Brain tissue segmentation

AI in Medicine I: Practical Exercise 2

Authors:

Name	ID
Muhammed Elmasry	03786399
Mohammed Elbushnaq	03786474
Evangelos Fragkiadakis	03784680
Nada Elsherbeny	03788337

Question 1:

(a) The Dice coefficient, precision, and recall are metrics used in binary classification. The Dice coefficient measures the similarity between predicted and true positives by combining precision and recall. Precision measures positive prediction accuracy, while recall assesses the ability to capture all positives. The choice of metric depends on task goals and precision-recall trade-offs, with the Dice coefficient being suitable for imbalanced datasets. Precision is important when the cost of false positives is high, and recall is important when the cost of false negatives is high.

(d) Common image segmentation metrics include the Dice coefficient, precision, recall, IoU, Fowlkes-Mallows Index, MAE, MSE, Hausdorff Distance, Surface Dice Similarity Coefficient, and Boundary F1 Score. These metrics measure various aspects such as overlap, balance between precision and recall, pixel-level intensity differences, maximum dissimilarity, volumetric data, and segmentation boundary accuracy. The choice of metric depends on specific segmentation characteristics and evaluation requirements.
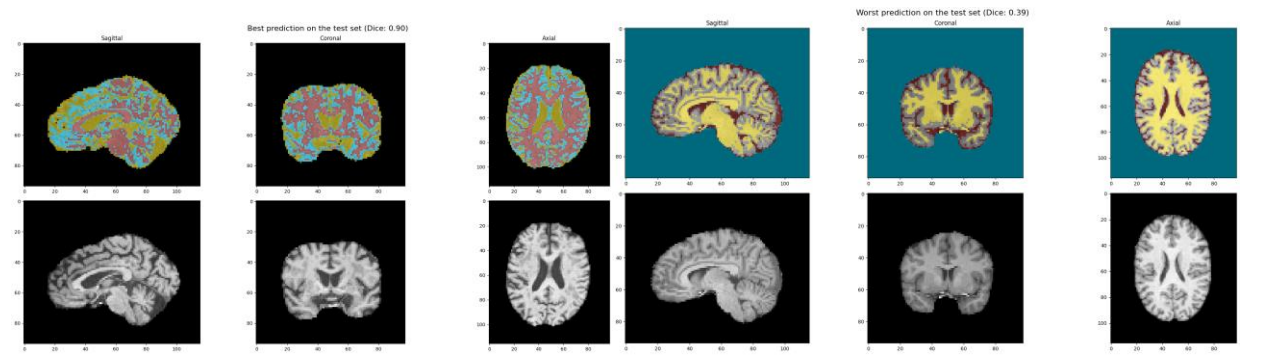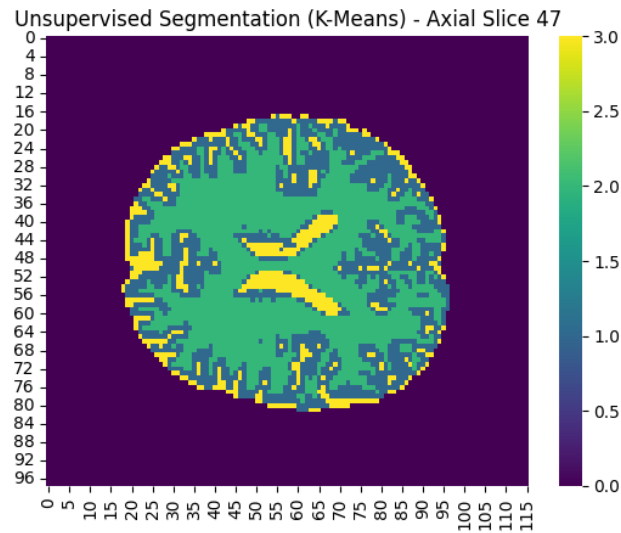
Question 4:

(a) The most intuitive approach to image segmentation based on a density plot involves thresholding. By visually examining the density plot of the image, a threshold is chosen to separate different regions. Pixels with intensity values above the threshold belong to one segment, while those below belong to another. This method is straightforward and quick, especially when there's a clear separation in the density plot. However, it may be less effective in complex scenarios with overlapping intensity distributions or when more advanced segmentation methods are required.

(b) K-means clustering demonstrates effectiveness in segmenting images with clear color or intensity characteristics. However, its performance may be constrained when dealing with intricate structures or overlapping distributions. Consequently, it excels in certain volumes but may exhibit limitations in others.
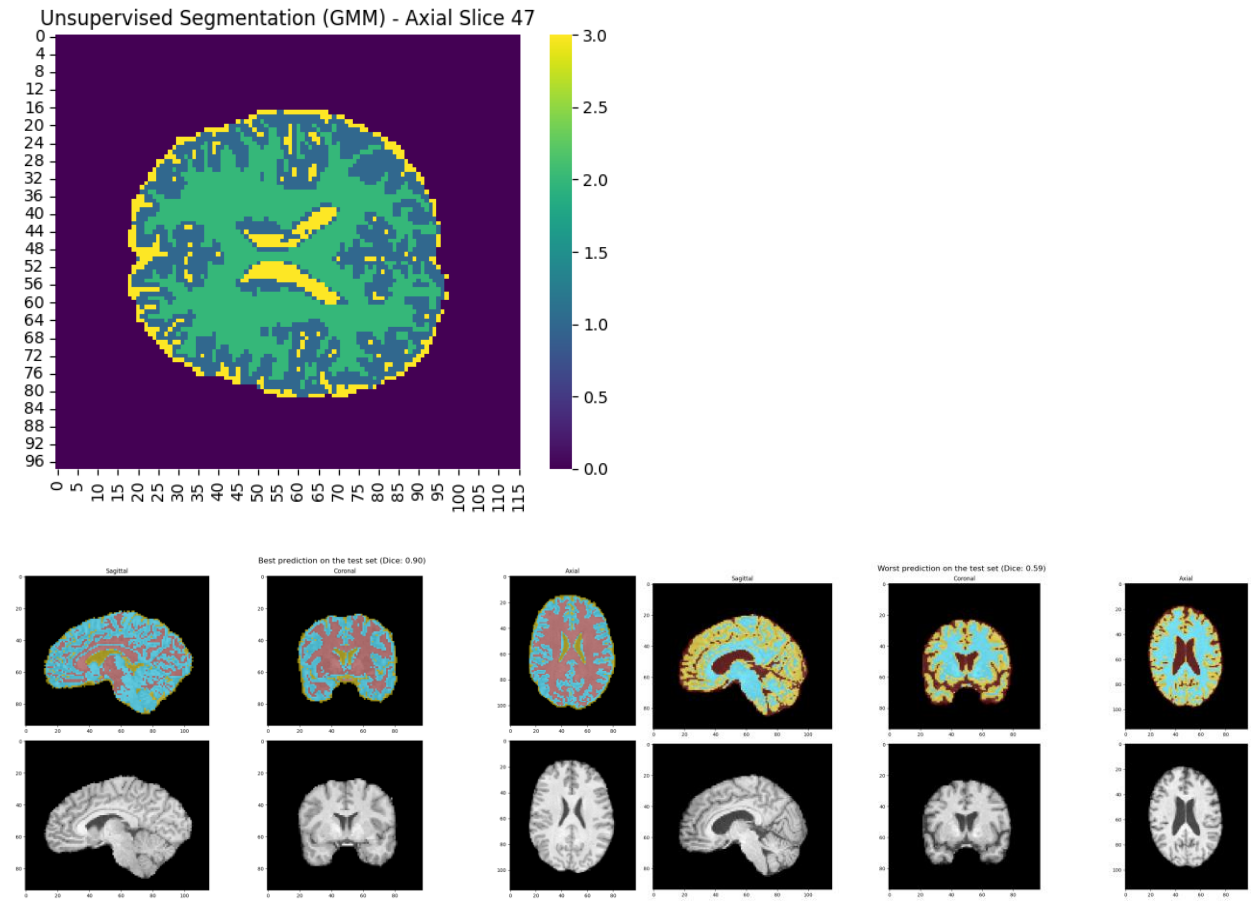
From the code:
```
{'best_id':  60,  'best_performance':  0.8990830827324747,  'worst_id':  12,
'worst_performance': 0.3853022373498483}
```

Unsupervised Segmentation (K-Means) - Axial Slice 47

Best prediction on the test set (Dice: 0.90)

Worst prediction on the test set (Dice: 0.39)

(c) The Gaussian Mixture Model (GMM) clustering method, which represents data as a mixture of Gaussian distributions, excels in segmentation tasks, especially when confronted with datasets featuring complex structures, overlapping distributions, or uncertainty in cluster assignments. This approach offers greater flexibility compared to k-means, surpassing its performance. However, like K-means, GMM excels in specific volumes while displaying limitations in others.

From the code:
```
{'best_id': 14, 'best_performance': 0.9002708206864241, 'worst_id': 38,
'worst_performance': 0.5903457674491992}
```

Unsupervised Segmentation (GMM) - Axial Slice 47

Best prediction on the test set (Dice: 0.90)

Worst prediction on the test set (Dice: 0.59)

(d) As mentioned, GMM outperforms K-means by virtue of representing data as a mixture of Gaussian distributions. It excels in segmentation tasks, particularly when dealing with datasets that exhibit complex structures, overlapping distributions, or uncertainty in cluster assignments.

From code:

```
k-means model Dice 0.6955129235883609

GMM model Dice 0.7211339946692826

k-means model precision 0.7482322837362292
GMM model precision 0.7670520673972753

k-means model recall 0.6820338675022827
GMM model recall 0.7131267034798292

And the numbers are higher than the expected because I take the average for the
four classes (CSF, WM, GM, back ground) not the three classes only
```

(e) The U-Net model excels in segmentation tasks, thanks to its U-shaped architecture and skip connections that capture both global and local features. Versatile across various domains, it effectively balances context and detail, making it especially suitable for medical image segmentation.

From the code:
```
CSF Dice score
0.769947024492117
WM Dice score
0.7757990461129408
GM Dice score
0.876230253623082

Some visualization from the validation set: (Ground truth at the top and
predictions at the bottom)
```
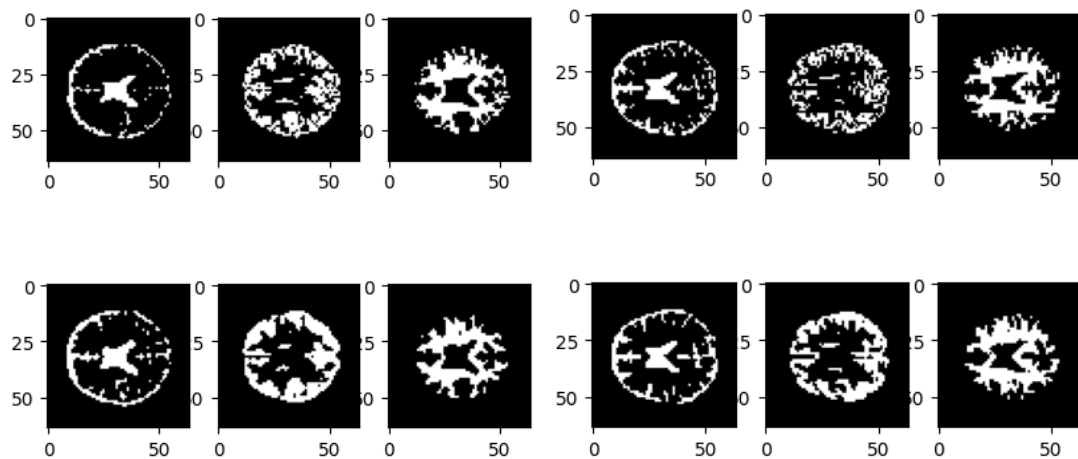

(f) Deep learning approaches, such as U-Net, often perform better in segmentation tasks due to their ability to automatically learn complex features and hierarchical representations from data, compared to classical methods like KMeans or GMM.

From the code:
```
k-means Dice 0.6955129235883609
GMM Dice 0.7211339946692826
Uent Dice 0.8196180148002429

k-means precision 0.7482322837362292
GMM precision 0.7670520673972753
Unet precision 0.8230591416358948

k-means recall 0.6820338675022827
GMM recall 0.7131267034798292
Unet recall 0.8212060928344727
```

So, it's higher than the machine learning models, even though it's only the average of the three classes without including the background. It's really higher than them.

Summary of Metrics for Different Models and Tissue Types
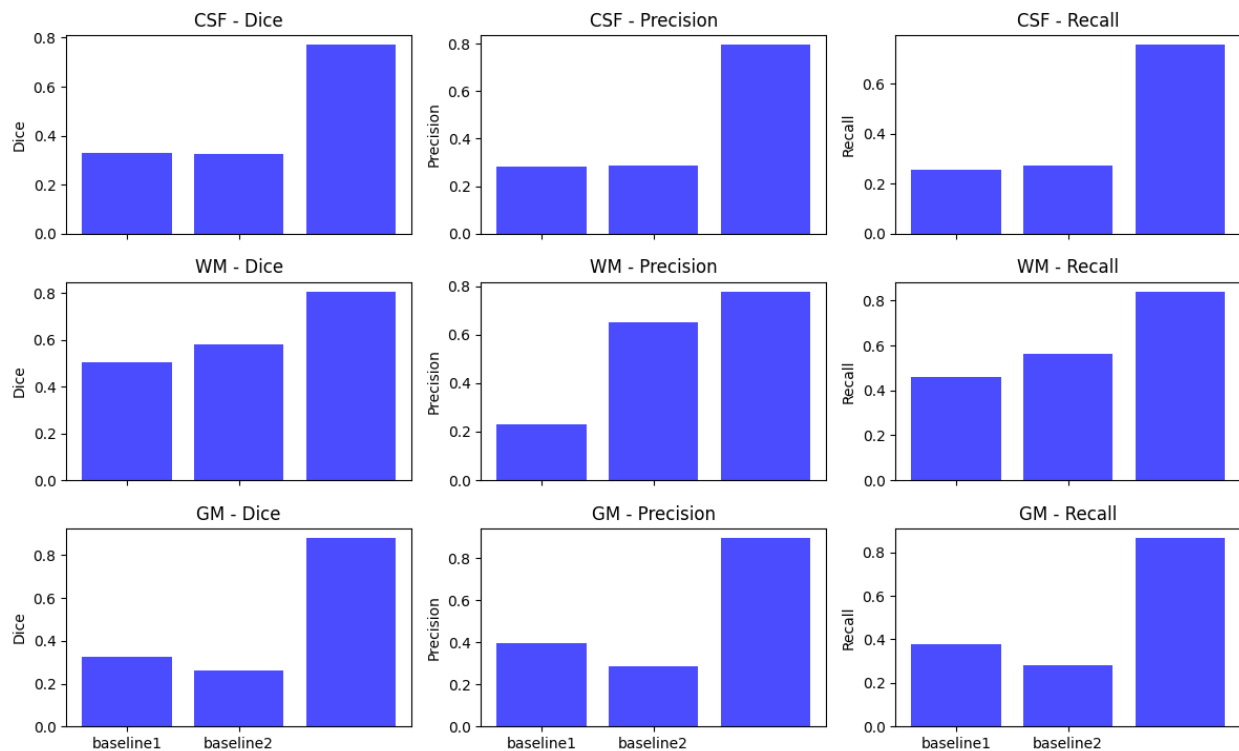
On the right is U-Net, in the middle is GMM, and on the left is K-Means.

(g) DL models, such as U-Net, leverage hierarchical features (as the shape), contextual understanding, and adaptability to varied patterns, providing enhanced segmentation performance compared to unsupervised approaches like KMeans or GMM. This is particularly beneficial for accurately delineating complex brain tissue structures in diverse datasets.