Supporting information (Figures)

Large-scale Modeling of Multi-Species Acute Toxicity Endpoints using Consensus of Multi-Task Deep Learning Methods

Sankalp Jain[a], Vishal B. Siramshetty[a], Vinicius M. Alves[b], Eugene N. Muratov[b], Nicole Kleinstreuer[c,d], Alexander Tropsha[b], Marc C. Nicklaus[e], Anton Simeonov[a], Alexey V. Zakharov[a,*]

[a] National Center for Advancing Translational Sciences (NCATS), National Institutes of Health, 9800 Medical Center Drive, Rockville, MD, 20850, United States

[b] UNC Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, United States

[c] Division of Intramural Research, Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences, 111 T.W. Alexander Drive, Durham, North Carolina 27709, United States

[d] National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods, National Institute of Environmental Health Sciences, 111 T.W. Alexander Drive, Durham, North Carolina 27709, United States

[e] Computer-Aided Drug Design (CADD) Group, Chemical Biology Laboratory, Center for Cancer Research, National Cancer Institute, National Institutes of Health, DHHS, NCI-Frederick, 376 Boyles St., Frederick, MD 21702, United States

*Corresponding author

Alexey V. Zakharov

Email: alexey.zakharov@nih.gov

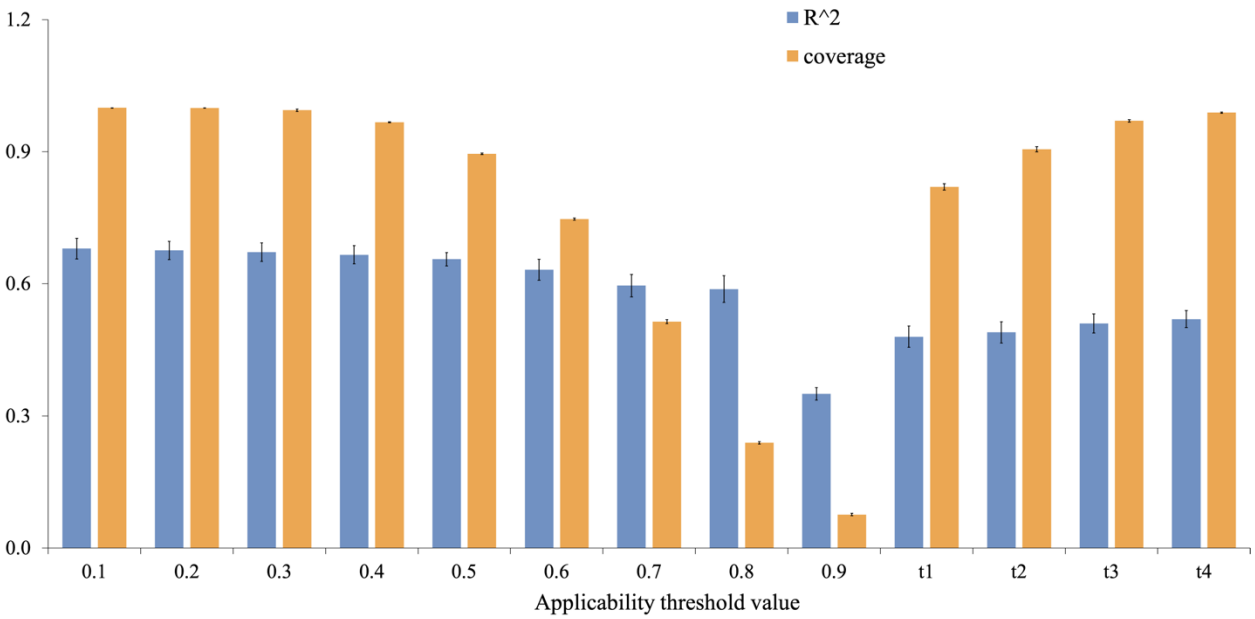
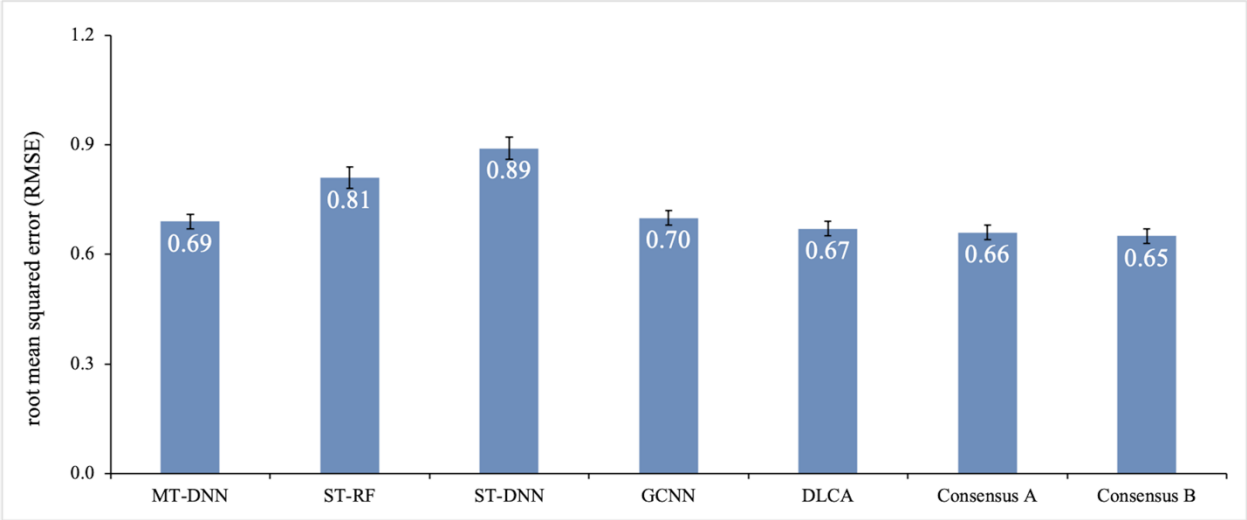


Figure S1: Distribution of the DLCA prediction results (R^2) over AD cut-offs and coverage values. The error bar represents the standard deviation of the average performance over five-folds.

(a)



(b)

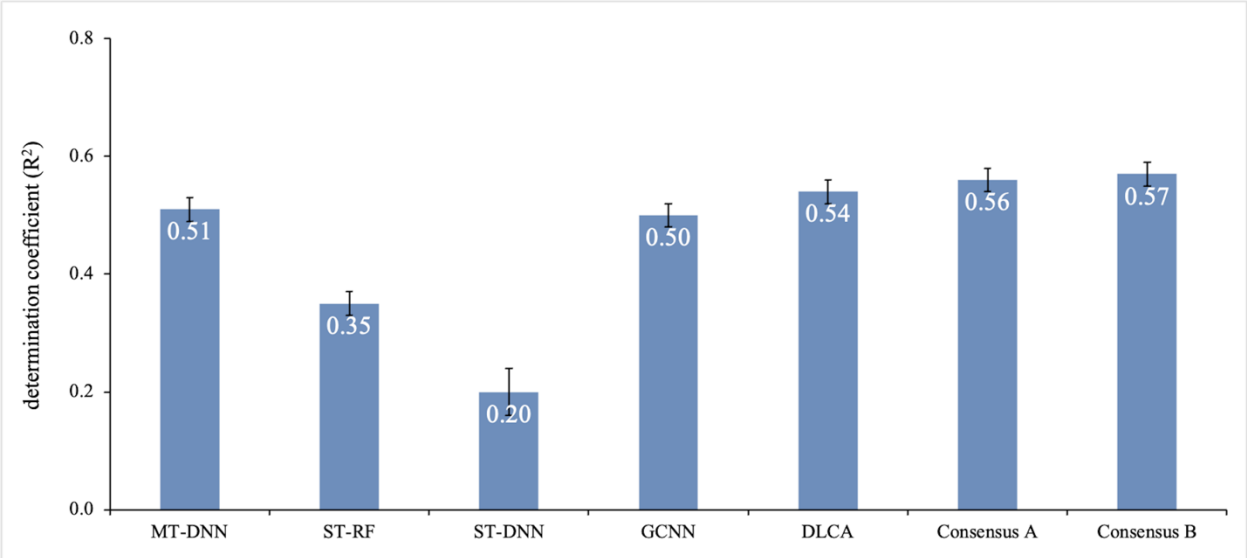


Figure S2. Average performance (a) RMSE, (b) R^2 of all 59 endpoints for each approach over five-fold cross-validation based on training and test data generated using random splitting. The error bar represents the standard error mean of the average performance over five-folds.