Master internship subject:

Automatic classification of plasmodium parasite species and stages of development from stained thin blood smears using machine learning

I. Context:

Malaria is a potentially fatal disease that is endemic in many world regions, and causes more than 400,000 deaths per year. It is a parasite infection (with Plasmodium parasite) that is transmitted from infected female mosquito bites to humans. More precisely, these parasites multiply using a sexual reproduction process in mosquitoes, and an asexual reproduction process in human hosts: first in the liver and then in the blood. Thus, malaria in humans can be evidenced by red blood cells infected with plasmodium parasites.

Nowadays, the "gold standard" for malaria diagnosis is still manual microscopic examination of Leishman/Giemsa stained thin and thick blood smear. But this method requires highly skilled microscopists, which are difficult to get in some remote areas, especially in developing countries. Even when such microscopists are available, this method is very time-consuming hence very costly.

This is why, recently, some authors started working on automatic malaria detection from stained thick or thin blood smears [1-4], using machine learning methods. An example of stained thick and thin blood smears is illustrated in Figure 1.

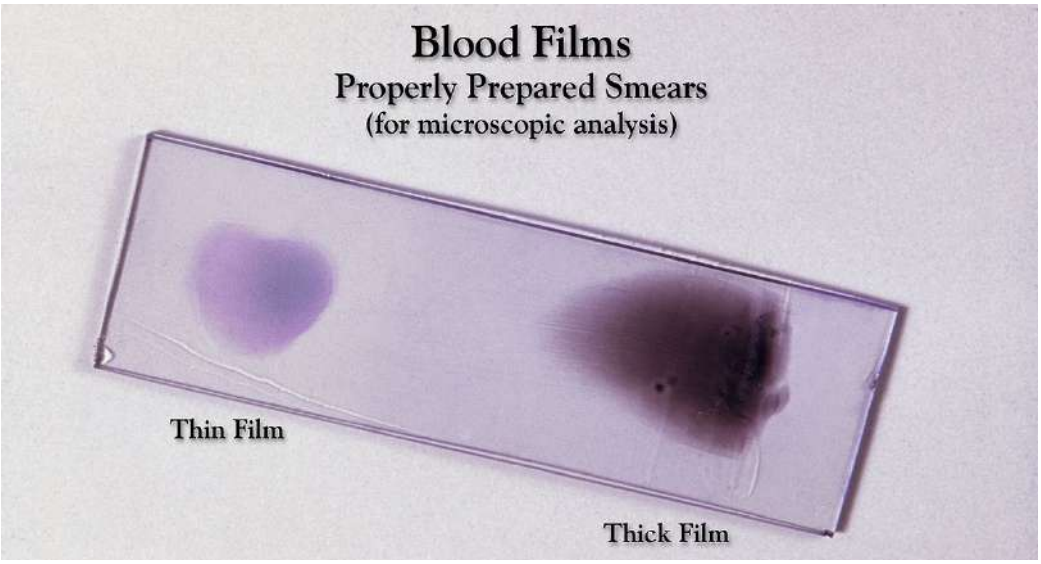

Figure 1: *(Image extracted from https://mltexpo.blogspot.com). Blood films are widely used for the investigation of blood disorders and blood parasites (e.g. malaria, filariasis…).*

Most of the existing work on automatic malaria analysis from blood smears focus on detecting the presence or absence of the parasite in red blood cells, and possibly counting them so as to quantify parasitemia, as illustrated in Figure 2.

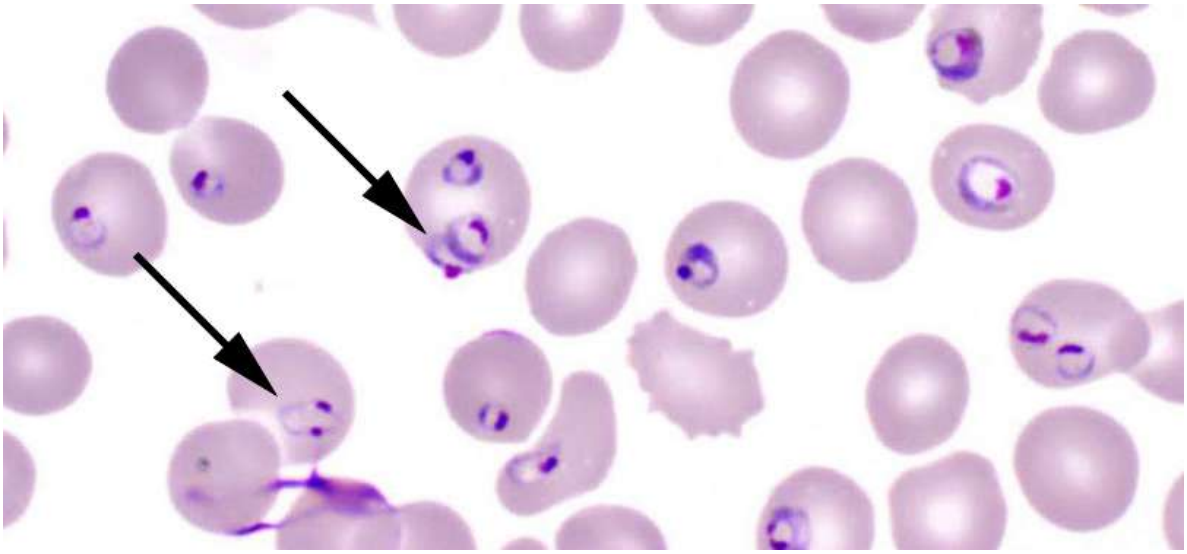

Figure 2: *(Image extracted from https://www.pathologyoutlines.com/) Detection of one of the species of plasmodium (plasmodium falciparum), at one of its stages of development (ring).*

This task can be performed with excellent accuracy by several image processing and machine learning algorithms [1-4].

There are five species of plasmodium parasites for humans: Plasmodium falciparum, Plasmodium vivax, Plasmodium malariae, Plasmodium ovale and Plasmodium knowlesi. Because Plasmodium falciparum is the most dangerous specie of the parasite for humans and causes most malaria-related deaths worldwide, most authors focus on the detection of P. falciparum.

However, medical doctors are also interested in knowing which species of plasmodium are infecting a given host, and their stage of development in the human blood. Indeed, these parasites go through a development cycle with different stages that give the parasites a different visual appearance. In chronological order, these stages are: ring, trophozoite, schizont, and gametocyte. Figure 3 illustrates the five species of parasites, and their different stages of development in the human blood.

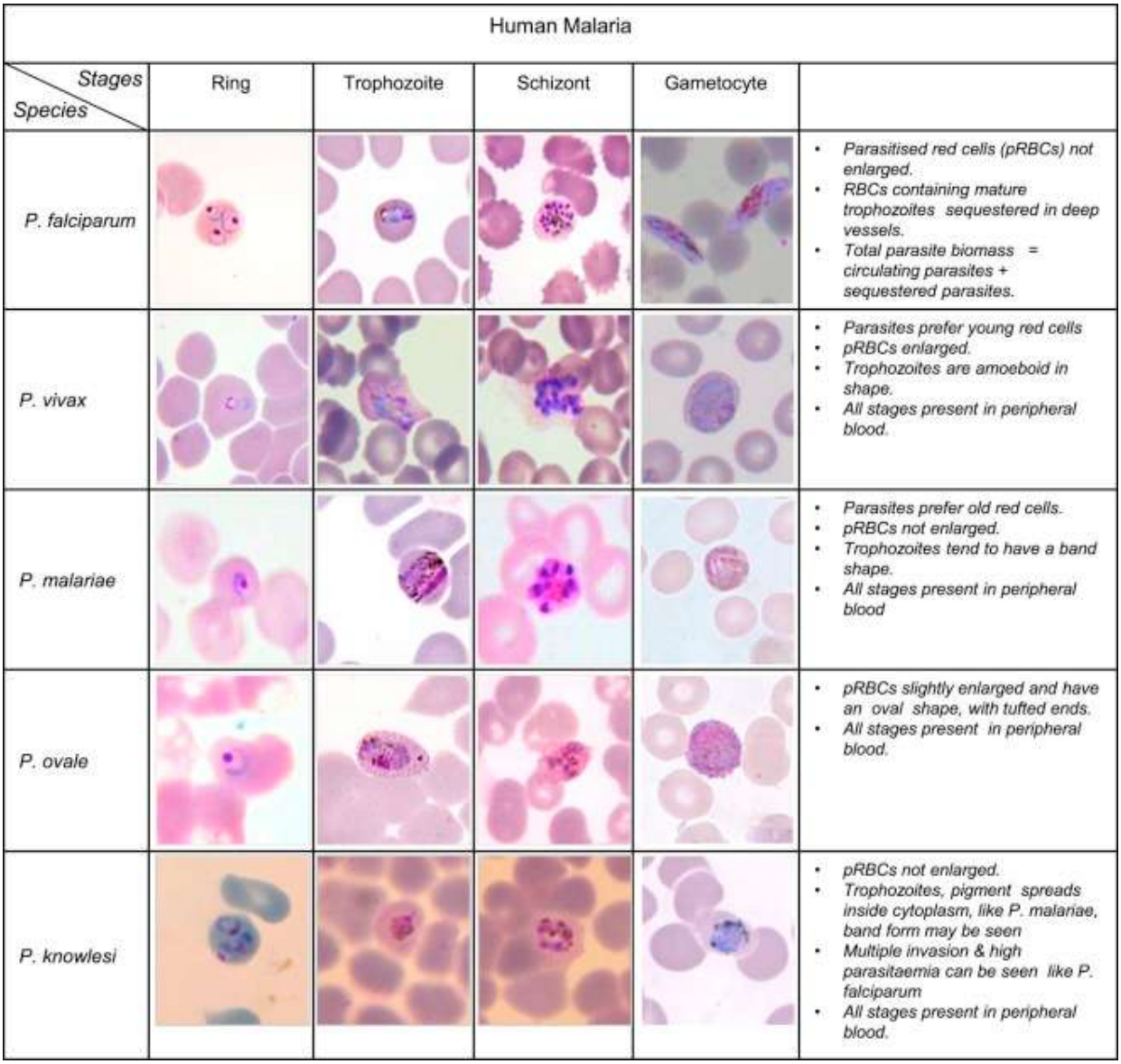

Figure 3: *Figure extracted from [4]. In rows: the five species of parasites; in columns: their different stages of development, from thin blood smear.*

II. Objective of this work:

The objective of this work is to go beyond most existing works which only focus on the plasmodium parasite detection (characterization of presence/absence), so as to be able to classify the different types of parasites and their stages of development.
As can be seen from Figure 3, the classification of different species of plasmodium with their stages of development (20-class classification problem) is much more complex than the usual case of plasmodium falciparum detection (binary classification problem).

Especially that other factors can make this classification task even more complex, some of them being illustrated in Figure 4:

- The presence of different parasite stages in the same thin blood slide image (or even of different parasite species in the case of co-infection);
- The presence of white blood cells which could be confused with parasites;
- The presence of parasites outside cells;
- Staining noise, etc.

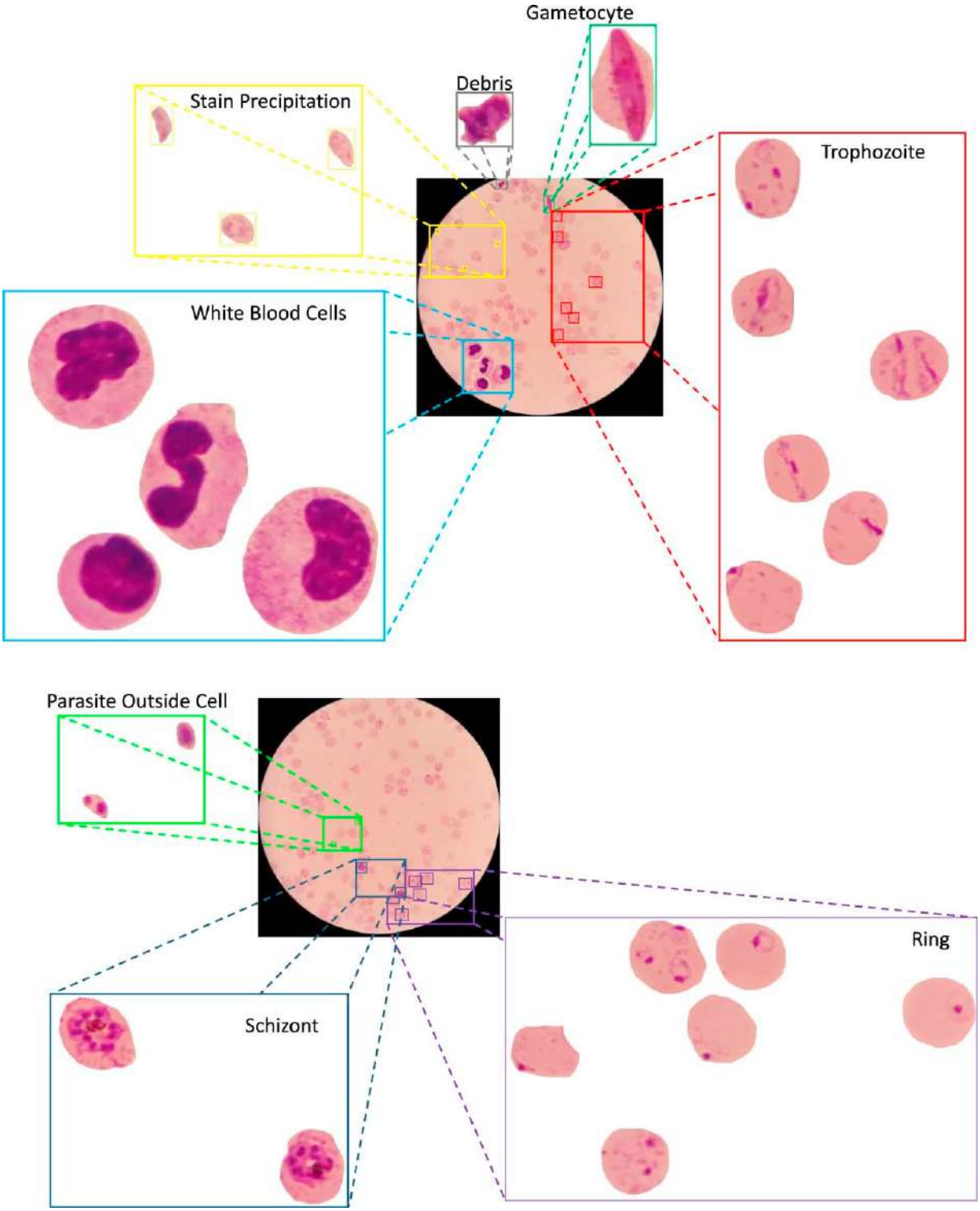

Figure 4: *Image extracted from [4]. Illustration of some difficulties for the 20-class classification (plasmodium species + stage of development).*

More precisely, the main objectives of this internship work are to:

- Make a literature review of existing approaches for plasmodium detection and recognition from stained blood smears.
- Reproduce the existing results obtained by the most recent methods for plasmodium parasite detection.
- Propose, implement and evaluate new methods for the classification of parasite species and their development stages. Both "traditional" machine learning approaches and deep learning approaches will be investigated.
- Publish the results obtained

III. Context of the internship

This internship will take place in the French Military Center for Epidemiology and Public Health (CESPA), located in Marseille. Your internship advisors will be Muriel Visani, PhD in computer science and specialist in machine learning applied to image analysis (especially classification), who recently joined CESPA, and Thierry Urruty, also a specialist of image processing and analysis, who will participate remotely in your supervision from Poitiers University.
The CESPA center has about 60 staff members, among which many military medical doctors with solid experience about vector-borne diseases such as malaria, and several collaborations with world-famous research centers specialized in the analysis and fight against vector-borne diseases.
The working conditions in the center are pleasant: public transportation to reach the center is good and there is a canteen on-site.

The expected duration of the internship is 4-6 months, but it can be adjusted depending on the student's constraints.
The internship stipend will follow the legal amount in 2023 (about 570 euros / month when the duration of the internship exceeds 2 months).

IV. Pre-requisites

We are looking for a **Master 2** student with:

- Motivation in research and innovation
- Good skills in programming (especially using Python)
- Good level of English (spoken and written)
- Knowledge in AI and machine learning
- Knowledge in image processing and analysis
- Some knowledge in deep learning would be a bonus

All applicants will have to go through a security clearance investigation before they can access the CESPA site.

V. How to apply ?

Please send as soon as possible your CV and motivation letter to:

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

Auditions will be organized after a first evaluation of the applications received.

VI. References

[1] 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*.

[2] 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.

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

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