Une approche pour regrouper des sujets atteints de cancers, sur la base des similarités des répartitions cellulaires au sein de leurs biopsies

Yassine El Ouahidi*, Matis Feller*
Matthieu Talagas**,***, Bastien Padeloup*,****

*IMT Atlantique, Technopole Brest-Iroise, France
prenom.nom@imt-atlantique.fr

**Univ Brest, LIEN, F-29200 Brest, France
matthieu.talagas@chu-brest.fr

***Département de Dermatologie, CHU de Brest, Brest, France

****Lab-STICC, UMR CNRS 6285, Brest F-29238, France

Résumé. Nous présentons une méthodologie interprétable pour regrouper des sujets souffrant de cancer, basée sur des attributs extraits d'images numériques de leurs biopsies. Nous proposons d'analyser des répartitions cellulaires à l'aide d'histogrammes, et de comparer les sujets via ceux-ci. Nous décrivons ici la méthodologie pour définir ces histogrammes et les exploiter à des fins de clustering. Nous illustrons notre approche sur une base de données de tissus colorés à l'hématoxyline et à l'éosine de sujets atteints d'un adénocarcinome pulmonaire de stade I, où nos résultats coïncident avec les connaissances existantes en matière d'estimation du pronostic de survie, avec une confiance statistique élevée.

1 Introduction

La médecine personnalisée en oncologie vise à améliorer le pronostic de survie d'un sujet, en prenant en compte des attributs spécifiques à son métabolisme. La plupart des travaux du domaine caractérisent les sujets sur la base d'attributs quantitatifs, tels que le nombre de lymphocytes infiltrants (TILs) (Mlecnik et al., 2011), leur densité et leur surface (Reichling et al., 2020), ou l'immunoscore (Galon et al., 2012). Cependant, il a été récemment montré que des organisations plus globales de cellules peuvent avoir un impact sur le pronostic vital (Fridman et al., 2012; Yener, 2016; Saltz et al., 2018).

Les biopsies sont numérisées via des scans à haute résolution de tissus cellulaires, préalablement colorés pour révéler les noyaux cellulaires. La coloration par hématoxyline et éosine (H&E) est la plus couramment utilisée pour des raisons financières et de simplicité d'acquisition. Des techniques plus complexes telles que l'immunohistochimie multiplex (mIHC) permettent de capturer des informations plus riches telles que la distinction des lymphocytes CD8/CD4/CD3, mais à un coût plus élevé. Une fois qu'une image haute résolution a été obtenue, une pratique courante consiste à effectuer sa segmentation et son phénotypage, afin de localiser et caractériser les cellules. Pour les images H&E, des outils utilisant des modèles