

Table of Contents

1. **Supplementary Table 1.** Detailed statistics for all acute toxicity endpoints in the dataset.
2. **Supplementary Table 2.** Average R^2 of 5-fold cross-validation over 59 acute toxicity endpoints.
3. **Supplementary Table 3.** Average RMSE of 5-fold cross-validation over 59 acute toxicity endpoints.
4. **Supplementary Table 4.** Average R^2 over small/large-sized endpoints (avg. +/- std.).
5. **Supplementary Table 5.** The acute toxicity evaluation performance (R^2) of ToxACoL with reducing training measurements of small-sized endpoints.

1. **Supplementary Figure 1.** Endpoint-wise RMSE achieved by different models.
2. **Supplementary Figure 2.** Performance of ToxACoL and its variant (without adjoint correlation) varying with the network depth.
3. **Supplementary Figure 3.** Comparison of average RMSE on small/large-sized endpoints.
4. **Supplementary Figure 4.** Acute toxicity evaluation performance of ToxACoL on small-sized endpoints with reducing training measurements.
5. **Supplementary Figure 5.** Acute toxicity evaluation performance of ToxACoL on large-sized endpoints and all endpoints with reducing training measurements.
6. **Supplementary Figure 6.** PCC values between human-oral-TDLo and other 56 animal endpoints.
7. **Supplementary Figure 7.** PCC values between women-oral-TDLo and other 56 animal endpoints.
8. **Supplementary Figure 8.** PCC values between man-oral-TDLo and other 56 animal endpoints.
9. **Supplementary Figure 9.** Pipeline of previous vanilla multi-task learning models for multi-condition acute toxicity evaluation.

Supplementary Table 1. Detailed statistics for all acute toxicity endpoints in the dataset.

Id	Experimental Species	Administration Route	Record Metric	Number of Measurements	Missing Rate of Measurements (%)	
1	mouse	intraperitoneal	lethal dose fifty	36295	54.68	Large-sized endpoint
2	mouse	oral	lethal dose fifty	23373	70.81	
3	mouse	intravenous	lethal dose fifty	16978	78.80	
4	rat	oral	lethal dose fifty	10190	87.28	
5	mouse	subcutaneous	lethal dose fifty	6769	91.55	
6	rat	intraperitoneal	lethal dose fifty	5021	93.73	
7	rat	intravenous	lethal dose fifty	2472	96.91	
8	rat	subcutaneous	lethal dose fifty	1896	97.63	
9	mouse	unreported	lethal dose fifty	1739	97.83	
10	rabbit	skin	lethal dose fifty	1495	98.13	
11	mammal (species unspecified)	unreported	lethal dose fifty	1129	98.59	
12	rabbit	oral	lethal dose fifty	894	98.88	
13	rat	skin	lethal dose fifty	835	98.96	
14	rat	unreported	lethal dose fifty	806	98.99	
15	guinea pig	oral	lethal dose fifty	793	99.01	
16	rabbit	intravenous	lethal dose fifty	792	99.01	
17	mammal (species unspecified)	oral	lethal dose fifty	674	99.16	
18	dog	oral	lethal dose fifty	649	99.19	
19	mouse	intramuscular	lethal dose fifty	571	99.29	
20	mammal (species unspecified)	intraperitoneal	lethal dose fifty	545	99.32	
21	dog	intravenous	lethal dose fifty	468	99.42	
22	dog	intravenous	lethal dose low	360	99.55	
23	chicken	oral	lethal dose fifty	353	99.56	
24	quail	oral	lethal dose fifty	352	99.56	
25	rabbit	intravenous	lethal dose low	346	99.57	
26	bird - wild	oral	lethal dose fifty	338	99.58	
27	rat	oral	lethal dose low	322	99.60	
28	rat	intraperitoneal	lethal dose low	318	99.60	
29	mouse	parenteral	lethal dose fifty	302	99.62	
30	rat	intramuscular	lethal dose fifty	300	99.63	
31	mouse	intraperitoneal	lethal dose low	266	99.67	
32	mouse	oral	lethal dose low	264	99.67	
33	cat	intravenous	lethal dose fifty	261	99.67	
34	mouse	subcutaneous	lethal dose low	252	99.69	
35	rabbit	oral	lethal dose low	249	99.69	
36	guinea pig	intraperitoneal	lethal dose fifty	248	99.69	
37	rabbit	subcutaneous	lethal dose low	241	99.70	
38	mouse	skin	lethal dose fifty	214	99.73	
39	duck	oral	lethal dose fifty	192	99.76	Small-sized endpoint
40	dog	oral	lethal dose low	187	99.77	
41	rabbit	skin	lethal dose low	181	99.77	
42	guinea pig	subcutaneous	lethal dose low	179	99.78	
43	guinea pig	skin	lethal dose fifty	176	99.78	
44	rat	subcutaneous	lethal dose low	174	99.78	
45	cat	oral	lethal dose fifty	171	99.79	
46	guinea pig	subcutaneous	lethal dose fifty	169	99.79	
47	man	oral	toxic dose low	163	99.80	
48	cat	intravenous	lethal dose low	159	99.80	
49	women	oral	toxic dose low	156	99.81	
50	rabbit	subcutaneous	lethal dose fifty	156	99.81	
51	guinea pig	intravenous	lethal dose fifty	153	99.81	
52	cat	oral	lethal dose low	142	99.82	
53	human	oral	toxic dose low	140	99.83	
54	rat	intravenous	lethal dose low	135	99.83	
55	rabbit	intraperitoneal	lethal dose fifty	131	99.84	
56	mammal (species unspecified)	subcutaneous	lethal dose fifty	125	99.84	
57	guinea pig	intravenous	lethal dose low	121	99.85	
58	frog	subcutaneous	lethal dose low	112	99.86	
59	mouse	intravenous	lethal dose low	102	99.87	
Total number of unique compounds: 80,081			Total number of measurements: 122,594			

Supplementary Table 2. Average R^2 of 5-fold cross-validation over 59 acute toxicity endpoints.

	Fold_0	Fold_1	Fold_2	Fold_3	Fold_4	5-CV (avg.)
ST-DNN	0.25 +/- 0.23	0.13 +/- 0.31	0.23 +/- 0.29	0.21 +/- 0.27	0.19 +/- 0.28	0.20 +/- 0.28
ST-RF	0.34 +/- 0.20	0.31 +/- 0.19	0.38 +/- 0.17	0.36 +/- 0.16	0.34 +/- 0.20	0.35 +/- 0.18
MT-DNN	0.51 +/- 0.18	0.50 +/- 0.18	0.52 +/- 0.16	0.52 +/- 0.13	0.51 +/- 0.16	0.51 +/- 0.16
MT-GCNN	0.50 +/- 0.18	0.48 +/- 0.18	0.50 +/- 0.18	0.53 +/- 0.13	0.49 +/- 0.16	0.50 +/- 0.17
MT-DLCA	0.54 +/- 0.16	0.51 +/- 0.17	0.55 +/- 0.16	0.55 +/- 0.12	0.54 +/- 0.14	0.54 +/- 0.15
ToxACoL	0.60 +/- 0.11	0.56 +/- 0.12	0.60 +/- 0.11	0.57 +/- 0.12	0.58 +/- 0.11	0.58 +/- 0.11

Supplementary Table 3. Average RMSE of 5-fold cross-validation over 59 acute toxicity endpoints.

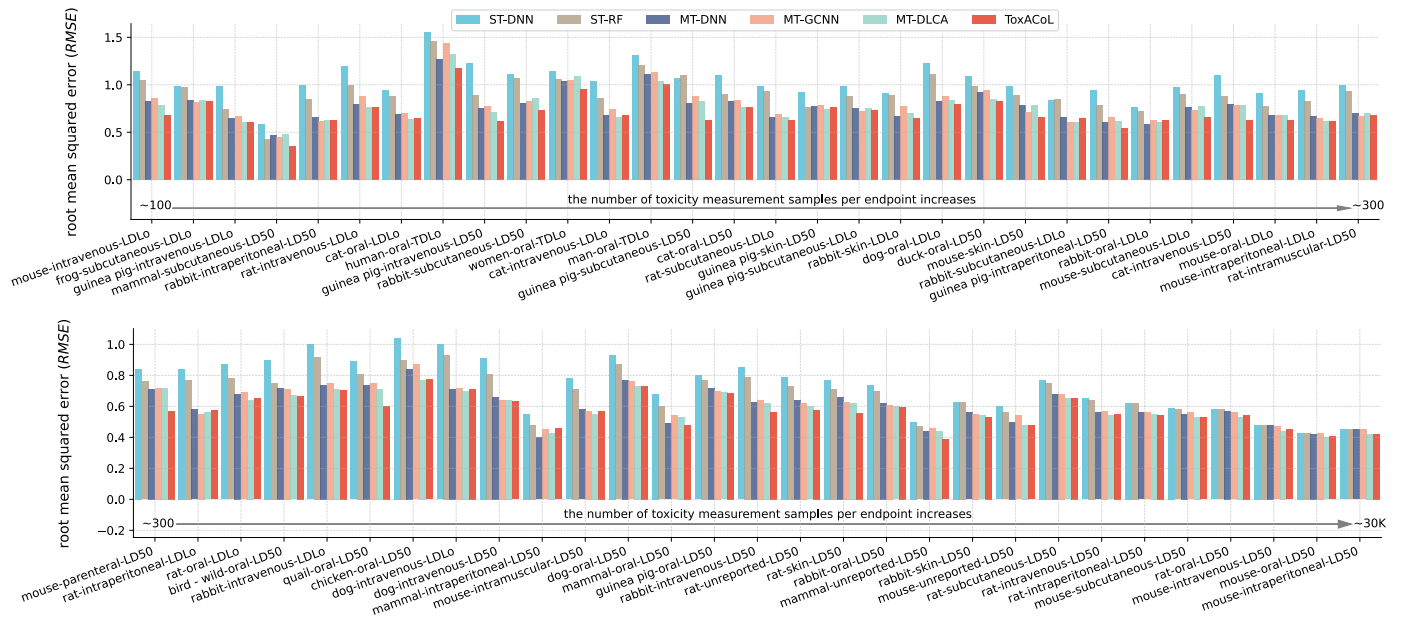
	Fold_0	Fold_1	Fold_2	Fold_3	Fold_4	5-CV (avg.)
ST-DNN	0.90 +/- 0.24	0.96 +/- 0.29	0.84 +/- 0.22	0.87 +/- 0.24	0.87 +/- 0.22	0.89 +/- 0.24
ST-RF	0.84 +/- 0.23	0.86 +/- 0.24	0.76 +/- 0.20	0.79 +/- 0.20	0.79 +/- 0.20	0.81 +/- 0.21
MT-DNN	0.72 +/- 0.20	0.72 +/- 0.18	0.66 +/- 0.17	0.67 +/- 0.16	0.68 +/- 0.17	0.69 +/- 0.18
MT-GCNN	0.73 +/- 0.21	0.74 +/- 0.19	0.67 +/- 0.19	0.67 +/- 0.16	0.70 +/- 0.18	0.70 +/- 0.19
MT-DLCA	0.70 +/- 0.19	0.71 +/- 0.19	0.64 +/- 0.17	0.66 +/- 0.17	0.66 +/- 0.17	0.67 +/- 0.18
ToxACoL	0.62 +/- 0.15	0.65 +/- 0.14	0.61 +/- 0.12	0.66 +/- 0.15	0.66 +/- 0.14	0.64 +/- 0.14

Supplementary Table 4. Average R^2 over small/large-sized endpoints (avg. +/- std.).

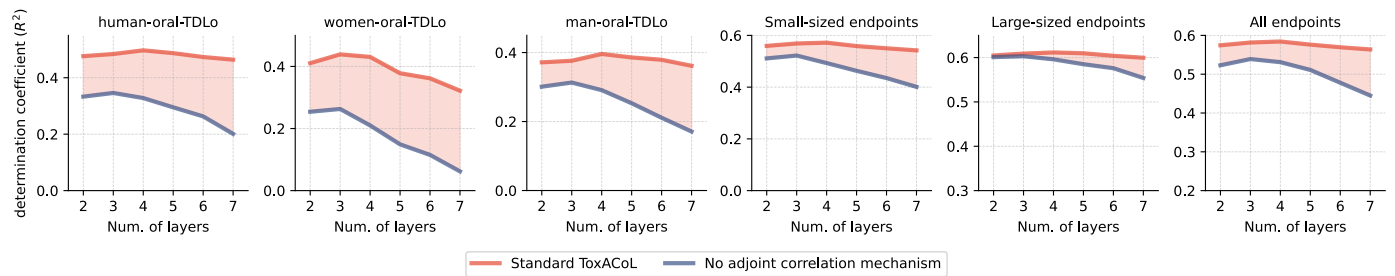
	21 small-sized endpoints	11 large-sized endpoints
ST-DNN	0.08 +/- 0.19	0.49 +/- 0.10
ST-RF	0.28 +/- 0.16	0.50 +/- 0.09
MT-DNN	0.48 +/- 0.15	0.57 +/- 0.07
MT-GCNN	0.44 +/- 0.16	0.55 +/- 0.07
MT-DLCA	0.50 +/- 0.16	0.60 +/- 0.07
ToxACoL	0.57 +/- 0.13	0.61 +/- 0.06

Supplementary Table 5. The acute toxicity evaluation performance (R^2) of ToxACoL with reducing training measurements of small-sized endpoints. The “Ratio” is the reduction ratio, and the numbers in parentheses indicate the number of training measurements per endpoint.

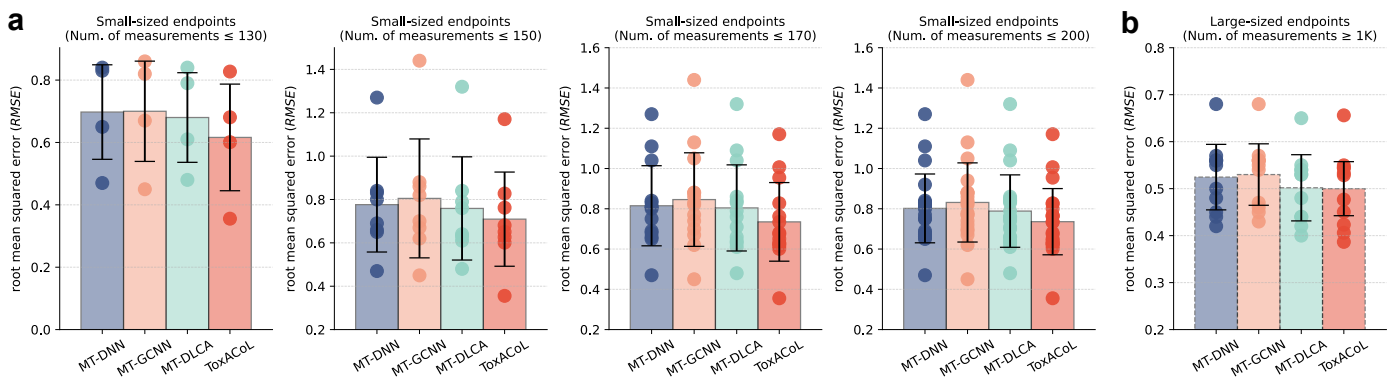
	Standard	Ratio=0.8	Ratio=0.5	Ratio=0.4	Ratio=0.3	Ratio=0.2	Ratio=0.1
human-oral-TDLo	0.50 (95 trn. meas.)	0.47 (76 trn. meas.)	0.45 (48 trn. meas.)	0.42 (38 trn. meas.)	0.38 (29 trn. meas.)	0.27 (19 trn. meas.)	0.24 (10 trn. meas.)
women-oral-TDLo	0.43 (106 trn. meas.)	0.37 (85 trn. meas.)	0.35 (53 trn. meas.)	0.25 (42 trn. meas.)	0.24 (32 trn. meas.)	0.20 (21 trn. meas.)	0.08 (11 trn. meas.)
man-oral-TDLo	0.40 (110 trn. meas.)	0.38 (88 trn. meas.)	0.33 (55 trn. meas.)	0.30 (44 trn. meas.)	0.25 (33 trn. meas.)	0.22 (22 trn. meas.)	0.15 (11 trn. meas.)
21 small-sized endpoints (avg+/- std)	0.57 +/- 0.15 (120 trn. meas.)	0.56 +/- 0.15 (96 trn. meas.)	0.53 +/- 0.15 (60 trn. meas.)	0.52 +/- 0.15 (48 trn. meas.)	0.51 +/- 0.15 (36 trn. meas.)	0.48 +/- 0.15 (24 trn. meas.)	0.43 +/- 0.15 (12 trn. meas.)



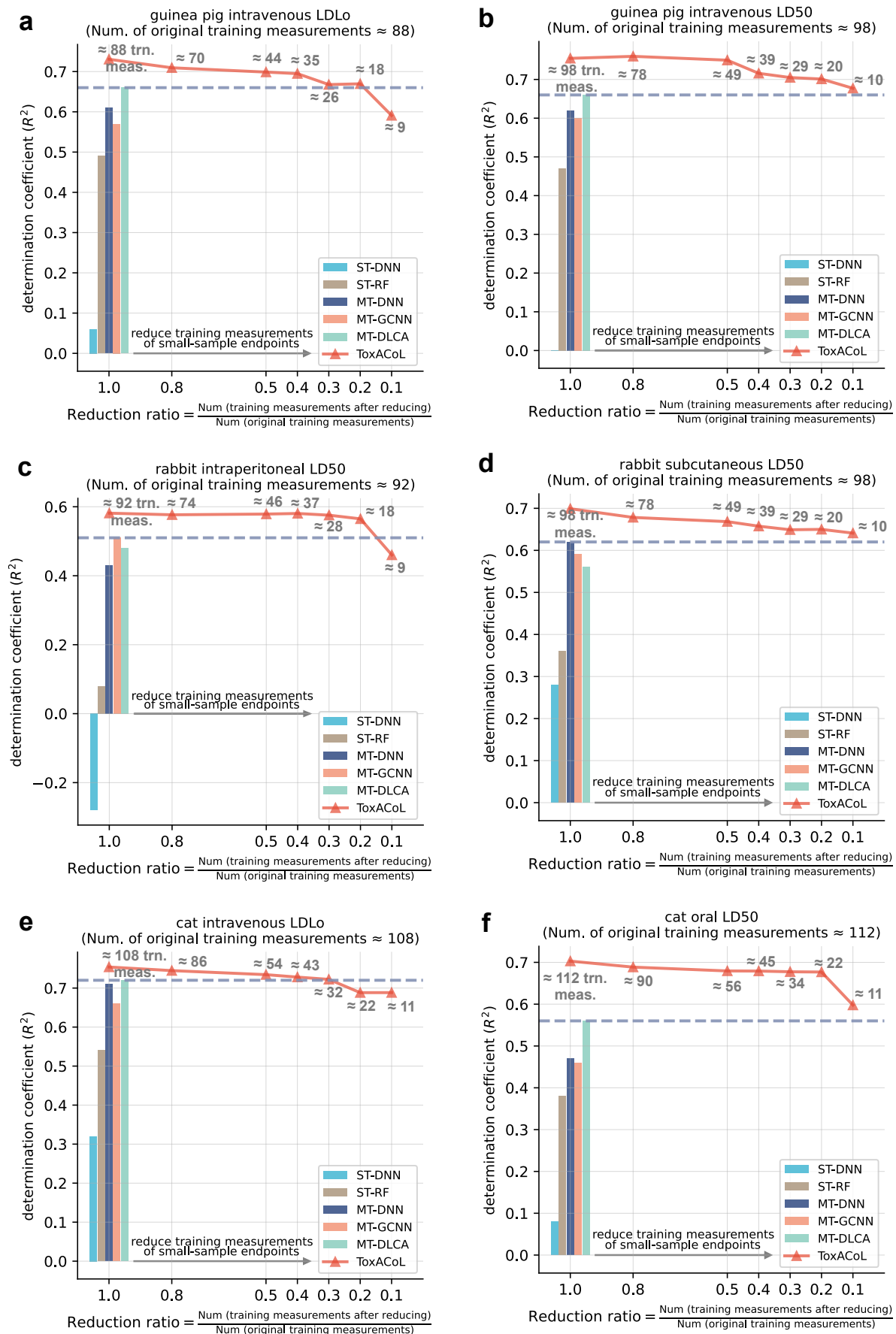
Supplementary Figure 1. Endpoint-wise RMSE achieved by different models. All toxicity endpoints were arranged from left to right in ascending order of their sample sizes of toxicity measurements.



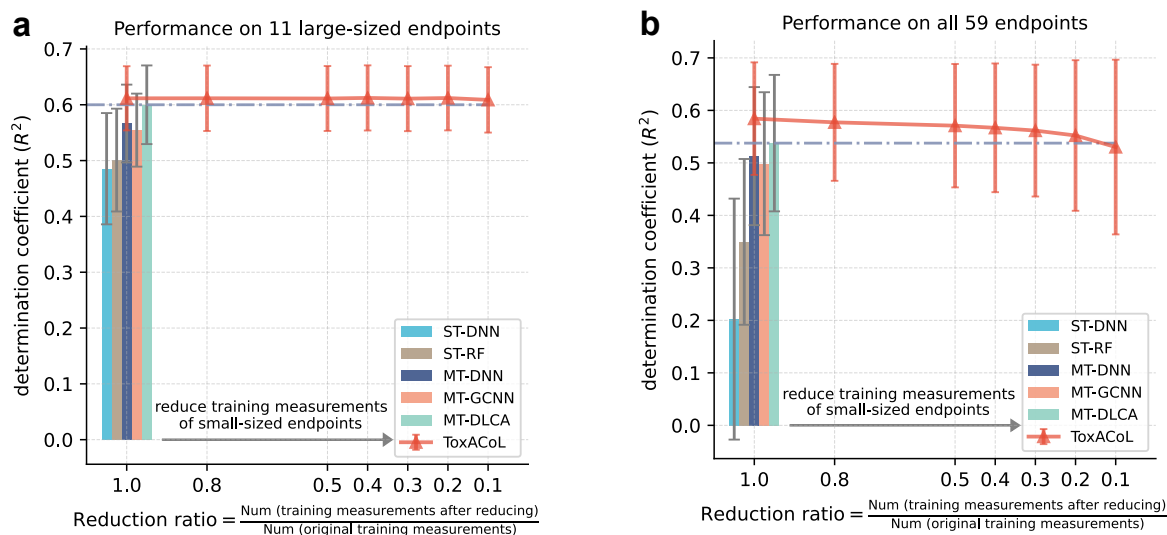
Supplementary Figure 2. Performance of ToxACoL and its variant (without adjoint correlation) varying with the network depth. Color-filled area represents the performance gap between ToxACoL and its variant. The performance was evaluated via the average R^2 after 5-fold cross-validation.



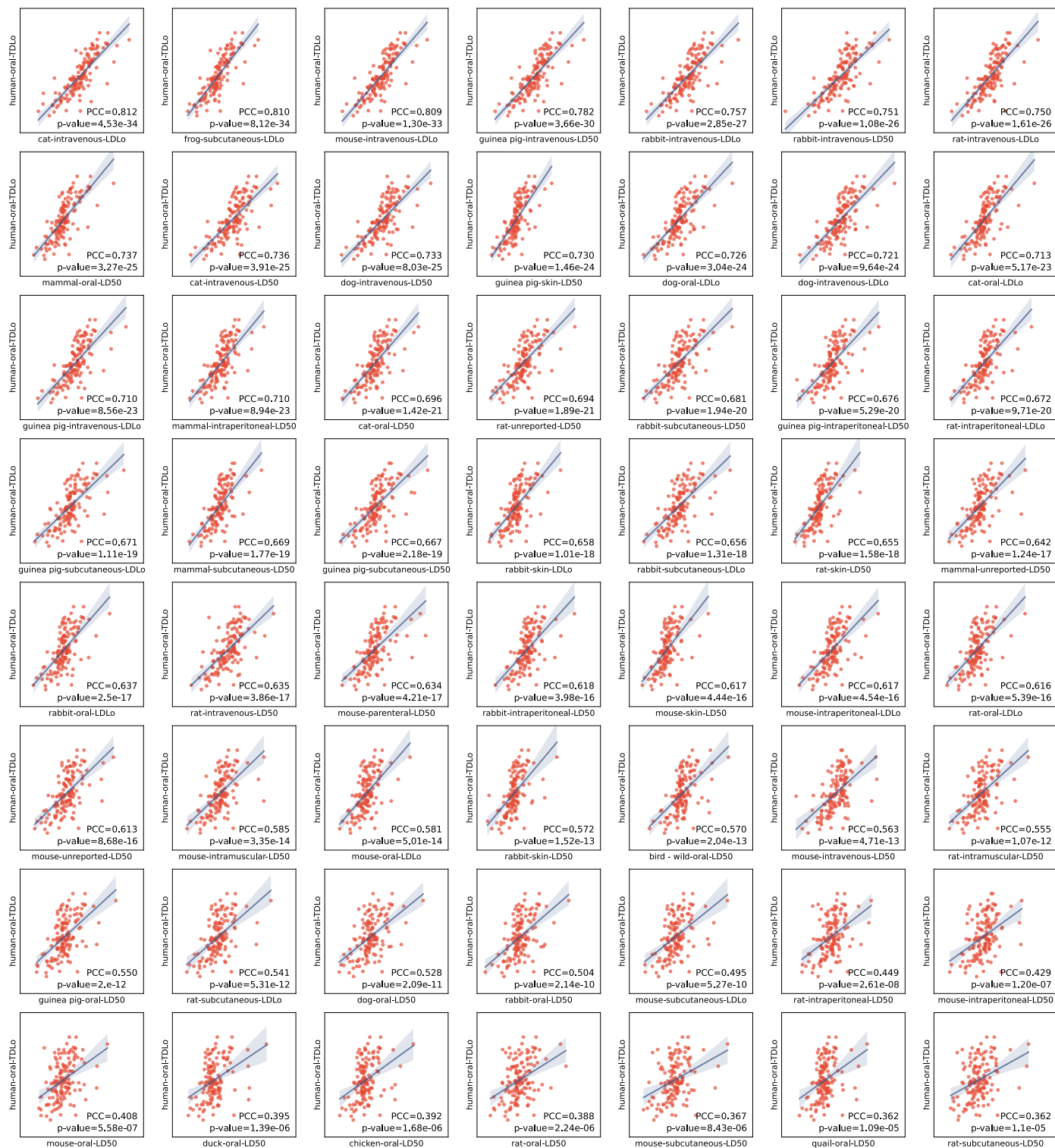
Supplementary Figure 3. **a**, Comparison of average RMSE on small-sized endpoints. Taking the first subgraph for example, it considered the 4 endpoints with a sample size of measurements less than 130, and so on for the following three subgraphs. The dots on histograms represent the RMSE values at single endpoints. **b**, Comparison of average RMSE on 11 large-sized endpoints.



Supplementary Figure 4. Acute toxicity evaluation performance of ToxACoL with reducing training measurements of small-sized endpoints. The performance at the six small-sized endpoints, including guinea pig (a-b), rabbit (c-d), and cat (e-f) related toxic endpoints, with the toxicity measurement samples used for training reduced proportionally.



Supplementary Figure 5. Acute toxicity evaluation performance of ToxACoL with reducing training measurements of small-sized endpoints. **a**, Average R^2 over 11 large-sized endpoints. **b**, Average R^2 over all 59 endpoints. Error bars denote the standard deviation among endpoints.



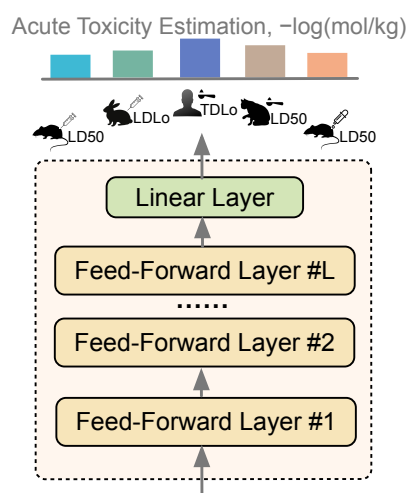
Supplementary Figure 6. Pearson correlation coefficient (PCC) values between human-oral-TDLo and other 56 animal endpoints. Note that there are a total of 140 compounds in the dataset that have toxicity measurement values at human-oral-TDLo endpoint. The missing toxicity intensity values of these 140 compounds at the 56 animal endpoints were filled in by the predicted intensity values of ToxACoL. Thus, the PCC value between two endpoints was calculated based on the toxicity intensity values of the 140 compounds at the two endpoints. The error bands in the correlation plots represent the confidence interval of 0.95 for linear regression.



Supplementary Figure 7. PCC values between women-oral-TDLo and other 56 animal endpoints. Note that there are a total of 156 compounds in the dataset that have toxicity measurement values at women-oral-TDLo endpoint. The missing toxicity intensity values of these 156 compounds at the 56 animal endpoints were filled in by the predicted intensity values of ToxACoL.



Supplementary Figure 8. PCC values between man-oral-TDLo and other 56 animal endpoints. Note that there are a total of 163 compounds in the dataset that have toxicity measurement values at man-oral-TDLo endpoint. The missing toxicity intensity values of these 163 compounds at the 56 animal endpoints were filled in by the predicted intensity values of ToxACoL.



Supplementary Figure 9. Pipeline of previous vanilla multi-task learning models for multi-condition acute toxicity evaluation. It extracts high-level representations of compounds using a shared network and deploys a multi-channel linear regression layer to predict the toxicity at various toxic endpoints.