Few-Shot Learning for Chemical Property Prediction

UNIVERSITY OF NOTRE DAME

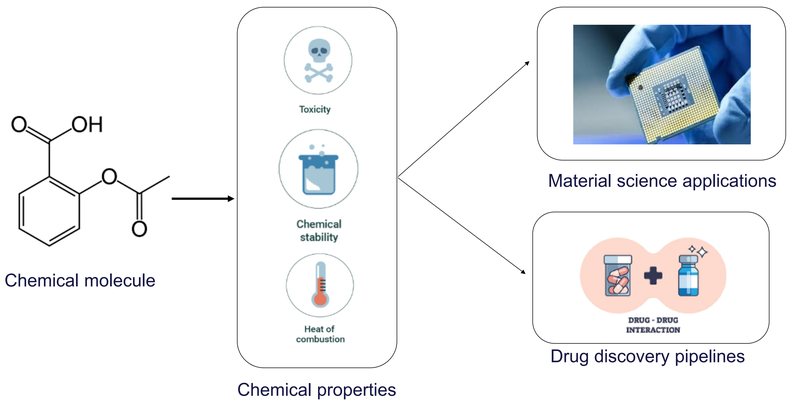
Aakarsh Misra ‡, Bryan Piguave † and Brett M Savoie † (bsavoie2@nd.edu)

†University of Notre Dame, Department of Chemical and Biomolecular Engineering

‡Texas Tech University , Department of Electrical and Computer Engineering

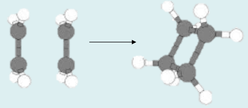
Introduction

Accurate prediction of chemical properties enables breakthroughs across various scientific domains, ranging from drug discovery to materials design



The evolving landscape of chemical property prediction

Conducting Chemical Experiments.

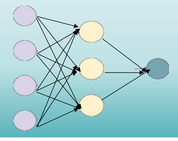


Time and cost intensive, ethically sensitive.

Statistical Methods and quantum Simulations

Struggle to scale with larger datasets and growing molecular complexity.

Current AI/ML based approaches



Fast and flexible, BUT need a lot of data for accurate predictions

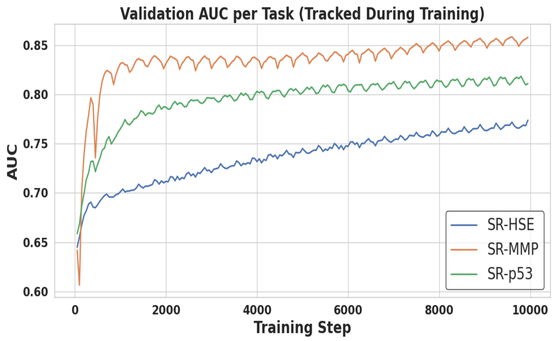
However, in chemical domains such as toxicology, labeled data is often scarce. In these low-data regimes, conventional AI and ML models tend to underperform and struggle to generalize

Funding Acknowledgement

This work was made possible by the Office of Naval Research (ONR) through support provided by the Energetic Materials Program (MURI grant number: N00014-21-1-2476, Program Manager: Dr. Chad Stoltz)

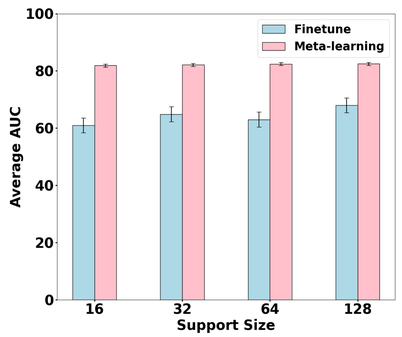
Results

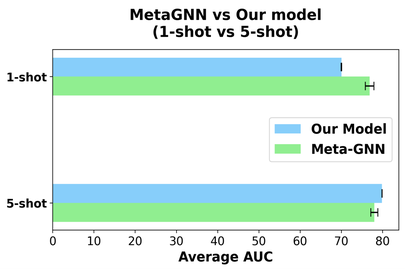
AUC progression and final accuracy metrics



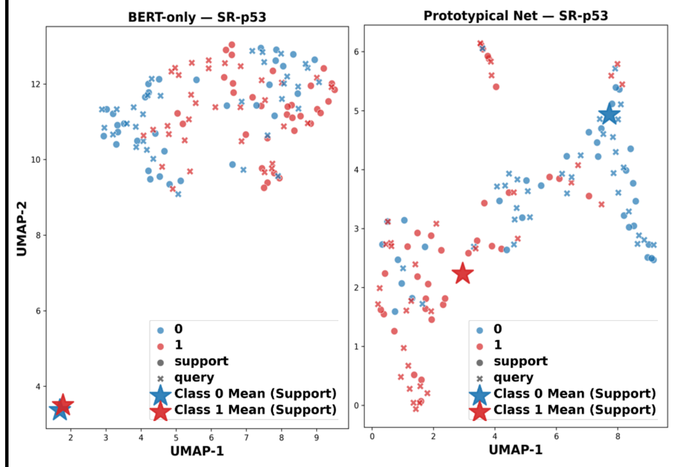
Support (2-way)	AVG AUC ± std dev	Accuracy ± std dev
16 shot	81.72 ± 0.04	78.87 ± 0.08
5 shot	79.88± 0.035	72.3 ± 0.035
1 shot	70.04 ± 0.06	61.76 ± 0.071

Positioning our model against Baselines





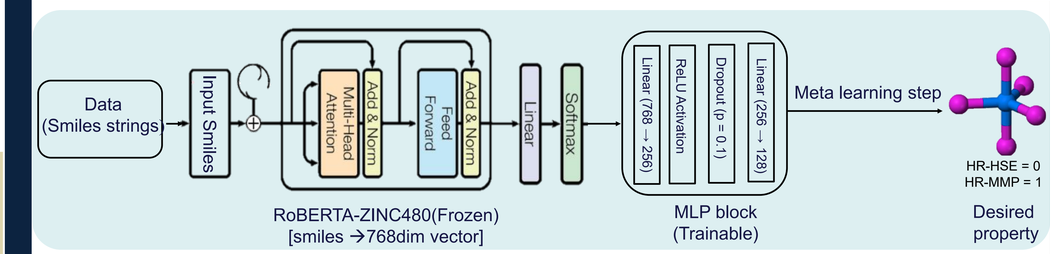
Pre and post Meta-Learning: UMAP latent space visualization



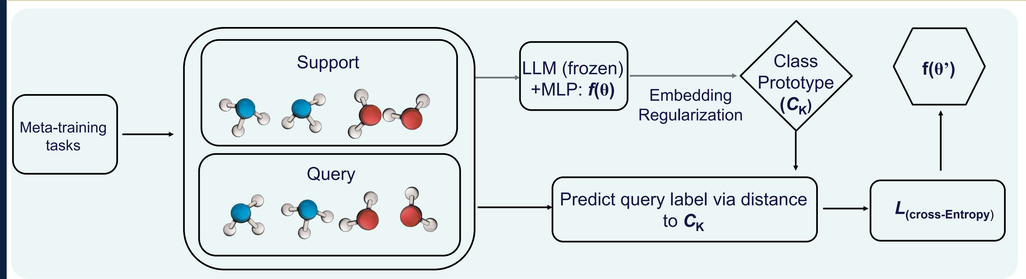
In conclusion, our Few-shot learning framework for class-based property prediction demonstrates competitive performance, often exceeding that of traditional baselines.

Methods

Our Few-shot Learning approach uses a Meta-learning framework based on prototypical networks, trained episodically to enable few-shot generalization. Our model combines a frozen Roberta-Zinc480M backbone for molecular embeddings with a trainable MLP head that adapts across property prediction tasks.



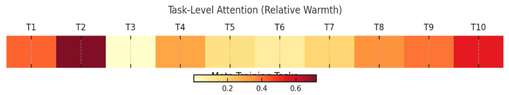
Below is a visual overview of the episodic Meta-learning process.



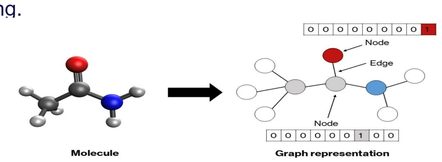
Future Work

Our meta-learning approach shows strong promise in the Few-shot class-based molecular property prediction domain. Future work includes:

1. Implement task-based attention mechanisms for improved cross-task generalization and explore deeper MLP's



2. Adapt our episodic training pipeline to Graph-based molecular encoders like EGAT, enabling enhanced structure-aware Few-shot learning.



Other Property prediction projects ongoing in the Savoie Group:

- Generalized property imputation
- Property prediction through GNN.