Grand Challenges in Medical Image Analysis

October 29, 2018
Benchmarking Algorithm Performance for Research
Amsterdam

Bram van Ginneken

Diagnostic Image Analysis Group, Radboud University Medical Center;
Fraunhofer MEVIS, Bremen; Thirona, Nijmegen

Radboudumc

Disclosures

- Developer CAD4TB (Delft Imaging Systems): royalties & funding
- Co-founder and CSO Thirona: stock, royalties & funding
- Developer Veolity (MeVis Medical Solutions) & DynaCAD Lung (InVivo): royalties & funding
- ScreenPoint is a spin-off from my group, led by Nico Karssemeijer; my group receives royalties
- Group funding: Canon Medical, Siemens Healthineers, Philips Medical, Sectra, Novartis

Tutorial on Benchmarking Algorithm Performance

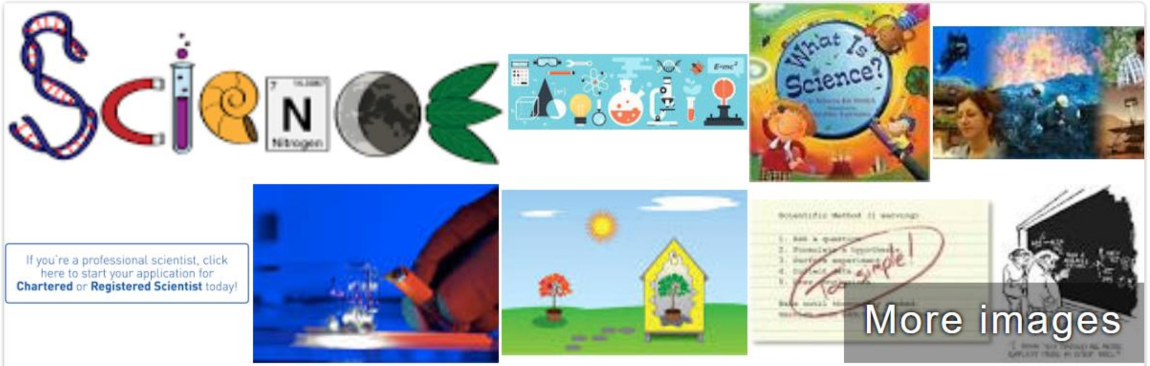
October 29th at the 14th IEEE International Conference on eScience 2018, Amsterdam, the Netherlands

View the Project on GitHub
NLeSC/IEEE-eScience-Tutorial-Designing-Benchmarks

About

One of the currently most well-known benchmarks for algorithm performance is ImageNet. Many challenges have been organized using this database, with the latest challenge now running on Kaggle. In various scientific disciplines there is a growing interest to benchmark algorithm performance on research data. Many algorithms are proposed in the literature, but there is a growing need to compare them on the same data, using the same metrics and ground truth to compare their performance for a specific task. Organizing these open online benchmarks, will not only increase insight into which algorithms perform best for a given task, but open up these tasks for a wider audience to test their algorithms on, which could lead to new breakthroughs in the field. This tutorial shows two research fields with a longer history in benchmarking algorithm performance, such as medical image analysis (Bram van Ginneken) and multimedia information retrieval (Maria Eskevich). Mike Lees will talk about how benchmarking is being introduced for slum detection on sattelite images, a field with strong restrictions on data sharing, and Kasper Marstal

Radboudumc



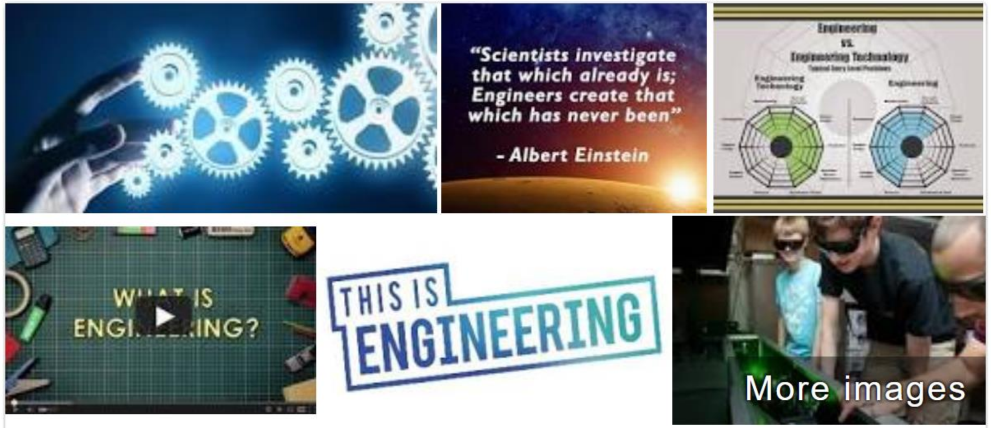
Science

Science is a systematic enterprise that builds and organizes knowledge in the form of testable explanations and predictions about the universe. The earliest roots of science can be traced to Ancient Egypt and Mesopotamia in around 3500 to 3000 BCE. Wikipedia

People also search for

View 10+ more

Research Chemistry Biology Mathemat… Technology



Engineering

Discipline

Engineering is the creative application of science, mathematical methods, and empirical evidence to the innovation, design, construction, operation and maintenance of structures, machines, materials, devices, systems, processes, and organizations for the benefit of humankind. Wikipedia

People also search for

View 10+ more

Technology Science Design Business Research

Why did I get involved in organizing challenges?

- Frustration about the typical output in my field: a paper where you present a new method and always claim it outperforms previously proposed methods. After almost a decade, I read too many of those papers that I did not believe.

Table 2. Segmentation error in comparison with a standard ASM and numbers from previous work (customized ASM [22] and deformable simplex mesh [5]), given as $\mu \pm \sigma$

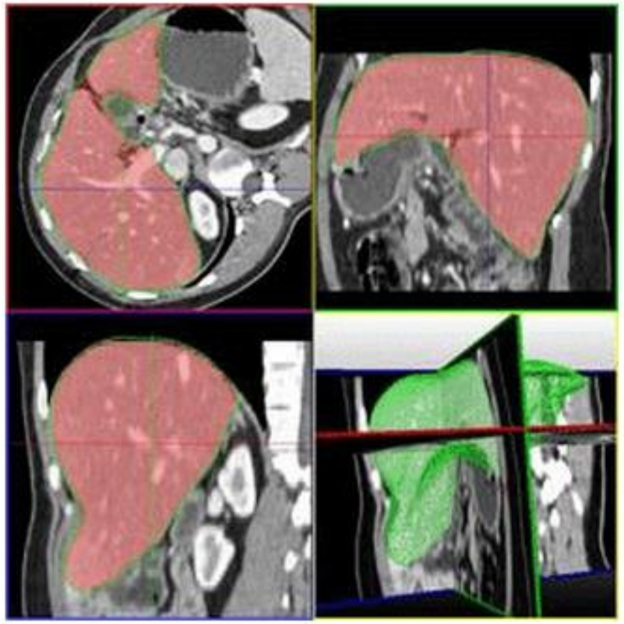
Segmentation method	$D_{avg}[mm]$	$D_{RMS}[mm]$	$V_D[\%]$
Deformable model	1.6 ± 0.5	3.3 ± 1.2	5.1 ± 1.4
Active Shape Model	2.9 ± 1.1	5.2 ± 2.3	8.9 ± 2.4
Lamecker et al. [22]	2.3 ± 0.3	3.1 ± 0.5	7.0 ± 1.8
Soler et al. [5]	$2 \pm ?$	*n.a.*	*n.a.*

Fig. 5. Transversal, sagittal and coronal slices for the image with median average surface error. The result of the deformable model is displayed in white, the manually traced reference contour in dark gray.

SLIVER07

Home Rules Results Register Download Submit FAQ

Segmentation of the Liver 2007



Welcome to the website of the Segmentation of the Liver Competition 2007 (SLIVER07). The goal of this competition is to compare different algorithms to segment the liver from clinical 3D computed tomography (CT) scans.

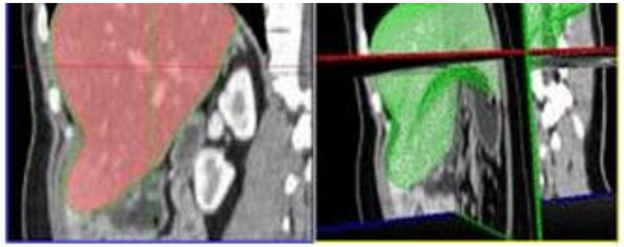
This competition started as part of the workshop **3D Segmentation in the Clinic: A Grand Challenge**, on October 29, 2007 in conjunction with **MICCAI 2007**. Teams that participated in

SLIVER07

Home Rules Results Register Download Submit FAQ

Segmentation of the Liver 2007

The workshop is over, but through this website, the liver segmentation competition continues. You can browse the **results** of various systems, and read papers and descriptions about the methods that have been applied to the SLIVER07 data set. If you want to join the competition, you can **register** a team, **download** training and test data, and **submit** the results of your own algorithms, provided you adhere to and agree with **the rules**. More information is available in the answers to **frequently asked questions**.



Welcome to the website of the Segmentation of the Liver Competition 2007 (SLIVER07). The goal of this competition is to compare different algorithms to segment the liver from clinical 3D computed tomography (CT) scans.

This competition started as part of the workshop **3D Segmentation in the Clinic: A Grand Challenge**, on October 29, 2007 in conjunction with **MICCAI 2007**. Teams that participated in

Radboudumc

Results

For a table with more detailed results, click on the rank of the submission you are interested in.

Rank	Team	System	Submission Date	Avg Total Score
1	LME Erlangen	Semi-automatic	2010-01-14	84.6
2	Niki-Lab	Semi-automatic	2012-11-02	84.5
3	LiverPlanner	Interactive	2008-02-25	82.1
4	Afifi	Semi-automatic	2011-11-24	81.8
5	liver sirD	Semi-automatic	2011-11-30	80.5
6	liver sirA	Semi-automatic	2011-11-02	80.0
7	Niki-Lab	Semi-automatic	2011-08-23	79.7
8	Eccet	Interactive	2007-10-29	77.9
9	ZIB-Charite	Automatic	2007-10-22	77.3
10	Eccet	Interactive	2007-10-29	77.0
11	LME Erlangen	Automatic	2009-03-11	76.8
12	LiverV	Semi-automatic	2007-10-26	76.4
13	Eccet	Interactive	2007-10-29	76.3
14	LivSpl	Automatic	2012-06-22	76.2
15	Eccet	Interactive	2007-10-24	76.1
16	amcmil	Semi-automatic	2007-10-22	75.5
17	Lara	Interactive	2007-10-19	75.1
18	Eccet	Interactive	2007-10-24	74.7
19	LiverPlanner	Interactive	2008-02-25	74.2
20	LME Erlangen	Automatic	2008-12-22	74.0
21	Eccet	Interactive	2007-10-24	73.8
22	VAVlab	Semi-automatic	2012-04-04	72.9
23	VMGPU	Interactive	2008-02-20	72.9
24	VMGPU	Interactive	2008-02-20	72.9
25	KMSeg	Automatic	2012-07-28	72.4
26	Afifi	Semi-automatic	2011-09-05	72.3
27	ABE_UPM	Semi-automatic	2011-12-22	70.5
28	vislib	Semi-automatic	2009-07-27	70.0
29	LRV	Automatic	2010-03-05	69.9
30	KMSeg	Automatic	2012-06-06	69.5
31	LME Erlangen	Semi-automatic	2008-01-24	69.4
32	Dieter Seghers	Automatic	2008-09-01	68.9
33	RITS_UT	Semi-automatic	2009-11-19	68.5
34	Team FFAL	Automatic	2010-03-01	68.4
35	HUJI-liver	Semi-automatic	2008-03-03	67.9

Table 2. Segmentation error in comparison with a standard ASM and numb previous work (customized ASM [22] and deformable simplex mesh [5]), given

Segmentation method	$D_{avg}[mm]$	$D_{RMS}[mm]$	$V_D[\%]$
Deformable model	1.6 ± 0.5	3.3 ± 1.2	5.1 ± 1.4
Active Shape Model	2.9 ± 1.1	5.2 ± 2.3	8.9 ± 2.4
Lamecker et al. [22]	2.3 ± 0.3	3.1 ± 0.5	7.0 ± 1.8
Soler et al. [5]	$2\pm?$	*n.a.*	*n.a.*

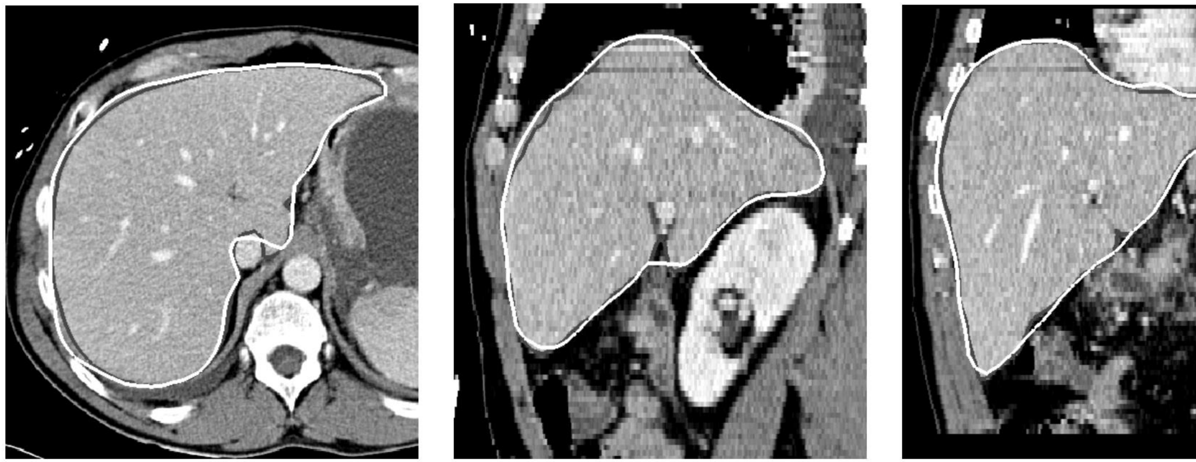


Fig. 5. Transversal, sagittal and coronal slices for the image with median surface error. The result of the deformable model is displayed in white, the r traced reference contour in dark gray.

Comparison and Evaluation of Methods for Liver Segmentation From CT Datasets

Tobias Heimann*, Bram van Ginneken, *Member, IEEE*, Martin A. Styner, *Member, IEEE*, Yulia Arzhaeva, Volker Aurich, Christian Bauer, Andreas Beck, Christoph Becker, Reinhard Beichel, *Member, IEEE*, György Bekes, Fernando Bello, *Member, IEEE*, Gerd Binnig, Horst Bischof, *Member, IEEE*, Alexander Bornik, Peter M. M. Cashman, Ying Chi, Andrés Córdova, Benoit M. Dawant, Márta Fidrich, Jacob D. Furst, Daisuke Furukawa, Lars Grenacher, Joachim Hornegger, *Member, IEEE*, Dagmar Kainmüller, Richard I. Kitney, Hidefumi Kobatake, Hans Lamecker, Thomas Lange, Jeongjin Lee, Brian Lennon, Rui Li, Senhu Li, Hans-Peter Meinzer, *Member, IEEE*, Gábor Németh, Daniela S. Raicu, Anne-Mareike Rau, Eva M. van Rikxoort, Mikaël Rousson, László Ruskó, Kinda A. Saddi, Günter Schmidt, Dieter Seghers, Akinobu Shimizu, *Member, IEEE*, Pieter Slagmolen, Erich Sorantin, Grzegorz Soza, Ruchaneewan Susomboon, Jonathan M. Waite, Andreas Wimmer, and Ivo Wolf

Radboudumc

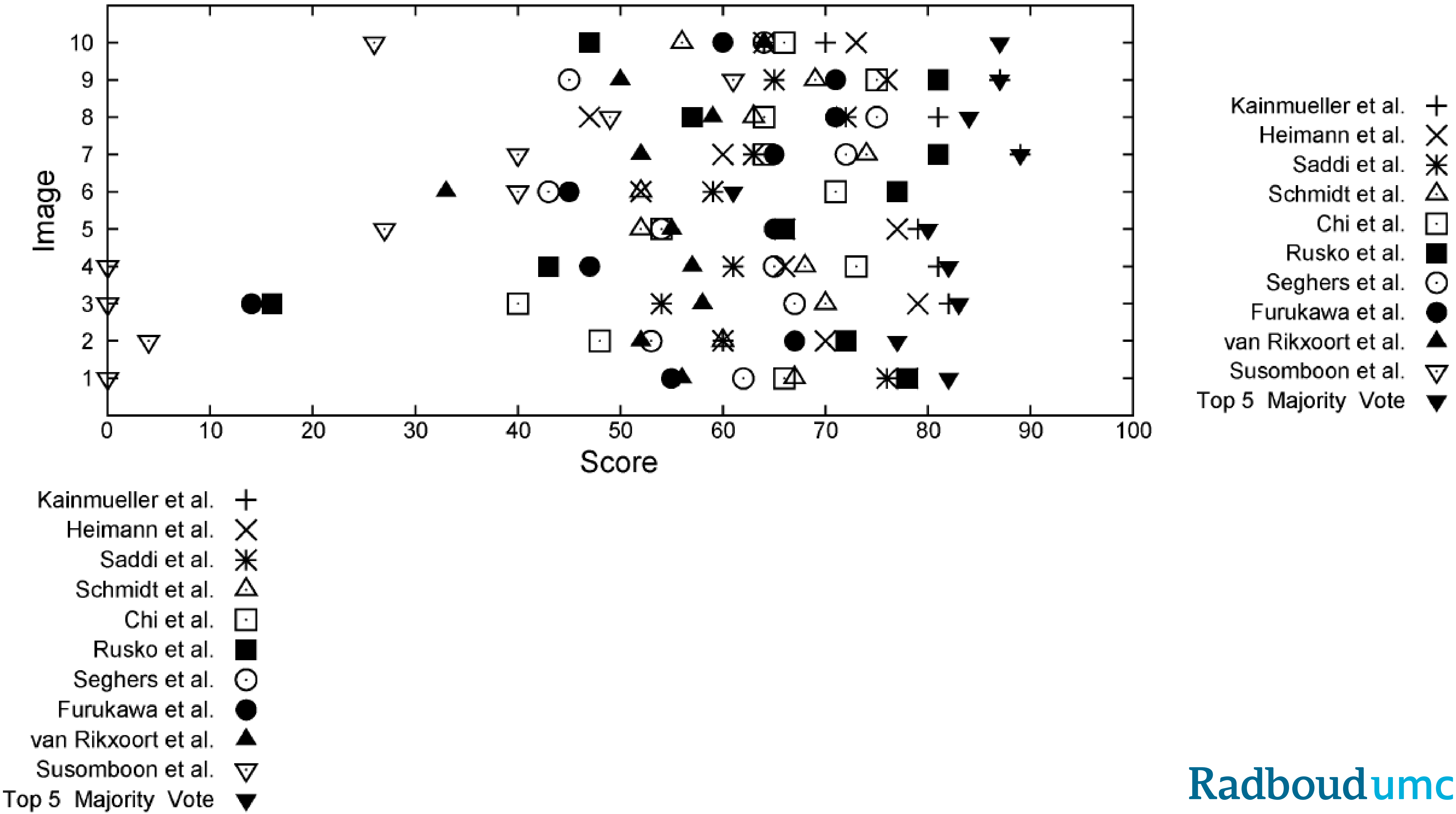


Image
Score
0
10
20
30
40
50
60
70
80
90
100
1
2
3
4
5
6
7
8
9
10
Kainmueller et al.
Heimann et al.
Saddi et al.
Schmidt et al.
Chi et al.
Rusko et al.
Seghers et al.
Furukawa et al.
van Rikxoort et al.
Susomboon et al.
Top 5 Majority Vote
Kainmueller et al.
Heimann et al.
Saddi et al.
Schmidt et al.
Chi et al.
Rusko et al.
Seghers et al.
Furukawa et al.
van Rikxoort et al.
Susomboon et al.
Top 5 Majority Vote

Radboudumc

Bram van Ginneken

FOLLOWING

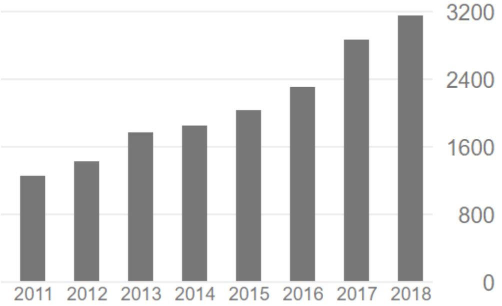
Radboudumc, Fraunhofer MEVIS, Thirona
Verified email at radboudumc.nl - Homepage

Medical Image Analysis · Medical Imaging · Deep Learning · Computer-Aided Diagnosis · Machine Learning

TITLE	CITED BY	YEAR
Ridge-based vessel segmentation in color images of the retina J Staal, MD Abràmoff, M Niemeijer, MA Viergever, B Van Ginneken IEEE transactions on medical imaging 23 (4), 501-509	2060	2004
Reflectance and texture of real-world surfaces KJ Dana, B Van Ginneken, SK Nayar, JJ Koenderink ACM Transactions On Graphics (TOG) 18 (1), 1-34	1597	1999
Comparative study of retinal vessel segmentation methods on a new publicly available database M Niemeijer, J Staal, B van Ginneken, M Loog, MD Abramoff Medical Imaging 2004: Image Processing 5370, 648-657	684	2004
Comparison and evaluation of methods for liver segmentation from CT datasets T Heimann, B Van Ginneken, MA Styner, Y Arzhaeva, V Aurich, C Bauer, ... IEEE transactions on medical imaging 28 (8), 1251-1265	660	2009
A survey on deep learning in medical image analysis G Litjens, T Kooi, BE Bejnordi, AAA Setio, F Ciompi, M Ghafoorian, ... Medical image analysis 42, 60-88	628	2017
Active shape model segmentation with optimal features B Van Ginneken, AF Frangi, JJ Staal, BM ter Haar Romeny, MA Viergever Medical Imaging, IEEE Transactions on 21 (8), 924-933	577	2002
Computer analysis of computed tomography scans of the lung: a survey I Sluimer, A Schilham, M Prokop, B Van Ginneken IEEE transactions on medical imaging 25 (4), 385-405	535	2006
Computer-aided diagnosis in chest radiography: a survey B Van Ginneken, BMTH Romeny, MA Viergever IEEE Transactions on medical imaging 20 (12), 1228-1241	522	2001
Automatic detection of red lesions in digital color fundus photographs M Niemeijer, B Van Ginneken, J Staal, MSA Suttorp-Schulten, ... IEEE Transactions on medical imaging 24 (5), 584-592	476	2005

Cited by · VIEW ALL

	All	Since 2013
Citations	21324	14042
h-index	70	57
i10-index	228	192

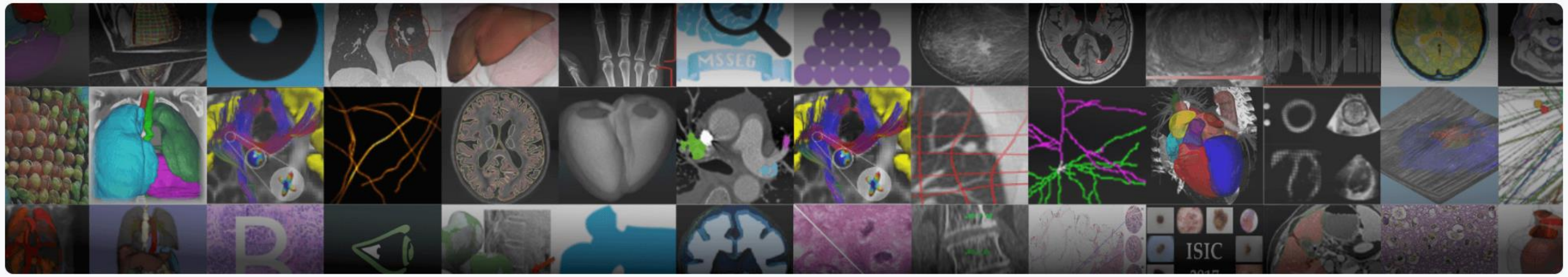


Co-authors · EDIT

- Mathias Prokop — Professor of Radiology, Radboud…
- Max Viergever — Professor of Medical Imaging, U…
- Meindert Niemeijer — CTO, IDx LLC
- Pim de Jong — UMC Utrecht
- Michael D. Abramoff — Professor of Ophthalmology and …
- Eva van Rikxoort — Senior researcher, Diagnostic Im…
- Clara I. Sánchez — Associate Professor, Diagnostic I…
- Ivana Isgum — Associate Professor, Image Scie…

What happened next?

- I urged others to organize challenges. It was fun and useful.
- They asked for our website scripts. They copied bits and pieces.
- There were more and more challenges organized
- I figured we needed an easy way to set up a challenge, and reuse components
- So I hired an enthusiastic programmer to build a general platform
- And we started to keep an overview of challenges on a website: grand-challenge.org



WHY CHALLENGES?

HOST A CHALLENGE

CONTRIBUTORS

ADMIN

Grand Challenges in Biomedical Image Analysis

Every year, thousands of papers are published that describe new algorithms to be applied to medical and biomedical images, and various new products appear on the market based on such algorithms. But few papers and products provide a fair and direct comparison of the newly proposed solution with the state-of-the-art. We believe that such comparisons can help the research community and industry to develop better algorithms. We support the organization of these comparative studies and the dissemination of their results.

Organizing and participating in challenges is not the only way to facilitate better comparisons between new and existing solutions. If it were easy to publish and share your data, and the code you used to evaluate your algorithm's performance on that data, and possibly the algorithm itself, others could directly compare their approach to yours, using the same test data and the same evaluation metrics. With this site we provide tools to make it as easy as possible for you to publish your data and your evaluation for any paper you've written.

Why Challenges? describes the rationale for organizing grand challenges, provides advice for those who want to organize such events, and discusses where we hope the field will move to next.

All Challenges provides an overview of all previous, ongoing and upcoming challenges in biomedical image analysis that we are aware of. Drop us a note if you want your event listed on this overview.

Create your own challenge explains how you can set up your own challenge site in a matter of minutes.

Grand challenges in medical image analysis would not be possible without the support and enthusiasm from all contributors.

All Challenges

https://grand-challenge.org/challenges/

Grand-Challenges ALL CHALLENGES

BramVanGinneken

Filters

> Modality

> Task type

> Structure

Displaying 173 of 173

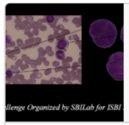
All Challenges

Here is an overview of all challenges that have been organized within the area of medical image analysis that we are aware of. If you know any study that would fit in this overview, or want to advertise your challenge, please send an email to support@grand-challenge.org and we will add the challenge to the list on this page.

+ Add an external challenge | Edit external challenges

Active filters: 0

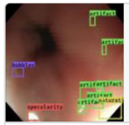
2019



C-NMC 2019

An effort to build an automated classifier for classification of normal vs malignant cells in B-ALL white blood cancer microscopic images that will overcome the problems associated with deploying sophisticated high-end machines with recurring reagent ...

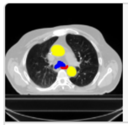
Patricipants: 4
Workshop: April 8, 2019
Associated with: ISBI 2019
Hosted on: grand-challenge.org



EAD2019

Endoscopic Artefact Detection (EAD) is a core problem and needed for realising robust computer-assisted tools. The EAD challenge has 3 tasks: 1) Multi-class artefact detection, 2) Region segmentation, 3) Detection generalisation.

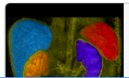
Patricipants: 23
Workshop: April 8, 2019
Associated with: ISBI2019
Hosted on: grand-challenge.org



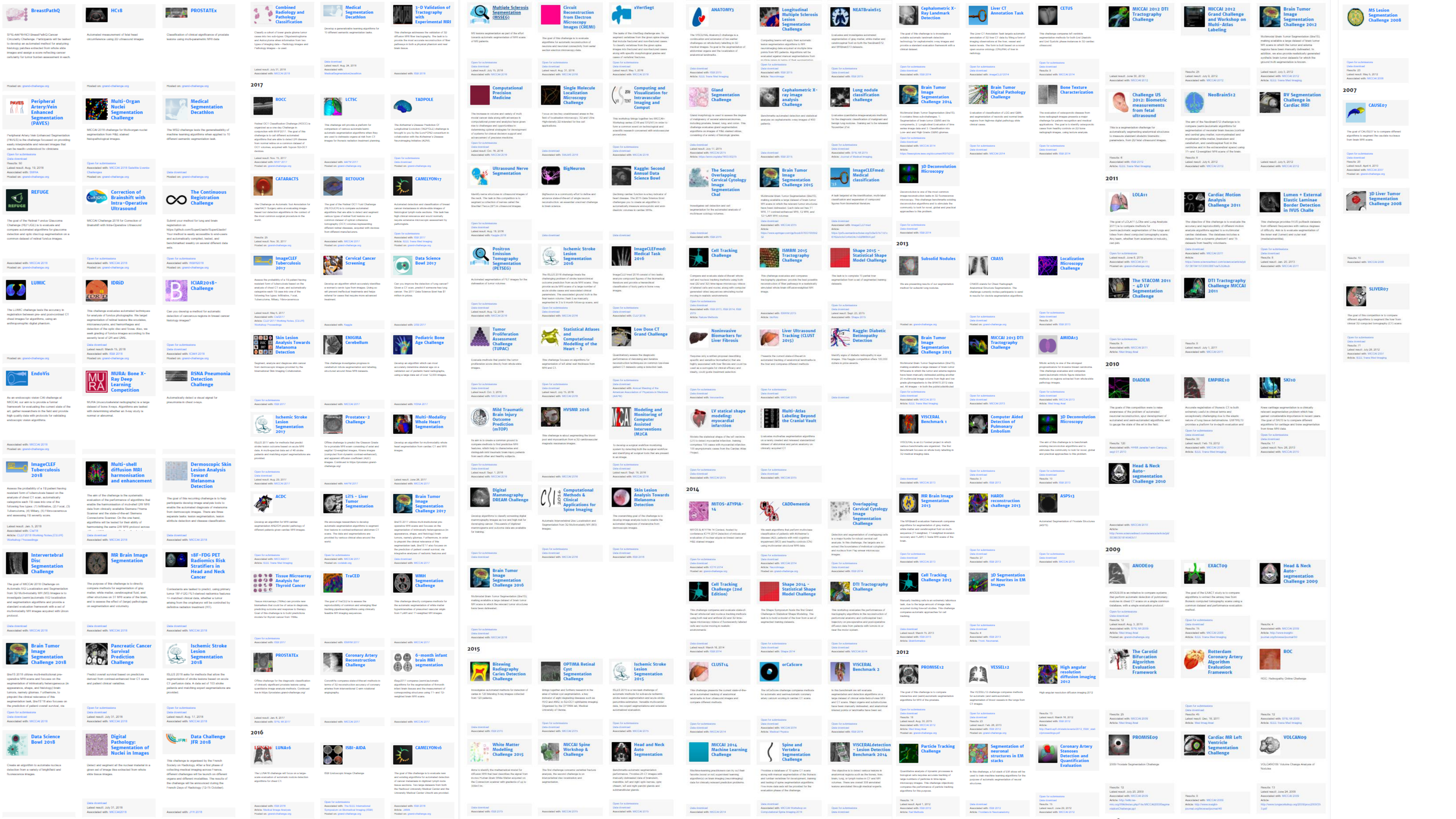
SegTHOR

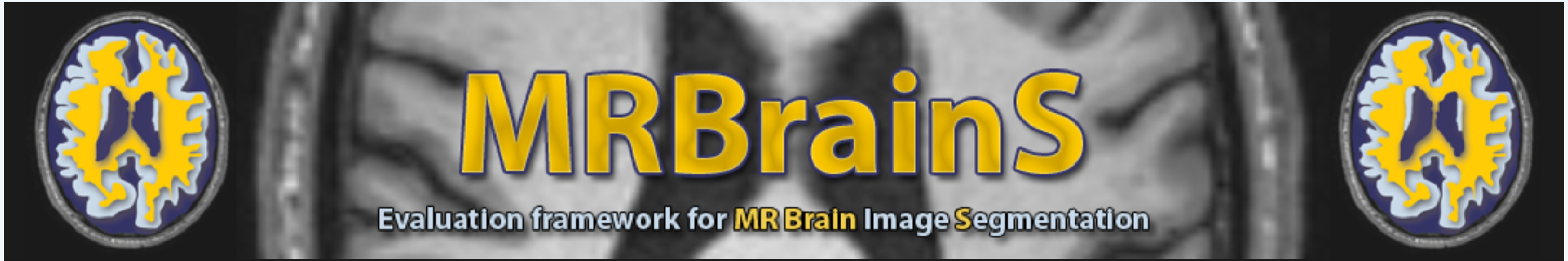
Segmentation of THoracic Organs at Risk in CT images

Workshop: April 8, 2019
Associated with: ISBI 2019
Hosted on: grand-challenge.org



CHAOS


BreastPathQ
HC18
PROSTATEx
Peripheral Artery:Vein Enhanced Segmentation (PAVES)
Multi-Organ Nuclei Segmentation Challenge
Medical Segmentation Decathlon
REFUGE
Correction of Brainshift with Intra-Operative Ultrasound
The Continuous Registration Challenge
LUMIC
IDRiD
ICIAR2018-Challenge
EndoVis
MURA: Bone X-Ray Deep Learning Competition
RSNA Pneumonia Detection Challenge
ImageCLEF Tuberculosis 2018
Multi-shell diffusion MRI harmonisation and enhancement
Dermoscopic Skin Lesion Analysis Toward Melanoma Detection
Intervertebral Disc Segmentation Challenge
MR Brain Image Segmentation
18F-FDG PET Radiomics Risk Stratifiers in Head and Neck Cancer
Brain Tumor Image Segmentation Challenge 2018
Pancreatic Cancer Survival Prediction Challenge
Ischemic Stroke Lesion Segmentation 2018
Data Science Bowl 2018
Digital Pathology: Segmentation of Nuclei in Images
Data Challenge JFR 2018
Combined Radiology and Pathology Classification
Medical Segmentation Decathlon
3-D Validation of Tractography with Experimental MRI
2017
ROCC
LCTSC
TADPOLE
CATARACTS
RETOUCH
CAMELYON17
ImageCLEF Tuberculosis 2017
Cervical Cancer Screening
Data Science Bowl 2017
Skin Lesion Analysis Towards Melanoma Detection
ENIGMA Cerebellum
Pediatric Bone Age Challenge
Ischemic Stroke Lesion Segmentation 2017
Prostatex-2 Challenge
Multi-Modality Whole Heart Segmentation
ACDC
LiTS - Liver Tumor Segmentation
Brain Tumor Image Segmentation Challenge 2017
Tissue Microarray Analysis for Thyroid Cancer
TraCED
WMH Segmentation Challenge
PROSTATEx
Coronary Artery Reconstruction Challenge
6-month infant brain MRI segmentation
2016
LUNA16
ISBI-AIDA
CAMELYON16
Multiple Sclerosis Segmentation (MSSEG)
Circuit Reconstruction from Electron Microscopy Images (CREMI)
xVertSeg
Computational Precision Medicine
Single Molecule Localization Microscopy Challenge
Computing and Visualization for Intravascular Imaging and Comput
Ultrasound Nerve Segmentation
BigNeuron
Kaggle: Second Annual Data Science Bowl
Positron Emission Tomography Segmentation (PETSEG)
Ischemic Stroke Lesion Segmentation 2016
ImageCLEFmed: Medical Task 2016
Tumor Proliferation Assessment Challenge (TUPAC)
Statistical Atlases and Computational Modelling of the Heart - S
Low Dose CT Grand Challenge
Mild Traumatic Brain Injury Outcome Prediction (mTOP)
HVSMR 2016
Modeling and Monitoring of Computer Assisted Interventions (M2CA)
Digital Mammography DREAM Challenge
Computational Methods & Clinical Applications for Spine Imaging
Skin Lesion Analysis Towards Melanoma Detection
Brain Tumor Image Segmentation Challenge 2016
2015
Bitewing Radiography Caries Detection Challenge
OPTIMA Retinal Cyst Segmentation Challenge
Ischemic Stroke Lesion Segmentation 2015
White Matter Modeling Challenge 2015
MICCAI Spine Workshop & Challenge
Head and Neck Auto Segmentation
ANATOMY3
Longitudinal Multiple Sclerosis Lesion Segmentation Challenge
NEATBrainS15
Gland Segmentation Challenge
Cephalometric X-ray image analysis challenge
Lung nodule classification challenge
The Second Overlapping Cervical Cytology Image Segmentation Challenge
Brain Tumor Image Segmentation Challenge 2015
ImageCLEFmed: Medical classification
Cell Tracking Challenge
ISMRM 2015 Tractography Challenge
Shape 2015 - Statistical Shape Model Challenge
Noninvasive Biomarkers for Liver Fibrosis
Liver Ultrasound Tracking (CLUST 2015)
Kaggle: Diabetic Retinopathy Detection
LV statical shape modeling: myocardial infarction
Multi-Atlas Labeling Beyond the Cranial Vault
2014
MITOS-ATYPIA-14
CADDementia
Overlapping Cervical Cytology Image Segmentation Challenge
Cell Tracking Challenge (2nd Edition)
Shape 2014 - Statistical Shape Model Challenge
DTI Tractography Challenge
CLUST14
orCaScore
VISCERAL Benchmark 2
MICCAI 2014 Machine Learning Challenge
Spine and Vertebra Segmentation Challenge
VISCERALdetection - Lesion Detection Benchmark 2014
Cephalometric X-Ray Landmark Detection
Liver CT Annotation Task
CETUS
Brain Tumor Segmentation Challenge 2014
Brain Tumor Digital Pathology Challenge
Bone Texture Characterization
3D Deconvolution Microscopy
2013
Subsolid Nodules
CRASS
Localization Microscopy Challenge
Brain Tumor Image Segmentation Challenge 2013
MICCAI 2013 DTI Tractography Challenge
AMIDA13
VISCERAL Benchmark 1
Computer Aided Detection of Pulmonary Embolism
3D Deconvolution Microscopy
MR Brain Image Segmentation 2013
HARDI reconstruction challenge 2013
ASPS13
Cell Tracking Challenge 2013
3D Segmentation of Neurites in EM Images
2012
PROMISE12
VESSEL12
High angular resolution diffusion imaging 2012
Particle Tracking Challenge
Segmentation of neuronal structures in EM stacks
Coronary Artery Stenoses Detection and Quantification Evaluation
MICCAI 2012 DTI Tractography Challenge
MICCAI 2012 Grand Challenge and Workshop on Multi-Atlas Labeling
Brain Tumor Image Segmentation Challenge 2012
Challenge US 2012: Biometric measurements from fetal ultrasound
NeoBrainS12
RV Segmentation Challenge in Cardiac MRI
2011
LOLA11
Cardiac Motion Analysis Challenge 2011
Lumen + External Elastic Laminae Border Detection in IVUS Challe
The STACOM 2011 - 4D LV Segmentation Challenge
DTI Tractography Challenge MICCAI 2011
2010
DIADEM
EMPIRE10
SKI10
Head & Neck Auto-segmentation Challenge 2010
2009
ANODE09
EXACT09
Head & Neck Auto-segmentation Challenge 2009
The Carotid Bifurcation Algorithm Evaluation Framework
Rotterdam Coronary Artery Algorithm Evaluation Framework
ROC
PROMISE09
Cardiac MR Left Ventricle Segmentation Challenge
VOLCANO09
MS Lesion Segmentation Challenge 2008
2007
CAUSE07
3D Liver Tumor Segmentation Challenge 2008
SLIVER07



Welcome to the MRBrainS website. On September 26th, 2013 we organized the **Grand Challenge on MR Brain Image Segmentation workshop** at the **MICCAI** in Nagoya, Japan, where we launched this evaluation framework. The aim of the MRBrainS evaluation framework is to compare (semi-)automatic algorithms for segmentation of grey matter, white matter and cerebrospinal fluid on multi-sequence (T1-weighted, T1-weighted inversion recovery and FLAIR) 3 Tesla MRI scans of the brain. More information about the workshop and the results of the workshop challenge can be found **here**.

Evaluation framework: The Challenge Continues

The MRBrainS website will remain open as an online evaluation framework for new segmentation algorithms or new versions of already applied algorithms. So feel free to register, download the training and test data and submit your results. Any team, whether from academia or industry, can join! You are welcome to submit your new unpublished method, but we would also like to encourage teams with well established published methods to register and submit their results. In this way, the MRBrainS challenge will present an overview of the performance of a wide variety of available brain structure segmentation methods.

How does it work?

The organisation of this challenge is similar to that of previous challenges described on **Grand Challenges in Medical Image Analysis**. Five datasets (MRI scans with manual segmentations) are provided for training and fifteen datasets (only MRI scans) are provided for testing. Teams can **register** on this website, sign the confidentiality agreement, and **download** the data. They can train their segmentation algorithm on the training data, apply it to the test data and **submit** the obtained results as well as a description of their algorithm. Each submission will be evaluated against the reference standard (manual segmentations) and the evaluation results will be sent to the team contact person by e-mail and published in the **results** section of this website.

More information can be found on the **details** page. If you have additional questions, feel free to sent an email to **mrbrains13@isi.uu.nl**.

Rank	Team name	Submission name	Date	Sum Scores	Sequences used	Speed	Doc
1	TailHot	Multi-modality aggregation network[3]	13-04-18	33	T1; T1_IR; FLAIR	~13 sec	
2	LIVIA_ETS	HyperDenseNet[2]	06-02-18	59	T1; T1_IR; FLAIR	~4 min	
3	CU_DL2	3D Deep Learning; voxnet2	28-06-16	69	T1; T1_IR; FLAIR	~2 min	
4	CU_DL	3D Deep Learning; voxnet1[3]	16-06-16	71	T1; T1_IR; FLAIR	~2 min	
5	MSL-SKKU	Deep Convolutional Neural Network	19-06-17	78	T1; T1-IR; FLAIR	~1.5 min	
6	LRDE	Fully Convolutional Network	20-12-16	78	T1	~2 sec	
7	MDGRU	Multi-Dimensional Gated Recurrent Units[3]	27-07-16	100	T1; T1_IR; FLAIR	~2 min	
8	FBI/LMB Freiburg	U-Net (3D)	01-05-16	105	T1-1mm; T1-IR; FLAIR	~2 min	
9	PyraMiD-LSTM2	NOCC with rounds[3]	23-05-16	109	T1; T1-IR; FLAIR	~2 min	
10	AOC	Atlas of Classifiers	24-12-17	117	T1	~6 sec	
11	IDSIA	PyraMiD-LSTM	05-06-15	124	T1; T1_IR; FLAIR	~2 min	
12	WTA2	3D Cascade convolutional architecture - Method 2	23-05-18	128	T1; T1_IR; FLAIR	~2 min	
13	STH	Hybrid ANN-based Auto-context method[2]	03-06-16	139	T1; T1-IR; FLAIR	~ 5 min	
14	ISI-Neonatology	Multi-stage voxel classification	31-05-14	143	T1	~1.5 hours	
15	UNC-IDEA	LINKS:Learning-based multi-source integration	09-02-15	151	T1; T1_IR; FLAIR	~3 min	
16	BCH_CRL_IMAGINE	3D patch-wise DenseNet and Patch Fusion[2]	24-05-18	169	T1; T1-IR; FLAIR	~2 min	
17	KSOM GHMF	ASeTs: MAP-Based with Manifold learning	13-05-14	173	T1; T1_IR; FLAIR	~23 min	
18	MNAB2	Random Forests	21-02-14	173	T1; T1_IR; FLAIR	~25 min	
19	WTA	3D Cascade convolutional architecture - Method 1[3]	15-05-18	190	T1; T1_IR; FLAIR	~5 min	
20	vicorob UdG T1_F	MSSEG using T1 + FLAIR (T1-IR skull)	14-01-16	193	T1; IR; FLAIR	~2 min	
21	VBM12	VBM12_r738 with WMHC=2	07-10-15	196	T1	~6 min	
22	BIGR2	Multi-Feature SVM Classification	26-09-13	197	T1; T1_IR; FLAIR	~35 min	
23	vicorob UdG T1	MSSEG using only T1 (T1-IR skull)	21-01-16	212	T1; IR	~2 min	
24	UofL BioImaging	MAP-Based Framework	26-09-13	222	T1	~6 sec	
25	CMIV	Model-guided Level Sets and Skeletons	26-09-13	241	T1; FLAIR	~3 min	
26	UB VPML Med	Multi-Atlas with Multiway Cut	26-09-13	246	T1; T1_IR; FLAIR	~30 min	
27	Narsil	Segmentation using Ensemble Trees	26-09-13	254	T1; FLAIR	< 2 min	
28	VIBOT10 M-T1+F	Convolutional Neural Network	27-02-17	259	T1; FLAIR	~3 min	
29	CVSS	Multi-atlas segmentation based on supervoxels	14-02-17	260	T1; T1_IR; FLAIR	~2 min	
30	bigr_neuro	Automatically Trained kNN Classifier	26-09-13	265	T1; FLAIR	~2 hours	
31	drai	VGG based Fully Convolutional Neural Network	20-01-17	266	T1	~2 sec	

Rank	Team name	Submission name	Date	Sum Scores	Sequences used	Speed	Doc
1	TailHot	Multi-modality aggregation network[3]	13-04-18	33	T1; T1_IR; FLAIR	~13 sec	
2	LIVIA_ETS	HyperDenseNet[2]	06-02-18	59	T1; T1_IR; FLAIR	~4 min	
3	CU_DL2	3D Deep Learning; voxnet2	28-06-16	69	T1; T1_IR; FLAIR	~2 min	
4	CU_DL	3D Deep Learning; voxnet1[3]	16-06-16	71	T1; T1_IR; FLAIR	~2 min	
5	MSL-SKKU	Deep Convolutional Neural Network	19-06-17	78	T1; T1-IR; FLAIR	~1.5 min	
6	LRDE	Fully Convolutional Network	20-12-16	78	T1	~2 sec	
7	MDGRU	Multi-Dimensional Gated Recurrent Units[3]	27-07-16	100	T1; T1_IR; FLAIR	~2 min	
8	FBI/LMB Freiburg	U-Net (3D)	01-05-16	105	T1-1mm; T1-IR; FLAIR	~2 min	
9	PyraMiD-LSTM2	NOCC with rounds[3]	23-05-16	109	T1; T1-IR; FLAIR	~2 min	
10	AOC	Atlas of Classifiers	24-12-17	117	T1	~6 sec	
11	IDSIA	PyraMiD-LSTM	05-06-15	124	T1; T1_IR; FLAIR	~2 min	
12	WTA2	3D Cascade convolutional architecture - Method 2	23-05-18	128	T1; T1_IR; FLAIR	~2 min	
13	STH	Hybrid ANN-based Auto-context method[2]	03-06-16	139	T1; T1-IR; FLAIR	~ 5 min	
14	ISI-Neonatology	Multi-stage voxel classification	31-05-14	143	T1	~1.5 hours	
15	UNC-IDEA	LINKS:Learning-based multi-source integration	09-02-15	151	T1; T1_IR; FLAIR	~3 min	
16	BCH_CRL_IMAGINE	3D patch-wise DenseNet and Patch Fusion[2]	24-05-18	169	T1; T1-IR; FLAIR	~2 min	
17	KSOM GHMF	ASeTs: MAP-Based with Manifold learning	13-05-14	173	T1; T1_IR; FLAIR	~23 min	
18	MNAB2	Random Forests	21-02-14	173	T1; T1_IR; FLAIR	~25 min	
19	WTA	3D Cascade convolutional architecture - Method 1[3]	15-05-18	190	T1; T1_IR; FLAIR	~5 min	
20	vicorob UdG T1_F	MSSEG using T1 + FLAIR (T1-IR skull)	14-01-16	193	T1; IR; FLAIR	~2 min	
21	VBM12	VBM12_r728 with WMHC=2	07-10-15	196	T1	~6 min	
22	BIGR2	Multi-Feature SVM Classification	26-09-13	197	T1; T1_IR; FLAIR	~35 min	
23	vicorob UdG T1	MSSEG using only T1 (T1-IR skull)	21-01-16	212	T1; IR	~2 min	
24	UofL BioImaging	MAP-Based Framework	26-09-13	222	T1	~6 sec	
25	CMIV	Model-guided Level Sets and Skeletons	26-09-13	241	T1; FLAIR	~3 min	
26	UB VPML Med	Multi-Atlas with Multiway Cut	26-09-13	246	T1; T1_IR; FLAIR	~30 min	
27	Narsil	Segmentation using Ensemble Trees	26-09-13	254	T1; FLAIR	< 2 min	
28	VIBOT10 M-T1+F	Convolutional Neural Network	27-02-17	259	T1; FLAIR	~3 min	
29	CVSS	Multi-atlas segmentation based on supervoxels	14-02-17	260	T1; T1_IR; FLAIR	~2 min	
30	bigr_neuro	Automatically Trained kNN Classifier	26-09-13	265	T1; FLAIR	~2 hours	
31	drai	VGG based Fully Convolutional Neural Network	20-01-17	266	T1	~2 sec	

Screenshot made a few months ago. As of today They rank #30

Automated Brain-Tissue Segmentation by Multi-Feature SVM Classification

Annegreet van Opbroek[1], Fedde van der Lijn[1] and Marleen de Bruijne[1,2]

[1] Biomedical Imaging Group Rotterdam, Departments of Medical Informatics and Radiology, Erasmus MC - University Medical Center Rotterdam, the Netherlands
[2] Department of Computer Science, University of Copenhagen, Denmark

Abstract. We present a method for automated brain-tissue segmentation through voxelwise classification. Our algorithm uses manually labeled training images to train a support vector machine (SVM) classifier, which is then used for the segmentation of target images. The classification incorporates voxel intensities from a T1-weighted scan, an IR scan, and a FLAIR scan; features to encode the voxel position in the image; and Gaussian-scale-space features and Gaussian-derivative features at multiple scales to facilitate a smooth segmentation.
An experiment on data from the MRBrainS13 brain-tissue-segmentation challenge showed that our algorithm produces reasonable segmentations in a reasonable amount of time.

Rank	Team name	Submission name	Date	Sum	Sequences used		
					T1, FLAIR	~2 hours	
31	drai	VGG based Fully Convolutional Neural Network	20-01-17	266	T1	~2 sec	

Rank	Team name	Submission name	Date	Sum Scores	Sequences used	Speed	Doc
1	TailHot	Multi-modality aggregation network[3]	13-04-18	33	T1; T1_IR; FLAIR	~13 sec	
2	LIVIA_ETS	HyperDenseNet[2]	06-02-18	59	T1; T1_IR; FLAIR	~4 min	
3	CU_DL2	3D Deep Learning; voxnet2	28-06-16	69	T1; T1_IR; FLAIR	~2 min	
4	CU_DL	3D Deep Learning; voxnet1[3]	16-06-16	71	T1; T1_IR; FLAIR	~2 min	
5	MSL-SKKU	Deep Convolutional Neural Network	19-06-17	78	T1; T1-IR; FLAIR	~1.5 min	
6	LRDE	Fully Convolutional Network	20-12-16	78	T1	~2 sec	
7	MDGRU	Multi-Dimensional Gated Recurrent Units[3]	27-07-16	100	T1; T1_IR; FLAIR	~2 min	
8	FBI/LMB Freiburg	U-Net (3D)	01-05-16	105	T1-1mm; T1-IR; FLAIR	~2 min	
9	PyraMiD-LSTM2	NOCC with rounds[3]	23-05-16	109	T1; T1-IR; FLAIR	~2 min	
10	AOC	Atlas of Classifiers	24-12-17	117	T1	~6 sec	
11	IDSIA	PyraMiD-LSTM	05-06-15	124	T1; T1_IR; FLAIR	~2 min	
12	WTA2	3D Cascade convolutional architecture - Method 2	23-05-18	128	T1; T1_IR; FLAIR	~2 min	
13	STH	Hybrid ANN-based Auto-context method[2]	03-06-16	139	T1; T1-IR; FLAIR	~ 5 min	
14	ISI-Neonatology	Multi-stage voxel classification	31-05-14	143	T1	~1.5 hours	
15	UNC-IDEA	LINKS:Learning-based multi-source integration	09-02-15	151	T1; T1_IR; FLAIR	~3 min	
16	BCH_CRL_IMAGINE	3D patch-wise DenseNet and Patch Fusion[2]	24-05-18	169	T1; T1-IR; FLAIR	~2 min	
17	KSOM GHMF	ASeTs: MAP-Based with Manifold learning	13-05-14	173	T1; T1_IR; FLAIR	~23 min	
18	MNAB2	Random Forests	21-02-14	173	T1; T1_IR; FLAIR	~25 min	
19	WTA	3D Cascade convolutional architecture - Method 1[3]	15-05-18	190	T1; T1_IR; FLAIR	~5 min	
20	vicorob UdG T1_F	MSSEG using T1 + FLAIR (T1-IR skull)	14-01-16	193	T1; IR; FLAIR	~2 min	
21	VBM12	VBM12_r738 with WMHC=2	07-10-15	196	T1	~6 min	
22	BIGR2	Multi-Feature SVM Classification	26-09-13	197	T1; T1_IR; FLAIR	~35 min	
23	vicorob UdG T1	MSSEG using only T1 (T1-IR skull)	21-01-16	212	T1; IR	~2 min	
24	UofL BioImaging	MAP-Based Framework	26-09-13	222	T1	~6 sec	
25	CMIV	Model-guided Level Sets and Skeletons	26-09-13	241	T1; FLAIR	~3 min	
26	UB VPML Med	Multi-Atlas with Multiway Cut	26-09-13	246	T1; T1_IR; FLAIR	~30 min	
27	Narsil	Segmentation using Ensemble Trees	26-09-13	254	T1; FLAIR	< 2 min	
28	VIBOT10 M-T1+F	Convolutional Neural Network	27-02-17	259	T1; FLAIR	~3 min	
29	CVSS	Multi-atlas segmentation based on supervoxels	14-02-17	260	T1; T1_IR; FLAIR	~2 min	
30	bigr_neuro	Automatically Trained kNN Classifier	26-09-13	265	T1; FLAIR	~2 hours	
31	drai	VGG based Fully Convolutional Neural Network	20-01-17	266	T1	~2 sec	

Rank	Team name	Submission name	Date	Sum Scores	Sequences used	Speed	Doc
1	TailHot	Multi-modality aggregation network[3]	13-04-18	33	T1; T1_IR; FLAIR	~13 sec	
2	LIVIA_ETS	HyperDenseNet[2]	06-02-18	59	T1; T1_IR; FLAIR	~4 min	
3	CU_DL2	3D Deep Learning; voxnet2	28-06-16	69	T1; T1_IR; FLAIR	~2 min	
4	CU_DL	3D Deep Learning; voxnet1[3]	16-06-16	71	T1; T1_IR; FLAIR	~2 min	
5	MSL-SKKU	Deep Convolutional Neural Network	19-06-17	78	T1; T1-IR; FLAIR	~1.5 min	
6	LRDE	Fully Convolutional Network	20-12-16	78	T1	~2 sec	
7	MDGRU	Multi-Dimensional Gated Recurrent Units[3]	27-07-16	100	T1; T1_IR; FLAIR	~2 min	
8	FBI/LMB Freiburg	U-Net (3D)	01-05-16	105	T1-1mm; T1-IR; FLAIR	~2 min	
9	PyraMiD-LSTM2	NOCC with rounds[3]	23-05-16	109	T1; T1-IR; FLAIR	~2 min	
10	AOC	Atlas of Classifiers	24-12-17	117	T1	~6 sec	
11	IDSIA	PyraMiD-LSTM	05-06-15	124	T1; T1_IR; FLAIR	~2 min	
12	WTA2	3D Cascade convolutional architecture - Method 2	23-05-18	128	T1; T1_IR; FLAIR	~2 min	
13	STH	Hybrid ANN-based Auto-context method[2]	03-06-16	139	T1; T1-IR; FLAIR	~ 5 min	
14	ISI-Neonatology	Multi-stage voxel classification	31-05-14	143	T1	~1.5 hours	
15	UNC-IDEA	LINKS:Learning-based multi-source integration	09-02-15	151	T1; T1_IR; FLAIR	~3 min	
16	BCH_CRL_IMAGINE	3D patch-wise DenseNet and Patch Fusion[2]	24-05-18	169	T1; T1-IR; FLAIR	~2 min	
17	KSOM GHMF	ASeTs: MAP-Based with Manifold learning	13-05-14	173	T1; T1_IR; FLAIR	~23 min	
18	MNAB2	Random Forests	21-02-14	173	T1; T1_IR; FLAIR	~25 min	
19	WTA	3D Cascade convolutional architecture - Method 1[3]	15-05-18	190	T1; T1_IR; FLAIR	~5 min	
20	vicorob UdG T1_F	MSSEG using T1 + FLAIR (T1-IR skull)	14-01-16	193	T1; IR; FLAIR	~2 min	
21	VBM12	VBM12_r738 with WMHC=2	07-10-15	196	T1	~6 min	
22	BIGR2	Multi-Feature SVM Classification	26-09-13	197	T1; T1_IR; FLAIR	~35 min	
23	vicorob UdG T1	MSSEG using only T1 (T1-IR skull)	21-01-16	212	T1; IR	~2 min	
24	UofL BioImaging	MAP-Based Framework	26-09-13	222	T1	~6 sec	
25	CMIV	Model-guided Level Sets and Skeletons	26-09-13	241	T1; FLAIR	~3 min	
26	UB VPML Med	Multi-Atlas with Multiway Cut	26-09-13	246	T1; T1_IR; FLAIR	~30 min	
27	Narsil	Segmentation using Ensemble Trees	26-09-13	254	T1; FLAIR	< 2 min	
28	VIBOT10 M-T1+F	Convolutional Neural Network	27-02-17	259	T1; FLAIR	~3 min	
29	CVSS	Multi-atlas segmentation based on supervoxels	14-02-17	260	T1; T1_IR; FLAIR	~2 min	
30	bigr_neuro	Automatically Trained kNN Classifier	26-09-13	265	T1; FLAIR	~2 hours	
31	drai	VGG based Fully Convolutional Neural Network	20-01-17	266	T1	~2 sec	

Rank	Team name	Submission name	Date	Sum Scores	Sequences used	Speed	Doc
1	TailHot	Multi-modality aggregation network[3]	13-04-18	33	T1; T1_IR; FLAIR	~13 sec	
2	LIVIA_ETS	HyperDenseNet[2]	06-02-18	59	T1; T1_IR; FLAIR	~4 min	
3	CU_DL2	3D Deep Learning; voxnet2	28-06-16	69	T1; T1_IR; FLAIR	~2 min	
4	CU_DL	3D Deep Learning; voxnet1[3]	16-06-16	71	T1; T1_IR; FLAIR	~2 min	
5	MSL-SKKU	Deep Convolutional Neural Network	19-06-17	78	T1; T1-IR; FLAIR	~1.5 min	
6	LRDE	Fully Convolutional Network	20-12-16	78	T1	~2 sec	
7	MDGRU	Multi-Dimensional Gated Recurrent Units[3]	27-07-16	100	T1; T1_IR; FLAIR	~2 min	
8	FBI/LMB Freiburg	U-Net (3D)	01-05-16	105	T1-1mm; T1-IR; FLAIR	~2 min	
9	PyraMiD-LSTM2	NOCC with rounds[3]	23-05-16	109	T1; T1-IR; FLAIR	~2 min	
10	AOC	Atlas of Classifiers	24-12-17	117	T1	~6 sec	
11	IDSIA	PyraMiD-LSTM	05-06-15	124	T1; T1_IR; FLAIR	~2 min	
12	WTA2	3D Cascade convolutional architecture - Method 2	23-05-18	128	T1; T1_IR; FLAIR	~2 min	
13	STH	Hybrid ANN-based Auto-context method[2]	03-06-16	139	T1; T1-IR; FLAIR	~ 5 min	
14	ISI-Neonatology	Multi-stage voxel classification	31-05-14	143	T1	~1.5 hours	
15	UNC-IDEA	LINKS:Learning-based multi-source integration	09-02-15	151	T1; T1_IR; FLAIR	~3 min	
16	BCH_CRL_IMAGINE	3D patch-wise DenseNet and Patch Fusion[2]	24-05-18	169	T1; T1-IR; FLAIR	~2 min	
17	KSOM GHMF	ASeTs: MAP-Based with Manifold learning	13-05-14	173	T1; T1_IR; FLAIR	~23 min	
18	MNAB2	Random Forests	21-02-14	173	T1; T1_IR; FLAIR	~25 min	
19	WTA	3D Cascade convolutional architecture - Method 1[3]	15-05-18	190	T1; T1_IR; FLAIR	~5 min	
20	vicorob UdG T1_F	MSSEG using T1 + FLAIR (T1-IR skull)	14-01-16	193	T1; IR; FLAIR	~2 min	
21	VBM12	VBM12_r738 with WMHC=2	07-10-15	196	T1	~6 min	
22	BIGR2	Multi-Feature SVM Classification	26-09-13	197	T1; T1_IR; FLAIR	~35 min	
23	vicorob UdG T1	MSSEG using only T1 (T1-IR skull)	21-01-16	212	T1; IR	~2 min	
24	UofL BioImaging	MAP-Based Framework	26-09-13	222	T1	~6 sec	
25	CMIV	Model-guided Level Sets and Skeletons	26-09-13	241	T1; FLAIR	~3 min	
26	UB VPML Med	Multi-Atlas with Multiway Cut	26-09-13	246	T1; T1_IR; FLAIR	~30 min	
27	Narsil	Segmentation using Ensemble Trees	26-09-13	254	T1; FLAIR	< 2 min	
28	VIBOT10 M-T1+F	Convolutional Neural Network	27-02-17	259	T1; FLAIR	~3 min	
29	CVSS	Multi-atlas segmentation based on supervoxels	14-02-17	260	T1; T1_IR; FLAIR	~2 min	
30	bigr_neuro	Automatically Trained kNN Classifier	26-09-13	265	T1; FLAIR	~2 hours	
31	drai	VGG based Fully Convolutional Neural Network	20-01-17	266	T1	~2 sec	

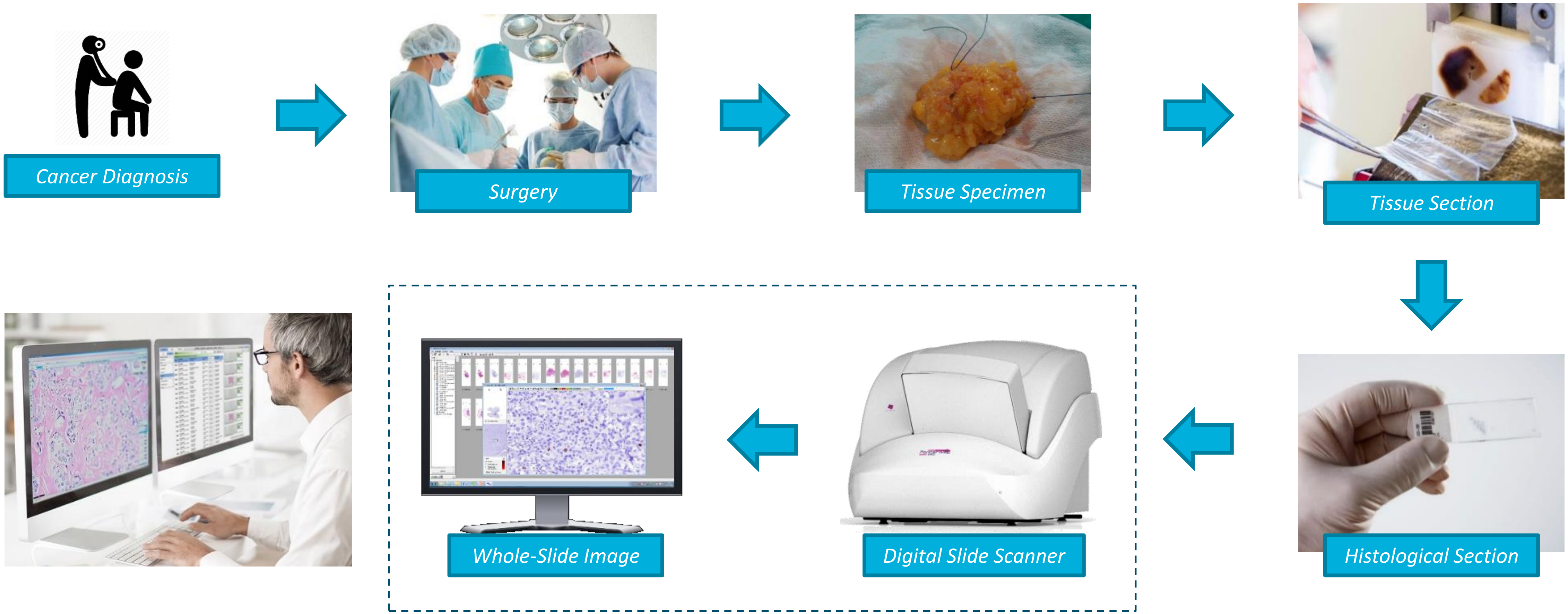
Good challenges run for years

- ...if the dataset is good or exceptional and/or the problem is very important and not solved

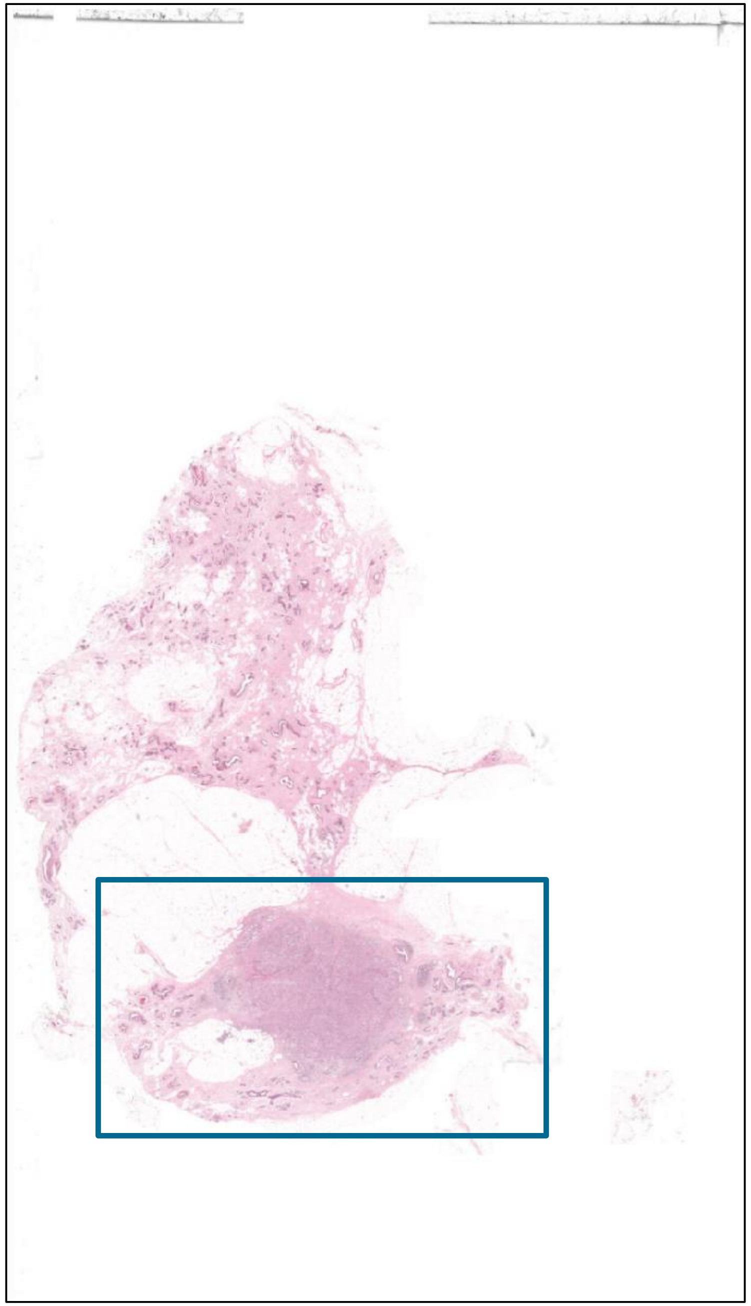
- Maintaining challenges is not easy
- grand-challenge.org provides automated evaluation (via docker and support of a metric library evalutils). More later.

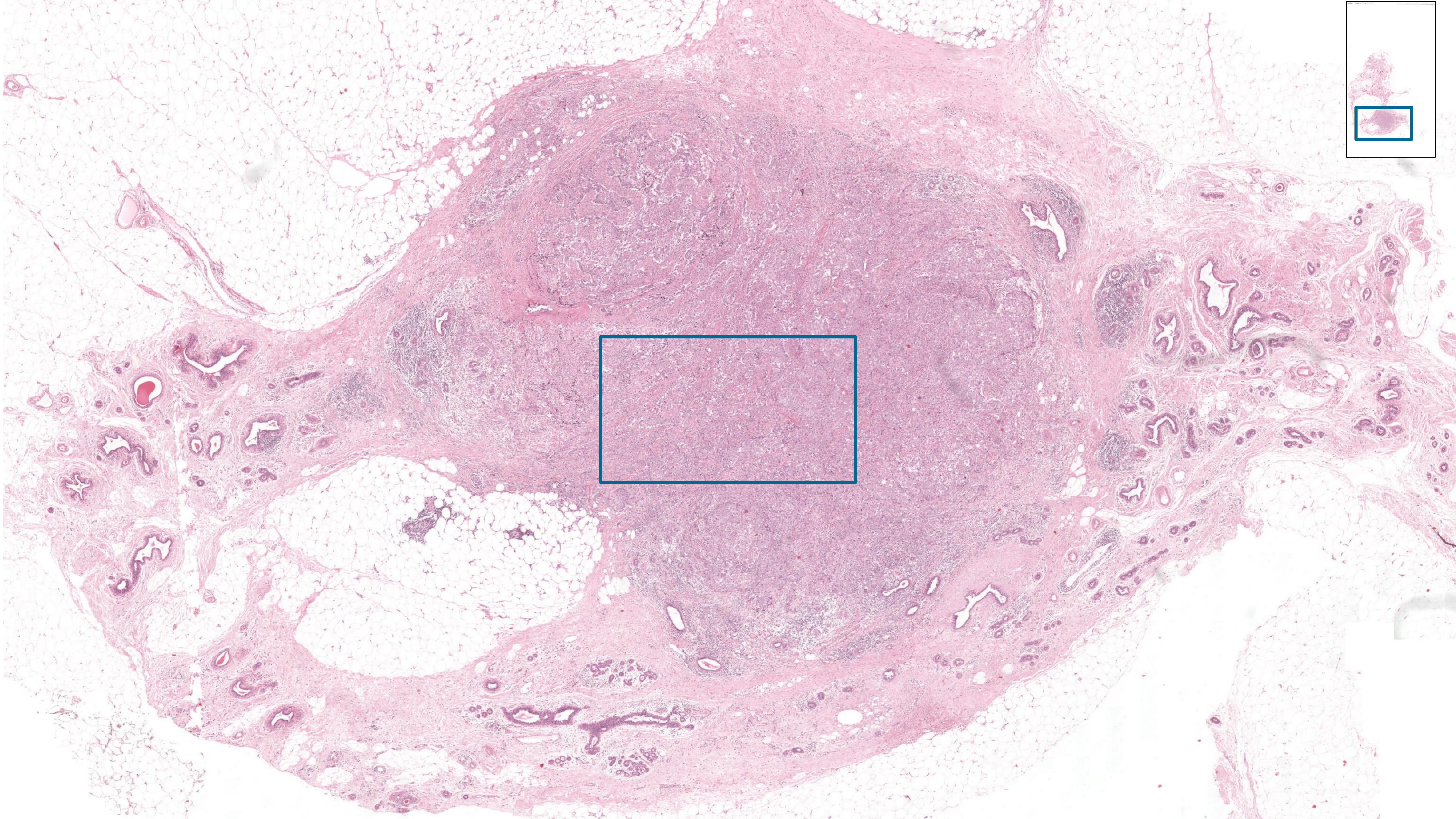
- The leaderboard and overview articles provide a ‘topical review’ of a field
- The leaderboards of a group of challenges show what are popular/powerful techniques in a field

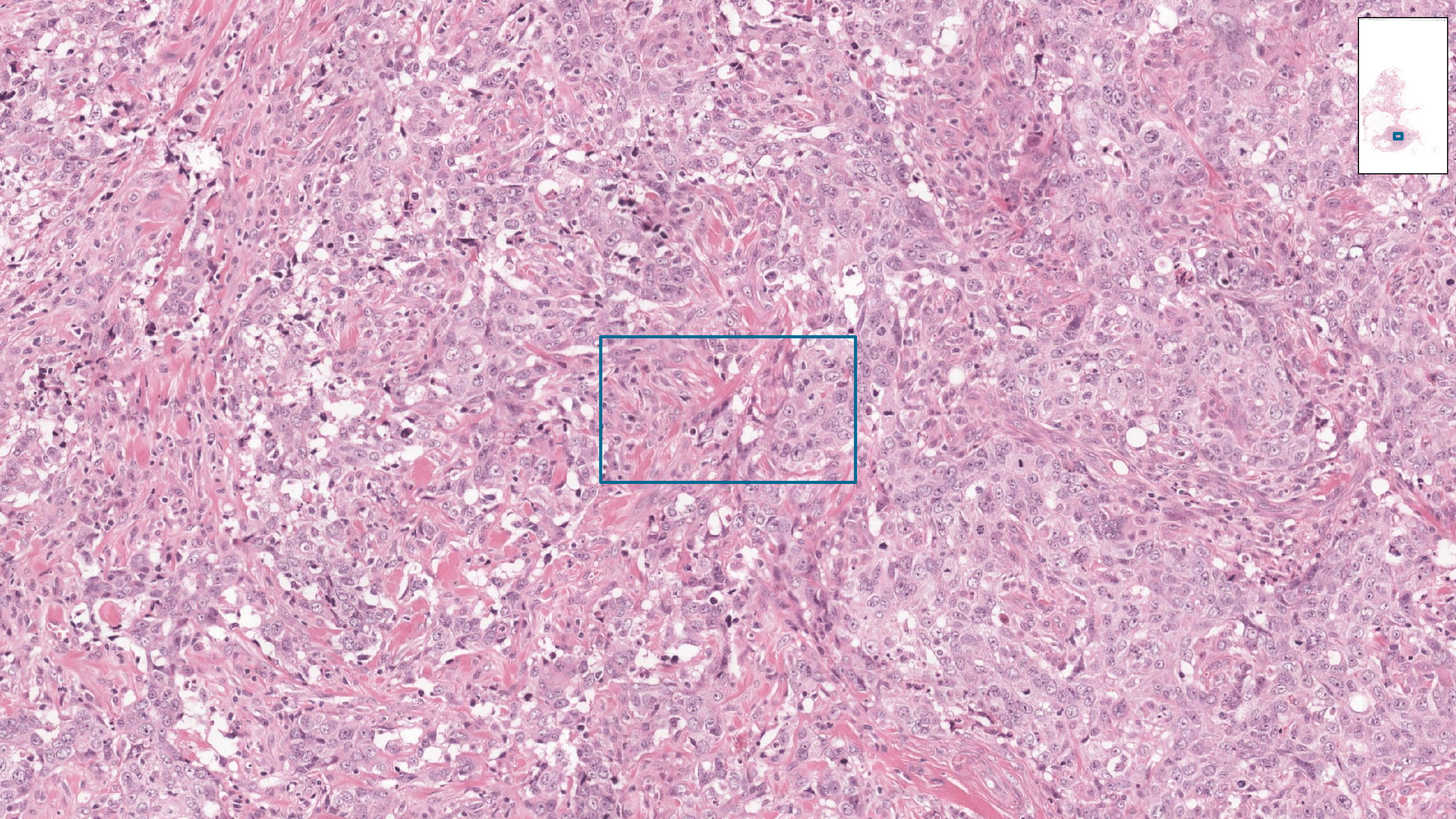
Digital Pathology

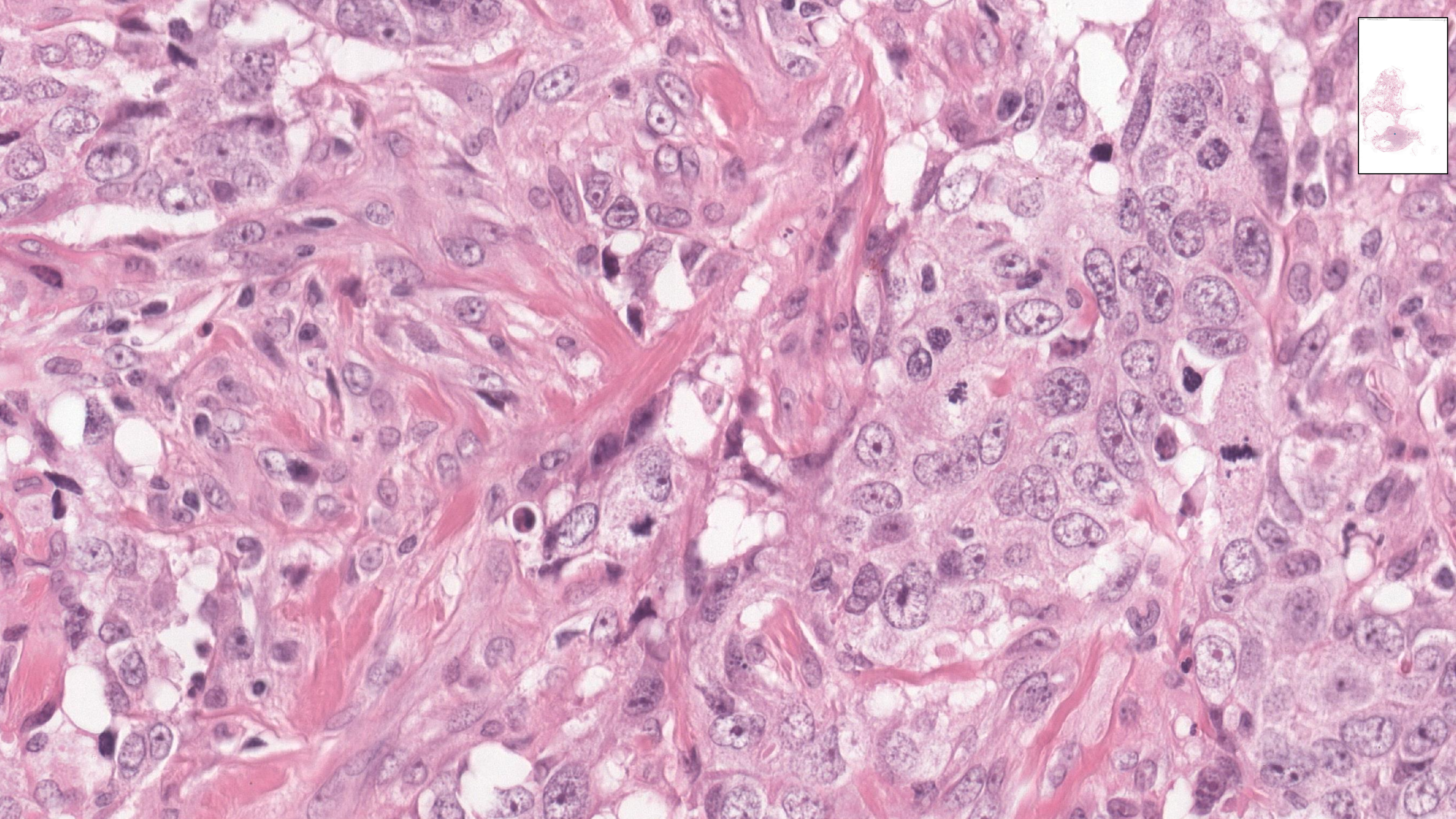


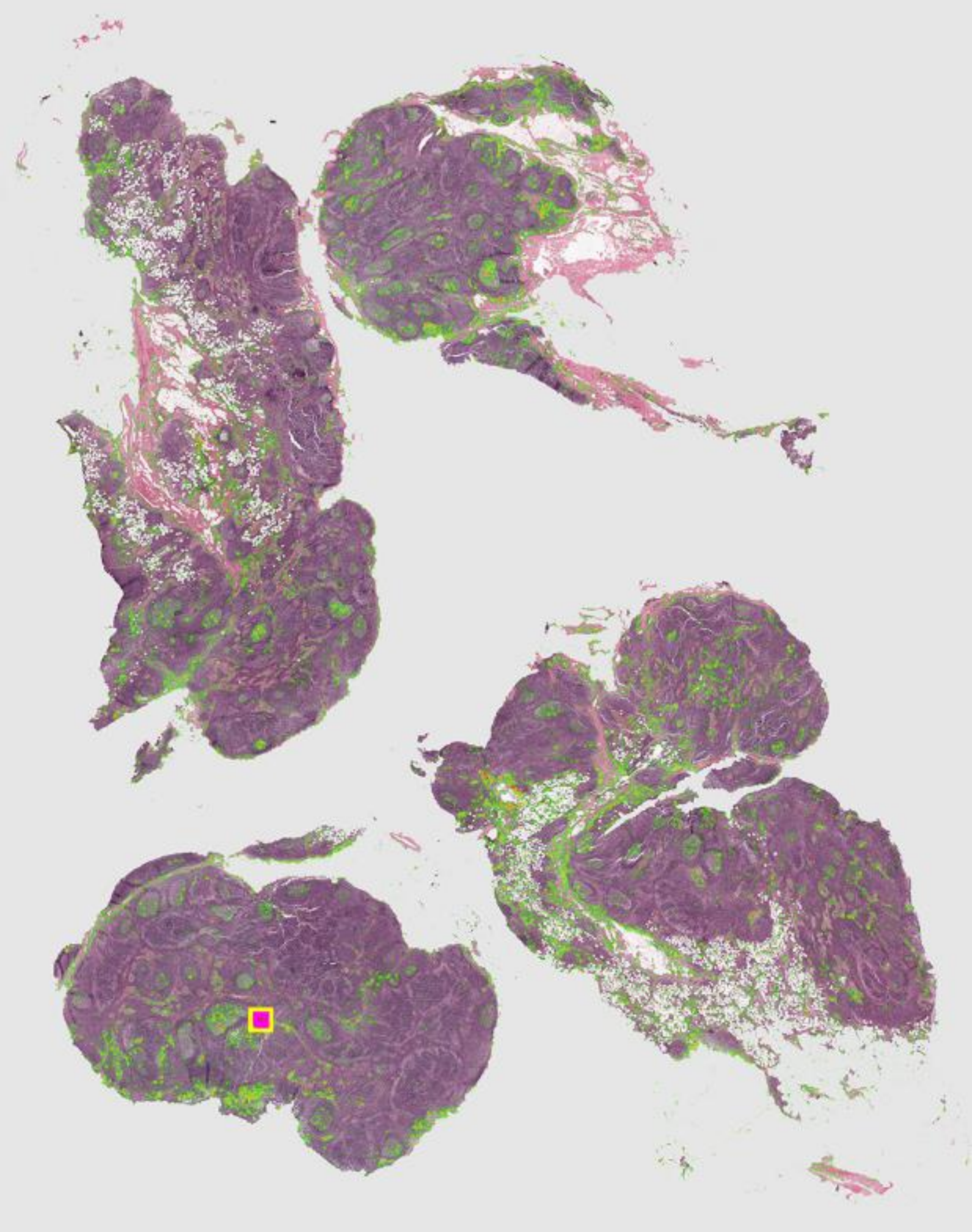
Radboudumc

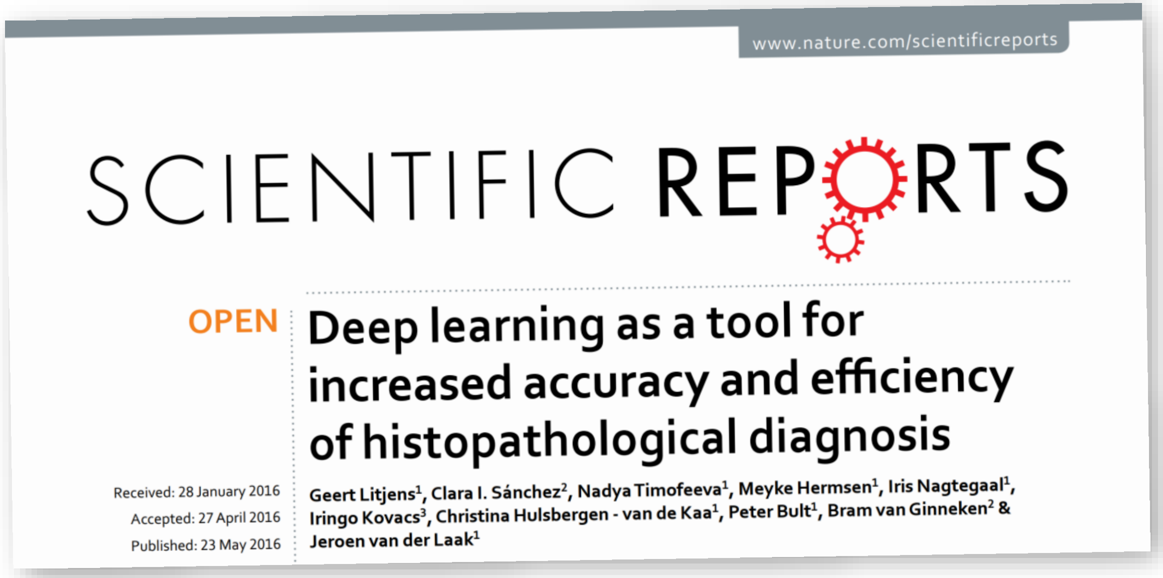












www.nature.com/scientificreports

SCIENTIFIC REPORTS

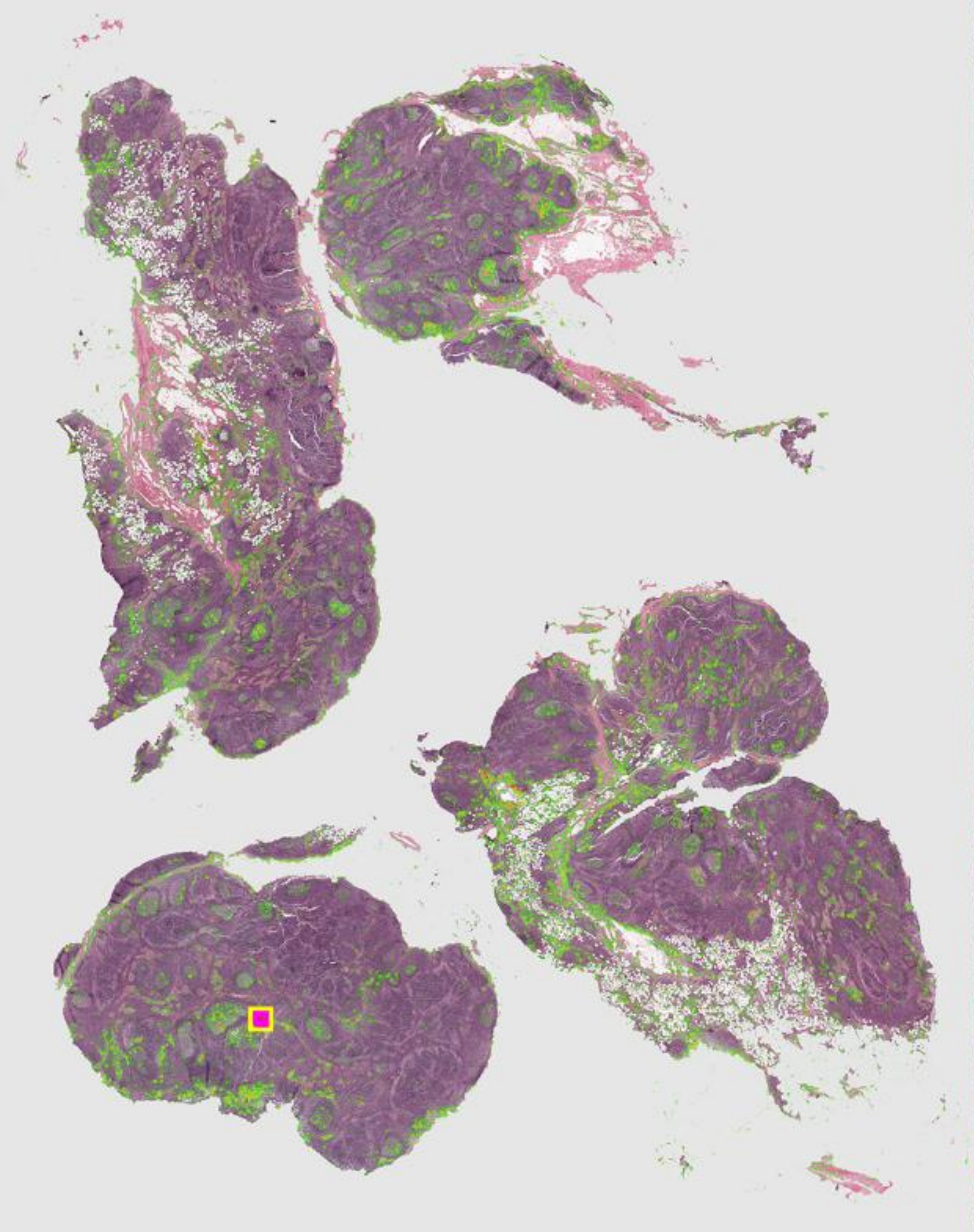
OPEN

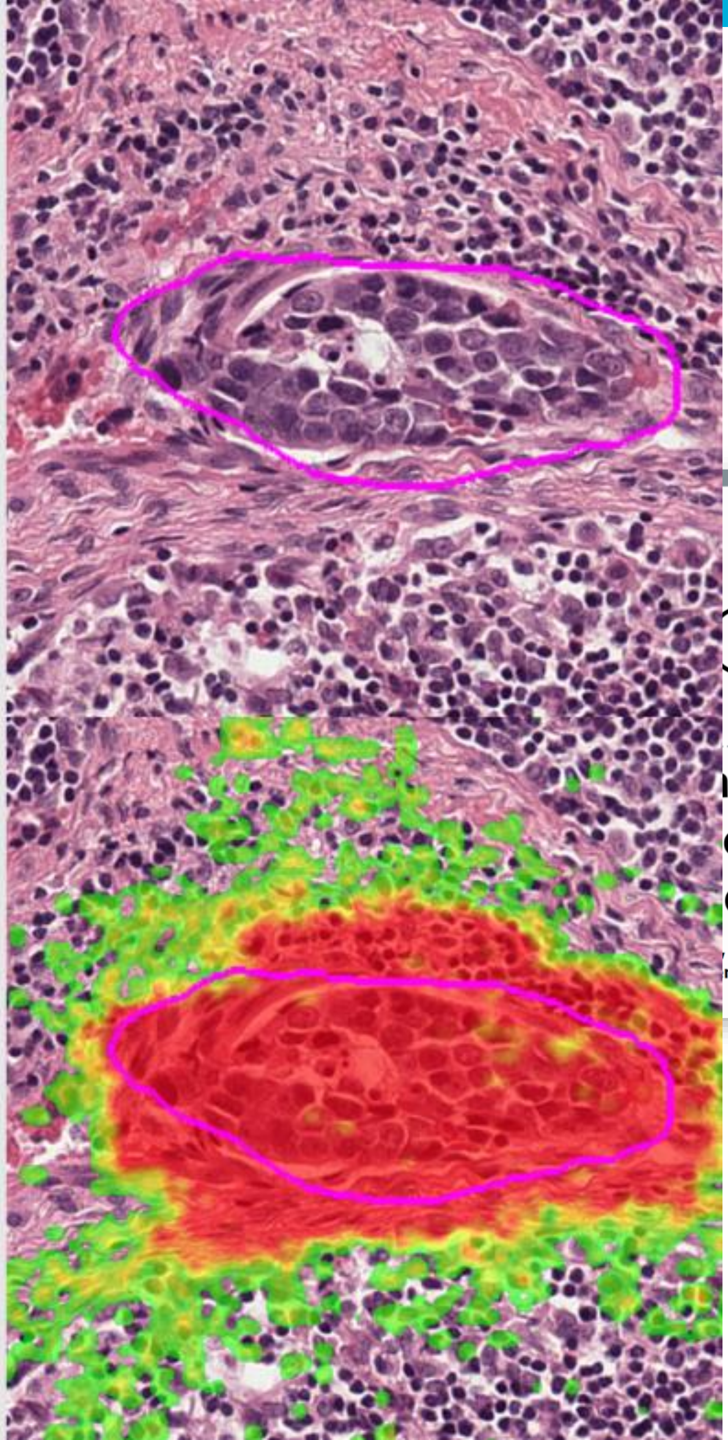
Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis

Received: 28 January 2016
Accepted: 27 April 2016
Published: 23 May 2016

Geert Litjens[1], Clara I. Sánchez[2], Nadya Timofeeva[1], Meyke Hermsen[1], Iris Nagtegaal[1], Iringo Kovacs[3], Christina Hulsbergen - van de Kaa[1], Peter Bult[1], Bram van Ginneken[2] & Jeroen van der Laak[1]

Radboudumc





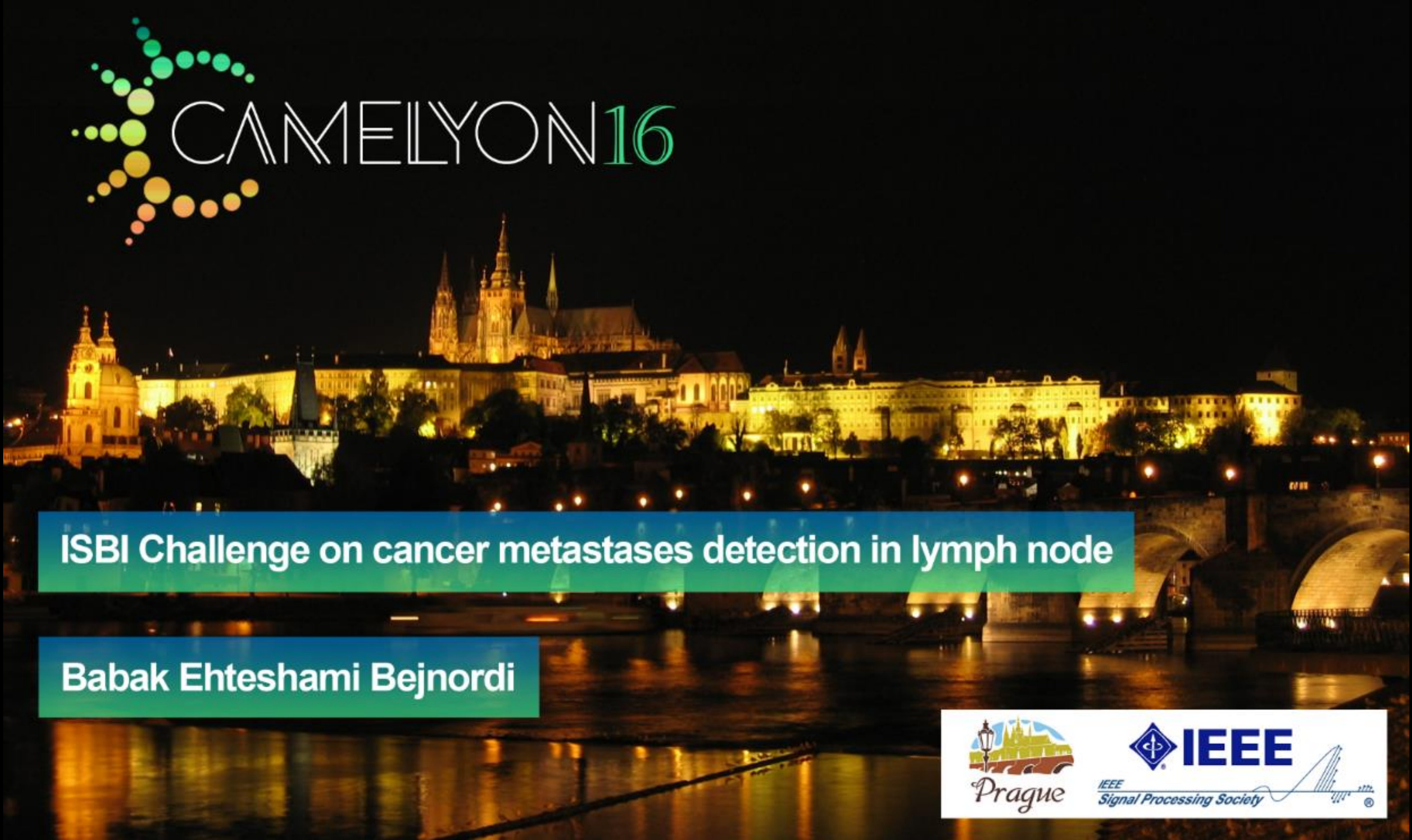
www.nature.com/scientificreports

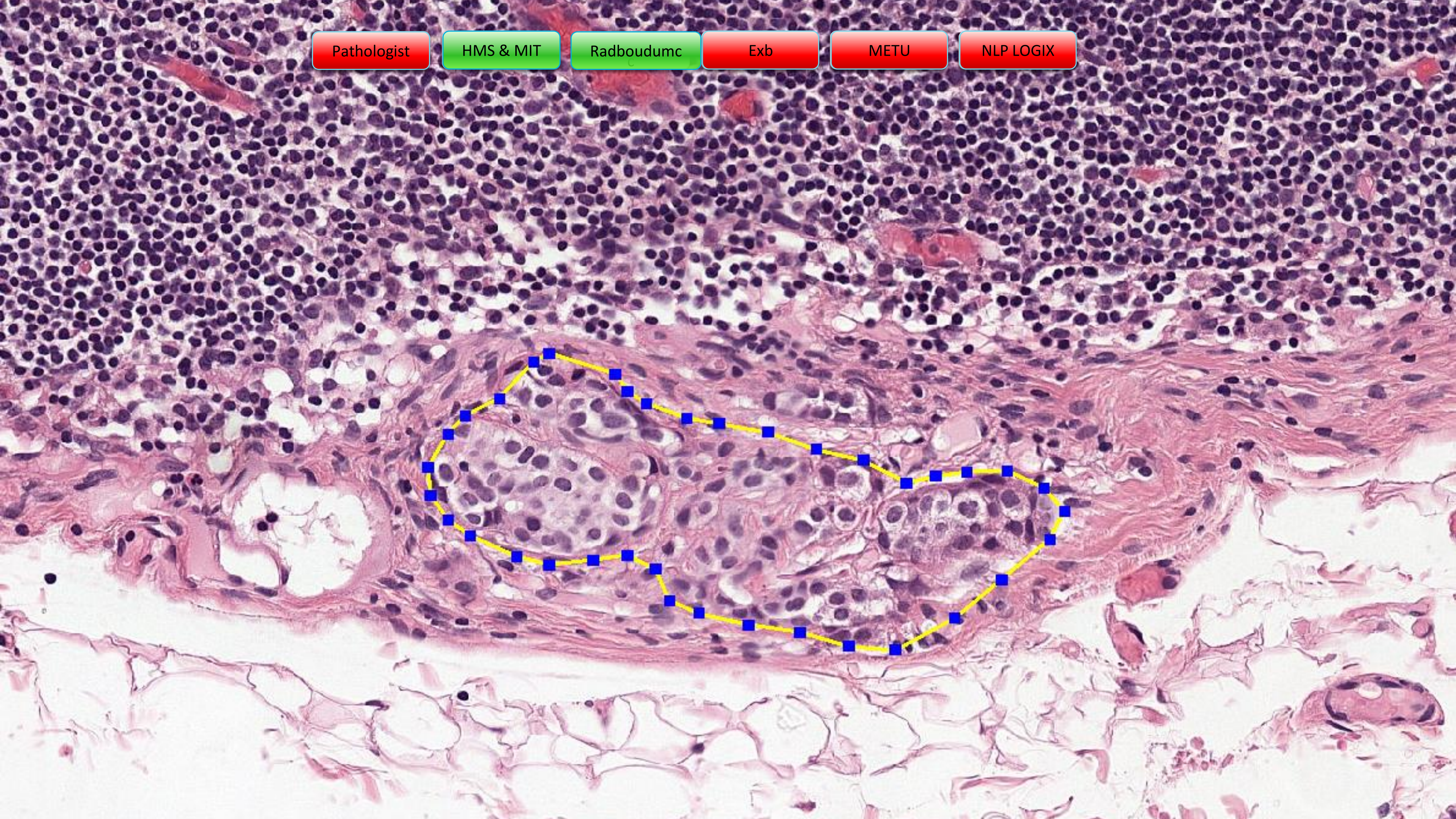
C REPORTS

ng as a tool for
curacy and efficiency
ological diagnosis

, Nadya Timofeeva[1], Meyke Hermsen[1], Iris Nagtegaal[1],
rgen - van de Kaa[1], Peter Bult[1], Bram van Ginneken[2] &

Radboudumc


CAMELYON16
ISBI Challenge on cancer metastases detection in lymph node
Babak Ehteshami Bejnordi
Prague
IEEE
IEEE Signal Processing Society


Pathologist
HMS & MIT
Radboudumc
Exb
METU
NLP LOGIX

THE NATIONAL ARTIFICIAL INTELLIGENCE RESEARCH AND DEVELOPMENT STRATEGIC PLAN

National Science and Technology Council

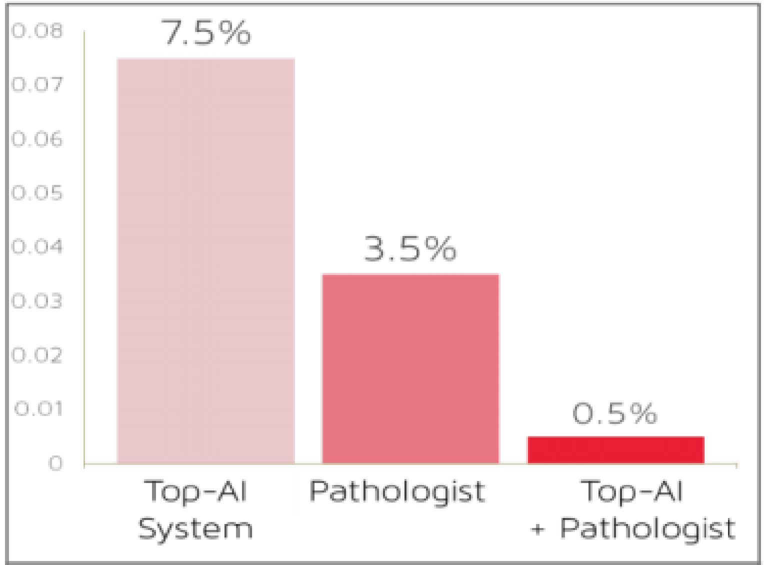
Networking and Information Technology Research and Development Subcommittee

October 2016

National Institutes of Health (NIH) grants-supported research

ARTIFICIAL INTELLIGENCE FOR COMPUTATIONAL PATHOLOGY

Image interpretation plays a central role in the pathologic diagnosis of cancer. Since the late 19^{th} century, the primary tool used by pathologists to make definitive cancer diagnoses is the microscope. Pathologists diagnose cancer by manually examining stained sections of cancer tissues to determine the cancer subtype. Pathologic diagnosis using conventional methods is labor-intensive with poor reproducibility and quality concerns. New approaches use fundamental AI research to build tools to make pathologic analysis more efficient, accurate, and predictive. In the 2016 Camelyon Grand Challenge for metastatic cancer detection,[69] the top-performing entry in the competition was an AI-based computational system that achieved an error rate of 7.5%.[70] A pathologist reviewing the same set of evaluation images achieved an error rate of 3.5%. Combining the predictions of the AI system with the pathologist lowered the error rate to down to 0.5%, representing an 85% reduction in error (see image).[71] This example illustrates how fundamental research in AI can drive the development of high performing computational systems that offer great potential for making pathological diagnoses more efficient and more accurate.



AI significantly reduces pathologist error rate in the identification of metastatic breast cancer from sentinel lymph node biopsies.

JAMA | Original Investigation

Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer

Babak Ehteshami Bejnordi, MS; Mitko Veta, PhD; Paul Johannes van Diest, MD, PhD; Bram van Ginneken, PhD; Nico Karssemeijer, PhD; Geert Litjens, PhD; Jeroen A. W. M. van der Laak, PhD; and the CAMELYON16 Consortium

IMPORTANCE Application of deep learning algorithms to whole-slide pathology images can potentially improve diagnostic accuracy and efficiency.

OBJECTIVE Assess the performance of automated deep learning algorithms at detecting metastases in hematoxylin and eosin–stained tissue sections of lymph nodes of women with breast cancer and compare it with pathologists' diagnoses in a diagnostic setting.

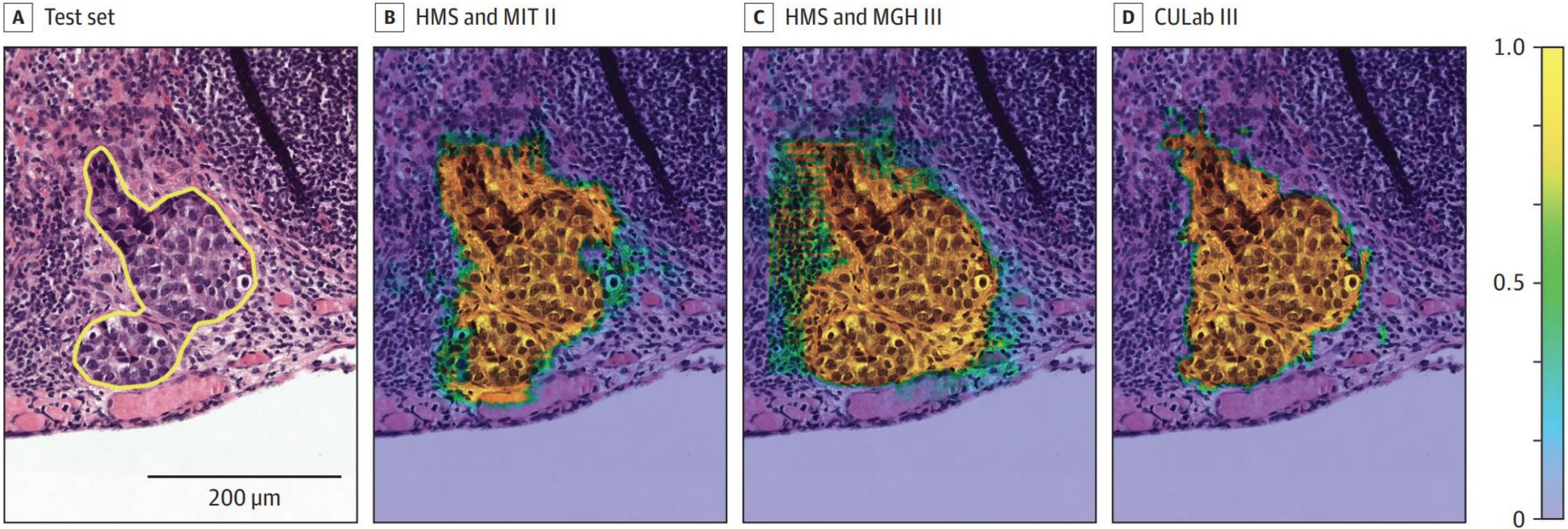
Editorial page 2184

Related articles page 2211 and page 2250

Supplemental content

CME Quiz at jamanetwork.com/learning and **CME Questions** page 2252

CONCLUSIONS AND RELEVANCE In the setting of a challenge competition, some deep learning algorithms achieved better diagnostic performance than a panel of 11 pathologists participating in a simulation exercise designed to mimic routine pathology workflow; algorithm performance was comparable with an expert pathologist interpreting whole-slide images without time constraints. Whether this approach has clinical utility will require evaluation in a clinical setting.



Advantages of challenges

- **Compare** algorithms from the same task in a **fair and consistent** way (same test data, same metric(s), same training data (if desired))
- Forces you to **think carefully** about aspects that relate to algorithm performance
- Leaderboard provides an **overview of techniques** that work well (are popular) for a task
- You can **combine** algorithms to see if they are complementary (even if you do not have access)
- Write an **overview article** to give a topical review enriched with real data
- Link the challenge to a **workshop** at a conference
- It is a way to **collaboratively solve an open research problem** and quickly move the field forward
- Challenges like ImageNet and Camelyon16 can have huge **impact**

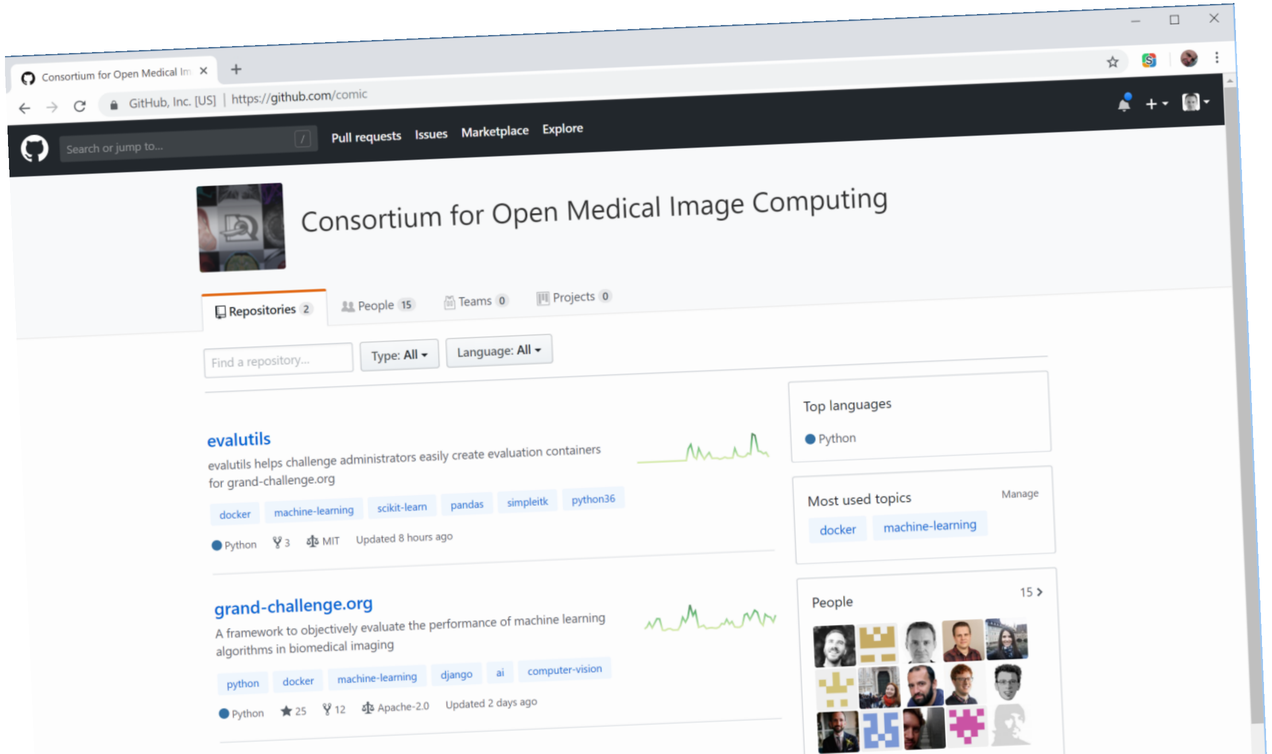
Every paper a challenge?

- Why not create a challenge around every algorithm paper you write?
- Allows new techniques to be tested on your data
- Will generate citations to your paper
- You're pretty much doing everything already
  - Posing a question
  - Gathering data and annotations
  - Splitting this into train/valid/test data
  - Deciding on and implementing evaluation metrics
- Drawback: It's a lot of work
  - Creating a site
  - Participant management
  - Data sharing
  - Metrics definition
  - Submission gathering and continuous evaluation
  - Visualizing and comparing results

Radboudumc

Simplifying the process with reusable code

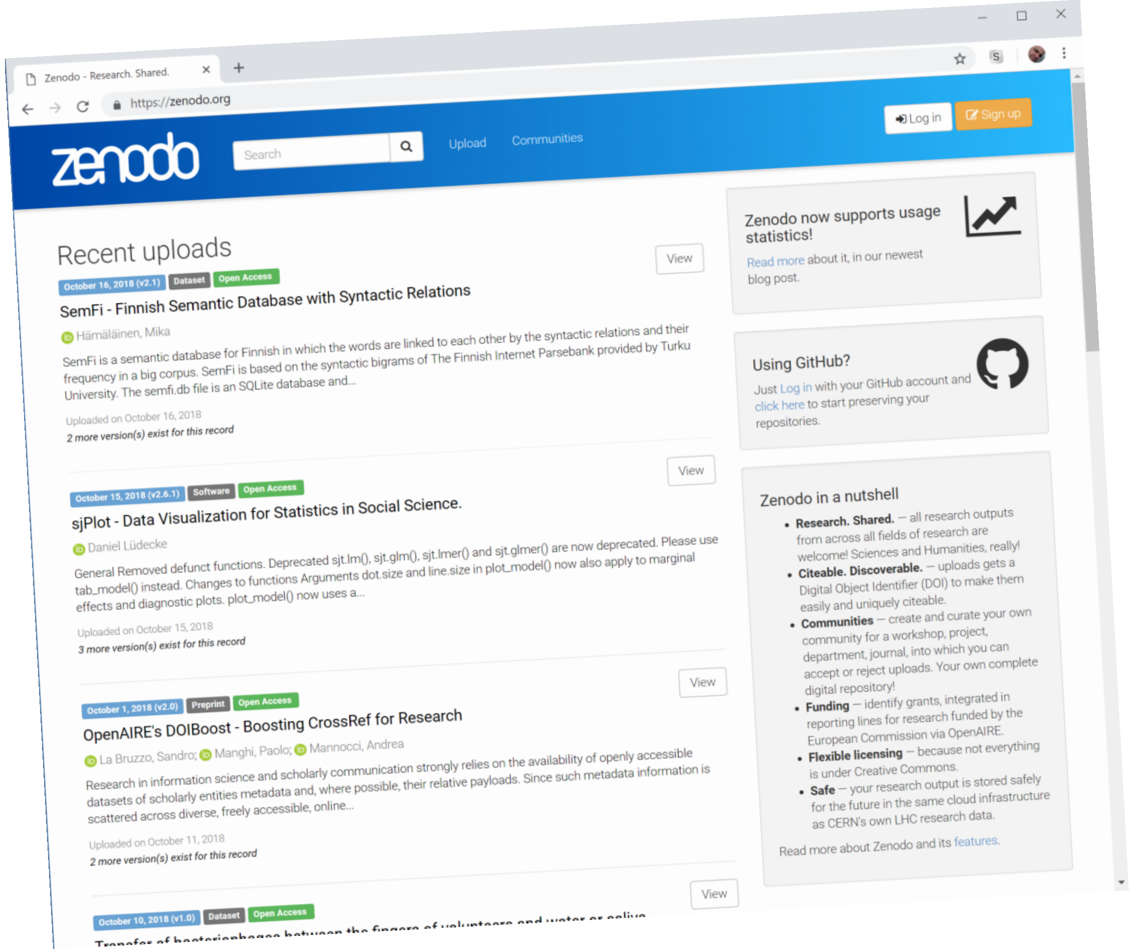
- Groups in Nijmegen, Bremen, Rotterdam, Leuven, London set up the Consortium for Open Medical Image Computing (**COMIC**) in 2012
- Goal: make it easy to set up challenge websites with reusable components
- No funding, slow progress
- Became a priority when we saw it could be our internal platform in most research projects (and became a success when James Meakin started to work on it)



Radboudumc

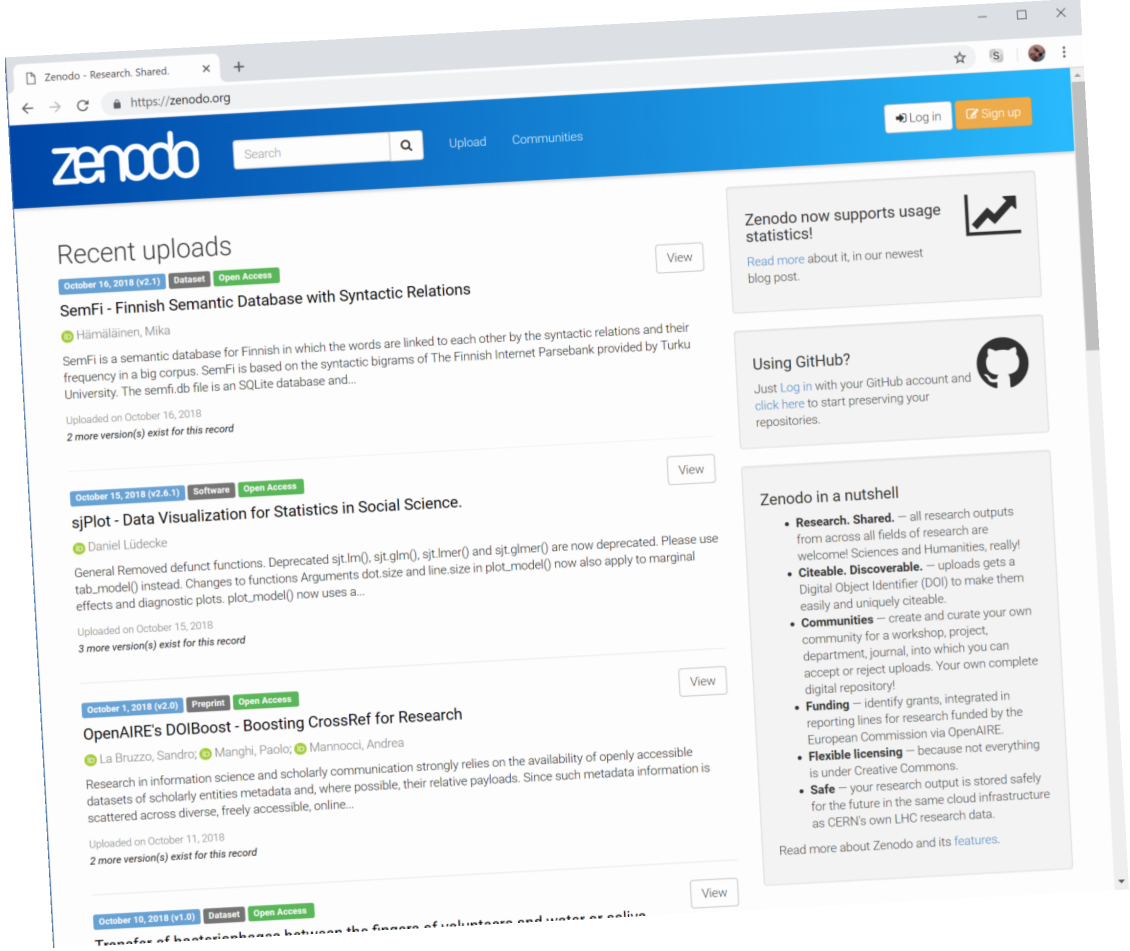
Open science

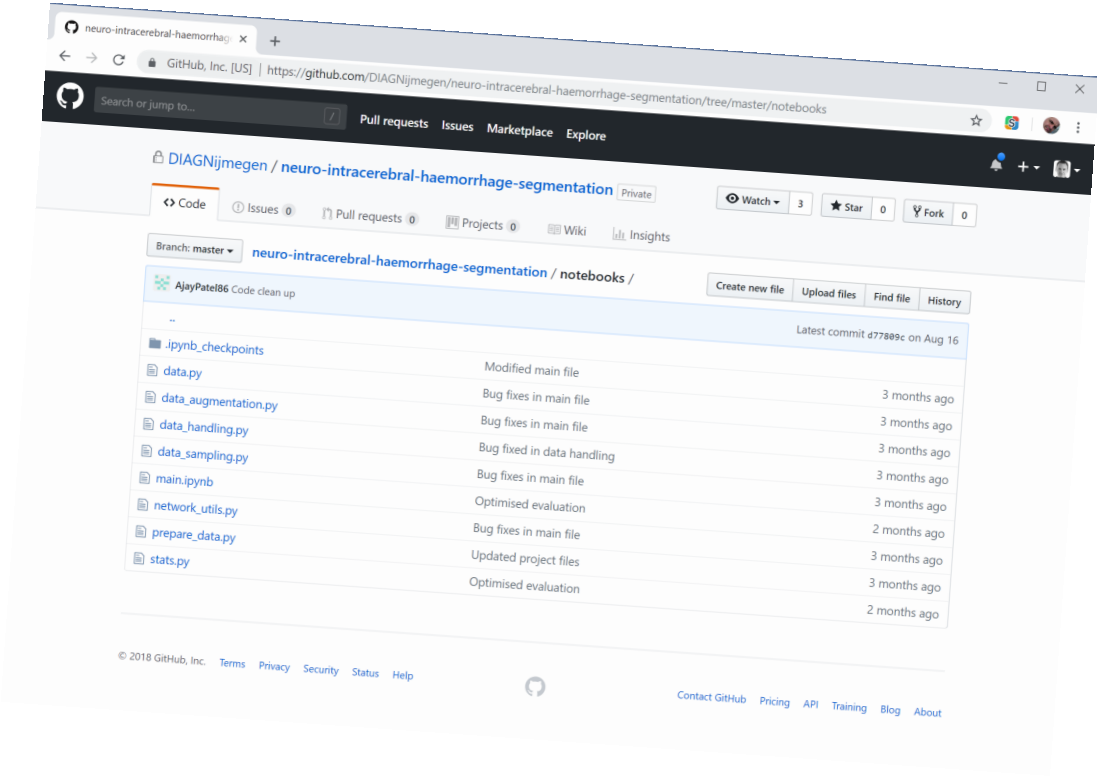
- Share data and annotations, evaluation code, algorithm code



Open science

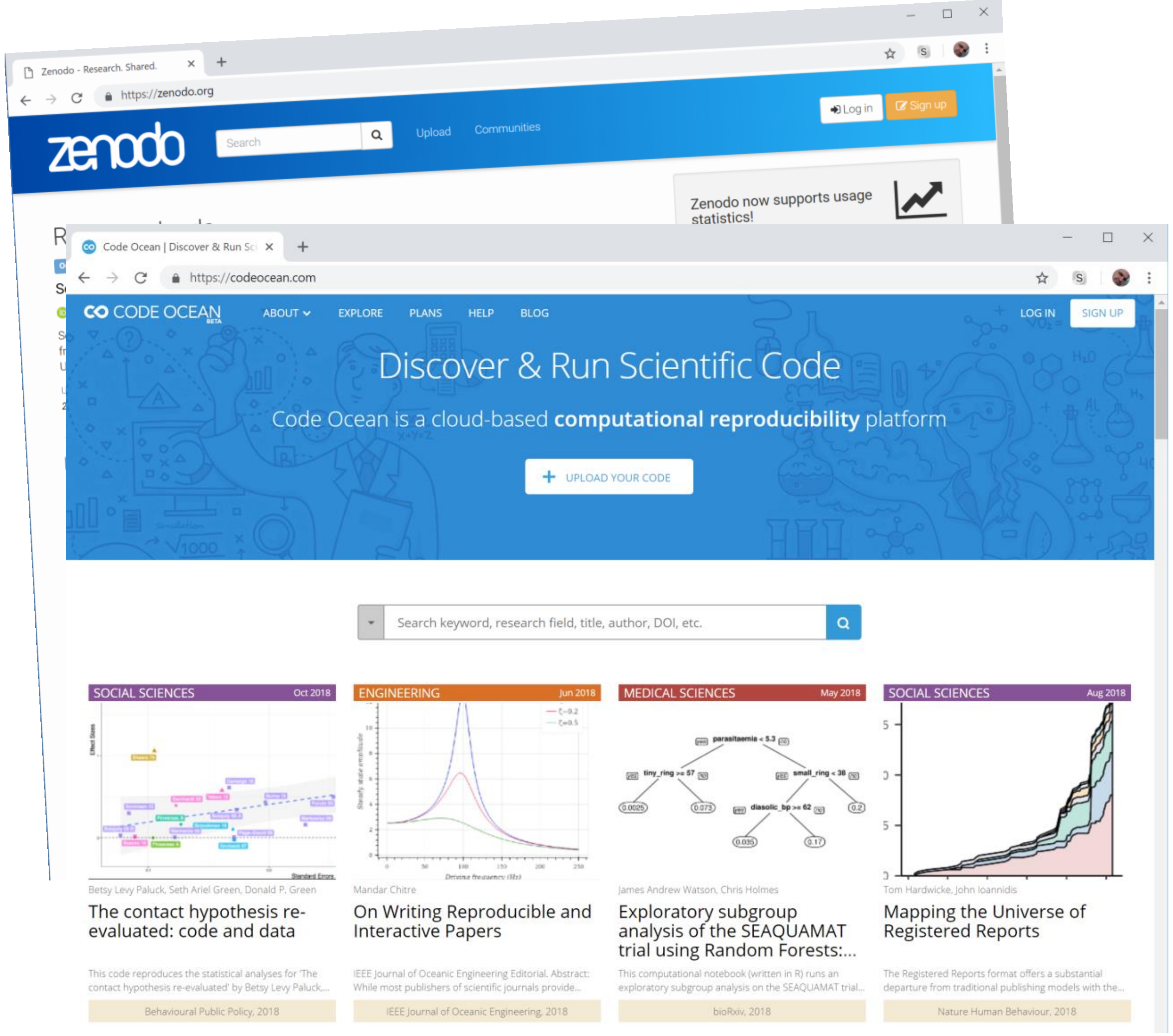
- Share data and annotations, evaluation code, algorithm code

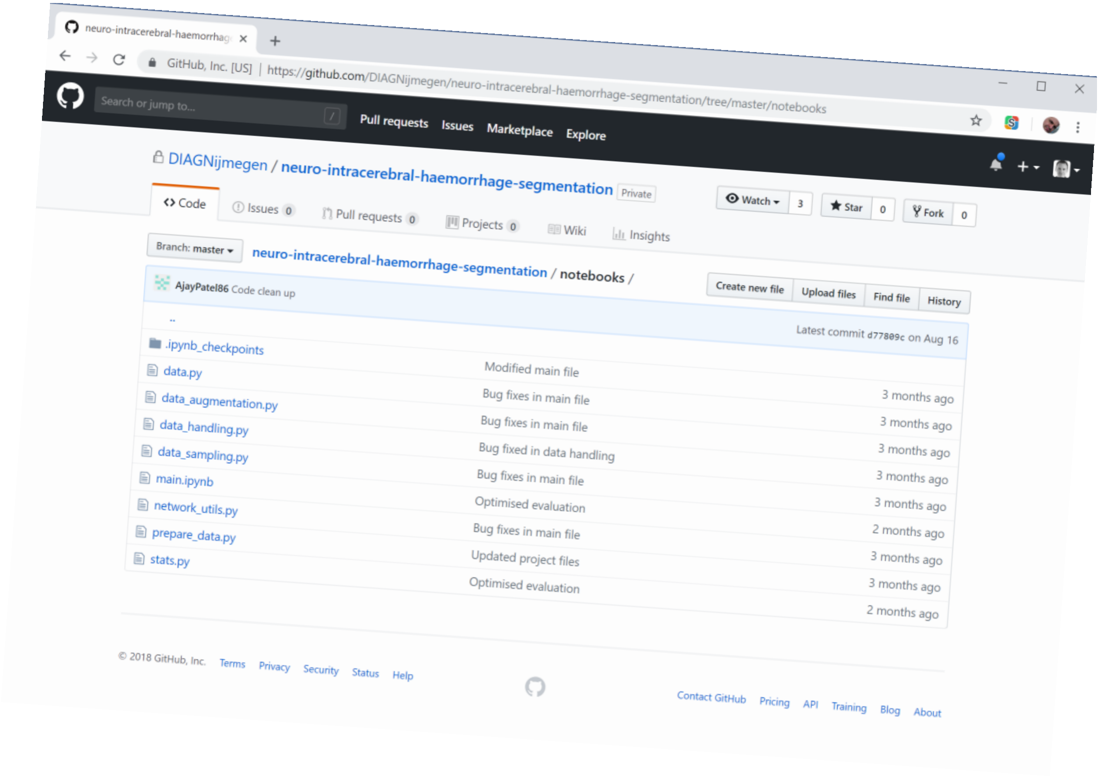




Open science

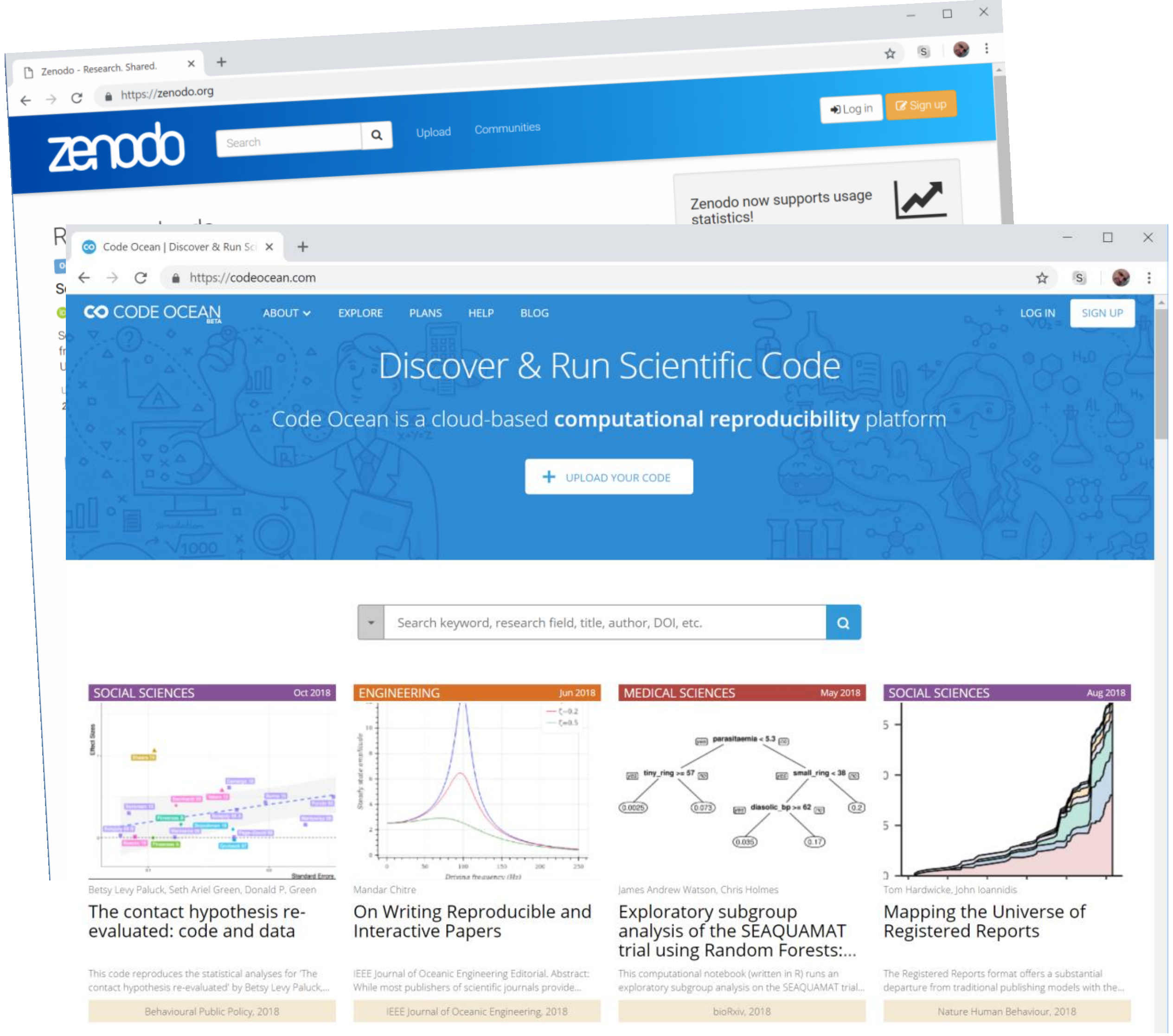
- Share data and annotations, evaluation code, algorithm code

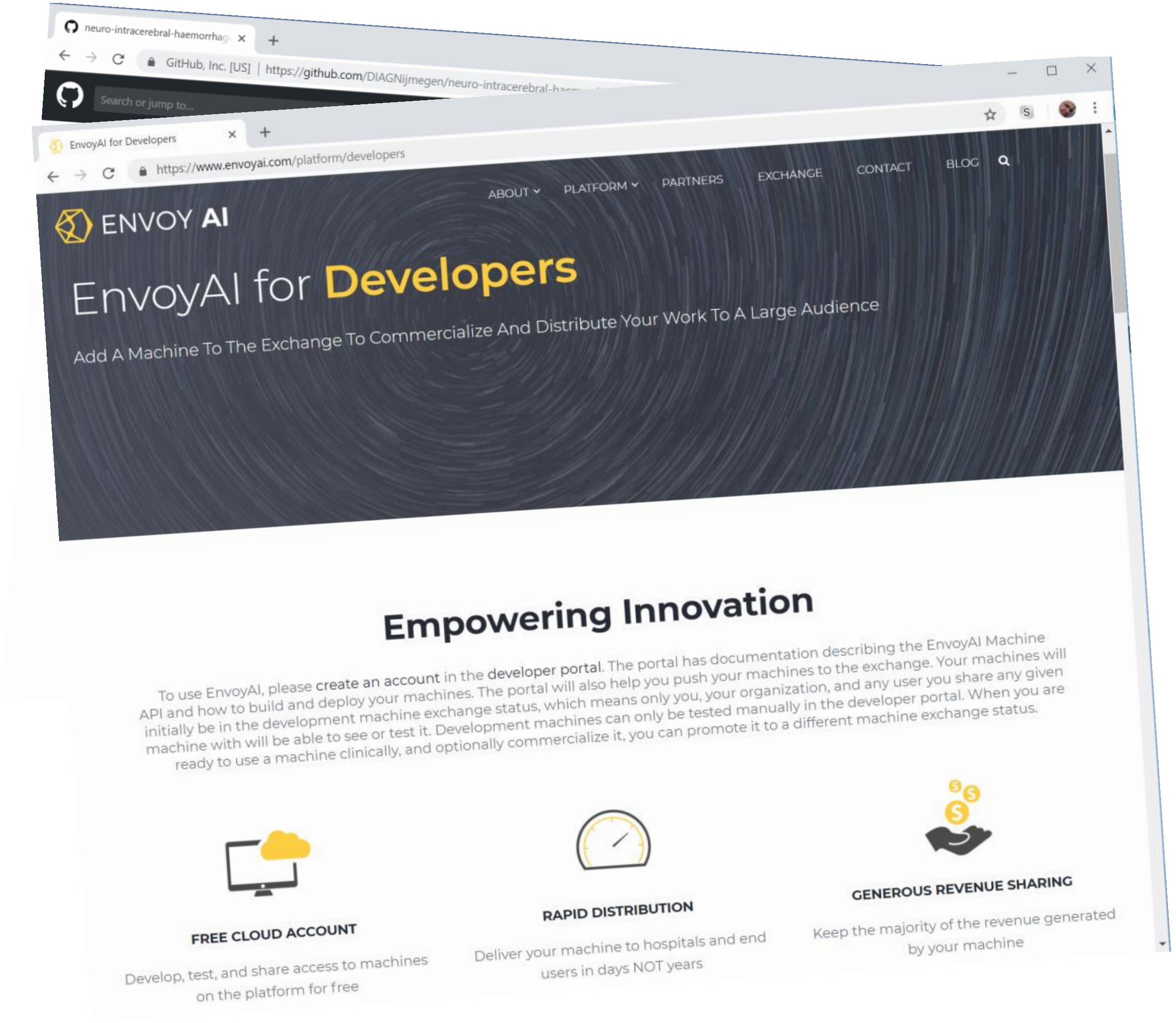




Open science

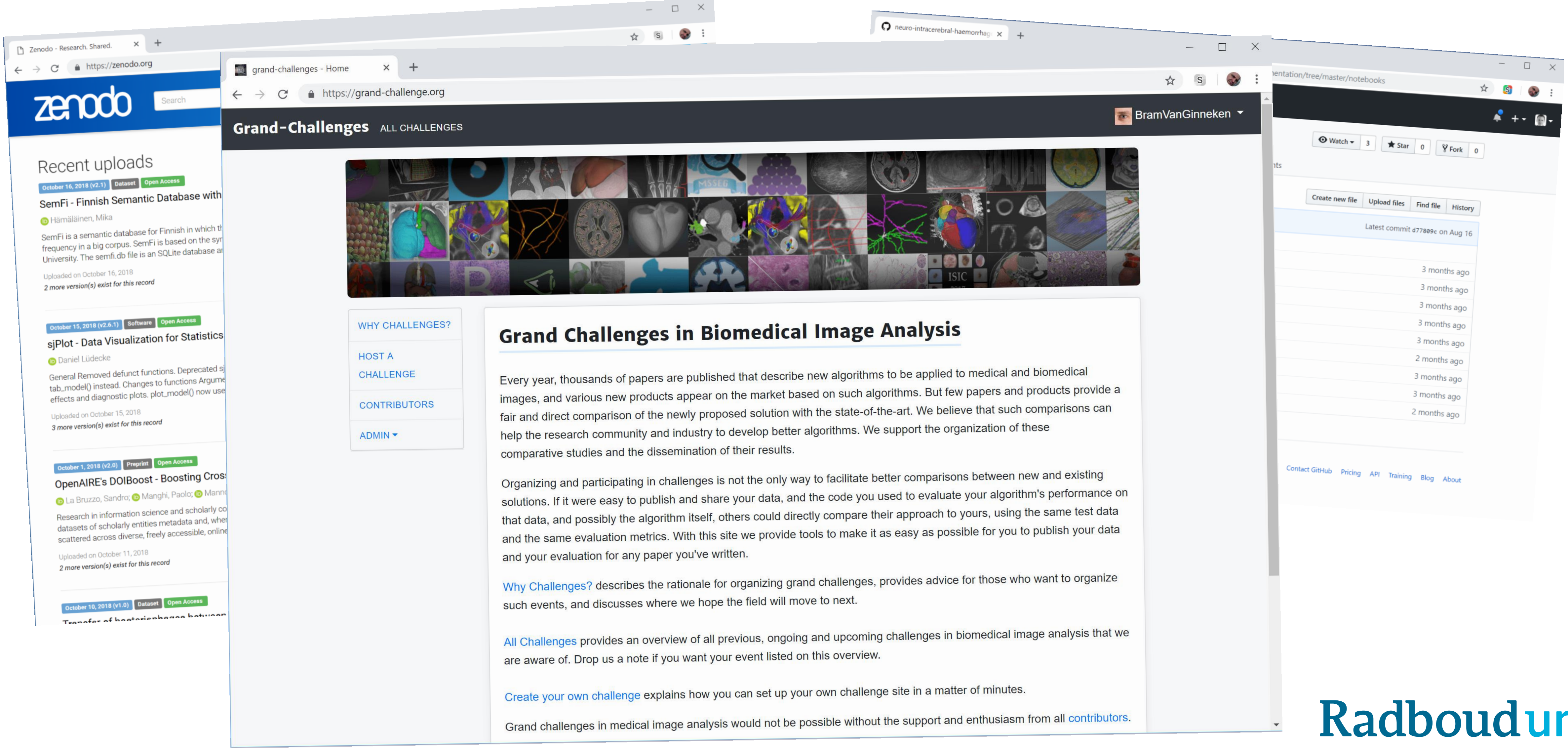
- Share data and annotations, evaluation code, algorithm code





Open science

- Share data and annotations, evaluation code, algorithm code



Radboudumc

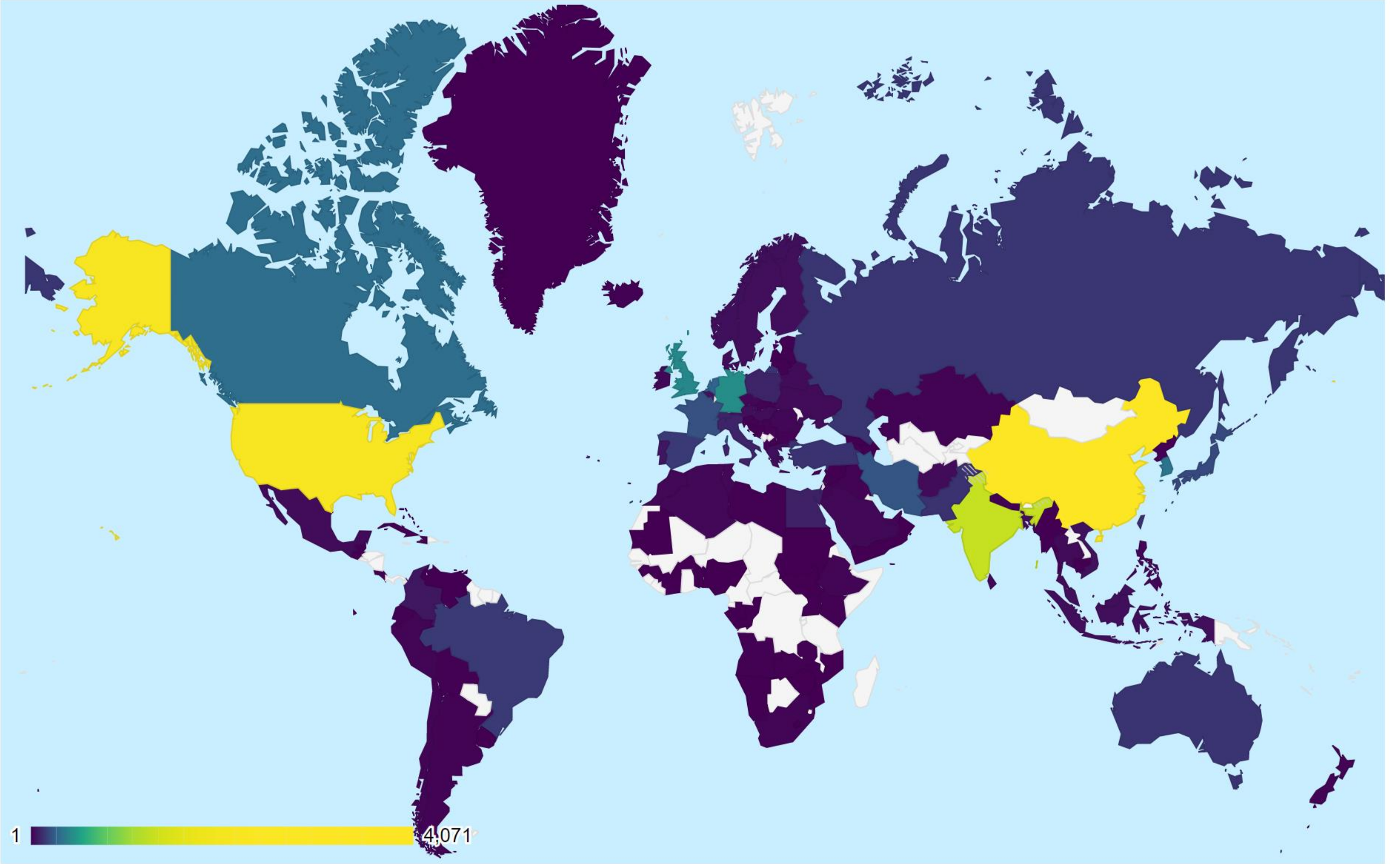
Statistics

https://grand-challenge.org/stats

Grand-Challenges ALL CHALLENGES BramVanGinneken

Site Statistics

Users



Total users: 17,939

New users (last 30 days): 892

Users logged in (last 30 days): 1,552

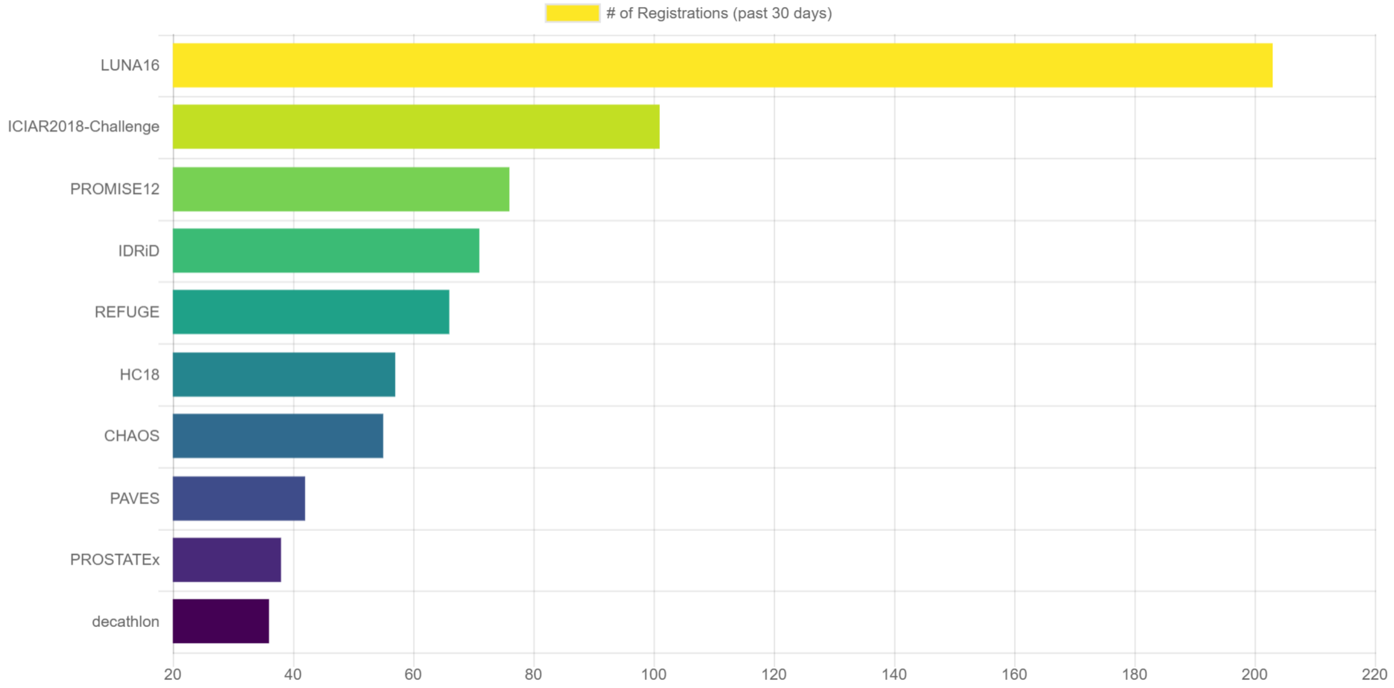
Statistics

https://grand-challenge.org/stats

Challenges

Public challenges: 35
Hidden challenges: 78
Public challenge with the most participants: LUNA16 (5,585 Participants)
Latest public challenge: ALL-challange

Registrations to public challenges in the past 30 days (top 10)



Evaluations

Challenges using automated evaluation: 45
Total submissions: 1,959
Total submissions (last 30 days): 177
Public challenge with the most submissions: decathlon (197 Submissions)
Latest public result: Result for ICIAR2018-Challenge, created 11 hours ago, 126th position on leaderboard.

Evaluation

Long term sustainability of challenges depends on automating the evaluation

- Dealing with user submissions
- Generating and updating the results tables
- Running the same evaluation code after n years

Automated Evaluation - our approach

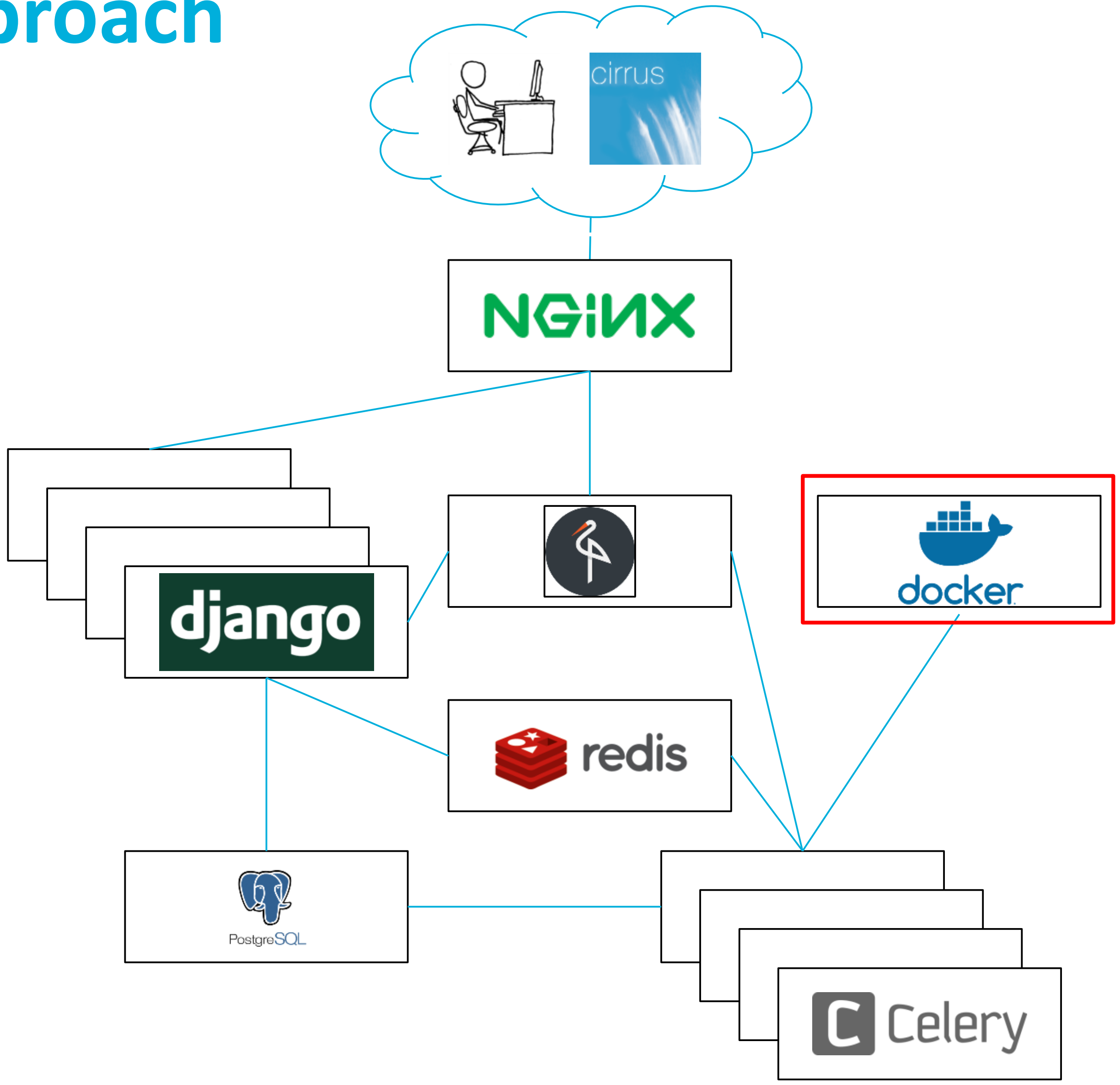
Use containers

- Challenge administrators can upload arbitrary code
- They are archived, and can run long term
- They are lightweight and can scale

Results stored in the database and tables dynamically updated

Define a clear API

- https://grand-challengeorg.readthedocs.io/en/latest/evaluation.html#evaluation-container-requirements
- The container will be executed on each submission, and only have access to this 1 submission
- The submission is available to the container on /input
- The container must produce /outputs/metrics.json
- Errors will be emailed to the user for them to fix



LOGGLY

Radboudumc

Problem

Now every challenge administrator needs to create an evaluation container

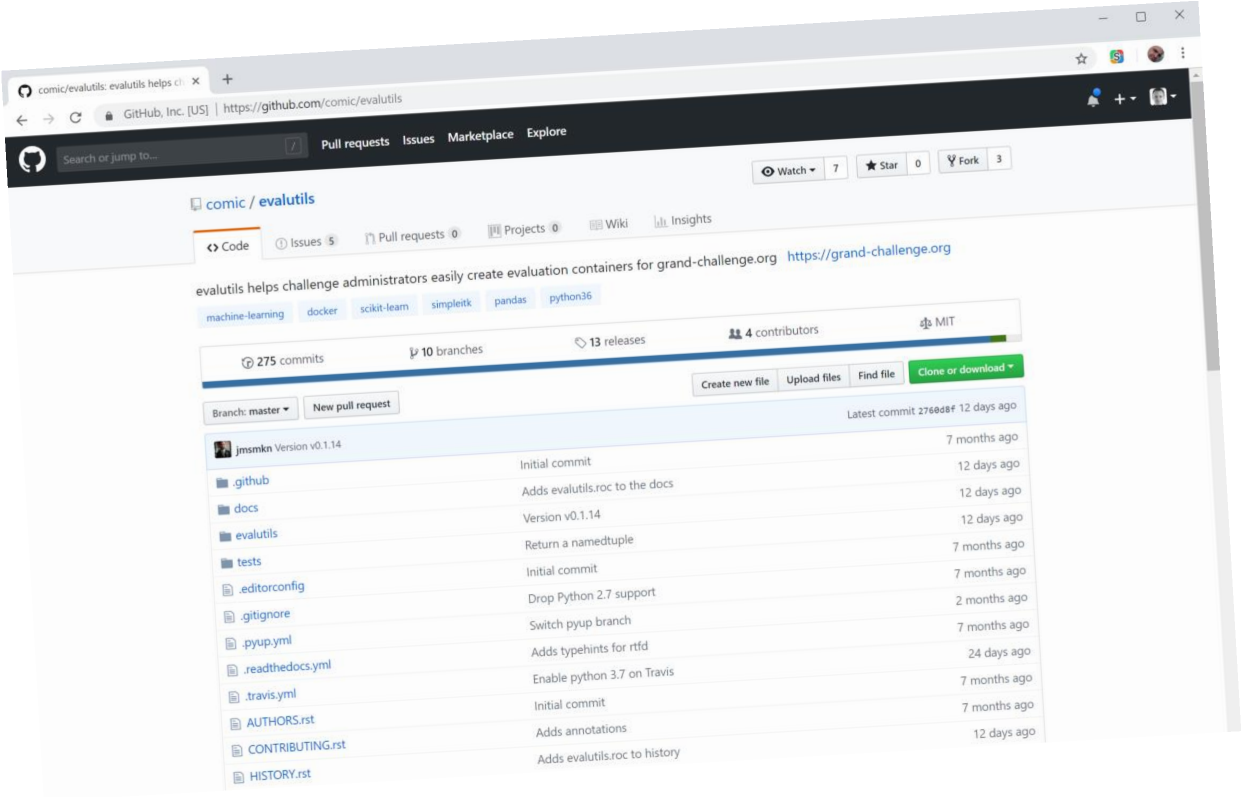
- Write a correct & maintainable dockerfile
- Load data
- Handle invalid submissions
- Repeated use of the same metrics across challenges
- Write the correct metrics to json

Evalutils

A tool to generate evaluation containers

- Boilerplate project generation
- Handles Classification, Segmentation and Detection tasks
- Pip installable
- IO module
- Validation module
- Metrics module

https://github.com/comic/evalutils



SPLEEN18

https://spleen18.grand-challenge.org/evaluation/methods/

Grand-Challenges ALL CHALLENGES

BramVanGinneken

SPLEEN18

- HOME
- EVALUATION
- JOIN
- TEAMS
- SUBMIT
- RESULTS
- ADMIN

Evaluation Methods

+ Add a new method

Show 10 entries

Search:

ID	Created	User	SHA
41162292-f53c-49be-a3cb-c719931d889e	Aug. 10, 2018, 12:53 p.m.	ghumpire	sha256:e34f8d7f26c5643fbc6f02ee0934ca85d62eed59e9b26b0212aaeb33effb2b
e5270bcf-97a0-4aa3-b870-a725e9b8ebde	Aug. 10, 2018, 1:44 p.m.	ghumpire	sha256:2a50d0b17c88363fa0c9725c55599068e9e3dc43d3737f0be24feb68c474c

Showing 1 to 2 of 2 entries

Previous 1 Next

SPLEEN18

https://spleen18.grand-challenge.org/evaluation/results/

Grand-Challenges ALL CHALLENGES

BramVanGinneken

SPLEEN18

- HOME
- EVALUATION
- JOIN
- TEAMS
- SUBMIT
- RESULTS
- ADMIN

Results

Show 10 entries

Search:

#	User (Team)	Created	Mean Dice	Max dice	Min dice	25pc dice	50pc dice	75pc dice	Mean jaccard	Mean volume similarity	Mean hausdorff distance	Comment
1st	ghumpire	10 Aug. 2018	0.9517	0.9792	0.7535	0.9492	0.9608	0.9703	0.9103	-0.0317	7.2430	

Showing 1 to 1 of 1 entries

Previous 1 Next

Annotations - SPLEEN18

https://spleen18.grand-challenge.org/datasets/annotations/b7e7d3e3-4a68-4f53-9ad0-ce6611edb025/

Grand-Challenges ALL CHALLENGES BramVanGinneken

SPLEEN18

- HOME
- EVALUATION
- JOIN
- TEAMS
- SUBMIT
- RESULTS
- ADMIN

Annotations

+ Add images to this annotation set

+ Set the labels on this annotation set

Update the images used in this annotation set

Kind

Prediction

Dataset

ImageSet object (1e68a48d-3e65-4e73-9102-23f5a9bd86ba)

Upload Sessions

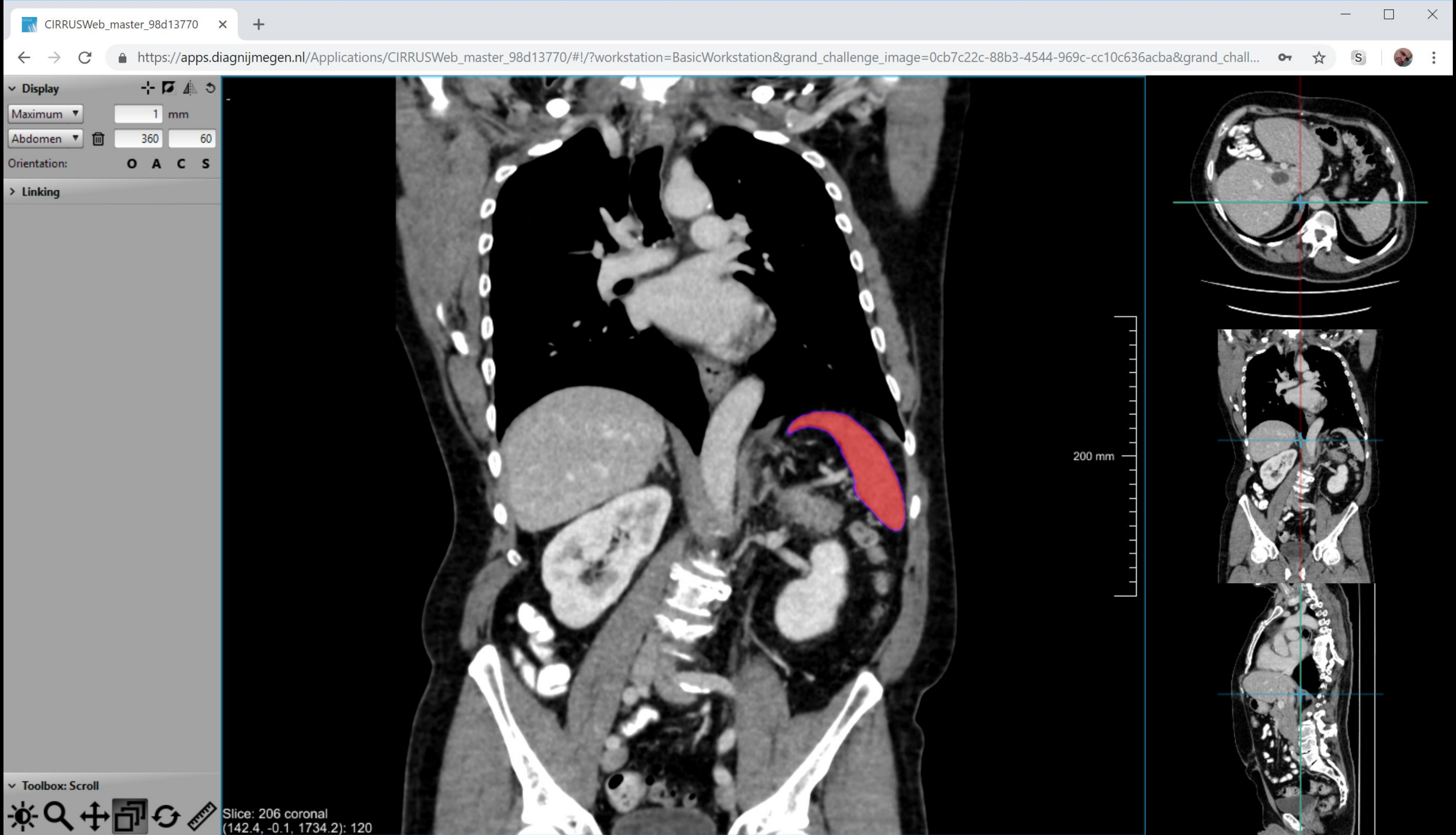
Upload Session <4b8ce781>, (stopped)

Matched Annotations

Show 10 entries

Search:

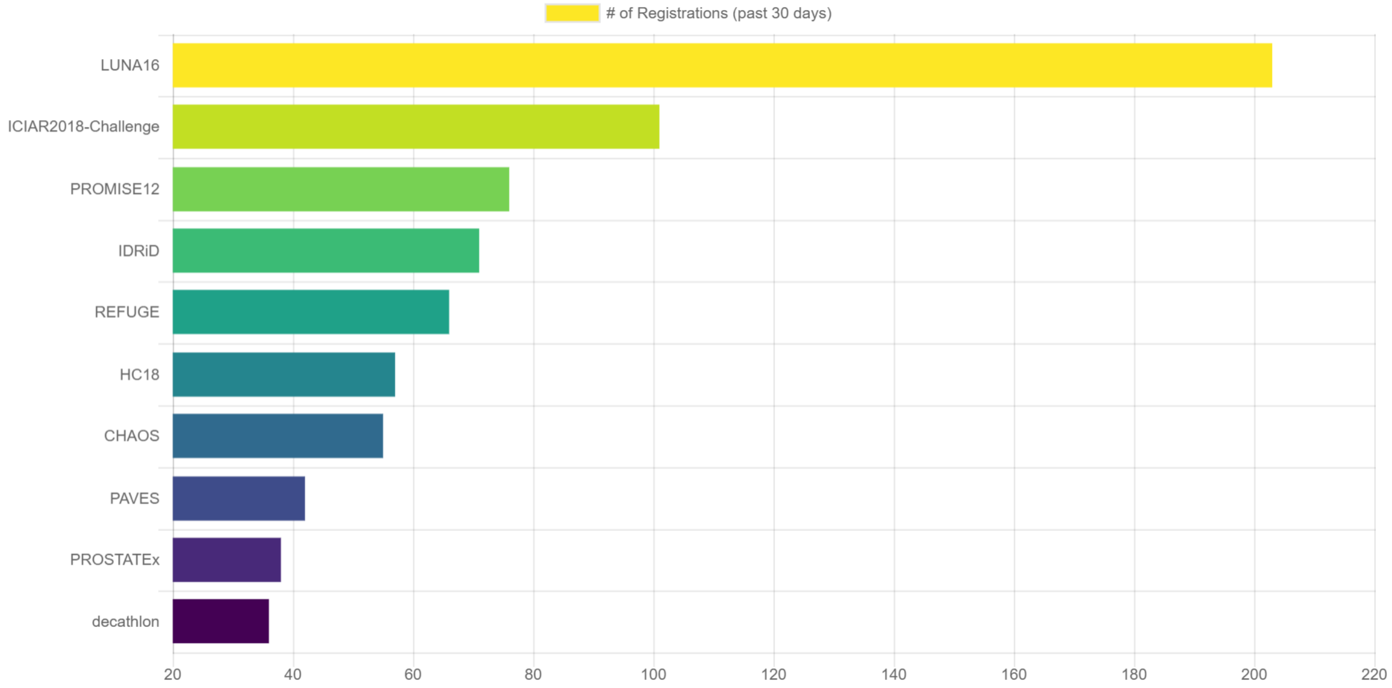
Key	Base Image
(' 23922000369116261372423319636142128464583060', 633, 412, 412)	Image 1.2.392.200036.9116.2.6.1.37.2423319636.1421284645.8306 [633, 412, 412]
(' 23922000369116261372423319636143571331610597', 601, 410, 410)	Image 1.2.392.200036.9116.2.6.1.37.2423319636.1435713316.1059 [601, 410, 410]
(' 23922000369116261372423319636144487632458827', 713, 400, 400)	Image 1.2.392.200036.9116.2.6.1.37.2423319636.1444876324.5882 [713, 400, 400]


CIRRUSWeb_master_98d13770
https://apps.diagnijmegen.nl/Applications/CIRRUSWeb_master_98d13770/#!/?workstation=BasicWorkstation&grand_challenge_image=0cb7c22c-88b3-4544-969c-cc10c636acba&grand_chall...
Display
Maximum
1
mm
Abdomen
360
60
Orientation:
O A C S
Linking
Toolbox: Scroll
200 mm
Slice: 206 coronal
(142.4, -0.1, 1734.2): 120

Challenges

Public challenges: 35
Hidden challenges: 78
Public challenge with the most participants: LUNA16 (5,585 Participants)
Latest public challenge: ALL-challange

Registrations to public challenges in the past 30 days (top 10)



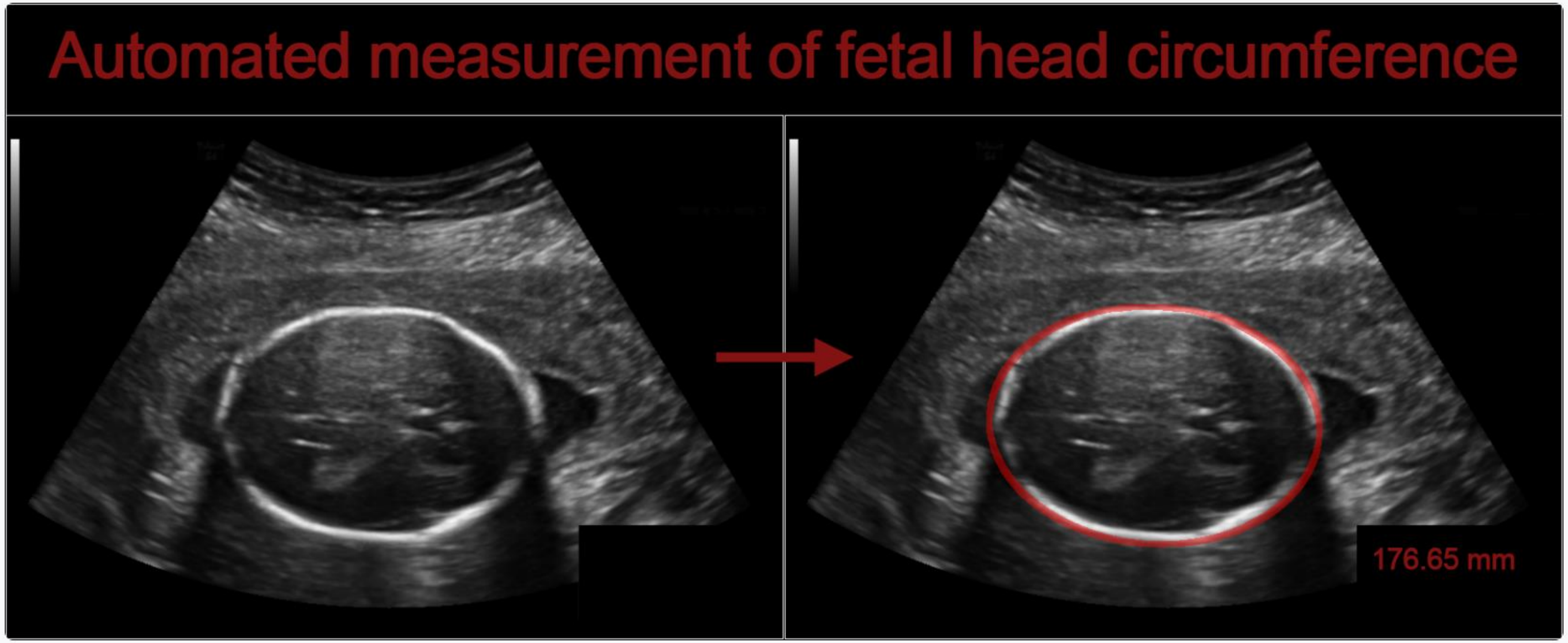
Evaluations

Challenges using automated evaluation: 45
Total submissions: 1,959
Total submissions (last 30 days): 177
Public challenge with the most submissions: decathlon (197 Submissions)
Latest public result: Result for ICIAR2018-Challenge, created 11 hours ago, 126th position on leaderboard.

HC18 - Home

https://hc18.grand-challenge.org

Grand-Challenges ALL CHALLENGES BramVanGinneken



- HOME
- JOIN
- SUBMIT
- RESULTS
- ADMIN

Background

During pregnancy, ultrasound imaging is used to measure fetal biometrics. One of these measurements is the fetal head circumference (HC). The HC can be used to estimate the gestational age and monitor growth of the fetus. The HC is measured in a specific cross section of the fetal head, which is called the standard plane. The dataset for this challenge contains a total of 1334 two-dimensional (2D) ultrasound images of the standard plane that can be used to measure the HC. This challenge makes it possible to compare developed algorithms for automated measurement of fetal head circumference in 2D ultrasound images.

Detailed description of the data

Show 10 entries

Search:

#	User	Created	Mean absolute difference (mm)	Std DICE (%)	Mean DICE (%)	Std difference (mm)	Mean difference (mm)	Std Hausdorff distance (mm)	Mean Hausdorff distance (mm)	Std absolu differe (mm)
1st	dra_v	9 Oct. 2018	2.3075	5.3681	95.2104	4.3083	0.1124	2.2359	2.4679	3.6378
2nd	thomasvandenheuvel	1 Oct. 2018	2.8332	2.7299	97.1010	4.2133	0.5603	1.6030	1.8319	3.1647
3rd	music	12 Oct. 2018	2.9370	7.2021	94.7254	5.5405	-1.6006	2.4486	2.6138	4.9614
4th	thomasvandenheuvel	1 Oct. 2018	3.0197	2.8001	97.0528	4.7317	0.7603	1.8327	1.9277	3.7179
5th	dra_v	1 Oct. 2018	3.4086	5.7906	94.5911	5.8620	0.3248	2.3650	2.8475	4.7765
6th	thomasvandenheuvel	1 Oct. 2018	3.6332	1.2309	97.2191	3.7182	2.8219	1.0795	1.7608	2.9278
7th	thomasvandenheuvel	1 Oct. 2018	4.1924	8.0644	95.4231	8.9749	1.3605	2.8286	2.3524	8.0484
8th	shenzexu	24 Oct. 2018	5.4308	10.0279	93.4925	10.0083	2.2773	3.6123	3.2561	8.7055
9th	631498444	11 Oct. 2018	6.3326	9.3412	94.6314	11.5555	4.8193	4.6598	3.2785	10.798
10th	631498444	9 Oct. 2018	6.4401	16.5091	87.8221	14.6160	1.2314	7.3371	7.4234	13.173

Showing 1 to 10 of 15 entries

Next steps

- We can use containers not only for the evaluation, but also for the algorithms
- Users/teams upload a docker container with an algorithm (and a license)
- Algorithm is applied to secret test data → leaderboard entry created

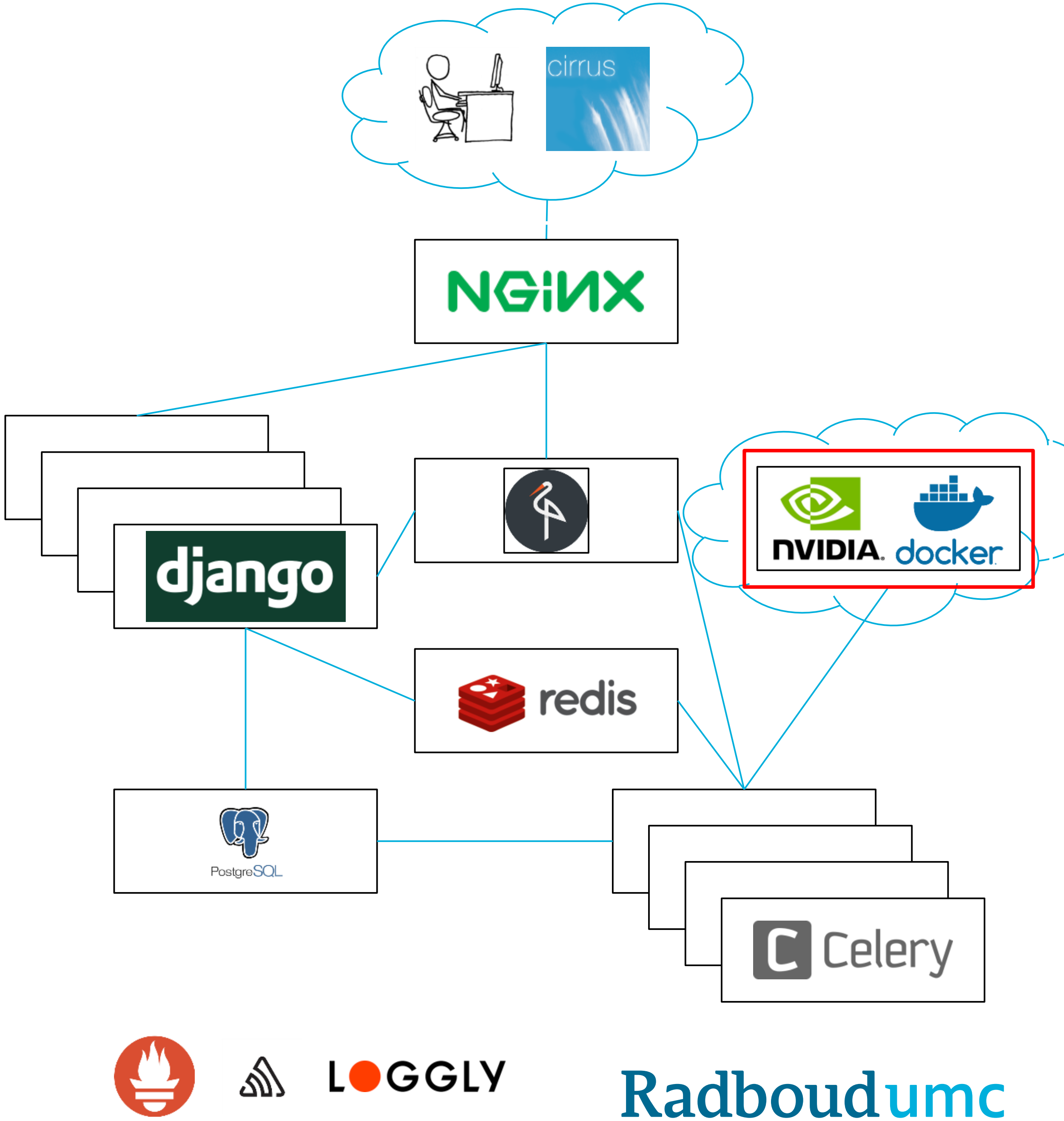
- Teams could also upload containers to **train** systems on secret training data and download trained model afterwards (e.g. Mammography DREAM challenge used such a set up)

- External user could upload data and run algorithms on it, possibly for a fee (app store)
- These cloud based algorithms/apps could be embedded in a cloud based viewing environment

- External users could also assist in annotating data; crowd-sourcing platform that integrates data storage, annotation tools, algorithms

Algorithm execution

- Run on DIAG infrastructure or AWS
  - Choice of CPU only, K80 or V100
- Improvements
  - Separate queues for evaluations and algorithm executions
  - Speed
    - Docker registry on AWS
    - Data replica on AWS/apps.diagnijmegen.nl
  - Scaling and cost: spot instances w/ cost tracking?
  - Algorithm permissions: who can run it?
  - Service based architecture for integration in apps?
  - Results interface



Algorithms - grand-challenges

https://grand-challenge.org/algorithms

Grand-Challenges ALL CHALLENGES BramVanGinneken

Algorithms

+ Add a new algorithm

Existing Algorithms

Show 50 entries Search:

Title	Created	Creator
Cardiomegaly Detection	Sept. 13, 2018, 1:31 p.m.	jamesmeakin.diag

Showing 1 to 1 of 1 entries

Previous 1 Next

Dataset

3 different datasets are used in the experiment. In particular:

1. In the training phase, SCR Dataset and JSRT are utilized. Both of the dataset are publicly avaible and can be downloaded. JSRT dataset consists of 247 PA chest x-rays. SCR dataset provides the segmentation result for the lung fields, heart and clavicles on 247 images takes from JSRT dataset.
2. In the evaluation phase, publicly avaible the ChestX-ray14 dataset dataset is utilized. It comprises 112,120 frontal-view chest X-ray images of 30,805 unique patients with 14 disease labels.

Try out this algorithm

Jobs for this algorithm

Show 50 entries

Search:

Created	Creator	Image	Status	Result
Oct. 25, 2018, 3:08 p.m.	pkcakeout248130100a1c4c02	Image cardiomegaly.mhd [1024, 1024]	The task executed successfully	{'out.mhd': {'cardiomegaly': {'errors': [], 'classification': True}}}
Sept. 14, 2018, 11:48 a.m.	pkcakeout248130100a1c4c02	Image 00000089_000.mhd [1024, 1024]	The task executed successfully	{'out.mhd': {'cardiomegaly': {'errors': [], 'classification': False}}}
Sept. 14, 2018, 11:48 a.m.	pkcakeout248130100a1c4c02	Image 00000088_000.mhd [1024, 1024]	The task executed successfully	{'out.mhd': {'cardiomegaly': {'errors': [], 'classification': False}}}
Sept. 14, 2018, 8:53 a.m.	jamesmeakin.diag	Image 00000008_002.mhd [1024, 1024]	The task executed successfully	{'out.mhd': {'cardiomegaly': {'errors': [], 'classification': True}}}
Sept. 14, 2018, 8:53 a.m.	jamesmeakin.diag	Image 00000008_001.mhd [1024, 1024]	The task executed successfully	{'out.mhd': {'cardiomegaly': {'errors': [], 'classification': False}}}

Run an algorithm

Select the images that you would like to run the algorithm on.

Image files*

Choose Files

Upload images for creating a new archive

Submit

Dataset

3 different datasets are used in the experiment. In particular:

1. In the training phase, SCR Dataset and JSRT are utilized. Both of the dataset are publicly avaible and can be downloaded. JSRT dataset consists of 247 PA chest x-rays. SCR dataset provides the segmentation result for the lung fields, heart and clavicles on 247 images takes from JSRT dataset.
2. In the evaluation phase, publicly avaible the ChestX-ray14 dataset dataset is utilized. It comprises 112,120 frontal-view chest X-ray images of 30,805 unique patients with 14 disease labels.

Try out this algorithm

Jobs for this algorithm

Show 50 entries   Search:

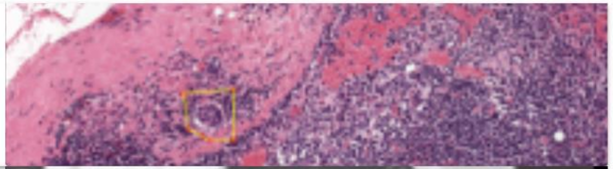
Created	Creator	Image	Status	Result
Oct. 25, 2018, 3:08 p.m.	pkcakeout248130100a1c4c02	Image cardiomegaly.mhd [1024, 1024]	The task executed successfully	{'out.mhd': {'cardiomegaly': {'errors': [], 'classification': True}}}
Sept. 14, 2018, 11:48 a.m.	pkcakeout248130100a1c4c02	Image 00000089_000.mhd [1024, 1024]	The task executed successfully	{'out.mhd': {'cardiomegaly': {'errors': [], 'classification': False}}}
Sept. 14, 2018, 11:48 a.m.	pkcakeout248130100a1c4c02	Image 00000088_000.mhd [1024, 1024]	The task executed successfully	{'out.mhd': {'cardiomegaly': {'errors': [], 'classification': False}}}
Sept. 14, 2018, 8:53 a.m.	jamesmeakin.diag	Image 00000008_002.mhd [1024, 1024]	The task executed successfully	{'out.mhd': {'cardiomegaly': {'errors': [], 'classification': True}}}
Sept. 14, 2018, 8:53 a.m.	jamesmeakin.diag	Image 00000008_001.mhd [1024, 1024]	The task executed successfully	{'out.mhd': {'cardiomegaly': {'errors': [], 'classification': False}}}

CIRRUS Applications

https://apps.diagnijmegen.nl

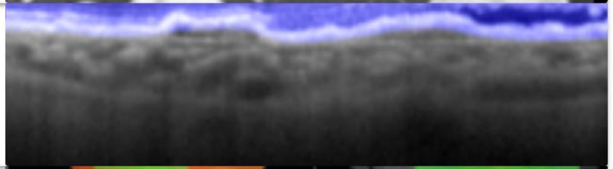
DIAG Applications

CIRRUS Pathology

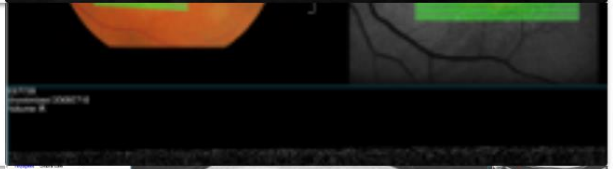


CIRRUS Oncology

AMI Ophthalmology AMD Workstation



AMI Ophthalmology Multimodal Workstation



CIRRUS Lung Screening

CIRRUS Lung Screening (Veolity branding)

Summary

- Benchmarking / challenges / competitions are useful for many reasons
- A paper or research project in which some algorithm is developed is a kind of challenge where the author is the only participant (and possibly some own implementations of prior work are shown to be outperformed)
- A challenge platform can be used for research projects (and for courses, etc etc)
- Excellent open source components exist today to build such a platform in a modular fashion
- Containers/dockers make sharing of running code easier; this will become more and more common
- Other use cases: app stores, could/web software for viewing and processing medical images, joint annotation and processing
- Please join and use/develop open source tools for benchmarking and reproducible science!

A Shape-Guided Deformable Model 11

Table 2. Segmentation error in comparison with a standard ASM and numbers from previous work (customized ASM [22] and deformable simplex mesh [5]), given as $\mu \pm \sigma$

Segmentation method	$D_{avg}[mm]$	$D_{RMS}[mm]$	$V_D[\%]$
Deformable model	1.6 ± 0.5	3.3 ± 1.2	5.1 ± 1.4
Active Shape Model	2.9 ± 1.1	5.2 ± 2.3	8.9 ± 2.4
Lamecker et al. [22]	2.3 ± 0.3	3.1 ± 0.5	7.0 ± 1.8
Soler et al. [5]	$2 \pm ?$	*n.a.*	*n.a.*