Digital Discovery

ROYAL SOCIETY OF CHEMISTRY

CORRECTION

View Article Online
View Journal | View Issue



Cite this: *Digital Discovery*, 2024, **3**, 1069

# Correction: Predicting small molecules solubility on endpoint devices using deep ensemble neural networks

Mayk Caldas Ramos and Andrew D. White *

DOI: 10.1039/d4dd90020k

rsc.li/digitaldiscovery

Correction for 'Predicting small molecules solubility on endpoint devices using deep ensemble neural networks' by Mayk Caldas Ramos and Andrew D. White, *Digital Discovery*, 2024, **3**, 786–795, https://doi.org/10.1039/D3DD00217A.

The header row in Table 2 is incorrect. The correct version of Table 2 is displayed below. Please note that the references are reproduced here as ref. 1–13.

Table 2 Metrics for the best models found in the current study (upper section) and for other state-of-the-art models available in the literature (lower section). Values were taken from the cited references. Missing values stand for entries that the cited authors did not study. SolChal columns stand for the solubility challenges. 2_1 represents the tight dataset (set-1), while 2_2 represents the loose dataset (set-2) as described in the original paper (see ref. 1). The best-performing metrics value are displayed in bold

| | SolChal1 | | SolChal2_1 | | SolChal2_2 | | ESOL | |
|---|---|---|---|---|---|---|---|---|
| Model | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE |
| RF | 1.121 | 0.914 | **0.950** | **0.727** | **1.205** | **1.002** | | |
| DNN | 1.540 | 1.214 | 1.315 | 1.035 | 1.879 | 1.381 | | |
| $DNN_{Aug}$ | 1.261 | 1.007 | 1.371 | 1.085 | 2.189 | 1.710 | | |
| $kde4_{Aug}^{LSTM}$ | 1.273 | 0.984 | 1.137 | 0.932 | 1.511 | 1.128 | 1.397 | 1.131 |
| $kde8_{Aug}^{LSTM}$ | 1.247 | 0.984 | 1.044 | 0.846 | 1.418 | 1.118 | 1.676 | 1.339 |
| $kde10_{Aug}^{LSTM}$ | **1.095** | **0.843** | 0.983 | 0.793 | 1.263 | 1.051 | **1.316** | **1.089** |
| Linear regression[2] | | | | | | | 0.75 | |
| UG-RNN[3] | 0.90 | 0.74 | | | | | | |
| RF w/CDF descriptors[4] | 0.93 | | | | | | | |
| RF w/Morgan fingerprints[5] | | **0.64** | | | | | | |
| Consensus[6] | **0.91** | | | | | | | |
| GNN[7] | ~1.10 | | **0.91** | | **1.17** | | | |
| SolvBert[8] | 0.925 | | | | | | | |
| SolTranNet[a,9] | | | 1.004 | | 1.295 | | 2.99 | |
| SMILES-BERT[b,10] | | | | | | | 0.47 | |
| MolBERT[b,11] | | | | | | | 0.531 | |
| RT[b,12] | | | | | | | 0.73 | |
| MolFormer[b,13] | | | | | | | **0.278** | |

[a] Has overlap between training and test sets. [b] Pre-trained model was fine-tuned on ESOL.

The Royal Society of Chemistry apologises for these errors and any consequent inconvenience to authors and readers.

## References

1 A. Llinas and A. Avdeef, Solubility Challenge revisited after ten years, with multilab shake-flask data, using tight (SD 0.17 log) and loose (SD 0.62 log) test sets, *J. Chem. Inf. Model.*, 2019, **59**, 3036–3040.

*Department of Chemical Engineering, University of Rochester, Rochester, NY 14627. E-mail: andrew.white@rochester.edu*

2 J. S. Delaney, ESOL: estimating aqueous solubility directly from molecular structure, *J. Chem. Inf. Comput. Sci.*, 2004, **44**, 1000–1005.
3 A. Lusci, G. Pollastri and P. Baldi, Deep architectures and deep learning in chemoinformatics: the prediction of aqueous solubility for drug-like molecules, *J. Chem. Inf. Model.*, 2013, **53**, 1563–1575.
4 J. L. McDonagh, N. Nath, L. De Ferrari, T. van Mourik and J. B. O. Mitchell, Uniting cheminformatics and chemical theory to predict the intrinsic aqueous solubility of crystalline druglike molecules, *J. Chem. Inf. Model.*, 2014, **54**, 844–856.
5 A. Tayyebi, A. S. Alshami, Z. Rabiei, X. Yu, N. Ismail, M. J. Talukder and J. Power, Prediction of organic compound aqueous solubility using machine learning: a comparison study of descriptor-based and fingerprints-based models, *J. Cheminf.*, 2023, **15**, 99.
6 S. Boobier, A. Osbourn and J. B. O. Mitchell, Can human experts predict solubility better than computers?, *J. Cheminf.*, 2017, **9**, 63.
7 G. Panapitiya, M. Girard, A. Hollas, J. Sepulveda, V. Murugesan, W. Wang and E. Saldanha, Evaluation of Deep Learning Architectures for Aqueous Solubility Prediction, *ACS Omega*, 2022, **7**, 15695–15710.
8 J. Yu, C. Zhang, Y. Cheng, Y.-F. Yang, Y.-B. She, F. Liu, W. Su and A. Su, SolvBERT for solvation free energy and solubility prediction: a demonstration of an NLP model for predicting the properties of molecular complexes, *Digital Discovery*, 2023, **2**, 409–421.
9 P. G. Francoeur and D. R. Koes, SolTranNet–A Machine Learning Tool for Fast Aqueous Solubility Prediction, *J. Chem. Inf. Model.*, 2021, **61**, 2530–2536.
10 H. Kim, J. Lee, S. Ahn and J. R. Lee, A merged molecular representation learning for molecular properties prediction with a web-based service, *Sci. Rep.*, 2021, **11**, 11028.
11 B. Fabian, T. Edlich, H. Gaspar, M. Segler, J. Meyers, M. Fiscato and M. Ahmed, Molecular representation learning with language models and domain-relevant auxiliary tasks, *arXiv*, 2020, preprint, arXiv:2011.13230, DOI: **10.48550/arXiv.2011.13230**.
12 J. Born and M. Manica, Regression Transformer enables concurrent sequence regression and generation for molecular language modelling, *Nat. Mach. Intell.*, 2023, **5**, 432–444.
13 J. Ross, B. Belgodere, V. Chenthamarakshan, I. Padhi, Y. Mroueh and P. Das, Large-scale chemical language representations capture molecular structure and properties, *Nat. Mach. Intell.*, 2022, **4**(12), 1256–1264.