Title: Scalable unsupervised subtle anomaly detection from longitudinal MR imaging data: Application to Parkinson's disease

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: florence.forbes@inria.fr, carole.lartizien@creatis.insa-lyon.fr, Michel.Dojat@inserm.fr

Main location: CREATIS Lyon – MYRIAD team, INRIA Grenoble, Statify team https://team.inria.fr/statify/ , GIN https://neurosciences.univ-grenoble-alpes.fr/fr

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. In the context of Parkinson's disease and newly diagnosed patients, the detection task is all the more challenging as abnormalities may be subtle and hardly visible in structural MR brain scans. Some preliminary work [Oudoumanessah et al 2023] has shown that structural abnormalities could be detected from MR image data in a way that is consistent with the disease progression, as accounted by the Hoehn and Yahr scale [Hoehn & Yahr 1998].

The goal of this project is to further improve the reliability of the detection 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 a disease evolves 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 MR 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 both to learn efficient and compact representation of 3D images and reconstruct the image from the latent variable. This model was shown to successfully model based on 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 tackle the issue of progression modelling on medical images by introducing a novel spatio-temporal 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.

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 ADNI database.
- Evaluate 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 (Baur et al), or support estimation of the normative distribution (Pinon et al)
-Transfer to the study of Parkinson disease progression based on the PPMI and a national cohort (FairPark).
-Train a model on ADNI (and/or PPMI) patients, perform a cluster analysis of the latent trajectory curves, to derive potential subtypes of disease progression. Use this clustering to analyse progression at the individual level and compare with unsupervised anomaly detection.

As a first direction of research, we propose to consider the modalities used in our previous work [Oudoumanessah et al 2023] and investigate the extension of the model and inference technique therein to multiple time data. A first idea would be to use analysis and results at previous times to inform analysis at subsequent times using a Bayesian approach as a way to incorporate information from one time to another.
In particular, we would like to adapt the model of [Sauty et al 2022] to images decomposed in smaller patches. As each patch/location in the brain could have a different time evolution, the mixed effect model may need to be extend to a mixture of mixed effect models to account for a potential heterogeneity in patch evolutions.

As a second direction of research, we will focus on accounting for possibly missing time sampling point, considering that the sample size of patients having performed all required analysis at regular time intervals, is often quite small. This task will aim at reporting on the uncertainties associated to the individual prediction in this case. The performances, strengths and weaknesses of two approaches will be compared. The first one will consist in making Bayesian predictions from the model already developed. The second will consist in exploring a Bayesian Deep learning approach [Kendal & Gal, 2017].

Data and computer environment:
The selected candidate will have access to data from the ADNI and PPMI database [Marek, 2018] and the NS-PARK cohort. Computer infrastructures available at Statify an d GriCAD will be used for algorithms development.

Skills and working environment

The selected candidate will be supervised by Florence Forbes and in constant collaboration with a PhD student, Geoffroy Oudoumanessah, working on this project. He/she will also benefit from the expertise of Michel Dojat from Grenoble Neurosciences Institute and Carole Lartizien from CREATIS Lyon.

References:

A. L. Young, N. P. Oxtoby, S. Garbarino, N. C. Fox, F. Barkhof, J. M. Schott, and D. C. Alexander,“Data-driven modelling of neurodegenerative disease progression: Thinking outside the blackbox”,Nature Reviews Neuroscience, vol. 25, no. 2, pp. 111–130, Feb. 2024,issn: 1471-0048.doi:10.1038/s41583-023-00779-6. [Online]. Available:https://doi.org/10.1038/s41583-023-00779-6

Raphael Couronne, Marie Vidailhet, Jean-Christophe Corvol, Stephane Lehericy, and Stanley Durleman. Learning disease progression models with longitudinal data and missing values. In ISBI 2019 - International Symposium on Biomedical Imaging, Venice, Italy, April 2019.

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

Hoehn, M. and Yahr, M. D. Parkinsonism: onset, progression, and mortality, Neurology 1998.

Kendall, A and Gal, Y. What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? NeurIPS 2017.

Oudoumanessah G, Lartizien C, Dojat M, Forbes F, Frugal unsupervised detection of subtle abnormalities in medical imaging, in: Greenspan H, Madabhushi A, Mousavi P, Salcudean S, James Duncan J, Syeda-Mahmood T, R T (Eds.) Miccai, Springer-Verlag AG Swizerland, Vancouver (Ca), 2023, pp. 411-421.

Kenneth Marek, Sohini Chowdhury, Andrew Siderowf, et al., “The parkinson’s progression markers initiative (ppmi) - establishing a pd biomarker cohort,” Annals of Clinical and Translational Neurology, p. 1460–1477, 2018.

Benoît Sauty, Stanley Durrleman . Progression models for imaging data with Longitudinal Variational Auto Encoders. MICCAI 2022, International Conference on Medical Image Computing and Computer Assisted Intervention, Sep 2022, Singapore, Singapore. hal-03701632

N. Pinon, G. Oudoumanessah, R. Trombetta, M. Dojat, F. Forbes, and C. Lartizien,Brainsubtle anomaly detection based on auto-encoders latent space analysis : Application to de novoparkinson patients, 2023. arXiv:2302.13593 [eess.IV]. [Online]. Available:https://arxiv.org/abs/2302.13593.[11]

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