Supplementary Material

Anonymous Author(s)
Affiliation
Address
email

1 Training details

For both experiments, toxicity estimation of PFAS molecules [1] [1] and biodegradability of general molecules [2], we divided the datasets into: 60% for training purposes, 20% for validation purposes, and 20% for testing purposes. In both cases, the divisions have been made following the original papers in order to have a fair comparison with the state-of-the-art algorithms.

Table 1 refers to the hyper-parameters considered for the toxicity estimation task.

Table 1: MULTIMODAL-MOLFORMER hyper-parameters for toxicity prediction considering the LDToxDB dataset.

Hyper-parameter	**Value**
Device	cuda
GPU	1
Batch Size	32
Number of heads	12
Number of embeddings	768
Dropout	0.1
Learning rate	3e-5
Number of workers	8
Number of epochs	2000
Dimension of the Feed-Forward network	[797 797 797 1]

Table 2 refers to the hyper-parameters considered for biodegradability classification..

Table 2: MULTIMODAL-MOLFORMER hyper-parameters for biodegradability prediction.

Hyper-parameter	**Value**
Device	cuda
GPU	1
Batch Size	32
Number of heads	12
Number of layers	12
Number of embeddings	768
Dropout	0.1
Learning rate	3e-5
Number of workers	8
Number of epochs	2000
Dimension of the Feed-Forward network	[773 773 773 1]

Both experiments have been conducted in the IBM computing cluster with 32 GB of RAM memory and a Nvidia V100 GPU.

The experiments conducted in this research are based on the recently introduced MOLFORMER approach which can be accessed through this link (`https://github.com/IBM/molformer`).

[1] AI4PFAS project `https://github.com/AI4PFAS/AI4PFAS`.

Submitted to 37th Conference on Neural Information Processing Systems (NeurIPS 2023). Do not distribute.

To extract the physico-chemical features based on the Mordred descriptors we used the library available at `http://mordred-descriptor.github.io/documentation/v0.1.0/introduction.html` [3]. To build the Markov-Blanket causal feature selection blocks we used the library available at `https://github.com/atif-hassan/PyImpetus` [4]. The default parameters were used for the physico-chemical extraction from SMILES and to build the Markov-Blanket causal graphs.

References

[1] J. Feinstein, G. Sivaraman, K. Picel, B. Peters, Á. Vázquez-Mayagoitia, A. Ramanathan, M. MacDonell, I. Foster, and E. Yan, "Uncertainty-informed deep transfer learning of perfluoroalkyl and polyfluoroalkyl substance toxicity," *Journal of chemical information and modeling*, vol. 61, no. 12, pp. 5793–5803, 2021.

[2] M. Lee and K. Min, "A comparative study of the performance for predicting biodegradability classification: the quantitative structure–activity relationship model vs the graph convolutional network," *ACS omega*, vol. 7, no. 4, pp. 3649–3655, 2022.

[3] H. Moriwaki, Y.-S. Tian, N. Kawashita, and T. Takagi, "Mordred: a molecular descriptor calculator," *Journal of cheminformatics*, vol. 10, no. 1, pp. 1–14, 2018.

[4] A. Hassan, J. H. Paik, S. Khare, and S. A. Hassan, "Ppfs: Predictive permutation feature selection," *arXiv preprint arXiv:2110.10713*, 2021.