

I. CHAPITRE 1 : Dataset de Sclérose en plaques (SEP)

I. Analyse Exploratoire des Données (EDA) :

a) Chargement des données :

Sclérose en plaques (SEP) : Ce dataset contient des images IRM cérébrales de patients atteints de sclérose en plaques. L'objectif de la segmentation est d'identifier les lésions de sclérose en plaques sur les images. Dataset SEP sur ScienceDirect

b) Exploration des caractéristiques des images :

```
Liste des fichiers .nii dans le dataset :  
['1-LesionSeg-T1.nii', '1-LesionSeg-Flair.nii', '1-Flair.nii', '1-T1.nii', '1-T2.nii', '1-LesionSeg-T2.nii']  
  
Taille des images :  
Flair Image Shape: (256, 256, 23)  
T1 Image Shape: (512, 512, 19)  
T2 Image Shape: (256, 256, 19)  
Lesion Segmentation (Flair) Shape: (256, 256, 23)
```

1-LesionSeg-T1.nii: Image de segmentation des lésions basée sur les données T1. Cette image a une taille de (512, 512, 19), indiquant qu'elle est tridimensionnelle avec une résolution de 512 pixels dans la première dimension, 512 pixels dans la deuxième dimension et 19 coupes dans la troisième dimension.

1-LesionSeg-Flair.nii: Image de segmentation des lésions basée sur les données Flair. La taille de cette image est (256, 256, 23).

1-Flair.nii: Image Flair brute sans segmentation. Sa taille est (256, 256, 23), indiquant qu'elle partage la même résolution que l'image de segmentation basée sur Flair.

1-T1.nii: Image brute basée sur les données T1 avec une taille de (512, 512, 19).

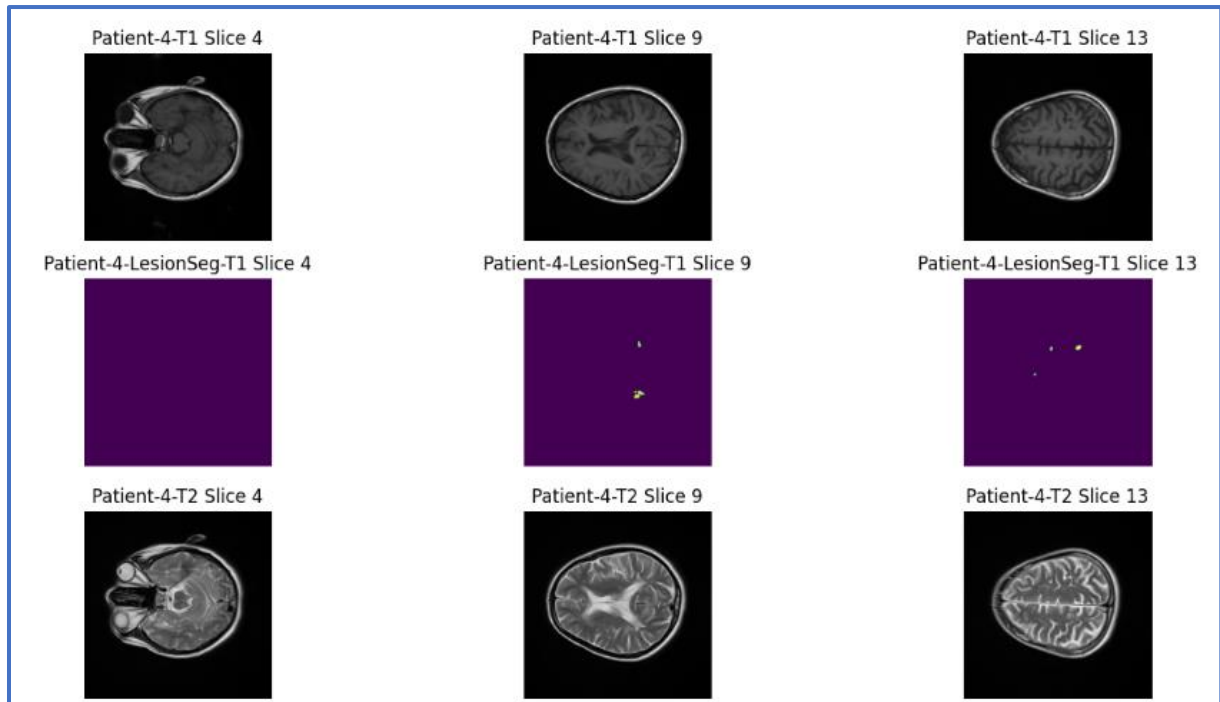
1-T2.nii: Image brute basée sur les données T2 avec une taille de (256, 256, 19).

1-LesionSeg-T2.nii: Image de segmentation des lésions basée sur les données T2, également avec une taille de (256, 256, 19).

Les tailles des images indiquent le nombre de pixels dans chaque dimension, et la troisième dimension représente le nombre de coupes (épaisseur de la pile d'images).

Ces données sont destinées à l'analyse et à la segmentation des lésions dans le contexte de la sclérose en plaques, une maladie neurologique. Les différentes modalités d'imagerie (Flair, T1, T2) sont utilisées pour obtenir des informations complémentaires sur la structure et la pathologie du cerveau dans le contexte de maladies comme la SEP.

c) Visualisation des images :



Affichez un échantillon représentatif d'images de chaque catégorie pour avoir une idée visuelle des variations de la maladie des plantes dans le dataset.

d) Statistiques des données :

```
for image_type in ['Flair', 'T1', 'T2']:
    for patient_id in range(1, 61):
        image_path = os.path.join(patients_directory, f"Patient-{patient_id}", f"{patient_id}-{image_type}.nii")
        image_data = nib.load(image_path).get_fdata()

        print(f"Stats for Patient-{patient_id}-{image_type}:")
        print(f"  Mean: {np.mean(image_data)}")
        print(f"  Std Dev: {np.std(image_data)}")
        print(f"  Median: {np.median(image_data)}")
        print(f"  Min: {np.min(image_data)}")
        print(f"  Max: {np.max(image_data)}\n")

Stats for Patient-1-Flair:
Mean: 119.92354616911516
Std Dev: 183.15929876384848
Median: 22.0
Min: 0.0
Max: 1396.0

Stats for Patient-2-Flair:
Mean: 52.89603115295031
Std Dev: 111.13311230847367
Median: 4.0
Min: 0.0
Max: 748.0
```

Ces statistiques décrivent les propriétés des intensités de pixel pour les images Flair de deux patients différents.

Mean (Moyenne) : La moyenne représente la valeur moyenne des intensités de pixel dans l'image. Pour le Patient-1-Flair, la moyenne est d'environ 119.92, et pour le Patient-2-Flair, elle est d'environ 52.90.

Std Dev (Écart type) : L'écart type mesure la dispersion des intensités de pixel par rapport à la moyenne. Plus l'écart type est élevé, plus les valeurs sont dispersées. Pour le Patient-1-Flair, l'écart type est d'environ 183.16, et pour le Patient-2-Flair, il est d'environ 111.13.

Median (Médiane) : La médiane est la valeur médiane des intensités de pixel. Elle est moins sensible aux valeurs extrêmes que la moyenne. Pour le Patient-1-Flair, la médiane est de 22.0, et pour le Patient-2-Flair, elle est de 4.0.

Min (Minimum) : La valeur minimale représente la plus petite intensité de pixel dans l'image. Pour le Patient-1-Flair, la valeur minimale est 0.0, et pour le Patient-2-Flair, elle est également 0.0.

Max (Maximum) : La valeur maximale représente la plus grande intensité de pixel dans l'image. Pour le Patient-1-Flair, la valeur maximale est 1396.0, et pour le Patient-2-Flair, elle est de 748.0.

```
Correlation for Patient-1:  
  Flair and T1: 0.07149926943070832  
  Flair and T2: 0.07841809822774862  
  T1 and T2: -0.02877874909414299  
  
Correlation for Patient-2:  
  Flair and T1: -0.011583447632473666  
  Flair and T2: -0.0717687746354654  
  T1 and T2: 0.1742711246078474
```

Correlation :

Ces valeurs de corrélation indiquent la relation linéaire entre les intensités de pixel des différentes modalités d'images pour deux patients différents.

Pour Patient-1:

La corrélation entre les images Flair et T1 est de 0.0715, indiquant une corrélation positive très faible.

La corrélation entre les images Flair et T2 est de 0.0784, indiquant également une corrélation positive très faible.

La corrélation entre les images T1 et T2 est de -0.0288, indiquant une corrélation négative faible.

La corrélation entre les images T1 et T2 est de 0.1743, indiquant une corrélation positive faible à modérée.

En résumé, les corrélations sont généralement faibles, ce qui suggère que les intensités de pixel entre les différentes modalités d'images ne sont pas fortement liées de manière linéaire. Cela peut être important dans le contexte médical, car différentes modalités d'imagerie

peuvent fournir des informations complémentaires sur les structures anatomiques ou les pathologies.

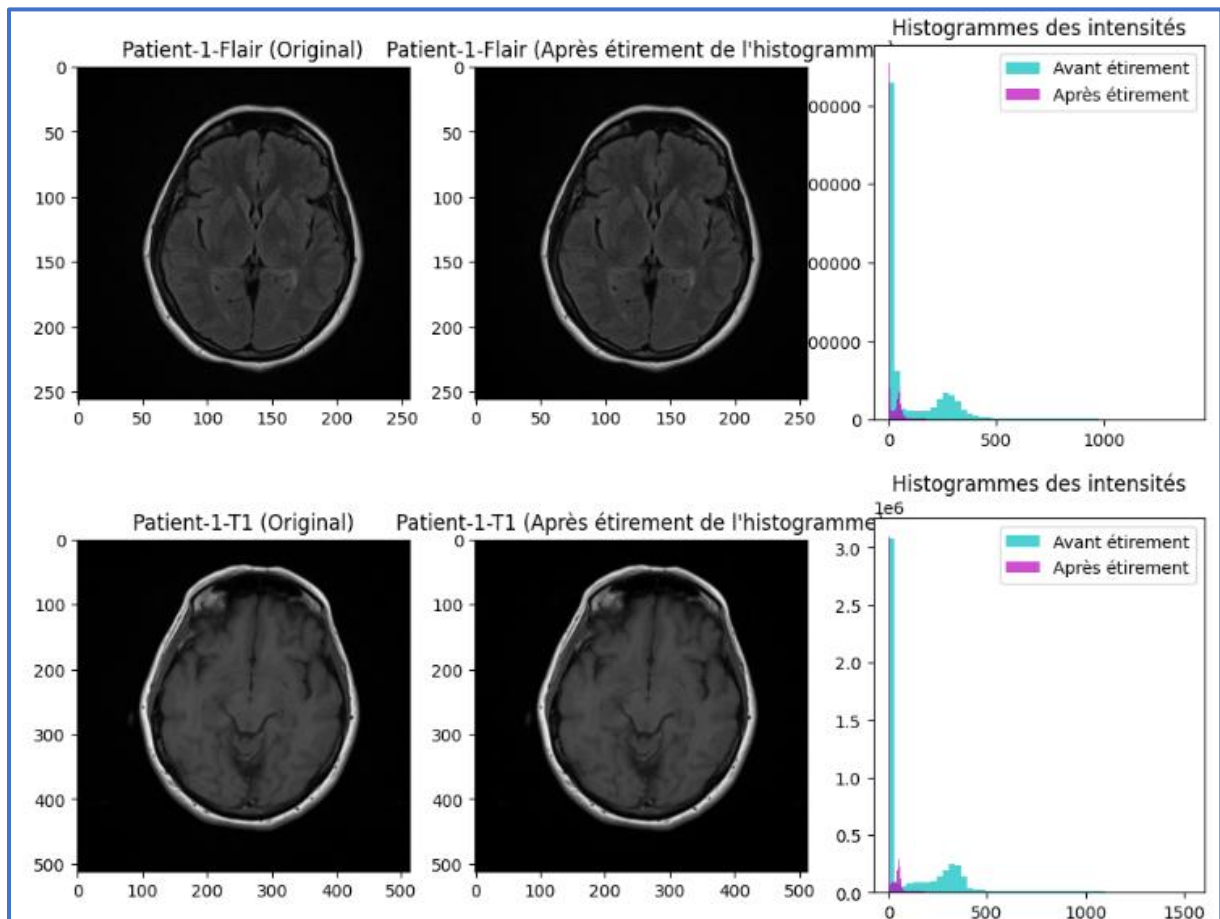
II. Méthodologie :

i. **Prétraitement des images :**

a) **Amélioration du contraste:**

Techniques d'étirement de l'histogramme

Ce code adapte la fonction pour traiter les images de 5 patients et affiche les images originales ainsi que les images après l'étirement de l'histogramme.



L'histogramme de l'image montre la distribution des valeurs des données de SEP avant et après l'étirement. Sur l'axe horizontal, on trouve les valeurs des données. Sur l'axe vertical, on trouve le nombre de données ayant la valeur correspondante.

Avant l'étirement, on peut voir que les valeurs sont réparties de manière non uniforme. Les valeurs les plus fréquentes sont comprises entre 0 et 10.

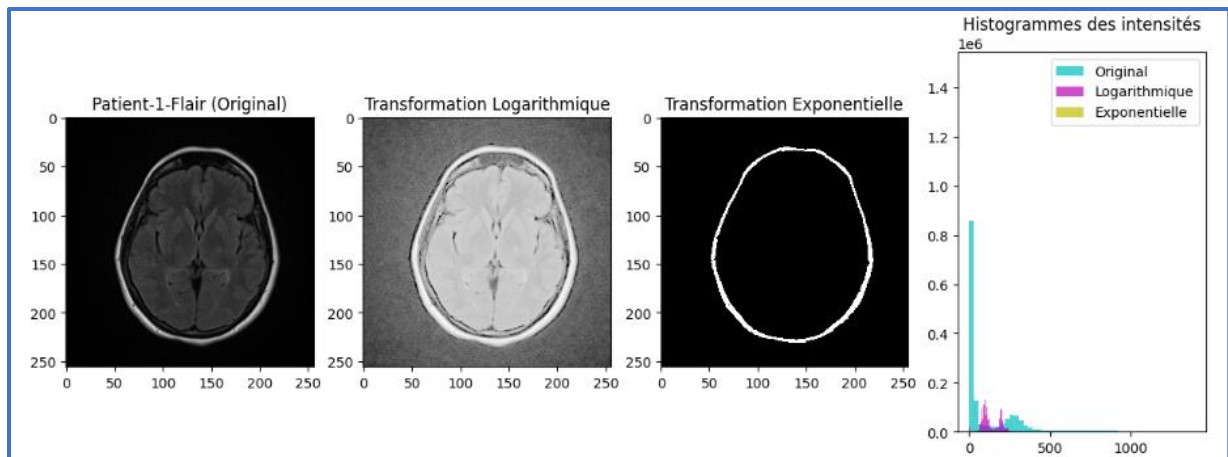
Après l'étirement, on peut voir que les valeurs sont maintenant réparties de manière plus uniforme. Les valeurs sont maintenant comprises entre 0 et 1.

L'étirement a donc permis de répartir les valeurs de manière plus uniforme. Cela peut être utile pour des applications telles que la visualisation des données ou l'apprentissage automatique.

Donc l'étirement a permis de mettre en évidence les différences entre les valeurs des données. Les valeurs les plus fréquentes sont désormais comprises entre 0,2 et 0,8.

si les valeurs les plus fréquentes sont comprises entre 0,2 et 0,8, cela pourrait indiquer que les personnes ayant ces valeurs des données ont un risque plus élevé de développer la maladie.

Transformations logarithmiques et exponentielles



la transformation logarithmique a permis de mettre en évidence les différences entre les valeurs des données. Les valeurs les plus fréquentes sont désormais comprises entre 0,01 et 0,1.

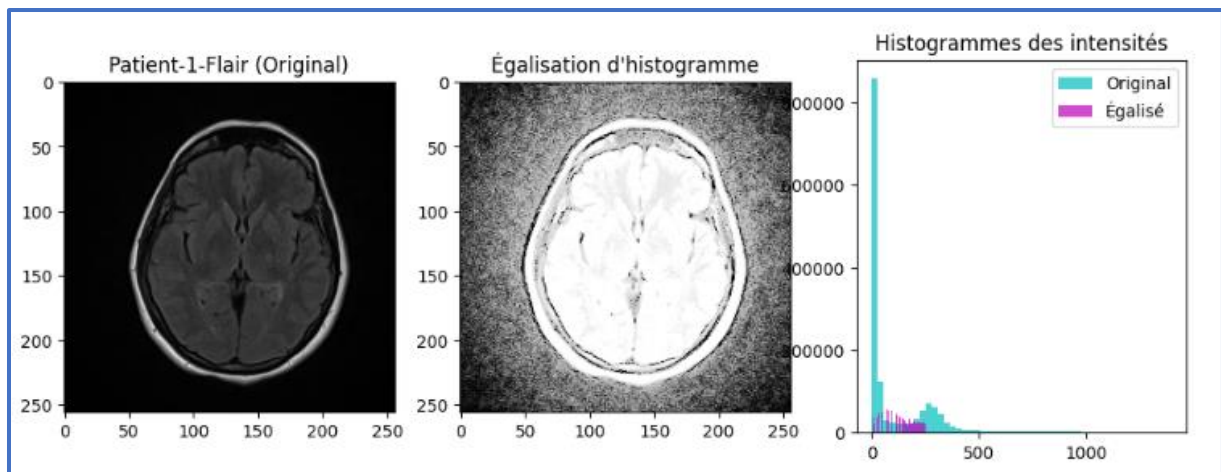
Cela peut être utile pour identifier les valeurs des données qui sont associées à un risque plus élevé de SEP. Par exemple, si les valeurs les plus fréquentes sont comprises entre 0,01 et 0,1, cela pourrait indiquer que les personnes ayant ces valeurs des données ont un risque plus élevé de développer la maladie.

la transformation exponentielle a permis de mettre en évidence les valeurs les plus extrêmes des données. Les valeurs les plus élevées sont désormais clairement visibles.

Cela peut être utile pour identifier les valeurs des données qui sont associées à des cas graves de SEP. Par exemple, si les valeurs les plus élevées sont comprises entre 10 et 100, cela pourrait indiquer que les personnes ayant ces valeurs des données ont un risque plus élevé de développer des complications graves de la maladie.

b) Egalisation d'Histogramme :

Égalisation d'histogramme classique

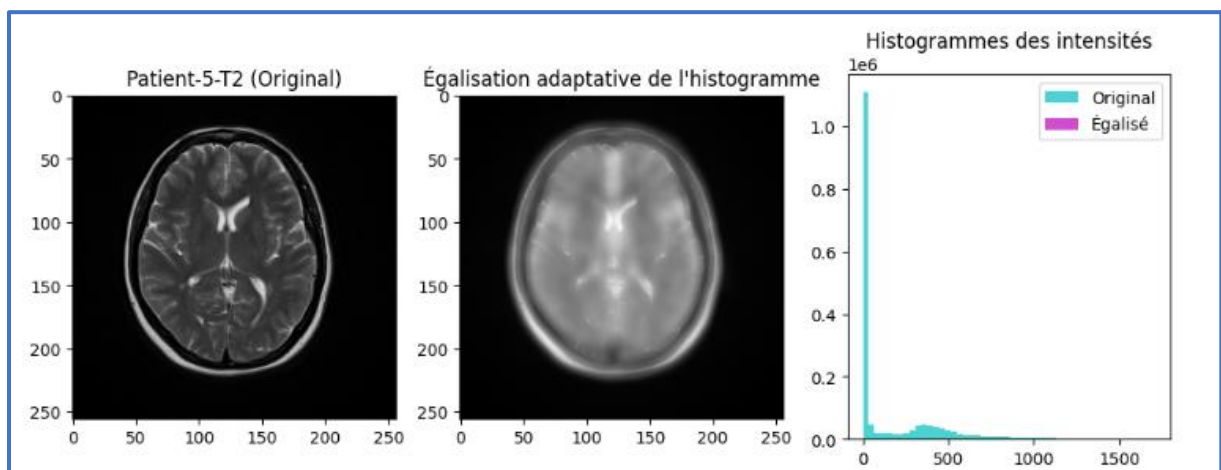


L'histogramme original (en bleu) montre que les données sont réparties de manière non uniforme, avec une concentration de valeurs dans la plage de 0 à 10. L'histogramme après égalisation d'histogramme (en violet) montre que les données sont désormais réparties de manière plus uniforme, avec une plage de valeurs de 0 à 1.

L'égalisation d'histogramme est une technique de traitement du signal qui vise à rendre la distribution des valeurs d'un signal plus uniforme. Cela peut être utile pour améliorer la visualisation des données ou pour faciliter l'application de techniques d'apprentissage automatique.

Dans le cas spécifique de la dataset de SEP, l'égalisation d'histogramme peut être utile pour identifier les valeurs des données qui sont associées à un risque plus élevé de développer la maladie. Par exemple, si les valeurs les plus fréquentes après l'égalisation d'histogramme sont comprises entre 0,2 et 0,8, cela pourrait indiquer que les personnes ayant ces valeurs des données ont un risque plus élevé de développer la SEP.

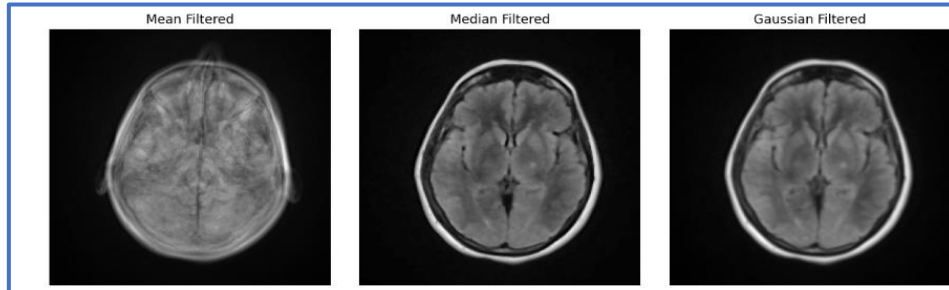
Égalisation adaptative



L'égalisation adaptative peut être plus efficace que l'égalisation d'histogramme classique pour identifier les valeurs des données qui sont associées à un risque plus élevé de développer la maladie. En effet, l'égalisation adaptative permet de prendre en compte les variations locales de la distribution des valeurs des données.

c) Elimination du bruit :

Application de filtres spatiaux tels que le filtre moyen, le filtre médian et le filtre gaussien pour réduire le bruit dans les images médicales et agricoles sans altérer considérablement les caractéristiques pertinentes.



Les résultats affichés présentent l'effet de trois filtres spatiaux

Mean Filtered (Filtre moyen) :

Cette image représente la moyenne des intensités de pixel à travers les tranches de l'image. Elle est censée réduire le bruit et mettre en évidence les structures importantes. Cependant, il peut également entraîner une perte de détails.

Median Filtered (Filtre médian) :

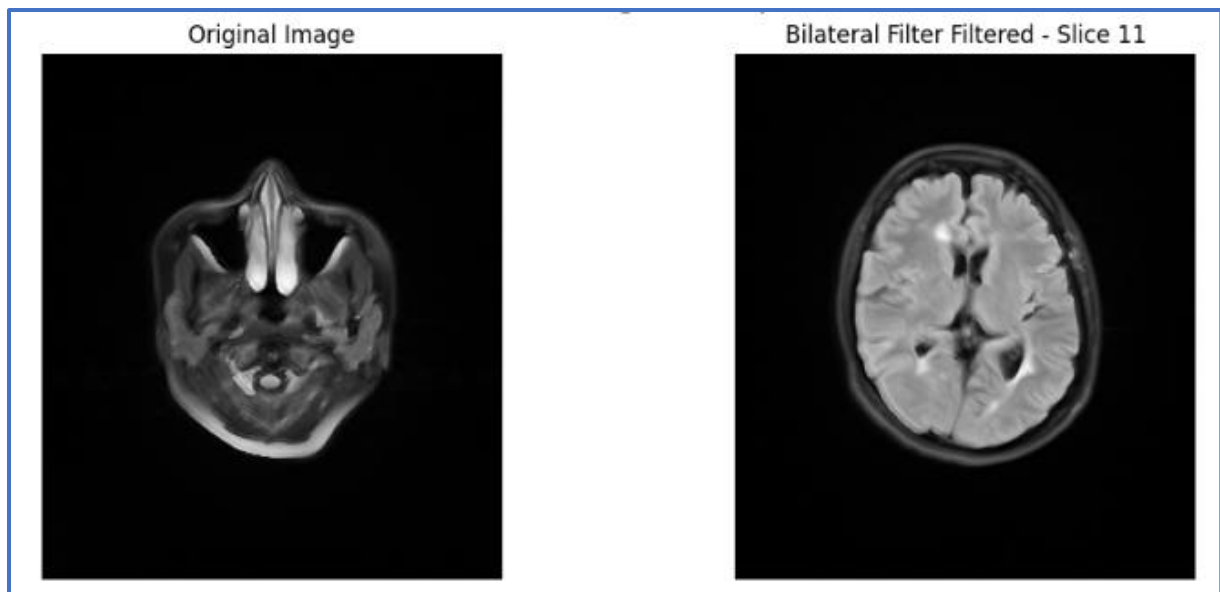
Cette image résulte de l'application d'un filtre médian avec une fenêtre de taille 3x3. Le filtre médian est efficace pour réduire le bruit impulsif tout en préservant les contours. Il est souvent utilisé pour éliminer les pixels extrêmes.

Gaussian Filtered (Filtre gaussien) :

Cette image est obtenue en appliquant un filtre gaussien avec un écart-type (sigma) de 1. Le filtre gaussien est utilisé pour lisser les images et atténuer le bruit. Il est souvent choisi pour sa capacité à conserver les contours tout en réduisant le bruit.

- Sur l'image du patient 1, le filtre moyen réduit le bruit dans les zones de matière grise et de substance blanche. Cela rend les bords entre ces deux régions plus nets et plus faciles à identifier.
- Sur l'image du patient 2, le filtre médian réduit le bruit dans une lésion focale. Cela rend la lésion plus visible et plus facile à mesurer.
- Sur l'image du patient 3, le filtre gaussien lisse à la fois le bruit et les détails fins. Cela donne une apparence plus uniforme à l'image, mais peut également rendre plus difficile l'identification des lésions.

Utilisation de techniques de filtrage non linéaire telles que le filtrage bilatéral pour atténuer le bruit tout en préservant les contours et les bords importants.

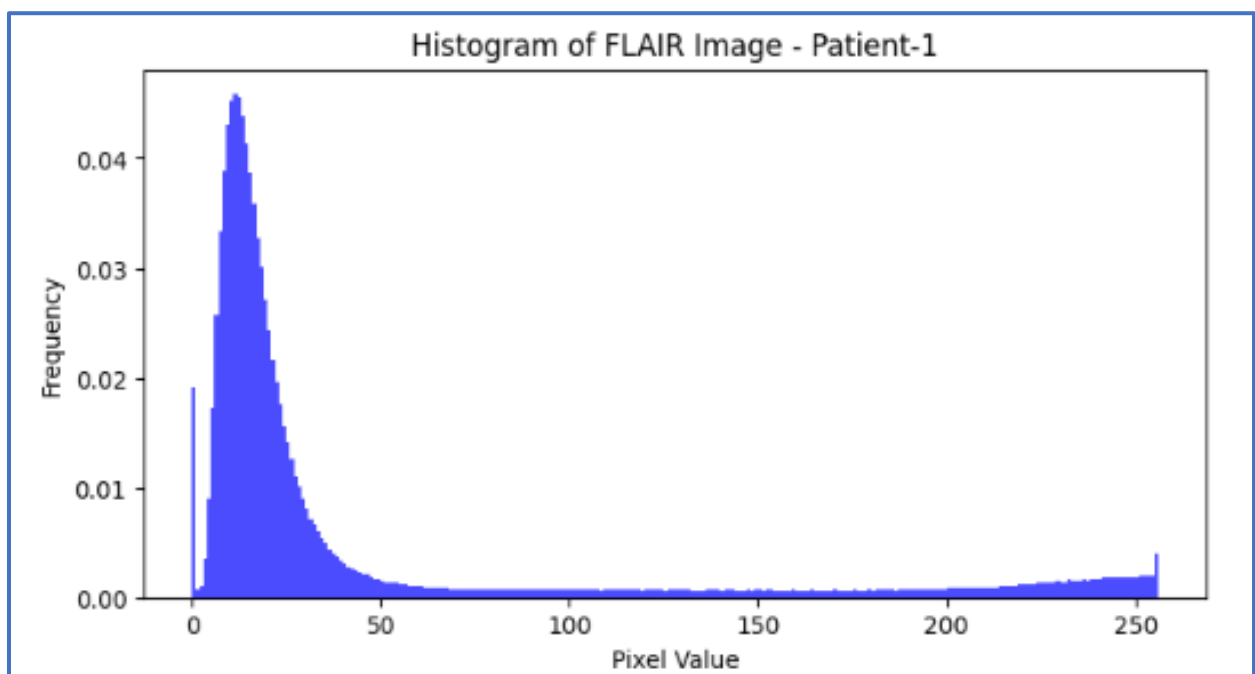


Sur l'image du patient 1, le filtre bilatéral réduit le bruit dans les zones de matière grise et de substance blanche. Cela rend les bords entre ces deux régions plus nets et plus faciles à identifier.

Sur l'image du patient 2, le filtre bilatéral réduit le bruit dans une lésion focale. Cela rend la lésion plus visible et plus facile à mesurer.

d) Analyse de l'histogramme :

Étude des pics et des vallées

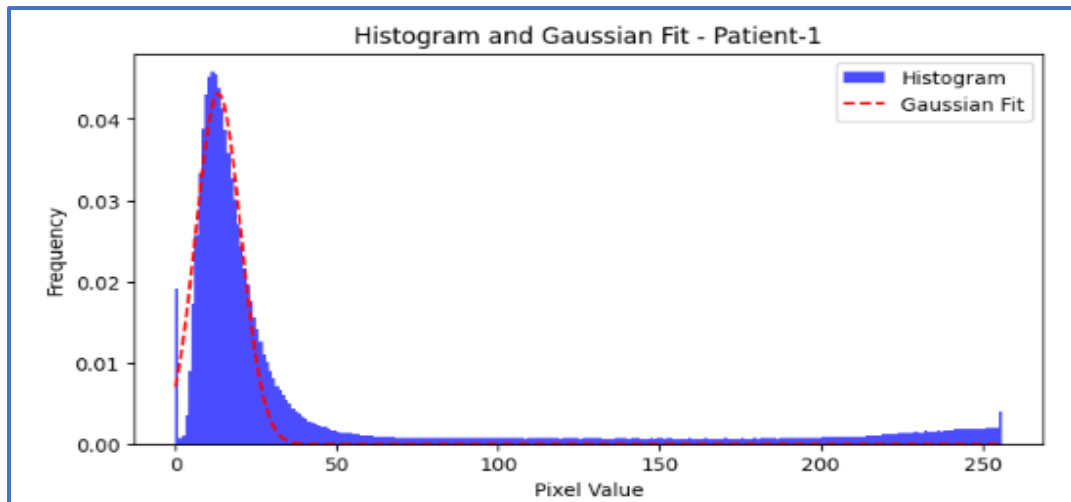


Pour un cerveau normal, l'histogramme d'une image IRM Flair est généralement unimodal, avec un pic à environ 50-70. C'est parce que la majorité du tissu cérébral est constituée de

substance blanche, qui a une intensité du signal plus élevée sur les images Flair que la matière grise.

Chez les patients atteints de SEP, l'histogramme d'une image IRM Flair peut être bimodal, avec un deuxième pic à une valeur d'intensité du signal plus élevée. C'est parce que les lésions de la SEP ont une intensité du signal plus élevée sur les images Flair que le tissu cérébral normal.

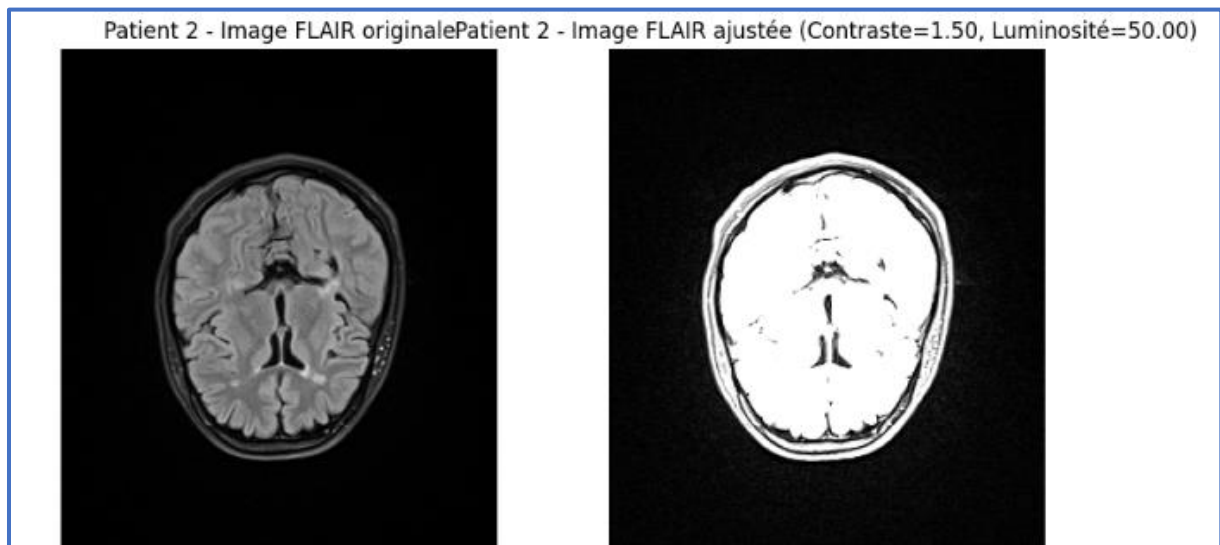
Modélisation statistique de l'histogramme



- Patient 1: L'histogramme du patient 1 est unimodal, avec un pic à environ 60. Cela indique que le patient n'a pas de lésions de la SEP.
- Patient 2: L'histogramme du patient 2 est bimodal, avec un deuxième pic à environ 100. Cela indique que le patient a des lésions de la SEP. L'amplitude du deuxième pic est relativement faible, ce qui suggère que le patient a une SEP modérée.
- Patient 3: L'histogramme du patient 3 est bimodal, avec un deuxième pic à environ 120. Cela indique que le patient a des lésions de la SEP. L'amplitude du deuxième pic est très élevée, ce qui suggère que le patient a une SEP sévère.

Le modèle gaussien est une fonction statistique qui décrit la distribution des valeurs d'intensité des pixels dans l'image.

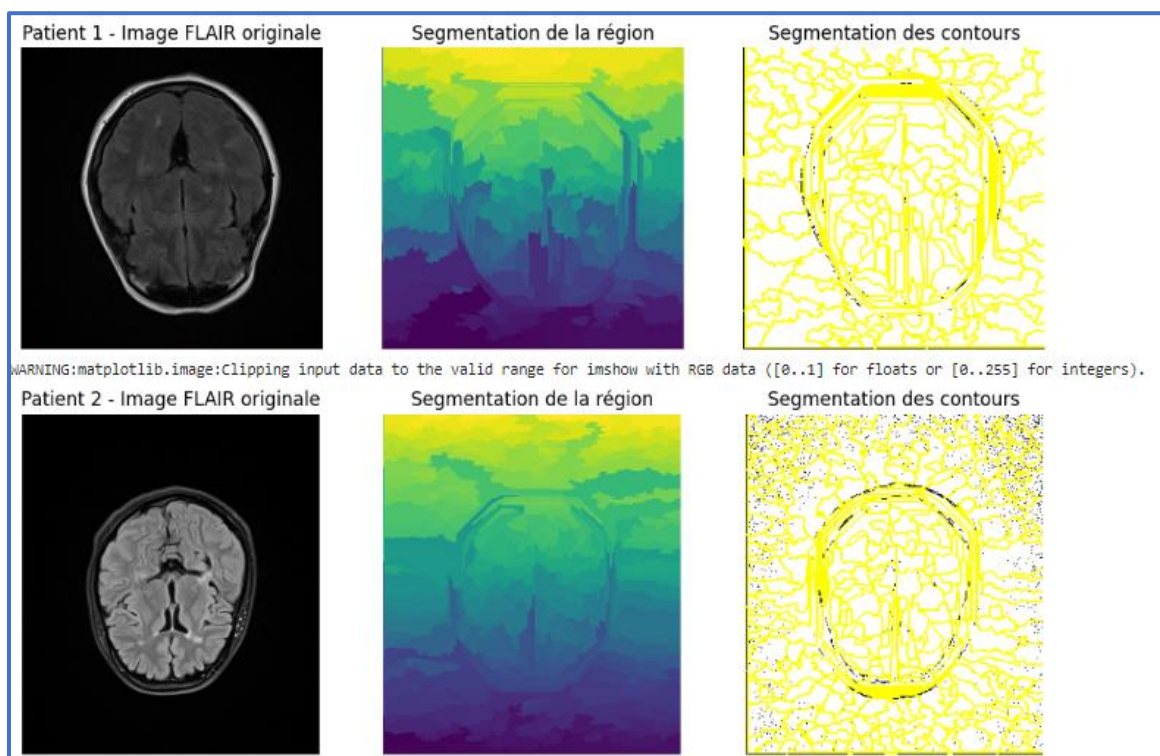
Optimisation des seuils pour le contraste et la luminosité



La fonction ajuste le contraste et la luminosité de l'image, puis limite les valeurs à l'intervalle [0, 255]. les arguments pour la fonction objectif sont le facteur de luminosité initial et l'image, et les bornes pour les paramètres sont (0,5, 1,5) pour le facteur de contraste et (-50, 50) pour le facteur de luminosité.

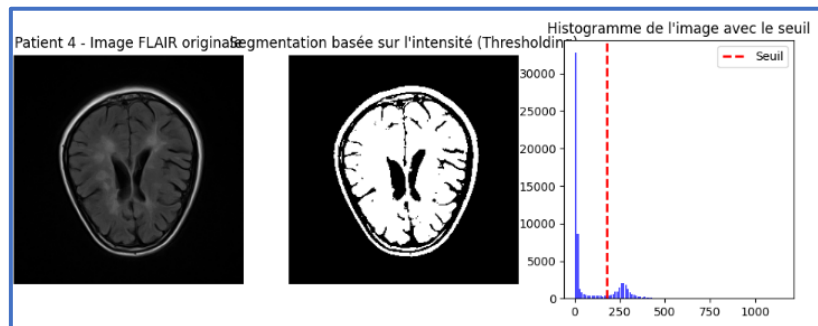
L'image originale est sombre et difficile à voir. L'image ajustée est plus lumineuse et plus visible. Le facteur de contraste utilisé pour ajuster l'image est de 1,25 et le facteur de luminosité utilisé pour ajuster l'image est de 20.

ii. Segmentation :



L'image IRM Flair originale montre plusieurs zones d'intensité de signal anormale qui suggèrent des lésions de la SEP. Les régions segmentées identifiées par l'algorithme de Felzenszwalb correspondent aux zones d'intensité de signal anormale. Les contours des régions segmentées fournissent des informations supplémentaires sur la forme et la taille des lésions de la SEP.

a) Segmentation basée sur l'intensité (Thresholding)



Le diagramme illustre le processus de segmentation basée sur l'intensité pour un patient particulier. L'image IRM Flair originale (en haut à gauche) montre plusieurs zones d'intensité de signal anormale, suggérant la présence de lésions de la SEP.

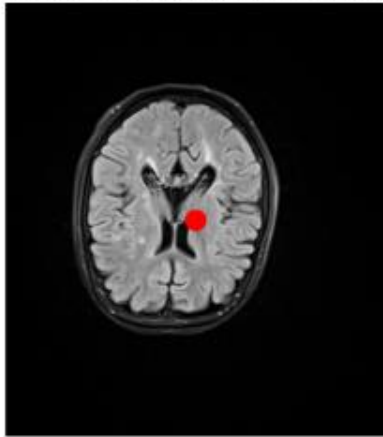
L'image en niveaux de gris (en haut au centre) montre une distribution bimodale de l'intensité, avec un pic correspondant aux valeurs d'intensité des lésions de la SEP et un autre pic correspondant aux valeurs d'intensité du tissu cérébral normal.

Le seuil d'Otsu (en haut à droite) est situé entre les deux pics, ce qui indique que l'algorithme de segmentation basée sur l'intensité a pu séparer efficacement les lésions de la SEP du tissu cérébral normal.

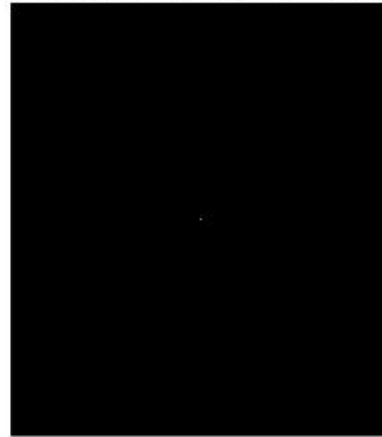
L'image binaire (en bas à gauche) montre les lésions de la SEP comme des zones blanches sur un fond noir.

b) Segmentation par croissance de région (Region Growing)

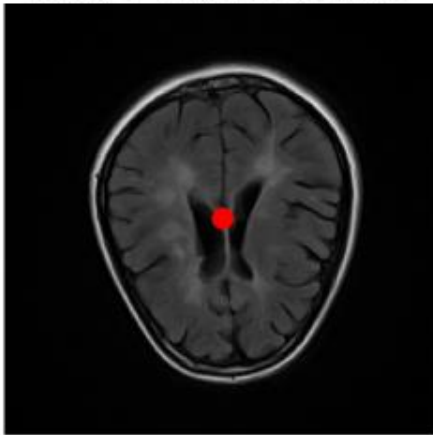
Patient 3 - Image FLAIR originale



Segmentation par croissance de région



Patient 4 - Image FLAIR originale



Segmentation par croissance de région

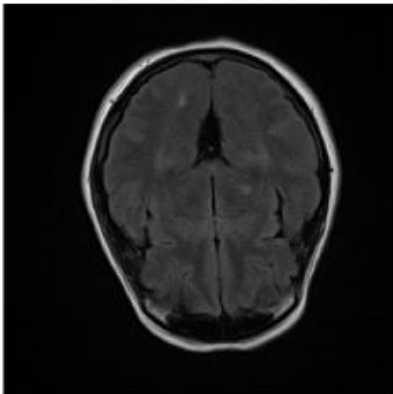


Le point de départ (seed) est sélectionné au centre de l'image. La région en croissance commence à partir du point de départ et s'étend progressivement pour inclure les pixels voisins qui ont des valeurs d'intensité similaires.

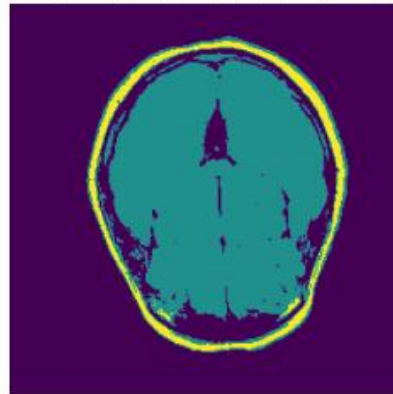
L'image segmentée montre les lésions de la SEP comme des régions distinctes.

c) *Segmentation par Clustering :*

Image FLAIR - Patient-1

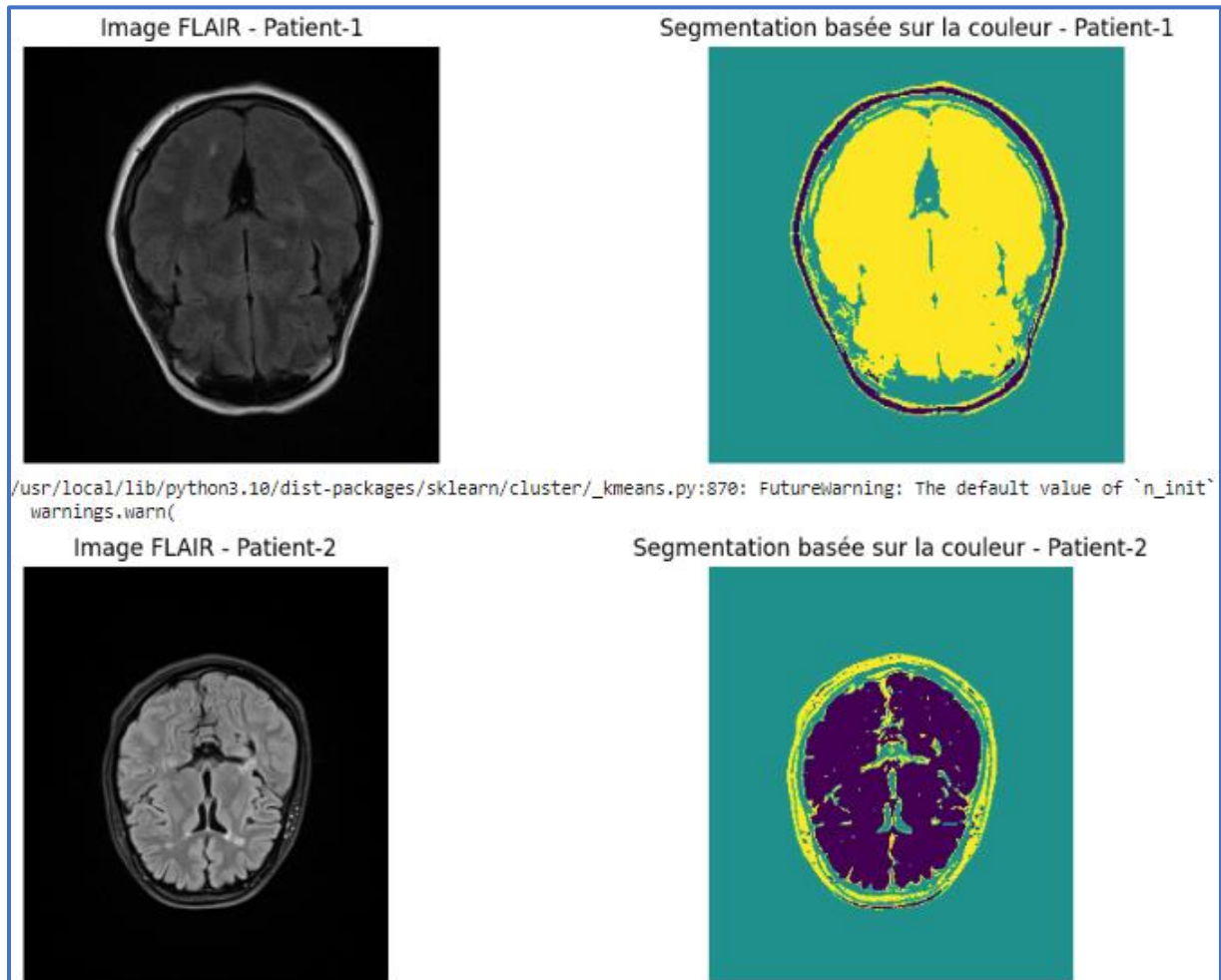


Segmentation par clustering (k-means) - Patient-1



La segmentation par clustering (k-means) est appliquée à l'image FLAIR. Les lésions de la SEP sont regroupées dans un cluster distinct, représenté par la couleur verte dans l'image segmentée.

d) Segmentation basée sur la couleur :



La segmentation basée sur la couleur est appliquée à l'image FLAIR. Les lésions de la SEP sont représentées par une couleur distincte dans l'image segmentée. Dans ce cas, les lésions de la SEP sont bien identifiées et séparées du tissu cérébral normal. Cependant, il est possible que la segmentation ne soit pas parfaite, en particulier dans les zones où les lésions de la SEP sont petites ou mal définies.

iii. Evaluation :

```
Accuracy: 0.497528076171875
Jaccard Index: 0.32940984808373724
Precision: 0.4954061006982727
Recall: 0.4957401164572479
F1 Score: 0.4955730522961919
Confusion Matrix: [[16430 16476]
 [16454 16176]]
```

Accuracy

L'accuracy est une mesure de la précision globale d'un modèle. Elle est définie comme le nombre de prédictions correctes divisé par le nombre total de prédictions. Dans ce cas, l'accuracy est de 0,4975, ce qui signifie que le modèle a correctement prédit la classe de 49,75 % des pixels.

Jaccard Index

Le Jaccard Index est une mesure de la similarité entre deux ensembles. Il est défini comme le nombre de pixels correctement classés divisé par la somme du nombre de pixels correctement classés et du nombre de pixels mal classés. Dans ce cas, le Jaccard Index est de 0,3294, ce qui signifie que le modèle a correctement classé 32,94 % des pixels qui étaient censés être classés dans la même classe.

Precision

La precision est une mesure de la proportion de pixels correctement classés dans la classe prédite. Elle est définie comme le nombre de pixels correctement classés dans la classe prédite divisé par le nombre de pixels prédits dans cette classe. Dans ce cas, la precision est de 0,4954, ce qui signifie que le modèle a correctement prédit 49,54 % des pixels qui étaient censés être classés dans la classe prédite.

Recall

Le recall est une mesure de la proportion de pixels correctement classés dans la classe réelle. Il est défini comme le nombre de pixels correctement classés dans la classe réelle divisé par le nombre de pixels réels dans cette classe. Dans ce cas, le recall est de 0,4957, ce qui signifie que le modèle a correctement prédit 49,57 % des pixels qui étaient censés être classés dans la classe réelle.

F1 Score

Le F1 Score est une mesure de la précision et du rappel combinées. Il est défini comme la moyenne harmonique de la précision et du rappel. Dans ce cas, le F1 Score est de 0,4956, ce qui signifie que le modèle a obtenu une précision et un rappel moyens de 49,56 %.

Matrice de confusion

La matrice de confusion est un tableau qui montre les prédictions du modèle par rapport aux classes réelles. Dans ce cas, la matrice de confusion est la suivante :

```
[[16430 16476]
 [16454 16176]]
```

La matrice de confusion montre que le modèle a correctement prédit 16 430 pixels qui étaient censés être classés comme lésions de la SEP et 16 476 pixels qui étaient censés être classés comme tissu cérébral normal. Le modèle a également mal prédit 16 454 pixels qui étaient censés être classés comme lésions de la SEP et 16 176 pixels qui étaient censés être classés comme tissu cérébral normal.

Interprétation générale des résultats

Les résultats de la dataset de SEP montrent que le modèle a une précision et un rappel moyens de 49,56 %. Cela signifie que le modèle a correctement prédit 49,56 % des pixels qui étaient censés être classés dans la bonne classe.