Brain tumor detection (and segmentation) in multispectral MRI data

László Szilágyi

Sapientia University of Transylvania

Sapientia University

- Founded in 2000, operates from 2001
- Four cities in Transylvania
- Faculty of Technical and Human Sciences of Tîrgu Mureş
- Department of Electrical Engineering
- Computational Intelligence Research Group (CIRG)

Medical problem

- Hundreds of thousands die of brain tumor yearly
- Early detection in case of any cancer increases the chances of survival

- Detection vs Segmentation

Detection (in early stage)

- Main goal: screening patients
- General philosophy:
  - Classify image voxels into various tissue types
  - Identify tissue types found
  - Decision making: is there any continuous area suspected to be tumor?
  - If so, point out the suspicious areas
- Procedure
  - Reliable detection
  - Fully automatic algorithm
  - Handle huge image data volumes
  - Runtime efficiency also matters

Segmentation

- Main goal: therapy planning; follow-up
- Difference from Detection
  - We already know the tumor is present
  - Differentiating between normal and tumor tissues
  - Identify various tissue types associated with tumors
  - Find the boundary of the tumor, quantify its size
- Procedure
  - Fine-tuned for tumor boundary detection
  - Fully automatic algorithm
  - Handle huge image data volumes
  - Runtime efficiency also matters

What can we do better?

- Detection
  - Early detection, better chances of survival
  - Implement it into screening procedure
- Segmentation
  - Radiation therapy can be designed to destroy tumor but not the surrounding normal tissues
  - Exact segmentation and quantification can assist exact follow-up monitoring

MRI

- Magnetic Resonance Imaging (MRI)
  - Paul Lauterbur, Peter Mansfield,
  - Imaging technique 1970’s
  - Nobel Prize in Physiology/Medicine 2003

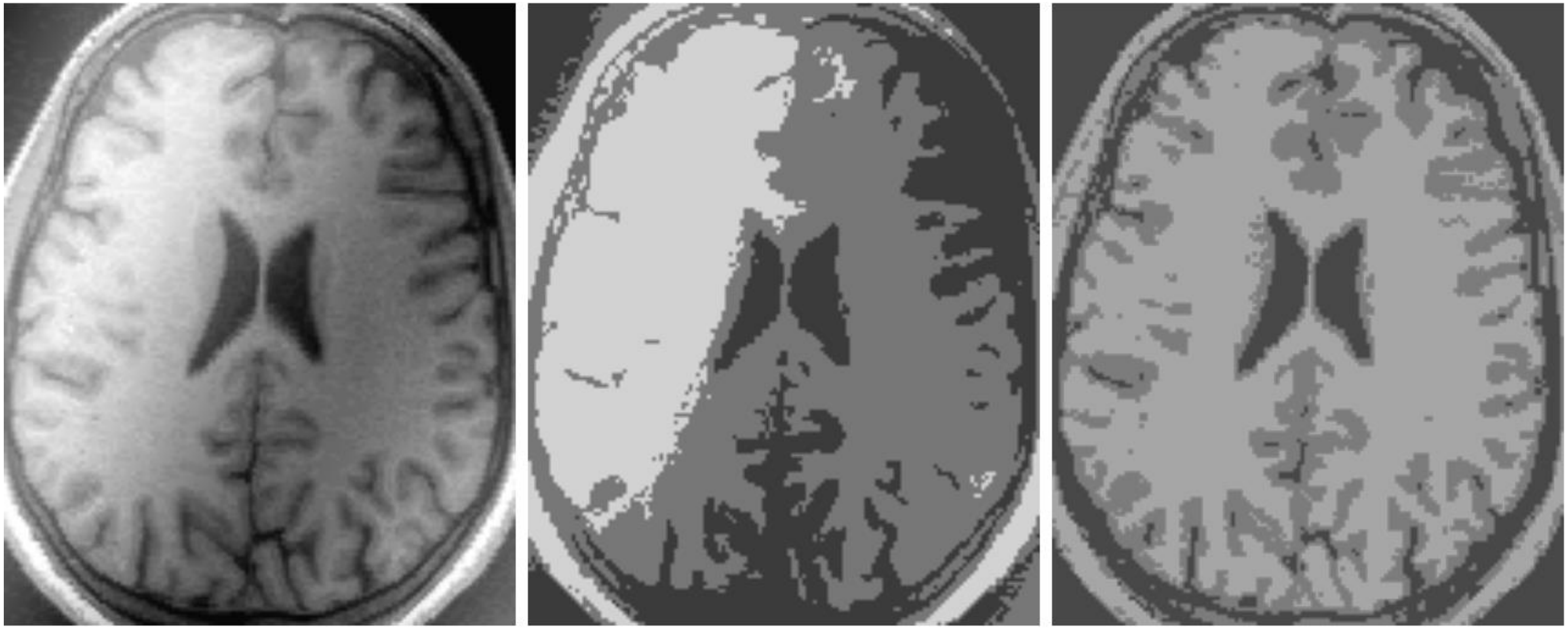
- Less harmful than X-Ray
- Better contrast
- Better resolution in intensity levels
- Multi-spectral: various weighting schemes

Difficulties

- Non-brain tissues (e.g. skull, eyes, etc.)
- Intensity non-uniformity or intensity inhomogeneity
- **Absolute intensity values**
- Registration of multiple data channels

Intensity non-uniformity (INU)

- Noise of low frequency and possibly high amplitude
- Compensation needed before or during segmentation
- Reviews:
  - Vovk et al, IEEE T Medical Imaging 26(3):405-421, 2007
  - Sled, in Brain Mapping. An encyclopedic reference, Acad. Press, 2015



MICCAI BRATS Challenge

- Medical Image Computation and Computer Assisted Interventions
- Brain Tumor Segmentation Challenge
  - http://braintumorsegmentation.org/
- Menze, Jakab, et al: IEEE T Med Imag 34:1993-2024, 2015
- 4 data channels
  - T1, T2, T1C (contrast enhanced T1)
  - FLAIR: Fluid-attenuated inversion recovery
- All channels registered to T1
- Resampled to cubic voxels of 1mm^3 size
- Skull and other non-brain tissues removed
- Intensity non-uniformity not present
- One volume: 176 slices of (160-216 x 216-236 pixels)
- 1.5 liters of brain = 1.5 million voxels

BRATS since 2012

- BRATS 2012:
  - 30 train volumes (20 high-grade, 10 low-grade)
  - Ground truth: **normal** tissue or **edema** or **tumor**
- Report on BRATS 2012/13: Menze et al, 2015
  - 2x10 methods provided by finalists, results
- BRATS 2015:
  - 200+ volumes
  - GT: differentiated tumor tissue types: tumor core, active tumor, necrosis

Impact of BRATS

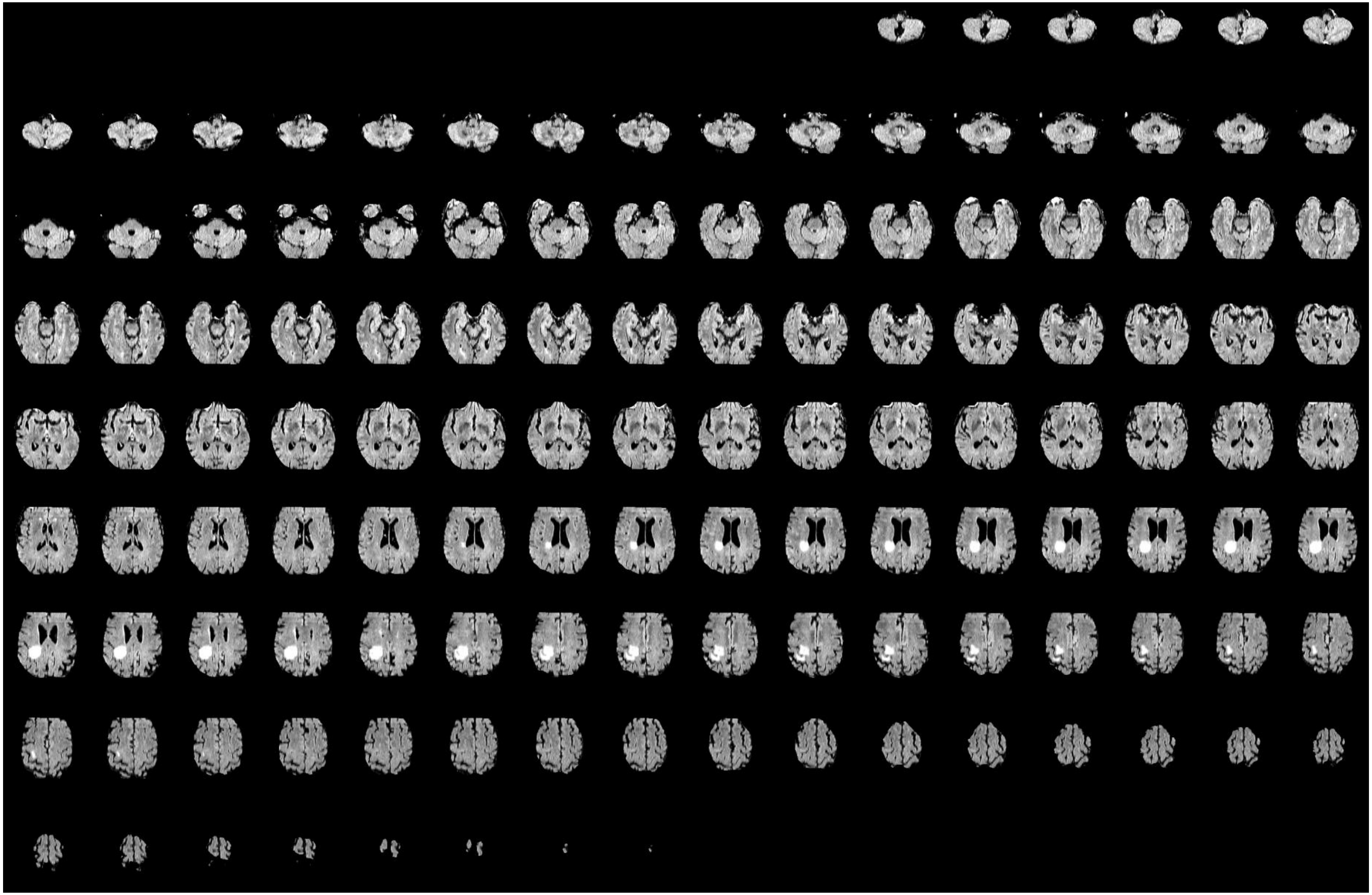
- Random Forest
  - Tustison et al: Neuroinformatics 13(2):209-225, 2015
- Greedy algorithm
  - Kanas et al: Biomed. Signal Proc. Contr. 22:19-30, 2015
- Deep learning NN
  - Havaei et al: Med. Imag. Analysis, in press
- CNN
  - Pereira et al: IEEE T Med. Imag. 35(5):1240-1251, 2016
- AdaBoost
  - Islam et al: IEEE T Biomed. Eng. 60(11):3204-3215, 2013
- Bayes network + EM
  - Menze et al: IEEE Trans. Med. Imag. 35(4):933-946, 2016
- Graph-based segmentation
  - Njeh et al: Comput. Med. Imag. Graph. 40:108-119, 2015
- Tumor growth model
  - Lê et al: IEEE T Med. Imag., in press
- Not yet: **SVM**

Input Data

13 selected volumes, each contains 176 slices, 4 data channels + ground truth

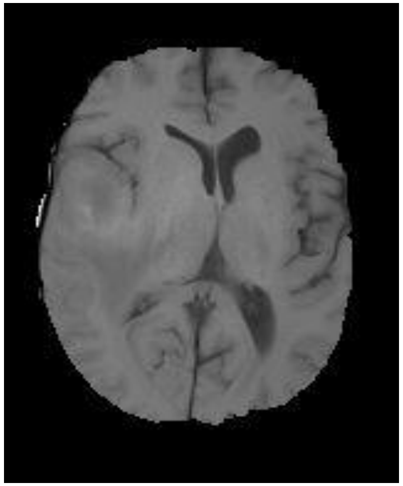
Volume	Size	Pixel count	Edema pixels	Tumor pixels	Missing data
HG01	160x216	1,435,938	52,073	56,743	428
HG02	160x216	1,533,860	39,672	9,029	598
HG03	176x216	1,702,191	133,029	20,558	66,402
HG04	160x216	1,198,268	46,054	60,635	62
HG05	160x216	1,469,666	35,545	24,659	4,215
HG06	216x236	1,577,073	103,781	75,337	538,771
HG07	176x216	1,569,615	60,181	28,156	150
HG09	160x216	1,255,130	135,135	83,564	44,426
HG11	176x216	1,333,310	98,356	45,806	81
HG12	176x216	1,508,141	13,928	6,577	106,754
HG13	176x216	1,451,294	6,350	4,537	65,428
HG14	176x216	1,590,059	26,081	115,286	198
HG15	160x216	1,617,051	108,706	72,310	214

Single channel of a volume

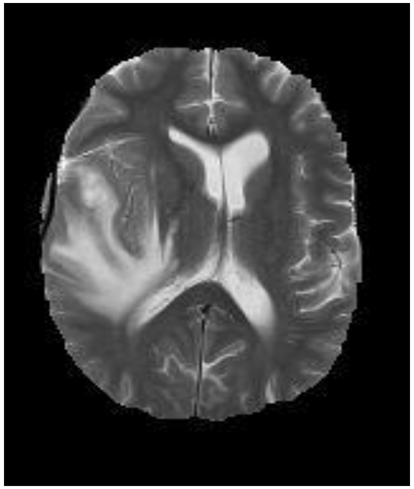


Single slice in 4 data channels + GT

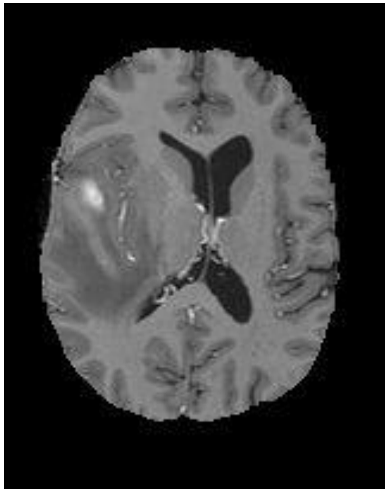
T1



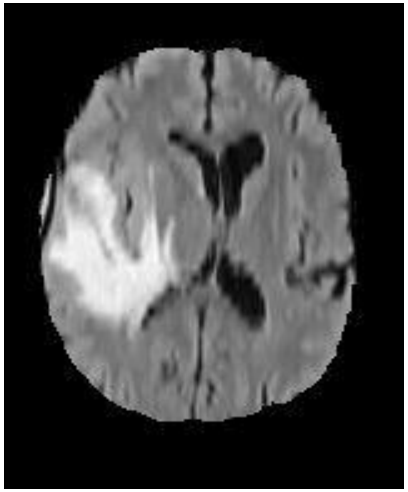
T2



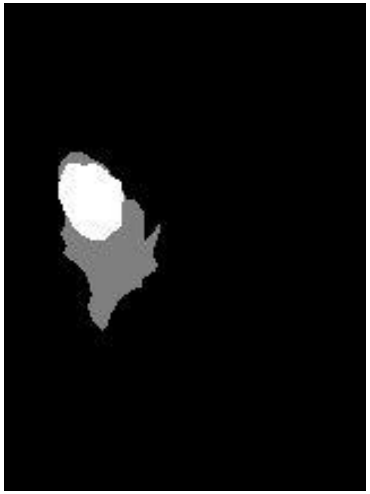
T1C



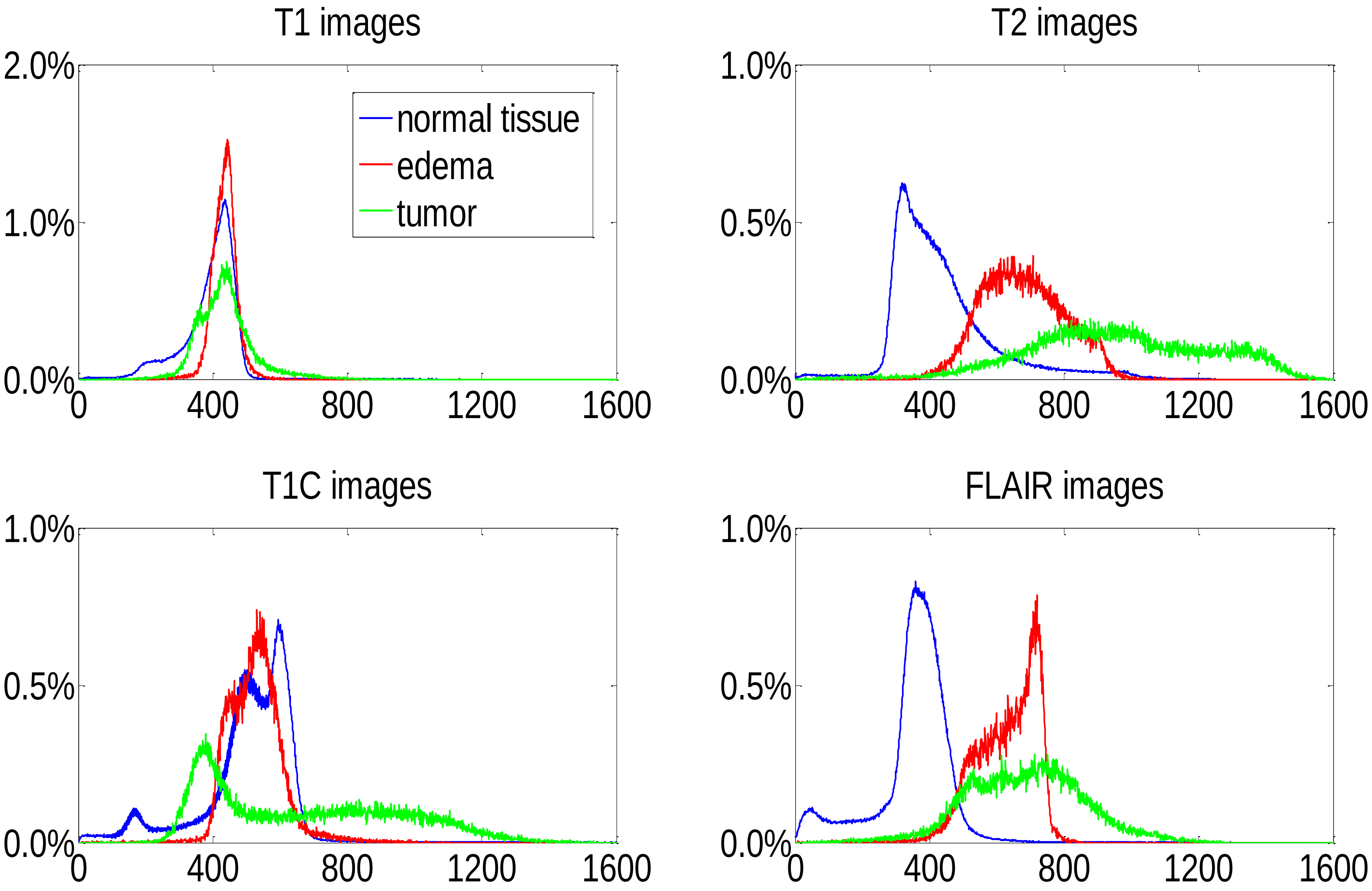
FLAIR



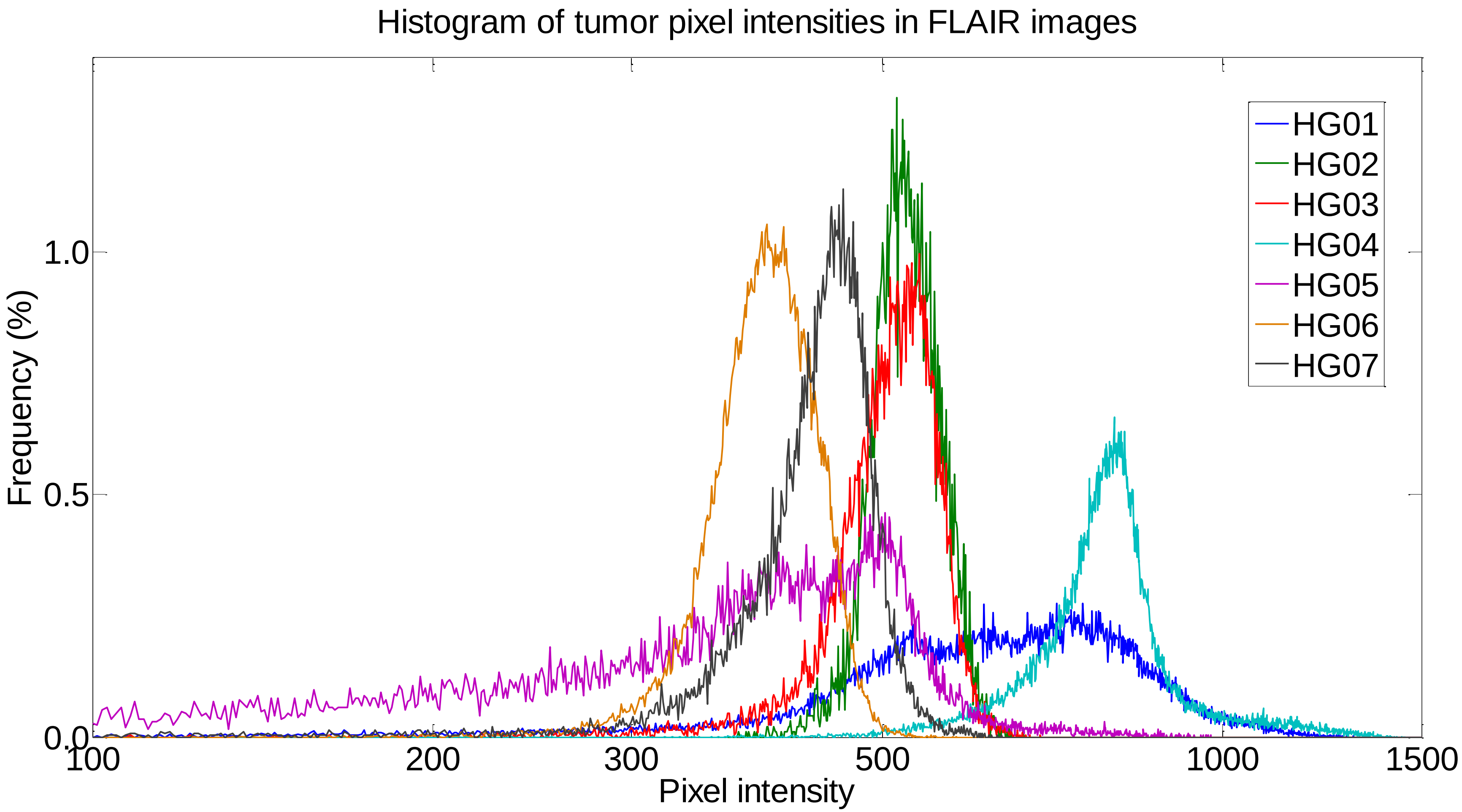
Ground truth



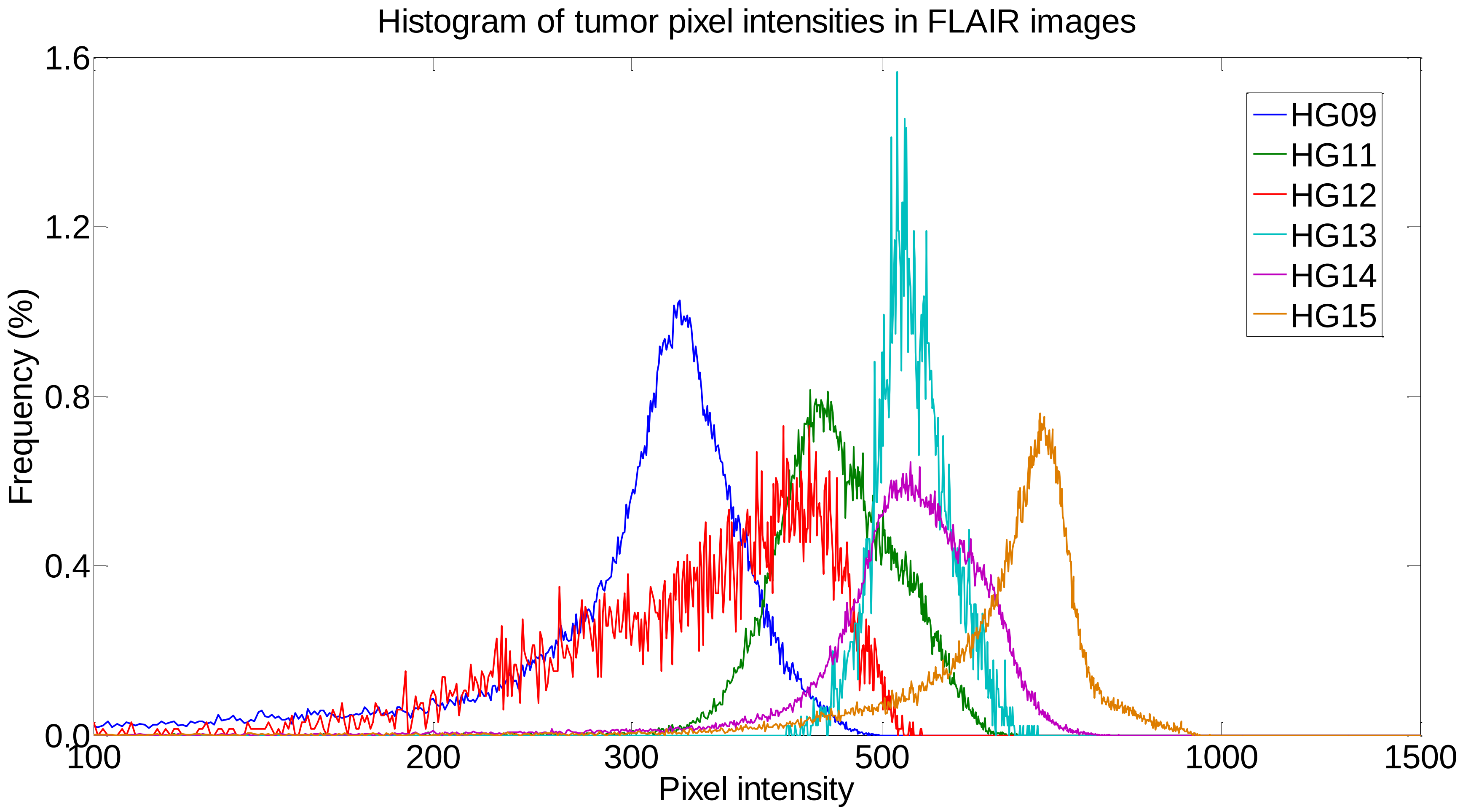
Intensity histogram of HG01



Intensity histogram of tumor pixels



Intensity histogram of tumor pixels



Handling variations in scanner sensitivity

- Histogram matching
  - Wang et al: Magn. Res. Med. 39(2):322-327, 1998
  - Nyúl et al: IEEE T Med. Imag. 19(2):143-150, 2000
- Gaussian mixture estimation
  - Hellier: ICIP 2003, pp. 1109-1112
- Histogram warping
  - Cox et al: ICIP 1995, pp. 2366-2369
- Kullback-Leibler divergence
  - Weisenfeld et al: ISBI 2004, pp. 101-104
- Symmetry-based criteria
  - Tustison et al: Neuroinformatics 13(2):209-225, 2015

A simple two-step alternative

- Treat each data channel (T1, T2, …) separately
- Linear transformation of voxel intensities
  - X becomes aX+b
  - Definition of a and b: middle fifty percent of voxel intensities would fall in a predefined interval, e.g. [600,800]
- Set up a lower and upper intensity limit, e.g.
  - All intensities below 200 are rounded up to 200
  - All intensities below 1200 are rounded down to 200
- Extreme (possibly noisy) data will not affect segmentation
- Easy to implement, quick, and it does help

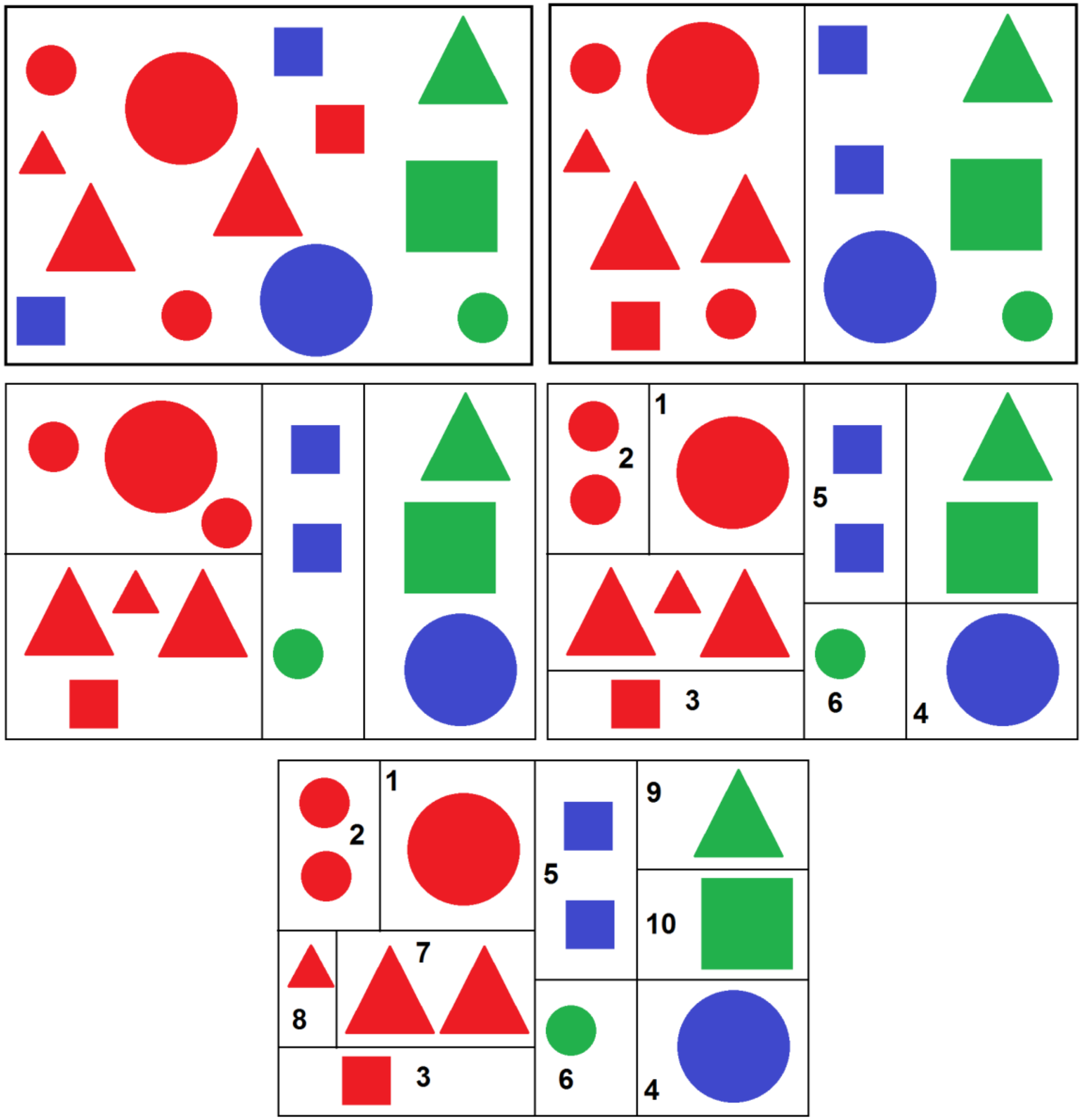
Feature extraction

- Feature vector: size does matter
- 4 intensity values is not enough
- Neighborhood information
  - Average filters, median filters, gradients
  - Morphological operations
  - Wavelet transform based texture descriptors
    - Demirhan et al: IEEE J BHI 19(4):1451-1458, 2015
    - Bendib et al: Pattern Analysis Applic. 18:829-843, 2015
  - Symmetry based features
    - Tustison et al: Neuroinformatics 13(2):209-225, 2015
  - Fractal based texture descriptors
    - Islam et al: IEEE T Biomed. Eng. 60(11):3204-3215, 2013
- The actual structure of the feature vector is usually not given by authors

Decision making

- Virtually any machine learning algorithm is suitable
- Mostly supervised learning
  - SVM, RF, CNN, DNN, ANN, etc.
- Some semi-supervised learning
  - e.g. based on SOM
  - Vishnuvarthanan et al: Appl. Soft. Comput 38:190-212, 2016
  - Demirhan et al: IEEE J Biomed. Health. Inform. 19(4):1451–1458, 2015

Binary decision tree at Kindergarten



```
+ Is it red?
Y-+ Is it circle?
| Y-+ Is it large?
| | Y-> Leaf 1
| | N-> Leaf 2
| N-+ Is it square?
|   Y-> Leaf 3
|   N-> Is it large?
|     Y-> Leaf 7
|     N-> Leaf 8
N-+ Is it large?
  Y-+ Is it green?
  | Y-+ Is it triangle?
  | | Y-> Leaf 9
  | | N-> Leaf 10
  | N-> Leaf 4
  N-+ Is it square?
    Y-> Leaf 5
    N-> Leaf 6
```

Rules of binary decision tree

- Each node compares a single feature with a threshold
- Two branches correspond to: >= or <
- Learning
  - Each decision locally minimizes an entropy function
  - We get trees of reduced depth
- Testing
  - Start from root and follow the decisions made at each node until a leaf is reached
  - Leaf assigns the label to test data

Forests

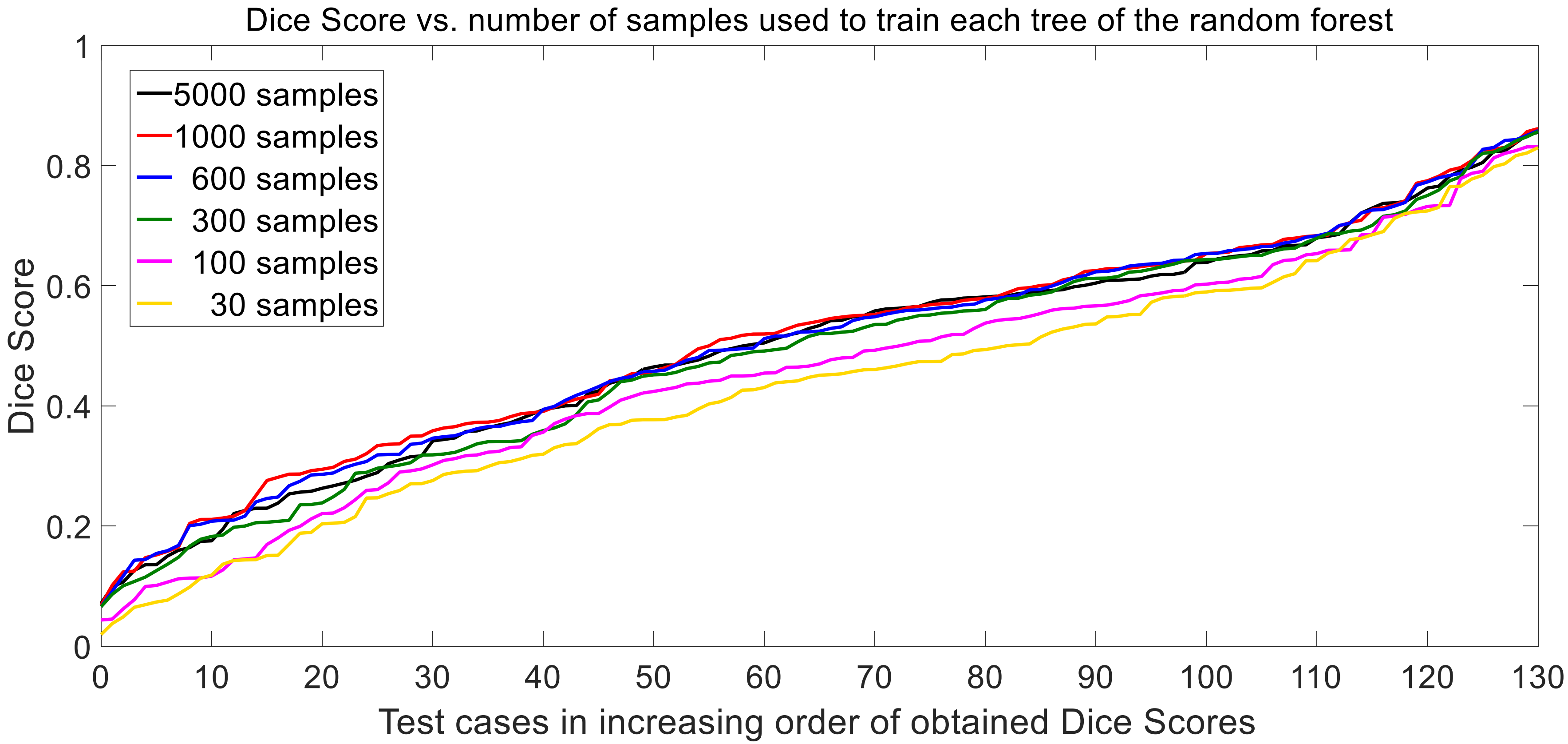
- Can a single tree learn a whole volume?
- Train several trees using randomly chosen small data set
  - Same amount of negative, edema, and tumor voxels
- Variable sample size from 3x30 to 3x5000 voxels
- How many trees in the forest?
- Testing
  - Each voxel receives a vote from each tree
  - Label assigned according to majority of votes

- Post-processing needed

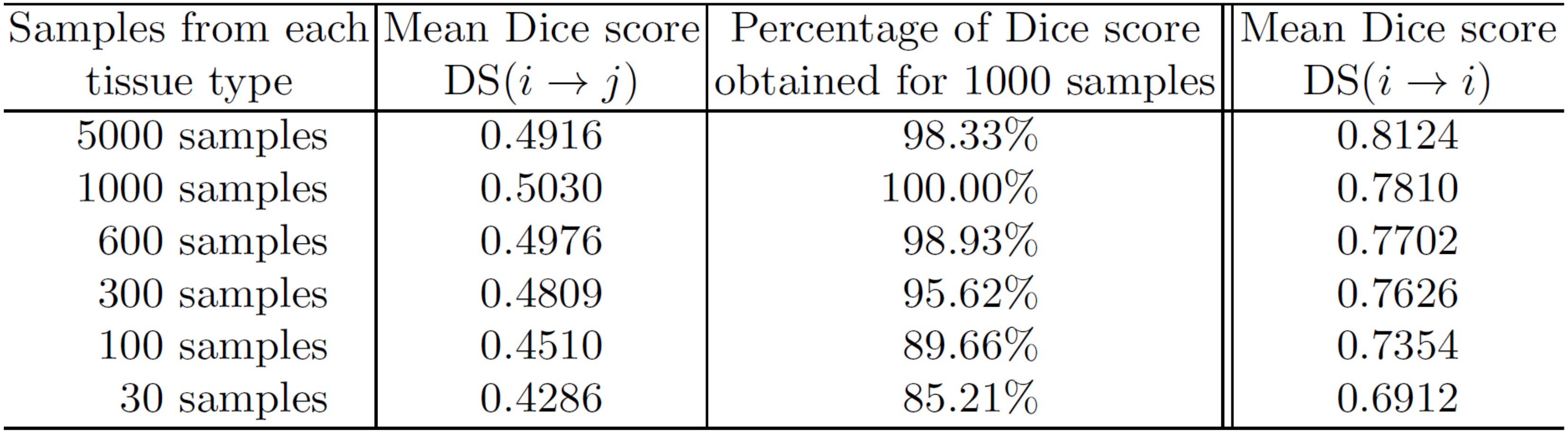
Measuring accuracy

- TP, TN, FP, FN
- Sensitivity: TP / (TP + FN)
- Specificity: TN / (TN + FP)
- Jaccard Index
  - JI = TP / (TP + FP + FN)
- Dice Score
  - DS = (2 x TP) / (2 x TP + FP + FN) = (2 x JI) / (1 + JI)

All DS(i→j) values

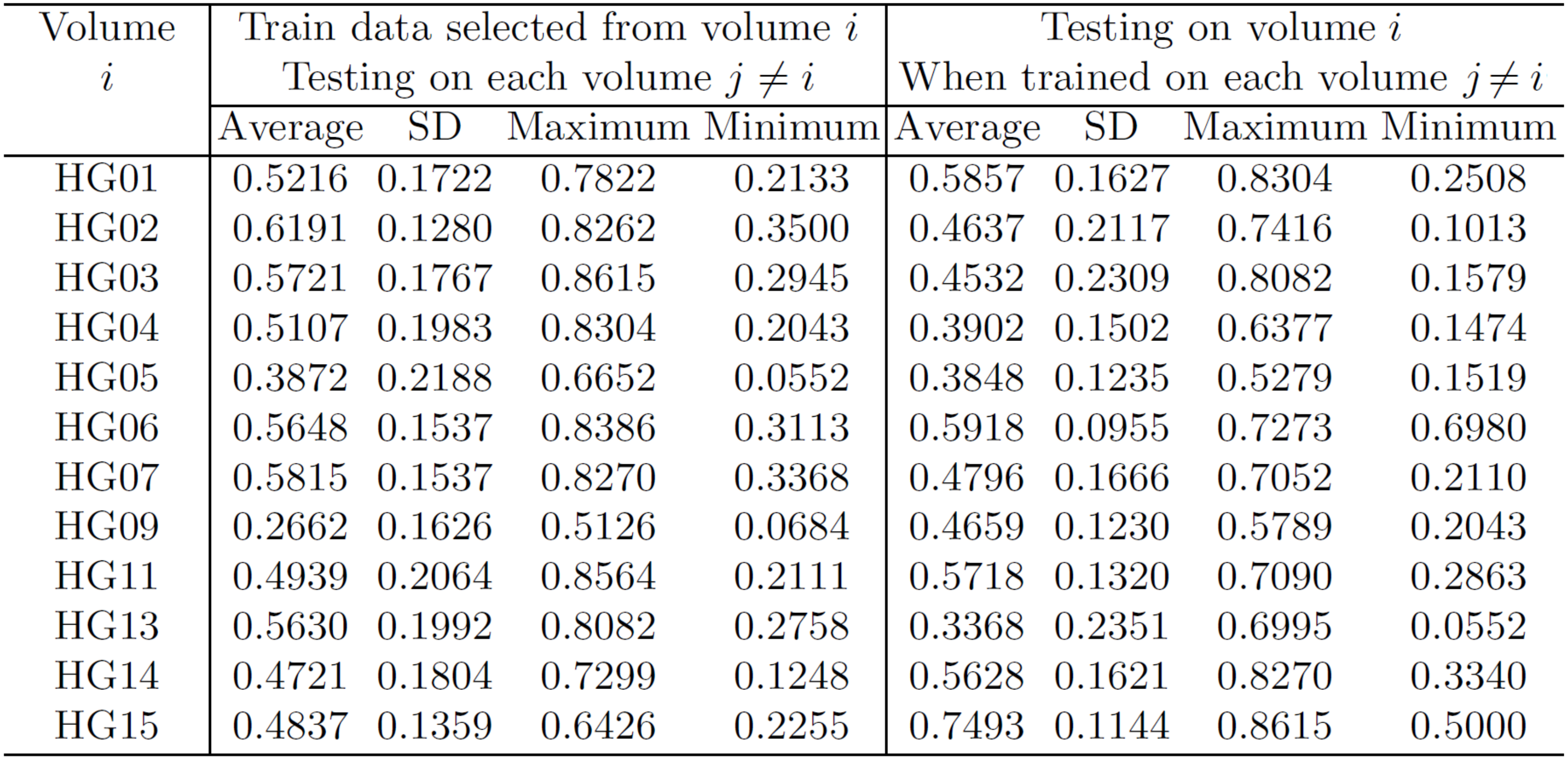


Grand mean of DS(i→j) values



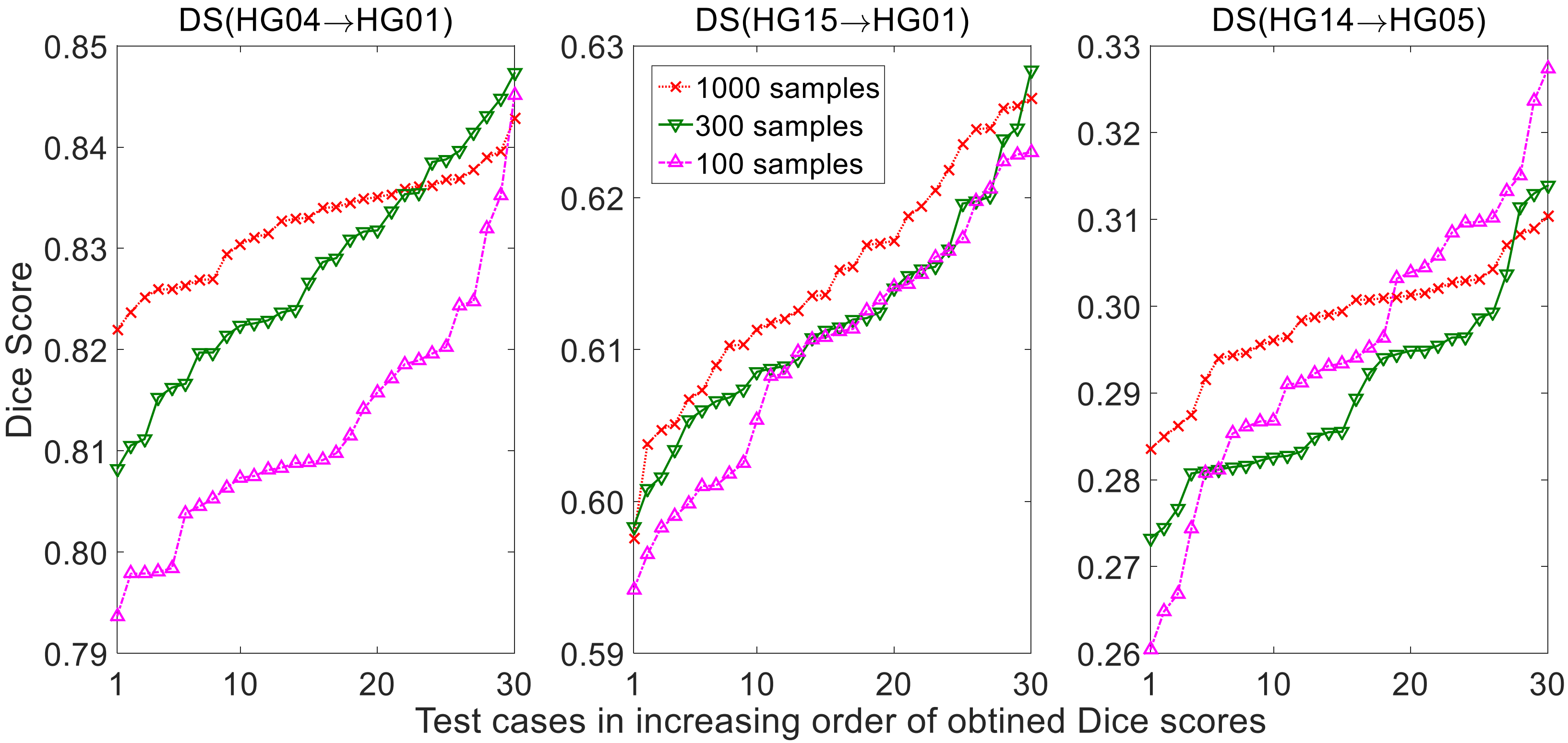
Samples from each tissue type	Mean Dice score $\mathrm{DS}(i \to j)$	Percentage of Dice score obtained for 1000 samples	Mean Dice score $\mathrm{DS}(i \to i)$
5000 samples	0.4916	98.33%	0.8124
1000 samples	0.5030	100.00%	0.7810
600 samples	0.4976	98.93%	0.7702
300 samples	0.4809	95.62%	0.7626
100 samples	0.4510	89.66%	0.7354
30 samples	0.4286	85.21%	0.6912

DS(i→j) values

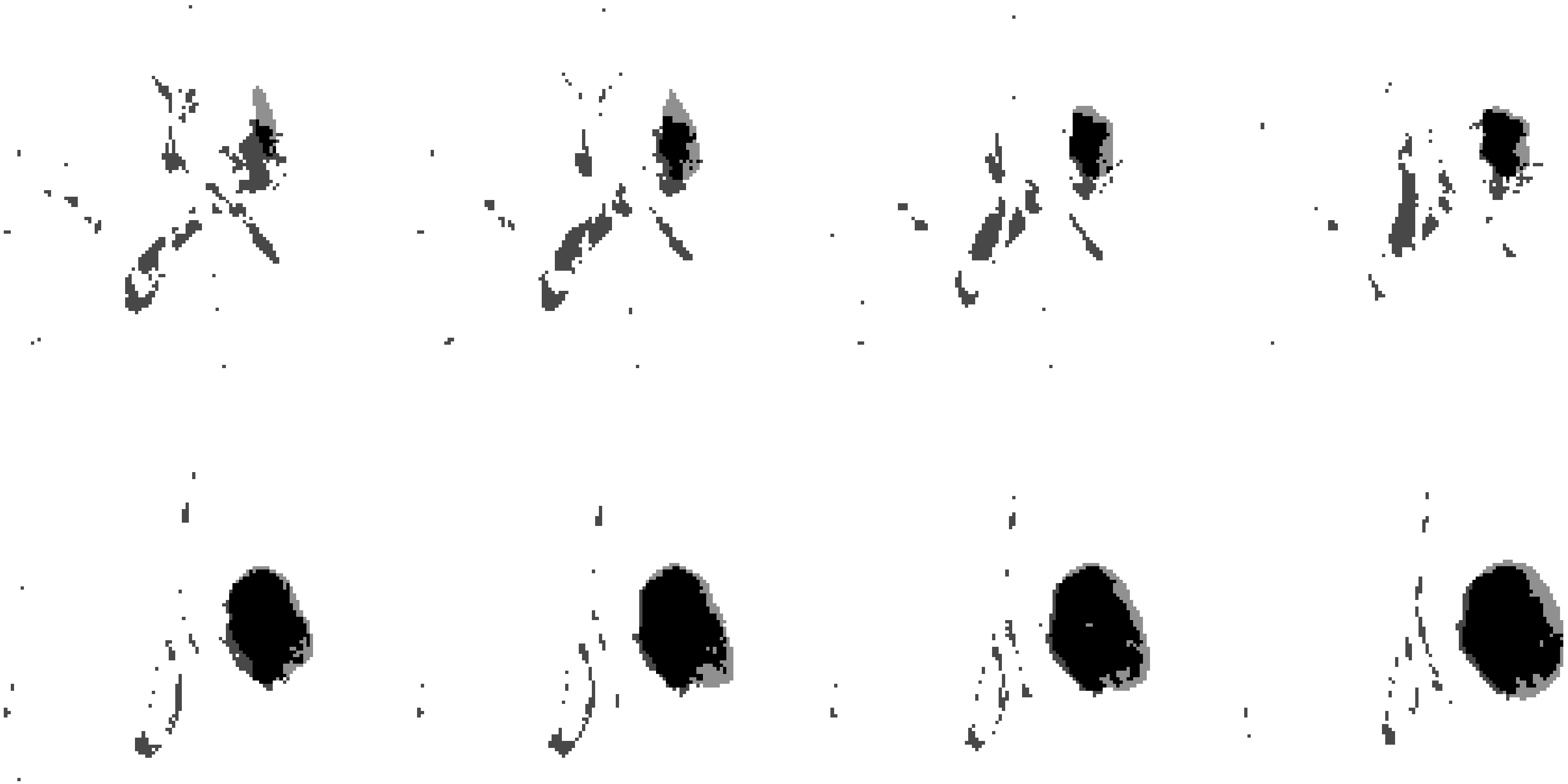


Volume i	Train data selected from volume i Testing on each volume $j \neq i$				Testing on volume i When trained on each volume $j \neq i$			
	Average	SD	Maximum	Minimum	Average	SD	Maximum	Minimum
HG01	0.5216	0.1722	0.7822	0.2133	0.5857	0.1627	0.8304	0.2508
HG02	0.6191	0.1280	0.8262	0.3500	0.4637	0.2117	0.7416	0.1013
HG03	0.5721	0.1767	0.8615	0.2945	0.4532	0.2309	0.8082	0.1579
HG04	0.5107	0.1983	0.8304	0.2043	0.3902	0.1502	0.6377	0.1474
HG05	0.3872	0.2188	0.6652	0.0552	0.3848	0.1235	0.5279	0.1519
HG06	0.5648	0.1537	0.8386	0.3113	0.5918	0.0955	0.7273	0.6980
HG07	0.5815	0.1537	0.8270	0.3368	0.4796	0.1666	0.7052	0.2110
HG09	0.2662	0.1626	0.5126	0.0684	0.4659	0.1230	0.5789	0.2043
HG11	0.4939	0.2064	0.8564	0.2111	0.5718	0.1320	0.7090	0.2863
HG13	0.5630	0.1992	0.8082	0.2758	0.3368	0.2351	0.6995	0.0552
HG14	0.4721	0.1804	0.7299	0.1248	0.5628	0.1621	0.8270	0.3340
HG15	0.4837	0.1359	0.6426	0.2255	0.7493	0.1144	0.8615	0.5000

Randomness vs Reproducibility



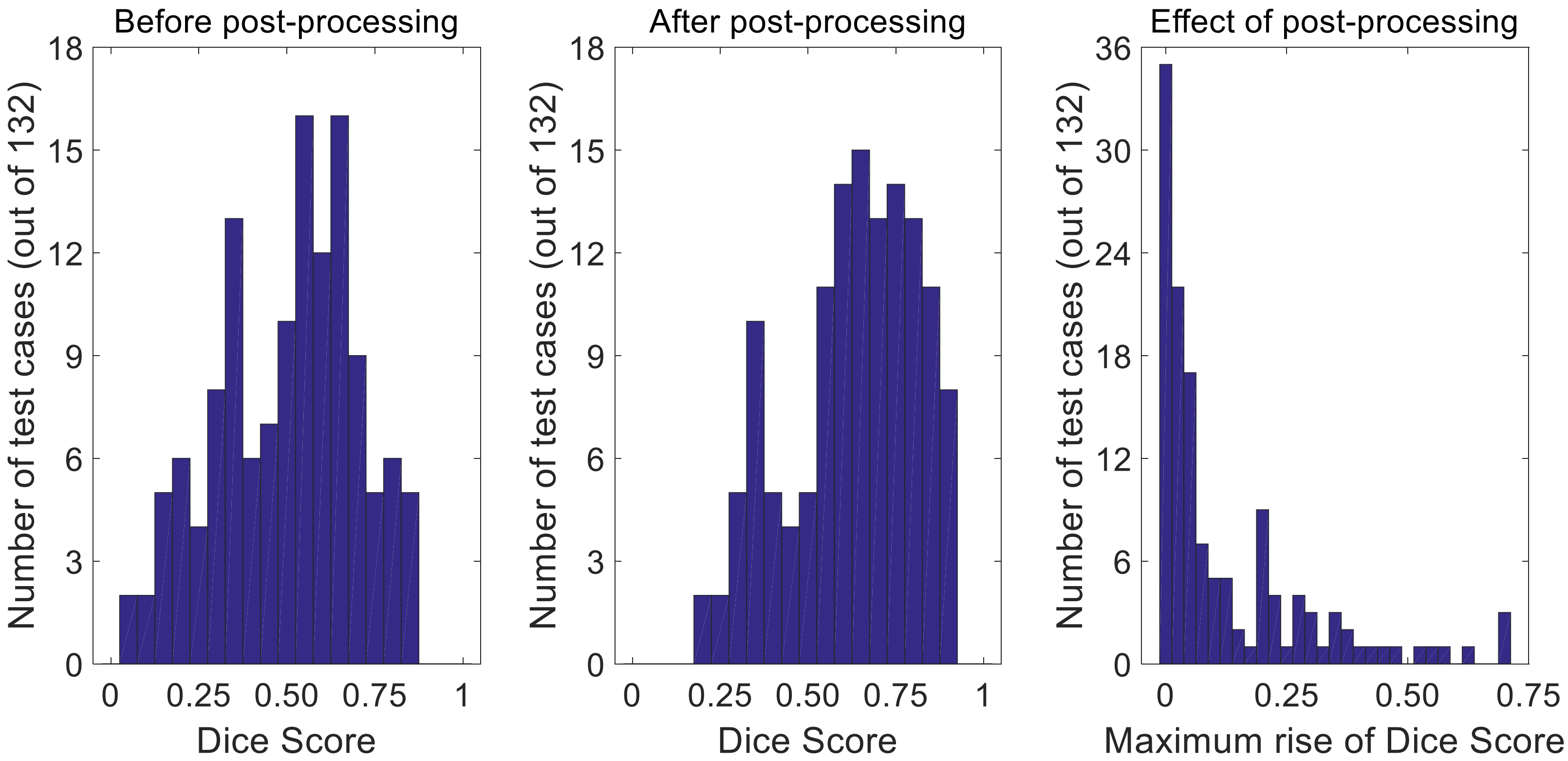
Classification result



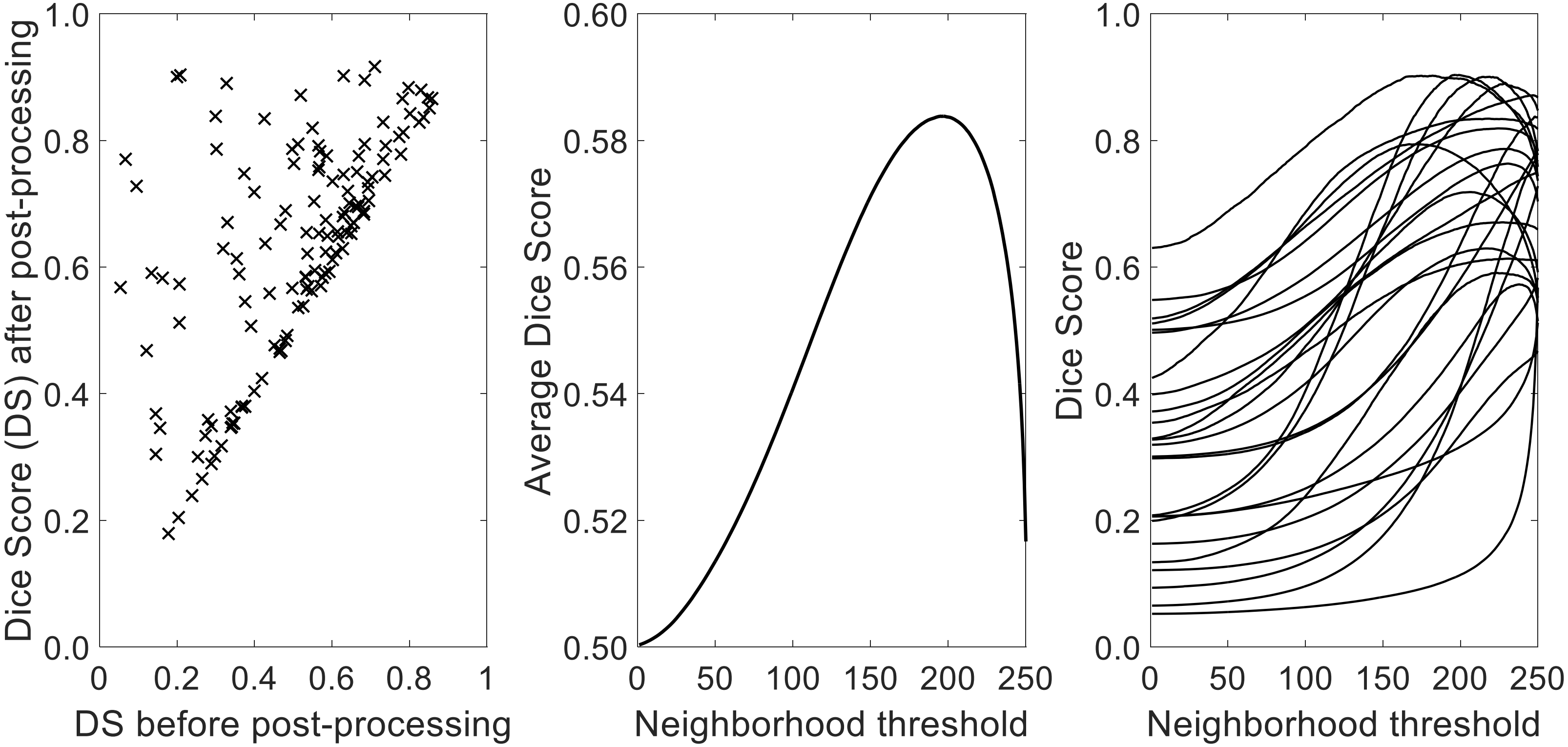
Validating tumor voxels

- Neighborhood of voxel
- Consider 7x7x7 neighborhood
- Euclidean distance < sqrt(15)
- 250 such neighbors: (5x5x5-1) + 6x(5x5-4)
- How many of them are labeled as tumor or edema?
- Threshold between 0 and 250
- Above threshold: tumor voxel validated
- Below threshold: tumor voxel discarded

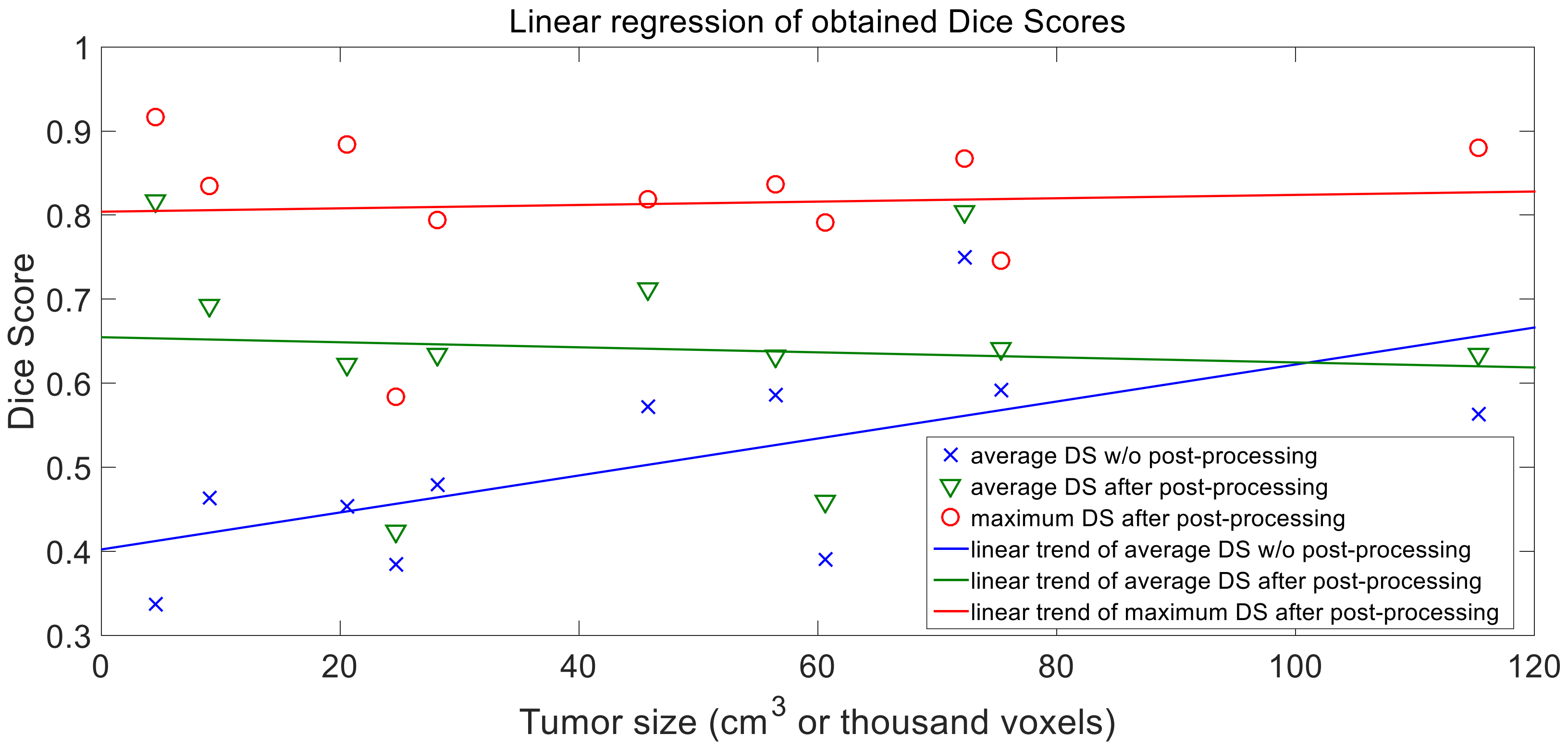
Post-processing



Post-processing



DS vs Tumor Size



Final Segmentation

Further development

- Build a single decision making system trained on several volumes
  - Choose the most independent volumes
  - Each tree should contain samples from several volumes
- Hierarchical forest?
- Reduced data

- Standalone application
  - Should also treat INU
  - Should also perform skull removal

Semi-supervised approach

- Clustering algorithm: fuzzy c-means (FCM)
- Cascade FCM
- Selection of clusters based on a decision tree
- Semi-supervised framework

Fuzzy c-means algorithm (FCM)

- Minimizes a quadratic objective function

$$J_{\mathrm{FCM}} = \sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik}^{m} ||\boldsymbol{x}_k - \boldsymbol{v}_i||^2$$

- Constrained by: $\sum_{i=1}^{c} u_{ik} = 1$
- Parameters: number of clusters c>1, fuzzy exponent m>1.0
- Problems:
  - fixed number of clusters, how many?
  - depends on initialization (seeding)
  - depends on outliers
  - sensitive to cluster sizes

FCM clustering outcome for single slice

c=10 classes, fuzzy exponent m=2.0

Cluster	**Cluster prototype intensities (log)**				**Pixels in cluster**		
	T1	T2	T1C	FLAIR	Normal	Edema	Tumor
1	6.094	5.767	6.414	5.874	5,627	1	
2	5.599	6.621	5.681	5.194	773		
3	4.099	3.955	4.779	3.669	88		
4	5.303	6.842	5.171	4.361	546		
5	6.027	5.982	6.278	5.996	3,816	9	
6	5.265	6.838	5.204	3.668	416		
7	5.976	6.713	6.155	6.532	207	1,149	256
8	5.995	6.215	6.254	6.120	2,758	236	27
9	5.814	6.379	6.095	5.774	1,522		
10	3.180	0.477	5.203	2.554	24		
				Total	**15,777**	**1,395**	**283**

Cascade FCM algorithm

- First: apply FCM to whole volume
    - fuzzy exponent m, c clusters
- Cluster selection: keep selected clusters, label as negative all others
- Second: apply FCM to pixels of kept clusters
    - fuzzy exponent m’, c’ clusters
- Cluster selection: positive and negative ones

- Selection: decision tree built upon clusters found in 3 further volumes

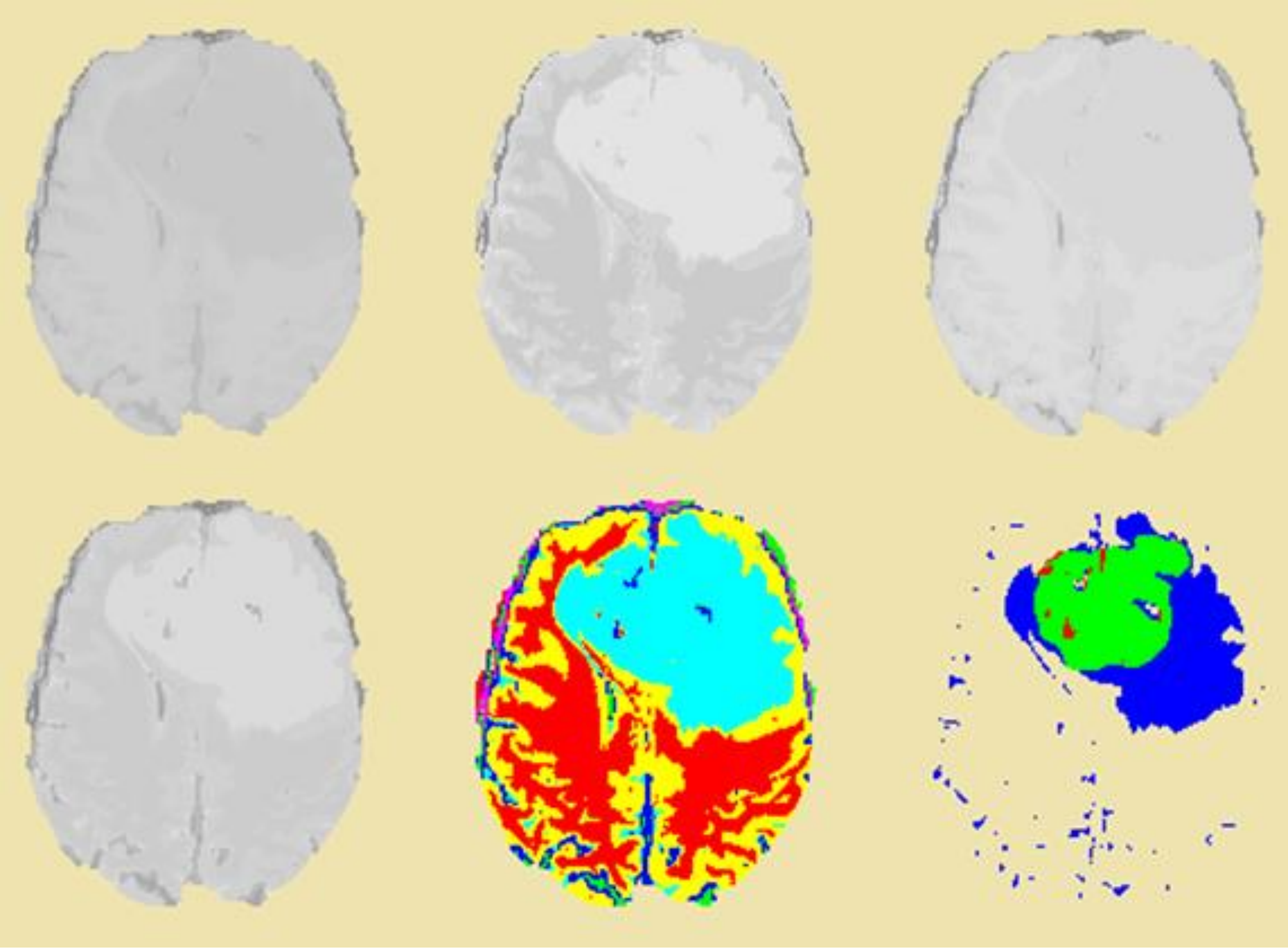
- Tests: m,m’=1.5:0.1:2.0 and c,c’=6:20

FCM initialization

- In every phase of the cascade, using actual data
- FCM clustering on each channel, c=3
  - T1: v_{11}, v_{12}, v_{13} T2: v_{21}, v_{22}, v_{23}
  - T1C: v_{31}, v_{32}, v_{33} FLAIR: v_{41}, v_{42}, v_{43}
- $3^4=81$ cluster prototype candidates
  - $[v_{1\alpha}, v_{2\beta}, v_{3\gamma}, v_{4\delta}]^T$ with $\alpha, \beta, \gamma, \delta = 1 \ldots 3$
- Ranking prototype candidates
  - Average distance to input vectors
- Select the best ranked c candidates

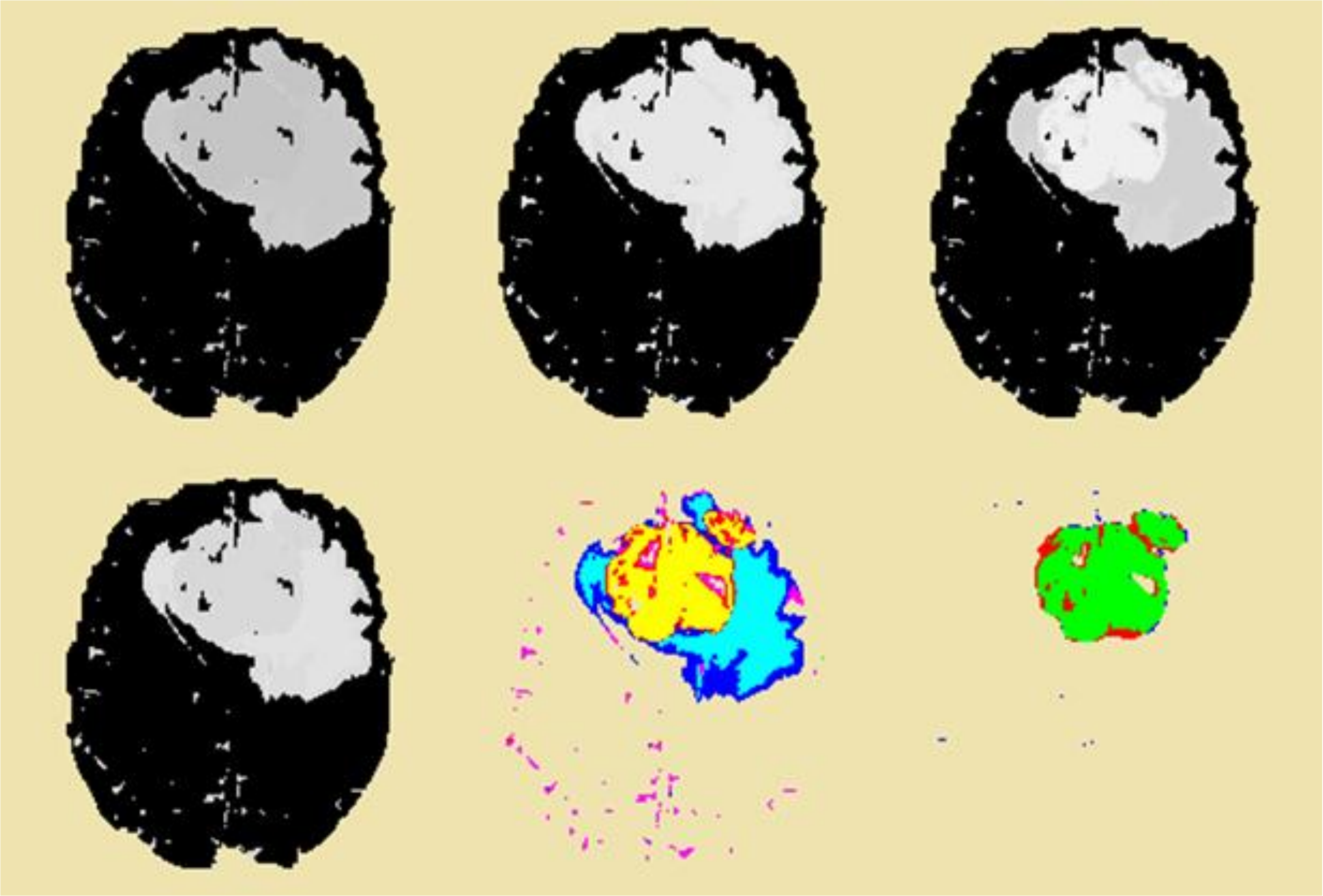
Outcome of Cascade FCM’s first step

c=6 classes, fuzzy exponent m=1.5



Outcome of Cascade FCM’s second step

c'=6 classes, fuzzy exponent m’=1.5



Measuring accuracy

- Based on ground truth
  - TP, TN, FP, FN
- Jaccard Index JI = TP/(TP+FP+FN)
- Dice Score DS = 2TP/(2TP+FP+FN) = 2JI/(1+JI)

- For each volume
  - $(6x15)^2 = 8100$ tests in unsupervised mode
  - Best Jaccard Index (MAX), Average JI (AVG)

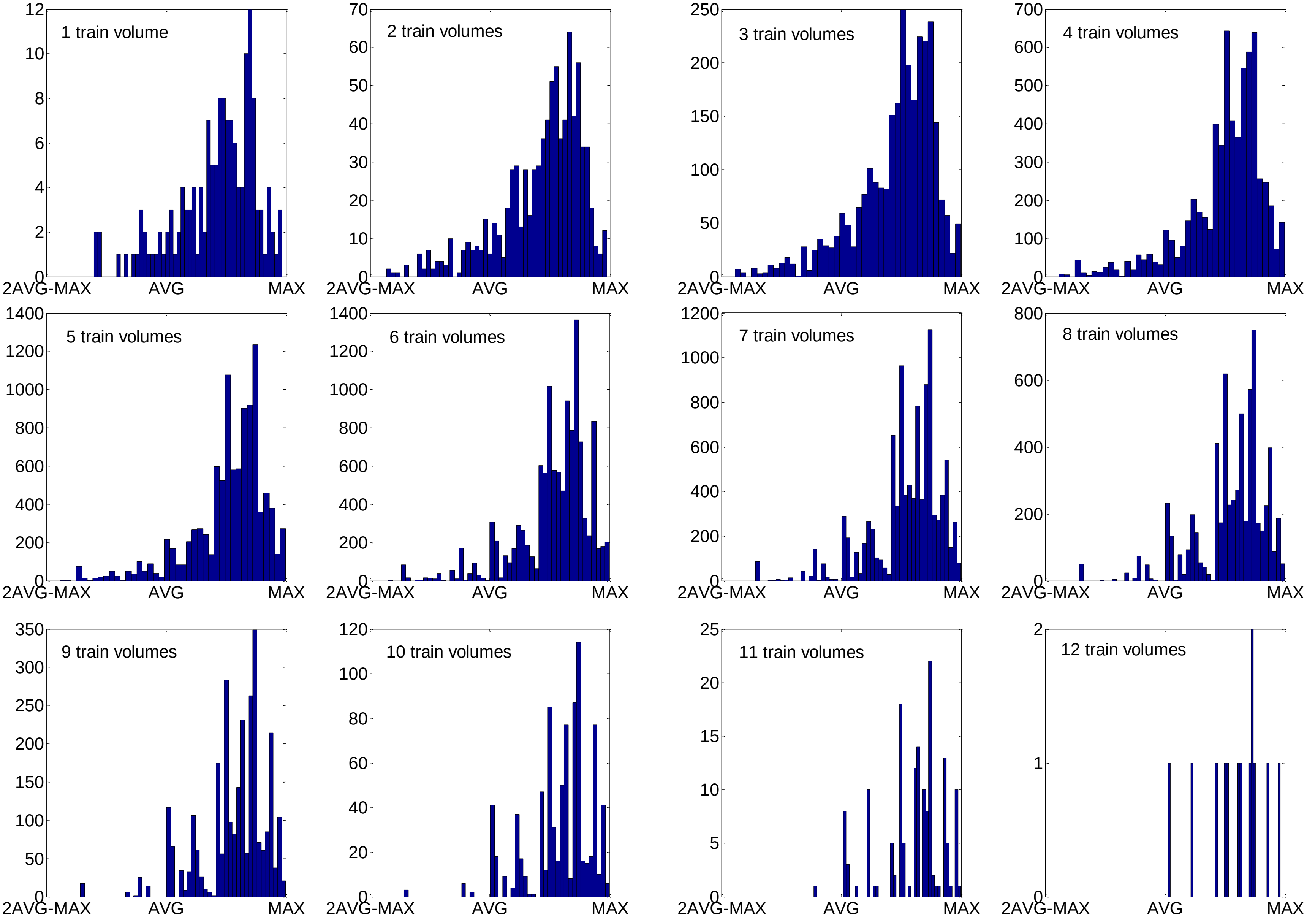
Average and maximum accuracy

Volume	Jaccard index		Dice score	
	AVG	MAX	AVG	MAX
HG01	0.7505	0.8058	0.8575	0.8925
HG02	0.3622	0.6097	0.5318	0.7576
HG03	0.7489	0.8155	0.8564	0.8984
HG04	0.6321	0.6498	0.7746	0.7878
HG05	0.2275	0.3143	0.3706	0.4783
HG06	0.5610	0.5969	0.7188	0.7476
HG07	0.3293	0.4358	0.4955	0.6070
HG09	0.3879	0.4929	0.5590	0.6603
HG11	0.4901	0.5954	0.6578	0.7464
HG12	0.0884	0.1422	0.1625	0.2490
HG13	0.2886	0.6073	0.4480	0.7557
HG14	0.5643	0.6468	0.7215	0.7855
HG15	0.7030	0.7343	0.8256	0.8468

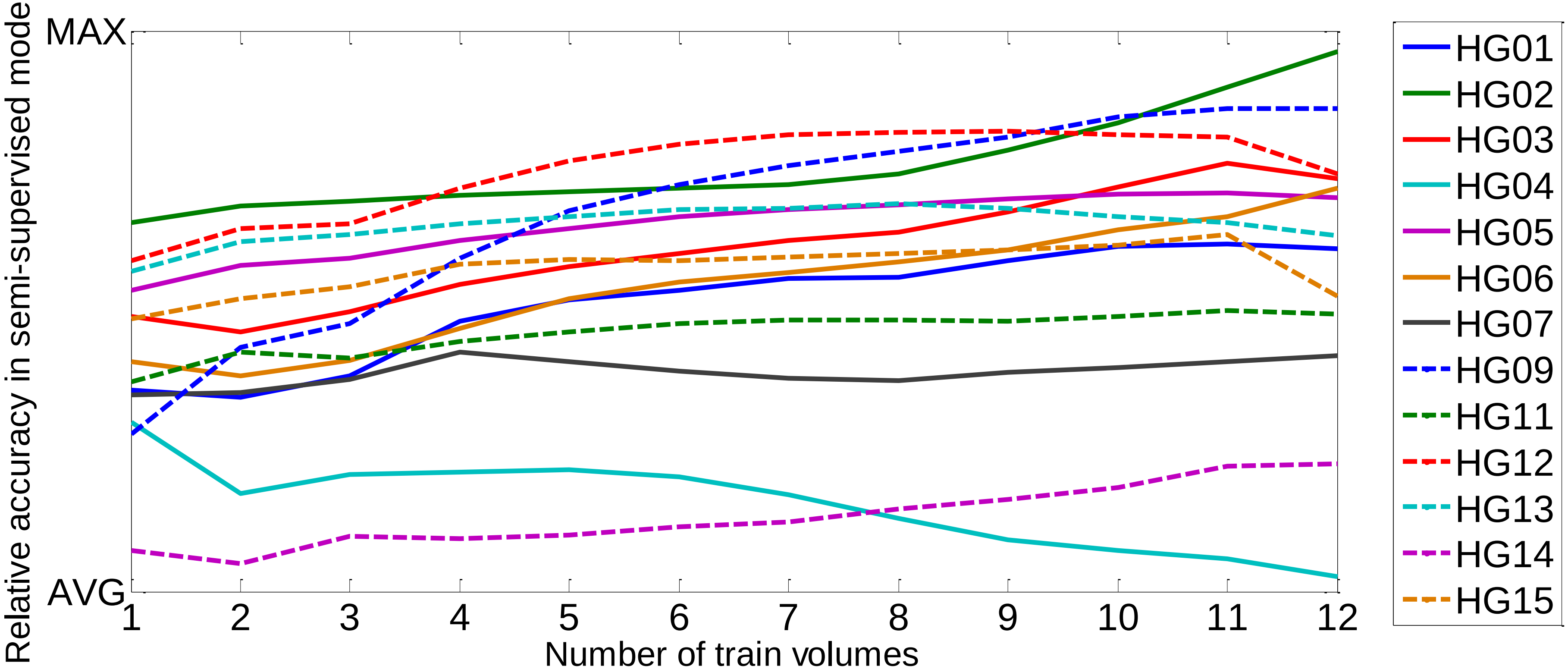
Semi-supervised clustering

- Test parameters (c,m,c’,m’)
- 13 volumes divided to: train data and test data
  - $2^{13}-2=8190$ possible cases
- Best parameter set obtained for train data were fed to test data clustering
- Jaccard Index vs. AVG, MAX
- For each volume and every number of train volumes, extracted average JI

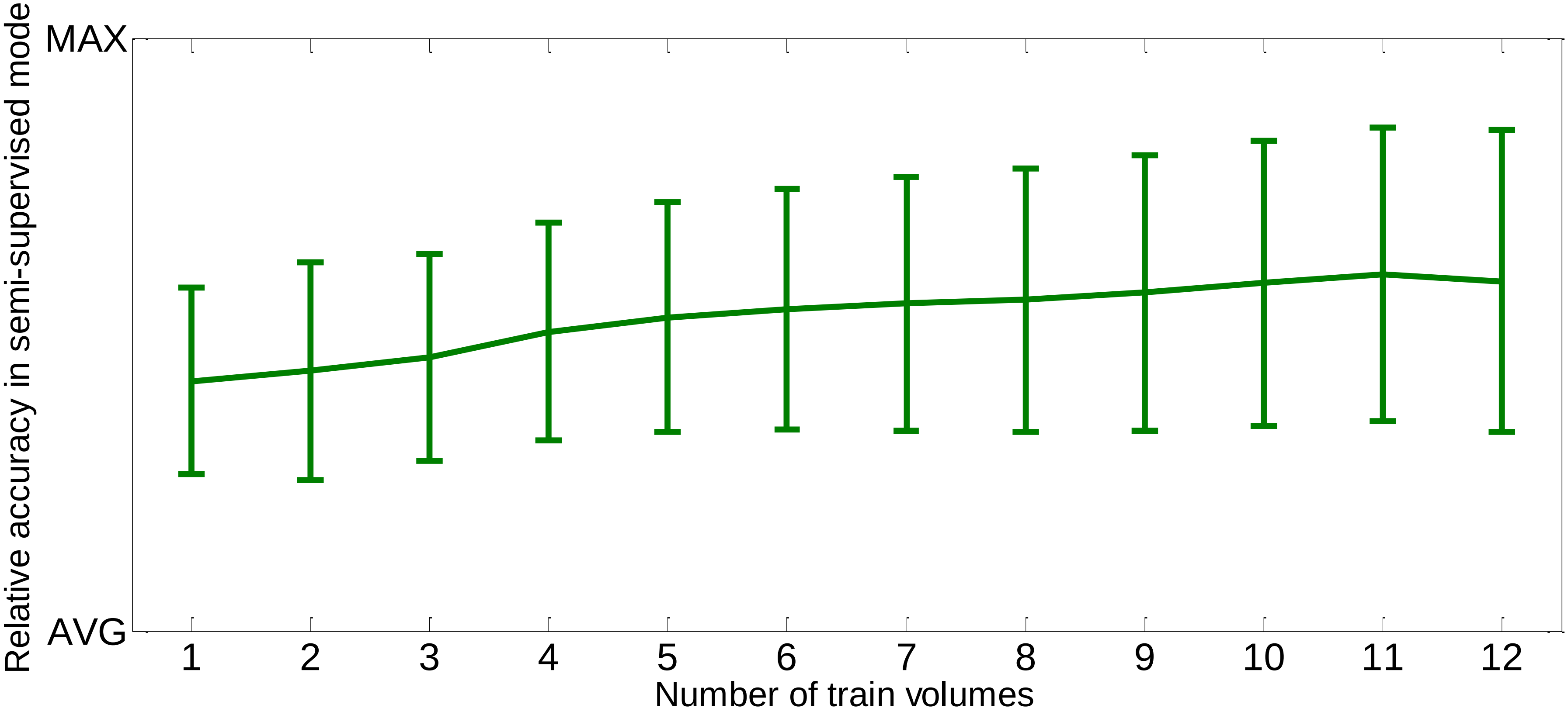
Histogram of Relative JI vs [AVG,MAX]



Relative accuracy vs. Train volume count

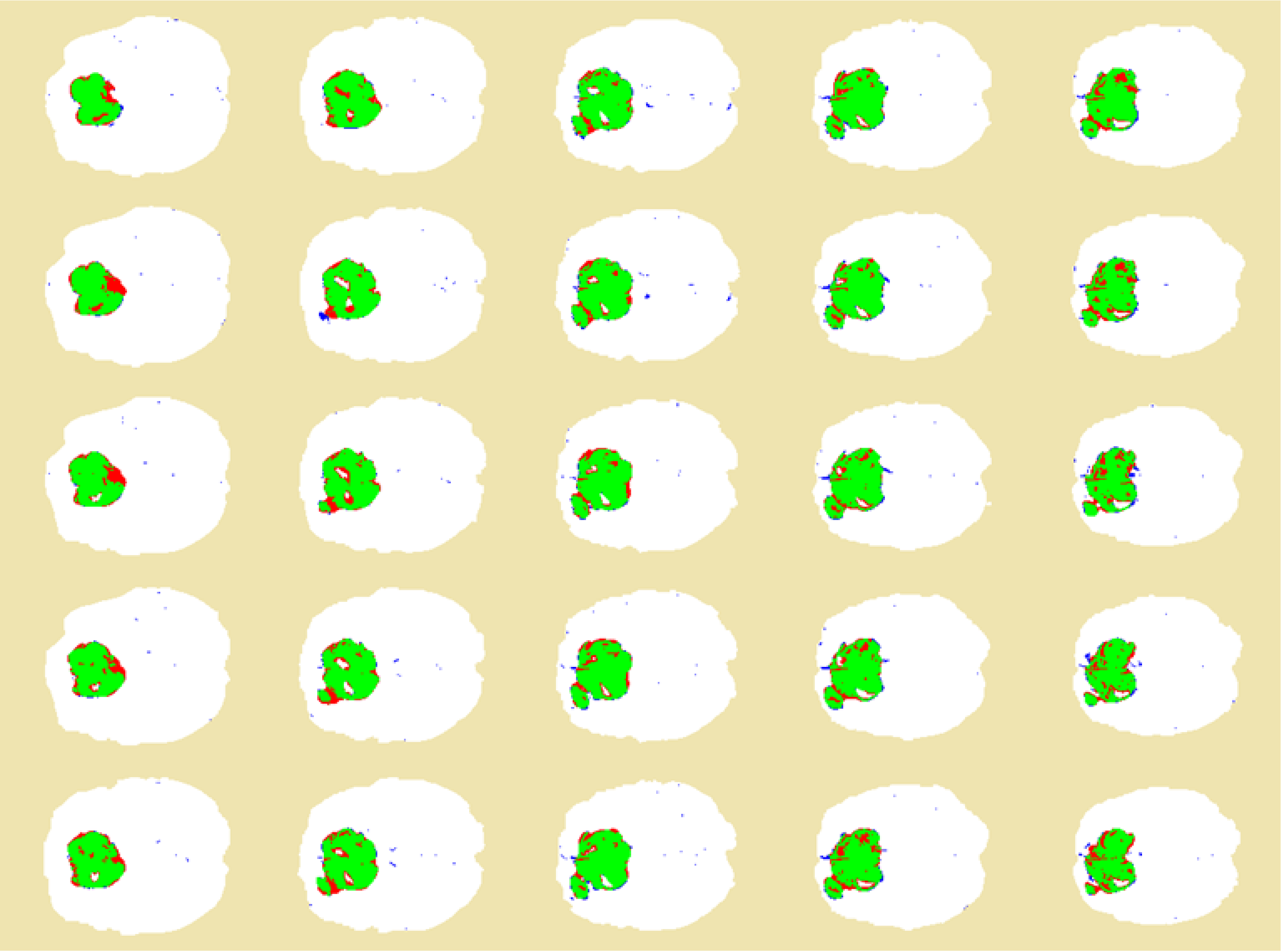


AVG & SD of accuracy vs. [AVG,MAX] interval



Results in consecutive slices

Volume HG15, slices 96-120, TP: green, FN: red, FP: blue, TN: white



Some more references

- Review on tumor segmentation methods
    - Gordillo et al.: Magn. Res. Imag. 31:1426-1438, 2013
- Two methods presented
    - Szilágyi L, et al., ICONIP 2015, LNCS 9492:174-181
    - Kapás Z, Szilágyi L, et al., MDAI 2016 (in press)

Conclusions

- Two approaches to tumor detection
- Preliminary results
- DS > 0.5 can detect most tumors
- Aim
  - High sensitivity utmost important
  - Specificity also matters
  - Detect smaller tumors
  - Detect low-grade tumors

Any questions?