A meta-graph approach for analyzing whole slide histopathological images of human brain tissue with Alzheimer's disease biomarkers

Gabriel Jiménez, Pablo Mas, Anuradha Kar, Léa Ingrassia, Susana Boluda, B. Delatour, Lev Stimmer, Daniel Racoceanu

A meta-graph approach for analyzing whole slide histopathological images of human brain tissue with Alzheimer's disease biomarkers

G. Jimenez, P. Mas, A. Kar, L. Ingrassia, S. Boluda, B. Delatour, L. Stimmer, D. Racoceanu

Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, Inserm, CNRS, APHP, Hôpital de la Pitié Saplêtrière, Department of Neuropathology, DMU Neuroscience, Paris, France

Keywords: Alzheimer's disease, Tau Proteins, Neurofibrillary Tangles, Neuritic Plaques, Graph Topology, Morphological Features, Deep Learning, Whole Slide Images, Segmentation

1. PURPOSE OF THE STUDY

Recently, high performance deep learning models have allowed automatic and precise analysis of high-content medical images. In digital histopathology, a typical challenge lies in analyzing whole slide images (WSI) due to their large dimensions which most often requires splitting them into small patches for feeding deep learning models. This leads to loss in global tissue level information and is particularly limiting to classification or clustering of patients based on tissue characteristics.

In this study, a meta-graph approach is developed for a semantic spatial analysis of histopathological Whole Slide Images (WSI) of human brain tissue containing tau protein aggregates, one of the hallmark lesions of Alzheimer disease (AD) in brain gray matter. We propose a pipeline that extracts morphological features of tau aggregates like neuritic plaques or neurofibrillary tangles using a pre-trained U-Net model and uses these to build a graph based on Delaunay triangulation at the WSI level, in order to extract topological features from them. This pipeline is generating morphological and topological tabular data from WSI for classification and clustering patients. Further, combining locally extracted morphological features - at the neuritic plaques or neurofibrillary tangle level - with the Delaunay graph constructed at the WSI level, allows constructing a meta-graph that can be efficiently fed to graph neural network models, instead of the voluminous WSI. This pipeline is developed and tested on a dataset of 60 WSIs from various cohorts of patients having classic and rapidly advancing AD. The purpose of this pipeline is to identify novel insights into AD evolution, as well as provide a generic framework for creating knowledge rich graphs for WSI characterization and analysis.

2. METHOD

2.1 Data preparation and pipeline workflow

Histological images of brain sections were acquired using HAMAMATSU C10730-12 NanoZoomer or HAMAMATSU C13210-01 NanoZoomer. The average size of an WSI is 1,000,000 x 60,000 pixels (ndpi format) and a resolution of 10x is used in this work. The initial dataset comprises 15 WSIs which have been fully annotated by anatomopathologist yielding a set of 10,000 annotated aggregates (50,000 patches after data augmentation), including neuritic plaques and neurofibrillary tangles. The preprocessing pipeline in this study consists of 5 primary steps: patch extraction, color normalization, image segmentation with a deep learning model, morphological feature extraction and graph building. These steps are described below:

a. *Patch extraction and color normalization*
For patch extraction from WSIs, expert-made annotations are used. For each neuritic plaque (or neurofibrillary tangle), the centroid is calculated from the annotation and the patch is extracted with the aggregate of interest

Further authors information: (send correspondence to Daniel Racoceanu)
D.R.: E-mail: daniel.racoceanu@sorbonne-universite.fr, Telephone: +33 (0)6 33 77 42 67

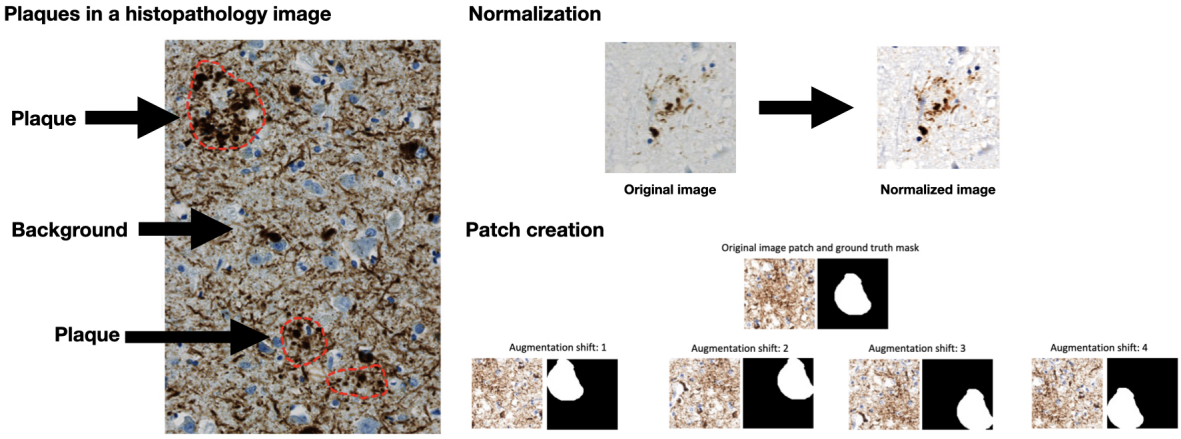

Figure 1. Identifying neuritic plaque objects in a WSI, patch creation and normalization.

at its center and specified dimensions (128x128). Finally, 4 other patches are extracted by placing the centroid in each of the four corners of the initial patch. Normalization approaches are used to reduce color fluctuations in histology images which may arise due to different staining times, solution content and pH, and the use of different digitization processes. In this work, Macenko normalization was applied using a reference image not containing any tangles or plaques but only background.

b. *UNet model architecture for segmentation*
A UNet model [9] is used for training and detection of plaques from WSIs. The model architecture consists of 3 downsampling and 3 upsampling convolutional blocks, in addition to the convolutional middle block. For the downsampling block, a leaky ReLU activation function and ReLU for the upsampling block is used. In both blocks, batch-normalization and dropout (with a probability of 0.5) were used. 8 WSIs were divided into 4 folds for the UNet training followed by 4-fold cross-testing and 3-fold cross-validation. Hyperparameters like best loss function (between Focal loss, BCEwithLogits, Dice, and BCE-Dice loss) and optimizers (SGD, Adam, RMSProp, and AdaDelta) were obtained through tuning.

c. *Morphology estimation*
The trained UNet model is used for detecting and segmenting plaques in 1 unseen WSI. Patches (128x128) were extracted from the WSI which are then used for model inference. The model generates binary masks for plaques segmented from each patch which can be used for extracting morphological parameters like centroid, area, perimeter and shape descriptors like convex area and eccentricity. These features are used to construct a topology graph for plaques in a WSI as shown in Fig 3 and Fig 4.

d. *Graph construction*
Using the detected plaques and plaque morphology features, meta-graphs may be created for the entire whole slide image. A node in the graph is the centroid of a detected plaque within a given WSI. To compute the graph from the detected plaque regions, first the centroid of each detected plaque is calculated and then the graph is computed using Delaunay triangulation.

From these graphs, topological features may be extracted which provide rich information on the distribution

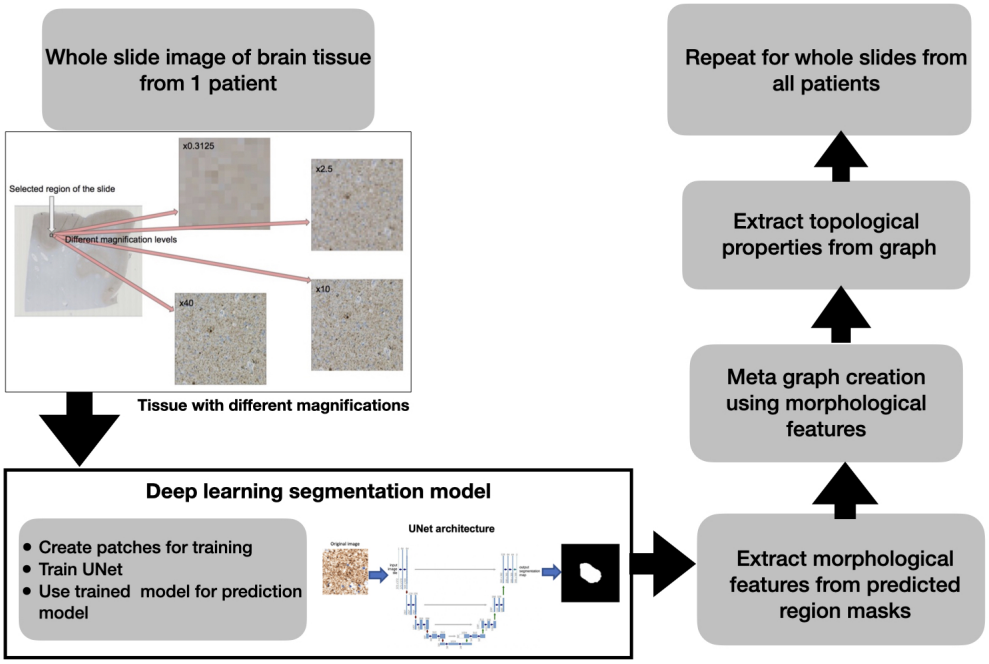

Figure 2. Using a deep learning model to detect neuritic plaques from WSIs to estimate morphological and topological features.

of tau aggregates on the entire whole slide, allowing a complete representation of tissue characteristics on a global level. Topological features such as average shortest path length, density and degree centrality may be derived from this graph. At the same time, morphological features of plaque regions may be computed and used (in conjunction with the graph) to generate a meta graph which uses morphological features as the graph attributes. The meta-graphs can then be used for clustering and classification tasks for understanding stratification of AD patients (where each graph belongs to each whole slide, i.e an individual patient).

3. RESULTS

Using our pipeline, a number of morphological parameters of plaques may be computed as shown in Fig 4. This plot shows that some features are strongly correlated.This can help us to identify and select a few of the significant features (from the whole cohort of derivable features) that can be used for further analysis and patient stratification. In the next step, a selected few morphological features such as centroid, area, eccentricity, extent and solidity are used to compute the graph as shown in Fig 4. In this way, the pipeline can be used to convert an entire whole slide image into a representative graph that can be used easily for further analysis such as clustering and classification.

4. NOVELTY OF THE STUDY

We present a novel end-to-end approach to process brain whole slide images based on creation of knowledge rich graphs that involves a Unet-based segmentation component. Using this pipeline, we are able to extract morphological features of tau plaque objects from WSIs of human brain tissue. Further, we create a topological signature using Delaunay graphs, by including - at the nodes level - these morphological features, representing the characteristics of objects distributed over the entire whole slide. Thus, using our approach, we are able to represent the required information from a computationally heavy WSI as an easily workable graph, which can be used for clusytering and classification of features within the WSI. The aim is to use these graphs in conjunction with computational models like SVM, KNN or Graph neural networks. This approach can also be extended

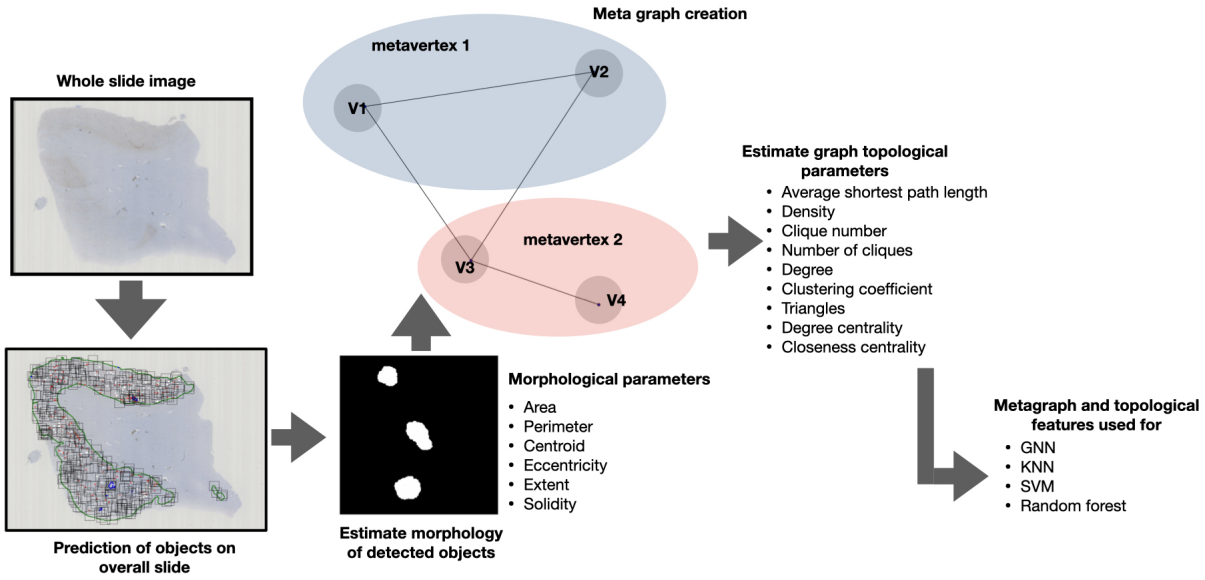

Figure 3. Complete pipeline for graph creation from WSIs and their applications.

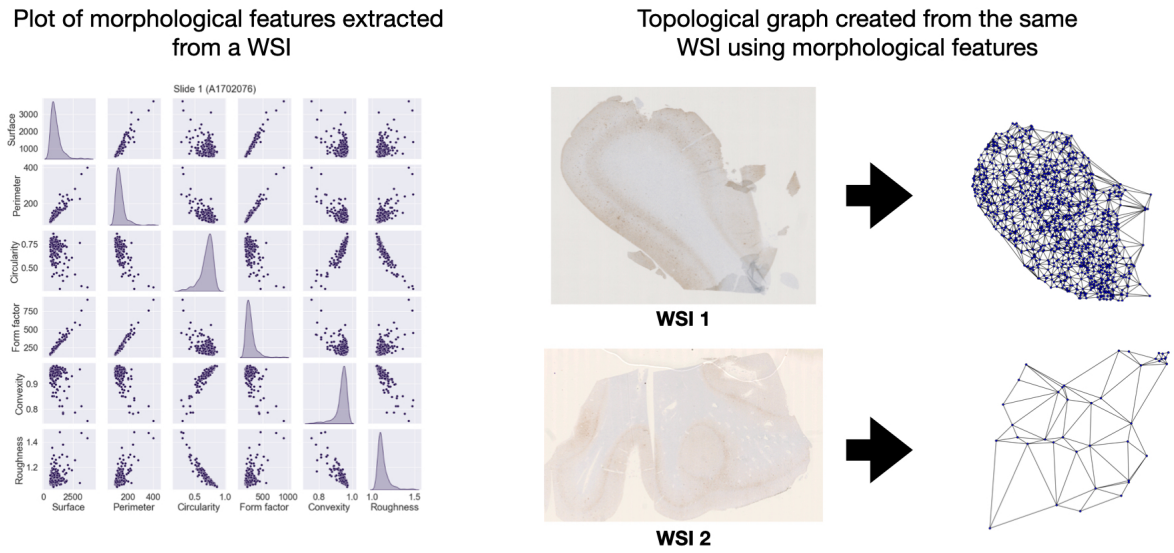

Figure 4. Morphological and topological interpretations obtained from the proposed pipeline.

to WSIs from any domain of biological research, thus providing an efficient way of computation and analysis of bulky WSIs. This approach allows considering both topological and morphological features within a sort of unique signatue of the WSI, dedicated to a precise study (in our case the study of the taupathies is Alzheimer's

disease) to work on massive WSI database and find meaningfull correlations. The pipeline is made available on line at: https://github.com/stratifIAD/graph_topology.git.

5. CONCLUSION

This study has shown that we can generate representative graphs from computationally heavy WSIs, through a sequential approach. First, we extract regions of interest within the WSI using a UNet model, then the morphological properties of these regions are estimated and used to create a representative graph for the entire whole slide. Future research will involve analysing these graphs with clustering and classification models such as SVMs, KNN or graph neural network and extending this pipelines to WSIs coming from diverse areas of digitial histopathology to create a generic framework for graph represented WSI analysis.

6. ACKNOWLEDGEMENT

This research was supported by Mr Jean-Paul Baudecroux and The Big Brain Theory Program - Paris Brain Institute (ICM). The human samples were obtained from the Neuro-CEB brain bank (https://www.neuroceb.org/en/) (BRIF Number 0033-00011), partly funded by the patients' associations ARSEP, ARSLA, "Connaître les Syndromes Cérébelleux", France-DFT, France Parkinson and by Vaincre Alzheimer Fondation, to which we express our gratitude. We are also grateful to the patients and their families.