Title: Predicting smell by physiologically inspired deep learning models

Research axis of the 3IA: Core Elements of AI / AI for Computational Biology and Bio-Inspired AI

Supervisor:
Marco Gori, 3IA int. chair holder and associate researcher of MASAI INRIA team marco.gori@unisi.it

Laboratory and/or research group:
MASAI Team (INRIA), https://team.inria.fr/maasai ; Siena Artificial Intelligence Lab https://sailab.diism.unisi.it/
Institut de Chimie de Nice (ICN), https://icn.univ-cotedazur.fr/

Scientific description:
Our brain perceives our volatile molecular environment through olfaction by triggering the activation of chemical sensors named olfactory receptors. These hundreds of receptors are differentially activated by a virtually infinite number of molecules suggesting a highly complex combinatorial code. Can a computer learn how to smell? The objective of the project is to push the boundaries of deep learning models in the framework of chemical senses and to understanding how chemicals code for a certain type of percept. Addressing these questions is a challenge for basic research and would be beneficial for the food, flavor and fragrance industry as well as for the pharmacology industry and public health since scent detection in medical diagnosis by animals and electronic nose holds promise for the future.
In this context, the use of machine learning and deep learning models emerge as relevant methods to decode olfaction: The recent Dream Olfaction challenge has demonstrated that machine learning models, although not perfect, accurately predict odor qualities [1]. A team of researchers from Google Brain has successfully trained Graph Neural Networks (GNN were initially developed in 2008 [2]) to predict odor descriptors [3,4]. We have recently shown that the application of machine learning in the context of virtual screening opens up the possibility of enlarging insect [5,6] and mammalian odorant chemical spaces [7]. Recently, we have also shown that deep learning models are able to predict the active OR-molecule pairs [8]. The originality of our approach lies in the combination of methods used: i) to encode the topology of an odorant molecule in the form of a molecular graph (GNN), *i.e.* by considering the atoms and the bonds respectively as nodes and edges of the network ; ii) the structural information of receptors extracted from their sequences using a strategy inspired by the field of natural language processing (NLP such as BERT or its variants).
These recent results reinforce the hypothesis that we may be able to predict the olfactory quality of a given molecule by describing its chemical structure and its interactions with olfactory receptors. However, several issues remain and will be addressed in this project, such as: i) in the preceding model [8], to keep the computational cost relatively low, we take advantage of the BERT's classification token which aggregates information about the entire protein. To identify the most important part of the protein driving the odorant recognition, we aim at developing a new approach where all the amino acids will be represented explicitly. ii) the two objects (the odorant and the olfactory receptor) can be viewed as two interacting graphs, leading to a third graph that can highlight the intermolecular interactions. Exploring and explaining such complex objects are other challenges from a theoretical point of view.
This PhD project will benefit from and benefit to the local ecosystem with an increasing interest for these specific topics of the food or flavor and fragrance industry, particularly well represented in the PACA region.

References: (in blue, article from co-PI)

[1] *Predicting human olfactory perception from chemical features of odor molecules.* A. Keller, R.C. Gerkin, Y. Guan, A. Dhurandhar, G. Turu, B. Szalai, J.D. Mainland, Y. Ihara, C.W. Yu, R. Wolfinger, C. Vens, l. Schietgat, K. De Grave, R. Norel, DREAM Olfaction Prediction Consortium, G. Stolovitzky, G.A. Cecchi, L.B. Vosshall, P. Meyer. *Science* **2017**, 355, 820-826.

[2] *The graph neural network model.* F. Scarselli, **M. Gori**, A.C. Tsoi, M. Hagenbuchner, G. Monfardini, *IEEE Trans. Neural Netw. Learn. Syst* **2008**, 20(1), 61-80.

[3] *Machine Learning for Scent: Learning Generalizable Perceptual Representations of Small Molecules.* B. Sanchez-Lengeling, J.N. Wei, B.K. Lee, R.C. Gerkin, A. Aspuru-Guzik, A.B. Wiltschko. arXiv:1910.10685 [stat.ML], **2019**.

[4] *A Principal Odor Map Unifies Diverse Tasks in Human Olfactory Perception.* B.K. Lee, E.J. Mayhew, B. Sanchez-Lengeling, J.N. Wei, W.W. Qian, K. Little, M. Andres, B.B. Nguyen, T. Moloy, J.K. Parker, R.C. Gerkin, J.D. Mainland, A.B. Wiltschko. bioRxiv 2022.09.01.504602, **2022**.

[5] *Machine learning decodes chemical features to identify novel agonists of a moth odorant receptor.* G. Caballero-Vidal, C. Bouysset, H. Grunig, **S. Fiorucci**, N. Montagné, J. Golebiowski, E. Jacquin-Joly. *Sci. Rep* **2020**, 10, 1655.

[6] *Reverse chemical ecology in a moth: machine learning on odorant receptors identifies new behaviorally active agonists.* G. Caballero-Vidal, C. Bouysset, J. Gévar, H. Mbouzid, C. Nara, J. Delaroche, J. Golebiowski, N. Montagné, **S. Fiorucci**, E. Jacquin-Joly. Cell. Mol. Life Sci., **2021**, 78, 6593-6603

[7] *Agonists of G Protein-Coupled Odorant Receptors are Predicted from Chemical Features.* C. Bushdid, C.A. de March, **S. Fiorucci**, H. Matsunami, J. Golebiowski. *J. Phys. Chem. Lett.* **2018**, 9, 2235-2240.

[8] *Matching receptor to odorant with protein language and graph neural networks*. M. Hladiš, M. Lalis, S. Fiorucci, J. Topin. International Conference on Learning Representation (ICLR), **2023.** https://openreview.net/forum?id=q9VherQJd8