

INRIA - EPITA Santé 2021

Rapport de Projet Classification de globules blancs

Paul VIALLET, Sylvain KEOSOUK, Antoine CHENG, Paviel SCHERTZER,
Keith AROUL



Inria

Février 2021

Table des matières

1	Introduction	3
2	Données	3
2.1	Dataset utilisé	3
2.2	Formes, couleurs et taille des globules rouges et globules blancs	5
3	Implémentation	6
3.1	YOLO V5	6
3.2	Cutter	7
3.3	Préprocessing avant classification	7
4	Résultats et discussion	9
5	Conclusion	9
6	Datasets et articles	10
6.1	Datasets	10
6.2	Articles de recherche	10

Table des figures

1	2 Images microscopiques de globules blancs et rouges dans le sang, zoom x2000	4
2	2 Images microscopiques de globules blancs neutrophiles, zoom x4000	4
3	Output visuel et textuel de YOLO	6
4	Fonctionnement du cutter	7
5	Image post cutter	7
6	Exemples de globules blancs préprocessés	8
7	Matrice de confusion	9
8	Rapport précision sur confiance	9

1 Introduction

Les globules rouges et blancs sont les composants principaux du sang produits par notre moelle osseuse.

Les globules rouges ou hématies permettant le transport de l'oxygène et du dioxyde de carbone dans le sang, elles sont donc indispensables pour l'oxygénation de l'organisme. Un faible nombre de ces cellules sanguines appelé anémie réduit la capacité du corps à transporter l'oxygène aux tissus.

Les globules blancs ou leucocytes sont les principales cellules responsables du système immunitaire pour combattre les infections. Un taux anormalement faible de globules blancs est appelé leucopénie.

Le comptage manuel de ces cellules est une tâche longue et pénible, et est sujette aux erreurs humaines, il faut donc utiliser une détection et segmentation automatique basée sur l'apprentissage de données.

Nous utiliserons le modèle YOLO V5 pour notre réseaux de neurones. Les modèles YOLO (You only look once) sont utilisés pour la détection d'objets avec une grande performance.

2 Données

2.1 Dataset utilisé

Pour les données, nous avons utilisées un dataset mis en ligne sur Kaggle¹, plateforme organisant des compétitions en data science.

Ce dataset contient 874 images de globules blancs, globules rouges et thrombocytes avec 4888 labels dans 3 classes différentes.

On a donc 765 images d'entraînement, 73 images de validation et 36 images de tests.

Le zoom sur les images est d'environ x2000 ce qui permet d'avoir une vue claire sur les globules avec des détails sur leur composition.

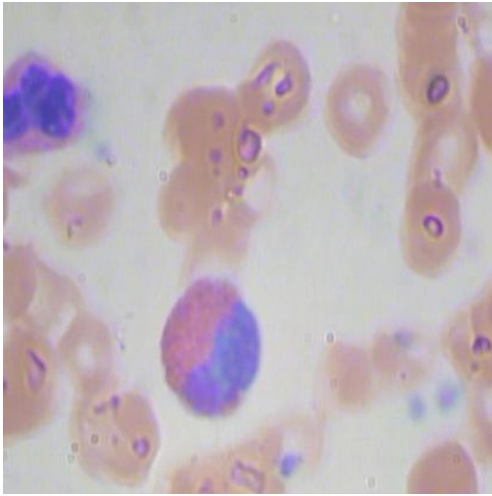
Nous avons choisi de garder les images dans leurs dimensions d'origine en 416x416.

Nous avons ensuite utilisé un deuxième dataset pour les différents globules blancs. Nous avons dans un premier temps choisi d'utiliser le dataset d'une université chinoise. Malheureusement après vérification il s'est avéré que celui-ci n'était pas fiable et que certaines cellules n'étaient pas correctement identifiées. Nous avons donc fait le choix de changer de dataset pour un plus récent ayant été clairement créé dans le but d'être utilisé pour l'entraînement d'algorithme de classification de globule blanc.²

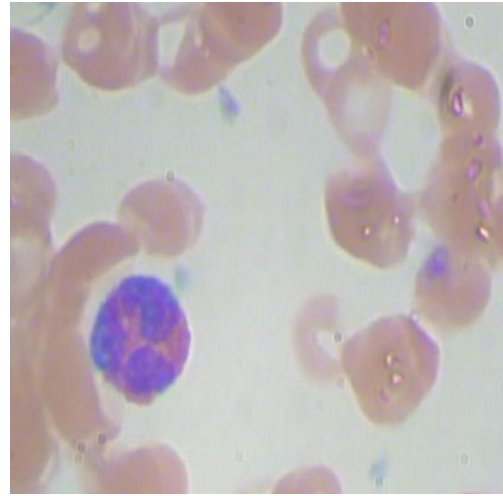
Le dataset utilisé est divisé en deux dossiers entraînement et test. On a 10176 images microscopiques de globules blancs avec 75% d'images d'entraînement et 25% validation, le dossier test contient 4430 images. Chaque dossier contient 5 sous-dossiers correspondant aux différents types de globules blancs que nous voulons prédire : neutrophiles, éosinophiles, basophiles, monocytes et lymphocytes. Chaque image a une résolution de 500x500 qui va être modifiée en 128x128 et permettra ainsi d'optimiser la mémoire GPU.

1. <https://www.kaggle.com/adhoppin/blood-cell-detection-dataset>

2. <https://raabindata.com/free-data/>



(a) Figure 1

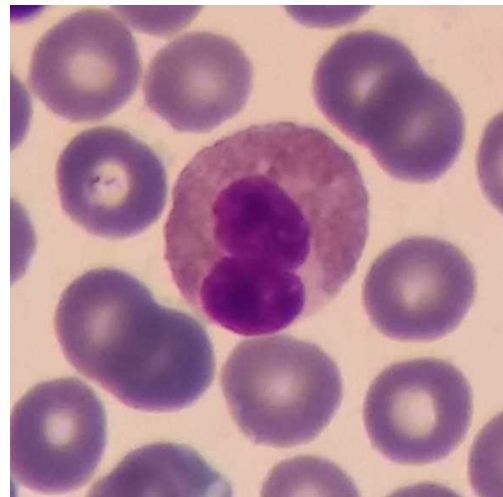


(b) Figure 2

FIGURE 1 – 2 Images microscopiques de globules blancs et rouges dans le sang, zoom x2000



(a) Figure 1



(b) Figure 2

FIGURE 2 – 2 Images microscopiques de globules blancs neutrophiles, zoom x4000

2.2 Formes, couleurs et taille des globules rouges et globules blancs

Les globules rouges ont une forme de disque biconcave d'un diamètre de 7 micromètres et ne possèdent pas de noyaux.

Les hématies sont reconnaissables par leur couleur rouge dû à la présence d'hémoglobine, protéine avec un pigment rouge, les globules rouges sont aussi appelés érythrocytes, composés des mots grecs erythros signifiant rouge et kutos signifiant cellule.

Les globules blancs possèdent un noyau et une couche intermédiaire, elles sont incolores avec une taille comprise entre 10 et 15 μm . Il n'y a qu'un seul type de globules rouges alors qu'il existe 3 grands types de globules blancs : Granulocytes, monocytes et lymphocytes. Ces granulocytes regroupent eux-même plusieurs sous-types selon leur coloration : les neutrophiles, les éosinophiles et les basophiles.

3 Implémentation

Lors de l'implémentation de notre projet, nous avons constaté qu'il nous faudrait passer les données dans deux réseaux différents pour obtenir une classification des types globules blancs à partir d'une image contenant des globules rouges, des globules blancs et du plasma, puisqu'il nous faut d'abord faire différencier globules blancs de globules rouges, puis faire apprendre les différents types de globules blancs à l'aide d'un jeu de données qui nous servirait de vérité terrain pour enfin obtenir notre classification de globules blancs.

3.1 YOLO V5

Nous avons utilisé YOLO V5 sans le jeu de données avec lequel il est présenté³, mais avons utilisé un guide d'utilisation⁴ pour l'adapter à notre jeu de données. L'algorithme présentant des résultats d'une précision et d'un taux de confiance d'environ 95%, après une centaine d'époques d'apprentissage nous n'avons pas tenté de pousser l'algorithme plus loin. Nous passons à l'algorithme des images de taille 416x416 et il nous sort un fichier texte pour chaque image avec le nom de l'image d'origine contenant une liste des plaquettes, des globules rouges et des globules blancs détectés, identifiés respectivement par 0, 1 ou 2, suivis des coordonnées du cadre dans lequel ils se trouvent.



FIGURE 3 – Output visuel et textuel de YOLO

3. <https://public.roboflow.com/object-detection/bccd>

4. <https://colab.research.google.com/drive/1gDZ2xcT0gR39tGGs-EZ6i3RTs16wmzZQ>

3.2 Cutter

Suite au passage dans l'algorithme YOLO V5, nous passons chaque image du jeu de données dans un algorithme de découpage. Cet algorithme prend en entrée les coordonnées de chaque globules blanc et l'extrait dans une nouvelle image.

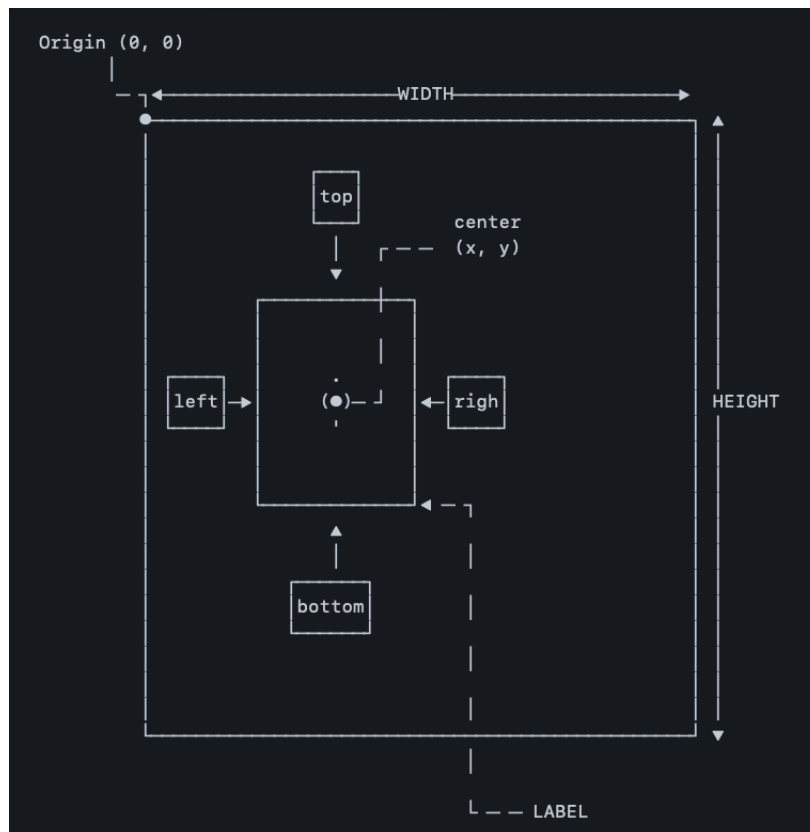


FIGURE 4 – Fonctionnement du cutter



FIGURE 5 – Image post cutter

3.3 Préprocessing avant classification

Avant d'utiliser les images dans nôtres réseaux il est nécessaire de leur appliquer un pré traitement. Différentes approches ont été utilisées et nous allons, ici en discuter. Le

problème des images utilisées en entrée du réseau, est qu'il y sont présents des globules rouges, en nombre plus ou moins important. Nous ne voulons pas que notre réseau prenne en compte ces cellules lors de la classification. Le but du pré traitement est d'isoler le noyau des cellules, qui est la partie qui permet de les différencier. Nous avons pris le temps d'inclure une explication plus détaillée du preprocessing dans le code source.

Une première approche à été de convertir l'image en niveau de gris puis de la binariser en prenant en compte un seuil. Cette approche permettait d'éliminer les globules rouges cependant il était possible d'utiliser d'autres outils afin de parvenir à un résultat plus correct. Pour ce faire, nous nous sommes intéressés aux mathématiques morphologiques. Ainsi, par le biais d'une alternance de dilations et d'érosions nous sommes parvenus à un bien meilleur résultat qui aurait pu suffire, mais qui pouvait encore être amélioré.

Nous avons donc exploré la bibliothèque de scikit afin d'analyser les différentes options présentes. Il s'est avéré que scikit utilise une méthode bien plus efficace que la nôtre et nous en avons donc utilisé les fonctions pour finaliser notre algorithme. Une limite de l'algorithme est que certains noyaux ne sont pas parfaitement isolés et cela peut poser des problèmes.

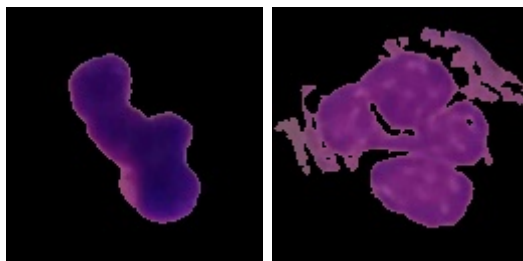


FIGURE 6 – Exemples de globules blancs préprocessés

Une fois cela fait, et pour indiquer au réseau quelles sont les classes de chaque image, nous avons fait le choix d'utiliser des DataGenerator. Le jeu de données choisi répartit chaque image de chaque type de cellule dans des dossiers distincts. Ainsi lors du chargement du fichier, et après avoir appliqué l'algorithme dessus, nous écrivons dans un fichier csv l'association du nom du fichier (son chemin plus exactement) associée à sa classe. La classe est un nombre compris entre 0 et 5.

- 0 Basophile
- 1 Eosinophil
- 2 Lymphocyte
- 3 Monocyte
- 4 Neutrophil

Le réseau en lui-même est un CNN, nous avons étudié d'autres possibilités évoquées dans plusieurs articles tels que le W-NET mais le CNN utilisé étant suffisant nous n'avons pas choisi d'étudier la question plus en profondeur. Notre réseau, après une centaine d'époques d'entraînement, affiche une 95% d'efficacité.

4 Résultats et discussion

Les résultats que nous avons eus sont présentés en deux étapes. Tout d’abord, les résultats de YOLO V5 sur notre jeu de données ont été très précis et ont répondu positivement à nos attentes. En effet, en une centaine d’époques d’entraînement, YOLO V5 présentait sur notre jeu de données une précision avoisinant les 92% pour la reconnaissance de globules blancs.

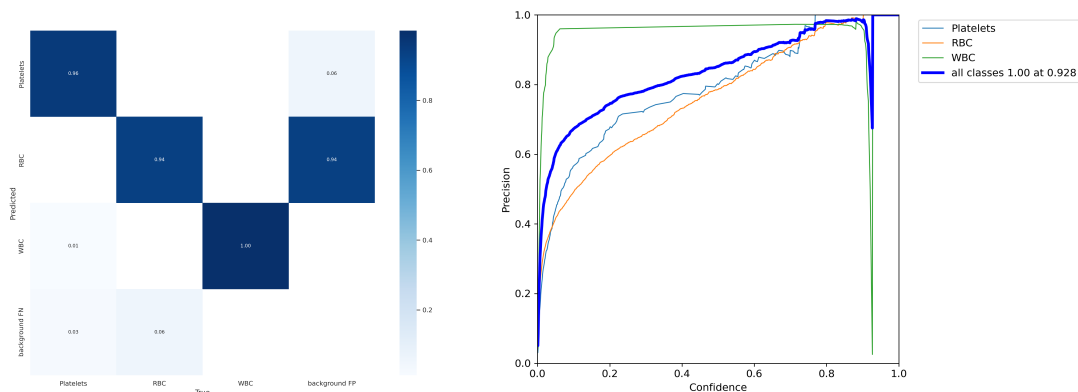


FIGURE 7 – Matrice de confusion

FIGURE 8 – Rapport précision sur confiance

Nous avons considéré que les résultats obtenus étaient suffisamment satisfaisant pour ne pas nécessiter d’approfondir ou d’améliorer l’algorithme, et nous avons donc procédé à la réalisation de l’algorithme de découpage puis du CNN pour l’étape finale du projet. Pour l’application du jeu de données au CNN final, les résultats présentés sont d’une précision de 95% en une centaine d’époques d’entraînement.

5 Conclusion

En conclusion, lors de ce projet, nous avons appris et approfondi nos connaissances en réseaux de neurones spécialisés en classification d’images. Nous avons pu également essayer de voir plus loin que les résultats d’un réseau de neurones et avons cumulé les résultats d’un réseau pour l’adapter à un réseau et un objectif plus large. En effet, dans le cadre de notre projet nous avons réussi à faire la différence entre globules rouges, globules blancs et plaquettes puis utilisé ces résultats pour faire la classification des globules blancs par type. Cela pourrait aider, avec un contexte donné, à comprendre les mécanismes qui agissent autour des différents types de globules blancs et estimer leur présence ainsi que leur quantification par section.

6 Datasets et articles

6.1 Datasets

<https://www.kaggle.com/adhoppin/blood-cell-detection-datatset>
<https://raabindata.com/free-data/>

6.2 Articles de recherche

Alam MM, Islam MT. Machine learning approach of automatic identification and counting of blood cells. *Healthc Technol Lett.* 2019 ;6(4) :103-108. Published 2019 Jul 17.

Yazan M. Alomari, Siti Norul Huda Sheikh Abdullah, Raja Zaharatul Azma, and Khai-ruddin Omar, Automatic Detection and Quantification of WBCs and RBCs Using Iterative Structured Circle Detection Algorithm, *Computational and Mathematical Methods in Medicine*

Zhengfen Jiang, Xin Liu, Zhuangzhi Yan, Wenting Gu, and Jiehui Jiang, "Improved detection performance in blood cell count by an attention-guided deep learning method," *OSA Continuum* 4, 323-333 (2021)

Tavakoli, S., Ghaffari, A., Kouzehkanan, Z.M. et al. New segmentation and feature extraction algorithm for classification of white blood cells in peripheral smear images. *Sci Rep* 11, 19428 (2021).

Drałus, Grzegorz, Damian Mazur, and Anna Czmil. 2021. "Automatic Detection and Counting of Blood Cells in Smear Images Using RetinaNet" *Entropy* 23, no. 11 : 1522.

Soumen Biswas, Dibyendu Ghoshal, Blood Cell Detection Using Thresholding Estimation Based Watershed Transformation with Sobel Filter in Frequency Domain, *Procedia Computer Science*, Volume 89, 2016, Pages 651-657

Elen, Abdullah and TURAN, M.. (2019). Classifying White Blood Cells Using Machine Learning Algorithms. *Uluslararası Muhendislik Arastirma ve Gelistirme Dergisi.* 141-152.

Miss. Madhuri G. Bhamare, Prof. D.S.Patil, 2013, Automatic Blood Cell Analysis by Using Digital Image Processing : A Preliminary Study, *INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH and TECHNOLOGY (IJERT)* Volume 02, Issue 09 (September 2013),

S. Pavithra, J. Bagyamani, White Blood Cell Analysis Using Watershed and Circular Hough Transform Technique, *International Journal of Computational Intelligence and Informatics*, Vol. 5 : No. 2, September 2015