

Deep learning informed multimodal fusion of radiology and pathology to predict outcomes in HPV-associated oropharyngeal squamous cell carcinoma



Bolin Song,^{a,n} Amaury Leroy,^{b,n} Kailin Yang,^c Tanmoy Dam,^a Xiangxue Wang,^d Himanshu Maurya,^a Tilak Pathak,^a Jonathan Lee,^e Sarah Stock,^e Xiao T. Li,^f Pingfu Fu,^g Cheng Lu,^h Paula Toro,ⁱ Deborah J. Chute,^j Shlomo Koyfman,^j Nabil F. Saba,^k Mihir R. Patel,^l and Anant Madabhushi^{a,m,*}



^aWallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, GA, USA

^bTherapancea, Paris, France

^cDepartment of Radiation Oncology, Holden Comprehensive Cancer Center, Iowa Neuroscience Institute, University of Iowa, Iowa City, IA, USA

^dInstitute of Artificial Intelligence in Medicine, School of Artificial Intelligence in Medicine, Nanjing University of Information Science and Technology, Nanjing, Jiangsu, China

^eDiagnostics Institute, Cleveland Clinic, Cleveland, OH, USA

^fDepartment of Radiology and Imaging Sciences, Emory University Hospital, Atlanta, GA, USA

^gDepartment of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH, USA

^hDepartment of Radiology, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Guangzhou, China

ⁱDepartment of Pathology, Cleveland Clinic, Cleveland, OH, USA

^jDepartment of Radiation Oncology, Taussig Cancer Center, Cleveland Clinic, Cleveland, OH, USA

^kDepartment of Hematology and Medical Oncology, Winship Cancer Institute, Atlanta, GA, USA

^lDepartment of Otolaryngology, Winship Cancer Institute, Atlanta, GA, USA

^mAtlanta Veterans Administration Medical Center, Atlanta, GA, USA

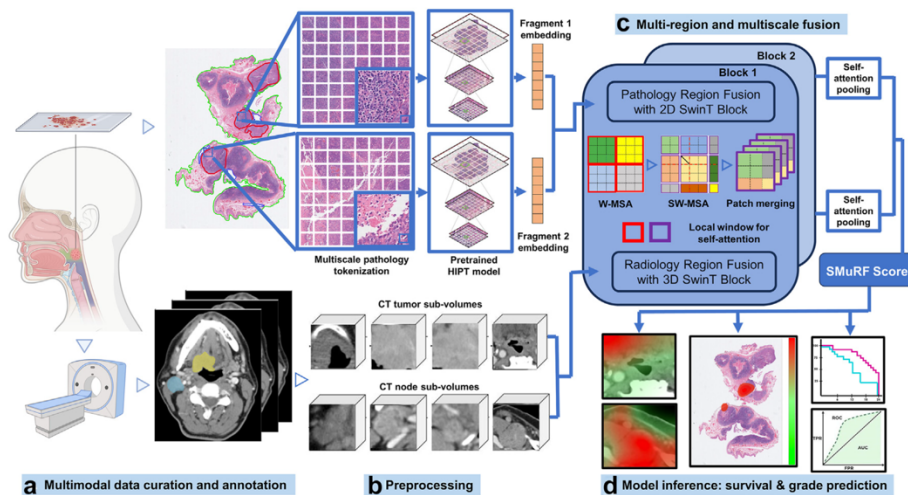


Fig. 1: Flowchart of this study: a) multimodal data curation and annotation; b) preprocessing on pathology WSI, fragment contours (green) are generated using the CLAM toolbox,³³ tumour annotations (red) are provided by pathologists. On radiology CT, primary tumour (yellow) and metastatic cervical lymph node annotations (blue) are provided by radiologists. c) multi-region and multiscale fusion with SwinT; d) model inference: survival and grade predictions. Red regions on CT and WSI indicate the prognostic relevant regions that the model is focusing on. W-MSA: window-based multi-head self-attention; SW-MSA: shifted window multi-head self-attention; SwinT: swin-transformer; HIPT: Hierarchical Image Pyramid Transformer.