



# **RECENT ADVANCES IN MEDICAL IMAGE SEGMENTATION AND CLASSIFICATION**

**László Szilágyi**

**Obuda University, Budapest, Hungary**

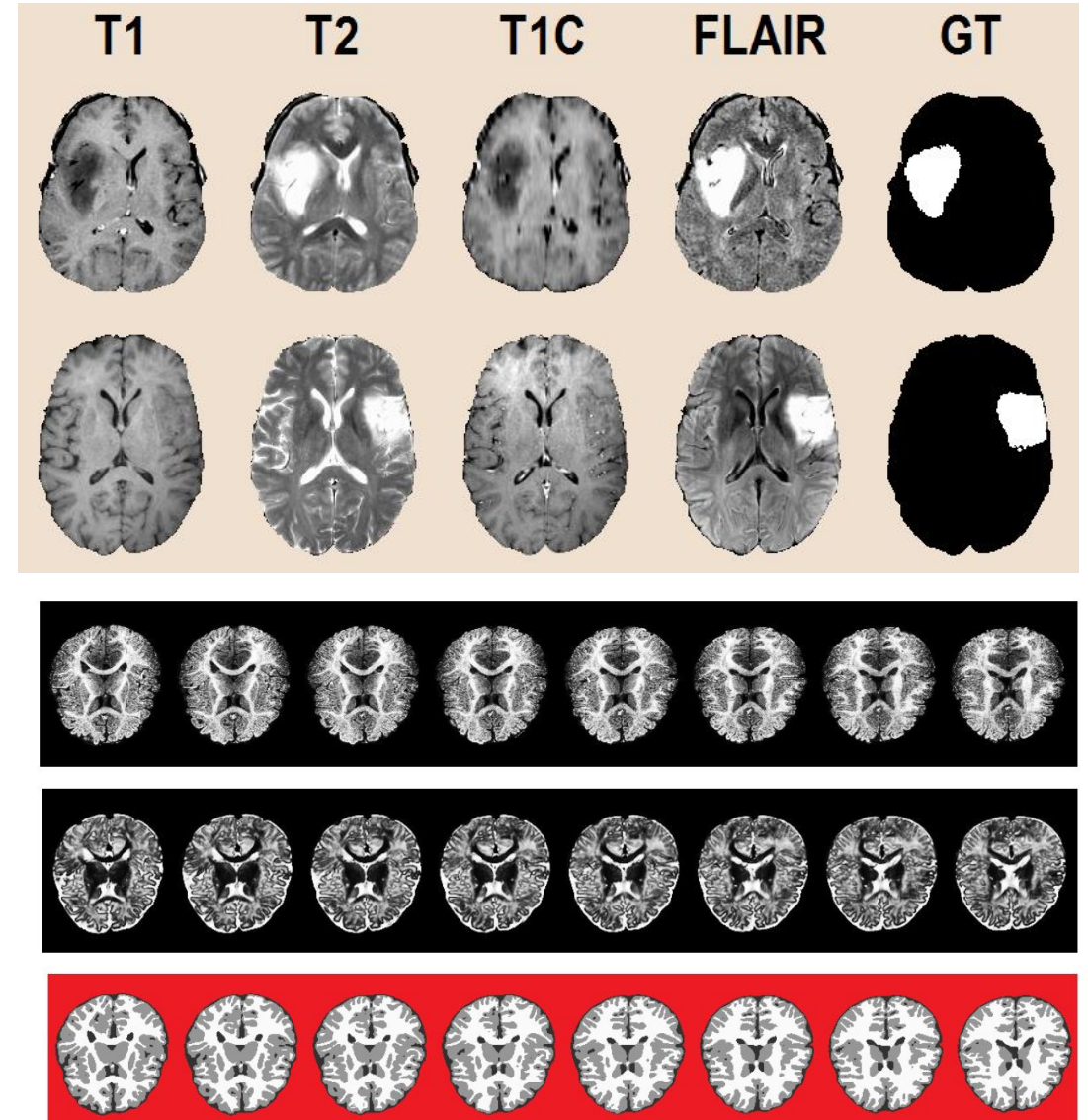
**Sapientia University of Transylvania, Romania**

# Why detect brain tumor?

- 100k++ people die of brain tumor yearly
- Early detection helps the survival
- More and more medical imaging devices
- Not so many more human experts
- Need for reliable automated algorithm
  - Draw attention to suspicious cases
  - Human expert has the last word
- Classical machine learning
- Convolutional neural networks and deep learning

# Input Data

- Medical Image Computation and Computer Aided Interventions (MICCAI)
- Brain Tumor Segmentation Challenge (BraTS) since 2012
- BraTS train dataset 2019
  - 76 low-grade (LG) and 259 high-grade (HG) volumes
- Multispectral (T1, T2, T1C, FLAIR)
- 155 x 240 x 240 image voxels
- Ground truth (negative, active tumor, necrotic tumor, tumor core, edema)
- iSeg-2017 challenge: T1, T2, and GT
- No tumor, just tissue segmentation



# Difficulties

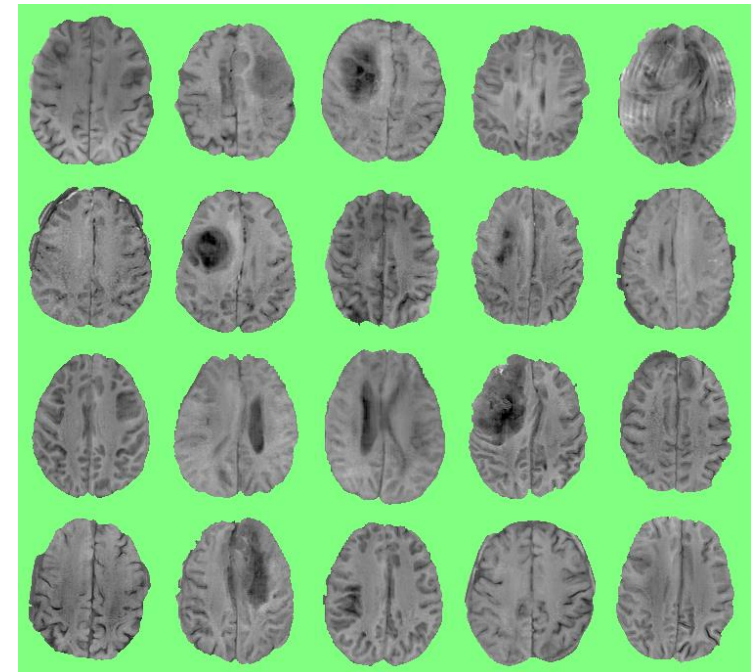
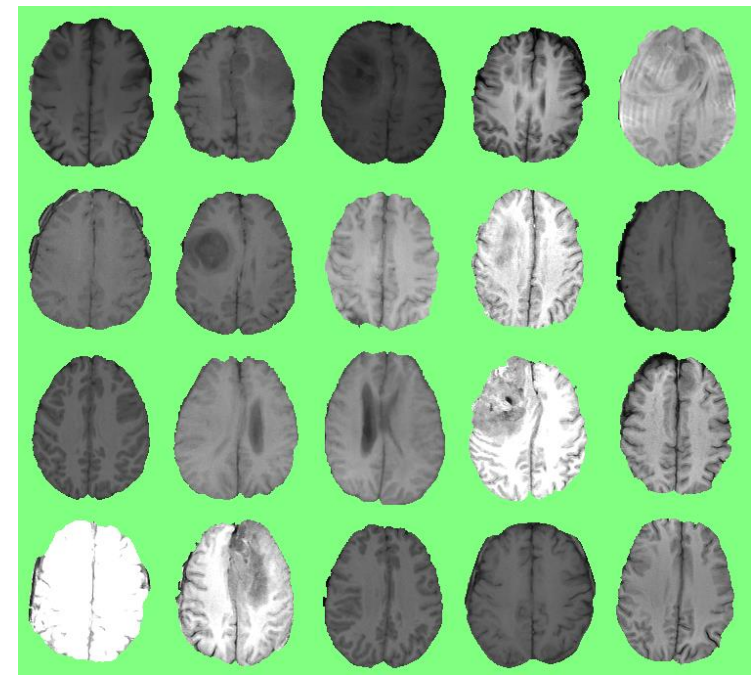
- Tumors have various locations, sizes, shapes, appearances
- Normal tissues are deformed, shifted
- Data channel registration is not perfect
- No standard scale of intensities
- Presence of noise, e.g. intensity non-uniformity

# Solutions

- Pre-BraTS era
  - Mostly unsupervised, image-processing methods (Gordillo 2013)
- BraTS era, 1<sup>st</sup> stage, classical machine learning methods
  - Classification of individual brain pixels, lots of features extracted after serious preprocessing, post-processing to improve coherence of decisions
- BraTS era, 2<sup>nd</sup> stage, CNN + deep learning
  - CNN architectures, no hand-crafted features, less pre-processing
  - Processing whole volumes, results may need regularization
- Our focus:
  - Contribute to both chapters
  - Optimize some supportive elements of segmentation techniques
  - Data enhancement (pre-processing)
  - Extra services (during segmentation or post-processing)

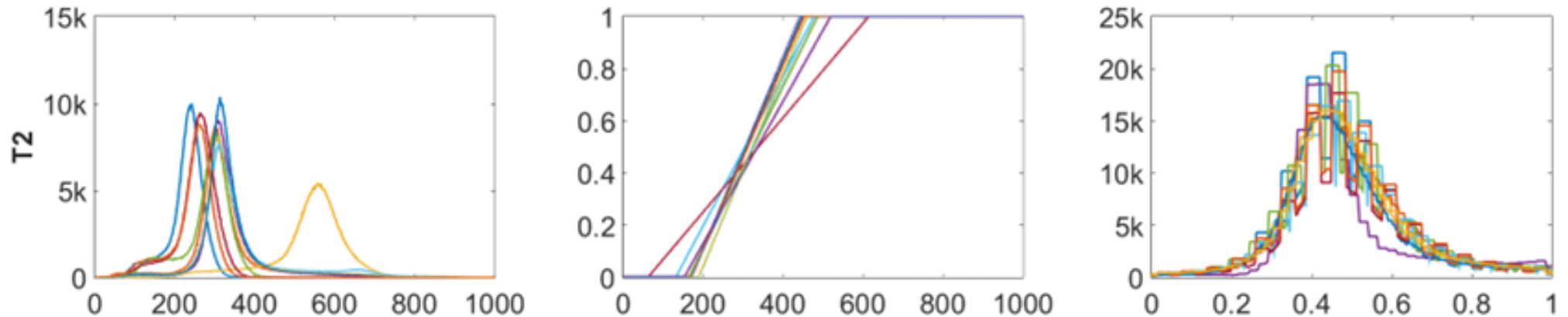
# No standard intensity scale

- Histograms need normalization
- Existing methods
  - Nyúl et al (2000) – piece-wise linear transform
  - Leung et al (2010) – segmentation, tissue-based alignment
  - Weisenfeld et al (2004) – Kullback-Leibler divergence
  - Shinohara et al (2011) – PCA
  - Jäger et al (2006) – hidden Markov random field
  - Linear transform

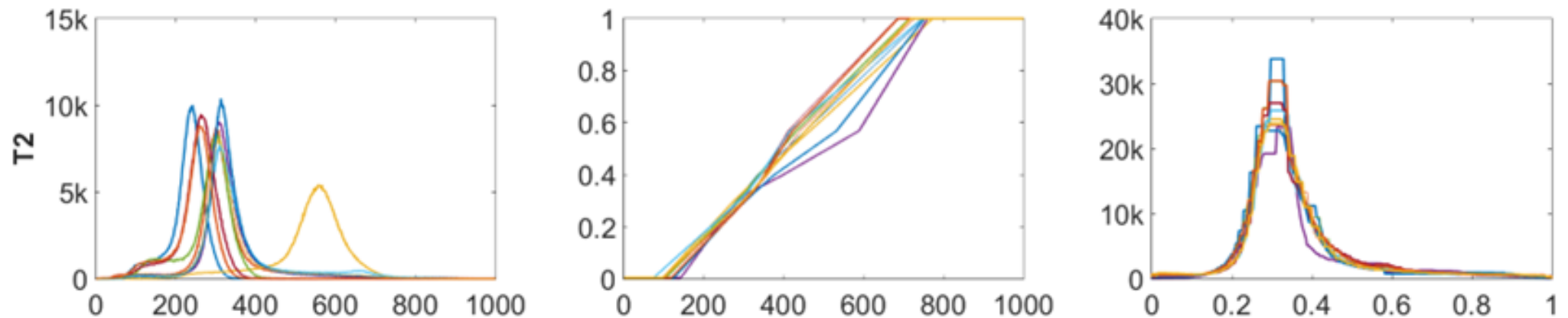


# Linear vs. piece-wise linear transform

- Linear



- Piece-wise linear: better alignment, but is it good for segmentation?





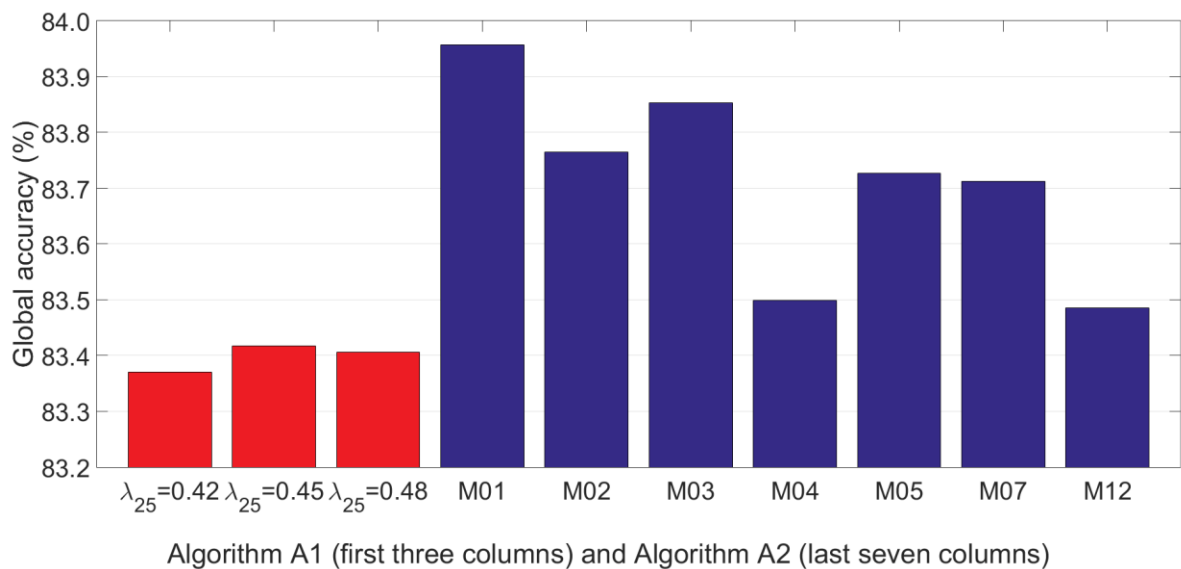
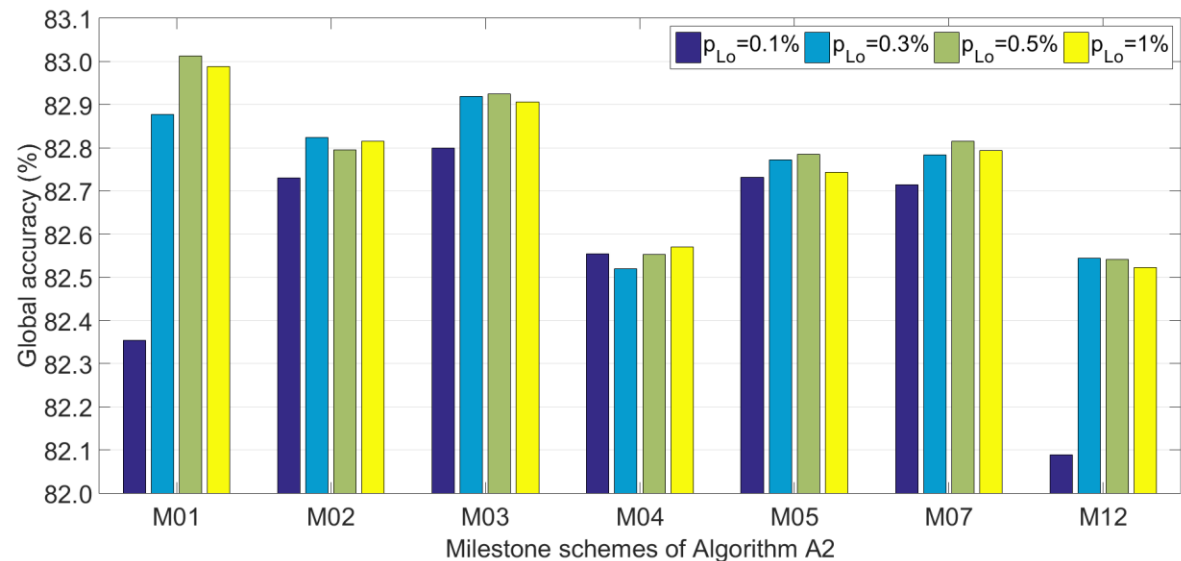
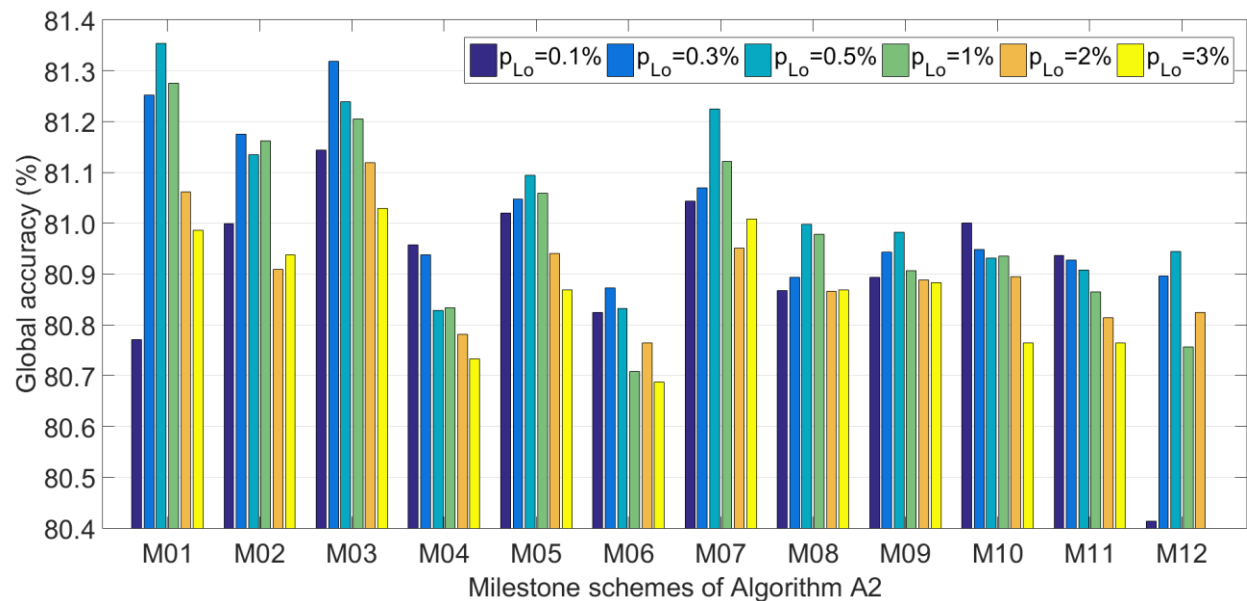
# Parameters

- Linear transform (Alg. A1)
  - single parameter  $\lambda_{25} \in [0.3, 0.5)$ ,  
 $p_{25} \rightarrow \lambda_{25}$  and  $p_{75} \rightarrow (1 - \lambda_{25})$
- Piece-wise linear transform (Alg. A2)
  - $p_{Lo} < 0.03$ ,  $p_{Hi} = 1 - p_{Lo}$  what part to cut off at both ends
  - Set of landmark points
    - fixed points to be aligned

| Scheme | Landmark points  |
|--------|--|
| M01    | $p_{Lo}, p_{50}, p_{Hi}$   |
| M02    | $p_{Lo}, p_{25}, p_{75}, p_{Hi}$   |
| M03    | $p_{Lo}, p_{25}, p_{50}, p_{75}, p_{Hi}$   |
| M04    | $p_{Lo}, p_{10}, p_{50}, p_{90}, p_{Hi}$   |
| M05    | $p_{Lo}, p_{20}, p_{40}, p_{60}, p_{80}, p_{Hi}$   |
| M06    | $p_{Lo}, p_{10}, p_{25}, p_{75}, p_{90}, p_{Hi}$   |
| M07    | $p_{Lo}, p_{20}, p_{35}, p_{50}, p_{65}, p_{80}, p_{Hi}$                                 |
| M08    | $p_{Lo}, p_{10}, p_{25}, p_{50}, p_{75}, p_{90}, p_{Hi}$                                 |
| M09    | $p_{Lo}, p_{10}, p_{25}, p_{40}, p_{60}, p_{75}, p_{90}, p_{Hi}$                         |
| M10    | $p_{Lo}, p_{10}, p_{25}, p_{40}, p_{50}, p_{60}, p_{75}, p_{90}, p_{Hi}$                 |
| M11    | $p_{Lo}, p_{10}, p_{20}, p_{30}, p_{40}, p_{60}, p_{70}, p_{80}, p_{90}, p_{Hi}$         |
| M12    | $p_{Lo}, p_{10}, p_{20}, p_{30}, p_{40}, p_{50}, p_{60}, p_{70}, p_{80}, p_{90}, p_{Hi}$ |



# Segmentation accuracy achieved with random forest classifier on iSeg-2017 data



# Before and after histogram alignment

Before

After

Before

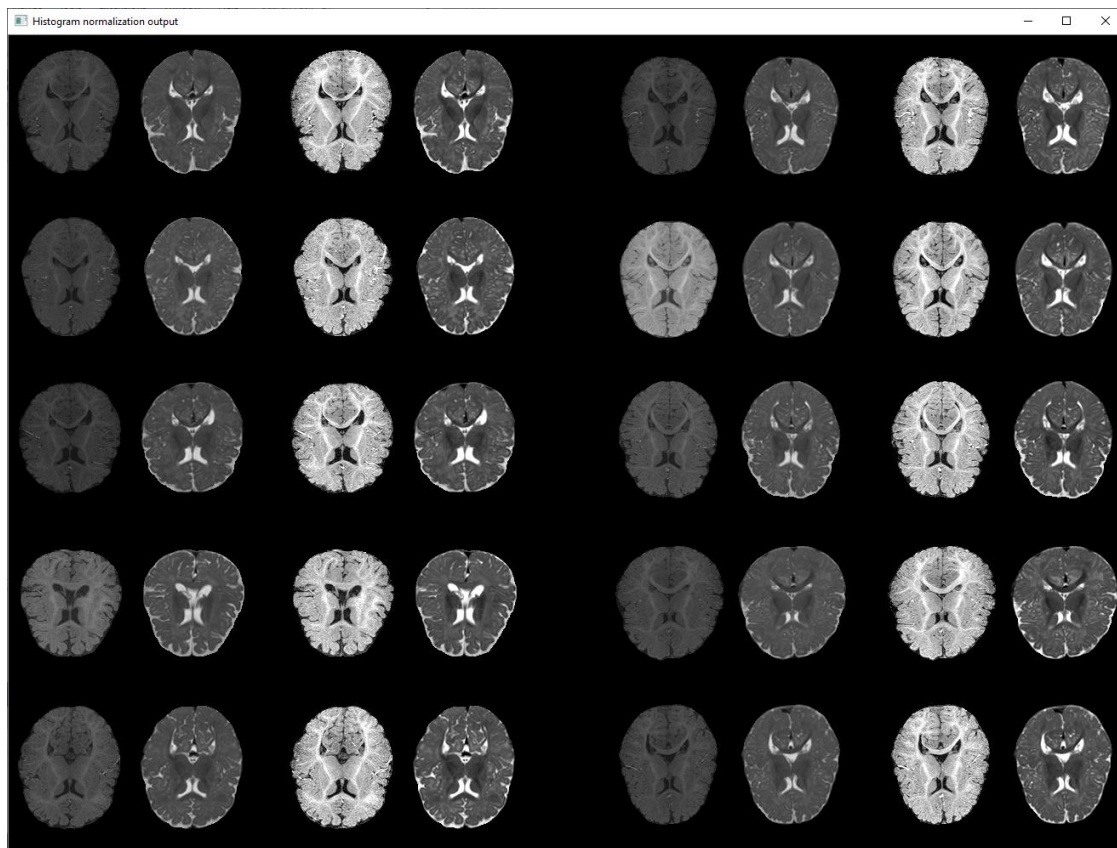
After

T1 T2

T1 T2

T1 T2

T1 T2

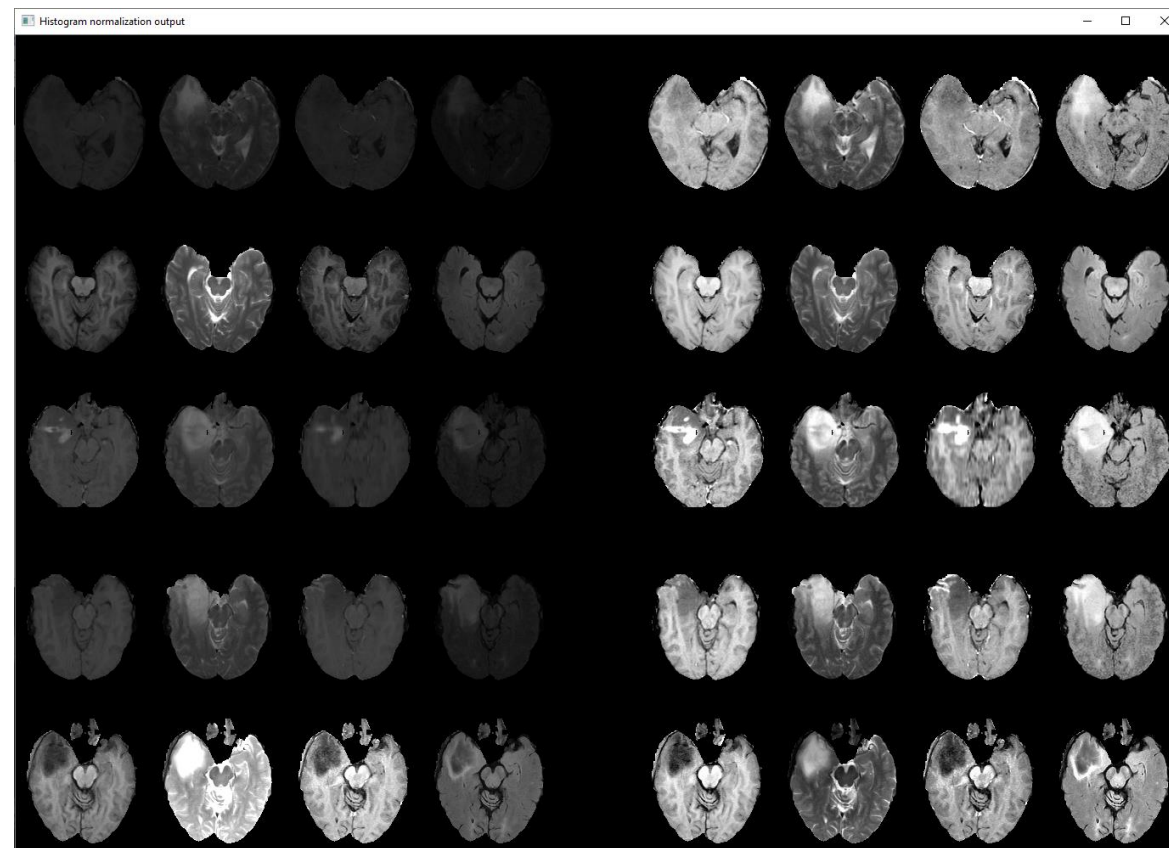


Before

After

T1 T2 T1c FLAIR

T1 T2 T1c FLAIR



# Recommendations

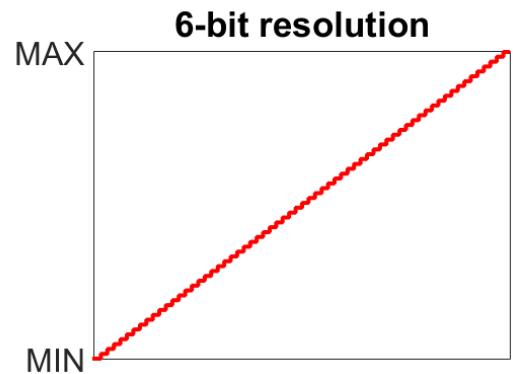
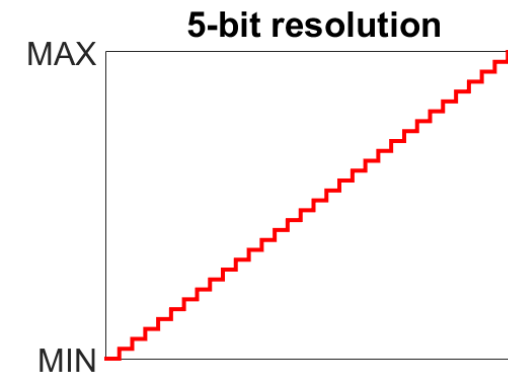
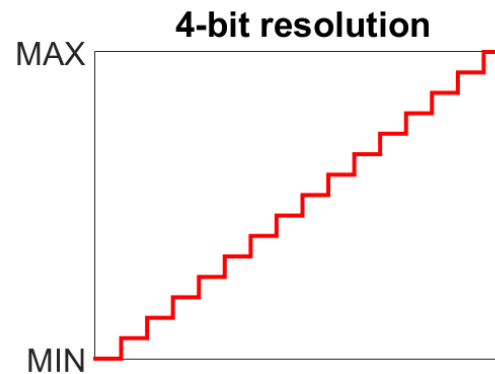
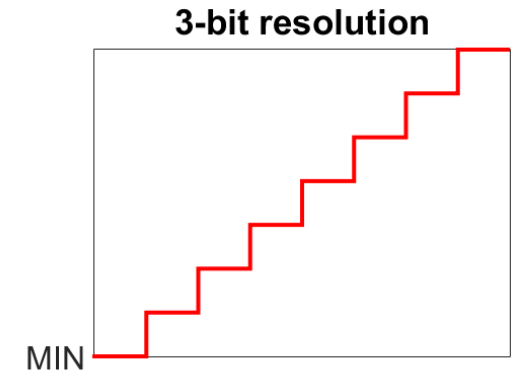
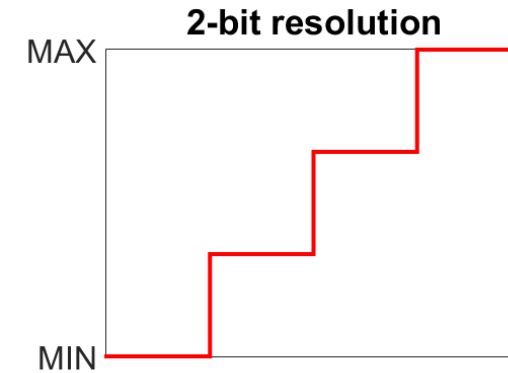
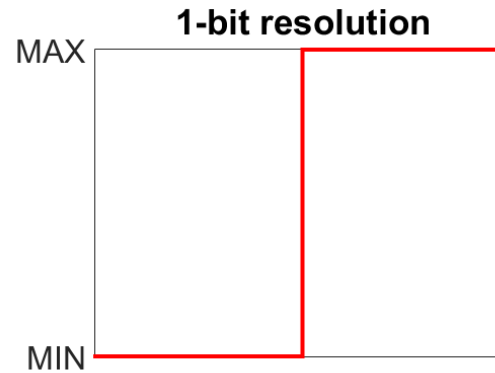
- Piece-wise linear transform (Nyúl et al) can perform better, but only if appropriately used
- Important issues
  - Not too many landmark points
  - Best schemes: M01, M03
  - No landmark points close to any ending of the histogram
  - better  $p_{20}$  than  $p_{10}$ , better  $p_{80}$  than  $p_{90}$
- Most participants at early BraTS competitions did not set it properly
- Better accuracy, higher Dice score, up to 2% difference

| Scheme | Landmark points  |
|--------|--|
| M01    | $p_{Lo}, p_{50}, p_{Hi}$   |
| M02    | $p_{Lo}, p_{25}, p_{75}, p_{Hi}$   |
| M03    | $p_{Lo}, p_{25}, p_{50}, p_{75}, p_{Hi}$   |
| M04    | $p_{Lo}, p_{10}, p_{50}, p_{90}, p_{Hi}$   |
| M05    | $p_{Lo}, p_{20}, p_{40}, p_{60}, p_{80}, p_{Hi}$   |
| M06    | $p_{Lo}, p_{10}, p_{25}, p_{75}, p_{90}, p_{Hi}$   |
| M07    | $p_{Lo}, p_{20}, p_{35}, p_{50}, p_{65}, p_{80}, p_{Hi}$                                 |
| M08    | $p_{Lo}, p_{10}, p_{25}, p_{50}, p_{75}, p_{90}, p_{Hi}$                                 |
| M09    | $p_{Lo}, p_{10}, p_{25}, p_{40}, p_{60}, p_{75}, p_{90}, p_{Hi}$                         |
| M10    | $p_{Lo}, p_{10}, p_{25}, p_{40}, p_{50}, p_{60}, p_{75}, p_{90}, p_{Hi}$                 |
| M11    | $p_{Lo}, p_{10}, p_{20}, p_{30}, p_{40}, p_{60}, p_{70}, p_{80}, p_{90}, p_{Hi}$         |
| M12    | $p_{Lo}, p_{10}, p_{20}, p_{30}, p_{40}, p_{50}, p_{60}, p_{70}, p_{80}, p_{90}, p_{Hi}$ |

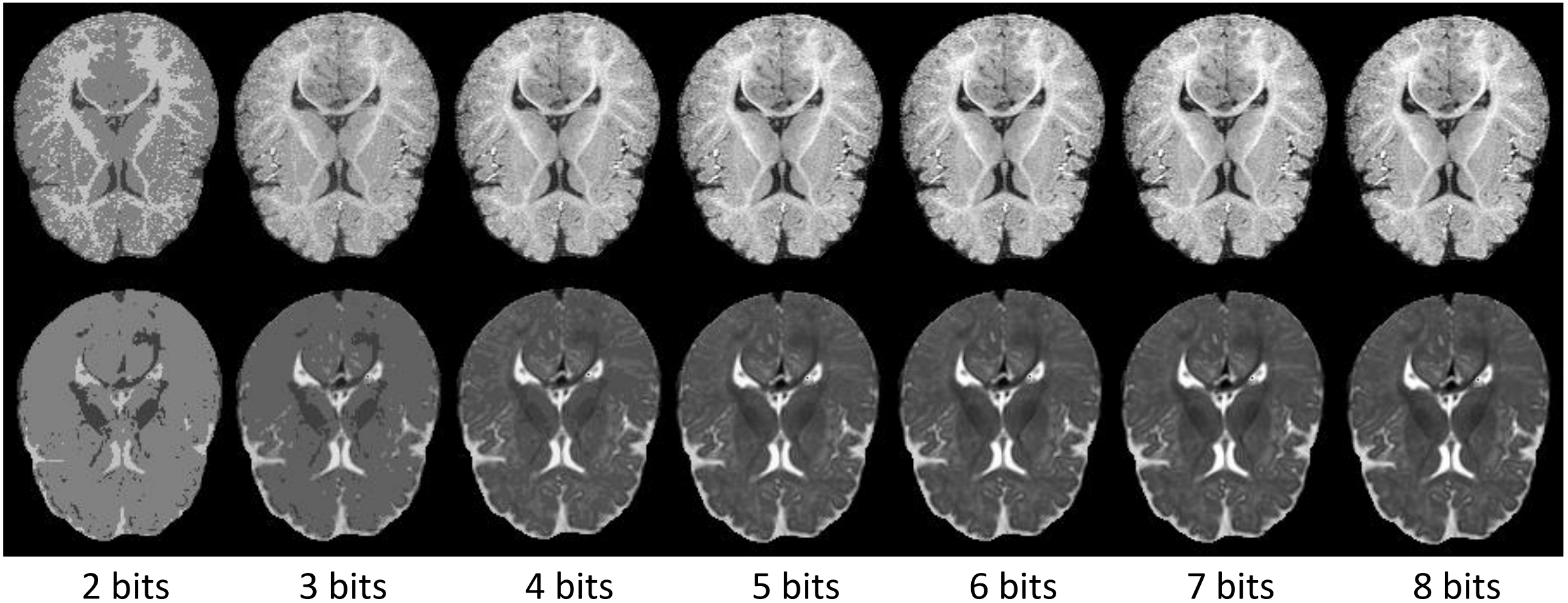
- CNN based methods only need enhancement of visibility
- Contrast Limited Adaptive Histogram Equalization (CLAHE)

# Spectral resolution

- What is the effect of color depth upon segmentation quality?
- How many bits are useful?
- Observed MRI data come at 16-bit resolution

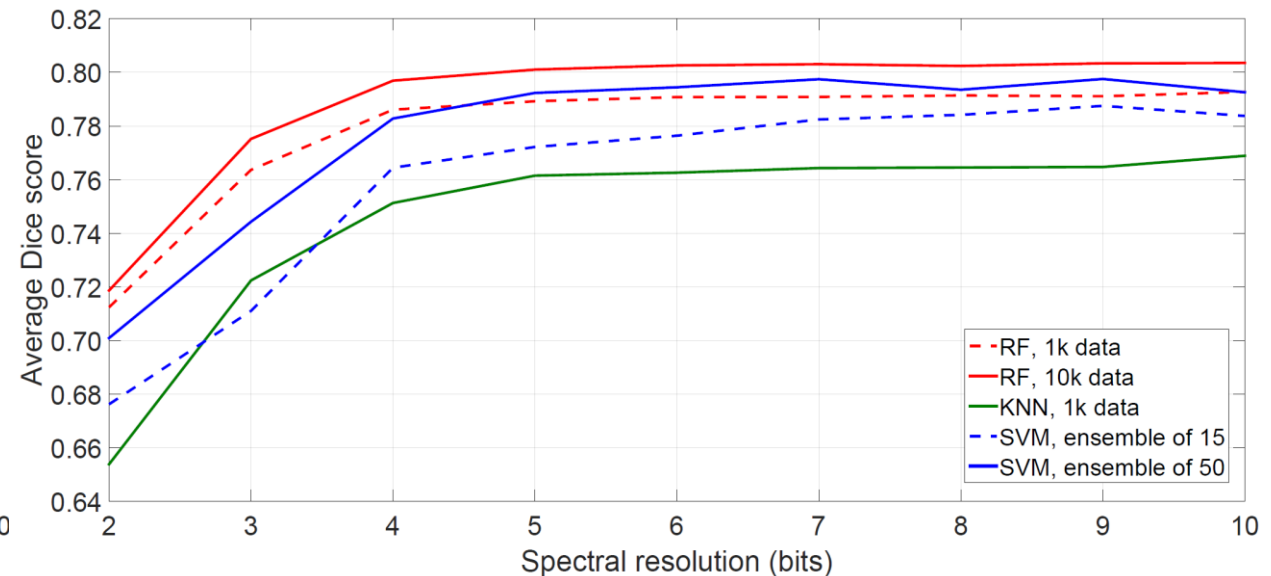
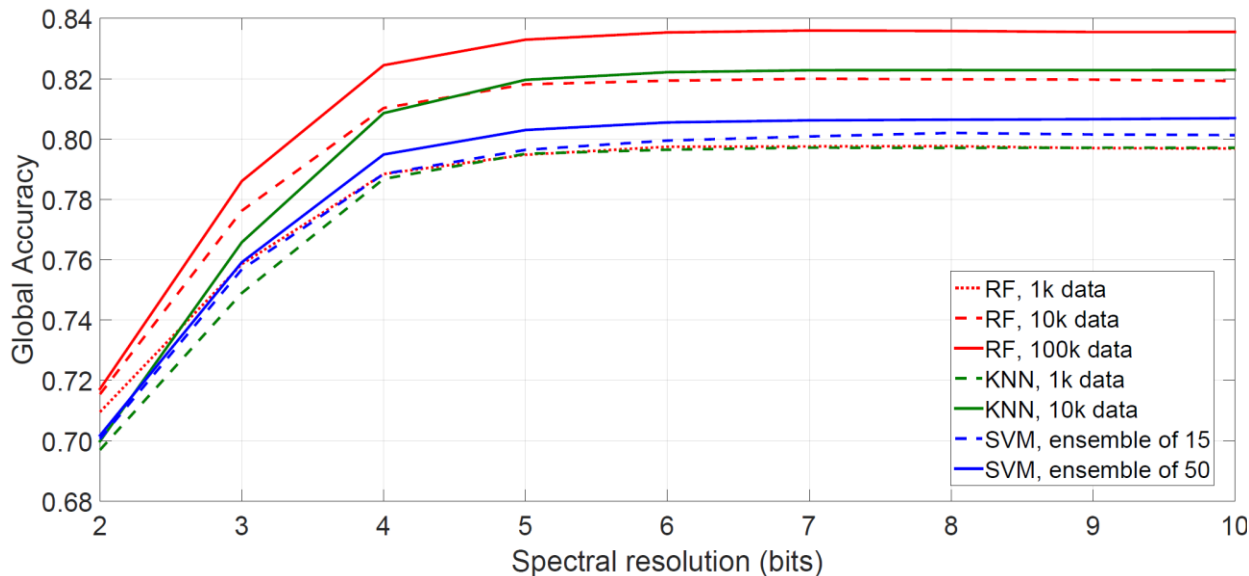


# MRI data represented at 2 to 8 bits spectral resolution



# Results

- Above 6-bit resolution the segmentation quality saturates
- Multi-channel preprocessed data can be efficiently stored in single byte per feature
- Reduce the archive storage space by 50%



# Feature generation and selection

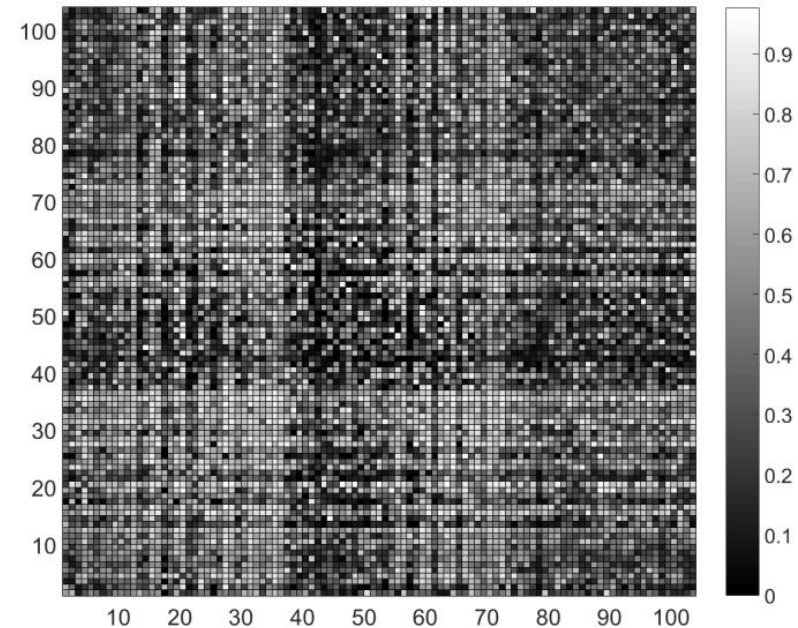
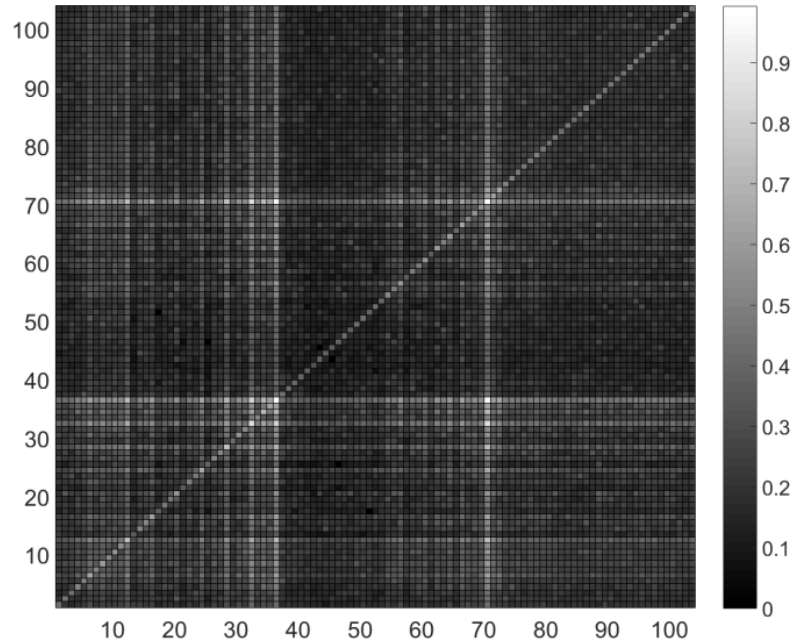
- CNN-based methods extract the features they use
- Classical machine learning based methods use handcrafted features
- 4 observed features (T1, T2, T1C, FLAIR)
- $4 \times 25 = 100$  computed features

| Neighborhood          | Average | Maximum | Minimum | Median | Gradient | Gabor | Total |
|-----------------------|---------|---------|---------|--------|----------|-------|-------|
| $3 \times 3 \times 3$ | 4       | 4       | 4       |        |          |       | 12    |
| $3 \times 3$          | 4       |         |         | 4      |          |       | 8     |
| $5 \times 5$          | 4       |         |         | 4      |          |       | 8     |
| $7 \times 7$          | 4       |         |         | 4      | 16       |       | 24    |
| $9 \times 9$          | 4       |         |         | 4      |          |       | 8     |
| $11 \times 11$        | 4       |         |         | 4      |          | 32    | 40    |
| Total                 | 24      | 4       | 4       | 20     | 16       | 32    | 100   |



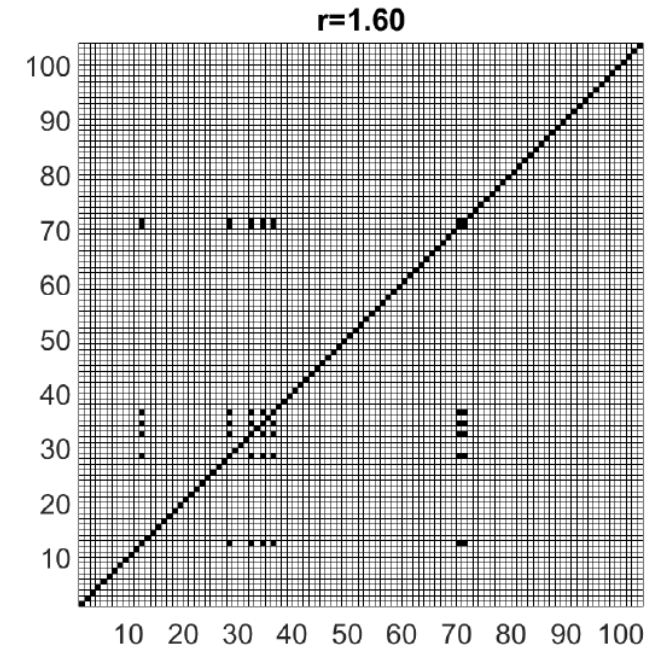
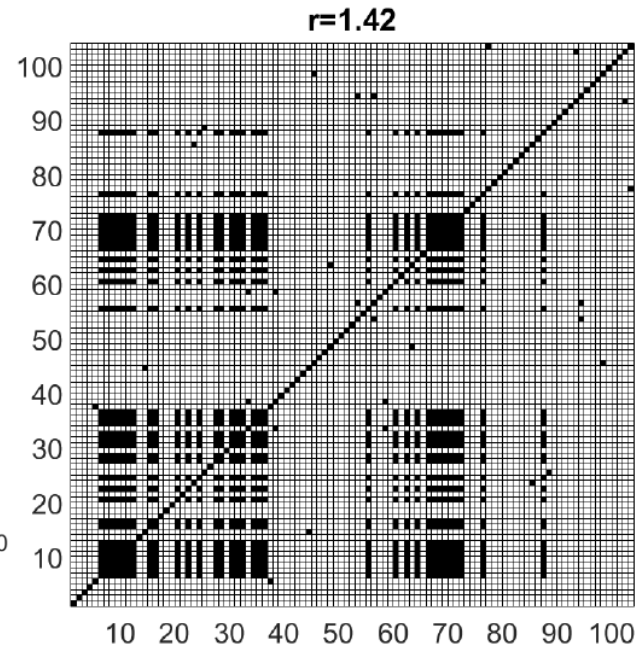
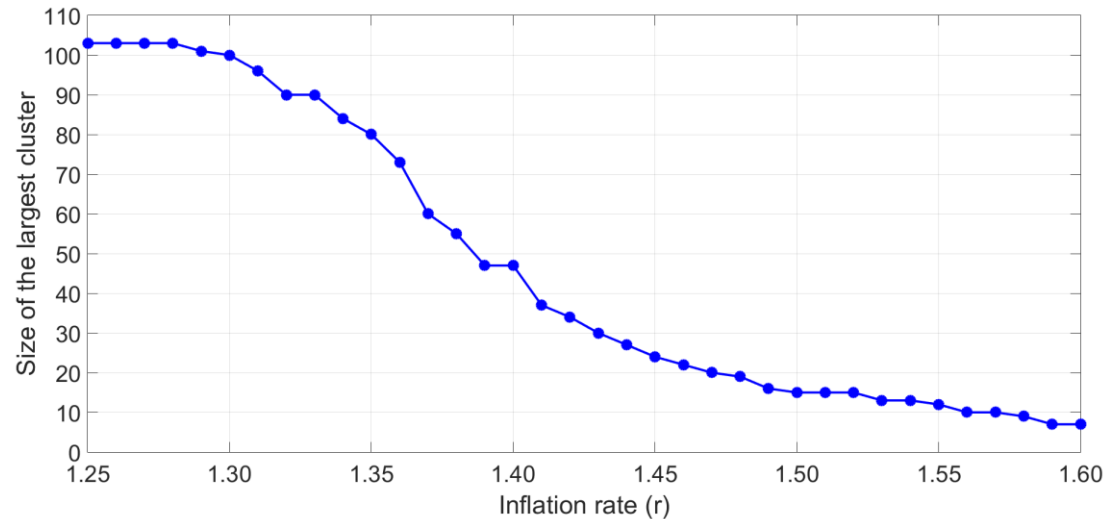
# Feature selection

- Markov clustering based on pairwise “similarity” data
- Similarity of features
  - How often they appear together in making a decision
  - How often these decisions are correct

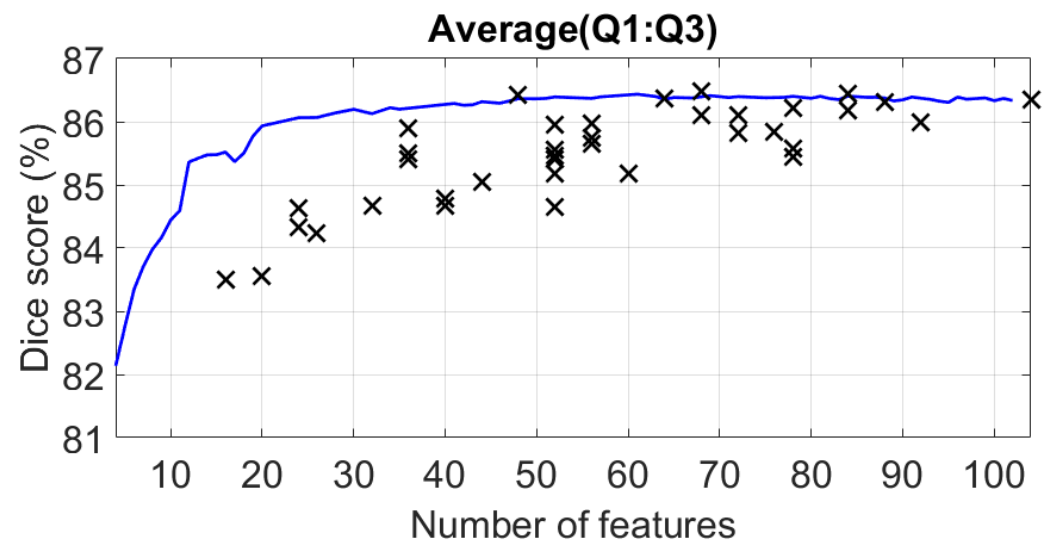
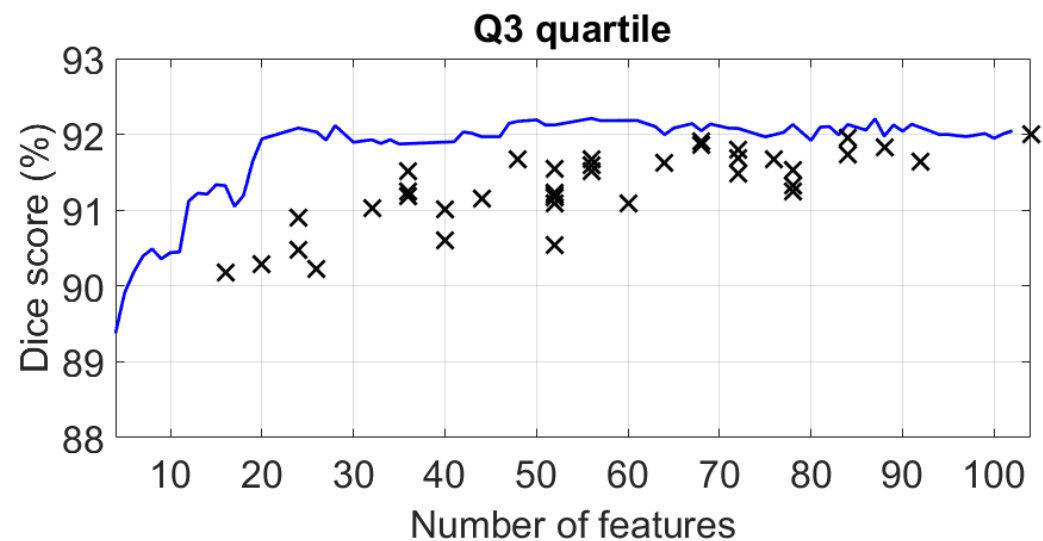
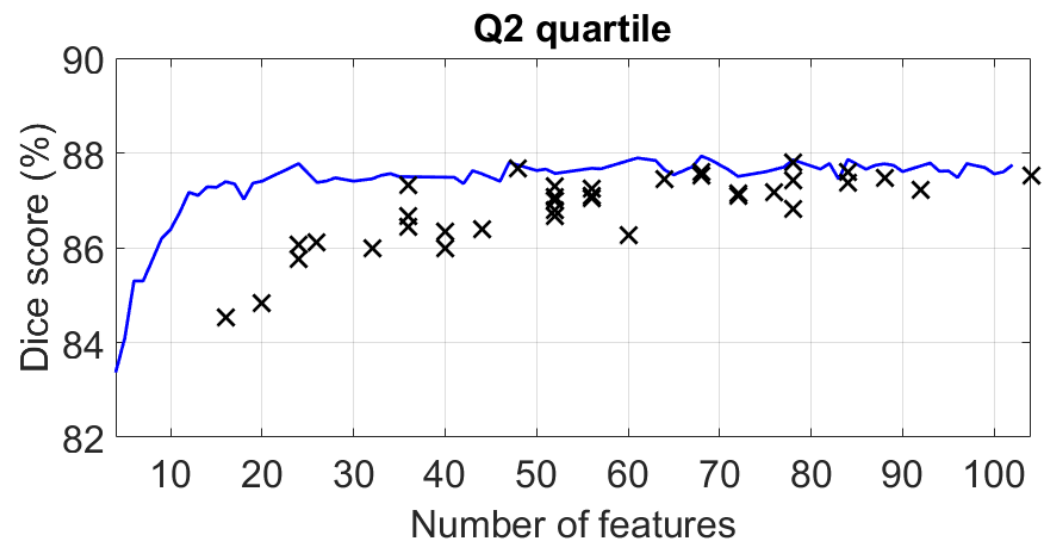
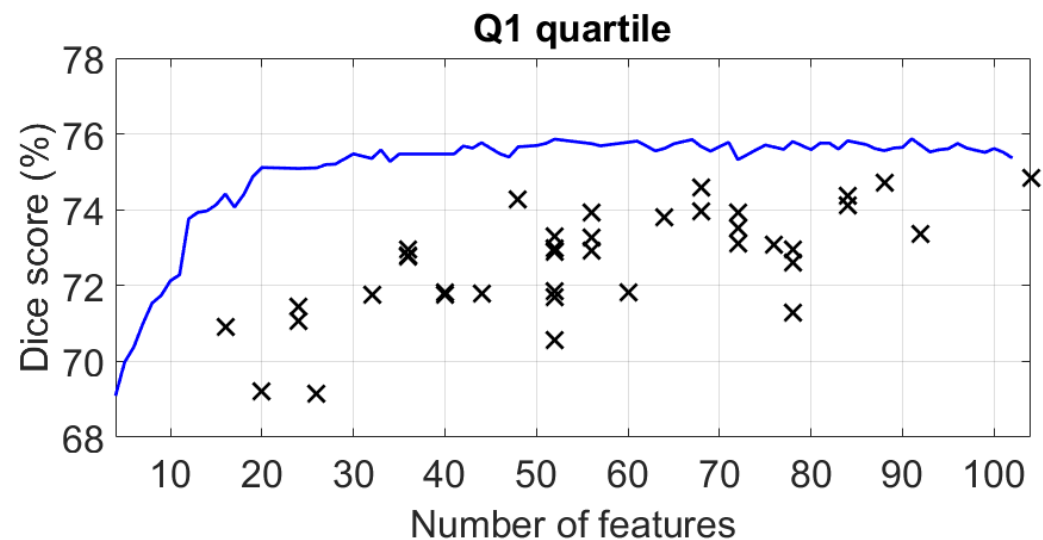


# Reduced sets of features

- Markov clustering has a single parameter: inflation rate  $r$
- Size of largest cluster of features depends on  $r$

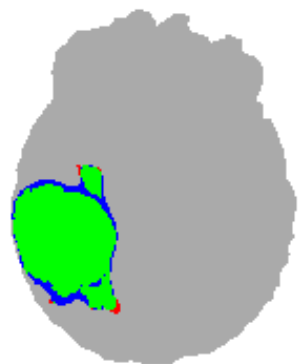


# Segmentation quality



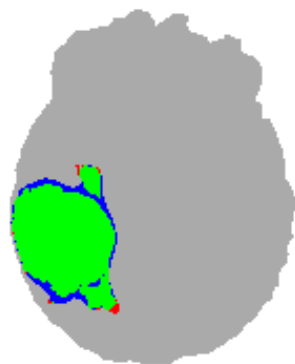
**False  
Positives**

**11 features**



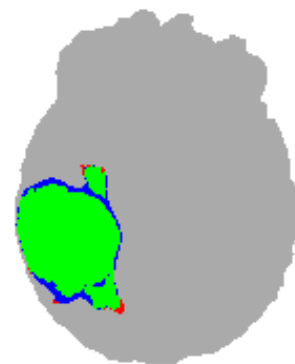
**336 2206 28**  
**DSC=92.38%**

**26 features**



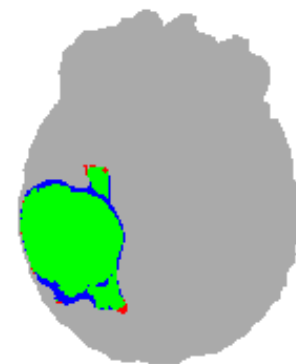
**309 2200 34**  
**DSC=92.77%**

**41 features**



**292 2197 37**  
**DSC=93.03%**

**104 features**

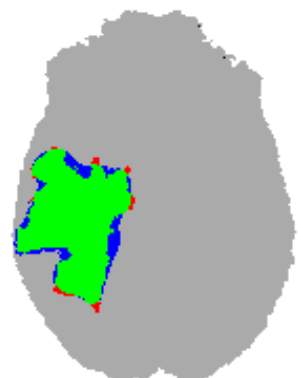


**295 2198 36**  
**DSC=92.99%**

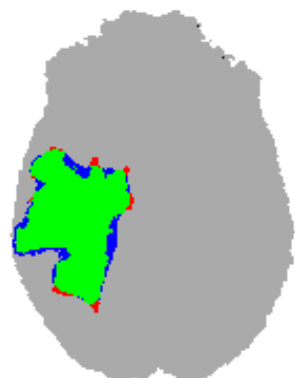
**True  
Positives**

**False  
Negatives**

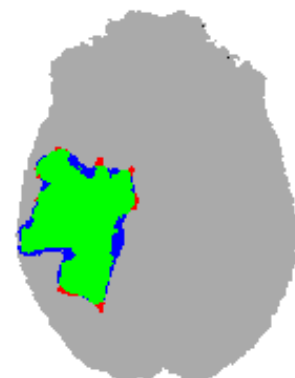
**Dice  
Score**



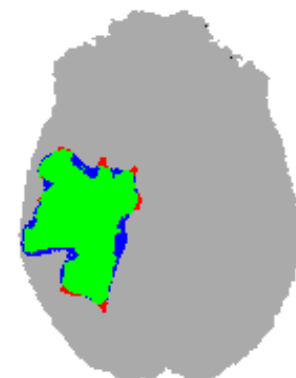
**348 2240 70**  
**DSC=91.47%**



**321 2229 81**  
**DSC=91.73%**

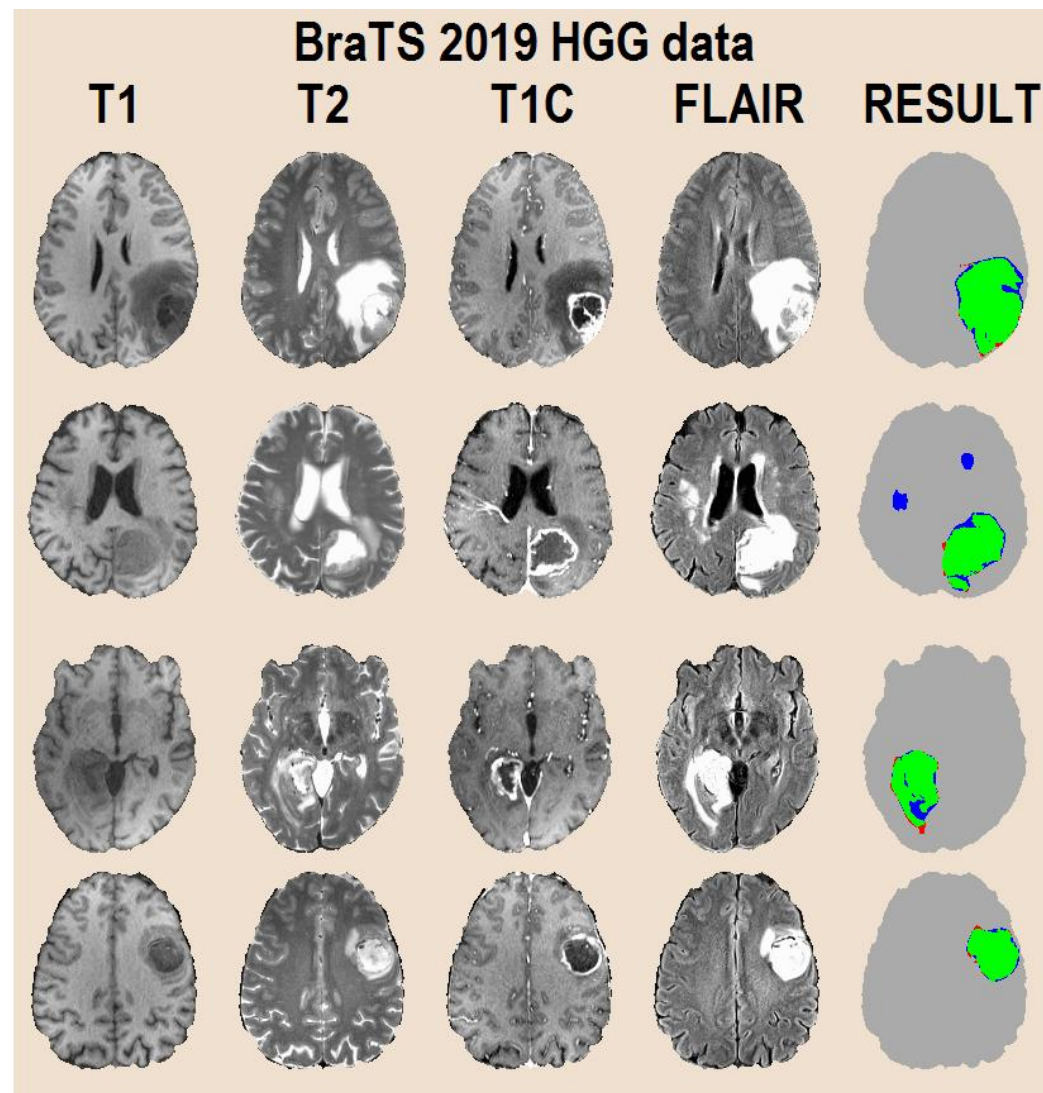
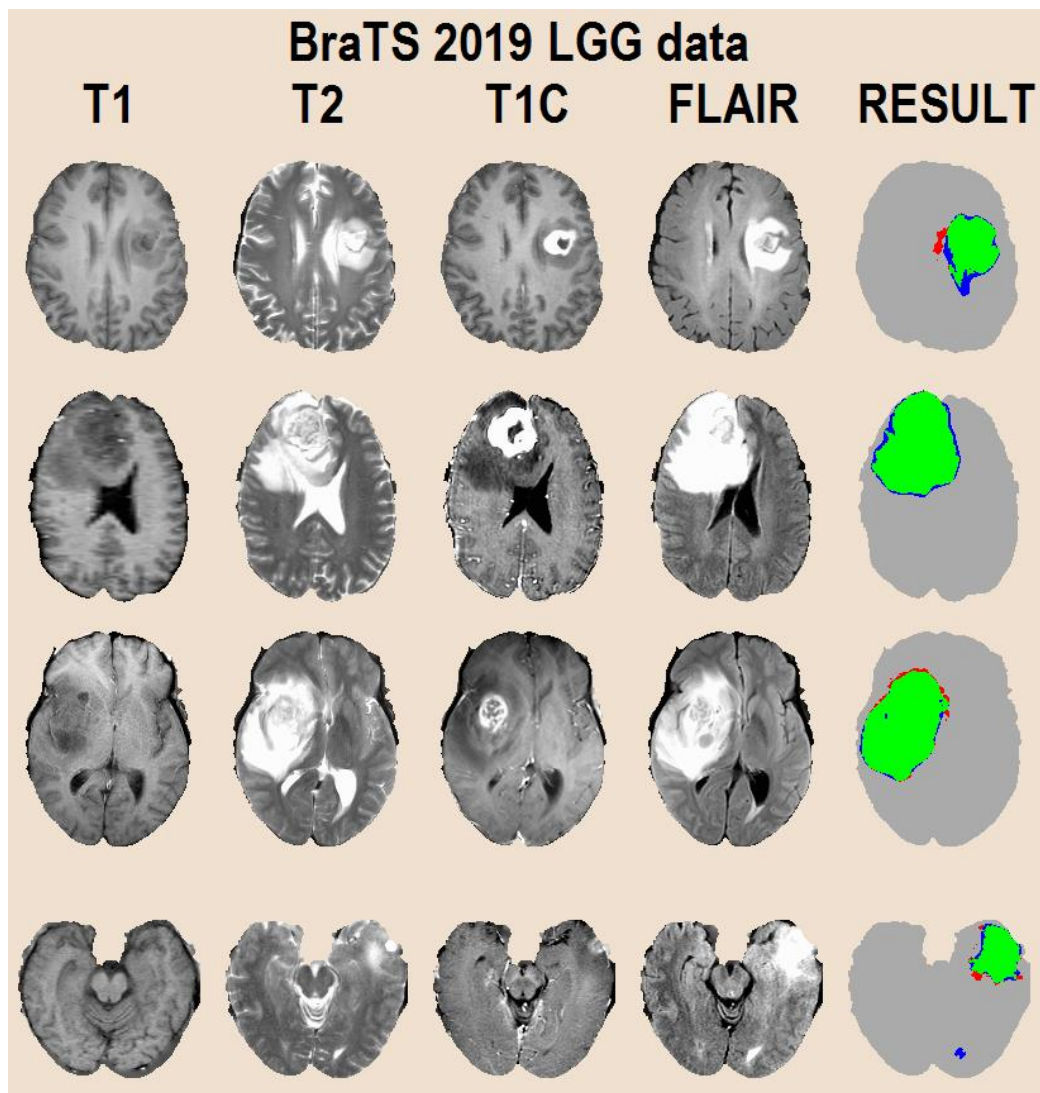


**337 2231 79**  
**DSC=91.47%**



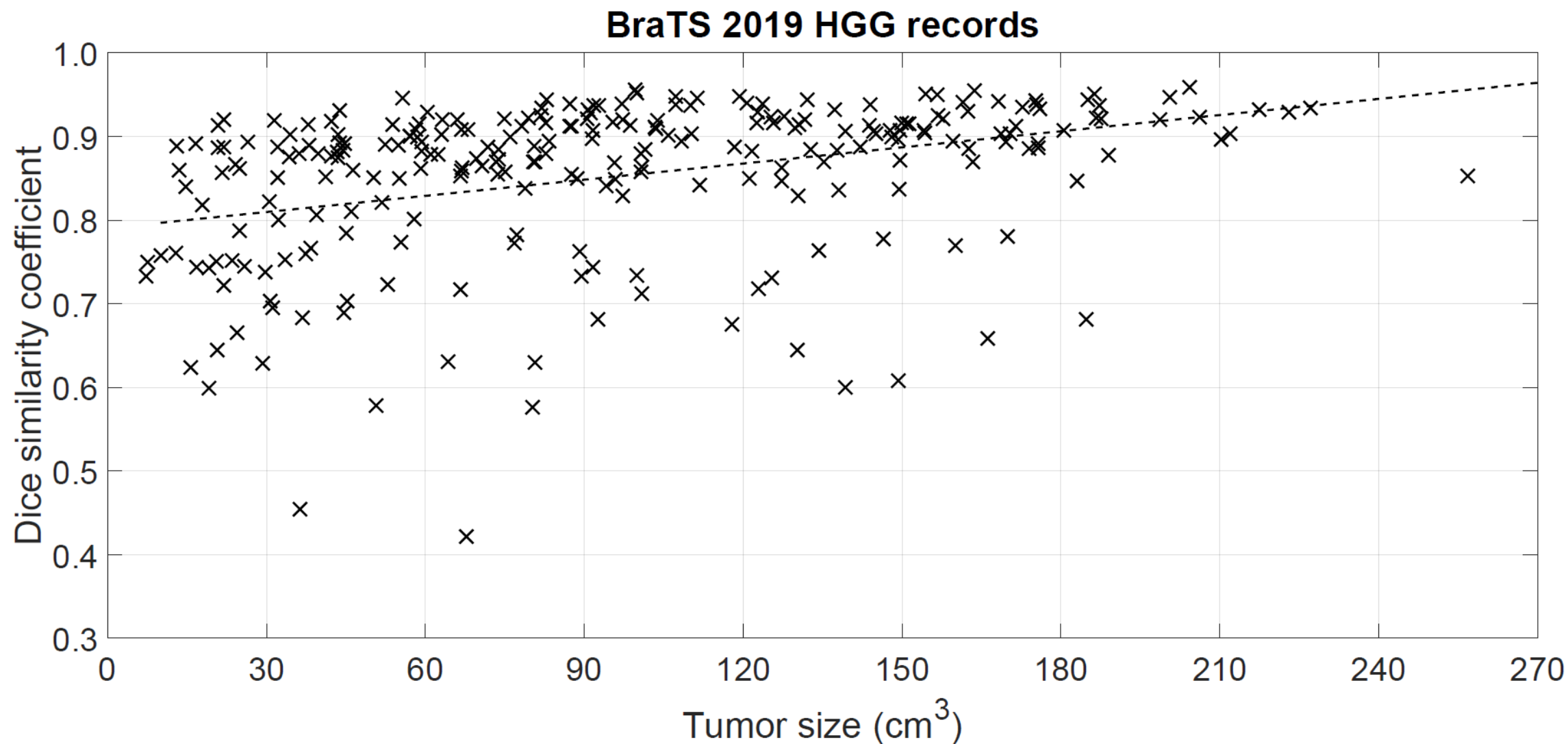
**322 2228 82**  
**DSC=91.69%**

# Segmentation results (TP FP FN)

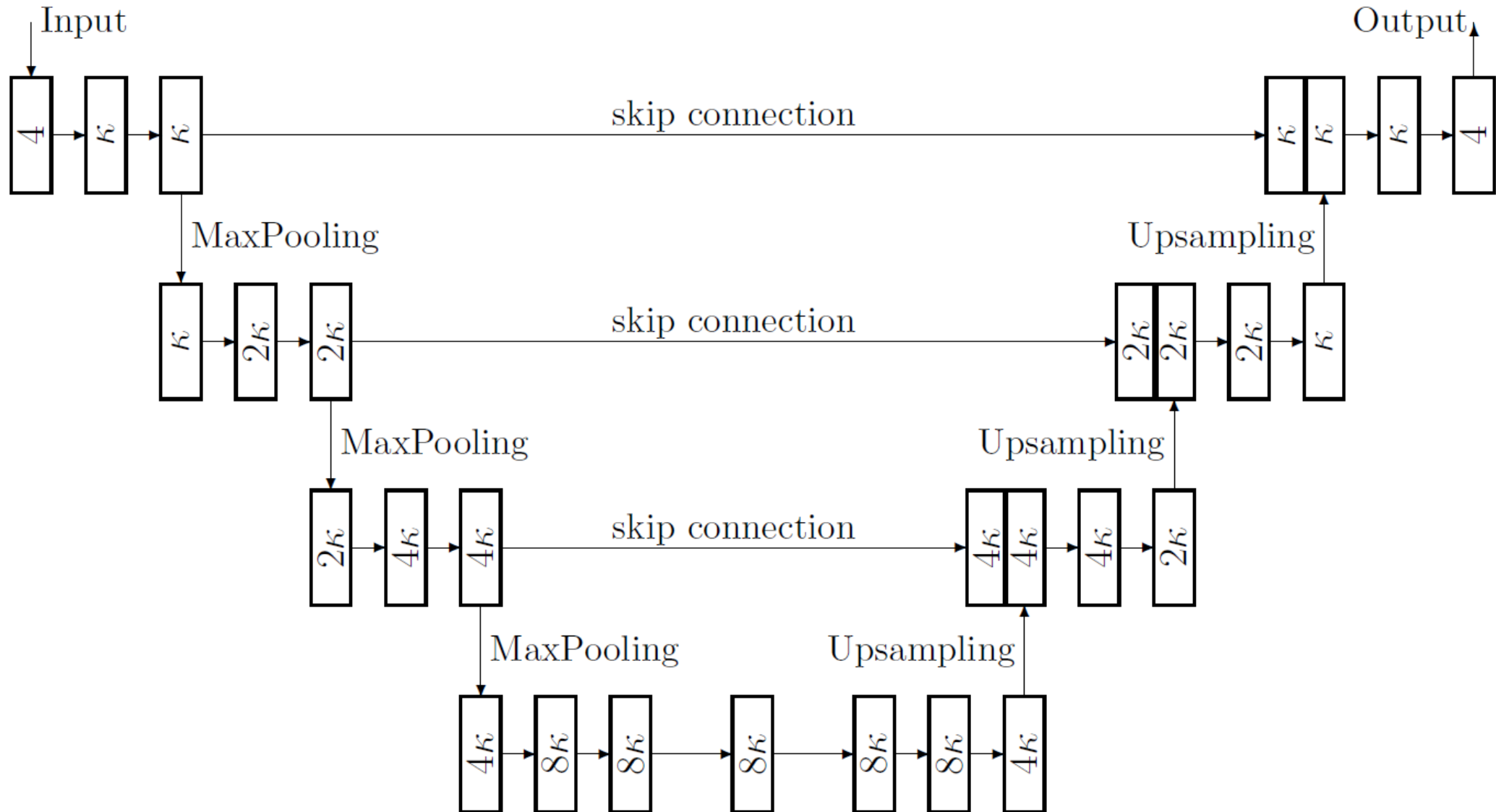




# Segmentation accuracy vs. tumor size (HGG)



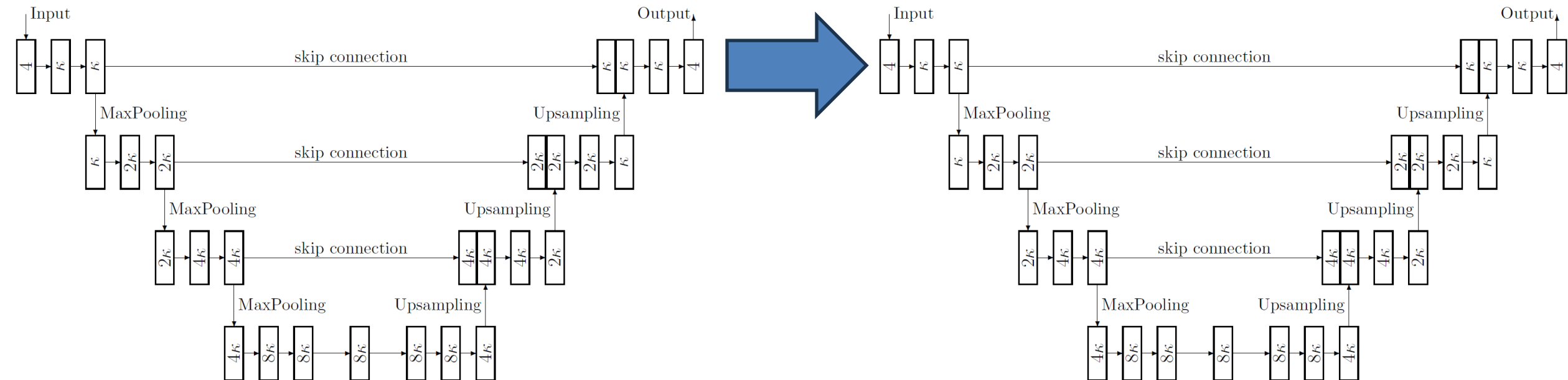
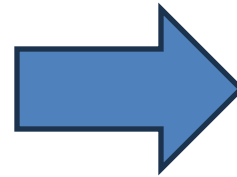
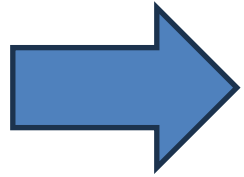
# U-net architecture





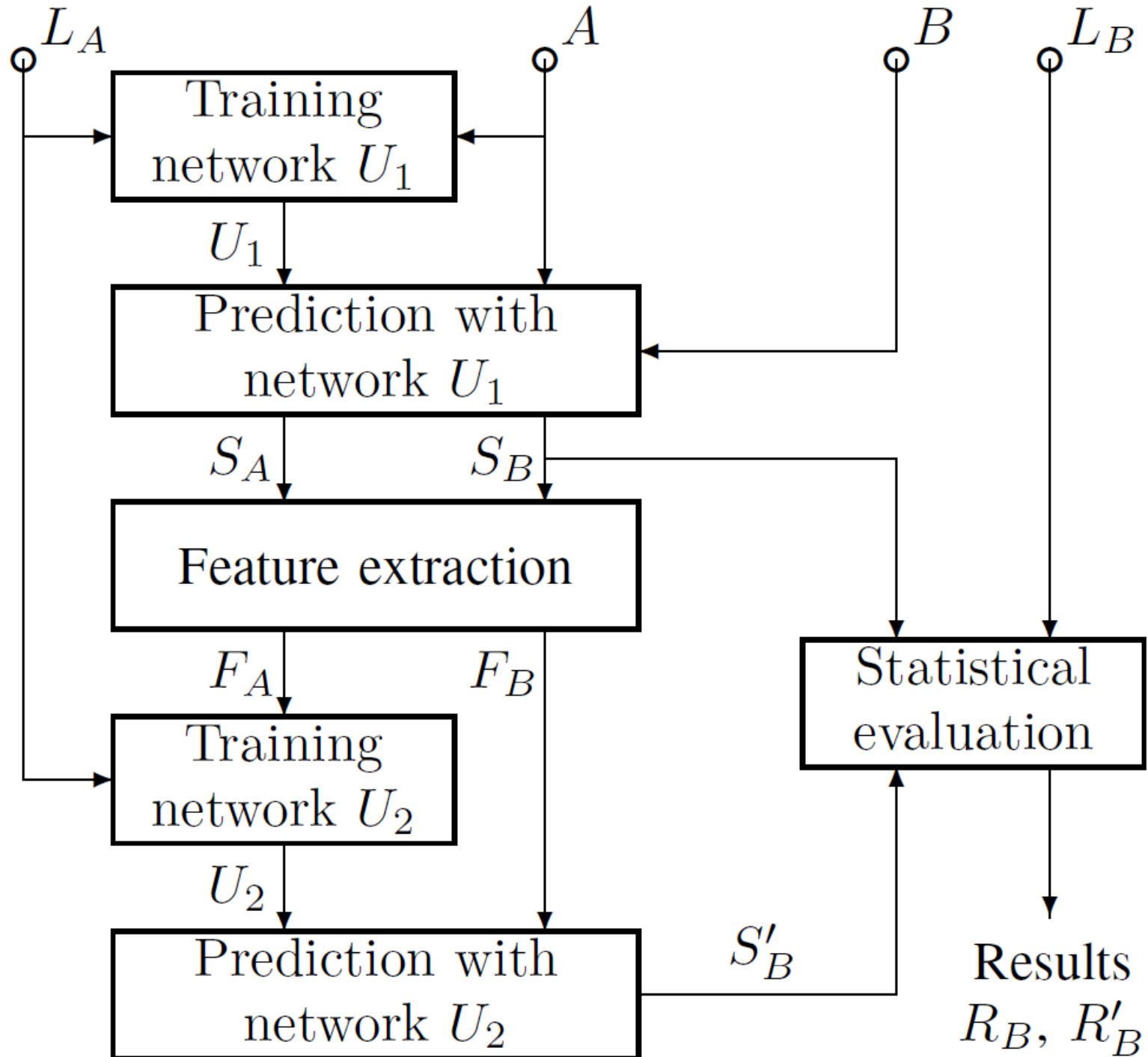
# Proposed cascade

- First U-net performs a segmentation of preprocessed MRI data
- Second U-net performs post-processing using features extracted from the first segmentation outcome
- Feature extraction: percentage of positive pixels within 3x3x3, 5x5x5, 7x7x7, 9x9x9 sized neighborhoods



# Data flow

- Dataset divided in two equal groups
- Data A, B
- Label  $L_A$ ,  $L_B$
- Initial segmentations  $S_A$ ,  $S_B$
- Features  $F_A$ ,  $F_B$
- Final segmentations  $S'_A$ ,  $S'_B$
- A and B can swap roles



# Measuring accuracy

- Based on ground truth and final decision: TP, TN, FP, FN
- Accuracy indicators: DSC, TPR, TNR, ACC

| Indicator                  | Formula   |
|----------------------------|---|
| True positive rate         | $TPR_i = \frac{TP_i}{TP_i + FN_i}$                          |
| True negative rate         | $TNR_i = \frac{TN_i}{TN_i + FP_i}$                          |
| Positive predictive value  | $PPV_i = \frac{TP_i}{TP_i + FP_i}$                          |
| Dice score or $F_1$ -score | $DSC_i = \frac{2 \times TP_i}{2 \times TP_i + FP_i + FN_i}$ |
| Accuracy                   | $ACC_i = \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i}$     |

- Individual value for each volume (record)
- Average, SD, quartiles

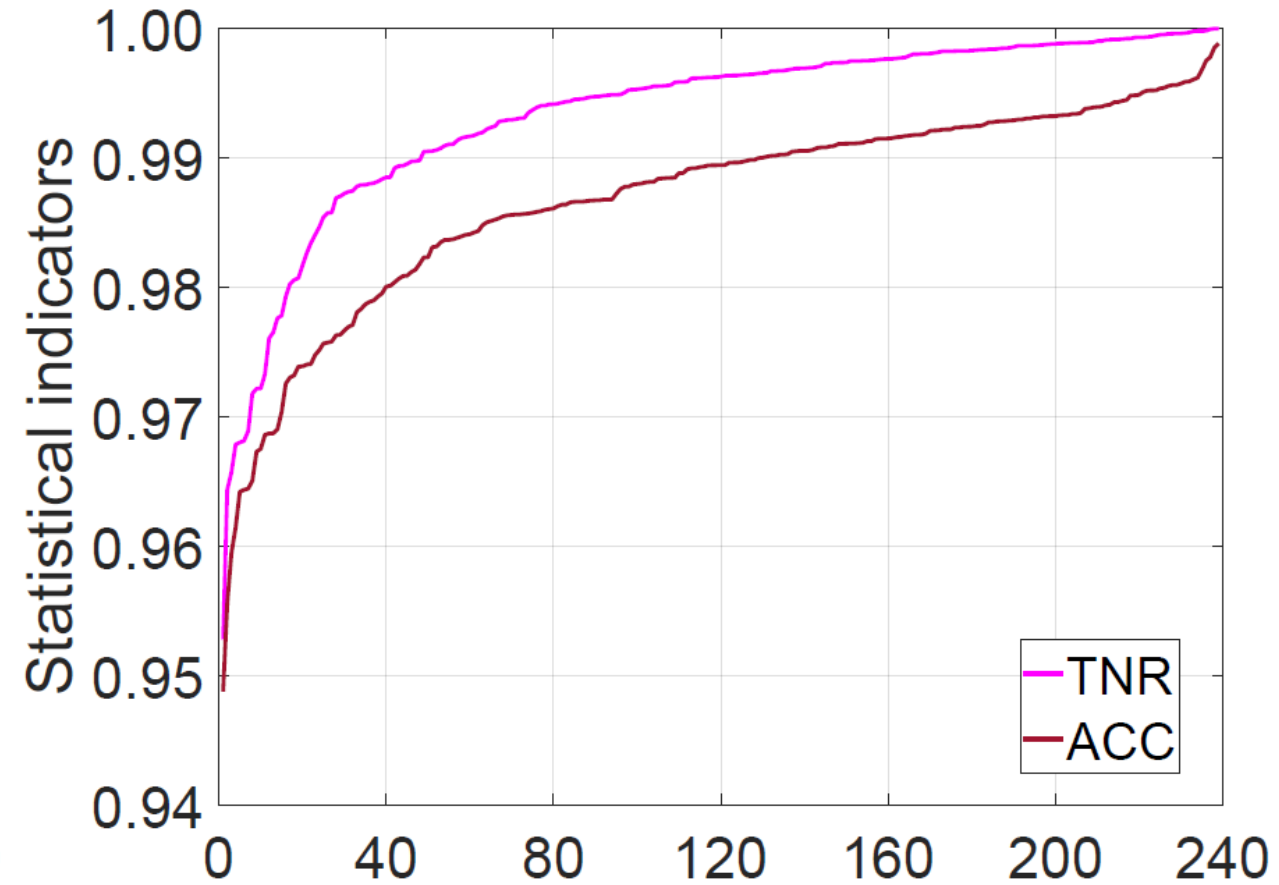
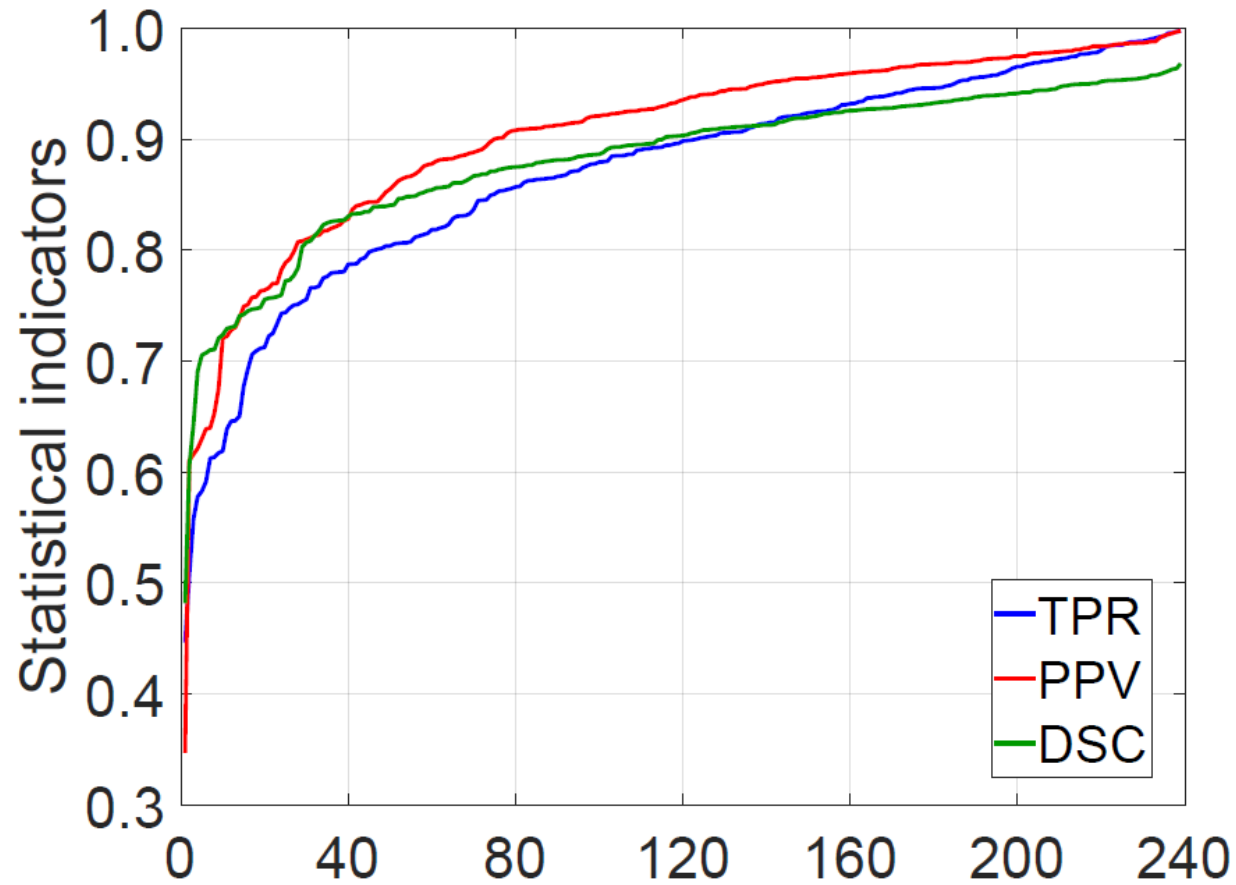
# Overall accuracy indicators

|           | Value        | DSC    | TPR    | PPV    | TNR    | ACC    |
|-----------|--------------|--------|--------|--------|--------|--------|
| Before PP | average      | 0.8768 | 0.8662 | 0.9057 | 0.9920 | 0.9864 |
|           | stdev        | 0.0776 | 0.1154 | 0.0961 | 0.0073 | 0.0094 |
|           | 1st quartile | 0.8503 | 0.8214 | 0.8751 | 0.9888 | 0.9836 |
|           | median       | 0.9024 | 0.8868 | 0.9338 | 0.9935 | 0.9891 |
|           | 3rd quartile | 0.9302 | 0.9548 | 0.9757 | 0.9973 | 0.9921 |
| After PP  | average      | 0.8879 | 0.8665 | 0.9251 | 0.9913 | 0.9870 |
|           | stdev        | 0.0767 | 0.1108 | 0.0819 | 0.0085 | 0.0094 |
|           | 1st quartile | 0.8612 | 0.8303 | 0.9070 | 0.9891 | 0.9834 |
|           | median       | 0.9116 | 0.8957 | 0.9531 | 0.9939 | 0.9901 |
|           | 3rd quartile | 0.9380 | 0.9398 | 0.9749 | 0.9964 | 0.9932 |

# Parameter selection

| DSC                   | Layer depth parameter $\kappa$ |        |        |        |        |        |        |
|-----------------------|--------------------------------|--------|--------|--------|--------|--------|--------|
|                       | 16                             | 32     | 48     | 64     | 80     | 96     | 112    |
| average               | 0.8568                         | 0.8702 | 0.8725 | 0.8629 | 0.8707 | 0.8768 | 0.8688 |
| stdev                 | 0.0868                         | 0.0757 | 0.0769 | 0.0944 | 0.0780 | 0.0774 | 0.0798 |
| 1st quartile          | 0.8255                         | 0.8509 | 0.8464 | 0.8386 | 0.8379 | 0.8503 | 0.8419 |
| median                | 0.8809                         | 0.8942 | 0.8956 | 0.8921 | 0.8942 | 0.9024 | 0.8910 |
| 3rd quartile          | 0.9164                         | 0.9215 | 0.9260 | 0.9240 | 0.9263 | 0.9302 | 0.9259 |
| ranked 1st            | 15                             | 29     | 33     | 19     | 23     | 58     | 62     |
| ranked 2nd            | 20                             | 39     | 26     | 31     | 52     | 52     | 19     |
| ranked 3rd            | 15                             | 23     | 41     | 33     | 59     | 33     | 35     |
| Total                 | 50                             | 91     | 100    | 83     | 134    | 143    | 116    |
| $\text{DSC}_i > 0.93$ | 31                             | 40     | 52     | 47     | 52     | 61     | 51     |
| $\text{DSC}_i > 0.9$  | 86                             | 109    | 110    | 105    | 112    | 122    | 107    |
| $\text{DSC}_i > 0.85$ | 160                            | 181    | 177    | 170    | 175    | 180    | 169    |
| $\text{DSC}_i > 0.8$  | 193                            | 201    | 203    | 196    | 204    | 202    | 200    |

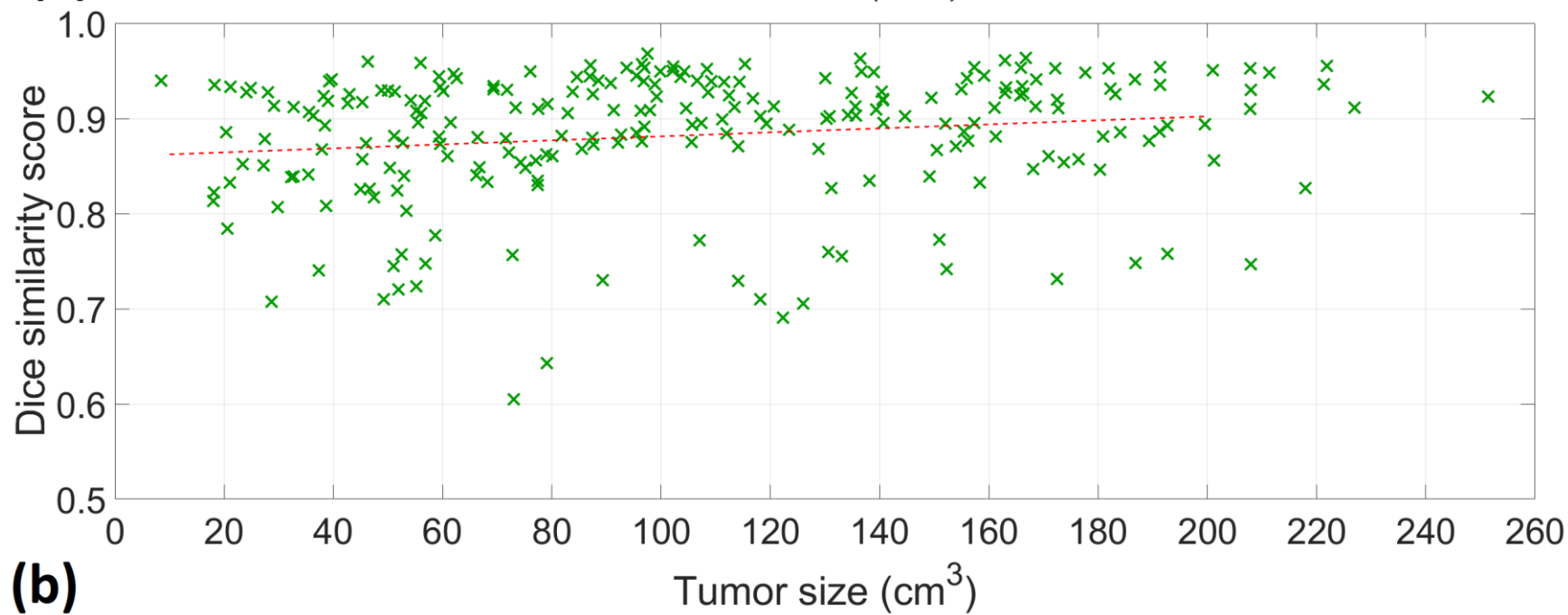
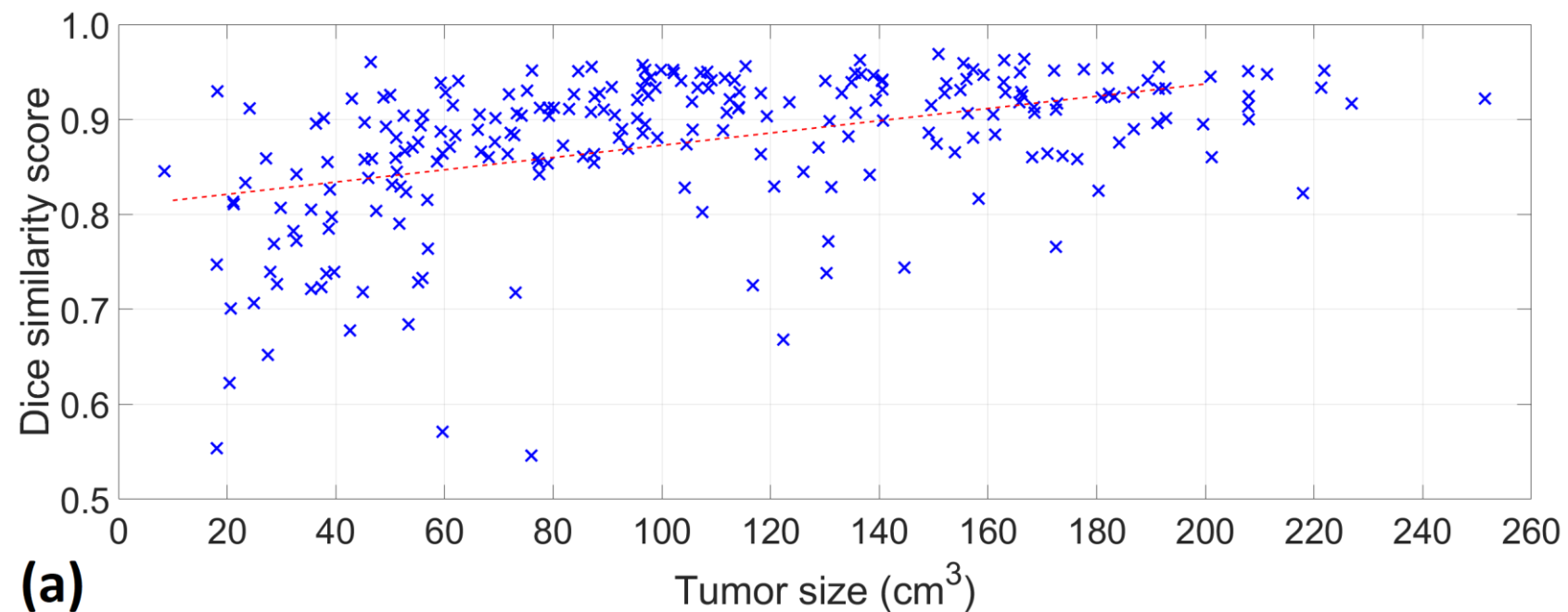
# Accuracy Indicators for Individual Volumes



Patients in increasing order of the statistical indicators

# Effect of Post- processing

| Tumor size | Before PP | After PP |
|------------|-----------|----------|
| 10 $cm^3$  | 0.8147    | 0.8622   |
| 20 $cm^3$  | 0.8212    | 0.8643   |
| 50 $cm^3$  | 0.8405    | 0.8706   |
| 100 $cm^3$ | 0.8728    | 0.8811   |



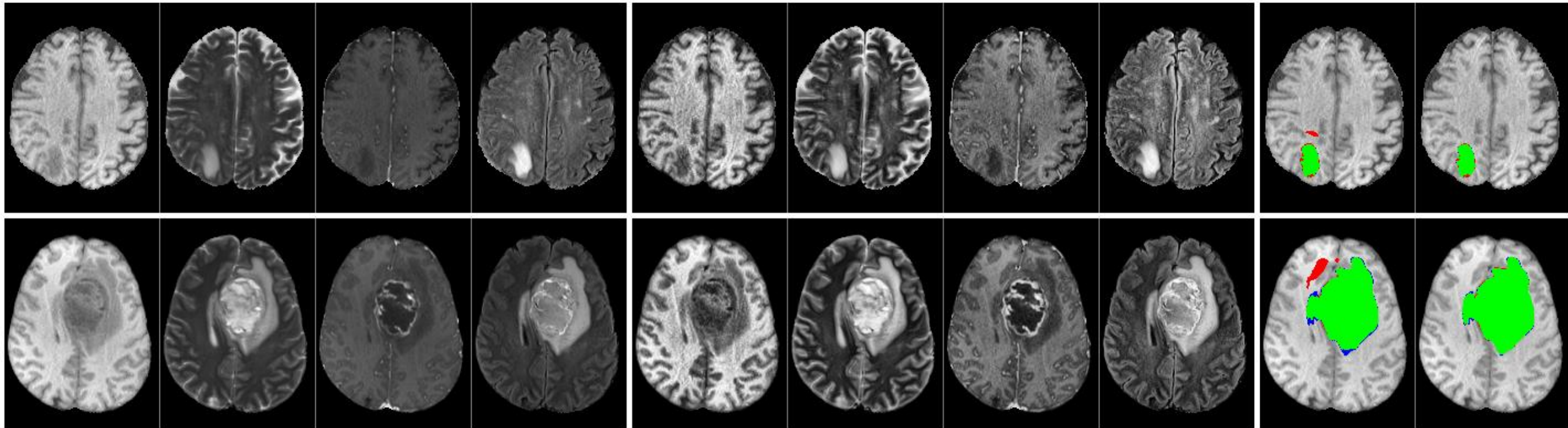


Detected tumors: TP FP FN

Input

Preprocessed

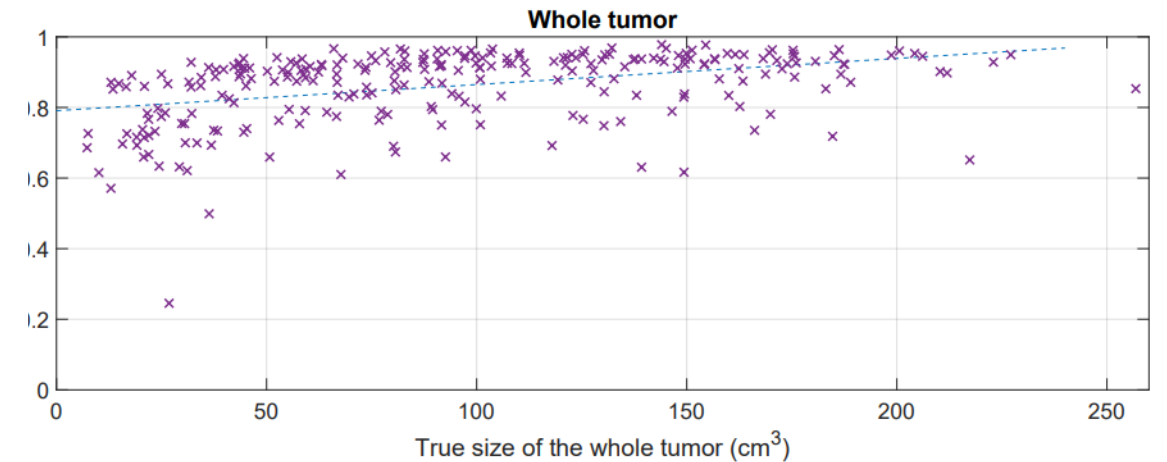
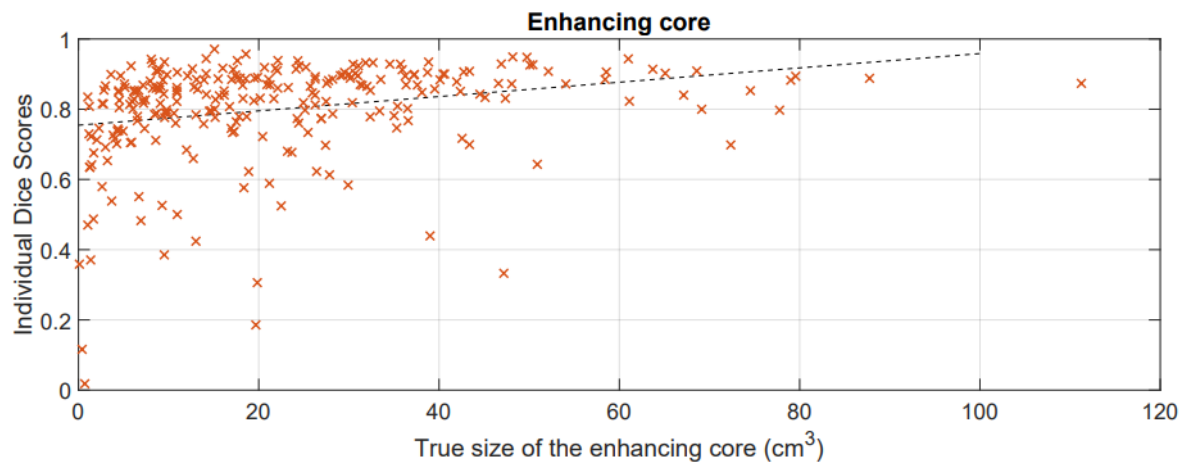
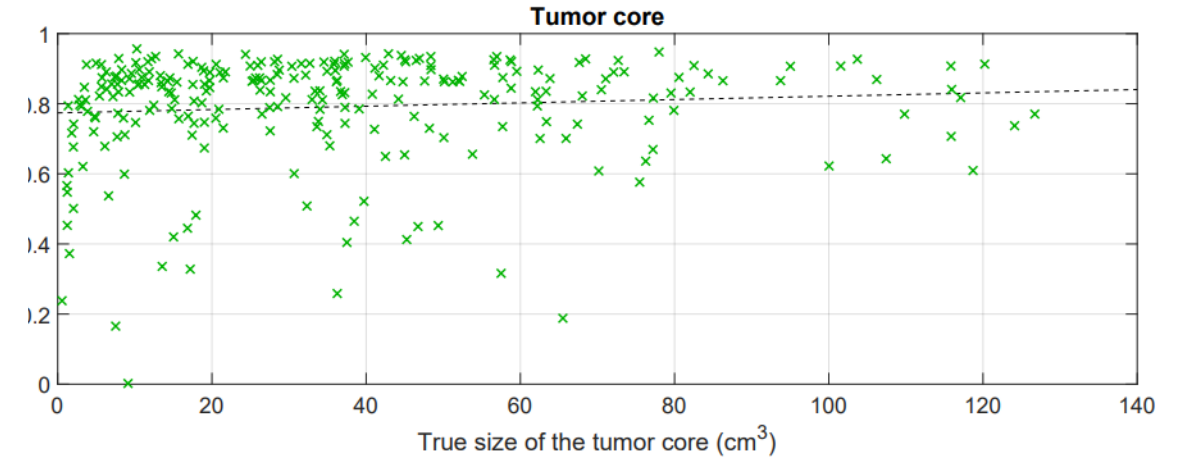
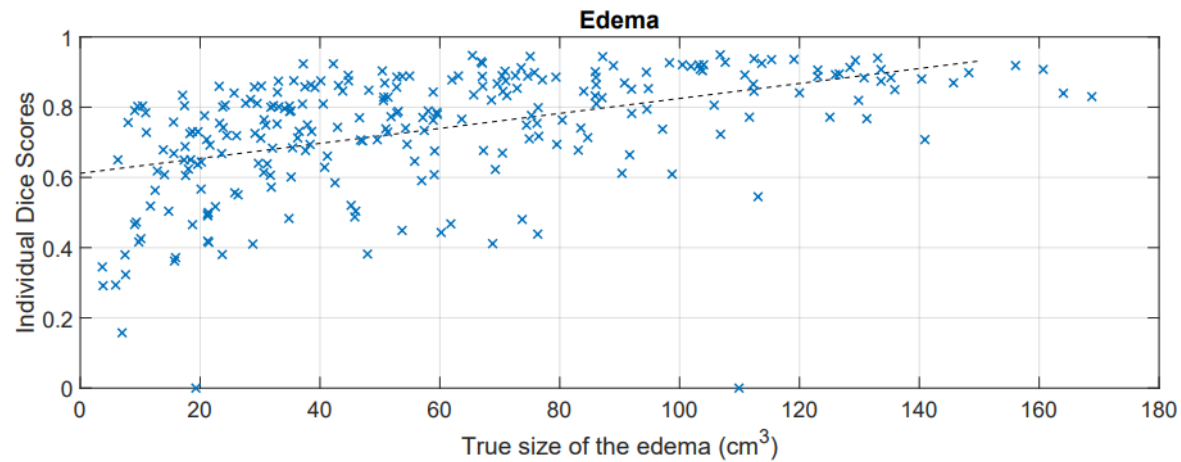
Segmented



# Comparison

| Paper                        | Classifier        | Data          | Mean DSC |
|------------------------------|-------------------|---------------|----------|
| Tustison et al. [10] (2015)  | RF, MRF           | BraTS 2013    | 0.87     |
| Pereira et al. [17] (2016)   | CNN               |               | 0.88     |
| Lefkovits et al. [30] (2017) | RF                |               | 0.868    |
| Havaei et al. [31] (2017)    | deep CNN          |               | 0.88     |
| Pinto et al. [14] (2018)     | ERT               |               | 0.85     |
| Pereira et al. [32] (2019)   | FCNN              |               | 0.86     |
| Pereira et al. [17] (2016)   | CNN               | BraTS 2015    | 0.78     |
| Kamnitsas et al. [20] (2017) | deep CNN          |               | 0.849    |
| Zhao et al. [18] (2018)      | FCNN, CRF         |               | 0.84     |
| Chen et al. [33] (2019)      | CNN               |               | 0.85     |
| Ding et al. [34] (2019)      | deep ResNet       |               | 0.86     |
| Wu et al. [19] (2020)        | CNN               |               | 0.83     |
| Györfi et al. [35] (2021)    | BDT ensemble      |               | 0.8355   |
| Bhalerao et al. [36] (2020)  | 3D Residual U-Net | BraTS 2019/20 | 0.85269  |
| Wang et al. [37] (2020)      | 3D U-Net          |               | 0.894    |
| Guo et al. [38] (2020)       | CNN + fusion      |               | 0.872    |
| Györfi et al. [35] (2021)    | BDT ensemble      |               | 0.8516   |
| Lefkovits et al. [39] (2022) | CNN ensemble      |               | 0.8780   |
| Proposed method              | U-net cascade     |               | 0.8879   |

# Dice similarity scores of various tumor parts

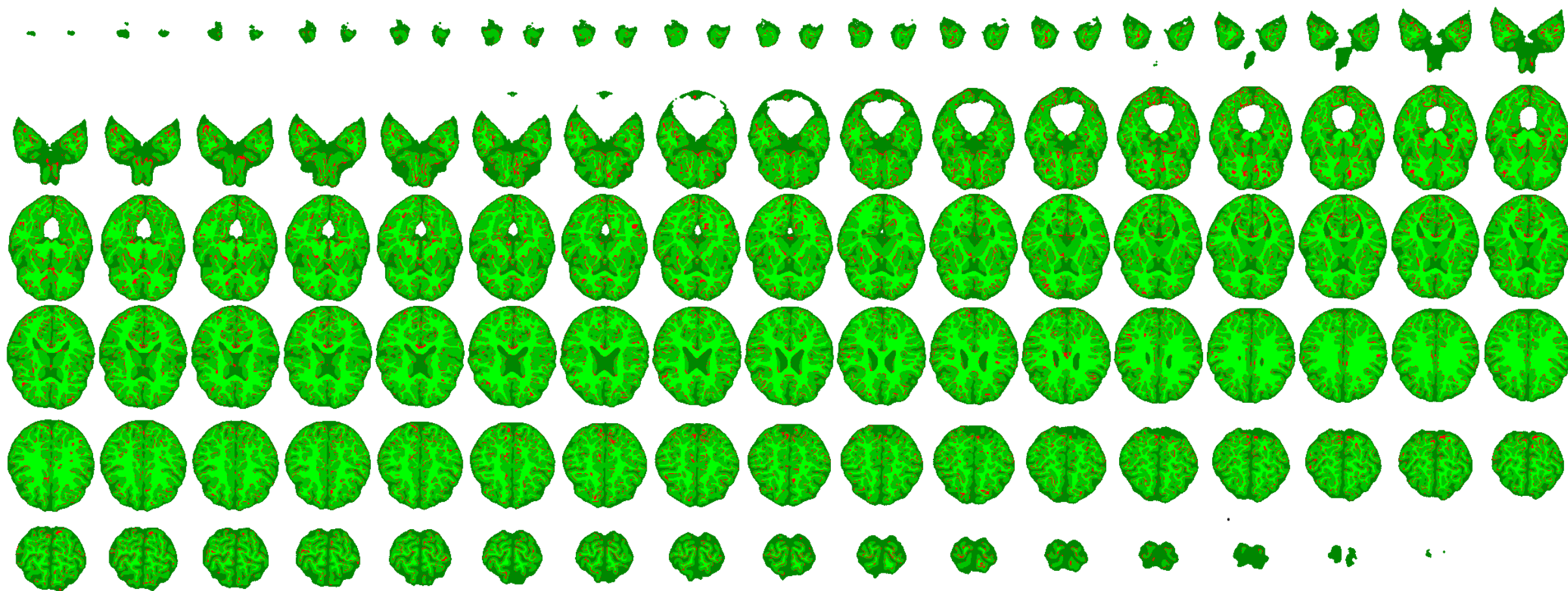


# Comparison

| Paper                        | Classifier    | Data set      | Test items | Average Dice score |        |        |        |
|------------------------------|---------------|---------------|------------|--------------------|--------|--------|--------|
|                              |               |               |            | ED                 | EC     | TC     | WT     |
| Csaholczi <i>et al.</i> [29] | random forest | BraTS15 train | 220        | 0.6566             | 0.6728 | 0.6554 | 0.7722 |
| Unpublished                  | XGBoost       | BraTS19 train | 259        | 0.7112             | 0.7895 | 0.7795 | 0.8339 |
| Kamnitsas <i>et al.</i> [16] | CNN ensemble  | BraTS15 train | 274        | N/A                | 0.728  | 0.754  | 0.901  |
| Kamnitsas <i>et al.</i> [16] | CNN ensemble  | BraTS15 test  | 110        | N/A                | 0.634  | 0.667  | 0.849  |
| Ding <i>et al.</i> [17]      | ResNet        | BraTS15 test  | 93         | N/A                | 0.63   | 0.71   | 0.86   |
| Bhalerao <i>et al.</i> [18]  | 3D Res U-net  | BraTS19 test  | 125        | N/A                | 0.697  | 0.772  | 0.828  |
| Wang <i>et al.</i> [30]      | 3D U-net      | BraTS19 test  | 125        | N/A                | 0.778  | 0.798  | 0.852  |
| Lefkovits <i>et al.</i> [31] | CNN ensemble  | BraTS19 train | 259        | 0.8005             | 0.7671 | N/A    | 0.878  |
| Proposed                     | U-net         | BraTS19 train | 259        | 0.7368             | 0.8005 | 0.7912 | 0.8612 |



# Brain tissue segmentation with U-net

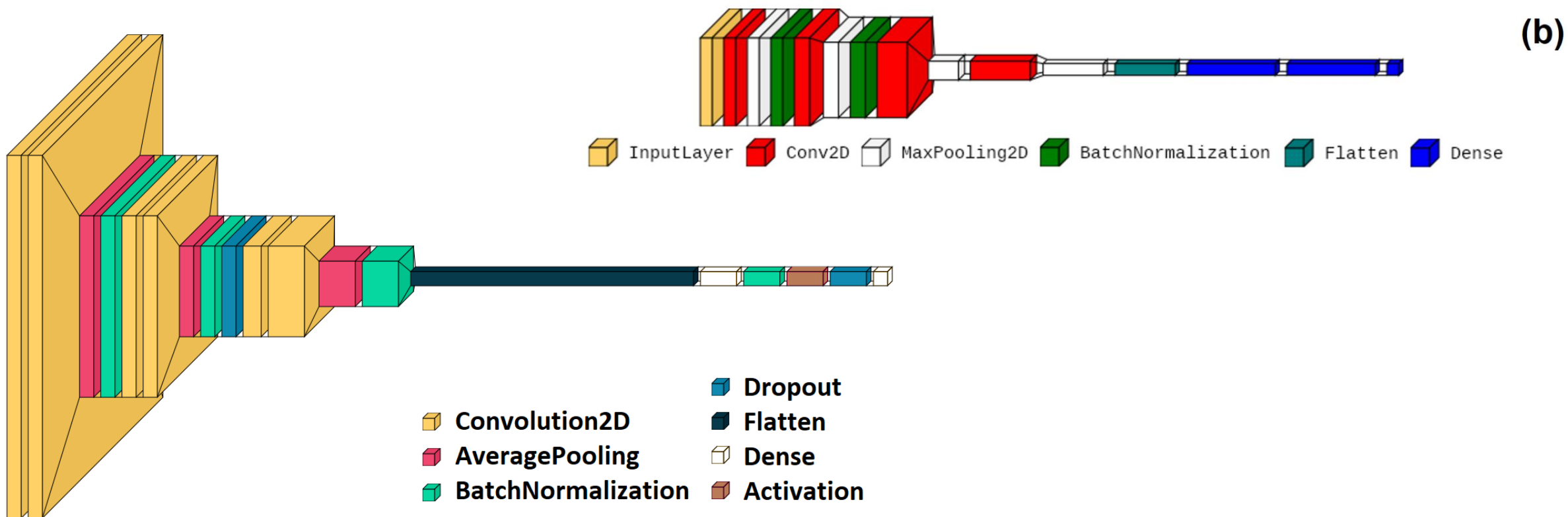
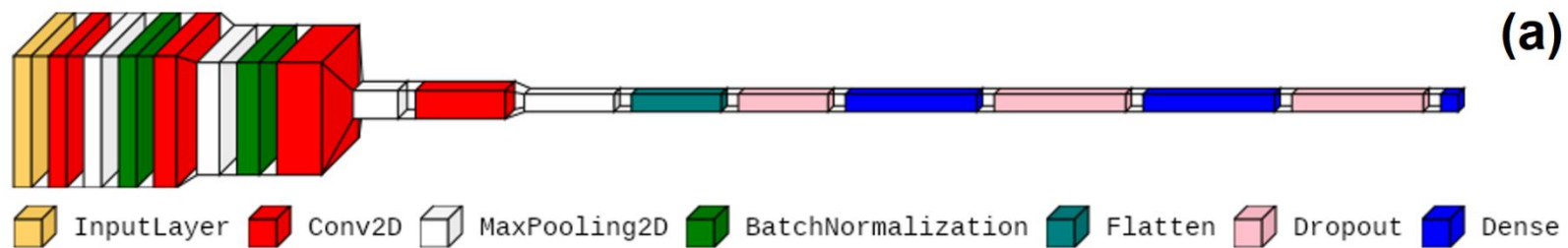


# Input Data for Tumor Classification

| Problem      | Glioma | Meningioma | Pituitary | Positives | Negatives | Total |
|--------------|--------|------------|-----------|-----------|-----------|-------|
| 2-class [26] | -      | -          | -         | 1536      | 1500      | 3036  |
| 3-class [27] | 1424   | 710        | 930       | -         | -         | 3064  |
| 4-class [28] | 826    | 822        | 827       | -         | 395       | 2870  |

- [26] [www.kaggle.com/datasets/ahmedhamada0/brain-tumor-detection](https://www.kaggle.com/datasets/ahmedhamada0/brain-tumor-detection), last visited on 14 December 2023.
- [27] M. Nickparvar, “Brain Tumor MRI Dataset [Data set],” Kaggle, 2021, <https://doi.org/10.34740/KAGGLE/DSV/2645886>
- [28] S. Bhuvaji, A. Kadam, P. Bhumkar, S. Dedge, and S. Kanchan, “Brain Tumor Classification (MRI) [Data set],” Kaggle, 2020, <https://doi.org/10.34740/KAGGLE/DSV/1183165>

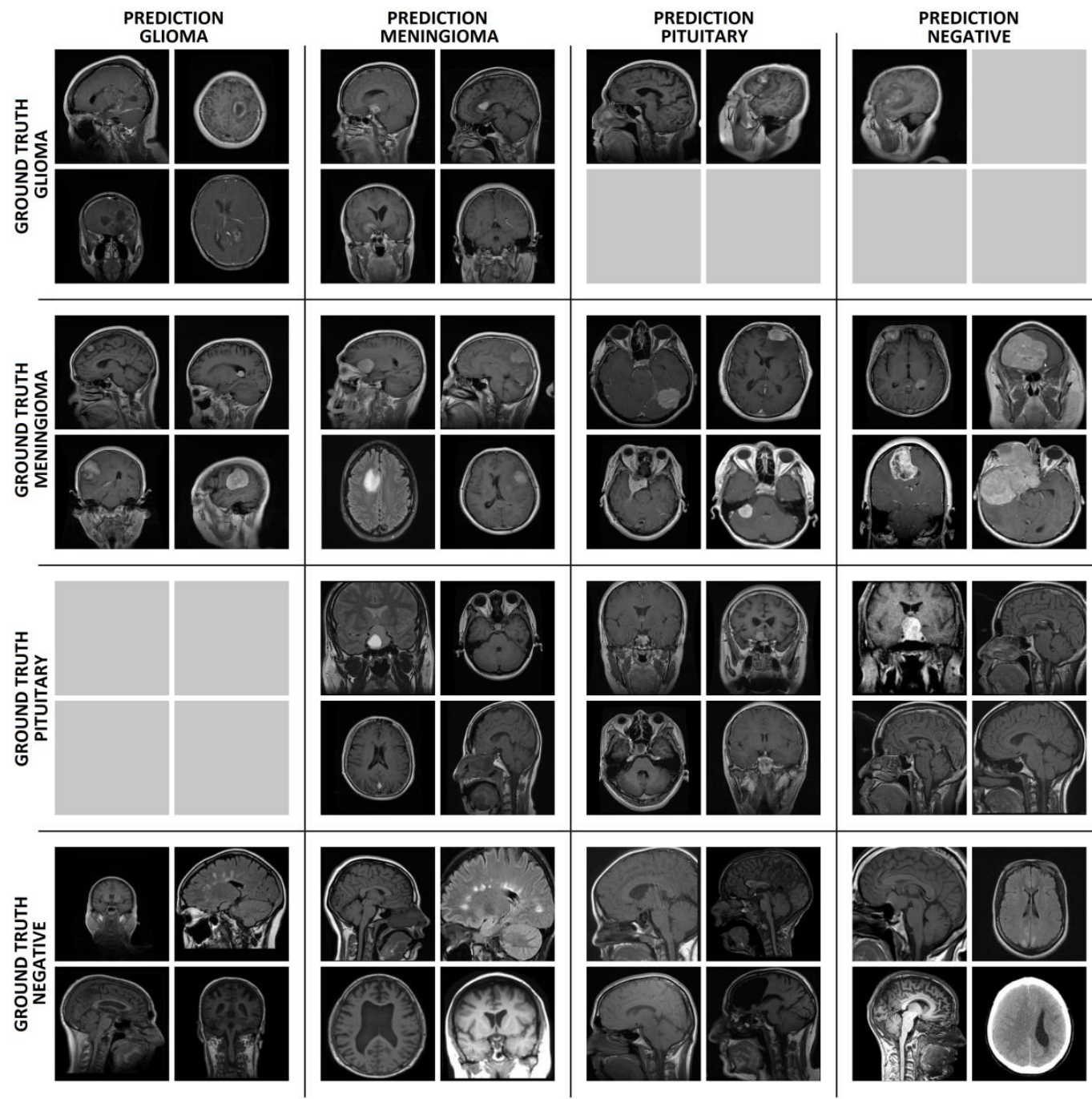
# Some network architectures





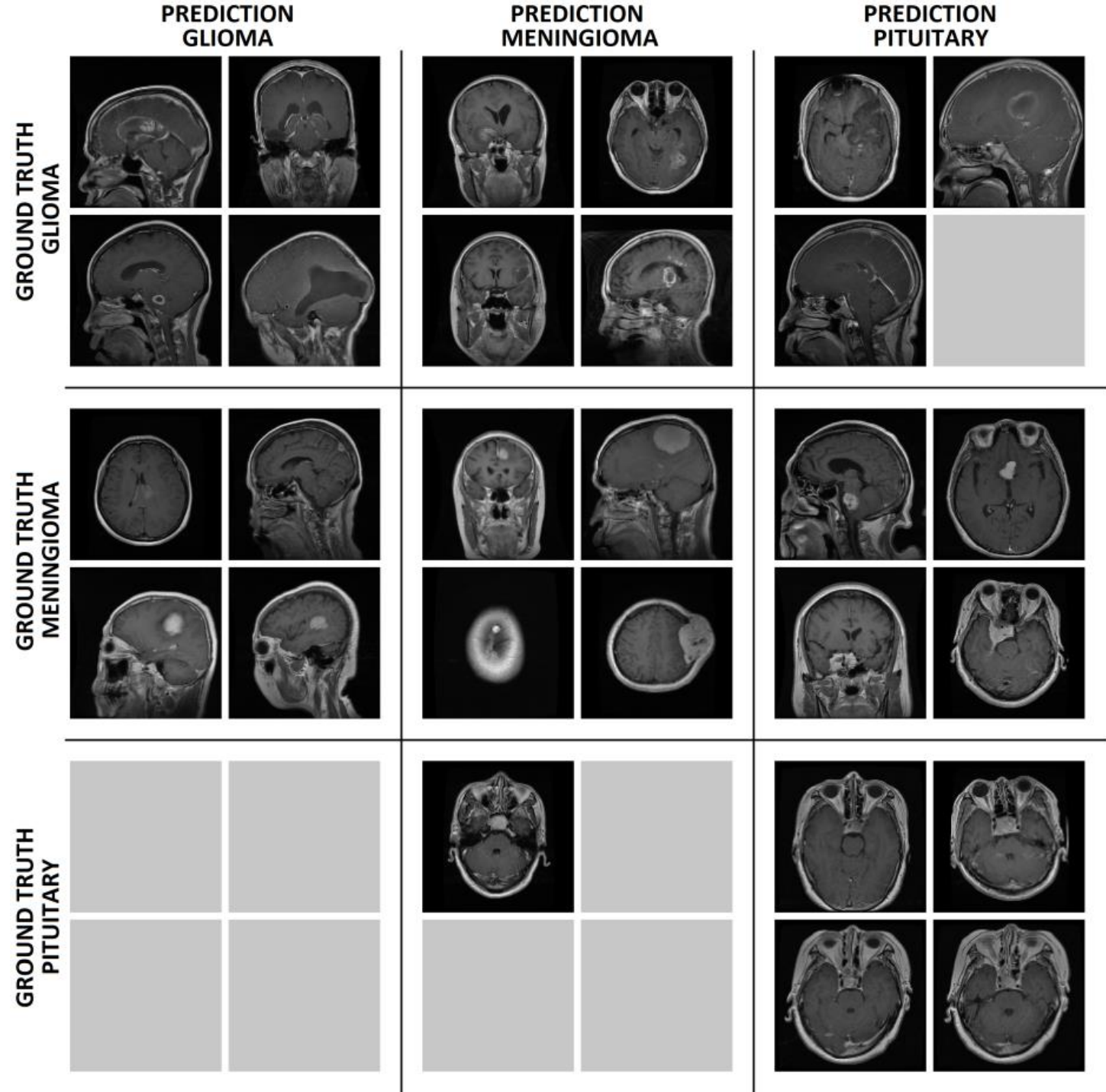
# Results – 4 classes

|            | Glioma | Meningioma | Pituitary | Negative |
|------------|--------|------------|-----------|----------|
| Glioma     | 804    | 19         | 2         | 1        |
| Meningioma | 16     | 762        | 14        | 30       |
| Pituitary  | 0      | 4          | 819       | 4        |
| Negative   | 6      | 11         | 8         | 370      |
| Recall     | 0.9734 | 0.9270     | 0.9903    | 0.9367   |
| Precision  | 0.9734 | 0.9573     | 0.9715    | 0.9136   |
| Dice score | 0.9734 | 0.9419     | 0.9808    | 0.9250   |
| Accuracy   | 0.9599 |            |           |          |



# Results – 3 classes

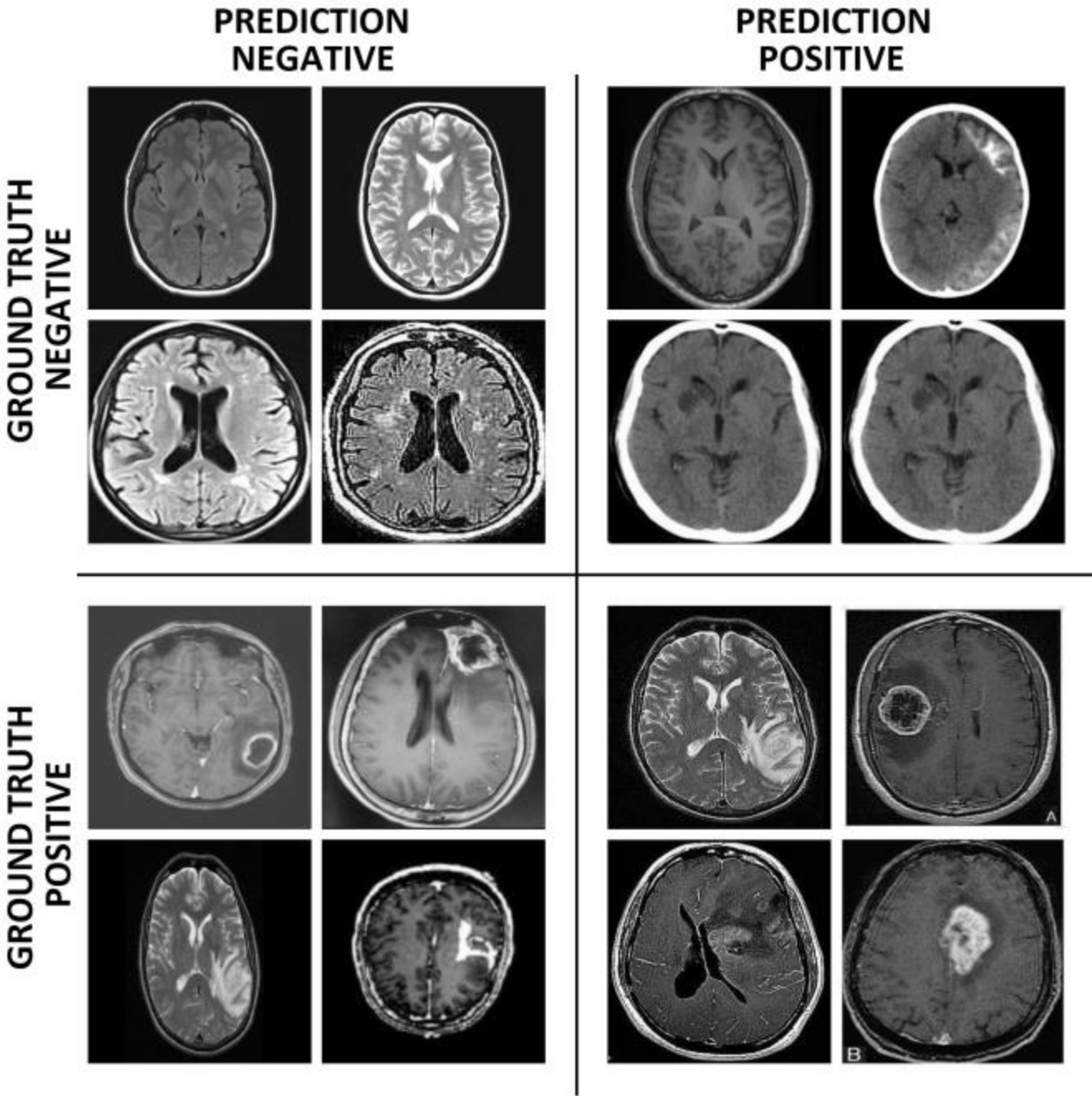
|            | Glioma | Meningioma | Pituitary |
|------------|--------|------------|-----------|
| Glioma     | 1402   | 19         | 3         |
| Meningioma | 34     | 672        | 4         |
| Pituitary  | 0      | 1          | 929       |
| Recall     | 0.9846 | 0.9465     | 0.9989    |
| Precision  | 0.9763 | 0.9711     | 0.9925    |
| Dice score | 0.9804 | 0.9586     | 0.9957    |
| Accuracy   | 0.9801 |            |           |



# Results – 2 classes

## ACCURACY BENCHMARKS OF THE 2-CLASS CLASSIFICATION PROBLEM

|            | Negative | Positive |
|------------|----------|----------|
| Negative   | 1486     | 14       |
| Positive   | 20       | 1516     |
| Recall     | 0.9907   | 0.9870   |
| Precision  | 0.9867   | 0.9908   |
| Dice score | 0.9887   | 0.9889   |
| Accuracy   | 0.9888   |          |





# Other study: best performing models

| Rank no. | CNN network | Dense neurons | Image size       | Parameters  | F1 score mean $\pm$ stdev | Accuracy mean $\pm$ stdev | AUC Meningioma mean $\pm$ stdev | AUC Glioma mean $\pm$ stdev | AUC Pituitary mean $\pm$ stdev |
|----------|-------------|---------------|------------------|-------------|---------------------------|---------------------------|---------------------------------|-----------------------------|--------------------------------|
| R1       | *VGG        | 256           | 256 $\times$ 256 | [MA][D]     | 0.9827 $\pm$ 0.0066       | 0.9827 $\pm$ 0.0066       | 0.9961 $\pm$ 0.0008             | 0.9981 $\pm$ 0.0013         | 0.9987 $\pm$ 0.0013            |
| R2       | *VGG        | 2048          | 256 $\times$ 256 | [MA][D][K9] | 0.9814 $\pm$ 0.0042       | 0.9814 $\pm$ 0.0042       | 0.9949 $\pm$ 0.0035             | 0.9973 $\pm$ 0.0023         | 0.9988 $\pm$ 0.0021            |
| R3       | *VGG        | 2048          | 256 $\times$ 256 | [MA][K9]    | 0.9814 $\pm$ 0.0029       | 0.9814 $\pm$ 0.0030       | 0.9957 $\pm$ 0.0020             | 0.9975 $\pm$ 0.0019         | 0.9995 $\pm$ 0.0008            |
| R4       | *VGG        | 1024          | 256 $\times$ 256 | [MA][D]     | 0.9808 $\pm$ 0.0037       | 0.9807 $\pm$ 0.0037       | 0.9967 $\pm$ 0.0010             | 0.9980 $\pm$ 0.0015         | 0.9988 $\pm$ 0.0011            |
| R5       | *VGG        | 1024          | 128 $\times$ 128 | [MA][D]     | 0.9795 $\pm$ 0.0018       | 0.9794 $\pm$ 0.0019       | 0.9940 $\pm$ 0.0044             | 0.9967 $\pm$ 0.0022         | 0.9994 $\pm$ 0.0008            |
| R6       | *VGG        | 256           | 256 $\times$ 256 | [MA]        | 0.9789 $\pm$ 0.0054       | 0.9788 $\pm$ 0.0056       | 0.9935 $\pm$ 0.0016             | 0.9968 $\pm$ 0.0016         | 0.9992 $\pm$ 0.0010            |
| R7       | *VGG        | 4096          | 256 $\times$ 256 | [MA][D]     | 0.9782 $\pm$ 0.0053       | 0.9781 $\pm$ 0.0054       | 0.9942 $\pm$ 0.0024             | 0.9972 $\pm$ 0.0020         | 0.9981 $\pm$ 0.0020            |
| R8       | *VGG        | 2048          | 128 $\times$ 128 | [MA][D][K9] | 0.9781 $\pm$ 0.0037       | 0.9781 $\pm$ 0.0038       | 0.9941 $\pm$ 0.0043             | 0.9963 $\pm$ 0.0017         | 0.9975 $\pm$ 0.0023            |
| R9       | *VGG        | 32            | 128 $\times$ 128 | [MA]        | 0.9778 $\pm$ 0.0047       | 0.9778 $\pm$ 0.0047       | 0.9943 $\pm$ 0.0018             | 0.9973 $\pm$ 0.0015         | 0.9991 $\pm$ 0.0009            |
| R10      | *VGG        | 4096          | 128 $\times$ 128 | [MA][D]     | 0.9776 $\pm$ 0.0046       | 0.9775 $\pm$ 0.0047       | 0.9930 $\pm$ 0.0025             | 0.9958 $\pm$ 0.0015         | 0.9985 $\pm$ 0.0012            |

|        |            | Predicted |      |     | Predicted |      |     | Predicted |      |     | Predicted |      |     | Predicted |      |     |
|--------|------------|-----------|------|-----|-----------|------|-----|-----------|------|-----|-----------|------|-----|-----------|------|-----|
|        |            | Class     |      |     | Class     |      |     | Class     |      |     | Class     |      |     | Class     |      |     |
| Actual | Meningioma | 691       | 10   | 7   | 681       | 17   | 10  | 681       | 19   | 8   | 685       | 13   | 10  | 687       | 14   | 7   |
|        | Glioma     | 24        | 1399 | 3   | 25        | 1401 | 0   | 25        | 1401 | 0   | 29        | 1396 | 1   | 36        | 1390 | 0   |
|        | Pituitary  | 5         | 4    | 921 | 3         | 2    | 925 | 3         | 2    | 925 | 3         | 3    | 924 | 4         | 2    | 924 |
|        |            | Rank R1   |      |     | Rank R2   |      |     | Rank R3   |      |     | Rank R4   |      |     | Rank R5   |      |     |
| Actual | Meningioma | 685       | 13   | 10  | 685       | 12   | 11  | 678       | 22   | 8   | 673       | 24   | 11  | 673       | 24   | 11  |
|        | Glioma     | 29        | 1396 | 1   | 34        | 1391 | 1   | 22        | 1401 | 3   | 28        | 1398 | 0   | 28        | 1398 | 0   |
|        | Pituitary  | 3         | 3    | 924 | 7         | 2    | 921 | 6         | 6    | 918 | 3         | 2    | 925 | 3         | 2    | 925 |
|        |            | Rank R6   |      |     | Rank R7   |      |     | Rank R8   |      |     | Rank R9   |      |     | Rank R10  |      |     |

# Comparison

| Paper                | Classif.<br>method | Test<br>images | Accuracy (test samples) |         |         |
|----------------------|--------------------|----------------|-------------------------|---------|---------|
|                      |                    |                | 2-class                 | 3-class | 4-class |
| Toğaçar et al [20]   | SVM                | 93             | 0.9677                  | –       | –       |
| Vankdothu et al [21] | RCNN               | 394            | 0.9517                  | –       | –       |
| Rajeev et al [22]    | LSTM               | 978            | 0.9949                  | –       | 0.9834  |
| Cheng et al [23]     | SPM                | 3064           | –                       | 0.9128  | –       |
| Isunuri et al [24]   | EfficientNet       | 1739           | –                       | 0.9835  | –       |
| Rahman et al [25]    | Parallel DCNN      | 200+           | –                       | 0.9610  | 0.9560  |
| Dénes-F. et al [19]  | VGG                | 3064           | –                       | 0.9822  | –       |
| Proposed             | CNN                | 2800+          | 0.9888                  | 0.9801  | 0.9599  |

# Conclusions

- Tumors of  $10\text{cm}^3$  size can be easily detected
- Most tumors can be segmented with 85-90% Dice score, that is approx. 98.5% accuracy of pixelwise decisions
- CNN + DL are the current state-of-the-art
  - perform better
  - work longer, decisions hardly explainable

# Some papers

- A. Györfi, L. Szilágyi, L. Kovács: A Fully Automatic Procedure for Brain Tumor Segmentation from Multi-Spectral MRI Records Using Ensemble Learning and Atlas-Based Data Enhancement. *Applied Sciences* 11(2):564, 2021.
- A. Györfi, L. Kovács, L. Szilágyi: A two-stage U-net approach to brain tumor segmentation from multi-spectral MRI records. *Acta Universitatis Sapientiae – Informatica*, 14(2):223–247, 2022.
- L. Dénes-Fazakas, G. Eigner, L. Kovács, L. Szilágyi: Two U-net Architectures for Infant Brain Tissue Segmentation from Multi-Spectral MRI Data. *IFAC World Congress 2023, IFAC PapersOnLine* 56(2):5637-5642, 2023.
- S. Csaholczi, L. Kovács, L. Szilágyi: Brain Tumor Classification Using Convolutional Neural Networks and Deep Learning. *ICCC 2024*, pp. 399-404.
- L. Szilágyi, Á. Györfi, L. Dénes-Fazakas, S. Csaholczi, I.M. Pisak-Lukáts, L. Kovács: Challenges and Difficulties of Multi-Spectral MRI Based Brain Tumor Detection and Segmentation, *ICHST 2023*, pp. 1-6.
- A. Köble, A. Györfi, S. Csaholczi, B. Surányi, L. Dénes-Fazakas, L. Kovács, L. Szilágyi: Identifying the most suitable histogram normalization technique for machine learning based segmentation of multispectral brain MRI data. *IEEE AFRICON 2021*, pp. 71–76.
- A. Györfi, S. Csaholczi, I.M. Pisak-Lukáts, L. Dénes-Fazakas, A. Köble, O. Shvets, Gy. Eigner, L. Kovács,, L. Szilágyi: Effect of spectral resolution on the segmentation quality of magnetic resonance imaging data. *INES 2022*, pp. 53–58.



# Team

- **PhD students**
  - Lehel Dénes-Fazakas
  - Ágnes Győrfi
  - Szabolcs Csaholczi
  - Ioan M. Pisak-Lukáts
- **Collaborators**
  - László Lefkovits
  - Szidónia Lefkovits
- **Special thanks**
  - Levente Kovács
  - Vladik Kreinovich