Post-doc/research engineer:

Generation of a latent representation of PET images using AI for cancer classification- ARAMIS

U1288 Inserm, Institut Curie www.lito-web.fr

Background

Assuming that the content of medical images is underexploited by visual analysis, sometimes complemented by measuring a few features on the images, the field of radiomics [Gillies-2016] proposes to automatically extract and analyze a large number of quantitative imaging features. The most commonly used approach is to calculate features derived from known mathematical equations, which reflect, for example, the distribution of gray levels or the shape of lesions but are not specifically designed to reflect biological characteristics. A second approach is to estimate these indices using deep learning (DL) techniques [Artesani-2024]. These techniques have the advantage of not requiring lesion segmentation before feature extraction, a time-consuming and operator-dependent step, and promote a more comprehensive analysis of the image content. Indeed, DL allows for exploration not only of the lesions but also of the surrounding tissue as well as other organs or tissues. Additionally, DL can identify complex "patterns," such as intra- and inter-tumoral heterogeneity or lesion dispersion, which may be poorly or insufficiently described by human experts and thus difficult to quantify using features derived from established mathematical equations.

Some unsupervised DL approaches have shown promising results in bioinformatics or histological imaging [Unger-2024] for modeling complex biological processes or stratifying cancer lesions. In particular, variational autoencoders (VAEs) can reveal complex information contained in the data through the latent space. For instance, by using the latent space representation of a VAE built from RNA sequencing data collected for 39 cancer types and 55 normal tissues, [Vibert-2021] developed a classifier to identify the primary site for cancers of unknown primary (CUP). The classifier had an accuracy of 96% in predicting the primary site in a test cohort of 48 patients.

Objectives

The goal of the project is to develop an artificial intelligence algorithm capable of automatically classifying positron emission tomography (PET) scans in oncology. By analyzing the images without prior assumptions, the algorithm will learn to extract information related to tumor lesions, such as their location, metabolic activity, number, or volume. It will also gather information from non-tumoral organs/tissues, such as the metabolism of the spleen, which may reflect the immune system's ability to influence treatment response. Additionally, the algorithm may identify entirely new and previously unreported information in the literature. The potential discovery of currently unknown phenotypic elements that provide insights into the disease and its progression is precisely the main motivation for the project.

Methods

The project will exploit a retrospective database comprising over 4,000 PET scans acquired before treatment, as part of the patient's care pathway. The images are classified into 7 distinct categories: lung cancer, breast cancer, head and neck cancer, lymphoma, melanoma, soft tissue sarcoma, or no diagnosed cancer. This database should enable an unsupervised learning algorithm to identify a representation space (latent space) that highlights the variety of PET scans, while structuring the common or differentiating information among the exams. This space will be examined to understand the elements used by the algorithm to characterize the images. Subsequently, this latent space will be leveraged to design a classifier capable of automatically identifying the organ of origin of the cancer.

Initially, the project will focus on identifying cancers with a known primary site, thus demonstrating the relevance of the latent space for solving relatively simple classification tasks. The developed classifier will undergo thorough evaluation through internal and external validation cohorts (>700 patients), using completely independent images that did not contribute to the initial design of the classifier. Once this phase is complete, we will explore the potential of the latent space to address other complex tasks, such as predicting treatment response and/or patient survival.

With the supportive research environment at Institut Curie and collaborations with oncologists, we will generalize our approach by enriching the model with new cancer types. We will then test the resulting model on PET scan images from patients with cancers of unknown primary (CUP). These cancers represent a significant challenge for oncologists, as treatment choices often depend on the organ of origin of the cancer. The positioning of patients in this latent space could also reveal links to prognosis, thereby enabling better treatment selection, leading to significant benefits in terms of survival and quality of life for cancer patients.

Scientific environment

The Translational Imaging Laboratory in Oncology (LITO) specializes in the production and analysis of medical images [Nioche-2018], particularly in metabolic and functional imaging (PET and MRI). Highly multidisciplinary, it includes around forty physicists, engineers, radiologists, nuclear medicine specialists, biologists, oncologists, PhD students, and post-doctoral researchers. Located at the Research Center of the Institut Curie and with many interactions with the Institut Curie Hospital, it benefits from a particularly favorable environment for advanced medical image analysis using the latest radiomics and artificial intelligence techniques.

Financial terms and conditions

Funding will be provided by the ARAMIS contract – Cancéropôle Ile-de-France, with the employer being the Research Center of the Institut Curie, for a 12-month fixed-term contract. The salary amount will depend on the candidate's experience. This contract may be extended.

Research valorization objectives: dissemination, publication and confidentiality, intellectual property rights, etc.

The postdoctoral researcher will have the opportunity to present the work through communications at national and international conferences in the field of medical imaging (EANM, SNMMI), oncology (ASCO, ESMO), and data science (e.g., MICCAI, ICML). They will also publish their results in international journals.

Skills required

- Deep learning
- Programming
- Statistics, machine learning
- Medical imaging, radiomics, image analysis

Contact

Fanny Orlhac : fanny.orlhac@curie.fr

References

[Artesani-2024] Empowering PET: harnessing deep learning for improved clinical insight. Artesani et al. *Eur Radiol Exp*. 2024;8:17. https://doi.org/10.1186/s41747-023-00413-1

[Gillies-2016] Radiomics: images are more than pictures, they are data. Gillies et al. *Radiology*. 2016;278:563-577. https://doi.org/10.1148/radiol.2015151169

[Nioche-2018] LIFEx: a freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity . Nioche et al. *Cancer Res*. 2018;78:4786-4789. https://doi.org/10.1158/0008-5472.CAN-18-0125

[Unger-2024] A systematic analysis of deep learning in genomics and histopathology for precision oncology. Unger et al. *BMC Med Genomics*. 2024;17:48. https://doi.org/10.1186/s12920-024-01796-9

[Vibert-2021] Identification of tissue of origin and guided therapeutic applications in cancers of unknown primary using deep learning and RNA sequencing (TransCUPtomics). Vibert et al. *J Mol Diagn*. 2021;23:1380-1392. https://doi.org/10.1016/j.jmoldx.2021.07.009