



Analyse par intelligence artificielle d'images de tomodensitométrie de patients traumatisés crâniens afin de prédire leur évolution neurologique

Clément Brossard

► To cite this version:

Clément Brossard. Analyse par intelligence artificielle d'images de tomodensitométrie de patients traumatisés crâniens afin de prédire leur évolution neurologique. Imagerie médicale. Université Grenoble Alpes [2020-..], 2022. Français. $\langle$ NNT : 2022GRALS036 $\rangle$. $\langle$ tel-04086886 $\rangle$

HAL Id: tel-04086886

<https://theses.hal.science/tel-04086886v1>

Submitted on 2 May 2023

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.



HAL Authorization

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : ISCE - Ingénierie pour la Santé la Cognition et l'Environnement

Spécialité : BIS - Biotechnologie, instrumentation, signal et imagerie pour la biologie, la médecine et l'environnement

Unité de recherche : Grenoble Institut des Neurosciences

Analyse par intelligence artificielle d'images de tomodensitométrie de patients traumatisés crâniens afin de prédire leur évolution neurologique

Artificial intelligence analysis of cerebral scanner-X images of brain injured patients to predict their neurological evolution

Présentée par :

Clément BROSSARD

Direction de thèse :

Benjamin LEMASSON

Chargé de recherche HDR, INSERM Délégation Auvergne-Rhône-Alpes

Directeur de thèse

Pierre BOUZAT

Professeur des Univ. - Praticien hosp., Université Grenoble Alpes

Co-directeur de thèse

Rapporteurs :

HABIB BENALI

Professeur, Concordia University

CAROLE FRINDEL

Maître de conférences HDR, INSA LYON

Thèse soutenue publiquement le **25 novembre 2022**, devant le jury composé de :

BENJAMIN LEMASSON

Chargé de recherche HDR, INSERM DELEGATION AUVERGNE-RHÔNE-ALPES

Directeur de thèse

HABIB BENALI

Professeur, Concordia University

Rapporteur

CAROLE FRINDEL

Maître de conférences HDR, INSA LYON

Rapporteuse

PIERRE BOUZAT

Professeur des Univ. - Praticien hosp., UNIVERSITE GRENOBLE ALPES

Co-directeur de thèse

MARIA A. ZULUAGA

Maître de conférences HDR, EURECOM

Examinatrice

LIONEL VELLY

Professeur des Univ. - Praticien hosp., AIX-MARSEILLE UNIVERSITE

Président

MICHEL DESVIGNES

Professeur des Universités, GRENOBLE INP

Examineur

Invités :

JULES GREZE

Praticien Hospitalier, CHU Grenoble Alpes

Remerciements

Tout d'abord, merci aux membres de mon jury d'avoir accepté de lire et d'évaluer ce travail. Les remarques et conseils de chercheurs aussi reconnus que vous lui donnent toute sa valeur. J'ai beaucoup apprécié nos échanges lors de la soutenance.

Évidemment, merci à mon mentor et encadrant principal, Benjamin Lemasson, sans qui ce travail n'existerait simplement pas. Merci pour ta confiance, ton écoute, tes idées, et ton optimisme à toute épreuve, depuis mon stage de M1 jusqu'à mon post-doctorat, en passant par les manip précliniques et les relectures de ce manuscrit. J'ai énormément appris pendant ces 5 années dont je me souviendrai longtemps et j'espère pouvoir plus tard pouvoir encadrer un étudiant aussi bien que tu l'as fait avec moi. Ce fut un plaisir et une chance de travailler avec toi.

J'ai aussi eu la chance de pouvoir pendant ces 3 ans de thèse être au contact direct et régulier des médecins du CHUGA qui m'ont aidé et guidé dans ce travail. Merci à Pierre Bouzat pour ses conseils et à Jules Grèze pour sa disponibilité, son écoute attentive et ses idées de recherche constamment tournées vers l'application concrète au bénéfice des patients. Merci aussi aux ARC et aux étudiants médecin avec qui j'ai eu le plaisir de travailler. Vous avez donné à ce travail tout son sens et son utilité concrète.

Je tiens à remercier toute l'équipe 5 du GIN, menée par Emmanuel Barbier, pour son aide et son soutien. Ce fut un cadre serein et stimulant et qui a permis le bon déroulement de cette thèse, malgré la pandémie ! Merci également aux autres étudiants ou jeunes diplômés de l'équipe, aussi appelés « précaires de l'ESR », qui ont contribué à rendre ces années des plus agréables, autour d'un café, d'un rat, d'un bug MP3, d'un karaoké, ou d'empannages plus ou moins fructueux sur Virtual Regatta !

Merci aux amis qui m'ont nourri, protégé du Kraken, supporté pendant la #VenteOM, payé des Ricards, ou simplement qui ont pris des nouvelles régulièrement, en faisant fi de mes délais de réponses qui feraient pâlir l'administration française.

Finalement, merci infiniment à ma famille, qui vient tout juste de s'agrandir, pour son soutien inconditionnel et sa compagnie inestimable. J'ai hâte de vous retrouver à Lyon, la Posso, Paris, Villejuif, Cassis, en Espagne, et bientôt au Brésil !

Table des matières

Table des matières.....	1
Glossaire.....	5
Introduction.....	11
Contexte.....	11
Structure du manuscrit.....	12
1 Chapitre 1 : État de l'art	15
1.1 Tomodensitométrie à rayon-X	15
1.1.1 Rayons X et atténuation.....	15
1.1.2 Principe du scanner X et projections.....	17
1.1.3 Reconstruction des images TDM à partir des projections	19
1.1.4 Quantitativité de la TDM et ordres de grandeurs	21
1.1.5 Sources de variabilités introduites sur les images TDM et harmonisation	24
1.1.6 Artefacts communs sur les images TDM.....	25
1.1.7 Effets des rayonnements ionisants des rayons X	25
1.1.8 Conclusion	28
1.2 Traumatisme crânien.....	28
1.2.1 Définition	29
1.2.2 Épidémiologie du TC	29
1.2.3 Mesurer l'état de conscience d'un patient : Score de Glasgow	31
1.2.4 Résumer la prise en charge médicale d'un patient : score TILsum	32
1.2.5 Mesurer le devenir d'un patient TC à long terme : score GOS.....	33
1.2.6 Brève anatomie de l'encéphale	34
1.2.7 Lésions typiques du TC	37
1.2.8 Sensibilité des différentes lésions à la tomodensitométrie.....	40
1.2.9 Résumer une image TDM : Scores d'imagerie.....	42
1.3 Intelligence artificielle	44
1.3.1 Définition	44
1.3.2 Apprentissage supervisé et non supervisé.....	44
1.3.3 Écueil du surapprentissage et méthodologies d'entraînement.....	45
1.3.4 Exemple d'algorithme de machine learning : Forêts aléatoires.....	50
1.3.5 Exemple d'algorithmes d'IA : Apprentissage profond	51
1.4 Quantification d'images TDM de TC.....	57
1.4.1 Classification et extraction de métriques.....	58
1.4.2 Segmentation d'une image TDM.....	62
2 Chapitre 2 : Objectifs de la thèse.....	68
3 Chapitre 3 : 3 cohortes utilisées en 3 ans.....	72
3.1 Radiomic-TBI : 29 patients à 3 temps, 84 images TDM	72
3.1.1 Critères d'inclusion	72
3.1.2 Données recueillies et stockage.....	73
3.1.3 Utilisation	74
3.2 RadioxyTC : 12 patients, 12 images TDM.....	75
3.2.1 Critères d'inclusion	75
3.2.2 Données recueillies et stockage.....	75
3.2.3 Utilisation	75
3.3 CT-Harmo : 987 patients, 987 images TDM	76

3.3.1	Critères d'inclusion	76
3.3.2	Données recueillies et stockage.....	76
3.3.3	Utilisation	76
4	<i>Chapitre 4 : Segmentation automatique des lésions cérébrales</i>	78
4.1	Évaluation de BLAST-CT sur nos données.....	79
4.1.1	Objectif.....	79
4.1.2	Matériel et méthodes.....	79
4.1.3	Résultats.....	83
4.1.4	Interprétation.....	84
4.2	Amélioration de la segmentation automatique à 4 classes par « fine-tuning » 85	
4.2.1	Objectif.....	85
4.2.2	Matériel et méthodes.....	85
4.2.3	Résultats.....	89
4.3	Développement d'une segmentation automatique à 7 classes par « Transfer Learning »	92
4.3.1	Objectif.....	92
4.3.2	Matériel et méthodes.....	92
4.3.3	Résultats.....	94
4.4	Interprétation	97
4.5	Limitations	97
4.6	Perspectives.....	98
5	<i>Chapitre 5 : Prédiction de l'intensité thérapeutique à partir de l'image TDM initiale.....</i>	100
5.1	Objectifs.....	100
5.2	Matériel et méthodes	101
5.2.1	Critère à prédire : Une journée d'intensité thérapeutique extrême au cours des 7 premiers jours en réanimation	101
5.2.2	Données utilisées	102
5.2.3	Quantification du profil lésionnel observé en TDM à l'admission	104
5.2.4	Classification	107
5.3	Résultats.....	114
5.3.1	Évaluation du pouvoir prédictif des métriques sur le jeu de développement	114
5.3.2	Interprétation des modèles	115
5.3.3	Évaluation des modèles et des segmentations automatiques sur le jeu de test.....	116
5.4	Interprétation	117
6	<i>Chapitre 6 : Validation externe et mise à disposition clinique.....</i>	122
6.1	Validation externe de nos outils de segmentation et de classification	123
6.1.1	Matériel et méthodes.....	123
6.1.2	Résultats.....	126
6.1.3	Interprétation.....	132
6.2	Mise à disposition clinique des outils développés.....	134
6.2.1	Première approche : Génération d'un rapport d'imagerie par patient.....	134
6.2.2	Mise à disposition des codes informatiques sur Github	135
6.2.3	Plateforme VIP : Virtual Imaging Platform.....	136
6.3	Conclusion.....	140
7	<i>Chapitre 7 : Étude de l'évolution lésionnelle afin de prédire l'expansion œdémateuse.....</i>	144

7.1	Prédiction du futur œdème : Analyse Radiomique	144
7.1.1	Données utilisées	145
7.1.2	Méthodes	145
7.1.3	Résultats	148
7.1.4	Interprétation et perspectives	150
7.2	Prédiction empirique de la segmentation lésionnelle à J3	151
7.2.1	Données utilisées	152
7.2.2	Méthode	153
7.2.3	Résultats et interprétation	156
7.2.4	Conclusion et perspectives de l'étude	161
7.3	Futurs projets	161
7.3.1	Génération de la segmentation à J3 à partir des images TDM à J0 et J1	162
7.3.2	Génération de l'image TDM à J3 grâce aux Autoencoder/Réseaux récurrents	163
7.4	Conclusion	164
8	Chapitre 8 : Étude de la variabilité des images TDM sur une cohorte de soin courant	168
8.1	Objectifs	169
8.2	Protocole clinique : cohorte CT-Harmo	169
8.3	Recueil et stockage des données	170
8.4	Méthode – Encadrement de stage : Mise en évidence de la variabilité inter machine	171
8.4.1	Prétraitement des images	171
8.4.2	Exclusion des sujets pathologiques : Segmentation des lésions	171
8.4.3	Segmentation de la matière blanche, matière grise et du liquide céphalo-rachidien	172
8.4.4	Détection automatique des images artefactées	173
8.4.5	Représentation graphique et quantification de la variabilité	173
8.5	Résultats	174
8.5.1	Détection des images pathologiques	174
8.5.2	Détection des images artefactées	175
8.5.3	Étude de la variabilité inter-machines/protocoles	177
8.6	Interprétation	180
8.6.1	Détection d'images pathologiques et artefactées	180
8.6.2	Étude de la variabilité inter-machines/protocoles	181
9	Chapitre 9 : Communications et tâches annexes	184
9.1	Contributions scientifiques	185
9.2	Développement et publication du logiciel MP3 et formations données	185
9.3	Travaux Pratiques/ Travaux dirigés	188
9.4	Encadrements	188
9.5	Financements obtenus	189
9.6	Protocoles recherche clinique	189
	Conclusion et perspectives	192
	Bibliographie	198
	Annexes	208
	Évaluation de la segmentation automatique : Distance de Hausdorff moyenne équilibrée	208

Définitions	208
Evaluation sur le jeu de test de Radiomic-TBI.....	209
Evaluation sur RadioxyTC	211

Glossaire

2D : 2 dimensions

3D : 3 dimensions

Acc : Exactitude

AIS : Abbreviated Injury Scale

ANN : Réseau de neurones artificiels

ANOVA : Analyse de la Variance

ANTS : Advanced Normalization Tools

AUC : Aire sous la courbe

AVD : Distance de Hausdorff moyenne

bAVD : Distance de Hausdorff moyenne équilibrée

BLAST-CT : Algorithme de segmentation automatique de 4 types de lésions du TC sur des images TDM (Brain Lesion Analysis and Segmentation Tool for Computed Tomography)

Ca : Noyau Caudé

Ce : Cervelet

CEA : Centre pour l'Energie Atomique

CHU : Centre Universitaire Hospitalier

CHUGA : Centre Universitaire Hospitalier Grenoble Alpes

CNN : Réseau de Neurones Convolutifs

CRASH : Corticosteroid Randomization after Significant Head Injury

CT-TIQUA : Computed tomography based Traumatic brain Injury Quantification

CV : Validation croisée

DICOM : Digital imaging and communications in medicine

DRCI : Direction de la Recherche Clinique et de l'Innovation

DSC : Score de DICE

ECR : European Congress of Radiology

EEC : Espace Extra Cérébral

EIF : Extended Isolation Forest

ESR : European Society of Radiology

FIDLE : Formation Introduction au Deep Learning

FLIRT : FMRIB's Linear Image Registration Tool

FN : Faux Négatifs

FP : Faux Positifs

FSL : FMRIB Software Library

FT_4C : Modèle de segmentation à 4 classes obtenu après fine-tuning.

GBM : Gradient Boosting Method

GCS : Glasgow Coma Score

GIN : Grenoble Institut des Neurosciences

GIT : Global Information Tracker
GLCM : Gray Level Co-occurrence Matrix
GLDM : Gray Level Dependence Matrix
GLRLM : Gray Level Run Length Matrix
GLSZM : Gray Level Size Zone Matrix
GOS : Glasgow Outcome Scale
GOSE : Glasgow Outcome Scale Extended
GRICAD : Grenoble Alpe Research - Scientific Computing and Data Infrastructure
HAL : Archive ouverte pluridisciplinaire
Hb : Hémoglobine
HDH : Health Data Hub
HEA : Hémorragie Extra Axiale
HED : Hémorragie Extra Durale
HIP : Hémorragie Intra Parenchymateuse
HIV : Hémorragie Intra Ventriculaire
HSA : Hémorragie Sous Arachnoïdienne
HSD : Hémorragie Sous Durale
IA : Intelligence Artificielle
IBSI : Image Biomarker Standardisation Initiative
IMPACT : International Mission for Prognosis and Clinical Trials in Traumatic Brain Injury
INSERM : Institut National de la Santé et de la Recherche Médicale
IRM : Imagerie par Résonance Magnétique
Is : Insula
J0 : Jour 0 - Jour de l'accident initial
J1 : Jour 1 – Lendemain de l'accident
J3 : Jour 3 – 4 jours après l'accident
JSON : Javascript Object Notation
LCR : Liquide Céphalo Rachidien
LF : Lobe Frontal
LO : Lobe Occipital
LP : Lobe Parietal
LSTM : Long Short Term Memory
LT : Lobe Temporal
M2 : Master 2
MAE : Mean Absolute Error
MATLAB : Langage de programmation interprété
MB : Matière Blanche
MG : Matière Grise
MLR : Régression Logistique Multivariée
MONAI : Medical Open Network for Artificial Intelligence
MP3 : Medical software for Processing Multi-Parametric images Pipelines
NCV : Validation croisée imbriquée

ND : Non défini
NGTDM : Neighbouring Gray Tone Difference Matrix
NIFTI : Neuroimaging Informatic Technology Initiative
O : Oedème
OMS : Organisation Mondiale de la Santé
P : Pétéchies
PACS : Picture Archive and Communication System
PAM : Pression Arterielle Moyenne
PIC : Pression Intra Crânienne
PLS-DA : Partial Least Squares – Discriminant Analysis
PO : Perilesionnal Oedème
PPC : Pression de Perfusion Cérébrale
Pr : Précision
Pu : Putamen
Random4C : Modèle de segmentation à 4 classes obtenu après initialisation aléatoire
Random7C : Modèle de segmentation à 7 classes obtenu après initialisation aléatoire
RC : Reste du Cerveau
RF : Random Forest
RLM : Regression Logistique Multivariée
RNN : Réseau de neurones récurrents
ROC : Receiver Operating Characteristic
ROI : Région d'Intérêt
RSB : Rapport Signal sur Bruit
Se : Sensibilité
Shanoir : Plateforme de stockage sécurisée d'images médicales
SN : Système nerveux
SNC : Système nerveux central
SNP : Système nerveux périphérique
Sp : Spécificité
SSIM : Mesure de similarité entre deux images
TBI : Traumatisme Crânien
TC : Traumatisme crânien
TDM : Tomodensitométrie / Tomodensitométrie
TEP : Tomographie par Emission de Positrons
TF : Transformée de Fourier
Th : Thalamus
TILsum : Therapeutic Intensity Level : Niveau d'intensité thérapeutique.
TILsum_bas : Se dit d'un patient qui atteint un TILsum supérieur ou égal à 11 au moins d'un des 8 premier jours en réanimation.
TILsum_haut : Se dit d'un patient qui n'atteint pas un TILsum supérieur ou égal à 11 au moins d'un des 8 premier jours en réanimation.
TL_7C : Modèle de segmentation à 7 classes obtenu après Transfer learning à partir du modèle BLAST-CT

TL_RE_7C : Modèle de segmentation à 7 classes obtenu après Transfer learning à partir du modèle FT_4C

TN : Vrais Négatifs

TP : Vrais Positifs

UGA : Université Grenoble Alpes

UH : Unités Hounsfield

UNSCEAR : Comité scientifique des Nations unies pour l'étude des effets des rayonnements ionisants

VIP : Virtual Imaging Platform

VPN : Valeur Prédictive Négative

Introduction

Contexte

Le traumatisme crânien (TC) est une pathologie grave et fréquente [1]. Il est estimé qu'une personne sur deux sera victime d'un TC au moins une fois dans sa vie dans le monde. De par ses causes, principalement les chutes et les accidents de la voie publique, le TC touche principalement les jeunes adultes [2], ce qui engendre un coût humain et financier important pour la société [3]. L'agression initiale entraîne des lésions cérébrales primaires inévitables, focales ou diffuses, de différentes natures, qui cohabitent au sein du cerveau et causent des lésions cérébrales secondaires, toutes aussi diverses, qui sans traitement engendrent des handicaps physiques, cognitifs, ou même la mort. Tout l'enjeu de la prise en charge du TC en réanimation réside dans la prévention et la prise en charge de ces lésions cérébrales secondaires. Le diagnostic de sévérité et l'évaluation pronostique des TC reposent aujourd'hui sur le score de Glasgow [4], introduit en 1974, qui évalue l'état de conscience du patient après l'accident, sur l'anomalie pupillaire, signe d'un engagement cérébral ou d'une lésion du tronc cérébral, et sur le suivi des lésions par imagerie médicale. Parmi les différentes modalités d'imagerie, la plus utilisée pour les patients TC, de par sa rapidité, son coût, et son accessibilité, est la tomodensitométrie (TDM) à rayons X, qui mesure le coefficient d'atténuation des tissus du cerveau, intrinsèquement reliée à la densité et la nature des tissus. Les patients sont donc imagés régulièrement par TDM durant leurs premiers jours d'hospitalisation. Si les images TDM de patients TC sont parfois quantifiées de manière grossière au travers de scores d'imagerie, qui résument l'état cérébral général à partir d'observations visuelles, ces outils restent imprécis, variables, et fastidieux à utiliser, ce qui les rend concrètement inemployés en routine clinique. De plus, ces analyses conduisent à un diagnostic qualitatif et non reproductible des lésions cérébrales et de leur évolution, ce qui, à l'heure de la médecine personnalisée et de l'intelligence artificielle, peut paraître anachronique.

Dans la littérature scientifique, il a été montré que ces scores d'imagerie ont un pouvoir prédictif non négligeable du devenir des patients, et sont corrélés au taux de mortalité ou de handicap [5], ce qui prouve que les images TDM contiennent une information capable de prédire l'état des patients à long terme. Des études récentes ont cherché à s'affranchir de ces scores d'imagerie pour exploiter directement les images TDM, à des fins diagnostiques [6] ou même pronostiques [7], mais à de rares exceptions [8], ces travaux portent sur la prédiction à long terme de l'état des patients. De tels critères de prédictions, à 3 ou 6 mois après le TC, ont un intérêt évident pour les patients et les familles, mais n'ont qu'une utilité limitée pour les cliniciens responsables de la prise en charge, qui aimeraient plutôt estimer à court

terme l'évolution lésionnelle des patients, de manière à intervenir de manière plus précoce et ainsi adapter la prise en charge thérapeutique.

Nous faisons l'hypothèse que les images TDM contiennent une information capable de distinguer de manière précoce les patients qui nécessitent un niveau de prise en charge thérapeutique extrême. Cette thèse exploratoire, débutée en octobre 2019, a tenté d'exploiter ces images TDM afin de les quantifier de manière automatique, via des outils d'intelligence artificielle, et de les associer à des critères cliniques pertinents pour les cliniciens. Plusieurs pistes ont été empruntées et plusieurs études de recherches, plus ou moins fécondes, ont été menées. Les principaux développements de ces 3 années de doctorat sont développés dans ce manuscrit.

Structure du manuscrit

Le manuscrit s'organise comme une suite des différentes études de recherche menées au cours de ces 3 ans de thèse. Chaque chapitre a été écrit, dans la limite du possible, de manière à être le plus indépendant possible des autres, ce qui permet au lecteur de naviguer dans le manuscrit à sa guise. Les différents chapitres qui composent ce manuscrit sont :

- Chapitre 1 : Ce chapitre définit et détaille les termes et notions du sujet de thèse. Il aborde le fonctionnement du scanner à rayon X, le traumatisme crânien, ou encore l'apprentissage statistique. Ce chapitre se conclut par un état de l'art des études utilisant des algorithmes d'apprentissage statistique à partir d'images tomographiques à rayons X cérébrales de patients traumatisés crâniens.
- Chapitre 2 : Ce chapitre détaille les objectifs initiaux de la thèse, leur positionnement par rapport aux besoins des cliniciens et à l'état de l'art.
- Chapitre 3 : Ce chapitre détaille les différentes cohortes de patients utilisées au cours des études présentées dans ce manuscrit. En particulier, on y trouvera les critères d'inclusions, le numéro d'enregistrement et les particularités de chaque cohorte.
- Chapitre 4 : Ce chapitre expose la première étude de ce manuscrit, qui a abouti à un outil de segmentation automatique des lésions typiques du TC à partir d'images TDM. Nous avons appliqué et comparé des méthodes d'apprentissage par transfert afin d'améliorer un réseau de neurones pour segmenter 7 différents types de lésions du TC.
- Chapitre 5 : Ce chapitre développe notre étude de prédiction automatique du niveau d'intensité thérapeutique dont un patient traumatisé crânien aura

besoin au cours des 8 premiers jours en réanimation à partir de l'image TDM acquise à son arrivée à l'hôpital. Nous avons pour cela extrait le volume et la localisation des différents types de lésions de l'image TDM et l'avons lié à cette intensité thérapeutique.

- Chapitre 6 : Ce chapitre a pour but de valider les outils de segmentation et de classification développés aux chapitres 4 et 5 sur une nouvelle cohorte. Ce chapitre présente également les moyens mis en œuvre pour mettre ces outils à disposition des cliniciens.
- Chapitre 7 : Dans ce chapitre nous développons deux études en cours qui utilisent les examens TDM répétés des patients afin d'exploiter l'évolution scanographique des lésions pour prédire les lésions œdémateuses futures.
- Chapitre 8 : Ce chapitre présente nos travaux portant sur l'étude de la variabilité introduite sur les images TDM par les différents protocoles d'acquisition des images. Cette étude a fait l'objet d'un stage de dernière année de Master que j'ai co-encadré.
- Chapitre 9 : Finalement, puisque la vie d'un chercheur académique n'est pas uniquement faite d'études scientifiques, j'ai listé dans ce dernier chapitre les différentes tâches extra scientifiques de la vie d'un chercheur que j'ai été amené à remplir. Vous y trouverez notamment des détails sur les demandes de financement et les protocoles de recherche clinique que j'ai pu écrire.

Bonne lecture !



Figure 1. Une image générée automatiquement par l'algorithme MidJourney [<https://www.midjourney.com/>] à partir de la phrase « Artificial intelligence analysis of the computed tomography evolution of brain injured patients to predict their neurological evolution », qui s'avère être une traduction de mon sujet de thèse. Une autre image du même type est disponible après la conclusion.

Chapitre 1 : État de l'art

Ce chapitre a pour but d'introduire et de définir les principaux domaines d'étude de cette thèse. Nous passerons donc en revue l'imagerie par tomodensitométrie (TDM), le traumatisme crânien, et l'intelligence artificielle, avant de s'intéresser à l'état de l'art des méthodes de quantification des images TDM de patients traumatisés crâniens. Le scientifique rompu à ces domaines peut aisément passer ce chapitre entier ou bien certains sous-chapitres.

1.1 Tomodensitométrie à rayon-X

La tomodensitométrie à rayon X (TDM), aussi appelée « scanner X » est une technique d'imagerie médicale qui permet de mesurer le coefficient d'atténuation aux rayons X de la matière imagée. Parmi les différentes modalités d'imagerie médicale (Imagerie par Résonance Magnétique (IRM), tomographie par Émission de Positons (TEP), etc.), la tomodensitométrie à rayons X (TDM), ou Computed Tomography en anglais (CT) est la plus utilisée dans le cas du Traumatisme crânien (TC) de par son accessibilité, son coût, et sa rapidité. Cette partie présente les rayons X, le fonctionnement d'un scanner, la manière d'obtenir une image TDM, les variabilités et les artefacts dont peuvent être victimes les images TDM, et finalement l'impact des rayonnements ionisants sur la santé.

1.1.1 Rayons X et atténuation

Les rayons X sont des rayonnements électromagnétiques de longueur d'onde comprise entre 10^{-11} m et 10^{-8} m, donc de fréquence comprise entre 3.10^{16} Hz et 3.10^{19} Hz. Les rayons X se situent donc entre les ultraviolets et les rayons gamma sur le spectre des rayonnements électromagnétiques Figure 2.

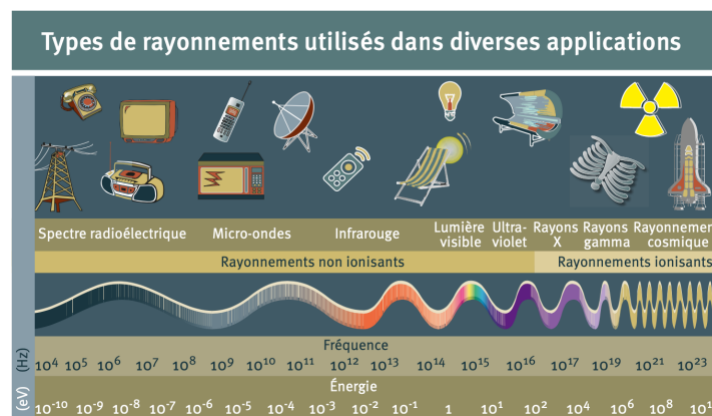


Figure 2. Spectre des différents rayonnements électromagnétiques en fonction de leur fréquence et applications courantes. Source : UNSCEAR [9]

Leur propriété principale est qu'ils traversent aisément les matières peu denses et constituées d'éléments légers, mais sont arrêtés par les matières denses et constituées d'éléments lourds, comme l'acier.

Tout rayon X traversant une matière est en partie absorbé et dévié, donc atténué, en fonction de la nature et de la densité de la matière traversée.

Dans le cas simplifié d'un rayon X monoénergétique, non réfléchi et non dévié, l'intensité I (en électronvolts - eV) du rayon d'intensité initiale I_0 (en eV) traversant un matériau de coefficient d'atténuation linéaire μ (m^{-1}) sur une épaisseur x (m), est donnée par la formule :

$$I = I_0 e^{-\mu x}$$

Le coefficient d'atténuation linéaire μ , exprimé en m^{-1} , résulte de différents types d'interactions entre les atomes et les photons, qui sont les particules constituant le rayon X, parmi lesquelles l'effet photoélectrique, l'effet Compton ou encore la création de paires. En fonction des atomes constitutifs de la matière traversée et de l'énergie du rayon incident, certains effets prennent le pas sur les autres. Dans le cadre de l'imagerie médicale, les rayons X utilisés ont une énergie généralement inférieure à 150 keV. De plus, les principaux atomes constituant le corps humain (Oxygène - 65%, Carbone - 18%, Hydrogène - 10%, Azote - 3%, Calcium - 1,5%, Phosphore - 1%) sont légers, avec un numéro atomique inférieur à 25 (numéros atomiques respectifs : 8, 6, 1, 7, 20, 15). Dans ces conditions, les deux effets dominants influant sur le coefficient d'atténuation sont l'effet photoélectrique μ_p et l'effet Compton μ_c , comme illustré sur la Figure 3, et le coefficient d'atténuation se rapproche de la somme de ces deux effets. Les phénomènes physiques en jeu dans ces deux effets ne seront pas abordés ici.

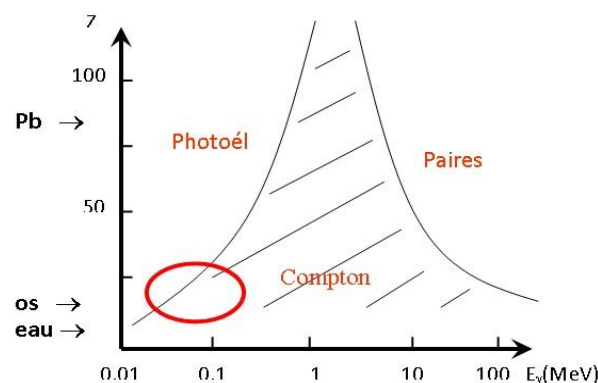


Figure 3. Illustration de la prédominance des interactions entre les atomes constitutifs de la matière à imager et les photons constitutifs des rayons X incidents en fonction de la nature des atomes et de l'énergie des rayons. La zone d'usage en radiologie est délimitée par l'ellipse rouge. Source :[10]

1.1.2 Principe du scanner X et projections

Un scanner dans sa version la plus simple est principalement constitué d'une table où l'on peut disposer la matière à imager, d'un anneau autour de la table où sont disposés un tube à rayons X, aussi appelé « émetteur », qui émet des rayons en éventail vers le centre de l'anneau, et des détecteurs qui mesurent l'intensité des rayons après avoir traversé la matière d'intérêt. Puisque les rayons sont envoyés en éventail dans le plan de l'anneau, plusieurs centaines de détecteurs couvrent un important arc de cercle de l'anneau, diamétralement opposé à l'émetteur, et ainsi captent un grand nombre de rayons ayant traversé la matière à imager. La détection d'un rayon émis avec un angle θ par rapport à la verticale est appelée une projection selon l'angle θ . L'anneau tourne sur lui-même pour pouvoir tirer des rayons X dans toutes les directions. Un demi-tour du couple émetteur/détecteur permet ainsi l'acquisition d'une coupe 2D de la matière à imager. Finalement, la table peut se déplacer sur l'axe central de l'anneau pour imager une autre tranche de la matière, ce qui permet au final d'obtenir une image 3D de la matière d'intérêt. Pour information, les scanners les plus récents sont appelés « scanner multibarrettes » car ils disposent de plusieurs couches de détecteurs qui leur permettent d'acquérir plusieurs coupes 2D à chaque demi-tour de l'émetteur.

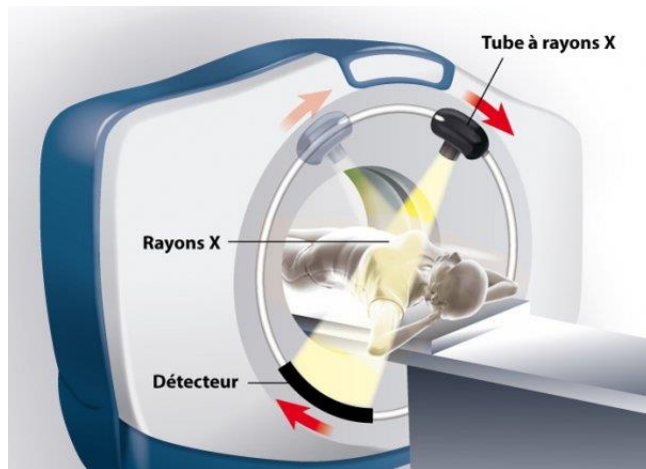


Figure 4. Schéma d'un scanner. Les détecteurs sont diamétralement opposés à l'émetteur sur l'anneau qui tourne autour de la matière à imager elle-même placée sur une table qui se translate sur un axe perpendiculaire à l'anneau. Source : Ramsay Santé – Clinique du Landy [11]

Dans le cas d'un scanner classique (monobarette), si l'on considère une table fixe, et l'émission de rayons X à 180 angles différents (avec un pas d'émission de 1 degré) autour de la matière à imager, les détecteurs captent 180 projections, comme illustré Figure 5.

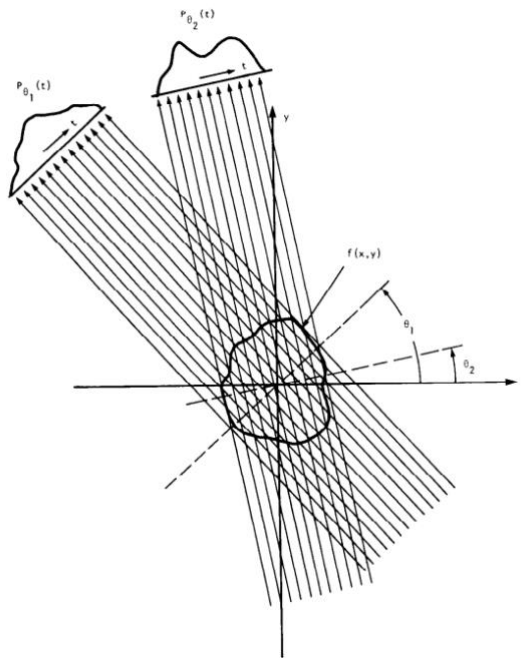


Figure 5. Schéma de deux projections selon deux angles. On fait l'hypothèse que les rayons arrivent parallèlement entre eux sur la matière à imager. Source : Thèse de Max Langer [12]

On peut rassembler les projections selon les 180 angles d'émission dans une figure appelée sinogramme, comme illustré Figure 6.

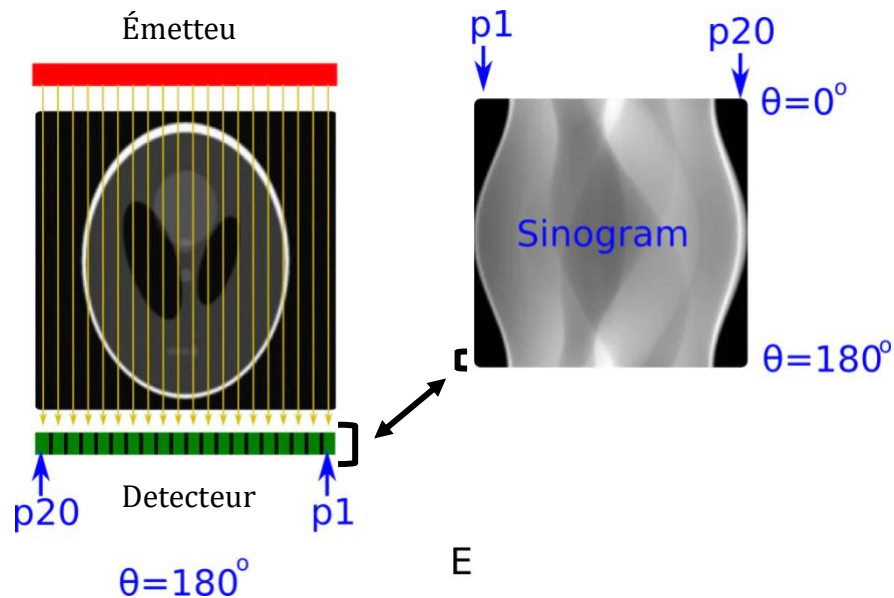


Figure 6. A gauche : Projection selon l'angle $\theta=180^\circ$ captée par 20 détecteurs. A droite, sinogramme représentant les valeurs mesurées sur les 20 détecteurs en fonction des 180 angles de projection. Source (image adaptée) : The Conversation ([13])

1.1.3 Reconstruction des images TDM à partir des projections

A partir d'un sinogramme, il est possible de reconstruire une image de la matière traversée par les rayons X. En 1917, le mathématicien autrichien Johann Radon a introduit la transformée de Radon, qui relie les projections $p(\theta, u)$ de l'espace du sinogramme à l'image reconstruite $f(x, y)$ selon la formule :

$$p(\theta, u) = \int_{-\infty}^{+\infty} f(x, y) dv$$

Avec u et v la base du référentiel du sinogramme illustrée Figure 7 et définie comme :

$$u = x \cos \theta + y \sin \theta$$

$$v = -x \sin \theta + y \cos \theta$$

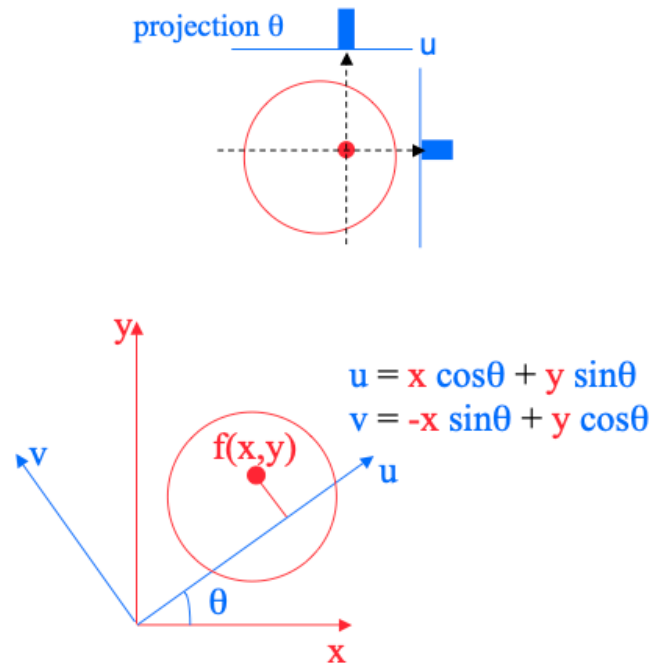


Figure 7. Illustration du changement de base permettant de passer du référentiel de la table du scanner à celui du sinogramme résultant d'un ensemble de projections. Source : Irene Buvat [14]

Ainsi, grâce à cette formule, on peut obtenir une image représentative du coefficient d'atténuation de la matière grâce notamment à la transformée de Radon inverse du sinogramme regroupant les projections des rayons X selon les différentes directions par rapport à la verticale :

$$f(x, y) = \int_0^{180} p(\theta, x \cos \theta + y \sin \theta) d\theta$$

Cette méthode est appelée rétroprojection, ou « backprojection ».

En pratique, pour un nombre fini de projections, il n'existe pas de solution unique, car il existe toujours plusieurs objets de taille ou coefficient d'atténuation différents capables de reproduire le sinogramme d'un nombre fini de projections. Ainsi, dans le cas réel d'un nombre fini de projections, la méthode de rétroprojection conduit à des images floues et avec peu de contraste, comme illustré sur la Figure 8.

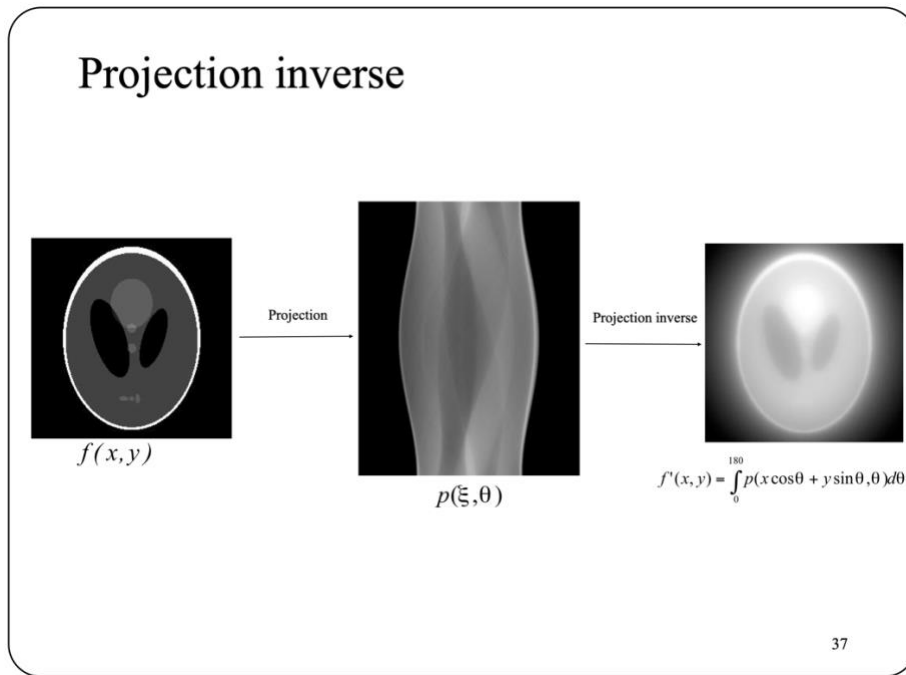


Figure 8. Illustration de la méthode de reconstruction d'une image par rétroprojection. Source [15]

Pour réduire ce bruit qui perturbe la reconstruction, on peut filtrer le sinogramme avant d'appliquer la rétroprojection et ainsi effectuer ce que l'on appelle une filtered backprojection. Plusieurs types de filtres peuvent être utilisés, comme un filtre de Hann, de Butterworth, ou bien un filtre gaussien, chacun avec ses avantages et inconvénients.

Une autre méthode de reconstruction qui consiste à utiliser la transformée de Fourier et le théorème de la tranche centrale peut être utilisée. Pour cela, il suffit de calculer la transformée de Fourier de la projection selon chaque angle, et d'appliquer la transformée de Fourier 2D inverse sur le sinogramme résultant exprimé en coordonnées polaires afin d'obtenir une approximation de l'image d'intérêt, comme illustré sur la Figure 9.

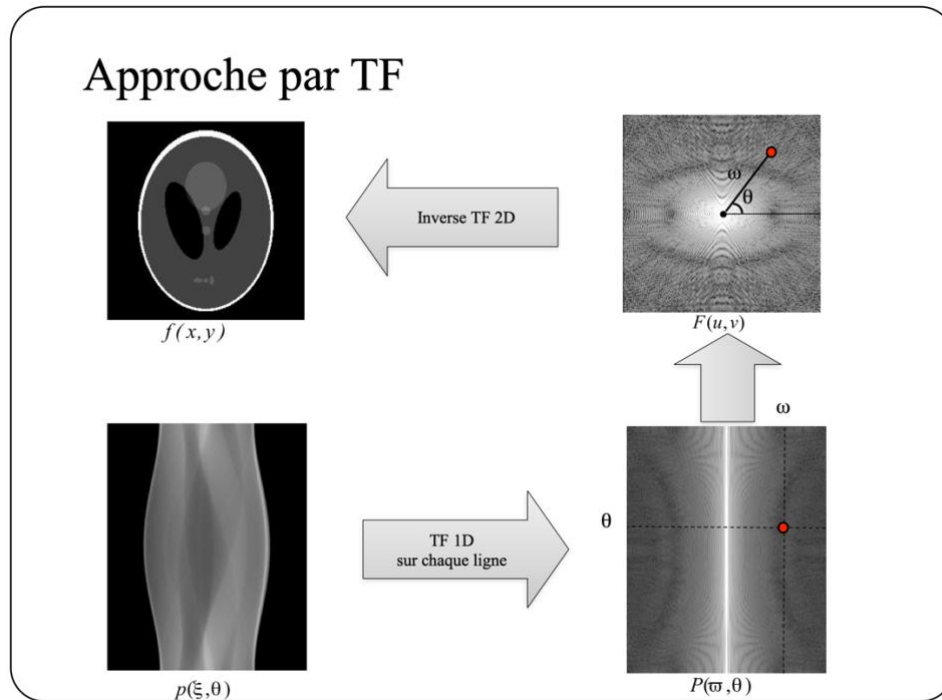


Figure 9. Illustration de la méthode de reconstruction d'une image par transformée de Fourier. Source [15]

Cette méthode fournit des images bien plus précises que la rétroprojection, et est également bien plus performante sur des reconstructions à partir d'un faible nombre d'angles de projection.

Les constructeurs de scanners travaillent constamment sur de nouvelles méthodes de reconstruction, afin d'obtenir des images plus nettes et plus résolues pour limiter au maximum le temps d'acquisition et donc de la dose de rayons X absorbée par le patient.

1.1.4 Quantitativité de la TDM et ordres de grandeurs

Les scanners sont régulièrement calibrés afin d'exprimer le coefficient d'atténuation de la matière imagée en unités Hounsfield (UH). L'échelle de Hounsfield a été introduite par Godfrey Newbold Hounsfield (Figure 10) et spécifie par définition que :

- Le coefficient d'atténuation de l'eau est nul
- Le coefficient d'atténuation de l'air est égal à -1000 UH

Ainsi, le coefficient d'atténuation en unités Hounsfield UH_x d'une matière de coefficient d'atténuation μ_x est exprimé en unités Hounsfield (UH) :

$$UH_x = 1000 \times \frac{\mu_x - \mu_{eau}}{\mu_{eau}}$$



Figure 10. Godfrey Hounsfield (1919-2004), lauréat avec Allan McLeod Cormack du prix Nobel de physiologie ou médecine en 1979. Source : Wikipedia [16]

L'échelle de Hounsfield est quantitative. Les différents éléments du corps humain ont ainsi pour coefficient d'atténuation une plage de valeurs bien spécifique, comme décrit dans le Tableau 1 :

Matériau	Unités Hounsfield
Air	-1000
Poumon	-500
Graisse	-100 à -50
Eau	0
Liquide Cérébro-spinal	15
Muscle	10 à 40
Matériau Blanche	20 à 30
Rein	30
Sang	30 à 45
Matériau Grise	37 à 45
Foie	40 à 60
Tissus mous	100 à 300
Os	700 (os spongieux) à 3000 (os dense)

Tableau 1. Valeurs caractéristiques du coefficient d'atténuation des tissus physiologiques en unités Hounsfield. Source : Wikipedia[17]

Comparé à l'imagerie par résonnance magnétique (IRM), autre technique d'imagerie médicale très utilisée en routine clinique, la tomodensitométrie (TDM) fournit des images certes quantitatives, mais avec un contraste moindre entre les différents tissus mous du corps humain. Le principal contraste de la TDM est entre les zones très denses comme les os et les tissus mous, ce qui permet d'identifier rapidement les fractures ou encore les objets métalliques étrangers.

Ce faible contraste, notamment au niveau cérébral, est compensé par une rapidité d'acquisition supérieure (quelques minutes seulement) et un coût plus faible que l'IRM. De plus, contrairement à l'IRM, les patients lourdement assistés peuvent plus facilement et rapidement être imagés, ce qui en fait la technique d'imagerie de choix pour les patients sévèrement traumatisés crâniens.

Le Tableau 2 regroupe et compare les principales caractéristiques d'un examen classique en TDM et en IRM

Modalité	TDM	IRM (3 Teslas, séquence anatomique)
Temps d'acquisition d'une image	2 minutes	5 minutes (Dépend des séquences)
Type et durée d'une session	Mono-paramétrique 10 minutes	Multiparamétrique 30 minutes
Résolution spatiale	0,5 mm	1 mm (Dépend des séquences)
Energie	100 keV	-
Coût	1 million €	3 millions €
Contraste	Élevé sur les os, faible sur les tissus mous	Élevé sur les tissus mous, faible sur les os
Invasif ?	Faibles rayonnements ionisants	Non
Quantitatif ?	Oui	Seulement après traitement mathématique
Contraintes d'utilisation	Déconseillé pour les femmes enceintes / jeunes enfants	Sensible aux matériaux ferromagnétiques donc non utilisable avec un pacemaker.
Poids d'une image 3D non compressée	100 Mo	10 Mo

Tableau 2. Tableau comparatif des principales grandeurs d'un examen classique TDM et IRM. Sources : [18] [19]

1.1.5 Sources de variabilités introduites sur les images TDM et harmonisation

Si les images TDM sont en effet quantitatives, elles dépendent bien évidemment de paramètres d'acquisition, comme par exemple la dose en rayons X envoyée. Si elle est trop faible, un très petit nombre de photon traversera la matière à imager et rendra l'image bruitée et sensible aux artefacts. Plusieurs autres paramètres jouent sur la qualité de l'image finale, comme la résolution de l'image, l'algorithme de reconstruction utilisé, ou encore les paramètres physiques du scanner utilisé, comme son nombre de détecteurs par exemple. L'influence de ces paramètres sur les images résultantes est nommée variabilité. La variabilité peut être inter-machine, inter-protocoles d'acquisition, ou encore temporelle, si deux images acquises à deux instants différents ne sont pas identiques. Ces trois types de variabilité sont détaillées dans cette partie.

Comme tout juste mentionné, les différents modèles de scanners n'ont pas forcément le même nombre ou les mêmes types de détecteurs, ou encore les mêmes composants électroniques. Mackin & al. [20] ont mesuré pour différents scanners la variation des valeurs UH pour différentes matières présentes dans un fantôme et ont remarqué que pour un même constructeur, une différence de modèle de scanners peut introduire une variation d'environ 10 unités Hounsfield (UH). Sachant que les tissus du cerveau ont globalement une absorption comprise entre 0 et 100 UH, une potentielle variation de 10% est loin d'être négligeable.

Les méthodes d'acquisitions des scanners utilisées dans les CHU sont optimisées pour les besoins du CHU ou des médecins en charge et ces différences impliquent elles aussi des modifications sur les images de sortie. Une étude de Lu & al. publiée en 2016 [21] montre que les paramètres d'acquisition, tels que la résolution ou encore les méthodes de reconstruction de l'image finale, ont une influence sur les propriétés de l'image. Quand certaines propriétés ne sont que peu influencées, comme la forme de la matière, d'autres sont entièrement modifiées comme par exemple la texture de l'image.

Concernant la variabilité temporelle des images, qui pourrait intervenir avec les mêmes paramètres d'acquisition et la même machine lors de 2 images acquises à plusieurs jours d'intervalle, elle peut être évaluée à environ 1 UH sur la moyenne des valeurs des voxels dans une ROI homogène [22] [23], ce qui est faible par rapport aux autres sources de variabilité précédemment citées.

Ces différentes sources de variabilité n'ont généralement que peu d'influence sur les diagnostics à l'œil nu émis par les radiologues. Néanmoins, ils peuvent

biaiser les algorithmes automatiques de quantification des images TDM. C'est pourquoi plusieurs méthodes d'harmonisation ont vu le jour, comme par exemple la méthode « Combat » [24][25], DualGAN [26] ou encore DeepHarmony [27]. Plus récemment, un doctorant de notre équipe a développé ImUnity, un algorithme d'harmonisation des images médicales qui semble prometteur [28].

1.1.6 Artefacts communs sur les images TDM

L'analyse des images TDM est régulièrement perturbée par des artefacts qui peuvent être dus à un défaut du matériel ou à des propriétés physiques de l'objet imagé. Aux dires des cliniciens, l'immense majorité des images TDM acquises au CHU sont artefactées, même si l'artefact ne touche pas toujours la zone d'intérêt sur laquelle se concentrent les radiologues. Ainsi, les images acquises en soin courant sont différentes de celles utilisées pour les protocoles de recherche, où les images sont plus propres et en général moins artefactées. Cette différence se ressent parfois sur les performances des outils de traitement automatique des images TDM, qui fonctionnent moins bien sur les données de soin courant que sur les cohortes des études de recherche.

Une étude de Boas & Fleischmann [29] recense et illustre les différents artefacts que l'on peut retrouver sur les images TDM et leurs causes. Les principaux artefacts sont liés à des objets de forte atténuation, comme par exemple du métal, qui peut être dû à une prothèse dentaire ou à une sonde quelconque, et qui créent sur l'image des stries de haute et de basses intensités autour de l'objet de forte atténuation. Une trop faible dose en rayons X peut également créer ou amplifier des artefacts sur les images, ainsi que réduire le rapport Signal sur Bruit (RSB) de l'image, dû au plus faible nombre de photons qui traversent la matière et sont mesurés par les détecteurs.

Ainsi, l'énergie des rayons X utilisés en imagerie médicale est généralement comprise entre 20 et 150 keV. Une énergie plus faible rend l'image inutilisable car trop sujette aux artefacts et une énergie plus haute abaisse le contraste entre les tissus mous et les os, qui deviennent donc plus « transparents ». De plus, une forte énergie implique une grande quantité de rayonnement ionisants, qui sont dangereux pour les patients.

1.1.7 Effets des rayonnements ionisants des rayons

X

Comme mentionné précédemment, lorsqu'un rayon X est envoyé sur une matière, il est atténué selon la matière et la densité de celle-ci. Cette atténuation est due en partie à une réflexion du rayon, mais principalement à une absorption du rayon par la matière. Cette absorption de rayons X, lorsqu'elle concerne une matière vivante, cause des dommages aux cellules traversées. Si l'ampleur des dommages est d'autant plus forte que l'exposition aux rayons X est prolongée et que l'énergie des rayons incidents est élevée, elle dépend également de la nature des tissus touchés. Ainsi on distingue la dose absorbée, mesurée en Grays (Gy), qui est commune à tous

les tissus, de la dose efficace, mesurée en Sieverts (Sv) et qui elle tient compte de la sensibilité aux rayonnements ionisants des tissus exposés. Les calculs des doses absorbée et efficace sont complexes et ne seront pas abordés dans ce manuscrit de thèse.

Les rayonnements ionisants sont émis par des sources naturelles ou artificielles, à des doses et durées d'expositions différentes. La Figure 11 illustre et compare les doses efficaces de différents phénomènes, naturels ou artificiels.

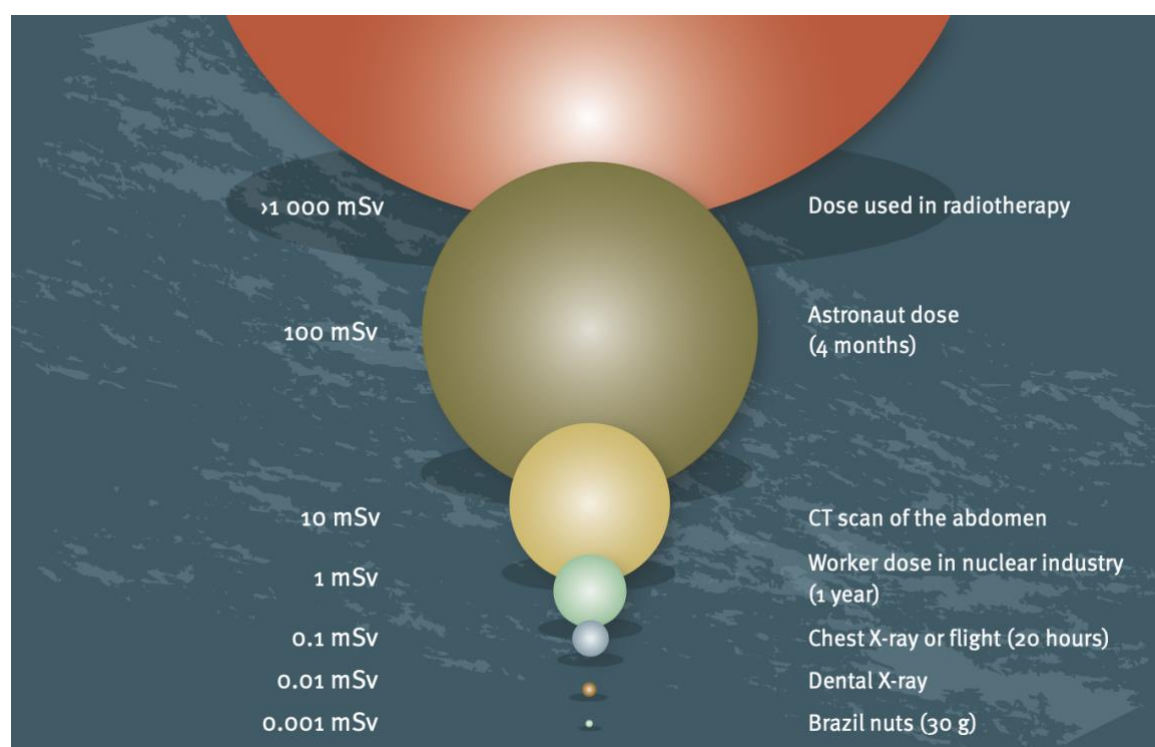


Figure 11. Doses efficaces de différents phénomènes, naturels ou artificiels, émetteurs de rayonnement ionisants.
Source : UNSCEAR [9]

A faible dose, les effets sont stochastiques, c'est-à-dire qu'ils se mesurent par une hausse de la probabilité d'événements anormaux à long terme. A forte dose, les effets sont déterministes, c'est-à-dire qu'ils sont cliniquement observables quasiment immédiatement.

1.1.7.1 Effets stochastiques

Les effets stochastiques ne sont mesurables qu'à l'échelle d'une population. Ils se déclarent à long terme et la dose reçue, tant qu'elle reste faible pour rester dans le domaine des effets stochastiques, n'a pas d'influence sur leur gravité mais sur leur probabilité d'apparition. Les principaux effets stochastiques de l'exposition aux rayons X sont les cancers. Comme illustré sur la Figure 12, le niveau de dose efficace le plus faible détecté à ce jour entraînant une hausse avérée du risque cancérogène est

de 100 mSv. A titre de comparaison, un examen tomodensitométrique cérébral délivre en moyenne 1,5 mSv.

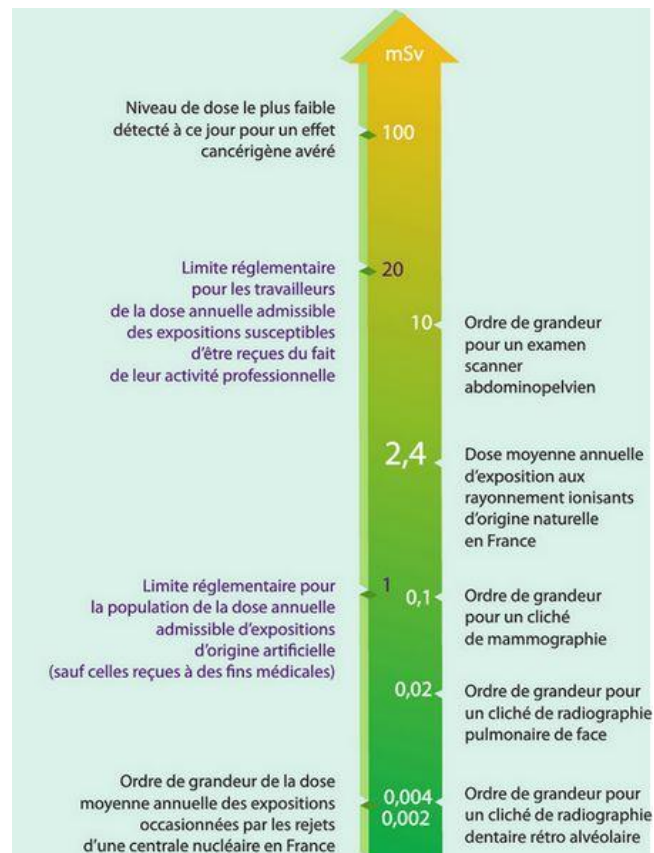


Figure 12. Échelle représentant la dose efficace délivrée par différents phénomènes émettant des rayons ionisants.
Source : IRSN ([30])

1.1.7.2 Effets déterministes

Les effets déterministes sont certains et quasiment immédiats. Ils apparaissent pour des doses absorbées élevées, à partir de 0,25 Gy. Les effets déterministes peuvent aller de rougeurs ou lésions cutanées, à la mort par atteinte du système nerveux, en passant par la stérilité définitive. Comme illustré sur la Figure 13, une dose de plus de 2,25 Gy, entraîne le décès d'une personne sur deux en l'absence de traitement [31].

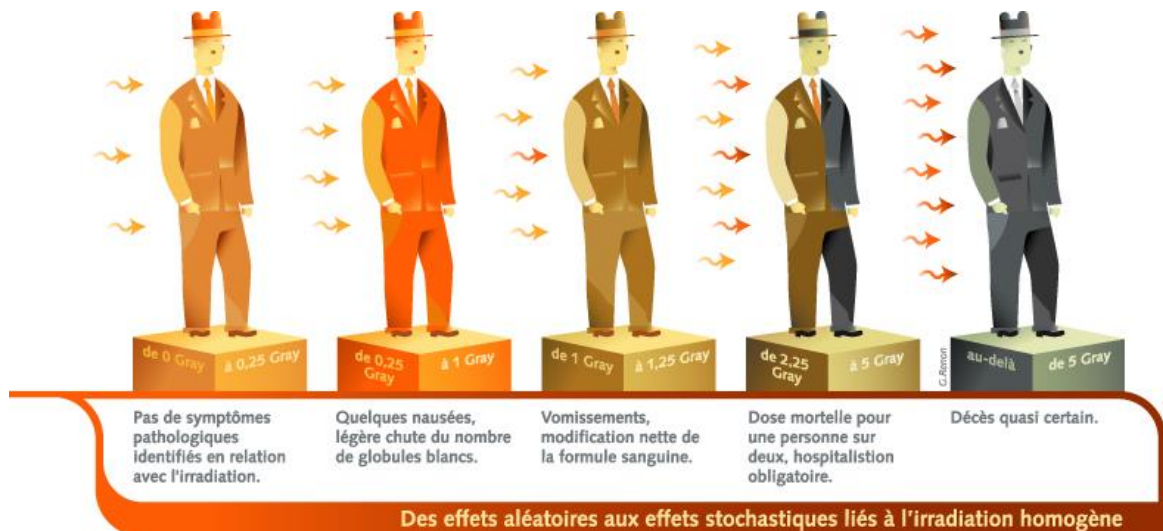


Figure 13. Effets sur l'organisme en fonction des doses pour une irradiation globale (corps entier). ©CEA/Guillaume Renon

1.1.8 Conclusion

La TDM, basée sur la physique des rayons X et sur les mathématiques du début du XXème siècle, est une formidable modalité d'imagerie qui permet de visualiser la densité et la nature de la matière imagée. Elle est rapide, facile d'accès et peu coûteuse, ce qui la rend très utilisée en routine clinique pour visualiser les fractures osseuses et en fait la modalité de choix pour le diagnostic du TC. Néanmoins les rayonnements ionisants induits par les rayons X sont dangereux pour les patients et la dose en rayons X est limitée autant que possible, ce qui a des conséquences non négligeables sur les images. Ces images sont ainsi sensibles aux artefacts mais également aux variabilités inhérentes aux différentes machines ou protocoles d'acquisition, ce qui peut rendre exigeante leur analyse par les radiologues. Cette variabilité a d'ailleurs été étudiée pendant ma thèse et est présentée au chapitre 8. Ces travaux sont actuellement poursuivis au sein de l'équipe dans le cadre d'une autre thèse.

1.2 Traumatisme crânien

Le traumatisme crânien (TC) est une pathologie grave et fréquente, responsable d'un coût financier et humain très important pour la société. Cette partie présente la définition et l'épidémiologie du TC, ainsi que différents scores, qui mesurent la gravité d'un TC, et les différentes lésions cérébrales engendrées par le TC, qui peuvent être observables en TDM.

1.2.1 Définition

Il y a plusieurs définitions du traumatisme crânien (TC). On peut en effet considérer que tout choc sur le neurocrâne, c'est-à-dire sur le crâne en excluant le visage, est un TC. Néanmoins, pour nos travaux, nous allons admettre la définition proposée par l'Organisation Mondiale de la Santé (OMS), et introduite par Thurman en 1995 [reef Thurman, 1995]. Ainsi, et en substance, le traumatisme crânien est une blessure à la tête suite à un traumatisme fermé, ouvert, ou à un phénomène d'accélération/décélération, qui entraîne une altération de l'état de conscience (ou une amnésie) associée ou non à des manifestations neurologiques, des fractures du crâne, ou des lésions intracrâniennes.

1.2.2 Épidémiologie du TC

De par sa nature, l'épidémiologie du TC est difficile à mesurer. En effet, le spectre de gravité est tellement large que la plupart des TC légers ne sont pas diagnostiqués, ce qui vaut parfois au TC d'être qualifié d'épidémie silencieuse, alors que les plus graves, entraînant le décès avant la prise en charge médicale, sont difficiles à répertorier. L'incidence du TC n'est donc pas homogène sur la planète [32] et dépend entre autres de la richesse des pays [33]. Les études d'épidémiologie ayant des méthodologies et des populations d'étude différentes, les chiffres calculés varient grandement et sont donc à prendre avec les précautions nécessaires.

Le TC entraîne chaque année l'admission aux urgences de plus de 260 personnes pour 100 000 en Europe et est responsable du décès d'environ 11 personnes pour 100 000 [34] [35]. Les études s'accordent pour estimer une répartition de 80% de TC légers, 10% de modérés et 10% de TC sévères [33], ces derniers engendrant un taux de décès compris entre 29 et 55% en Europe [34]. Aux États-Unis, ces chiffres augmentent et l'incidence d'admission aux urgences, d'hospitalisation, et de décès, sont respectivement de 468, 94, et 17 pour 100 000 personnes, selon une étude menée en 2010 [36]. Il existe très peu d'études sur les causes du TC en France. Une étude de 2001 menée dans l'ancienne région Aquitaine a estimé l'implication des accidents de la voie publique et des chutes dans respectivement 50% et 40% des cas des TC [37]. Le Tableau 3 résume les résultats de cette étude sur les causes du TC.

	Survenue en %	Incidence/10 ⁵ /an	Âge médian (p25-p75)	Mortalité en %
Accidents de trafic	48,3	8,4	-	
<i>Piétons</i>	15,5	-	65 (29-75)	37,8
<i>Cyclistes</i>	11,7	-	47 (20-63)	28,6
<i>Deux roues motorisées</i>	19,7	-	28 (21-38)	36,2
<i>Occupants de véhicules à moteur</i>	53,1	-	28 (21-43)	21,9
Chutes de sa propre hauteur	24,5	4,2	71 (53-82)	33,6
Chutes de grande hauteur	17,3	3,0	49 (27-71)	26,7
Coups à la tête	6,4	1,1	39 (22-52)	18,8
Armes à feu	2,8	0,5	46 (26-50)	71,4
Divers	0,6	0,1	-	-
Total	100	17,3	44 (23-69)	30,2

Tableau 3. Résultats détaillés de l'étude de Masson & al [37] sur la cause de survenue du TC. Étude menée sur moins de 500 patients de la région Aquitaine, en France et en 1995.

Une étude menée aux États-Unis a comparé l'incidence d'admission aux urgences, d'hospitalisation et de décès dus au TC en fonction de l'âge et du sexe des patients [36] (Figure 14). Il ressort que les hommes sont bien plus exposés aux TC que les femmes, et que les 3 tranches d'âge les plus touchées sont ainsi les jeunes enfants (0-4 ans), les adolescents/jeunes adultes (15-24 ans), et les personnes âgées (65 ans et plus). Cette pathologie touche donc fortement les populations jeunes (< 25 ans), avec de fortes conséquences sur la qualité de vie des patients, ce qui engendre un coût humain mais aussi financier important pour la société. A titre d'illustration, l'impact financier du TC aux États Unis a été estimé en 2007 à 396 000\$ par patient, en prenant en compte le coût de la prise en charge médicale, de la rééducation, et de la perte de productivité [3].

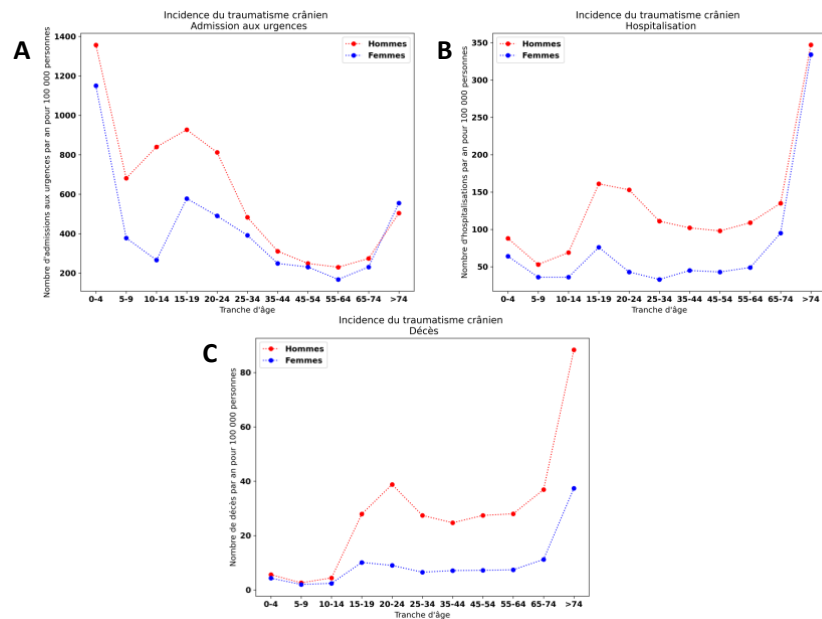


Figure 14. Épidémiologie du traumatisme crânien par sexe et tranche d'âge. A : Nombre d'admissions aux urgences. B : Nombre d'hospitalisations. C : Nombre de décès. Données issues de [36]

1.2.3 Mesurer l'état de conscience d'un patient : Score de Glasgow

Le score de Glasgow (ou GCS pour Glasgow Coma Score), introduit en 1974 par Teasdale et Jennet [4], permet d'évaluer l'état de conscience d'un patient. Il est souvent utilisé comme première estimation de l'atteinte neurologique d'une victime dans un cas d'urgence, comme après un accident, et permet d'orienter la prise en charge initiale. Il se calcule en mesurant la réponse motrice et verbale d'une victime, ainsi que l'ouverture de ses yeux, comme illustré Tableau 4.

Réponse motrice	Réponse verbale	Ouverture des yeux
1 : Nulle	1 : Nulle	1 : Nulle
2 : Extension stéréotypée	2 : Incompréhensible	2 : A la douleur
3 : Flexion stéréotypée	3 : Inappropriée	3 : A la demande
4 : Évitement	4 : Confuse	4 : Spontanée
5 : Orientée	5 : Normale	
6 : Aux ordres		

Tableau 4. Calcul du score de Glasgow – Somme des scores des 3 colonnes

Le GCS correspond à la somme des notes des 3 types de réponses, donnant un chiffre entre 3 et 15. Un score de Glasgow de 15 est parfaitement normal, alors qu'un GCS de 3 indique un coma profond ou bien la mort. Ce score permet également de classer les traumatismes crâniens en :

-
- Sévère (GCS entre 3 et 8)
 - Modéré (GCS entre 9 et 12)
 - Léger (GCS entre 13 et 15)

Si le score de Glasgow est corrélé à la mortalité [38] il reste peu précis et soumis à de nombreux paramètres non liés au patient : alcoolisation aigue, drogue, hypothermie profonde.... De plus, il n'est pas utilisable après injection de sédatifs, ce qui empêche son utilisation pour suivre l'état d'un patient au cours du temps. Ainsi le score de Glasgow mesure l'état de conscience d'un patient, qui peut être altérée par les lésions cérébrales consécutives au TC. Intuitivement, on voit bien que la localisation et la taille d'une lésion cérébrale va influencer sur cet état de conscience, car les différentes zones du cerveau sont impliquées dans les voies de la conscience et remplissent des rôles différents. C'est pourquoi le meilleur moyen de surveiller l'état neurologique d'un patient est de le maintenir éveillé. En cas de nécessité de sédation, il y a une perte d'informations cliniques et donc une nécessité de surveiller l'évolution lésionnelle par d'autres approches, comme la TDM, qui apporte des informations différentes et complémentaires sur l'état du patient et sur l'évolution de ses lésions [39].

1.2.4 Résumer la prise en charge médicale d'un patient : score TILsum

Dès l'admission à l'hôpital, les victimes de TC graves sont admises dans le service de déchocage, unité de soins d'urgence dont les premiers objectifs sont la stabilisation des patients et le premier diagnostic. Puis, le patient subit un examen TDM et est ensuite transféré en réanimation ou au bloc opératoire selon les observations faites. Le diagnostic précis et l'évolution des constantes du patient conditionnent les interventions subies en réanimation [40]. Elles peuvent être modérées, comme par exemple une sédation simple ou extrêmes, comme une craniectomie, c'est-à-dire une ablation partielle de la boîte crânienne afin de réduire la pression intra crânienne (PIC).

La prise en charge en réanimation d'un patient traumatisé crânien peut être résumée et quantifiée grâce à un score qui dépend des interventions et traitements subis par ce patient. Ce score est nommé TILsum, pour Therapy Intensity Level summary, et est détaillé sur le Tableau 5.

CALCULATING SUMMARY TIL

	Assignment of scores to correspond with PED version	
	Score	Max score
Head elevation for ICP control	1	1
Nursed flat (180°) for CPP management	1	
Sedation (low dose as required for mechanical ventilation)	1	8
Higher dose sedation for ICP control (not aiming for burst suppression)	2	
Metabolic suppression for ICP control with high dose barbiturates or propofol	5	
Neuromuscular blockade (paralysis)	3	
CSF drainage <120 ml/day (<5 ml/hour)	2	3
CSF drainage ≥ 120 ml/day (≥ 5 ml/hour)	3	
Fluid loading for maintenance of cerebral perfusion	1	2
Vasopressor therapy required for management of cerebral perfusion	1	
Mild hypocapnia for ICP control [PaCO ₂ 4.6-5.3 kPa (35-40 mmHg)]	1	4
Moderate hypocapnia for ICP control [PaCO ₂ 4.0-4.5 kPa (30-35 mmHg)]	2	
Intensive hypocapnia for ICP control (PaCO ₂ < 4 kPa (30 mmHg))	4	
Hyperosmolar therapy with mannitol up to 2 g/kg/24 hours	2	6
Hyperosmolar therapy with hypertonic saline up to 0.3 g/kg/24 hours	2	
Hyperosmolar therapy with mannitol > 2 g/kg/24 hours	3	
Hyperosmolar therapy with hypertonic saline > 0.3 g/kg/24 hours	3	
Treatment of fever (temp > 38°C or spontaneous temp < 34.5°C)	1	5
Mild hypothermia for ICP control with a lower limit of 35°C	2	
Hypothermia below 35°C	5	
Intracranial operation for progressive mass lesion, not scheduled on admission	4	9
Decompressive craniectomy	5	
Total maximal score:		38

Tableau 5. Calcul du score TILsum, Tableau issu de la publication de Zuercher & al. [41]

Ce score est habituellement calculé en prenant la valeur maximale de chaque élément constitutif du TILsum durant 4 ou bien 24 heures [41]. L'évolution du TILsum au cours du temps permet ainsi de suivre l'évolution de l'intensité thérapeutique dont le patient a besoin pour être stabilisé. Ce score étant calculé à postériori, il est inutile à la prise en charge et n'est donc pas calculé en routine clinique, mais seulement à des fins de recherche.

1.2.5 Mesurer le devenir d'un patient TC à long terme : score GOS

Le TC peut engendrer de lourdes séquelles parfois mortelles [42]. On mesure généralement le devenir des patients à 3 ou 6 mois, par la mortalité ou grâce au score du devenir de Glasgow [43] (ou Glasgow Outcome Score - GOS). On peut aussi utiliser l'échelle étendue du score du devenir de Glasgow (ou Glasgow Outcome Score Extended - GOSE) défini par le Tableau 6 [44]. Le devenir à long terme des patients dépend bien évidemment de la gravité du TC, de la localisation des lésions cérébrales [45], mais également d'autres variables qu'il est pour l'instant difficile de quantifier, comme la rééducation ou encore des facteurs sociologiques comme

l'accompagnement des familles ou l'hygiène de vie. Ces variables compliquent la prédiction à long terme du devenir neurologique des patients TC à partir des simples données médicales habituelles obtenues pendant la prise en charge initiale des patients.

Score du devenir de Glasgow (GOS)	Score du devenir de Glasgow étendu (GOSE)	État du patient	
1	1	Mort	
2	2	État végétatif	
3	3	Besoin d'assistance dans la vie quotidienne	Handicap sévère
	4	Besoin partiel d'assistance dans la vie quotidienne	
4	5	Indépendant, mais pas de reprise du travail possible	Handicap modéré
	6	Reprise partielle du travail possible	
5	7	Déficit mineur physique ou mental qui affecte la vie quotidienne	Handicap faible
	8	Aucun handicap ou très léger	

Tableau 6. Définition du score du devenir de Glasgow, dans sa version classique et étendue [46].

Les 3 scores présentés jusqu'à présent mesurent l'état de conscience d'un patient (GCS), l'intensité de sa prise en charge thérapeutique (TILsum), et son devenir à long terme (GOS). Ces scores ne dépendent pas directement des lésions cérébrales observables par TDM. Pour compléter ces scores cliniques, il existe des scores d'imagerie, basés sur l'analyse des images TDM, capables de quantifier la charge lésionnelle cérébrale d'un patient. Avant de présenter ces scores d'imagerie, nous allons introduire brièvement quelques notions et mots clés de l'anatomie de l'encéphale sain, puis nous allons lister les principales lésions cérébrales du TC ainsi que leur localisation au sein de l'encéphale.

1.2.6 Brève anatomie de l'encéphale

Chez l'Homme, le système nerveux (SN), qui coordonne les actions avec l'environnement extérieur et la communication entre les différentes parties du corps, est habituellement séparé en système nerveux central (SNC) et périphérique (SNP). Le SNC est constitué de l'encéphale, à l'intérieur de la boîte crânienne, et de la moelle

épineière, prolongement de l'encéphale contenu dans le canal spinal de la colonne vertébrale. Les tissus du SNC, composés des neurones, sont regroupés dans ce qu'on appelle la substance grise ou matière grise (MG), par opposition à la substance blanche ou matière blanche (MB), essentiellement composée des axones, les fibres nerveuses qui conduisent le signal électrique entre les neurones.

A l'intérieur de la boîte crânienne, l'encéphale est protégé par un ensemble de membrane : les méninges, constituées de la dure-mère, l'arachnoïde, et la pie-mère, comme illustré Figure 15. L'ensemble baigne dans le liquide cébrospinal (LCS), produit dans des cavités nommées ventricules et résorbé dans les granulations de Pacchioni, dans l'espace sous-dural entre la dure-mère et l'arachnoïde.

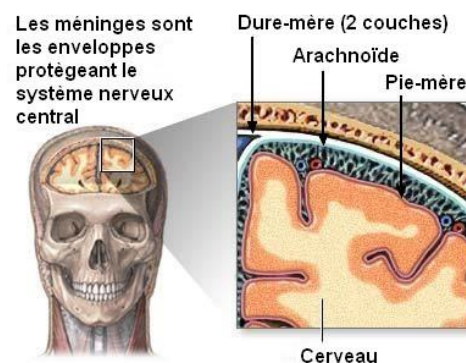


Figure 15. Par Grook Da Oger — Travail personnel, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=16090617>

On peut diviser l'encéphale en 3 parties, comme illustré Figure 16 :

- Le cerveau, lui-même constitué du télencéphale et du diencephale [47], [48]:
 - Diencephale : constitué entre autres du thalamus, de l'hypothalamus, de l'épithalamus, et du sous-thalamus, il contribue à la régulation du système hormonal et du sommeil.
 - Télencéphale (détaillé ci-dessous) : constitué des deux hémisphères cérébraux, droit et gauche, qui sont rattachés de part et d'autre au diencephale et permettent de développer des facultés intellectuelles ou d'effectuer un mouvement quand on le désire.
- Le tronc cérébral, prolongement de la moelle épinière, constitué du bulbe rachidien, du pont, et du mésencéphale, permet le contrôle des poumons et du cœur.
- Le cervelet, situé sous le cerveau et derrière le tronc cérébral, permet la coordination des mouvements du corps.

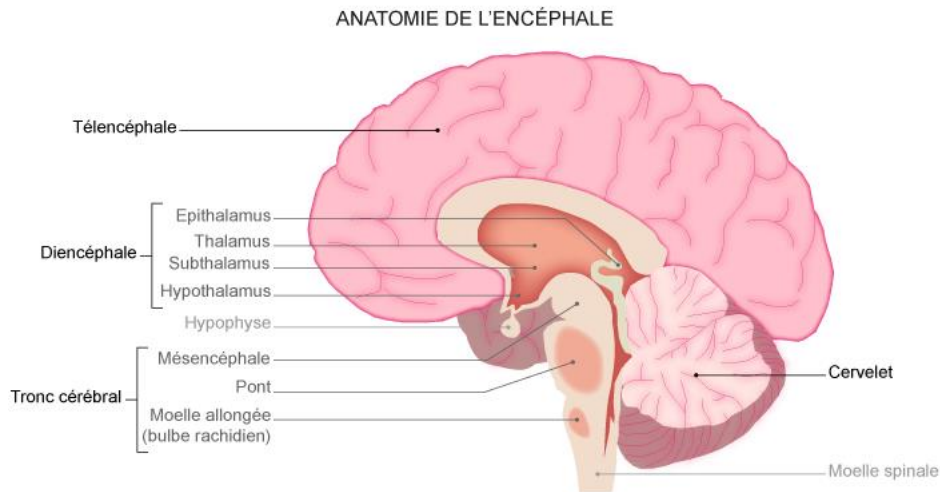


Figure 16. Anatomie de l'encéphale. Source : M. Pitte (2019)[49]

Le télencéphale est la zone la plus volumineuse du cerveau. On peut la décomposer en 4 lobes externes [48][Lefebvre, 2006], présentés Figure 17, et 2 lobes internes :

- Le lobe frontal (externe), qui intervient essentiellement dans la planification, la prise de décision, le raisonnement, le langage et le mouvement volontaire.
- Le lobe temporal (externe), dont la partie supérieure traite les informations auditives et la partie inférieure les informations visuelles. Il est aussi impliqué dans la mémoire.
- Le lobe pariétal (externe), qui intègre les informations sensorielles et qui est impliqué dans la perception de l'espace et dans l'attention.
- Le lobe occipital (externe), centre visuel qui permet la reconnaissance des orientations et des contours des images.
- Le lobe limbique (interne), qui traite les informations concernant les émotions.
- Le lobe de l'insula (interne), permettant de traiter la douleur, les odeurs, et le goût.

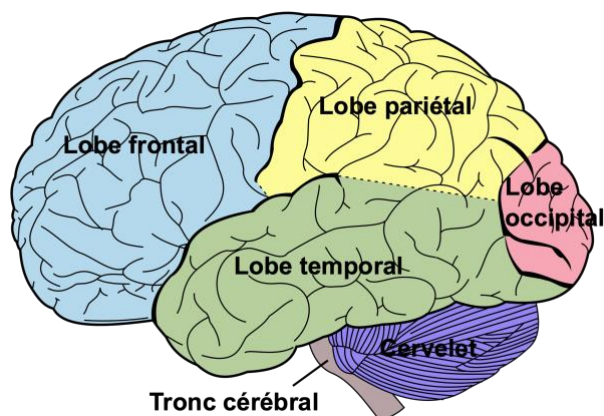


Figure 17. Lobes externes du cerveau humain. Par Mysid — Vectorisé sur CorelDraw par Mysid, Domaine public, <https://commons.wikimedia.org/w/index.php?curid=4424792>

Les ventricules sont eux au nombre de 4. Deux ventricules latéraux (un par hémisphère), un troisième ventricule en forme d'entonnoir situé entre les thalamus droit et gauche, au-dessus du tronc cérébral, et un quatrième ventricule qui longe le tronc cérébral, relié au troisième ventricule par l'aqueduc cérébral.

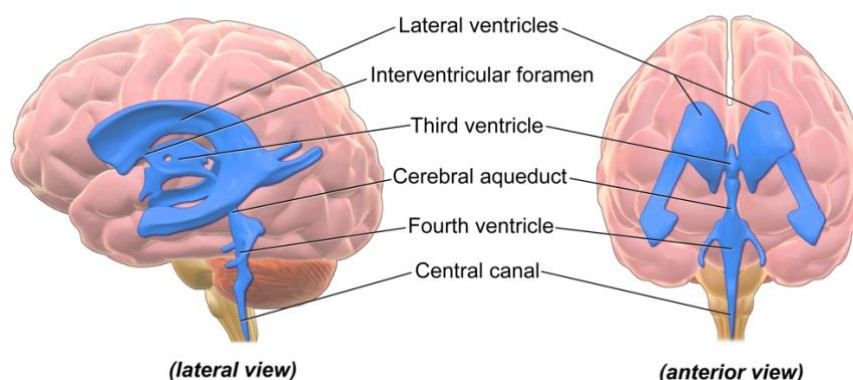


Figure 18. Système ventriculaire cérébral. Par BruceBlaus — Travail personnel, CC BY 3.0, <https://commons.wikimedia.org/w/index.php?curid=28761845>

1.2.7 Lésions typiques du TC

Résultante d'une violente agression sur la boîte crânienne, le TC engendre des lésions cérébrales multiples et de nature différente. Ces lésions coexistent au sein de l'encéphale, influents les uns sur les autres, et requièrent une prise en charge médicale différenciée. On distingue les lésions primaires, quasi immédiates après l'accident, et inévitables, des lésions secondaires, résultantes des lésions primaires, qui aggravent le pronostic et qui sont l'objet des traitements et interventions

médicales. Les principales lésions cérébrales causées par le traumatisme crânien sont détaillées ci-dessous.

1.2.7.1 Fracture

La fracture du crâne est la lésion cérébrale typique du TC la plus évidente. Elle est la cause d'un choc sur le neurocrâne. Elle est généralement accompagnée d'autres lésions cérébrales, même si ce n'est pas toujours le cas. A l'inverse, elle n'est pas systématique, même en cas de TC grave. Un exemple de fracture visible et entourée en orange sur une image TDM est visible sur la Figure 19.



Figure 19. Illustration d'une fracture du crâne, entourée en orange. Le fort contraste entre les os et les autres tissus cérébraux permet de détecter rapidement les fracture par TDM. Fenêtrage entre 0 et 500 UH.

1.2.7.2 Hémorragies / hématomes

Une hémorragie est un saignement causé par une rupture d'un vaisseau sanguin. Une fois l'hémorragie tarie, le sang collecté forme un hématome. Hématome et hémorragie sont donc intimement liés et nous utiliserons l'un ou l'autre sans distinction dans la suite de ce document. On classe habituellement les hémorragies en fonction de leur localisation. Ainsi, on distingue :

- L'hémorragie intra parenchymateuse (HIP), ou intra-cérébral : hémorragie localisée dans le parenchyme cérébral, c'est-à-dire dans les tissus constitutifs de l'encéphale
- L'hémorragie intra ventriculaire (HIV) : hémorragie localisée dans les ventricules, habituellement remplis de LCS.
- L'hémorragie extra axiale (HEA) : hémorragie au niveau des méninges, un ensemble de membranes qui enveloppent l'encéphale. En fonction de la position de l'hémorragie par rapport à ces couches de membranes, illustrées sur la Figure 20, on distingue :

- L'hémorragie sous-durale (HSD) : hémorragie localisée entre la dure mère et l'arachnoïde
 - L'hémorragie extra-durale (HED), ou épidurale : hémorragie localisée entre l'os du crâne et la dure mère
 - L'hémorragie sous-arachnoïdienne (HSA) : hémorragie localisée entre l'arachnoïde et la pie mère, qui est la membrane méningée qui enveloppe entièrement l'encéphale
- Les lésions axonales diffuses : lésions hémorragiques dues à un phénomène d'accélération/décélération brutale du parenchyme cérébral. Elles apparaissent généralement à l'interface entre matière grise et blanche. En effet, ces deux matières ayant des densités différentes, une accélération brutale met ces deux matières en mouvement à une vitesse différente et crée ainsi une force de cisaillement à l'interface qui peut rompre les axones et créer de petites hémorragies. Ces lésions sont parfois repérables au scanner par de petites hyperintensités hémorragiques appelées pétéchies. Elles sont souvent multiples et diffuses, touchant un grand nombre de régions encéphaliques. Ces lésions ont un faible impact sur la prise en charge à court terme des patients mais une grande influence à long terme sur la récupération des fonctions fonctionnelles des patients.

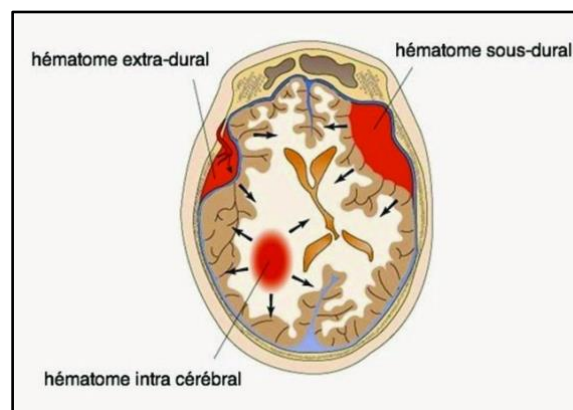


Figure 20. Représentation de 3 types d'hématomes : intra-parenchymateux, sous-dural, et extra-dural. Source : *ophthalmologie.pro* ou Thèse de Héloïse Baillet [50]

1.2.7.3 Autres lésions

Parmi les autres lésions, non hémorragiques, on peut lister l'œdème, l'ischémie cérébrale, et le vasospasme, détaillées ci-dessous :

-
- Un œdème est un gonflement d'un tissu dû à une accumulation de liquide. On peut distinguer l'œdème cytotoxique, où l'accumulation de liquide se fait au sein des cellules, de l'œdème vasogénique, qui se caractérise par une augmentation du liquide dans l'espace interstitiel.
 - Une ischémie représente une souffrance cellulaire suite à une inadéquation entre apports et besoin du métabolisme cellulaire en oxygène. Si elle dure trop longtemps elle engendre la mort des cellules.
 - Un vasospasme est un rétrécissement d'un vaisseau, engendrant une diminution du flux sanguin et par conséquent une ischémie.

1.2.7.4 Cercle vicieux œdémateux

Les différentes lésions détaillées ci-dessus interagissent entre elles. Nous allons décrire ici l'enchaînement physiopathologique appelé cercle vicieux œdémateux qui a des conséquences délétères pour le patient et dont l'évitement est l'enjeu principal de la prise en charge des patients TC. Les lésions hémorragiques ou œdémateuses entraînent mécaniquement une hausse de volume dans une boîte crânienne inextensible, ce qui fait augmenter la pression intra crânienne (PIC). Or, pour une pression artérielle moyenne (PAM) constante, la pression de perfusion cérébrale (PPC), définie comme la différence entre la PAM et la PIC, ne pourra que diminuer. Une baisse de la PPC entraînera à son tour une diminution du niveau d'oxygénation insuffisante des tissus cérébraux, qui si elle n'est pas rapidement résolue, engendrera la mort cellulaire. Cette mort entraîne une hausse du liquide interstitiel et donc un œdème, qui conduit à une augmentation de la PIC. Ce cercle vicieux œdémateux crée donc une expansion des zones lésées et conduit à la mort cérébrale.

Pour conclure, les lésions du TC diffèrent donc par leur nature et par leur influence les unes sur les autres. Leurs caractéristiques ainsi que leurs multiples localisations possibles génèrent toute la complexité nosologique du TC. Ces profils lésionnels hétérogènes ont des conséquences différentes sur l'état des patients. Le suivi de l'évolution de ces lésions est donc crucial pour le diagnostic et la prise en charge de ces patients. Ce suivi peut être fait par imagerie médicale, et notamment par tomodensitométrie à rayons X.

1.2.8 Sensibilité des différentes lésions à la tomodensitométrie

Les lésions induites lors d'un TC se matérialisent par une modification plus ou moins forte de la densité des tissus, ce qui permet de les visualiser par imagerie médicale.

En 2008, Saatman & al. [51] ont recensé la sensibilité de différentes techniques d'imagerie médicale à quelques-unes des lésions du TC. Le Tableau 7, tirée de cet article, résume ces informations.

TABLE 2. IMAGING MODALITIES FOR PATHOANATOMIC AND PATHOPHYSIOLOGIC SEQUELAE

Modality	Hematoma	Contusion	SAH	DAI	Ischemia	Vasospasm
CT	+++	++	+++	+/0	+/0	0
CTA	0	0	+	0	+	+++
CTP	+	+	0	0	++	++
T2 FLAIR	+	+++	+++	+	+	0
1.5T T2* GRE	+	+	+++	+	0	0
3T T2* GRE	+	+	+++	++	0	0
SWI	+++	++	++	+++	0	0
DTI	+	+	0	+++	+++	0
PWI	+	+	0	+	++	++
MRSI	0	0	0	++	+	0
MRA	0	0	+	0	++	+++
fMRI	0	0	0	?	0	0
PET	0	+	0	?	+++	0

Scoring: little/no data, ?; insensitive, 0; minimally sensitive, +/0; mildly sensitive, +; moderately sensitive ++; highly sensitive, +++.

CT, computed tomography; CTA, computed tomography angiography; CTP, computed tomography perfusion; DAI, diffuse axonal injury; DTI, diffusion tensor imaging; fMRI, blood oxygenation level dependent (BOLD) functional magnetic resonance imaging; MRA, 3D time-of-flight MR angiography; MRSI, MR spectroscopic imaging; PET, positron emission tomography; PWI, dynamic susceptibility contrast perfusion-weighted imaging; SAH, subarachnoid hemorrhage; SWI, susceptibility-weighted imaging; T2 FLAIR, T2-weighted fluid attenuated inversion recovery; T2*GRE, T2*-weighted gradient echo.

Tableau 7. Sensibilité de différentes modalités d'imagerie à différentes lésions cérébrales. Source : [51]

La TDM est donc l'une des techniques d'imagerie les plus sensibles aux lésions du TC après l'IRM. Mais comparé aux autres modalités, de par son coût relativement faible, son accessibilité, et sa rapidité d'acquisition, la TDM est devenue la modalité de choix pour le diagnostic précoce et le suivi des lésions induites par un TC. La Figure 21 présente des exemples de chacune des principales lésions présentées ci-dessus sur des images de TDM.

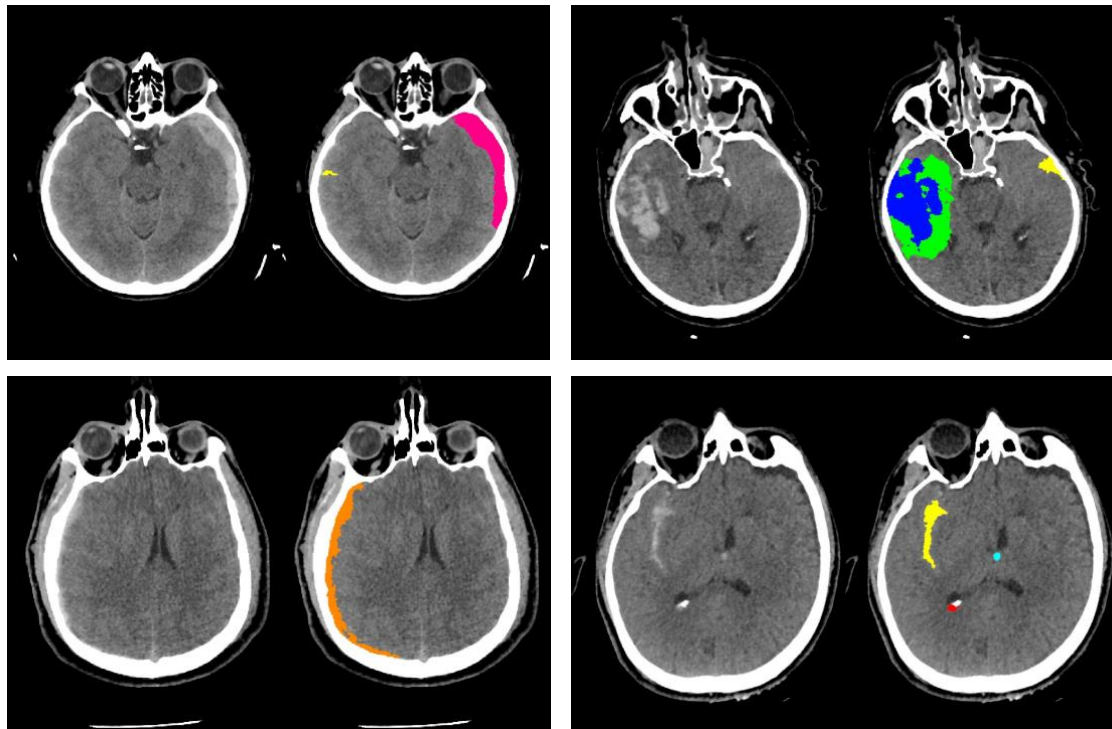


Figure 21. Illustration des principales lésions post TC imagées par TDM acquises sur 4 patients. Fenêtrage entre -15 et 115 UH. Légende : Bleu foncé – Hémorragie intra parenchymateuse, Orange – Hémorragie sous durale, Rose – Hémorragie extra durale, Rouge – Hémorragie intra ventriculaire, Jaune – Hémorragie sous arachnoidienne, Bleu clair – Pétéchies, Vert – Œdème

1.2.9 Résumer une image TDM : Scores d'imagerie

Si la plupart des images TDM de patients traumatisés crâniens sont analysées à l'œil nu par des neuroradiologues spécialisés, il existe des moyens de résumer une image TDM grâce à quelques questions appelant une réponse binaire afin de quantifier l'image. Cette quantification observationnelle sera appelée « score d'imagerie » dans la suite de ce document.

Il existe plusieurs scores permettant de classifier le profil lésionnel d'une image TDM. Les principaux sont le score de Marshall [52] (Tableau 8), de Rotterdam [53] (Tableau 9), de Stockholm [54] ou d'Helsinki [55]. Bien que ces scores diffèrent par leur méthode de quantification, ils sont tous corrélés à la mortalité et au devenir neurologique des patients [55][56][57][5]. Malgré ce pouvoir prédictif évident, ces scores sont peu utilisés en routine clinique car ils doivent être calculés manuellement, et sont donc supplantés par l'analyse à l'œil nu, plus rapide et informative. De plus, il existe une variabilité inter-observateur dans la quantification à l'œil nu des lésions sur image TDM [58], pouvant mener à un diagnostic et une prise en charge différente. Finalement, ces scores restent imprécis, notamment à cause des seuils sur les mesures de distance ou de volume. En effet, si la ligne médiane du cerveau a subi une déviation de 4mm, cette déviation ne sera pas prise en compte dans le calcul des scores de

Marshall ou Rotterdam car le seuil de déviation est fixé à 5mm. Cette information est donc perdue alors que le diagnostic aurait pu être affiné et la prise en charge peut-être modifiée.

Grade	Définition
I (pas de lésion visible)	Pas de signe d'atteinte cérébrale
II	Citernes de la base présentes Déviation de la ligne médiane < 5mm Pas de lésion hyperdense ou en mosaïque > 25mL
III (gonflement)	Citernes comprimées ou absentes Déviation de la ligne médiane > 5mL Pas de lésion hyperdense ou en mosaïque > 25mL
IV (engagement)	Déviation de la ligne médiane > 5mL Pas de lésion hyperdense. Ou en mosaïque > 25mL
V (lésion focale évacuée)	Toute lésion évacuée chirurgicalement
VI (lésion focale non évacuée)	Lésion hyperdense ou en mosaïque > 25mL non évacuées

Tableau 8. Score de Marshall[52]

Citernes de la base	Déplacement de la ligne médiane	Hémorragie extra durale	Hémorragie sous arachnoïdienne ou intra ventriculaire
0 : Normales	0 : Absent ou de moins de 5mm	0 : Présente	0 : Absente
1 : Compressées	1 : De plus de 5mm	1 : Absente	1 : Présente
2 : Absentes			

Tableau 9. Le score de Rotterdam est obtenu en ajoutant les chiffres associés aux 4 réponses et en ajoutant 1 à cette somme.[53]

Une analyse automatique des images TDM par des outils d'intelligence artificielle (IA) pourrait non seulement résoudre cette sous exploitation de l'imagerie acquise en routine clinique mais elle pourrait aussi améliorer la reproductibilité de ces quantifications. La suite de ce manuscrit va donc introduire d'un point de vue théorique quelques notions des algorithmes d'intelligence artificielle, puis, dans un second temps, présentera l'état de l'art de l'utilisation de ces algorithmes appliqués aux images TDM dans le contexte du TC.

1.3 Intelligence artificielle

La récente amélioration des capacités de stockage, de calcul, et la création de grandes bases de données structurées a révélé l'intérêt et la performance des algorithmes d'IA dans le domaine de l'imagerie médicale afin d'améliorer le diagnostic, le pronostic, ou la prise en charge des patients. Cette partie passe en revue les différents types d'apprentissages, les procédures d'entraînements, de validation, et de test, ainsi que le fonctionnement de quelques algorithmes utilisés au cours de cette thèse.

1.3.1 Définition

A cause des implications philosophiques de la définition d'intelligence, la définition d'intelligence artificielle (IA) reste vague. Pour nos travaux, on adoptera la définition de [59] de l'intelligence artificielle : « a field of computer science dedicated to the creation of systems performing tasks that usually require human intelligence », qui peut se traduire par « un domaine de l'informatique dédié au développement de systèmes effectuant des tâches qui nécessitent généralement l'intelligence humaine », et de l'apprentissage statistique (ou machine learning) : « a sub-field of AI that includes all those approaches that allow computers to learn from data without being explicitly programmed », qui peut se traduire par « un champ d'étude de l'IA qui regroupe les approches qui permettent aux ordinateurs d'apprendre à partir de données, sans avoir été explicitement programmés pour ».

1.3.2 Apprentissage supervisé et non supervisé

On peut classer les méthodes d'apprentissage statistique en différentes classes selon leur mode d'apprentissage, dont les deux principaux sont l'apprentissage supervisé et non supervisé.

1.3.2.1 Apprentissage supervisé

L'apprentissage supervisé consiste à demander à un algorithme d'apprendre la relation entre un type d'entrées X et un type de sorties Y . On dispose pour cela des données X et des sorties Y , sans forcément connaître une relation analytique entre X et Y .

C'est le mode d'apprentissage le plus utilisé en imagerie médicale afin de prédire par exemple le devenir d'un patient (Y) à partir de son diagnostic à l'entrée à l'hôpital (X) [60]. Si Y est une variable continue, on parle de régression, et si c'est une variable discrète on parle de classification. Parmi les principaux algorithmes utilisés pour l'apprentissage supervisé, on peut citer le Random Forest, ou la méthode des plus proches voisins.

1.3.2.2 Apprentissage non supervisé

L'apprentissage non supervisé repose sur la minimisation d'une relation sur les données d'entrée elles-mêmes, et donc sur la mise en évidence par l'algorithme de structures sous-jacentes au sein des données. Ainsi on ne dispose pas de « vérité » permettant d'entraîner l'algorithme ou bien de calculer un score d'exactitude. Parmi les principaux algorithmes d'apprentissage non supervisé, on peut citer le k-means clustering ou bien la réduction de dimension par analyse en composantes principales.

1.3.3 Écueil du surapprentissage et méthodologies d'entraînement

L'entraînement d'un algorithme dépend des données d'apprentissage. Plus ces données sont nombreuses et variées, plus le modèle résultant de l'apprentissage sera robuste et pourra être utilisé sur de nouvelles données sans baisse de performance. À l'inverse, peu de données d'entraînement entraînera probablement des instabilités dans le modèle et sera difficilement utilisable sur de nouvelles données.

Même en disposant d'un nombre important de données, il est possible de créer un modèle non généralisable, notamment à cause du surapprentissage, l'un des écueils principaux lors de l'entraînement.

1.3.3.1 Sous apprentissage et surapprentissage

Le sous apprentissage consiste à utiliser un modèle trop simple pour la relation à estimer, ce qui conduit à une mauvaise estimation de la structure des données, et le surapprentissage consiste à utiliser un modèle trop complexe, ce qui conduit à apprendre la structure mais également le bruit inhérent aux données d'entraînement, ce qui résulte en un modèle difficilement généralisable.

Par exemple, plaçons-nous dans le cas d'une tâche de régression dans lequel le but est d'estimer par un polynôme la relation f qui relie les entrées X aux sorties Y selon l'équation $Y = f(X)$. Mettons que l'apprentissage est effectué grâce à la méthode des moindres carrés, qui nécessite comme hyperparamètre l'ordre du polynôme de la relation à estimer.

3 hyperparamètres sont testés : un ordre 1, un ordre 3, et un ordre plus élevé. La Figure 22 représente les points du jeu de données d'entraînement et la relation estimée entre X et Y issues des trois modèles entraînés.

Le graphique de gauche montre que la relation polynomiale d'ordre 1 apprise sur le jeu d'entraînement est trop simple pour bien estimer Y grâce à X : c'est le sousapprentissage. Le graphique de droite montre que la relation polynomiale d'ordre élevé est trop complexe par rapport à la relation entre X et Y et modélise

certes la relation structurelle mais aussi le bruit inhérent aux données d'entraînement : c'est le surapprentissage. Le graphique du centre montre lui une bonne estimation de la structure qui lie les données X et Y : c'est l'apprentissage conforme. Si l'on se concentre uniquement sur la performance sur le jeu d'entraînement, la relation qui donnera les meilleurs résultats sera la relation d'ordre élevé. Ainsi il faut trouver une autre méthode pour sélectionner le meilleur modèle à utiliser.

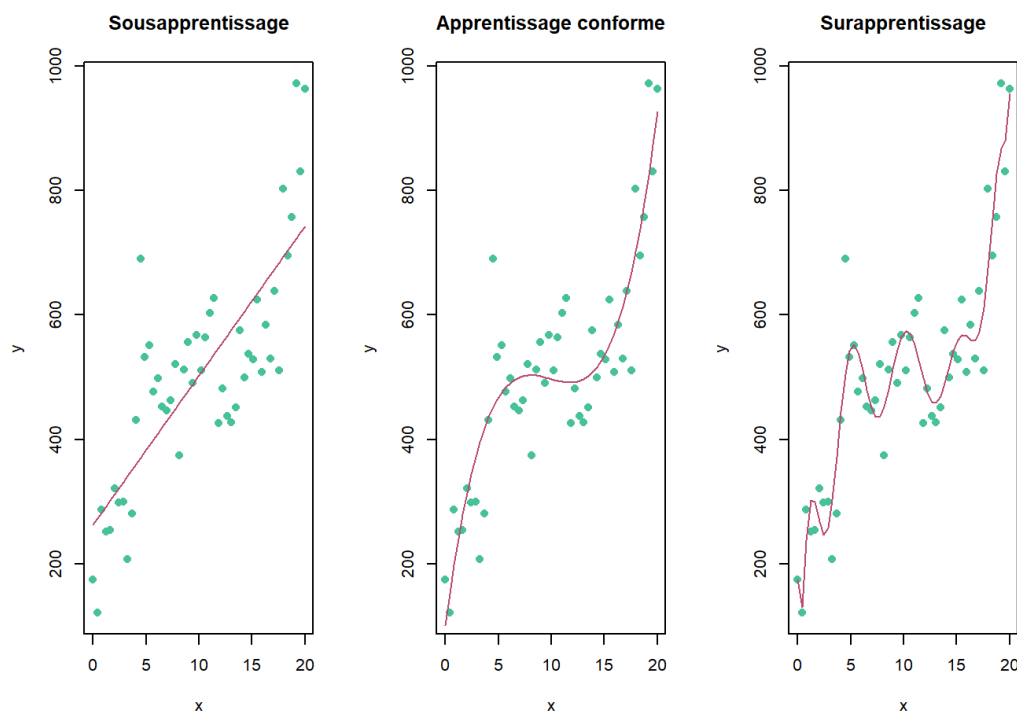


Figure 22. Illustration du sousapprentissage, apprentissage conforme, et surapprentissage d'un modèle sur une tâche de régression. En vert les points du jeu d'entraînement et en rouge la relation estimée. Source [61]

Afin d'éviter l'écueil du surapprentissage et obtenir des résultats généralisables, il faut que les 3 étapes fondamentales du développement d'un modèle que sont l'entraînement, la sélection, et l'évaluation du modèle soient menées sur des jeux de données différents. Il existe pour cela plusieurs méthodologies, qui dépendent du nombre de données disponibles et du but recherché.

1.3.3.2 Séparation en 3 jeux de données statiques

La méthode la plus classique est de séparer l'ensemble des données en 3 jeux de données : un jeu d'entraînement, de validation et de test. Le jeu d'entraînement contient les données que l'on va donner à l'algorithme pour qu'il puisse apprendre sa tâche. Le jeu de validation va permettre de sélectionner les meilleurs hyperparamètres du modèle tout en évitant le surapprentissage. Finalement, le jeu de test permet d'évaluer de manière non biaisée le modèle préalablement sélectionné, et ainsi conclure sur les performances du modèle.

1.3.3.3 Validation croisée

La validation croisée, ou cross-validation en anglais, permet d'effectuer deux des 3 étapes fondamentales sur le même jeu de données [62], ce qui est utile lorsque l'on dispose d'un petit jeu de données, puisque cela permet d'utiliser tous les échantillons du jeu de données. En pratique, cette méthode est très utilisée pour entraîner et sélectionner un modèle, qui sera ensuite évalué sur un autre jeu de données.

Prenons l'exemple d'une validation croisée utilisée pour entraîner et sélectionner un modèle sur un jeu de données nommé D, et qui sera ensuite évalué sur un jeu de données nommé T. Premièrement, on va diviser D en K ensembles (folds en anglais) contenant chacun un nombre équivalent d'échantillons.

On va sélectionner l'un des ensembles et décréter que cet ensemble servira de jeu de validation. Le jeu d'entraînement sera le reste de D, c'est-à-dire les K-1 ensembles restants. Ainsi on va entraîner plusieurs modèles sur le jeu d'entraînement, avec plusieurs valeurs d'hyperparamètres, et calculer leur performance sur le jeu de validation.

Cette procédure va ensuite être répétée en utilisant l'un des autres ensembles comme jeu de validation, et ainsi de suite avec tous les ensembles. On peut ensuite calculer la moyenne des performances de chacun des modèles sur tous les ensembles de validation pour aboutir à la performance globale des modèles et ainsi sélectionner le modèle qui fournit la meilleure performance moyenne. Ce modèle sélectionné pourra ensuite être évalué sur le jeu de données T, préalablement mis de côté.

Là où la méthode du paragraphe précédent impliquait 1 entraînement par jeu d'hyperparamètres, cette méthode implique K entraînements par jeu d'hyperparamètres, ce qui augmente le temps de calcul.

	Jeu de données D				
Répartition	Ensemble 1	Ensemble 2	Ensemble 3	Ensemble 4	Ensemble 5
Répartition 1	P Entraînements car P jeux d'hyperparamètres et donc P modèles à valider				Validation
Répartition 2	Entraînement			Validation	Entraînement
Répartition 3	Entraînement		Validation	Entraînement	
Répartition 4	Entraînement	Validation	Entraînement		
Répartition 5	Validation	Entraînement			

Figure 23. Illustration de la division d'un jeu de données en ensembles d'entraînement et de validation dans le cas d'une validation croisée avec K=5 ensembles et P jeux d'hyperparamètres à entraîner.

1.3.3.4 Validation croisée imbriquée

Si la validation croisée permet d'effectuer 2 des 3 étapes fondamentales sur le même jeu de données, la 3ème reste à effectuer sur un autre jeu de données. La validation croisée imbriquée, ou *nested cross-validation* en anglais, consiste à imbriquer 2 boucles de validation croisée afin d'effectuer les 3 opérations sur l'ensemble d'un même jeu de données [62]. Ainsi, on dispose d'une boucle interne, où se déroule la validation croisée d'entraînement et de sélection du modèle, et d'une boucle externe, où se déroule la validation croisée d'évaluation, comme illustré Figure 24.

La boucle interne permet d'entraîner et sélectionner des modèles évalués au sein de la boucle externe. Puisque la boucle interne est répétée un grand nombre de fois, les modèles sélectionnés n'ont pas forcément les mêmes valeurs d'hyperparamètres. Ainsi, ce qu'on évalue avec la boucle externe, ce n'est pas vraiment la performance d'un modèle, mais plutôt la performance des métriques d'entrée à prédire un critère grâce à un pipeline de classification.

Bien évidemment, le temps de calcul est là aussi augmenté, car pour une validation croisée interne de X ensembles et une validation croisée de Y ensembles, sans répétitions, il faut faire $X * Y$ entraînements par jeu d'hyperparamètres. Cette augmentation considérable du temps de calcul rend la validation croisée imbriquée raisonnablement utilisable uniquement sur de petits jeux de données, et c'est d'ailleurs là qu'elle est le plus utile car elle permet de limiter le surapprentissage [63].

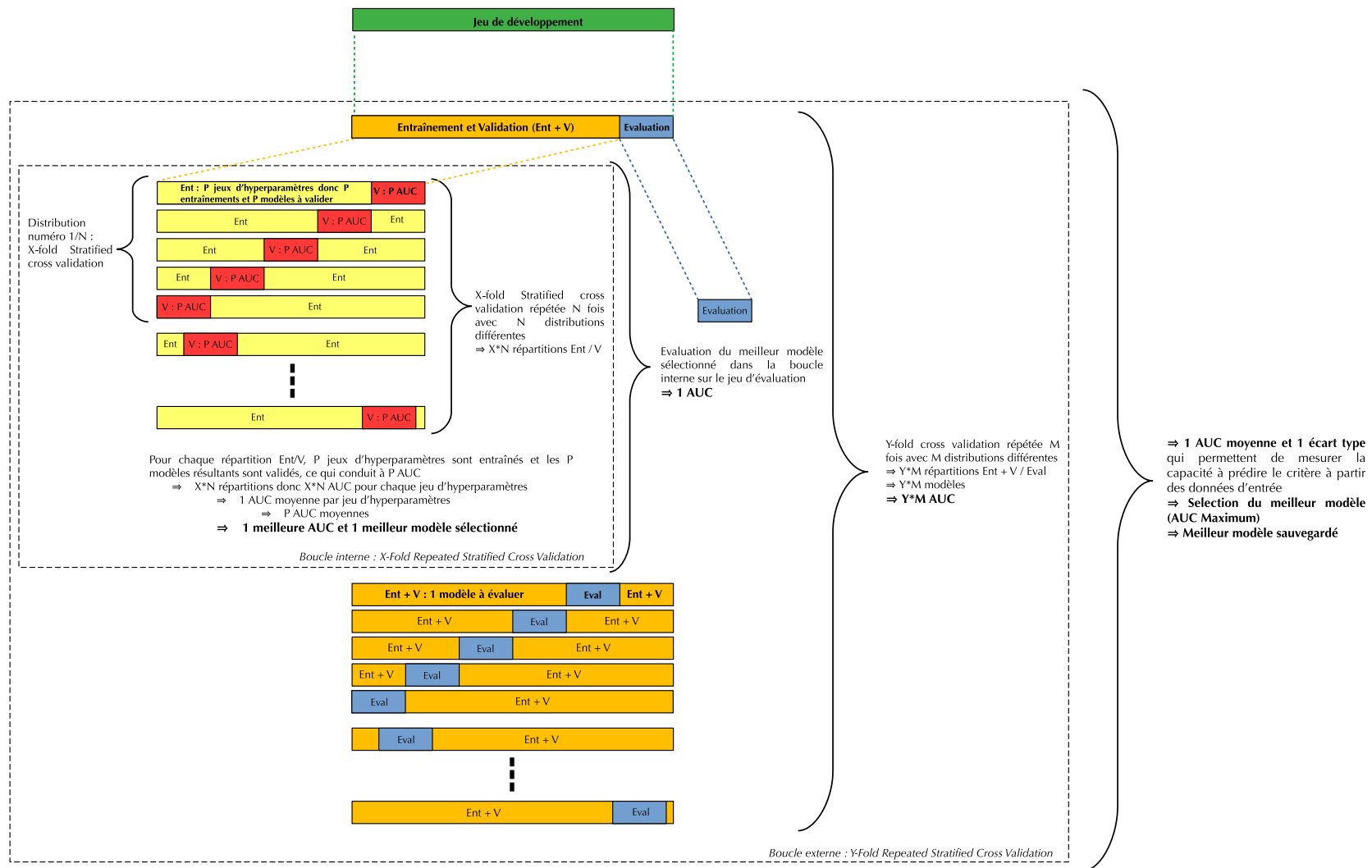


Figure 24. Exemple de la procédure validation croisée imbriquée avec comme boucle interne une validation croisée à $X=5$ ensembles répétée N fois et comme boucle externe une validation croisée à $Y=6$ ensembles répétée M fois. Dans cet exemple, P jeux de paramètres sont testés. Le nombre total d'entraînements effectués dans cet exemple est de $P \times X \times N \times Y \times M$. Sur cette figure, la métrique utilisée pour sélectionner et évaluer les performances des modèles est l'AUC, c'est-à-dire l'aire sous la courbe ROC, définie plus loin.

1.3.4 Exemple d'algorithme de machine learning : Forêts aléatoires

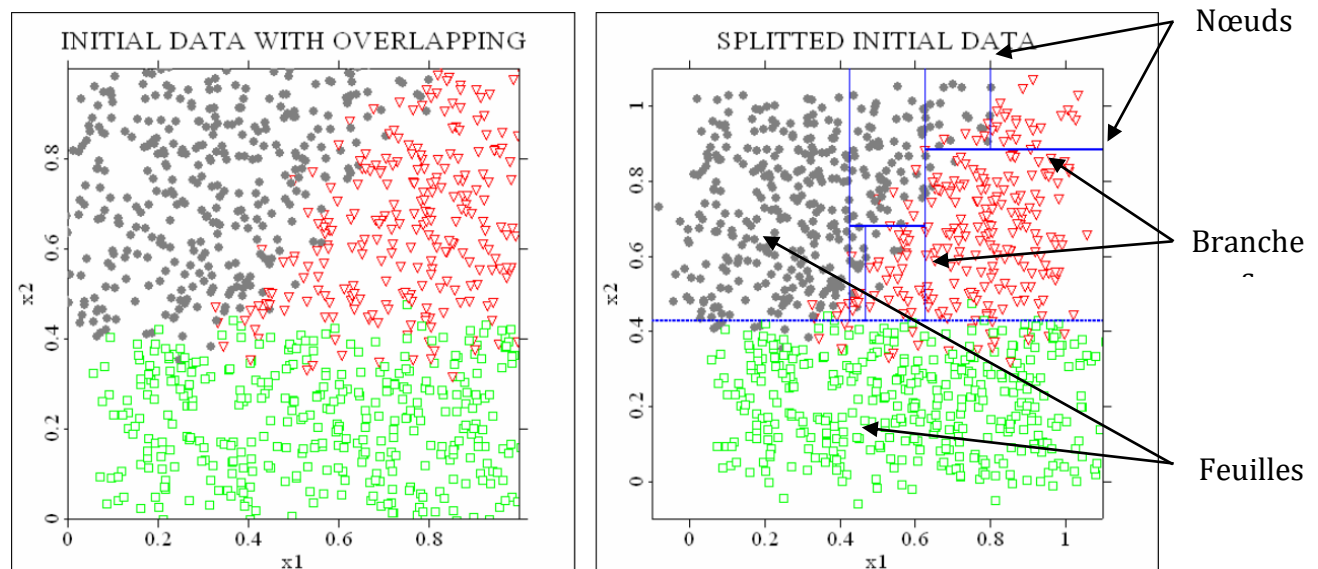
L'un des algorithmes d'apprentissage statistique les plus connus est la forêt aléatoire, ou random forest en anglais. Comme son nom le suggère, une forêt aléatoire est la combinaison de nombreux arbres décisionnels.

1.3.4.1 L'arbres décisionnel

Un arbre décisionnel est un algorithme d'apprentissage supervisé. Nous allons illustrer ce concept à travers un problème de classification visant à attribuer aux entrées X une classe Y . X est composé de plusieurs observations et pour chaque observation de plusieurs métriques. Le principe est de séparer les observations en deux grâce à une condition sur l'une des métriques. Cette condition peut être calculée de différentes manières mais en substance, elle est calculée afin de maximiser la coexistence d'échantillons appartenant à la même classe après séparation des observations.

Chaque séparation est appelée un nœud. Les nœuds sont reliés par des branches, et chaque branche qui ne termine pas par un nœud se termine par ce qu'on appelle une feuille.

Chaque nœud sépare les observations en deux grâce à un critère sur l'une des métriques, ce qui conduit à séparer l'espace des données d'entrées en rectangles, dans une représentation 2D des données, comme illustré sur la Figure 25.



1.3.4.2 La forêt aléatoire

La forêt aléatoire est un algorithme d'apprentissage supervisé basé sur l'agrégation d'arbres de décision.

Pour chaque arbre, le jeu de données d'entraînement est sous échantillonné de manière aléatoire sur les observations mais aussi sur les métriques et l'arbre décisionnel est ensuite entraîné sur ce sous-ensemble.

Un certain nombre d'arbres N sont ainsi entraînés, ce qui conduit à N prédictions différentes. Ces prédictions sont agrégées et la prédiction de la forêt aléatoire est la prédiction majoritaire des N arbres décisionnels.

1.3.5 Exemple d'algorithmes d'IA : Apprentissage profond

L'apprentissage profond, ou deep learning, est un ensemble de méthodes de machine learning, supervisées ou non, basées sur des réseaux de neurone artificiels. Depuis les années 2010, ces méthodes surpassent et remplacent les algorithmes de machine learning « classique ». En particulier, dans le domaine de la vision par ordinateur, les réseaux de neurones profonds convolutifs obtiennent des résultats spectaculaires et constituent l'immense majorité des algorithmes vainqueurs des différents challenges internationaux. Cette sous partie a pour but de présenter succinctement le fonctionnement de ces réseaux.

1.3.5.1 Neurones artificiels

Inspirés des neurones naturels, les neurones artificiels sont des systèmes transformant un certain nombre de signaux d'entrée en un signal de sortie. Comme illustré sur la Figure 26, les signaux d'entrée X_i sont chacun pondérés par les poids w_i puis sommés entre eux. Un terme de biais b est ensuite ajouté et le tout est finalement transformé par une fonction d'activation non linéaire σ en un signal de sortie $\hat{y}$.

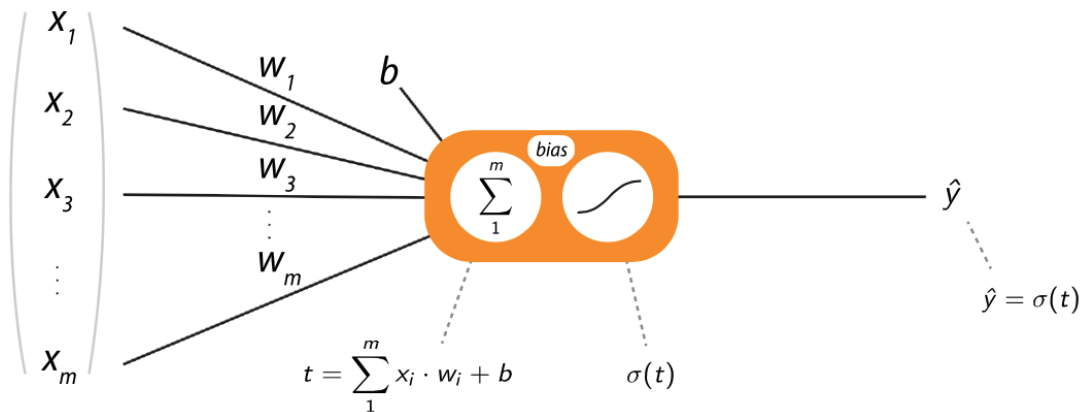


Figure 26. Illustration d'un neurone artificiel. Source : FIDLE, Jean Luc Parouty [65]

Le neurone artificiel est l'élément de base d'un réseau de neurones.

1.3.5.2 Réseaux de neurones

Un réseau de neurones est un ensemble de neurones artificiels structurés en couches reliées entre elles. La première couche contient les signaux d'entrée et la dernière les signaux de sortie. Les couches du milieu sont les couches profondes du réseau.

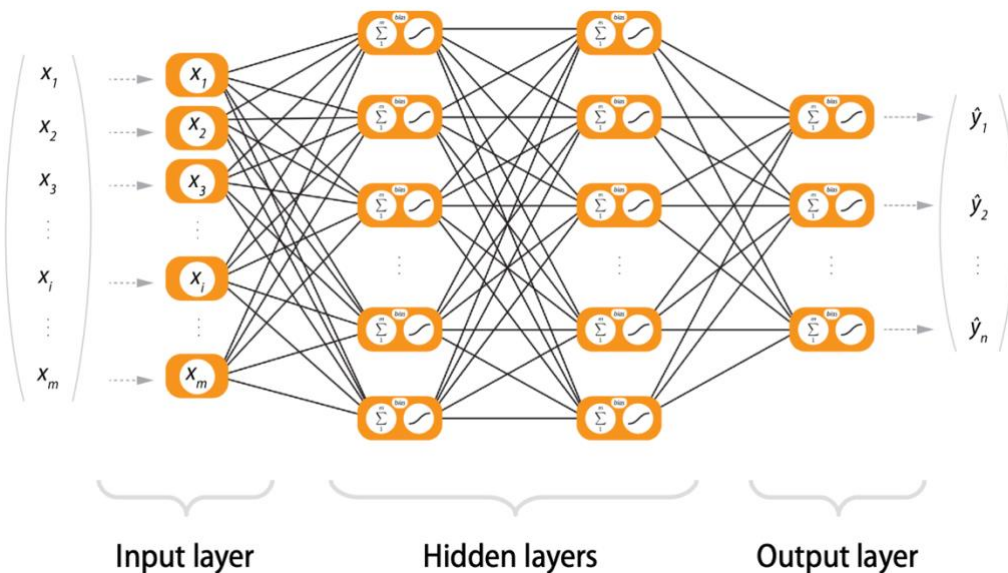


Figure 27. Illustration d'un réseau de neurones artificiels. Source : FIDLE, Jean Luc Parouty [65]

Le nombre de neurones des couches d'entrée et de sortie est fixé par les signaux d'entrée et de sortie, mais le nombre de couches profondes, le nombre de neurones de ces couches, et la manière de les relier entre elles est arbitraire. L'étude de ces différentes architectures et leurs implications est un champ de recherche à part

entière. Figure 27 illustre un réseau dont chaque neurone d'une couche profonde est connecté à l'ensemble des neurones de la couche précédente et à l'ensemble des neurones de la couche suivante. Ce type de réseau est appelé dense (ou fully connected).

1.3.5.3 Neurones convolutifs

Dans le domaine de la vision par ordinateur, lorsque les signaux d'entrée sont des images, l'information spatiale de l'image est exploitée grâce aux réseaux de neurones convolutifs (CNN), une variante des réseaux de neurones qui permet d'extraire les motifs de l'image grâce à des couches de convolution.

Un neurone convolutif transforme l'image d'entrée en une autre à travers une convolution par un filtre appelé noyau, l'ajout d'un terme de biais, et l'application d'une fonction d'activation, généralement non linéaire, comme illustré Figure 28.

Les paramètres de la convolution, tels que la taille du noyau, l'ajout de padding et le pas de déplacement du filtre (strides en anglais), définissent la taille de l'image de sortie, qui peut être ou non de la même taille que celle d'entrée.

Notons que là où le neurone artificiel classique, présenté plus tôt, applique un poids, c'est-à-dire un hyperparamètre, pour chaque pixel d'entrée, le neurone convolutif n'applique aucun poids à la sortie de la convolution. Par conséquent les seuls hyperparamètres de ce neurone sont les valeurs du noyau de convolution et le terme de biais, ce qui conduit généralement à moins d'hyperparamètres qu'un neurone artificiel.

Notons également que si l'image d'entrée est en 3 dimensions, on peut effectuer la même opération avec un noyau de convolution en 3 dimensions.

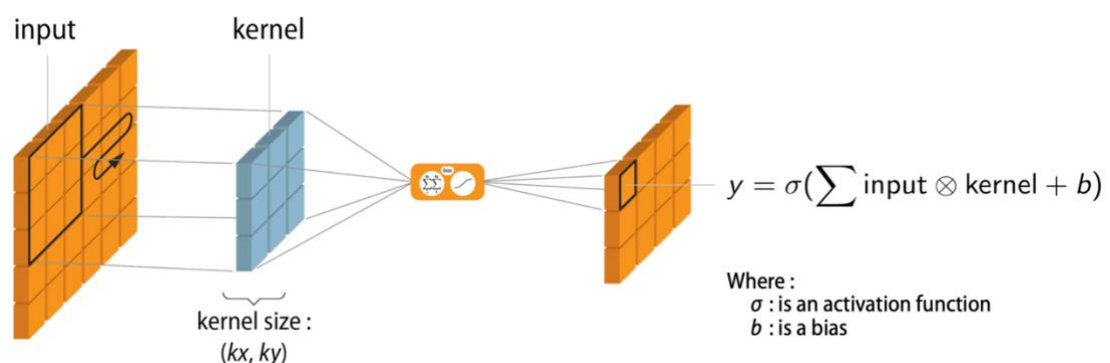


Figure 28. Illustration du passage d'une image à travers un neurone convolutif. Ici, $\otimes$ signifie le produit de Hadamard. Source : FIDLE, Jean Luc Parouty [65]

1.3.5.4 Réseau de neurones convolutifs

Un réseau de neurones convolutifs est un réseau de neurones dont au moins la première couche est une couche de neurones convolutifs. Les autres types de couches généralement associées aux couches convolutives sont les couches de pooling, qui découpe d'image d'entrée en petits patches et ne sélectionne qu'une partie de l'information contenue dans chaque patch, les couches nommées «flat layers», qui transforment les images 3D en 2D en concaténant la 3ème dimension, ou encore les couches de neurones artificiels classiques.

Les couches convolutives permettent d'extraire des motifs de l'image d'entrée. En fonction des noyaux de celles-ci, on extrait des propriétés statistiques de l'image, comme les contours par exemple. Ces propriétés sont assimilables aux métriques Radiomiques, détaillées plus tard dans ce document.

Comme illustré Figure 29, il est courant d'appliquer plusieurs neurones convolutifs au sein d'une couche, afin d'utiliser différents noyaux de convolution et ainsi extraire plusieurs propriétés de l'image.

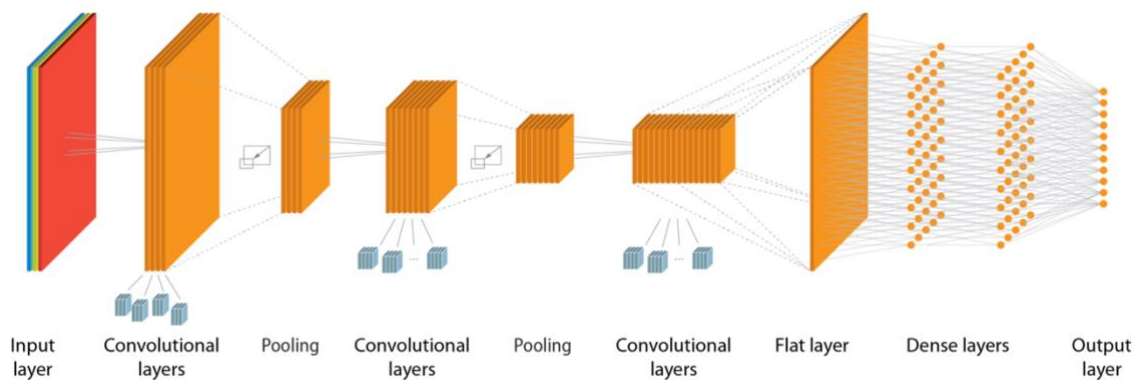


Figure 29. Illustration d'un réseau de neurones convolutif. Source : FIDLE, Jean Luc Parouty [1]

Les valeurs des noyaux de convolution de ces neurones sont aléatoirement initialisés et mis à jour à chaque itération de l'apprentissage du réseau grâce à une méthode nommée rétropropagation.

1.3.5.5 Réseau DeepMedic

DeepMedic est une architecture de réseau de neurones convolutifs bien connue, initialement publiée pour la segmentation de tumeurs à partir d'images IRM [66] [67]. La particularité de DeepMedic, qui travaille sur des patches 3D, est de prendre en compte l'information contenue dans le voisinage de ce patch. Plus précisément il prend en entrée le patch à segmenter mais aussi son voisinage, à l'aide de patches plus grands, à la résolution native ou sous échantillonnée. La Figure 30 décrit un réseau de type DeepMedic tel qu'il pourrait être utilisé avec $N = 4$ couches convolutives, un noyau de convolution isotrope de taille $k = 5$, et deux voies d'analyse,

l'une à la résolution initiale (donc sous échantillonnée d'un facteur $D1 = 1$), et l'autre sur l'image sous échantillonnées d'un facteur $D2 = 3$. Pour segmenter un patch de sortie 3D isotrope de taille $P_{out} = 9$ et centré sur un point C , ce réseau analyse donc un patch d'entrée 3D isotrope centré sur C de taille $P_{in} = 33$, car P_{in} calculée par la formule :

$$P_{in} = P_{out} + \max(D1, D2) \times N \times (k - 1)$$

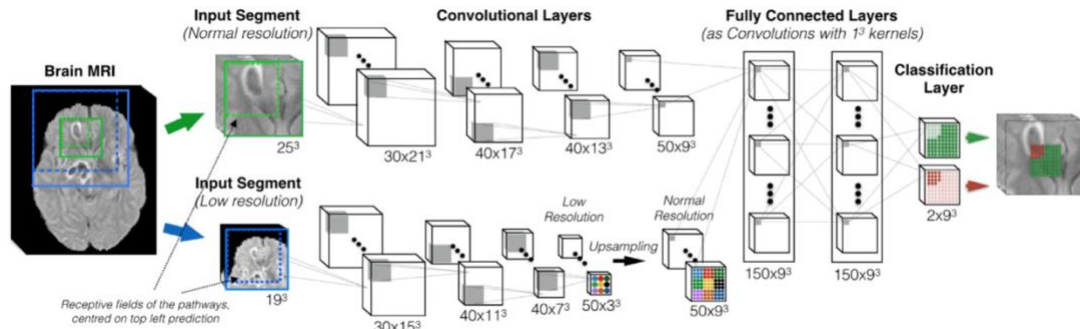


Figure 30. Architecture du réseau DeepMedic [66]

Les deux voies d'analyse permettent d'étudier le voisinage du patch à segmenter à différentes résolutions, ce qui permet de repérer des motifs de taille différents :

- Pour la voie d'analyse sur l'image initiale, un patch centré sur C de taille $P_{in_or} = 17$ ($P_{out} + N \times (k - 1)$) passe à travers les 4 couches convolutives et se transforme en un patch de taille 9, soit la taille du patch de sortie.
- Pour la voie d'analyse sur l'image sous échantillonnée, ce patch d'entrée de taille $P_{in} = 33$ est sous échantillonné d'un facteur $D2 = 3$ pour obtenir un patch de taille $P_{in_ds} = 11$ qui passe à travers les 4 couches de convolution avec des noyaux de taille $k=3$ (sans padding et avec un pas de déplacement égal à 1) et se transforme donc en un patch de taille 3 ($P_{in_ds} - N \times (k - 1)$). Ce patch est ensuite suréchantillonné par le facteur $D2$ pour atteindre une taille égale à 9, soit la taille du patch de sortie.

Ces deux patches 3D isotropes de taille 9, issues des deux voies d'analyses en parallèles, sont ensuite sommés à l'entrée de la première couche dense, qui est connectée à une seconde couche dense puis à une couche de classification, qui est la couche de sortie et qui permet d'obtenir le patch 3D isotrope de taille 9 de la segmentation finale.

1.3.5.6 Mise à jour automatique des hyperparamètres d'un réseau : la rétropropagation

L'apprentissage d'un réseau se fait, comme pour un algorithme d'apprentissage statistique classique, en minimisant une fonction de coût, bien choisie de manière à converger vers un état voulu du réseau. Dans le cas d'un apprentissage supervisé, on peut imaginer comme fonction de coût la différence entre la prédiction et la vérité. La minimisation de cette fonction va forcer le réseau à converger vers un état où la prédiction sera proche de la vérité. Ainsi cette fonction sert à quantifier la performance du réseau.

Un réseau possède des paramètres fixes et d'autres variables. Les paramètres fixes sont le nombre de couches, de neurones, l'architecture du réseau et la fonction d'activation, tandis que les hyperparamètres, variables, sont les valeurs des poids pondérant les signaux d'entrée de chaque neurone, les valeurs de biais et les filtres de convolution (dans le cas d'une couche de neurones convolutifs). Les hyperparamètres vont être modifiés automatiquement par le réseau durant son apprentissage.

Durant l'apprentissage, la fonction de coût va être calculée à partir de la prédiction du réseau. Cette valeur devant être la plus proche possible de 0, elle quantifie la distance entre l'état actuel du réseau et l'état optimal souhaité. En fonction de cette valeur, le réseau va modifier chacun de ses paramètres variables afin de réduire la valeur de la fonction de coût. En pratique, cette modification des paramètres variables est effectuée par « rétropropagation », c'est-à-dire que chaque paramètre est modifié en fonction de son influence dans le calcul de la fonction de coût. Cette influence est obtenue par le calcul de la dérivée partielle de la fonction de coût par le paramètre en question.

Ainsi, l'apprentissage d'un réseau de neurones peut se simplifier en 3 étapes :

- Calcul de la valeur de la fonction de coût de l'état du réseau
- Calcul de l'influence de chaque paramètre variable dans la fonction de coût
- Mise à jour des paramètres variables de manière à réduire la valeur de la fonction de coût

Cette séquence est répétée à chaque nouvel état du réseau, jusqu'à l'arrêt automatique ou manuel de l'entraînement.

1.3.5.7 Profiter d'un réseau déjà entraîné comme initialisation : Transfer learning

L'apprentissage par transfert, ou Transfer Learning, est une méthode d'apprentissage de machine learning qui consiste à initialiser les paramètres d'un réseau à l'aide des paramètres d'un modèle déjà entraîné sur une autre tâche plutôt que par des paramètres aléatoires. Ainsi, si la tâche à apprendre n'est pas trop différente de la tâche déjà apprise, le réseau tire parti de ce pré entraînement, converge généralement plus vite, avec moins de données, ou affiche de meilleures performances. Lorsque la tâche reste la même, on ne parle pas de Transfer Learning mais plutôt de fine-tuning.

Dans le domaine des réseaux de neurones, les paramètres transférés sont les poids du réseau. Ainsi, un réseau entraîné à reconnaître des chats pourra servir d'initialisation à un apprentissage pour reconnaître des chiens. Les couches de traitement d'image extrayant par exemple les oreilles ou la queue des chats pourront être réutilisées et légèrement adaptées afin de fonctionner sur des images de chiens. Il existe plusieurs variantes dans lesquelles on gèle certains des poids du réseau afin de ne réentraîner qu'une partie du réseau.

Ces méthodes ont déjà été utilisées avec succès en imagerie médicale [68] [69] sur des tâches de segmentation [70] ou de classification [71].

1.4 Quantification d'images TDM de TC

Les progrès des algorithmes d'IA et la mise à disposition de bases de données d'imagerie ont permis d'automatiser la quantification des images TDM de patients TC et/ou la prédiction du futur neurologique des patients.

La quantification d'une image TDM de patient TC signifie d'extraction de métriques représentant certaines propriétés de l'image. Ainsi, le calcul des scores d'imagerie introduits section 1.2.9 est une quantification d'image TDM. Pour remédier aux limites de cette quantification manuelle précédemment décrites, on peut quantifier automatiquement une image en extrayant des métriques statistiques calculées automatiquement, soit à partir de l'image entière, soit à partir de régions d'intérêt (ROI) préalablement segmentées, manuellement ou automatiquement. Les métriques extraites peuvent ensuite servir de variables d'entrée à des algorithmes de machine learning de classification, capables de poser un diagnostic, ou bien de prédire le futur neurologique des patients TC. Le déroulement d'une étude d'imagerie classique est illustré sur la Figure 31.

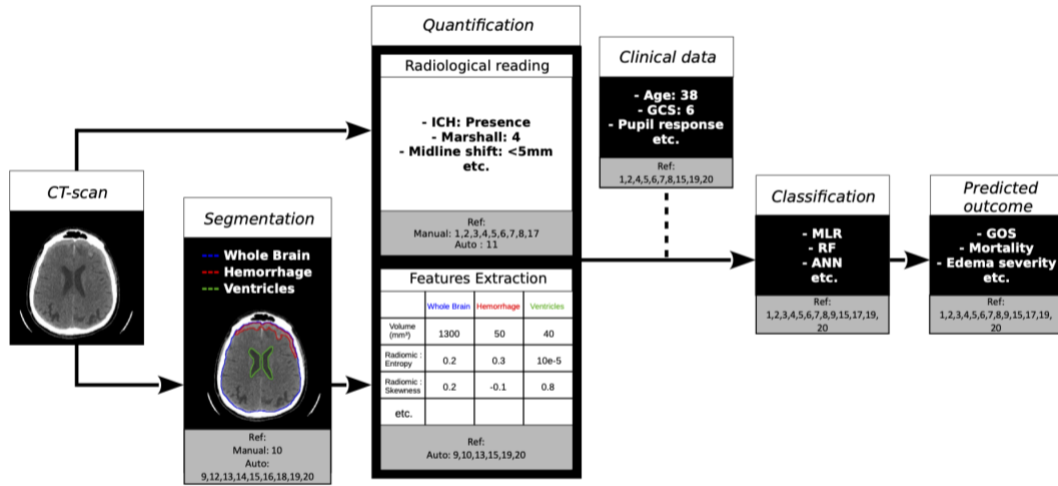


Figure 31. Vue d'ensemble des études d'imagerie classiques. Premièrement on quantifie les images, de manière manuelle ou non, en passant par une segmentation préalable ou non, puis on ajoute ou pas des données cliniques, pour enfin prédire un critère clinique grâce à des algorithmes d'apprentissage supervisé. Les numéros Ref correspondent aux numéros des articles regroupés dans le Tableau 10. CT : Computed Tomography, ICH : Intracranial Hemorrhage, GCS : Glasgow Coma Score, MLR : Multivariate Logistic Regression, RF : Random Forest, ANN : Artificial Neural Network, GOS : Glasgow Outcome Scale. Image tirée de [72]

L'analyse par d'IA d'images TDM de patients TC a fait l'objet d'une publication d'une revue de la littérature publiée début 2021 [72]. Cet article recense et compare les différentes études de classification et de segmentation faisant appel à des algorithmes d'IA. La suite de cette sous partie sera donc largement inspirée de cet article. Les principaux articles recensés dans cette étude bibliographique, ainsi que quelques ajours récents, sont regroupés dans le Tableau 10.

1.4.1 Classification et extraction de métriques

1.4.1.1 Définition et métriques d'évaluation

La classification est la manière dont un algorithme attribue une classe à chaque objet d'entrée, qui est défini par une ou plusieurs variables. Il existe différentes métriques pour évaluer les performances d'une classification. Supposons un algorithme de classification binaire visant à prédire une pathologie et considérons son évaluation sur une base de données de test. On définit comme TP le nombre de vraies prédictions positives, TN le nombre de vraies prédictions négatives, FP le nombre de fausses prédictions positives, et FN le nombre de fausses prédictions négatives. L'exactitude (Acc) de l'algorithme, qui représente le pourcentage de bonnes prédictions, peut être calculée comme suit :

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}$$

Pour séparer les cas de fausse prédiction sur des sujets sains ou pathologiques, on peut définir la Sensibilité (Se), la Spécificité (Sp), et la Précision (Pr) :

$$Se = \frac{TP}{TP + FN}$$

$$Sp = \frac{TN}{TN + FP}$$

$$Pr = \frac{TP}{TP + FP}$$

Modifier le seuil de discrimination d'un algorithme de classification permet de pondérer le coût de la prédiction de la pathologie. Il est en effet plus coûteux de prédire la normalité pour un patient que la pathologie pour un sujet sain. Il permet de diminuer le FN, mais par conséquent d'augmenter le FP, et de modifier le TP et le TN. La modification de ce seuil conduit à de nouvelles valeurs de TP, TN, FP, et FN, et on peut résumer ces valeurs en représentant Se en fonction de (1 - Sp), définissant la courbe récepteur-opérateur (ROC). La mesure de l'aire sous la courbe (AUC) de la courbe ROC représente les performances globales de l'algorithme de classification, quel que soit son seuil de discrimination, et est une métrique largement utilisée pour évaluer les algorithmes de classification. Une AUC de 1 décrit un classifieur parfait, et une valeur de 0,5 signifie que le classificateur ne fait pas mieux que le hasard.

On peut également définir l'exactitude équilibrée, qui se définit comme la moyenne de la sensibilité et de la spécificité. Elle est particulièrement utile sur des bases de données où le nombre d'échantillons des deux classes est déséquilibré. Finalement, on peut calculer la Valeur Prédicative Négative (VPN), qui représente la probabilité d'avoir effectivement un résultat négatif sachant que le test est négatif :

$$VPN = \frac{Sp(1 - Pr)}{Sp(1 - Pr) + Pr(1 - Se)}$$

1.4.1.2 État de l'art

1.4.1.2.1 Quantification manuelle de l'image TDM

Les premières études impliquant des outils d'IA avaient comme objectif de prédire le devenir d'un patient après un TC. Le devenir des patients à 6 et 12 mois est défini comme favorable ou défavorable, c'est-à-dire le décès et le handicap sévère, selon le score GOS. D'un point de vue méthodologique, ces études sont basées sur l'analyse de métriques hétérogènes, représentant la démographie (âge, sexe, etc.), la physiologie (réponse de la pupille à la lumière, score GCS, concentration de glucose dans le sang, etc.), et les TDM quantifiés manuellement par les radiologues en utilisant des scores d'imagerie. Ces paramètres ont été agrégés pour construire des modèles pronostiques à l'aide d'approches de régression logique multivariée, comme l'ont examiné Perel et al. [73]. Deux grandes bases de données de patients souffrant de TC ont produit des modèles prédictifs : CRASH [74] et IMPACT [75] [76] [77]. Ces modèles ont ensuite été validés en externe avec une bonne performance dans la prédiction de la mortalité et du devenir défavorable (AUC entre 0,65 et 0,87) [78]. La prévalence des paramètres a été estimée et, si les principaux prédicteurs sont l'âge ou l'état de conscience général [79], les métriques extraites manuellement à partir des TDM contiennent également des informations cruciales permettant de prédire le devenir du traumatisme crânien [74] [80] [81]. Ces métriques issues de la TDM ont également été évaluées individuellement et ont montré une capacité pronostique importante pour la prédiction du devenir du TC [82] [5].

Plus récemment, des algorithmes complexes ont été évalués au lieu de la régression logique multivariée (MLR). En 2018, 9 algorithmes de machine learning ont été testés sur 232 patients TC afin de prédire le devenir et la mortalité à partir de 14 métriques parmi lesquelles les informations extraites des comptes rendus radiologiques à l'admission et le lendemain de l'accident [83]. Les meilleurs modèles ont été obtenus en utilisant l'algorithme Random Forest (RF), affichant une AUC de 0,895 et en utilisant la régression « ridge », affichant une AUC de 0,875 pour la prédiction de la mortalité, tous deux sur un jeu de test dédié.

D'autres équipes ont étudié les réseaux neuronaux artificiels (ANN) [84]. Les modèles basés sur les ANN ont montré des performances contrastées dans la classification des résultats à partir de métriques quantifiées manuellement, avec des études obtenant des résultats supérieurs à l'état de l'art, affichant une AUC = 0,946 pour la prédiction du GOS dichotomisé (GOS<4 /GOS>3 à 6 mois)[85] ou une Acc = 0,92 pour la prédiction de la mortalité à 6 mois [86], et des études affichant des performances inférieures pour la prédiction de la mortalité hospitalière (AUC = 0,706) [87] ou du résultat dichotomisé (AUC = 0,78) [88]. Ces différences peuvent s'expliquer par l'écart entre le nombre de sujets de l'étude (tous jeux de données

confondus) (565, 2059, 11 026, 12 576, respectivement à l'ordre de citation), la proportion de sujets pathologiques et sains (6, 10, 26, et 47%, respectivement) ou par la méthode de validation (interne, interne, externe, externe, respectivement). Ces différences indiquent que la méthodologie est une étape cruciale pour mesurer réellement les performances des modèles et faire confiance à leurs prédictions [89].

1.4.1.2.2 Quantification automatique de l'image TDM

Outre les paramètres cliniques, le scanner fournit des images du cerveau à haute résolution spatiale qui contiennent beaucoup plus d'informations que celles résumées dans les scores d'imagerie. Plusieurs chercheurs ont récemment commencé à utiliser des algorithmes automatiques pour exploiter cette grande quantité de données dans le but de trouver des biomarqueurs capables de prédire l'issue d'un traumatisme crânien.

1.4.1.2.2.1 Histogrammes de distribution des valeurs Hounsfield dans le parenchyme

A notre connaissance, le premier article à étudier le TC directement sur les images TDM est celui de Kim et al. [8]. Grâce à un pipeline automatique d'extraction de cerveau de sur images TDM pédiatriques de patients TC, les auteurs ont montré que la proportion voxels cérébraux dont la valeur est comprise entre 17 et 24 UH est un bon prédicteur de la sévérité d'un œdème (AUC = 0,85). Ce travail a démontré que (i) les images TDM contiennent des informations quantitatives précieuses liées à l'évolution de l'état du cerveau et (ii) que des outils automatiques peuvent être utilisés pour extraire ces caractéristiques quantitatives.

1.4.1.2.2.2 Métriques radiomique pour quantifier les propriétés de l'image

La distribution des valeurs des voxels, représentée par un histogramme, est une métrique simple pour caractériser quantitativement une image. Mais on peut aller plus loin et extraire des métriques plus complexes pour représenter la forme, la texture ou le contraste d'une image [90]. Ce domaine de recherche est appelé radiomique [91], et alors que sa signification biologique est actuellement discutée [92], il a prouvé sa pertinence dans le diagnostic des tumeurs [93]. Chaque métrique radiomique peut être calculée sur une région d'intérêt (ROI) en 2D ou 3D, ce qui permet d'extraire un grand nombre de métriques caractérisant la ROI. Cette méthode a été utilisée chez des patients souffrant d'un TC [94], suggérant que l'extraction de métriques représentatives de statistiques de premier ordre, de texture et de forme à partir de la ROI de lésion manuellement segmentée pouvait discriminer la nature de la blessure, parmi les hématomes épiduraux (EDH), les hématomes sous-duraux aigus (ASH) et les contusions, avec une Acc = 89,7.

Une autre façon d'utiliser la radiomique consiste à extraire chaque métrique sur une fenêtre glissante qui se déplace sur une image TDM, ce qui conduit à générer

autant de cartes paramétriques que de métriques utilisées. Cette méthode a été utilisée par Muschelli et al. [95] pour segmenter les hémorragies. Cependant, à notre connaissance, elle n'a pas encore été utilisée spécifiquement pour le TC.

1.4.1.2.2.3 Réseau de neurones convolutifs

En vision par ordinateur et en imagerie médicale, en raison de leur capacité à analyser les pixels et leur voisinage, les CNN sont les réseaux les plus utilisés dans les tâches de classification et de segmentation [96] [97]. Dans le domaine des traumatismes crâniens, pour classifier les blessures à partir d'images TDM, Chilamkurthy et al. [6] ont construit plusieurs algorithmes DL (et RF), chacun capable de détecter la présence ou l'absence d'un type de lésion sur une image TDM. Ces algorithmes ont été validés de manière externe avec une AUC de 0,92 à 0,97.

1.4.2 Segmentation d'une image TDM

1.4.2.1 Définition et métriques d'évaluation

La segmentation est, en traitement d'image, une opération visant à diviser une image en 2 ou plusieurs ROI. Supposons un algorithme de segmentation visant à segmenter une lésion cérébrale et son évaluation sur une base de données de test. Pour comparer la segmentation de sortie X à une segmentation de la vérité-terrain Y , on peut calculer le coefficient de similarité de Dice (DSC), défini comme suit, pour $|X|$ le nombre de pixels dans la segmentation X :

$$DSC = 2 \times \frac{|X \cap Y|}{|X| + |Y|}$$

Ce coefficient mesure le nombre de pixels communs entre 2 segmentations par rapport à la somme du nombre de pixels des 2 segmentations. Certains algorithmes sont évalués sur des images comprenant des images sans lésion, donc sans segmentation de vérité terrain. Dans ces cas, si l'algorithme prédit une lésion, même très petite, le DSC est égal à 0 et affecte fortement le DSC moyen, bien que l'erreur de segmentation soit assez faible. C'est pourquoi dans certaines études, les petites lésions sont exclues des mesures de DSC, ce qui conduit à de meilleures valeurs de DSC que les études mesurant le DSC sur toutes les lésions. Il existe plusieurs autres métriques d'évaluation d'une segmentation, comme par exemple la distance de Hausdorff, ou l'index de Jaccard [98].

1.4.2.2 État de l'art

Parmi les études sur le TC, Jadon et al. ont comparé plusieurs architectures de CNN et construit 3 algorithmes, chacun segmentant un type de lésions parmi l'hémorragie intraparenchymateuse (IPH), l'hémorragie extra axiale (EAH), et les contusions hémorragiques à partir d'une image TDM 2D. Leur DSC varie de 0,85 pour l'EAH à 0,95 pour l'IPH. De leur côté, Jain et al. [99] ont développé un algorithme basé sur la segmentation de la MB, de la MG et du LCS pour segmenter les lésions intracrâniennes aiguës (DSC = 0,73) et les ventricules (DSC = 0,70 pour les citernes de plus de 5 ml).

Kuo et al. [100][101] ont démontré la supériorité d'un apprentissage sur des patchs aléatoires de 240*240 pixels et d'une évaluation sur des fenêtres glissantes sur une image TDM pour détecter et segmenter des hémorragies intracrâniennes par rapport à un CNN plus classique, qui prend en entrée une image TDM 2D entière. L'algorithme résultant affiche une AUC de 0,966 sur validation externe et un DSC de 0,72 sur validation interne.

Comme il existe différentes façons de construire un CNN, Yao et al. [7] ont construit un CNN multi-vue destiné à segmenter les hématomes aigus, qui analyse l'image TDM à différentes résolutions pour aider à détecter les petites lésions. Il affiche un DSC de 0,69 sur un jeu de test interne contenant 20 sujets.

Plus récemment, Monteiro et al. [102] ont mis à disposition un modèle nommé « Brain Lesion Analysis and Segmentation Tool for Computed Tomography » (BLAST-CT), destiné à segmenter et à classer quatre types de lésions typiques du TC : 3 hémorragies (IPH, EAH, IVH), ainsi que l'œdème périlésionnel (PO). Ils ont construit un CNN, inspiré de l'architecture de Deepmedic [66], et l'ont entraîné sur plus de 200 images TDM 3D de TC où les quatre types de lésions ont été segmentés par des spécialistes. La segmentation du modèle résultant a été validée sur une cohorte interne de 655 scans, conduisant à un DSC de 0,36 (comme mentionné précédemment, le DSC est très sensible aux petits volumes et pourrait ne pas être la meilleure métrique pour évaluer la segmentation de petites lésions), tandis que la classification a été validée sur une cohorte externe de 490 scans, fournissant une AUC entre 0,83 et 0,95 pour les différents types de lésions à classer.

De manière analogue, en juin 2022 est sorti un article de Inkeaw & al. [103] qui décrit une segmentation automatique de HSD, HED et HIP à partir là encore du réseau de neurones convolutif DeepMedic, atteignant un score de Dice médian global de 0,5 sur une validation interne.

Finalement, une étude disponible en preprint début août 2022 a pu utiliser les segmentations automatiques de BLAST-CT pour ensuite extraire une information de localisation et de type de lésions pour prédire la mortalité d'environ 600 patients[104]. Ce travail étant très proche de l'une des études menée au cours de ma thèse et détaillée dans ce manuscrit, elle sera comparée à nos travaux à la fin du chapitre 4.

Numéro	Année	Auteurs	Tâche	Entrée	Sortie	Nb données	Type de données	Algorithme	Validation	Métrique d'évaluation	Performance	Modèle disponible ?
1/	2008	MRC CRASH Trial Collaborators [74]	Classif	Données cliniques + comptes rendus radiologique	GOS binaire	18517	GCS $\leq$ 14	RLM	E	AUC	77%	Oui
2/	2006	Steyerberg & al. [75]	Classif	Données cliniques + comptes rendus radiologique	GOS binaire	14781	GCS $\leq$ 12	RLM	E	AUC	80%	Oui
3/	2014	Raj & al. [55]	Classif	Comptes rendus radiologique	GOS binaire	869	TC léger, modéré et sévère	RLM	I	AUC	75%	Non
4/	2020	Matsuo & al. [83]	Classif	Données cliniques + comptes rendus radiologique	GOS binaire	232	Compte rendu radiologique anormal	RF	I	AUC	89,5%	Non
5/	2018	Hale & al. [85]	Classif	Données cliniques + comptes rendus radiologique	GOS binaire	565	TC modéré et sévère	ANN	I	AUC	94,6%	Non
6/	2018	Rau & al. [86]	Classif	Données cliniques + comptes rendus radiologique	Mortalité	2059	AIS $\geq$ 3	RLM	I	Acc	93,5%	Non
7/	2016	V.d. Ploeg & al. [87]	Classif	Données cliniques + comptes rendus radiologique	Mortalité	11026	TC modéré et sévère	RLM	E	AUC	76,4%	Non
8/	2020	Gravesteijn & al. [88]	Classif	Données cliniques + comptes rendus radiologique	Mortalité	12576	TC modéré et sévère	GBM	E	AUC	83%	Non
8/	2020	Gravesteijn & al. [88]	Classif	Données cliniques + comptes rendus radiologique	GOS binaire	12576	TC modéré et sévère	ANN	E	AUC	78%	Non
9/	2014	Kim & al. [8]	Classif	Images TDM	Œdème sévère ou modéré	70	TC pédiatrique	Proportion des voxels $\in [17,24]$ UH et tests non paramétriques	ND	AUC	85%	Non
9/	2014	Kim & al. [8]	Classif	Images TDM	Œdème futur ou modéré	70	TC pédiatrique	Proportion des voxels $\in [17,24]$ UH et tests non paramétriques	ND	AUC	75%	Non
10/	2019	Rosa & al. [94]	Classif	Images TDM + segmentation manuelle des lésions	HED + HSD + Contusions	155	Présence de lésion	Métriques radiomique et PLS-DA	I	Acc	89,7%	Non

11/	2018	Chilamkurthy & al. [6]	Classif	Images TDM	HIP + fractures + effets de masse + déviation de la ligne médiane	313809	ND	CNN	E	AUC	92,16 – 97,31%	Non
12/	2020	Jadon & al.[105]	Seg	Images TDM 2D	Hémorragies	40000	ND	CNN	ND	DSC	85,78 – 94,24%	Non
13/	2019	Jain & al. [99]	Seg	Images TDM	Lésions intracrâniennes	144	Center-TBI	CNN	I	DSC	73%	Non
14/	2019	Kuo & al. [100]	Seg	Images TDM	HIP	791	ND	CNN	E	DSC	76,6%	Non
15/	2020	Yao & al. [7]	Seg	Images TDM	Hématomes	828	GCS ∈ [4, 12]	CNN	I	DSC	69,7%	Non
15/	2020	Yao & al. [7]	Classif	Données cliniques + Images TDM	Mortalité	828	GCS ∈ [4, 12]	RF	I	AUC	85,3%	Non
16/	2020	Monteiro & al. [102]	Seg	Images TDM	HIP + HEA + O + HIV	839	Center-TBI	CNN	I	DSC	36%	Oui
16/	2020	Monteiro & al. [102]	Classif	Images TDM	HIP + HEA + O + HIV	490	Center-TBI + CQ500	CNN	E	AUC	83 – 95%	Oui
17/	2021	Khaki & al. [5]	Classif	Scores d'imagerie (Marshall, Rotterdam, Helsinki, Stockholm)	GOS binaire	158	TC admis en réanimation	RLM	I	AUC	70%	Non
18/	2022	Inkeaw & al. [103]	Seg	Images TDM	HSD + HED + HIP	153	Patients suspectés d'un TC à l'admission	CNN	I	DSC	50%	Non
19/	2022	Rosnati & al. [104]	Classif	Données cliniques + Images TDM	GOS binaire	646	TC admis en réanimation (base CREACTIVE)	RF	E	AUC	87,7%	Non
20/	2022	Brossard & al. [106]	Seg	Images TDM	HIP + HSD + HED + HSA + O + HIV + P	41	AIS≤3	CNN	E	DSC	41%	Oui
20/	2022	Brossard & al [106]	Classif	Images TDM	TILsum binaire	41	AIS≤3	RF	E	Acc	83%	Non

Tableau 10.: Résumé et comparaison des principaux articles utilisant les images TDM et des algorithmes d'IA dans le domaine du TC. Légende : Classif – Classification, Seg – Segmentation, TDM – Tomodensitométrie, GOS – Glasgow Outcome Score, HED – Hémorragie Extra Durale, HSD – Hémorragie Sous Durale, HIP – Hémorragie Intra Parenchymateuse, HEA – Hémorragie Extra Axial, O – Œdème, HIV – Hémorragie Intra Ventriculaire, HSA – Hémorragie sous arachnoïdienne, P – Pétéchies, TILsum – Therapy Intensity Level summary, GCS – Glasgow Coma Score, TC – Traumatisme Crânien, AIS – Abbreviated Injury Score, ND – Non Défini, Center-TBI et CQ500 – Bases de donnée publiques contenant des images TDM de patients TC, RLM – Régression Logistique Multivariée, RF – Random Forest, ANN – Réseau de Neurones Artificiels, GBM – Amplification de gradient, UH – Unités Hounsfield, PLS-DA – Analyse discriminante par les moindres carrés partiels, CNN – Réseau de Neurones Convolutifs, E – Externe, I – Interne, AUC – Aire sous la courbe Récepteur-Opérateur (ROC), ACC – Exactitude, DSC – Score de Dice

Chapitre 2 : Objectifs de la thèse

Cette thèse a débuté en octobre 2019, en collaboration entre le GIN et le CHUGA. Si plusieurs études sur le TC ont déjà été menées entre le CHUGA et le GIN, elles se sont limitées soit au stade préclinique, soit à la modalité d'imagerie par résonnance magnétique, rendant compliqué le transfert des outils développés pour ces projets en routine clinique. Afin de coller aux contraintes cliniques, nous nous sommes penchés sur l'analyse des images TDM, bien plus utilisées en routine cliniques pour suivre les patients TC.

Cette thèse a pour objectif de développer l'analyse automatique des images TDM de patients TC, et de trouver de véritables biomarqueurs d'imagerie capable de prédire le futur à court terme d'un patient en réanimation pour TC. Au début de cette thèse, en 2019, peu d'articles avaient étudié la prédiction d'un critère clinique relatif au TC à partir d'une analyse automatique d'images TDM. Cette thèse a donc été pensée comme un travail exploratoire, s'appuyant sur une petite cohorte de patients TC imagés au scanner le jour de l'arrivée au CHU (J0), le lendemain (J1) et 2 jours plus tard (J3), cohorte nommée Radiomic-TBI. A partir de ces données, le but était de prédire l'intensité thérapeutique nécessitée par un patient à court terme après son accident, sans plus de détails sur les méthodes à utiliser. Quelques pistes ressortaient comme l'utilisation des métriques Radiomiques pour quantifier les propriétés des images TDM ou encore la quantification de l'évolution scanographique d'un patient entre J0 et J1. Finalement, nous souhaitions arriver à développer un outil capable de quantifier automatiquement une image TDM de TC, valider sa pertinence sur deux cohortes différentes, et le mettre à disposition des médecins du CHU, afin qu'ils puissent le tester et l'évaluer qualitativement sur des patients de soin courant, et ainsi formuler des hypothèses pour l'améliorer. Ces travaux s'inscrivent ainsi dans le développement de l'analyse d'images TDM sur le TC, qui doit devenir dans les prochaines années un sujet de recherche important du CHUGA et du GIN. Les résultats de ces travaux serviront donc de base aux prochains projets de recherche dans ce domaine.

Hormis un exposé des données utilisées pendant ces 3 ans dans le prochain chapitre, la suite de ce manuscrit présente les résultats probants des expériences menées, c'est-à-dire les résultats qui ont fait ou vont faire l'objet de publications scientifiques, mais aussi certaines pistes en cours d'étude en cours de développement.

En particulier, les chapitres 4, 5 et 6 concernent des études dont les résultats ont été ou sont en cours de publication :

- Chapitre 4 : Développement par transfer learning d'un outil de segmentation automatique de 7 types de lésions sur des images TDM de patients TC.
- Chapitre 5 : Développement d'un outil de quantification automatique du profil lésionnel d'un patient TC qui permet de prédire l'intensité thérapeutique nécessitée par un patient durant sa première semaine d'hospitalisation.
- Chapitre 6 : Validation externe et mise à disposition des deux outils des chapitres 4 et 5.

Les chapitres 7 et 8 détaillent des études en cours, qui doivent être poursuivies et perfectionnées :

- Chapitre 7 : Exploitation de l'évolution lésionnelle des patients afin de prédire l'expansion d'un œdème cérébral.
- Chapitre 8 : Étude de la variabilité introduite sur les images TDM par les différents protocoles d'acquisition.

Finalement, puisque le doctorat est une formation au métier de chercheur, qui est certes composé d'études scientifiques, mais aussi de tout un ensemble de tâches administratives, d'enseignement ou d'encadrement, le dernier chapitre présentera les principales tâches extrascientifiques mais indispensables au métier de chercheur que j'ai pu mener.

Chapitre 3 : 3 cohortes utilisées en 3 ans

Les études présentées dans ce manuscrit font appel à 3 différentes cohortes cliniques, une cohorte prospective (Radiomic-TBI) et deux rétrospectives, (RadioxyTC, et CT-Harmo). Pour chacune de ces 3 cohortes, nous détaillons dans ce chapitre les critères d'inclusions, les données récupérées, et la manière dont nous avons utilisé ces données.

3.1 Radiomic-TBI : 29 patients à 3 temps, 84 images TDM

La cohorte prospective Radiomic-TBI (Numéro d'autorisation : NCT04058379) est composée de 30 patients traumatisés crâniens admis au CHUGA entre Janvier 2020 et Avril 2021. Cette cohorte sera démographiquement caractérisée au chapitre 5.

3.1.1 Critères d'inclusion

Les critères d'inclusions incluent, en plus du consentement éclairé du patient :

- Age du patient supérieur à 18 ans
- Patient victime d'un traumatisme crânien fermé
- Admission primaire au CHUGA
- Image TDM initiale retrouvant des lésions cérébrales classées au moins 3 par l'Abbreviated Injury Score (AIS), correspondant à une lésion visible sur l'image
- Patient hospitalisé en réanimation pour une durée attendue d'au moins 48h
- Patient affilié à un système de sécurité sociale

Parmi les 30 patients inclus, l'un fut inclus à tort car admis au préalable dans un autre centre que le CHUGA, comme illustré dans le diagramme de flux de l'étude Figure 32.

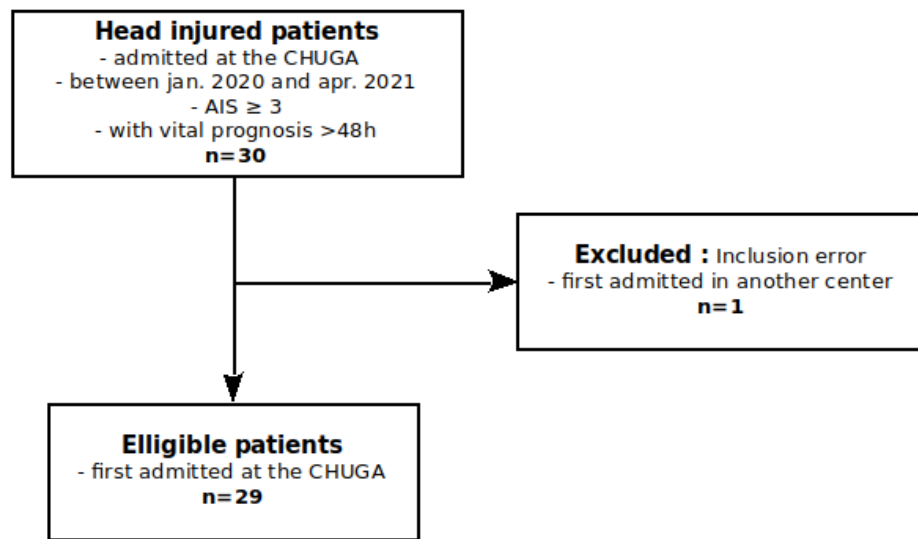


Figure 32. Diagramme de flux de l'étude Radiomic-TBI

3.1.2 Données recueillies et stockage

Sur les 29 patients restants, les données recueillies incluent :

- Image TDM le jour de l'admission au CHUGA : J0
- Image TDM le lendemain : J1
- Image TDM à J3
- Données cliniques à J0 : Age, Sexe, Poids, Taille, GCS à l'admission, Pression artérielle moyenne (en mmHg), Taux d'Hémoglobines (en g/l), Présence d'antiagrégants, score de Marshall, score de Rotterdam
- Données descriptives de l'intensité thérapeutique quotidienne (Therapeutic Intensity Level Summary - TILsum) de J0 à J7.
- Données de mesure du devenir clinique : Statut-vital à 3 mois et GOSE à 6 mois après l'accident.

Le recueil de ces données est illustré Figure 33. Au total, 84 images TDM furent recueillies parmi les 29 patients, car 3 patients n'ont pas été imaged à J3 par manque d'accessibilité du scanner ou par erreur.

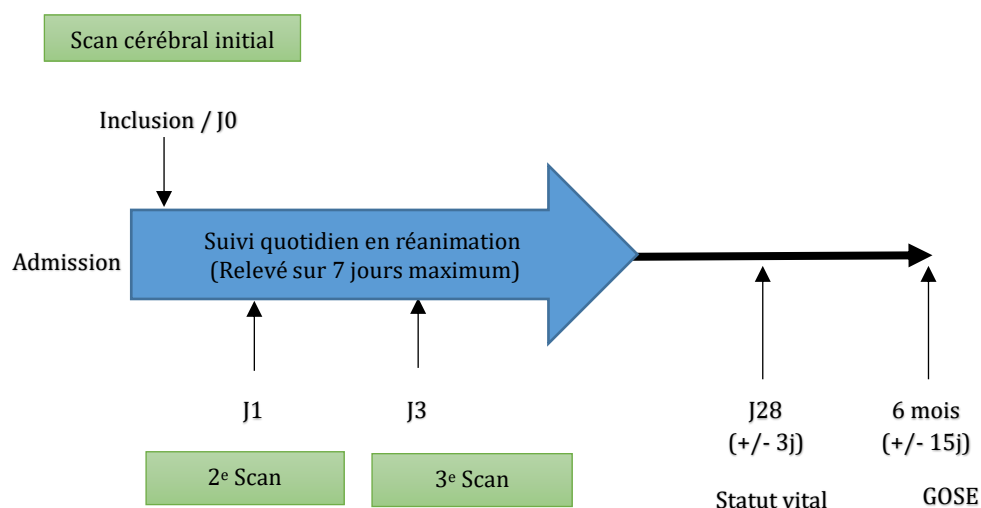


Figure 33. Schéma des données recueillies sur les patients de la cohorte Radiomic-TBI. Source : Marion Richard.

Ces images ont été stockées sur Shanoir, une plateforme d'anonymisation, d'organisation, et de stockage sécurisé de données d'imagerie médicales [107]. Elle est certifiée « Hébergeur de Données de Santé » et a été développée par l'Institut National de recherche en informatique et en automatique (Inria) avec l'aide du plan France Life Imaging (FLI).

3.1.3 Utilisation

La richesse de cette base de données réside dans son aspect longitudinal et dans le suivi quotidien de l'intensité thérapeutique demandée par un patient en réanimation. Cette cohorte a servi à :

- Réentraîner un algorithme de segmentation automatique des lésions sur les images TDM (Chapitre 4).
- Développer un algorithme de prédiction du niveau d'intensité thérapeutique maximum nécessité par un patient au cours de la première semaine en réanimation à partir de métriques extraites de l'image TDM acquise à J0 (Chapitre 5).
- Valider de manière interne la segmentation et la classification (Chapitre 5).
- Etudier l'évolution longitudinale des patients entre J0, J1 et J3 (Chapitre 7).

3.2 RadioxyTC : 12 patients, 12 images TDM

La cohorte rétrospective RadioxyTC (enregistrée sous le numéro F20220207212747 sur le Health Data Hub) est une sous cohorte de l'étude OxyTC menée au CHUGA. Elle est composée de 12 patients traumatisés crâniens admis au CHUGA entre août 2018 et Avril 2021. Cette cohorte sera démographiquement caractérisée au chapitre 6.

3.2.1 Critères d'inclusion

Les critères d'inclusions incluent, en plus du consentement éclairé du patient :

- Patient déjà inclus dans l'étude OxyTC
- Patient non inclus dans l'étude Radiomic-TBI

3.2.2 Données recueillies et stockage

Sur les 12 patients restants, les données recueillies incluent :

- Image TDM le jour de l'admission au CHUGA : J0
- Données cliniques à J0 : Age, Sexe, Poids, Taille, GCS à l'admission, Pression artérielle moyenne (en mmHg), Taux d'Hémoglobines (en g/l), Présence d'antiagrégants, score de Marshall, score de Rotterdam
- Données descriptives de l'intensité thérapeutique quotidienne (Therapeutic Intensity Level Summary - TILsum) de J0 à J4.
- Le compte rendu d'hospitalisation en réanimation, afin d'estimer le TILsum entre J5 et J7.

Ces données ont été stockées sur des disques durs internes au laboratoire.

3.2.3 Utilisation

Cette cohorte a servi de validation externe aux algorithmes développés sur la cohorte Radiomic-TBI (Chapitre 6). Les patients de cette cohorte sont déjà inclus dans l'étude OxyTC, qui recueille entre autres nos données d'intérêt, ce qui a facilité l'accès aux données.

3.3 CT-Harmo : 987 patients, 987 images TDM

La cohorte rétrospective CT-Harmo (enregistrée sous le numéro F20201127123438 sur le Health Data Hub) est composée de 987 patients traumatisés crâniens admis au CHUGA entre janvier 2019 et juillet 2020.

3.3.1 Critères d'inclusion

Les critères d'inclusions incluent, en plus du consentement éclairé du patient :

- Age du patient supérieur à 18 ans
- Patient admis au déchocage
- Patient imagé par TDM au CHUGA

3.3.2 Données recueillies et stockage

Sur les 987 patients restants, les données recueillies incluent :

- Image TDM le jour de l'admission au déchocage
- Compte rendu radiologique, rédigé par le radiologue qui a analysé l'image TDM

Ces images TDM ont été stockées sur Shanoir, comme les données de la cohorte Radiomic-TBI.

3.3.3 Utilisation

La richesse de cette base de données réside d'une part dans sa taille, et d'autre part dans la nature des images. En effet, une partie des images TDM de cette base de données sont des images de cerveau sain. En effet, la plupart des patients passés au déchocage subissent systématiquement une imagerie TDM mais ne sont pas admis au déchocage pour un traumatisme crânien ou une pathologie cérébrale. Du fait de l'aspect invasif de l'imagerie TDM, les images TDM de cerveaux sains sont rares, et donc précieuses. Cette cohorte a servi à :

- Étudier la variabilité introduite sur les images TDM par les différents protocoles d'acquisition. Cette problématique a fait l'objet du stage de Master 2 de Vaëa TESAN que j'ai co-encadré durant le premier semestre 2021 (Chapitre 8).
- Entraîner des réseaux de neurones convolutifs (CNN).

La première cohorte à laquelle nous avons eu accès est la cohorte Radiomic-TBI. Le projet s'est construit autour d'elle et c'est donc naturellement celle que nous avons utilisé en premier, en nous permettant de développer un algorithme de segmentation automatique de lésions cérébrales à partir d'images TDM, détaillé au chapitre suivant.

Chapitre 4 : Segmentation automatique des lésions cérébrales

Afin de quantifier une image TDM de patient TC, on peut essayer de résumer son profil lésionnel en recensant les différentes lésions cérébrales, c'est-à-dire en segmentant et en identifiant les lésions visibles à l'œil nu sur les images TDM. Cette opération, faite manuellement, est longue et fastidieuse et n'offre donc pas une alternative pertinente aux scores d'imagerie (Marshall, Rotterdam ; cf chapitre 1) qui eux aussi quantifient à leur manière le profil lésionnel des patients. En effet la segmentation manuelle en 3 dimensions dessinée au voxel près sur les images TDM conduit, à cause de la haute résolution des images TDM et des larges lésions induites par un TC, à de nombreuses heures de travail [108][109].

Si cette segmentation pouvait être automatique, elle permettrait d'offrir une information quantitative aux radiologues et pourrait servir d'aide à la décision lors d'un diagnostic de TC. Dans la littérature scientifique, les outils qui affichent les meilleures performances sur la segmentation de lésions à partir d'images médicales sont basés sur les réseaux de neurones convolutifs. Dans le cas du traumatisme crânien et des images TDM, le réseau BLAST-CT, publié dans The Lancet Digital Health en 2020, est le modèle disponible le plus avancé. Le but du développement de ces méthodes est de fournir des segmentations automatiques les plus proches possibles des segmentations manuelles.

Ce chapitre, présente le développement d'un outil de segmentation automatique des lésions traumatiques observables par TDM et s'organise comme décrit ci-dessous :

- Nous avons tout d'abord cherché à évaluer le réseau de neurones convolutifs BLAST-CT sur les données de la cohorte Radiomic-TBI. Pour rappel, ce réseau publié en 2020 segmente automatiquement 4 lésions du TC sur des images TDM (cf. chapitre 1).
- Dans un deuxième temps, devant les limites observées et quantifiées de ce réseau à segmenter 4 types de lésions sur nos données, nous avons cherché à améliorer ses performances à l'aide d'un réentraînement par une approche nommée « fine-tuning ».
- Finalement, nous avons fait évoluer ce réseau afin de segmenter automatiquement non plus 4 mais 7 types de lésions cérébrales post-TC observables en TDM. Pour ce faire, nous avons modifié la dernière couche de l'architecture de ce réseau et nous l'avons entraîné par une approche d'apprentissage par transfert, ou « Transfer Learning ».

4.1 Évaluation de BLAST-CT sur nos données

4.1.1 Objectif

La première tâche fut d'évaluer l'algorithme BLAST-CT sur nos données, afin de conclure sur la pertinence de son utilisation pour quantifier automatiquement les lésions cérébrales post-TC sur une image TDM.

Si les principaux résultats sont détaillés dans cette partie, une étude plus poussée des performances de BLAST-CT a été menée par Dr Jules-Arnaud de Busschère dans le cadre de sa thèse de médecine que j'ai co-encadrée [110].

4.1.2 Matériel et méthodes

4.1.2.1 Données utilisées

Pour cette étude nous avons utilisé la cohorte Radiomic-TBI : 84 images TDM de 29 patients TC admis au CHUGA (cf Chapitre 3). Sur ces 84 images, 7 différents types de lésions ont été segmentées manuellement par Dr Jules-Arnaud de Busschère, puis corrigées par Dr Arnaud Attyé, respectivement anesthésiste et neuroradiologue au CHUGA avec 3 et 10 années d'expérience. Les lésions segmentées sont :

- L'hématome intra parenchymateux (HIP)
- L'hémorragie sous durale (HSD)
- L'hémorragie extra durale (HED)
- L'hémorragie intra ventriculaire (HIV)
- L'hémorragie sous arachnoïdienne (HSA)
- Les pétéchies (P)
- L'œdème (O)

Parmi ces 7 classes, on peut regrouper l'HSD, l'HED et l'HSA en Hémorragie extra axiale (HEA), et fondre les pétéchies avec les HIP, ce qui conduit à une segmentation à 4 classes, qui regroupe les mêmes voxels, mais qui est moins précise sur le type de lésions.

Dans la suite du manuscrit, et comme illustré sur le Tableau 11, on parlera soit de segmentation à :

- 4 classes : HIP+P, HEA (HSD + HED + HSA), HIV et O qui correspond aux 4 classes publié par Monteiro et al.
- 7 classes : HIP, HSD, HED, HIV, HSA, P et O

Segmentation à 4 classes	Segmentation à 7 classes
Hémorragie intra parenchymateuse (HIP)	Hémorragie intra parenchymateuse (HIP)
	Pétéchies (P)
Œdème (O)	Œdème (O)
Hémorragie intra ventriculaire (HIV)	Hémorragie intra ventriculaire (HIV)
Hémorragie extra axiale (HEA)	Hémorragie sous durale (HSD)
	Hémorragie extra durale (HED)
	Hémorragie sous arachnoïdienne (HSA)

Tableau 11. Résumé et lien entre les lésions de la segmentation à 4 classes et à 7 classes.

Les segmentations manuelles ont été tracées sur le logiciel ITK-SNAP, à partir des images au format NIfTI à leur résolution initiale. Pour faciliter la segmentation, qui est une tâche longue et fastidieuse, les images TDM d'un même patient ont été automatiquement recalées sur la première image TDM de ce patient, acquise à l'arrivée au CHUGA, grâce à l'algorithme FLIRT d'FSL. Ainsi, la segmentation d'une image TDM d'un patient a pu être aidée par les images précédentes (ou suivantes) de ce même patient. Les segmentations ont ensuite été rééchantillonnées à 1mm³ afin de correspondre à la résolution spatiale d'entrée du réseau BLAST-CT.

Nous avons réparti les 84 images TDM de la cohorte Radiomic-TBI en un jeu d'entraînement, de validation, et de test, selon la répartition décrite dans le Tableau 12. Pour des questions d'indépendance entre les jeux de données, toutes les images TDM d'un patient font obligatoirement partie du même jeu de données. De plus, grâce à la segmentation manuelle associée à ces images, ce partage a été choisi afin d'obtenir une répartition équilibrée du volume de chaque type de lésion dans les 3 jeux. La répartition des patients est la suivante :

- Jeu d'entraînement : P01, P05, P06, P12, P13, P14, P15, P17, P18, P19, P20, P22, P24, P25, P26, P27, P30
- Jeu de validation : P03, P07, P09, P10, P23, P28
- Jeu de test : P02, P04, P08, P11, P16, P29

Les images TDM ont été extraites du PACS du CHUGA au format DICOM et converties au format NIfTI grâce au convertisseur « MRI file Manager » (<https://irmage.univ-grenoble-alpes.fr/services/traitement-dimages/mri-file-manager>) inclus dans le logiciel MP3 [111]. Les images ont ensuite été rééchantillonnées à 1mm³, afin d'avoir la même résolution que les segmentations manuelles.

	Jeu d'entraînement	Jeu de validation	Jeu de test
Nombre de patients	17	6	6
Nombre d'images TDM	50	17	17
HIP Volume % du dataset Nombre d'images contenant la lésion	475,6 cm3 68% (n=36)	109,1 cm3 16% (n=14)	109,9 cm3 16% (n=12)
HSD	810,0 cm3 61% (n=39)	229,5 cm3 17% (n=14)	282,7 cm3 21% (n=13)
HED	235,1 cm3 65% (n=15)	66,2 cm3 18% (n=1)	59,6 cm3 17% (n=3)
HIV	12,2 cm3 64% (n=17)	3,3 cm3 17% (n=7)	3,7 cm3 19% (n=6)
HSA	211,5 cm3 68% (n=34)	49,9 cm3 16% (n=17)	48,0 cm3 16% (n=12)
Pétéchies	5,6 cm3 62% (n=15)	1,9 cm3 20% (n=8)	1,6 cm3 18% (n=11)
Œdème	1177,4 cm3 62% (n=37)	334,2 cm3 18% (n=13)	373,9 cm3 20% (n=11)
Volume lésionnel total	2927,6 cm3 64% (n=50)	794,2 cm3 17% (n=17)	879,4 cm3 19% (n=17)

Tableau 12. Caractérisation des jeux d'entraînement, de validation et de test utilisés au cours de ce chapitre. Les données sont exprimées en Volume de lésion (cm3), en pourcentage du volume de la lésion dans le jeu de données en question par rapport à l'ensemble des données (%), et en nombre d'images TDM contenant cette lésion (n=).

4.1.2.2 Réseau de segmentation multi-classe automatique : BLAST-CT

Comme mentionné au chapitre 1, BLAST-CT est un réseau de neurones profonds capable de segmenter et d'identifier 4 types de lésions cérébrales à partir d'images TDM de patients TC, publié en 2020 dans le journal Lancet Digital Health [102]. Les 4 lésions segmentées sont l'hémorragie intra-parenchymateuse (HIP), l'hémorragie extra-axiale (HEA), l'œdème périlésionnel (O) et l'hémorragie intra-ventriculaire (HIV).

4.1.2.2.1 Architecture

L'architecture de ce réseau est de type DeepMedic, présenté au Chapitre 1. L'image d'entrée est analysée par 3 voies : l'une à la résolution 1mm³, considérée comme la résolution initiale, l'une sur l'image sous échantillonnée d'un facteur 3, et l'une sur l'image sous échantillonnée d'un facteur 5. Le nombre de couches convolutives est de 8 pour les 3 voies, et les kernels de convolution sont de taille 3x3x3. Le réseau segmente l'image en patches de sortie de taille 30x30x30, ce qui conduit selon l'équation définie section 1.3.5.5 à des patches d'entrée de taille 110x110x110. La taille des couches denses est de 250x30x30x30 pour aboutir à 5 patches de taille 30x30x30 contenant chacun la probabilité d'appartenance de chaque voxel à chacune des 5 classes (Tissu sain, HIP, HEA, O, HIV). Afin de tenir compte des dernières avancées dans le domaine de l'apprentissage profond, Miguel Monteiro a légèrement modifié cette architecture en ajoutant des connexions résiduelles [112], des blocks de pré-activation [113], et une normalisation par batch [114], que nous n'aborderons pas ici.

4.1.2.2.2 Entraînement

Le réseau BLAST-CT a été entraîné sur près de 200 images TDM de TC provenant de multiples centres. L'apprentissage étant supervisé, des segmentations manuelles ont été délimitées par des spécialistes afin d'entraîner le réseau. Le prétraitement subi par les images contient un rééchantillonnage à 1mm³, un seuillage entre -15 et 100 UH, et une normalisation des voxels entre -1 et 1.

L'entraînement consiste en 1200 epochs, durant lesquelles à chaque epoch, et pour chaque image du jeu d'entraînement, 40 voxels de l'image sont sélectionnés autour desquels 40 patches de 110x110x110 voxels sont extraits de l'image. Autour des mêmes voxels sur les images des segmentations manuelles, des patches de 30x30x30 sont extraits. Afin de pallier la sur-représentation des patches ne contenant aucune lésion cérébrale, les 40 voxels sélectionnés initialement ne sont pas choisis complètement aléatoire mais selon une probabilité d'appartenance aux lésions à segmenter.

L'entraînement comprend une phase d'augmentation de données, qui consiste à modifier légèrement et aléatoirement les patches en y ajoutant un bruit aléatoire afin de rendre l'algorithme résultant plus robuste aux perturbations des images.

Ainsi, à chaque epoch, une fonction de coût, en l'occurrence la fonction CrossEntropy, est calculée entre la prédiction du réseau et la vérité terrain. Cette fonction mesure l'erreur du réseau, et le but de l'entraînement est de minimiser cette fonction. Les poids du réseau sont ensuite mis à jour par rétro propagation jusqu'à l'epoch suivante.

Toutes les 100 epochs a lieu une validation, c'est-à-dire un calcul de la fonction de coût sur un jeu de données de validation. Cette étape permet de suivre l'apprentissage du réseau et notamment de vérifier l'absence de surapprentissage.

4.1.2.2.3 Combinaison des prédictions

Lors d'un apprentissage classique, les poids initiaux du réseau sont aléatoires. Cette initialisation peut avoir une influence sur l'état final du réseau ainsi que sur sa convergence. Ainsi, afin de limiter l'aspect stochastique de la prédiction, le réseau BLAST-CT a été entraîné 12 fois avec une initialisation différente, afin d'aboutir à 12 modèles différents. La prédiction finale de BLAST-CT est ainsi obtenue en combinant les 12 prédictions des 12 modèles. Cette combinaison est tout simplement la moyenne des probabilités d'appartenance à chacune des classes à segmenter pour chacun des 12 modèles.

4.1.2.2.4 Application et évaluation

Les 12 réseaux CNN entraînés du modèle BLAST-CT étant généreusement mis à disposition (<https://github.com/biomed-mira/blast-ct>), nous avons pu facilement l'appliquer sur nos données et en particulier sur le jeu de test. Nous avons ainsi pu calculer le score de DICE entre les segmentations manuelles à 4 classes et les segmentations issues de BLAST-CT.

Afin de faciliter la lecture, dans ce chapitre, nous parlerons de « CNN » lorsque nous ferons référence à un réseau entraîné, et de « Modèle » lorsque nous ferons référence à un algorithme combinant plusieurs CNN pour effectuer une prédiction globale. Exemple : « Les 12 CNN entraînés forment le modèle BLAST-CT ».

4.1.3 Résultats

D'après la publication initiale du réseau de neurones, sur un jeu de test interne de près de 600 images TDM, BLAST-CT affiche des scores de DICE moyens de 0.376, 0.287, 0.277, 0.127 pour respectivement l'HIP, HEA, l'œdème, et l'HIV. Globalement, le score de DICE sur une lésion quelconque est de 0.353. La comparaison entre les résultats obtenus sur notre jeu de donnée test et ceux publiés est résumée sur le Tableau 13. Une illustration visuelle sur un patient issu de la cohorte Radiomic-TBI de la segmentation obtenue grâce à BLAST-CT et de la segmentation manuelle associée est disponible Figure 34.

	Ensemble des lésions	HIP	HEA	Œdème	HIV
BLAST-CT sur notre jeu de test	0.34 (0.05)	0.24 (0.04)	0.33 (0.06)	0.11 (0.04)	0.10 (0.04)
BLAST-CT publication	0.35 (0.01)	0.38 (0.02)	0.29 (0.01)	0.28 (0.02)	0.13 (0.01)

Tableau 13. Moyenne et erreur standard des scores de DICE entre la segmentation automatique BLAST-CT et la segmentation manuelle obtenus sur notre jeu de test et sur un jeu de test contenant 599 images TDM (Résultats tirés de la publication de BLAST-CT).

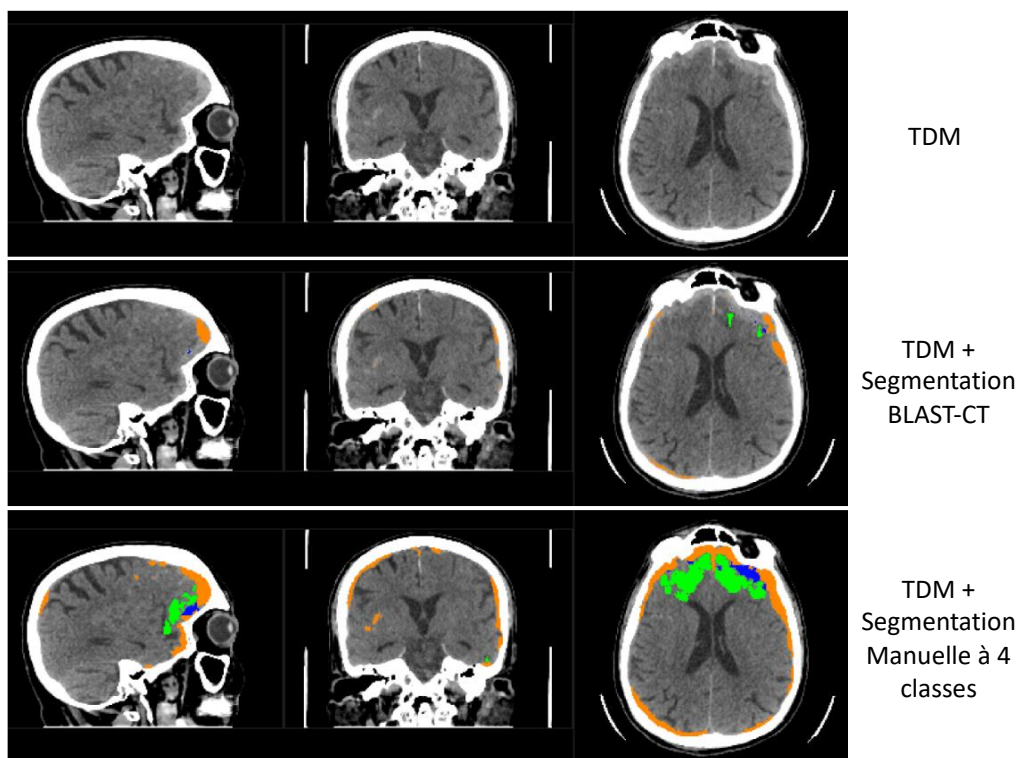


Figure 34. Image TDM d'un patient de la cohorte Radiomic-TBI, issu du jeu de test. De haut en bas : image TDM du patient numéro 8 acquise à J0, TDM et segmentation issue de BLAST-CT superposée, TDM et segmentation manuelle superposée. Légende couleur : Bleu – HIP, Orange – HEA, Vert – Œdème, Rouge – HIV.

4.1.4 Interprétation

Si le score de DICE moyen sur l'ensemble des lésions obtenu sur nos données est du même ordre de grandeur que celui affiché dans la publication, la performance de BLAST-CT diminue pour la segmentation de l'œdème et de l'HEA sur nos données par rapport aux résultats présentés dans l'article de BLAST-CT. Cette différence entre le score de Dice moyen et celui sur lésions individuelles peut s'expliquer par une mauvaise identification des lésions.

Visuellement, BLAST-CT semble sous-estimer la plupart des lésions. Cette différence de performance sur nos données pourrait s'expliquer par le fait que nos données proviennent d'un centre dont ne provient aucune des images utilisées pour entraîner BLAST-CT. Cette hypothèse impliquerait l'existence d'une variabilité intersites sur les images de TDM ou alors d'une différence de nature de données, qui pourrait s'expliquer par des lésions plus ou moins sévères ou bien par une différence de population entre nos données et celles utilisées pour l'entraînement de BLAST-CT. Une autre hypothèse serait une différence dans la manière de segmenter manuellement les lésions cérébrales entre nos segmentations et celles utilisées pour l'entraînement de BLAST-CT. Cette dernière hypothèse est difficilement vérifiable car les segmentations utilisées par BLAST-CT ne sont pas mises à disposition par les auteurs.

4.2 Amélioration de la segmentation automatique à 4 classes par « fine-tuning »

4.2.1 Objectif

Face aux mauvaises performances de BLAST-CT sur nos données, notamment en ce qui concerne la segmentation de l'œdème, qui est l'une des lésions les plus importantes dans l'évolution du TC car elle conditionne grandement la prise en charge, nous avons cherché à améliorer ces performances sur nos données, provenant d'un nouveau site. Plutôt que d'entraîner un nouveau réseau de neurones à partir de zéro, ce qui nous apparaissait voué à l'échec vu le faible nombre de données dont nous disposons, nous avons voulu tirer parti du réseau de BLAST-CT déjà entraîné afin de l'améliorer. Nous nous sommes ainsi tournés vers le transfer learning, qui permet d'apprendre à un réseau déjà entraîné une nouvelle tâche. Ici, nous souhaitons adapter le réseau de BLAST-CT à nos données. La tâche restant la même, on nomme plutôt cette approche « fine-tuning ».

Dans cette expérience, nous avons donc réentraîné le réseau de BLAST-CT par fine-tuning sur nos données, mais nous avons aussi entraîné un réseau en partant de 0, c'est-à-dire de poids aléatoirement initialisés, sur nos données, afin d'estimer de façon objective l'impact du fine-tuning sur les segmentations à 4 classes. Finalement, ces 2 modèles résultants ont été évalués sur le jeu de test et comparés avec les performances de BLAST-CT.

Notre hypothèse est que la segmentation automatique du modèle BLAST-CT sera plus proche de la segmentation manuelle que celle du modèle entraîné à partir de poids aléatoirement initialisés, mais sera moins proche de la segmentation manuelle qu'avec le modèle entraîné par « fine tuning ».

4.2.2 Matériel et méthodes

4.2.2.1 Données utilisées

Pour cette expérience, nous avons utilisé les mêmes données que pour l'évaluation de la performance de BLAST-CT, c'est-à-dire la cohorte Radiomic-TBI répartie en jeux d'entraînement, de validation, et de test comme détaillé précédemment lors de l'étude sur l'évaluation de BLAST-CT.

4.2.2.2 Réentraînement - Fine tuning

Pour réentraîner le réseau de BLAST-CT, nous avons tout d'abord téléchargé les 12 CNN différemment initialisés et pré-entraînés qui composent la prédiction finale de BLAST-CT. Ces CNN sont accessibles sur <https://github.com/biomed-mira/blast-ct>. Ensuite, nous avons développé un script d'entraînement, en utilisant le framework MONAI [115], basé sur PyTorch, en suivant la documentation de BLAST-CT et en exploitant son dépôt github. Nous avons choisi MONAI pour profiter des transformations d'images disponibles et afin de s'inscrire dans un projet maintenu et développé spécialement pour le traitement d'images médicales par apprentissage profond. Ainsi nous avons codé le prétraitement des données, le chargement du réseau DeepMedic et l'entraînement par rétropropagation. Nous avons conservé la même fonction de coût (CrossEntropy) et le même optimiseur (RMSprop) que BLAST-CT. La différence principale avec l'entraînement de BLAST-CT est notre critère d'arrêt et la sélection du CNN final. En effet, BLAST-CT effectue 1200 epochs quoi qu'il arrive. Nous avons préféré calculer la fonction de coût à chaque étape de validation et introduire une condition d'arrêt de l'entraînement sur la variation de cette fonction de coût. Si pendant 100 validations la fonction de coût de validation n'atteint pas un nouveau minimum, l'entraînement s'arrête. Étant donné que nous effectuons une validation toutes les 2 epochs, l'entraînement est ainsi arrêté 200 epochs après la plus basse valeur de la fonction de coût de validation. Concernant la sélection du CNN final, BLAST-CT conserve le dernier CNN, obtenu après 1200 epochs. Nous avons préféré calculer le score de DICE lors de chaque validation (sur les patches issus du jeu de validation) et conserver le CNN présentant les meilleurs scores de DICE.

Nous avons ainsi entraîné 24 CNN selon notre procédure d'entraînement :

- Les 12 CNN de BLAST-CT ont été réentraînés pour former le modèle nommé fin tuning 4 classes ou « FT_4C ».
- 12 CNN initialisés avec 12 jeux de poids aléatoires ont été entraînés pour former le modèle nommé « Random4C ».

Les entraînements des 3 modèles étudiés dans cette expérience (les 2 entraînés plus le modèle BLAST-CT de la publication) sont résumés dans le Tableau 14.

Nom du modèle	Initialisation des poids du réseau	Entraîné sur nos données ?
1/ BLAST	Poids de BLAST-CT	NON
2/ Random4C	Aléatoire	OUI
3/ FT_4C	Poids de BLAST-CT	OUI

Tableau 14. Initialisation et entraînement des modèles étudiés.

4.2.2.3 Exécutions sur une grille de calcul : GRICAD

Chaque entraînement dure en moyenne 30 heures sur une carte graphique Nvidia Tesla A100 (32Go). Afin d'accélérer la campagne de calcul, nous avons utilisé la grille de calcul du mésocentre Grenoblois appelé GRICAD (<https://gricad-doc.univ-grenoble-alpes.fr/hpc/>), qui dispose de plusieurs GPU. Les entraînements ont ainsi pu être lancés en parallèle, réduisant grandement de temps total de la campagne de calcul.

4.2.2.4 Combinaison des modèles

Appliqués au jeu de données de test, chacun des 12 CNN obtenus fournit pour chaque voxel de l'image TDM, une probabilité d'appartenance à chacune des 5 classes possibles (les 4 lésions plus l'absence de lésion). La classe ayant la plus forte probabilité est ensuite attribuée au voxel en question, ce qui permet d'aboutir à la segmentation lésionnelle de l'image.

Afin de combiner les segmentations des 12 CNN, Monteiro & al. ont moyenné les probabilités d'appartenance aux 5 classes pour chaque voxel, et attribué au voxel la classe de probabilité maximale. Cette méthode de combinaison sera nommée pour la suite du chapitre la « Combinaison moyennée » (Average_Probability).

Dans le cadre de cette thèse nous avons également testé plusieurs autres méthodes de combinaison et évalué leurs performances. Les autres méthodes testées sont :

- « Combinaison par vote majoritaire » (Majority_Voting) : Chacun des 12 CNN attribue une classe à chacun des voxels et c'est la classe majoritaire qui l'emporte lors de la prédiction finale. S'il y a égalité du nombre de votes, la première classe l'emporte (fonction python `argmax`)
- « Combinaison par moyenne pondérée ordonnée » (Ranked_Averaging) : On attribue à chacune des probabilités d'un CNN un coefficient C et on moyenne les probabilités pondérées des 12 CNN. On attribue ensuite au voxel sa classe de probabilité maximale. Chaque CNN est classé par ordre croissant en fonction du score de DICE moyen obtenu par le CNN lors de la validation de fonction de coût minimale. Le coefficient C est ensuite égal à la position du CNN dans le classement divisé par la somme des entiers jusqu'au nombre total de CNN. Ainsi, le réseau affichant le moins bon DICE de validation sera pondéré par le coefficient égal à $\frac{1}{1+2+\dots+11+12}$, soit $\frac{1}{78}$. A l'inverse, le CNN affichant le meilleur DICE de validation sera pondéré par le coefficient égal à $\frac{12}{78}$.
- « Combinaison par moyenne pondérée » (Weighted_Averaging) : On attribue à chacune des probabilités d'un CNN un coefficient et on moyenne les probabilités pondérées des 12 CNN. On attribue ensuite au voxel sa classe de probabilité maximale. Le coefficient est égal au score de DICE moyen obtenu par le CNN lors de la validation de fonction coût minimale.

4.2.2.5 Méthode d'évaluation

Une fois les segmentations automatiques obtenues sur le jeu de données de test, nous les avons comparées à la segmentation manuelle. Pour quantifier les différences, nous avons utilisé le score de Dice, défini au chapitre 1. Nous avons ainsi calculé ce score pour chacune des images contenues dans le jeu de test, et pour chacune des 4 lésions mais aussi sur l'ensemble de tous les types de lésions, c'est-à-dire sur le tissu lésionnel global, ce qui permet d'étudier la segmentation sans l'identification des lésions. Finalement, nous avons calculé le score de Dice moyen sur chacune des lésions et sur le tissu lésionnel global.

Dans 3 cas particuliers, le score de Dice peut prendre des valeurs extrêmes et ainsi perturber le calcul de la moyenne :

- Lorsque la segmentation manuelle et la segmentation automatique sont vides, le score de Dice n'est pas défini, mais pourrait être fixé à 1 car la prédiction est parfaite. Nous avons néanmoins décidé de retirer dans le cas ce score dans le calcul de la moyenne.
- Lorsque la segmentation manuelle est vide mais pas la segmentation automatique, le score de Dice est égal à 0, et ce, quel que soit le volume segmenté automatiquement. Puisque nous pensons qu'il faut dans le domaine médical privilégier les faux positifs aux faux négatifs, nous avons dans ce cas décidé là encore de retirer ce score dans le calcul de la moyenne.
- Lorsque la segmentation automatique est vide mais pas la segmentation manuelle, le score de Dice est égal à 0, et ce, quel que soit le volume segmenté manuellement. Puisque nous pensons qu'il faut dans le domaine médical privilégier les faux positifs aux faux négatifs, nous avons dans ce cas décidé de conserver ce score égal à 0 dans le calcul de la moyenne, et donc de pénaliser la moyenne car la lésion a été manquée, même dans le cas où elle était de faible volume.

Finalement, pour chacun des types de lésions et pour le tissu lésionnel global, nous avons comparé les scores de Dice moyen obtenus sur le jeu de données de test par les 3 différentes segmentations automatiques. Un test statistique a ensuite été appliqué pour déterminer la significativité de la différence entre les scores de Dice. Pour certains types de lésions, le faible nombre de scores de Dice calculés ne permet pas d'établir la normalité de l'échantillon. Nous avons donc décidé d'appliquer le test bilatéral des rangs signés de Wilcoxon [116], un test non-paramétrique pour échantillons appariés qui permet de tester la différence des médianes des échantillons. La significativité des résultats a été classée en 3 types en fonction de la p-valeur résultante du test statistique :

- La p-valeur est inférieure à 0.05 : significativité faible
- La p-valeur est inférieure à 0.01 : significativité modérée
- La p-valeur est inférieure à 0.001 : significativité forte

4.2.3 Résultats

Les résultats chiffrés sont disponibles sur le Tableau 15 et sont représentés sous forme d'histogrammes sur la Figure 35. Les segmentations obtenues par les 3 modèles sur les images représentatives d'un patient du jeu de données de test peuvent être visuellement comparées sur la Figure 36.

	Lésion globale	HIP	HEA	O	HIV
BLAST-CT	0,34 (0,05)	0,24 (0,04)	0,33 (0,06)	0,11 (0,04)	0,10 (0,04)
Random4C	0,53 (0,07)	0,35 (0,05)	0,43 (0,08)	0,47 (0,09)	0,38 (0,10)
FT_4C	0,63 (0,07)	0,45 (0,06)	0,48 (0,08)	0,57 (0,08)	0,51 (0,10)

Tableau 15. Moyenne et erreur standard des scores de Dice calculés entre les segmentations automatiques à 4 classes et les segmentations manuelles sur le jeu de test de la cohorte Radiomic-TBI.

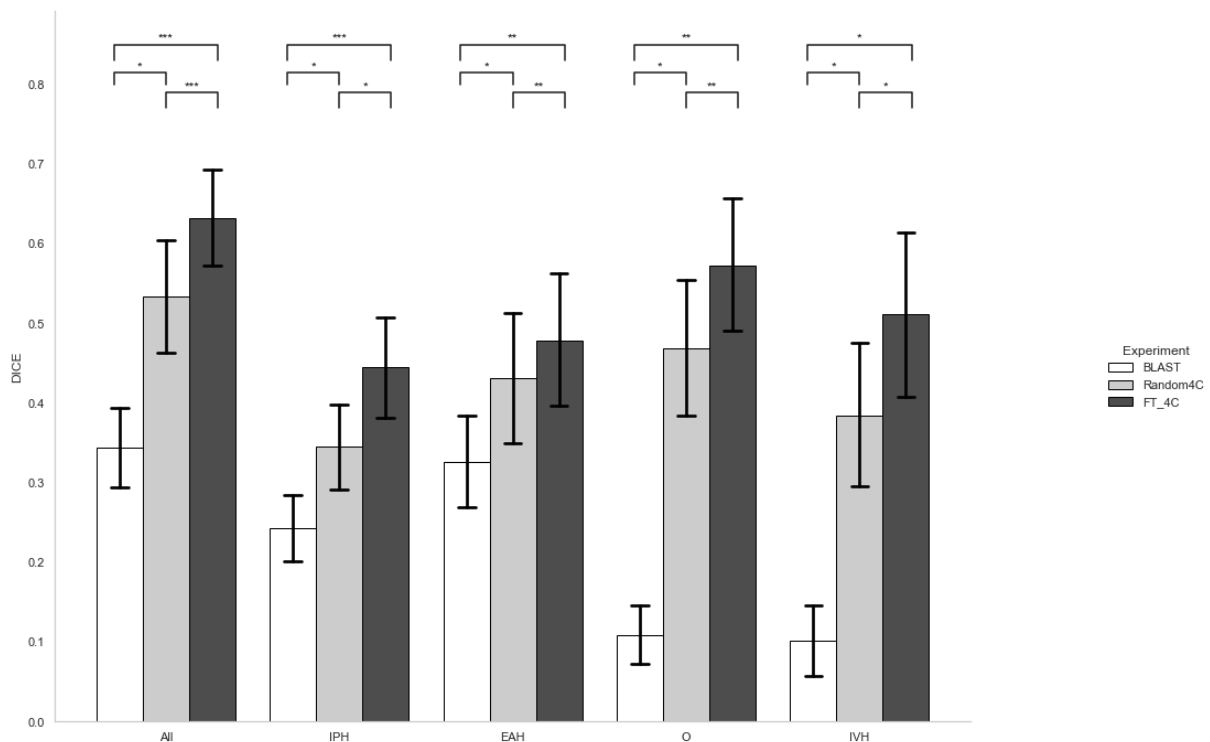


Figure 35. Scores de Dice moyens et erreurs standards par lésion (IPH/EAH/O/IVH) et pour le tissu lésionnel global (All) calculés sur le jeu de données de test pour 3 segmentations automatiques par rapport à la segmentation manuelle. Significativité entre les échantillons via le test bilatéral des rangs signés de Wilcoxon. * : faiblement significatif, ** : modérément significatif, *** : fortement significatif.

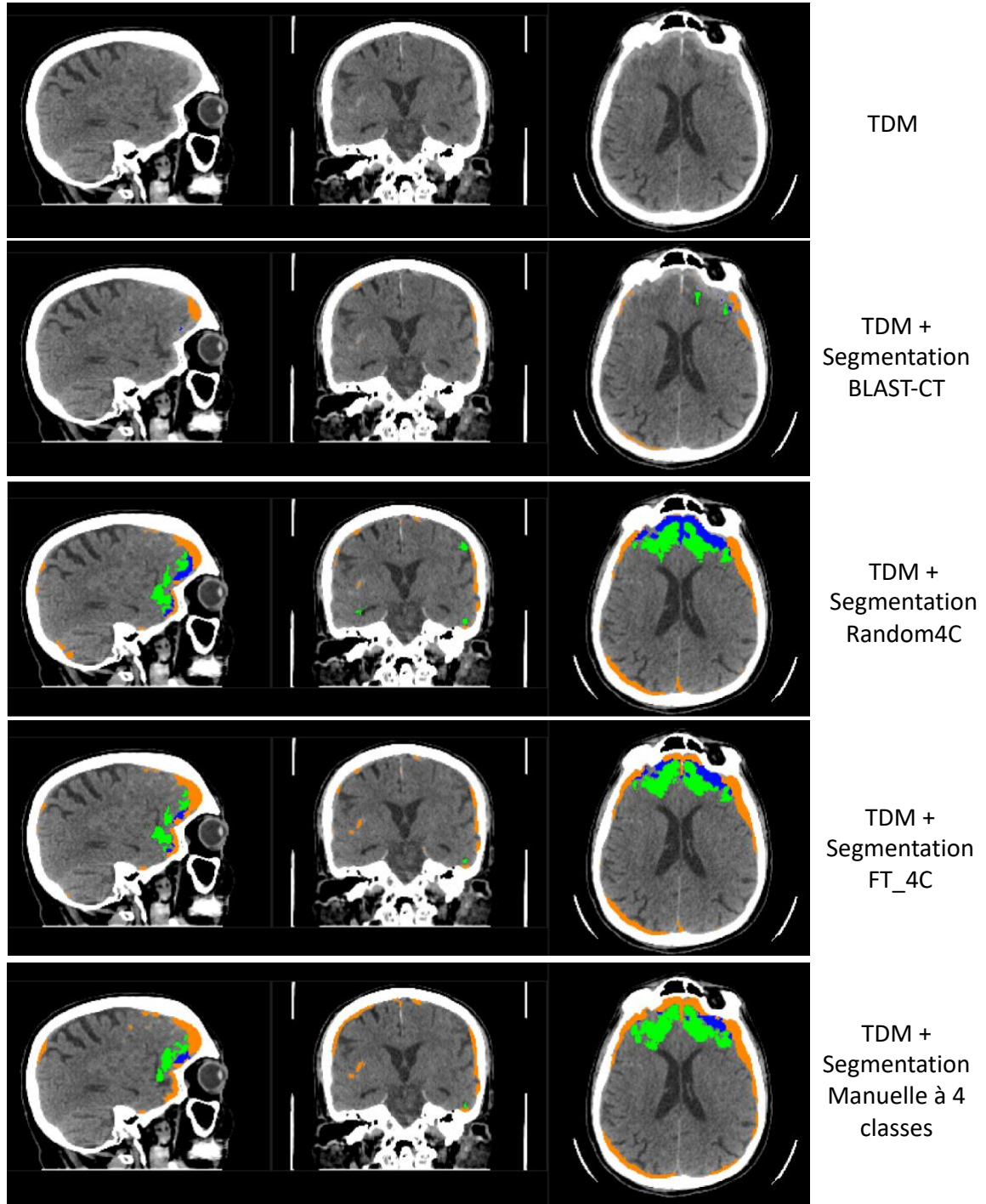


Figure 36. Comparaison sur une image du jeu de test des segmentations automatiques par rapport à la segmentation manuelle. Légende : Bleu – HIP, Orange – HEA, Vert – Œdème, Rouge – HIV.

Contrairement à notre hypothèse initiale, les résultats montrent une supériorité significative des 2 modèles entraînés sur nos données (Random4C et FT_4C) par rapport au modèle initial BLAST-CT, et ce, pour le tissu lésionnel global comme pour chacune des lésions. Comme prévu, le modèle FT_4C conduit à des scores de Dice significativement meilleurs que le modèle Random4C, pour toutes les lésions et pour le tissu lésionnel global.

Visuellement, sur le patient Figure 36, on voit que la sous-estimation récurrente des lésions du modèle BLAST-CT a disparue sur les modèles Random4C et FT_4C, ce qui est probablement la source de l'amélioration du score de Dice moyen. On peut également remarquer que le modèle FT_4C semble segmenter plus ou moins le même volume que le modèle Random4C mais mieux identifier les différents types de lésion. Ainsi, on remarque une plus faible erreur d'identification de l'HEA (vert) en HIP (orange) avec FT_4C qu'avec Random4C.

En ce qui concerne la combinaison des 12 CNN, après avoir appliqué les 4 méthodes pour obtenir 4 différentes segmentations globales pour le modèle FT_4C, nous avons obtenu les résultats organisés sous forme d'histogrammes sur la Figure 37.

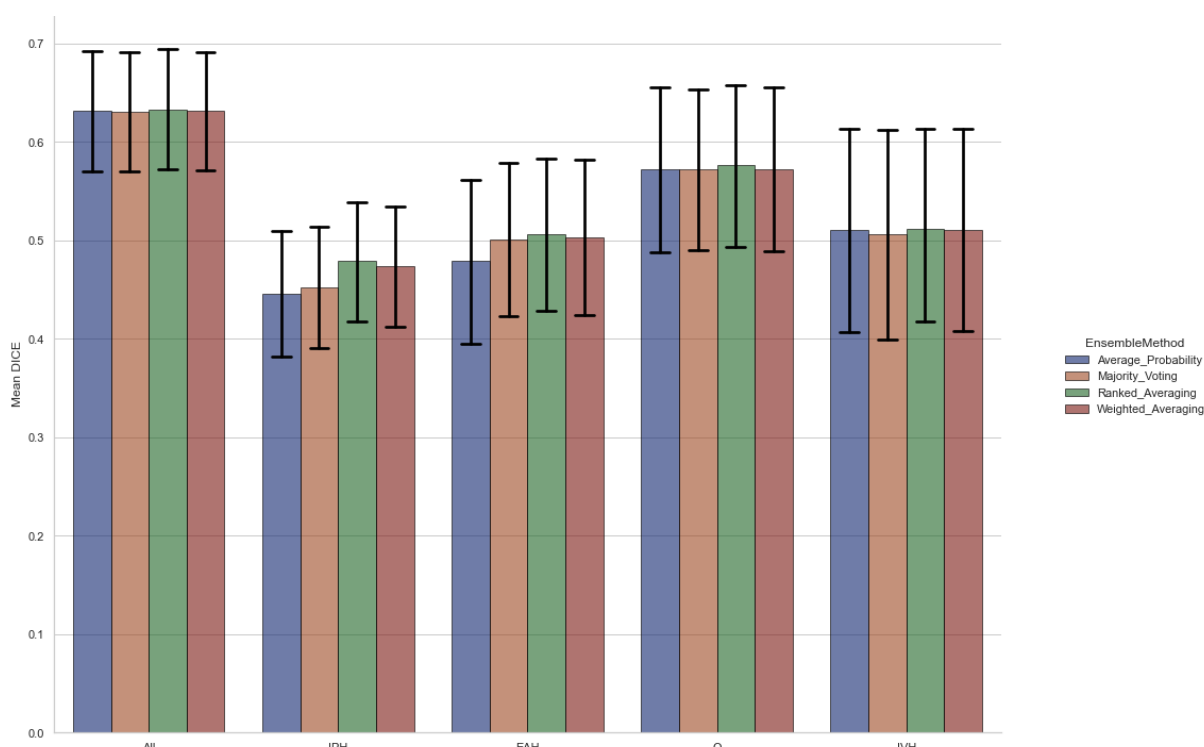


Figure 37. Comparaison des scores de Dice entre la segmentation automatique du modèle FT_4C et la segmentation manuelle à 4 classes pour les 4 différentes manières de combiner les 12 CNN du modèle. Légende : Bleu – méthode « Average_Probability », Orange – méthode « Majority_Voting », Vert – méthode « Ranked_Averaging », Rouge – méthode « Weighted_Averaging ». Les histogrammes représentent la moyenne des scores de Dice sur les images du jeu de test, et les barres représentent l'erreur standard.

Les 4 méthodes fournissent globalement des résultats similaires. La méthode « Ranked_Averaging » affiche de légèrement meilleurs résultats à la fois sur le tissu lésionnel global et sur les lésions prises individuellement. Cette méthode pondère l'influence de chaque CNN dans la prédiction globale en fonction de sa performance durant l'entraînement sur le jeu de validation. Si cette information est manquante, comme c'est par exemple le cas pour les 12 CNN de BLAST-CT, cette méthode n'est pas utilisable.

Vu les faibles différences entre les méthodes, nous avons décidé de conserver les modèles combinés avec la méthode classique « Average_Probability ».

4.3 Développement d'une segmentation automatique à 7 classes par « Transfer Learning »

4.3.1 Objectif

Au vu des bons résultats donnés par le réentraînement à 4 classes, nous avons naturellement souhaité aller plus loin et exploiter les segmentations manuelles à 7 classes afin de construire un modèle de segmentation automatique à 7 classes. En effet, nous pensons qu'il est pertinent de distinguer les 3 hémorragies confondues dans l'hémorragie extra axiale. Puisque la prise en charge d'un patient diffère s'il souffre d'une hémorragie sous arachnoïdienne, sous durale, ou extra durale. De plus, nous avons choisi de distinguer les pétéchies des hémorragies intra parenchymateuses car elles sont généralement associées aux lésions axonales diffuses, qui ont un impact bien différent des hémorragies intra parenchymateuses.

Dans cette expérience, nous avons adapté l'architecture DeepMedic afin de segmenter automatiquement les 7 types de lésions, puis entraîné cette architecture sur nos données, à partir de poids aléatoirement initialisés, puis à partir des poids de BLAST-CT et enfin à partir des poids du modèle précédemment fine-tuné lors de la précédente expérience (FT_4C). Les 3 modèles ainsi obtenus ont été évalués et comparés sur le jeu de test. Au vu des résultats du réentraînement à 4 classes, notre hypothèse est que le modèle entraîné à partir de poids aléatoirement initialisés fournira une segmentation moins proche de la segmentation manuelle que les modèles entraînés avec apprentissage par transfert.

4.3.2 Matériel et méthodes

4.3.2.1 Données utilisées

Pour cette expérience, nous avons utilisé les mêmes données que pour le réentraînement à 4 classes, à l'exception des segmentations manuelles, pour lesquelles nous avons bien naturellement utilisé les segmentations manuelles à 7 classes. La répartition des données en jeux d'entraînement, de validation, et de test reste la même.

4.3.2.2 Transfert des poids du réseau : Transfer Learning

Tel qu'illustré dans le chapitre 1, le nombre de classes segmentées par l'architecture DeepMedic est uniquement dépendante de la taille de la dernière couche, la couche de classification. Ainsi, il est facile de modifier cette dernière couche afin de segmenter 3 classes de plus. C'est d'ailleurs un paramètre modifiable dans le fichier json qui permet de configurer le réseau de BLAST-CT. Nous avons donc pu obtenir une architecture DeepMedic prenant une image 3D en entrée et segmentant 7 différentes lésions.

Évidemment, le modèle BLAST-CT étant entraîné pour segmenter 4 lésions, les tailles des couches ne sont pas compatibles avec celles de la nouvelle architecture à 7 classes.

Ainsi, lors de notre approche par transfer learning, nous avons transféré sur l'architecture à 7 classes les poids de l'architecture à 4 uniquement sur les couches dont les tailles étaient compatibles, c'est-à-dire l'ensemble des couches à l'exception de la dernière. Sur la dernière couche, les poids ont donc toujours été aléatoirement initialisés. S'il est parfois courant, dans le domaine du transfer learning, de geler une partie des poids, par exemples ceux des premières ou des dernières couches, afin que l'entraînement ne les modifie pas, nous avons préféré réentraîner l'ensemble des poids du réseau. En effet, de précédents travaux, non publiés, de l'équipe semblaient montrer de meilleurs résultats avec cette approche.

4.3.2.3 Entraînement

Nous avons utilisé la même méthodologie que pour l'entraînement à 4 classes à l'exception de la fonction de coût, que nous avons remplacé par la fonction DICECrossEntropy, qui revient simplement à ajouter « 1 moins le score de Dice » à la fonction de coût CrossEntropy classique. En effet, puisque l'évaluation de nos modèles se fait sur le score de Dice, autant l'inclure dans la fonction de coût afin de maximiser ce score.

Une fois encore, les entraînements ont été exécutés sur la grille de calcul GRICAD. Les entraînements des 3 modèles étudiés (et donc des 36 CNN) de cette expérience sont présentés sur le Tableau 16.

Nom du modèle	Initialisation des poids du réseau	Entraîné sur nos données ?
4/ Random7C	Aléatoire	OUI
5/ TL_7C	Poids de BLAST-CT	OUI
6/ TL_RE_7C	Poids du modèle FT_4C	OUI

Tableau 16. Initialisation et entraînement des modèles étudiés à 7 classes.

4.3.3 Résultats

Les résultats chiffrés sont disponibles sur le Tableau 17 et sont représentés sous forme d'histogrammes sur la Figure 38. Les segmentations obtenues par les 3 modèles sur un patient représentatif du jeu de données de test sont présentées sur la Figure 39.

	Lésion globale	HIP	HSD	HED	HIV	HSA	P	O
Random7C	0,55 (0,07)	0,41 (0,06)	0,41 (0,08)	0,68 (0,12)	0,35 (0,09)	0,14 (0,05)	0,14 (0,05)	0,46 (0,08)
TL_7C	0,63 (0,06)	0,50 (0,08)	0,48 (0,08)	0,66 (0,16)	0,42 (0,11)	0,19 (0,10)	0,24 (0,07)	0,54 (0,09)
TL_RE_7C	0,64 (0,06)	0,48 (0,08)	0,48 (0,07)	0,61 (0,17)	0,48 (0,10)	0,20 (0,07)	0,26 (0,06)	0,57 (0,08)

Tableau 17. Moyenne et erreur standard des scores de Dice calculés entre les segmentations automatiques à 7 classes et les segmentations manuelles sur le jeu de test de la cohorte Radiomic-TBI.

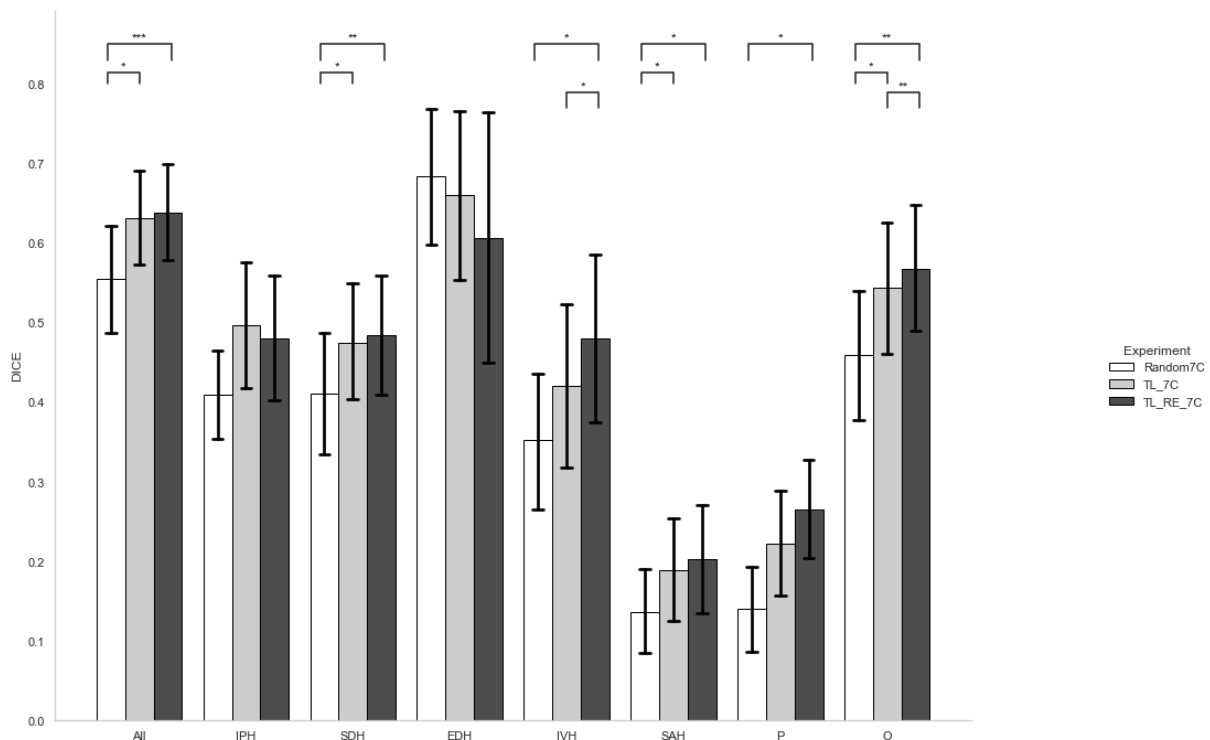


Figure 38. Moyenne et erreur standard des scores de Dice calculés entre les segmentations automatiques à 7 classes et les segmentations manuelles sur le jeu de test de la cohorte Radiomic-TBI.

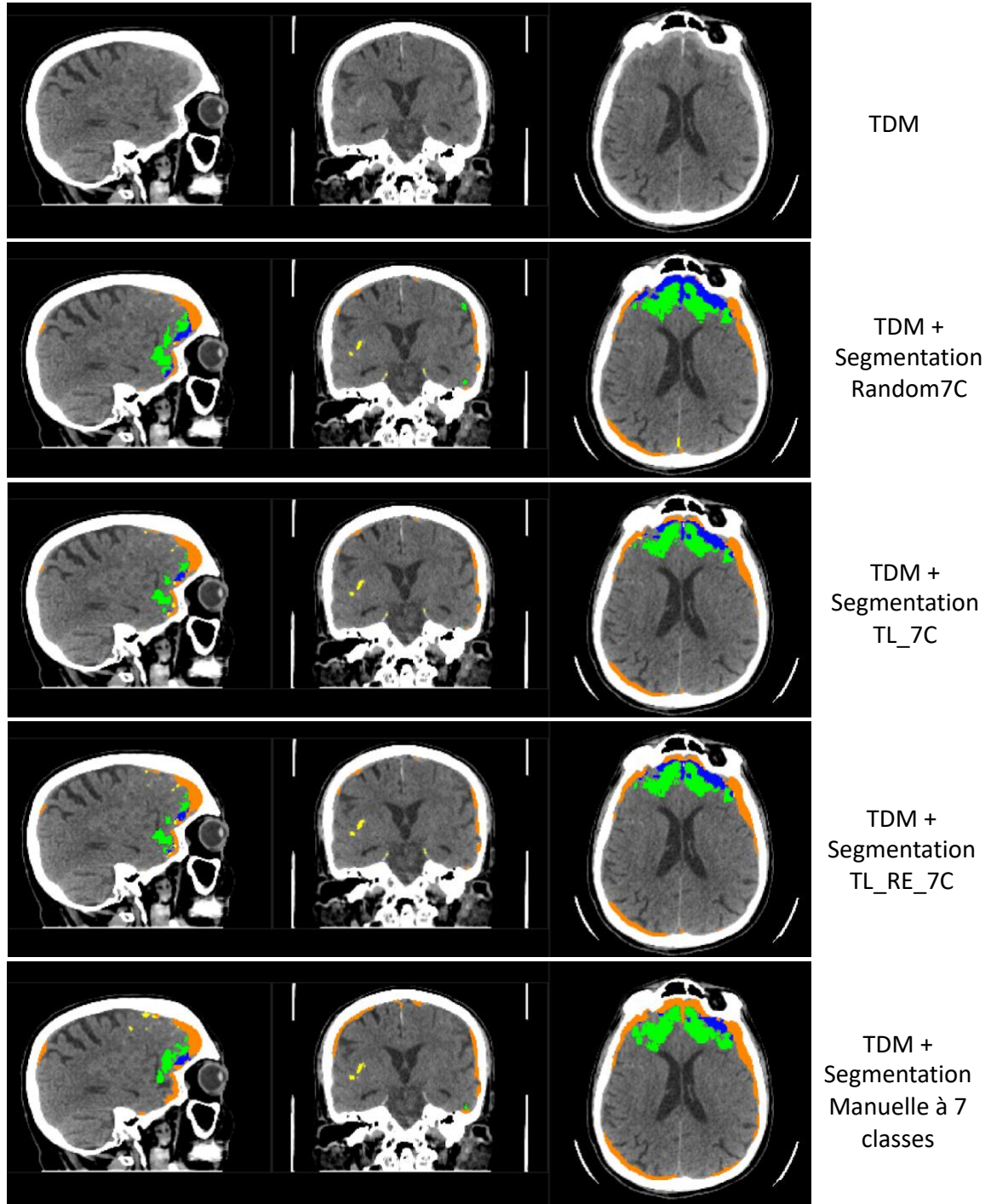


Figure 39. Illustration sur un patient du jeu de test de la cohorte Radiomic-TBI des 3 modèles de segmentation automatiques à 7 classes et segmentation manuelle.

Les résultats montrent une supériorité significative des 2 modèles entraînés avec Transfer Learning (TL_7C et TL_RE_7C) par rapport au modèle initialisé aléatoirement Random7C et ce, pour le tissu lésionnel global comme pour chacune des lésions. Les résultats des deux modèles entraînés avec Transfer Learning fournissent des résultats similaires. Visuellement, sur la Figure 39, on voit que les segmentations des 2 modèles entraînés par Transfer Learning sont très proches. On peut également remarquer que le

modèle Random_7C semble segmenter plus ou moins le même volume que les deux autres modèles mais classe moins bien les voxels segmentés dans les 7 différentes catégories. En ce qui concerne la combinaison des 12 CNN, après avoir appliqué les 4 méthodes pour obtenir 4 différentes segmentations globales pour le modèle TL_RE_7C, nous avons obtenu les résultats organisés sous forme d'histogrammes sur la Figure 40.

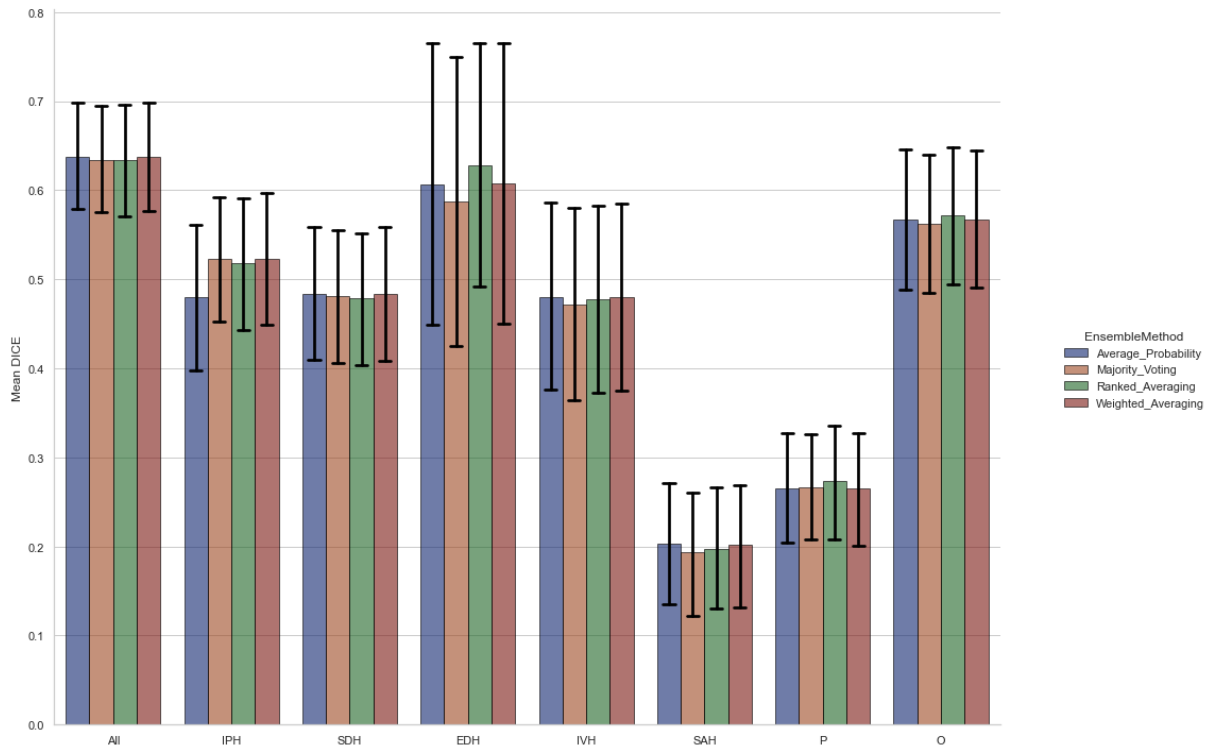


Figure 40. Comparaison des scores de Dice entre la segmentation automatique du modèle TL_RE_7C et la segmentation manuelle à 7 classes pour les 4 différentes manières de combiner les 12 CNN du modèle. Légende : Bleu – méthode « Average_Probability », Orange – méthode « Majority_Voting », Vert – méthode « Ranked_Averaging », Rouge – méthode « Weighted_Averaging ». Les histogrammes représentent la moyenne des scores de Dice sur les images du jeu de test, et les barres représentent l'erreur standard.

Les 4 méthodes fournissent une fois encore des résultats similaires. Nous avons donc décidé de conserver les modèles combinés avec la méthode classique « Average_Probability », comme nous l'avons fait pour le réentraînement à 4 classes.

4.4 Interprétation

Les résultats de ces deux expériences montrent une supériorité des approches par Transfer Learning pour adapter à une tâche précise un modèle entraîné soit sur une autre tâche soit sur des données d'une autre nature.

Ainsi, lorsqu'on souhaite utiliser un réseau publié, nos résultats suggèrent qu'il est plus pertinent de réentraîner ce réseau sur un échantillon de nos données plutôt que de l'appliquer tel quel.

De la même manière, si le modèle initialement publié est facilement accessible, il semble plus pertinent de commencer le réentraînement en partant des poids de ce réseau plutôt qu'avec des poids aléatoires.

Concernant les méthodes de combinaison des CNN d'un modèle, aucune des méthodes testées n'a fourni de bien meilleurs résultats que la méthode classique qui consiste à moyenner, pour chaque voxel, la probabilité d'appartenir à telle ou telle classe des 12 CNN. On peut néanmoins remarquer qu'une étude sortie en pré-print en septembre 2022 a semble-t-il trouvé un moyen de combiner en un seul CNN les poids de 2 CNN entraînés sur des données différentes, afin d'obtenir un unique CNN affichant de meilleures performances à celles des 2 CNN pris séparément [117]. On pourrait imaginer combiner les 12 CNN résultants de nos entraînements d'une manière similaire.

Il est à noter que ces résultats ont été obtenus à partir d'un petit jeu de données, seulement 84 images TDM provenant de 29 patients différents. On peut ainsi imaginer que le réentraînement d'un réseau devienne une étape du prétraitement inhérent à toute étude impliquant des données d'imagerie médicale.

4.5 Limitations

Comme noté au paragraphe précédent, le jeu de données utilisé pour ces expériences est faible en nombre. Les résultats seraient plus précis et plus fiables avec une évaluation sur une grande base de données, mais cela implique la segmentation manuelle d'un grand nombre d'images TDM, ce qui est une opération fastidieuse et rébarbative, qui doit être effectuée par des médecins spécialistes, qui ne disposent naturellement pas de beaucoup de temps à consacrer à ces travaux.

L'évaluation de ces expériences est basée sur le score de Dice, qui n'est pas sans limites, comme exposé plus haut. Il serait bon de vérifier la cohérence de notre interprétation à la lumière d'évaluations basées sur d'autres métriques, comme par le taux de lésions vraies positives et fausses positives [118], qui s'intéresse non pas au volume des lésions mais au fait qu'elle ait été détectée ou non. Nous avons néanmoins calculé la distance moyenne de Hausdorff équilibrée sur le jeu de test de la cohorte Radiomic-TBI pour les différentes segmentations automatiques et ces résultats sont disponibles en annexe.

4.6 Perspectives

En plus de l'évaluation avec de nouvelles métriques, nous pourrions explorer les raisons de la mauvaise performance de BLAST-CT sur nos données. Est-ce dû à une différence de population ? Est-ce dû à une variabilité introduite par les différents protocoles d'acquisition sur les images TDM ? Est-ce dû à la manière de segmenter nos images, et donc d'estimer les lésions ? Nous pourrions monter une étude pour répondre à cette question, avec par exemple des données issues de notre centre et d'autres centres, et des segmentations manuelles dessinées par nos spécialistes mais aussi par d'autres médecins.

Nous pourrions également entraîner un réseau à segmenter plus de lésions, comme par exemple l'ischémie, et explorer les mécanismes d'oubli catastrophique : est-ce que le transfer learning implique une perte du savoir déjà acquis ? Le réentraînement sur des données de soin courant permettrait également d'améliorer la robustesse du réseau aux artefacts. Ce point va être abordé dans le cadre de la thèse de Vaëa TESAN, doctorante de première année dans l'équipe.

Finalement, pour améliorer la segmentation des différentes lésions, nous pourrions séparer les tâches et entraîner un réseau par type de lésion à segmenter, et ensuite combiner les modèles pour obtenir une segmentation multiclasse. Nous pourrions également tester plusieurs architectures de réseaux convolutifs, afin de trouver pour chaque lésion l'architecture qui permet d'obtenir les meilleurs résultats.

Chapitre 5 : Prédiction de l'intensité thérapeutique à partir de l'image TDM initiale

5.1 Objectifs

La majorité des articles prédisant le futur des patients après un TC se concentrent sur des critères à moyen terme, comme l'état neurologique mesuré par le GOS ou bien la mortalité à 3 ou 6 mois. Ces critères, calculés plusieurs mois après l'accident initial, dépendent naturellement de variables non prises en compte parmi lesquelles la rééducation et l'accompagnement familial. De plus, du point de vue des réanimateurs, une prédiction de l'état d'un patient plusieurs mois après sa sortie de l'hôpital est inutile pour la prise en charge du patient. En revanche, les cliniciens ont un réel besoin d'aide pour la prise en charge à court terme d'un patient (< 10 jours). En effet, il existe des patients dont l'évolution est extrêmement difficile à prévoir, et ce même en analysant les constantes des patients et leurs images TDM. Comme illustré Figure 41 et Figure 42, certains patients au profil lésionnel similaire ont des évolutions très différentes à court terme et ne nécessitent pas la même prise en charge. A l'heure actuelle, ces différents profils d'évolutions ne sont pas prédictibles par l'analyse visuelle et qualitative des images TDM réalisées en routine clinique.

Nous faisons l'hypothèse que les informations contenues dans les images TDM permettent d'estimer de manière précoce le niveau d'intensité thérapeutique nécessitée par un patient. Dans le but de vérifier cette hypothèse et d'apporter de nouvelles options aux cliniciens, l'objectif de cette étude vise à développer un biomarqueur d'imagerie TDM, couplé ou non à des données cliniques, pouvant prédire l'intensité thérapeutique nécessaire à la prise en charge des patients durant les premiers jours en réanimation. Nous nous sommes concentrés sur la prédiction de ce critère à partir des données acquises dès l'admission au CHUGA. Cet objectif nous semble accessible car, contrairement aux critères habituellement prédits dans la littérature scientifique, l'intensité thérapeutique ne dépend pas de facteurs extra médicaux.

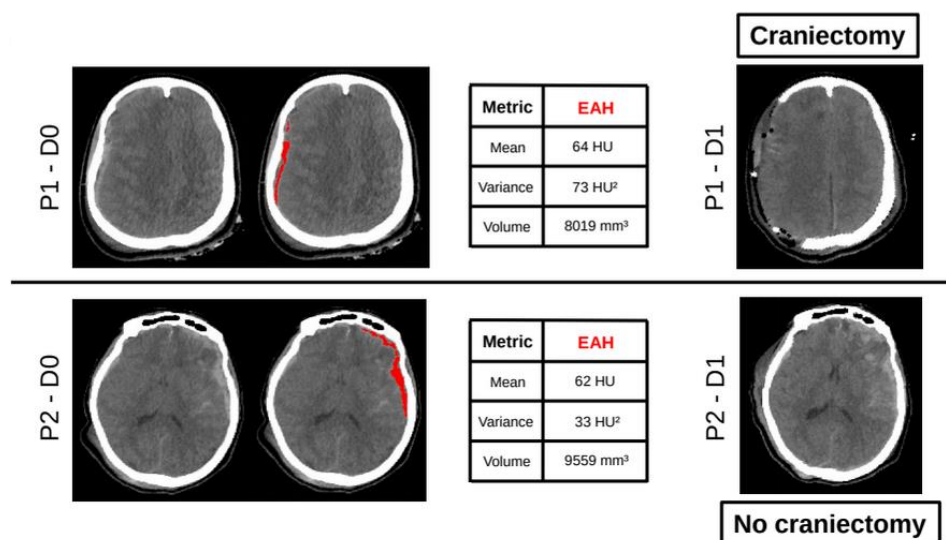


Figure 41. Illustration de 2 patients de la cohorte Radiomic-TBI de profil lésionnel similaire à l'arrivée à l'hôpital et qui n'ont pas nécessité la même prise en charge car le lendemain, l'un a été craniectomisé (P11 en haut, TILsum max = 20) et pas l'autre (P12 en bas, TILsum max = 7). Source : [72]

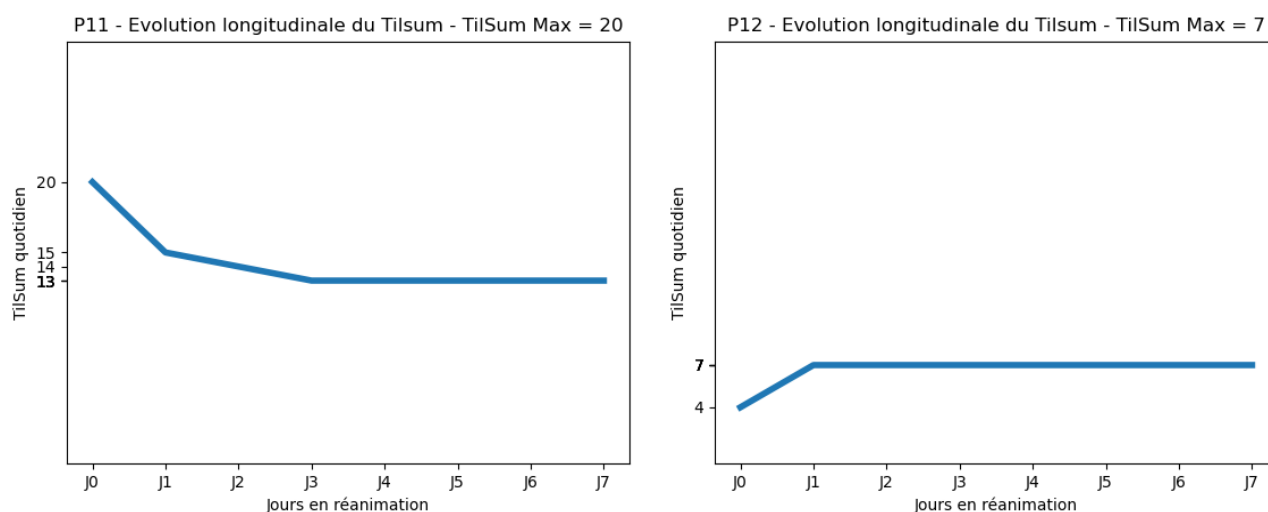


Figure 42.Évolution du TILsum quotidien de J0 à J7 de deux patients de la cohorte Radiomic-TBI (P11 à gauche et P12 à droite). Ces deux patients sont les mêmes qu'à la figure 41.

5.2 Matériel et méthodes

5.2.1 Critère à prédire : Une journée d'intensité thérapeutique extrême au cours des 7 premiers jours en réanimation

Comme présenté au Chapitre 1, le TILsum est généralement calculé quotidiennement, en recensant toutes les interventions ou traitements subis par un patient au cours des 24h de la journée. Ce score est compris entre 0 et 38. Selon [119]

[120], on peut classer la prise en charge d'un patient grâce à ce score en basique, légère, modérée ou extrême, selon le Tableau 18 ci-dessous :

Critère à prédire	TILsum	Prise en charge
TILsum_bas	0-3	Basique, naturelle en réanimation
	4-6	Légère
	8-10	Modérée
TILsum_haut	11-38	Extrême

Tableau 18. Classification de la prise en charge à partir du TILsum.

Dans notre étude nous souhaitons prédire, à partir de l'image TDM acquise à l'admission à l'hôpital, si un patient va subir au moins une journée de prise en charge extrême pendant les 8 premiers jours en réanimation. Si c'est le cas, et pour la suite de ce travail, le patient sera classé comme TILsum_haut, sinon il sera TILsum_bas. Attention, si le TILsum est un critère calculé quotidiennement, en revanche, le TILsum_haut ou TILsum_bas est un critère binaire calculé sur les 8 premiers jours en réanimation, donc calculé par patient. Nous allons classer les patients en fonction de leur appartenance aux groupes TILsum_haut / TILsum_bas. Puisque ce critère est prospectif, cette classification est donc une prédiction.

5.2.2 Données utilisées

Pour cette étude, nous avons utilisé la cohorte Radiomic-TBI, plus précisément les données cliniques à l'admission ainsi que l'image TDM acquise à l'arrivée au CHU. La cohorte contient 29 patients donc nous avons utilisé 29 images TDM, que nous avons réparties en 2 jeux : un jeu de développement (23 patients), et un jeu de test (6 patients), selon la répartition suivante :

- Jeu de développement : P01, P05, P06, P12, P13, P14, P15, P17, P18, P19, P20, P22, P24, P25, P26, P27, P30, P03, P07, P09, P10, P23, P28
- Jeu de test : P02, P04, P08, P11, P16, P29

Notons que le jeu de développement est simplement la concaténation des jeux d'entraînement et de validation présentés au chapitre précédent dans l'étude sur la segmentation, et que le jeu de test est strictement le même qu'au chapitre précédent.

Concernant les données cliniques, nous avons collecté et utilisé seulement toutes les données qui étaient disponibles pour tous les patients afin de ne pas exclure de patient. Ces données cliniques, relevées à l'admission au CHUGA, sont :

- Age
- Sexe (seulement utilisé pour la démographie)

- Poids (seulement utilisé pour la démographie)
- Taille (seulement utilisé pour la démographie)
- Score de Glasgow (GCS)
- Pression artérielle moyenne (PAM)
- Taux d'hémoglobine (Hb)
- Présence ou non d'antiagrégants

De plus, nous avons calculé pour les 29 images TDM le score de Marshall et de Rotterdam associé (cf. chapitre 1). Pour rappel, ces scores ont pour objectif de quantifier visuellement les anomalies cérébrales observées sur les images TDM, comme par exemple une compression des citernes de la base ou une déviation de la ligne médiane de plus de 5 mm.

La caractérisation des deux sous-groupes, TILsum_haut et TILsum_bas, de la cohorte Radiomic-TBI est disponible sur le Tableau 19.

	TILsum_bas	TILsum_haut
Nombre de patients	20	9
Age (années)	48,3 +-23,2 [18-79]	42,1 +-18,6 [22-73]
Sexe	13M / 7F	8M / 1F
Poids (kg)	70,8 +-13,6 [51-100]	76,7 +-6,1 [68-86]
Taille (cm)	175,3 +-9,8 [160-197]	175,5 +-5,8 [169-185]
GCS	9,3 +-4,3 [3-15]	6,9 +-4,0 [3-14]
PAM (mmHg)	87,8 +-17,9 [41-115]	92,7 +-14,6 [67-106]
Taux d'Hb (g/l)	125 +-27,7 [44-155]	129 +-20,6 [95-164]
Présence d'antiagrégants	4/20	0/9
Score de Marshall	2,6 +-1,3 [2-6]	4,9 +-1,1 [3-6]
Score de Rotterdam	3,0 +-0,8 [2-5]	4,4 +-1,2 [2-6]
TILsum	6,3 +-2,6 [1-10]	18,9 +-4,5 [11-28]

Tableau 19. Caractérisation des deux sous-groupes TILsum_Haut et TILsum_bas de la cohorte Radiomic-TBI (Moyenne +- Écart Type [Min-Max])

5.2.3 Quantification du profil lésionnel observé en TDM à l'admission

Nous avons mis en place un pipeline de traitement qui permet d'extraire le volume, le type et la localisation spatiale de différentes lésions cérébrales des patients. Ce pipeline de quantification d'une image TDM est détaillé ci-dessous et illustré Figure 43.

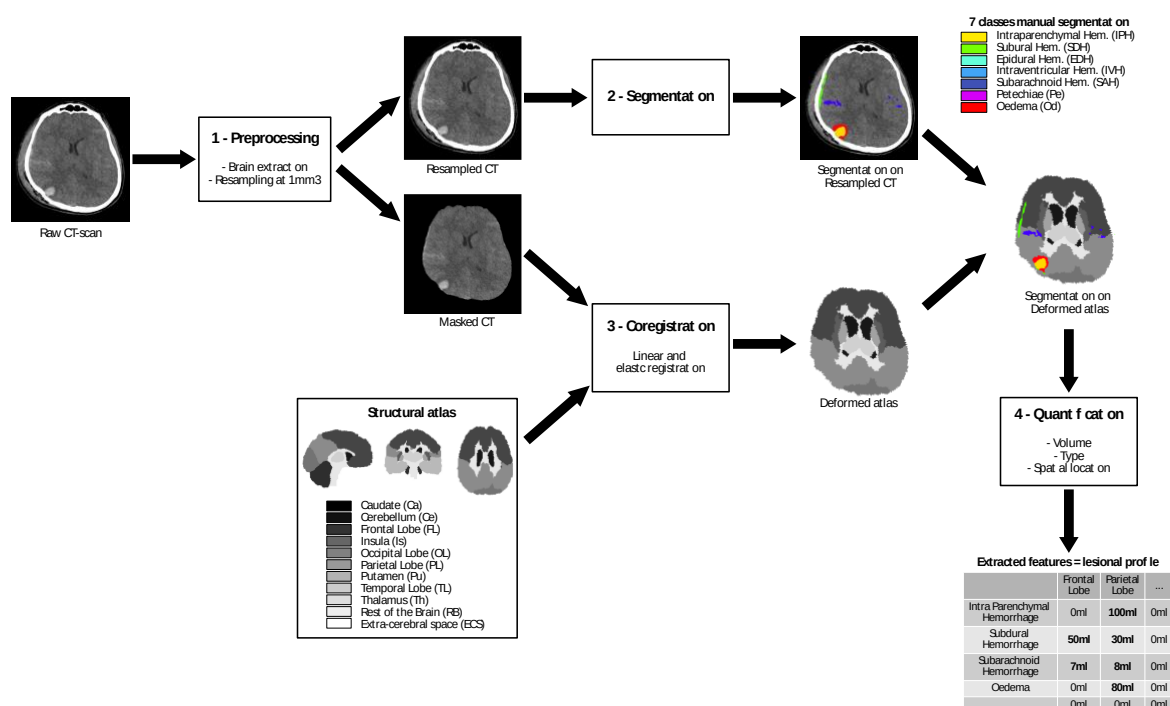


Figure 43. Vue d'ensemble du pipeline de quantification automatique du profil lésionnel d'un patient à partir d'une image TDM.

5.2.3.1 Pré-traitement

Comme pour l'étude sur la segmentation, les images TDM ont été extraites du PACS du CHUGA au format DICOM et converties au format NIfTI grâce au convertisseur « MRI file Manager » (<https://irmage.univ-grenoble-alpes.fr/services/traitement-dimages/mri-file-manager>) qui est directement intégré dans le logiciel MP3 [111]. Elles ont ensuite été rééchantillonnées à 1mm³ et le cerveau a été extrait des images grâce à un algorithme codé en langage MATLAB inspiré de l'outil de segmentation proposé par Najm et al. [121] auquel on a ajouté une sélection de la ROI de plus grande taille et remplissage des trous.

5.2.3.2 Volume et type de lésion : Segmentation

Afin d'obtenir le volume et le type de chaque lésion, la segmentation et identification des lésions semble être l'option la plus pertinente. Nous avons donc segmenté les lésions de manière automatique grâce à l'algorithme BLAST-CT, mais aussi de manière manuelle comme déjà décrit au chapitre précédent et rappelé sur le Tableau 20. Ainsi, nous disposons pour chacune des images TDM de 3 segmentations : La segmentation automatique à 4 classes de BLAST-CT (issue de la publication de Monteiro [102]), la segmentation manuelle à 4 classes et la segmentation manuelle à 7 classes.

Segmentation à 4 classes	Segmentation à 7 classes
Hémorragie intra parenchymateuse (HIP)	Hémorragie intra parenchymateuse (HIP)
	Pétéchies (P)
Œdème (O)	Œdème (O)
Hémorragie intra ventriculaire (HIV)	Hémorragie intra ventriculaire (HIV)
Hémorragie extra axiale (HEA)	Hémorragie sous durale (HSD)
	Hémorragie extra durale (HED)
	Hémorragie sous arachnoïdienne (HSA)

Tableau 20. Résumé et lien entre les lésions de la segmentation à 4 classes et à 7 classes.

5.2.3.3 Localisation spatiale : Recalage d'un atlas structurel

Afin d'obtenir la localisation spatiale des lésions, nous avons recalé un atlas structurel sur les images, ainsi chaque voxel du cerveau se voit attribuer une zone. Premièrement nous avons téléchargé 2 fichiers : un template TDM [122], c'est-à-dire un cerveau moyen, sur lequel nous avons extrait le cerveau, et un atlas structurel, qui délimite le cerveau en 9 zones [123] :

- Le noyau caudé (Ca)
- Cervelet (Ce)
- Lobe Frontal (LF)
- Insula (Is)
- Lobe Occipital (LO)
- Lobe Parietal (LP)
- Putamen (Pu)
- Lobe Temporal (LT)
- Thalamus (Th)

Pour couvrir le reste du cerveau (RC), nous avons manuellement rajouté une zone couvrant le centre du cerveau, majoritairement composée des ventricules. Finalement, en

comptant l'espace extra cérébral (EEC), on arrive à un atlas qui décompose l'espace en 11 zones, illustré Figure 44.

Cet atlas a ensuite été manuellement recalé sur le template.

Puis, ce template a été recalé sur chacune des 29 images TDM, de manière linéaire grâce à l'algorithme FLIRT du logiciel FSL [123], puis de manière élastique, avec ANTS [124]. Les déplacements calculés ont finalement pu être appliqués à l'atlas, afin d'obtenir cet atlas recalé sur chacune des 29 images TDM. Ainsi, chaque voxel des images TDM appartient à l'une des 11 zones de l'atlas.

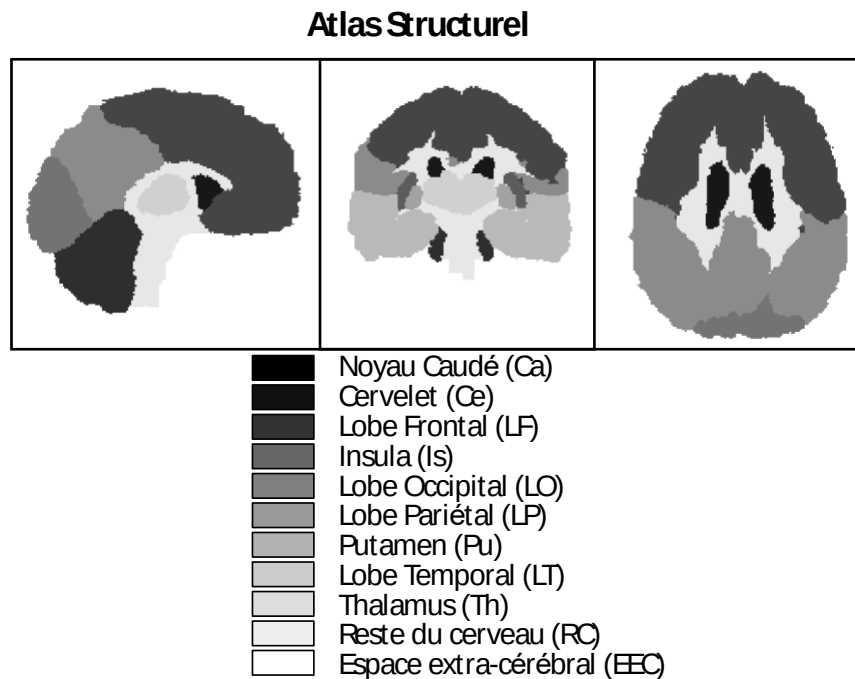


Figure 44. Atlas structurel [123] modifié afin de segmenter 10 zones du cerveau.

5.2.3.4 Quantification

Après la segmentation des lésions et le recalage de l'atlas structurel, à chaque voxel de l'image est attribué une zone de l'atlas et un type de lésion (ou une absence de lésion). On peut ainsi extraire facilement le volume de chaque type de lésion dans chacune des zones de l'atlas et réunir ces volumes dans un tableau qui représente le profil lésionnel d'un patient. Pour un atlas à 11 zones et une segmentation à 4 classes, on obtient 44 métriques différentes par patient (par exemple : 100ml d'hémorragie intraparachymateuse dans le lobe pariétal). Pour une segmentation à 7 classes on passe à 77 métriques.

5.2.4 Classification

A partir des métriques extraites, représentant le profil lésionnel des patients sous forme du volume, type et localisation des lésions, nous avons prédit notre critère d'intensité thérapeutique, c'est-à-dire l'appartenance d'un patient au groupe TILsum_haut ou TILsum_bas. Nous avons donc mené une étude de classification impliquant un grand nombre de métriques pour un petit nombre d'observations. Pour cela, nous avons réalisé différentes expériences de classifications employant différentes méthodes statistiques afin d'obtenir un modèle de classification aussi robuste et généralisable que possible. Ces expériences sont détaillées ci-dessous.

5.2.4.1 Expériences menées

Dans le but d'étudier la capacité de prédiction du critère en fonction de l'information contenue dans les métriques, nous avons mené plusieurs expériences de classification avec des jeux de métriques différents pour prédire notre critère : l'appartenance des patients au groupe TILsum_haut ou TILsum_bas. En effet, nous pouvons combiner les métriques extraites par notre pipeline de traitement afin de contrôler l'information utilisée. On peut, par exemple, sommer les volumes de toutes les zones de l'atlas pour aboutir au volume par type de lésion dans l'ensemble de cerveau. On aura ainsi effacé l'information de localisation pour ne garder que celle du volume par type des lésions.

Nous avons ainsi mené 2 expériences utilisant des métriques cliniques associées ou non à des scores d'imagerie (Tableau 21; Exp 1 et 2), puis 5 expériences principales basées sur les volumétries des lésions (Tableau 21; Exp 3 --> 7), auxquelles se rajoutent 5 expériences secondaires, qui ne sont aussi basées sur les volumétries des lésions auxquelles on a ajouté des métriques cliniques (Tableau 21; Exp 3_bis --> 7_bis). Ces 12 expériences sont détaillées sur le Tableau 21.

Expérience	Métriques utilisées	Nombre de métriques
Exp1	Age + GCS + PAM, + Hb + Antiaggregant	5
Exp2	Métriques de l'Exp1 + score de Marshall + score de Rotterdam	7
Exp3	Volume lésionnel global dans le cerveau (Ex : 321 mm3 de lésion)	1
Exp4	Volume de lésion par type (Ex : 321 mm3 d'HIP)	4 pour une segmentation à 4 classes et 7 pour une segmentation à 7 classes

Exp5	Volume de lésion par localisation (Ex : 321 mm3 de lésion dans le LF)	11
Exp6	Métriques de l'Exp4 + Métriques de l'Exp 5	15 pour une segmentation à 4 classes et 18 pour une segmentation à 7 classes
Exp7	Volume de lésion par type et localisation (Ex : 321 mm3 d'HIP dans le LF)	44 pour une segmentation à 4 classes et 77 pour une segmentation à 7 classes
Exp3_bis	Métriques de l'Exp3 + Métriques de l'Exp1	6
Exp4_bis	Métriques de l'Exp4 + Métriques de l'Exp1	9 pour une segmentation à 4 classes et 12 pour une segmentation à 7 classes
Exp5_bis	Métriques de l'Exp5 + Métriques de l'Exp1	16
Exp6_bis	Métriques de l'Exp6 + Métriques de l'Exp1	20 pour une segmentation à 4 classes et 23 pour une segmentation à 7 classes
Exp7_bis	Métriques de l'Exp7 + Métriques de l'Exp1	49 pour une segmentation à 4 classes et 82 pour une segmentation à 7 classes

Tableau 21. Résumé des 12 expériences menées : Nombre et nom des métriques utilisées

5.2.4.2 Pipeline de classification

Pour convertir les métriques d'entrée en une prédiction l'appartenance des patients au groupe TILsum_haut ou TILsum_bas nous avons mis en place un pipeline de classification, illustré Figure 45, constitué de 4 étapes :

1/ Une normalisation des métriques entre 0 et 1

Pour que les métriques occupent des plages de valeurs comparables, on les normalise de manière à ce que, pour chaque métrique, la valeur de l'échantillon de valeur maximum soit fixée à 1 et celle de l'échantillon de valeur minimum soit fixée à 0. On obtient les valeurs des échantillons après normalisation Y à partir des valeurs initiales X par la formule :

$$Y = \frac{X - \min(X)}{\max(X) - \min(X)}$$

2/ Un sous échantillonnage aléatoire

Puisque la répartition du nombre de patient appartenant aux groupes TILsum_haut ou TILsum_bas dans le jeu de développement n'est pas égale (TILsum_haut $n=7$ et TILsum_bas = 16) nous avons mis en place une stratégie pour pallier ce biais. Ainsi, pour avoir le même nombre d'échantillons dans les deux groupes à classer, à chaque entraînement, on a exclu de manière aléatoire un certain nombre d'échantillons de la classe majoritaire (TILsum_bas) de manière à équilibrer le nombre d'échantillon dans chaque classe.

3/ Une sélection de métriques

Afin de supprimer une information superflue et ainsi limiter le surapprentissage, on a sélectionné les métriques les plus liées au critère à prédire, selon une fonction qui calcule un score par rapport au critère. Nous avons utilisé la fonction `f_classif`, de la bibliothèque Python `scikit learn`, qui effectue une analyse de la variance (ANOVA) et qui revient pour chaque métrique à calculer la moyenne des valeurs de la métrique pour les deux groupes à prédire et à calculer la variance entre ces deux moyennes. Plus la variance est élevée, plus cette métrique discrimine les deux groupes. Ainsi, les 4 métriques qui discriminent le mieux les deux groupes à prédire sont conservées, et les autres exclues.

4/ L'entraînement d'un classifieur

Une fois que les métriques et les échantillons sont préparés, on peut entraîner un classifieur pour estimer une relation entre les valeurs d'entrée et le critère à prédire. Il existe plusieurs types de classifieurs, et l'un des plus connus et plus utilisés est la forêt aléatoire (random forest en anglais). C'est le classifieur Random Forest de la bibliothèque `scikit learn` que nous avons utilisé, avec les valeurs des hyperparamètres par défaut, à l'exception du nombre d'arbres, du nombre minimum de feuilles, et du nombre minimum d'échantillons pour séparer les données par un nœud. Nous avons fait varier ces 3 hyperparamètres. Ils peuvent prendre comme valeur un entier entre, respectivement, [5,9], [1,2], [2,4] comme illustré Figure 45. Ces plages de valeurs ont été fixées de manière arbitraire. Les différentes combinaisons de ces 3 hyperparamètres sont au nombre de $15 \times 2 \times 3 = 90$. Nous avons donc 90 jeux d'hyperparamètres à entraîner et comparer.

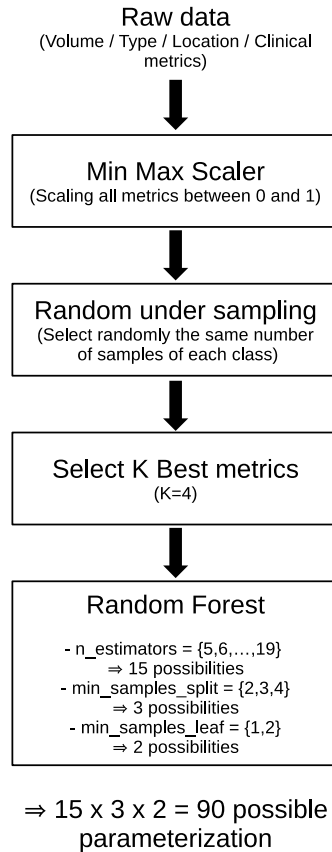


Figure 45. Illustration du pipeline de classification utilisé pour prédire notre critère. Figure issue de [106]

5.2.4.3 Procédure de validation croisée imbriquée

Puisque nous disposons d'un petit jeu de données (23 patients dans le jeu de développement), nous avons mis en place une procédure de validation croisée imbriquée [62] [63] pour entraîner, sélectionner et évaluer sur ces 23 patients un modèle pour chacune des 12 expériences menées (cf. Tableau 21). La boucle interne est une validation croisée à 5 ensembles stratifiée et répétée 10 fois (5-fold Repeated Stratified cross validation), et la boucle externe est une validation croisée à 6 ensembles stratifiée et répétée 10 fois (6-fold Repeated Stratified cross validation).

Le fait qu'une validation croisée soit stratifiée signifie que pour chaque répartition du jeu de données initial en jeu d'entraînement et de validation, le pourcentage d'échantillons des deux groupes est préservé par rapport à celui du jeu de données initial.

Le fait qu'elle soit répétée signifie que l'on répète un certain nombre de fois la validation croisée avec une répartition différente des échantillons.

Comme détaillé dans le premier chapitre et illustré sur la Figure 46 (partie de gauche : jeu de développement), la procédure de validation croisée imbriquée sélectionne les modèles affichant les meilleures performances sur les jeux de validation et sont

ensuite évalués sur les jeux d'évaluation, conduisant à 60 modèles. Puisque nous avons choisi d'utiliser l'AUC comme métrique de sélection et d'évaluation de nos modèles, aux 60 modèles s'associent 60 AUC, ce qui permet de calculer une AUC moyenne, qui représente la capacité moyenne du pipeline de classification à prédire le critère voulu à partir des métriques d'entrée, et un écart type. On peut aussi sauvegarder le meilleur modèle de classification, afin de l'utiliser plus tard pour prédire le futur d'autres patients. Cet entraînement a été effectué avec la bibliothèque Python PhotonAI [125], qui interface les principales fonctions de la bibliothèque scikit learn et permet de mener facilement et rapidement une classification au travers d'une validation croisée imbriquée.

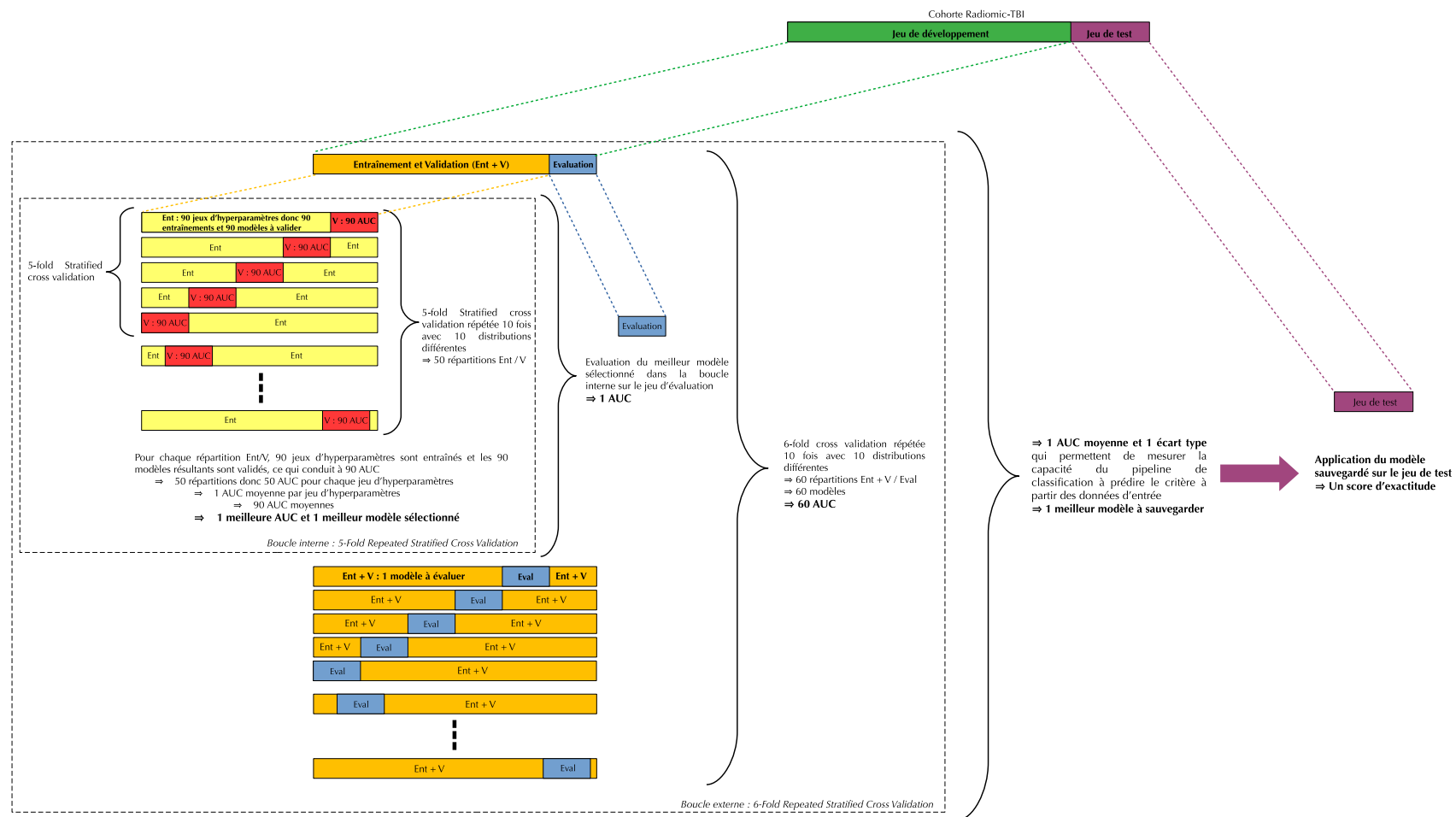


Figure 46. Illustration de la procédure de validation croisée imbriquée utilisée. 90 jeux d'hyperparamètres testés, avec comme boucle interne une validation croisée à 5 ensembles répétée 10 fois et comme boucle externe une validation croisée à 6 ensembles répétée 10 fois. Figure issue de l'article [106]

5.2.4.4 Interprétation des modèles sauvegardés

Une fois le meilleur modèle sauvegardé, on peut l'étudier pour repérer les métriques d'entrée qui contribuent le plus à la prédiction. On peut notamment calculer l'importance de chacune des métriques comme l'importance de gini, aussi appelée Diminution moyenne des impureté (Mean Decrease Impurity en anglais) [126] [127] afin d'obtenir le pourcentage d'importance des 4 métriques sélectionnées par le pipeline de classification, ce qui permet de repérer les principales métriques utilisées par le modèle.

5.2.4.5 Évaluation sur le jeu de test

A la suite de la validation croisée imbriquée, nous avons sauvegardé le modèle de l'expérience qui a fourni le meilleur résultat avec une segmentation à 4 classes et celui qui a fourni le meilleur résultat avec une segmentation à 7 classes.

Dans le but de les évaluer sur le jeu de test préalablement mis de côté (cf. Figure 46 partie droite), nous avons extrait les métriques de volume type et localisation des 6 patients du jeu de test à partir de 6 différentes segmentations :

- BLAST-CT
- TL_BLAST4C, le meilleur modèle de segmentation à 4 classes, obtenu après fine-tuning, développé au chapitre 3
- Segmentation manuelle à 4 classes
- Random_7C, modèle obtenu après entraînement aléatoirement initialisé à 7 classes, développé au chapitre 3
- TL_RE_7C, le meilleur modèle de segmentation à 7 classes, obtenu après transfer learning, développé au chapitre 3
- Segmentation manuelle à 7 classes

Il doit être noté que les modèles de segmentation que nous avons développés au chapitre 3 ont été entraînés et validés sur le jeu de développement. Par conséquent ils peuvent être appliqués sur ce jeu de test sans induire de biais dans l'analyse. Nous ne pouvons par contre pas les utiliser pour appliquer la validation croisée imbriquée sur le jeu de développement, car ces segmentations automatiques auraient été biaisées.

Une fois les métriques extraites nous les avons utilisés pour prédire l'appartenance des 6 patients au groupe TILsum_haut ou TILsum_bas. Nous avons ainsi pu calculer et comparer l'exactitude de prédiction de chacun des modèles en fonction de la segmentation des lésions utilisée.

5.3 Résultats

Les résultats de la validation croisée imbriquée sur le jeu de développement, de l'interprétation des modèles, et de l'évaluation sur le jeu de test sont détaillés et décrits ci-dessous.

5.3.1 Évaluation du pouvoir prédictif des métriques sur le jeu de développement

Les résultats de l'évaluation du notre pipeline de classification à travers la validation croisée imbriquée sur le jeu de développement sont disponibles sous forme d'AUC moyenne +- écart type sur le Tableau 22. Ce tableau contient les résultats des 12 expériences, à partir des 3 différentes segmentations.

	Seg. auto 4-classes : BLAST-CT	Seg. manuelle 4-classes	Seg. manuelle 7-classes
Exp1	54 +-24		
Exp2	66 +-24		
Exp3	59 +-24	63 +-22	
Exp4	61 +-27	69 +-24	75 +-27
Exp5	65 +-25	71 +-25	
Exp6	64 +-25	68 +-26	73 +-24
Exp7	60 +-23	74 +-26	89 +-17
Exp3_bis	52 +-27	54 +-27	
Exp4_bis	56 +-29	63 +-26	70 +-25
Exp5_bis	66 +-25	71 +-21	
Exp6_bis	61 +-27	68 +-26	75 +-23
Exp7_bis	65 +-25	73 +-25	86 +-18

Tableau 22. AUC moyenne +- écart type (%) des performances des modèles entraînés validés et évalués sur le jeu d'entraînement suite à une validation croisée imbriquée. Résultats des 12 expériences pour les 3 segmentations utilisées. Meilleure AUC en vert. Meilleure AUC à 4 classes en bleu.

La première observation possible est que les métriques cliniques utilisées ne permettent pas à elles seules de prédire l'appartenance d'un patient au groupe TILsum_haut ou TILsum_bas (Tableau 22 -Exp1 : AUC=54+-24%). L'ajout des scores d'imagerie (score de Marshall et de Rotterdam) permet d'arriver à une AUC de 66%, ce qui n'est néanmoins pas satisfaisant (Tableau 22 - Exp2). Rajouter séparément des informations de volume global, par type de lésion ou de localisation, permet d'atteindre au maximum une AUC de 75 +-27% pour l'Exp4 (Tableau 22), et on obtient des résultats similaires en concaténant les volumes par type de lésion et ceux par zones de l'atlas (Tableau 22- Exp6 et 6_bis). Les meilleurs résultats sont de loin obtenus pour les expériences 7 et 7_bis, qui font intervenir

les volumes de chaque type de lésion dans chaque zone de l'atlas (Tableau 22 - Exp7 et 7_bis ; AUC = 89 % +-17 et 86 +-18 ; respectivement). Il est également à noter que les écarts types des AUC, calculés sur les AUC des 60 jeux de test de la boucle externe de la validation croisée imbriquée, sont très élevés, même si les deux écarts types les plus faibles sont observés pour les AUC les plus élevées (Expériences 7 et 7_bis segmentation manuelle à 7 classes).

La deuxième observation est que l'ajout des données cliniques aux métriques extraites par notre pipeline de quantification (Expériences « _bis ») n'améliore globalement pas les performances de classification.

La troisième observation concerne la précision des segmentations utilisées pour quantifier le profil lésionnel des patients. Les résultats provenant d'une segmentation à 7 classes sont systématiquement meilleurs que ceux provenant d'une segmentation à 4 classes. De plus, la segmentation automatique issue de BLAST-CT fournit systématiquement les moins bons résultats que les segmentations manuelles.

5.3.2 Interprétation des modèles

Le calcul de l'importance des métriques n'étant pertinent que sur les modèles qui parviennent à de bonnes performances de classification, nous présentons ici les métriques les plus influentes sur les expériences et les segmentations dont l'AUC résultante est supérieure ou égale à 70%. Sur le Tableau 23, les deux métriques les plus importantes sont présentées, sauf si l'influence de la 3ème métrique la plus importante égale ou dépasse les 20%, auquel cas cette métrique est également affichée.

	Segmentation manuelle 4-classes	Segmentation manuelle 7-classes
Exp4	-	1/ HSD (53%) 2/ HSA (26%) 3/ HIV (20 %)
Exp5	1/ LF (55%) 2/ IS (20%)	
Exp6	-	1/ HSD (50%) 2/ LF (26%)
Exp7	1/ HEA dans le LF (38%) 2/ HEA dans le LO (31%) 3/ HIV dans le RC (25%)	1/ HSD dans le LP (47%) 2/ HSD dans le LF (33%)
Exp4_bis	-	1/ HSD (60%) 2/ Hb (19%)
Exp5_bis	1/ LF (44%) 2/ LO (23%) 3/ IS (20%)	
Exp6_bis	-	1/ HSD (50%) 2/ IS (27%) 3/ LF (22%)

Exp7_bis	1/ HEA dans le LF (38%) 2/ HEA dans le LO (31%) 3/ HIV dans le RC (25%)	1/ HSD dans le LP (53%) 2/ HSD dans le LF (25%)
-----------------	---	--

Types de lésions – HSD : Hémorragie sous durale, HSA : Hémorragie sous arachnoïdienne, HIV : Hémorragie intra ventriculaire, HEA : Hémorragie extra axiale.

Localisation – LF : Lobe frontal, LP : Lobe Pariétal, LO : Lobe occipital, IS : Insula, RC : Reste du cerveau

Métriques cliniques – Hb : Taux d'hémoglobine

Légende couleur de police : **Volume d'un type de lésion** / **Volume lésionnel dans une zone structurale de l'atlas** / **Volume d'un type de lésion dans une zone de l'atlas** / **Métrique clinique**

Légende couleur de remplissage – **Vert** : Meilleure AUC moyenne et expérience pour laquelle le meilleur modèle a été sauvegardé pour devenir notre modèle de prédiction à partir d'une segmentation à 7 classes, **Bleu** : Meilleure AUC moyenne à 4 classes et expérience pour laquelle le meilleur modèle a été sauvegardé pour devenir notre modèle de prédiction à partir d'une segmentation à 4 classes.

Tableau 23. Métriques les plus importantes du meilleur modèle pour les expériences et segmentations dont l'AUC égale ou dépasse 70%.

Le calcul de l'importance des métriques utilisées par les meilleurs modèles fait ressortir l'influence des volumes d'hémorragie sous durale dans les lobes frontal et pariétal, qui à eux deux font 80% de la prédiction des deux meilleurs modèles (Exp 7 et 7_bis avec segmentation manuelle à 7 classes). Dans le cas de la segmentation manuelle à 4 classes, les métriques les plus importantes sont les volumes l'hémorragie extra axiale dans les lobes frontal et occipital (environ 70% de la prédiction avec ces deux métriques).

5.3.3 Évaluation des modèles et des segmentations automatiques sur le jeu de test

Les deux modèles sauvegardés sont tous les deux issus de l'expérience 7 et ont été entraînés à partir des segmentations manuelles sur le jeu de développement. Les résultats relatifs au modèle à 4 classes sont surlignés en bleu clair et ceux relatifs au modèle à 7 classes sont surlignés en vert clair.

Appliqués sur le jeu de test et pour les 6 segmentations différentes, ces deux modèles fournissent des scores d'exactitudes sur la prédiction, illustrés sur le Tableau 24, entre 50% et 83%.

Les scores d'exactitude équilibrée (corrigeant le déséquilibre du nombre d'échantillons entre les deux groupes), de précision, de sensibilité et de spécificité sont également présentés sur le Tableau 24.

Modèle de prédiction	Segmentation	Jeu de test : 6 patients (4/2)				
		Exactitude	Exactitude équilibrée	VPN	Spécificité	Sensibilité
Modèle à 4 classes sauvegardé	BLAST-CT	0,50	0,38	0,60	0,75	0
	TL_BLAST4C	0,67	0,75	1	0,50	1
	Manuelle 4 classes	0,67	0,63	0,75	0,75	0,50
Modèle à 7 classes sauvegardé	Random_7C	0,83	0,88	1	0,75	1
	TL_RE_7C	0,83	0,88	1	0,75	1
	Manuelle 7 classes	0,83	0,88	1	0,75	1

Tableau 24. Résultats d'exactitude, d'exactitude équilibrée, de précision, de sensibilité, de spécificité, et d'AUC sur la prédiction de l'intensité thérapeutique des 2 modèles sauvegardés à partir de 6 segmentations différentes sur le jeu de test de la cohorte Radiomic-TBI.

Les meilleurs résultats sont obtenus avec le modèle à 7 classes.

On s'aperçoit que la segmentation BLAST-CT ne peut pas se substituer à la segmentation manuelle à 4 classes mais que la segmentation issue du fine-tuning de BLAST fournit des résultats équivalents à la segmentation manuelle (67%). C'est également le cas pour le modèle à 7 classes où les segmentations issues du transfer learning de BLAST affichent des performances aussi bonnes que la segmentation manuelle à 7 classes (83%). Le seul patient sur lequel la segmentation à 7 classes commet une erreur est le patient P08, qui a un TILsum maximum pendant les 7 premiers jours en réanimation égal à 8, et donc appartient au groupe TILsum_bas, alors qu'il a été prédit comme appartenant au groupe TILsum_haut.

5.4 Interprétation

Les résultats de cette étude montrent qu'il est possible de prédire le besoin pour un patient TC de subir une journée d'intensité thérapeutique extrême au cours de la première semaine en réanimation à partir de l'image TDM acquise à l'arrivée à l'hôpital, et que les segmentations automatiques développées au chapitre 4 permettent de rendre cette prédiction entièrement automatique.

Plus précisément, il semblerait que le volume d'hémorragie sous durale dans les lobes frontal et pariétal soient les 2 métriques les plus reliées à la nécessité d'une prise en charge thérapeutique extrême. Si les informations de type de lésion et de localisation sont naturellement corrélées, c'est en combinant les deux informations que les meilleurs résultats prédictifs sont atteints. A partir de la segmentation à 4 classes, les métriques les plus importantes font intervenir l'hémorragie extra axiale, qui englobe l'hémorragie sous durale. Les résultats sont toutefois bien meilleurs lorsque l'hémorragie sous durale est discriminée. Cette importance de l'HSD, principalement dans les lobes frontal et pariétal, est cohérente avec la littérature [128] [129] [130] et celle de l'HEA dans le lobe frontal est même confirmée par une étude d'imagerie très récente [104], menée en partie par l'équipe qui a développé l'algorithme BLAST-CT, dont le préprint a été publié en août 2022, et qui fait écho à nos travaux. Cette étude a cherché à prédire si le score GOS étendu d'un patient sera inférieur ou égal à 4, six mois après le traumatisme. Ils ont pour cela obtenu les images TDM acquises à l'admission à l'hôpital de 646 patients et ont segmenté l'HIP, HEA, œdème et HIV grâce à l'algorithme BLAST-CT. Ils ont en outre estimé la localisation des lésions en divisant les images TDM en 64 cubes et en mesurant le volume lésionnel dans chacun de ces cubes. Ils ont ainsi montré que cette information de volume lésionnel par type et localisation, associée à des données cliniques (Age, composante motrice du GCS, réponse pupillaire, hypoxie et hypotension) était capable de prédire l'état du patient à 6 mois avec une AUC de 0,88. Ils ont également trouvé que ce qui influence le plus la prédiction (du GOS inférieur ou égal à 4 ou bien supérieur) est le volume d'hémorragie extra axiale dans la zone frontale.

Cette étude, sortie après la réalisation mais avant la publication de nos travaux, conforte nos résultats sur l'importance de l'HEA dans les lobes du cerveau. Nous pensons à ce sujet être allés plus loin en montrant de manière plus précise que la lésion responsable d'une intensité thérapeutique élevée est surtout l'HSD. Cette étude de Rosnati [104] est plus robuste et puissante que la nôtre car elle fait intervenir 20 fois plus de patients que nous avons pu en étudier au sein de la cohorte Radiomic-TBI. Néanmoins, le fait d'aboutir à des résultats similaires nous conforte sur les procédures mises en place pour pallier ce petit nombre de données, comme le réentraînement des algorithmes de segmentation afin de les affiner sur la nature des données étudiées ou la procédure de validation croisée imbriquée, qui permet d'utiliser tous les échantillons de la cohorte pour à la fois entraîner, valider, et évaluer les performances.

Intuitivement, on voit bien que l'ajout de données clinique devrait permettre une meilleure prédiction, pourtant les métriques cliniques que nous avons ajoutées en support des métriques extraites de l'image TDM n'ont rien apporté. Cela pourrait s'expliquer par le fait que les données cliniques ajoutées ne sont pas les plus pertinentes et auraient pu être complétées par d'autres comme par exemple les données de pupillométrie.

Finalement, si l'étude sur la segmentation au chapitre 3 a permis de mettre en évidence les différences flagrantes entre la segmentation manuelle à 4 classes et celle issue de BLAST-CT, cette étude de classification montre que le manque de précision de la

segmentation issue de BLAST-CT est tel qu'il ne permet simplement pas de correctement prédire notre critère.

Ces résultats ont été confirmés lors de l'évaluation des meilleurs modèles de classification sur le jeu de test. Cette évaluation interne a également permis de montrer que les segmentations automatiques développées au chapitre 3 étaient suffisamment proches des segmentations manuelles pour obtenir les mêmes résultats de classification que les segmentations manuelles.

Bien évidemment, cette étude a plusieurs limitations. Premièrement le faible nombre de patients qui a été compensé par des méthodes de classifications robustes aux petits jeux de données, mais qui résultent en une instabilité des modèles, visible à travers les valeurs élevées des écarts types entre les AUC des 60 modèles sur les jeux de test. Pour pallier cette limitation, nous avons validé les deux meilleurs modèles de classification sur une base de données d'une autre étude. Cette validation externe est détaillée au chapitre 6.

De plus, les segmentations manuelles sont subjectives et bien qu'elles aient fait l'objet d'un consensus entre deux observateurs, des variabilités inter observateurs existent et donc remettent en question l'existence d'une vérité terrain indiscutable.

Finalement, les images TDM sont toutes issues du même centre et de la même étude, ce qui ne permet pas de conclure sur la performance de ces outils sur des images provenant d'un autre centre ou d'une autre étude. Pour évaluer cette influence, nous avons mené une étude de la variabilité introduite sur les images TDM par les différents protocoles d'acquisition. Cette étude est détaillée au chapitre 8.

Chapitre 6 : Validation externe et mise à disposition clinique

Nous avons exposé dans les chapitres 4 et 5 le développement de deux outils : i/ un outil de segmentation des lésions typiques du TC à partir d'images TDM et ii/ un outil de prédiction de l'intensité thérapeutique nécessitée par un patient pendant ses 7 premiers jours en réanimation à partir de son image TDM acquise le jour de son arrivée au CHU. Ces deux outils ont été évalués sur un sous ensemble de la cohorte Radiomic-TBI, d'où proviennent également les données utiliser pour entraîner ces deux outils. Par conséquent l'évaluation de ces outils a jusqu'ici été faite de manière interne. De plus, cette évaluation a été menée sur un faible nombre de patients (6 patients pour la classification, soit 17 images TDM pour la segmentation). Dans le but d'évaluer les performances de ces outils sur des données de nature différente, nous avons évalué de manière externe ces deux outils sur 12 images TDM de 12 patients appartenant à une seconde étude. Nous faisons l'hypothèse que les résultats obtenus sur le jeu de test de la cohorte Radiomic-TBI seront vérifiés.

Après cette étude de validation externe de nos outils, nous avons travaillé à la mise à disposition de ces outils au monde médical. Cet objectif affiché a été possible grâce à la plateforme Virtual Imaging Platform (VIP) qui permet d'exécuter en ligne, à distance, et sur des ressources de calcul adaptées des pipelines de traitement d'image complexes de façon extrêmement simple. Ainsi, nos outils sont aujourd'hui disponibles sur VIP et peuvent être utilisés facilement par toute personne le désirant.

Ce chapitre se décompose donc en deux parties, la première consacrée à la validation externe de nos outils de segmentation et de prédiction et la deuxième à l'intégration sur VIP de nos outils.

6.1 Validation externe de nos outils de segmentation et de classification

Notre objectif est de valider sur une seconde cohorte, externe, notre algorithme de segmentation automatique des différentes lésions typiques post-TC observées en TDM (chapitre 4) ainsi que notre modèle statistique de stratification des patients en fonction du niveau d'intensité thérapeutique nécessaire à leur prise en charge pendant leur première semaine d'hospitalisation (chapitre 5).

6.1.1 Matériel et méthodes

6.1.1.1 Données utilisées

Nous souhaitons récupérer des images TDM de patients TC imagés au CHUGA afin de tester notre outil de segmentation. Pour pouvoir également tester notre outil de classification sur les mêmes patients, il fallait être en mesure de récupérer pour chaque patient le TILsum associé aux premiers jours en réanimation. Puisque le TILsum est un score uniquement utilisé en recherche et non en routine clinique, nous avons récupéré les images TDM d'une autre étude clinique menée au CHUGA sur le TC, pour laquelle les patients ont été imagés au scanner à l'arrivée au CHU et suivis en réanimation pendant plusieurs jours avec un calcul quotidien du TILsum.

Nous avons ainsi écrit un protocole de recherche clinique rétrospective, nommée RadioxyTC, afin de collecter les images TDM, les données cliniques, et les TILsum de 12 patients admis au CHUGA pour TC entre août 2018 et avril 2021, ayant été inclus dans un autre projet de recherche clinique appelé OxyTC mais non inclus dans l'étude Radiomic-TBI.

La catégorisation des patients en TILsum_bas/TILsum_haut n'a pas pu être réalisé selon la même méthodologie que pour la cohorte Radiomic-TBI (utilisée pour développer, valider et tester notre modèle de prédiction ; chapitre 5). En effet, pour 12 patients d'OxyTC, les données cliniques collectées nous ont permises de calculer le TILsum que pour les 5 premiers jours après l'accident alors qu'il avait été calculé pour les 8 premiers jours pour les patients de Radiomic-TBI. Afin de pallier ce manque de données, le Dr Jules Grèze, anesthésiste réanimateur avec 10 ans d'expérience, a lu les comptes rendus de fin d'hospitalisation de ces 12 patients, et a ainsi pu estimer le TILsum des 3 derniers jours en fonction de la prise en charge de ces patients. Grâce à cette estimation, nous avons pu déterminer quels patients avaient subi un jour d'intensité thérapeutique extrême durant les 8 premiers jours en réanimation après l'accident initial et ainsi classer les patients dans les deux groupes TILsum_bas et TILsum_haut de façon similaire à l'étude du chapitre 5.

Puis, sur ces 12 images TDM, Dr Arnaud Attyé, Dr Jules-Arnaud de Busschère et Florian Dhaussy Tornior, respectivement neuroradiologue, anesthésiste réanimateur, et interne en anesthésie réanimation, ont segmenté manuellement les 7 mêmes types de lésion qu'étudié précédemment pour le développement de notre outil de segmentation automatique :

- L'hématome intra parenchymateux (HIP)
- L'hémorragie sous durale (HSD)
- L'hémorragie extra durale (HED)
- L'hémorragie intra ventriculaire (HIV)
- L'hémorragie sous arachnoïdienne (HSA)
- Les pétéchies (P)
- L'œdème (O)

Ici encore, on peut fusionner l'HSD, l'HED et l'HSA pour former l'Hémorragie Extra Axiale (HEA) et fondre les Pétéchies dans l'HIP afin d'obtenir une 2ème segmentation manuelle : la segmentation manuelle à 4 classes, comme illustré sur le Tableau 25.

Segmentation à 4 classes	Segmentation à 7 classes
Hémorragie intra parenchymateuse (HIP)	Hémorragie intra parenchymateuse (HIP)
	Pétéchies (P)
Œdème (O)	Œdème (O)
Hémorragie intra ventriculaire (HIV)	Hémorragie intra ventriculaire (HIV)
Hémorragie extra axiale (HEA)	Hémorragie sous durale (HSD)
	Hémorragie extra durale (HED)
	Hémorragie sous arachnoïdienne (HSA)

Tableau 25. Résumé et lien entre les lésions de la segmentation à 4 classes et à 7 classes.

6.1.1.2 Évaluation de la segmentation automatique

Sur ces 12 images TDM, nous avons appliqué 6 modèles de segmentation automatique étudiés au chapitre 4 :

- BLAST-CT, la segmentation de l'article originale de Monteiro [102].
- Random_4C, la segmentation à 4 classes obtenue après entraînement du réseau BLAST-CT sur nos données.
- FT_4C, la segmentation à 4 classes obtenue après réentraînement du modèle BLAST-CT sur nos données, par fine-tuning.
- Random_7C, la segmentation à 7 classes obtenue après entraînement du réseau BLAST-CT sur nos données.
- TL_7C, la segmentation à 7 classes obtenue après réentraînement du modèle BLAST-CT sur nos données, par transfer learning.

- TL_RE_7C, la segmentation à 4 classes obtenue réentraînement du modèle TL_BLAST4C sur nos données, par transfer learning.

Ces 6 segmentations automatiques ont été comparées aux segmentations manuelles (segmentation manuelle à 4 classes ou segmentation manuelle à 7 classes) grâce au score de Dice. Ainsi nous comparons 3 segmentations automatiques à 4 classes entre elles et 3 segmentations automatiques à 7 classes entre elles. Le test bilatéral des rangs signés de Wilcoxon a été utilisé pour estimer une différence significative entre les segmentations. La significativité des résultats a été classée en 3 types en fonction de la p-valeur résultante du test statistique :

- La p-valeur est inférieure à 0.05 : significativité faible
- La p-valeur est inférieure à 0.01 : significativité modérée
- La p-valeur est inférieure à 0.001 : significativité forte

6.1.1.3 Évaluation de la classification

Lors de notre étude de prédiction du niveau d'intensité thérapeutique nécessitée par un patient durant sa première semaine en réanimation (Chapitre 5), nous avons évalué le pouvoir de prédiction de différents jeux de métriques, et nous avons montré que le jeu de métriques responsable des meilleurs scores de prédiction est celui rassemblant les volumes lésionnels en fonction des différents types de lésions et des différentes localisations des lésions. Nous avons également sauvegardé le meilleur modèle de classification utilisant les segmentations à 4 classes et celui utilisant les segmentations à 7 classes. Pour rappel, ces deux modèles sont résumés sur le Tableau 26.

Nom du modèle	Jeu de métriques	Métriques les plus influentes	AUC moyenne durant la validation croisée imbriquée sur le jeu de développement	Exactitude de prédiction sur le jeu de test de la cohorte Radiomic-TBI
Modèle à 4 classes sauvegardé	Volume de lésion par type et localisation	1/ HEA dans le LF (38%) 2/ HEA dans le LO (31%) 3/ HIV dans le RC (25%)	74% +-26%	67%
Modèle à 7 classes sauvegardé		1/ HSD dans le LP (47%) 2/ HSD dans le LF (33%)	89% +-17%	83%

Tableau 26. Résumé des 2 modèles sauvegardés au chapitre 5 : Le meilleur modèle utilisant les segmentations à 4 classes (en bleu) et celui utilisant les segmentations à 7 classes (en vert).

Nous avons évalué ces deux modèles sur la cohorte RadioxyTC de la même manière que nous l'avons fait sur le jeu de test de la cohorte Radiomic-TBI au chapitre 5.

Nous avons donc premièrement extrait les métriques représentant le profil lésionnel des patients grâce à notre pipeline de quantification à partir de 6 différentes segmentations (2 manuelles et 4 automatiques) :

- BLAST-CT
- FT_4C
- Segmentation manuelle à 4 classes
- Random_7C
- TL_RE_7C
- Segmentation manuelle à 7 classes

Ainsi, nous avons recalé l’atlas structurel présenté au chapitre 5 sur chacune des 12 images de la cohorte RadioxyTC et pour chacune de ces segmentations, nous avons extrait le volume de chaque type de lésion dans chacune des zones de l’atlas. Nous avons ensuite appliqué à ces métriques le modèle sauvegardé correspondant, selon s’il prend en entrée des métriques extraites à partir d’une segmentation à 4 ou à 7 classes. Ainsi, pour chaque segmentation, nous avons obtenu la prédiction d’appartenance de chacun des 12 patients aux groupes TILsum_haut/ TILsum_bas. Grâce à ces prédictions, nous avons pu calculer l’exactitude de prédiction, ainsi que l’exactitude équilibrée, la spécificité, la sensibilité et la précision, qui ont été définies au Chapitre 1.

6.1.2 Résultats

6.1.2.1 Caractérisation de la cohorte RadioxyTC et comparaison avec la cohorte Radiomic-TBI

La caractérisation de la cohorte par groupe est disponible sur le Tableau 27. Il est à noter que la valeur moyenne du TILsum maximum dans le groupe TILsum_haut est de 13. Ce chiffre indique que de nombreux patients de ce groupe ont un TILsum maximum égal ou légèrement supérieur à 11 ; ce qui est la limite entre les groupes TILsum_bas et TILsum_haut.

Ce résultat signifie que les patients de la cohorte RadioxyTC qui font partie du groupe TILsum_haut ont globalement eu besoin d’un niveau d’intensité thérapeutique plus faible que ceux de la cohorte Radiomic-TBI.

	Cohorte RadioxyTC		Cohorte Radiomic-TBI	
	TILsum_bas	TILsum_haut	TILsum_bas	TILsum_haut
Nb Patients	7	5	20	9
Age (année)	32,3 +-14,4 [19-55]	44,6 +-12,7 [27-60]	48,3 +-23,2 [18-79]	42,1 +-18,6 [22-73]
Sexe	7M / 0F	3M / 2F	13M / 7F	8M / 1F
Poids (kg)	76,1 +-7,0 [70-90]	86,0 +-14,9 [72-110]	70,8 +-13,6 [51-100]	76,7 +-6,1 [68-86]
Taille (cm)	175,-7 +-4,5 [170-180]	174,0 +-9,6 [165-190]	175,3 +-9,8 [160-197]	175,5 +-5,8 [169-185]
GCS	5,4 +-1,8 [3-7]	6,6 +-0,5 [6-7]	9,3 +-4,3 [3-15]	6,9 +-4,0 [3-14]
TILsum max	6,6 +-2,4 [4-10]	13,0 +-3,5 [11-19]	6,3 +-2,6 [1-10]	18,9 +-4,5 [11-28]

Tableau 27. Caractérisation de la cohorte RadioxyTC et de la cohorte Radiomic-TBI. (Moyenne +-ecart type [Min-Max])

Le volume des lésions manuellement segmentées sur la cohorte RadioxyTC est comparé avec celui de la cohorte Radiomic-TBI sur le Tableau 28.

	Cohorte RadioxyTC	Cohorte Radiomic-TBI
Nombre de patients	12	29
Nombre d'images TDM	12	84
HIP Volume % des lésions Nombre d'images contenant la lésion	18,1 cm3 10,1% (n=7)	694,6 cm3 15,1% (n=62)
HSD	103,3 cm3 57,4% (n=5)	1322,2 cm3 28,7% (n=66)
HED	3,0 cm3 1,7% (n=2)	360,9 cm3 7,8% (n=19)
HIV	1,9 cm3 1,1% (n=3)	19,2 cm3 0,4% (n=30)
HSA	28,1 cm3	309,4 cm3

	15,6% (n=9)	6,7% (n=63)
Pétéchies	0,8 cm3 0,4% (n=6)	9,1 cm3 0,2% (n=34)
Œdème	24,6 cm3 13,7% (n=4)	1885,5 cm3 40,1% (n=61)
Volume lésionnel total	179,9 cm3 100% (n=11)	4601,2 cm3 100% (n=84)

Tableau 28. Caractérisation des volumes lésionnels manuellement délimités au sein de la cohorte RadioxyTC et de la cohorte Radiomic-TBI - Volume de lésion (Pourcentage du volume de la lésion dans le volume lésionnel total) (Nombre d'images TDM contenant la lésion)

On peut remarquer que le profil lésionnel des patients de la cohorte RadioxyTC diffère largement de celui des patients de la cohorte Radiomic-TBI. En particulier, il y a proportionnellement moins d'HIP, d'HED et d'œdème dans la cohorte RadioxyTC que dans la cohorte Radiomic-TBI. On peut également noter qu'une image TDM de la cohorte RadioxyTC ne présente aucune lésion. Ces différences sont probablement liées aux différences cliniques remarquées sur le Tableau 27 et prouvent la nature différente de ces deux cohortes

6.1.2.2 Évaluation de la segmentation

Les résultats de la comparaison du score de Dice moyen des 3 segmentations automatiques à 4 classes sont disponibles sur le Tableau 29 et sont représentés sous forme d'histogrammes sur la Figure 47.

	Lésion globale	HIP	HEA	O	HIV
BLAST-CT	0,33 (0,06)	0,29 (0,08)	0,17 (0,06)	0,16 (0,10)	0,15 (0,15)
Random4C	0,30 (0,07)	0,25 (0,09)	0,25 (0,08)	0,31 (0,16)	0,31 (0,22)
FT_4C	0,35 (0,07)	0,42 (0,11)	0,30 (0,08)	0,33 (0,17)	0,45 (0,22)

Tableau 29. Moyenne et erreur standard des scores de Dice calculés entre les segmentations automatiques à 4 classes et les segmentations manuelles sur la cohorte RadioxyTC.

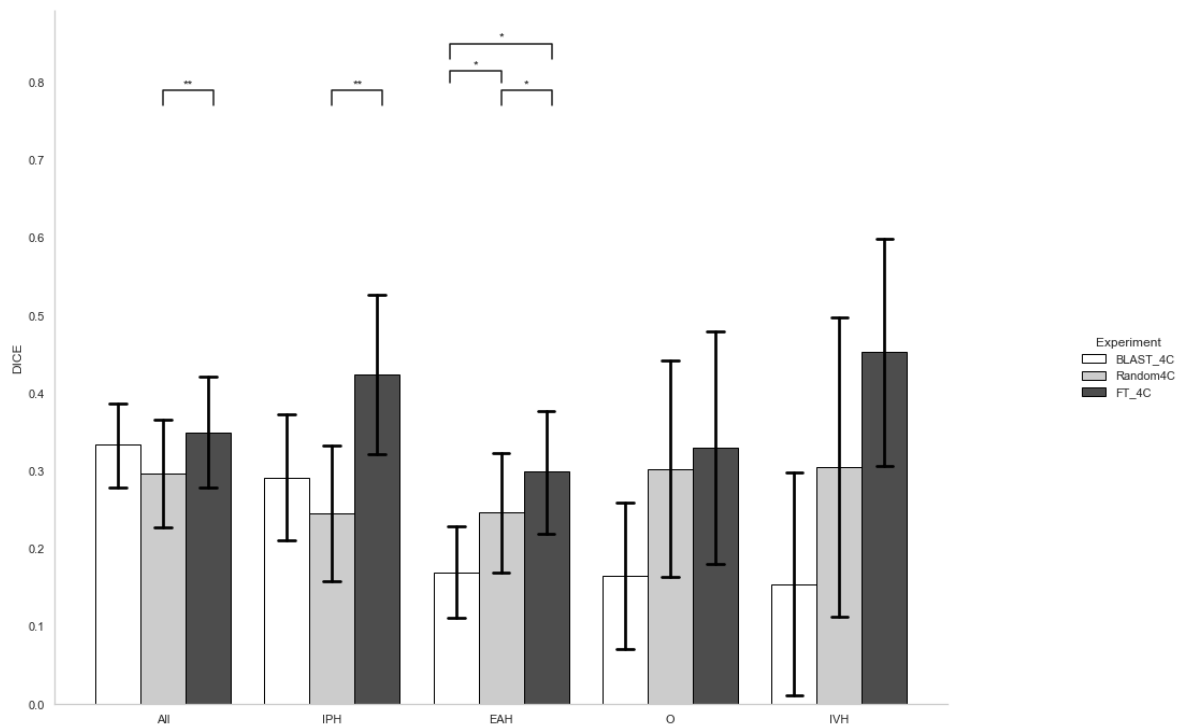


Figure 47. Représentation des résultats du Tableau 29 sous forme d'histogrammes. Significativité calculée par le test de Wilcoxon.

Comme lors de l'évaluation sur le jeu de test de la cohorte Radiomic-TBI, les meilleurs scores de Dice sont obtenus pour la segmentation FT_4C, issue du réentraînement par fine tuning du modèle BLAST-CT. Pour toutes les lésions les scores de Dice de la segmentation FT_4C sont meilleurs que ceux de BLAST-CT. Néanmoins cette fois la différence n'est pas systématiquement significative selon le test de Wilcoxon. Si les différences sur la lésion globale sont faibles (0,35 contre 0,33, donc +0,02), elles sont bien plus importantes sur les lésions prises individuellement (de +0,13 à +0,30).

La segmentation issue du modèle Random4C est constamment moins bonne que FT_4C, et est aussi moins bonne que BLAST-CT sur la lésion globale, ce qui n'était pas le cas lors de l'évaluation sur le jeu de test de la cohorte Radiomic-TBI.

Les résultats de la comparaison du score de Dice moyen des 3 segmentations automatiques à 7 classes sont disponibles sur le Tableau 30 et sont représentés sous forme d'histogrammes sur la Figure 48.

	Lésion globale	HIP	HSD	HED	HIV	HSA	P	O
Random_7C	0,32 (0,08)	0,31 (0,12)	0,41 (0,11)	0,15 (0,15)	0,35 (0,22)	0,11 (0,03)	0,32 (0,11)	0,29 (0,15)
TL_7C	0,40 (0,07)	0,47 (0,13)	0,45 (0,12)	0,09 (0,09)	0,40 (0,22)	0,13 (0,03)	0,41 (0,11)	0,34 (0,17)
TL_RE_7C	0,41 (0,07)	0,48 (0,13)	0,46 (0,12)	0,14 (0,14)	0,44 (0,22)	0,14 (0,04)	0,41 (0,10)	0,35 (0,17)

Tableau 30. Moyenne et erreur standard des scores de Dice calculés entre les segmentations automatiques à 7 classes et les segmentations manuelles sur la cohorte RadioxyTC.

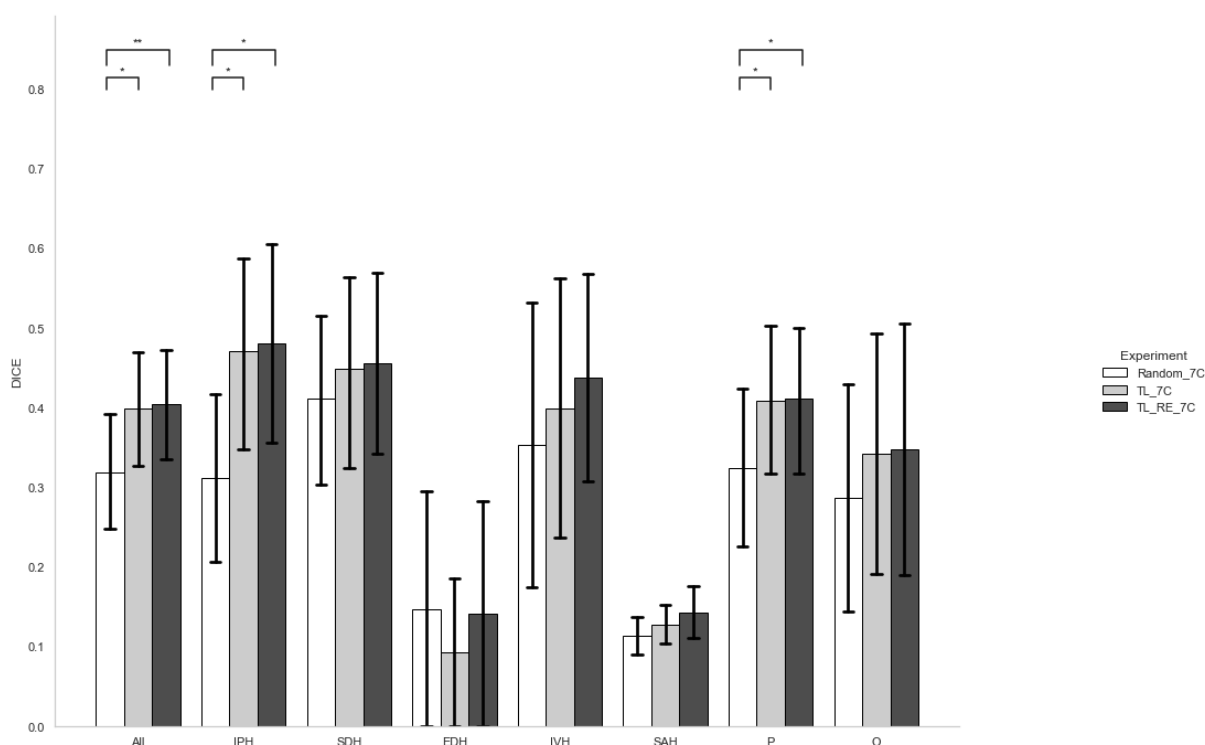


Figure 48. Représentation des résultats du Tableau 30 sous forme d'histogrammes. Significativité calculée par le test de Wilcoxon.

Comme lors de l'évaluation sur le jeu de test de la cohorte Radiomic-TBI, les meilleurs scores de Dice sont obtenus pour la segmentation TL_RE_7C, issue du réentraînement par transfert learning du modèle FT_4C. Les performances de la segmentation TL_7C, issue du réentraînement par transfert learning du modèle BLAST-CT sont proches mais moins bonnes que celles de TL_RE_7C, notamment sur l'HEH et l'HIV, bien qu'aucune différence ne soit significative. La segmentation Random_7C, issue d'un entraînement à partir d'un réseau non pré-entraîné sur nos données, est significativement moins bonne que les 2 autres sur la lésion globale, ce qui est également le cas sur l'HIP et les Pétéchies. Sur l'HED, cette segmentation affiche de meilleurs scores de Dice que les deux autres segmentations

obtenues par transfer learning, ce qui pourrait s'expliquer par le faible nombre d'images TDM contenant de l'HED au sein de la cohorte (2/12).

6.1.2.3 Évaluation de la classification

Les résultats de classification des deux modèles sauvegardés sur la cohorte RadioxyTC sont disponibles sur le Tableau 31.

Modèle de prédiction	Segmentation	Validation externe : 12 patients (Répartition 7/5)				
		Exactitude	Exactitude équilibrée	VPN	Spécificité	Sensibilité
Modèle à 4 classes sauvegardé	BLAST-CT	0,67 (8 sur 12)	0,60	0,64	1	0,20
	FT_4C	0,67 (8 sur 12)	0,63	0,67	0,86	0,40
	Manuelle 4 classes	0,67 (8 sur 12)	0,60	0,64	1	0,20
Modèle à 7 classes sauvegardé	Random_7C	0,75 (9 sur 12)	0,70	0,70	1	0,40
	TL_RE_7C	0,83 (10 sur 12)	0,80	0,78	1	0,60
	Manuelle 7 classes	0,67 (8 sur 12)	0,60	0,64	1	0,20

Tableau 31. Résultats d'exactitude, de précision, de sensibilité, et d'AUC sur la prédiction de l'intensité thérapeutique des 2 modèles sauvegardés à partir de 6 segmentations différentes sur la cohorte RadioxyTC.

Si on se concentre sur l'exactitude, on s'aperçoit, comme pour l'évaluation sur le jeu de test de la cohorte Radiomic-TBI, que le modèle de classification à partir d'une segmentation à 7 classes permet de mieux classer les patients que celui à partir d'une segmentation à 4 classes. Si sur le modèle à 4 classes, les 3 segmentations affichent le même score d'exactitude (0,67), sur le modèle à 7 classes, les métriques extraites des deux segmentations automatiques entraînées sur nos données prédisent mieux l'intensité thérapeutique que celles extraites à partir de la segmentation manuelle. La segmentation affichant le meilleur score d'exactitude est la segmentation TL_RE_7C, qui classe bien 10 patients sur 12.

L'exactitude équilibrée et la précision se comportent sensiblement de la même manière que l'exactitude. On peut observer que si la spécificité est toujours supérieure ou égale à 0,70, la sensibilité, elle, ne dépasse pas les 0,60. On a donc tendance à trop classer les patients dans le groupe TILsum_bas, et ainsi rater parfois quelques patients du groupe TILsum_haut. Il est à noter que dans le domaine médical, on préférerait privilégier la sensibilité à la spécificité, car le coût humain et financier d'un faux diagnostic pathologique est plus faible que celui d'un faux diagnostic sain. Finalement, l'AUC est sensiblement la même pour les modèles à 4 et à 7 classes et toutes les segmentations à l'exception notable de la segmentation FT_4C.

6.1.3 Interprétation

Sur une cohorte de patients de nature très différente que celle sur laquelle ont été entraînés nos outils de segmentation et de classification, les résultats de l'application de ces outils ont confirmé ceux obtenus sur le jeu de test de la cohorte Radiomic-TBI. Les résultats de la segmentation montrent encore une fois une supériorité des modèles obtenus par les approches de réentraînement (fine-tuning et transfer learning) par rapport aux modèles entraînés de manière classique ou même au modèle BLAST-CT, et ce, bien que les patients de la cohorte RadioxyTC aient un profil lésionnel bien différent de ceux de la cohorte Radiomic-TBI utilisée pour l'entraînement.

Les résultats de classification à 4 classes sont identiques pour les 3 segmentations testées, qu'elle soit automatique ou manuelle (0,67). Cette valeur d'exactitude de prédiction est la même que la performance du modèle sur le jeu de test de la cohorte Radiomic-TBI. La segmentation manuelle à 7 classes n'améliore pas la prédiction par rapport à celle à 4 classes. Ce résultat pourrait s'expliquer par le fait que les patients du groupe TILsum_haut de la cohorte RadioxyTC ont un score TILsum maximum moyen de 13, ce qui est proche de la limite qui sépare les deux groupes (qui est de 11) par rapport à au score TILsum maximum moyen de la cohorte Radiomic-TBI (qui est de 19), qui a servi de base de données d'entraînement aux modèles de classification. Ainsi la tâche de classification sur RadioxyTC est probablement plus ardue que sur Radiomic-TBI, il n'est donc pas choquant qu'un modèle de classification entraîné sur Radiomic-TBI fournisse des résultats contrastés sur RadioxyTC.

Les segmentations automatiques à 7 classes permettent d'obtenir de meilleurs scores d'exactitude de prédiction, même si cette amélioration ne peut pas être significative sur toutes les lésions au vu du faible nombre d'images évaluées. Cet effet pourrait être mis en relation avec les scores de Dice moins bons qu'escomptés. Puisque le score de Dice est calculé par rapport à la segmentation manuelle, et puisque contrairement à nos attentes la segmentation manuelle fournit des résultats de classification moins bons que la segmentation automatique, on pourrait imaginer que la

segmentation manuelle est défailante. A titre d'exemple, sur l'un des 2 patients bien classés avec la segmentation TL_RE_7C mais mal classé avec la segmentation manuelle à 7 classes, nous avons observé, à posteriori, que la segmentation manuelle a sous-estimé une partie d'HSD qui pourtant est bien détectée avec la segmentation TL_RE_7C. Une illustration de ce patient est disponible Figure 49. De plus, cette partie d'HSD manquée est, d'après l'atlas structurel recalé, localisée dans le Lobe Frontal, ce qui est une des deux métriques les plus importantes du modèle de classification, ce qui aurait tendance à confirmer cette hypothèse. On peut également rappeler que les segmentations manuelles de la cohorte RadioxyTC ont été dessinées par 3 médecins, alors que celles de la cohorte Radiomic-TBI ont-elles été dessinées par 2 de ces 3 médecins. Peut-être que cette différence d'opérateur a engendré une différence de nature entre ces deux jeux de segmentations manuelles.

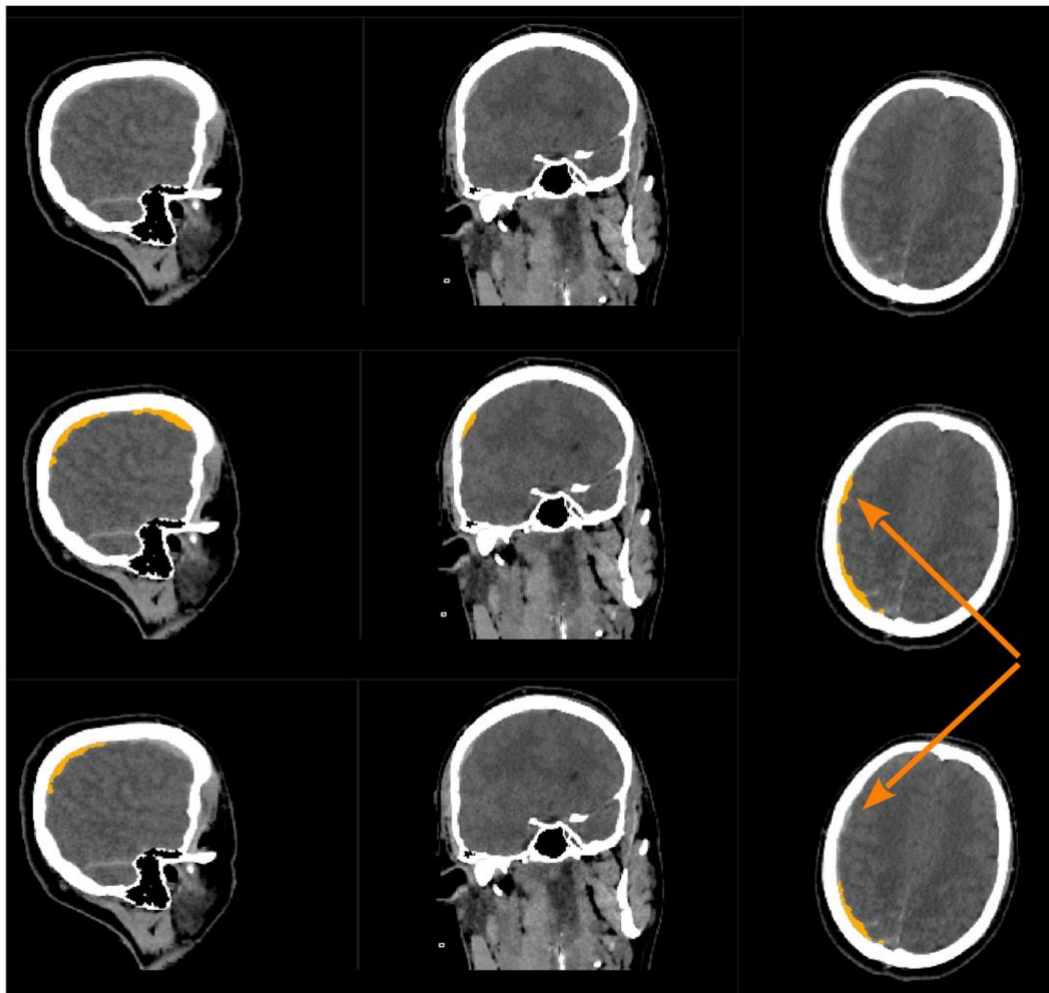


Figure 49. Comparaison de la segmentation de l'HSD entre la segmentation automatique fournie par le modèle TL_RE_7C (ligne du milieu) et la segmentation manuelle à 7 classes (ligne du bas) sur l'un des deux patients bien classés à partir de la segmentation TL_RE_7C mais mal classé à partir de la segmentation manuelle à 7 classes. Les flèches pointent une différence de segmentation d'HSD dans le lobe frontal, qui est une métrique fondamentale de notre modèle de prédiction de l'intensité thérapeutique. Patient numéro 12 de la cohorte RadioxyTC.

Finalement on peut dire que si les segmentations automatiques ne sont évidemment pas exactement similaires aux segmentations manuelles, au vu des performances de classification on peut dire qu'elles permettent d'obtenir des résultats de classification au moins aussi bon qu'avec les segmentations manuelles. Concernant les limitations, une fois encore, le nombre de données est trop faible pour que ces résultats puissent servir de justification pour utiliser ces outils en routine clinique. De plus, toutes les images TDM utilisées ont été acquises au CHUGA. En aucun cas ces outils ont été évalués à partir de données provenant d'un autre site. Ainsi, nous ne savons pas si ces outils se généralisent bien à des données d'un autre type de populations, d'un autre site, utilisant des équipements différents, ou bien sur des données fortement artefactées.

6.2 Mise à disposition clinique des outils développés

Dès le début de la thèse, nous avons eu comme objectif d'aboutir à un outil d'aide à la prise en charge et de le mettre à la disposition des cliniciens. Non pas pour qu'il soit utilisé en routine clinique, car comme expliqué au paragraphe précédent il faut encore valider ces outils sur des cohortes prospectives plus grosses et plus hétérogènes, mais pour que les cliniciens puissent le tester, l'évaluer de manière qualitative en soin courant et nous guider pour l'améliorer.

6.2.1 Première approche : Génération d'un rapport d'imagerie par patient

Nous avons tout d'abord commencé par générer de manière automatique un rapport PDF pour chaque patient, contenant entre autres :

- L'image TDM d'une tranche de cerveau à J0, J1 et J3
- Une segmentation automatique superposée sur les images TDM
- Un histogramme empilé du volume des différentes lésions à J0, J1 et J3, comme illustré Figure 50
- Quelques métriques extraites de l'image, par exemple quelques métriques Radiomiques
- Un résumé des données démographiques du patient (Age Sexe Poids Taille)
- Un résumé des données cliniques du patient (GCS, scores d'imagerie, score TILsum)

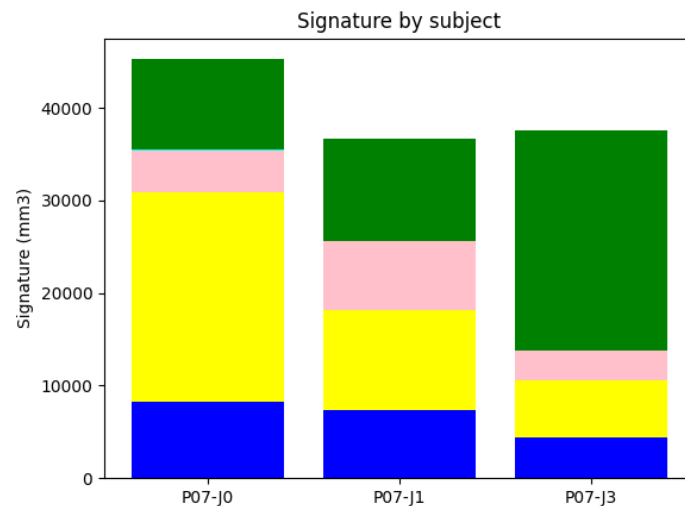


Figure 50. Histogramme empilé du volume de 7 différents types de lésion du patient numéro 7 de la cohorte Radiomic-TBI sur l'image TDM acquise à J0, J1 et J3. Légende : Bleu foncé – HIP, Jaune – HSD, Rose – HSA, Bleu clair – Pétéchies, Vert – Œdème

Ces rapports étaient générés sur ma machine de calcul à l'Institut des Neurosciences de Grenoble et étaient ensuite envoyées aux médecins impliqués dans ce projet. La principale raison de l'abandon de ces rapports est qu'ils incluaient des images 2D ce qui était trop limitant pour les médecins. Le choix des tranches avait beau être automatiquement déterminé de manière à afficher les tranches contenant le plus de lésion, bien souvent, afficher assez de tranches et fournir un rapport concis étaient deux objectifs incompatibles.

6.2.2 Mise à disposition des codes informatiques sur Github

Au fur et à mesure du développement de nos outils de segmentation automatique et de prédiction de l'intensité thérapeutique d'un patient à partir de l'image TDM acquise à l'arrivée au CHU, nous avons cherché à proposer ces outils aux cliniciens.

Tout d'abord nous avons mis nos codes informatiques de traitement des images TDM publiquement à disposition sur un dépôt Github, afin qu'ils puissent être exécutés directement par les médecins au CHU. Néanmoins, l'exécution s'est heurtée aux contraintes informatiques du CHU, c'est à dire à des machines de faible puissance, sans cartes graphiques, partagées entre plusieurs utilisateurs, et où il est difficile d'obtenir les droits pour installer un nouveau logiciel. Pour pallier ces limitations, nous nous sommes tournés vers une plateforme adaptée.

6.2.3 Plateforme VIP : Virtual Imaging Platform

La plateforme nommée Virtual Imaging Platform (VIP) est une interface web permettant d'exécuter à distance et sur des ressources informatiques dédiées des applications de traitement des images médicales [131]. Elle a été développée par le laboratoire CREATIS, à Lyon, et est accessible à l'adresse <https://vip.creatis.insa-lyon.fr>. L'utilisation de VIP est gratuite et ouverte à tous, mais il faut au préalable créer un compte qui soit validé par l'équipe de développement de VIP.

6.2.3.1 Outils mis à disposition

Nous avons décidé de mettre à disposition au sein de VIP notre pipeline de quantification d'une image TDM détaillé au chapitre 5 et illustré par la Figure 51. Ce pipeline traite une image TDM de patient TC au format NIfTI, et fournit 3 fichiers :

- La segmentation à 7 classes fournie par notre modèle TL_RE_7C, obtenu par transfer learning, qui a montré les meilleures performances durant la validation externe.
- L'atlas recalé sur l'image TDM d'entrée, ce qui permet de contrôler la qualité du recalage effectué.
- Un fichier au format csv qui recense le volume de chacune des lésions dans chacune des zones de l'atlas. Ce sont ces métriques qui ont fourni les meilleurs résultats de prédiction du niveau d'intensité thérapeutique.

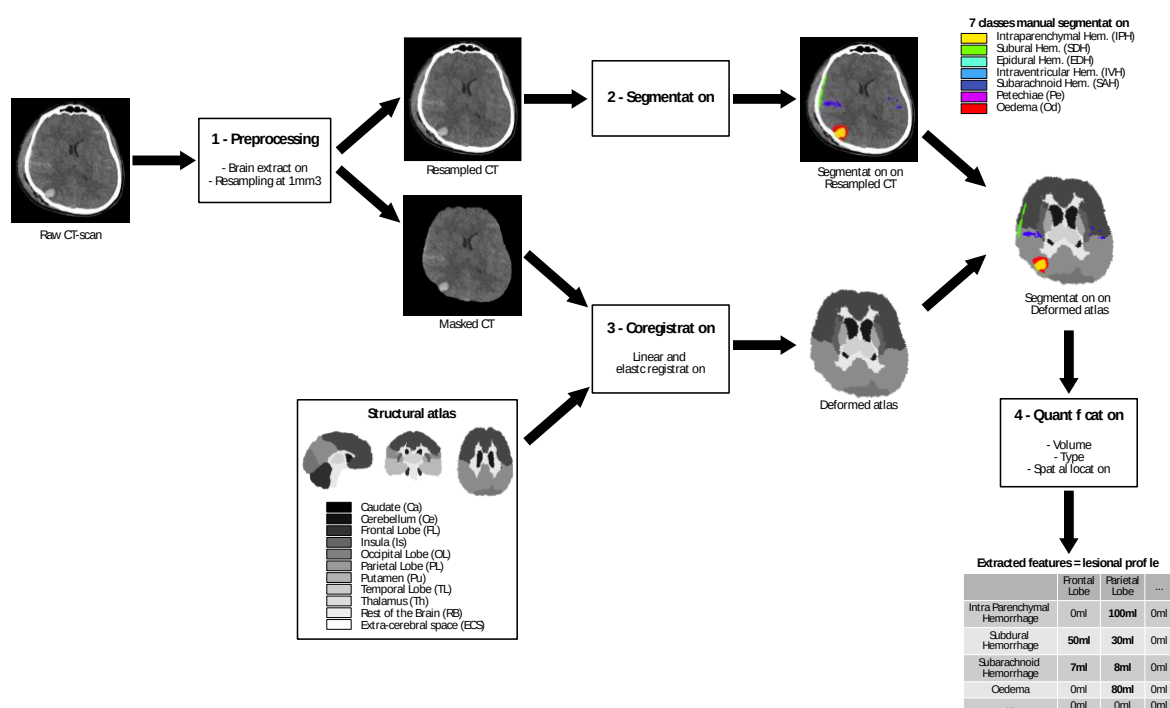


Figure 51. Pipeline de quantification mis à disposition sur la plateforme VIP

6.2.3.2 Encapsulation Docker et intégration à VIP

Le pipeline de quantification est constitué de plusieurs étapes qui ont toutes leurs particularités :

- Un rééchantillonnage des images TDM grâce à la bibliothèque python « nibabel »
- Une extraction de cerveau grâce à un algorithme codé en langage MATLAB
- Un recalage rigide de l'atlas structurel grâce à l'algorithme FLIRT du logiciel FSL codé en langage C++
- Un recalage élastique de l'atlas structurel grâce à la bibliothèque python « AntsPyx »
- Une segmentation automatique grâce au code source du dépôt GIT de BLAST-CT, codé en langage Python et qui peut être exécutée sur processeur classique (CPU) ou sur processeur graphique (GPU)
- Une extraction des métriques volumétriques et l'écriture dans un fichier CSV

Pour exécuter le pipeline de quantification, qui fait appel à toutes ces étapes, et que nous avons nommé CT-TIQUA, il faut installer les logiciels, bibliothèques, ou langages de programmation nécessaires. Ces installations doivent être faites pour chacune des machines de calculs que l'on souhaite utiliser et elles dépendent également du type de machine et du système d'exploitation de l'ordinateur.

Toutes ces contraintes sont de réelles limites au déploiement de notre outil et à sa mise à disposition aux cliniciens. Pour pallier ces limites, nous avons utilisé un « conteneur docker », qui est une sorte de petite machine virtuelle, dans laquelle on peut installer tous les outils nécessaires au code informatique que l'on veut exécuter. Ainsi, dans un conteneur docker nous avons installé FSL, les bibliothèques python nécessaires, l'outil « MATLAB Runtime », qui permet d'exécuter du code Matlab compilé sans avoir besoin de licence matlab, et le code d'exécution de notre pipeline. Ce conteneur, dans sa version la plus récente (v1.4) à l'heure où j'écris ces lignes, a été publiquement déposé en ligne et est disponible à l'adresse :

<https://hub.docker.com/repository/docker/brosscle/ct-tiqua>

La plateforme VIP exécute les applications disponibles grâce à ce qu'on appelle un descripteur Boutique [132]. Ce descripteur est un fichier json qui contient une description de l'application, des entrées et sorties du pipeline, et qui contient également, si nécessaire, le lien vers le docker à utiliser. C'est ce descripteur qui est intégré dans VIP et qui permet à la plateforme de facilement exécuter de la même manière toutes les applications disponibles.

Une fois ce descripteur intégré, il est visible sur la plateforme, comme l'atteste la capture d'écran sur la Figure 52.

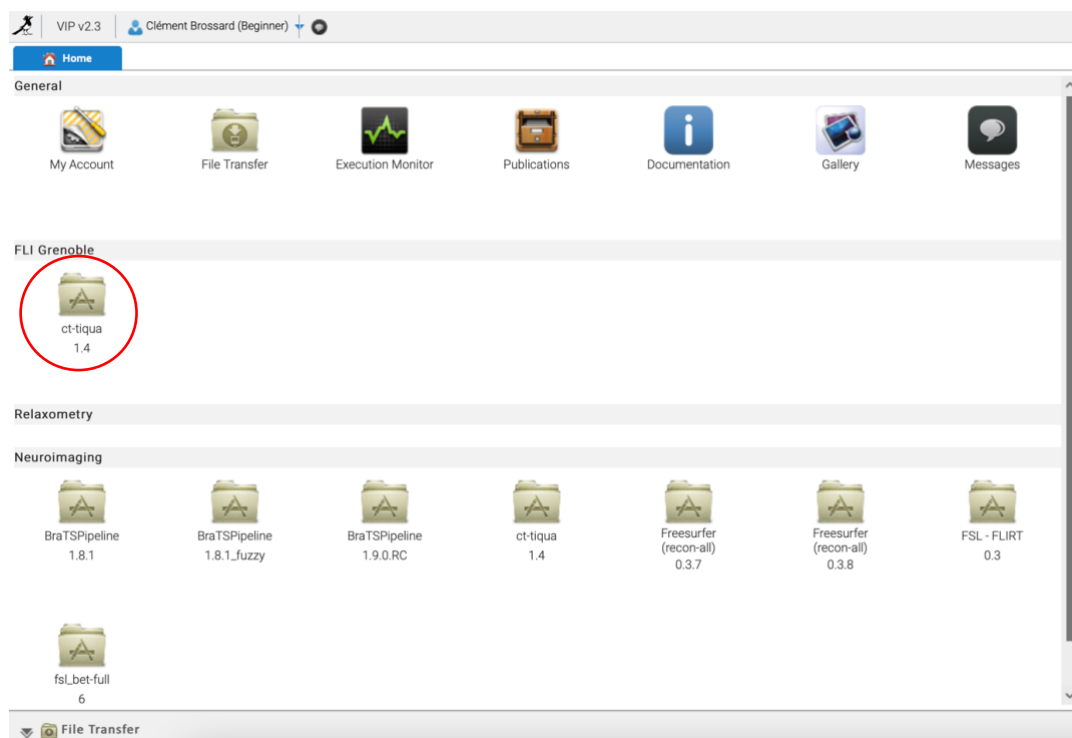


Figure 52. Capture d'écran de l'accueil de la plateforme VIP après la connexion. En rouge est entouré l'application CT-TIQUA 1.4.

La plateforme VIP dispose de ressources de stockage sur lesquelles on peut déposer les fichiers à traiter. Les fichiers de sortie du pipeline seront également stockés sur VIP et pourront ensuite être téléchargés. Une fois les fichiers à traiter déposés, on peut paramétrer le pipeline en remplissant les champs obligatoires, comme illustré sur la capture d'écran Figure 53.

VIP v2.3 | Clément Brossard (Beginner)

Home | ct-tiqua 1-4

Documentation and Terms of Use

Mandatory inputs :

Execution name*

Results directory*

Input file*
 Input image (e.g. img.nii, img.nii.gz) to be processed. It must be a brain CT-scan.

Optional inputs :

☐ **Ensemble CNN models**
 For the segmentation step: Whether to use all the 12 CNN models rather than only one (slower but more precise).

☐ **Keep temporary files**
 Do not remove temporary files at the end of the pipeline, ie. include them in the output archive.

Figure 53. Capture d'écran de la paramétrisation du pipeline CT-TIQUA

Une fois le pipeline paramétré, on peut l'exécuter en cliquant sur « Launch ». Il est possible de lancer en parallèle plusieurs exécutions. Dans la version académique de VIP, les calculs sont exécutés sur une grille de calcul à Lyon et l'équipe de développement peut également mettre à disposition des cartes graphiques pour accélérer les exécutions. Étant donné que les calculs se font à distance, et que la plateforme est accessible en ligne via une interface web, il est possible de lancer une exécution à partir d'un ordinateur peu puissant, ou même à partir d'un téléphone. Il est également possible d'éteindre son ordinateur pendant l'exécution, ce qui rend l'utilisation de CT-TIQUA compatible avec les contraintes des cliniciens.

6.2.3.3 Exécution à grande échelle

Pour tester la bonne exécution de notre pipeline sur la plateforme VIP, nous avons traité 100 images TDM de patients TC sur GPU. D'un clic nous avons commandé les 100 jobs mais elles ont été exécutées l'une après l'autre.

Si 3 jobs sur 100 ont subi une erreur car les fichiers d'entrée étaient corrompus, les 97 autres se sont déroulés sans soucis, fournissant les résultats attendus. Une fois les exécutions terminées, on a accès à quelques statistiques, parmi lesquelles le temps d'exécution de chaque job sous forme d'histogramme, comme illustré Figure 54.

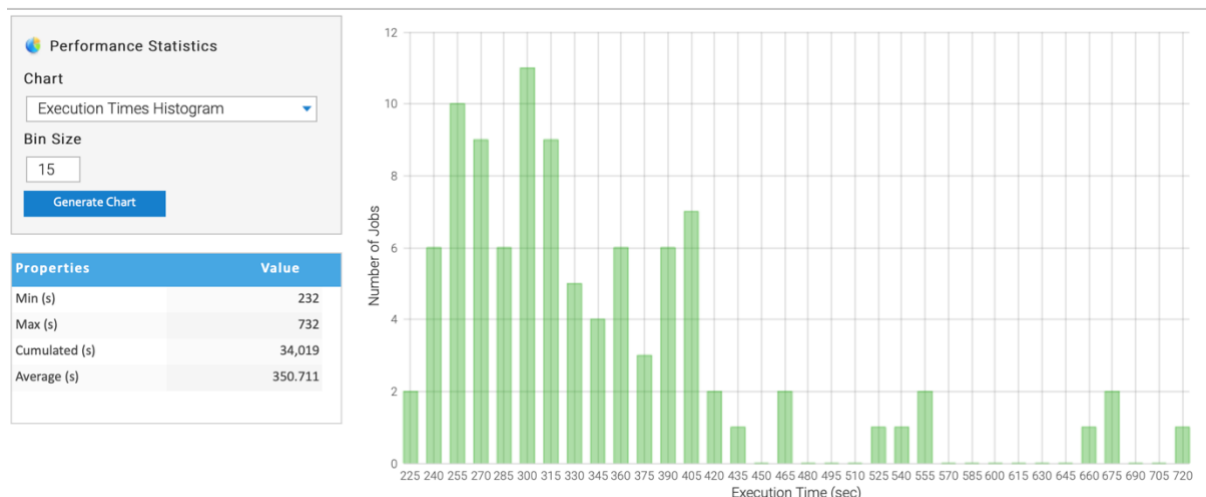


Figure 54. Histogramme du temps d'exécution de chacun des 97 jobs sur la plateforme VIP.

Le temps moyen de l'exécution d'un job est de moins de 6 minutes. Ainsi on a obtenu pour chaque image TDM une segmentation de 7 types de lésions, un atlas recalé et un rapport volumétrique en moins de 6 minutes en moyenne, ce qui rend compatible cet outil avec une pratique clinique de routine.

Il est à noter que si l'exécution avait été faite sur CPU, le temps d'exécution aurait été bien plus long, probablement de plusieurs heures par patient.

6.2.3.4 Interface avec Shanoir

La seule limite restante à l'utilisation de CT-TIQUA en routine clinique est l'anonymisation préalable et la conversion des données du format DICOM au format NIfTI, qui doit être faite en dehors de la plateforme VIP.

Pour résoudre cette ultime barrière, une interface va être développée entre VIP et Shanoir, plateforme de stockage sécurisée de données d'imagerie médicale [107]. Ainsi, les images du PACS du CHU pourront être stockées sur Shanoir, qui se chargera d'anonymiser et de convertir les images du format NIfTI, et qui proposera automatiquement l'exécution sur VIP des différents pipelines de traitement. Cette dernière étape a été entamé lors de ma dernière année de thèse, mais en raison des délais de contractualisation des contrats de collaboration entre le CHU, l'INSERM, l'université Grenoble Alpes et de Shanoir, elle n'est pas disponible à l'heure actuelle.

6.3 Conclusion

Ce chapitre a tout d'abord présenté une validation externe de nos modèles de segmentation et de classification sur une 2eme cohorte, ce qui a permis de quantifier la

généralisation de nos algorithmes à des données de nature différentes. Puis nous avons détaillé la solution technique qui nous a permis de mettre à disposition notre pipeline complexe de quantification, qui fait appel, entre autres, à notre modèle de segmentation par réseau de neurone. Les cliniciens du CHUGA vont désormais pouvoir facilement utiliser notre pipeline de quantification du profil lésionnel d'un patient TC sur des données de soin courant et nous guider pour son amélioration et pour les futurs développements de l'équipe. Maintenant maîtrisée au laboratoire, la procédure de mise à disposition d'un pipeline complexe sur la plateforme VIP pourra ensuite être rapidement répétée pour tout nouvel outil développé dans le laboratoire (en lien ou non avec des images TDM et/ou des patients TC).

Chapitre 7 : Étude de l'évolution lésionnelle afin de prédire l'expansion œdémateuse

Dans le cadre de la prise en charge du TC, les différents traitements et interventions ont pour principal objectif de limiter la création et la progression d'œdème cérébraux. En effet, les hémorragies initiales du TC peuvent conduire, selon le cercle vicieux décrit au chapitre 1, à une poussée œdémateuse cérébrale. Ainsi au cours du temps l'œdème peut se développer chez les patients et prend une place importante dans l'ensemble des lésions visibles sur les images TDM. Ce chapitre est consacré aux premiers développements visant à prédire l'apparition d'un œdème à partir d'images TDM. Ce chapitre présentera deux approches différentes visant à étudier cette problématique, une basée sur une analyse Radiomique et la seconde sur une analyse empirique brièvement détaillée ci-dessous :

- L'analyse Radiomique vise à discriminer les tissus de même état lésionnel à J0 mais d'un état lésionnel différent à J1, grâce aux métriques radiomiques extraites de ces régions d'intérêt. Concrètement, nous avons utilisé un classifieur pour classer ces régions d'intérêt de même état lésionnel à J0 en 3 catégories selon si elles stagnent, se transforment en œdème, ou bien redevenir du tissu sain à J1, dans le but de déterminer si ces évolutions sont contenues dans les images TDM dès J0.
- L'approche empirique consiste à étudier, à l'échelle du voxel, la fréquence d'évolution en œdème d'une lésion en fonction de son état à J0 et J1 résultant de la segmentation. En fonction de ces observations, nous avons généré ce qu'on a appelée des segmentation empirique représentant les régions œdémateuses à J3.

7.1 Prédiction du futur œdème : Analyse Radiomique

Notre hypothèse est qu'il existe des différences biologiques au sein d'un même tissu lésionnel qui conduisent, 24 heures plus tard, à des tissus différents : stagnation lésionnelle, transformation œdémateuse, ou transformation en tissu sain, et que ces

différences biologiques induiraient des modifications de la densité tissulaire observables et quantifiables par imagerie TDM. Par définition, les métriques radiomiques permettent de quantifier les propriétés d'une image, à travers des métriques calculées à différentes échelles spatiales et à différents moments d'ordre statistique.

L'objectif de cette étude est donc de déterminer s'il était possible d'exploiter ces métriques Radiomiques extraites des images TDM acquise à J0 pour prédire les zones lésionnelles qui se transformeront en œdème de celles qui se transformeront en tissu sain et de celles qui stagneront à J1.

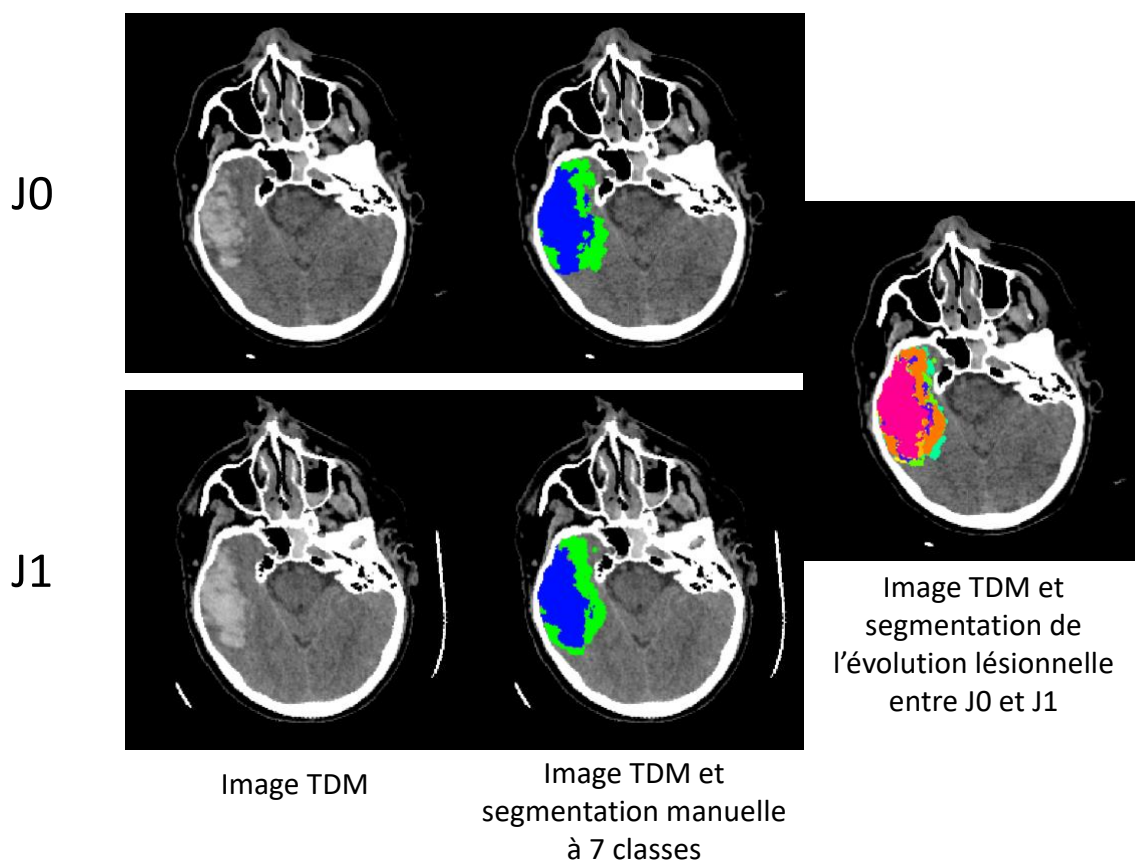
7.1.1 Données utilisées

Pour cette étude nous avons utilisé la cohorte Radiomic-TBI, car c'est la seule qui dispose d'un suivi scanographique longitudinal des patients. Pour rappel, elle est constituée de 84 images TDM de 29 patients. Chaque patient est imagé à l'arrivée à l'hôpital (J0), le lendemain (J1), et 2 jours plus tard (J3). Pour cette étude, nous avons uniquement utilisé les images acquises à J0 et à J1, soit 58 images TDM. Nous avons utilisé comme vérité lésionnelle les segmentations manuelles à 7 classes déjà utilisées aux chapitres 3 et 4. Pour chaque patient, les images acquises à J1 ont été recalées de manière rigide (algorithme FLIRT du logiciel FSL) sur les TDM de J0.

7.1.2 Méthodes

7.1.2.1 Quantification de l'évolution lésionnelle

Grâce à la segmentation manuelle à 7 classes, et puisque les images d'un même patient sont recalées entre elles, chaque voxel appartient à une classe lésionnelle (au sens large, en comprenant l'absence de lésion) à J0 et à J1. Si cette classe n'est pas la même, le voxel lésionnel s'est transformé. A l'inverse, si la classe est la même entre J0 et J1, le voxel a stagné. On peut ainsi regrouper tous les voxels qui partagent la même évolution et définir une nouvelle segmentation qui délimite les zones de même évolution lésionnelle entre J0 et J1. Par exemple, tous les voxels étant classés comme de l'HIP à J0 et comme de l'œdème à J1 seront regroupés dans une catégorie, et ainsi de suite avec les autres évolutions. Puisqu'il y a 8 classes différentes possibles (7 lésions et la classe « absence de lésion ») pour chacun des temps, la nouvelle segmentation, que l'on appellera « segmentation de l'évolution lésionnelle entre J0 et J1 », peut contenir entre 1 et 64 classes différentes. Un exemple de ce type de segmentation est disponible sur la Figure 55.



Légende de la segmentation manuelle à 7 classes : Bleu – HIP, Vert – Œdème.

Légende des évolutions : Rose – HIP/HIP, Vert – Tissu Sain/Œdème, Orange foncé – Œdème/Œdème, Orange clair – Œdème/HIP, Jaune – HIP/Tissu Sain, Violet – HIP/Œdème, Cyan – Œdème/Tissu Sain

Figure 55. Illustration de la segmentation à J0 et J1 d'un patient de la cohorte Radiomic-TBI et de la segmentation de son évolution entre J0 et J1.

7.1.2.2 Définition des régions d'intérêt et extraction des métriques Radiomiques

Puisque notre but est de prédire la transformation de lésions déjà existantes en œdème, nous nous sommes concentrés, pour chaque type de lésion et pour chaque patient, sur 3 différentes zones :

- Les régions lésionnelles à J0 qui deviennent saines à J1
- Les régions lésionnelles à J0 qui ne se transforment pas à J1
- Les régions lésionnelles à J0 qui deviennent œdème à J1

Ainsi puisque l'on dispose de 7 lésions, on obtient 21 régions d'intérêt. En réalité, on a seulement 20 ROI différentes car les régions œdémateuses à J0 qui stagnent à J1 et celles qui se transforment en œdème à J1 sont les mêmes. Sur chacune de ces 20 zones, nous avons extrait des métriques Radiomiques sur les images TDM à J0, qui représentent les

propriétés statistiques des voxels contenus dans ces zones. Notons que si sur une image existe plusieurs zones de même évolution qui ne sont pas connectées, nous extrayons les métriques radiomiques sur l'union des zones de même évolution, ce qui revient à moyenner les métriques qui seraient calculées sur chacune des zones.

Les métriques Radiomiques ont été extraites grâce à la bibliothèque Python « pyradiomics » [133], dont la plupart des métriques proposées respectent la convention Image Biomarker Standardisation Initiative (IBSI), qui vise à harmoniser le calcul des métriques radiomiques extraites par les différents outils disponibles.

Parmi les métriques disponibles, nous avons extrait à partir de l'image TDM les métriques relatives aux groupes :

- First order statistics (18 métriques)
- Gray Level Cooccurrence Matrix – GLCM (24 métriques)
- Gray Level Run Length Matrix – GLRLM (16 métriques)
- Gray Level Size Zone Matrix – GLSZM (16 métriques)
- Neighbouring Gray Tone Difference Matrix – NGTDM (5 métriques)
- Gray Level Dependence Matrix – GLDM (14 métriques)

Le seul groupe de métriques que nous n'avons pas utilisé ont celles relatives à la forme de la zone considérée, car puisque les zones sont issues de la segmentation manuelle, la forme des zones a été manuellement créé, et nous souhaitons plutôt analyser la capacité des propriétés intrinsèques de l'image afin de prédire la transformation œdémateuse des lésions. Les métriques relatives au groupe First Order Statistics, qui se basent en partie sur l'histogramme de la distribution des voxels dans la zone étudiée ont été extraites après une translation de 1000 UH de l'histogramme afin d'éviter les valeurs négatives des images TDM. La taille des compartiments des histogrammes (« bins » en anglais), a été fixée à 10 UH.

7.1.2.3 Classification

Dans le but de prédire l'appartenance d'une zone à l'une des 3 classes d'évolution différente grâce aux métriques radiomiques, nous avons tout simplement mis en place une procédure de classification similaire à celle du chapitre 5, c'est-à-dire un pipeline de classification et une validation croisée imbriquée sur l'ensemble de la cohorte Radiomic-TBI.

A l'exception de sous échantillonnage, que nous avons supprimé cette fois car nous disposons de plus de données, le pipeline de classification est le même qu'au chapitre 5 :

- Une normalisation des métriques entre 0 et 1
- Sélection des 4 métriques les plus liées au critère à prédire
- Un classifieur Forêt Aléatoire avec les 90 mêmes jeux d'hyperparamètres qu'au chapitre 5.

La procédure de validation croisée imbriquée a une fois encore été effectuée grâce à la bibliothèque python PhotonAI [125]. La boucle interne a cette fois consisté en une validation croisée à 3 ensembles (3-Fold cross validation) répétée 5 fois et la boucle externe a quant à elle consisté en une validation croisée à 4 ensembles (4-Fold cross validation) répétée 5 fois. Puisque l'AUC est définie pour une classification binaire et non multiclasse comme c'est le cas ici, nous avons utilisé l'exactitude de prédiction pour sélectionner et évaluer nos modèles de classification.

Nous avons ainsi pu comparer la capacité des métriques radiomiques à prédire l'évolution d'une lésion en fonction du type de lésion. Nous avons également évalué l'importance des métriques dans la prédiction des meilleurs modèles de classification.

7.1.3 Résultats

Le Tableau 32 ci-dessous récapitule les volumes moyens des 3 zones à classifier pour les 7 différentes lésions. La tâche de classification à 3 classes devient une classification à 2 classes pour l'œdème.

Volume lésionnel moyen	Qui se transforme en tissu sain à J1	Qui stagne à J1	Qui se transforme en œdème à J1
d'HIP à J0 (mm3)	1640 +-1577 n=20	6731 +-7859 n=20	1119 +-1716 n=18
d'HSD à J0 (mm3)	24627 +-34515 n=21	8168 +-10108 n=20	868 +-1125 n=11
d'HED à J0 (mm3)	12009 +-19790 n=8	9851 +-11018 n=6	40 +-3 n=2
d'HIV à J0 (mm3)	439 +-444 n=10	291 +-3225 n=8	26 +-20 n=2
d'HSA à J0 (mm3)	3372 +-4897 n=23	1356 +-1553 n=18	246 +-325 n=8
de pétéchies à J0 (mm3)	133 +-156 n=13	174 +-218 n=9	106 +-91 n=3
d'œdème à J0 (mm3)	6680 +-5765 n=20	7173 +-6347 n=17	

Tableau 32. Volume moyen de chacun des types de lésion à J0 selon son évolution à J1 +- écart type (Nombre de patient où la zone est présente). Il faut lire le tableau comme : « Il y a en moyenne 1640 mm3 d'HIP à J0 qui se transforme en tissu sain à J1 par patient ».

Les résultats de la classification sont présentés sur le Tableau 33. Les scores d'exactitude sont comparés au pourcentage de la classe majoritaire dans le jeu de données, ce qui

permet de relativiser les scores d'exactitude en fonction de la distribution des échantillons.

Lésion	Exactitude de prédiction	Pourcentage de la classe majoritaire dans le jeu de données	Métriques les plus importantes
HIP à J0	74% +- 14%	34%	GLRLM_LongRunEmphasis (28%) GLSZM_SmallAreaEmphasis (24%) GLRLM_RunVariance (24%) GLSZM_SizeZoneNonUniformityNormalized (23%)
HSD à J0	64% +- 10%	40%	GLSZM_SizeZoneNonUniformityNormalized (32%) GLSZM_GrayLevelNonUniformityNormalized (30%) Firstorder_Entropy (24%)
HED à J0	Pas assez d'images avec transformation en Œdème à J1 (n=2)		
HIV à J0	Pas assez d'images avec transformation en Œdème à J1 (n=2)		
HSA à J0	48% +- 13%	47%	GLDM_DependenceVariance (31%) GLSZM_SmallAreaEmphasis (24%) GLRLM_RunEntropy (24%) GLRLM_RunPercentage (21%)
Pétéchie s à J0	Pas assez d'images avec transformation en Œdème à J1 (n=3)		
Œdème à J0	59% +- 12%	54%	Firstorder_RootMeanSquared (33%) Firstorder_Range (26%) Firstorder_Maximum (26%)

Tableau 33. 2ème colonne – Exactitude de classification des modèles entraînés pour prédire l'évolution à J1 de zones lésionnelles à J0. 3ème colonne – Exactitude d'un classifieur prédisant toujours la classe majoritaire au sein du jeu de données. 4ème colonne : Nom des métriques dont l'importance dépasse les 20% dans la classification.

Pour 3 types de lésion (HED, HIV, Pétéchies), le nombre d'images comprenant une zone se transformant en œdème est inférieur à 4, c'est-à-dire inférieur au nombre d'ensembles de la boucle externe de la validation croisée imbriquée, ce qui rend la classification impossible. Nous nous sommes donc concentrés sur les 4 autres lésions.

En ce qui concerne l'HSA et l'œdème, les scores d'exactitude sont très proches du pourcentage de la classe majoritaire, ce qui laisse penser que les résultats de la prédiction sont proches du hasard.

Sur l'HSD, le score d'exactitude est de 64%, ce qui est supérieur de 24 points au pourcentage de la classe majoritaire. Les 2 principales métriques responsable de cette prédiction sont issues du groupe GLSZM.

Finalement, le meilleur score est obtenue sur l'HIP, où on atteint 74% (+-14%) d'exactitude comparé à un pourcentage de la classe majoritaire de 34, soit plus du double. Les métriques les plus influentes dans cette prédiction sont issues des groupes GLSZM et GLRLM.

7.1.4 Interprétation et perspectives

S'il est apparemment impossible de différencier une zone d'œdème à J0 qui reste œdème à J1 d'une zone qui se transforme en tissu sain grâce aux métriques radiomiques, nos résultats suggèrent qu'il serait possible de déterminer le futur de l'HIP. Effectivement, il semble que certaines métriques radiomiques sont capables de différencier une zone d'HIP à J0 qui va, à J1, rester de l'HIP, se transformer en œdème, ou bien se transformer en tissus sain. On peut voir ce résultat d'une autre manière : nos résultats montrent qu'une partie de l'HIP qui va se transformer sous 24 heures en œdème est un tissu au profil tomодensitométrique différent de celle qui va rester de l'HIP et de celle qui va se transformer en tissu sain. S'il n'est pas possible de le remarquer à l'œil nu, cette différence de tissu induit une différence des propriétés statistiques de l'image, que l'on peut quantifier grâce aux métriques radiomiques, comme illustré sur la Figure 56.

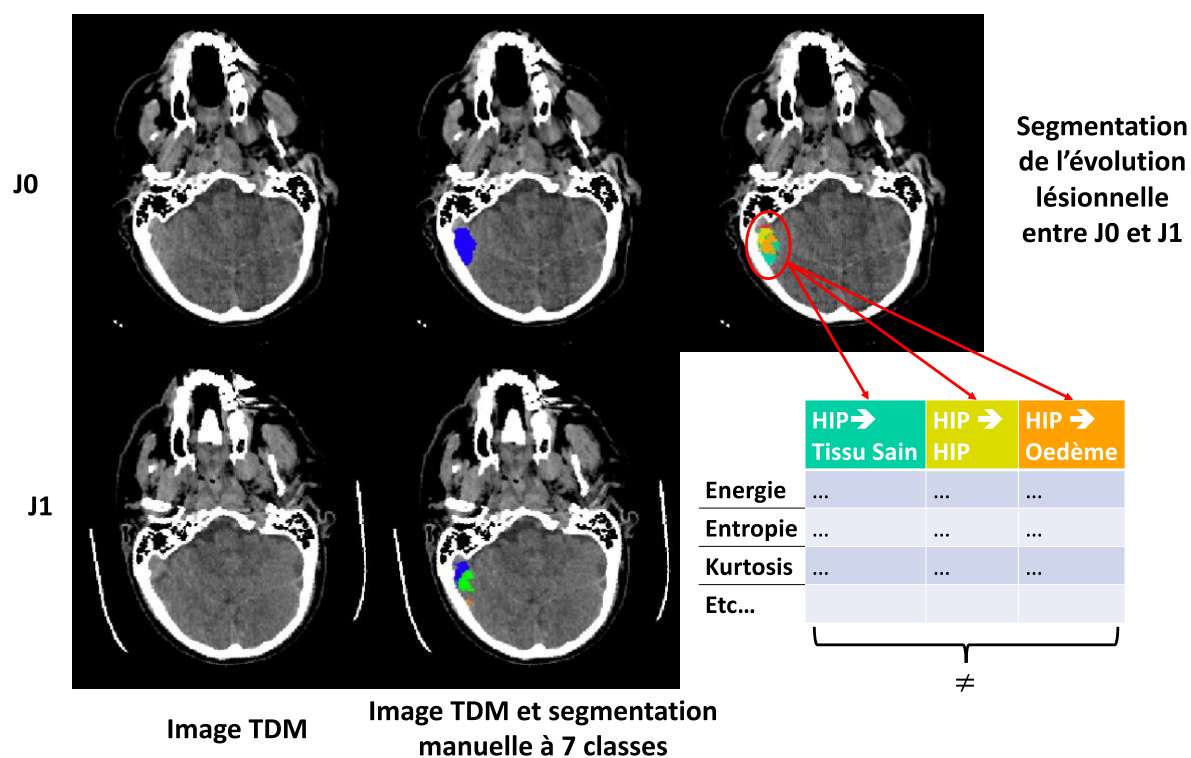


Figure 56. Illustration de la capacité des métriques radiomiques à discriminer les hémorragies intra-parenchymateuses à J0 qui se transforment en tissu sain, qui stagnent, ou qui se transforment en œdème à J1.

Les métriques radiomiques qui permettent de mettre en évidence cette différence sont issues des groupes GLSZM et GLRLM, qui représentent tous les deux la taille des régions de même niveau de gris au sein de la région d'intérêt sur laquelle sont calculées les métriques. Ainsi, nos résultats suggèrent qu'il existe un biomarqueur d'imagerie capable de détecter de manière précoces les transformations œdémateuses de l'HIP.

Cette étude a été menée à partir des régions d'intérêt d'évolutions différents. Notre outil de classification ne permet donc pas de déterminer à priori les zones de l'HIP à J0 qui se transformeront en œdème à J1. On pourrait pour cela calculer les métriques radiomiques les plus importantes sur des petits patchs parcourant l'image TDM acquise à J0, afin d'obtenir plusieurs nouvelles images représentant chacune des métriques radiomiques précédentes sur tout le cerveau (comme illustré dans Muschelli & al. [95]). Une fois ces images obtenues, on pourrait appliquer un algorithme d'apprentissage non supervisé comme un clustering afin de regrouper les voxels de l'HIP en fonction de leur profil radiomique, c'est-à-dire en fonction de la valeur de ce voxel sur chacune des images radiomiques. Le développement et la validation d'un tel outil pourrait ainsi aboutir à une segmentation automatique de l'HIP qui va se transformer sous 24h en œdème, ce qui faciliterait grandement les cliniciens dans la prise en charge et l'estimation de l'évolution de l'état des patients. Ce travail n'est pas encore entamé mais nous disposons de toutes les données et tous les outils nécessaires pour mener à bien ce projet.

7.2 Prédiction empirique de la segmentation lésionnelle à J3

L'étude précédente a permis de différencier les zones d'HIP qui pourraient, sous 24h, se transformer en œdème, en tissu sain, ou qui vont rester de l'HIP. Pourtant sur la Figure 57, qui représente le diagramme de Sankey des évolutions lésionnelles entre J0, J1 et J3 des 26 patients de la cohorte Radiomic-TBI dont nous disposons des 3 images, on remarque que le volume d'HIP à J0 qui se transforme en œdème à J1 ne compte que comme une infime partie du volume total d'œdème à J1 (courbe en rouge Figure 57). Ainsi, la majeure partie de l'œdème n'est pas prise en compte. Pourtant, nous pensons que la transformation des lésions entre J0 et J1 contient une information précoce sur l'évolution lésionnelle globale à J3. Cette partie a pour but de développer une approche différente de celle utilisant les métriques radiomiques, issue des statistiques fréquentistes, qui vise également à prédire l'apparition d'un œdème cérébral à partir de l'état lésionnel aux temps précédents. Plus précisément, nous allons présenter une approche empirique dans laquelle nous allons calculer la segmentation de l'évolution lésionnelle entre J0, J1, et J3 des lésions et de leur voisinage, et utiliser cette information pour dégager des tendances, des évolutions qui se répètent sur une grande partie des patients, et qu'on peut interpréter comme des « règles », résultat des interactions lésionnelles au sein du

cerveau. Ces règles pourront ensuite être appliqués à de nouveaux patients afin de prédire leur profil lésionnel à J3 à partir des segmentations à J0 et J1. Une telle prédiction pourrait grandement améliorer la prise en charge des patients, car elle permettrait d'intervenir de manière plus précoce.

Evolution lésionnelle entre J0 (gauche), J1 (milieu), et J3 (droite) de la cohorte Radiomic-TBI

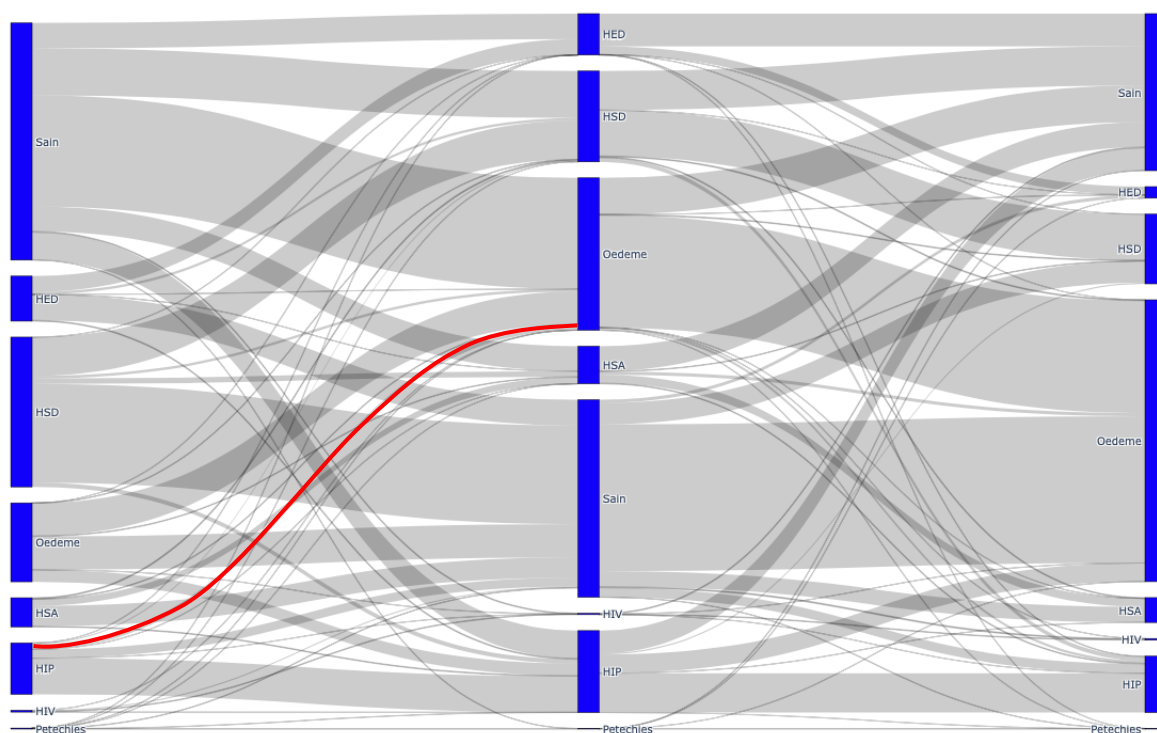


Figure 57. Diagramme de Sankey représentant l'évolution lésionnelle entre J0 (à gauche), J1 (au milieu), et J3 (à droite) des 26 patients de la cohorte Radiomic-TBI dont nous disposons des 3 images. La courbe en rouge représente le volume d'HIP à J0 qui évolue en œdème à J1.

7.2.1 Données utilisées

Pour cette étude nous avons encore une fois utilisé la cohorte Radiomic-TBI, car c'est la seule qui dispose d'un suivi scanographique longitudinal des patients. Pour rappel, elle est constituée de 84 images TDM de 29 patients. Chaque patient est imagé à l'arrivée à l'hôpital (J0), le lendemain (J1), et 2 jours plus tard (J3). Contrairement à l'analyse Radiomique, nous avons ici utilisé les images acquises à J0, à J1, et à J3. Trois des 29 patients n'ont pas eu d'image TDM à J3, principalement pour des raisons d'accessibilité du scanner. Ces 3 patients ont donc été exclus et nous avons étudié les 26 patients restants et les 78 images associées. Là encore, nous avons utilisé comme vérité lésionnelle les segmentations manuelles à 7 classes déjà utilisées aux chapitres 3 et 4. Pour rappel, toutes les images d'un patient ont été recallées de manière rigide sur l'image TDM à J0 grâce à l'algorithme FLIRT du logiciel FSL.

7.2.2 Méthode

7.2.2.1 Division de la cohorte Radiomic-TBI en jeu de développement et jeu de test.

Pour cette étude, qui est toujours en cours, nous avons séparé la cohorte Radiomic-TBI en un jeu de développement et un jeu de test. La répartition est la même que lors de l'étude sur la prédiction du TILsum détaillée au chapitre 5, à l'exception de l'exclusion des 3 patients qui ne disposent pas d'une image TDM à J3 (P03, P04, P20) :

- Jeu de développement (21 patients) : P01, P05, P06, P12, P13, P14, P15, P17, P18, P19, P22, P24, P25, P26, P27, P30, P07, P09, P10, P23, P28
- Jeu de test (5 patients) : P02, P08, P11, P16, P29

7.2.2.2 Rajout de classes à la segmentation : Voisinages

Puisque les images TDM de chaque patient ont déjà été recalées sur l'image acquise à J0, la segmentation manuelle à 7 classes permet d'attribuer à chaque voxel du cerveau une évolution entre J0, J1 et J3. Par exemple, un voxel peut être HIP à J0, rester HIP à J1, et évoluer en œdème à J3. Puisque nous pensons que les différentes lésions influent les unes sur les autres et que les lésions ont tendance à s'étendre plutôt qu'à apparaître spontanément, nous avons ajouté, aux segmentations manuelles de 7 types de lésions, des régions représentant le voisinage immédiat de ces lésions.

Ainsi nous avons appliqué une opération morphologique de dilatation sur les segmentations de chaque type de lésion afin de délimiter leur voisinage immédiat. La dilatation a été effectuée avec un élément structuel de 3 dimensions et de connectivité maximum, et a été répétée 2 fois, de manière à « entourer » les lésions avec une profondeur de quelques millimètres. Puisque les lésions sont parfois proches les unes des autres, il arrive que le voisinage de deux lésions se chevauche. Dans tous ces cas, nous créons de nouvelles classes de la segmentation qui délimitent ces zones appartenant au voisinage de deux ou de plusieurs lésions. Au final, à partir d'une segmentation à 7 classes, nous avons délimité 7 nouvelles classes représentant le voisinage de chacun de ces types de lésions, ainsi que 120 combinaisons de ces voisinages. Cette nouvelle segmentation, comprenant donc 135 différentes classes (en comptant l'absence de lésion ou de voisinage), est nommée « segmentation manuelle à 7 classes étendue ». Finalement, nous avons masqué cette segmentation par la segmentation du cerveau obtenue précédemment lors de l'étude de classification détaillée au chapitre 5, afin de ne considérer que les voxels à l'intérieur du parenchyme. Une illustration de cette segmentation manuelle à 7 classes étendue est disponible Figure 58.

Nous avons calculé ces nouvelles segmentations pour les 26 patients de notre cohorte sur les images acquises à J0 et à J1. Quant à la segmentation à J3, nous avons uniquement considéré la segmentation manuelle initiale, sans s'intéresser aux voisinages des lésions.

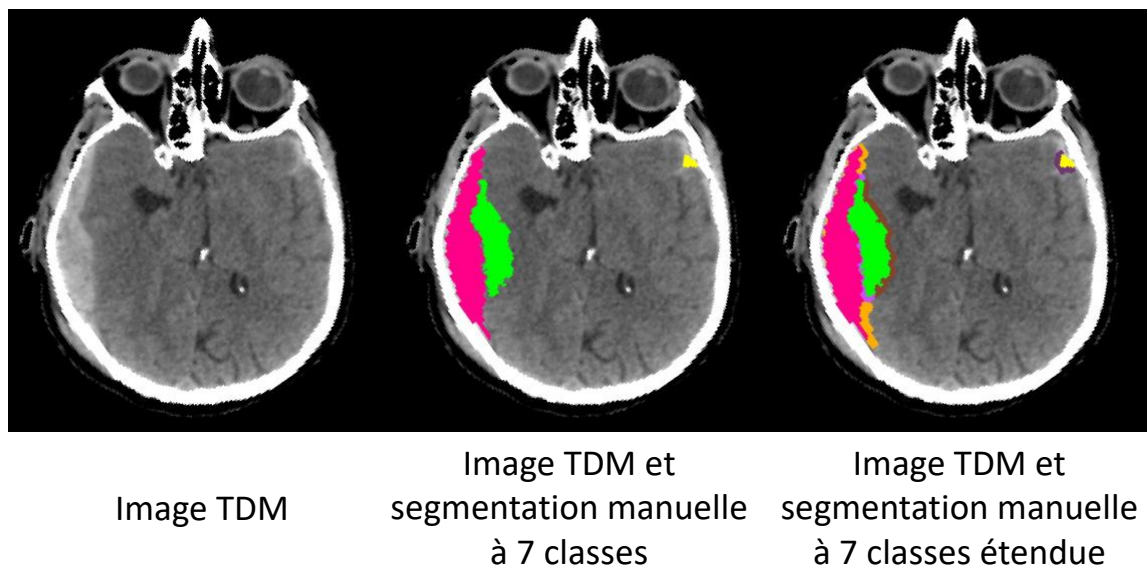


Figure 58. Comparaison entre la segmentation manuelle à 7 classes et sa version étendue. Légende couleur : Rose foncé – HED, Vert – Œdème, Jaune – HSA, Orange – Voisinage de l'HED, Violet clair – Voisinage de l'HED et de l'œdème, Marron – Voisinage de l'œdème, Violet foncé – Voisinage de l'HSA.

7.2.2.3 Calcul de la segmentation d'évolution entre J0, J1 et J3.

A partir de ces segmentations étendues à J0 et à J1, nous avons pu calculer la segmentation d'évolution entre J0, J1, et J3 de la même manière qu'exposé au début de ce chapitre et illustré Figure 55. Cette fois nous avons 135 différentes classes à J0, 135 à J1, et 8 classes à J3, soit 145 800 possibilités d'évolution différentes, donc 145 800 classes différentes au sein de la segmentation d'évolution entre J0, J1, et J3. Bien évidemment, il est très peu probable d'un point de vue physiologique que ces 145 800 évolutions existent au sein de notre cohorte. Par exemple, qu'une hémorragie intra ventriculaire se transforme en hémorragie sous durelle semble impossible. L'immense majorité de ces classes seront donc très peu ou pas représentées.

7.2.2.4 Création du dictionnaire et de la fréquence d'apparition

Pour chaque patient de notre cohorte, on dispose donc d'une segmentation à 145 800 classes, représentant les 145 800 évolutions possibles (exemple : (J0) voisinage d'un œdème → (J1) œdème → (J3) tissu sain) pour chacun des voxels. Nous avons ensuite ajouté le nombre de voxels partageant la même évolution pour chaque patient du jeu de développement. Cette somme nous a permis d'obtenir le nombre de voxels suivant chacune des 145 800 évolutions recensées à l'échelle du jeu de développement. Afin de

limiter le bruit introduit lors de la segmentation manuelle ou bien lors du recalage, nous avons décidé de conserver uniquement les évolutions partagées par plus de 50 voxels pour chaque patient (soit 50 mm³ au total dans le cerveau du patient).

L'ensemble des voxels suivant une évolution entre J0 et J1 se répartit donc en 8 états différents à J3 : l'une des 7 lésions ou bien le tissu sain. Grâce au nombre de voxels dans chacun des 8 états à J3, on peut calculer les fréquences d'apparition de ces états sachant l'évolution entre J0 et J1. On peut donc assimiler ces fréquences aux probabilités pour un voxel de se transformer dans l'un des 8 états à J3 en fonction de son évolution entre J0 et J1. On peut également recenser pour chacune des 145 800 évolutions le nombre de patients qui présentent cette évolution. Pour chaque évolution, les deux principaux chiffres qui la caractérisent sont donc : la probabilité d'apparition et le nombre de patients qui présentent cette évolution. La liste des évolutions associées à ces deux chiffres est conservée dans ce que l'on appelle un dictionnaire des évolutions.

7.2.2.5 Élaboration des « règles d'évolution »

Afin d'éliminer le bruit mais aussi les évolutions entre J0 et J1 pour lequel aucun état à J3 n'est largement plus probable que les autres, nous avons classé les évolutions en fonction de seuils sur les chiffres de probabilité d'apparition et de nombre de patients qui présentent l'évolution.

Ainsi nous avons défini 4 types d'évolutions de façon arbitraire :

- Les évolutions « quasi-certaines » : Celles qui présentent une probabilité d'apparition supérieure à 90% et qui sont partagées par plus de 10 patients
- Les évolution « très probables » : Celles qui présentent une probabilité d'apparition supérieure à 80% et qui sont partagées par plus de 8 patients
- Les évolutions « probables » : Celles qui présentent une probabilité d'apparition supérieure à 70% et qui sont partagées par plus de 5 patients
- Les évolutions « contingentes » : Toutes les autres évolutions, c'est-à-dire celles dont la probabilité est inférieure ou égale à 70%, ou bien qui ne sont partagées que par 5 patients ou moins.

On peut lire ces évolutions comme des « règles » qui régissent les processus biologiques et les interactions entre les lésions. On peut donc définir un ensemble de règles « quasi-certaines », « très probables », « probables », et « contingentes » en fonction de leur probabilité d'apparition.

7.2.2.6 Évaluation sur le jeu de test

A partir de l'évolution d'un voxel entre J0 et J1, on peut exploiter les fréquences d'apparition afin de prédire l'état de ce voxel à J3, en lui attribuant l'état de plus grande probabilité en fonction des évolutions stockées dans le dictionnaire crée à partir des évolutions des patients du jeu de développement. Le fait d'exploiter ces fréquences pour obtenir une segmentation à J3 est ce qu'on a appelé « segmenter empiriquement » les lésions à partir de l'évolution lésionnelle à J0 et J1. Ainsi, dans le but d'évaluer cette méthode, nous avons segmenté empiriquement les images TDM du jeu de test préalablement mis de côté, afin d'obtenir la segmentation empirique à J3 sur ces images.

Afin d'évaluer la segmentation empirique, nous avons calculé, sur les 5 patients du jeu de test, le score de Dice entre cette segmentation et la segmentation manuelle à J3, pour le tissu lésionnel global et pour les 7 lésions prises de manière individuelle.

7.2.3 Résultats et interprétation

7.2.3.1 Définition des règles empiriques

Après la génération des segmentations du voisinage des lésions et des combinaisons de voisinages à partir des segmentations manuelles à J0 et J1, nous avons pu extraire les évolutions longitudinales des états lésionnels des voxels des patients du jeu de développement de la cohorte Radiomic-TBI. Nous avons donc pu calculer les fréquences d'apparition des différentes évolutions et recenser les différentes évolutions qui respectent nos critères sur la fréquence d'apparition et le nombre de patients concernés par l'évolution. Les règles quasi-certaines, très probables, et probables sont détaillées sur le Tableau 34 ci-dessous :

Numéro de règle	État lésionnel à J0	État lésionnel à J1	État lésionnel à J3	Probabilité d'apparition	Nombre de patients concernés	Volume partageant l'évolution (mm3)
Règles quasi-certaines						
1	Tissu sain	Tissu sain	Tissu sain	99,93	21	254757432
2	Voisinage d'HSD	Tissu sain	Tissu sain	97,22	16	222772
3	Voisinage d'HSA	Voisinage d'HSA	Tissu sain	93,30	12	44815

4	Voisinage d'HSD	Voisinage d'HSD	Tissu sain	93,01	11	67560
5	Tissu sain	Voisinage d'HSD	Tissu sain	90,20	15	171253
6	Tissu sain	Voisinage d'HSA	Tissu sain	90,11	15	132119
Règles très probables						
7	Voisinage d'HIP	Tissu sain	Tissu sain	94,25	9	5553
8	HSD	Tissu sain	Tissu sain	89,50	16	180757
9	Voisinage d'HSA	Tissu sain	Tissu sain	88,58	15	109633
10	HSD	Voisinage d'HSD	Tissu sain	85,69	11	27668
11	Voisinage d'œdème	Tissu sain	Tissu sain	85,45	12	60749
12	Voisinage d'HIP et d'œdème	Tissu sain	Tissu sain	84,93	11	4829
13	Voisinage d'HIP	Voisinage d'HIP	Tissu sain	83,70	10	4041
14	HIP	Tissu sain	Tissu sain	83,17	11	4442
15	Tissu sain	HSA	Tissu sain	81,37	16	44285
16	Œdème	Tissu sain	Tissu sain	80,22	14	27009
Règles probables						
17	Voisinage d'HIV	Tissu sain	Tissu sain	99,35	7	27688
18	Voisinage d'HSD	Voisinage d'HSA	Tissu sain	95,91	8	16835
19	HED	Tissu sain	Tissu sain	93,47	7	66927
20	Voisinage d'HSD	HSA	Tissu sain	92,43	6	5371
21	Voisinage d'HSA	Voisinage d'HSD	Tissu sain	89,36	6	5140

22	Voisinage d'HSD	Voisinage d'HIP	Tissu sain	82,74	6	1074
23	Voisinage d'HSD et d'HSA	Voisinage d'HSA	Tissu sain	81,08	7	2537
24	Voisinage d'HSA	HSA	Tissu sain	79,82	12	8578
25	Voisinage d'HSD et d'œdème	Tissu sain	Tissu sain	79,18	6	1993
26	HSA	Tissu sain	Tissu sain	78,67	17	36832
27	HSA	Voisinage d'HSA	Tissu sain	78,41	11	9619
28	Tissu sain	Voisinage d'HIP	Tissu sain	78,35	16	29694
29	HSD	Voisinage d'HIP et d'HSD	Tissu sain	78,19	6	853
30	Voisinage d'HSA	Voisinage d'HIP	Tissu sain	78,10	6	938
31	Œdème	Œdème	Œdème	76,05	15	82531
32	Voisinage d'HSA	Œdème	Œdème	75,41	7	2681
33	Voisinage d'HIP	Œdème	Œdème	74,97	6	2079
34	Voisinage d'œdème	Œdème	Œdème	72,53	11	28880
35	Voisinage d'HIP et d'œdème	Œdème	Œdème	72,24	14	11757
36	Voisinage d'œdème	Voisinage d'HSD	Tissu sain	72,10	6	5984
37	Tissu sain	Œdème	Œdème	70,68	14	158267
38	Voisinage d'HIP et d'œdème	Voisinage d'HIP	Tissu sain	70,28	10	3715

Tableau 34. Définition des règles quasi-certaines, très probables et probables, et probabilité d'apparition, nombre de patients avec présence de l'évolution, et nombre de voxels concernés associés aux règles. Sont surlignées en bleu les cellules concernant un voisinage de lésions. Les œdèmes à J3 sont mis en évidence en gras. Les règles sont classées au sein de leur groupe par probabilité d'apparition décroissante.

Après établissement du dictionnaire et application des seuils sur la probabilité d'apparition et le nombre de patients concernés, on obtient 38 règles : 6 règles quasi-certaines, 10 règles très probables et 22 règles probables. Le reste des évolutions possibles sont dites contingentes. On peut tout d'abord noter que la règle 1, la plus fréquente, est la stagnation du tissu sain de J0 à J1 et J3. Le grand nombre de voxels suivant cette évolution rend cette règle évidente.

On peut également remarquer que parmi les 16 évolutions quasi-certaines et très probables, aucune ne permet de prédire une transformation lésionnelle à J3. Néanmoins, on peut interpréter les règles numéro 2, 4, 5, 8, 10, 29 comme le fait que l'HSD a tendance à se résorber, tout comme l'HSA (règles numéro 3, 6, 9, 15, 23, 24, 26, 27) et l'HIP (règles numéro 7, 12, 13, 14, 38). On retrouve au sein des règles probables 6 règles qui aboutissent à un voxel œdémateux à J3. Toutes ces règles ont la particularité de concerner des voxels déjà œdémateux à J1. Néanmoins, au vu de l'état de ces voxels à J0, on peut conjecturer la tendance des œdèmes à s'étendre. Tous ces points sont cohérents avec ce qu'observent les médecins en clinique et cohérent avec les grandes évolutions à l'échelle de la cohorte visibles sur le diagramme de Sankey Figure 57.

7.2.3.2 Segmentation empirique

Si l'on utilise les fréquences d'apparition de l'ensemble des évolutions du dictionnaire (et non plus seulement celles des 38 règles définies plus tôt) afin d'attribuer à chaque voxel des images à J3 des patients du jeu de test l'état lésionnel le plus fréquent, on peut générer les segmentations empiriques générées à partir du dictionnaire et des segmentations étendues à J0 et J1. Nous avons généré ces segmentations et calculé le score de Dice entre la segmentation empirique et la segmentation manuelle à J3, sur l'ensemble du tissu lésionnel ainsi que sur chacune des 7 segmentations individuelles. Ces résultats sont rassemblés sur le Tableau 35, et une illustration de la segmentation empirique d'un patient du jeu de test est disponible sur la Figure 59.

	Lésion globale	HIP	HSD	HED	HIV	HSA	P	O
Segmentation empirique à J3	0,42 (0,11)	0,57 (0,03)	0,18 (0,10)	0 (0)	0 (0)	0 (0)	0,50 (0,12)	0,45 (0,14)

Tableau 35. Moyenne et erreur standard des scores de Dice calculées entre la segmentation empirique à J3 et la segmentation manuelle à J3, sur le jeu de test de la cohorte Radiomic-TBI.

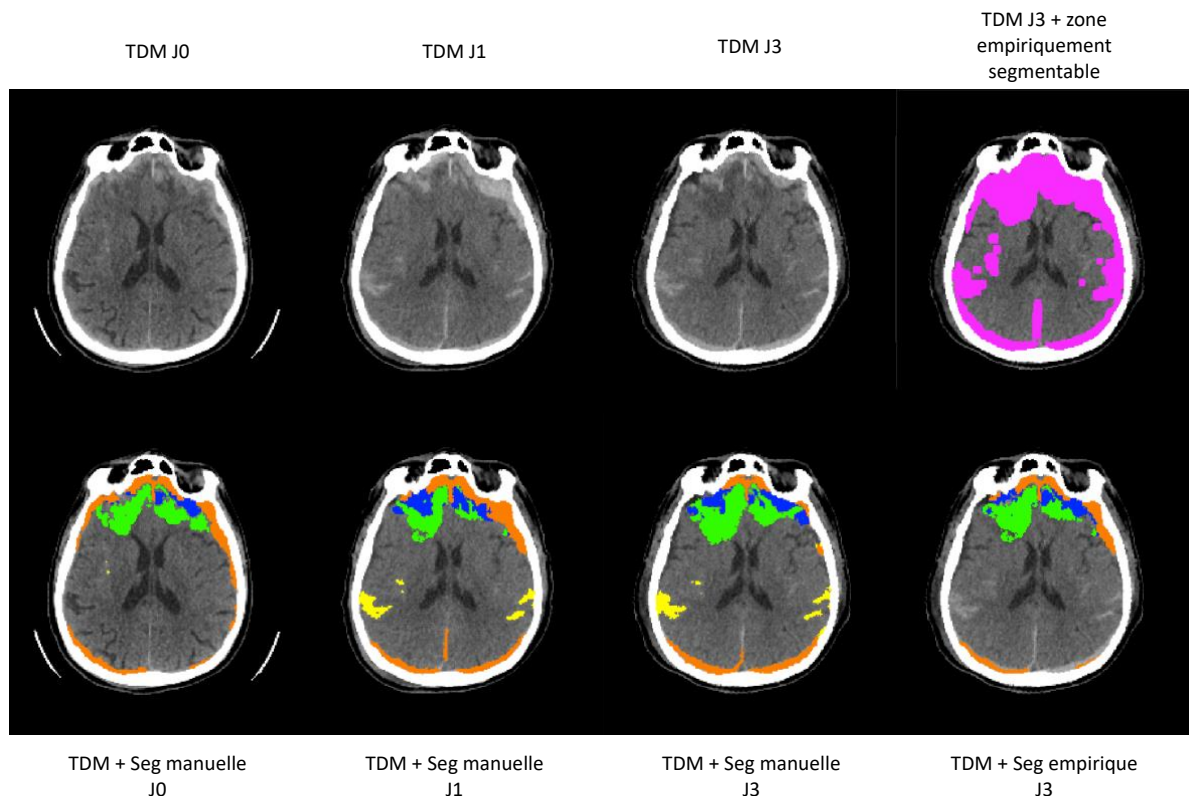


Figure 59. Illustration des images TDM d'un patient de test de la cohorte Radiomic-TBI, à J0, J1 et J3. Sur les 3 premières colonnes, les segmentations manuelles ont été superposées aux images. Sur la 4ème colonne, la ROI violette représente les voxels concernés par notre segmentation empirique à J3 (en haut) ainsi que le résultat de la segmentation empirique à J3 (en bas). Légende : Bleu – HIP, Orange – HSD, Jaune – HSA, Vert – Œdème, Violet clair – Union des voisinages et des lésions à J0 et à J1 (donc la zone empiriquement segmentable)

On remarque premièrement la grande disparité entre les scores de Dice des lésions individuelles. Si la segmentation paraît performante sur l'HIP, les pétéchiées et l'œdème, elle est tout simplement inexistante sur l'HED, l'HIV et l'HSA. Nos données montrent qu'il n'existe pas d'évolution typique entre J0 et J1 qui mènent vers ces trois types de lésions à J3. Ainsi, les évolutions qui conduisent à ces lésions sont contingentes et ne semblent pas être prédictibles à partir de l'état lésionnel à J0 et J1. C'est ce qu'on observe également sur la Figure 59, où l'HSA, pourtant présente à J1 et à J3, est totalement oubliée par la segmentation empirique, qui prédit un tissu sain. Néanmoins, sur l'œdème et l'HIP, on obtient des résultats cohérents avec la segmentation manuelle à J3. Il est toutefois important de noter que la forme de la segmentation empirique est très similaire à la forme de la segmentation à J1. Cette observation est la conséquence directe de la principale limitation de cette méthode. En effet, puisque par définition la segmentation dépend uniquement de la segmentation à J0 et J1 (et de leurs voisinages), il n'est pas possible de prédire une apparition lésionnelle spontanée. Sur la Figure 59, on peut visualiser en violet la zone empiriquement segmentable, qui correspond à l'union de toutes les classes de la segmentation étendue à J0 et à J1, c'est-à-dire l'ensemble des voxels dont l'évolution entre J0 et J1 n'est pas Tissu Sain → Tissu Sain. Une lésion apparaissant spontanément en dehors de cette zone empiriquement segmentable n'est actuellement pas prédictible par notre méthode. Finalement, cette approche sera d'autant plus puissante que la

segmentation du cerveau sera fine, avec par exemple d'introduction de nouvelles classes représentant les voisinages proches ou lointains des lésions et que le nombre de patients utilisés pour créer le dictionnaire sera grand.

7.2.4 Conclusion et perspectives de l'étude

Nous pensons, après l'analyse de ces résultats préliminaires, que l'attribution à un voxel de J3 de l'état lésionnel le plus fréquent en fonction de son évolution entre J0 et J1 n'est pas forcément la méthode la plus adaptée, étant donné le poids de la classe « Tissu Sain » au sein des évolutions. Par exemple, un voxel qui suit une évolution entre J0 et J1 associée avec une probabilité de 50% de devenir du Tissu Sain, 40% un œdème et 10% toute autre lésion à J3 va être prédit comme Tissus Sain (probabilité majoritaire), alors qu'il a tout de même une grande chance d'être œdémateux. Pour pallier cette limitation, nous pourrions décréter qu'en dessous d'une certaine probabilité de se transformer en tissu sain, on attribue à un voxel la classe lésionnelle majoritaire, sans prendre en compte le tissu sain. Nous pourrions par exemple utiliser un jeu de validation, afin de déterminer le seuil optimal à appliquer sur la probabilité pour maximiser le score de Dice entre la segmentation empirique et la segmentation manuelle. Une autre amélioration possible pourrait consister à étendre les voisinages des lésions afin d'étudier l'ensemble du cerveau et ainsi disposer d'évolutions plus précises. Néanmoins si on augmente le nombre d'évolutions possibles on diminue le nombre de voxels partageant la même évolution et donc on érode la puissance des règles ce qui impliquerait une augmentation du nombre de données nécessaires pour mener à bien cette étude.

Cette segmentation empirique est donc à améliorer, mais, si elle était calculée sur une cohorte bien plus importante, pourrait approximer de manière pertinente la fréquence des différentes transformations lésionnelles des patients TC et ainsi fournir une information descriptive qui n'existe à notre connaissance pas dans la littérature scientifique à l'heure actuelle.

7.3 Futurs projets

Jusqu'ici, nous avons mis en évidence la différenciation des zones HIP qui se transforment en œdème, et nous avons testé une approche empirique pour générer la segmentation à J3 à partir des segmentations manuelles à J0 et J1. Ces deux études doivent encore être consolidées et validées. C'est ce que nous allons faire dans les prochains mois, dans le cadre d'un contrat post-doctoral de quelques mois que je vais effectuer dans l'équipe.

Pour aller encore plus loin, nous allons également tenter de générer la segmentation totale à J3, et non plus seulement l'œdème, à partir des images TDM à J0 et J1. Ce travail n'a pas encore débuté, mais la suite de ce chapitre liste les deux pistes que je vais explorer pour prédire la segmentation totale à J3.

7.3.1 Génération de la segmentation à J3 à partir des images TDM à J0 et J1

Comme détaillé dans le chapitre 4, nos travaux sur l'amélioration d'un réseau de neurones convolutifs (CNN) se basent sur l'architecture DeepMedic, qui est très utilisée pour la segmentation automatique d'images médicales. Nous avons ainsi utilisé cette architecture pour générer la segmentation à 7 classes d'une image TDM à partir de cette même image TDM. Entraînée, cette architecture a permis d'obtenir de bons résultats de segmentations, c'est-à-dire qu'elle a automatiquement trouvé des caractéristiques uniques qui, extraites de l'image, mettent en évidence les lésions, ce qui permet ensuite de bien les segmenter. L'extraction de ces caractéristiques par le CNN qui correspondent principalement aux noyaux des couches de convolution du réseau peuvent être conceptuellement assimilées aux calculs de métriques Radiomiques dont la formule est déterminée par le réseau. Puisque notre étude sur la différenciation des zones de l'HIP qui se transformeront en œdème à partir de métriques Radiomiques affiche des résultats encourageants, nous pensons qu'il peut exister des métriques de types Radiomiques contenues dans les images TDM à J0 et J1 qui permettraient de générer la segmentation à 7 classes à J3.

Ainsi, nous souhaitons entraîner l'architecture DeepMedic à générer la segmentation à 7 classes à J3 à partir des images TDM acquises à J0 et à J1. Nous espérons que l'entraînement permettra de trouver les métriques capables de mettre en évidence cette information d'évolution. S'il fonctionne, un tel outil permettrait de prédire le futur lésionnel d'un patient à 48h dès le premier jour après l'accident.

Nous pourrions également exploiter les images TDM acquises à J0 et J1 à l'échelle du voxel et de son voisinage, en extrayant des métriques représentant les propriétés de l'image au niveau local, comme par exemple les métriques radiomiques sur un petit patch autour du voxel d'intérêt, ou bien en encodant l'information spatio-temporelle du voisinage de chaque voxel du cerveau puis en calculant le motif binaire local associé, ce qui permettrait, après entraînement d'un classifieur, de prédire l'état lésionnel d'un voxel à J3 à partir des images TDM à J0 et J1. Cette dernière méthode a déjà été utilisée avec succès sur des images IRM de perfusion afin de segmenter la lésion ischémique après un accident vasculaire cérébral [134]. Bien que la résolution temporelle des images IRM de perfusion soit sans commune mesure avec celle de notre cohorte Radiomic-TBI, cette méthode pourrait nous fournir un moyen rapide de combiner les informations spatiales et temporelles de notre cohorte.

7.3.2 Génération de l'image TDM à J3 grâce aux Autoencoder/Réseaux récurrents

Après avoir généré la segmentation à J3 à partir de l'image TDM à J3 (chapitre 4), après avoir étudié la génération de la segmentation à J3 à partir de la segmentation à J0 et J1, nous aimerions étudier la génération de l'image TDM à J3 à partir des images TDM à J0 et J1.

Entraîner un CNN à générer un contraste à partir d'un autre contraste peut sembler irréaliste. Pourtant, des études récentes ont montré qu'il était par exemple possible de générer automatiquement une image IRM après injection de Gadolinium à partir d'une image sans injection comme par exemple dans cet article de Preetha et al. publié dans The Lancet Digital Health en 2021 [135]. Notre équipe a déjà utilisé ce genre de techniques et j'ai notamment co-encadré plusieurs travaux pratiques sur le sujet dans le cadre de l'école d'hiver AI4Health (cf chapitre 9).

Parmi la multitude d'architecture de CNN existantes nous souhaitons nous concentrer sur deux types : les autoencodeurs et les réseaux neuronaux récurrents (RNN). Les autoencodeurs sont un type de réseaux de neurones artificiels qui prend une image en entrée et qui essaye de reconstituer cette même image en sortie. L'architecture de ce type de réseau comprenant un goulot d'étranglement (ou Bottleneck), au centre du réseau dont le nombre de paramètres dans lesquels est résumée l'image est faible par rapport au nombre de voxels de l'image d'entrée. L'image est donc résumée en un petit nombre de métriques, appelées variables latentes. L'entraînement du réseau permet de s'assurer que les variables latentes contiennent assez d'information pour pouvoir bien reconstruire l'image initiale. Une illustration de l'architecture d'un autoencodeur est disponible Figure 60, issue de des travaux de Veronica Muñoz-Ramirez, ancienne doctorante dans notre équipe [136]. De leur côté, les réseaux de neurones récurrents (RNN) sont un type de réseaux de neurones artificiels capable d'exploiter la relation temporelle des données d'entrées et de générer la suite de l'entrée grâce à des cellules LSTM (Long short-term memory) qui permet de garder en mémoire l'information temporelle des données d'entrée.

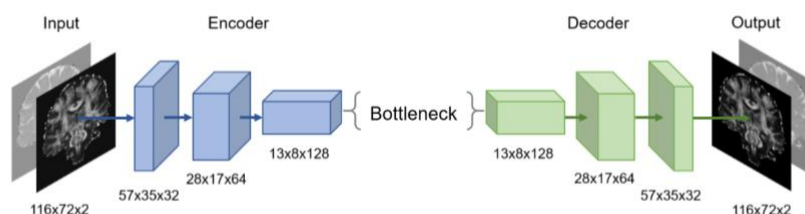


Figure 60. Illustration d'un autoencodeur. Image tirée de [136]

Afin de générer l'image TDM à J3 à partir de celles de J0 et J1, nous souhaitons donc combiner ces deux architectures autoencodeurs et RNN. Plus précisément, nous voulons ainsi exploiter l'importante cohorte CT-Harmo (987 patients) pour entraîner un autoencodeur et ensuite appliquer ce réseau sur la cohorte Radiomic-TBI, afin de résumer ses images TDM en un faible nombre de paramètres. Une fois ces images résumées dans un vecteur latent, nous souhaitons exploiter la longitudinalité en entraînant un RNN capable de générer le vecteur latent de J3 à partir de celui de J0 et celui de J1. Une fois le vecteur latent généré, il pourra être décodé grâce à l'autoencodeur préalablement entraîné. Puis, nous pourrions comparer l'image prédite à l'image réelle grâce à des métriques comme la Mean Absolute Error (MAE) ou encore la similarité structurelle (SSIM). Cette approche combinant deux réseaux séparés est illustrée sur la Figure 61. On pourrait même aller jusqu'à imaginer utiliser l'image TDM de J3 générée afin de segmenter les 7 types de lésions grâce à notre algorithme développé au chapitre 4, et ainsi comparer nos différentes méthodes de génération de la segmentation à J3.

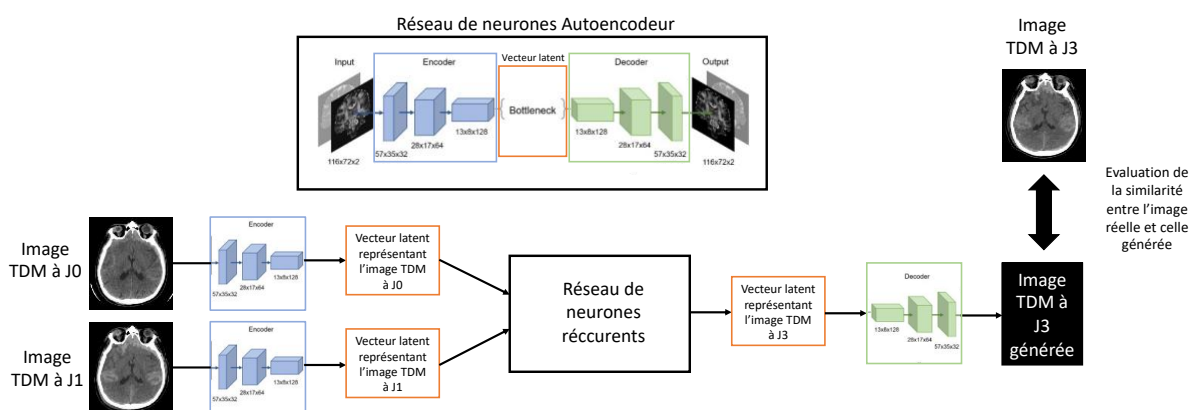


Figure 61. Illustration de l'approche de génération de l'image TDM à partir d'un autoencodeur et d'un réseau récurrent que nous souhaitons mettre en place. La figure de l'autoencodeur provient de [136].

7.4 Conclusion

La richesse de la cohorte Radiomic-TBI réside, en outre, dans sa longitudinalité. Nous faisons l'hypothèse que les mécanismes biologiques qui résultent en un état lésionnel à J3 sont déjà à l'œuvre à J0 et à J1 et qu'ils peuvent être quantifiés sur les images TDM. Nous pensons donc qu'il est possible de prédire le futur lésionnel d'un patient à partir soit des images TDM de ce patient à J0 et à J1, soit à partir des segmentations des lésions à J0 et à J1.

Nous avons déjà mis en œuvre 2 méthodes d'estimation de l'œdème à J3 : la classification par métrique radiomiques et la segmentation empirique. Si les résultats sont encourageants, ils doivent être consolidés. Ce sera l'objet de mon futur post-doctorat dans l'équipe. Durant ces quelques mois, nous allons également tester des méthodes bien plus

ambitieuses de génération de la segmentation à J3, voire de la génération de l'image TDM à J3.

Néanmoins, pour exploiter les évolutions scanographiques entre J0 et J1, il faut pouvoir comparer les images TDM à J0 et J1. Les images n'étant pas forcément obtenues avec les mêmes protocoles d'acquisition ou sur les mêmes machines, se pose la question de la variabilité potentiellement introduite sur les images par ces différences de machine et de protocoles d'acquisition. C'est pourquoi en parallèle des objectifs principaux de ma thèse, je me suis intéressé à l'étude de la variabilité inter machines et inter protocoles sur les images TDM acquises au CHUGA. Cette étude est détaillée dans le chapitre suivant.

Chapitre 8 : Étude de la variabilité des images TDM sur une cohorte de soin courant

Pour poursuivre l'étude longitudinale des lésions, nous souhaitons mesurer non plus seulement l'évolution des segmentations des lésions, mais directement celle des images TDM. Ainsi, nous souhaitons trouver des motifs d'évolution sur les images TDM entre J0 et J1 capables de prédire l'état des lésions à J3, ce qui implique de comparer, pour chaque patient, les images TDM de J0 et de J1, que cela soit sous forme de métriques radiomiques ou tout simplement d'unités Hounsfield. Une telle comparaison entre deux images sous-entend que ces images sont comparables, c'est-à-dire, qu'elles sont quantitatives. Si les images TDM sont naturellement quantitatives, elles sont toutefois sujettes à différentes sources de variabilité, comme détaillé dans le chapitre 1. Les deux principales sources de variabilités sur ces images sont les variabilités inter-machines et inter-protocoles d'acquisition. Une variabilité introduite sur les images pourrait impacter les performances d'algorithmes automatiques et pourraient également créer artificiellement ou masquer un réel effet sur les images qui biaiserait les analyses et conclusions d'une étude.

Les patients de la cohorte Radiomic-TBI ayant été imagés sur différentes machines et avec différents protocoles d'acquisition, nous avons souhaité étudier et quantifier la variabilité des images acquises au CHUGA avant de poursuivre l'analyse longitudinale des images TDM. Si cette variabilité sur les images TDM a déjà été montrée ex-vivo [20] ou in-vivo sur les tumeurs pulmonaires [25], aucune n'a étudié les tissus du parenchyme, et encore moins les cerveaux sains. En effet, le rayonnement des scanners étant ionisant leurs effets nocifs limitent l'utilisation de la TDM sur des sujets sains. Ainsi la plupart des études d'image TDM cérébrales concernent des sujets pathologiques. Ces pathologies sont de natures différentes et font apparaître sur les images TDM des lésions aux multiples profils comme illustré dans les études précédentes. Cette variabilité pathologique se rajoute à une potentielle variabilité inter-machines et inter-protocoles, ce qui biaiserait la quantification de ces dernières. La présence d'artefacts, récurrente, pourrait également introduire un biais capable de créer artificiellement ou de masquer un effet. L'idéal serait donc d'étudier la variabilité inter-machines et inter-protocoles sur des images TDM de cerveaux sains non artefactées. Nous avons donc déposé un protocole de recherche

clinique rétrospective qui nous a permis de collecter une cohorte, appelée CT-harmo, d'environ 1000 images TDM de patients de soin courant du CHUGA comprenant à la fois des sujets sains et pathologiques plus ou moins artéfactés acquis sur différentes machines avec différents protocoles d'acquisitions.

8.1 Objectifs

Cette étude avait deux objectifs. Premièrement : développer un pipeline de sélection automatique des images TDM saines dans une cohorte de soin courant et deuxièmement : étudier une potentielle variabilité introduite sur les images TDM par les différents protocoles et machines d'acquisition du CHUGA à l'aide des métriques radiomiques. Cette étude s'est déroulée dans le cadre d'un stage de Master 2 que j'ai co-encadré.

8.2 Protocole clinique : cohorte CT-Harmo

Le CHUGA dispose de 3 machines d'acquisition scanner différentes. Pour chaque machine, plusieurs protocoles d'acquisition sont utilisés, avec différents paramètres d'acquisition, comme la résolution de l'image de sortie ou l'énergie des rayons X envoyés. Ces différents protocoles correspondent habituellement aux différents manipulateurs radio ou aux différents médecins radiologues du service.

Afin de réaliser cette étude sur la caractérisation de la variabilité des images TDM, nous avons écrit un protocole de recherche clinique rétrospective nommé CT-Harmo, visant à récupérer plus de 1036 images TDM passés par le service de déchocage du CHUGA entre janvier 2019 et Juillet 2020, ainsi que le compte rendu radiologique écrit par les radiologues ayant analysé ces images. Dans ce protocole, nous avons décrit notre étude, en détaillant les méthodes qui seront décrites dans les paragraphes suivants. Ce protocole a été accepté par la Direction de la Recherche Clinique et de l'Innovation (DRCI) et a été enregistré sur le Health Data Hub (HDH) sous le numéro F20201127123438.

Parmi les patients passés au déchocage, l'immense majorité passe un examen TDM cérébral. Pourtant, il est estimé que seulement la moitié des patients présente des lésions cérébrales. En effet, on peut très bien passer au déchocage pour une blessure qui n'impacte pas le cerveau, et pourtant passer un examen TDM par précaution. Ainsi, une grande partie des images TDM de la cohorte CT-Harmo seront des images TDM sans lésions cérébrales, assimilées à des images TDM saines. Puisque les images TDM induisent des rayonnements ionisants pour les patients, il est rare de disposer de TDM de cerveau sain. C'est ce qui fait la principale richesse de cette cohorte.

8.3 Recueil et stockage des données

Afin de recueillir de manière rétrospective des données médicales, la loi française oblige les investigateurs de l'étude à informer les patients du type de données récupérées, de l'utilisation de leurs données, et du but de l'étude à laquelle ils participent. Une fois informés par courrier, les patients ont un mois pour s'opposer à l'utilisation de leurs données. En cas de retour du courrier, par exemple à cause d'un déménagement non connu par le CHU, ou en cas de non réponse, le patient accepte par défaut l'utilisation de ses données.

Les 1036 patients majeurs passés au déchocage entre janvier 2019 et Juillet 2020 et imagés par TDM au CHUGA ont été contactés par courrier fin Décembre 2020. Début 2021, après le délai réglementaire d'un mois, nous avons pu accéder aux données. Les images ont été automatiquement extraites du PACS du CHUGA au format DICOM, anonymisées et envoyées sur la plateforme Shanoir, spécialisée en stockage sécurisé de données d'imagerie médicale [107]. Les comptes rendus radiologiques au format pdf ont quant à eux été automatiquement anonymisés et stockés sur un serveur au sein de notre équipe. Ce travail de sélection, d'anonymisation, de transfert et de conversion, a été effectué par Jean-Côme Douteau, ingénieur Shanoir travaillant dans nos locaux au GIN. La Figure 62 récapitule les étapes principales de la mise en place de l'étude de recherche CT-Harmo.

Quelques patients se sont opposés à l'utilisation de leurs données, certaines images n'ont pas été retrouvées, de même que certains comptes rendus. Au final, 987 images TDM et comptes rendus ont été récupérés et constituent donc la cohorte CT-Harmo.

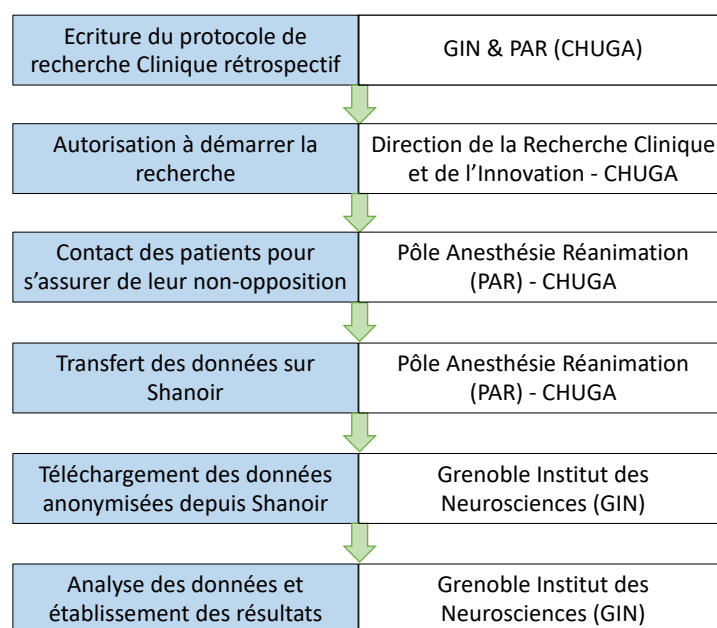


Figure 62. Circuit des étapes de la mise en place d'une étude de recherche clinique rétrospective, depuis l'établissement du protocole jusqu'à l'analyse des données. A droite, les structures impliquées à chaque étape

8.4 Méthode – Encadrement de stage : Mise en évidence de la variabilité inter machine

Afin d'avancer en parallèle sur cette étude et sur les études présentées dans les chapitres précédents, nous avons fait de cette étude un sujet de stage de Master 2. Nous avons fait passer une petite dizaine d'entretiens et nous avons finalement choisi de recruter Vaëa Tesan, étudiante en master 2 pour mener à bien cette étude entre février et juillet 2021 et que j'ai personnellement encadrée. Depuis la fin de son stage, Vaëa est doctorante dans notre équipe et continue de travailler sur ces problématiques d'analyse des données de soin courant pour améliorer la robustesse des algorithmes de quantification automatique des images TDM de patients TC. Le travail présenté dans la fin de ce chapitre reprend une partie du stage de M2 de Vaëa TESAN. Le rapport de stage écrit par Vaëa à la fin de son stage, qui présente plus en détails cette étude, est disponible sur la plateforme HAL [137].

8.4.1 Prétraitement des images

Les résultats présentés ici se basent sur une analyse partielle de la cohorte CT-Harmo. En effet, pour des questions de temps de calcul, de retards dans la mise à disposition de la cohorte et de durée limitée du stage seulement 243 patients sur les 987 que compte cette cohorte ont été analysées,

Les 243 images TDM ont été téléchargées de la plateforme Shanoir au format DICOM. Nous avons ensuite utilisé le logiciel MRI_conv (<https://irmage.univ-grenoble-alpes.fr/services/traitement-dimages/mri-file-manager>), développé dans l'équipe, pour convertir les DICOM en fichiers NIfTI et json. Ainsi les images seront stockées dans les fichiers NIfTI et tous les paramètres d'acquisition des images seront contenus dans le fichier json comme le modèle de scanner utilisé ou le protocole d'acquisition.

Toutes ces images TDM ont été rééchantillonnées à 1mm³, et le cerveau a été extrait grâce à un algorithme déjà existant [121] que nous avons modifié afin de ne garder que la zone de voxels connectés la plus volumineuse et de remplir les potentiels trous au sein de cette zone.

8.4.2 Exclusion des sujets pathologiques : Segmentation des lésions

Au début du stage de Vaëa, nous n'avions pas encore accès aux compte rendus radiologiques qui permettent de statuer sur la présence de lésions cérébrales ou pas sur les images TDM. Ainsi, dans le but d'identifier et de retirer les images pathologiques de la cohorte, qui pourraient biaiser nos analyses, nous avons segmenté de manière

automatique 4 types de lésions du TC grâce à l'algorithme BLAST-CT [102]. Les 4 lésions segmentées sont :

- L'hémorragie intra-parenchymateuse (HIP)
- L'hémorragie extra-axiale (HEA)
- L'œdème
- L'hémorragie intra-ventriculaire (HIV)

Il est important de noter que cette étude a été réalisée avant la fin de l'étude sur l'amélioration de BLAST-CT (chapitre 4). C'est pourquoi, bien que non optimal, comme montré au Chapitre 4, nous avons utilisé ici l'algorithme natif de BLAST-CT. A partir de cette segmentation, et afin de limiter le bruit potentiel, nous avons conservé l'ensemble de voxels connectés le plus volumineux. Ainsi seule la lésion la plus volumineuse a été conservée.

Pour chaque image TDM, nous avons ensuite mesuré le volume de chacun des 4 types de lésions. Chacun de ces volumes a ensuite été divisé par le volume total du cerveau pour obtenir un ratio de tissu lésionnel au sein du parenchyme cérébral.

Pour pallier la sur-segmentation induite par BLAST-CT, comme l'HEA qui est régulièrement surestimée au niveau des sinus veineux, nous avons fixé un seuil sur chacun de ces ratios, au-delà duquel l'image TDM est déclarée pathologique et donc exclue du reste de l'analyse. Les valeurs de ces seuils ont été fixées arbitrairement après un examen visuel.

8.4.3 Segmentation de la matière blanche, matière grise et du liquide céphalo-rachidien

Afin de mesurer la variabilité potentiellement introduite par les protocoles d'acquisition sur les images TDM, nous avons dû éviter au maximum de mesurer la variabilité physiologique entre les patients, qui ont tous des cerveaux différents. Nous avons donc segmenté le cerveau en 3 zones plus ou moins homogènes : la Matière Blanche (MB), Matière Grise (MG), et le Liquide Céphalo Rachidien (LCR). Cette segmentation a été effectuée grâce à l'algorithme FAST du logiciel FSL. Comme conseillé par Cauley & al. [138], les images ont au préalable été filtrées par un filtre gaussien. La valeur de d'écart-type du filtre gaussien a été estimée au préalable à 0,3 afin de maximiser la distance entre les histogrammes des 3 régions d'intérêt (ROI) : MB/MG/LCR.

Nous avons ensuite post-traité ces régions d'intérêt :

- Les 3 ROI ont été masquées par le masque du cerveau préalablement calculé afin d'exclure des segmentations de potentiels voxels de l'espace extra-cérébral.
- Les 3 ROI ont ensuite été morphologiquement érodées et fermées afin d'exclure de potentiels voxels à l'interface entre deux tissus et donc sensibles aux effets de volumes partiels.
- Finalement, nous avons conservé pour chacune des 3 ROI l'ensemble de voxels connectés le plus volumineux, afin de supprimer les voxels non connectés.

8.4.4 Détection automatique des images artefactées

Après avoir exclu les images pathologiques, nous souhaitons exclure les images fortement artefactées. Puisque nous ne disposons d'aucune information dans les métadonnées de l'image nous permettant de soupçonner la présence d'artefacts, nous avons analysé l'image TDM sous forme de métriques radiomiques. Ainsi pour toutes les images TDM, et sur chacune des 3 ROI, nous avons extrait 98 métriques Radiomiques grâce à la bibliothèque python pyRadiomics [133]. Nous avons ensuite exploité les métriques radiomiques précédemment extraites sur les 3 ROI afin de détecter automatiquement les valeurs aberrantes de ces métriques, que nous avons associé à des artefacts.

Pour chaque image TDM, nous disposons donc de 98 métriques radiomiques sur les 3 ROI, soit 294 métriques radiomiques. A partir de ces valeurs, nous avons appliqué un algorithme de détection automatique de valeurs aberrantes nommé Extended Isolation Forest (EIF) [139]. Cet algorithme non supervisé est inspiré de l'algorithme Isolation Forest et spécialement développé pour traiter des valeurs en grande dimension. Ainsi, nous avons pu détecter les images associées aux valeurs radiomiques aberrantes, ce que nous avons assimilé en regardant les images exclues, à des images artefactées.

8.4.5 Représentation graphique et quantification de la variabilité

Après avoir exclu de nos données les images pathologiques et artefactées, nous avons pu utiliser les métriques radiomiques déjà extraites sur les images TDM restantes afin d'estimer la variabilité inter machine/protocole. Ainsi, nous disposons de 98 métriques radiomiques pour chacune des 3 ROI. Premièrement, nous avons appliqué une analyse en composantes principales afin de combiner ces 98 métriques radiomiques et de conserver les 3 composantes principales. Ainsi, chaque ROI de chaque image TDM peut être représentée dans un espace à 3 dimensions.

A chaque point, on peut associer un tissu (MB, MG, ou LCR), mais on peut également associer un couple protocole/machine. Afin d'estimer la variabilité physiologique entre ces 3 tissus, nous avons calculé la position du barycentre de chacun des nuages de points appartenant au même tissu. Ainsi, nous avons déterminé le point représentant le centre du nuage de points de chaque tissu et donc la position moyenne des points associés à ce tissu. Ces 3 points sont appelés barycentres des tissus. Nous avons finalement pu calculer les 3 distances euclidiennes entre ces 3 barycentres, nommées distances inter-tissus, afin d'estimer la variabilité sur les métriques radiomiques entre les 3 ROI.

Afin d'estimer la variabilité introduite par les différents couples protocole/machine nous avons, pour chaque tissu, calculé le barycentre du nuage de points partageant le même couple protocole/machine, appelé barycentre du protocole X

associé au tissu Y. La distance euclidienne entre ces points et le barycentre du tissu associé, nommée distance protocole/tissu a ensuite été calculée pour chaque couple protocole/machine et nommée distance du protocole au tissu.

Finalement, pour chaque ROI, les distances protocole/tissu ont été comparées aux distances inter-tissus, ce qui nous a permis de mesurer la variabilité introduite par les différents protocoles sur les images par rapport à la différence entre les tissus.

La Figure 63 représente le pipeline de traitement global, du recueil des images jusqu'à la quantification de la variabilité inter-protocoles.

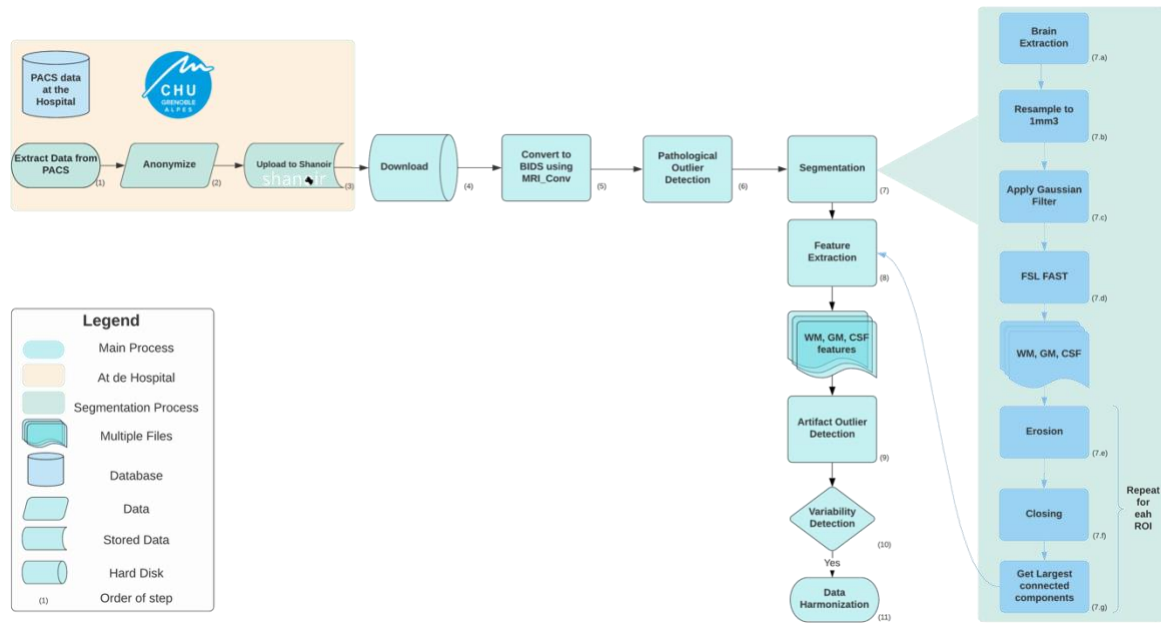


Figure 63. Pipeline de traitement des images TDM, à partir du recueil des données au CHU. Image issue du M2 de Vaëa Tesan.

8.5 Résultats

8.5.1 Détection des images pathologiques

Sur les 243 images TDM initiales, BLAST-CT a été appliqué afin de segmenter les 4 types de lésions. Ces lésions ont été divisées par le volume total du cerveau de manière à obtenir des ratios et ces ratios ont été représentés sous forme d'histogramme. L'histogramme de la lésion HEA est montré sur la Figure 64.

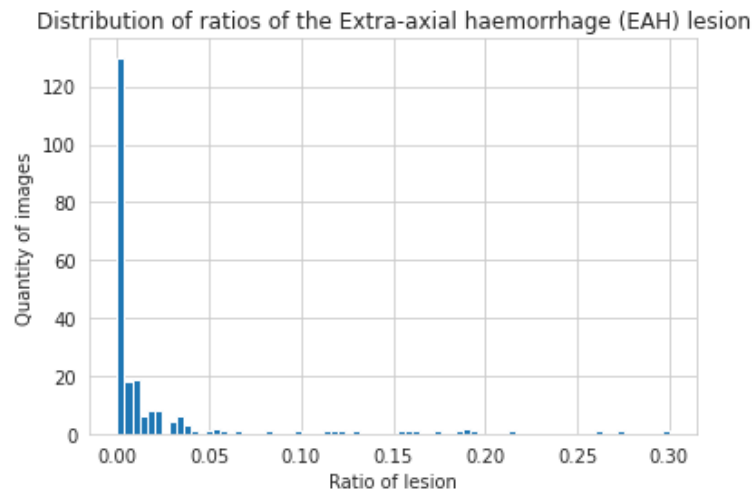


Figure 64. Histogramme de la répartition des ratios d'HEA pour les 243 images TDM. Image issue du M2 de Vaëa Tesan.

A partir de ces ratios, le seuil de détection d'images pathologiques a été arbitrairement fixé à 0,04 pour l'HEA et à 0,01 pour les 3 autres lésions. Ce seuil a été estimé en regardant les images correspondantes et en analysant si les volumes segmentés sont dus au bruit lié à une sur-segmentations des lésions par BLAST-CT ou à une réelle lésion. Le seuil est plus élevé pour l'HEA car cette lésion est fréquemment surestimée par BLAST-CT, notamment au niveau du sinus veineux.

Grâce à ces seuils, sur 243 images TDM, 106 ont été détectées comme étant pathologiques, soit plus de 43% du total. Ces 106 images ont donc été exclues du reste de l'analyse qui a été poursuivie sur les 137 images restantes.

8.5.2 Détection des images artefactées

Sur ces 137 images restantes, nous avons segmenté les 3 ROI MB/MG/LCR grâce à l'algorithme FAST et à notre post-traitement. Une illustration des ROI obtenues à la fin du post traitement des segmentations est disponible Figure 65.

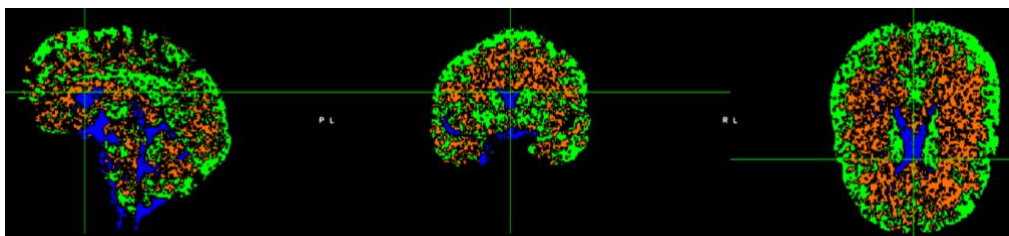


Figure 65. Illustration de la segmentation de la matière grise (vert), de la matière blanche (orange) et du liquide céphalo rachidien (bleu). Image issue du M2 de Vaëa Tesan.

Après application de l'algorithme EIF, dont le but est de détecter les valeurs aberrantes, sur les métriques radiomiques extraites sur chaque image TDM à partir des 3 ROI MG/MB/TDM, 10 images ont été détectées comme ayant des propriétés aberrantes. Visuellement, ces 10 images, dont une est illustré sur la Figure 66, sont bien artefactées.

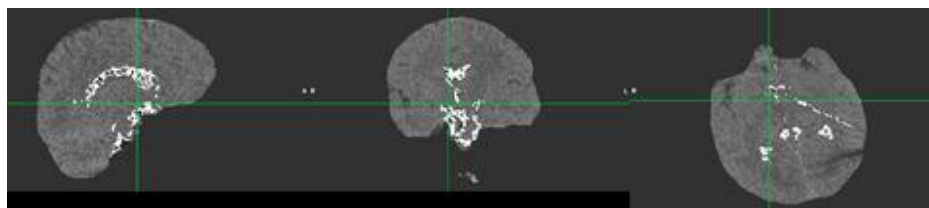


Figure 66. Image artefactée exclue après analyse des métriques Radiomiques par l'algorithme EIF. On remarque en blanc la ROI du LCR, biaisée par l'artefact, et dont les métriques radiomiques extraites sur cette ROI ont conduit à exclure cette image. Image issue du M2 se Vaëa Tesan.

Une fois ces images exclues, nous avons poursuivi l'analyse sur les 127 images restantes. Le flux des images TDM de la cohorte CT-Harmo, depuis l'établissement du protocole jusqu'à l'exclusion des images artefactées, est disponible sur la Figure 67.

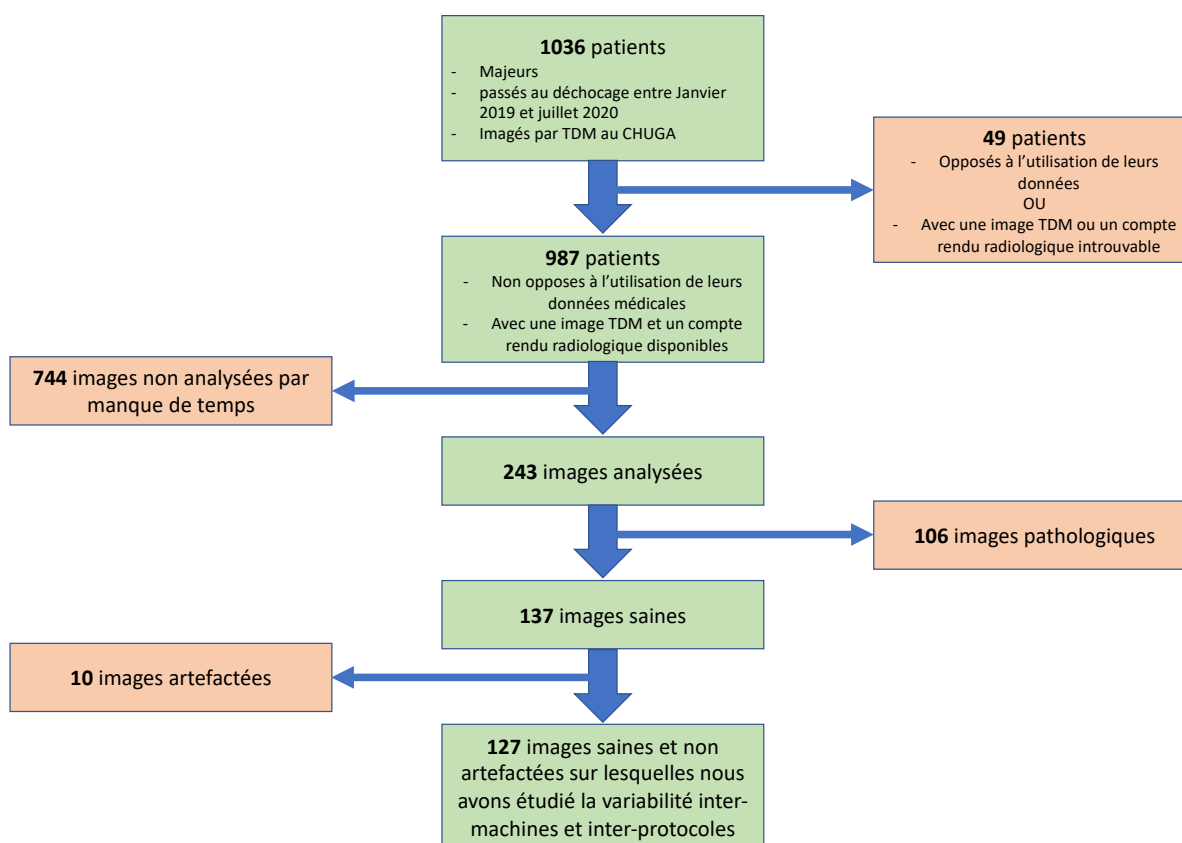


Figure 67. Flux des images de la cohorte CT-Harmo, depuis celles qui remplissent les critères d'inclusion jusqu'à celles utilisées pour l'analyse de la variabilité durant le stage de Vaëa Tesan.

8.5.3 Étude de la variabilité inter-machines/protocoles

Sur les 127 images TDM saines et non artefactées, nous avons récupéré le nom du protocole et de la machine utilisée au moment de l'acquisition de l'image. La répartition des images par couple protocole/machine est représentée sous forme de diagramme camembert sur la Figure 68.

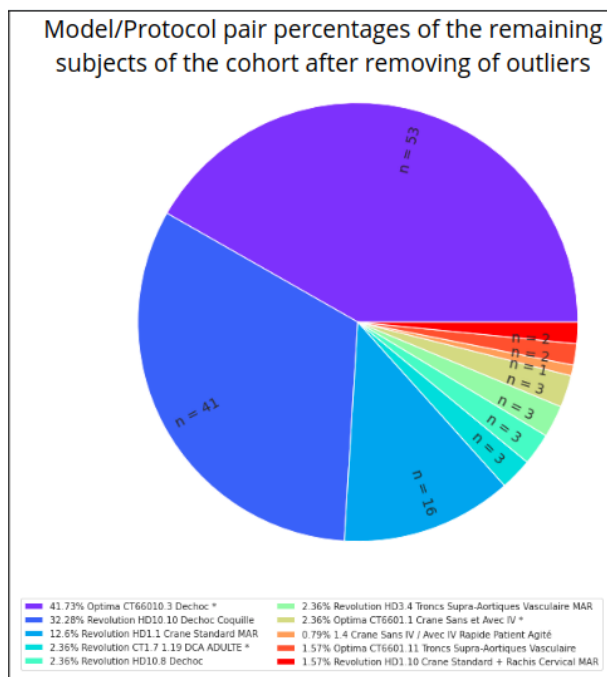


Figure 68. Répartition des 127 images TDM par couple protocole d'acquisition /machine d'acquisition. Image issue du M2 de Vaëa Tesan.

On voit la prépondérance de trois couples protocole/machine qui à eux trois rassemblent 110 images sur 127. Les 17 autres images se partagent entre 7 couples protocole/machine différents. Sur ces 127 images, on dispose de 3 jeux de métriques radiomiques : un par ROI. Ces 381 jeux de métriques peuvent être représentés en 3 dimensions après leur réduction de dimension par analyse en composantes principales, comme sur la Figure 69.

3 component PCA without outliers

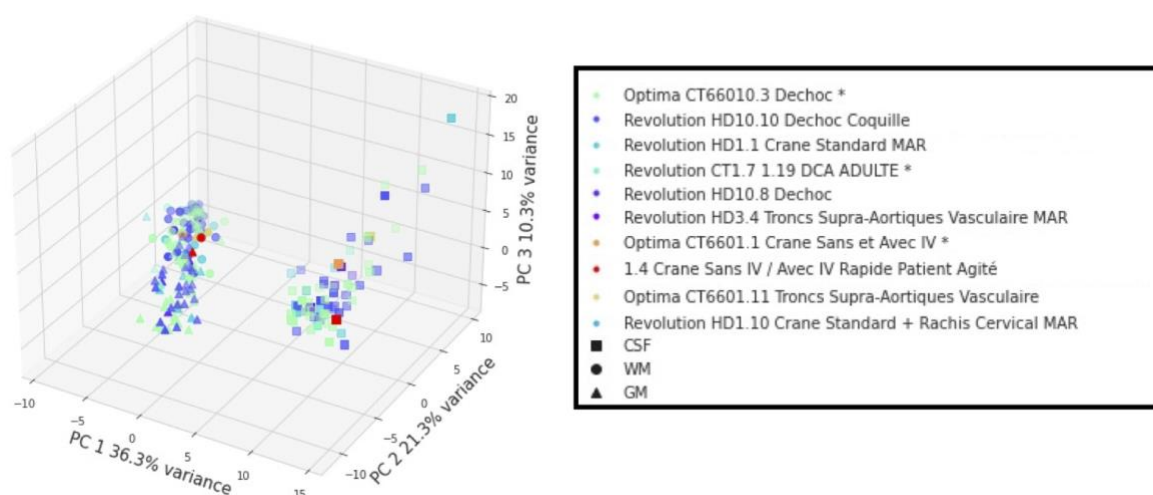


Figure 69. Représentation en 3 dimensions (après analyse en composantes principales) des 381 points représentant les métriques radiométriques extraites des images des sujets sains non artéfactés. Chaque forme de marqueur représente l'une des 3 ROI : rond pour la matière blanche, triangle pour la matière grise, et carré pour le liquide céphalo rachidien. Chaque couleur représente l'un des 10 couples protocole/machine. Image issue du M2 de Vaëa Tesan.

Grâce à ces points, nous avons pu calculer, pour chacun des 3 tissus, la position du barycentre du nuage de points. Ces barycentres des tissus sont représentés à gauche sur la

Figure 70, ainsi que les 3 distances euclidiennes entre ces 3 barycentres. La distance inter-tissus entre les 3 tissus est également disponible sur le Tableau 36. On confirme bien quelque chose d'intuitif : la MB et la MG sont plus proches les unes des autres que du LCR.

Distance LCR ⇔ MB	Distance LCR ⇔ MG	Distance MG ⇔ MB	Distance moyenne inter-tissus
14,46	12,47	10,56	12,50

Tableau 36. Distances euclidiennes entre les 3 barycentres représentant les 3 tissus et distance moyenne inter-tissus.

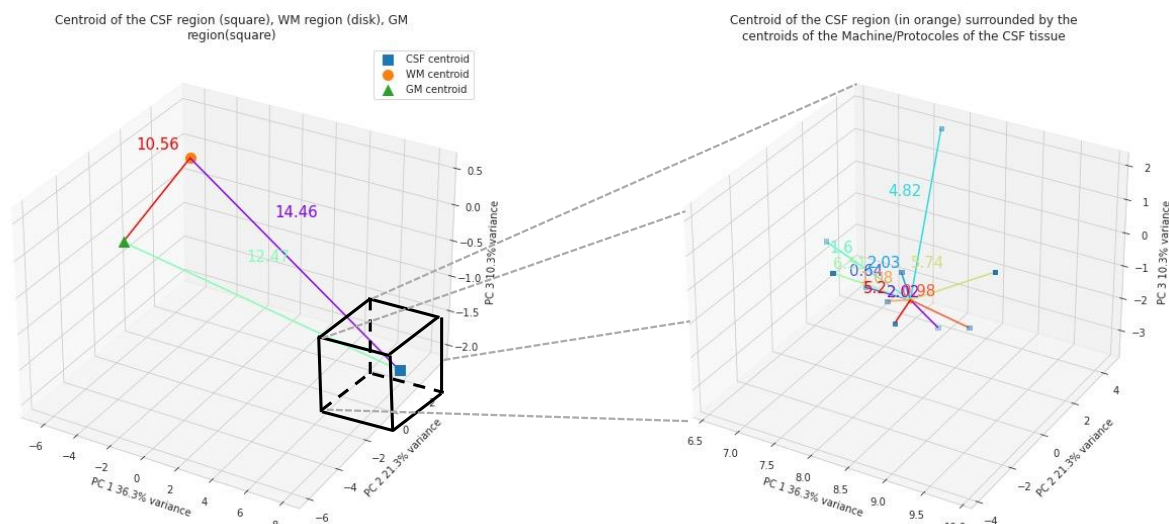


Figure 70. Gauche – Position des 3 barycentres correspondants aux 3 tissus et distances euclidiennes entre ces 3 barycentres. Droite – Position des barycentres de chacun des 10 protocoles par rapport au tissu LCR. La distance euclidienne entre ces points et le barycentre du LCR préalablement calculé est également représentée. Légende identique à celle de la Figure 69. Image issue du M2 de Vaëa Tesan.

Ensuite, pour chaque tissu, nous avons pu calculer la position du barycentre de tous les points appartenant au même couple protocole/machine. On obtient ainsi 10 barycentres relatifs aux protocoles pour chaque tissu, et on peut donc calculer la distance euclidienne entre chacun de ces points au barycentre du tissu. Un exemple sur le LCR est visible à droite sur la

Figure 70. Les distances euclidiennes barycentres/tissus pour les 10 protocoles et les 3 tissus étudiés, ainsi que quelques statistiques, sont disponibles sur le Tableau 37.

	Couples machine/protocole (n=10)										Moyenne	Ecart-type	Max	Min
	1	2	3	4	5	6	7	8	9	10				
LCR	2,02	0,64	2,03	4,82	1,60	6,51	5,74	1,08	0,98	5,20	3,06	2,12	6,51	0,64
MB	2,50	2,90	3,70	2,59	3,79	1,64	0,77	3,72	3,81	0,94	2,64	1,12	3,81	0,77
MG	4,74	2,45	1,79	3,55	0,77	2,85	4,11	7,02	1,68	2,37	3,13	1,72	7,02	0,77

Tableau 37. Distances euclidiennes protocole/tissu entre chacun des 10 couples protocole/machine et les barycentres des tissus associés.

8.6 Interprétation

8.6.1 Détection d'images pathologiques et artefactées

Sur 243 images TDM de la sous-cohorte CT-Harmo étudiée ici, 116 ont été exclues par nos deux étapes de sélection de patients, soit quasiment 48%. Dans le détail, 106 ont été déclarées pathologiques, soit plus de 43 %, par l'utilisation de BLAST-CT. A l'issue de cette étape, seulement 137 images considérées non-pathologiques mais potentiellement artefactées ont été conservées. Puis, après l'étape de détection d'artefact par l'analyse de l'EIF sur les métriques radiomiques, 10 images ont été déclarées artefactées, soit plus de 7% de ces 137 images. Si la proportion de sujets pathologiques est cohérente avec l'estimation clinique des médecins en réanimation, qui était environ d'une image sur deux initialement, la proportion d'images artefactées est bien plus faible que l'estimation. En effet, en soin courant, la quasi-totalité des images sont artefactées. Cette différence peut avoir plusieurs explications, premièrement, il est probable que l'algorithme de segmentation BLAST-CT ait segmenté comme lésion une partie des artefacts, ce qui aurait eu comme conséquence de classer comme pathologiques des sujets sains mais artefactés. Deuxièmement, nous avons uniquement étudié les propriétés de l'image sur les 3 tissus du parenchyme MG/MG/LCR, donc les artefacts au niveau de la mâchoire ou des cervicales ne pourront pas être détectés par notre algorithme. Le fait que les manipulateurs radio positionnent les patients de manière à diriger les potentiels artefacts (généralement dans le plan de l'anneau contenant les émetteurs/détecteurs à rayons X) vers des zones extra cérébrales nous pousse à penser que la plupart des artefacts sur les images de soin courant ne touchent pas le parenchyme cérébral.

Cette détection automatique de sujets pathologiques et artefactés doit être améliorée. Les seuils sur les ratios de lésion arbitrairement fixés doivent être raffinés, tout comme les hyper-paramètres de l'algorithme EIF qui influent sur le nombre de valeurs aberrantes détectées. Concernant l'étape de détection des images pathologiques, depuis la fin du stage, nous avons enfin eu accès aux comptes rendu radiologiques, sur lesquels figurent toutes les informations permettant de classer de manière sûre une image en tant que pathologique ou « saine ». L'utilisation de ces données pourrait permettre d'optimiser cette étape. Concernant la détection d'artefacts, les images TDM ont désormais été relues par des étudiants en médecine afin d'établir la vérité terrain, ce qui pourrait nous permettre de tester nos algorithmes de détection de manière plus optimale et ainsi donc de les améliorer.

Ce travail est poursuivi actuellement en thèse par Vaëa Tesan sur la cohorte entière de CT-Harmo. Elle travaille notamment sur l'amélioration des réseaux de neurones de segmentations afin qu'ils soient plus robustes aux données de soin courant. D'autres approches vont également être testées afin de créer, grâce à la vérité terrain des comptes rendus, un outil automatique de triage des images TDM. Un tel outil pourrait être très utile pour extraire des sous cohortes homogènes au sein d'une cohorte plus large, ou bien pour trier de manière plus rapide les sujets pathologiques versus sain en routine clinique.

8.6.2 Étude de la variabilité inter-machines/protocoles

Au final, l'analyse de la variabilité a été faite sur 127 images considérée non-pathologique et non-artéfactées. Concernant la mesure de la variabilité sur les images déclarées comme saines et non artéfactées, nous avons pu estimer la distance moyenne inter-tissus à 12,50 et la distance moyenne des barycentres des 10 couples protocole/machine au barycentre du tissu associé à 2,94. On peut donc estimer la variabilité inter protocoles à plus de 4 fois moins que la variabilité inter tissus. Néanmoins, certains protocoles affichent une distance qui dépasse la valeur de 7 (7,02 pour le couple protocole/machine numéro 8 sur la matière grise), ce qui se rapproche de la distance moyenne entre la matière grise et la matière blanche (7,13). On pourrait donc imaginer un cas extrême où un algorithme entraîné sur des images acquises avec un certain protocole d'acquisition pourrait confondre la MG et la MB lorsqu'il serait appliqué sur des données acquises avec un autre protocole. Ces valeurs de même ordre de grandeur tendraient à confirmer l'impact de certains protocoles sur les propriétés des images TDM.

Cette étude peut néanmoins être grandement améliorée en analysant plus de données et en utilisant des métriques d'évaluation plus précises, comme par exemple la distance de Wasserstein [140] sur l'ensemble des dimensions, sans passer par une analyse en composante principale qui exclue une partie de l'information contenue dans les métriques radiomiques. Nous pourrions également améliorer la segmentation du cerveau en MG/MB/LCR, par exemple en recalant un atlas sur les images TDM. Nous pensons également que nous devrions travailler non pas à l'échelle des protocoles d'acquisition, mais au niveau des paramètres d'acquisitions. En effet, deux protocoles d'acquisition ont peut-être un nom différent mais des paramètres d'acquisitions égaux ou très proches. Ainsi nous pourrions rassembler des protocoles similaires et mieux caractériser l'influence de tel ou tel paramètre d'acquisition.

Finalement, si l'impact de la variabilité inter protocoles est confirmé, nous devons étudier les différentes manières d'harmoniser les images ou les métriques extraites des images. Plusieurs pistes sont déjà envisagées, comme une approche par réseau de neurones développée au laboratoire, ImUnity [28], pour harmoniser directement les images ou l'application de l'algorithme neuroCombat [24] pour harmoniser les métriques extraites des images. Cette étude sur la variabilité inter protocoles a de grandes limites et doit être consolidée, en analysant l'ensemble des patients de la cohorte CT-Harmo et en utilisant une méthodologie plus robuste. Vaëa Tesan, qui a mené cette étude durant son stage de M2, va poursuivre cette étude lors de son doctorat dont le but, plus large, est d'étudier la robustesse de nos algorithmes de traitement d'image aux données de soin courant, qui peuvent différer des données des cohortes de recherche par la quantité d'artefacts, la qualité des images, ou encore la présence d'anciennes lésions cérébrales qui se superposent et influent sur de nouvelles lésions causées par un TC.

Chapitre 9 :

Communications et tâches annexes

Au cours des 3 ans de thèse, j'ai certes effectué un travail scientifique, décrit dans ces pages, mais j'ai également eu l'occasion de pleinement participer aux tâches annexes de la vie d'un chercheur. Une frise temporelle regroupant quelques-unes de ces tâches est disponible Figure 71. Ce chapitre passe en revue les travaux qui soit sont du domaine extrascientifique, soit ont été effectués en partie avant le début de mon contrat doctoral, soit n'entrent pas directement dans le narratif développé dans ces pages.

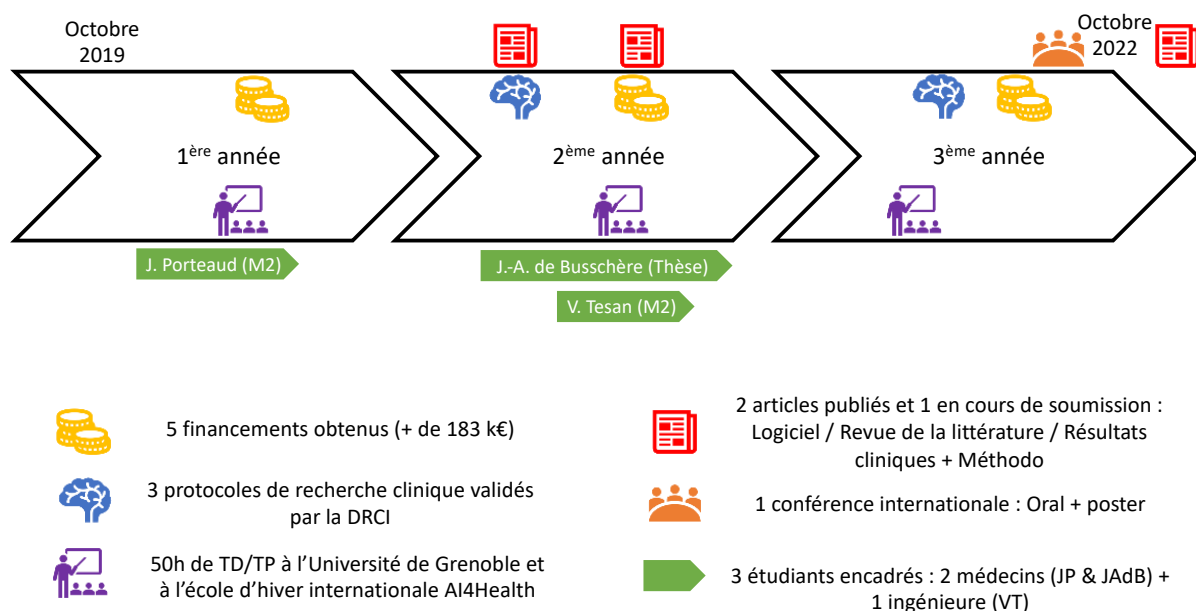


Figure 71. Frise temporelle de mes trois années de thèse.

9.1 Contributions scientifiques

J’ai pu publier 2 articles de recherche dans des journaux à comité de lecture. Un article décrivant le logiciel MP3 (Medical Software for Processing Multi-Parametric Images Pipelines) et une revue de la littérature en lien directe avec les travaux présentés dans cette thèse et qui a donc logiquement largement été reprise dans ces pages. L’article principal de ma thèse, qui rassemble les résultats des chapitres 4, 5 et 6, sera soumis dans les prochains jours.

Nos travaux ont été présentés dans des manifestations internationales. J’ai notamment pu présenter un poster sur notre étude de classification développée au chapitre 5 et une communication orale sur notre étude de segmentation développée au chapitre 4 au congrès ECR 2022 de la Société Européenne de Radiologie (ESR).

J’ai également eu l’occasion de participer au processus d’évaluation d’un article scientifique avant sa publication dans un journal scientifique à comité de lecture. Plus précisément j’ai pu participer à l’écriture du rapport d’évaluation de cet article avec mon directeur de thèse Benjamin Lemasson, qui a été initialement contacté pour faire cette évaluation.

9.2 Développement et publication du logiciel MP3 et formations données

Au cours d’un contrat d’ingénieur d’étude d’un an et 9 mois au sein de l’équipe avant le début de ma thèse, j’ai notamment co-développé le logiciel MP3, que j’ai publié au cours de ma première année de thèse [111]. Ce logiciel, développé en langage Matlab, permet de mener une étude de recherche en imagerie médicale à partir de la conversion des données brutes jusqu’au post-traitement et l’établissement des résultats de l’étude, et ce, sur de grandes cohortes, de manière robuste et facilement reproductible.

MP3 est constitué de 3 interfaces graphiques, illustrés sur la Figure 72:

- Un convertisseur capable de convertir les images médicales des formats DICOM, Bruker ou Philips au format NIfTI pour l’image et json pour les métadonnées de l’image. Ce convertisseur est également disponible en dehors du logiciel MP3 à l’adresse <https://irmage.univ-grenoble-alpes.fr/services/traitement-dimages/mri-file-manager>.
- Un visualisateur, permettant d’exploiter le format NIfTI pour facilement visualiser et comparer des images médicales acquises avec une résolution, une orientation, ou un champ de vue différents. Ce visualisateur permet également de dessiner, modifier et visualiser des régions d’intérêt sur les images, et d’organiser la base de données rassemblant toutes les images de l’étude.

- Un « pipeline manager », qui permet de facilement créer, modifier, sauvegarder, et exécuter des pipelines de traitement complexes, constitués d'un ou de plusieurs modules de traitement. Ces pipelines sont exécutables en parallèles et permettent de traiter facilement de la même manière toute une base de données. L'exécution est « intelligente » et permet de tenir compte des exécutions précédentes et des données manquantes.

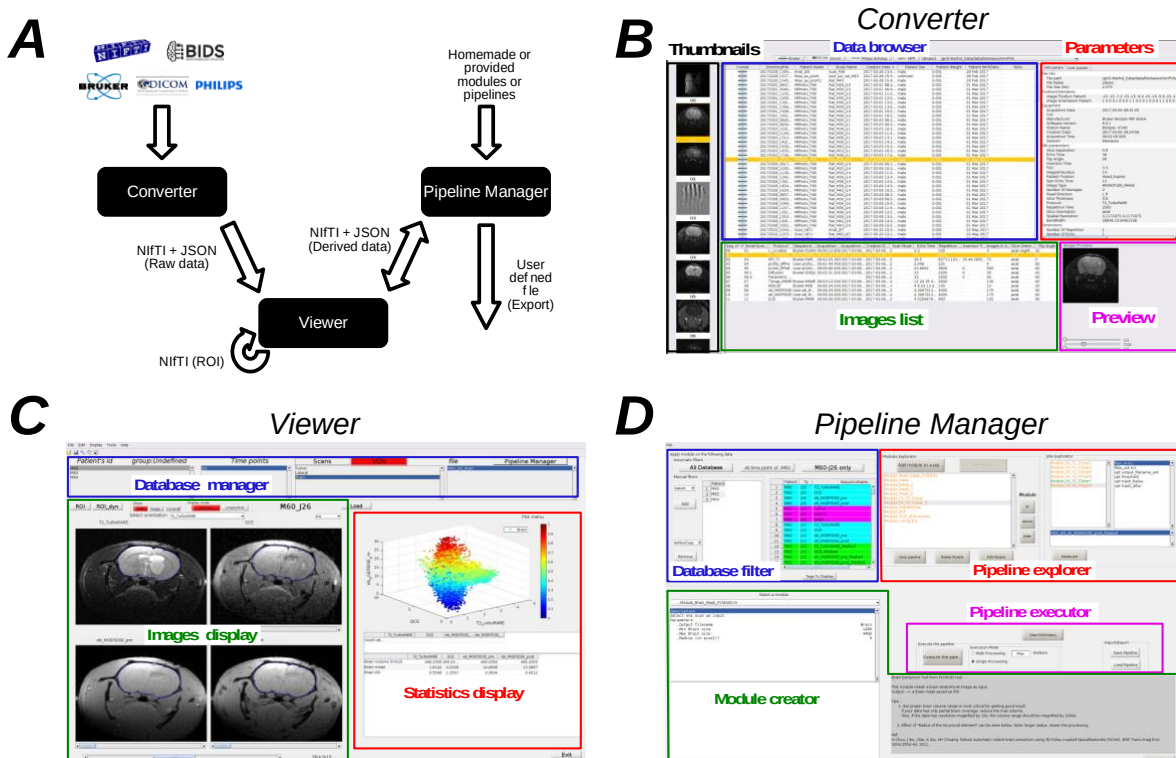


Figure 72. Vue d'ensemble des 3 interfaces graphiques du logiciel MP3 et flux de données.

Un pipeline de traitement d'images scanners sur la cohorte Radiomic-TBI est illustré sur la Figure 73. Ce pipeline comprend une étape de rééchantillonnage, un pré-traitement décrit par [141] pour extraire le cerveau à partir des images TDM, un recalage des images acquises à J1 et J3 sur l'image acquise à J0, et un calcul d'une carte radiomique représentant l'entropie, calculée par patches comme illustré dans [95].

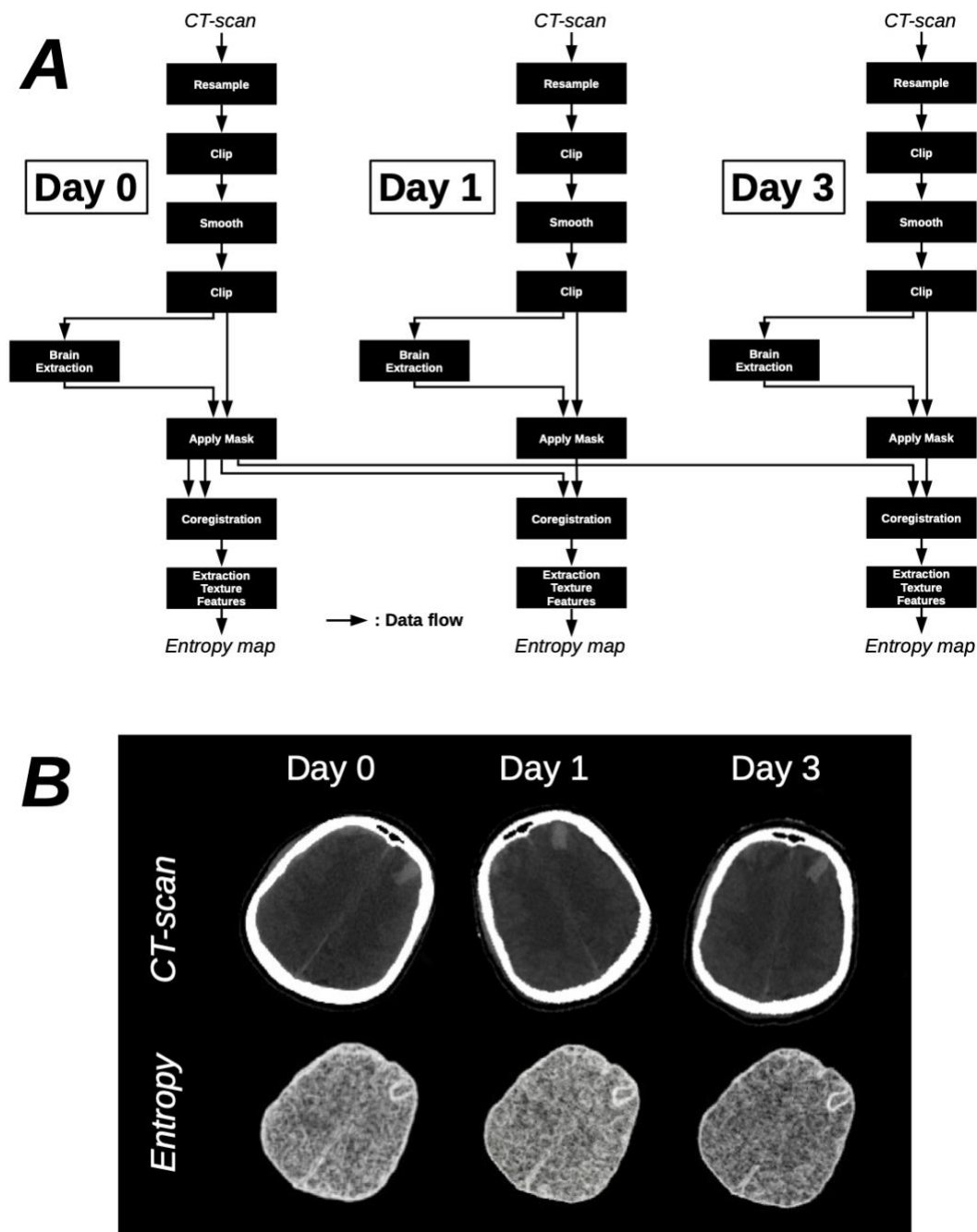


Figure 73. A : Exemple d'un pipeline complexe de traitement d'images CT. B : La ligne du haut représente les images d'entrée du pipeline et la ligne du bas les images de sortie.

MP3 est téléchargeable à l'adresse <https://github.com/nifm-gin/MP3>, et la documentation du dépôt GIT est complétée par un ensemble de 4 vidéos illustrant les principales fonctionnalités d'MP3. Ces vidéos sont disponibles à l'adresse <https://www.youtube.com/playlist?list=PL-Tj6Wc9aE9x7i6s-RLetvNE0isnEsFm7>. MP3 est utilisé quotidiennement dans notre équipe et a déjà été cité plusieurs fois par des équipes internationales. Durant ces 3 ans, j'ai participé à l'organisation de plusieurs formations. La dernière, qui a regroupé environ 30 personnes, remonte au 19 mai 2022

et a été donnée dans le cadre du réseau national « Small Animal Imaging Network » (SAIN).

9.3 Travaux Pratiques/ Travaux dirigés

Outre les formations MP3, j'ai eu l'occasion de co-animer 2 sessions de travaux pratiques (TP) à l'école d'hiver 2021 AI4Health, organisée par le Health Data Hub, qui a l'ambition de centraliser les données médicales à des fins de recherche au niveau national. Le TP, nommé « Deep learning for brain MRI: Adversarial generation, contrastive and transfer learning », avait pour but de mettre en œuvre et de comparer plusieurs méthodes d'entraînement afin de générer automatiquement une image IRM clinique pondérée T1 après injection de gadolinium à partir de l'image T1 avant injection de gadolinium. Grâce à ces méthodes nous avons ainsi généré automatiquement l'effet de l'agent de contraste sur des images IRM de patients porteurs d'une tumeur cérébrale.

Ces TP, sous forme de notebook Jupyter, et qui utilisent la base de données BRATS [142] [143] [144], sont disponibles sur demande à l'adresse <https://drive.google.com/drive/folders/1B9QdNkBHy97oOBOQRFlg9w9hrG-VOVEa>. Nous avons l'ambition de publier ces résultats mais le temps nous a jusqu'ici manqué.

J'ai également pu donner 37,5 heures de travaux dirigés (TD) en 3ème année de licence biotechnologie à l'UGA dans l'unité d'enseignement nommée « Questions d'Actualité en Biologie ». Ces TD consistaient à réfléchir par petits groupes à des questions éthiques d'actualité dans le domaine du biomédical et à organiser et restituer ces réflexions sous forme de petits exposés.

9.4 Encadrements

Durant ces 3 années, et dans le cadre de mon projet de thèse, j'ai été amené à travailler avec et à encadrer plusieurs étudiants, 2 étudiants en médecine et une étudiante ingénieure.

- Jordan Porteaud, étudiant en médecine, encadré durant son Master 2 « Neuropsychologie & Neurosciences cliniques » à l'université Grenoble Alpes de décembre 2019 à Juin 2020. Bien que cette période de stage ait été réduite à seulement 3 mois suite à la réquisition de Jordan en tant qu'interne en anesthésie-réanimation dans le cadre de la pandémie Covid-19, Jordan a pu veiller à la bonne acquisition des données de la cohorte Radiomic-TBI et a étudié la distribution de l'intensité des voxels du cerveau des premiers patients de cette cohorte, en essayant de trouver un lien visuel avec la gravité neurologique.
- Jules-Arnaud de Busschère, étudiant en médecine, durant sa thèse de médecine entre octobre 2020 et septembre 2021. Jules-Arnaud a évalué l'algorithme de segmentation automatique BLAST-CT de manière approfondie. Il a également

participé à la segmentation manuelle des lésions des cohortes Radiomic-TBI et RadioxyTC et a calculé les scores de Marshall et Rotterdam sur les images de la cohorte Radiomic-TBI.

- Vaëa Tesan, étudiante ingénieure, durant son stage de M2, entre février et juin 2021. Vaëa a travaillé sur la variabilité inter-protocoles sur les images TDM de la cohorte CT-Harmo. Son travail est résumé au chapitre 8 et son rapport de stage est disponible sur HAL [137].

9.5 Financements obtenus

Mon doctorat a été financé par la fondation des Gueules Cassées. Chaque année, nous faisons une demande de financement pour financer mon année de doctorat. Nous avons ainsi obtenu respectivement 32 000€ / 35 000€ / 34 000€ pour mes 3 années de thèse. La fondation m'a également attribué 35 000€ pour une prolongation de thèse et quelques mois de post-doctorat et a attribué au CHUGA un financement de 47 600€ pour un projet dans lequel nous sommes prestataires qui a pour but de développer un outil de prédiction de la survenue du vasospasme après un hémorragie sous-arachnoïdienne traumatique. Ce projet utilisera l'outil de segmentation automatique de 7 types de lésions post-traumatiques développé et validé aux chapitres 4 et 6.

J'ai donc au cours de ces 3 ans obtenu en mon nom propre plus de 100 000€ (2ème et 3ème année de thèse et postdoc) et participé à l'obtention d'environ 80 000€ supplémentaires (1ère année de thèse et projet vasospasme).

9.6 Protocoles recherche clinique

En partenariat avec le CHUGA, j'ai rédigé 2 protocoles d'étude de recherche rétrospective et participé à la rédaction d'un protocole d'étude de recherche prospective, tous les 3 acceptés par la DRCI, qui ont rendu possible la récupération de plus de 1000 images TDM provenant du CHUGA, ainsi que de nombreuses données cliniques associées. Ces deux cohortes sont détaillées au chapitre 3 :

- Cohorte CT-Harmo : 987 images TDM de 987 patients passés au déchocage au CHUGA. Objectif primaire : Étudier la variabilité inter-protocoles sur les images TDM. Résultats au chapitre 8.
- Cohorte RadioxyTC : 12 images TDM de 12 patients imagés au CHUGA et pour lesquels le score TILsum quotidien a été calculé les 5 premiers jours en réanimation. Objectif primaire : valider nos outils automatiques de classification et de segmentation. Résultats au chapitre 6.

Conclusion et perspectives

Les traumatismes crâniens (TC) sont fréquents et engendrent de lourdes séquelles ou même la mort chez les victimes, qui sont souvent dans la force de l'âge, ce qui a un coût économique et humain important pour les sociétés. Aujourd'hui le diagnostic du TC est basé sur des données cliniques, via des scores cliniques, et/ou sur une analyse visuelle des images tomodensitométriques (TDM), via des scores des scores d'imagerie qui quantifient de manière sommaire l'état lésionnel des patients. Les images TDM sont donc exploitées de manière qualitative et peu reproductible. Nous pensons qu'une analyse automatique et quantitative de ces images, par des outils de traitement d'image ou d'intelligence artificielle, permettrait d'améliorer le diagnostic précoce des patients TC et leur prise en charge en exploitant de manière plus fine les informations qu'elles contiennent. C'est avec cet objectif en tête que j'ai débuté ma thèse en octobre 2019. Plus précisément, nous avons comme objectif de développer des outils d'analyse d'images TDM faciles à utiliser, capables d'assister les cliniciens dans le diagnostic et pronostic à court terme des patients TC.

Ce manuscrit présente les différentes études menées dans le cadre de cette thèse, résultants des problématiques rencontrées au long de ces 3 ans. L'objectif principal de la thèse, la prédiction du futur niveau d'intensité thérapeutique que nécessite un patient en réanimation a été en partie atteint au chapitre 5, avec le développement d'un modèle de prédiction. Ce modèle a été rendu utilisable de manière automatique après le développement d'un outil de segmentation automatique présenté au chapitre 4. Ces deux outils ont été validés et mis à disposition des cliniciens au chapitre 6. Puis, l'aspect longitudinal de l'évolution lésionnelle des patients a été exploité, afin d'estimer l'existence et l'ampleur d'une poussée œdémateuse chez un patient, ce qui apparaît aujourd'hui imprévisible à partir de l'analyse à l'œil nu des images TDM (chapitre 7). Finalement, et avec en ligne de mire l'idée d'exploiter l'évolution scanographique, et non plus seulement l'évolution lésionnelle, nous avons étudié l'impact sur les images TDM des différents paramètres d'acquisitions (chapitre 8), afin de pouvoir comparer de manière pertinente les images TDM d'un même patient acquises à différents moments. Toutes ces études forment un ensemble cohérent et pertinent dans notre objectif exploratoire d'exploiter les images TDM de patients TC afin d'apporter aux cliniciens une aide quantitative au diagnostic et à la prise en charge des patients.

Les principaux résultats de ces travaux concernent la segmentation automatique et la prédiction du niveau d'intensité thérapeutique que nécessite un patient en réanimation. Concernant la segmentation, nous avons amélioré un réseau de neurones déjà publié grâce à des méthodes d'apprentissage par transfert. Nous l'avons

premièrement rendu significativement plus précis sur nos données (score de Dice = 0,63 vs 0,34). Dans un second temps, nous avons modifié l'architecture du réseau afin de segmenter 7 différentes lésions et non plus 4, par une approche d'apprentissage par transfert, ce qui nous a permis d'obtenir un réseau de segmentation automatique à 7 classes, ce qui à notre connaissance n'a jamais été fait dans la littérature scientifique. En ce qui concerne les résultats de prédiction du niveau d'intensité thérapeutique, nous avons étudié le pouvoir prédictif du volume, du type et de la localisation spatiale des lésions, ce qui nous a permis de mettre en évidence l'importance de l'hémorragie sous dure dans l'intensité de la prise en charge à court terme des patients. Nous avons ainsi pu développer un modèle de prédiction affichant un taux d'exactitude de prédiction de 83% sur un jeu de données externe à partir d'une quantification automatique des images TDM basée sur notre algorithme de segmentation automatique.

Concernant les résultats de prédiction, le plus frustrant est sans doute de ne pas pouvoir se confronter à la littérature scientifique à cause de l'absence d'études de prédiction du niveau d'intensité thérapeutique. Nous aimerions prochainement tenter de prédire un critère classique, comme la mortalité ou le GOSE à 6 mois par exemple, afin de pouvoir comparer la capacité prédictive de nos métriques extraites des images à celles des autres études, bien qu'une telle prédiction ait bien moins de valeur pour les cliniciens dans leur quotidien.

Bien que prometteur, nos résultats cliniques sont peu puissants du fait du faible nombre de patients étudiés. Ils doivent donc être validés sur des cohortes plus importantes. Cependant, le fait que nos résultats concernant l'importance de l'hémorragie sous dure, et spécialement dans les lobes externes du parenchyme, aient été confirmés par une récente étude sur plus de 600 patients, met en exergue la qualité des procédures que nous avons mises en place pour pallier le faible nombre de données, que cela soit la validation croisée imbriquée ou l'apprentissage par transfert d'un réseau de neurone. En particulier, nos travaux sur l'amélioration d'un algorithme de segmentation par apprentissage par transfert montrent qu'il est facile d'ajuster un modèle déjà entraîné à de nouvelles données à partir d'un petit nombre d'échantillons. Ces résultats vont dans le sens de modèles entraînés de manière locale et combinés afin de fonctionner de manière globale, comme décrit récemment par Ainsworth & al. [117].

Si nous n'avons pas eu le temps de mener l'étude longitudinale aussi loin que nous aurions aimé, nos résultats préliminaires sont toutefois encourageants. La méthode de segmentation par empirisme est facilement explicable, très peu chère en temps de calcul, et généralisable par itérations (on peut améliorer le dictionnaire sans repartir de 0). Si elle doit évidemment être améliorée et évaluée sur d'autres données et d'autres critères, elle pourrait être d'une grande aide pour les cliniciens en générant le profil lésionnel futur des patients TC. En ce sens, nos prochains développements vont se concentrer sur l'exploitation longitudinale. Comme développé à la fin du chapitre 7, plusieurs pistes sont à l'étude comme la génération par réseaux de neurones convolutifs de la segmentation

lésionnelle à J3 ou encore directement l'image TDM à J3. Ces deux pistes sont les objectifs de ma prolongation de quelques mois dans l'équipe en tant que post-doctorant.

Finally, pendant ces 3 ans, nous avons mis en place des outils qui seront utiles à l'équipe dans ses recherches sur l'analyse quantitative des images TDM cliniques de TC. De tels outils commencent à voir le jour dans la littérature scientifique. Les articles de revues scientifiques analysant les images TDM de patients TC par intelligence artificielle se multiplient. L'arrivée de tels outils en routine clinique n'est donc probablement qu'une question de temps. Néanmoins, il faudra davantage d'études sur la robustesse des algorithmes à la variabilité inter-machines/protocoles, aux artefacts, aux populations d'étude, en d'autres termes, les rendre robustes aux variabilités des images TDM de soin courant. Il faut pour cela traiter des cohortes de grande taille, avec bien plus de patients que les 29 de Radiomic-TBI, la principale cohorte de ce manuscrit.

Les principales limites à l'établissement et au traitement de ces cohortes résident dans les contraintes administratives, éthiques et techniques. Les procédures de récupération de données médicales doivent être facilitées, la structuration des données au sein des hôpitaux doit être épurée et centralisée au niveau local. Les outils d'anonymisation totale doivent également être améliorés, afin de limiter tout lien rétroactif avec l'identité des patients. Les capacités de stockage sécurisé des données médicales doivent être plus facilement accessibles et intégrées aux capacités de calcul, qui doivent elles aussi être optimisées, plus accessibles, et étendues.

Si ces conditions, largement accessibles, sont remplies, on peut imaginer l'arrivée en routine clinique d'outils d'aide au diagnostic et à la prise en charge bénéfique aux cliniciens et aux patients TC à partir des images TDM dans les 5 ans. Le travail mené dans le cadre de cette thèse appliquée et exploratoire a permis, grâce aux nombreuses études entreprises, d'ouvrir de nouveaux pans de recherches à explorer. Ces travaux doivent être poursuivis, dans le cadre de nouvelles thèses et à plus court terme dans le cadre de mon futur post-doctorat.



Figure 74. Image générée automatiquement par l'algorithme MidJourney [<https://www.midjourney.com/>] à partir des mots « Computed tomography traumatic brain injury artificial intelligence ».

Bibliographie

- [1] É. Bayen, C. Jourdan, P. Azouvi, J.-J. Weiss, et P. Pradat-Diehl, « Prise en charge après lésion cérébrale acquise de type traumatisme crânien », *Inf. Psychiatr.*, vol. 88, n° 5, p. 331-337, 2012, doi: 10.3917/inpsy.8805.0331.
- [2] SPF, « Épidémiologie des traumatismes crâniens en France et dans les pays occidentaux : Synthèse bibliographique, avril 2016 ». <https://www.santepubliquefrance.fr/import/epidemiologie-des-traumatismes-craniens-en-france-et-dans-les-pays-occidentaux-synthese-bibliographique-avril-2016> (consulté le 2 octobre 2022).
- [3] M. Faul, M. M. Wald, W. Rutland-Brown, E. E. Sullivent, et R. W. Sattin, « Using a cost-benefit analysis to estimate outcomes of a clinical treatment guideline: testing the Brain Trauma Foundation guidelines for the treatment of severe traumatic brain injury », *J. Trauma*, vol. 63, n° 6, p. 1271-1278, déc. 2007, doi: 10.1097/TA.0b013e3181493080.
- [4] G. Teasdale et B. Jennett, « Assessment of coma and impaired consciousness. A practical scale », *Lancet Lond. Engl.*, vol. 2, n° 7872, p. 81-84, juill. 1974, doi: 10.1016/s0140-6736(74)91639-0.
- [5] D. Khaki, V. Hietanen, A. Corell, H. O. Hergès, et J. Ljungqvist, « Selection of CT variables and prognostic models for outcome prediction in patients with traumatic brain injury », *Scand. J. Trauma Resusc. Emerg. Med.*, vol. 29, n° 1, p. 94, déc. 2021, doi: 10.1186/s13049-021-00901-6.
- [6] S. Chilamkurthy *et al.*, « Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study », *The Lancet*, vol. 392, n° 10162, p. 2388-2396, déc. 2018, doi: 10.1016/S0140-6736(18)31645-3.
- [7] H. Yao, C. Williamson, J. Gryak, et K. Najarian, « Automated hematoma segmentation and outcome prediction for patients with traumatic brain injury », *Artif. Intell. Med.*, vol. 107, p. 101910, juill. 2020, doi: 10.1016/j.artmed.2020.101910.
- [8] H. Kim *et al.*, « Quantitative analysis of computed tomography images and early detection of cerebral edema for pediatric traumatic brain injury patients: retrospective study », p. 16, 2014.
- [9] UNSCEAR, « Radiation: Effects and Sources ». Consulté le: 3 octobre 2022. [En ligne]. Disponible sur: <http://www.unscear.org/unscear/en/publications/radiation-effects-and-sources.html>
- [10] B. Van Oystaeyen, « Atténuation ». <https://www.phys4med.be/generalites-sur-les-rayons-x/absorption> (consulté le 3 octobre 2022).
- [11] « Scanner | Clinique du Landy ». <https://clinique-du-landy-saint-ouen.ramsaysante.fr/vous-etes-patient-pourquoi-choisir-notre-etablissement-tous-nos-soins/scanner-38> (consulté le 3 octobre 2022).
- [12] M. Langer, « Phase retrieval in X-ray in-line phase tomography -- developments and applications », mars 2014.
- [13] M. Dimmock, « The science of medical imaging: X-rays and CT scans », *The Conversation*. <http://theconversation.com/the-science-of-medical-imaging-x-rays-and-ct-scans-15029> (consulté le 3 octobre 2022).
- [14] I. Buvat, « Reconstruction tomographique », p. 98.
- [15] Themefisher, « SCIL », *SCIL*. <https://scil.usherbrooke.ca/> (consulté le 3 octobre 2022).

- [16] « Godfrey Hounsfield », *Wikipédia*. 4 mars 2022. Consulté le: 3 octobre 2022. [En ligne]. Disponible sur: https://fr.wikipedia.org/w/index.php?title=Godfrey_Hounsfield&oldid=191607322
- [17] « Échelle de Hounsfield », *Wikipédia*. 4 mars 2021. Consulté le: 3 octobre 2022. [En ligne]. Disponible sur: https://fr.wikipedia.org/w/index.php?title=%C3%89chelle_de_Hounsfield&oldid=180542102
- [18] E. Lin et A. Alessio, « What are the basic concepts of temporal, contrast, and spatial resolution in cardiac CT? », *J. Cardiovasc. Comput. Tomogr.*, vol. 3, n° 6, p. 403-408, 2009, doi: 10.1016/j.jcct.2009.07.003.
- [19] C. Jin *et al.*, « DC2Anet: Generating Lumbar Spine MR Images from CT Scan Data Based on Semi-Supervised Learning », *Appl. Sci.*, vol. 9, p. 2521, juin 2019, doi: 10.3390/app9122521.
- [20] D. Mackin *et al.*, « Measuring CT scanner variability of radiomics features », *Invest. Radiol.*, vol. 50, n° 11, p. 757-765, nov. 2015, doi: 10.1097/RLI.0000000000000180.
- [21] L. Lu, R. C. Ehmke, L. H. Schwartz, et B. Zhao, « Assessing Agreement between Radiomic Features Computed for Multiple CT Imaging Settings », *PLOS ONE*, vol. 11, n° 12, p. e0166550, déc. 2016, doi: 10.1371/journal.pone.0166550.
- [22] E. Husby, E. D. Svendsen, H. K. Andersen, et A. C. T. Martinsen, « 100 days with scans of the same Catphan phantom on the same CT scanner », *J. Appl. Clin. Med. Phys.*, vol. 18, n° 6, p. 224-231, nov. 2017, doi: 10.1002/acm2.12186.
- [23] T. E. Plautz, C. Zheng, G. Noid, et X. A. Li, « Time stability of delta-radiomics features and the impact on patient analysis in longitudinal CT images », *Med. Phys.*, vol. 46, n° 4, p. 1663-1676, avr. 2019, doi: 10.1002/mp.13395.
- [24] J.-P. Fortin *et al.*, « Harmonization of multi-site diffusion tensor imaging data », *NeuroImage*, vol. 161, p. 149-170, nov. 2017, doi: 10.1016/j.neuroimage.2017.08.047.
- [25] F. Orlhac, F. Frouin, C. Nioche, N. Ayache, et I. Buvat, « Validation of A Method to Compensate Multicenter Effects Affecting CT Radiomics », *Radiology*, vol. 291, n° 1, p. 53-59, janv. 2019, doi: 10.1148/radiol.2019182023.
- [26] Z. Yi, H. Zhang, P. Tan, et M. Gong, « DualGAN: Unsupervised Dual Learning for Image-to-Image Translation », *ArXiv170402510 Cs*, oct. 2018, Consulté le: 7 février 2020. [En ligne]. Disponible sur: <http://arxiv.org/abs/1704.02510>
- [27] B. E. Dewey *et al.*, « DeepHarmony: A deep learning approach to contrast harmonization across scanner changes », *Magn. Reson. Imaging*, vol. 64, p. 160-170, 2019, doi: 10.1016/j.mri.2019.05.041.
- [28] S. Cackowski, E. L. Barbier, M. Dojat, et T. Christen, « ImUnity: a generalizable VAE-GAN solution for multicenter MR image harmonization ». arXiv, 14 septembre 2021. doi: 10.48550/arXiv.2109.06756.
- [29] F. E. Boas et D. Fleischmann, « CT artifacts: causes and reduction techniques », *Imaging Med.*, vol. 4, n° 2, p. 229-240, avr. 2012.
- [30] « La radioactivité et ses effets sur l'homme ». <https://www.irs.fr/FR/connaissances/Sante/rayonnements-ionisants-effets-radioprotection-sante/effets-rayonnements-ionisants/Pages/sommaire.aspx#.Yzotii8iu9Y> (consulté le 3 octobre 2022).
- [31] CEA, « L'homme et les rayonnements », *CEA/Médiathèque*, 16 avril 2014. <https://www.cea.fr/multimedia/Pages/editions/livrets-thematiques/l-homme-et-les-rayonnements.aspx> (consulté le 13 septembre 2022).
- [32] A. I. R. Maas *et al.*, « Traumatic brain injury: integrated approaches to improve prevention, clinical care, and research », *Lancet Neurol.*, vol. 16, n° 12, p. 987-1048, déc. 2017, doi: 10.1016/S1474-4422(17)30371-X.
- [33] M. C. Dewan *et al.*, « Estimating the global incidence of traumatic brain injury », *J.*

- Neurosurg.*, vol. 130, n° 4, p. 1080-1097, avr. 2019, doi: 10.3171/2017.10.JNS17352.
- [34] W. Peeters *et al.*, « Epidemiology of traumatic brain injury in Europe », *Acta Neurochir. (Wien)*, vol. 157, n° 10, p. 1683-1696, oct. 2015, doi: 10.1007/s00701-015-2512-7.
- [35] M. Majdan *et al.*, « Epidemiology of traumatic brain injuries in Europe: a cross-sectional analysis », *Lancet Public Health*, vol. 1, n° 2, p. e76-e83, déc. 2016, doi: 10.1016/S2468-2667(16)30017-2.
- [36] M. Faul, L. Xu, M. M. Wald, et V. Coronado, « Traumatic Brain Injury in the United States », p. 74, 2010.
- [37] F. Masson *et al.*, « Epidemiology of severe brain injuries: a prospective population-based study », *J. Trauma*, vol. 51, n° 3, p. 481-489, sept. 2001, doi: 10.1097/00005373-200109000-00010.
- [38] B. Singh *et al.*, « Meta-analysis of Glasgow Coma Scale and Simplified Motor Score in predicting traumatic brain injury outcomes », *Brain Inj.*, vol. 27, n° 3, p. 293-300, mars 2013, doi: 10.3109/02699052.2012.743182.
- [39] H. Nayeibaghayee et T. Afsharian, « Correlation between Glasgow Coma Scale and brain computed tomography-scan findings in head trauma patients », *Asian J. Neurosurg.*, vol. 11, n° 1, p. 46-49, mars 2016, doi: 10.4103/1793-5482.165780.
- [40] T. Geeraerts *et al.*, « Management of severe traumatic brain injury (first 24hours) », *Anaesth. Crit. Care Pain Med.*, vol. 37, n° 2, p. 171-186, avr. 2018, doi: 10.1016/j.accpm.2017.12.001.
- [41] P. Zuercher *et al.*, « Reliability and Validity of the Therapy Intensity Level Scale: Analysis of Clinimetric Properties of a Novel Approach to Assess Management of Intracranial Pressure in Traumatic Brain Injury », *J. Neurotrauma*, vol. 33, n° 19, p. 1768-1774, oct. 2016, doi: 10.1089/neu.2015.4266.
- [42] J. H. Olver, J. L. Ponsford, et C. A. Curran, « Outcome following traumatic brain injury: a comparison between 2 and 5 years after injury », *Brain Inj.*, vol. 10, n° 11, p. 841-848, janv. 1996, doi: 10.1080/026990596123945.
- [43] B. Jennett, « ASSESSMENT OF OUTCOME AFTER SEVERE BRAIN DAMAGE A Practical Scale », *The Lancet*, vol. 305, n° 7905, p. 480-484, mars 1975, doi: 10.1016/S0140-6736(75)92830-5.
- [44] G. M. Teasdale, L. E. I. Pettigrew, J. t. L. Wilson, G. Murray, et B. Jennett, « Analyzing Outcome of Treatment of Severe Head Injury: A Review and Update on Advancing the Use of the Glasgow Outcome Scale », *J. Neurotrauma*, vol. 15, n° 8, p. 587-597, août 1998, doi: 10.1089/neu.1998.15.587.
- [45] V. Perlberg, L. Puybasset, E. Tollard, S. Lehericy, H. Benali, et D. Galanaud, « Relation between brain lesion location and clinical outcome in patients with severe traumatic brain injury: a diffusion tensor imaging study using voxel-based approaches », *Hum. Brain Mapp.*, vol. 30, n° 12, p. 3924-3933, déc. 2009, doi: 10.1002/hbm.20817.
- [46] A. Olsen, « Cognitive Control Function and Moderate-to-Severe Traumatic Brain Injury: Functional and Structural Brain Correlates », 2014.
- [47] M. Manto, « Neuroanatomie et sémiologie neurologique – Cours, QCM et exercices corrigés », p. 18, 2015.
- [48] L. Lefebvre, « La réorganisation fonctionnelle des aires cérébrales du langage et de l'organisation visuospatiale. Influence des régularités environnementales lors d'un entraînement non verbal: études en IRMf », p. 428, 2006.
- [49] M. Pitte, « EspaceSoignant.com: cours et procédures de soins en ligne ». <https://espacesoignant.com/> (consulté le 3 octobre 2022).
- [50] H. Baillet, « Approche dynamique de l'utilisation d'un cheval mécanique au service de la rééducation posturale de patients cérébrolésés », 2018.
- [51] K. E. Saatman, A.-C. Duhaime, R. Bullock, A. I. R. Maas, A. Valadka, et G. T. Manley,

- « Classification of Traumatic Brain Injury for Targeted Therapies », *J. Neurotrauma*, vol. 25, n° 7, p. 719-738, juill. 2008, doi: 10.1089/neu.2008.0586.
- [52] L. F. Marshall *et al.*, « A new classification of head injury based on computerized tomography », *J. Neurosurg.*, vol. 75, n° Supplement, p. S14-S20, nov. 1991, doi: 10.3171/sup.1991.75.1s.0s14.
- [53] A. I. R. Maas, C. W. P. M. Hukkelhoven, L. F. Marshall, et E. W. Steyerberg, « Prediction of outcome in traumatic brain injury with computed tomographic characteristics: a comparison between the computed tomographic classification and combinations of computed tomographic predictors », *Neurosurgery*, vol. 57, n° 6, p. 1173-1182; discussion 1173-1182, déc. 2005, doi: 10.1227/01.neu.0000186013.63046.6b.
- [54] D. W. Nelson *et al.*, « Extended analysis of early computed tomography scans of traumatic brain injured patients and relations to outcome », *J. Neurotrauma*, vol. 27, n° 1, p. 51-64, janv. 2010, doi: 10.1089/neu.2009.0986.
- [55] R. Raj, J. Siironen, M. B. Skrifvars, J. Hernesniemi, et R. Kivisaari, « Predicting outcome in traumatic brain injury: development of a novel computerized tomography classification system (Helsinki computerized tomography score) », *Neurosurgery*, vol. 75, n° 6, p. 632-646; discussion 646-647, déc. 2014, doi: 10.1227/NEU.0000000000000533.
- [56] E. P. Thelin *et al.*, « Evaluation of novel computerized tomography scoring systems in human traumatic brain injury: An observational, multicenter study », *PLOS Med.*, vol. 14, n° 8, p. e1002368, août 2017, doi: 10.1371/journal.pmed.1002368.
- [57] R. Pargaonkar, V. Kumar, G. Menon, et A. Hegde, « Comparative study of computed tomographic scoring systems and predictors of early mortality in severe traumatic brain injury », *J. Clin. Neurosci.*, vol. 66, p. 100-106, août 2019, doi: 10.1016/j.jocn.2019.05.011.
- [58] K. A. Chun *et al.*, « Interobserver Variability in the Assessment of CT Imaging Features of Traumatic Brain Injury », *J. Neurotrauma*, vol. 27, n° 2, p. 325-330, févr. 2010, doi: 10.1089/neu.2009.1115.
- [59] F. Pesapane, M. Codari, et F. Sardanelli, « Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine », *Eur. Radiol. Exp.*, vol. 2, n° 1, p. 35, oct. 2018, doi: 10.1186/s41747-018-0061-6.
- [60] M. Komorowski, « Artificial intelligence in intensive care: are we there yet? », *Intensive Care Med.*, vol. 45, n° 9, p. 1298-1300, sept. 2019, doi: 10.1007/s00134-019-05662-6.
- [61] S.-É. Parent, *11 Introduction à l'autoapprentissage | Analyse et modélisation d'agroécosystèmes*. Consulté le: 3 octobre 2022. [En ligne]. Disponible sur: <https://essicolo.github.io/ecologie-mathematique-R/chapitre-ml.html>
- [62] S. Raschka, « Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning », *ArXiv181112808 Cs Stat*, nov. 2020, Consulté le: 18 février 2022. [En ligne]. Disponible sur: <http://arxiv.org/abs/1811.12808>
- [63] S. Varma et R. Simon, « Bias in error estimation when using cross-validation for model selection », *BMC Bioinformatics*, vol. 7, n° 1, p. 91, déc. 2006, doi: 10.1186/1471-2105-7-91.
- [64] « Cours - Arbres de décision — Cnam — UE RCP209 ». <https://cedric.cnam.fr/vertigo/cours/ml2/coursArbresDecision.html> (consulté le 3 octobre 2022).
- [65] « Home · Wiki · Talks / Fidle · GitLab », *GitLab*. <https://gricad-gitlab.univ-grenoble-alpes.fr/talks/fidle/-/wikis/home> (consulté le 3 octobre 2022).
- [66] K. Kamnitsas *et al.*, « DeepMedic for Brain Tumor Segmentation », in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, vol. 10154, A. Crimi, B. Menze, O. Maier, M. Reyes, S. Winzeck, et H. Handels, Éd. Cham: Springer International Publishing, 2016, p. 138-149. doi: 10.1007/978-3-319-55524-9_14.
- [67] K. Kamnitsas *et al.*, « Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation », *Med. Image Anal.*, vol. 36, p. 61-78, févr. 2017, doi:

10.1016/j.media.2016.10.004.

[68] N. Tajbakhsh *et al.*, « Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning? », *IEEE Trans. Med. Imaging*, vol. 35, n° 5, p. 1299-1312, mai 2016, doi: 10.1109/TMI.2016.2535302.

[69] V. Cheplygina, M. de Bruijne, et J. P. W. Pluim, « Not-so-supervised: A survey of semi-supervised, multi-instance, and transfer learning in medical image analysis », *Med. Image Anal.*, vol. 54, p. 280-296, mai 2019, doi: 10.1016/j.media.2019.03.009.

[70] S. Takahashi *et al.*, « Fine-Tuning Approach for Segmentation of Gliomas in Brain Magnetic Resonance Images with a Machine Learning Method to Normalize Image Differences among Facilities », *Cancers*, vol. 13, n° 6, Art. n° 6, janv. 2021, doi: 10.3390/cancers13061415.

[71] M. A. Morid, A. Borjali, et G. Del Fiol, « A scoping review of transfer learning research on medical image analysis using ImageNet », *Comput. Biol. Med.*, vol. 128, p. 104115, janv. 2021, doi: 10.1016/j.combiomed.2020.104115.

[72] C. Brossard *et al.*, « Contribution of CT-Scan Analysis by Artificial Intelligence to the Clinical Care of TBI Patients », *Front. Neurol.*, vol. 12, 2021, doi: 10.3389/fneur.2021.666875.

[73] P. Perel, P. Edwards, R. Wentz, et I. Roberts, « Systematic review of prognostic models in traumatic brain injury », *BMC Med. Inform. Decis. Mak.*, vol. 6, n° 1, p. 38, nov. 2006, doi: 10.1186/1472-6947-6-38.

[74] MRC CRASH Trial Collaborators *et al.*, « Predicting outcome after traumatic brain injury: practical prognostic models based on large cohort of international patients », *BMJ*, vol. 336, n° 7641, p. 425-429, févr. 2008, doi: 10.1136/bmj.39461.643438.25.

[75] E. W. Steyerberg *et al.*, « Predicting outcome after traumatic brain injury: development and international validation of prognostic scores based on admission characteristics », *PLoS Med.*, vol. 5, n° 8, p. e165; discussion e165, août 2008, doi: 10.1371/journal.pmed.0050165.

[76] A. Marmarou *et al.*, « IMPACT Database of Traumatic Brain Injury: Design And Description », *J. Neurotrauma*, vol. 24, n° 2, p. 239-250, févr. 2007, doi: 10.1089/neu.2006.0036.

[77] A. I. R. Maas, A. Marmarou, G. D. Murray, S. G. M. Teasdale, et E. W. Steyerberg, « Prognosis and clinical trial design in traumatic brain injury: the IMPACT study », *J. Neurotrauma*, vol. 24, n° 2, p. 232-238, févr. 2007, doi: 10.1089/neu.2006.0024.

[78] B. Roozenbeek *et al.*, « Prediction of outcome after moderate and severe traumatic brain injury: External validation of the International Mission on Prognosis and Analysis of Clinical Trials (IMPACT) and Corticoid Randomisation After Significant Head injury (CRASH) prognostic models* », *Crit. Care Med.*, vol. 40, n° 5, p. 1609-1617, mai 2012, doi: 10.1097/CCM.0b013e31824519ce.

[79] A. Rovlias et S. Kotsou, « Classification and Regression Tree for Prediction of Outcome after Severe Head Injury Using Simple Clinical and Laboratory Variables », <https://home-liebertpub-com.sid2nomade-1.grenet.fr/neu,> 25 août 2004. <http://www.liebertpub.com/doi/abs/10.1089/0897715041526249> (consulté le 1 janvier 2021).

[80] G. D. Murray *et al.*, « Multivariable Prognostic Analysis in Traumatic Brain Injury: Results from The IMPACT Study », *J. Neurotrauma*, vol. 24, n° 2, p. 329-337, févr. 2007, doi: 10.1089/neu.2006.0035.

[81] H. F. Lingsma, B. Roozenbeek, E. W. Steyerberg, G. D. Murray, et A. I. Maas, « Early prognosis in traumatic brain injury: from prophecies to predictions », *Lancet Neurol.*, vol. 9, n° 5, p. 543-554, mai 2010, doi: 10.1016/S1474-4422(10)70065-X.

[82] G. W. Zhu, F. Wang, et W. G. Liu, « Classification and prediction of outcome in traumatic brain injury based on computed tomographic imaging », *J. Int. Med. Res.*, vol. 37, n° 4, p. 983-995, août 2009, doi: 10.1177/147323000903700402.

[83] K. Matsuo, H. Aihara, T. Nakai, A. Morishita, Y. Tohma, et E. Kohmura, « Machine Learning to Predict In-Hospital Morbidity and Mortality after Traumatic Brain Injury », *J.*

Neurotrauma, vol. 37, n° 1, p. 202-210, janv. 2020, doi: 10.1089/neu.2018.6276.

[84] N. S. E. Mohd Noor et H. Ibrahim, « Predicting Outcomes in Patients with Traumatic Brain Injury Using Machine Learning Models », in *Intelligent Manufacturing and Mechatronics*, Singapore, 2020, p. 12-20. doi: 10.1007/978-981-13-9539-0_2.

[85] A. T. Hale *et al.*, « Machine-learning analysis outperforms conventional statistical models and CT classification systems in predicting 6-month outcomes in pediatric patients sustaining traumatic brain injury », *Neurosurg. Focus*, vol. 45, n° 5, p. E2, nov. 2018, doi: 10.3171/2018.8.FOCUS17773.

[86] C.-S. Rau, P.-J. Kuo, P.-C. Chien, C.-Y. Huang, H.-Y. Hsieh, et C.-H. Hsieh, « Mortality prediction in patients with isolated moderate and severe traumatic brain injury using machine learning models », *PLOS ONE*, vol. 13, n° 11, p. e0207192, nov. 2018, doi: 10.1371/journal.pone.0207192.

[87] T. van der Ploeg, D. Nieboer, et E. W. Steyerberg, « Modern modeling techniques had limited external validity in predicting mortality from traumatic brain injury », *J. Clin. Epidemiol.*, vol. 78, p. 83-89, oct. 2016, doi: 10.1016/j.jclinepi.2016.03.002.

[88] B. Y. Gravesteijn *et al.*, « Machine learning algorithms performed no better than regression models for prognostication in traumatic brain injury », *J. Clin. Epidemiol.*, vol. 122, p. 95-107, juin 2020, doi: 10.1016/j.jclinepi.2020.03.005.

[89] E. Christodoulou, J. Ma, G. S. Collins, E. W. Steyerberg, J. Y. Verbakel, et B. Van Calster, « A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models », *J. Clin. Epidemiol.*, vol. 110, p. 12-22, juin 2019, doi: 10.1016/j.jclinepi.2019.02.004.

[90] T. Gong *et al.*, « Classification of CT Brain Images of Head Trauma », in *Pattern Recognition in Bioinformatics*, Berlin, Heidelberg, 2007, p. 401-408. doi: 10.1007/978-3-540-75286-8_38.

[91] P. Lambin *et al.*, « Radiomics: Extracting more information from medical images using advanced feature analysis », *Eur. J. Cancer*, vol. 48, n° 4, p. 441-446, mars 2012, doi: 10.1016/j.ejca.2011.11.036.

[92] M. R. Tomaszewski et R. J. Gillies, « The Biological Meaning of Radiomic Features », *Radiology*, p. 202553, janv. 2021, doi: 10.1148/radiol.2021202553.

[93] R. J. Gillies, P. E. Kinahan, et H. Hricak, « Radiomics: Images Are More than Pictures, They Are Data », *Radiology*, vol. 278, n° 2, p. 563-577, févr. 2016, doi: 10.1148/radiol.2015151169.

[94] E. d l Rosa, D. M. Sima, T. V. Vyvere, J. S. Kirschke, et B. Menze, « A Radiomics Approach to Traumatic Brain Injury Prediction in CT Scans », in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, avr. 2019, p. 732-735. doi: 10.1109/ISBI.2019.8759229.

[95] J. Muschelli, E. M. Sweeney, N. L. Ullman, P. Vespa, D. F. Hanley, et C. M. Crainiceanu, « PIcHPERFeCT: Primary Intracranial Hemorrhage Probability Estimation using Random Forests on CT », *NeuroImage Clin.*, vol. 14, p. 379-390, 2017, doi: 10.1016/j.nicl.2017.02.007.

[96] M. R. Arbabshirani *et al.*, « Advanced machine learning in action: identification of intracranial hemorrhage on computed tomography scans of the head with clinical workflow integration », *Npj Digit. Med.*, vol. 1, n° 1, p. 1-7, avr. 2018, doi: 10.1038/s41746-017-0015-z.

[97] M. Kim *et al.*, « Deep Learning in Medical Imaging », *Neurospine*, vol. 16, n° 4, p. 657-668, déc. 2019, doi: 10.14245/ns.1938396.198.

[98] A. A. Taha et A. Hanbury, « Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool », *BMC Med. Imaging*, vol. 15, n° 1, p. 29, août 2015, doi: 10.1186/s12880-015-0068-x.

[99] S. Jain *et al.*, « Automatic Quantification of Computed Tomography Features in Acute

- Traumatic Brain Injury », *J. Neurotrauma*, vol. 36, n° 11, p. 1794-1803, juin 2019, doi: 10.1089/neu.2018.6183.
- [100] W. Kuo, C. Häne, E. Yuh, P. Mukherjee, et J. Malik, « PatchFCN for Intracranial Hemorrhage Detection », *ArXiv180603265 Cs*, avr. 2019, Consulté le: 10 janvier 2020. [En ligne]. Disponible sur: <http://arxiv.org/abs/1806.03265>
- [101] W. Kuo, C. Häne, P. Mukherjee, J. Malik, et E. L. Yuh, « Expert-level detection of acute intracranial hemorrhage on head computed tomography using deep learning », *Proc. Natl. Acad. Sci.*, vol. 116, n° 45, p. 22737-22745, nov. 2019, doi: 10.1073/pnas.1908021116.
- [102] M. Monteiro *et al.*, « Multiclass semantic segmentation and quantification of traumatic brain injury lesions on head CT using deep learning: an algorithm development and multicentre validation study », *Lancet Digit. Health*, vol. 2, n° 6, p. e314-e322, juin 2020, doi: 10.1016/S2589-7500(20)30085-6.
- [103] P. Inkeaw *et al.*, « Automatic hemorrhage segmentation on head CT scan for traumatic brain injury using 3D deep learning model », *Comput. Biol. Med.*, vol. 146, p. 105530, juill. 2022, doi: 10.1016/j.compbiomed.2022.105530.
- [104] M. Rosnati *et al.*, « Automatic lesion analysis for increased efficiency in outcome prediction of traumatic brain injury », *arXiv*, arXiv:2208.04114, août 2022. doi: 10.48550/arXiv.2208.04114.
- [105] S. Jadon *et al.*, « A comparative study of 2D image segmentation algorithms for traumatic brain lesions using CT data from the ProTECTIII multicenter clinical trial », *Med. Imaging 2020 Imaging Inform. Healthc. Res. Appl.*, p. 48, mars 2020, doi: 10.1117/12.2566332.
- [106] C. Brossard *et al.*, « Prediction of therapeutic intensity from automatic multiclass segmentation of TBI lesions on CT-scans ». [En ligne]. Disponible sur: To be submitted. Provided on demand.
- [107] M. Kain *et al.*, « Small Animal Shanoir (SAS) A Cloud-Based Solution for Managing Preclinical MR Brain Imaging Studies », *Front. Neuroinformatics*, vol. 14, mai 2020, doi: 10.3389/fninf.2020.00020.
- [108] F. Vaassen *et al.*, « Evaluation of measures for assessing time-saving of automatic organ-at-risk segmentation in radiotherapy », *Phys. Imaging Radiat. Oncol.*, vol. 13, p. 1-6, janv. 2020, doi: 10.1016/j.phro.2019.12.001.
- [109] A. Koç, Ö. S. Sezgin, et S. Kayıpmaz, « Comparing different planimetric methods on volumetric estimations by using cone beam computed tomography », *Radiol. Med. (Torino)*, vol. 125, n° 4, p. 398-405, avr. 2020, doi: 10.1007/s11547-019-01131-8.
- [110] J.-A. de Busschère, « EVALUATION DE L'EXACTITUDE D'UNE APPROCHE DE SEGMENTATION AUTOMATIQUE PAR APPRENTISSAGE PROFOND DES LESIONS TOMODENSITOMÉTRIQUES DU TRAUMATISÉ CRANIEN », 2021. [En ligne]. Disponible sur: Disponible sur demande.
- [111] C. Brossard *et al.*, « MP3: Medical Software for Processing Multi-Parametric Images Pipelines », *Front. Neuroinformatics*, vol. 14, 2020, doi: 10.3389/fninf.2020.594799.
- [112] K. He, X. Zhang, S. Ren, et J. Sun, « Deep Residual Learning for Image Recognition », 2016, p. 770-778. Consulté le: 23 septembre 2022. [En ligne]. Disponible sur: https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html
- [113] K. He, X. Zhang, S. Ren, et J. Sun, « Identity Mappings in Deep Residual Networks », in *Computer Vision – ECCV 2016*, Cham, 2016, p. 630-645. doi: 10.1007/978-3-319-46493-0_38.
- [114] S. Ioffe et C. Szegedy, « Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift », in *Proceedings of the 32nd International Conference on Machine Learning*, juin 2015, p. 448-456. Consulté le: 23 septembre 2022. [En ligne]. Disponible sur: <https://proceedings.mlr.press/v37/ioffe15.html>

- [115] M. Consortium, « MONAI: Medical Open Network for AI ». Zenodo, 25 novembre 2021. doi: 10.5281/zenodo.5728262.
- [116] F. Wilcoxon, « Individual Comparisons by Ranking Methods », *Biom. Bull.*, vol. 1, n° 6, p. 80-83, 1945, doi: 10.2307/3001968.
- [117] S. K. Ainsworth, J. Hayase, et S. Srinivasa, « Git Re-Basin: Merging Models modulo Permutation Symmetries ». arXiv, 11 septembre 2022. Consulté le: 22 septembre 2022. [En ligne]. Disponible sur: <http://arxiv.org/abs/2209.04836>
- [118] A. Carass *et al.*, « Longitudinal Multiple Sclerosis Lesion Segmentation: Resource & Challenge », *NeuroImage*, vol. 148, p. 77-102, mars 2017, doi: 10.1016/j.neuroimage.2016.12.064.
- [119] TBI-IMPACT, « THERAPY INTENSITY LEVEL », 2010. http://www.tbi-impact.org/cde/mod_templates/T_TIL.9.1.pdf
- [120] A. I. R. Maas *et al.*, « Standardizing data collection in traumatic brain injury », *J. Neurotrauma*, vol. 28, n° 2, p. 177-187, févr. 2011, doi: 10.1089/neu.2010.1617.
- [121] M. Najm *et al.*, « Automated brain extraction from head CT and CTA images using convex optimization with shape propagation », *Comput. Methods Programs Biomed.*, vol. 176, p. 1-8, juill. 2019, doi: 10.1016/j.cmpb.2019.04.030.
- [122] D. Rajashekar *et al.*, « High-resolution T2-FLAIR and non-contrast CT brain atlas of the elderly », *Sci. Data*, vol. 7, n° 1, Art. n° 1, févr. 2020, doi: 10.1038/s41597-020-0379-9.
- [123] M. Jenkinson, C. F. Beckmann, T. E. J. Behrens, M. W. Woolrich, et S. M. Smith, « FSL », *NeuroImage*, vol. 62, n° 2, p. 782-790, août 2012, doi: 10.1016/j.neuroimage.2011.09.015.
- [124] B. B. Avants, N. J. Tustison, G. Song, P. A. Cook, A. Klein, et J. C. Gee, « A reproducible evaluation of ANTs similarity metric performance in brain image registration », *NeuroImage*, vol. 54, n° 3, p. 2033-2044, févr. 2011, doi: 10.1016/j.neuroimage.2010.09.025.
- [125] R. Leenings *et al.*, « PHOTONAI—A Python API for rapid machine learning model development », *PLOS ONE*, vol. 16, n° 7, p. e0254062, juill. 2021, doi: 10.1371/journal.pone.0254062.
- [126] L. Breiman, *Classification And Regression Trees*. New York: Routledge, 2017. doi: 10.1201/9781315139470.
- [127] E. Scornet, « Trees, forests, and impurity-based variable importance », novembre 2020. Consulté le: 24 août 2022. [En ligne]. Disponible sur: <https://hal.archives-ouvertes.fr/hal-02436169>
- [128] F. Al-Mufti et S. A. Mayer, « Neurocritical Care of Acute Subdural Hemorrhage », *Neurosurg. Clin. N. Am.*, vol. 28, n° 2, p. 267-278, avr. 2017, doi: 10.1016/j.nec.2016.11.009.
- [129] M. Aromatario *et al.*, « Traumatic Epidural and Subdural Hematoma: Epidemiology, Outcome, and Dating », *Medicina (Mex.)*, vol. 57, n° 2, Art. n° 2, févr. 2021, doi: 10.3390/medicina57020125.
- [130] K. Haselsberger, R. Pucher, et L. M. Auer, « Prognosis after acute subdural or epidural haemorrhage », *Acta Neurochir. (Wien)*, vol. 90, n° 3, p. 111-116, sept. 1988, doi: 10.1007/BF01560563.
- [131] T. Glatard *et al.*, « A Virtual Imaging Platform for Multi-Modality Medical Image Simulation », *IEEE Trans. Med. Imaging*, vol. 32, n° 1, p. 110-118, janv. 2013, doi: 10.1109/TMI.2012.2220154.
- [132] T. Glatard *et al.*, « Boutiques: a flexible framework to integrate command-line applications in computing platforms », *GigaScience*, vol. 7, n° 5, p. giy016, mai 2018, doi: 10.1093/gigascience/giy016.
- [133] J. J. M. van Griethuysen *et al.*, « Computational Radiomics System to Decode the Radiographic Phenotype », *Cancer Res.*, vol. 77, n° 21, p. e104-e107, 01 2017, doi: 10.1158/0008-5472.CAN-17-0339.

-
- [134] M. Giacalone *et al.*, « Local spatio-temporal encoding of raw perfusion MRI for the prediction of final lesion in stroke », *Med. Image Anal.*, vol. 50, p. 117-126, déc. 2018, doi: 10.1016/j.media.2018.08.008.
- [135] C. J. Preetha *et al.*, « Deep-learning-based synthesis of post-contrast T1-weighted MRI for tumour response assessment in neuro-oncology: a multicentre, retrospective cohort study », *Lancet Digit. Health*, vol. 3, n° 12, p. e784-e794, déc. 2021, doi: 10.1016/S2589-7500(21)00205-3.
- [136] V. M. Ramírez, V. Kmetzsch, F. Forbes, et M. Dojat, « Deep Learning Models to Study the Early Stages of Parkinson's Disease », in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, avr. 2020, p. 1534-1537. doi: 10.1109/ISBI45749.2020.9098529.
- [137] V. Tesan, C. Brossard, et B. Lemasson, « Automatic evaluation of inter-machine and interprotocol variability introduced on healthy brain CT images. », sept. 2022. [En ligne]. Disponible sur: hal-03788569
- [138] K. A. Cauley, J. Och, P. J. Yorks, et S. W. Fielden, « Automated Segmentation of Head Computed Tomography Images Using FSL », *J. Comput. Assist. Tomogr.*, vol. 42, n° 1, p. 104-110, févr. 2018, doi: 10.1097/RCT.0000000000000660.
- [139] S. Hariri, M. C. Kind, et R. J. Brunner, « Extended Isolation Forest », *IEEE Trans. Knowl. Data Eng.*, vol. 33, n° 4, p. 1479-1489, avr. 2021, doi: 10.1109/TKDE.2019.2947676.
- [140] L. V. Kantorovich, « Mathematical Methods of Organizing and Planning Production », *Manag. Sci.*, juill. 1960, doi: 10.1287/mnsc.6.4.366.
- [141] J. Muschelli, N. L. Ullman, W. A. Mould, P. Vespa, D. F. Hanley, et C. M. Crainiceanu, « Validated automatic brain extraction of head CT images », *NeuroImage*, vol. 114, p. 379-385, juill. 2015, doi: 10.1016/j.neuroimage.2015.03.074.
- [142] B. H. Menze *et al.*, « The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS) », *IEEE Trans. Med. Imaging*, vol. 34, n° 10, p. 1993-2024, oct. 2015, doi: 10.1109/TMI.2014.2377694.
- [143] S. Bakas *et al.*, « Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features », *Sci. Data*, vol. 4, p. 170117, sept. 2017, doi: 10.1038/sdata.2017.117.
- [144] S. Bakas *et al.*, « Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge », nov. 2018, doi: 10.48550/arXiv.1811.02629.
- [145] E. A. AlBadawy, A. Saha, et M. A. Mazurowski, « Deep learning for segmentation of brain tumors: Impact of cross-institutional training and testing », *Med. Phys.*, vol. 45, n° 3, p. 1150-1158, 2018, doi: 10.1002/mp.12752.
- [146] M. R. Liechti *et al.*, « Manual prostate cancer segmentation in MRI: interreader agreement and volumetric correlation with transperineal template core needle biopsy », *Eur. Radiol.*, vol. 30, n° 9, p. 4806-4815, sept. 2020, doi: 10.1007/s00330-020-06786-w.
- [147] O. U. Aydin *et al.*, « On the usage of average Hausdorff distance for segmentation performance assessment: hidden error when used for ranking », *Eur. Radiol. Exp.*, vol. 5, p. 4, janv. 2021, doi: 10.1186/s41747-020-00200-2.

Annexes

Évaluation de la segmentation automatique : Distance de Hausdorff moyenne équilibrée

De la même manière que nous avons présenté les résultats de comparaison de nos segmentations automatiques par rapport à la segmentation manuelle dans les chapitres 4 et 6 grâce au score de Dice, voici les mêmes tableaux et histogrammes pour la distance de Hausdorff moyenne équilibrée, pour le jeu de test de la cohorte Radiomic-TBI et pour la cohorte RadioxyTC.

Définitions

Soient 2 segmentations G et S contenant respectivement NG et NS voxels et $dh(G \rightarrow S)$ la distance de Hausdorff dirigée moyenne de G vers S . On définit la distance de Hausdorff moyenne AVD comme :

$$AVD = \frac{\frac{dh(G \rightarrow S)}{NG} + \frac{dh(S \rightarrow G)}{NS}}{2}$$

Cette distance permet de mesurer la similarité entre les formes des deux segmentations et est habituellement utilisée pour évaluer des segmentations médicales [145] [146]. Une étude de Aydin & al [147] a montré le biais inhérent à cette métrique pour le classement de méthodes de segmentations. Pour pallier ce biais, les auteurs ont introduit la distance de Hausdorff moyenne équilibrée, définie comme :

$$bAVD = \frac{\frac{dh(G \rightarrow S)}{NG} + \frac{dh(S \rightarrow G)}{NG}}{2}$$

Evaluation sur le jeu de test de Radiomic-TBI

	Lésion globale	HIP	HEA	O	HIV
BLAST-CT	5,11(1,90)	14,46 (8,69)	8,20 (2,81)	8,11 (3,05)	62,00 (47,00)
Random4C	58,83 (32,66)	76,29 (53,40)	111,46 (49,84)	92,33 (79,46)	9,32 (6,22)
FT_4C	7,40(3,51)	9,04 (5,23)	577,46 (542,81)	19,70 (16,00)	3,26 (0,89)

Tableau 38. Moyenne et erreur standard des distances de Hausdorff moyennes équilibrées calculées entre les segmentations automatiques à 4 classes et les segmentations manuelles sur le jeu de test de la cohorte Radiomic-TBI.

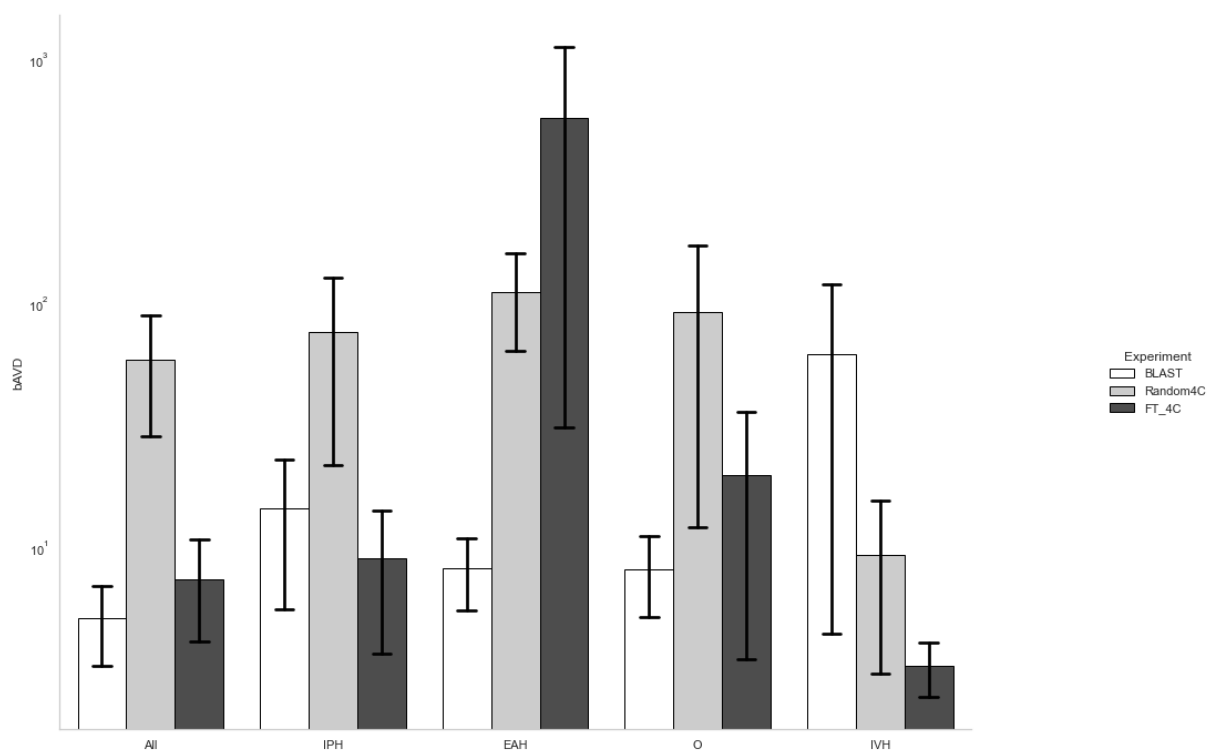


Figure 75. Moyenne et erreur standard des distances de Hausdorff moyennes équilibrées calculées entre les segmentations automatiques à 4 classes et les segmentations manuelles sur le jeu de test de la cohorte Radiomic-TBI. Échelle logarithmique.

	Lésion globale	HIP	HSD	HED	HIV	HSA	P	O
Random_7C	64,22 (46,02)	6,65 (2,40)	143,09 (122,76)	2,69 (1,99)	8,58 (5,21)	72,75 (39,71)	17,03 (6,27)	45,39 (37,96)
TL_7C	6,69 (3,41)	4,13 (2,09)	51,92 (42,56)	4,06 (2,67)	7,02 (4,71)	84,81 (41,83)	33,29 (12,96)	7,55 (4,75)
TL_RE_7C	9,31 (4,83)	2,15 (0,67)	72,59 (59,83)	4,15 (2,92)	4,51 (2,37)	105,71 (47,72)	40,18 (18,23)	4,38 (3,07)

Tableau 39. Moyenne et erreur standard des distances de Hausdorff moyennes équilibrées calculées entre les segmentations automatiques à 7 classes et les segmentations manuelles sur la cohorte Radiomic-TBI.

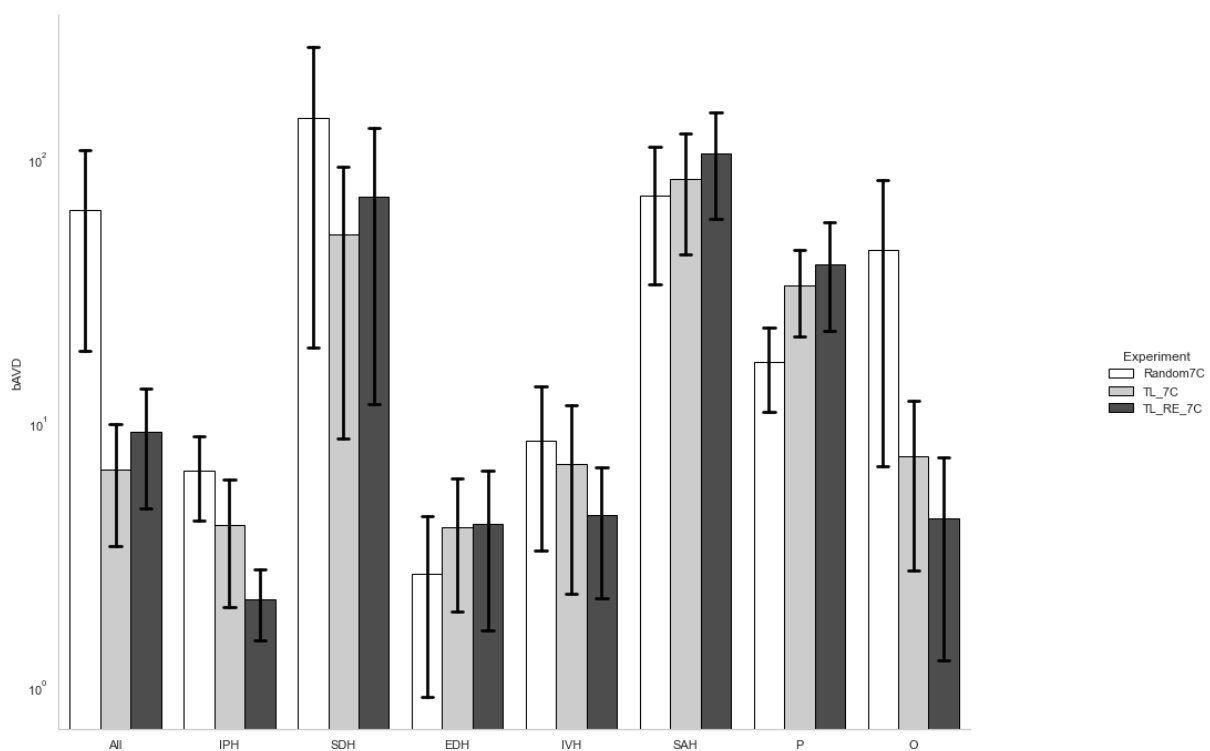


Figure 76. Moyenne et erreur standard des distances de Hausdorff moyennes équilibrées calculées entre les segmentations automatiques à 7 classes et les segmentations manuelles sur la cohorte Radiomic-TBI. Echelle logarithmique.

Evaluation sur RadioxyTC

	Lésion globale	HIP	HEA	O	HIV
BLAST-CT	4,80 (1,87)	45,77 (27,38)	5,50 (1,62)	5,95 (3,13)	9,64 (5,64)
Random4C	451,25 (236,07)	1879,20 (1374,43)	78,23 (33,73)	2111,04 (1944,83)	29,17 (20,57)
FT_4C	111,14 (62,02)	256,89 (149,02)	25,59 (10,45)	1020,00 (854,80)	13,45 (10,73)

Tableau 40. Moyenne et erreur standard des distances de Hausdorff moyennes équilibrées calculées entre les segmentations automatiques à 4 classes et les segmentations manuelles sur le jeu de test de la cohorte RadioxyTC.

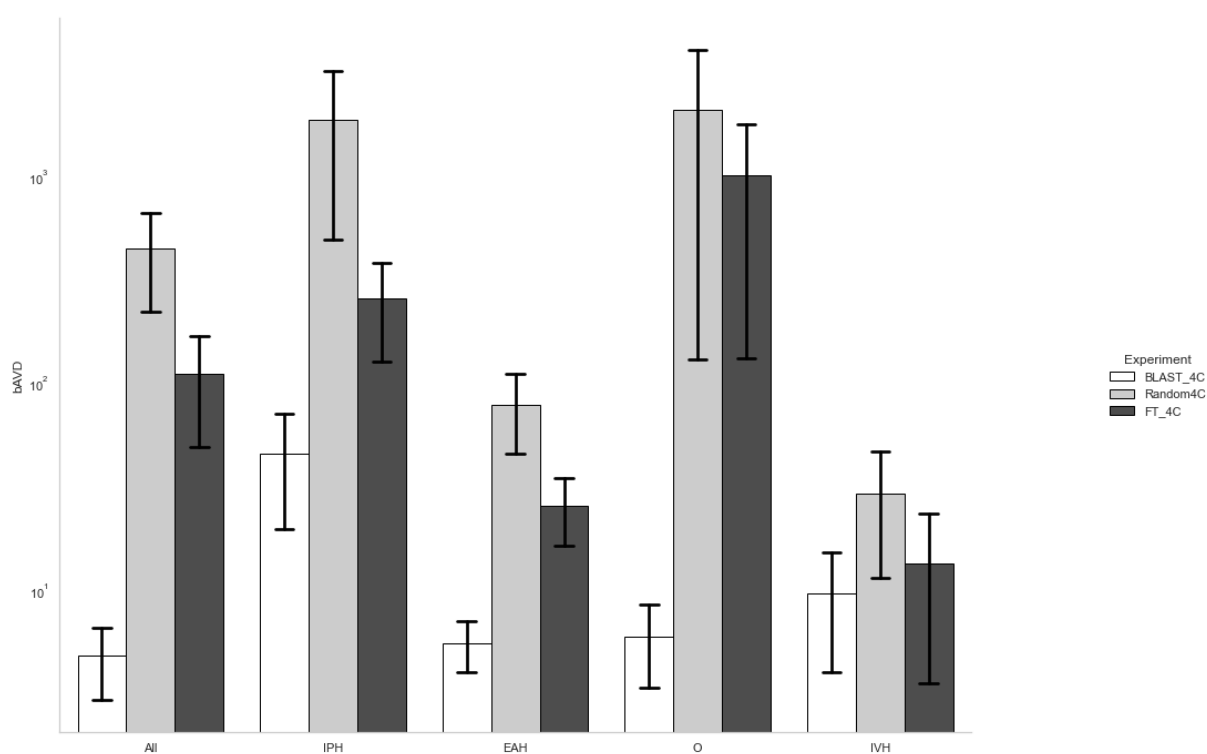


Figure 77. Moyenne et erreur standard des distances de Hausdorff moyennes équilibrées calculées entre les segmentations automatiques à 4 classes et les segmentations manuelles sur le jeu de test de la cohorte RadioxyTC. Echelle logarithmique.

	Lésion globale	HIP	HSD	HED	HIV	HSA	P	O
Random_7C	561,82 (363,92)	945,32 (342,13)	451,66 (440,63)	19,93 (0)	37,27 (32,26)	15,48 (3,57)	256,15 (247,33)	2490,23 (2295,77)
TL_7C	88,32 (44,00)	120,81 (92,80)	348,65 (343,38)	4,33 (0)	17,02 (14,30)	15,64 (3,73)	203,04 (190,66)	431,82 (349,39)
TL_RE_7C	70,25 (34,18)	127,29 (110,94)	425,18 (418,15)	56,02 (46,14)	16,98 (14,96)	19,96 (5,88)	275,62 (258,09)	429,35 (342,89)

Tableau 41. Moyenne et erreur standard des distances de Hausdorff moyennes équilibrées calculées entre les segmentations automatiques à 7 classes et les segmentations manuelles sur la cohorte RadioxyTC.

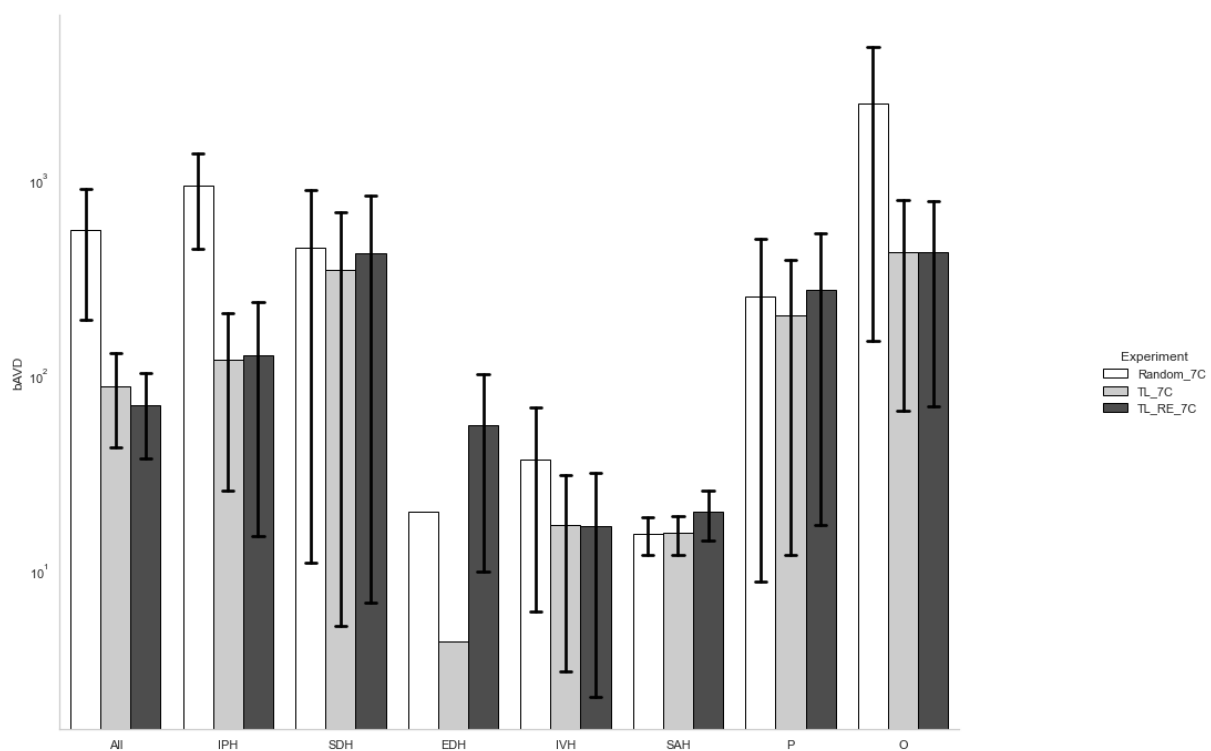


Figure 78. Moyenne et erreur standard des distances de Hausdorff moyennes équilibrées calculées entre les segmentations automatiques à 7 classes et les segmentations manuelles sur la cohorte RadioxyTC. Echelle logarithmique.

Résumé :

Le traumatisme crânien (TC) est une pathologie fréquente et grave, responsable d'une morbi-mortalité importante. Il est estimé que la moitié de la population mondiale va être victime d'un TC au moins une fois dans sa vie. Le TC peut être considéré comme un ensemble complexe d'entités nosologiques d'évolutions différentes qui conditionnent la prise en charge des patients. Le diagnostic de sévérité et l'évaluation pronostique des TC repose aujourd'hui sur le score de Glasgow initial, introduit en 1974, séparant les patients en TC grave, modéré ou léger, et sur le suivi qualitatif des lésions par imagerie tomodensitométrique (TDM). Ce suivi semi quantitatif ne permet pas de distinguer initialement les différents profils d'évolutions lésionnelles, et la prise en charge repose donc sur les symptômes induits par les lésions (hypertension intra-crânienne, hydrocéphalie, ...). L'objectif principal de cette thèse de sciences est d'exploiter quantitativement les informations contenues dans les examens TDM précoces des patients victimes de TC grave afin de prédire le niveau d'intensité thérapeutique nécessitée par ces patients au cours des 7 jours suivant l'agression initiale. Cette étude a permis, entre autres, d'améliorer un algorithme de segmentation automatique des lésions typiques du TC grâce à une approche d'apprentissage par transfert, de développer un outil de quantification du profil lésionnel d'un patient à partir de son image TDM à l'admission au CHU (J0) et de prédire le niveau d'intensité thérapeutique de sa prise en charge durant sa première semaine en réanimation. De plus, nous avons étudié l'évolution scanographique des patients en exploitant les images TDM répétées lors des premiers jours en réanimation (J0, J1 et J3). Finalement, une attention particulière a été portée à la mise à disposition de ces outils au personnel médical afin de faciliter leur utilisation en routine clinique. Les résultats sont prometteurs et laissent penser que l'analyse automatique des images TDM par intelligence artificielle fera partie intégrante de la prise en charge des TC dans un futur proche.

Summary:

Traumatic Brain Injury (TBI) is a frequent and serious pathology, responsible for a significant morbi-mortality. It is estimated that half of the world's population will suffer a head injury at least once in their lifetime. TBI can be considered as a complex group of nosological entities with different evolutions which condition the management of patients. The diagnosis of severity and prognostic evaluation of TBI is still based on the initial Glasgow score, introduced in 1974, which separates patients in severe, moderate or mild TBI, and on the qualitative follow-up of lesions by Computed Tomography (CT). This semi-quantitative diagnosis does not allow to distinguish early the different profiles of lesion evolution, and the management is then based on the symptoms induced by the lesions (intracranial hypertension, hydrocephalus, ...). The main objective of this thesis is to quantitatively exploit the information contained in the early CT examinations of severe TBI patients in order to predict the level therapeutic intensity required by these patients during the 7 days after the initial aggression. This study allowed, among other things, to improve an automatic segmentation algorithm of TBI lesions on CT images thanks to a transfer learning approach, to develop a tool to quantify the lesion profile of a patient from his CT image at admission to the University Hospital (D0) and to predict the level of therapeutic intensity of his management during his first week in the ICU. Moreover, we studied the scanographic evolution of the patients by exploiting the repeated CT images during the first days in the ICU (D0, D1 and D3). Finally, particular attention was paid to the availability of these tools to medical staff in order to facilitate their use in clinical routine. The results are promising and suggest that automatic analysis of CT images by artificial intelligence will be an integral part of TBI management in the near future.