



UNIVERSITE SAAD DAHLAB DE BLIDA 1



Faculté de Technologie
Département d'Electronique

THESE DE DOCTORAT EN SCIENCES

Filière Génie Biomédical

IMAGERIE DE DIFFUSION DEDIEE A LA TRACTOGRAPHIE

Par

HOURIA Latifa

Devant le jury composé de :

Kara Kamel	Pr	Université Blida 1	Président
Djendi Mohammed	Pr	Université Blida 1	Examineur
Meziane Abdelkrim	Dr	CERIST-Alger	Examineur
Belkhama Nouredine	MCA	Université Blida 1	Directeur de Thèse
Bouzouad-Cherfa Assia	Pr	Université Blida 1	Co-directeur de Thèse

Blida, 22-01-2024

REMERCIEMENTS

Permettez-moi de remercier Allah le Tout-Puissant, de m'avoir donnée la foi, la force et la détermination nécessaires pour terminer cette thèse de doctorat.

Je tiens à remercier toutes les personnes, qui ont contribué au succès de cette thèse de doctorat, et toute l'équipe pédagogique de l'université de Blida 1.

Je tiens à remercier le Pr. Y Cherfa, Pr. A Cherfa et mon Directeur de thèse Mr N. Belkhamza pour leurs encouragements tout au long de mon travail, ainsi que pour leur accueil au sein de laboratoire LASICOM.

J'exprime ma gratitude aux membres de jury : Pr Kara Kamel, Pr Djendi Mohammed, et Dr Meziane Abdelkrim pour l'intérêt qu'ils ont porté à mon travail en acceptant de l'examiner.

Enfin, je remercie tous ceux qui, de prêt ou de loin, ont contribué à la réalisation de ce travail.

Je dédie ce travail à :

Ma mère et mon père

Toutes ma famille

Ceux que j'aime et à ceux qui m'aiment

RESUME

L'imagerie du tenseur de diffusion (DTI), est devenue une technique non invasive puissante, depuis son introduction en neuroimagerie. Il s'agit d'une nouvelle modalité d'imagerie, par résonance magnétique (MRI), qui donne un aperçu de la microstructure de la connectivité de la substance blanche (WM), aidant à identifier les perturbations de la WM, causées par des maladies neurodégénératives, telles que la maladie d'Alzheimer (AD), et son stade précoce, appelé «trouble cognitif léger» (MCI). La maladie d'Alzheimer est un trouble neurologique irréversible et progressif, qui entraîne une perte de mémoire et de capacité de réflexion. De nombreuses méthodes et algorithmes sont appliqués, pour extraire des modèles, à partir de données de neuroimagerie, afin de distinguer les différents stades de l'AD. Cependant, la similitude des schémas cérébraux, chez les personnes âgées, et à différents stades, rend leur classification difficile pour les chercheurs. Dans cette thèse, nous proposons un extracteur de caractéristiques, à réseaux de neurones convolutionnels profonds (CNN) bidimensionnels, et un classifieur de machine à vecteurs de support (SVM), permettant de classer le stade de la maladie. Tout d'abord, les données sont collectées, à partir de la base de données ADNI. Elles sont prétraitées efficacement, pour éviter tout bruit. Ensuite, les caractéristiques sont extraites, des deux cartes DTI (anisotropie fractionnelle (FA), et diffusivité moyenne (MD)), à l'aide du modèle CNN, puis classées par le SVM. La méthode proposée confirme l'efficacité du système de diagnostic assisté par ordinateur, en utilisant des cartes DTI, lorsqu'il est appliqué à la classification de l'AD. De plus, nous intégrons la DTI et l'imagerie par résonance magnétique structurale (sMRI), comme modalités d'imagerie complémentaires, pour la détection précoce de l'AD, afin de créer un outil de diagnostic assisté par ordinateur, plus efficace. Par conséquent, sMRI détecte l'atrophie corticale, et la DTI détecte les altérations de la substance blanche. Le cadre de fusion, consiste à fusionner des caractéristiques, extraites des métriques scalaires DTI (FA et MD), et de la matière grise (GM), segmentées à partir des images sMRI, à l'aide de 2D-CNN, et de les transmettre au SVM, pour classer, le stade AD vs le stade cognitif normal (CN), AD vs MCI, et MCI vs CN. L'utilisation de l'imagerie MRI multimodale, montre une amélioration significative de la classification, de 99,79 %, 99,6 % et 97,00 %, pour AD/CN, AD/MCI et MCI/CN, respectivement.

Mots clés : AD, imagerie de tenseur de diffusion (DTI), réseau de neurones convolutionnels profonds (CNN), classification , SVM.

ABSTRACT

Diffusion tensor imaging (DTI) has become a powerful non-invasive technique since its introduction in neuroimaging. It is a new Magnetic Resonance Imaging (MRI) modality that provides insight into the microstructure of white matter (WM) connectivity, helping to identify WM disruption caused by neurodegenerative diseases such as Alzheimer's disease (AD) and its early stage called “mild cognitive impairment” (MCI). AD is an irreversible and progressive neurological disorder that causes loss of memory and thinking skills. Many different methods and algorithms are applied to extract patterns from neuroimaging data to distinguish different stages of AD. However, the similarity of brain patterns in older people and at different stages makes classification a challenge for researchers. In this thesis, we propose a two-dimensional deep convolutional neural network (CNN) feature extractor and a support vector machine (SVM) classifier. First, data is collected from the ADNI database. They are strictly pretreated to avoid any noise. Then, the features are extracted from the two DTI maps (fractional anisotropy (FA) and mean diffusivity (MD)) using the CNN model and classified by SVM. The proposed method confirms the effectiveness of the computer-aided diagnosis system using DTI maps when applied to the classification of AD. Furthermore, we integrate DTI and Structural Magnetic Resonance Imaging (sMRI) as complementary imaging modalities for the early detection of AD to create a more efficient computer-assisted diagnostic tool. Hence, sMRI detects cortical atrophy and DTI detects white matter alterations. The fusion framework consists of merging features extracted from DTI scalar metrics (FA and MD) and gray matter (GM) segmented from the sMRI images using 2D-CNN. The SVM is then used to classify AD stage versus the normal cognitive stage (CN), AD versus MCI and MCI versus CN. The use of multimodal MRI feature shows a significant classification improvement of 99.79%, 99.6% and 97.00% for AD/CN, AD/MCI and MCI/CN respectively.

Key words: Diffusion tensor imaging (DTI), Convolutional neural network (CNN), Alzheimer's disease (AD), classification, SVM.

ملخص

أصبح التصوير الموتر للانتشار (DTI) تقنية قوية غير جراحية منذ إدخالها في التصوير العصبي. إنها طريقة جديدة للتصوير بالرنين المغناطيسي (MRI) توفر نظرة ثاقبة للبنية الدقيقة لاتصال المادة البيضاء (WM)، مما يساعد على تحديد اضطراب WM الناجم عن الأمراض التنكسية العصبية مثل مرض الزهايمر (AD) ومرحلته المبكرة التي تسمى "الضعف الإدراكي المعتدل" (MCI). مرض الزهايمر (AD) هو اضطراب عصبي تدريجي لا رجعة فيه يسبب فقدان الذاكرة ومهارات التفكير. يتم تطبيق العديد من الطرق والخوارزميات المختلفة لاستخراج الأنماط من بيانات التصوير العصبي للتمييز بين المراحل المختلفة لمرض الزهايمر. ومع ذلك، فإن تشابه أنماط الدماغ لدى كبار السن وفي مراحل مختلفة يجعل تصنيفهم تحدياً للباحثين. في هذه الأطروحة، نقترح مستخرج ميزة الشبكة العصبية التلافيفية العميقة ثنائية الأبعاد (CNN) ومصنف جهاز ناقل الدعم (SVM). أولاً، يتم جمع البيانات من قاعدة بيانات ADNI. بعد ذلك، يتم معالجتهم بشكل صارم لتجنب أي ضوضاء. بعد ذلك، يتم استخراج الميزات من خريطتي DTI (التباين الجزئي (FA) والانتشار المتوسط (MD)) باستخدام نموذج CNN وتصنيفها بواسطة SVM. تؤكد الطريقة المقترحة فعالية نظام التشخيص بمساعدة الكمبيوتر باستخدام خرائط DTI عند تطبيقها على تصنيف مرض الزهايمر. علاوة على ذلك، نقوم بدمج DTI والتصوير بالرنين المغناطيسي الهيكلي (sMRI) كطرائق تصوير تكميلية للكشف المبكر عن مرض الزهايمر لإنشاء أداة تشخيص أكثر كفاءة بمساعدة الكمبيوتر. ومن ثم، يكتشف التصوير بالرنين المغناطيسي (sMRI) الضمور القشري ويكتشف DTI تغيرات المادة البيضاء. يتكون إطار الاندماج من دمج الميزات المستخرجة من مقاييس DTI القياسية (MD و FA) والمواد الرمادية (GM) مقسمة من صور sMRI باستخدام D-CNN2 وتميرها إلى SVM لتصنيف مرحلة AD مقابل المرحلة المعرفية العادية (CN)، مقابل MCI و MCI مقابل CN. يُظهر استخدام ميزة التصوير بالرنين المغناطيسي المتعدد الوسائط تحسناً كبيراً في التصنيف بنسبة 99.79% و 99.6% و 97.00% لكل من AD / CN و AD / MCI و MCI / CN على التوالي.

الكلمات المفتاحية: التصوير الموتر للانتشار، الشبكة العصبية التلافيفية العميقة، التصنيف

LISTES DES ABREVIATIONS

AD : Alzheimer's Disease.

ADNI: Alzheimer's Disease Neuroimaging Initiative.

ANN : réseaux de neurones artificiels.

BN : Batch Normalization.

CAE: Convolutional Auto Encoder.

CN : sujet cognitivement normal.

CNN : Convolutional Neural Network.

DTI : Imagerie du tenseur de diffusion.

DWT: Discrete Wavelet Transform.

DWI: Diffusion Weighted imaging.

EMCI: Early Mild Cognitive Impairment.

FA: Fractional Anisotropy.

FC: fully connected.

FID: Free Induction Decay.

fMRI: Fonctionel Magnetic resonance Imaging.

GM: Gray Matter.

KNN: K plus proches voisins.

LD: Longitudinal diffusivity.

LMCI: Late Mild Cognitive Impairment.

LCR : liquide céphalo-rachidien.

MCI: Mild cognitive Impairment.

MD: Mean Diffusivity.

MMSE: Mini-examen of state mental.

ML: Machine Learning.

MO: Mode of Diffusivity.

MRI: Magnetic Resonance Imaging.

OMS : Organisation mondial de la santé.

PCA: Principal Component Analysis.

RD : Radial Diffusivity.

ReLU : Unité linéaire rectifiée.

RF : Randon Forest.

RMN : Résonance Magnétique Nucléaire.

ROI : la région d'intérêt.

sMRI : Structural Magnetic resonance Imaging.

SVM : Support Vector Machine.

SIFT: Scale-Invariant Feature Transform.

SURF: Speeded-Up Robust Features.

TBSS: Algorithme Tract-Based Spatial Statistics.

TE : Temps d'écho.

TEP : Tomographie par Emission de Positrons.

TR : Temps de répétition.

VBM : Voxel-based morphometry.

WM : White Matter.

TABLE DES MATIERES

Remerciements	i
RESUME	iii
ABSTRACT.....	v
ملخص.....	vi
Listes des abréviations	vii
TABLE DES MATIERES.....	ix
LISTE DES ILLUSTRATIONS GRAPHIQUES	xiii
LISTE DES ILLUSTRATIONS DES TABLEAUX.....	xvii
INTRODUCTION GENERALE	1
CHAPITRE 1: la malAdie d’Alzheimer et L’imagerie de diffusion	5
1.1 Introduction.....	5
1.2 La maladie d’Alzheimer	5
1.2.1 Changements cérébraux associés à la maladie d'Alzheimer.....	6
1.2.2 Différents stades de la maladie d'Alzheimer.....	7
1.2.3 Diagnostic de la maladie d'Alzheimer	8
1.3 imagerie par Résonance Magnétique	9
1.3.1 Résonance magnétique nucléaire.....	9
1.3.2 Influence d’un champ magnétique B_0	10
1.3.3 Perturbation par une onde de radiofréquence B1 (la résonance)	11
1.3.4 Phénomène de relaxation	11
1.3.5 Du signal RMN à la formation d’une image.....	13
1.3.6 Imagerie structurale (sMRI).....	14
1.4 MRI de diffusion.....	15
1.4.1 Diffusion moléculaire dans le cerveau humain.....	16
1.4.2 Acquisition des données de diffusion par MRI.....	17
1.5 Imagerie de tenseur de diffusion (DTI)	18

1.5.1 Notion DTI.....	18
1.5.2 Interprétation géométrique du tenseur de diffusion	19
1.5.3 Caractéristiques scalaires du tenseur de diffusion	20
1.5.4 Caractéristiques vectorielles du tenseur de diffusion.....	22
1.6 Conclusion	23
CHAPITRE 2: Détection de la maladie d'Alzheimer par MRI : etat de l'art.....	24
2.1 Introduction.....	24
2.2 La contribution de l'imagerie MRI, structurale et de diffusion, dans le diagnostic de la maladie d'Alzheimer :.....	24
2.2.1 Epaisseur corticale	25
2.2.2 Atrophie du volume de l'hippocampe	25
2.2.3 Cortex entorhinal	26
2.2.4 Élargissement des ventricules	26
2.2.5 Liquide céphalorachidien.....	27
2.3 approches pour l'évaluation de la maladie d'alzheimer	27
2.3.1 Méthodes basées sur les volumes	27
2.3.2 Méthodes basées sur les régions d'intérêts (ROI).....	28
2.3.3 Méthodes basées sur les coupes	28
2.3.4 Méthodes basées sur les voxels.....	29
2.4 Algorithmes de classification automatique pour la maladie d'Alzheimer	29
2.4.1 Approches monomodales, utilisant l'imagerie MRI structurale :.....	31
2.4.2 Approches monomodales utilisant l'imagerie de diffusion	35
2.4.3 Approche multimodale.....	39
2.5 Discussion	42
CHAPITRE 3: Méthodes d'apprentissage en profondeur pour la classification d'objets	44
3.1 Introduction.....	44
3.2 Apprentissage Automatique.....	45

3.2.1 Apprentissage supervisé :	45
3.2.2 Apprentissage non supervisé :	46
3.2.3 Apprentissage semi-supervisé (ou apprentissage semi-non supervisé) :	46
3.3 Réseaux de neurone artificiels	46
3.4 Fonction d'activation :	48
3.5 Neurone formel (Perceptron)	49
Perceptron multicouche (MLP)	49
3.6 Deep learning	50
3.6.1 Réseaux de neurones convolutifs	50
3.6.2 Optimisation :	56
3.6.3 Régularisation :	56
3.6.4 Autres architectures populaires des CNN :	58
3.7 Conclusion	60
CHAPITRE 4 Méthodologies	62
4.1 Introduction :	62
4.2 Présentation de la base de donnée ADNI	62
4.3 Prétraitement	64
4.3.1 Prétraitement des images de diffusion	66
4.3.2 Extraction du cerveau	66
4.3.3 Correction de la distorsion induite par les courants de Foucault et génération des cartes FA et MD	67
4.3.4 Extraction de cartes entières de matière grise	67
4.3.5 Recalage des images DTI-FA et DTI-MD	70
4.4 Sélection de coupes 2D	72
4.5 Extraction des caractéristiques	72
4.6 Classification à l'aide de la machine à vecteurs de support (SVM)	73
4.6.1 Machine à vecteurs de support (SVM)	74
4.6.2 Processus de fusion MRI multimodale	77

4.7 Conclusion	77
CHAPITRE 5 Résultats et discussion.....	78
5.1 Introduction.....	78
5.2 Evaluation de la classification	78
5.2.1 Métriques d'évaluation.....	79
5.3 Expérimentation et résultats.....	81
5.3.1 Résultat	85
5.3.2 Comparaison avec les études précédentes	92
5.4 Discussion.....	93
5.5 Conclusion	95
Conclusion Générale.....	96
Références.....	99
BIBLIOGRAPHIE DE L'AUTEUR	112

LISTE DES ILLUSTRATIONS GRAPHIQUES

Figure 1.1: Coupe coronale d'un cerveau illustrant la différence entre un cerveau sain et un cerveau atteint par l'Alzheimer (en ligne).	6
Figure 1.2 : Neurones sain et neurones non fonctionnels.	7
Figure 1.3: Moment magnétique du spin d'un noyau d'hydrogène.....	10
Figure 1.4: Influence du champ magnétique externe B_0 ; les atomes s'orientent aléatoirement dans des circonstances normales, et s'alignent selon une direction, en présence d'un champ magnétique.....	11
Figure 1.5: Relaxation longitudinale M_z	12
Figure 1.6: Relaxation transversale M_{xy}	13
Figure 1.7: Séquence écho de spin.....	15
Figure 1.8: Diffusion restreinte dans la SB. La diffusion, dans la direction des fibres $D_{ }$, est trois fois plus importante que la diffusion perpendiculaire D_{\perp} [21].	17
Figure 1.9: Schéma d'une séquence d'écho de spin, à gradient d'impulsion introduit par [22]. La durée de l'impulsion est notée δ et la différence de temps entre les impulsions de gradient, est notée Δ	18
Figure 1.10: Surfaces d'iso probabilité, dérivée du champ du tenseur de diffusion. On peut remarquer que dans chaque voxel, la surface d'iso probabilité est un ellipsoïde, qui est définie de manière unique, par les vecteurs propres, et les valeurs propres des tenseurs [25].	19
Figure 1.11: Diffusion isotrope et anisotrope. Dans le cas isotrope, la diffusion est similaire dans toutes les directions. Cependant, lorsque le champ est anisotrope, la diffusion est plus importante dans un sens que dans l'autre [26].	20
Figure 1.12: Un exemple des surfaces d'iso probabilités, dérivées du champ du tenseur de diffusion.	22
Figure 1.13: Coupes axiales d'un cerveau humain montrant les cartes quantitatives des mesures DTI. De gauche à droite : diffusivité moyenne (MD), anisotropie fractionnelle (FA), direction du vecteur propre majeur, indiqué par la carte des couleurs RVB.	23

Figure 2.1: Illustration de la mesure de l'épaisseur corticale par imagerie MRI.	25
Figure 2.2: Effets de la maladie d'Alzheimer sur le cerveau ; A : sMRI pondérées en T1, d'un sujet âgé, cognitivement normal (CN), d'un sujet souffrant de troubles cognitifs légers (MCI), et d'un sujet atteint de la maladie d'Alzheimer (AD) ; B : Ventricules latéraux agrandissement (vert) ; C : Atrophie progressive de la structure de l'Hippocampe (Rouge). .	26
Figure 2.3: Organigramme d'un algorithme d'apprentissage pour la classification des images	30
Figure 2.4: Méthode de représentation des données DTI :	36
Figure 2.5: Procédure générale d'un système d'aide au diagnostic en utilisant la DTI, consistant en quatre étapes : prise d'un ensemble de données d'imagerie par DTI, ou d'images multimodales, extraction de caractéristiques, à partir de l'ensemble de données, classification par apprentissage automatique en fonction des caractéristiques les plus significatives, diagnostic automatisé, obtenu en classant les scans individuels, dans une classe clinique spécifique [11].	38
Figure 3.1: Représentation graphique de la relation entre l'intelligence artificielle, l'apprentissage automatique (Machine Learning) et l'apprentissage profond (Deep Learning)[120].	44
Figure 3.2: Un neurone biologique (a) comparé à un neurone artificiel (b). L'analogie est claire : les deux neurones reçoivent plusieurs signaux, et les agrègent en une seule sortie qui est transférée plus loin.	46
Figure 3.3: Un exemple simple de réseau de neurones artificiels. Il n'a qu'une seule couche cachée composée de 5 unités, donc la profondeur du modèle est un. La couche cachée est entièrement connectée.....	48
Figure 3.4: Fonctions d'activation populaires : (a) : RELU, (b) : Sigmoides.....	49
Figure 3.5: Un réseau neuronal simple et un réseau neuronal convolutif.....	51
Figure 3.6: Vue d'un filtre glissant de la convolution. Le noyau, de taille 3*3, est décalé sur l'entrée (matrice de taille 4*4), et à chaque position, les cellules qui se chevauchent sont multipliées et additionnées. Chaque cellule, de la carte des caractéristiques, est le résultat d'une convolution, d'un voisinage spatial correspondant, appelé le champ récepteur.	52
Figure 3.7: Exemple simple d'arithmétique de convolution [131].....	52
Figure 3.8: une matrice 4x4 à zéro-padding	53

Figure 3.9: Mécanisme de la fonction RELU.	54
Figure 3.10: Exemple de pooling, avec les différentes stratégies (maximum, moyen et somme). La taille des fenêtres, ainsi que le pas utilisé dans cet exemple, sont de 2x2.	54
Figure 3.11: Couche entièrement connectée sur une matrice.	55
Figure 3.12: Exemple de la couche Softmax.	56
Figure 3.13: Exemple de couche dropout (la probabilité de 50 % apparaît à droite) [134].....	57
Figure 3.14: Un réseau de neurones, avant et après la normalisation par lots [136].	58
Figure 3.15: Illustration de l'arrêt précoce.	58
Figure 3.16: Un exemple de l'architecture LeNet.	59
Figure 3.17: Architecture AlexNet.	60
Figure 4.1: Organigramme du système de fusion multimodalité proposé, utilisant l'approche 2DCNN-SVM, pour l'identification de l'AD.....	64
Figure 4.2: Les images pondérées en diffusion (DWI) ;.....	67
Figure 4.3 : Le prétraitement effectué aux images T1, pour segmenter la matière grise.....	68
Figure 4.4: Les détails de la normalisation spatiale des images sMRI, à l'aide d'un modèle..	68
Figure 4.5: Images des trois plans (coronale, sagittale et axiale) ; (a) T1w originales (b) après utilisation de VBM à l'espace natif (visualisé à l'aide de la fonction Check-Reg du logiciel SPM).	69
Figure 4.6: les images résultantes des trois plans prétraitées, segmentées et recalées dans l'espace MNI ; (a) Tw1, (b) la matière grise, (c) la matière blanche (visualisé à l'aide de la fonction Check-Reg du logiciel SPM).	70
Figure 4.7 : Schéma du processus de recalage des images MD.....	71
Figure 4.8 : Le pipeline de GM-CNN, proposé avec la méthode SVM, pour distinguer AD et CN.	74
Figure 4.9:Classification avec la machine à vecteurs de support.	74
Figure 4.10: Représentation des données. (a) Données linéairement séparables. (b) Données non linéairement séparables.	75
Figure 5.1: Schéma de validation croisée à 5 folders.	79

Figure 5.2: La courbe d'apprentissage des CNN, pour la classification binaire CN vs AD, en utilisant : (a) les images DTI-FA, (b) les images DI-MD, (c) les images GM.	83
Figure 5.3: La courbe d'apprentissage des CNN, pour la classification binaire CN vs MCI, en utilisant : (a) les images DTI-FA, (b) les images DI-MD, (c) les images GM.	84
Figure 5.4: La courbe d'apprentissage des CNN, pour la classification binaire AD vs MCI, en utilisant : (a) les images DTI-FA, (b) les images DI-MD, (c) les images GM.	85
Figure 5.5: La courbe ROC de la classification AD vs CN.	86
Figure 5.6 : La courbe ROC de la classification AD vs MCI.	87
Figure 5.7 : La courbe ROC de la classification AD vs MCI.	87
Figure 5.8: Matrices de confusion de la classification AD vs MCI, en utilisant des caractéristiques, extraites de (a) FA, (b) MD, (c) GM, (d) FA+MD, (e) GM+FA, (f) MD+GM, (g) FA+MD+GM.	89
Figure 5.9: Matrices de confusion de la classification CN vs MCI, en utilisant des caractéristiques, extraites de (a) FA, (b) MD, (c) GM, (d) FA+MD, (e) GM+FA, (f) MD+GM, (g) FA+MD+GM.	90
Figure 5.10: Comparaison de la performance de la technique de classification binaire proposée AD vs. CN, AD vs. MCI, and CN vs. MCI.	92

LISTE DES ILLUSTRATIONS DES TABLEAUX

Tableau 2.1: Résumé des différentes approches de détections de la maladie d'Alzheimer. ...	40
Tableau 4.1: Données statistiques liées aux sujets.....	63
Tableau 4.2: Les différents outils contenus dans le logiciel FSL :	65
Tableau 4.3: Propriétés des couches pour l'architecture 2DCNN proposée	73
Tableau 5.1: Une représentation indicative d'une matrice de confusion pour un classifieur binaire. Toutes les prédictions correctes sont situées dans la diagonale de la matrice.	80
Tableau 5.2: Performance de l'ensemble de validation	86
Tableau 5.3: Performance evaluation of the proposed 2DCNN-SVM technique on the test dataset.	91
Tableau 5.4: Comparaison des résultats avec les techniques de pointe appliquées à la détection de l'AD.	92

INTRODUCTION GENERALE

Les principales sources d'information, dans le domaine médical, sont les images. Ces images sont obtenues, à travers une variété de modalités, dont chacune a des propriétés uniques. Parmi ces modalités, l'imagerie par résonance magnétique (MRI), est une modalité de base, dont nous allons donner les principes.

Aujourd'hui, le MRI est un outil incontournable, dans un bon nombre d'examen cliniques, plus particulièrement en neurosciences cognitives. MRI offre un large éventail de possibilités d'exploitation, et nous pouvons citer : MRI anatomique ou bien structurale (sMRI), pour la visualisation des tissus, constituant le cerveau, et l'imagerie du tenseur de diffusion (DTI), pour le suivi de la connectivité anatomique du cerveau. Le principe général sur lequel se base la DTI, est l'observation de la diffusion des molécules d'eau, dont les déplacements sont contraints, par l'organisation tissulaire, au sein de l'organisme. Rapidement, la DTI suscita un grand intérêt, en imagerie du cerveau, et représente maintenant un outil d'aide au diagnostic, d'un grand nombre de pathologies, touchant la matière blanche (WM). La maladie d'Alzheimer en est un cas typique, qui va faire le sujet de cette thèse.

La maladie d'Alzheimer (AD) est une maladie neurodégénérative, et la cause la plus fréquente de démence dans le monde. Selon des statistiques récentes, cette maladie a touché environ 50 millions de personnes en 2018, et ce nombre devrait tripler d'ici 2050 [1]. En général, l'AD affecte la partie du cerveau, qui contrôle le langage et la mémoire, provoquant la mort cellulaire, et un déclin progressif de la fonction de la capacité cérébrale, entraînant, éventuellement, des troubles de la mémoire, et des troubles cognitifs. La transition, d'un état sain à l'AD, prend plusieurs années, en commençant par un trouble cognitif léger (MCI). De nos jours, la prévalence de la maladie d'Alzheimer, augmente considérablement chaque année, ce qui rend essentiel l'identification de moyens efficaces, pour ralentir, et prévenir la maladie. Le taux moyen, de conversion de MCI en AD, est d'environ 10 à 12 %. Ainsi, le diagnostic de MCI peut aider à une intervention précoce, contre la maladie d'Alzheimer.

L'AD se caractérise, par une perte progressive de matière grise (GM), qui survient de manière asymptomatique, dans certaines structures neuroanatomiques [2]. La sMRI est la modalité de neuroimagerie la plus utilisée, pour détecter l'atrophie cérébrale. Elle a déjà mis en évidence, de nombreux biomarqueurs de la maladie d'Alzheimer, en particulier, l'atrophie de structures, telles que l'hippocampe, l'amygdale et le thalamus [3]. En fait, l'atrophie hippocampique, chez les patients prodromiques, s'est avérée être le meilleur prédicteur structurel, de la progression de la maladie d'Alzheimer. Dans ce contexte, de nombreuses études, sur la phase prodromique de l'AD, ont concentré leurs recherches, sur l'hippocampe. Cependant, elle est associée à un grand nombre de pathologies neurodégénératives, limitant ainsi sa spécificité à la maladie d'Alzheimer [4].

Néanmoins, certaines autres structures semblent intéressantes, comme le volume de l'amygdale, qui pourrait être un prédicteur structurel aussi puissant, voir plus efficace, que le volume de l'hippocampe, pour prédire le MCI [5, 6]. De plus, il existe des changements dans la substance blanche, qui précède l'atrophie de la substance grise, mais qui ne sont pas détectables par la sMRI [7]. L'introduction de l'imagerie en tenseur de diffusion, a permis l'identification de ces modifications, quand le patient présente encore un MCI [8], par conséquent, elle fournit des informations supplémentaires, à la modalité sMRI. L'anisotropie fractionnelle (FA) et la diffusivité moyenne (MD), font partie des mesures, dérivées du tenseur de diffusion, liées à l'intégrité de la matière blanche (WM). FA désigne le degré d'anisotropie de la diffusion de l'eau, et MD désigne son amplitude. Elles décrivent la perturbation de la connectivité, entre les régions du cerveau, causée par les dommages axonaux, conduisant au processus de dégénérescence dans l'AD [9]. L'efficacité de l'utilisation de FA et MD, pour détecter le vieillissement physiologique, chez les patients MCI et AD, a été démontrée dans plusieurs études [10]. En général, ils ont constaté, que les patients atteints de l'AD, avaient un MD plus élevé, et une FA plus faible, que les sujets cognitivement normaux (CN), ainsi qu'un MD plus élevé, dans les deux hippocampes (droite et gauche) du MCI.

Les techniques traditionnelles d'évaluation de l'AD sont l'analyse visuelle, et la mesure manuelle des structures, par les radiologues. Cependant, avec l'avancement des techniques de classification automatique, il est désormais possible, d'évaluer rapidement les modifications de l'AD, chez un grand nombre de patients. En raison de ses excellentes performances, dans le domaine médical, l'apprentissage automatique (ML), et les techniques, basées sur l'apprentissage en profondeur, ont suscité une attention accrue dans ce domaine [11].

L'évaluation des changements physiopathologiques, par neuroimagerie, serait essentielle, pour prédire la maladie. Dans ce contexte, nous cherchons à détecter, des modèles de changements micro et macrostructuraux, dans les différents stades de l'AD, en utilisant la modalité DTI. Nous développons un nouveau système d'aide au diagnostic, basé sur l'utilisation des deux cartes DTI-FA et DTI-MD. Nous proposons un nouveau réseau de neurones convolutifs (CNN), pour extraire les caractéristiques, et le classifieur SVM, pour classer : AD vs MCI, AD vs CN et MCI vs CN. Après cela, nous ajoutons la modalité sMRI, comme modalité complémentaire à la DTI. Nous extrayons les fonctionnalités, à partir d'images de la matière grise. Ces fonctionnalités sont concaténées avec celles des FA et MD, puis transmises au SVM, pour les classer.

Structure du document

Après une introduction générale, le présent manuscrit est organisé autour de cinq chapitres. Le premier chapitre, qui constitue une entrée en la matière, expose le contexte médical et technologique. Un aperçu sur la maladie d'Alzheimer, et ses différents stades est donné, de même que quelques notions sur l'imagerie par résonance magnétique utilisée sont présentées. Nous commencerons par les techniques d'imagerie MRI, en termes de physique, nous expliquerons la modalité adoptée, qui est l'imagerie de diffusion. Nous verrons comment est exploitée la diffusion des molécules d'eau dans l'organisme, et nous passerons ensuite à l'explication de la DTI.

Dans le deuxième chapitre, nous passons en revue, les méthodes existantes, dans la littérature, de la classification de la maladie d'Alzheimer, en utilisant des images monomodales, et des images multimodales sMRI-DTI.

Le troisième chapitre explique l'apprentissage automatique, et l'apprentissage en profondeur. Ces deux techniques sont les plus récentes, et les plus utilisées dans la littérature, et sont importantes, pour appréhender la suite de la thèse.

Dans le quatrième chapitre, nous expliquons en détail la méthode proposée. Nous commençons par l'étape de prétraitement, puis l'étape de la sélection des coupes, et par l'utilisation d'une région spécifique du cerveau. Ensuite, nous présentons l'architecture CNN utilisée, et le classifieur SVM. Nous étendons l'étude au cadre multimodale, où nous ajoutons la modalité sMRI, par un processus de fusion des caractéristiques.

Le dernier chapitre est consacré à l'étude expérimentale. Nous présentons nos résultats, par l'utilisation des deux cartes DTI (FA et MD), et la carte GM, ainsi que la fusion de ces cartes,

et nous faisons une comparaison de nos résultats avec ceux des travaux récents dans la littérature. Nous terminons le chapitre par une discussion des résultats.

Nous clôturons cette thèse par une conclusion générale concernant la méthode proposée, et les recherches futures dans le domaine, pour améliorer la détection précoce, et la classification de cette pathologie énigmatique.

CHAPITRE 1

LA MALADIE D'ALZHEIMER ET L'IMAGERIE DE DIFFUSION

1.1 INTRODUCTION

Aujourd'hui, l'Imagerie par Résonance Magnétique est un outil incontournable dans un bon nombre d'examens cliniques, plus particulièrement en neurologie, en neurochirurgie, et dans la recherche en neurosciences cognitives. MRI offre un large éventail de possibilités d'exploitation, et nous pouvons citer l'imagerie structurale (sMRI), pour la visualisation des tissus constituant le cerveau, et l'imagerie de diffusion, pour le suivi de la connectivité anatomique du cerveau.

Dans ce chapitre, nous vous invitons à découvrir la maladie d'Alzheimer, ainsi que la différence entre le vieillissement normal et le vieillissement pathologique. Puis nous présenterons la théorie et le concept de base de sMRI et du DTI, et terminons par une conclusion.

1.2 LA MALADIE D'ALZHEIMER

La maladie d'Alzheimer est une maladie neurodégénérative, progressive et irréversible, qui touche principalement les personnes âgées, de plus de 65 ans. Elle est causée par des lésions des cellules nerveuses, dans certaines parties du cerveau. Elle commence au niveau de l'hippocampe, puis s'étend au reste du cerveau, affectant la mémoire et les capacités cognitives d'une personne, ce qui perturbe sa vie quotidienne.

Au début, la maladie est silencieuse, voire discrète, marquée par des symptômes banals, qui, avec le temps, finie par s'aggraver. Parmi ces symptômes, nous distinguons :

- Des troubles de la mémoire (amnésie), qui sont souvent une perte de la mémoire à court terme, qui peut s'étendre à une perte de la mémoire à long terme, qui peut durer des mois, ou même des années.
- Des troubles du comportement, qui se traduisent par une réduction de l'activité, une incapacité à reconnaître sa propre maladie (anosognosie), et qui peut mener à des problèmes psychiatriques, voire à une dépression.

- Des troubles du langage (aphasie), où le malade a des difficultés, pour formuler des mots simples. La destruction neurale continue, entraînant une perte d'autonomie, et par la suite la mort.

Selon l'organisation mondiale de la santé (OMS), il s'agit de la sixième cause de décès aux États-Unis. Cependant, si aucune précaution n'est prise, le taux d'incidence augmentera, il y aurait 13,8 millions de patients, atteints de l'AD en 2050 dans le monde [1].

La maladie d'Alzheimer est connue, comme la principale cause de démence [12], provoquée par des dommages des cellules cérébrales. À la suite de ces dommages, les cellules cérébrales ne sont pas capables de communiquer correctement entre elles, ce qui entraîne la perte du fonctionnement cognitif, et des capacités comportementales de la personne. La démence comporte différents stades, du stade le plus bénin, où la personne commence à perdre certaines fonctions cognitives, au stade le plus grave, où la personne dépend complètement des autres pour vivre [13].

À l'heure actuelle, aucune prévention ou traitement efficace n'a été trouvé. Plus important encore, cela coûte beaucoup d'argent et de main-d'œuvre, pour guérir (prendre en charge) les patients atteints de l'AD. Une détection précoce de cette maladie, est devenue une tâche impérative. Une visualisation d'un cerveau atteint d'Alzheimer, et un cerveau sain est illustrée dans la figure 1.1, montrant clairement l'atrophie du cortex.

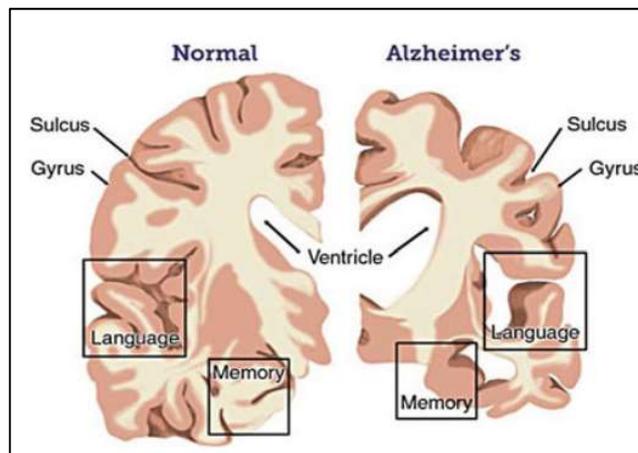


Figure 1.1: Coupe coronale d'un cerveau illustrant la différence entre un cerveau sain et un cerveau atteint par l'Alzheimer (en ligne).

1.2.1 Changements cérébraux associés à la maladie d'Alzheimer

À l'intérieur de chaque cerveau humain, adulte en bonne santé, il y a environ 100 milliards de neurones, avec des extensions ramifiées, qui leur permettent de communiquer entre eux. Les connexions entre les neurones appelées synapses, libèrent et détectent des substances

chimiques, pour faciliter le flux d'informations entre les neurones. Le cerveau contient environ 100 000 milliards de synapses. Ces synapses assurent la transmission des signaux, à travers les circuits neuronaux du cerveau, pour créer la base cellulaire des souvenirs, des pensées, des sensations, des émotions, des mouvements et des compétences.

L'augmentation du niveau de la forme anormale de protéine tau (tau ou enchevêtrements neurofibrillaires), à l'intérieur des neurones, et l'augmentation du niveau de protéine bêta-amyloïde (plaques bêta-amyloïdes), à l'extérieur des neurones, sont deux des principaux changements dans un cerveau touchés par la maladie d'Alzheimer (Figure 1.2). Les plaques bêta-amyloïdes interrompent la communication entre les neurones, et provoquent la mort cellulaire, tandis que les enchevêtrements de tau, bloquent la réception des nutriments et d'autres molécules essentielles, à l'intérieur des neurones [14-16]. En raison de ces changements, le cerveau des personnes atteintes de la maladie d'Alzheimer grave rétrécit, en raison de la perte de cellules. Dans les scanners cérébraux, on peut voir les débris et l'inflammation répandus à l'intérieur du cerveau à la suite de la mort des neurones.

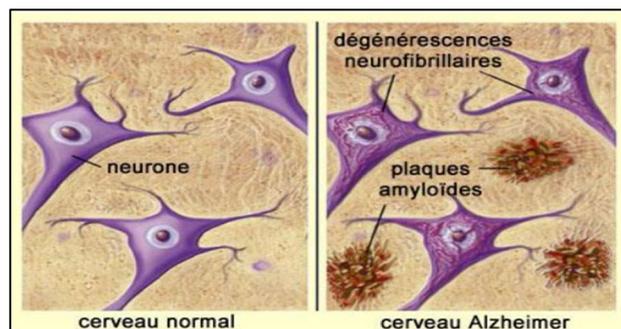


Figure 1.2 : Neurones sain et neurones non fonctionnels.

1.2.2 Différents stades de la maladie d'Alzheimer

La maladie d'Alzheimer évolue sur plusieurs années, et sa progression varie d'une personne à une autre. On sait que les premières lésions apparaissent dans le cerveau, au moins 10 ans à 15 ans, avant les premiers symptômes [12]. Ceux-ci apparaissent généralement, après l'âge de 60 ans. En moyenne, une fois que la maladie est déclarée, l'espérance de vie est de 8 ans à 12 ans. Plus la maladie survient à un âge avancé, plus elle tend à s'aggraver rapidement.

Trois stades de la maladie ont été identifiés :

- **AD préclinique** : environ la moitié des personnes, dans cette phase, ne signalent pas de troubles cognitifs, quelques années avant le diagnostic, car les dégénérescences cellulaires, associées à l'AD, commencent des années, voire des décennies avant que les sujets ne présentent, pour la première fois des symptômes cliniques. En effet, des

changements biologiques sont en cours dans le corps, avant que les symptômes de la maladie n'apparaissent. Il est difficile de quantifier les modèles de changement structurel, dans les premiers stades de l'AD, ou pendant les stades cliniquement normaux. Ainsi, un diagnostic précis de la maladie en phase clinique, n'est pas encore possible.

- **Déficience cognitive légère ou Mild cognitive impairment en anglais (MCI) :** la déficience cognitive légère est la phase transitoire, de la plupart des patients, avant de sombrer dans l'AD. À ce stade, le sujet peut présenter des problèmes de mémoire, bien avant d'obtenir un diagnostic d'Alzheimer. La maladie dans cette phase n'est pas assez grave, pour perturber la vie d'une personne. Le MCI est une étape difficile et confuse, car dans cette phase, le sujet n'est pas encore considéré, comme atteint de l'AD.
- **Diagnostic clinique de l'AD :** c'est le stade avancé de la maladie d'Alzheimer, qui peut également se qualifier de « sévère ». Les sujets à ce stade présentent une diminution des capacités mentales, une perte totale de la fonction cognitive, en raison de la dégénérescence des cellules du cerveau, et finalement cela provoque la mort.

1.2.3 Diagnostic de la maladie d'Alzheimer

Le diagnostic de la maladie d'Alzheimer nécessite un examen médical complet et exhaustif. Il est assez facile, pour les médecins, de déterminer si un patient est atteint de démence, mais il est difficile d'identifier les causes de la démence. C'est un long processus pour les médecins, de poser leur diagnostic, sur la base des tests des patients, et des résultats d'examen. Il existe de nombreuses façons de diagnostiquer la maladie d'Alzheimer. Les médecins, avec l'aide de neurologues et de gériatres, utilisent différentes approches et outils, pour établir un diagnostic. Certains de ces approches et outils sont les suivants :

- Antécédents médicaux et familiaux des patients, y compris les antécédents psychiatriques et cognitifs ;
- Observation de changements de comportement, et de capacité de réflexion des patients ;
- Réalisation d'un mini-examen de l'état mental (MMSE), et d'examens physiques et neurobiologiques ;
- Test sanguin et imagerie cérébrale, pour déterminer d'autres causes de démence, y compris les carences en vitamines ou les tumeurs.

Il existe de nombreuses techniques de neuroimagerie différentes, pour aider les médecins à mieux comprendre les progrès du patient, et à poser un diagnostic plus précis. L'une des techniques les plus utilisées est l'imagerie par résonance magnétique, en tant qu'élément essentiel de l'évaluation clinique. L'étude de la MRI d'un patient est désormais un biomarqueur valable, pour diagnostiquer la maladie, en particulier dans les stades prodromiques (c'est-à-dire le stade de la déficience cognitive légère (MCI)). Les mesures de l'atrophie, montrent des lésions neuronales cumulatives dans le cerveau, comme l'une des principales raisons de l'état clinique. En comparaison avec d'autres biomarqueurs, l'atrophie cérébrale a une forte corrélation avec le déclin cognitif [3]. Cependant, certains types inhabituels de l'AD ont des schémas d'atrophie irréguliers et, dans certains cas, les schémas d'atrophie peuvent se chevaucher avec d'autres maladies. La sMRI n'est pas en mesure de préciser le comportement moléculaire, ni d'identifier directement les autres causes de l'AD, telles que les plaques amyloïdes et les enchevêtrements de tau, ni d'étudier les anomalies au niveau de la matière blanche. Par conséquent, de nouvelles techniques sont devenues disponibles dans le domaine de la neuroimagerie, telles que la tomographie par émission de positrons (TEP), l'imagerie de diffusion, et l'imagerie par résonance magnétique fonctionnelle (fMRI) [17, 18].

Dans cette thèse, nous utilisons des données d'imagerie MRI (structurale et de diffusion), qui détectent des changements micro-macrostructuraux.

1.3 IMAGERIE PAR RESONANCE MAGNETIQUE

L'imagerie par Résonance Magnétique est une technique d'imagerie médicale, non invasive, qui permet l'exploration in vivo, des tissus mous, comme le cerveau, le cœur, la moelle épinière et les muscles dans l'organisme. Cette technique repose sur le phénomène de la résonance magnétique nucléaire (RMN), décrit en 1946, par l'équipe de Félix Bloch de l'université de Stanford et celle d'Edward M. Purcell à Harvard.

1.3.1 Résonance magnétique nucléaire

Nos tissus corporels contiennent plus que 75% d'eau. Le phénomène de la résonance magnétique nucléaire (RMN) est basé sur le fait que, le noyau de l'atome d'hydrogène est composé d'un unique proton ($1H$), chargé positivement, qui est généralement en mouvement (se comporte comme une petite planète (la terre) qui tourne constamment autour d'un axe). Il peut être considéré comme un petit aimant, appelé 'spin', qui est en rotation autour de lui-

même, et qui induit un moment cinétique (μ), représenté par un vecteur S comme le montre la figure 1.3.

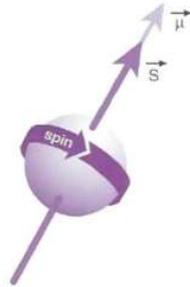


Figure 1.3: Moment magnétique du spin d'un noyau d'hydrogène

1.3.2 Influence d'un champ magnétique B_0

A l'état de repos, les moments magnétiques de spin ont des directions aléatoires (la résultante magnétique ($\sum \vec{\mu}$) d'un ensemble d'atome est donc nulle). En présence d'un champ magnétique extérieur intense (les protons sont placés dans un champ magnétique B_0 intense), les moments magnétiques s'alignent parallèlement ou antiparallèlement, à la direction de ce champ (oz). Cela produit l'apparition d'un vecteur d'aimantation macroscopique \vec{M}_z appelé, magnétisation longitudinale, la composante dans le plan transversal \vec{M}_z étant nulle.

La plupart des protons sont alignés, suivant la direction du champ magnétique, avec une énergie potentielle stable et faible, c'est-à-dire d'une façon parallèle à la direction de B_0 . Cependant, la faible proportion des protons restants, est disposée antiparallèlement au champ magnétique (Figure 1.4). Pendant ce temps, ils précessent autour de B_0 , avec un angle donné, avec une fréquence caractéristique, appelée fréquence de Larmor f_0 , telle que (eq.1.1) :

$$f_0 = \gamma B_0 \quad (1.1)$$

La constante γ est le rapport gyromagnétique qui vaut 42,58 MHz/Tesla pour le noyau d'Hydrogène.

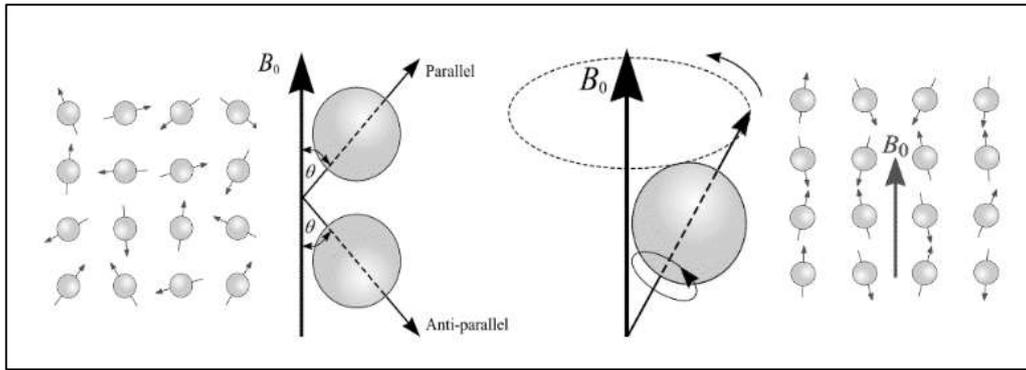


Figure 1.4: Influence du champ magnétique externe B_0 ; les atomes s'orientent aléatoirement dans des circonstances normales, et s'alignent selon une direction, en présence d'un champ magnétique.

1.3.3 Perturbation par une onde de radiofréquence B_1 (la résonance)

Selon l'orientation et l'alignement des spins, la particule a deux niveaux d'énergie. L'un est le niveau d'énergie élevé, et l'autre est le niveau d'énergie faible.

Pour perturber l'état d'équilibre (M_z), une impulsion radiofréquence B_1 (ou deuxième champ magnétique dit tournant), est appliquée, perpendiculairement au champ primaire B_0 , à la fréquence de Larmor. Ceci conduit à un transfert d'énergie au niveau atomique, connu par le phénomène de résonance, qui fait basculer un proton, d'un état énergétique faible, à un état élevé. Néanmoins, ce transfert se fait à une certaine fréquence, bien déterminée. L'utilisation d'une onde radiofréquence provoque un basculement de l'aimantation, suivant différents angles 30° , 90° , 180° possibles (fonction de l'amplitude de l'onde, et sa durée). Les spins s'alignent alors également avec B_1 , et tournent autour de B_0 . Ceci permet l'apparition de la composante transversale (M_{xy}), et la disparition de la composante longitudinale (M_z).

1.3.4 Phénomène de relaxation

L'arrêt du champ magnétique exciteur B_1 , permet un retour à l'état d'équilibre, ou de relaxation. Elle est observée, par une diminution de l'aimantation transversale (M_{xy}), jusqu'à son annulation, et une augmentation de l'aimantation longitudinale (M_z), jusqu'à ce qu'elle atteigne sa valeur maximale. Cela induit l'apparition d'un signal, appelé signal de précession libre, ou Free Induction Decay (FID). On peut distinguer deux types de relaxation, la relaxation longitudinale et la relaxation transversale.

La relaxation longitudinale, aussi appelée relaxation spin-réseau, est caractérisée par le temps T_1 , qui correspond aux interactions des spins avec leur environnement. Il en résulte une

croissance de l'aimantation longitudinale, due à la repousse de l'aimantation magnétique M_z , à son état initial.

Cette croissance suit une loi exponentielle croissante exprimée par l'équation de Bloch (eq.1.2) :

$$M_z = M_0(1 - e^{(-t/T_1)}) \quad (1.2)$$

Le temps T_1 représente le temps nécessaire, pour que la magnétisation longitudinale M_z , récupère 63 % de sa valeur finale, après une impulsion de 90° (Figure 1.5).

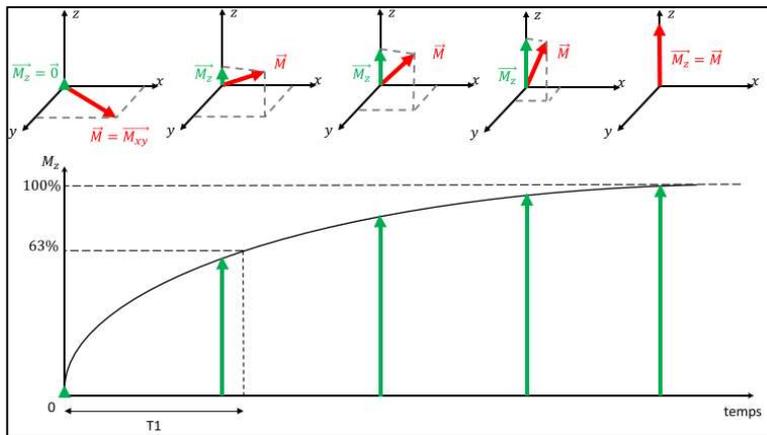


Figure 1.5: Relaxation longitudinale M_z

La relaxation transversale, ou relaxation spin-spin, est caractérisée par le temps T_2 . L'interaction électromagnétique des spins eux-mêmes, provoque un déphasage des protons, qui entraîne une chute progressive de l'aimantation transversale M_{xy} . En effet, les spins donnent leur énergie aux molécules voisines, et aux spins voisins non excités. On parle alors de la relaxation « spin-spin »

Cette chute est modélisée par une loi exponentielle décroissante, donnée par la formule (eq.1.3) :

$$M_{xy} = M_0 e^{(-t/T_2)} \quad (1.3)$$

Le temps T_2 correspond au temps pour lequel l'aimantation transversale retrouve 37 % de sa valeur initiale, après une impulsion de 90° (Figure 1.6).

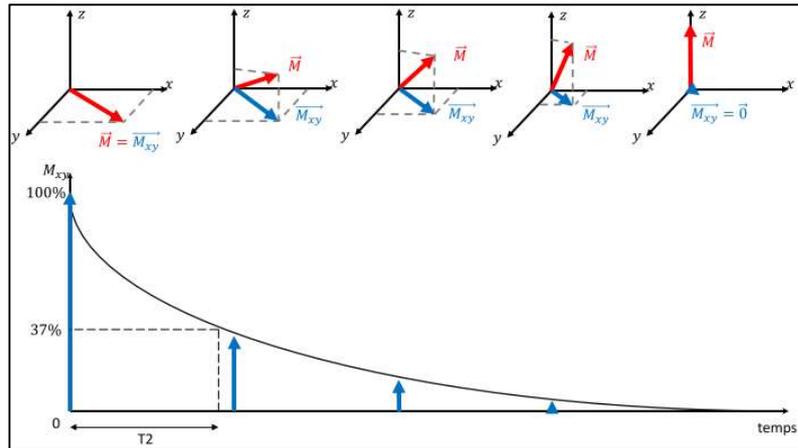


Figure 1.6: Relaxation transversale M_{xy}

1.3.5 Du signal RMN à la formation d'une image

La MRI est basée sur la technique RMN, et permet de cartographier l'ensemble des excitations. Le signal RMN produit des échos, qui proviennent d'un volume élémentaire d'un échantillon (le corps humain en routine clinique). La reconstruction de l'image se fait à partir de ces échos, qui codent l'information spatiale. Ceci est réalisé par l'application des trois gradients de champ magnétique. Un gradient de champ est un champ magnétique, de faible intensité et variable dans l'espace, qui s'ajoutera linéairement et progressivement, au champ B_0 .

Pour une image 2D, il est nécessaire de localiser son origine spatiale, le plus précisément possible. Pour atteindre cet objectif, trois étapes sont nécessaires :

- Il faut d'abord sélectionner une coupe du volume observé, on parle ici de la première étape, nommée codage de sélection de coupe.
- Ensuite, il faut sélectionner une ligne de cette coupure (codage en fréquence).
- Puis, une partie de cette ligne est sélectionnée (codage en phase).

Le gradient de sélection de coupe G_{coupe} , permet la sélection d'une coupe, à l'intérieur d'un volume d'intérêt. A cet effet, une onde de radiofréquence, correspondant à celle de précession des protons du plan d'intérêt, est appliquée. Elle permet de n'exciter, que les protons de ce plan, les autres n'étant pas excités, rendant ainsi le signal RMN créé, caractéristique de la coupe sélectionnée.

Le gradient de codage de phase G_{phase} intervient dans un temps très limité. Il s'applique, perpendiculairement au gradient de sélection de coupe, permettant le codage des différentes lignes, du plan de coupe sélectionné. Par conséquent, la fréquence sera modifiée, en fonction de chaque ligne, le long de ce gradient, ce qui induira un déphasage différent, selon la ligne

qui persistera à la fin de l'application de ce gradient. A la fin de l'application du gradient, les spins retrouvent la même vitesse de précession, mais avec un déphasage, dépendant de la ligne à laquelle il appartient. Ce déphasage persistera jusqu'à ce que le signal soit collecté.

Le gradient de fréquence, ou gradient de lecture, est appliqué perpendiculairement au gradient d'encodage en phase G_{phase} . Il permet la modification des fréquences de précession, créant des colonnes de protons, ayant une vitesse de précession identique.

L'utilisation de ces trois gradients, permet le codage spatial du volume, dans les trois directions de l'espace. Si l'on considère x , l'axe de codage fréquentiel, et y , l'axe de codage de phase, on obtient N_y signaux, acquis avec différentes valeurs de gradient de codage de phase, et chaque signal comprend N_x points échantillonnés, en présence du gradient de lecture. Ainsi, les données recueillies, se situent dans le domaine fréquentiel, appelé espace k .

L'ensemble des signaux obtenus, est ainsi encodé dans l'espace de Fourier (eq.1.4), qui est une matrice formée de $N_x \times N_y$ points. Le transformée de Fourier 2D inverse (eq.1.5), permet le décodage du signal, à savoir le passage du domaine fréquentiel au domaine spatial, compréhensible par l'œil humain, et qui donne lieu à une image en niveaux de gris.

$$TF[f(x, y)] = F(v_x, v_y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) e^{-i2\pi(v_x x + v_y y)} dx dy \quad (1.4)$$

$$TF^{-1}[F(v_x, v_y)] = f(x, y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} F(v_x, v_y) e^{i2\pi(v_x x + v_y y)} dv_x dv_y \quad (1.5)$$

Avec $f(x, y)$ fonction continue, x et y sont les coordonnées spatiales, et v_x et v_y représentent les coordonnées spectrales.

Les basses fréquences de l'image « l'information de contraste », se trouvent au centre du plan de Fourier. A l'inverse, les détails de l'image, correspondant aux hautes fréquences, se retrouvent à la périphérie du plan de Fourier.

1.3.6 Imagerie structurale (sMRI)

En faisant varier les deux paramètres d'acquisition fondamentaux, le temps d'écho TE et le temps de répétition TR, des différents types de pondération (T1, T2) sont obtenues.

Temps d'écho (TE) : c'est la durée qui sépare l'application de l'onde RF d'excitation, du milieu du temps de lecture (Figure 1.7).

Temps de répétition (TR) : c'est l'intervalle de temps, entre deux impulsions excitatrices successives, de 90° (Figure 1.7).

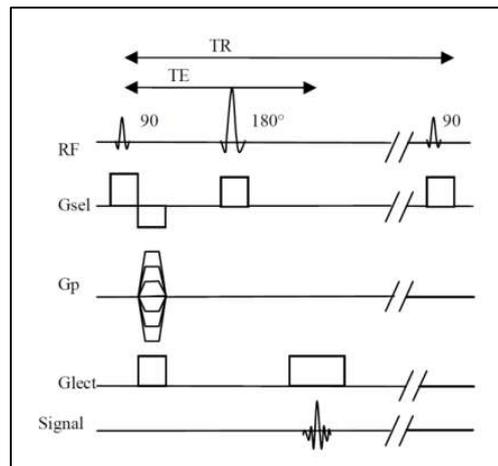


Figure 1.7: Séquence écho de spin.

- **Pondération T1** : la pondération T1 (également connue sous le nom de temps de relaxation spin-réseau T1), offre un bon contraste, entre les tissus de la matière grise (gris foncé), et de la substance blanche (gris plus clair), tandis que le liquide céphalo-rachidien (LCR) est dépourvu de signal (noir). Le contraste de ces images permet une différenciation précise, des structures cérébrales [19]. La MRI pondérée en T1 est construite, avec un TE court et un TR court.
- **Pondération T2** : contrairement à l'image pondérée en T1, la MRI pondérée en T2 est construite avec un TE long et un TR long ($TR > 2\ 000\text{ ms}$, $TE > 80\text{ ms}$). Le LCR est clair, et le tissu cérébral (sombre), la matière grise est en gris clair, et la matière blanche est en gris foncée. Le contraste, en pondération T2, permet aux radiologues, de mieux voir les anomalies, dans les ventricules et le cortex cérébral, que sur les images pondérées en T1, en raison de la meilleure mesure de la teneur en eau. Cependant, les limites de la substance blanche ne sont pas aussi clairement définies, que dans les images pondérées en T1[19].

1.4 MRI DE DIFFUSION

En 1827, le botaniste anglais Robert Brown observe la marche aléatoire de grains de pollen, immergés dans l'eau. Cette marche, appelée mouvement Brownien, servira de modèle, pour décrire le phénomène de diffusion. Adolph Fick propose en 1855, des lois, inspirées de la loi de Fourier pour la chaleur (1822), permettant de décrire le phénomène. Ces lois seront démontrées ensuite par Albert Einstein en 1905, et confirmées par Jean Perrin en 1908.

En 1994, l'équipe de Basser et al. [20], introduit un modèle, permettant de caractériser, in vivo, le phénomène de diffusion moléculaire dans le cerveau, la structure microscopique des tissus vivants, ainsi que leur intégrité. Ce modèle, appelé modèle de tenseur de diffusion, se base sur l'hypothèse, que la diffusion dans le cerveau est libre, tolérant ainsi, sa modélisation par une distribution gaussienne. Bien que, cette hypothèse ne soit pas conforme à la réalité anatomique des structures biologiques, l'imagerie tensorielle de diffusion ou Diffusion tensor Imaging (DTI), reste très utiles, dans l'exploration de la diffusion dans le cerveau humain, et s'introduit aujourd'hui en routine clinique.

1.4.1 Diffusion moléculaire dans le cerveau humain

Dans un milieu biologique, la diffusion libre est perturbée, en raison des différents obstacles comme les membranes cellulaires ou les fibres, qui composent ce milieu. Dans ce cas, on parle de diffusion restreinte, et le coefficient de diffusion, ne dépend plus seulement de la viscosité du milieu, mais également de la géométrie de celui-ci.

Le cerveau présente trois structures, qui modulent la diffusion : le liquide céphalo-rachidien, la substance grise et la substance blanche. La diffusion des molécules d'eau dépend de la composition et de l'agencement des tissus, qui le composent. L'observation de la diffusion de l'eau, permet donc de donner des indices, sur la structure, et l'organisation des tissus neuronaux.

Le LCR est assimilé à la diffusion en milieu homogène de faible viscosité. Le coefficient de diffusion est donc important (environ $3,2 \times 10^{-3} \text{ mm}^2/\text{s}$), et identique dans toutes les directions de l'espace.

GM et WM, qui sont de nature plus visqueuse, ont des coefficients de diffusion plus faibles, compris entre $0,6 \times 10^{-3} \text{ mm}^2/\text{s}$ et $0,8 \times 10^{-3} \text{ mm}^2/\text{s}$. Cependant, la structure spécifique de chacun de ces tissus, génère des propriétés de diffusion très différentes. La GM est constituée d'un maillage dense de corps neuronaux, de dendrites et d'axones, ce qui confère à la diffusion un caractère quasi-isotrope.

La diffusion est anisotrope, au niveau de la WM, en raison de l'organisation spécifique en fibres parallèles serrées, des faisceaux axonaux de la WM, ce qui affecte, de manière significative, la diffusion, qui est plus rapide dans le sens des fibres, que dans le sens transversal (Figure 1.8).

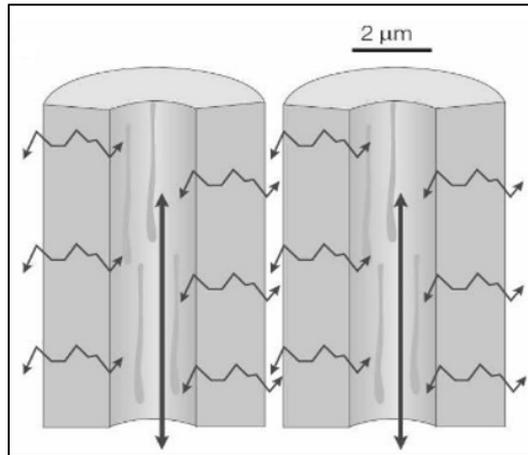


Figure 1.8: Diffusion restreinte dans la SB. La diffusion, dans la direction des fibres $D_{||}$, est trois fois plus importante que la diffusion perpendiculaire D_{\perp} [21].

1.4.2 Acquisition des données de diffusion par MRI

La MRI de diffusion vise à obtenir des images, en explorant les micromouvements des molécules d'eau à l'intérieur des voxels. Cette technique permet de quantifier l'anisotropie de la diffusion, afin d'avoir des informations, sur la structure des tissus, et plus précisément, de construire l'orientation globale des fibres nerveuses, de la matière blanche.

Pour mesurer la diffusion par la MRI, une paire de forts gradients de champs, est ajoutée à une séquence MRI conventionnelle, basée sur la technique écho de spin [22]. Le premier gradient de diffusion, marque les spins suivant leurs positions au temps $t = 0$. Après l'impulsion radiofréquence de rephasage de 180° , le deuxième gradient de diffusion est appliqué, et marque à nouveau les spins, suivant leur position à $t = \Delta$. Si les spins n'ont pas bougé, les marques s'annulent, alors que, si les spins ont bougé, cela engendre une perte du signal, proportionnelle au déplacement des spins. L'espacement entre les impulsions, et la durée d'impulsion sont donnés par, Δ et δ , respectivement (Figure (1.9)).

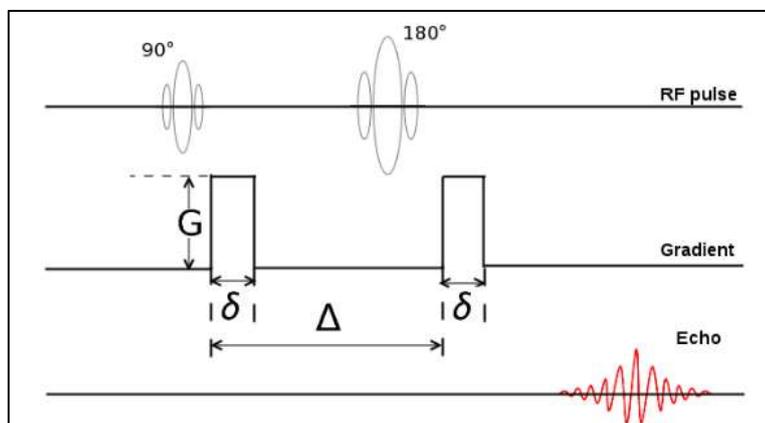


Figure 1.9: Schéma d'une séquence d'écho de spin, à gradient d'impulsion introduit par [22]. La durée de l'impulsion est notée δ et la différence de temps entre les impulsions de gradient, est notée Δ .

1.5 IMAGERIE DE TENSEUR DE DIFFUSION (DTI)

1.5.1 Notion DTI

L'imagerie par tenseur de diffusion (DTI) est une technique de résonance magnétique, relativement nouvelle [9] et une extension de la MRI de diffusion.

Elle mesure la quantité de diffusion d'eau, à différentes positions du cerveau. Elle est considérée comme un modèle tridimensionnel, permettant de représenter le degré d'anisotropie dans chaque voxel (pixels volumétriques), dans l'espace 3D.

En milieu isotrope, le phénomène de diffusion peut être entièrement décrit par un scalaire, correspondant au coefficient de diffusion. Mais en milieu anisotrope, un seul coefficient ne suffit plus, il nous faut donc mesurer le coefficient de diffusion, dans plusieurs directions de l'espace. Plus de trois directions de gradient orthogonales, doivent être acquises.

En DTI, le signal de résonance magnétique de diffusion est exprimé par [23] (eq.1.6) :

$$S(b) = S_0 e^{-bD} \quad \text{et} \quad b = (\gamma G \delta)^2 \left(\nabla - \frac{\delta}{3} \right); \quad (1.6)$$

Où S_0 est l'intensité de l'image d'origine, au niveau du voxel (sans la pondération de diffusion), lorsque $b=0$; S est l'intensité mesurée, après l'application du gradient de diffusion, dans la direction unitaire, b caractérise les impulsions de gradient utilisées, dans la séquence MRI, y compris l'amplitude de gradient G , la durée d'impulsion de gradient et le temps de diffusion Δ . Dans la séquence d'écho de spin [22], γ désigne le rapport gyromagnétique du proton ^1H , et le tenseur de diffusion est donné par (eq.1.7) :

$$D = \begin{pmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{pmatrix} \quad (1.7)$$

Selon les lois thermodynamiques, le tenseur de diffusion des molécules d'eau, est considéré comme symétrique, $D_{ij} = D_{ji}$, avec $i, j = x, y, z$. Les éléments diagonaux D_{xx} , D_{yy} et D_{zz} sont les coefficients de diffusion, le long des axes x , y , et z .

Le tenseur de diffusion étant symétrique, le signal de diffusion devient [23] (eq. 1.8) :

$$S(b) = S_0 e^{-(b_{xx}D_{xx} - b_{yy}D_{yy} - b_{zz}D_{zz} - 2b_{xy}D_{xy} - 2b_{xz}D_{xz} - 2b_{yz}D_{yz})} \quad (1.8)$$

1.5.2 Interprétation géométrique du tenseur de diffusion

La DTI est souvent utilisée, pour détecter les orientations préférées, dans les tissus. Mathématiquement, un tenseur est souvent représenté par un ellipsoïde [9], avec ses trois axes principaux, appelés vecteurs propres (e_1 , e_2 et e_3) (Figures 1.9 et 1.10), obtenus par une diagonalisation de la matrice D. Chaque vecteur propre correspond à une valeur propre, exprimée par λ . On suppose que la direction du vecteur propre associé à la plus grande valeur propre du tenseur, correspond à l'orientation locale du tissu (Figure 1.9). L'orientation de diffusion, tirée du DTI, est couramment utilisée, pour suivre les fibres dans le cerveau [21, 24].

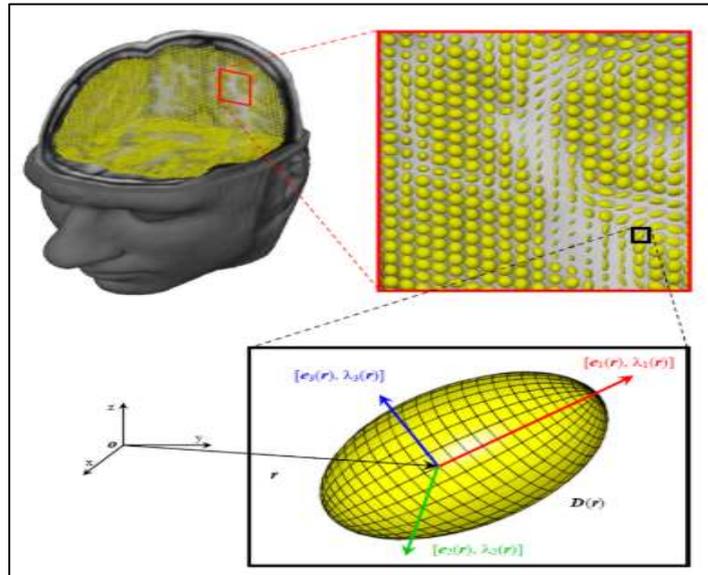


Figure 1.10: Surfaces d'iso probabilité, dérivée du champ du tenseur de diffusion. On peut remarquer que dans chaque voxel, la surface d'iso probabilité est un ellipsoïde, qui est définie de manière unique, par les vecteurs propres, et les valeurs propres des tenseurs [25].

Dans le cas tridimensionnel, et dans un repère cartésien (O, x, y, z), trois cas particuliers du tenseur de diffusion sont distingués, comme le montre la figure 1.11:

- **Cas linéaire :** La diffusion dans ce type de zones, s'effectue uniquement, suivant une seule direction avec $\lambda_1 \geq \lambda_2 \approx \lambda_3$, correspondant à l'élément propre non nul (λ_1).
- **Cas planaire :** On dit que la diffusion est restreinte dans un plan avec $\lambda_1 \approx \lambda_2 \gg \lambda_3$, engendré par les deux éléments propres non nuls (λ_1 et λ_2).

- **Cas sphérique :** C'est le cas d'une diffusion isotrope, correspondant au rang 3 du tenseur de diffusion, avec $\lambda_1 \approx \lambda_2 \approx \lambda_3$.

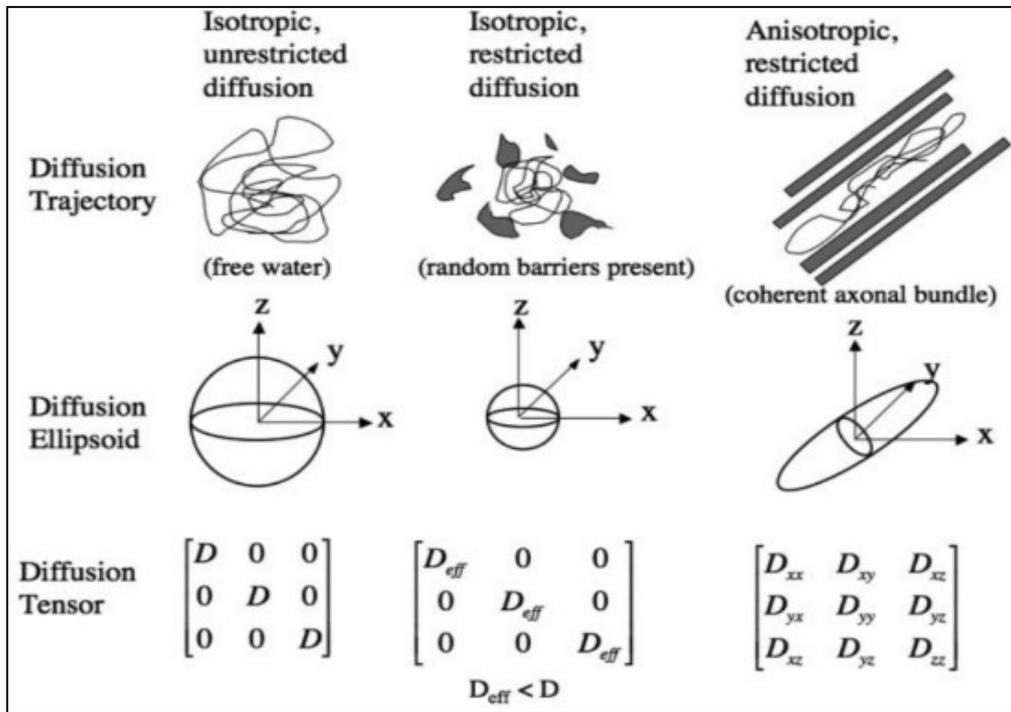


Figure 1.11: Diffusion isotrope et anisotrope. Dans le cas isotrope, la diffusion est similaire dans toutes les directions. Cependant, lorsque le champ est anisotrope, la diffusion est plus importante dans un sens que dans l'autre [26].

1.5.3 Caractéristiques scalaires du tenseur de diffusion

Il existe un certain nombre d'indices scalaires, basées sur les mesures tensorielles, qui résument les attributs de diffusion dans un voxel, qui donnent une vue d'ensemble de la diffusion dans le cerveau.

- **Anisotropie fractionnelle (FA) :** ce paramètre quantifie le degré d'anisotropie du processus de diffusion, et fournit des informations utiles, sur la densité des fibres, et la myélinisation dans la WM, qui varie dans l'intervalle [0, 1].
 - La diffusion est isotrope, si $FA = 0$, avec ($\lambda_1 \approx \lambda_2 \approx \lambda_3$) ;
 - Elle est entièrement anisotrope, suivant une direction, si $FA = 1$, avec ($\lambda_1 > \lambda_2 \approx \lambda_3$). Elle est donnée par la relation (eq 1.9) :

$$FA = \sqrt{\frac{1}{2} \frac{\sqrt{(\lambda_1^2 - \lambda_2^2)^2 + (\lambda_1^2 - \lambda_3^2)^2 + (\lambda_2^2 - \lambda_3^2)^2}}{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}} \quad (1.9)$$

- **La diffusivité moyenne (MD)** : cet indice représente l'amplitude moyenne du déplacement moléculaire par diffusion. Elle correspond à la moyenne de trois valeurs propres. MD élevée, indique une intégrité réduite de la substance blanche, due à une dégradation d'axones ou de myéline. Elle calculée par l'équation suivante (eq 1.10) :

$$MD = \bar{\lambda} = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3} \quad (1.10)$$

- **Mode d'anisotropie (MO)** : cet indice précise la forme de l'anisotropie, il varie entre [-1,1], et donne des informations supplémentaires à FA :
 - L'anisotropie est planaire (mode = -1), si ($\lambda_1 \approx \lambda_2 > \lambda_3$) ;
 - Orthotrope (mode=0) ou linéaire anisotrope (mode = 1), si ($\lambda_1 > \lambda_2 \approx \lambda_3$).

L'indice MO se calcule par (eq 1.10) :

$$MO = \frac{\lambda_1^2 \times \lambda_2^2 \times \lambda_3^2}{\sqrt{(\lambda_1 - \bar{\lambda})^2 + (\lambda_2 - \bar{\lambda})^2 + (\lambda_3 - \bar{\lambda})^2}} \quad (1.11)$$

- **Diffusivité axiale (AD)** : le coefficient de diffusion axiale, est l'amplitude du mouvement de diffusion, parallèle aux faisceaux de fibres axonales. La diffusivité axiale, est d'avantage liée aux caractéristiques intrinsèques des axones, ou aux modifications de l'espace extracellulaire. Une lésion axonale, traduit une diminution de diffusivité axiale. Ce coefficient est défini par la formule suivante : $AD = \lambda_3$
- **Diffusivité radiale (RD)** : La diffusivité radiale capture l'amplitude de la diffusion perpendiculaire aux faisceaux de fibres. Une diffusivité radiale plus élevée, indique une perte de myéline, ou une perte d'axones. Elle est calculée par la formule (eq 1.12) :

$$RD = \frac{\lambda_2 + \lambda_3}{2} \quad (1.12)$$

L'ensemble des cartes obtenues est illustré dans la figure 1.12.

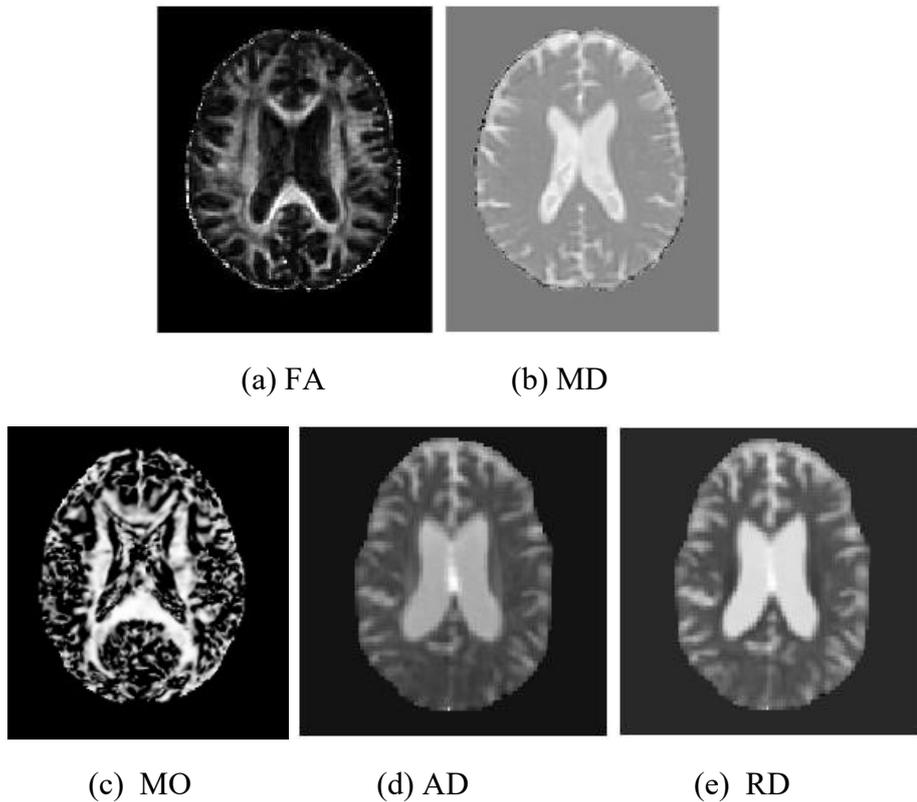


Figure 1.12: Un exemple des surfaces d'iso probabilités, dérivées du champ du tenseur de diffusion.

1.5.4 Caractéristiques vectorielles du tenseur de diffusion

Il est également possible, d'extraire des caractéristiques vectorielles, basées sur la DTI. On choisit généralement, le vecteur propre principal du tenseur D . Parmi les caractéristiques vectorielles les plus utilisées, nous pouvons citer, la carte d'orientation en couleur RGB (Red, Green, Blue). Elle revient à imager la direction principale, au niveau de chaque voxel. La couleur rouge indique que les fibres de ce voxel, sont orientées dans la direction gauche-droite, le bleu indique la direction inférieure-supérieure (bas-haut), et le vert signifie qu'elles sont dirigées dans la direction antéro-postérieure (avant-arrière), comme le montre la figure 1.13.

Les faisceaux de fibres nerveuses de la WM peuvent être visualisés par une reconstruction avec des courbes tridimensionnelles, appelée tractographie [27].

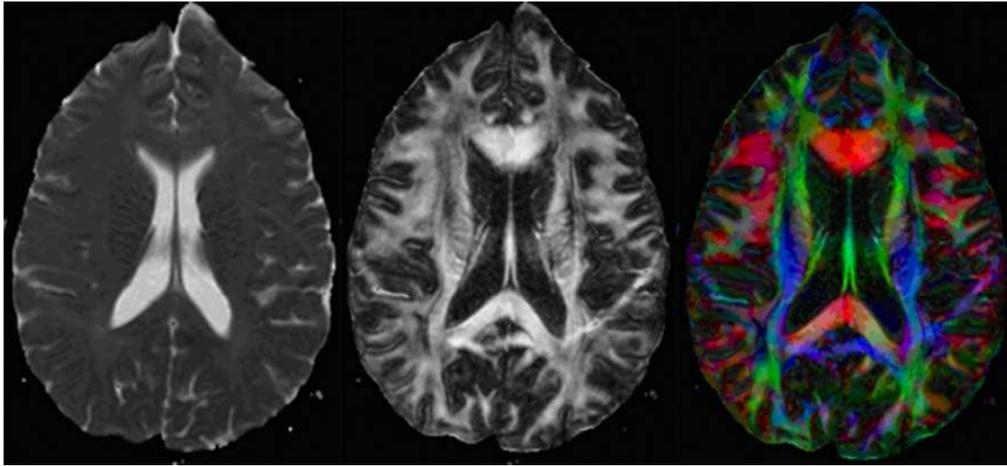


Figure 1.13: Coupes axiales d'un cerveau humain montrant les cartes quantitatives des mesures DTI. De gauche à droite : diffusivité moyenne (MD), anisotropie fractionnelle (FA), direction du vecteur propre majeur, indiqué par la carte des couleurs RVB.

1.6 CONCLUSION

L'imagerie de tenseur de diffusion donne un aperçu, de la microstructure complexes, du système nerveux central, in vivo. Elle donne des indications sur d'éventuelles anomalies des fibres nerveuses de la WM, non visibles en imagerie conventionnelle. L'imagerie tensorielle de diffusion, est actuellement utilisée pour l'étude de pathologies, qui entraînent des modifications de la structure de la WM, et donc des caractéristiques de la diffusion (FA et MD), rendant possible l'identification des perturbations de la WM, causées aussi par des maladies neurodégénératives, telles que la maladie d'Alzheimer.

Dans cette thèse, nous nous concentrerons sur l'utilisation des cartes FA et MD, pour classer les sujets, avec et sans AD. Les mesures structurelles sMRI et de la DTI, promettent d'aider au diagnostic, et au suivi du traitement du MCI et d'AD, en quantifiant les modèles d'atrophie cérébrale structurelle.

CHAPITRE 2 :

DETECTION DE LA MALADIE D'ALZHEIMER PAR MRI : ETAT DE L'ART

2.1 INTRODUCTION

Les progrès de la neuroimagerie ont permis d'extraordinaires avancées, dans le domaine des neurosciences. Grâce à elle, il est possible d'obtenir des images anatomiques et fonctionnelles du cerveau. Elle permet ainsi de mieux connaître le fonctionnement normal ou pathologique du cerveau, et possède de nombreuses applications en neurosciences, telles que l'étude des maladies neurodégénératives, comme la maladie d'Alzheimer.

Dans ce chapitre, on s'intéresse à la contribution de la DTI et de la sMRI, dans le diagnostic de la maladie Alzheimer. Nous décrirons brièvement les mesures, largement utilisées dans l'atrophie cérébrale, et enfin, nous présenterons un bref aperçu des méthodes basées sur ces techniques (DTI/sMRI), pour le diagnostic de la maladie d'Alzheimer.

2.2 LA CONTRIBUTION DE L'IMAGERIE MRI, STRUCTURELLE ET DE DIFFUSION, DANS LE DIAGNOSTIC DE LA MALADIE D'ALZHEIMER :

L'analyse et l'identification, des changements macro et microstructuraux dans le cerveau, à l'aide d'images sMRI et DTI, sont largement utilisées, pour faciliter le diagnostic et le suivi du traitement du MCI et de l'AD. En effet, ces changements induisent une perte cellulaire, qui est spatialement distribuée, dans de nombreuses régions du cerveau. Cette perte cellulaire se manifeste par une atrophie corticale, qui commence au niveau des structures du lobe temporal médian, précisément, le cortex entorhinal et l'hippocampe, et se propage à l'ensemble du cortex, affectant ainsi des structures temporelles latérales et inférieures, et les cingulaires, antérieur et postérieur [3, 28-30]. Les biomarqueurs d'imagerie, les plus utilisés

dans la littérature, pour diagnostiquer et prédire la maladie d'Alzheimer, à un stade précoce, sont :

2.2.1 Epaisseur corticale

Le cortex cérébral est formé de matière grise (GM) périphérique aux hémisphères du cerveau, mesurant généralement 2-3 mm. De nombreuses études ont rapporté, que les patients atteints de l'AD, ou de MCI, ont une atrophie du cortex cérébral [31, 32]. Au fur et à mesure que la maladie progresse, le cortex cérébral se rétrécit dans plusieurs zones, endommageant les zones responsables de la pensée, du jugement, du langage et de la mémoire (Figure 2.1).

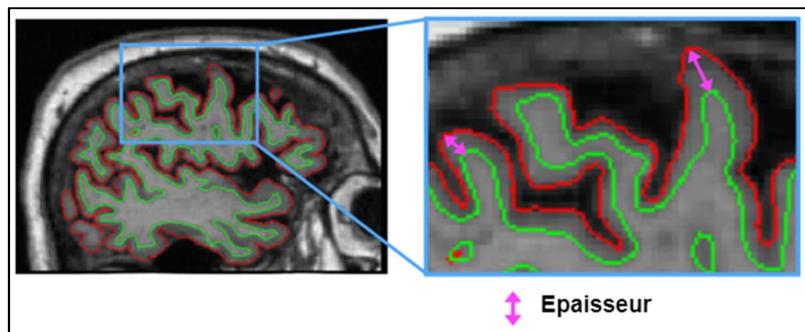


Figure 2.1: Illustration de la mesure de l'épaisseur corticale par imagerie MRI.

2.2.2 Atrophie du volume de l'hippocampe

L'atrophie hippocampique est un bon biomarqueur pour détecter la maladie d'Alzheimer, à un stade précoce, par l'imagerie MRI, qui offre des images de haute résolution (figure 2.2). L'hippocampe est l'indicateur, le plus utilisé par les chercheurs, car c'est la région responsable de la mémoire. Ils ont remarqué une réduction du volume des régions hippocampiques, gauche et droite, chez les patients atteints de l'AD [33, 34], et des troubles cognitifs légers (MCI) [3], par rapport aux personnes cognitivement normales (CN). Il existe une relation entre le taux d'atrophie hippocampique, et les troubles de la mémoire. La mesure de la progression de la perte hippocampique, peut être utilisée comme substitut potentiel, pour la mesure de l'efficacité des interventions thérapeutiques.

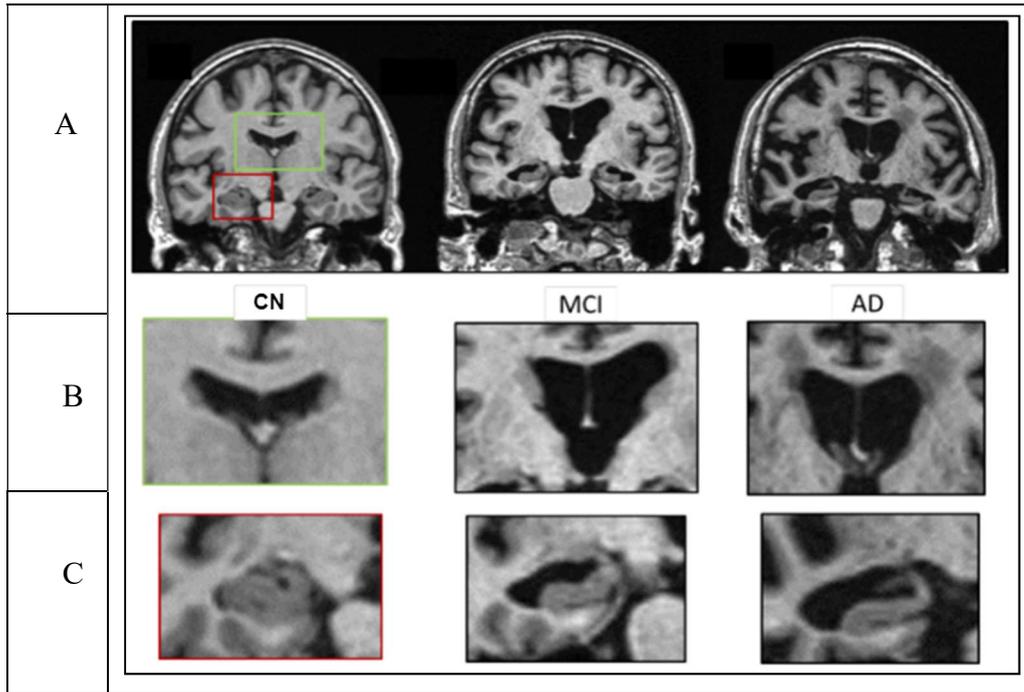


Figure 2.2: Effets de la maladie d'Alzheimer sur le cerveau ; A : sMRI pondérées en T1, d'un sujet âgé, cognitivement normal (CN), d'un sujet souffrant de troubles cognitifs légers (MCI), et d'un sujet atteint de la maladie d'Alzheimer (AD) ; B : Ventricules latéraux agrandissement (vert) ; C : Atrophie progressive de la structure de l'Hippocampe (Rouge).

2.2.3 Cortex entorhinal

Le cortex entorhinal est la région, située à côté de l'hippocampe dans le lobe médial, et responsable aussi, de l'attention, de la mémoire et de la cognition, comme l'hippocampe. En effet, de nombreuses études ont démontré que les structures des lobes temporaux médiaux, sont les premières zones du cerveau, qui subissent des pertes de volume, spécifiquement au niveau de cortex entorhinal et de l'hippocampe [29, 35-37]. Le suivi de la progression de l'AD, par l'imagerie structurale, montre des changements atrophiques du cortex d'entorhinal, dans un premier temps, d'où ils se progressent vers l'hippocampe, puis le système limbique, les lobes temporaux et pariétaux, et finalement les lobes frontaux [38-40]. L'analyse volumétrique de l'imagerie MRI indique que, dans la phase MCI, le cortex entorhinal et l'hippocampe, montrent des réductions de volume, de 20 à 25%, par rapport au contrôle normal (CN).

2.2.4 Élargissement des ventricules

L'élargissement ventriculaire (expansion) est également un biomarqueur visuel, utile pour la détection de la maladie d'Alzheimer (voir figure 2.2). Sur la base d'images MRI, de

nombreuses études, montrent que la maladie d'Alzheimer provoque une hypertrophie (élargissement) des ventricules, dans le cerveau des personnes atteintes, de MCI ou de l'AD, par rapport aux sujets témoins [41, 42]. À un stade avancé de la maladie d'Alzheimer, une expansion plus importante est observée. En plus, une atrophie importante de l'hippocampe, et du cortex entorhinal, est détectable en MRI [43]. Cependant, l'expansion des ventricules latéraux seule, n'est pas une mesure suffisante, pour spécifier un MCI ou un AD. Cela nécessite des informations supplémentaires, sur les changements structurels dans le cerveau, qui pourraient être utilisés pour diagnostiquer et identifier l'AD, en particulier dans les premiers stades de la maladie.

2.2.5 Liquide céphalorachidien

Le liquide céphalo-rachidien est un liquide du système nerveux central, qui entoure le cerveau et la moelle épinière, et qui assure la protection du cerveau, contre les chocs externes. Le LCR remplace le tissu cérébral endommagé, en raison de la dégénérescence des cellules neuronales, ou de la perte de volume d'une région, principalement l'hippocampe, provoqué par la maladie d'Alzheimer. Par conséquent, une augmentation du fluide LCR, est perçue comme un biomarqueur, pour diagnostiquer l'AD, dans plusieurs études [44-46]. Dans les images MRI structurelles, la distribution du fluide de LCR est donnée par des zones sombres.

2.3 APPROCHES POUR L'EVALUATION DE LA MALADIE D'ALZHEIMER

Auparavant, nous avons exposé les biomarqueurs d'imagerie conventionnels de la maladie d'Alzheimer. Cependant, pour analyser et évaluer la progression de la maladie, nous avons besoin de méthodes de mesure, robustes et efficaces, pour suivre les changements structurels, à l'aide de l'imagerie MRI du cerveau. Dans cette section, nous présentons certaines des approches les plus utilisées, pour la classification de l'AD. Ceci permet d'étudier les variations de toutes les régions cérébrales, liées à l'AD, telles que l'épaisseur corticale, l'élargissement du ventricule, et le volume de perte de l'hippocampe.

2.3.1 Méthodes basées sur les volumes

L'extraction de caractéristiques peut être effectuée, soit au niveau du cerveau entier, soit au niveau de certaines régions locales spécifiques. Dans la littérature concernant la détection de l'AD, de nombreuses méthodes, utilisant le cerveau complet, basées sur différentes approches, pour extraire des caractéristiques volumétriques, représentatives et pertinentes pour la maladie de l'AD, sont rapportées. Ces méthodes consistent à impliquer toutes les régions, pour discriminer les patients des stades AD/MCI [47].

2.3.2 Méthodes basées sur les régions d'intérêts (ROI)

Les méthodes de description, basées sur la région d'intérêt, se concentrent sur la mesure des volumes anatomiques, dans des zones prédéfinies du cerveau. Etant donné que, l'AD affecte les régions, liées au système de mémoire, et aux fonctions cognitives, il est intuitif de se concentrer, exclusivement sur les régions identifiées comme biomarqueurs AD ; ils peuvent refléter le stade de la maladie, à savoir le stade précoce. De nombreux travaux ont utilisé des méthodes, basées sur la 'ROI', dans le diagnostic de l'AD, impliquant différentes régions, pour étudier la progression de la maladie, telles que le volume de l'hippocampe [48-50], l'épaisseur corticale [51, 52], et la densité tissulaire [53, 54], dans des régions cérébrales spécifiques, en utilisant l'imagerie MRI. Cependant, la conception d'algorithmes, se basant sur les méthodes ROI, nécessite une connaissance à priori des zones anormales, d'un point de vue structurel ou/et fonctionnel, pour déterminer les régions du cerveau. En revanche, la maladie peut commencer dans certaines régions, à un stade précoce, et peut progresser, pour s'étendre à d'autres régions, au fil du temps. Néanmoins, compte du, ces méthodes sont toujours meilleures que les méthodes prenant en compte le cerveau entier.

2.3.3 Méthodes basées sur les coupes

L'utilisation des méthodes, basées sur les coupes, réduit les données volumétriques, en une représentation bidimensionnelle. Il y a sans doute une perte d'information, puisque la morphologie originelle des tissus, est sous une forme tridimensionnelle. L'avantage des méthodes basées sur des coupes complètes, est qu'elles ne nécessitent pas de segmentation tissulaire ; ils ne prennent qu'une coupe 2D, et ignorent le reste. Cependant, les méthodes de sélection des coupes, diffèrent les unes des autres. De nombreuses études ont utilisé leur méthode spécifique pour extraire des coupes d'images 2D, à partir de scanners cérébraux, tandis que d'autres considèrent des projections standards, de modalités de neuroimagerie, à savoir, les plans axial, coronal et sagittal [50, 55]. Néanmoins, aucune des études de cette catégorie, n'a effectué une analyse complète du cerveau, car une coupe d'image 2D, ne peut pas inclure toutes les informations du scanner cérébral. Dans les travaux d'évaluation de l'AD, le plan axial est le plus largement utilisé, suivi du coronal. Cependant, ce dernier couvre les trois régions cérébrales les plus critiques liées à l'AD (hippocampe, cortex et ventricules) [56]. En effet, de nombreuses études utilisent la coupe axiale pour analyser les régions cérébrales liées à l'AD, principalement les structures de la GM. Par exemple, les auteurs, dans [57], ont utilisé la méthode, basée sur les coupes, sur les volumes GM, en analysant la vue axiale. De la même manière, les auteurs dans [58], ont utilisé des coupes

axiales médianes la MRI. De plus, cette méthode est plus adaptée, pour mesurer les biomarqueurs, d'épaisseur corticale, puisque le cortex cérébral peut être bien représenté.

2.3.4 Méthodes basées sur les voxels

Les méthodes volumétriques sont utilisées, pour extraire des mesures tridimensionnelles (3D), de structures cérébrales spécifiées. La nature 3D des approches, basées sur des volumes 3D, permettra une exploitation complète des informations spatiales, sur les analyses. L'une des approches les plus populaires, est celle nommée « voxel-based morphometry » (VBM) [59]. VBM compare les schémas régionaux du cerveau, entre des groupes de sujets, en effectuant des tests statistiques, sur tous les voxels. Cependant, cela nécessite généralement un recalage spatial (coregistration en anglais), et une normalisation, où les images cérébrales sont recalées, sur un espace commun. La plupart des études, qui impliquent cette méthode, effectuent des recherches sur le cerveau complet, en mode monomodale ou multimodale. Dans d'autres travaux [60], la segmentation tissulaire, basée sur la GM, la ROI est réalisée sur des images MRI, avant l'étape de classification. La mise en œuvre de modèles de classification, à l'aide de la méthode basée sur les voxels, peut nécessiter une technique de réduction de la dimension des caractéristiques, en particulier pour les méthodes, utilisant des caractéristiques visuelles, mais n'est pas nécessairement utile, dans les structures profondes. Néanmoins, diverses méthodes existent, pour surmonter le problème de la dimensionnalité élevée des caractéristiques, lors de l'utilisation de méthodes d'apprentissage automatique classiques [61].

2.4 ALGORITHMES DE CLASSIFICATION AUTOMATIQUE POUR LA MALADIE D'ALZHEIMER

Ces dernières années, un grand nombre de travaux concernant la classification de l'AD, se sont basés sur les méthodes d'apprentissage automatique.

La figure 2.3 montre les phases d'apprentissage automatique, pour la classification. La première étape est la phase d'apprentissage, où l'algorithme d'apprentissage automatique est entraîné, avec les images d'entrée et leurs étiquettes correspondantes. Dans la classification par apprentissage automatique, l'extraction de caractéristiques est une étape essentielle, pour extraire de nouvelles caractéristiques des images. La deuxième étape est la phase de prédiction (test), qui utilise le classificateur, entraîné dans la phase précédente, pour prédire les étiquettes des images de test (images inédites). Dans cette phase, c'est le même algorithme d'extraction de caractéristiques, qui est appliqué aux nouvelles images [62].

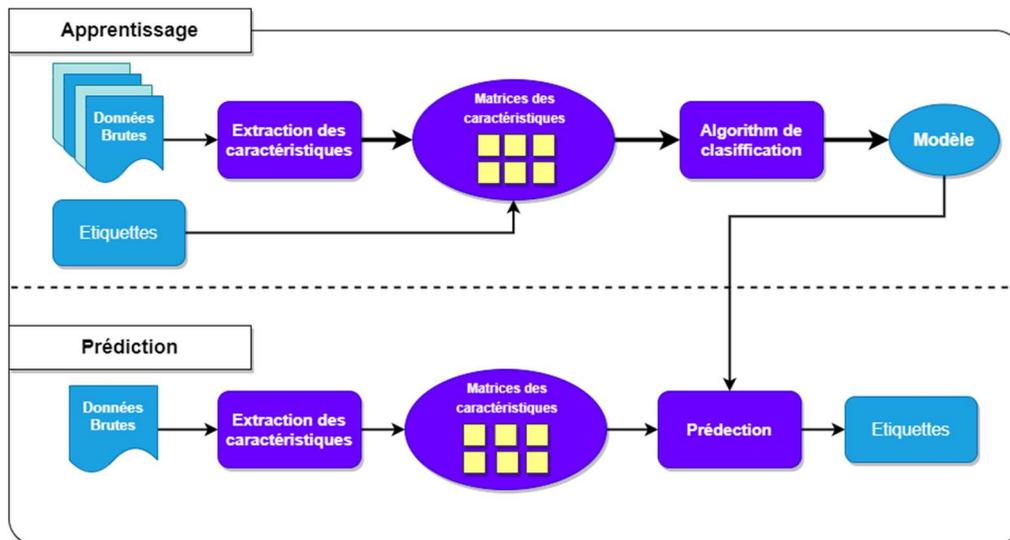


Figure 2.3: Organigramme d'un algorithme d'apprentissage pour la classification des images

Pour les ensembles de données, plus volumineux, les chercheurs utilisent souvent une nouvelle classe d'apprentissage automatique, appelée apprentissage en profondeur [63]. L'apprentissage en profondeur est une classe de réseaux de neurones artificiels (ANN), avec de nombreuses couches de traitement. La principale différence entre les modèles d'apprentissage automatique traditionnels, et le modèle d'apprentissage profond, est l'extraction automatique de caractéristiques, qui pour l'algorithme d'apprentissage en profondeur, joue un rôle important dans l'amélioration de la précision, dans différents problèmes [64]. Une autre caractéristique qui distingue l'apprentissage en profondeur, des autres techniques d'apprentissage automatique, est la profondeur du modèle.

En général, la classification automatique des images cérébrales, de différentes modalités, peut être appliquée à l'ensemble du cerveau [11, 16, 17, 65-67], ou réalisée, en utilisant les connaissances du domaine, sur des régions d'intérêt spécifiques (ROI). Par exemple, l'hippocampe est la région la plus utilisée, dans l'étude de la maladie d'Alzheimer [45, 46]. Suite au progrès de la vision par ordinateur et de la recherche sur la récupération d'images, basées sur le contenu, un certain nombre de travaux ont été développés, qui peuvent être divisés en deux classes :

- **Les méthodes, dites basées sur les caractéristiques** : elles se composent de quatre étapes : l'extraction des caractéristiques, la sélection des caractéristiques, la réduction de la dimensionnalité, et la sélection de l'algorithme de classification, basée sur les caractéristiques. Un ensemble de caractéristiques extraites des scanners cérébraux, ont permis une classification efficace, avec une charge de travail de calcul réduite. Les

vecteurs des caractéristiques obtenus, sont classés à l'aide d'algorithmes d'apprentissage automatique, comme le montre la figure (2.3).

- **Les méthodes dites basées sur le deep learning** : dernièrement, avec le développement des réseaux de neurones, l'approche basée sur les fonctionnalités est devenue moins populaire, et est progressivement remplacée par des réseaux de neurones convolutionnels, d'architectures différentes. Le deep learning est capable d'extraire des caractéristiques pertinentes, et de les classer, sans une intervention humaine, à l'aide des couches de convolution. Une explication détaillée sur les réseaux de neurones et le deep learning est donné dans le chapitre 3.

Nous proposons ici, une revue systématique des travaux sur ce sujet, en tenant compte des travaux, qui utilisent l'approche monomodale 'DTI' ou 'sMRI', et l'approche multimodale. En outre, ces méthodes visent deux niveaux de classification AD : la classification binaire dans laquelle la technique classifie CN par rapport à AD ou MCI, et la multi classification, qui classe la maladie dans un des niveaux (CN, AD, MCI, ...etc.).

2.4.1 Approches monomodales, utilisant l'imagerie MRI structurale :

- Ding et al [68] suivent le pipeline de prétraitement d'image standard de la base de données Alzheimer's Disease Neuroimaging Initiative (ADNI). Le point fort de ce travail, est l'utilisation de la matrice de cooccurrence en niveaux de gris, et du filtre de Gabor, pour extraire à la fois les caractéristiques morphométriques, et les caractéristiques de texture. Les caractéristiques combinées, fournissent un meilleur résultat, avec une augmentation de 6 % de la précision de la classification, par rapport à celles des deux méthodes, lorsqu'elles sont mises en œuvre séparément. Ces travaux sont destinés uniquement, à la classification d'AD et CN.
- Un autre descripteur de caractéristiques locales prometteur, connu sous le nom de Scale-invariant feature transform (SIFT), est présenté dans [69]. L'invariance d'échelle et de rotation, sont les principales caractéristiques de SIFT. Les caractéristiques SIFT, sont extraits efficacement du domaine fréquentiel, ainsi que du domaine spatial. SIFT est justement mis en œuvre, pour le diagnostic précoce de l'AD dans [70, 71].
- Ben Ahmed et al., [45] Les auteurs ont extrait automatiquement les informations visuelles des images, en utilisant l'approximation du signal d'image par les fonctions harmoniques circulaires (CHF). Pour représenter efficacement les informations de contenu extraites de la ROI de l'hippocampe, ils ont utilisé l'approche Bag-of-Visual-Words (BoVW) [72]. Ils ont proposé un schéma de fusion tardive, où les sorties

probabilistes des classifications sur les caractéristiques locales et la quantité de LCR sont fusionnées pour effectuer la classification finale de l'AD.

- Mathieu et al. [73] ont utilisé un sous-ensemble de données pour des expériences ayant 151 patients, dont 71 témoins normaux, et 87 patients AD. L'extraction des caractéristiques est réalisée par une analyse en composantes principales (PCA : Principal Component Analysis), et la transformée en ondelettes discrètes (DWT : Discrete Wavelet Transform). Les caractéristiques extraites, sont ainsi transmises à la SVM, pour la classification. Ils ont atteint une précision de 84 % et 91 %, pour AD/ CN et MCI/ CN, respectivement.
- Hosseini et al. [74] ont proposé un 3D-CNN, entraîné sur des données, contenant 70 AD, 70 MCI et 70 CN, obtenues à partir de la base de données ADNI. Les caractéristiques locales sont extraites de l'image d'entrée 3D, à l'aide de CAE (Convolutional Auto Encoder). Le modèle est entraîné sur l'ensemble des données, contenant des images MRI, pondérées en T1, d'AD, de CN et de MCI. L'isolation du cerveau, et la normalisation spatiale, sont effectuées en tant que pré-traitement. Les caractéristiques extraites sont utilisées comme biomarqueur, pour détecter la MA sur un ensemble de données, en utilisant une approche fine tuning. La classification est effectuée à l'aide d'une technique de validation croisée, de 10 blocs, et a atteint 97,6 % de précision, pour AD vs. CN.
- Dans [66], les auteurs ont proposé l'architecture DemNet, qui est une version modifiée du CNN à 16 couches, VGG16. La contribution de leur travail est, qu'ils ont entraîné leur réseau, sur 20 coupes MRI sélectionnées au hasard, de chaque sujet, et ont pu surpasser plusieurs classificateurs.
- Les auteurs dans [75] ont fait un travail similaire, mais ont sélectionné des coupes, basées sur l'entropie, plutôt que de sélectionner des données par intuition. Ensuite, ils ont utilisé l'apprentissage par transfert sur deux architectures : VGG16 et Inception V4, et ont comparé les performances des deux architectures. Ils ont pu atteindre une précision de 96,25 %, lors de la classification binaire, par l'architecture Inception V4, sur l'ensemble de données OASIS.
- Glzman et al. [76] ont utilisé une approche d'apprentissage par transfert, où ils ont affiné l'architecture AlexNet pré entraîné, et ont également effectué une augmentation des données, telles que la transformation en miroir, sur les données collectées, à partir de la base de données ADNI, pour augmenter la quantité de données, pour l'apprentissage.

- Dans [41], Aderghal et al., ont utilisé 3 coupes centrales, dans chaque projection (axiale, sagittale et coronale), d'une ROI hippocampique. L'architecture du réseau représentait trois réseaux convolutifs 2D (un réseau par projection), réunis dans la dernière couche entièrement connectée. La précision atteinte pour la classification AD / CN, AD / MCI, et MCI / CN, est de 91.4%, 69.5% et 65.6%, respectivement, sur un sous-ensemble de la base de données ADNI.
- Dans [77], les auteurs ont développé une approche d'apprentissage profond, basée sur les CNN, pour détecter les sujets AD. Le CNN proposé, est constitué de huit couches, qui comprennent des couches « Leaky Rectified Linear unit », et qui utilisent le max pooling. L'approche a atteint une précision de 97,65 %, en classification binaire (AD vs. CN).
- Dans [78], les ondelettes, le filtre de Gabor, et la gaussienne des descripteurs locaux, sont utilisés comme outils, pour l'extraction de caractéristiques, avec un classifieur SVM . Le système combine trois SVM différentes, qui sont formées par différents descripteurs de caractéristiques.
- Les auteurs dans [79], ont proposé un apprentissage par transfert. Ils ont utilisé le VGG-16 pré-entraîné, pour classer AD/CN/MCI. La précision obtenue est de 95,73 %.
- Lee et al. [80] ont proposé un schéma de permutation de données, pour obtenir les trois plans. Puis, ils ont sélectionné, pour chaque sujet, et sur chaque plan, les 10 coupes ayant l'entropie la plus élevée. Ces coupes sont combinées, avant d'être transmises à AlexNet-CNN. Leurs résultats expérimentaux ont montré que, leur schéma de permutation des données, améliorerait la précision globale de la classification. Les précisions de la classification AD/CN et AD/CN/MCI, sont de 98,74 %, et de 98.06%, respectivement, sur l'ensemble des données ADNI.
- Maqsood et al. [81] ont appliqué la technique du « transfert learning », sur les trois dernières couches, du modèle pré-entraîné AlexNet, pour effectuer une classification binaire, et multi-classes. Le modèle est entraîné, avec les images pondérées en T1 non segmentées, puis sur les images segmentées de la GM, la WM, et le LCR, issues de la base de données OASIS. Les meilleurs résultats ont été obtenus, par le modèle entraîné avec les images non segmentées.
- Jiang et al.,[82] ont suggéré un système de détection précoce de l'AD. Le modèle pré-entraîné VGG16, est utilisé pour l'extraction des caractéristiques. L'algorithme LASSO est utilisé, pour la sélection des caractéristiques, et la SVM pour classer les

EMCI (Early MCI) des CN. L'étude est réalisée sur la base de données ADNI, avec 70 EMCI et 50 CN, et donne des résultats intéressants, avec une précision de 89.4%.

- Puente-Castro et al., [83] ont utilisé le modèle ResNet18, comme un extracteur de caractéristiques, et SVM comme classifieur. L'étude est réalisée, sur les images 2D du plan sagittal, et comparée à celle du plan axial, des deux bases de données ADNI et OASIS. La contribution principale des auteurs, est l'ajout de l'information d'âge, et de sexe, au vecteur de caractéristiques.
- Feng et al.,[84] ont effectué une comparaison, entre une architecture CNN 2D et 3D, avec une couche Softmax, puis un classifieur SVM. L'architecture comprend 4 couches de convolutions. Le modèle 3D-CNN-SVM a donné les meilleurs résultats, comparant avec les modèles 2DCNN, 2DCNN-SVM, et 3DCNN, en utilisant la base de données ADNI.
- Bansal et al., [85] ont détecté, et extrait les caractéristiques, avec l'approche « Speed Up Robust Features (SURF) ». Ensuite, ils ont construit un dictionnaire visuel, pour entraîner le classifieur SVM.
- Nawaz et al., [86] ont proposé trois modèles, et les ont comparés, pour déterminer lequel d'entre eux, atteint une grande précision. Dans le premier modèle, les images sont prétraitées, pour extraire des caractéristiques. Ils effectuent ensuite une classification, à l'aide de SVM, du k-plus proche voisin, et de la forêt aléatoire. Dans le deuxième modèle, ils utilisent l'apprentissage en profondeur CNN. Dans le troisième modèle, AlexNet-CNN, est utilisé, pour extraire des caractéristiques profondes, et leur classification est faite, à l'aide de SVM, le k-plus proche voisin, et la forêt aléatoire. En comparant les trois modèles, le modèle basé sur les caractéristiques profondes, a obtenu la meilleure précision, avec le classifieur SVM, à savoir 99,21%.
- Dans une autre étude, Naz et al., [87] ont utilisé les caractéristiques profondes, extraites par des modèles CNN pré-entraînés, par exemple, GoogLeNet, AlexNet, DenseNet, ResNet, Inceptionv3, InceptionResNet, et VGG (16 et 19). La classification se fait par le classifieur SVM. Les expériences ont été réalisées, à l'aide de la base de données ADNI. Les meilleurs résultats ont été obtenus, par le modèle VGG+SVM, avec une précision de 99.27%, pour (MCI/AD), 98.89% pour (AD/CN), et 97.06% pour (MCI/CN).
- Serkan et al., [88] ont proposé une approche de classification multiple (AD/CN/MCI), à partir de MRI cérébrales. Différents modèles, utilisant l'architecture CNN, comme

AlexNet, ZFNet, ResNet 50-101-152-50V2-101V2-152V2, et VGG 16-19, ont utilisés pour classer 2182 images, obtenues à partir de la base de données ADNI. Le modèle EfficientNetB0, a fourni la plus grande précision, avec un taux de précision de 92,98 %.

2.4.2 Approches monomodales utilisant l'imagerie de diffusion

Plusieurs études ont appliqué la technique DTI, pour la caractérisation de l'intégrité de la WM, dans l'AD. En particulier, des études basées, sur la DTI, ont montré que les patients AD, présentent des valeurs FA et MD aberrantes, dans la substance blanche de régions cérébrales spécifiques [89]. De plus, d'autres études ont trouvé des changements similaires, mais moins sévères, de ces valeurs, chez les patients MCI [90]. En particulier, des études, basées sur les voxels, ont montré que les sujets AD et MCI, ont une anisotropie fractionnelle (FA) réduite, dans plusieurs régions WM postérieures [8] et une diffusivité moyenne (MD) accrue, dans le cortex occipital-pariétal postérieur, et le gyrus supramarginal pariétal droit [91]. Les études basées sur la ROI, ont démontré une MD plus élevée et/ou une FA plus faible, dans l'hippocampe [92, 93] et le cingulaire postérieur [94, 95]. Notamment, les résultats d'une étude précédente ont montré que, les mesures de diffusivité extraites de l'hippocampe, sont de meilleurs prédicteurs de la conversion du MCI en AD, que son volume [95]. Dans l'ensemble, ces résultats suggèrent que les biomarqueurs obtenus à partir de la technique DTI, peuvent être utilisés, pour la classification de l'AD, grâce à des méthodes de classification avancées [96].

- Grana et al., [97] ont entraîné un SVM, à l'aide de mesures DTI, pour classer les patients AD et CN. Les images des scans DTI, ont été prétraitées, afin d'extraire FA et MD. Différentes méthodes, de validation croisée ont été employées, et la prédiction la plus précise, est obtenue par la méthode d'exclusion. Avec les caractéristiques FA, une précision, une sensibilité, et une spécificité de 100 %, ont été obtenues, tandis qu'avec les caractéristiques MD, ces indices ont atteint des valeurs inférieures.
- Patil et al., [98] ont déterminé certaines régions de la substance blanche, comme des régions d'intérêt. La classification, entre AD et CN, est effectuée, par l'algorithme Adaptive Boosting (AdaBoost). En considérant les mesures FA, et un ensemble de 10 caractéristiques, sélectionnées par un algorithme génétique, les scores de la précision, de la sensibilité, et de la spécificité étaient, respectivement, de 84,5 %, 80,2 % et 85,2 %. Si l'ensemble des caractéristiques n'est pas réduit, ces valeurs diminuent (la précision = 75,3 %, prouvant ainsi, que la réduction des caractéristiques améliore la

précision de la classification, en supprimant la redondance. On peut remarquer que, en prenant MD à la place de FA, aucun changement significatif dans la précision n'a été observé, ce qui suggère que FA est un paramètre efficace, pour la classification AD/CN.

- Dyrba et al., [99] emploient une approche d'apprentissage automatique, pour faire la distinction entre AD et CN, à l'aide des valeurs d'anisotropie FA, comme caractéristiques d'entrée, atteignant 80 % de précision.
- Ben Ahmed et al., [100] ont utilisé les informations de diffusion du motif visuel, basées sur les cartes MD, pour la classification binaire, à l'aide de classifieur SVM. Ils ont atteint une précision de 86,73 % (AD vs CN), 77 % (MCI vs CN), et 73,1 % (AD vs MCI)
- Lee et al., [101] ont extrait le squelette de la WM, des cartes FA et MO, à l'aide de l'algorithme Tract-Based Spatial Statistics (TBSS) (Figure 2.4). L'algorithme de sélection des caractéristiques, relief est utilisé, pour éliminer les voxels non discriminatoires, afin de les classifier par le SVM.

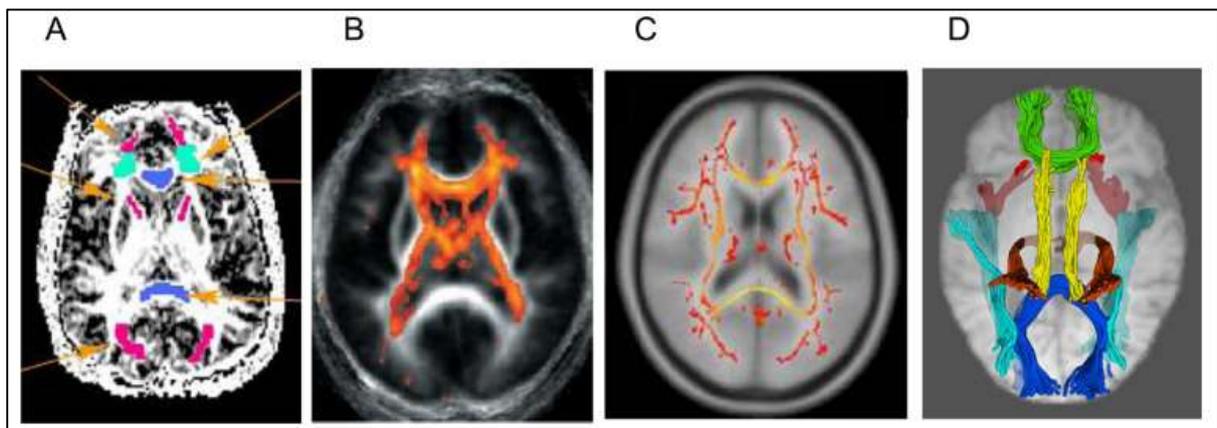


Figure 2.4: Méthode de représentation des données DTI :

A = régions d'intérêt (en couleur), placées directement sur l'image DTI ;

B = morphométrie, à base de voxels (VBM) ;

C = squelette moyen des voies de la matière blanche, à partir des statistiques spatiales, basées sur les voies (TBSS) ;

D = suivi des fibres des voies de la substance blanche [102].

- Dans des recherches similaires, Maggipinto et al. [103] ont proposé une approche de classification, à partir de l'imagerie DTI. Les cartes FA et MD, sont extraites, puis appliquées comme entrées, dans le modèle d'analyse TBSS. Pour sélectionner les

caractéristiques les plus discriminantes, les auteurs ont utilisé le test de la somme des rangs de Wilcoxon, et l'algorithme Relief. La classification est faite, par la forêt aléatoire.

- Eldeeb et al., [104] ont proposé une nouvelle méthode, pour extraire les marqueurs pertinents, associés à FA et MD. Après le prétraitement des données DTI, les cartes FA et MD des ROI, sont déterminées, à l'aide d'un modèle Bag of Words. Ce modèle est utilisé, pour modéliser les modèles de cartes de diffusivité de l'hippocampe, en regroupant les caractéristiques extraites de l'hippocampe, où le nombre de caractéristiques change d'une coupe à l'autre. Les caractéristiques SURF et les caractéristiques SIFT sont extraits. Avec ces cartes FA et MD, un SVM est ensuite entraîné, pour classer les différents groupes de sujets. La classification est effectuée, pour chaque paire de groupes, puis entre toutes les classes, en résolvant un problème multi classe. Les meilleures précisions sont obtenues, avec la carte MD, en utilisant un descripteur de caractéristiques SIFT, et sont rapportées comme suit : 98,3 % AD/CN, 93,6 % MCI/CN, 92,0 % AD/MCI et 89,0 % multi classe.
- Dalboni da Rocha et al., [105] ont classé AD, MCI et CN, via un SVM, appliqué aux cartes FA, en se concentrant sur les zones cérébrales, fréquemment associées aux anomalies AD. L'analyse est répétée pour le cerveau entier, et dans des zones cérébrales spécifiques (ROI), avec et sans étape de sélection des caractéristiques. Comme prévu, les résultats obtenus, sans sélection des caractéristiques, étaient inférieurs. Parmi toutes les zones cérébrales considérées, deux d'entre elles ont montré un plus grand pouvoir discriminant (FA systématiquement plus faible), entre AD et CN : le cingulum bilatéral et le gyrus parahippocampique.
- Lella et al., [106] ont développé une approche d'apprentissage automatique d'ensemble, utilisant les cartes de mesures de diffusion, FA, MD, LD, et RD. L'algorithme TBSS est utilisé, pour extraire l'information squelettique de chaque carte, pour classer CN, par rapport à AD. Ces informations sont utilisées, pour entraîner trois classificateurs : SVM, RF, et perceptron multicouches (MLP). Les meilleurs classificateurs pour chaque carte, sont combinés, à l'aide d'un vote majoritaire.

Habituellement, la matière grise est fragmentée, en un ensemble de régions anatomiques, et la connectivité entre deux régions données, est calculée, sur la base du tractogramme. À cette fin, différentes mesures sont utilisées, notamment le nombre de fibres, ou le FA moyen, le

long de la connexion. Ces caractéristiques caractérisent, la topologie locale et globale du réseau de connectivité, et sont transmises à un classifieur (voir figure 2.5).

- Amoroso et al., [107], ont utilisé des caractéristiques basées, sur les réseaux de connectivité dérivés de la DTI, pour la classification de l'AD. L'étude est réalisée sur 99 sujets, obtenus à partir de la base de données ADNI, comprenant les deux stades, CN et AD. Une tractographie probabiliste est réalisée, pour obtenir la matrice de connectivité de chaque sujet.
- Ebadi et ces collègues, [108] ont acquis des images DTI, de 15 patients AD, 15 patients MCI, et 15 volontaires sains. Après les étapes de prétraitement, la carte FA est calculée. Les auteurs ont considéré 41 zones Brodmann, puis ont calculé les matrices de connectivité pour ces zones, puis ont généré un graphique de connectivité. Le graphique est décrit, avec le vecteur de caractéristiques, puis classé avec l'ensemble des classificateurs (régression logistique, forêt aléatoire, Bayes natif gaussien, SVM), en utilisant une validation croisée 5 fois. Les auteurs ont atteint les niveaux de précision de 80%, 83.3%, 70%, pour AD/CM, AD/MCI, et MCI/CN, sur leur base de données personnalisée.

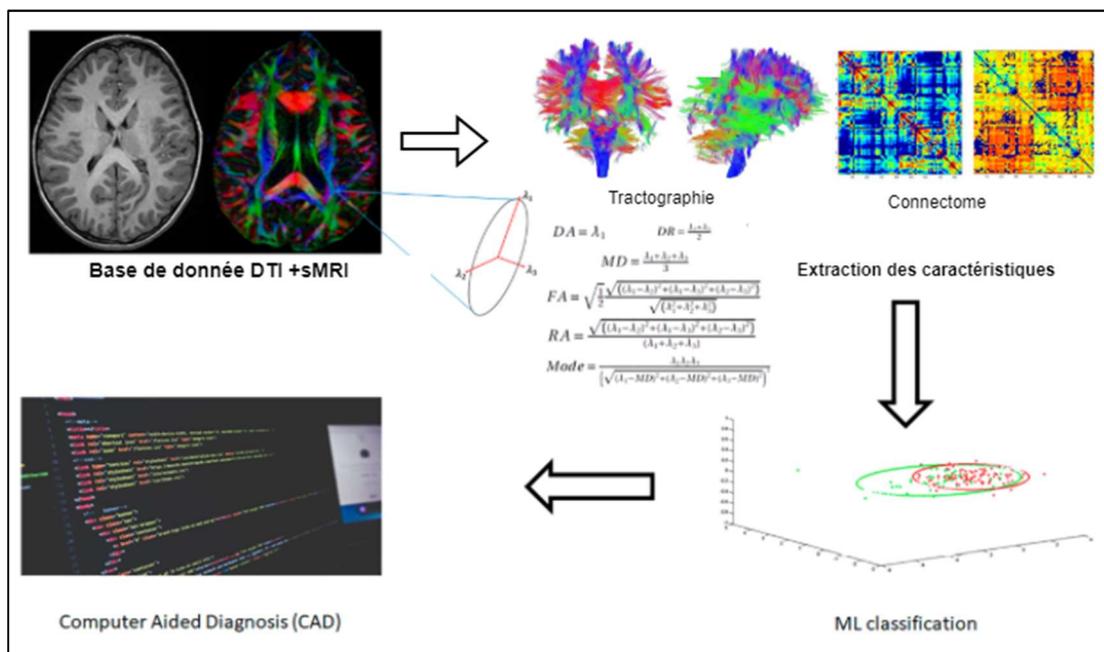


Figure 2.5: Procédure générale d'un système d'aide au diagnostic en utilisant la DTI, consistant en quatre étapes : prise d'un ensemble de données d'imagerie par DTI, ou d'images multimodales, extraction de caractéristiques, à partir de l'ensemble de données, classification par apprentissage automatique en fonction des caractéristiques les plus significatives,

diagnostic automatisé, obtenu en classant les scans individuels, dans une classe clinique spécifique [11].

2.4.3 Approche multimodale

L'approche multimodale a été largement utilisée récemment. Les chercheurs essaient de combiner les différentes modalités, principalement sMRI + DTI ou sMRI + PET.

Des études [19, 46, 55], basées sur ML et CNN utilisant le multimodal, ont montré leur capacité et leur efficacité, à atténuer les limites de l'utilisation, des données unimodales, et ainsi à améliorer la précision des modèles de prédiction, dans le diagnostic de la maladie [25, 26]. La plupart des études se sont concentrées sur la combinaison des cartes DTI et de la sMRI, dans le dépistage de l'AD.

- Marzban et al., [109] ont utilisé des mesures dérivées du DTI (FA, MD et le mode de diffusivité (MO)), et les cartes GM segmentées, à partir de sMRI, pondérées en T1. L'étude est développée sur les régions de l'hippocampe, et du cortex entorhinal, extraites des données de prétraitement. Les auteurs ont entraîné le CNN, en utilisant FA, MO, MD, et l'hippocampe segmente de la GM séparément, puis en cascade comme entrée. Le CNN proposé, utilise une seule couche de convolution 2D. Les volumes MD et GM en cascade en tant qu'entrées, ont atteint la meilleure précision globale à savoir, 88,9 % et 79,6 %, respectivement, pour AD/CN et MCI/CN.
- Ahmed et al., [46] ont calculé les caractéristiques locales de la région d'intérêt hippocampe (hippocampe-ROI), segmentée à partir des volumes de MD et T1w, en utilisant les fonctions harmoniques de Gauss-Laguerre (GL-CHF). L'apprentissage multiple du noyau (MKL), est utilisé, pour combiner ces caractéristiques avec la quantité LCR. Ils ont rapporté, dans leur précédente étude [45], appliquée aux images sMRI, que les GL-CHF sont plus efficaces, que les descripteurs locaux Scale-Invariant Feature Transform (SIFT), et Speeded-Up Robust Features (SURF), contrairement à Bansal et al.[110].
- Aderghal et al., [111] ont proposé un CNN, de type LeNet, basé sur des images sMRI et DTI-MD. Ils ont sélectionné la coupe médiane de l'hippocampe, et ses deux voisines, dans chaque projection (axiale, sagittale et coronale). Le CNN proposé est entraîné, sur la base de données MNIST. Ensuite, il est d'abord entraîné sur sMRI, puis sur les images DTI-MD. Ils ont atteint une précision de classification de 86,83 % pour AD vs CN, 69,85 % pour MCI vs CN, et 71,75 % pour AD vs MCI.

- Kang et al., [112] ont combiné les images FA, MD, et T1w. Ensuite, l'extraction de caractéristiques profondes, est effectuée, via le réseau VGG16, et la classification des MCI des premiers stades (EMCI) du CN, est faite, à l'aide d'un SVM.
- Massalimova et al., [55] utilisent trois réseaux Resnet-18, en cascade, avec trois entrées (FA, MD et T1w), pour classer AD vs CN vs MCI. Une précision de 96,00 % est obtenue, en utilisant la base de données OASIS.
- Perez-Gonzalez et al. [113] ont calculé les caractéristiques pertinentes, à partir de sMRI (T1w), et du DTI, et les ont combinées avec des scores neuropsychologiques, avant de les transmettre au classificateur RF. Ils ont atteint une précision de 88,8 %, en utilisant les caractéristiques de l'hippocampe, du thalamus et de l'amygdale, pour les hémisphères droit et gauche, afin de distinguer MCI et CN.
- Fang et al. [114] ont proposé une approche d'apprentissage par re-transfert, basée sur le modèle AlexNet, pour classer AD vs EMCI, AD vs LMCI, AD vs CN, et EMCI vs CN. AlexNet est entraîné avec les images sMRI, et les poids du modèle obtenu, sont ensuite utilisés, pour entraîner les images DTI (FA et MD).
- Agostinho et al., [65] ont comparé les performances de l'association sMRI+DTI, et sMRI+Tomographie par Emission de Positrons (TEP). Ils ont utilisé divers atlas de marquage anatomique, pour diviser le cerveau en ROI, puis ont extrait différentes caractéristiques. Après cela, SVM a été utilisée, pour les classer CN vs. AD. Les résultats des expériences ont atteint une précision de 97,00 % et de 98,00 %, en utilisant sMRI + DTI, et sMRI+ TEP, respectivement.

Un récapitulatif des travaux antérieurs est présenté dans le tableau 2.1.

Tableau 0.1: Résumé des différentes approches de détections de la maladie d'Alzheimer.

Auteur	Année	Modalité	Méthode	Base de données
Daliri et al. [70]	2012	sMRI	SIFT	ADNI
Lee et al. [115]	2015	DTI	TBSS+ SVM	LONI
Ding et al. [68]	2015	sMRI	Matrice de cooccurrence en niveaux de gris (GLMC), et du filtre de Gabor	ADNI
Ben Ahmed et al., [45]	2015	sMRI	Les fonctions harmoniques circulaires (CHF), et la quantité de pixels CSF+ SVM	ADNI

Ebadi et al. [108]	2017	DTI	Graphe de connectivité+ ensemble de classificateurs (régression logistique, forêt aléatoire, Bayes natif gaussien, SVM)	Base de données personnalisée
Maggipinto et al. [116]	2017	DTI	TBSS+ forêt aléatoire	ADNI
Ben Ahmed et al. [46]	2017	sMRI+ DTI	Fonctions harmoniques de Gauss-Laguerre (GL-CHF) +MKL	ADNI
Hosseini et al. [74]	2018	sMRI	3DCNN	ADNI
Eldeeb et al. [104]	2018	DTI	SURF/ SIFT+SVM	ADNI
Maqsood et al. [81]	2019	sMRI	AlexNet	ADNI
Marzban et al. [109]	2020	sMRI+ DTI	2DCNN	ADNI
Jiang et al. [117]	2020	sMRI	VGG16+SVM	ADNI
Puente-Castro et al. [83]	2020	sMRI	ResNet-18+SVM	ADNI/ OASIS
Bansal et al. [85]	2020	sMRI	SURF+SVM	ADNI
Feng et al. [118]	2020	sMRI	2D CNN+SVM/ 3D CNN+SVM	ADNI
Aderghal et al. [111]	2020	sMRI+DTI	CNN (LeNet5)	ADNI
Kang et al. [112]	2020	sMRI+DTI	VGG-16+SVM	ADNI
Naz et al. [87]	2021	sMRI	GoogLeNet, lexNet, DenseNet, ResNet, Inceptionv3,	ADNI

			InceptionResNet et VGG (16 et 19) et classifieur par le SVM	
Lella et al. [119]	2021	DTI	TBSS+ensemble learning	ADNI
Massalimova et al.[55]	2021	sMRI+DTI	ResNet-18	OASIS
Serkan et al. [88]	2022	sMRI	AlexNet, ZFNet, ResNet 50-101-152-50V2-101V2-152V2, VGG 16-19	ADNI
Fang et al. [114]	2022	sMRI+ DTI (FA, MD)	AlexNet	ADNI
Agostinho et al. [65]	2022	sMRI+ DTI/ sMRI+TEP	ROI, basée sur atlas+SVM	ADNI

2.5 DISCUSSION

Dans cet état de l'art, nous avons examiné les méthodes existantes, dans la détection de la maladie d'Alzheimer. Ces études peuvent être classées selon les critères suivants :

- **Modalité étudiée :** alors que la majorité des chercheurs, n'utilisait que la sMRI structurale, d'autres utilisent la DTI, et certains utilisent des données multimodales (combinant la DTI avec la sMRI (T1w)).
- **Type de caractéristiques :** Nous avons séparé, entre les articles, utilisant des métriques de caractéristiques DTI, telles que l'anisotropie fractionnaire (FA) et la diffusivité moyenne (MD), et ceux utilisant des caractéristiques plus avancées, telles que les caractéristiques, extraites par la TBSS, ou basé sur le réseau de fibres (matrice de connectivité). La plupart des études sMRI utilisent des CNN, que ce soit, pour extraire les caractéristiques, ou pour la classification directe.
- **Classifieurs :** Les plus couramment utilisés, sont les machines à vecteurs de support (SVM), mais les forêts aléatoires (RF), la régression logistique (LR), les k plus proches voisins (KNN), ou les Bayes naïfs (NB) sont également utilisées dans certaines études. Dans d'autres études, les CNN sont les plus utilisés.

- **Ensemble de données :** le jeu de données le plus couramment utilisé, est celui de la base ADNI. La plupart des articles utilisent cette base de données, avec trois stades AD, CN et MCI.
- **Tâche de classification :** certaines études se sont concentrées sur la discrimination, entre les patients AD et les sujets sains (CN), tandis que d'autres ont abordé la classification des patients, atteints de MCI (trouble cognitif léger), ou la prédiction de la progression vers l'AD, chez les patients MCI. Il y a plus de travaux, qui visent la classification binaire, que les travaux qui visent la multi classification (AD/CN/MCI).

CHAPITRE 3 :

METHODES D'APPRENTISSAGE EN PROFONDEUR POUR LA CLASSIFICATION D'OBJETS

3.1 INTRODUCTION

Ces dernières années, l'intelligence artificielle est de plus en plus utilisée dans divers domaines, tel que la reconnaissance d'images, plus particulièrement, l'apprentissage automatique (*machine learning* en anglais), qui fait partie de l'intelligence artificielle, et son sous domaine, nommé apprentissage profond (Deep Learning en anglais). La figure 3.1 explique la relation entre ces trois domaines.

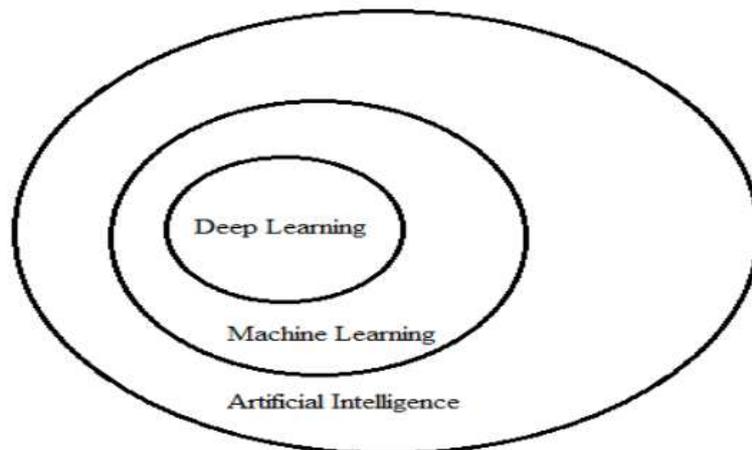


Figure 3.1: Représentation graphique de la relation entre l'intelligence artificielle, l'apprentissage automatique (Machine Learning) et l'apprentissage profond (Deep Learning)[120].

Dans ce chapitre, nous présentons les concepts fondamentaux de l'apprentissage profond, en particulier, l'architecture des réseaux de neurones convolutionnels (Convolutional Neural Network (CNN)), puisqu'ils sont les plus utilisés, dans la classification des stades de la

maladie d'Alzheimer, comme nous l'avons vu dans le chapitre précédent. Nous commencerons par donner un aperçu sur les trois formes de l'apprentissage automatique : supervisés, semi supervisés et non supervisés. Nous présenterons ensuite, brièvement, les réseaux de neurones traditionnels, du perceptron au perceptron multi-couches, puis nous introduirons l'élément principal, utilisé dans les réseaux de neurones profonds : l'approche CNN. Nous verrons ces différents composants, et leurs paramètres, les fonctions d'activation les plus connues, ainsi que les méthodes d'optimisation et de régularisations, et nous décrirons les modèles CNN, populaires dans la littérature, pour conclure ce chapitre.

3.2 APPRENTISSAGE AUTOMATIQUE

Ce travail se concentre sur l'extraction de caractéristiques, à partir d'images, directement avec l'apprentissage automatique, plutôt que de s'appuyer sur des extracteurs de caractéristiques classiques. L'apprentissage automatique est une science, visant à amener les machines, à apprendre des solutions à des problèmes spécifiques, sans être explicitement programmées pour le faire [121, 122]. De nombreuses solutions, d'apprentissage automatique, sont dérivées de notre connaissance, du cerveau humain.

L'apprentissage automatique est un processus, utilisé pour créer des modèles, capables d'extraire des informations, à partir de données, pour résoudre un problème donné, et par conséquent, améliore automatiquement leurs performances. Une perspective intéressante, qui peut être utilisée, est de considérer l'apprentissage automatique, comme une forme de compression d'informations, où le modèle d'apprentissage automatique, essaie d'extraire des informations, à partir des données d'entrée, de manière à ce que la quantité de données utilisées pour la classification soit réduite, tandis que les informations qu'elles contiennent soient préservées. Il existe trois types d'apprentissage automatique :

3.2.1 Apprentissage supervisé :

L'apprentissage supervisé s'intéresse aux données étiquetées. L'objectif est de prédire l'étiquette (inconnue), associée à une nouvelle observation, à partir de la connaissance, fournie par les N observations étiquetées, du jeu de données. En effet, la plupart des algorithmes d'apprentissage, adaptent le modèle, de manière à optimiser les performances de prédiction, sur le jeu de données, dont ils disposent, appelé base d'apprentissage ou d'entraînement. L'apprentissage est basé sur la mesure de l'écart, entre les « vraies » étiquettes, et les étiquettes prédites.

3.2.2 Apprentissage non supervisé :

Le modèle apprend uniquement, à partir des données d'entrée. L'objectif est d'identifier automatiquement, des caractéristiques communes aux observations. Cette approche est particulièrement utile, dans la pratique, car les données non étiquetées sont abondantes, tandis que les données étiquetées sont plus rares, et nécessitent beaucoup d'efforts pour être collectées. Cependant, en raison de sa caractéristique non supervisée, cette technique ne peut pas être utilisée directement, pour la classification. Elle est plutôt utilisée, pour découvrir automatiquement, des structures dans l'espace d'entrée, telles que la présence de clusters.

3.2.3 Apprentissage semi-supervisé (ou apprentissage semi-non supervisé) :

Les deux types de données sont utilisés, pour entraîner le modèle. Le modèle est d'abord pré-entraîné, à l'aide de données non supervisées, puis amélioré avec des données supervisées. Il y a généralement, plus de données non supervisées disponibles, et cette approche utilise cet avantage. Il est également possible, de former un modèle en utilisant les deux types de données à la fois, mais cela est plus complexe, et rarement utilisé.

3.3 RESEAUX DE NEURONE ARTIFICIELS

Les réseaux de neurones artificiels (RNA), ou bien Artificiel neural networks (ANNs), sont à l'origine des systèmes, inspirés des réseaux de neurones biologiques (Figure 3.2). Ils étaient destinés, à simuler le cerveau des animaux. Ainsi, par analogie, ils sont constitués de neurones, et de connexions entre eux [123, 124].

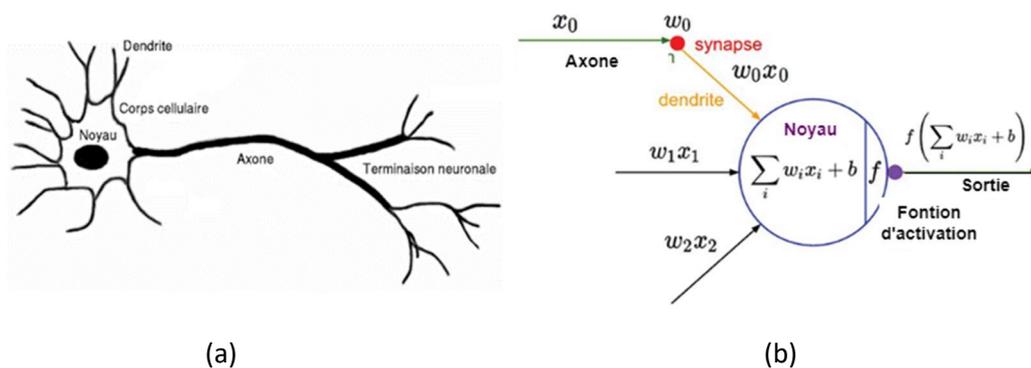


Figure 0.2: Un neurone biologique (a) comparé à un neurone artificiel (b). L'analogie est claire : les deux neurones reçoivent plusieurs signaux, et les agrègent en une seule sortie qui est transférée plus loin.

Le neurone artificiel est conçu, comme un modèle mathématique, du neurone biologique. Il s'agit d'une fonction, transformant x "stimuli" d'entrée réels, en une seule "valeur de réponse"

$$f : R^n \rightarrow R .$$

La définition la plus courante du neurone est donnée, par équation suivante (eq 3.1) :

$$f(x) = \phi(w^T x + b) \tag{3.1}$$

Où $w \in R^n, b \in R$ sont respectivement, les poids et les biais du neurone, et $\phi : R \rightarrow R$ est la fonction d'activation, qui est décrite ci-dessous. Les poids et les biais sont les paramètres du neurone.

- **Poids** : lorsqu'un signal (valeur) arrive à un neurone, il est multiplié par une valeur de poids. Si un neurone à N entrées, il a N valeurs de poids, qui peuvent être ajustées, pendant la phase d'entraînement.
- **Biais** : c'est une entrée supplémentaire, pour les neurones, toujours égale à 1, et elle puisque vous parlez de l'entrée a son propre poids de connexion. Cela garantit que, même lorsque toutes les entrées sont nulles (toutes des 0), il y aura une activation dans le neurone.

Les ANN sont généralement décrites, en termes de couches, car ils sont plus pratiques, en raison de la plus grande échelle, et de l'abstraction, par rapport aux neurones artificiels. La couche est une collection de neurones, qui sont regroupés, pour traiter les entrées multidimensionnelles, en sorties multidimensionnelles.

Un exemple notable de couche, est une couche entièrement connectée (Fully connected layers en anglais (FC)). Dans une couche entièrement connectée, les neurones reçoivent une entrée, de chaque neurone de la couche précédente. Cela permet un calcul simultané pratique, de toutes les sorties neuronales de la couche, via la multiplication matricielle (eq 3.2) :

$$f(x) = \phi(wx + b) ; \tag{3.2}$$

Où w et b sont les poids, et les biais des neurones individuels, collectés respectivement, dans une matrice et un vecteur.

La topologie de réseau, est une configuration de plusieurs couches connectées. Deux couches spéciales, sont toujours présentes dans le réseau : celles d'entrée et de sortie. Une couche d'entrée contient les valeurs du vecteur d'entrée. Une couche de sortie est la dernière couche de la chaîne, et ses sorties sont le résultat final du traitement d'entrée, effectué par le réseau. Les couches entre la couche d'entrée et la couche de sortie, sont appelées couches cachées, car leurs sorties ne sont qu'intermédiaires, et ne sont pas utilisées, en dehors du modèle. Le

nombre de couches cachées, représente la profondeur du réseau. Le nombre de neurones dans la couche, indique sa largeur (Figure 3.3).

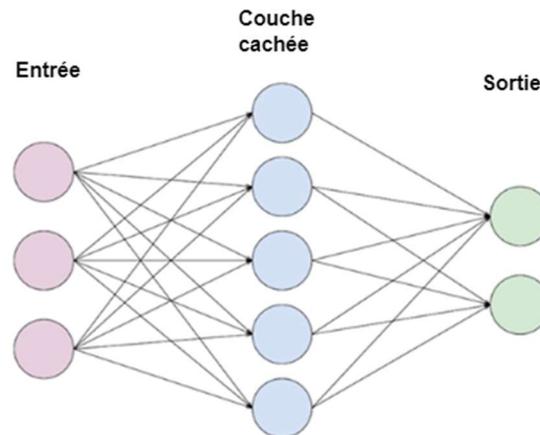


Figure 3.3: Un exemple simple de réseau de neurones artificiels. Il n'a qu'une seule couche cachée composée de 5 unités, donc la profondeur du modèle est un. La couche cachée est entièrement connectée.

3.4 FONCTION D'ACTIVATION :

Les fonctions d'activation, sont également appelées, fonctions de transfert. Le but d'une fonction d'activation, est d'introduire de la non-linéarité, dans le modèle, afin d'apprendre des fonctions non linéaires complexes.

Sans l'hypothèse des fonctions d'activation, l'espace d'un réseau de neurones, ne serait constitué que de fonctions linéaires. Des exemples populaires de fonction d'activation, incluent ReLU (unité linéaire rectifiée), et sigmoïde (Figure 3.4).

a. Fonction RELU

ReLU est un exemple de fonction d'activation, qui peut être utilisée, en conjonction avec d'autres couches. Depuis son introduction, elle est devenue la fonction d'activation, la plus fréquemment utilisée, en apprentissage profond, et est devenue plus populaire, que la tangente hyperbolique, et les fonctions sigmoïdes logistiques [125].

Elle est donnée par (eq 3.3) :

$$ReLU(x) = \max(0, x) \quad (3.3)$$

Où x est l'entrée de ReLU. Elle tronque toutes les entrées négatives à zéro.

b. Fonction sigmoïde

Elle est donnée par (eq3.4) :

$$\sigma(x) = \frac{e^x}{e^x + 1} \quad (3.4)$$

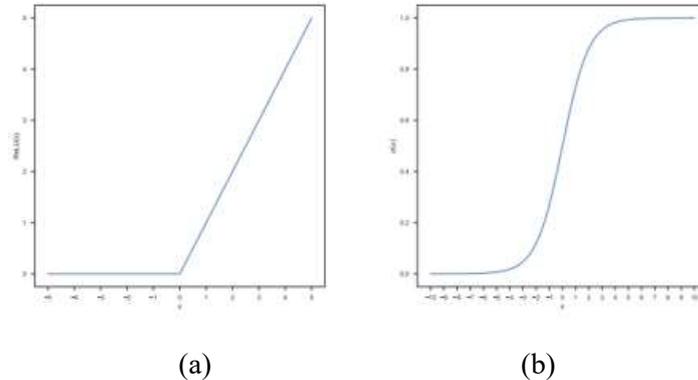


Figure 3.4: Fonctions d'activation populaires : (a) : RELU, (b) : Sigmoïde

3.5 NEURONE FORMEL (PERCEPTRON)

Le modèle original de réseau de neurones, appelé le perceptron, a été conçu en 1958 [126]. Ce modèle est la forme la plus simple, de réseau de neurones, avec deux couches, et une fonction d'activation très simple, basée sur l'addition et la multiplication. Ce type de réseau de neurones, a un potentiel d'apprentissage très limité. En raison de l'absence de couches cachées, les sorties ne sont qu'une combinaison linéaire d'entrées, de sorte que le perceptron, à une seule couche, ne pourra qu'approximer les relations linéaires [127]. La fonction du perceptron est la même, que pour le neurone artificiel. Les perceptrons peuvent être entraînés, par un algorithme d'apprentissage simple, qui est généralement appelé, la règle delta [128]. Il calcule les erreurs, entre la sortie calculée, et les données de sortie de l'échantillon, et les utilise, pour créer un ajustement des poids.

Perceptron multicouche (MLP)

Un perceptron multicouche (MLP) est un réseau neuronal artificiel, à rétropropagation, qui génère un ensemble de sorties, à partir d'un ensemble d'entrées. Un MLP se compose, d'au moins trois couches de nœuds : une couche d'entrée, une couche cachée, et une couche de sortie [129]. À l'exception des nœuds d'entrée, chaque nœud est un neurone, qui utilise une fonction d'activation non linéaire. Un MLP est caractérisé par plusieurs couches de nœuds d'entrée, connectés comme un graphe orienté, entre les couches d'entrée et de sortie.

Leur processus d'apprentissage, est basé sur la mise à jour des poids. Le réseau commence par ses poids initiaux, puis les données sont introduites dans le réseau (feed-forward), et une prédiction est faite. La prédiction est ensuite comparée à la vérité terrain, et l'erreur est calculée via une fonction de coût, et propagée vers le réseau (rétropropagation). La rétropropagation décrit, le processus de calcul des gradients d'erreur, et de mise à jour des poids. Ce processus est continuellement répété, jusqu'à ce que l'erreur de prédiction devienne très petite, et les prédictions convergent vers la vérité terrain.

3.6 DEEP LEARNING

Compte tenu de la grande quantité, de travaux de recherches, sur l'apprentissage en profondeur, une introduction complète serait impossible à inclure dans cette thèse. Au lieu de cela, nous présenterons simplement, un aperçu de chaque couche, du réseau de neurones convolutifs (CNN), utilisé au chapitre 4, et nous recommanderons au lecteur le livre "Deep Learning" de Goodfellow, Bengio et Courville [120], pour une formulation mathématique approfondie.

3.6.1 Réseaux de neurones convolutifs

Il existe un large éventail, de réseaux de neurones spécialisés. L'un des plus connus, est le réseau de neurones convolutifs (CNN) [120,130], qui a connu un énorme succès, dans les problèmes de vision par ordinateur, tels que la reconnaissance, la détection et la segmentation d'objets. Les CNN sont similaires aux réseaux de neurones classiques, décrits dans la section précédente, en ce sens, qu'ils comportent de nombreuses couches de neurones, avec des poids et des biais, qui peuvent être modifiés, par l'apprentissage.

Pour comprendre le fonctionnement d'un CNN, on peut imaginer, comment l'information visuelle, circule de la rétine, à travers le nerf optique, vers le cerveau. En bref, les neurones recherchent des couleurs, et des motifs spécifiques, dans les informations visuelles d'entrée. Tout comme le cortex visuel a plusieurs couches d'abstraction, les CNN sont conçus sur la base d'une idée similaire. Cela signifie qu'ils empilent plusieurs couches, afin de pouvoir apprendre de plusieurs abstractions différentes, des données d'entrée.

La puissance des CNN, vient du fait que, l'opération de convolution opère également sur la profondeur du réseau. Cela permet au CNN, de former des abstractions riches, et de haut niveau, des fonctionnalités de bas niveau, dans les couches précédentes, en convoluant une hiérarchie de cartes de fonctionnalités. Ainsi, le réseau est capable d'identifier des formes

complexes, telles qu'un chien ou un chat, à travers des interactions compliquées, entre des formes simples, telles que des arêtes et des gouttes.

En particulier, ils apprennent automatiquement ces abstractions, en utilisant des filtres de convolution, qui sont mis à jour, pendant le processus d'apprentissage, par rétropropagation de gradient. La procédure automatique, est justifiée par le fait que, ces réseaux apprennent les caractéristiques représentatives, de bas, moyen et hauts niveaux, des données d'entrée, directement à partir de l'image, sans avoir besoin de caractéristiques d'image, fabriquées manuellement.

De plus, ils ne nécessitent aucune connaissance préalable du domaine, ce qui signifie qu'ils peuvent apprendre à effectuer une tâche, simplement, en apprenant à partir des données d'apprentissage.

Par rapport aux réseaux de neurones, un CNN n'a qu'à apprendre un certain nombre de petits filtres, dans chaque couche. Pour une image d'un mégapixel, un réseau de neurones doit apprendre, un million de poids par nœud, dans chaque couche, tandis qu'un CNN n'a qu'à apprendre 9 poids, plus un biais, pour un filtre 3×3 . Ce rend les CNN beaucoup plus efficaces, pour gérer les données d'image (Figure 3.5).

Les couches principales d'un CNN se composent de l'entrée, de la convolution, de l'activation, du pooling et de la couche entièrement connectée. Enfin, nous rassemblerons tous ces éléments, pour former l'architecture LeNet classique, développée dans [131].

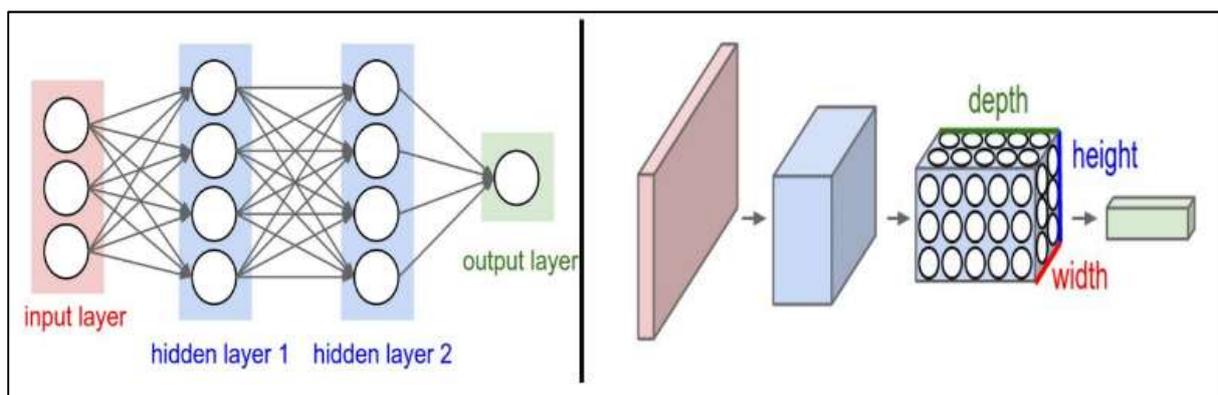


Figure 3.5: Un réseau neuronal simple et un réseau neuronal convolutif.

a. Couche de convolution :

Son but, est d'effectuer une extraction de caractéristiques, sur l'image, et de fournir une représentation, basée sur les caractéristiques de l'ensemble de données [132]. Son mécanisme

est simple. Chaque couche de convolution, consiste en un ensemble de noyaux/filtres de convolution (par exemple 1, 128), dont le nombre est défini par l'utilisateur. La profondeur des filtres, dépend du canal de l'entrée. Ensuite, la couche de convolution convolue, entre l'entrée, et chacun de ces filtres, en glissant sur l'image, afin de créer la carte caractéristique/activation en sortie. Dans cette étape, la multiplication, entre les valeurs de l'image et les valeurs du filtre, se produit. Mathématiquement, l'opération de convolution est commutative, et peut être notée par (eq 3.5) :

$$conv(x, y) = (I * K)(x, y) \sum_i \sum_k I(i, k) K(x-i, y-k) \quad (3.5)$$

Où x et y représentent, les coordonnées en pixels d'une image, Conv est la sortie de l'opération de convolution, I est l'image d'entrée, K est le filtre, et $i \times k$ représente la taille du filtre. En parallèle, le stride détermine comment le noyau va glisser, en termes d'orientation, en termes de pixels. Les figures 3.6 et 3.7 illustrent un exemple d'opération de convolution.

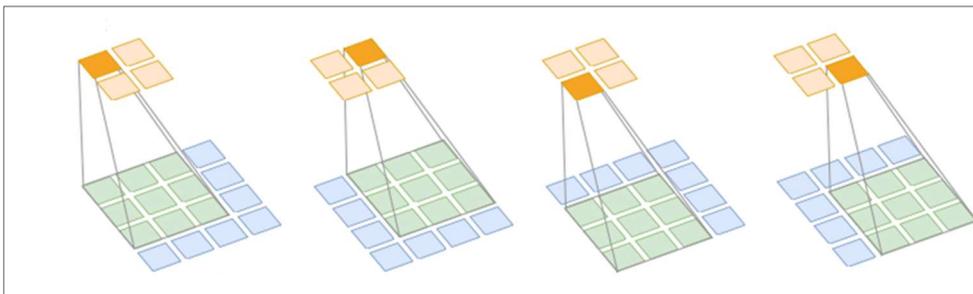


Figure 3.6: Vue d'un filtre glissant de la convolution. Le noyau, de taille 3*3, est décalé sur l'entrée (matrice de taille 4*4), et à chaque position, les cellules qui se chevauchent sont multipliées et additionnées. Chaque cellule, de la carte des caractéristiques, est le résultat d'une convolution, d'un voisinage spatial correspondant, appelé le champ récepteur.

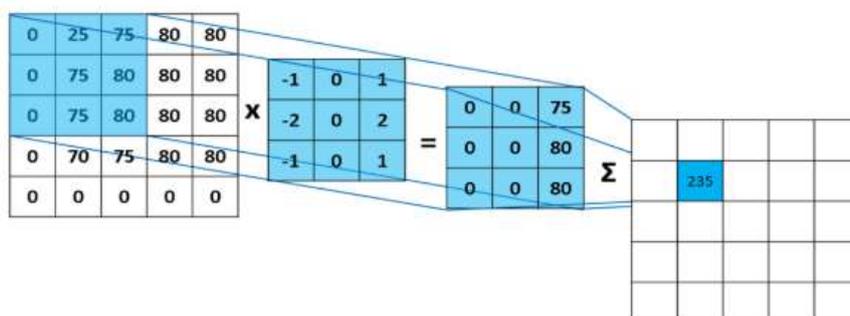


Figure 3.7: Exemple simple d'arithmétique de convolution [131].

Le volume de sortie d'une convolution, est contrôlé par : le nombre de filtres (K), la taille du filtre (F), le pas (S), et le zéro-padding (P). Nous avons déjà évoqué, le champ récepteur, qui correspond aux dimensions spatiales du filtre. Le nombre de filtres, correspondra alors à la profondeur, du volume de sortie. La taille du noyau (N), détermine la façon, dont nous déplaçons le filtre, sur l'image, et S dictera ensuite, le nombre de pixels, que nous sauterons, lors du calcul des entrées nettes, dans la carte des caractéristiques.

Parfois, nous pouvons vouloir remplir le volume d'entrée, avec des zéros, afin de manipuler la dimensionnalité du volume de sortie. C'est ce qu'on appelle le zéro-padding. Le zéro-padding découple la taille du filtre, et la taille du volume de sortie, nous permettant de contrôler l'un, indépendamment de l'autre (figure 3.8).

0	0	0	0	0	0
0	35	19	25	6	0
0	13	22	16	53	0
0	4	3	7	10	0
0	9	8	1	3	0
0	0	0	0	0	0

Figure 3.8: une matrice 4x4 à zéro-padding

Supposons que nous ayons une entrée, de dimensions $W_{in} \times H_{in} \times D_{in}$, où W désigne la largeur, H désigne la hauteur, et D désigne la profondeur. Alors les dimensions du volume de sortie $W_{out} \times H_{out} \times D_{out}$ sont données par (eq 3.6) :

$$\begin{aligned}
 W_{out} &= \frac{W_{in} - F + 2P}{S} + 1 \\
 H_{out} &= \frac{H_{in} - F + 2P}{S} + 1 \\
 D_{out} &= K
 \end{aligned}
 \tag{3.6}$$

Dumoulin et ces collègues [132] donne aperçu exhaustif de l'arithmétique des couches convolutives.

b. Couche d'activation

Dans la couche d'activation, les réponses de sortie des couches de convolution, sont cartographiées, via une fonction d'activation non linéaire. Un choix populaire, parmi les fonctions d'activation, est la fonction ReLU (§3.4. a).

Selon la figure 3.9, si l'entrée est négative, elle sera convertie en 0, et le neurone ne sera pas activé. Cela signifie, qu'à la fois, seuls quelques neurones sont activés, ce qui rend le réseau, clairsemé et très efficace.

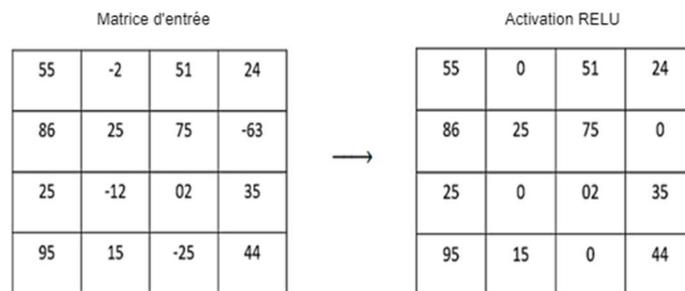


Figure 3.9: Mécanisme de la fonction RELU.

c. Couche de pooling :

La couche de pooling, est responsable de la réduction, de la dimensionnalité des caractéristiques convoluées et, par conséquent, de la réduction du nombre de paramètres du modèle. Ceci est réalisé, par la combinaison de caractéristiques, spatialement proches ou redondantes, et conduit à une représentation plus compacte, qui reste insensible, aux petits changements dans une image. Le pooling moyen et maximale, sont parmi les types les plus courants, et sont illustrés par la figure 3.10. Il existe d'autres types de pooling comme la somme, ou bien : il existe un autre type de pooling, nommé somme qui est le moins utilisé.

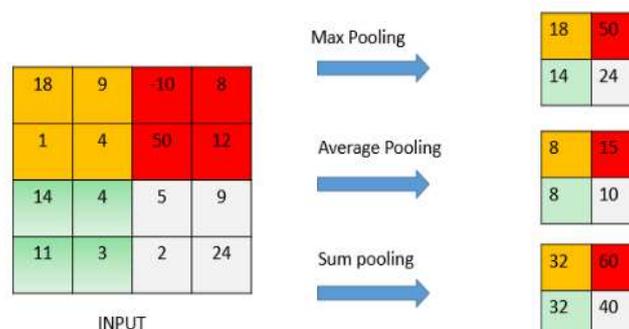


Figure 3.10: Exemple de pooling, avec les différentes stratégies (maximum, moyen et somme). La taille des fenêtres, ainsi que le pas utilisé dans cet exemple, sont de 2x2.

d. Couche entièrement connectée :

La couche entièrement connectée, est identique à la couche du réseau de neurones, entièrement connectée, qui a déjà été décrite. Son entraînement suit également le processus déjà décrit (Figure 3.11).

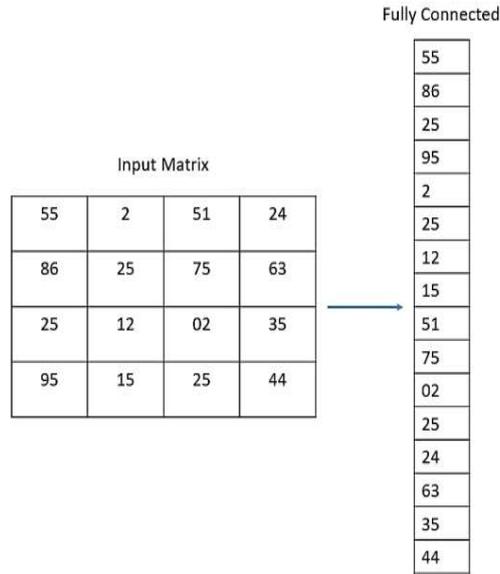


Figure 3.11: Couche entièrement connectée sur une matrice.

e. Couche Softmax

La fonction softmax est couramment utilisée, dans la dernière couche d'un CNN, et mappe la sortie de la dernière couche, entièrement connectée, à une distribution de probabilité. Elle peut être exprimée mathématiquement par (eq 3.7) :

$$f(x_i) = \frac{e^{x_i}}{\sum_{j=1}^k e^{x_j}} \quad (3.7)$$

Avec $x_i = (x_1, \dots, x_k) \in R$ représentent les activations unitaires de sortie, de la dernière couche

FC, K représente le nombre de classes, et $\sum_{j=1}^k e^{x_j}$ est le terme de normalisation, qui garantit

que toutes les valeurs de sortie de la fonction, totaliseront 1, et seront toujours comprises dans la plage (0, 1), constituant ainsi, une distribution de probabilité valide (Figure 3.12).

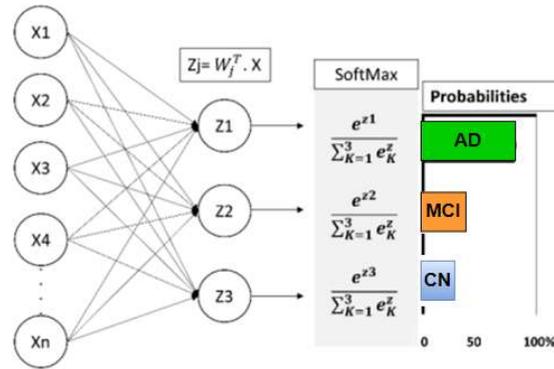


Figure 3.12: Exemple de la couche Softmax.

3.6.2 Optimisation :

L'optimisation est une partie essentielle du processus de l'apprentissage, dans les algorithmes DL, qui fait référence à la minimisation/maximisation d'une fonction, en fonction de la tâche à accomplir. Lorsque la tâche nécessite une minimisation, la fonction à utiliser est généralement appelée, fonction de coût, de perte ou d'erreur.

a. Descente de gradient

En termes d'algorithmes d'optimisation, le plus courant, pour les réseaux de neurones, est la descente de gradient. Le rôle de la descente de gradient, est de minimiser la fonction de perte. Cela peut être fait, en utilisant la rétropropagation, où les dérivées partielles de l'erreur sont calculées, par rapport à chaque poids. Le taux d'apprentissage est un paramètre, de descente de gradient, et détermine le nombre de pas nécessaires, pour atteindre un minimum (local).

b. Adaptive Moment Estimation (Adam)

C'est un algorithme d'optimisation stochastique, qui nécessite des gradients de premier ordre, et calcule des taux d'apprentissage adaptatifs, pour chaque paramètre, à partir d'estimations des premiers et seconds moments des gradients [133]. L'algorithme permet une mise en œuvre simple, il est efficace en termes de calcul, nécessite peu de mémoire, et convient parfaitement aux tâches abondantes, en données. Adam sera utilisé, dans le modèle CNN présenté, dans cette thèse.

3.6.3 Régularisation :

La régularisation consiste à bien ajuster une fonction de résolution, pendant l'entraînement tout en évitant le sur-apprentissage du système. De nombreuses techniques sont utilisées, pour éviter le sur-apprentissage, lors de la phase d'entraînement.

- **Augmentation des données** : c'est une méthode, pour augmenter la taille de l'ensemble d'apprentissage, afin que le modèle ne puisse pas tout mémoriser. Cela peut prendre plusieurs formes, selon le jeu de données, en créant une distorsion géométrique et de couleur, sur les images originales.
- **Modification des architectures de réseau** : elle est réalisée, par modification directe de l'architecture, par l'ajout ou la suppression des couches, et l'application des tests, jusqu'à l'obtention d'une architecture convenable.
- **Couche dropout** : un succès récent est démontré, avec une technique de régularisation, appelée Dropout [134]. L'idée, est de fixer au hasard, un certain pourcentage d'activations, dans chaque couche à 0. Pendant l'entraînement, les neurones doivent apprendre, de meilleures représentations, sans s'adapter les uns aux autres, en étant actifs. Par exemple, si nous fixons le dropout à 50 %, alors 2 neurones sur 4 seront inactivés pendant la rétropropagation (Figure 3.13).

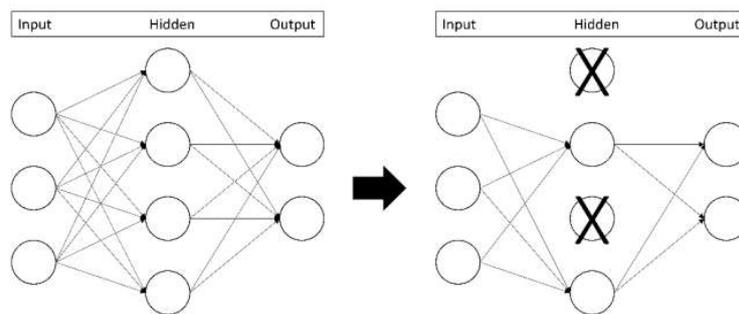


Figure 3.13: Exemple de couche dropout (la probabilité de 50 % apparaît à droite) [134].

- **Régularisation L2** : elle est également utilisée, pour ajouter une pénalité, à la fonction de coût ($C(x)$), et introduire une décroissance des poids, comme indiqué dans l'équation suivante (eq 3.8)

$$C(x) = C(x) + \lambda \sum_{j=1}^p w_j^2 \quad (3.8)$$

Où λ est le paramètre de régularisation (hyperparamètre défini), w est le ou les poids correspondants, pour $i = 1, \dots, p$.

- **Normalisation par lots** : la normalisation par lots (Batch normalisation (BN) en anglais) est une méthode, pour réduire le changement de covariable interne, dans les réseaux de neurones, conduisant à l'utilisation possible, de taux d'apprentissage plus élevés. En principe, le procédé ajoute une étape supplémentaire, entre les couches, dans laquelle la sortie de la couche avant, est normalisée. BN empêche, en outre, de petites modifications

des paramètres à amplifier, et permet ainsi, des taux d'apprentissage plus élevés, ce qui rend le réseau encore plus rapide (Figure 3.14). La normalisation par lots (BN) consiste en deux algorithmes. Le premier algorithme est la transformation de l'entrée d'origine d'une couche x en la valeur y décalée et normalisée. Le deuxième algorithme est l'entraînement global d'un réseau normalisé par lots [135].

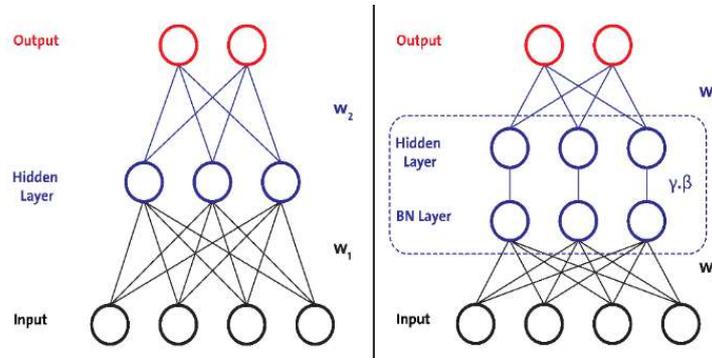


Figure 3.14: Un réseau de neurones, avant et après la normalisation par lots [136].

- **Technique d'arrêt précoce (Early stopping)** : elle consiste à arrêter l'apprentissage, avant que le modèle ne commence à diverger (Figure 3.15). En pratique, il est beaucoup utilisé, lors de l'apprentissage des réseaux de neurones [137].

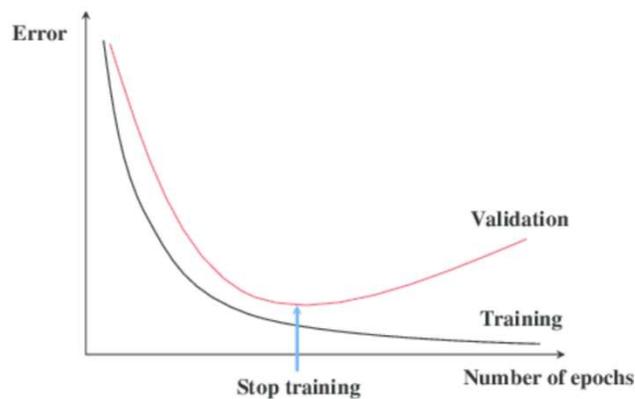


Figure 3.15: Illustration de l'arrêt précoce.

3.6.4 Autres architectures populaires des CNN :

a. LeNet : un réseau neuronal convolutif simple

En résumé, le bloc de construction principal d'un CNN, est la couche convolutive. Tout d'abord, l'opération de convolution, calcule une somme pondérée d'entrées, qui passent par une fonction d'activation non linéaire. Il en résulte une carte de caractéristiques 2D d'activations, qui sont regroupées, pour produire une carte de caractéristiques, de résolution inférieure.

L'un des CNN les plus anciens et les plus célèbres, est la famille de modèles LeNet [138], développé par Yann LeCun et ses collègues, pour identifier les chiffres manuscrits.

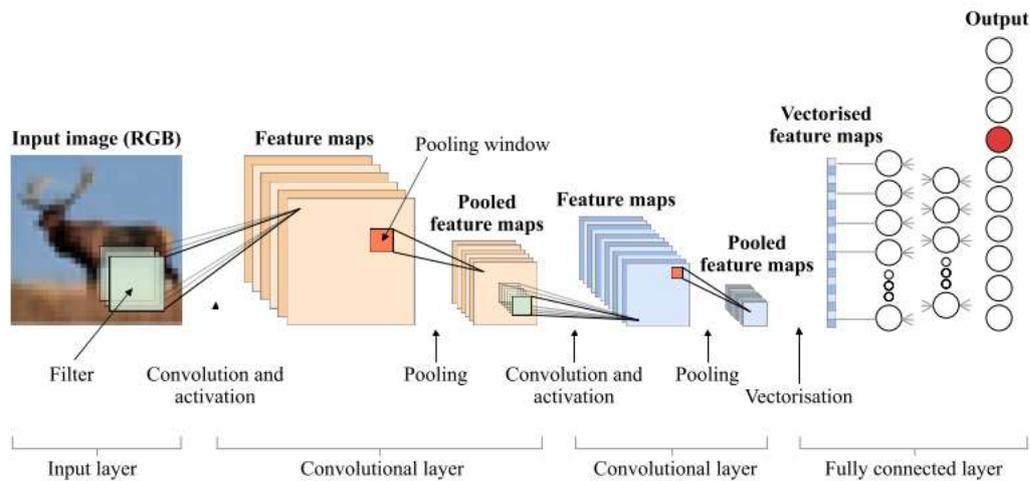


Figure 3.16: Un exemple de l'architecture LeNet.

La figure (3.16) montre l'architecture générale d'un modèle LeNet appliqué à une entrée d'image couleur. Deux couches convolutionnelles sont empilées, l'une sur l'autre. Le nombre de cartes d'entités, dépend du nombre de filtres spécifiés dans chaque couche. Le volume final de cartes de caractéristiques, est déroulé dans un long vecteur de caractéristiques, qui est introduit dans une couche, entièrement connectée (fully connected). La couche de sortie applique généralement une fonction d'activation softmax, et le neurone de sortie, avec la plus grande activation, détermine la prédiction de classe.

b. AlexNet

AlexNet est une architecture CNN profonde proposée en 2012, en tant que lauréate du challenge l'ILSVRC. Elle était capable de battre, toutes les méthodes traditionnelles d'apprentissage automatique [139]. Ce réseau comporte 8 couches, avec 60 millions de paramètres. Les cinq premières couches du réseau, sont des couches convolutionnelles, et les trois dernières couches, sont des couches entièrement connectées. La dernière couche, entièrement connectée, est connectée à un softmax, qui génère 1000 valeurs de probabilité, pour 1000 étiquettes de classe. La figure (3.17) montre l'architecture d'AlexNet.

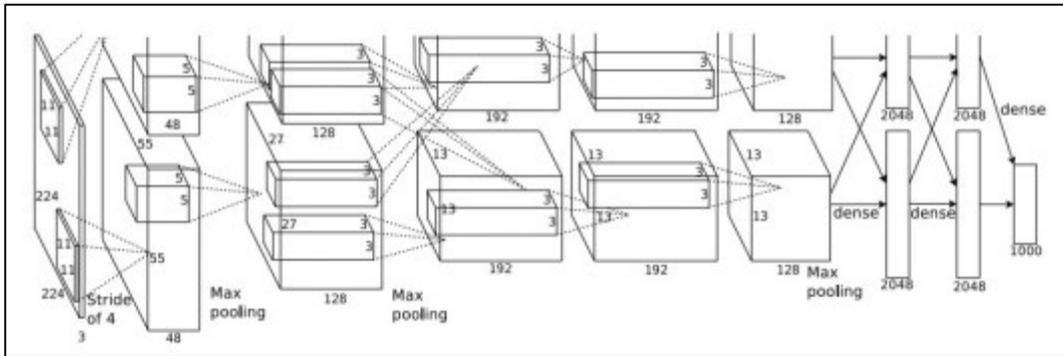


Figure 3.17: Architecture AlexNet.

c. VGG :

Ce CNN profond, a été proposé en 2014, par Visual Geometry Group. Il mentionne que l'augmentation de la profondeur du réseau peuvent aboutie à des valeurs de précision de classification, plus élevées [50]. Trois modèles VGGNet sont proposés, avec les noms VGG-11, VGG-16 et VGG-19, et les numéros à la fin des noms indiquent le nombre de couches du réseau.

d. GoogLeNet

Ce CNN est un autre réseau de neurones, à convolution profonde, proposé en 2014 [51]. C'est une forme particulière, de l'architecture Inception, et comprend 22 couches. La caractéristique importante de ce réseau, est la réduction de dimension. Comparé à AlexNet et VGGNet, le nombre de paramètres est beaucoup plus faible dans GoogLeNet, ce qui réduit la complexité des calculs. Concrètement, par rapport à l'architecture AlexNet, proposée en 2012, GoogLeNet fournit 12 fois moins de paramètres.

E. ResNet

Ce réseau, basé sur l'apprentissage résiduel, est proposé en 2015, dans différentes versions, telles que ResNet18, ResNet34, ResNet101, ResNet152, qui ont différents nombres de couches [63]. ResNet50 est l'un des modèles ResNet, les plus populaires, avec 50 couches de convolution, et 25,5 millions de paramètres.

3.7 CONCLUSION

Dans ce chapitre, nous avons présenté une vue d'ensemble, de l'approche d'apprentissage en profondeur, commençant par la représentation des théories fondamentales, des réseaux de neurones, de la perception, en passant par la définition du réseau multicouche, aux réseaux convolutionnels profonds. La limitation de non linéarité des perceptrons, est résolue par

l'introduction d'une fonction d'activation non linéaire, comme la fonction RELU, et un réseau de neurones multicouches, au lieu d'un seul neurone. Nous avons présenté les différents éléments qui entrent dans la construction des CNN tels que : la couche de convolution, la couche de pooling, la couche de regroupement, et la fonction Softmax. Nous sommes passés ensuite à l'algorithme utilisé, pour ajuster les paramètres d'apprentissage, et la régularisation. Finalement, nous avons conclu le chapitre par une revue sur les architectures CNN les plus populaires. Ce chapitre était une sorte de rappel, pour l'introduction de nos solutions dans le problème de classification AD, en utilisant les modalités MRI.

Dans le chapitre suivant, nous présenterons nos approches de classification AD, et leurs implémentations, en utilisant les méthodes CNN.

CHAPITRE 4

METHODOLOGIES

4.1 INTRODUCTION :

Dans les chapitres précédents, nous avons passé en revue l'état de l'art de la recherche, liée à la classification de la maladie l'Alzheimer, en mettant en évidence les travaux les plus récents, utilisant des méthodes d'apprentissage en profondeur. Nous avons expliqué en détail la méthode d'apprentissage en profondeur, en se concentrant beaucoup plus sur les CNN, que nous avons utilisés dans notre thèse. Dans ce chapitre, nous proposons l'approche que nous avons adopté, pour la détection précoce de l'AD, par l'imagerie de diffusion, en utilisant l'approche CNN. Nous sommes passés à une extension au cadre multimodale, puis nous avons effectué une comparaison, entre l'étude monomodale par DTI, et l'étude par MRI multimodale (DTI et sMRI). Nous présentons en premier lieu la base de données ADNI utilisée, puis les étapes de prétraitement, et la méthode de sélection des coupes utilisées. Nous décrivons l'architectures CNN utilisée, pour l'extraction des caractéristiques, et nous terminons par l'étape de classification par SVM.

4.2 PRESENTATION DE LA BASE DE DONNEE ADNI.

Les données utilisées, dans la préparation de ce travail, ont été téléchargées, à partir de la base de données Alzheimer's Disease Neuroimaging Initiative (ADNI)¹. L'ADNI a été lancée en 2004, par National Institute of Bio-medical Imaging and Bio-engineering (NIBIB), Food and Drug Administration (FDA), et des sociétés pharmaceutiques privées, et des organisations à but non lucratif, en tant que partenariat public-privé, de 60 millions de dollars, sur 5 ans. ADNI est le résultat, des efforts de nombreux chercheurs, d'un large éventail d'établissements universitaires, et de sociétés privées. Les sujets ont été recrutés dans plus de 50 sites, à travers les États-Unis et le Canada. Depuis les dates de lancement, il y a plus de dix ans, le projet comporte désormais trois phases, ADNI-1, ADNI-GO (Grand Opportunities), ADNI-2 et ADNI-3.

¹ <http://adni.loni.usc.edu/>

L'ensemble de données, sélectionnées dans cette thèse, comprend à la fois des images cérébrales, pondérées en diffusion (DWI), et des images pondérées en T1, de 150 personnes des deux sexes (50 AD, 50 CN et 50 MCI), avec des âges variant de 55 à 90 ans, acquis par les scanners, du système médical GE. Les sujets sélectionnés sont issus des phases, ADNI-GO et ADNI-2.

Les étiquettes existantes sont :

- AD (Alzheimer's disease) : patients ayant la maladie d'Alzheimer ;
- EMCI (early MCI) : patients au début de la phase prodromique ;
- LMCI (late MCI) : patients en fin de phase prodromique (similaire au label MCI d'ADNI-1);
- CN (Cognitively normal) : participants témoins âgés.

Les deux phases prodromiques, sont regroupées dans notre étude, et sont labellisées par MCI, de manière similaire, aux différents travaux antérieurs, donc les 50 sujets MCI regroupent 25 patients EMCI et 25 LMCI.

Les informations démographiques, des participants employés dans cette étude, sont répertoriées dans le tableau 1, où le mini-examen de l'état mental (MMSE), est un test mental, qui mesure la fonction cognitive. Le score MMSE inférieur (tableau 4.1), en tant qu'indice de diagnostic auxiliaire, implique une faible capacité cognitive.

Les sMRI 3D brutes, pondérées en T1, sont acquises, avec $256 \times 256 \times 196$ voxels par volume, et une taille de voxel de $1,0 \times 1,0 \times 1,2 \text{ mm}^3$.

Les données DWI brutes, ont été acquises avec une matrice de 128×128 , une taille de voxel de $2,7 \times 2,7 \times 2,7 \text{ mm}^3$, et 41 directions de gradient ($b = 1000 \text{ s} / \text{mm}^2$). En plus de ces images, 5 images pondérées en T2 sans diffusion ($b = 0$), sont utilisées comme référence. Plus de détails sur le protocole d'acquisition, peuvent être trouvés dans le protocole ADNI2.

Tableau 4.1: Données statistiques liées aux sujets

Classes	CN	MCI	AD
Nombre	50	50	50
Sexe (F/M)	27/23	20/30	28/22
Age	72.5±6.1	74.40 ± 7.47	75.60 ± 8.63
MMSE	28.93±1.18	27.4 ± 1.99	23.0 ± 2.42

Approche proposée

La stratégie que nous avons adoptée, consiste en un prétraitement, une sélection de coupe 2D, une extraction de caractéristiques, et une classification. Nous travaillons sur les mesures DTI (FA, MD), et la matière grise, segmentée à partir d'images MRI du cerveau, pondérées en T1, pour classer (CN vs AD), (AD vs MCI), et (CN vs MCI). La nouvelle architecture 2DCNN, est entraînée par un ensemble de données, au niveau des coupes (uniquement les 32 coupes pertinentes sélectionnées à partir des images FA, MD, GM), pour extraire les caractéristiques pertinentes, des cartes DTI et GM. Les modèles FA-CNN, MD-CNN et GM-CNN sont adaptés, pour extraire les fonctionnalités de la dernière couche entièrement connectée. Après cela, les caractéristiques de chaque coupe, de l'ensemble de données, au niveau du sujet (FA, MD, GM), sont extraites par leur modèle optimal (FACNN, MD-CNN et GM-CNN). Ces caractéristiques sont ensuite fusionnées, et transférées au classificateur SVM, pour améliorer les performances (figure 4.1). La description détaillée, se trouve dans les sous-sections suivantes.

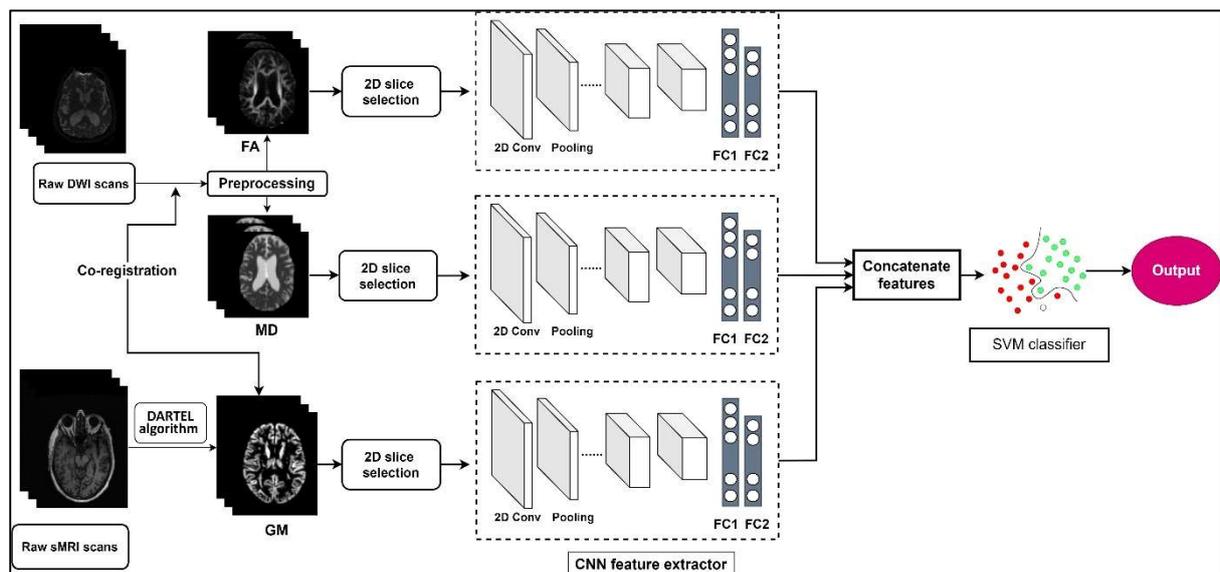


Figure 4.1: Organigramme du système de fusion multimodalité proposé, utilisant l'approche 2DCNN-SVM, pour l'identification de l'AD.

4.3 PRETRAITEMENT

Les images MRI peuvent souffrir de divers artefacts. Ceux-ci peuvent résulter, de sources physiologiques (par exemple, mouvement de la tête, respiration et anxiété), ou du scanner lui-

même (par exemple, distorsions géométriques et pertes de signal). Généralement, la procédure de prétraitement des images de diffusion comprend les étapes principales suivantes :

- La correction du champ de polarisation,
- La normalisation de l'intensité,
- L'extraction du cerveau,
- La correction de la distorsion induite par les courants de Foucault,
- Le recalage.

Plusieurs logiciels statistiques, sont devenus largement disponibles, et couramment utilisés, dans le prétraitement, et récemment dans l'analyse d'images, comme la cartographie paramétrique statistique (SPM) [140], et FMRIB Software Library (FSL)[141]. Ces deux logiciels sont utilisés dans notre thèse. Ils contiennent des outils de prétraitement et d'analyse.

➤ **SPM**

SPM [142] est un logiciel, librement distribuée, conçu pour construire des cartes statistiques paramétriques, pour l'analyse des séquences de données, d'imagerie cérébrale. Les séquences peuvent être, une série d'images, de différents plans, ou des séries temporelles, du même sujet. La version actuelle est implémentée, pour l'analyse d'images, de différentes natures : résonance magnétique fonctionnelle, tomographie par émission de positrons, tomographie par émission de photons uniques, électroencéphalographie et magnétoencéphalographie.

➤ **FSL**

FSL [141] est une autre bibliothèque, open source, d'outils d'analyse pour les données d'imagerie cérébrale fonctionnelle, structurale et de diffusion. FSL dispose d'un ensemble d'outils majeurs, qui peuvent être utilisés directement (tableau 2.1).

Tableau 4.2: Les différents outils contenus dans le logiciel FSL :

Nom de l'outil	Utilisation
FDT	Diffusion : tractographie probabiliste
TBSS	Diffusion : analyse multi-sujets voxel-Wise
BET	Extraction du cerveau
FAST	Segmentation de type tissulaire (MG/MB/LCR)
FIRST	Segmentation sous-corticale
FLIRT and FNIRT	Recalage linéaire et non linéaire

FSL-VBM	Analyse Voxel de la densité de matière grise
SIENA and SIENAX	Atrophie cérébrale totale (longitudinale et transversale)

Dans cette section, nous présentons les étapes de prétraitement, les plus importantes, pour les images structurales et de diffusion, utilisées dans notre thèse. Nous nous intéressons à la détection de l'atrophie GM, et des anomalies FA et MD. Tout d'abord, les scans sMRI bruts, et les scans DWI, sont convertis, du format DICOM au format Neuroimaging InFormatics Technology Initiative (NIFTI), à l'aide du logiciel MRICro (<http://www.mricro.com/>). Le format NIFTI est le format d'images utilisé par les deux logiciels FSL et SPM12.

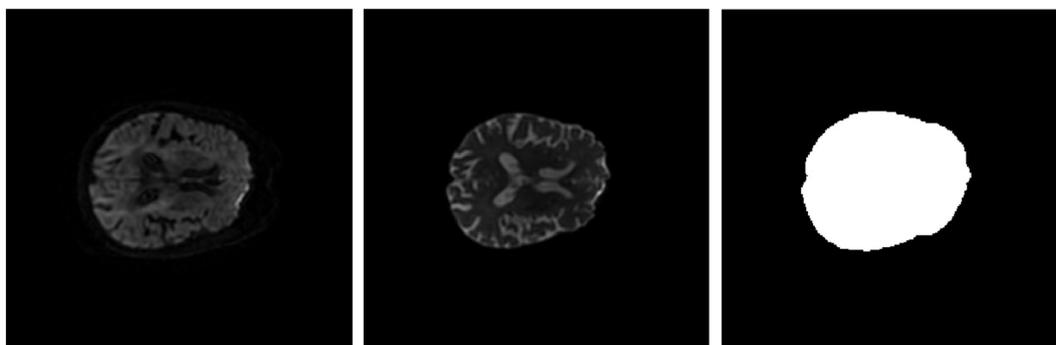
4.3.1 Prétraitement des images de diffusion

Les volumes DWI sont prétraités, à l'aide de la bibliothèque, de logiciels (FSL), d'imagerie par résonance magnétique fonctionnelle du cerveau (FMRIB).

Les premiers scans DWI sont corrigés, pour les distorsions de courant de Foucault, et les artefacts de susceptibilité. L'outil d'extraction du cerveau de FSL, est utilisé, pour retirer le crâne cérébral. Les calculs du tenseur de diffusion sont effectués, par le dtifit FSL, à chaque voxel des scans DWI fixes. Les valeurs propres du tenseur de diffusion (λ_1 , λ_2 , λ_3), sont utilisées, pour obtenir des cartes, d'anisotropie scalaire, et de diffusivité FA et MD.

4.3.2 Extraction du cerveau

Les tissus non cérébraux, peuvent constituer un obstacle, pour les algorithmes d'analyse d'images. Un grand nombre de méthodes sont développées, pour l'extraction du cerveau, et beaucoup sont implémentées, dans des outils logiciels, tels que le Brain Extraction Tool (BET) [143], utilisé dans notre méthode (Figure 4.2).



(a)

- (b)

(c)

Figure 4.2: Les images pondérées en diffusion (DWI) ;

(a) Images DWI avec $b=1000$,

(b) image DWI avec $b=0$,

(c) image masque résultante avec l'outil BET.

4.3.3 Correction de la distorsion induite par les courants de Foucault et génération des cartes FA et MD

Le champ hors résonance, induit par les courants de Foucault, est fréquemment présent, dans les images écho planaires (EPI). Le protocole d'acquisition MRI, le plus rapide disponible, est causée, par les gradients de codage de diffusion, à commutation rapide, qui est une source supplémentaire, de hors résonance. Le champ magnétique, changeant rapidement, entraîne des courants de Foucault, dans les conducteurs, à l'intérieur de l'alésage, induisant ainsi à son tour, un champ magnétique. De nombreuses études, ont porté sur la correction de cet artefact. La technique, la plus couramment utilisée, est l'outil Eddy correct de FSL. Nous avons utilisé cet outil, qui intègre les corrections des distorsions, induites par les courants de Foucault, et les mouvements du sujet. L'outil dtifit FSL est utilisé, pour générer les cartes FA et MD.

4.3.4 Extraction de cartes entières de matière grise

Pour la MRI T1w anatomique, le pipeline de prétraitement, est basé sur SPM12 [142] implémentée dans MATLAB R2021a. Tout d'abord, la procédure de segmentation [144] est utilisée, pour effectuer simultanément, la segmentation des tissus, la correction du biais, et la normalisation spatiale de l'image d'entrée (figure 4.3). Ensuite, un modèle de groupe est créé, à l'aide de DARTEL, un algorithme de recalage d'images diffeomorphes [145], à partir des cartes de probabilité tissulaire des sujets, sur l'espace natif, généralement les tissus utilisés sont GM, WM et LCR, obtenus à l'étape précédente. Enfin, la méthode DARTEL à MNI [145] est appliquée, fournissant un recalage des images de l'espace natif à l'espace MNI

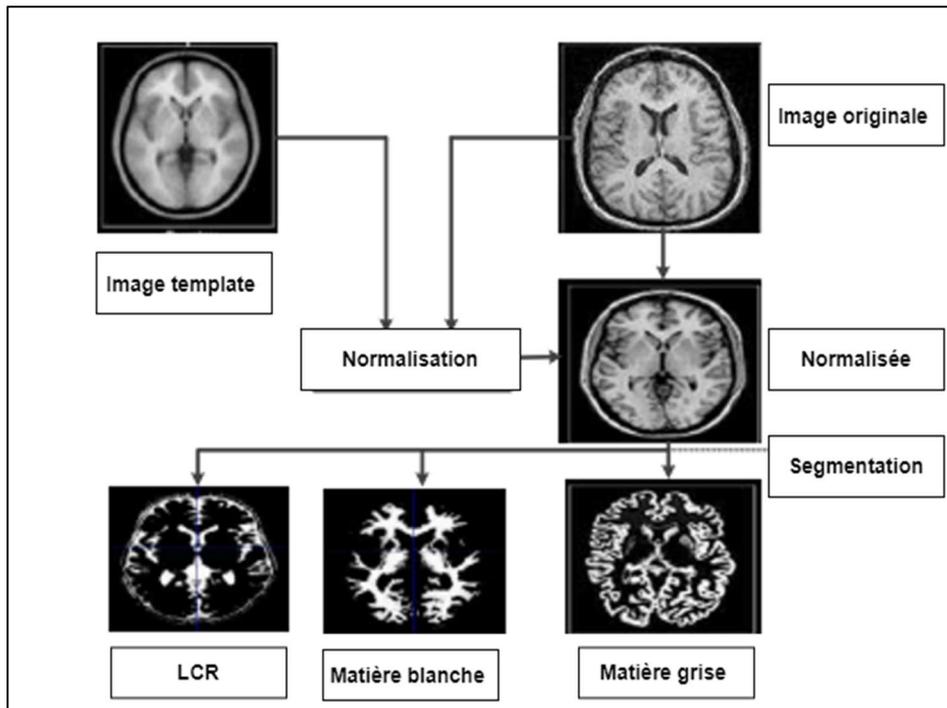


Figure 4.3 : Le prétraitement effectué aux images T1, pour segmenter la matière grise.

Les principales étapes du prétraitement sont présentées dans ce qui suit.

➤ **Normalisation spatiale :**

Le but de la normalisation spatiale, est de fournir un alignement des images MRI, dans l'espace standard (modèle), afin d'établir une correspondance voxel à voxel, entre les sujets [146]. La figure 4.4 montre les détails de la normalisation spatiale.

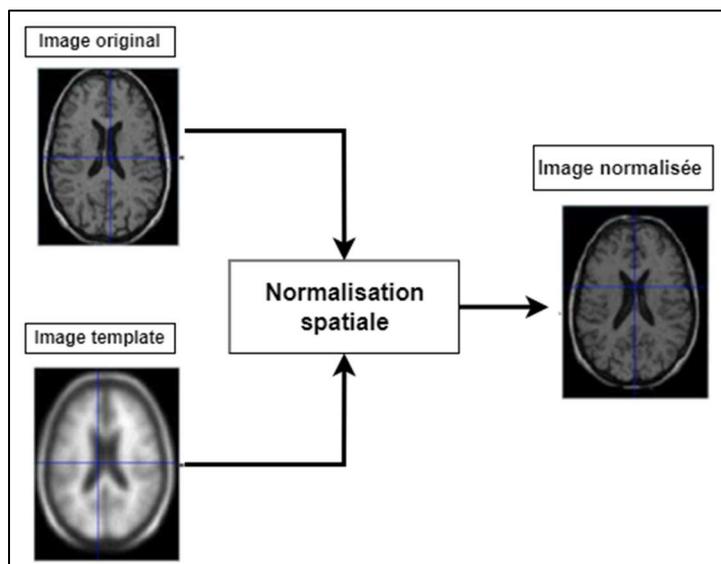


Figure 4.4: Les détails de la normalisation spatiale des images sMRI, à l'aide d'un modèle.

➤ **Segmentation :**

Le but de la segmentation, est de segmenter les images MRI normalisées, en composants matière grise, et matière blanche. La figure 4.5 illustre le processus de segmentation.

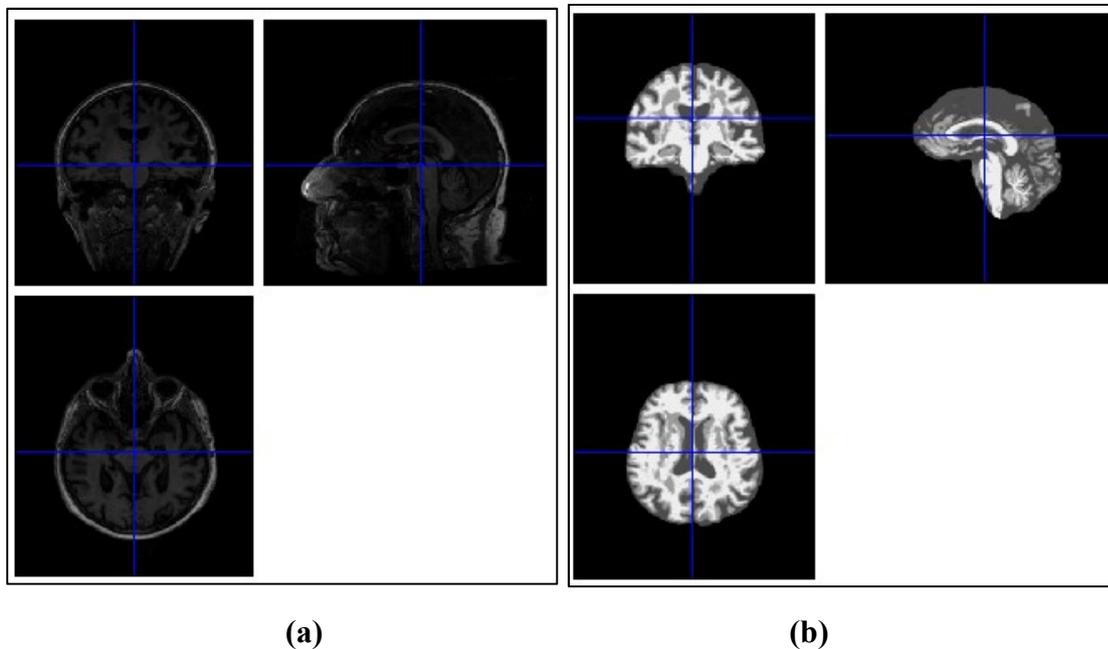


Figure 4.5: Images des trois plans (coronale, sagittale et axiale) ; (a) T1w originales (b) après utilisation de VBM à l'espace natif (visualisé à l'aide de la fonction Check-Reg du logiciel SPM).

➤ **Recalage**

Le recalage, dans l'espace standard de l'Institut neurologique de Montréal (MNI), est un processus important, qui contient une transformation affine linéaire, et une déformation non linéaire, en utilisant la normalisation DARTEL, à haute dimension. Ce processus implique l'utilisation du modèle DARTEL, généré à partir de 550 participants témoins sains (définis par les paramètres par défaut de VBM8) [147] (Figure 4.6).

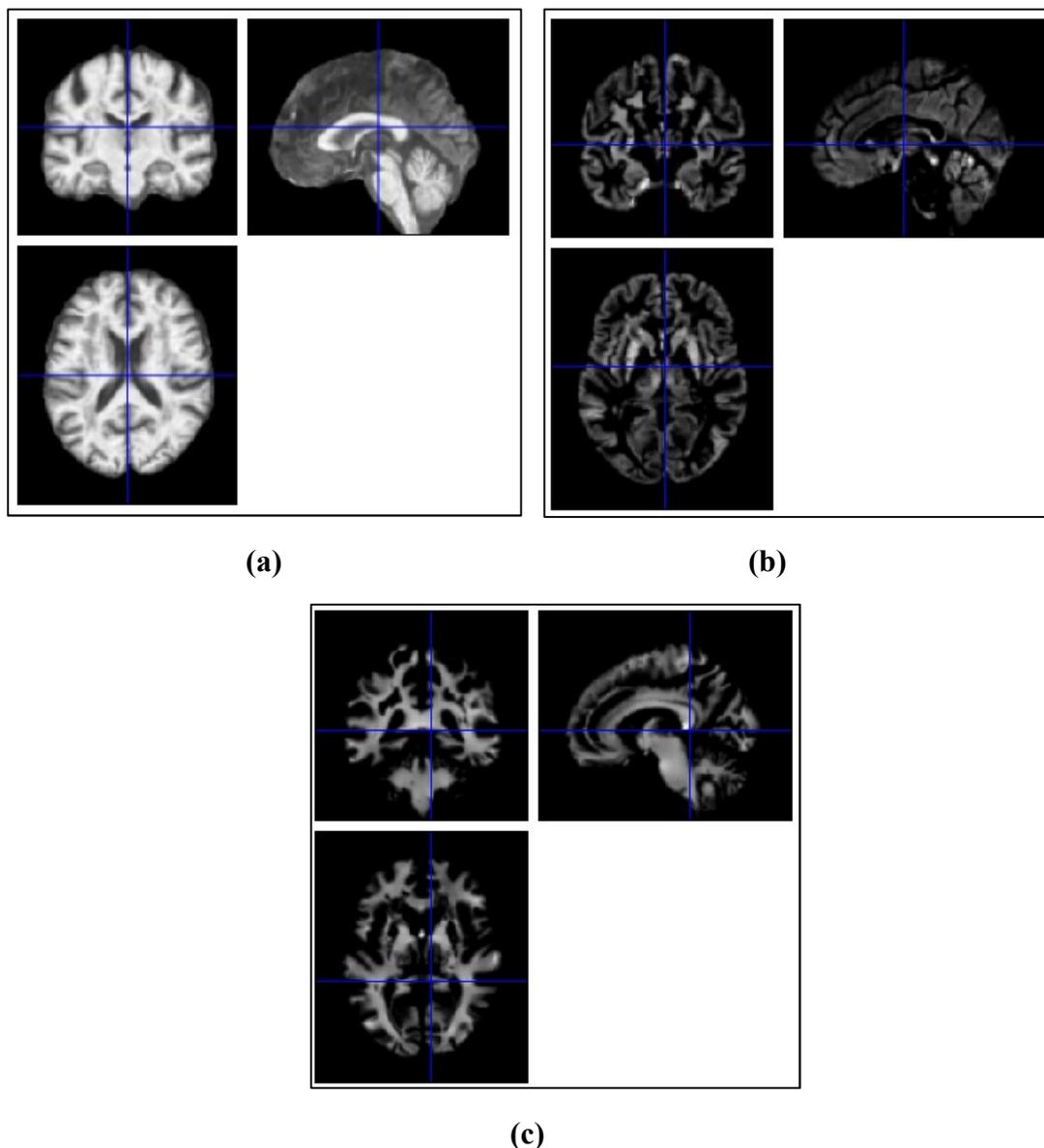


Figure 4.6: les images résultantes des trois plans prétraitées, segmentées et recalées dans l'espace MNI ; (a) Tw1, (b) la matière grise, (c) la matière blanche (visualisé à l'aide de la fonction Check-Reg du logiciel SPM).

4.3.5 Recalage des images DTI-FA et DTI-MD

FA et MD sont recalés avec les scans sMRI correspondants, en utilisant SPM12, et chaque scan contient $121 \times 145 \times 121$ voxels. Nous utilisons la méthode de recalage de SPM12, avec une approche « information mutuelle normalisée ». L'approche maximise les informations mutuelles, entre deux images. Cela permet de superposer les cartes MD et FA, sur la propre anatomie d'un même individu, et de vérifier les valeurs MD et FA, qui se superposent à l'espace structurel (figure 4.7). En effet, la méthode maximise l'information mutuelle, dans l'histogramme 2D, en un nombre limité d'itérations. Du fait qu'il y a un nombre limité

d'itérations, il est important, qu'avant le recalage, les images se trouvent approximativement, au même endroit. La correspondance, entre les cartes MD et FA normalisées-recalées obtenues, et leurs scans anatomiques correspondants normalisés, est illustrée dans la figure 4.7.

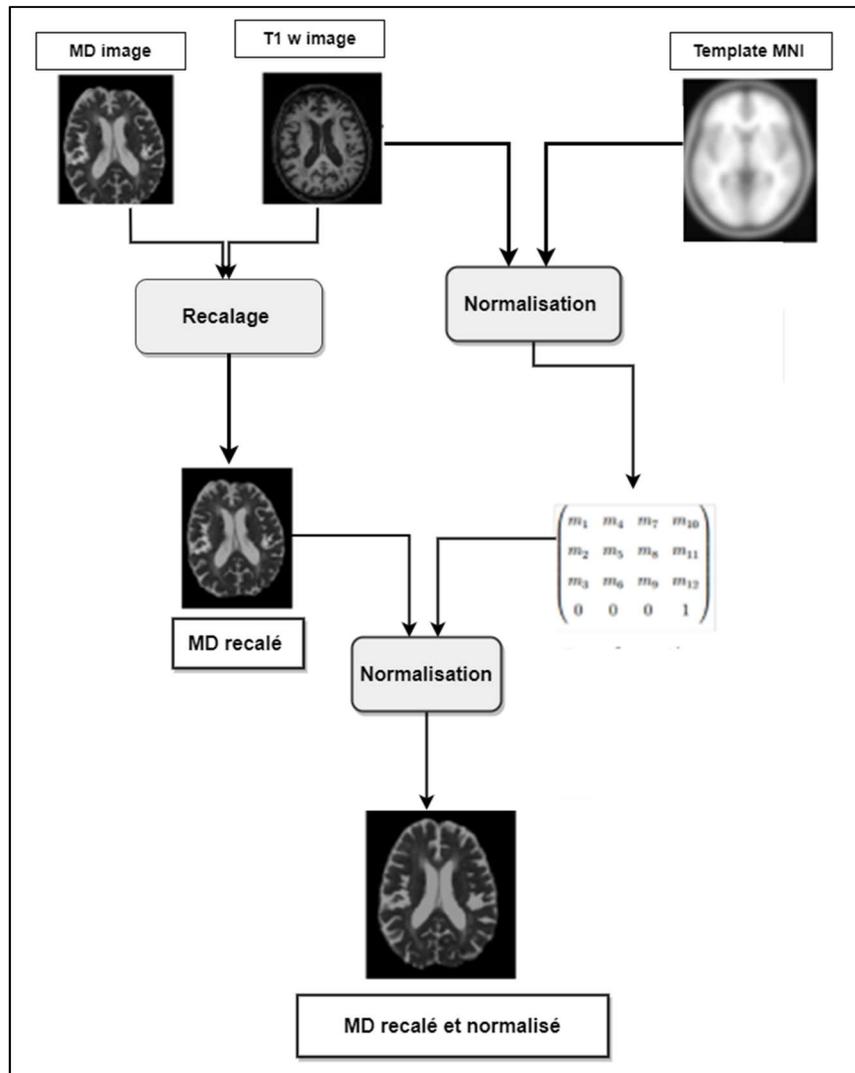


Figure 4.7 : Schéma du processus de recalage des images MD.

➤ Mesure de similarité de l'information mutuelle

L'information mutuelle (IM), est une mesure de similarité, introduite par Maes et al. [148], pour faire du recalage, entre deux images, de modalités différentes [149] (comme sMRI, la carte FA, et MD, dans notre cas). L'IM consiste à mesurer, l'entropie de l'histogramme conjoint, des deux images. Selon la théorie de l'information, l'entropie est la quantité d'information, que contient une image. L'entropie d'une image A est définie par (eq 4.1) :

$$H(A) = \sum_i p(i) \log p(i); \quad (4.1)$$

Où i sont les valeurs d'intensité, dans A , et $p(i)$ est la fonction de distribution de probabilité marginale (PDF) de i . La quantité d'informations combinées, de deux images est mesurée par leur entropie conjointe. Pour deux images A et B , leur entropie conjointe, est définie par (eq 4.2):

$$H(A, B) = \sum_{i,j} p(i, j) \log p(i, j); \quad (4.2)$$

Où i est la PDF de A , et j est la PDF de B . Par conséquent, le MI de A et B , est donné par (eq 4.3):

$$IM(A, B) = H(A) + H(B) - H(A, B); \quad (4.3)$$

Où $H(A)$ et $H(B)$ sont l'entropie de l'image individuelle, et $H(A, B)$ est l'entropie conjointe. Cependant, Studholme et ses collègues, ont prouvé, que la mesure IM est sensible au chevauchement des changements, et ont proposé les mesures d'information mutuelle normalisée, (IMN) comme alternative, pour surmonter ce problème [150]. Le IMN est défini comme (eq 4.4) :

$$IMN = \frac{H(A) + H(B)}{H(A, B)} \quad (4.4)$$

4.4 SELECTION DE COUPES 2D

Chaque volume FA, MD, et GM est décomposé, en coupes 2D, le long de la vue axiale, pour mettre en évidence, les caractéristiques, les plus distinctives, et assurer une efficacité, de classification améliorée. Nous sélectionnons 32 coupes, de chaque sujet, avec une entropie plus élevée (coupes avec des indices 34-65)[82]. Les coupes sélectionnées, sont associées, à la plupart des régions cérébrales, détériorées de l'AD mentionnées, dans la littérature, telles que l'hippocampe, le cortex entorhinal, et le thalamus. En conséquence, un total de, 1600 (32×50) coupe, de chaque classe (CN, MCI et AD) sont sélectionnés.

4.5 EXTRACTION DES CARACTERISTIQUES

L'extraction de caractéristiques classiques était le principal problème, des algorithmes d'apprentissage automatique traditionnels, ce qui est difficile, et prend du temps. Comme nous l'avons expliqué en détail, dans le chapitre précédent, l'un des avantages des CNNs, est qu'ils peuvent effectuer cette tâche, automatiquement, sans intervention humaine.

En bref, l'architecture 2DCNN, utilisée dans notre approche, se compose de trois couches convolutives, avec des filtres de taille 3×3 . Chaque couche convolutive, est suivie d'une couche RELU, de couches de normalisation par lots (BN), d'une couche max pooling, de deux couches, entièrement connectées, d'une couche softmax, et d'une couche de sortie. La couche RELU, définit les valeurs négatives sur zéro, et BN accélère le processus d'entraînement. Plus de détails sont présentés dans le tableau 4.3.

Tableau 4.3: Propriétés des couches, pour l'architecture 2DCNN proposée

Couche No	Nom de la Couche	Propriétés de la Couche
1	Couche entrée	$145 \times 121 \times 1$ images
2	Couche de convolution 1	$32 \ 3 \times 3 \times 1$ convolutions, avec un pas [1 1]
3	Batch Normalization BN	32
4	ReLu	ReLu
5	Max pooling	2×2 max pooling, avec un pas [2 2]
6	Couche de convolution 2	$16 \ 3 \times 3 \times 32$ convolutions, avec un pas [1 1]
7	Batch Normalization	16
8	ReLu	ReLu
9	Max pooling	2×2 max pooling, avec un pas [2 2]
10	Couche de convolution 3	$8 \ 3 \times 3 \times 16$ convolutions, avec un pas [1 1]
11	Batch Normalization	8
12	ReLu	ReLu
13	Fully Connected	256 neurones cachés, dans la couche fully connected
14	Fully Connected	2 neurones cachés, dans la couche fully connected
15	Softmax	Softmax
16	Classification sortie	CN/ MCI, CN/AD, AD/MCI

4.6 CLASSIFICATION A L'AIDE DE LA MACHINE A VECTEURS DE SUPPORT (SVM)

Les modèles FA-CNN, MD-CNN et GM-CNN sont adoptés, pour extraire les caractéristiques. Celles-ci sont ensuite transmises au classifieur SVM, au lieu de la couche Softmax, pour performer la classification. Ces caractéristiques extraites des images FA, MD et GM, constituent une matrice, dont la taille est le nombre de coupes, multiplié par le nombre de caractéristiques, sélectionnées dans chaque coupe. Pour 32 coupes, de chaque sujet, la représentation des caractéristiques à la dimension de 32×2 .

Pour tous les sujets (100), la sortie de chaque modèle est une matrice de 100×32 . Ils sont ensuite concaténés, dans une matrice de caractéristiques totales, de dimension 3200×2 . Le

classifieur SVM est entraîné et testé, à l'aide des caractéristiques, extraites en profondeur (figure 4.8).

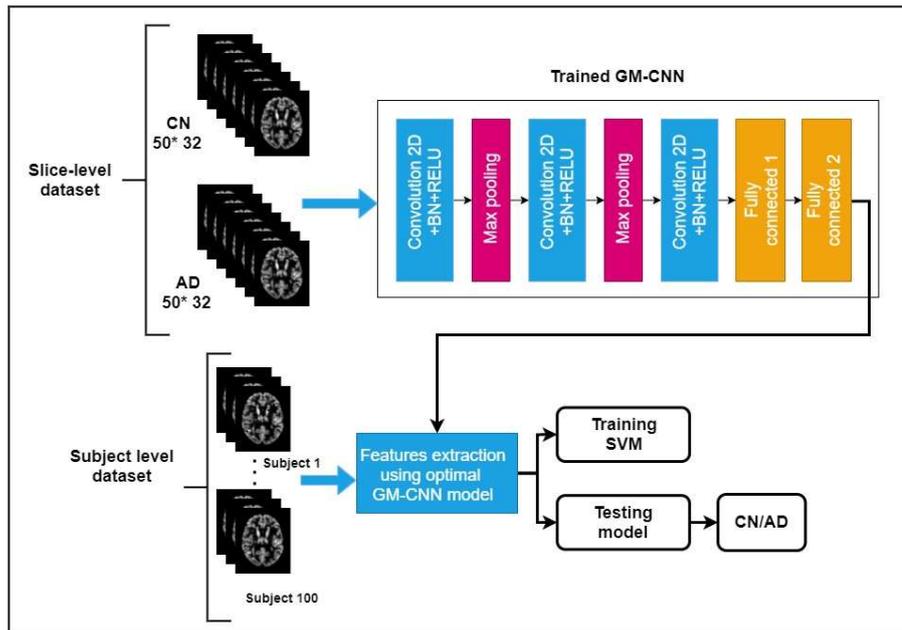


Figure 4.8 : Le pipeline de GM-CNN, proposé avec la méthode SVM, pour distinguer AD et CN.

4.6.1 Machine à vecteurs de support (SVM)

La méthode de machine à vecteurs de support (SVM), est une technique d'apprentissage supervisé, introduite en 1992 par Boser, Guyon et Vapnik [151]. SVM est basée sur l'hypothèse, qu'il existe un plan de séparation optimal, entre les points de données, appartenant à différentes classes. SVM trouvera le meilleur plan de séparation, et sa marge associée, marquée par les deux lignes, parallèles à la ligne de séparation (figure 4.9).

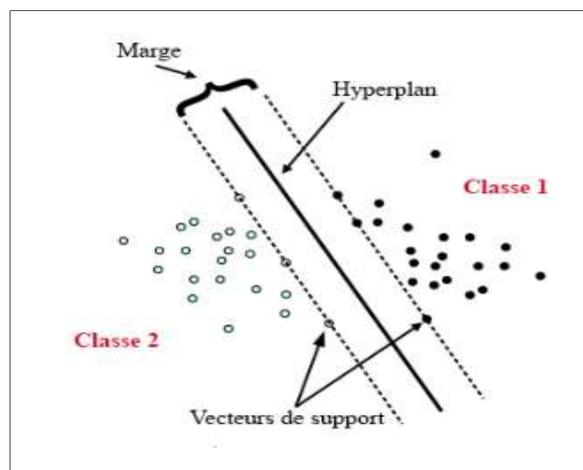


Figure 4.9: Classification avec la machine à vecteurs de support.

Étant donné un ensemble de données (x_i, y_i) , $i = 1, \dots, n$; où $x_i \in \mathbb{R}^n$ et $y_i \in [-1,1]$. x_i est le vecteur de caractéristiques, à n dimensions, qui décrit le point de données, et y_i est l'étiquette correspondante de x_i , dont nous avons besoin, pour créer un hyperplan, entre les points de données et leurs étiquettes. L'hyperplan est un classificateur linéaire, qui permet la séparation des deux classes.

Si on considère une classification binaire, x_i est positif, son y_i vaut 1 ; sinon, y_i vaut -1. La solution du problème consiste à trouver une fonction $f(x)$, capable à prédire le y , pour chaque x donné.

Les données sont linéairement séparables, ou bien qu'elles ne soient pas linéairement séparables (figure 4.10).

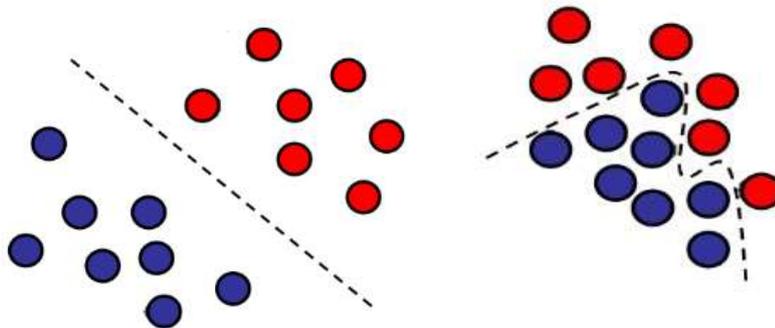


Figure 4.10: Représentation des données. (a) Données linéairement séparables. (b) Données non linéairement séparables.

La surface de décision d'un classifieur SVM linéaire est décrite par (eq. 4.5) :

$$y(x) = w x - b = 0. \quad (4.5)$$

Le vecteur de poids des caractéristiques w , et le seuil b sont choisis de telle sorte que la marge, ou distance entre les vecteurs supports, soit maximisée (figure 4.9). Les vecteurs supports reposent sur deux hyperplans parallèles décrits par $y(x) = 1$ et $y(x) = -1$, tels que la distance entre eux est de $2/ \|w\|$.

La phase d'apprentissage se fait, en résolvant le problème d'optimisation primal suivant (eq. 4.6) :

$$\left\{ \begin{array}{l} \text{minimiser} \quad \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \\ \text{sous contraintes} \\ y_i(w x_i + b) \geq 1 - \xi_i, \quad i = 1, \dots, N \\ \xi_i \geq 0 \end{array} \right. \quad (4.6)$$

Où ξ_i sont appelées variables d'écart, et C représente un hyperparamètre, qui contrôle le degré d'erreur de classification. Le problème d'optimisation sous sa forme duale lagrangienne est (eq. 4.7) :

$$L(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j x_i \cdot x_j \quad (4.7)$$

Sous les contraintes :

$$C \geq \alpha_i \geq 0, \quad i = 1 \dots N ; \quad \sum_{i=1}^N \alpha_i \alpha_j = 0$$

Avec α est le multiplicateur de Lagrange.

Après avoir résolu le problème d'optimisation double, à savoir trouver le α_i , w et b sont calculés par (eq. 4.8) :

$$w = \sum_{i=1}^N \alpha_i y_i x_i \quad (4.8)$$

Sous les contraintes : $b = w \cdot x_k - y_k$, pour $\alpha_i \geq 0$

La méthodologie SVM, peut facilement être étendue, aux tâches de classification non linéaires. En remplaçant chaque vecteur de données x_i , par une application non linéaire $\varphi(x_i)$, et en définissant un noyau comme, $K(x, y) = \varphi(x) \cdot \varphi(y)$, le problème d'optimisation, sous sa forme duale lagrangienne, peut être réécrit comme (eq. 4.9) :

$$L(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad (4.9)$$

Le problème d'optimisation consiste à maximiser $L(\alpha)$ sous les contraintes :

$$C \geq \alpha_i \geq 0, \quad i = 1 \dots N$$

$$\sum_{i=1}^N \alpha_i \alpha_j = 0$$

L'un des noyaux, les plus couramment utilisés dans la littérature, est le noyau de la fonction de base gaussienne ou radiale, défini comme suit :

$$K(x, z) = \exp\left(-\frac{\|x - z\|^2}{2\sigma^2}\right) \quad (4.10)$$

4.6.2 Processus de fusion MRI multimodale

L'algorithme de fusion de détection automatique, développé à l'aide de la MRI multimodale, consiste à fusionner les vecteurs de caractéristiques, extraites de chaque carte (FA, MD et GM), dans un seul vecteur. Les trois CNN optimaux (FA-CNN, MD-CNN et GM-CNN), sont utilisés, pour extraire ces caractéristiques, d'où plusieurs expériences de procédures de fusion, (FA et MD), (FA et GM), (MD et GM) et (FA et MD et GM), sont effectuées, pour choisir la meilleure carte. En conséquence, la taille de la matrice de caractéristiques fusionnées, pour (FA et MD), (FA et GM), (MD et GM), est de 3200×4 , et de 3200×6 , pour FA + MD + GM.

4.7 CONCLUSION

Dans ce chapitre, nous avons décrit la méthode utilisée, pour classifier AD vs CN, AD vs MCI et CN vs MCI, à l'aide de DTI et sMRI. Ces deux modalités sont complémentaires, d'où le sMRI détecte l'atrophie, au niveau de la matière grise, et les changements structuraux, au niveau du volume du cerveau. De plus, l'imagerie DTI est un marqueur de prédiction, utile pour la détection précoce, de la détérioration de la WM. Dans ce contexte, nous visons à détecter, des modèles de changements, micro et macrostructuraux, dans les différents stades de l'Alzheimer, en utilisant un processus de fusion multimodale. Nous proposons dans cette section, une nouvelle méthodologie, qui consiste en un nouveau CNN, pour extraire les caractéristiques, pertinentes des mesures DTI, et des images GM prétraitées. Après cela, ces caractéristiques sont fusionnées, et transmises au classifieur SVM, pour identifier AD de MCI, AD de CN et MCI de CN.

CHAPITRE 5

RESULTATS ET DISCUSSION

5.1 INTRODUCTION

Comme nous l'avons vu au chapitre 4, notre approche consiste en un nouveau système d'aide au diagnostic, composé d'une étape de prétraitement, d'une étape de sélection de coupes, d'une étape d'extraction de caractéristiques, et d'une étape de classification. Cependant, notre objectif principal, est de voir, si l'utilisation des images DTI est plus précieuse et plus puissante, que l'imagerie structurelle, ou si l'inverse, ou si la combinaison de ces deux modalités est la meilleure.

Ce chapitre synthétise les résultats de notre approche, en commençant par une explication des méthodes d'évaluation utilisées, puis l'expérimentation menée, avec les résultats obtenus. Nous terminons le chapitre par une discussion, à partir de laquelle, nous concluons, si la mono-modalité est meilleure, que la multimodalité, ou vice versa.

5.2 EVALUATION DE LA CLASSIFICATION

Une autre partie très importante, de l'apprentissage automatique, est la façon de savoir, si un programme informatique a donné des résultats appropriés, ou s'il a induit à des erreurs.

L'évaluation des tâches de classification, se fait normalement, en divisant l'ensemble des données, en un ensemble de données d'apprentissage, et en un ensemble de données de test (dans le cas de l'apprentissage supervisé). L'algorithme d'apprentissage automatique, est ensuite entraîné sur le premier, tandis que l'ensemble des données de test, est utilisé, pour calculer des indicateurs de performance, afin d'évaluer la qualité de l'algorithme.

Un problème courant, pour les algorithmes d'apprentissage automatique, réside dans le fait que, l'accès aux données de test et d'entraînement, est limité. Afin de résoudre ce problème, une approche courante, consiste à utiliser une validation croisée (X-Folder). La validation croisée, décrit le processus de division, de l'ensemble des données, en X parties, et l'utilisation de chacune d'elles, séquentiellement, comme ensemble de données de test, tout en

combinant les autres, aux données d'apprentissage (Figure 5.1). Ensuite, les indicateurs de performance sont moyennés, sur tous les processus de validation. Il n'y a pas d'indicateur parfait, pour chaque sujet, concernant l'évaluation des algorithmes d'apprentissage automatique, car chaque méthode a ses défauts et ses avantages.

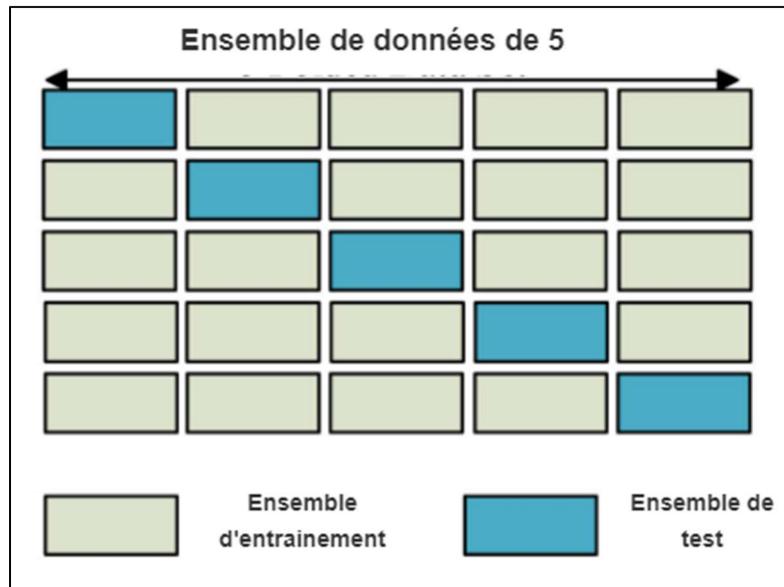


Figure 5.1: Schéma de validation croisée à 5 folders.

5.2.1 Métriques d'évaluation

Les facteurs les plus importants, pour évaluer les performances d'un programme d'apprentissage automatique, sont :

- la matrice de confusion (en tant qu'outil de visualisation des performances de classification, à partir de laquelle des métriques d'évaluation peuvent être calculées) ;
- la précision ;
- le rappel (sensibilité) ;
- la spécificité ;
- la caractéristique de fonctionnement du récepteur (ROC).

Matrice de confusion : La matrice de confusion, fournit un résumé des instances, classées sous la forme d'un tableau (tableau 5.1), où les lignes représentent les instances d'une classe prédite, et les colonnes rapportent les instances, qui appartiennent à une classe vrai/faux. Dans le cas d'un problème de classification binaire, les instances peuvent être catégorisées, en positives ou négatives. Les termes qui sont trouvés dans une matrice de confusion sont: les

vrais positifs (VP), les vrais négatifs (VN), et les faux positifs (FP), faux négatifs (FN), respectivement.

Plus précisément, VP désigne des instances positives, qui sont correctement classées, comme positives, tandis que FP fait référence à des instances négatives, qui sont incorrectement classées, comme positives. De même, VN désigne les instances négatives, qui sont correctement classées, comme négatives, tandis que FN fait référence à des instances positives, qui sont incorrectement classées, comme négatives.

Tableau 5.1: Une représentation indicative d'une matrice de confusion pour un classifieur binaire. Toutes les prédictions correctes sont situées dans la diagonale de la matrice.

Valeurs réelles	Valeurs prédites	
	Positive	Négative
Positive	VP	FN
Négative	FP	VN

La précision (Accuracy) : est une mesure, qui calcule la somme des instances, qui ont été correctement classées, sur le nombre total de toutes les instances.

$$Accuracy = \frac{\text{Prédictions correctes}}{\text{le nombre total des images}} = \frac{VP + VN}{VP + VN + FP + FN}$$

En général, plus ce nombre est proche de 1, meilleure est la performance. Cela est vrai non seulement pour le score F1, mais aussi pour le rappel, la précision et l'exactitude. Cependant, la qualité d'un modèle ne doit pas être basée uniquement sur ces métriques, mais également, être comparée à une autre ligne de base, ou à des études antérieures.

La spécificité : est la fraction du nombre d'instances, de vrais négatifs, sur la somme des instances de vrais négatifs, et de faux positifs.

$$Spécificité = \frac{VN}{VN + FP}$$

La sensibilité : est la fraction du nombre d'instances, de vrais positifs, sur la somme des instances de vrais positifs, et de faux négatifs.

$$Sensibilité = \frac{VP}{VP + FN}$$

Les courbes ROC : sont des illustrations graphiques, qui évaluent la performance d'un test diagnostique. Le graphique d'une courbe ROC est produit, par le tracé de la sensibilité, sur l'axe des ordonnées, par rapport à la spécificité, sur l'axe des abscisses, pour différents seuils.

Les courbes ROC sont largement utilisées, pour la prise de décision médicale [109], comme outil de diagnostic complémentaire, car la précision de la classification seule, n'est souvent pas une métrique d'évaluation objective. Traditionnellement, les courbes ROC sont utilisées, pour la classification binaire, mais il est possible de les appliquer également, dans les tâches de classification multi-classes. Le seul changement, est le nombre de classes.

Afin de mesurer la capacité d'un test, à comparer, si une condition spécifique, est présente ou non, l'aire sous la courbe ROC (AUC) est calculé. L'AUC signifie la probabilité, que le classifieur classe une instance positive, choisie au hasard, plus haut qu'une instance négative, choisie au hasard" [152]. En termes de performances moyennes, le classifieur, avec la plus grande AUC, est considéré comme le meilleur, 1 étant la valeur la plus élevée, pouvant être obtenue.

5.3 EXPERIMENTATION ET RESULTATS

L'évaluation des changements physiopathologiques, par neuroimagerie, est essentielle, pour la détection précoce de l'AD. La mono-modalité ne peut pas fournir suffisamment d'informations, par conséquent, MRI multimodale peut être utile, dans la détection de la maladie d'Alzheimer. sMRI et le DTI ont reçu plus d'attention, ces dernières années, pour étudier la progression de la maladie d'Alzheimer, comme nous l'avons vu dans le chapitre 2. Ces deux modalités sont complémentaires : le sMRI détecte l'atrophie de la matière grise, et les changements dans le volume du cerveau, et le DTI est un marqueur de prédiction, utile pour détecter la détérioration de la WM. Dans ce contexte, nous visons à détecter des modèles de changements, micro et macrostructuraux, dans les différents stades de la maladie d'Alzheimer, en utilisant le processus de fusion multimodalité MRI (sMRI et DTI).

Dans ce travail, plusieurs expériences sont réalisées, pour valider l'efficacité de la méthode que nous avons proposée, pour classer (AD vs. CN), (CN vs. MCI) et (AD vs. MCI).

➤ **Dans la première expérience**, nous avons effectué une classification unimodale directe des caractéristiques, extraites de FA, MD et GM. Cela nous donne des informations, sur la meilleure modalité et carte.

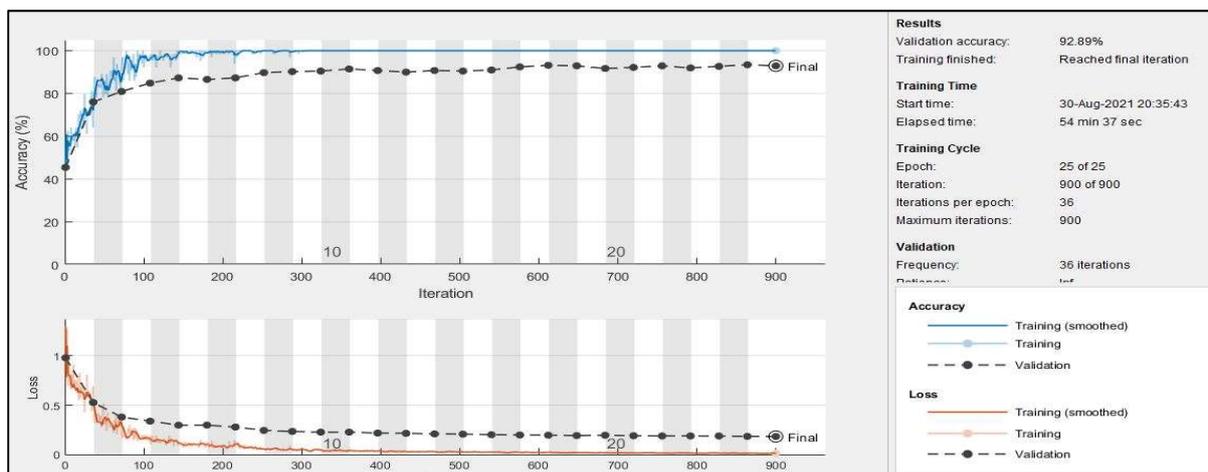
- **Dans la deuxième expérience**, nous étudions, si la multimodalité augmente ou non les performances, et permet une meilleure discrimination, entre les différentes classes. Ceci est réalisé, en étudiant l'impact de la fusion des caractéristiques des deux modalités.

Le 2DCNN-SVM proposé est implémenté, en utilisant MATLAB ver. R2019a, et fonctionnant sur un processeur Intel-i7, à 3,1 GHz, 16 Go de RAM.

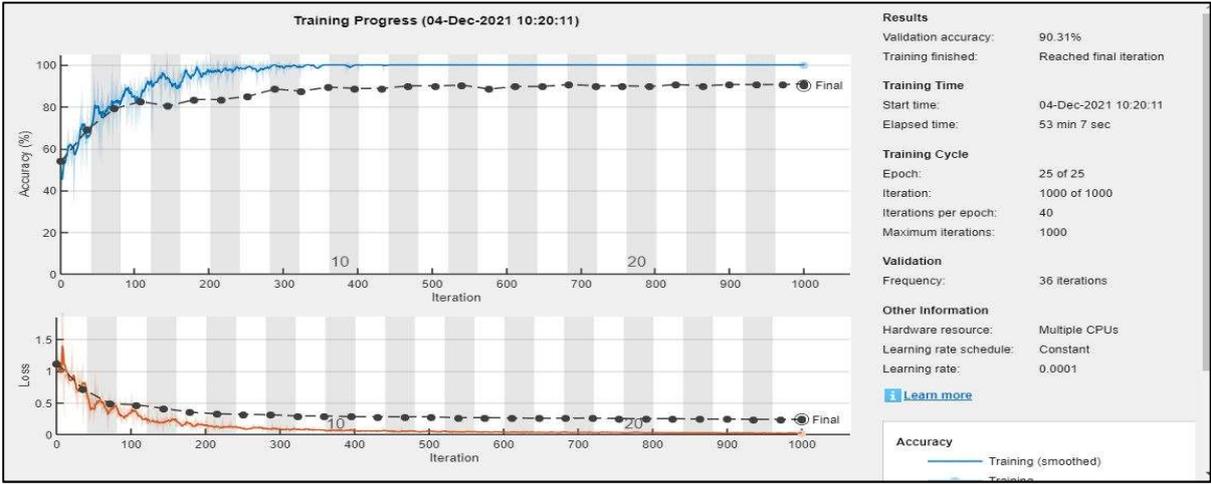
Le modèle CNN est entraîné, à l'aide d'un SGDM (Stochastic Gradient Descent Momentum), optimisé, en utilisant l'algorithme de rétropropagation, et l'entropie croisée, comme fonction de perte. La taille du lot est de 64, le taux d'apprentissage est de 0,0001 pour 25 époques.

Il y a un total de 3200 images, de chaque carte (FA, MD et GM), 1600 images pour chaque classe. L'ensemble de données est divisé, en 70 % pour l'entraînement, 15 % pour la validation, et les 15 % restants, pour le test de SVM. La même architecture CNN est utilisée, pour entraîner les coupes FA, les coupes MD, et les coupes GM.

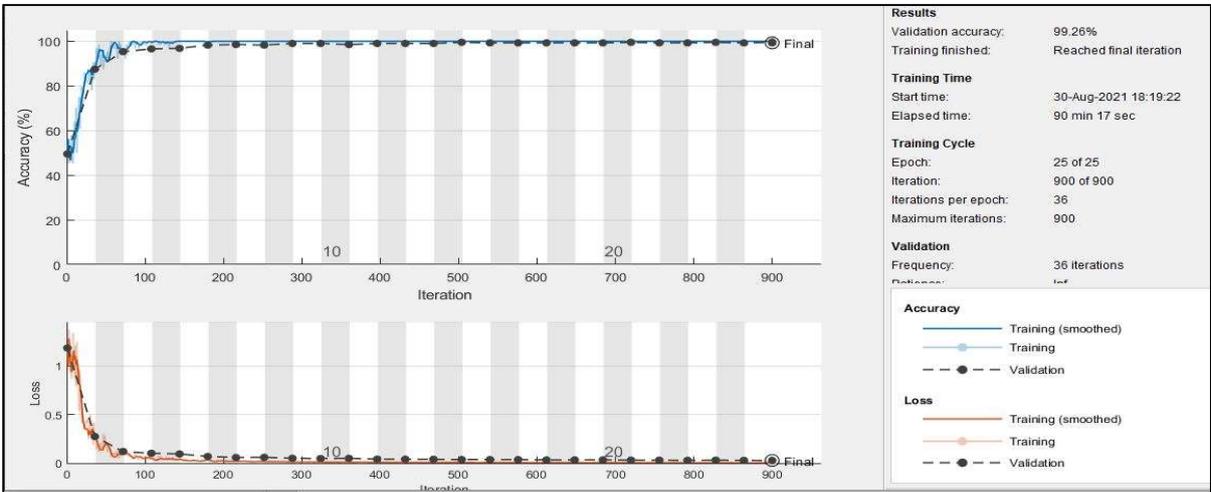
Les figures 5.2, 5.3, et 5.3 illustrent la phase d'apprentissage des CNN, avec 70% de la base de données utilisées, d'où, pour chaque tâche de classification binaire (AD vs CN, AD vs MCI, CN vs MCI), l'apprentissage est effectué, en utilisant les images FA, MD et GM. Ce qui nous donne 3 modèles CNN entraînés, pour chaque classification binaire.



(a)

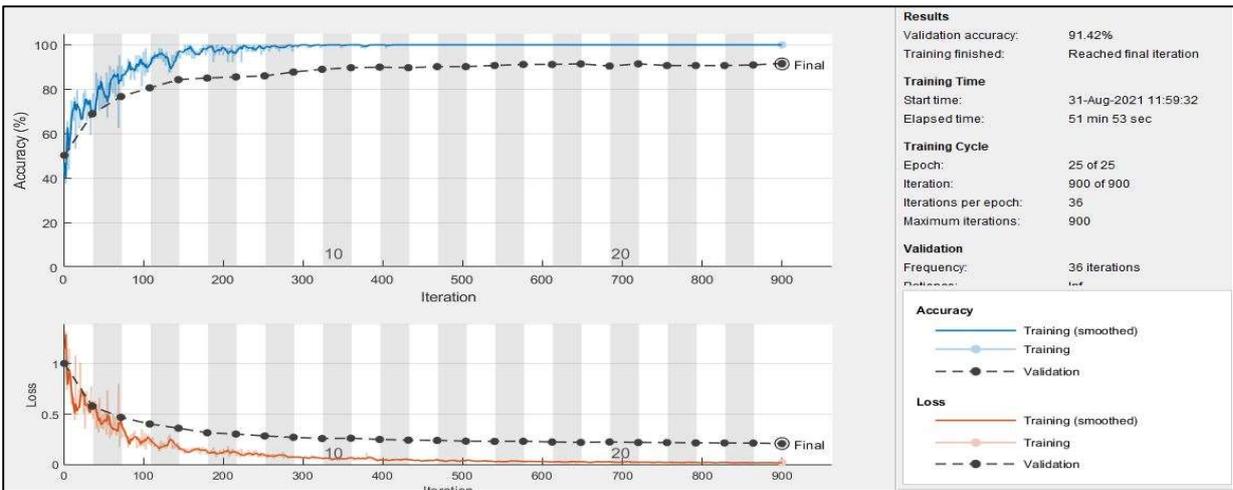


(b)

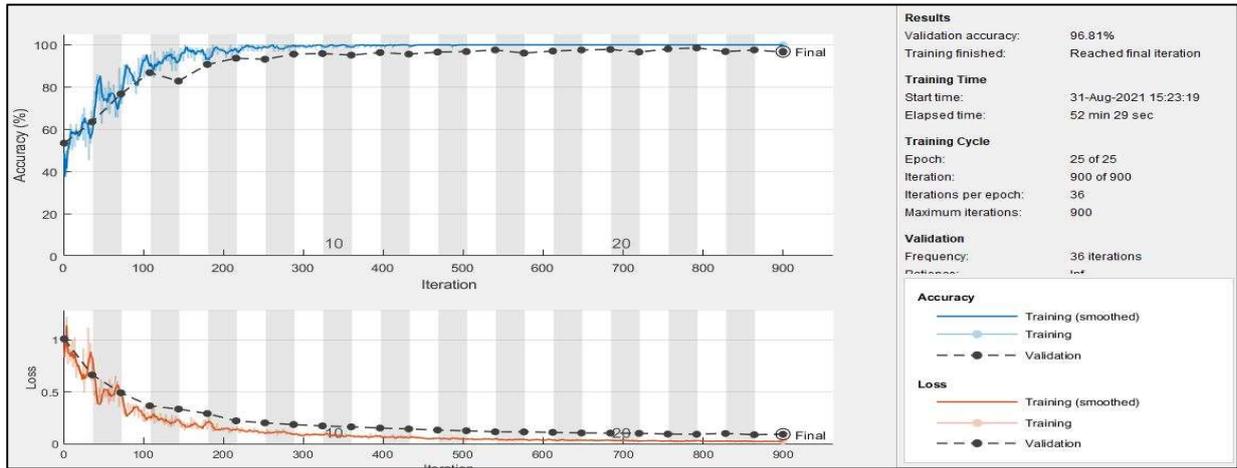


(c)

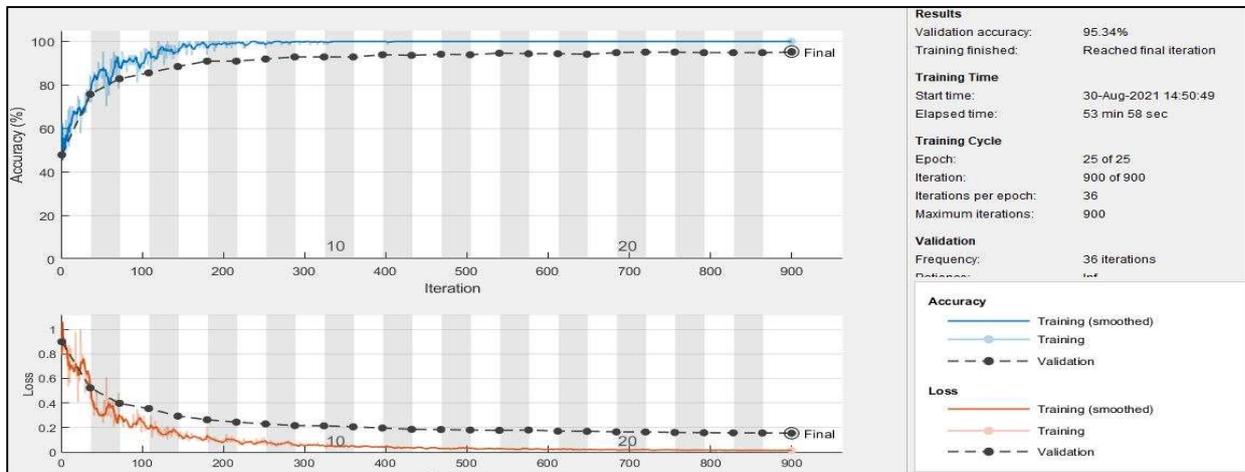
Figure 5.2: La courbe d'apprentissage des CNN, pour la classification binaire CN vs AD, en utilisant : (a) les images DTI-FA, (b) les images DI-MD, (c) les images GM.



(a)

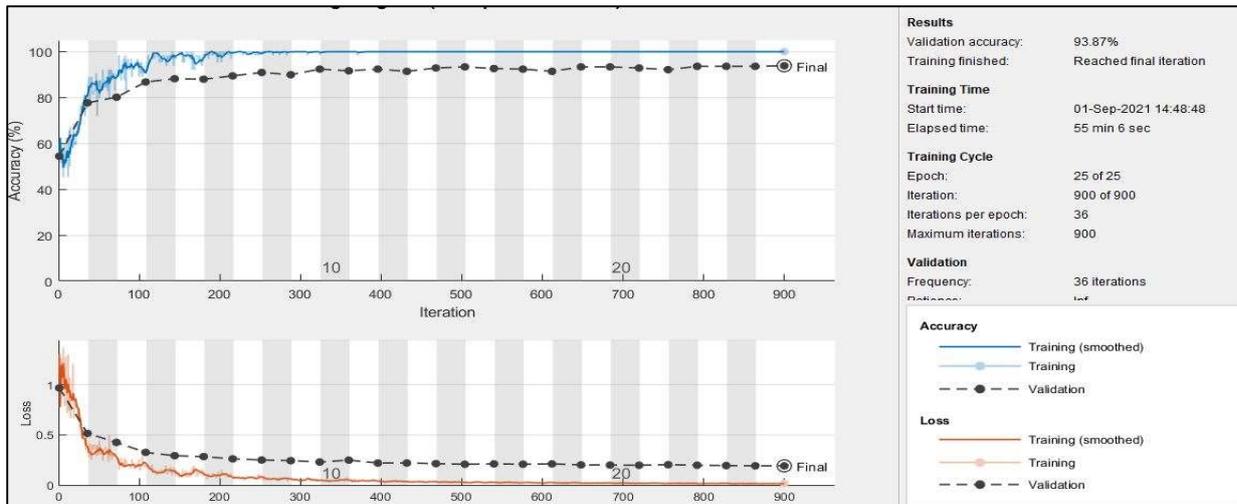


(b)

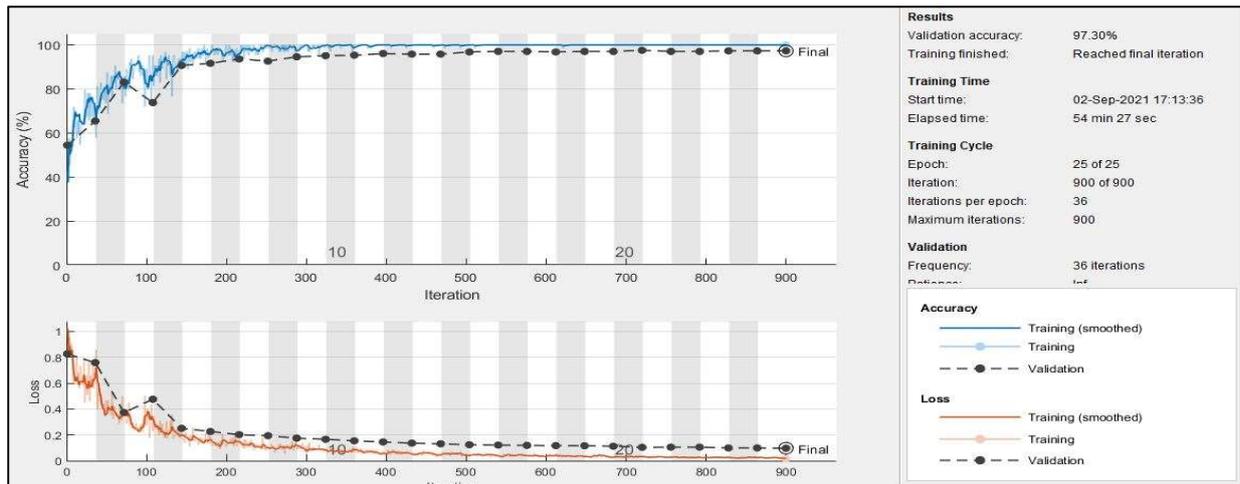


(c)

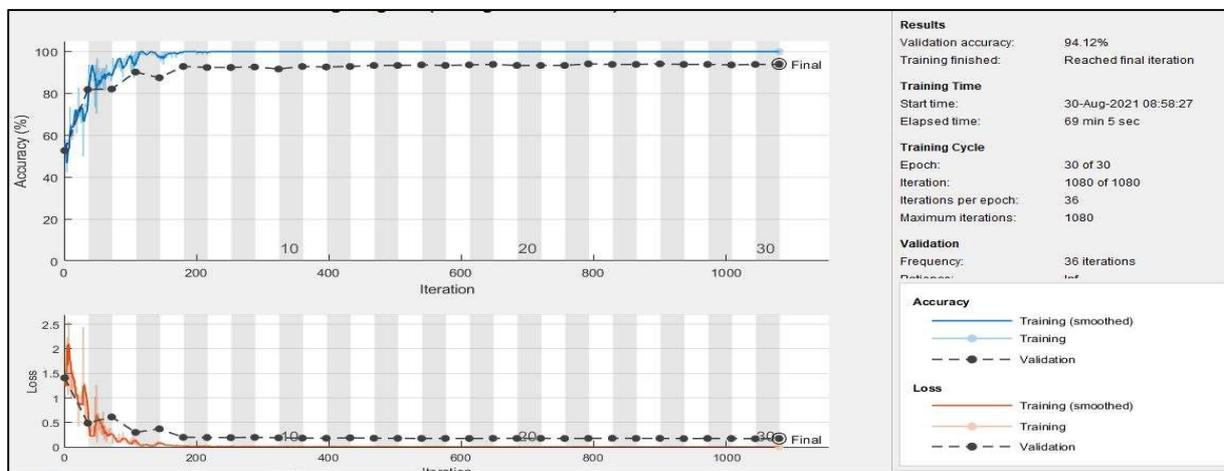
Figure 5.3: La courbe d'apprentissage des CNN, pour la classification binaire CN vs MCI, en utilisant : (a) les images DTI-FA, (b) les images DI-MD, (c) les images GM.



(a)



(b)



(c)

Figure 5.4: La courbe d'apprentissage des CNN, pour la classification binaire AD vs MCI, en utilisant : (a) les images DTI-FA, (b) les images DI-MD, (c) les images GM.

Pour la classification par SVM, les données extraites sont classées, en données d'apprentissage, de validation, et de test. Nous avons utilisé les caractéristiques, extraites de 2720 images, pour l'apprentissage et de 480 images, pour le test. Le meilleur score est obtenu, par l'utilisation de SVM, avec un noyau gaussien, et une validation croisée à 10 folders.

5.3.1 Résultat

Les performances de notre méthode, sont validées, à l'aide de la précision, et de l'aire sous la courbe caractéristique de fonctionnement du récepteur (AUC). Les résultats de validation

sont illustrés dans le tableau 3, et les courbes ROC de la validation croisée 10 fois, sont présentées dans les figures (5.5, 5.6 et 5.7).

Tableau 5.2: Performance de l'ensemble de validation

Classes	Map	FA	MD	FA+MD	GM	GM+FA	GM+MD	GM+FA+MD
AD/CN	Acc	095.48	97.94	97.79	96.91	99.12	98.49	99.52
	AUC	098.79	99.52	98.91	99.51	99.96	98.97	99.99
AD/MCI	Acc	92.98	96.21	96.95	96.87	98.49	98.68	99.08
	AUC	96.42	99.02	99.54	99.31	99.98	99.93	99.96
CN/MCI	Acc	93.31	92.76	94.85	93.03	97.17	95.99	97.79
	AUC	96.43	95.18	98.83	97.72	99.06	98.65	99.60

La fusion des caractéristiques FA, MD et GM a amélioré le résultat, et a surpassé la modalité unique, et la procédure de fusion sMRI + MD, adoptée dans de nombreuses études antérieures [30, 34, 35].

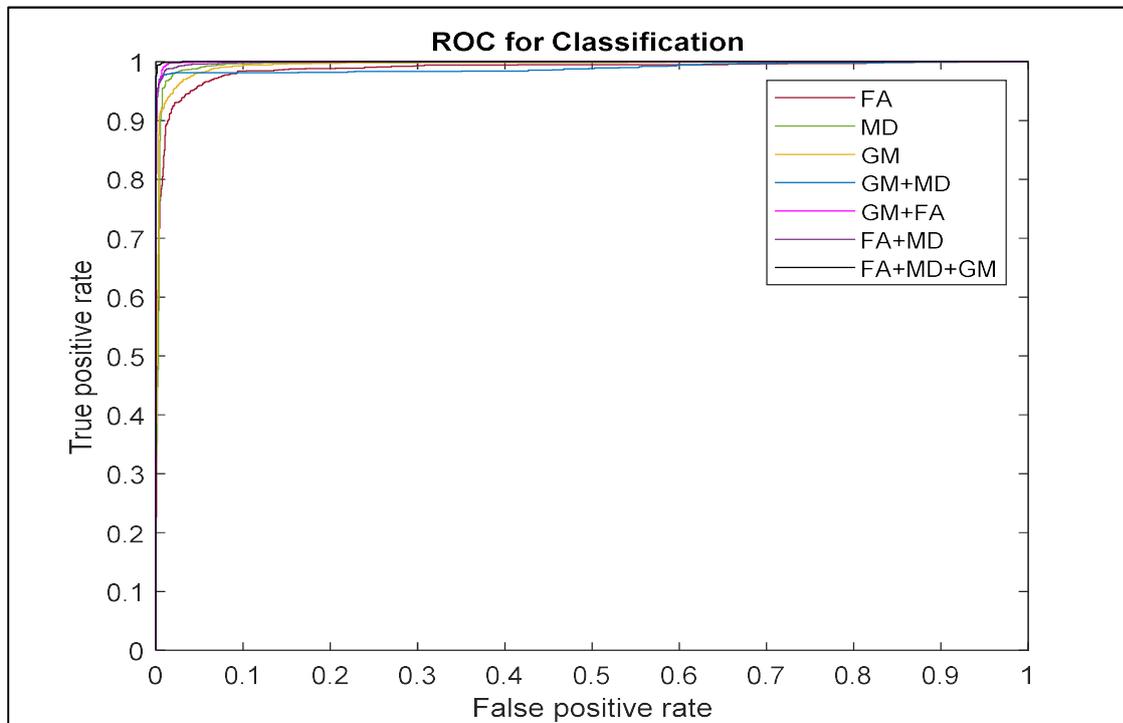


Figure 5.5: La courbe ROC de la classification AD vs CN.

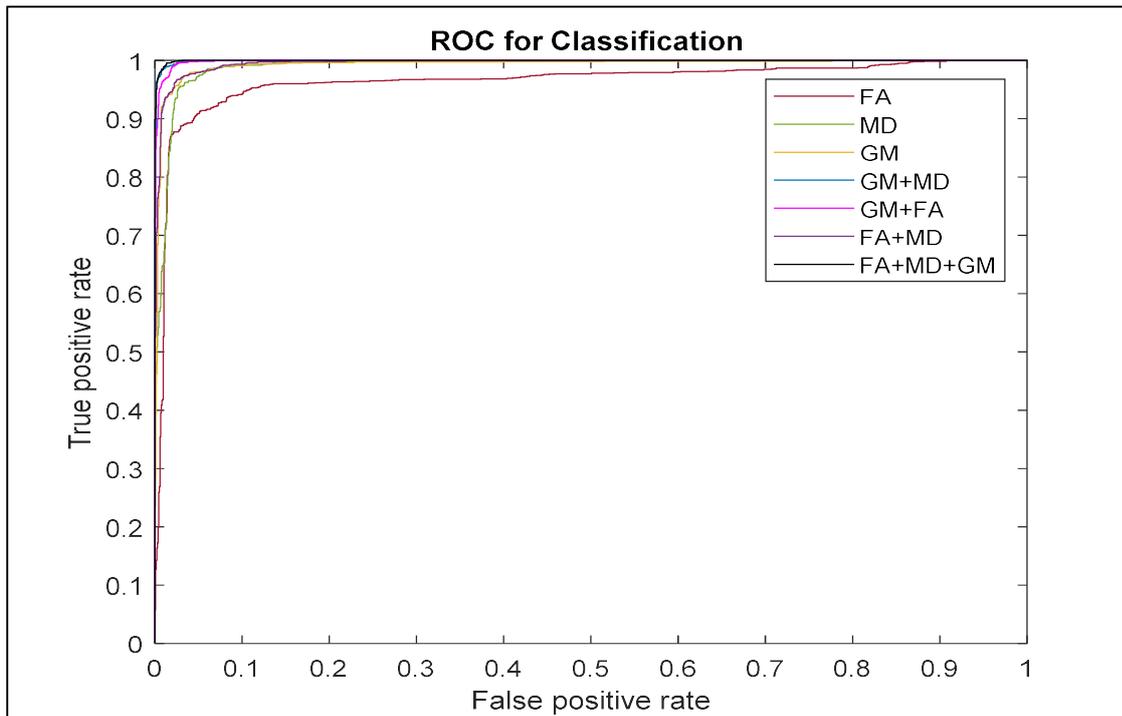


Figure 5.6 : La courbe ROC de la classification AD vs MCI.

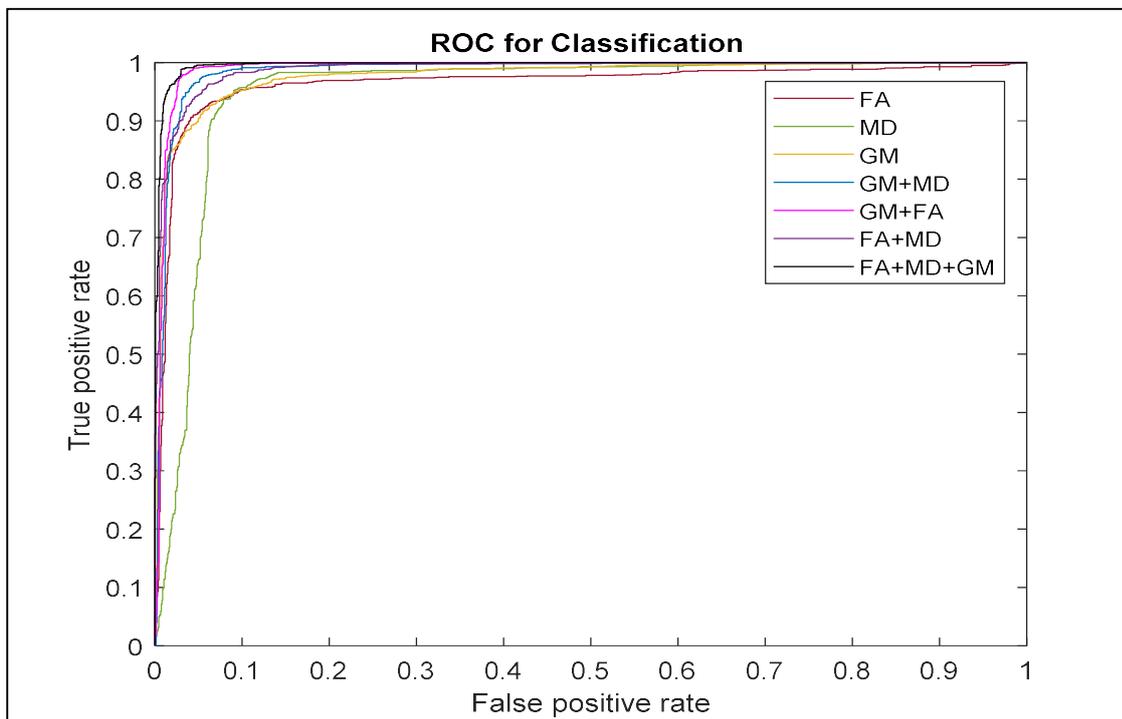
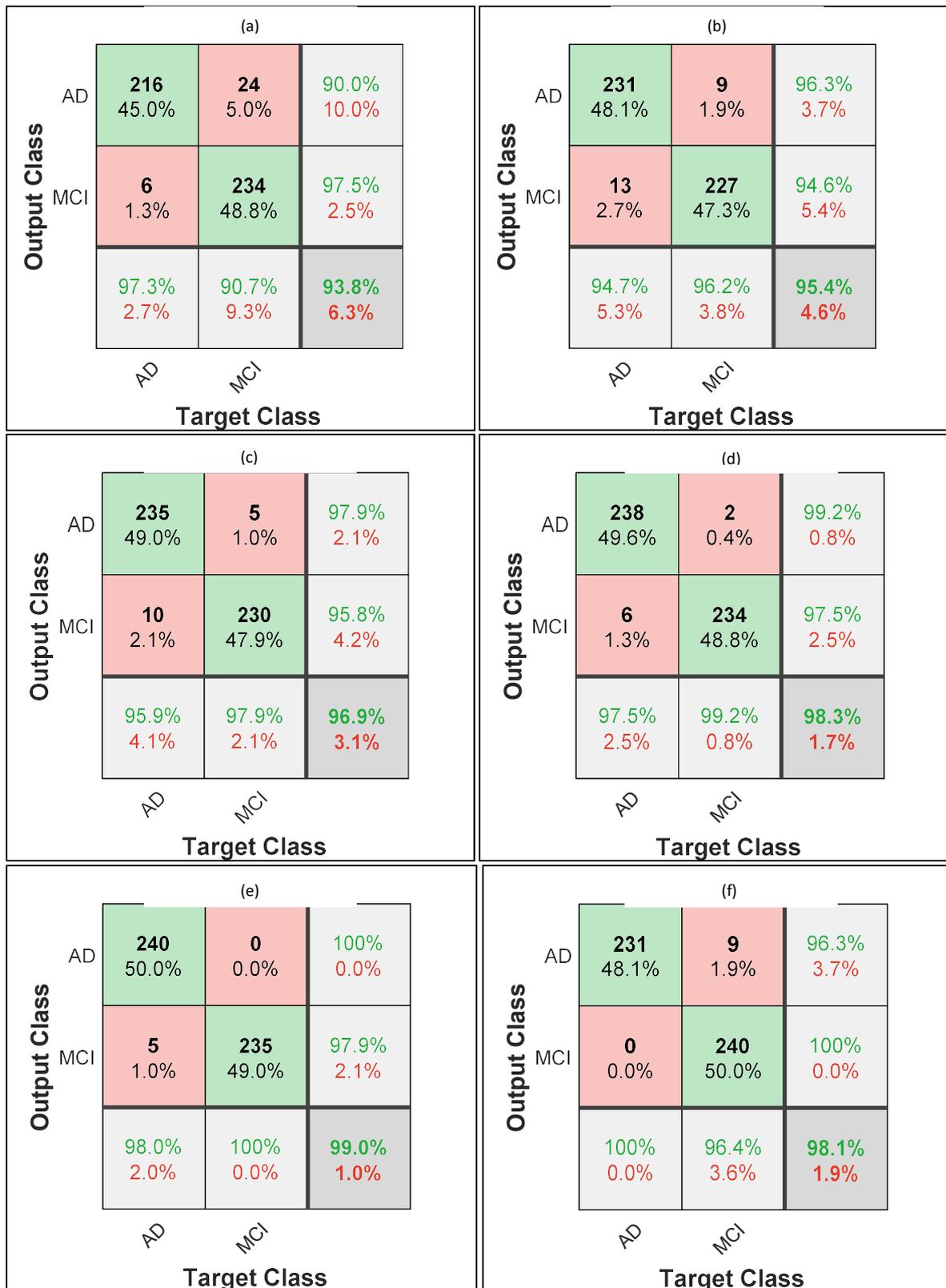


Figure 5.7 : La courbe ROC de la classification AD vs MCI.

Nous avons testé notre méthode, en utilisant 240 images AD, 240 images CN, et 240 images MCI. Les métriques d'évaluation utilisées sont, la précision, la sensibilité et la spécificité,

déterminées par les matrices de confusion. Les matrices de confusion sont présentées dans les figures (5.8 et 5.9). Tous les résultats des tests sont résumés dans le tableau 4.



		(g)		
Output Class	AD	240 50.0%	0 0.0%	100% 0.0%
	MCI	2 0.4%	238 49.6%	99.2% 0.8%
		99.2% 0.8%	100% 0.0%	99.6% 0.4%
		AD	MCI	
		Target Class		

Figure 5.8: Matrices de confusion de la classification AD vs MCI, en utilisant des caractéristiques, extraites de (a) FA, (b) MD, (c) GM, (d) FA+MD, (e) GM+FA, (f) MD+GM, (g) FA+MD+GM

		(a)			(b)		
Output Class	CN	224 52.0%	16 3.7%	93.3% 6.7%	227 52.7%	13 3.0%	94.6% 5.4%
	MCI	17 3.9%	174 40.4%	91.1% 8.9%	21 4.9%	170 39.4%	89.0% 11.0%
		92.9% 7.1%	91.6% 8.4%	92.3% 7.7%	91.5% 8.5%	92.9% 7.1%	92.1% 7.9%
		CN	MCI		CN	MCI	
		Target Class			Target Class		

		(c)			(d)		
Output Class	CN	229 53.1%	11 2.6%	95.4% 4.6%	231 53.6%	9 2.1%	96.3% 3.7%
	MCI	17 3.9%	174 40.4%	91.1% 8.9%	15 3.5%	176 40.8%	92.1% 7.9%
		93.1% 6.9%	94.1% 5.9%	93.5% 6.5%	93.9% 6.1%	95.1% 4.9%	94.4% 5.6%
		CN	MCI		CN	MCI	
		Target Class			Target Class		



Figure 5.9: Matrices de confusion de la classification CN vs MCI, en utilisant des caractéristiques, extraites de (a) FA, (b) MD, (c) GM, (d) FA+MD, (e) GM+FA, (f) MD+GM, (g) FA+MD+GM.

Le tableau 5.3, calculé par l'utilisation des matrices de confusion, présentées dans les figures 5.8 et 5.9, montre que les cartes FA, MD et GM, sont utiles et importantes, pour la discrimination des différents stades AD. Pour l'utilisation de FA, MD, GM indépendamment, nous rapportons que, MD a obtenu le meilleur résultat, dans le cas d'AD vs CN, avec une précision de 98,96 %. Cependant, la carte GM donne de meilleurs résultats, dans la classification AD vs MCI, et CN vs MCI, avec une précision de 96,88%, et 93,50%, respectivement. Nous avons étudié la meilleure combinaison de caractéristiques (FA et MD, FA et GM, et MD et GM). FA et MD fusionnés, ont surpassé les autres caractéristiques combinées avec une précision de 99,98 % et 98,33 %, dans la classification AD vs CN, et AD vs MCI.

Tableau 5.3: Évaluation des performances de la technique 2DCNN-SVM proposée sur l'ensemble de données de test.

Classes	Performances	FA	MD	FA+MD	GM	GM+FA	GM+MD	GM+FA+MD
AD/CN	Précision	97.50	98.96	99.89	96.25	98.50	99.38	99.79
	Sensitivité	97.50	99.00	98.75	96.25	98.55	99.40	99.80
	Spécificité	97.50	99.00	98.75	96.25	98.50	99.40	99.80
AD/MCI	Précision	93.75	95.42	98.33	96.88	98.96	98.13	99.58
	Sensitivité	94.00	95.45	98.35	96.90	99.00	98.20	99.60
	Spécificité	93.75	95.45	98.35	96.85	98.95	98.15	99.60
CN/MCI	Précision	92.30	92.10	94.40	93.50	95.60	97.00	97.00
	Sensitivité	92.25	92.20	94.50	93.60	95.90	97.20	96.95
	Spécificité	92.20	91.80	94.20	93.25	95.20	96.80	97.00

D'autre part, les GM et MD fusionnés, ont obtenu des résultats plus élevés, pour classer CN par rapport à MCI, avec une précision de 97,00 %, une sensibilité de 97,20 %, et une spécificité de 96,80 %. Par rapport à l'utilisation de la modalité unique, la fusion FA, MD et GM a entraîné une augmentation d'environ 3,54 %, 5,83 %, 4,1 % pour la précision, 3,55 %, 5,85 %, 4,75 % pour la sensibilité, et 2,8 %, 5,85 %, 5,2 % pour la spécificité dans les trois cas de classification (AD vs CN), (AD vs MCI), et (CN vs MCI), respectivement. Ainsi, la multimodalité donne les meilleures performances, comme le montre clairement la figure 5.10.

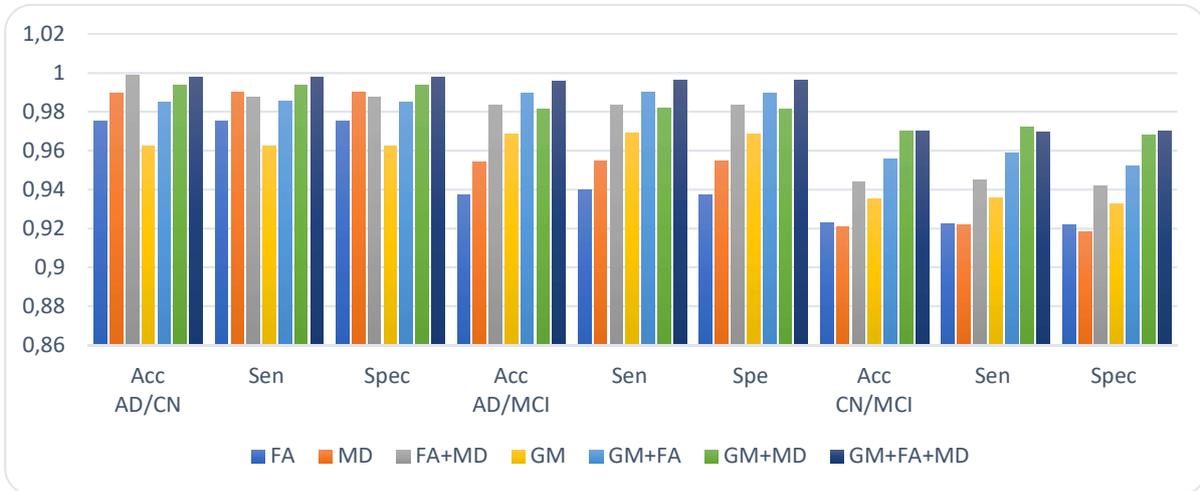


Figure 5.10: Comparaison de la performance de la technique de classification binaire proposée AD vs. CN, AD vs. MCI, and CN vs. MCI.

5.3.2 Comparaison avec les études précédentes

Pour valider la performance et l'efficacité de notre nouveau workflow, nous l'avons comparé aux approches précédentes, présentées dans la littérature, et traitant des mêmes bases de données (ADNI), et des mêmes modalités (sMRI et DTI). Nos résultats ont obtenu une plus grande précision, dans la détection de la maladie d'Alzheimer, par rapport à d'autres études, comme le montre le tableau 5.4. Cependant, la comparaison avec les travaux précédents n'était pas fiable, car la plupart d'entre eux utilisent, soit la modalité DTI, soit l'imagerie structurale. D'autre part, la plupart des travaux antérieurs combinent l'imagerie structurale, avec les images MD- DTI, et négligent la carte FA et ne se limitent pas à la région d'intérêt.

Tableau 5.4: Comparaison des résultats avec les techniques de pointe appliquées à la détection de l'AD.

Travaux	Modalités	Nombre de sujets			Méthode	Précision(%)		
		AD	CN	MCI		AD/CN	CN/MCI	MCI/AD
Lella et al. 2017 [67]	DTI	40	-	40	FA, MD, RF SVM, Bayes naïf	78.0	-	-
Maggipinto et al. 2017 [103]	DTI	89	90	90	TBSS pour FA et MD, RF	87.0	81.0	-
Ahmed et al. 2017 [46]	DTI+sMRI	45	52	58	Hippocampe (ROI) (sMRI,MD), CSF volume, SVM	90.2	79.42	76.63
Aderghal et al. 2020 [111]	DTI+sMRI	64	399	273	hippocampe ROI (sMRI,MD), CNN – LeNet	92.3	78.48	79.16
Jiang et al. 2020	sMRI	50-70(EMCI)			VGG16-SVM	-	89.4	-
Marzban et al. 2020	DTI+sMRI	115	185	106	MD, FA, MO, GM, 2DCNN	93.5	-	79.6

[109]							
Kang et al. 2020 [112]	DTI+sMRI	- 50 70	FA+MD+sMRI, VGG16, LASSO, SVM	-	94.2	-	
Feng et al. 2020 [84]	sMRI	130 135 133	3D-CNN-SVM	99.1	98.90	89.40	
Lella et al. 2021 [106]	DTI	40 - 40	FA+MD+RD+LD; SVM, RF, PML	88.5			
Naz et al. 2021 [153]	sMRI	95 146 138	VGG16, VGG19	98.89	97.06	99.27	
Notre approche	DTI+sMRI	50 50 50	FA, MD, GM 2DCNN-SVM	99.79	99.58	97.00	

5.4 DISCUSSION

Afin de proposer une méthode de détection de la maladie d'Alzheimer, sur les images DTI, ou sMRI, nous sommes passés par plusieurs étapes.

➤ **Prétraitement**

Le prétraitement est une étape cruciale, dans la création des systèmes d'aide à la détection. Elle facilite la tâche, et elle est décisive, pour les phases de classification, et d'interprétation, car elle permet de produire une image plus lisible, et moins bruitée. Le prétraitement est effectué d'une façon standard, à l'aide de FSL et SPM, qui sont les deux boîtes à outils les plus utilisées en neuroimagerie.

➤ **Sélection des coupes**

En général, nos résultats, concernant la détection précoce de l'AD, impliquent l'existence de processus physiopathologiques distincts. En fait, l'hippocampe est connu, pour être l'une des structures les plus précoces, et la plus gravement endommagée, affectée par l'AD. Cependant, il existe d'autres structures, impliquées dans la détection de l'AD, telles que l'amygdale, le thalamus et le putamen [3].

La sélection des coupes pertinentes, semble une méthode puissante, et plus facile, que la segmentation de l'hippocampe, ou d'autres régions du cerveau, qui nécessite l'intervention d'un expert humain. Notre réseau apprend les modèles complexes d'atrophie cérébrale, à partir de sections pertinentes, qui contiennent, presque toutes, les régions touchées par AD,

mentionnées dans la littérature, pour chaque patient. Cela élimine le processus de segmentation de l'hippocampe, et d'autres régions du cerveau. De plus, une sélection ultérieure des caractéristiques, les plus discriminantes, est évitée dans notre approche.

➤ **Architecture CNN**

L'apprentissage par transfert, basé sur la sMRI, a donné des résultats impressionnants [153, 154]. Généralement, les résultats des travaux, qui utilisent les modèles VGG16 et VGG19, ont surpassé les autres modèles pré-entraînés [153]. Récemment, certains auteurs [82] ont réussi à utiliser ces modèles pré-entraînés, comme un extracteur automatique des caractéristiques, et ont utilisé SVM, pour la classification ; ils ont atteint une plus grande précision, qu'avec l'utilisation de la couche Softmax du CNN.

Cependant, la technique d'apprentissage par transfert, repose généralement sur des images naturelles, dont les modèles sont entraînés, à l'aide de la base de données Imagenet. A l'inverse, nos réseaux simples apprennent, et extraient de zéro, les caractéristiques les plus pertinentes. L'architecture proposée est simple à réaliser, et n'exige pas de hardware puissant, pour effectuer l'apprentissage, comme les architectures populaires, décrites dans le chapitre 3.

➤ **Application sur les images FA-DTI, MD-DT, GM**

Nos résultats confirment l'efficacité de la mesure DTI-FA et DTI-MD, dans la classification AD vs CN, AD vs MCI et CN vs MCI, ce qui est cohérent avec les travaux antérieurs [103, 106]. L'utilisation des cartes FA est moins utilisée, par rapporte à l'utilisation de la carte MD. Cependant, nos résultats confirment l'utilité de cette carte, dans la détection précoce de la maladie d'Alzheimer. Ces résultats n'annulent pas l'efficacité de la carte MD, et les résultats obtenus, par la combinaison de ces deux cartes surpassent leur utilisation séparée.

De plus, l'atrophie GM en sMRI est d'un grand intérêt, pour les chercheurs, pour la détection précoce de l'AD. Pour cette raison, nous n'avons segmenté que la matière grise, pour concentrer notre étude, uniquement sur cette partie du cerveau. En parallèle, la DTI est utilisée, pour l'étude de la matière blanche. Nos résultats confirment la puissance de la GM, dans la classification de la maladie d'Alzheimer, surtout dans la classification binaire AD/CN. Cela revient à la grande différence qu'il y a, entre un cerveau sain et un cerveau atteint de l'Alzheimer, due au changement au niveau du volume du cerveau, et bien sûre à l'atrophie corticale, comme nous l'avons déjà mentionné dans le chapitre 2.

➤ **Extension au cadre MRI multimodal**

Au cours des dernières années, les multi modalités (DTI-MD et sMRI), ont été rapportées, par de nombreux chercheurs. Ils ont proposé différentes techniques de combinaison, pour assurer la meilleure classification [19] [46, 109]. Cependant, ils ont atteint une précision, inférieure à celle que nous avons obtenue, soit plus de 97 %. Cela est probablement dû, à la petite taille de l'échantillon que nous avons utilisé, par rapport à eux, ou au fait que nous n'avons pas travaillé sur une région spécifique, comme hippocampe, qui est la région la plus utilisée, dans la littérature, ou à l'impact de l'ajout de FA. Ceci car la plupart des travaux dans la littérature, combinent, soit les images structurales T1w, soit la GM, avec les images DTI-MD. Cependant la fusion des cartes FA, MD et GM ont donné de meilleurs résultats, que si elles sont utilisées séparément.

En résumé, les cartes de diffusion et la GM, sont des éléments puissants et importants, pour la discrimination des stades de l'AD. Le processus de fusion multimodalité (FA+MD+GM), semble être la meilleure technique, pour améliorer les performances de la classification AD.

5.5 CONCLUSION

Nous avons proposé, un diagnostic assisté par ordinateur, pour la classification de la maladie d'Alzheimer, en utilisant l'imagerie MRI multimodale. Un nouveau CNN est utilisé, pour extraire les caractéristiques des images DTI-FA, DTI-MD et GM. Les caractéristiques profondes, sont concaténées, et transmises au classificateur SVM. L'approche CNN-SVM proposée, démontre l'efficacité de l'utilisation de la MRI multimodale, en atteignant une précision de classification de 99,79 %, 99,85 % et 97,00 %, pour AD/CN, AD/MCI et CN/MCI, respectivement. Les résultats récents montrent que l'approche proposée offre une valeur ajoutée, à la profession de la santé, en permettant une classification précise de la maladie d'Alzheimer, et surtout un diagnostic précoce.

CONCLUSION GENERALE

Dans cette thèse, nous avons essayé d'étudier la classification de la maladie d'Alzheimer, à l'aide de la technique d'imagerie multimodale MRI. Cet outil d'imagerie, pourrait diagnostiquer avec précision, l'état de la maladie d'Alzheimer et son évolution, pendant un certain temps, que ce soit en évaluation clinique ou en recherche.

Les algorithmes d'apprentissage automatique, en particulier son sous-domaine "Deep Learning", ont produit des résultats émergents, pour divers domaines d'application ; le domaine médical en fait partie. À ce stade, il n'y a aucun doute, sur les avantages et les bénéfices de l'utilisation de modèles (Deep Learning), pour la classification AD. Ainsi, dans le travail actuel, nous avons proposé une approche CAD, pour la classification de l'AD, qui intègre des modules de classification hautement automatiques, pour aider les cliniciens à comprendre la pathologie de l'AD, et à prendre des décisions précises et rapides. Ces méthodes consistent à développer des architectures supervisées, robustes de bout en bout, qui ne nécessitent aucun processus préalable d'extraction de caractéristiques, et à les adopter pour l'imagerie MRI cérébrale. Nous avons, spécifiquement, considéré trois catégories, dans l'évaluation de la maladie, et le but est de classer les sujets, en normaux (CN), en pathologiques (AD), ou en déficience cognitive légère (MCI).

Au cours de ce travail, notre objectif était de montrer l'efficacité de l'utilisation de la DTI, dans la détection précoce de l'AD. Nous avons engagé, en premier lieu, les cartes dérivées du DTI (FA et MD), puis les images de la matière grise segmentée, des images T1, pour développer des modèles intelligents, pour la classification AD.

L'imagerie de diffusion est à ce jour, le seul moyen non invasif, permettant l'étude des faisceaux de fibres de la matière blanche du cerveau. Elle fournit des indications précieuses, sur les pathologies, altérant les fibres nerveuses de la substance blanche, non visibles en imagerie conventionnelle. D'autre part, l'atrophie de la matière grise, représente le meilleur indicateur, pour la détection de l'AD.

L'évaluation des changements physiopathologiques, par neuroimagerie, serait essentielle, pour prédire l'AD. La modalité unique, ne peut pas fournir suffisamment d'informations. sMRI et la DTI ont reçu plus d'attention ces dernières années, pour étudier la progression de

la maladie d'Alzheimer. Ces deux modalités sont complémentaires ; le sMRI détecte l'atrophie de la matière grise, et les modifications du volume cérébral, alors que le DTI est /un marqueur de prédiction, utile pour détecter la détérioration de la matière blanche. Dans ce contexte, nous visons à détecter, des modèles de changements, micro et macrostructuraux, dans les différents stades de l'AD, en utilisant le processus de fusion multimodale (sMRI et DTI).

Tout d'abord, la méthode proposée consiste, en une étape de prétraitement des ensembles de données T1-w sMRI et DTI. Les volumes de matière grise sont segmentés, à partir de T1w et FA, et les métriques scalaires MD sont ajustées, à partir de DTI. Dans un second temps, chaque scan 3D est transformé en 2D vues axiales. Dans notre étude, les 32 coupes, ayant l'entropie la plus élevée, sont sélectionnées. Contrairement aux études antérieures, qui n'utilisent que l'hippocampe comme ROI, et ignorent les autres structures. Nous avons proposé un nouveau CNN, pour extraire les caractéristiques pertinentes, des mesures DTI, et des images GM séparément. Après cela, ces fonctionnalités sont concaténées, et transmises à SVM, pour identifier, AD à partir de MCI, AD à partir de CN et MCI à partir de CN.

L'approche CNN-SVM démontre l'efficacité de l'utilisation de l'imagerie MRI multimodale, en atteignant une précision de classification de 99,79 %, 99,85 % et 97,00 %, pour AD/CN, AD/MCI et CN/MCI, respectivement. Les découvertes récentes prouvent que l'approche proposée offre une valeur ajoutée ou un plus, à la profession de soins de santé, en tenant compte de la classification précise de la maladie d'Alzheimer.

Cependant, ce travail présente quelques limites telles que l'intégration des autres stades de la déficience cognitive, pour vérifier l'efficacité, et la robustesse, de la méthode proposée.

Dans le futur, une nouvelle technique de fusion sera présente. Nous intégrerons d'autres techniques de neuroimagerie, ou mesures DTI, telles que les cartes LD, MO et RD. Les trois plans (axial, sagittal et coronal) peuvent également être fusionnés, au lieu de limiter l'étude au plan axial.

Une autre approche serait intéressante à investiguer, celle des réseaux de neurones récurrents (R-NN). Ce dernier a été introduit dans de nombreuses applications de recherche, qui nécessitent des processus de reconnaissance récursifs, tels que le traitement de texte, et la reconnaissance vocale. Cela permettrait de développer des systèmes plus compréhensibles, et plus précieux, qui apportent le concept de séquence, ou de dimensions temporelles, pour bien estimer la variation cérébrale des régions AD, plus sensibles au fil du temps. La mémoire

longue à court terme (LSTM), et ses méthodes de variations, par exemple, semblent une voie prometteuse pour relever ce défi.

RÉFÉRENCES

1. Alzheimer's, A., 2019 Alzheimer's disease facts and figures. *Alzheimer's & dementia*, 2019. 15(3): p. 321-387.
2. Fox, N.C., et al., Presymptomatic hippocampal atrophy in Alzheimer's disease: a longitudinal MRI study. *Brain*, 1996. 119(6): p. 2001-2007.
3. Pini, L., et al., Brain atrophy in Alzheimer's disease and aging. *Ageing research reviews*, 2016. 30: p. 25-48.
4. Van de Pol, L., et al., Hippocampal atrophy in subcortical vascular dementia. *Neurodegenerative Diseases*, 2011. 8(6): p. 465-469.
5. Bell-McGinty, S., et al., Differential cortical atrophy in subgroups of mild cognitive impairment. *Archives of neurology*, 2005. 62(9): p. 1393-1397.
6. Klein-Koerkamp, Y., et al., Amygdalar atrophy in early Alzheimer's disease. *Current Alzheimer Research*, 2014. 11(3): p. 239-252.
7. Ringman, J.M., et al., Diffusion tensor imaging in preclinical and presymptomatic carriers of familial Alzheimer's disease mutations. *Brain*, 2007. 130(7): p. 1767-1776.
8. Medina, D., et al., White matter changes in mild cognitive impairment and AD: A diffusion tensor imaging study. *Neurobiology of aging*, 2006. 27(5): p. 663-672.
9. Basser, P.J., J. Mattiello, and D. LeBihan, MR diffusion tensor spectroscopy and imaging. *Biophysical journal*, 1994. 66(1): p. 259-267.
10. Cherubini, A., et al., Aging of subcortical nuclei: microstructural, mineralization and atrophy modifications measured in vivo using MRI. *Neuroimage*, 2009. 48(1): p. 29-36.
11. Billeci, L., et al., Machine learning for the classification of alzheimer's disease and its prodromal stage using brain diffusion tensor imaging data: A systematic review. *Processes*, 2020. 8(9): p. 1071.
12. Patterson, C., The state of the art of dementia research: New frontiers. *World Alzheimer Report*, 2018. 2018.

13. Alzheimer's, A., 2017 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, 2017. 13(4): p. 325-373.
14. Qiang, W., et al., Structural variation in amyloid- β fibrils from Alzheimer's disease clinical subtypes. *Nature*, 2017. 541(7636): p. 217-221.
15. Selkoe, D.J. and J. Hardy, The amyloid hypothesis of Alzheimer's disease at 25 years. *EMBO molecular medicine*, 2016. 8(6): p. 595-608.
16. Eftekharzadeh, B., et al., Tau protein disrupts nucleocytoplasmic transport in Alzheimer's disease. *Neuron*, 2018. 99(5): p. 925-940.
17. Ramzan, F., et al., A deep learning approach for automated diagnosis and multi-class classification of Alzheimer's disease stages using resting-state fMRI and residual neural networks. *Journal of medical systems*, 2020. 44(2): p. 1-16.
18. Risacher, S.L. and A.J. Saykin, Neuroimaging and other biomarkers for Alzheimer's disease: the changing landscape of early detection. *Annual review of clinical psychology*, 2013. 9: p. 621.
19. Aderghal, K., Classification of multimodal MRI images using Deep Learning: Application to the diagnosis of Alzheimer's disease. Doctoral dissertation, Université de Bordeaux; 2021.
20. Basser, P.J., J. Mattiello, and D.J.B.j. LeBihan, MR diffusion tensor spectroscopy and imaging. *Biophysical journal*. 1994. 66(1): p. 259-267.
21. Le Bihan, D., Looking into the functional architecture of the brain with diffusion MRI. *Nature reviews neuroscience*, 2003. 4(6): p. 469-480.
22. Stejskal, E.O. and J.E. Tanner, Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. *The journal of chemical physics*, 1965. 42(1): p. 288-292.
23. Le Bihan, D. and S.J. Warach, Diffusion and perfusion magnetic resonance imaging: applications to functional MRI. LWW, 1995.
24. Mori, S., et al., Imaging cortical association tracts in the human brain using diffusion-tensor-based axonal tracking. *Magnetic Resonance in Medicine: An Official*

Journal of the International Society for Magnetic Resonance in Medicine, 2002. 47(2): p. 215-223.

25. Jeurissen, B., Improved analysis of brain connectivity using high angular resolution diffusion MRI. Doctoral dissertation, Universiteit Antwerpen (Belgium), 2012.

26. Mukherjee, P., et al., Diffusion tensor MR imaging and fiber tractography: theoretic underpinnings. American journal of neuroradiology, 2008. 29(4): p. 632-641.

27. Mori, S., et al., Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society, 1999. 45(2): p. 265-269.

28. Braak, H., E. Braak, and J. Bohl, Staging of Alzheimer-related cortical destruction. European neurology, 1993. 33(6): p. 403-408.

29. Dickerson, B.C., et al., MRI-derived entorhinal and hippocampal atrophy in incipient and very mild Alzheimer's disease. Neurobiology of aging, 2001. 22(5): p. 747-754.

30. Nir, T.M., et al., Effectiveness of regional DTI measures in distinguishing Alzheimer's disease, MCI, and normal aging. NeuroImage: clinical, 2013. 3: p. 180-195.

31. Brun, A. and L. Gustafson, Distribution of cerebral degeneration in Alzheimer's disease. Archiv für Psychiatrie und Nervenkrankheiten, 1976. 223(1): p. 15-33.

32. Thompson, P.M., et al., Mapping cortical change in Alzheimer's disease, brain development, and schizophrenia. Neuroimage, 2004. 23: p. S2-S18.

33. Kesslak, J.P., O. Nalcioglu, and C.W. Cotman, Quantification of magnetic resonance scans for hippocampal and parahippocampal atrophy in Alzheimer's disease. Neurology, 1991. 41(1): p. 51-51.

34. McKhann, G.M., et al., The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's & dementia, 2011. 7(3): p. 263-269.

35. Braak, H. and E. Braak, Neuropathological staging of Alzheimer-related changes. Acta neuropathologica, 1991. 82(4): p. 239-259.

36. Du, A.T., et al., Magnetic resonance imaging of the entorhinal cortex and hippocampus in mild cognitive impairment and Alzheimer's disease. *Journal of Neurology, Neurosurgery & Psychiatry*, 2001. 71(4): p. 441-447.
37. Xu, Y., et al., Usefulness of MRI measures of entorhinal cortex versus hippocampus in AD. *Neurology*, 2000. 54(9): p. 1760-1767.
38. Du, A.T., et al., Higher atrophy rate of entorhinal cortex than hippocampus in AD. *Neurology*, 2004. 62(3): p. 422-427.
39. De Leon, M.J., et al., MRI and CSF studies in the early diagnosis of Alzheimer's disease. *Journal of internal medicine*, 2004. 256(3): p. 205-223.
40. Jack, C.R., et al., Comparison of different MRI brain atrophy rate measures with clinical disease progression in AD. *Neurology*, 2004. 62(4): p. 591-600.
41. Aël Chetelat, G. and J.-C. Baron, Early diagnosis of Alzheimer's disease: contribution of structural neuroimaging. *Neuroimage*, 2003. 18(2): p. 525-541.
42. Ridha, B.H., et al., Volumetric MRI and cognitive measures in Alzheimer disease. *Journal of neurology*, 2008. 255(4): p. 567-574.
43. Ferrarini, L., et al., Shape differences of the brain ventricles in Alzheimer's disease. *Neuroimage*, 2006. 32(3): p. 1060-1069.
44. Blennow, K., et al., Cerebrospinal fluid and plasma biomarkers in Alzheimer disease. 2010. *Nature Reviews Neurology*, 6(3): p. 131-144.
45. Ben Ahmed, O., et al., Classification of Alzheimer's disease subjects from MRI using hippocampal visual features. *Multimedia Tools and Applications*, 2015. 74(4): p. 1249-1266.
46. Ahmed, O.B., et al., Recognition of Alzheimer's disease and Mild Cognitive Impairment with multimodal image-derived biomarkers and Multiple Kernel Learning. *Neurocomputing*, 2017. 220: p. 98-110.
47. Wu, H., et al., 3D transfer learning network for classification of Alzheimer's disease with MRI. *International Journal of Machine Learning and Cybernetics*, 2022: p. 1-15.
48. Jack, C.R., et al., MR-based hippocampal volumetry in the diagnosis of Alzheimer's disease. *Neurology*, 1992. 42(1): p. 183-183.

49. Jack, C.R., et al., Prediction of AD with MRI-based hippocampal volume in mild cognitive impairment. *Neurology*, 1999. 52(7): p. 1397-1397.
50. Aderghal, K., et al. Classification of Alzheimer disease on imaging modalities with deep CNNs using cross-modal transfer learning. in 2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS). 2018. IEEE.
51. Fischl, B. and A.M. Dale, Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences*, 2000. 97(20): p. 11050-11055.
52. Cuingnet, R., et al., Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. *neuroimage*, 2011. 56(2): p. 766-781.
53. Yuan, L., et al., Multi-source feature learning for joint analysis of incomplete multiple heterogeneous neuroimaging data. *NeuroImage*, 2012. 61(3): p. 622-632.
54. Zhu, X., H.-I. Suk, and D. Shen, A novel matrix-similarity based loss function for joint regression and classification in AD diagnosis. *NeuroImage*, 2014. 100: p. 91-105.
55. Massalimova, A. and H.A. Varol. Input Agnostic Deep Learning for Alzheimer's Disease Classification Using Multimodal MRI Images. in 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). 2021. IEEE.
56. Gunawardena, K., R.N. Rajapakse, and N.D. Kodikara. Applying convolutional neural networks for pre-detection of alzheimer's disease from structural MRI data. in 2017 24th International Conference on Mechatronics and Machine Vision in Practice (M2VIP). IEEE.
57. Farooq, A., et al. A deep CNN based multi-class classification of Alzheimer's disease using MRI. in 2017 IEEE International Conference on Imaging systems and techniques (IST). 2017. IEEE.
58. Valliani, A. and A. Soni. Deep residual nets for improved Alzheimer's diagnosis. *Proceedings of the 8th ACM international conference on bioinformatics, computational biology, and health informatics*. 2017.
59. Ashburner, J. and K.J. Friston, Voxel-based morphometry—the methods. *Neuroimage*, 2000. 11(6): p. 805-821.

60. Wen, J., et al. Comparison of DTI Features for the Classification of Alzheimer's Disease: A Reproducible Study. in OHBM 2018-Organization for Human Brain Mapping Annual Meeting. 2018.
61. Zebari, R., et al., A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction. *Journal of Applied Science and Technology Trends*, 2020. 1(2): p. 56-70.
62. Segaran, T., *Programming collective intelligence: building smart web 2.0 applications*. O'Reilly Media, Inc.2007.
63. He, K., et al. Deep residual learning for image recognition. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
64. Plis, S.M., et al., Deep learning for neuroimaging: a validation study. *Frontiers in neuroscience*, 2014. 8: p. 229.
65. Agostinho, D., et al., Combined Structural MR and Diffusion Tensor Imaging Classify the Presence of Alzheimer's Disease With the Same Performance as MR Combined With Amyloid Positron Emission Tomography: A Data Integration Approach. *Frontiers in neuroscience*, 2021. 15:p. 638175.
66. Billones, C.D., et al. DemNet: a convolutional neural network for the detection of Alzheimer's disease and mild cognitive impairment. in *2016 IEEE region 10 conference (TENCON)*. 2016. IEEE.
67. Lella, E., et al. Machine learning for the assessment of Alzheimer's disease through DTI. in *Applications of Digital Image Processing XL*. 2017. International Society for Optics and Photonics.
68. Ding, Y., et al. Classification of Alzheimer's disease based on the combination of morphometric feature and texture feature. in *2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. 2015. IEEE.
69. Lowe, D.G., Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 2004. 60(2): p. 91-110.
70. Daliri, M.R., Automated diagnosis of Alzheimer disease using the scale-invariant feature transforms in magnetic resonance images. *Journal of medical systems*, 2012. 36(2): p. 995-1000.

71. Altaf, T., et al., Multi-class Alzheimer's disease classification using image and clinical features. *Biomedical Signal Processing and Control*, 2018. 43: p. 64-74.
72. Nowak, E., F. Jurie, and B. Triggs. Sampling strategies for bag-of-features image classification. In *Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006, Proceedings, Part IV 9* (pp. 490-503). Springer Berlin Heidelberg.
73. Mathew, J., et al. Robust algorithm for early detection of Alzheimer's disease using multiple feature extractions. in *2016 IEEE Annual India Conference (INDICON)*. 2016. IEEE.
74. Hosseini-Asl, E., R. Keynton, and A. El-Baz. Alzheimer's disease diagnostics by adaptation of 3D convolutional network. in *2016 IEEE international conference on image processing (ICIP)*. 2016. IEEE.
75. Hon, M. and N.M. Khan. Towards Alzheimer's disease classification through transfer learning. in *2017 IEEE International conference on bioinformatics and biomedicine (BIBM)*. 2017. IEEE.
76. Glozman, T. and O. Liba. Hidden cues: Deep learning for Alzheimer's disease classification. CS331B project final report. 2016, no.
77. Wang, S.-H., et al., Classification of Alzheimer's disease based on eight-layer convolutional neural network with leaky rectified linear unit and max pooling. *Journal of medical systems*, 2018. 42(5): p. 1-11.
78. Nanni, L., et al., Texture descriptors and voxels for the early diagnosis of Alzheimer's disease. *Artificial intelligence in medicine*, 2019. 97: p. 19-26.
79. Jain, R., et al., Convolutional neural network based Alzheimer's disease classification from magnetic resonance brain images. *Cognitive Systems Research*, 2019. 57: p. 147-159.
80. Lee, B., et al., Using deep CNN with data permutation scheme for classification of Alzheimer's disease in structural magnetic resonance imaging (sMRI). *IEICE TRANSACTIONS on Information and Systems*, 2019. 102(7): p. 1384-1395.
81. Maqsood, M., et al., Transfer learning assisted classification and detection of Alzheimer's disease stages using 3D MRI scans. *Sensors*, 2019. 19(11): p. 2645.

82. Jiang, J., et al., Deep learning based mild cognitive impairment diagnosis using structure MR images. *Neuroscience letters*, 2020. 730: p. 134971.
83. Puente-Castro, A., et al., Automatic assessment of Alzheimer's disease diagnosis based on deep learning techniques. *Computers in biology and medicine*, 2020. 120: p. 103764.
84. Feng, W., et al., Automated MRI-based deep learning model for detection of Alzheimer's disease process. *International Journal of Neural Systems*, 2020. 30(06): p. 2050032.
85. Bansal, D., et al., Classification of Magnetic Resonance Images using Bag of Features for Detecting Dementia. *Procedia Computer Science*, 2020. 167: p. 131-137.
86. Nawaz, H., et al., A deep feature-based real-time system for Alzheimer disease stage detection. *Multimedia Tools and Applications*, 2021. 80(28): p. 35789-35807.
87. Naz, S., A. Ashraf, and A.J.M.S. Zaib, Transfer learning using freeze features for Alzheimer neurological disorder detection using ADNI dataset. *Multimedia Systems*, 2021: p. 1-10.
88. Savaş, S., Detecting the stages of Alzheimer's disease with pre-trained deep learning architectures. *Arabian Journal for Science and Engineering*, 2022. 47(2): p. 2201-2218.
89. Naggara, O., et al., Diffusion tensor imaging in early Alzheimer's disease. *Psychiatry Research: Neuroimaging*, 2006. 146(3): p. 243-249.
90. Zhang, Y., et al., Diffusion tensor imaging of cingulum fibers in mild cognitive impairment and Alzheimer disease. *Neurology*, 2007. 68(1): p. 13-19.
91. Rose, S.E., et al., Diffusion indices on magnetic resonance imaging and neuropsychological performance in amnesic mild cognitive impairment. *Journal of Neurology, Neurosurgery & Psychiatry*, 2006. 77(10): p. 1122-1128.
92. Fellgiebel, A., et al., Predicting conversion to dementia in mild cognitive impairment by volumetric and diffusivity measurements of the hippocampus. *Psychiatry Research: Neuroimaging*, 2006. 146(3): p. 283-287.
93. Müller, M.J., et al., Functional implications of hippocampal volume and diffusivity in mild cognitive impairment. *Neuroimage*, 2005. 28(4): p. 1033-1042.

94. Fellgiebel, A., et al., Color-coded diffusion-tensor-imaging of posterior cingulate fiber tracts in mild cognitive impairment. *Neurobiology of aging*, 2005. 26(8): p. 1193-1198.
95. Choo, I.L.H., et al., Posterior cingulate cortex atrophy and regional cingulum disruption in mild cognitive impairment and Alzheimer's disease. *Neurobiology of aging*, 2010. 31(5): p. 772-779.
96. Selnes, P., et al., Diffusion tensor imaging surpasses cerebrospinal fluid as predictor of cognitive decline and medial temporal lobe atrophy in subjective cognitive impairment and mild cognitive impairment. *Journal of Alzheimer's disease*, 2013. 33(3): p. 723-736.
97. Graña, M., et al., Computer aided diagnosis system for Alzheimer disease using brain diffusion tensor imaging features selected by Pearson's correlation. *Neuroscience letters*, 2011. 502(3): p. 225-229.
98. Patil, R.B., R. Piyush, and S. Ramakrishnan. Identification of brain white matter regions for diagnosis of Alzheimer using diffusion tensor imaging. In 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 6535-6538). IEEE.
99. Dyrba, M., et al., Robust automated detection of microstructural white matter degeneration in Alzheimer's disease using machine learning classification of multicenter DTI data. *PloS one*, 2013. 8(5): p. e64925.
100. Ahmed, O.B., et al. Features-based approach for Alzheimer's disease diagnosis using visual pattern of water diffusion in tensor diffusion imaging. In 2015 IEEE International Conference on Image Processing (ICIP) (pp. 2840-2844). IEEE.
101. Lee, G. and H. Fujita, *Deep learning in medical image analysis: challenges and applications*. Vol. 1213. 2020: Springer Nature.
102. Madden, D.J., et al., Diffusion tensor imaging of cerebral white matter integrity in cognitive aging. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 2012. 1822(3): p. 386-400.
103. Maggipinto, T., et al., DTI measurements for Alzheimer's classification. *Physics in Medicine & Biology*, 2017. 62(6): p. 2361.
104. Eldeeb, G.W., N. Zayed, and I.A. Yassine. Alzheimer'S Disease Classification Using Bag-Of-Words Based on Visual Pattern of Diffusion Anisotropy for DTI Imaging. in 2018

40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2018. IEEE.

105. Dalboni da Rocha, J.L., et al., Fractional Anisotropy changes in parahippocampal cingulum due to Alzheimer's Disease. *Scientific reports*, 2020. 10(1): p. 1-8.

106. Lella, E., et al., An ensemble learning approach based on diffusion tensor imaging measures for Alzheimer's disease classification. *Electronics*, 2021. 10(3): p. 249.

107. Amoroso, N., et al., Topological measurements of DWI tractography for Alzheimer's disease detection. *Computational and mathematical methods in medicine*, 2017. 2017.

108. Ebadi, A., et al., Ensemble classification of Alzheimer's disease and mild cognitive impairment based on complex graph measures from diffusion tensor images. *Frontiers in neuroscience*, 2017. 11: p. 56.

109. Marzban, E.N., et al., Alzheimer's disease diagnosis from diffusion tensor images using convolutional neural networks. *PloS one*, 2020. 15(3): p. e0230409.

110. Bansal, D., et al., Classification of magnetic resonance images using bag of features for detecting dementia. *Procedia Computer Science*, 2020. 167: p. 131-137.

111. Aderghal, K., et al., Improving Alzheimer's stage categorization with Convolutional Neural Network using transfer learning and different magnetic resonance imaging modalities. *Heliyon*, 2020. 6(12): p. e05652.

112. Kang, L., et al., Identifying early mild cognitive impairment by multi-modality MRI-based deep learning. *Frontiers in aging neuroscience*, 2020. 12: p. 206.

113. Perez-Gonzalez, J., et al., Mild cognitive impairment classification using combined structural and diffusion imaging biomarkers. *Physics in Medicine & Biology*, 2021. 66(15): p. 155010.

114. Fang, M., et al., Re-transfer learning and multi-modal learning assisted early diagnosis of Alzheimer's disease. *Multimedia Tools and Applications*, 2022: p. 1-17.

115. Lee, W., B. Park, and K. Han. SVM-based classification of diffusion tensor imaging data for diagnosing Alzheimer's Disease and mild cognitive impairment. in *International Conference on Intelligent Computing*. 2015. Springer.

116. Maggipinto, T., et al., DTI measurements for Alzheimer's classification. *Physics in Medicine & Biology*, 2017. 62(6): p. 2361.
117. Jiang, J., et al., Deep learning based mild cognitive impairment diagnosis using structure MR images. *Neuroscience letters*, 2020. 730: p. 134971.
118. Feng, W., et al., Automated MRI-based deep learning model for detection of Alzheimer's disease process. *International Journal of Neural Systems*, 2020. 30(06): p. 2050032.
119. Lella, E., et al., An Ensemble Learning Approach Based on Diffusion Tensor Imaging Measures for Alzheimer's Disease Classification. *Electronics*, 2021. 10(3): p. 249.
120. Goodfellow, I., Y. Bengio, and A. Courville, *Deep learning*. MIT press. 2016, 800 pp, ISBN: 0262035618. *Genetic programming and evolvable machines*, 19(1-2), 305-307.
121. Samuel, A.L., Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, 2000. 44(1.2): p. 206-226.
122. Robert, C., *Machine learning, a probabilistic perspective*. Taylor & Francis, 2014. 62-63.
123. Haykin, S., *Neural networks and learning machines*, 3/E. 2009: Pearson Education India.
124. Krose, B. and P.v.d. Smagt, *An introduction to neural networks*. MIT press. 2011.
125. Agarap, A.F., *Deep learning using rectified linear units (relu)*. arXiv preprint arXiv:1803.08375, 2018.
126. Rosenblatt, F., The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 1958. 65(6): p. 386.
127. Minsky, M.L. and S.A. Papert, *Perceptrons: expanded edition*. 1988, MIT press.
128. Widrow, B. and M.E. Hoff, Adaptive switching circuits. In *IRE WESCON convention record (Vol. 4, No. 1, pp. 96-104)*. 1960
129. Noriega, L., *Multilayer perceptron tutorial*. School of Computing. Staffordshire University, 2005. 4: p. 5.

130. LeCun, Y., I. Kanter, and S. Solla, Second order properties of error surfaces: Learning time and generalization. *Advances in neural information processing systems*, 1990. 3.
131. LeCun, Y., LeNet-5, convolutional neural networks. URL: <http://yann. lecun. com/exdb/lenet>, 2015. 20(5): p. 14.
132. Dumoulin, V. and F. Visin, A guide to convolution arithmetic for deep learning. *arXiv preprint arXiv:1603.07285*, 2016.
133. Kingma, D.P. and J. Ba, Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
134. Srivastava, N., et al., Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 2014. 15(1): p. 1929-1958.
135. Ioffe, S. and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. in *International conference on machine learning*. 2015. PMLR.
136. Hossain, T., et al. Brain tumor detection using convolutional neural network. in *2019 1st international conference on advances in science, engineering and robotics technology (ICASERT)*. 2019. IEEE.
137. Bishop, C.M. and N.M. Nasrabadi, *Pattern recognition and machine learning*. Vol. 4. 2006. New York: springer.
138. LeCun, Y., et al., Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998. 86(11): p. 2278-2324.
139. Krizhevsky, A., I. Sutskever, and G.E. Hinton, Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 2017. 60(6): p. 84-90.
140. Penny, W.D., et al., *Statistical parametric mapping: the analysis of functional brain images*. *Statistical parametric mapping: the analysis of functional brain images*, Elsevier.2011
141. Jenkinson, M., et al., *Fsl. Neuroimage*, 2012. 62(2): p. 782-790.
142. Ashburner, J., et al., *SPM12 manual*. Wellcome Trust Centre for Neuroimaging, London, UK, 2014. 2464: p. 4.
143. Smith, S.M., Fast robust automated brain extraction. *Human brain mapping*, 2002. 17(3): p. 143-155.

144. Mechelli, A., et al., Voxel-based morphometry of the human brain: methods and applications. *Current Medical Imaging*, 2005. 1(2): p. 105-113.
145. Ashburner, J., A fast diffeomorphic image registration algorithm. *Neuroimage*, 2007. 38(1): p. 95-113.
146. Greve, D.N. An absolute beginner's guide to surface-and voxel-based morphometric analysis. In *Proceedings of the International Society for Magnetic Resonance in Medicine (Vol. 19, p. 33)*.
147. Cousijn, J., et al., Grey matter alterations associated with cannabis use: results of a VBM study in heavy cannabis users and healthy controls. *Neuroimage*, 2012. 59(4): p. 3845-3851.
148. Maes, F., et al., Multimodality image registration by maximization of mutual information. *IEEE transactions on Medical Imaging*, 1997. 16(2): p. 187-198.
149. West, J.B., et al. Comparison and evaluation of retrospective intermodality image registration techniques. In *Medical Imaging 1996: image processing (Vol. 2710, pp. 332-347)*. SPIE.
150. Studholme, C., et al., Deformation-based mapping of volume change from serial brain MRI in the presence of local tissue contrast change. *IEEE transactions on Medical Imaging*, 2006. 25(5): p. 626-639.
151. Boser, B.E., I.M. Guyon, and V.N. Vapnik. A training algorithm for optimal margin classifiers. in *Proceedings of the fifth annual workshop on Computational learning theory*. 1992.
152. Fawcett, T., An introduction to ROC analysis. *Pattern recognition letters*, 2006. 27(8): p. 861-874.
153. Naz, S., A. Ashraf, and A. Zaib, Transfer learning using freeze features for Alzheimer neurological disorder detection using ADNI dataset. *Multimedia Systems*, 2022. 28(1): p. 85-94.
154. Ashraf, A., et al., Deep transfer learning for alzheimer neurological disorder detection. *Multimedia Tools and Applications*, 2021. 80(20): p. 30117-30142.

BIBLIOGRAPHIE DE L'AUTEUR

- Houria, L., Belkhamza, N., Cherfa, A., & Cherfa, Y. (2022). Multi-modality MRI fusion for Alzheimer's disease detection using deep learning. *Physical and Engineering Sciences in Medicine*, 1-11.
- Houria, L., Belkhamza, N., Cherfa, A., & Cherfa, Y. (2023). Multimodal magnetic resonance imaging for Alzheimer's disease diagnosis using hybrid features extraction and ensemble support vector machines. *International Journal of Imaging Systems and Technology*, 33(2), 610-621.
- Houria, L., Belkhamza, N., Cherfa, A. and Cherfa, Y. (xxxx) 'Diffusion tensor imaging for Alzheimer's disease classification using a bag of features and majority voting', *Int. J. Medical Engineering and Informatics*, Vol. X, No. Y, pp.xxx-xxx.
- Houria, L., Belkhamza, N., Cherfa, A. and Cherfa, Y. (xxxx) 'A CNN and spatial tract-based statistics- based approach for the diagnosis of Alzheimer's disease', *Int. J. Medical Engineering and Informatics*, Vol. X, No. Y, pp.xxx-xxx.