

# Devoir Maison

Student name: *Durand Enzo 21107517*

---

Course: *Algorithmes sur les arbres et les graphes en bio-informatique (AAGB)*  
Due date: *December 10th, 2021*

---

## Sujet : analyse d'une base de données en rapport avec la bio-informatique

Lien de la base de données : ADNI

ADNI (Alzheimer's Disease Neuroimaging Initiative) est un projet d'étude créé afin de mieux comprendre la maladie d'Alzheimer. C'est une maladie neuro-dégénérative qui a pour symptôme une perte de la mémoire et une dégénérescence des capacités intellectuelles (aphasie, apraxie, agnosie *etc...*). Cette maladie peut être due à des prédispositions familiales mais aussi plusieurs facteurs de risques dont l'âge, l'hypertension artérielle, le surpoids, le tabagisme *etc....* Ce projet s'intéresse principalement au dépistage de la maladie afin de prévoir un traitement rapide. Il peut donc être intéressant d'étudier Alzheimer de différentes façons :

- La première est une approche génétique basée sur l'analyse de séquences ADN. Il est possible d'utiliser des techniques d'alignement de séquence pour trouver des similitudes entre les différents patients atteints de la maladie. Il est aussi possible de trouver des portions de séquence qui diffèrent suivant si le sujet est atteint d'Alzheimer, cela permet de détecter les zones en relation directe avec cette maladie. La partie **genetic** de la base de données est donc utile pour ces études.
- Une seconde approche est l'analyse des données comprenant les facteurs de risques. Les dernières techniques de régression ou encore de deep learning peuvent être très performantes pour analyser des données tabulaires. Avec cette base de données, il est possible d'entraîner des réseaux de neurones avec en entrée les différents facteurs de risques et en sortie une prédiction binaire : sain ou malade. La partie **clinical** et **biospecimen** pourra nous être utile pour constituer nos données.
- La dernière est une analyse de données d'imagerie. En effet Alzheimer est une maladie neuro-dégénérative, ce qui implique qu'on peut observer des IRM (Imagerie par Résonance Magnétique) et des TEP (Tomographie par Émission de Positons) afin de la détecter. Pour ce type d'analyse, le deep learning est encore une solution très performante. On peut utiliser des réseaux de neurones convolutifs avec en entrée des images sous la forme de matrice représentant les pixels et en sortie une prédiction binaire : sain ou malade. Des techniques d'analyse d'image peuvent permettre d'améliorer nos données afin d'obtenir de meilleures prédictions ou encore d'augmenter le nombre de données. C'est donc la partie **MRI image** et **PET image** qui va nous intéresser pour ce problème.

En combinant ces différentes techniques d'analyse et donc ces différents modèles, on peut obtenir un super-modèle très performants dans la detection de la maladie. On peut ensuite utiliser nos études sur la prédiction de la maladie afin d'affiner nos connaissances sur Alzheimer. Ce projet a pour avantage d'avoir une base de données très diversifiée. Il est difficile de trouver des défauts car les données sont aussi quantitatives que qualitatives. Cependant pour certaines d'entre elles il y a un nombre d'exemples probablement trop faible pour la phase d'apprentissage des algorithmes de deep learning que nous pourrions utiliser.