

Thèse en analyse d’images et apprentissage profond : Apport de l’intelligence artificielle dans le diagnostic du myélome multiple sur myélogramme et prédiction du risque cytogénétique

Direction de thèse Nicolas Normand LS2N¹, Marion Eveillard²

1 Université de Nantes, CNRS, LS2N, F-44000 Nantes, France

2 Pharm D, PhD, MCU-PH, CHU de Nantes (service d’hématologie biologique), UFR de Pharmacie (Département de biologie médicale), équipe 10 REMOVE-B :« Research on Molecular Vulnerabilities of tumor Escape in mature B-cell malignancies » INSERM U1232, CRCINA, CNRS ERL6001, IRS-UN

Encadrement Nicolas Normand nicolas.normand@univ-nantes.fr, Marion Eveillard marion.eveillard@chu-nantes.fr, Perrine Paul-Gilloteaux perrine.paul-gilloteaux@univ-nantes.fr

Lieu de la thèse LS2N Nantes et IRSUN Nantes

Cette thèse est proposée dans le contexte de l’appel d’offre : AIBY4 – AI made by Humans for Humans (ls2n.fr).

Sujet

Le myélome multiple (MM) est une hémopathie B mature représentant 1 % des cancers. Environ 4500 à 5000 cas sont diagnostiqués de novo en France par an. Les progrès thérapeutiques ont, depuis ces dernières années, considérablement allongé l’espérance de vie des patients, transformant cette maladie aiguë en maladie chronique. Le MM est caractérisé par la prolifération au sein de la moelle osseuse de plasmocytes malins clonaux. Cette prolifération tumorale s’accompagne :

- De la sécrétion d’une immunoglobuline monoclonale complète ou de la chaîne légère
- De lésions osseuses lytiques

Le diagnostic de cette hémopathie repose sur plusieurs critères biocliniques, dont la présence d’au moins 10 % de plasmocytes parmi l’ensemble des cellules présentes au myélogramme (frottis médullaire). La cytomorphologie est une discipline basée sur l’expertise d’un biologiste médical et elle reste à ce jour le gold standard pour porter un diagnostic d’hémopathie maligne. Elle constitue une première étape avant une éventuelle orientation des prélèvements pour analyse biologique complémentaire comme l’immunophénotypage, la cytogénétique et la biologie moléculaire. La reproductibilité inter/intra-observateur de la classification d’une cellule sur frottis médullaire n’est que de 75 à 90 % et le temps de la procédure et le volume des examens à analyser fait que le plus souvent seul un nombre limité de cellules sont observées (de 200 à 500 cellules).

Dans le cadre de ce projet, les lames sont virtualisées par un scanner de lames.

À partir des images numérisées du frottis médullaire, les objectifs de la thèse sont de développer des approches automatiques pour :

1. Identifier les zones d'intérêt de manière automatique sur les scans virtuels des lames prélevées.
2. Détecter les leucocytes parmi l'ensemble des cellules du myélogramme
3. Identifier les plasmocytes parmi les leucocytes
4. Prédire le risque cytogénétique ($t(11;14)$ ou $t(4;14)$) en fonction de la morphologie des plasmocytes (GARAND et al. 2003).

La thèse s'appuiera sur une banque de données annotées par les cytologistes du laboratoire d'hématologie du CHU de Nantes, qui sera organisée dans une base de données image et d'annotation sur un système client serveur pour faciliter les échanges intersites.

Les phases de la thèse qui pourraient être envisagées sont :

1. préparation des données pour l'apprentissage, et définition de la stratégie d'augmentation des données
2. détection automatique de la zone de lecture et automatisation associée du microscope pour l'acquisition.
3. évaluation de modèles existants et comparaison avec des approches de segmentation par traitement d'image, définition d'une architecture mixte pour la classification des cellules et la segmentation d'instances (KROMP et al. 2021 ; SegPC2021).
4. étude au niveau global de l'image (classification d'image) d'un diagnostic clinique, à comparer avec le diagnostic posé basée sur la classification automatique des cellules
5. validation de l'algorithme développé en multicentrique et comparaison à la variabilité inter-observateur humaine.

Procédure de candidature : envoyer CV, lettre de motivation et notes de master (ou équivalent) aux 3 porteurs du projet par e mail (cf contact).

Références

- GARAND, R et al. (sept. 2003). "t(11;14) and t(4;14) translocations correlated with mature lymphoplasmacytoid and immature morphology, respectively, in multiple myeloma". In : *Leukemia* 17.10, p. 2032-2035. DOI : 10.1038/sj.leu.2403091.
- KROMP, Florian et al. (2021). "Evaluation of Deep Learning architectures for complex immunofluorescence nuclear image segmentation". In : *IEEE Transactions on Medical Imaging*, p. 16. DOI : 10.1109/tmi.2021.3069558.
- Segmentation of multiple myeloma plasma cells in microscopic images* (2021). URL : <https://segpc-2021.grand-challenge.org>.