Automated MRI Quality Assessment in Clinical Data Warehouses: A Transfer Learning Approach Relying on Artefact Simulation

Position:	**Internship for a Master student**
Start:	**Spring 2025**
Duration:	**6 months**
Financial support:	**Regular internship gratification**
Location:	**1. Équipe Epione, Inria Sophia-Antipolis, 2004 Route des Lucioles, Antibes**
	2. Aramis Lab, Institut du Cerveau, 187 rue du Chevaleret, 75013 Paris

Keywords: Deep learning, Medical image analysis, Quality control, Image generation, Domain randomisation, Clinical data warehouses, Python, PyTorch

The topic

In recent years, very large clinical data warehouses (CDW) have been created containing the medical data of millions of patients. The AP-HP (Assistance Publique-Hôpitaux de Paris) CDW brings together medical imaging data from all hospitals in the Paris region. This resource provides a fantastic opportunity to train and test effective machine learning models on large clinical cohorts for a variety of neuroimaging tasks, ranging from routine checks to the study of rare pathologies.

Unlike research datasets, where acquisition protocols are well standardised, the nature and quality of CDW images is very heterogeneous. First, many MRIs are unusable because they are corrupted by various artefacts (noise, motion, poor tissue contrast…). Then, their meta-data is often inaccurate, which means that automatic data retrieval based on some criteria (sequence, resolution, pathology) is not trustable. Hence, it is crucial to detect these effects directly from the images themselves. However, visual checking by human raters is impossible due to the large volume of images. Therefore, there is a need for automatic tools that can reliably perform data quality control (QC) on CDW images and extract information from them.

The Aramis Lab has developed a deep learning model to automatically check the quality of T1-weighted MRI scans and thus select the data that can be used to train deep learning models based on the generation of synthetic artefacts [1]. Recently, we proposed a generalisation of this tool to a new sequence of MRI (FLAIR) based on domain adaptation techniques [2].

The project

This project consists of developing a new tool for automatic quality control and image-based information retrieval from MRIs scans. In line with our previous work [1], we propose here to rely on simulation strategies to alleviate the need for manual ground truths for the presence/strengths of artefacts. Moreover, instead of adapting our QC deep learning network to multiple MRI types (i.e., domains), we propose here to adopt a domain randomisation approach, best represented by SynthSeg [3]. In this paradigm, parametric generative models are used to synthesise extremely variable images of randomised aspect that are then used to train a domain-agnostic AI system.

A review of the literature will be necessary to identify (a) state-of-the-art domain randomisation techniques [4] based on deep learning tools; (b) the best ways of generating synthetic datasets containing varied artefacts. Once identified, the techniques will be implemented and tested on research data before

being applied to images from the CDW. Finally, experiments will be carried out to determine whether this method leads to a better estimate of the overall image quality than our current approach [2].

Overall, this project combines a strong image generation component with state-of-the-art deep learning architectures (diffusion models, CNNs, etc.). If successful, the methodological developments will be integrated into the open-source software ClinicaDL (https://clinicadl.readthedocs.io) designed to enable reproducible neuroimaging processing with deep learning.

A vibrant scientific, technological and clinical environment

This project will take place between the Epione team in Antibes and the Aramis Lab in Paris, two of the world-leading research teams in medical image analysis and neuroscience. You are free to choose your main work location, with possibilities of funded travels between the two.

The Epione team (around 60 people) is part of Inria (the French National Institute for Research in Digital Science and Technology), and is located in the technological cluster of Sophia-Antipolis. Its expertise is on the analysis of medical images and is at the crossroads between machine learning, physics-based modelling and geometrical statistics.

The Aramis Lab is ideally located at the heart of the Pitié-Salpêtrière hospital, downtown Paris. The ARAMIS Lab (around 40 people) is a joint team between CNRS and Inria, and is dedicated to the development of new computational approaches for the analysis of large neuroimaging and clinical data sets.

In both cases, the Epione team and the Aramis lab have a strong multidisciplinary composition, bringing together researchers in machine learning and medical doctors. This project will give you the opportunity to interact with the PhD students and engineers of the lab, as well as our medical collaborators at the Pitié-Salpêtrière hospital.

Your profile

- Master's degree or engineering degree with computer science, image analysis and/or applied mathematics profile.
- Strong interest for medical applications.
- Knowledge of deep learning.
- Knowledge in digital image processing and medical imaging.
- Good programming skills in Python.
- Good writing skills.
- Good relational and communication skills to interact with professionals from various backgrounds.

Ready to take up the challenge?

Send your CV and a cover letter explaining your motivations to Benjamin Billot (benjamin.billot@inria.fr) and Ninon Burgos (ninon.burgos@cnrs.fr).

References

[1] Loizillon, Bottani, Maire, Ströer, Dormont, Colliot, Burgos. Transfer learning from synthetic to routine clinical data for motion artefact detection in brain T1-weighted MRI. In *SPIE Medical Imaging 2023: Image Processing*, 2023. https://hal.inria.fr/hal-03831746

[2] Loizillon, Colliot, Chougar, Stroer, Jacob, Maire, Dormont, Burgos. Semi-supervised Domain Adaptation for Automatic Quality Control of FLAIR MRIs in a Clinical Data Warehouse. In DART 2023 - 5th MICCAI Workshop on Domain Adaptation and Representation Transfer, 2023. https://inria.hal.science/hal-04273997

[3] Billot, Greve, Puonti, Thielscher, Van Leemput, Fischl, Dalca, Iglesias. SynthSeg: Segmentation of brain MRI scans of any contrast and resolution without retraining. In *Medical Image Analysis*, 2023. https://arxiv.org/abs/2107.09559

[4] Chalcroft, Crinion, Price, Ashburner. Domain-Agnostic Stroke Lesion Segmentation Using Physics-Constrained Synthetic Data. In *MICCAI,* 2025. https://arxiv.org/abs/2412.03318