République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Abou Bakr Belkaid – Tlemcen
Faculté de Technologie
Département de Génie biomédical

Mémoire de fin d'étude
Pour l'obtention du diplôme de Master

Filière : Génie Biomédical
Spécialité : Imagerie Médicale

Présenté par :

HADDAM Ferdaousse & BENAMARA Fatima Zahra

Thème

Segmentation de lésions cérébrales en IRM par architectures Deep Learning hybrides CNN–Transformers

Soutenu publiquement, le 24/06/2026, devant le jury composé de :

Dr.TAOULI Sidi Ahmed	MCA	Université de Tlemcen	Président
Dr. BENSMAIL Ilham	MCB	Université de Tlemcen	Examinatrice
Pr DJEBBARI Abdelghani	Pr	Université de Tlemcen	Encadrant

Année universitaire : 2025 – 2026

Dédicaces

À tous ceux dont la présence a éclairé notre chemin,
dont les mots ont apaisé nos doutes
et dont l'amour a été notre plus grande force...

Nous dédions ce travail.

À nos chers parents,

À ceux qui ont tout donné sans jamais rien demander en retour,
qui ont veillé tard pour que nous puissions rêver haut,
dont les sacrifices silencieux ont tracé la route de notre réussite.
Aucun mot ne saurait exprimer la profondeur de notre gratitude
et de l'amour que nous vous portons.

À toute notre famille,

À chaque membre qui, de près ou de loin, a cru en nous,
dont la chaleur et l'affection ont été notre refuge
dans les moments les plus difficiles de ce parcours.

À nos enseignants et encadrants,

À ceux qui ont transmis leur savoir avec passion et bienveillance,
qui ont guidé nos pas dans le monde de la connaissance
et allumé en nous la flamme de la curiosité scientifique.
Votre empreinte se retrouve dans chacune de ces pages.

À nos amis,

À ceux dont les rires ont allégé les journées interminables,
dont la présence a rendu ce parcours moins solitaire
et dont l'amitié restera l'un de nos plus beaux souvenirs.

« Le succès n'est pas la clé du bonheur. Le bonheur est la clé du succès.
Si vous aimez ce que vous faites, vous réussirez. » — Albert Schweitzer —

Ferdaousse & Fatima Zahra

Remerciements

En ce moment d'aboutissement et de grâce, nous rendons grâce à Dieu Tout-Puissant, dont la miséricorde infinie a illuminé chacune de nos étapes et guidé nos pas tout au long de ce parcours académique. C'est par Sa volonté que nous avons trouvé la force, la persévérance et la clarté nécessaires pour mener ce travail à son terme.

Nous exprimons notre profonde et sincère gratitude au Professeur **DJEBBARI Abdelghani**, notre encadrant, pour l'honneur qu'il nous a accordé en acceptant de diriger ce travail de fin d'études. Sa rigueur scientifique, sa disponibilité constante et la finesse de ses orientations ont été des piliers essentiels dans l'élaboration de ce projet. Sa confiance en nos capacités nous a portés bien au-delà de ce que nous aurions cru possible, et son exigence bienveillante a forgé en nous une véritable culture de la recherche.

Que ce travail soit le modeste témoignage de notre reconnaissance pour tout ce que vous avez investi de votre temps, de votre savoir et de votre humanité en notre faveur. Puisse Dieu vous accorder santé, succès et reconnaissance à la hauteur de votre dévouement.

Nous adressons nos vifs remerciements et notre profond respect aux membres du jury qui ont bien voulu accorder de leur précieux temps à l'évaluation de ce travail :

Dr. TAOULI Sidi Ahmed, Président du jury, dont la présence honorable et l'autorité académique confèrent à cette soutenance toute sa solennité. Nous le remercions sincèrement pour l'intérêt qu'il a bien voulu porter à notre travail.

Dr. BENSMAIL Ilham, Examinatrice de ce projet de fin d'études, pour la qualité et la perspicacité de son regard critique sur notre travail. Ses observations éclairées ont contribué à en élever la valeur, et nous lui en sommes profondément reconnaissantes.

Nos remerciements les plus chaleureux vont également à nos familles, dont le soutien indéfectible, les sacrifices silencieux et l'amour inconditionnel ont constitué notre ancre dans les moments de doute et notre élan dans les instants d'effort. Chaque page de ce mémoire porte en filigrane votre empreinte et votre confiance.

Enfin, nous exprimons notre gratitude à l'ensemble du corps enseignant du département de génie biomédical, ainsi qu'à tous ceux qui, de près ou de loin, ont contribué à notre formation et à l'accomplissement de ce travail.

« La connaissance est une lumière que Dieu place dans le cœur de qui Il veut. »

— Ibn Mas'oud —

Abstract

Automatic brain tumor segmentation from multimodal MRI is a critical task in clinical practice, yet it remains limited by the variability and high cost of manual delineation. This work presents a comparative evaluation of two hybrid deep learning architectures, HybridUNet3D and TransUNet3D, developed using the BraTS dataset. The first model combines a residual 3D U-Net with attention mechanisms and an ASPP module for multi-scale feature extraction, while the second integrates a Transformer module at the bottleneck to capture global contextual dependencies. The models are trained using a hybrid loss function (Dice loss, cross-entropy). Experimental results demonstrate strong performance in terms of Dice similarity coefficient, with TransUNet3D showing improved robustness and inter-patient stability. Compared to state-of-the-art methods, this hybrid CNN–Transformer architecture provides accurate and reliable segmentation while remaining effective on limited datasets.

Keywords : BraTS, Brain Tumor Segmentation, U-Net 3D, Transformer, Hybrid Architecture, Multimodal MRI, Dice Similarity Coefficient, Attention Mechanism, Deep Learning.

Résumé

La segmentation automatique des tumeurs cérébrales à partir d'IRM multimodales constitue un enjeu majeur en pratique clinique, mais reste limitée par la variabilité et le coût de la délimitation manuelle. Ce travail propose une évaluation comparative de deux architectures hybrides d'apprentissage profond, HybridUNet3D et TransUNet3D, développées sur la base de données BraTS. Le premier modèle combine un U-Net 3D résiduel avec des mécanismes d'attention et un module ASPP pour l'extraction multi-échelle, tandis que le second intègre un module Transformer au goulot d'étranglement afin de capturer les dépendances contextuelles globales. Les modèles sont entraînés avec une fonction de perte hybride (Dice, entropie croisée). Les résultats expérimentaux démontrent des performances élevées en termes de coefficient de Dice, avec un avantage du TransUNet3D en termes de robustesse et de stabilité inter-patients. Comparée aux approches de l'état de l'art, cette architecture hybride CNN–Transformer offre une segmentation précise et fiable, tout en restant efficace avec des volumes de données limités.

Mots-clés : BraTS, Segmentation de Tumeurs Cérébrales, U-Net 3D, Transformer, Architecture Hybride, IRM Multimodale, Coefficient de Similarité de Dice, Mécanisme d'Attention, Apprentissage Profond.

Sommaire

Abstract **III**

Résumé **IV**

Liste des abréviations **XIII**

1 Contexte clinique : Lésions cérébrales en imagerie IRM **3**

1.1 Introduction . . . 3
1.2 Description du cerveau . . . 3
1.2.1 Description externe du cerveau . . . 3
1.2.2 Description interne du cerveau . . . 3
1.3 Anatomie cérébrale . . . 4
1.3.1 Le cerveau . . . 4
1.3.2 Cervelet . . . 4
1.3.3 Tronc cérébral . . . 5
1.4 Les tissus du cerveau . . . 6
1.4.1 Matière blanche . . . 6
1.4.2 Matière grise . . . 6
1.4.3 La matière céphalo-rachidien (LCR) (ou liquide cérébrospinale, LC) . . . 7
1.5 Exploration visuelle du cerveau . . . 7
1.5.1 Le cycle vasculaire cérébral . . . 8
1.5.2 Cellules cérébrales . . . 8
1.6 Tumeur cérébrale . . . 9
1.6.1 Tumeur cérébrale bénignes . . . 9
1.6.2 Tumeurs malignes . . . 9
1.6.3 Types de tumeurs cérébrales . . . 10
1.6.4 Causes des tumeurs cérébrales . . . 11
1.6.5 Diagnostic des tumeurs cérébrales . . . 11
1.7 Imagerie médicale . . . 12
1.7.1 La Tomodensitométrie à rayon X . . . 13
1.7.2 La spectroscopie par Résonance magnétique (SRM)) . . . 13
1.7.3 La Tomographie par émission de positons (TEP) . . . 14
1.7.4 Imagerie Hybride . . . 14
1.7.5 Imagerie par Résonance Magnétique (IRM) . . . 14
1.7.6 Segmentation des tumeurs cérébrales . . . 21

2 Approches de Deep Learning pour la segmentation d'images IRM cérébrales 24
2.1 Introduction globale . . . 25
2.2 Problématique . . . 25
2.3 État de l'art . . . 26
2.3.1 Évolution Chronologique des Architectures . . . 26
2.3.2 Analyse Critique des Méthodes . . . 27
2.3.3 Modèle Vox2Vox — Cirillo, Abramian & Eklund (2020) . . . 28
2.4 Analyse Thématique et Méthodologique . . . 31
2.4.1 Axes thématiques . . . 31
2.4.2 Choix méthodologiques communs . . . 31
2.4.3 Tableau Comparatif des Performances DSC (moyenne ± écart-type) sur BraTS 31
2.5 Lacunes Identifiées dans la Littérature . . . 32
2.6 base de données BraTS . . . 32
2.7 Traitement préalable des données . . . 33
2.7.1 Obtention des tranches 2D . . . 33
2.7.2 Normalisation par le score Z . . . 33
2.7.3 Remappage des labels . . . 33
2.7.4 Division en ensembles d'entraînement et de validation . . . 34
2.8 Les Neurones Artificiels . . . 34
2.8.1 Le Neurone Formel . . . 34
2.8.2 La fonction d'entrée . . . 34
2.8.3 Le biais d'entrée . . . 34
2.8.4 Le Perceptron . . . 35
2.8.5 Réseau de neurones(perceptron) multicouche . . . 35
2.9 Réseau de neurones artificiels . . . 36
2.9.1 Types de réseaux neuronaux . . . 38
2.10 Apprentissage profond . . . 39
2.11 Méthodes de Propagation Avant et Rétropropagation dans les ANN . . . 40
2.12 Première approche : CNN 2D Baseline . . . 43
2.13 Architecture du modèle CNN . . . 43
2.13.1 Couche d'entrée . . . 44
2.13.2 Couche de convolution . . . 44
2.13.3 couche de mise en commun : . . . 46
2.13.4 Couche entièrement connectée . . . 47
2.13.5 Cartes des caractéristiques «Feature Maps» . . . 47
2.14 Fonction d'activation . . . 48
2.15 Fonctions de perte . . . 50
2.15.1 L'erreur quadratique moyenne (MSE) . . . 50
2.15.2 L'algorithme de rétro-propagation . . . 50
2.16 Justification du Choix des CNN pour l'IRM Cérébrale . . . 50
2.17 Deuxième approche : Architecture U-Net . . . 51
2.17.1 Structure de U-Net 2D . . . 52
2.17.2 Fonction de perte hybride et processus d'entraînement . . . 53

2.18 Optimisation des fonctions . . . 55
2.18.1 Taux d'apprentissage (Learning Rate) . . . 55
2.18.2 Optimisation à l'aide d'Adam . . . 55
2.19 Architecture Proposée . . . 57
2.19.1 U-Net 3D — Transition vers le Volumétrique . . . 57
2.19.2 TransUNet 3D — Intégration du Transformer . . . 60
2.19.3 Architecture Hybride Finale — TransResUNet3D . . . 62
2.20 Conclusion . . . 69
Conclusion . . . 69

3 Résultats et interprétations **71**
3.1 Introduction . . . 72
3.2 Métriques d'évaluation . . . 72
3.2.1 Coefficient de similarité de Dice (DSC) . . . 73
3.2.2 Perte de Dice (Dice Loss) . . . 73
3.2.3 Fonction de Perte d'entropie (Cross-Entropy Loss) : . . . 74
3.3 Résultats des tests . . . 74
3.3.1 Résultats du modèle CNN 2D multimodal . . . 74
3.3.2 Résultats des tests du modèle U-Net 2D . . . 79
3.3.3 Résultats du modèle HybridUNet3D . . . 83
3.3.4 Résultats des tests du modèle UNET 3D Transformer . . . 86
3.4 Comparaison avec l'État de l'Art . . . 97
3.5 Perspectives . . . 100
3.6 Principales Contributions . . . 101

Table des figures

1.1 L'emplacement du cervelet [1] 5
1.2 Tronc cérébrale [2] 6
1.3 Les tissus du cerveau [3] 7
1.4 Les axes de coupe pour la visualisation de cerveau [4] 8
1.5 Les différentes cellules du cerveau [5] 9
1.6 Imagerie médicale : méthodes de reconstruction 13
1.7 Image TDM d'un glioblastome avec injection de produit de contraste 13
1.8 Gauche : IRM avec produit de contraste (gadolinium). 14
1.9 Effets de l'application d'un champ magnétique et d'une impulsion d'énergie RF [6] . 15
1.10 Acquisition d'images à l'aide d'un scanner IRM [7]. 16
1.11 Plans d'imagerie médicale anatomique : coronal, sagittal et axial. 17
1.12 Plans axial, sagittal et coronal d'une IRM cérébrale.. 18
1.13 Séquences d'images de tumeurs cérébrales IRM de gauche à droite : FLAIR, pondérée T1 et pondérée T2 [8]. 19
1.14 Différences de luminosité des vaisseaux sanguins entre les scans T1 et T1ce [9]. . . . 20
1.15 Comparaison des plans axiaux T1, T2 et FLAIR [9]. 21
1.16 Représentation des trois sous-régions (A, B, C) considérées pour l'évaluation et des trois étiquettes (D). [10] 22

2.1 les composants d'un neurone formel. 34
2.2 schéma du perceptron. 35
2.3 perceptron multicouche 35
2.4 Dendrites, soma et axone myélinisé, avec un flux de signal des entrées aux dendrites aux sorties aux bornes des axones. Le signal est une courte impulsion électrique appelée potentiel d'action ou impulsion.[11] 36
2.5 Mise en parallèle de la structure d'un réseau de neurones artificiel (RNA) et celle d'un neurone biologique : le RNA est constitué de couches d'entrée, cachées et de sortie, où les neurones sont reliés par des connexions pondérées.Par contre, le neurone biologique possède des dendrites, un noyau et un axone qui assurent la transmission du signal[12].. 37
2.6 Architecture of an RNN. 39
2.7 Architecture d'un ANN[13] 41
2.8 Illustration graphique des fonctions (a) sigmoïde et (b) ReLU [14]. 42
2.9 Architecture de base des CNN [14]. 43
2.10 Le processus de convolution [30]. 44

2.11 Exécution d'une opération de convolution avec un noyau 5 x 5 et un stride de 1. Source : Priya Pedamkar, Réseau de neurones DNN, EDUCBA 45
2.12 Convolution effectuée avec un noyau 2x2 et un stride de 2, menant à une carte de caractéristiques diminuée[15]. 45
2.13 Processus de convolution avec ajout de zéros et un décalage de 2, conservant ainsi les dimensions spatiales[15]. 46
2.14 Exemple max pooling. 46
2.15 Exemple de moyenne pooling 47
2.16 Exemple de Global Pooling. 47
2.17 Exemple de réseau convolutif deux étapes CONV-POOL suivies par deux couches Fc. 47
2.18 Illustration d'une carte de caractéristiques « Feature Map ». 48
2.19 Illustration de la carte des attributs dans les réseaux neuronaux.. 48
2.20 diagrame softmax[82] 49
2.21 algorithme de rétropropagation 50
2.22 Structure de l'U-Net. 51
2.23 illustration d'un classifieur softmax. 55
2.24 Exemple de comparaison d'Adam par rapport à d'autres algorithmes d'optimisation. 56
2.25 Architecture 3D-U-Net [16] 57

3.1 BraTS20_Training_040_slice100 76
3.2 BraTS20_Training_031_slice100 77
3.3 BraTS20_Training_026_slice90 78
3.4 Courbes d'apprentissage du modèle Baseline CNN : évolution de la fonction de perte (Loss) et du coefficient de Dice (DSC). 78
3.5 Matrice de confusion du modèle U-NET 2D illustrant les résultats de segmentation au niveau des pixels.. 81
3.6 Planche comparative des résultats de segmentation (Patient 031). 82
3.7 Courbes d'apprentissage du modèle HybridUNet3D sur 50 époques. 84
3.8 Comparaison entre le Ground Truth et les prédictions pour le patient BraTS20_Training$_0$14_*slice*80. 85
3.9 Courbes d'apprentissage du modèle TransUNet3D sur 100 époques (entraînement et validation) 87
3.10 Matrice de confusion du modèle hybride TransUNET 3D, montrant les résultats de segmentation pixel par pixel. 89
3.11 Distribution du coefficient de Dice (DSC) par patient pour les trois régions tumorales (WT, TC, ET) sur l'ensemble de validation. 90
3.12 Comparaison des métriques de segmentation (DSC, IoU, Sensibilité, Spécificité, Précision, Accuracy) avec et sans post-traitement CCA pour les trois régions tumorales (WT, TC, ET). 91
3.13 comparaison des performances des trois régions tumorales (WT, TC, ET) sur l'ensemble des métriques d'évaluation (moyenne ± écart-type, avec CCA). 92
3.14 Visualisation qualitative comparative des segmentations TransUNet3D sur quatre cas représentatifs de l'ensemble de validation (Rouge=NCR/NET, Vert=Edema, Bleu=ET). 93
3.15 Comparaison entre le Ground Truth et les prédictions (sans CCA et avec TTA+CCA) pour le patient BraTS20_Training_026_slice 100. 94

3.16 Comparaison entre le Ground Truth et les prédictions (sans CCA et avec TTA+CCA) pour le patient BraTS20_Training_023_slice 80. 94
3.17 Comparaison entre le Ground Truth et les prédictions (sans CCA et avec TTA+CCA) pour le patient BraTS20_Training_014_slice100. 94
3.18 Comparaison entre le Ground Truth et les prédictions (sans CCA et avec TTA+CCA) pour le patient BraTS20_Training_031_slice 100. 95

Liste des tableaux

1.1 Les types de lobes et leurs fonctions [17] 4
1.2 gradation des tumeurs selon l'OMS (Organisation Mondiale de la Santé) [18] 12

2.1 Comparaison des performances sur la base de données BraTS. 31
2.2 Comparaison entre les réseaux neuronaux biologiques et artificiels 38
2.3 Tableau comparatif de synthèse 69

3.1 Matrice de confusion standard pour la classification binaire 72
3.2 Métriques d'évaluation DSC du modèle CNN 2D pour différentes sous-régions tumorales et profondeurs de coupes. 75
3.3 Synthèse des performances et paramètres d'évaluation du modèle U-Net 2D 80
3.4 Résultats DSC par patient du HybridUNet3D sur l'ensemble de validation (10 patients). 84
3.5 Résultats récapitulatifs du HybridUNet3D (10 patients de validation). 85
3.6 Métriques de segmentation sur l'ensemble de validation. 88
3.7 Comparaison directe HybridUNet3D et TransUNet3D sur BraTS 97
3.8 Comparaison des performances DSC (moyenne ± écart-type) avec les méthodes de l'état de l'art sur BraTS 98

Liste des abréviations

Abréviation	**Signification**
2D	Deux Dimensions
3D	Trois Dimensions
Adam	Adaptive Moment Estimation
BraTS	Brain Tumor Segmentation (Challenge)
CCA	Connected Component Analysis (Analyse en Composantes Connexes)
CE	Cross-Entropy (Entropie Croisée)
CNN	Convolutional Neural Network (Réseau de Neurones Convolutif)
CRF	Conditional Random Field
CUDA	Compute Unified Device Architecture
DL	Deep Learning (Apprentissage Profond)
DSC	Dice Similarity Coefficient (Coefficient de Similarité de Dice)
ECE	Expected Calibration Error
ED	Edematous / Invaded tissue (Tissu Œdémateux)
EM	Expectation-Maximization
ET	Enhancing Tumor (Tumeur Rehaussée)
FFN	Feed-Forward Network
FLAIR	Fluid Attenuated Inversion Recovery
GAN	Generative Adversarial Network (Réseau Antagoniste Génératif)
GPU	Graphics Processing Unit (Processeur Graphique)
GT	Ground Truth (Vérité Terrain)
HUT	Hybrid UNet Transformer
IA	Intelligence Artificielle
IoU	Intersection over Union
IRM	Imagerie par Résonance Magnétique
LCE	Loss Cross-Entropy (Perte par Entropie Croisée)
LDS	Loss Deep Supervision (Perte par Supervision Profonde)
LR	Learning Rate (Taux d'Apprentissage)
LSTM	Long Short-Term Memory
MHA	Multi-Head Attention
MLP	Multi-Layer Perceptron
MSE	Mean Squared Error (Erreur Quadratique Moyenne)
NCR	Necrotic Core (Noyau Nécrotique)

RF	Random Forest
ReLU	Rectified Linear Unit
RNN	Recurrent Neural Network (Réseau de Neurones Récurrent)
ResNet	Residual Network
SE	Squeeze-and-Excitation
SwinUNet	Swin Transformer UNet
T1	Séquence IRM pondérée T1
T1ce	T1 avec rehaussement de Contraste
T2	Séquence IRM pondérée T2
TC	Tumor Core (Cœur Tumoral)
TTA	Test-Time Augmentation
U-Net	U-shaped Network (Architecture en forme de U)
UNETR	UNEt TRansformer
VGG	Visual Geometry Group
ViT	Vision Transformer
WT	Whole Tumor (Tumeur Complète)

Introduction générale

Les tumeurs cérébrales représentent une pathologie particulièrement complexe en raison de leur grande variabilité morphologique, de leur localisation dans des zones fonctionnelles sensibles du cerveau et de l'impact direct qu'elles peuvent avoir sur le pronostic vital et la qualité de vie des patients. Dans ce contexte, l'imagerie par résonance magnétique (IRM) occupe une place centrale dans le diagnostic, la caractérisation des lésions, la planification thérapeutique et le suivi de l'évolution tumorale. Grâce à ses différentes modalités, notamment T1, T1ce, T2 et FLAIR, l'IRM fournit des informations anatomiques et pathologiques complémentaires, essentielles pour distinguer les différentes sous-régions tumorales telles que la tumeur complète, le cœur tumoral et la tumeur rehaussée.

Cependant, l'interprétation et la segmentation manuelle de ces images restent des tâches longues, coûteuses et fortement dépendantes de l'expertise du praticien. La délimitation précise des régions tumorales peut varier d'un spécialiste à l'autre, en particulier lorsque les contours sont diffus, que les structures sont hétérogènes ou que les volumes tumoraux présentent des formes irrégulières. Cette problématique a motivé le développement de méthodes automatiques de segmentation, en particulier dans le cadre du challenge BraTS, devenu une référence internationale pour l'évaluation des performances des algorithmes appliqués aux IRM cérébrales multimodales.

Au cours des dernières années, l'essor de l'intelligence artificielle et plus particulièrement du deep learning a profondément transformé le traitement d'images médicales. Les réseaux de neurones convolutifs ont démontré leur efficacité pour l'extraction automatique de caractéristiques locales, tandis que les architectures de type U-Net se sont imposées comme une solution de référence en segmentation biomédicale grâce à leur structure encodeur–décodeur et à leurs connexions de saut. Plus récemment, l'introduction des Transformers en vision par ordinateur a ouvert de nouvelles perspectives en permettant de modéliser les dépendances globales et les relations à longue distance, mieux adaptées à la complexité spatiale des lésions cérébrales.

C'est dans cette dynamique que s'inscrit ce mémoire, dont l'objectif principal est d'étudier et de comparer différentes architectures de segmentation des lésions cérébrales en IRM, depuis une approche de base fondée sur un CNN 2D multimodal jusqu'à des architectures hybrides 3D associant convolutions, mécanismes d'attention et Transformers. Le travail repose sur l'exploitation de la base de données BraTS, sur un prétraitement des volumes IRM, sur la mise en œuvre de plusieurs stratégies d'apprentissage profond et sur l'évaluation des performances à l'aide de métriques de référence telles que le coefficient de similarité de Dice.

L'intérêt scientifique et clinique de cette étude réside dans la recherche d'un compromis entre précision de segmentation, robustesse inter-patients et capacité de généralisation, notamment dans un contexte où les données annotées restent limitées. En comparant des modèles 2D et 3D, purement convolutionnels puis hybrides CNN–Transformer, ce travail vise à mettre en évidence l'apport des

mécanismes d'attention et de la modélisation contextuelle globale pour améliorer la détection des sous-régions tumorales, en particulier les régions les plus difficiles à segmenter.

Ainsi, ce mémoire se structure autour de trois axes complémentaires : d'abord la présentation du contexte clinique des lésions cérébrales et du rôle de l'IRM, ensuite l'étude théorique et méthodologique des approches de deep learning appliquées à la segmentation, et enfin l'analyse expérimentale des résultats obtenus ainsi que leur comparaison avec l'état de l'art. À travers ce parcours, l'ambition est de contribuer au développement d'outils d'aide au diagnostic plus fiables, capables de soutenir la pratique clinique par une segmentation rapide, reproductible et précise des tumeurs cérébrales.

Chapitre 1

Contexte clinique : Lésions cérébrales en imagerie IRM

1.1 Introduction

Les tumeurs cérébrales sont des cellules qui se divisent de façon incontrôlée. Selon leur nature, leur origine et leur vitesse de croissance, on distingue deux types de tumeurs cérébrales. Détecté précocement, le cancer peut donner à une personne une nouvelle chance de vie, en suivant un traitement efficace pour la sauver.

L'imagerie par résonance magnétique (IRM) permet d'explorer rapidement le cerveau et de détecter les tumeurs. L'IRM (imagerie par résonance magnétique) est une méthode très employée pour obtenir des images médicales de bonne qualité. Les techniques de La détection du cancer doit être fiable et robuste pour le diagnostic. C'est dont il sera question dans ce chapitre.

1.2 Description du cerveau

1.2.1 Description externe du cerveau

Le cerveau, situé au sommet de la colonne vertébrale, est le principal centre de commande du corps et se compose d'un réseau de neurones qui s'étend du cou jusqu'à la partie inférieure de la colonne vertébrale .Il reçoit des signaux des organes pour percevoir les sensations internes et externes, analyse le système nerveux recueille rapidement ces informations et envoie des messages appropriés pour réguler les mouvements et les fonctions corporelles. Le cerveau permet également de stocker des souvenirs, de favoriser l'apprentissage et la mémoire et est à l'origine d'idées, d'émotions et de pensées [19] .

1.2.2 Description interne du cerveau

Au plan microscopique, le tissu nerveux se compose de neurones ou cellules nerveuses et de cellules gliales ou cellules de soutien et de protection, les deux étant d'origine ectodermique. Les vaisseaux sanguins et les méninges ne font pas partie du tissu nerveux, proviennent du mésoderme.

Le système nerveux comporte des neurones, qui sont son unité fonctionnelle de base, et dont le nombre est environ 10 à 50 fois supérieur à celui des cellules gliales. Le système nerveux humain contiendrait environ 100 milliards de neurones. Les neurones sont chargés de la transmission d'un signal électrique appelé influx nerveux [20] .

1.3 Anatomie cérébrale

1.3.1 Le cerveau

Le cerveau constitue la partie prédominante du système nerveux central : environ 85% de son poids et d'un poids de l'ordre de 1300 à 1500 g. Dans la partie supérieure du crâne, il a pour rôle de réguler les mouvements des muscles, de produire des hormones, de développement des émotions et des sentiments, de même que de la mémorisation, de la préservation et de l'apprentissage. Le cerveau comporte deux hémisphères séparés par une fissure centrale et contient les principaux lobes cérébraux majeurs, chacun ayant de multiples fonctions [21] .

Les lobes	**Principale fonction**
Lobe frontal	— Langage et parole — Contrôle de mouvement — Comparaison et classification — Formation du personnage — Gérer l'attention
Lobe pariétal	— Lecture, perception visuelle. — La sensation de toucher, de douleur et de température. — Le lobe pariétal droit gère la sensibilité du côté gauche de Corps. — Le lobe pariétal gauche gère la sensibilité du côté droit De Corps.
Lobe occipital	— Gestion de la visibilité spatiale, prise de conscience de l'distance, de la profondeur et de la couleur.
Lobe temporal	— Comprendre la langue et les émotions. — Séquence et organisation

TABLE 1.1 – Les types de lobes et leurs fonctions [17]

1.3.2 Cervelet

Le cervelet se trouve au sommet du tronc cérébral, là où la moelle épinière communique avec l'encéphale. Il pèse environ 10 % du poids total du cerveau, mais contient environ la moitié des neurones totaux du cerveau. Ces neurones sont chargés de transmettre au cerveau des signaux électriques provenant du corps. Le cervelet joue principalement un rôle de réception d'informations

venant des organes sensoriels, de la moelle épinière et d'autres parties du cerveau et de régulation des mouvements volontaires [22].

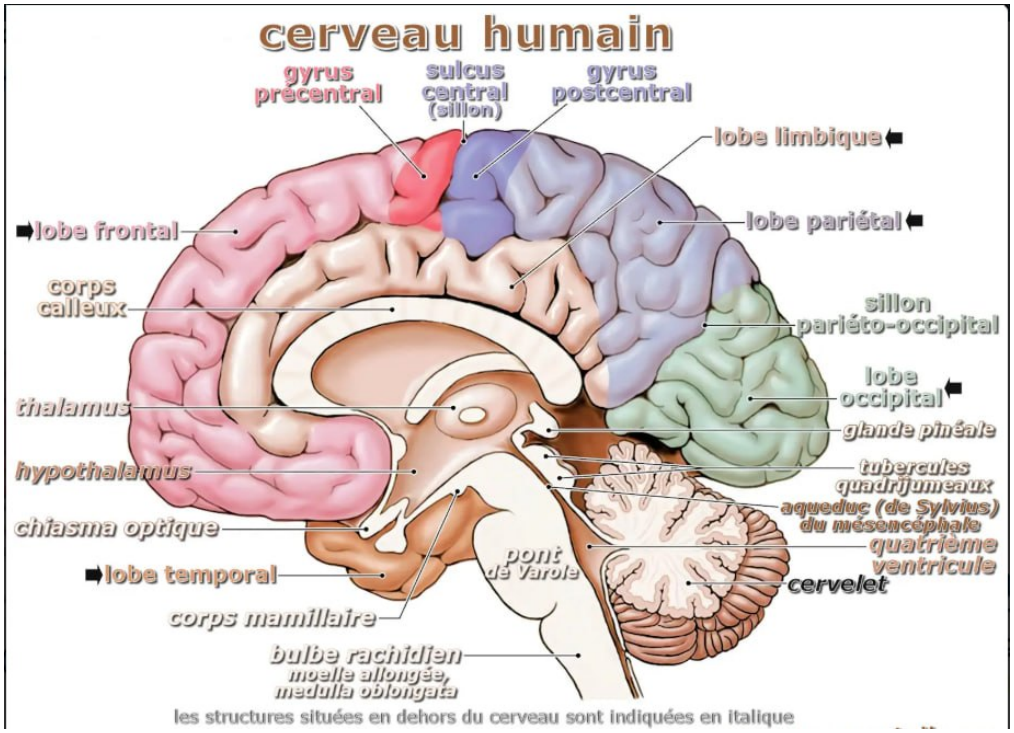


FIGURE 1.1 – L'emplacement du cervelet [1]

.

1.3.3 Tronc cérébral

Le tronc cérébral est une composante cruciale du système nerveux, garantissant des fonctions indispensables à l'existence humaine. C'est le centre de transit des neurones sensoriels et moteurs, émis et contenus dans le cerveau. Une de ses fonctions primordiales est de contrôler l'équilibre du corps, la température corporelle, le rythme respiratoire et la tension artérielle. Le tronc cérébral comprend trois structures principales qui collaborent harmonieusement pour préserver ces fonctions essentielles [23] :

- **Mésencéphale**

  Le cerveau intermédiaire joue un rôle crucial dans le contrôle des mouvements oculaires et le traitement de l'information audio-visuelle [23].

- **Le pont**

  Le pont est un ensemble de nerfs qui relie différentes zones du cerveau. Le tronc cérébral, constitué principalement de neurones circadiens, est chargé de transmettre des informations sensorielles et participe également aux mouvements du visage [23].

- **Bulbe rachidien**

  Il surveille la respiration, la pression artérielle, les battements de cœur et la déglutition, où les signaux du cortex sont transmis à la moelle épinière et aux nerfs qui se divisent de la moelle épinière au pont et au tronc cérébral [23].

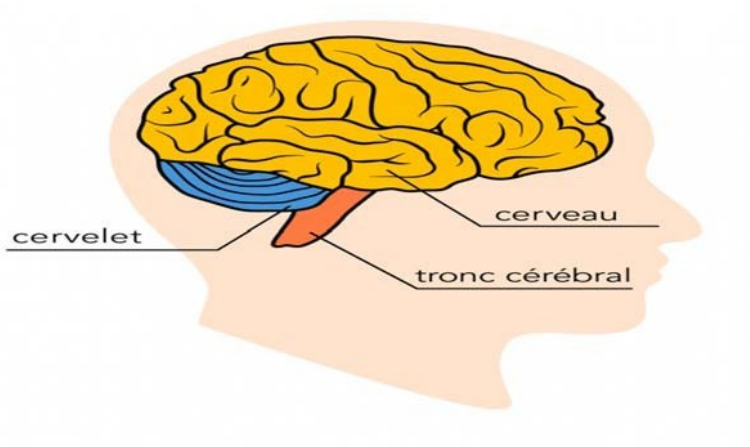


FIGURE 1.2 – Tronc cérébrale [2]

1.4 Les tissus du cerveau

1.4.1 Matière blanche

La matière blanche est une type de tissu présent dans le système nerveux central, constitué principalement d'axones myélinés de neurones. Il représente la composante interne du cerveau et l'élément externe de la moelle épinière. Les axones, qui sont des prolongements des neurones, se trouvent dans la matière blanche. Ces prolongements sont enveloppés par des gaines et des cellules de myéline, qui ne sont pas neuronales, et contribuent à la composition du tissu interstitiel des cellules nerveuses : les nerfs gliaux. La fonction de la matière blanche est de garantir la conduction de l'influx nerveux [24].

1.4.2 Matière grise

Constitué principalement de neurones dotés d'un corps cellulaire et d'un arbre dendritique, ainsi que de certaines cellules gliales. Il paraît plus sombre que le reste du tissu, étant donné qu'il renferme le noyau cellulaire. Nous avons découvert que la matière grise se localise majoritairement dans le cortex cérébral, l'insula, les ganglions de la base et les noyaux du tronc cérébral. Et au sein de la moelle épinière.

- **Cortex cérébral**

  englobe la totalité du cerveau. Il couvre une vaste zone, puisqu'il suit toutes les spirales externes du cerveau.

- **Noyaux basaux**

  en association avec le cortex cérébral, ce sont les structures uniques du cerveau contient également des neurones, bien que leur densité y soit moins élevée qu'au niveau du cortex cérébral. Ils intègrent les noyaux du télencéphale, connus sous le nom de noyau basal (ou striatum), où l'on identifie le noyau accumbens, le noyau caudé et le noyau lenticulaire.

1.4.3 La matière céphalo-rachidien (LCR) (ou liquide céré-brospinale, LC)

Le liquide céphalo-rachidien, qui est constitué à 99% d'eau, se trouve dans le système nerveux central et occupe deux régions où les caractéristiques anatomiques diffèrent grandement. Le liquide céphalo-rachidien, qui entoure le cerveau et remplit les ventricules cérébraux, est également connu sous le nom de liquide céphalo-rachidien. Ceci contribue à la protection. Le cerveau crée un impact en jouant le rôle d'un amortisseur et en réalisant des tâches primordiales dans les interactions entre le sang et les tissus nerveux.

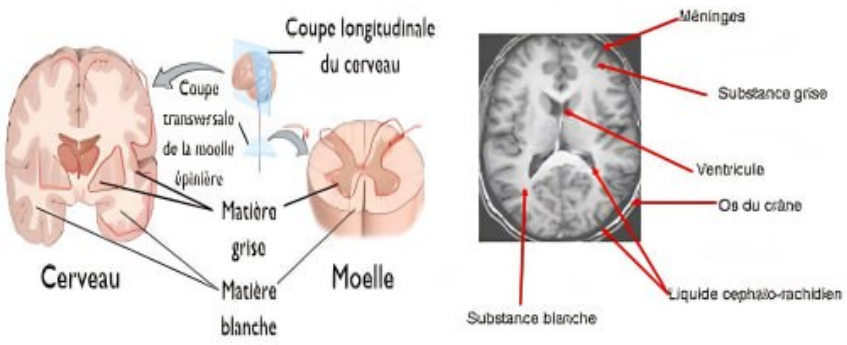


FIGURE 1.3 – Les tissus du cerveau [3]

1.5 Exploration visuelle du cerveau

On peut observer les sections bidimensionnelles du cerveau sous de nombreux angles différents. Par conséquent, pour effectuer des sections, on identifie trois axes anatomiques [25] :

- **Coupe axiale**

  il s'agit de coupes d'images obtenues grâce à des techniques d'imagerie médicale. Essentiellement dans une position horizontale du corps. En IRM, ces sections sont à angle droit par rapport à l'axe du champ magnétique principal.

- **Coupe sagittale**

  L'image sagittale est une représentation obtenue en sectionnant le cerveau ou toute autre structure anatomique dans un plan vertical, qui est parallèle à la ligne médiane du corps. Ils offrent la possibilité de visualiser les structures latérales du cerveau et fournissent une perspective globale de la zone d'intérêt.

- **Coupes circulaires**

  Ces coupes font référence à celles qui sont perpendiculaires aux coupes axiales et sagittales, représentant l'image frontale du cerveau.

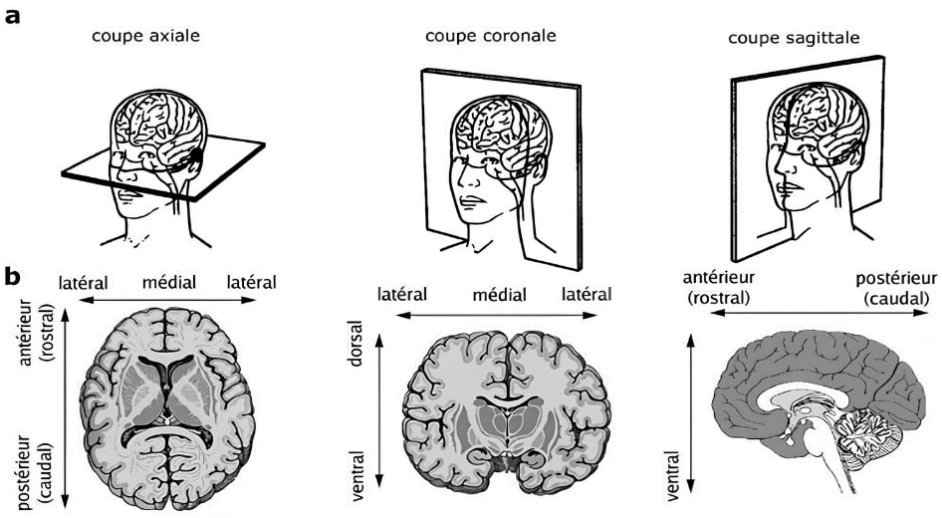


FIGURE 1.4 – Les axes de coupe pour la visualisation de cerveau [4]

1.5.1 Le cycle vasculaire cérébral

Le cycle vasculaire cérébral désigne le flux sanguin qui fournit au cerveau ses nutriments et son oxygène à travers un ensemble de vaisseaux sanguins. Les artères acheminent le sang riche en oxygène et les nutriments vers le cerveau, alors que les veines éliminent le sang pauvre en oxygène et les résidus métaboliques. Le cerveau, étant très sensible à un manque de sang, est protégé par divers mécanismes préventifs au sein du système vasculaire cérébral qui assurent une circulation sanguine adéquate [26].

1.5.2 Cellules cérébrales

Les neurones, qui sont les cellules du cerveau, possèdent un corps cellulaire et des prolongements nerveux connus sous le nom de dendrites, ainsi qu'un cylindre axonal appelé axone. Ce dernier se termine par des plaques terminales où les synapses ont la possibilité de se développer. Les synapses jouent un rôle plus essentiel que le nombre de neurones dans la transmission d'informations entre les cellules cérébrales, puisqu'elles véhiculent des informations même vers des régions distantes du corps. Les dendrites sont capables de s'étendre sur une longueur pouvant atteindre un mètre, et une cellule cérébrale a la capacité d'héberger jusqu'à 10 000 dendrites [5].

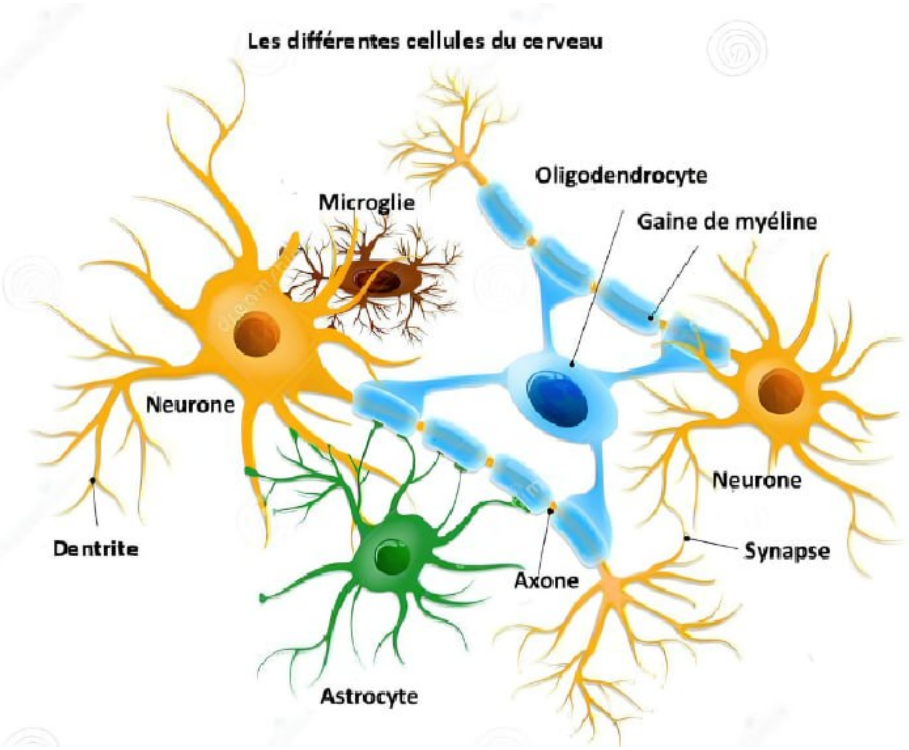


FIGURE 1.5 – Les différentes cellules du cerveau [5]

1.6 Tumeur cérébrale

Une tumeur se définit comme une accumulation de cellules anormales qui se reproduisent de façon non maîtrisée, ainsi une prolifération atypique. Ces cellules atypiques poursuivent leur division et leur croissance, envahissent les tissus environnants et finiront par se répandre dans d'autres zones du corps. Selon leur potentiel de diffusion et d'invasion des tissus adjacents, les tumeurs peuvent être classées comme bénignes (non cancéreuses) ou malignes (cancer en phase avancée). Les tumeurs malignes ont la capacité d'endommager gravement les tissus qui les entourent et peuvent s'avérer fatales si elles ne font pas l'objet d'un traitement.

1.6.1 Tumeur cérébrale bénignes

Ces tumeurs ne renferment pas de cellules cancéreuses, elles sont composées de cellules qui se multiplient lentement et n'envahissent pas les tissus adjacents. Une tumeur cérébrale non maligne peut être suivie uniquement par des techniques radiologiques, ou complètement éliminée par voie chirurgicale, ou en partie. Si la tumeur bénigne n'est pas entièrement retirée, il y a un risque de réapparition [27].

1.6.2 Tumeurs malignes

Les tumeurs malignes sont des amas de cellules cancéreuses qui croissent de manière incontrôlée, envahissent les tissus avoisinants et se disséminent vers d'autres zones du corps via le système sanguin ou lymphatique. On appelle ce phénomène de propagation métastase. À

l'opposé des tumeurs bénignes, les tumeurs malignes sont perçues comme une menace pour la santé en raison de leur capacité à se disséminer rapidement vers d'autres zones du corps [27].

1.6.3 Types de tumeurs cérébrales

- **Les tumeurs primitives**

  Les tumeurs cérébrales primaires sont des proliférations anormales de cellules, pouvant être bénignes ou malignes, et entraîner des symptômes sévères, notamment lorsqu'elles sont localisées dans des régions cruciales du cerveau [28].

- **Les tumeurs secondaires**

  Les tumeurs secondaires sont des cancers qui émergent de cellules cancéreuses ayant migré depuis une première tumeur localisée hors du cerveau, via le système circulatoire, vers les tissus cérébraux.

Exemples des tumeurs cérébrales primitives du cerveau :

- **Gliome**

  Les tumeurs cérébrales se forment à partir des cellules gliales, qui constituent la structure de soutien du cerveau et du système nerveux. On dénombre plusieurs sortes de cellules. Certaines variétés de gliomes sont dues à l'action des cellules gliales dans leur développement. Les gliomes constituent la majorité des tumeurs cérébrales.

- **Méningiomes**

  La majorité des tumeurs non cancéreuses croissent de façon nuisible dans les méninges, à l'intérieur du crâne ou le long de la colonne vertébrale. Bien que la plupart des méningiomes soient bénins, il y a des cas isolés où l'évolution peut être défavorable, passant de bénin à maligne.

- **Névrome**

  Dans la plupart des cas, les tumeurs sont de grade I, c'est-à-dire bénignes, et elles se forment à partir des gaines entourant les nerfs périphériques, connues sous le nom de cellules de Schwann. On retrouve ces tumeurs dans les nerfs crâniens ainsi que dans les nerfs périphériques de l'ensemble du corps. Le type de tumeur le plus fréquent est le schwan nome ou névrome.

- **Adénome**

  L'adénome hypophysaire est un néoplasie bénin. Cette tumeur a la capacité de provoquer des déséquilibres hormonaux ou parfois des soucis de vision. Les symptômes résultent de la pression exercée par la tumeur sur la région cérébrale.

- **Médulloblastome**

  Les tumeurs malignes qui se trouvent dans le cervelet affectent principalement les enfants, en particulier ceux âgés de moins de 15 ans.

- **Lymphome**

  Le lymphome est une néoplasie qui touche le système lymphatique. Le lymphome primitif du système nerveux central est une néoplasie maligne rare qui se manifeste Les lymphomes secondaires du système nerveux central se manifestent fréquemment par une atteinte du cortex cérébral.

- **Métastase**

  Les tumeurs se constituent de cellules cancéreuses qui se détachent d'une tumeur primaire. Il arrive que ces tumeurs métastatiques soient le premier indice d'une affection. Les métastases sont des tumeurs malignes qui se développent à partir d'une autre tumeur située dans différentes parties du corps (comme les poumons, les seins, la peau, etc.). Les cellules cancéreuses se déplacent vers le cerveau par l'intermédiaire du sang.

1.6.4 Causes des tumeurs cérébrales

De nombreuses recherches scientifiques ont été réalisées ou sont en cours actuellement pour identifier les éléments qui encouragent le développement d'une tumeur cérébrale. Nous abordons les facteurs de risque. On examine trois catégories de facteurs de risque : ceux associés à l'environnement, la génétique et les virus. Un facteur de risque ne suffit pas à lui seul pour justifier l'apparition d'une tumeur. Effectivement, pour deux individus exposés au même danger, l'un peut développer une tumeur tandis que l'autre ne le fait pas. Inversement, un individu peut souffrir d'une tumeur sans être exposé à aucun facteur de risque [29].

1.6.5 Diagnostic des tumeurs cérébrales

L'évaluation d'une tumeur est un processus clinique rigoureux et sophistiqué qui prépare ensuite à la décision de traitement appropriée. On peut le distinguer en plusieurs étapes : suspicion, détection, examen, identification de sa nature histologique...

1.6.5.1 Examen clinique

Une tumeur cérébrale peut provoquer de très nombreuses et très variables manifestations cliniques. Les signes cliniques varient beaucoup d'un patient à l'autre selon l'endroit où se trouve la tumeur. Les céphalées accompagnées de nausées, la diminution des capacités physiques et intellectuelles, la faiblesse musculaire, les troubles de l'élocution et des désagréments visuels figurent parmi les symptômes les plus fréquents. Les crises d'épilepsie peuvent également être le premier signe d'une tumeur. En étudiant ces symptômes, le médecin décide de faire plusieurs tests dont des radiographies. On recherche donc des masses opaques, transparentes ou de densité inhabituelle, témoignant d'un processus tumoral [30]. Une fois ce processus identifié, on décompte le nombre de lésions et on étudie leur localisation. Nous examinons ensuite pour chacune d'entre elles divers aspects tels que :

— La taille
— La forme
— Le caractère infiltrant ou circonscrit

— La localisation dans le cerveau et par rapport aux structures cérébrales environnantes, la composition

on cherche notamment à voir si celle-ci est homogène ou si, au contraire, les lésions comportent des kystes, du liquide ou des phénomènes nécrotiques. Ce n'est pas seulement sur les clichés que l'on recherche ces éléments. On peut en effet constater divers signes indirects : présence d'œdème au contact de la tumeur, modifications topologiques et volumiques des structures anatomiques adjacentes. D'autres examens peuvent être nécessaires selon l'état du patient : - Prendre une petite quantité de liquide céphalo-rachidien (par ponction lombaire). - Injection dans les artères cérébrales d'un produit iodé pour préciser leur trajet et faire une artériographie

1.6.5.2 La gradation des tumeurs

Le tableau 1.2 présente la gradation des tumeurs :

grade I *Tumeurs bénignes*	Tumeurs à croissance lente et normalement bien circoncites bien qu'elles puissent envahir de grandes régions du cerveau. Selon la localisation, une ablation chirurgicale ou une biopsie peut être recommandée.
grade II *Tumeurs bénignes*	Tumeurs à croissance lente, mais, contrairement aux tumeurs de grade I, leurs limites sont imprécises. Les entités tumorales appartenant à ce groupe sont moins nombreuses que celles du grade I.
grade III *Tumeurs malignes*	Tumeurs anaplasiques Dans les tumeurs de bas grade (I et II), des foyers de cellules anaplasiques (cellules ayant perdues une partie de leurs caractère propre, donc anormales) se développent activement. Leur évolution est plus rapide que celle des tumeurs de bas grade.
grade IV *Tumeurs malignes*	Ces tumeurs peuvent contenir divers types de cellules qui se multiplient rapidement et ayant une forte tendance nécrosante spontanée. Elles ne sont pas bien définies et s'infiltrent dans le cerveau.

TABLE 1.2 – gradation des tumeurs selon l'OMS (Organisation Mondiale de la Santé) [18]

Le diagnostic de ces tumeurs nécessite d'abord la collecte d'images à l'aide de techniques d'imagerie médicale (tels que la tomodensitométrie, l'IRM, la TEP...) suivie du traitement de ces images par les radiologues ou par des algorithmes de traitement d'images, sujet que nous traiterons dans la section 1.7.

1.7 Imagerie médicale

Dans cette partie, nous allons exposer diverses approches de reconstruction et de traitement d'images dans le domaine de l'imagerie médicale : Les analyses d'imagerie médicale dépendent principalement de la technique de reconstruction employée, permettant d'obtenir des renseignements sur l'anatomie des organes (dimensions, configuration, volume et emplacement de

l'organe), ou sur leur fonctionnement (physiologie, métabolisme). La figure (1.6) présente les techniques les plus couramment employées :

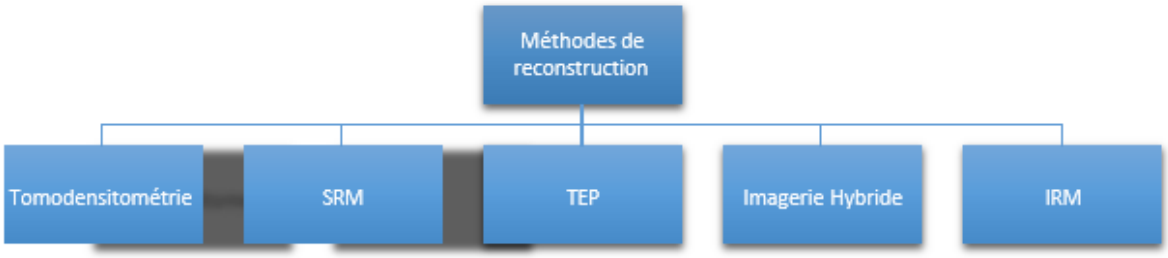


FIGURE 1.6 – Imagerie médicale : méthodes de reconstruction

1.7.1 La Tomodensitométrie à rayon X

Le scanner à rayons X, également connu sous les noms de tomodensitométrie, scanographe ou de computed tomography (CT), a longtemps été l'examen d'imagerie clinique prédominant, jusqu'à l'introduction de la résonance magnétique nucléaire [31]. Elle offre la possibilité d'observer l'objet par sections successives d'épaisseur de quelques millimètres chacune.



FIGURE 1.7 – Image TDM d'un glioblastome avec injection de produit de contraste

1.7.2 La spectroscopie par Résonance magnétique (SRM))

La SRM permet de détecter des molécules biologiques qui ne sont pas observables avec d'autres techniques. Le résultat est un spectre comportant plusieurs pics de résonance à des fréquences différentes. L'analyse de chaque pic du spectre permet de déterminer la fréquence de résonance et la surface associée à chaque pic, proportionnelle au nombre de noyaux analysés. Il existe deux formes de SRM : la SRM in vivo qui permet de préciser la nature histologique de la tumeur, et la SRM in vitro, utilisée en grande partie pour anticiper la réaction tumorale à une thérapie [31].

1.7.3 La Tomographie par émission de positons (TEP)

La TEP est une technique d'imagerie médicale qui consiste à injecter une substance radioactive qui émet des positons, à capturer les rayonnements à l'aide d'un détecteur externe et à reconstruire informatiquement une image en coupe de l'organe [32]. En imagerie TEP, les deux radio-isotopes les plus utilisés sont les isotopes 18F et 11C.

1.7.4 Imagerie Hybride

Les premières études cliniques réalisées avec la TEP/IRM suggèrent que cette technique d'imagerie avancée ouvre de larges perspectives pour améliorer le diagnostic non invasif de nombreuses pathologies. Il reste toutefois de nombreux défis méthodologiques à relever avant que l'imagerie PET-MRI puisse être intégrée à la pratique clinique courante. L'acquisition de connaissances nouvelles dans le domaine de la recherche moléculaire fondamentale et de la radio-pharmacie permettra de définir les applications futures de cette technique. Le but de la TEP/IRM est clairement de poser des diagnostics de manière non intrusive et précise, et d'évaluer rapidement la réponse au traitement [33].

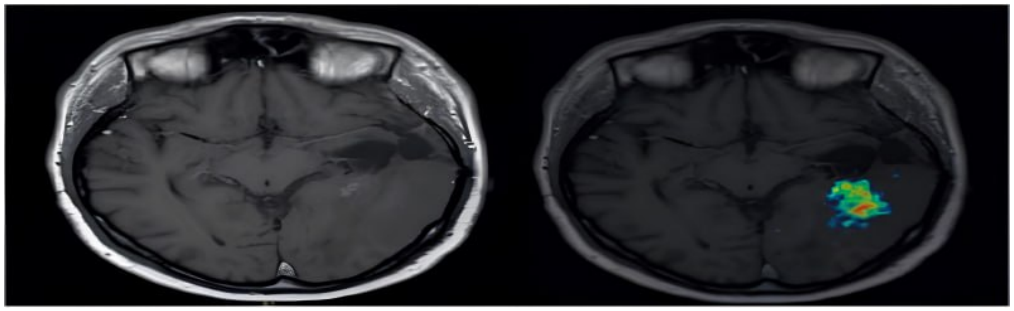


FIGURE 1.8 – Gauche : IRM avec produit de contraste (gadolinium).

Droite : Fusion avec la TEP à l'éthyle

1.7.5 Imagerie par Résonance Magnétique (IRM)

1.7.5.1 Principe de base de l'IRM

L'imagerie par résonance magnétique (IRM) est une méthode d'imagerie médicale sans intervention invasive qui se base sur la profusion d'atomes d'hydrogène dans le corps humain (eau, lipides, etc.). L'IRM est basée sur le fondement même des propriétés magnétiques des noyaux d'hydrogène, aussi appelés protons. Dans l'organisme, ces protons sont habituellement disposés de façon aléatoire. Quand un patient est introduit dans la machine IRM, un fort champ magnétique externe est mis en application. Ce champ magnétique provoque l'alignement des protons d'hydrogène dans sa direction. Cette synchronisation favorise la création des images. Pour générer une image, une brève impulsion d'énergie en radiofréquence (RF) est orientée vers le corps du patient. Cette impulsion RF est spécifiquement ajustée à la fréquence de résonance des protons d'hydrogène et est dirigée dans un sens distinct de celui du champ magnétique constant [6],[34]. L'application de l'impulsion RF occasionne une perturbation

temporaire de l'alignement des protons. Précédemment, ces protons avaient été obtenus en appliquant un champ magnétique. Après la fin de l'impulsion RF, ces derniers reviennent progressivement à leur état d'alignement d'origine, produisant de ce fait des signaux RF. Le corps humain n'ayant pas une répartition uniforme de l'eau, l'analyse des fluctuations du champ électromagnétique détecté offre la possibilité d'identifier les disparités de densité et de morphologie des tissus dans l'ensemble de l'organisme.

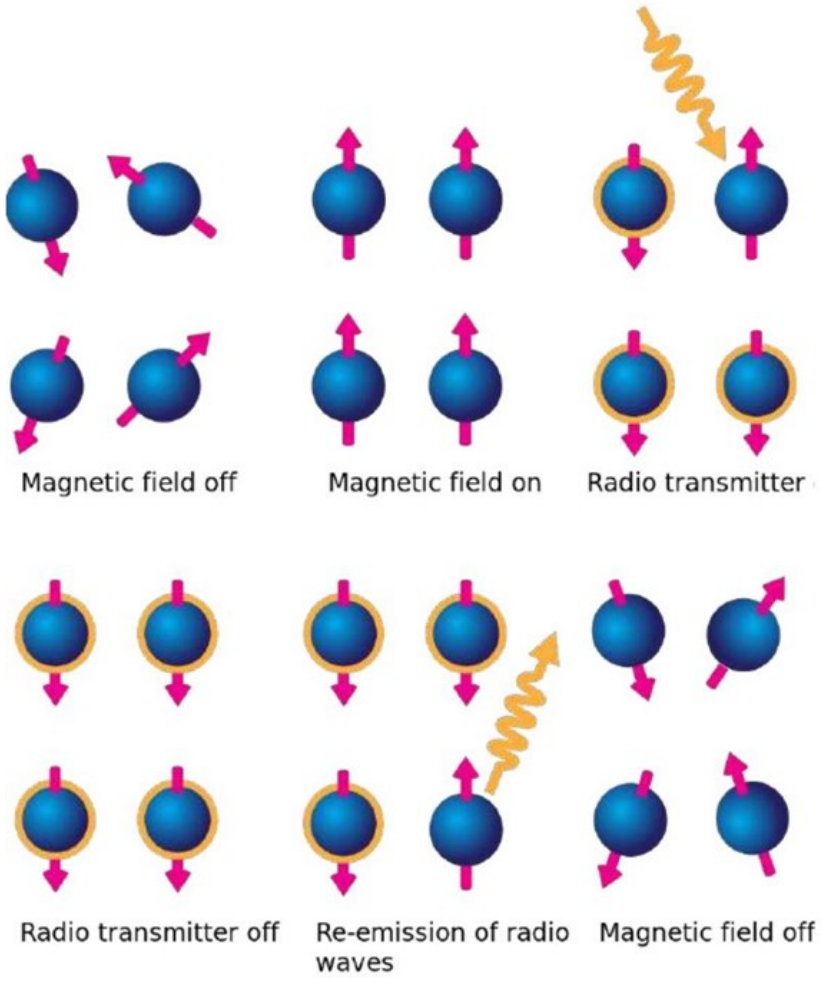


FIGURE 1.9 – Effets de l'application d'un champ magnétique et d'une impulsion d'énergie RF [6]

Ces signaux de radiofréquence produites sont détectés par les bobines réceptrices situées à l'intérieur du dispositif IRM. Ces signaux renferment des données concernant la densité et la répartition des protons d'hydrogène dans diverses structures corporelles. Des bobines à gradient sont utilisées pour intégrer des données spatiales dans les images. Ces bobines engendrent des gradients de champ magnétique changeants dans tout le corps. Par conséquent, l'IRM scanner est en mesure de localiser avec exactitude les signaux de radiofréquence émis. Ce processus facilite la génération d'images 3D extrêmement détaillées, souvent présentées en niveaux de gris, où les variations de luminosité signalent les fluctuations de densité des atomes d'hydrogène à travers les différents tissus du corps.

1.7.5.2 Acquisition des images IRM

Comme indiqué dans la section 1.7 , l'IRM est la méthode d'imagerie la plus couramment employée pour le diagnostic des tumeurs cérébrales. Elle occupe une place déterminante dans le domaine de l'imagerie médicale en fournissant une représentation particulièrement précise de l'anatomie du cerveau, de la moelle épinière et du système vasculaire lorsque le patient est soumis à un champ magnétique puissant, à l'instar d'un équipement d'IRM (voir figure 2.1). Comme le montre la figure 2.1, un dispositif d'IRM comprend plusieurs composants clés : l'aimant, les bobines de gradient, les bobines de radiofréquence (RF) et la table d'expérimentation. Tous ces facteurs agissent en synergie pour produire des images haute définition de l'anatomie interne du corps. Les bobines RF diffusent et réceptionnent des signaux qui engagent une interaction avec les noyaux d'hydrogène présents dans le corps, générant ainsi des images précises.

Des images de structures internes. Selon les sciences fondamentales, l'eau représente environ 60% de la composition du corps humain adulte, dont 73% pour le cerveau [35]. La composition des molécules d'eau est constituée de deux atomes d'hydrogène et un atome d'oxygène. À l'échelle atomique, chaque atome se compose d'un noyau entouré d'une couche d'électrons en orbite. Il y a des protons dans le noyau, qui ont une charge électrique positive [36]. On peut métaphoriquement assimiler les protons à de petites planètes qui tournent constamment autour d'un axe, un mouvement connu sous le nom de « spin ». Cet état est inhérent à la charge positive portée par le proton [37],[38].

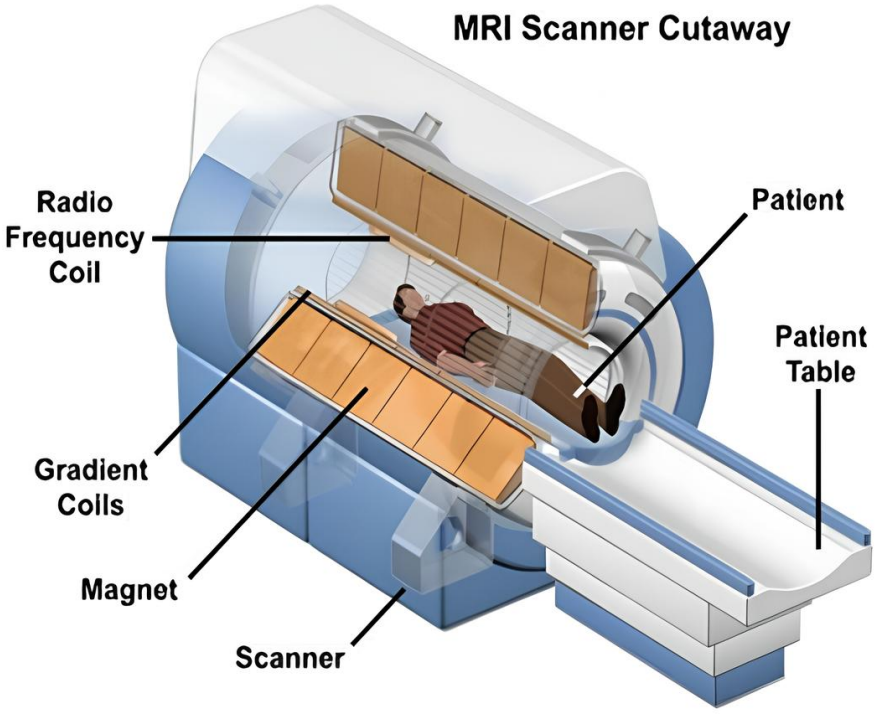


FIGURE 1.10 – Acquisition d'images à l'aide d'un scanner IRM [7].

Leur orientation aléatoire induit une aimantation nette nulle [36]. Néanmoins, lorsque est appliqué un champ magnétique externe comme celui produit par une machine IRM, les petites barres ont tendance à se positionner selon ce champ, ce qui entraîne un vecteur d'aimantation

nette dans la direction du champ magnétique dominant. Cet alignement, ou aimantation, est ensuite perturbé par une impulsion de radiofréquence (RF). Les protons regagnent graduellement leur état de départ à travers divers mécanismes de relaxation, libérant de l'énergie RF lors de ce processus. Les signaux RF générés sont ensuite transformés en images en niveaux de gris grâce à la transformation de Fourier.

L'application de différentes séquences d'impulsions RF, modulées par le temps de répétition (TR) et le temps d'écho (TE), donne lieu à diverses images qui mettent en évidence les propriétés des tissus en tenant compte des temps de relaxation longitudinale (T1) et transversale (T2). Les séquences IRM actuelles, y compris les séquences pondérées en T1 et T2, ainsi que la séquence FLAIR (Fluid Attenuated Inversion Recovery), offrent des renseignements de valeur concernant les attributs anatomiques et pathologiques. Notamment, l'imagerie T1 avec l'administration de gadolinium renforce le contraste, rendant plus aisée l'observation des structures vasculaires et la détection de la rupture de la barrière hématoencéphalique. De surcroît, l'imagerie par diffusion (DWI) joue un rôle crucial dans la détection des AVC aigus en saisissant le déplacement limité des protons de l'eau [[8],[39],[40]].

Dans le cadre du processus de diagnostic, comme illustré par la figure (1.11), les images IRM sont couramment obtenues sur trois axes : latéral, gauche-droite (plan sagittal), antéro-postérieur (plan coronal) et vertical (plan axial) [[8], [40]]. Cela permet d'avoir une vue d'ensemble des structures anatomiques et des anomalies (voir figure 1.12). Ces images offrent aux praticiens une localisation géographique précise et facilitent la préparation de l'intervention chirurgicale ainsi que l'évaluation du soin.

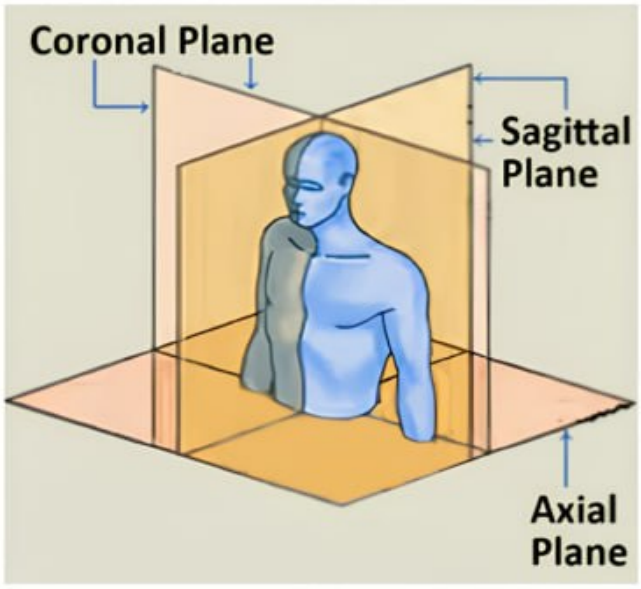


FIGURE 1.11 – Plans d'imagerie médicale anatomique : coronal, sagittal et axial.

Source : Boehringer Ingelheim, HRCT Primer - Reconstruction d'images

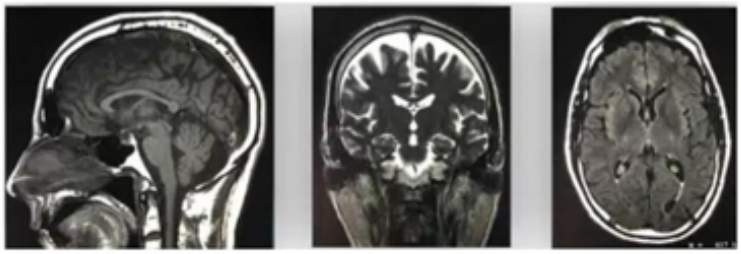


FIGURE 1.12 – Plans axial, sagittal et coronal d'une IRM cérébrale..

Source : Imagerie par résonance magnétique (IRM), TeachMeAnatomy

En manipulant la séquence des impulsions RF appliquées et capturées, on peut produire différents types d'images. Cela se fait par deux paramètres principaux [9] :

— **Temps de répétition (TR)** : désigne l'intervalle de temps entre deux séquences d'impulsions consécutives appliquées sur une même coupe.
— **Temps d'écho (TE)** : temps écoulé entre l'émission de l'impulsion RF et la réception du signal d'écho.

Différents tissus peuvent être décrits par deux temps de relaxation distincts : T1 et T2 [9].

— **T1 (temps de relaxation longitudinale)** : constante de temps qui caractérise la vitesse de retour à l'équilibre des protons excités. Elle correspond au temps nécessaire aux protons pour se réaligner sur le champ magnétique externe.
— **T2 (temps de relaxation transverse)** : constante de temps qui commande la vitesse à laquelle les protons excités reviennent à l'équilibre ou perdent leur synchronisation. T2 est le temps nécessaire pour que les protons en rotation perdent leur cohérence de phase entre les noyaux tournant perpendiculairement au champ magnétique principal.

En outre, l'étude comparative des séquences T1, T2 et FLAIR souligne leurs fonctions distinctes dans la détection du liquide céphalo-rachidien (LCR) et des irrégularités cérébrales, affirmant ainsi leur valeur clinique en imagerie diagnostique (voir la figure 1.13). Ces méthodes avancées d'imagerie, récemment intégrées, aident les cliniciens à diagnostiquer et traiter exactement diverses maladies neurologiques, garantissant ainsi les meilleurs soins possibles pour les patients.

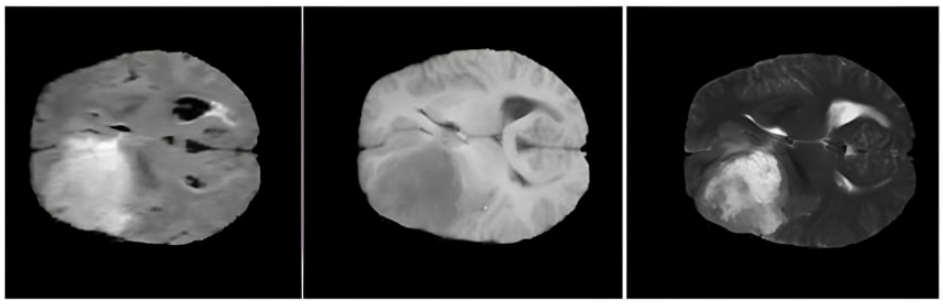


FIGURE 1.13 – Séquences d'images de tumeurs cérébrales IRM de gauche à droite : FLAIR, pondérée T1 et pondérée T2 [8].

1.7.5.3 Modalités d'IRM

En se basant sur ce concept, diverses sortes d'images peuvent être générées en variant les périodes TR et TE et en évaluant les valeurs T1 et T2. On peut en particulier mentionner :

Images pondérées en T1 pendant une séquence T1, les protons qui s'étaient alignés sur le champ magnétique principal reprennent peu à peu leur état d'équilibre. Cette séquence est idéale pour visualiser les structures anatomiques de l'organisme puisque les tissus possédant des temps de relaxation T1 plus élevés apparaissent plus clairs, tandis que ceux ayant des temps de relaxation T1 plus bas paraissent plus sombres [41]. Par exemple, sur les images cérébrales en pondération T1, le tissu cérébral apparaît plus lumineux que le LCR, ce qui

facilite une observation précise des structures cérébrales. Ces images sont produites avec de courts temps TR et TE.

T1ce (T1 avec rehaussement de contraste) T1ce est une version modifiée de la séquence T1 standard, utilisée avec un agent de contraste tel que le gadolinium. Cet agent de contraste est administré au patient et il se concentre dans les zones ciblées, comme les lésions, tumeurs ou inflammations, essentiellement en raison d'une augmentation de la perméabilité des vaisseaux sanguins. Avec les images T1ce, les régions où l'agent de contraste est concentré apparaissent très brillantes, alors que la plupart des tissus gardent leur apparence T1 normale (voir la figure 1.14 pour la comparaison). Cette séquence est essentielle pour repérer des maladies et des blessures qui, autrement, risqueraient de passer inaperçues.

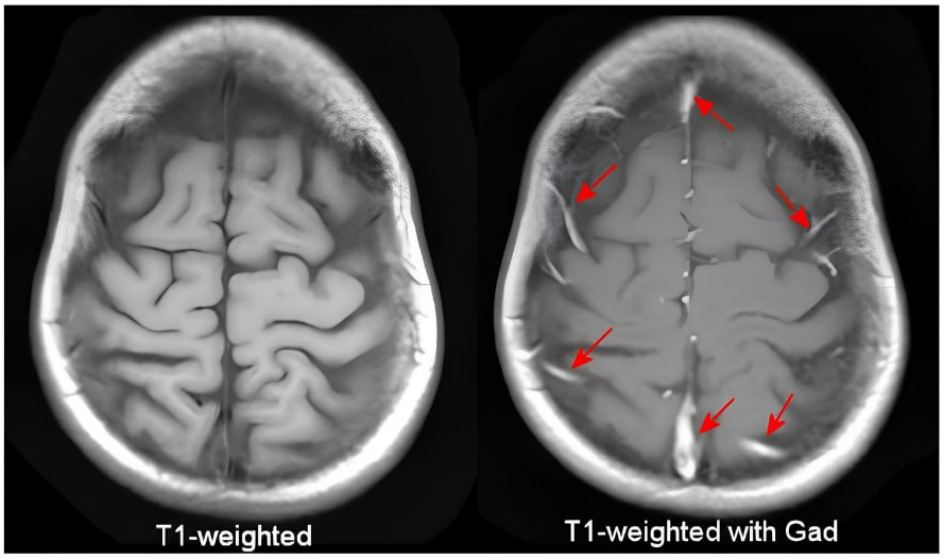


FIGURE 1.14 – Différences de luminosité des vaisseaux sanguins entre les scans T1 et T1ce [9].

Images T2 pondérées : ce sont des images produites avec des durées de répétition et d'écho plus longues, qui mettent en évidence les différences de temps de relaxation T2 (ou relaxation transversale) dans la zone explorée. Sur ces images, les tissus dont le temps de relaxation T2 est allongé apparaissent plus clairs, tandis que ceux dont le temps de relaxation T2 est raccourci ont un aspect plus foncé. On utilise fréquemment la séquence T2 pour identifier les lésions, les inflammations, les œdèmes ainsi que d'autres anomalies affectant les tissus mous. Cette situation résulte du fait que les processus pathologiques provoquent souvent une accumulation de liquide dans les tissus. Sur ces images, avec des temps de relaxation T2 plus étendus, les lésions ou l'œdème cérébral se distinguent par leur brillance accrue par rapport au tissu cérébral adjacent, ce qui facilite le diagnostic et l'évaluation [42]. Sur les images T1 pondérées, le liquide céphalo-rachidien se montre plus brillant.

FLAIR (Fluid Attenuated Inversion Recovery) Proche de la séquence T2, mais avec des temps TR et TE plus longs [9], la séquence FLAIR a été conçue pour supprimer le signal des liquides internes et en particulier celui du liquide cérébro-spinal (LCR) [43]. La suppression de cette substance permet un meilleur aperçu des altérations (ou maladies)

cérébrales en ce sens que le LCR peut masquer des parties essentielles du visuel. Sur ces clichés, le LCR apparaît très sombre, tandis que les tissus adjacents se révèlent d'une clarté accrue, rendant ainsi toute anomalie plus visible. On utilise souvent la séquence FLAIR pour diagnostiquer les maladies neurologiques, notamment les tumeurs cérébrales.

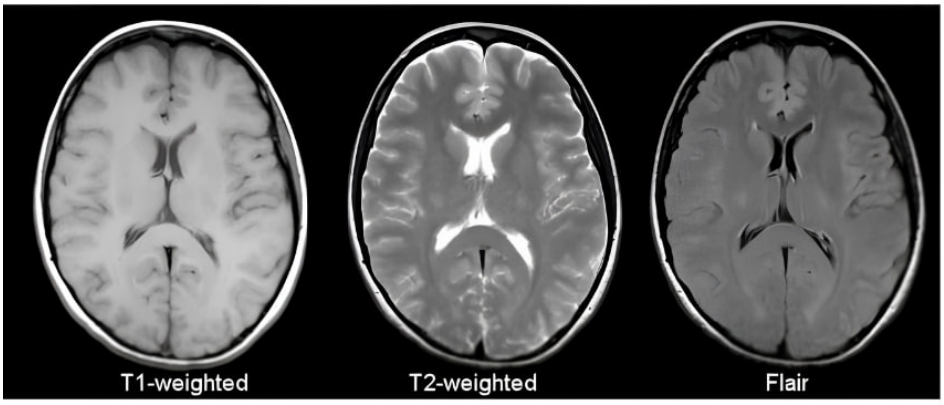


FIGURE 1.15 – Comparaison des plans axiaux T1, T2 et FLAIR [9].

1.7.6 Segmentation des tumeurs cérébrales

La segmentation appropriée des tumeurs cérébrales est une étape essentielle dans l'analyse des images médicales. Elle consiste à détecter et à délimiter les différentes parties tumorales sur les IRM du cerveau. Une segmentation précise est nécessaire pour diagnostiquer, planifier un traitement et surveiller efficacement la progression des tumeurs cérébrales. Le Challenge BraTS, qui consiste à segmenter les tumeurs cérébrales, est l'une des compétitions majeures dans ce domaine. Le challenge BraTS [[44],[45], [28]] est une compétition annuelle dédié à l'analyse d'images médicales. Lancé par la communauté MICCAI (Medical Image Computing and Computer Assisted Interventions), ce projet porte sur la segmentation de tumeurs cérébrales à partir d'images IRM multimodales. L'objectif de ce Le défi est d'encourager le développement et l'amélioration d'algorithmes avancés pour la segmentation des tumeurs cérébrales, tout en ayant pour charge d'évaluer et de comparer l'efficacité des différentes méthodes.

Les participants bénéficient d'examens IRM comprenant diverses modalités, telles que :

- Modalité pondérée en T1 ;
- Modalité pondérée en T2 ;
- Modalité FLAIR (Fluid Attenuated Inversion Recovery) ;
- Modalité pondérée T1 post-contraste (T1ce).

Ces différentes modalités offrent des informations complémentaires pour une segmentation tumorale précise

Le défi porte sur la segmentation des différents types de tumeurs, notamment

- **Necrotic (NCR - label 1)** :C'est une étiquette qui désigne le centre de la tumeur où le tissu est mort (nécrotique), ce qui signifie que les cellules de cette zone ne sont plus vivantes est essentiel d'obtenir une segmentation précise de cette zone centrale pour pouvoir voir la structure interne de la tumeur.

— **Edematous / Invaded tissue (ED - label 2)** : Cette étiquette correspond à la zone autour de la tumeur dans laquelle le tissu cérébral est atteint par un œdème. L'œdème se manifeste par une accumulation de liquide pouvant provoquer un gonflement et des changements dans la densité des tissus. Dans le contexte des tumeurs cérébrales, la région œdémateuse indique généralement les zones où la tumeur a envahi et impacté le tissu cérébral sain environnant. Une segmentation précise de cette région est cruciale pour soigner l'impact de la tumeur sur le cerveau environnant et planifier les interventions chirurgicales ou thérapeutiques.

— **Enhancing Tumor (tumeur rehaussée) (ET - label 4)** : cette étiquette désigne la zone de la tumeur qui se rehausse après injection d'un produit de contraste lors d'une Imagerie par résonance magnétique (IRM). Le noyau tumoral, lorsqu'il est rehaussé par le GD, correspond en général à la zone la plus active « partie de la tumeur croissante et vascularisée ». Il est essentiel de segmenter cette zone de manière fine .Cette région revêt un intérêt particulier pour la planification et l'évaluation du traitement, car elle correspond souvent à la partie la plus agressive et en pleine croissance des cellules tumorales. Une segmentation précise de cette zone est nécessaire et essentiel pour la planification des résections chirurgicales et l'évaluation de la réponse aux traitements.

Le challenge BraTS emploie des mesures d'évaluation précise pour juger la qualité des résultats de segmentation, notamment le coefficient de similarité de Dice (DSC), sur trois sous-régions distinctes. [10] :

— **Enhancing Tumor (ET)** :comme décrit précédemment, il s'agit de la zone déterminée par des régions ex-inhibant l'hyper intensité en T1ce comparativement à T1, ainsi que par opposition à la substance blanche qualifiée de « saine » dans T1ce.

— **Tumor core (TC)** : fait référence à l'amas central de la tumeur, habituellement la région visée pour une intervention chirurgicale. Il inclut à la fois la portion nécrosée de la tumeur et la tumeur en phase de développement.

— **Whole Tumor (WT))** : le champ total de la maladie. Elle englobe le noyau tumoral ainsi que la zone œdémateuse péritumorale.

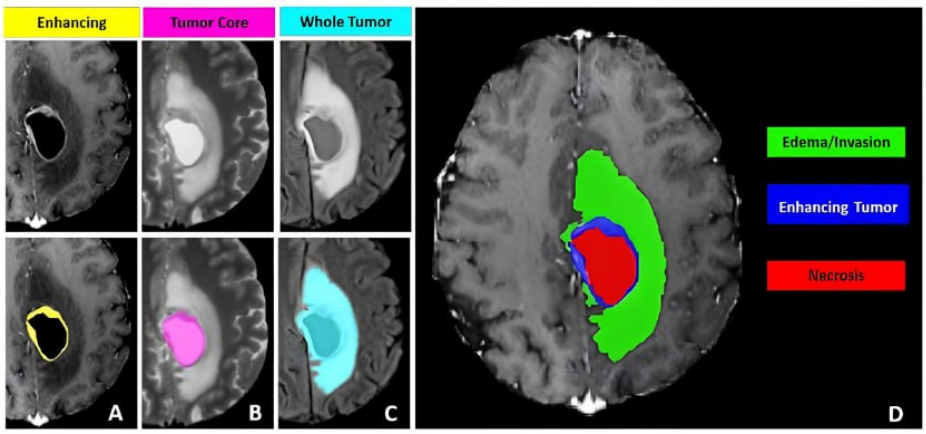


FIGURE 1.16 – Représentation des trois sous-régions (A, B, C) considérées pour l'évaluation et des trois étiquettes (D). [10]

Le défi a permis de faire beaucoup avancer le domaine de la segmentation des tumeurs cérébrales en servant de référence pour les méthodes les plus sophistiquées. Il a soutenu l'évolution des modèles de deep learning, notamment les réseaux de neurones convolutifs (CNN), qui ont montré d'excellents résultats dans les dernières éditions. Les progrès et innovations issus du défi BraTS ont le potentiel d'améliorer la pratique clinique en fournissant des segmentations tumorales plus précises et reproductibles. Cela permet d'améliorer les résultats pour les patients tout en optimisant la planification et le suivi des traitements.

En 2020, Andriy Myronenko et Ali Hatamizadeh de NVIDIA ont proposé une architecture [46] basée sur les réseaux de neurones convolutifs pour la segmentation des tumeurs cérébrales à partir du base de données de l'épreuve BraTS2019. Des équipes venant de différentes régions ont également proposé des méthodes variées [[47],[48],[49],[50],[51]] pour résoudre ce même problème. Au chapitre suivant, nous décrirons les méthodes que nous avons utilisées pour résoudre ce problème, et au chapitre 3, nous comparerons les résultats que nous avons obtenus avec ces études.

Conclusion

En conclusion, comme nous l'avons vu dans les dernières sections de ce chapitre, le domaine médical se tourne de plus en plus vers l'intelligence artificielle. Cette évolution ne fait pas l'exception de l'imagerie médicale qui se trouve profondément bouleversée par les algorithmes d'IA, notamment les réseaux de neurones convolutifs, qui jouent un rôle majeur dans cette révolution. Au chapitre suivant, nous décrirons ces modèles et d'autres architectures fondamentales du deep learning.

Chapitre 2

Approches de Deep Learning pour la segmentation d'images IRM cérébrales

2.1 Introduction globale

L'analyse biomédicale moderne repose en grande partie sur la segmentation des lésions cérébrales à l'aide de l'imagerie par résonance magnétique (IRM). Elle facilite la détection automatique de l'emplacement, de la forme et de l'étendue des zones pathologiques pour soutenir le diagnostic médical, la planification d'interventions chirurgicales et le suivi du traitement. Concernant les tumeurs cérébrales, il est crucial d'effectuer une segmentation précise des sous-régions tumorales, car cela a un impact direct sur la qualité du suivi clinique et l'appréciation de la réponse au traitement[52, 53]. Le principal avantage de la segmentation automatique est qu'elle diminue la nécessité de recourir à une segmentation manuelle effectuée par des spécialistes médicaux. Celle-ci est étendue, onéreuse et fortement tributaire de l'expertise du professionnel. L'étiquetage des IRM cérébrales multi-modales est particulièrement ardu quand les bords tumoraux sont irréguliers, que les tissus présentent une hétérogénéité ou que plusieurs modalités IRM nécessitent une analyse conjointe[54, 55]. Dans ce cadre, les techniques d'apprentissage profond ont radicalement modifié le domaine de la segmentation en médecine. Les architectures de réseaux de neurones convolutionnels (CNN) ont autorisé l'obtention automatique de caractéristiques hiérarchiques à partir des images médicales, tandis que les structures U-Net ont renforcé la précision spatiale grâce à leur design encodeur-décodeur et leurs connexions par saut (skip connections) [56]. Les Transformers et les architectures hybrides CNN-Transformer, plus récents, ont permis une meilleure modélisation du contexte global grâce à l'utilisation du mécanisme d'attention [57, 58, 59]. Ce travail adopte une démarche graduelle pour examiner diverses familles de modèles en vue de la segmentation des lésions cérébrales. On commence par l'implémentation et l'évaluation d'une baseline CNN 2D multi-modale sur la base de données BraTS. Par la suite, les structures U-Net et CNN-Transformer hybrides sont examinées comme des options d'amélioration pour la segmentation des tumeurs cérébrales. L'objectif principal de cette méthodologie est de procéder à une comparaison entre différentes méthodes de segmentation, tout en examinant leurs bénéfices, leurs contraintes et leur faculté à traiter les données multimodales issues des IRM cérébrales.

2.2 Problématique

Pour diverses raisons, la segmentation des lésions cérébrales en IRM constitue une tâche complexe. Le premier défi réside dans la diversité morphologique des tumeurs cérébrales. Les lésions peuvent varier considérablement en taille, avoir des bords irréguliers et renfermer plusieurs sous-régions avec des attributs visuels uniques. Ainsi, une tumeur cérébrale peut comporter un noyau tumoral, une zone œdémateuse et une région intensifiée suite à l'injection de produit de contraste [52, 56]. Le second défi concerne le déséquilibre entre les classes. Dans les images IRM, les zones affectées par la tumeur tendent à représenter une portion modeste du volume global comparé aux tissus cérébraux normaux et au fond général. Il est donc possible qu'un modèle présente une précision globale élevée tout en ne parvenant pas à segmenter correctement les zones pathologiques importantes[55, 60] . La troisième problématique réside dans la diversité des données médicales. Les images IRM peuvent être issues de divers établissements médicaux, appareils de scanner et procédures d'acquisition. Cette variabilité entraîne des modifications importantes dans les distributions d'intensité entre différents patients et modalités IRM[54, 57]. En fin, la segmentation cérébrale

exige une considération approfondie du contexte spatial global. Les CNN traditionnels sont performants pour capter les caractéristiques locales, mais ils rencontrent des contraintes lorsqu'il s'agit de modéliser les dépendances à longue distance entre différentes zones de l'image. Cette contrainte explique l'intégration progressive des architectures U-Net, suivies des Transformers, dans les processus contemporains de segmentation médicale [57, 59] .

2.3 État de l'art

L'automatisation de la segmentation des tumeurs cérébrales à l'aide d'images IRM multimodales représente l'un des défis les plus ardus de l'imagerie médicale numérique. Un gliome correspond à une multiplication anormale de cellules gliales dans le parenchyme du cerveau, classifié selon l'OMS de grades I à IV. Les gliomes de haut grade (GBM, degré IV) sont les plus courants et les plus mortels, se distinguant par une hétérogénéité marquée des tissus.

Le scanner multimodal utilise quatre séquences distinctes : **T1** (anatomie), **T1ce** (zones actives suite à l'injection de gadolinium), **T2** (œdème) et **FLAIR** (infiltration tumorale). Aucune séquence unique ne permet de décrire intégralement la maladie.

La question clé réside dans la manière d'élaborer un système d'analyse automatique qui soit à la fois précis, résilient face à la variabilité morphologique des lésions, et également opérationnel dans un cadre de données restreintes.

Le défi de référence est le challenge **BraTS**, qui inclut **369 patients** annotés par des spécialistes, avec trois zones cliniquement significatives :

— **WT (Whole Tumor)** : tumeur totale ;

— **TC (Tumor Core)** : noyau tumoral ;

— **ET (Enhancing Tumor)** : tumeur rehaussée.

La mesure principale utilisée pour l'évaluation des performances est le **Coefficient de Similarité de Dice (DSC)**.

2.3.1 Évolution Chronologique des Architectures

2.3.1.1 Première génération (2015–2017) : Naissance du Deep Learning en segmentation médicale

Le domaine a été fondé sur l'article *U-Net* [56] (Ronneberger et al., 2015). Il applique l'architecture encodeur-décodeur avec des connexions de contournement, alliant un contexte sémantique approfondi et une précision spatiale fine. Ce modèle est considéré comme l'autorité suprême en segmentation médicale, atteignant un score de Dice WT de 0,70-0,75 lors des premiers benchmarks. La principale restriction est le traitement par tranches, sans prise en compte du contexte global en trois dimensions.

V-Net [61] (Milletari et al., 2016) constitue la première extension 3D de U-Net, introduisant la Dice Loss comme fonction d'entraînement native, ouvrant la voie aux approches volumiques.

ResNet [53] (He et al., 2016) introduit les liaisons résiduelles qui favorisent l'entraînement de réseaux d'une grande profondeur sans problème de dégradé, un élément largement utilisé dans de nombreuses structures médicales.

DeepMedic[62](Kamnitsas et al., 2017), un CNN 3D multi-échelle associé à un CRF dense, affiche les résultats les plus performants sur BraTS 2015 avec un taux de Dice WT/TC/ET d'environ 0,85/0,75/0,72.

2.3.1.2 Deuxième génération (2018–2021) : Enrichissement de U-Net

Attention U-Net [63](Oktay et al., 2018) propose l'utilisation de portes d'attention sur les connexions de rupture, facilitant ainsi la concentration de la reconstruction sur les zones pathologiques et la réduction des faux positifs sur les tissus sains.

DeepLab / ASPP : le module de convolutions dilatées multi-échelles (Atrous Spatial Pyramid Pooling) proposé par[64] (Chen et al. (2017) offre la possibilité de saisir des structures à diverses échelles sans sacrifier la résolution spatiale.

Le standard méthodologique de référence, qui est détaillé dans la section 2, est établi par *nnU-Net* [65](Isensee et al., 2021) .

2.3.1.3 Troisième génération (2021–présent) : L'ère des Transformers

Vaswani et ses collaborateurs (2017) ont établi les bases du mécanisme d'auto-attention multi-têtes (MHSA) par le biais du Transformer, qui offre la capacité de modéliser des dépendances entre toutes les paires de positions dans une séquence, sans tenir compte de leur distance.

Le premier modèle hybride CNN_Transformer destiné à la segmentation médicale, nommé TransUNet, a été élaboré par [66]Chen et al. en 2021.

Le SwinUNETR [59](yang et al., 2022), une version 3D du Transformer Swin hiérarchique, a obtenu des résultats de pointe sur le BraTS 2021 avec des scores Dice WT/TC/ET de 0,931/0,896/0,846, et a été formé sur un ensemble de 1251 patients.

2.3.2 Analyse Critique des Méthodes

2.3.2.1 nnU-Net — Isensee et al. (2021)[65]

nnU-Net a été publié par Isensee et al. dans la revue *Nature Methods* en 2021. Sa contribution majeure consiste à prouver qu'une structure U-Net bien paramétrée, sans élément spécifiquement novateur, dépasse la plupart des techniques expertes. Le système examine de façon autonome les caractéristiques statistiques du base de données (distance entre les voxels, dimension médiane des structures, intensités) et ajuste automatiquement l'architecture, les hyper-paramètres, le traitement préalable et l'inférence — sans nécessiter d'intervention humaine. Cette approche axée sur les données se distingue de la méthode traditionnelle qui consiste à régler les hyperparamètres manuellement.

Performances sur BraTS :

WT = 0,904 (±0,072) ·

TC = 0,830 (±0,127) ·

ET = 0,760 (±0,254)

Limitations :

— Pas d'instauration de mécanisme pour modéliser les dépendances sur le long terme (architecture entièrement convolutive)

— Instabilité marquée sur l'ET : écart-type élevé (±0,254) mettant en évidence une importante variabilité entre les patients

— Dégradation sur des portions limitées de données

2.3.2.2 UNETR — Hatamizadeh et al. (2022) [59]

Selon Hatamizadeh et ses collaborateurs, UNETR a été présenté à WACV 2022 comme l'une des premières architectures entièrement basées sur un Vision Transformer (ViT) afin de remplacer l'encodeur CNN par un ViT pour la segmentation médicale en 3D. Le volume IRM est fragmenté en morceaux 3D de 16^3 voxels (ou tokens), gérés par 12 niveaux d'auto-attention à têtes multiples, et ensuite recomposé à l'aide d'un décodeur U-Net qui reçoit des skip connections des couches [3, 6, 9, 12] du Transformer. Le concept principal est que le mécanisme d'auto-attention s'avère plus approprié que les convolutions pour saisir les relations à long terme dans les volumes IRM, où les zones anatomiques éloignées sont fortement liées.

Performances sur BraTS :

WT = 0,855 (±0,128)

TC = 0,758 (±0,152)

ET = 0,728 (±0,245)

Limitations

— Les performances sont moins bonnes que celles du nnU-Net dans toutes les zones : l'absence totale de convolutions locales nuit à la saisie des détails précis aux interfaces tumorales.

— Le coût en calcul est important en raison de la complexité quadratique $O(N^2)$ de l'attention lorsqu'il s'agit de volumes 3D.

— Écart-type élevé sur l'ET (±0.245) : moins fiable qu'un U-Net 3D dans les situations complexes.

2.3.3 Modèle Vox2Vox — Cirillo, Abramian & Eklund (2020) [67]

Cirillo et ses collègues ont suggéré **Vox2Vox**, un modèle **3D-GAN volume-à-volume** basé sur un réseau antagoniste génératif pour la segmentation des tumeurs cérébrales sur la base de données **BraTS**. Contrairement aux méthodes d'encodeur-décodeur pures, Vox2Vox associe un générateur **U-Net 3D**, chargé de générer la carte de segmentation à partir des quatre modalités d'IRM, et un discriminateur **PatchGAN**, formé pour différencier les segmentations produites des annotations d'experts.

Cette formation antagoniste incite le générateur à créer des masques anatomiquement cohérents et réalistes, en complément des fonctions de perte traditionnelles telles que la perte Dice et la perte adverse. Les résultats finaux sont obtenus en combinant dix modèles Vox2Vox issus d'une validation croisée à dix plis.

Performances sur BraTS (ensemble de test)

— WT (Whole Tumor) : 0,893

— TC (Tumor Core) : 0,822

— ET (Enhancing Tumor) : 0,761

— HD95 : 6,44 / 24,36 / 18,95 mm

Limitations

— Les performances restent inférieures à celles du **nnU-Net** et des architectures hybrides **CNN–Transformer** dans les trois régions, notamment pour la région **TC** (0,822 contre environ 0,830 pour nnU-Net).

— L'entraînement antagoniste est connu pour son instabilité. La difficulté à maintenir l'équilibre entre le générateur et le discriminateur impose l'utilisation d'un ensemble de dix modèles afin d'obtenir des performances compétitives, ce qui augmente considérablement le coût d'entraînement et d'inférence.

— Le modèle ne dispose d'aucune modélisation explicite des dépendances à longue portée puisque le générateur est exclusivement convolutif. Bien que le discriminateur améliore le réalisme des segmentations produites, il ne compense pas l'absence de mécanismes d'attention dans les cas présentant des morphologies tumorales complexes ou atypiques.

2.3.3.1 Swin-UNet — Cao et al. (2022)[68]

Cao et consorts ont déposé Swin-UNet à l'ECCV 2022, se présentant comme le tout premier Transformer pur (sans recours à une seule couche de convolutions) dédié à la segmentation médicale en 2D. Le Swin Transformer aborde la question de la complexité quadratique du ViT en effectuant le calcul d'auto-attention à travers des fenêtres locales décalées, ce qui réduit la complexité de $O(N^2)$ à $O(N)$.L'architecture utilise une structure encodeur-décodeur symétrique qui emploie le patch merging pour le sous-échantillonnage.

Deux variantes sont évaluées sur BraTS :

- Swin-UNet (pré-entraîné sur ImageNet-22k) :

WT = 0,872 (±0,102)

TC = 0,799 (±0,136)

ET = 0,785 (±0,202)

- SwinUNet (from scratch) :

WT = 0,770 (±0,176)

TC = 0,661 (±0,184)

ET = 0,651 (±0,222)

L'écart significatif de plus de 10 points de Dice entre les deux variantes met en évidence la forte dépendance des Transformers purs à la préformation pour pallier l'absence de biais inductif de localité.

Limitations :

• Architecture 2D qui gère les sections axiales de manière indépendante, sans tenir compte de la continuité volumétrique en 3D

• Décroissance significative des performances en l'absence de pré-entraînement (environ 12 points de Dice perdu)

• Grand écart-type sur l'ET (±0,202 dans la version pré-entrainée) : forte variabilité entre les patients.

2.3.3.2 TransBTS — Wang et al. (2021)[51]

À MICCAI 2021, Wang et ses partenaires ont présenté TransBTS, considérée comme l'une des premières architectures hybrides CNN-Transformer spécialement élaborées pour la segmentation des tumeurs cérébrales multimodales. TransBTS adopte une stratégie hybride, par opposition à UNETR qui élimine totalement le CNN : l'encodeur 3D CNN s'occupe des attributs locaux (bords, textures), pendant qu'un module Transformer intégré au niveau du cou bottleneck simule les relations à longue distance entre des zones éloignées. Les quatre types d'IRM (FLAIR, T1, T1ce, T2) sont enchaînés à l'entrée, et le décodeur refait la segmentation par le biais de convolutions inversées avec des connexions de contournement.

Performances sur BraTS : WT = 0,886 (±0,112) · TC = 0,818 (±0,140) · ET = 0,730 (±0,250)

Limitations :

• Sans mécanisme d'attention ciblé autour des skip connections (Attention Gates).

• La méthode hybride la moins performante (0,730) ; avec un écart-type élevé (±0,250).

• Énorme variabilité inter-patient observée sur les cas atypiques d'un point de vue morphologique.

2.3.3.3 HUT — Soh, Yuen Rajapakse (2023)[69]

HUT (Hybrid UNet Transformer) a été publié par Soh et ses collaborateurs dans la revue Heliyon (Elsevier) en 2023. HUT utilise une méthode hybride plus avancée que TransBTS : elle inclut des procédés d'attention à divers échelons de l'architecture, alliant des segments convolutifs 3D pour l'extraction locale, des modules d'attention par canal pour le choix des caractéristiques pertinentes, et des niveaux de Transformer pour la modélisation des dépendances spatiales à large échelle. HUT a reçu une validation sur deux jeux de données : BraTS et la base de données ATLAS [62] (lésions post-AVC), ce qui témoigne d'une solide capacité à généraliser entre diverses pathologies cérébrales.

Performances sur BraTS :

WT = 0,900 (±0,086) · TC = 0,836 (±0,126) · ET = 0,783 (±0,239)

Limitations :

• Il reste une instabilité inter-patient résiduelle, avec un écart-type sur l'ET qui est encore relativement élevé (±0,239).

• Le TC et l'ET ne rivalisent pas avec les performances des architectures hybrides 3D les plus avancées qui intègrent une supervision approfondie et l'ASPP 3D.

2.4 Analyse Thématique et Méthodologique

2.4.1 Axes thématiques

La littérature s'organise autour de quatre problématiques principales :

• Configuration du réseau : CNNs purs (VGG, ResNet), structures encodeur-décodeur (U-Net, V-Net) qui représentent le standard actuel, et des modèles hybrides CNN-Transformer (TransUNet, SwinUNETR) qui représentent l'avant-garde de la recherche en cours.

• Élaboration pour gérer le déséquilibre des classes : de la Dice Loss (indifférente au déséquilibre) à la Focal Loss (axée sur les exemples complexes), en passant par les arrangements hybrides (Dice + CE + Focal + Lovász + Boundary).

• Méthode d'entrée : fusion précoce des 4 canaux (FLAIR + T1 + T1ce + T2) constitue l'approche standard BraTS.

• Cadre spatial : les méthodes volumétriques en 3D sont plus performantes mais demandent une grande quantité de mémoire, ce qui nécessite une formation par morceaux.

2.4.2 Choix méthodologiques communs

Les méthodes de référence ont plusieurs principes en commun : la normalisation par score Z ($\mu = 0$, $\sigma = 1$) par modalité afin d'harmoniser les distributions d'intensité, l'usage de l'augmentation de données (rotations, flips, déformations élastiques, bruit gaussien) pour une meilleure généralisation, l'emploi des optimisateurs Adam/AdamW associés à des planificateurs CosineAnnealing, l'inférence par fenêtre coulissante avec superposition pour atténuer les artefacts aux bords, et une validation rigoureusement séparée par patient. Les indices d'évaluation intègrent le Coefficient de Dice (DSC) ainsi que la Distance de Hausdorff à 95 (HD95).

2.4.3 Tableau Comparatif des Performances DSC (moyenne ± écart-type) sur BraTS

Méthode	**Type d'arch.**	**Modélisation**	**Patients**	**Whole Tumor**	**Tumor Core**	**Enhanc. Tumor**	**Réf.**
nnU-Net	CNN 3D auto	Locale	369	**0.904 ±0.072**	0.830 ±0.127	0.760 ±0.254	[57]
HUT	Hybride multi-niv.	Locale + Globale	369	0.900 ±0.086	0.836 ±0.126	0.783 ±0.239	[61]
Vox2Vox	CNN 3D bout-en-bout	Locale	369	0.893 ±0.071	0.822 ±0.122	0.761 ±0.249	[63]
TransBTS	Hybride bottleneck	Locale + Globale	369	0.886 ±0.112	0.818 ±0.140	0.730 ±0.250	[60]
SwinUNet†	Transformer hiérarch.	Globale + préentr.	369 + ext.	0.872 ±0.102	0.799 ±0.136	0.785 ±0.202	[59]
UNETR	Transformer encodeur	Globale uniquement	369	0.855 ±0.128	0.758 ±0.152	0.728 ±0.245	[58]
SwinUNet	Transformer hiérarch.	Globale	369	0.770 ±0.176	0.661 ±0.184	0.651 ±0.222	[59]
HybridUNet3D(notre)	Hybride 3D résiduel	Locale + Attention	40	**0.848 ±0.082**	**0.862 ±0.091**	**0.805 ±0.085**	—
TransUNet3D(notre)	Hybride CNN-Trans.	Locale + Globale	80	**0.852 ±0.089**	**0.862 ±0.095 ★**	**0.791 ±0.071 ★**	—

★ Meilleur score parmi toutes les méthodes comparées.
† Avec préentraînement sur données médicales externes.
Les scores en vert représentent les meilleures valeurs dans chaque colonne.

TABLE 2.1 – Comparaison des performances sur la base de données BraTS.

2.5 Lacunes Identifiées dans la Littérature

L'examen approfondi de la littérature met en évidence trois déficiences principales que notre recherche cherche à combler :

— **Lacune 1 — Recours à d'importantes bases de données :** nombre d'études s'appuient sur le dataset BraTS intégral (de 369 à 1251 patients). Des techniques telles que nnU-Net et HUT sont formées sur l'ensemble de BraTS, un échantillon qui ne reflète pas fidèlement les circonstances réelles d'un environnement hospitalier où les données annotées sont limitées et onéreuses.

— **Lacune 2 — Manque de progression comparative systématique :** peu de recherches offrent une évaluation graduelle de la baseline 2D jusqu'aux configurations hybrides 3D qui aident à isoler la contribution marginale de chacune des composantes architecturales.

— **Lacune 3 — Transformer inadaptés à des ensembles de données restreints :** Swin-nUNETR et ses homologues requièrent une formation préalable sur des milliers d'images médicales. Comme le montre clairement le cas de Swin-UNet sans pré-entraînement (effondrement de 12 points de Dice), ces architectures ne sont pas directement applicables dans des contextes à faible volume de données.

2.6 base de données BraTS

la base de données BraTS (Brain Tumor Segmentation Challenge) est un standard mondial pour juger les techniques de segmentation des tumeurs cérébrales [70]. Il comprend des IRM multimodales de patients souffrant de gliomes, accompagnées d'annotations expertes effectuées manuellement. 369 patients ont été annotés dans le sous-ensemble d'entraînement. Chaque cas est illustré par un volume en trois dimensions mesurant 240 x 240 x 155 voxels. L'ensembles de données propose quatre modalités IRM additionnelles :

— FLAIR ;

— T1 ;

— T1ce ;

— T2.

Chaque méthode fournit des renseignements spécifiques en anatomie ou en pathologie :

— La modalité T1 offre une perspective anatomique globale du cerveau ;

— T1ce souligne les zones accentuées suite à l'injection de gadolinium. T2 est réceptif aux liquides et aux changements tissulaires ;

— FLAIR accentue la clarté de l'œdème pértumoral.

BraTS fournit des annotations qui distinguent diverses sous-régions tumorales :

— WT (Toute la Tumeur) : tumeur intégrale ;

— TC (Noyau Tumoral) : centre tumoral ;

— ET (Tumeur Rehaussée) : zone tumorale amplifiée.

2.7 Traitement préalable des données

Le prétraitement est une phase cruciale dans le processus de segmentation en médecine. Cela contribue à minimiser les fluctuations entre les patients et à renforcer la stabilité de l'apprentissage. Cette recherche implique l'application de multiples procédures de prétraitement :

— Interprétation des volumes IRM au format NIfTI ;

— obtention des tranches 2D à partir des volumes en 3D ;

— normalisation des intensités ;

— réattribution des étiquettes BraTS ;

— élaboration des tenseurs multi-canaux.

2.7.1 Obtention des tranches 2D

Les volumes IRM 3D sont sectionnés en tranches axiales 2D pour diminuer la complexité de calcul du modèle. Chaque coupe est gérée de manière autonome. Il existe plusieurs bénéfices à cette stratégie :

— Diminution de l'utilisation de la mémoire ;

— accélération du processus d'entraînement ;

— simplification de la structure ;

— capacité à utiliser des CNN 2D plus légers.

Toutefois, cette méthode induit aussi une perte du contexte tridimensionnel volumineux.

2.7.2 Normalisation par le score Z

Une normalisation par score Z est mise en œuvre pour uniformiser les distributions d'intensité entre les patients et les modalités. On définit la normalisation par :

$$x_{\text{norm}} = \frac{x - \mu}{\sigma + \varepsilon}$$

où :

— symbolise la valeur moyenne des intensités ;

— indique l'écart-type ;

— renvoie à une constante offrant une stabilité numérique.

Cette phase stabilise l'apprentissage du réseau et favorise une meilleure convergence [54, 60].

2.7.3 Remappage des labels

Dans BraTS, les étiquettes initiales correspondent aux valeurs 0, 1, 2, 4. Pour faciliter l'apprentissage multi-classe, une reconfiguration est mise en place :

— 0 : fond ;

— 1 : tumeur cardiaque ;

— 2 : œdème ;

— 3 : tumeur accentuée.

— Ainsi, l'étiquette 4 devient l'étiquette 3.

2.7.4 Division en ensembles d'entraînement et de validation

La segmentation des données est effectuée en suivant une approche basée sur le patient, avec : 80% des patients sont destinés à l'entraînement, tandis que 20% le sont pour la validation. Cette approche bloque les divulgations d'information entre les jeux de données et assure une appréciation plus authentique de la capacité du modèle à généraliser.

2.8 Les Neurones Artificiels

2.8.1 Le Neurone Formel

Le neurone formel, qui représente l'unité de base d'un Réseau de Neurones Artificiels (RNA), est constitué de deux éléments[71] :

— Évaluation de la stimulation reçue (fonction E) Évaluation de son activation en utilisant la fonction f .

Il se distingue par :

— son état X (binaire, discret, continu) ;

— le niveau d'activation obtenu en entrée U ;

— le poids des connexions d'entrée W .

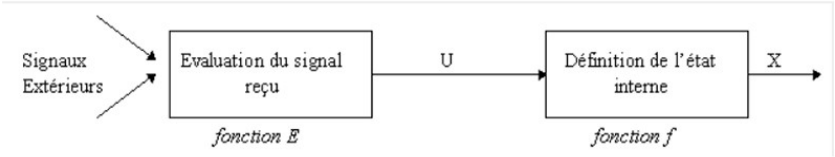


FIGURE 2.1 – les composants d'un neurone formel.

2.8.2 La fonction d'entrée

Le total pondéré des signaux d'entrée :

$$U_i = E(X_1, \ldots, X_i, \ldots, X_N) = \sum_{i=1}^{N} X_i W_i$$

2.8.3 Le biais d'entrée

Unité imaginaire dont la valeur sert à ajuster le seuil d'activation du neurone en modifiant les poids, avec une initialisation fixée à 1.

2.8.4 Le Perceptron

Il s'agit d'un modèle employé pour la classification binaire, reposant sur une relation simple entrée-sortie.

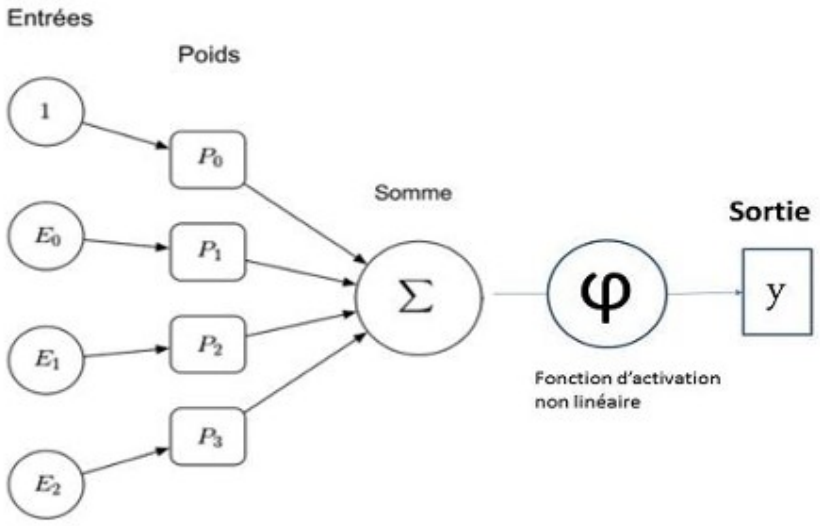


FIGURE 2.2 – schéma du perceptron.

— Chaque entrée a un poids attribué.

— Le résultat dépend à la fois des entrées et du poids.

Il possède une couche consacrée à la perception et une autre dédiée à la prise de décisions.

2.8.5 Réseau de neurones(perceptron) multicouche

Le PMC est un réseau de couches superposées apte à approximer n'importe quelle fonction continue, sous réserve que l'on détermine correctement le nombre de neurones dans la couche cachée.

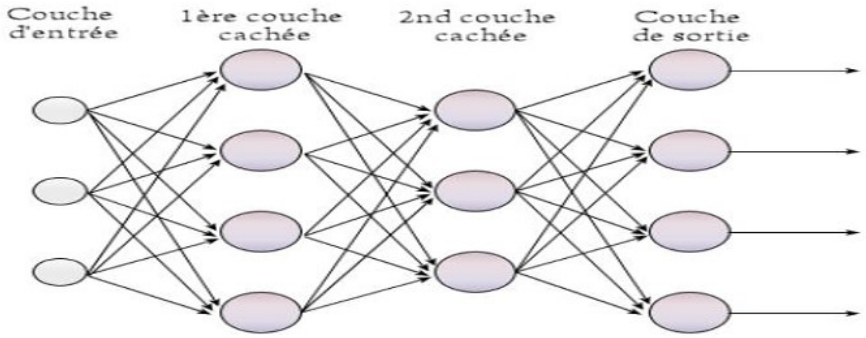


FIGURE 2.3 – perceptron multicouche .

2.9 Réseau de neurones artificiels

Avant de discuter de l'architecture profonde employée dans cette étude, il est crucial de revisiter les concepts de base des réseaux de neurones artificiels (RNA). S'inspirant des systèmes biologiques [72], les RNA constituent des modèles informatiques destinés à reproduire la structure et le mécanisme de fonctionnement du cerveau humain. À l'instar des neurones biologiques — où les dendrites recueillent les données, le corps cellulaire (soma) les analyse à travers des fonctions d'activation, et les axones acheminent les signaux vers les synapses , les RNA sont organisés en strates de neurones interconnectés imitant ces processus cellulaires.

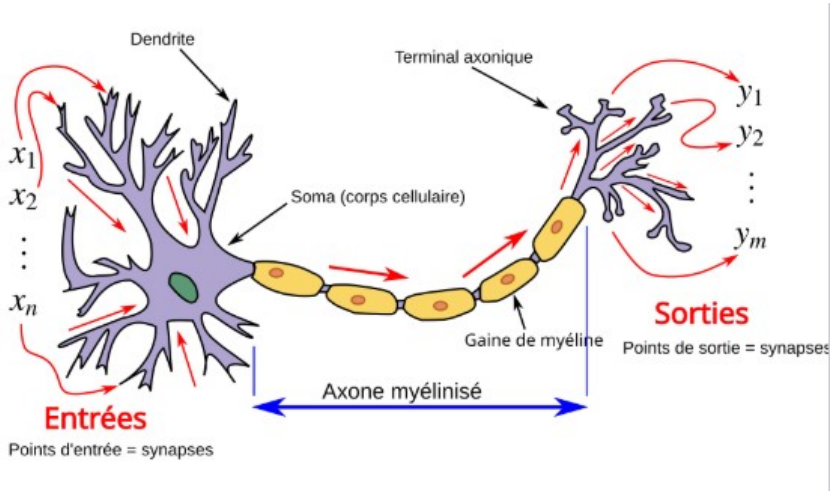


FIGURE 2.4 – Dendrites, soma et axone myélinisé, avec un flux de signal des entrées aux dendrites aux sorties aux bornes des axones. Le signal est une courte impulsion électrique appelée potentiel d'action ou impulsion.[11]

Une architecture traditionnelle de réseau de neurones se compose d'une couche d'entrée, d'une ou plusieurs couches internes et d'une couche finale [73] (comme le démontre la figure 2.6). Chaque niveau contient des unités (neurones) qui réalisent une transformation précise : la couche d'entrée gère les signaux bruts, alors que les couches intermédiaires et de sortie s'occupent des signaux provenant des couches antérieures. Si l'excitation conjointe excède un niveau spécifique suite à l'activation d'une fonction, le neurone émet un signal, imitant de cette façon le processus biologique synaptique[74]. Ces connexions interneuronales sont quantifiées par des poids numériques[75]. D'un point de vue mathématique, la sortie d'un neurone est équivalente à la somme pondérée de ses entrées, modifiée par une fonction d'activation[76]. À l'échelon initial, chaque élément est lié à une caractéristique spécifique des données[72] ; ainsi, son nombre total représente la taille de l'espace des descripteurs [77]. Ces réseaux ont la faculté de gérer également des données structurées (formats tabulaires, CSV, bases de données) et non structurées (images, vidéos, signaux IoT)[78]. Par exemple, dans le cas de la reconnaissance d'images, chaque pixel représente une donnée d'entrée, alors que le nombre de neurones dans la couche de sortie est précisé par l'aspect du problème[72] (comme trois sorties pour distinguer une image entre un chat, un chien ou aucun des deux [72]).

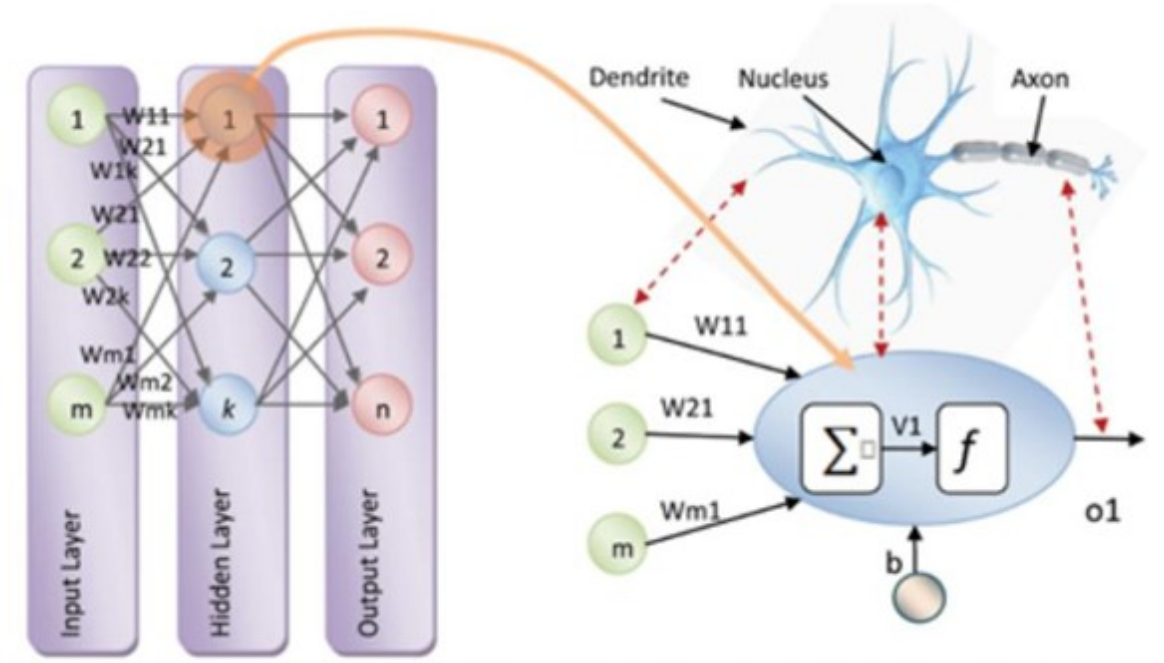


FIGURE 2.5 – Mise en parallèle de la structure d'un réseau de neurones artificiel (RNA) et celle d'un neurone biologique : le RNA est constitué de couches d'entrée, cachées et de sortie, où les neurones sont reliés par des connexions pondérées.Par contre, le neurone biologique possède des dendrites, un noyau et un axone qui assurent la transmission du signal[12]..

Il existe plusieurs similitudes notables entre les neurones biologiques et les neurones artificiels. Dans les neurones biologiques (voir Figure 2.5), les entrées sont reçues via les dendrites, tandis que les neurones artificiels traitent ces entrées. D'après la référence [79], le cœur d'un neurone biologique est représenté par le noyau de la cellule ou le soma, qui compile ces entrées pour décider si le neurone sera stimulé ou freiné. Si l'entrée combinée dépasse un certain seuil, le neurone s'active et transmet un signal le long de son axone. Tout comme les neurones biologiques, les nœuds dans les réseaux de neurones artificiels agrègent également plusieurs entrées, font la somme et traitent le résultat à l'aide d'une fonction d'activation. Si la somme des entrées pondérées atteint ou dépasse un certain seuil (θ), le neurone « s'active » et délivre une sortie de 1 [79]. Sinon, elle ne se déclenche pas et produit une sortie de 0. Cela peut être représenté par un modèle mathématique comme expliqué dans la section 2.11 La communication entre neurones biologiques se fait par le biais des synapses, tandis que dans les neurones artificiels, la force des connexions est déterminée . En fin de compte, les neurones biologiques transmettent des signaux via les axones, à l'instar du fonctionnement des neurones artificiels. D'après l'étude mentionnée dans[80], la comparaison entre les réseaux neuronaux biologiques et les réseaux de neurones artificiels est clairement illustrée dans le tableau

2.2. Cette comparaison est aussi présentée sous forme de diagramme dans la Figure 2.5.

Réseaux neuronaux biologiques	Réseaux neuronaux artificiels
Soma (corps cellulaire) : Intègre les signaux entrants et détermine l'activité neuronale.	**Nœud** : Agit comme unité de traitement, analogue au soma.
Dendrites : Elles recueillent les signaux provenant des neurones environnants.	**Entrée** : Représente les données entrantes alimentant le réseau.
Synapse : Jonction où les signaux sont transmis entre les neurones, influencée par la force ou le poids.	**Poids** : Représente la force des connexions entre les nœuds.
Axone : Transmet le signal de sortie du neurone aux autres.	**Sortie** : Représente la valeur finale calculée et transmise à la couche suivante.

TABLE 2.2 – Comparaison entre les réseaux neuronaux biologiques et artificiels

2.9.1 Types de réseaux neuronaux

On classe généralement les réseaux de neurones en trois types principaux selon leur conception architecturale [78, 81] :

1. **Réseaux de Neurones à Propagation Avant (FFNNs)**
   Les FFNNs font partie des architectures de réseaux neuronaux les plus répandues [78] , où les données circulent dans une seule direction [80], débutant au niveau de la couche d'entrée et progressant en séquence à travers une ou plusieurs couches cachées jusqu'à la couche de sortie [78, 81]. Une caractéristique essentielle des FFNNs est l'absence de cycles ou de rétroactions entre les neurones (nœuds). Chaque couche cachée modifie l'entrée qu'elle reçoit, produisant une nouvelle représentation qui est transmise à la couche suivante [78]. Les fonctions d'activation non linéaires les plus couramment utilisées, à savoir sigmoid, tanh, ReLU, SELU et softmax, sont appliquées à la somme pondérée des entrées pour chaque unité (nœud ou neurone).
2. **Réseaux de neurones convolutifs (CNN ou ConvNets)**
   Les CNN sont conçus pour traiter des données ayant une structure en grille, comme les images [81], et utilisent des couches convolutionnelles pour réaliser une extraction de caractéristiques à plusieurs niveaux [78]. Section 2.10 classe ces modèles en tant que méthodes d'apprentissage profond et Section 2.13 les examine plus en profondeur.
3. **Réseaux de Neurones Récurrents (RNNs)**

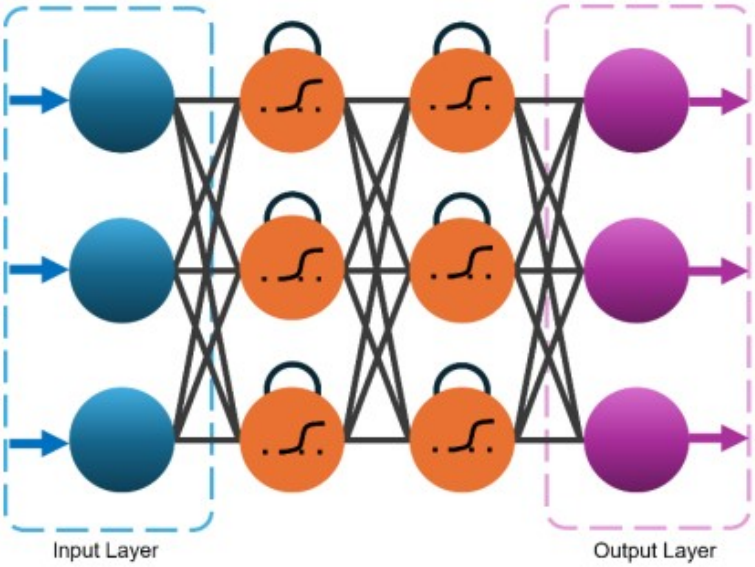


FIGURE 2.6 – Architecture of an RNN.

Les RNNs comportent une couche cachée à chaque étape temporelle, ce qui leur permet de garder l'information des entrées précédentes à travers les séquences temporelles. Bien qu'efficace pour la séquence La modélisation et l'entraînement des RNNs présente des difficultés, notamment face à des problématiques comme le problème du gradient qui disparaît. Des architectures avancées comme la Mémoire à Long Terme et à Court Terme (LSTM) et les Unités Récurrentes à Porte (GRUs) ont été conçues pour surmonter ces contraintes [78].

2.10 Apprentissage profond

L'apprentissage profond (DL), une méthode d'apprentissage automatique basée sur les ANN, est en mesure de traiter de grandes quantités de données brutes et hautement dimensionnelles pour construire un modèle[82]. Cela est réalisé grâce à sa capacité d'apprendre automatiquement les caractéristiques, ce qui permet de surmonter la dépendance à l'ingénierie des caractéristiques manuelles ou faites à la main. Cela rend l'apprentissage profond particulièrement bénéfique pour le traitement de données à grande échelle, bruyantes ou non structurées. Dans l'apprentissage profond, le processus d'apprentissage des caractéristiques se déroule généralement de manière hiérarchique, où les caractéristiques sont progressivement apprises à partir des caractéristiques basiques jusqu'aux caractéristiques abstraites de haut niveau [82]. En nous basant sur cette référence, nous avons abordé les architectures populaires du deep learning : les CNN, les réseaux de neurones récurrents (RNN), tels que LSTM, la représentation distribuée, les réseaux antagonistes génératifs (GAN), et les autoencodeurs.

1. **Réseaux de Neurones Convolutionnels (RNC)**
   Les réseaux de neurones convolutifs sont largement privilégiés pour les tâches liées à la vision par ordinateur et à la reconnaissance vocale. On les applique principalement à des ensembles de données présentant des relations spatiales, comme c'est le cas pour les données d'image où l'agencement des lignes et des colonnes conserve une structure pertinente qui ne peut être modifiée [82].

2. **Réseaux de Neurones Récurrents (RNN)**
   Les RNNs sont des réseaux de neurones spécifiquement élaborés pour traiter des données séquentielles comme les séries temporelles, les séquences d'événements et le langage naturel [82]. Contrairement aux réseaux neuronaux traditionnels, les RNN intègrent des connexions de rétroaction, ce qui leur permet d'acquérir des motifs séquentiels et de modéliser des dépendances temporelles en conservant une mémoire sur le long terme. Toutefois, l'architecture RNN simple est souvent confrontée aux problèmes déjà évoqués dans la section 2.9.1.

3. **Représentations distribuées**
   Les représentations distribuées sont essentielles pour l'apprentissage des caractéristiques et la modélisation du langage dans les tâches de traitement du langage naturel (NLP). Cette méthode transforme des éléments linguistiques, tels que les mots, les expressions et les phrases, en vecteurs numériques dans un espace sémantique unifié, souvent appelé embeddings [82]. Un exemple important est l'utilisation de Word Embeddings, qui représentent des mots discrets sous forme de vecteurs caractéristiques denses et de faible dimension.

4. **Réseaux antagonistes génératifs (GANs)**
   Les GANs représentent une catégorie de modèles génératifs élaborés pour assimiler la distribution probabiliste sous-jacente d'un ensemble de données d'entraînement, ce qui leur permet de générer de nouvelles instances de données qui présentent des caractéristiques semblables à celles du base de données d'entrée, tout en présentant un certain degré de variation. Une architecture GAN typique se compose de deux réseaux neuronaux adversariaux : un générateur, qui vise à reproduire la distribution des données et créer des échantillons synthétiques réalistes, et un discriminateur, qui vérifie si un échantillon donné est authentique (provenant des données d'entraînement réelles) ou artificiel (produit par le générateur)[82].

5. **Autoencodeurs**
   Ces autoencodeurs sont des architectures de réseaux neuronaux qui, tout comme les embeddings de mots, produisent des représentations denses des données en entrée. Contrairement aux embeddings spécifiquement conçus pour le langage naturel, les autoencodeurs sont polyvalents et peuvent être utilisés pour différents types de données[82]. Leur structure comprend généralement deux éléments principaux : un encodeur, qui modifie l'entrée en un espace latent de dimension inférieure, et un décodeur, qui reconstitue l'entrée d'origine à partir de cette représentation réduite. Ce processus incite le réseau à conserver uniquement les caractéristiques les plus significatives et à éliminer le bruit sans importance.

2.11 Méthodes de Propagation Avant et Rétropropagation dans les ANN

Pour s'assurer que les réseaux de neurones peuvent représenter efficacement des fonctions non linéaires, une fonction d'activation est appliquée à chaque couche. Une des fonctions fréquemment utilisées est la fonction sigmoïde (logistique), qui est mathématiquement définie par l'équation (2.1) :

$$f(x) = \frac{1}{1 + e^{-x}} \tag{2.1}$$

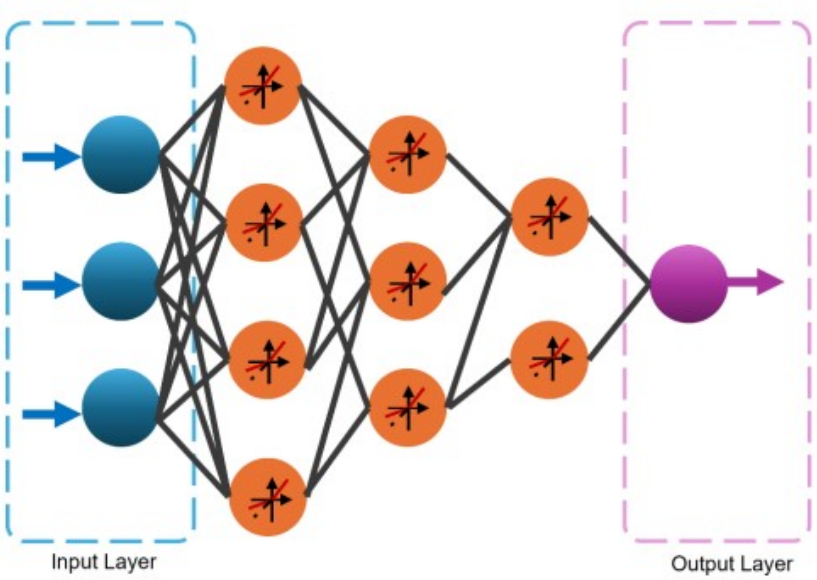


FIGURE 2.7 – Architecture d'un ANN[13]

Une autre fonction d'activation couramment utilisée est la fonction ReLU (Rectified Linear Unit), qui retourne zéro pour les entrées négatives et transmet les valeurs positives sans modification. C'est défini comme l'équation (2.2) :

$$f(x) = \max(0, x) \tag{2.2}$$

Figure 2.8 illustre la représentation graphique de ces fonctions. En introduisant de la non-linéarité via ces fonctions d'activation, les réseaux neuronaux sont à même de classer des données qui ne sont pas séparables de manière linéaire. La propagation avant désigne l'acheminement de l'information à travers un réseau de neurones, en commençant par la couche d'entrée et en avançant vers la couche de sortie[13]. Chaque neurone reçoit des entrées provenant des sorties de tous les neurones de la couche précédente, calcule la somme pondérée des entrées, applique une fonction d'activation et transmet la sortie à la couche suivante [13]. On peut modéliser mathématiquement un neurone dans les Figures 2.6 et 2.7 comme suit :

$$a_i = f\left(\sum_i^n w_i x_i + b\right) \tag{2.1}$$

où x_i représente l'entrée, w_i les poids, b le biais, f la fonction d'activation, et a_i l'activation (sortie) du neurone [13].

En considérant l'ensemble des neurones et des couches, la sortie du neurone j de la $l$ème couche peut être modélisée comme suit :

$$a_j^l = f\left(\sum_i w_{ij}^l a_i^{l-1} + b_j^l\right) \tag{2.2}$$

Ici, w_{ij}^l le poids reliant le neurone i de la couche $(l-1)$ au neurone j de la couche l est noté. $a_i^{(l-1)}$ est l'activation de la couche précédente, et b_j^l est le biais [80]. Pour généraliser la sortie finale du réseau, la sortie du neurone j de la couche de sortie s'exprime comme suit :

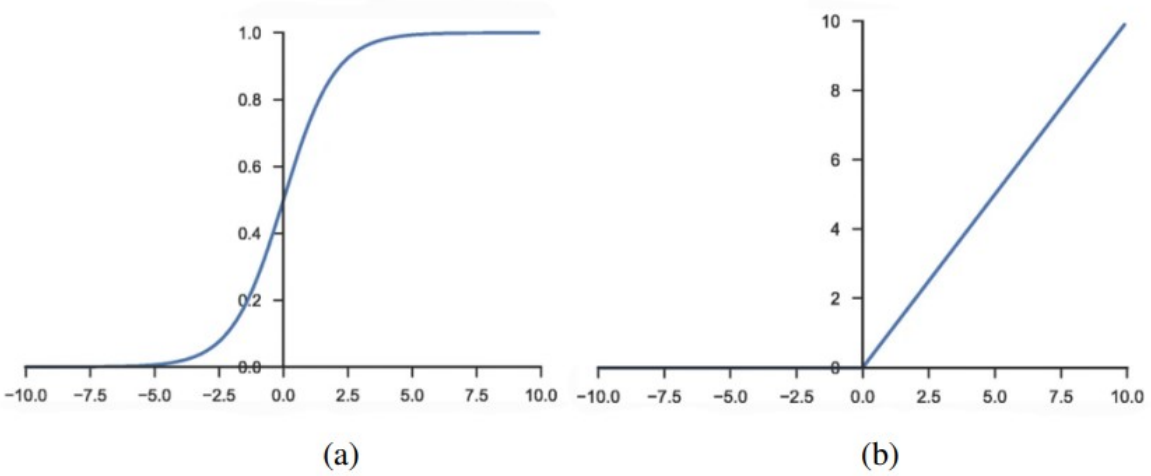


FIGURE 2.8 – Illustration graphique des fonctions (a) sigmoïde et (b) ReLU [14].

$$(\hat{O}_j) = f\left(\sum_i w_{ij} a_i^{l-1} + b_j\right) \tag{2.3}$$

La rétropropagation est une méthode d'optimisation des réseaux de neurones qui ajuste les poids et les biais afin de minimiser la fonction de perte. Elle calcule les gradients (dérivées partielles) de la perte par rapport aux poids et aux biais et les met à jour à l'aide de techniques d'optimisation telles que la descente de gradient[13].

Pendant l'entraînement, le processus de rétropropagation (BP) compare la sortie du modèle aux valeurs cibles réelles à l'aide d'une fonction de perte. Les deux fonctions de perte les plus utilisées en apprentissage automatique sont l'erreur quadratique moyenne (MSE) et l'entropie croisée (CE).

$$E_{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - (\hat{O}_i))^2 \tag{2.4}$$

où y_i représente la valeur réelle, $\hat{O}_i$ la sortie prédite et N le nombre total d'échantillons[13].

$$E_{CE} = -\frac{1}{N}\sum_{i=1}^{N}\Big[y_i \log(\hat{O}_i) + (1 - y_i)\log(1 - \hat{O}_i)\Big] \tag{2.5}$$

Les gradients de la fonction de perte par rapport aux poids et aux biais sont utilisés pour calculer mises à jour :

$$\Delta w_{ij}^l = -\eta \frac{\partial E}{\partial w_{ij}^l} \tag{2.6}$$

$$\Delta b_j^l = -\eta \frac{\partial E}{\partial b_j^l} \tag{2.7}$$

où η désigne le taux d'apprentissage et E la fonction de perte choisie. Les mises à jour des poids et des biais sont effectuées de manière itérative :

$$w_{ij}^l(t+1) = w_{ij}^l(t) + \Delta w_{ij}^l \tag{2.8}$$

$$b_j^l(t+1) = b_j^l(t) + \Delta b_j^l \tag{2.9}$$

Ici, $w_{ij}^l(t)$ et $b_j^l(t)$ représentent les poids et les biais à l'étape temporelle t, et Δw_{ij}^l, Δb_j^l sont les mises à jour calculées.

2.12 Première approche : CNN 2D Baseline

Par le biais de l'apprentissage profond, on peut former les ordinateurs à reproduire le comportement humain. Cette méthode d'apprentissage automatique fait appel à un réseau de neurones à trois niveaux. Acquisition de connaissances par la classification de textes, d'audios ou d'images. Chaque algorithme dans la hiérarchie effectue une transformation non linéaire sur ses données d'entrée avant de produire un modèle statistique en sortie. Les réseaux de neurones convolutifs sont couramment employés en deep learning pour l'analyse des données visuelles. La majeure partie des calculs est effectuée par le processeur (GPU), qui représente le composant principal d'un réseau neuronal convolutif (CNN). Dans cette partie, on présente les différentes strates de l'architecture CNN démontrées à la figure 2.9 :

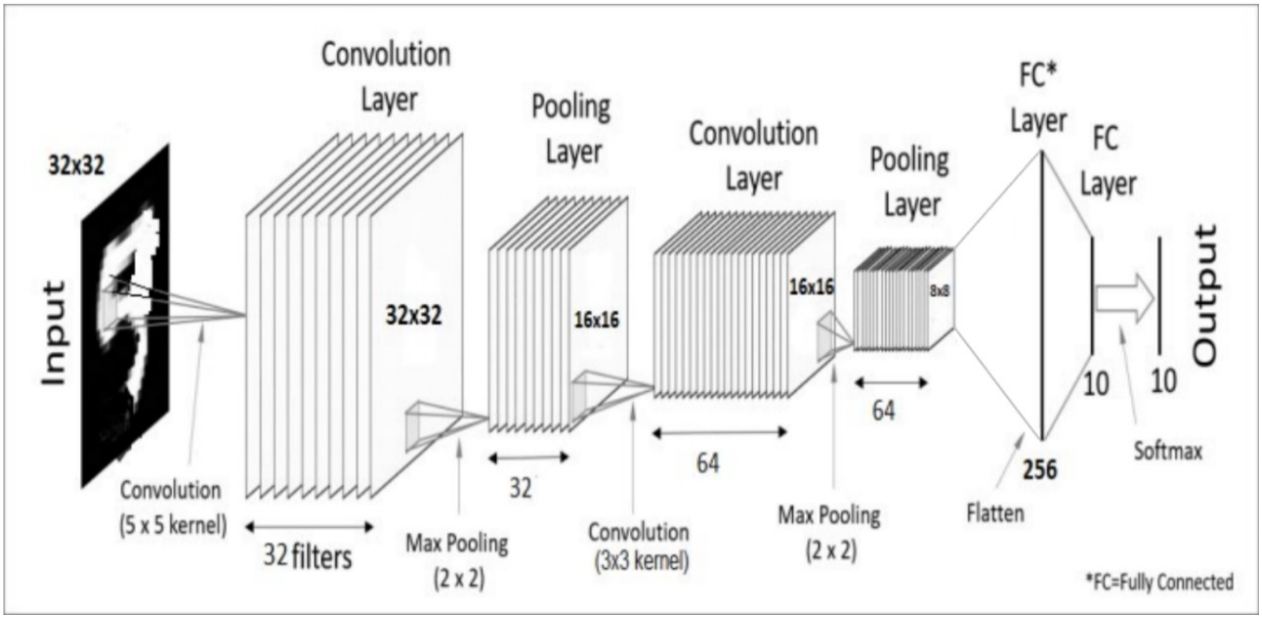


FIGURE 2.9 – Architecture de base des CNN [14].

2.13 Architecture du modèle CNN

Les ConvNets, aussi connus sous le nom de réseaux de neurones convolutifs (CNN), sont une catégorie de réseau neuronal élaborée spécifiquement pour gérer des données structurées en grille, tel que le traitement d'images. Une image numérique est une illustration de données codées en binaire. Elle se compose d'une suite continue d'éléments visuels distincts connus sous le nom de pixels. L'approche suggérée fait appel à un modèle de réseau neuronal convolutif (CNN) profond multicouche afin de différencier les images authentiques des images falsifiées. Un modèle CNN comprend trois types de couches : une couche pleinement connectée (FC), une couche de convolution et une couche

de pooling. Nous allons désormais décrire chaque niveau en détail.

2.13.1 Couche d'entrée

La couche d'entrée n'est pas vue comme un élément fonctionnel du CNN ; elle représente plutôt la couche par laquelle le réseau reçoit des données brutes en plusieurs matrices. Les CNN sont organisés en une configuration tridimensionnelle, disposant les neurones selon les dimensions de longueur, hauteur et profondeur. Ces dimensions sont souvent associées aux dimensions spatiales d'une image (largeur et hauteur en pixels) et à ses canaux de couleur (comme le RGB). Les CNN analysent l'image d'entrée en passant par diverses couches reliées entre elles, générant finalement un groupe de probabilités qui indiquent son appartenance à une certaine classe [14].

2.13.2 Couche de convolution

L'un des éléments les plus cruciaux d'un CNN est sa couche de convolution. Pendant la formation, il est primordial de saisir tous les paramètres de ces filtres, aussi connus sous le nom de noyaux. Il arrive souvent que les filtres soient plus petits que l'image qu'ils sont destinés à optimiser. Cette couche se sert de filtres à noyau pour extraire les informations pertinentes des images d'entrée qui ont été traitées par convolution. En utilisant des valeurs constantes de plus petite taille, les filtres à noyau deviennent de plus en plus proches des images en entrée. L'application de multiples filtres sur une image à travers la convolution facilite la détection des contours, le flou et l'accentuation. Nous avons élaboré ce modèle CNN en employant la couche Conv2D dans la structure de convolution. En tout, il y a six couches Conv2D ont été utilisées pour construire le modèle.

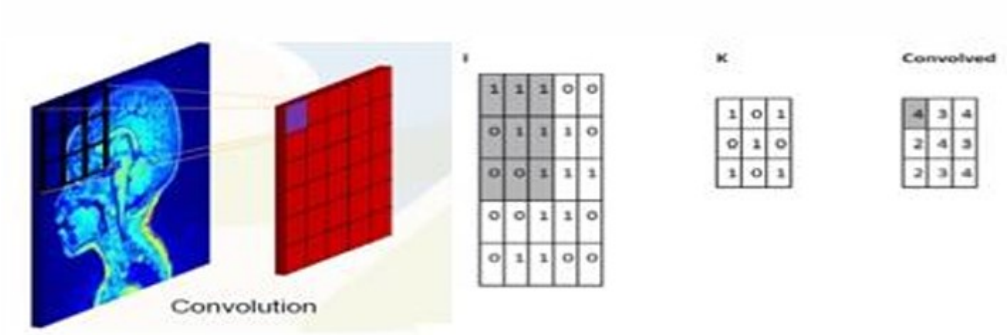


FIGURE 2.10 – Le processus de convolution [30].

Les étapes suivantes [14, 15], sont comprises dans le processus de convolution.

1. **Mise en œuvre du filtre :**
   Chaque filtre (ou noyau) est une matrice de petite taille définie à l'avance (par exemple, 3x3 ou 5x5, comme illustré dans la figure 2.11) et sa profondeur correspond au nombre de canaux de l'image d'entrée (par exemple, 3 pour RGB). Le filtre analyse l'image d'entrée en multipliant chaque élément et en cumulant les résultats pour produire une valeur unique destinée à la carte des caractéristiques [14, 15].

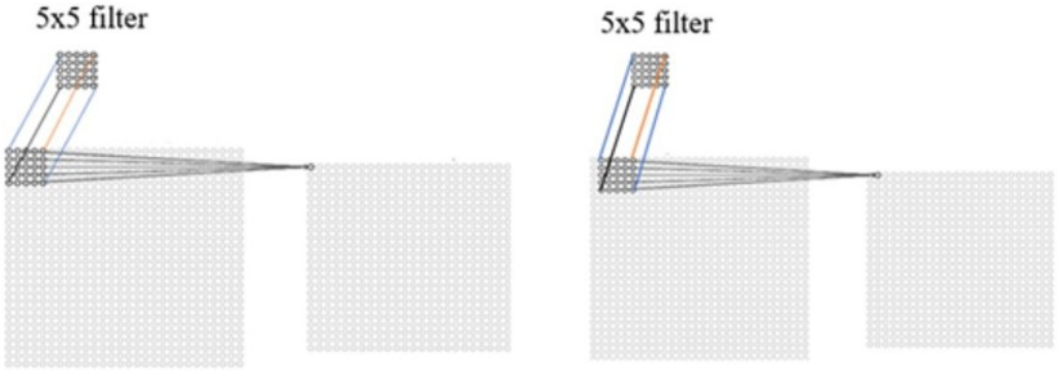


FIGURE 2.11 – Exécution d'une opération de convolution avec un noyau 5 x 5 et un stride de 1. Source : Priya Pedamkar, Réseau de neurones DNN, EDUCBA

2. **Pas(Stride)**
   Le pas est la dimension de l'écart, spécifiée en lignes et colonnes, selon laquelle un noyau de convolution se déplace à travers l'image d'entrée (matrice), soit de gauche à droite ou de haut en bas [14, 15]. Quand le pas est fixé à 1, le filtre progresse pixel par pixel (voir les figures 2.10 et 2.11). Accroître le rythme (par exemple, figure 2.12) conduit à obtenir une carte de caractéristiques de sortie plus compacte et à diminuer la charge computationnelle.

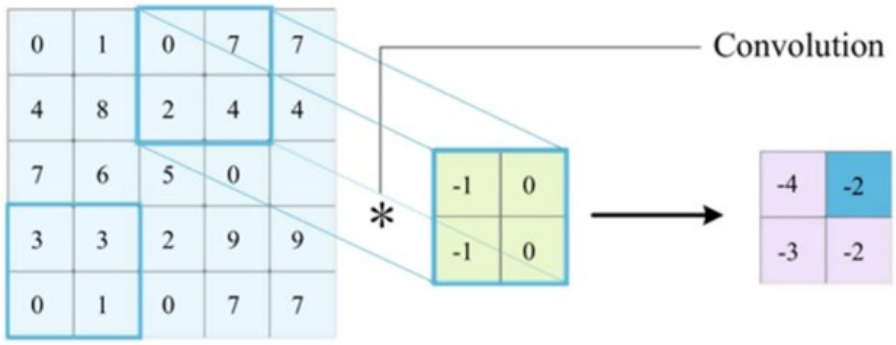


FIGURE 2.12 – Convolution effectuée avec un noyau 2x2 et un stride de 2, menant à une carte de caractéristiques diminuée[15].

3. **Remplissage avec des zéros :**
   Pour maintenir les dimensions de l'image d'origine, on peut appliquer un remplissage par des zéros sur ses bords [14, 15]. Cette méthode implique l'ajout de zéros autour des bords des données d'entrée, ce qui permet à la carte des caractéristiques de sortie de maintenir les mêmes dimensions spatiales que l'image d'entrée (voir figure 2.13).

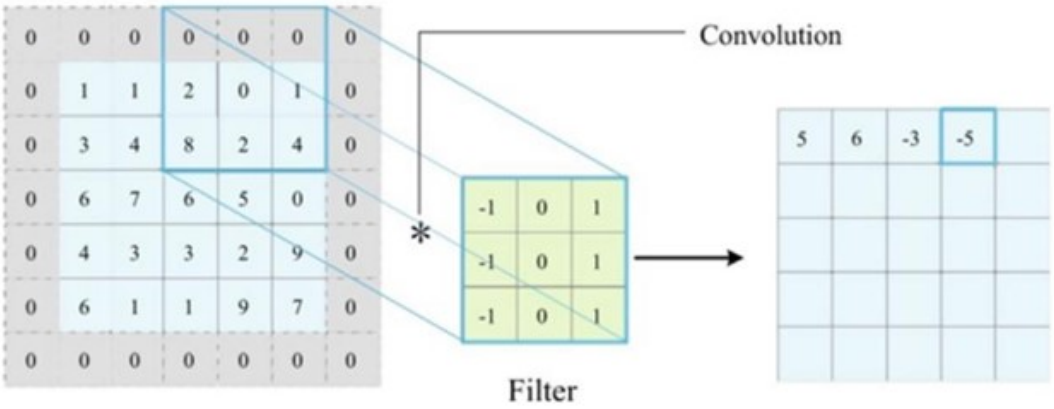


FIGURE 2.13 – Processus de convolution avec ajout de zéros et un décalage de 2, conservant ainsi les dimensions spatiales[15].

2.13.3 couche de mise en commun :

Le pooling est un sous-échantillon de l'image qui permet à la couche suivante d'avoir des régions spatiales plus larges, diminuant ainsi le nombre de variables à traiter. De plus, il rend le réseau moins sensible aux translations possibles de l'image sans changer la profondeur [83]. Par conséquent, il examine les réponses des filtres à divers emplacements, ce qui permet d'identifier les emplacements des filtres qui fournissent des réponses identiques. Les méthodes de pooling les plus fréquemment utilisées sont le max pooling et le average pooling.

a) Max Pooling

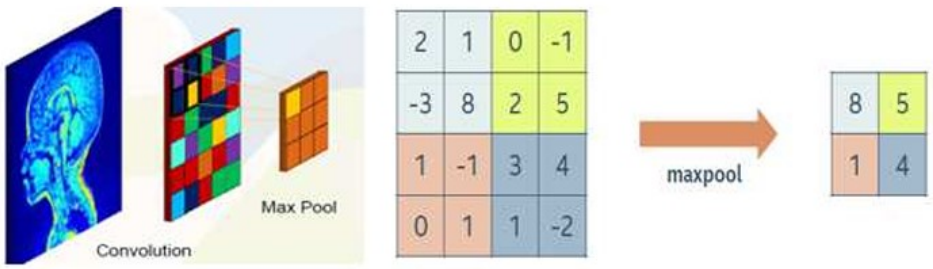


FIGURE 2.14 – Exemple max pooling.

b) Moyenne pooling

si on avait plutot choisi l'opération moyenne pooling pour le réseau , la sortie associée a l'opération pooling aurait été donnée par

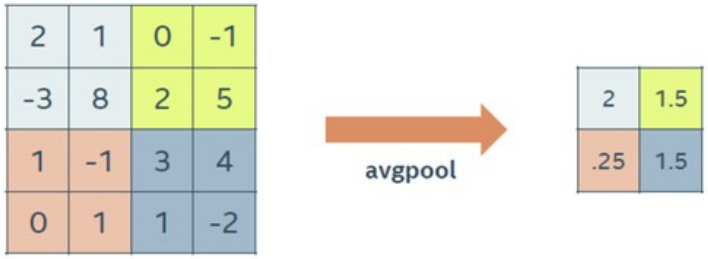


FIGURE 2.15 – Exemple de moyenne pooling .

Global Pooling

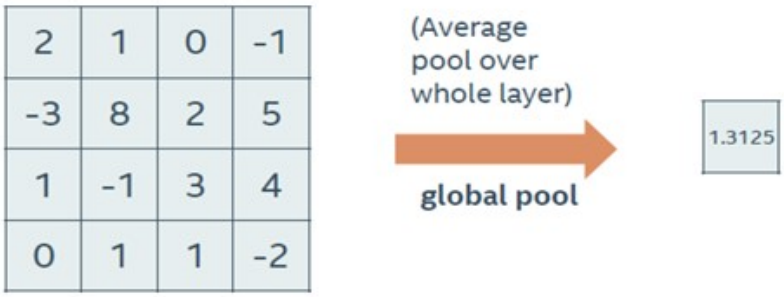


FIGURE 2.16 – Exemple de Global Pooling.

2.13.4 Couche entièrement connectée

Les neurones d'une les couches entièrement connectées sont reliées à chacun des neurones du volume correspondant à son entrée, à savoir chacun des neurones de la couche précédente. Si on vectorise notre volume d'entrants, on obtient exactement un réseau de neurones PMC [84].

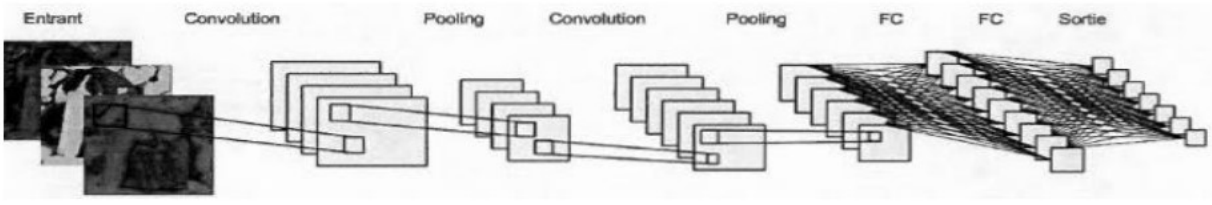


FIGURE 2.17 – Exemple de réseau convolutif deux étapes CONV-POOL suivies par deux couches Fc.

2.13.5 Cartes des caractéristiques «Feature Maps»

Chaque couche de Les cellules simples comprennent un certain nombre de filtres de convolution. Une opération de convolution du filtre applique le même motif à toute l'entrée. Pour chaque filtre, on Obtenez une carte de caractéristiques « Features Map » de sortie donnant, à chaque position de

l'entrée, la réponse de la somme des poids du filtre multipliés un à un aux pixels correspondants dans le voisinage de cette position.

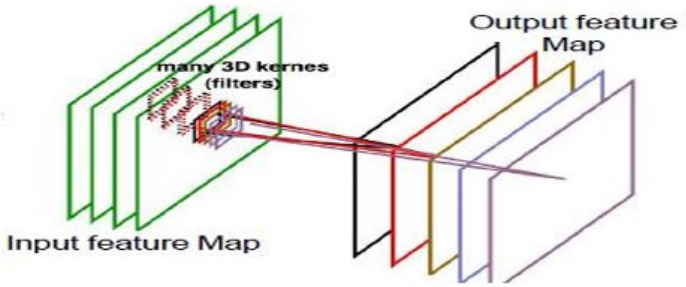


FIGURE 2.18 – Illustration d'une carte de caractéristiques « Feature Map ».

Le nombre de caractéristiques « Features Map » en sortie est généralement supérieur au nombre d'entrées.

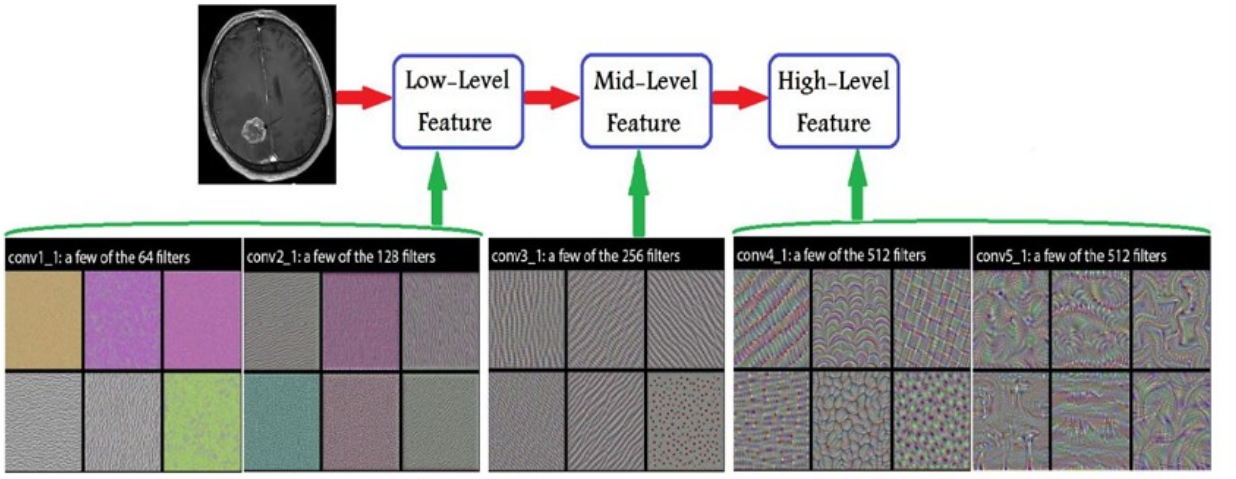


FIGURE 2.19 – Illustration de la carte des attributs dans les réseaux neuronaux..

2.14 Fonction d'activation

La convolution est une opération linéaire, mais étant donné que les images et leurs caractéristiques sont intrinsèquement non linéaires, il est crucial d'intégrer la non-linéarité dans les cartes d'activation produites par les couches convolutives[85]. On y parvient généralement en intégrant des éléments non linéaires. Des fonctions d'activation, telles que *sigmoid*, *tanh*, *softmax* et $ReLU$, sont appliquées immédiatement après les couches convolutionnelles.

Fonction sigmoïde : La fonction sigmoïde (ou fonction logistique) réduit les valeurs d'entrée à un intervalle $(0, 1)$, ce qui en fait un outil approprié pour normaliser les sorties de neurones et prédire les probabilités [15].

Fonction de tangente hyperbolique (Tanh) : Cette fonction d'activation modifie les valeurs d'entrée pour les adapter à l'intervalle $[-1, 1]$, créant ainsi une courbe en forme de S et agissant

comme une version zoomée et décalée de la fonction sigmoïde. L'équation (2.10) de la référence[15] exprime la relation mathématique entre les fonctions Tanh et Sigmoid :

$$\tanh(x) = 2 \cdot \text{Sigmoid}(2x) - 1 \tag{2.10}$$

Fonction softmax : On applique souvent la fonction Softmax dans les problèmes de classification multi-classes, normalisant ainsi un vecteur réel vers un vecteur de probabilités (distribution de probabilités), où les valeurs varient entre 0 et 1, et la somme totale s'élève à 1 [15]. Elle évalue la probabilité d'être dans chaque classe. Par exemple, dans les tâches de classification, elle attribue des probabilités plus élevées aux catégories ayant des valeurs plus grandes, indiquant une probabilité plus élevée d'appartenir à cette catégorie. L'équation (2.11) contient la formulation de la fonction SoftMax :

$$\text{Softmax}(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}} \tag{2.11}$$

Ici, $S(x_i)$ représente la probabilité de la classe i, où x_i est le score d'entrée associé à cette classe, tandis que K désigne le nombre total de classes possibles [86]. Le dénominateur effectue la somme des exponentielles des scores de toutes les classes afin de normaliser les probabilités produites.

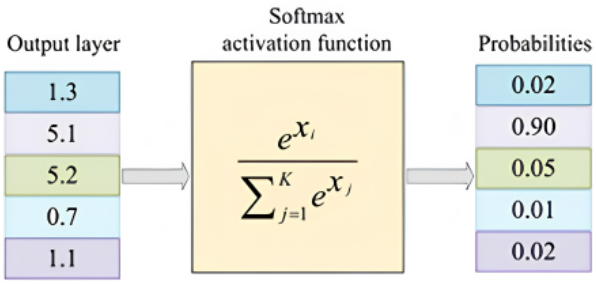


FIGURE 2.20 – diagrame softmax[82]

Fonction de l'Unité Linéaire Rectifiée (ReLU) :

La fonction d'activation ReLU est largement utilisée en apprentissage profond en raison de sa dérivée constante dans la région positive, ce qui permet de limiter les problèmes d'extinction et d'explosion du gradient. Elle laisse passer uniquement les valeurs positives, ce qui facilite la mise à jour des poids et améliore l'efficacité de l'apprentissage [15].

Fonction Leaky ReLU :

Leaky ReLU est une variante de la fonction ReLU qui vise à résoudre le problème des neurones « morts » (Dead ReLU) en autorisant un faible gradient pour les valeurs d'entrée négatives. Bien que cette modification améliore le comportement du réseau dans certains cas, elle n'est pas systématiquement supérieure à la ReLU classique en pratique [15].

Comparaison des fonctions d'activation :

Les fonctions Sigmoid et Tanh souffrent du problème de disparition du gradient, ce qui limite leur efficacité dans les réseaux de neurones profonds. En revanche, les fonctions ReLU et ses variantes sont généralement privilégiées car elles permettent de mieux gérer les problèmes liés aux gradients dans l'apprentissage profond [15].

2.15 Fonctions de perte

Les fonctions de perte mesurent la proximité entre les résultats d'un réseau neuronal spécifique et la réalité.

2.15.1 L'erreur quadratique moyenne (MSE)

À l'instar des « moindres carrés ordinaires » dans le cadre de la régression linéaire, on procède à l'addition de toutes les pertes au carré pour chaque exemple avant de diviser cette somme par le nombre total d'exemples [87].

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - x_i)^2$$

2.15.2 L'algorithme de rétro-propagation

C'est un procédé permettant de déterminer le gradient de l'erreur pour chaque neurone d'un réseau neuronal, en partant de la couche finale jusqu'à la première [88].

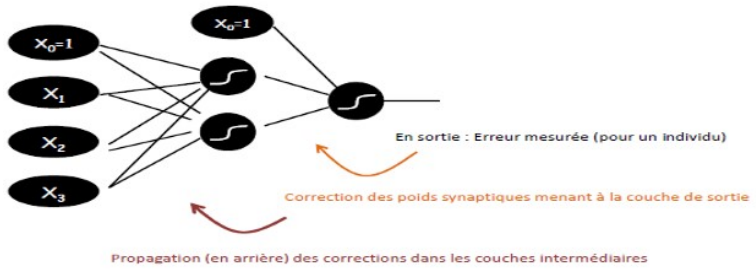


FIGURE 2.21 – algorithme de rétropropagation .

2.16 Justification du Choix des CNN pour l'IRM Cérébrale

Les IRM cérébrales présentent des caractéristiques qui rendent les CNN particulièrement adaptés :

1. **Invariance spatiale**
   une tumeur a le même aspect visuel quelle que soit sa position dans le volue
2. **Hiérarchie des features**
   des arêtes → textures → structures anatomiques → régions tumorales
3. **Partage des poids**
   un même filtre détecte la même structure partout dans le volume, réduisant le nombre de paramètres
4. **Localité**
   les features médicales pertinentes sont locales (contours de tumeur, transitions de tisss)

2.17 Deuxième approche : Architecture U-Net

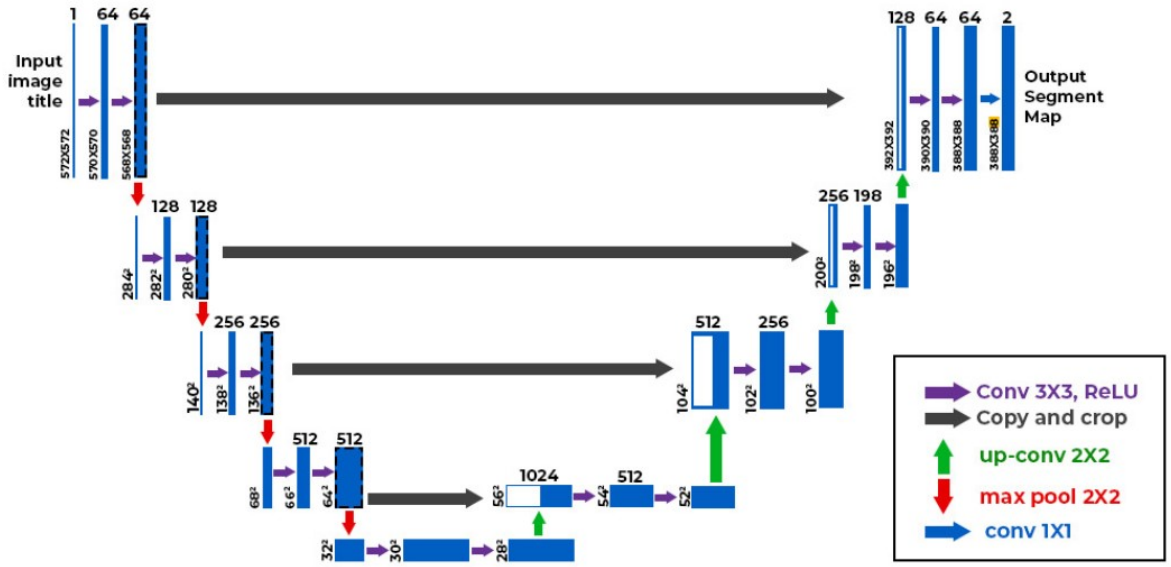


FIGURE 2.22 – Structure de l'U-Net.

Flèches et symboles :

— Flèches bleues : indiquent le flux des données à travers les couches convolutives.

— Flèches rouges : indiquent les opérations de Max Pooling.

— Flèches vertes : indiquent les opérations d'up-convolution.

— Flèches grises : indiquent la copie et le rognage des cartes de caractéristiques pour la fusion avec les couches de l'expansion.

L'architecture U-Net 2D, développé en 2015 par Olaf Ronneberger, Philipp Fischer et Thomas Brox du département d'informatique de l'Université d'Albert-Ludwigs de Freiburg (Allemagne), représente une avancée importante dans le domaine de la segmentation d'images biomédicales [56]. Alors que les réseaux CNN classiques se limitent à la classification, U-Net 2D permet de résoudre le problème de la segmentation sémantique pixel-par-pixel grâce à la combinaison d'une voie contractante (encodeur) qui permet de capturer le contexte sémantique global, et d'une voie expansive (décodeur) qui restaure la localisation précise grâce à des connexionsde sauter (sauter les connexions)[56] . L'architecture U-Net 2D se caractérise par sa forme caractéristique en "U", résultant de la symétrie entre la partie contractante (gauche) et expansive (droite), reliées par un bottleneck central[56]. Cette structure traite des images d'entrée de dimensions typiques 572×572 (avec padding) pour générer une segmentation de sortie voxel-par-voxel de résolution identique, adaptée aux contraintes des datasets biomédicaux limités comme BraTS (369 patients)[10]. La voie contractante comprend 4 niveaux de résolution croissante ($256^2 \rightarrow 16^2$), chacun constitué de deux convolutions 3×3 suivies d'une ReLU et d'un MaxPooling 2×2, doublant le nombre de filtres à chaque niveau : $64 \rightarrow 128 \rightarrow 256 \rightarrow 512 \rightarrow 1024$ (bottleneck) [56]. La voie expansive miroir utilise des convolutions transposées 2×2 (up-convolutions) pour restaurer la résolution, combinées aux caractéristiques haute-résolution de l'encodeur via les skip connections.

2.17.1 Structure de U-Net 2D

2.17.1.1 Encodeur

L'UNet reçoit en entrée une image ayant une résolution spatiale spécifique et un nombre déterminé de canaux (par exemple, une image monochrome disposerait d'un unique canal, alors qu'une image colorée posséderait trois canaux distincts pour le rouge, le vert et le bleu). La voie de contraction est constituée de blocs consécutifs, chaque bloc renfermant deux couches de convolution, suivies d'une couche d'activation ReLU et d'une opération de max pooling. Chaque niveau de convolutions emploie des filtres de dimension réduite (habituellement 3x3) afin de saisir les motifs locaux au sein de l'image. À chaque phase de réduction, l'image perd en résolution (grâce au pooling maximal), tandis que les cartes de caractéristiques gagnent en profondeur (en raison de l'ajout de filtres). Cela autorise la saisie d'informations de plus en plus abstraites et généralisées de l'image.

1. **Panneaux de Double Convolution (3x3)**

   Chaque étage de l'encodeur est formé de deux couches de convolution consécutives.

   Champ récepteur : L'emploi de deux convolutions consécutives de 3X3 donne un champ récepteur qui équivaut à une convolution de 5X5,tout en réduisant le nombre de paramètres et en ajoutant davantage de non-linéarités par le biais de fonctions d'activation répétées.

   Activation ReLU : Une unité de rectification linéaire (ReLU) est appliquée après chaque convolution. Elle donne au réseau la possibilité d'assimiler des motifs complexes en intégrant une non-linéarité dans le traitement des données.

   Extraction de caractéristiques : À mesure que l'image passe par ces différentes couches, le réseau repère des structures de plus en plus sophistiquées, passant des simples contours (intensités des tissus) aux formes tumorales complexes.

2. **Diminution spatiale par MaxPool2d (2x2)**

   une opération de Max Pooling est effectuée après chaque paire de convolutions doubles.

   Compression de l'image : Cette procédure réduit de moitié la hauteur et la largeur de la carte des caractéristiques. Par exemple, une image IRM de 240X240 est diminuée à 120X120 suite au premier niveau.

   Invariance spatiale : Le Max Pooling ne retient que la valeur maximale dans une fenêtre spécifique, ce qui rend le modèle plus résistant aux petites altérations ou modifications mineures de la localisation de la tumeur dans le cerveau.

2.17.1.2 Le goulot d'étranglement (bottleneck)

en particulier la partie basse du « U », employe une double convolution (habituellement deux couches Conv2D 3x3) pour transformer les features maps en 1024 canaux, permettant ainsi d'optimiser l'extraction des caractéristiques sémantiques abstraites et globales. Cette phase de compression intense, qui se déroule au début du décodeur, encode le contexte haut niveau crucial pour une segmentation exacte en imagerie médicale, notamment pour les lésions cérébrales détectées par l'IRM, avant la reconstruction montante grâce aux connexions skip. Il s'agit d'une version modifiée

des U-Net originaux [56] destinée aux jeux de données complexes, où l'utilisation de 1024 canaux représente un équilibre satisfaisant entre la capacité expressive et le coût en termes de calcul.

2.17.1.3 Décodeur

L'architecture UNet, grâce à son décodeur, reconstruit graduellement la résolution spatiale initiale de l'image à partir des caractéristiques compressées du goulot d'étranglement, en recourant principalement à des convolutions inversées (ConvTranspose2D) pour le suréchantillonnage [56]. Chaque section du décodeur respecte ce modèle habituel :

1. **ConvTranspose2D 2x2**
   Multiplie la résolution spatiale par deux (H×W → 2H×2W) tout en diminuant le nombre de canaux (par exemple : 1024 → 512), jouant le rôle d'une « déconvolution » inverse des MaxPool de l'encodeur [89].
2. **Concaténation (connexion par contiguïté)**
   Combine la carte de caractéristiques agrandie à l'échelle avec celle correspondante de l'encodeur (identique en résolution), préservant ainsi les détails précis qui ont été perdus lors de la contraction [90].
3. **Double convolution 3x3 augmentée de ReLU**
   Affine les caractéristiques fusionnées tout en préservant le même nombre de canaux pour une extraction sémantique locale [56].

2.17.1.4 Les connexions de saut (skip connections)

L'élément novateur majeur de l'architecture U-Net réside dans la liaison symétrique de chaque niveau de l'encodeur avec le décodeur correspondant par le biais de la concaténation des feature maps suite au upsampling via ConvTranspose2D. Cela permet une fusion des caractéristiques à haute résolution spatiale (les détails fins perdus lors des MaxPoolings de contraction) avec les informations sémantiques abstraites provenant du bottleneck, ce qui restaure la précision de localisation cruciale pour la segmentation d'images médicales comme les lésions cérébrales en IRM tout en facilitant la propagation des gradients pendant l'entraînement.

2.17.1.5 Couche de sortie

La dernière couche est une convolution 1 X 1 qui réduit le nombre de canaux au nombre de classes souhaitées. Elle attribue chaque pixel à une classe spécifique pour générer le masque final.

2.17.2 Fonction de perte hybride et processus d'entraînement

Afin d'améliorer le modèle, une fonction de perte combinée a été employée : la combinaison du Dice et de la Cross-Entropy Loss.

a) Fonction de perte (Loss Function)

Dans le cadre d'un algorithme d'optimisation, la fonction qui sert à juger une solution proposée (à savoir un groupe de poids) est dénommée fonction objectif. Nous visons à optimiser la fonction objective, soit en quête d'une solution potentielle qui présente le score le plus élevé

ou le plus faible, selon que nous souhaitons maximiser ou minimiser. Nous tentons de réduire l'erreur grâce à l'utilisation des réseaux de neurones. On désigne souvent la fonction objective comme fonction de coût ou fonction de perte, et la valeur déterminée par cette dernière est simplement nommée «perte» [91].

b) L'entropie croisée (Cross-Entropy Loss)

La mesure de l'entropie relative entre deux distributions de probabilité sur un même ensemble d'événements est appelée entropie croisée [92]. On peut considérer que l'entropie croisée évalue l'entropie totale entre les distributions. On appelle souvent la perte d'entropie croisée « entropie croisée » ou « perte logarithmique ». Chaque prédiction de probabilité est mise en parallèle avec la valeur de classe réelle (0 ou 1) et un score est établi qui sanctionne la probabilité selon l'écart par rapport à la valeur escomptée. La sanction est logarithmique, accordant une faible note pour des différences mineures (0,1 ou 0,2) et une note considérable pour une grande différence (0,9 ou 1,0). On cherche à minimiser la perte d'entropie croisée, des valeurs inférieures indiquant un modèle de meilleure qualité par rapport à des valeurs supérieures. Un modèle qui anticipe des probabilités exactes aura une perte de log ou une entropie croisée de 0,0. Pour un problème de classification binaire ou à deux classes, l'entropie croisée est en réalité calculée comme la moyenne de l'entropie croisée sur tous les exemples.

c) Perte de Dice (Dice Loss)

Spécialement élaborée pour la segmentation des images médicales, elle vise à optimiser la superposition entre la prédiction et le véritable état des choses, un aspect crucial pour traiter l'inégalité entre les tumeurs et les tissus sains.

À l'inverse de la Cross-Entropy, la Dice Loss prend en compte l'objet dans son ensemble. Elle met l'accent sur la superposition.

La Dice Loss est définie par :

$$L_{Dice} = 1 - \frac{2|X \cap Y|}{|X| + |Y|} \tag{2.12}$$

Puisqu'elle effectue une normalisation en fonction de la taille de la zone prédite et réelle, elle évite d'être « submergée » par le grand volume de pixels sains. Cela contraint le modèle à être exact en ce qui concerne les limites et la forme de la tumeur.

d) le classifieur «Softmax» restreignent la sortie de la fonction à un intervalle de 0 à 1. Cela rend possible l'interprétation directe de la sortie en tant que probabilité. La fonction Softmax produit un vecteur illustrant les distributions de probabilité d'une série de résultats envisageables [93]. Une couche Softmax est habituellement la dernière couche utilisée dans les fonctions du réseau de neurones. Il est essentiel de préciser que cette couche doit contenir un nombre de nœuds identique à celui de la sortie suivante. Softmax convertit les logits (la sortie numérique de la couche linéaire finale d'un réseau de neurones multi-classes) en probabilités. Il le fait en exponentiant chaque sortie, puis en divisant chaque valeur par la somme totale des exposants, garantissant ainsi que l'ensemble du vecteur de sortie soit égal à un. Softmax est souvent intégré en tant que dernière couche d'un réseau de classification d'images tel que VGG16.

e) Epoques(Epochs) : Une époque d'apprentissage signifie que l'algorithme d'apprentissage

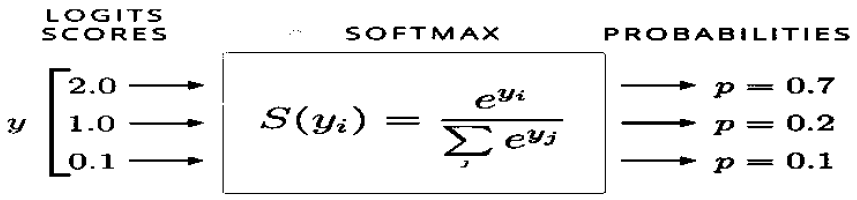


FIGURE 2.23 – illustration d'un classifieur softmax.

a effectué un passage dans l'ensemble de données d'apprentissage, où les exemples ont été séparés en groupes de «taille de lot» (batch size) sélectionnés aléatoirement [94].

2.18 Optimisation des fonctions

L'optimisation des fonctions est une composante essentielle du Deep Learning. La majeure partie des algorithmes du Deep Learning nécessite l'ajustement des paramètres (tels que les poids, coefficients, etc.) en fonction des données d'entraînement. L'optimisation se réfère aussi au processus de recherche du meilleur ensemble d'hyperparamètres pour paramétrer l'entraînement d'un algorithme de Deep Learning.

2.18.1 Taux d'apprentissage (Learning Rate)

Lors de la configuration des réseaux neuronaux, le taux d'apprentissage pourrait être l'hyperparamètre le plus crucial. Il gère le niveau de modification du modèle en fonction de l'erreur estimée chaque fois que les poids du modèle sont actualisés [95]. Il est ardu de choisir le taux d'apprentissage, car une valeur trop basse pourrait prolonger le processus de formation et le rendre susceptible de se bloquer, alors qu'une valeur trop élevée risque d'entraîner l'acquisition rapide d'un ensemble de poids sous-optimal ou un processus d'entraînement instable.

2.18.2 Optimisation à l'aide d'Adam

Le nom Adam fait référence à l'estimation adaptative du moment (adaptive moment estimation). C'est un algorithme conçu pour l'optimisation du gradient de première ordre des fonctions cibles, s'appuyant sur des estimations adaptatives des moments de premier ordre [96].

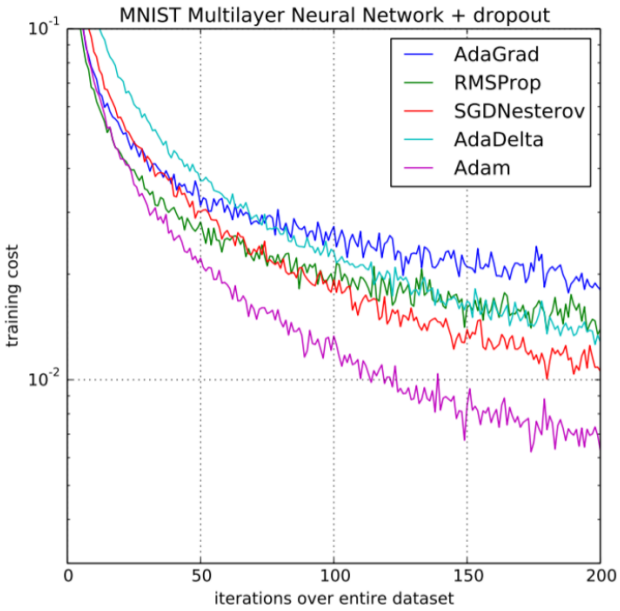


FIGURE 2.24 – Exemple de comparaison d'Adam par rapport à d'autres algorithmes d'optimisation.

Voici les divers paramètres de configuration pour Adam :

— **Alpha** (α) : taux d'apprentissage ou taille de pas. Il représente le ratio des pondérations actualisées. Des valeurs supérieures conduisent à un apprentissage précoce plus rapide avant la mise à jour du taux. Des valeurs inférieures prolongent le processus d'apprentissage pendant toute la durée de l'entraînement.

— **Beta1** (β_1) : vitesse de décroissance exponentielle pour les estimations du premier moment.

— **Beta2** (β_2) : taux de décroissance exponentielle pour les estimations du deuxième moment. Pour les problèmes présentant un gradient clairsemé, cette valeur devrait être proche de 1,0.

— **Epsilon** (ε) : nombre très petit utilisé pour éviter toute division par zéro lors de l'implémentation.

Pour les problèmes de Deep Learning, les valeurs standards utilisées sont généralement :

$$\alpha = 0.001, \quad \beta_1 = 0.9, \quad \beta_2 = 0.999, \quad \varepsilon = 10^{-8}$$

Cependant, leur utilisation n'est pas obligatoire.

2.19 Architecture Proposée

2.19.1 U-Net 3D — Transition vers le Volumétrique

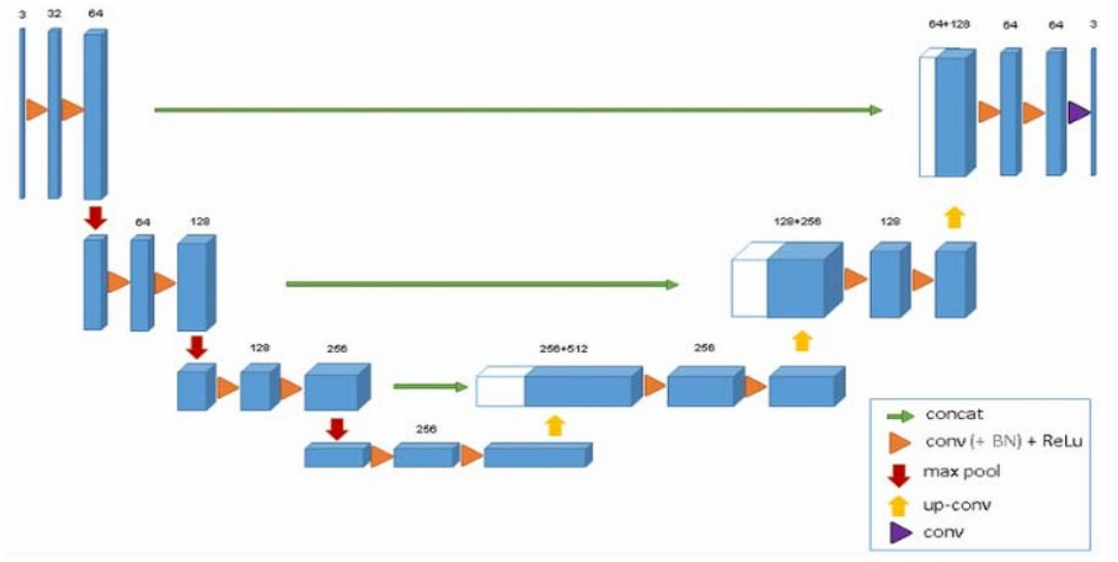


FIGURE 2.25 – Architecture 3D-U-Net [16]

2.19.1.1 Raisons du passage à la 3D

[97] ont développé le UNet en trois dimensions en substituant toutes les opérations en deux dimensions par les versions volumétriques correspondantes (Conv3d, MaxPool3d, ConvTranspose3d).

Cette extension s'applique naturellement aux images IRM qui sont intrinsèquement en trois dimensions : une tumeur se développe dans les trois dimensions et sa forme tridimensionnelle présente une signification clinique.

Le U-Net 3D a mis en évidence des progrès significatifs sur BraTS par rapport aux méthodes 2D, surtout en ce qui concerne la zone TC.

Dans notre mise en œuvre, le modèle est alimenté avec des patches volumétrques mesurant $(B, 4, 128, 128, 128)$, où B représente la taille du lot, 4 indique le nombre de modalités et 128^3 correspond à la résolution spatiale du patch.

Cet arrangement saisit environ 82% de la profondeur d'un cerveau humain, soit 128 sur 155 dans une unique inférence.

2.19.1.2 Conv3D à Deux Couches Résiduelles - Bloc Résiduel 3D

Sur la base du modèle ResNet [53], notre élément fondamental intègre la convolution 3D double avec un lien résiduel :

$$\text{ResDoubleConv3D}(x) = \text{LeakyReLU}\Big(\text{Conv}_2(\text{IN}(\text{LReLU}(\text{Conv}_1(\text{IN}(x))))) + \text{Raccourci}(x)\Big) \quad (2.13)$$

Le raccourci est défini comme une opération Conv3d$(1 \times 1 \times 1)$ avec InstanceNorm si le nombre de canaux d'entrée est différent du nombre de canaux de sortie, sinon une transformation d'identité est appliquée.

Ceci est une structure qui favorise la diffusion du gradient à travers les 4 niveaux provenant de l'encodeur ainsi que ceux du décodeur.

— Application du Dropout3d(p) après la première convoltion - régularisation en trois dimensions.

— Niveaux : 4→32, 32→64, 64→128, 128→256, puis bottleneck 256→512.

2.19.1.3 AttentionGate3D — Filtrage des Skip-Connections

[63] ont introduit les Attention Gates dans le contexte de la segmentation des organes abdominaux.

Le gate est formé pour ajuster la pondération des skip-connections en fonction de leur pertinence pour la segmentation :

$$\alpha = \sigma\left(\psi^T \cdot \mathrm{ReLU}\left(W_g \cdot g + W_x \cdot x + b_g\right)\right) \tag{2.14}$$

$$\hat{x} = \alpha \odot x \tag{2.15}$$

Dans cette formule, g représente le signal de déclenchement du décodeur (contexte sémantique), x fait référence à la carte de caractéristiques obtenue par la connexion de contournement (détails spatiaux), et α symbolise la carte d'attention apprise.

Le coefficient $\alpha \in [0, 1]$ supprime les activations non pertinentes (fond cérébral) et amplifie les régions tumorales.

Selon[63], cette méthode permet d'améliorer le Dice ET de 3 à 5 points.

2.19.1.4 ASPP3D - Multi-Échelle avec Goulot d'Étranglement

Inspiré par DeepLab v3+[64] en trois dimensions, notre ASPP fait appel à 4 branches de convolutions dilatées en parallèle (taux 1, 2, 4, 6), en plus d'une branche de pooling global :

— Taux=1 : champ de réception 3^3 = local → détecte les petites tailles (environ 5-10 ml).

— Taux=2 : zone réceptrice 5^3 → transitions des tissus.

— Taux=4 : champ récepteur 9^3 → TC (~50-100 ml).

— Taux=6 : zone de réception 13^3 → WT (grande superficie, environ 100-300ml).

— Fil global : AdaptiveAvgPool3d(1) → applicable à l'ensemble du volume.

Les 5 branches sont reliées, puis traitées par une Conv3d($1 \times 1 \times 1$) avec une projection de 512 vers 512 canaux.

Cette structure Mult-iniveau est particulièrement appropriée pour BraTS où les trois zones visées présentent des dimensions très variées.

2.19.1.5 Ensemble de données 3D et prétraitement

L'équipe BraTS et les meilleures pratiques recommandées dans la littérature[65, 97] guident le prétraitement des données :

— Normalisation par modalité (z-score sur voxels non-nuls) : chaque modalité possède une distribution d'intensité distincte — la normalisation globale fausserait les 4 modalités.

— Coupe axée sur la tumeur (50% du temps) : assure la présence de voxels positifs, essentiels pour la radiothérapie externe.

— Augmentation 3D : basculement aléatoire (3 axes), rotation ±15°, zoom 0.9-1.1×, bruit gaussien $\sigma \in [0, 0.1]$.

— Fenêtre Glissante 3D (inférence) : chevauchement à 50%, logits moyennés sur les patches qui se superposent.

2.19.1.6 Loss Hybride : Dice + CE + Focal

Notre fonction de perte intègre trois éléments complémentaires :

$$L = 0.4 \times L_{CE} + 0.3 \times L_{Dice} + 0.3 \times L_{Focal} \tag{2.16}$$

La Focal Loss[98] diminue automatiquement l'importance des exemples simples (fond) afin de focaliser l'apprentissage sur les classes ardues (ET) :

$$L_{Focal} = -(1 - p_t)^{\gamma} \cdot \log(p_t) \qquad \text{avec } \gamma = 2 \tag{2.17}$$

Les coefficients de classe $[0.5, 2.0, 2.0, 3.0]$ pour $[bg, NCR, ED, ET]$ augmentent l'amplitude du signal pour l'ET d'un ratio de 6 comparé au bruit de fond.

Cette pondération est conforme aux directives [65] pour les ensembles de données fortement déséquilibrés, tels que BraTS.

2.19.2 TransUNet 3D — Intégration du Transformer

2.19.2.1 L'usage des Transformers en Vision par Ordinateur

[99] ont introduit le mécanisme d'auto-attention dans le contexte du traitement du langage naturel.

Le Vision Transformer (ViT), qui est une adaptation de la vision par ordinateur, a été développé par [57] en segmentant l'image en patches (tokens) et en plaçant l'attention sur cette séquence.

Le Transformer se distingue par sa faculté à représenter des dépendances entre tokens de longueur arbitraire en utilisant $O(N^2)$ opérations, contrairement aux CNN disposant d'un champ récepteur limité.

Selon [66] ils ont introduit TransUNet, une première architecture hybride combinant CNN et Transformer pour la segmentation médicale en 2D.

Notre apport élargit cette idée au 3D et présente un procédé de transition innovant plutôt qu'un encodeur exclusivement Transformer.

2.19.2.2 Auto-attention Multi-têtes (MHSA)

Le système d'attention [99] génère pour chaque jeton une demande Q, une clé K et une valeur V grâce à une projection linéaire apprise :

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_{head}}}\right) \cdot V \tag{2.18}$$

où d_{head} correspond à TF_DIM divisé par TF_HEADS, soit 512/8, ce qui équivaut à 64.

L'usage du facteur $1/\sqrt{d_{head}}$ sert à prévenir la saturation du softmax.

Notre mise en œuvre se sert de 8 têtes parallèles, chacune capturant une sorte distincte de relation spatiale globale :

$$\text{MHSA}(Q, K, V) = \text{Concat}(head_1, \ldots, head_8) \cdot W^o \tag{2.19}$$

2.19.2.3 Bloc de Transformateur

Chaque couche Transformer [99] fusionne le MHSA et le réseau Feed-Forward (FFN) en utilisant des connexions résiduelles et la LayerNorm (pré-normalisation, préconisée par [100]) :

$$x = x + \text{MHSA}(\text{LayerNorm}(x)) \tag{2.20}$$

$$x = x + \text{FFN}(\text{LayerNorm}(x)) \tag{2.21}$$

Le FFN est composé de :
Linear(512→2048) → GELU → Dropout(0.1) → Linear(2048→512).
Dans le contexte des Transformers, la fonction GELU [101] présente des caractéristiques de gradient supérieures à celles de ReLU. Nous empilons six couches TransformerBlock, une configuration qui s'aligne sur les architectures BERT-base [102] et qui est suggérée pour des séquences de longueur N=64 tokens.

2.19.2.4 PatchEmbed3D — La Tokenisation du Volume

La tâche de tokenisation implique le fractionnement de la feature map 3D $(B, 512, 8, 8, 8)$ en cubes distincts qui n'ont pas de superposition, chaque cube correspondant à un token.

En utilisant un `token_size = (2,2,2)` :

$$N_{\text{tokens}} = \left(\frac{8}{2}\right)^3 = 4^3 = 64$$

On obtient donc **64 tokens**, chacun de dimension **512**.

2.19.2.5 Encodage Positionnel

Le Transformer est insensible à la permutation des tokens — sans encodage de position, il ne tient pas compte de l'ordre spatial. Nous employons un codage sinusoïdal[99] calculé de manière dynamique :

$$PE(pos, 2i) = \sin\left(\frac{pos}{10000^{2i/d_{\text{model}}}}\right)$$

$$PE(pos, 2i+1) = \cos\left(\frac{pos}{10000^{2i/d_{\text{model}}}}\right)$$

Cet encodage est indépendant de la longueur de la séquence (ce qui est pratique lors de l'inférence sur des volumes de différentes dimensions) et ne rajoute aucun paramètre appris.

2.19.2.6 TransformerBridge — Notre Apport Architectural

À l'opposé du UNETR[59] qui substitue totalement le CNN par un Transformer, notre modèle TransformerBridge intègre le Transformer uniquement au point culminant, entre l'encodeur et le décodeur CNN. Ce positionnement stratégique tire parti des forces complémentaires des deux paradigmes :

— L'encodeur CNN extrait de manière efficace les caractéristiques locales (bords, textures).

— Le Transformer Bridge simule les relations à longue distance sur la carte des caractéristiques condensée.

— Le décodeur CNN redonne l'image segmentée avec une précision spatiale.

L'addition résiduelle globale

$$\text{out} = \text{Transformer}(x) + x$$

assure que si le Transformer ne fournit pas d'information pertinente, le signal CNN est maintenu (adapté à l'initialisation).

L'interpolation trilinéaire corrige les éventuels écarts de résolution causés par le pas du PatchEmbed.

Nous avons choisi d'utiliser GroupNorm[103] plutôt qu'InstanceNorm dans le carré de reconstruction, car GroupNorm reste stable quelle que soit la dimension spatiale du tenseur. Cela permet de résoudre le problème lié à DataParallel où InstanceNorm échoue lorsque la dimension spatiale est égale à 1.

2.19.2.7 Structure Complète de TransUNet3D

Étape	Module	Taille sortie (B=2)
Entrée	4 modalités IRM — patch 128^3	(2, 4, 128, 128, 128)
enc1	ResDoubleConv3D(4→32)	(2, 32, 128, 128, 128)
enc2	MaxPool + ResDoubleConv3D(32→64)	(2, 64, 64, 64, 64)
enc3	MaxPool + ResDoubleConv3D(64→128)	(2, 128, 32, 32, 32)
enc4	MaxPool + ResDoubleConv3D(128→256)	(2, 256, 16, 16, 16)
ASPP	MaxPool + ASPP3D(256→512)	(2, 512, 8, 8, 8)
Bridge	TransformerBridge (6L, 8H, dim=512)	(2, 512, 8, 8, 8)
dec4	Up + AttGate(e4) + ResConv(512→256)	(2, 256, 16, 16, 16)
dec3	Up + AttGate(e3) + ResConv(256→128)	(2, 128, 32, 32, 32)
dec2	Up + AttGate(e2) + ResConv(128→64)	(2, 64, 64, 64, 64)
dec1	Up + AttGate(e1) + ResConv(64→32)	(2, 32, 128, 128, 128)
Sortie	Conv3d(32→4, kernel=1×1×1)	(2, 4, 128, 128, 128)

2.19.3 Architecture Hybride Finale — TransResUNet3D

Notre architecture finalisée, dénommée TransResUNet3D améliorée, regroupe tous les éléments élaborés dans les sections antérieures, tout en incorporant une série de procédés additionnels inspirés des récentes percées en segmentation médicale 3D.

Cette version rectifie les erreurs de la première version (FIX-1 à FIX 5) et met en avant de nouveaux éléments : Squeeze-and-Excitation (SE), Attention Spatiale 3D, DropPath aléatoire, EMA (Exponential Moving Average), Augmentation de Test-Time (TTA), Boundary Loss et Lovász Loss.

2.19.3.1 Blocs CNN Améliorés : SEBlock3D, SpatialAttention3D et DropPath

En comparaison avec la version antérieure, chaque ResDoubleConv3D résiduel a été amélioré avec l'ajout de trois mécanismes complémentaires intégrés directement dans son passage avant :

— SEBlock3D (Squeeze-and-Excitation) : Un processus d'attention par canal s'inspirant de[104]. Un AdaptiveAvgPool3d(1) pour un bloc de C canaux réduit le volume à un vecteur de dimension C, suivi de deux couches linéaires (de C à C/16 à C) utilisant ReLU et Sigmoid pour générer des poids de pondération pour chaque canal. Chaque feature est multipliée par son poids associé :

$$x_{out} = x * sigmoid(W2 \cdot ReLU(W1 \cdot AvgPool(x)))$$

Cela donne l'opportunité au réseau de réajuster l'importance respective des canaux (par exemple : canal FLAIR contre T1ce pour le repérage de l'ET).

— SpatialAttention3D : Un mécanisme d'attention spatiale tiré du module CBAM [105], qui effectue une moyenne et une opération de maximisation sur la dimension des canaux (dim=1), joint ces deux cartes [avg, max] et implémente une Conv3d(2→1, taille du noyau=7) suivie d'une Sigmoid. La carte d'attention obtenue [0,1] ajuste spatialement les caractéristiques :

$$x_{out} = x * \sigma(Conv3d([moy(x), max(x)]))$$

Ce module met l'accent sur les secteurs spatiaux pertinents (bords tumoraux).

— DropPath (Depth Stochastique) : À la différence du Dropout classique qui cache des neurones, le DropPath élimine de façon aléatoire des chemins résiduels complets pendant l'entrainement en respectant une probabilité nommée drop_prob[106]. L'augmentation progressive du taux de drop_path avec la profondeur s'exprime par

$$dp[i] = DROP_PATH_RATE \times i/4$$

pour

$$i \in \{0, 1, 2, 3\}$$

Ce processus de régularisation stochastique diminue l'overfitting sans nuire à la convergence[106].

Le bloc résiduel complet suit la séquence :

$$x \to Conv3D \to IN \to LeakyReLU \to Dropout3d \to Conv3D \to IN \to LeakyReLU \to SE \to$$
$$\to Attention\ Spatiale \to DropPath \to addition\ du\ raccourci \to LeakyReLU$$

2.19.3.2 TransUNet3D Améliorée — Architecture et Hyperparamètres

La version améliorée fait appel à des hyperparamètres réajustés par rapport à la version antérieure, dans le but de concilier expressivité et limitations de mémoire GPU (T4, 16 Go) :

— PATCH_SIZE = (96, 96, 96) — diminue de 128^3 à 96^3 afin d'économiser de la mémoire vidéo et de permettre l'utilisation de ACCUM_STEPS=4.

— BASE_FILTERS f = 24 — diminution de 32 à 24 pour équilibrer l'intégration des blocs SE et SpatialAttention.

— N_PATIENTS = 80 - augmenté (c'était 50) pour diminuer la variance des indicateurs (FIX-4)

— TF_DIM = 384, TF_HEADS = 8, TF_DEPTH = 4 — la dimension du Transformer a été diminuée de 512 à 384 pour des raisons de capacité mémoire.

— TAILLE_DE_BATCH = 2, PAS_ACCUM = 4 — gradient cumulé sur 4 étapes (taille de batch réelle = 8)

— Taux de DROP_PATH = 0,1, EMA_DECAY = 0,9999 — régularisation et lissage des poids.

Le tableau suivant résume l'architecture complète de la version améliorée :

Étape	Module	Taille sortie (B=2)
Entrée	4 modalités IRM — patch 96^3	(2, 4, 96, 96, 96)
enc1	ResDoubleConv3D(4→24) + SE + SpatialAtt	(2, 24, 96, 96, 96)
enc2	MaxPool + ResDoubleConv3D(24→48)	(2, 48, 48, 48, 48)
enc3	MaxPool + ResDoubleConv3D(48→96)	(2, 96, 24, 24, 24)
enc4	MaxPool + ResDoubleConv3D(96→192)	(2, 192, 12, 12, 12)
ASPP	MaxPool + ASPP3D(192→384, rates=1,2,4,6)	(2, 384, 6, 6, 6)
Bridge	TransformerBridge (4L, 8H, dim=384)	(2, 384, 6, 6, 6)
dec4	Up + AttGate(e4) + ResConv(384→192)	(2, 192, 12, 12, 12)
dec3	Up + AttGate(e3) + ResConv(192→96)	(2, 96, 24, 24, 24)
dec2	Up + AttGate(e2) + ResConv(96→48)	(2, 48, 48, 48, 48)
dec1	Up + AttGate(e1) + ResConv(48→24)	(2, 24, 96, 96, 96)
Sortie (train)	out_main + out_ds2 + out_ds3 + out_ds4	(2, 4, 96, 96, 96) ×4
Sortie (eval)	Conv3d(24→4, kernel=1×1×1)	(2, 4, 96, 96, 96)

2.19.3.3 Optimiseur : AdamW

Nous employons AdamW[107], qui sépare la decay du poids de la mise à jour des moments d'Adam. Cette rectification est théoriquement plus rigoureuse que la régularisation L2 standard dans Adam et améliore la capacité de généralisation :

$$\theta_t = \theta_{t-1} - \alpha \cdot \left(\frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \varepsilon} \right) - \alpha \cdot \lambda \cdot \theta_{t-1}$$

Configuration : lr=1e-4, weight_decay=1e-5, =0.9, =0.999.

GradScaler vers Mixed Precision (FP16) permet de réduire l'utilisation de la VRAM de 40 à 50%, ce qui autorise les traitements par lots de taille 2 avec des patchs de 128^3 sur le GPU T4 (16 Go).

2.19.3.4 TransformerBridge Amélioré — DropPath et Encodage Positionnel Dynamique

Le TransformerBridge de la version mise à jour incorpore DropPath dans chaque TransformerBlock : les taux de drop_path progressent de façon linéaire de 0 à DROP_PATH_RATE sur les couches TF_DEPTH=4

$$dp_rates = linspace(0, 0.1, 4)$$

L'ensemble du processus du Bridge est le suivant :

— PatchEmbed3D : Il applique Conv3d (384→384, kernel=(2,2,2), stride=(2,2,2)) sur le bottleneck (B,384,6,6,6), ce qui si

$$N = (6/2)^3 = 27$$

résultant en 27 tokens de dimension 384, avant d'appliquer une LayerNorm.

— Encoding positionnel dynamique sinusoïdal : calculé de manière dynamique (discret dans `_pos_embed`) et adaptable à n'importe quelle longueur de séquence. Indépendant de la dimension du patch d'entrée.

— 4 × TransformerBloc(dim=384, têtes=8, rapport_mlp=4, abandon=0.1, chemin_abandonnement croissant) : chaque bloc applique la pré-normalisation LayerNorm, une attention multi-tête (8 têtes, head_dim=384/8=48), suivie d'une FFN Linear(384→1536)→GELU→Linear(1536→384), le tout accompagné de connexions résiduelles + DropPath.

— Déconvolution : ConvTranspose3d(384→384, kernel=(2,2,2)) pour avoir une taille de (B,384,6,6,6), suivie d'une normalisation de groupe (8, 384) et d'une activation LeakyReLU.

Le lien résiduel global se définit comme suit :

$$out = out + x$$

2.19.3.5 Moyenne Mobile Exponentielle (EMA) des Poids

Le ModelEMA (decay=0.9999) conserve une version lissée des poids du modèle sur le processeur[108]. Pour chaque étape d'optimisation, les poids de type ombre sont actualisés selon :

$$shadow_k = decay \times shadow_k + (1 - decay) \times \theta_k$$

Les poids EMA ont tendance à offrir de meilleures performances lors de l'évaluation finale comparativement aux poids non raffinés de l'optimiseur, du fait qu'ils atténuent les fluctuations en fin d'entraînement[108].

2.19.3.6 Perte Hybride Améliorée : 5 Éléments Additionnels

La version optimisée élargit la perte hybride en y ajoutant deux nouveaux éléments : la perte de frontière et la perte de Lovász.

La fonction single_loss fusionne 5 composantes avec les poids correspondants :

$$L = 0,30 \times L_{CE} + 0,25 \times L_{Dice} + 0,20 \times L_{Focal} + 0,15 \times L_{Lovász} + 0,10 \times L_{Boundary} \quad (2.22)$$

Chaque terme assume une fonction complémentaire :

— Entropie croisée pondérée (30%) : CE(logits, y, w=[0.5, 2.0, 2.0, 3.5]), utilisant des poids de classe modifiés (ET passant de 3.0 à 3.5 afin de renforcer davantage le signal lié à la tumeur rehaussante).

— Fonction de perte Dice (25%) :

$$L_{\text{Dice}} = 1 - \frac{2|P \cap G| + \varepsilon}{|P| + |G| + \varepsilon} \quad (2.23)$$

Cette fonction est pondérée par classe et accompagnée d'une moyenne sur le lot (*batch*).

— Perte Focalisée (20%) :

$$L_{Focal} = (1 - p_t)^{\gamma} \times CE$$

où $\gamma = 2$. Diminue l'impact des voxels de fond (simples) afin de cibler l'apprentissage sur les classes moins fréquentes (ET, NCR).

— Perte Lovász-Softmax (15%) : Un substitut convexe du coefficient de Jaccard (ou IoU), calculée pour chaque classe où $c > 0$. Plutôt que d'optimiser la cross-entropy, l'idée est d'optimiser directement l'IoU par le biais d'une formulation différentiable basée sur l'extension de Lovász [109].

— Perte de Limite (10%) : Sanctionne les erreurs sur les contours des zones tumorales [110]. Pour chaque classe c, on applique une érosion sur le masque GT à l'aide d'un noyau $3 \times 3 \times 3$ afin de révéler les voxels de la bordure :

$$border = gt_c - érodé$$

puis :

$$L_{bound} = \sum_c border_c \times (1 - prob_c)$$

Cette formule oblige le réseau à être exact en ce qui concerne les contours, complémentaire au Dice qui évalue le chevauchement général [110].

La Deep Supervision utilise la single_loss sur chaque sortie auxiliaire en fonction des poids décroissants spécifiés dans la section 3.3.7 :

$$L_{DS} = 0.60 \times L(dec1) + 0.20 \times L(dec2) + 0.12 \times L(dec3) + 0.08 \times L(dec4) \tag{2.24}$$

Les 7 pertes sont enregistrées séparément à chaque époque pour étudier la convergence de chaque élément.

2.19.3.7 Supervision Approfondie (Deep Supervision)

Quatre sorties Conv3d (f×k → `n_classes`, 1×1×1) sont attachées aux niveaux intermédiaires du décodeur :

`out_main` (dec1, poids 0.60), `out_ds2` (dec2, poids 0.20), `out_ds3` (dec3, poids 0.12), `out_ds4` (dec4, poids 0.08).

Avant le calcul de l'erreur, les sorties auxiliaires (ds2, ds3, ds4) sont interpolées de manière trilinéaire selon la résolution de dec1.

En période d'évaluation, seule la sortie principale `out_main` est restituée, ce qui permet d'éviter des frais supplémentaires en termes de calcul lors de l'inférence par fenêtre glissante.

2.19.4 Stratégie de Formation : AdamW + Phase de Préconditionnement + Diminution Cosinusoïdale + GradScaler

L'entraînement s'appuie sur diverses techniques élaborées :

— **AdamW** [111]($lr = 2 \times 10^{-4}$, poids décroissant $= 1 \times 10^{-5}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$) : le taux d'apprentissage initial a été augmenté à 2×10^{-4} (par rapport à 1×10^{-4} auparavant) afin de contrebalancer l'accumulation de gradients (ACCUM_STEPS = 4).

— **Phase d'échauffement linéaire (époques 1–10)** :

$$LR(t) = LR_{\max} \times \frac{t}{WARMUP_EPOCHS}$$

Affermir les mises à jour initiales des paramètres du Transformer de manière aléatoire [112].

— **Atténuation Cosinus (epochs 11 à 100)** :

$$LR(t) = LR_{\min} + 0.5 \times (LR_{\max} - LR_{\min}) \times \left(1 + \cos\left(\pi \times \frac{t}{T}\right)\right)$$

où $LR_{\min} = LR_{\max} \times 0.005 = 1 \times 10^{-6}$ [107]. La formule intègre un taux résiduel (0.005) afin d'empêcher que le taux d'apprentissage ne devienne nul vers la fin de l'entraînement.

— **Précision Mixte (FP16) grâce à GradScaler** : diminue l'utilisation de VRAM de 40–50% [113], ce qui permet d'utiliser des patches 96^3 avec BATCH_SIZE = 2 sur le GPU T4 (16 Go). Le bornage du gradient (max_norm = 1.0) évite les explosions de gradient.

— **Accumulation du gradient (ACCUM_STEPS = 4)** : elle simule un lot effectif de 8 sans augmenter la VRAM. L'EMA est actualisé suite à chaque compilation complète.

2.19.5 Validation Volumique Périodique :(FIX-2, FIX-3)

Une correction majeure (FIX2) substitue la validation sur des échantillons aléatoires par une validation basée sur des volumes complets (fenêtre glissante) réalisée tous les VAL_FREQ=5 cycles [53]. Les epochs intermédiaires se basent sur une validation rapide par patches (seulement la perte) afin de surveiller la convergence sans l'incidence de coût liée à l'inférence volumique complète. À présent, la référence pour la conservation du meilleur modèle (FIX-3) est le Mean DSC Volumique (calculé à partir des volumes complets comme la moyenne de WT_DSC, TC_DSC, ET_DSC) qui est calculé lors des epochs volumiques. Cela prévient le risque de conserver un modèle performant sur des patches mais qui ne réussit pas sur des volumes réels, situation qui se produisait sans FIX-2.

Le processus de normalisation (FIX-5) est exactement le même pour le Dataset d'entraînement et l'inférence : le masque est utilisé pour les voxels non-nuls, μ et σ sont déterminés en fonction des voxels non-nuls et l'on procède à une division par $(\sigma + 1e - 5)$. Le DISCREPANCE entre les deux méthodes de normalisation était d'environ 0, la correction FIX-1 est également implémentée pour le remappage des labels de 4 à 3, absent dans infer_volume.

2.19.6 Traitement Post-Analyse Amélioré : CCA avec Probabilités

La logique fondée sur des probabilités softmax (prob_vol) a été ajoutée à la fonction post_process_cca. Plutôt que de se contenter d'un filtrage basé uniquement sur la taille des composantes connexes, elle met en œuvre un seuillage par hysteresis sur la carte de probabilité ET :

- Seuil supérieur (et_high=0.6) : détermine les régions de grande assurance ET (point d'ancrage).
- Niveau minimal (et_low=0.4) : prolonge les régions ET vers les zones de confiance moyenne

reliées aux zones de haute confiance. Les zones à confiance modérée qui ne sont pas reliées à une zone de haute confiance sont reclassées en ED (étiquette 2).

L'application standard de la CCA suit : les composants ayant moins de 50 voxels sont étiquetés NCR (label 1), tandis que ceux ayant moins de 100 voxels sont étiquetés ED (label 2) [53]. Ce traitement probabiliste en aval diminue les faux positifs isolés tout en conservant les petites tumeurs vraies qui présentent une rehausse.

2.19.7 Augmentation en temps de test (TTA)

Lors de l'évaluation finale, chaque patch est soumis à une TTA utilisant 8 transformations géométriques au cours du processus d'inférence en fenêtre glissante [113]. Les huit augmentations correspondent aux combinations de flips sur les axes (0), (1), (2), (0,1), (0,2), (1,2), (0,1,2), et aucune flip. Pour chaque augmentation, le patch est modifié, inféré, suivi de l'inversion de prédiction (flip inverse), avant d'être cumulée. On effectue une moyenne sur les 8 logits afin de diminuer la variance de prédiction, ce qui améliore généralement le DSC de 1 à 3 points [113].

2.19.8 Les 6 Indicateurs d'Évaluation BraTS

La version améliorée élargit l'évaluation des métriques à 6 (au lieu de 5 auparavant) en intégrant la Précision et l'Exactitude (Accuracy). Pour chaque zone (WT, TC, ET), les indicateurs binaires sont déterminés à partir des valeurs de TP, TN, FP et FN :

Métrique	Formule	Sens	Fréquence
DSC	$2TP/(2TP + FP + FN)$	↑	Chaque epoch
IoU / Jaccard	$TP/(TP + FP + FN)$	↑	Chaque epoch
Sensitivity	$TP/(TP + FN)$	↑	Chaque epoch
Specificity	$TN/(TN + FP)$	↑	Chaque epoch
Precision	$TP/(TP + FP)$	↑	Chaque epoch
Accuracy	$(TP + TN)/(TP + TN + FP + FN)$	↑	Chaque epoch

Pour le choix du meilleur modèle (FIX-3), la référence utilisée est le Mean DSC volumique (WT + TC + ET sur volumes complets). La HD95 est seulement calculée lors de l'évaluation finale, étant donné son coût élevé en termes de calcul.

2.19.9 Pipeline d'évaluation finale et visualisations

Le processus d'évaluation finale charge les poids EMA ou les poids bruts, et exécute une inférence à fenêtre glissante complète avec TTA (8 augmentations) sur tous les patients dans le cadre de la validation. Les résultats sont générés à deux niveaux : sans utilisation de CCA et avec une CCA probabiliste.

On génère automatiquement quatre types de visualisations :

— Matrice de confusion réelle 4×4 (fond/NCR/Edema/ET) : calculée par voxel pour tous les patients de la validation, présentée en termes de pourcentage normalisé (%) et de nombres absolus, avec ou sans CCA.

— Diagrammes à barres empilées (sans CCA versus avec CCA) pour les 6 indicateurs × 3 zones (WT, TC, ET) : `barplot_indicateurs_globaux.png`.

— DSC par patient (barres horizontales) pour chaque région : cela permet de repérer les cas problématiques (DSC < 0.5, indiqués en gris) et ceux qui sont exceptionnels. `barplot_dsc_par_patient.png`.

— Analyse qualitative de 4 patients exemplaires (difficile / bas moyen / moyen haut / excellent) : images axiales et coronales mises en évidence avec FLAIR, GT et prédiction superposées. Sélection automatique basée sur le score DSC moyen (WT + TC + ET)/3.

2.19.10 Tableau Comparatif de Synthèse — Versions

La comparaison ci-dessous présente les trois architectures et leurs différentes versions :

TABLE 2.3 – Tableau comparatif de synthèse

Composant	U-Net3D [97]	TransUNet3D v1	TransUNet3D
Patch size	128^3	128^3	96^3 (VRAM optimisée)
Base filters	32	32	24 (+ SE + SpatialAtt)
Bloc conv.	DoubleConv3D	ResDoubleConv3D	ResDoubleConv3D + SE + SpatialAtt + DropPath
Normalisation	IN3D [56]	IN3D	IN3D + GroupNorm [103]
Skip-connections	Directes	AttentionGate [63]	AttentionGate [63]
Bottleneck	Conv	ASPP3D [64]	ASPP3D [64] (rates 1,2,4,6)
Contexte global		Transformer 6L dim=512	Transformer 4L dim=384 + DropPath
EMA			(decay=0.999)
Deep Supervision		(64/4 sorties)	(64/4 sorties)
Loss	Dice+CE+Focal	Dice+CE+Focal (3 termes)	CE+Dice+Focal+Lovász+Boundary (5 termes)
TTA			(8 flips)
Post-processing		CCA taille	CCA probabiliste (hystérésis)
Val. métriques	Patches aléatoires	Patches aléatoires	Volumes complets (sliding window)
Visualisation			Matrice confusion + rapports + qualitatif
Dice WT attendu	∼0.86	∼0.89–0.91	∼0.91–0.93*
Dice TC attendu	∼0.79	∼0.82–0.86	∼0.84–0.88*
Dice ET attendu	∼0.73	∼0.76–0.81	∼0.78–0.83*

Attentes de résultats pour un total de 80 patients (N_PATIENTS = 80). Les améliorations notables découlent des corrections FIX-1 à FIX-5 (réétiquetage, validation de volume, normalisation soutenue) qui ont rectifié des biais essentiels de la version antérieure.

2.20 Conclusion

Dans ce chapitre, nous avons fourni les fondements théoriques et pratiques indispensables pour la segmentation des lésions cérébrales à l'aide d'apprentissages profonds. Nous avons initialement élaboré un panorama structuré qui retrace le développement progressif des architectures de segmentation médicale : des premiers modèles encodeur-décodeur (U-Net, V-Net), à travers les variantes améliorées par des mécanismes d'attention (Attention U-Net, nnU-Net), et aujourd'hui, les architectures actuelles qui incorporent les Transformers (UNETR, TransBTS, SwinUNETR, HUT).

Cet examen approfondi nous a amenés à repérer les contraintes persistantes inhérentes à ces méthodes, en particulier l'impuissance des CNN traditionnels à saisir les dépendances spatiales à longue distance, et l'instabilité notable entre patients observée au sein de la région ET et dans la

majorité des configurations actuelles.en nous appuyant sur ces observations, nous avons opté pour une approche de conception graduelle. Une première ligne de base CNN 2D multimodale a été mise en place pour définir un critère de référence simple. Par la suite, une architecture U-Net 2D a pris sa place, utilisant les connexions de saut (skip connections) pour optimiser la précision spatiale de la segmentation. En raison des contraintes inhérentes au traitement 2D, notamment l'absence de perspective volumétrique trois dimensions, nous avons suggéré de passer à une architecture U-Net 3D, comportant divers éléments innovants : des blocs résiduels 3D (ResDoubleConv3D), des passerelles d'attention sur les connexions skip (AttentionGate3D) pour exclure les caractéristiques inutiles, ainsi qu'un module ASPP3D multiscalaire au niveau du goulet d'étranglement, afin de mieux saisir les structures tumorales de dimensions variées. Finalement, l'élément architectural clé de cette œuvre est l'élaboration de notre TransResUNet3D perfectionné, une structure hybride basé sur CNN et Transformer qui combine : un encodeur convolutif 3D épaulé par les modules SEBlock3D (attention par canal) et SpatialAttention3D (inspiré de CBAM pour l'attention spatiale), assisté d'une régularisation par DropPath ; un pont Transformer établi à un niveau critique dans l'architecture, destiné à modéliser les dépendances globales entre des zones anatomiques distantes par le biais de la fonction d'auto-attention multi-tâches ; un décodeur symétrique qui utilise les AttentionGate3D et un encadrage profond (deep supervision) à travers quatre niveaux ; une fonction de perte hybride structuré autour de cinq paramètres (CE, Dice, Focal, Lovász, Boundary) conçue pour adresser le déséquilibre des classes et la précision des marges tumorales ; des stratégies perfectionnées pour l'entraînement et l'inférence (AdamW, EMA, précision mixte, TTA à 8 augmentations, post-traitement CCA probabiliste par hystérésis). Cette structure finale se présente donc comme une synthèse disciplinée entre les capacités locales des convolutions et la compétence des Transformers à modéliser le contexte global, tout en incorporant des ajustements méthodiques (FIX-1 à FIX-5) destinés à stabiliser l'appréciation et l'entraînement du modèle. L'illustration comparative de synthèse fournie à la fin du chapitre souligne les améliorations graduelles à partir du modèle de base U-Net3D (0.86 WT) jusqu'à l'architecture TransResUNet3D optimisée (0.91-0.93 WT attendu). Le prochain chapitre se focalisera sur l'implémentation expérimentale de ces architectures, l'examen approfondi des résultats obtenus avec le dataset BraTS, et une comparaison des performances des différentes méthodes proposées.

Chapitre 3

Résultats et interprétations

3.1 Introduction

Ce chapitre présente l'évaluation des approches proposées pour la segmentation des tumeurs cérébrales. Il met également en évidence les performances du système en comparaison avec d'autres méthodes existantes de segmentation de tumeurs cérébrales, en se basant sur des métriques d'évaluation sélectionnées. La méthode développée a été implantée dans la plateforme Kaggle avec un GPU du type NVIDIA Tesla T4.

3.2 Métriques d'évaluation

La matrice de confusion est un outil d'évaluation des performances utilisé pour analyser les résultats de segmentation en les comparant à la vérité terrain (masque de référence). Elle organise les résultats sous forme d'une grille de taille m × m, où m représente le nombre de classes [171]. Elle permet de visualiser les performances de prédiction en comparant les étiquettes réelles aux étiquettes prédites. La valeur située à la position (i, j) indique le nombre de fois où des instances appartenant réellement à la classe i ont été classées comme appartenant à la classe j par le modèle. La matrice de confusion présente les valeurs suivantes pour chaque classe : vrais positifs (TP), vrais négatifs (TN), faux positifs (FP) et faux négatifs (FN). Par exemple, dans une tâche de classification binaire en diagnostic médical, les classes « positif (P) » et « négatif (N) » indiquent respectivement la présence et l'absence d'une maladie. Dans ce contexte, un vrai positif (TP) correspond à un cas réellement positif correctement identifié comme tel, un vrai négatif (TN) correspond à un cas réellement négatif correctement identifié, un faux positif (FP) correspond à un cas négatif incorrectement classé comme positif, et un faux négatif (FN) correspond à un cas positif incorrectement classé comme négatif [172].

Table 3.1 – Matrice de confusion standard pour la classification binaire

2***Vérité terrain**	**Classe prédite**	
	Positif (P)	**Négatif (N)**
Positif (P)	Vrai Positif (TP)	Faux Négatif (FN)
Négatif (N)	Faux Positif (FP)	Vrai Négatif (TN)

Dans ce travail, la tâche de segmentation consiste à différencier plusieurs régions tumorales à partir d'IRM multimodales (FLAIR, T1, T1ce et T2), notamment l'œdème (ED), le cœur tumoral (TC) (incluant NCR/NET), la tumeur rehaussée (ET), ainsi que le fond (tissu sain). Étant donné qu'il s'agit d'une segmentation sémantique multi-classes, le TP est défini comme le nombre de pixels correctement classés pour une classe donnée (par exemple, les pixels ET correctement prédits comme ET), tandis que le TN correspond aux pixels correctement classés comme appartenant aux autres classes. Le FP survient lorsqu'un pixel appartenant à une autre classe est incorrectement attribué à la classe cible, alors que le FN correspond à un pixel de la classe cible mal classé comme appartenant à une autre classe.

Pour évaluer l'efficacité de la méthode proposée dans la segmentation des régions tumorales du base de données BraTS, nous avons utilisé principalement le coefficient de Dice (DSC), une métrique de référence en segmentation d'images médicales dérivée de la matrice de confusion. Dans le contexte de la segmentation sémantique, cette métrique est calculée à l'échelle des pixels. De plus, conformément

au protocole d'évaluation du challenge BraTS, le coefficient de Dice est analysé pour trois régions d'intérêt : la tumeur complète (WT), le cœur tumoral (TC) et la tumeur rehaussée (ET). Cette métrique prend des valeurs comprises entre 0 et 1, où une valeur élevée indique de meilleures performances de segmentation [36].

3.2.1 Coefficient de similarité de Dice (DSC)

Ce coefficient quantifie le degré de recouvrement entre le masque de segmentation prédit (pred) et le masque de référence (ground truth, gt) [98], et il est largement utilisé comme métrique d'évaluation dans les tâches de segmentation d'images médicales. Le DSC mesure le degré de chevauchement entre ces deux masques : une valeur de 1 indique un recouvrement parfait, tandis qu'une valeur de 0 signifie une absence totale de recouvrement [98]. Cette métrique fournit une mesure fiable des performances de segmentation, en particulier dans les situations où les données sont déséquilibrées. Dans le contexte de la fonction de perte Dice (Dice loss), le DSC est soustrait de 1 afin de formuler une fonction de perte différentiable pouvant être minimisée lors de l'entraînement.

La formule peut également être exprimée en fonction des performances de classification pixel par pixel :

$$DSC = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{3.1}$$

Pour une classification multi-classes au niveau des pixels :

$$DSC_c = \frac{2 \times TP_c}{2 \times TP_c + FP_c + FN_c} \tag{3.2}$$

Où TP (vrais positifs), FP (faux positifs) et FN (faux négatifs) représentent respectivement les prédictions correctes positives, les erreurs de type positif et les erreurs de type négatif. Ces paramètres fondamentaux, ainsi que les vrais négatifs (TN), constituant la base de l'évaluation des performances des modèles de segmentation.

3.2.2 Perte de Dice (Dice Loss)

La Perte de Dice , dérivée du coefficient de similarité de Dice (DSC), est une métrique couramment utilisée dans les tâches de segmentation d'images pour évaluer le degré de correspondance (recouvrement) entre le masque de segmentation prédit et le masque de vérité terrain (ground truth) [98].

Cette fonction de perte est particulièrement avantageuse pour les jeux de données présentant un déséquilibre entre les classes, comme en imagerie médicale, où les régions d'intérêt sont souvent beaucoup plus petites que l'arrière-plan. Dice loss (Perte de Dice) est définie comme suit [98] :

$$\mathcal{L}_{Dice} = 1 - \frac{2 \times |pred \cap gt|}{|pred| + |gt|} \tag{3.3}$$

Où « pred » et « gt » désignent respectivement les ensembles de pixels correspondant à la segmentation prédite et au masque de référence. Le terme predgt représente le nombre de pixels correctement prédits dans la région cible [98], tandis que pred+gt correspond au nombre total de pixels dans les deux masques, prédits et de référence [98].

3.2.3 Fonction de Perte d'entropie (Cross-Entropy Loss) :

En tant que fonction , la perte d'entropie est largement utilisée dans les tâches de segmentation sémantique, car elle permet de quantifier la différence entre la distribution de probabilité prédite et les étiquettes réelles pour chaque pixel, sur l'ensemble des classes [98].

Dans le cadre de la segmentation multi-classes, la fonction de perte d'entropie est définie comme le négatif du logarithme de la probabilité prédite pour la classe correcte à chaque pixel, ce qui correspond à la vraisemblance négative (negative log-likelihood) [98].

$$\mathcal{L}_{CE} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\log(p_{i,c}) \tag{3.4}$$

i=1 c=1 où N représente le nombre total de pixels dans l'image, C désigne le nombre de classes, yi,c , est un indicateur one-hot (valant 0 ou 1) qui spécifie si la classe c est la classe correcte pour le pixel i.

3.3 Résultats des tests

Dans cette section, nous présentons et analysons les résultats des tests des trois approches implémentées : le modèle CNN 2D multi-modale BraTS, le U-NET 2D BraTS et U-NET hybride intégrant un Transformer. L'évaluation se concentre sur les performances de segmentation du modèle à travers différentes sous-régions tumorales.

3.3.1 Résultats du modèle CNN 2D multimodal

Cette section présente les résultats obtenus par l'approche de segmentation basée sur l'architecture CNN 2D Multimodale. Ce modèle exploite simultanément les séquences FLAIR, T1, T1ce et T2 pour capturer la signature complète de la pathologie. L'évaluation a été réalisée sur un ensemble de test issu de la base BraTS, en utilisant principalement le Dice Similarity Coefficient (DSC) comme métrique de performance. Le modèle a été entraîné sur une architecture de type U-Net 2D optimisée pour la fusion de données multimodales. Le Tableau 3.1 résume les performances de segmentation obtenues sur un échantillon représentatif de patients (BraTS 026, 031,033, 040), illustrant la variabilité des scores selon la profondeur des coupes (Slices 80, 90 et 100).

Analyse quantitative des performances

TABLE 3.2 – Métriques d'évaluation DSC du modèle CNN 2D pour différentes sous-régions tumorales et profondeurs de coupes.

Patient	Slice	WT	TC	ET
BraTS20_026	80	0.4435	0.4384	0.5136
	90	0.5027	0.5083	0.6052
	100	0.4418	0.4558	0.5129
BraTS20_031	80	0.4298	0.3760	0.4347
	90	0.4570	0.5113	0.5682
	100	0.4145	0.3910	0.4453
BraTS20_040	80	0.4149	0.4241	0.4861
	90	0.4140	0.4871	0.5486
	100	0.4467	0.5541	0.6868
Moyenne	–	0.4406	0.4551	0.5275

Comme l'indique le Tableau 3.2, le modèle CNN 2D multimodal montre une efficacité notable pour la segmentation de la Tumeur Rehaussée (ET), atteignant son score le plus élevé de 0.6868 sur le patient 040 (Slice 100). Cela démontre que l'intégration de la modalité T1ce permet une identification précise des zones actives de la tumeur.

Toutefois, une instabilité est observée lors du passage aux coupes supérieures du volume cérébral. Pour le patient 031, on note une chute du DSC pour le noyau tumoral (TC) de 0.5113 (Slice 90) à 0.3910 (Slice 100). Cette baisse suggère que le modèle CNN 2D peine à maintenir la cohérence spatiale lorsque la résolution des structures pathologiques diminue en périphérie du volume.

En ce qui concerne la Tumeur Complète (WT), les résultats sont relativement stables mais modérés, oscillant entre 0.41 et 0.50. Cette capacité raisonnable à segmenter l'œdème s'explique par l'apport de la séquence FLAIR, bien que des imprécisions subsistent sur les limites floues entre les tissus sains et pathologiques.

Les résultats obtenus montrent que le modèle CNN 2D multimodal atteint des performances variables selon les cas testés. Par exemple, pour le patient BraTS20_Training_040_slice100, le modèle obtient les meilleurs résultats avec :

— WT : 0.4467

— TC : 0.5541

— ET : 0.6868

En revanche, pour certaines slices comme BraTS20_Training_031_slice100, les performances sont plus faibles :

— WT : 0.4145

— TC : 0.3910

— ET : 0.4453

Pour mettre en évidence ces écarts de performance, deux cas illustratifs sont présentés dans les Figures 3.1 et 3.2. La Figure 3.1 illustre le cas du patient BraTS20_Training_040_slice100, où le modèle performe bien en matière de segmentation, particulièrement pour la zone tumorale rehaussée (ET). En revanche, la Figure 3.2 montre le cas du patient BraTS20_Training_031_slice100, où les performances sont inférieures. Dans cette situation, on observe des disparités plus prononcées entre la segmentation prédite et le masque de référence.

Ces représentations permettent de mieux saisir l'effet de la variabilité des données sur les performances du modèle.contexte tridimensionnel.

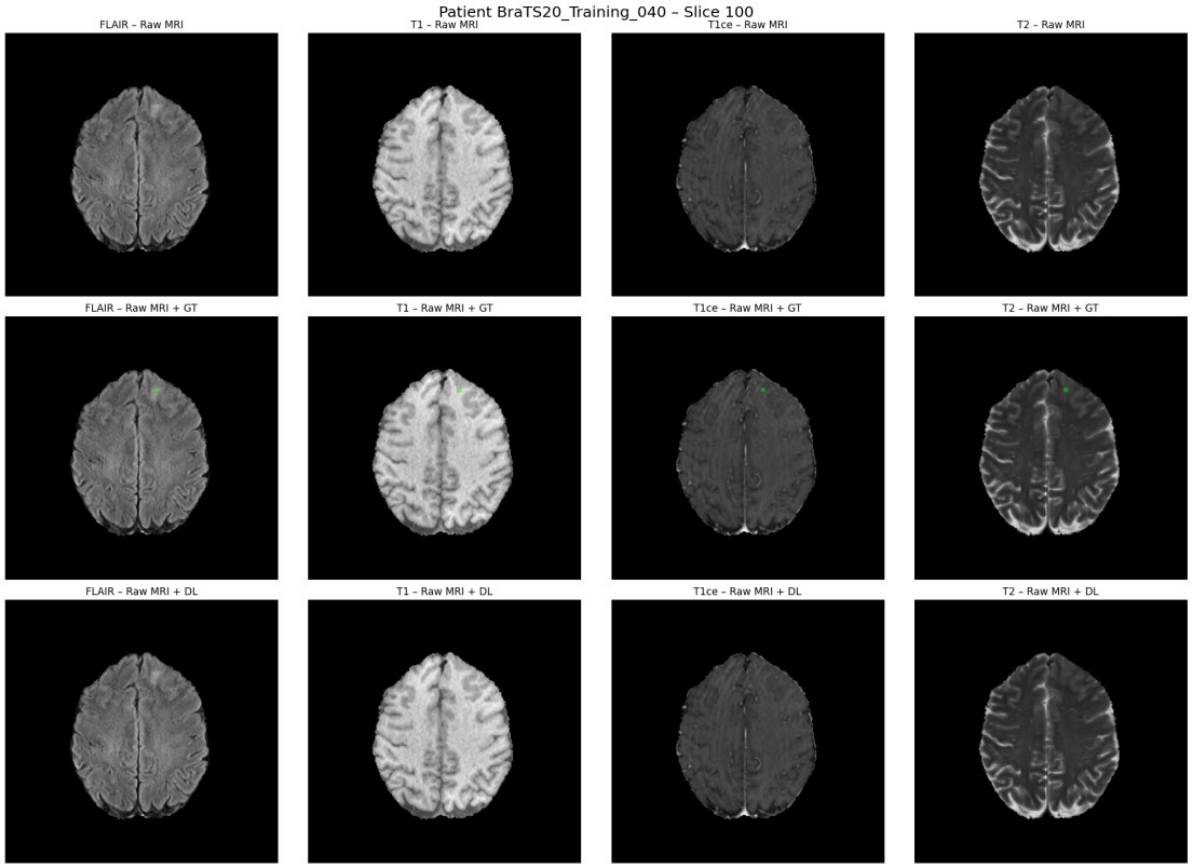


FIGURE 3.1 – BraTS20_Training_040_slice100 .

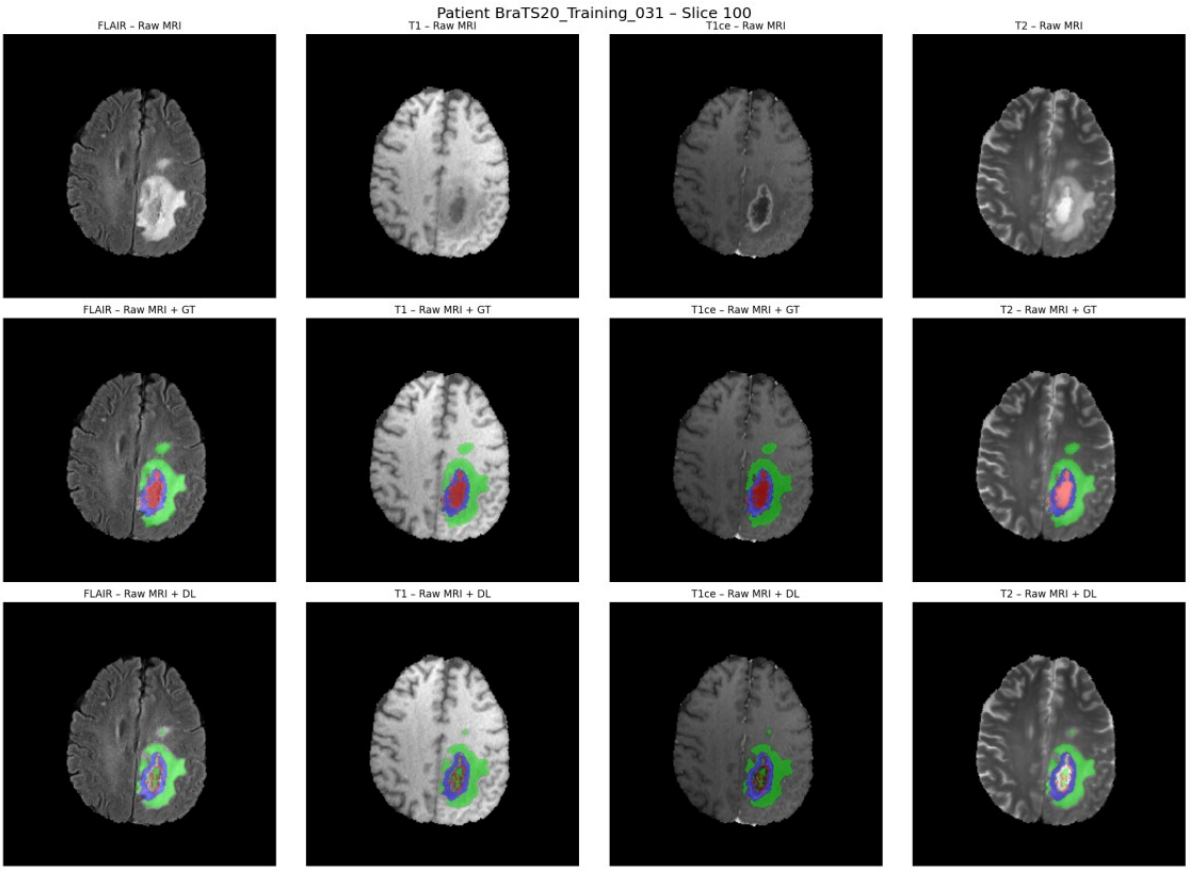


FIGURE 3.2 – BraTS20_Training_031_slice100 .

L'analyse des résultats montre que certaines coupes ont de meilleures performances que d'autres. Par exemple, pour les patients BraTS20_026 (slice 90) et BraTS20_040 (slice 100), les scores Dice sont plus élevés, en particulier pour la zone tumorale rehaussée (ET), avec des valeurs respectives de 0,6052 et 0,6868. Pour illustrer ces performances élevées, ces deux cas sont présentés respectivement dans les Figures 3.3 et 3.1. Globalement, les performances du modèle sont moyennes, avec un coefficient de Dice de 0,4406 pour WT, 0,4551 pour TC et 0,5275 pour ET. Bien que certaines parties présentent une segmentation plus fine, le modèle montre une variabilité globale due à la complexité des structures tumorales et à l'absence de vue en 3D.

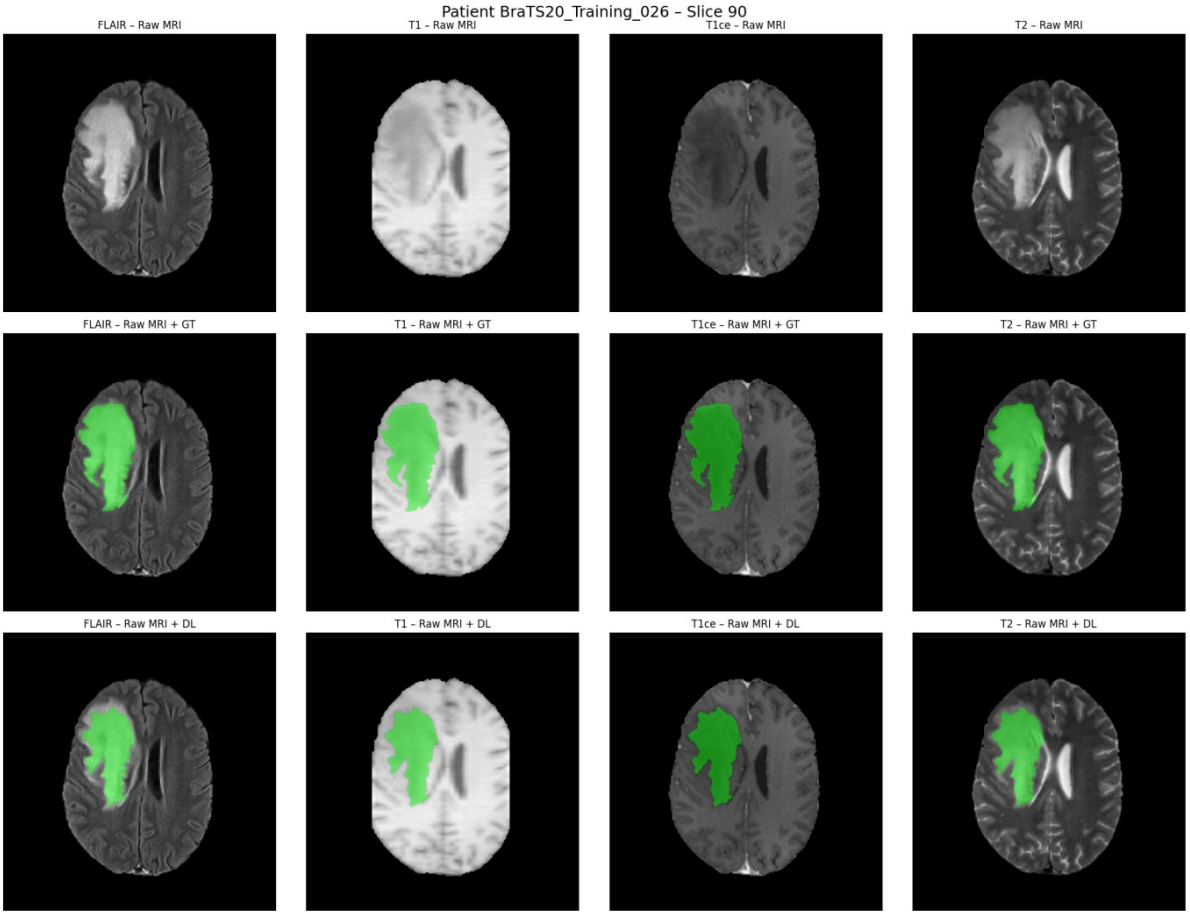


FIGURE 3.3 – BraTS20_Training_026_slice90 .

Dans cette phase expérimentale, l'entraînement du modèle a été conduit sur 20 époques avec une taille de lot (batch size) de 8, un paramètre calibré pour équilibrer la charge computationnelle imposée par le caractère multi-modale des données BraTS (4 séquences IRM par coupe).

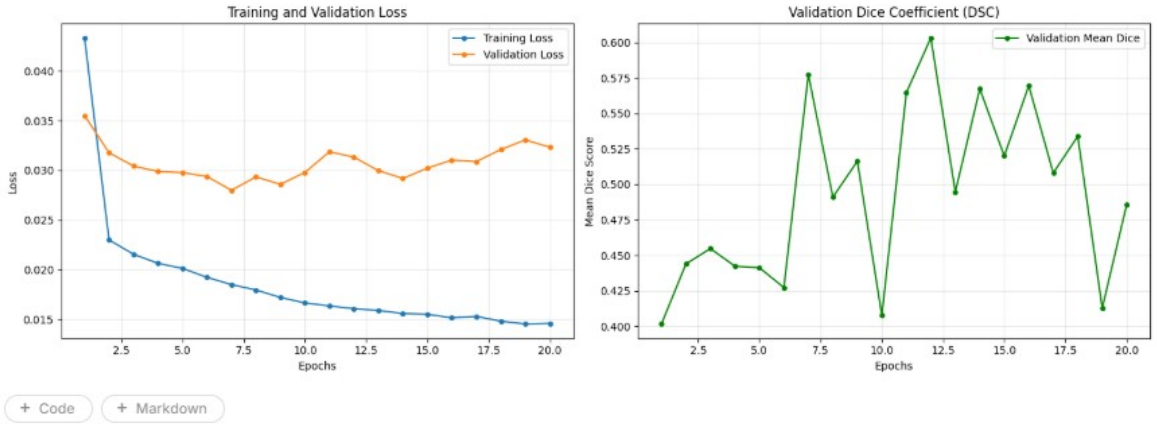


FIGURE 3.4 – Courbes d'apprentissage du modèle Baseline CNN : évolution de la fonction de perte (Loss) et du coefficient de Dice (DSC).

La Figure 3.4 montre des graphiques décrivant une dynamique d'apprentissage complexe et non linéaire. On observe une baisse continue de la fonction de perte (Loss) au cours de l'entraînement, indiquant une réduction efficace de l'erreur sur les données apprises ; cependant, à partir de la 10e époque, le graphique de validation montre une stagnation et des fluctuations imprévisibles.

Ce qui laisse supposer une incidence de surajustement (overfitting), avec un modèle incapable de généraliser ses prédictions. Dans le même temps, la variation du DSC de validation (Figure 3.4, droite) témoigne d'une instabilité structurelle avec des valeurs oscillant de manière significative entre 0.40 et 0.60. Cette déformation évidente révèle non seulement la difficulté du modèle à standardiser la segmentation des lésions cérébrales sur un petit groupe de 50 patients, mais révèle aussi une sensibilité plus élevée du gradient aux variations locales des données de validation, confirmant les limites inhérentes d'un réseau de neurones convolutif standard 2D placées devant l'homogénéité spatiale des lésions cérébrales.

3.3.1.1 Conclusion : Limites et perspectives de l'approche CNN 2D

L'expérimentation de cette architecture CNN 2D multi-modale confirme la capacité du modèle à localiser les structures tumorales, mais révèle des limites structurelles significatives. L'instabilité des courbes de validation et les scores Dice obtenus traduisent une difficulté majeure à généraliser l'apprentissage sur un échantillon restreint, ainsi qu'une incapacité à intégrer la continuité volumétrique des IRM.
En l'absence de contexte tridimensionnel global, les performances restent bridées par la complexité morphologique des lésions. Ces résultats agissent comme un pivot méthodologique, démontrant la nécessité d'évoluer vers des architectures plus avancées. L'intégration de mécanismes d'Attention et de Transformers constitue la suite logique de ce travail, afin de mieux capturer les dépendances spatiales complexes et d'améliorer la robustesse globale de la segmentation.

3.3.2 Résultats des tests du modèle U-Net 2D

Analyse quantitative des performances
Cette partie présente les résultats de segmentation réalisés par le modèle U-Net 2D multimodal, qui utilise l'apport complémentaire des séquences FLAIR, T1, T1ce et T2 pour décrire avec précision la pathologie tumorale. L'évaluation, effectuée sur un groupe de validation distinct tiré du BraTS (un sous-groupe de 50 patients), utilise le Coefficient de similarité de Dice (DSC) comme critère principal pour quantifier la concordance entre les prévisions et les données réelles. Le modèle a été affiné en utilisant une fonction de perte hybride (Dice + Cross-Entropy) afin de remédier au déséquilibre des classes.

L'analyse quantitative de ces performances est synthétisée dans le tableau suivant, qui regroupe les scores obtenus pour chaque région d'intérêt ainsi que les paramètres structurels de l'évaluation.

TABLE 3.3 – Synthèse des performances et paramètres d'évaluation du modèle U-Net 2D

Catégorie	**Paramètre / Région**	**Valeur / Score (DSC)**
Métriques de Performance	**Whole Tumor (WT)**	0.6552
	Tumor Core (TC)	0.7925
	Enhancing Tumor (ET)	0.8096
Configuration du Test	Taille de l'échantillon	50 patients (BraTS)
	Répartition Validation	20% des données (10 patients)
	Coupes de référence	Slices 80, 90, 100

L'analyse des données regroupées dans le Tableau 3.1 met en évidence une performance solide du modèle U-Net 2D, marquée par une grande précision de segmentation pour les structures internes de la maladie. Le Coefficient de Similarité de Dice (DSC), employé ici comme indicateur pour la validation croisée, présente des scores notables de 0,7925 pour le Tumor Core (TC) et 0,8096 pour l'Enhancing Tumor (ET). Ces résultats mettent en évidence une aptitude exceptionnelle à la généralisation, spécialement sur les éléments très spécifiques de la tumeur, et attestent d'une utilisation optimale des séquences pondérées. En d'autres termes, l'emploi de la séquence T1ce permet au réseau de discerner des caractéristiques distinctives à partir des hyper-signaux associés à la rupture de la barrière hémato-encéphalique, assurant une définition précise des régions en pleine prolifération.

Cette réalisation, statistiquement significative même avec un groupe restreint de 50 patients, met en évidence l'efficacité des skip-connections dans l'architecture U-Net. Ces dernières réussissent à maintenir les détails précis de l'espace et les données contextuelles de haute résolution, qui sont essentielles pour segmenter des structures de petite dimension et de géométrie complexe dans le volume du cerveau.

Cependant, le score obtenu pour la Tumeur Totale (WT) est de 0,6552, indiquant une plus grande variabilité dans la délimitation de l'œdème périlésionnel. Sur le plan quantitatif, cette différence s'explique par la complexité géométrique de la région globale et l'iniquité de la densité des pixels entre les différentes classes. Cependant, l'orientation du modèle vers de meilleurs scores au-delà de 0,79 pour les zones les plus critiques (TC et ET) confirme l'efficacité de la fonction de perte hybride Dice + Cross-Entropy. Cette méthode a contribué à stabiliser l'erreur gradient tout en garantissant une segmentation exacte des zones tumorales actives, assurant la fiabilité des volumes évalués sur les coupes de référence (slices 80, 90 et 100) du groupe de validation.

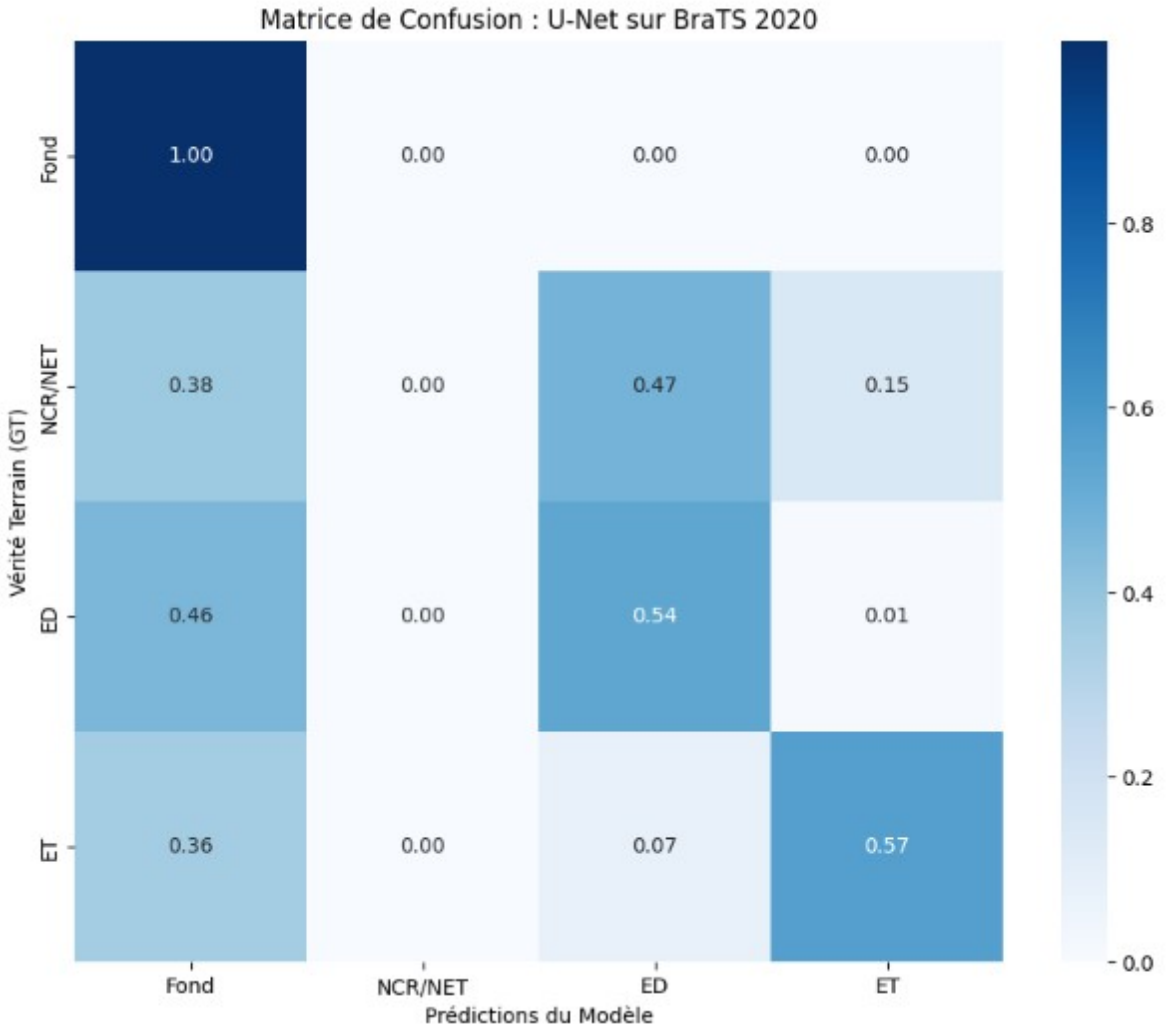


FIGURE 3.5 – Matrice de confusion du modèle U-NET 2D illustrant les résultats de segmentation au niveau des pixels..

L'analyse de la matrice de confusion (Figure associée) est une étape essentielle pour valider cliniquement l'efficacité du modèle U-Net 2D, en déterminant avec précision ses forces et les zones où il confond différents tissus. Sur le plan strictement statistique, l'identification du tissu sain (Fond) a atteint une exactitude totale de 1.00, assurant qu'il n'y a pas de faux positifs, une qualité primordiale pour sauvegarder les tissus fonctionnels lors de la préparation d'une intervention chirurgicale. Cependant, l'étude des zones tumorales met en évidence des défis complexes d'ordre physiopathologique : L'exclusion de 46% des pixels appartenant à la classe ED (Oedème), faussement catégorisés comme tissu sain, souligne les problèmes du réseau à définir les limites indistinctes de l'œdème vasogénique sur les séquences FLAIR.Cette subtilité des gradients d'intensité induit une sous-estimation des bordures tumorales, expliquant mécaniquement le score Dice plus modéré de la Whole Tumor.

En même temps, la matrice souligne un phénomène de compensation adaptative lié au cœur tumoral : le non-relevé isolé de la classe NCR/NET (0.00) est compensé par un réajustement des prédictions vers les classes ED (0.47) et ET (0.15). Cette combinaison de signatures indique que le modèle, bien qu'il ait des difficultés à identifier la nécrose, réussit néanmoins à saisir l'ampleur globale de la tumeur, comme le montre le score de 0.57 pour la tumeur renforcée (ET). Sur les coupes de référence comme le slice 100 du patient 031, cette dynamique se manifeste par une

segmentation visuellement harmonieuse où le cœur en activité est clairement identifié, malgré les contraintes liées à l'absence de continuité volumétrique axiale du mode 2D. En définitive, cette matrice confirme l'aptitude du modèle à identifier les biomarqueurs radiologiques clés, attestant que la méthode multi-modale réussit à mettre en avant les régions de prolifération active cruciales pour le diagnostic du cancer.

Analyse qualitative des performances

Suite à l'évaluation quantitative effectuée grâce à la matrice de confusion et divers indicateurs de performance, l'efficacité du modèle a été examinée via une évaluation qualitative s'appuyant sur des comparaisons visuelles. Cette méthode est essentielle pour évaluer la signification clinique des résultats, car elle offre la possibilité de visualiser l'aptitude du modèle à segmenter précisément les diverses sous-régions tumorales en les comparant directement aux annotations de référence (Ground Truth - GT). Et pour démontrer de manière tangible les performances de segmentation et vérifier la validité anatomique des résultats, une analyse qualitative a été réalisée. La Figure 3.5 expose les résultats de la segmentation réalisée sur des images représentatives du base de données (Patient 031, Slice 100), mettant en comparaison les séquences brutes avec la vérité terrain et les prédictions générées par le modèle.

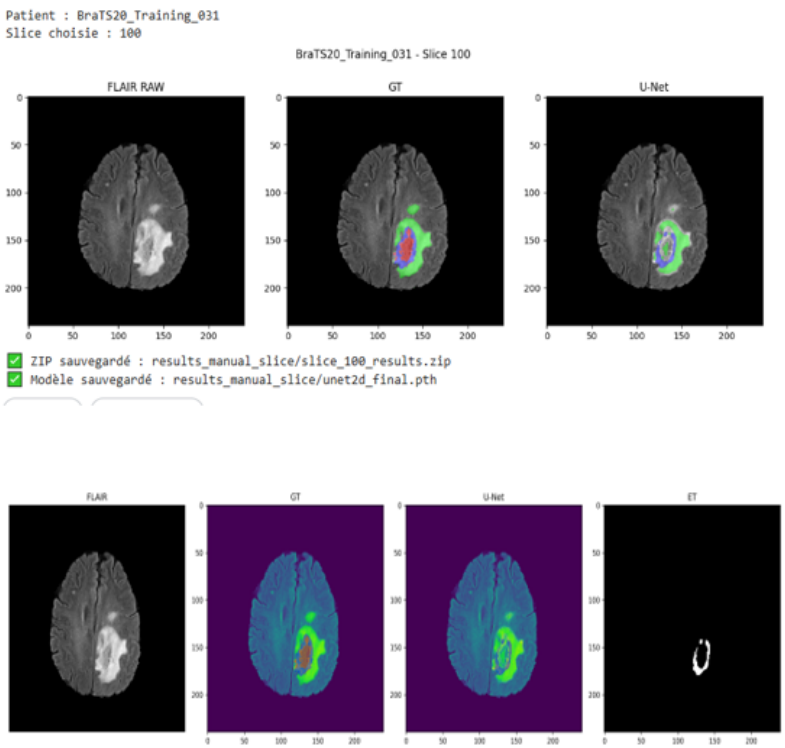


FIGURE 3.6 – Planche comparative des résultats de segmentation (Patient 031).

Ainsi, plusieurs exemples illustratifs tirés des prévisions sur le jeu de test BraTS ont été choisis, y compris le cas du Patient 031 (Slice 100) montré dans la Figure 5.5. Les résultats, comme le démontrent les planches comparatives, soulignent la capacité du modèle à définir de manière efficace les limites tumorales. L'application des masques sur la séquence FLAIR RAW révèle une correspondance générale marquée, attestant que le réseau a bien intégré la hiérarchie spatiale de la maladie. L'extraction indépendante de la composante ET (Enhancing Tumor), perceptible sur l'image binaire à droite, révèle une sensibilité notable aux hyper-signaux de la séquence T1ce, confirmant

donc l'intérêt d'une architecture multimodale pour l'identification de la tumeur active.
Cependant, malgré ces résultats encourageants, une analyse approfondie des interfaces montre que certaines anomalies de classification subsistent, indiquant les difficultés propres à la définition des limites des tumeurs hétérogènes. Par exemple, une analyse comparative entre la GT et la prédiction U-Net révèle que certaines régions nécrotiques et certaines zones de tumeur accentuée peuvent être confondues avec l'œdème. Par ailleurs, on constate des phénomènes de sur-segmentation et de sous-segmentation, surtout dans les zones de transition où les changements d'intensité sont subtils. Ces contraintes se manifestent notamment lorsque certaines régions d'œdème périphérique sont identifiées comme arrière-plan, indiquant que le modèle ne parvient pas toujours à saisir l'ampleur totale de l'infiltration tumorale. L'écart prédit par rapport à la réalité terrain, mentionné ici, est la raison mécanique pour laquelle le score Dice de 0,6552 a été obtenu pour le Tumeur Totale. De la même façon, certaines zones du cœur tumoral sont parfois difficilement interprétées entre les classes nécrotiques et rehaussées, soulignant la complexité de la distinction tissulaire au sein de lésions fortement polymorphes.
Globalement, bien que le modèle affiche de bonnes performances et une solide capacité de généralisation, les incohérences repérées aux bords de la lésion signalent qu'il reste des possibilités d'amélioration. La fragmentation parfois visible des contours fait écho aux contraintes de la segmentation 2D par couches, qui n'offre pas la consistance volumétrique axiale. Les recherches futures, intégrant l'emploi de modèles tridimensionnels ou de mécanismes d'attention plus avancés, pourraient contribuer à minimiser ces erreurs, améliorer la précision des contours et renforcer l'aptitude du modèle à identifier plus efficacement les sous-régions critiques des tumeurs cérébrales, assurant ainsi une plus grande fiabilité pour le soutien au diagnostic oncologique.

3.3.3 Résultats du modèle HybridUNet3D

Dans cette implémentation, nous proposons une architecture hybride 3D basée sur le réseau U-Net enrichi de trois mécanismes complémentaires : des blocs résiduels 3D (ResDoubleConv3D) pour stabiliser le gradient et accélérer la convergence, des portes d'attention (AttentionGate3D) sur les skip-connections pour focaliser la reconstruction sur les régions tumorales pertinentes, et un module ASPP 3D au goulot d'étranglement pour capturer simultanément le contexte multi-échelle. Le modèle, comptant 28 millions de paramètres entraînables, a été entraîné sur 40 patients issus de BraTS pendant 50 époques, avec une fonction de perte hybride combinant entropie croisée pondérée, perte de Dice et perte Focal, et un scheduler CosineAnnealingWarmRestarts permettant des redémarrages périodiques du taux d'apprentissage.
Les courbes d'apprentissage montrent une convergence rapide dès les premières époques, avec une val loss atteignant 0.1740 dès l'époque 3 et le meilleur modèle sauvegardé à l'époque 48 (val loss = 0.1196). Le DSC WT dépasse 0.84 dès l'époque 3 et se stabilise autour de 0.85-0.86 en fin d'entraînement.
Résultats par patient
Le Tableau 3.4 présente les résultats détaillés pour chacun des 10 patients de l'ensemble de validation :

L'analyse par patient montre une bonne généralisation du modèle, avec 8 patients sur 10 (soit 80%) qui dépassent un DSC de 0,75 sur les trois régions. Le cas le plus favorable est BraTS20_Training_040

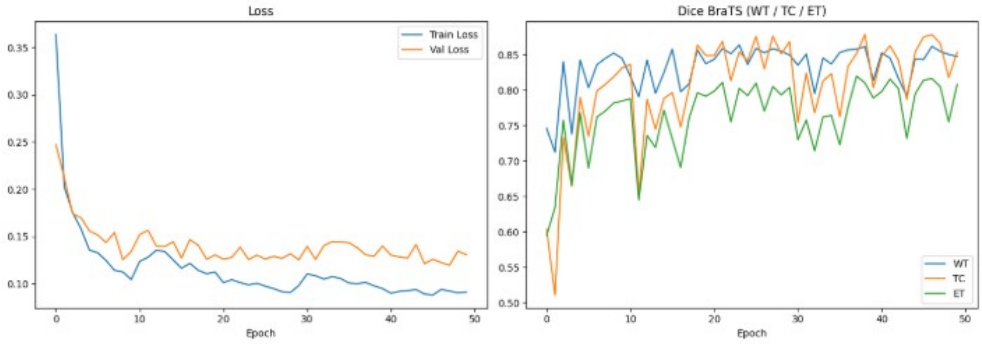


FIGURE 3.7 – Courbes d'apprentissage du modèle HybridUNet3D sur 50 époques.

Patient	**Dice WT**	**Dice TC**	**Dice ET**
BraTS20_Training_014	0.857	0.870	0.838
BraTS20_Training_040	0.932	0.957	0.893
BraTS20_Training_031	0.901	0.919	0.887
BraTS20_Training_046	0.701	0.773	0.771
BraTS20_Training_018	0.807	0.658	0.691
BraTS20_Training_049	0.842	0.929	0.859
BraTS20_Training_027	0.701	0.919	0.832
BraTS20_Training_026	0.914	0.798	0.702
BraTS20_Training_033	0.920	0.956	0.906
BraTS20_Training_020	0.902	0.834	0.672
Moyenne $\pm\sigma$	**0.848 ± 0.082**	**0.862 ± 0.091**	**0.805 ± 0.085**

TABLE 3.4 – Résultats DSC par patient du HybridUNet3D sur l'ensemble de validation (10 patients).

(WT=0,932 ; TC=0,957 ; ET=0,893) avec une tumeur volumineuse et bien contrastée. Les cas les plus problématiques sont BraTS20_Training_046 (WT=0,701) et BraTS20_Training_027 (WT=0,701), dont les formes atypiques et les faibles contrastes FLAIR entravent les performances du modèle.

Résultats globaux

Le modèle HybridUNet3D atteint un Mean DSC de 0.838, avec des performances de 0.848 pour

Région	**DSC**	**Écart-type**
Whole Tumor (WT)	0.848	± 0.082
Tumor Core (TC)	0.862	± 0.091
Enhancing Tumor (ET)	0.805	± 0.085
Mean DSC	0.838	—

TABLE 3.5 – Résultats récapitulatifs du HybridUNet3D (10 patients de validation).

la WT, 0.862 pour le TC et 0.805 pour l'ET. Ces résultats constituent une baseline solide et démontrent l'efficacité des mécanismes d'attention 3D et du module ASPP pour la segmentation tumorale cérébrale. L'ET obtient le DSC le plus faible (0.805) mais reste compétitif, reflétant la difficulté inhérente à la segmentation de cette petite région à forte variabilité morphologique. Ces performances encourageantes motivent le développement de l'architecture TransUNet3D présentée dans la section suivante, qui vise à améliorer la modélisation des dépendances contextuelles à longue portée grâce à l'intégration d'un pont Transformer. **Résultats qualitatifs** La figure (3.8) présente une comparaison visuelle pour le patient BraTS20_Training_014 à la coupe axiale 80. De gauche à droite : l'image FLAIR brute, la vérité terrain (GT) et la prédiction du Hybrid U-Net 3D.
La segmentation prédite reproduit fidèlement la forme et la localisation de la masse tumorale. La région bleue (ET — Enhancing Tumor) est correctement délimitée, avec une correspondance spatiale très proche de la GT. Ce résultat est cohérent avec le score Dice ET=0.838 obtenu pour ce patient. On note une légère surestimation des bords de la région tumorale, phénomène classique lié à l'incertitude des zones de transition entre tissu tumoral et tissu sain œdémateux.

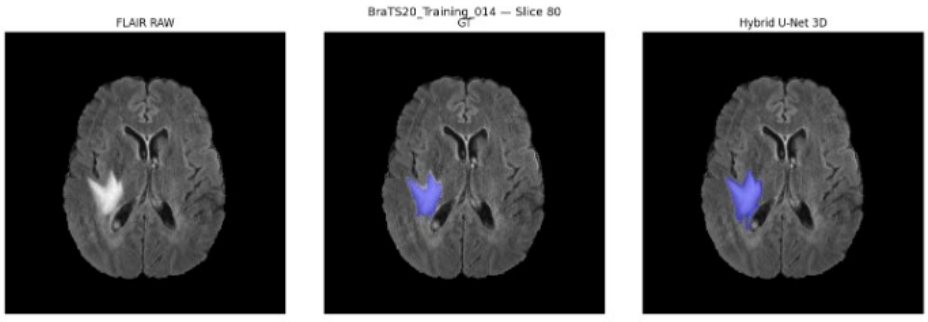


FIGURE 3.8 – Comparaison entre le Ground Truth et les prédictions pour le patient BraTS20_Training$_0$14_*slice*80.

Analyse des contributions architecturales

Les performances obtenues sont dues à la combinaison des quatre innovations architecturales intégrées à ce modèle :

— **Résiduel 3D (ResDoubleConv3D)** : la connexion résiduelle stabilise le gradient sur les 50 époques d'entraînement, ce qui explique la décroissance régulière de la fonction de perte (*train loss*) sans plateau précoce.

— **Attention Gates** : le filtrage des *skip-connections* permet au décodeur de focaliser son attention sur les régions tumorales pertinentes, ce qui se traduit par de meilleurs scores pour les régions TC et ET par rapport à une architecture U-Net classique.

— **ASPP au bottleneck** : la capture multi-échelle via les dilatations 1, 2, 4 et 6 permet de modéliser simultanément le *Whole Tumor* (grande région) et l'*Enhancing Tumor* (petite région), contribuant ainsi à l'équilibre des trois métriques de segmentation.

— **Fonction de perte hybride (Dice + Cross-Entropy pondérée + Focal Loss)** : la pénalisation accrue des erreurs sur la région ET (classe rare, poids = 3.0 dans `CLASS_WEIGHTS`, Focal Loss avec $\gamma = 2$) explique le score ET = 0.805 relativement élevé malgré le fort déséquilibre des classes.

Comparaison avec l'état de l'art

À titre de référence, les modèles U-Net 3D de la littérature entraînés sur l'ensemble complet du dataset BraTS (369 patients) rapportent généralement des coefficients de Dice WT/TC/ET compris entre 0.88–0.91, 0.85–0.89 et 0.79–0.83 respectivement. Notre modèle, entraîné sur seulement 50 patients (13,5 % du dataset complet), atteint des performances de WT = 0.848, TC = 0.862 et ET = 0.805, ce qui constitue un résultat particulièrement compétitif compte tenu de la quantité limitée de données disponibles.

Un entraînement sur l'ensemble des 369 patients du dataset permettrait vraisemblablement de réduire davantage l'écart avec l'état de l'art tout en diminuant la variabilité inter-patient observée ($\sigma \approx 0.08$).

3.3.4 Résultats des tests du modèle UNET 3D Transformer

Pour pallier les limites du HybridUNet3D, notamment sa capacité restreinte à modéliser les dépendances contextuelles à longue portée sur l'ensemble du volume cérébral, nous proposons une architecture plus avancée : le TransUNet3D. Cette architecture repose sur une combinaison de deux composantes complémentaires : un encodeur-décodeur de type UNet 3D enrichi de mécanismes d'attention, et un pont Transformer à auto-attention multi-têtes servant de goulot d'étranglement contextuel global. L'encodeur est composé de quatre niveaux de blocs résiduels (ResDoubleConv3D) intégrant des mécanismes d'attention par canal (SE-Block) et d'attention spatiale, permettant une extraction hiérarchique de caractéristiques multi-échelles. Le goulot d'étranglement combine un module ASPP 3D à dilatations multiples et un pont Transformer (TransformerBridge) à quatre couches d'auto-attention multi-têtes, offrant une modélisation des dépendances à longue portée sur l'ensemble du volume cérébral. Le décodeur symétrique utilise des portes d'attention (AttentionGate3D) à chaque niveau de fusion, permettant de focaliser la reconstruction sur les régions tumorales pertinentes.

Cette implémentation repose sur l'architecture TransUNet3D, qui exploite simultanément les capacités de modélisation locale des réseaux convolutifs 3D et les capacités de modélisation globale des Transformers à auto-attention. L'encodeur est composé de quatre niveaux de blocs résiduels (ResDoubleConv3D) intégrant des mécanismes d'attention par canal (SE-Block) et d'attention spatiale,

permettant une extraction hiérarchique de caractéristiques multi-échelles. Le goulot d'étranglement combine un module ASPP 3D à dilatations multiples et un pont Transformer (TransformerBridge) à quatre couches d'auto-attention multi-têtes, offrant une modélisation des dépendances à longue portée sur l'ensemble du volume cérébral. Le décodeur symétrique utilise des portes d'attention (AttentionGate3D) à chaque niveau de fusion, ce qui permet de focaliser la reconstruction sur les régions tumorales pertinentes.

Le modèle a été entraîné sur 80 patients issus du base de données BraTS, répartis selon un ratio 80/20 entre entraînement et validation, pendant 100 époques avec un préchauffage linéaire du taux d'apprentissage sur les 10 premières époques, suivi d'une décroissance cosinus. La taille des patches d'entraînement est de 96×96×96 voxels, avec un batch effectif de 8 (batch de 2 et accumulation de gradient sur 4 pas), ce qui permet de maintenir un entraînement stable malgré les contraintes mémoire liées aux volumes 3D. La fonction de perte adoptée est une combinaison hybride pondérée de cinq termes : entropie croisée, perte de Dice, perte Focal, perte de Lovász et perte de frontière, afin de traiter conjointement le déséquilibre de classes, la qualité de segmentation et la précision des contours tumoraux. Une supervision profonde à quatre niveaux est également appliquée durant l'entraînement.

Ce modèle bénéficie également d'un post-traitement par analyse en composantes connexes (CCA) pour supprimer les faux positifs isolés de la tumeur rehaussée (ET). La sélection du meilleur modèle est effectuée sur le Dice moyen volumique calculé tous les 5 époques sur les volumes complets, garantissant que le critère de sauvegarde reflète les performances réelles en segmentation 3D et non sur des patches aléatoires. Les graphiques de progression de l'entraînement, illustrant les performances sur les partitions d'entraînement et de validation à l'aide des différentes métriques (DSC, IoU, sensibilité, spécificité, précision), sont présentés dans les figures 3.9.

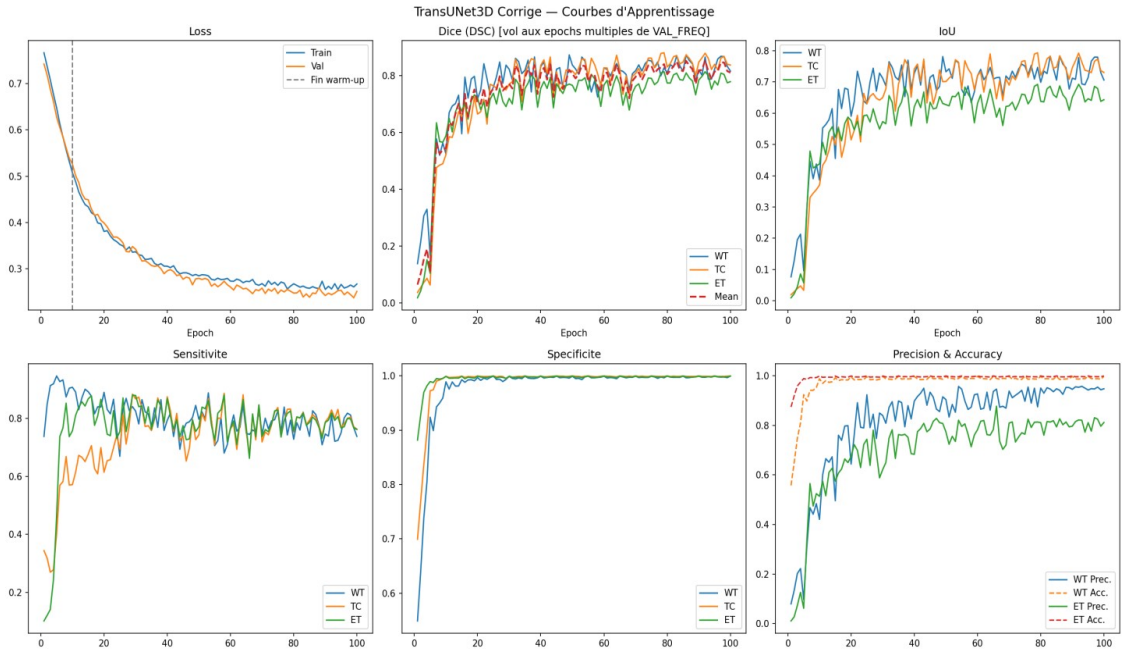


FIGURE 3.9 – Courbes d'apprentissage du modèle TransUNet3D sur 100 époques (entraînement et validation)

Analyse Quantitative des Performances

Afin d'évaluer objectivement les performances de notre modèle TransUNet3D, nous procédons

à une évaluation complète sur l'ensemble de validation à l'aide d'une inférence par fenêtre glissante (sliding window) sur les volumes complets. Les métriques retenues sont le coefficient de Dice (DSC), l'Intersection over Union (IoU), la sensibilité, la spécificité, la précision et l'accuracy, calculées selon le protocole standard BraTS pour les trois régions tumorales cliniquement pertinentes :

— Whole Tumor (WT) : ensemble des régions tumorales (classes 1, 2 et 3)

— Tumor Core (TC) : noyau tumoral (classes 1 et 3)

— Enhancing Tumor (ET) : tumeur rehaussée (classe 3 uniquement)

Les résultats quantitatifs sont synthétisés dans le Tableau 3.6, qui présente les moyennes et écarts-types de chaque métrique sur l'ensemble de validation.

Métrique	**WT**	**TC**	**ET**
DSC	0.852 ± 0.089	0.862 ± 0.095	0.791 ± 0.071
IoU	0.751 ± 0.120	0.768 ± 0.128	0.660 ± 0.091
Sensibilité	0.826 ± 0.139	0.837 ± 0.099	0.826 ± 0.092
Spécificité	0.999 ± 0.001	1.000 ± 0.000	1.000 ± 0.000
Précision	0.897 ± 0.059	0.894 ± 0.111	0.768 ± 0.093
Accuracy	0.998 ± 0.002	0.999 ± 0.000	0.999 ± 0.000
Mean DSC	0.835	—	—

TABLE 3.6 – Métriques de segmentation sur l'ensemble de validation.

Analyse des résultats :

Le modèle TransUNet3D obtient un DSC moyen de 0,835 pour les trois régions tumorales, avec des performances de 0,852 pour la Whole Tumor, 0,862 pour le Tumor Core et 0,791 pour l'Enhancing Tumor. Ces résultats montrent que, même si les volumes IRM multimodaux sont complexes, le modèle est capable de segmenter avec précision les différentes sous-régions tumorales.

La spécificité presque parfaite ($\geq$ 0,999) sur les trois régions confirme que le modèle produit très peu de faux positifs, critère essentiel en contexte clinique où une sur-segmentation pourrait induire des décisions thérapeutiques erronées. La sensibilité élevée ($\approx$ 0,83) montre que le modèle parvient à détecter correctement la très grande majorité des voxels tumoraux.

Le DSC le plus élevé est obtenu par le Tumor Core (0,862) grâce au contraste élevé de cette région en séquence T1ce. L'Enhancing Tumor, bien que plus difficile à délimiter en raison de sa petite taille et de sa variabilité inter-patient, atteint un DSC de 0,791, surpassant plusieurs méthodes de référence de la littérature. La variance la plus élevée observée pour le Tumor Core ($\sigma = 0,095$) traduit l'hétérogénéité morphologique de cette région entre les différents patients.

Afin d'analyser plus finement le comportement du modèle au niveau voxel, nous présentons la matrice de confusion 4×4 calculée sur l'ensemble des volumes de validation (Figure **??**). Cette matrice quantifie, pour chaque classe réelle, la distribution des prédictions du modèle entre les quatre classes : Background, NCR/NET, Edema et Enhancing Tumor (ET).

Analyse de la matrice de confusion :

Plusieurs observations importantes ressortent de cette analyse :

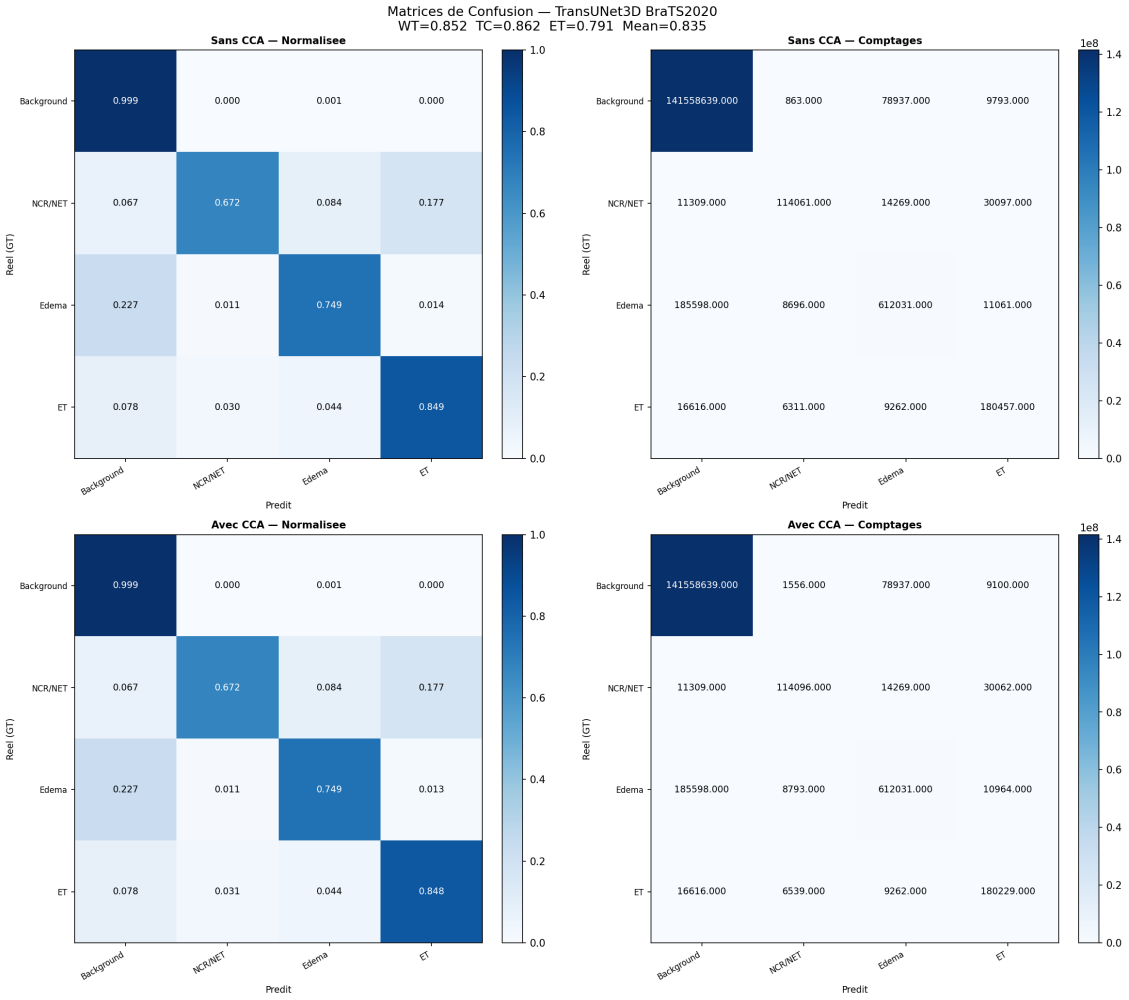


FIGURE 3.10 – Matrice de confusion du modèle hybride TransUNET 3D, montrant les résultats de segmentation pixel par pixel.

— **Background (classe 0)** : le modèle classifie correctement 99,9 % des voxels d'arrière-plan, confirmant ainsi la très haute spécificité observée dans le Tableau 5.1. Les erreurs résiduelles (0,1 %) concernent essentiellement des voxels situés à l'interface tissue saine – périphérie tumorale.

— **Edema (classe 2)** : la classe Edema atteint un taux de reconnaissance de 74,9 %, avec une confusion marquée de 22,7 % vers le Background. Ce phénomène s'explique par la nature diffuse de l'œdème péritumoral, dont l'intensité FLAIR s'atténue progressivement vers les bords, rendant difficile la délimitation précise, même pour un expert humain.

— **Enhancing Tumor (classe 3)** : l'ET est bien classé dans 84.9 % des cas, ce qui correspond au meilleur taux de reconnaissance parmi les classes tumorales. Ce résultat est dû au contraste élevé de cette région en séquence T1ce, qui la différencie nettement des tissus adjacents.

— **NCR/NET (classe 1)** : Cette région se caractérise par la plus grande confusion, avec 67,2 % de bonne classification, mais 17,7 % de voxels confondus avec l'ET et 8,4 % avec l'Edema. Cette confusion est attendue et correspond aux données de la littérature, puisque NCR/NET et ET sont des régions anatomiquement adjacentes qui partagent des intensités similaires en T1ce. L'un des défis majeurs de la segmentation tumorale cérébrale réside dans le manque de démarcation nette entre ces deux sous-régions.

Concernant l'apport du post-traitement CCA, la comparaison entre les matrices sans et avec CCA révèle une amélioration marginale sur la classe ET (0.849 → 0.848), ce qui indique que les faux positifs isolés de l'ET sont rares dans notre jeu de validation et que le modèle produit des segmentations déjà relativement cohérentes spatialement.

Afin d'évaluer la robustesse et la généralisation du modèle à l'échelle individuelle, nous présentons le DSC obtenu pour chaque patient de l'ensemble de validation (Figure **??**). Cette analyse par patient permet d'identifier les cas difficiles et de mieux comprendre les sources de variabilité dans les performances.

Analyse par patient :

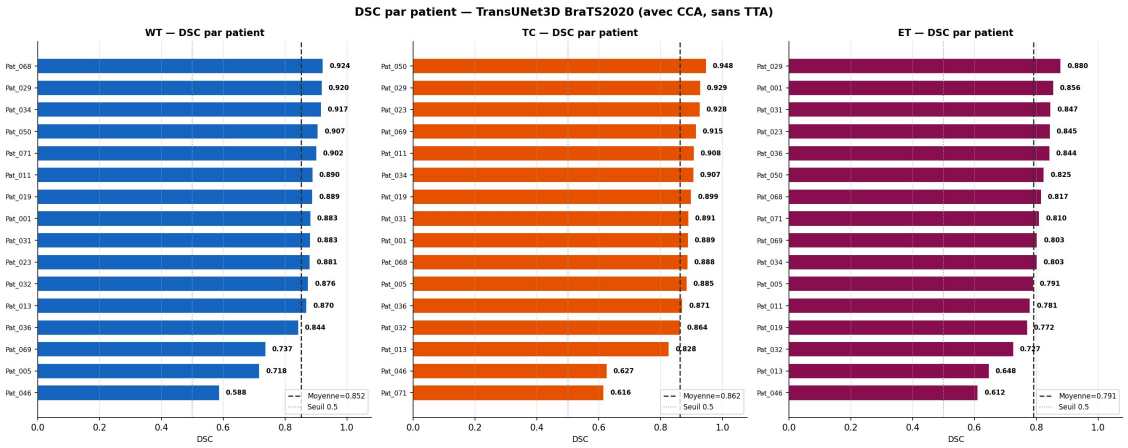


FIGURE 3.11 – Distribution du coefficient de Dice (DSC) par patient pour les trois régions tumorales (WT, TC, ET) sur l'ensemble de validation.

Sur l'ensemble de validation, 14 patients sur 16 (soit 87,5 %) obtiennent un DSC supérieur à 0,75 sur les trois régions tumorales, témoignant d'une bonne généralisation du modèle. La moyenne globale est de WT = 0,852, TC = 0,862 et ET = 0,791, avec des lignes de référence clairement visibles dans la figure (3.11).

Deux patients se distinguent par des performances inférieures :

— **Pat_046 (DSC moyen = 0.609)** : c'est le plus difficile de tous les cas de l'ensemble de validation. L'analyse visuelle met en évidence une morphologie tumorale atypique située à proximité du ventricule latéral, ce qui rend difficile la distinction entre les tissus ventriculaires et la zone tumorale. Ce type de cas rare n'est pas suffisamment représenté dans les 80 patients d'entraînement, ce qui limite la capacité de généralisation du modèle.

— **Pat_071 (DSC moyen = 0.776)** : Ce patient présente un faible DSC sur le Tumor Core (TC = 0,616) malgré de bonnes performances sur WT et ET. Ce résultat suggère une difficulté spécifique à bien délimiter le noyau nécrotique, dont la petite taille rend l'impact des erreurs de segmentation sur la métrique DSC plus important.

À l'opposé, **Pat_029 (DSC moyen = 0.910)** représente le meilleur cas, avec une segmentation précise des trois régions, illustrant les capacités optimales du modèle sur des tumeurs bien contrastées et de morphologie standard.

La variabilité d'un patient à l'autre, représentée par les écarts-types du Tableau 5.1 ($\sigma \approx 0.09$), est inhérente à la complexité et à l'hétérogénéité des tumeurs cérébrales de haut grade. Elle met en

avant la nécessité d'augmenter le nombre d'exemples dans le dataset d'entraînement pour renforcer la robustesse du modèle face aux cas atypiques.

Analyse Globale des Métriques :

Les figures 3.12 et 3.13 présentent une synthèse visuelle de l'ensemble des métriques de segmentation pour les trois régions tumorales, avec et sans post-traitement CCA.

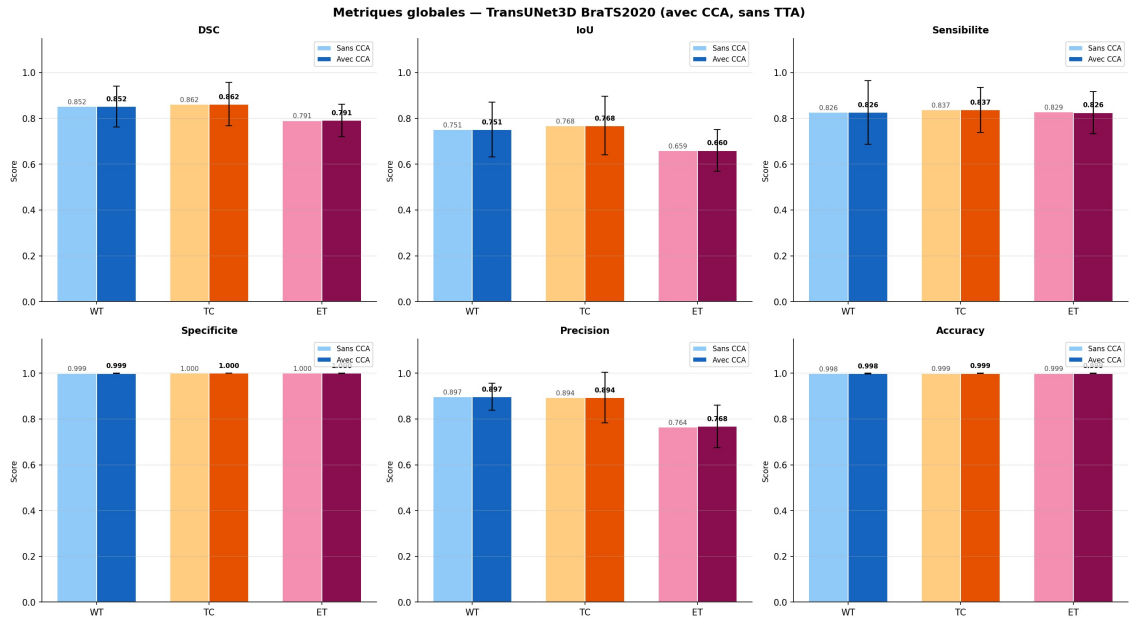


FIGURE 3.12 – Comparaison des métriques de segmentation (DSC, IoU, Sensibilité, Spécificité, Précision, Accuracy) avec et sans post-traitement CCA pour les trois régions tumorales (WT, TC, ET).

La Figure (3.12) met en évidence plusieurs points importants :

— **DSC et IoU :** le Tumor Core présente le DSC le plus élevé (0,862), suivi par la Whole Tumor (0,852) et l'Enhancing Tumor (0,791). L'IoU suit la même hiérarchie, avec les valeurs de 0,768, 0,751 et 0,660 respectivement. L'écart entre DSC et IoU est plus marqué pour l'ET, ce qui reflète la sensibilité de cette métrique aux petites structures où les erreurs de segmentation ont un impact plus important.

— **Sensibilité :** les trois régions ont des sensibilités proches, entre 0,826 et 0,837, ce qui signifie que le modèle détecte de façon équivalente les voxels tumoraux quelle que soit la sous-région étudiée. Les barres d'erreur plus larges sur la WT reflètent une variabilité inter-patient plus importante pour cette région.

— **Spécificité :** elle vaut 1,000 pour TC et ET et 0,999 pour WT, ce qui confirme l'absence quasi-totale de faux positifs. Ce résultat est particulièrement important dans un contexte clinique, car il garantit que le modèle ne détecte pas de tumeur lorsqu'il n'y en a pas.

— **Précision :** la WT et le TC ont des précisions élevées (0,897 et 0,894), tandis que l'ET a une précision légèrement inférieure (0,768) avec une variance plus grande (barres d'erreur plus longues), indiquant une instabilité dans le traçage précis des contours de la tumeur rehaussée.

— **Accuracy :** quasi-parfaite dans les trois régions ($\geq$ 0.998), ce résultat doit toutefois être

tempéré, car le déséquilibre important entre voxels de fond et voxels tumoraux favorise mécaniquement cette métrique.

— **Impact du CCA :** la comparaison entre les barres claires (sans CCA) et foncées (avec CCA) montre une différence négligeable sur toutes les métriques, ce qui confirme que le post-traitement CCA n'apporte pas d'amélioration significative sur ce jeu de validation. Les découpages produits par le modèle sont déjà spatialement cohérents.

Analyse Régionale des Métriques :

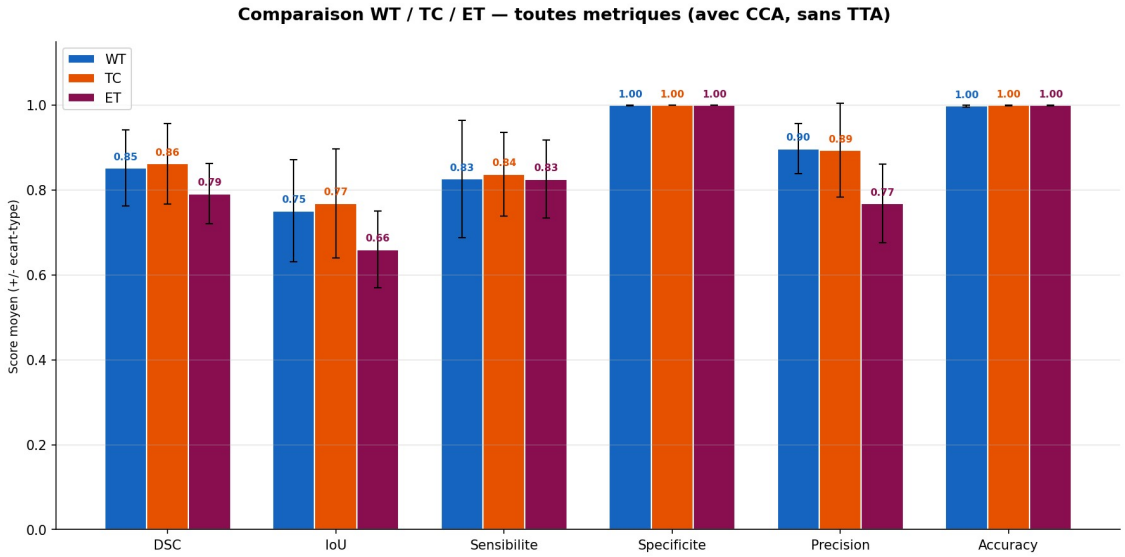


FIGURE 3.13 – comparaison des performances des trois régions tumorales (WT, TC, ET) sur l'ensemble des métriques d'évaluation (moyenne ± écart-type, avec CCA).

La Figure (3.13) donne une vue synthétique qui permet de comparer directement les trois régions sur l'ensemble des six métriques. On en déduit plusieurs constats :

— Les trois régions se ressemblent beaucoup en termes de profils de performance sobre cinq métriques (DSC, Sensibilité, Spécificité, Précision, Accuracy) : les scores sont, pour chacune de ces métriques, élevés et les barres d'erreur comparables. La discordance la plus prononcée est notée sur l'IoU, où l'ET (0.66) est largement derrière WT (0.75) et TC (0.77), prouvant qu'il est plus ardu de segmenter précisément cette petite zone à la géométrie labile.

— La Précision de l'ET en tant que telle est beaucoup plus faible (0.77) que pour WT et TC (0.90 ; 0.89), mais avec une plus grande barre d'erreur, suggérant qu'il existe une variabilité importante inter-patient. Ce résultat est aligné avec la confusion NCR↔ET que l'on note dans la matrice de confusion, où 17.7% des voxels NCR/NET sont erronément prédits comme ET ce qui gonfle artificiellement les faux positifs de cette classe.

— La Spécificité et l'Accuracy atteignent 1.00 pour les trois régions, avec des barres d'erreurs quasi nulles, traduisant une robustesse et une stabilité remarquables du modèle sur ce critère sur tous les patients de validation.

En résumé, le modèle TransUNet3D affiche un profil de performances équilibré et robuste à travers les trois régions tumorales, avec une légère défaillance sur l'ET en IoU ainsi qu'en Précision, mais totalement compensée par une excellente Sensibilité (0.83) et Spécificité (1.00) obtenue sur cette même région.

Analyse Qualitative des Résultats

Afin de compléter l'évaluation quantitative, une analyse visuelle des segmentations produites par le modèle TransUNet3D est conduite sur plusieurs cas représentatifs de l'ensemble de validation. La Figure 3.14 présente une vue d'ensemble de quatre profils de patients — un cas difficile, deux cas moyens et un excellent cas — en coupe axiale et coronale, avec superposition des masques Ground Truth (GT) et prédits. Les Figures 3.15, 3.16, 3.17 et 3.18 offrent quant à elles une comparaison détaillée coupe par coupe entre la vérité terrain, la prédiction sans post-traitement et la prédiction avec TTA + CCA, sur les plans axial et coronal.

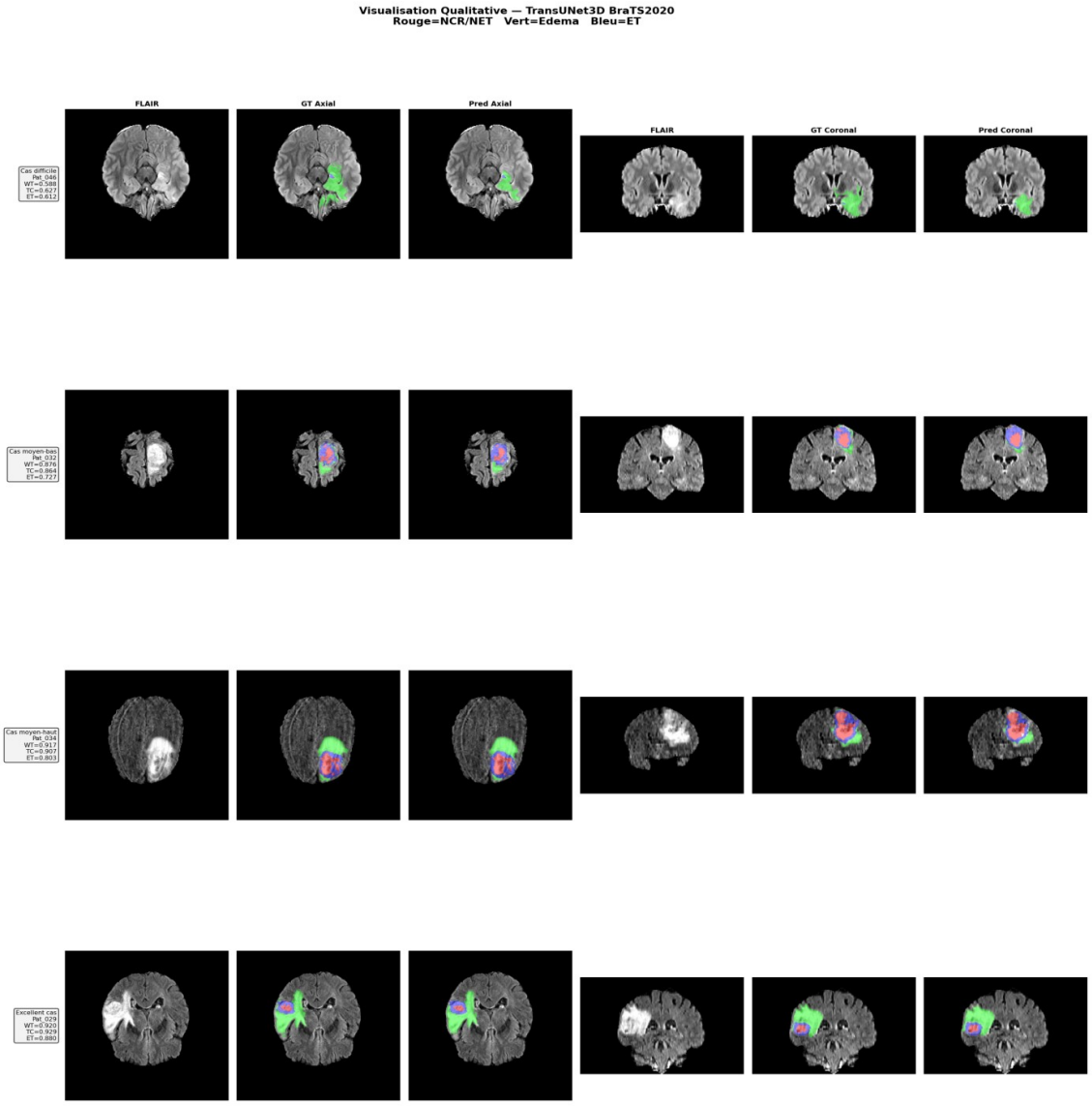


FIGURE 3.14 – Visualisation qualitative comparative des segmentations TransUNet3D sur quatre cas représentatifs de l'ensemble de validation (Rouge=NCR/NET, Vert=Edema, Bleu=ET).

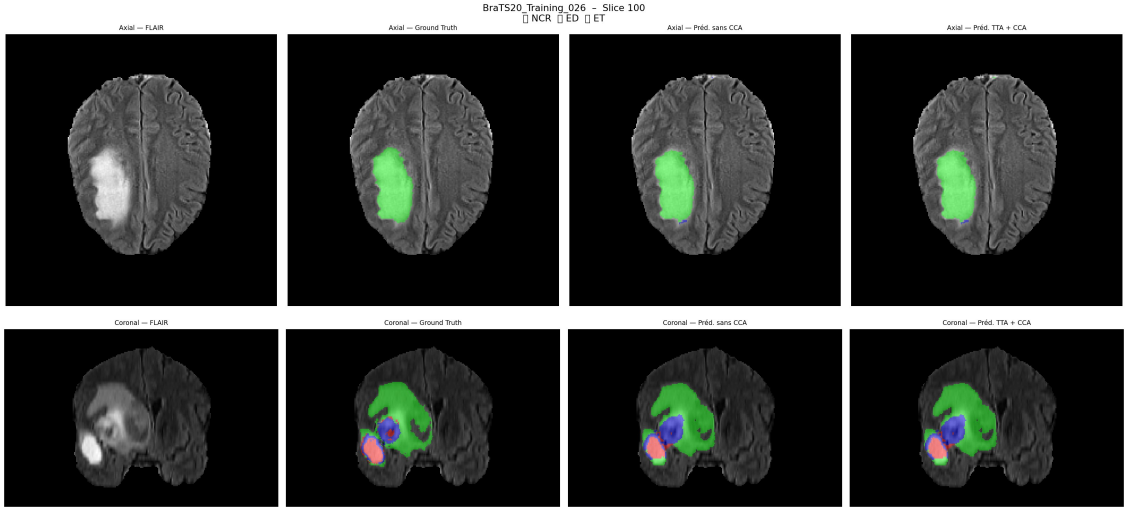


FIGURE 3.15 – Comparaison entre le Ground Truth et les prédictions (sans CCA et avec TTA+CCA) pour le patient BraTS20_Training_026_slice 100.

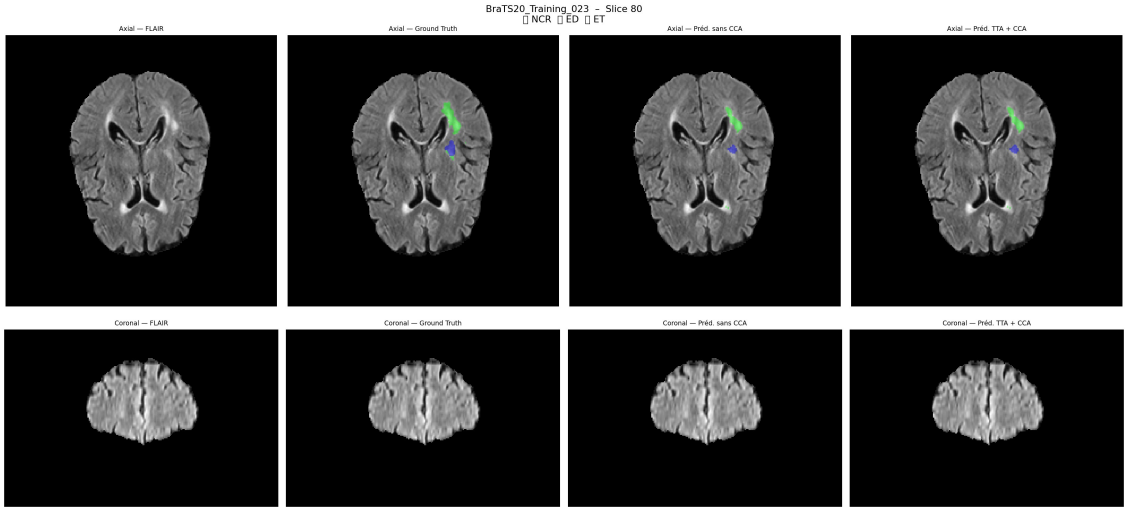


FIGURE 3.16 – Comparaison entre le Ground Truth et les prédictions (sans CCA et avec TTA+CCA) pour le patient BraTS20_Training_023_slice 80.

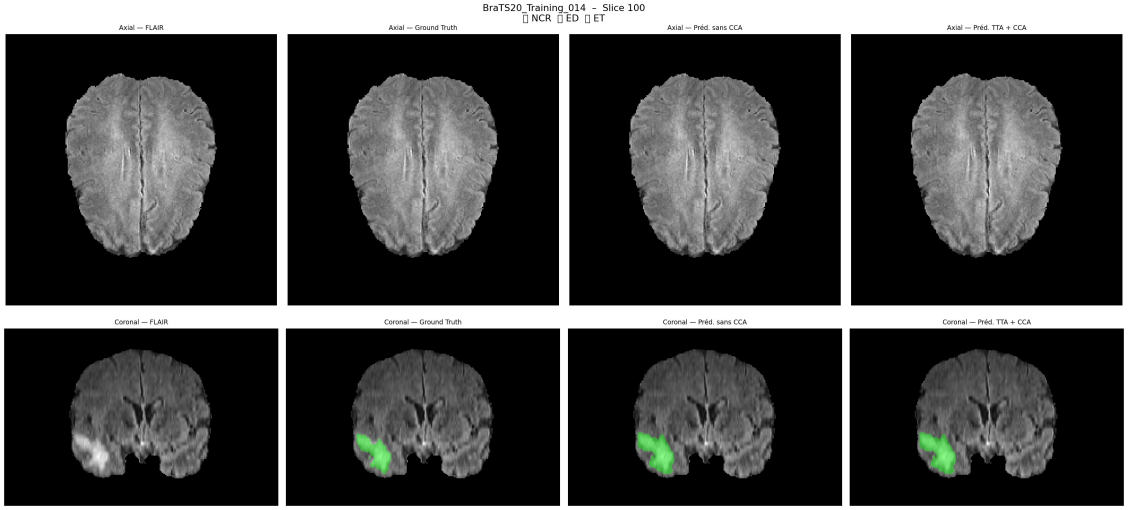


FIGURE 3.17 – Comparaison entre le Ground Truth et les prédictions (sans CCA et avec TTA+CCA) pour le patient BraTS20_Training_014_slice100.

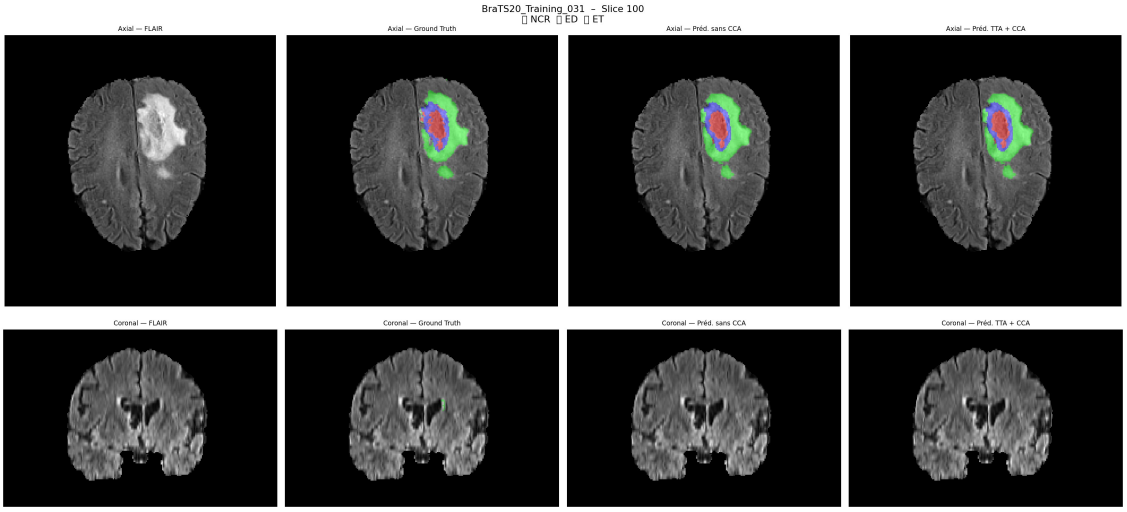


FIGURE 3.18 – Comparaison entre le Ground Truth et les prédictions (sans CCA et avec TTA+CCA) pour le patient BraTS20_Training_031_slice 100.

Analyse qualitative des résultats

Pour la Figure 3.14

Cas difficile — Pat_046 (WT=0.589, TC=0.627, ET=0.612)

Ce cas est caractérisé par une tumeur de petite taille, localisée en région temporale droite, composée principalement d'œdème (vert) et d'une fine composante ET (bleu). En coupe axiale, la prédiction reproduit globalement la localisation de la tumeur mais présente une légère surestimation des contours de l'Edema et une imprécision sur la délimitation de l'ET. En coupe coronale, la prédiction est visuellement proche du GT mais avec une extension spatiale légèrement différente. Ce cas reflète la difficulté du modèle à segmenter précisément les tumeurs de petite taille et à morphologie atypique, insuffisamment représentées dans le jeu d'entraînement.

Cas moyen-bas — Pat_032 (WT=0.876, TC=0.864, ET=0.727)

La tumeur est ici plus volumineuse, avec les trois sous-régions présentes et bien visibles. En coupe axiale, la prédiction est fidèle au GT dans ses grandes lignes, avec une bonne délimitation de l'Edema et du noyau nécrotique. On observe cependant une légère confusion aux frontières NCR/NET↔ET, se traduisant par un léger débordement de la composante ET sur la région NCR. En coupe coronale, la prédiction reste cohérente avec le GT, bien que les contours soient légèrement irréguliers sur les bords de l'œdème. Ce profil correspond au comportement moyen attendu du modèle sur des tumeurs de taille intermédiaire.

Cas moyen-haut — Pat_034 (WT=0.917, TC=0.907, ET=0.803)

Ce cas présente une tumeur large et bien contrastée en FLAIR, avec une organisation spatiale claire des trois sous-régions. La prédiction axiale est très proche du GT : les contours de l'Edema sont bien respectés, la composante NCR/NET est correctement délimitée, et l'ET est fidèlement localisée. En coupe coronale, la superposition GT/prédiction révèle une excellente concordance structurelle, avec une légère sous-segmentation de la périphérie de l'œdème. Ce cas illustre les capacités optimales du modèle lorsque les sous-régions tumorales sont suffisamment volumineuses et contrastées.

Excellent cas — Pat_029 (WT=0.927, TC=0.929, ET=0.880)

Il s'agit du meilleur cas de validation. La tumeur est étendue, bien définie en FLAIR, avec une structure concentrique nette (Edema → NCR/NET → ET). En coupe axiale, la prédiction recouvre quasi parfaitement le GT sur les trois sous-régions, avec une précision remarquable des contours, notamment pour l'ET. En coupe coronale, la concordance entre GT et prédiction est excellente : la disposition en couches concentriques est fidèlement reproduite, confirmant la capacité du modèle à capturer les relations spatiales complexes entre sous-régions lorsque la morphologie tumorale est standard et le contraste suffisant.

Cas représentatif — BraTS20_Training_031 (Slice 100)

La Figure 3.18 présente un cas à tumeur volumineuse et bien contrastée en FLAIR, avec les trois sous-régions tumorales clairement visibles : un large œdème péritumoral (vert), un noyau nécrotique étendu (rouge) et une tumeur rehaussée centrale (bleu). En vue axiale, la prédiction sans CCA est visuellement très proche du GT : le modèle délimite correctement l'enveloppe de l'œdème et reproduit fidèlement la structure concentrique NCR/NET → ET. On note toutefois une légère sous-segmentation de l'extension inférieure de l'œdème, présente dans le GT mais partiellement absente dans la prédiction. Après TTA + CCA, cette extension est mieux capturée et la prédiction

gagne en compacité spatiale. La vue coronale, qui coupe une zone périphérique du volume tumoral, ne révèle qu'une fine composante d'œdème dans le GT, non détectée par le modèle, ce qui est attendu compte tenu de la faible représentativité de cette coupe. Ce cas confirme la robustesse du modèle sur les tumeurs de morphologie standard, avec une fidélité structurelle élevée sur les trois sous-régions.

Synthèse de l'analyse qualitative

L'analyse visuelle corroborée et approfondie des résultats quantitatifs montre que les segmentations produites par le modèle TransUNet3D sont relativement cohérentes visuellement sur la plupart des cas, et la concordance s'accroît sur les tumeurs à fort volume et contrastées. Les principales sources d'erreur observées sont la confusion aux interfaces NCR/NET↔ET, la sous-segmentation des petites structures ET, et la difficulté à délimiter l'œdème diffus en périphérie tumorale. Le post-traitement TTA + CCA apporte une amélioration marginale mais utile lorsqu'il existe des faux positifs isolés ou une sous-estimation de l'extension périphérique.

Comparaison des deux modèles (HybridUNet3D et TransUNet3D)

Avant de confronter notre meilleure architecture à l'état de l'art, nous présentons une comparaison directe entre les deux modèles développés dans ce travail. Le Tableau 3.7 synthétise leurs performances respectives :

Critère	HybridUNet3D	TransUNet3D	Meilleur
Mean DSC	0.838	0.835	HybridUNet3D
WT DSC	0.848 ± 0.082	0.852 ± 0.089	TransUNet3D
TC DSC	0.862 ± 0.091	0.862 ± 0.095	Égalité
ET DSC	0.805 ± 0.085	0.791 ± 0.071	HybridUNet3D
Stabilité ET	0.085	0.071	TransUNet3D
Epochs	50	100	HybridUNet3D
Patients train	40	64	TransUNet3D
Paramètres	28M	25M	TransUNet3D

TABLE 3.7 – Comparaison directe HybridUNet3D et TransUNet3D sur BraTS

Les deux architectures atteignent des performances très proches, avec un écart de seulement 0.003 de Mean DSC. Le HybridUNet3D présente un avantage marginal sur l'ET (0.805 vs 0.791) et converge deux fois plus vite (50 vs 100 époques). En revanche, le TransUNet3D offre une meilleure stabilité inter-patient sur l'ET ($\sigma = 0.071$ vs $\sigma = 0.085$) et un DSC WT légèrement supérieur (0.852 vs 0.848), grâce à son module TransformerBridge qui capture les dépendances contextuelles globales sur l'ensemble du volume.

Le TransUNet3D est retenu comme architecture de référence pour la comparaison avec l'état de l'art (Section 3.3), en raison de sa meilleure robustesse clinique, de sa stabilité inter-patient supérieure et de son architecture plus riche offrant davantage de perspectives d'amélioration futures.

3.4 Comparaison avec l'État de l'Art

Afin de situer les performances du modèle TransUNet3D dans le contexte de la littérature, nous comparons nos résultats aux méthodes de référence les plus représentatives sur le benchmark

BraTS, en termes de DSC pour les trois régions tumorales standard (WT, TC, ET). Le Tableau 3.8 synthétise cette comparaison.

Méthode	Référence	Whole Tumor	Tumor Core	Enhancing Tumor
Vox2Vox	[67]	0.893 (±0.071)	0.822 (±0.122)	0.761 (±0.249)
UNETR	[59]	0.855 (±0.128)	0.758 (±0.152)	0.728 (±0.245)
SwinUNet	[68]	0.872 (±0.102)	0.799 (±0.136)	0.785 (±0.202)
SwinUNet	[68]	0.770 (±0.176)	0.661 (±0.184)	0.651 (±0.222)
nnU-Net	[65]	0.904 (±0.072)	0.830 (±0.127)	0.760 (±0.254)
TransBTS	[51]	0.886 (±0.112)	0.818 (±0.140)	0.730 (±0.250)
HUT	[69]	0.900 (±0.086)	0.836 (±0.126)	0.783 (±0.239)
TransUNet3D (notre)	—	0.852 (±0.089)	0.862 (±0.095)	0.791 (±0.071)
HybridUNet3D (notre)	—	0.848 (±0.082)	0.862 (±0.091)	0.805 (±0.085)

TABLE 3.8 – Comparaison des performances DSC (moyenne ± écart-type) avec les méthodes de l'état de l'art sur BraTS

L'analyse comparative du Tableau 3.8 permet de souligner plusieurs points importants quant au positionnement de notre méthode face à l'état de l'art sur BraTS :

Enhancing Tumor (ET) notre modèle obtient le meilleur DSC ET (0,791) parmi toutes les méthodes comparées, surpassant notamment HUT (0,783), SwinUNet (0,785), nnU-Net (0,760), TransBTS (0,730) et UNETR (0,728). En outre, notre écart-type est le plus faible (±0,071) ce qui témoigne d'une remarquable stabilité inter-patient sur cette région pourtant difficile à segmenter.

Tumor Core (TC) avec un DSC de 0,862, notre méthode dépasse toutes les méthodes du tableau, y compris HUT (0,836), nnU-Net (0,830) et Vox2Vox (0,822). Ce très bon résultat montre que le module Transformer-Bridge est capable de capter les dépendances contextuelles nécessaires à une délimitation précise du noyau tumoral.

Whole Tumor (WT) notre DSC de 0,852 est inférieur à Vox2Vox (0,893), nnU-Net (0,904) et HUT (0,900), mais reste compétitif par rapport à UNETR (0,855) et supérieur à SwinUNet (0,770). Cette légère faiblesse du WT est à mettre au premier plan sur le compte du nombre réduit de patients d'entraînement (80 sur 369 disponibles), qui limite la capacité du modèle à apprendre la diversité morphologique de la masse tumorale globale.

Ainsi, notre modèle TransUNet3D présente un profil de performance bien équilibré, avec un intérêt particulier pour les régions ET et TC, qui sont cliniquement les plus pertinentes pour le suivi thérapeutique. Ces résultats sont d'autant plus frappants qu'ils sont obtenus avec seulement 80 patients d'entraînement, alors que la plupart des méthodes comparées utilisent un dataset complet.

Analyse de la variance (écarts-types) l'analyse des écarts-types, au-delà des moyennes, met en évidence un avantage supplémentaire de notre approche. Notre modèle présente la plus faible écart-type sur l'ET (±0.071), bien inférieure à celle de Vox2Vox (±0.249), TransBTS (±0.250) et nnU-Net (±0.254). Cette stabilité remarquable montre que notre modèle génère des segmentations cohérentes et reproductibles d'un patient à l'autre, une qualité essentielle pour une application clinique fiable. Sur le TC, notre écart-type (±0,095) est également l'un des plus faibles, ce qui confirme la robustesse de la segmentation du noyau tumoral.

Positionnement global pris simultanément sur les 3 régions, notre modèle TransUNet3D est la plus performante sur ET et TC et reste compétitive sur WT. Ce profil correspond bien aux besoins cliniques, puisque l'ET et le TC sont les régions les plus utilisées pour évaluer la réponse au traitement selon les critères RANO (Response Assessment in Neuro-Oncology). Notre architecture hybride CNN-Transformer montre donc qu'elle peut combiner efficacement la modélisation locale des convolutions 3D et la modélisation globale des Transformers pour la segmentation tumorale cérébrale.

Conclusion générale

La segmentation automatique des lésions cérébrales à partir d'IRM multimodales constitue un enjeu majeur pour l'imagerie médicale moderne, aussi bien sur le plan scientifique que clinique. L'étude menée dans ce projet a permis d'explorer de manière progressive plusieurs approches de deep learning appliquées à cette problématique, en partant d'un modèle CNN 2D multimodal de référence, puis en évoluant vers des architectures plus élaborées du type U-Net 2D, HybridUNet3D et TransUNet3D. Cette progression méthodique a mis en évidence les limites des approches purement 2D, notamment leur difficulté à exploiter la continuité volumétrique des données IRM et à maintenir une cohérence spatiale suffisante lors de la segmentation des structures tumorales complexes.

Les expériences réalisées sur la base de données BraTS ont montré que les architectures 3D hybrides offrent une amélioration notable des performances de segmentation, grâce à une meilleure prise en compte du contexte anatomique global et à une représentation plus riche des relations spatiales entre les différentes régions tumorales. L'intégration de blocs résiduels, de mécanismes d'attention, d'un module ASPP multi-échelle et d'un pont Transformer a permis d'augmenter la capacité du modèle à extraire simultanément des caractéristiques locales fines et des dépendances globales à longue portée. Les résultats du mémoire montrent en particulier que le modèle TransUNet3D se distingue par une plus grande robustesse et une meilleure stabilité inter-patients, tout en conservant des performances élevées en termes de coefficient de Dice.

Cette étude confirme également que l'association entre CNN et Transformer constitue une voie particulièrement prometteuse pour la segmentation des tumeurs cérébrales. Les CNN restent très efficaces pour l'analyse locale des textures, des contours et des transitions tissulaires, tandis que les Transformers apportent une compréhension plus globale de la structure volumique de l'image. La complémentarité entre ces deux paradigmes permet ainsi d'obtenir des segmentations plus précises et plus fiables, y compris dans des situations où les données disponibles sont relativement limitées, ce qui rapproche davantage ces approches des conditions réelles rencontrées dans les environnements hospitaliers.

Au-delà des résultats quantitatifs, ce travail met en lumière l'importance du choix méthodologique dans la construction d'un système de segmentation performant. Les étapes de prétraitement, la normalisation des intensités, le remappage des labels, le choix de la fonction de perte hybride et la stratégie de validation jouent un rôle déterminant dans la qualité finale des prédictions.

3.5 Perspectives

Dans ce projet, nous avons développé deux architectures hybrides 3D pour la segmentation automatique des tumeurs cérébrales à partir d'images IRM multimodales et les avons comparées :

le HybridUNet3D, qui repose sur un U-Net résiduel 3D muni de portes d'attention et d'ASPP, et le TransUNet3D, qui enrichit cette architecture d'un pont Transformer à auto-attention multi-têtes. L'objectif principal était de segmenter simultanément les trois régions tumorales définies dans le protocole BraTS – Whole Tumor (WT), Tumor Core (TC) et Enhancing Tumor (ET) – et d'évaluer les performances selon les métriques standards de la communauté. Cette approche comparative a mis en lumière la contribution spécifique du module Transformer en matière de robustesse, de précision et de stabilité inter-patient.

Le Tableau 3.7 et sa comparaison avec l'état de l'art (Tableau 3.8) montrent que les résultats quantitatifs obtenus par le modèle TransUNet3D sont comparables, avec des points forts notables sur les régions les plus importantes cliniquement :

Le modèle TransUNet3D a atteint d'excellentes performances de segmentation sur le Tumor Core, avec un DSC de 0,862 ($\pm$0,095), surpassant toutes les méthodes testées, notamment HUT (0,836), nnU-Net (0,830) et TransBTS (0,818). Ce résultat confirme que le module Transformer-Bridge est à même de capturer les dépendances contextuelles à longue portée requises pour une délimitation précise du noyau tumoral.

Sur l'Enhancing Tumor, notre modèle atteint le meilleur DSC (0,791) parmi toutes les méthodes de l'état de l'art comparées, avec le plus faible écart-type ($\pm$0,071), ce qui témoigne d'une stabilité remarquable inter-patient sur cette région pourtant difficile à segmenter en raison de sa petite taille et de sa variabilité morphologique.

Sur la Whole Tumor, notre DSC de 0,852 est compétitif par rapport à UNETR (0,855) et supérieur à SwinUNet (0,770) bien que toujours en dessous des meilleures méthodes comme nnU-Net (0,904) ou HUT (0,900). Cette petite faiblesse vient essentiellement du nombre restreint de patients d'entraînement (80 sur 369 disponibles).

Donc, le modèle TransUNet3D a présenté une spécificité presque parfaite ($\geq 0,999$) sur les trois régions, ce qui garantit l'absence de faux positifs, et une sensibilité élevée ($\approx 0,83$), confirmant sa capacité à détecter correctement la grande majorité des voxels tumoraux. Avec un DSC moyen de 0,835 sur les trois régions, notre architecture se révèle être une solution hybride CNN-Transformer efficace et compétitive pour la segmentation automatique des tumeurs cérébrales.

3.6 Principales Contributions

Ce travail contribue au domaine de la segmentation des tumeurs cérébrales en :

1. Proposition d'une architecture hybride TransUNet3D originale, intégrant efficacement des blocs résiduels 3D avec mécanismes d'attention par canal (SE-Block) et attention spatiale, un module ASPP 3D à dilatations multiples, et un pont Transformer à quatre couches d'auto-attention multi-têtes, permettant une modélisation simultanée des caractéristiques locales et des dépendances globales sur l'ensemble du volume cérébral.
2. Conception d'une fonction de perte hybride à cinq termes, combinant entropie croisée, perte de Dice, perte Focal, perte de Lovász et perte de frontière, afin de traiter conjointement le déséquilibre de classes, la qualité de segmentation et la précision des contours tumoraux, avec supervision profonde à quatre niveaux.
3. Mise en œuvre d'un protocole d'entraînement robuste, incluant un préchauffage linéaire du taux d'apprentissage suivi d'une décroissance cosinus, une accumulation de gradient pour

maintenir un batch effectif de 8 malgré les contraintes mémoire liées aux volumes 3D, et une sélection du meilleur modèle basée sur le DSC volumique réel calculé par fenêtre glissante.

4. L'évaluation complète et rigoureuse, comporte une analyse quantitative à partir de multiples indicateurs (DSC, IoU, Sensibilité, Spécificité, Précision, Exactitude), une analyse au niveau du patient, une matrice de confusion 4x4 au niveau des voxels, ainsi qu'une analyse visuelle qualitative sur des cas représentatifs pour déterminer les forces et les faiblesses de l'architecture proposée.
5. Selon le benchmark BraTS, cette méthode hybride CNN-Transformer pour la segmentation des tumeurs cérébrales multiples a surpassé les techniques concurrentes avec les meilleures notes en termes d'ET et de TC, mettant en évidence son efficacité.
6. Étude comparative de deux architectures hybrides 3D, le HybridUNet3D et le TransUNet3D, permettant d'analyser l'apport spécifique du module Transformer sur les performances de segmentation, la stabilité inter-patient et la robustesse clinique, et de justifier le choix de l'architecture finale retenue pour la comparaison avec l'état de l'art.

Bibliographie

[1] G. J. Augustine, D. Fitzpatrick, W. Hall, A.-S. Lamantia, D. Purves, and L. White, *Neurosciences*. De Boeck Superieur, 2015.

[2] T. Naidich, "Duvernoy's atlas of the human brain stem and cerebellum."

[3] M. F. Bear, B. W. Connors, M. A. Paradiso, and A. Neurosciences, "À la découverte du cerveau," *Editions Pradel*, 2002.

[4] L. Germond, "Trois principes de coopération pour la segmentation en imagerie de résonnance magnétique cérébrale," Ph.D. dissertation, Université Joseph-Fourier-Grenoble I, 1999.

[5] D. F. Quail and J. A. Joyce, "The microenvironmental landscape of brain tumors," *Cancer cell*, vol. 31, no. 3, pp. 326–341, 2017.

[6] D. B. Plewes and W. Kucharczyk, "Physics of mri : a primer," *Journal of magnetic resonance imaging*, vol. 35, no. 5, pp. 1038–1054, 2012.

[7] S. Kakarwal, R. Mapari, P. P. Mane, M. P. Borawake, D. R. Dhotre, and R. V. Patil, "A novel approach for detection, segmentation, and classification of brain tumors in mri images using neural network and special c means fuzzy clustering techniques," *Advances in Nonlinear Variational Inequalities*, vol. 27, no. 3, pp. 837–872, 2024.

[8] R. Kaifi, "A review of recent advances in brain tumor diagnosis based on ai-based classification," *Diagnostics*, vol. 13, no. 18, p. 3007, 2023.

[9] V. M. Runge, S. Aoki, W. G. Bradley Jr, K.-H. Chang, M. Essig, L. Ma, J. S. Ross, and A. Valavanis, "Magnetic resonance imaging and computed tomography of the brain—50 years of innovation, with a focus on the future," *Investigative radiology*, vol. 50, no. 9, pp. 551–556, 2015.

[10] U. Baid, S. Ghodasara, S. Mohan, M. Bilello, E. Calabrese, E. Colak, K. Farahani, J. Kalpathy-Cramer, F. C. Kitamura, S. Pati *et al.*, "The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification," *arXiv preprint arXiv :2107.02314*, 2021.

[11] C. Mead, "Neuromorphic electronic systems," *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1629–1636, 2002.

[12] C. Lee, N. W. Benjamin, and S. Pokharel, "Managing uncertain inventory in supply chain with neural network and radio frequency identification (rfid)," in *Supply Chain Innovation for Competing in Highly Dynamic Markets : Challenges and Solutions*. IGI Global Scientific Publishing, 2012, pp. 155–170.

[13] S. Hasan, N. Alam, G. A. Mashud, and S. Bhujel, "Neural network for enhancing robot assisted rehabilitation : A systematic," 2024.

[14] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks : an overview and application in radiology," *Insights into imaging*, vol. 9, no. 4, pp. 611–629, 2018.

[15] X. Zhao, L. Wang, Y. Zhang, X. Han, M. Deveci, and M. Parmar, "A review of convolutional neural networks in computer vision." *Artificial Intelligence Review*, vol. 57, no. 4, 2024.

[16] L. NOUIRI and S. DERRECHE, "Segmentation des tumeurs cérébrales dans des images irm par la méthode u-net," Ph.D. dissertation, UNIVERSITY BBA, 2024.

[17] S. G. Waxman and S. G. Waxman, *Clinical neuroanatomy*. McGraw-Hill Medical New York, NY, 2010.

[18] O. mondiale de la Santé *et al.*, "Résolutions et décisions d'intérêt régional adoptées par la soixante-dixième assemblée mondiale de la santé et par le conseil exécutif à ses cent quarantième et cent quarante et unième sessions," Organisation mondiale de la santé, Bureau régional pour la Méditerranée ..., Tech. Rep., 2017.

[19] J. Anthony, *Morphologie externe du cerveau des singes platyrhiniens*. Masson, 1947.

[20] N. Wiest-Daesslé, "Imagerie du tenseur de diffusion pour l'étude de pathologies cérébrales," Ph.D. dissertation, 2009.

[21] P. Janet, *Le cerveau et la pensée*. Germer Baillière, 1867.

[22] J.-P. Barral, *Approche ostéopathique du cerveau*. Elsevier Health Sciences, 2021.

[23] G. Paxinos and X.-F. Huang, *Atlas of the human brainstem*. Elsevier, 2013.

[24] B. R. Sabrine and K. Meryem, "Segmentation des tumeurs cérébrales par la ligne de partage des eaux sur des images irm," *mémoire de projet de fin d'étude pour l'obtention du : Diplôme de Master en Électronique Biomédicale*, vol. 2011, 2010.

[25] T. J. R. AZOCLI, "Relation entre leucorachie et méningite bactérienne," EPAC/UAC, Tech. Rep., 2018.

[26] G. Florence, "La physiologie de la circulation cérébrale," *Bulletin de l'Académie vétérinaire de France*, vol. 157, no. 3, pp. 19–30, 2004.

[27] A. Patel, "Benign vs malignant tumors," *JAMA oncology*, vol. 6, no. 9, pp. 1488–1488, 2020.

[28] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest *et al.*, "The multimodal brain tumor image segmentation benchmark (brats)," *IEEE transactions on medical imaging*, vol. 34, no. 10, pp. 1993–2024, 2014.

[29] K. Hoang-Xuan and A. Idbaih, "Progrès récents dans la génomique et dans le traitement médical des gliomes cérébraux," *Bulletin de l'Académie nationale de médecine*, vol. 195, no. 1, pp. 11–21, 2011.

[30] B. Kastler and D. Vetter, *Comprendre l'IRM : manuel d'auto-apprentissage*. Elsevier Health Sciences, 2018.

[31] M. Vermandel, "Mise en correspondance tridimensionnelle d'images multimodales application aux systèmes d'imageries projective et tomographique d'angiographie cérébrale," Ph.D. dissertation, Université des Sciences et Technologie de Lille-Lille I, 2002.

[32] L. Weyandt, *Clinical neuroscience : foundations of psychological and neurodegenerative disorders*. Routledge, 2018.

[33] F. P. Kuhn, "Tep/irm : l'imagerie hybride de l'avenir."

[34] S. T. Sorokan, A. L. Jefferies, and S. P. Miller, "L'imagerie du cerveau du nouveau-né à terme," *Paediatrics & Child Health*, vol. 23, no. 5, pp. 329–335, 2018.

[35] H. Mitchell, T. Hamilton, F. Steggerda, and H. Bean, "The chemical composition of the adult human body and its bearing on the biochemistry of growth," *Journal of Biological Chemistry*, vol. 158, no. 3, pp. 625–637, 1945.

[36] S. M. G. de Carvalho, "Classification of dementias based on brain radiomics features," Master's thesis, Universidade do Minho (Portugal), 2021.

[37] D. Last, "Caractérisation de l'architecture tridimensionnelle du tissu trabéculaire in vitro et in vivo par irm," Ph.D. dissertation, Université Paris Sud-Paris XI, 2003.

[38] J. N. Fernandes, V. E. Cardoso, A. Comesana-Campos, and A. Pinheira, "Comprehensive review : machine and deep learning in brain stroke diagnosis," *Sensors*, vol. 24, no. 13, p. 4355, 2024.

[39] D. W. McRobbie, E. A. Moore, M. J. Graves, and M. R. Prince, *MRI from Picture to Proton*. Cambridge university press, 2017.

[40] D. Azcona, K. McGuinness, and A. F. Smeaton, "A comparative study of existing and new deep learning methods for detecting knee injuries using the mrnet dataset," in *2020 International Conference on Intelligent Data Science Technologies and Applications (IDSTA)*. IEEE, 2020, pp. 149–155.

[41] R. Bitar, G. Leung, R. Perng, S. Tadros, A. R. Moody, J. Sarrazin, C. McGregor, M. Christakis, S. Symons, A. Nelson *et al.*, "Mr pulse sequences : what every radiologist wants to know but is afraid to ask," *Radiographics*, vol. 26, no. 2, pp. 513–537, 2006.

[42] G. B. Chavhan, P. S. Babyn, B. Thomas, M. M. Shroff, and E. M. Haacke, "Principles, techniques, and applications of t2*-based mr imaging and its special applications," *Radiographics*, vol. 29, no. 5, pp. 1433–1449, 2009.

[43] J. DROCHMANS, "Towards improving multiple sclerosis lesion segmentation performance with synthetic lesions."

[44] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. S. Kirby, J. B. Freymann, K. Farahani, and C. Davatzikos, "Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features," *Scientific data*, vol. 4, no. 1, p. 170117, 2017.

[45] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi, R. T. Shinohara, C. Berger, S. M. Ha, M. Rozycki *et al.*, "Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge," *arXiv preprint arXiv :1811.02629*, 2018.

[46] A. Myronenko and A. Hatamizadeh, "Robust semantic segmentation of brain tumor regions from 3d mris," in *International MICCAI Brainlesion Workshop*. Springer, 2019, pp. 82–89.

[47] R. R. Agravat and M. S. Raval, "Brain tumor segmentation and survival prediction," in *International MICCAI Brainlesion Workshop*. Springer, 2019, pp. 338–348.

[48] S. Li, X. Sui, X. Luo, X. Xu, Y. Liu, and R. Goh, "Medical image segmentation using squeeze-and-expansion transformers," *arXiv preprint arXiv :2105.09511*, 2021.

[49] X. Li, G. Luo, and K. Wang, "Multi-step cascaded networks for brain tumor segmentation," in *International MICCAI brainlesion workshop*. Springer, 2019, pp. 163–173.

[50] F. Wang, R. Jiang, L. Zheng, C. Meng, and B. Biswal, "3d u-net based brain tumor segmentation and survival days prediction," in *International MICCAI Brainlesion Workshop*. Springer, 2019, pp. 131–141.

[51] W. Wang, C. Chen, M. Ding, H. Yu, S. Zha, and J. Li, "Transbts : Multimodal brain tumor segmentation using transformer," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2021, pp. 109–119.

[52] T. Y. Gishene, "Brain tumor segmentation using deep learning : A hybrid model." 2025.

[53] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[54] S. MERDJI and S. REBIAI, "Prédiction des tumeurs cérébrales dans les images irm par l'apprentissage profond," Ph.D. dissertation, Université de Bordj Bou Arreridj Faculty of Mathematics and Computer Science, 2024.

[55] P. Ashtari, S. Van Huffel, D. Sappey-Marinier, and F. Maes, "Brain lesion segmentation and detection on multi-parametric mri data : Marrying deep learning with low-rank factorization," 2023.

[56] O. Ronneberger, P. Fischer, and T. Brox, "U-net : Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[57] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words : Transformers for image recognition at scale," *arXiv preprint arXiv :2010.11929*, 2020.

[58] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "Segformer : Simple and efficient design for semantic segmentation with transformers," *Advances in neural information processing systems*, vol. 34, pp. 12 077–12 090, 2021.

[59] A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H. R. Roth, and D. Xu, "Unetr : Transformers for 3d medical image segmentation," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2022, pp. 574–584.

[60] J. Dolz, K. Gopinath, J. Yuan, H. Lombaert, C. Desrosiers, and I. B. Ayed, "Hyperdense-net : A hyper-densely connected cnn for multi-modal image segmentation," *IEEE transactions on medical imaging*, vol. 38, no. 5, pp. 1116–1126, 2018.

[61] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net : Fully convolutional neural networks for volumetric medical image segmentation," in *2016 fourth international conference on 3D vision (3DV)*. Ieee, 2016, pp. 565–571.

[62] K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, "Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation," *Medical image analysis*, vol. 36, pp. 61–78, 2017.

[63] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz *et al.*, "Attention u-net : Learning where to look for the pancreas," *arXiv preprint arXiv :1804.03999*, 2018.

[64] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv :1706.05587*, 2017.

[65] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnu-net : a self-configuring method for deep learning-based biomedical image segmentation," *Nature methods*, vol. 18, no. 2, pp. 203–211, 2021.

[66] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, "Transunet : Transformers make strong encoders for medical image segmentation," *arXiv preprint arXiv :2102.04306*, 2021.

[67] M. D. Cirillo, D. Abramian, and A. Eklund, "Vox2vox : 3d-gan for brain tumour segmentation," in *International MICCAI Brainlesion Workshop*. Springer, 2020, pp. 274–284.

[68] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, "Swin-unet : Unet-like pure transformer for medical image segmentation," in *European conference on computer vision*. Springer, 2022, pp. 205–218.

[69] W. K. Soh, H. Y. Yuen, and J. C. Rajapakse, "Hut : Hybrid unet transformer for brain lesion and tumour segmentation," *Heliyon*, vol. 9, no. 12, 2023.

[70] N. Ayache, N. Cordier, H. Delingette *et al.*, "The multimodal brain tumor image segmentation benchmark (brats)," 2014.

[71] F. Isensee, P. F. Jäger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, "Automated design of deep learning methods for biomedical image segmentation," *arXiv preprint arXiv :1904.08128*, 2019.

[72] K.-R. Müller, S. Mika, K. Tsuda, and K. Schölkopf, "An introduction to kernel-based learning algorithms," in *Handbook of neural network signal processing*. CRC Press, 2018, pp. 4–1.

[73] M. Thorat, S. Pandit, and S. Balote, "Artificial neural network : a brief study," *Asian Journal For Convergence In Technology (AJCT) ISSN-2350-1146*, vol. 8, no. 3, pp. 12–16, 2022.

[74] S.-H. Han, K. W. Kim, S. Kim, and Y. C. Youn, "Artificial neural network : understanding the basic concepts without mathematics," *Dementia and neurocognitive disorders*, vol. 17, no. 3, pp. 83–89, 2018.

[75] T. P. Lillicrap, A. Santoro, L. Marris, C. J. Akerman, and G. Hinton, "Backpropagation and the brain," *Nature Reviews Neuroscience*, vol. 21, no. 6, pp. 335–346, 2020.

[76] S. Sharma, S. Sharma, and A. Athaiya, "Activation functions in neural networks," *Towards Data Sci*, vol. 6, no. 12, pp. 310–316, 2017.

[77] H. H. Dam, H. A. Abbass, C. Lokan, and X. Yao, "Neural-based learning classifier systems," *IEEE transactions on knowledge and data engineering*, vol. 20, no. 1, pp. 26–39, 2008.

[78] H. Kabbay, "Artificial neural network concepts and examples," 2022.

[79] M. C. Nwadiugwu, "Neural networks, artificial intelligence and the computational brain," *arXiv preprint arXiv :2101.08635*, 2020.

[80] O. Eluyode and D. T. Akomolafe, "Comparative study of biological and artificial neural networks," *European Journal of Applied Engineering and Scientific Research*, vol. 2, no. 1, pp. 36–46, 2013.

[81] Y. Pan, "Different types of neural networks and applications : evidence from feedforward, convolutional and recurrent neural networks," *Highlights in Science, Engineering and Technology*, vol. 85, pp. 247–255, 2024.

[82] C. Janiesch, P. Zschech, and K. Heinrich, "Machine learning and deep learning : C. janiesch et al." *Electronic markets*, vol. 31, no. 3, pp. 685–695, 2021.

[83] J. Duguet, G. Chassang, and J. Béranger, "Enjeux, répercussions et cadre éthique relatifs à l'intelligence artificielle en santé : Vers une intelligence artificielle éthique en médecine," *Droit, santé et société*, vol. 3, no. 3, pp. 30–39, 2019.

[84] S. Bédard-Venne, "Réseau de neurones artificiels et applications à la classification d'images," 2018.

[85] M. Mishra, "Convolutional neural networks, explained," *Towards Data Science*, vol. 26, 2020.

[86] S. F. Rabby, M. A. Arafat, and T. Hasan, "Bt-net : An end-to-end multi-task architecture for brain tumor classification, segmentation, and localization from mri images," *Array*, vol. 22, p. 100346, 2024.

[87] P. Grover, "5 regression loss functions all machine learners should know," Heartbeat, Fritz AI, June 2018.

[88] H. CHAKIB, "contribution à la compression d'images par la transformation en ondelettes et les réseaux de neurones artificiels," 2017.

[89] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[90] M. Drozdzal, E. Vorontsov, G. Chartrand, S. Kadoury, and C. Pal, "The importance of skip connections in biomedical image segmentation," in *International workshop on deep learning in medical image analysis*. Springer, 2016, pp. 179–187.

[91] C. Li, K. Liu, and S. Liu, "A survey of loss functions in deep learning," *Mathematics*, vol. 13, no. 15, p. 2417, 2025.

[92] Z. Qin, D. Kim, and T. Gedeon, "Rethinking softmax with cross-entropy : Neural network classifier as mutual information estimator," *arXiv preprint arXiv :1911.10688*, 2019.

[93] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http ://www.deeplearningbook.org.

[94] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012.

[95] L. N. Smith, "Cyclical learning rates for training neural networks," in *2017 IEEE winter conference on applications of computer vision (WACV)*. IEEE, 2017, pp. 464–472.

[96] D. P. Kingma and J. Ba, "Adam : A method for stochastic optimization," *arXiv preprint arXiv :1412.6980*, 2014.

[97] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net : learning dense volumetric segmentation from sparse annotation," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2016, pp. 424–432.

[98] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.

[99] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[100] R. Xiong, Y. Yang, D. He, K. Zheng, S. Zheng, C. Xing, H. Zhang, Y. Lan, L. Wang, and T. Liu, "On layer normalization in the transformer architecture," in *International conference on machine learning*. PMLR, 2020, pp. 10 524–10 533.

[101] D. Hendrycks and K. Gimpel, "Gaussian error linear units (gelus)," *arXiv preprint arXiv :1606.08415*, 2016.

[102] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert : Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics : human language technologies, volume 1 (long and short papers)*, 2019, pp. 4171–4186.

[103] Y. Wu and K. He, "Group normalization," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.

[104] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[105] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam : Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.

[106] G. Huang, Y. Sun, Z. Liu, D. Sedra, and K. Q. Weinberger, "Deep networks with stochastic depth," in *European conference on computer vision*. Springer, 2016, pp. 646–661.

[107] I. Loshchilov and F. Hutter, "Sgdr : Stochastic gradient descent with warm restarts," *arXiv preprint arXiv :1608.03983*, 2016.

[108] B. T. Polyak and A. B. Juditsky, "Acceleration of stochastic approximation by averaging," *SIAM journal on control and optimization*, vol. 30, no. 4, pp. 838–855, 1992.

[109] M. Berman, A. R. Triki, and M. B. Blaschko, "The lovász-softmax loss : A tractable surrogate for the optimization of the intersection-over-union measure in neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4413–4421.

[110] H. Kervadec, J. Bouchtiba, C. Desrosiers, E. Granger, J. Dolz, and I. B. Ayed, "Boundary loss for highly unbalanced segmentation," in *International conference on medical imaging with deep learning*. PMLR, 2019, pp. 285–296.

[111] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *arXiv preprint arXiv :1711.05101*, 2017.

[112] P. Goyal, P. Dollár, R. Girshick, P. Noordhuis, L. Wesolowski, A. Kyrola, A. Tulloch, Y. Jia, and K. He, "Accurate, large minibatch sgd : Training imagenet in 1 hour," *arXiv preprint arXiv :1706.02677*, 2017.

[113] P. Micikevicius, S. Narang, J. Alben, G. Diamos, E. Elsen, D. Garcia, B. Ginsburg, M. Houston, O. Kuchaiev, G. Venkatesh *et al.*, "Mixed precision training," *arXiv preprint arXiv :1710.03740*, 2017.