

THÈSE DE DOCTORAT DE

NANTES UNIVERSITÉ

ÉCOLE DOCTORALE N° 641
*Mathématiques et Sciences et Technologies
de l'Information et de la Communication*
Spécialité : *MASTIC - Informatique*

Par

Cédric DE ALMEIDA BRAGA

Apport de l'apprentissage profond pour le diagnostic en hématologie

Applications du sang à la moelle

Thèse présentée et soutenue à Nantes, le 20 décembre 2024

Unité de recherche : UMR6004 LS2N Laboratoire des Sciences du Numérique de Nantes

Rapporteurs avant soutenance :

Thomas WALTER Professeur de l'école desMines - IMT - Mines Paris PSL
Sarah BERTOLI Maître de Conférences Praticienne Hospitalière - Université Toulouse III

Composition du Jury :

Président :
Examineur : Nicolas COURTY Professeur des Universités - Université Bretagne Sud
Rapporteur.ices : Thomas WALTER Professeur de l'école desMines - IMT - Mines Paris PSL
Sarah BERTOLI Maître de Conférences Praticienne Hospitalière - Université Toulouse III
Dir. de thèse : Nicolas NORMAND Professeur des Universités - Polytech Nantes
Co-dir. de thèse : Marion EVEILLARD Professeur des Universités Praticienne Hospitalière - Nantes Université
Encadrante : Perrine PAUL-GILLOTEAUX Ingénieure de Recherche CNRS - SFR Bonamy

Invité(s) :

Cyrille Touzeau Professeur des Universités Praticienne Hospitalier - Nantes Université

Titre : Apport de l'apprentissage profond pour le diagnostic en hématologie

Mot clés : Apprentissage profond, Myélogramme, Hémogramme, Classification, Comptage

Résumé : Ces dernières années, l'apprentissage profond pour l'aide au diagnostic médical a connu un essor dans de nombreux domaines. En hématologie, le diagnostic de pathologies repose en grande partie sur l'observation visuelle du frottis sanguin et du myélogramme d'un patient. De nombreux travaux récents proposent d'automatiser ces processus, mais se heurtent à deux limites. D'une part, la disponibilité limitée de données annotées pour de nombreuses pathologies freinent le développement d'applications diagnostiques. D'autre part, les applications de santé nécessitent d'être interprétables et placées sous la supervision des praticiens. Au cours de cette thèse, nous avons développé plusieurs contributions visant à améliorer l'automatisation de l'aide au diagnostic hé-

matologique. Nous avons démontré la capacité des réseaux de neurones à relever une suspicion du syndrome de VEXAS sur la base des anomalies morphologiques des cellules du sang. Nous avons ensuite amélioré les modèles d'aide au diagnostic des syndromes myélodysplasiques en ajoutant la représentation latente des images du sang des patients aux paramètres sanguins utilisés traditionnellement. Enfin, nous avons consacré nos efforts à l'élaboration d'un jeu de données multi-modal pour l'étude de la moelle des patients atteints de myélome multiple, numérisé sous différentes modalités. Nous présentons sur la base de ce jeu de données les performances de l'état-de-l'art pour le comptage cellulaire et discutons les perspectives offertes pour des recherches futures.

Title: Deep learning for computer-aided diagnosis in haematology

Keywords: Deep learning, Bone marrow smear, Blood smear, Classification, Cell counting

Abstract: In recent years, deep learning for computer-aided diagnosis has seen a rise in numerous medical fields. In hematology, the diagnosis largely relies on the visual examination of a patient's blood smear and bone marrow aspirate. Many recent studies aim to automate these processes, but face two key limitations. On one hand, the limited availability of annotated data for many pathologies hinders the development of diagnostic applications. On the other hand, healthcare applications need to be interpretable and remain under the supervision of practitioners. Throughout this thesis, we developed several contributions aimed at improving the automation

of hematological diagnostic. We demonstrated the ability of neural networks to raise suspicion of VEXAS syndrome based on morphological abnormalities of blood cells. We then improved diagnostic models for myelodysplastic syndromes by adding the latent representation of patients' blood images to the traditionally used blood parameters. Finally, we dedicated our efforts to developing a multi-modal dataset for the study of bone marrow in patients with multiple myeloma, digitized under different modalities. Based on this dataset, we present the performance of state-of-the-art methods for cell counting and discuss the perspectives for future research.