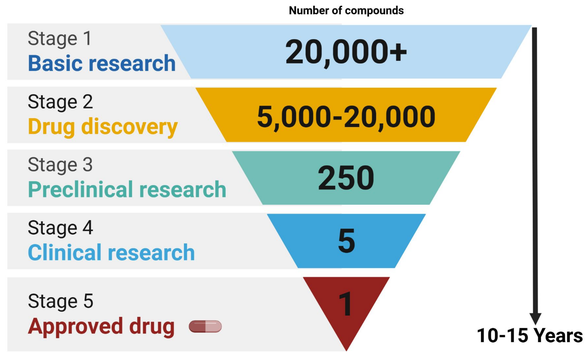
Machine Learning for SARS-CoV-2 Antiviral Drug Discovery

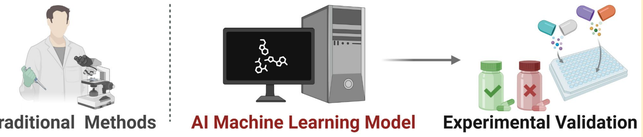
Zhexiang(Bruce) Li, Emily Nguyen and Yan Liu
Department of Computer Science, USC, Los Angeles, CA

Motivation & Background

- Drug discovery is costly and time-intensive process.



- AI can accelerate development before experimental testing



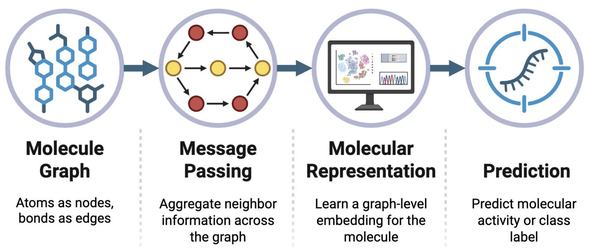
Methods

- Benchmarked 7 binary classification datasets from TDC across HTS, toxicity, and ADMET.
- Used 3 scaffold-based splits and reported AUROC mean +/- std.

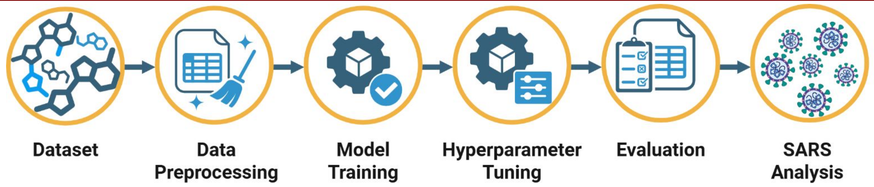
Dataset	Domain	Molecules
SARSCoV2_3CLPro_Diamond	HTS	879
ogbg-molhiv	HTS	40983
hERG	Toxicity	622
DILI	Toxicity	475
AMES	Toxicity	7255
HIA_Hou	ADMET	578
Bioavailability_Ma	ADMET	640

- Compared RDKit-descriptor baselines with Chemprop, a message passing neural network (MPNN).
- For SARS-CoV-2 3CLPro, we also analyzed descriptor importance and property distributions.

Chemprop / MPNN Overview



Main Results

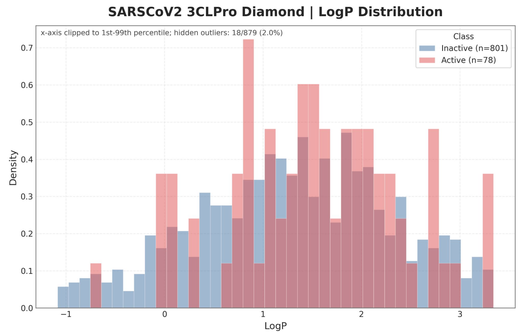


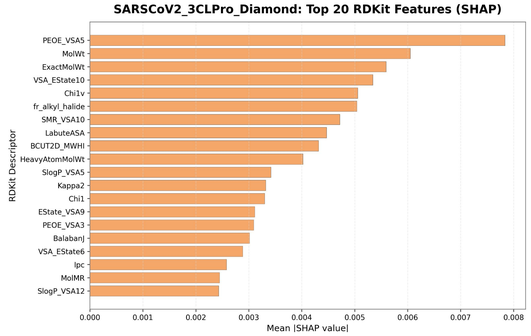
- Tree-based descriptor models consistently outperformed the Chemprop runs across benchmarks datasets.
- Extra Trees achieved the best average AUROC (bold). Random Forest was a close second (underlined).

Model	SARS	AMES	Bio_Ma	DILI	HIA_Hou	hERG	ogbg-molhiv
Extra Trees	**90.0 +/- 6.4**	84.6 +/- 2.9	**74.2 +/- 2.8**	**86.1 +/- 3.7**	**98.5 +/- 1.4**	**85.0 +/- 3.1**	82.9 +/- 0.9
Random Forest	87.9 +/- 7.4	**84.8 +/- 3.2**	74.1 +/- 0.5	84.6 +/- 4.6	98.3 +/- 1.6	84.7 +/- 2.9	**84.3 +/- 1.9**
Grad. Boosting	80.1 +/- 11.2	84.0 +/- 2.8	72.9 +/- 6.9	81.3 +/- 5.7	95.7 +/- 4.8	79.6 +/- 2.4	71.0 +/- 13.3
Log. Reg.	74.8 +/- 0.9	82.8 +/- 1.1	74.2 +/- 2.4	83.1 +/- 3.3	98.8 +/- 1.4	76.6 +/- 2.9	65.3 +/- 18.9
AdaBoost	70.4 +/- 8.8	80.1 +/- 4.1	68.1 +/- 6.6	82.9 +/- 3.9	95.2 +/- 5.3	80.1 +/- 1.3	67.0 +/- 12.6
Chemprop	44.9 +/- 7.9	76.8 +/- 3.0	69.6 +/- 8.6	80.2 +/- 11.4	39.5 +/- 8.5	58.8 +/- 10.9	54.8 +/- 15.9
Chemprop+RDKit	36.9 +/- 11.7	71.8 +/- 2.0	38.4 +/- 9.4	63.9 +/- 2.3	53.4 +/- 5.7	62.6 +/- 4.9	56.5 +/- 9.1

SARS Case Study

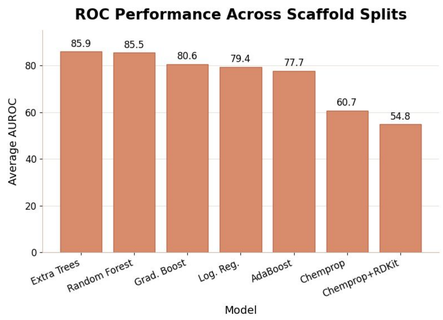
- Descriptor-level interpretation and property-shift analysis support chemically meaningful screening behavior.
- These results support RDKit-descriptor baselines for practical SARS screening.
  * SARS refers to the TDC dataset SARSCoV2_3CLPro_Diamond.





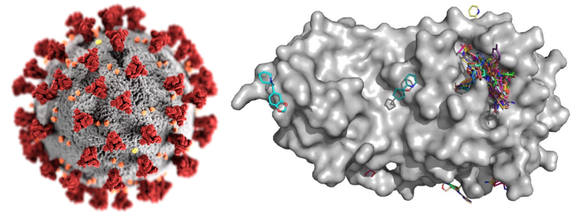
- Extra Trees and Random Forest delivered consistent high AUROC performance across scaffold splits.

Model	Split 1	Split 2	Split 3	Mean ± Std
Extra Trees	**99.0**	**85.5**	**85.5**	**90.0 ± 6.4**
Random Forest	98.5	82.8	82.6	87.9 ± 7.4
Grad. Boost	95.9	72.9	71.4	80.1 ± 11.2
Log. Reg.	75.5	73.4	75.3	74.8 ± 0.9
AdaBoost	82.7	62.4	66.0	70.4 ± 8.8
Chemprop	33.7	50.5	50.5	44.9 ± 7.9
Chemprop+RDKit	20.4	45.1	45.2	36.9 ± 11.7



Overall Model Performance:

- Best overall model: Extra Trees, average AUROC 85.9
- Second-best overall model: Random Forest, average AUROC 85.5
- Best SARS result: Extra Trees, 90.0 +/- 6.4



Left: SARS-CoV-2 virus. Right: SARS-CoV-2 3CL protease, the target used in our SARS case study. Dataset from a high-resolution XChem fragment screen (MIT AiCures).

Limitations & Future Directions:

- Chemprop was trained under a modest setup with short training and limited tuning.
- Current deep-learning results should be interpreted as baseline Chemprop results, not fully optimized GNN performance.
- Feature interpretation should be rerun directly on the best-performing SARS model.
- Future work should improve SARS-focused Chemprop training and revisit model-aligned interpretation.

Conclusions

- Classical RDKit-descriptor baselines were strong across the benchmark.
- Extra Trees and Random Forest were the top-performing models overall.
- For SARS-CoV-2 3CLPro, classical ML clearly outperformed the current Chemprop variants.
- Strong descriptor-based baselines remain essential in molecular screening, and SARS-CoV-2 3CLPro provides a clear HTS test case.

Acknowledgement

- This work was conducted at the University of Southern California. Figures were created using BioRender.
- Thanks to Emily Nguyen and Yan Liu for project guidance and feedback.