Title: Scalable unsupervised subtle anomaly detection from longitudinal medical imaging data: Applications to brain and cardiac imaging

Key words: Statistical and deep learning, Longitudinal analysis, Clustering, Mixed effect model, variational autoencoders, Biomarkers.

Theme / Domain / Context: The main topics of this proposal are in statistical learning and big longitudinal data.

Skills required: computer science, applied mathematics, interest for statistics applied to medical data.

Contact: carole.lartizien@creatis.insa-lyon.fr, nicolas.duchateau@creatis.insa-lyon.fr

Main location: CREATIS Lyon – MYRIAD team

Context:

Anomaly detection in medical imaging is a challenging task in contexts where abnormalities are not annotated and difficult to detect even for experts. This problem can be addressed through unsupervised anomaly detection (UAD) methods, which identify features that do not match with a reference model of normal profiles.

The goal of this project is to further improve the reliability of the UAD by leveraging additional information coming from longitudinal data. Longitudinal data [Hedeker & Gibbons 2006] consist in the repeated observations of patients over time. In practice, we expect to analyse image data at a few different times corresponding to successive visits of patients. Their analysis informs us on the progression of the disease through the evolution of abnormalities, both in size, numbers, or locations. More specifically, when applied to anomaly detection, the expectation is the confirmation of uncertain detections or the discovery of new ones, not visible at early stages.

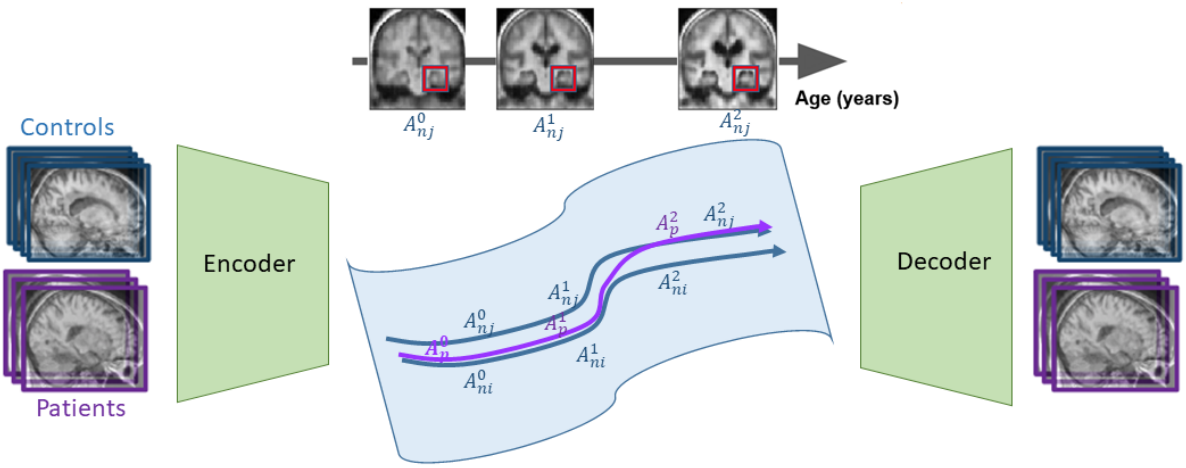
Modelling longitudinal data presents different types of challenges. First are the methodological challenges related to the design of relevant models to handle all the data and disease's characteristics in order to answer the statistical and medical questions. These modelling difficulties cannot be separated from challenges arising from data with very different modalities and time dependencies, in particular involving different acquisition time-sets and different scales of patient screening, resulting on possibly partially missing data [Couronne et al 2019].

Young et al. data [Young et al 2024] recently performed an exhaustive review of data-driven generative models of how neurodegenerative diseases evolve over time. Such models use a generative disease progression model and a set of constraints informed by human insight to infer a data-driven disease time axis and the shape of biomarker trajectories along it.

Within this framework, Sauty et al [Sauty et al 2022] recently investigated a way to model such longitudinal effects directly in the medical images by training a linear mixed effect model in the latent representation space of a longitudinal variational autoencoder. This design enables to combine the robustness of mixed-effects modelling of clinical biomarkers progression with missing data and, for any timepoint, with that of autoencoders to both learn efficient and compact representation of 3D images and reconstruct the image from the latent variable. This

model was shown to successfully model 3D T1w MRI normal brains and disease progression in Alzheimer patients.

In the same line, Puglisi et al. data [Puglisi et al 2024] also recently tackled the issue of progression modelling on medical images by introducing a novel spatiotemporal model that combines a latent diffusion model (LDM) with a ControlNet to generate individualized brain MRIs conditioned on subject-specific data. Similarly to Sauty et al, this model was shown to successfully model healthy and Alzheimer patients' brains.



Graphical abstract of the proposed methodological framework that consists first in modeling patient and normal aging trajectories in a latent low dimensional representation space and then in detecting pathological trajectories (purple curve) in this latent space.

Targeted applications:

This project will evaluate the relevance of such methods for brain and cardiac imaging applications.

For brain imaging, anomalies will consist of markers of neurodegenerative diseases, mainly Parkinson and Alzheimer diseases to be identified from magnetic resonance (MR) brain scans. In both cases, we will focus on newly diagnosed patients, where the detection task is all the more challenging as abnormalities may be subtle and hardly visible in structural magnetic resonance (MR) brain scans.

For cardiac imaging, we target the detection of myocardial shape and/or deformation abnormalities during a stress protocol (e.g. during effort) from 2D echocardiographic sequences. This may be critical for trained athletes, who have an increased risk of arrhythmias, for which geometrical and functional remodeling may be the underlying substrate [Gabrielli et al. 2016]

Directions of research:

- Review the state-of-the art in the domain of deep generative progression models, e.g. based on the review by Young et al.
- Select and implement some promising models of the literature and try to replicate reported performance on T1w MRI of the public ADNI database.
- Evaluate the performance of unsupervised anomaly detection: learn a model of normal aging evolution on control subject population, infer normal progression of ADNI patient, and derive anomaly maps by comparing pseudo-normative and patient MR exam at the same time point.

- Compare with standard UAD based on reconstruction error or support estimation of the normative distribution [Pinon et al. 2023]
- Transfer to the study of shape and functional remodelling on 2D echocardiographic data.

Skills and working environment

The selected candidate will be co-supervised by C. Lartizien (CREATIS CNRS Senior Researcher) and N. Duchateau (CREATIS Associate Professor)
The candidate should have a background either in machine learning and/or deep learning or image processing, as well as good programming skills. He/she will have access to computing resources (CREATIS and/or CNRS supercomputer) as well as to public brain datasets (ADNI) and to a private cardiac databases. The candidate will benefit from the stimulating research environment of the MYRIAD team working in the field of deep machine learning for medical image analysis.

Application
Interested applicants are required to send a cover letter, CV and any other relevant documents (reference letter, recent transcripts of marks, ...) to: carole.lartizien@creatis.insa-lyon.fr and Nicolas.duchateau@creatis.insa-lyon.fr

References:

Couronne R, Vidailhet M, Corvol JC, Lehericy S, Durrleman S. Learning disease progression models with longitudinal data and missing values. Proc. ISBI 2019 - International Symposium on Biomedical Imaging, Venice, Italy, April 2019.

Gabrielli L, Bijnens B, Bambrila C, Duchateau N, Marin J, Sitges-Serra I, Mont L, Brugada J, Sitges M. Differential atrial performance at rest and during exercise in highly trained athletes: a potential trigger for developing atrial dysfunction? Scandinavian Journal of Medicine and Science in Sports 2016;26:1444-54.

Hedeker D & Gibbons RD. Longitudinal data analysis. John Wiley & Sons, Inc, New Jersey, 2006.

Sauty B, Durrleman S. Progression models for imaging data with Longitudinal Variational Auto Encoders. Proc. MICCAI 2022, Singapore. p.3-13.

Pinon N, Oudoumanessah G, Trombetta R, Dojat M, Forbes F, Lartizien C. Brain subtle anomaly detection based on auto-encoders latent space analysis : Application to de novo parkinson patients, 2023. Proc. ISBI 2023.

Puglisi L, Alexander DC, Ravì D, Enhancing spatiotemporal disease progression models via latent diffusion and prior knowledge, 2024. arXiv:2405.03328 [cs.CV]. [Online]. Available:https://arxiv.org/abs/2405.03328

Young AL, Oxtoby NP, Garbarino S, Fox NC, Barkhof F, Schott JM, Alexander DC, Data-driven modelling of neurodegenerative disease progression: Thinking outside the blackbox, Nature Reviews Neuroscience, vol. 25, no. 2, pp. 111–130, 2024