

*République Algérienne Démocratique et Populaire*  
*Ministère de l'Enseignement Supérieur et de la Recherche Scientifique*  
*Université 20Aout 1955-Skikda*



Faculté des Sciences  
Département d'Informatique

كلية العلوم  
قسم الإعلام الآلي

### Thèse

Présentée pour obtenir le diplôme de

**Doctorat 3<sup>ième</sup> cycle**

Spécialité : Informatique

---

# Approches collectives basées datamining et machine-learning pour l'interprétation d'images et reconnaissance d'objets

---

Présentée par

**Nadjat BOUCHAOUR**

Directeur de thèse

**Prof. Smaine MAZOUZI**

Devant le Jury :

Président	MOHAMMED REDJIMI	<i>Professeur, Université de Skikda</i>
Rapporteur	SMAINE MAZOUZI	<i>Professeur, Université de Skikda</i>
Examineurs	MOHAMED BENMOHAMMED	<i>Professeur, Université de Constantine 2</i>
	BACHIR BOUCHEHAM	<i>Professeur, Université de Skikda</i>

# Dédicaces

*Je dédie cette thèse à*

Mes chers parents ;

Mon mari et mes enfants ;

Mes frères et mes soeurs ;

Ma grande famille ;

Mes amies, mes collègues de travail et à toutes mes connaissances.

**Nadjet BOUCHAOUR**

# Remerciements

Je voudrais tout d'abord remercier grandement mon directeur de thèse, Monsieur Smaïne MAZOUZI, Professeur à l'Université de Skikda, qui m'a encadré tout au long de cette thèse et qui m'a fait partager ses brillantes intuitions. Qu'il soit aussi remercié pour sa gentillesse, sa disponibilité permanente et pour les nombreux encouragements qu'il m'a prodigués.

J'adresse tous mes remerciements à Monsieur Mohammed REDJIMI , Professeur à l'Université de Skikda, pour l'honneur qu'ils m'a fait en acceptant de présider le jury de soutenance.

J'exprime également ma gratitude à Monsieur Mohamed BENMOHAMMED, Professeurs à l'Université de Constantine 2, et à Monsieur Bachir BOUCHEHAM, professeur à l'université de Skikda , qui ont bien voulu être examinateurs et pour avoir accepté d'évaluer mon travail.

Sans oublier de remercier, toute personne ayant contribué de près ou de loin à ce travail, ou m'ayant facilité de le mener à bien.

**Nadjet BOUCHAOUR**

## ملخص

في التصوير الطبي ، كما هو الحال في معظم المجالات التي تتعامل مع الصور الرقمية ، تحتل التجزئة مكانًا بالغ الأهمية ، وتعتمد عليها جودة نتائج التفسير بشدة. ومع ذلك ، كانت تجزئة الصور وما زالت تمثل مشكلة عويصة، ليس فقط بسبب عدم اليقين وعدم دقة البيانات أثناء الحصول على الصورة ، ولكن أيضًا لصعوبة نمذجة بياناتها حتى تتمكن المعالجة من تحصيل النتائج المتوقعة من قبل المستخدمين . صور الرنين المغناطيسي (MRI) بالإضافة إلى الكثير من المشاكل الشائعة في معالجة الصور مثل الضوضاء والتشوهات، تشوبها أيضا عوامل مؤثرة معينة مما يعقد تجزئتها، تتمثل في مشكلة عدم انتظام الشدة (INU: Intensity Non Uniformity) ومشكلة تأثير الحجم الجزئي

من أجل المساهمة في حل مشكلة تجزئة التصوير بالرنين المغناطيسي ، الهيكلية والمرضية ، اتجهنا من خلال هذه الأطروحة لاقتراح نماذج من البيانات ونماذج من التصنيفات تسمح بتصنيف وحدات الفوكسل Voxel الخاصة بالرنين المغناطيسي بشكل صحيح بالاعتماد على الأنسجة التي تنتمي إليها. للقيام بذلك، بدأنا من فرضية أن الأنماط التي قد توجد ضمن بيانات التصوير بالرنين المغناطيسي تسمح بتصنيفها بشكل صحيح. أيضًا ، يجب أن تعزز السمات التي تمثل كلاً من شدة وحدات الفكسل ، وكذلك التفاعلات الموجودة بينها ، القوة التمييزية للمصنفات المستخدمة. وأخيرًا ، وبسبب الأحجام الكبيرة لبيانات التصوير بالرنين المغناطيسي ، افترضنا أن المصنفات الوصفية للمجموعات مثل Adaboost مناسبة تمامًا للبيانات المعالجة. أدت هذه الافتراضات وهذه الانعكاسات إلى الخروج بعملين في إطار هذه الأطروحة. العمل الأول و يتمثل في تصميم نموذج لتمثيل بيانات التصوير بالرنين المغناطيسي ، استنادًا إلى فكرة الانتروبيا المكانية spatial entropy داخل مجموعة من الفوكسل المتجاورة. يتم استخدام السمات القائمة على فكرة الانتروبيا المكانية

وشدة الفوكسل نفسها كمدخلات في خوارزمية Adaboost ، عن طريق اختبار اثنين من المصنفات الضعيفة ، وهما Naive Bayesian و SVM في العمل الثاني ، وباستخدام Adaboost مع مصنف CNN خفيف ، اعتبرنا مجموعة من السمات لتصنيف فوكسل معين. تتكون هذه المجموعة من صورة فرعية ثنائية الأبعاد لشريحة التصوير بالرنين المغناطيسي حول الفوكسل المراد تصنيفه. تم إجراء التجارب لكلا النموذجين وتمت مقارنة النتائج التي تم الحصول عليها مع تلك الخاصة بالأعمال الحديثة في الأدبيات. أظهرت النتائج التجريبية التي تم الحصول عليها الإمكانيات القوية للنماذج المقترحة ، سواء للتمثيل أو التصنيف من أجل تجزئة فعالة للتصوير بالرنين المغناطيسي ، وهذا في كلتا الحالتين التصوير بالرنين المغناطيسي الهيكلي والتصوير بالرنين المغناطيسي للورم. سمحت لنا هذه النتائج بتأكيد فرضيتنا ، وهي أن استخدام السمات التي تمثل الأنماط ، مثل الانتروبيا المكانية spatial entropy أو الصور الفرعية sub-images ، وكذلك استخدام المصنفات الوصفية للمجموعة ، مثل Adaboost يمكن اعتباره إطارًا عامًا لكفاءة و فعالية تجزئة التصوير بالرنين المغناطيسي.

**الكلمات المفتاحية :** تجزئة التصوير بالرنين المغناطيسي ، Adaboost ، spatial entropy ،

. CNN

# Résumé

En imagerie médicale, comme pour la quasi-totalité de domaines traitant de l'image numérique, la segmentation occupe une place critique, et dont dépend fortement la qualité des résultats d'interprétation. Cependant, la segmentation d'image a été et demeure encore un problème mal-posé, du fait de l'incertitude et de l'imprécision des données lors de l'acquisition des images, mais aussi de la difficulté de modéliser ses données afin que les traitements puissent produire les résultats escomptés par les utilisateurs. Les images à résonance magnétique, (IRM) et en plus du lot de problèmes communs en traitement d'images tels que le bruit et les déformations, elles sont entachées d'artéfacts spécifiques qui compliquent leur segmentation, il s'agit du problème de la non uniformité des intensités (INU : Intensity Non Uniformity) et le problème de l'effet du volume partiel.

Afin de contribuer à la résolution du problème de segmentation d'IRM, structurelles et pathologique, nous nous sommes penchés le long de cette thèse à proposer des modèles de données et des modèles de classification, permettant de bien étiqueter les voxels d'une IRM en fonction des tissus auxquels ils appartiennent. Pour ce faire, nous avons démarré d'une hypothèse, postulant que des patterns qui peuvent exister au sein des données IRM permettent de bien classifier ces dernières. Aussi, des attributs qui représentent à la fois les intensités des voxels, mais aussi les interactions qui existent entre eux, devraient renforcer le pouvoir discriminatoire des classifieurs utilisés. Et enfin, et à cause des tailles importantes des volumes IRM, nous avons supposé que les méta-classifieurs ensemblistes tels que l'Adaboost sont bien appropriés aux données traitées. Ces hypothèse et ces réflexions nous ont conduit à mener deux travaux dans le cadre de cette thèse.

Le premier travail consiste à concevoir un modèle de représentation des données IRM, basé sur la notion d'entropie spatiale au sein des voxels dans un voisinage local. Les attributs basés entropie obtenus, et l'intensité du voxel lui même sont utilisé en entrée de l'algorithme Adaboost, en testant deux classifieurs faibles, à savoir le Naïf bayésien, et le

SVM. Au second travail, et en utilisant l'Adaboost avec un classifieur CNN léger, nous avons considéré un ensemble d'attributs pour classifier un voxel donné. Cet ensemble consiste en une sous image 2D de coupe IRM autour du voxel à étiqueter. Pour les deux modèles, des expérimentations ont été menées et les résultats obtenus ont été comparés avec ceux de travaux récents de la littérature. Les résultats expérimentaux obtenus ont montré le fort potentiel des modèles proposées, soit de représentation, soit de classification pour une segmentation efficace des IRM, et ce dans les deux cas : IRM structurale et IRM tumorale.

Ces résultats nous ont conféré de confirmer notre hypothèse, à savoir que l'utilisation d'attributs représentant des patterns, tels que l'entropie spatiale ou les sous-images, ainsi que le recours à des méta-classifieurs d'ensemble, tel que Adaboost pourra être considéré comme un Framework général pour une segmentation efficace d'IRM.

**Mots-clés :** Segmentation d'IRM, Entropie Spatial, Classifieur d'ensemble Adaboost, CNN, Segmentation tumorale, Segmentation basée Pattern.

# Abstract

In medical imaging, like most of the fields dealing with digital images, segmentation takes a critical place, and on which the quality of the interpretation results strongly depends. However, image segmentation has been and still remains an ill-posed problem, due to the uncertainty and imprecision of the data during image acquisition, but also due to the difficulty of modeling its data so that processing can produce the results expected by users. Magnetic resonance images (MRI) and in addition to the set of common problems in image processing such as noise and deformations, are affected by specific artéfacts which complicate their segmentation, namely the problem of non-uniformity of intensities (INU : Intensity Non Uniformity) and the problem of the partial volume effect.

In order to contribute to the resolution of the problem of MRI segmentation, structural and pathological, we aimed in this thesis to propose models of data and models of classifications, allowing to correctly label the voxels of an MRI according to the tissues to which they belong. To do this, we started from the hypothesis, postulating that patterns that may exist within MRI data allow the latter to be properly classified. Also, attributes that represent both intensities of voxels, and also the interactions that exist between them, should reinforce the discriminatory power of the used classifiers. Finally, and because of the large sizes of the MRI volumes, we assumed that the ensemble meta-classifiers such as Adaboost are well suited to such processed data. These assumptions and these reflections led us to carry out two works during this thesis.

The first work consists in designing a model for representing MRI data, based on the notion of spatial entropy within voxels in a local neighborhood. The obtained entropy-based attributes, and the intensity of the voxel itself are used as input to the Adaboost algorithm, where we have tested two weak classifiers, namely the Naive Bayesian and the SVM. In the second work, and using Adaboost with a light CNN classifier, we considered a set of attributes to classify a given voxel. This set consists of a 2D sub-image of an

MRI slice around the voxel to be labelled. For both models, experiments were conducted and the obtained results were compared to those of recent works in the literature. The experimental results showed the strong potential of the proposed models, either of representation or of classification for an effective segmentation of MRIs, and this in both cases : structural MRI and tumoral MRI.

Obtained results allowed us to confirm our hypothesis, namely that the use of attributes representing patterns, such as spatial entropy or sub-images, as well as the use of ensemble meta-classifiers, such as Adaboost can be considered as a general framework for an efficient segmentation of MRI.

**Keywords :** MRI Segmentation, Spatial Entropy, Adaboost Classifier, CNN, Tumor Segmentation, Pattern-based Segmentation.

# Table des matières

Dédicaces	i
Remerciements	ii
Résumé en arabe	iv
Résumé	vi
Abstract	viii
Liste des Tables	4
Liste des Figures	7
Liste des abréviations	8
<b>1 Introduction générale</b>	<b>10</b>
1.1 Contexte, problématique et contributions . . . . .	11
1.2 Travaux réalisés et résultats . . . . .	14
1.3 Plan du manuscrit . . . . .	16
<b>2 IRM : Principe, Artéfacts et Traitements</b>	<b>18</b>
2.1 Introduction . . . . .	19
2.2 Acquisition d'IRM . . . . .	19
2.2.1 Modalité d'acquisition et codage en IRM . . . . .	20
2.2.2 Conditions d'acquisition . . . . .	23
2.3 Artéfacts en IRM . . . . .	24
2.4 IRM cérébrales . . . . .	26
2.4.1 Traitements des IRM cérébrales . . . . .	26

2.5	Segmentation d'IRM . . . . .	29
2.5.1	Segmentation d'IRM par classification de données . . . . .	32
2.5.2	Revue de la littérature . . . . .	34
2.6	Conclusion . . . . .	38
<b>3</b>	<b>Apprentissage automatique : Origines et aboutissement</b>	<b>39</b>
3.1	Introduction . . . . .	40
3.2	Une brève histoire de l'intelligence artificielle et de l'apprentissage automatique . . . . .	41
3.3	Apprentissage automatique : Définition et taxonomie . . . . .	43
3.3.1	Définition . . . . .	43
3.3.2	Une taxonomie des méthodes d'apprentissage automatique . . . . .	43
3.4	Quelques techniques d'apprentissage supervisé . . . . .	45
3.4.1	Les $k$ plus proches voisins . . . . .	46
3.4.2	Machine à vecteurs de support . . . . .	46
3.4.3	Classifieur Bayésien Naïf . . . . .	47
3.4.4	Réseaux de neurones artificiels . . . . .	48
3.5	Classification ensembliste . . . . .	50
3.5.1	Bagging . . . . .	51
3.5.2	Boosting . . . . .	51
3.5.3	Random Forest . . . . .	52
3.6	Apprentissage profond . . . . .	53
3.6.1	Réseaux de neurones convolutifs (CNN) . . . . .	54
3.6.2	Réseaux de neurones récurrents (RNN) . . . . .	58
3.7	Conclusion . . . . .	61
<b>4</b>	<b>Une approche ensembliste basée Entropie pour la segmentation d'IRM cérébrales</b>	<b>62</b>
4.1	Introduction . . . . .	63
4.2	Entropie de l'information . . . . .	64
4.3	Classification basée entropie pour la segmentation structurelle en IRM . . . . .	65
4.3.1	Les données IRM . . . . .	66
4.3.2	Classification basée entropie . . . . .	67

4.3.3	L'AdaBoost basé entropie pour la segmentation IRM . . . . .	68
4.4	Expérimentation et évaluation . . . . .	70
4.4.1	Sélection des paramètres . . . . .	70
4.4.2	Évaluation des performances . . . . .	70
4.4.3	Résultats expérimentaux . . . . .	71
4.4.4	Analyse des résultats et discussion . . . . .	78
4.5	Conclusion . . . . .	78
<b>5</b>	<b>Un modèle profond basé pattern pour la segmentation des tumeurs cérébrales</b>	<b>80</b>
5.1	Introduction . . . . .	81
5.2	Segmentation d'image avec prise en charge des patterns . . . . .	82
5.3	Segmentation profonde basée pattern de tumeurs cérébrales . . . . .	84
5.3.1	Les données d'IRM tumorales . . . . .	85
5.3.2	L'AdaBoost basé pattern pour la classification des données IRM .	86
5.3.3	Le CNN léger proposé pour l'étiquetage des voxels dans les IRM .	88
5.4	Expérimentation et évaluation . . . . .	91
5.4.1	Sélection des paramètres . . . . .	91
5.4.2	Évaluation des performances . . . . .	92
5.4.3	Résultats expérimentaux . . . . .	93
5.4.4	Analyse des résultats et discussion . . . . .	96
5.5	Conclusion . . . . .	96
<b>6</b>	<b>Conclusion générale</b>	<b>98</b>
6.1	Conclusion générale . . . . .	99
	<b>Bibliography</b>	<b>102</b>

# Liste des tableaux

4.1	Résultats de segmentation selon l'indice de Dice pour les différentes IRM et les différents tissus cérébraux (WM, GM et CSF). Les caractéristiques de classification par le SVM sont l'intensité du voxel et ses entropies spatiales.	73
4.2	Résultats de segmentation selon l'indice de Dice pour les différentes IRM et les différentes matières cérébrales (WM, GM et CSF). L'Adaboost avec le classifieur faible Naïve Bayes utilisant les entropies spatiales du voxel et son intensité comme caractéristiques de classification.	74
4.3	Comparaison des résultats selon l'indice de Dice avec certains travaux de la littérature.	77
5.1	Résultats de segmentation selon l'indice Dice pour les différentes IRM et les différents tissus tumoraux (noyau (Co), partie active (Ac), œdème (Oe) et tissu sain (He)).	94
5.2	Comparaison des résultats selon l'indice de Dice avec certains travaux de la littérature.	95

# Table des figures

2.1	Principe de la formation d'IRM [1] . . . . .	20
2.2	Les 3 vues IRM selon les 3 principaux axes [2]. . . . .	21
2.3	Variation des contrastes en fonction de T1 et T2 [3] . . . . .	22
2.4	Les trois pondérations en IRM[4] . . . . .	22
2.5	Bruit dans une IRM [5]. . . . .	24
2.6	Illustration de l'effet du volume partiel [6]. . . . .	25
2.7	INU dans une IRM due à un champ de biais [7]. . . . .	26
2.8	Matière cérébrale dans une IRM [8]. . . . .	27
2.9	Elimination du crane dans une IRM [9] . . . . .	28
2.10	Evolution du modèle déformable pour s'ajuster au cerveau [10] . . . . .	28
2.11	Correction de l'INU en IRM [11]. . . . .	29
2.12	Détection de contours basée ICA en IRM : a, a' IRM, b, b' Contours par le détecteur de Canny, c, c' contours par la méthode basée ICA [12] . . .	31
2.13	Région de la tumeur cérébrale selon les trois axes (sagittal, coronal, et axial) extraite par un accroissement multi-agent, présentée dans [12] . . .	32
2.14	Améliorations du FCM par l'introduction de l'information spatiale : (a) FCM standard, b) SFCM, (c) MS-FCM [13] . . . . .	33
2.15	Framework markovien pour la segmentation des tumeurs cérébrales proposé dans [14] . . . . .	34
2.16	Délimitation des tumeurs cérébrales dans des IRM. Extrait de l'article [15]	36
2.17	Enchaînement des tâches pour la segmentation des tumeurs cérébrales selon [16] . . . . .	36
2.18	Segmentation en régions anatomiques par une méthode profonde, proposée dans [17] . . . . .	37

3.1	Types d'apprentissage automatique . . . . .	45
3.2	Exemple de classification avec $k$ NN [18] . . . . .	46
3.3	Principe du classifieur SVM [19]. . . . .	47
3.4	Le neurone formel [20] . . . . .	49
3.5	Le perceptron multicouche [21] . . . . .	49
3.6	Un exemple montrant comment la classification est opérée au sein d'un random-forest [22] . . . . .	53
3.7	Architecture type de CNN [23] . . . . .	54
3.8	Exemple d'opération de convolution [24]. . . . .	55
3.9	Exemple d'opération de pooling [25]. . . . .	55
3.10	Fonctions ReLU et SoftMax [26] . . . . .	56
3.11	Réseau LeNet [27] . . . . .	57
3.12	Réseau ResNet [28] . . . . .	57
3.13	Réseau VGG16 [29]. . . . .	58
3.14	Exemple de réseau de neurones récurrent (RNN) avec plusieurs connexions récurrentes dans les couches cachées [30] . . . . .	58
3.15	Schéma général d'une unité computatioennelle dans un LSTM [31] . . . . .	59
3.16	Unité GRU comparée à l'unité LSTM et à l'unité de base d'un réseau RNN [32] . . . . .	60
3.17	Module d'attention d'un Transformer (Q : Requête, V : Valeur et K : clé) [33] . . . . .	60
4.1	Modèle de mixture gaussienne en IRM cérébrale [34] . . . . .	67
4.2	L'application du $k$ -means sur l'ensemble de voisinage . . . . .	68
4.3	Principe de la classification proposée basée entropie . . . . .	69
4.4	Un exemple d'IRM de la base de données "Brain Web". (a) IRM brute avec un niveau de bruit de 3% et un niveau d'INU de 20%, (b) Cerveau extrait par "Brain Extraction Tool" de FSL. . . . .	72
4.5	Résultats de la segmentation de l'IRM avec niveau de bruit défini sur 3% et INU défini sur 20%. (a), (b), (c) sont respectivement la matière blanche, la matière grise et le CSF pour le cas où le classifieur faible est un SVM. . . . .	72

4.6	Résultats de la segmentation de l'IRM avec niveau de bruit défini sur 3% et INU défini sur 20%. (a), (b), (c) sont respectivement la matière blanche, la matière grise et le CSF pour le cas où le classifieur faible est un Naïve Bayes. . . . .	73
4.7	Variation de la précision de la classification basée sur le classifieur faible SVM en fonction du nombre d'itérations d'Adaboost.(a) :WM,(b) :GM,(c) :CSF. 76	76
4.8	Variation de la précision de la classification basée sur le classifieur faible Naïve Bayes en fonction du nombre d'itérations d'Adaboost.(a) :WM,(b) :GM,(c) :CSF. 77	77
5.1	Les 32 block de patterns de base utilisés dans [35] . . . . .	83
5.2	Amélioration introduite par OSP en comparaison avec l'algorithme de base : PSM [36] . . . . .	84
5.3	Un échantillon d'IRM de la base de données BraTS : (a) :modalités T1, (b) :T1c, (c) :T2, (d) :Flair et (e) :segmentation de réalité terrain de la tumeur. . . . .	86
5.4	La sous-image en entré du CNN léger . . . . .	87
5.5	Principe de l'ensemble de classifieurs profonds basés pattern proposé pour la segmentation des tumeurs en IRM. . . . .	89
5.6	L'architecture du CNN léger, utilisé comme classificateur faible avec Adaboost. . . . .	90
5.7	Variation de la précision de la segmentation tumorale en fonction du nombre d'itérations de l'algorithme Adaboost.(a) :Co,(b) :Ac,(c) :Oe. . . . .	92
5.8	Résultats de segmentation présentés sur deux tranches différentes. De gauche à droite : IRM Flair, Réalité terrain, et résultats obtenus de segmentation de la tumeur. . . . .	93

# Liste des abréviations

<b>AA</b>	Apprentissage Automatique
<b>Ac</b>	Active region
<b>Adaboost</b>	Adaptive Boosting
<b>AG</b>	Algorithme Génétique
<b>BET</b>	Brain Extraction Tool
<b>BraTS</b>	Brain Tumor Segmentation
<b>CNN</b>	Convolutional Neural Network
<b>Co</b>	Core
<b>CRF</b>	Conditional Random Field
<b>CSF</b>	Cerebro Spinal Fluid
<b>2D</b>	Two-Dimensional
<b>3D</b>	Tridimensionnel
<b>Fast</b>	Features from Accelerated Segment Test
<b>FC</b>	Fully connected
<b>FCM</b>	Fuzzy C-Mean
<b>FLAIR</b>	Fluid-Attenuated Inversion Recovery
<b>FMRIB</b>	Functional Magnetic Resonance Imaging of the Brain
<b>FP</b>	False Positif
<b>FSL</b>	FMRIB Software Library
<b>GM</b>	Gray Matter
<b>GMM</b>	Gaussian Mixture Model
<b>GPU</b>	Graphical Processing Unit
<b>GRU</b>	Gated Recurrent Unit
<b>He</b>	Healthy brain tissue
<b>IA</b>	Intelligence Artificielle
<b>IBM</b>	International Business Machines Corporation
<b>IBSR</b>	Internet Brain Segmentation Repository
<b>ICA</b>	Independent Component Analysis
<b>INU</b>	Intensity Non Uniformity
<b>IRM</b>	Imagerie par Résonance Magnétique
<b><math>k</math>NN</b>	$k$ Nearest Neighbors
<b>LCR</b>	Liquide Céphalo-Rachidien
<b>LSTM</b>	Long Short Term Memory
<b>MB</b>	Matière Blanc
<b>MCCNN</b>	Multi-Cascaded Convolutional Neural Networks
<b>MG</b>	Matière Grise
<b>MS-FCM</b>	Modified Spatial FCM
<b>NL-FCM</b>	Non local-FCM
<b>Oe</b>	Œdema
<b>OSP</b>	Object Segmentation using block based Patterns
<b>PD</b>	Proton density weighted
<b>PMC</b>	Perceptron MultiCouche
<b>PSM</b>	Pattern Split-and-Merge
<b>ReLU</b>	Rectified Linear Unit
<b>ResNet</b>	Residual Network
<b>RNN</b>	Recurrent neural networks
<b>ROI</b>	Region Of Interest

<b>SFCM</b>	Spatial FCM
<b>SMP5</b>	Statistical Parametric Mapping (2005)
<b>SVM</b>	Support Vector Machine
<b>TDM</b>	Tomodensitométrie
<b>TN</b>	True Negatif
<b>TP</b>	True Positif
<b>TSBTS</b>	Task Structure for Brain Tumor Segmentation
<b>WM</b>	White Matter

# Chapitre 1

## Introduction générale

## 1.1 Contexte, problématique et contributions

En 2022, cent vingt sept ans étaient écoulés depuis la découverte des rayons X par Wilhelm Röntgen [37], évènement considéré comme le premier début de l'imagerie médicale. Wilhelm Röntgen a mérité le premier prix Nobel de physique en 1901 pour cette découverte monumentale, et la médecine a vu depuis, des progrès considérables en termes de diagnostic de pathologies et de recherche scientifique médicale. Durant le 20<sup>ième</sup> siècles, différentes procédés en physique, ont permis de mettre en place d'autres nouvelles techniques d'imagerie médicales, basées sur différents phénomènes physiques. Sans qu'on soit exhaustifs, il s'agit de la radioactivité, notamment artificielle, et l'électromagnétisme. En effet, ce dernier a permis la mise en oeuvre de l'imagerie par résonance magnétique (IRM), basée sur la magnétisation des molécules d'eau dans les tissus biologiques et où le signal résultant, en diverses intensités, de la précession des molécules d'eau permet de les différencier sur les clichés de coupes obtenus. Par ailleurs, et depuis l'avènement de l'informatique, notamment les disciplines de traitement d'images et de reconnaissance d'objets, les différentes techniques en imagerie médicales se sont dotées d'outils logiciels d'aides au diagnostic, permettant aux praticiens médecins et radiologues une assistance à la détermination des différentes régions d'intérêt dans les images obtenues, et ainsi de poser les diagnostics nécessaires en moins de temps et en plus de fiabilité. Actuellement, un volume tridimensionnel d'une IRM est formé de centaines de coupes, et les faire passer une après l'autre à la recherche d'une anomalie donnée, en prendra des heures du temps du praticien.

Aussi, et dû à plusieurs facteurs inhérents aux images médicales, tels que le bruit, le faible contraste et la non uniformité des intensités, le praticien peut passer à coté d'une anomalie qui devrait être détectée, et les conséquences qui en découlent de ce

type de faux négatif de diagnostic est sans doute très préjudiciable pour le patient et pour l'évolution de sa maladie. De ce fait, il a été remarqué l'importance de l'aide au diagnostic basée imagerie médicale aux mains des médecins et des radiologues, et des quantités indénombrables de travaux de recherche étaient menés depuis plus de quatre décennies, qui ont permis aujourd'hui d'aboutir à des techniques fiables pour le traitement et aussi la détection de différents objets d'intérêt diagnostic dans les images médicales [38].

Si nous considérons l'IRM comme technique d'imagerie médicale par excellence, par sa nature non-invasive, et ses bonnes qualités en contraste et en résolution, des milliers de chercheurs se sont penchés pour proposer des méthodes et des outils pour leur traitement automatique, en améliorant les images produites ou en procédant à leur segmentation dans le but de détecter les objets qui ont un intérêt diagnostic pour les médecins. Cependant, et malgré le progrès considérable en traitement et en diagnostic automatique, la fiabilité des techniques élaborées restent en-deçà des attentes des praticiens. Néanmoins, le progrès parallèle en traitement de données, notamment la constante évolution des techniques d'intelligence artificielle d'une part, et les architectures matérielles qui lui sont dédiées d'autre part, ne cessent de pousser les horizons de recherche afin de mieux fiabiliser les outils de traitement et d'aide au diagnostic basé sur l'imagerie médicale [39].

Dans la présente thèse, nous nous sommes intéressés à l'interprétation d'images médicales, et nous nous sommes penchés plus particulièrement sur la segmentation de données en imagerie à résonance magnétique (IRM). Néanmoins, les méthodes que nous présentons dans ce manuscrit demeurent applicables à tout type d'images où les mêmes problématiques de traitement et d'interprétation sont posées. En effet, pour l'interprétation d'IRM ainsi que d'autres images médicales, la segmentation des organes, des tissus, ou des lésions pathologiques est une étape cruciale, et d'une difficulté avérée [40]. Par la nature à la fois incertaine et imprécise des données en imagerie médicale, due à divers facteurs relevant principalement du dispositif et des conditions d'acquisition d'images, il en résulte qu'une méthode simple de partitionnement de données ne sera pas efficace et ne produira pas les résultats escomptés par les utilisateurs. Une IRM comme tout type d'image réelle, est entachée de plusieurs artéfacts dont le bruit, et le faible contraste des tissus. Elle est également frappée par deux artéfacts spécifiques, à savoir la non uniformité des intensités, et le problème du volume partiel, dus respectivement aux champs magnétiques lors de l'acqui-

sition, et à la faible résolution lors de l'opération d'échantillonnage, produisant les voxels du volume de l'IRM. Malgré le nombre élevé de travaux ayant abordé la segmentation d'IRM, les résultats obtenus restent insuffisants, par rapport aux attentes des médecins, et par rapport à la criticité de la décision médicale qui en découle. Divers paradigmes ont été adoptés, allant du traitement du signal en arrivant aux méthodes d'intelligence artificielle, en passant par les méthodes statistiques, de data mining, d'optimisation ... etc.

Il a été aussi constaté que certains traitements conventionnels appliqués aux IRM, tels que le filtrage de bruit, n'est pas efficace avec les données IRM, voire il dégrade la qualité de l'image vis-à-vis d'un aspect articulé. En effet, un prétraitement de bruit par n'importe quelle méthode n'améliore guère la non-uniformité des intensités dans une IRM, étant donné que les variations des intensités sont très lentes. Il a été signalé également que le traitement de bruit par différents filtres basés lissage, aggrave le problème de l'effet du volume partiel [41]. Ce problème se produit principalement au niveau des frontières entre les régions caractérisées par des intensités différentes, quand l'intensité d'un voxel donné sur une frontière est remplacée par une valeur obtenue par une convolution selon le filtre utilisé. La nouvelle intensité d'un voxel sur une frontière comprend des intensités issues de régions différentes. Ceci résulte en un éloignement du voxel en question de la région pour laquelle il devrait appartenir. Une application sélective des voxels à inclure dans le calcul de la nouvelle intensité d'un voxel n'est pas facile à réaliser, car a priori nous ne disposons pas des distributions statistiques des intensités des différentes régions, et même si nous les avons, les intervalles de confiances, au sens statistiques, se chevauchent considérablement et empêchent une bonne sélection.

Nous nous sommes attaqués dans nos travaux de thèse à la segmentation de données IRM sous contrainte de présence de bruit, de non-uniformité d'intensité, et de volume partiel, sans qu'aucun traitement préalable de ces artéfacts ne soit opéré. Nous avons posé l'hypothèse au début de la thèse qu'un bon choix des modèles de classification, et qu'une bonne représentation des données brutes des IRM devrait nous permettre de produire des résultats de segmentation comparables, voire meilleures que ceux de la littérature, sans qu'un prétraitement de bruit, ou de non-uniformité ne soit effectué. Pour ce faire, nous avons d'une part, fait recours aux classifieurs d'ensemble utilisant des classifieurs classiques et des classifieurs profonds pour proposer de nouveaux modèles dédiés à la segmentation structurelle d'IRM (IRM sans tumeurs) et à la segmentation de tumeurs

cérébrales dans les IRM pathologiques. D'autre part, nous avons cherché des modèles de représentation des données IRM qui vont au delà de l'information d'intensité du voxel et qui intègrent les interactions et les configurations au sein de sous-ensembles de voxels voisins. Nous avons supposé que ces derniers aspects, longtemps négligés dans les travaux de recherche, nous permettrons de booster la fiabilité des méthodes proposées, et ce afin de produire une segmentation exacte des structures des IMR ou des tumeurs.

## 1.2 Travaux réalisés et résultats

Différents modèles de représentation et différents modèles de classification ont été proposés dans le cadre de cette thèse. Comme nous l'avons mentionné plus haut, nous avons émis l'hypothèse que les interactions et les configurations de voxels des différents tissus dans un voisinage donné, va permettre de mieux classifier les voxels d'une IRM. Pour ce là, nous avons procédé par une approche basée pattern pour la représentation de l'information IRM. Dans l'ensemble des travaux, que nous citons ci-après, différents voisinages d'un voxel, de tailles variables ont été considérés comme attributs pour les différents modèles de classification proposés. En plus, et afin de bien capturer les interactions de voxels au sein d'un voisinage local, nous avons opté pour une représentation basée énergie, et qui consiste en l'entropie spatiale de voxels voisins. Ainsi, des ensembles d'attributs servant comme entrées aux classifieurs proposés seront composés, pour un voxel à classifier, de l'intensité du voxel lui-même, des intensités des voxels voisins, et de la grandeur de l'entropie spatiale des voxels du voisinage. En premier lieu, nous avons considéré un voisinage étroit d'un voxel ( $3 \times 3$  et  $5 \times 5$ ) et nous avons fait recours à l'entropie spatiale dans ce voisinage. L'entropie par sa nature à estimer le degré de désordre dans un système devrait nous permettre d'exprimer les interactions entre les voxels voisins en supposant l'existence d'un pattern local au niveau du voxel à classifier. En seconde lieu, et par leur pouvoir d'apprendre les patterns locaux dans les images, les CNN sont utilisés pour considérer le pattern local au voisinage du voxel. Pour ce cas, le voisinage est étendu à  $21 \times 21$  autour du voxel en question.

En termes de modèles de classification, différents modèles ont été conçus, implémentés et testés. Il s'agit en premier lieu d'un modèle basé sur un algorithme d'ensemble : AdaBoost, avec lequel nous avons testé deux classifieurs faibles : le SVM et le Bayésien

naïf. Pour ces classifieurs, différents ensembles d'attributs incluant le voxel, ses voisins, et l'entropie spatiale ont été testés et évalués pour la segmentation d'IRM structurale. Pour la segmentation de tumeurs cérébrales dans une IRM, un modèle ensembliste basé AdaBoost et utilisant une architecture de CNN comme classifieur faible a été utilisée. Pour ce modèle, il a été considéré également le voxel à classifier et son voisinage, mais sans l'entropie spatiale, étant donné qu'il s'agit d'un attribut de synthèse, peu adéquat avec le paradigme des classifieurs profonds. En restant fidèles au principe de l'apprentissage profond, les intensités brutes des voxels appartenant à la sous-image dont le voxel en question est au centre, sont utilisés pour prédire la classe de ce dernier. Nous avons remarqué qu'un voisinage de  $21 \times 21$  voxels permet de capturer les interactions et de prendre en compte le pattern géométrique qui y existe.

Pour les différents modèles proposés, des évaluations de performance en utilisant la base d'images adéquate et l'indice de Dice ont été effectuées. Aussi, des comparaisons avec des résultats de travaux récents de la littérature ont été effectuées, et ce pour la segmentation structurales des IRM saines et pour la segmentation des tumeurs cérébrales en IRM pathologique. Pour le modèle utilisé en segmentation structurale, il a été évalué sur la base de données Brain Web [42]. Pour le modèle utilisé pour la segmentation de tumeurs cérébrales dans les IRM, il a été testé en utilisant la base d'images BraTS'2017 [43]. Il est à noter que dans les 2 cas, la base d'image contient une réalité terrain de la segmentation, ce qui nous a permis de procéder avec des méthodes d'apprentissage automatique supervisées (Bayésien naïf, SVM et CNN), et aussi d'évaluer quantitativement et rigoureusement les méthodes de segmentation proposées.

Il en ressort de l'ensemble des évaluations et des différentes comparaisons que les modèles de représentation des données IRM, notamment basés pattern, servant à capturer les interactions entre voxels, ont permis d'obtenir des résultats meilleurs que ceux de la littérature, même en l'absence de prétraitement de bruit et de non uniformité d'intensités. En termes de complexité de calcul, les temps d'exécution sont raisonnables et ce à cause des architectures simples des classifieurs faibles y compris celle du classifieur profond, étant donné qu'il opère sur une sous-image de faible taille. Pour ce dernier point, et contrairement à l'ensemble des travaux de la littérature que nous avons passé en revue, l'utilisation de sous-images de tailles justes suffisante autour du voxel à classifier, permet de prendre en compte le pattern local et d'effectuer l'apprentissage et la prédiction en

temps très court.

### 1.3 Plan du manuscrit

Le premier chapitre du manuscrit est intitulé "**IRM : Principe, Artéfacts et Traitements**". Nous présentons dans ce chapitre le domaine auquel nous nous sommes intéressés, à savoir l'imagerie médicale et plus particulièrement l'imagerie par résonance magnétique (IRM). Après une introduction de la technique IRM, nous présentons les artéfacts qui affectent spécifiquement les IRM. Nous abordons ensuite la problématique de segmentation en général, et puis la segmentation d'IRM, où nous présentons une revue de la littérature de ce domaine en survolant les travaux récents et bien référencés, et en nous focalisons sur les travaux ayant utilisé l'intelligence artificielle et l'apprentissage automatique pour la segmentation structurelle d'IRM et la segmentation de tumeurs cérébrales.

Le second chapitre est intitulé "**Apprentissage automatique : Origines et aboutissement**" présente les paradigmes auxquels nous avons fait recours pour proposer les modèles de représentation et de classification pour la segmentation en IRM. En première partie de ce chapitre, nous présentons le paradigme d'apprentissage automatique, en abordant ses différents types, à savoir l'apprentissage classique, l'apprentissage profond et l'apprentissage par renforcement. Nous présentons également la classification d'ensemble via les méta algorithmes de classification dont le "boosting de classifieurs", que nous utilisons dans le cadre de cette thèse.

Notre première contribution est introduite au chapitre 3, intitulé "**Une approche ensembliste basée Entropie pour la segmentation d'IRM cérébrales**". Nous commençons ce chapitre par la présentation de l'entropie relevant de la théorie d'information, dont nous adoptons dans nos travaux pour une représentation basée pattern de l'information IRM. Nous considérons une représentation basée sur des patterns où les voxels voisins sont agrégés pour former des patterns, puis utilisons ces derniers comme caractéristiques pour la classification. Pour construire le pattern qui sera utilisé comme caractéristique de classification, nous utilisons un codage basé sur l'énergie, en supposant que l'énergie définie dans le voisinage d'un voxel représente bien la séparation des voxels. Par conséquent, le modèle est basé sur la définition au sein d'un voisinage d'un voxel d'une fonction d'éner-

gie exprimée par l'entropie spatiale des intensités des voxels. Une telle fonction d'énergie permet de considérer, en plus de l'intensité de l'image, les relations spatiales géométriques locales qui existent dans les données IRM. Nous montrons que l'agrégation des données selon le pattern considéré, notamment avec le codage basé énergie, ainsi que la classification de ces patterns avec l'ensemble de classificateurs le plus utilisé : l'Adaboost, permet d'améliorer significativement les résultats de classification des données IRM, que nous présentons et nous discutons à la fin du chapitre.

Le dernier chapitre du manuscrit présente notre seconde contribution, et est intitulé "**Un modèle profond basé pattern pour la segmentation des tumeurs cérébrales**". Dans ce chapitre, nous nous intéressons à la segmentation des tumeurs cérébrales par classification des données IRM pathologiques à l'aide d'une méthode d'apprentissage d'ensemble, où nous avons opté pour des CNN légers comme classificateurs faibles. Chaque instance du modèle CNN proposé ne fonctionne que sur une sous-image partielle, sélectionnée parmi l'une des tranches qui composent l'IRM. Ainsi, le CNN léger proposé représente le classificateur faible utilisé au sein du classificateur d'apprentissage d'ensemble. De tels CNN légers se traduisent par un faible nombre de paramètres, ce qui permet un entraînement rapide et une prédiction rapide. Contrairement à la plupart des travaux publiés, et au lieu d'utiliser les données d'image brutes, prises individuellement au niveau des différents voxels, nous avons suivi une approche basée pattern où les voxels voisins sont agrégés pour former des patterns, puis utilisons ces derniers pour la classification. Le modèle de classification est basé sur l'hypothèse qu'un voxel central au centre d'une sous-image, sera bien classé en prenant en compte les voxels environnants dans un voisinage large. Selon ce schéma, et en plus des intensités des voxels, les formes géométriques locales dans les sous-images sont considérées pour l'étiquetage des voxels. Des expériences ont été réalisées sur la base de données BraTS (Brain Tumor Segmentation), et les résultats obtenus montrent que le modèle proposé permet à la fois une détection et une segmentation rapides et précises des tumeurs cérébrales dans les IRM.

Nous terminons le manuscrit par une "**Conclusion générale**" dans laquelle nous revenons sur l'essentiel de nos contributions et les résultats qui ont été obtenus. Nous y traçons également quelques potentielles perspectives à ce travail de thèse.

## Chapitre 2

# IRM : Principe, Artéfacts et Traitements

## 2.1 Introduction

L'imagerie par résonance magnétique (IRM) est une technologie d'imagerie non invasive, utilisée pour étudier l'anatomie et les fonctions du corps, à la fois en situation de bonne santé ou en situation de maladie, et ce sans utilisation de rayonnement ionisant dommageable. Elle est souvent utilisée pour la détection des maladies, le diagnostic et le suivi du traitement. Elle est basée sur une technologie sophistiquée qui excite et détecte les changements dans les protons composant les molécules d'eau qui forment les tissus vivants. Les scanners d'IRM sont particulièrement adaptés à l'imagerie des parties non osseuses ou des tissus mous du corps. Le cerveau, la moelle épinière et les nerfs, ainsi que les muscles, les ligaments et les tendons sont vus beaucoup plus clairement avec l'IRM qu'avec les radiographies régulières ou la tomodensitométrie (TDM). Dans le cerveau, l'IRM peut différencier la matière blanche de la matière grise et peut également être utilisée pour diagnostiquer les anévrismes et les tumeurs. Nous nous intéressons dans notre thèse à la segmentation structurelle des IRM du cerveau pour en extraire les différents tissus cérébraux. Mais aussi, nous nous intéressons à la segmentation des tumeurs cérébrales en leurs différents tissus tumoraux. Ce chapitre présentera donc notre contexte de recherche, à savoir les IRM cérébrales, notamment leur segmentation, soit de structures cérébrales soit de tumeurs.

## 2.2 Acquisition d'IRM

La formation d'images (données) IRM résulte de l'enregistrement de la résonance des atomes d'hydrogène, contenus dans l'eau abondante dans le corps humain, et qui entre dans la composition des tissus des organes cibles [44, 45]. De ce fait, plus la concentra-

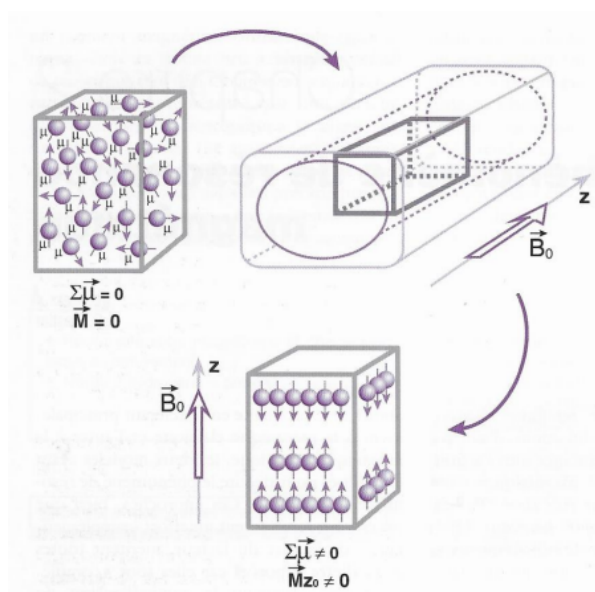


FIGURE 2.1 – Principe de la formation d'IRM [1]

tion de l'eau est élevée dans les tissus, plus la qualité de l'IRM obtenue est meilleure. Le procédé technologique permettant d'obtenir une IRM consiste à faire appliquer sur l'organisme un champ magnétique puissant, ce qui induit l'unification de la polarisation des molécules d'eau, qui se dirigent toutes dans la direction du champ (Figure 2.1). Le retour à l'état initial des molécules d'eau engendre un signal qui sera reçu et traité par un dispositif informatique pour former le volume de données IRM.

La technique IRM permet donc la visualisation des tissus mous de l'organisme contenant de l'eau. Ces tissus émettent des signaux d'intensités différentes ce qui permet d'avoir des écarts de contraste après leur représentation. Le principe de l'IRM diffère de celui du scanner, car cette dernière technique est basée sur l'absorption des tissus des rayons 'X' qui traversent l'organisme. Étant donné que la masse de données obtenue par une IRM est un volume tridimensionnel de forme rectangulaire 3D, il est possible de la visualiser selon les 3 principaux axes qui forment le référentiel 3D. On obtient trois vues nommées respectivement coronale, axiale et sagittale (voir Figure 2.2).

### 2.2.1 Modalité d'acquisition et codage en IRM

Différents contrastes peuvent être obtenus, en fonction des paramètres d'acquisition, appelés communément modalités IRM. Il en résulte ainsi trois types de pondération, à savoir T1, T2, et DP pour densité de protons. On énumère deux paramètres d'acquisition

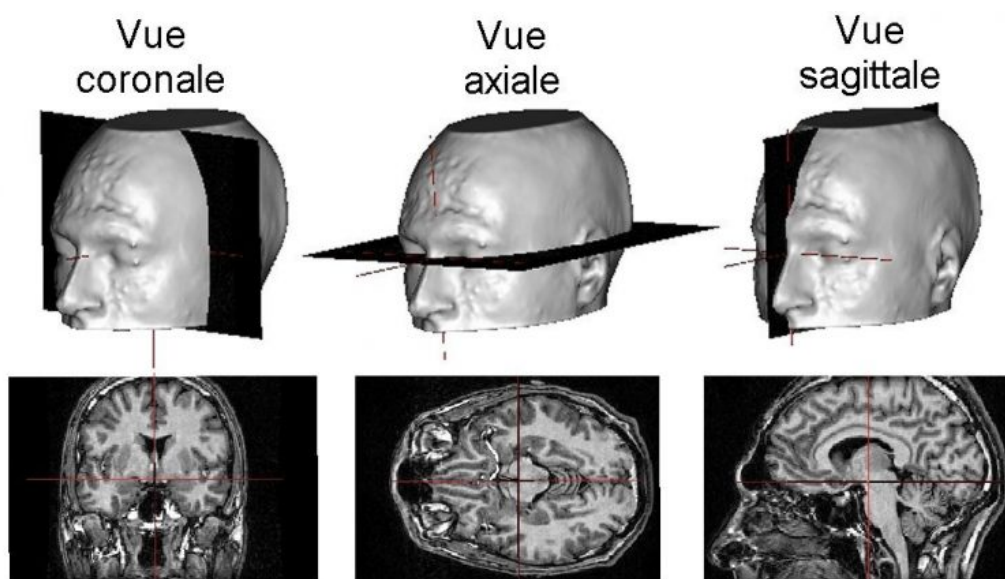


FIGURE 2.2 – Les 3 vues IRM selon les 3 principaux axes [2].

qui définissent les types de pondération [44] :

1. TR : Temps de répétition entre deux impulsions électromagnétiques successives ;
2. TE : Temps d'écho, séparant l'instant de l'impulsion et l'instant d'acquisition du signal.

En fonction de ces deux paramètres, il en résulte les trois pondérations possibles en IRM :

- T1 : correspond à un TR et un TE, les deux courts ;
- T2 : correspond à un TR et un TE, les deux longs ;
- DP : correspond à un TR long et un TE court.

Chacune des pondérations précédentes correspond à des contrastes différents entre les divers tissus qui peuvent exister dans une IRM. Pour l'IRM cérébrale, il s'agit des trois tissus respectivement de la matière grise (MG), de la matière blanche (MB), et du liquide céphalo-rachidien (LCR). La figure 2.3 montre les dégradés des niveaux de gris en fonction des temps TE et TR, ce qui donne en résultat les trois pondérations T1, T2 et DP [3].

La figure 2.4 montre les trois pondérations possibles d'une même prise d'IRM. On peut constater que pour une pondération donnée, certains détails sont mis en valeur, comparés aux autres pondérations.

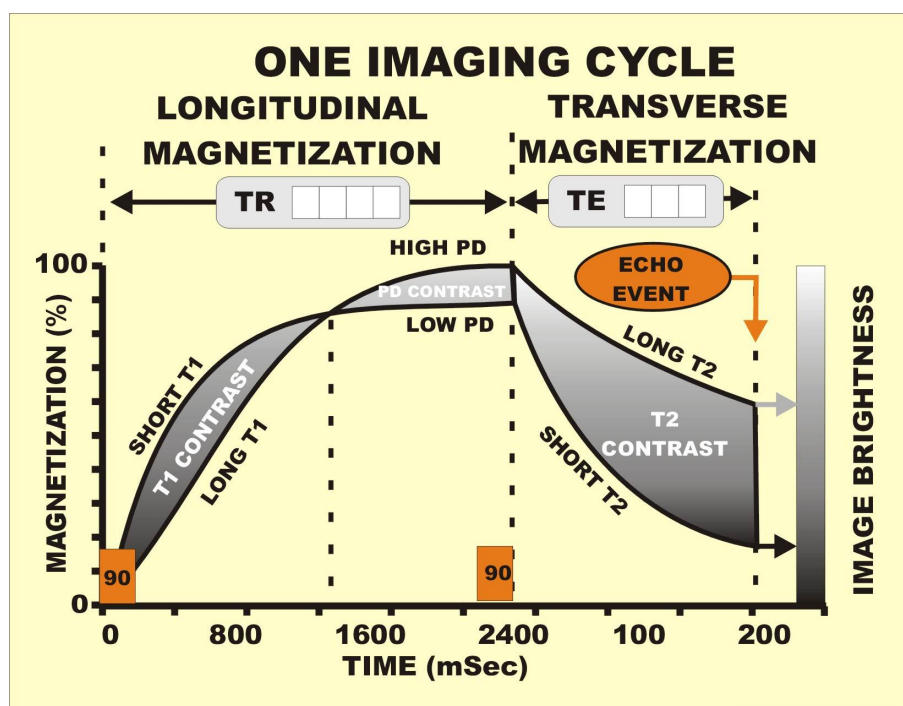


FIGURE 2.3 – Variation des contrastes en fonction de T1 et T2 [3]

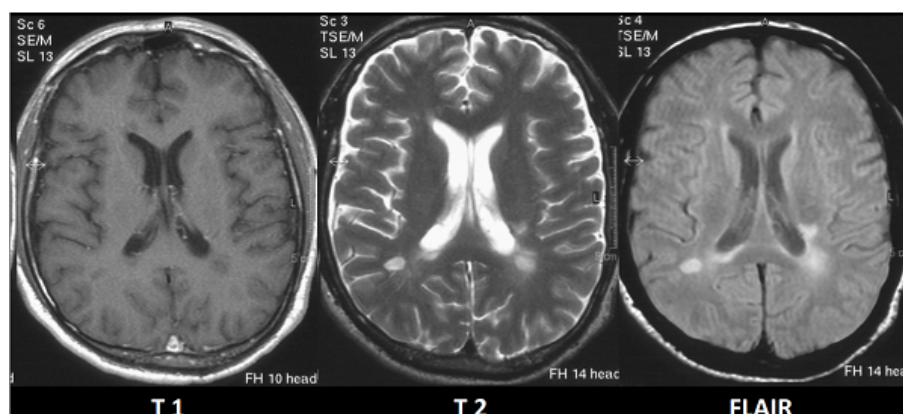


FIGURE 2.4 – Les trois pondérations en IRM[4]

## 2.2.2 Conditions d'acquisition

Pour obtenir une image IRM, un patient est placé à l'intérieur d'un gros aimant, et doit rester très immobile pendant le processus d'imagerie afin de ne pas flouter l'image. Des agents de contraste (souvent contenant l'élément Gadolinium) peuvent être administrés à un patient par voie intraveineuse avant ou pendant l'IRM pour augmenter la vitesse à laquelle les protons se réalignent avec le champ magnétique. Plus les protons se réalignent vite, plus l'image est brillante. Bien que l'IRM n'émette pas le rayonnement ionisant que l'on trouve dans les radiographies et les tomodensitogrammes, elle utilise un fort champ magnétique. Le champ magnétique s'étend au-delà de la machine et exerce des forces très puissantes sur les objets en fer, certains aciers, et d'autres objets magnétisables ; il est assez fort pour jeter un fauteuil roulant à travers la pièce. Les patients doivent informer leur médecin de toute forme de traitement médical ou d'implant avant de procéder à une IRM. Lors d'une IRM, les éléments suivants doivent être pris en considération :

- Les personnes portant des implants, en particulier ceux qui contiennent du fer
  - des stimulateurs cardiaques, des stimulateurs nerveux vagues, des défibrillateurs cardioverters implantables, des enregistreurs de boucles, des pompes à insuline, des implants cochléaires, des stimulateurs cérébraux profonds et des capsules provenant d'une endoscopie à capsule ne devraient pas entrer dans un appareil d'IRM.
- Bruit : bruit fort communément appelé cliquetis et bip, ainsi que l'intensité sonore jusqu'à 120 décibels dans certains scanners IRM, peuvent nécessiter une protection auditive spéciale.
- Les agents de contraste : les patients atteints d'insuffisance rénale grave qui ont besoin de dialyse peuvent présenter un risque de maladie rare mais grave appelée fibrose systémique néphrogène qui peut être liée à l'utilisation de certains agents contenant du gadolinium, comme le gadodiamide et d'autres. Bien qu'un lien de causalité n'ait pas été établi, les lignes directrices en vigueur aux États-Unis, par exemple, recommandent que les patients en dialyse ne reçoivent que des agents de gadolinium lorsqu'ils sont essentiels, et que la dialyse devrait être effectuée le plus tôt possible après l'examen pour retirer rapidement l'agent du corps.
- Grossesse : Bien qu'aucun effet n'ait été démontré sur le fœtus, il est recommandé d'éviter les examens IRM à titre préventif, surtout au cours du premier trimestre de la grossesse, lorsque les organes du fœtus se forment, et les produits de contraste,

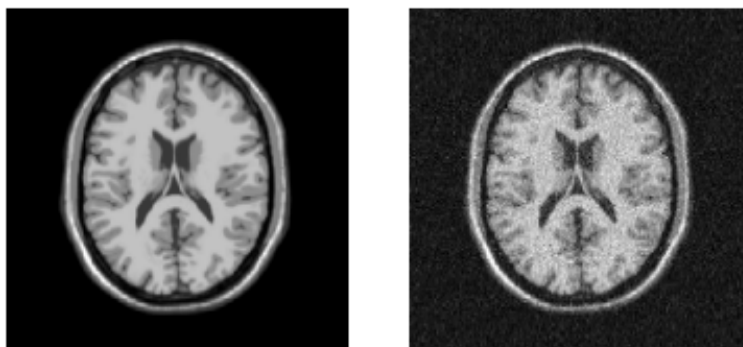


FIGURE 2.5 – Bruit dans une IRM [5].

s'ils sont utilisés, pourraient pénétrer dans le sang du fœtus.

## 2.3 Artéfacts en IRM

Un artéfact est une caractéristique apparaissant dans une image qui n'est pas présente dans l'objet original. Selon leur origine, les artéfacts sont généralement classés comme liés au patient, au traitement du signal ou au matériel (machine). En réalité, une dizaine d'artéfact peuvent affecter une IRM [46], cependant nous nous intéressons qu'au artéfacts informationnels qui peuvent être corrigés en post-acquisition par des procédés computationnels. Il s'agit du bruit, de l'effet du volume partiel, et de la non-uniformité des intensités.

### Bruit

Le bruit dans une IRM, ou image en général, résulte des altérations aléatoires de l'information enregistrée dans l'image. La valeur de chaque voxel/pixel de l'image est altérée par l'ajout d'une valeur aléatoire, plus ou moins élevée, qui pourra être négative. Ceci résulte en une IRM à qualité réduite [47], où l'interprétation manuelle ou automatique peut être sujette à des erreurs, et ce qui conduit à son tour à un faux diagnostic dans le cas d'images médicales.

La figure 2.5 montre à gauche une coupe d'IRM exempte du bruit, et à droite la même IRM affectée par un bruit gaussien additif important.

L'atténuation du bruit en IRM est opérée par diverses méthodes, empruntées des méthodes filtrage des images 2D. En supposant que le bruit est additif et gaussien, un pré-traitement de bruitage basé sur un filtre de Gauss est effectué. Ceci consiste en un filtre

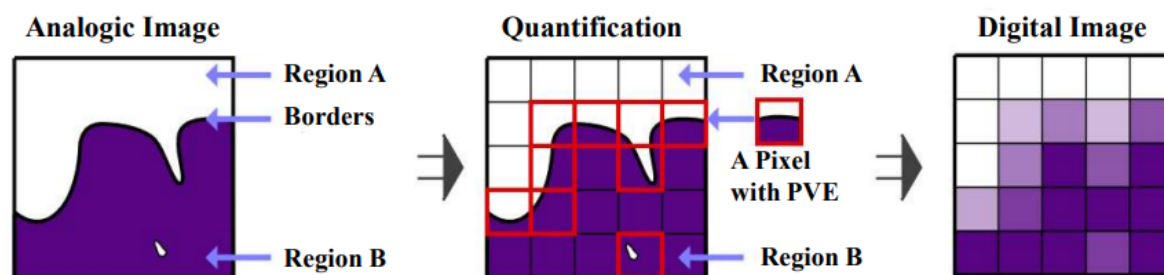


FIGURE 2.6 – Illustration de l'effet du volume partiel [6].

passé-bas, permettant de lisser le bruit notamment présent sous forme d'impulsions dans le signal image. Certains auteurs ont combiné plusieurs filtres tels que le filtre gaussien et le filtre médian pour le lissage de bruit dans les IRM [48]. Plusieurs travaux ont intégré le filtrage de bruit et la segmentation d'IRM en adoptant une méthode conjointe [49].

### Effet du volume partiel

L'artéfact du volume partiel, dit effet du volume partiel, provient de la taille du voxel sur lequel le signal est moyenné. Les objets plus petits que les dimensions des voxels perdent leur identité, et la perte de détails et de résolution spatiale s'en résulte. La réduction de ces artéfacts est opérée en utilisant une taille de pixel plus petite et/ou une épaisseur de coupe plus petite [50].

La figure 2.6 illustre comment le mélange de l'information s'opère sur les pixels situés sur les bordures entre les régions et ce à cause de la taille spatiale élevée du pixel.

Le problème de l'effet du volume partiel est résolu généralement par l'estimation d'apport de chaque tissu à un point donné de l'image [51]. De telles estimations des apports des différents tissus par point de l'IRM, permettent de considérer ceci dans les méthodes de segmentation basées classification, notamment en classification floue [52].

### Non-uniformité des intensités

La non-uniformité des intensités (INU), dite aussi inhomogénéité des intensités, provient du fait que des ensembles de voxels/pixels appartenant à un même tissu (région homogène) auront des valeurs d'intensité différentes. Concrètement, il s'agit d'un biais d'intensité qui affecte l'image d'une manière très lente, qui ne peut se constater qu'à l'échelle globale de l'image. Cette déformation particulière a impliqué des méthodes spécifiques pour la correction [53]. Les raisons de l'INU sont multiples en IRM cependant, elle

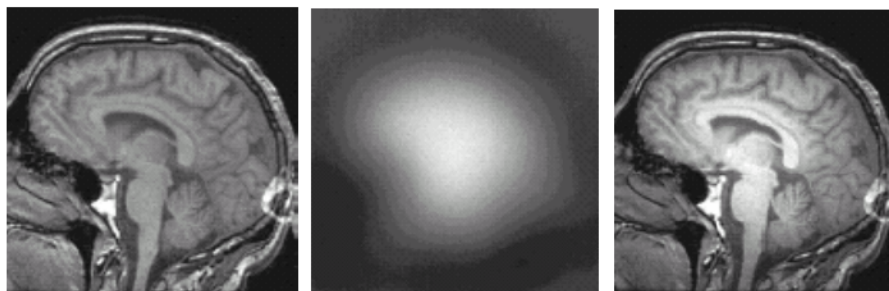


FIGURE 2.7 – INU dans une IRM due à un champ de biais [7].

est due principalement à l’inhomogénéité au sein des mêmes tissus et l’inhomogénéité du champs magnétique produit par la machine IRM.

La figure 2.7 montre comment une IRM (image à gauche) est affectée par un biais du champ (image au centre) pour produire une IRM affectée par l’INU (image à droite).

Par sa difficulté à la détecter localement dans les IRM, la non uniformité des intensités est traitée par des méthodes particulières qui lui sont dédiées [53]. Cependant, les premières méthodes était basées filtrage en considérant l’INU comme une sorte de bruit [54, 55]. La plupart des méthodes de la littératures [53] ont procéder à une estimation du champs de biais par une surface, puis en utilisant cette dernière, les intensités sont corrigées [56, 53]. Récemment, l’apprentissage profond est utilisé pour estimer le champs de biais et puis corriger les intensités [57].

## 2.4 IRM cérébrales

Une IRM cérébrale est réalisée sur l’encéphale humain, où les médecins s’intéressent généralement à la matière cérébrale composant le cerveau humain. Il s’agit respectivement de la matière grise (GM pour Gray Matter), situé sur le cortex du cerveau, la matière blanche (WM pour White Matter) situé au centre du cerveau, et le liquide céphalo-rachidien (CSF pour Cerebro Spenal Fluid), qui entoure la matière grise et la matière blanche (Figure 2.8).

### 2.4.1 Traitements des IRM cérébrales

Certains traitements en IRM sont généraux, et dont les plus importants sont respectivement, le débruitage, le recalage (ou appariement), et la segmentation. Cependant, certains traitements sont spécifiques aux IRM cérébrales, dont l’élimination du crane et

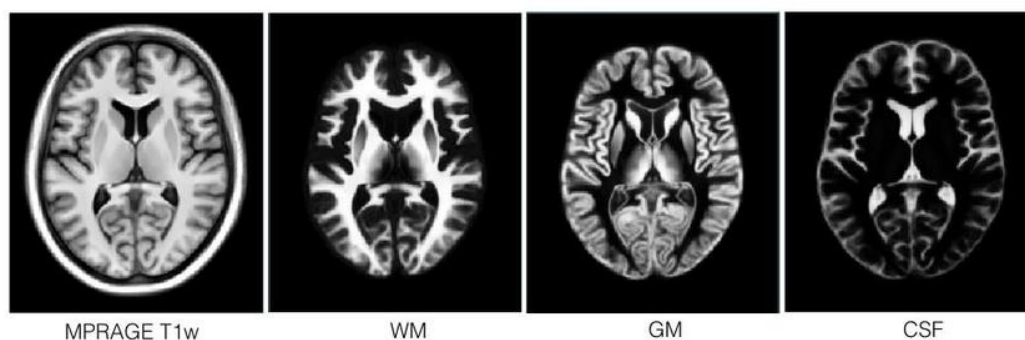


FIGURE 2.8 – Matière cérébrale dans une IRM [8].

la détection et la segmentation de tumeurs cérébrales. Malgré qu'il s'agit d'un traitement commun aux images en général, les méthodes de segmentation d'IRM devraient prendre en considération les artéfacts spécifiques aux IRM. Généralement, les traitements d'une IRM sont menés d'une manière séquentielle, en commençant par les prétraitements d'amélioration de la qualité, dont le lissage de bruit et la correction des intensités. Cependant, plusieurs travaux et après avoir formulé conjointement la formation du signal et les artéfacts à considérer, ont proposé des méthodes conjointes [58]. Nous présentons dans la suite de la section, les principaux prétraitements qui sont exclusivement spécifiques aux IRM cérébrales.

### Élimination du crane

Traduit de l'anglais "Skull Stripping", ce traitement consiste à supprimer de l'IRM tous les tissus entourant le cerveau, dont le crane principalement. Il s'agit d'un traitement délicat même quand il est réalisé manuellement par le radiologue, car toute erreur de l'extraction du cerveau peut engendrer des fausses interprétations et de faux diagnostics. La figure 2.9 montre une IRM avant (à gauche) et après (à droite) l'élimination du crane.

Une méthode de référence est proposée par Smith en 2002 [10], est à la base d'un outil célèbre pour l'extraction du cerveau dans une IRM cérébrale, dit BET pour "Brain Extraction Tool". La méthode est basée sur un modèle déformable qui s'initialise et tend à s'ajuster à la surface du cerveau (voir Figure 2.10) en utilisant un ensemble de modèles de forces locales.

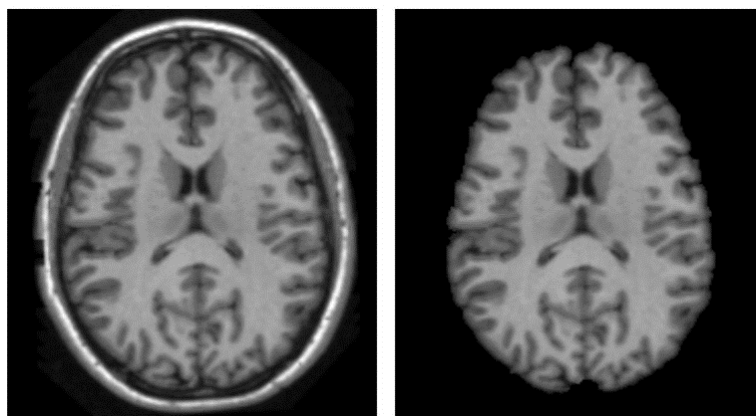


FIGURE 2.9 – Elimination du crane dans une IRM [9]

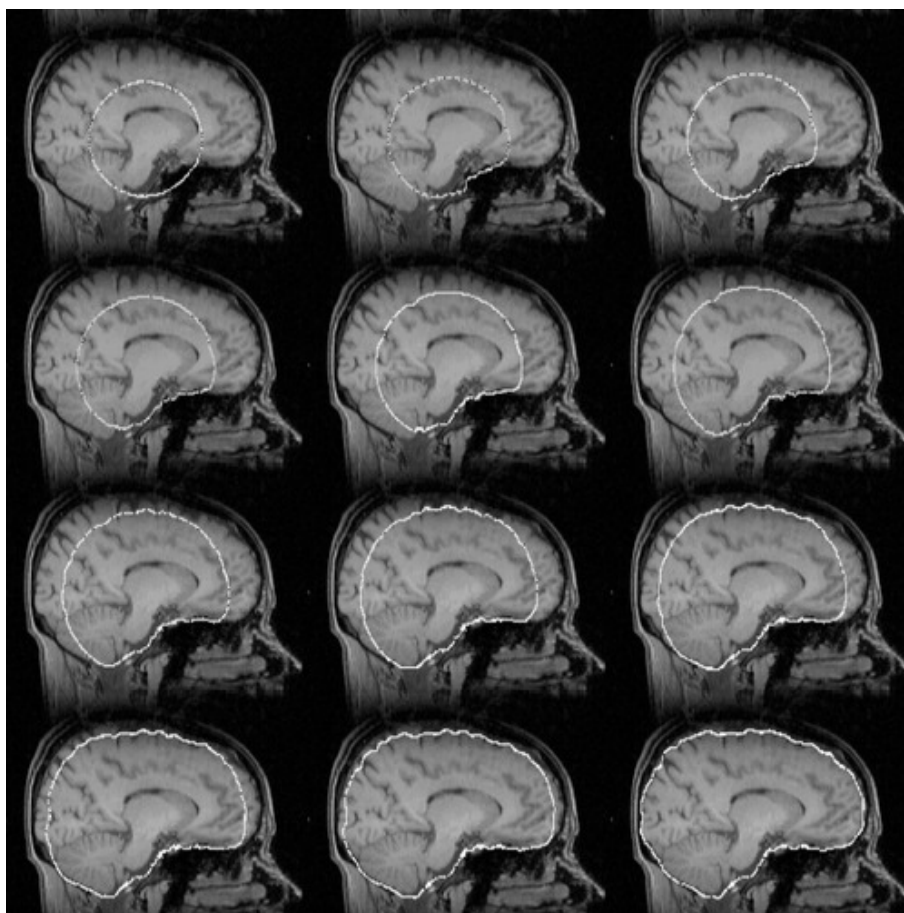


FIGURE 2.10 – Evolution du modèle déformable pour s'ajuster au cerveau [10]



FIGURE 2.11 – Correction de l'INU en IRM [11].

### Correction de l'INU

Comme nous l'avons mentionné plus haut, la correction de l'INU est délicate et ne peut être assimilée à un débruitage de l'image. La variation lente du champ de biais rendent tout traitement local inefficace, car localement, la non-uniformité ne peut pas être constaté ni fiablement quantifiée [53]. De ce fait, plusieurs approches, basées sur divers paradigmes computationnels ont été proposées dans la littérature. Cependant, la modélisation du champ de biais par une surface est l'approche la plus utilisée [53]. La figure 2.11 montre comment la non-uniformité (remarquée dans l'IRM à gauche) a été corrigée (dans l'IRM à droite), avec la modélisation du champ de biais (dans l'image du centre).

## 2.5 Segmentation d'IRM

La segmentation des images en général, est l'une des tâches les plus importantes dans le processus de reconnaissance des formes à partir de données visuelles. Ça consiste à subdiviser l'image en régions distinctes et homogènes, qui correspondent à des objets intelligibles.

Formellement, une image  $I$  est une fonction qui associe à chaque point de l'espace  $(x, y, z)$  de l'image une grandeur scalaire représentant l'intensité du point ou un vecteur représentant sa couleur. Une segmentation de l'image  $I$  est une partition de ses points en plusieurs régions  $\{R_i, i = 1..N\}$  disjointes et dont l'union forme l'image en entier. Au sein d'une même région les points partagent une propriété commune, dite critère d'homogénéité de la région, noté  $(C)$ .

Cette définition nous conduit à formuler la segmentation comme suit [59] :

$$\bigcup_{i=1}^N R_i = I$$

$$R_i \cap R_j = \emptyset, \forall i, j, i \neq j$$

$$C(R_i) = True, \forall i$$

$$C(R_i \cup R_j) = False, \forall i, j, i \neq j$$

Cette formulation exprime une segmentation idéale, où tous les points de l'images ont été affectés à leurs régions respectives et qu'un point n'est affecté qu'à une et une seule région. Par contre et avec les images réelles, un ensemble plus ou moins important de points ne se retrouvent pas affectés à des régions, on parle dans ce cas de points de bruit. Il est aussi difficile de définir un critère d'homogénéité qui permet une vraie partition disjonctive des régions et même dans la réalité les régions peuvent se chevaucher, où des frontières floues peuvent les séparer.

Durant un demi siècle de recherche dans le domaine, des dizaines de méthodes de segmentation différentes ont été développées. Cependant, toutes ces méthodes peuvent être classées selon trois familles principales :

1. Méthodes basées sur les contours : Le principe commun de ces méthodes est de détecter les discontinuités dans les données visuelles. Ces discontinuités représentent les bords des régions. Les bords qui en résultent de la segmentation sont généralement disjoints et ouverts, et doivent donc être joints et fermés pour une utilisation appropriée dans le processus de reconnaissance ultérieur. Ces méthodes sont rares pour la segmentation IRM, car elles sont très sensibles au bruit et aux autres artéfacts. Cependant, les contours actifs et la méthode du partage des eaux ont été utilisées pour la segmentation des IRM [60, 61], car ils sont connus pour leur faible sensibilité aux artéfacts IRM, à savoir le bruit, la non uniformité d'intensité (INU) et l'effet du volume partiel. Certains auteurs ont combiné plusieurs techniques pour résoudre les problèmes de détection de contours dans les IRM, dus principalement à la complexité des structures anatomiques révélés dans les IRM, notamment les IRM cérébrales. Par exemple, Han et Chen [12], et afin de pouvoir proposer une méthode robuste pour la détection de contours dans les images médicales, ont utilisé la technique d'analyse de données ICA (Independent Component Analysis). Les résultats de détection ainsi obtenus sont nettement supérieurs à ceux obtenus par le détecteur de référence Canny (voir figure 2.12).

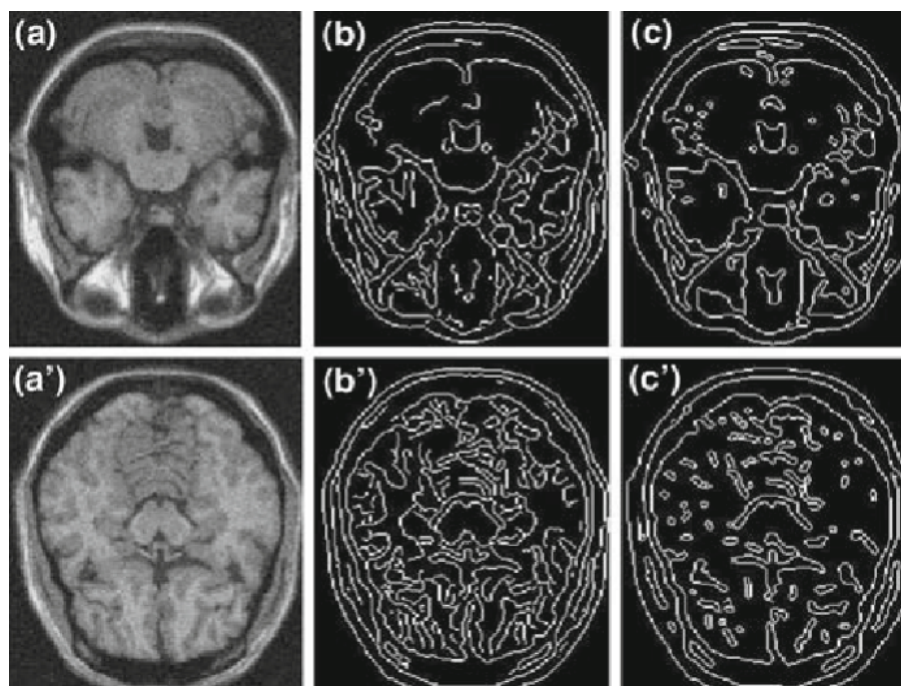


FIGURE 2.12 – Détection de contours basée ICA en IRM : a, a' IRM, b, b' Contours par le détecteur de Canny, c, c' contours par la méthode basée ICA [12]

2. Méthodes basées régions : leur principe consiste à regrouper les pixels/voxels de l'image, ayant les mêmes caractéristiques, en sous-ensembles disjoints mais homogènes selon un certain critère d'homogénéité. Ces sous-ensembles homogènes sont appelés régions. Les méthodes à base de régions sont encore utilisées pour la segmentation d'IRM et la détection des tumeurs, comme les travaux introduits par Bennai et al. [41], où les auteurs dans ce travail ont utilisé l'accroissement de région 3D pour extraire les trois régions formant les tumeurs cérébrales 2.13. L'accroissement des régions est distribué en utilisant un système multi-agent. Les régions extraites par les différents agents peuvent être fusionnées en utilisant un protocole de négociation adéquat.
3. Méthodes basées sur l'apprentissage automatique : leur principal atout est qu'elles permettent l'apprentissage à partir des données, où la plupart des méthodes sont basées sur la classification supervisée. La segmentation par classification consiste à attribuer une étiquette à chaque pixel/voxel de l'image à l'aide d'un classifieur (simple ou d'ensemble). Étant donné que nous nous intéressons à cette dernière famille dans cette thèse, nous consacrons le reste de la présente section à l'introduction de certaines méthodes de segmentation de l'IRM par classification des données.

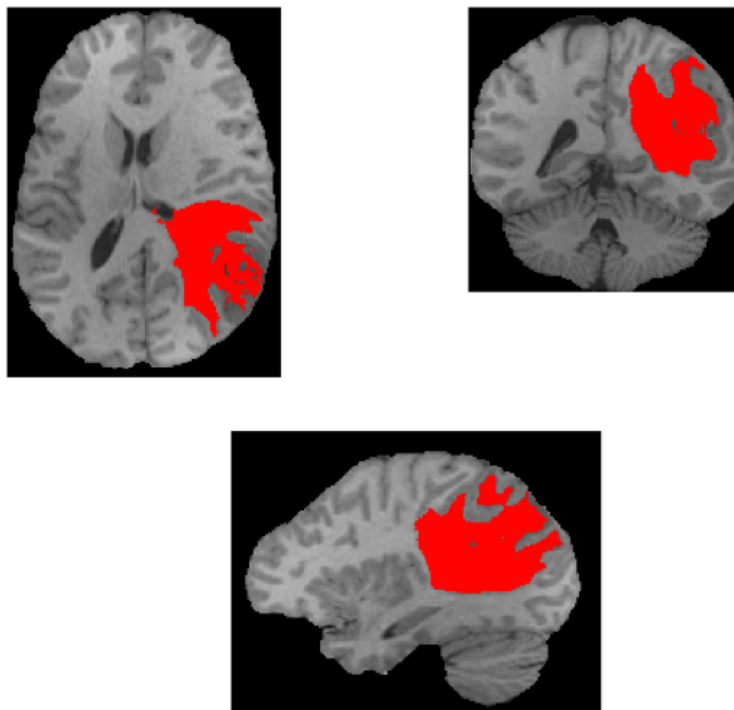


FIGURE 2.13 – Région de la tumeur cérébrale selon les trois axes (sagittal, coronal, et axial) extraite par un accroissement multi-agent, présentée dans [12]

### 2.5.1 Segmentation d'IRM par classification de données

Bien que les méthodes basées sur les contours ou basées sur les régions aient été les premières utilisées pour la segmentation d'IRM, la segmentation basée sur la classification est maintenant largement adoptée. L'émergence de ce type de méthode est à attribuer principalement à la mise à disposition des chercheurs de plusieurs bases d'IRM avec réalité terrain de segmentation, opérée par plusieurs médecins et radiologues. Selon notre point de vue, les méthodes de segmentation d'IRM par classification peuvent à leur tour être divisées en deux sous-ensembles distincts : Méthodes heuristiques et méthodes par apprentissage automatique.

#### Méthodes heuristiques

On qualifie une méthode de telle, lorsqu'une ou plusieurs heuristiques sont considérées pour définir un critère d'étiquetage des pixels/voxels. L'heuristique considère un préalable donné, lié à l'image, au bruit, ou aux distorsions que l'image pourrait subir lors de son acquisition. Par exemple, on peut citer l'algorithme Fuzzy C-Mean [62], où l'hypothèse de classification consiste à considérer pour les pixels/voxels aux frontières des régions et

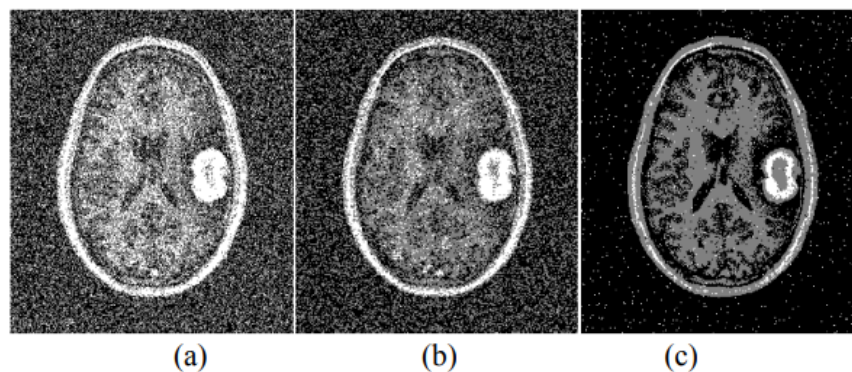


FIGURE 2.14 – Améliorations du FCM par l'introduction de l'information spatiale : (a) FCM standard, b) SFCM, (c) MS-FCM [13]

ailleurs qu'il y a un mélange d'informations, chacune se rapportant à l'une des classes de données. Plusieurs travaux ont proposé des méthodes FCM modifiées pour la segmentation des tumeurs cérébrales [63]. En effet, l'information spatiale était introduite par plusieurs auteurs pour améliorer les performances du clustering des voxels dans les IRM. Dans son travail, Hamed Shamsi [13] a pu introduire deux améliorations de l'algorithme du FCM, baptisées respectivement 'SFCM' pour 'Spatial FCM' et 'MS-FCM' pour 'Modified Spatial FCM'. Dans les deux cas une fonction spatiale dont l'utilisation dans l'algorithme du FCM a permis de bien accentuer les degrés d'appartenance des voxels à leurs classes respectives. Les résultats des améliorations présentés par l'auteur sont nettement meilleurs que ceux du FCM standard (voir figure 2.14).

Les méthodes markoviennes [64] considèrent le préalable de lissage, où les données sont considérées comme homogènes par parties, et toute partie correspond à une région homogène de l'image. De plus, les représentations markoviennes peuvent exprimer certaines contraintes spatiales que les données doivent respecter. Un travail typique pour la détection des tumeurs cérébrales dans les IRM utilisant des modèles markoviens, est le travail introduit par Subbanna et al. [14]. Dans ce travail, les auteurs ont combiné le filtre de Gabor et le Framework bayésien markovien (voir figure 2.15). Les résultats de segmentation de tumeurs cérébrales, obtenus avec la base de données BraTS, ont été très concluantes.

### Méthodes par Apprentissage automatique

Le principe commun de ces méthodes est de procéder en apprenant les classificateurs à l'aide de données étiquetées, appelées ensemble d'apprentissage, puis d'utiliser les classi-

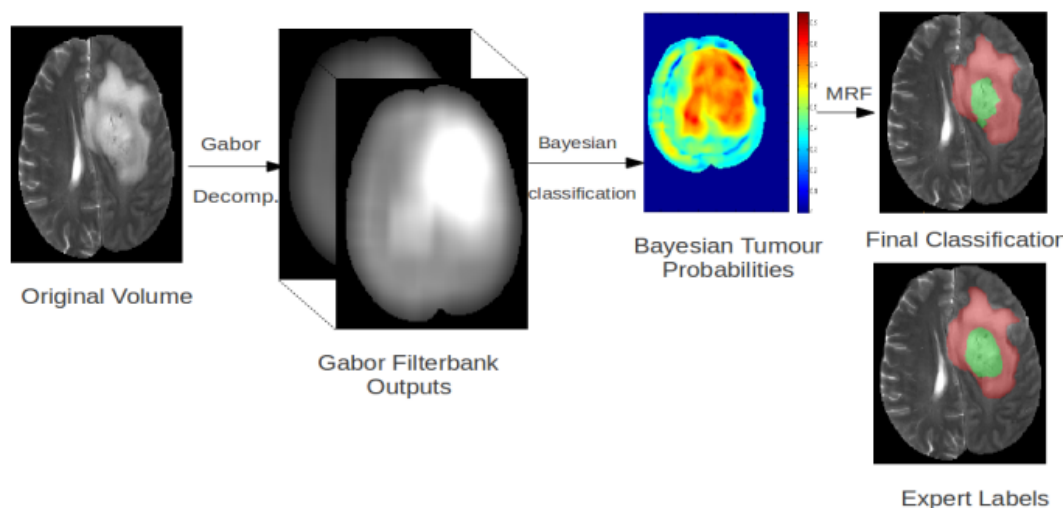


FIGURE 2.15 – Framework markovien pour la segmentation des tumeurs cérébrales proposé dans [14]

fiereurs entraînés pour classer les données, dans ce cas appelées ensemble de test. Selon une telle approche, plusieurs nouvelles méthodes basées sur la combinaison des classificateurs ont émergé. L'ensemble des classificateurs est l'une de ces méthodes, et elle est basée sur l'agrégation d'un ensemble de classificateurs, appelés classificateurs faibles, formés séparément selon un échantillonnage particulier, permettre d'une part l'amélioration de la performance de la classification et d'autre part, traiter certains problèmes inhérents à l'apprentissage automatique, dont le plus connu est le problème de surapprentissage.

## 2.5.2 Revue de la littérature

La segmentation IRM afin d'extraire des tissus et d'établir un diagnostic reste un domaine de recherche actif jusqu'à nos jours, [65, 41]. Pour une IRM, sa segmentation consiste à extraire les principaux tissus qui intéressent principalement les médecins et les radiologues. En général, les tissus d'intérêt dans la segmentation structurale des IRM, sont les trois matières qui composent les tissus cérébraux, la matière grise, la matière blanche, et le liquide céphalo-rachidien. Pour la segmentation des tumeurs cérébrales, les tissus d'intérêt concernent la tumeur et ses environs et sont respectivement le noyau, la partie active et l'œdème, et le reste est considéré comme des tissus sains.

## Méthodes par apprentissage d'ensemble

Plusieurs méthodes de segmentation par IRM ont été publiées, les plus anciennes étant basées sur les contours ou les régions. Plus tard, plusieurs travaux ont proposé des méthodes à base d'apprentissage automatique pour la segmentation IRM, où le principal attribut utilisé était l'intensité du voxel. De plus, seuls quelques travaux ont utilisé des classifieurs d'ensemble, où dans la plupart des cas, la sélection et l'extraction des attributs ont été effectuées. Rajasree et al. ont considéré une représentation fractale des données IRM, en utilisant la technique du mouvement brownien [66]. Les attributs obtenus sont ensuite utilisés avec l'algorithme Adaboost pour détecter les tumeurs dans les données IRM. Gustavo et al. ont combiné les algorithmes génétiques (AG) et le clustering Adaboost pour détecter la zone tumorale dans les IRM [15] (figure 2.16). Après un seuillage des données à l'aide de l'algorithme génétique afin de délimiter la zone tumorale, Adaboost est entraîné à l'aide de la première classification obtenue par l'algorithme génétique, puis utilisé pour finalement détecter la tumeur comme le plus grand composant connexe de toute l'image. Fulop et al. ont proposé une méthode basée sur l'apprentissage d'ensemble pour la segmentation des tumeurs cérébrales. Ils s'appuient sur un multi-atlas, et utilisent des arbres de décision binaires comme classifieurs faibles [16]. La figure 2.17 montre le diagramme d'enchaînement des tâches selon la méthode proposée par les auteurs.

## Méthode par apprentissage profond

Récemment, des techniques d'apprentissage profond, principalement en utilisant les réseaux de neurones convolutifs (CNN) ont été largement proposées pour le traitement des données IRM [67, 68, 69, 70]. Leur grand avantage est qu'ils n'ont pas besoin de représentation, de sélection et d'extraction de caractéristiques. Dans de telles techniques, les données IRM à l'entrée sont convoluées en noyaux dans les couches intermédiaires. Ainsi, les attributs sont automatiquement produits et utilisés par les couches entièrement connectées pour étiqueter les voxels [71, 17]. Parmi les premiers travaux ayant fait recours à l'apprentissage automatique, est celui de Alexandre de Brebisson et Giovanni Montana [17], où ces auteurs ont proposé une méthode profonde pour la segmentation d'IRM en régions anatomiques (voir figure 2.18). Pour pouvoir s'en passer du recalage des IRM, les auteurs ont utilisés des patches 3D et des patches longitudinales pour détecter le contexte local. Coupé et al. ont proposé un cadre à base d'un grand ensemble de CNN pour la seg-

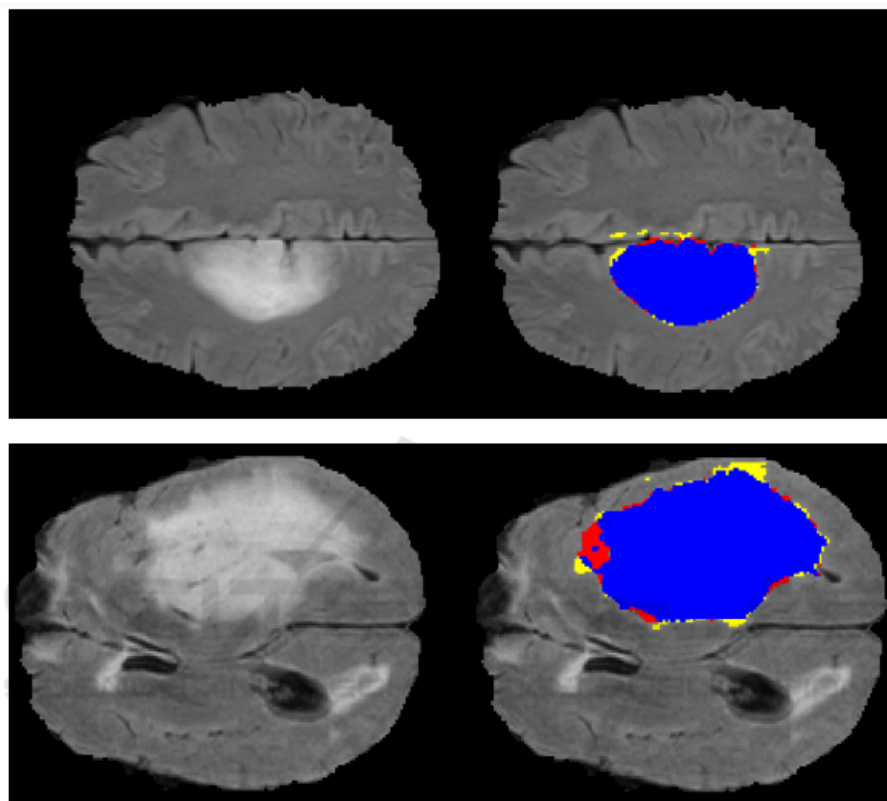


FIGURE 2.16 – Délimitation des tumeurs cérébrales dans des IRM. Extrait de l'article [15]

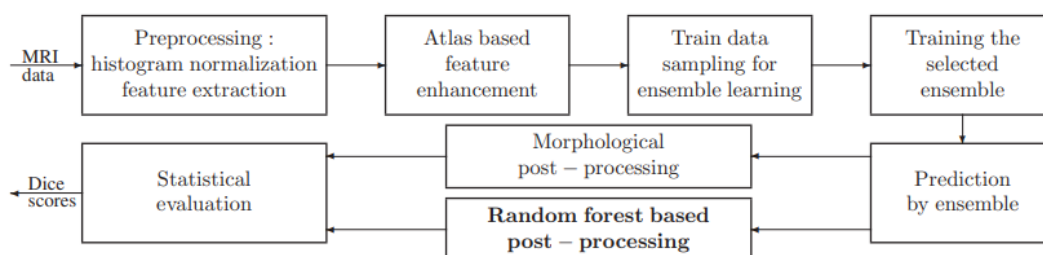


FIGURE 2.17 – Enchaînement des tâches pour la segmentation des tumeurs cérébrales selon [16]

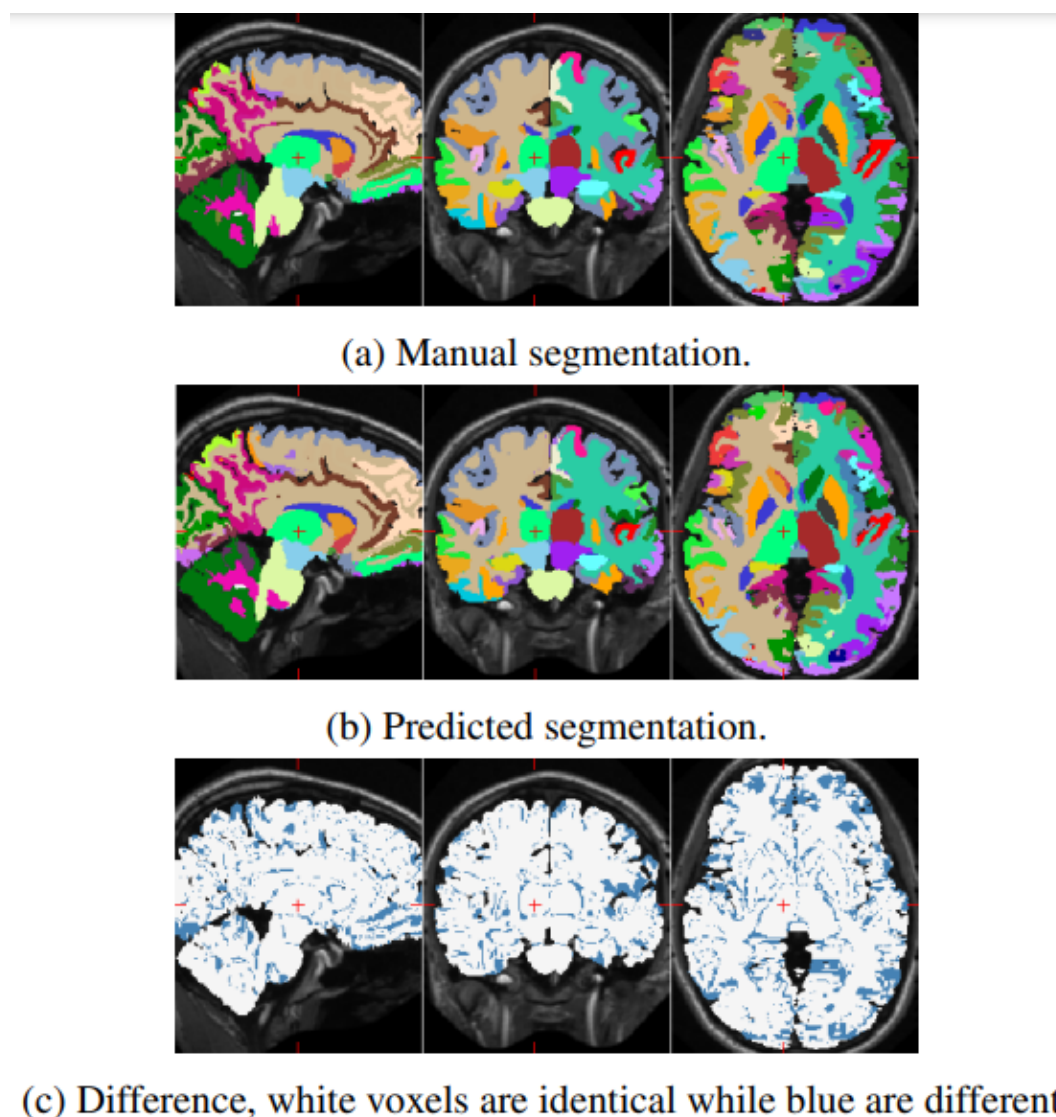


FIGURE 2.18 – Segmentation en régions anatomiques par une méthode profonde, proposée dans [17]

mentation IRM du cerveau entier [72]. Les auteurs de ce travail ont utilisé plusieurs CNN afin de couvrir les différentes zones anatomiques du volume d'IRM du cerveau. Plusieurs autres travaux ont proposé différentes méthodes basées sur l'apprentissage automatique pour la segmentation des tumeurs cérébrales, y compris des méthodes profondes et utilisant des CNN sont publiés dans la littérature.

En résumé de cette courte revue de littérature, nous pouvons constaté que le long des dernières dizaines d'années, toute la panoplie des méthodes de segmentation d'images a été appliquée aux IRM. Par la nature spécifique de l'anatomie du cerveau qui représenté pas des régions distinctes de tailles variables dans les volumes IRM, les méthodes basées régions étaient les plus utilisées durant les années 1990 et les années 2000. Les méthodes

basées contours sont peu utilisées, car le niveau de bruit est important dans les IRM, et toute tentative pour y remédier impliquera une dégradation de l'acuité de l'image et par conséquent une dégradation de la qualité de sa segmentation. Avec l'avènement des premières bases d'images (Ex. Brain web pour IRM de simulation, et IBSR pour IRM réelles), ainsi que l'enrichissement de ses bases par des segmentations réalité terrain, des méthodes à bases d'apprentissage automatiques ont commencé à voir le jour à la fin des années 2000 et au début des années 2010. A cet époque, et en se basant sur ces bases d'images, plusieurs méthodes de segmentation à base d'Atlas ont été proposées. Les Atlas, et les multi-altas, ont permis la représentation des connaissances a priori, ouvrant ainsi la porte à l'utilisation de méthodes bayésienne et markoviennes pour la segmentation d'IRM. A la fin des années 2010 et au début des années 2020, les bases de données d'IRM sont devenue si volumineuses pour pouvoir utiliser des modèles d'apprentissage profond pour divers traitements sur les IRM dont la segmentation. Par sa nature bien appropriée aux données matricielles, comportant des motifs à apprendre et à utiliser pour prédire, les CNN se sont imposées comme modèles d'excellence pour la segmentation d'IRM, que ce soit saines ou pathologique.

## 2.6 Conclusion

Ces dernières années l'IRM est devenue la technique radiologique la plus utilisée pour détecter les différentes lésions qui affectent les organes humains, et principalement pour le diagnostic des tumeurs. Dans ce premier chapitre de notre thèse, nous nous sommes penché sur les IRM, leur technique d'acquisition, leur représentation et les différents artéfacts qui leur sont associés, qui rend leur traitement problématique. Nous avons également présenté les différents traitements des IRM, nécessaire l'interprétation de ces dernières, dont certains traitements qui sont spécifiques aux IRM tels que l'élimination du crane et la correction des non-uniformités des intensités. La segmentation des IRM est finalement présentée avec les différentes approches, dont l'approche par apprentissage et apprentissage profond. Au chapitre suivant, nous le consacrons au paradigme de segmentation utilisé dans nos travaux à savoir l'apprentissage automatique, où une importante partie est consacrée aux réseaux de neurones convolutifs, outils que nous avons utilisé dans nos travaux.

## Chapitre 3

# Apprentissage automatique : Origines et aboutissement

## 3.1 Introduction

Les années 2010 ont vu l'émergence et la généralisation de l'utilisation de l'apprentissage automatique (AA), comme sous-domaine de l'intelligence artificielle (IA) dans divers domaines de la science et de ses applications. Les années d'avant, l'apprentissage automatique selon ses modèles classiques, dont les différents réseaux de neurones, était à la base de plusieurs modèles de résolution de problème, principalement, en reconnaissance de formes, en aide à la décision, et en classification de données. En vision par ordinateur, qui relève, dans son essence de la reconnaissance de formes, l'apprentissage automatique était utilisé dans les différentes étapes du processus de reconnaissance d'objets, selon le modèle standard de David Marr [73], et ce en partant des pré-traitements des données visuelles et en arrivant à l'étape de reconnaissance et d'interprétation. Les modèles classiques d'apprentissage automatique souffrent de deux problèmes majeurs : d'une part la définition des caractéristiques des données qu'on devrait utiliser et leur sélection, et d'autre part, l'inadéquation de ses modèles aux données massives, qui commencent à être présentes après des décennies de numérisation et d'archivage. L'avènement d'un certain type de processeurs, dits graphiques (GPU : Graphical Processing Unit) a permis, au début des années 2010, d'expérimenter des modèles de réseaux de neurones profonds et les résultats ne s'attardaient pas à devenir spectaculaires. Ainsi, l'air de l'apprentissage profond fut commencé. Dans ce chapitre, nous survolons les techniques d'apprentissage automatique, et nous montrons à la fois la spécificité de ces techniques et leurs avantages et inconvénients respectifs.

## 3.2 Une brève histoire de l'intelligence artificielle et de l'apprentissage automatique

Si en 9<sup>ème</sup> siècle, Mohamed Ibn Moussa Al-khawarsmi, le célèbre algébriste de la civilisation arabo-musulmane, fut écrit les premiers algorithmes en devenant le premier à avoir ouvert la mécanisation des mathématiques, et si en dix-neuvième siècle, Ada, la comtesse de Lovelace, fut écrit le premier programme sur la machine de George Babbage, et devenant ainsi le premier informaticien de l'histoire, ce n'est qu'en 1956 que John McCarthy, et Marvin Minsky ont donné le nom « Intelligence artificielle » à une discipline qui commence à naître et qui consiste à proposer des modèles et construire des machines imitant les facultés intellectuelles de l'homme. Partant de ce but constitutif, l'intelligence artificielle est définie actuellement comme la science qui traite des approches, des méthodes et des techniques, qui permettent à la machine d'imiter les facultés intellectuelles de l'être humain [74]. Cependant, c'est bien avant McCarthy et Minsky du Dartmouth College, que le célèbre britannique, Alan Turing, en travaillant sur la calculabilité des problèmes en mettant en œuvre des automates de calcul, qu'on appelle actuellement "machine de Turing", a mis le célèbre test d'intelligence. Ce test consiste à questionner une machine supposée intelligente, et un être humains, se trouvant les deux derrière un écran. Si celui qui les questionne, n'arrive pas à distinguer les réponses de l'homme de celles de la machine, on conclue que cette dernière est intelligente.

En effet, c'est avant même l'invention des ordinateurs, que les scientifiques des années 1940 ont essayé de modéliser l'esprit humain en inventant le concept de cybernétique, dont Norbert Wiener fut le pionnier [75]. A cette époque même, et avec les avancées dans le domaine de la neurologie, certains scientifiques ont commencé à modéliser le cerveau humain par des neurones formels (Warren McCulloch et Walter Pitts) [76]. A ce temps reculé de l'histoire de l'informatique, fut donc l'avènement du connexionisme en IA. Malgré le fondement biologique du connexionisme, il s'est rapidement avéré qu'il ne sera pas implémentable sur des machines, vu le nombre hallucinant de neurones nécessaires dont on devrait modéliser, et en prenant également en comptes les réseaux synaptiques qui en découlent avec. À cet effet le cognitivisme vient en alternative, comme approche de modélisation de l'intelligence et du comportement humain. Ainsi, la pensée, qui représente la faculté ultime de l'intelligent, consistera en des manipulations de symboles, en

considérant l'hypothèse d'existence de liens étroits entre la pensée et le langage.

Au fil des années, l'intelligence artificielle s'est scindée en deux principaux courants : l'IA connexionniste, empruntant des fondements de la neurologie et du computationnel, et l'IA cognitive, qui emprunte de la logique mathématique. Les réalisations en IA connexionniste se sont heurtées à la faiblesse des calculateurs des années 60 et 70. Ceci a permis à l'IA cognitive (symbolique) de prendre les devants, régner durant les deux décennies suivantes, et dont la principale réalisation était les systèmes experts. Cependant, et avec les avancées technologiques qui ont permis d'augmenter considérablement la puissance de calcul des ordinateurs, le courant connexionniste est renaît et l'apprentissage automatique a commencé à se développer.

Historiquement, l'apprentissage automatique (Machine Learning) est cité la première fois par Arthur Samuel au début des années 1950 [77], quand il a pu implémenter un programme sur un IBM qui joue au jeu de dames, mais il s'améliore en jouant. L'apprentissage automatique s'est vu popularisé suite au succès de Deep Blue, machine IBM, qui a vaincu le champion du monde en jeu d'échec, Garry Kasparov en 1997. Cependant, les années 2010 ont vu l'explosion des modèles et des applications de l'apprentissage automatique, avec l'avènement des architectures GPU et l'implémentation dessus de modèles de réseaux de neurones profonds, jusqu'au là non implémentable sur des machines conventionnelles, et ce à cause du nombre de paramètres des réseaux à estimer, et qui se comptent en nombre de millions et de dizaines de millions, même pour des applications simples [78].

L'intelligence artificielle a vécu deux crises majeures durant son courte histoire, et qui n'a débuté que depuis 60 ans. Malgré l'enthousiasme qui a accompagné ses débuts, qui s'est occasionné avec la construction de calculateurs modernes et à la mise au point de langages de programmation évolués, les pionniers de la discipline se sont heurtés à diverses problèmes, dont les principaux sont : la complexité des calculs, et l'ambiguïté des connaissances manipulées. Durant les décennies 60 et 70, plusieurs échecs se sont enregistrés, notamment en compréhension du langage naturel et en traduction automatique. Durant ces décennies, l'IA s'est réduite aux systèmes experts et à la représentation des connaissances avec des succès modérés. En parallèle, les réseaux de neurones et par la difficulté de concevoir des réseaux profonds, ont vécu un déclin durant les années 80 et 90. Et ce n'est qu'à partir de 2000 qu'ont vu un rebond, avec la mise sur le marché de

processeurs puissants et de mémoires à grandes capacités. Si une crise future de l'IA est prévue, elle serait certainement de nature éthique. En effet, plusieurs chercheurs dans différents domaines ont attiré l'attention à de néfastes conséquences de l'IA sur la condition humaine et sur son cours de développement [79].

### 3.3 Apprentissage automatique : Définition et taxonomie

#### 3.3.1 Définition

L'apprentissage automatique, ou l'apprentissage machine, est la branche de l'intelligence artificielle qui traite des approches, méthodes et techniques basées sur les mathématiques et les statistiques et qui permettent à des ordinateurs d'apprendre et d'améliorer leur performance et exploitant les données disponibles [80]. Classiquement, et selon une vision basée modèle, l'ordinateur résout les problèmes qui lui sont posés en exécutant un programme préalablement écrit et immuable dans le sens où sa structure et son comportement ne changent pas dans le temps ni en fonction des données. En apprentissage automatique et selon cette optique, l'ordinateur arrive à traiter et décider devant des situations qui n'ont pas été prévues par son programmeur. La notion de pattern est fondamentale en apprentissage automatique, car selon ce concept, l'ordinateur procède à la décision suite à la reconnaissance des patterns cachés dans les données à traiter [81]. De ce fait, l'apprentissage automatique implique la notion d'entraînement des modèles d'inférence ou de décision. Pour ce faire, on devrait disposer de données en quantités suffisantes pour que les modèles qui en résultent de l'apprentissage puissent satisfaire les critères de performance requises.

#### 3.3.2 Une taxonomie des méthodes d'apprentissage automatique

Selon le mode d'apprentissage des modèles, les méthodes en apprentissage automatique, sont classifiées selon les catégories suivantes :

## Apprentissage Supervisé

Les méthodes d'apprentissage automatique supervisées devrait disposer des données d'apprentissage sous formes d'entrées et de sorties associées. Autrement dis, on devrait avoir pour les données d'entraînement, de leurs classes respectives. Ces classes sont obtenues par des experts ou des oracles, préalablement aux processus d'apprentissage [82]. En segmentation d'IRM à titre d'exemple, les données d'entraînement sont composée pour chaque voxel, de son intensité et d'une étiquette indiquant la classe à laquelle appartient ce voxel.

## Apprentissage non-supervisé

En apprentissage non-supervisé, l'objectif est généralement de partitionner les éléments en entrée, en sous ensembles homogènes vis à vis d'un critère donné de similarité, présent dans les données. Le partitionnement peut se faire en considérant la distance entre les éléments, projetés dans un hyper-espace de dimension égale aux nombre d'attributs des données (Moyennes mobiles) [83]. Comme il peut se faire en se basant sur la répartition des éléments selon une distribution de probabilités, dont on suppose que les variables suivent (Maximisation de vraisemblance) [84]. En segmentation d'IRM, l'apprentissage non supervisé est utilisé pour regrouper les voxels dans des parties (pas nécessairement connectées) et qui partagent des caractéristiques similaires.

## Apprentissage par renforcement

L'apprentissage par renforcement postule un agent autonome qui explore un environnement, et dans lequel il entreprend des actions qui peuvent lui être bénéfiques ou non. Lors de son exploration l'agent cherche à maximiser sa récompense totale, en cumulant les récompenses positives qu'il obtient après toute action réussite le menant à son objectif, et en évitant les actions qui lui résultent en une pénalisation négative. Ainsi, l'agent apprend de son expérience et au fil du temps, l'agent se procura une stratégie décisionnelle qui consiste à bien choisir l'action à exécuter en fonction de son propre état et l'état de l'environnement, à la place où il est situé [85].

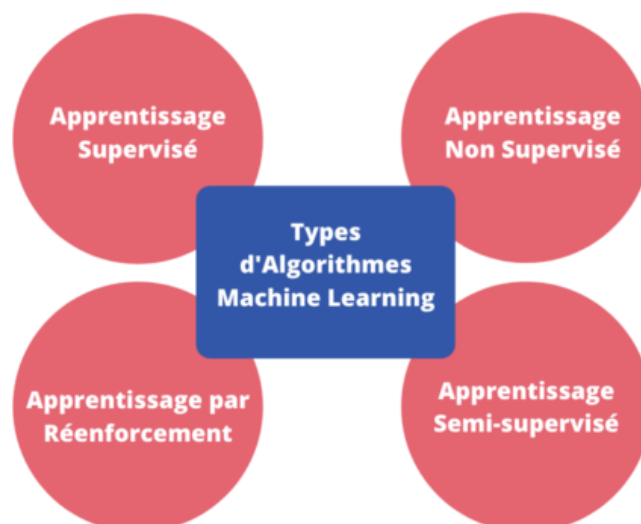


FIGURE 3.1 – Types d'apprentissage automatique

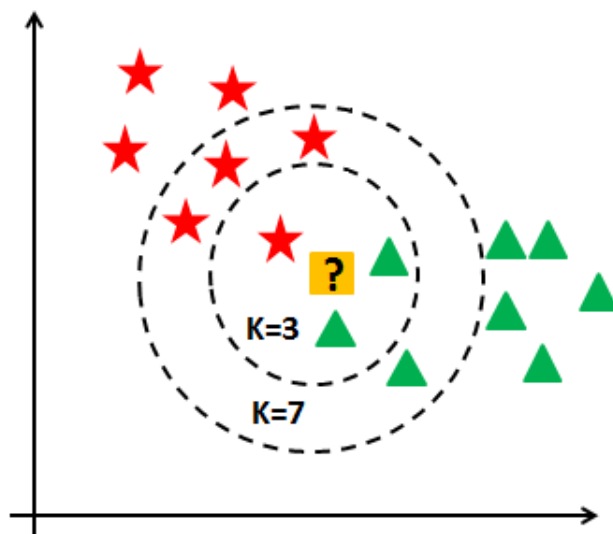
### Apprentissage semi-supervisé

Dans ce mode d'apprentissage on utilise un ensemble de données étiquetées, comme en apprentissage supervisé, et un ensemble de données non-étiquetées, comme en apprentissage non supervisé. En effet, l'étiquetage de données est généralement effectué par des humains experts, ce qui rend la tâche difficile surtout si on souhaite que les données étiquetées soient nombreuses, ce qui est nécessaire pour l'entraînement des modèles profonds par exemple. C'est en cette situation que les données non étiquetées sont utilisées. L'expert n'a qu'à étiqueter qu'un sous ensemble de taille raisonnable, et le reste est effectué automatiquement selon différentes méthodes mathématiques, ou issu de l'apprentissage automatique [86].

La figure 3.1 montre les différents types d'apprentissage, sous lesquelles se placent les différentes techniques, et dont nous présentons les plus importantes dans les sections qui suivent.

## 3.4 Quelques techniques d'apprentissage supervisé

Les techniques d'apprentissage supervisé sont nombreuses et nous ne pouvons pas leur réserver toute une présentation dans ce manuscrit. Cependant, nous essayons d'introduire dans la suite de cette section, les techniques les plus importantes, en termes de leur utilisation au sein de la communauté d'apprentissage automatique, mais aussi en termes de leurs fondements respectifs et dont nombreuses d'autres techniques se sont basées sur

FIGURE 3.2 – Exemple de classification avec  $k$ NN [18]

ces dernières.

### 3.4.1 Les $k$ plus proches voisins

Connue par le nom de  $k$ NN, pour  $k$  Nearest Neighbors, il s'agit d'une technique d'apprentissage supervisé, dont le principe consiste à affecter à l'élément à étiqueter, la classe des  $k$  éléments d'apprentissage qui leur sont les plus proches [87]. En considérant le principe de cette technique, il est clair qu'aucune phase d'apprentissage préalable à la classification n'est affectée. Les données d'apprentissage sont présentes, mais elles ne seront utilisées que lors de la prédiction. Il est également à noter que les  $k$  éléments les plus proches n'ont pas forcément la même classe. Donc un vote majoritaire, ou une autre technique équivalente, est à considérer pour déterminer la classe finale (voir figure 3.2).

Le choix du paramètre  $k$  est problématique. En effet, la classe qui en résulte dépend de ce paramètre, et les résultats de test varient considérablement d'une valeur à une autre. Dans la figure 3.2, le changement de  $k$  de 3 à 7 a modifié la classe de l'élément à étiqueter.

### 3.4.2 Machine à vecteurs de support

Il s'agit d'une technique supervisée largement utilisée, qu'on abrégé par SVM (pour Support Vector Machine) et qu'on appelle également, Séparateur à Vaste Marge. Un SVM

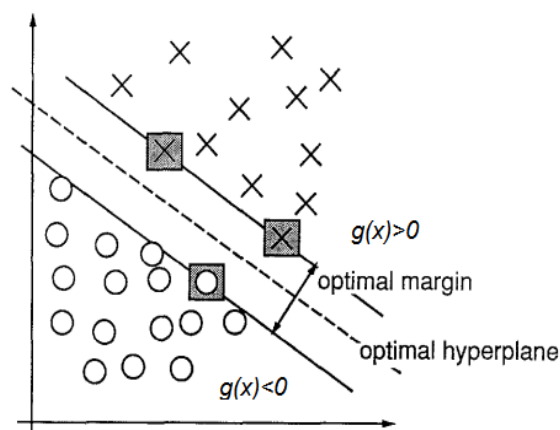


FIGURE 3.3 – Principe du classifieur SVM [19].

ou un classifieur SVM [19] est nativement défini pour la discrimination en deux classes, et dont le principe consiste à calculer un hyper-plan séparant les éléments des deux classes. Comme le nombre d'hyper-plans pourra être infini, le SVM calcule un plan optimal où la marge entre les échantillons des deux classes, d'une part et ce plan d'autre part, est maximale (voir figure 3.3). Le problème de la détermination de l'hyper-plan devient donc un problème d'optimisation quadratique. En fait, la marge est recherchée en utilisant les échantillons les plus proches de l'hyper-plan, et qu'on appelle les vecteurs supports.

Dans la figure 3.3, l'hyper-plan est représenté par la droite  $g(x) = 0$ , où on peut noter les marges maximales des deux côtés de la droite vers un certain nombre d'échantillons des deux classes (vecteurs supports).

### 3.4.3 Classifieur Bayésien Naïf

Le Bayésien naïf est basé sur le théorème de Bayes en probabilités. L'appellation 'naïf' provient du fait qu'on suppose que les variables aléatoires impliquées dans le calcul des probabilités sont indépendantes les unes des autres, et ce qui est rare dans les problèmes réels, ou du moins difficile à vérifier qu'elles le sont réellement [88]. Le classifieur bayésien est paramétrique dans le sens où son apprentissage consiste à déterminer les paramètres statistiques des différentes classes. Ces paramètres seront utilisés pour le calcul de certaines probabilités, nécessaires à la décision sur la classe à affecter à l'élément à classer par ce classifieur.

Si nous disposons des probabilités a priori des classes  $P(C_i)$ , et d'une fonction de densité de probabilités de ces classes qui nous permet de calculer des probabilités de

vraisemblance  $P(X/C_i)$ , il est possible, en utilisant la loi de Bayes, de calculer les probabilités à posteriori des classes  $P(C_i/X)$ , et dont la classe qui en résulte est celle avec le maximum des probabilités a posteriori (voir équation 3.1).

$$P(C_i/X) = \frac{P(X/C_i) \times P(C_i)}{P(X)} \quad (3.1)$$

Comme  $P(X)$ , dite 'évidence', est invariante par rapport à  $C_i$ , son calcul est inutile, et la classe qui en résulte ( $C^*$ ) pourra être exprimée comme suit :

$$C^* = \text{ArgMax} P(X/C_i) \times P(C_i) \quad (3.2)$$

Pour plusieurs problèmes, on suppose que la distribution de probabilités est normale, et par conséquent on utilise la loi normale pour calculer les probabilités de vraisemblance, et ce en considérant la densité  $f$  :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp -\frac{1}{2} \frac{(x - \mu)^2}{\sigma^2} \quad (3.3)$$

Dans ce cas, l'apprentissage consiste à calculer les paramètres statistiques des classes, à savoir les moyennes  $\mu_i$  et les écart-types  $\sigma_i$ .

### 3.4.4 Réseaux de neurones artificiels

Les neurones artificiels, ou dits formels, sont inspirés des neurones biologiques, et dont on a noté depuis des années 40 leur capacité de mémorisation, d'apprentissage et d'inférence. En effet, c'était en 1943 que Warren McCulloch et Walter Pitts ont présenté le premier neurone formel [76]. Et quelques années plus tard, en 1949, Donald Hebb a monté la capacité d'apprentissage du neurone formel, ouvrant ainsi la voie à un nouveau domaine de l'intelligence artificielle, dit le connexionnisme.

#### Neurone formel

Le neurone formel, inspiré du neurone biologique, pourra être modélisé par un automate, ayant des entrées et une sortie et doté d'une fonction de transfert ou d'activation. Cette dernière transfère les entrées en la sortie avec un calcul bien précis. Conventionnellement, un neurone formel additionne ses entrées après les avoir pondérées. Ensuite, il

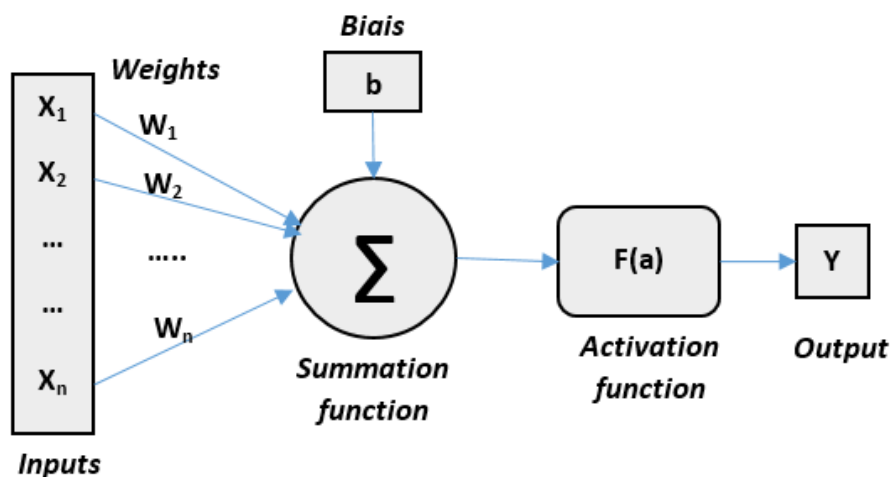


FIGURE 3.4 – Le neurone formel [20]

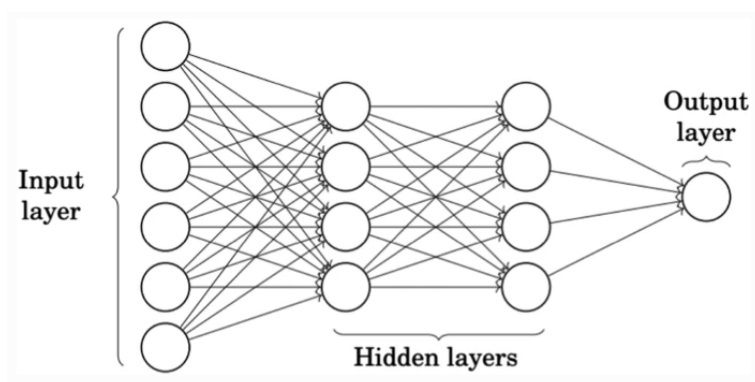


FIGURE 3.5 – Le perceptron multicouche [21]

active sa sortie si la somme pondérée est supérieure à un seuil donné (voir figure 3.4). Plus tard, en 1957, Frank Rosenblatt a mis en oeuvre une règle d'apprentissage permettant au neurone formel d'apprendre pour discriminer deux classes, et ainsi est né le premier perceptron.

### Perceptron multicouche

Nous présentons dans cette sous-section le perceptron multicouche (PMC) comme réseau de neurones de référence dans un contexte d'apprentissage supervisé. Selon notre point de vue, il s'agit du réseau le plus utilisé et qui a été le plus étudié dans la littérature du domaine. Le PMC est un réseau organisé en plusieurs couches de neurones. L'information se propage des couches d'entrée vers les couches de sortie (mode feedforward). Chaque couche est composée d'un nombre propre à elle de neurones. Un neurone en une couche donnée est connecté à tous les neurones de la couche suivante (voir figure 3.5).

L'ajustement des poids par la rétropropagation du gradient de l'erreur a permis au perceptron multicouche un apprentissage performant et qui lui a, à son tour, accordé de résoudre des problèmes non linéaires, et dont les réseaux précédents n'ont pas pu le faire. En effet, il a été montré dans plusieurs études, que le perceptron multicouche avec un nombre de couches des nombres de neurones dans les différentes couches, permet d'imiter toute fonction non linéaire à plusieurs variables, quelque soit sa complexité. Cependant, il faut disposer de données d'apprentissage en quantités suffisante pour pouvoir le faire.

Sachant les sorties correspondantes aux entrées dans un jeu d'apprentissage, l'erreur est calculée au niveau des neurones de sorties, puis propagée vers les couches cachées. Les erreurs sont ainsi calculées au fur et à mesure qu'on recule dans les couches, et par conséquent tout poids synaptique est corrigé en fonction du gradient de l'erreur. Ainsi les poids des neurones qui ont contribué plus à l'erreur vont voir leur valeur significativement corrigées. Un paramètre particulier, dit taux d'apprentissage permet de réguler la vitesse de correction dans le but que la configuration des poids ne tombera pas dans des minima locaux.

### 3.5 Classification ensembliste

La classification ensembliste consiste à combiner plusieurs classifieurs entraînés sur des sous ensembles d'apprentissage différents, et ce dans le but d'améliorer le pouvoir de prédiction du méta-classifieur qui en résulte, et ce en comparaison avec les classifieurs élémentaires, dits dans ce cas classifieurs faibles. L'idée sous-jacente de l'intérêt de la classification d'ensemble vient du fait, qu'un seul classifieur élémentaire, quelle que soit la force de son modèle, ne peut produire la même performance en modifiant le jeu d'apprentissage sur lequel il est entraîné. D'autre part, il est évident qu'on ne peut pas réduire les ensembles d'apprentissage pour les faire caler sur les modèles choisis. Devant ce dilemme, il est possible de faire entraîner plusieurs classifieurs faibles sur des parties distinctes des données d'apprentissage, et ensuite combiner ces classifieurs faibles pour produire les prédictions relatives aux données à classifier.

Statistiquement parlant, un méta-classifieur corrige deux types d'erreurs relatives à tout modèle d'apprentissage automatique : la première est l'erreur du "biais" avec laquelle un classifieur génère une moyenne de prédictions différente de la moyenne observée dans

les données de test. La seconde erreur est dite de la variance, indiquant le degré de sensibilité du classifieur aux données d'apprentissage : Pour un léger écart de données en entrée, le classifieur produit des prédictions différentes. Et c'est ce type d'erreur qui produit le phénomène de sur-apprentissage, car les classifieurs qui en résultent seront très dépendants des données d'apprentissage, et leur utilisation pour d'autres ensembles de test produit des résultats de classification médiocres.

### 3.5.1 Bagging

La technique de bagging est basée sur un re-échantillonnage des données d'apprentissage avec remise, et d'utiliser chaque sous-ensemble pour entraîner un classifieur faible [89]. La combinaison des résultats des différents classifieurs faibles se fait en général par vote majoritaire ou par moyennes quand il s'agit de grandeurs quantitatives dans un cadre de régression.

L'algorithme suivant schématise les étapes du bagging et montre comment les résultats des classifieurs faibles sont combinés.

---

**Algorithm 1:** Bagging

---

Input :

Training set S

Training Algorithm A

Number of samples T

Test data x

Output : Metaclassifier C\*

**for**  $t=1, \dots, T$  **do**

    |  $S_i$  = Sample from S with replacement

    |  $C_i$  = weak classifier Using algorithm A and sample  $S_i$

**end**

$C^*(x) = \operatorname{argmax}(\sum_{C_i(x)=y} 1)$

---

### 3.5.2 Boosting

Le boosting de classifieurs est basé sur le principe que les classifieurs faibles ayant des performances élevées sur les données d'apprentissage seront pondérés pour contribuer plus aux résultats du métaclassifieur comparés aux classifieurs faibles ayant obtenus des scores moindres. L'algorithme d'AdaBoost est d'une des implémentations les plus populaires. Il consiste en un ensemble de classifieurs, appelés faibles dans ce cas, qui améliorent de ma-

nière adaptative les performances du classifieur d'ensemble [90]. Les sorties de ce dernier sont combinées selon un ensemble de poids obtenus en fonction des données d'apprentissage, fournissant le classifieur AdaBoost final. L'adaptation dans un tel méta algorithme fait référence au fait que les classifieurs faibles sont ajustés par les sous ensembles de données d'apprentissage qui ont été mal classés lors des itérations précédentes. L'algorithme ci-dessous est une implémentation d'Adaboost avec un classifieur faible binaire  $h$ . Après l'apprentissage du classifieur d'ensemble à l'aide de l'ensemble d'apprentissage  $(x_1, y_1), \dots, (x_m, y_m)$  l'hypothèse complète du classifieur d'ensemble est exprimée par la fonction  $H(x)$ , où  $x$  est une instance du jeu de test à étiqueter  $y \in \{-1, +1\}$ .

---

**Algorithm 2:** Adaboost
 

---

**Result:**  $H(x)$

Given  $:(x_1, y_1), \dots, (x_m, y_m)$  where  $x_i \in X, y_i \in \{-1, +1\}$  ;

Initialisation  $:D_1(i) = \frac{1}{m}$  for  $i = 1, \dots, m$ ;

**for**  $t=1, \dots, T$  **do**

    Train weak learner using distribution  $D_t$ ;

    Get weak hypothesis  $h_t : X \rightarrow \{-1, +1\}$ ;

    Aim : select  $h_t$  with low weighted error :

$$\varepsilon = \Pr_{i \sim D_t}[h_t(x_i) \neq y_i];$$

    Choose  $\alpha = \frac{1}{2} \ln \left( \frac{\varepsilon}{1-\varepsilon} \right)$ ;

    Update, for  $i = 1, \dots, m$  :

$$D_{t+1}(i) = \frac{D_t(i) \exp(-\alpha_t y_i h_t(x_i))}{Z_t};$$

    where  $Z_t$  is a normalization factor (chosen so that  $D_{t+1}$  will be a distribution).

**end**

Output the final hypothesis :

$$H(x) = \text{sign} \left( \sum_{t=1}^T \alpha_t h_t(x) \right)$$


---

### 3.5.3 Random Forest

Les random-forests qu'on traduit dans la littérature en langue française par forêt d'arbres décisionnels ou forêts aléatoires tout court, est considéré comme des classifieurs ensemblistes, car ils permettent d'intégrer plusieurs classifieurs faibles, type arbre de décision, selon le principe du bagging (par rapport aux données d'apprentissage). Cependant, et contrairement au bagging et au boosting, les random-forests procèdent par une sélection aléatoire d'un sous ensemble d'attributs pour chacun des arbres impliqués dans l'ensemble [91]. À l'instar des autres méthodes ensemblistes les random-forests visent à

réduire la variance des prévisions, ce qui permet d'améliorer substantiellement les performances des classifieurs d'ensemble qui en résultent.

La figure 3.6 montre un exemple illustratif comment les données d'apprentissage sont aléatoirement sélectionnées au début, puis comment les résultats partiels obtenus au niveau des différents arbres sont intégrés par une méthode de fusion, typiquement la technique du vote majoritaire. Dans la figure, chaque sous arbre s'est limitée à un sous-ensemble des features.

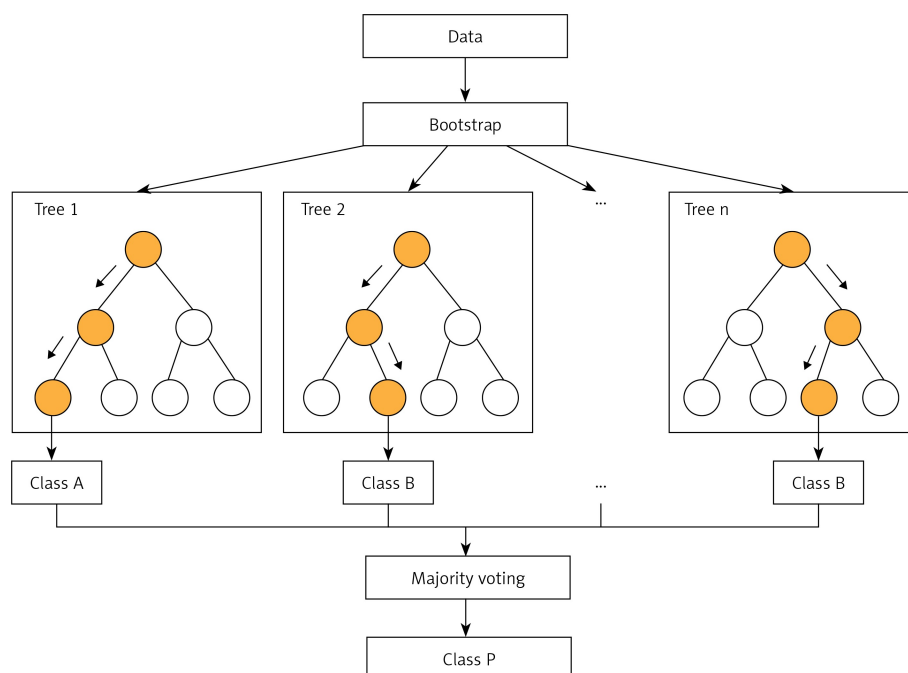


FIGURE 3.6 – Un exemple montrant comment la classification est opérée au sein d'un random-forest [22]

## 3.6 Apprentissage profond

L'apprentissage profond (Deep Learning) relève de l'apprentissage automatique, dont le principe commun de l'ensemble de ses techniques consiste à une représentation abstraite des données en entrées en utilisant une série de transformations non linéaires [92]. L'apprentissage profond a commencé avec ses premiers succès durant les années 2000. Cependant, il a connu un grand progrès durant la décennie suivante, avec des investissements importants de plusieurs leaders mondiaux en informatique, tels que Google, Apple, Microsoft, et autres, ainsi que plusieurs équipes de recherche universitaires qui se sont spécialisées dans le domaine. L'apprentissage profond est basé exclusivement sur les réseaux

de neurones profonds qui diffèrent des réseaux de neurones en général par leur nombre élevé de couches cachées, qui peut atteindre des dizaines ou des centaines de couches. Ainsi, et avec le progrès du domaine et des applications, différentes architectures-types de réseaux de neurones ont été développées et utilisées.

### 3.6.1 Réseaux de neurones convolutifs (CNN)

Il s'agit d'un type particulier de réseaux de neurones profonds, bien adapté à la reconnaissance et à la classification dans les images. Contrairement aux diverses méthodes d'apprentissage automatique, où on devrait définir manuellement ou semi-automatiquement les caractéristiques (features) à considérer, dans ce type de réseau, les caractéristiques sont extraites automatiquement des données d'apprentissage et sont présentées aux couches de classifications, qui sont les dernières couches du réseau [93, 94]. Typiquement, les réseaux de neurones convolutifs sont conçus pour la classification des images en fonction de leurs contenus visuels. Pour pouvoir réaliser ceci, avec des tailles élevées en pixels des images, on devrait d'une part extraire les bonnes caractéristiques pertinentes à la classification, et réduire autant que possible le nombre de ses caractéristiques sans perdre l'information pertinente, et ce pour pouvoir entraîner le réseau en un temps raisonnable.

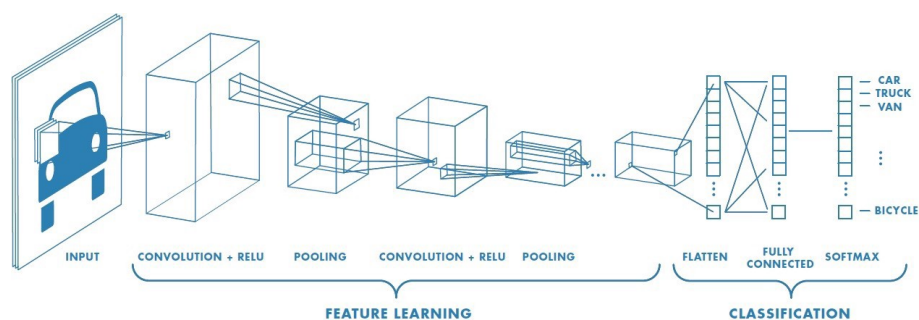


FIGURE 3.7 – Architecture type de CNN [23]

En considérant ces deux points, les réseaux de neurones convolutifs se composent de trois types de couches dont les objectifs sont différents : Les couches de convolution, les couches de Pooling, et les couches entièrement connectées (Fully connected). Le premier type de couche, à savoir les couches de convolution permettent de ressortir les caractéristiques spécifiques aux images à classifier, en découpant l'image en sous-régions, appelées tuiles, et l'analyser par un noyau de convolution (Figure 3.8). Les couches de Pooling quand à elles permettent de réduire la taille des cartes de caractéristiques et ce par échan-

tillonnage spatial. Cela réduit le temps de calcul et diminue le risque de sur-apprentissage (Figure 3.9). Et finalement, les couches fully connected permettent la classification des images selon les caractéristiques au nombre réduit produit par les couches précédentes. Cette partie peut être vue à elle seule comme un perceptron multicouches classique avec deux ou plus de couches cachées (Figure 3.7). La sortie d'un CNN est un vecteur de réels ou chaque élément  $i$  représente la probabilité que l'image en entrée est de type représenté par  $i$ .

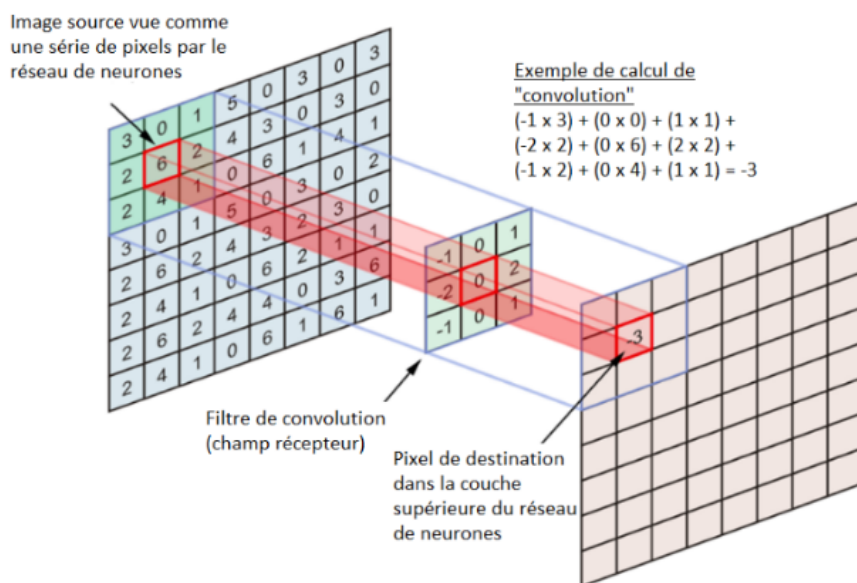


FIGURE 3.8 – Exemple d'opération de convolution [24].

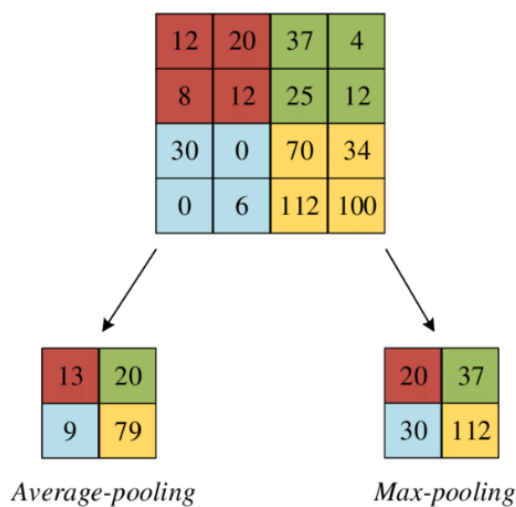


FIGURE 3.9 – Exemple d'opération de pooling [25].

On peut voir sur la figure 3.7 la disposition des différentes couches dans un CNN.

Typiquement, on réserve plusieurs couches de convolution pour couvrir les différentes caractéristiques qui peuvent être présentes dans l'image, dont chaque sortie est appliquée la fonction ReLU, qui permet d'écartier les valeurs négatives (Figure 3.10). L'image en sortie de chaque couche de convolution est réduite en utilisant une couche de Pooling. À la fin de cette succession de couches de convolution et de couches de Pooling, l'image réduite qui en résulte est linéarisée (Flaten) pour être présentée en entrée au sous-réseau composé de deux ou plus de couches totalement connectées. La fonction *softmax* appliquée aux probabilités obtenues dans le vecteur de sortie permet de sélectionner une seule classe, qui sera retenue comme la classe à laquelle appartient l'image ou le motif en entrée.

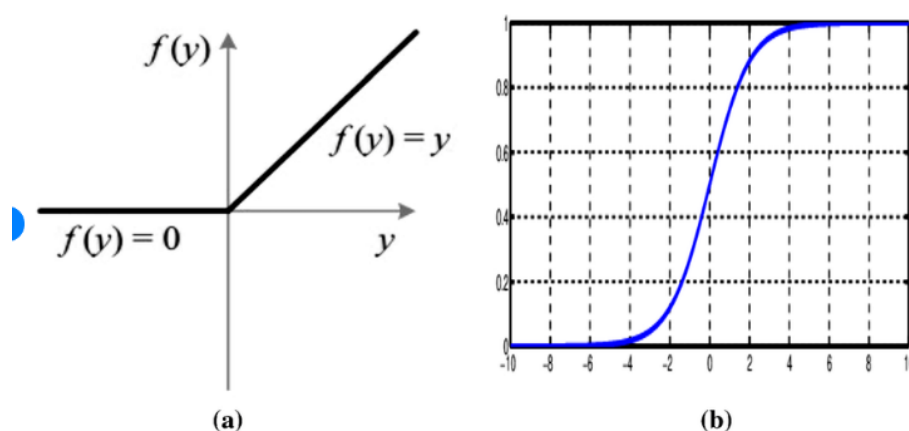


FIGURE 3.10 – Fonctions ReLU et SoftMax [26]

### Diverses architectures de CNN

La dernière décennie a vu la proposition de diverses architectures de réseaux CNN, où des appellations ont été données à ces architectures. Étant donné qu'elles sont au nombre de dizaines, nous présentons dans ce bref aperçu les architectures qui ont été jugées les plus représentatives et aussi les plus distinctes [95].

- LeNet : Il s'agit du réseau CNN le plus populaire par le fait qu'il est premier qui a été développé en 1998 [27], originalement pour la reconnaissance des chiffres manuscrits. On retrouve dans ce réseau les trois types de couches, dont la couche de Pooling, appelée couche de ré-échantillonnage (Figure 3.11).

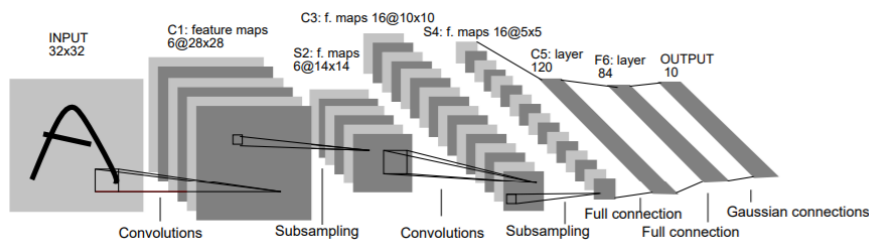


FIGURE 3.11 – Réseau LeNet [27]

- ResNet : Pour "Residual Network", il a introduit pour la première fois la notion d'omettre des connexions, ce qui permet la mise en oeuvre de réseaux CNN avec des dizaines voire des centaines de couches cachées [96]. Le Saut de certaines couches en l'établissement de connexions directes entre couches lointaines (Figure 3.12), permet d'éviter le problème d'effacement du gradient, qui influença négativement les performances des réseaux très profonds.

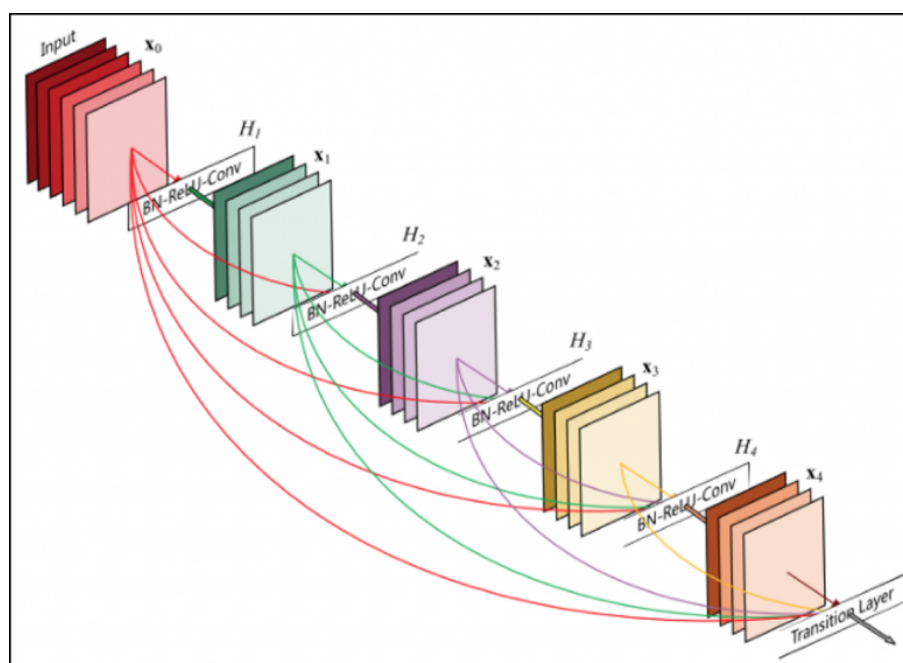


FIGURE 3.12 – Réseau ResNet [28]

- VGG16 : A travers ce réseau [97], il a été introduit le concept de réseaux denses en utilisant des filtres de tailles minimales soit pour la convolution soit pour le Pooling. La figure 3.13 montre l'architecture du réseau VGG16 originale, mise en oeuvre par ses développeurs Simonyan and Zisserman. Comme les filtres sont de petites tailles, leur nombre au sein d'une même couche est élevé, ce qui donne l'aspect dense à ce type de CNN.

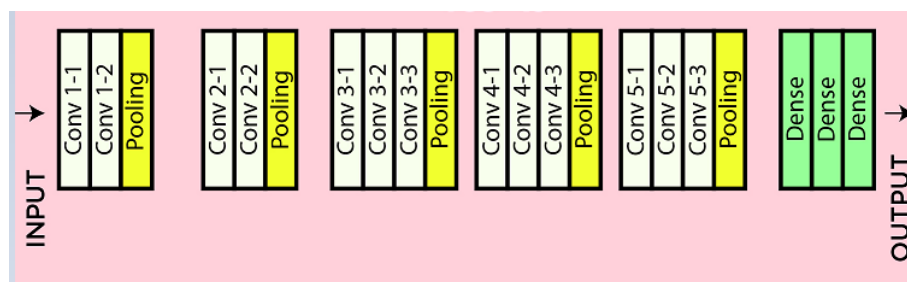


FIGURE 3.13 – Réseau VGG16 [29].

### 3.6.2 Réseaux de neurones récurrents (RNN)

Si les connexions dans réseaux de neurones sont généralement linéaires, résultant en un schéma "feed forward", les réseaux de neurones récurrents présentent des connexions récurrentes, dans le sens où des chemins cycliques (boucles) se créent au sein du réseau (figure 3.14).

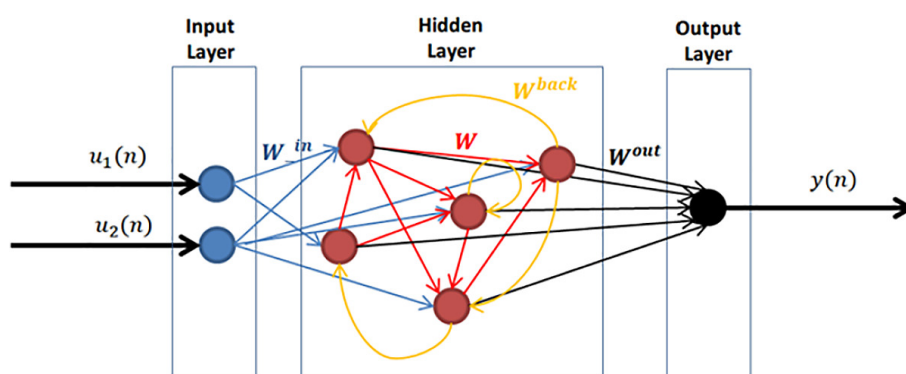


FIGURE 3.14 – Exemple de réseau de neurones récurrent (RNN) avec plusieurs connexions récurrentes dans les couches cachées [30]

Par leur capacité à représenter la récurrence, les réseaux de neurones sont bien adaptés aux applications où des série d'objets ou de phénomènes sont à traiter [98]. Typiquement il s'agit des séries temporelles, avec des applications en reconnaissance de la parole et la reconnaissance de l'écriture manuscrite. Ils sont également utilisés en traitement du langage naturel, avec toutes ses applications dont la traduction automatique [99].

Dans leur version de base, les réseaux de neurones récurrents sont heurtés au problème de disparition du gradient (Gradient Vanishing). Ce problème qui consiste à la diminution exponentielle des valeurs du gradients qui serviront à ajuster les poids des neurones. Pour un nombre élevé de récurrences le gradient tend vers zéro, et les poids ne peuvent donc plus être modifiés. Ce phénomène a conduit les chercheurs à améliorer le schéma de base

des réseaux de neurones récurrents, et ce en proposant différents modèles dont les plus réussis sont le réseau LSTM et le réseau GRU.

## Réseau LSTM

Les réseaux de neurones récurrents à mémoire court et long terme (Long Short Term Memory : LSTM) sont introduits dès 1997 pour résoudre le problème de la disparition du gradient [100, 101]. Actuellement, il s'agit du modèle de réseau récurrent le plus utilisé [102]. Dans les LSTM on utilise pour chaque unité computationnelle une cellule mémoire (voir figure 3.15), dont l'état se transfert (à gain constant) entre les époques. Ainsi le problème de la disparition du gradient est résolu.

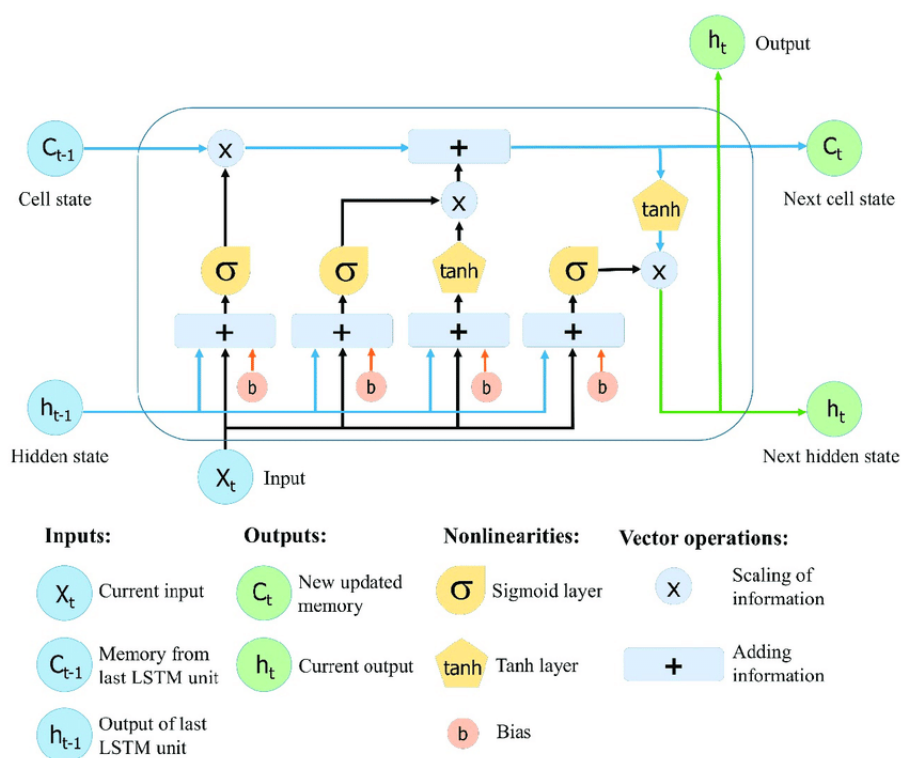


FIGURE 3.15 – Schéma général d'une unité computationnelle dans un LSTM [31]

## Réseau GRU

Une unité récurrente à portes (Gated Recurrent Unit : GRU) est à la base des réseaux GRU. Une unité GRU consiste en une extension du réseau LSTM, introduite en 2014 [103]. Il s'agit d'une simplification de modèle LSTM, avec peu de paramètres à apprendre. La simplification opérée consiste en la fusion de l'état caché et de l'état mémoire, aussi les portes d'entrée et les portes d'oubli sont fusionnées. La figure 3.16 montre comment l'unité GRU est

obtenue à partir de l'unité LSTM, les deux comparés à l'unité de base dans un RNN.

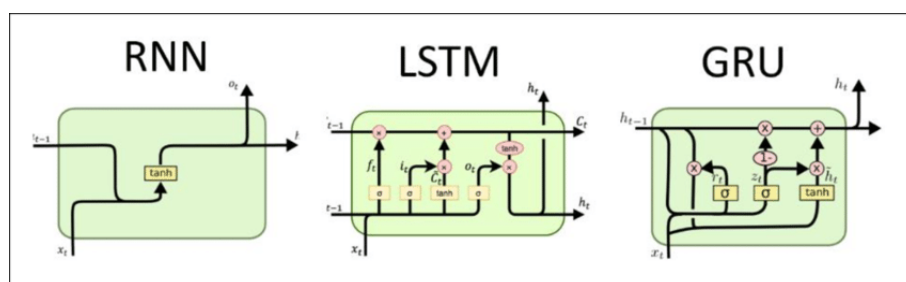


FIGURE 3.16 – Unité GRU comparée à l'unité LSTM et à l'unité de base d'un réseau RNN [32]

## Transformers

Il s'agit de modèles d'apprentissage profond, basés sur le concept d'attention qu'ils formalisent et utilisent initialement pour le traitement du langage naturel, notamment pour la traduction automatique [33]. Les modules d'attention dans les Transformers assurent des associations entre des ensembles et non pas entre des séquences comme pour les LSTM et les GRU. Ceci confère aux Transformers l'indifférence quant à l'ordre des entrées, et le plus important est la possibilité du traitement parallèle des unités d'entrée (voir figure 3.17).

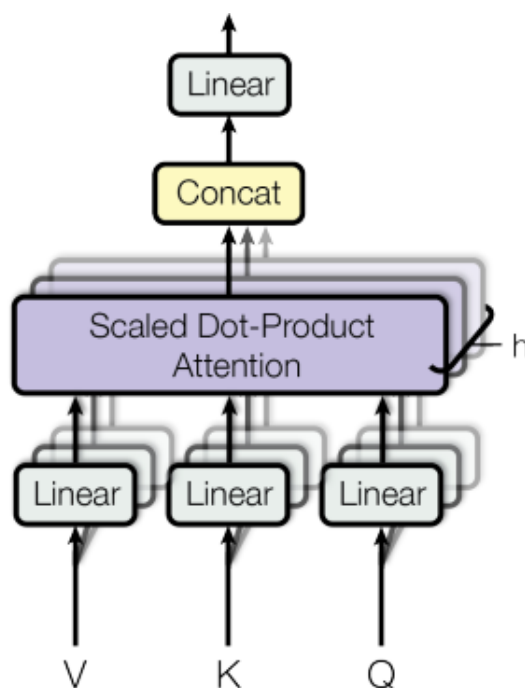


FIGURE 3.17 – Module d'attention d'un Transformer (Q : Requête, V : Valeur et K : clé) [33]

## 3.7 Conclusion

L'apprentissage automatique se situe au carrefour de plusieurs disciplines scientifiques, allant des statistiques et arrivant aux neurosciences et sciences cognitives. Comme ce domaine inclus l'ensemble des techniques que nous avons utilisées dans notre travail de thèse, nous lui ont consacré ce chapitre, en présentant tout d'abord ses origines avant de survoler à la fois ses techniques classiques et modernes. En, effet en plus des méthodes inspirées de l'analyse et la fouille des données, nous avons consacré une partie du chapitre aux techniques avancées de l'apprentissage automatique, et qui font l'essence des travaux de recherche durant la dernière décennie. Il s'agit des méthodes de classification ensemblistes d'une part, et les réseaux de neurones profonds d'autre part. Ces techniques et méthodes nous étaient d'une grande aide pour proposer de nouvelles méthodes pour la segmentation d'IRM et l'extraction de tumeurs cérébrales.

# Chapitre 4

Une approche ensembliste basée

Entropie pour la segmentation d'IRM

cérébrales

## 4.1 Introduction

Selon l'approche de classification basée pattern, non seulement les données sont utilisées pour produire des caractéristiques, mais également les patterns qui existent dans les données. Ces patterns sont également utilisés pour la classification des données et la définition de leurs attributs. Il a été indiqué dans plusieurs travaux [81] que la fouille de patterns peut aider à améliorer la classification des données, principalement pour les données structurées, telles qu'en reconnaissance d'objets et en analyse d'images. Par conséquent, pour plusieurs applications dans divers domaines, les patterns peuvent être définis comme des séquences de graphes dans les données brutes [104, 105]. Ainsi, définir des modèles de classification basés pattern permet sans aucun doute d'améliorer la classification, plutôt que d'utiliser des modèles basés uniquement sur des données brutes. De plus, combiner les classifieurs et les entraîner avec une bonne partition des données d'apprentissage a permis d'améliorer les résultats globaux de classification. Aussi, tous les travaux ayant utilisé les classifieurs d'ensemble ont donné de meilleurs résultats que ceux utilisant des classifieurs simples et ce quel que soit le classifieur élémentaire faible qui a été utilisé comme élément de construction de l'ensemble. Il a été indiqué dans plusieurs travaux de la littérature que la combinaison de classifieurs selon diverses méthodes d'ensemble permet d'éviter plusieurs problèmes classiques dans la classification des données. Le problème le plus connu est le problème de surapprentissage, constaté principalement lorsque le volume de données d'apprentissage est trop important. De plus, la combinaison de classifieurs permet de traiter le problème de la convergence précoce de l'algorithme lorsque la classification est basée sur l'optimisation comme dans les méthodes neuronales [106, 107].

Dans ce chapitre, nous définissons une caractéristique basée pattern qui sera testée

pour la classification des tissus cérébraux dans les IRM. Un ensemble de classifieurs faibles (Support Vector Machine (SVM) ou Naïve Bayes) sont combinés selon l'algorithme de boosting à entraîner et à utiliser pour étiqueter les voxels du volume IRM. Néanmoins, et avant d'entrer dans le vif de notre approche, nous présentons tout d'abord le concept de l'entropie, notamment l'entropie de l'information, et dont nous nous sommes servis pour représenter nos patterns.

## 4.2 Entropie de l'information

L'entropie est un concept introduit en physique statistique pour exprimer l'ordre de désorganisation au sein d'un système [108]. La désorganisation est associée également à l'imprédictibilité, et donc l'entropie indique à quel point les états d'un système sont imprédictibles. Selon le 2<sup>ième</sup> principe de la thermodynamique de Sadi Carnot [109], l'entropie est une fonction de la masse et du volume de la matière dont un système est composé. En 1865, Rudolf Clausius a montré que le rapport de la quantité de chaleur  $Q$  reçue par un système, à sa température  $T$  est toujours inférieur à une fonction d'état qu'il a appelée entropie, et a notée  $S$  [110]. Ensuite, et d'après la thermodynamique statistique, il a été établi que plus l'entropie d'un système est élevée plus il y a du désordre dans ce système. Ludwig Boltzmann l'a exprimé en fonction du nombre d'état microscopiques au sein d'un système (nombre de configuration)  $\Omega$ , par l'expression suivante :

$$S = k_B \ln \Omega \quad (4.1)$$

où  $k_B$  est la constante de Boltzmann ( $k_B = 1.38110^{-23} J.K^{-1}$ ).

Il a été également établi qu'un système thermodynamique isolé, évolue vers un état d'équilibre où les probabilités de ces micro-états deviennent égales, et son entropie devient ainsi maximale.

A l'origine et selon la théorie de l'information, l'entropie de Shannon, introduite par Claud Shannon en 1948 [111], mesure la quantité d'information émise par une source de données. Ceci est exprimée dans le sens où à chaque fois que la source émet des informations différentes, la quantité d'information est plus importante et par conséquent l'entropie également. Formellement, on exprime l'entropie de l'information en supposant que la source est représentée par une variable aléatoire discrète  $X$  à  $n$  symboles (moda-

lités)  $x_1, \dots, x_n$ . En supposant que chaque symbole  $x_i$  a une probabilité d'apparition  $P_i$ , l'entropie d'information de la source  $X$  est exprimée par la formule suivante :

$$H(X) = - \sum_{i=1}^n P_i \text{Log} P_i \quad (4.2)$$

Historiquement, Claud Shannon a pu, en utilisant son concept d'entropie de l'information, calculer la quantité d'information qui pourra être transmise par un canal de communication. Les applications de l'entropie de Shannon sont nombreuses, et cette entropie est à la base de plusieurs méthodes computationnelles en traitements d'images [112]. Il s'agit principalement du recalage d'images [113, 114], en segmentation d'images [115], et en évaluation de la qualité d'images [116].

### 4.3 Classification basée entropie pour la segmentation structurelle en IRM

La diversité est un atout en traitement de données en général et en particulier pour la classification de ces dernières. Cependant, il est difficile de capturer et de représenter cette diversité pour en bénéficier après. Pour en tirer profit, nous avons mené un travail dont le but est d'exploiter la diversité des données IRM cérébrales pour la segmentation des tissus qui composent le cerveau. Dans notre cas, la diversité est bi-dimensionnelle, dans le sens où nous utilisons un ensemble de classifieurs selon l'algorithme 'Adaboost', ce qui représente une diversité de traitement. D'autre part, nous utilisons l'entropie spatiale de l'information IRM pour capturer et représenter la diversité des données. Pour nous, la seconde diversité est matérialisée par la multitude de patterns locaux qui peuvent exister dans une IRM. La prise en compte de ces patterns permet de bien étiqueter les voxels, car en addition des intensités des voxels, l'interaction entre ses derniers est prise en compte. Le boosting est bénéfique pour la segmentation des données IRM car il permet d'éviter le problème de surapprentissage en répartissant les données d'apprentissage sur plusieurs classifieurs faibles. De plus, l'algorithme Adaboost pour lequel nous avons opté dans ce travail, permet une distribution adaptative des données qui est bien appropriée pour les données IRM, étant donné que le modèle de distribution des données IRM n'est pas connu à l'avance. Dans un premier temps, l'image IRM est prétraitée à l'aide d'un

algorithme d'extraction de l'encéphale de la boîte crânienne (élimination du crane), à savoir FSL Brain Extraction Tool (BET) [117, 118], pour éliminer les tissus non cérébraux. Cependant, nous avons préféré éviter le filtrage du bruit, étant donné que, premièrement, les données IRM sont généralement altérées sur les frontières entre les différents tissus par l'effet du volume partiel, où une moyenne des intensités à ces voxels aggrave cet effet. Deuxièmement, parce que le pattern proposé permet la réduction de l'effet du bruit sur la classification des voxels même sans lissage des données IRM. Une telle caractéristique de la méthode proposée s'explique par le fait que le pattern basé sur l'entropie défini autour d'un voxel donné exprime les interactions au sein de l'ensemble des voxels appartenant au voisinage du voxel en question [119, 120].

Avant de présenter en détail la manière dont notre méthode procède pour étiqueter les voxels, nous présentons d'abord comment l'algorithme "Adaboost" traite les problèmes de classification à plus de deux classes, car nous nous intéressons à trois classes de tissus en addition d'une classe de fond (tissus non cérébraux). En effet, pour  $M$  classes,  $M - 1$  classifieurs faibles binaires sont utilisés pour substituer un classifieur faible de  $M$  classes. Il existe deux méthodes bien connues qui permettent de combiner des classifieurs binaires pour construire un classifieur multi-classes. La première méthode procède selon le principe *winner takes all* [121], où le  $k^{\text{ième}}$  classifieur permet de distinguer la classe  $k$  de toutes les autres classes. Pour la deuxième méthode, un classifieur est dédié à chaque couple de classes [122], où son rôle est de traiter une paire de deux classes parmi l'ensemble de toutes les classes. Nous constatons que les deux méthodes sont implémentées dans la plupart des plateformes d'apprentissage automatique.

### 4.3.1 Les données IRM

Le volume IRM obtenu, après l'élimination du crâne, est un ensemble de voxels pouvant chacun appartenir à l'un des trois tissus restants, à savoir le liquide céphalo-rachidien (CSF : Cerebro Spinal Fluid), la matière grise (GM : Gray Matter) et la matière blanche (WM : White Matter). Chacun d'eux est caractérisé par son intensité moyenne et son écart-type correspondant  $(\mu_c, \sigma_c)$ ,  $c \in \{CSF, GM, WM\}$ . Nous supposons également que la distribution d'intensité dans chaque tissu est gaussienne (voir la formule 4.3).

$$f_c(x_i, \mu_c, \sigma_c) = \frac{1}{\sigma_c \sqrt{2\pi}} e^{-\frac{1}{2}(x_i - \mu_c)^2 / 2\sigma_c^2} \quad (4.3)$$

où  $x_i$  est l'intensité du voxel  $i$ .

Sur la base de ces distributions gaussiennes, plusieurs auteurs ont utilisé le modèle de mixture gaussienne pour proposer des méthodes de segmentation d'IRM [53]. La figure 4.1 montre les distributions conjointes des trois tissus cérébraux à savoir : le liquide céphalo-rachidien (CSF), la matière grise (GM), et la matière blanche (WM).

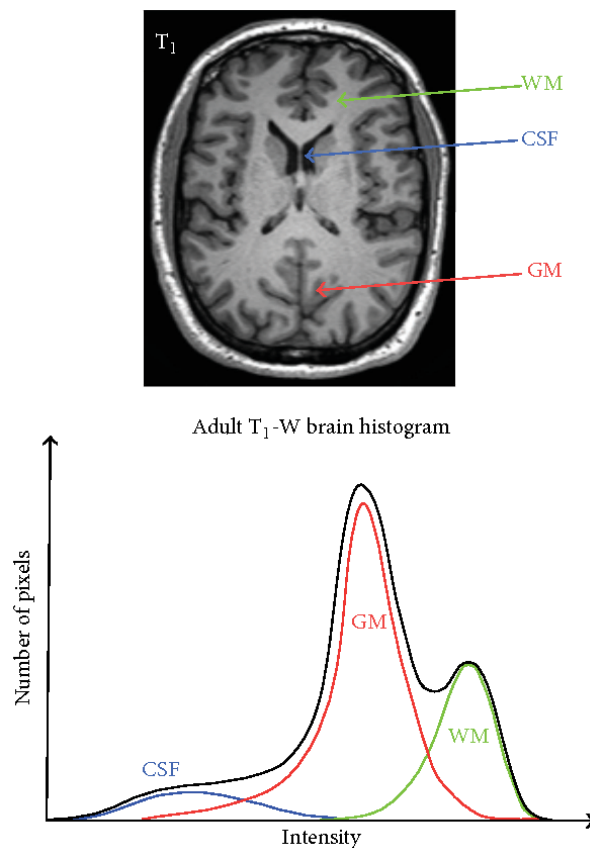
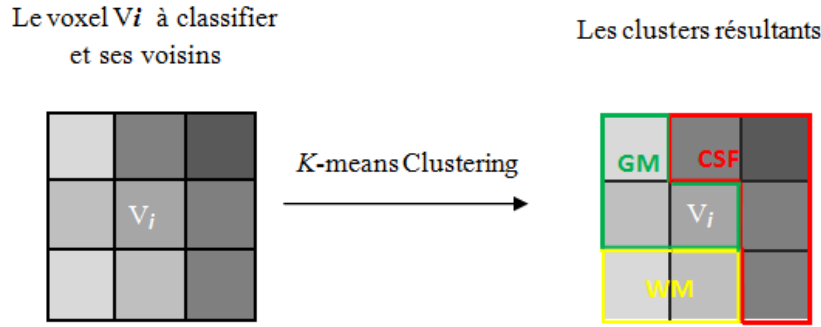


FIGURE 4.1 – Modèle de mixture gaussienne en IRM cérébrale [34]

### 4.3.2 Classification basée entropie

Notre modèle proposé basé sur l'entropie, vise à capturer les interactions entre les voxels appartenant à un voisinage local. De telles interactions peuvent être représentées selon une fonction d'énergie. Ainsi, le modèle proposé pour un voxel donné  $i$  dans le volume IRM est un vecteur de trois composantes, où chacune représente l'entropie spatiale des intensités des voxels similaires dans le voisinage. De tels sous ensembles de voxels similaires sont obtenus par l'algorithme  $k$ -means (voir figure 4.2), appliqué dans le voisinage du voxel avec trois classes (CSF,GM,WM) (voir équation 4.4).

FIGURE 4.2 – L'application du  $k$ -means sur l'ensemble de voisinage

$$E_c = - \sum_{D_c} P_i \times \log_2 P_i \quad (4.4)$$

où  $D_c$  désigne l'ensemble des voxels appartenant à la classe  $c$ , et  $P_i$  est la probabilité que le voxel appartienne à la classe  $c$  (voir équation 4.5) :

$$P_i = \frac{\frac{1}{\sigma_c \sqrt{2\pi}} e^{\frac{1}{2}(x_i - \mu_c)^2 / 2\sigma_c^2}}{\sum_c \frac{1}{\sigma_c \sqrt{2\pi}} e^{\frac{1}{2}(x_i - \mu_c)^2 / 2\sigma_c^2}} \quad (4.5)$$

$\mu_c$ ,  $\sigma_c$  sont respectivement la moyenne et l'écart-type des intensités des voxels appartenant à la classe  $c$  et situés au voisinage du voxel considéré ( $i$ ). Ainsi, un clustering par l'algorithme  $k$ -means est effectué dans le voisinage du voxel, donc les trois sous-ensembles de voxels et leurs couples respectifs de  $(\mu_c, \sigma_c)$ ,  $c \in \{CSF, GM, WM\}$  sont obtenus. Nous remarquons que pour les IRM d'entraînement nous n'avons pas besoin de clustering dans le voisinage des voxels car nous disposons de la réalité terrain qui fournit les voxels de chaque cluster, ce qui permet de calculer les trois entropies spatiales sans clustering des voxels.

### 4.3.3 L'AdaBoost basé entropie pour la segmentation IRM

Nous avons considéré une implémentation multiclasse d'Adaboost (voir l'algorithme 3),  $hcls$  est le classifieur faible à entraîner avec l'ensemble d'apprentissage  $(pattern_1, y_1), \dots, (pattern_m, y_m)$ , le vecteur de caractéristiques constituant le pattern est  $[x_i, E_{1i}, E_{2i}, E_{3i}]$  où  $x_i$  est l'intensité du voxel  $i$  et  $E_{1i}, E_{2i}, E_{3i}$  sont les entropies spatiales calculés à partir du voisinage du voxel  $i$ .

**Algorithm 3:** Adaboost basé entropie

---

Given :  $(Pattern_1, y_1), \dots, (Pattern_m, y_m)$  where  
 $Pattern_i \in Patterns, y_i \in \{CSF, GM, WM\}$  ;  
**Result:**  $H(pattern)$   
 $K = 3$  // 3 classes  
weights initialisation  
 $w_i = \frac{1}{m}; i = 1..m$   
**for**  $t=1, \dots, T$  **do**  
    Train a weak classifier  $hcls^t$  using wights  $\{w_i\}$   
     $err^t = \sum_{i=1}^m w_i \cdot \mathcal{I}(label_i \neq hcls^t(pattern_i))$   
     $\alpha^t = \log \frac{1-err^t}{err^t} + \log(K-1)$   
     $w_i = w_i \cdot \exp(\alpha^t \cdot \mathcal{I}(label_i \neq hcls^t(pattern_i))), i = 1..m$   
     $w_i = w_i / \sum_{j=1}^m w_j, i = 1..m$   
**end**  
 $H(pattern) = \operatorname{argmax}_k \sum_{t=1}^T \alpha^t \cdot \mathcal{I}(hcls^t(pattern) = k)$

---

Comme on peut le remarquer sur la Figure 4.3, le vecteur de caractéristiques qui sera utilisé pour la classification par le SVM ou le Naïve Bayes est composé, en plus de l'intensité du voxel  $x_i$ , des trois caractéristiques d'entropies spatiales  $E_1, E_2$  et  $E_3$ , obtenues en fonction du clustering de l'ensemble des voxels formant le voisinage local. Un tel pattern capture bien les interactions des voxels et exprime bien les contraintes spatiales qui existent dans les données IRM. Les entropies  $E_1, E_2$ , et  $E_3$  permettent de distinguer les cas où le voxel est au voisinage d'une bordure d'un tissus ou non. De plus, elles permettent de distinguer si un voxel est affecté par un fort écart dû au bruit. Évidemment, l'intensité du voxel en question est prise en compte pour la classification, de sorte que la classe résultante est probablement celle du tissu dont l'intensité moyenne est la plus proche, mais ajustée si nécessaire par l'interaction avec les voxels dans le voisinage, exprimée par leurs entropies spatiales respectives.

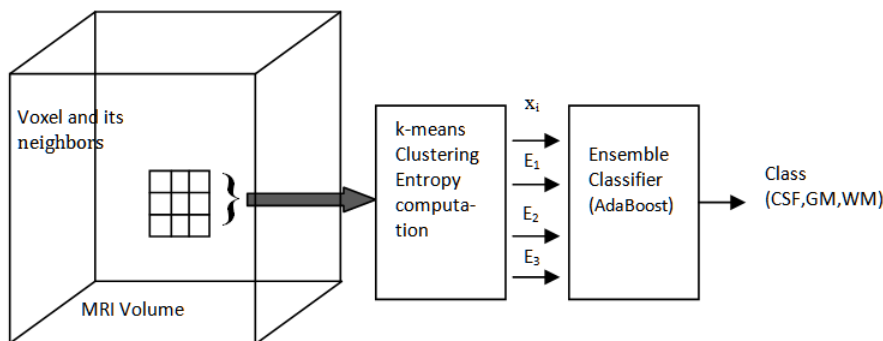


FIGURE 4.3 – Principe de la classification proposée basée entropie

## 4.4 Expérimentation et évaluation

L'expérimentation du pattern proposé a été réalisée à l'aide de volumes IRM de la base de données bien connue "Brain web" [42]. Cette base de données fournit un large ensemble de volumes d'IRM avec leur étiquetage de réalité terrain, permettant aux auteurs d'utiliser des méthodes basées sur l'apprentissage automatique supervisé et d'évaluer quantitativement les méthodes proposées. De plus, les IRM peuvent être obtenues selon différents niveaux d'artéfacts, à savoir le bruit et l'INU (Non uniformité des intensités). Toutes les IRM ont une taille de  $181 \times 217 \times 181$  voxels. Dans ce travail, nous nous sommes contentés uniquement des IRM selon la modalité T1. Comme dans la plupart des travaux de la littérature du domaine, nous avons utilisé 70% des données pour l'entraînement et 30% pour les tests. Il est à noter également que lors de nos expérimentations, nous avons considéré des tranches (slices) 2D au lieu de travailler sur le volume en entier (3D). En effet, nous avons constaté lors de toute notre expérimentation que le voisinage 2D d'un voxel donné permet amplement de l'étiqueter correctement en utilisant le pattern proposé (intensité et entropies).

### 4.4.1 Sélection des paramètres

Un seul paramètre est à définir pour la méthode proposée. Ça consiste au nombre d'itérations de l'algorithme Adaboost, qui correspond aussi au nombre de classifieurs faibles qui seront combinés pour classer les voxels d'IRM. Ce paramètre est ajusté quantitativement en testant un ensemble de valeurs, puis en choisissant celle à partir de laquelle il n'y a plus d'amélioration significative de la précision de la classification, exprimée par l'indice de Dice (voir sous section "Évaluation des performances"). Enfin, nous voulons souligner que la taille du voisinage pour le calcul de l'entropie est fixée à  $3 \times 3$ .

### 4.4.2 Évaluation des performances

Deux indices principaux sont généralement calculés pour évaluer et comparer les méthodes de segmentation basées sur la classification et le clustering. Il s'agit notamment des indices de Jaccard et de Dice. Basé sur les instances d'étiquetage "vrai positif" ( $TP$ ), "vrai négatif" ( $TN$ ) et "faux positif" ( $FP$ ), le coefficient de Jaccard est exprimé comme suit :

$$Jaccard = \frac{TP}{TP + TN + FP} \quad (4.6)$$

Le coefficient de Dice est exprimé de sa part par :

$$Dice = \frac{2TP}{2TP + TN + FP} \quad (4.7)$$

Nous avons opté pour l'indice de Dice pour deux principales raisons : la première est qu'il est le plus cité dans la littérature et la seconde est qu'il a été considéré par des travaux avec lesquels nous comparons nos résultats.

### 4.4.3 Résultats expérimentaux

Comme il a été cité ci-dessus, nous avons expérimenté les fonctionnalités proposées en utilisant deux classifieurs faibles différents, à savoir le classifieur SVM et le classifieur Naïve Bayes. En effet, nous avons commencé notre expérimentation en testant plusieurs modèles de classifieurs et nous avons remarqué que les classifieurs retenus donnaient de meilleurs scores. Ainsi, les résultats introduits correspondent à l'utilisation de deux ensembles de classifieurs faibles, combinés chacun selon l'algorithme AdaBoost. Nous avons considéré deux classifieurs différents, et nous avons comparé nos résultats à ceux de travaux antérieurs, et ce afin de montrer que l'amélioration des résultats n'est pas due au classifieur lui-même, mais aux caractéristiques proposées, à savoir les entropies spatiales.

Tout d'abord, nous introduisons dans les figures 4.4, 4.5 et 4.6 respectivement l'IRM obtenue avec un niveau de bruit de 3% et un niveau d'INU de 20%, les trois tissus cérébraux obtenus en utilisant le SVM comme classificateur faible, et les trois tissus cérébraux en utilisant le Naïve Bayes comme classificateur faible. On peut constater visuellement, en comparant avec l'image IRM, que les tissus obtenus sont bien délimités et correspondent bien à la réalité terrain.

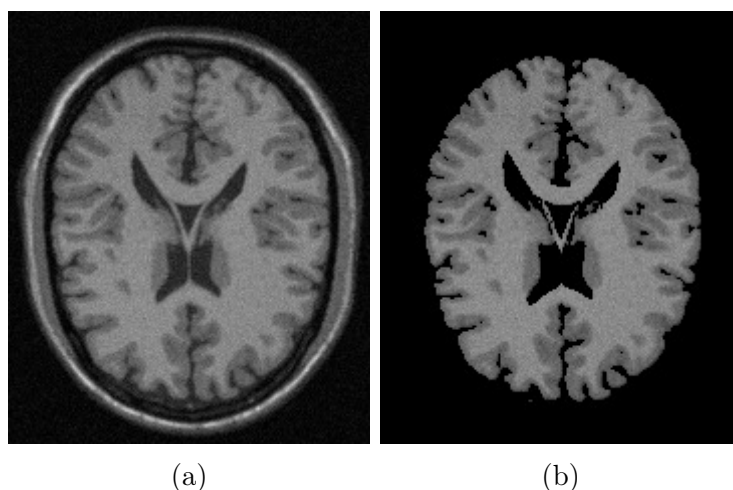


FIGURE 4.4 – Un exemple d'IRM de la base de données "Brain Web". (a) IRM brute avec un niveau de bruit de 3% et un niveau d'INU de 20%, (b) Cerveau extrait par "Brain Extraction Tool" de FSL.

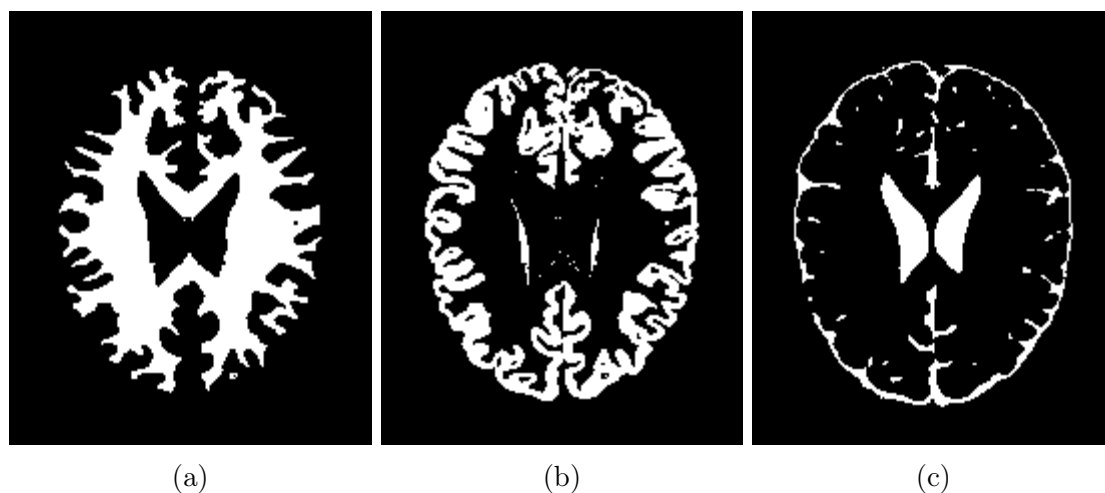


FIGURE 4.5 – Résultats de la segmentation de l'IRM avec niveau de bruit défini sur 3% et INU défini sur 20%. (a), (b), (c) sont respectivement la matière blanche, la matière grise et le CSF pour le cas où le classifieur faible est un SVM.

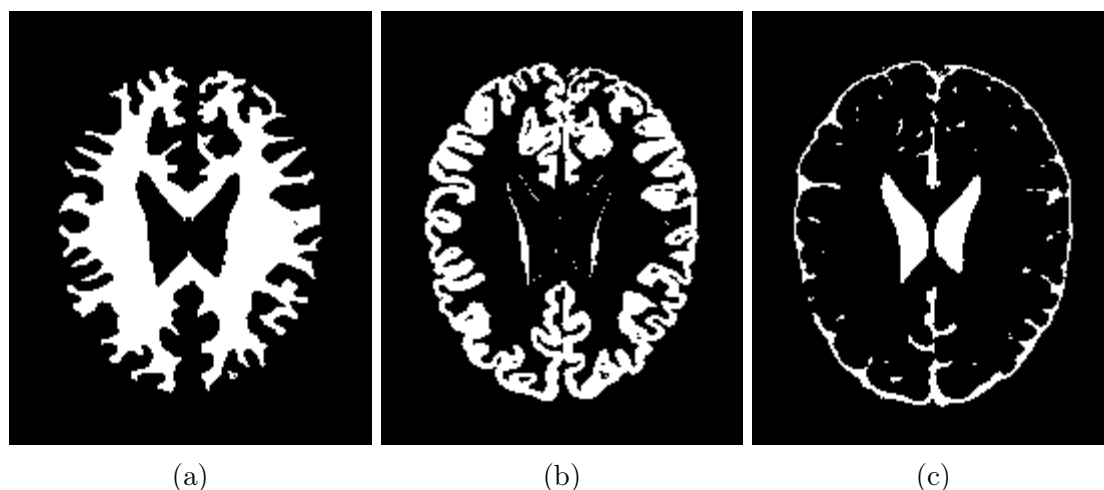


FIGURE 4.6 – Résultats de la segmentation de l'IRM avec niveau de bruit défini sur 3% et INU défini sur 20%. (a), (b), (c) sont respectivement la matière blanche, la matière grise et le CSF pour le cas où le classifieur faible est un Naïve Bayes.

Le tableau 4.1 montre les résultats de la segmentation selon l'indice de Dice en utilisant le classifieur SVM comme classifieur faible avec l'intensité du voxel et ses entropies spatiales comme caractéristiques de classification.

TABLE 4.1 – Résultats de segmentation selon l'indice de Dice pour les différentes IRM et les différents tissus cérébraux (WM, GM et CSF). Les caractéristiques de classification par le SVM sont l'intensité du voxel et ses entropies spatiales.

N INU	WM				GM			
	1	3	5	7	1	3	5	7
0%	99.17	98.10	96.60	94.08	98.23	96.21	93.53	89.25
20%	98.40	97.70	96.26	94.30	96.86	95.51	93.05	89.47
40%	97.02	96.55	94.95	92.75	94.59	93.67	90.80	87.02
60%	93.58	91.79	89.48	87.3	86.37	83.78	79.41	73.97
90%	90.40	89.08	87.42	85.78	83.05	80.17	76.66	71.44
N INU	CSF							
	1	3	5	7				
0%	98.46	97.09	95.60	93.63				
20%	97.78	96.85	95.59	93.73				
40%	96.92	96.29	94.69	92.99				
60%	87.94	87.07	85.32	81.04				
90%	87.89	86.32	84.04	80.38				

Le tableau 4.2 montre les résultats de la segmentation selon l'indice de Dice en

utilisant un Naïve Bayes comme classifieur faible avec l'intensité du voxel et ses entropies spatiales comme caractéristiques de classification.

TABLE 4.2 – Résultats de segmentation selon l'indice de Dice pour les différentes IRM et les différentes matières cérébrales (WM, GM et CSF). L'Adaboost avec le classifieur faible Naïve Bayes utilisant les entropies spatiales du voxel et son intensité comme caractéristiques de classification.

	WM				GM			
N INU	1	3	5	7	1	3	5	7
0%	98.90	97.53	95.17	91.48	97.41	94.88	90.30	82.45
20%	98.08	97.14	94.32	91.55	95.94	94.15	88.98	82.57
40%	96.19	94.52	92.30	89.67	92.56	89.54	84.63	78.68
60%	90.09	88.58	87.08	86.24	77.73	73.94	69.88	68.08
90%	86.18	85.70	85.24	84.89	67.67	66.91	65.69	65.31
	CSF							
N INU	1	3	5	7				
0%	97.36	96.05	94.18	91.96				
20%	96.72	95.73	94.45	92.15				
40%	95.39	94.78	92.95	91.72				
60%	86.29	85.47	83.97	80.68				
90%	83.97	84.05	82.16	79.96				

Selon les résultats présentés dans les tableaux 4.1 et 4.2, nous remarquons les scores élevés de la segmentation lorsque l'entropie spatiale est utilisée comme caractéristique de classification. Cette amélioration s'explique par la capture des interactions entre les voxels voisins. Ces interactions s'expriment grâce à une énergie, formulée comme une entropie spatiale. Un voxel n'est pas classé uniquement en fonction de sa valeur (niveau d'intensité) mais en fonction de la force de ses interactions avec ses voxels voisins.

Selon les résultats présentés, nous pouvons remarquer que la méthode a bien étiqueté les voxels, et qu'elle présente une forte robustesse contre le bruit et l'INU, où l'indice de Dice n'a pas considérablement chuté avec des niveaux élevés de ces deux artefacts. Par exemple, pour la matière blanche, la variation de l'indice de Dice est de 98,40 à 96,26 avec le classifieur SVM pour une variation du bruit de 1% à 5%, avec INU fixée à 20%. On peut remarquer la même robustesse vis-à-vis de l'INU. L'écart est de 99,17 à 93,58 pour l'INU variant de 0% à 60% avec un niveau de bruit fixé à 1%. Avec le classifieur

Naïve Bayes, la méthode obtient également de bons résultats et présente également une bonne robustesse contre le bruit et l'INU. Pour le CSF, la variation de l'indice de Dice est de 97,78 à 95,59 avec le classifieur SVM pour une variation de bruit de 1% à 5%, avec INU fixée à 20%. Aussi, la même robustesse est constatée face à l'INU. La variation est de 98,46 à 96,92 pour l'INU variant de 0% à 40% avec un niveau de bruit fixé à 1%.

Les figures 4.7 et 4.8 présente les variations de l'indice Dice en fonction du nombre d'itérations de l'algorithme AdaBoost, respectivement dans les deux cas : en utilisant le SVM, ou en utilisant le Naïve Bayes comme classifieur faible. On peut remarquer que l'indice de Dice croît uniformément avec le classifieur SVM pour les trois tissus, et devient stable à partir de la 9<sup>ième</sup> itération. Il en est de même pour le classifieur Naïve Bayes, cependant ce dernier devient stable un peu plus tôt que celui du SVM, à savoir à partir de la 8<sup>ième</sup> itération. Selon les deux cas, on peut conclure que l'algorithme Adaboost ne nécessite pas beaucoup d'itérations (moins de 10 itérations) pour être entraîné. Un si faible nombre d'itérations permet de conclure que la classification des voxels selon la méthode proposée est rapide et envisageable pour être utilisée par les praticiens de la santé.

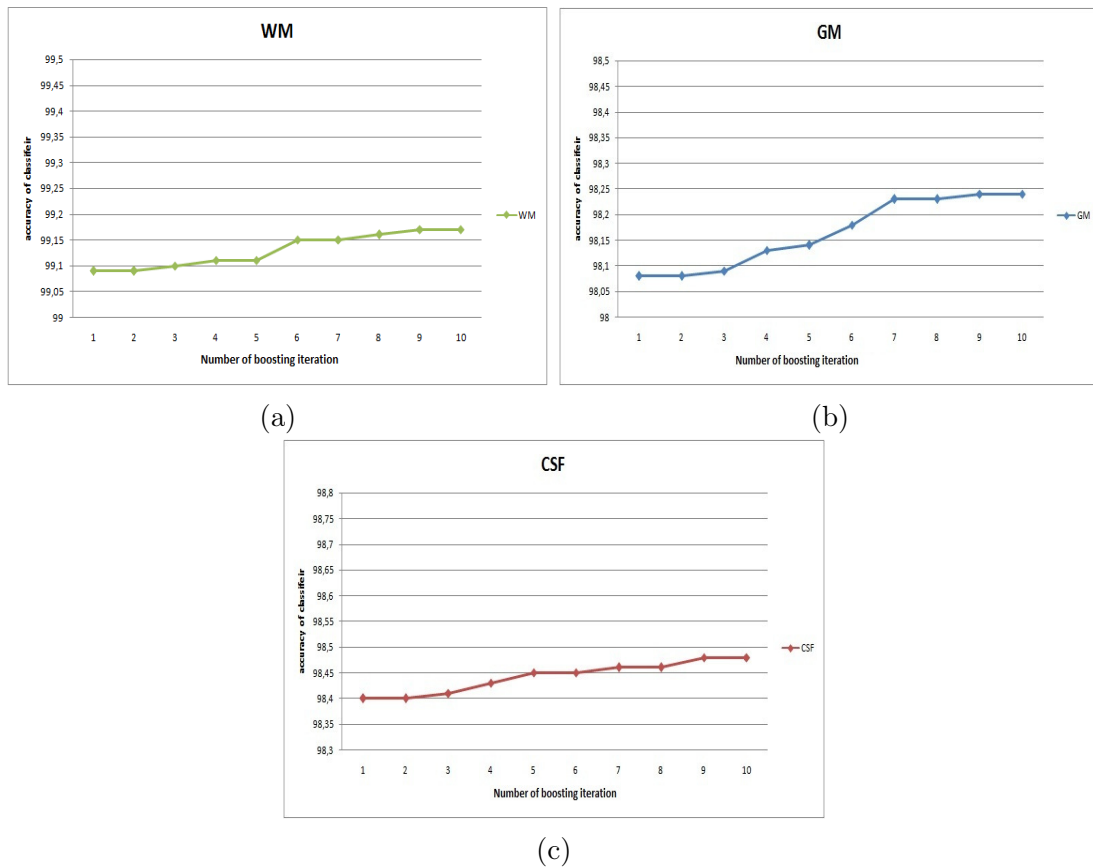


FIGURE 4.7 – Variation de la précision de la classification basée sur le classifieur faible SVM en fonction du nombre d'itérations d'Adaboost.(a) :WM,(b) :GM,(c) :CSF.

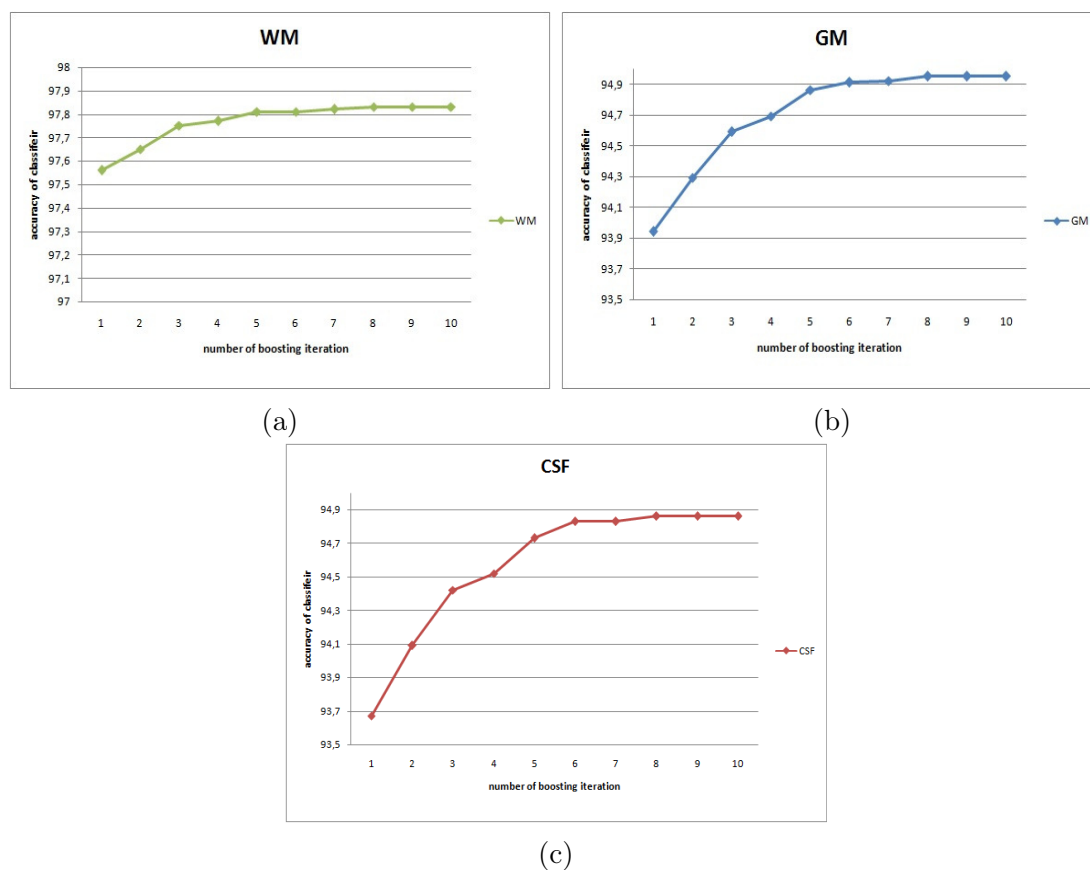


FIGURE 4.8 – Variation de la précision de la classification basée sur le classifieur faible Naïve Bayes en fonction du nombre d’itérations d’Adaboost.(a) :WM,(b) :GM,(c) :CSF.

Afin de montrer l’efficacité des caractéristiques proposées, nous introduisons dans le tableau 4.3 une comparaison entre les résultats de classification des données IRM obtenus avec SVM comme classifieur faible et ceux de certains travaux bien cités dans la littérature. Nous avons considéré l’IRM avec 20% d’INU et différents niveaux de bruit, et nous avons comparé les résultats pour les tissus MW et GM.

TABLE 4.3 – Comparaison des résultats selon l’indice de Dice avec certains travaux de la littérature.

Méthode	WM				GM			
	1	3	5	7	1	3	5	7
Fast%	97	95	94	92	96	94	91	91
SMP5%	94	94	90	86	93	92	90	87
NL-FCM%	94	91	90	83	94	93	90	87
Boosted Spatial Entropy (with SVM)%	98	98	96	94	97	96	93	89

Nous pouvons remarquer dans le tableau précédent que notre méthode obtient des

résultats presque meilleurs que toutes les méthodes impliquées dans la comparaison. Ils sont plus proches de Fast [123], considéré comme la meilleure, par rapport aux autres [58, 124]. De tels résultats confirment que les caractéristiques proposées lorsqu'elles sont utilisées avec un classificateur d'ensemble, comme l'Adaboost, sont bien appropriées pour représenter des patterns locaux dans les données IRM, en particulier lorsque ces dernières sont à segmenter.

#### 4.4.4 Analyse des résultats et discussion

Selon les différents résultats présentés ci-dessous, nous remarquons la forte amélioration des résultats de segmentation lorsque l'entropie spatiale est utilisée comme caractéristique de classification en plus des intensités des voxels. L'utilisation de l'entropie spatiale du voxel avec son intensité a permis une forte augmentation des valeurs de l'indice de Dice, même avec des niveaux élevés de bruit et d'INU. Ceci peut s'expliquer par le fait que l'entropie spatiale proposée, outre sa capacité à considérer le voxel et son voisinage, et comme elle est de nature énergétique, exprime la force d'interaction entre les voxels dans les IRM. Ainsi, un voxel n'est pas classé uniquement en fonction de son intensité, mais aussi selon sa force d'interaction avec les voxels dans son voisinage. La prise en compte de l'entropie spatiale a montré une meilleure robustesse vis-à-vis du bruit et de l'INU, notamment lorsque les niveaux de ces artefacts sont élevés. Un tel résultat peut s'expliquer par la prise en compte du champ d'interaction du voxel et de ne pas se limiter à la juste valeur d'intensité de ce dernier.

## 4.5 Conclusion

Dans ce chapitre, nous avons introduit un nouvel ensemble de caractéristiques des données IRM basées patterns et comment ces caractéristiques ont été utilisées avec un classifieur d'ensemble pour la classification des voxels. Les caractéristiques introduites consistent, en plus de l'intensité du voxel, des entropies spatiales des intensités de voxels similaires dans le voisinage du voxel. De telles entropies sont intéressantes pour capturer l'interaction entre voxels voisins, afin qu'elles permettent de mieux classer ces derniers. En effet, étiqueter un voxel en IRM ne dépend pas seulement de son intensité mais aussi de son interaction avec son voisinage. Avec l'algorithme AdaBoost, deux classifieurs

faibles ont été testés, à savoir le SVM et le Naïve Byes. L'expérimentation de la méthode, en faisant varier les différents niveaux d'artéfacts, a montré une forte amélioration des résultats de classification dans les deux cas. De plus l'algorithme Adaboost converge en moins de 10 itérations, ce qui rend l'étape d'apprentissage trop rapide. Dans les travaux futurs, les caractéristiques proposées peuvent être testées avec différents classifieurs, et une optimisation supplémentaire pourrait être appliquée à l'algorithme AdaBoost, en tenant compte des a priori sur les données d'IRM cérébrales.

# Chapitre 5

Un modèle profond basé pattern pour  
la segmentation des tumeurs cérébrales

## 5.1 Introduction

Une détection et une segmentation précises des tumeurs cérébrales sont nécessaires au succès d'un traitement dont dépend l'état futur du patient et ses capacités mentales et physiques. La détection tardive des tumeurs cérébrales complique l'acte chirurgical et conduit souvent à perdre plusieurs capacités du patient. De plus, la classification basée pattern et les réseaux de neurones convolutifs (CNNs) ont tous deux suscité plus d'intérêt au cours de la dernière décennie. L'approche basée pattern pour la classification de données procède par des ensembles d'attributs capturant des configurations ou des interactions au sein des données. Dans cette optique, plusieurs auteurs ont souligné l'efficacité d'une classification de données quand des patterns sont définis et utilisés[125]. Typiquement un pattern pourra être défini comme des séquences de configurations bien distinguées dans les données [104, 105]. Ces configurations permettent de définir des ensembles d'attributs dont l'utilisation pour la classification améliore nettement les résultats. Par ailleurs et récemment, les réseaux de neurones, et en particulier les réseaux de neurones convolutifs (CNNs pour "Convolutional Neural Networks") ont été largement utilisés pour l'analyse et la compréhension d'images, y compris la segmentation d'images et la reconnaissance d'objets [126, 68, 127], et ont produit les meilleurs résultats par rapport aux méthodes classiques. En outre, tous les travaux ayant utilisé des classifieurs d'apprentissage d'ensemble ont donné de meilleurs résultats que ceux utilisant des classifieurs simples, et ce indépendamment des classifieurs élémentaires faibles qui ont été utilisés comme éléments de construction de l'ensemble. La classification d'ensemble selon différents méta-algorithmes d'agrégation de classifieurs permet de faire bénéficier à l'ensemble des classifieurs faibles de la diversité au sein des données, mais aussi sans tomber dans le problème de sur apprentissage, notamment si les volumes de données d'appren-

tissage sont trop importants. Il est aussi à noter que les classifieurs d'ensemble, et par leur prise en charge efficace de la diversité des données, permettent d'éviter le problème de la convergence précoce des algorithmes d'apprentissage quand les modèles utilisés sont basés sur la recherche d'un optimum global d'une fonction objectif, comme pour le cas de réseaux de neurones [106, 107] et des modèles markoviens [64].

Dans ce chapitre, nous définissons un CNN et un schéma de classification ensembliste basé pattern qui seront testés pour la classification des tissus tumoraux dans les IRMs pathologiques. Un ensemble de classifieurs faibles CNNs sont combinés selon l'algorithme de boosting à entraîner et à utiliser pour étiqueter les voxels du volume IRM, supposés être affectés d'une tumeur cérébrale. Cependant, et avant de présenter les détails de la méthode proposée, nous introduisons une courte littérature sur la segmentation d'images avec prise en charge des patterns qui y sont dedans.

## 5.2 Segmentation d'image avec prise en charge des patterns

Nous rappelons d'abord que l'ultime objectif d'une segmentation d'image est d'en extraire les objets qui peuvent y figurer. Ces objets peuvent être géométriques, bien délimités, comme ils peuvent être des régions d'intérêt dans les images traitées. La principale raison de l'échec des algorithmes de segmentations type "éclatement-fusion (Split and merge)" sont la grande variabilité des formes, des tailles, des orientations et des intensités et couleurs qui caractérise les objets dans une même image [35].

Les techniques de reconnaissance de patterns sont utilisées dans les opérations d'éclatement et de fusion des régions dans le but d'identifier les objets qui figurent dans les images. Selon les auteurs dans [35], la stabilité des régions est utilisée pour classifier une région, soit appartenant à un objet, soit au fond de l'image, soit mixte. Ainsi, les régions mixtes sont récursivement éclatées jusqu'à atteindre une taille minimale de  $16 \times 16$  pixels. Les petites régions obtenues (homogènes) sont ensuite appariées aux patterns dans l'objectif de bien placer les contours des objets de l'image (voir figure 5.1). Lors des fusions des régions, les formes de ces dernières sont prises en comptes. Les régions ne sont fusionnées que si les formes des contours qui les délimitent sont compatibles, et que le résultat de fusion satisfait le critère de stabilité. Les résultats expérimentaux sont concluants selon

une comparaison avec l'algorithme du FCM.

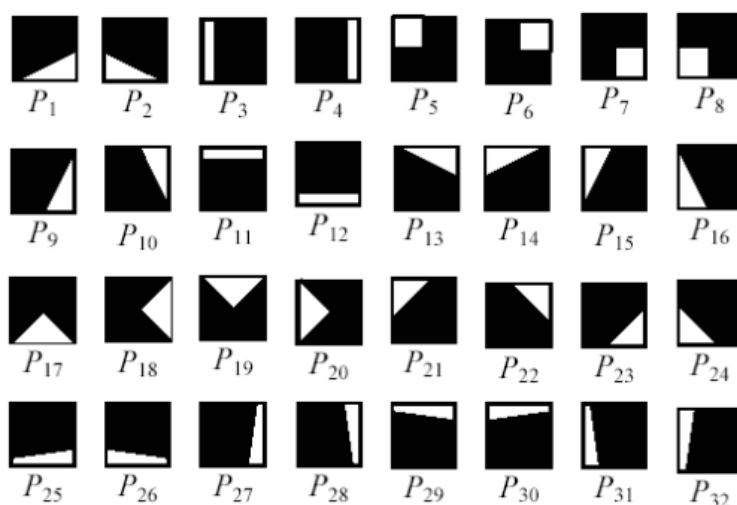


FIGURE 5.1 – Les 32 block de patterns de base utilisés dans [35]

Partant de la limitation de l'algorithme introduit dans [35] (PSM : Pattern Split-and-Merge), à savoir son incapacité à éclater des régions d'objets connectés, les auteurs dans [36] ont proposé une extension de l'algorithme de base, qu'il ont baptisé (OSP : Object Segmentation using block based Patterns) pour que ce dernier puisse séparer ce type de régions. Ils se sont basés sur la stabilité des régions (définie dans [35]), sur l'intra et l'inter-variabilité des régions et sur la perception visuelle. Ils ont pu améliorer le pouvoir de séparation de l'algorithme de base (voir figure 5.2).

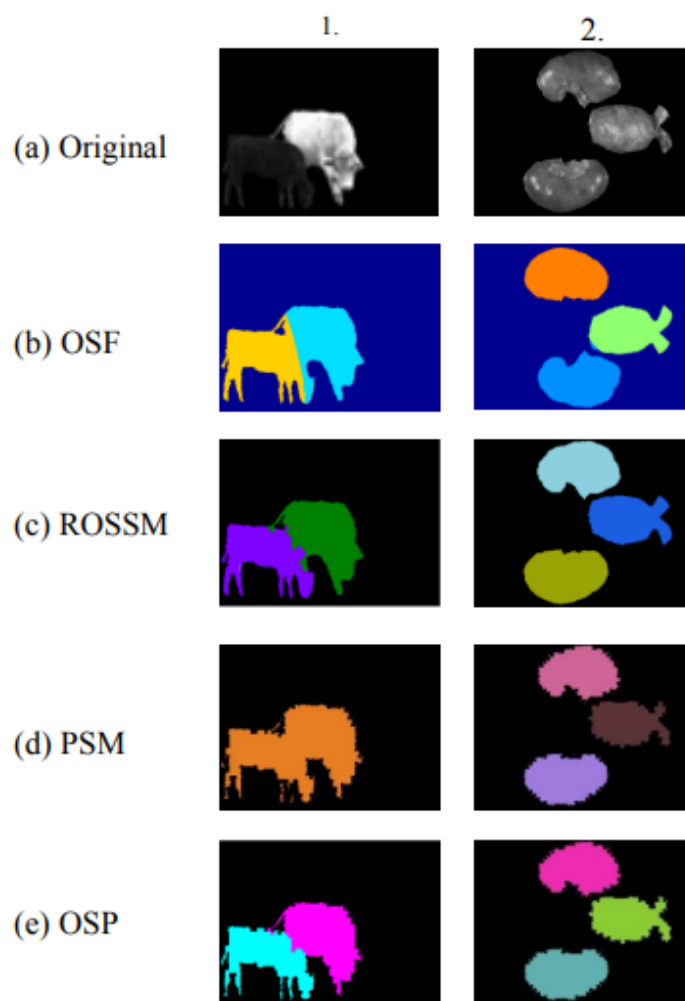


FIGURE 5.2 – Amélioration introduite par OSP en comparaison avec l’algorithme de base : PSM [36]

### 5.3 Segmentation profonde basée pattern de tumeurs cérébrales

Comme nous l’avons mentionné plus haut et au chapitre précédent, le boosting de classifieurs permet de renforcer les résultats de classification et d’éviter certains problèmes classiques dont principalement le surapprentissage. Ce problème est bien constaté dans les données IRM, car elles sont de volumes très importants de l’ordre de plusieurs mégaoctets par image. Dans ce travail, comme pour le précédent, nous avons choisi l’algorithme Adaboost, et ce pour son adaptabilité de distribution des données d’apprentissage sur les classifieurs faibles de sorte à équilibrer leur performance et d’en assurer un résultat global meilleur. Habituellement, une tâche de prétraitement est effectuée, cependant, nous avons

préférée ne pas appliquer un filtrage de bruit pour ne pas aggraver le problème de l'effet du volume partiel qui consiste en un flou de l'IRM qui complique sa segmentation. Cependant la prise en compte d'un voisinage large, en addition de l'utilisation d'un CNN comme classifieur de base, permettra de surmonter l'effet du bruit, mais sans que ce dernier ne soit filtré. Nous commençons la présente section par la présentation des données d'IRM tumorales, et les classes pour lesquelles nous nous intéressons. Ensuite, nous présentons une implémentation de l'algorithme Adaboost, adaptée à notre problème de classification en multi-classes [128].

### 5.3.1 Les données d'IRM tumorales

Une IRM cérébrale est un volume de voxels appartenant chacun à une partie anatomique du cerveau. Les données IRM sont obtenues selon l'une des quatre techniques de pondération, à savoir T1, Tc1, T2 et FLAIR (voir Figure. 5.3). Dans les IRM cérébrales montrant des tumeurs, les tissus d'intérêt sont respectivement le noyau (Co pour "Core") de la tumeur, sa partie active (Ac pour "Active") et l'œdème (Oe pour "Oedema") qui l'entoure la tumeur, en plus des tissus sains (He pour "Healthy") composant la matière cérébrale. Afin de réduire le grand nombre de paramètres nécessaires aux réseaux de neurones profonds dont les CNNs, nous avons procédé, lors de l'apprentissage et de la prédiction, au traitement des IRM, tranche par tranche. Ainsi, une tranche d'IRM est une image 2D de pixels, où chacun est représenté par une valeur scalaire comprise entre [0..255]. Nous avons procédé ainsi pour deux raisons : 1) en 3D, les CNNs devient complexes en apprentissage et en prédiction, 2) via des tests, nous avons constaté que le passage de la 2D à la 3D n'améliore pas significativement les résultats de segmentation, cependant les temps d'apprentissage et de tests ont été grandement réduits.

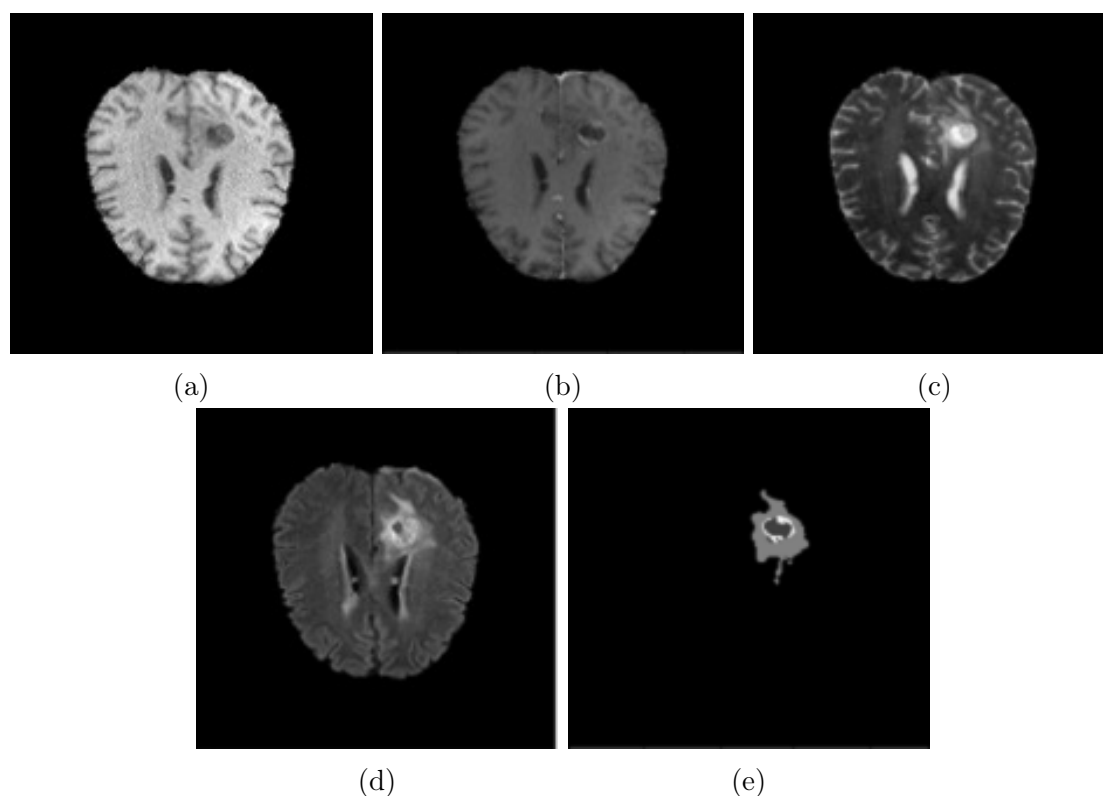


FIGURE 5.3 – Un échantillon d’IRM de la base de données BraTS : (a) :modalités T1, (b) :T1c, (c) :T2, (d) :Flair et (e) :segmentation de réalité terrain de la tumeur.

### 5.3.2 L’AdaBoost basé pattern pour la classification des données IRM

Pour notre problème, nous avons considéré une implémentation multi-classes d’AdaBoost, introduite par l’algorithme ci-dessous. Nous utilisons un classificateur faible CNN multi-classes *hcnn*. Après l’apprentissage du classifieur d’ensemble à l’aide de l’ensemble d’apprentissage  $(pattern_1, label_1), \dots, (pattern_m, label_m)$ , l’hypothèse complète du classificateur d’ensemble est exprimée par la fonction  $HCNN(pattern)$ , où  $pattern$  est une instance de données, représentée par la sous-image entourant le voxel à étiqueter (voir Figure 5.4), et  $label \in \{Co, Ac, Oe, He\}$ .

Dans notre cas, les patterns ne sont pas explicitement représentés comme pour les peu de travaux de la littérature ayant traité de la segmentation d’image en définissant des patterns au sein de l’image. Selon notre méthode, un voxel est mieux classifié si on considère le pattern présent dans une sous-image minimale quelle le contient (voir Figure 5.4). Par ailleurs, et pour rester fidèle au paradigme de l’apprentissage profond, qui s’en passe de la représentation des caractéristiques des images, nous avons supposé l’existence

**Algorithm 4:** CNNAdaboost

---

Inputs :  $(pattern_1, label_1), \dots, (pattern_m, label_m)$  where  
 $pattern_i \in \mathbf{Patterns}, label_i \in \{Co, Ac, Oe, He\}$  ;  
**Result:**  $H(pattern)$   
 $K = 4$  // 4 classes  
weights initialisation  
 $w_i = \frac{1}{m}$  ;  $i = 1..m$   
**for**  $t=1, \dots, T$  **do**  
    Train a weak classifier  $hcnn^t$  using wights  $\{w_i\}$   
     $err^t = \sum_{i=1}^m w_i \cdot \mathcal{I}(label_i \neq hcnn^t(pattern_i))$   
     $\alpha^t = \log \frac{1-err^t}{err^t} + \log(K - 1)$   
     $w_i = w_i \cdot \exp(\alpha^t \cdot \mathcal{I}(label_i \neq hcnn^t(pattern_i)))$ ,  $i = 1..m$   
     $w_i = w_i / \sum_{j=1}^m w_j$ ,  $i = 1..m$   
**end**  
 $H(pattern) = \operatorname{argmax}_k \sum_{t=1}^T \alpha^t \cdot \mathcal{I}(hcnn^t(pattern) = k)$

---

d'un pattern autour d'un voxel sans le rechercher, de sorte que ce pattern sera pris en compte lors de l'apprentissage du CNN (dit léger) pour mieux classifier le voxel au centre de la sous-image.

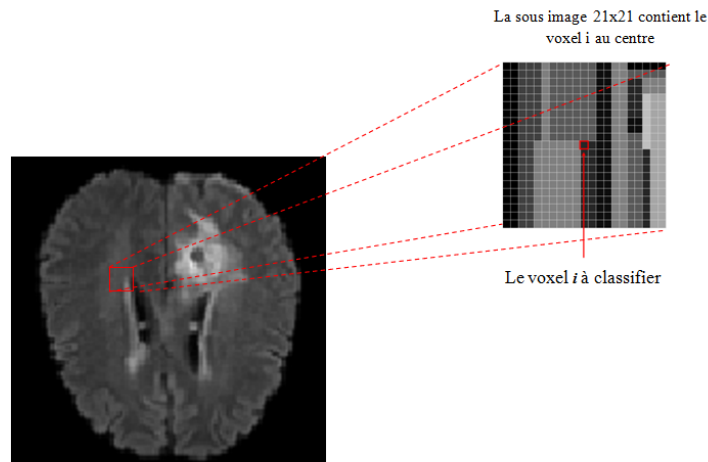


FIGURE 5.4 – La sous-image en entré du CNN léger

Il a été soulevé dans plusieurs travaux de la littérature de segmentation d'image médicales [41, 53] que les taux d'échecs d'attribution des pixels/voxels à leurs classes respectives se produit au niveau des frontières entre les tissus définissant les structures anatomiques. Quelque soit la puissance des méthodes proposées pour la classification de ce type de pixels/voxels, la considération du pixel/voxel en question ou son voisinage seuls n'a pas permis de bien les affecter correctement à leur classes. Nous supposons que l'élargissement du voisinage permet de prendre en compte le pattern local au sein de quoi

se situe le pixel/voxel, et de ce fait, le mieux étiqueter.

D'autre part et pour tirer profit de la diversité des données, en termes d'intensités et de patterns, nous avons opter pour une classification d'ensemble. Ce paradigme de classification, via ses familles d'algorithmes, tels que le boosting et le bagging, a contribué à améliorer significativement les résultats de classification des classifieurs conventionnels, dits dans ce contexte, classifieurs faibles. Partant de ce constat, et vu la grande diversité, en intensités et en formes, qui caractérisent les images médicales en général, et les IRM en particulier, nous avons opté pour ce paradigme, et en considérant un voisinage large du voxel, et ce pour pouvoir prendre en compte le pattern qui y se trouve.

### **5.3.3 Le CNN léger proposé pour l'étiquetage des voxels dans les IRM**

Afin d'entraîner, puis d'utiliser les CNNs pour étiqueter les voxels dans une IRM pathologique, nous nous concentrons sur chaque voxel dans une tranche, et nous considérons que le pattern formé par le voxel en question et les voxels qui forment son large voisinage sur la tranche, permet de classifier précisément le voxel (Figure 5.5). De plus, nous supposons qu'un ensemble de classifieurs basé sur un tel CNN léger permet d'améliorer la précision de la classification des voxels, étant donné que les éventuelles erreurs qui se produisent avec les classifieurs individuels sont corrigées de manière adaptative par l'ensemble.

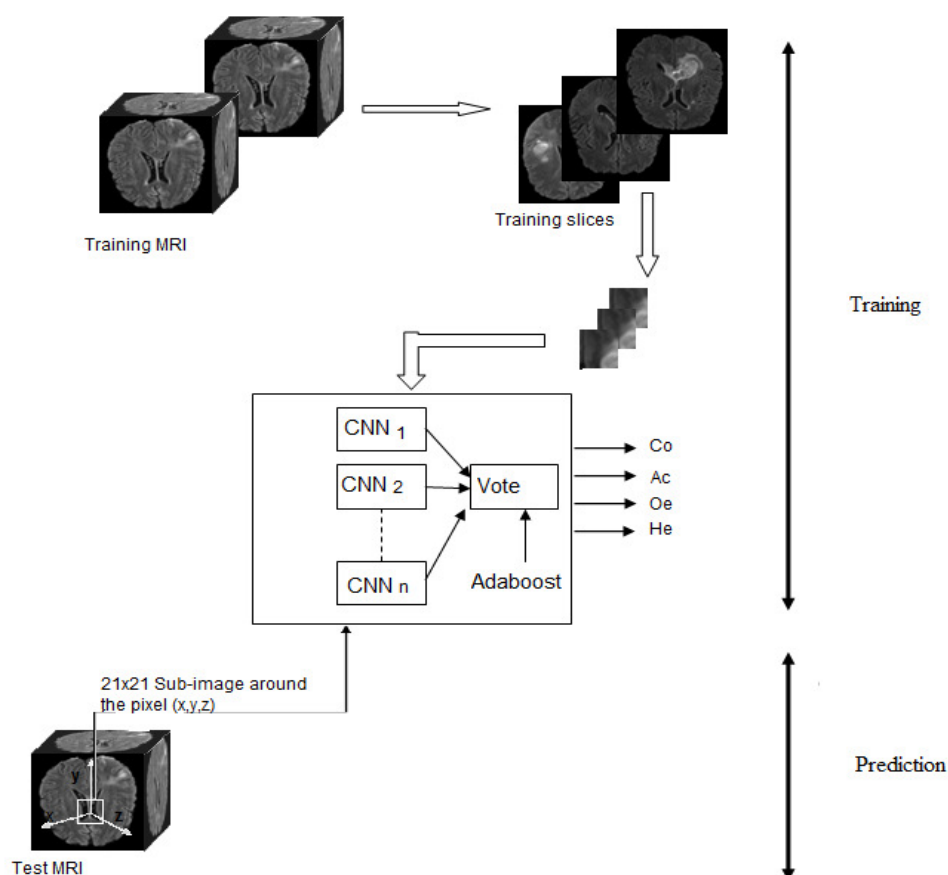


FIGURE 5.5 – Principe de l’ensemble de classifieurs profonds basés pattern proposé pour la segmentation des tumeurs en IRM.

L’apprentissage d’un CNN est une opération très coûteuses à la fois coté du mémoire et coté calcul. Cependant, l’avènement des unités de traitement graphique GPU (Graphics Processing Unit) a aidé à réduire le problème de la complexité de calcul. Mais aussi la mise en disposition des chercheurs et des professionnels de plusieurs CNN adaptés aux différents problèmes, et qui sont entraînés sur des grandes bases des images, permet de raccourcir l’étape d’entraînement des réseaux de neurones convolutifs. Cependant ce n’est pas le cas dans notre approche, nos données d’entrée sont des sous-images représentant le pixel/voxel à classifier et son voisinage élargi (voir Figure 5.4), la sous-image d’entrée peut contenir un pattern dans le voisinage du pixel à classifier. L’architecture de l’ensemble utilisé de réseaux de neurones convolutif est représentée dans la Figure 5.6, et expliquée ci-dessous.

L’entrée du réseaux de neurone convolutif est une sous-image de taille  $21 \times 21 \times 1$  (le 1 indique que c’est une image en niveau de gris (scalaire)).

— Les deux premières couches sont des couches de convolution, qui utilisent 64 filtres

de taille  $3 \times 3$  et un pas de 1 pixel. Chaque couche de convolution a une fonction de correction  $\text{ReLU Max}(x,0)$ , ce qui résulte en  $21 \times 21 \times 64$  paramètres. Suivis par une couche de Pooling avec un max-pool de taille  $2 \times 2$ , ce qui réduit le volume de données de  $21 \times 21 \times 64$  à  $10 \times 10 \times 64$ .

- Deux autres couches de convolution avec 128 filtres de taille  $3 \times 3$  et un pas de 1 pixel avec une fonction de correction  $\text{ReLU}(x) = \text{Max}(x,0)$ , donc on obtient  $10 \times 10 \times 128$  réduit par une couche de pooling au volume de  $5 \times 5 \times 128$ .
- Une couche entièrement connectée (FC) avec 512 canaux suivi d'une autre entièrement connectée de 256 canaux, et enfin une couche de 128 canaux.
- La dernière couche entièrement connectée avec une fonction de softmax renvoie un vecteur de taille égale au nombre de classes, 4 dans notre cas chacune de ses valeurs représente la probabilité d'appartenance à la classe correspondante.

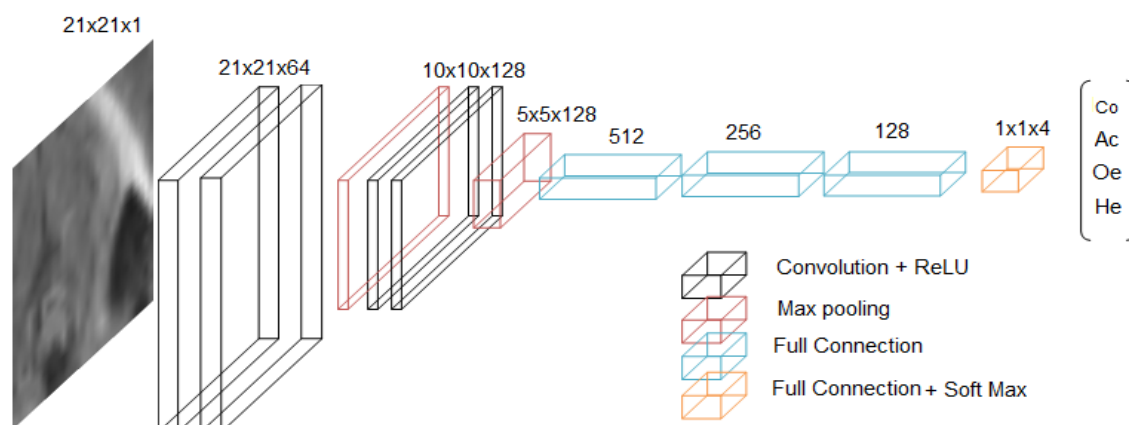


FIGURE 5.6 – L'architecture du CNN léger, utilisé comme classificateur faible avec AdaBoost.

Comme on peut le remarquer sur la Figure 5.5, l'ensemble des IRM d'entraînement est divisé en tranches. Ces dernières sont également subdivisées en sous-images 2D de  $21 \times 21$  voxels. Chacune de ces dernières forme l'ensemble des traits bruts dont il dépend la classe du voxel au centre de la sous-image. Selon un tel schéma, seuls quelques volumes d'IRM seront suffisants pour générer des dizaines de milliers de sous-images de  $21 \times 21$  pour l'entraînement. Après plusieurs tests, le nombre de classifieurs CNNs faibles qui seront utilisés avec l'algorithme AdaBoost a été défini expérimentalement (voir Figure 5.7). Ainsi, l'amélioration de la précision de l'étiquetage des voxels est due à deux facteurs :

premièrement, l'utilisation de CNN, où la largeur de la sous-image autour du voxel central, et la capture du pattern correspondant, permettent d'étiqueter avec précision ce dernier. Deuxièmement, l'utilisation de l'AdaBoost avec les CNNs comme classifieurs faibles. En effet, ceci permet de pallier au problème des voxels mal étiquetés, dû aux différents artefacts qui affectent les IRM, principalement le bruit et la non uniformité des intensités (INU). Dans notre cas, et grâce au schéma adopté, nous sommes en mesure de traiter ces artefacts, sans altérer les données brutes par filtrage du bruit ou correction de l'INU. Par ailleurs, le faible nombre de pixels dans la sous-image autour du pixel à classifier, permet un apprentissage rapide du CNN léger, et aussi une inférence en temps réel des classes des voxels. A noter que les CNN sont indépendants, et de ce fait, leur apprentissage, et l'inférence des classes peuvent être totalement parallèles.

## 5.4 Expérimentation et évaluation

Pour l'expérimentation de la méthode proposée nous avons fait recours aux volumes d'IRM de la base de données, bien connue pour la segmentation des tumeurs cérébrales : BraTS [43]. Cette base de données fournit un large ensemble de volumes IRM contenant des tumeurs cérébrales à différents stades et de différentes tailles. Les volumes IRM sont fournis avec leur étiquetage de réalité terrain, permettant aux auteurs d'utiliser des méthodes basées sur l'apprentissage automatique et d'évaluer quantitativement les méthodes proposées. Toutes les IRM ont une taille de  $240 \times 240 \times 155$  voxels. Dans ce travail, elles sont considérées uniquement les IRM avec modalité Flair (voir Figure 5.3).

### 5.4.1 Sélection des paramètres

Un seul paramètre doit être défini pour la méthode proposée. Il s'agit du nombre d'itérations de l'algorithme Adaboost, qui correspond aussi au nombre de classifieurs faibles ( $T$ ) (voir Algorithme 4) qui seront combinés pour classer les voxels IRM. Ce paramètre a été ajusté quantitativement en testant un ensemble de valeurs, puis nous avons considéré la valeur à partir de laquelle il n'y a plus d'amélioration significative de la précision de la classification, exprimée par l'indice Dice (voir sous-section ci-dessous). Concrètement nous avons varié  $T$  dans l'ensemble  $[1, 2, \dots, 15]$ , où nous avons remarqué que la précision de la classification pour un ensemble d'IRM d'entraînement ne s'améliore

pas à partir de  $T = 9$  (voir Figure 5.7). Donc, nous fixons  $T$  à 10 pour la suite de l'expérimentation. Cette faible valeur du paramètre  $T$  et son unicité nous permettent d'affirmer que la méthode proposée pourrait être caractérisée par un apprentissage rapide des paramètres. Enfin, nous soulignons que la taille des sous-images a été initialement fixée à  $21 \times 21$  voxels, en supposant qu'une telle taille permet de représenter tous les motifs locaux potentiellement existant autour d'un voxel dans une IRM.

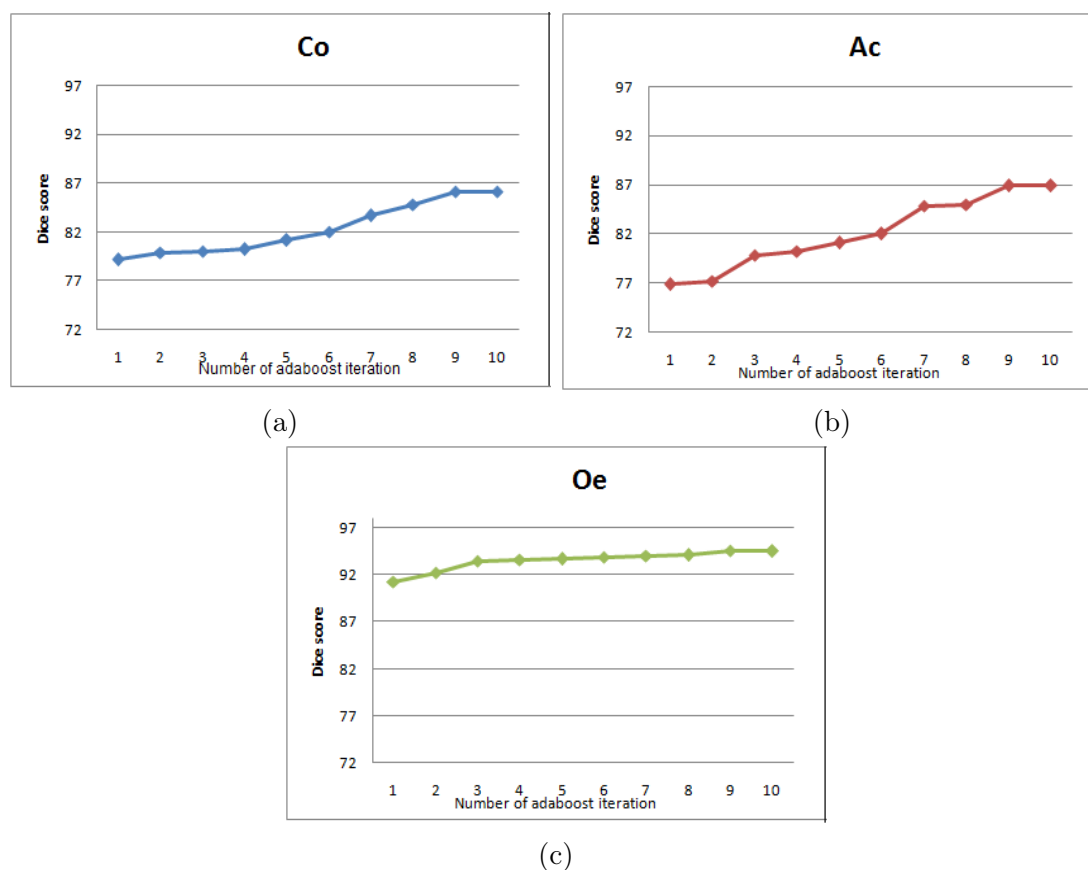


FIGURE 5.7 – Variation de la précision de la segmentation tumorale en fonction du nombre d'itérations de l'algorithme Adaboost.(a) :Co,(b) :Ac,(c) :Oe.

## 5.4.2 Évaluation des performances

A l'instar du travail sur la segmentation structurale des IRM, présenté au chapitre précédent, nous avons utilisé l'indice de Dice pour évaluer notre méthode de segmentation des tumeurs et pour comparer nos résultats à ceux de la littérature. Basé sur les instances d'étiquetage vrai positif ( $TP$ ), vrai négatif ( $TN$ ) et faux positif ( $FP$ ), l'indice de Dice est exprimé comme suit :

$$Dice = \frac{2TP}{2TP + TN + FP} \quad (5.1)$$

### 5.4.3 Résultats expérimentaux

Tout d'abord, nous introduisons sur la Figure 5.8 deux IRM différentes, chacune contenant une tumeur cérébrale, où l'on peut remarquer ses différentes parties. Dans celui-ci, nous avons introduit de gauche à droite : la modalité Flair des données IRM, la réalité terrain, et les résultats selon notre méthode. Notre méthode a pu extraire les différents types de tissus qui composent la tumeur, à savoir le noyau (Co), la partie active (Ac), l'œdème (Oe), en plus des tissus sains (He) extérieurs à la tumeur. En comparant visuellement les régions obtenues par la méthode proposée à la réalité terrain, on peut remarquer que les tissus d'intérêt dans la tumeur ont été efficacement délimités, représentant les régions d'intérêt pour les médecins, à savoir le cœur de la tumeur, sa partie active, et l'œdème autour.

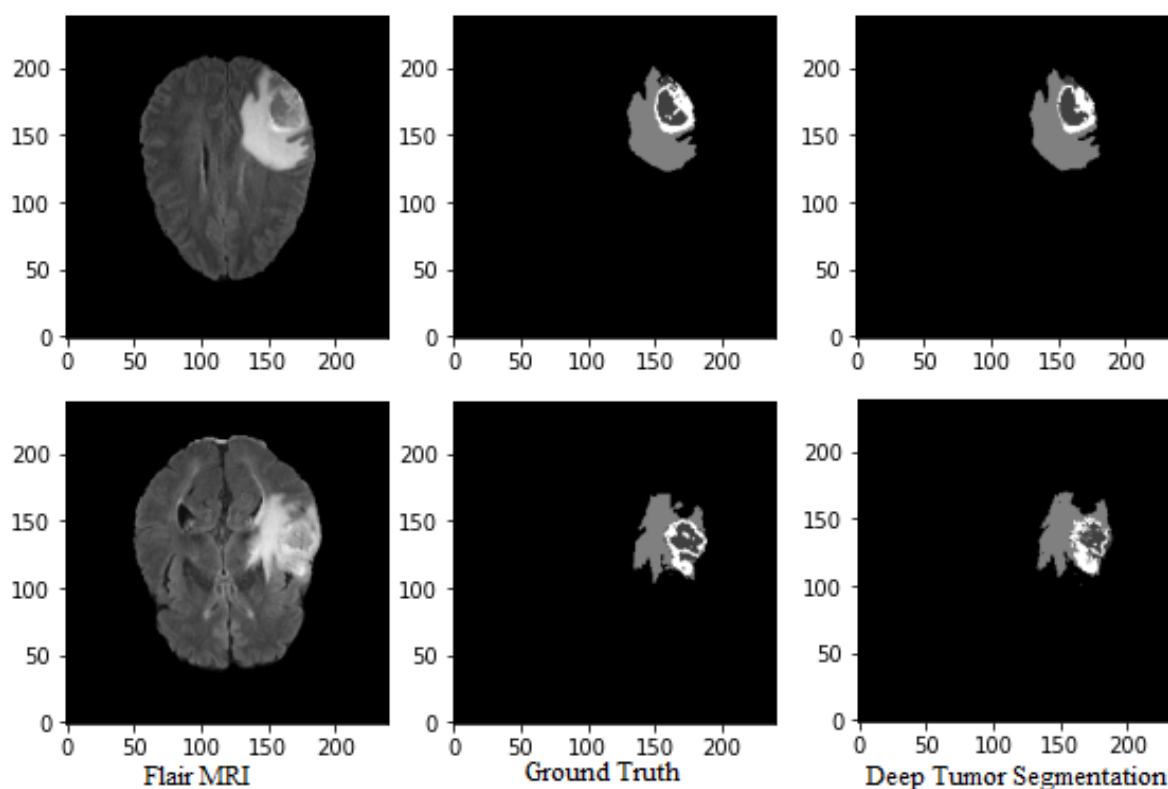


FIGURE 5.8 – Résultats de segmentation présentés sur deux tranches différentes. De gauche à droite : IRM Flair, Réalité terrain, et résultats obtenus de segmentation de la tumeur.

Le tableau 5.1 montre les résultats de la segmentation de l'ensemble des IRM de la

base de données BraTS, selon l'indice Dice en utilisant la méthode proposée. Pour chaque tissu de la tumeur, en plus du tissu sain, nous avons calculé la moyenne de l'indice de Dice pour toutes les IRM impliquées dans l'expérimentation, et nous avons également introduit l'écart-type correspondant. L'écart-type montre comment les résultats sont stables à travers l'ensemble des IRM testés. D'après les résultats présentés dans le tableau 5.1,

TABLE 5.1 – Résultats de segmentation selon l'indice Dice pour les différentes IRM et les différents tissus tumoraux (noyau (Co), partie active (Ac), œdème (Oe) et tissu sain (He)).

	Co	Ac	Oe	He
Moyenne	83.47	82.59	93.63	99.63
Écart type	2.07	3.12	1.62	0.19

on remarque les scores élevés de la segmentation pour les différents tissus tumoraux. De plus, l'écart-type était relativement faible pour tous les tissus. Cette amélioration peut s'expliquer par la capture de l'ensemble des patterns qui entourent un voxel dans une sous-image. Ces patterns sont capturés par le classifieur CNN, et l'amélioration de la précision globale est obtenue grâce à la classification d'ensemble réalisée par l'algorithme Adaboost. Contrairement à d'autres travaux ayant utilisé des schémas complexes afin d'améliorer la segmentation tumorale [129, 130], le nôtre utilise uniquement l'algorithme Adaboost, et dans notre cas il ne nécessite pas un nombre élevé d'itérations. On remarque sur la figure 5.7 les variations de l'indice de Dice en fonction du nombre d'itérations de l'algorithme AdaBoost. La performance maximale de l'algorithme est atteinte à partir de la 9<sup>ième</sup> itération. Un si faible nombre d'itérations permet d'affirmer que la classification des voxels tumoraux selon la méthode proposée est rapide et envisageable pour des applications réelles. Afin de montrer l'efficacité du schéma profond proposé, nous introduisons dans le tableau 5.2 une comparaison entre les résultats de classification des données IRM obtenus avec le modèle proposé et ceux de certains travaux récents de la littérature, ayant utilisé la base de données BraTS, et pour chacun d'eux nous présentons une brève revue ci-après. Visant à segmenter et à enregistrer des scintigraphies cérébrales avec gliome, Kwon et al. [131] ont combiné plusieurs techniques, principalement la croissance de régions à graines multiples, contrôlée par un atlas, la marche aléatoire pour générer la forme préalable de la tumeur, et enfin l'algorithme de maximisation de vraisemblance pour produire la cartographie entre l'atlas et les tranches de l'IRM. Pereira et al. [69] ont utilisé un petit CNN basé sur un noyau de  $3 \times 3$  pour segmenter les IRM cérébrales avec gliome,

avec incorporation de la normalisation de l'intensité. Ils affirment qu'un tel choix permet de proposer une architecture plus profonde qui traite le problème de surapprentissage et fonctionne avec moins de paramètres. Chen et al. [129] ont proposé un Framework supervisé en deux étapes pour la segmentation des lésions cérébrales. Dans la première étape, la région d'intérêt (ROI) est extraite à l'aide d'un classificateur de forêt aléatoire, entraîné par une carte de probabilité du modèle de mixture gaussienne (GMM). Dans la deuxième étape, la région d'intérêt est segmentée en utilisant également un classificateur forêt aléatoire et la carte de probabilité optimisée. Une méthode de segmentation des tumeurs cérébrales basée sur des réseaux de neurones convolutifs en cascade (MCCNN) et des champs aléatoires conditionnels (CRF) entièrement connectés a été proposée par Hu et al. [132]. Le CNN multi-cascade permet de produire une segmentation grossière, puis une segmentation fine est obtenue par les CRFs. Afin de tenir compte des règles sous-jacentes à la pratique clinique, Zhang et al. [130] ont proposé un nouveau réseau de segmentation des tumeurs cérébrales structuré en tâches (réseau TSBTS). Les auteurs font référence à deux propriétés liées aux pratiques des médecins, qui sont respectivement la structure tâche-modalité et la structure tâche-tâche. Dans un autre travail, Zhang et al. [133] utilise des données d'IRM multimodales et des réseaux de neurones convolutifs pour proposer une nouvelle méthode de segmentation des tumeurs cérébrales. L'idée principale est d'effectuer une fouille de modèles pour traiter le problème de données insuffisante.

TABLE 5.2 – Comparaison des résultats selon l'indice de Dice avec certains travaux de la littérature.

Méthode	Noyau	Partie active
Pereira et al. ([69])	83.00	77.00
Kwon et al. ([131])	83.00	72.00
Chen et al. ([129])	79.00	75.00
Zhang et al. (1) ([130])	82.40	78.20
Zhang et al. (2) ([133])	83.60	90.30
Hu et al. ([132])	74.81	71.78
Notre schéma du CNN d'ensemble	86.46	88.11

Nous pouvons remarquer dans le tableau 5.2 que notre méthode obtient de meilleurs résultats comparée à toutes les méthodes impliquées, et ce pour la délimitation des noyaux des tumeurs. Pour la partie active, notre méthode obtient de meilleurs résultats que la plupart des méthodes examinées, à l'exception de la méthode introduite par Zhang et al. qui utilise l'exploration de modèles de données IRM multimodalité avec des CNNs [133],

qui obtient un score de 90,30, par rapport au nôtre qui est de 88,11. De tels résultats confirment que malgré l'architecture simple du CNN utilisé et la faible complexité de calcul qui en résulte, le large éventail de caractéristiques brutes, représentées par les modèles sous-jacents dans les sous-images, a permis de détecter et de segmenter efficacement les tumeurs cérébrales dans les IRM.

#### 5.4.4 Analyse des résultats et discussion

Selon les différents résultats introduits ci-dessus, on remarque des scores élevés pour les différentes parties des tumeurs cérébrales, grâce au classificateur d'ensemble proposé avec des CNN comme classificateurs faibles. L'utilisation de petites sous-images pour définir un large ensemble de caractéristiques brutes pour l'étiquetage des voxels a permis une forte augmentation des valeurs de l'indice Dice, même sans filtrage du bruit ni correction de l'INU. Cela peut s'expliquer par le fait que les CNNs utilisés permettent de capturer les patterns autour des voxels, ce qui conduit à une grande précision de classification, en plus de la capacité de l'ensemble de classificateurs à corriger automatiquement toute erreur d'étiquetage qui pourra être produite par un CNN unique. Selon le schéma proposé, un voxel n'est pas classifié uniquement selon son intensité, mais aussi selon le pattern défini dans son large voisinage. Comme les données d'entraînement sont organisées en un ensemble de sous-images de petites tailles, nous avons pu générer un nombre suffisant de sous-images nécessaires à l'entraînement des CNNs profonds proposés. De plus, la petite taille des sous-images considérées a permis de pallier au problème de complexité calculs lors de l'apprentissage des CNNs. Une telle caractéristique de notre modèle profond permet d'utiliser la méthode avec peu d'IRM avec leur étiquetage de vérité terrain, et permet une segmentation rapide et efficace des tumeurs.

## 5.5 Conclusion

Dans ce chapitre, nous avons introduit un nouveau modèle basé pattern pour la segmentation des tumeurs cérébrales en IRM. Un ensemble de classifieurs CNN profonds a été utilisé pour étiqueter les voxels en fonction des différents tissus des tumeurs cérébrales. Les fonctionnalités introduites pour un CNN consistent en l'ensemble de voxels qui entourent le voxel considéré pour l'étiquetage, dans un large voisinage. De telles ca-

ractéristiques sont suffisantes pour prendre en compte les patterns qui y existent et qui permettent de bien étiqueter le voxel en question. En effet, étiqueter un voxel en IRM ne dépend pas seulement de son intensité mais aussi de ses interactions avec le voisinage. Avec l'algorithme AdaBoost, un classifieur faible a été utilisé, à savoir un CNN léger. L'expérimentation de la méthode proposée sur une base de données dédiée, a montré de bons scores de classification des différents tissus tumoraux. De plus, l'algorithme AdaBoost converge en moins de 10 itérations, ce qui rend l'étape d'apprentissage trop rapide. Dans les travaux futurs, les autres modalités du codage IRM peuvent être conjointement utilisées, ce qui augmente d'avantage la diversité des données. Par ailleurs, les fonctionnalités proposées peuvent être testées avec d'autres classifieurs neuronaux profonds, et une optimisation supplémentaire pourrait être appliquée à l'algorithme AdaBoost, en tenant compte des a priori factuels sur les IRM cérébrales pathologiques.

# Chapitre 6

## Conclusion générale

## 6.1 Conclusion générale

Après quatre décennies de recherche intense, la segmentation d'images en générale, et la segmentation d'images médicales en particulier, demeure problématique. Malgré le progrès technologique considérable dans le domaine d'acquisition des images, ces dernières continuent à être entachées de bruit et affectées par diverses déformations, rendant leur traitement automatique, notamment leur segmentation et interprétation, insatisfaisants pour les utilisateurs. Par ailleurs, les images médicales, sont reconnues pour leur degré élevé des différents artéfacts qui les affectent. En plus du bruit et de la non-uniformité des intensités dans certains types d'images médicales, tels que les IRM, dûs aux appareils d'acquisition, certains artéfacts sont dûs aux conditions d'acquisition, telles que le mouvement des patients et de leurs organes ainsi que l'anatomie du corps humains, qui engendre des images à faible contraste.

Pour contribuer à résoudre le problème de segmentation d'images à résonance magnétique (IRM), nous avons proposé dans le cadre de cette thèse différents modèles de représentation des données IRM et différents modèles de classification de ces données, et ce, dans le but d'améliorer les résultats de segmentation, tout en assurant un temps de calcul raisonnable, acceptable par les praticiens de santé. Pour ce faire, nous sommes penchés sur la recherche des attributs fortement représentatifs, permettant de bien discriminer les voxels et de les classifier fiablement par rapport aux tissus biologiques représentés dans les IRM cérébrales. Après avoir constaté que la seule intensité du voxel est loin de pouvoir représenter la diversité des tissus cérébraux, nous avons jugé nécessaire que les attributs devraient pouvoir capturer les interactions entre les voxels dans des voisinages plus au moins étendus. A cette fin, nous nous sommes arrivés à proposer un attribut, consistant en l'entropie spatiale des voxels formant le voisinage du voxel à

classifier. Cette entropie nous a permis de bien caractériser les patterns qui se forment autour des voxels et qui permettrons, à leur tour, de classifier correctement les voxels d'une IRM.

Ainsi, lors de notre première contribution, nous avons proposé une nouvelle méthode ensembliste basée sur l'entropie spatiale des intensités pour une segmentation structurelle des IRM. En effet, en plus de la puissance de l'entropie pour caractériser les patterns locaux, l'algorithme ensembliste Adaboost nous a permis d'une part de bien équilibrer les données d'apprentissage, étant très volumineuses pour ce type d'images, et d'autre part bénéficier de diversité en données et entraînement en utilisant des classifieurs classiques faibles dont l'ensemble forme le méta classifieur. Diverses configurations des attributs, dépendant du voisinage du voxel et incluant l'entropie ont été testés. Les expérimentations ont été menées sur la base d'IRM "Brain Web" où l'objectif était de segmenter la matière cérébrale en ses trois tissus constitutifs, à savoir : la matière grise, la matière blanche, et le liquide céphalo-rachidien. Le recours à cette base est motivé par la diversité d'IRM avec des niveaux connus au préalable des artéfacts du bruit et la non-uniformité des intensités, et ce contrairement aux autres bases qui ne fournissent pas ces niveaux, telle que IBSR. Les résultats obtenus nous ont permis de conclure que l'attribut de l'entropie spatiale, utilisé avec l'algorithme Adaboost, offre un framework efficace pour une segmentation fiable d'IRM, et en un temps d'exécution raisonnable.

En seconde contribution, et en suivant une approche de segmentation basée patterns locaux, nous avons proposé un modèle profond pour la classification des tissus formant les tumeurs cérébrales. En adoptant le principe de capture des interactions entre voxels dans les voisinages, nous avons proposé un modèle ensembliste utilisant l'Adaboost, et en adaptant un CNN léger comme classifieur faible. L'ensemble des attributs à l'entrée de tout classifieur faible consiste en la sous-image 2D de taille  $21 \times 21$ , centrée sur le voxel à classifier. Une telle sous-image est considérée suffisamment large pour pouvoir caractériser les patterns qui sont dedans, et qui permettent de bien représenter le voxel à classifier. De toute évidence, nous n'avons pas fait recours à aucun attribut de synthèse, tel que l'entropie, car selon une approche basée classifieurs profonds s'il y a un attribut caché, il sera inféré et représenté par les couches profondes du classifieur.

Le modèle proposé a été testé en faisant recours à la base d'évaluation de segmentation de tumeurs cérébrales, à savoir BraTS'17 (Brain Tumor Segmentation). Nous avons

opté pour l'indice de Dice pour l'évaluation de performance, et ce pour pouvoir comparer nos résultats à ceux des travaux récents ayant traité du même sujet. Les résultats expérimentaux obtenus sur l'ensemble des IRM de la base BraTS nous ont permis de conclure de l'opportunité d'utilisation d'un CNN léger comme classifieur faible pour un méta-classifieur ensembliste pour l'étiquetage des voxels d'une tumeur selon son appartenance aux différents tissus tumoraux, à savoir le coeur de la tumeur, sa partie active, et l'oedème qui l'entoure.

De plus, le modèle léger du CNN nous a permis de réduire considérablement le nombre de variables à ajuster lors de son apprentissage. En effet, une taille d'image de  $21 \times 21$  nécessitera un temps d'apprentissage largement inférieur à ceux des méthodes profondes qui considèrent toute une coupe complète d'une IRM. Cette taille réduite de la sous-image permet aussi d'étiqueter un voxel en un temps très court, et donc la segmentation de tout le volume de l'IRM, incluant la tumeur, se fera en un temps très acceptable par les médecins et les radiologues.

Nous pensons que nos travaux dans le cadre de cette thèse, basés sur l'hypothèse que la représentation et l'utilisation de patterns locaux autour des voxels d'une IRM, ou d'un pixel en général dans une image quelconque, ont ouvert une nouvelle piste, prometteuse, pour proposer de nouvelles méthodes de segmentation plus fiables. En effet, que ce soit en utilisant une représentation explicite par attributs avec des classifieurs classiques, ou en adoptant une approche profonde avec des classifieurs profonds légers, les méta-classifieurs ensemblistes semblent être prometteurs, notamment si les images sont de volumes importants. De cela, les perspectives à nos travaux sont nombreuses dont on peut citer certains : La recherche de patterns plus concis ainsi que l'adaptation de différents classifieurs profonds pour qu'ils soient utilisés comme classifieurs faibles pour des méta-classifieurs ensemblistes. D'autre part, l'entropie dans l'espace dual des attributs est un chemin potentiel pour renforcer les attributs à considérer pour la classification en segmentation d'IRM, et d'image en général.

# Bibliographie

- [1] B. Kastler and D. Vetter, *Comprendre l'IRM : manuel d'auto-apprentissage*. Elsevier Health Sciences, 2018.
- [2] devoteam/Paroles d'experts, "Diagnostiquer une pathologie," 2016. <https://tpelyceeblaisepascalorsayirm.wordpress.com/lirm-clinique-2/diagnostiquer-une-pathologie/>, Last accessed on November 2, 2022.
- [3] P. Sprawls, *Magnetic Resonance Imaging : Principles, Methods, and Techniques*. Medical Physics Publishing Corporation, 2000.
- [4] F. Durand-Dubief, "Place de l'irm dans le suivi de la sclérose en plaques," 2013. <http://www.rhone-alpes-sep.org/wp-content/uploads/2013/02/La-place-de-lIRM-dans-le-suivi-de-la-SEP-Dr-Françoise-Durand-Dubief.pdf>, Last accessed on November 2, 2022.
- [5] P. Coupé, J. V. Manjón, E. Gedamu, D. Arnold, M. Robles, and D. L. Collins, "Robust rician noise estimation for mr images," *Medical image analysis*, vol. 14, no. 4, pp. 483–493, 2010.
- [6] B. Scherrer, *Segmentation des tissus et structures sur les IRM cérébrales : agents markoviens locaux coopératifs et formulation bayésienne*. Theses, Institut National Polytechnique de Grenoble - INPG, Dec. 2008.
- [7] M. S. Cohen, R. M. DuBois, and M. M. Zeineh, "Rapid and effective correction of rf inhomogeneity for high field magnetic resonance imaging," *Human brain mapping*, vol. 10, no. 4, pp. 204–211, 2000.
- [8] Y. Xiao, V. Fonov, M. M. Chakravarty, S. Beriault, F. Al Subaie, A. Sadikot, G. B. Pike, G. Bertrand, and D. L. Collins, "A dataset of multi-contrast population-averaged brain mri atlases of a parkinsons disease cohort," *Data in Brief*, vol. 12, pp. 370–379, 2017.

- [9] J. Wang, C. Vachet, A. Rumble, S. Gouttard, C. Ouziel, E. Perrot, G. Du, X. Huang, G. Gerig, and M. Styner, “Multi-atlas segmentation of subcortical brain structures via the autoseg software pipeline,” *Frontiers in neuroinformatics*, vol. 8, p. 7, 2014.
- [10] S. M. Smith, “Fast robust automated brain extraction,” *Human brain mapping*, vol. 17, no. 3, pp. 143–155, 2002.
- [11] Y. Xu, Y. Wang, S. Hu, and Y. Du, “Deep convolutional neural networks for bias field correction of brain magnetic resonance images,” *The Journal of Supercomputing*, pp. 1–26, 2022.
- [12] X.-H. Han and Y.-W. Chen, “A robust method based on ica and mixture sparsity for edge detection in medical images,” *Signal, Image and Video Processing*, vol. 5, pp. 39–47, 2011.
- [13] H. Shamsi and H. Seyedarabi, “A modified fuzzy c-means clustering with spatial information for image segmentation,” *International journal of computer theory and engineering*, vol. 4, no. 5, p. 762, 2012.
- [14] N. K. Subbanna, D. Precup, D. L. Collins, and T. Arbel, “Hierarchical Probabilistic Gabor and MRF Segmentation of Brain Tumours in MRI Volumes,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013* (K. Mori, I. Sakuma, Y. Sato, C. Barillot, and N. Navab, eds.), (Berlin, Heidelberg), pp. 751–758, Springer Berlin Heidelberg, 2013.
- [15] G. C. Oliveira., R. Varoto., and A. C. Jr., “Brain Tumor Segmentation in Magnetic Resonance Images using Genetic Algorithm Clustering and AdaBoost Classifier,” in *Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies*, vol. 2, pp. 77–82, INSTICC, SciTePress, 2018.
- [16] T. Fülöp, Á. Györfi, B. Surányi, L. Kovács, and L. Szilágyi, “Brain tumor segmentation from mri data using ensemble learning and multi-atlas,” in *2020 IEEE 18th World Symposium on Applied Machine Intelligence and Informatics (SAMi)*, pp. 111–116, 2020.
- [17] A. de Brébisson and G. Montana, “Deep Neural Networks for Anatomical Brain Segmentation,” pp. 20–28, 2015.

- [18] M. F. Ak, “A comparative analysis of breast cancer detection and diagnosis using data visualization and machine learning applications,” *Healthcare*, vol. 8, p. 111, 04 2020.
- [19] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [20] M. Saber, E. R. Abdessamad, S. Rachid, H. Kharraz Aroussi, and M. Wahbi, “Artificial neural networks, support vector machine and energy detection for spectrum sensing based on real signals,” *International Journal of Communication Networks and Information Security*, vol. 11, 04 2019.
- [21] S. Ghoshal and S. Roberts, “Thresholded convnet ensembles : neural networks for technical forecasting,” *Neural Computing and Applications*, vol. 32, 09 2020.
- [22] C. Xu, J. Wang, T. Zheng, Y. Cao, and Y. Fan, “Prediction of prognosis and survival of patients with gastric cancer by weighted improved random forest model,” *Archives of Medical Science*, vol. 18, 04 2021.
- [23] M. Khairandish, M. Sharma, V. Jain, J. Chatterjee, and N. Jhanjhi, “A hybrid cnn-svm threshold segmentation approach for tumor detection and classification of mri brain images,” *IRBM*, 2021.
- [24] devoteam/Paroles d’experts, “Mieux comprendre le deep learning appliqué à la reconnaissance d’images,” 2021. <https://france.devoteam.com/paroles-dexperts/mieux-comprendre-le-deep-learning-applique-a-la-reconnaissance-dimages/>, Last accessed on November 2, 2022.
- [25] I. S. Mohamed, *Detection and Tracking of Pallets using a Laser Rangefinder and Machine Learning Techniques*. PhD thesis, 09 2017.
- [26] V. N. and J. Manjula, “High-performance visual geometric group deep learning architectures for mri brain tumor classification,” *The Journal of Supercomputing*, vol. 78, 07 2022.
- [27] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” in *Proceedings of the IEEE*, vol. 86, pp. 2278–2324, 1998.

- [28] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, 2017.
- [29] S. Mahale and K. Shah, “Detection of plant diseases using convolutional neural network architectures,” *Pulse*, vol. 2, no. 2, 2021.
- [30] D. Ibanez-Soria, A. Soria-Frisch, J. Garcia-Ojalvo, and G. Ruffini, “Characterization of the non-stationary nature of steady-state visual evoked potentials using echo state networks,” *PLOS ONE*, vol. 14, pp. 1–13, 07 2019.
- [31] X. H. Le, H. Ho, G. Lee, and S. Jung, “Application of long short-term memory (lstm) neural network for flood forecasting,” *Water*, vol. 11, p. 1387, 07 2019.
- [32] T. Toharudin, R. Pontoh, R. Caraka, S. Zahroh, Y. Lee, and R.-C. Chen, “Employing long short-term memory and facebook prophet model in air temperature forecasting,” *Communication in Statistics- Simulation and Computation*, pp. 1–24, 2020.
- [33] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, “Attention is all you need,” in *Advances in Neural Information Processing Systems* (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds.), vol. 30, Curran Associates, Inc., 2017.
- [34] I. Despotovic, B. Goossens, and W. Philips, “Mri segmentation of the human brain : Challenges, methods, and applications,” *Computational and Mathematical Methods in Medicine*, vol. 2015, 2015.
- [35] Z. Karim, N. R. Paiker, M. A. Ali, G. Sorwar, and M. M. Islam, “Pattern based object segmentation using split and merge,” in *2009 IEEE International Conference on Fuzzy Systems*, pp. 2166–2169, 2009.
- [36] M. M. Hasan, M. A. Ali, M. H. Kabir, and G. Sorwar, “Object segmentation using block based patterns,” in *TENCON 2009 - 2009 IEEE Region 10 Conference*, pp. 1–6, 2009.
- [37] W. R. Nitske, *The life of Wilhelm Conrad Rontgen, discoverer of the X ray*. Tucson, University of Arizona Press, 1971.

- [38] Y. Y. Abdallah, “History of Medical Imaging,” *Archives of Medicine and Health Sciences*, vol. 5, no. 2, pp. 275–278, 2017.
- [39] S. Rooppakhun, N. Chantarapanich, and K. Sitthiseripratip, “Advanced medical imaging and reverse engineering technologies in craniometric study,” in *Forensic Medicine* (D. N. Vieira, ed.), ch. 14, Rijeka : IntechOpen, 2011.
- [40] D. Béréziat and I. Herlin, “Solving ill-posed Image Processing problems using Data Assimilation,” *Numerical Algorithms*, vol. 56, pp. 219–252, Feb. 2011.
- [41] M. T. Bennai, Z. Guessoum, S. Mazouzi, S. Cormier, and M. Mezghiche, “A stochastic multi-agent approach for medical-image segmentation : Application to tumor segmentation in brain MR images,” *Artificial Intelligence in Medicine*, vol. 110, p. 101980, 2020.
- [42] C. Cocosco, V. Kollokian, R.-S. Kwan, and A. Evans, “Simulated brain database homepage.” <https://brainweb.bic.mni.mcgill.ca/brainweb>, Last Accesse on 2020-06-13.
- [43] B. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, L. Lanczi, E. Gerstner, M.-A. Weber, T. Arbel, B. Avants, N. Ayache, P. Buendia, L. Collins, N. Cordier, J. Corso, A. Criminisi, T. Das, H. Delingette, C. Demiralp, C. Durst, M. Dojat, S. Doyle, J. Festa, F. Forbes, E. Geremia, B. Glocker, P. Golland, X. Guo, A. Hamamci, K. Iftekharruddin, R. Jena, N. John, E. Konukoglu, D. Lashkari, J. Antonio Mariz, R. Meier, S. Pereira, D. Precup, S. J. Price, T. Riklin-Raviv, S. Reza, M. Ryan, L. Schwartz, H.-C. Shin, J. Shotton, C. Silva, N. Sousa, N. Subbanna, G. Szekely, T. Taylor, O. Thomas, N. Tustison, G. Unal, F. Vasseur, M. Wintermark, D. Hye Ye, L. Zhao, B. Zhao, D. Zikic, M. Prastawa, M. Reyes, and K. Van Leemput, “The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS),” *IEEE Transactions on Medical Imaging*, p. 33, 2014. <https://hal.inria.fr/hal-00935640>.
- [44] M. H. Levitt, *Spin dynamics : basics of nuclear magnetic resonance*. New York, NY : Wiley, 2001.
- [45] W. Zhang, L. Liu, H. Chen, K. Hu, I. Delahunty, S. Gao, and J. Xie, “Surface impact on nanoparticle-based magnetic resonance imaging contrast agents,” *Theranostics*, vol. 8, pp. 2521–2548, 2018.

- [46] M. Hoff, J. Andre, and B. Stewart, *Artifacts in Magnetic Resonance Imaging*, pp. 165–190. CRC Press, 2016.
- [47] K. Krupa and M. Bekiesińska-Figatowska, “Artifacts in magnetic resonance imaging,” *Polish Journal of Radiology*, vol. 80, pp. 93 – 106, 2015.
- [48] M. Beladgham, F. Derraz, and M. Khélif, “Segmentation d’images médicales IRM par la méthode d’ensembles de niveaux (Levelsets),” *Abou – Bekr Belkaid university – Tlemcen*, 2005.
- [49] A. Iftikhar, A. Jalil, S. Rathore, A. Ali, and M. Hussain, “Brain MRI denoizing and segmentation based on improved adaptive nonlocal means,” *International journal of imaging systems and technology*, vol. 23, pp. 235–248, 09 2013.
- [50] N. Richard, *Une approche située, coopérative et décentralisée pour l’interprétation d’images cérébrales par RMN*. PhD thesis, 2004. Thèse de doctorat dirigée par Dojat, Michel et Garbay, Catherine Imagerie, vision, robotique Grenoble INPG 2004.
- [51] S. Widz, K. Revett, and D. Ślęzak, “A rough set-based magnetic resonance imaging partial volume detection system,” in *International Conference on Pattern Recognition and Machine Intelligence*, vol. 3776, pp. 756–761, Springer, 12 2005.
- [52] J. Wang, J. Kong, Y. Lu, M. Qi, and B. Zhang, “A modified fcm algorithm for mri brain image segmentation using both local and non-local spatial constraints,” *Computerized Medical Imaging and Graphics*, vol. 32, no. 8, pp. 685–698, 2008.
- [53] D. Azzouz and S. Mazouzi, “A hyper-surface-based modeling and correction of bias field in mr images,” *Jordanian Journal of Computers and Information Technology (JJCIT)*, vol. 07, no. 03, pp. 223–238, 2021.
- [54] P. A. Narayana and A. Borthakur, “Effect of radio frequency inhomogeneity correction on the reproducibility of intra-cranial volumes using mr image data,” *Magnetic Resonance in Medicine*, vol. 33, no. 3, pp. 396–400, 1995.
- [55] J. W. Murakami, C. E. Hayes, and E. Weinberger, “Intensity correction of phased-array surface coil images,” *Magnetic Resonance in Medicine*, vol. 35, no. 4, pp. 585–590, 1996.

- [56] A. Madabhushi, J. K. Udupa, and A. Souza, “Generalized scale : Theory, algorithms, and application to image inhomogeneity correction,” *Computer Vision and Image Understanding*, vol. 101, no. 2, pp. 100–121, 2006.
- [57] X. Dai, Y. Lei, Y. Liu, T. Wang, L. Ren, W. J. Curran, P. R. Patel, T. Liu, and X. Yang, “Intensity non-uniformity correction in MR imaging using residual cycle generative adversarial network,” *Physics in Medicine & Biology*, vol. 65, no. 21, p. 215025, 2020.
- [58] J. Ashburner and K. J. Friston, “Unified segmentation,” *Neuroimage*, vol. 26, no. 3, pp. 839–851, 2005.
- [59] S. W. Zucker, “Region growing : Childhood and adolescence,” *Computer Graphics and Image Processing*, vol. 5, no. 3, pp. 382–399, 1976.
- [60] A. B. Rabeh, F. Benzarti, and H. Amiri, “Segmentation of brain MRI using active contour model,” *Int. J. Imaging Syst. Technol.*, vol. 27, no. 1, pp. 3–11, 2017.
- [61] M. Kaleem, M. Sanaullah, M. A. Hussain, M. A. Jaffar, and T.-S. Choi, “Segmentation of Brain Tumor Tissue Using Marker Controlled Watershed Transform Method,” in *Emerging Trends and Applications in Information Communication Technologies*, (Berlin, Heidelberg), pp. 222–227, Springer Berlin Heidelberg, 2012.
- [62] J. Bezdek, R. Ehrlich, and W. E. Full, “FCM : The fuzzy c-means clustering algorithm,” *Computers et Geosciences*, vol. 10, pp. 191–203, 1984.
- [63] N. Gordillo, E. Montseny, and P. Sobrevilla, “State of the art survey on MRI brain tumor segmentation,” *Magnetic Resonance Imaging*, vol. 31, no. 8, pp. 1426–1438, 2013.
- [64] S. Li, “Markov Random Field Modeling in Image Analysis,” in *Advances in Pattern Recognition*, 2009.
- [65] N. Yamanakkanavar, J. Y. Choi, and B. Lee, “MRI Segmentation and Classification of Human Brain Using Deep Learning for Diagnosis of Alzheimer’s Disease : A Survey,” *Sensors*, vol. 20, no. 11, p. 3243, 2020.
- [66] R. Rajasree and C. C. Columbus, “Brain tumour image segmentation and classification system based on the modified AdaBoost classifier,” *International Journal of Applied Engineering Research*, vol. 10, no. 14, 2015.

- [67] C. L. Choudhury, C. Mahanty, R. Kumar, and B. K. Mishra, "Brain tumor detection and classification using convolutional neural network and deep neural network," in *2020 International Conference on Computer Science, Engineering and Applications (ICCSEA)*, pp. 1–4, 2020.
- [68] J. Amin, M. Sharif, M. Yasmin, and S. L. Fernandes, "Big data analysis for brain tumor detection : Deep convolutional neural networks," *Future Generation Computer Systems*, vol. 87, pp. 290–297, 2018.
- [69] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1240–1251, 2016.
- [70] Y. Pan, W. Huang, Z. Lin, W. Zhu, J. Zhou, J. Wong, and Z. Ding, "Brain tumor grading based on Neural Networks and Convolutional Neural Networks," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 699–702, 2015.
- [71] W. Zhang, R. Li, H. Deng, L. Wang, W. Lin, S. Ji, and D. Shen, "Deep convolutional neural networks for multi-modality isointense infant brain image segmentation," *NeuroImage*, vol. 108, pp. 214–224, 2015.
- [72] P. Coupé, B. Mansencal, M. Clément, R. Giraud, B. D. de Senneville, V.-T. Ta, V. Lepetit, and J. V. Manjon, "Assemblynet : A large ensemble of cnns for 3d whole brain mri segmentation," *NeuroImage*, vol. 219, p. 117026, 2020.
- [73] D. Marr, *Vision : A Computational Investigation into the Human Representation and Processing of Visual Information*. New York, NY, USA : Henry Holt and Co., Inc., 1982.
- [74] E. Charniak and D. McDermott, *Introduction to Artificial Intelligence*. Addison Wesley, 1985.
- [75] N. Wiener, *Cybernétique et société*. 1971.
- [76] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics*, vol. 5, no. 4, pp. 115–133, 1943.
- [77] J. McCarthy and E. A. Feigenbaum, "In memoriam : Arthur samuel - pioneer in machine learning.," *AI Magazine*, vol. 11, no. 3, pp. 10–11, 1990.

- [78] M. A. Nielsen, “Neural networks and deep learning.” <http://neuralnetworksanddeeplearning.com/>, Last accessed on November 2, 2022.
- [79] N. J. Nilsson, “The quest for artificial intelligence : A history of ideas and achievements,” 2009.
- [80] T. M. Mitchell, *Machine learning*, vol. 1. McGraw-hill New York, 1997.
- [81] B. Bringmann, S. Nijssen, and A. Zimmermann, “Pattern-Based Classification : A Unifying Perspective,” *CoRR*, vol. abs/1111.6191, 2011.
- [82] E. Alpaydin, *Introduction to Machine Learning*. Adaptive Computation and Machine Learning, Cambridge, MA : MIT Press, 3 ed., 2014.
- [83] J. A. Hartigan and M. A. Wong, “A k-means clustering algorithm,” *JSTOR : Applied Statistics*, vol. 28, no. 1, pp. 100–108, 1979.
- [84] T. K. Moon, “The expectation-maximization algorithm,” *IEEE Signal Process. Mag.*, vol. 13, pp. 47–60, 1996.
- [85] R. S. Sutton and A. G. Barto, *Reinforcement Learning : An Introduction*. The MIT Press, second ed., 2018.
- [86] O. Chapelle, B. Schölkopf, and A. Zien, eds., *Semi-Supervised Learning*. The MIT Press, 2006.
- [87] A. Mucherino, P. J. Papajorgji, and P. M. Pardalos, *k-Nearest Neighbor Classification*, pp. 83–106. New York, NY : Springer New York, 2009.
- [88] G. I. Webb, *Naïve Bayes*, pp. 713–714. Boston, MA : Springer US, 2010.
- [89] L. Breiman, “Bagging predictors,” *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.
- [90] Y. Freund and R. Schapire, “A short introduction to boosting,” *Journal-Japanese Society For Artificial Intelligence*, vol. 14, no. 771-780, p. 1612, 1999.
- [91] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [92] I. J. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA : MIT Press, 2016. <http://www.deeplearningbook.org>.
- [93] S. Khan, H. Rahmani, S. A. A. Shah, M. Bennamoun, and EBSCOhost, *A guide to convolutional neural networks for computer vision*. Morgan Claypool [San Rafael, California], 2018.

- 
- [94] F. Millstein, *Convolutional Neural Networks In Python : Beginner's Guide To Convolutional Neural Networks In Python*. North Charleston, SC, USA : CreateSpace Independent Publishing Platform, 2018.
- [95] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks : Analysis, applications, and prospects," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–21, 2021.
- [96] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015.
- [97] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014.
- [98] F. Salem, *Recurrent Neural Networks : From Simple to Gated Architectures*. Springer International Publishing, 2022.
- [99] S. K. Mahata, D. Das, and S. Bandyopadhyay, "Mtil2017 : Machine translation using recurrent neural network on statistical machine translation," *Journal of Intelligent Systems*, vol. 28, no. 3, pp. 447–453, 2019.
- [100] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [101] J. Brownlee, *Long Short-term Memory Networks with Python : Develop Sequence Prediction Models with Deep Learning*. Jason Brownlee, 2017.
- [102] K. Smagulova and A. James, "A survey on lstm memristive neural network architectures and applications," *The European Physical Journal Special Topics*, vol. 228, 05 2019.
- [103] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," in *NIPS 2014 Workshop on Deep Learning, December 2014*, 2014.
- [104] C. Zhou, B. Cule, and B. Goethals, "Pattern Based Sequence Classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, pp. 1285–1298, 2016.
- [105] A. A. Roma, A. Diaz De Vivar, K. J. Park, I. Alvarado-Cabrero, G. Rasty, J. G. Chanona-Vilchis, Y. Mikami, S. R. Hong, N. Teramoto, R. Ali-Fehmi, J. K. L. Rutgers, D. Barbuto, and E. G. Silva, "Invasive endocervical adenocarcinoma : a

- new pattern-based classification system with important clinical significance,” *The American journal of surgical pathology*, vol. 39, p. 667–672, May 2015.
- [106] K. Hornik, “Approximation Capabilities of Multilayer Feedforward Networks,” *Neural Netw.*, vol. 4, p. 251–257, Mar. 1991.
- [107] J. Park and I. W. Sandberg, “Approximation and Radial-Basis-Function Networks,” *Neural Comput.*, vol. 5, pp. 305–316, Mar. 1993.
- [108] R. M. Gray, *Entropy and Information Theory*. Springer Publishing Company, Incorporated, 2nd ed., 2011.
- [109] *Fundamentals of Thermodynamics*, ch. 1, pp. 1–60. John Wiley Sons, Ltd, 2020.
- [110] R. Clausius and F. Folie, *Théorie mécanique de la chaleur*. No. vol. 2 in Bibliothèque des professions industrielles et agricoles. Série B, Eugène Lacroix, 1869.
- [111] C. E. Shannon, “A mathematical theory of communication,” *The Bell System Technical Journal*, vol. 27, pp. 379–423, 1948.
- [112] A. C. Sparavigna, “Entropy in image analysis,” *Entropy*, vol. 21, p. 502, May 2019.
- [113] A. Sedghi, L. J. O’Donnell, T. Kapur, E. Learned-Miller, P. Mousavi, and W. M. Wells, “Image registration : Maximum likelihood, minimum entropy and deep learning,” *Medical Image Analysis*, vol. 69, p. 101939, 2021.
- [114] M. R. Sabuncu and P. J. Ramadge, “Spatial information in entropy-based image registration,” in *Biomedical Image Registration, Second International Workshop, WBIR 2003, Philadelphia, PA, USA, June 23-24, 2003, Revised Papers* (J. C. Gee, J. B. A. Maintz, and M. W. Vannier, eds.), vol. 2717 of *Lecture Notes in Computer Science*, pp. 132–141, Springer, 2003.
- [115] C. Lang and H. Jia, “Kapur’s entropy for color image segmentation based on a hybrid whale optimization algorithm,” *Entropy*, vol. 21, no. 3, 2019.
- [116] X. Yang, F. Li, W. Zhang, and L. He, “Blind image quality assessment of natural scenes based on entropy differences in the dct domain,” *Entropy*, vol. 20, no. 11, 2018.
- [117] S. Smith, “Fast robust automated brain extraction,” *Human Brain Mapping*, vol. 17, 2002.

- [118] M. JENKINSON, “BET2 : MR-Based Estimation of Brain, Skull and Scalp Surfaces,” *Eleventh Annual Meeting of the Organization for Human Brain Mapping, 2005*, 2005.
- [119] N. Bouchaour and S. Mazouzi, “Ensemble Classification Using Entropy-Based Features for MRI Tissue Segmentation,” in *Pattern Recognition and Artificial Intelligence : 4th Mediterranean Conference, MedPRAI 2020, Hammamet, Tunisia, December 20–22, 2020, Proceedings 4*, pp. 125–139, Springer, 2021.
- [120] N. Bouchaour and S. Mazouzi, “Pattern-based Classification Using Entropy Coding For MRI Data Classification,” in *IAM’20 : Third conference on informatics and applied mathematics, 21–22 October 2020, Guelma, ALGERIA*, pp. 51–63, 2020.
- [121] B. Schölkopf, C. Burges, and V. Vapnik, “Extracting Support Data for a Given Task,” in *KDD*, 1995.
- [122] S. Knerr, L. Personnaz, and G. Dreyfus, “Single-Layer Learning Revisited : A Step-wise Procedure for Building and Training a Neural Network,” in *Neurocomputing : Algorithms, Architectures and Applications* (F. Fogelman Soulié and J. Héroult, eds.), vol. F68 of *NATO ASI Series*, pp. 41–50, Springer-Verlag, 1990.
- [123] Y. Zhang, M. Brady, and S. Smith, “Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm,” *IEEE Transactions on Medical Imaging*, vol. 20, pp. 45–57, 2001.
- [124] B. Caldairou, N. Passat, P. Habas, C. Studholme, and F. Rousseau, “A non-local fuzzy segmentation method : Application to brain MRI,” *Pattern Recognition*, vol. 44, no. 9, pp. 1916–1927, 2011.
- [125] B. Bringmann, S. Nijssen, and A. Zimmermann, “From local patterns to classification models,” in *Inductive Databases and Constraint-Based Data Mining* (S. Dzeroski, B. Goethals, and P. Panov, eds.), pp. 127–154, Springer, 2010.
- [126] N. Kumar, N. Kaur, and D. Gupta, “Major Convolutional Neural Networks in Image Classification : A Survey,” in *Proceedings of International Conference on IoT Inclusive Life (ICIIL 2019), NITTTR Chandigarh, India* (M. Dutta, C. R. Krishna, R. Kumar, and M. Kalra, eds.), (Singapore), pp. 243–258, Springer Singapore, 2020.
- [127] S. Yousaf, H. RaviPrakash, S. M. Anwar, N. Sohail, and U. Bagci, “State-of-the-art in brain tumor segmentation and current challenges,” in *Machine Learning in*

- Clinical Neuroimaging and Radiogenomics in Neuro-oncology - Third International Workshop, MLCN 2020, and Second International Workshop, RNO-AI 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4-8, 2020, Proceedings* (S. M. Kia, H. Mohy-ud-Din, A. Abdulkadir, C. Bass, M. Habes, J. M. Rondina, C. M. W. Tax, H. Wang, T. Wolfers, S. Rathore, and M. Ingalhalikar, eds.), vol. 12449 of *Lecture Notes in Computer Science*, pp. 189–198, Springer, 2020.
- [128] N. Bouchaour and S. Mazouzi, “Deep pattern-based tumor segmentation in brain MRIs,” *Neural Computing and Applications*, pp. 1–10, 2022.
- [129] G. Chen, Q. Li, F. Shi, I. Rekiq, and Z. Pan, “RFDCR : Automated brain lesion segmentation using cascaded random forests with dense conditional random fields,” *NeuroImage*, vol. 211, p. 116620, 2020.
- [130] D. Zhang, G. Huang, Q. Zhang, J. Han, J. Han, Y. Wang, and Y. Yu, “Exploring Task Structure for Brain Tumor Segmentation From Multi-Modality MR Images,” *IEEE Transactions on Image Processing*, vol. 29, pp. 9032–9043, 2020.
- [131] D. Kwon, R. T. Shinohara, H. Akbari, and C. Davatzikos, “Combining generative models for multifocal glioma segmentation and registration,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014* (P. Golland, N. Hata, C. Barillot, J. Hornegger, and R. Howe, eds.), (Cham), pp. 763–770, Springer International Publishing, 2014.
- [132] K. Hu, Q. Gan, Y. Zhang, S. Deng, F. Xiao, W. Huang, C. Cao, and X. Gao, “Brain Tumor Segmentation Using Multi-Cascaded Convolutional Neural Networks and Conditional Random Field,” *IEEE Access*, vol. 7, pp. 92615–92629, 2019.
- [133] D. Zhang, G. Huang, Q. Zhang, J. Han, J. Han, and Y. Yu, “Cross-modality deep feature learning for brain tumor segmentation,” *Pattern Recognition*, vol. 110, p. 107562, 2021.