Pan-cancer explainable deep survival model on routine histology slides to predict benefit from immunotherapy

Supervisor : Hamid Ladjal ; **Co-supervisor** : Nazim Benzerdjeb

Contact : hamid.ladjal@univ-lyon1.fr
Laboratory : LIRIS (UMR 5205 CNRS), Université Claude Bernard Lyon 1
Master 2 : Master 2 Data Science Lyon 1
Application deadline : January 31, 2027
Starting date : September 1, 2027

Abstract : Immune checkpoint inhibitors have transformed the treatment of many cancers, yet most patients do not benefit, and PD-L1 immunohistochemistry, the main biomarker used in practice, predicts benefit poorly. We first developed a deep learning model that predicts the outcome of immunotherapy in head and neck cancer from the routine histology slide: in 410 patients from three centres (Hospices Civils de Lyon, Centre Léon Bérard, Saint-Étienne University Hospital), it reached an AUC of 0.74 on a held-out test set and separated overall survival where PD-L1 did not, showing that the information is present in the tissue. We now aim to extend this approach to multiple cancers, and have collected for this purpose a pan-cancer cohort of 2,000 patients treated with immunotherapy (head and neck, skin, lung, digestive, peritoneal, breast and gynaecological cancers). The Master 2 student will (1) build a prediction model based on the latest pathology foundation models and a Cox survival head, to predict benefit from immunotherapy more accurately; (2) predict, from the slide alone, the spatial expression of key proteins to create virtual protein maps, compared with the model's attention heatmaps and predicted risk; and (3) validate these virtual maps by immunohistochemistry on serial sections. The project offers hands-on training in computer vision, survival analysis and explainable artificial intelligence.

Project description :

Context. Immune checkpoint inhibitors (anti-PD-1/PD-L1, anti-CTLA-4) are now used in many tumour types, but only a minority of patients benefit, while all are exposed to immune toxicity and costly treatment. PD-L1 immunohistochemistry is assay-, observer- and organ-dependent and poorly predicts benefit [1]. Deep learning applied to whole-slide images (WSI) can capture the spatial organisation of the tumour-immune microenvironment that shapes the response, from the routine slide and without additional staining.

Preliminary data. With ethics approval (MR-004), we developed a first model in 410 patients with head and neck squamous cell carcinoma treated with immunotherapy between 2015 and 2024 at three centres (HCL, CLB, Saint-Étienne). Pre-treatment slides were encoded by frozen pathology foundation models [2] and pooled by attention-based multiple-instance learning (ABMIL) [3]. Trained to classify responders versus non-responders, the model reached an AUC of 0.74 on a held-

out test set (Fig. 1a) and its prediction separated overall survival (p = 0.028), whereas the PD-L1 combined positive score did not (p = 0.78) in the same patients (Fig. 1b-c). Attention maps pointed to tumour-stroma interfaces and lymphocyte-rich areas. These results show that benefit from immunotherapy can be predicted from the tissue.

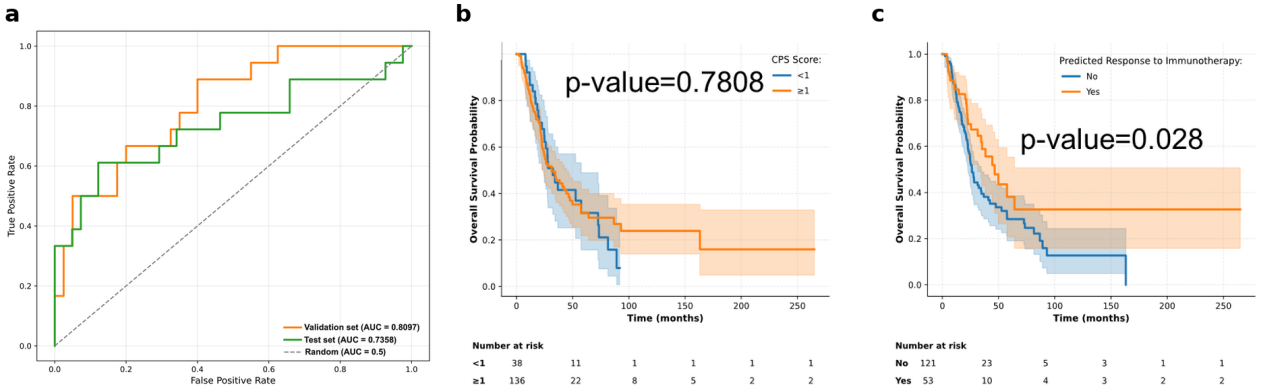


Figure 1. Preliminary results in head and neck cancer (n = 410). a, ROC curves. b, Overall survival by PD-L1 CPS. c, Overall survival by model prediction.

Pan-cancer cohort. To extend this approach, we have collected a pan-cancer cohort of 2,000 patients treated with immunotherapy, including head and neck, skin, lung, digestive, peritoneal, breast and gynaecological cancers, with digitised pre-treatment slides, PD-L1 status and clinical follow-up (response, progression-free and overall survival).

Aim 1 – Pan-cancer survival model. The student will build a prediction model based on the latest generation of pathology foundation models (e.g. UNI2, CONCH v1.5, H-optimus-1) and ABMIL, with a Cox head (DeepSurv loss) [4] trained on survival under immunotherapy. Unlike a binary responder classification, the Cox model uses all patients, their follow-up and censoring, which brings precision and stability. Performance will be assessed by Uno's C-index, time-dependent AUC and calibration, across and within cancer types, and compared with PD-L1.

Aim 2 – Virtual protein maps. The student will train models that predict, from the H&E slide alone, the spatial expression of key proteins (e.g. CD8, CD20, FOXP3, CD163, FAP, PD-L1), creating virtual protein maps. These maps will be compared with the attention heatmaps and the predicted risk to identify which cell populations and niches (T-cell infiltration or exclusion, tertiary lymphoid structures, fibroblast-rich stroma) drive the prediction, and whether they are shared across cancers.

Aim 3 – Immunohistochemical validation. These predictions will be checked by immunohistochemistry on serial sections, quantified by digital image analysis and compared region by region with the virtual protein maps, in order to validate both the predicted expression and the explanation of the model.

Environment and profile. The student will join a team of pathologists, data scientists and computer-vision researchers, with GPU servers and a curated, ready-to-use dataset. We seek a student in data science, computer vision or biomedical engineering, fluent in Python (PyTorch).

References. [1] Burtness B, Lancet 2019;394:1915. [2] Chen RJ, Nat Med 2024;30:850. [3] Ilse M, ICML 2018. [4] Katzman JL, BMC Med Res Methodol 2018;18:24.