

Sujet de stage de master :
IA générative pour aider à la classification automatique des stades de développement du *Plasmodium falciparum* dans des images de frottis sanguins

/!\ À noter que ce sujet de stage s'adresse préférentiellement à des étudiants de M2 mais, en fonction de leur profil et de leur expérience, des étudiants de M1 pourraient être acceptés.
/!\ Parmi les pré-requis explicités ci-dessous, nous attirons l'attention des potentiels candidats sur les contraintes en termes de nationalité liées à ce poste.

I. Contexte :

Le paludisme est une maladie potentiellement mortelle qui est endémique dans de nombreuses régions du monde et qui cause plus de 400 000 décès par an. Il s'agit d'une infection parasitaire (par le parasite *Plasmodium*) qui se transmet par des piqûres de moustiques femelles infectées aux humains. Plus précisément, ces parasites se multiplient en utilisant un processus de reproduction sexuée chez les moustiques, et un processus de reproduction asexuée chez les hôtes humains : d'abord dans le foie puis dans le sang, comme illustré en Figure 1. Ainsi, le paludisme chez l'homme peut être mis en évidence par des globules rouges infectés par des parasites de *Plasmodium*.

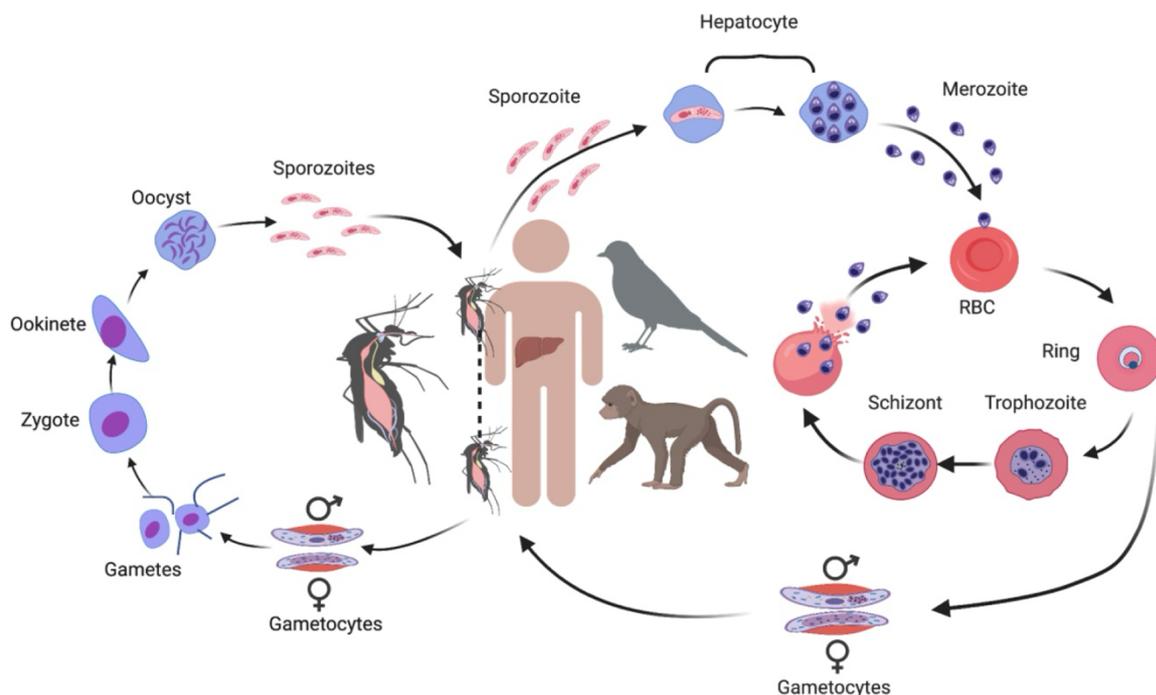


Figure 1: (Image extraite de [1]). Cycle de vie du parasite *Plasmodium*.

La référence pour le diagnostic du paludisme est l'examen microscopique manuel de frottis sanguins (ou de gouttes épaisses) avec ajout de colorant Leishman/Giemsa. Mais cette méthode nécessite des microscopistes hautement qualifiés, qui sont malheureusement souvent difficiles à trouver dans les

régions reculées les plus concernées par cette infection. Même lorsque de tels microscopistes sont disponibles, cette méthode prend beaucoup de temps et est donc très coûteuse.

C'est pourquoi, récemment, certains auteurs ont commencé à travailler sur la détection automatique du paludisme à partir de frottis sanguins ou de gouttes épaisses [2-5], en utilisant des méthodes d'apprentissage automatique et notamment de *deep learning*. La Figure 2 illustre un exemple de frottis sanguins et de goutte épaisse avec colorant.

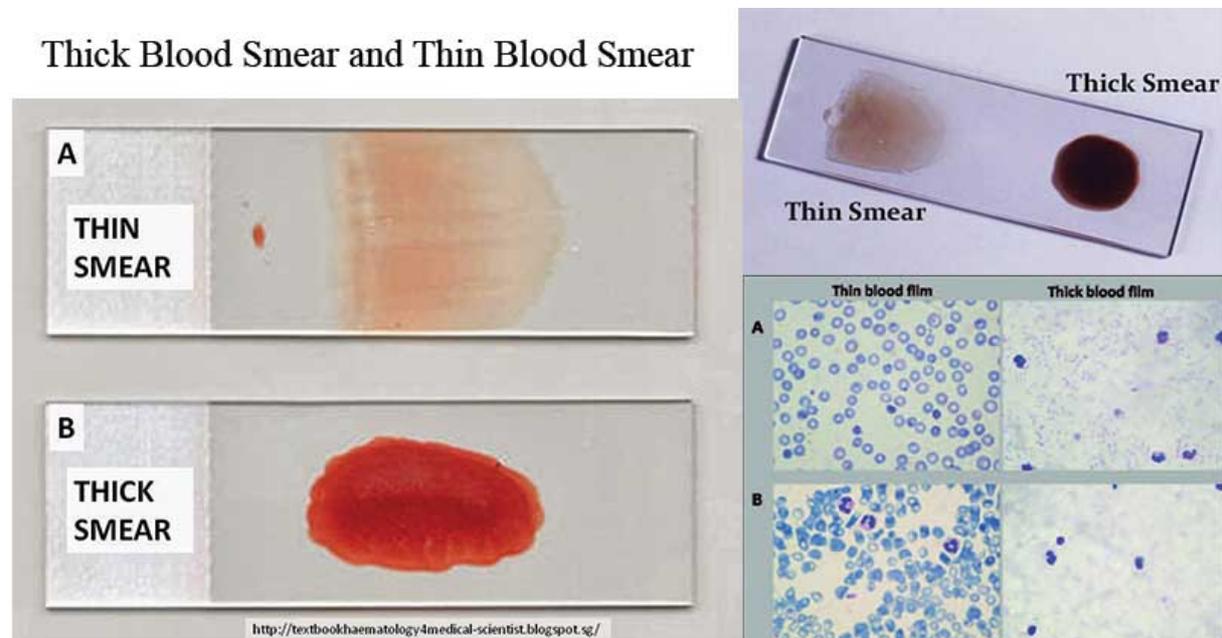


Figure 2 : (Image extraite de <https://microbenotes.com/>). Les frottis sanguins (*thin blood smear*) et les gouttes épaisses (*thick blood smears*) sont largement utilisés pour l'investigation des troubles sanguins et des parasites sanguins (par exemple le paludisme, la filariose...).

La plupart des travaux existants sur l'analyse automatique du paludisme à partir de frottis sanguins se concentrent sur la détection de la présence ou de l'absence du parasite dans les globules rouges, et éventuellement leur comptage afin de quantifier la parasitémie, comme l'illustre la Figure 3.

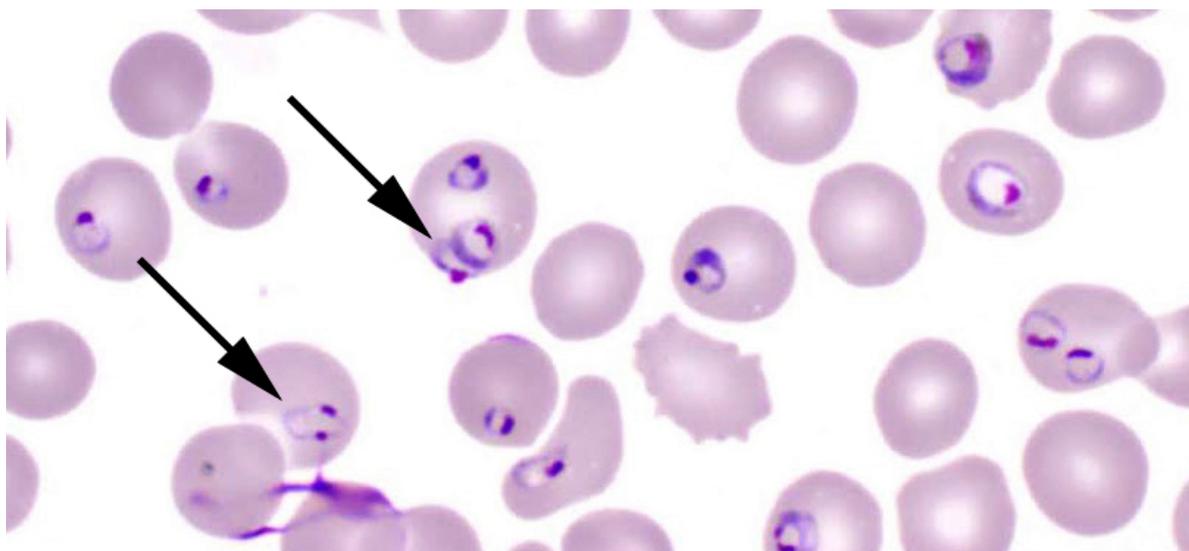


Figure 3 : (Image extraite de <https://www.pathologyoutlines.com/>) Détection d'une des espèces de *Plasmodium* (*Plasmodium falciparum*), à l'un de ses stades de développement (trophozoïte jeune a.k.a « ring »).

Cette tâche peut être effectuée avec une excellente précision par plusieurs algorithmes de traitement d'images et d'apprentissage automatique [2-5].

Jusqu'à présent, moins d'attention a été portée à la reconnaissance des stades de développement [6-9], qui est pourtant très utile à des fins de recherche biomédicale, entre autres pour :

- le suivi quotidien des cultures conduites en laboratoire, que ce soit des cultures d'isolats cliniques ou de clones de laboratoire, en conditions standard ou sous pression thérapeutique ;
- l'optimisation des expérimentations nécessitant une synchronisation des cultures ;
- l'évaluation de l'efficacité de molécules connues ou en développement pour lutter contre l'infection en bloquant le développement du parasite à un stade donné ou qui est à déterminer (trophozoïte ou schizonte par exemple).

II. Objectif de ce travail :

C'est à la reconnaissance des stades de développement du *Plasmodium* que nous nous intéresserons dans le cadre de ce stage.

Il existe cinq espèces de parasites de *Plasmodium* pour l'homme mais, parce que *Plasmodium falciparum* est l'espèce la plus dangereuse pour l'homme et cause la plupart des décès liés au paludisme dans le monde, nous nous concentrons ici sur le *Plasmodium falciparum*. La figure 4 illustre ses différents stades de développement dans le sang humain.

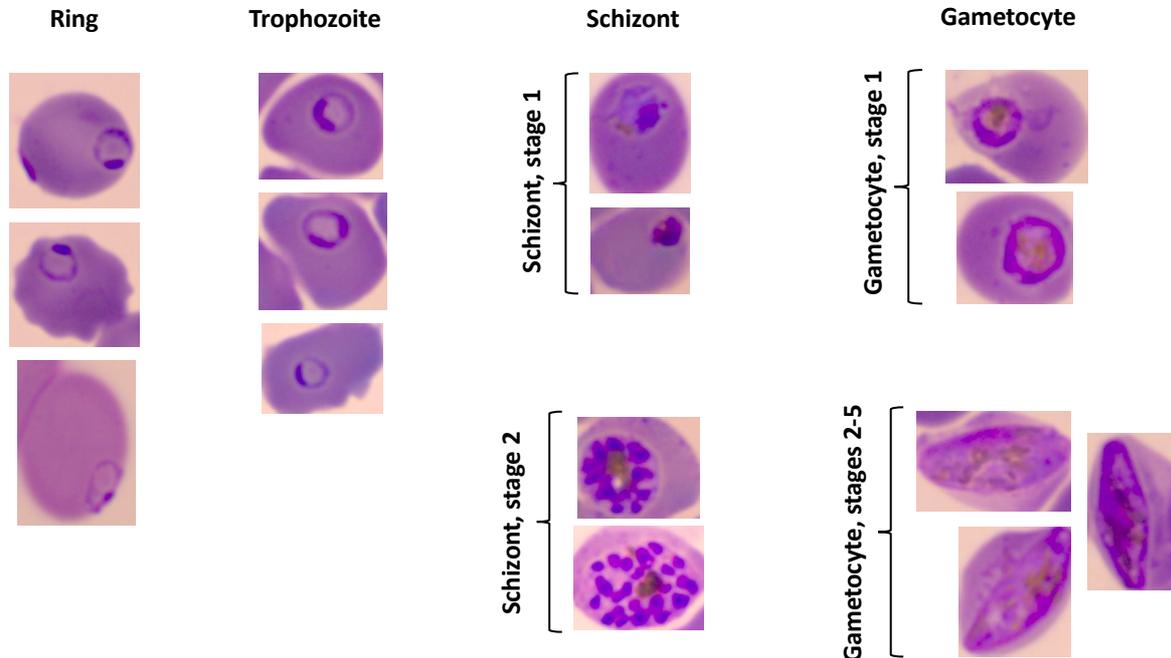


Figure 4 : Illustration des différents stades de développement du *Plasmodium falciparum*.

Comme on peut le voir en comparant la Figure 4 et la Figure 2, la classification des stades de développement (problème de classification multi-classes) est beaucoup plus complexe que le cas habituel de la détection de *Plasmodium falciparum*.

Ceci, d'autant plus que certains stades de développement sont visuellement très similaires, et que ce problème de classification souffre de classes très déséquilibrées : dans un frottis sanguin infecté, on compte typiquement entre 20 et 100 fois plus de globules rouges non infectés que de globules rouges infectés.

De plus, d'autres facteurs peuvent rendre cette tâche de classification encore plus complexe, comme par exemple :

- La présence de globules rouges multi-parasités (sachant que, par exemple, un globule rouge multi-parasité par des trophozoïtes peut ressembler à un schizonte) ;
- La présence de différents stades parasitaires dans le même globule rouge ;
- La présence de globules blancs ou d'artefacts, débris etc. qui pourraient être confondus avec des parasites ;
- La présence de parasites à l'extérieur des globules rouges (mérozoïtes).

Les principaux objectifs de ce stage sont de :

- Reprendre la revue de la littérature des approches existantes pour la reconnaissance des stades de développement du *Plasmodium falciparum* effectuée dans le cadre de ce projet, et reproduire les résultats existants obtenus par ces approches sur des bases publiques ;
- Évaluer ces approches sur notre propre base d'images ;
- Effectuer un état de l'art des approches génératives en-contexte pouvant être utilisées pour équilibrer les effectifs par classe ;
- Évaluer quantitativement et qualitativement l'impact de ces approches génératives, à la fois sur les bases d'images publiques, et sur la base développée dans le cadre de ce projet.

III. Contexte du stage

Ce stage se déroulera au Centre d'Epidémiologie et de Santé Publique des Armées (CESPA), situé à Marseille. Vos encadrants de stage seront Muriel Visani, docteure en informatique et spécialiste en apprentissage automatique appliqué à l'analyse d'images (notamment la classification), qui a récemment rejoint le CESPA, et Thierry Urruty, également spécialiste du traitement et de l'analyse d'images, qui participera à distance à votre encadrement depuis l'université de Poitiers.

Le centre CESPA compte environ 60 membres du personnel, parmi lesquels de nombreux médecins militaires ayant une solide expérience des maladies à transmission vectorielle telles que le paludisme, et plusieurs collaborations avec des centres de recherche de renommée mondiale spécialisés dans l'analyse et la lutte contre les maladies à transmission vectorielle (incluant les chercheurs de l'IRBA).

Ce stage se déroulera dans le cadre d'un projet de recherche mené conjointement entre le CESPA et l'IRBA (Institut de Recherche Biomédicale des Armées), pour lequel une base d'images de frottis sanguins est en cours de finalisation.

Pour cette raison, le stagiaire pourra être amené à passer du temps sur le site de l'IHU (La Timone), qui héberge les chercheurs de l'IRBA impliqués dans ce projet.

Les conditions de travail sont bonnes : les transports en commun permettent de se rendre facilement au CESPA et il y a une cantine (très peu chère) sur place.

La durée prévue du stage est de 4 à 6 mois, mais elle peut être ajustée en fonction des contraintes de l'étudiant.

L'allocation de stage suivra le montant légal en 2025 (environ 580 euros/mois lorsque la durée du stage dépasse 2 mois).

D'un point de vue purement administratif, il est à noter qu'il faudra que l'organisme de formation du stagiaire (université, école d'ingénieur, ou autres) accepte de signer une unique version de la convention de stage : celle proposée par le Service de Santé des Armées.

IV. Pré-requis

Nous recherchons un(e) **étudiant(e)** de Master 2 ou de Master 1, de **nationalité Française**, avec :

- Motivation dans la recherche et l'innovation
- Bonnes compétences en programmation (en particulier avec Python)
- Bonnes compétences en IA et en apprentissage automatique
 - o Une première expérience du *deep learning* serait un atout
- Très bon niveau d'anglais (parlé et écrit)
- Connaissances en traitement et analyse d'images

Tous les demandeurs devront se soumettre à une enquête d'habilitation de sécurité avant de pouvoir accéder au site du CESP.A.

V. Comment postuler ?

Merci d'envoyer dès que possible votre CV, vos relevés de notes de Master et de licence et, si vous en avez, des lettres de recommandation, à :

- muriel.visani@univ-lr.fr
- muriel.visani@def.gouv.fr
- thierry.urruty@univ-poitiers.fr

Des auditions seront organisées après une première évaluation des candidatures reçues.

VI. Références

[1] Su, X.Z et Wu, J., 2021. Zoonotic transmissions and host switches of malaria parasites. *Zoonoses (Burlington, Mass.)*, vol. 1(1), 2021, NIH Public Access.

[2] Alharbi, A.H., Lin, M., Ashwini, B., Jabarulla, M.Y. and Shah, M.A., 2022. Detection of peripheral malarial parasites in blood smears using deep learning models. *Computational Intelligence and Neuroscience*, 2022.

[3] Abdurahman, F., Fante, K.A. and Aliy, M., 2021. Malaria parasite detection in thick blood smear microscopic images using modified YOLOV3 and YOLOV4 models. *BMC bioinformatics*, 22(1), pp.1-17.

[4] Maqsood, A., Farid, M.S., Khan, M.H. and Grzegorzec, M., 2021. Deep malaria parasite detection in thin blood smear microscopic images. *Applied Sciences*, 11(5), p.2284.

[5] Mahdieh Poostchi, Kamolrat Silamut, Richard J. Maude, Stefan Jaeger, George Thoma, Image analysis and machine learning for detecting malaria, *Translational Research*, vol. 194, 2018, Pages 36-55.

[6] Q. A. Arshad, M. Ali, S.-u. Hassan, *et al.*, 2022, A dataset and benchmark for malaria life-cycle classification in thin blood smear images, *Neural Computing and Applications*, vol. 34(6), pp. 4473–4485.

[7] M. S. Davidson, C. Andradi-Brown, S. Yahiya, *et al.*, 2021, Automated detection and staging of malaria parasites from cytological smears using convolutional neural networks, *Biological imaging*, vol. 1(2).

[8] S. Li, Z. Du, X. Meng, and Y. Zhang, Multi-stage malaria parasite recognition by deep learning, 2021, *GigaScience*, vol. 10(6).

[9] F. Araujo, N. Colares, U. Carvalho, C. F. Costa Filho, and M. G. Costa, 2023, Plasmodium life cycle-stage classification on thick blood smear microscopy images using deep learning: A contribution to malaria diagnosis, 19th IEEE International Symposium on Medical Information Processing and Analysis (SIPAIM), pp. 1–4.