Télécom Paris and **Eidgenössische Technische Hochschule Zürich**

Internship project

Subject

Machine-learning for detection of mechano-transduction from cellular images

Description of the internship

Supervision

Collin Ewald (https://www.ewaldlab.com/)

Pavlo Mozharovskyi (https://perso.telecom-paristech.fr/mozharovskyi/)

Location and dates of the internship

Address 1: Télécom Paris, 19 Place Marguerite Perey, 91120 Palaiseau (place of internship)

Address 2: Schorenstrasse 16, 8603 Schwerzenbach-Zürich, Switzerland (online)

Date of the beginning of the internship: early 2022

Teams where the thesis will be written

Department Information, Data, Signal; Telecom Paris; Institut Polytechnique de Paris

Institute of Translational Medicine; Department of Health Sciences and Technology; ETH Zürich

Keywords

Machine learning, image analysis, mechano-transduction, extra-cellular matrix, https://www.rxrx.ai/rxrx1 data base.

Detailed subject

Recursion has recently released the RxRx1 dataset to kickstart a flurry of innovation in machine learning on large biological datasets to impact drug discovery and development. RxRx1 is a dataset consisting of 296 GB of 16-bit fluorescent microscopy images, the result of the same experimental design being run multiple times with the primary differences between experiments being technical noise unrelated to the underlying biology. As such, RxRx1 provides a significant sample of controlled biological variability that is prime for training models to discern classes of cell morphology, independent from experimental batch variation. It is important to note that RxRx1 has been created in a controlled manner to provide the appropriate data for discerning biological variation in its common context of changing experimental conditions.

The main goal of this internship is to confirm the mechano-transduction for a number of chemical compounds, which were identified during the previous collaboration of the internship supervisors ([4] or [5] could be an example). More precisely, based on cellular images, a connection shall be established between a set of compounds and changes in morphological cellular changes, and in particular cytoskeleton and nuclear changes.

In order to achieve this, the expected approach shall rely on large-scale neural networks twinning the biological mechanisms by predicting certain features. Further, interpretable machine leaning tools (see,

e.g., [1] or [2], and [3] for an application example) shall be employed – and if necessary improved – to obtain (visual) explanations analyzable from biological point of view.

Candidate profile

Student having master 2 (preferably research) with:

- Good knowledge of statistical learning, bases of algorithms/optimization

- Good level of programming (Python with NN frameworks and optionally R, C/C++)

- Good command of English

Application

To send on pavlo.mozharovskyi@telecom-paris.fr:

- Curriculum Vitae

- Personalized motivation letter that explains interest of the candidate in the subject (can be directly in the body of the email)

- Grade reports for recent years

- Contact of a person willing to give recommendation

Incomplete applications will not be considered.

References

[1] Lundberg, S. M., and Lee, S.-I. (2017): A unified approach to interpreting model predictions. *NeurIPS 2017 Conference*.

[2] Parekh, J., Mozharovskyi, P., and D'Alché-Buc, F. (2021): A framework to learn with interpretation. *NeurIPS 2021 Conference*.

[3] Qiu, W., Chen, H., Dincer, A. B., Lundberg, S., Kaeberlein, M., and Lee, S.-I. (2021): Interpretable machine learning prediction of all-cause mortality. DOI: https://doi.org/10.1101/2021.01.20.21250135

[4] Statzer, C. and Ewald, C. Y. (2020): The extracellular matrix phenome across species. *Matrix Biology Plus*, 8, 100039.

[5] Statzer, C., Jongsma, E., Liu, S. X., Dakhovnik, A., Wandrey, F., Mozharovskyi, P., Zülli, F., and Ewald, C. Y. (2021). Youthful and age-related matreotypes predict drugs promoting longevity. *Aging Cell*, 20, e13441.