

PFE - Document de spécification

Classification de tumeurs sur images histologiques



Pierre Minier
pminier@bordeaux-inp.fr

Laboratoire eVida
Université de Deusto - Bilbao

Début du stage: 1 février 2023



Résumé

Ce document de «spécification» présente le cadre de travail, le sujet et les avancements de mon stage au laboratoire eVida après six semaines. L'objectif est d'utiliser des réseaux de neurones convolutifs (CNNs) pour identifier la gravité de tumeurs affectant les neuroblastes, des cellules souches du système nerveux. Bien que populaires pour des problématiques de classification, les CNNs n'ont pas encore été mis en oeuvre pour l'étude automatique d'image histologique de neuroblastes.

Mots clefs : CNN, classification, histologie, neuroblaste.

Abstract

This document presents the framework, the subject and the progress of my internship in the eVida laboratory after six weeks. The aim is to use convolutional neural networks (CNNs) to determine the severity of tumours affecting neuroblasts, stem cells of the nervous system. Although popular for classification issues, CNNs have not yet been implemented for automatic diagnostic of histological images of neuroblasts.

Keywords: CNN, classification, histology, neuroblast.

1. Cadre de travail

1.1. Le laboratoire eVida

Le laboratoire eVida est un groupe de recherche situé dans la province de Biscaye, au Nord-Ouest de l'Espagne, qui est rattaché à l'Université de Deusto. Les travaux menés ont pour objectif d'améliorer la santé et le bien-être des personnes.

Dans son ensemble, l'équipe se concentre en particulier sur les trois domaines suivants : le traitement de signaux EEG (Électroencéphalogramme) notamment pour l'étude de l'épilepsie, le développement de jeux pédagogiques (serious games) pour des enfants autistes et des personnes en réadaptation physique, ainsi que le traitement numérique des images pour aider à poser un diagnostic médical plus précis.

L'effectif du laboratoire eVida est d'environ 15 personnes, mais cela varie en fonction du nombre d'étudiants et de doctorants présents. Actuellement, deux personnes travaillent sur des problématiques

de traitement d'image : moi-même ainsi qu'un autre étudiant-stagiaire de l'ENSEIRB.

1.2. Travail attendu

Dans ce stage, il est demandé d'entraîner un modèle de Deep Learning, basé sur de la convolution 2D, afin de classer des images de tumeurs en fonction du niveau de gravité de celles-ci. Une base de donnée encore jamais utilisée auparavant est fournie. Il est donc nécessaire de réaliser les pré-traitements et les régularisations requises.

1.3. Environnement de travail

Des réunions sont tenues avec le maître de stage ainsi que d'autres personnes du laboratoire externes au projet. Elles ont lieu une fois toutes les deux semaines, et fixées selon les disponibilités de chacun. L'objectif est de discuter de l'avancement du projet et de justifier des choix devant des personnes ayant

plus de recul, et pouvant apporter un regard neuf sur une situation incertaine.

La base de données est accessible depuis Google Drive. Elle est mise à jour de manière hebdomadaire par un chercheur ayant travaillé au laboratoire eVida, et maintenant basé dans un hôpital pour enfant au Mexique où une grande partie des images sont obtenues et labélisées.

Les développements se réalisent exclusivement sous Python avec principalement le module PyTorch pour architecturer et entraîner les réseaux de neurones. Les entraînements sont menés localement sur CPU, ou sur les GPU de Google Colab lorsque ces ressources sont disponibles. L'utilisation de Google Colab impose le format du notebook comme interface de développement.

2. Sujet de stage

2.1. Présentation

Le neuroblastome est un type de cancer se développant à partir de cellules immatures du système nerveux sympathique, et souvent diagnostiqué chez les enfants de moins de 5 ans [1]. Ce cancer est relativement rare, mais il peut être agressif et se propager rapidement à d'autres parties du corps par métastases. Le diagnostic est généralement réalisé par imagerie histologique (Figure 1), permettant d'évaluer la composition des cellules tumorales, et notamment de déterminer si elles sont matures ou immatures. Les cellules dites immatures sont plus susceptibles de se propager à d'autres parties du corps, et donc d'aggraver la maladie.

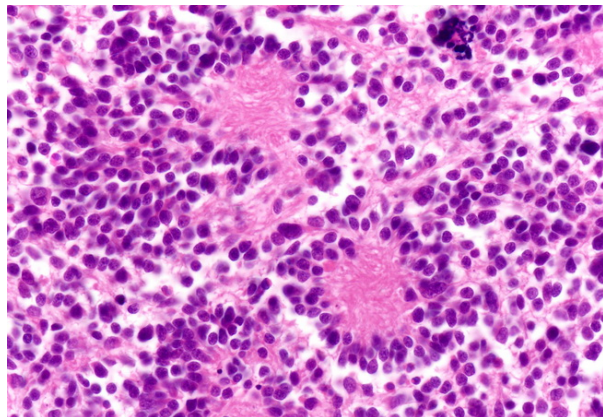


Figure 1. Exemple d'image histologique, montrant des cellules immatures (en violet).

Pour aider les médecins à évaluer la gravité du neuroblastome, un système de classification nommé

Shimada a été développé à la fin des années 90 [2]. Ce système se base sur l'aspect des cellules tumorales observées, en prenant en compte leur morphologie et leur organisation. Les cellules sont alors classées selon plusieurs catégories, allant de celles qui sont très matures (peu agressives), à celles qui sont très immatures (très agressives). Cependant, l'analyse histologique est une tâche complexe et fastidieuse, nécessitant à la fois une expertise et du temps.

L'objectif de ce stage est double. Premièrement l'accélération du processus de diagnostic pour une prise en charge plus efficace des enfants atteints de cette maladie. Deuxièmement, la réduction des erreurs humaines par l'amélioration de la précision de la classification.

Afin d'atteindre ces objectifs, la solution étudiée dans ce stage concerne l'utilisation d'algorithmes d'apprentissage automatique, et en particulier les réseaux de neurones convolutifs (CNNs). Les CNNs sont des algorithmes capables de traiter de grandes quantités d'images, et de reconnaître des motifs complexes dans les données. Ces algorithmes sont donc particulièrement bien adaptés à l'analyse histologique, car ils peuvent automatiquement identifier les cellules tumorales et les classer en fonction de leur stade de développement. Les caractéristiques extraites par CNNs ne sont pas nécessairement les mêmes que celles utilisées par les pathologistes. Il est donc possible d'améliorer la précision et la robustesse des diagnostics, contrairement à des approches plus classiques développant des algorithmes imitant la démarche des docteurs.

2.2. Étapes établies

Le projet est divisé en 3 parties :

- 1. L'état de l'art** sur les méthodes et performances atteintes vis-à-vis du type de tumeurs étudiées, ainsi que sur les technologies liées aux réseaux de neurones CNNs.
- 2. Le pré-traitement de la base de données** pour former les datasets nécessaires à l'entraînement. Des problématiques de nettoyage, de régularisation et d'indépendance des données sont présents. De plus, comme la base de données peut évoluer au cours du stage via l'ajout de nouvelles images et/ou de nouvelles classes, il est important d'automatiser le plus possible cette phase de pré-traitement.
- 3. La recherche de la bonne architecture** de modèle CNN permettant la classification. Il s'agit de la partie la plus chronophage. Des expérimentations devront être menées sur la base des données fournies.

2.3. Avancement actuel

En six semaines de stage, l'état de l'art ainsi que la partie pré-traitement des données ont été réalisés. Les premières expériences sur la base de *transfer learning* sont en cours. L'idée est d'identifier, parmi les réseaux de neurones de type CNN, des architectures pré-entraînées sur des bases d'images conséquentes qui capturent des caractéristiques intéressantes pour le problème de classification posé. Cette approche permet de résoudre en partie la faible quantité de données disponibles. La famille des ResNet (Residual Network) entraînée sur ImageNet est actuellement étudiée. D'autres familles de réseaux pourront être analysées suivant le temps restant et les résultats déjà obtenus.

3. Conclusion

Après six semaines au laboratoire, ce projet de classification du neuroblastome semble bien lancé. La mise en place d'une approche automatisée pour le preprocessing de la base de données permet de gagner du temps lorsque celle-ci est mise à jour. On peut ainsi se concentrer sur l'étape cruciale de la recherche d'une architecture. Les premières expérimentations menées avec un backbone ResNet 18 montrent que l'approche de *transfer learning* est pertinente. Ces résultats incitent à poursuivre dans cette direction.

Références

- [1] National Cancer INSTITUTE. *Neuroblastoma Treatment (PDQ) - Health Professional Version*. <https://www.cancer.gov/types/neuroblastoma/patient/neuroblastoma-treatment-pdq>. (Visité le 09/03/2023).
- [2] Hiroyuki SHIMADA et al. "The International Neuroblastoma Pathology Classification (the Shimada system)". In : *Cancer* 86.2 (1999), p. 364-372. DOI : [https://doi.org/10.1002/\(SICI\)1097-0142\(19990715\)86:2<364::AID-CNCR21>3.0.CO;2-7](https://doi.org/10.1002/(SICI)1097-0142(19990715)86:2<364::AID-CNCR21>3.0.CO;2-7).