., vol. 63, no. 17, pp. 5496–5512, Sep. 2023, doi: 10.1021/acs.jcim.3c00768.

[19] D. Weininger, “SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules,” J. Chem. Inf. Comput. Sci., vol. 28, no. 1, pp. 31–36, Feb. 1988, doi: 10.1021/ci00057a005.

[20] D. Bajusz, A. Rácz, and K. Héberger, “Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations?,” J. Cheminformatics, vol. 7, no. 1, p. 20, May 2015, doi: 10.1186/s13321-015-0069-3.

[21] K.-D. Luong and A. Singh, “Application of Transformers in Cheminformatics,” J. Chem. Inf. Model., vol. 64, no. 11, pp. 4392–4409, Jun. 2024, doi: 10.1021/acs.jcim.3c02070.

[22] W. J. Geldenhuys, A. S. Mohammad, C. E. Adkins, and P. R. Lockman, “Molecular determinants of blood–brain barrier permeation,” Ther. Deliv., vol. 6, no. 7, pp. 961–971, Aug. 2015, doi: 10.4155/tde.15.32.

[23] G. Landrum et al., rdkit/rdkit: 2024_09_5 (Q3 2024) Release. (Jan. 31, 2025). Zenodo. doi: 10.5281/zenodo.14779836.

[24] “Licensing and Commercial Uses — deepchem 2.8.1.dev documentation.” Accessed: Mar. 24, 2025. [Online]. Available: https://deepchem.readthedocs.io/en/latest/development_guide/licence.html

[25] L. van der Maaten and G. Hinton, “Visualizing Data using t-SNE,” J. Mach. Learn. Res., vol. 9, no. 86, pp. 2579–2605, 2008.

BMES
BIOMEDICAL ENGINEERING SOCIETY