DEVELOPMENT OF GRAPH MODELS FOR REPRESENTATION LEARNING AND TOPOLOGICAL ANALYSIS OF ALZHEIMER'S DISEASE SIGNATURES FROM WHOLE SLIDE IMAGES

R&D MASTER INTERNSHIP

Contacts: Gabriel Jimenez – gabriel.jimenez@icm-institute.org
Anuradha Kar - anuradha.kar@icm-institute.org
Daniel Racoceanu - daniel.racoceanu@sorbonne-universite.fr

Application deadline: **December 10th, 2022**
Beginning of the internship: February/March, 2023 (according to the university regulation)
Duration: **6 months** (longer duration possible condition to an agreement with the University/School)
Financial support: **regular internship gratification / StratifIAD project (BBT3-ICM)**
Locations of the internship: **Paris Brain Institute** (**Institut du Cerveau-ICM), ARAMIS team, Hôpital Pitié Salpêtrière, 47 Bd. Hôpital, 75013 Paris**

Keywords: Alzheimer Disease, TAU and Aß positive markers, Computational Pathology, Whole Slide Images, Graph theory, Graph topology, Graph Neural Networks

Context of the internship:

Alzheimer's disease (AD), the most frequent neurodegenerative disease, is defined by the misfolding and accumulation of Aß peptides and of tau proteins in the brain. The clinical presentation of the patients is more heterogeneous and different subtypes or clusters of brain lesions have been identified. The aim of this project is to understand to which extent the topography and morphology of the different peptide aggregates present in the brain can predict the diversity of symptoms observed in the patients. The research is based on an extensively human-annotated and unique set of histological images of postmortem brains from the very rare form of rapidly progressive AD patients (rpAD) and most common forms of AD. Within these images, Aβ accumulations are seen in the form of focal deposits or diffuse plaques (Figure 1a) present over the grey matter. In order to study the topography and morphology of these aggregates and better understand the morphological substratum of AD heterogeneity on large databases of patients, there is the need to develop software systems for the automatic segmentation, annotation, and quantitation of brain lesions in histopathological whole slide images. These form the core goals of the STRATIFIAD (Refining Alzheimer Disease Patients' stratification using explicable artificial intelligence in computational histopathology) project.

As part of this project, a deep learning pipeline has been developed which uses UNet and attention UNet models for detection and segmentation of tau aggregates in WSIs. With the pipeline it is currently possible to detect a majority of the tau aggregates and determine aggregate boundaries with high accuracy. Other derivable parameters include morphological features of the aggregates such as surface area, perimeter, circularity, convexity, proximity and density of the aggregates. Example results of segmentation of WSI patches using an Attention-UNet model are shown in Figure 1b. Using the segmented aggregate ROIs obtained from this pipeline, the next step in this project is to estimate topological distributions of the tau aggregates over entire whole slide regions in larger brain image databases containing hundreds of WSIs.

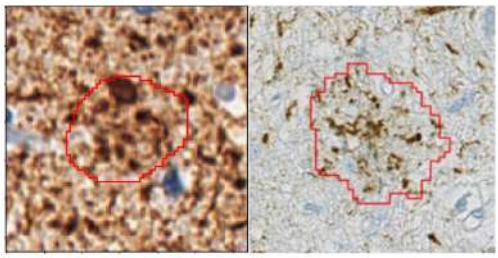

Figure 1a. Neuritic plaques

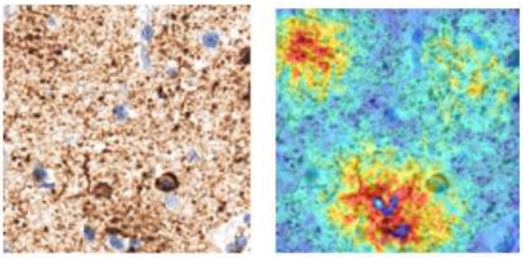

Figure 1b. Plaques (left) segmented using Att-Unet(right)

The goal of this internship is to develop a graph theory based analytical paradigm that utilizes the segmented tau aggregates from WSIs for mapping their topology. In this work, an entire whole slide image will be transformed into a tau-object based-graph, where nodes and edges of the graph will denote the detected tau aggregates and inter-aggregate relationships. This will transform a pixel based whole slide image into a knowledge- based graph characterized by the type of aggregates and their attributes. These graphs can then be used for studying aggregate topology, apply machine learning algorithms and graph neural networks on them for clustering and classification studies. Figure 2a shows an example of a graph obtained from a whole slide image patch after applying SLIC super-pixel algorithm on the image. Figure 2b shows an example of a region adjacency graph generated on a WSI (applied on the SICAPv2 database containing prostate histology whole slide images).

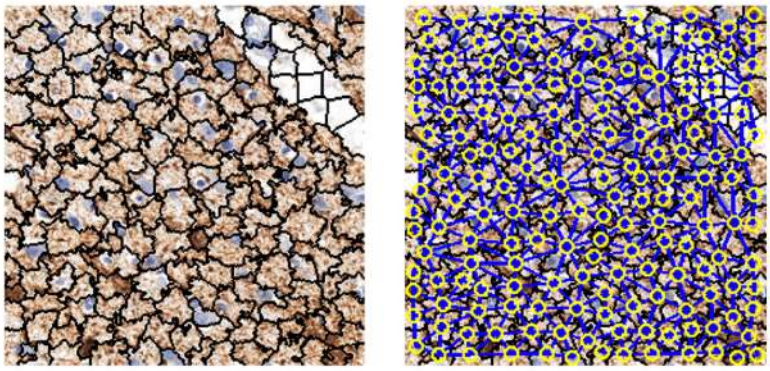

Figure 2a. (Left) A patch from a whole slide image and boundaries of the superpixels generated by the SLIC algorithm (Right)

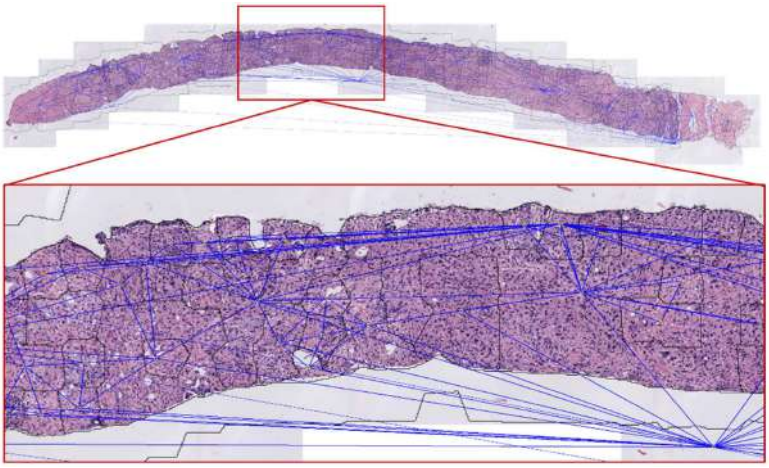

Figure 2b. Example of a region adjacency graph generated on a WSI

This internship will take place at the interface of two teams with complementary expertise:

- The Brain Development team (Prof. Daniel Racoceanu) who has a long-run expertise in digital/computational pathology, being at the origin of the first digital pathology challenges (MITOS @ ICPR 2012) and involved in the European Society for Digital and Integrative Pathology (ESDIP: https://digitalpathologysociety.org/) supporting from 15 years the European Congress on Digital Pathology. Prof. Racoceanu in particular develops image analysis and pattern recognition methods with application in computational pathology.

- The team Alzheimer's and prion diseases (Dr. Benoît Delatour and Dr. Lev Stimmer), a world-leading team for the study of the molecular and cellular mechanisms responsible for prion and prion-like diseases like AD. Together with the cellular imaging core facility of the ICM, they have developed a unique expertise in the acquisition and analysis of histologic images. Images are acquired and analyzed within the ICM.

Internship Project – Development of graph models for representation learning and topological analysis of Alzheimer's disease signatures from whole slide images

The internship aims at developing graph-theory based methods for characterizing the topology of the tau and Aß aggregates in histopathological whole slide images of the brain. The topography and morphology of the aggregates is heterogeneous with Aß accumulation taking the form of focal deposits or diffuse plaques; tau lesions forming the so-called neurofibrillary tangles but also with different morphologies in dendrites or axons. The graphs developed from WSIs will be further used for clustering and classification studies for AD patient stratification. Overall, the wsi-to-graph creation framework will help in simplifying the analysis of large datasets of voluminous WSIs and codifying their key information contents in easy to process graph models.

The candidate will work with a large dataset of whole slide images (average size: 1,000,000x60,000 pixels, ndpi format). The images have been fully annotated by neuropathologist with a degree of confidence for neuritic plaque objects.

The tasks planned for the internship include the following:

- Develop a meta-graph representation of a whole slide histopathology image to model the topological distribution of AD plaque features in the slide.
- Encode AD plaque morphological features within the graph as node attributes.
- Develop a framework wherein a WSI is represented as a fully connected meta-graph of AD features (plaques) and their attributes. Through such a framework each plaque object's appearance can be captured as a multi-attributed vertex feature within the graph. The edges could represent spatial relations between neighboring AD plaques denoted as edge features within the graph.

- **Quantitative evaluation:**
- a) Apply the framework to estimate topological meta-graph representations of whole slides from a large AD WSI dataset of 60 slides.

- b) Demonstrate the effectiveness of the proposed methodology in stratifying WSIs through graph representations of each WSI and using them in conjunction with machine learning models for clustering and classification.
- c) Apply meta-graphs to graph NN models for mapping meta graph features to different types of AD for identifying relevant patient groups.

- **Qualitative evaluation:**
- Demonstration of salient regions in a whole slide image using topological meta-graph representation. Compare these with views from pathologists. Study the correlation between the patient's characteristics and the meta-graph statistics from the WSI to understand the evolution of the patients' disease.

The candidate will work in close interaction with the histologists and pathologists to understand their needs and the precise type of information that needs to be provided. The method developed will be tested and evaluated by our experts in the cellular imaging core facility. Depending on the results, the software may be deployed routinely in the core facility and proposed as a solution to similar research organization worldwide.

Competencies (selection) required to reinforce our R&D projects:

- Graph theory, knowledge about Python libraries for building and analyzing graphs
- Knowledge on network science and graph-based topology analysis is a plus
- Image Classification, Pattern recognition, Machine and Deep Learning (CNN, U-net, GAN, MIL)

Applicant profile:

- University Master or Engineering School student (last year of study) with computer science, image analysis and/or applied mathematics profile;
- Interest for multidisciplinary projects, curiosity, learning capability and creativity are qualities.
- Interest for neuroscience research and brain diseases

We do appreciate;

- Positive spirit, communication skills and ability to work in a team, if necessary;
- Autonomy, dynamism and motivation to advance his/her own part of the project;
- Excellent methodological and hands-on computer programming skills.
- Programming language: Python. Libraries: PyTorch, Open CV, CUDA;
- Facility of understanding and manipulating mathematical models in a biological context.

Expected deliverables:

- Development of software components (data collection, expertise formalization/modelling, study of the state of the art and technology intelligence, design, test, validation);
- Proof of concept in interaction with the partners, according to the progress of the project;
- Possible publications and patents, with the prior consent of ICM and partners;
- Internship report (including methods used, results and perspectives) according to university guidelines;
- Consistent and effective user manual of the software/code developed.

Remarks:

- A careful assessment of general, methodological and programming skills will be carried out by e-conference or face-to-face (depending on availability);
- Regular (weekly) meetings will be organized, with synthetic presentations of the last advances, problems encountered, potential and proposed solution(s) as the necessary support;
- Periodic (monthly) meetings will be organized by involving a larger group of partners, according to the projects' advances and perspectives;

Concerning ICM:

The project will be carried out at the ICM (Paris Brain Institute). The research in this center is devoted to the study and treatment of neurological disorders, with a strategy, pluridisciplinary by nature, integrating cellular biology, neurophysiology, neuropathology, behavioral analysis, neuroimaging, mathematical modelling, and molecular-genetic approaches. This center is equipped with cutting edge technology and scientific expertise required for the completion of the project.