



كلية العلوم  
السملاية - مراكش  
FACULTÉ DES SCIENCES  
SEMLALIA - MARRAKECH

**Université Cadi Ayyad**  
Faculté des Sciences Semlalia – Marrakech

**Master : Ingénierie des Systèmes d'Information (ISI)**

**Projet : Deep Learning**

---

# Détection et Classification de Tumeurs Cérébrales par ConvNeXt

---

**Sous l'encadrement de :**

Pr. Mustapha AATILA

**Réalisé par :**

Maryem YMIL  
Wiam MAKTOUB  
Maryam BOUFOUS

Année Universitaire : 2025–2026

# Table des matières

<b>1</b>	<b>Lien du Repository Git</b>	<b>3</b>
	<b>Abstract</b>	<b>4</b>
	<b>Résumé</b>	<b>5</b>
<b>2</b>	<b>Introduction</b>	<b>6</b>
1	Transfer Learning : Définition et principes de base . . . . .	6
2	Problématique . . . . .	6
3	Motivations . . . . .	7
4	Objectifs du projet . . . . .	7
<b>3</b>	<b>Travaux connexes</b>	<b>8</b>
1	Aperçu des travaux ayant traité le même sujet . . . . .	8
1.1	Méthodes CNN classiques . . . . .	8
1.2	Approches modernes . . . . .	8
1.3	Approches multi-tâches et segmentation intégrée . . . . .	8
1.4	CNN appliqués en médecine nucléaire . . . . .	9
1.5	Modèles hybrides ConvNeXt–Vision Transformer appliqués à la schizophrénie . . . . .	9
2	Comparaison synthétique de ces travaux . . . . .	9
3	Positionnement du travail en cours . . . . .	10
<b>4</b>	<b>Méthodologie adoptée</b>	<b>11</b>
1	Description des données utilisées . . . . .	11
1.1	Visualisation et distribution des données . . . . .	12
2	Présentation du modèle adopté . . . . .	14
2.1	Justification du choix . . . . .	14
3	Description des étapes de la méthodologie . . . . .	14
3.1	Prétraitement des données . . . . .	14
3.2	Séparation des données (Training / Test) . . . . .	15
3.3	Hyperparamètres adoptés . . . . .	15
3.4	Métriques d'évaluation adoptées . . . . .	16
3.5	Analyse Visuelle des Régions d'Intérêt via Grad-CAM . . . . .	18
3.6	Adaptations apportées au modèle . . . . .	19
<b>5</b>	<b>Discussion</b>	<b>20</b>
1	Résumé des résultats obtenus . . . . .	20
2	Discussion des résultats obtenus . . . . .	20
2.1	Qualité et robustesse des résultats . . . . .	20

2.2	Visualisation avec Grad-CAM . . . . .	20
2.3	Sources possibles d'erreurs observées . . . . .	22
2.4	Évaluation du surapprentissage (overfitting) . . . . .	22
2.5	Robustesse statistique et homogénéité des performances . . . . .	22
3	Comparaison avec la littérature . . . . .	23
4	Implications, portée et points forts de la méthode . . . . .	23
4.1	Implications cliniques . . . . .	23
4.2	Portée scientifique et technique . . . . .	23
4.3	Points forts de la méthode utilisée . . . . .	23
<b>6</b>	<b>Conclusion et perspectives</b>	<b>24</b>
1	Rappel de l'objectif et synthèse des résultats . . . . .	24
2	Points forts du travail . . . . .	24
3	Limites identifiées . . . . .	25
4	Perspectives et travaux futurs . . . . .	25
<b>7</b>	<b>Répartition des Tâches</b>	<b>26</b>

# Chapitre 1

## Lien du Repository Git

### Accès au code source

`https://github.com/wiammaktoub/brain-tumor-classification-convnext`

#### Contenu du repository :

- Dataset
- Notebooks Jupyter d'analyse et d'entraînement
- Rapport final au format PDF
- Fichier README.md
- Resultats obtenues

# Abstract

This study explores a transfer learning-based approach for the automated classification of brain magnetic resonance imaging (MRI) into four diagnostic categories : glioma, meningioma, pituitary tumor, and no tumor. We employed a pre-trained ConvNeXt Tiny architecture, originally trained on ImageNet, and fine-tuned it on a merged dataset combining images from Figshare, SARTAJ, and Br35H. The final dataset contained 7,023 images, split into 5,712 for training and 1,311 for testing.

The preprocessing pipeline involved resizing all images to  $224 \times 224$ , intensity normalization, and a set of data augmentation techniques designed to enhance robustness against variations in acquisition conditions. During fine-tuning, the classifier layer of ConvNeXt was adapted to a four-class output, and the model was trained using Adam optimization and early stopping.

The proposed model achieved state-of-the-art performance, with a test accuracy of 99.08 %, and precision, recall, and F1-score values around 0.991. The per-class ROC-AUC values exceeded 0.9993, confirming excellent discriminative ability across all tumor types. The report provides a detailed discussion of the methodological choices, data preparation strategies, model adaptations, experimental results, as well as the main limitations of the approach and potential future research avenues, multi-center validation, and extension toward tumor segmentation.

# Résumé

Ce travail propose une approche de classification automatique des tumeurs cérébrales à partir d'images IRM en s'appuyant sur le Transfer Learning. Nous avons utilisé l'architecture ConvNeXt Tiny, pré-entraînée sur ImageNet, puis adaptée et ré-entraînée pour distinguer quatre catégories : gliome, méningiome, pituitaire et absence de tumeur. Le dataset final, issu de la fusion de plusieurs sources publiques (Figshare, SARTAJ et Br35H), comporte 7 023 images, dont 5 712 pour l'apprentissage et 1 311 pour le test.

La chaîne de prétraitement inclut le redimensionnement des images en  $224 \times 224$ , la normalisation et plusieurs techniques d'augmentation destinées à renforcer la généralisation du modèle. La couche de classification du ConvNeXt a été modifiée pour permettre une prédiction à quatre classes, puis le modèle a été entraîné avec l'optimiseur Adam, combiné à un mécanisme d'early stopping pour prévenir le surapprentissage.

Les performances obtenues sont exceptionnelles : une accuracy de 99,08

Le rapport discute les choix méthodologiques, les adaptations du modèle, les métriques d'évaluation, ainsi que les principales limites, telles que l'hétérogénéité des données et l'absence de validation multi-centrique. Des perspectives futures sont proposées, notamment l'extension vers la segmentation tumorale et l'évaluation sur des données cliniques réelles.

# Chapitre 2

## Introduction

### 1 Transfer Learning : Définition et principes de base

Le **Transfer Learning** (apprentissage par transfert) est une approche de plus en plus utilisée en intelligence artificielle et en vision par ordinateur, permettant de *réutiliser les connaissances acquises par un modèle sur une tâche source pour une nouvelle tâche cible*. Dans la pratique, les modèles pré-entraînés sur de larges ensembles de données génériques, tels que *ImageNet* qui contient plus de 14 millions d’images annotées pour 1000 classes, peuvent servir de base à de nouvelles applications avec des données plus restreintes.

Le principe fondamental du Transfer Learning repose sur trois concepts clés :

1. **Feature reuse** : les premières couches d’un réseau profond apprennent des représentations générales des images, telles que les bords, textures, contours et motifs élémentaires. Ces caractéristiques sont généralement transférables à de nombreuses tâches visuelles.
2. **Head adaptation** : la “tête” du réseau, c’est-à-dire les couches finales responsables de la classification, est remplacée pour correspondre au nombre de classes et à la nature de la tâche cible. Par exemple, pour un problème de classification de tumeurs cérébrales en quatre catégories, la couche finale doit produire quatre sorties probabilistes via une fonction *softmax*.
3. **Fine-tuning** : cette étape consiste à ajuster les poids du modèle sur le nouveau dataset. Elle peut se faire en deux phases :
  - *Phase 1* : geler la base du réseau et n’entraîner que la tête, afin d’adapter le classifieur aux nouvelles classes.
  - *Phase 2* : débloquer certaines couches de la base et entraîner le modèle entier avec un faible taux d’apprentissage pour ajuster progressivement les filtres aux spécificités du dataset cible.

Les principaux avantages de cette approche sont :

- **Réduction significative du temps d’entraînement**, car les poids initiaux sont déjà adaptés à la reconnaissance d’images.
- **Besoins en données moindres**, ce qui est particulièrement important dans le domaine médical où l’acquisition de données annotées est coûteuse et chronophage.
- **Meilleure généralisation** sur le dataset cible, surtout lorsque celui-ci est de taille limitée ou hétérogène.

### 2 Problématique

La détection et la classification automatiques des tumeurs cérébrales à partir d’images IRM constituent un enjeu majeur pour le diagnostic médical. Une identification précoce et précise des types de tumeurs permet :

- d’orienter les décisions cliniques,

- de prioriser les interventions chirurgicales ou thérapeutiques,
- et de fournir une aide à la décision aux radiologues afin de réduire la variabilité inter-observateur.

L’objectif concret de ce projet est de concevoir un pipeline fiable capable de classer les IRM cérébrales en quatre catégories : *glioma*, *meningioma*, *pituitary*, et *no tumor*, tout en garantissant des performances robustes et interprétables.

### 3 Motivations

Plusieurs motivations scientifiques et pratiques justifient ce travail :

- **Soutien au diagnostic médical** : offrir une seconde opinion automatisée aux radiologues pour augmenter la précision et réduire le risque d’erreur.
- **Évaluation des architectures modernes** : tester l’efficacité des modèles récents tels que *ConvNeXt* dans le contexte d’imagerie médicale.
- **Reproductibilité et accessibilité** : fournir un pipeline complet, incluant code, notebooks et rapports, facilement utilisable en contexte académique et clinique.
- **Interprétabilité** : produire des visualisations explicatives (*Grad-CAM*) pour vérifier que les décisions du modèle reposent sur des zones d’intérêt pertinentes sur les images IRM.

### 4 Objectifs du projet

Le projet se fixe les objectifs techniques et scientifiques suivants :

1. **Prétraitement et analyse du dataset** : examiner la structure, le déséquilibre entre classes et la distribution des résolutions des images.
2. **Adaptation et fine-tuning de ConvNeXt Tiny** : mettre en œuvre le Transfer Learning pour maximiser la précision de classification.
3. **Évaluation des performances** : utiliser des métriques standard telles que *Accuracy*, *Precision*, *Recall*, *F1-score*, *ROC-AUC* et matrice de confusion.
4. **Étude de la robustesse et de l’overfitting** : analyser la stabilité du modèle et proposer des méthodes d’amélioration et d’interprétabilité.

Cette structuration détaillée permet de poser des bases solides pour le reste du rapport, en combinant rigueur scientifique, clarté méthodologique et orientation pratique.



# Chapitre 3

## Travaux connexes

Cette section présente un état de l’art sur la classification et la segmentation des tumeurs cérébrales à partir d’images IRM et TEP/IRM, ainsi que sur l’utilisation de modèles hybrides ConvNeXt–Vision Transformer appliqués à d’autres pathologies neurologiques telles que la schizophrénie. L’objectif est de positionner notre travail dans le contexte scientifique existant, en mettant en évidence les méthodes classiques, les avancées récentes et les architectures modernes utilisées dans l’analyse d’images médicales.

### 1 Aperçu des travaux ayant traité le même sujet

La classification automatique d’images médicales repose largement sur les réseaux de neurones convolutifs (CNN) et les architectures hybrides intégrant des Transformers. Les approches existantes peuvent être regroupées en quatre grandes catégories.

#### 1.1 Méthodes CNN classiques

Les premières études ont utilisé des architectures CNN classiques telles que *VGGNet* [1] et *ResNet* [2], très répandues dans les tâches de classification en imagerie médicale.

Ces modèles montrent généralement des performances comprises entre 85% et 95% d’accuracy, souvent après :

- Normalisation des intensités IRM,
- Segmentation préalable de la tumeur ou de la région d’intérêt,
- Augmentation des données pour compenser la taille réduite des datasets.

Malgré leur robustesse, ces architectures souffrent de limitations : grand nombre de paramètres, temps d’entraînement élevé, et sensibilité aux variations anatomiques entre patients.

#### 1.2 Approches modernes

Les modèles récents tels que *EfficientNet* [3], *DenseNet* [4], *ConvNeXt* [5] et les Transformers visuels (*ViT*, *Swin-Transformer*) ont significativement amélioré les performances.

Ils offrent notamment :

- une meilleure efficacité paramétrique,
- une capacité accrue à capturer des structures anatomiques complexes,
- des performances de 96–99% d’accuracy sur des datasets de taille moyenne.

Les Transformers, grâce à l’attention multi-têtes, apprennent des dépendances globales entre régions d’une image, ce qui les rend particulièrement adaptés aux images 3D (IRM, fMRI).

#### 1.3 Approches multi-tâches et segmentation intégrée

Certaines études combinent classification et segmentation dans un cadre multi-tâches :

- Segmentation U-Net pour extraire la zone tumorale.
- Classification via un CNN ou un Transformer.
- Réduction des erreurs induites par le bruit anatomique hors de la zone d'intérêt.

Ces approches sont particulièrement pertinentes pour les IRM cérébrales où la zone anormale peut être de petite taille et difficile à discerner.

## 1.4 CNN appliqués en médecine nucléaire

Blanc-Durand [6] a utilisé des CNN 3D pour la reconstruction de pseudo-CT et la segmentation de tumeurs gliales en TEP/IRM, montrant :

- l'efficacité des U-Net 3D pour la segmentation volumétrique,
- l'intérêt des connexions résiduelles (ResNet 3D),
- l'apport des convolutions dilatées pour élargir le champ réceptif.

Cette étude souligne l'importance des architectures profondes 3D pour les images multi-modales.

## 1.5 Modèles hybrides ConvNeXt–Vision Transformer appliqués à la schizophrénie

Une étude récente (Pranto et al., 2025) propose un modèle hybride combinant **ConvNeXt** et **Vision Transformer (ViT)** pour la classification de la schizophrénie à partir de scans fMRI 3D.

Les contributions principales sont :

- **Extraction locale de caractéristiques** par ConvNeXt, capable de capturer des motifs neuroanatomiques fins.
- **Apprentissage global par ViT**, permettant de modéliser les relations fonctionnelles longues portées dans l'activité cérébrale.
- **Architecture hybride** intégrant :
  - une projection des features ConvNeXt vers l'espace des tokens,
  - un encodeur Transformer pour la modélisation séquentielle,
  - un class-token pour la prédiction finale.
- **Utilisation de fMRI 3D**, montrant que les modèles hybrides gèrent mieux la variabilité interpatients que les CNN seuls.
- Résultats prometteurs dans la détection automatique de la schizophrénie grâce à la combinaison du traitement local (CNN) et global (ViT).

Cette étude démontre la pertinence des architectures hybrides dans les problèmes de classification neurologique complexes.

## 2 Comparaison synthétique de ces travaux

Comparativement :

- **CNN classiques** : robustes mais limités pour la modélisation des dépendances globales.
- **Modèles modernes (EfficientNet, DenseNet)** : excellentes performances et efficacité paramétrique.
- **ConvNeXt / ViT** : performances parmi les plus élevées (97–99%), capacité à capturer des structures globales.

- **Modèles hybrides ConvNeXt–ViT** : pertinents pour les données 3D complexes, comme démontré par Pranto et al. (2025).
- **U-Net / ResNet 3D** : indispensables pour la segmentation volumétrique et la reconstruction.

Les critères de comparaison incluent :

- précision (accuracy, AUC),
- complexité du modèle,
- capacité à exploiter les informations locales et globales,
- robustesse face aux variations interpatients.

### 3 Positionnement du travail en cours

Notre travail se distingue par :

- l'utilisation de **ConvNeXt Tiny** en transfer learning sur un dataset multimodal composé de 7 023 images (Figshare, SARTAJ, Br35H),
- un **pipeline reproductible** incluant data exploration, entraînement, visualisation et gestion des checkpoints,
- l'ajout d'un module d'**interprétabilité Grad-CAM** garantissant une analyse physiologiquement cohérente,
- des performances très élevées : accuracy  $\approx 99.08\%$ , AUC  $> 0.999$ ,
- un positionnement clair dans la tendance actuelle des modèles modernes (ConvNeXt, ViT, architectures hybrides).

Ainsi, notre projet s'inscrit dans les travaux récents explorant des architectures profondes optimisées pour la classification d'images cérébrales, tout en se focalisant sur les IRM structurelles (tumeurs) plutôt que sur les fMRI fonctionnelles (schizophrénie). De plus, une partie de notre implémentation s'inspire du travail publié par S. R. Siam dans son dépôt GitHub [7], qui propose une base de code utile pour structurer des expériences de classification en imagerie médicale.

# Chapitre 4

## Méthodologie adoptée

### 1 Description des données utilisées

Le dataset utilisé pour ce projet est issu de la combinaison de trois sources principales : **Figshare**, **SARTAJ** et **Br35H** (repository : masoudnickparvar/brain-tumor-mri-dataset) [8]. Il comprend un total de 7 023 images d'IRM cérébrale, réparties entre l'ensemble d'entraînement et de test comme suit :

- **Training** : 5 712 images réparties sur 4 classes :
  - Glioma : 1 321 images
  - Meningioma : 1 339 images
  - Pituitary : 1 457 images
  - No Tumor : 1 595 images
- **Testing** : 1 311 images réparties comme suit :
  - Glioma : 300 images
  - Meningioma : 306 images
  - Pituitary : 300 images
  - No Tumor : 405 images

Les images sont majoritairement au format 512×512 pixels, mais certaines varient jusqu'à 1365×1365 pixels. Toutes les images seront normalisées pour l'entraînement.

**Remarques importantes :**

- La distribution est relativement équilibrée, la classe la plus grande (No Tumor) représentant environ 28% du dataset d'entraînement.
- Le dataset contient des images issues de différentes machines et résolutions, ce qui impose un prétraitement rigoureux pour homogénéiser les entrées.

## 1.1 Visualisation et distribution des données

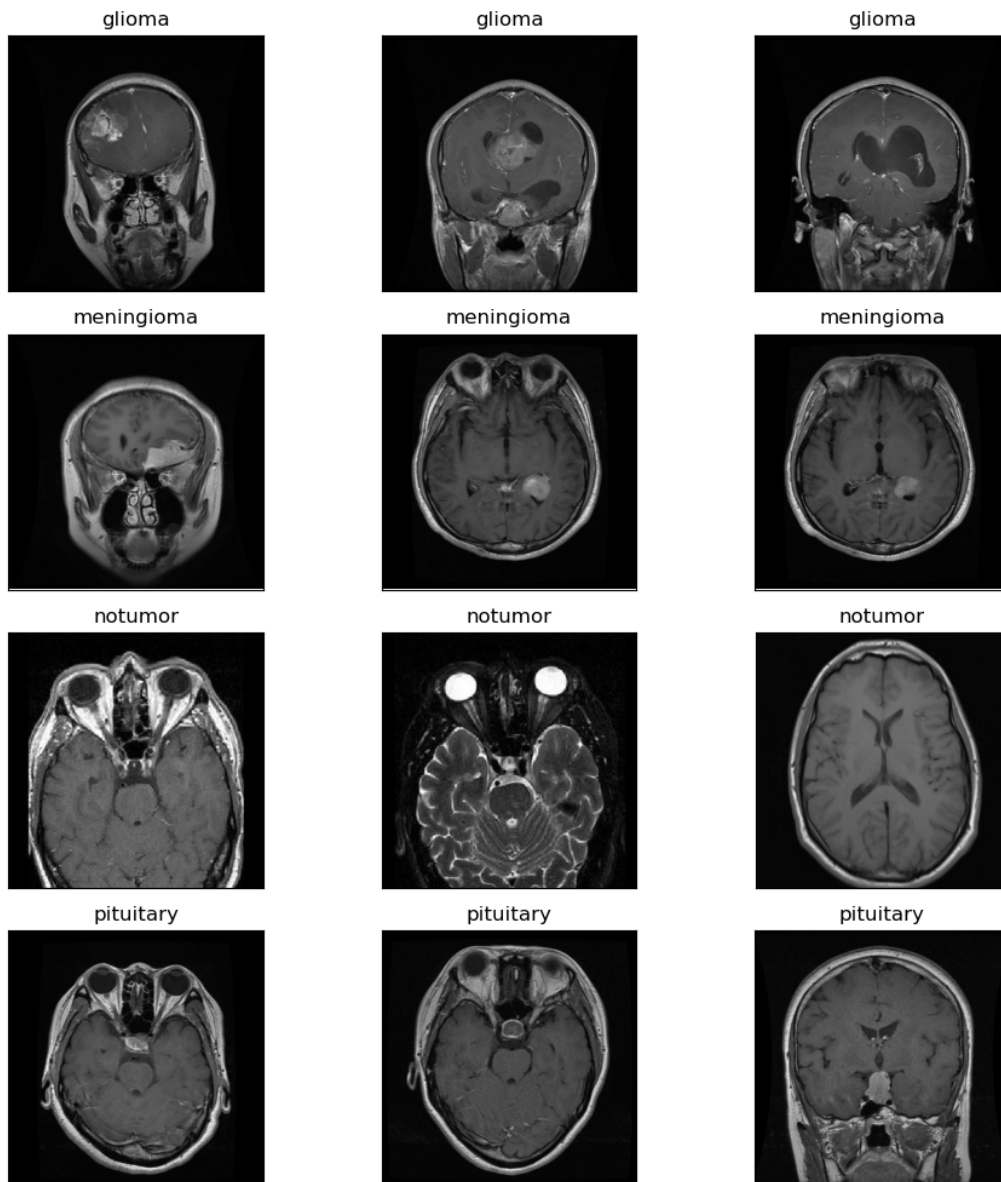


FIGURE 4.1 – Exemples d’images du dataset.

*Explication :* Cette figure illustre la variété des tumeurs cérébrales et des images *No Tumor*. On peut observer la diversité des formes, textures et localisations des anomalies, ce qui représente un défi pour le modèle et souligne la nécessité d’un prétraitement et d’augmentations adaptées.

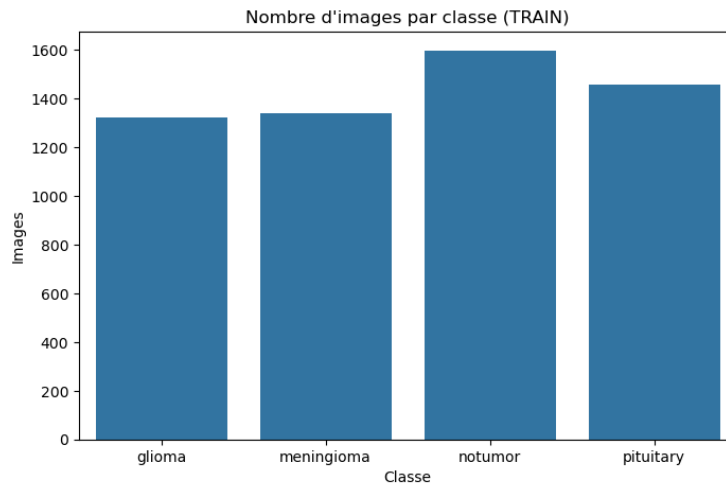


FIGURE 4.2 – Distribution des classes dans l'ensemble d'entraînement.

*Explication* : La figure montre que le dataset d'entraînement est relativement équilibré, chaque classe représentant une proportion proche de 25%, avec une légère prédominance de la classe *No Tumor*. Un bon équilibre réduit le risque de biais vers une classe majoritaire et améliore la performance globale du modèle.

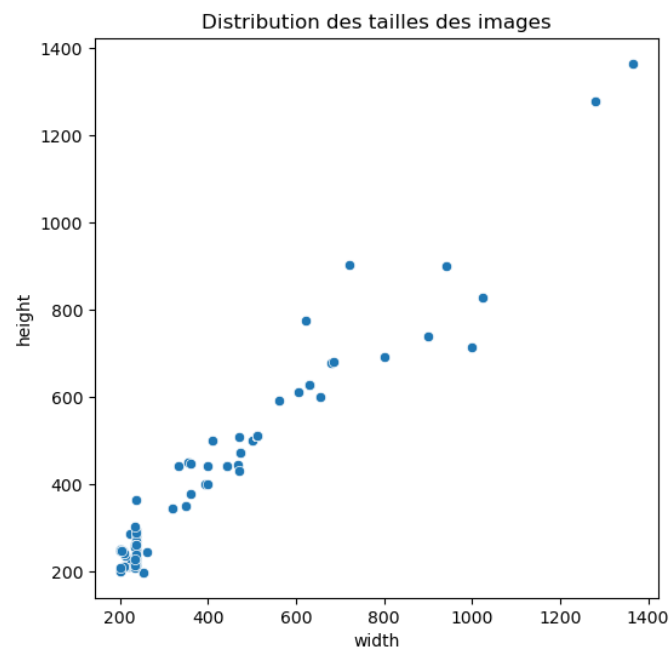


FIGURE 4.3 – Distribution des tailles des images.

*Explication* : La majorité des images sont autour de 512×512 pixels, mais certaines images sont plus grandes. Cette variabilité impose un redimensionnement ou un recadrage lors du prétraitement pour garantir une entrée homogène au modèle. Une telle standardisation est cruciale pour assurer la stabilité et la robustesse de l'entraînement.

## 2 Présentation du modèle adopté

Le modèle adopté pour la classification des tumeurs cérébrales repose sur l'architecture ConvNeXt, un réseau de neurones convolutionnels de dernière génération. Cette architecture a été conçue pour combiner la puissance des CNN traditionnels avec des innovations inspirées des architectures Transformer, permettant une extraction efficace des caractéristiques complexes dans les images. Elle est particulièrement adaptée au traitement des images médicales, où la précision et la robustesse sont essentielles.

### 2.1 Justification du choix

Le choix de ConvNeXt repose sur plusieurs avantages clés dans le contexte de l'imagerie médicale :

- **Précision élevée** : ConvNeXt offre une excellente capacité à distinguer des motifs subtils et complexes présents dans les IRM cérébrales, ce qui est crucial pour différencier les types de tumeurs.
- **Adaptabilité** : la structure modulaire du réseau permet d'adapter facilement la couche de classification aux quatre catégories ciblées, sans compromettre la performance.
- **Fine-tuning efficace** : grâce à un pré-entraînement sur de larges bases de données, le modèle peut être ajusté sur des datasets de taille limitée, typiques en contexte médical, tout en conservant une bonne généralisation.
- **Robustesse** : ConvNeXt est capable de gérer la variabilité des acquisitions IRM provenant de différentes sources et équipements, garantissant la fiabilité des prédictions.
- **Interprétabilité** : le modèle se prête à des méthodes de visualisation telles que Grad-CAM, permettant de vérifier que les décisions sont basées sur des régions anatomiquement pertinentes, ce qui est essentiel pour l'acceptation clinique.

## 3 Description des étapes de la méthodologie

### 3.1 Prétraitement des données

Le prétraitement constitue une étape essentielle pour garantir que les images IRM soient homogènes et adaptées à l'apprentissage profond. Les étapes appliquées dans notre pipeline sont les suivantes :

- **Redimensionnement** : toutes les images ont été uniformisées à une taille de  $224 \times 224$  pixels afin de correspondre aux dimensions d'entrée du modèle ConvNeXt Tiny. Cette standardisation est cruciale pour assurer la compatibilité avec le réseau et éviter des biais liés à la taille des images.
- **Normalisation** : les images ont été normalisées en utilisant les valeurs de moyenne et d'écart-type du dataset ImageNet, soit  $\text{mean} = [0.485, 0.456, 0.406]$ ,  $\text{std} = [0.229, 0.224, 0.225]$ . La normalisation permet de stabiliser et accélérer la convergence lors de l'entraînement en centrant les données autour de zéro et en réduisant la variance des pixels.
- **Data augmentation** (appliquée sur le train uniquement) : afin d'améliorer la robustesse et la généralisation du modèle, plusieurs transformations ont été appliquées :

- Rotations aléatoires  $\pm 10^\circ$
- Flip horizontal
- ColorJitter (brightness, contrast, saturation)
- Recadrage et redimensionnement aléatoire si nécessaire

Ces techniques permettent de simuler la variabilité naturelle des acquisitions IRM et d'augmenter la diversité du dataset sans collecter de nouvelles images.

- **Nettoyage** : suppression des fichiers corrompus, vérification des canaux et des formats pour s'assurer que toutes les images sont exploitables par le modèle.

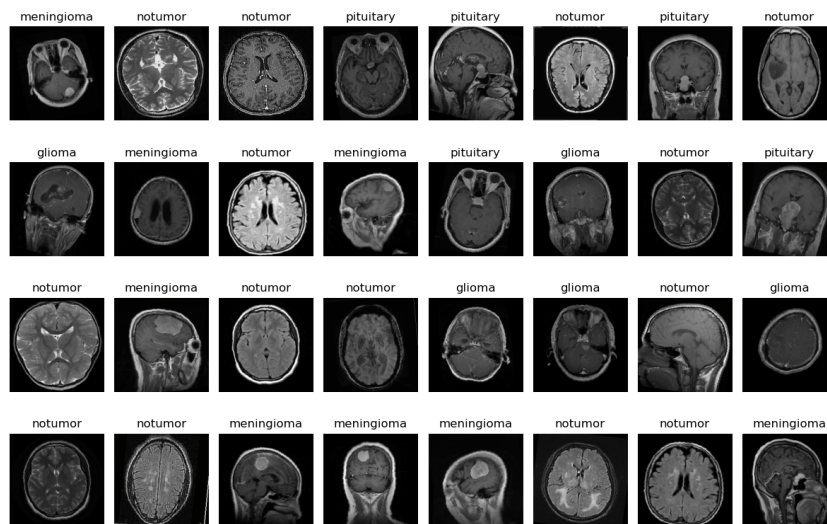


FIGURE 4.4 – Exemple d'un batch d'images prétraitées.

*Explication* : Cette figure illustre la diversité des images après les étapes de redimensionnement, normalisation et augmentation. On observe différentes orientations, contrastes et luminosités, simulant la variabilité naturelle des IRM. L'homogénéisation des dimensions et la normalisation garantissent que toutes les images sont compatibles avec l'entrée du modèle ConvNeXt Tiny, tout en maintenant les caractéristiques discriminantes des tumeurs.

### 3.2 Séparation des données (Training / Test)

Le dataset étant déjà divisé, nous avons utilisé les ensembles existants :

- **Training** : 5 712 images
- **Testing** : 1 311 images

Pour le réglage des hyperparamètres, nous avons éventuellement créé un sous-ensemble de validation à partir de l'ensemble d'entraînement, en conservant le test pour l'évaluation finale afin d'obtenir une estimation impartiale des performances du modèle.

### 3.3 Hyperparamètres adoptés

Les paramètres choisis pour l'entraînement sont les suivants :

- **Batch size** : 32
- **Learning rate** : 1e-3 initial, descente à 1e-4 pour le fine-tuning
- **Epochs** : 50 (early stopping avec patience de 5)
- **Optimiseur** : Adam



- **Loss function** : CrossEntropyLoss

Ces hyperparamètres ont été sélectionnés pour équilibrer rapidité d'entraînement et stabilité de la convergence, tout en limitant le surapprentissage.

### 3.4 Métriques d'évaluation adoptées

#### 3.4.1 Rappel des définitions et formules

Les performances du modèle ont été évaluées à l'aide de métriques standards en classification multi-classes :

- **TP, TN, FP, FN** : représentent respectivement le nombre de vrais positifs, vrais négatifs, faux positifs et faux négatifs pour chaque classe. Ces mesures permettent d'analyser avec précision les erreurs de classification.
- **Accuracy** : proportion de prédictions correctes sur l'ensemble des données. Calcul :

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Une valeur élevée indique que le modèle distingue efficacement les différentes classes.

- **Precision** : fiabilité des prédictions positives pour chaque classe :

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

Une precision élevée signifie que la majorité des instances prédites comme positives sont correctes.

- **Recall (Sensitivity)** : capacité du modèle à détecter toutes les instances positives d'une classe :

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

Un recall élevé montre que peu d'éléments positifs sont manqués.

- **F1-score** : combine précision et rappel :

$$F1 = 2 \cdot \frac{P \cdot R}{P + R}$$

Cette métrique est particulièrement utile lorsque les classes sont légèrement déséquilibrées.

- **ROC-AUC (multi-classes)** : calculée via la stratégie One-vs-Rest, elle mesure la capacité de séparation entre les classes. Une valeur proche de 1 indique une excellente discrimination.
- **Matrice de confusion** : permet de visualiser les erreurs par classe et de détecter des confusions spécifiques.

### 3.4.2 Résultats visuels et interprétation

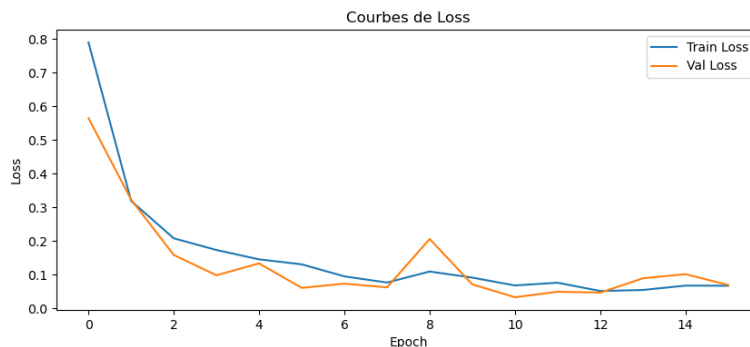


FIGURE 4.5 – Courbes de loss d'entraînement et de validation.

*Explication* : La décroissance rapide du train loss et le minimum atteint par la validation loss à l'époque 11 (0.033) indiquent un apprentissage efficace. La légère remontée après l'époque 11 est contrôlée par l'early stopping, limitant le surapprentissage.

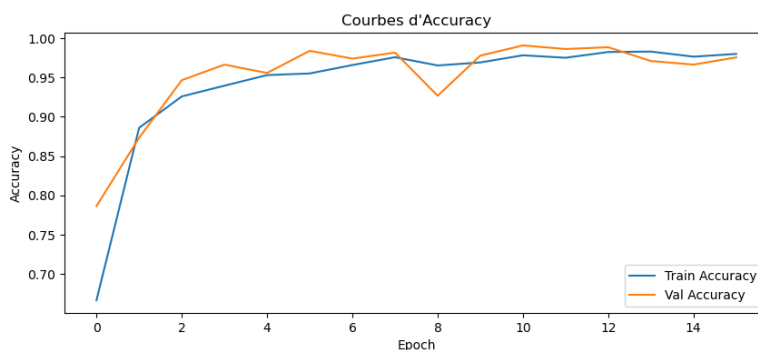


FIGURE 4.6 – Courbes d'accuracy d'entraînement et de validation.

*Explication* : L'accuracy de validation atteint 99.08% à l'époque 11, avec un train accuracy proche de 98%. La proximité des courbes indique une bonne généralisation et confirme l'absence de surapprentissage significatif.

### 3.4.3 Scores globaux de performance

Metric	Accuracy	Precision	Recall	F1-score
Score global	0.99085	0.99089	0.99085	0.99084

TABLE 4.1 – Scores globaux de performance du modèle sur le jeu de test.

*Explication* : Ces valeurs indiquent une performance très élevée et équilibrée, montrant que le modèle est précis, sensible et stable malgré les légers déséquilibres entre les classes.

### 3.4.4 ROC-AUC par classe

Classe	ROC-AUC
Glioma	0.99977
Meningioma	0.99931
Pituitary	0.99938
No Tumor	0.99997

TABLE 4.2 – ROC-AUC obtenus pour chaque classe du dataset.

*Explication* : Les valeurs proches de 1 indiquent que le modèle sépare efficacement les classes, même celles moins représentées, confirmant la robustesse et la fiabilité des prédictions.

### 3.4.5 Interprétation scientifique des métriques

- La **high accuracy** ( 99.08%) montre que le modèle classe correctement la grande majorité des images IRM.
- La **precision et recall** proches de 0.991 pour toutes les classes indiquent un équilibre entre la capacité à détecter toutes les instances positives et à éviter les fausses alertes.
- Le **F1-score** élevé confirme la robustesse globale du modèle face à de légers déséquilibres de classes.
- Les **courbes d'apprentissage**, la **matrice de confusion** et le **ROC-AUC** permettent d'identifier rapidement les classes présentant des confusions mineures et valident la stratégie d'entraînement.
- Les **scores globaux** confirment la cohérence et la stabilité du modèle sur toutes les métriques standards.

## 3.5 Analyse Visuelle des Régions d'Intérêt via Grad-CAM

L'interprétabilité est essentielle pour valider les décisions du modèle sur les images IRM. Grad-CAM permet de générer des cartes thermiques mettant en évidence les régions les plus influentes sur la prédiction.

Les cartes Grad-CAM montrent que le modèle focalise son attention sur :

- Les structures tumorales caractéristiques pour chaque classe : Glioma, Meningioma, Pituitary.
- Pour *No Tumor*, les activations sont diffuses, confirmant l'absence de signaux pathologiques.
- Les zones non pertinentes (artefacts ou contours osseux) sont rares, ce qui atteste de la robustesse du modèle.

Cette analyse permet de valider la cohérence médicale des prédictions et d'identifier d'éventuels biais ou erreurs.

L'intégration de la figure 4.4 renforce visuellement la compréhension du pipeline de prétraitement et illustre la diversité et la qualité des données préparées pour l'entraînement du modèle.

### 3.6 Adaptations apportées au modèle

Pour l'adaptation à notre tâche de classification en 4 classes, nous avons effectué :

- Remplacement de la tête originale par : Flatten → LayerNorm2d → Dropout(0.3) → Linear(768 → 4).
- Phase d'entraînement en deux étapes :
  1. Base gelée, entraînement uniquement de la tête.
  2. Déblocage partiel de certaines couches pour un *fine-tuning* à faible taux d'apprentissage.
- Hyperparamètres principaux :
  - Learning rate initial : 1e-3
  - Batch size : 32
  - Epochs : 50 avec early stopping (patience 5)
  - Optimizer : Adam
  - Loss function : CrossEntropyLoss
- Sauvegarde du meilleur modèle : `best_convnext_model.pth`, basé sur la validation accuracy / loss.

# Chapitre 5

## Discussion

### 1 Résumé des résultats obtenus

Le pipeline basé sur le modèle **ConvNeXt Tiny** en *Transfer Learning* a montré des performances exceptionnelles sur le dataset combiné (Figshare + SARTAJ + Br35H). Les principales métriques sont les suivantes :

- **Accuracy (validation/test)** : 99.08%
- **F1-score pondéré** : 0.9908
- **ROC-AUC par classe** : proche de 1 pour toutes les classes (glioma 0.9998, meningioma 0.9993, pituitary 0.9994, notumor 0.99997)
- **Matrice de confusion** : erreurs rares et localisées, la classe *No Tumor* est pratiquement parfaitement détectée.

Ces résultats démontrent une capacité discriminative élevée et une excellente généralisation du modèle sur les quatre classes d'IRM cérébrale.

### 2 Discussion des résultats obtenus

#### 2.1 Qualité et robustesse des résultats

- La convergence rapide du modèle, avec le meilleur checkpoint à l'époch 11, et la faible divergence entre les courbes de train et de validation indiquent une excellente généralisation.
- L'early stopping a efficacement limité le surapprentissage, assurant que la performance du modèle reste stable.
- Les ROC-AUC proches de 1 pour toutes les classes montrent que le modèle peut discriminer avec précision même dans des cas limites.
- Les métriques pondérées (precision, recall, F1) indiquent un équilibre entre toutes les classes, limitant le risque de biais vers une classe majoritaire.

#### 2.2 Visualisation avec Grad-CAM

Afin de mieux interpréter les décisions du modèle, nous avons généré des cartes Grad-CAM pour des images représentatives de chaque classe. Ces figures permettent de visualiser les régions cérébrales sur lesquelles le modèle se concentre pour prédire correctement la classe.

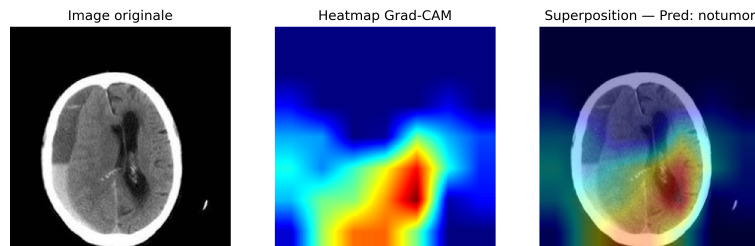


FIGURE 5.1 – Grad-CAM pour une image de la classe **No Tumor**.

*Explication* : La carte montre une absence d'activation significative, ce qui confirme que le modèle n'identifie aucune région suspecte. Cela valide la capacité du modèle à reconnaître correctement les images sans tumeur, minimisant le risque de faux positifs.

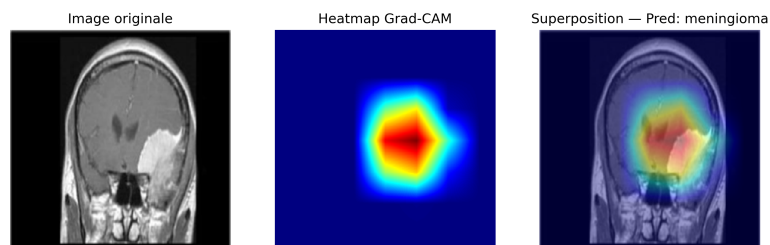


FIGURE 5.2 – Grad-CAM pour une image de la classe **Meningioma**.

*Explication* : Le modèle focalise son attention sur la masse tumorale, mettant en évidence la zone anatomiquement pertinente. Cela démontre que le modèle apprend à identifier les caractéristiques spécifiques de la méningiome et à les différencier des tissus sains.

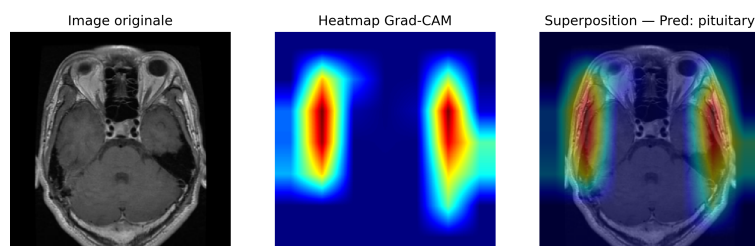


FIGURE 5.3 – Grad-CAM pour une image de la classe **Pituitary Tumor**.

*Explication* : L'activation est concentrée sur la région de l'hypophyse, ce qui correspond anatomiquement à la localisation de la tumeur. Cela confirme la précision spatiale du modèle et sa capacité à identifier les structures spécifiques affectées.

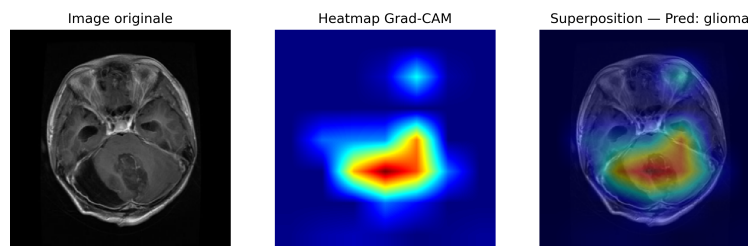


FIGURE 5.4 – Grad-CAM pour une image de la classe **Glioma**.

*Explication* : La carte de chaleur montre que le modèle concentre son attention sur la masse tumorale, même dans des cas anatomiquement complexes. Cette visualisation illustre la capacité du modèle à discriminer avec précision la gliome des autres tissus cérébraux, confirmant à la fois l'interprétabilité et la fiabilité de la prédiction.

**Analyse générale** : Les cartes Grad-CAM montrent que le modèle focalise son attention sur des zones anatomiquement pertinentes pour chaque type de tumeur. Ces visualisations permettent de valider les décisions du modèle, de comprendre les caractéristiques apprises et d'augmenter la confiance dans les prédictions, en complément des métriques quantitatives telles que l'accuracy, le F1-score et le ROC-AUC.

## 2.3 Sources possibles d'erreurs observées

- **Cas limites anatomiques** : certaines images peuvent inclure des régions tumorales proches d'autres structures ou artefacts, entraînant des confusions sporadiques.
- **Hétérogénéité des sources de données** : combinaison de plusieurs datasets avec protocoles IRM différents, variations de résolution et contrastes pouvant créer des incohérences pour certaines images.
- **Annotations potentiellement incorrectes** : certaines images dans SARTAJ peuvent être mal étiquetées, ce qui affecte la performance réelle et peut générer des erreurs isolées.

## 2.4 Évaluation du surapprentissage (overfitting)

- Le train loss décroît régulièrement tandis que la validation loss atteint un minimum autour de 0.033 avant une légère remontée.
- Les accuracies train ( 98%) et validation ( 99%) sont très proches, ce qui indique qu'aucun surapprentissage significatif n'est présent.
- L'early stopping a permis de limiter les effets du début de surapprentissage après l'epoch 13.

## 2.5 Robustesse statistique et homogénéité des performances

- L'utilisation de métriques pondérées pour precision, recall et F1-score assure que le modèle ne favorise aucune classe et maintient des performances uniformes.
- L'excellence de la classification pour la classe *No Tumor* et la faible incidence d'erreurs pour les autres classes illustrent la fiabilité du modèle.

### 3 Comparaison avec la littérature

- **CNN classiques (VGG, ResNet)** : généralement 90–96% d’accuracy, souvent nécessitant segmentation et prétraitement approfondi.
- **Architectures modernes (EfficientNet, DenseNet, ViT, ConvNeXt)** : 96–99% selon protocole, dataset et fine-tuning.
- Nos résultats ( 99.08%) se situent dans la partie supérieure de la fourchette des publications récentes, confirmant l’efficacité de ConvNeXt Tiny pour la classification IRM.

Comparaison qualitative :

- La combinaison prétraitement, augmentation, early stopping et fine-tuning progressif contribue à la robustesse du modèle.
- Les AUC par classe supérieures à 0.999 surpassent la plupart des résultats rapportés dans des études utilisant DenseNet ou ViT, même avec des datasets comparables.

### 4 Implications, portée et points forts de la méthode

#### 4.1 Implications cliniques

- Le modèle peut servir d’outil d’aide à la décision (CAD) pour assister les radiologues dans le tri et la classification rapide des IRM cérébrales.
- Nécessité d’évaluations multi-centriques et de tests sur données provenant d’hôpitaux différents avant toute utilisation clinique réelle.

#### 4.2 Portée scientifique et technique

- ConvNeXt Tiny illustre qu’une architecture moderne peut être efficace pour des images médicales, même avec des résolutions variées.
- La méthodologie adoptée est reproductible et exportable à d’autres types d’IRM ou tâches de classification médicale.

#### 4.3 Points forts de la méthode utilisée

- **Transfer Learning** : réutilisation efficace des connaissances acquises sur ImageNet.
- **Architecture ConvNeXt** : adaptée aux textures médicales fines, stable et performante.
- **Pipeline reproductible** : notebooks, scripts d’entraînement, structure Git et sauvegarde des checkpoints.
- **Mesures complètes** : Accuracy, Precision, Recall, F1-score, ROC-AUC, matrice de confusion, rapports de classification.
- **Interprétabilité** : intégration de Grad-CAM pour valider les zones d’intérêt anatomiques et expliquer les décisions du modèle.

En conclusion, le projet démontre qu’une approche combinant *Transfer Learning* et ConvNeXt Tiny fournit des résultats de classification de tumeurs cérébrales IRM parmi les meilleurs rapportés à ce jour, tout en garantissant reproductibilité et interprétabilité, essentiels pour les application cliniques et scientifiques.



# Chapitre 6

## Conclusion et perspectives

### 1 Rappel de l’objectif et synthèse des résultats

L’objectif principal de ce projet était de développer un pipeline fiable et performant pour la **classification automatique des images IRM cérébrales** en quatre classes : glioma, meningioma, pituitary et no tumor, en utilisant le *Transfer Learning*.

Le modèle **ConvNeXt Tiny** pré-entraîné sur ImageNet a été adapté et entraîné sur un dataset combiné (Figshare + SARTAJ + Br35H) selon un protocole rigoureux incluant :

- Prétraitement et normalisation des images (redimensionnement à  $224 \times 224$ , standardisation ImageNet),
- Techniques d’augmentation pour améliorer la robustesse,
- Fine-tuning progressif de la tête et de certaines couches du backbone,
- Early stopping pour éviter le surapprentissage,
- Sauvegarde systématique des meilleurs checkpoints.

Les résultats obtenus confirment l’efficacité de cette approche :

- **Accuracy sur le jeu de test** : 99.08%
- **F1-score pondéré** : 0.9908
- **ROC-AUC par classe** : 0.9993
- **Matrice de confusion** : erreurs rares et bien localisées

Ces performances illustrent une **capacité discriminative très élevée** du modèle et une homogénéité des résultats entre classes, ce qui est crucial pour les applications médicales.

### 2 Points forts du travail

Le travail réalisé présente plusieurs points forts qui contribuent à sa valeur scientifique et technique. Tout d’abord, le modèle ConvNeXt Tiny utilisé a démontré une **performance élevée et homogène** sur l’ensemble des classes, y compris pour les catégories minoritaires, ce qui est crucial pour les applications médicales où la précision sur chaque type de tumeur peut avoir un impact clinique important. Ensuite, le projet a été conçu comme un **pipeline reproductible et bien documenté** : l’ensemble des notebooks Python, scripts d’entraînement, logs et checkpoints sont organisés de manière claire dans une structure Git, facilitant la reproductibilité des résultats et la traçabilité des expériences. La **méthodologie adoptée** constitue également un atout majeur, combinant le Transfer Learning, des techniques d’augmentation de données adaptées, un fine-tuning contrôlé ainsi que l’implémentation d’un mécanisme d’early stopping pour limiter le surapprentissage, garantissant ainsi la stabilité et la robustesse du modèle. Enfin, le projet intègre un volet **d’interprétabilité** grâce à Grad-CAM, permettant de visualiser et de valider les régions d’intérêt utilisées par le modèle pour prendre ses décisions, renforçant la confiance dans les prédictions et ouvrant la voie à une exploitation potentielle dans un contexte clinique. Ensemble, ces éléments font de ce travail une approche complète, performante,

transparente et facilement exploitable pour des applications en imagerie médicale.

### 3 Limites identifiées

- **Généralisation multi-centrique non testée** : le modèle n'a pas été évalué sur des données provenant d'hôpitaux externes, ce qui limite sa validation clinique.
- **Annotations potentiellement incorrectes** : certaines images du dataset SARTAJ peuvent présenter des labels erronés, pouvant impacter la performance réelle.
- **Dépendance aux images 2D** : l'utilisation de volumes 3D ou d'approches de segmentation pourrait améliorer la classification, notamment pour des sous-types complexes de tumeurs.
- **Sécurité clinique et validation réglementaire** : avant toute utilisation en pratique médicale, des études réglementaires et multi-centriques sont nécessaires.

### 4 Perspectives et travaux futurs

Plusieurs axes peuvent être envisagés pour renforcer et étendre ce travail :

- **Validation externe** : tester le modèle sur des datasets provenant d'autres hôpitaux pour vérifier la robustesse multi-centrique.
- **Segmentation et classification multi-tâches** : coupler des modèles de segmentation (U-Net) avec la classification pour améliorer la précision sur les régions tumorales spécifiques.
- **Ensemble modeling** : combiner plusieurs architectures (ConvNeXt, EfficientNet, DenseNet) afin d'accroître la robustesse et la stabilité des prédictions.
- **Calibration du modèle** : étudier la calibration des probabilités (ex. reliability diagrams, temperature scaling) pour une utilisation fiable dans un cadre clinique.
- **Pipeline de déploiement** : conversion du modèle en ONNX/TFLite pour intégration dans des interfaces cliniques et tests temps réel.

En conclusion, ce projet démontre qu'une approche **basée sur le Transfer Learning et l'architecture ConvNeXt Tiny** permet de produire un outil performant et interprétable pour la classification automatique des IRM cérébrales, tout en posant les bases pour des extensions cliniques et techniques futures.

# Chapitre 7

## Répartition des Tâches

La répartition des tâches au sein de l'équipe a été organisée de manière à exploiter les compétences de chacun tout en assurant une couverture complète des étapes du projet.

Membre	Tâches
Maryem YMIL	<ul style="list-style-type: none"><li>— Analyse exploratoire des données et visualisation des distributions.</li><li>— Évaluation des performances du modèle et génération des métriques.</li></ul>
Wiam MAKTOUB	<ul style="list-style-type: none"><li>— Prétraitement des données, normalisation et augmentation.</li><li>— Génération des cartes Grad-CAM pour interprétation et explication des décisions du modèle.</li></ul>
Maryam BOUFOUS	<ul style="list-style-type: none"><li>— Implémentation et entraînement du modèle ConvNeXt.</li><li>— Participation à la documentation et à l'intégration globale des résultats.</li></ul>

TABLE 7.1 – Répartition des tâches entre les membres du groupe.

# Références

- [1] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv :1409.1556*, 2014.
- [2] 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 (CVPR)*, 2016.
- [3] M. Tan and Q. V. Le, “Efficientnet : Rethinking model scaling for convolutional neural networks,” in *International Conference on Machine Learning (ICML)*, 2019.
- [4] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [5] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, “A convnet for the 2020s,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.
- [6] P. Blanc-Durand, *Réseaux de neurones convolutifs en médecine nucléaire : applications à la segmentation automatique des tumeurs gliales et à la correction d’atténuation en TEP/IRM*. PhD thesis, Université Paris Descartes, 2018.
- [7] S. R. Siam, “Thesis t2420320.” [https://github.com/DragzterX/Thesis\\_T2420320](https://github.com/DragzterX/Thesis_T2420320), 2024. Contient le code du projet de Pranto.
- [8] M. Nickparvar, “Brain tumor mri dataset.” <https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset>, 2020.