

Probabilistic Multiple-Instance Learners for Computational Pathology

Research Internship Proposal

Biomathematics team, MICS Laboratory, CentraleSupélec, University Paris-Saclay

- ▶ **Research axis:** Artificial intelligence for oncology
- ▶ **Keywords:** Computations pathology, deep learning, attention networks, multiple-instance learning, breast cancer, biomarker detection
- ▶ **Groups:** MICS Laboratory, CentraleSupélec, Université Paris-Saclay & National Precision Medicine Cancer Center (PRISM), Gustave Roussy
- ▶ **Advisors:** Stergios Christodoulidis (stergios.christodoulidis@centralesupelec.fr), Maria Vakalopoulou (maria.vakalopoulou@centralesupelec.fr)
- ▶ **Duration:** 5 to 6 months (*with the possibility for an extension to a PhD project*)

Context

Cancer is the leading cause of death worldwide and it naturally adds a heavy socioeconomic and mortality burden on modern societies. There are multiple fronts against cancer that span from early prevention strategies, multiple screening protocols, novel patient stratification methods and innovative therapies. In recent decades, precision medicine is shifting the cancer care standards by providing personalized treatment plans for individual patients that either utilize the immune system or directly target cancer cells, with promising outcomes [1, 2]. Appropriate patient selection strategies can therefore greatly impact the treatment outcomes and the life expectancy of the patients.

Histopathological tissue assessment is the gold standard for the characterization of the cancer and the identification of typical prognostic and treatment factors and therefore treatment selection. Lately, clinical centers have invested in the digitization of such tissue slides in order to enable automatic processing as well as promoting research studies to elucidate the underlying biological processes. The resulting images are typically referred to as whole slide images (WSIs) and are of gigapixel size making their automatic processing quite challenging.

Deep learning methods have emerged as a powerful tool for processing tissue slides with examples stemming from the availability of multiple resolutions within each WSI. On the WSI level, weakly supervised systems for the processing of whole slide images for classification tasks [3] have also been proposed. These methods typically rely on multiple instance learning (MIL) schemes [4], where the vast WSIs are typically annotated with a single global label. In such cases, a whole slide image is modeled as a collection of tissue patches to be aggregated and classified. To this end, robust and explainable MIL models are of great importance and could have a tremendous impact on both clinical routines and patient lives.

Research Directions

The scope of this internship is to investigate novel MIL models based on probabilistic formulations that can enhance the performance and explainability for current architectures. Specifically, due to the computational constraints, WSIs are typically tiled and embedded using a pretrained network (e.g., ResNet) in a deterministic manner [4]. Furthermore, attention based MIL schemes aggregate the sequence of representations using attention modules [5] also in a deterministic manner. On this ground, the goal of this internship is to investigate alternatives that introduce a probabilistic formulation and replace the deterministic aspects similar to the very recent efforts that have been presented in [6,7]. This internship will build on top of these ideas, exploring ways on how such formulations can be used for uncertainty estimation and the explainability of the resulting models.

This internship will require adapting existing models under a probabilistic formulation, curating datasets, establishing experimental configuration and evaluating the performance and efficacy of the proposed methods. For the purposes of this project a number of different databases are already available both from the public domain (e.g., TCGA, TIGER, CAMELYON, etc.) as well as private in-house data from our clinical partners (e.g, Gustave Roussy). Furthermore the intern will have access to computational resources that are available in the MICS lab in order to run all the experiments (i.e., Mesoscenter Paris-Saclay).

Such formulations can be directly applicable also in different problems like remote sensing, natural images, or other medical data that are using weakly supervised schemes or MIL. As such, additional interesting experiments can be carried out while a publication could be also pursued.

Expected Work and Deliverables

The student is expected to implement novel modules in tensorflow or pytorch that can be integrated in different architectures and perform comparisons with the baselines on different databases. A comprehensive experimental configuration is expected as well as clean and well documented code that will be revisioned in the MICS's gitlab repository. Continuous feedback will be provided both for the research and the clinical aspects of the project by the supervisors as well as our clinical collaborators.

Positioning

MICS is a research laboratory of CentraleSupélec focusing on Mathematics and Computer Science. Among MICS's scientific teams the biomathematics team specializes in mathematical modeling, statistics and artificial intelligence applied in life sciences, notably medical applications, and more specifically oncology. There are a number of active collaborations between academic and industrial partners (e.g., Gustave Roussy, AP-HP, Therapanacea, University of Bern, Stony Brook University, ETS Montréal). This project will be supported by an ANR project as well as the [PRISM institute](#).

Apply

If you are interested please send your CV at:

- Stergios Christodoulidis - stergios.christodoulidis@centralesupelec.fr
- Maria Vakalopoulou - maria.vakalopoulou@centralesupelec.fr

References

1. Reck, M., Rodríguez-Abreu, D., Robinson, A.G., Hui, R., Csósz, T., Fülöp, A., Gottfried, M., Peled, N., Tafreshi, A., Cuffe, S. and O'Brien, M., 2016. Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. *New England Journal of Medicine*, 375(19), pp.1823-1833.
2. Slamon, D.J., Leyland-Jones, B., Shak, S., Fuchs, H., Paton, V., Bajamonde, A., Fleming, T., Eiermann, W., Wolter, J., Pegram, M. and Baselga, J., 2001. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *New England journal of medicine*, 344(11), pp.783-792.

3. Lu, M.Y., Williamson, D.F., Chen, T.Y., Chen, R.J., Barbieri, M. and Mahmood, F., 2021. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nature biomedical engineering*, 5(6), pp.555-570.
4. Carbonneau, M.A., Cheplygina, V., Granger, E. and Gagnon, G., 2018. Multiple instance learning: A survey of problem characteristics and applications. *Pattern Recognition*, 77, pp.329-353.
5. Ilse, M., Tomczak, J. and Welling, M., 2018, July. Attention-based deep multiple instance learning. In *International conference on machine learning* (pp. 2127-2136). PMLR.
6. Cui, Y., Liu, Z., Liu, X., Liu, X., Wang, C., Kuo, T.W., Xue, C.J. and Chan, A.B., 2023, May. Bayes-MIL: A New Probabilistic Perspective on Attention-based Multiple Instance Learning for Whole Slide Images. In *11th International Conference on Learning Representations (ICLR 2023)*.
7. Schmidt, A., Morales-Álvarez, P. and Molina, R., 2023. Probabilistic Attention Based on Gaussian Processes for Deep Multiple Instance Learning. *IEEE Transactions on Neural Networks and Learning Systems*.
8. Shao, Z., Bian, H., Chen, Y., Wang, Y., Zhang, J. and Ji, X., 2021. Transmil: Transformer based correlated multiple instance learning for whole slide image classification. *Advances in neural information processing systems*, 34, pp.2136-2147.