**Technical University Munich**
Department of Informatics

I31 - AI in Medicine and Healthcare
I32 - Computational Imaging in AI and Medicine

## Practical exercise 2 — 14. Nov. 2024

# Brain tumor segmentation

Submission deadline: 29. Nov. 2024, 11:59 p.m.

Please submit your solutions via Moodle.
The corresponding tutorial session is

14. Nov. 2024, 4-6 p.m. in lecture hall 5901.EG.051

For questions regarding this exercise sheet, please contact: `alina.dima@tum.de`, `can.erdur@tum.de`
For general questions, please contact: `course.aim-lab@med.tum.de`

Brain tumor segmentation is an important step in patient care, allowing for more sophisticated downstream tasks, such as predicting patient survival or tumor progression. The goal of the coursework is to implement classical and deep learning approaches for the segmentation of different tumor tissue types in Fluid Attenuated Inversion Recovery scans (FLAIR) of the brain, i.e., background, necrotic tumor core (NCT), edemataous/invaded tissue (ED), and enhancing tumor (ET). We provide data from a total of **484** subjects with glioma tumors, that is split into different development sets and a hold-out test set on which you will evaluate your final segmentation accuracy. Each approach will require a processing pipeline with different components that you will need to implement using methods that were discussed in the lectures and tutorials. There are dedicated parts in the Jupyter notebook that contain some detailed instructions and helper code. Please fill up the cells in the jupyter notebook. You will also need to write a report based on your experiments and observations.

1. (15 + 2 points) **Evaluation and Visualization**. Your first task is to implement methods to quantitatively and qualitatively evaluate the performance of the segmentation algorithms.
   (a) (5 points) **Evaluation metrics**. Implement the Dice similarity coefficient, precision and recall to evaluate the performance of segmentation algorithms.
   What is the relationship between the Dice coefficient, precision and recall? What are their pros and cons? Provide your answer in the report.
   **Hint**: Make sure to exclude the background class in the dice computation, and make sure that your score can handle multiple classes.
   (b) (5 points) **Visual inspection of best/worst predictions**. Given a set of predictions, visualize the images with the best and worst Dice scores across the entire test set. Complete the corresponding code block in the Jupyter notebook.
   (c) (5 points) **Results summary**. Given a set of predictions for a number of models, create a summary figure of your choice (e.g. table, bar plot) which visualizes all of the metric values. Visualize multiple metrics (Dice coefficient, precision, and recall), each computed for every tissue type. A sample metric dictionary is provided in the *Jupyter* notebook.
   (d) (2 bonus points) **Other metrics**. What other evaluation metrics are there for image segmentation? How do they differ from the ones implemented in Task 1a?

2. (30 points) **Unsupervised Segmentation**. The first segmentation approach that you are going to implement is based on the intensity distribution of the different brain tissues. Implement two different unsupervised learning methods to leverage the different intensity profiles of the tissues.

   (a) (15 points) **Unsupervised method 1**. Implement an unsupervised learning-based segmentation method of your choice.

   (b) (15 points) **Unsupervised method 2**. Implement a second unsupervised learning-based segmentation method of your choice.

   Please include in the report a description of each of the two methods implemented, how they work, your approach to choosing the hyperparameters (max. 1 page).
   **Hints:**

   - *sklearn* has implementations of unsupervised methods.
   - Some of the unsupervised methods might swap classes. Please make sure that you can remap the classes to the correct labels.

3. (20 points) **Deep Supervised Segmentation**. The second approach is to use deep supervised training to segment the different tissue types. Implement, train and evaluate a U-Net to segment the 4 tissues (NCT, ED, ET, and background). The followings apply:

   - Feel free to choose the number of layers, the number of features within convolutional layers, the number of convolutions within each layer, concatenation strategy..
   - You may use pre-defined models (e.g., from torchvision), but the networks have to be trained from scratch (no pre-training).
   - Do not use a library that obfuscates the training logic (e.g. PyTorch Lightning).
   - After you are done training, save the final model to a file *model.pth*, and update the jupyter notebook to read that file, run inference on the test dataset and generate the results figure. The results should be reproducible given the model file.
   - Please include in the report a summary of your network architecture (number of layers, types of convolutions, number of feature channels, concatenation strategy..), training scheme, hyperparameters, training observations, loss curves (max 2 pages).

   **Hints**:

   - Each channel of your predictions should represent 1 class, so make sure you are one-hot encoding your predictions and labels.
   - Analyze the training curves, check for overfitting.
   - Take a look at the results, qualitatively and quantitatively. What kind of errors is the network consistently making?
   - Make sure that the network output is normalized to a range compatible to the chosen training loss.

4. (35 + 2 points) **Analysis**. After you have implemented a few segmentation methods, it's time to analyze the results. Answer all of the following questions in the report.

   (a) (5 points) What is the most intuitive approach to segment the images based on the density plot of the input (Hint: take a look a the density plot in the first block of Task 2 in the *Jupyter* notebook)?

   (b) (5 points) How did method 2a perform? Comment based on the quantitative and qualitative results.

   (c) (5 points) How did method 2b perform? Comment based on the quantitative and qualitative results.

   (d) (5 points) Which unsupervised method performed better? Why?

   (e) (5 points) How did method 3 perform? Comment based on the quantitative and qualitative results.

(f) (5 points) Which approach (classical or DL) performed better? Why?

(g) (5 points) What additional information in the volumes is used by the DL models compared to the unsupervised approaches in Task 2? Why is it helpful?

(h) (3 bonus points) The official BraTS Glioma Segmentation challenge (source of the data) merges the tumor classes into non-mutually exclusive labels and creats a multi-label segmentation problem (see the last cell in the Jupyter notebook). What could be the reason of this label re-mapping?

Submit an archive **group_XYZ.zip** containing the following files:

- a Jupyter notebook **notebook.ipynb** containing your code, and relevant visualizations saved in the notebook cells.
- a file named **info.txt** containing the group number on the first line, as well as all of the group members' full names and matriculation numbers, one member per line.
- the model weights of your trained model: **model.pth**.
- a report file **report.txt** containing the answer to the theoretical questions, as well as all of the intermediate results and justifications for the practical questions.

Failure to comply with the submission format will result in a 10 point deduction.
For questions 2 and 3, full points will not be awarded in the absence of a corresponding report section.