ACS AUTHORCHOICE
Erratum

JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

Cite This: J. Chem. Inf. Model. 2019, 59, 5304−5305

pubs.acs.org/jcim

Correction to Analyzing Learned Molecular Representations for Property Prediction

Kevin Yang,*[,†] Kyle Swanson,*[,†] Wengong Jin,[†] Connor Coley,[‡] Philipp Eiden,[¶] Hua Gao,[§] Angel Guzman-Perez,[§] Timothy Hopper,[§] Brian Kelley,[∥] Miriam Mathea,[¶] Andrew Palmer,[¶] Volker Settels,[¶] Tommi Jaakkola,[†] Klavs Jensen,[‡] and Regina Barzilay[†]

[†]Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, Massachusetts 02139, United States

[‡]Department of Chemical Engineering, MIT, Cambridge, Massachusetts 02139, United States

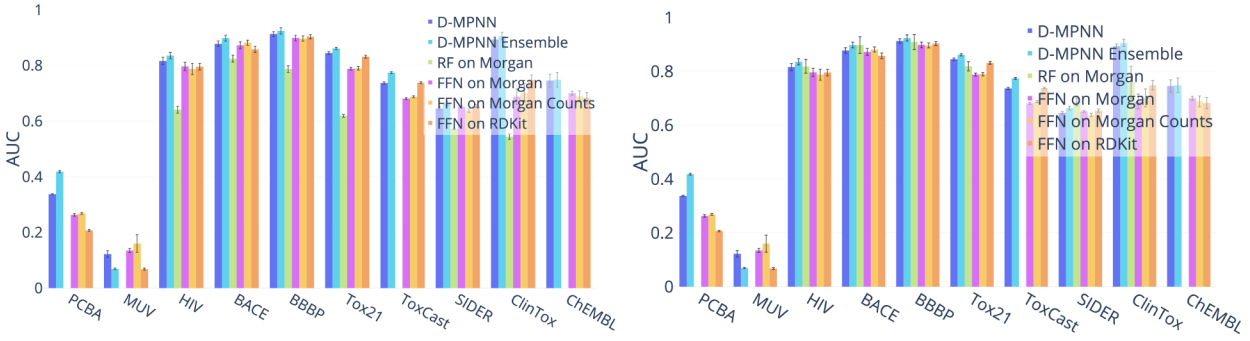
[¶]BASF SE, Ludwigshafen 67063, Germany

[§]Amgen Inc., Cambridge, Massachusetts 02141, United States

[∥]Novartis Institutes for BioMedical Research, Cambridge, Massachusetts 02139, United States

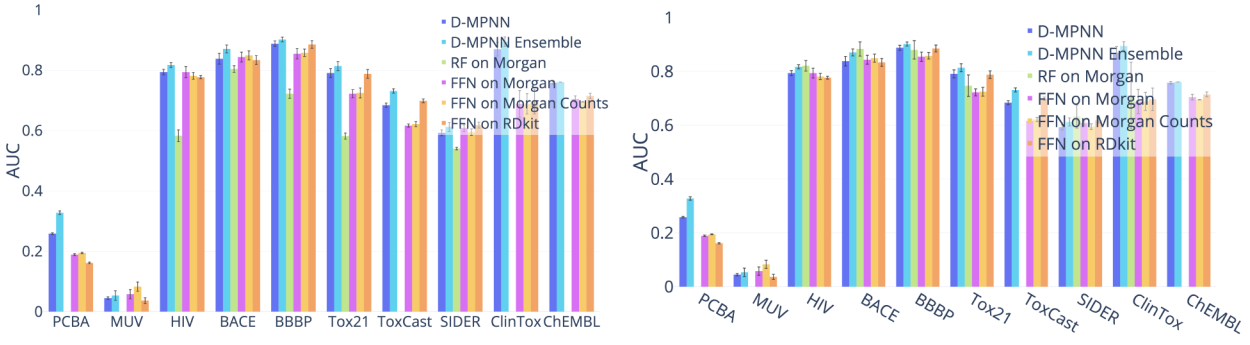
J. Chem. Inf. Model. **2019**, *59* (8), 3370−3388. DOI: 10.1021/acs.jcim.9b00237

Due to an error in the processing of the random forest model's predictions on classification data sets, our original random forest AUC numbers were incorrect on six public classification data sets—HIV, BACE, BBBP, Tox21, SIDER, and ClinTox—and on one proprietary classification data set—hPXR (class). We fixed the error and reran the random forest experiments. After the fix, the random forest model performs better than previously reported, though our D-MPNN continues to outperform it on some classification data sets and on all but one of the regression data sets. Additionally,



(a) Comparison to baselines (original). (b) Comparison to baselines (fixed).

Figure 1. (random split, higher = better) Comparison to baselines on public data sets with original (left) and fixed (right) random forest numbers using a random split.

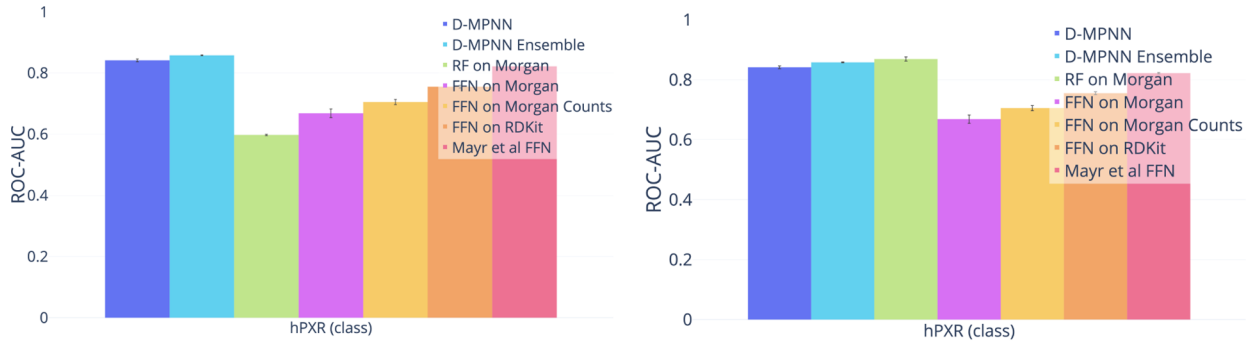


(a) Comparison to baselines (original). (b) Comparison to baselines (fixed).

Figure 2. (scaffold split, higher = better) Comparison to baselines on public data sets with original (left) and fixed (right) random forest numbers using a scaffold split.

Published: December 9, 2019

DOI: 10.1021/acs.jcim.9b01076
J. Chem. Inf. Model. 2019, 59, 5304−5305



(a) Comparison to baselines (original). (b) Comparison to baselines (fixed).

Figure 3. (time split, higher = better) Comparison to baselines on Amgen data set with original (left) and fixed (right) random forest numbers using a time split.

Table 1. (Random Split, Higher = Better) Comparison to Baselines on Public Datasets with Original and Fixed Random Forest Numbers Using a Random Split

data set	metric	D-MPNN	D-MPNN ensemble	RF on Morgan (original)	RF on Morgan (fixed)
HIV	ROC-AUC	0.816 ± 0.023	0.836 ± 0.020 (+2.40% p = 0.01)	0.641 ± 0.022 (−21.45% p = 0.00)	0.819 ± 0.025 (+0.31% p = 0.97)
BACE	ROC-AUC	0.878 ± 0.032	0.898 ± 0.034 (+2.31% p = 0.00)	0.825 ± 0.039 (−6.08% p = 0.00)	0.898 ± 0.031 (+2.26% p = 1.00)
BBBP	ROC-AUC	0.913 ± 0.026	0.925 ± 0.036 (+1.23% p = 0.01)	0.788 ± 0.038 (−13.77% p = 0.00)	0.909 ± 0.028 (−0.42% p = 0.19)
Tox21	ROC-AUC	0.845 ± 0.015	0.861 ± 0.012 (+1.95% p = 0.00)	0.619 ± 0.015 (−26.75% p = 0.00)	0.819 ± 0.017 (−3.06% p = 0.00)
SIDER	ROC-AUC	0.646 ± 0.016	0.664 ± 0.021 (+2.79% p = 0.01)	0.572 ± 0.007 (−11.38% p = 0.00)	0.687 ± 0.014 (+6.35% p = 1.00)
ClinTox	ROC-AUC	0.894 ± 0.027	0.906 ± 0.043 (+1.33% p = 0.05)	0.544 ± 0.031 (−39.13% p = 0.00)	0.759 ± 0.060 (−15.12% p = 0.00)

Table 2. (Scaffold Split, Higher = Better) Comparison to Baselines on Public Datasets with Original and Fixed Random Forest Numbers Using a Scaffold Split

data set	metric	D-MPNN	D-MPNN ensemble	RF on Morgan (original)	RF on Morgan (fixed)
HIV	ROC-AUC	0.794 ± 0.016	0.817 ± 0.013 (+2.94% p = 0.00)	0.583 ± 0.034 (−26.59% p = 0.00)	0.821 ± 0.020 (+3.42% p = 0.99)
BACE	ROC-AUC	0.838 ± 0.056	0.871 ± 0.041 (+3.89% p = 0.00)	0.804 ± 0.035 (−4.04% p = 0.01)	0.884 ± 0.026 (+5.43% p = 1.00)
BBBP	ROC-AUC	0.888 ± 0.029	0.902 ± 0.024 (+1.56% p = 0.01)	0.722 ± 0.049 (−18.68% p = 0.00)	0.880 ± 0.034 (−0.88% p = 0.45)
Tox21	ROC-AUC	0.791 ± 0.047	0.814 ± 0.047 (+2.89% p = 0.00)	0.582 ± 0.031 (−26.42% p = 0.00)	0.747 ± 0.040 (−5.54% p = 0.00)
SIDER	ROC-AUC	0.593 ± 0.032	0.612 ± 0.047 (+3.31% p = 0.03)	0.540 ± 0.013 (−8.79% p = 0.00)	0.632 ± 0.043 (+6.75% p = 1.00)
ClinTox	ROC-AUC	0.870 ± 0.072	0.895 ± 0.050 (+2.86% p = 0.01)	numerically unstable	0.711 ± 0.123 (−18.24% p = 0.00)

Table 3. (Time Split, Higher = Better) Comparison to Baselines on Amgen Dataset with Original and Fixed Random Forest Numbers Using a Time Split

data set	metric	D-MPNN	D-MPNN ensemble	RF on Morgan (original)	RF on Morgan (fixed)
hPXR (class)	ROC-AUC	0.842 ± 0.008	0.858 ± 0.002 (+1.95%)	0.598 ± 0.004 (−28.98%)	0.869 ± 0.007 (+3.28%)

Table 4. Number of Public Datasets Where D-MPNN is Statistically Significantly Better than, Equivalent to, or Worse than Random Forest

baseline	D-MPNN is *better*	D-MPNN is the *same*	D-MPNN is *worse*	no. data sets
RF on Morgan (original)	14	0	1	15
RF on Morgan (fixed)	9	1	4	15

since the fixed random forest model is better than our D-MPNN on BACE and hPXR (class), our D-MPNN now achieves comparable or better performance than all baseline models on 11 rather than 12 of the 19 public data sets and on 15 rather than 16 of the 16 proprietary data sets. The results of the other 800+ experiments we report in the paper are unaffected. The tables and figures included here show the changes.

Separately, on page 3372 the learned matrix $W_a \in \mathbb{R}^{h\times h}$ should be $W_a \in \mathbb{R}^{h\times h_a}$ where $\text{cat}(x_v, m_v) \in \mathbb{R}^{h_a}$.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: yangk@mit.edu.
*E-mail: swansonk@mit.edu.

ORCID

Kyle Swanson: 0000-0002-7385-7844
Connor Coley: 0000-0002-8271-8723
Klavs Jensen: 0000-0001-7192-580X

DOI: 10.1021/acs.jcim.9b01076
J. Chem. Inf. Model. 2019, 59, 5304−5305