



HAL
open science

Analyse de données RMN multimodales par intelligence artificielle pour la discrimination binaire du grade du gliome

Paul Dequidt

► **To cite this version:**

Paul Dequidt. Analyse de données RMN multimodales par intelligence artificielle pour la discrimination binaire du grade du gliome. Imagerie médicale. Université de Poitiers, 2021. Français. NNT : 2021POIT2255 . tel-03257844

HAL Id: tel-03257844

<https://theses.hal.science/tel-03257844>

Submitted on 11 Jun 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

POUR L'OBTENTION DU GRADE DE
DOCTEUR DE L'UNIVERSITÉ DE POITIERS

FACULTÉ DES SCIENCES FONDAMENTALES ET APPLIQUÉES
DIPLOME NATIONAL - ARRÊTÉ DU 25 MAI 2016

École Doctorale : Sciences et Ingénierie des Systèmes, Mathématiques, Informatique - SISMI
Secteur de Recherche : Traitement du Signal et des Images

Présentée par:

Paul DEQUIDT

**Analyse de données RMN multimodales par intelligence artificielle
pour la discrimination binaire du grade du gliome**

Directeurs de Thèse:

Christine FERNANDEZ-MALOIGNE

Rémy GUILLEVIN

Soutenue le 5 février 2021

Devant la Commission d'Examen

JURY

Christine Fernandez-Maloigne, Professeur, Université de Poitiers Directeur
Rémy Guillevin, Professeur, Université de Poitiers, CHU de Poitiers Co-directeur
Pascal Bourdon, Maître de conférence, Université de Poitiers Encadrant
Benoit Tremblais, Maître de conférence, Université de Poitiers Encadrant
Michèle Rombaut, Professeur, Université Grenoble Alpes Rapporteur
Su Ruan, Professeur, Université de Rouen Rapporteur
Olivier Colot, Professeur, Université de Lille Examineur
Christophe Habas, Professeur, CHNO des 15-20 Examineur

Remerciements

Après trois années de travail, voici donc venu le moment des remerciements. Exercice difficile, s'il en est, tant il est délicat. J'ai en effet à cœur d'affirmer que l'expérience scientifique est avant tout un travail collectif, et qu'une liste à la Prévert ne parviendrait pas à retranscrire l'émulation si particulière que constitue l'ambiance d'un laboratoire de recherche. C'est pourquoi, j'entame ces remerciements par une pensée pour toutes les personnes qui, de près comme de loin, m'ont accompagné ces dernières années. Vous avez contribué à mon évolution et pouvez donc voir, en ce manuscrit, une concrétisation de vos soutiens.

Merci aux membres du jury d'avoir accepté d'évaluer mon travail. Puisse-t-on nous retrouver, je l'espère, dans de futures collaborations.

En particulier, merci à Michèle Rombaut et Su Ruan d'avoir accepté d'être les rapporteuses de ce manuscrit. Vos critiques et vos retours ont permis de consolider le formalisme mathématique et le style rédactionnel de ce travail.

Merci à Olivier Colot d'avoir accepté le rôle de Président du jury.

Merci à Christophe Habas d'avoir été examinateur et membre du jury.

J'adresse de chaleureux remerciements à ma directrice de thèse, Christine Fernandez-Maloigne. Toujours disponible malgré une myriade de responsabilités, d'une énergie positive et d'une volonté certaines, je vous prie de recevoir ici mes respects, mon admiration et mon amitié.

Je remercie avec la même intensité mon co-directeur de thèse, Rémy Guillevin. Modèle de savoir, je vous sais amateur de bons mots. C'est pourquoi, vous aimez vous décrire comme *Homme sans qualités* (sic). Soyez pourtant assuré que j'ai vu en vous la précision du scientifique, l'humanisme du médecin et l'habileté du dirigeant d'équipe.

À mes encadrants, Pascal Bourdon et Benoit Tremblais, recevez ici mes remerciements nostalgiques. J'ai pu compter sur votre soutien lors des moments difficiles, quand la recherche piétine malgré la sueur et le labeur. Quand les nuits deviennent courtes alors qu'elles sont trop longues. Vous m'avez toujours fait part de vos idées, vos solutions, et votre confiance en mon travail. Merci à vous.

À Carole Guillevin, j'adresse les mêmes mots de sympathie. Merci pour ton encadrement, énergique parfois mais salutaire. Sois assurée de mon affection.

Cette thèse n'aurait pas été possible sans les équipes de Siemens Healthineers. Aussi, je remercie Matthieu Lepetit-Coiffe et Christian Bert, pour la confiance qu'ils m'ont accordée. J'adresse une pensée à Andrea Plank, Atika Boudaoud et Céline Dinis. Je remercie également Marc Lapert pour son expertise et ses conseils.

J'adresse une pensée aux permanents d'XLIM, BIATSS, MCF et Professeurs, en particulier l'inénarrable Noël Richard, Clency Perrine, Philippe Carré, Olfa Ben Ahmed, Yoann Simon, Bruno Mercier et Sophie Hardouin. Merci également à Jocelyne Attab Fazilleau, du LMA et Sylvie Duclaud, de l'Université de Poitiers. Pour l'attention dont ils font part au quotidien et leur enthousiasme à expliquer leur travail, merci aux manipulateurs IRM du CHU, CCV et 7T, et à Ludovic Blanchard.

Merci également aux radiologues qui m'ont assisté dans cette aventure et que je n'ai pas encore cités : Jean-Noël Vallée, Jean-Philippe Cottier, Claire Boutet et Benoît Gianelli.

Aux doctorants qui m'ont précédé et à ceux qui vont suivre, j'adresse les quelques mots suivants :

À Mathieu Naudin, merci pour ton écoute, tes conseils éclairés, et ton incroyable joie de vivre. Tu es un vrai soleil du labo.

À Angélique Perrillat-Mercerot, ma mathématicienne préférée, sois assurée de mon respect, de mes encouragements et de mon affection. Tu as une énergie incroyable et toute mon estime.

À mon cher co-bureau, Alexandre Fenneteau. Je te souhaite le meilleur pour ta fin de thèse. Merci d'avoir subi et même parfois partagé mes terribles jeux de mots.

À William Koch, qui sait déjà tout le bien que je pense de lui mais que je réitère à nouveau. Merci pour cette amitié sans faille depuis toutes ces années.

Je remercie également Wassim Rharbaoui, Devendra Vyas, Bastien Rimbault, les membres de la chorale universitaire de Poitiers, et les déménajoueurs.

Merci enfin à ma famille. À ma mère, à mon père, mon frère et ma sœur, pour leur soutien, surtout lors de mes longues nuits de rédaction. À ma grand-mère qu'il me tarde de revoir. Et, en particulier, à ma tante Sylvie.

Merci à tous.

Table des matières

Table des figures	0
Liste des tableaux	0
Introduction	1
1 Notions élémentaires et état de l'art	8
1 Anatomie cérébrale et types cellulaires	9
1.1 Introduction	10
1.2 Encéphale et système nerveux central	11
1.3 Zones fonctionnelles	13
1.4 Plasticité cérébrale	15
1.5 Types cellulaires composant l'encéphale	16
1.5.1 Le neurone biologique	16
1.5.2 Les cellules gliales	18
2 Le gliome, tumeur cérébrale	20
2.1 Histopathologie	21
2.2 Épidémiologie	22
2.3 Dynamique tumorale	22
2.4 Classification des gliomes et grades	23
2.5 Diagnostic	29
2.6 Suivi thérapeutique	29
3 Physique de l'IRM	32
3.1 Introduction	33
3.2 Magnétisme nucléaire	34

3.3 Résonance magnétique nucléaire	34
3.4 Relaxations	37
3.5 Séquences RMN : spin-echo	41
3.6 Reconstruction de l'image	44
3.7 Qualité du signal, qualité d'image	48
3.8 Modalités IRM et bilan lésionnel	54
3.8.1 IRM anatomique	56
3.8.2 IRM de diffusion	58
3.8.3 IRM de perfusion	63
3.8.4 Spectroscopie RMN	65
3.8.5 IRM fonctionnelle	70
4 Apprentissage automatique	73
4.1 Introduction	74
4.2 Anatomie d'un schéma d'apprentissage automatique	76
4.3 Apprentissage automatique classique	77
4.3.1 Statistiques, mesures et représentations	78
4.3.2 Types d'apprentissages	86
4.3.3 Les problèmes d'optimisation	87
4.3.4 Divisions de la base de données	90
4.3.5 Complexité du modèle : le sous-apprentissage et le sur-apprentissage	91
4.3.6 La validation croisée	92
4.3.7 Paramétrage fin : paramètres et hyperparamètres	94
4.3.8 Matrice de confusion et métriques d'évaluation d'un classifieur	95
4.3.9 Exemples d'algorithmes	96
4.4 Apprentissage profond (<i>deep learning</i>)	103
4.4.1 Analyse structuraliste des réseaux d'apprentissage profond	105

2 Méthodologie et applications 113

5 Classification du grade du gliome par apprentissage automatique sur	
données RMN : avancées récentes	115
5.1 Introduction	117
5.2 Prétraitement et cadre applicatif	118
5.2.1 Un désir de données RMN multiparamétriques	119
5.2.2 Recalage multimodal	120
5.2.3 Normalisation en intensité	120
5.2.4 Segmentation tumorale	121
5.2.5 Augmentation de données	121
5.3 Descripteurs	122
5.3.1 Descripteurs anatomiques	122
5.3.2 Descripteurs de diffusion, perfusion et spectroscopie	122
5.3.3 Algorithmes de hiérarchisation de descripteurs	123
5.4 Classifieurs	124
5.5 Synthèse	126
5.6 Mise à jour du travail présenté	127
5.7 Conclusion	128

6 Une nouvelle vérité terrain à base de critères radiologiques pour la	
base de données BraTS	131
6.1 Avant-propos	132
6.2 Introduction	134
6.3 Cadre du travail	135
6.3.1 Analyse radiomique de données IRM anatomiques	135
6.3.2 Classification du grade par apprentissage automatique	136
6.4 Classification binaire et analyse des résultats	137
6.4.1 Données utilisées	137
6.4.2 Descripteurs	138

6.4.3	Résultats et cohérence des critères radiologiques	140
6.4.4	Analyse radiomique des faux positifs et faux négatifs	149
6.5	Une nouvelle vérité terrain pour le BraTS	150
6.5.1	Obtention d'étiquettes radiologiques	150
6.5.2	Évaluation d'un classifieur SVM entraîné sur la vérité terrain ra-	
	diologique	151
6.6	Discussion et conclusion	156
7	Application d'un classifieur sur les données multimodales du CHU de	
	Poitiers	158
7.1	Introduction	159
7.2	Analyse d'imagerie multimodale et bases de données utilisées	161
7.3	État de l'art et performances actuelles des classifieurs	163
7.4	Pipeline d'apprentissage automatique	163
7.5	Classifieur	167
7.6	Analyse des performances	168
7.7	Analyse de la hiérarchisation des descripteurs	171
7.7.1	Analyse du meilleur classifieur	171
7.7.2	Analyse sur 100 entraînements successifs	179
8	Conclusions et perspectives	189
8.1	Conclusion	190
8.2	Perspectives	193
	Bibliographie	195

Table des figures

1.1	<i>Vue axiale, vue sagittale, vue coronale.</i>	11
1.2	<i>Coupe coronale de l'encéphale montrant substance grise et substance blanche.</i>	12
1.3	<i>Reconstruction 3D du système ventriculaire.</i>	12
1.4	<i>Encéphale et lobes fonctionnels.</i>	13
1.5	<i>Cas d'hydrocéphalie présentant des ventricules sur-développés.</i>	16
1.6	<i>Schémas d'un d'un neurone unipolaire et d'une synapse.</i>	18
1.7	<i>Neurones, astrocytes et oligodendrocytes.</i>	19
2.1	<i>Arbre de décision diagnostic selon l'OMS (2016).</i>	28
3.1	<i>Mouvement de simple précession du proton dans le champ \vec{B}_0.</i>	35
3.2	<i>Illustration d'une relaxation en 3D.</i>	37
3.3	<i>Relaxations T_1 et T_2.</i>	38
3.4	<i>Décroissance du signal T_2^*.</i>	40
3.5	<i>Illustration du principe de l'acquisition spin-écho.</i>	41
3.6	<i>Chronogramme représentant le bloc fondamental de la séquence spin-écho.</i>	42
3.7	<i>Représentation hybride montrant à la suite relaxation T_1 puis relaxation T_2.</i>	43
3.8	<i>Chronogramme représentant l'application des gradients de codage lors de la séquence spin-écho.</i>	45
3.9	<i>Exemples d'utilisation de l'espace de Fourier.</i>	46
3.10	<i>Illustration des séries de Fourier.</i>	47
3.11	<i>Deux images de même résolution (même nombre de lignes et de colonnes), mais de résolutions (qualité d'image) différentes pour un radiologue.</i>	50
3.12	<i>Signal réel, signal théorique, et bruit.</i>	51
3.13	<i>Exemple d'un artefact de repliement.</i>	53

3.14 <i>Bilan lésionnel du CHU de Poitiers.</i>	54
3.15 <i>Acquisitions d'IRM anatomiques.</i>	56
3.16 <i>Diffusion isotropique et anisotropique.</i>	59
3.17 <i>Cartographies d'IRM de diffusion.</i>	61
3.18 <i>Cartographies d'IRM de perfusion.</i>	63
3.19 <i>Spectre à TE court et métabolites.</i>	66
3.20 <i>Une acquisition SVS.</i>	68
3.21 <i>Une acquisition CSI.</i>	68
4.1 <i>Un schéma blocs d'apprentissage automatique.</i>	76
4.2 <i>Trois représentations et interprétations d'un même objet.</i>	78
4.3 <i>Nuage de points et modélisation gaussienne.</i>	80
4.4 <i>Illustration des notions de précision et justesse.</i>	84
4.5 <i>Résultat d'une régression linéaire et de deux régressions polynomiales (de-</i>	
<i>gré 2 et 17) sur un jeu de données.</i>	85
4.6 <i>Trois descentes de gradients sur une fonction de coût en 2D.</i>	90
4.7 <i>Un exemple de sous-apprentissage et de sur-apprentissage.</i>	92
4.8 <i>Évolution du taux d'erreur sur la base d'apprentissage et la base de test</i>	
<i>en fonction de la complexité du modèle.</i>	93
4.9 <i>Exemple de validation croisée à 4 divisions.</i>	93
4.10 <i>Exemple d'application de l'algorithme des K-moyennes avec deux centroïdes.</i>	97
4.11 <i>Exemple d'application de l'algorithme des KPPV avec K croissant.</i>	98
4.12 <i>Un arbre de décision.</i>	99
4.13 <i>Illustration du principe de vaste marge.</i>	100
4.14 <i>Problèmes non linéairement séparables.</i>	101
4.15 <i>Illustration du perceptron simple.</i>	104
4.16 <i>Réseau de neurones artificiels dense.</i>	106
4.17 <i>Trois fonctions d'activation.</i>	108
4.18 <i>Application d'un noyau convolutif sur une image d'entrée. Il en résulte</i>	
<i>une carte de description spécifique du noyau appliqué.</i>	109

4.19 Schéma de la structure GAN.	110
4.20 Schéma de la structure U-net.	111
5.1 Nombre de publications par années incluant les mots clés "glioma grade classification learning".	118
6.1 Représentation d'une ellipsoïde en 3 dimensions.	139
6.2a Histogramme des scores de prédiction de la population des LGG _{BraTS} .	148
6.2b 4 des 9 patients les plus à gauche de l'histogramme des scores de prédiction.	149
6.3 Histogramme des scores de prédiction des patients LGG _{WHO} .	153
6.4 Histogrammes des scores de prédiction obtenus sur la population des HGG _{BraTS} et sur la population des HGG _{WHO} .	155
7.1 Adaptation de la base BraTS pour établir un point de référence.	159
7.2 Un patient présentant une lésion de haut grade inhomogène en T2 FLAIR et très contrastée en T1c.	161
7.3 Illustration du principe de recalage entre deux séquences.	164
7.4 Résultats de l'extraction d'encéphale par FSL BET-Surf.	165
7.5 Validation visuelle de la segmentation du U-net à 3 séquences anatomiques.	166
7.6 Comparaison des valeurs d'ADC entre un gliome de bas grade et un gliome de haut grade.	174
7.7 Hiérarchisation des descripteurs produite par le meilleur classifieur.	175
7.8 Hiérarchisation, analyse par type de séquence RMN.	176
7.9 Hiérarchisation, analyse par type de descripteur.	177
7.10 Nombre d'apparitions des 124 descripteurs dans le sous-ensemble du SVM-RFE sur 100 itérations.	180
7.11 Comparaison de présence des 23 descripteurs les plus fréquents sur 100 itérations et des 23 descripteurs sélectionnés par le meilleur classifieur.	181
7.12 Classement par type de modalité RMN (Imagerie anatomique, de diffusion, de perfusion, et SRM) – 23 descripteurs les plus récurrents.	183

7.13 Classement par type de descripteur (forme, intensité, texture, mesures manuelles de SRM, rCBV et rCBF) – 23 descripteurs les plus récurrents.	183
7.14 Classement par type de modalité RMN (imagerie anatomique, de diffusion, de perfusion, et SRM) – 23 descripteurs les moins fréquents. 185
7.15 Classement par type de descripteur (forme, intensité, texture, mesures manuelles de SRM, rCBV et rCBF) – 23 descripteurs les moins fréquents.	185

Liste des tableaux

2.1	<i>Classification des gliomes selon l'OMS (WHO 2000).</i>	26
3.1	<i>Valeurs et représentations des composantes longitudinales et transversales selon l'impulsion utilisée.</i>	37
4.1	<i>Matrice de confusion dans un cas de classification binaire.</i>	95
4.2	<i>Plusieurs métriques issues de la matrice de confusion.</i>	96
4.3	<i>Trois fonctions d'activation fréquemment utilisées.</i>	107
5.1	<i>Récapitulatif des différentes publications traitées</i>	130
6.1	<i>Grades des gliomes selon l'OMS (WHO) et la base de données BraTS.</i>	132
6.2	<i>Descripteurs de forme utilisés.</i>	141
6.3	<i>Descripteurs d'intensité utilisés.</i>	142
6.4	<i>Descripteurs de texture utilisés.</i>	143
6.5	<i>Performances du classifieur sur la vérité terrain du BraTS.</i>	144
6.6	<i>Sigles utilisés dans la matrice de confusion.</i>	146
6.7	<i>Analyse des votes des radiologues en vue d'une construction de vérité terrain radiologique.</i>	151
6.8	<i>Comparaison des performances entre les différentes vérités terrains.</i>	152
7.1	<i>Tableau des résultats.</i>	170

Glossaire

2D, 3D	2 dimensions, 3 dimensions
ACP	Analyse en composantes principales
ADC	Coefficient de diffusion anisotropique
ADN	Acide désoxyribo-nucléique
AIF	Fonction d'entrée artérielle
AIR	<i>Automatic Image Registration software</i>
ASL	Labellisation de spin artériel
ATRX	<i>Transcriptional regulator ATRX</i>
BAcc	<i>Balanced accuracy</i>
BraTS	<i>Brain Tumor Segmentation Challenge</i>
BOLD	<i>Blood Oxygen-Level Dependant</i>
CBF	Flux sanguin cérébral
CBV	Volume sanguin cérébral
Cho	Choline
CHU	Centre Hospitalier Universitaire
CIFRE	Convention Industrielle de Formation par le REcherche
CNN	Réseaux neuronaux convolutifs
CNRS	Centre National de la Recherche Scientifique
CoIFA	Anisotropie fractionnelle colorée
Cr	Créatine
CSI	<i>Chemical shift imaging</i>
DACTIM-MIS	<i>Data Analysis and Computation Through Imaging and Modeling - Maths Images Santé</i>
DHG	Désoxyhémoglobine
DSC	Contraste dynamique de susceptibilité
DTI	<i>Diffusion Tensor Imaging</i>
EPI-SE	<i>Echo Planar Imaging - Spin Echo</i>
FA	Anisotropie fractionnelle
FID	<i>Free induction decay</i>
FLAIR	<i>FLuid Attenuated Inversion Recuperation</i>
FN, FP	Faux négatif, faux positif
FSL	<i>FMRIB Software Library</i>
GABA	Acide γ -AminoButyrique
GAN	Réseaux antagonistes génératifs

GBM	Glioblastome multiforme
GDBG	Gliome diffus de bas grade
GLCM	<i>Gray Level Co-occurrence Matrix</i>
Gln	Glutamine
GLRLM	<i>Gray Level Run Length Matrix</i>
Glu	Glutamate
Glx	Complexe glutaminergique
HGG	<i>High Grade Glioma</i> , gliome de haut grade
HGG_{BraTS}	Gliomes de hauts grades de la base BraTS, grades IV uniquement
HGG_{WHO}	Gliomes de hauts grades selon l'OMS, grades III et IV
I3M	Imagerie Métabolique Multi-noyaux Multi-organes
IA	Intelligence artificielle
ICABME	<i>International Conference on Advances in Biomedical Engineering</i>
ICONES	Image, COuleur, mouvemeNts, rElief, Surface
IDH	Isocitrate Dehydrogénase
IPTA	<i>International Conference on Image Processing, Tools and Applications</i>
IRBM	<i>Innovation and Research in BioMedical engineering</i>
IRM	Imagerie par résonance magnétique
IRMf	Imagerie par résonance magnétique fonctionnelle
ISBI	<i>Imaging Biomarker Standardization Initiative</i>
ISZ	<i>Intensity Size Zone</i>
ITK	<i>Image Tool Kit</i>
KPPV	K plus proches voisins
LabCom	Laboratoire commun
Lac	Lactate
Lac35, Lac135	Lactate à TE court, Lactate à TE long
LCR	Liquide céphalo-rachidien
LGG	<i>Low Grade Glioma</i> , gliomes de bas grade
LGG_{BraTS}	<i>Lower grade gliomas</i> de la base BraTS, regroupant grades I, II et III
LGG_{WHO}	Gliomes de bas grades selon l'OMS, grades I et II
Lip	Lipides
LMA	Laboratoire de Mathématiques et Applications
LSTM	Long short-term memory
MD	<i>Mean Diffusivity</i>
mI	Myo-inositol
MIRES	Mathématiques et leurs Interactions, Images et information numérique, Réseaux et Sécurité

mRMR	<i>minimum Redundancy Maximum Relevance</i>
MTT	Temps de transit moyen
NAA	N-Acetyl Aspartate
OMS	Organisation Mondiale de la Santé
pH	Potentiel hydrogène
rCBF, RCBF	CBF relatif
rCBV, RCBV	CBV relatif
RFE	<i>Recursive feature elimination</i>
ReLu	<i>Rectified Linear Unit</i>
RF (IRM)	Radio-fréquences
RF (IA)	Forêts d'arbres aléatoires
RGB	Canaux couleurs Rouge, Vert, Bleu
RNN	Réseau de neurones récurrents
RMN	Résonance magnétique nucléaire
ROC	Receiver operating characteristic
ROI	Région d'intérêt
SMO	<i>Sequential Minimal Optimization</i>
SNC	Système nerveux central
SNP	Système nerveux périphérique
SRM	Spectroscopie par résonance magnétique
STIR	<i>Short TI Inversion Recovery</i>
SVM	Machines à vecteur de support
SVS	<i>Single voxel spectroscopy</i>
SWI	<i>Susceptibility Weigthed Imaging</i>
T1c	T1 <i>contrast-enhanced</i> , T1 post injection de gadolinium
TCIA	<i>The Cancer Imaging Archive</i>
TE	Temps d'écho
TI	Temps d'inversion
TN, TP	Vrai négatif, vrai positif
TNR	Spécificité, taux de vrais négatifs
TP53	<i>Tumor Protein 53</i>
TPR	Sensibilité, taux de vrais positifs
TR	Temps de répétition
TTP	Temps de pic
UMR	Unité Mixte de Recherche
VT	Vérité terrain
VT_radiologues	Vérité terrain établie par 5 experts radiologues
VT_BraTS	Vérité terrain de la base BraTS
VGG-16, VGG-19	<i>Visual Geometry Group</i>
WHO	<i>World Health Organisation</i>

A man wanted to know about mind, not in nature, but in his private,
large computer. He asked it (no doubt in his best Fortran),

"Do you compute that you will ever think like a human being ?"

The machine is then set to work and analyze its own computational habits.

Finally, the machine printed its answer on a piece of paper,
as such machines do.

The man ran to get the answer and found, neatly typed, the words :

THAT REMINDS ME OF A STORY .

Gregory Bateson, *Mind and Nature*

Introduction

Contexte et problématique

Avec 8,8 millions de morts dans le monde en 2015, le cancer est la seconde cause de mortalité selon l'Organisation Mondiale de la Santé (OMS) [1]. En France, pour l'année 2018, un rapport de Santé Publique France estime à presque 5900 le nombre de nouveaux cas de tumeurs malignes du système nerveux central (SNC) [31]. Sur la même année, le glioblastome avec confirmation histologique représente presque 60% des tumeurs du SNC, ce qui fait de cette tumeur le type le plus fréquent. Associée à une forte mortalité et à une courte durée de survie, son incidence est en constante augmentation depuis 1990, pour les deux sexes, bien qu'une prédominance soit observée chez l'homme.

Le glioblastome est le stade d'évolution ultime d'une famille de tumeurs, les gliomes, qui se développent à partir du tissu de soutien cérébral : la glie. L'évolution du gliome présente plusieurs phases, associées à des dynamiques de développement tumoral distinctes. Ainsi, l'OMS propose de classer les gliomes selon un critère, le grade, représentant l'agressivité croissante de ces tumeurs. Sont ainsi distingués d'une part les gliomes de bas grades, grades I et II; et d'autre part les gliomes de hauts grades, grades III et IV. Le gliome de bas grade présente un développement lent, parfois silencieux, doublé d'un caractère diffus et infiltrant. Il se propage au sein du tissu de soutien cérébral, contaminant ainsi les régions saines environnantes. Le gliome se développe jusqu'à ce que les tissus envahis ne puissent plus assurer leurs fonctions biologiques. Surviennent alors phénomènes épileptiques, céphalées ou pertes fonctionnelles.

Les parcours de soin envisagés diffèrent selon différents critères, comme le grade observé, la localisation de la tumeur, l'étendue de la zone d'infiltration, ou encore l'âge du patient. Radiothérapie, chimiothérapie et chirurgie représentent les voies d'intervention principales. La chirurgie, avec des résections totales voire supra-totales, reste une voie privilégiée pour améliorer la survie à long terme et la qualité de vie.

L'intervention chirurgicale nécessite un bilan d'imagerie riche permettant de caractériser au mieux l'état des tissus et la dynamique tumorale. Depuis 2012, le Centre Hospitalier Universitaire de Poitiers (CHU) est entré dans une démarche de modernisation radicale de ses techniques d'imagerie, notamment l'imagerie par résonance magnétique (IRM), dans un but de diagnostic, d'intervention clinique et de recherche. Cette démarche comprends l'amplification des capacités d'imagerie avec l'acquisition de nouveaux imageurs de pointe, et la diversification des informations recueillies lors d'un examen. La démarche peut se résumer en termes simples : "voir plus, voir mieux".

Voir plus, c'est mettre en place un bilan lésionnel complet pour caractériser au mieux la tumeur. Ce bilan regroupe plusieurs modalités IRM. Sont ainsi acquises des données anatomiques, mais aussi des informations diffusionnelles, perfusionnelles et fonctionnelles. De surcroît, ce bilan est complété par les informations métaboliques issues de la spectroscopie par résonance magnétique (SRM) multinoyaux. L'ensemble de ces modalités offre un niveau de description rare.

En 2017, le CHU de Poitiers se dote du premier imageur à haut champ (Imageur Siemens Skyra 3 Tesla) permettant d'effectuer des examens per-opératoire. L'imageur, installé à côté de la salle d'intervention, permet d'offrir au neuro-chirurgien des images sur la lésion et de décider, le cas échéant, de poursuivre l'opération. Grâce à ces images, le neuro-chirurgien et le neuro-radiologue peuvent travailler de concert pour améliorer la qualité de l'intervention chirurgicale. C'est dans ce contexte qu'une première thèse portant sur l'aide à la décision pour la neuronavigation per-opératoire a été défendue par Mathieu Naudin en décembre 2018 [\[82\]](#). Ce travail s'est effectué dans le cadre d'une première collaboration entre le laboratoire XLIM, le CHU de Poitiers et Siemens Healthineers via un financement de type Convention Industrielle de Formation par la REcherche (CIFRE).

Poursuivant sur cette lancée, le CHU de Poitiers acquiert en 2019 un imageur ultra-haut champ (Imageur Siemens Terra 7 Tesla). Cet imageur est le premier en France à bénéficier d'une accréditation à la fois pour la recherche et pour l'activité clinique. L'acquisition ultra haut champ permet de suivre de nouveaux métabolites en SRM tout en offrant une meilleure qualité d'image IRM. Voir plus, voir mieux.

Ces nouvelles acquisitions produisent à chaque examen de grandes quantités de données, ce qui soulève la question de l'automatisation de leur traitement. De nos jours, l'intelligence artificielle (IA) offre des possibilités de traitement sur de grandes bases de données. C'est pourquoi le champ de l'IA appliquée à la Santé est en plein développement. Segmentation automatique, caractérisation tumorale ou aide au diagnostic, ces applications demandent une expertise importante en traitement du signal et des images.

C'est ainsi qu'une collaboration de recherche s'est développée entre la société SIEMENS Healthineers, le CHU et l'Université de Poitiers, avec 2 équipes CNRS des laboratoires XLIM (UMR CNRS 7252) et LMA (UMR CNRS 7348), au sein de la fédération régionale de recherche en Mathématiques et Sciences des données MIRES (FR CNRS 3423).

Cette collaboration a débouché en février 2019 sur la création d'un Laboratoire Commun CNRS entreprise, dit LabCom I3M (Imagerie Métabolique Multi-noyaux Multi-organes) au sein duquel la présente thèse s'est déroulée. Tout comme pour le travail de Mathieu Naudin, cette thèse a bénéficié d'un financement de type CIFRE de Siemens Healthineers.

Son inscription administrative a été effectuée au sein du laboratoire XLIM, dont l'équipe ICONES (Image, COuleur, mouvemeNts, rElief, Surface) est spécialisée dans l'analyse des images et données multivariées. Les membres de cette équipe sont familiers des algorithmes d'apprentissage automatique, et portent depuis plusieurs années un intérêt grandissant aux problématiques d'imagerie médicale. Ils collaborent avec le LMA, via l'équipe DACTIM-MIS composée de membres des personnels du CHU de Poitiers et de mathématiciens, spécialisée dans la réalisation de modèles biostatistiques, trouvant de nombreuses applications pour la modélisation du métabolisme du cerveau, notamment à travers l'analyse de la dynamique tumorale.

C'est donc dans ce cadre que le travail proposé explore comment les données multimodales issues de l'IRM peuvent être traitées puis analysées par intelligence artificielle dans un but d'aide au diagnostic.

Pour cela, nous proposons des solutions techniques permettant de différencier auto-

matiquement un gliome de bas grade d'un gliome de haut grade. Puis, nous explorons via des techniques novatrices d'analyse de données comment les informations contenues dans l'image sont hiérarchisées par l'intelligence artificielle. Partant de cette hiérarchie d'informations, nous montrons en quoi l'analyse par intelligence artificielle est complémentaire de l'analyse d'imagerie multimodale du radiologue. Cette complémentarité ouvre de nombreuses perspectives dans le domaine de l'aide au diagnostic assisté par intelligence artificielle.

Principales contributions

Cette thèse a fait l'objet de deux publications pour des conférences internationales à comité de relecture. À l'heure de la rédaction de ce manuscrit, deux autres publications sont en cours d'écriture. Le travail de cette thèse a été présenté lors de deux conférences internationales et une conférence nationale :

- Le premier article, *Recent advances glioma grade classification using machine and deep learning* [32], a été présenté à la cinquième *International Conference on Advances in Biomedical Engineering* (ICABME) en 2019 ;
- Le deuxième article, *Assigning a new glioma grade label ground-truth for the BraTS dataset using radiologic criteria* [33], a été présenté à la dixième édition de la *International Conference on Image Processing Theory, Tools and Applications* (IPTA) en 2020 ;
- Une communication sans publication a été donnée lors de la conférence nationale *Interplay between Oncology, Mathematics and Numerics : focus on pre-treatment studies* (IbOMaN) en 2020.

Citons également la parution d'un article de popularisation des sciences, *Des données biologiques aux modèles et inversement*, co-écrit avec Angélique Perrillat-Mercerot, sur le site Images de Mathématiques du CNRS en 2019 [34] ; et l'obtention du 5ème prix de l'Université de Poitiers lors du concours *Ma thèse en 180 secondes* en 2019.

Les principaux résultats de ce travail de thèse sont :

- Une revue des méthodes récentes utilisées dans le champ de la classification binaire et automatique du grade du gliome. Nous y présentons les classifieurs implémentés, les pré-traitements effectués et les bases de données utilisées lors des travaux récents ;
- Un outil d'exploration de données et d'analyse de performances de classifieurs. Cet outil est indépendant des métriques issues de la matrice de confusion et permet d'identifier des sous-groupes d'individus similaires dans un jeu de données ;
- Une remise en question de la base de données la plus utilisée pour la classification binaire du grade du gliome, qui n'est pas adaptée à cette tâche selon les standards de l'OMS ;
- Un étiquetage alternatif de cette base via un consensus d'experts, dans la limite de l'IRM anatomique ;
- Une chaîne de traitement complète, permettant d'automatiser le traitement des données multimodales du CHU de Poitiers, de l'acquisition à la classification automatique ;
- Une étude de l'impact de l'ajout des données de diffusion, de perfusion et de spectroscopie sur les performances de classification ;
- Enfin, deux méthodes d'analyse permettant d'obtenir une hiérarchisation de descripteurs, et qui montrent comment cette hiérarchie renforce et complète l'analyse visuelle du radiologue.

Organisation du manuscrit

Le contenu de ce manuscrit s'articule autour de deux parties, chacune composée de quatre chapitres.

Dans la première partie, nous présentons les éléments nécessaires à la compréhension du travail effectué. Ainsi, le premier chapitre aborde des notions relatives à l'organisation du cerveau sain et à son fonctionnement. Nous commençons par des notions d'anatomie générales, et finissons par une description des différents types cellulaires composant l'encéphale, notamment les cellules gliales. Suite à cette présentation, un deuxième chapitre détaille l'histopathologie du gliome, la dynamique tumorale et les stratégies mises en place pour le suivi thérapeutique, parmi lesquelles l'imagerie par résonance magnétique tient un grand rôle. Cela nous amène à notre troisième chapitre qui décrit comment l'IRM peut être utilisée comme outil d'exploration multimodal par la production d'images multivariées. Enfin, un quatrième chapitre revient sur les notions essentielles à l'application d'algorithmes d'apprentissage automatique. Nous couvrons ainsi plusieurs éléments relatifs aux statistiques descriptives, avant de détailler plusieurs méthodes d'apprentissage, qu'il s'agisse d'apprentissage classique ou d'apprentissage profond.

La deuxième partie de ce manuscrit présente le travail accompli lors des trois années de thèse. Nous revenons ainsi sur les différentes contributions et nos propositions techniques à travers quatre chapitres. Ainsi, le cinquième chapitre de cette thèse présente une analyse des travaux récents concernant la tâche de classification du grade du gliome. Nous montrons en quoi cette tâche particulière est devenu un champ scientifique actif au cours des dernières années à travers une revue de littérature récente ayant donné cours à une publication en 2019 [32]. Dans le chapitre suivant, nous entraînons un premier classifieur sur la base de données la plus utilisée dans la littérature récente. Nous proposons une analyse du comportement du classifieur via l'utilisation de l'histogramme des scores de prédiction. Cette analyse nous amène à pointer des erreurs au sein de cette base de données, via des labels de grade inadaptés. Nous proposons une rectification de ces labels suite à un travail collaboratif avec plusieurs radiologues experts, et montrons le bon fonctionnement de notre classifieur sur ces labels rectifiés. Ce travail a donné lieu à une seconde publication en 2020 [33]. Au cours du septième chapitre, nous présentons un schéma complet d'apprentissage automatique permettant d'utiliser les données multimodales du CHU de Poitiers. Nous entraînons à nouveau un classifieur et fournissons un point de référence permettant d'évaluer le gain de performance issu de l'ajout

des informations multimodales, comparé à la simple imagerie anatomique. Nous proposons ensuite deux méthodes permettant d'obtenir une hiérarchisation informationnelle des descripteurs images. Riche de ces hiérarchisations, nous terminons par un retour à l'imagerie dans le but d'enrichir l'analyse du radiologue. Un huitième chapitre fait office de conclusion sur le travail, en offrant différentes perspectives et applications des outils et méthodes présentés.

1

PARTIE

Notions élémentaires et état de l'art

Anatomie cérébrale et types cellulaires

Contenu du chapitre

1.1 Introduction	10
1.2 Encéphale et système nerveux central	11
1.3 Zones fonctionnelles	13
1.4 Plasticité cérébrale	15
1.5 Types cellulaires composant l'encéphale	16
1.5.1 Le neurone biologique	16
1.5.2 Les cellules gliales	18

Avant d’aborder la description des modifications cérébrales engendrées par le gliome, nous proposons de décrire l’organisation et le fonctionnement du cerveau sain. Cette partie s’organise à travers différents niveaux de lecture, commençant par une description anatomique générale du cerveau, qui nous amènera à évoquer les principaux lobes et zones fonctionnelles composant l’encéphale. Nous amènerons la vision fonctionnelle existante en détaillant le phénomène de plasticité cérébrale, phénomène que nous retrouverons impliqué dans la dynamique tumorale dans le chapitre suivant. Nous terminerons ce chapitre par la description des différents types cellulaires composant l’encéphale, en portant une attention toute particulière aux différentes cellules gliales.

1.1 Introduction

Au cours de l’histoire, le cerveau a fait l’objet de nombreuses représentations. Ces dernières portaient aussi bien sur sa description anatomique que sur son fonctionnement ou son rôle dans l’organisme [92]. Le cerveau a ainsi été appréhendé comme un régulateur de flux par Aristote, puis comme le siège de l’esprit dans la pensée dualiste de Descartes. Bien que son rôle sensori-moteur ait été appréhendé par Galien dès le II^{ème} siècle, une compréhension plus fine n’émergera qu’à la Renaissance, grâce aux dissections. Mais, il faudra attendre les techniques d’imagerie modernes pour pouvoir observer le cerveau et son fonctionnement *in vivo*, notamment grâce à l’imagerie par résonance magnétique fonctionnelle (IRMf). Dans cette partie, nous verrons, via une description anatomique et fonctionnelle du cerveau, comment cet organe est aujourd’hui devenu un objet dynamique en constante adaptation. Pour ce faire, nous allons détailler les principales structures anatomiques, puis les grandes zones fonctionnelles. Nous finirons par une description du neurone comme unité fonctionnelle de l’encéphale. Cette description sera réutilisée lors de l’explication des algorithmes d’apprentissage automatique profond.

1.2 Encéphale et système nerveux central

Le corps humain est régi par deux systèmes d'information : le système hormonal et le système nerveux. Ce dernier est divisé en deux parties : le système nerveux central (SNC) et le système nerveux périphérique (SNP). Le SNC comprend la moelle spinale et l'encéphale, terme que l'on préférera au nom générique "cerveau". L'encéphale est le principal organe du SNC. Il se décompose en télencéphale (ou hémisphères cérébraux), diencéphale, tronc cérébral et cervelet. Les deux hémisphères cérébraux sont reliés par le corps calleux et sont séparés au-dessus par la faux du cerveau, une membrane verticale. Ils sont également séparés du cervelet par une membrane horizontale appelée tente du cervelet, délimitant l'espace supra-tentorial et l'espace infra-tentorial. L'exploration de l'encéphale se fait selon trois axes de coupe équivalents aux trois plans de l'espace (Figure 1.1).

L'encéphale est composé de deux types de cellules : les neurones et les cellules gliales. Une coupe de l'encéphale révèle deux types de tissus : la substance grise et la substance blanche (Figure 1.2). La substance grise, ou cortex, est principalement composée des corps cellulaires des neurones, tandis que la substance blanche est composée des axones neuronaux, regroupés parmi les principaux faisceaux de fibres blanches, dont fait partie le corps calleux.

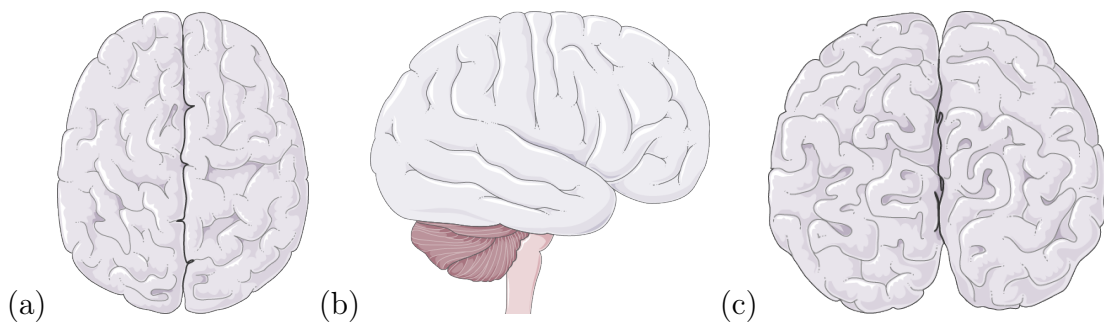


FIGURE 1.1 : *Les trois axes de vue de l'encéphale : la vue axiale (a), la vue sagittale (b) et la vue coronale (c) (Servier Medical Art - CC BY 3.0)*

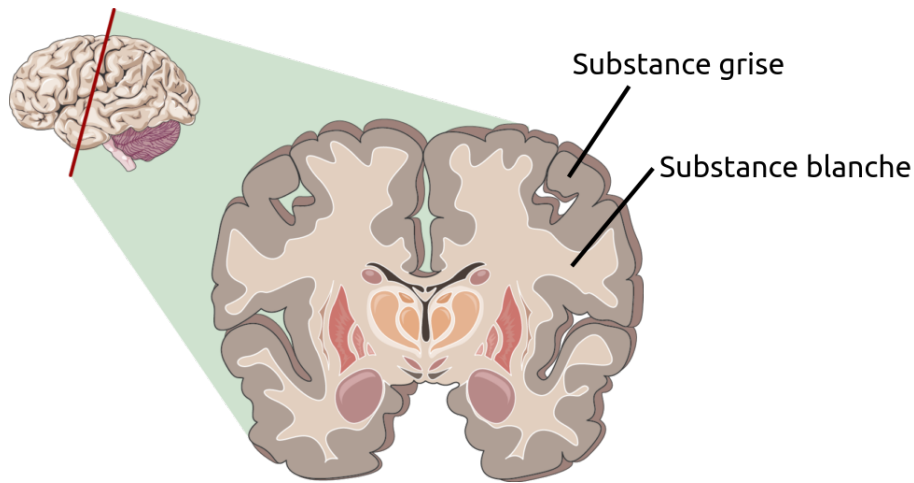


FIGURE 1.2 : Coupe coronale de l'encéphale montrant substance grise et substance blanche (Servier Medical Art - CC BY 3.0)

Au sein du cerveau, quatre ventricules reliés entre eux forment le système ventriculaire, représenté Figure 1.3. Les ventricules sont recouverts de cellules gliales ciliées qui sécrètent et mettent en mouvement le liquide céphalo-rachidien (LCR). Principalement composé d'eau, ce liquide, dans lequel baigne le cerveau et la moelle, permet une protection mécanique en absorbant les mouvements et les chocs. C'est également par lui que sont évacués les déchets du cerveau. Il joue aussi un rôle immunologique, et permet le transport de nutriments et d'hormones.

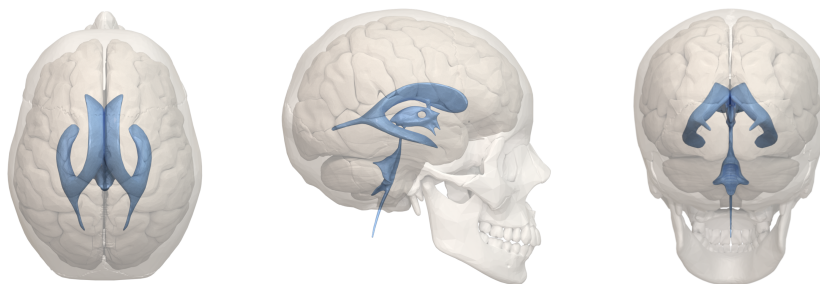


FIGURE 1.3 : Reconstruction 3D du système ventriculaire (en bleu), vu selon les trois axes. (BodyParts3D by DBCLS - CC BY 2.1 Japan)

Le cerveau est un organe fortement vascularisé. Cette vascularisation répond aux besoins importants en oxygène et autres métabolites nécessaires au bon fonctionnement

cérébral. Les voies d'apport comprennent en avant les deux artères carotides internes et en arrière le tronc basilaire. A la base de l'encéphale, une structure d'anastomoses, c'est-à-dire d'interconnexions, appelée polygone de Willis regroupe ces voies d'apport avant de donner en sortie les différentes artères cérébrales. Une part des artères cérébrales forment le trajet basal, qui permet la vascularisation des structures cérébrales profondes. Une autre partie des artères cérébrales forment le trajet périphérique, qui peut jouer un rôle de suppléance en cas de chute du débit sanguin cérébral.

La vascularisation permet l'apport de métabolites. Ceux-ci ne peuvent pénétrer les tissus qu'en franchissant la barrière hémato-encéphalique. Cette barrière, qui tapisse les capillaires, constitue un filtre qui protège le cerveau des agents pathogènes tout en permettant l'homéostasie cérébrale. Une atteinte de cette barrière, comme par exemple une lésion, provoque des risques importants pour le patient.

1.3 Zones fonctionnelles

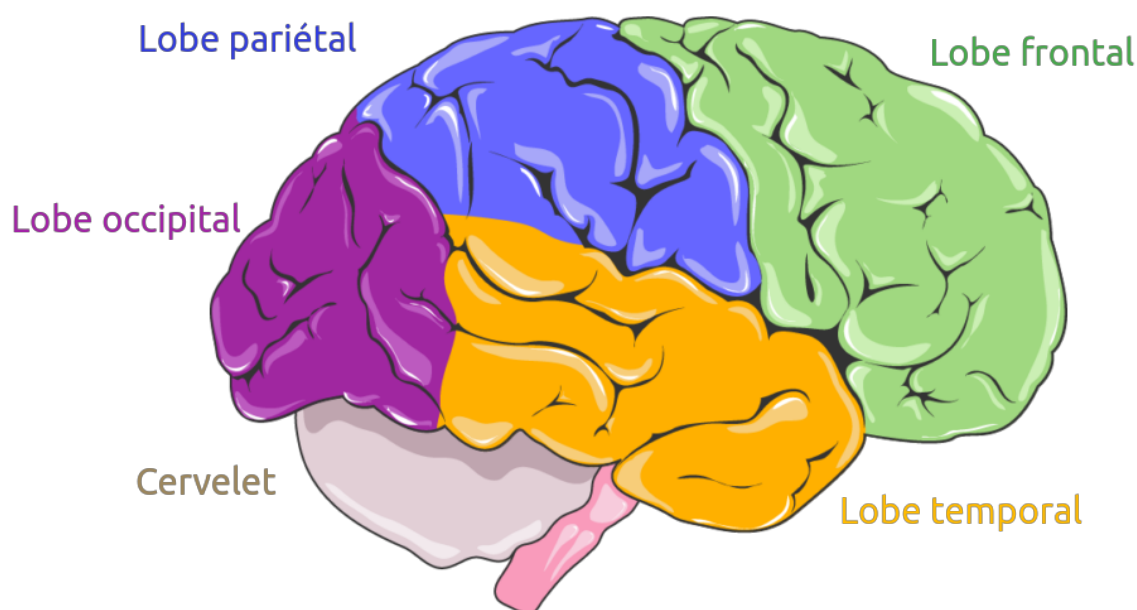


FIGURE 1.4 : *Encéphale et lobes fonctionnels.* (Servier Medical Art - CC BY 3.0)

L'idée selon laquelle le cerveau serait organisé en zones fonctionnelles est à manipuler avec précaution. Cette intuition remonte à la fin du XVIIIème siècle avec la phrénologie de Frank Joseph Gall. La phrénologie explique que le cerveau est organisé en unités fonctionnelles - ce qui est une idée juste. Mais elle postule également que le développement de ces zones provoque des bosses mesurables sur la surface du crâne. Les applications furent houleuses, certains y cherchant une justification des thèses racistes. Mais la mesure d'aspérités sur le crâne n'offrant aucune corrélation avec un développement du cortex sous-jacent, la pratique phrénologique disparut progressivement. Néanmoins, certaines expressions comme "avoir la bosse des maths" continuent d'exister, comme traces de cette errance pseudoscientifique.

Il faut reconnaître cependant que la compréhension du cerveau comme un ensemble de régions fonctionnelles offre un premier niveau de compréhension de l'organisation cérébrale. Dans son livre *L'erreur de Descartes*, Antonio Damasio utilise l'exemple historique de Phineas Gage, dont une partie du cerveau vint à manquer suite à un accident, pour illustrer le fonctionnement en régions de l'encéphale. Cependant, face à cette simplification, l'auteur insiste sur l'importance de comprendre non seulement les systèmes qui composent ces régions, mais aussi leurs interactions [28].

Ainsi, nous allons décrire la surface des deux hémisphères de l'encéphale comme divisés en quatre lobes chacun, représentés Figure 1.4. À chaque lobe est attribué un ensemble de fonctions (avec la distance que l'on vient d'instaurer). Ainsi, on distingue :

- **Le lobe frontal** : Limité en arrière par le sillon central. Il intervient dans la planification, le langage et le mouvement volontaire ;
- **Le lobe pariétal** : Situé en arrière du lobe frontal, on lui attribue un rôle dans la perception de l'espace ou les mécanismes de l'attention, comme les saccades oculaires ;
- **Le lobe temporal** : Situé au niveaux des tempes, il est séparé du lobe frontal et pariétal par le sillon latéral. Il représente la partie inférieure et latérale du cerveau.

Il remplit plusieurs fonctions cognitives, dont l'audition, le langage, la mémoire et la vision des formes complexes ;

- **Le lobe occipital** : C'est le lobe le plus en arrière du cerveau. Il joue un rôle dans la vision, notamment par la reconnaissance de l'orientation et des contours des images. Le cortex consacré à l'analyse visuelle se prolonge jusqu'aux lobes pariétaux et temporaux.

Le cerveau comprend aussi deux lobes internes :

- **Le lobe limbique** : Identifié par Broca en 1878, sa délimitation exacte est ambiguë. Concernant ses fonctions, on lui associe les émotions, la mémoire ou encore un rôle dans l'odorat ;
- **Le cortex insulaire ou insula** : Il est situé au fond du sillon latéral, et serait impliqué dans la conscience du soi, l'homéostasie, le contrôle moteur ou les émotions sociales.

1.4 Plasticité cérébrale

La notion de zone fonctionnelle est aussi à nuancer par le phénomène de plasticité cérébrale. En effet, des études de cas ont montré que le cerveau était un objet adaptatif, sensible aux effets d'entraînement. Ce phénomène est important pour comprendre la compensation qui peut exister d'un hémisphère sur l'autre lors d'un développement tumoral. Le cerveau est capable d'adaptation. Ce phénomène a été particulièrement mis en évidence en étudiant l'hippocampe, zone liée à la mémoire, chez des conducteurs de taxi londoniens [71]. Comparés à un groupe contrôle, les conducteurs de taxi, qui mobilisent beaucoup leur mémoire spatiale, présentaient un développement plus important de cette zone que le groupe contrôle. Cette plasticité cérébrale intervient également dans des situations pathologiques. Ainsi, en 2007, le cas d'un patient atteint d'une hydrocéphalie sévère a montré des ventricules sur-développés, et une bande de cortex résiduelle en

périphérie du crâne [40] (voir Figure 1.5). Malgré ce cerveau fortement impacté et déformé, le patient ne présentait qu'un fonctionnement faiblement dégradé. Ce cas illustre à l'extrême les capacités de la plasticité cérébrale.

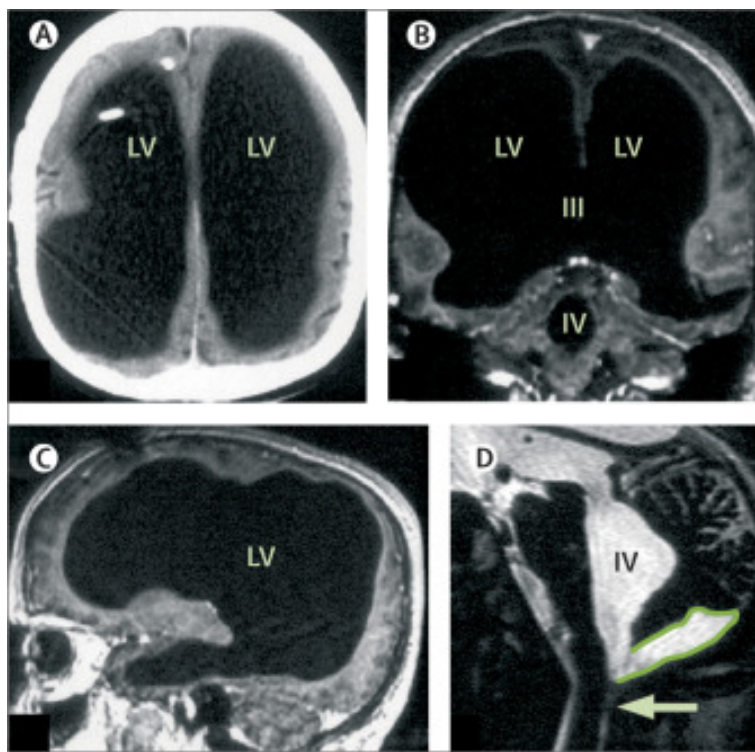


FIGURE 1.5 : Cas d'hydrocéphalie présentant des ventricules sur-développés. Image reproduite depuis l'article original de Feuillet et al. [40].

1.5 Types cellulaires composant l'encéphale

Le cerveau sain est composé de deux types de cellules : les neurones et les cellules gliales. Nous allons présenter successivement une description anatomique et fonctionnelle du neurone, puis les différentes cellules gliales.

1.5.1 Le neurone biologique

Le cerveau sain est composé de deux types de cellules : les neurones et les cellules gliales. Un cerveau adulte pèse environ 1300 grammes [102] et contiendrait environ 100 milliards

de neurones et 10 fois plus de cellules gliales, bien que ces chiffres se rapprocheraient plus du mythe que d'une estimation fiable selon une revue de littérature récente [19].

Un neurone, comme celui représenté sur la Figure 1.6.a, est une cellule excitable composé d'un corps cellulaire appelé soma, d'un axone et de dendrites.

Le soma contient le noyau et les différents organites nécessaires à la machinerie cellulaire. C'est dans le corps cellulaire que seront synthétisées les molécules nécessaires à la communication entre neurones : les neurotransmetteurs.

Les dendrites correspondent à la partie réceptrice de la cellule. C'est ainsi qu'un neurone est connecté à son environnement neuronal et va être capable de recevoir puis transmettre une excitation.

L'axone est une structure au sein de laquelle se propage l'influx nerveux : on parle de potentiel d'action. L'axone est entouré d'une gaine de myéline qui accélère la transmission de l'influx nerveux. Aux terminaisons de l'axone se trouvent les synapses : ce sont les interfaces entre les neurones. Une synapse est représentée Figure 1.6.b. À travers la synapse, une terminaison axonale se connecte aux dendrites d'un autre neurone.

Le déclenchement du potentiel d'action s'effectue lorsque les dendrites captent un signal au travers de leurs synapses. Si ces signaux sont suffisamment forts, à partir d'un certain seuil, le neurone va lui aussi s'exciter et émettre un potentiel d'action le long de son axone. Le potentiel d'action est un message électrique issu d'un gradient électrochimique. Il est la conséquence d'une différence de concentration entre des ions sodium et potassium et se propage tout le long de l'axone. L'arrivée du potentiel d'action à la synapse provoque la sécrétion de neurotransmetteurs. Ces derniers vont servir de support de transmission de l'information aux autres neurones. Ainsi, l'information va se transmettre de neurone en neurone par excitations successives.

Ces informations seront réutilisées lors de la présentation des neurones artificiels, structure informatique mise en œuvre lors de l'apprentissage automatique profond.

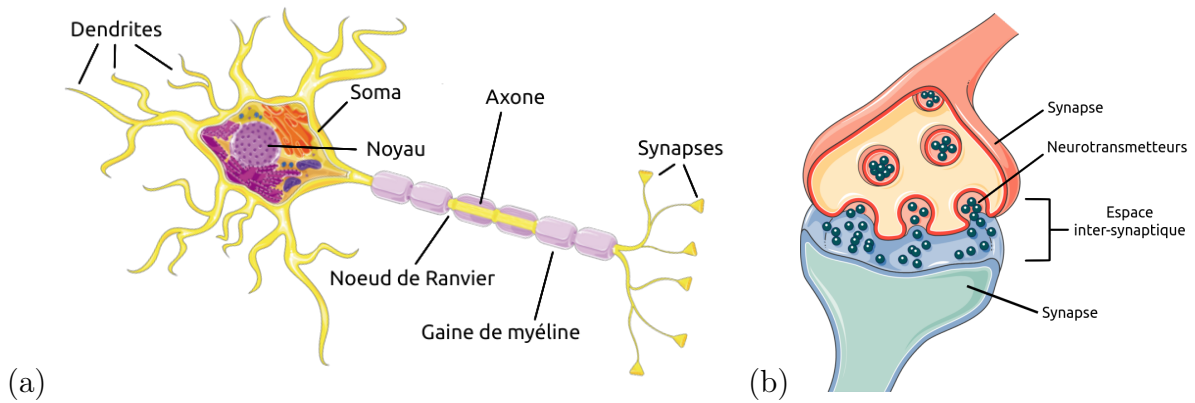


FIGURE 1.6 : Schémas : (a) d'un neurone unipolaire, (b) d'une synapse (Servier Medical Art - CC BY 3.0)

1.5.2 Les cellules gliales

L'ensemble des cellules gliales forme la glie, tissu de soutien cérébral. Parmi les cellules gliales, on trouve les astrocytes, les oligodendrocytes, les épendymocytes et la microglie. La microglie est composée de macrophages et assure un rôle dans la réponse immunitaire du système nerveux central. Hors de ces cellules, on trouve la matrice extracellulaire et le liquide interstitiel.

Les astrocytes sont des larges cellules et tirent leur nom de leur forme étoilée. Ils ont un rôle de soutien biochimique des cellules endothéliales formant la barrière hémato-encéphalique. De par leur proximité avec le système vasculaire, ils permettent aussi l'évacuation des déchets et l'apport d'oxygène et de glucose du sang vers les cellules. Ils forment 20 à 40% des cellules gliales.

Les oligodendrocytes ont un rôle de protection des neurones. En effet, ils s'enroulent autour des axones, formant une couche lipidique appelée gaine de myéline. La gaine de myéline permet aussi une transmission plus rapide des influx nerveux. Tout au long de l'axone, la gaine de myéline s'amincit parfois pour former une structure appelée noeud de Ranvier. Grâce à ces structures, lors du passage de l'influx nerveux, le potentiel d'action saute de noeud de Ranvier en noeud de Ranvier, ce qui accélère la transmission du signal.

Les épendymocytes tapissent les parois des ventricules. Ce sont des cellules ciliées qui produisent et brassent le LCR.

Les astrocytes et les oligodendrocytes peuvent dégénérer et rentrer dans une dynamique tumorale. On appelle gliome une tumeur de la glie. Nous allons développer plus en détail les différentes caractéristiques et catégories de gliomes dans la partie suivante.

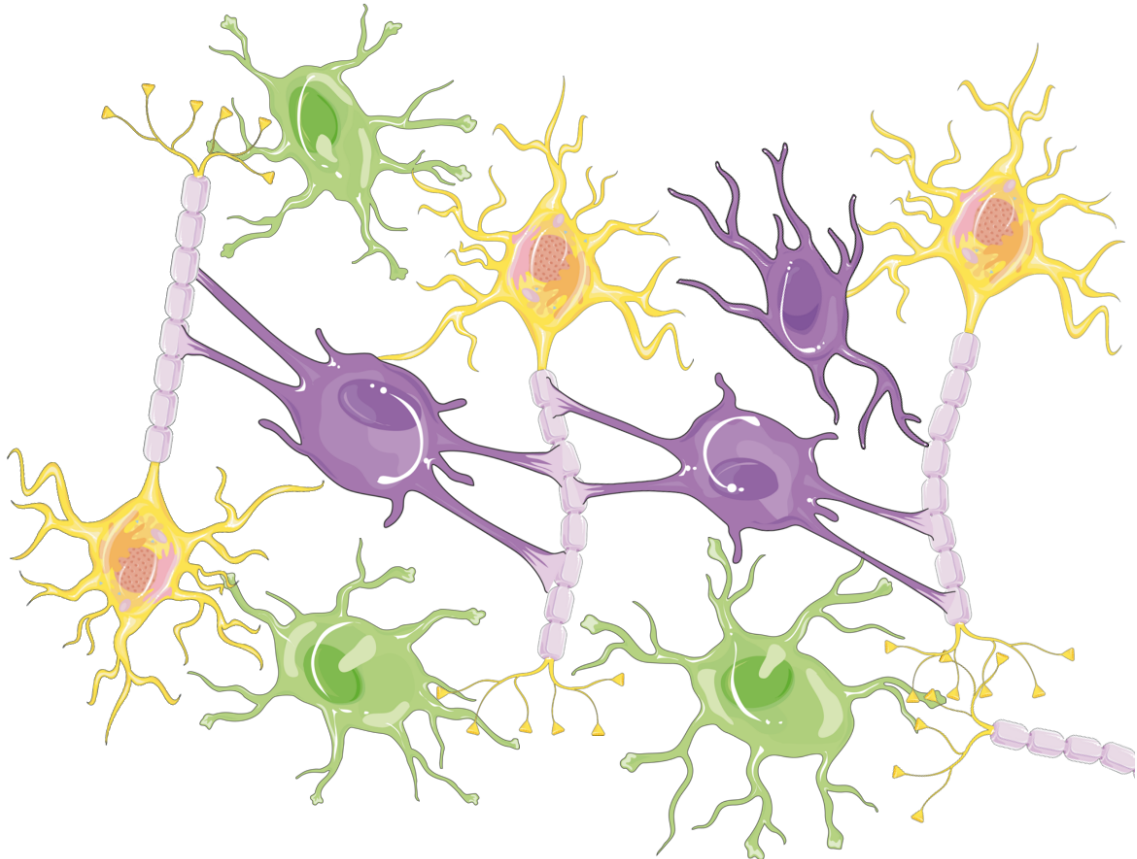


FIGURE 1.7 : Neurones (jaunes), astrocytes (verts) et oligodendrocytes (violets). (Composition d'images de Servier Medical Art - CC BY 3.0)

Le gliome, tumeur cérébrale

Contenu du chapitre

2.1 Histopathologie	21
2.2 Épidémiologie	22
2.3 Dynamique tumorale	22
2.4 Classification des gliomes et grades	23
2.5 Diagnostic	29
2.6 Suivi thérapeutique	29

Dans ce chapitre, nous allons présenter le gliome selon différents aspects. Seront évoqués des notions d'histopathologie et d'épidémiologie des gliomes, puis nous détaillerons la dynamique tumorale. Cela nous amènera à détailler la classification de l'OMS, qui attribue un grade selon l'état d'évolution de la pathologie.

Une tumeur, du latin *tumere* ("enfler") est le fruit d'une succession de divisions cellulaires incontrôlée. On identifie le type oncologique par la nature des cellules à l'origine du processus tumoral. Issu de la dégénérescence de cellules gliales, le gliome est aussi caractérisé dans son évolution par un grade. La classification du gliome évolue avec l'état des connaissances : la dernière mise à jour issue de l'OMS date de 2016 [68]. La principale distinction s'effectue entre gliome de bas grade et gliome de haut grade, deux entités ayant des dynamiques tumorales et des conséquences dans le suivi thérapeutique très différentes. Un gliome de bas grade évolue inexorablement vers un gliome de haut grade, au développement plus agressif et donc plus dangereux pour le patient.

2.1 Histopathologie

Bien que "gliome" soit un terme englobant regroupant toute tumeur de la glie, une description plus fine est possible, partant de la nature cellulaire retrouvée dans la masse tumorale. Ainsi, on parlera d'astrocytomes pour une tumeur provenant des astrocytes et d'oligodendrogliomes pour une tumeur issue des oligodendrocytes. Des gliomes mixtes peuvent exister, regroupant astrocytes et oligodendrocytes : on les appelle oligoastrocytomes.

Au cours de la dégénérescence cellulaire, les cellules tumorales perdent peu à peu leurs caractères différenciés. Cette perte de différenciation s'appelle anaplasie cellulaire. L'anaplasie est un marqueur supplémentaire de l'état d'avancement du gliome qui s'observe par l'observation d'un prélèvement histologique. Ce prélèvement peut être effectué par biopsie ou depuis l'analyse d'une pièce chirurgicale. Ces deux méthodes sont donc invasives pour le patient. Le stade ultime d'évolution du gliome est le glioblastome multiforme (GBM). Il s'agit de la forme la plus agressive, associée au pronostic le plus sombre pour le patient.

2.2 Épidémiologie

En fonction de leur stade de développement, les gliomes sont associés à des durées de survie très variables, allant dans les meilleurs cas de 10 ans à 20 mois dans le cas d'un GBM. Parmi les facteurs de risques, on trouve une prévalence plus importante pour les individus de sexe masculin et une influence de l'âge [86, 79]. Ainsi, les gliomes anaplasiques et GBM sont plus fréquents chez les 75-84 ans, mais il faut noter que les oligodendrogliomes sont détectés majoritairement chez les 35-44 ans selon une étude de 2018 [88]. Conjointement, le gliome se classe comme deuxième cause de mortalité des personnes âgées de 20 à 40 ans [14]. Face à la différence de survie constatée entre un gliome non-anaplasique et un gliome anaplasique, il est important d'étudier la dynamique tumorale afin d'identifier les facteurs à l'origine de la transformation anaplasique.

2.3 Dynamique tumorale

Partons de la cellule saine jusqu'au glioblastome. Dans une cellule saine, certains gènes et protéines permettent de corriger les mutations génétiques ou d'enrayer un emballement, en utilisant notamment la voie apoptotique pour éviter la multiplication frénétique tumorale. Des mutations génétiques entraînant des modifications de ces facteurs de contrôle peuvent être à l'origine de la dynamique tumorale. Ainsi, l'étude génétique apporte sa part d'information dans l'analyse oncologique.

Des mutations apparaissent suite à des prédispositions génétiques ou des facteurs environnementaux, et un gliome se développe. Ce gliome va connaître un développement en deux phases : une phase lente et silencieuse, suivi d'une accélération avec une expansion rapide, une transformation nécrotique et d'importants symptômes neurologiques pour le patient.

La première phase est silencieuse, c'est à dire asymptomatique pour le patient. De part son développement lent, les fonctions cérébrales altérées sont progressivement compensées par l'hémisphère controlatéral, grâce au phénomène de plasticité cérébrale. Durant ce développement lent, le gliome présente un fort caractère infiltrant. Des cellules

tumorales vont migrer dans les tissus environnants, provoquant inflammation et œdème dans une zone étendue et très diffuse. Certains symptômes peuvent survenir comme des épisodes épileptiques. La crise d'épilepsie est d'ailleurs le premier symptôme qui amène les patients à passer un examen IRM. Cependant, certains gliomes restent asymptomatique malgré l'infiltration et peuvent continuer leur développement pendant la phase silencieuse sans qu'un traitement ne soit envisagé. Les gliomes présentant ce développement silencieux sont sous-diagnostiqués, faisant parfois l'objet de découverte fortuite [94].

Dans un second temps, le gliome va mobiliser localement le système vasculaire pour son apport énergétique. Apparaît alors le phénomène de néoangiogenèse, sur des foyers localisés. On peut alors identifier ces foyers comme cœurs tumoraux, par opposition à la zone d'infiltration. Cette mobilisation énergétique s'accompagne d'une augmentation du nombre de divisions cellulaires. À terme, ces divisions cellulaires peuvent rompre la barrière hémato-encéphalique, propageant des cellules tumorales via le système vasculaire. Finalement, dans son stade le plus avancé, les cellules composant le cœur tumoral se dégradent et meurent, laissant place à une zone nécrotique entourée de néoangiogénèse en couronne. Les mécanismes précis de cette dégradation nécrotique sont flous, mais des indices amènent à penser qu'une défaillance de la voie glutaminergique en serait une des origines possibles [84].

2.4 Classification des gliomes et grades

Classification OMS : Depuis 1979, l'OMS propose une classification des tumeurs du système nerveux central. Cette classification sert de référence et a évolué avec l'avancée des connaissances. La version actuelle date de 2016 et est la cinquième édition de cette classification [68]. La précédente datait de 2007 [67]. La mise à jour de cette classification est un bel exemple de l'organisation du débat scientifique. En effet, chaque itération provoque son lot de débat dont les thèses finissent intégrées ou non à l'édition suivante. Ce fût par exemple la transition de la quatrième à la cinquième version. Plusieurs auteurs pointaient l'intérêt d'ajouter à la classification tumorale des facteurs issus de l'analyse

généétique. La classification actuelle utilise donc des critères histologiques et génétiques. Conséquence de ces débats : d'une version à l'autre, différentes entités diagnostiques apparaissent ou disparaissent, comme par exemple la gliomatose, qui a disparue entre la quatrième et la cinquième version, remplacée par des sous-ensembles génotypiques d'oligodendrogliomes.

Grades (concept) : La classification OMS s'articule autour du concept de grade. Le grade est un marqueur de l'agressivité et donc de l'évolution du gliome. On distingue deux familles divisées en sous catégories : les gliomes de bas grade composés des grades I et II, et les gliomes de haut grade composés des grades III et IV. Le grade est la discrétisation temporelle d'une évolution continue. C'est pourquoi un débat est ouvert sur la création d'un grade II+, intermédiaire entre le grade II et III dont la transformation est lente [98]. La méthode de référence d'établissement du grade est l'analyse histopathologique issue d'une biopsie ou de l'examen d'une pièce chirurgicale. Ces techniques invasives ne permettent l'analyse qu'une petite partie du gliome : il s'agit donc d'un sous-échantillonnage dont on tire une information globale. Ce sous-échantillonnage soulève des problèmes de variabilité : erreur d'échantillonnage [55] et variabilité inter-observateur [117]. De plus, des risques existent pour le patient dans le cas de biopsies de gliomes difficilement atteignable [55].

Bas grades : La classification actuelle ne reconnaît que quelques cas rares de gliomes de grade I. Citons par exemple le gliome angiocentrique touchant le jeune enfant [61]. Le reste des gliomes de bas grade sont des gliomes de grade II, conformes aux catégories oncologiques évoquées précédemment. La catégorie des bas grades de l'OMS est contestée et une autre appellation de "gliome diffus de bas grade" (GDBG) est parfois utilisée [37].

Hauts grades : La transformation anaplasique est le marqueur oncologique du haut grade. Ainsi, le grade III contient les astrocytomes anaplasiques et les oligodendrogliomes anaplasiques. Le grade IV contient les glioblastomes multiformes. Cependant, cette dynamique d'évolution tumorale, qui part d'un gliome de bas grade pour aboutir à un haut

grade, est minoritaire dans l'incidence des gliomes de haut grade. En effet, une majorité des gliomes sont des hauts grades *de novo*, qui apparaissent spontanément à un grade élevé. Cette différence entre un gliome primaire (apparition *de novo*) et secondaire (évolution anaplasique d'un gliome de bas grade) a été établie dès 1940 par H. J. Scherer [85, 103], bien en avance sur les institutions actuelles. L'OMS donnera raison à cette démarcation, notamment par l'intégration des facteurs génétiques en 2016.

Afin d'être constant dans notre propos, nous faisons le choix de rester aligné sur les grades de l'OMS : nous utiliserons "bas grades" pour les grades I et II, et "hauts grades" pour les grades III et IV de l'OMS.

Le Tableau 2.4 récapitule les caractéristiques histologiques des différents gliomes selon la classification OMS.

Type	Différentiation	Densité cellulaire	Atypies cytonucléaires	Activité mitotique	Nécrose	Prolifération capillaire
Astrocytome grade II	Elevée	Modérée	Occasionnelles	<1 mitose	Absente	Absente
Oligodendrogliome grade II	Elevée	Modérée	Possible	Occasionnelle	Absente	Absente
Oligoastrocytome grade II	Elevée	Faible ou modérée	N/A ¹	Absente ou faible	Absente	Absente
Astrocytome grade III	Anaplasie focale ou élevée	Augmentation de façon diffuse ou focale	Présentes	Présente	Absente	Absente
Oligodendrogliome grade III	Anaplasie focale ou dispersée	Augmentée	Possibles	Possiblement forte	Possible	Possible
Oligoastrocytome grade III	N/A ¹	Augmentée	Possibles	Possiblement forte	Possible	Possible
Glioblastome grade IV	Faible	Elevée	Marquées	Marquée	Présente	Présente

¹Données manquantes

TABLE 2.1 : *Classification des gliomes selon l'OMS (WHO 2000).*

Source : thèse de Mathieu Naudin [82]

Facteurs génétiques : Les facteurs génétiques intègrent la classification OMS en 2016. En particulier, quatre facteurs sont identifiés : les mutations des gènes codant les protéines IDH, ATRX et TP53 ; et la codélétion chromosomique 1p/19q. Voici leurs caractéristiques :

- **Isocitrate Deshydrogénase (IDH)** est une enzyme intervenant dans le cycle de Krebs, élément essentiel de l'énergétique cellulaire. La mutation peut porter soit sur le gène de l'IDH1, soit de l'IDH2. L'influence de sa mutation dans la dynamique du gliome est identifiée dès 2012 [56] et donne lieu dans la classification de 2016 aux GBM IDH *mutant* et GBM IDH *wildtype*.
- **La Tumor Protein 53 (TP53)** est une protéine qui joue un rôle dans la régulation des divisions cellulaires. Elle intervient dans le cycle cellulaire, l'autophagie et l'apoptose. De par ce rôle clé, sa mutation possède le potentiel de déclencher une dynamique tumorale. La mutation de la TP53 est présente dans 60% des astrocytomes de bas grade. Elle constitue ainsi un indice permettant la différenciation entre gliome primaire et gliome secondaire [85].
- **La Transcriptional regulator ATRX (ATRX)** est une protéine dont le rôle dans la dynamique du gliome bien compris [81]. Cette protéine intervient lors de la réplication de l'acide désoxyribonucléique (ADN). Tout comme la TP53, sa mutation a le potentiel de déclencher une dynamique tumorale. On retrouve cette mutation dans les astrocytomes de bas grade.
- **La codélétion 1p/19q** est un marqueur génétique présent dans les oligodendrogliomes de bas grade. Il s'agit de la délétion du bras court du chromosome 1 et du bras long du chromosome 19. Tous les gliomes 1p/19q codélétés sont également porteurs de la mutation IDH [62].

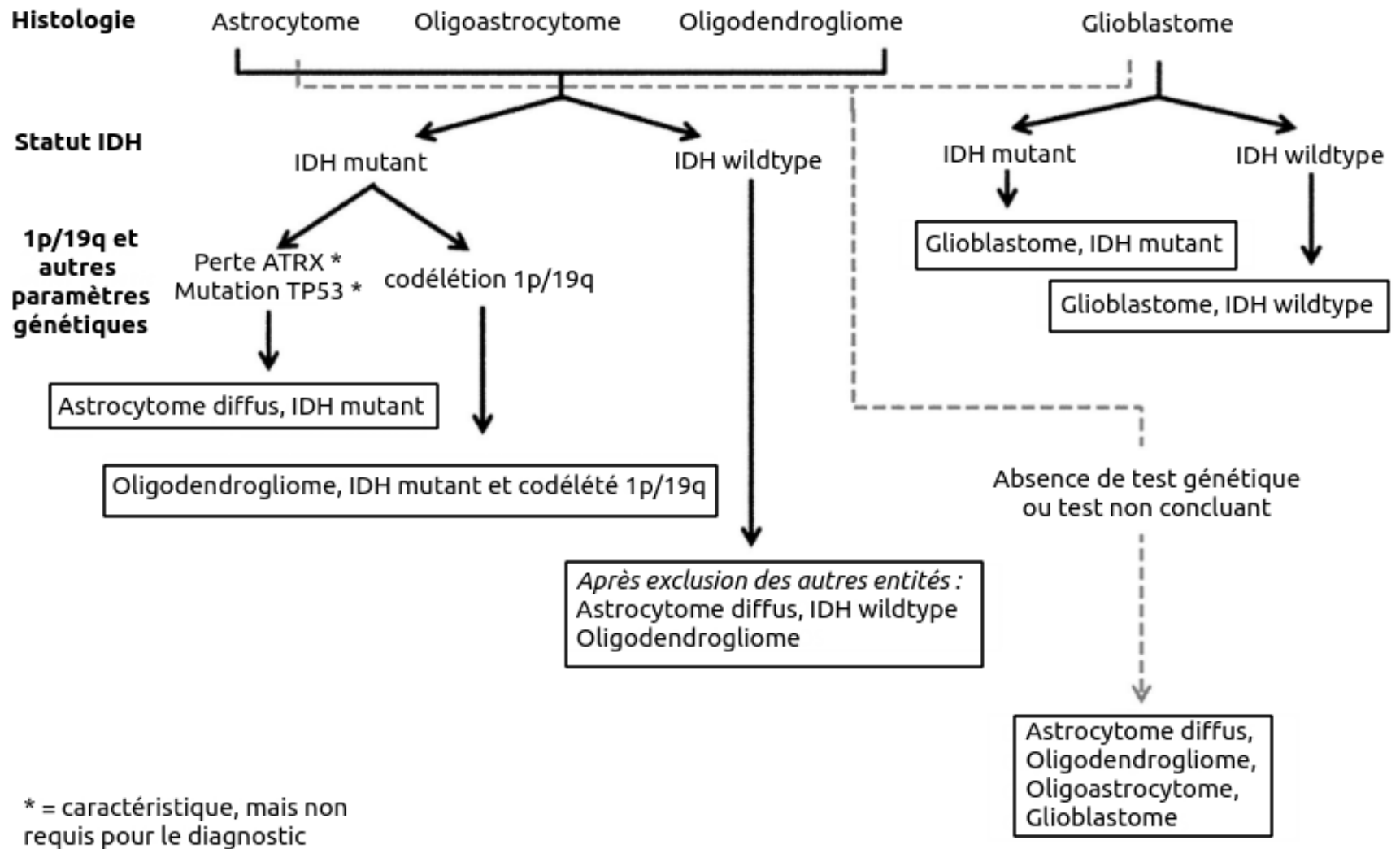


FIGURE 2.1 : Arbre de décision diagnostic selon l'OMS (2016).

En intégrant la biopsie et les données moléculaires et génomiques, la publication de référence de l’OMS en 2016 érige l’arbre de décision diagnostic représenté Figure [2.1](#). Ce schéma permet d’obtenir, depuis le type histologique observé, la qualification oncologique précise selon le statut IDH, la présence ou l’absence de codélétion 1p/19q et les autres indices génétiques que nous avons détaillé.

2.5 Diagnostic

Les signes évocateurs d’un gliome sont des maux de tête, des nausées et vomissements, un déficit moteur, des troubles affectant la parole, la vision, le comportement, l’humeur ou la mémoire et des crises d’épilepsies.

Concernant l’imagerie, lorsqu’un patient se présente aux urgences du CHU de Poitiers pour des maux de tête ou des pertes fonctionnelles, la première intention est d’effectuer un scanner. Le scanner est une acquisition irradiante qui permet d’obtenir les premières images rapidement. Si une lésion est identifiée, alors on proposera au patient un examen IRM. Une biopsie peut être réalisée. Elle permettra l’analyse histopathologique et moléculaire. Suite à quoi, le patient sera orienté vers différentes stratégies thérapeutiques.

2.6 Suivi thérapeutique

Face à l’apparition d’un gliome, plusieurs stratégies thérapeutiques existent et dépendent de l’avancée de la maladie, de l’étendue de la zone d’infiltration, ou encore de la localisation de la tumeur dans le cerveau. L’objectif est double : maximiser la durée de survie du patient tout en préservant sa qualité de vie. Lorsqu’elle est possible, la résection tumorale qui est envisagée en première. Elle est complétée par des traitements médicamenteux. La stratégie thérapeutique peut également inclure des séances de radiothérapie.

La résection tumorale est un acte chirurgical consistant à ouvrir la boîte crânienne pour retirer les tissus tumoraux. C’est une technique très invasive, mais qui donne les meilleurs résultats thérapeutiques à long terme, notamment car elle permet d’endiguer la propagation tumorale. Lors d’une chirurgie conventionnelle, le patient est sous anesthésie

générale durant toute l'opération. Le chirurgien utilise un outil de neuro-navigation pour se repérer dans le cerveau du patient. Cet outil intègre les images du bilan pré-opératoire, réalisé quelques jours avant l'intervention. La chirurgie peut être partielle, totale, ou supra-totale. Cela dépend de la position de la tumeur, des fonctions cérébrales présentes dans les zones environnantes et de leur facilité d'accès. Ainsi, la résection totale est plus facilement envisageable sur une tumeur très localisée et peu infiltrante que sur une tumeur très diffuse ayant infiltrée de grandes zones fonctionnelles. La résection supra-totale, quand elle est possible, vise à empêcher au maximum les récurrences. Cette technique consiste à retirer le tissu cérébral au-delà des frontières visibles de la tumeur, afin d'éliminer les cellules cancéreuses infiltrées dans les tissus sains environnants. Mais l'acte chirurgical n'est jamais anodin et la balance bénéfice-risque doit minimiser la dégradation fonctionnelle liée à l'acte chirurgical.

C'est pourquoi, pour limiter la perte fonctionnelle, il est possible d'avoir recours à des techniques dites de chirurgie éveillée, où le patient est réveillé durant l'opération pour effectuer des tâches en rapport avec la zone endommagée (élocution, mouvement d'un membre...). Le neurochirurgien peut inhiber momentanément des zones fonctionnelles grâce à un outil à impulsions électriques, ce qui lui permet de simuler, dans le repère subjectif du patient, l'éventuelle perte fonctionnelle de l'acte chirurgical. Par ce processus, le chirurgien se crée sa propre carte fonctionnelle patient spécifique, ce qui aide à raffiner les possibilités de résection tout en minimisant la dégradation fonctionnelle du patient.

De surcroît, il est également possible au CHU de Poitiers d'effectuer un examen IRM per-opératoire. Cet examen permet au chirurgien d'obtenir de nouvelles informations sur l'état des tissus pendant l'opération, après un premier geste chirurgical. Ainsi, selon les résultats de l'imagerie, il peut ensuite décider de poursuivre éventuellement l'opération pour accroître la résection ou non.

Selon l'évolution de la tumeur, la stratégie thérapeutique doit s'adapter. Cela demande de prédire l'évolution de la dynamique tumorale à partir des examens antérieurs, voire à partir du simple examen initial. Un suivi régulier du patient est alors nécessaire pour surveiller l'évolution des cas sensibles et acquérir plus de données temporelles. Ce

suivi, dit longitudinal, permet d'évaluer la dynamique du gliome mais aussi la réponse thérapeutique. Il s'effectue principalement par IRM et permet le recueil d'informations multiples de manière non invasive.

Physique de l'IRM

Contenu du chapitre

3.1 Introduction	33
3.2 Magnétisme nucléaire	34
3.3 Résonance magnétique nucléaire	34
3.4 Relaxations	37
3.5 Séquences RMN : spin-echo	41
3.6 Reconstruction de l'image	44
3.7 Qualité du signal, qualité d'image	48
3.8 Modalités IRM et bilan lésionnel	54
3.8.1 IRM anatomique	56
3.8.2 IRM de diffusion	58
3.8.3 IRM de perfusion	63
3.8.4 Spectroscopie RMN	65
3.8.5 IRM fonctionnelle	70

Après avoir introduit les tumeurs de la glie et détaillé leur évolution, nous devons présenter les moyens d’investigation et d’exploration que nous allons utiliser pour les caractériser. De par sa grande variété d’informations, l’imagerie par résonance magnétique (IRM) offre une voie d’exploration privilégiée. C’est pourquoi nous allons dans un premier temps revenir sur les principes physiques en jeu lors d’une acquisition IRM, du magnétisme nucléaire à la reconstruction de l’image. Puis, nous allons présenter les différentes modalités IRM utilisées au CHU de Poitiers lors du bilan lésionnel.

3.1 Introduction

Il est rare, dans l’histoire des sciences, qu’une même découverte advienne simultanément en deux endroits séparés du globe. Parmi les grands exemples, citons la découverte par Darwin et Wallace de la théorie de l’évolution en 1858, ou encore l’invention du calcul infinitésimal par Leibniz et Newton, respectivement en 1669 et 1674. Le phénomène de résonance magnétique nucléaire (RMN) rentre dans cette catégorie de découvertes : en 1946, il est à la fois décrit par Felix Bloch à Stanford et par Edward M. Purcell à Harvard [16, 95]. Ces deux chercheurs obtiendront ensemble le prix Nobel de physique en 1952. Aussitôt, la RMN devient un champ d’étude en expansion, d’abord via l’usage de la spectroscopie RMN entre 1950 et 1970, puis de l’imagerie les années suivantes, avec deux autres prix Nobel à la clef. Bien que l’écho de spin, principe fondateur des techniques d’imagerie, ait été connu depuis 1949, il faut attendre 1973 pour que Paul C. Lauterbur réalise les premières images RMN [63]. Richard Ernst quant à lui met au point en 1975 les techniques d’encodage de phase et de fréquence toujours en vigueur aujourd’hui [60]. L’année suivante, Peter Mansfield met au point l’échoplanar, permettant des acquisitions rapides [72]. Pour ces travaux, Richard Ernst recevra le prix Nobel de chimie en 1991 ; Paul C. Lauterbur et Peter Mansfield celui de médecine en 2003 [43].

Dans ce chapitre, nous allons exposer comment, partant de la structure fondamentale de la matière, le phénomène de résonance magnétique permet de recueillir un signal physique. Ce signal sera ensuite converti en une image dont nous expliquerons les caractéristiques visuelles. Enfin, nous détaillerons comment l’ajout de différentes modalités

d'acquisition RMN transforme l'IRM en un outil d'exploration multivarié et riche en informations. Une grande partie des informations de ce chapitre provient de la 8ème édition du livre *Comprendre l'IRM, Manuel d'auto-apprentissage*, de B. Kastler et D. Vetter [58].

3.2 Magnétisme nucléaire

Le magnétisme nucléaire est une propriété atomique. Les atomes, dont le noyau est composé de protons et de neutrons, sont porteurs de charges. Le noyau d'hydrogène est composé d'un unique proton possédant des propriétés magnétiques intrinsèques très élevées. Le proton tourne sur lui-même : on parle de spin nucléaire.

Ce spin peut prendre deux valeurs. selon sa direction, il peut être positif ou négatif. La charge du proton, couplée à la rotation du noyau, contribue à créer un moment magnétique propre au proton. Ce moment magnétique est une grandeur vectorielle, colinéaire à l'axe de rotation du proton. Ainsi, de part sa charge, sa rotation et son moment magnétique, le proton est assimilable à un dipôle magnétique. Soumis à un champ électromagnétique, il s'oriente comme un aimant.

3.3 Résonance magnétique nucléaire

La résonance magnétique nucléaire étudie les variations d'aimantation des noyaux en présence de deux champs électromagnétiques \vec{B}_0 et \vec{B}_1 . \vec{B}_0 est un champ fixe et \vec{B}_1 est un champ tournant. Les protons soumis à \vec{B}_0 décrivent un mouvement de simple précession : c'est l'état d'équilibre. L'application de \vec{B}_1 , sous forme d'impulsion, vient perturber les protons orientés selon \vec{B}_0 . Nous allons dans un premier temps décrire le comportement du proton dans \vec{B}_0 , puis détailler les différents effets de l'application du champ \vec{B}_1 .

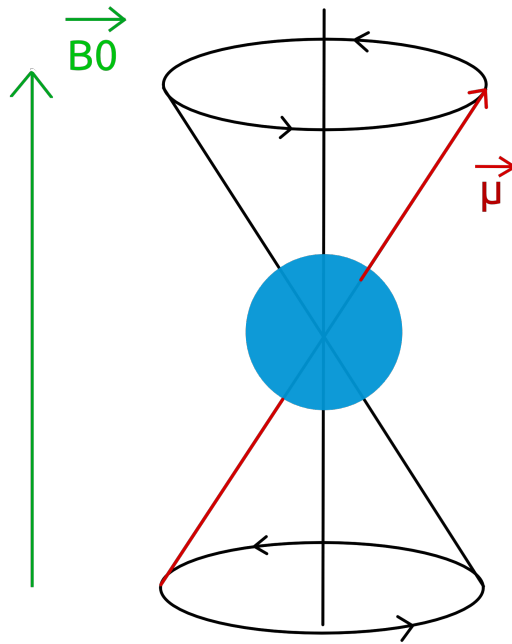


FIGURE 3.1 : *Mouvement de simple précession du proton dans le champ \vec{B}_0 . Le petit vecteur $\vec{\mu}$ indique le moment magnétique nucléaire du proton.*

Plongé dans un champ magnétique \vec{B}_0 , les protons s'alignent et décrivent un mouvement de simple précession (Figure 3.1). Tout en pivotant sur eux-mêmes, les protons tournent autour de \vec{B}_0 comme le ferait une toupie en fin de course. Ce mouvement décrit un double cône, orienté positivement ou négativement selon le spin nucléaire.

Le mouvement de simple précession est décrit par l'équation de Larmor. Il est représenté par la Figure 3.1

$$\omega_0 = \gamma \times \vec{B}_0 \quad (3.1)$$

Avec : ω_0 la fréquence angulaire de Larmor, ou fréquence angulaire de précession (en radians par seconde)

γ le rapport gyromagnétique

\vec{B}_0 l'intensité du champ magnétique \vec{B}_0 (en Tesla)

Décrivons cet état d'équilibre. Dans un tissu biologique, les moments magnétiques de tous les protons orientés selon \vec{B}_0 s'additionnent pour créer un vecteur d'aimanta-

tion macroscopique \vec{M} . Ce vecteur peut être décrit comme la somme d'une composante longitudinale \vec{M}_z , orientée dans la direction de \vec{B}_0 , et d'une composante transversale \vec{M}_{xy} située dans le plan orthogonal à \vec{B}_0 . Comme une majorité des protons possède un spin positif, \vec{M}_z est positive et pointe dans le sens de \vec{B}_0 . A l'état d'équilibre, les spins sont déphasés et la composante transversale \vec{M}_{xy} est nulle. Notons qu'une plus grande densité de protons provoque un vecteur d'aimantation macroscopique de norme plus importante, ce qui se traduira par une plus grande quantité de signal.

Le champ \vec{B}_1 vient perturber la précession simple décrite précédemment. A l'aide d'une impulsion sous forme d'onde de radio-fréquence (onde RF), \vec{B}_1 apporte de l'énergie au système. Pour que cet apport puisse se produire, il faut que la fréquence de rotation du champ magnétique tournant, ω_r , soit égale à ω_0 . Cette condition permet l'apparition du phénomène de résonance magnétique nucléaire : il y a transfert d'énergie de l'onde RF aux protons.

Tout en précessant autour de \vec{B}_0 , les protons vont alors précesser autour de \vec{B}_1 . On parle alors de double précession. C'est lors de ce double mouvement que le proton va accumuler de l'énergie qu'il restituera sous forme de signal.

En fonction de la durée des impulsions RF d'excitation de \vec{B}_1 , le vecteur \vec{M} va plus ou moins précesser autour de son axe transverse. En pratique, on utilise des impulsions provoquant des bascules de 90° ou 180° . Les variations de \vec{M}_z et \vec{M}_{xy} selon chaque type d'impulsion sont décrites dans le Tableau [3.1](#).

Pour une impulsion de 90° , l'impulsion du champ \vec{B}_1 met en phase des spins, ce qui provoque l'apparition de la composante transversale \vec{M}_{xy} . Cette synchronisation n'apparaît pas lors d'une impulsion à 180° , où la valeur initiale de \vec{M}_{xy} est nulle. Dans cette partie, nous n'allons considérer uniquement que les effets d'une impulsion de 90° .

Lorsque l'impulsion \vec{B}_1 s'arrête, le vecteur d'aimantation M va progressivement retourner à l'état d'équilibre (Figure [3.2](#)). L'énergie accumulée lors de la résonance va être restituée sous la forme d'une onde : c'est le phénomène de relaxation.

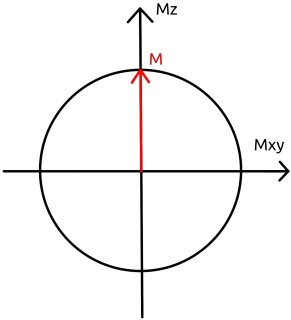
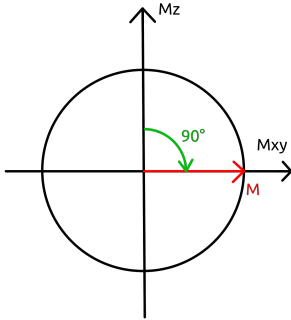
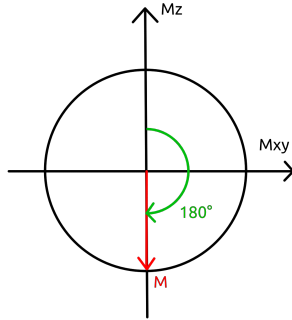
Composante	État d'équilibre	Impulsion de 90°	Impulsion de 180°
Longitudinale	+Mz (Maximale)	Nulle	-Mz (Minimale)
Transversale	Nulle	+Mxy (Maximale)	Nulle
			

TABLE 3.1 : Valeurs et représentations des composantes longitudinales et transversales selon l'impulsion utilisée.

3.4 Relaxations

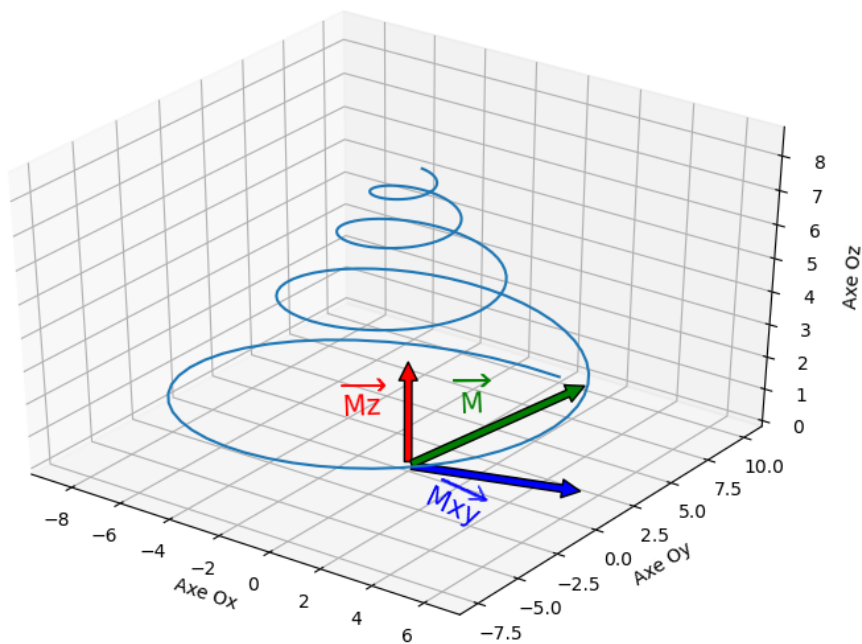


FIGURE 3.2 : Relaxation 3D : le vecteur d'aimantation \vec{M} (vert) et ses composantes transversales (bleu) et longitudinale (rouge). Lors des relaxations T1 et T2, \vec{M} retourne à l'état d'équilibre, en simple précession autour de \vec{B}_0 .

Le phénomène de relaxation décrit le retour à l'état d'équilibre et la restitution de l'énergie accumulée lors de la résonance magnétique nucléaire. Pour la composante longitudinale, le retour à l'équilibre se traduit par une repousse de la composante. Cette repousse s'appelle relaxation T1. Pour la composante transversale, le retour à l'équilibre provoque la disparition de la composante : c'est la relaxation T2.

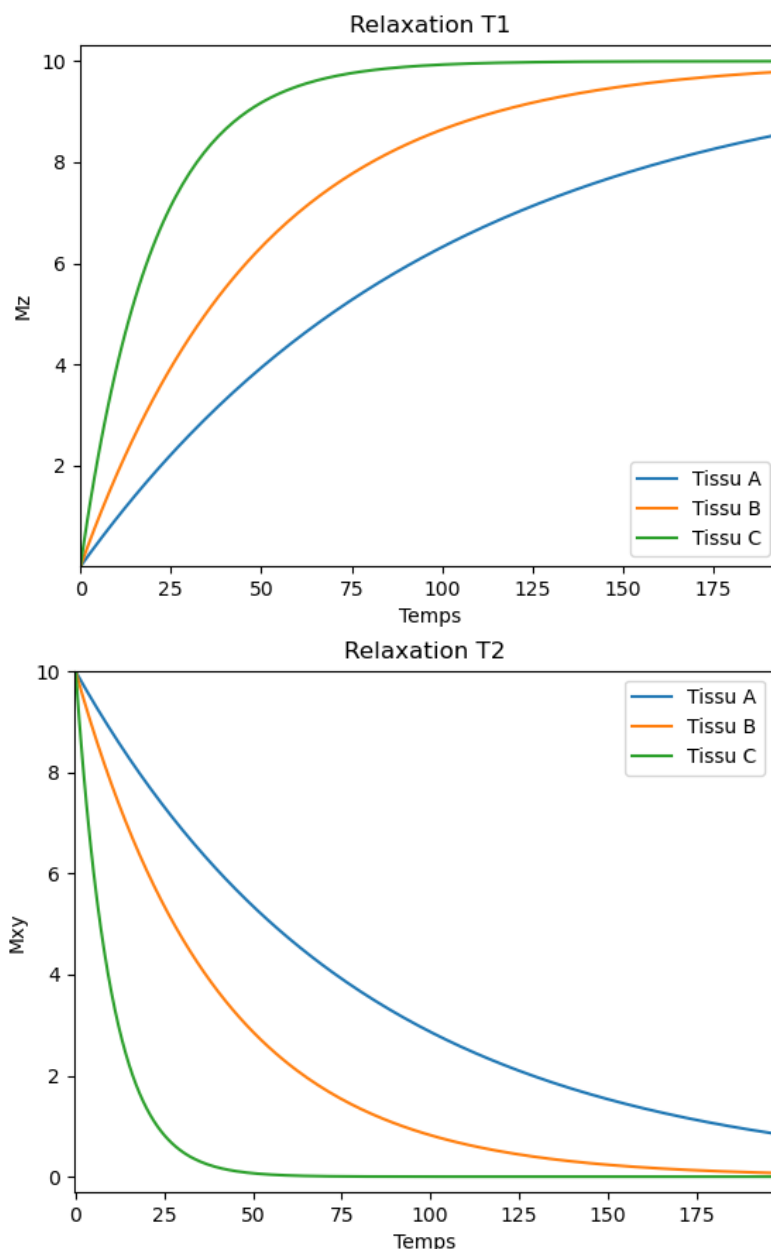


FIGURE 3.3 : Relaxations T1 (en haut) et T2 (en bas). Des tissus différents ont des constantes de temps différentes, donc des relaxations différentes.

Ces deux évolutions suivent des lois exponentielles, l'une croissante, l'autre décroissante (Figure 3.3). En physique générale, les évolutions exponentielles sont caractérisées par une constante de temps τ , propre à chaque phénomène. Cette constante équivaut au temps nécessaire à l'accomplissement de 63% de l'évolution observée. Autrement dit, on appellera T1 le temps nécessaire à ce que 63% de la repousse longitudinale ait eu lieu, et T2 le temps nécessaire à ce que 63% de la décroissance transversale s'accomplisse.

Les équations des relaxations sont les suivantes :

Relaxation T1 :

$$M_z(t) = M_z(0) \times \left(1 - e^{-\frac{t}{T1}} \right) \quad (3.2)$$

Relaxation T2 :

$$M_{xy}(t) = M_{xy}(0) \times e^{-\frac{t}{T2}} \quad (3.3)$$

Avec : $M_z(0)$ et $M_{xy}(0)$ les valeurs initiales des vecteurs \vec{M}_z et \vec{M}_{xy}

t le temps (en seconde)

$T1$ et $T2$ les constantes de temps (en seconde)

Les relaxations T1 et T2 sont des phénomènes distincts. Pour un même tissu, les temps T1 et T2 sont différents. Avec un champ magnétique de 3 Tesla, l'ordre de grandeur du T2 des tissus cérébraux est d'environ 70 à 100 ms, soit environ 15 à 20 fois plus court que le T1 de ces mêmes tissus (112).

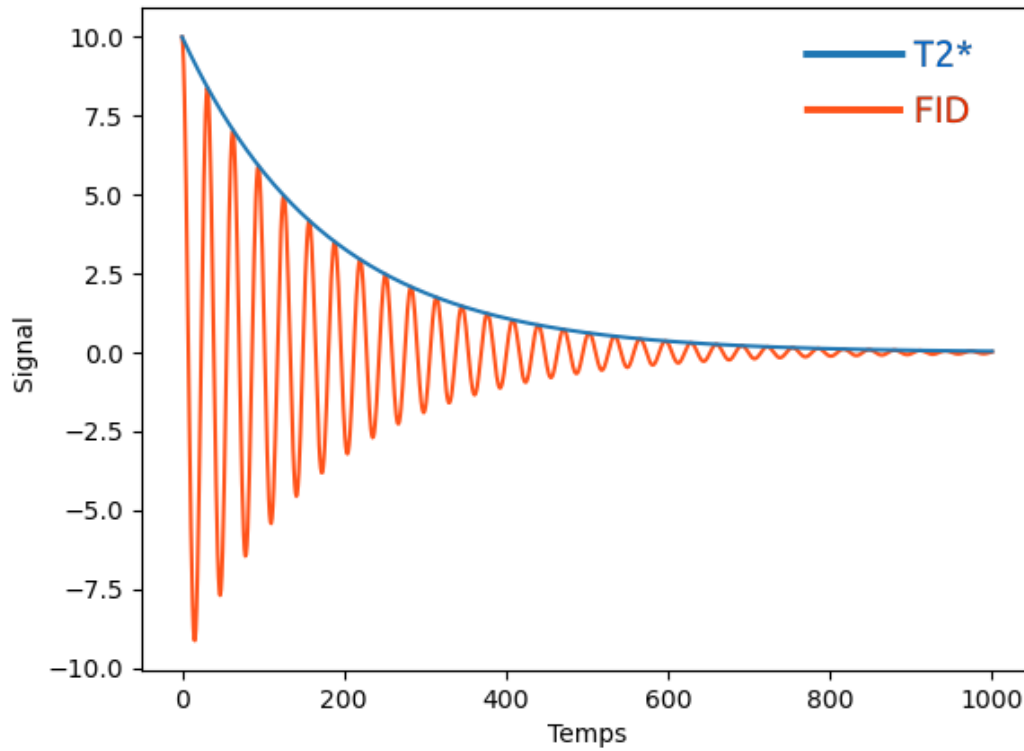


FIGURE 3.4 : *Décroissance du signal $T2^*$.*

La relaxation provoque la libération de l'énergie accumulée. Cette énergie se libère sous la forme d'une onde électromagnétique. On peut mesurer cette énergie à l'aide d'une bobine positionnée sur l'axe Ox. L'acquisition résultante montre l'oscillation décroissante réalisée par la composante transversale. Ce signal appelé FID (Free Induction Decay) est le signal fondamental de l'acquisition RMN.

3.5 Séquences RMN : spin-echo

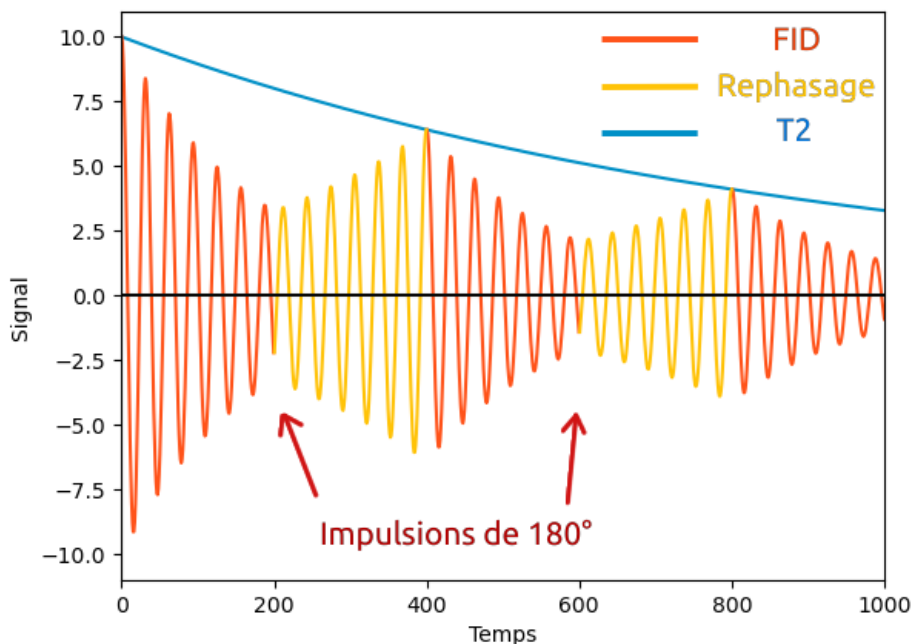


FIGURE 3.5 : Illustration du principe de l'acquisition spin-écho : le signal décroît plus rapidement (décroissance $T2^*$) que le $T2$ réel. L'impulsion de 180° permet de rephaser les spins. On mesure des points de l'enveloppe pour obtenir la décroissance $T2$.

Dans un monde parfait, la décroissance que suit le signal FID est équivalente à la décroissance $T2$ du matériau étudié. Cependant, des inhomogénéités microscopiques de champ, issues d'imperfections irréductibles au niveau de la qualité de l'aimant et de son champ magnétique, provoquent un déphasage accéléré des spins, et donc une décroissance accélérée du signal. L'enveloppe mesurée correspond alors à un $T2$ amendé appelé $T2^*$ (*T2 star*, *T2 étoile*), représentée sur la Figure 3.4.

Pour retrouver la véritable enveloppe $T2$, on utilise une impulsion RF de 180° , ce qui a pour effet, de rephaser les spins (Figure 3.5). Le signal transversal connaît alors un maximum local, avant que les spins ne se redéphasent à nouveau. Ainsi, une répétition d'impulsions similaires permet de mesurer le signal $T2$, diminué de son enveloppe décroissante. Suite à chaque impulsion, le signal revient périodiquement comme un écho.

Ainsi, on appellera temps d'écho (TE) le temps entre la première impulsion de 90° et la resynchronisation des spins suite à l'impulsion de 180° . Du fait de la construction du signal, l'impulsion de 180° a lieu à $TE/2$.

On définit également le temps de répétition (TR), qui est le temps séparant deux impulsions de 90° . Le motif d'impulsions RF qui a lieu durant un TR constitue un bloc fondamental, qui va ensuite se répéter jusqu'à la fin de l'acquisition. Une acquisition complète, composée d'une succession de ces blocs fondamentaux, s'appelle une séquence. Différentes séquences, composées de différents blocs fondamentaux, vont permettre de mesurer différents signaux. Le bloc fondamental de la séquence spin-echo est illustré dans la Figure 3.6.

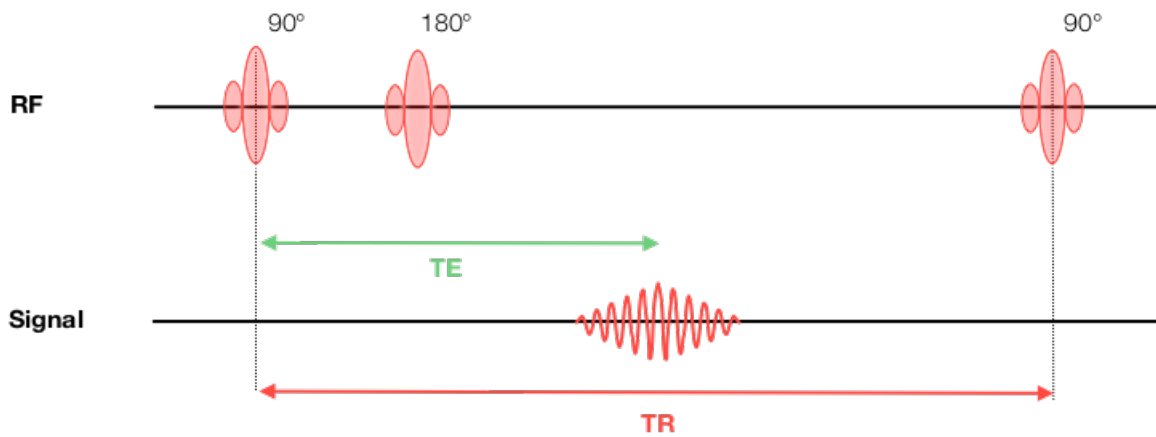


FIGURE 3.6 : Chronogramme représentant le bloc fondamental de la séquence spin-echo. Le signal réapparaît et est mesuré à TE, suite à l'impulsion de 180° , déclenchée à $TE/2$. Deux impulsions de 90° sont séparées par un TR.

Source : thèse de Mathieu Naudin [82]

Les tissus biologiques ont des relaxations T1 et T2 et des densités protoniques différentes. De plus, les choix du TE et du TR déterminent la quantité de signal disponible, en laissant plus ou moins de temps aux différentes relaxations d'advenir. Le signal est mesuré lors de son écho après l'impulsion de 180° . Des couples de TR et TE spécifiques permettent l'acquisition de signaux représentatifs de la relaxation T1 ou T2 des tissus :

on parle alors de pondération. Une pondération T1 s'acquière à TR court et TE court. La pondération T2 s'acquiert à TR long et TE long.

Pour comprendre les différences en intensité de signal, il est possible de faire une représentation hybride, Figure 3.7, représentant sur un même axe repousse T1 et relaxation T2. Cet axe représente la quantité de signal disponible dans le tissu à l'instant T. La mesure s'effectue lorsque l'écho du signal apparaît, tous les $n \cdot TE$. La différence de signal entre les deux courbes nous donne le contraste de l'acquisition.

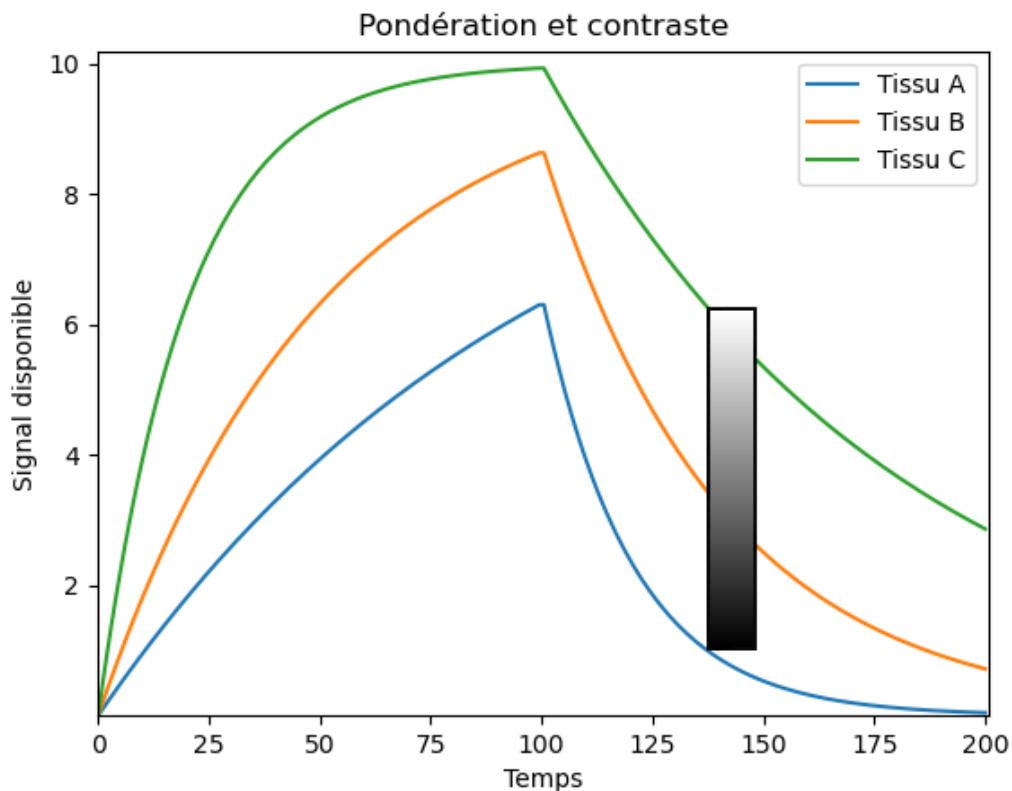


FIGURE 3.7 : Représentation hybride montrant à la suite relaxation T1 puis relaxation T2. Un bon paramétrage du TE permet d'obtenir une plus grande différenciation des tissus lors de l'acquisition, donc un meilleur contraste de l'image, représenté ici par un gradient de niveaux de gris.

En pondération T1, la substance blanche est blanche, la substance grise est grise et le LCR est noir. Lors d'une pondération T2, le LCR est très blanc, la substance grise est blanche et la substance blanche est grise.

3.6 Reconstruction de l'image

Nous avons expliqué la nature du signal RMN et le mécanisme selon lequel la relaxation T2 se libère de l'énergie accumulée. Mais ce signal n'est pas encore une image. Pour cela, on utilise un codage particulier à l'aide de trois gradients, et une opération mathématique : la transformée de Fourier inverse.

L'acquisition IRM explore un volume 3D. Avant de devenir une image facilement manipulable, trois gradients magnétiques sont utilisés lors de l'acquisition pour localiser le signal dans l'espace.

Le gradient de sélection de coupe (GSC) : Il permet de définir l'épaisseur de la coupe. Puis, dans la coupe ou le volume sélectionné, deux gradients seront successivement appliqués pour distinguer coder les protons en ligne et en colonne. On obtient ainsi une matrice dont on identifiera les lignes et les colonnes par les deux gradients suivants.

Le gradient de codage de phase (GCP) : Il permet de coder les lignes en attribuant à chaque proton du volume une phase différente par ligne.

Le gradient de codage de fréquence (GCF) : Il permet de coder les colonnes en attribuant une fréquence à chaque colonne de protons.

L'application des différents gradients est représentée dans la Figure [3.8](#). Après leur application, chaque voxel du volume sélectionné est identifiable par sa phase et sa fréquence.

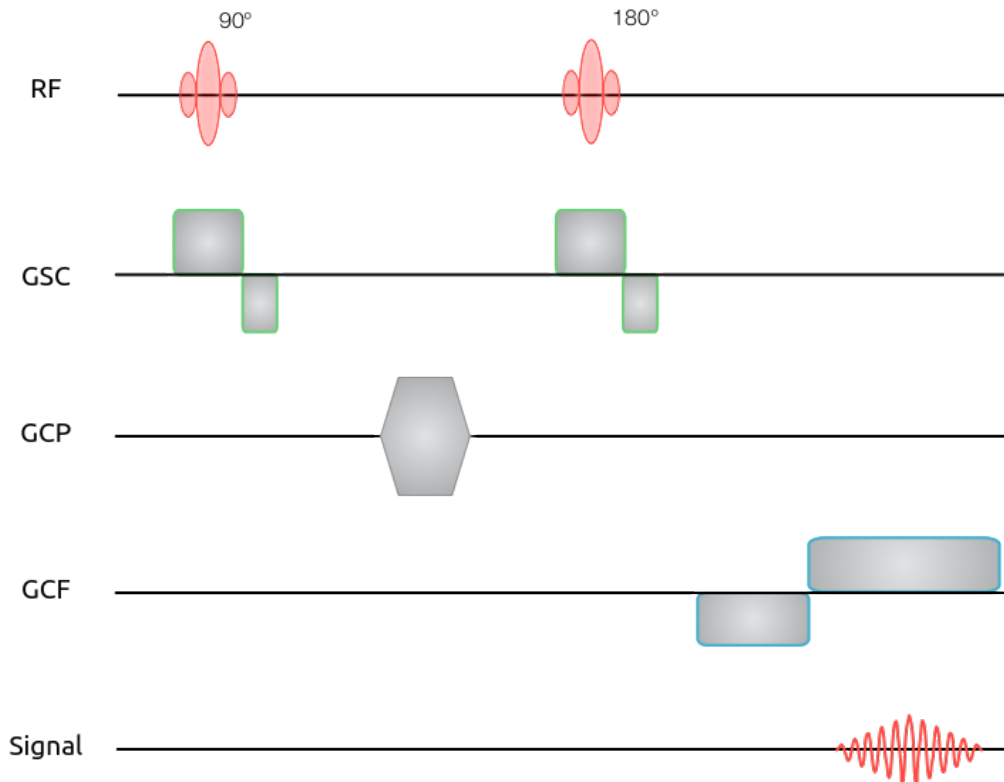


FIGURE 3.8 : Chronogramme représentant l'application des gradients de codage lors de la séquence spin-echo. Le signal réapparaît et est mesuré à TE , suite à l'impulsion de 180° , déclenchée à $TE/2$. Deux impulsions de 90° sont séparées par un TR .

Source : thèse de Mathieu Naudin [82]

Cependant, l'acquisition ne remplit pas la matrice image mais un autre espace appelé espace des K , ou espace de Fourier. L'espace de Fourier contient la représentation fréquentielle de l'image, avec au centre de l'espace les basses fréquences, et les hautes fréquences dans la périphérie (Figure 3.9). Cet espace est tel que chaque voxel contient une part de l'information globale de l'image. On retrouve ici une notion fondamentale du traitement du signal : la décomposition d'un signal en séries de Fourier.

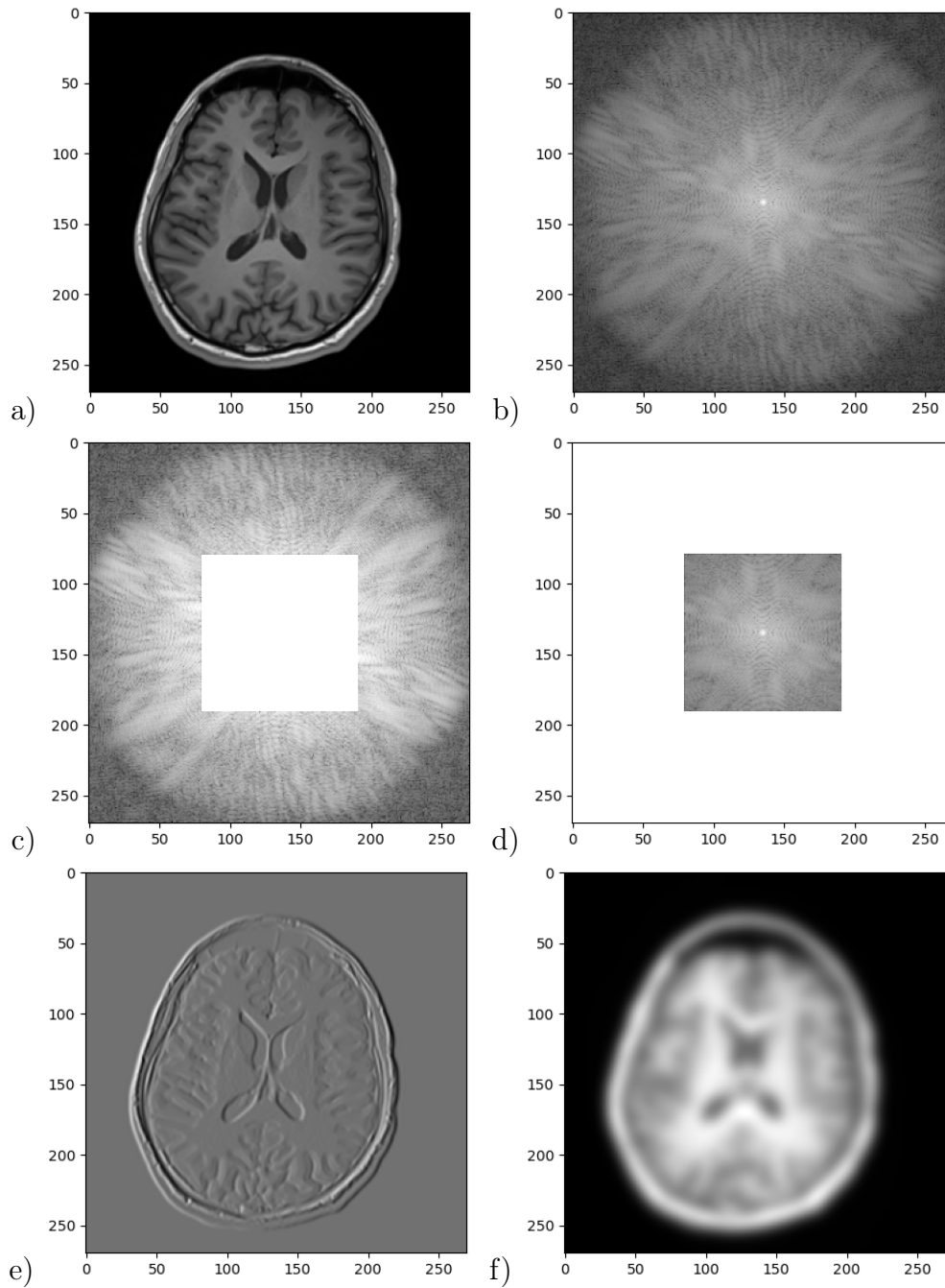


FIGURE 3.9 : Exemples d'utilisation de l'espace de Fourier, avec (a) un cerveau de docteur et (b) son espace de Fourier. (c) En filtrant le centre de cet espace (d) ou les bords, on obtient (e) uniquement les hautes fréquences ou (f) les basses fréquences de l'image. Les hautes fréquences correspondent aux variations brusques de l'image (bords) tandis que les basses fréquences représentent des zones homogènes et continues (effet de flou).

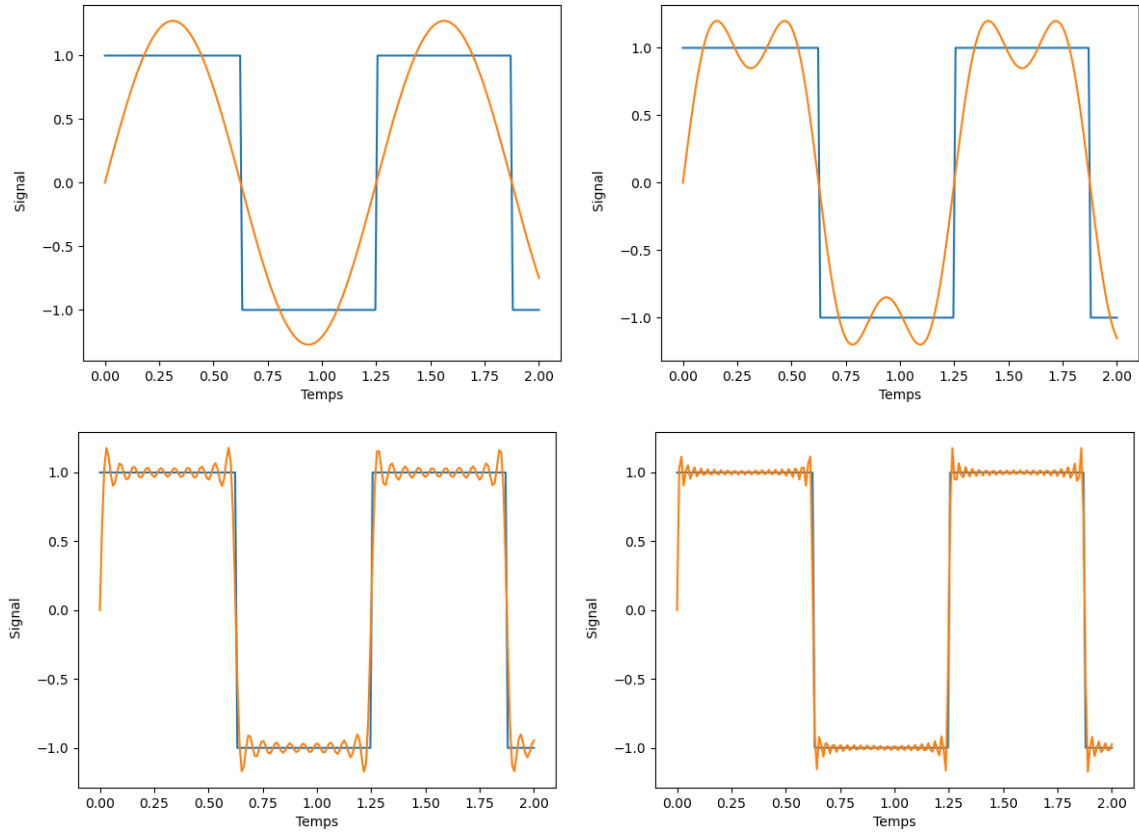


FIGURE 3.10 : *Illustration des séries de Fourier : on peut reconstruire n'importe quel signal à partir d'une somme de sinusoides. Ici, 1, 2, 10 et 20 sinus sur un signal carré.*

Il est possible de représenter n'importe quel signal à l'aide d'une somme de sinus et cosinus. Ces deux fonctions apparaissent comme des atomes que l'on peut recombinaison et pondérer pour former tout type de signal. Cette décomposition, dite en séries de Fourier, s'adapte aussi bien aux signaux à une dimension qu'aux acquisitions en deux dimensions ou plus. Un exemple est donné sur la Figure [3.10](#).

Lors de sa réception, le signal RMN, codé en phase et en fréquence, va être assigné à une valeur du plan de Fourier. La question du parcours de l'espace des K se pose alors. Lors d'une acquisition spin-echo, chaque TR permet de remplir une ligne du plan de Fourier. Plusieurs parcours de l'espace des K sont possibles. Le choix du parcours influence le type de séquence utilisé. Pour un parcours linéaire, on peut choisir de commencer à remplir les lignes centrales avant de remplir la périphérie de l'espace. Il est également pos-

sible de parcourir l'espace des K non linéairement, par exemple suivant un motif spiralé partant du centre. Des techniques de remplissage de l'espace des K peuvent raccourcir le temps d'acquisition. On peut par exemple citer les récentes évolutions concernant la technique de *finger-printing* [70].

Une fois l'espace de Fourier rempli, l'image peut être formée par transformée de Fourier inverse, dont l'équation est la suivante :

$$s(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} S(w)e^{i\omega t} d\omega \quad (3.4)$$

Avec : $s(t)$ le signal à reconstruire (l'image)

$S(w)$ le signal fréquentiel

t une variable continue

Dans le domaine discret, cette formule devient :

$$s(n) = \frac{1}{N} \sum_{k=0}^{N-1} S(k)e^{i2\pi \frac{kn}{N}} \quad (3.5)$$

Avec : $s(n)$ le signal numérique de N échantillons

$S(k)$ le signal fréquentiel

n une variable discrète

À ce point, partant de l'objet biologique "cerveau", on a sous-échantillonné physiquement (par la mesure) puis on a transformé cette mesure en un objet numérique : l'image.

3.7 Qualité du signal, qualité d'image

Ayant obtenu une image, il nous faut à présent évoquer les facteurs influant sur la qualité d'image. En effet, ces notions sont importantes lors de la mise en place d'un système d'apprentissage automatique. Pour qu'un système soit performant, il faut être en mesure

de lui fournir de bonnes images, ce qui nécessite de définir ce qu'est une image de bonne qualité.

La qualité de l'image numérique dépend de plusieurs composantes interdépendantes. Il s'agit ici de la représentation échantillonnée d'un signal physique. C'est pourquoi on identifiera des contraintes liées à la quantité de signal disponible ainsi qu'aux différents bruits qui peuvent s'y être ajoutés. L'obtention d'une image demande un échantillonnage spatial. Le volume d'acquisition 3D est alors discrétisé en un nombre fini de voxels, faisant apparaître un premier paramètre qualitatif mesurable : la résolution.

La résolution : Pour un ingénieur en traitement du signal, il s'agit du nombre d'échantillons numérisés, soit pour image 2D le nombre de lignes multiplié par le nombre de colonnes. Mais pour un radiologue, la "résolution" désigne la qualité visuelle subjective de l'image. Ainsi, une image gigantesque avec un cerveau représenté sur peu de voxels aura une très bonne résolution numérique, mais mauvaise résolution qualitative. Un exemple caricatural est montré Figure [3.11](#). Ainsi, on souhaite en réalité que les voxels soient les plus petits possibles physiquement pour décrire au mieux les variations de structures anatomiques.

Mais augmenter la résolution spatiale a un impact sur la quantité de signal présente en chaque voxel. Physiquement, un petit voxel contient moins de protons qu'un grand voxel. L'utilisation d'imageur ultra haut champ à 7 Tesla permet d'augmenter la résolution des images [\[44\]](#). Le champ \vec{B}_0 initial étant plus fort, la quantité de signal disponible augmente, ce qui permet en retour d'augmenter la résolution des images.

À champ magnétique constant, la quantité de signal disponible pour chaque voxel se réduit lorsque la taille des voxels diminue. Pour un voxel donné, cette baisse de quantité de signal recueilli impacte l'intensité du signal mesurée. C'est pourquoi un deuxième paramètre rentre en jeu dans la qualité des images : le contraste.

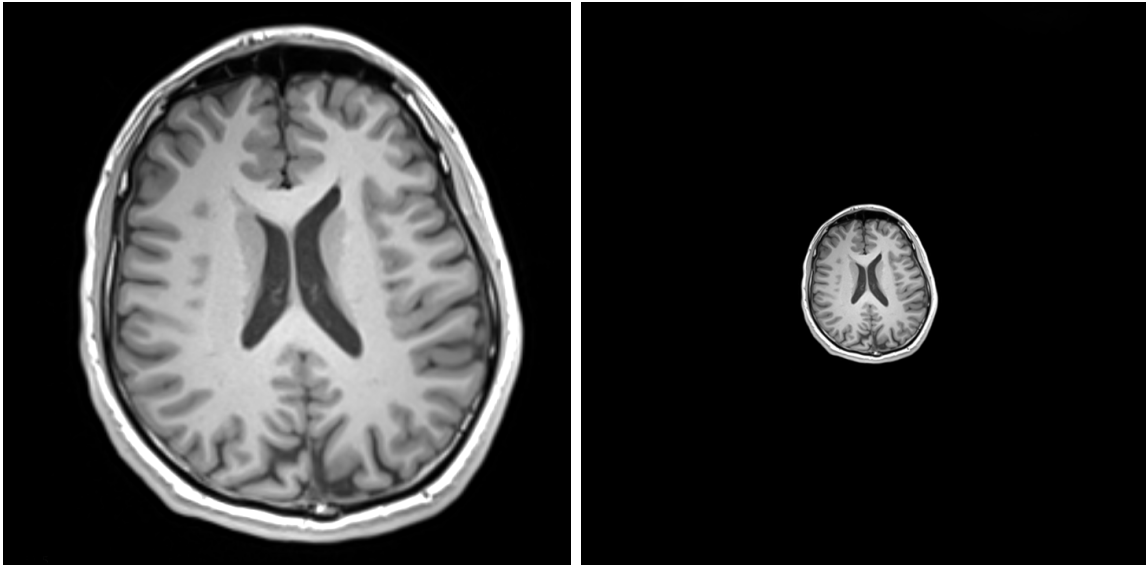


FIGURE 3.11 : Deux images de même résolution (même nombre de lignes et de colonnes), mais de résolutions (qualité d'image) différentes pour un radiologue.

$$C = \max(I(x, y)) - \min(I(x, y)) \quad (3.6)$$

Avec : C le contraste

$I(x, y)$ une image à deux dimensions x et y

Le contraste : En traitement d'image, cette notion décrit la dynamique en intensité de l'image. Lorsque le contraste est bon, son effet direct est de permettre de différencier visuellement deux structures différentes d'intensités proches. Le contraste IRM dépend de nombreux paramètres, parmi lesquels on retrouve la densité protonique, les temps de relaxation T1 et T2 des tissus, les variations du champ magnétique, ou encore les paramètres d'acquisition TE et TR.

Le bruit : Une mesure comporte toujours sa part de bruit (Figure 3.12). Sur le plan métrologique, le bruit est une information parasite issue d'un ensemble de contraintes physiques. Parmi les bruits, on distingue d'une part des bruits aléatoires issus des systèmes électroniques ou des mouvements moléculaires et les artefacts, qui sont de fortes

perturbations structurées. Dans les deux cas, ces perturbations viennent s'ajouter au signal théorique attendu et altèrent l'image résultante.

On évalue la perturbation du signal à l'aide du rapport signal sur bruit. À bruit constant, si l'on dispose de peu de signal (par exemple par choix d'une petite taille de voxel), le rapport signal sur bruit devient élevé et la qualité visuelle de l'image est dégradée. Ainsi, l'un des buts de l'acquisition consiste à trouver les paramètres optimaux réduisant le rapport signal sur bruit tout en conservant une haute résolution radiologique.

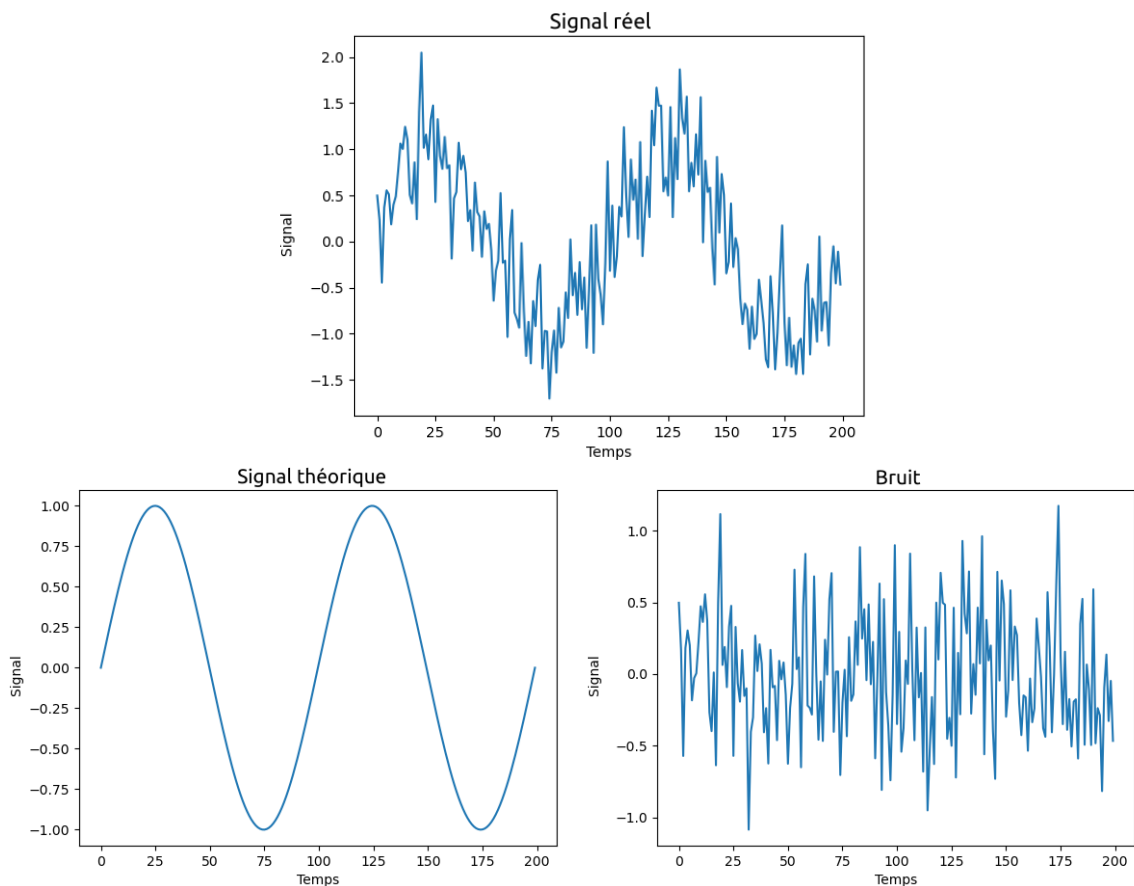


FIGURE 3.12 : *Un signal réel (haut) peut se décomposer en une composante pure (théorique, à gauche) sur laquelle est ajoutée un bruit (à droite)*

Les artefacts : Ce sont de grandes perturbations structurées. Elles sont de plusieurs types. Chaque perturbation laisse une signature caractéristique dans l'image. En IRM, on reconnaît :

- **Les artefacts métalliques** : Il s'agit d'une distorsion du champ magnétique provoquée par la présence de matériaux ferromagnétiques (appareil dentaire, prothèse) dans le volume d'acquisition. Cela provoque une absence de signal au cœur de la perturbation et une déformation non-linéaire en périphérie de perturbation. Certaines séquences luttent spécifiquement contre les artefacts métalliques.
- **Les artefacts de mouvement** : Les acquisitions IRM nécessitent que le patient reste immobile. Certains mouvements sont irréductibles, comme la respiration du patient. Le mouvement de respiration peut être corrigé par une acquisition adaptée calée sur l'expiration du patient. Pour des patients de bas âge, la sédation est souvent nécessaire. L'artefact de mouvement provoque un flou dans la zone concernée ainsi que l'apparition d'images fantômes dans le sens du codage de phase.
- **Les artefacts de repliement** : Ils apparaissent lorsque l'objet étudié est plus grand que le champ de vue. On observe alors un repliement de la partie hors-champ du côté contro-latéral de l'image. À champ de vue constant, il est possible de faire disparaître ces artefacts en changeant la taille des voxels, au détriment de la résolution spatiale, ou en appliquant un filtre anti-aliasing. Un exemple d'artefact de repliement est donné Figure [3.13](#).
- **Les artefacts de troncature** : Ils surviennent au niveau d'une rupture brutale d'intensité entre deux structures différentes, comme par exemple la transition entre l'os et la graisse. L'artefact de troncature produit une alternance de bandes sombres et claires dans le sens du codage de phase.
- **Les artefacts de flux** : Provoqués par le flux sanguin, ils ressemblent aux artefacts de mouvement en créant des images fantômes, soit de renforcement, soit d'annulation du signal. L'utilisation de bandes de saturation, ou de séquences avec gradient de compensation de flux, permettent d'atténuer ces artefacts.
- **Les artefacts de déplacement chimique** : Ils sont provoqués par les variations de la fréquence de résonance entre différents tissus, ce qui a pour effet une superposition des signaux. Cela concerne les interfaces entre tissus, comme par exemple

une transition entre la graisse et l'eau. On peut faire disparaître cet artefact par l'augmentation de la largeur de bande passante, au détriment du rapport signal sur bruit.

- **Les artefacts de susceptibilité magnétique :** Ils apparaissent aux interfaces de deux milieux ayant des sensibilités magnétiques différentes, comme par exemple la transition de l'hémoglobine aux tissus, ou de l'os au tissu. À ces interfaces, un gradient de champ magnétique intrinsèque existe, ce qui provoque des zones hypo-intenses d'autant plus marquées que le voxel est grand.
- **Le phénomène d'excitation croisée :** L'excitation des protons sur une coupe ne se limite pas aux bords de la coupe elle-même. L'impulsion RF affecte aussi les protons adjacents. Si les coupes d'acquisition sont collées l'une à l'autre, les protons se retrouvent excités à plusieurs reprises, faussant le signal. C'est pourquoi, un espacement des coupes est nécessaire pour éviter cette excitation croisée.

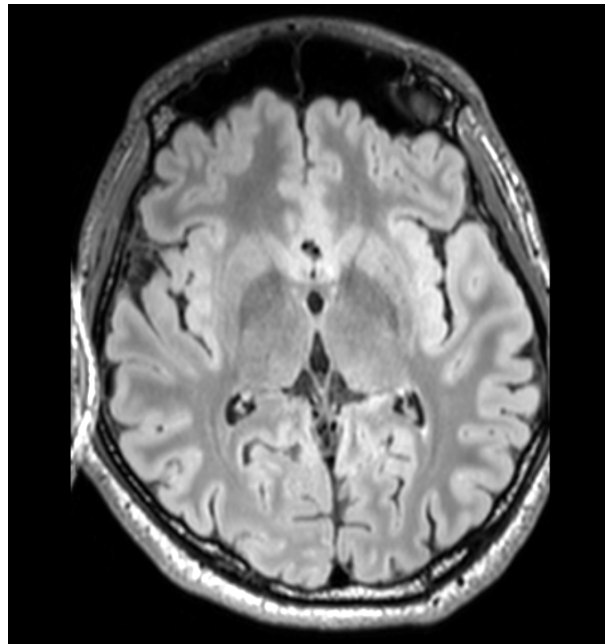


FIGURE 3.13 : *Exemple d'un artefact de repliement. La partie coupée hors du champ de vue, à droite, se reproduit sur la gauche de l'image.*

3.8 Modalités IRM et bilan lésionnel

Jusqu'à présent, on a présenté comment la formation du signal et son acquisition pouvaient donner lieu à une image. Il nous faut à présent détailler les grands types d'acquisition IRM permettant d'obtenir une grande variété d'information. Il existe un très grand nombre de séquences IRM, chacune paramétrée pour mesurer un certain type de signal. Ces signaux de valeur informationnelle différente se regroupent en familles de signaux RMN : on parlera de modalité IRM. Nous allons différencier l'IRM anatomique, de diffusion, de perfusion, l'imagerie métabolique avec la spectroscopie RMN et l'imagerie fonctionnelle. Chaque modalité IRM apporte son lot d'information pour caractériser les différents types de gliomes.

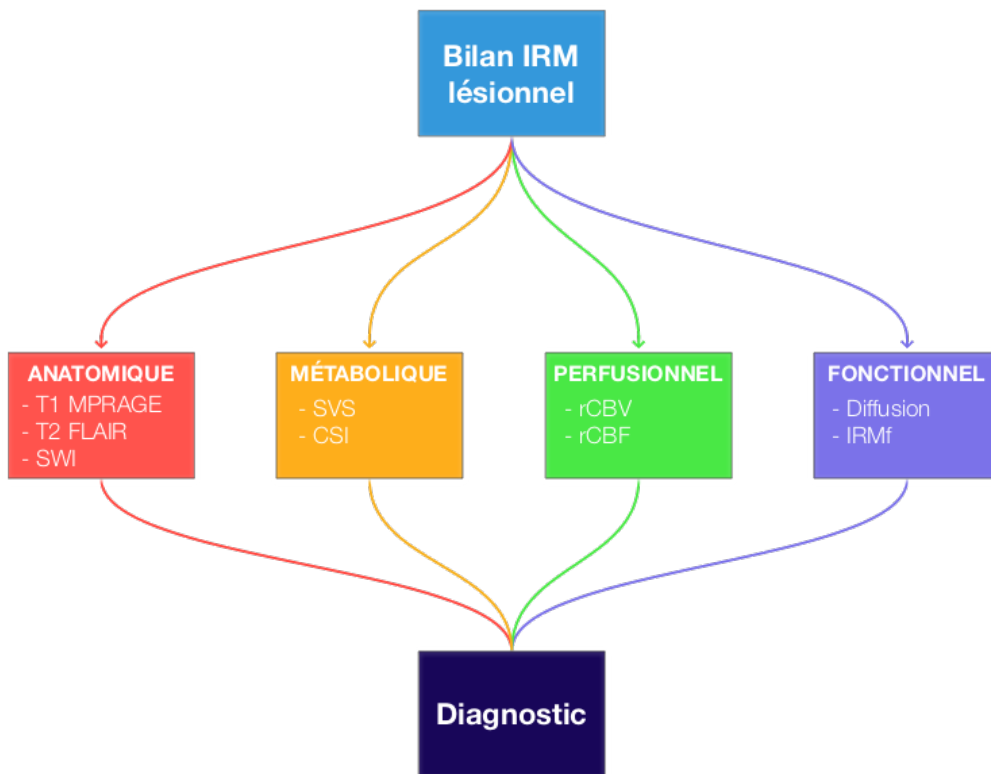


FIGURE 3.14 : Le bilan lésionnel du CHU de Poitiers combine les différentes modalités IRM pour aboutir au diagnostic.

Source : thèse de Mathieu Naudin [82]

Ces différentes modalités sont utilisées au CHU de Poitiers lors du bilan lésionnel initial et lors du suivi du patient (Figure 3.14). L'enjeu de cette diversité de séquences est de pouvoir combiner les différentes informations pour mieux caractériser les lésions et leur évolution. Conséquence de ce bilan : le temps d'examen est long, demandant presque une heure et demi par patient pour un bilan complet. C'est pourquoi, un traitement sur ces nombreuses données doit permettre d'explorer la pertinence de ces séquences, qu'il s'agisse de soutenir leur présence dans le bilan par de nouveaux arguments, ou au contraire de pointer un faible apport au diagnostic.

3.8.1 IRM anatomique

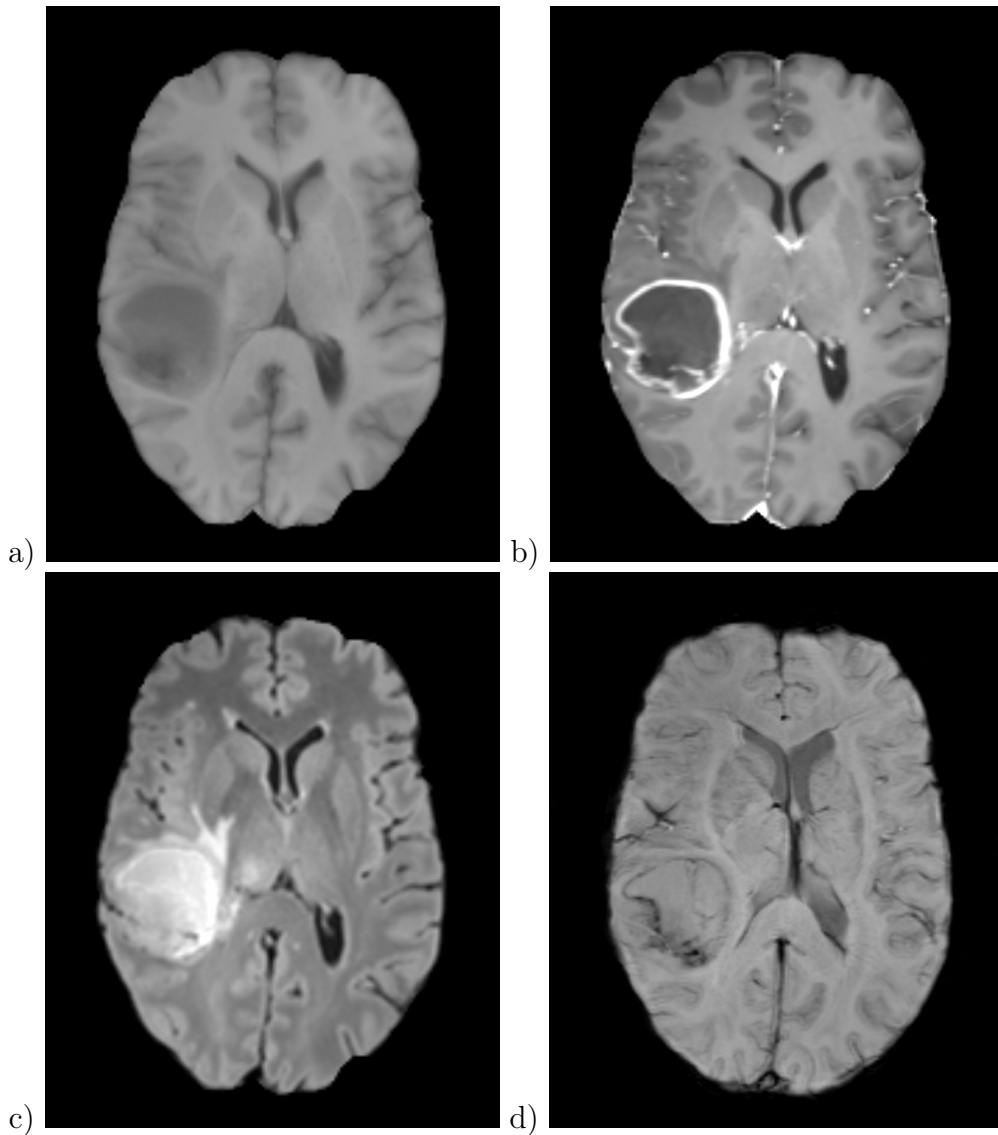


FIGURE 3.15 : *Acquisitions T1 (a), T1 post injection de gadolinium (b), T2 FLAIR (c) et SWI (d).*

Les séquences d'IRM anatomiques permettent d'obtenir les informations structurelles topologiques de la lésion. On trouve dans cette modalité les premières acquisitions de l'examen. Elles permettent entre autre d'identifier la présence d'une lésion et de caractériser son étendue, ainsi que fournir un premier niveau de différenciation tissulaire à

l'intérieur de la lésion même. Les cartes de pondération T1 et T2 font partie des séquences anatomiques. Un T1 après injection de gadolinium, un agent de contraste, est également acquis. Nous allons aussi présenter la séquence FLAIR utilisée pour l'imagerie lésionnelle. On finira par une séquence caractérisant la susceptibilité magnétique des tissus : la cartographie *Susceptibility Weighted Imaging* (SWI).

Une perfusion de gadolinium Gd^{3+} est utilisée comme produit de contraste. Elle altère les temps T1 et T2 des tissus dans lesquels le gadolinium se propage, ce qui permet de faire apparaître le système vasculaire.

La séquence *FLuid Attenuated Inversion Recupération* (FLAIR) utilise le principe d'Inversion Récupération pour supprimer le signal de l'eau libre et du LCR [29]. Les séquences utilisant l'inversion récupération font basculer totalement l'aimantation longitudinale par une impulsion RF de 180° . Lors de la repousse T2, cette aimantation va passer par un zéro. On appelle TI, Temps d'Inversion, le temps entre l'impulsion de 180° et le passage à 0 de la repousse longitudinale d'un tissu. Le TI est dépendant du T2 d'un tissu, c'est pourquoi un ciblage spécifique permet d'exclure spécifiquement un tissu lors du recueil du signal. La séquence FLAIR supprime l'eau libre, mais d'autres paramètres permettent de supprimer par exemple le signal de la graisse (séquence STIR pour *Short TI Inversion Recovery*).

Caractéristiques visuelles : Trois exemples d'acquisitions anatomiques sont données sur la Figure 3.15. Sur une carte T1, la substance blanche est blanche, la substance grise est grise et le LCR est noir. Inversement, lors d'une pondération T2, la substance grise apparaît blanche, la substance blanche est sombre et le LCR est très blanc. Le caractère infiltrant du gliome, de par sa propagation diffuse au sein des tissus, provoque l'apparition d'œdème. Cet œdème apparaît blanc en pondération T2. Pour bien le distinguer de l'hypersignal T2 du LCR, la séquence FLAIR permet d'annuler le signal du LCR, ce qui fait ressortir les contours de la lésion. De ce fait, la cartographie FLAIR est la séquence de référence de l'imagerie lésionnelle, permettant d'obtenir une estimation de la segmentation tumorale complète. Les séquences anatomiques permettent aussi de visualiser l'effet de masse, c'est à dire la déformation des tissus environnants due à la

croissance tumorale.

Caractériser le cœur tumoral se fait à l'aide des séquences T1 et T1 post-gadolinium. En effet, lors de l'injection de gadolinium, l'agent le contraste se propage dans le système vasculaire, créant sur son passage un hypersignal T1. Le cœur tumoral, au fur et à mesure de son évolution, va progressivement mobiliser le système vasculaire pour maximiser son apport énergétique. Cette mobilisation va créer de nouveaux vaisseaux proche du cœur tumoral. Il peut également y avoir rupture de la barrière hémato-encéphalique. Le passage du gadolinium va laisser un fort hypersignal T1 sur l'acquisition post-gadolinium. Ainsi, la comparaison entre le T1 initial et le T1 après injection va permettre d'identifier la zone active de la tumeur. De plus, la présence de nécrose, très sombre en T1 après injection, contraste visuellement fortement avec la périphérie active du cœur tumoral. On observe alors dans les gliomes avancés une couronne de réhaussement entourant un hyposignal nécrotique caractéristique d'une lésion agressive.

Enfin, la séquence SWI donne au radiologue des informations sur la vascularisation de la lésion. A haut champ, cette information peut être corrélée avec les données de perfusion, ce qui permet d'estimer le grade des lésions gliales. [90]

3.8.2 IRM de diffusion

Découverte en 1985, l'IRM de diffusion permet d'étudier le déplacement de l'eau dans le cerveau [64]. Selon les structures cérébrales, comme les fibres, la diffusion peut présenter une direction préférentielle. Ainsi, cette modalité IRM permet non seulement de décrire la diffusion cérébrale via des cartes paramétriques, mais aussi de reconstruire des grands faisceaux de fibres du cerveau.

Dans l'encéphale, l'eau peut être libre comme dans les ventricules, ou se déplacer dans une direction contrainte comme dans les fibres de la substance blanche. Les molécules d'eau suivent un mouvement brownien : elles avancent en ligne droite jusqu'à percuter une paroi ou une autre molécule. En l'absence de contraintes, les molécules peuvent se déplacer dans toutes les directions. Il s'agit d'une diffusion isotropique. Mais ce mouvement peut être contraint par l'environnement dans lequel évolue les molécules,

ce qui peut avoir pour effet de privilégier certaines directions. On parle alors de diffusion anisotrope. Ces deux diffusions sont représentées dans la Figure [3.16](#)

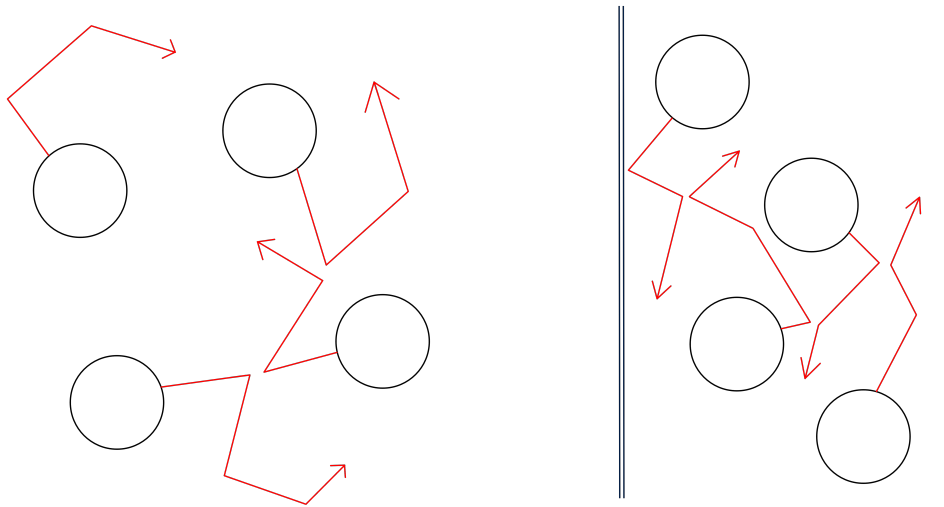


FIGURE 3.16 : (À gauche) diffusion isotropique : Les molécules évoluent librement dans toutes les directions. (À droite) diffusion anisotropique : dans un environnement contraint, une direction de diffusion préférentielle apparaît.

Pour capter le signal représentatif de la diffusion, il faut appliquer des gradients supplémentaires appelés gradients de diffusion, décrits par Stejskal et Tanner [\[113\]](#). L'effet de ces gradients est caractérisé par un facteur de gradient "b", tel que :

$$b = (\gamma G \tau)^2 \left(\Delta T - \frac{\tau}{3} \right) \quad (3.7)$$

Avec : b le facteur de gradient (en $s.mm^{-2}$)

γ le rapport gyromagnétique

G l'amplitude du gradient

τ la durée d'application du gradient (en s)

ΔT Le temps entre deux gradients de diffusion (en s)

La séquence de référence pour l'IRM de diffusion est l'*Echo Planar Imaging - Spin Echo* (EPI-SE). Il s'agit d'une technique d'imagerie ultrarapide sensible aux mouve-

ments de flux et micro-flux. Cette sensibilité renforce l'observation du phénomène. De plus, lorsqu'il prend de hautes valeurs, le facteur de gradient b accroît la sensibilité de la séquence au phénomène de diffusion. Toutefois, cette augmentation de b , qui peut passer de 0 mm.s^{-2} à 5000 mm.s^{-2} selon l'imageur, sollicite les machines de manière importante.

L'application des gradients doit être effectuée dans de multiples directions pour décrire toutes les variations d'anisotropie. A partir de six directions, il est possible de calculer le tenseur de diffusion. Ce tenseur décrit une ellipsoïde 3D et est une structure mathématique dont on peut calculer des vecteurs propres et des valeurs propres de diffusion :

$$\overline{\overline{D}} = \begin{pmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{pmatrix} \rightarrow \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} \quad (3.8)$$

Avec : les $D_{..}$ les valeurs de diffusion dans les différentes directions
les λ_n les valeurs propres de diffusion

À partir du tenseur de diffusion, il est possible de calculer des cartes paramétriques propres aux informations diffusionnelles (Figure 3.17). On parlera alors d'imagerie du tenseur de diffusion (DTI pour *Diffusion Tensor Imaging*).

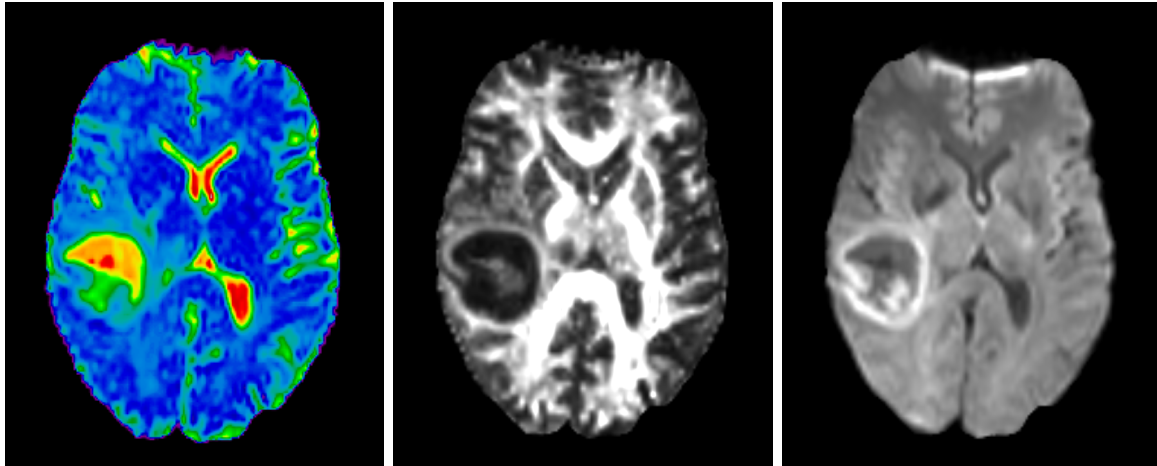


FIGURE 3.17 : Cartographies d'IRM de diffusion. De gauche à droite : ADC, de FA et image Trace.

Le coefficient de diffusion apparent (ADC) : Cette première cartographie représente en chaque point le coefficient de diffusion apparent (ADC pour Apparent Diffusion Coefficient). Ce coefficient est une grandeur physique, exprimée en mm^2/s . Il exprime l'intensité de la diffusivité de l'eau à travers les tissus. Sa valeur est grande lorsque la diffusivité est élevée. Il est calculé à partir de deux valeurs de b par la formule suivante :

$$ADC = \frac{\ln\left(\frac{S_0}{S_1}\right)}{(b_1 - b_0)} \quad (3.9)$$

Avec : b_0 et b_1 deux valeurs du facteur de gradient
 S_0 et S_1 les valeurs d'un pixel à b_0 et b_1

L'anisotropie fractionnelle (FA) : Valeur comprise entre 0 et 1, elle exprime l'intensité des contraintes à la diffusion en un voxel. Cette cartographie peut être calculée en deux versions : une version en niveaux de gris exprimant un pourcentage global d'anisotropie, et une version colorée où chaque canal couleur RGB code une direction de l'espace. On parle alors d'Anisotropie Fractionnelle colorée (ou ColFA). Il en résulte une carte où les grands faisceaux sont visibles, et dont l'orientation est identifiable via un code couleur. Par convention, le canal bleu code l'axe inférieur-supérieur, le canal vert

l'axe antéropostérieur, et le canal rouge l'axe gauche-droite. Le calcul de la cartographie implique les valeurs propres du tenseur de diffusion et est le suivant :

$$FA = \sqrt{\frac{3}{2}} \times \sqrt{\frac{(\lambda_1 - \bar{\lambda})^2 + (\lambda_2 - \bar{\lambda})^2 + (\lambda_3 - \bar{\lambda})^2}{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}} \quad (3.10)$$

Avec : λ_n les valeurs propres issues du tenseur de diffusion
 $\bar{\lambda}$ la valeur moyenne des lambda.

De cette information colorée, il est possible de reconstruire l'architecture des fibres dans le cerveau. Cette technique, issue d'un tracking des fibres s'appelle tractographie et mets en avance les grands faisceaux, supports de la connexité macroscopique dans le cerveau.

Image Trace : Enfin, l'application des gradients de diffusion dans les trois directions de l'espace permet de calculer une image de diffusion isotrope, appelée image trace. L'image trace devient hyperintense lorsque la diffusion diminue.

Intérêt diagnostic : L'imagerie de diffusion est particulièrement pertinente pour l'étude des tumeurs infiltrantes, permettant d'identifier des directions préférentielles de propagation. Elle permet aussi de mettre en évidence les évolutions à l'interface entre tumeur et tissu sain : déviation, infiltration et rupture de faisceaux apparaissent comme des indicateurs du développement tumoral.

3.8.3 IRM de perfusion

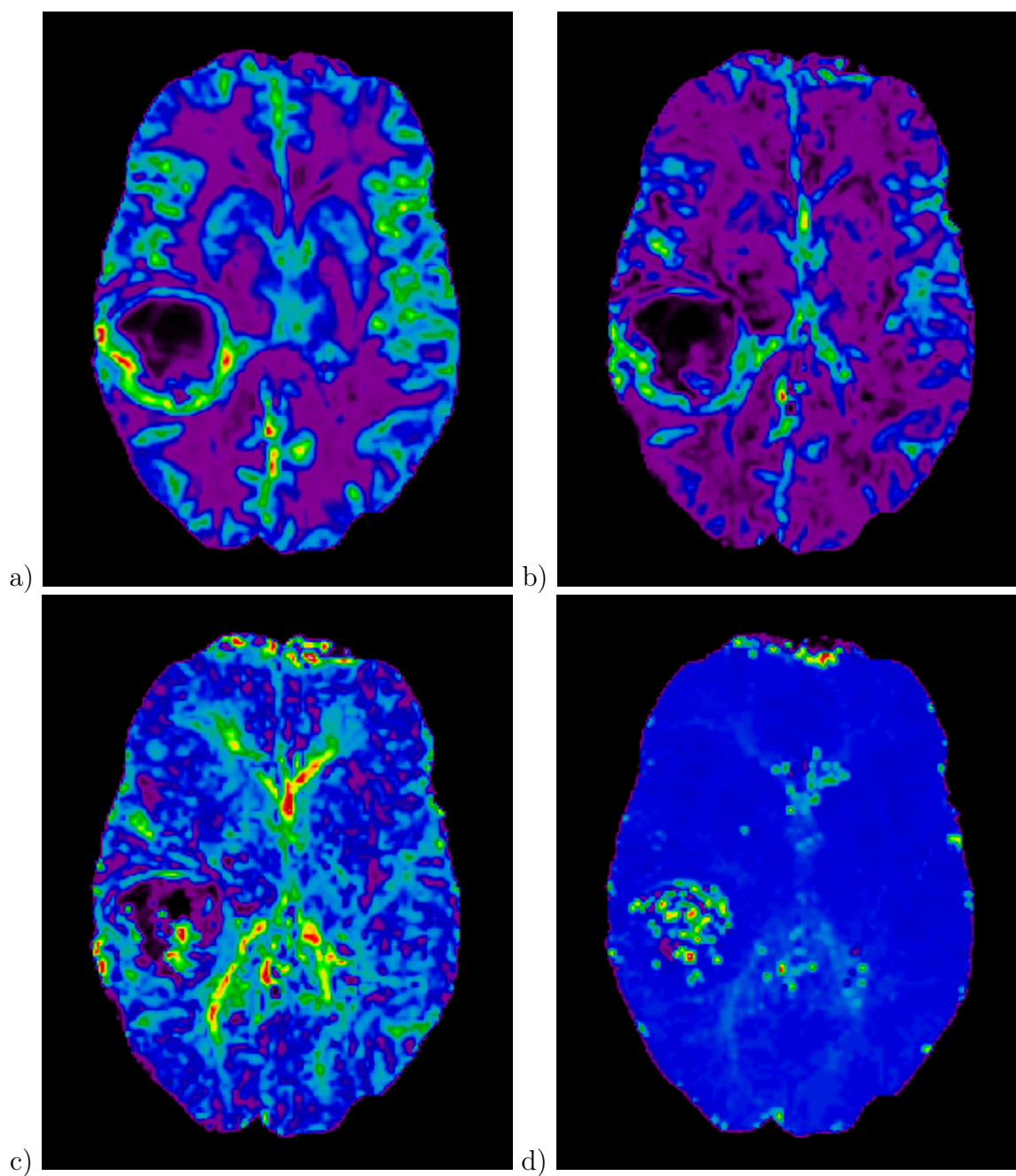


FIGURE 3.18 : Cartographies d'IRM de perfusion. Ces cartes paramétriques sont calculées à partir de la modélisation de la courbe de premier passage du bolus de gadolinium, avec : (a) le flux sanguin cérébral (CBF) , (b) le volume sanguin cérébral (CBV), (c) le temps de transit moyen (MTT) et (d) temps de pic (TTP)

L'IRM de perfusion permet de caractériser le flux sanguin cérébral dans l'encéphale au travers de cartes paramétriques. L'obtention de ces informations demande une technique d'acquisition particulière, invasive lorsqu'il y a injection d'un produit de contraste.

Certaines séquences utilisent une perturbation de la susceptibilité magnétique via l'injection d'un produit de contraste. Le volume injecté, dit bolus, va circuler dans le système vasculaire, altérant les propriétés magnétiques des tissus environnants sur son passage. La mesure de la perturbation magnétique va être mesurée via la fonction d'entrée artérielle (AIF). Cette fonction permet de revenir à la concentration du produit de contraste en une zone donnée par rapport au temps. Cette courbe est caractérisée par plusieurs paramètres que l'on va détailler. Le calcul de ces paramètres permet de générer des cartes paramétriques : ce sont les cartes de perfusion.

Parmi les paramètres calculés, on trouve le volume sanguin cérébral (CBV), le flux sanguin cérébral (CBF), le temps de transit moyen (MTT) et le temps de pic (TTP). Bien que ces cartes paramétriques soient analogues à des images en niveaux de gris, on les représente souvent avec une échelle de fausses couleurs lors de l'interprétation.

Ces cartes ne sont pas quantitatives mais qualitatives. Leur interprétation se fait à travers l'utilisation de zones d'intérêts (ROI pour *Region of Interest*), en comparant les valeurs dans la lésion puis dans la zone contro-latérale, en symétrique par rapport au plan interhémisphérique (à une variation près). Le rapport entre les deux valeurs (ROI lésionnelle et controlatérale) qui informe le radiologue sur le paramètre perfusionnel de la lésion.

Le caractère invasif de l'IRM de perfusion n'est pas obligatoire. Citons comme exemple de technique non invasive la labellisation de spin artériel (ASL). Au CHU de Poitiers, c'est le contraste dynamique de susceptibilité (DSC) qui est utilisé, avec une injection de Gadolinium.

Apport au diagnostic : Comme présenté section [2.3](#), le gliome va mobiliser le système vasculaire pour maximiser son apport énergétique. Par conséquent, la zone tumorale va, au cours de l'évolution du gliome, passer d'une hypoperfusion à une hyperperfusion. Cette évolution est visible sur la carte de CBV. Un rapport peut être calculé entre la

valeur de CBV dans la zone lésionnelle et zone controlatérale. Ce rapport, le CBV relatif (rCBV) est un marqueur pour le gliome de haut grade s'il est supérieur à 1,75 [3]. Le rapport du CBF relatif (rCBF) peut également être calculé, et devient un indice de haut grade pour un seuil de 1,25 [49].

Aspect des images : Lors de l'acquisition, la séquence de perfusion cherche à caractériser un évènement bref : le passage du bolus dans le système vasculaire. Cela demande d'acquérir plusieurs images à des temps rapprochés. Par conséquent, la résolution de la séquence est faible et le rapport signal sur bruit est bas. Les cartes paramétriques présentent donc un bruit important.

3.8.4 Spectroscopie RMN

La spectroscopie RMN est une modalité d'acquisition qui permet de mesurer la concentration de différentes molécules dans un tissu. Dans le milieu biologique, ces molécules sont des métabolites. Dans le champ magnétique, les protons réagissent différemment selon leur environnement moléculaire proche. La répartition des charges au sein des différents atomes composant une molécule altère localement ses propriétés magnétiques. Par conséquent, chaque molécule présente une fréquence de résonance qui lui est propre, ce qui permet de l'identifier.

Cette variation de fréquence porte le nom de déplacement chimique. Le déplacement chimique δ est une grandeur sans dimension dont l'unité cardinale est la partie par million (ppm), calculée par rapport à une molécule de référence : le tétraméthylesilane $Si - (CH_3)_4$.

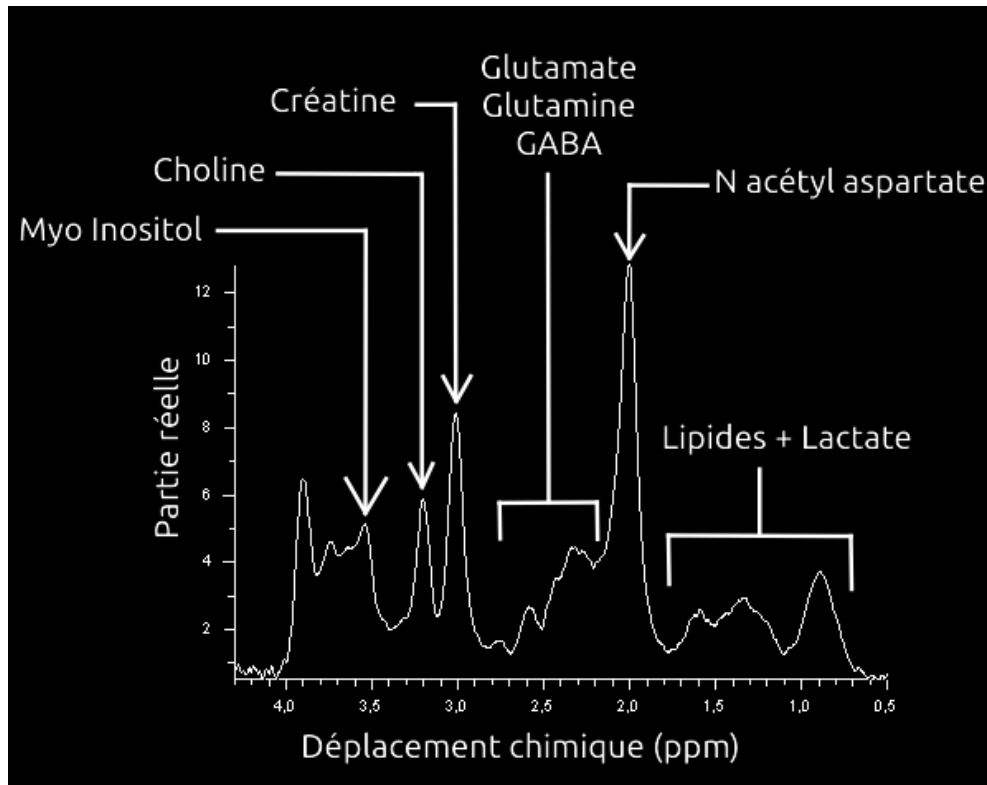


FIGURE 3.19 : Une acquisition de spectroscopie par résonance magnétique (SRM), montrant les différents métabolites cérébraux identifiables à TE court. L'aire sous chaque pic indique la concentration du métabolite étudié. Pour une comparaison inter-patient, on utilise des rapports métaboliques, comme par exemple le ratio Choline/Créatine.

L'acquisition ne produit pas sur une image, mais un spectre (Figure 3.19). Ce spectre, qui se lit par convention de droite à gauche, fait apparaître les signaux de résonance des molécules à un déplacement chimique donné sous la forme d'un pic. Ainsi, chaque pic correspond à une molécule particulière, et l'amplitude du pic, liée à la quantité de signal mesurée, permet de quantifier la concentration moléculaire du tissu étudié. Notons qu'une même molécule peut contenir plusieurs protons, eux-mêmes situés dans différents environnements moléculaires. Cela a pour effet de dédoubler les pics de résonance. On parlera alors de doublets, de triplets, voire quadruplets, tandis qu'un simple pic produira un singulet.

$$\delta = \frac{\omega_n - \omega_{ref}}{\omega_{ref}} \times 10^6 \quad (3.11)$$

Avec : δ le déplacement chimique

ω_n la fréquence de résonance d'un métabolite

ω_{ref} la fréquence de résonance de référence, le $Si - (CH_3)_4$

Pour être analysable, un spectre nécessite que le pic de l'eau, dont l'amplitude est la plus forte due à la grande concentration en eau des tissus biologiques, soit supprimé. Le signal du pic de l'eau, situé à 4.7 ppm, est tellement fort qu'il écrase visuellement les autres métabolites. Sans sa suppression, l'interprétation de l'information résiduelle ne permet pas la mesure des différentes concentrations.

La résolution fréquentielle du spectre dépend de l'intensité du champ magnétique. Ce dernier doit être supérieur à 1.5 T pour permettre la spectroscopie *in vivo*. Mais travailler à plus haut champ apporte son lot de contraintes techniques, en particulier concernant l'uniformité du champ magnétique. Une inhomogénéité de champ peut perturber la ligne de base du spectre, ce qui dégrade sa qualité lors de l'interprétation.

Les métabolites visibles ne sont pas les mêmes selon la valeur du TE. Deux valeurs sont utilisées au CHU de Poitiers, correspondant à un TE court (30 ms) ou long (135 ms).

Les acquisitions de spectroscopie sont de deux types : soit une acquisition d'un seul voxel (SVS pour *Single Voxel Spectroscopy*), soit l'acquisition d'une coupe qui peut donner lieu à une image paramétrique (CSI pour *Chemical Shift Imaging*). Un exemple de SVS est montré Figure [3.20](#). Une CSI est montrée Figure [3.21](#).

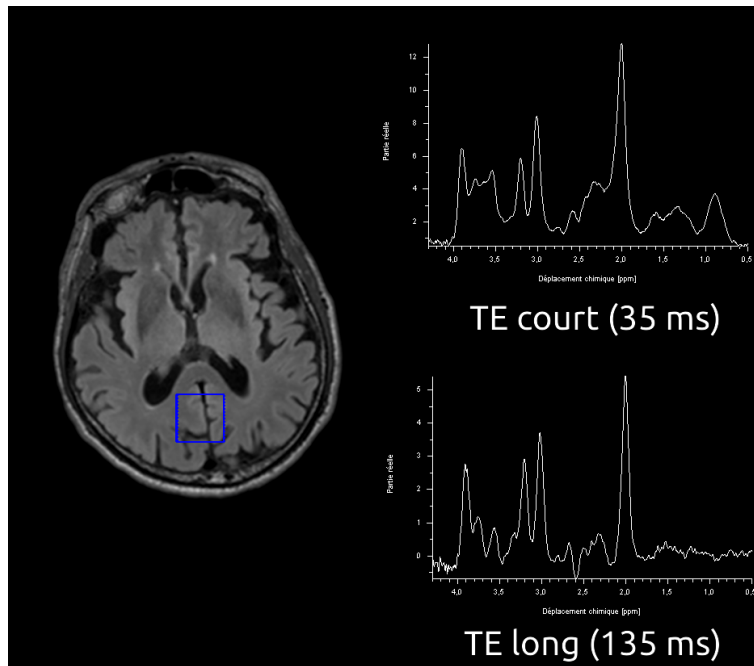


FIGURE 3.20 : Une acquisition SVS et son spectre à TE court et TE long. Le voxel d'acquisition, unique, est ici situé dans la partie médiane postérieure.

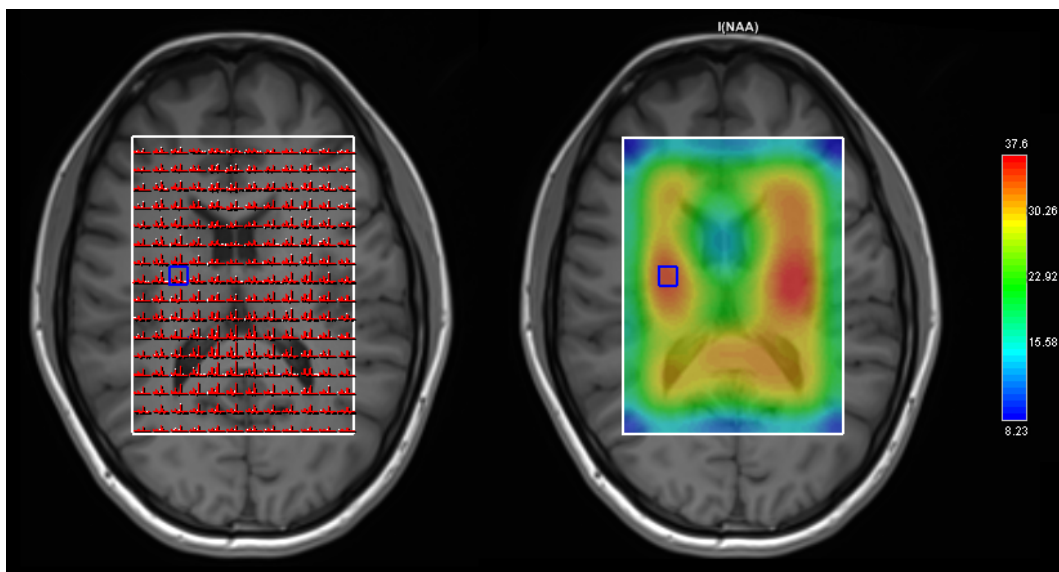


FIGURE 3.21 : Une acquisition CSI. Relever les valeurs métaboliques dans chaque voxel permet de créer une carte paramétrique de concentration. (a) Chaque voxel contient une acquisition spectrale. (b) carte paramétrique montrant la concentration en NAA dans un cerveau sain.

Métabolites étudiés : En spectroscopie du proton, les principaux métabolites observables sont :

- **Le N-Acetyl Aspartate (NAA) :** Il résonne à 2.01 ppm et résonne aussi bien à TE court qu'à TE long. Le NAA est un acide aminé transporté dans les axones. Une forte concentration suggère une dégradation neuronale : c'est donc un indicateur de la souffrance neuronale.
- **La choline (Cho) :** Elle résonne à 3.26 ppm. Il s'agit d'un phospholipide membranaire, molécule de base de la bicouche lipidique cellulaire. Sa présence en grande quantité indique un renouvellement membranaire accru, qui est un indice de la forte multiplication cellulaire engendrée par la tumeur.
- **La créatine (Cr) :** Elle résonne par deux singulets éloignés, l'un à 3.03 et l'autre à 3.93 ppm. Elle est source de phosphate et joue donc un rôle dans l'énergétique cellulaire. Son pic n'est pas affecté dans un contexte pathologique et sert donc de référence. En particulier, la créatine est utilisée dans le calcul de rapports métaboliques, ce qui permet une normalisation et une comparaison inter-patient des données de spectroscopie.
- **Le lactate (Lac) :** Il résonne par un doublet centré sur 1.33 ppm. Les pics du doublet sont proches : séparés par 7Hz, ils ne sont distants que de quelques centièmes de ppm à 3 Tesla. Le lactate est un des produits de la glycolyse anaérobie, c'est à dire en absence d'oxygène. Ainsi, sa concentration augmente rapidement en cas d'hypoxie ou d'ischémie cérébrale.
- **Les lipides (Lip) :** Ils résonnent par deux singulets distants, à 1.05 et 1.20 ppm. Ces lipides sont marqueurs de la nécrose cellulaire, un phénomène présent dans les gliomes de stades avancés.
- **Le complexe glutaminergique (Glx) :** Il s'agit d'un ensemble de molécules résonnant par plusieurs pics proches. On identifie la glutamine (Gln), le glutamate (Glu) et l'Acide γ -AminoButyrique (GABA). Ces métabolites résonnent entre 2.00

et 2.46 ppm. L'identification de chaque molécule du complexe glutaminergique demande une grande résolution spectrale, donc un haut champ magnétique.

- **Le myo-Inositol (mI)** : Il résonne à 3.5 ppm et est un marqueur de souffrance gliale impliqué dans les échanges transmembranaires. Son augmentation est un signe d'une gliose, c'est à dire une multiplication des cellules de la glie.

Enfin, la spectroscopie RMN permet d'étudier la résonance de différents atomes selon l'antenne utilisée. On l'a dit, le phénomène de résonance magnétique peut concerner plusieurs noyaux, bien que ce soit le noyau d'hydrogène qui soit le plus étudié. Mais l'acquisition d'antennes supplémentaires permet d'explorer la résonance du phosphore, du sodium ou du carbone. En particulier, les spectres du phosphore permettent d'analyser le fonctionnement énergétique de la cellule et le pH cellulaire.

Les informations de spectroscopie permettent d'explorer le tissu cérébral au niveau moléculaire. Ce niveau d'information nous renseigne sur la dynamique cellulaire (multiplication, apoptose), ce qui permet d'affiner le diagnostic. Ces informations sont en effet très complémentaires de l'analyse obtenue par imagerie. En particulier, les variations du niveau moléculaire s'inscrivent dans une temporalité différente des changements macroscopiques. Plusieurs études montrent que les changements moléculaires précèdent les évolutions tissulaires. L'intégration des informations de spectroscopie donne donc une dimension prédictive au diagnostic.

3.8.5 IRM fonctionnelle

Citons enfin brièvement l'IRM fonctionnelle. Dernière modalité IRM utilisée au CHU de Poitiers, l'IRM fonctionnelle (IRMf) permet d'explorer l'activation des zones cérébrales lors d'une tâche demandée ou au repos. Cette technique d'imagerie utilise l'effet Blood Oxygen-Level Dependant (effet BOLD). Cet effet apparaît lorsqu'une zone du cerveau s'active pour répondre à une tâche. La zone consomme alors plus d'oxygène pour répondre à ses besoins énergétiques. Cet oxygène est apporté par l'hémoglobine. Sa réduction donne la désoxyhémoglobine (DHG), dont la concentration augmente dans

la zone sollicitée. La DHG est paramagnétique : elle s'aligne dans le champ magnétique, ce qui permet de mesurer indirectement le signal de la zone activée.

Deux types de cartographies sont alors possibles. La première consiste à mesurer l'activité cérébrale au repos. On parle de *resting state*, littéralement "état de repos". Le *resting state* montre l'activation des zones actives même lorsqu'aucune tâche particulière n'est effectuée. On appelle ces zones actives à l'état d'éveil le réseau par défaut (*default network*) [73]. Il s'agit de la référence par rapport à laquelle on va ensuite pouvoir mesurer une activité lorsqu'une tâche est réalisée par le patient. Ces informations permettent de cartographier spatialement des zones fonctionnelles mobilisées par le patient pour répondre à un stimulus. Ainsi, le neurochirurgien disposera d'une cartographie patient-spécifique lors de l'opération, ce qui lui permettra de limiter plus finement les pertes fonctionnelles (parole, motricité...) pour le patient.

De part l'intégration de toutes ces différentes modalités, l'IRM sort de son statut de simple imageur anatomique et devient un outil d'exploration multimodal et multiparamétrique. Chaque modalité supplémentaire ouvre une dimension propre dans l'analyse et le suivi des pathologies. Cependant, cet élargissement du champ de la mesure s'accompagne de coûts supplémentaires. Ces coûts sont de trois natures : coût matériel, coût temporel et coût dans la prise en charge du patient. Concernant le coût matériel, il faut noter que l'équipement, comprenant l'imageur et les différentes antennes, est parfois onéreux. Le coût en temps, dit "temps machine", comprend la durée d'acquisition des séquences mais aussi le temps de prise en charge du matériel et du patient. Accélérer le temps d'acquisition constitue un axe de recherche privilégié dans la réduction du temps machine. Enfin, la capacité du patient à rester immobile durant de longues périodes dans un environnement étroit, tout en requérant parfois sa participation active (retenir sa respiration, effectuer une tâche d'IRMf...), créent des contraintes supplémentaires à sa bonne prise en charge.

Il y a donc un intérêt à dégager, parmi ce bilan multimodal, la valeur informationnelle ajoutée par ces acquisitions, dans un arbitrage coûts-bénéfices. Cette valeur informationnelle intègre l'intérêt scientifique propre à la recherche (découvrir ce qui est nouveau),

mais aussi l'impérieuse nécessité clinique de l'utilité de ces informations dans le parcours de soin.

Apprentissage automatique

Contenu du chapitre

4.1 Introduction	74
4.2 Anatomie d'un schéma d'apprentissage automatique	76
4.3 Apprentissage automatique classique	77
4.3.1 Statistiques, mesures et représentations	78
4.3.2 Types d'apprentissages	86
4.3.3 Les problèmes d'optimisation	87
4.3.4 Divisions de la base de données	90
4.3.5 Complexité du modèle : le sous-apprentissage et le sur-apprentissage	91
4.3.6 La validation croisée	92
4.3.7 Paramétrage fin : paramètres et hyperparamètres	94
4.3.8 Matrice de confusion et métriques d'évaluation d'un classifieur	95
4.3.9 Exemples d'algorithmes	96
4.4 Apprentissage profond (<i>deep learning</i>)	103
4.4.1 Analyse structuraliste des réseaux d'apprentissage profond	105

Dans cette partie, nous présentons les outils que nous allons appliquer sur les données d'IRM pour analyser et décrire l'évolution des gliomes. Ces outils proviennent du champ scientifique de l'intelligence artificielle, que nous aborderons via le terme d'apprentissage automatique. Nous différencions les méthodes d'apprentissage automatique classique des méthodes d'apprentissage automatique profond. Après une présentation de notions essentielles de statistiques et de traitement des données, nous détaillerons plusieurs algorithmes d'apprentissage automatique classiques, ainsi que plusieurs structures d'apprentissage automatique profond.

4.1 Introduction

Dans son *Invitation à la théorie de l'information* [36], Emmanuel Dion classe les révolutions techniques en deux catégories : les révolutions liées à l'exploitation de nouvelles formes d'énergie d'une part et les révolutions liées à un nouveau traitement de l'information d'autre part. Dans la première catégorie, on identifie par exemple la maîtrise du feu ou l'invention de la machine à vapeur. La seconde catégorie regroupe l'invention de l'écriture, de l'imprimerie, ou encore des techniques modernes de télécommunication. L'apprentissage automatique, que nous appellerons aussi *machine learning*, entre dans cette seconde catégorie liée à une exploitation nouvelle de l'information.

En effet, là où la robotique et l'automatique permettent une rétroaction mécanique (par l'asservissement des systèmes), le concept d'apprentissage machine ouvre la possibilité à une rétroaction informationnelle. En 1983, Herbert A Simon définit l'apprentissage selon les termes suivants : "*L'apprentissage désigne des changements adaptatifs d'un système, dans le sens où ils permettent au système d'effectuer une même tâche [...] plus efficacement que la fois précédente*" [107]. Cette définition offre une approche conceptuelle plus simplifiée et abordable que celle décrite par les termes "intelligence artificielle", le concept d'intelligence étant lui-même un signifiant abstrait et polysémique [89]. Car le concept d'intelligence, désignant étymologiquement la capacité à créer du lien entre plusieurs éléments (*inter-ligare*), prend différentes significations selon son champ d'application, qu'il s'agisse de psychologie, d'éthologie ou encore de traitement du signal.

Conséquence directe : la locution "intelligence artificielle" recoupe de nombreuses activités scientifiques et souffre d'un manque de spécificité. C'est pourquoi notre utilisation de ce terme sera restreinte à au termes "apprentissage automatique" tel que définit précédemment. Ainsi, un système qui apprend est un système qui s'adapte par rétroaction informationnelle, en vue de produire un résultat plus performant, plus adapté à sa tâche.

Cette dernière définition induit le besoin de mesurer la qualité d'un apprentissage. Cela se fait via le taux d'erreur du système, que l'on va chercher à minimiser. Ainsi, nous pouvons inscrire l'apprentissage automatique comme un champ d'étude à l'intersection de la statistique descriptive et des problèmes d'optimisation.

À ce titre, le champ applicatif du *machine learning* est large, une même série d'outils pouvant être appliquée sur des problèmes très divers. C'est d'ailleurs pour cela que les techniques d'apprentissage automatique ont envahi tous les champs scientifiques, pour peu que ces derniers aient un problème d'optimisation à résoudre. Plus récemment, le développement de l'apprentissage profond, ou *deep learning*, basé sur l'utilisation de réseaux de neurones artificiels, a intensifié l'utilisation de telles techniques. Ce développement a été rendu possible par l'amélioration de la puissance de calcul, les réseaux de neurones artificiels étant particulièrement nécessaires en temps et en puissance calculatoire. Par opposition sémantique, tout le *machine learning* qui n'emploie pas ces grands réseaux neuronaux artificiels peut être décrit comme étant un apprentissage automatique classique.

Le qualificatif *Deep learning* est un terme utilisé de paire avec *big data*, un terme populaire pour désigner une grande quantité de données. L'exploitation des données de santé, par la multiplication des appareils de mesure (compteur de pas, mesure du rythme cardiaque, objets connectés...), constitue un champ applicatif privilégié de l'apprentissage automatique. La fouille des données de santé en vue de produire diverses prédictions sur l'apparition de certaines maladies soulève des enjeux économiques, des questions de santé publique, mais aussi des questions d'ordre bioéthique et politique. C'est donc un terrain délicat.

Dans un premier temps, nous allons présenter les étapes générales qui composent un système d'apprentissage automatique. Nous reviendrons bloc par bloc sur chacun des

éléments. Nous présenterons dans un second temps les outils statistiques fondamentaux permettant de comprendre les différentes techniques d'apprentissage automatique, puis décrirons plus précisément les différentes catégories et familles d'algorithmes. Nous établirons une différence entre les algorithmes d'apprentissage automatique superficiel et l'apprentissage automatique profond. Enfin, cet apprentissage profond sera traité en fin de chapitre.

4.2 Anatomie d'un schéma d'apprentissage automatique

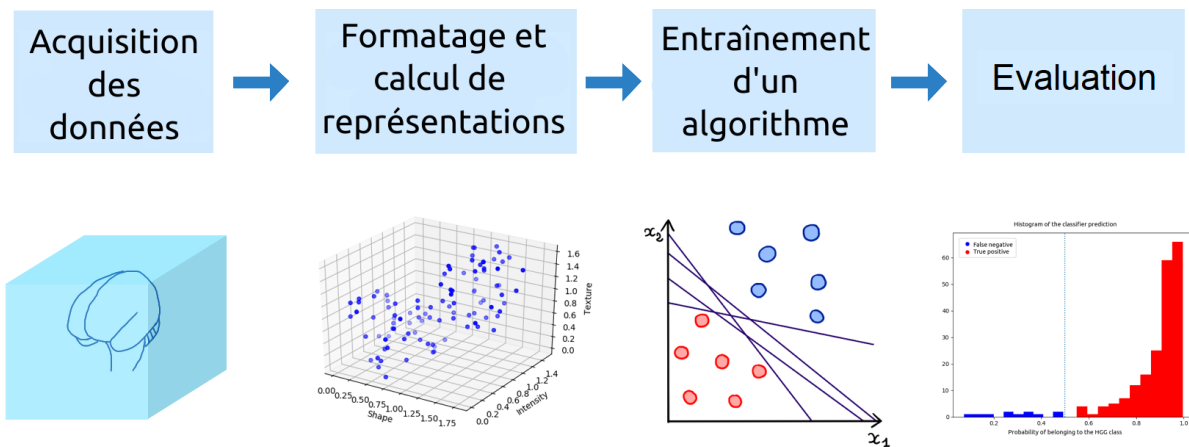


FIGURE 4.1 : Un schéma blocs d'apprentissage automatique.

Un système d'apprentissage automatique peut être résumé par les quatre blocs illustrés en Figure 4.1 : acquisition, prétraitement et calcul de représentations, application de l'algorithme et évaluation des résultats. Chaque bloc contient un part de choix de la part de l'utilisateur. Passons en revue les différents blocs :

- **L'acquisition** : Cette étape comprend le recueil de données via un appareil de mesure. S'appliquent alors les principes de la métrologie et les contraintes des

chaînes d'acquisitions. Il s'agit de l'étape qui rattache les objets numériques sur lesquels nous allons travailler au monde sensible, à la réalité biologique. Suite à l'acquisition, le signal est devenu un objet numérique facilement manipulable ;

- **Le prétraitement et le calcul de représentations** : Le signal va devoir être travaillé en amont de l'algorithme. Le prétraitement peut comprendre des étapes de filtrages ou de normalisation. Puis, des descriptions supplémentaires vont être calculées afin d'enrichir la base de données et faciliter le travail de l'algorithme pour atteindre le résultat souhaité. Ces descriptions peuvent être des variables continues (des mesures informatiques) ou discrètes (des labels, des catégories) ;
- **L'entraînement d'un algorithme** : Plusieurs familles d'algorithmes existent et chacune fournit un résultat qui lui est propre. Ainsi, le choix de l'algorithme n'est pas un choix anodin dans la production du résultat. Différentes méthodes comme la validation croisée (*cross-validation*) ou la recherche quadrillée (*grid search*) permettent d'optimiser les paramètres et hyperparamètres de l'algorithme pour un ajustement plus fin, ou plus robustes par rapport aux bruits inhérents aux données réelles ;
- **L'évaluation du système** : Une fois que l'algorithme a appris, il faut pouvoir l'évaluer. Différentes métriques peuvent être choisies pour évaluer la performance ;
- **La prédiction** : Une fois le système paramétré, on utilise ce dernier sur des données inconnues.

4.3 Apprentissage automatique classique

L'apprentissage automatique cherche à identifier des motifs dans un jeu de données. L'identification du motif, qu'elle soit guidée ou spontanée, relève d'un mécanisme d'inférence sémantique : l'algorithme apporte un sens, organise, une représentation numérique. Un exemple est donné Figure [4.2](#). Dans cet exemple, un cerveau présente une tumeur.

La compréhension de cet objet se fait via différentes représentations, chacune à un niveau de lecture différent. L'ordinateur interprète l'image comme une matrice de valeurs. Le radiologue interprète la même image comme une tumeur dont il peut identifier les contours. Ce faisant, il a projeté sur l'image une information supplémentaire : c'est une inférence sémantique. Tout l'enjeu de l'apprentissage automatique consiste à reproduire le mécanisme par lequel le radiologue passe de l'image à son interprétation. Il s'agit d'une modélisation qui automatise la production d'une information nouvelle, exploitable et cohérente pour l'utilisateur humain.

Pour cela, nous devons disposer d'outils permettant la description de catégories. C'est pourquoi, l'apprentissage automatique utilise les outils de la statistique descriptive, que nous allons maintenant présenter.

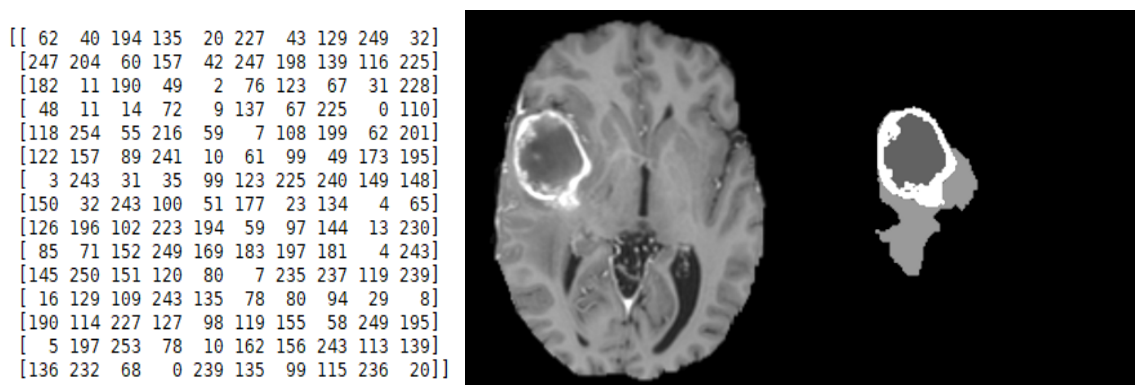


FIGURE 4.2 : *Trois représentations et interprétations d'un même objet : un cerveau présentant une tumeur cérébrale. (a) Informatiquement, il s'agit d'un ensemble de valeurs. (b) Ce dernier représente un cerveau pour un observateur humain, (c) le radiologue expert est capable d'identifier les contours d'une tumeur. L'algorithme d'intelligence artificielle doit automatiser l'interprétation sémantique du radiologue à la représentation numérique.*

4.3.1 Statistiques, mesures et représentations

Statistique descriptive : L'apprentissage automatique repose sur un socle de connaissances mathématiques issues de la statistique descriptive. En effet, tout problème d'ap-

prentissage automatique nécessite un jeu de données que l'on va chercher à caractériser, structurer, reproduire... Pour cela, nous utilisons les notions courantes permettant de caractériser une distribution, comme la moyenne, la variance, l'écart-type afin d'identifier la loi de distribution qui régit nos différents échantillons.

Le théorème central limite nous indique que la somme de variables aléatoires converge vers une distribution normale. Ainsi, à défaut de connaître la distribution d'un motif apparent, on peut émettre une hypothèse naïve lors de l'approche d'un problème, en supposant que ce dernier est régit par une loi normale (Figure 4.3) telle que :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \times e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (4.1)$$

Avec : $f(x)$ la fonction gaussienne

x une variable aléatoire

σ l'écart-type de la distribution

μ la moyenne de la distribution

Mesures et données : Les algorithmes d'apprentissage automatique s'appliquent sur des représentations numériques. Ces représentations s'obtiennent suite à des mesures, qui comportent leur part de bruit. Il est donc important de ne pas traiter l'information comme un signal informatique théorique épuré, mais comme le résultat d'un processus d'acquisition. Il faut pour cela connaître et être vigilant aux différents bruits et artefacts présents dans les données. Concernant l'IRM, ces différents types de bruits et artefacts ont été présentés dans le chapitre correspondant. Ainsi, la connaissance physique du signal et de sa chaîne d'acquisition est nécessaire avant d'aborder un problème d'apprentissage automatique.

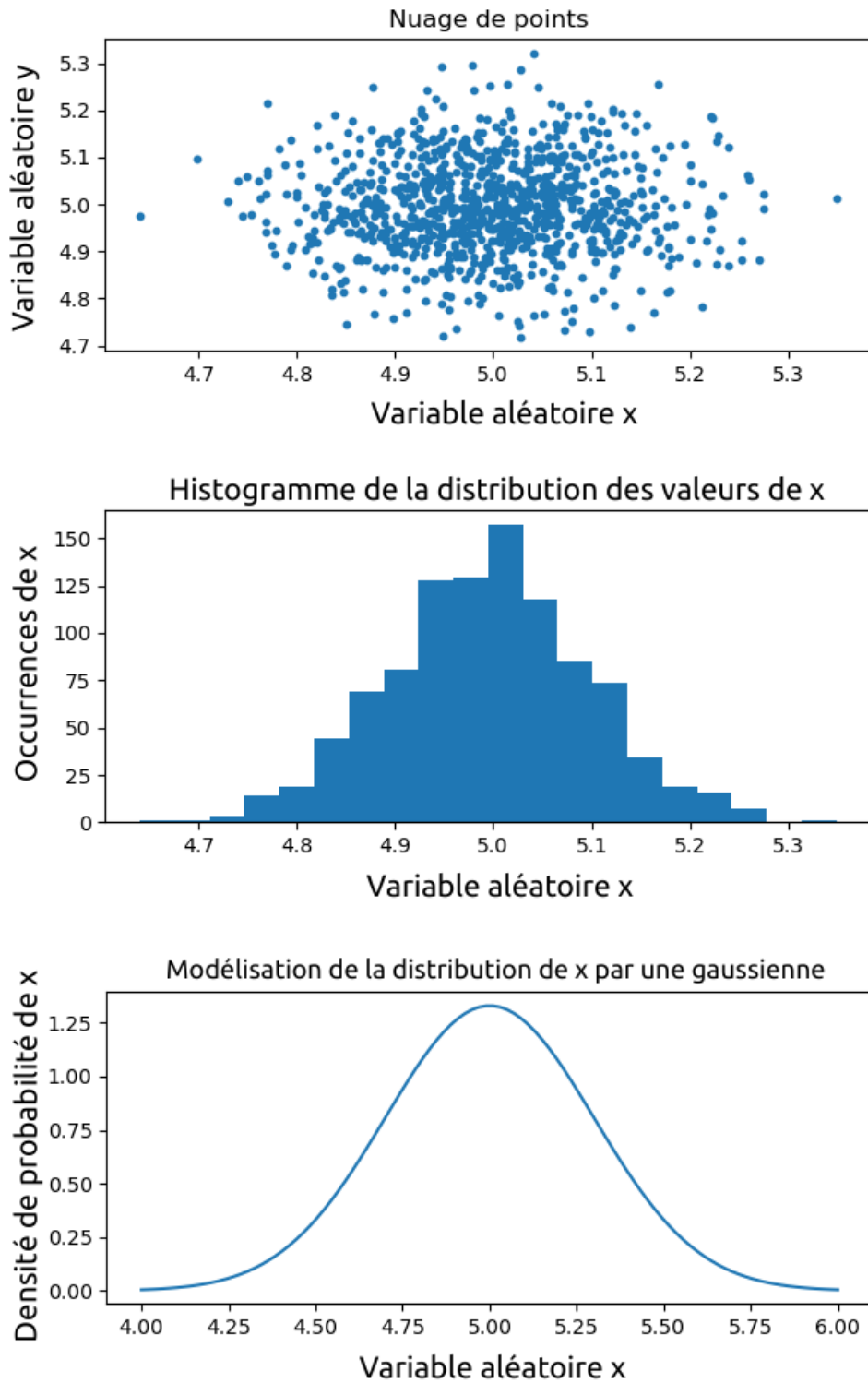


FIGURE 4.3 : (haut) Un nuage de points créé par deux variables aléatoires x et y ; (milieu) l'histogramme de la distribution des valeurs de x ; (bas) En supposant une loi normale, il est possible de caractériser la distribution de x par une gaussienne de moyenne 5 et d'écart-type 0,3.

L'acquisition transforme un signal physique en un signal informatique. Ainsi, l'organe "cerveau" devient un objet informatique dont la représentation est une matrice à trois dimensions spatiales, chaque voxel représentant numériquement une intensité de signal. Chaque patient constitue un individu statistique et chaque modalité IRM mesurée fournit une description brute. Mais il est rare en apprentissage automatique superficiel d'utiliser directement la donnée brute en entrée. On utilise à la place un ensemble de descripteurs numériques, calculés depuis la donnée brute. Ce sont ces descripteurs, ou *features*, qui vont offrir une nouvelle représentation de l'objet numérique.

Un descripteur est une mesure. Sa valeur peut être obtenue physiquement par un instrument de mesure, ou calculée à partir d'une représentation numérique. Hormis lorsqu'ils sont estimés automatiquement comme cela peut être le cas en apprentissage profond, les descripteurs sont choisis par l'utilisateur. Nous appellerons "matrice de description" la structure qui recueille en ligne les individus et en colonne chacun des descripteurs choisis. Chaque description supplémentaire correspond à une colonne supplémentaire dans la matrice de description, et à chaque description est associé un type de variable. Ainsi, il est possible de représenter dans un espace nos données sous forme de nuage de points.

Au-delà de trois dimensions, la représentation graphique du nuage de points n'est plus intuitive. Mais il est commun en apprentissage automatique de travailler avec un grand nombre de dimensions, car chaque descripteur agit comme un axe de représentation supplémentaire.

Avoir un trop grand nombre de dimensions provoque la malédiction de la dimension. Chaque axe rajoute de l'information, mais augmente aussi l'espace et la distance entre les points. Si bien qu'à partir d'un trop grand nombre de descripteurs, l'espace devient vide de points, et les groupes à identifier s'éparpillent, disparaissent. Il n'est alors plus possible d'identifier les motifs que l'on souhaite reconnaître et les algorithmes ne peuvent plus apprendre correctement. Ils perdent donc en performance. Il y a donc un marchandage (*tradeoff*) entre l'envie de l'utilisateur d'augmenter le nombre de descripteurs et la limitation technique, conséquence de la malédiction de la dimension. Pour résoudre ce problème, plus de descriptions doit s'accompagner de plus d'individus, ce qui n'est pas

toujours possible dans un contexte médical.

C'est pourquoi, pour comprendre les nuages de points sur lesquels on travaille, on utilise des techniques issues de la visualisation de données (*data visualisation*). Ce champ scientifique s'intéresse aux moyens de produire les représentations les plus riches en informations possibles, et les plus intelligibles possibles pour les utilisateurs. Un exemple pratique peut être d'effectuer une réduction de dimensions avec un algorithme, en combinant les informations sur un même axe, par exemple avec une analyse en composantes principales (ACP). Certains algorithmes permettent également de hiérarchiser les descripteurs lors de l'apprentissage pour en retirer les moins importants, c'est le cas notamment de l'algorithme d'élimination récursive de descripteurs (RFE pour *Recursive Feature Elimination*).

Données creuses : Une base de données est constituée de plusieurs individus statistiques regroupant chacun différentes acquisitions. Il peut arriver que, suite à des contraintes diverses, certaines acquisitions soient manquantes pour certains individus. Pour désigner ces valeurs manquantes, on parle de données creuses. A l'inverse, une base de données complète sera désignée comme dense.

Manipuler des données creuses soulève différents problèmes, qu'il s'agisse de représenter l'individu, ou d'effectuer des calculs sur l'ensemble de la base de données. Comment, par exemple, calculer la moyenne d'un descripteur dont certaines valeurs de la série sont manquantes ? Un traitement particulier doit être choisi et appliqué aux individus présentant des données creuses. Le but est de produire une représentation dense.

Selon l'ampleur du problème, plusieurs solutions peuvent être envisagées. La plus simple consiste à filtrer tous les individus comportant des données creuses. Cette méthode a un désavantage immédiat : elle diminue le nombre d'individus présents dans la base de données, ce qui peut causer des problèmes de représentativité du jeu de données final.

Pour conserver tous les individus, il faut assigner des valeurs aux données creuses. Plusieurs propositions existent. Il est répandu de d'affecter une valeur statistique aux données creuses, qu'il s'agisse de la valeur moyenne du descripteur, la valeur médiane, ou la valeur la plus récurrente de la série. Il est également possible de simplement affecter

une valeur nulle ou du bruit blanc.

Augmentation de données : Un jeu de données peut être de trop petite taille pour représenter fidèlement toute la complexité d'un problème. La meilleure solution pour améliorer la représentativité d'une population est d'augmenter le nombre d'individus, mais cela n'est pas toujours possible. Il faut alors recourir à des techniques d'augmentation de données, qui vont créer artificiellement de nouveaux individus à partir de ceux existants.

Ces techniques sont utilisées, par exemple, dans le champ de la vision par ordinateur. Dans ce champ, une étude de 2019 par C.Shorten [106] regroupe d'une part les techniques à base de manipulation d'image et d'autre part les techniques à base d'apprentissage profond. Dans les manipulations d'image, on trouve l'application de filtres, de transformations géométriques comme des symétries ou des rotations, des transformations de l'espace de couleur, ou encore des suppressions aléatoires de morceaux de l'image. Dans les méthodes d'apprentissage profond, on demande directement à un algorithme de générer une image à partir d'exemples. En particulier, nous expliquerons plus en aval la structure des réseaux antagonistes génératifs (GAN).

Avec l'application de ces techniques d'augmentation de données, il est possible de multiplier d'un facteur important le nombre d'individus que l'on peut présenter en entrée d'un système d'apprentissage automatique.

Le compromis biais-variance : L'erreur commise lors d'une prédiction peut-être mesurée. La métrologie nous dit qu'une erreur se compose en trois parties :

- un biais, qui nous renseigne sur la justesse du système. Il s'agit de la différence moyenne entre la valeur obtenue et la valeur réelle souhaitée ;
- une variance, qui nous renseigne sur la précision du système. La variance reflète la dispersion des valeurs obtenues autour de la moyenne ;
- l'erreur irréductible de Bayes : il s'agit de l'erreur minimale à laquelle on peut s'attendre compte tenu de la variabilité d'un jeu de données.

Biais et variance sont illustrés sur la Figure 4.4. En pratique, un compromis existe entre le biais et la variance. En effet, on souhaite que les résultats d'un apprentissage puissent se généraliser à l'ensemble des cas qui ne sont pas dans la base d'apprentissage. Il faut donc mitiger le degré de spécialisation d'un algorithme à sa base d'apprentissage.

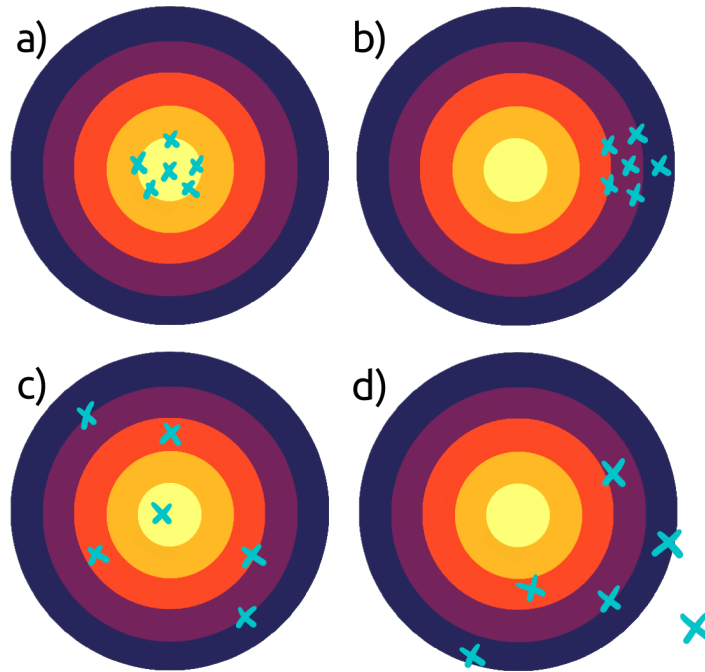


FIGURE 4.4 : *Illustration des notions de précision et justesse. Un système de prédiction peut être : (a) précis et juste, (b) précis mais biaisé, (c) juste mais peu précis, (d) peu précis et biaisé. Image inspirée d'un article de popularisation des sciences co-écrit avec Angélique Perrillat-Mercerot [34].*

Régression et classification : Deux grands types de problèmes existent en apprentissage automatique : les problèmes de régression et les problèmes de classification. Le type de problème est déterminé par le résultat recherché.

La régression cherche à déterminer la valeur d'une variable de sortie quantitative (dite variable expliquée) en fonction des variables d'entrée (dites variables explicatives). L'obtention de la variable expliquée se fait via une fonction de lien. Cette fonction peut avoir une forme prédéfinie grâce à la connaissance *a priori* du problème. On parle alors de régression paramétrique. Autrement, on parle de régression non paramétrique.

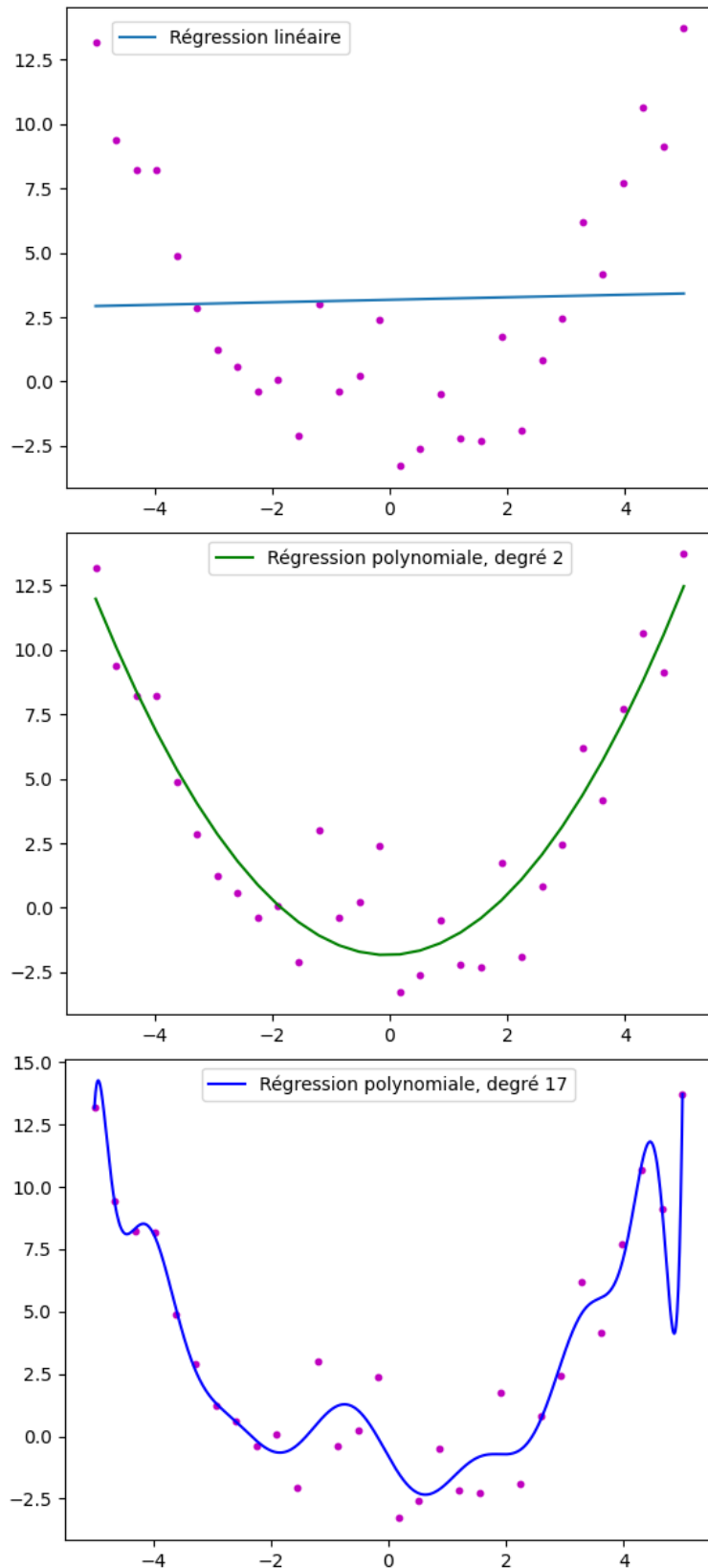


FIGURE 4.5 : Résultat d'une régression linéaire et de deux régressions polynomiales (degré 2 et 17) sur un jeu de données. Une connaissance a priori permet de choisir le meilleur modèle de régression.

Par exemple, un ensemble de points peut être modélisé par une régression linéaire ou polynomiale comme montré sur la Figure 4.5. Les trois algorithmes de l'exemple fournissent chacun leur meilleur résultat, mais c'est bien une connaissance *a priori*, compte tenu de la forme du nuage de points, qui nous fera préférer une régression polynomiale de degré 2 dans cet exemple.

La classification cherche à associer une étiquette ou un label à un individu inconnu via un modèle appelé classifieur. Là où la régression cherchait à représenter la forme du nuage de points par une fonction, un classifieur va trouver des frontières de séparation dans l'espace de données. Ces frontières peuvent optimiser différents critères :

- Maximiser la ressemblance intra-classe : les individus doivent être rattachés à d'autres points qui leur sont semblables ;
- Maximiser la distance inter-classes : des classes différentes doivent être les plus distantes les unes des autres.

4.3.2 Types d'apprentissages

Les systèmes d'apprentissage automatique peuvent s'adapter à de nombreux problèmes scientifiques. Selon Sebastian Raschka dans son livre *Python Machine Learning* [97], trois grandes stratégies d'apprentissage existent en fonction de l'application : l'apprentissage supervisé, l'apprentissage non supervisé et l'apprentissage par renforcement.

L'apprentissage supervisé : Il s'agit de la grande majorité des applications. Lors d'un apprentissage supervisé, on dispose d'une base de données contenant les différents échantillons d'une part et d'une série de mesure ou labels appelée vérité terrain d'autre part. Cette dernière constitue l'objectif que l'on va chercher à modéliser à partir de nos données. L'algorithme va apprendre par présentation successive d'exemples. Il est alors possible de calculer l'erreur de prédiction dès lors qu'un individu est présenté en entrée. Comme la vérité terrain est le plus souvent établie par un expert, le modèle résultant de l'algorithme va chercher à modéliser la réponse de l'expert.

L'apprentissage non supervisé : Par opposition à l'apprentissage supervisé, l'apprentissage non-supervisé désigne une famille de méthodes qui fonctionnent sans nécessiter de vérité terrain. Il s'agit notamment des algorithmes de clustering, où les individus vont être regroupés en sous-groupes selon leur similarité. Le clustering est une technique d'exploration des données qui renseigne l'utilisateur sur la structure de son nuage de points. Ainsi, ce type d'algorithme ne cherche pas à reproduire le résultat d'un expert.

L'apprentissage par renforcement : L'apprentissage par renforcement constitue une troisième catégorie de méthodes. Un système apprenant, dit agent, va tenter de maximiser une récompense en effectuant une action. Selon le résultat de l'action, un signal de récompense sera calculé et permettra au système de s'ajuster pour optimiser ses performances. Contrairement à l'apprentissage supervisé, le signal de récompense n'est pas calculé par rapport à une vérité terrain fixée : il est issu d'une mesure de la performance elle-même. Il est courant d'illustrer ces algorithmes avec des tâches de robotique, comme par exemple l'apprentissage de la marche pour robot bipède, où un robot doit apprendre à parcourir la plus longue distance sans tomber [78]. La distance parcourue est la mesure à améliorer, et le robot découvre par une série d'essais et d'erreurs comment ajuster son équilibre au fil de ses pas.

4.3.3 Les problèmes d'optimisation

Pour faire converger les résultats de l'algorithme sur les valeurs attendues, il faut définir une métrique du taux d'erreur appelée fonction de coût. L'un des enjeux de l'apprentissage revient à chercher le minimum global de la fonction de coût via l'optimisation des paramètres de l'algorithme. Ces paramètres sont des poids. Pour ce faire, un ensemble de méthodes numériques peuvent être mises en place afin de chercher le minimum de la fonction de coût par un ajustement successif des poids de l'algorithme. Mais cette fonction, dont les variations ne sont pas connues, peut présenter des minima locaux, qui peuvent piéger l'algorithme dans un état éloigné du minimum global recherché. C'est pourquoi les méthodes d'optimisation incorporent une part d'aléatoire (stochastique)

dans leur résolution. En incorporant des éléments aléatoires, on permet à l'algorithme de tester différentes valeurs de poids initiaux, qui convergeront vers des solutions différentes, comme représenté Figure 4.6. On augmente ainsi la probabilité d'atteindre le minimum global tout en évitant les minima locaux.

Minimisation d'erreur et fonction de coût : Un exemple de fonction de coût provient du calcul des moindres carrés. Il s'agit d'une estimation de l'erreur moyenne au carré entre chaque point et un modèle de régression. L'intérêt de cette méthode vient de sa symétrie : elle traite de la même manière une erreur positive (sur-évaluation) ou négative (sous-estimation). Le calcul effectué est le suivant :

$$J(\vec{w}) = \frac{1}{2} \sum_{i=0}^N \left(y_i - \tilde{y}_i(\vec{w}) \right)^2 \quad (4.2)$$

Avec : $J(\vec{w})$ la fonction de coût pour le vecteur de poids \vec{w}

N le nombre d'échantillons estimés

y_i la réponse attendue

$\tilde{y}_i(\vec{w})$ la réponse obtenue avec le vecteur de poids \vec{w}

Une fonction de coût décrit la performance d'un algorithme appliqué à un jeu de données. Pour obtenir le meilleur apprentissage possible, il faut minimiser le coût, en atteignant la valeur la plus basse possible de la fonction. Pour un modèle optimal, le minimum de la fonction correspond à l'erreur irréductible de Bayes. Nous pouvons tendre vers cet optimum par l'application de méthodes numériques comme la descente de gradient.

La descente de gradient : Le gradient d'une fonction est le vecteur des dérivées partielles en un point de la fonction. Il est orienté vers le maximum le plus proche. Ainsi, en suivant le gradient dans sa direction opposée, on peut converger vers une valeur plus basse de la fonction. Cette descente du gradient se fait par une adaptation des poids du modèle telle que :

$$\vec{w} = \vec{w} + \overrightarrow{\Delta w} \quad (4.3)$$

et

$$\overrightarrow{\Delta w} = -\eta \times \overrightarrow{\nabla J(w)} \quad (4.4)$$

Avec : \vec{w} un poids à ajuster

$\overrightarrow{\Delta w}$: la correction du poids à appliquer

η : le pas de progression avec $\eta > 0$

$\overrightarrow{\nabla J(w)}$: le gradient de la fonction de coût

Dans ces équations, le $\overrightarrow{\Delta w}$ indique dans quelle direction effectuer la descente de gradient. Un paramètre η , le pas de progression, détermine l'amplitude d'ajustement du poids à chaque itération. Plus ce paramètre est élevé, plus le système peut converger rapidement, mais risque d'être instable. En revanche, un pas de progression plus petit évoluera de manière très stable sur la fonction de coût, mais il ne permettra que de petits déplacements. Il faudra alors un grand nombre d'itérations pour atteindre le minima local.

Appliquée à une fonction de coût, cette méthode permet d'atteindre un minima local ou un plateau. La Figure 4.6 montre trois initialisations aléatoires sur une même fonction de coût, ici en 2D. L'algorithme de descente de gradient classique est déterministe. Sans l'initialisation aléatoire, il convergerait toujours au même résultat. C'est pourquoi, afin d'éviter les minima locaux, une part d'aléatoire est introduite. Cet élément stochastique permet d'évaluer la fonction de coût en des points différents, qui, selon les entraînements, amèneront à un minima satisfaisant.

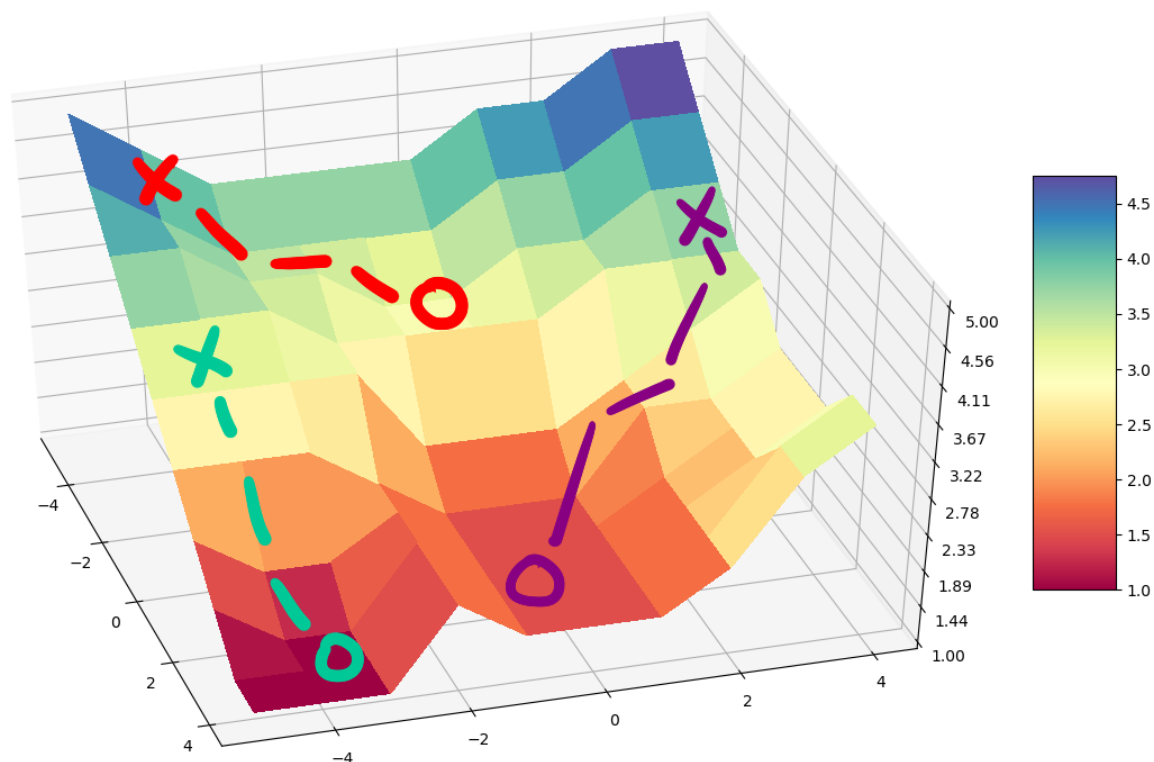


FIGURE 4.6 : Trois descentes de gradients sur une fonction de coût en 2D. Les croix montrent les points de départ initiaux et les cercles l'arrêt de la descente de gradient (minima locaux à pente nulle). Selon son initialisation, la descente de gradient peut atteindre des minima locaux (exemple des tracés rouge et violet). C'est pourquoi, les algorithmes d'apprentissage automatique reposent sur une part d'aléatoire (stochastique).

4.3.4 Divisions de la base de données

Une fois que l'algorithme a appris, il faut pouvoir évaluer la qualité de son apprentissage. Ce dernier doit non seulement donner de bons résultats -selon la métrique choisie- sur la base de données dont on dispose, mais il doit aussi pouvoir s'extrapoler à de nouvelles données inconnues. En effet, un des problèmes pouvant émerger à ce stade est le problème de sur-apprentissage, où le modèle entraîné est devenu trop spécifique aux données que nous lui avons présentées. Pour contrer cela, il est commun de diviser la base de données en trois parties :

La base d'apprentissage : C'est celle sur laquelle nous allons entraîner notre modèle. Elle doit être suffisamment fournie pour permettre une bonne représentation du problème à modéliser.

La base de validation : Elle sert à tester les performances du modèle issu de la base d'apprentissage. Cette première base permet de tester la première extrapolation du modèle à des données inconnues. Si l'évaluation du modèle n'est pas satisfaisante, il faut recommencer l'apprentissage sur la base d'apprentissage, peut-être avec des paramètres différents, puis évaluer le nouveau modèle sur la base de validation. Ainsi, la base de validation doit elle aussi contenir suffisamment d'individus pour être représentative du problème à étudier.

La base de test : Il s'agit d'une seconde base de données qui ne sert à évaluer que le meilleur algorithme issu de la base de validation. Les échantillons de cette base sont considérés comme une part des individus inconnus auxquels notre modèle serait confronté une fois hors du schéma d'apprentissage. Il est donc important que ces données soient suffisamment nombreuses pour pouvoir représenter la variabilité et la complexité du monde réel.

Les rapports de proportion permettant de diviser une base de données entre base d'apprentissage, base de validation ou base de test dépendent des problèmes à traiter.

4.3.5 Complexité du modèle : le sous-apprentissage et le sur-apprentissage

Cette division en trois bases permet d'éviter le phénomène de sur-apprentissage. Selon le niveau d'apprentissage, qui dépend de la complexité du modèle, comme par exemple du degré d'un polynôme lors d'une régression, le modèle résultant va être plus ou moins spécifique à la base d'apprentissage. Un modèle qui généralise trop fortement la base de données disponible sous-apprend, tandis qu'un modèle qui devient trop spécifique des données d'apprentissage sur-apprend. Deux exemples sont montrés en Figure [4.7](#). Dans

les deux cas, l'extrapolation du modèle à des données inconnues, comme la base de test, produit une grande erreur.

On peut représenter l'objectif de l'apprentissage par la Figure 4.8. Dans cette figure, on représente le taux d'erreur d'un modèle sur la base de test et la base d'apprentissage en fonction de sa complexité. L'enjeu est de trouver une méthode permettant d'obtenir le plus bas taux d'erreur sur les deux bases de données.

4.3.6 La validation croisée

Lorsque le nombre d'échantillons est faible, on ne peut pas diviser la base de données en base d'apprentissage, base de validation et base de test. La validation croisée est une technique qui permet d'apprendre sur tous les échantillons (hors base de test) par un système de roulement. La base d'apprentissage est mélangée puis divisée en N portions égales, et un modèle va être entraîné sur $N-1$ portions. La portion restante va servir de base de validation. Puis, par roulement, l'apprentissage va être réitéré sur chaque arrangement possible, comme montré en Figure 4.9.

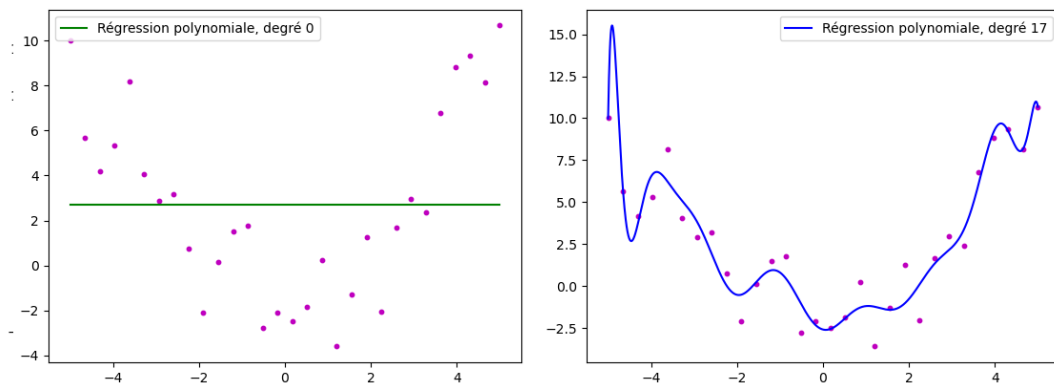


FIGURE 4.7 : Un exemple de sous-apprentissage (à gauche) et de sur-apprentissage (à droite). Le sous-apprentissage généralise trop le jeu de données (ici, en ne prenant que la valeur moyenne), tandis que le sur-apprentissage devient trop spécifique aux données, par un modèle trop complexe. Il en résulte une plus grande erreur lors de l'extrapolation du modèle à des données inconnues.

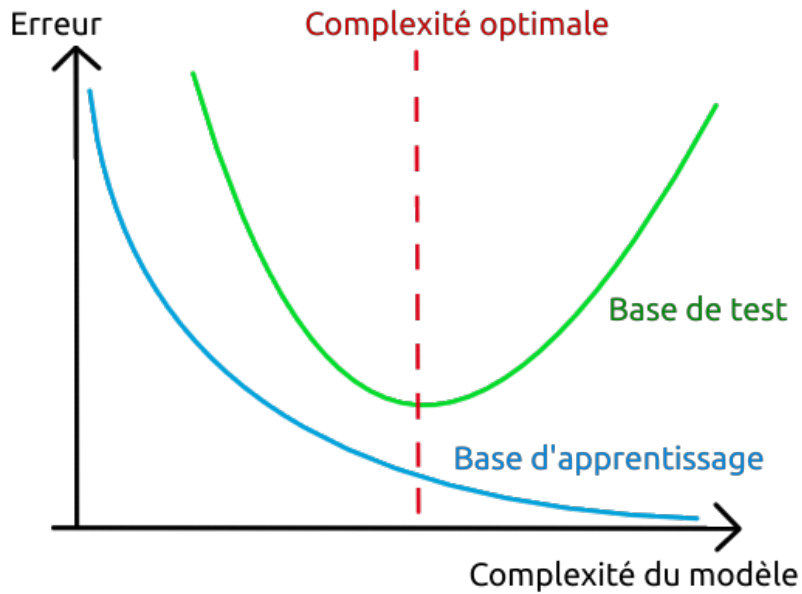


FIGURE 4.8 : Évolution du taux d'erreur sur la base d'apprentissage et la base de test en fonction de la complexité du modèle. Il existe une complexité optimale du modèle qui minimise le taux d'erreur et sur la base de test et sur la base d'apprentissage. Un modèle moins complexe sous-apprend, et un modèle plus complexe sur-apprend.

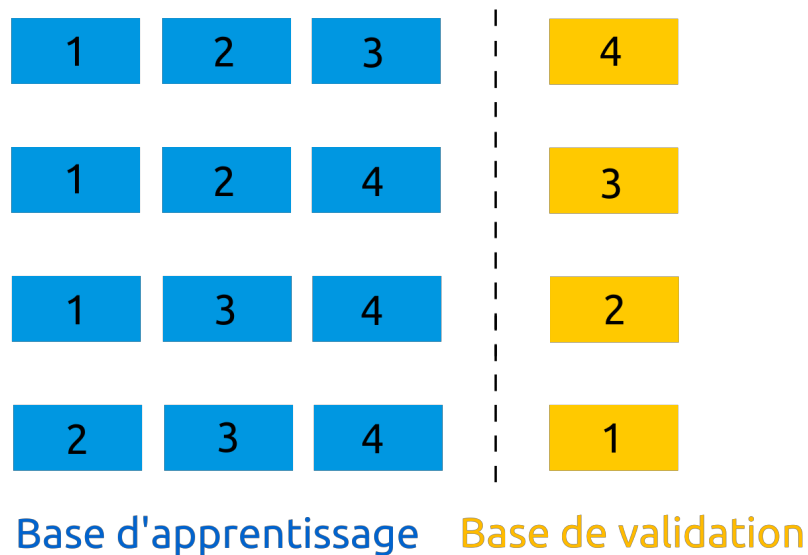


FIGURE 4.9 : Exemple de validation croisée à 4 divisions. A chaque arrangement correspondra un modèle dont on évaluera les performances.

La validation croisée génère une instance du modèle par apprentissage effectué. Chaque instance aura ses performances propres, certaines ayant par exemple une meilleure précision que d'autres. Ainsi, pour connaître les performances, on peut prendre la valeur moyenne des N essais.

Cette méthode s'appelle validation croisée en K partitions (*K-folds cross validation*), où K est le nombre de divisions de notre base de données. Un cas particulier, la validation croisée à un seul individu (*leave-one-out cross validation*), existe lorsque K est égal au nombre d'individus de la base d'apprentissage. Un seul individu est mis de côté pour la validation, ce qui a pour effet de présenter presque toute la base de données lors des apprentissages successifs. Cette technique permet d'avoir un apprentissage plus représentatif de la base de données, mais demande d'entraîner un algorithme autant de fois qu'il y a d'individus, ce qui peut être prohibitif au regard du temps de calcul.

Enfin, si la base de données initiale présente un déséquilibre de classe, cas où une classe contient beaucoup moins d'individus qu'une autre, il est possible d'utiliser une validation croisée stratifiée (*stratified cross-validation*). Cette méthode permet de s'assurer que chaque base de validation contient la même proportion d'individus de chaque classe à identifier que la base d'apprentissage. Ainsi, nous évitons, par exemple, les cas où toute la base de validation ne contient qu'une seule des classes à identifier.

4.3.7 Paramétrage fin : paramètres et hyperparamètres

Nous avons vu dans nos exemples précédents que certains paramètres jouent sur la complexité du modèle à entraîner. Ces paramètres sont donnés à l'initialisation de l'algorithme, et ce dernier va converger lors de son apprentissage. Cette convergence se fait par l'ajustement de ses paramètres internes (ordonnées à l'origine, pente, poids des différents membres d'un polynôme...). On appelle paramètres ces informations que l'algorithme va optimiser de lui-même lors de l'apprentissage, par opposition aux hyperparamètres, qui eux ne changent pas lors de l'apprentissage (comme le degré du polynôme que l'on choisit d'entraîner).

Nous avons montré, avec l'exemple de la Figure [4.5](#) ou [4.8](#), que ces hyperparamètres

influaient fortement sur la qualité des résultats obtenus. Mais tester chaque valeur d’hyperparamètre peut s’avérer laborieux. C’est pourquoi une méthode comme la recherche quadrillée (*GridSearch*) permet de tester différentes combinaisons d’hyperparamètres, suggérées par l’utilisateur. Combiné à l’application d’une validation croisée, ces deux outils permettent un paramétrage fin des performances du classifieur, en échange d’un temps de calcul important.

4.3.8 Matrice de confusion et métriques d’évaluation d’un classifieur

Pour évaluer les performances d’un classifieur, on utilisera également la matrice de confusion, qui, dans un cas de classification binaire, compte les individus selon quatre catégories représentées dans le Tableau 4.1. De cette matrice, il est possible de calculer plusieurs métriques de performance, dont les principales sont détaillées dans le Tableau 4.2.

Dans ce tableau, nous soulignons la différence entre des termes anglais et français qui peuvent être trompeurs. C’est le cas entre la métrique de précision (*accuracy*) et la valeur prédictive positive (*precision*). Pour éviter toute confusion, nous utiliserons la terminologie française dans la présentation de nos travaux.

En plus des métriques issues de la matrice de confusion, le couple sensibilité et spécificité permet d’établir la courbe ROC (pour *receiver operating characteristic*) dont l’aire sous la courbe est aussi utilisée comme mesure de performance d’un classifieur binaire.

	Test positif	Test négatif
Individu positif	Vrai Positif (TP)	Faux négatif (FN)
Individu négatif	Faux Positif (FP)	Vrai négatif (TN)

TABLE 4.1 : *Matrice de confusion dans un cas de classification binaire. Après un test, la matrice de confusion répartit les individus en quatre catégories selon leur résultat attendu et leur résultat réel.*

Précision (<i>Accuracy</i>)	$\frac{TP+TN}{TP+TN+FP+FN}$
Sensibilité, taux de vrais positifs (TPR)	$\frac{TP}{TP+FN}$
Spécificité, taux de vrais négatifs (TNR)	$\frac{TN}{TN+FP}$
Valeur prédictive positive (<i>Precision</i>)	$\frac{TP}{TP+FP}$
Précision équilibrée (<i>Balanced accuracy</i>)	$\frac{TPR+TNR}{2}$
Dice score, ou score F1	$\frac{2TP}{2TP+FP+FN}$

TABLE 4.2 : *Plusieurs métriques issues de la matrice de confusion.*

4.3.9 Exemples d’algorithmes

Nous allons à présent illustrer les notions que nous venons d’aborder par la présentation de plusieurs algorithmes de classification. Nous commencerons par présenter les méthodes de clustering, une famille d’algorithmes non-supervisés, à travers l’exemple des K-moyennes [17]. Puis nous détaillerons des algorithmes d’apprentissage supervisé : les K plus proches voisins (KPPV) [4], les forêts d’arbres aléatoires (RF) [51] et les machines à vecteur de support (SVM) [25]. Dans chacun de ces exemples, le problème est le même : partant d’une base de données, nous souhaitons assigner un label à un individu inconnu.

Les K-moyennes (clustering) : Le clustering est une technique d’apprentissage non supervisé qui permet d’explorer un ensemble de données en regroupant les individus proches en sous-groupes appelés clusters. L’algorithme des K-moyennes (*K-means*) est un exemple d’algorithme de clustering. Cet algorithme permet d’illustrer le rôle des hyperparamètres : le résultat des K-moyennes dépend directement du nombre de classes que l’on cherche à identifier.

La structure de cet algorithme est illustrée Figure 4.10, et est la suivante :

- Partant d’un jeu de données non labellisé, on initialise aléatoirement K centroïdes. Ces centroïdes agissent comme le centre de gravité de leur classe ;
- Ainsi, selon une métrique de distance, on rattache chaque point du jeu de données

à son centroïde le plus proche ;

- La position des centroïdes est alors mise à jour comme étant le barycentre du nuage de point formé par leur classe ;
- On réitère l'algorithme jusqu'à convergence du résultat.

L'algorithme des K-moyennes emploie deux hyperparamètres : K, le nombre de classes que l'on souhaite identifier, et la métrique de distance utilisée. Lorsqu'il est utilisé avec une distance euclidienne, l'algorithme des K-moyennes converge vers une solution.

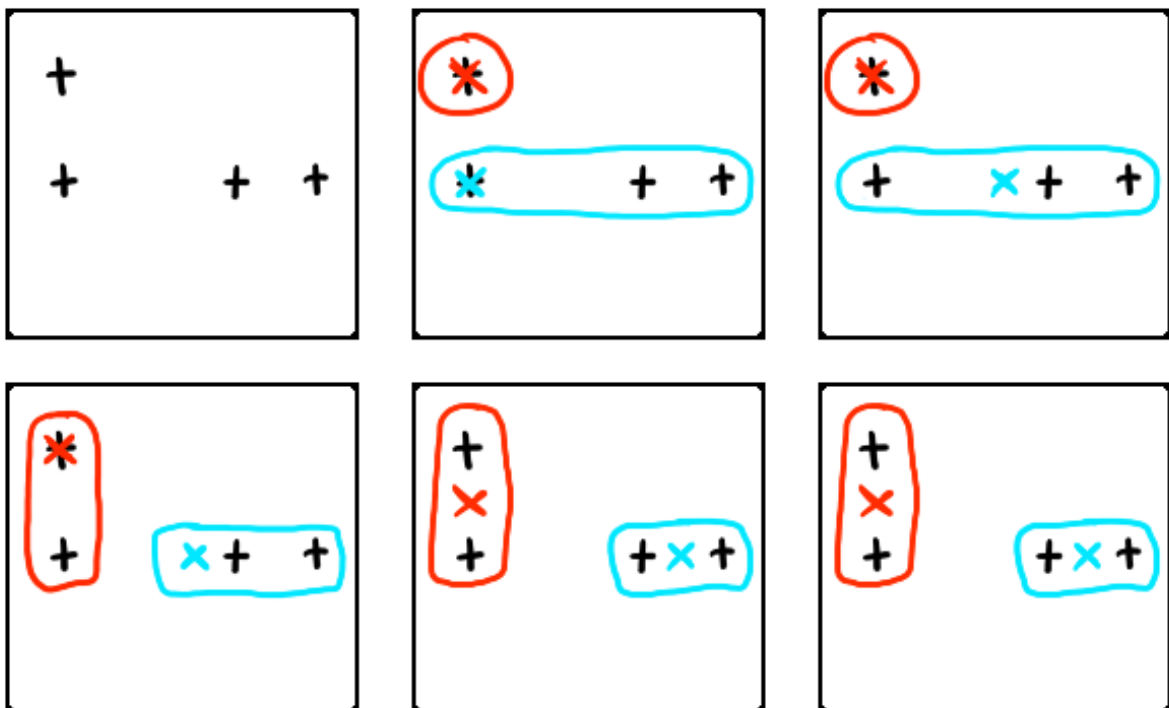


FIGURE 4.10 : Exemple d'application de l'algorithme des K-moyennes avec deux centroïdes. Lorsque la distance utilisée est une distance euclidienne, l'algorithme converge forcément vers une solution.

Les K plus proches voisins : L'algorithme des K plus proches voisins (KPPV) est un cas particulier de *lazy learner* : il ne produit pas une fonction permettant la discrimination entre plusieurs groupes, mais mémorise directement l'ensemble du jeu de données.

Cet algorithme, illustré Figure 4.11, est simple : on assigne à un point la classe majoritaire de ses K voisins les plus proches. Cet algorithme possède deux hyperparamètres : le nombre K de voisins à regarder, et la fonction utilisée pour calculer une distance entre deux points. Cette métrique de distance peut être une distance euclidienne.

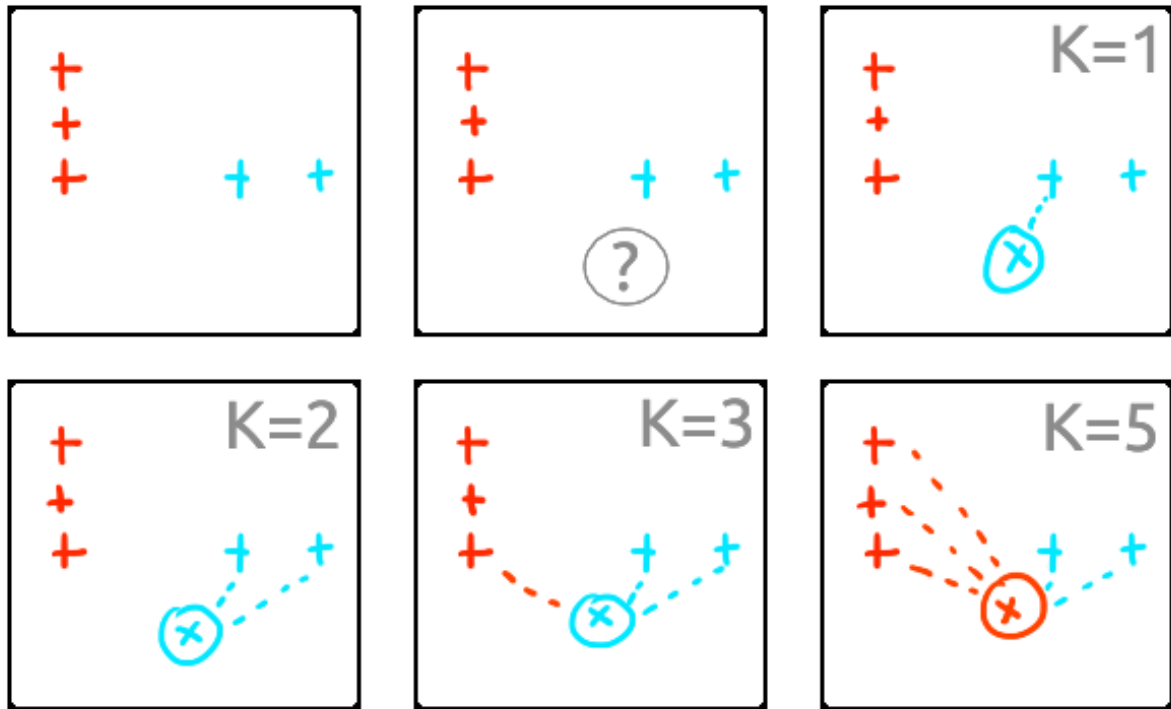


FIGURE 4.11 : Exemple d'application de l'algorithme des KPPV avec K croissant. Partant d'un jeu de données labellisé, on souhaite attribuer une classe au nouvel individu gris. En cas d'égalité, pour $K=4$, une loi choisit à quelle classe attribuer l'individu inconnu. On voit également que selon K , le résultat obtenu peut ne pas correspondre à ce qui semble intuitif pour l'œil humain : le point est rouge pour $K=5$, alors que l'œil humain attacherait plutôt le point inconnu à la classe bleue.

Les forêts d'arbres aléatoires (RF) : Les forêts d'arbres aléatoires (RF pour *random forests*) constituent une famille de méthodes particulièrement efficaces pour structurer et explorer un jeu de données. Elles reposent sur le concept d'arbre de décision. Un arbre de décision, dont un exemple est donné Figure 4.12, organise les individus via

une série de tests définis par des seuils. La construction d'un arbre de décision s'effectue en estimant à chaque nœud quel descripteur offre la meilleure discrimination de nos individus.

Dans une forêt aléatoire, seul un sous-ensemble de descripteurs et un sous-ensemble d'individus sont présents lors de la construction de l'arbre, ce qui induit de la variabilité. Il est alors possible de construire un grand nombre d'arbres dont on va pouvoir évaluer les performances. La performance globale s'évalue en utilisant tous les arbres créés. Cette méthode permet de multiplier les classifieurs et d'aggréger leurs résultats. On parle de *bagging* pour décrire ce regroupement de classifieurs.

Comme les arbres de décisions se construisent sur une hiérarchisation d'un sous-ensemble de descripteurs, il est possible de récupérer ces informations pour mieux comprendre la valeur informationnelle de nos descripteurs.

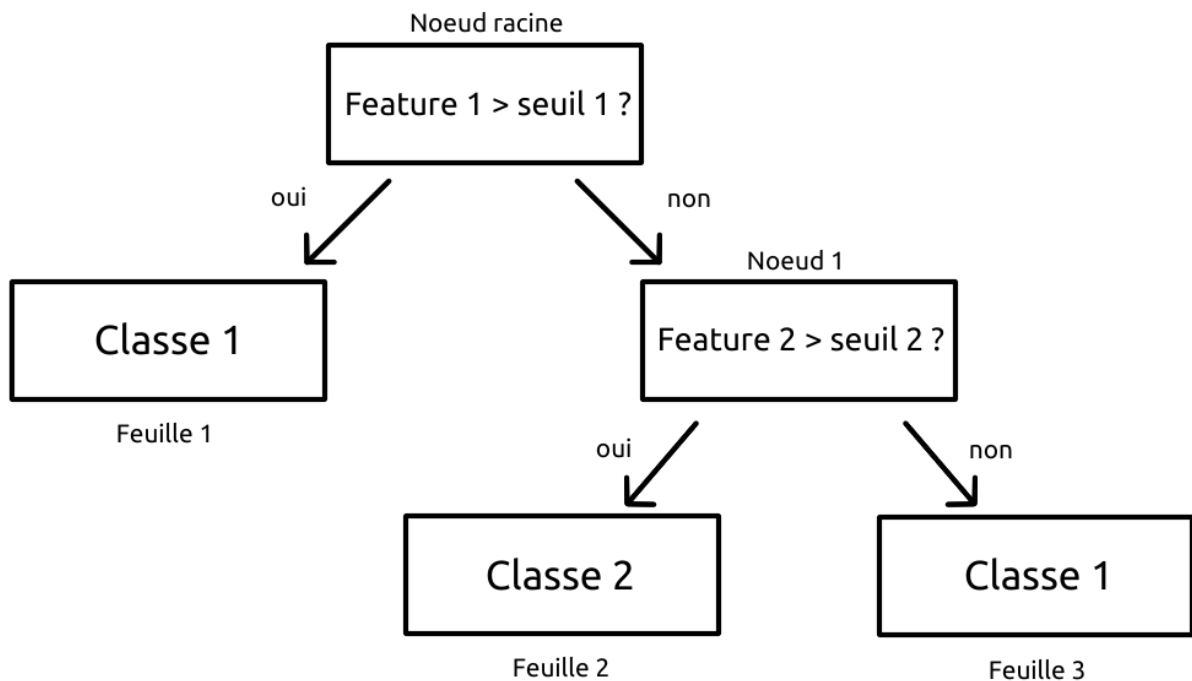


FIGURE 4.12 : Un arbre de décision. Les individus sont groupés selon une série de tests définis par des seuils. Plusieurs chemins peuvent amener à la même classe.

Les machines à vecteur de support (SVM) : Lors d'un problème de classification, un exemple de frontière de séparation entre classes peut être une droite, un plan ou un hyperplan. On parle alors de classifieur linéaire. Mais ce classifieur linéaire n'est peut-être pas le meilleur, comme représenté sur la Figure 4.13. Les SVM proposent une généralisation des classifieurs linéaires, en trouvant, grâce à deux principes, le meilleur plan ou hyperplan séparant les différentes classes. Ces deux principes sont la vaste marge et la marge souple.

Le principe de vaste marge force les SVM à maximiser la distance euclidienne entre l'hyperplan et le ou les plus proches individus de chaque classe de la base d'apprentissage. La marge souple, elle, permet de tolérer un certain taux d'erreur de classification pour maximiser les performances globales du classifieur. La combinaison de ces deux principes fait des SVM un algorithme robuste.

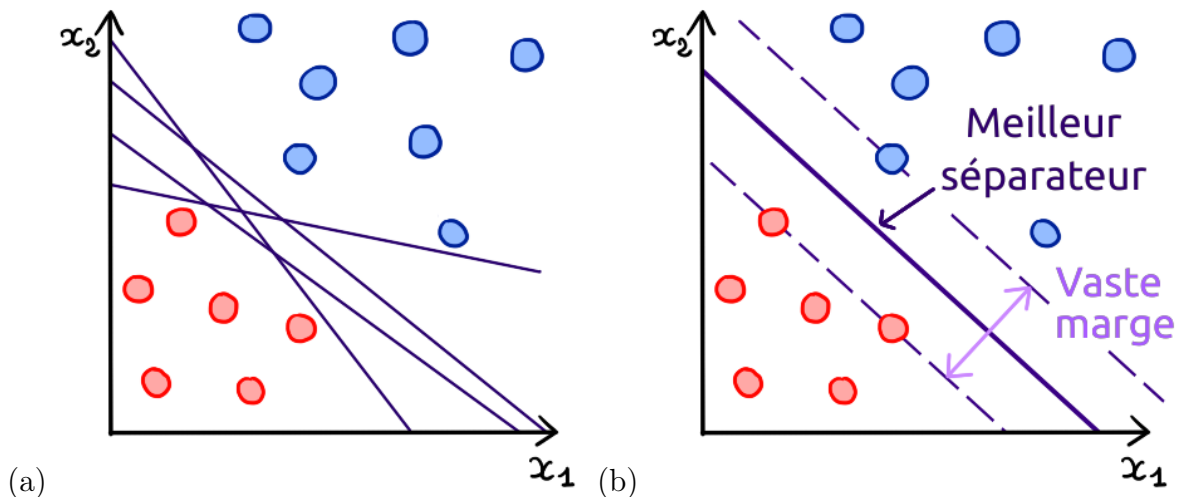


FIGURE 4.13 : (a) Plusieurs classifieurs linéaires peuvent séparer deux nuages de points. (b) En utilisant les principes de vaste marge et de marge souple, les SVM peuvent fournir le plan ou l'hyperplan qui maximise la distance entre les classes. Ici, le principe de vaste marge est représenté sur la figure.

Il est possible d'adapter les SVM à des problèmes non linéairement séparables. En effet, les SVM utilisent un noyau (*kernel*). Cet hyperparamètre permet à l'algorithme de calculer une transformation sur le jeu de données. Ainsi, des noyaux adaptés permettent de rendre linéairement séparables des problèmes qui ne l'étaient pas initialement. Un

exemple de ces transformations est donné Figure 4.14, où la deuxième composante du nuage de points, Y , est remplacée par une variable composée Z telle que $Z = X^2 + Y^2$. Une telle transformation permet de transformer notre exemple non linéairement séparable en un problème bien plus abordable par un classifieur linéaire.

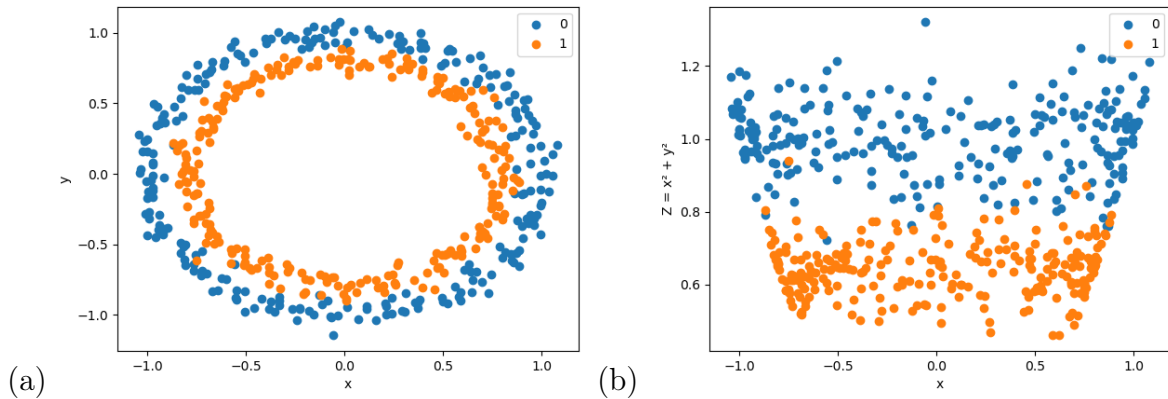


FIGURE 4.14 : Avec un changement de noyau, un problème initialement non linéairement séparable (a) peut le devenir (b). Le SVM peut utiliser différents noyaux pour revenir à un problème linéairement séparable.

Il existe plusieurs noyaux : linéaire, polynomiaux, à fonction radiale... Ici, l'utilisation de recherche quadrillée et de validation croisée permet de tester différents noyaux facilement, ce qui rend les SVM particulièrement adaptatifs aux différents problèmes où ils sont appliqués.

De plus, l'hyperplan tracé par le SVM peut donner, pour un noyau linéaire, une estimation de l'importance de l'information contenue dans chaque descripteur. Ceci permet d'élaborer des algorithmes de hiérarchisation et de sélection de descripteurs, comme le *SVM-Recursive Feature Elimination* (SVM-RFE).

Formalisation du SVM-RFE : Nous avons utilisé le SVM-RFE dans ce travail de thèse, en particulier section 7.7. C'est pourquoi, nous allons ici détailler le formalisme de cet algorithme, tel que décrit dans Guyon et al. [48]. Cette publication, définie dans le cas de classification binaire, est à l'origine de l'implémentation que nous avons utilisée.

Guyon et al. décrivent la tâche de classification de la manière suivante. Un vecteur

d'entrée à n composantes, correspondant à n descripteurs, est présenté en entrée du classifieur. Ce vecteur représente un individu x_i au sein d'un espace de représentation n -dimensionnel. Comme il s'agit d'un problème de classification binaire, cet individu peut appartenir à une classe positive (labellisée $+$) ou négative (labellisée $-$).

À la base d'apprentissage composée de l individus, $\{x_1, x_2, \dots, x_k, \dots, x_l\}$, est associé un vecteur de labels $\{y_1, y_2, \dots, y_k, \dots, y_l\}$, avec $y_k \in \{-1, +1\}$.

Une fonction de décision $D(x)$ divise la population telle que :

$$D(x) > 0 \Rightarrow x \in class(+)$$

$$D(x) < 0 \Rightarrow x \in class(-)$$

$$D(x) = 0, \text{ frontière de décision}$$

Pour un individu x_i , ce formalisme pose la fonction de décision comme suit :

$$D(x_i) = \vec{w} \cdot x_i + b \quad (4.5)$$

Avec : \vec{w} un vecteur de poids

b le biais

Il résulte de cette dernière équation que la frontière de séparation $D(x)$ est un plan divisant l'espace de représentation n -dimensionnel et orienté selon le vecteur de poids \vec{w} . La fonction de décision est donc directement orientée selon des poids attribués aux descripteurs formant l'espace de représentation.

Pour évaluer la contribution de chaque descripteur, Guyon et al. utilisent le coefficient de Golub [45], ω_i qui exprime la corrélation entre la valeur d'un poids w_i et les classes (+) ou (-). Ce coefficient est le suivant :

$$\omega_i = \frac{(\mu_i(+)) - \mu_i(-)}{(\sigma_i(+)) + \sigma_i(-)} \quad (4.6)$$

Avec : μ_i la moyenne des x_i appartenant à la classe (+) (ou la classe (-))

σ_i la variance des x_i appartenant à la classe (+) (ou la classe (-))

Avec ce coefficient, de grandes valeurs de ω_i indiquent une forte corrélation avec la classe positive et inversement. S'affranchissant du signe, les $(\omega_i)^2$ permettent une hiérarchisation des descripteurs. Selon Guyon et al., les coefficients ω_i peuvent être utilisés comme poids du classifieur. Réciproquement, le vecteur de poids \vec{w} multipliant les entrées du classifieur peut être utilisé comme coefficients de hiérarchisation. En effet, toujours selon ces auteurs, les entrées fortement pondérées influencent fortement la décision de classification. Ainsi, les descripteurs associés à des poids élevés sont des descripteurs de grande valeur informationnelle. On retrouve alors la relation exprimée par l'équation [4.5](#)

Avec ces éléments, l'élimination récursive de descripteurs s'établit alors en trois étapes :

- Entraînement du classifieur (optimisation des poids ω_i au regard d'une fonction de coût J);
- Calculer le critère de hiérarchisation $(\omega_i)^2$ pour chaque descripteur;
- Exclure le descripteur contribuant le moins à la classification.

À chaque itération, le SVM-RFE produit un sous-ensemble de descripteurs hiérarchisé, avec lequel il réapprend. Par conséquent, cette méthode peut être coûteuse compte tenu du nombre d'apprentissages successifs à effectuer. Elle présente cependant un avantage dans une problématique d'exploration de données. En effet, l'algorithme peut s'arrêter de deux manières : soit l'utilisateur indique la taille du sous-ensemble à atteindre, soit l'algorithme s'arrête de lui-même lorsque, par validation croisée, les performances baissent lors de l'entraînement du classifieur. Le précédent sous-ensemble est alors retourné.

4.4 Apprentissage profond (*deep learning*)

Parmi les méthodes d'apprentissage, les modèles d'apprentissage profond (ou *deep learning*) constituent un champ d'activité scientifique de plus en plus actif. En raison de leurs excellents résultats et leur grande applicabilité aux différents problèmes scientifiques liés

à l'optimisation, ces techniques jouissent d'une popularité grandissante depuis le début des années 2010.

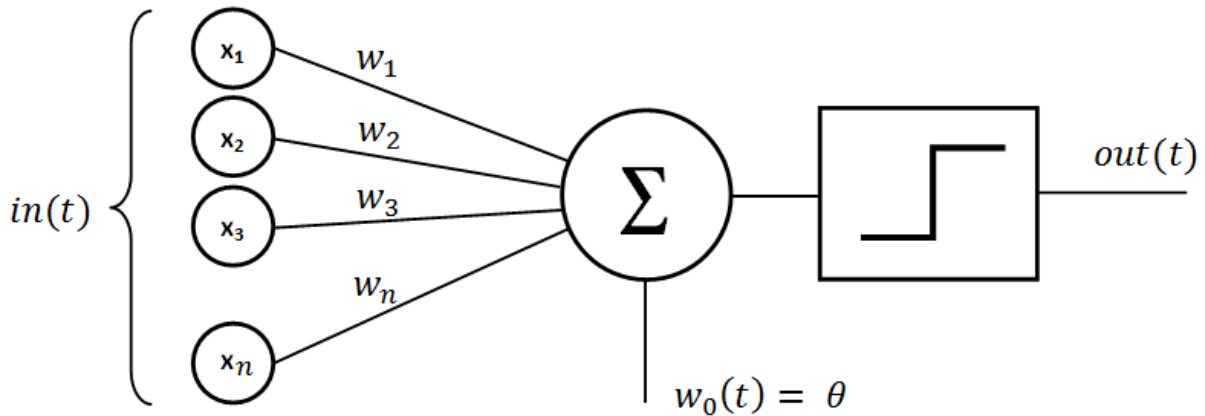


FIGURE 4.15 : *Illustration du perceptron simple. Ce dernier modélise le comportement d'un neurone biologique. Des échantillons sont présentés à des dendrites en entrées. Leurs signaux sont pondérés selon leur importance et sont sommés dans le noyau central. Si la valeur dépasse un certain seuil, alors un signal est envoyé au neurone suivant. Un neurone de ce type correspond à un classifieur linéaire. Ainsi, plusieurs neurones peuvent être organisés en réseaux pour atteindre une grande puissance combinatoire.*

Il est intéressant de voir comment l'apprentissage profond trouve ses racines dans un contexte historique favorable à l'émergence de l'informatique moderne et des neurosciences. En effet, les neurosciences apparaissent comme discipline universitaire à la fin des années 50 [26], et dès 1958, on voit dans les travaux de Von Neumann se tisser un parallèle entre le fonctionnement cérébral et le fonctionnement des ordinateurs [120]. Aussi est-ce sans surprise que ce soit dans ce contexte que Frank Rosenblatt propose le perceptron (Figure 4.15) comme structure de modélisation neuronale. Les théories de l'esprit de l'époque, computationnalistes, amenaient à aborder la cognition comme un problème de calcul. C'est ainsi que dans la modélisation proposée par Rosenblatt, le neurone biologique s'est retrouvé traduit en un neurone informatique émettant un signal binaire, équivalent du potentiel d'action, en réponse à une somme de stimuli,

traduction des multiples capteurs dendritiques [100]. L'implémentation d'un seuil de déclenchement, dépendant des entrées pondérées, et donnant une sortie binaire, permet un fonctionnement analogue à celui d'un classifieur linéaire.

Une organisation avec de multiples perceptrons, sur une même couche puis multicouches, en réseaux, permet une combinaison pondérée de ces classifieurs linéaires, résultant en des frontières de séparation de plus en plus complexes. La technologie des années 1960 ne répondait pas aux besoins en puissance de calcul de ces structures et devra attendre le début des années 2000 pour voir ses premières applications sur de grands jeux de données. On appelle apprentissage profond ou *deep learning* l'application de cette puissance de discrimination des réseaux de neurones à d'importants jeux de données.

4.4.1 Analyse structuraliste des réseaux d'apprentissage profond

Un point de vue structuraliste amène à penser que différentes structures vont produire différents effets. Si cette formulation peut sembler tautologique, elle nous amène cependant à considérer comment l'organisation des couches successives de neurones et leurs interconnexions font émerger des propriétés propres à une architecture de réseau, comme par exemple un effet de mémoire. L'analyse structuraliste permet d'organiser la pensée face à la grande technodiversité du *deep learning*. Les réseaux d'apprentissage automatique profond sont organisés autour de séquences de couches de convolution, d'activation, ou de couches denses. Dans leur travail, Anaraki et al. expliquent les effets de chaque couche de façon très détaillée [6]. Ainsi, on peut comprendre dans quel but certaines des structures de référence ont été élaborées pour répondre à des caractéristiques spécifiques de leur(s) problème(s) à traiter.

Les réseaux denses : La première de ces structures est le réseau dense, équivalent du perceptron multicouches, représenté Figure 4.16. On identifie la couche d'entrée, les couches cachées et la couche de sortie. Son fonctionnement lors de l'apprentissage est simple : un vecteur d'entrée représentant un individu est présenté à la couche d'entrée, et

le signal se propage de couche en couche, d'activation en activation, jusqu'à la couche de sortie. Entre la présentation d'un échantillon à la couche d'entrée et le calcul du résultat à la couche de sortie, il n'y a pas de rétroaction d'une couche vers une couche antérieure. Le signal avance, de couche en couche, dans un réseau de complexité d'analyse grandissante. La couche de sortie délivre le signal de réponse estimée, qui est ensuite comparé à la sortie attendue. Une erreur est calculée, puis rétropropagée le long du réseau avec l'application d'un facteur correctif. On parle de rétropropagation du gradient. Ainsi, la présentation d'exemples successifs ajuste les pondérations de chaque neurone pour correspondre de plus en plus aux valeurs de sorties désirées.

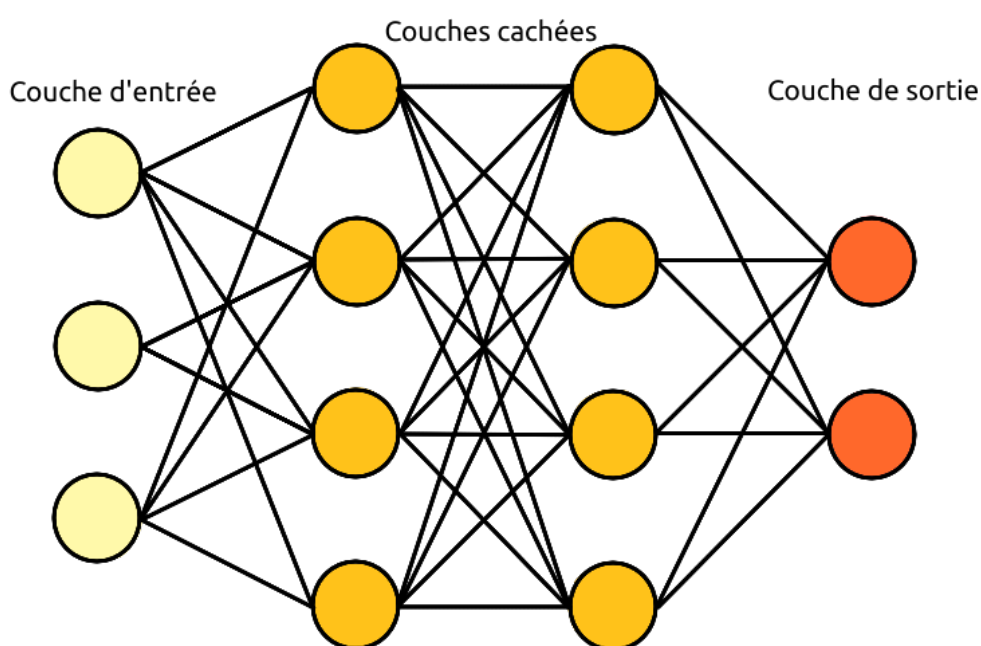


FIGURE 4.16 : *Un réseau dense est composé d'une couche d'entrée, de couches cachées et d'une couche de sortie. À chaque lien correspond un poids : ce sont les paramètres que le modèle va devoir apprendre par rétropropagation du gradient.*

Dans le réseau, chaque neurone émet un signal, fonction de la somme de ses entrées pondérées. Dans le perceptron simple, le signal sortant du coeur neuronal était simplement seuillé entre 0 et 1 selon un seuil paramétrable. L'application de ce seuil s'appelle fonction d'activation, car cette fonction transmet l'information ou non, tout comme le neurone transmet le potentiel d'action ou non selon ses entrées. En pratique, les réseaux

denses utilisent d'autres fonctions d'activation, dont plusieurs exemples sont donnés dans la Figure 4.17 et le Tableau 4.3.

Nom de la fonction	Équation
<i>Rectified Linear Unit</i> (ReLU)	$ReLU(x) = \max(0, x)$
<i>Leaky Relu</i>	$LeakyReLU(x) = \begin{cases} \alpha x, & \text{si } x < 0, \text{ avec } \alpha > 0 \\ x, & \text{si } x > 0 \end{cases}$
Tangente hyperbolique	$\tanh(x) = \frac{2}{1+e^{-2x}} - 1$

TABLE 4.3 : *Trois fonctions d'activation fréquemment utilisées.*

Les réseaux de neurones récurrents : L'introduction de rétroactions d'une couche vers une couche antérieure introduit un effet de mémoire, par récurrence du signal. Ces réseaux sont appelés réseaux de neurones récurrents (RNN) et peuvent être utilisés pour des applications présentant un aspect séquentiel. En particulier, on peut citer la structure *Long short-term memory* (LSTM) [52], utilisée dans la reconnaissance de langage.

Les réseaux neuronaux convolutifs : L'ajout d'une dimension supplémentaire au vecteur d'entrée a permis l'application des réseaux neuronaux au traitement d'image. Des techniques comme les cartes auto-organisatrices [59] permettaient déjà une interprétation spatiale du signal, mais ce sont les réseaux neuronaux convolutifs (CNN) proposés par Yann LeCun en 1995 [65] qui révolutionnèrent le plus le champ de l'analyse d'image par apprentissage automatique. Les réseaux neuronaux convolutifs utilisent l'opération de convolution pour traiter directement des signaux bi- voire multidimensionnels par l'application d'un noyau (*kernel*) sur l'image. Un exemple de convolution est représentée sur la Figure 4.18.

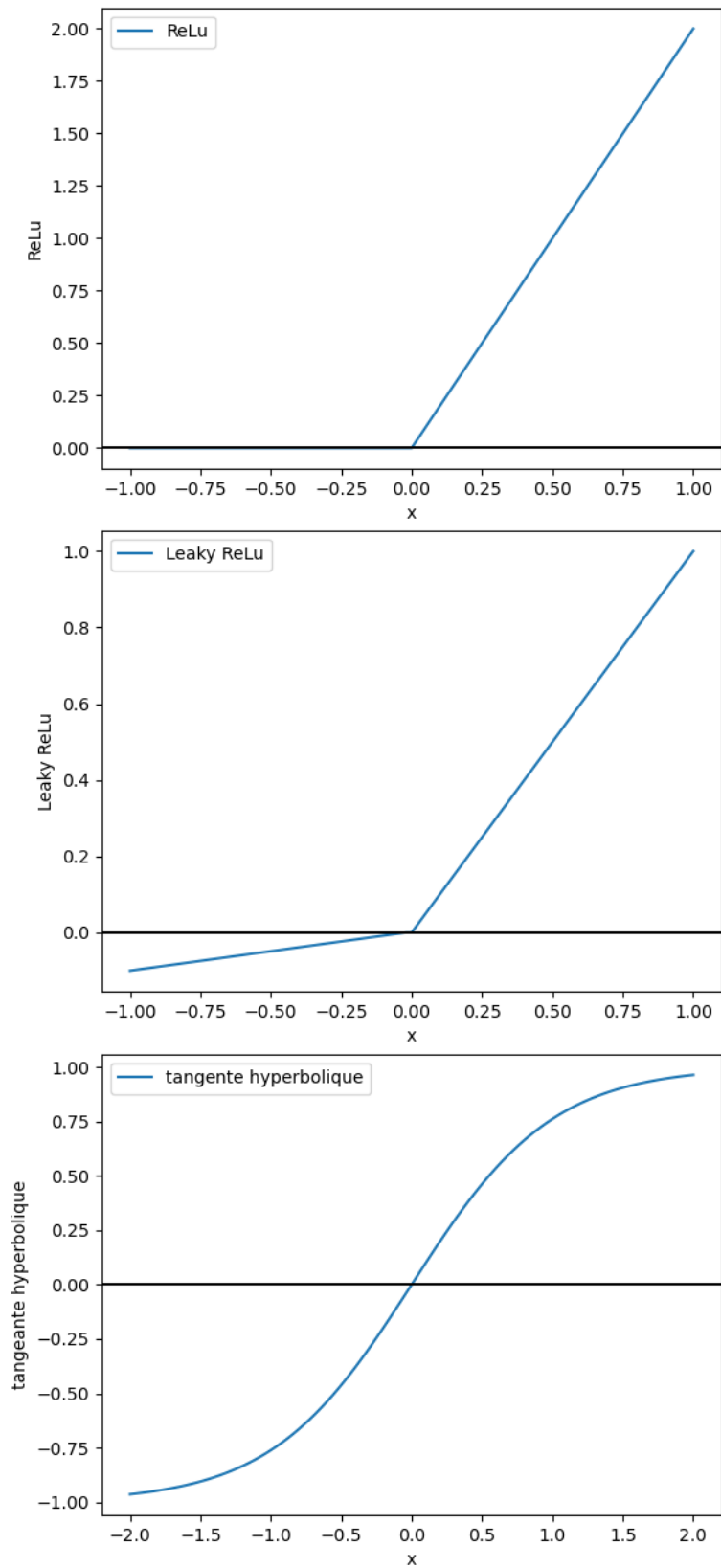


FIGURE 4.17 : Trois fonctions d'activation. De haut en bas : ReLu, Leaky ReLu et tangente hyperbolique.

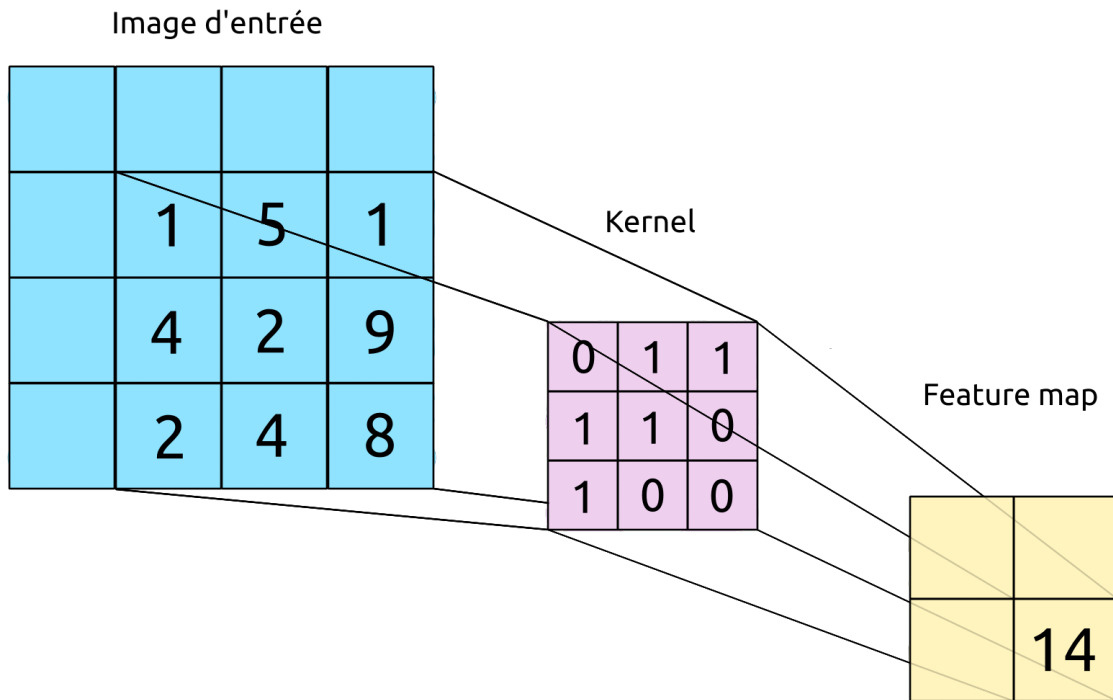


FIGURE 4.18 : Application d'un noyau convolutif sur une image d'entrée. Il en résulte une carte de description spécifique du noyau appliqué.

Le résultat de la convolution par le noyau est une image spécifique du noyau appelée carte de description (*feature map*). Ces *features maps* ressemblent à des cartes d'activation qui vont pouvoir être recombinaées entre elles par de nouvelles couches convolutives pour produire une analyse d'image de plus en plus complexe et fine au fur et à mesure de l'avancée du signal dans le réseau. Le réseau apprend en modifiant les poids des noyaux d'activation par rétropropagation du gradient. L'un des intérêts des réseaux convolutifs vient de leur capacité à identifier et hiérarchiser par eux-même les informations pertinentes pour interpréter l'image et répondre à la tâche demandée avec des performances souvent supérieures aux algorithmes d'apprentissage automatique superficiels. De plus, l'analyse des *features map* et des noyaux d'activation à chaque couche peut fournir à l'utilisateur des bribes de compréhension sur le fonctionnement réel du réseau.

Les réseaux adverses génératifs : Proposés en 2014 par Ian Goodfellow [46], les réseaux adverses génératifs (GAN) combinent un réseau générateur d'image et un réseau classifieur, comme représenté Figure 4.19. Le générateur reçoit un bruit aléatoire en entrée et génère une image artificielle. Le but du classifieur est de différencier une image générée par ordinateur d'une image naturelle. Le générateur et le classifieur sont mis en concurrence l'un avec l'autre, le premier essayant de produire des images capables de tromper le second. Cette organisation permet de produire, après entraînement, des images factices ressemblant aux acquisitions naturelles. C'est pourquoi, on retrouve ces structures dans les techniques d'augmentation de données.

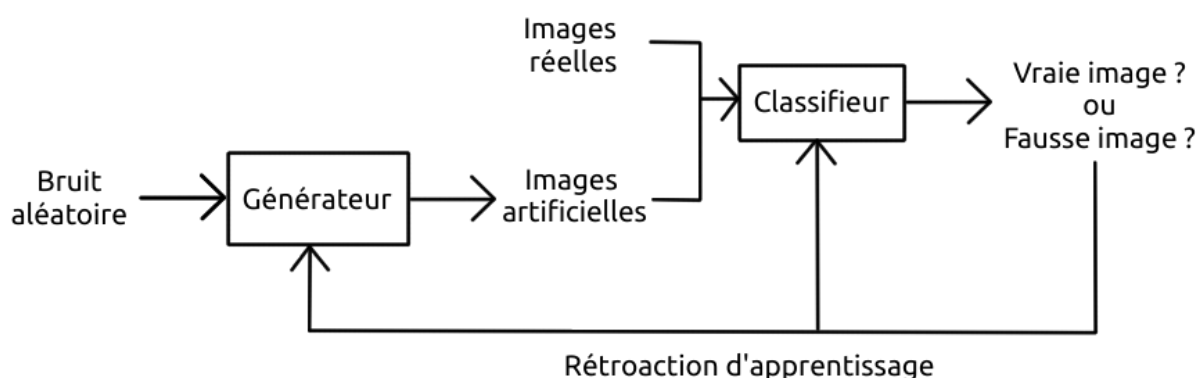


FIGURE 4.19 : Schéma de la structure GAN. Un générateur d'image et un classifieur sont mis en compétition pour produire des images artificielles proches d'images réelles.

Le réseau en U (U-net) : Cette structure proposée en 2015 [99] est utilisée pour des tâches de segmentation d'images biomédicales. Le réseau est organisé en deux parties : une voie descendante d'analyse multi-échelle d'image, et une voie ascendante générative permettant de fournir une carte de segmentation de la tâche étudiée. L'analyse se fait à chaque échelle par des convolutions successives à une même profondeur. Le changement d'échelle s'effectue par un sous-échantillonnage (*pooling*). La construction de l'image de sortie s'effectue par sur-échantillonnage. La résultante est combinée avec l'information de la voie descendante à profondeur équivalente.

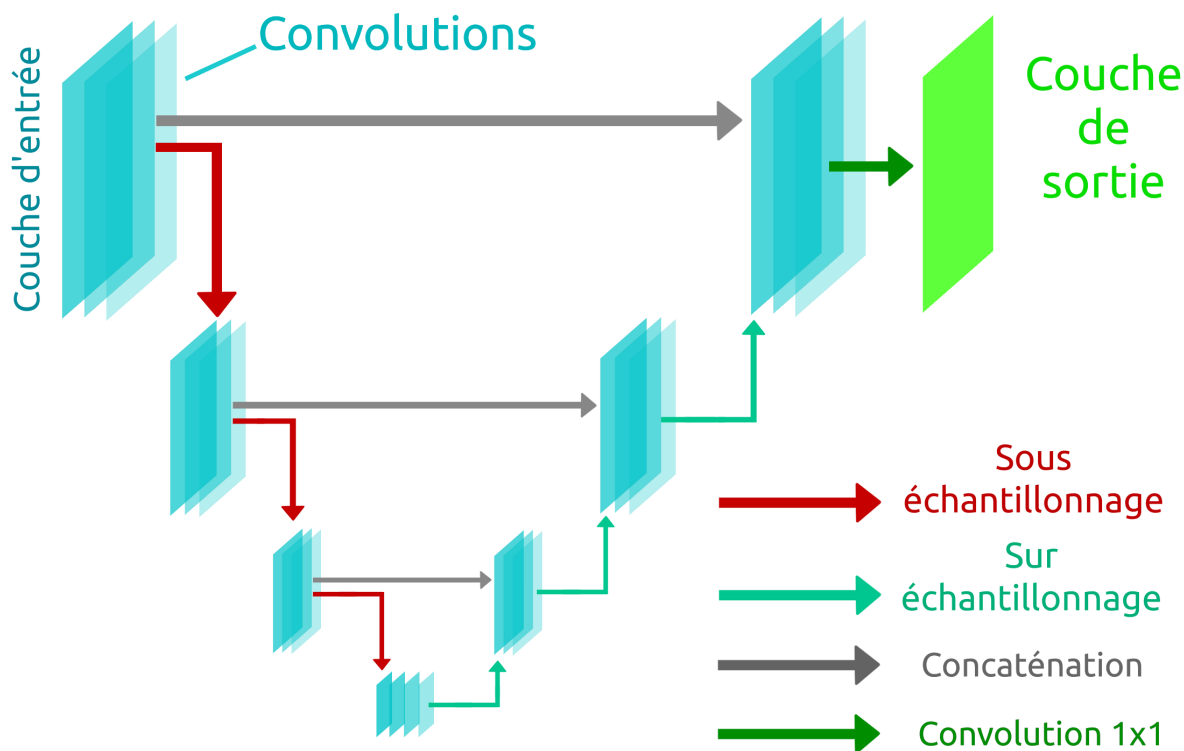


FIGURE 4.20 : Schéma de la structure U-net. Des sous-échantillonnages successifs et des convolutions permettent une analyse multi-échelle des images.

L'apprentissage automatique profond requiert de grandes quantités de données et une puissance de calcul importante. Il est courant d'avoir recours à des supercalculateurs, des baies de calcul réservées spécifiquement à l'entraînement des algorithmes. Ces ressources ne sont pas toujours disponibles. Pour contourner ce problème, des techniques d'apprentissage par transfert sont utilisées. Le principe est d'utiliser les poids d'un réseau pré-entraîné (par exemple un réseau spécialisé dans l'analyse d'images naturelles) et de les figer pour ne réapprendre que les poids proches de la couche de sortie. On bénéficie alors de la puissance d'analyse du réseau pré-entraîné au travers ses couches cachées figées, mises au service d'un problème spécifique via l'entraînement des poids proches de la couche de sortie. L'apprentissage par transfert permet de réduire à la fois le nombre de données nécessaires à l'obtention de bons résultats, et le temps de calcul de la phase d'apprentissage.

Tout au long de cette première partie, nous avons présenté les notions essentielles à la bonne compréhension de ce travail de thèse, partant de la description anatomique du cerveau, de la dynamique tumorale du gliome, et explicitant les différents outils d'exploration, IRM multimodale et algorithmes d'apprentissage automatique, que nous allons utiliser dans la deuxième partie pour caractériser automatiquement le grade du gliome et offrir ainsi un outil d'aide au diagnostic.

Méthodologie et applications

Avant-propos

Cette deuxième partie présente le travail accompli au cours des trois années de thèse. Ce travail a donné lieu à deux publications dans des journaux à comité de relecture. A l'heure de l'écriture de ce manuscrit, un troisième article est également en cours de rédaction. Le premier chapitre de cette partie présente une revue de l'état de l'art portant sur la classification du grade du gliome par apprentissage automatique. Ce champ scientifique connaît depuis 2016 une activité grandissante. Cette étude nous servira de base pour développer notre propre classifieur.

Le chapitre suivant présente nos premiers résultats. Ayant appliqué un classifieur d'apprentissage automatique classique sur la base de données la plus utilisée dans notre revue de littérature, nous nous sommes rendus compte avec un outil d'analyse de performance innovant que les labels de grades associés à cette base n'étaient pas en accord avec la division haut grade et bas grade de l'OMS. Par conséquent, nous proposons une nouvelle vérité terrain en accord avec l'OMS, dans les limites posées par l'analyse de l'imagerie anatomique. Ce travail a fait l'objet d'une seconde publication.

Un troisième chapitre explore les performances d'un algorithme entraîné sur les données d'IRM multimodales du CHU de Poitiers. Nous proposons pour cela le développement d'une chaîne de traitement complète, allant du fichier issu de l'acquisition machine au classifieur automatique. Nous utilisons ensuite deux méthodes d'analyse de hiérarchisation de descripteurs. L'analyse de ces résultats nous permet de revenir sur l'analyse d'imagerie du radiologue, mais en y apportant cette fois-ci un éclairage nouveau issu de l'apprentissage automatique. Le travail et les résultats de ce chapitre font l'objet d'un article en cours de rédaction.

Enfin, un dernier chapitre permet de revenir sur les résultats présentés tout au long de manuscrit, tout en offrant différentes perspectives et applications des outils développés lors de ce travail de thèse.

Classification du grade du gliome par apprentissage
automatique sur données RMN : avancées récentes

Contenu du chapitre

5.1 Introduction	117
5.2 Prétraitement et cadre applicatif	118
5.2.1 Un désir de données RMN multiparamétriques	119
5.2.2 Recalage multimodal	120
5.2.3 Normalisation en intensité	120
5.2.4 Segmentation tumorale	121
5.2.5 Augmentation de données	121
5.3 Descripteurs	122
5.3.1 Descripteurs anatomiques	122
5.3.2 Descripteurs de diffusion, perfusion et spectroscopie	122
5.3.3 Algorithmes de hiérarchisation de descripteurs	123
5.4 Classifieurs	124
5.5 Synthèse	126
5.6 Mise à jour du travail présenté	127
5.7 Conclusion	128

Le CHU de Poitiers dispose d'une base de données multimodales de gliomes de bas grades et de hauts grades. Dans ce travail de thèse, nous proposons des méthodes à base d'apprentissage automatique dans un but d'aide au diagnostic. Nous souhaitons différencier, depuis les images issues du bilan initial multimodal, si le patient est atteint d'un haut grade ou d'un bas grade. On a vu que cette démarcation était majeure dans la prise en charge thérapeutique. C'est pourquoi nous allons présenter dans un premier temps quelles méthodes sont appliquées et quelles séquences RMN sont utilisées de nos jours pour la classification du grade des gliomes.

La classification du grade du gliome depuis les données de résonance magnétique nucléaire (RMN) par des techniques d'apprentissage automatique est un champ scientifique de plus en plus actif. La Figure [5.1](#) montre une intensification du nombre de travaux portant sur ce sujet depuis 2016-2017. Ce chapitre passe en revue une sélection des méthodes les plus récentes, publiées entre 2018 et 2019. Depuis la parution initiale de ce travail, nous avons rajouté section [5.6](#) un retour sur les travaux les plus marquants de 2019 et 2020.

Nous y détaillons les prétraitements techniques de chaque auteur, et le cadre applicatif de leurs travaux, incluant les différentes modalités et séquences RMN utilisées dans leur jeu de données. Nous regroupons ensuite les différentes approches en comparant leurs schémas d'apprentissage.

Alors que l'apprentissage automatique classique reste fréquemment utilisé, de plus en plus d'auteurs appliquent des réseaux neuronaux convolutifs avec un certain succès. L'ajout de modalités RMN supplémentaires, telles que l'imagerie de perfusion, l'imagerie de diffusion ou la spectroscopie RMN, octroie une plus grande diversité d'informations. Nous observerons dans ce chapitre que les travaux utilisant ces modalités supplémentaires atteignent une précision de classification plus élevée que ceux reposant uniquement sur l'IRM anatomique.

5.1 Introduction

Commençons par rappeler, comme vu lors section [2.4](#), que la technique de référence pour l'établissement du grade est l'analyse histologique issue d'une biopsie ou de l'examen d'une pièce chirurgicale. Ces méthodes sont invasives et sont soumises à la variabilité inter-observateur et à l'erreur d'échantillonnage. Nous cherchons à créer, en utilisant les données multimodales du CHU de Poitiers, une alternative non-invasive et robuste. En effet, comme décrit section [3.8](#), l'examen IRM offre une grande variété d'informations utilisées par les radiologues pour estimer le grade du gliome. La combinaison des séquences anatomiques, de l'imagerie de diffusion et de perfusion et de la SRM permettent au radiologue d'affiner son diagnostic. Cette variété d'information fait de l'IRM un outil multimodal pour l'analyse oncologique.

L'apprentissage automatique basé sur l'analyse radiomique, c'est à dire issue des descripteurs d'imagerie, permet le traitement de ces grandes quantités de données. Comme nous l'avons détaillé dans section [4.1](#), nous faisons une distinction entre les approches d'apprentissage automatique classique et les méthodes d'apprentissage automatique profond. Les premières procèdent par une extraction de descripteurs choisis par l'utilisateur avant d'être traités par un classifieur. Par opposition, l'apprentissage profond repose sur des réseaux neuronaux convolutifs (CNN), capables d'apprendre leurs propres descripteurs.

Une revue de littérature de 2019 [\[66\]](#) résume le champ d'application des techniques d'apprentissage automatique sur les données RMN de gliomes à deux tâches principales : l'analyse d'image dans un but de diagnostic et la segmentation des tumeurs dans un but de suivi de l'évolution tumorale. En particulier, la tâche de classification automatique du grade du gliome est une tâche en pleine expansion. Le nombre de publication consacrées à ce sujet est en augmentation et montre une accélération depuis 2016, comme représenté sur la Figure [5.1](#). Dans ce chapitre, nous présenterons les avancées récentes dans la tâche de classification du grade des gliomes par apprentissage automatique classique et profond, à travers 11 articles publiés entre 2018 et 2019. Nous verrons ainsi quels

traitements communs sont actuellement appliqués dans ce champ scientifique.

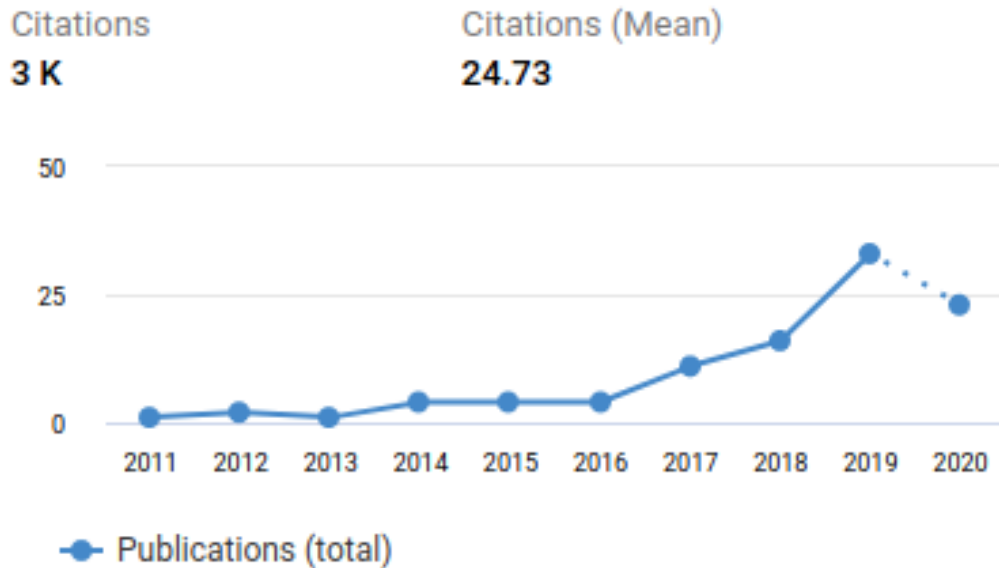


FIGURE 5.1 : Nombre de publications par années incluant les mots clés "glioma grade classification learning". Graphique généré par Dimension.ai [104]. La classification automatique du grade par apprentissage automatique apparaît comme une tâche de plus en plus active depuis 2016.

Dans une première partie, nous présentons quelles bases de données et quelles séquences RMN sont actuellement utilisées pour la classification du grade des gliomes. Une deuxième partie détaille comment les données RMN sont pré-traitées et quels descripteurs sont extraits avant l'étape d'apprentissage. Nous décrirons ensuite les différents classifieurs et structures de réseaux convolutifs utilisés. Enfin, une dernière partie propose une synthèse des résultats des travaux présentés et une discussion.

5.2 Prétraitement et cadre applicatif

En fonction des bases de données disponibles, les travaux examinés se développent dans des cadres applicatifs techniques différents. Par exemple, certains jeux de données disponibles présentent des acquisitions où l'encéphale a été préalablement extrait et où

les séquences RMN sont déjà recalées entre elles. Les étapes de prétraitement détaillées dans les différents articles dépendent de ces prérequis. C'est pourquoi il existe une diversité dans les prétraitements appliqués, dépendants ultimement des bases de données utilisées. Ainsi, des étapes comme le recalage multimodal, la normalisation en intensité ou l'application de certains filtres comme l'extraction de l'encéphale ou la correction de biais de champ sont à la charge de l'utilisateur.

5.2.1 Un désir de données RMN multiparamétriques

Concernant la tâche de classification du grade des gliomes, il est difficile de trouver une base de données fournissant abondance et diversité d'acquisition RMN. Il existe des bases de données anatomiques contenant des centaines de patients, comme le propose *The Cancer Imaging Archive* (TCIA) [24] ou le challenge de segmentation tumorale BraTS (*Brain Tumor Segmentation Challenge*) [74, 8, 10]. Ce dernier comprend 4 séquences anatomiques, à savoir les pondérations T1 et T2, T2 FLAIR et un T1 post-injection de gadolinium (T1c). Les données BraTS sont recalées entre elles et l'encéphale y a été segmenté. Ces deux étapes préalablement réalisées simplifient le nombre de prétraitement à appliquer avant classification. Mais ce jeu de données est limité aux séquences anatomiques. La création d'un ensemble de données utilisant imagerie de diffusion, de perfusion ou la SRM relève d'un processus coûteux en temps machine. Cependant, de telles données commencent à être utilisées dans lors de travaux concernant les gliomes de bas grades, comme le ciblage tumoral multimodal [83]. Pour la classification du grade, les publications récentes utilisant ces modalités génèrent leurs propres jeux de données, issus de leur propre protocole d'acquisition [116, 96, 23]. La mise en oeuvre d'un tel protocole limite la taille de ces jeux de données propriétaires, puisqu'on trouve parmi ces travaux récents des jeux de données n'allant que de 40 patients (contenant séquences anatomiques, de diffusion, de perfusion et SRM) [116], à un maximum de 112 patients (ne disposant ici que de SRM seule) [96].

5.2.2 Recalage multimodal

Lorsqu'il n'est pas préalablement appliqué au jeu de données, le recalage est une étape importante de prétraitement. Cette opération crée une correspondance anatomique de voxel à voxel entre deux volumes de RMN différents par une transformation géométrique. Une étude récente présente 3 méthodes différentes de recalage [77] : la cartographie paramétrique statistique (SPM), l'application FSL-FLIRT de la FMRIB Software Library 5.0 (FSL) [12] et le logiciel de recalage d'image automatique (AIR pour *Automatic Image Registration software*). Dans notre revue d'articles, le logiciel AIR est utilisé pour le recalage de l'imagerie de diffusion tandis que la SPM et FSL-FLIRT sont utilisés pour du recalage anatomique. Mais la seule librairie FSL peut aussi être utilisée pour recaler l'intégralité des modalités, tout en ré-échantillonnant les séquences en un volume isotrope dont les voxels font 1 mm d'épaisseur [116].

5.2.3 Normalisation en intensité

Les inhomogénéités de champ magnétique peuvent créer des artefacts lors des acquisitions RMN. Parmi nos références traitées, l'algorithme N4 [7] de la bibliothèque de traitement d'image ITK (*Image Tool Kit* [15, 122, 69]) a été utilisé à deux reprises [101, 20], et celle implémentée dans FSL a été utilisée sur des données multimodales [116].

En ce qui concerne la normalisation en intensité, plusieurs méthodes peuvent être appliquées. Le traitement le plus simple, centrer et réduire indépendamment pour chaque séquence, est d'ailleurs appliqué dans une publication [101]. Mais pour permettre une meilleure correspondance en intensité entre les séquences, il est possible d'effectuer au préalable une correspondance d'histogramme grâce à la bibliothèque ITK. Cette correspondance permet d'aligner l'histogramme de chaque volume en un histogramme de référence choisi, avant de centrer et réduire. Ce traitement est appliqué dans une seule de nos publications [20]. Un autre auteur a simplement normalisé chaque canal d'entrée entre 0 et 1 [6]. Ainsi, nous voyons que, dans les différentes publications, le processus de normalisation en l'intensité varie, sans qu'un consensus technique ne semble émerger.

5.2.4 Segmentation tumorale

Une segmentation tumorale peut être utilisée pour extraire une région d'intérêt (ROI) contenant la tumeur. Quand cette segmentation n'est pas fournie avec le jeu de données, cette dernière peut être effectuée à l'aide de CNN [101, 20]. La segmentation donne aux auteurs une estimation de la tumeur segmentée, éliminant ainsi tous les tissus sains du cerveau. La base de données BraTS fournit une carte de segmentation pour chaque patient. Afin d'augmenter la saillance des régions tumorales, Ge et al. propose de réduire d'un tiers l'intensité de chaque voxel sain dans ses deux travaux [41, 42]. Cet auteur montre ainsi que, dans son cas, cette augmentation de la saillance tumorale améliore la précision de la classification de plus de 5% [42].

5.2.5 Augmentation de données

Pour des raisons cliniques, les données de patients bas grades sont moins fréquentes que les données de patients hauts grades [87], ce qui crée un déséquilibre de classe pour les algorithmes d'apprentissage automatique. Les bases de données de petite taille peuvent demander de recourir à des techniques d'augmentation de données pour améliorer les performances. L'augmentation de données la plus riche se trouve dans le travail de Sajjad [101], où 8 types de transformations différents sont appliqués aux données : des transformations linéaires comme des rotations, des symétries gauche-droite, l'application d'inclinaisons ; mais aussi du filtrage avec la génération d'images filtrées par flou gaussien, par réhaussement de contours. L'auteur applique aussi des transformations comme la détection de bords ou le calcul de gradients directionnels (*emboss*). En appliquant chaque transformation avec 4 valeurs différentes, l'auteur indique augmenter jusqu'à 30 fois la taille de son jeu de données. Il prouve ensuite que cette augmentation des données améliore la précision de classification de plus de 3%. L'apprentissage automatique profond est également utilisé pour générer des données artificielles à l'aide de réseaux GAN (voir section 4.4.1) [50]. La mise en œuvre de telles techniques pourrait aider à résoudre le problème du déséquilibre de classes pour les jeux de données anatomiques ou multimodaux et à améliorer ensuite les performances.

5.3 Descripteurs

5.3.1 Descripteurs anatomiques

Les algorithmes d'apprentissage automatique classiques ont besoin de descripteurs calculés préalablement avant l'application d'un classifieur. Ces descripteurs sont spécifiques des modalités de RMN utilisées dans les travaux examinés. Pour ces algorithmes, un cadre commun se dégage : calculer des centaines de descripteurs, puis appliquer une méthode de hiérarchisation pour sélectionner les caractéristiques les plus discriminantes pour les présenter en entrée du classifieur. Deux de nos auteurs travaillant sur la base BraTS utilisent le logiciel open-source PyRadiomics [118] pour calculer respectivement 420 [20] et 468 descripteurs [21]. Les 420 descripteurs de la première publication sont divisés en 105 descripteurs, calculés sur les 4 séquences disponibles. Cela comprend : 18 descripteurs d'intensité, 13 de forme et 74 de texture. Dans la deuxième publication, les 468 descripteurs sont : 24 descripteurs de forme, 228 descripteurs basées sur des histogrammes et 216 descripteurs de texture, extraits de la matrice de cooccurrence des niveaux de gris (GLCM pour *Gray Level Cooccurrence Matrix*) et de matrices de zones d'intensité (ISZ pour *Intensity Size Zone*). Nous voyons que les descripteurs se répartissent en trois catégories, décrivant intensité, forme et textures.

5.3.2 Descripteurs de diffusion, perfusion et spectroscopie

Parmi les travaux récents que nous étudions, peu utilisent des modalités supplémentaires à l'IRM anatomique. Deux articles ajoutent à l'IRM anatomique des données de diffusion, de perfusion et de spectroscopie RMN [116, 23]. Pour être complet dans notre revue de littérature multimodale, nous ajoutons une référence qui ne s'appuie que sur les données de SRM [96]. Ces trois articles utilisent des zones d'intérêt (ROI) pour l'extraction de leurs descripteurs. Pour l'imagerie de perfusion, le traitement s'effectue sur les cartes paramétriques, où sont extraites les valeurs moyennes de différentes ROI. Les cartes paramétriques utilisées sont le volume sanguin cérébral (CBV), le flux sanguin

cérébral (CBF) et le temps de transit moyen (MTT). Pour l'imagerie de diffusion, les cartes paramétriques calculées sont la diffusivité moyenne (MD pour *Mean Diffusivity*), l'anisotropie fractionnelle (FA), l'isotropie pure (p) et l'anisotropie pure (q). Les acquisitions SRM ne donnent pas de cartes paramétriques mais un spectre d'où l'on peut extraire des concentrations et des rapports métaboliques. Ce sont ces informations qui sont utilisées comme descripteurs. Dans la référence utilisant uniquement la spectroscopie [96], l'auteur utilise 26 descripteurs métaboliques. Les descripteurs communs aux trois publications utilisant la spectroscopie sont la concentration en choline (Cho), en créatine (Cr), en Acide N-Acétylaspartique (NAA) et les rapports métaboliques Cho/Cr et NAA/Cr. Les lipides jouent également un rôle d'importance et sont mesurés dans les publications [116] et [96].

5.3.3 Algorithmes de hiérarchisation de descripteurs

Dans le champ de l'apprentissage automatique, un algorithme de sélection de descripteurs réduit la dimensionnalité du problème abordé en limitant le nombre de paramètres impliqués dans la classification. Dans notre série de publications étudiées, deux algorithmes différents sont utilisés : les SVM avec élimination récursive de descripteurs (l'algorithme SVM-RFE, tel que présenté section 4.3.9) et la sélection par pertinence maximale et redondance minimale (mRMR pour *minimum Redundancy Maximum Relevance*). Ces méthodes génèrent une hiérarchie parmi les descripteurs, à partir de laquelle l'auteur choisit un sous-ensemble contenant les meilleurs éléments de description, ce qui permet de réduire la dimensionnalité. Par exemple, Chen et al. [20] passent de 420 à 20 descripteurs en utilisant le SVM-RFE, tandis que la référence Cho et al. [21] passent de 468 à seulement 5 descripteurs après le mRMR. Parmi les publications étudiées, l'algorithme SVM-RFE est le plus couramment utilisé lors de la sélection de descripteurs. En raison des différentes séquences disponibles dans les bases de données, les descripteurs les plus discriminants diffèrent d'une publication à l'autre. C'est pourquoi il est difficile d'établir par comparaison la liste des descripteurs les plus pertinents au regard de la tâche de classification binaire. Citak-Er et al. [23] indiquent qu'à chaque applica-

tion du SVM-RFE, 6 séquences et descripteurs persistent à chaque fois : la valeur de CBV relative (rCBV), les cartographies MTT, ADC et FA, et les rapports métaboliques NAA/Cr et Cho/NAA. Parmi les publications utilisant des données multimodales, le travail de Vamvakas et al. [116] donne un classement de 21 descripteurs. Parmi les 10 caractéristiques les mieux classées, on trouve un rapport métabolique (Lipides/Cr), 8 descripteurs de perfusion et 2 descripteurs anatomiques calculés sur T1c. Les descripteurs de perfusion et anatomiques proviennent de mesures basées sur l’histogramme et l’analyse de texture. Ce résultat fait écho à d’autres travaux où les descripteurs de forme et de texture apparaissent comme étant les plus discriminants [21]. Dans cette publication, la description de la texture produite par la GLCM ressort comme étant un descripteur intéressant lorsque ce dernier est calculé depuis une séquence T1c. La pertinence de cette séquence apparaît à nouveau dans le travail de Chen et al. [20], où la séquence T1c apparaît 12 fois dans la liste des 20 descripteurs les plus discriminants.

En ce qui concerne la SRM, la publication reposant uniquement sur la spectroscopie [96] fait elle aussi apparaître un rôle discriminant des lipides. Elle identifie comme meilleurs descripteurs les concentrations en lipides, en NAA, et les rapports Lipides/NAA et Cho/Cr. Cette référence évalue l’apport de chaque descripteur via un test de Student. Ainsi, les lipides apparaissent comme un descripteur SRM d’importance pour la classification du grade.

5.4 Classifieurs

Les travaux utilisant l’apprentissage automatique classique font appel à des classifieurs issus des SVM, des forêts d’arbres aléatoires (RF) ou de leurs dérivés. Ainsi, trois publications emploient des SVM [96, 23, 116]. On notera que la publication [116] utilise une optimisation minimale séquentielle (SMO pour *Sequential Minimal Optimization*) qui est un dérivé des SVM [116]. Deux publications reposent sur des forêts aléatoires [21, 20] et leurs dérivés, comme l’algorithme XGBoost [19] utilisé dans le travail de Chen et al. [20]. Cho et al. [21] comparent trois classifieurs : une régression logistique, les SVM et les forêts aléatoires. Les performances de ces trois algorithmes y apparaissent comme

similaires, avec une précision comprise entre 88% et 89% pour les trois méthodes. De cet ensemble de publications utilisant l'apprentissage automatique classique, nous observons que le SVM semble être le classifieur le plus couramment utilisé pour la classification du grade du gliome.

Concernant l'apprentissage profond, les structures de CNN sont diverses. Citons à nouveau le travail d'Anaraki et al. [6] qui détaille les effets de chaque type de couche. Les CNN impliquent un très grand nombre de paramètres, et chacun d'entre eux doit être ajusté lors de l'apprentissage. C'est pourquoi la mise en place de telles structures est coûteuse en temps et nécessite de grands jeux de données. Pour contourner ces problèmes, le recours à l'apprentissage par transfert a été utilisé à trois reprises dans les travaux récents [101, 13, 105]. Ces trois auteurs utilisent une structure commune : le groupe géométrique visuel (VGG pour *Visual Geometry Group*) [108]. Il s'agit d'une structure développée par l'Université d'Oxford, pré-entraînée sur des images naturelles et disponible à 11, 13, 16 ou 19 couches. Parmi nos auteurs, deux utilisent la structure VGG-16 [13, 105] et un la structure VGG-19 [101].

Les publications qui utilisent de l'apprentissage profond ne calculent pas de descripteurs préalablement. Elles utilisent directement la donnée brute en entrée de leurs réseaux. Banerjee et al. [13] comparent trois formatages d'entrées différentes : l'utilisation de patches, de coupes 2D ou de volumes 3D. L'auteur conclut de cette comparaison que l'utilisation de volumes 3D donne de meilleurs résultats.

La base de données BraTS est utilisée à de nombreuses reprises [42, 13, 105, 41]. On peut s'interroger sur la méthode d'intégration des différentes séquences. Un exemple de traitement multiséquentiel est proposé par Ge et al. [41]. Les 4 séquences anatomiques sont présentées à un CNN à entrées multiples. Chaque entrée du réseau reçoit une séquence qui est ensuite traitée par 7 couches de convolution. Une couche d'agrégation permet ensuite de fusionner chaque flux et l'envoyer sur une succession de couches denses.

Les problèmes de grande dimensionnalité peuvent provoquer une disparition du gradient par malédiction de la dimension. Pour résoudre ce problème, plutôt que de réduire le nombre de dimensions comme on l'a vu en apprentissage automatique classique, le travail de Shahzadi et al. [105] propose d'utiliser une structure récurrente avec effet

mémoire (LSTM pour *Long short-term memory*) avant d'envoyer les données dans un VGG-16.

Enfin, citons le travail de Anaraki et al. [6], dont l'architecture du réseau de neurones est évolutive au cours de l'apprentissage. L'auteur utilise un Algorithmes Génétiques (GA) [76] qui génère de lui-même une architecture par une série d'essais. Ce mécanisme permet au réseau d'évoluer vers un meilleur CNN en auto-régulant ses propres paramètres, tels que le nombre de couches de convolution. A la fin de l'apprentissage, leur meilleure architecture comprend cinq couches convolutives et une couche dense de sortie.

On voit ici la grande diversité de structures applicables en apprentissage profond pour traiter le problème de classification du grade du gliome.

5.5 Synthèse

Différents métriques d'évaluation ont été utilisées dans les publications pour évaluer le classifieur, les plus courantes étant la précision, la sensibilité et la spécificité. Lorsque cela était possible, nous avons ajouté la précision équilibrée (BAcc pour *Balanced accuracy*), qui est la moyenne entre la sensibilité et la spécificité. La BAcc est pertinente pour l'évaluation des classifieurs sur les jeux de données déséquilibrés [18]. Un récapitulatif de chaque publication analysée est donnée dans le Tableau 5.7. La diversité des séquences de RMN semble améliorer la qualité de discrimination, le meilleur résultat étant obtenu par une publication multimodale avec une précision de 95,5% [116]. Pour les séquences anatomiques, on voit que 4 des 5 publications reposant sur des CNN atteignent une précision supérieure à 90%. L'apprentissage machine classique donne également de bons résultats, le SVM étant le classifieur le plus utilisé.

Dans les discussions, certains auteurs affirment qu'une classification binaire n'est pas suffisamment pertinente et devrait être étendue à une discrimination entre le grade II, le grade III et le grade IV [6] ou une classification entre chaque grade individuellement [101].

Nous notons que la base de données la plus utilisée est la base BraTS. Cette base ne

dispose que de données anatomiques, avec un déséquilibre de classe entre bas grades et hauts grades. L'utilisation de GAN pourrait résoudre le problème du déséquilibre entre les classes par augmentation de données.

5.6 Mise à jour du travail présenté

Depuis la parution originale de ce travail, de nouvelles publications ont été proposées pour répondre à la tâche de classification du grade du gliome. Nous proposons ici d'en détailler les plus marquantes, pour leurs structures d'apprentissage ou les données qu'elles utilisent.

Pour sa structure d'apprentissage, citons le travail de Zang et al. [123], qui propose d'appliquer, sur les données de diffusion uniquement, un classifieur hybride, entre apprentissage automatique classique et apprentissage profond. Les descripteurs images sont extraites par un réseau convolutif sur la zone tumorale préalablement segmentée manuellement. Sont ajoutés à ces descripteurs issus du réseau convolutif 27 descripteurs de forme et 43 descripteurs de texture. La classification se fait par un SVM, suite à une sélection de descripteurs. Ce travail combine ainsi la puissance analytique du réseau convolutif avec un classifieur classique. Avec une base composée de 108 patients (43 bas grades et 53 hauts grades), les auteurs atteignent une précision de 94%. Ils testent également la classification des grades III contre les grades IV et obtiennent une précision de 98%, ce qui est particulièrement élevé.

Le travail de Sun et al. [114] compare 16 méthodes de sélection de descripteurs et 15 classifieurs différents. L'auteur utilise la bibliothèque Pyradiomics pour calculer ses descripteurs sur la segmentation tumorale fournie par la base BraTS 2018. Ces descripteurs sont calculés sur la tumeur complète, mais aussi sur des zones intra-tumorale, comme le cœur tumoral seul, la zone de prise de contraste, la zone nécrotique ou l'œdème. A l'issue de son travail, le meilleur résultat de classification est obtenu en utilisant un SVM linéaire pour la sélection de descripteurs et un réseau dense pour la classification. L'auteur note que les descripteurs les plus récurrents suite à la sélection de descripteurs sont les descripteurs d'intensité issus de l'histogramme et les descripteurs de textures issus

de la GLCM et de la GLRLM (*gray level run length matrix*). La combinaison du SVM linéaire et du réseau dense permet d'atteindre une précision de classification de 94,4%.

Citons enfin le travail de Huang et al. [53] qui propose d'entraîner, sur les données de la base BraTS, un réseau GAN qui génère, à partir de la seule séquence T1, les images T1c, T2 et T2 FLAIR. Ce réseau GAN calcule depuis le T1 un ensemble de descripteurs lui servant ensuite à générer les autres séquences manquantes. Partant de ces descripteurs, les auteurs implémentent une cascade de couches convolutives capable de prédire le grade avec une précision de 87,9%. Le même réseau entraîné sur les 4 séquences BraTS atteint 89,4% de précision. Ce travail offre des perspectives intéressantes pour augmenter les bases de données à partir d'une unique séquence T1, ce qui permettrait de réduire le temps machine d'acquisition pour chaque patient.

Parmi les travaux de 2019 et 2020, nous notons la grande utilisation de la base de données BraTS [27, 5] et une absence de publications combinant imagerie anatomique, de diffusion, de perfusion et SRM. Ce dernier point souligne l'intérêt de notre démarche d'inclure ces modalités au sein d'un même travail à l'aide des données du CHU de Poitiers.

5.7 Conclusion

La classification du grades du gliome par apprentissage automatique sur des données RMN est une tâche intéressante où s'exprime une grande diversité technique. Les travaux récents de 2018 à 2020 sont différents dans leur approche technique du problème et dans leur cadre applicatif, conditionné par leur jeu de données. Ainsi, des bases de données, selon leur cadre applicatif, nécessitent des prétraitements adaptés. Les schémas classiques d'apprentissage automatique classique utilisent majoritairement les SVM comme classifieur. Concernant l'apprentissage profond, une grande diversité de structure existe. L'application d'un VGG avec un apprentissage par transfert domine. Les CNN donnent des résultats intéressants, bien qu'ils n'aient pas encore été utilisés sur des données multimodales, combinant des séquences de diffusion, de perfusion et de spectroscopie. Pour l'apprentissage classique, les jeux de données multimodales donnent

de meilleurs résultats que les bases de données limitées aux seules données anatomiques. Enfin, nous notons que certains articles discutent de la pertinence d'une classification binaire, là où une classification à quatre grades pourrait s'avérer plus intéressante pour les radiologues.

Après avoir passé en revue les méthodes récentes, nous proposons tout d'abord d'implémenter un classifieur intelligible sur la base de données BraTS. Cette implémentation doit nous permettre d'obtenir un point de référence des performances de notre algorithme avant d'y ajouter les données multimodales du CHU de Poitiers. Nous observerons ainsi clairement si l'ajout de ces données permet d'augmenter la précision de classification. Nous faisons le choix d'utiliser un classifieur d'apprentissage automatique classique car ces derniers sont plus intelligibles que les classifieurs d'apprentissage automatique profond. L'implémentation de ce classifieur simple et l'analyse de ses résultats sur les données BraTS font l'objet du chapitre suivant.

Le travail présenté dans ce chapitre a fait l'objet d'une publication [32] pour la conférence IEEE ICABME 2019 (*International Conference on Advances in Biomedical Engineering*).

Référence	Modalités	Classifieur	Précision	Sensibilité	Spécificité	Précision équilibrée
Citak-Er [23]	Multimodales ¹	SVM	93%	96.4%	86.7%	91.55%
Vamvakas [116]	Multimodales ²	SMO (SVM)	95.5%	95%	96%	95.5%
Qi [96]	Spectroscopie	SVM	76.31%	89.5%	63.2%	76.35%
Sajjad [101]	Anatomique	VGG-19	94.58%	88.41%	96.12%	92.27%
Chen [20]	Anatomique	XGBoost (RF)	91.27%	NA ⁴	NA ⁴	NA ⁴
Anaraki [6]	Anatomique ³	CNN	93.94%	96.6%	88.76%	92.68%
Ge [42]	Anatomique	CNN	89.47%	90.48%	86.67%	88.58%
Cho [21]	Anatomique	RF	88.77%	94.29%	73.33%	83.81%
Banerjee [13]	Anatomique	VGG-16	95%	NA ⁴	NA ⁴	NA ⁴
Shahzadi [105]	Anatomique	VGG-16	84%	NA ⁴	NA ⁴	NA ⁴
Ge [41]	Anatomique	CNN	90.87%	91.27%	90.48%	90.87%

¹Jeu de données propriétaire. Inclus des données anatomiques, de diffusion, de perfusion et de SRM.

²Jeu de données propriétaire. Inclus des données anatomiques, de diffusion et de SRM.

³L'auteur n'utilise que les séquences T1, T1c et T2.

⁴Résultat non disponible.

TABLE 5.1 : *Récapitulatif des différentes publications traitées*

Une nouvelle vérité terrain à base de critères radiologiques pour
la base de données BraTS

Contenu du chapitre

6.1 Avant-propos	132
6.2 Introduction	134
6.3 Cadre du travail	135
6.3.1 Analyse radiomique de données IRM anatomiques	135
6.3.2 Classification du grade par apprentissage automatique	136
6.4 Classification binaire et analyse des résultats	137
6.4.1 Données utilisées	137
6.4.2 Descripteurs	138
6.4.3 Résultats et cohérence des critères radiologiques	140
6.4.4 Analyse radiomique des faux positifs et faux négatifs	149
6.5 Une nouvelle vérité terrain pour le BraTS	150
6.5.1 Obtention d'étiquettes radiologiques	150
6.5.2 Évaluation d'un classifieur SVM entraîné sur la vérité terrain radiologique	151
6.6 Discussion et conclusion	156

6.1 Avant-propos

Ce chapitre occupe une place très particulière par rapport au travail effectué durant cette thèse. En effet, alors que nous entraînions un classifieur d'apprentissage automatique classique sur les données de la base BraTS, que nous avons choisie comme base de référence tant cette dernière est utilisée dans l'état de l'art récent, nous nous sommes rendus compte que cette base de données ne répondait pas aux critères de division haut grades et bas grades de l'OMS. En effet, cette base est divisée de manière binaire, mais avec d'une part des glioblastomes multiformes (grades IV) et d'autre part un regroupement des gliomes "de grades inférieurs [au grade IV]", sous l'appellation *Lower Grade Glioma*. Cette appellation est à l'origine d'une confusion avec le terme *Low Grade Glioma*, désignant les bas grades de l'OMS, et partageant le même acronyme : LGG. Cette différence de classification est représentée dans le Tableau [6.1](#)

Grade I	Grade II	Grade III	Grade IV
LGG _{WHO}		HGG _{WHO}	
LGG _{BraTS}			HGG _{BraTS}

TABLE 6.1 : *Grades des gliomes selon l'OMS (WHO) et la base de données BraTS.*

De nombreux auteurs utilisent la base BraTS lors de leurs travaux en pensant produire une classification alignée sur l'OMS [\[42, 20, 27, 22, 80, 125\]](#) (échantillons représentatifs et non exhaustifs). Dès que la base BraTS est utilisée, la confusion entre les LGG de l'OMS et les *Lower grade gliomas* du BraTS est presque systématique.

Nous nous sommes rendu compte de ce décalage alors que nous développons un outil d'analyse de performances du classifieur, l'histogramme des scores de prédiction, qui fait d'ailleurs l'objet d'une contribution importante de ce chapitre (section [6.4.3](#)). L'analyse de cet histogramme nous indiquait un comportement étrange de plusieurs patients bas grades du BraTS (LGG_{BraTS}), que notre classifieur étiquetait comme hauts grades (HGG_{BraTS}) avec un haut niveau de confiance. L'analyse des images de ces patients montre de nombreux signes radiologiques d'appartenance aux hauts grades de

l’OMS (HGG_{WHO}). Notre classifieur, entraîné avec des descripteurs modélisant l’analyse radiologique, retrouvait de lui-même le biais présent dans l’étiquetage du BraTS.

Ainsi, il était possible de présenter ce chapitre de deux manières : soit d’une manière chronologique, tel que décrit dans cet avant-propos ; soit en étudiant comment des critères d’imagerie cohérents avec l’analyse radiologique nous amènent à souligner une incohérence de vérité terrain. Pour des raisons rédactionnelles, nous avons choisi cette deuxième solution. Le récit chronologique, bien que séduisant par son côté narratif, devient rapidement confus dès lors que l’on change la signification des différents sigles liés à la base BraTS en cours de chapitre.

C’est ainsi que dans ce chapitre, nous proposons d’implémenter un classifieur d’apprentissage automatique classique intelligible et d’étudier ses performances à travers un outil novateur : l’histogramme des scores de prédiction. Dans un premier temps, nous allons détailler les choix de notre implémentation suite à la revue de littérature proposée lors du chapitre précédent. Afin d’établir un point de référence des performances avec la littérature, nous entraînerons notre classifieur sur la base de données BraTS, qui est la base de données la plus utilisée pour la classification binaire du grade.

Grâce à l’histogramme des scores de prédiction, nous montrerons que cette base de données n’est pas adaptée pour une tâche de classification binaire haut grade contre bas grade, tels que définis par l’OMS. Nous montrerons que la base de données BraTS n’est pas divisée en grades I, II d’une part et III, IV d’autre part. Elle sépare d’un côté les glioblastomes multiformes (grades IV) et regroupe de l’autre les gliomes "de grades inférieurs [au grade IV]" sous l’appellation *Lower grades gliomas*. Dans ce chapitre, nous exprimerons cette distinction par les sigles LGG_{WHO} et HGG_{WHO} pour parler de la classification de l’OMS binarisée, et LGG_{BraTS} et HGG_{BraTS} pour la séparation opérée dans la base de données BraTS.

Nous proposons ensuite un alignement des étiquettes de cette base de données sur le système de l’OMS binarisé. Nous feront appel pour cela à cinq radiologues experts, auxquels nous soumettons à classification, via une tâche radiologique, les images de la

base BraTS. Par l'application d'un vote majoritaire, nous proposons une vérité terrain radiologique pour la base BraTS, alignée sur l'OMS, dans la limite de l'analyse d'imagerie anatomique.

Enfin, nous entraînons à nouveau notre classifieur sur cette nouvelle vérité terrain, et montrons comment ce ré-étiquetage impacte ses performances. Ce résultat nous servira d'étalon des performances avant d'intégrer, lors du chapitre suivant, les données multimodales du CHU de Poitiers.

6.2 Introduction

Comme expliqué section [2.3](#), le développement lent ou asymptotique des LGG_{WHO} provoque un sous-diagnostic de ces derniers, comparés aux HGG_{WHO} . Cette absence de diagnostic crée un déséquilibre de classe : les HGG_{WHO} sont sur-représentés dans les bases de données. C'est pourquoi l'acquisition d'une base de données larges et diverse est un défi qui se pose actuellement.

La technique de référence pour évaluer le grade des gliomes est l'analyse histologique et moléculaire depuis une biopsie ou une pièce chirurgicale (voir section [2.4](#)). Il s'agit d'une technique invasive qui permet de classer les tumeurs en fonction de critères cellulaires, comme le degré de différenciation. Depuis 2016, les informations histologiques sont combinées avec l'analyse de différents facteurs génétiques, pour identifier des mutations telles que la mutation IDH et la codéletion 1p/19q (voir section [2.4](#)).

Des méthodes basées sur l'imagerie par résonance magnétique (IRM) se développent pour proposer une alternative non invasive et robuste. On l'a vu au chapitre précédent, l'apprentissage automatique est utilisé sur les données IRM pour discriminer LGG_{WHO} et HGG_{WHO} , à l'aide de classifieurs simples comme les SVM [\[23\]](#) ou les forêts aléatoires [\[21\]](#). L'apprentissage profond avec les réseaux neuronaux convolutifs (CNN) a également été appliqué sur des données anatomiques pour la classification des grades [\[121\]](#). Bien que la classification des quatre grades OMS ait été testée [\[6\]](#), la plupart des publications se concentrent sur une classification binaire des grades [\[32\]](#) (voir section [5.5](#)). Ces approches reposent sur la modélisation de l'analyse radiologique telle qu'effectuée

par le radiologue. Ainsi, ces algorithmes extraient des critères radiomiques représentatifs de la classification tumorale comme la forme et l'hétérogénéité de la tumeur, la longueur du premier et du deuxième axe, la présence de nécrose, ou encore le rehaussement de contraste visible dans les séquences T1 post injection de gadolinium (T1c). Ces critères, utilisés par les radiologues [115], peuvent être traduits en descripteurs numériques, puis injectés un système d'apprentissage automatique.

Après un retour sur l'état actuel des connaissances en matière d'analyse radiomique et de classification binaire du grade des gliomes par apprentissage automatique, nous entraînerons un classifieur SVM avec des descripteurs d'imagerie pour produire une classification binaire des gliomes. Partant de ce classifieurs, nous utiliserons les scores de prédiction pour analyser la base de données BraTS et identifier, parmi ses *Lower grade gliomas*, des patients radiologiquement HGG_{WHO}. Suite à quoi, nous avons proposé à 5 experts radiologues de participer à une tâche radiologique permettant d'identifier les tumeurs LGG_{WHO} et HGG_{WHO} de la base BraTS. Par vote majoritaire, nous proposons une vérité terrain qui aligne les grades de la base BraTS sur la classification de l'OMS. Nous évaluons également les performances de notre classifieur sur cette nouvelle vérité terrain. Une discussion sur les résultats est donnée en conclusion.

6.3 Cadre du travail

6.3.1 Analyse radiomique de données IRM anatomiques

Bien que l'imagerie de perfusion et la spectroscopie RMN fournissent des informations importantes pour la classification du grade des gliomes et sont nécessaires pour compléter le diagnostic radiologique [35], nous savons que l'imagerie anatomique donne accès à certaines caractéristiques utilisées dans la classification des gliomes. Par exemple, des critères comme l'effet de masse, la formation de kystes ou la présence de nécrose sont des indices de malignité [30].

Pour l'identification des lésions et de l'œdème, la séquence FLAIR est une séquence de référence. Elle permet d'annuler le signal du liquide céphalo-rachidien. Les lésions et

œdèmes y apparaissent par un hypersignal (voir section [3.8.1](#)). Les séquences à réhaussement de contraste comme le T1c montrent la rupture de la barrière hémato-encéphalique, qui est un marqueur de haut grade, même si jusqu'à un tiers des gliomes de haut grade ne présentent pas ce signal de réhaussement [\[115\]](#).

Après la publication de référence de l'OMS 2016, certaines études ont établi un lien entre le phénotype visible sur l'IRM anatomique et le génotype sous-jacent. Ces critères permettent d'enrichir la base de descripteurs utilisables sur l'IRM anatomique. Citons par exemple qu'un bord de tumeur net est plus souvent présent lors d'une codéletion 1p/19q [\[111\]](#), ou qu'un réhaussement en couronne apparaît parfois sur le T2 FLAIR lors d'une mutation IDH [\[91\]](#). Ainsi, ces caractères visuels de la lésion T2 FLAIR semblent fournir une information sur les marqueurs génétiques du gliome [\[47, 57\]](#). L'analyse de descripteurs de texture peut être utilisée dans un schéma d'apprentissage automatique pour une discrimination de grade binaire [\[109\]](#). Par conséquent, en combinant ces éléments macroscopiques et ces indices génotypiques, l'IRM anatomique permet de se rapprocher de la classification des gliomes de l'OMS.

6.3.2 Classification du grade par apprentissage automatique

Dans notre revue de littérature du chapitre précédent, nous avons vu que les implémentations d'apprentissage automatique profond comme le VGG-16 ou VGG-19 donnent des résultats intéressants, avec d'une précision allant jusqu'à 95% [\[13\]](#). Mais ces structures souffrent d'un effet boîte noire qui les rend peu intelligibles. C'est pourquoi nous allons nous tourner vers les classifieurs d'apprentissage automatique classiques, comme les SVM ou les forêts aléatoires (RF). On a vu dans le Tableau [5.7](#) que les SVM étaient plus couramment utilisés que les RF, notamment sur les données d'IRM multimodales, où ces derniers donnent le meilleur résultat de notre revue de littérature.

La base de données la plus utilisée pour la classification binaire du grade des gliomes est la base BraTS [\[75, 9, 11\]](#). Cette base de données, publiée à l'origine pour une tâche de segmentation, a été largement utilisée pour la classification du grade des gliomes. Dans sa version 2018, l'ensemble de données BraTS propose 285 gliomes, ce qui en fait l'une

des bases les plus riches disponibles.

Dans sa version 2018, les 285 gliomes du BraTS sont divisés en 210 glioblastomes multiformes (HGG_{BraTS} , grade IV) et 75 "gliomes de grade inférieur" (LGG_{BraTS} , grades I, II et III). En anglais, "gliomes bas grade" et "gliome de grade inférieur" partagent le même acronyme (*Low grade glioma* et *Lower Grade Glioma*). Nous avons montré que cette confusion amène de nombreux auteurs travaillant sur la tâche de classification à entraîner des classifieurs binaires sur la division BraTS, tout en présentant leur travail comme basé sur la discrimination fournie par l’OMS. Il est donc nécessaire de clarifier cette distinction, en identifiant les cas HGG_{WHO} parmi les LGG_{BraTS} . Les étiquettes des grades BraTS (LGG_{BraTS} et HGG_{BraTS}) ont été établies par dépistage histologique. Dans la base BraTS, pour chaque patient, 4 séquences anatomiques sont disponibles : T1, T1c, T2 et T2 FLAIR, ainsi qu’une carte de segmentation tumorale établie par 5 experts radiologues. Ces séquences sont recalées entre elles, et l’encéphale y a été extrait.

6.4 Classification binaire et analyse des résultats

Dans cette partie nous allons étudier la cohérence des étiquettes de grade BraTS avec des descripteurs radiologiques de discrimination binaire. Grâce à un classifieur SVM, nous allons souligner la pertinence de ces critères pour une classification binaire alignée sur l’OMS. Cette analyse se fera grâce à un outil d’évaluation des performances du classifieur : les scores de prédiction dont nous allons afficher l’histogramme pour la population des patients LGG_{BraTS} .

6.4.1 Données utilisées

Comme expliqué précédemment (section [6.1](#)), les étiquettes BraTS ne sont pas divisées en LGG et HGG comme définies par l’OMS. Le sigle LGG_{BraTS} fait référence aux *Lower grade glioma*, qui regroupent les grades I, II et III. Néanmoins, nous avons entraîné un SVM sur la division BraTS, car elle contient déjà partiellement la division finale alignée sur l’OMS que nous souhaitons obtenir. En effet, contrairement à la popu-

lation LGG_{BraTS} , chaque patient du groupe HGG_{BraTS} est très fortement représentatif du groupe HGG_{WHO} de l’OMS. Par conséquent, nous proposons d’entraîner un classifieur et d’analyser ses performances en matière de score de prédiction sur la population LGG_{BraTS} , afin d’analyser comment les patients HGG_{WHO} présents dans la catégorie LGG_{BraTS} ont été triés. Nous utilisons la base de données BraTS de 2018.

6.4.2 Descripteurs

En utilisant la carte de segmentation donnée avec les données BraTS, nous avons calculé 51 descripteurs à l’aide de la bibliothèque PyRadiomics [118] pour chaque patient : 7 descripteurs de forme, 6 descripteurs d’intensité basés sur l’histogramme et 5 descripteurs de texture. Cette bibliothèque propose des descripteurs définis par le standard *Imaging Biomarker Standardization Initiative* (IBSI) [126]. Ces descripteurs sont calculés depuis une région d’intérêt (ROI pour *Region of interest*) correspondant à la segmentation tumorale fournie dans les données BraTS. Tous les descripteurs utilisés sont détaillés dans les Tableaux 6.2, 6.3 et 6.4. Nous pouvons revenir sur le choix des descripteurs et décrire l’effet recherché de par leur implémentation.

Les descripteurs de forme sont décrits dans le Tableaux 6.2. Ces descripteurs sont calculés soit depuis un maillage triangulaire de la ROI tumorale, soit depuis l’ellipsoïde englobant la tumeur. Un exemple d’ellipsoïde est représenté Figure 6.1. Cette ellipsoïde est définie par trois demi-axes. Ces demi-axes sont orientés selon trois vecteurs propres \vec{e}_1 , \vec{e}_2 et \vec{e}_3 , eux-mêmes rattachés à trois valeurs propres λ_1 , λ_2 et λ_3 , avec $\lambda_1 > \lambda_2 > \lambda_3$. Ces valeurs propres servent à calculer le grand axe, le petit axe, la planéité (*flatness*) et l’allongement (*elongation*). Nous ajoutons à cette description le diamètre 3D maximal, qui est calculé par la distance euclidienne (norme L2) des deux voxels les plus distants de ROI tumorale. Deux autres descripteurs, l’aire de surface et la sphéricité sont calculés depuis le maillage triangulaire de la ROI. La sphéricité nécessite un calcul de volume du maillage triangulaire : celui-ci s’effectue par une modélisation à base de tétraèdres. Avec ces descripteurs, nous cherchons à mesurer l’étendue de la tumeur au sein du cerveau.

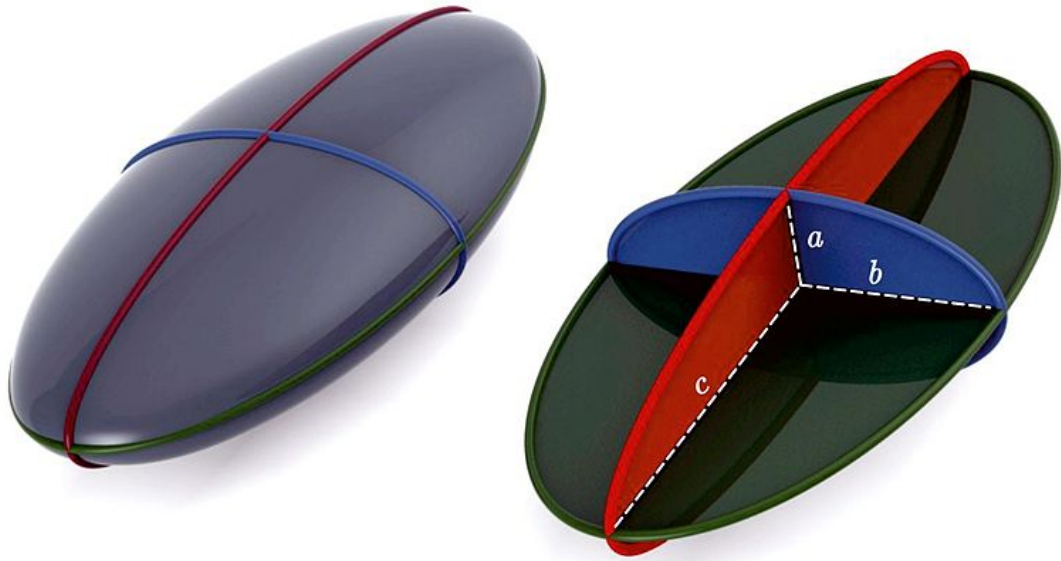


FIGURE 6.1 : Représentation d'une ellipsoïde en 3 dimensions. L'ellipsoïde est définie par trois demi-axes a , b et c .

Image par JoshDif, publiée sous licence *Creative Commons Attribution-Share Alike 3.0*.

Les descripteurs d'intensité sont décrits dans le Tableau [6.3](#). Ils sont calculés depuis l'histogramme de la ROI formée par la segmentation tumorale. Plusieurs mesures simples permettent de caractériser une distribution : la valeur moyenne, l'asymétrie (*skewness*) et l'aplatissement de l'histogramme (*kurtosis*). Il est facile de se figurer en quoi de tels descripteurs peuvent être pertinents. Par exemple, dans le cas d'un T2 FLAIR, on peut imaginer que ces mesures permettent de faire la différence entre une lésion très diffuse, présentant un gradient de niveaux de gris ; et une lésion plus uniforme, très visible. Le descripteur de contraste est calculé comme une variation d'intensité locale à partir de la matrice de co-occurrence des niveaux de gris (GLCM). Il permet de mettre en évidence de zones hétérogènes comme le sont la prise de de contraste et la nécrose d'une séquence T1 post-gadolinium. Enfin, des descripteurs comme l'énergie et l'entropie donnent une information sur la diversité des niveaux de gris présents dans la zone d'étude, ce qui permet à nouveau de pointer l'homogénéité ou l'hétérogénéité de la région d'intérêt.

La caractérisation locale de l'hétérogénéité ou de l'homogénéité se traduit efficace-

ment avec les descripteurs de texture, détaillés dans le Tableau 6.4. Ces descripteurs permettent d'étudier les variations locales dans la région d'intérêt. Ainsi, la GLCM répertorie le nombre d'occurrence des différents couples d'intensité de voxels voisins. La mesure de la corrélation de la GLCM exprime donc la dépendance linéaire qui peut exister entre un niveau de gris et ses voisins. La rugosité (*coarseness*) exprime la différence moyenne d'intensité entre un voxel et son entourage, et offre une mesure plus globale des variations en intensité. Une faible valeur de rugosité caractérise une zone uniforme. L'homogénéité locale se mesure également par le moment de différence inverse et la mesure de complexité. Enfin, la force est une mesure des primitives de l'image. Elle retourne une grande valeur pour une zone ayant une lente variation en intensité.

Nous disposons ainsi de 7 descripteurs de forme, 6 d'intensité et 5 de texture, calculés sur la ROI tumorale. Notons que les descripteurs de forme ne sont calculés qu'une seule fois par patient, leur valeurs dépendant uniquement de la ROI de segmentation tumorale 3D fournie. En revanche, les descripteurs d'intensité et de texture doivent être calculés pour chaque séquence disponible. Avec les 4 séquences anatomiques de la base de données BraTS, cela nous fait un total de 51 descripteurs pour les 285 patients.

6.4.3 Résultats et cohérence des critères radiologiques

Nous souhaitons entraîner un classifieur simple et intelligible afin d'analyser comment ce dernier se comporte vis-à-vis des patients LGG_{BraTS} . Nous avons entraîné et optimisé un classifieur SVM linéaire. Le SVM est le type de classifieur le plus présent dans notre revue de littérature et semble offrir de bons résultats une fois entraîné sur des données IRM multimodales (voir section 5.5), ce qui nous sera utile lorsque nous utiliserons les données du CHU de Poitiers.

Descripteur	Formule	Détails
Grand axe	$4\sqrt{\lambda_1}$	λ_1 : la plus grande valeur propre de l'ellipsoïde englobant la ROI
Petit axe	$4\sqrt{\lambda_2}$	λ_2 : 2ème valeur propre de l'ellipsoïde
Planéité	$\sqrt{\frac{\lambda_3}{\lambda_1}}$	λ_3 : 3ème valeur propre de l'ellipsoïde
Allongement	$\sqrt{\frac{\lambda_1}{\lambda_2}}$	/
Diamètre 3D maximal	$\ (\vec{v}_i, \vec{v}_j)\ _2$	v_i et v_j : coordonnées des deux voxels les plus éloignés spatialement $\ (\vec{v}_i, \vec{v}_j)\ _2$: la norme L2, distance euclidienne
Aire de surface	$A_i = \frac{ a_i b_i \times a_i c_i }{2}$ $A = \sum_{i=1}^{N_f} A_i$	$a_i b_i$ et $a_i c_i$: les bords du i ème triangle du maillage triangulaire, formé des vertex a_i, b_i et c_i A_i : l'aire d'un triangle du maillage triangulaire de la ROI N_f : Le nombre de triangles du maillage
Sphéricité	$\sqrt[3]{\frac{36\pi V^2}{A}}$ Avec : $V = \sum_{i=1}^{V_f} V_i$	A : l'aire de surface V : le volume de la ROI V_i : le volume d'un tétraèdre du maillage triangulaire

TABLE 6.2 : Descripteurs de forme utilisés.

Descripteur	Formule	Détails
Moyenne	$\frac{1}{N_p} \sum_{i=1}^{N_p} X(i)$	X : l'ensemble des N_p voxels de la ROI
Asymétrie	$\frac{\frac{1}{N_p} \sum_{i=1}^{N_p} (X(i) - \bar{X})^3}{\left(\sqrt{\frac{1}{N_p} \sum_{i=1}^{N_p} (X(i) - \bar{X})^2}\right)^3}$	\bar{X} : la valeur moyenne de X
Aplatissement	$\frac{\frac{1}{N_p} \sum_{i=1}^{N_p} (X(i) - \bar{X})^4}{\left(\frac{1}{N_p} \sum_{i=1}^{N_p} (X(i) - \bar{X})^2\right)^2}$	/
Contraste	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i - j)^2 p(i, j)$	N_g : le nombre de niveaux de gris de l'image $p(i, j)$ la GLCM normalisée à i lignes et j colonnes
Énergie	$\sum_{i=1}^{N_p} X(i)^2$	/
Entropie	$-\sum_{i=1}^{N_g} p(i) \log_2(p(i))$	$p(i)$: l'histogramme d'intensité à N_g divisions

TABLE 6.3 : *Descripteurs d'intensité utilisés.*

Descripteur	Formule	Détails
Corrélation de la GLCM	$\frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j)ij - \mu_X \mu_Y}{\sigma_X(i)\sigma_Y(j)}$	<p>$p(i, j)$ la GLCM normalisée à i lignes et j colonnes</p> <p>X et Y les probabilités marginales de $P(i, j)$ en ligne et en colonne</p> <p>μ_X, μ_Y : la moyenne des séries X et Y</p> <p>σ_X, σ_Y : l'écart-type de X et Y</p>
Moment de différence inverse	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{p(i, j)}{1 - i - j ^2}$	N_g le nombre de niveaux de gris
Rugosité	$\frac{1}{\sum_{i=1}^{N_g} p_i s_i}$ <p>avec $s_i = \sum_{i=1}^{n_i} i - \bar{A}$</p>	<p>p_i la probabilité d'occurrence du niveau de gris i</p> <p>\bar{A} l'intensité moyenne</p> <p>n_i le nombre de voxels d'intensité i</p>
Complexité	$\frac{1}{N_v} \sum_{i=1}^{n_i} \sum_{j=1}^{n_i} i - j \frac{p_i s_i + p_j s_j}{p_i + p_j}$	N_v le nombre de voxels du voisinage
Force	$\frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (p_i + p_j)(i - j)^2}{\sum_{i=1}^{N_g} s_i}$	/

TABLE 6.4 : *Descripteurs de texture utilisés.*

Nous avons tenu compte du déséquilibre de classe en corrigeant ce dernier durant l'apprentissage par des pondérations automatiques fonctions du nombre d'individus présents dans la classe via la formule [6.1](#) :

$$w(c) = \frac{N_{population}}{N_{classe} \times N_{individu}(c)} \quad (6.1)$$

Avec : $w(c)$ la pondération pour la classe c

$N_{population}$ le nombre d'individus dans la population

N_{classe} le nombre de classes dans la vérité terrain

$N_{individu}(c)$ le nombre d'individus appartenant à la classe c

Pour éviter le sur-apprentissage, nous avons utilisé une validation croisée à 5 divisions. Les performances du classifieur sont présentées dans le Tableau [6.5](#). Nous obtenons une précision de 84,1%, une sensibilité de 87,0% et une spécificité de 75,9% sur l'ensemble de données BraTS. Ces premières performances se classent dans la fourchette basse de l'état de l'art, tout en restant comparables avec les travaux présentés dans la section [5.5](#). Avec ce classifieur, nous pouvons explorer les résultats de classification de la population LGG_{BraTS} , et analyser comment ont été traités les patients HGG_{WHO} qui en font partie.

Vérité terrain BraTS		
Précision	Sensibilité	Spécificité
84,1%	87,0%	75,9%

TABLE 6.5 : Performances du classifieur sur la vérité terrain du BraTS.

Afin d'analyser le comportement du classifieur, nous proposons d'utiliser le score de prédiction, tel que définit par Platt et al. [93](#). Il s'agit d'un score exprimant, pour chaque patient, la probabilité d'appartenance à une classe donnée, de type $P(\text{classe}|\text{entrée}) \in [0; 1]$.

Détaillons comment est construit ce score de prédiction. Nous avons présenté section [4.3.9](#) l'équation [4.5](#), établissant la fonction de décision $D(x)$ dans le cas d'une classification binaire. La sortie de cette fonction est définie sur \mathfrak{R} . Pour obtenir une probabilité,

il faut donc trouver une transformation T qui, définie sur \mathfrak{R} , donne une sortie s avec $s \in [0; 1]$. Cette sortie doit être monotone, c'est à dire constamment croissante ou décroissante, et refléter qu'une grande valeur de $D(x)$ entraîne un haut score de prédiction.

Pour cela, Platt et al. proposent de transformer le résultat du SVM par une sigmoïde, dont l'équation générale est la suivante :

$$sig(x) = \frac{1}{1 + e^{-x}} \quad (6.2)$$

Combinée avec l'équation [4.5](#), on obtient :

$$T(x) = \frac{1}{1 + e^{A*D(x)+B}} \quad (6.3)$$

Avec : $T(x)$ la transformation recherchée
 $D(x)$ la fonction de décision du SVM
 A et B deux paramètres à estimer

Cette réécriture fait apparaître deux paramètres, A et B , qu'il faut estimer afin que les variations de la sigmoïde $T(x)$ représentent les variations de $D(x)$. La valeur de A influe sur la pente de la sigmoïde au point d'inflexion. La valeur de B influe sur la position de la courbe en x , par translation horizontale du point d'inflexion. Ce point d'inflexion, obtenu pour $T(x) = 0.5$, est l'expression du cas où $D(x) = 0$. Dans notre cas binaire, il s'agit donc de la frontière de discrimination. Dans l'implémentation de Platt et al., les paramètres A et B sont obtenus par estimation d'un maximum de vraisemblance à partir de la base d'apprentissage.

Nous allons interpréter ce score de prédiction comme l'expression d'un taux de confiance produit par le classifieur lors de l'affectation d'un patient à une classe. En effet, ce score exprime si un individu se situe proche de la fonction de séparation $D(x)$ ou non. Dans le cas où l'individu est éloigné de l'hyperplan du SVM, la valeur absolue $|D(x)|$ est élevée, et l'individu génère après $T(x)$ une valeur proche de 0 ou de 1. Dans

		Vérité terrain du BraTS	
		LGG _{BraTS}	HGG _{BraTS}
SVM	LGGsvm	Vrais négatifs	Faux négatifs
	HGGsvm	Faux positifs	Vrais positifs

TABLE 6.6 : *Sigles utilisés dans la matrice de confusion.*

le cas où l'individu est proche de la fonction de décision, $D(x)$ est proche de 0, et $T(x)$ retourne une valeur proche de 0,5. L'algorithme était alors proche de l'erreur. On peut alors lire la réponse du classifieur comme associée à une faible certitude, un faible taux de confiance.

Pour étudier la cohérence de la prédiction, nous pouvons associer le score de prédiction d'un patient et l'apparence des images de son examen IRM. Prenons l'exemple d'un patient HGG_{BraTS} présentant de nombreux critères radiologiques cohérent avec sa classe, comme une prise de contraste T1c marquée ou une zone nécrotique. Si un tel patient est classé par le classifieur (réponse HGG_{SVM}) avec un score de prédiction élevé, alors le classifieur donne une réponse correcte avec un haut taux de confiance. Dans ce cas, le score de prédiction montre une cohérence entre le label attribué par le classifieur, la classe du patient et l'analyse d'image. Le même raisonnement peut être appliqué à un patient LGG_{WHO} associé à une réponse LGG_{SVM} élevée.

Pour notre test, nous définissons les vrais positifs et les vrais négatifs comme indiqué dans le Tableau 6.6. Les vrais positifs sont les patients HGG_{BraTS} classés comme HGG_{SVM} par notre classifieur. Les vrais négatifs sont les patients LGG_{BraTS} classés comme LGG_{SVM}. Nous voulons discriminer les populations LGG_{WHO} et HGG_{WHO} présents parmi les LGG_{BraTS}. Comme certains HGG_{WHO} se retrouvent dans le groupe LGG_{BraTS}, nous devons étudier les patients classés comme vrais négatifs et faux positifs.

Nous avons tracé l'histogramme des scores de prédiction pour tous les patients LGG_{BraTS} (Fig 6.2a). Cet histogramme nous donne des informations sur la qualité de la discrimination du classifieur. En effet, plus un patient est proche de 1, plus le classifieur est confiant dans sa prédiction lorsqu'il affecte un patient à un groupe. Un patient dont

le score de prédiction est proche de 0,5 indique que le classifieur affecte ce patient à une classe de manière proche de l'aléatoire. Nous pouvons analyser la forme de cet histogramme comme une distribution bimodale avec un mode dans la partie des vrais négatifs (LGG_{BraTS} classés comme LGG_{SVM}) et un autre dans la partie des faux positifs en rouge (LGG_{BraTS} classés comme HGG_{SVM}). Pour mieux comprendre le classifieur, nous pouvons analyser les patients les plus à gauche de l'histogramme : ce sont les LGG_{BraTS} qui ont été prédits comme HGG_{SVM} par le classifieur avec un niveau de confiance très élevé.

Ce sont donc les patients pour lesquels le classifieur s'est fortement trompé.

Histogramme des scores de prédiction - Population des LGG BraTS

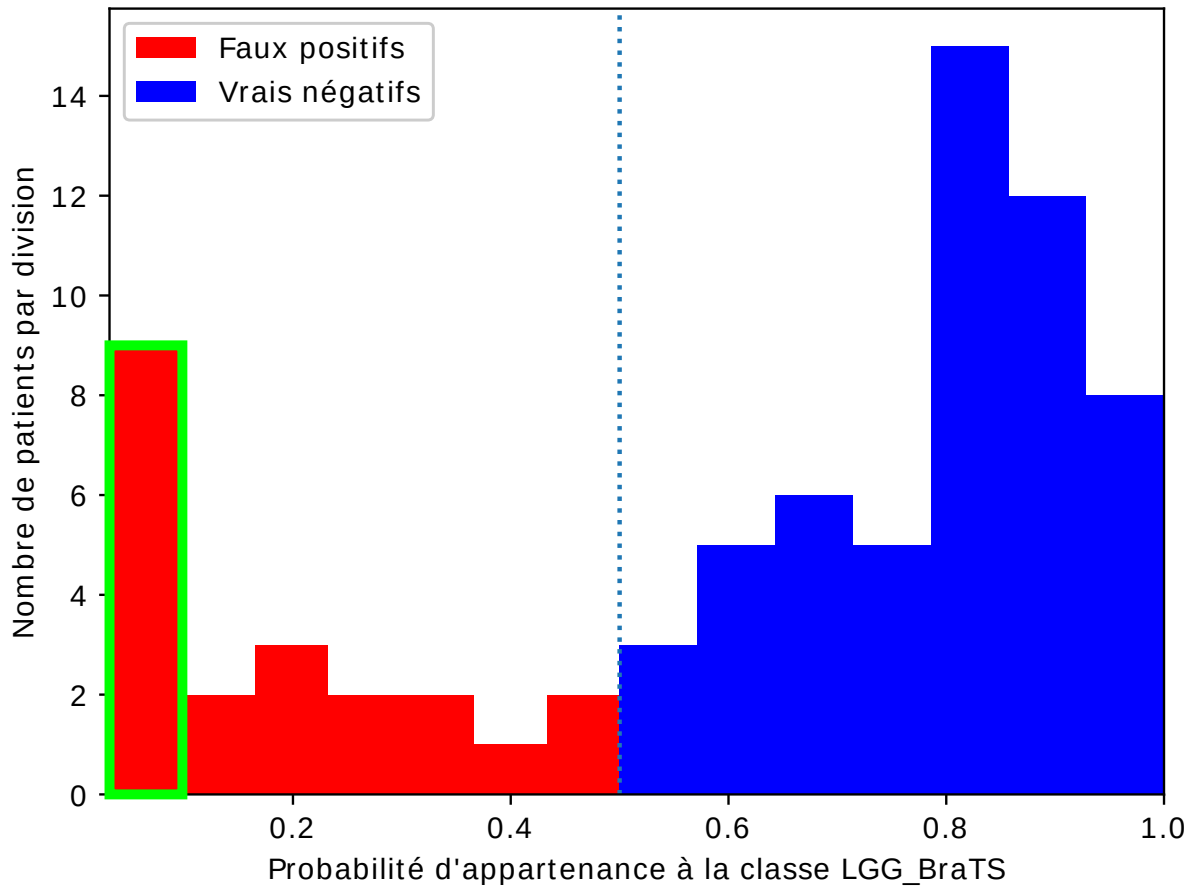


FIGURE 6.2a : *Histogramme des scores de prédiction de la population des LGG_{BraTS}. Les faux positifs, sous le seuil de 0,5% sont représentés en rouge. Les vrais négatifs, dont les scores sont supérieurs à 0,5, sont représentés en bleu. Les faux positifs les plus à gauche de l'histogramme présentent des caractéristiques radiologiques cohérentes avec les patients HGG_{WHO} (voir Figure 6.2b)*

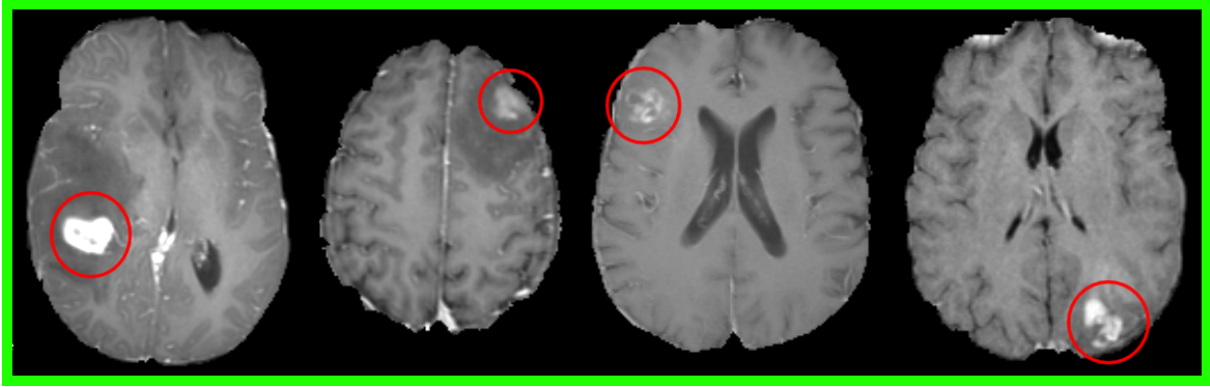


FIGURE 6.2b : 4 des 9 patients les plus à gauche de l’histogramme des scores de prédiction. Ils présentent, dans la zone rouge, des marqueurs radiologiques caractéristiques des HGG_{WHO} , comme la nécrose ou le réhaussement de contraste après injection de gadolinium.

6.4.4 Analyse radiomique des faux positifs et faux négatifs

L’analyse des scores de prédiction de chaque patient permet de savoir à quel point le classifieur était sûr de lui lors de son attribution de classe, et donc à quel point ce dernier s’est aussi trompé. Nous pouvons relier le score de prédiction à l’aspect visuel des images. La plupart des faux positifs montrent de nombreux critères qui les feraient normalement appartenir au groupe HGG_{WHO} . Quatre exemples de faux positifs sont présentés en Figure [6.2b](#). Ces patients présentent un réhaussement de contraste post injection de gadolinium et/ou de la nécrose, qui sont des signes radiologiques de d’appartenance au groupe HGG_{WHO} . Ceci illustre le fait que ce classifieur, entraîné à analyser des images avec des descripteurs issus de la modélisation de l’analyse radiologique pour discriminer entre LGG_{WHO} et HGG_{WHO} , fait apparaître un sous-groupe à l’intérieur de la population LGG_{BraTS} . Ce sous-groupe apparaît dans les faux positifs et correspond radiologiquement aux patients HGG_{WHO} . Ainsi, un observateur extérieur peut identifier les patients radiologiquement HGG_{WHO} parmi les faux positifs du classifieur. Nous pouvons également supposer que le groupe HGG_{BraTS} utilisé lors de l’apprentissage est représentatif de la population HGG_{WHO} , car il est composé exclusivement de glioblastomes

multiformes. C’est pourquoi le classifieur regroupe ensemble les patients de grades les plus élevés. Ainsi, la frontière de classification coupe à travers la population LGG_{BraTS} en séparant d’une part les LGG_{WHO} et les patients présentant les critères visuels des HGG_{WHO} . Ce résultat confirme que les descripteurs utilisés sont bien cohérents avec l’analyse radiologique utilisée lors de la classification de grade selon l’OMS.

6.5 Une nouvelle vérité terrain pour le BraTS

6.5.1 Obtention d’étiquettes radiologiques

Nous voulons réaligner la base BraTS sur la division de l’OMS. C’est pourquoi nous avons créé une tâche d’étiquetage destinée à des radiologues experts, afin de générer une nouvelle vérité terrain. Nous avons développé un module Python pour le logiciel Slicer3D [38][2], permettant de présenter chaque patient dans un ordre aléatoire. Des boutons permettent au radiologue de changer la séquence affichée parmi les 4 séquences disponibles dans la base BraTS, et de donner leur réponse de classification binaire. Nous avons choisi de présenter 100 patients pour que la tâche dure environ une heure. Cela nous a permis de présenter les 75 patients LGG_{BraTS} et 25 patients HGG_{BraTS} en complément. Un nombre impair de 5 experts ont analysé les images et ont choisi d’assigner chaque patient dans la catégorie LGG_{WHO} ou HGG_{WHO} . Chacun d’entre eux a ensuite rempli une liste de critères, nous permettant de savoir quels étaient les descripteurs visuels utilisés par les experts pour cette classification, ce qui nous permet de comparer l’homogénéité des réponses des radiologues. À partir de cette liste, nous pouvons dire que les radiologues experts ont utilisé en priorité la nécrose et l’augmentation du contraste du T1c, puis se sont appuyés sur divers critères tels que l’hémorragie intralésionnelle, l’inhomogénéité FLAIR ou l’effet de masse pour une classification plus fine. Le processus étant randomisé, ni le radiologue ni l’expérimentateur ne connaissaient le grade original du patient présenté dans l’ensemble de données BraTS, ce qui fait de cette tâche un processus en double aveugle. Comme nous avons demandé à un nombre impair de radiologues de participer à cette tâche, nous avons utilisé un vote majoritaire comme règle

de décision pour créer une nouvelle vérité terrain cohérente avec l'analyse radiologique.

Les détails de la répartition des votes sont indiqués dans le Tableau 6.7. Nous pouvons voir que les radiologues sont parvenus à un consensus de 5 votes sur 5 dans 92% des cas lors de l'étiquetage d'un patient HGG_{BraTS} . Seulement 8% des HGG_{BraTS} n'ont obtenu que 4 voix contre 1. Ces chiffres élevés mettent en avant la grande cohérence de la population HGG_{BraTS} , composée uniquement de glioblastomes multiformes. Cette cohérence du groupe HGG_{BraTS} explique également pourquoi aucune étiquette n'a été modifiée par le processus de vote : tous les HGG_{BraTS} ont été affectés à la catégorie HGG_{WHO} après le vote. Parmi les 75 LGG_{BraTS} , seuls 65,3% des patients ont obtenu un consensus complet des experts, qu'il s'agisse d'un consensus pour confirmer le label LGG_{WHO} d'origine ou pour attribuer une nouvelle étiquette HGG_{WHO} au patient. 20% ont obtenu 4 voix sur 5 et 14,6% se sont trouvés dans une situation plus ambiguë, n'obtenant que 3 voix sur 5.

Le vote majoritaire a changé l'étiquette de grade de 44 patients LGG_{BraTS} sur 75. Sur ces 44 changements, nous précisons que 29 patients LGG_{BraTS} ont été classés comme HGG_{WHO} suite à un consensus de 5 voix sur 5 ; 8 suite à un résultat de 4 voix contre 1 ; et 7 suite à un résultat de 3 voix contre 2.

	LGG_{BraTS}			HGG_{BraTS}			Test complet		
Vote majoritaire	5/5	4/5	3/5	5/5	4/5	3/5	5/5	4/5	3/5
% des cas	65,3%	20%	14,7%	92%	8%	0%	72%	17%	11%
Grades changés	44/75			0/25			44/100		

TABLE 6.7 : Analyse des votes des radiologues en vue d'une construction de vérité terrain radiologique.

6.5.2 Évaluation d'un classifieur SVM entraîné sur la vérité terrain radiologique

Cette vérité terrain nous donne de nouveaux groupes de LGG_{WHO} et HGG_{WHO} , présentant une cohérence radiologique, mais marqués par un fort déséquilibre de classe. En effet, après le vote majoritaire des radiologues, notre nouvelle vérité terrain ne compte

plus que 31 patients LGG_{WHO} contre 254 patients HGG_{WHO} . Ainsi, nous voulons savoir s'il est toujours possible d'entraîner un classifieur sur cette base réalignée. C'est pourquoi, à titre de preuve de concept, nous avons entraîné à nouveau un classifieur SVM de la même manière que présenté section 6.4.3, et évalué les performances de ce nouveau classifieur sur cette nouvelle vérité terrain. Malgré ce déséquilibre de classe important, nous observons une précision similaire, une sensibilité légèrement inférieure et une spécificité en baisse. Les résultats détaillés sont présentés dans le Tableau 6.8

Vérité terrain BraTS			Vérité terrain radiologique		
Précision	Sensibilité	Spécificité	Précision	Sensibilité	Spécificité
84,1%	87,0%	75,9%	82,4%	83,9%	70,6%

TABLE 6.8 : Comparaison des performances entre les différentes vérités terrains.

Nous avons à nouveau tracé l'histogramme des scores de prédiction pour chaque patient LGG_{WHO} (Figure 6.3). Nous pouvons voir sur ce nouvel histogramme que nous n'avons que 3 patients dans les faux positifs proches de 0. Dans les faux positifs, le mode apparaît cette fois-ci proche de 0,3, ce qui signifie que la réponse du classifieur est plus incertaine que lorsque le mode est proche de 0,0. Ce déplacement du mode des faux positifs vers la droite de l'histogramme traduit une amélioration dans le comportement du classifieur. De plus, contrairement à la Figure 6.3, lorsque nous regardons les patients présents dans le mode des faux positifs, aucun ne présente de signes radiologiques de patients HGG_{WHO} . Par conséquent, malgré le petit nombre de patients LGG_{WHO} dont nous disposons, nous pouvons dire que la nouvelle vérité terrain permet au classifieur de donner des résultats cohérents sur le plan radiologique.

Enfin, nous pouvons étudier l'effet de la nouvelle vérité terrain sur le classement des hauts grades. Pour permettre la comparaison, les deux histogrammes des scores de prédictions, l'un avec la vérité terrain du BraTS, l'autre avec la vérité terrain radiologique, sont représentés Figure 6.4. On observe que les deux classifieurs identifient les hauts grades avec un haut niveau de confiance, car les divisions modales de chaque histogramme sont très proches de 1,0. Entraîné sur la vérité terrain radiologique, le classifieur

semble davantage sûr de lui : presque tous les patients HGG_{WHO} se retrouvent dans la division la plus à droite.

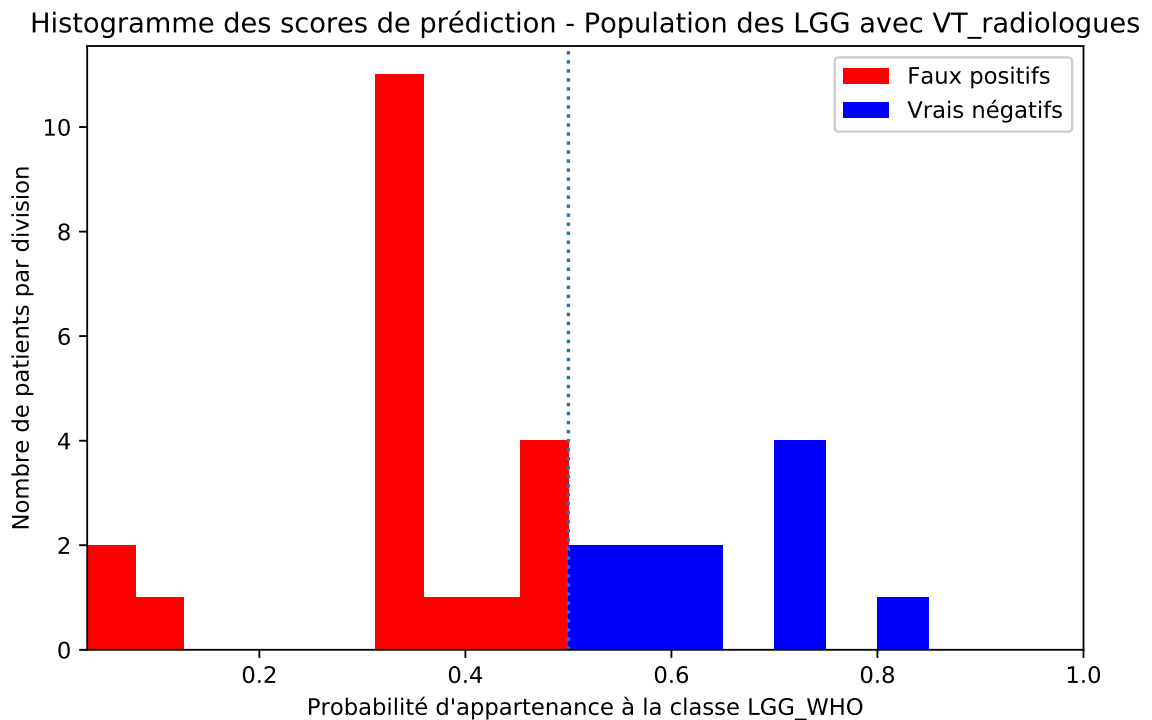


FIGURE 6.3 : *Histogramme des scores de prédiction des patients LGG_{WHO} . Comparé à l'apprentissage sur la vérité terrain du BraTS, la population de patients radiologiquement HGG_{WHO} proche de 0 a disparu. De plus, tous les patients du mode proche de 0,3 sont cohérents avec une appartenance à la classe LGG_{WHO} . Il s'agit donc de vraies erreurs du classifieur.*

Cette observation montre une chose intéressante : même si les performances de classification avec la vérité terrain radiologique semblent légèrement inférieure (voir Tableau 6.8), nous voyons au travers de l'histogramme des scores de prédiction que le classifieur entraîné sur cette vérité terrain classe mieux les gliomes de hauts grades, avec un plus haut taux de confiance. Cet effet échappe à l'analyse lorsque l'on regarde uniquement les valeurs de précision, sensibilité et spécificité, c'est-à-dire les métriques issues de la matrice de confusion. Cet effet nous montre que l'histogramme des scores de prédiction

peut être utilisé pour obtenir une analyse qualitative de la classification. Ainsi, en plus d'être un outil d'analyse de données comme nous l'avons montré Figure [6.2a](#), l'histogramme des scores de prédiction peut aussi être utilisé comme outil de comparaison des performances de deux classifieurs.

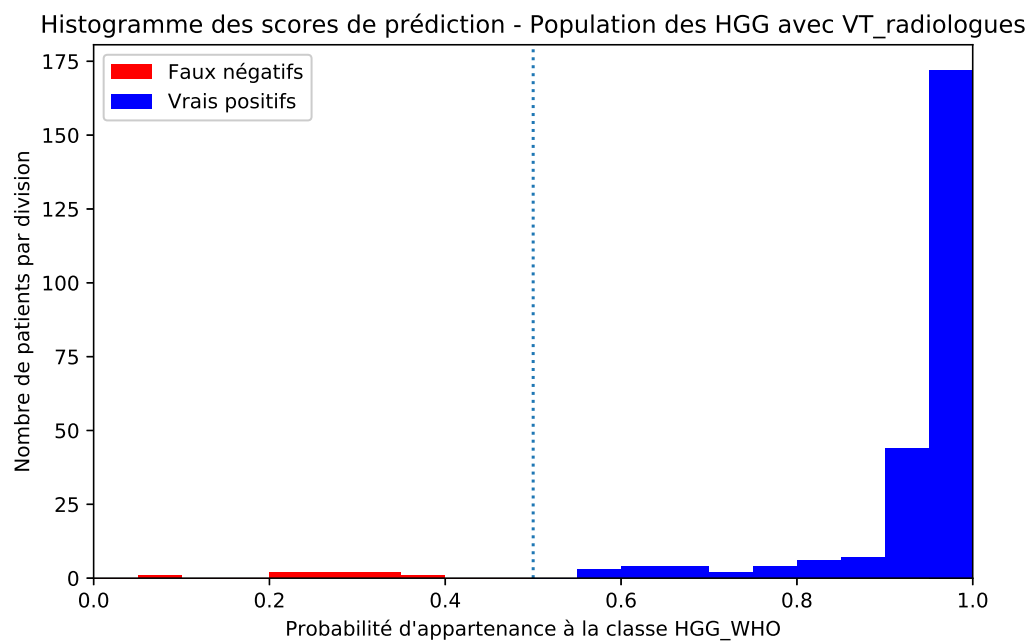
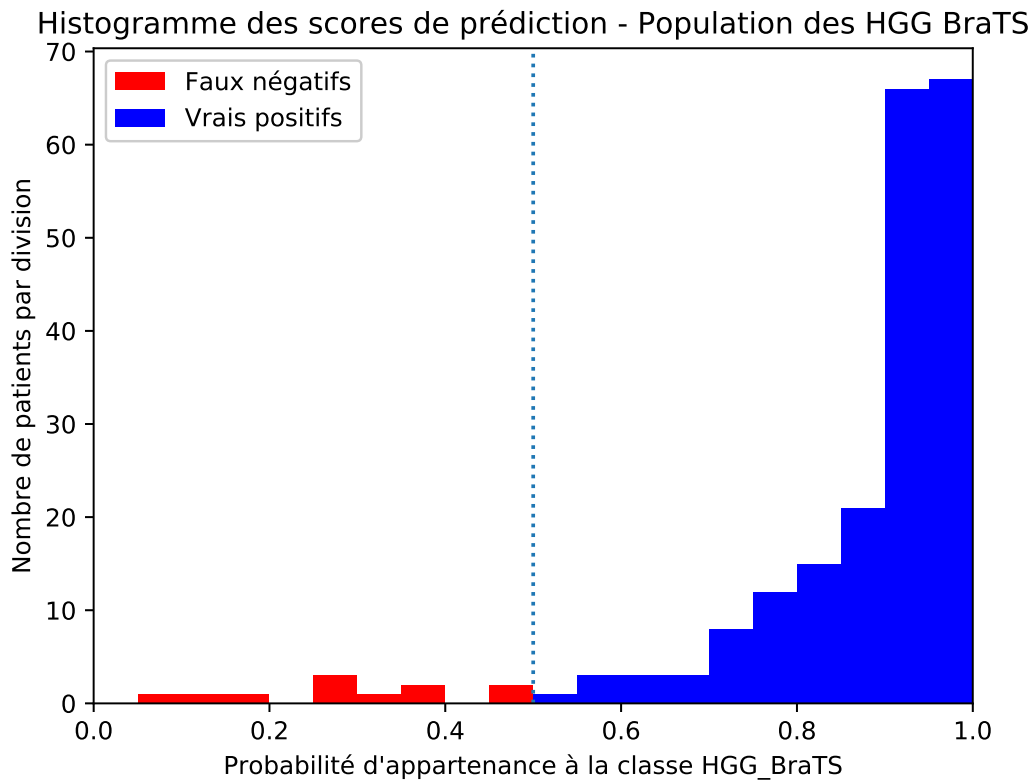


FIGURE 6.4 : (haut) Histogrammes des scores de prédiction obtenus sur la population des HGG_{BraTS} ; (bas) sur la population des HGG_{WHO} .

6.6 Discussion et conclusion

Dans ce chapitre, nous avons montré comment les scores de prédiction d'un classifieur SVM ayant appris depuis la vérité terrain BraTS peuvent être utilisés pour identifier des sous-groupes au sein de l'histogramme des performances du classifieur. L'étude de la position de chaque patient le long de l'axe de prédiction peut donner plus d'informations que l'évaluation habituelle basée sur la précision, la sensibilité et la spécificité. Dans notre exemple, nous avons vu un mode apparaître dans les faux positifs, proche de 0. Ce mode exprimait la grande confiance de notre classifieur à attribuer une étiquette fausse à ces patients.

L'analyse radiologique détaillée montre que les patients faux positifs présentent des critères radiologiques du groupe HGG_{WHO} , tels que la nécrose et la prise de contraste du T1c. Nous avons demandé à 5 radiologues experts d'étiqueter chaque patient, ce qui nous a permis de générer une nouvelle vérité terrain radiologique dont les grades sont cohérents avec l'OMS.

Un nouvel apprentissage automatique sur cette vérité terrain radiologique montre une amélioration de la répartition des faux positifs, plus proches de 0,5, cela avec des performances de précision similaires, malgré un déséquilibre de classe plus important. Ainsi, cette amélioration visible dans l'histogramme des scores de prédiction ne se voit pas dans les mesures habituelles de performance que sont la précision ou la sensibilité.

Cependant, il faut faire attention à la classification des gliomes lorsqu'on se base uniquement sur des séquences anatomiques. L'accès à la réalité biologique, par une biopsie, reste nécessaire à la production du diagnostic. Il est possible également d'augmenter la quantité d'information disponible, notamment grâce à un examen IRM multiparamétrique. Notamment, l'imagerie de perfusion et la spectroscopie RMN donnent au radiologue plus d'informations pour la classification du grade. Par exemple, l'imagerie anatomique a du mal à identifier un patient HGG_{WHO} qui ne présente pas de réhaussement de contraste en T1c. C'est pourquoi la vérité terrain que nous avons produite, limitée par les 4 seules séquences disponibles dans la base de données BraTS, ne peut

s'appuyer que sur un ensemble d'information restreint. Par conséquent, les grades que nous proposons par cette vérité terrain radiologique, bien que basés sur l'imagerie anatomique, ne peuvent que s'approcher du grade défini par l'OMS, sans en avoir la robustesse ou la spécificité liées à un véritable diagnostic.

Ainsi, nous disposons à présent d'un classifieur dont nous pouvons évaluer ses performances sur la base BraTS, ce qui nous fournit une valeur de référence avec la littérature. La prochaine étape de ce travail de thèse est donc d'augmenter la qualité de description avec une base de données multimodale : celle du CHU de Poitiers. Nous pourrions ainsi voir l'impact sur les performances de l'ajout de l'IRM de diffusion, de perfusion, et la SRM. C'est l'objet du travail présenté lors du chapitre suivant.

Le travail présenté dans ce chapitre a fait l'objet d'une publication [33] pour la conférence IEEE IPTA 2020 (*International Conference on Image Processing, Tools and Applications*). Une version détaillée de cet article est en cours de rédaction pour le numéro spécial de la revue Elsevier *Innovation and Research in BioMedical engineering* (IRBM) associé.

Application d'un classifieur sur les données multimodales du
CHU de Poitiers

Contenu du chapitre

7.1 Introduction	159
7.2 Analyse d'imagerie multimodale et bases de données utilisées	161
7.3 État de l'art et performances actuelles des classifieurs	163
7.4 Pipeline d'apprentissage automatique	163
7.5 Classifieur	167
7.6 Analyse des performances	168
7.7 Analyse de la hiérarchisation des descripteurs	171
7.7.1 Analyse du meilleur classifieur	171
7.7.2 Analyse sur 100 entraînements successifs	179

Dans ce chapitre, nous souhaitons étudier comment l’ajout de données IRM multimodales, en plus de la simple imagerie anatomique, impacte les performances d’un classifieur d’apprentissage automatique.

7.1 Introduction

Dans ce chapitre, nous proposons d’entraîner un classifieur sur les données multimodales du CHU de Poitiers. Dans la section [5.2.1](#), nous avons montré que peu de publications récentes utilisaient des données multimodales. Ces jeux de données sont souvent de taille réduite, issus d’un processus d’acquisition laborieux car coûteux en temps machine. De plus, il s’agit de jeux de données propriétaires, dont l’accès est souvent restreint. L’absence de base de données multimodale large et ouverte nous empêche d’avoir un point de référence des performances permettant la comparaison entre les différentes publications.

C’est pourquoi, nous proposons ici d’établir un point de comparaison des performances de classification à l’aide de la base BraTS, de sa vérité terrain d’origine, et de la vérité terrain radiologique établie lors du chapitre précédent. L’établissement de ce point de référence est détaillé sur la Figure [7.1](#). À chaque étape de cette Figure, un classifieur est entraîné et évalué.

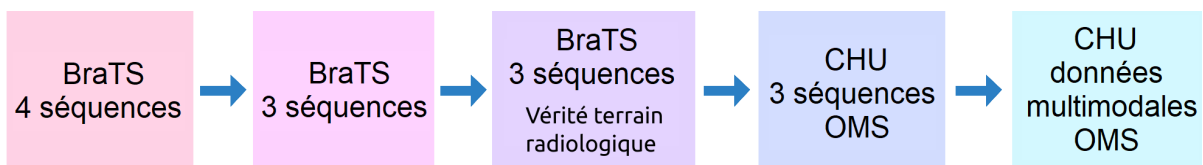


FIGURE 7.1 : Schéma de transformation de la base BraTS permettant d’établir un point de référence pour comparer les performances de classification. À chaque étape, un classifieur est entraîné et évalué. Le CHU de Poitiers ne dispose que de 3 séquences anatomiques (voir section [7.2](#)). L’établissement du point de référence s’effectue donc sur ces 3 séquences, avec la vérité terrain radiologique du BraTS d’un côté, et les données anatomiques du CHU de Poitiers de l’autre. L’ajout des données multimodales permet de montrer l’influence des différentes modalités IRM sur les performances de classification.

Nous utilisons la base de BraTS car il s'agit de la base la plus couramment utilisée. C'est pourquoi elle fournit un point de référence pertinent pour comparer nos résultats avec l'état de l'art concernant la classification binaire sur des données d'IRM anatomique.

L'établissement de ce point de référence doit permettre de comparer les performances de classification sur la base BraTS avec les performances obtenues sur les données du CHU de Poitiers. Nous l'avons vu précédemment, la base BraTS est composée de 4 séquences anatomiques. Le CHU de Poitiers dispose de 3 séquences anatomiques (T1, T1c et T2 FLAIR). C'est pourquoi, le point de référence doit être établi en apprenant sur ces mêmes 3 séquences anatomiques. Nous utiliserons dans un premier temps la vérité terrain d'origine du BraTS, puis la vérité terrain radiologique telle qu'établie lors du chapitre précédent. Le grade des patients du CHU a été établi par analyse histomoléculaire, il est donc aligné avec les standards de l'OMS.

Après avoir évalué les performances sur les données anatomiques du CHU de Poitiers, nous augmenterons la qualité de description à l'aide des différentes modalités IRM disponibles dans le bilan lésionnel du CHU. Nous étudierons ainsi le gain en performances provoqué par l'ajout de séquences de perfusion, de diffusion, et les rapports métaboliques de SRM du proton.

Intégrer les données multimodales du CHU de Poitiers demande de mettre en œuvre une chaîne de traitement complète allant de l'acquisition au classifieur automatique. Nous passerons en revue toutes les étapes de cette chaîne de traitement.

Dans un premier temps nous allons revenir sur les performances actuelles des travaux de classification binaire du grade du gliome et sur les techniques utilisées. Puis nous détaillerons notre chaîne d'acquisition et les différents traitements appliqués avant l'entraînement du classifieur. Nous reviendrons sur les descripteurs utilisés et expliquerons comment ont été gérées les données creuses que nous avons rencontrées. Les performances du classifieur seront alors évaluées sur la base de données du CHU de Poitiers ainsi que sur la base BraTS pour fournir un point de référence des performances. Nous analyserons la hiérarchisation des descripteurs produite par notre classifieur. En discussion, nous verrons comment l'entraînement de classifieur d'apprentissage automatique peut fournir une plus grande compréhension des problèmes abordés.

7.2 Analyse d'imagerie multimodale et bases de données utilisées

Un radiologue peut donner une estimation binaire du grade d'un gliome depuis les données d'imagerie. Un exemple est donné Figure 7.2, où un gliome de haut grade présente de nombreuses caractéristiques comme une prise de contraste en T1c, une nécrose, ou encore une inhomogénéité de la lésion en T2 FLAIR. Les données multimodales viennent enrichir le diagnostic, comme présenté section 3.8. Pour ces données multimodales, nous souhaitons émuler l'analyse du radiologue à travers deux types de descripteurs : les descripteurs d'image présentés section 6.4.2 et des descripteurs issus de mesures manuelles, notamment pour les rapports métaboliques de SRM du proton (voir section 3.8.4) et les valeurs de CBV et CBF relatives controlatérales (voir section 3.8.3), notées rCBF et rCBV. Ces valeurs sont issues de relevés manuels dont le traitement n'est pas automatisé. Il s'agit donc d'une tâche laborieuse dont la question de l'automatisation se pose.

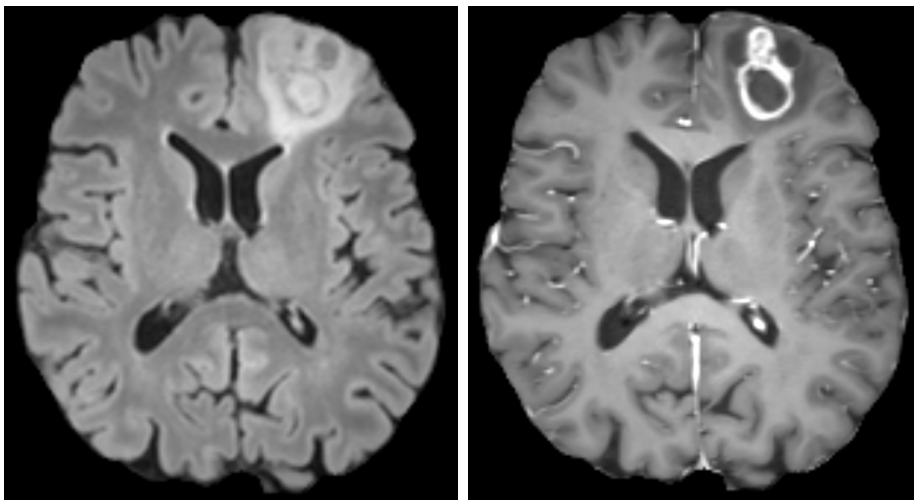


FIGURE 7.2 : Un patient présentant une lésion de haut grade inhomogène en T2 FLAIR (à gauche) et très contrastée en T1c (à droite)

Au CHU de Poitiers, nous disposons de 123 patients répartis en 46 bas grades et 77 hauts grades. Les grades ont été établis par analyse histomoléculaire. Les images ont

été acquises via un imageur Magnetom Skyra 3 Tesla (Siemens Healthineers, Erlangen, Allemagne). Pour chaque patient, nous faisons l'acquisition des modalités et séquences RMN suivantes : imagerie anatomique avec les séquences T1, T1 post injection de gadolinium (T1c), T2 FLAIR ; imagerie du tenseur de diffusion avec les cartes paramétriques ADC, FA, image Trace ; imagerie de perfusion, dont nous tirons les cartographies CBF, CBV, MTT et TTP ; et SRM du proton à TE court (35 ms) et TE long (135 ms). En calculant les mêmes descripteurs images que section [6.4.3](#) sur les séquences anatomiques, de diffusion et de perfusion, nous obtenons 106 descripteurs. S'y ajoutent les relevés manuels rCBV et rCBF, ainsi que les rapports métaboliques suivants : Cho/Cr, Cho/NAA, Lipides/Cr à TE court, Lactate/Cr à TE long et NAA/Cr à TE long. Ainsi, chaque individu se retrouve décrit par 124 descripteurs.

À cause des contraintes d'acquisition, certaines modalités peuvent être manquantes pour certains patients, ce qui crée des données creuses. Nous choisissons de remplir ces valeurs creuses en attribuant, pour un descripteur manquant, la valeur moyenne du descripteur calculée sur le reste de la population où la modalité RMN a pu être acquise.

En plus de cette base multimodale, nous utilisons comme base de référence pour évaluer la performance du classifieur la base de données du challenge BraTS 2018. On rappelle que cette base fournit 4 séquences anatomiques (T1, T1c, T2 et T2 FLAIR) pour 285 patients : 210 glioblastomes multiformes et 75 *Lower Grades Gliomas*. Avec la vérité terrain radiologique établie section [6.5.1](#), le BraTS 2018 se divise en 31 LGG_{WHO} d'une part et 254 HGG_{WHO} d'autre part. Le travail effectué lors du chapitre précédent montre que nous pouvons apprendre sur cette nouvelle division malgré l'important déséquilibre de classe (voir section [6.5.2](#)), et obtenir des résultats exploitables. Cependant, nous ne disposons pas au CHU de Poitiers des informations T2. C'est pourquoi, pour établir un point de référence des performances, nous n'entraînerons le classifieur que sur les trois séquences T1, T1c et T2 FLAIR du BraTS. Calculés sur ces 3 séquences, nous obtenons un ensemble de 40 descripteurs par patient.

7.3 État de l’art et performances actuelles des classifieurs

Dans la section [5.5](#), nous avons détaillé 11 publications récentes portant sur la tâche de classification du grade du gliome. Nous notons que les classifieurs issus de l’apprentissage automatique classique pouvaient fournir des performances comparables aux réseaux d’apprentissage profond. Les métriques d’évaluation utilisées dans ce champ d’étude sont la précision, la sensibilité et la spécificité. En particulier, la précision s’étend selon la méthode de 84% à 95,5%. Comme montré section [6.4.3](#), nous utilisons des méthodes d’apprentissage automatique classique pour disposer d’un classifieur intelligible, là où un réseau d’apprentissage profond souffre parfois d’un effet de boîte noire rendant difficile sa compréhension.

L’extraction de nombreux descripteurs a pour conséquence d’augmenter la dimensionnalité du problème. C’est pourquoi, plusieurs méthodes d’apprentissage automatique permettent d’obtenir une hiérarchisation des descripteurs. C’est notamment le cas des SVM, via l’algorithme *Recursive Feature Elimination* (SVM-RFE) [\[48\]](#). Comme présenté section [4.3.9](#), le SVM-RFE permet d’obtenir un sous-ensemble de descripteurs de haute valeur informationnelle et de faible dimensionnalité. Nous avons vu section [5.3.3](#) que cette implémentation du SVM avait déjà été utilisée pour la classification binaire du grade du gliome. C’est pourquoi nous avons choisi d’utiliser le SVM-RFE au cours de l’implémentation proposée dans ce chapitre.

7.4 Pipeline d’apprentissage automatique

Utiliser des données IRM multimodales demande l’implémentation d’une chaîne de traitement complète. Citons parmi les étapes le recalage multimodal (comprenant transformations rigides et sur-échantillonnage), l’extraction de l’encéphale, la normalisation en intensité ou encore la segmentation tumorale. Les bibliothèques ITK [\[15\]](#), [\[122\]](#), [\[69\]](#) et FSL [\[12\]](#) offrent des outils satisfaisants pour les opérations de recalage, d’extraction

d'encéphale, de sur-échantillonnage et de normalisation en intensité. Pour chaque étape du pipeline, une validation visuelle des résultats a été effectuée à l'aide d'un module python sur Slicer3D.

Passons à présent en revue chaque étape de la chaîne de traitement mise en place pour traiter les données du CHU de Poitiers.

Recalage multimodal : Ce pipeline commence par le recalage multimodal intrapatient tel que figurant dans les travaux de thèse de Mathieu Naudin [82]. Il est question d'aligner géométriquement dans un même espace toutes les acquisitions d'un même patient, comme représenté Figure 7.3. Nous utilisons le recalage rigide d'ITK pour recaler et ré-échantillonner les images sur un volume commun. Ce volume a la même résolution que les acquisitions du BraTS 2018.

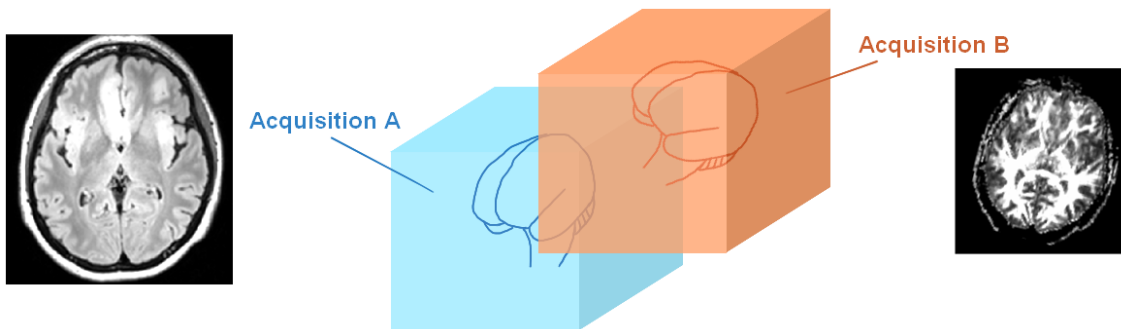


FIGURE 7.3 : *Illustration du principe de recalage entre deux séquences. Le recalage rigide comprend des rotations, translations et homothétie. Un sur-échantillonnage est nécessaire pour mettre les images à une même résolution.*

Extraction d'encéphale : L'extraction d'encéphale est effectuée via l'algorithme BET (pour *Brain Extraction Tool*) [110] de la librairie FSL. Sur les données du CHU de Poitiers, la meilleure extraction que nous obtenons est en appliquant cet algorithme dans sa version BET-Surf sur une séquence de T1c. La segmentation est effectuée sur cette séquence, puis le masque résultant de la segmentation d'encéphale est appliqué à toutes les autres séquences du patient par simple multiplication. Un exemple d'extraction d'encéphale par BET-Surf est donné Figure 7.4.

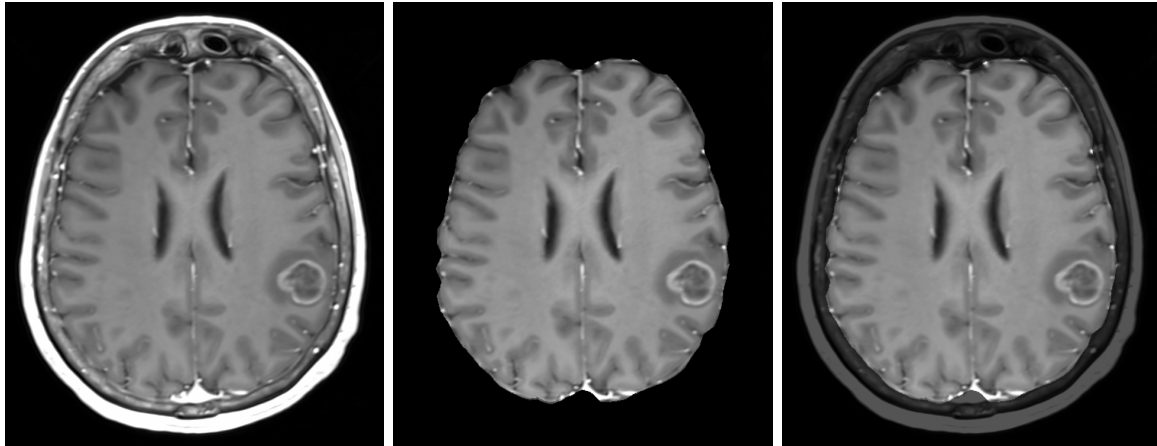


FIGURE 7.4 : Résultats de l'extraction d'encéphale par FSL BET-Surf, avec : (gauche) le T1c de départ, (milieu) le résultat, (droite) une superposition transparente pour validation visuelle.

Normalisation en intensité : Nous utilisons la mise en correspondance d'histogramme de la bibliothèque ITK, telle que présentée section 5.2.3. Nous choisissons l'histogramme de référence à partir du patient présentant la plus grande dynamique en intensité de la base BraTS.

Segmentation tumorale : Nous avons besoin d'une segmentation tumorale afin de calculer les descripteurs, car ceux-ci sont extraits depuis une ROI tumorale. Jusqu'à présent, dans ce travail de thèse, cette ROI était fournie directement avec les données du challenge BraTS. Pour appliquer notre classifieur aux données du CHU de Poitiers, il nous faut donc produire une segmentation. Le challenge de segmentation tumorale BraTS fait office de référence pour la tâche de segmentation tumorale. Les différentes participations y sont évaluées à l'aide du *Dice score* (voir section 4.3.8). De bons résultats de segmentation sont atteints à l'aide de réseaux d'apprentissage profond [124], utilisant en particulier la structure U-net [99]. A titre d'exemple, en 2017, Isensee et al. [54] proposent un U-net pour le challenge BraTS. Leur implémentation se classe parmi les meilleurs résultats de 2017 et dépasse les meilleures performances atteintes par les participants du challenge en 2015, avec un *Dice score* moyen de 0,896.

Mais, l'absence d'acquisition T2 parmi les données du CHU de Poitiers nous empêche d'utiliser les poids de cette solution pré-existante. Par conséquent, nous avons ré-entraîné le U-net proposé par Isensee et al. sur les données BraTS, mais en n'utilisant que les 3 séquences anatomiques dont nous disposons au CHU de Poitiers, c'est à dire T1, T1c et T2 FLAIR. Pour alléger l'occupation mémoire, nous utilisons l'extraction de patches 2D de Fenneteau et al. [39]. Une fois entraîné, ce réseau peut segmenter les données du CHU de Poitiers. Avec une séquence en moins, le *Dice score* moyen que nous obtenons est de 0,804. Les performances sont donc inférieures à celles obtenues par Isensee avec 4 séquences, néanmoins elles offrent un résultat visuel correct, comme montré via la Figure 7.5. Cette segmentation a été validée visuellement par un radiologue.

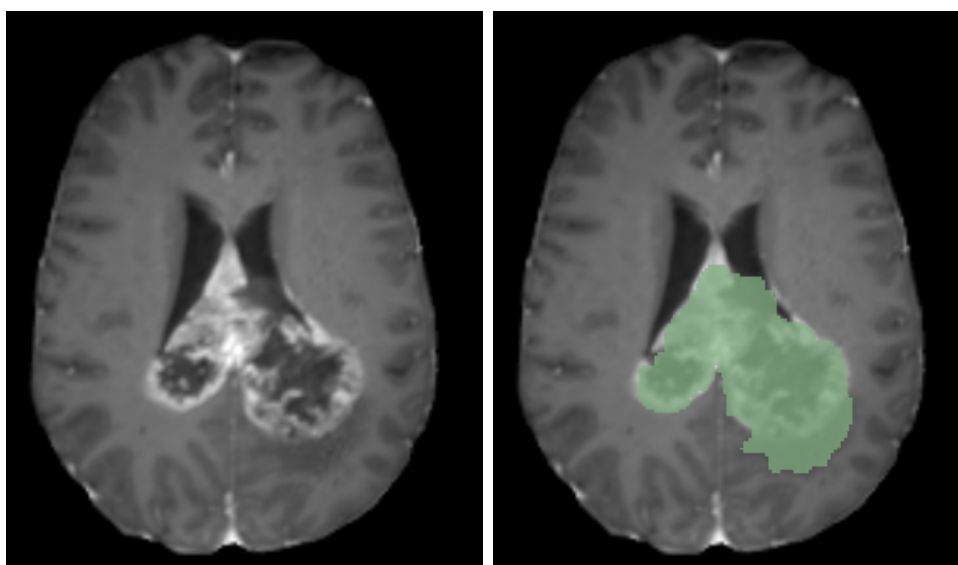


FIGURE 7.5 : *Validation visuelle de la segmentation du U-net à 3 séquences anatomiques. La segmentation est en vert transparent.*

En sortie de pipeline, nous définissons une base de test de 20% des individus pour évaluer le meilleur classifieur en sortie d'apprentissage. La population présente dans la base de test dépend de l'apprentissage effectué (on rappelle les différentes étapes du travail Figure 7.1). Nous n'avons pas mélangé les individus entre la base BraTS et la base de données du CHU. Ainsi, si l'apprentissage a lieu sur la base BraTS, alors la base de test contient 20% de patients BraTS. Pour un apprentissage sur les données du CHU,

la base de test contient 20% de patients du CHU. Dans tous les cas, nous calculons les descripteurs présentés section [6.4.2](#).

7.5 Classifieur

Nous souhaitons utiliser un classifieur intelligible et offrant une hiérarchisation de descripteurs. C'est pourquoi nous effectuons une première passe pour sélectionner les meilleurs descripteurs avec un SVM-RFE. Puis, en utilisant le sous-ensemble produit, nous entraînons un SVM à l'aide d'une recherche quadrillée pour optimiser les hyperparamètres. Le SVM-RFE est entraîné par validation croisée à 5 divisions et fournit de manière exploratoire, comme décrit section [4.3.9](#), un sous-ensemble de descripteurs les plus discriminants. Dans la recherche quadrillée, nous testons deux noyaux (linéaire et à fonction radiale) et 5 valeurs d'un hyperparamètre de régularisation $C \{0, 1; 0, 5; 1, 0; 1, 5; 2\}$.

Nous évaluons les performances du classifieur SVM avec les métriques suivantes : précision, sensibilité et spécificité. Mais nous observons une grande volatilité des résultats d'un entraînement à l'autre, indépendamment des hyperparamètres testés. En effet, le sous-ensemble de descripteurs retourné par le SVM-RFE varie d'un apprentissage à l'autre. Cette variabilité du SVM-RFE trouve peut-être son origine dans le faible nombre d'individus de la base d'apprentissage.

C'est pourquoi, nous proposons de lisser la variabilité de nos résultats en étudiant les performances moyennes (précision, sensibilité et spécificité) issues de 100 apprentissages indépendants. À chaque apprentissage, de nouvelles bases d'apprentissage et de test de la population étudiée sont générées, le SVM-RFE fournit un nouveau sous-ensemble de descripteurs, et le SVM effectue une validation croisée pour évaluer les performances. Nous gardons trace des performances obtenues sur ces 100 apprentissages et pouvons ainsi fournir une valeur moyenne, dont la variabilité a été lissée, ainsi qu'un écart-type qui renseigne sur la variabilité rencontrée.

Nous présentons chaque valeur accompagnée de plus ou moins un écart-type dans le Tableau [7.1](#). Nous y ajoutons également la valeur de précision du meilleur classifieur SVM obtenu sur les 100 apprentissages.

Comme présenté Figure [7.1](#), nous souhaitons établir un point de référence d'évaluation des performances, en entraînant notre algorithme d'une part sur le BraTS 2018, puis sur les données multiparamétriques du CHU. Comme ces dernières ne disposent pas d'une acquisition T2, nous établissons les performances du classifieur SVM sur le BraTS 2018 à 4 séquences (T1, T1c, T2, T2 FLAIR), puis à uniquement 3 séquences (T1, T1c, T2 FLAIR). Ces deux valeurs nous servent de référence vis-à-vis de la littérature d'une part, mais aussi pour entraîner un classifieur multimodal, d'autre part. Enfin, pour chaque test effectué sur le BraTS, nous testons aussi les performances avec la vérité terrain radiologique que nous avons établie précédemment section [6.5.1](#)

7.6 Analyse des performances

Vis-à-vis de la littérature existante, c'est-à-dire entraînée sur BraTS 2018 avec ses quatre séquences anatomiques, la méthode que nous avons implémentée obtient une précision moyenne de $85,8 \pm 4,1\%$. Ce résultat se situe dans la fourchette basse de l'état de l'art, qui atteint fréquemment les 90% de précision. Entraîné sur la vérité terrain radiologique, on retrouve les résultats présentés section [6.5.2](#) : une légère baisse de la précision et de la sensibilité, et une baisse plus importante de la spécificité.

Le passage à trois séquences, avec le retrait de l'acquisition T2, ne perturbe pas les performances de l'algorithme : la précision moyenne reste de $86,0 \pm 4,1\%$, sensibilité et spécificité restent équivalentes aux résultats d'apprentissage sur quatre séquences. Entraîné sur la vérité terrain radiologique, on observe à nouveau les mêmes variations : une baisse légère de la précision et de la sensibilité moyenne, et une baisse plus importante de la spécificité, qui passe de $75,9\% \pm 9,6\%$ à $67,1\% \pm 9,7\%$. Ce sont les valeurs issues de ces deux derniers entraînements, à trois séquences, qui peuvent servir de point de référence de notre algorithme entre les données BraTS et les données du CHU de Poitiers.

Entraîné uniquement avec les trois séquences anatomiques du CHU, le classifieur montre une précision moyenne basse : seulement $77,5 \pm 6,8\%$. La sensibilité moyenne est comparable aux essais précédents avec $88,5 \pm 6,2\%$. La spécificité est également basse, avec seulement $60,2 \pm 14,0\%$. La valeur de précision du meilleur classifieur est aussi peu

élevée avec 81,1% de précision. Nous voulons voir comment évoluent ces performances lors de l'ajout des données multimodales.

Avec l'ajout des données de diffusion, de perfusion, des rapports métaboliques et des mesures de rCBV et rCBF, un gain en précision moyenne semble exister, avec une évolution de $77,5 \pm 6,8\%$ à $78,6 \pm 6,6\%$. Ce gain est faible, mais trouve son expression la plus manifeste dans la précision du meilleur classifieur obtenu sur 100 essais, qui progresse de 81,1% à 86,5%. La valeur de sensibilité décroît légèrement, passant de $88,5 \pm 6,2\%$ à $85,9 \pm 7,9\%$. Le plus fort gain en performance se trouve dans l'évolution de la spécificité moyenne, qui passe de $60,2 \pm 14,0\%$ à $67,3 \pm 15,0\%$. Ainsi, à écart-type constant, la spécificité augmente en valeur moyenne sur 100 entraînements. L'usage des données multiparamétriques semble donc avoir une influence positive sur les performances du classifieur, mais ces gains ne sont pas aussi impressionnants que ce à quoi on pouvait s'attendre selon l'état de l'art, où les SVM entraînés sur données multimodales dépassent les 90% de précision (voir section 5.5). Dans la suite de ce travail, nous souhaitons analyser plus en détail le comportement du meilleur classifieur entraîné sur les données multimodales du CHU de Poitiers.

Vérité terrain d'apprentissage	Précision ¹	Sensibilité ¹	Spécificité ¹	Meilleure précision ²
BraTS 2018 4 séquences	85,8 ± 4,1%	89,5 ± 4,5%	75,4 ± 9,7%	89,5%
BraTS 2018 - radiologique 4 séquences	83,7 ± 5,7%	86,3 ± 6,3%	62,9 ± 17,2%	88,4%
BraTS 2018 3 séquences	86,0 ± 4,2%	89,4 ± 4,5%	75,9 ± 9,6%	87,7%
BraTS 2018 - radiologique 3 séquences	82,1 ± 4,1%	84,0 ± 5,5%	67,1 ± 9,7%	84,2%
Données CHU de Poitiers 3 séquences	77,5 ± 6,8%	88,5 ± 6,2%	60,2 ± 14,0%	81,1%
Données CHU de Poitiers multiparamétriques	78,6 ± 6,6%	85,9 ± 7,9%	67,3% ± 15,0%	86,5%

¹Résultats moyens sur 100 essais

²Meilleur classifieur sur 100 essais

TABLE 7.1 : *Tableau des résultats. Chaque ligne montre le résultat d'un apprentissage (suivant la Figure 7.1). Partant du BraTS à 4 séquences anatomiques entraîné sur sa vérité terrain d'origine, le point de référence pour la comparaison des performances est obtenu en entraînant le BraTS à 3 séquences anatomiques sur la vérité terrain radiologique. Les deux dernières lignes montrent les entraînements effectués sur les données du CHU de Poitiers. Les valeurs de précision, sensibilité et spécificité sont moyennées sur 100 apprentissages indépendants, tel qu'expliqué section 7.5. La dernière colonne indique la meilleure précision obtenue sur ces 100 apprentissages.*

7.7 Analyse de la hiérarchisation des descripteurs

Nous pouvons produire deux niveaux d'analyse de la hiérarchisation des descripteurs. Un premier niveau concerne le meilleur classifieur obtenu à l'issue de la validation croisée. Ce classifieur produit une hiérarchisation efficace pour la tâche de discrimination. Mais nous avons montré que s'intéresser à un classifieur seul peut être sujet à une variabilité des performances. C'est pourquoi, nous proposons un deuxième niveau d'analyse à l'aide d'une hiérarchisation plus globale des descripteurs issue des 100 entraînements indépendants effectués.

7.7.1 Analyse du meilleur classifieur

Commençons par le cas du meilleur classifieur. Le SVM-RFE produit un sous-groupe de descripteurs, ne gardant que ceux dont l'information est la plus pertinente pour le problème abordé. Le SVM-RFE ne produit pas de hiérarchisation à l'intérieur du sous-groupe qu'il retourne : il s'agit uniquement des descripteurs qui n'ont pas été éliminés lors de l'élimination récursive. Tous les descripteurs présents dans le sous-groupe sont donc considérés comme égaux : ils sont tous ex-aequo, en 1ère place, dans la sélection produite. Nous pouvons donc faire une analyse individuelle du meilleur classifieur issu de la validation croisée et de son jeu de descripteurs associé. Suite au SVM-RFE, le meilleur classifieur obtenu emploie un ensemble de 23 descripteurs.

Tel quel, nous ne pouvons pas analyser la hiérarchie interne de ces 23 descripteurs. Mais, suite à la recherche quadrillée, le meilleur classifieur dispose d'un noyau linéaire. Un tel noyau nous offre la possibilité d'approfondir l'analyse de la hiérarchie de descripteurs en regardant les coefficients, les poids, qui paramètrent l'orientation du vecteur normal à l'hyperplan calculé. Cette hiérarchisation correspond aux poids du classifieur linéaire [48]. Ainsi, nous pouvons afficher pour chaque descripteur son importance via son poids associé, comme représenté sur la Figure 7.7. De plus, afin de rendre comparables les valeurs de ce type de représentation pour différents classifieurs, nous avons normalisé les poids en divisant par la norme du vecteur normal à l'hyperplan.

À partir de cet ensemble, nous voyons quels descripteurs et quelles séquences RMN sont les plus pertinentes dans la tâche de classification selon le classifieur SVM. Ici, les 5 meilleurs descripteurs sont : la corrélation de la GLCM du T1c, l'entropie du T1c, la valeur de rCBV, l'asymétrie de l'histogramme du T1 et la complexité de la cartographie CBV. Nous allons analyser ce classement par séquence RMN puis par type de descripteur.

Analyse par séquence : Dans la hiérarchie créée par le meilleur classifieur, nous commençons par noter que près de la moitié des descripteurs de ce classement sont des descripteurs issus de séquences anatomiques, majoritairement calculés sur la séquence T1c (4 descripteurs, contre 3 pour le T1 et 3 pour le T2 FLAIR). Plus précisément, parmi les 5 descripteurs en tête, 3 sont issus des séquences anatomiques : le contraste du T1c, l'entropie du T1c et l'asymétrie de l'histogramme du T1. Le contraste et l'entropie du T1c sont d'ailleurs respectivement en tête de classement.

Pour certains de ces descripteurs, il est aisé de remonter à l'analyse d'imagerie. Par exemple, le contraste du T1c est très élevé pour un gliome de haut grade, présentant de la nécrose sombre et un hypersignal fort dans la zone de prise de contraste, d'où une amplitude en niveaux de gris élevée. Un gliome de bas grade, ne présentant pas de réhaussement ni de nécrose, apparaît sur en T1c comme beaucoup plus homogène avec des niveaux de gris uniformes. C'est dans ce sens qu'on peut également comprendre la mesure d'entropie en tête de classement : l'information est plus diverse, plus riche, dans une lésion de haut grade que dans une lésion de bas grade.

Nous notons ensuite la faible présence de descripteurs issus des séquences de diffusion : seuls quatre descripteurs sont présents, et uniquement dans la deuxième moitié du classement. Deux sont issus de l'anisotropie fractionnelle, en 18 et 19ème positions. Il s'agit de la valeur de la corrélation de la GLCM et de l'aplatissement de l'histogramme (*kurtosis*). Les deux autres descripteurs de diffusion sont issus de la séquence Trace. Il s'agit de l'asymétrie de l'histogramme (13ème position) et l'entropie de la lésion (22ème position). En revanche, tous les descripteurs issus de l'ADC sont absents de ce classement. Or, nous aurions pu nous attendre à ce que certains descripteurs d'ADC soient présents, car dans les zones nécrotiques, présentes uniquement dans les gliomes de hauts

grades, la valeur d'ADC est très élevée (voir Figure [7.6](#)).

Pour les descripteurs issus des séquences de perfusion, nous notons qu'en comptant la valeur de rCBV, 6 descripteurs sur 8 proviennent de la cartographie CBV, ce qui vient soutenir la pertinence de cette acquisition pour l'analyse du radiologue. Parmi ces valeurs issues du CBV, le rCBV, descripteur non issu de l'analyse d'image, se retrouve en tête de classement. Cela souligne la pertinence de cette valeur et de cette séquence dans l'analyse du radiologue. En revanche, on ne retrouve qu'un seul descripteur des cartographies CBF et TTP, et aucun de la cartographie MTT, ce qui interroge sur la pertinence de ces trois séquences pour le diagnostic.

Pour finir, notons également qu'un rapport de spectroscopie est présent : il s'agit de la valeur de Lac/Cr à TE Long.

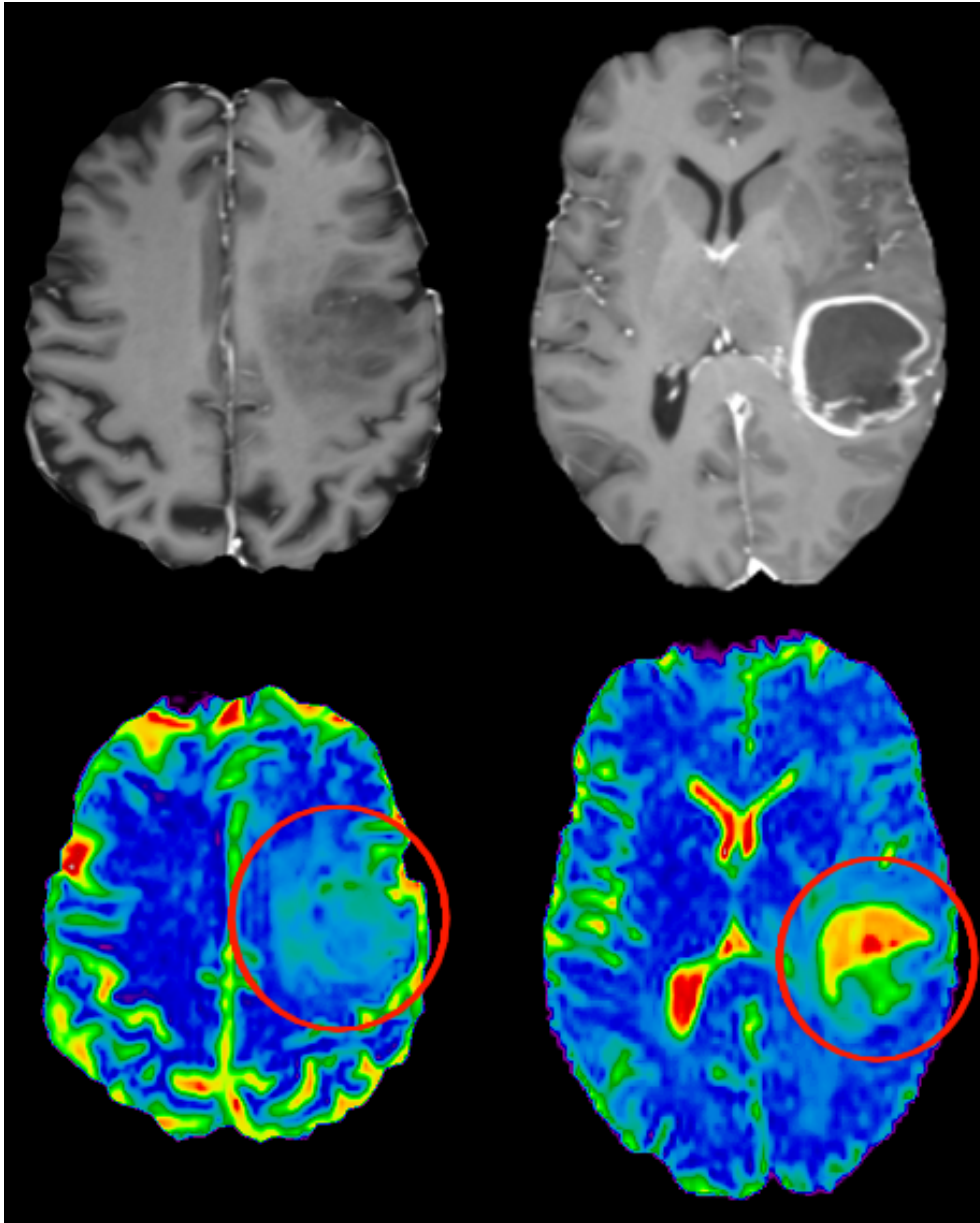
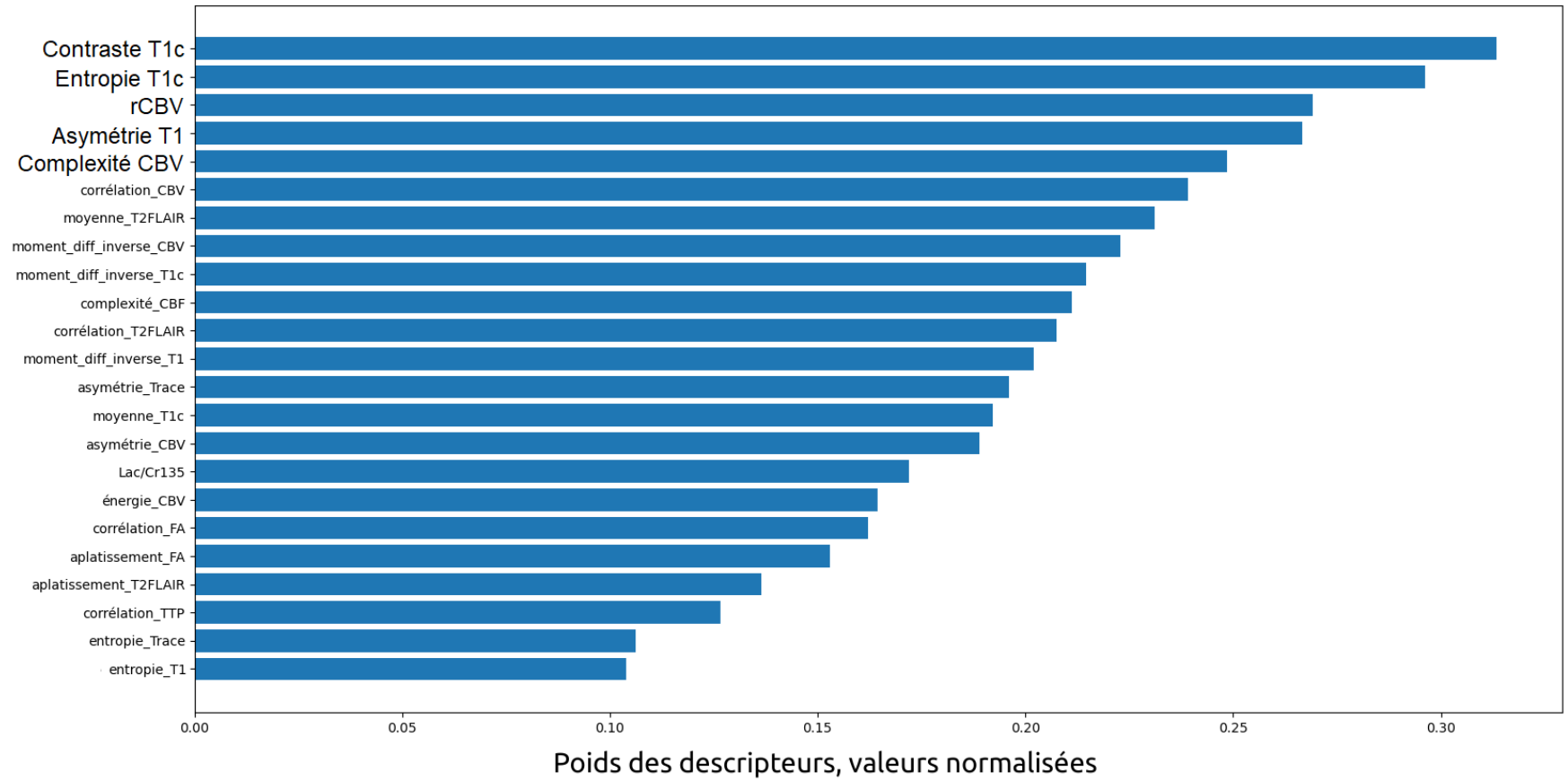


FIGURE 7.6 : *Comparaison des valeurs d'ADC entre (à gauche) un gliome de bas grade et (à droite) un gliome de haut grade. Les acquisitions T1c sont données en haut pour référence. Dans un gliome de bas grade, la valeur d'ADC intra-lésionnelle (zone rouge) est peu élevée. Dans un gliome de haut grade, l'ADC montre une très forte diffusion, notamment dans une zone nécrotique.*

Hierarchisation des descripteurs, classifieur entraîné sur données multimodales



175

FIGURE 7.7 : Hierarchisation des descripteurs produite par le meilleur classifieur.

Hiérarchisation des descripteurs, classifieur entraîné sur données multimodales Analyse par type de séquence

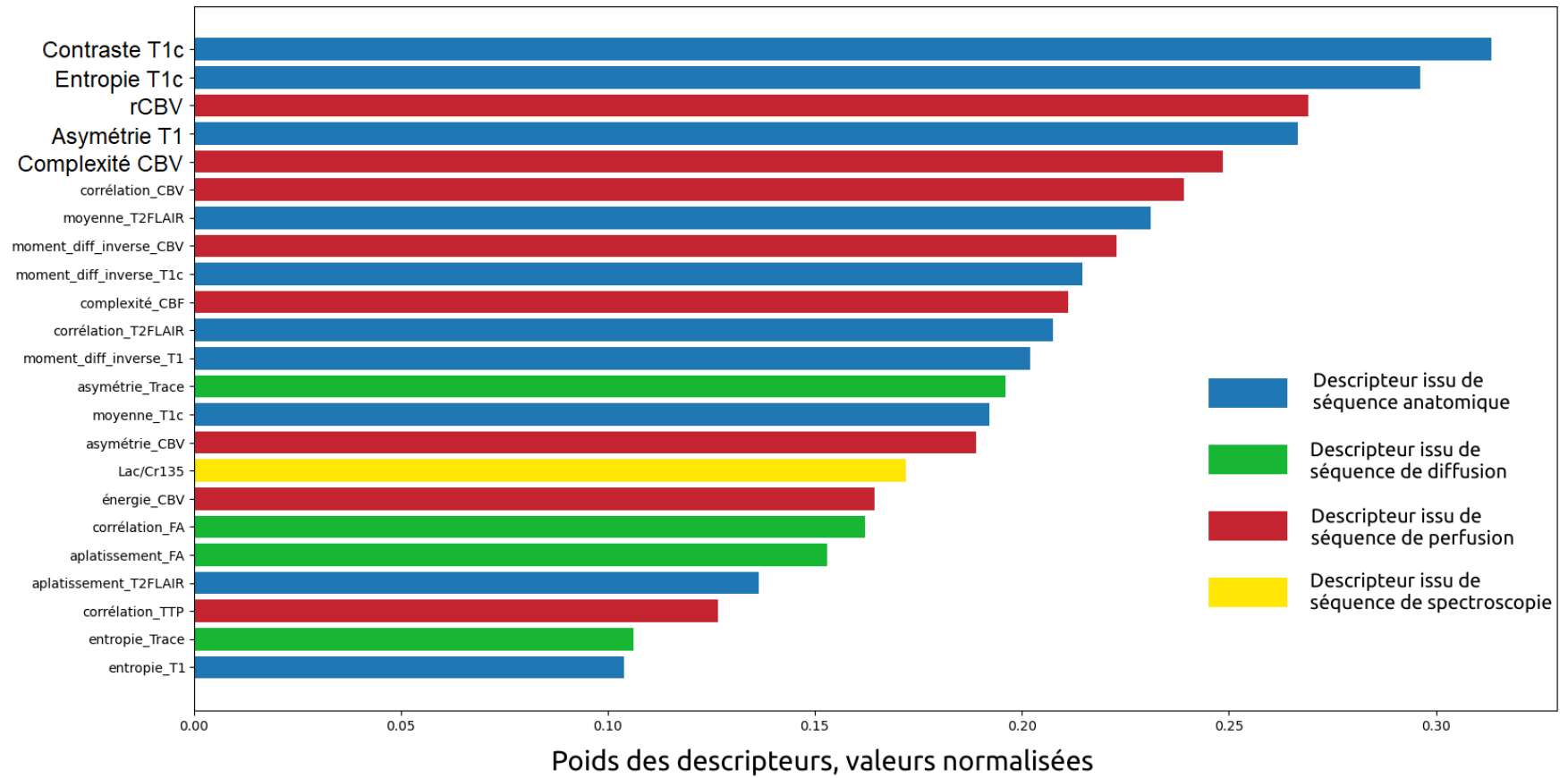


FIGURE 7.8 : Hiérarchisation des descripteurs produite par le meilleur classifieur, analyse par type de séquence RMN.

Hiérarchisation des descripteurs, classifieur entraîné sur données multimodales Analyse par type de descripteur

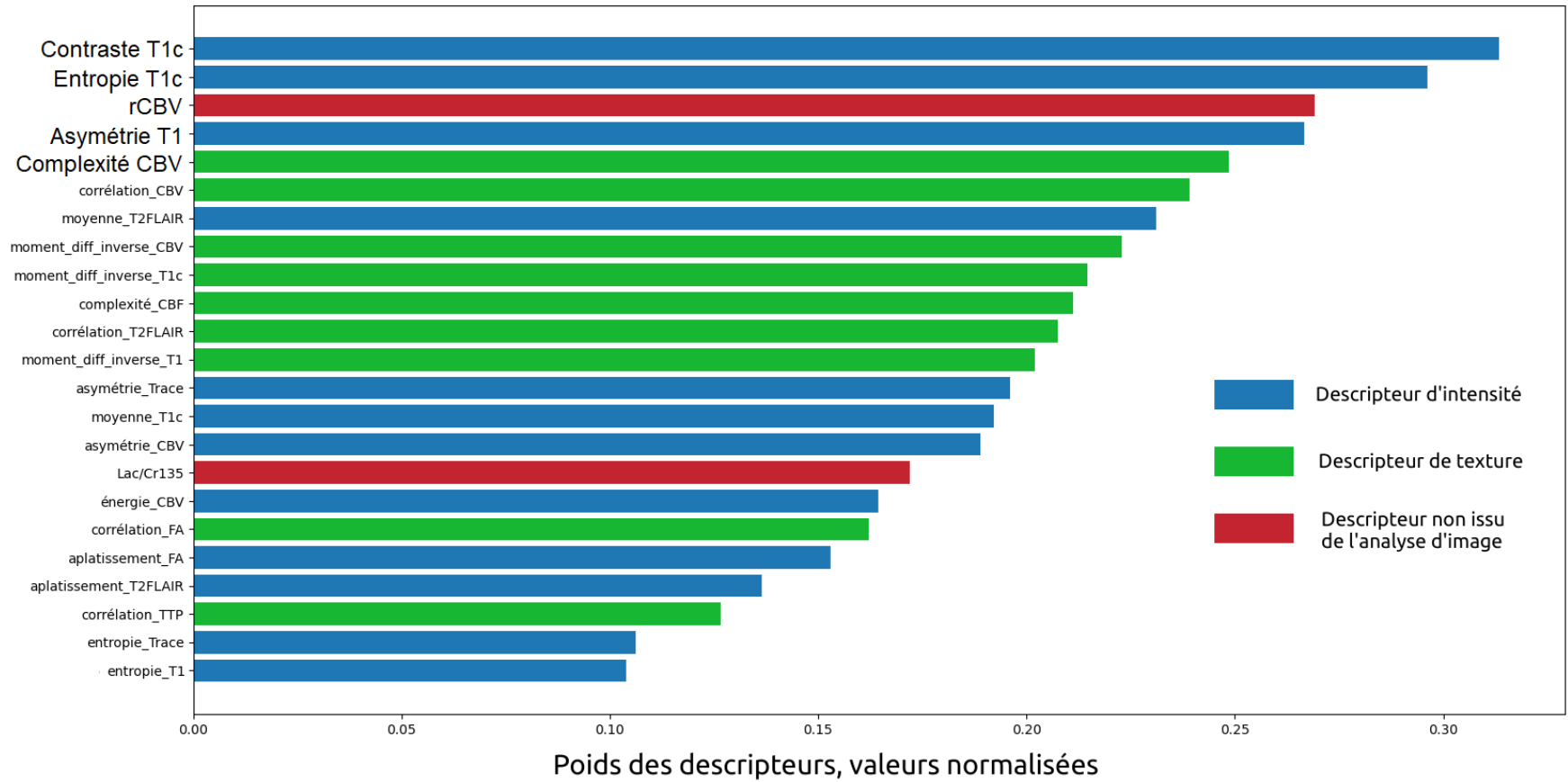


FIGURE 7.9 : Hiérarchisation des descripteurs produite par le meilleur classifieur, analyse par type de descripteur.

Analyse par type de descripteur : Regardons à présent quels types de descripteurs composent ce classement. Bien que les descripteurs d'intensité soient les plus présents avec 12 apparitions, ce sont surtout les descripteurs de texture qui composent la première moitié de la hiérarchisation. Parmi ces descripteurs de texture, on retrouve 4 valeurs de corrélation de la GLCM, pour le CBV, le T2 FLAIR, la FA et le TTP ; ainsi que 3 moments de différence inverse, calculés sur le CBV, le T1c et le T1. On trouve également deux valeurs de complexité, pour le CBV et le CBF.

Les descripteurs d'intensité sont majoritaires dans la hiérarchisation. On les retrouve en tête et dans la seconde moitié du classement. Ils sont issus principalement des séquences anatomiques (7 descripteurs sur les 12 présents). Notons la présence en haut de classement de la valeur moyenne du T2 FLAIR, en 7ème position. La présence de ce descripteur est un peu surprenante, car on pourrait imaginer que la lésion apparaît avec un hypersignal identique qu'il s'agisse d'un LGG ou d'un HGG. La valeur moyenne du T1c est également présente, ce qui est déjà plus compréhensible compte tenu des forts hypersignaux présents dans la zone de prise de contraste. Tous les descripteurs de la hiérarchisation issus du T1c sont des descripteurs d'intensité, ce qui peut souligner l'intérêt de ces informations pour la classification, mais aussi interroger, en creux, sur la pertinence de l'analyse de texture de cette séquence.

Nous l'avons dit précédemment, deux descripteurs non issus de l'analyse d'image sont présents : la valeur de rCBV et le rapport métabolique Lac/Cr à TE long.

Enfin, aucun descripteur de forme n'est présent dans la hiérarchie proposée. De même, ni la rugosité ni la force n'apparaissent dans ce classement. Le descripteur d'énergie n'apparaît qu'une seule fois, pour la cartographie CBV, en deuxième moitié de classement. Le contraste n'est présent qu'une seule fois, mais il s'agit cette fois-ci du contraste du T1c, qui est le meilleur descripteur de la hiérarchie.

Nous avons présenté la hiérarchisation des descripteurs produite par le meilleur classifieur. Mais nous pouvons également produire une analyse plus globale de l'intégralité des descripteurs utilisés. C'est ce que nous allons détailler dans la partie suivante.

7.7.2 Analyse sur 100 entraînements successifs

Nous avons vu avec le Tableau 7.1 que les résultats de l'apprentissage sont soumis à une grande variabilité. Aussi, l'analyse d'un seul classifieur, fût-il le meilleur issu de la validation croisée, ne donne qu'une information partielle. Au cours de l'apprentissage, d'autres classifieurs utilisant d'autres ensemble hiérarchisés de descripteurs sont produits. C'est pourquoi, afin d'avoir une vue plus globale des performances, nous avons proposés des métriques moyennées sur 100 entraînements indépendants. Chaque entraînement a produit son sous-ensemble de descripteurs issus du SVM-RFE. Nous pouvons ainsi recueillir, pour les 100 entraînements indépendants, quels descripteurs font partie du sous-ensemble du SVM-RFE, et incrémenter une liste comptant l'apparition de chaque descripteur dans ces sous-ensembles.

Cette méthode permet de quantifier les descripteurs les plus récurrents lors de la sélection du SVM-RFE. On verra que cette approche permet une vision globale du problème, et produit donc des résultats plus généralisables.

La Figure 7.10 représente le nombre d'apparitions de chaque descripteur dans le sous-ensemble du SVM-RFE sur les 100 entraînements indépendants. Nous observons, par exemple, que le descripteur le plus souvent présent est apparu 99 fois sur 100. Cet histogramme permet de voir que, sur les 100 entraînements, presque tous les descripteurs apparaissent au moins une fois dans le sous-ensemble, à l'exception de 15 d'entre eux. Deuxièmement, il n'y a pas de rupture nette dans l'enveloppe de l'histogramme permettant de séparer un sous-groupe clair de descripteurs essentiels. Certains descripteurs sont bien plus souvent conservés que d'autres, ce qui souligne leur importance vis à vis de la tâche étudiée, mais il n'est pas possible de déterminer visuellement la taille du meilleur sous-ensemble. Le meilleur classifieur obtenu sur 100 entraînements utilisait un ensemble de 23 descripteurs. C'est pourquoi, nous allons analyser dans la suite de ce travail uniquement les 23 premiers et les 23 derniers descripteurs de ce classement global.

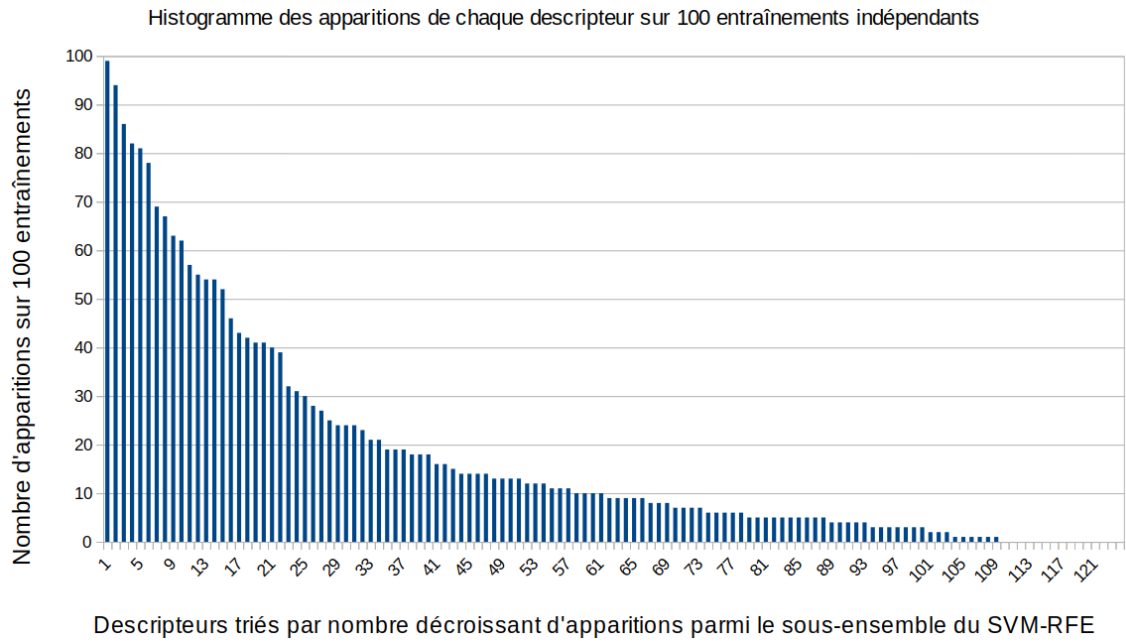


FIGURE 7.10 : *Nombre d'apparitions des 124 descripteurs dans le sous-ensemble du SVM-RFE sur 100 itérations.*

Analyse des 23 meilleurs descripteurs sur 100 itérations : Commençons par un propos général : 14 descripteurs sur 23 sont communs entre la hiérarchisation de descripteurs issus du meilleur classifieur et la hiérarchisation produite par l'analyse sur 100 itérations. Une comparaison est disponible Figure [7.11](#). Nous pouvons noter que les descripteurs communs aux deux listes semblent respecter la même hiérarchie à quelques variations près. Par exemple, en tête de liste, on trouve le contraste du T1c, la valeur de rCBV, l'entropie du T1c. Les trois valeurs de moment de différence inverse du T1c, CBV et T1 se retrouvent également proches les uns des autres. Et, en fin de hiérarchisation, on retrouve l'entropie du T1 ou l'aplatissement de l'histogramme de la FA. Cette correspondance que la fréquence d'apparition des descripteurs sur 100 apprentissages indépendants semble reproduire en partie la hiérarchisation du meilleur classifieur obtenu. C'est pourquoi nous proposons de lire cette liste d'apparition fréquentielle comme une seconde hiérarchisation de descripteurs, moins sujette aux variations de performances que nous avons exposé dans le Tableau [7.1](#).

Tout comme pour le meilleur classifieur, nous pouvons analyser cette hiérarchisation

23 descripteurs les plus fréquents sur 100 itérations	N apparitions	Hierarchie des 23 descripteurs du meilleur classifieur	comparaison
aplatissement_FA	32	entropie_T1	présent
moyenne_T2FLAIR	39	entropie_Trace	absent
complexité_Trace	40	corrélation_TTP	
asymétrie_T1c	41	aplatissement_T2FLAIR	
force_Trace	41	aplatissement_FA	
entropie_T1	42	corrélation_FA	
corrélation_MTT	43	énergie_CBV	
moment_diff_inverse_FA	46	Lac/Cr135	
moyenne_T1c	52	asymétrie_CBV	
corrélation_FA	54	moyenne_T1c	
corrélation_CBF	54	asymétrie_Trace	
rCBF	55	moment_diff_inverse_T1	
aplatissement_CBV	57	corrélation_T2FLAIR	
aplatissement_T1c	62	complexité_CBF	
asymétrie_T1	63	moment_diff_inverse_T1c	
moment_diff_inverse_CBV	67	moment_diff_inverse_CBV	
moment_diff_inverse_T1c	69	moyenne_T2FLAIR	
Lac/Cr135	78	corrélation_CBV	
moment_diff_inverse_T1	81	complexité_CBV	
corrélation_CBV	82	asymétrie_T1	
entropie_T1c	86	rCBV	
rCBV	94	entropie_T1c	
contraste_T1c	99	contraste_T1c	

FIGURE 7.11 : Comparaison de présence des 23 descripteurs les plus fréquents sur 100 itérations et des 23 descripteurs sélectionnés par le meilleur classifieur. La fréquence d'apparition et la hiérarchisation du meilleur classifieur sont présentées de façon croissante, de haut en bas.

selon le type de descripteur utilisé (forme, intensité, texture, relevé manuel pour la SRM et les valeurs de rCBV et rCBF) ou selon la modalité d'imagerie sur laquelle ces descripteurs sont calculés (IRM anatomique, de diffusion, de perfusion, ou SRM). Les 23 descripteurs les plus récurrents parmi les 100 itérations sont représentés par type de séquence RMN sur la Figure 7.12 et par type de descripteur sur la Figure 7.13.

L'analyse par modalité RMN nous montre une nouvelle fois que l'IRM anatomique est la modalité la plus présente en tête de classement. Au total, on dénombre 10 descripteurs issus de l'IRM anatomique contre 5 pour l'IRM de diffusion et 7 pour l'imagerie de perfusion, dont les deux valeurs rCBF et rCBV. Tout comme pour la hiérarchisation produite par le meilleur classifieur, on observe que la première moitié des descripteurs en tête de classement repose à la fois sur l'IRM anatomique et l'IRM de perfusion. Les

descripteurs issus de l'IRM de diffusion n'apparaissent que dans la deuxième moitié du classement.

Nous pouvons noter que le rapport métabolique Lac/Cr à TE long est présent plus en amont dans la hiérarchisation sur 100 itérations que sur la hiérarchisation du meilleur classifieur car il apparaît en 6ème position (contre 16 dans le classement du meilleur classifieur, voir Figure 7.11). Cela peut nous amener à penser que le meilleur classifieur obtenu sur 100 entraînements a tendance à sous-estimer l'information contenue dans ce rapport métabolique. Nous pouvons dire la même chose de la valeur du rCBF qui était absente du meilleur classifieur, et apparaît ici en 12ème position, avec 55 apparitions sur 100, c'est à dire dans plus de la moitié des cas. Nous soulignons l'absence des cartographie de TTP et d'ADC. Cette dernière était déjà absente du meilleur classifieur.

Nous observons également que le T2 FLAIR est moins présent dans ce nouveau classement : il n'apparaît qu'en 22ème position avec la valeur moyenne de la lésion. Dans le meilleur classifieur, le T2 FLAIR apparaissait à 3 reprises, avec sa valeur moyenne, l'aplatissement de son histogramme et la valeur de corrélation de la GLCM. Nous pouvons supposer que l'absence partielle des descripteurs de T2 FLAIR en tête de classement vient du fait que nous extrayons nos descripteurs d'une ROI de segmentation tumorale. Or, la séquence T2 FLAIR est intéressante en imagerie par l'hypersignal qu'elle affiche lorsqu'une lésion est présente. Cet hypersignal est repérable pour l'observateur humain par l'effet de contraste qu'il crée entre la lésion, claire, et le tissu sain, plus sombre. Ainsi, des descripteurs calculés uniquement sur la ROI tumorale ne font que peu ressortir cette différence avec le tissu sain, ce qui peut expliquer la relative absence de descripteurs issus du T2 FLAIR.

L'analyse par type de descripteur utilisé, Figure 7.13, ne donne pas beaucoup d'information : nous avons vu Figure 7.11 que les deux hiérarchisations sont très proches dans leur première moitié de classement. La deuxième moitié montre une coexistence entre descripteurs d'intensité et de texture, ce qui correspond également à la hiérarchisation du meilleur classifieur. Les descripteurs les plus présents sont le moment de différence inverse et la corrélation de la GLCM. Ces deux descripteurs étaient déjà parmi les plus présents dans la hiérarchie du meilleur classifieur.

23 descripteurs les plus fréquents sur 100 itérations	N apparitions	Modalité IRM
contraste_T1c	99	Anatomique
rCBV	94	Perfusion
entropie_T1c	86	Anatomique
corrélation_CBV	82	rCBF et rCBV
moment_diff_inverse_T1	81	Anatomique
Lac/Cr135	78	SRM
moment_diff_inverse_T1c	69	Anatomique
moment_diff_inverse_CBV	67	rCBF et rCBV
asymétrie_T1	63	Anatomique
aplatissement_T1c	62	Anatomique
aplatissement_CBV	57	rCBF et rCBV
rCBF	55	rCBF et rCBV
corrélacion_FA	54	Diffusion
corrélacion_CBF	54	rCBF et rCBV
moyenne_T1c	52	Anatomique
moment_diff_inverse_FA	46	Diffusion
corrélacion_MTT	43	rCBF et rCBV
entropie_T1	42	Anatomique
asymétrie_T1c	41	Anatomique
force_Trace	41	Diffusion
complexité_Trace	40	Diffusion
moyenne_T2FLAIR	39	Anatomique
aplatissement_FA	32	Diffusion

FIGURE 7.12 : Classement par type de modalité RMN (Imagerie anatomique, de diffusion, de perfusion, et SRM) – 23 descripteurs les plus récurrents.

23 descripteurs les plus fréquents sur 100 itérations	N apparitions	Type de descripteur
contraste_T1c	99	Forme
rCBV	94	rCBF et rCBV
entropie_T1c	86	Intensité
corrélation_CBV	82	rCBF et rCBV
moment_diff_inverse_T1	81	rCBF et rCBV
Lac/Cr135	78	SRM
moment_diff_inverse_T1c	69	rCBF et rCBV
moment_diff_inverse_CBV	67	rCBF et rCBV
asymétrie_T1	63	Forme
aplatissement_T1c	62	Forme
aplatissement_CBV	57	rCBF et rCBV
rCBF	55	rCBF et rCBV
corrélacion_FA	54	Intensité
corrélacion_CBF	54	rCBF et rCBV
moyenne_T1c	52	Forme
moment_diff_inverse_FA	46	Intensité
corrélacion_MTT	43	rCBF et rCBV
entropie_T1	42	Intensité
asymétrie_T1c	41	Forme
force_Trace	41	Intensité
complexité_Trace	40	Intensité
moyenne_T2FLAIR	39	Forme
aplatissement_FA	32	Forme

FIGURE 7.13 : Classement par type de descripteur (forme, intensité, texture, mesures manuelles de SRM, rCBV et rCBF) – 23 descripteurs les plus récurrents.

Nous notons à nouveau l'absence des descripteurs de forme. La rugosité est elle aussi absente de ce nouveau classement. La force, qui n'apparaissait pas dans le meilleur classifieur, apparaît une fois en bas de classement : il s'agit de la force de l'image Trace. Inversement, l'énergie, qui était présente une fois dans le meilleur classifieur, n'apparaît plus dans cette hiérarchisation. On peut donc s'interroger sur la pertinence de ces descripteurs.

À nouveau, nous observons que tous les descripteurs issus du T1c présents dans cette hiérarchie sont des descripteurs d'intensité. Ce fait est encore plus marqué que lors de l'analyse du meilleur classifieur, car tous les descripteurs d'intensité à l'exception de l'énergie sont présents. Cela souligne à nouveau que seule l'information d'intensité du T1c semble pertinence pour la classification binaire du grade.

Analyse des 23 derniers descripteurs sur 100 itérations : Avec 100 apprentissages indépendants, nous disposons d'un classement de tous les descripteurs. C'est pourquoi nous proposons d'analyser lesquels n'ont pas été sélectionnés par le SVM-RFE. Tous comme pour les descripteurs les plus récurrents, nous proposons d'analyser les 23 descripteurs les moins présents sur les 100 itérations. Ces derniers sont représentés par type de descripteur sur la Figure [7.14](#) et par type de séquence RMN sur la Figure [7.15](#).

Nous savons d'ores et déjà par la Figure [7.10](#) que parmi ces descripteurs, 15 d'entre eux n'ont jamais été sélectionnés dans le sous-ensemble du SVM-RFE. Nous voyons sur la Figure [7.14](#) qu'il s'agit très majoritairement de descripteurs de texture, en particulier la rugosité et la force, dont nous avons noté l'absence en tête de classement. Ces deux descripteurs représentent à eux seuls 16 des 23 descripteurs les moins fréquents. La Figure [7.14](#) nous montre que ces deux descripteurs se retrouvent en queue de classement quelle que soit la modalité IRM étudiée. Parmi les descripteurs n'ayant jamais été sélectionnés par le SVM-RFE, nous trouvons deux descripteurs d'intensité : le contraste du T2 FLAIR et le contraste du CBV. La présence du contraste du T2 FLAIR est surprenante, car nous avons montré Figure [7.5](#) que les lésions de haut grade pouvaient donner un aspect hétérogène du T2 FLAIR, ce qui résulte en une augmentation du contraste de

la lésion.

23 descripteurs les moins fréquents sur 100 itérations	N apparitions	Type de descripteur
rugosité_T1	0	Forme
force_T1	0	Forme
rugosité_T1c	0	Forme
contraste_T2FLAIR	0	Intensité
rugosité_T2FLAIR	0	Forme
force_T2FLAIR	0	Forme
rugosité_ADC	0	Forme
force_ADC	0	Forme
rugosité_FA	0	Forme
force_FA	0	Forme
rugosité_CBF	0	Forme
contraste_CBV	0	Intensité
rugosité_CBV	0	Forme
rugosité_MTT	0	Forme
force_MTT	0	Forme
énergie_T1	1	Intensité
contraste_T1	1	Intensité
force_T1c	1	Forme
énergie_T2FLAIR	1	Intensité
force_CBF	1	Forme
rugosité_TTP	1	Forme
moyenne_CBV	2	Intensité
contraste_TTP	2	Intensité

FIGURE 7.14 : Classement par type de modalité RMN (imagerie anatomique, de diffusion, de perfusion, et SRM) – 23 descripteurs les moins fréquents.

23 descripteurs les moins fréquents sur 100 itérations	N apparitions	Modalité IRM
rugosité_T1	0	Anatomique
force_T1	0	Anatomique
rugosité_T1c	0	Anatomique
contraste_T2FLAIR	0	Diffusion
rugosité_T2FLAIR	0	Anatomique
force_T2FLAIR	0	Anatomique
rugosité_ADC	0	Diffusion
force_ADC	0	Diffusion
rugosité_FA	0	Diffusion
force_FA	0	Diffusion
rugosité_CBF	0	Perfusion
contraste_CBV	0	Perfusion
rugosité_CBV	0	Perfusion
rugosité_MTT	0	Perfusion
force_MTT	0	Perfusion
énergie_T1	1	Anatomique
contraste_T1	1	Anatomique
force_T1c	1	Anatomique
énergie_T2FLAIR	1	Diffusion
force_CBF	1	Perfusion
rugosité_TTP	1	Perfusion
moyenne_CBV	2	Perfusion
contraste_TTP	2	Perfusion

FIGURE 7.15 : Classement par type de descripteur (forme, intensité, texture, mesures manuelles de SRM, rCBV et rCBF) – 23 descripteurs les moins fréquents.

Enfin, nous avons utilisé le coefficient de corrélation de Pearson (équation 7.1) pour étudier si des variables fortement corrélées pouvaient se dissimuler l'une l'autre dans le classement du SVM-RFE. Deux variables corrélées portent la même information. C'est pourquoi nous pouvons craindre un effet de dissimulation par compétition entre deux descripteurs différents corrélés. L'analyse croisée du coefficient de corrélation de la p-value associée montre que ces effets de dissimulation ne semblent pas exister.

Nous pouvons argumenter ce point avec l'exemple de la corrélation entre les valeurs de rCBV et rCBF. Cette corrélation est très élevée ($r=0.95$) et très significative (p-value de $10e-63$). Pourtant les deux descripteurs se retrouvent dans le classement du SVM-RFE sur 100 apprentissages, avec le rCBV en 2ème position, sélectionné 94 fois sur 100, et le rCBF en 11ème position, apparu 55 fois sur 100. La forte apparition croisée de ces deux descripteurs malgré leur forte corrélation permet donc d'écarter l'hypothèse de dissimulation entre variables corrélées. Ce fait vient souligner à nouveau le caractère robuste du classement obtenu sur 100 apprentissages indépendants.

$$r(x, y) = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}} \quad (7.1)$$

Avec : r le coefficient de Pearson, avec $r \in [-1; 1]$

x et y deux variables aléatoires

\bar{x} et \bar{y} les valeurs moyennes de x et y

N le nombre d'individus

De cette double comparaison de hiérarchisation, nous avons obtenu les informations suivantes :

- Les descripteurs de forme ne sont jamais apparus, ni dans les meilleurs descripteurs, ni dans les moins fréquents ;
- Les descripteurs de rugosité et de force apparaissent très majoritairement parmi les descripteurs les moins fréquents. On peut donc s'interroger sur leur pertinence, et ainsi peut-être épargner du temps de calcul ;

- L'énergie n'apparaît qu'une seule fois dans le meilleur classifieur, est absente des descripteurs les plus fréquents, et se retrouve deux fois parmi les moins fréquents. De la même manière que pour la rugosité et la force, l'intérêt du calcul de ce descripteur se pose ;
- Seuls les descripteurs d'intensité du T1c apparaissent parmi les meilleurs descripteurs, et ce pour le meilleur classifieur, ou pour la hiérarchie sur 100 itérations. L'intérêt de ces descripteurs est donc fortement soulignée. En revanche, les descripteurs de texture du T1c sont absents des descripteurs les plus fréquents ;
- Les descripteurs de texture apparaissant en tête de classement sont le moment de différence inverse et la corrélation de la GLCM. Ces deux descripteurs apparaissent aussi bien pour le meilleur classifieur que pour les 100 entraînements indépendants ;
- Enfin, concernant les séquences IRM, on peut s'interroger sur la pertinence des cartographies d'ADC, de MTT et de TTP. En effet, l'ADC a été absente des deux têtes de classement. Le TTP n'est apparu qu'une seule fois dans le classement du meilleur classifieur, et le MTT n'est apparu qu'une seule fois parmi les descripteurs les plus fréquents sur 100 entraînements. Ces séquences ne semblent pas porter d'information pertinente pour la classification du grade. Ainsi, ne pas acquérir, calculer, ou étudier ces cartographies peut permettre de sauver du temps d'acquisition machine ou du temps humain d'analyse du radiologue.

Dans ce chapitre, nous avons présenté une chaîne de traitement complète, allant de l'acquisition RMN à la classification. Cette chaîne permet d'automatiser le recalage multimodal, l'extraction de l'encéphale, la normalisation en intensité, la segmentation tumorale, et l'extraction des descripteurs. Seuls restent non automatisés le relevé des rapports métaboliques et des valeurs de rCBV et rCBF.

Contrairement à toutes les publications dont nous ayons connaissance utilisant des données multimodales, nous fournissons un point de référence pour l'évaluation des performances du classifieur, par le biais de la base BraTS. Pour cela, nous avons évalué un classifieur SVM sur les données BraTS 2018 avec 4 puis 3 séquences anatomiques.

Comme nous savons que l'étiquetage de cette base ne correspond pas à la division des grades de l'OMS, nous avons entraîné des classifieurs sur les deux vérités terrains dont nous disposions.

Concernant les données du CHU de Poitiers, nous avons entraîné un classifieur sur les données anatomiques pour obtenir un point de référence des performances, puis nous avons enrichi la description de des individus à l'aide des séquences de diffusion, de perfusion et de SRM du proton disponibles au CHU de Poitiers. Un faible gain d'intensité moyenne a été observé entre le passage de la seule utilisation de l'IRM anatomique à l'IRM multimodale. Nous notons également gain pour la valeur moyenne de spécificité : l'ajout d'informations a donc un effet bénéfique sur les performances du classifieur.

Nous avons proposé dans une seconde partie une analyse des différents descripteurs utilisés. Ceci s'est fait via deux méthodes : l'analyse de la hiérarchisation de descripteurs produite par le noyau linéaire du meilleur classifieur d'une part, et l'analyse de l'occurrence des descripteurs en sortie du SVM-RFE à l'issue de 100 entraînements indépendants. Nous avons montré que ces hiérarchies comportaient de nombreux points similaires, et nous proposons même d'utiliser l'analyse sur 100 entraînements comme une hiérarchisation globale des descripteurs, là où le seul SVM-RFE retourne un ensemble de descripteurs tous classés ex-aequo. Ces deux méthodes permettent d'approfondir l'analyse des descripteurs. Enfin, dans notre analyse, nous sommes remontés, depuis la hiérarchisation des descripteurs, à l'analyse d'image du radiologue et avons identifié quels descripteurs et quelles séquences présentaient des informations pertinentes ou non.

Conclusions et perspectives

Contenu du chapitre

8.1 Conclusion	190
8.2 Perspectives	193

8.1 Conclusion

Les gliomes, originant du tissu de soutien cérébral, sont les tumeurs du système nerveux central les plus fréquentes. Certaines présentent un développement silencieux, doublé d'une infiltration lente et diffuse dans les tissus sains environnants. D'autres surviennent *de novo* avec un développement rapide et agressif. Dans tous les cas, ces tumeurs sont liées à une haute mortalité et la question de la durée de survie pour le patient se pose à court ou moyen terme.

Selon l'agressivité de la lésion, la stratégie thérapeutique s'adapte. Radiothérapie, chimiothérapie et chirurgie sont les principales voies d'intervention. Ce choix nécessite de connaître et de caractériser la dynamique tumorale.

Pour cela, des outils d'exploration à base d'imagerie par résonance magnétique sont développés. Le CHU de Poitiers a fait l'acquisition de plusieurs imageurs permettant l'acquisition de données multimodales, très diverses et riches en informations, afin de produire un diagnostic et une analyse les plus complets possibles. Se retrouvent ainsi regroupées des informations anatomiques, de diffusion, de perfusion et de spectroscopie multi-noyaux. La combinaison de ces informations permet d'affiner le diagnostic et d'étudier le développement tumoral.

Le grade tumoral, information standardisée par l'Organisation Mondiale de la Santé (OMS), est un indicateur de la dynamique tumorale et de la dangerosité de la lésion. Sont ainsi distingués les gliomes de bas grades, grades I-II, et les gliomes de hauts grades, grades III-IV. L'établissement du grade se fait par biopsie, ou sur l'analyse de pièce chirurgicale, qui requièrent des techniques invasives pour le patient. Développer une alternative robuste et non-invasive depuis l'imagerie est alors un sujet porteur de nombreux enjeux.

De part sa grande adaptabilité aux problèmes de forte dimensionnalité, l'apprentissage automatique est couramment utilisé lors du traitement et de l'exploration des données de santé. Mais dans la tâche de caractérisation du grade tumoral, les données manquent : rares sont les travaux publiés qui ont recours à toute la diversité offerte par

l'examen IRM multimodal.

C'est dans ce cadre que ce travail de thèse apporte plusieurs contributions :

- Nous présentons un panorama des méthodes utilisées dans le champ scientifique de la classification binaire et automatique du grade du gliome. Nous y présentons les classifieurs implémentés, les pré-traitements effectués et les bases de données utilisées dans les travaux récents de la littérature ;
- Nous développons, via l'implémentation d'un classifieur automatique classique, un outil d'analyse de performances de classifieur avec l'utilisation de l'histogramme des scores de prédiction. Cet outil est indépendant des métriques issues de la matrice de confusion comme la précision, la sensibilité ou la spécificité. Il renseigne l'utilisateur sur le degré de confiance du classifieur en sa réponse ;
- Nous montrons que ce même outil permet une analyse de base de données, faisant apparaître des groupes d'individus partageant des caractéristiques similaires ;
- Par ce biais, nous pointons que la base de données la plus utilisée pour la classification binaire du grade du gliome, la base de données BraTS, n'est pas adaptée à cette tâche selon les standards de l'OMS ;
- Par une tâche d'étiquetage randomisée, nous proposons un ré-alignement de la base BraTS sur la classification binaire de l'OMS, dans les limites posées par l'imagerie anatomique. Nous montrons que malgré un déséquilibre de classe marqué, apprendre sur cette nouvelle vérité terrain n'affecte que faiblement les performances de notre classifieur ;
- Nous présentons une chaîne de traitement complète, permettant d'automatiser le traitement des données multimodales du CHU de Poitiers, de l'acquisition à la classification automatique ;
- Nous étudions comment l'ajout des données de diffusion, de perfusion et de spectroscopie vient impacter les performances de nos classifieurs. Nous fournissons un point de référence permettant de comparer nos performances avec l'état de l'art ;

- Nous présentons deux méthodes d'analyse permettant d'obtenir une hiérarchisation des descripteurs, et donc de l'information contenue par séquence et par type de descripteur d'image. La première permet d'étudier la hiérarchisation produite par un classifieur seul. La seconde montre qu'une hiérarchisation produite sur N entraînements indépendants vient compléter la hiérarchisation produite par un classifieur seul, notamment en faisant remonter dans la hiérarchie des descripteurs ignorés par le classifieur unique ;
- Cette analyse de hiérarchisations de descripteurs nous permet de souligner la pertinence des descripteurs d'intensité du T1c ; d'éliminer les descripteurs de rugosité, de force et d'énergie ; et de pointer la faible valeur informationnelle des cartographies d'ADC, de MTT et de TTP pour la tâche de classification binaire du grade du gliome.

Ces propositions ont fait l'objet de deux publications scientifiques dans des conférences internationales à comité de relecture [32, 33]. Une version détaillée du travail présenté section 6 est en cours de rédaction pour le numéro spécial de la revue Elsevier *Innovation and Research in BioMedical engineering* (IRBM) associé. Une quatrième publication, relatant le travail effectué section 7 est aussi en court d'écriture.

Pour le CHU de Poitiers, ces travaux de thèse permettent tels quels d'obtenir une forme d'aide au diagnostic automatisé pour les radiologues. En effet, pour l'activité de recherche, chaque maillon de la chaîne de traitement fonctionne, dans la limite des performances présentées. Ainsi, un nouveau patient ajouté à la base de données peut être extrait, recalé, normalisé, segmenté, puis classé par notre classifieur. Ce dernier peut même fournir deux types de réponses : le résultat de prédiction via l'attribution d'une étiquette, ou simplement le score de prédiction associé à ce patient. Ainsi, le radiologue peut juger de la performance de l'algorithme et de son accord avec l'examen d'imagerie.

8.2 Perspectives

Établissement d'une grande base de données multimodale : La première des perspectives est de proposer une grande base de données multimodales. En effet, s'émanciper de la base BraTS est une nécessité. Nous avons montré dans ce travail de thèse combien cette base était problématique pour la tâche de classification binaire. De plus, proposer une base de données IRM haut champ et ultra haut champ permettra de disposer d'images de meilleures résolutions. Avec ces résolutions viendront de meilleurs outils pour l'analyse d'image RMN, qui permettront de meilleures segmentations tumorales et de meilleurs résultats pour l'aide au diagnostic. En cela, la plateforme d'imagerie du CHU de Poitiers, notamment via l'IRM 7 Telsa et ses acquisitions, ouvrent la voie permettant ces études.

Classification selon les quatre grades de l'OMS : La deuxième perspective à envisager suite à ce travail est d'ouvrir le champ de la classification aux quatre grades de l'OMS. Plusieurs travaux commencent à s'intéresser à des discriminations grade II, grade III, grade IV. L'analyse d'image automatique, à travers certains marqueurs, semble capable d'extraire depuis les acquisitions RMN, les informations moléculaires et génétiques. Ainsi, une classification plus fine, multiclasse, est un pas supplémentaire vers la création d'une biopsie virtuelle et robuste.

Utilisation des scores de prédiction pour l'apprentissage profond : L'obtention des scores de prédiction n'est pas limitée aux SVM. Aussi, il serait intéressant de coupler cet outil d'analyse à l'utilisation des algorithmes d'apprentissage profond, notamment les réseaux de neurones à couches denses. En effet, lors d'une tâche de classification, chaque neurone de la couche de sortie dispose d'un score de prédiction propre. Ainsi, l'analyse des scores de prédiction peut fournir des informations sur le comportement du réseau et ainsi aider à rendre plus intelligible les structures d'apprentissage. Ceci pose une deuxième question, sur laquelle il serait intéressant de travailler : comment étendre l'analyse proposée par l'histogramme des scores de prédiction à une classification à K

classes ? L'interprétation de ces histogrammes devient plus difficile lorsque le nombre de classes augmente. En effet, le seuil permettant de discriminer entre partie gauche et partie droite de l'histogramme n'est plus fixe à 0,5, mais devient flottant. Il y a donc un travail à fournir sur la généralisation multiclasse de cet outil.

Les perspectives que nous proposons illustrent les deux champs de recherche à l'intersection du travail de cette thèse : imagerie médicale et apprentissage automatique. Ainsi, nous clôturons cette thèse en nous appuyant sur les deux questions qui portent ce travail : comment faire pour mieux décrire et explorer la réalité biologique et la dynamique des tumeurs cérébrales ? Et comment peut-on sortir les applications d'apprentissage automatique de leur effet de boîte noire, pour les rendre intelligible et ainsi mieux les interpréter ? Nous espérons que nos travaux aient pu en esquisser un début de réponse.

Bibliographie

- [1] Cancer - organisation mondiale de la santé, disponible à <https://www.who.int/fr/news-room/fact-sheets/detail/cancer>, dernier accès le 24 novembre 2020.
- [2] Logiciel 3d slicer, disponible à <https://www.slicer.org/>, dernier accès le 29 septembre 2020.
- [3] Jill M Abrigo, Daniel M Fountain, James M Provenzale, Eric K Law, Joey SW Kwong, Michael G Hart, and Wilson Wai San Tam. Magnetic resonance perfusion for differentiating low-grade from high-grade gliomas at first presentation. *Cochrane Database of Systematic Reviews*, (1), 2018.
- [4] Naomi S Altman. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3) :175–185, 1992.
- [5] Javaria Amin, Muhammad Sharif, Mudassar Raza, Tanzila Saba, Rafiq Sial, and Shafqat Ali Shad. Brain tumor detection : a long short-term memory (lstm)-based learning model. *Neural Computing and Applications*, 32(20) :15965–15973, 2020.
- [6] Amin Kabir Anaraki, Moosa Ayati, and Foad Kazemi. Magnetic resonance imaging-based brain tumor grades classification and grading via convolutional neural networks and genetic algorithms. *Biocybernetics and Biomedical Engineering*, 39(1) :63–74, 2019.
- [7] Brian B Avants, Nick Tustison, and Gang Song. Advanced normalization tools (ants). *Insight j*, 2(365) :1–35, 2009.
- [8] Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin S Kirby, John B Freymann, Keyvan Farahani, and Christos Davatzikos. Ad-

- vancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific data*, 4 :170117, 2017.
- [9] Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin S Kirby, John B Freymann, Keyvan Farahani, and Christos Davatzikos. Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific data*, 4 :170117, 2017.
- [10] Spyridon Bakas, Mauricio Reyes, Andras Jakab, Stefan Bauer, Markus Rempfler, Alessandro Crimi, Russell Takeshi Shinohara, Christoph Berger, Sung Min Ha, Martin Rozycki, et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge. *arXiv preprint arXiv :1811.02629*, 2018.
- [11] Spyridon Bakas, Mauricio Reyes, Andras Jakab, Stefan Bauer, Markus Rempfler, Alessandro Crimi, Russell Takeshi Shinohara, Christoph Berger, Sung Min Ha, Martin Rozycki, et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge. *arXiv preprint arXiv :1811.02629*, 2018.
- [12] P Bandettini, M Jenkinson, CF Beckmann, TE Behrens, MW Woolrich, and SM Smith. Fsl. *NeuroImage*, 62(2) :782–790, 2012.
- [13] Subhashis Banerjee, Sushmita Mitra, Francesco Masulli, and Stefano Rovetta. Deep radiomics for brain tumor detection and classification from multi-sequence mri. *arXiv preprint arXiv :1903.09240*, 2019.
- [14] Luc Bauchet. Epidemiology of diffuse low grade gliomas. In *Diffuse low-grade gliomas in adults*, pages 13–53. Springer, 2017.
- [15] Richard Beare, Bradley Lowekamp, and Ziv Yaniv. Image segmentation, registration and characterization in r with simpleitk. *Journal of statistical software*, 86, 2018.

- [16] Felix Bloch. Nuclear induction. *Physical review*, 70(7-8) :460, 1946.
- [17] Hans-Hermann Bock. Origins and extensions of the k-means algorithm in cluster analysis. *Electronic Journal for History of Probability and Statistics*, 4(2) :1–18, 2008.
- [18] Kay Henning Brodersen, Cheng Soon Ong, Klaas Enno Stephan, and Joachim M Buhmann. The balanced accuracy and its posterior distribution. In *2010 20th International Conference on Pattern Recognition*, pages 3121–3124. IEEE, 2010.
- [19] Tianqi Chen and Carlos Guestrin. Xgboost : A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.
- [20] Wei Chen, Boqiang Liu, Suting Peng, Jiawei Sun, and Xu Qiao. Computer-aided grading of gliomas combining automatic segmentation and radiomics. *International journal of biomedical imaging*, 2018, 2018.
- [21] Hwan-ho Cho, Seung-hak Lee, Jonghoon Kim, and Hyunjin Park. Classification of the glioma grading using radiomics analysis. *PeerJ*, 6 :e5982, 2018.
- [22] Hwan-ho Cho and Hyunjin Park. Classification of low-grade and high-grade glioma using multi-modal image radiomics features. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 3081–3084. IEEE, 2017.
- [23] Fusun Citak-Er, Zeynep Firat, Ilhami Kovanlikaya, Ugur Ture, and Esin Ozturk-Isik. Machine-learning in grading of gliomas based on multi-parametric magnetic resonance imaging at 3t. *Computers in biology and medicine*, 99 :154–160, 2018.
- [24] Kenneth Clark, Bruce Vendt, Kirk Smith, John Freymann, Justin Kirby, Paul Koppel, Stephen Moore, Stanley Phillips, David Maffitt, Michael Pringle, et al. The cancer imaging archive (tcia) : maintaining and operating a public information repository. *Journal of digital imaging*, 26(6) :1045–1057, 2013.

- [25] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3) :273–297, 1995.
- [26] W Maxwell Cowan, Donald H Harter, and Eric R Kandel. The emergence of modern neuroscience : some implications for neurology and psychiatry. *Annual review of neuroscience*, 23(1) :343–391, 2000.
- [27] Ge Cui, Jiwoong Jeong, Bob Press, Yang Lei, Hui-Kuo Shu, Tian Liu, Walter Curran, Hui Mao, and Xiaofeng Yang. Machine-learning-based classification of lower-grade gliomas and high-grade gliomas using radiomic features in multi-parametric mri. *arXiv preprint arXiv :1911.10145*, 2019.
- [28] Antonio R Damasio. *L’erreur de Descartes : la raison des émotions*. Odile Jacob, 2006.
- [29] Beatrice De Coene, Joseph V Hajnal, Peter Gatehouse, Donald B Longmore, Susan J White, Angela Oatridge, JM Pennock, IR Young, and GM Bydder. Mr of the brain using fluid-attenuated inversion recovery (flair) pulse sequences. *American journal of neuroradiology*, 13(6) :1555–1564, 1992.
- [30] Bruce L Dean, Burton P Drayer, C Roger Bird, Richard A Flom, John A Hodak, Stephen W Coons, and Russell G Carey. Gliomas : classification with mr imaging. *Radiology*, 174(2) :411–415, 1990.
- [31] Gautier Defossez, Sandra Le Guyader-Peyrou, Zoé Uhry, Pascale Grosclaude, Marc Colonna, Emmanuelle Dantony, Patricia Delafosse, Florence Molinié, Anne-Sophie Woronoff, Anne-Marie Bouvier, et al. Estimations nationales de l’incidence et de la mortalité par cancer en france métropolitaine entre 1990 et 2018. *Etude à partir des registres des cancers du réseau Francim. Résultats préliminaires. Synthèse. Saint-Maurice (Fra) : Santé Publique France*, page 19, 2019.
- [32] Paul Dequidt, Pascal Bourdon, Olfa Ben Ahmed, Benoit Tremblais, Carole Guillevin, Mathieu Naudin, Christine Fernandez-Maloigne, and Remy Guillevin. Recent advances in glioma grade classification using machine and deep learning on mr

- data. In *2019 Fifth International Conference on Advances in Biomedical Engineering (ICABME)*, pages 1–4. IEEE, 2019.
- [33] Paul Dequidt, Pascal Bourdon, Benoît Tremblais, Carole Guillevin, Benoît Gianneli, Claire Boutet, Jean-Philippe Cottier, Jean-Noël Vallée, Christine Fernandez-Maloigne, and Rémy Guillevin. Assigning a new glioma grade label ground-truth for the brats dataset using radiologic criteria. In *Tenth International Conference on Image Processing Theory, Tools and Applications (IPTA) - in proceedings*, pages 1–5. IEEE, 2020.
- [34] Paul Dequidt and Angélique Perrillat-Mercerot. Des données biologiques aux modèles et inversement. *Images des Mathématiques*, 2019.
- [35] Jean-Louis Dietemann. *Neuro-imagerie diagnostique*. Elsevier Health Sciences, 2018.
- [36] Emmanuel Dion. Invitation à la théorie de l’information. 1997.
- [37] Hugues Duffau. *Diffuse low-grade gliomas in adults*. Springer, 2017.
- [38] Andriy Fedorov, Reinhard Beichel, Jayashree Kalpathy-Cramer, Julien Finet, Jean-Christophe Fillion-Robin, Sonia Pujol, Christian Bauer, Dominique Jennings, Fiona Fennessy, Milan Sonka, et al. 3d slicer as an image computing platform for the quantitative imaging network. *Magnetic resonance imaging*, 30(9) :1323–1341, 2012.
- [39] A Fenneteau, P Bourdon, D Helbert, C Fernandez-Maloigne, and CHU de Poitiers. Apprentissage d’un cnn sur la segmentation des lésions de sclérose en plaques avec auto-supervision.
- [40] Lionel Feuilleux, Henry Dufour, and Jean Pelletier. Brain of a white-collar worker. *The Lancet*, 370(9583) :262, 2007.
- [41] Chenjie Ge, Irene Yu-Hua Gu, Asgeir Store Jakola, and Jie Yang. Deep learning and multi-sensor fusion for glioma classification using multistream 2d convolutional

- networks. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5894–5897. IEEE, 2018.
- [42] Chenjie Ge, Qixun Qu, Irene Yu-Hua Gu, and Asgeir Store Jakola. 3d multi-scale convolutional networks for glioma grading using mr images. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 141–145. IEEE, 2018.
- [43] Tal Geva. Magnetic resonance imaging : historical perspective. *Journal of cardiovascular magnetic resonance*, 8(4) :573–580, 2006.
- [44] ER Gizewski, C Mönninghoff, and M Forsting. Perspectives of ultra-high-field mri in neuroradiology. *Clinical neuroradiology*, 25(2) :267–273, 2015.
- [45] Todd R Golub, Donna K Slonim, Pablo Tamayo, Christine Huard, Michelle Gassenbeek, Jill P Mesirov, Hilary Coller, Mignon L Loh, James R Downing, Mark A Caligiuri, et al. Molecular classification of cancer : class discovery and class prediction by gene expression monitoring. *science*, 286(5439) :531–537, 1999.
- [46] Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Advances in neural information processing systems*, 3(06), 2014.
- [47] Anshit Goyal, Yagiz U Yolcu, Aakshit Goyal, Panagiotis Kerezoudis, Desmond A Brown, Christopher S Graffeo, Sandy Goncalves, Terence C Burns, and Ian F Parney. The t2-flair-mismatch sign as an imaging biomarker for idh and 1p/19q status in diffuse low-grade gliomas : a systematic review with a bayesian approach to evaluation of diagnostic test performance. *Neurosurgical focus*, 47(6) :E13, 2019.
- [48] Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. Gene selection for cancer classification using support vector machines. *Machine learning*, 46(1-3) :389–422, 2002.
- [49] B Hakyemez, C Erdogan, I Ercan, N Ergin, S Uysal, and S Atahan. High-grade and low-grade gliomas : differentiation by using perfusion mr imaging. *Clinical radiology*, 60(4) :493–502, 2005.

- [50] Changhee Han, Hideaki Hayashi, Leonardo Rundo, Ryosuke Araki, Wataru Shimoda, Shinichi Muramatsu, Yujiro Furukawa, Giancarlo Mauri, and Hideki Nakayama. Gan-based synthetic brain mr image generation. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 734–738. IEEE, 2018.
- [51] Tin Kam Ho. Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition*, volume 1, pages 278–282. IEEE, 1995.
- [52] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8) :1735–1780, 1997.
- [53] Pu Huang, Dengwang Li, Zhicheng Jiao, Dongming Wei, Guoshi Li, Qian Wang, Han Zhang, and Dinggang Shen. Coca-gan : Common-feature-learning-based context-aware generative adversarial network for glioma grading. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 155–163. Springer, 2019.
- [54] Fabian Isensee, Philipp Kickingereder, Wolfgang Wick, Martin Bendszus, and Klaus H Maier-Hein. Brain tumor segmentation and radiomics survival prediction : Contribution to the brats 2017 challenge. In *International MICCAI Brainlesion Workshop*, pages 287–297. Springer, 2017.
- [55] Robert J Jackson, Gregory N Fuller, Dima Abi-Said, Frederick F Lang, Ziya L Gokaslan, Wei Ming Shi, David M Wildrick, and Raymond Sawaya. Limitations of stereotactic biopsy in the initial management of gliomas. *Neuro-oncology*, 3(3) :193–200, 2001.
- [56] Yuchen Jiao, Patrick J Killela, Zachary J Reitman, B Ahmed Rasheed, Christopher M Heaphy, Roeland F de Wilde, Fausto J Rodriguez, Sergio Rosemberg, Sueli Mieko Oba-Shinjo, Suelly Kazue Nagahashi Marie, et al. Frequent atrx, cic, fubp1 and idh1 mutations refine the classification of malignant gliomas. *Oncotarget*, 3(7) :709, 2012.

- [57] Tareq A Juratli, Shilpa S Tummala, Angelika Riedl, Dirk Daubner, Silke Hennig, Tristan Penson, Amir Zolal, Christian Thiede, Gabriele Schackert, Dietmar Krex, et al. Radiographic assessment of contrast enhancement and t2/flair mismatch sign in lower grade gliomas : correlation with molecular groups. *Journal of neuro-oncology*, 141(2) :327–335, 2019.
- [58] Bruno Kastler and Daniel Vetter. *Comprendre l'IRM : manuel d'auto-apprentissage*. Elsevier Health Sciences, 2018.
- [59] Teuvo Kohonen. The self-organizing map. *Proceedings of the IEEE*, 78(9) :1464–1480, 1990.
- [60] Anil Kumar, Dieter Welti, and Richard R Ernst. Nmr fourier zeugmatography. *Journal of magnetic resonance*, 213(2) :495–509, 2011.
- [61] Manoj Kumar, Rahu Ramakrishnaiah, and Rohan Samant. Angiocentric glioma, a recently added who grade-i tumor. *Radiology case reports*, 8(4) :782, 2013.
- [62] M Labussiere, A Idbaih, X-W Wang, Y Marie, B Boisselier, C Falet, S Paris, J Laffaire, C Carpentier, E Criniere, et al. All the 1p19q codeleted gliomas are mutated on idh1 or idh2. *Neurology*, 74(23) :1886–1890, 2010.
- [63] Paul C Lauterbur. Image formation by induced local interactions : examples employing nuclear magnetic resonance. *nature*, 242(5394) :190–191, 1973.
- [64] Denis Le Bihan, Eric Breton, Denis Lallemand, Philippe Grenier, Emmanuel Cabanis, and Maurice Laval-Jeantet. Mr imaging of intravoxel incoherent motions : application to diffusion and perfusion in neurologic disorders. *Radiology*, 161(2) :401–407, 1986.
- [65] Yann LeCun, Yoshua Bengio, et al. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10) :1995, 1995.

- [66] Eyal Lotan, Rajan Jain, Narges Razavian, Girish M Fatterpekar, and Yvonne W Lui. State of the art : machine learning applications in glioma imaging. *American Journal of Roentgenology*, 212(1) :26–37, 2019.
- [67] David N Louis, Hiroko Ohgaki, Otmar D Wiestler, Webster K Cavenee, Peter C Burger, Anne Jouvett, Bernd W Scheithauer, and Paul Kleihues. The 2007 who classification of tumours of the central nervous system. *Acta neuropathologica*, 114(2) :97–109, 2007.
- [68] David N Louis, Arie Perry, Guido Reifenberger, Andreas Von Deimling, Dominique Figarella-Branger, Webster K Cavenee, Hiroko Ohgaki, Otmar D Wiestler, Paul Kleihues, and David W Ellison. The 2016 world health organization classification of tumors of the central nervous system : a summary. *Acta neuropathologica*, 131(6) :803–820, 2016.
- [69] Bradley Christopher Lowekamp, David T Chen, Luis Ibáñez, and Daniel Blezek. The design of simpleitk. *Frontiers in neuroinformatics*, 7 :45, 2013.
- [70] Dan Ma, Vikas Gulani, Nicole Seiberlich, Kecheng Liu, Jeffrey L Sunshine, Jeffrey L Duerk, and Mark A Griswold. Magnetic resonance fingerprinting. *Nature*, 495(7440) :187–192, 2013.
- [71] Eleanor A Maguire, David G Gadian, Ingrid S Johnsrude, Catriona D Good, John Ashburner, Richard SJ Frackowiak, and Christopher D Frith. Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences*, 97(8) :4398–4403, 2000.
- [72] Peter Mansfield. Multi-planar image formation using nmr spin echoes. *Journal of Physics C : Solid State Physics*, 10(3) :L55, 1977.
- [73] Malia F Mason, Michael I Norton, John D Van Horn, Daniel M Wegner, Scott T Grafton, and C Neil Macrae. Wandering minds : the default network and stimulus-independent thought. *Science*, 315(5810) :393–395, 2007.

- [74] Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, 34(10) :1993–2024, 2014.
- [75] Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, 34(10) :1993–2024, 2014.
- [76] Melanie Mitchell. *An introduction to genetic algorithms*. MIT press, 1998.
- [77] Geethu Mohan and M Monica Subashini. Mri based medical image analysis : Survey on brain tumor grade classification. *Biomedical Signal Processing and Control*, 39 :139–161, 2018.
- [78] Jun Morimoto, Gordon Cheng, Christopher G Atkeson, and Garth Zeglin. A simple reinforcement learning algorithm for biped walking. In *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA'04. 2004*, volume 3, pages 3030–3035. IEEE, 2004.
- [79] Erin S Murphy, Charles M Leyrer, Michael Parsons, John H Suh, Samuel T Chao, S Yu Jennifer, Rupesh Kotecha, Xuefei Jia, David M Peereboom, Richard A Prayson, et al. Risk factors for malignant transformation of low-grade glioma. *International Journal of Radiation Oncology* Biology* Physics*, 100(4) :965–971, 2018.
- [80] Hiba Mzoughi, Ines Njeh, Ali Wali, Mohamed Ben Slima, Ahmed BenHamida, Chokri Mhiri, and Kharedine Ben Mahfoudhe. Deep multi-scale 3d convolutional neural network (cnm) for mri gliomas brain tumor classification. *Journal of Digital Imaging*, 2020.
- [81] Pravanya Nandakumar, Alireza Mansouri, and Sunit Das. The role of atrx in glioma biology. *Frontiers in oncology*, 7 :236, 2017.

- [82] Mathieu Naudin. *Visualisation et aide à la décision pour la neuro-navigation per-opératoire*. PhD thesis, Poitiers, 2018.
- [83] Mathieu Naudin, Benoit Tremblais, Carole Guillevin, Rémy Guillevin, and Christine Fernandez-Maloigne. Diffuse low grade glioma nmr assessment for better intra-operative targeting using fuzzy logic. In *International Conference on Advanced Concepts for Intelligent Vision Systems*, pages 200–210. Springer, 2018.
- [84] Evan Noch and Kamel Khalili. Molecular mechanisms of necrosis in glioblastoma : the role of glutamate excitotoxicity. *Cancer biology & therapy*, 8(19) :1791–1797, 2009.
- [85] Hiroko Ohgaki and Paul Kleihues. Genetic pathways to primary and secondary glioblastoma. *The American journal of pathology*, 170(5) :1445–1453, 2007.
- [86] Quinn T Ostrom, David J Cote, Mustafa Ascha, Carol Kruchko, and Jill S Barnholtz-Sloan. Adult glioma incidence and survival by race or ethnicity in the united states from 2000 to 2014. *JAMA oncology*, 4(9) :1254–1262, 2018.
- [87] Quinn T Ostrom, Haley Gittleman, Peter Liao, Toni Vecchione-Koval, Yingli Wolinsky, Carol Kruchko, and Jill S Barnholtz-Sloan. Cbtrus statistical report : primary brain and other central nervous system tumors diagnosed in the united states in 2010–2014. *Neuro-oncology*, 19(suppl_5) :v1–v88, 2017.
- [88] Quinn T Ostrom, Haley Gittleman, Lindsay Stetson, Selene Virk, and Jill S Barnholtz-Sloan. Epidemiology of intracranial gliomas. In *Intracranial Gliomas Part I-Surgery*, volume 30, pages 1–11. Karger Publishers, 2018.
- [89] David Panzoli. *Proposition de l'architecture " Cortexionist " pour l'intelligence comportementale de créatures artificielles*. PhD thesis, Université de Toulouse, Université Toulouse III-Paul Sabatier, 2008.
- [90] MJ Park, HS Kim, G-H Jahng, C-W Ryu, SM Park, and SY Kim. Semiquantitative assessment of intratumoral susceptibility signals using non-contrast-enhanced

- high-field high-resolution susceptibility-weighted imaging in patients with gliomas : comparison with mr perfusion imaging. *American Journal of Neuroradiology*, 30(7) :1402–1408, 2009.
- [91] Sohil H Patel, Laila M Poisson, Daniel J Brat, Yueren Zhou, Lee Cooper, Matija Snuderl, Cheddi Thomas, Ana M Franceschi, Brent Griffith, Adam E Flanders, et al. T2–flair mismatch, an imaging biomarker for idh and 1p/19q status in lower-grade gliomas : a tcga/tcia project. *Clinical Cancer Research*, 23(20) :6078–6085, 2017.
- [92] Angélique Perrillat-Mercerot. *Modélisation et étude du métabolisme énergétique cérébral. Applications à l'imagerie des gliomes diffus de bas grade*. PhD thesis, Poitiers, 2019.
- [93] John Platt et al. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers*, 10(3) :61–74, 1999.
- [94] Matthew B Potts, Justin S Smith, Annette M Molinaro, and Mitchel S Berger. Natural history and surgical management of incidentally discovered low-grade gliomas. *Journal of neurosurgery*, 116(2) :365–372, 2012.
- [95] Edward M Purcell, Henry Cutler Torrey, and Robert V Pound. Resonance absorption by nuclear magnetic moments in a solid. *Physical review*, 69(1-2) :37, 1946.
- [96] Chong Qi, Yiming Li, Xing Fan, Yin Jiang, Rui Wang, Song Yang, Lanxi Meng, Tao Jiang, and Shaowu Li. A quantitative svm approach potentially improves the accuracy of magnetic resonance spectroscopy in the preoperative evaluation of the grades of diffuse gliomas. *NeuroImage : Clinical*, 23 :101835, 2019.
- [97] Sebastian Raschka and Vahid Mirjalili. *Python machine learning*. Packt Publishing Ltd, 2017.

- [98] Valérie Rigau. Towards an intermediate grade in glioma classification. In *Diffuse Low-Grade Gliomas in Adults*, pages 101–108. Springer, 2017.
- [99] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net : Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
- [100] Frank Rosenblatt. The perceptron : a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6) :386, 1958.
- [101] Muhammad Sajjad, Salman Khan, Khan Muhammad, Wanqing Wu, Amin Ullah, and Sung Wook Baik. Multi-grade brain tumor classification using deep cnn with extensive data augmentation. *Journal of computational science*, 30 :174–182, 2019.
- [102] A Salami, M Ajani, I Orhorho, G Ogun, A Adeoye, C Okolo, A Oluwasola, and J Ogunbiyi. Brain weights in adult africans. *Journal of Morphological Sciences*, 34(4) :223–225, 2017.
- [103] HJ Scherer. Cerebral astrocytomas and their derivatives. *The American Journal of Cancer*, 40(2) :159–198, 1940.
- [104] Digital Science. Dimensions [software] available from <https://app.dimensions.ai>. accessed on (2020-11-13), under licence agreement.
- [105] Iram Shahzadi, Tong Boon Tang, Fabrice Meriadeau, and Abdul Quyyum. Cnn-lstm : Cascaded framework for brain tumour classification. In *2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, pages 633–637. IEEE, 2018.
- [106] Connor Shorten and Taghi M Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1) :60, 2019.
- [107] Herbert A Simon. Why should machines learn ? In *Machine learning*, pages 25–37. Elsevier, 1983.

- [108] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv :1409.1556*, 2014.
- [109] Karoline Skogen, Anselm Schulz, Johann Baptist Dormagen, Balaji Ganeshan, Eirik Helseth, and Andrès Server. Diagnostic performance of texture analysis on mri in grading cerebral gliomas. *European journal of radiology*, 85(4) :824–829, 2016.
- [110] Stephen M Smith. Fast robust automated brain extraction. *Human brain mapping*, 17(3) :143–155, 2002.
- [111] Marion Smits and Martin J van den Bent. Imaging correlates of adult glioma genotypes. *Radiology*, 284(2) :316–331, 2017.
- [112] Greg J Stanisiz, Ewa E Odrobina, Joseph Pun, Michael Escaravage, Simon J Graham, Michael J Bronskill, and R Mark Henkelman. T1, t2 relaxation and magnetization transfer in tissue at 3t. *Magnetic Resonance in Medicine : An Official Journal of the International Society for Magnetic Resonance in Medicine*, 54(3) :507–512, 2005.
- [113] Edward O Stejskal and John E Tanner. Spin diffusion measurements : spin echoes in the presence of a time-dependent field gradient. *The journal of chemical physics*, 42(1) :288–292, 1965.
- [114] Pan Sun, Defeng Wang, Vincent Ct Mok, and Lin Shi. Comparison of feature selection methods and machine learning classifiers for radiomics analysis in glioma grading. *IEEE Access*, 7 :102010–102020, 2019.
- [115] N Upadhyay and AD3473894 Waldman. Conventional mri evaluation of gliomas. *The British journal of radiology*, 84(special_issue_2) :S107–S111, 2011.
- [116] A Vamvakas, SC Williams, K Theodorou, E Kapsalaki, K Fountas, C Kappas, K Vassiou, and I Tsougos. Imaging biomarker analysis of advanced multiparametric mri for glioma grading. *Physica Medica*, 60 :188–198, 2019.

- [117] Martin J van den Bent. Interobserver variation of the histopathological diagnosis in clinical trials on glioma : a clinician’s perspective. *Acta neuropathologica*, 120(3) :297–304, 2010.
- [118] Joost JM Van Griethuysen, Andriy Fedorov, Chintan Parmar, Ahmed Hosny, Nicole Aucoin, Vivek Narayan, Regina GH Beets-Tan, Jean-Christophe Fillion-Robin, Steve Pieper, and Hugo JWL Aerts. Computational radiomics system to decode the radiographic phenotype. *Cancer research*, 77(21) :e104–e107, 2017.
- [119] Christopher S von Bartheld, Jami Bahney, and Suzana Herculano-Houzel. The search for true numbers of neurons and glial cells in the human brain : A review of 150 years of cell counting. *Journal of Comparative Neurology*, 524(18) :3865–3895, 2016.
- [120] John Von Neumann and Ray Kurzweil. *The computer and the brain*. Yale University Press, 2012.
- [121] Yang Yang, Lin-Feng Yan, Xin Zhang, Yu Han, Hai-Yan Nan, Yu-Chuan Hu, Bo Hu, Song-Lin Yan, Jin Zhang, Dong-Liang Cheng, et al. Glioma grading on conventional mr images : a deep learning study with transfer learning. *Frontiers in neuroscience*, 12 :804, 2018.
- [122] Ziv Yaniv, Bradley C Lowekamp, Hans J Johnson, and Richard Beare. Simpleitk image-analysis notebooks : a collaborative environment for education and reproducible research. *Journal of digital imaging*, 31(3) :290–303, 2018.
- [123] Zhiwei Zhang, Jingjing Xiao, Shandong Wu, Fajin Lv, Junwei Gong, Lin Jiang, Renqiang Yu, and Tianyou Luo. Deep convolutional radiomic features on diffusion tensor images for classification of glioma grades. *Journal of Digital Imaging*, pages 1–12, 2020.
- [124] Tongxue Zhou, Su Ruan, and Stéphane Canu. A review : Deep learning for medical image segmentation using multi-modality fusion. *Array*, 3 :100004, 2019.

- [125] Ying Zhuge, Holly Ning, Peter Mathen, Jason Y Cheng, Andra V Krauze, Kevin Camphausen, and Robert W Miller. Automated glioma grading on conventional mri images using deep convolutional neural networks. *Medical Physics*, 2020.
- [126] Alex Zwanenburg, Stefan Leger, Martin Vallières, Steffen Löck, et al. Image biomarker standardisation initiative-feature definitions. *arXiv preprint arXiv :1612.07003*, 2016.

Résumé

Les gliomes sont des tumeurs cérébrales dont l'évolution de bas grade à haut grade signe un diagnostic sombre pour le patient. Le grade du gliome est connu via des techniques invasives : analyse de pièce chirurgicale ou biopsie. Le CHU de Poitiers propose une alternative non-invasive via un bilan d'imagerie par résonance magnétique multimodal, regroupant des données anatomiques, de diffusion, de perfusion et de spectroscopie. Dans ce travail de thèse, nous proposons d'implémenter des outils de classification automatique depuis l'analyse de ces images multivariées. Nous déployons pour cela des outils novateurs permettant d'analyser le comportement de classifieurs automatiques. Cela nous amène à pointer des incohérences au sein de la base de données la plus utilisée sur la tâche de classification binaire du grade du gliome. Nous proposons une alternative via un consensus de cinq experts radiologues. Puis, nous créons une chaîne de traitement complète allant de l'acquisition à la classification, et explorons comment les données multimodales améliorent les performances de classification automatique.

Mots clefs : Gliome, IRM, Aide au diagnostic automatique, Exploration de données

Abstract

Gliomas are brain tumors whose evolution from low to high grade is a sign of a darkened outcome. The grade is assessed through invasive techniques such as or surgery or biopsy. The Poitiers University Hospital develops a non-invasive alternative through multimodal magnetic resonance imaging, including anatomical, diffusion, perfusion and spectroscopy data. In this thesis, we have proposed new tools to analyse an automatic classifier's behaviour. This help us underline inconsistencies within the most popular database used for glioma binary grade classification. We have proposed an alternative through a consensus of five expert radiologists. Then, we have designed a pipeline for automatic data processing, from the acquisition to the classification, and explore how multimodal data can enhance the classifier's performances.

Keywords : Glioma, MRI, Computed-aided diagnosis, Data exploration