Predicting Potentially Hazardous Chemical Reactions Using Explainable Neural Network

Juhwan Kim, ‡[a] Geun Ho Gu, ‡[a] Juhwan Noh, ‡[a] Seongun Kim,[b] Suji Gim,[c] Jaesik Choi*[b] and Yousung

[a] Department of Chemical and Biomolecular Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea. E-mail: ysjn@kaist.ac.kr

[b] Graduate School of Artificial Intelligence, KAIST Daejeon: 291 Daehak-ro, N24, Yuseong-gu, Daejeon 34141, Republic of Korea. E-mail: jaesik.choi@kaist.ac.kr

[c] Environment & Safety Research Center, 1, Samsungjeonja-ro, Hwasung-si, Gyeonggi-do, Republic of Korea. E-mail : suji.gim@samsung.com

‡G. G. and J. N. contributed equally to this work.

Jung*[a]

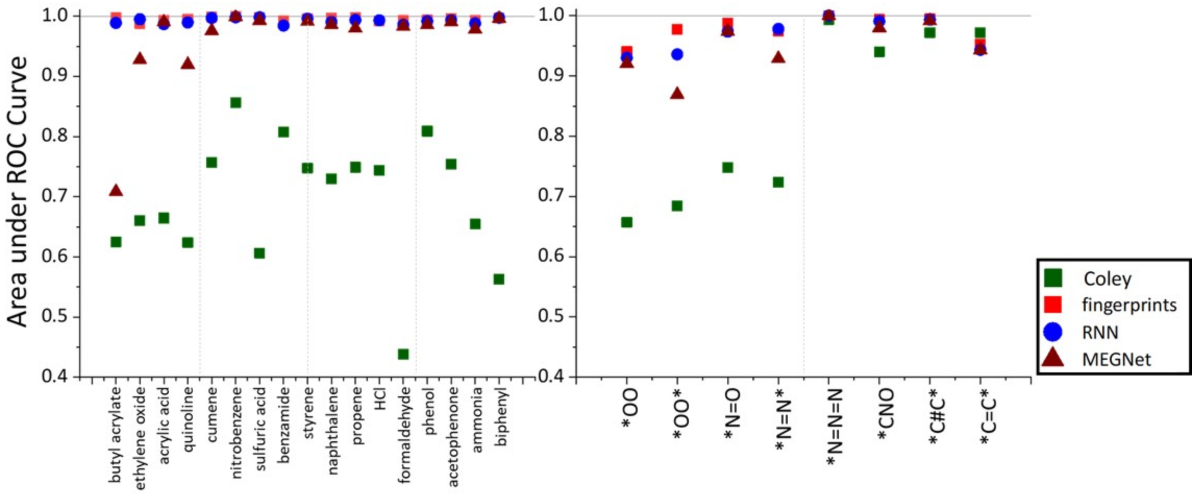

Figure 1 The area under receiver operating characteristic curve (AUC) of fingerprint-based binary prediction models. The receiver operating characteristic curve of the fingerprints-based model, RNN-based model, and MEGNet is plotted with a various thresholds. The receiver operating characteristic curve of the Coley model used the 'n' of the top n model as a threshold. The binary models trained with our data show higher AUC than the product prediction model of Coley et al[1].

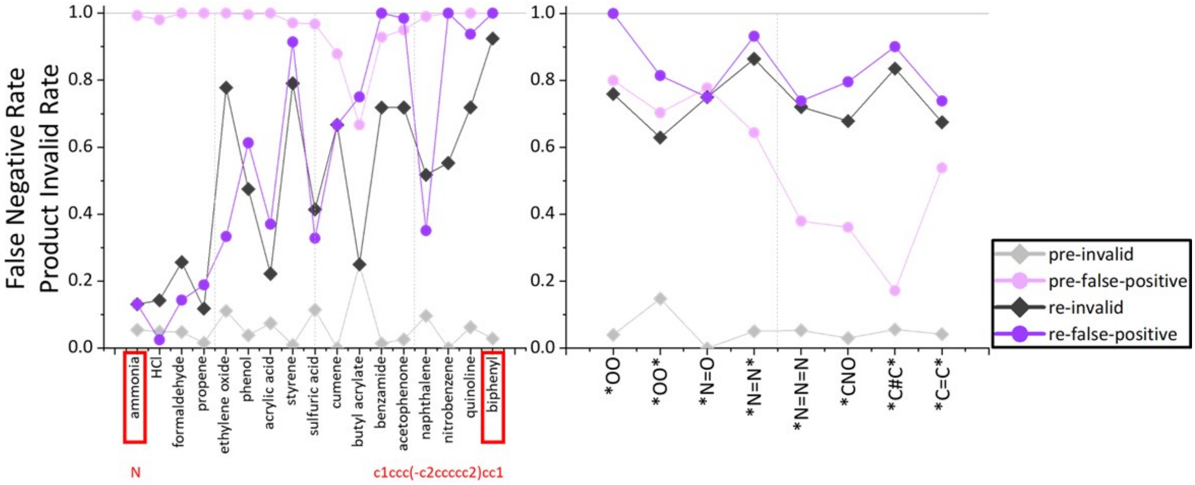

Figure 2 The false negative rate is plotted with the pink for the pre-trained Molecular Transformer and the purple for the re-trained Molecular Transformer. The product invalid rate is plotted with gray for the pre-trained Molecular Transformer and black for the re-trained Molecular Transformer. The product invalid rate is the ratio of invalid SMILES among products. The product invalid rate and the false negative rate have a similar tendency. The false negative rate and product invalid rate increase as the size of the molecule increases. The re-trained Molecular transformer shows lower accuracy than the pre-trained Molecular transformer for large molecules, and potentially explosive materials because of the high product invalid rate.

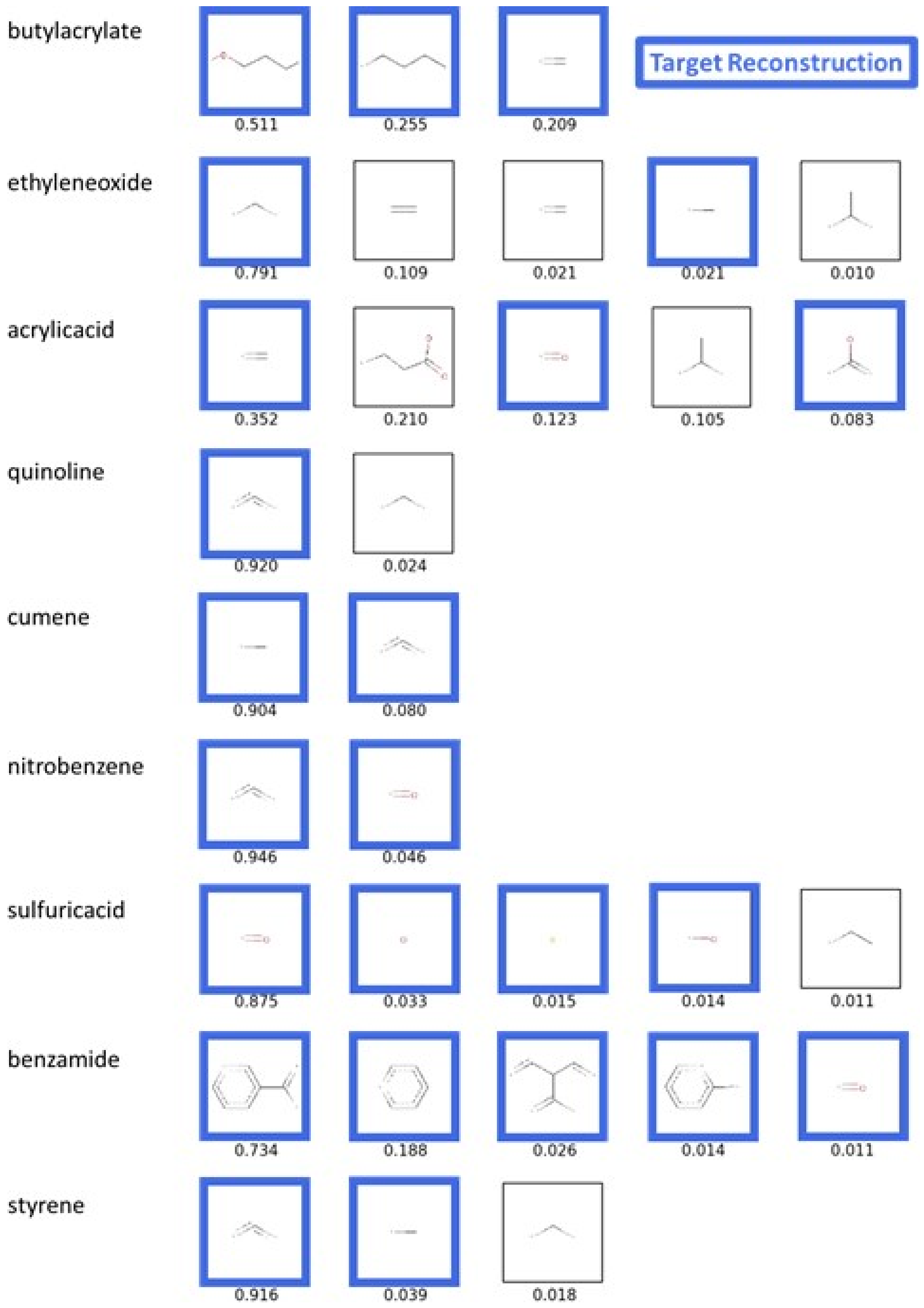

Figure 3 – 1 The most frequently highlighted substructure of the toxin prediction models is plotted. Only substructures with a frequency higher than 1% are plotted. The target reconstruction substructures are emphasized with blue edges.

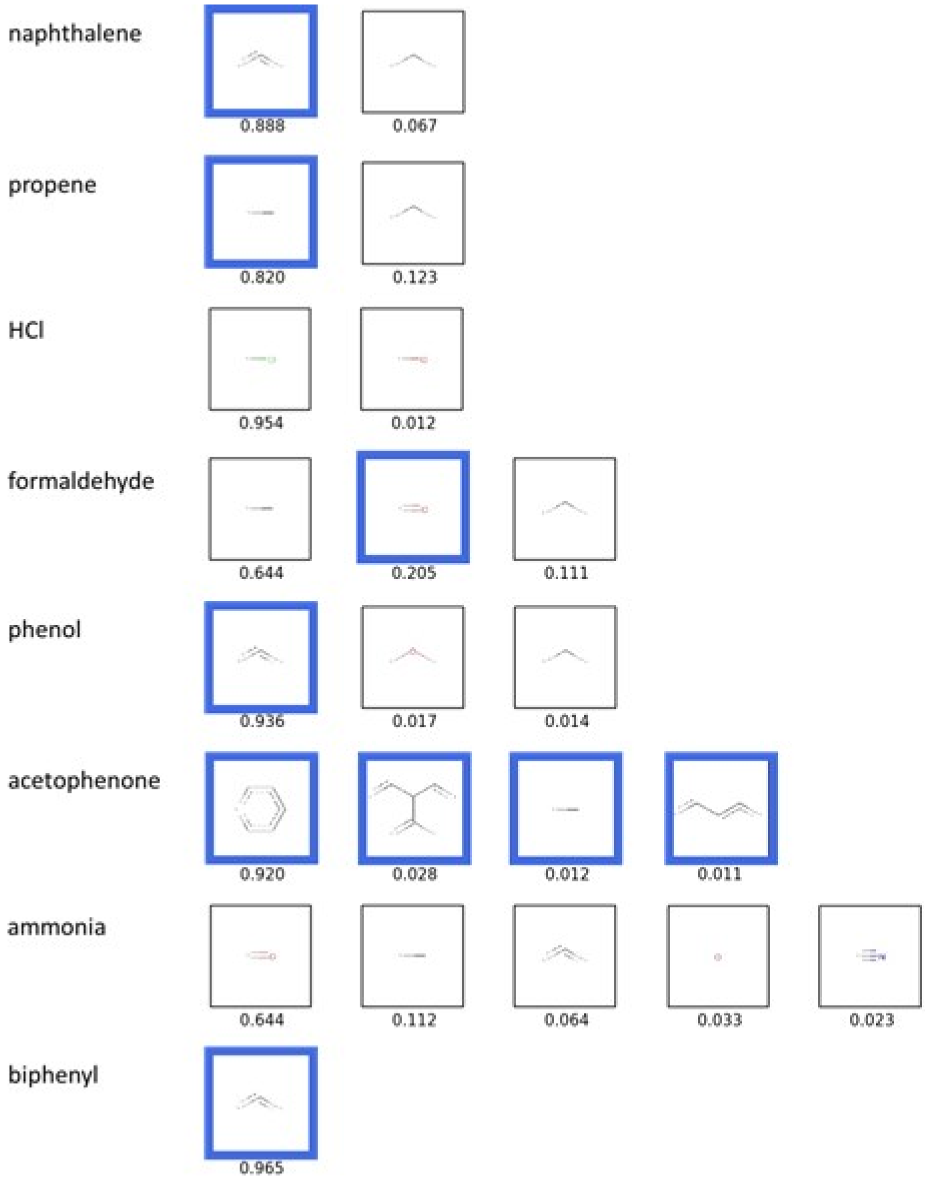

Figure 3 – 2 The most frequently highlighted substructure of the toxin prediction models is plotted. Only substructures with a frequency higher than 1% are plotted. The target reconstruction substructures are emphasized with blue edges.

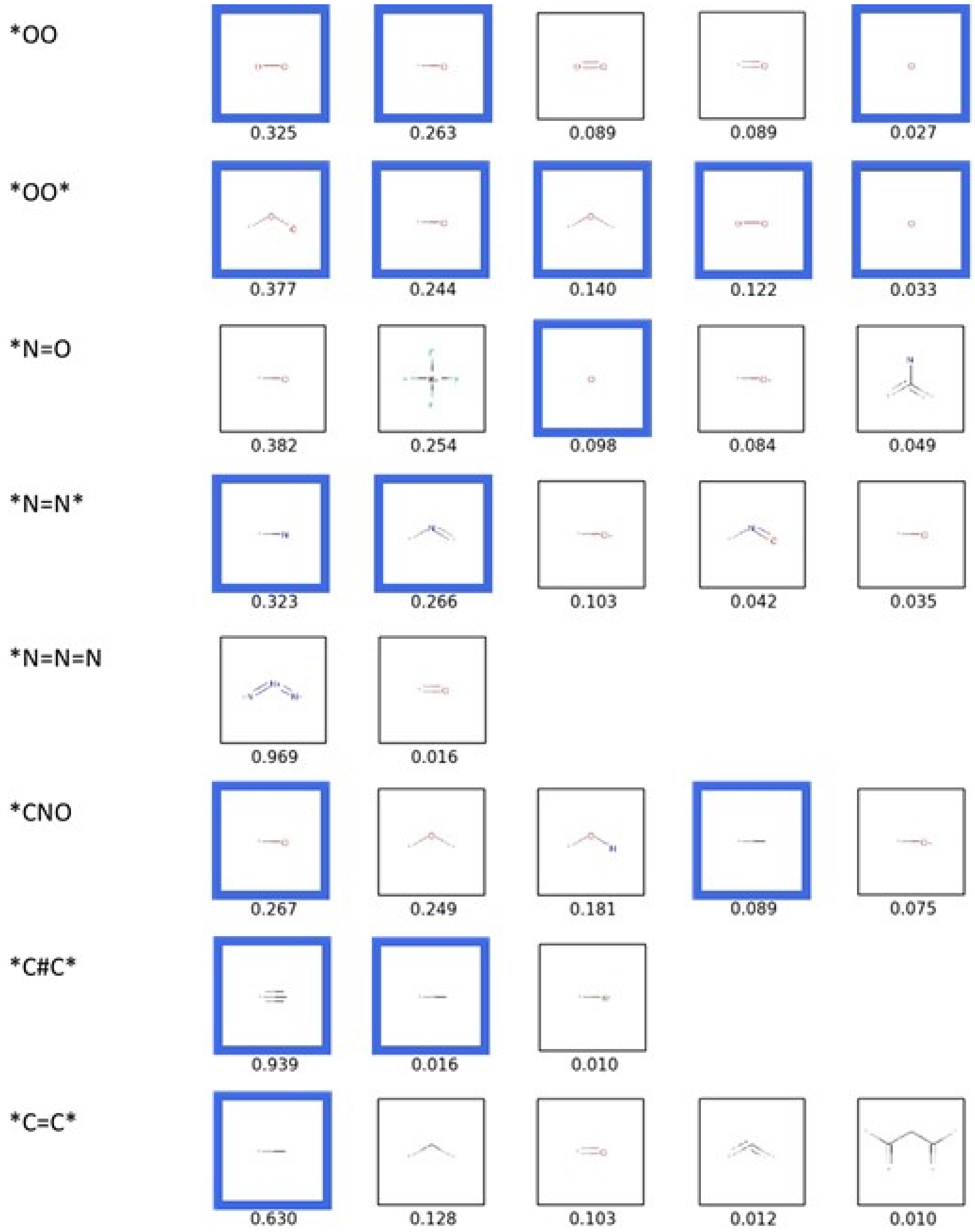

Figure 4 The most frequently highlighted substructure of the potentially explosive prediction models is plotted. Only substructures with a frequency higher than 1% are plotted. The target reconstruction substructures are emphasized with blue edges.

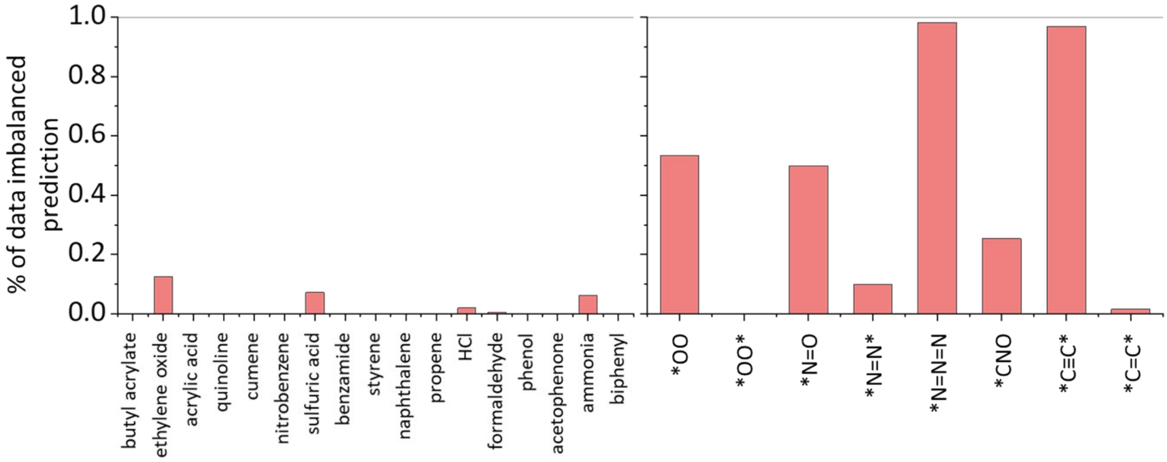

Figure 5 The percentage of the potentially biased predictions by data-imbalance for the chosen toxin and explosive substructures identified by the LRP analysis. The substructure is defined to have potential data imbalance if the ratio of the given substructure occurrence in the positive training data to those in the negative data is more than 2:1, instead of 5:1 in main text.

References

1. C. W. Coley, W. Jin, L. Rogers, T. F. Jamison, T. S. Jaakkola, W. H. Green, R. Barzilay and K. F. Jensen, *Chemical science*, 2019, **10**, 370-377.