Multi-task QSPR Modelling of the Antioxidant Activity of Organic Compounds

Nadine BENZIADA[1], Gilles Marcou[1]

[1] Laboratory of Chemoinformatics, UMR 7140 CNRS / Université de Strasbourg, 4 rue Blaise Pascal, 67000 Strasbourg, France.

Antioxidant activity data from in vitro assays are notoriously heterogeneous: diverse units and protocols systematically shift the measured potency of the same compound, which limits the transferability of single-protocol QSAR models. To address this issue, we propose to build antioxidant activity QSAR models that are protocol-aware. We collected and curated a paper-verified, protocol-aware multi-assay antioxidant dataset (DPPH, ABTS, ORAC, FRAP, HRSA, NOSA, MCA, CUPRAC) from the Antioxidant Open Database (AODB) [1] and expert-curated primary literature. As an illustration, we focus on the DPPH radical-scavenging activity data, which were organized into six condition-dependent subsets based on solvent and temperature. Molecular structures were standardized and embedded using ISIDA fragment descriptors[2]. Each subset has been the target of single-task learning models and benchmarked with a multi-task learning approach[3]. The best performance was obtained with MTL Trace. It exemplifies that sharing information across related assay conditions improves predictive performance. This approach is extended to cover other antioxidant potency measures, independently or together in a larger multi-task approach. Protocol-aware curation combined with multi-task modelling improves both the prediction and interpretation of antioxidant activity.

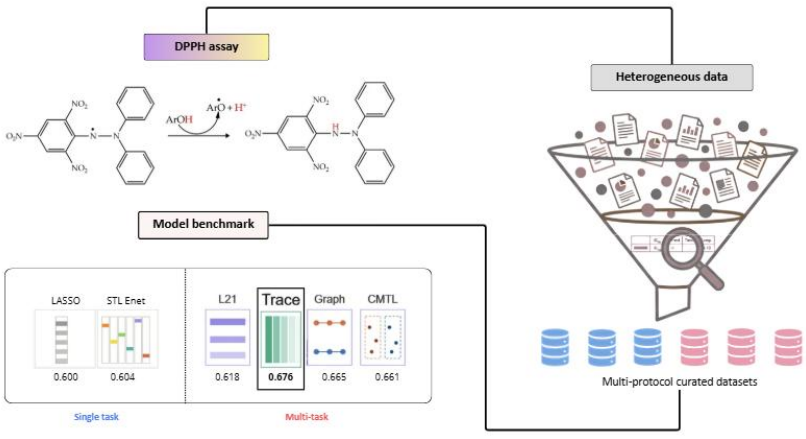

***Figure 1:* Workflow of the protocol-aware multi-task QSAR Antioxidant activity pipeline**: *heterogeneous DPPH antioxidant data are curated into protocol-aware subsets, encoded with ISIDA fragment descriptors, and modelled with single-task and multi-task learning with MTL Trace achieving the best performance (Q^2 = 0.676).*

Bibliography :

[1] Deng, W.; Chen, Y.; Sun, X.; Wang, L.; AODB: A Comprehensive Database for Antioxidants Including Small Molecules, Peptides and Proteins ; Food Chem. 418 (2023) 135992.
[2] Ruggiu, F.; Marcou, G.; Varnek, A.; Horvath, D.; ISIDA Property-Labelled Fragment Descriptors ; Mol. Inform. 29 (2010) 855–868.
[3] Cao, H.; Zhou, J.; Schwarz, E.; RMTL: An R Library for Multi-Task Learning ; Bioinformatics 35 (2019) 1797–1798.