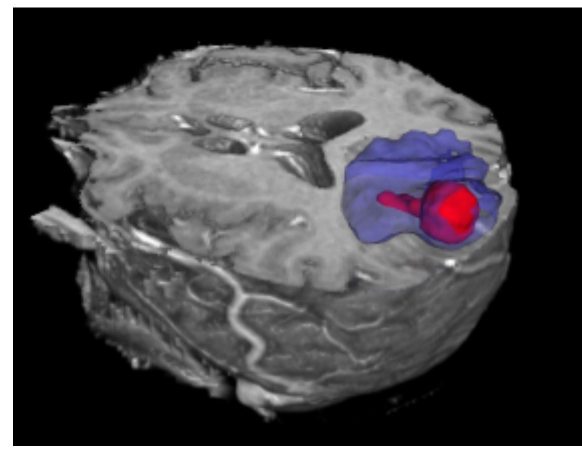
Mauricio Reyes,
Medical Image Analysis Group
Institute for Surgical Technology and Biomechanics
Univ. of Bern, Switzerland

BraTS

- Aim: Objective benchmark of different algorithms on a standardized dataset
- Low- and high-grade gliomas
- 65 synthetic datasets
- 65 patient datasets (multiple centers, different scanners & acquisition protocols)
- Standardized pre-processing (skull-stripping, rigid registration, isotropic resampling)
- Groundtruth from fused manual segmentations by 7 independent experts (Bern, Debrecen, Boston)



Concept BraTS

Miccai Segmentation challenge

1. Submission of paper contrib.
2. Training phase
3. On-site evaluation
4. Paper preparation

Concept BraTS

Miccai Segmentation challenge – Organization point of view

1. Submission of paper contrib.
2. Training phase
3. On-site evaluation
4. Paper preparation

Data selection

Data pre-processing

Data segmentation

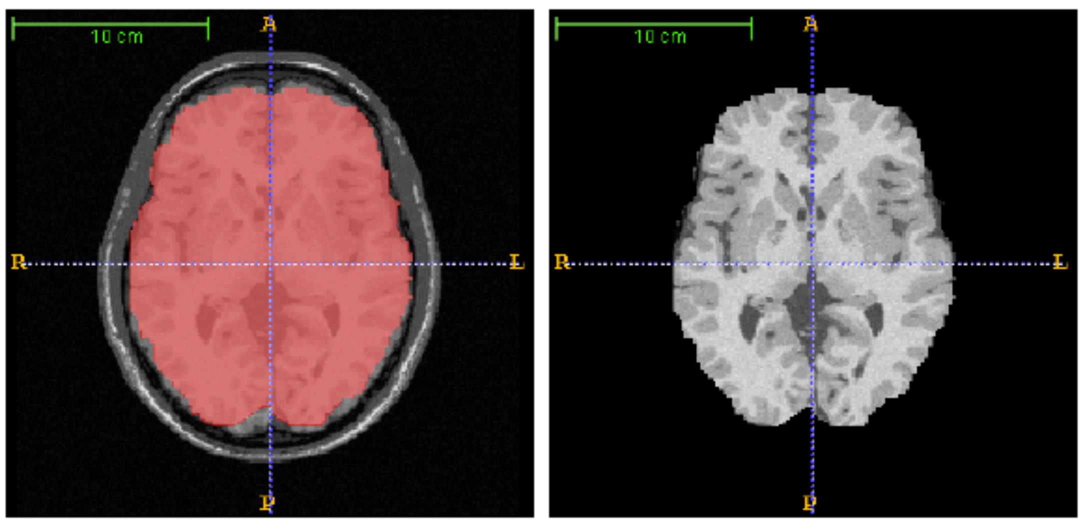
Checking & uploading

MRI modalities

- T1-weighted, native image, sagittal or axial 2D acquisitions, with 1-6mm slice thickness.
- T1-weighted, post Gd image, with 3D acquisition and 1 mm isotropic voxel size for most patients.
- T2-weighted image, axial 2D acquisition, with 2-6 mm slice thickness.
- T2-weighted FLAIR image, axial, coronal, or sagittal 2D acquisitions, 2-6 mm slice thickness.

Pre-processing

– Alignment to T1
– Skullstripping (semiautomatic)
– 1mm isotropic resampling



BraTS 2012

- 13 total participants
- 7 on-site participants
- Mix of automic and semi-automic methods
- Evaluation of complete tumor and core (2 compartments)
- Computation time not counted for ranking

BRATS 2012 Real data Dice (in %)	whole	LG/HG	core	LG/HG
Bauer	60	34 / 70	29	39 / 26
Geremia	61	58 / 63	23	29 / 20
Hamamci	69	46 / 78	37	43 / 35
Shin	32	44 / 27	9	0 / 12
Subbanna	29	0 / 32	27	0 / 30
Zhao (I)	34	na / 34	37	na / 37
Zikic	70	49 / 77	25	28 / 24

BRATS 2012 Synthetic data Dice (in %)	whole	LG/HG	core	LG/HG
Bauer	87	87 / 88	81	86 / 78
Geremia	83	83 / 82	62	54 / 66
Hamamci	82	74 / 85	69	46 / 80
Shin	8	4 / 10	3	2 / 4
Subbanna	81	81 / 81	41	42 / 40
Zhao	na	na / na	na	na / na
Zikic	91	88 / 93	86	84 / 87

BraTS 2013

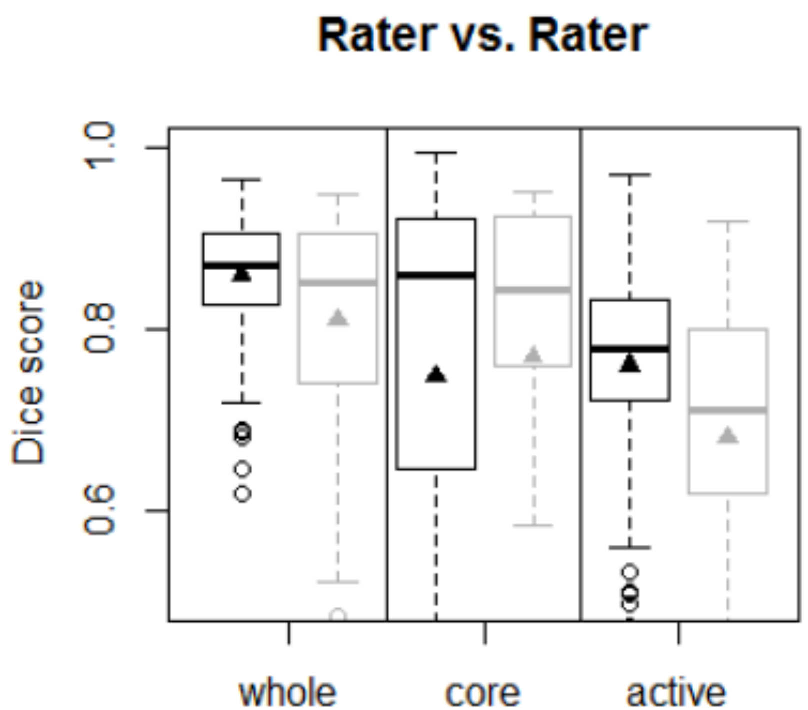
MICCAI 2013 NAGOYA JAPAN

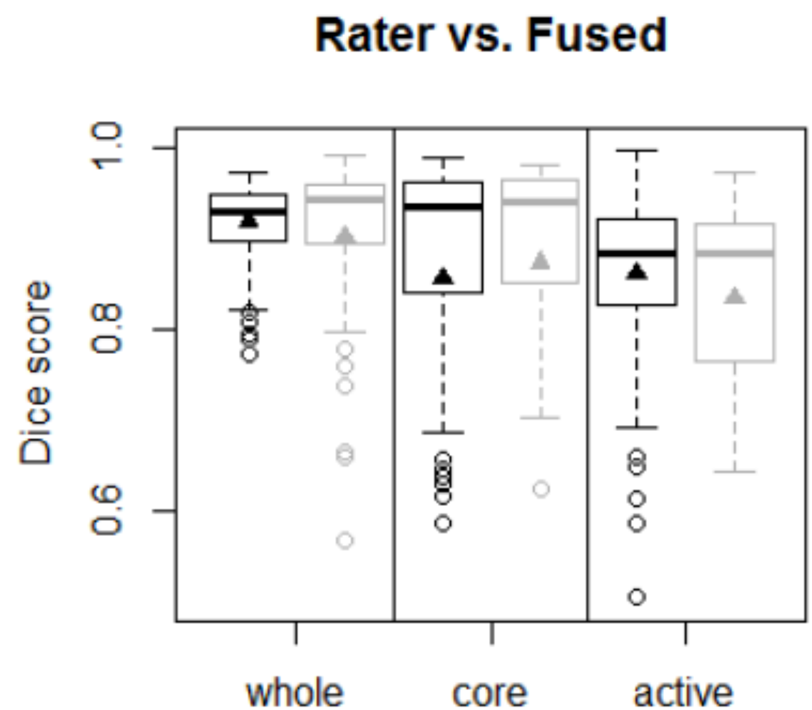
VIRTUALSKELETON database

- 11 total participants
- 7 on-site participants
- Only automatic methods
- Mostly decision forest – based methods (due to their success in 2012)
- Evaluation of complete tumor, core and enhancing tumor (3 compartments)
- Data from BraTS 2012 + additional NCIA images
- Computation time not counted for ranking

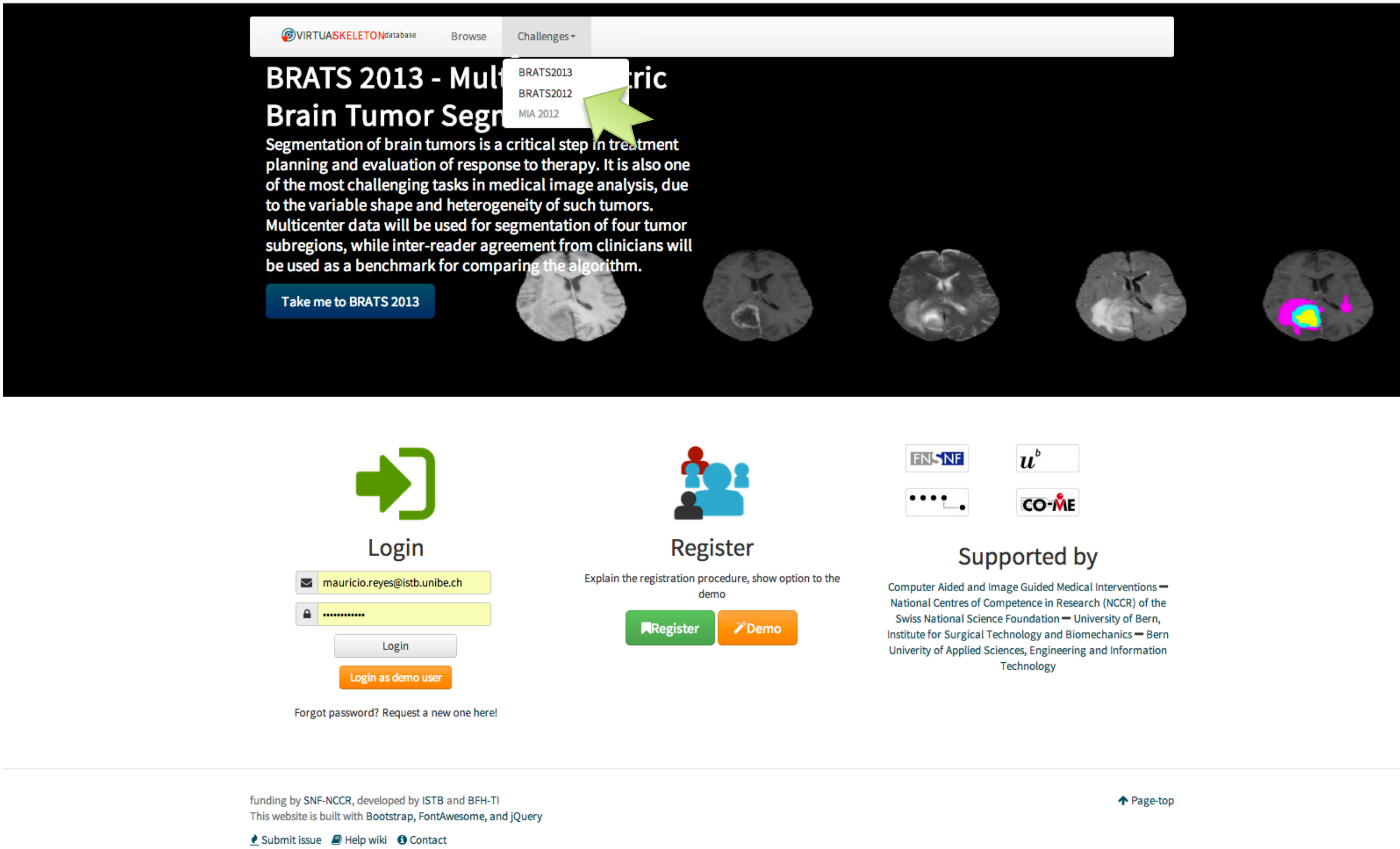
BRATS 2013			
Real data Dice (in %)	whole *HG only*	core *HG only*	active
Cordier	84	68	65
Doyle	71	46	52
Festa	72	66	67
Meier	82	73	69
Reza	83	72	72
Tustison	87	78	74
Zhao (II)	84	70	65

Manual Segmentations





Evaluation platform – Virtual Skeleton Database (VSD)



VSD Features

- User management
- File management
  - File Sharing
  - Folder structure + sharing
- Multiple file-type support
  - Image formats (dicom, hdr, mhd, mha, nii)
  - Statistical models (HDF5)
- Full-text search
- Meta-data (incl. FMA ontology)
- Integration of external libraries - ITK and HDF5
- System integration (webdav)
- Discussions (forum), rating and comments on datasets
- De-identification
- Advanced upload
  - Segmentation recognition
  - Dataset recognition
  - Version handling
  - Auto-linking of objects
- Hosting MICCAI segmentation challenge

Evaluation metrics

- Dice coeff.
- Specificity
- Sensitivity
- Average/Haussdorf distance
- Kappa

- Limitation: no “clever” combination of metrics. Used aggregation of rankings

Ranking system

- Per-metric ranking
- Divided by regions (complete, core, enhancing)

Dice		
complete	core	enhancing
0.79 (1)	0.65 (1)	0.53 (1)

Sensitivity		
complete	core	enhancing
0.81 (2)	0.73 (1)	0.66 (1)

Positive Predictive Value		
complete	core	enhancing
0.83 (1)	0.70 (1)	0.51 (1)

Improvements to labels (after BraTS 2012)

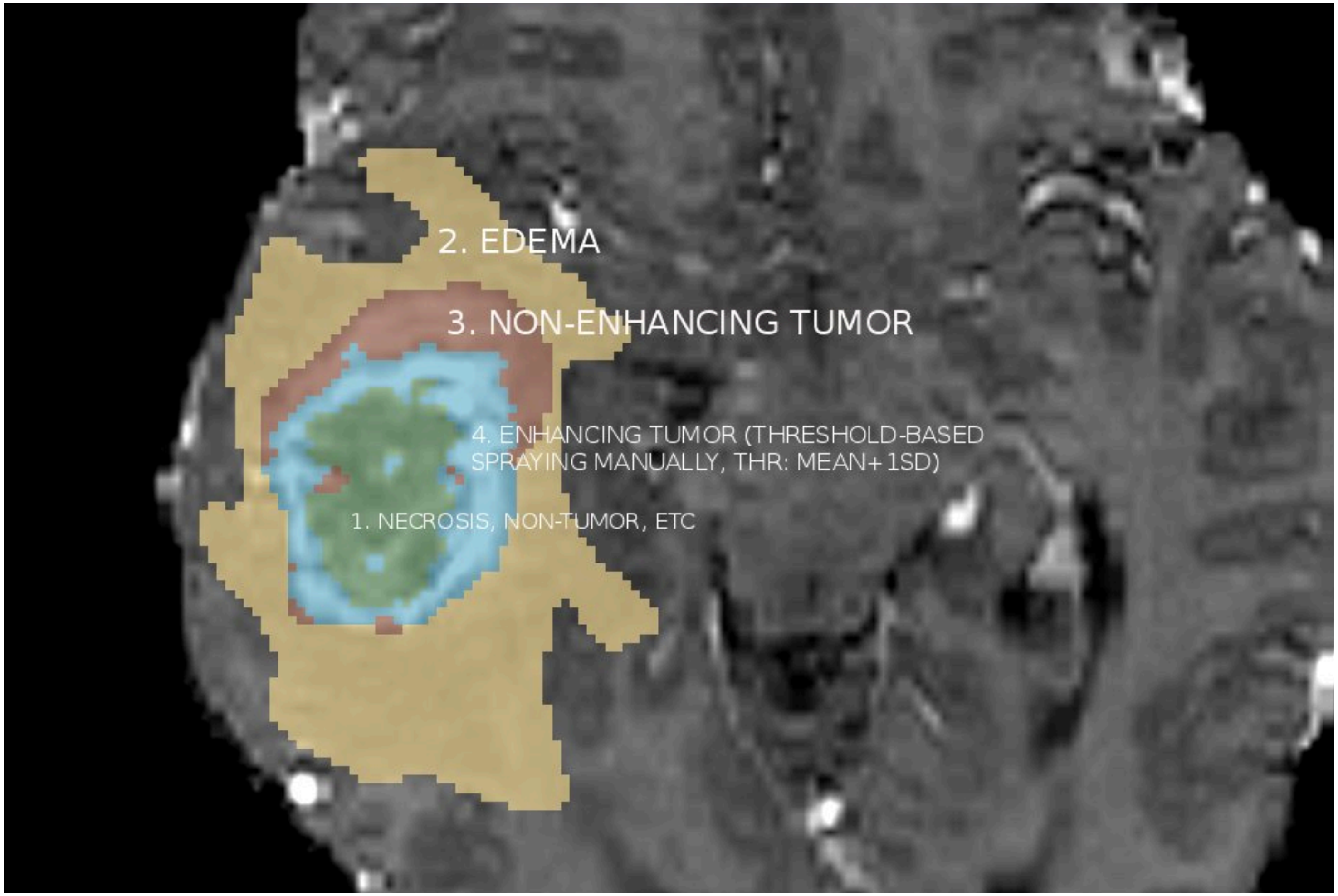
- 2-labels:

  1 for edema
  2 for active tumor

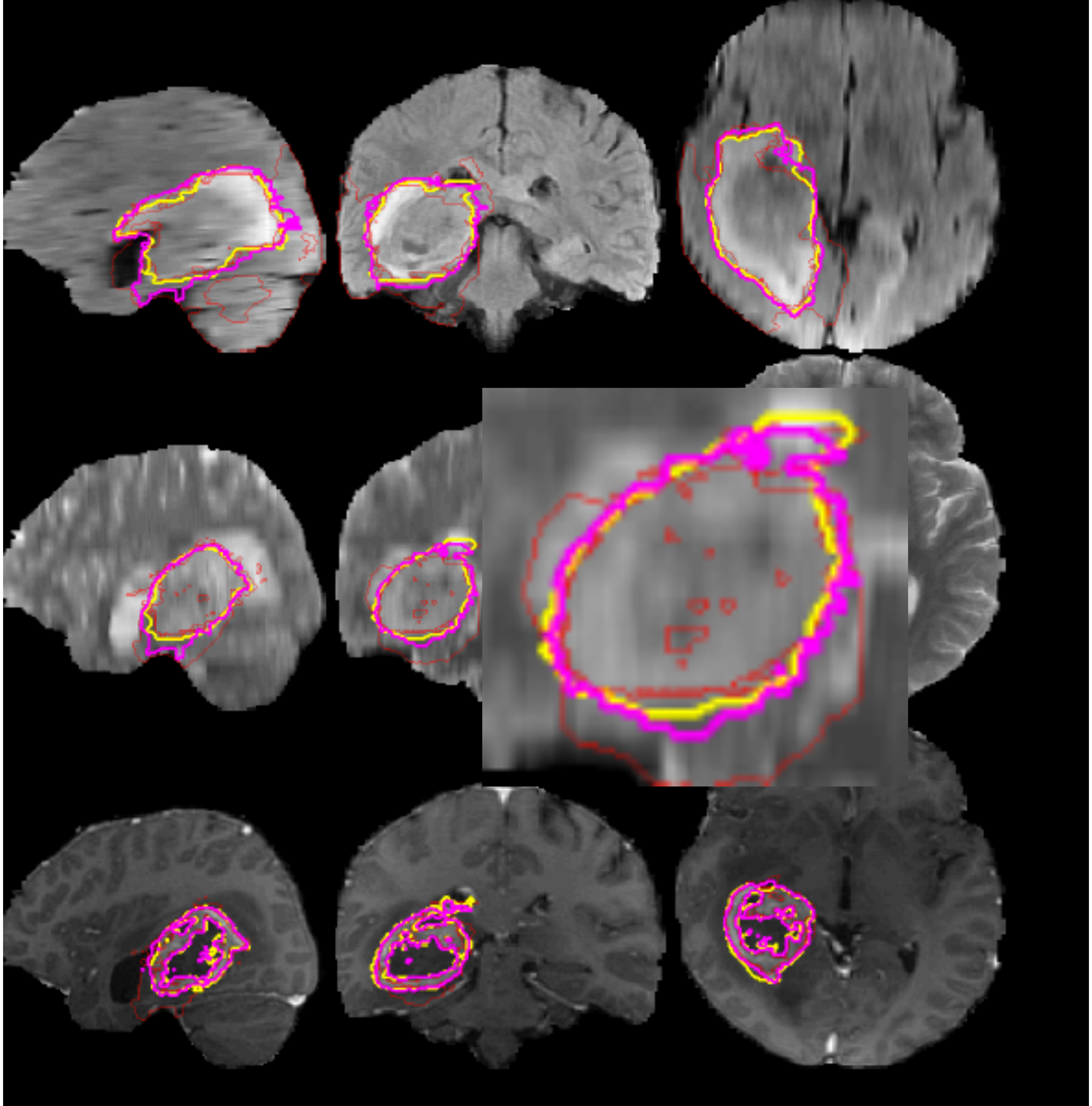
- 4-labels:

  1 for necrosis
  2 for edema
  3 for non-enhancing tumor
  4 for enhancing tumor
  0 for everything else

Evaluation



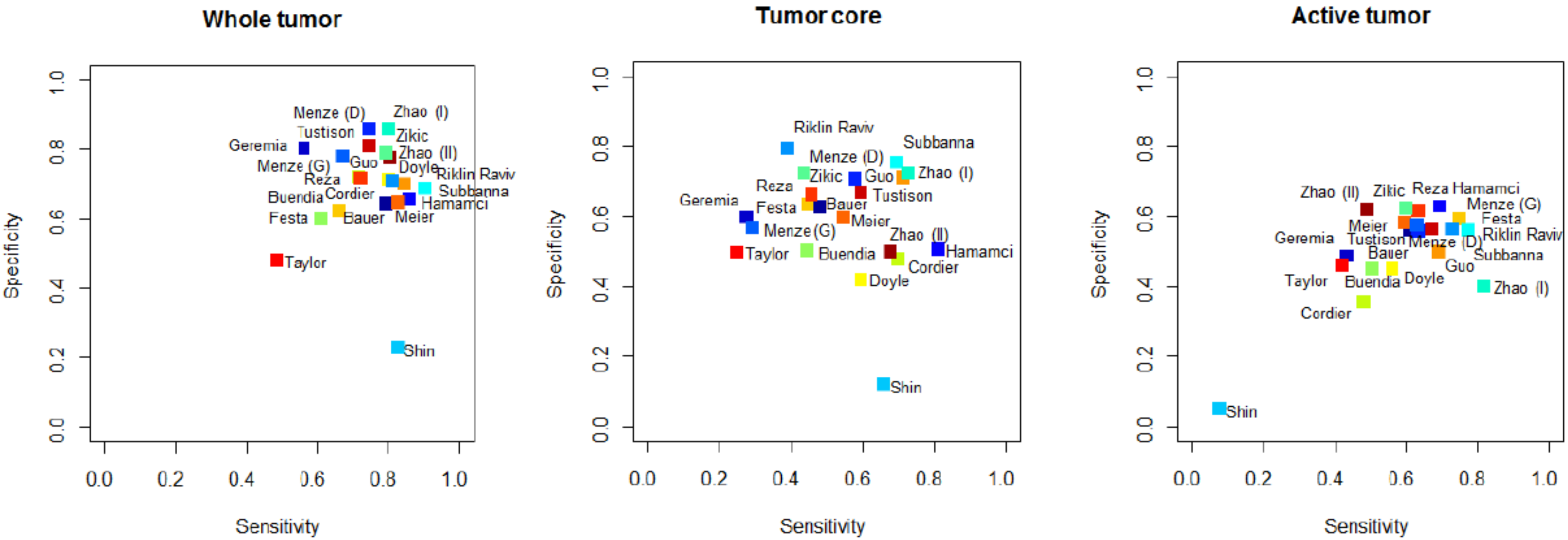
Region 1: complete tumor (labels 1+2+3+4 for patient data, labels 1+2 for synthetic data
Region 2: Tumor core (labels 1+3+4 for patient data, label 2 for synthetic data)
Region 3: Enhancing tumor (label 4 for patient data, n.a. for synthetic data



Results

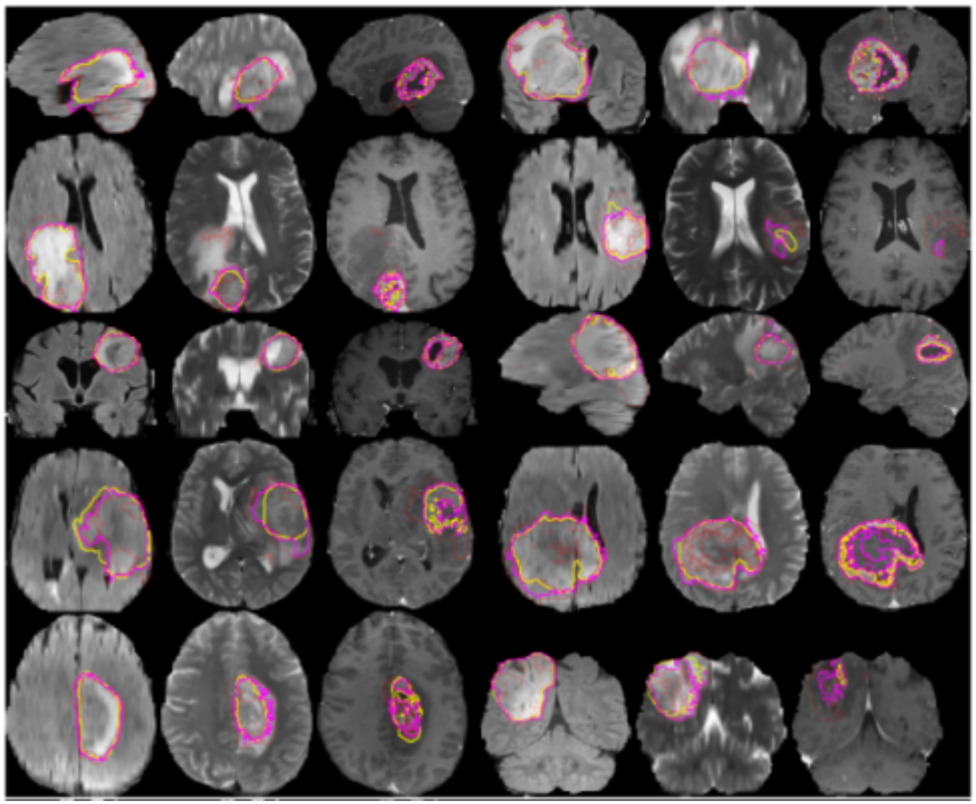
	whole		core		active	time (min) (arch).
Dice (in %)		*LG/HG*		*LG/HG*		
Bauer	68	*49/74*	48	*30/54*	**57**	8 (CPU)
Buendia	57	*19/71*	42	*8/54*	**45**	X(XX)
Cordier	68	*60/71*	51	*41/55*	39	200 (CPU)
Doyle	**74**	*63/78*	44	*41/45*	42	X (XX)
Festa	62	*24/77*	50	*33/56*	**61**	30 (CPU)
Geremia	62	*55/65*	32	*34/31*	**42**	X (XX)
Guo	**74**	*71/75*	**65**	*59/67*	**49**	X (XX)
Hamamci	**72**	*55/78*	57	*40/63*	**59**	20 (CPU)
Meier	**69**	*46/77*	50	*36/55*	**57**	6 (CPU)
Menze (D)	**78**	*81/76*	**58**	*58/59*	**54**	20 (CPU)
Menze (G)	69	*48/77*	33	*9/42*	**53**	10 (CPU)
Reza	70	*52/77*	47	*39/50*	**55**	X (XX)
Riklin Raviv	74	*na/74*	50	*na/50*	**58**	8 (CPU)
Shin	30	*28/31*	17	*22/15*	5	8 (Cluster)
Subbanna	**75**	*55/82*	**70**	*54/75*	**59**	70 (CPU)
Taylor	44	*24/51*	28	*11/34*	41	1 (Cluster)
Tustison	**75**	*68/78*	**55**	*42/60*	**52**	100 (CPU)
Zhao (I)	**82**	*78/84*	**66**	*60/68*	**49**	X (XX)
Zhao (II)	**76**	*67/79*	51	*42/55*	**52**	20 (CPU)
Zikic	**75**	*62/80*	47	*33/52*	**56**	2 (CPU)
Best Combination	88	*86 / 89*	78	*66 / 82*	71	
Fused_4	82	*68 / 87*	73	*62 / 77*	65	

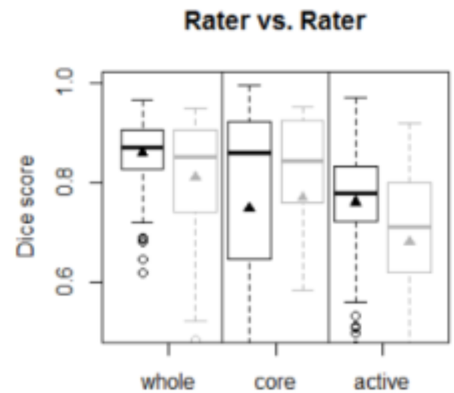
Results



Whole tumor
Tumor core
Active tumor
Specificity
Sensitivity
0.0
0.2
0.4
0.6
0.8
1.0
Menze (D)
Zhao (I)
Tustison
Zikic
Geremia
Zhao (II)
Menze (G)
Guo
Doyle
Riklin Raviv
Reza
Subbanna
Buendia
Cordier
Hamamci
Festa
Bauer
Meier
Taylor
Shin

BraTS Conclusion





	whole		core		active	time (min) (arch).
Dice (in %)		*LG/HG*		*LG/HG*		
Bauer	68	*49/74*	48	*30/54*	**57**	8 (CPU)
Buendia	57	*19/71*	42	*8/54*	**45**	X(XX)
Cordier	68	*60/71*	51	*41/55*	39	200 (CPU)
Doyle	**74**	*63/78*	44	*41/45*	42	X (XX)
Festa	62	*24/77*	50	*33/56*	**61**	30 (CPU)
Geremia	62	*55/65*	32	*34/31*	**42**	X (XX)
Guo	**74**	*71/75*	**65**	*59/67*	**49**	X (XX)
Hamamci	**72**	*55/78*	57	*40/63*	**59**	20 (CPU)
Meier	**69**	*46/77*	50	*36/55*	**57**	6 (CPU)
Menze (D)	**78**	*81/76*	**58**	*58/59*	**54**	20 (CPU)
Menze (G)	69	*48/77*	33	*9/42*	**53**	10 (CPU)
Reza	70	*52/77*	47	*39/50*	**55**	X (XX)
Riklin Raviv	74	*na/74*	50	*na/50*	**58**	8 (CPU)
Shin	30	*28/31*	17	*22/15*	5	8 (Cluster)
Subbanna	**75**	*55/82*	**70**	*54/75*	**59**	70 (CPU)
Taylor	44	*24/51*	28	*11/34*	41	1 (Cluster)
Tustison	**75**	*68/78*	**55**	*42/60*	**52**	100 (CPU)
Zhao (I)	**82**	*78/84*	**66**	*60/68*	**49**	X (XX)
Zhao (II)	**76**	*67/79*	51	*42/55*	**52**	20 (CPU)
Zikic	**75**	*62/80*	47	*33/52*	**56**	2 (CPU)
Best Combination	88	*86 / 89*	78	*66 / 82*	71	
Fused_4	82	*68 / 87*	73	*62 / 77*	65	

Recommendations

- Long-term impact => sustainability and consistency of the comparisons
- Using training and evaluation datasets. Publications
- Cross-evaluation and cross-evaluation…
- Close or far away from clinical scenarios?

BraTS

Multimodal Brain Tumor Segmentation Challenge

Thanks for your attention!