



Enseirb-
matmeca

RAPPORT DE PROJET VISION ARTIFICIELLE
3A INFORMATIQUE

Détection d'Alzheimer à partir d'IRM du cerveau

Equipe :

MOUGOU YASSINE
LEGER CORENTIN
KINI WALID
Kafak JACQUES

Encadrant :

MICHAËL CLÉMENT

December 19, 2022



RÉSUMÉ

Dans notre article nous présenterons les méthodes mises en œuvre afin de prédire le stade de développement de la maladie d'Alzheimer par analyse d'IRM 3D. Afin d'y parvenir, nous avons utilisé le jeu de données ADNI1 qui contient des images 3D d'IRM de patients, et avons entraîné différents réseaux de neurones à classer ces images en deux catégories : CN (cognitively normal) ou AD (Alzheimer's Disease). Ces modèles entraînés ont ensuite servi à réaliser une classification sur des IRM de patients étant dans un état intermédiaire (MCI) entre CN et AD.

INTRODUCTION

La maladie d'Alzheimer (MA) est un trouble neurologique progressif qui affecte les capacités cognitives et la mémoire d'une personne. C'est la cause la plus fréquente de démence chez les personnes âgées et elle peut avoir un impact significatif sur la qualité de vie de la personne concernée et de son aidant. La détection et le diagnostic précoces de la MA sont importants pour la gestion et le traitement de la maladie.

Une façon de classer les stades de la MA est d'utiliser l'échelle d'évaluation de la démence clinique (REC), qui comprend quatre catégories : pas de déficience cognitive (CN), démence douteuse (REC=0,5), déficience cognitive légère (REC=1) et démence (REC=2). Dans la catégorie de la déficience cognitive légère (DCL), il existe deux sous-types : la DCL subjective (DCLS), dans laquelle la personne est consciente de son déclin cognitif, et la DCL objective (DCLP), dans laquelle le déclin est détectable par une évaluation clinique mais n'est pas signalé par la personne.

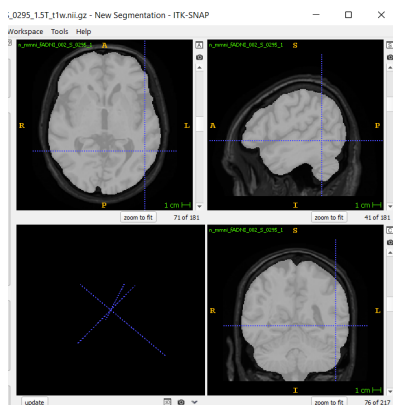
L'objectif de ce projet est de classer les individus dans l'une des quatre catégories du PCEM (CN, sMCI, pMCI, AD) en utilisant des techniques d'apprentissage automatique. Pour ce faire, nous avons recueilli des données sur diverses caractéristiques cliniques et démographiques de personnes atteintes de la maladie d'Alzheimer et de troubles cognitifs mineurs, ainsi que d'un groupe témoin de personnes sans déficience cognitive.

Dans les sections suivantes, nous décrirons les données utilisées dans cette étude, les méthodes employées pour l'analyse et les résultats obtenus. Nous discuterons également des implications de ces résultats et suggérerons des orientations pour les recherches futures.

Présentation du dataset ADNI1

Le jeu de données ADNI1 est un jeu de données utilisé pour la classification de la maladie d'Alzheimer. Il s'agit d'un jeu de données qui contient des scans d'imagerie par résonance magnétique (IRM) structurelle du cerveau, ainsi que des données cliniques telles que l'âge, le sexe et les scores cognitifs. Ce jeu de données est souvent utilisé dans les études de recherche pour évaluer les performances des algorithmes d'apprentissage automatique pour le diagnostic de la maladie d'Alzheimer.

Le jeu de données ADNI1 contient des scanners IRM d'une grande cohorte de personnes, comprenant à la fois des témoins sains et des personnes atteintes de la maladie d'Alzheimer.



En plus des scanners MRI, des masques sont également fournis dans le jeu de données ADN1. Ces masques peuvent être appliqués aux scanners pour isoler et se concentrer sur des régions spécifiques du cerveau, tout en ignorant d'autres parties du scanner qui ne sont pas pertinentes pour l'analyse.

Le jeu de données ADN1 comprend des données structurées, qui sont des données organisées dans un format ou une structure prédéfinis. Ce type de données est généralement plus facile à analyser et à manipuler que les données non structurées, qui n'ont pas de format prédéterminé.

Dans le jeu de données ADN1, les données structurées comprennent des informations cliniques telles que l'âge, le sexe et les scores cognitifs de chaque individu du jeu de données. Ces données sont généralement recueillies à l'aide de tests et d'évaluations normalisés.

Les données structurées dans le jeu de données ADN1 sont stockées sous forme de fichier csv, chaque ligne représentant un individu et chaque colonne représentant une information spécifique par exemple: âge, sexe, scores cognitifs, groupe(AD, CN, MCI) et conversion (permet de faire la différence entre les sMCI et les pMCI). Cela permet d'accéder et d'analyser facilement les données à l'aide

d'outils tels que des tableurs et des logiciels statistiques.

Approche 2D + epsilon

Principe:

Cette stratégie utilise principalement des coupes 2D avec un décalage inter-slices (epsilon). En effet, c'est une légère représentation 3D mais avec une perte d'informations. Dans cette approche on se concentre essentiellement sur la région d'intérêt (région liée au système de mémoire et qui est impactée par la maladie d'alzheimer). Nos tranches d'images 2D sont donc focalisées sur l'hippocampe. Cette partie du cerveau reflète les premières lésions de la maladie d'alzheimer pour un patient AD. Ce qui est bien pour cette approche c'est qu'il y pas d'informations perturbatrices provenant de côtés aberrants puisque la région d'intérêt prend en compte seulement la tranche centrale et ses voisins

L'idée de cette méthode est de faire 3 coupes voisines 2D de notre région d'intérêt (hippocampe), chacune de taille 40*40. On applique ce principe pour les 3 axes suivants: sagittal, axial, coronal. Pour chaque IRM on dispose donc de 3 images 2D chacune de dimensions (40,40,3). Nous allons prendre des coupes successives (epsilon vaut 1). Le choix d'une petite valeur du paramètre de décalage inter-slices est important, sinon d'autres zones hors région d'intérêt interviendront et baisserons les performances de notre modèle

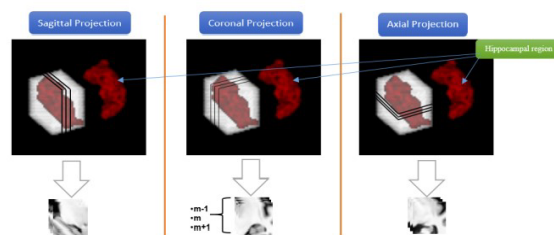
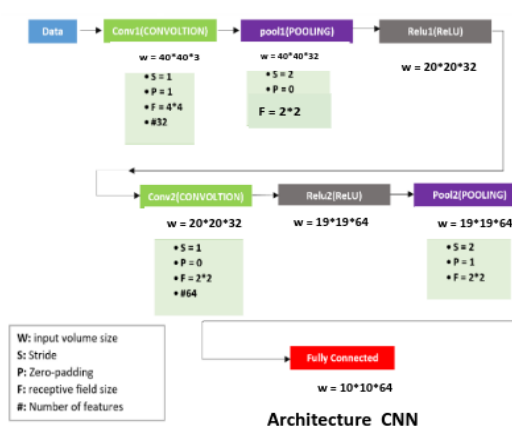


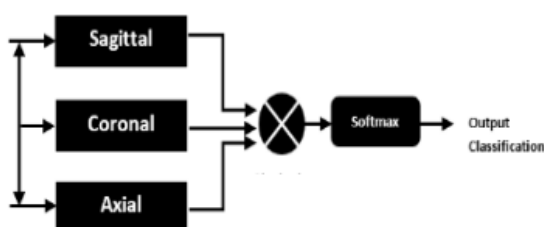
FIGURE 4 – Exemples des coupes 2D augmentées sur l'hippocampe

Modèle adopté:

Le modèle utilisé pour cette approche consiste à prendre en entrée séparément des coupes pour chaque type de projection (sagittal, coronal et axial). Pour chaque axe, on a créé un réseau de neurones CNN indépendant constitué de deux couches de convolutions 2D **SeparableConv2D** (32 filtres puis 64), deux **maxpooling** et suivies par la fonction d'activation **relu** avec un **fully-connected** à la fin.



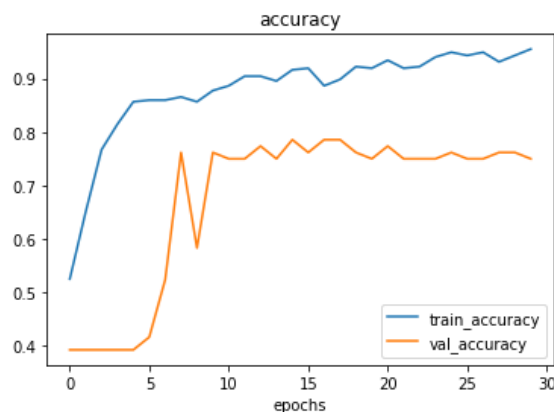
Ensuite nous avons considéré une méthode de fusion des 3 réseaux de neurones de projection qui est une concaténation des trois réseaux au niveau de la couche FC et on a ensuite appliqué une couche **Dense** avec sa fonction d'activation **softmax** pour générer la classe prédite.



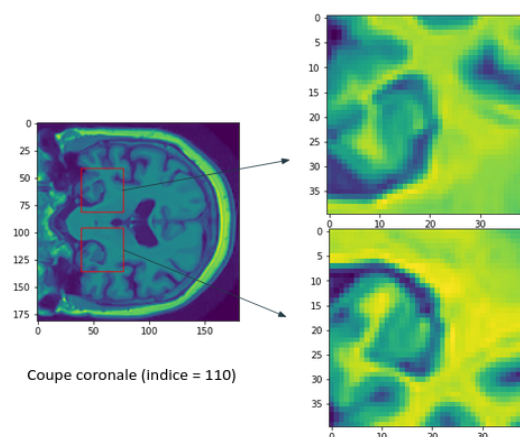
On a utilisé l'optimizer **Adam** comme algorithme de descente de gradient avec un **learning rate** de $1e-4$ et **binary_crossentropy** pour calculer la perte d'entropie à la fin.

Entraînement et résultats

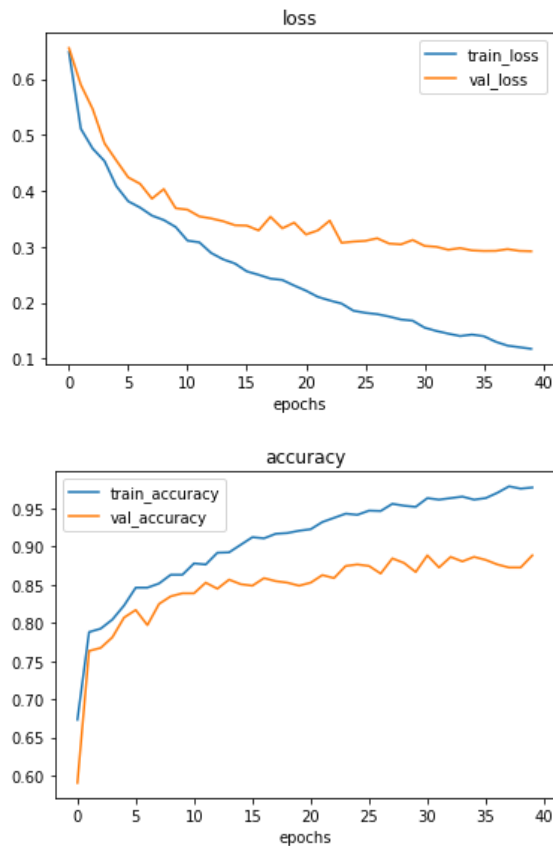
Le nombre de patients qui sont soit AD soit CN vaut 419. On applique le principe précédent et on ne prend en compte qu'un seul hippocampe, les dimensions de notre entrée sont (419,3,40,40,3). On sépare nos données en 80% pour l'entraînement et 20% pour la validation et on fit notre modèle. On obtient les métriques suivantes:



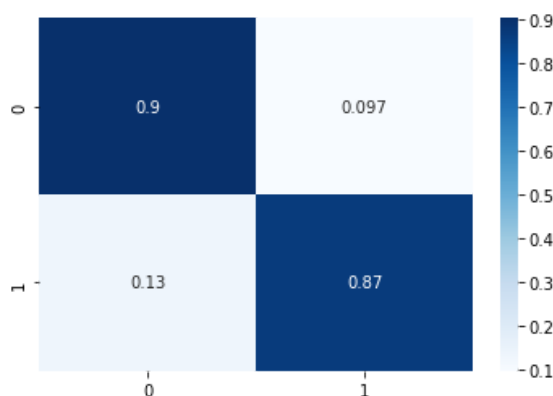
On obtient une faible accuracy sur les données de validation de 75%. Ceci s'explique par le fait que la taille de nos données est trop faible. Pour palier à ce problème et améliorer la performance de notre modèle, on a eu recours à des techniques d'augmentation de la data d'apprentissage. Pour ce faire, on a répété le processus de génération des coupes 2D plusieurs fois en coupant à des indices différents. Ainsi on a pris en considération les deux hippocampes (droite et gauche).



Si on fit notre modèle avec la data d'apprentissage augmentée durant 40 epochs, on obtient les métriques suivantes:



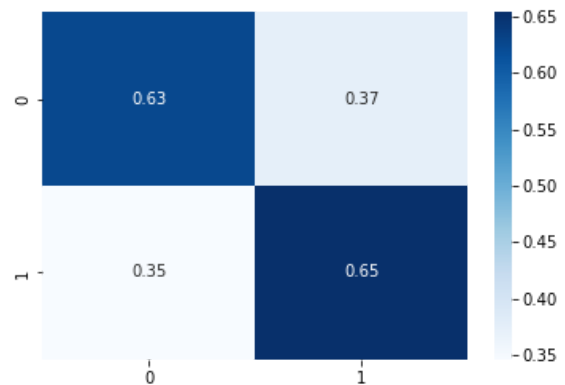
On obtient une bonne accuracy sur les données de validation de plus que 88%. On remarque que la courbe de loss est stable et converge (pas de overfitting).



On constate que 90% des patients CN ont été prédites sains et que 87% des patients malades ont été prédites AD.

Ensuite, on a testé une prédiction sur les données de patients étant dans un état intermédiaire (MCI). Pour se faire, on a mis les

labels des pMCI à 1 considérés comme AD et les sMCI à 0 supposés des CN. Les performances d'accuracy de notre modèle dépassent 64% pour les MCI.



Convolutions 3D sur le cerveau entier :

Inconvénient : On perd en résolution et donc en information

Les réseaux de neurones avec à base de convolutions 3D peuvent être utilisés pour analyser des images 3D telles que les IRM car ils permettent de capturer les caractéristiques spatiales des images en trois dimensions. Ces opérations de filtrage permettent de détecter des motifs spécifiques dans les images à l'aide de filtres 3D, ce qui est particulièrement utile pour l'analyse d'images volumétriques comme celles obtenues par IRM. Ces réseaux peuvent ici être utiles pour réaliser une classification, ce qui peut être très utile pour le diagnostic médical.

Un des avantages de l'utilisation de cette méthode est qu'on utilise toute l'information sur une image 3D sans biais par rapport à l'axe choisi, comme avec les convolutions 2D-epsilon. Le problème est cependant que les données d'IRM 3D sont très volumineuses (187*219*187), on choisit donc de ne prendre qu'un pixel sur 4, et on réduit donc leur taille par 64 (on perd donc aussi en information).

Le réseau de neurones que nous avons utilisé se compose donc de 3 blocs constitués des

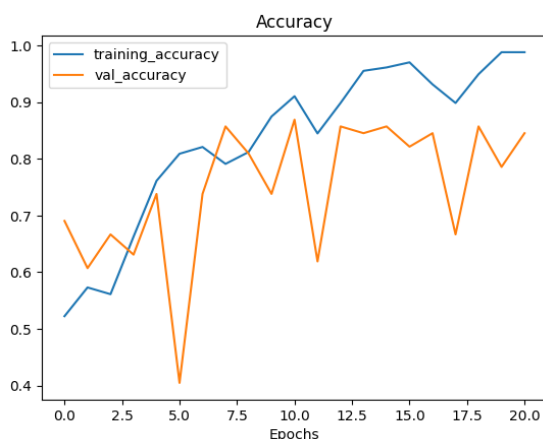
éléments suivants : une couche de convolution 3D (pour extraire de l'information 3D), une couche de MaxPooling3D (pour réduire la dimension des images), une couche de Batch Normalization (accélérer l'apprentissage et améliorer la performance) et une couche de Dropout (pour limiter l'overfitting).

Nous avons ensuite utilisé deux couches denses et une sortie avec une activation sigmoid pour réaliser la classification binaire en sortie : AD ou CN.

Entraînement et résultats

Lors de l'entraînement, nous avons également utilisé de l'early-stopping afin d'améliorer nos performances et tenter d'éviter le surapprentissage.

En faisant un train test split de nos données (20% de validation), nous obtenons donc une accuracy sur les données de validation de 87% (*model1*). Voici ci dessous les métriques obtenues sur l'accuracy :



Il nous a cependant semblé que les valeurs d'accuracy et de loss pour les données de validation étaient assez instables. Cela pourrait s'expliquer par la très faible quantité de données utilisées pour l'entraînement et la validation, en plus de la taille des batchs extrêmement faible. La précision de validation peut varier considérablement d'une itération à l'autre en raison de la variance statistique.

Second training (model 2) :

Il n'est pas nécessairement pertinent avec 100% des données pour faire une prédiction ensuite. En général, il est recommandé de réserver un pourcentage des données pour tester le modèle pendant son entraînement, afin de pouvoir évaluer sa performance et ajuster les choix des paramètres pour obtenir de meilleurs résultats.

Cependant dans notre cas, nous avons également tenté d'entraîner notre modèle avec l'ensemble des données disponibles au vu de la très faible quantité de données dont on dispose (*model2*). Cependant nous prenons le risque de faire un surapprentissage.

Pour le nombre d'epochs à fixer, nous n'avons pas utilisé d'early-stopping car cela n'a pas de sens sans données de validations, mais avons entraîné le modèle durant 20 epochs, en nous référant à la courbe d'apprentissage sur le *model1*.

Prédictions

Après avoir entraîné nos deux modèles sur nos données AD et CN, nous avons ensuite testé une prédiction sur les données de patients étant dans un état intermédiaire (MCI). Pour se faire, nous avons considéré que les pMCI (1) étaient des AD et les sMCI (0) des CN.

Voici nos performances d'accuracy sur les prédictions des deux modèles :

model1 : 0.63

model2 : 0.61

Convolutions 3D sur l'hippocampe :

1 - Modèle Initial

Après avoir réalisé notre prédiction sur un volume 3D contenant les IRM des cerveaux entiers de patients, nous avons décidé de

tester une prédiction en utilisant uniquement une partie spécifique du cerveau : l'hippocampe. En effet, le volume de cette structure cérébrale réduit avec l'âge et cette diminution est associée aux troubles de mémoire. l'accélération de ce phénomène est une des manifestations de la maladie d'Alzheimer.

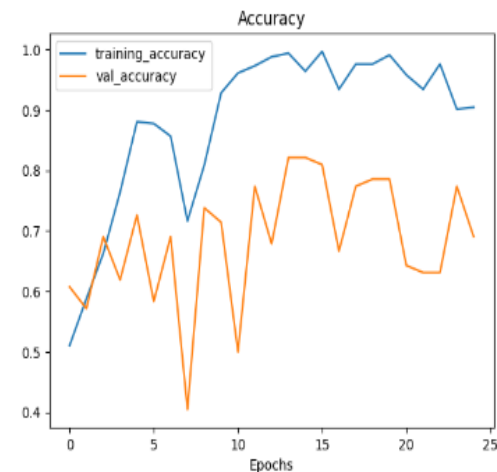
En étudiant uniquement les deux hippocampes des patients, on réduit drastiquement le volume des données (chaque hippocampe est contenu dans un volume de 40*40*40) tout en ciblant notre analyse sur une région qui semble être déterminante dans la détection de la maladie d'Alzheimer. Cependant, il résulte tout de même de ce choix une perte d'information sur les autres aires du cerveau qui pourraient également être déterminantes dans la détection de la maladie.

Nous avons fait le choix de faire une prédiction non pas pour les deux hippocampes de chaque patient, mais plutôt sur chaque hippocampe individuellement. Afin de faire la prédiction sur un individu, nous devons donc juste faire une fusion des résultats de ses deux hippocampes. Nous espérons que cette méthode nous permette artificiellement d'augmenter notre jeu de données.

Cependant, nous avons fait une erreur au niveau du train test split, en faisant un mélange aléatoire des données, ce qui fait que certains hippocampes de patients peuvent se retrouver à la fois en données d'entraînement mais aussi de test, ce qui vient biaiser les résultats obtenus, notamment au niveau de l'accuracy et de la loss. Cela ne devrait par contre pas avoir de répercussions sur la prédiction finale sur les données MCI.

Pour l'entraînement du modèle, nous avons utilisé le même modèle et la même procédure que pour les convolutions 3D sur le cerveau entier. Nous avons également entraîné deux modèles avec premièrement un train test split avec 20% de validation (*model1*) et un second modèle avec 100% de nos données (*model2*).

Voici ci-dessous les métriques de performance que nous avons obtenu pendant l'entraînement du *model1* :



Nous obtenons finalement une accuracy finale de 84% sur les données de validation. Cependant, les courbes de performance sont encore plus instables que pour le modèle précédent. Plusieurs pistes comme l'architecture du réseau de neurones ou bien la data augmentation sont donc à explorer afin de pallier ce phénomène.

Prédictions :

Voici les résultats obtenus grâce à nos deux modèles sur les hippocampes des patients MCI:

model1 : 0.61

model2 : 0.66

En fusionnant les résultats des deux hippocampes de chaque patient, on peut ensuite faire une prédiction pour le patient lui-même, ce qui nous donne les résultats ci-dessous :

model2 : 0.65

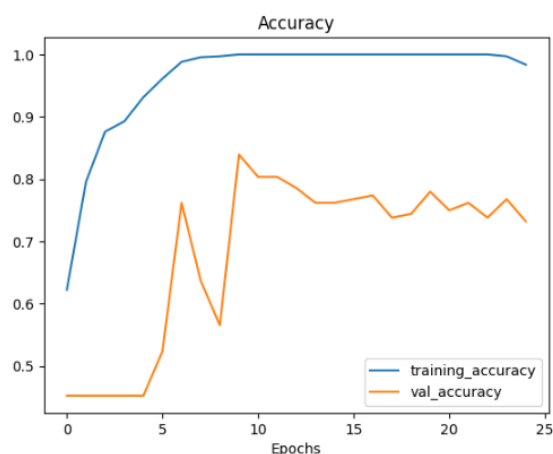
model2 : **0.7**

2 - Modèle avec réduction du réseau de neurones

Le choix des hyperparamètres du réseau de neurones, comme le taux d'apprentissage ou le nombre de couches et de neurones, peut affecter la stabilité de la précision sur l'ensemble de validation. On a donc tenté de modifier l'architecture de notre réseau de neurones afin d'observer l'évolution des performances.

Pour se faire, nous avons supprimé le troisième bloc de Conv3D, MaxPolling3D, BatchNorm et Dropout.

Cependant, nous observons des performances inférieures en utilisant cette méthode, comme le montre le graphique ci-dessous :



La meilleure accuracy obtenue sur les données de validation a cependant été de 84%.

Au niveau de la prédiction sur les MCI, les *model1* (80% des données en entraînement) et le *model2* (100% des données en entraînement) ont respectivement obtenu des scores d'accuracy de 66 et de 65%. 3 - Modèle Initial avec data augmentation

Nous avons également tenté d'augmenter la taille de notre jeu de données. La méthode utilisée est un simple 'flip' des hippocampes. Cela permet donc de doubler le jeu de données.

Cela nous a permis de légèrement améliorer

nos performances, en obtenant une accuracy de 67% pour les deux modèles créés

3D Convolutions - Zone aléatoire

Après ces différentes expériences, il nous a semblé que l'hippocampe joue un rôle important dans la maladie d'Alzheimer. Afin de vérifier si son analyse était vraiment déterminante pour la prédiction de la maladie, nous avons entraîné un autre modèle sur deux cubes de la même taille que ceux qui contenaient les hippocampes, dans une région au-dessus de l'hippocampe. Voici les résultats obtenus pour ce dernier modèle pour la classification de la région cérébrale choisie :

model1 : 0.51

model2 : 0.56

Voici la classification pour les patients (même méthode que pour la classification avec les hippocampes) :

model1 : 0.53

model2 : 0.56

Comme pour la prédiction avec les hippocampes, on trouve la prédiction pour les patients en fusionnant les résultats pour ses deux cubes de région 'aléatoire' du cerveau.

Pour rendre ce processus réellement aléatoire, il aurait fallu sélectionner aléatoirement des parties de 40*40*40 voxels dans le cerveau différentes de l'hippocampe, en faisant attention à ne pas sélectionner des voxels étant à l'extérieur des IRM du cerveau des patients.

On voit cependant ici que les résultats diminuent drastiquement en étudiant une autre région que l'hippocampe, ce qui semble confirmer son importance dans la prédiction de la maladie d'Alzheimer.

Resultats et Pistes d'amélioration

Les résultats des différents modèles et techniques utilisés peuvent être résumé par ce tableau:

Model	2D + Epsilon	3D whole brain	3D hippo	3D rand region
AD/CN Acc	0.88	0.87	0.86	0.68
MCI Acc	0.64	0.63	0.7	0.56

Les résultats de cette étude montrent que les différents modèles de machine learning évalués ont des performances variables pour la classification de la maladie d'Alzheimer et de la déficience cognitive légère (MCI). Le modèle 2D + Epsilon a obtenu la meilleure précision sur l'ensemble de données AD/CN, avec une précision de 88%. Le modèle 3D whole brain a obtenu la deuxième meilleure précision sur cet ensemble de données, avec une précision de 87%.

Sur l'ensemble de données DCL, le modèle 3D hippo a obtenu la meilleure précision, avec une précision de 70%. Le modèle 2D + Epsilon a obtenu la deuxième meilleure précision sur cet ensemble de données, avec une précision de 64%.

Bien que ces résultats soient satisfaisants, nous avons identifié plusieurs pistes d'amélioration pour améliorer les résultats de prédiction de notre projet. Voici une liste de ces pistes:

- Entraîner les modèles avec des données AD/CN, mais également avec des données MCI
- Utiliser d'autres métriques pour évaluer les résultats, comme le F-score et le taux de rappel
- Appliquer une meilleure stratégie de data augmentation, comme MixUp
- Trouver d'autres régions du cerveau importantes pour la classification de la maladie, en plus de l'hippocampe
- Intégrer les métadonnées des patients dans notre analyse, comme l'âge et le sexe
- Utiliser les masques fournis avec les images 3D pour cibler les parties du cerveau à analyser."

Références:

[12] H.-D. Nguyen, M. Clément, B. Mansencal, and P. Coupé, "Deep grading based on collective artificial intelligence for ad diagnosis and prognosis", Workshop on Interpretability of Machine Intelligence in Medical Image Computing (iMIMIC) at MICCAI, 2021.

<https://theses.hal.science/tel-03191293>