Integrating Bilinear Transduction with Message Passing Neural Networks for Improved ADMET Property Prediction

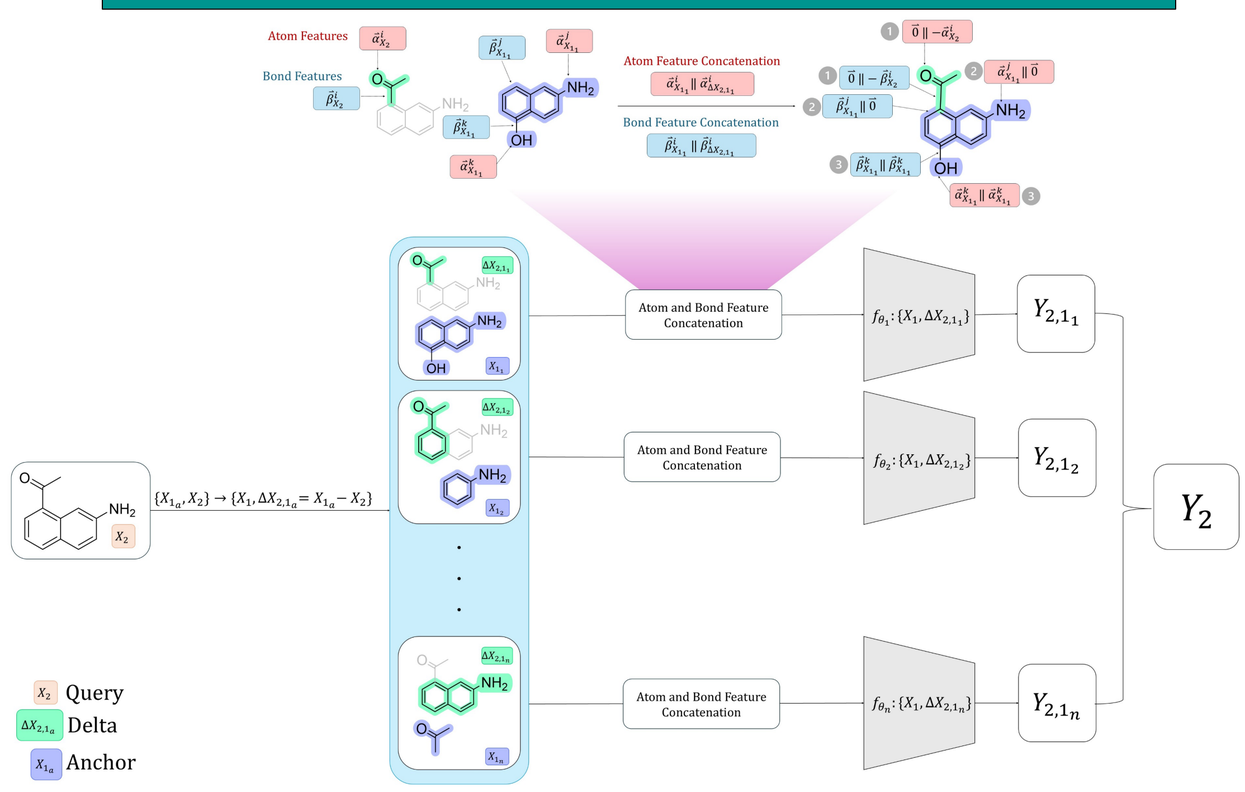
Claire Suen[1], Alan C Cheng[1]

[1]Modeling and Informatics, Discovery Chemistry, Merck & Co., Inc., South San Francisco, Ca, USA.

Abstract

- **The challenge:** Deep learning in drug discovery often face censored molecular property datasets[4]. This occurs because measurement limitations in pharmaceutical assays mean exact values beyond certain thresholds aren't recorded.
- **Current limitations:** Standard deep learning methods struggle with this censoring, leading to systematic prediction errors, even for in-distribution molecules.
- **Our solution:** We propose to integrate bilinear transduction[2,3] into Chemprop's[1] message-passing neural network. This builds on Chemprop's strengths and allows us to leverage domain-specific structural relationships between molecules.

Model Architecture



Model Performance with Baselines

Naturally Censored Internal Datasets

Assay	Training Label	R^2 ↑		RMSE ↓	
		BT	D-MPNN	BT	D-MPNN
CYP 3A4	50K	0.30±0.03	0.26±0.06	0.47±0.01	0.48±0.02
	100K	**0.32±0.04**	0.22±0.04	**0.46±0.01**	0.49±0.01
	224,593	0.40±0.04	0.29±0.07	**0.43±0.01**	0.47±0.02
CYP 2D6	50K	**0.19±0.05**	-0.02±0.06	**0.45±0.01**	0.51±0.02
	100K	**0.30±0.04**	0.12±0.04	**0.42±0.01**	0.47±0.01
	221,745	**0.32±0.04**	0.11±0.03	**0.41±0.01**	0.47±0.01
CYP 2C9	50K	**0.08±0.03**	-0.07±0.10	0.50±0.01	0.54±0.03
	100K	**0.19±0.04**	0.01±0.04	**0.47±0.01**	0.52±0.01
	225,026	**0.24±0.06**	0.05±0.06	**0.46±0.02**	0.51±0.02
CAV 1.2	50K	0.10±0.06	0.05±0.06	0.27±0.01	0.28±0.01
	100K	**0.11±0.05**	-0.06±0.05	0.27±0.01	0.29±0.01
HERG MK499	50K	0.18±0.05	0.16±0.03	0.48±0.01	0.48±0.01
	100K	0.23±0.04	0.25±0.04	0.46±0.01	0.46±0.01

Synthetically Censored Internal Datasets

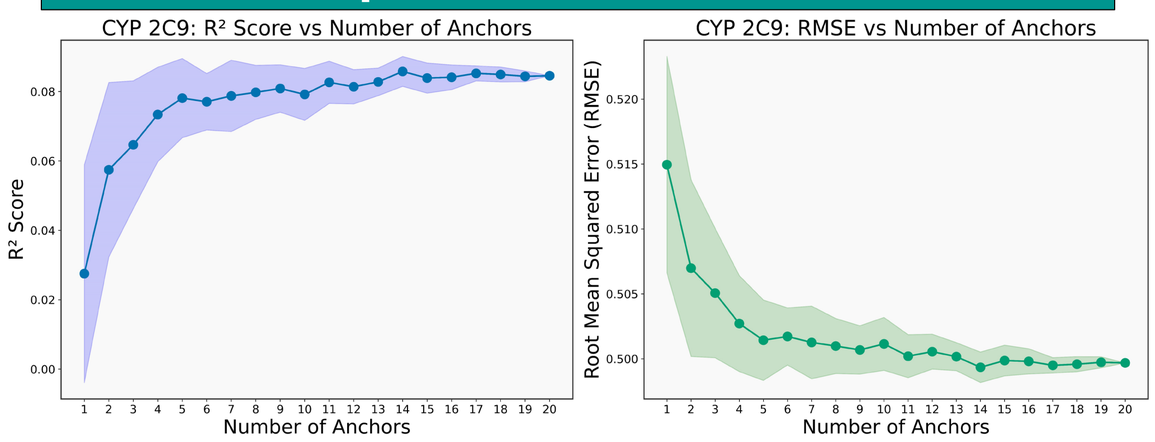
Assay	Training Label	R^2 ↑		RMSE ↓	
		BT	D-MPNN	BT	D-MPNN
P-GP, rat	25TH	**0.57±0.02**	0.50±0.01	**0.35±0.01**	0.37±0.00
	50TH	**0.40±0.05**	0.18±0.12	**0.29±0.01**	0.34±0.02
	75TH	**-0.03±0.06**	-0.50±0.26	**0.22±0.01**	0.26±0.02
rat $F_{u,p}$	25TH	0.57±0.02	0.56±0.01	0.30±0.01	0.31±0.00
	50TH	0.51±0.02	0.47±0.02	0.23±0.01	0.24±0.00
	75TH	**0.43±0.02**	0.27±0.04	**0.16±0.00**	0.18±0.00

Synthetically & Naturally Censored Public Datasets

Assay	Training Label	R^2 ↑		RMSE ↓	
		BT	D-MPNN	BT	D-MPNN
MS (Human)	Base	0.24±0.04	0.33±0.06	0.46±0.01	0.43±0.02
	50TH	**0.21±0.05**	0.08±0.07	0.38±0.01	0.41±0.02
	75TH	**0.03±0.04**	-0.06±0.02	0.29±0.01	0.30±0.00
MS (Rat)	Base	0.48±0.02	0.48±0.02	0.46±0.01	0.46±0.01
	50TH	0.14±0.04	0.09±0.02	0.35±0.01	0.36±0.00
	75TH	**0.06±0.02**	-0.56±0.12	**0.23±0.00**	0.30±0.01
CYP 3A4	4,403	0.52±0.02	0.53±0.02	**0.58±0.01**	0.61±0.01

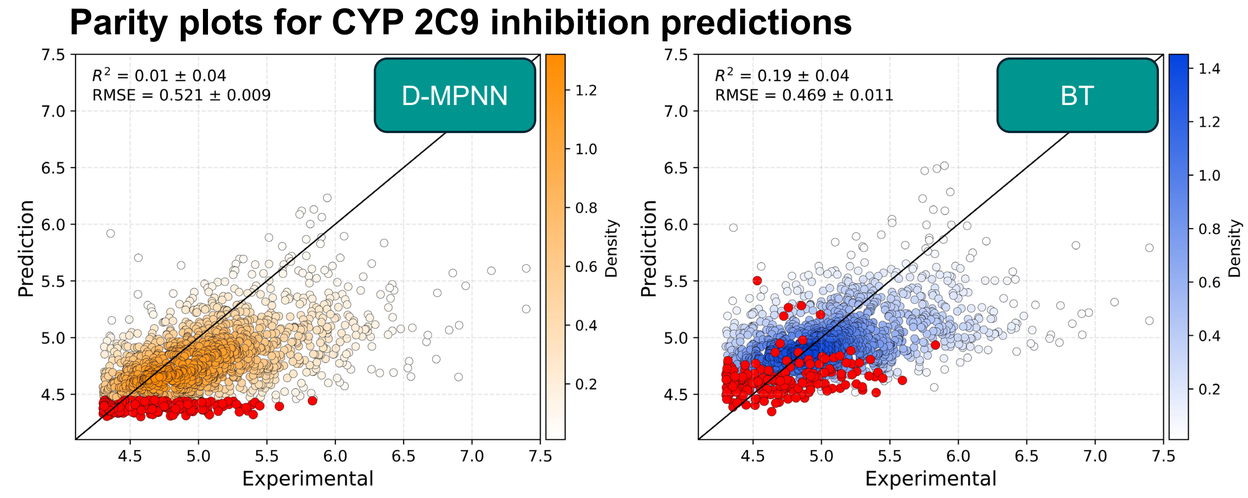
- All performance metrics for BT were found to be equal to or better than D-MPNN
- Most significant performance gains were observed for CYP enzyme prediction tasks

Optimal Anchor Size



Predictive Performance of BT

Parity plots for CYP 2C9 inhibition predictions



Anchor Error Distribution Comparison

Top 5 Similar Anchor Error Distributions

Anchor X	Anchor Y
Anchor 2677 — CYP 3A4 = 4.591	TS = 0.13, R^2 = 0.93 — Anchor 299 — CYP 3A4 = 1.699
Anchor 2677 — CYP 3A4 = 4.591	TS = 0.11, R^2 = 0.93 — Anchor 3135 — CYP 3A4 = 4.602
Anchor 2240 — CYP 3A4 = 4.055	TS = 0.10, R^2 = 0.92 — Anchor 3330 — CYP 3A4 = 4.23
Anchor 299 — CYP 3A4 = 1.699	TS = 0.13, R^2 = 0.92 — Anchor 3135 — CYP 3A4 = 4.602
Anchor 2240 — CYP 3A4 = 4.055	TS = 0.11, R^2 = 0.92 — Anchor 2551 — CYP 3A4 = 3.826

Top 5 Dissimilar Anchor Error Distributions

Anchor X	Anchor Y
Anchor 1053 — CYP 3A4 = 4.064	TS = 0.15, R^2 = 0.79 — Anchor 1192 — CYP 3A4 = 3.286
Anchor 1053 — CYP 3A4 = 4.064	TS = 0.11, R^2 = 0.81 — Anchor 1225 — CYP 3A4 = 3.626
Anchor 1192 — CYP 3A4 = 3.286	TS = 0.15, R^2 = 0.81 — Anchor 2677 — CYP 3A4 = 4.591
Anchor 1192 — CYP 3A4 = 3.286	TS = 0.11, R^2 = 0.81 — Anchor 2895 — CYP 3A4 = 2.672
Anchor 1192 — CYP 3A4 = 3.286	TS = 0.11, R^2 = 0.82 — Anchor 3135 — CYP 3A4 = 4.602

Conclusion

- Our approach significantly improves in-distribution prediction accuracy, particularly for datasets with censored labels where measurement limitations prevent recording exact values.
- We show that while standard Chemprop models can handle small percentages of censored data, they degrade significantly at ~50% censoring, predicting values clustered around the censoring threshold even when true values lie well above.
- We analyze the impact of different anchor selection strategies, showing that performance improvement plateaus around 8-10 anchor molecules while variance continues to decrease with larger anchor sets.

[1]Yang, K., Swanson, K., Jin, W., Coley, C., Eiden, P., Gao, H., Guzman-Perez, A., Hopper, T., Kelley, B., Mathea, M., Palmer, A., Settels, V., Jaakkola, T., Jensen, K., & Barzilay, R. (2019). Analyzing learned molecular representations for property prediction. *Journal of Chemical Information and Modeling*, *59*(8), 3370–3388. https://doi.org/10.1021/acs.jcim.9b00237

[2]Segal, N., Netanyahu, A., Greenman, K. P., Agrawal, P., & Gomez-Bombarelli, R. (2025). Known unknowns: Out-of-distribution property prediction in materials and molecules. *arXiv preprint* arXiv:2502.05970. https://arxiv.org/abs/2502.05970

[3]Netanyahu, A., Gupta, A., Simchowitz, M., Zhang, K., & Agrawal, P. (2023). Learning to extrapolate: A transductive approach. *arXiv preprint* arXiv:2304.14329. https://arxiv.org/abs/2304.14329

[4]Svensson, E., Friesacher, H. R., Winiwarter, S., Mervin, L., Arany, A., & Engkvist, O. (2025). Enhancing uncertainty quantification in drug discovery with censored regression labels. *Artificial Intelligence in the Life Sciences*, *7*, 100128. https://doi.org/10.1016/j.ailsci.2025.100128