

Deep Imaging Genetics: using deep learning latent representations of image-derived endophenotypes to strengthen associations with genetics

Research topic proposal

Jonas Richiardi, Translational Machine Learning Lab, Department of Radiology, Lausanne University Hospital and University of Lausanne

Jonas.richiardi@chuv.ch

<https://unil.ch/tml>

Imaging genetics refers to the use of image-derived phenotypes (complex traits) as targets for genetic association studies. These form quantitative measures and can thus be regression targets for genome-wide association studies, heritability studies, polygenic risk scores, and more. However, medical images such as brain or heart images obtained from magnetic resonance images form large and complicated signals (3D, 4D or 5D), with several hundreds of thousands of voxels exhibiting spatio-temporal autocorrelation patterns.

To encode and represent organ structure and function compactly from medical images, deep learning has emerged as a tool of choice and is showing tremendous promise. Neural network architecture as varied as variational autoencoders, full convolutional networks, generative adversarial networks, or vision transformers can be used to extract low-dimensional latent-space representation of medical images, typically around 100 or 200 dimensions being a reasonable choice.

In this project, the goal will be to explore the ways in which these latent representations can be used to perform genetic studies. A common approach is to use each of the dimension independently as a regression target, but this independence assumption is clearly wrong. Using large-scale cardiac imaging data and brain imaging data (e.g. UK Biobank and Philadelphia neurodevelopmental cohort), together with associated genotype data, we will explore ways in which image-derived latent spaces can be used as endophenotypes for diseases or traits of interest, and whether this provides a gain in power.