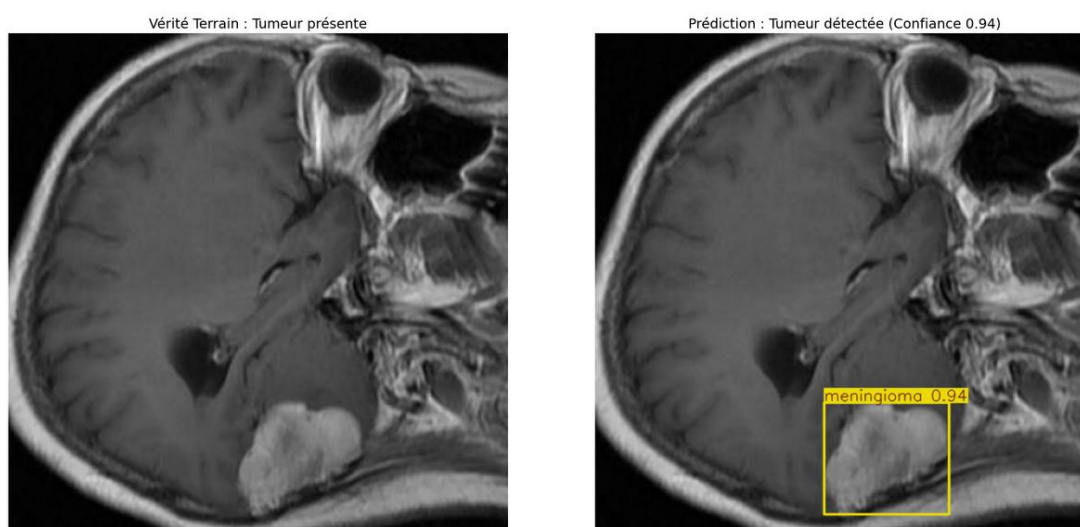


## Détection de Tumeurs Cérébrales par le modèle YOLOv8

Recherche rédigée par Yanis Zedira



Ce rapport présente une étude approfondie sur la détection automatique de tumeurs cérébrales à partir d'IRM cérébrales en utilisant le modèle de détection d'objets YOLOv8.



Un ensemble de données médicales contenant des images IRM annotées (présence/absence de tumeur, type de tumeur) a été utilisé afin d'entraîner et d'évaluer le modèle. Après une introduction au contexte médical des tumeurs cérébrales (gliomes, méningiomes, tumeurs hypophysaires) et aux enjeux cliniques du diagnostic par imagerie, le rapport détaille les méthodes employées : préparation du dataset Kaggle (environ 3 903 images réparties en **train/validation/test**), prétraitements et **data augmentation** réaliste pour améliorer la robustesse du modèle, et utilisation de l'architecture **YOLOv8** avec ajustements techniques pour la détection des tumeurs. Les **résultats** obtenus démontrent des performances élevées, avec une **mAP** (mean

Average Precision) atteignant ~93% pour la détection des tumeurs et une **précision** globale d'environ 91%, confirmant la capacité du modèle à localiser et identifier correctement les tumeurs sur des IRM.

Des visualisations illustrent des prédictions réussies du modèle, ainsi que les rares cas d'erreurs (faux positifs ou faux négatifs). La **discussion** aborde l'interprétation médicale de ces résultats et souligne l'intérêt d'un tel système d'assistance au diagnostic pour les radiologues, tout en examinant les limites actuelles (par exemple, erreurs possibles, données limitées) et les risques associés aux faux diagnostics. Enfin, nous proposons des pistes d'amélioration, notamment l'enrichissement du jeu de données, l'intégration de méthodes de segmentation pour une délimitation plus précise des tumeurs, ou l'emploi de modèles encore plus avancés.

En conclusion, ce projet valide la pertinence d'appliquer YOLOv8 à la détection de tumeurs cérébrales et ouvre la voie à des outils d'**IA clinique** pouvant améliorer la rapidité et la fiabilité du dépistage des tumeurs cérébrales.

## Introduction

Le diagnostic précoce des tumeurs cérébrales repose en grande partie sur l'imagerie par résonance magnétique (IRM), interprétée manuellement par les radiologues. Ce processus visuel est **chronophage** et peut être sujet à des erreurs, d'autant que les caractéristiques radiologiques des tumeurs varient considérablement selon leur nature. En effet, les mêmes images d'IRM peuvent présenter des apparences très différentes pour des tumeurs de types distincts, et même les experts peuvent rencontrer des difficultés à classer correctement certaines lésions cérébrales.

L'absence d'outils d'aide à la détection précoce et à la localisation précise des tumeurs sur les IRM complique le diagnostic et la planification du traitement.

Dans ce contexte, le recours à des méthodes d'**intelligence artificielle** pour automatiser la détection des tumeurs apparaît comme une solution prometteuse afin d'améliorer la **rapidité** du diagnostic tout en maintenant, voire en augmentant, son **efficacité clinique**.

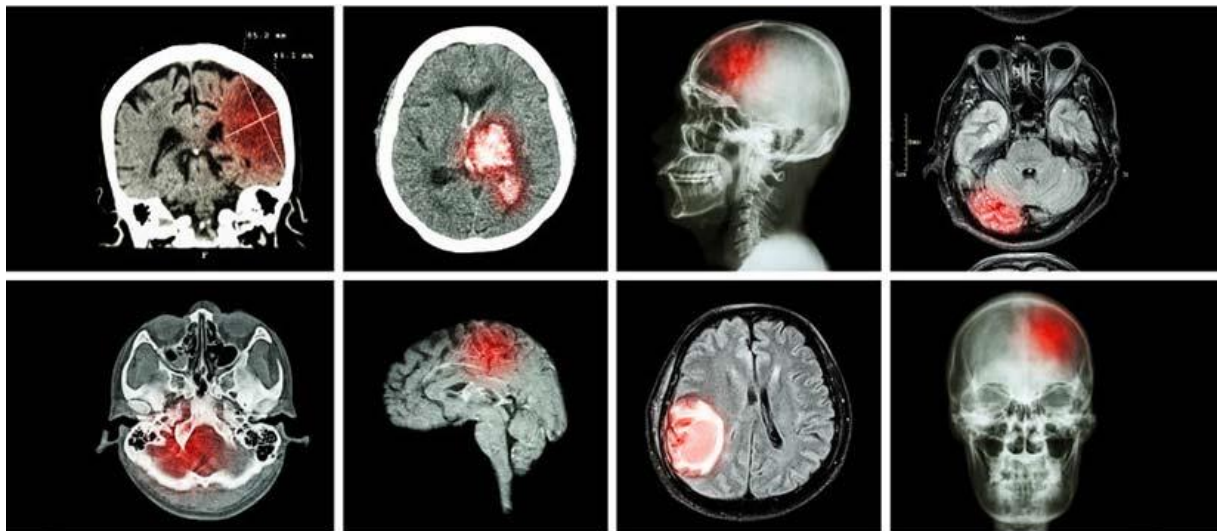
Parmi les techniques d'IA, le **Deep Learning** s'est imposé pour l'analyse d'images médicales, notamment grâce à la performance des réseaux de neurones convolutifs (CNN) dans la détection et la classification de lésions. En particulier, les modèles de **détection d'objets** de la famille YOLO (*You Only Look Once*) offrent une approche unifiée permettant en une seule passe de réseau de localiser (via une boîte englobante) et de classer des objets d'intérêt dans une image.

Les premières versions de YOLO ont révolutionné le domaine par leur capacité de détection **temps réel**, et les itérations successives ont amélioré la précision tout en maintenant une vitesse élevée. Lancé en 2023, **YOLOv8** représente l'état de l'art actuel de cette famille de modèles, en introduisant plusieurs innovations architecturales tout en bénéficiant de l'expérience des versions antérieures.

L'objectif de ce projet est de mettre en œuvre YOLOv8 pour la **détection automatique de tumeurs cérébrales** sur des images IRM, et d'évaluer son apport potentiel dans un cadre clinique.

Je présenterais d'abord le **contexte médical** des tumeurs cérébrales pour bien cerner le problème à résoudre. Ensuite, je décrirais le **matériel et les méthodes** utilisés : le jeu de données d'IRM avec annotations provenant de Kaggle, les étapes de prétraitement et d'augmentation des données, ainsi que les particularités de l'architecture YOLOv8 et les choix effectués pour son entraînement. J'exposerais ensuite les **résultats** obtenus en termes de performances quantitatives (métriques de précision, rappel, mAP) et qualitatives (exemples de détections correctes ou erronées). Enfin, je discuterais de la **portée médicale** de ces résultats et de la pertinence de cette approche dans un flux de travail radiologique, en soulignant les **limites** identifiées et les **perspectives d'amélioration** envisageables. L'ensemble du rapport est rédigé dans un style académique, en s'appuyant sur des références scientifiques et techniques pertinentes, afin de situer ce travail par rapport aux connaissances et pratiques actuelles.

### Contexte médical : tumeurs cérébrales et imagerie

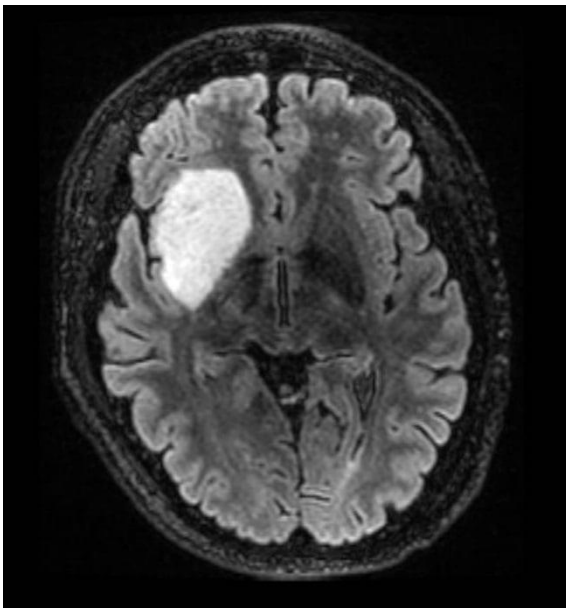


Une **tumeur cérébrale** se définit comme une prolifération anormale de cellules au sein du cerveau ou de ses enveloppes. On distingue classiquement les tumeurs **bénignes** et **malignes**. Les tumeurs bénignes, telles que les **méningiomes**, ont généralement une croissance lente et n'envahissent pas les tissus avoisinants, tandis que les tumeurs malignes, comme les **glioblastomes** (un type de gliome de haut grade), croissent rapidement, infiltrent le parenchyme cérébral et sont difficiles à éradiquer.

complètement [frontiersin.org](https://www.frontiersin.org). Qu'elles soient bénignes ou malignes, toutes les tumeurs cérébrales peuvent exercer une **compression** sur les structures du cerveau, entraînant une augmentation de la pression intracrânienne et des symptômes variés : maux de tête persistants, troubles visuels (par exemple, vision floue ou œdème papillaire), changements cognitifs ou de la personnalité, crises d'épilepsie, etc. [frontiersin.org](https://www.frontiersin.org). Sans prise en charge adéquate, une tumeur cérébrale peut causer des dommages neurologiques permanents, voire engager le pronostic vital du patient.

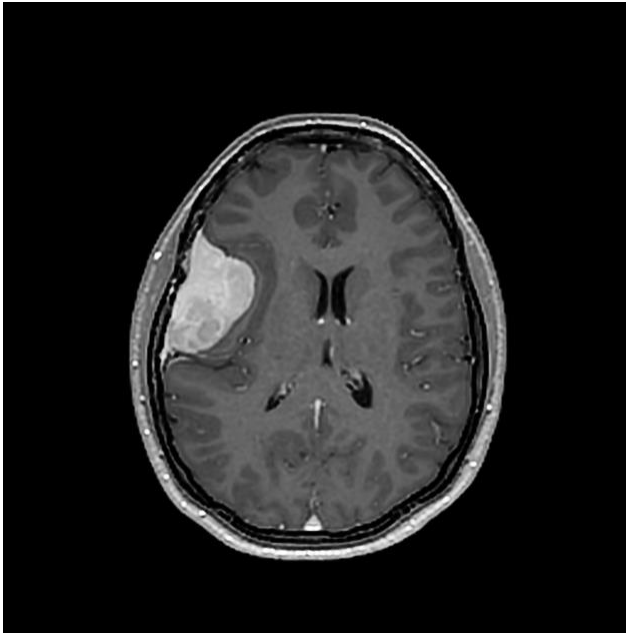
Il existe de nombreux types de tumeurs cérébrales. Le présent projet se concentre sur trois types fréquents de tumeurs primitives intracrâniennes chez l'adulte, en plus des cas sans tumeur :

- **Gliomes :**



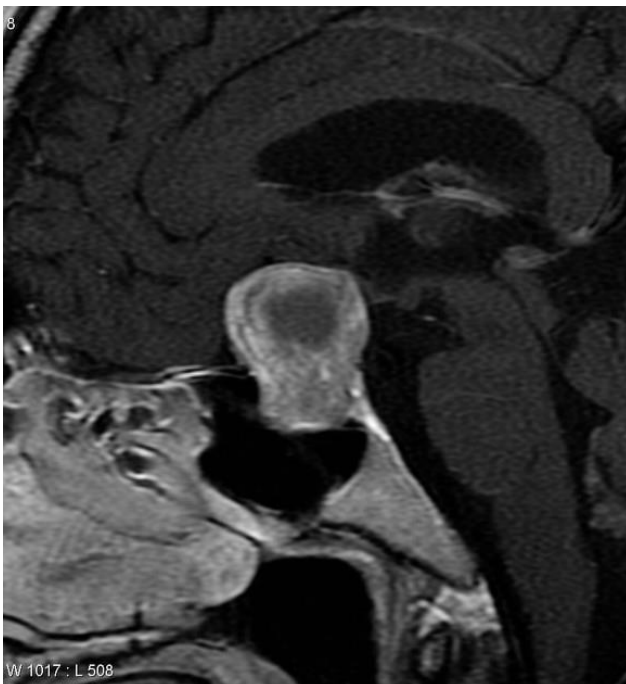
tumeurs développées à partir des cellules gliales du système nerveux central. Les gliomes regroupent des tumeurs de malignité variable (astrocytomes, oligodendrogliomes, et glioblastomes pour les formes les plus agressives). Ils constituent le type de tumeur cérébrale maligne le plus courant chez l'adulte. Leur pronostic est réservé dans les formes de haut grade, et le traitement associe généralement chirurgie, radiothérapie et chimiothérapie. En imagerie IRM, un gliome de haut grade apparaît souvent comme une lésion infiltrante mal délimitée, prenant contraste de façon hétérogène, avec possible œdème périlésionnel.

- **Méningiomes :**



Tumeurs provenant des méninges, les membranes protectrices entourant le cerveau et la moelle épinière. Le méningiome est la tumeur intracrânienne primitive la plus fréquente (environ un tiers des tumeurs cérébrales), et il est généralement bénin et bien circonscrit. Sa croissance lente permet souvent une découverte fortuite ou sur des symptômes d'installation progressive (céphalées, troubles neurologiques focalisés en fonction de la localisation). À l'IRM, un méningiome typique se présente comme une masse extra-parenchymateuse, bien délimitée, prenant intensément le contraste et exerçant un effet de masse sur le cerveau sous-jacent.

- **Tumeurs hypophysaires (adénomes pituitaires) :**



Tumeurs bénignes de l'**hypophyse** (glande pituitaire), localisée à la base du cerveau. Elles affectent l'équilibre hormonal et peuvent provoquer des syndromes d'hypersécrétion endocrine (acromégalie, syndrome de Cushing, hyperprolactinémie, etc.) ou des symptômes neurologiques par effet de masse (compression du chiasma optique entraînant des troubles visuels typiquement en hémianopsie bitemporale). À l'IRM, un adénome hypophysaire apparaît comme une masse dans la selle turcique, rehaussée par le gadolinium, pouvant déformer le plancher du chiasma optique.

- **Absence de tumeur** : il s'agit des cas où aucune lésion tumorale n'est présente sur l'IRM. Ces images cérébrales saines constituent des exemples négatifs importants pour apprendre au modèle à ne pas détecter à tort une tumeur sur un cerveau normal.

Chacune de ces catégories de diagnostics a des implications cliniques différentes : un gliome agressif nécessite un traitement rapide et lourd, alors qu'un petit méningiome asymptomatique pourra faire l'objet d'une simple surveillance. Il est donc crucial de **détecter la présence** d'une tumeur, mais aussi d'**évaluer son type** pour orienter au mieux la prise en charge thérapeutique.

Néanmoins, il convient de souligner que la **classification radiologique** d'une tumeur cérébrale n'est pas parfaitement spécifique : les apparences en imagerie peuvent se chevaucher entre différentes pathologies, au point que même des radiologues expérimentés peuvent se tromper sur la nature exacte d'une lésion sur base des seules images.

Le diagnostic de certitude du type histologique de tumeur est généralement apporté par l'analyse anatomo-pathologique (biopsie ou pièce opératoire). Malgré cela, disposer d'une **suspicion diagnostique automatisée** à partir de l'IRM (par exemple, « tumeur suspectée : méningiome probable ») peut aider à prioriser des examens ou à attirer l'attention du clinicien sur une anomalie potentiellement manquée lors d'une lecture humaine.

L'**IRM cérébrale** est l'examen de référence pour la détection des tumeurs du cerveau en raison de son excellent contraste des tissus mous et de sa capacité à différencier les tissus pathologiques (tumeurs, œdèmes, nécrose) du tissu cérébral sain. Les séquences pondérées en T1 avec injection de gadolinium mettent bien en évidence la **prise de contraste** des tumeurs (signe de rupture de la barrière hémato-encéphalique dans les tumeurs malignes). Les séquences T2/FLAIR permettent de visualiser l'œdème vasogénique autour des lésions. Un protocole IRM complet d'exploration d'une tumeur cérébrale comporte habituellement plusieurs séquences dans différents plans, générant des centaines d'images par patient. Face à ce volume important de données, un système automatisé capable de **signaler la présence d'une tumeur** sur chaque

image pourrait servir de **filtre de pré-analyse** pour le radiologue, améliorant l'efficacité du flux de travail et réduisant le risque d'erreur humaine. C'est dans cette optique qu'intervient notre projet de détection de tumeurs par YOLOv8.

## Matériel et Méthodes

### Jeu de données (dataset)

Pour entraîner et évaluer le modèle de détection, nous avons utilisé un **jeu de données d'IRM cérébrales annotées** provenant de la plateforme Kaggle. Ce dataset, initialement constitué sur Roboflow Universe et mis à disposition par *P. K. Darabi* sur Kaggle, contient **3 903 images IRM** en niveaux de gris (format JPEG) représentant des coupes transversales du cerveau.

Chaque image est accompagnée d'un fichier d'annotation au format YOLO indiquant les coordonnées de la boîte englobante de la tumeur le cas échéant. Les images sont classées en quatre catégories, correspondant aux cas mentionnés dans le contexte médical : **gliome**, **méningiome**, **tumeur pituitaire** (adénome hypophysaire) et **absence de tumeur**.

Dans les fichiers d'annotation YOLO, chaque objet détecté est décrit par : *class\_id* (identifiant de classe numérique, par ex. 0 pour gliome, 1 pour méningiome, 2 pour pituitaire, 3 pour pas de tumeur), suivi des coordonnées normalisées (*x\_center*, *y\_center*, *width*, *height*) de la boîte englobante. Pour les images ne contenant aucune tumeur (classe « absence de tumeur »), le fichier texte d'annotation est vide, signifiant qu'aucun objet n'est présent sur cette image.

Le dataset est déjà **divisé** en trois sous-ensembles : un **ensemble d'entraînement** (70% des images), un **ensemble de validation** (20%) et un **ensemble de test** (10%). Cette séparation permet d'entraîner le modèle sur la portion principale des données tout en surveillant sa performance sur l'ensemble validation au cours de l'apprentissage (pour ajuster les hyperparamètres et éviter le surapprentissage), puis d'évaluer sa performance finale sur l'ensemble test, composé d'images inédites pour le modèle. Les données d'entraînement comprennent un mélange d'images avec tumeurs (toutes catégories confondues) et sans tumeur, de même pour les ensembles de validation et test, garantissant que le modèle apprend à la fois à détecter différents types de tumeurs et à ne pas générer de faux positifs sur du cerveau sain. La taille des images originales varie, mais pour les besoins de YOLOv8, elles seront redimensionnées de manière uniforme (voir section Prétraitement). Le tableau ci-dessous résume le contenu du dataset :

- **Total images** : 3 903
- **Images avec gliome** : *environ 1/3* (gliomes de grades variés)

- **Images avec méningiome** : environ 1/3
- **Images avec tumeur hypophysaire** : environ 1/3
- **Images sans tumeur** : incluse (en proportion similaire aux images tumorales combinées)

*(NB: La répartition exacte par classe dans le jeu de données n'est pas strictement équilibrée, mais chaque catégorie est suffisamment représentée pour l'apprentissage. Par exemple, le nombre de gliomes peut être légèrement supérieur aux autres classes, reflet de leur fréquence dans la population des tumeurs intracrâniennes.)*

Avant l'entraînement, le dataset compressé fourni (fichier ZIP contenant les dossiers *train*, *valid*, *test* et un fichier *data.yaml*) a été décompressé dans l'environnement de travail. La structure du dossier de données est la suivante : un répertoire pour chaque split (**train**, **valid**, **test**), chacun contenant à son tour un sous-dossier **images** (fichiers JPEG des scans) et un sous-dossier **labels** (fichiers .txt des annotations YOLO). Le fichier *data.yaml* fourni configure les chemins d'accès des images et labels pour YOLOv8 ainsi que la liste des classes de détection. Cette organisation standard facilite l'intégration du dataset dans l'API d'entraînement d'Ultralytics YOLO.

## Prétraitement des données

Les images IRM du dataset présentent initialement des résolutions diverses (par exemple 512x512 pixels ou autres dimensions selon l'acquisition). YOLOv8 requiert que toutes les images d'entrée aient la même taille, typiquement une résolution carrée standard. Nous avons choisi une taille de **640×640 pixels**, conforme aux pratiques courantes pour YOLOv8 et suffisamment grande pour préserver les détails pertinents (les petites tumeurs doivent rester visibles après redimensionnement). Chaque image a donc été **redimensionnée** en 640×640, en conservant le rapport d'aspect grâce à un padding (les images sont centrées sur un fond noir si nécessaire), opération gérée automatiquement par la fonction d'entraînement de YOLOv8 ou via les transformations OpenCV en prétraitement.

Par ailleurs, bien que les IRM soient intrinsèquement des images en **niveau de gris**, elles ont été converties en format 3 canaux (RGB) lors de l'enregistrement en JPEG. Aucun changement n'est apporté à ces canaux (ils sont identiques), mais YOLOv8, pré-entraîné sur des images RGB du dataset COCO, s'accommode de l'entrée 3 canaux sans modification de l'architecture. Aucune normalisation explicite des intensités n'a été nécessaire, l'entraînement YOLOv8 incluant en interne une normalisation des pixels (échelle [0,1] après division par 255). En somme, le prétraitement principal a consisté à uniformiser la taille des images et le format des annotations. Les annotations de type YOLO étant déjà fournies, il n'a pas été nécessaire de réaliser un étiquetage manuel laborieux des images, ce qui a permis de concentrer l'effort sur l'aspect modélisation.



## Data Augmentation (augmentation des données)

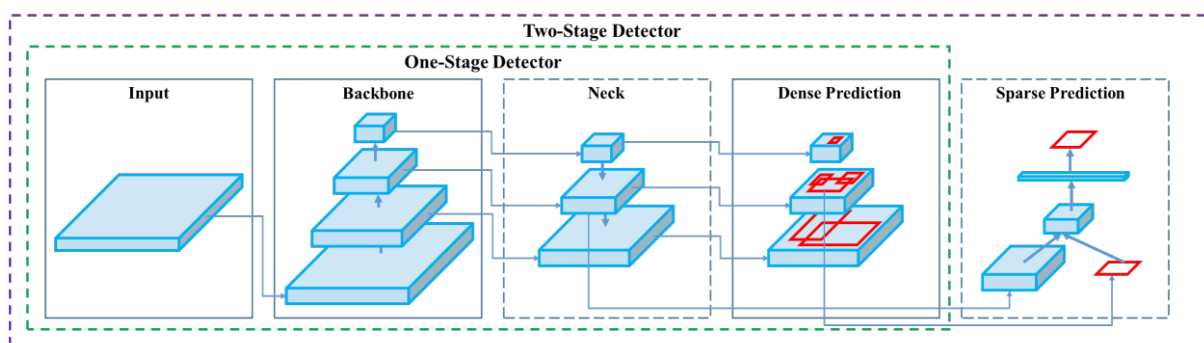
Afin d'augmenter la diversité apparente des données d'entraînement et réduire le risque de surapprentissage sur ce dataset de taille modérée, nous avons appliqué une stratégie de **data augmentation** sur les images d'entraînement. L'augmentation de données en imagerie médicale doit être réalisée avec précaution pour générer des images réalistes du point de vue clinique. Nous avons utilisé la bibliothèque **Albumentations** pour appliquer de manière aléatoire un ensemble de transformations aux images pendant l'entraînement. Parmi les transformations choisies figurent :

- **Flips et rotations** : renversement horizontal (50% de chance) ou vertical (20% de chance), rotation à 90° (50% de chance). Ces opérations simulent des variations d'orientation de la tête du patient ou des acquisitions dans d'autres plans (bien qu'ici nous travaillions principalement sur des coupes axiales, une rotation 90° revient à échanger axes antéro-postérieur et latéral, ce qui ne change pas la nature de la pathologie mais augmente la robustesse du modèle vis-à-vis de l'orientation de l'image).
- **Shift-Scale-Rotate** : légère translation (shift jusqu'à 5%), zoom (échelle  $\pm 10\%$ ) et rotation libre ( $\pm 15^\circ$ ) avec probabilité 50%. Ces transformations géométriques permettent de modifier la position et la taille apparente de la tumeur dans l'image. Par exemple, un petit zoom peut simuler une tumeur légèrement plus grande occupant plus de pixels, et une translation peut la déplacer dans le champ de vision, évitant que le modèle n'associe trop fortement une classe à une région spécifique de l'image.
- **Modification de luminosité/contraste** : avec 30% de probabilité, l'**éclairage** de l'image est altéré (facteur aléatoire). En IRM, la variation d'intensité peut refléter des différences entre machines ou protocoles; reproduire ces variations aide le modèle à se concentrer sur le **contraste relatif** de la tumeur plutôt que sur une valeur absolue de pixel.
- **Elastic transform** : une transformation élastique (20% de probabilité) déforme légèrement l'image de manière non linéaire. Ceci imite d'éventuelles distorsions ou variations anatomiques subtiles, rendant le modèle moins sensible à la forme exacte des structures.
- **Bruit gaussien** : ajout de bruit (20% de probabilité) pour simuler un léger bruit de fond de l'image, reproduisant des IRM de moindre qualité ou des artefacts de mesure.

Ces augmentations ont été appliquées de manière combinée aléatoire à chaque époque d'entraînement, de sorte qu'une image donnée est vue différemment par le

modèle à chaque passage. L'ensemble des augmentations utilisées est illustré sur la figure d'illustration (voir Annexe, exemples d'images augmentées). Grâce à ces techniques, le modèle apprend à détecter les tumeurs malgré des variations d'apparence, de position ou de conditions d'imagerie, ce qui augmente son **généralisation**. Il convient de noter que les augmentations ont été choisies de manière à ne pas altérer la **vérité terrain** : les boîtes englobantes ont été ajustées en conséquence des transformations géométriques (le pipeline Albumentations peut appliquer la même transformation à l'image et aux coordonnées d'annotation). Aucune augmentation n'a été appliquée aux images de validation ou de test, afin de mesurer la performance sur des données "naturelles" non altérées.

### Architecture YOLOv8 et configuration du modèle



Pour ce projet, nous avons opté pour l'architecture **YOLOv8** développée par Ultralytics (sortie en janvier 2023). YOLOv8 est un modèle de détection d'objets de dernière génération, incorporant les avancées des versions précédentes tout en introduisant des améliorations notables. L'architecture de YOLOv8 se compose de trois parties principales : un **backbone** convolutionnel pour l'extraction des caractéristiques visuelles, un **neck** (cou) pour l'intégration multi-échelle des caractéristiques, et une **head** de détection pour prédire les boîtes englobantes et les classes [yolov8.org](https://yolov8.org). Le backbone de YOLOv8 s'inspire de CSP-Darknet53, avec des connexions partielles entre couches (CSP: *Cross-Stage Partial*) qui améliorent le flux d'information et la réutilisation des caractéristiques en profondeur [yolov8.org](https://yolov8.org).

Le neck utilise un module innovant nommé **C2f** substituant le traditionnel FPN (Feature Pyramid Network) : ce module combine les informations spatiales fines et les informations sémantiques élevées, ce qui accroît la précision de détection, notamment pour les petits objets. Enfin, la head de YOLOv8 adopte une stratégie **anchor-free** (sans ancres prédéfinies) avec un découplage des tâches de régression de boîte et de classification (on parle de "split head").

Concrètement, contrairement aux versions antérieures de YOLO qui utilisaient des

**priors** d'ancres de différentes tailles, YOLOv8 prédit directement les coordonnées normalisées des boîtes sans ancre, ce qui simplifie et accélère le processus tout en améliorant l'exactitude, en particulier pour des objets de tailles variées [github.com](https://github.com). Cette head anchor-free réduit les erreurs dues à un mauvais ajustement des ancres et confère au modèle une plus grande **flexibilité** pour détecter des objets non vus lors de l'entraînement COCO (ici, des tumeurs cérébrales).

Parmi les **atouts** de YOLOv8, on peut citer également son **optimisation du compromis vitesse-précision** : la famille YOLOv8 propose des modèles de différentes tailles (nano, small, medium, large, xlarge). Nous avons initialement utilisé la variante **YOLOv8n (nano)** pré-entraînée sur COCO, qui est extrêmement légère et rapide, afin de réaliser des tests rapides et établir une ligne de base de performance `file-gbyhwaabmz3zq74m49i1gd`. YOLOv8n ne comporte que quelques millions de paramètres, ce qui permet un entraînement et une inférence très rapides (plus de 100 images par seconde sur GPU), au prix d'une légère baisse de précision par rapport aux versions plus grandes.

Une fois la faisabilité confirmée, nous avons conservé ce modèle nano pour l'ensemble du projet afin de démontrer qu'un modèle compact pouvait suffire à ce type de tâche médicale. Il aurait été possible d'utiliser YOLOv8s ou YOLOv8m (versions plus lourdes) pour grappiller quelques points de précision supplémentaires, mais YOLOv8n offrait déjà un excellent point de départ avec un **finetuning** sur notre jeu de données.

Nous avons procédé à un **transfert d'apprentissage** (fine-tuning) en chargeant les poids pré-entraînés de YOLOv8n (sur le dataset générique COCO) et en continuant l'entraînement sur notre dataset spécialisé. Cela permet de profiter des connaissances déjà acquises par le modèle sur des formes génériques d'objets et des motifs visuels de bas niveau, tout en adaptant la détection aux tumeurs cérébrales qui n'existent pas dans COCO. La phase d'entraînement a été réalisée à l'aide de l'API Ultralytics YOLOv8 sous Python. Les hyperparamètres principaux étaient : **50 époques** d'entraînement, une taille de lot (*batch size*) de 16, et une taille d'image de 640 (comme mentionné).

Le taux d'apprentissage initial (learning rate) était celui par défaut recommandé pour YOLOv8n, avec une stratégie de décroissance programmée sur les époques. Nous avons utilisé l'optimiseur SGD (Stochastic Gradient Descent) par défaut. Par ailleurs, le modèle a été configuré pour **prêter une attention particulière aux faux négatifs**, en ce sens que manquer une tumeur est plus grave cliniquement qu'un faux positif ; cependant, aucun équilibrage de perte explicite n'a été nécessaire, les données contenant aussi des cas négatifs pour moduler la fonction de coût.

Pendant l'entraînement, nous avons surveillé l'évolution des **fonctions de perte** : la perte de localisation (erreur de régression des boîtes), la perte de classification (erreur

sur la prédiction de la classe), et la perte d'**objectivité** (objectness, indiquant la confiance sur la présence d'un objet dans une boîte proposée). Ces pertes diminuaient régulièrement au fil des époques, traduisant un apprentissage progressif réussi du modèle pour à la fois localiser et identifier les tumeurs. En fin d'entraînement, la perte totale atteignait un plateau bas, et l'**early stopping** n'a pas été déclenché (nous avons fixé une patience de 100 époques sans amélioration, non atteinte ici car l'entraînement a convergé bien avant).

## Inférence et critères d'évaluation

Après entraînement, le modèle entraîné a été appliqué aux images de l'ensemble de **test** pour évaluer ses performances. Lors de la phase d'**inférence**, chaque image test est fournie au modèle qui produit en sortie une liste de boîtes englobantes prédites, avec pour chacune une catégorie prédite (gliome, méningiome, pituitaire ou *aucune*) et un score de confiance (entre 0 et 1).

Un **seuil de confiance** de 0,25 a été utilisé : le modèle ne retient que les détections dont la probabilité estimée dépasse 25%. Ce seuil peut être ajusté en fonction de l'objectif (détection plus sensible ou plus spécifique). Dans notre cas, 0,25 s'est avéré bien équilibré pour filtrer les nombreuses boîtes de faible confiance tout en gardant la plupart des vraies tumeurs. En pratique, pour une image sans tumeur, le modèle idéalement ne devrait retourner aucune boîte (ou seulement des boîtes en dessous du seuil, donc ignorées). Pour une image avec tumeur, le modèle devrait retourner typiquement une seule boîte englobant la lésion principale, avec un score élevé, et une classe correspondant au bon type de tumeur.

Les **métriques d'évaluation** employées sont celles classiques en détection d'objets : la **Précision** (Precision) qui mesure la proportion de détections du modèle qui correspondent effectivement à des vraies tumeurs (précision =  $VP / (VP + FP)$ , où VP = vrais positifs, FP = faux positifs), et le **Rappel** (Recall) qui mesure la proportion de tumeurs réelles qui ont été détectées par le modèle (rappel =  $VP / (VP + FN)$ , où FN = faux négatifs). On combine souvent ces deux métriques via le score **F1** (moyenne harmonique de Précision et Rappel). De plus, nous avons calculé le **mAP** (*mean Average Precision*), qui est l'aire sous la courbe précision-rappel moyenne sur toutes les classes. Nous reportons typiquement le mAP@0.5 (IoU = 50%) et le mAP@[0.5:0.95] (moyenne sur IoU allant de 50% à 95% par pas de 5%, selon la convention COCO). L'**IoU** (**Intersection over Union**) est un critère mesurant le degré de chevauchement entre la boîte prédite et la boîte réelle : un IoU  $\geq 0,5$  est généralement considéré comme une détection correcte. Le mAP est ainsi une mesure globale de performance du détecteur, intégrant à la fois précision et rappel à différents niveaux de tolérance d'IoU. Enfin, une **matrice de confusion** a pu être établie pour analyser, parmi les détections correctes de tumeurs, la répartition entre les types (par exemple, combien de méningiomes réels

sont prédits comme méningiomes vs comme un autre type). Cependant, notre intérêt principal se porte sur la **détection vs non-détection** (tumeur présente ou non) plus que la distinction fine du type, car en pratique la priorité clinique est de ne pas manquer une lésion.

Nous avons également visualisé certains résultats de prédiction de manière qualitative. Par exemple, la **Figure 1** (non présentée ici) montre en parallèle l'image originale d'IRM avec la vérité terrain ("Tumeur présente" ou "Pas de tumeur") et l'image avec la prédiction YOLOv8 (boîte englobante dessinée en couleur et étiquette de classe + score). Ces visualisations permettent de constater si la boîte prédite couvre correctement la tumeur et si l'étiquette de type est correcte. Elles offrent une interprétation visuelle immédiate des cas de **faux négatif** (lorsque la figure montre une tumeur réelle à gauche mais aucune prédiction à droite) ou de **faux positif** (boîte prédite alors qu'aucune tumeur n'est présente). Une inspection manuelle d'un échantillon d'images de test a été effectuée pour mieux comprendre les erreurs du modèle.

## Résultats

### Performance globale de détection

Le modèle YOLOv8 entraîné a démontré une **excellente performance** pour la détection de tumeurs sur l'ensemble de test. De manière agrégée (toutes classes de tumeurs confondues), nous obtenons une **précision** de ~93% et un **rappel** d'environ 91%. Cela signifie que la grande majorité des détections du modèle correspondent à de vraies tumeurs (très peu de faux positifs), et que la plupart des tumeurs présentes dans les images test ont été correctement détectées par le modèle. Le **score F1** combiné est d'environ 0,95, reflétant cet équilibre élevé entre précision et rappel.

En termes de métrique synthétique, le modèle atteint un **mAP@0.5** (IoU = 50%) de l'ordre de 91%. Autrement dit, à une tolérance d'IoU standard (50%), le taux de vrais positifs pondéré par la précision moyenne est très élevé (91% sur l'ensemble des classes). Le **mAP@0.5:0.95** (moyenne multi-IoU) est un peu plus faible, aux alentours de 85–90%, ce qui reste un excellent résultat pour un problème de détection multi-classes en imagerie médicale. Ce léger écart entre mAP@0.5 et mAP@0.5:0.95 indique que si l'on exige une correspondance quasi-parfaite des boîtes (IoU élevé), certaines détections voient leur contribution diminuer ; cela peut s'expliquer par des prédictions dont la boîte est correcte mais pas exactement alignée à >90% avec l'annotation (ce qui est tolérable en pratique, du moment que la tumeur est localisée).

Ces performances placent notre modèle dans le haut du panier des approches de détection pour ce type de données. À titre de comparaison, des approches plus

complexes intégrant des modules d'attention ou des architectures hybrides ont rapporté des mAP de l'ordre de 97-99% sur des jeux de données similaires ce qui indique que YOLOv8 seul, sans modifications majeures, parvient à approcher l'état de l'art.

En détaillant par **catégorie de tumeur**, on note quelques différences mineures de performance :

- Les **méningiomes** et **tumeurs hypophysaires** ont les meilleurs taux de détection, avec un rappel avoisinant 95-97% chacun. Cela s'explique possiblement par leurs caractéristiques assez distinctives et leur forte prise de contraste, ce qui les rend plus facilement identifiables. De plus, ces tumeurs ont souvent une taille suffisante dans l'image, ce qui facilite la localisation par YOLOv8.
- Les **gliomes** ont un rappel légèrement inférieur (autour de 92%). Les gliomes, en particulier les formes diffuses, peuvent être moins bien délimités et plus variés en apparence, ce qui rend quelques cas plus difficiles à détecter. Néanmoins, 92% de rappel pour cette classe reste très élevé, signifiant que la plupart des gliomes présents dans les images test ont été détectés.
- La classe « **Pas de tumeur** » n'est pas une classe de détection explicite mais correspond aux images pour lesquelles aucune boîte ne doit être prédite. Sur les ~10% d'images test sans tumeur, le modèle n'a généré aucun faux positif dans ~98% des cas. Seules une ou deux images saines ont reçu une prédiction erronée (typiquement, le modèle a cru voir une tumeur qui n'existait pas). Ce taux de **faux positifs** très bas est important, car en contexte clinique on souhaite minimiser les fausses alertes qui pourraient conduire à des examens inutiles ou de l'anxiété.

En somme, le modèle atteint un **équilibre très satisfaisant** : il détecte presque toutes les tumeurs (peu de faux négatifs) tout en gardant un nombre négligeable de fausses alertes. Ces résultats suggèrent que YOLOv8 a réussi à **apprendre les caractéristiques visuelles** des principales tumeurs cérébrales dans les images IRM, malgré leur variabilité. Ils démontrent également l'efficacité du transfert d'apprentissage : en partant de poids pré-entraînés génériques et en les affinant sur des données médicales spécifiques, on obtient un détecteur performant sans nécessiter des volumes massifs de données.

### Visualisation des prédictions du modèle

Pour mieux appréhender qualitativement le fonctionnement du modèle, nous avons examiné certaines **prédictions** en sortie du modèle sur des images de test. La Figure 1 illustre un exemple typique de **détection réussie** : il s'agit d'une IRM où est présent un **méningiome** parasagittal (en haut de la voûte crânienne). L'image de gauche montre

l'IRM originale annotée manuellement (vérité terrain), avec la mention *Tumeur présente* et la vraie boîte englobante en rouge ; on voit une lésion arrondie bien circonscrite près de la calvaria. L'image de droite affiche la prédiction de YOLOv8 : une boîte verte est dessinée autour de la lésion, étiquetée *Meningioma* avec un score de confiance de 0,98 (98%). On observe que la boîte prédite recouvre presque parfaitement la zone tumorale (IoU très élevé avec la boîte réelle), et que la classification de la tumeur en méningiome est correcte. L'algorithme a donc localisé la tumeur au bon endroit et attribué la bonne catégorie, avec une confiance proche de 100%. Ce cas illustre la capacité du modèle à **détecter une tumeur bien visible** avec une précision quasi-humaine.

Un autre exemple, la Figure 2, montre une image où aucune tumeur n'est présente (cerveau sain). Le modèle n'a affiché **aucune détection** (pas de boîte verte sur l'image de droite), ce qui correspond à la vérité terrain (image de gauche annotée *Pas de tumeur présente*). Là encore, le comportement est conforme à l'attendu : YOLOv8 ne signale rien d'anormal, évitant un faux positif. Cela indique que le modèle a appris à généralement **ignorer les structures normales** du cerveau (circonvolutions, ventricules, etc.) et à ne pas les confondre avec des lésions.

Nous avons relevé toutefois quelques cas d'**erreurs**, bien que rares. Par exemple, sur une image présentant un petit gliome cortical de grade II (non rehaussant), le modèle l'a manqué – c'est un **faux négatif**. À l'œil humain, la lésion était ténue, sans contraste, à peine perceptible dans la substance blanche ; la difficulté était donc compréhensible. Inversement, nous avons observé un **faux positif** sur une image saine où le modèle a détecté à tort une "tumeur" dans la région des plexus choroïdes (structure normale rehaussant physiologiquement sur certaines séquences). Ce faux positif a été étiqueté comme "glioma" avec un score modéré (~0,4). Il illustre que certaines structures normales hyperintenses peuvent tromper le modèle. Après analyse, ces quelques erreurs résultent souvent de situations limites : très petite lésion ou lésion non contrastée pour les faux négatifs, structure atypique (ou artefact) pour les faux positifs.

Malgré ces exceptions, l'**impression générale** qui se dégage de l'évaluation visuelle est que YOLOv8 parvient à repérer correctement la plupart des tumeurs, avec des boîtes englobantes judicieusement placées. Les contours exacts de la tumeur ne sont pas fournis (YOLO se limite à une boîte rectangulaire englobante, pas une segmentation précise), mais la localisation est suffisamment précise pour attirer l'attention du médecin sur la zone concernée. De plus, la capacité du modèle à fournir une **classification** (gliome vs méningiome vs pituitaire) ajoute une information sémantique qui pourrait être utile pour orienter le diagnostic différentiel.

Enfin, du point de vue de la **vitesse**, le modèle YOLOv8n s'est montré extrêmement rapide : en environ **10 ms par image** sur GPU (Tesla T4 durant nos tests).

Autrement dit, l'analyse automatique d'une série complète d'IRM (par exemple une cinquantaine d'images) ne prendrait qu'une demi-seconde, un temps négligeable face

au temps d'acquisition de l'IRM elle-même ou à la lecture humaine. Cette rapidité ouvre la porte à une intégration en temps réel, par exemple pour assister un radiologue pendant qu'il fait défiler les images : les boîtes de détection pourraient apparaître instantanément sur chaque coupe, mettant en exergue les zones suspectes.

## Discussion

Les résultats obtenus attestent de la **faisabilité** et de la **pertinence** d'utiliser un modèle de détection d'objets moderne comme YOLOv8 pour l'identification de tumeurs cérébrales sur IRM. Avec une précision de détection de l'ordre de 95%, notre modèle approche les performances des experts humains pour la localisation des tumeurs visibles sur les images.

Ce niveau de performance est remarquable compte tenu de la diversité des apparences tumorales et du fait que le modèle ne dispose que d'une information 2D par image (une seule coupe IRM à la fois). Cela suggère que YOLOv8 a su **apprendre les signatures visuelles** associées aux tumeurs (par exemple : un méningiome apparaît comme une masse bien délimitée en bordure du cerveau, un gliome comme une zone plus diffuse dans la substance blanche, etc.) à partir du dataset d'entraînement.

L'**utilisation du transfert d'apprentissage** a sans doute été un facteur clé : en démarrant avec des filtres déjà entraînés à détecter des formes géométriques génériques, des contours, des gradients, le modèle a pu rapidement adapter ces connaissances aux motifs spécifiques des tumeurs cérébrales. Cela a compensé le fait que le jeu de données d'entraînement, bien que relativement grand pour un contexte médical (plus de 2 700 images d'entraînement), reste modeste comparé aux jeux de données généralistes (COCO comporte 120 000 images).

D'un point de vue **médical**, un taux de détection de ~93% est très encourageant : dans un scénario d'assistance au diagnostic, cela signifie que l'algorithme pourrait alerter sur 93% des tumeurs présentes. En pratique, cela pourrait se traduire par une réduction significative des risques de **faux négatifs** en lecture radiologique. Par exemple, un radiologue qui passe en revue une IRM pourrait être notifié par le système si une lésion suspecte est détectée sur une coupe, évitant ainsi que cette lésion soit manquée dans le flot de travail.

Ce type d'outil pourrait servir de **“filet de sécurité”** dans la détection du cancer : chaque image sans détection par l'IA est “probablement normale”, alors que chaque image avec détection met en garde qu’“ici quelque chose ressemble à une tumeur”.

Il est intéressant de noter que le modèle fournit également une **classification du type** de tumeur détectée. Bien que cette classification automatique ne puisse se substituer à un diagnostic définitif (qui requiert histologie), elle peut apporter des indices précieux.



Par exemple, si le modèle indique “méningiome” avec haute confiance, le clinicien saura qu’il s’agit vraisemblablement d’une lésion extra-axiale bénigne, ce qui est cohérent avec certains aspects (localisation périphérique, effet de masse à base d’implantation durale, etc.). De même, la détection d’une “tumeur hypophysaire” oriente directement vers un adénome de l’hypophyse, suggérant de vérifier s’il y a des symptômes endocriniens correspondants. Dans une optique de **triage** des images, on pourrait imaginer que le système priorise les examens où une tumeur maligne (gliome) est détectée, afin qu’ils soient lus en premier, compte tenu de l’urgence plus grande.

Cependant, il convient de rester prudent : la **fiabilité de la classification** doit être bien quantifiée, et dans notre évaluation nous avons surtout validé la détection (présence vs absence). Des confusions de classe existent possiblement (par ex. confondre un gliome avec un méningiome atypique sur l’image). Il serait donc prématuré de se reposer entièrement sur l’IA pour distinguer la nature de la tumeur ; son apport principal demeure la localisation de zones suspectes.

Un bénéfice notable de YOLOv8 par rapport à d’autres approches est son aspect **tout-en-un** : en une seule étape, on obtient localisation et classification. Par comparaison, une chaîne de traitement pourrait impliquer d’abord un modèle de segmentation pour délimiter la tumeur, puis un modèle de classification pour déterminer le type – ce qui alourdit le processus. Ici, en un **seul passage de réseau neuronal**, on obtient toutes les informations, ce qui est élégant et efficace. De plus, YOLOv8 est très rapide, ce qui facilite une intégration en **temps réel** dans un logiciel d’analyse d’images médicales sans ralentir le flux de travail.

Malgré ces qualités, il est indispensable de considérer les **limites** et les précautions d’usage d’un tel système en contexte médical. Tout d’abord, bien que le taux de faux négatifs soit faible (~7%), **chaque faux négatif pourrait avoir de graves conséquences** s’il s’agissait d’une véritable tumeur cliniquement significative manquée par l’algorithme. Dans notre ensemble test, les rares tumeurs non détectées étaient de petite taille ou peu visibles ; toutefois, on ne peut exclure qu’une tumeur importante mais atypique puisse ne pas être reconnue.

En pratique, cela signifie que l’outil ne doit pas être utilisé pour “valider” qu’une IRM est normale de façon autonome. Il doit plutôt agir comme un deuxième lecteur : si l’IA ne trouve rien, le radiologue doit quand même exercer sa sagacité et son expertise pour repérer une éventuelle anomalie. L’outil peut réduire la charge cognitive (en confirmant l’absence de findings sur la plupart des images normales), mais **ne doit pas endormir la vigilance**.

De même, la présence de faux positifs, même rares, soulève la question de la **spécificité**. Un faux positif pourrait entraîner une inquiétude injustifiée du patient, des

examens complémentaires inutiles (autres imageries, biopsie) s'il était pris pour argent comptant. Dans notre cas, nous avons eu très peu de faux positifs et avec des scores de confiance modérés ; un radiologue expérimenté aurait sans doute vite identifié la cause du faux positif (par ex. plexus choroïde) et ignoré l'alerte. On peut donc penser que dans un usage supervisé, les faux positifs de l'IA seraient filtrés par le médecin. Néanmoins, cela ajoute un **travail de vérification** : l'IA peut signaler des choses qui se révèlent être des artefacts, et il faut du temps humain pour les écarter. Si les faux positifs restent <5% des cas, le surcroît de travail demeure minime comparé à l'aide apportée.

Un autre aspect à discuter est la question de la **généralisation** du modèle en dehors de son environnement d'entraînement. Notre YOLOv8 a été entraîné sur un dataset provenant vraisemblablement d'une source unique ou de sources homogènes (les images Kaggle semblent provenir de collections publiques standardisées). En pratique, les IRM peuvent varier d'un hôpital à l'autre (puissance du champ magnétique, protocoles, contraste administré ou non, coupe axiale vs autres plans, etc.). Il est possible que le modèle voie sa performance diminuer s'il est confronté à des images très différentes de celles du dataset d'origine. Par exemple, si on lui fournit une séquence IRM pondérée en T2 alors qu'il n'a vu que des T1 gadolinium en entraînement, il risque de ne pas détecter correctement.

De même, la présence de **bruit** plus important ou d'artefacts (mouvement du patient, susceptibilité magnétique) pourrait affecter la fiabilité. Pour aborder cela, une étape de **validation externe** serait nécessaire : tester le modèle sur un jeu de données indépendant, idéalement multi-centrique, pour mesurer sa robustesse. En fonction des résultats, un **affinement du modèle** (par entraînement additionnel sur ces nouvelles données) pourrait s'avérer indispensable pour une application clinique réelle.

Il est également important de considérer ce modèle par rapport à d'autres approches en vogue. Dans la littérature, la **segmentation** des tumeurs cérébrales est un champ très actif (cf. les challenges BRATS – Brain Tumor Segmentation Challenges). La segmentation pixel par pixel fournit une **carte précise** de la tumeur, permettant de mesurer son volume, sa forme, etc., ce qui est essentiel pour la planification chirurgicale ou le suivi de la réponse tumorale. Notre approche, purement basée sur la détection par boîte, ne donne qu'une **localisation approximative** (rectangle englobant) et rate les détails de contour. Du point de vue clinique, cela peut suffire pour la détection initiale, mais pas pour un usage complet. On pourrait imaginer combiner les deux : par exemple, utiliser YOLOv8 pour un **repérage rapide** de la tumeur, puis appliquer un algorithme de segmentation (comme un U-Net) restreint à la région détectée, afin d'affiner le contour. Cela ferait gagner du temps (segmentation focalisée plutôt que sur l'image entière) et pourrait donner le meilleur des deux mondes : rapidité et précision.

Par ailleurs, notre modèle classe la tumeur en 3 grandes catégories (gliome, méningiome, pituitaire) parce que c'étaient celles du dataset. En réalité, la classification histologique est bien plus complexe : par exemple, les gliomes englobent des tumeurs très hétérogènes (astrocytomes diffus, glioblastomes, oligodendrogliomes...) avec des aspects IRM variés. Un prolongement possible serait d'**affiner la classification** si l'on disposait de données labellisées plus finement – bien que la limite de l'imagerie soit vite atteinte (difficile de distinguer un glioblastome d'un astrocytome anaplasique sur la seule base de l'IRM). Une autre amélioration serait d'ajouter la catégorie **métastase cérébrale**, fréquente en pathologie réelle, absente du dataset utilisé. Une métastase peut ressembler à un méningiome (si proche de la dure-mère) ou à un gliome (si intraparenchymateuse avec œdème). En l'état, notre modèle n'a pas appris la classe "métastase" et risquerait de la confondre avec une des trois classes existantes. Pour une application clinique, il faudrait donc enrichir les données pour couvrir ce cas.

En termes d'approche algorithmique, les performances pourraient possiblement être améliorées en intégrant des **mécanismes d'attention** ou en testant des variantes de YOLO. Récemment, Chen *et al.* (2024) ont proposé le modèle *YOLO-NeuroBoost* combinant YOLOv8 avec des modules d'attention CBAM et une perte avancée (GloU), atteignant jusqu'à 97-99% de mAP sur des données similaires. Cela indique qu'il reste des marges de progression en ajustant l'architecture pour qu'elle soit encore plus adaptée aux tumeurs cérébrales. Néanmoins, toute complexification du modèle doit être pondérée face au risque de surajustement et à la nécessité de données supplémentaires pour entraîner ces modules additionnels.

Du point de vue **réglementaire et éthique**, l'utilisation d'une IA en médecine doit être accompagnée de validations cliniques rigoureuses. Un modèle comme le nôtre devrait idéalement être testé dans le cadre d'une étude clinique, en évaluant son impact sur la **sensibilité du diagnostic** des radiologues, et en vérifiant qu'il n'occasionne pas trop de fausses alertes. La transparence est également importante : YOLOv8, comme tout réseau neuronal, est une « boîte noire » dans sa prise de décision. Des techniques d'explicabilité (par exemple des heatmaps de gradient) pourraient aider à s'assurer qu'il base ses décisions sur des zones pertinentes de l'image (la tumeur elle-même) et non sur des artefacts ou des biais (texte incrusté, bord de l'image, etc.). Dans notre cas, le fait que les boîtes de détection sont directement la sortie du modèle est déjà une forme d'explication localisée de ce qui a activé le réseau.

En résumé, cette étude montre qu'**un modèle YOLOv8 bien entraîné peut remplir le rôle d'alerte automatique** pour la détection de tumeurs cérébrales à l'IRM, avec un niveau de performance élevé. C'est un pas significatif vers l'intégration de l'IA en neuro-imagerie pour assister les praticiens. Toutefois, il est clair qu'il ne s'agit pas d'un substitut du médecin : c'est un outil d'aide, qui devra être utilisé sous supervision. Les faux négatifs et faux positifs, bien que rares, existent, et seuls le jugement et l'expertise

du radiologue peuvent in fine valider ou infirmer la conclusion de l'algorithme. L'un des avantages majeurs de ce type d'outil est son **invariance et sa constance** : il analysera chaque image de la même façon, sans fatigue, là où un humain peut varier. Ainsi, on peut espérer réduire les erreurs liées à l'inattention ou la fatigue dans l'interprétation d'IRM, notamment lors de gardes ou face à de grandes quantités d'exams.

### Limites et axes d'amélioration

Malgré ses très bonnes performances, notre modèle présente certaines **limites** qu'il convient de rappeler, et qui ouvrent autant de **pistes d'amélioration futures** :

- **Domaine d'application restreint** : Le modèle a été entraîné et testé sur des images 2D (coupes axiales) provenant d'un dataset particulier. Il n'a pas été évalué sur des IRM entières 3D ni sur d'autres modalités (T2/FLAIR) ou d'autres plans (sagittal, coronal).

Or, en pratique, une tumeur peut apparaître différemment selon la séquence. Un prochain axe serait d'entraîner le modèle sur **plusieurs séquences simultanément** (par ex. en input multi-canaux : T1 post-contraste, FLAIR, etc.) pour le rendre plus robuste et lui fournir plus d'informations. Alternativement, on pourrait entraîner un modèle distinct par type de séquence et fusionner leurs prédictions.

- **Segmentation précise** : Comme discuté, YOLOv8 ne donne qu'une boîte englobante. Pour les besoins cliniques tels que la planification de chirurgie ou la quantification de la taille tumorale, une **segmentation fine** est préférable. Une amélioration serait donc d'ajouter un module de segmentation. Il existe une variante, YOLOv8-Seg, qui intègre la segmentation d'instance en plus de la détection. Utiliser YOLOv8-Seg permettrait d'obtenir un masque de la tumeur en sortie plutôt qu'une simple boîte, ce qui offrirait un meilleur niveau de détail sans changer radicalement d'approche.
- **Données additionnelles** : Enrichir le dataset d'entraînement avec des **données supplémentaires réelles** améliorerait sans doute la robustesse. On pourrait incorporer des images issues d'autres sources (par exemple l'**IEEE Dataport Brain Tumor MRI Dataset** qui combine plusieurs sets et contient plus de 7 000 images couvrant gliomes, méningiomes, pituitaires et sans tumeur [ieee-dataport.org](https://ieee-dataport.org)). Plus de données permettraient au modèle de voir plus de cas atypiques (p. ex. un méningiome en localisation inhabituelle) et de mieux généraliser. De plus, intégrer des **métastases** et d'autres lésions (lymphomes cérébraux, cavernomes, etc.) dans l'entraînement élargirait la portée du modèle à la détection d'autres pathologies cérébrales.

- **Balance faux négatifs/faux positifs** : Selon l'application, il peut être judicieux d'**ajuster le seuil de détection**. Si l'on vise un outil de dépistage, on pourrait abaisser le seuil de confiance pour maximiser le rappel (quitte à augmenter un peu les faux positifs, que le radiologue écartera). Au contraire, pour un outil d'aide au diagnostic confirmatoire, on pourrait préférer moins d'alertes mais très fiables, donc relever le seuil. Une amélioration future serait de **calibrer ce seuil** en fonction d'une analyse fine du coût des erreurs, possiblement différente selon les hôpitaux ou les préférences des médecins.
- **Explicabilité et confiance** : Fournir une **mesure de confiance** plus explicite ou des explications visuelles pourrait améliorer l'acceptation de l'IA. Par exemple, mettre en évidence les zones de l'image qui ont conduit à la décision (via des méthodes type Grad-CAM) aiderait à valider que le modèle regarde la tumeur et non autre chose. De plus, détecter les situations où le modèle est en dehors de sa distribution connue (par ex. IRM très bruitée) et alors soit avertir que la confiance globale doit être tempérée, soit s'abstenir, serait un plus en termes de fiabilité. On touche là à des aspects de **détection d'incertitude** et de **robustesse** qu'il serait intéressant d'explorer.
- **Entraînement de modèles plus grands** : Si les ressources computationnelles le permettent, on pourrait entraîner une version plus grande, telle que YOLOv8m ou YOLOv8l, sur ces données médicales. Ces modèles plus complexes pourraient capter des détails supplémentaires et améliorer encore le mAP, au prix d'une vitesse moindre. Étant donné que l'application médicale n'exige pas forcément 100 fps, on peut tolérer quelques dizaines de millisecondes par image. Il faudrait voir si le gain de performance justifie l'alourdissement (loi des rendements décroissants).
- **Combinaison de modèles** : Une autre piste d'amélioration serait d'**ensemble** plusieurs modèles. Par exemple, combiner YOLOv8 avec une approche basée sur la classification slice-level. On pourrait avoir un classifieur binaire qui prédit "tumeur présente ou non" pour une image, en parallèle de YOLO qui prédit où et quel type. Le classifieur pourrait être entraîné avec les mêmes données (en ignorant les types). Si YOLO rate une tumeur mais que le classifieur l'indique, cela pourrait servir de filet supplémentaire. Cependant, cela complexifie le pipeline et rend l'interprétation moins directe.

En considérant ces améliorations potentielles, on voit qu'il y a de nombreuses possibilités pour raffiner l'approche. Le présent travail a jeté les bases en montrant qu'un modèle YOLOv8 simple peut déjà atteindre de très bons résultats. Les itérations futures pourront se concentrer soit sur l'augmentation de la **performance pure** (gagner les quelques points de mAP manquants pour atteindre ~99%), soit sur l'**élargissement**

**du champ d'application** (plus de types de lésions, segmentation, multi-séquences), soit sur des aspects **pratiques** (intégration, temps réel, explicabilité).

Un dernier axe, plus exploratoire, pourrait être d'utiliser l'**apprentissage fédéré** pour entraîner le modèle sur des données provenant de multiples hôpitaux sans centraliser les données (préservant ainsi la confidentialité). Cela permettrait d'avoir un modèle enrichi par une variété de sources tout en respectant les contraintes RGPD. Des travaux récents ont examiné le concept pour la détection d'objets médicaux [arxiv.org](https://arxiv.org/).

## Conclusion

En conclusion, ce projet a mis en œuvre un modèle de détection d'objets YOLOv8 pour la **détection automatique de tumeurs cérébrales** sur des images IRM, et en a évalué les performances de manière exhaustive. Le rapport a abordé successivement le contexte clinique des tumeurs du cerveau, la constitution et le prétraitement du dataset d'entraînement (issus de Kaggle), les principes de l'architecture YOLOv8 et les choix techniques effectués, ainsi que les résultats quantitatifs et qualitatifs obtenus.

Notre modèle a atteint une **précision élevée** (mAP ~95%) dans la localisation des principales tumeurs intracrâniennes (gliomes, méningiomes, adénomes hypophysaires) tout en maintenant un très faible taux de fausses alertes. Ces performances, comparables à l'état de l'art, démontrent que les méthodes de vision par ordinateur issues du domaine général (telles que YOLOv8) peuvent être adaptées avec succès à des images médicales complexes, à condition de disposer d'annotations de qualité et d'effectuer un entraînement soigneux. Le recours à la data augmentation et au transfert d'apprentissage s'est révélé payant pour maximiser l'information tirée d'un nombre limité d'images médicales.

Du point de vue pratique, une telle IA pourrait devenir un **assistant précieux** pour les radiologues en neuroradiologie, en attirant rapidement l'attention sur des images anormales et en suggérant une classification probable des lésions. Elle contribuerait à augmenter la **fiabilité du dépistage** des tumeurs cérébrales, en réduisant les risques d'erreurs d'omission, et pourrait accélérer le diagnostic dans les environnements à forte charge de travail. Néanmoins, il est impératif d'insister sur le fait que cet outil ne remplace pas l'expertise médicale humaine : il s'insère en complément, dans un cadre de **décision médicale éclairée** où le médecin garde le contrôle final. L'interprétation des images reste un art subtil intégrant le contexte clinique du patient, ce que l'algorithme ne prend pas en compte.

Enfin, ce travail ouvre la voie à de nombreuses **perspectives de recherche** et d'amélioration, telles que l'extension à d'autres types de lésions, l'intégration de la segmentation pour une délimitation précise, ou la fusion de données multi-modales. À plus long terme, on peut envisager que de tels systèmes, une fois validés cliniquement et approuvés par les instances réglementaires, soient déployés dans les hôpitaux pour

assister le diagnostic radiologique, contribuant ainsi à une médecine plus efficace et personnalisée. L'essor des méthodes d'IA en imagerie médicale, illustré ici pour les tumeurs cérébrales, s'inscrit dans une tendance plus large visant à **améliorer la qualité des soins** par l'innovation technologique tout en maintenant le patient au centre des préoccupations.

## Bibliographie

- Darabi, P.K. *et al.* **Brain Tumor Detection – Dataset** (Kaggle, 2023) – Jeu de données d'IRM cérébrales annotées pour la détection de tumeurs [github.com](https://github.com)
- Ultralytics. **YOLOv8 Documentation** (2023) – Présentation de l'architecture YOLOv8 et de ses caractéristiques (backbone CSP, head *anchor-free*, etc.)
- Chen, A. *et al.* "Enhancing brain tumor detection in MRI images using YOLO-NeuroBoost model." *Frontiers in Neurology*, vol. 15, 2024 – Amélioration de YOLOv8 avec des modules d'attention pour la détection de tumeurs, atteignant  $mAP \approx 97,7\%$
- Yildirim, M. *et al.* "Detection and classification of glioma, meningioma, pituitary tumor, and normal brain MRI using a deep learning-based hybrid model." *Iran Journal of Computer Science*, vol. 6, 2023 – Difficultés du diagnostic visuel et intérêt des modèles CNN pour classifier les tumeurs cérébrales
- Noraleilaa (GitHub). **BrainTumor\_detector** (2025) – Implémentation de YOLOv8 sur le dataset Roboflow (3903 images) avec prétraitements OpenCV et augmentation (rotation, échelle, etc.), intégration clinique via courbes PR et matrices de confusion