6TH EUROCC VILNIUS WORKSHOP ON USING HPC

Abstract book

January 22, 2026

Vilnius, Lithuania

Published by Vilnius University Press

https://doi.org/10.15388/EUROCC.2026.6

eISSN 2669-0233

Vilnius University Proceedings

HIDRA – a Hierarchical DFT accuracy transformer model to Reconstruct Atomic geometry

Rokas Garbacauskas, Stepas Toliautas

Vilnius University, Faculty of Physics, Institute of Chemical Physics

E-mail: rokas.garbacauskas@ff.vu.lt

Predicting molecular properties from string representations enables high-throughput screening without expensive 3D structure generation. We present a systematic comparison of SMILES [1] and SELFIES [2] representations using transformer encoders [3] trained on 3.9 million DFT-optimized molecules from the PubChemQC database, predicting seven structural and symmetry properties.

This work addresses three questions: which string representation better encodes molecular structure for property prediction, how transformer depth should scale given limited quantum chemistry data, and whether multi-task learning improves performance through shared representations. We evaluate three architectural variants shown in Fig. 1: single-task models with one property per encoder, parallel multi-task with all prediction heads at the final layer, and hierarchical multi-task with heads at different encoder depths and cross-attention feedback between properties.

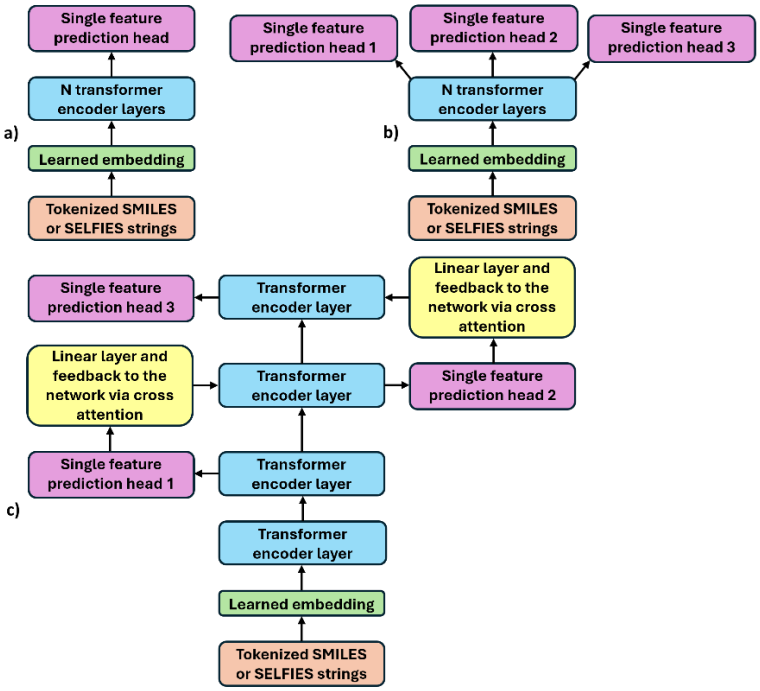

Fig. 1. (a) Single-property, (b) parallel multi-property, and (c) hierarchical multi-property transformer encoder networks.

REFERENCES

[1] D. Weininger, SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules, J. Chem. Inf. Comput. Sci. 28 (1988) 31–36.

[2] M. Krenn, F. Häse, A. Nigam, P. Friederich, A. Aspuru-Guzik, Self-referencing embedded strings (SELFIES): A 100% robust molecular string representation, Mach. Learn.: Sci. Technol. 1 (2020) 045024.

[3] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Adv. Neural Inf. Process. Syst. 30 (2017).