

Clinical and Translational Imaging Informatics Project (CTIIP)

Speaker Series Presentation

July 6, 2016

Objective

Standards, Infrastructure for
Pre-Clinical and Co-Clinical Studies



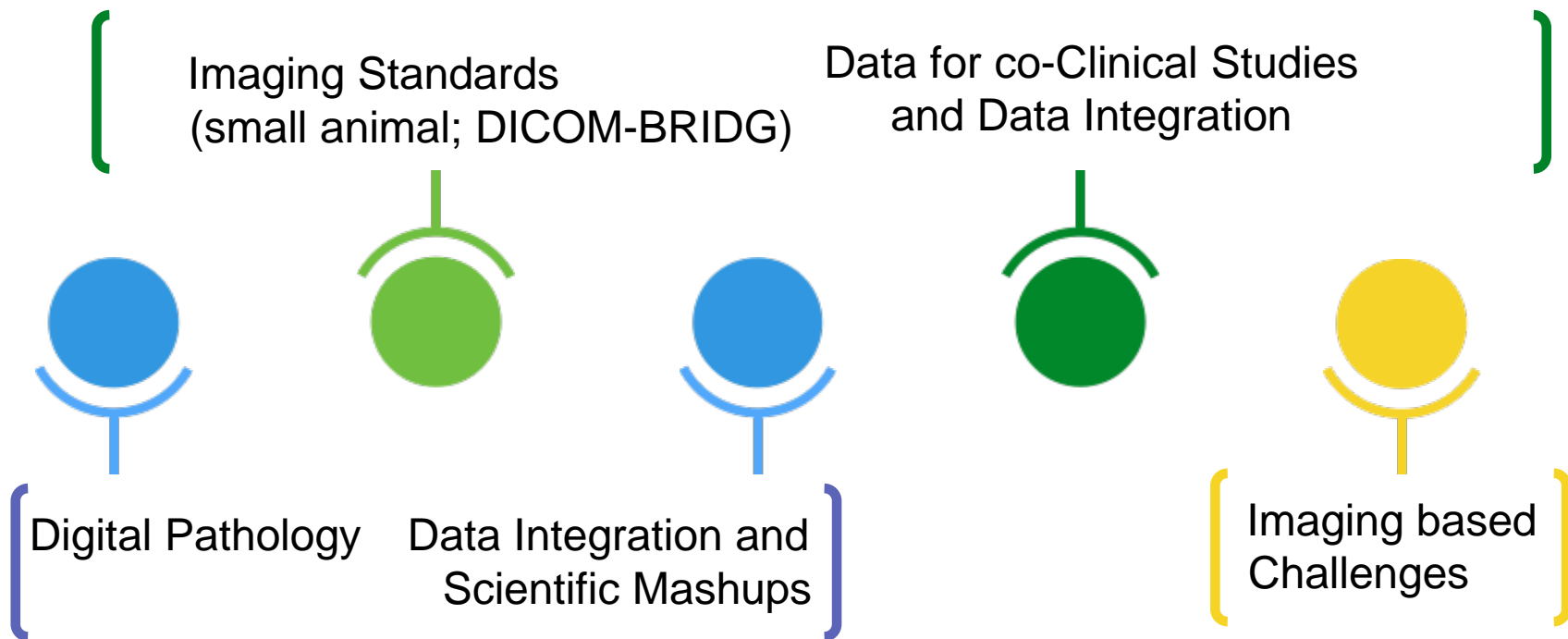
Interoperability and Integration
with Digital Pathology Studies



Evaluate with Integrative
Pilot Challenges



Approach



Aim 1



Improve Digital Pathology Tools and establish an Integrative Query System

- Enable caMicroscope to
 - Directly serve whole slide pathology images from the majority of digital pathology vendors.
 - Execute basic image analysis algorithms
 - Propose and develop specifications and tools for image annotation and markup.
- Explore data mashups between image-derived information, clinical, and molecular data

PI: Dr. Ashish Sharma, Emory University

Co-PIs: Dr. Joel Saltz, Stonybrook University

Dr. Fred Prior, University of Arkansas Medical Sciences

Aim 2



- Improving Standards in Imaging
 - Enable DICOM* compliance of small-animal imaging by including it in the DICOM standard
 - Coordination of two ISO standards by including DICOM* references in the BRIDG** Imaging subdomain

PI: Dr. David Clunie, Pixelmed

- Provide data including DICOM-compliant animal images for development and evaluation of data integration strategies

PI: Dr. Robert Cardiff, University of California Davis

* Digital Imaging and Communications in Medicine (DICOM) Standard

** Biomedical Research Integrated Domain Group (BRIDG) Model.
BRIDG is a collaborative effort engaging stakeholders from CDISC, HL7, ISO, NCI, and FDA

Aim 3



Establish an Imaging Challenge Management System utilizing existing tools and execute pilot challenges in

- ▣ Clinical Imaging
- ▣ Preclinical/co-clinical Imaging
- ▣ Digital Pathology

PI: Dr. Jayashree Kalpathy-Cramer, Massachusetts General Hospital

Co-PIs: Dr. Daniel Rubin, Stanford University
Dr. Ashish Sharma, Emory University

MedICI: an Infrastructure for Challenges

Jayashree Kalpathy-Cramer, PhD

Athinoula A. Martinos Center for Biomedical Imaging,
Massachusetts General Hospital, Boston, MA



MGH/HST Athinoula A. Martinos
Center for Biomedical Imaging

Aim 3



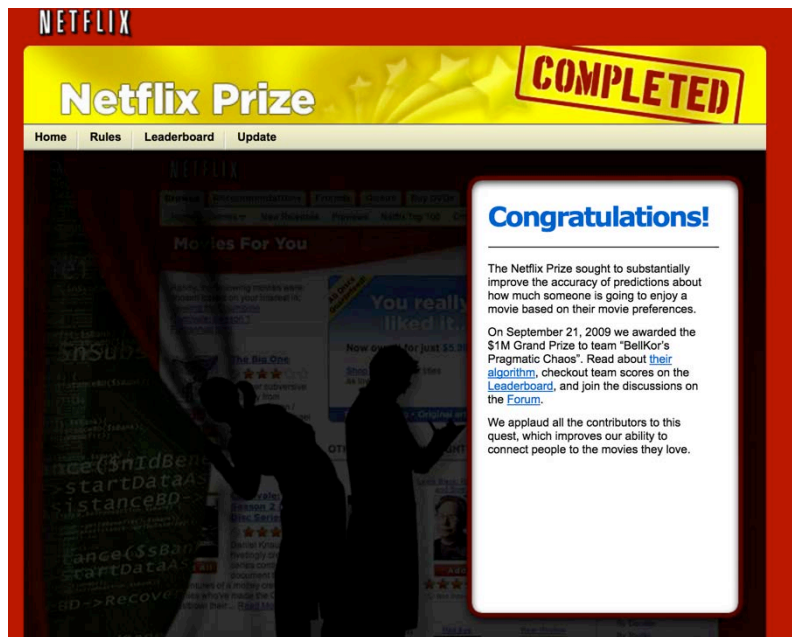
Establish an Imaging Challenge Management System utilizing existing tools and execute pilot challenges in

- ▣ Clinical Imaging
- ▣ Preclinical/co-clinical Imaging
- ▣ Digital Pathology







PI: Dr. Jayashree Kalpathy-Cramer, Massachusetts General Hospital
Artem Mamanov, Karl Helmer, Massachusetts General Hospital

Co-PIs: Dr. Daniel Rubin, Stanford University
Dr. Ashish Sharma, Emory University

Challenges are a popular means of engaging the community in problem solving



<http://www.netflixprize.com/index>

Competition Name	Reward	Teams	Deadline
 Heritage Health Prize Identify patients who will be admitted to a hospital within the next year using historical claims data. (Enter by 06:59:59 UTC Oct 4 2012)	\$500,000	1353	3 years ago
 GE Flight Quest Think you can change the future of flight?	\$250,000	173	3 years ago
 Flight Quest 2: Flight Optimization, Milestone Phase Optimize flight routes based on current weather and traffic.	\$250,000	129	2 years ago
 Flight Quest 2: Flight Optimization, Main Phase Optimize flight routes based on current weather and traffic.	\$220,000	121	2 years ago
 Flight Quest 2: Flight Optimization, Final Phase Final Phase of Flight Quest 2	\$220,000	33	2 years ago
 Second Annual Data Science Bowl Transforming How We Diagnose Heart Disease	\$200,000	192	3 months ago

<https://www.kaggle.com/competitions>

Self driving cars

THE DARPA GRAND CHALLENGE: TEN YEARS LATER

March 13, 2014

Autonomous vehicle challenge led to new technologies and invigorated the prize challenge model of promoting innovation

At the break of dawn on March 13, 2004, 15 vehicles left a starting gate in the desert outside of Barstow, Calif., to make history in the DARPA Grand Challenge, a first-of-its-kind race to foster the development of self-driving ground vehicles. The immediate goal: autonomously navigate a 142-mile course that ran across the desert to Primm, Nev. The longer-term aim was to accelerate development of the technological foundations for autonomous vehicles that could ultimately substitute for men and women in hazardous military operations, such as supply convoys.

- Ushered in the self driving era

Why (prize) challenges?

- Why does DARPA sometimes turn to prizes instead of conventional means of encouraging progress, such as by funding activities through contracts and grants?
 - Prizes Encourage Thinking Outside the Box: Unlike grants and contracts, which are awarded in the hope that the recipient will be successful, prizes allow a funder to establish an ambitious goal without having to predict who or which approach is most likely to succeed, making way for novel approaches that might otherwise seem too risky to pursue.
 - Prizes Encourage Broad Participation: Prizes attract a wide array of potential solvers to tackle a problem and not just the usual experts in a given field.
 - The Economics Are Great: Prize purses are paid out only if someone succeeds, and in many cases, the amount of time and money invested by multiple teams as they vie for a prize exceeds the size of the prize purse itself.

THE OPEN GOVERNMENT PARTNERSHIP

SECOND OPEN GOVERNMENT NATIONAL ACTION PLAN FOR THE UNITED STATES OF AMERICA

December 5, 2013

5. Promote Innovation Through Collaboration and Harness the Ingenuity of the American Public

Creating a more Open Government and addressing our Nation's most challenging issues requires an informed and active citizenry. Recognizing the value of the American public as a strategic partner in addressing some of the country's most pressing challenges, the United States will work to more effectively harness the expertise, ingenuity, and creativity of the American public by enabling, accelerating, and scaling the use of open innovation methods across the Federal Government, including commitments to:

- **Create an Open Innovation Toolkit.** In 2014, the Administration will convene an interagency group to develop an "open innovation toolkit" for Federal agencies that will include best practices, training, policies, and guidance on authorities related to open innovation, including approaches such as incentive prizes, crowdsourcing, and citizen science.
- **New Incentive Prizes and Challenges on Challenge.gov.** The U.S. Government champions the use of challenges, prizes, and competitions to catalyze breakthroughs in national priorities. Launched on September 2010, Challenge.gov has hosted more than 300 crowdsourcing competitions, and the platform has been used by more than 50 Federal departments and agencies. The website will continue to provide public listings of new competitions offered by the Administration to engage citizens in solving difficult problems to help agencies achieve their missions.
- **Increased Crowdsourcing and Citizen Science Programs.** Public participation in scientific research, one type of crowdsourcing known as "citizen science", allows the public to make

[CHALLENGES](#)[PRIZEWIRE](#)[ABOUT](#)[CONTACT](#)[LOG IN](#)[SEE HOW IT WORKS](#)[NEWEST CHALLENGES](#)Sort 
Search

702 Competitions Found

2016 CDFI Fund Prize Competition

**\$1,000,000 in prizes**

The CDFI Fund is seeking new ways to increase CDFI Investments in Underserved Rural Communities

Open Until**Jul 29, 2016**

Posted by:

Department of the Treasury

Vascular Tissue Challenge

**\$500,000 in prizes**

A first-to-demonstrate challenge to create thick, human vascularized organ tissue surviving 30 calendar days

Open Until**Sep 30, 2019**

Posted by:

National Aeronautics and Space Administration

\$100,000 for Start a SUD Startup

**National Institute on Drug Abuse****\$100,000 in prizes**

The goal of this Challenge is to enable biomedical scientists to test the hypothesis that their research idea can be fostered into...

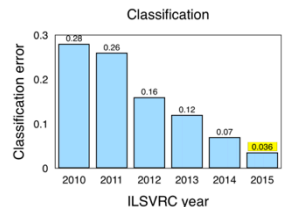
Open Until**Sep 16, 2016**

Posted by:

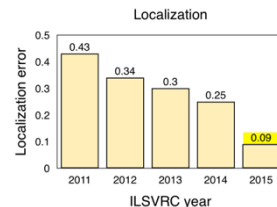
National Institute on Drug Abuse

ImageNet Large Scale Visual Recognition Challenge (ILSVRC)

- ▣ Widely successful challenge for image classification
- ▣ ImageNet database
 - ▣ Over 15M labeled high resolution images
 - ▣ 21841 categories
 - ▣ Collected from web and labeled by Amazon Mechanical Turk
- ▣ ILSVRC
 - ▣ Annual competition of image classification at large scale
 - ▣ 1.2M images in 1K categories
 - ▣ 10x improvement in 5 years



1.9x ↓



2.8x ↓

Challenges have been successful in a number of areas including cancer genomics

Boutros *et al. Genome Biology* 2014, **15**:462
<http://genomebiology.com/2014/15/9/462>



OPINION

Toward better benchmarking: challenge-based methods assessment in cancer genomics

Paul C. Boutros^{1,2,3}, Adam A. Margolis^{4,5}, Joshua A. Starr⁶, Andrew C. Collins⁷ and Cristian G. Cantu^{8*}

Abstract

Rapid technological development has created an urgent need for improved evaluation of algorithms for the analysis of cancer genomics data. We outline how challenge-based assessment may help fill this gap by leveraging crowd-sourcing to distribute effort and reduce bias.

Why “challenges” in medicine?

- Reproducibility is an issue in all aspects of medicine
 - Algorithm performance often not replicated by other sites
- Access to clinical data of sufficient variety can be a challenge for (computational) scientists developing algorithms
- Can evaluate the performance of techniques on real, noisy clinical data
- Test data (sequestered) can provide indication of algorithm generalizability to unseen data
- Allows for cross-pollination of methods from other domains
- Best algorithms can be translated into commercial products

Coding4Cancer

Challenges for improving cancer screening

[The Digital Mammography ...](#) » 2 - Challenge Overview

2 - Challenge Overview

The Digital Mammography DREAM Challenge



Out of every 1000 women screened, only 5 will have breast cancer. But 100 will be recalled for further testing.

We can do better.

Build a model to help reduce the recall rate for breast cancer screening.

Calling all coders to join the Challenge.

Up to a **\$1,000,000** in cash prizes for winning models.

May the best model win.

<https://www.synapse.org/#!Synapse:syn4224222>

What is a “challenge”?

- Impartial group of scientists (govt./commercial org) organizes a ‘challenge’ to solve a (clinical relevant) problem
 - Meaningful question
 - Well curated, representative dataset
 - Well established evaluation metrics
- Typically split into a training dataset, a validation dataset (optional) and test dataset
 - Test data withheld from challenge participants and used for final evaluation
- Leaderboards can provide real-time feedback to participants based on the validation dataset
- Final results based on (gold-standard, preferably independent) test dataset
- Such a design closely reflects the actual difficulties faced by real-world users trying to determine whether **an algorithm generalizes to unseen cases**

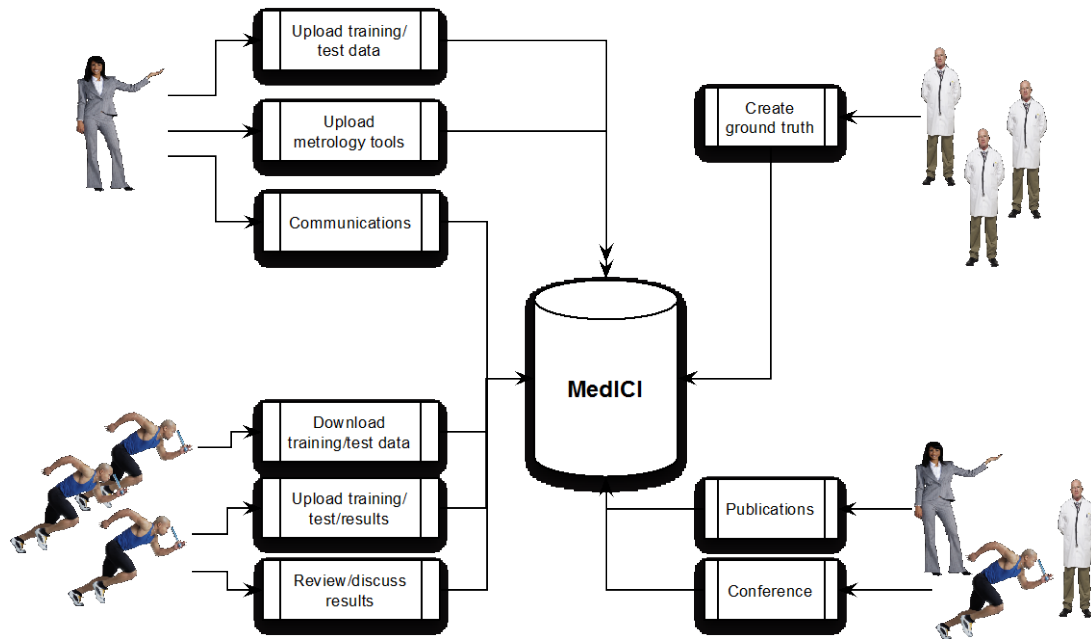
Where are challenges conducted?

- At annual conferences sponsored by scientific societies
 - The Medical Image Computing and Computer Assisted Intervention Society (MICCAI) has held “grand challenges” since 2007
 - International Symposium on Biomedical Imaging (ISBI)
 - SPIE
- On commercial platforms
 - Kaggle
 - TopCoder
 - Sage/Synapse
- Within Organizations such as Quantitative Imaging Network (NCI) and Quantitative Imaging Biomarker Alliance (RSNA)

How are challenges typically conducted?

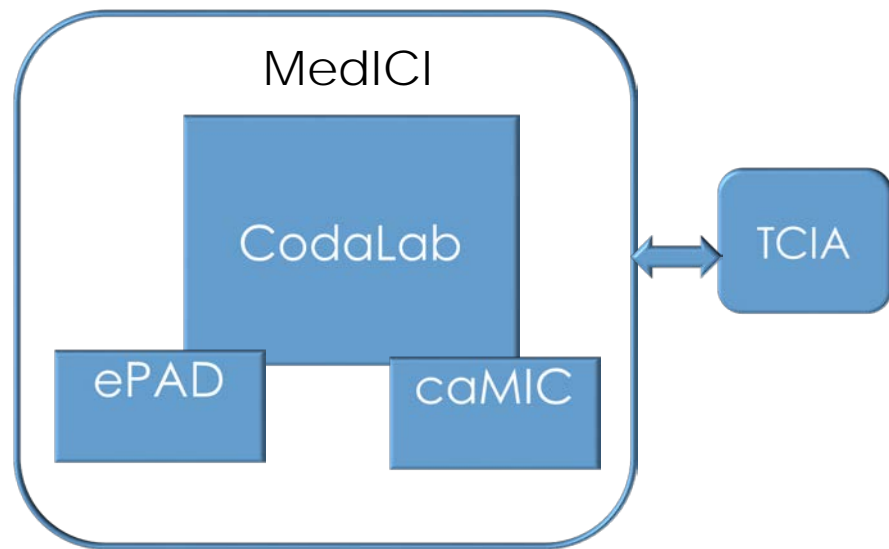
- ▣ Organizers identify a "challenge"
- ▣ Organizers generate/identify "ground truth" for training and test data
- ▣ Organizers define evaluation metrics
- ▣ Organizers announce challenge
- ▣ Interested participants register and download data
- ▣ Participants apply their methods to the training data and obtain results. If satisfactory, participants apply methods to test data
- ▣ Upload results test data results
- ▣ Results made available to community ("leader board"), perhaps at conference
- ▣ Prizes can be awarded

System overview



MedICI: a challenge management software suite

- Goals:
 - Facilitate challenge organization
 - Facilitate challenge participation
 - Support imaging (radiology, pathology) and genomic challenges
 - Statistical analysis
 - Visualization of results
 - Open source, flexible
 - Support integration with TCIA/TCGA



CodaLab

- Open source project created by Microsoft Research and now owned by OuterCurve Foundation
- Worksheets and Competitions
- Hosted on github
- Based on Python/Django framework
- Developed based on work from machine learning community
- Good for user management
- Hosted version available at CodaLab.org

Challenge organization



Front page

MICCAI 2015 CBTC

[Worksheets](#)

[Competitions](#)

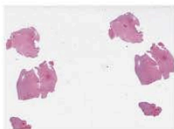
[Help](#)



[Sign In](#)

Competitions

Search



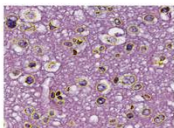
Combined Radiology and Pathology Classification

Organized by artem

This challenge is to evaluate the performance of automated classification algorithms when information from two types of imaging data – Radiology ...

Jul 01, 2015-*No end date*

2 participants



Segmentation of Nuclei in Pathology Images.

Organized by artem

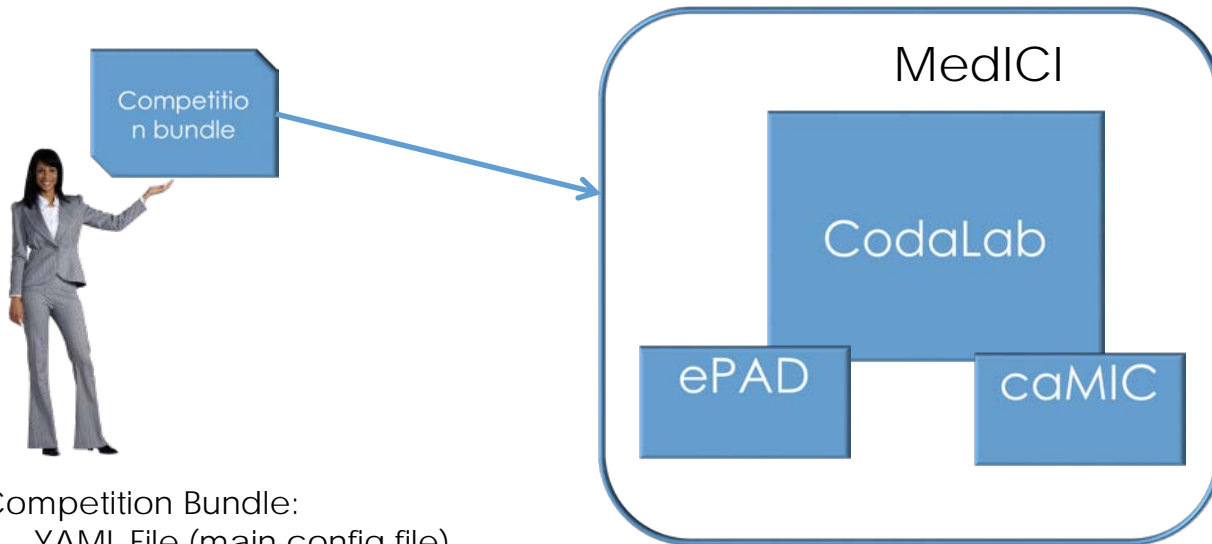
This challenge is to evaluate the performance of algorithms for detection and segmentation of nuclear material in a tissue image.

Jul 01, 2015-*No end date*

3 participants

[Learn how to create your own competition.](#)

Setting up a competition by an organizer



Competition Bundle:

- YAML File (main config file)
- HTML files describing various pages
- Data (optional at this stage)
- Ground truth (optional at this stage)
- Phases (optional at this stage)
- Scoring Metrics(optional at this stage)
- Enter name of TCIA Shared list in competition setup (optional at this stage)
- All setup is done in CodaLab, no specific setup of ePAD and caMIC is needed

Organizing a challenge

- Upload basic “bundle”
 - Include yaml config file
- Upload evaluation
 - Ground truth
 - Evaluation program (executable, python script)

```
competition.zip
|- competition.yaml
|- data.html
|- evaluation.html
|- logo.jpg
...
```

Here are the contents of the reference.zip file:

```
reference.zip
|- answer.txt (Contains: 3.14159265359)
|- metadata   (Contains: This is the authoritative result.)
```

Here are the contents of the program.zip file:

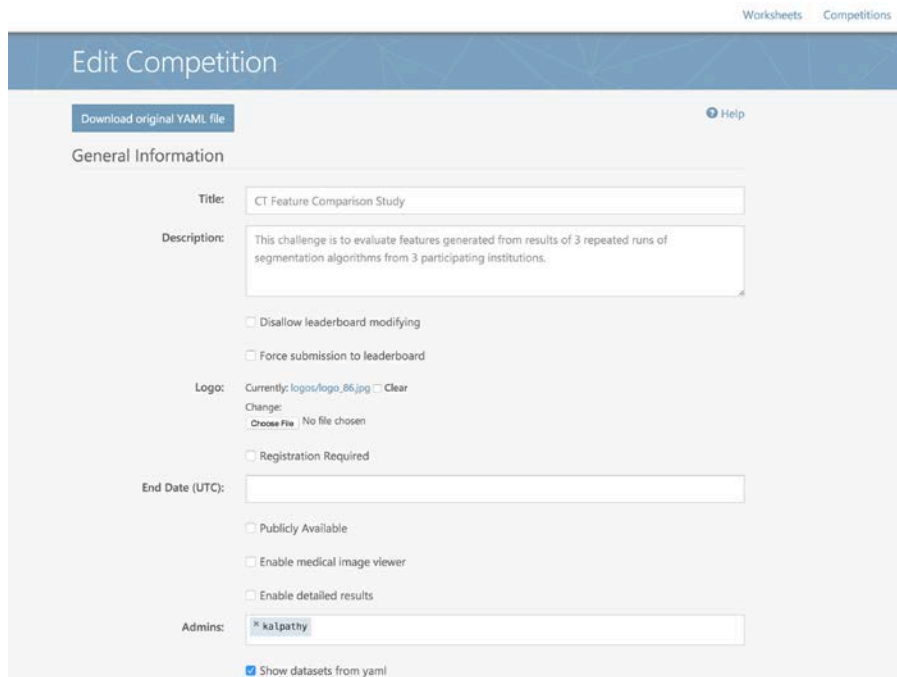
```
program.zip
|- evaluate.py (The actual evaluation code to run)
|- metadata    (Syntax and information needed to run)
|- readme.txt  (Contains notes about the evaluation program)
|- setup.py    (Enables py2exe to build a windows executable of the evaluate.py script)
|- Supporting modules and libraries (if required).
```

The program.zip metadata file contains command syntax to use, along with a short description:

```
command: python $program/evaluate.py $input $output
description: Example competition evaluation program.
```

Editing a challenge

- Can be edited through web interface



The screenshot displays the 'Edit Competition' web interface. At the top, there are links for 'Worksheets' and 'Competitions'. The main heading is 'Edit Competition'. Below this, there is a button labeled 'Download original YAML file' and a 'Help' icon. The 'General Information' section contains the following fields and options:

- Title:** A text input field containing 'CT Feature Comparison Study'.
- Description:** A text area containing 'This challenge is to evaluate features generated from results of 3 repeated runs of segmentation algorithms from 3 participating institutions.'
- Disallow leaderboard modifying:** An unchecked checkbox.
- Force submission to leaderboard:** An unchecked checkbox.
- Logo:** A section showing 'Currently: logos/logo_86.jpg' with a 'Clear' button. Below it, a 'Change:' label is followed by a 'Choose File' button and the text 'No file chosen'.
- Registration Required:** An unchecked checkbox.
- End Date (UTC):** An empty text input field.
- Publicly Available:** An unchecked checkbox.
- Enable medical image viewer:** An unchecked checkbox.
- Enable detailed results:** An unchecked checkbox.
- Admins:** A text input field containing 'kalpathy' with a dropdown arrow on the left.
- Show datasets from yaml:** A checked checkbox.

Publish/unpublish

- Challenge can be created without making it visible to public
- Only other organizers can view it

The screenshot shows a web interface for managing competitions. At the top, there are tabs: "Competitions", "Competitions I'm In", "Competitions I'm Running", and "My Datasets". Below these is a "Create Competition" button. The main area displays a list of four competitions, each with a thumbnail image, title, organizer, description, status, date, and participant count. The status bar for each competition includes buttons for "Delete", "Edit", "Publish", "Unpublish", "Participants", "Submissions", and "Leaderboard".

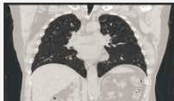
Image	Title	Organized by	Description	Status	Date	Participants
	Spine Vertebrae Detection and Localization	Organized by artem	The goal of this competition is to detect vertebrae visible in the input CT scan and estimate their locations accurately.	Publish	Nov 01, 2013-No end date	1 participant
	Stanford Moist run Challenge	Organized by artem	This is a test competition for Stanford Moist Run	Unpublish	Mar 05, 2015-No end date	9 participants
	Stanford Moist Run Challenge New	Organized by artem	This is a test for Stanford Moist Run Challenge	Unpublish	Apr 01, 2015-No end date	1 participant
	QIN Lung CT Feature Comparison Study	Organized by kirbyju	The goal of the CT Feature Comparison Study is to evaluate features of segmentations generated from the QIN Lung Segmentation ...	Publish	Mar 05, 2015-No end date	6 participants

Integration of caMicroscope with CodaLab

medici4.cloudapp.net

MEDICI

Worksheets Competitions Help artem

 caMicroscope integration DEMO

Organized by artem - Current server time: April 4, 2016, 3:03 p.m. UTC

► Current End


Test Competition Ends

March 17, 2016, midnight UTC Never

Learn the Details Phases Participate Results

Test

Phase description
None

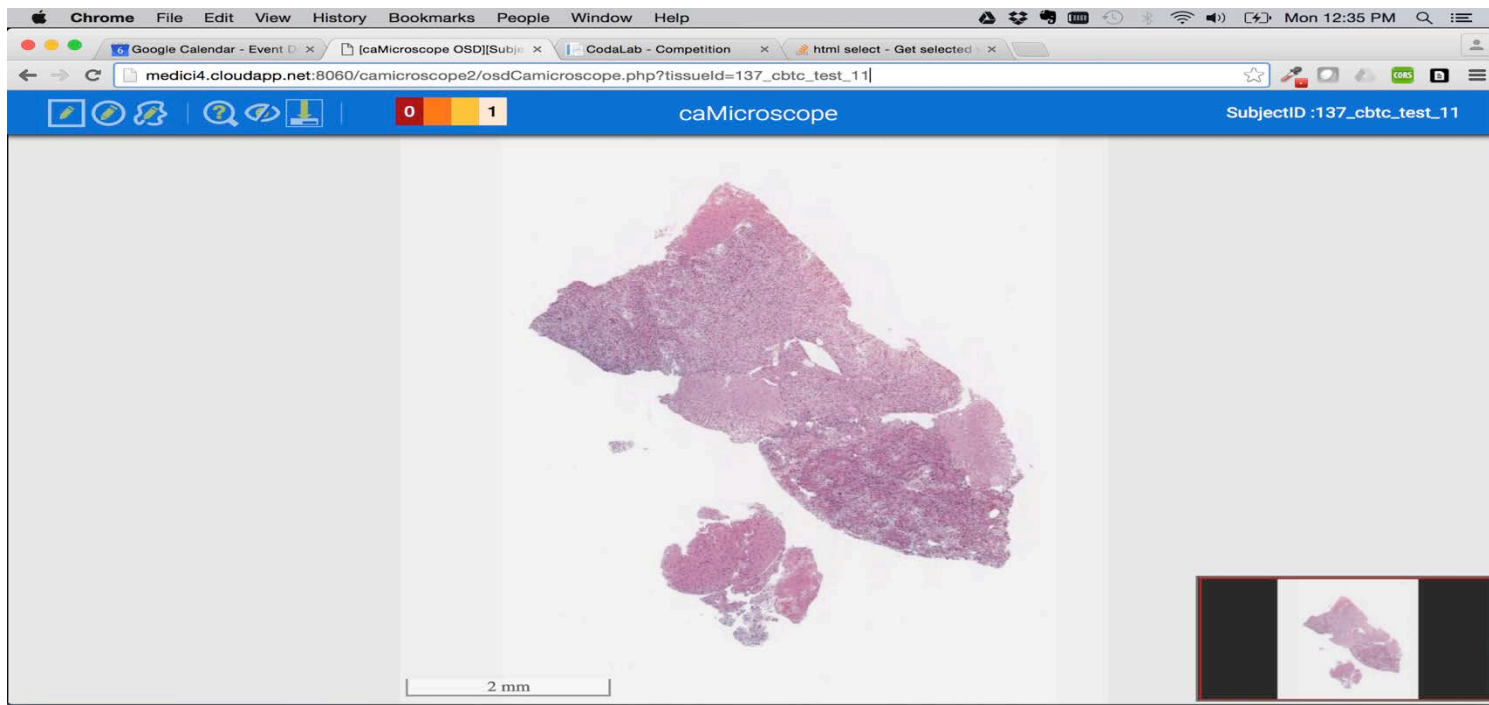
 Download CSV Download all submissions on leaderboard

Results

	User	DIFF	Image Viewer
1	artem	1.000 (1)	View results

About Forum Survey Privacy and Terms

Integration of caMicroscope with CodaLab


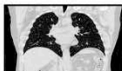
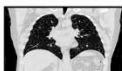

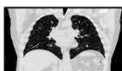
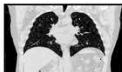


ePAD integration

medici4.cloudapp.net:8000/competitions/

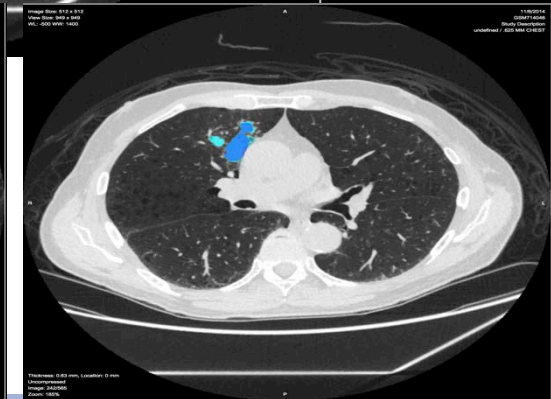
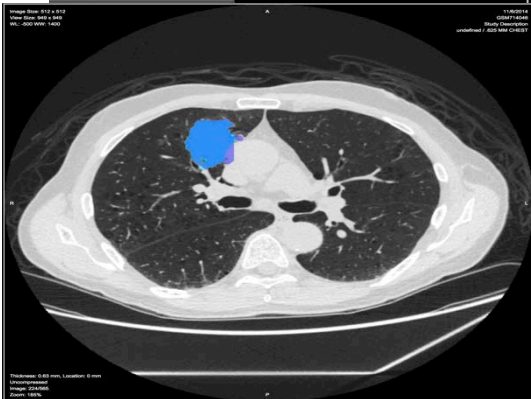
Apps Bookmarks dataVis ShortcutMapper - Ki ctspedia.org/wiki/pu Quatitative Imaging 6.869 Advances in C HART BIRN Pathology Wor https://www.nmr.mgl Capture Reference Design and analysis Other Bookmarks

MEDICI Worksheets Competitions Help kalpathy

	Example Competition for pi value evaluation Organized by tester This is a competition to test the competition bundle system. It should be able to create a competition from this ...	Mar 05, 2015- <i>No end date</i> 3 participants
	Stanford Moist run Challenge Organized by artem This is a test competition for Stanford Moist Run	Mar 05, 2015- <i>No end date</i> 8 participants
	Standord Moist Run Challenge New Organized by artem This is a test for Stanford Moist Run Challenge	Apr 01, 2015- <i>No end date</i> 1 participant
	QIN Lung CT Feature Comparison Study Organized by kirbyju The goal of the CT Feature Comparison Study is to evaluate features of segmentations generated from the QIN Lung Segmentation ...	Mar 05, 2015- <i>No end date</i> 4 participants
	Standord Moist Run Challenge Demo Organized by artem Stanford Moist Run Challenge Demo	Apr 01, 2015- <i>No end date</i> 1 participant
	ePAD integration demo Organized by artem Test of ePAD integration	Aug 15, 2015- <i>No end date</i> 2 participants

ePAD on Azure VM using Docker





TCIA

Ground Truth Annotation

TCIA Shared List:

Upload to ePAD

Add Annotator

Create annotator accounts in ePAD

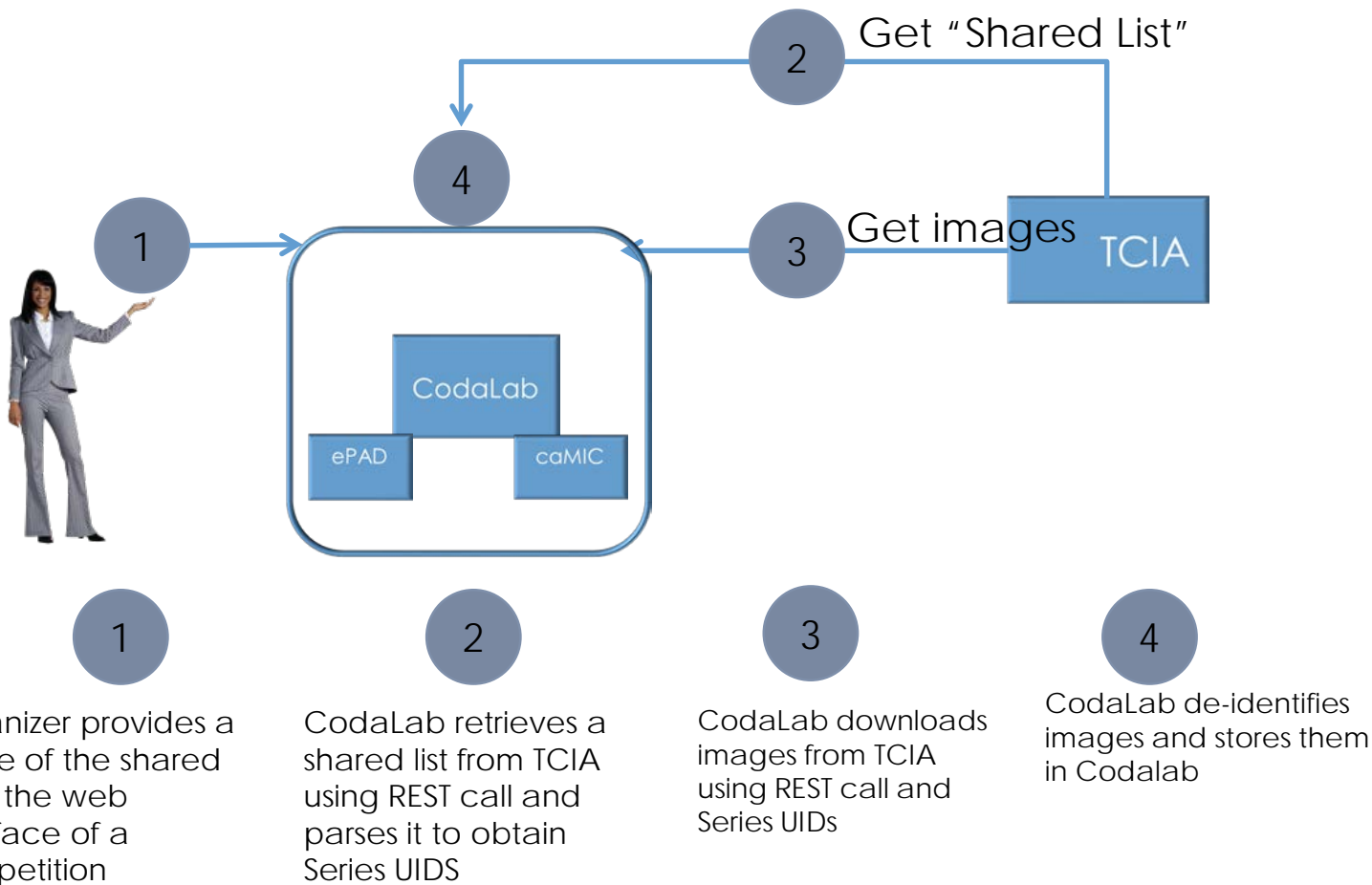
Send email to all annotators

■ Integration with TCIA

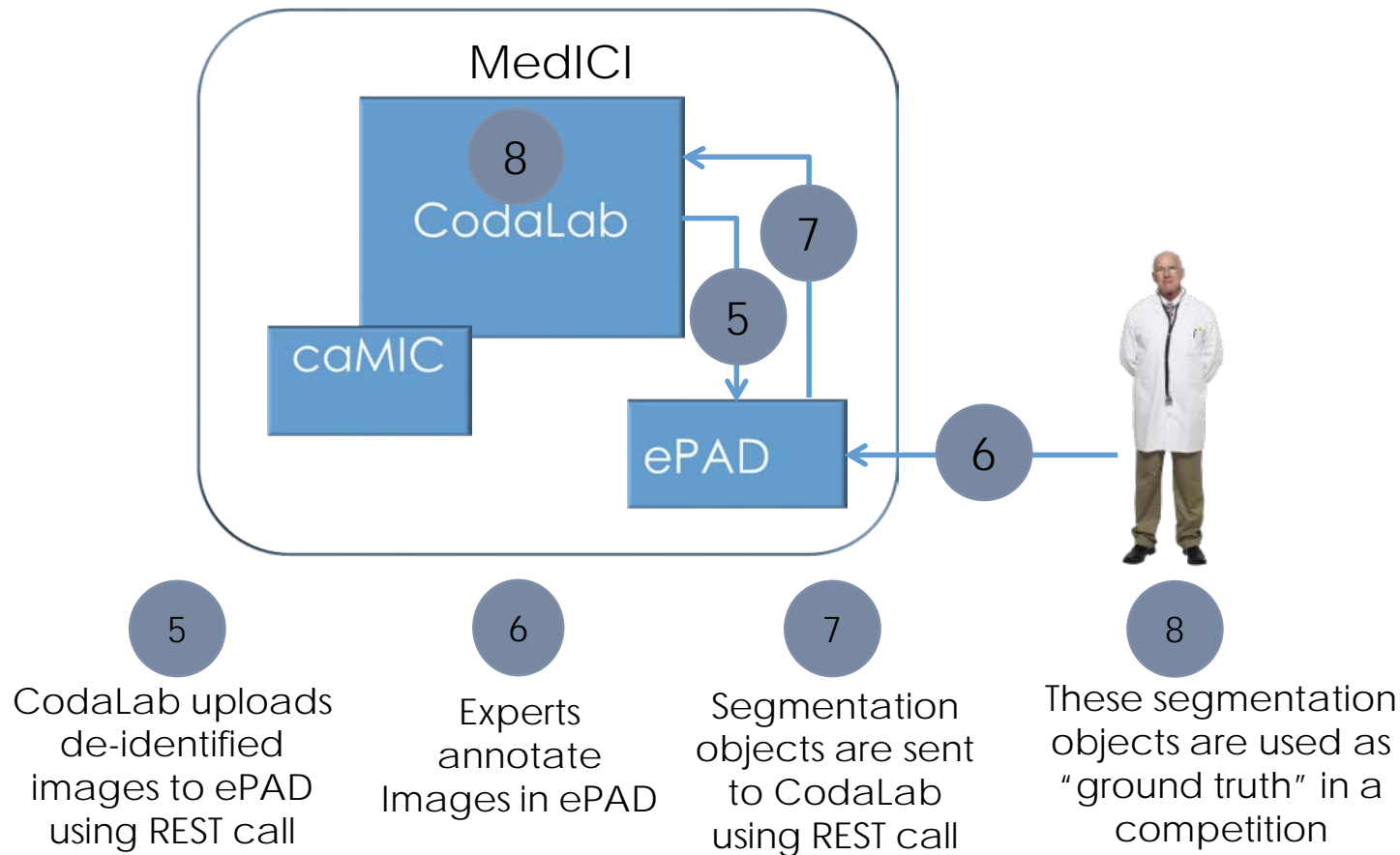
Integration with TCIA

```
amamonov@bryce: ~/Dataset-Bel...GE_CT_Archive_9407_scans_ZIPs  less  azureuser@cbibop: ~  
# Build an example competition  
---  
title: Standard Moist Run Challenge Demo  
description: Stanford Moist Run Challenge Demo  
image: logo.jpg  
has_registration: False  
end_date:  
enable_detailed_results: True  
tcia_shared_list: my_tcia_list  
annotators:  
  1:  
    first_name: Artem  
    last_name: Mamonov  
    email: artmnv@gmail.com  
  2:  
    first_name: Jayashree  
    last_name: Kalpathy  
    email: kalpathy@nmr.mgh.harvard.edu  
  3:  
    first_name: Daniel  
    last_name: Rubin  
    email: dlrubin@stanford.edu  
html:  
  overview: overview.html  
  evaluation: evaluation.html  
  terms: terms_and_conditions.html  
  data: data.html  
phases:  
  1:  
    phasenum: 0  
:
```

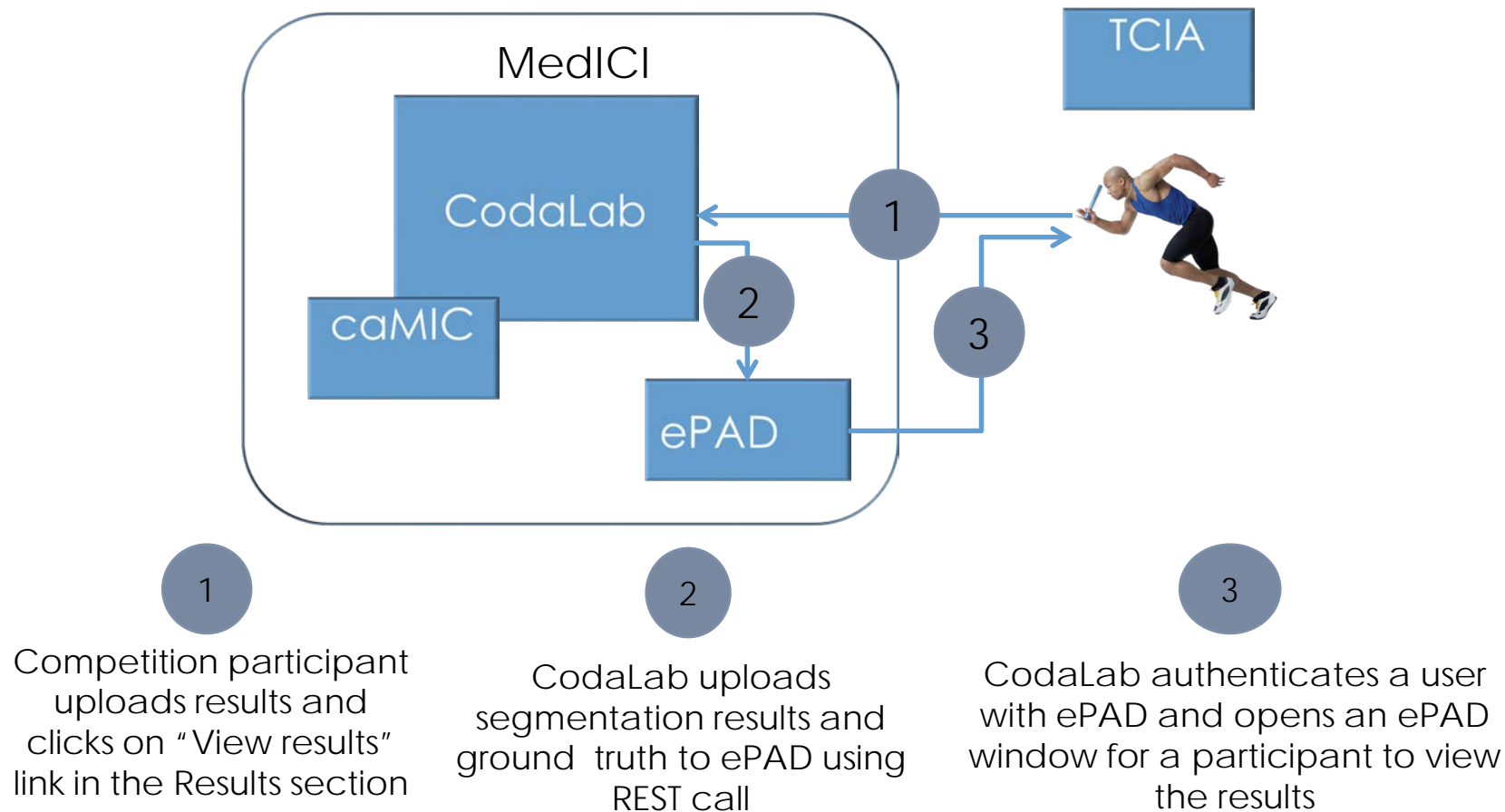
Creating Ground Truth



Creating Ground Truth, cont'd




Visualizing results



Submitting results for automatic evaluation

MICCAI 2015 CBTC

Worksheets Competitions Help  cbtc.organizing.committee

Learn the Details Phases Participate Results Forums

Get Data

Submit / View Results

Training Test

Phase description
None

Click the Submit button to upload a new submission.

Optionally add more information about this submission

Submit

Here are your submissions to date (✓ indicates submission on leaderboard):

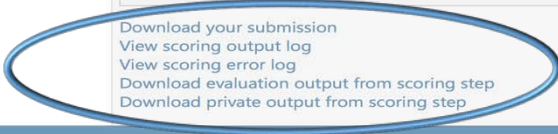



	FILENAME	SUBMISSION DATE	STATUS	✓
1	reference.zip	09/24/2015 17:00:04	Finished	✓

Description:

- Download your submission
- View scoring output log
- View scoring error log
- Download evaluation output from scoring step
- Download private output from scoring step

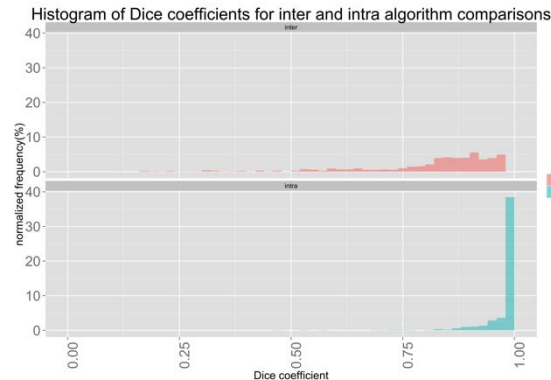
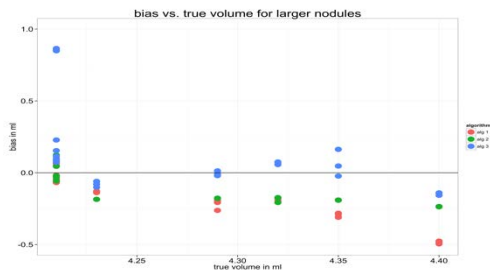
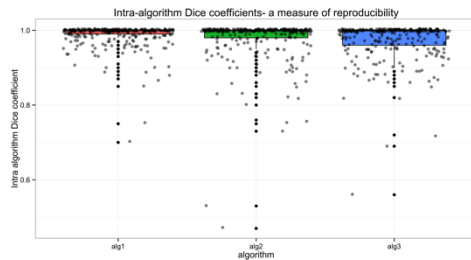
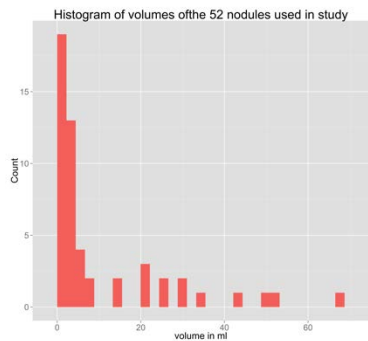
Submit to Leaderboard

Based on CodaLab



Statistical package integration

Integration with R for statistical analysis and interactive data visualization




Leaderboard


miccai.cloudapp.net

H CSS Web S... 2.5 Using A... Web Develo... 8. Errors an... 11. Recursi... T-Mobile Ce... Python Tuto... Rabix My Competition Goo... +

MICCAI 2015 CBTC

Worksheets Competitions Help  cbtc.organizing.committee

Organized by cbtc.organizing.committee - Current server time: Oct. 28, 2015, 5:48 p.m. UTC



Current

Training

Sept. 21, 2015, midnight UTC

Next


Test

Dec. 1, 2015, midnight UTC

Learn the Details Phases Participate Results Forums

Training Test

Phase description
None

 Download CSV

Download all submissions on leaderboard

Results

	User	Submit User	File name	SCORE
1	cbtc.organizing.committee	Czakon	Output.txt.zip	0.750 (1)
2	cbtc.organizing.committee2	joelcarlson	JoelCarlson_results.zip	0.600 (2)
3	cbtc.organizing.committee3	yson1723	test_result_SONG.zip	0.500 (3)

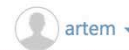
Based on CodaLab

Forum

QINLABS

My Competitions

Help



emails

QIN BMMR - Breast MRI Metrics of Response Forum

[Go back to competition](#)

[Start a new topic](#)

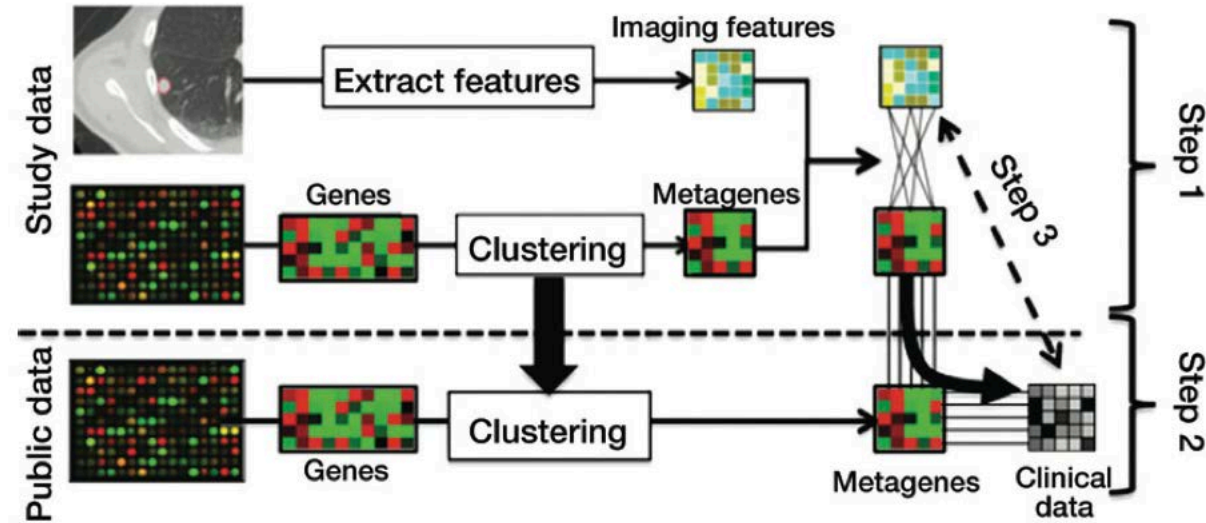
Title	Started by	Date created	Latest post	Posts
QIN_BMMR_Training DCEonly_HRknown	dIrubin	Jun 28, 2016	6 days, 8 hours	4
ROI (Lesion Segmentation)	acheerla	Jun 27, 2016	1 week	2
Tumor Segmentation	acheerla	Jun 28, 2016	1 week	2
Metadata / directories: too many or not enough T1 directories	FiReTiTi	Jun 17, 2016	1 week, 5 days	6
44 exams not in shared list	UC	Jun 14, 2016	2 weeks, 5 days	5
Inter-regimen cases 1224 and 1230 slices mis-aligned between pre and post images	UC	Jun 14, 2016	2 weeks, 5 days	4
Test set: problem in number of pre and post contrast files	baishali	May 28, 2016	3 weeks, 1 day	5
question about number of result submissions for test set	UC	May 27, 2016	1 month	2
Information about the directories downloaded	FiReTiTi	May 07, 2016	1 month, 2 weeks	5
DCE exams with unexpected number of image files	UC	May 13, 2016	1 month, 2 weeks	2
BR-15 MR1 seems corrupted	UC	May 10, 2016	1 month, 3 weeks	2

er

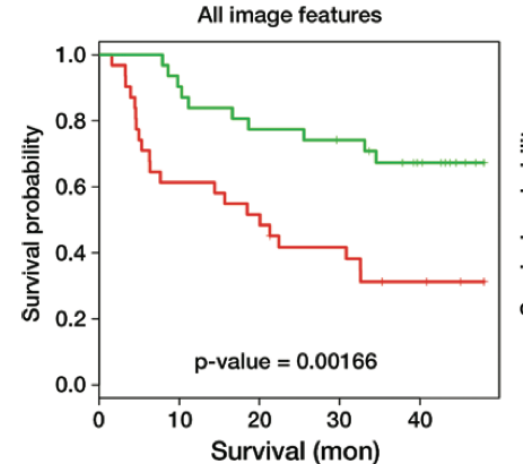
Radiogenomics (Radiopathagenomics...)

■ Radiomics

■ Genomics



■ Clinical prediction/out comes



How are image features derived?

- ▣ Clinicians provide semantic input
 - ▣ Binary
 - ▣ Scale
 - ▣ e.g. LIDC, VASARI
- ▣ Mathematical descriptors (based on tumor segmentation)
 - ▣ Texture, shape, intensity, SIFT...
 - ▣ Segmentation can be manual or automatic

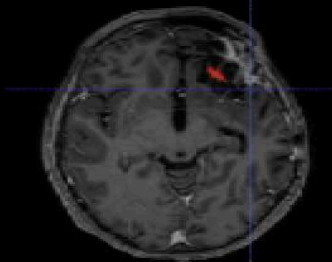
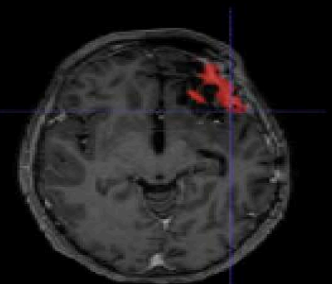
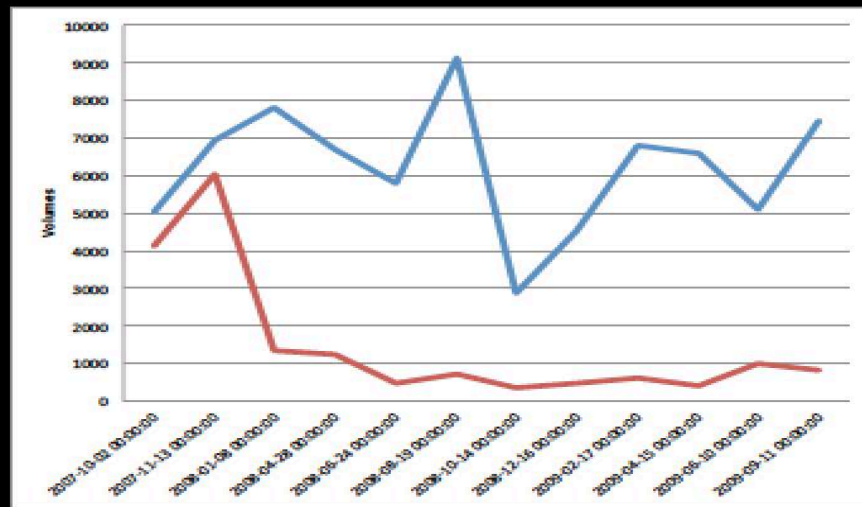
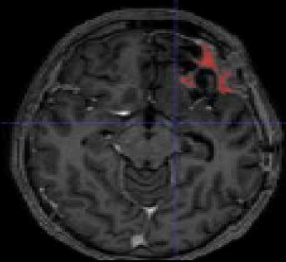
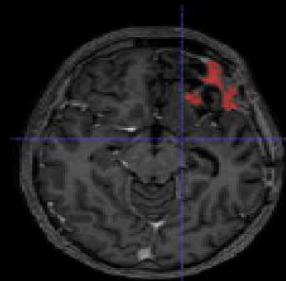
Barriers to scaling up radiomics/radiogenomics studies

- ▣ Access to large datasets
 - ▣ TCIA/TCGA have lowered the barrier
- ▣ Interoperability of annotations
 - ▣ AIM/DICOM-SEG/DICOM-SR
 - ▣ Controlled terminology (e.g. RadLex)
- ▣ Generation of imaging annotations
 - ▣ Segmentation often basis of feature extraction
 - ▣ Need scalable means of generating segmentations and annotations

Human annotations

- Challenging to scale up process of human annotations
 - Although very successfully demonstrated for TCIA datasets, not typical
- Time consuming and subject to lack of inter-rater agreement
 - LIDC demonstrated some of these issues and the need for adequate QA and process well
 - "71 lesions received "nodule >3 mm" marks from at least one radiologist; however, all four radiologists assigned such marks to only 24 (33.8%) of these lesions. "
- Generate consensus for experts?
- Explore use crowd sourcing to get more annotations?
 - Cell Slider (CRUK) demonstrated the power for "citizen science"

Example case: Would the choice of reader affect the response assessment/radiomics features?



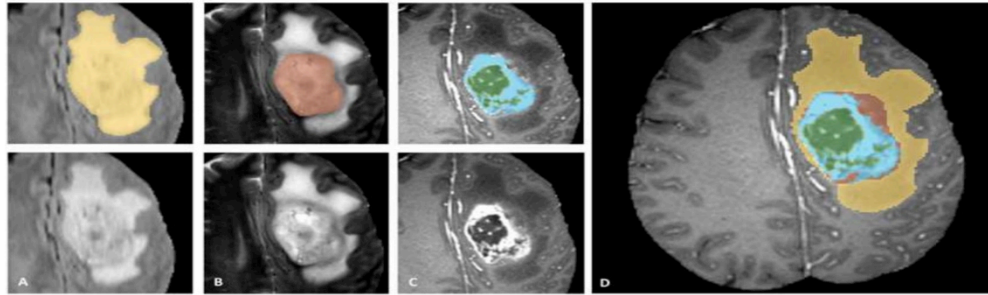
Why automated?

- Manual annotations are
 - Subjective
 - Time consuming
 - 1D or 2D typically
 - Expensive
- Computational techniques and hardware are evolving rapidly
- Can extract more characteristics than just size

MICCAI Brain tumor segmentation challenge

- Organized by academics, NCI
 - Dozens have participated at MICCAI and used data subsequently
 - 100+ citations since 2015
- Being run since 2012
 - Had 3-4 experts label volumes
- In 2014, added data from TCIA-GBM collection
 - Test labels were machine generated (not optimal)
- 2015 added expert labels
 - 2 experts
- 2016 -longitudinal data, additional datasets
- "deep learning" demonstrating excellence performance

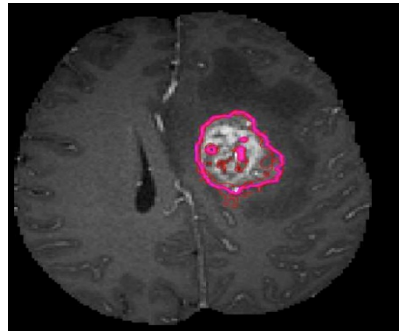
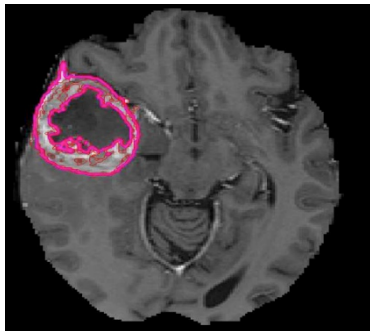
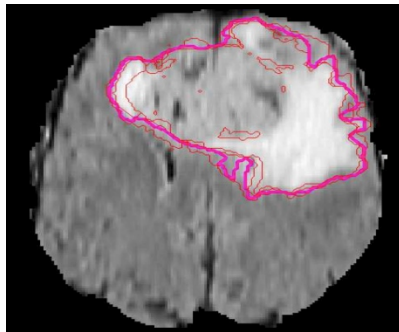
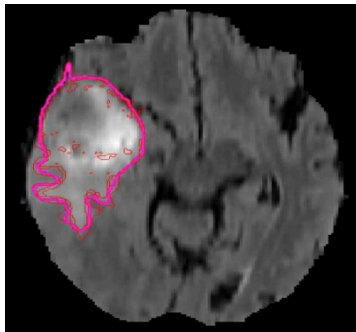
Manual Segmentation (4 labels)



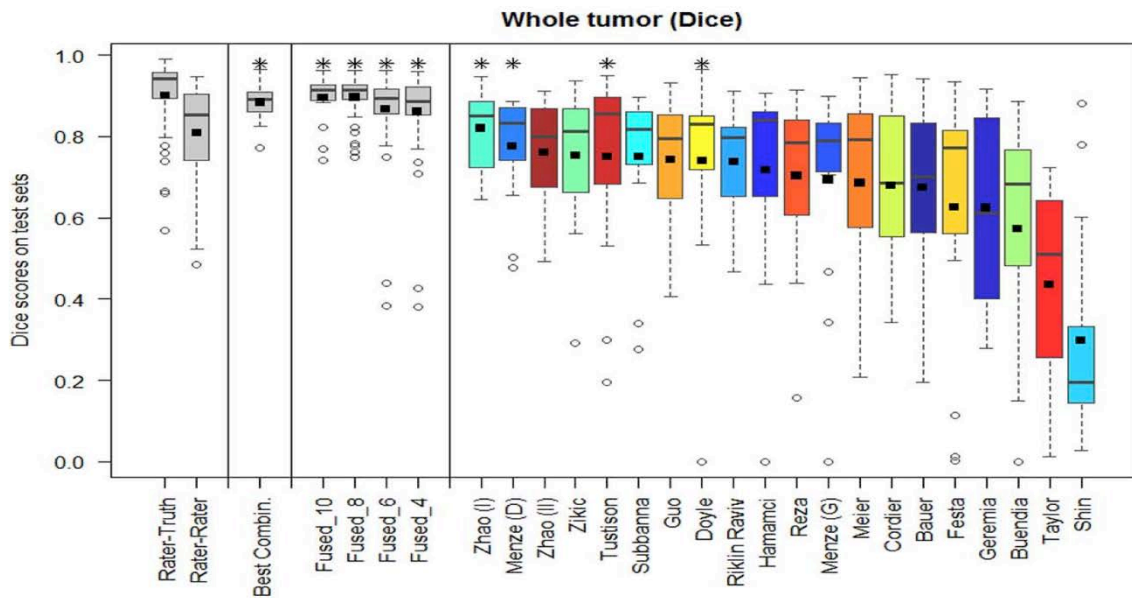
- Manual annotation through expert raters:
- the whole tumor visible in FLAIR (A),
- the tumor core visible in T2 (B),
- the enhancing active tumor visible in T1c (blue),
- surrounding the cystic/necrotic components of the core (green) (C). The segmentations are combined to generate the final labels (D): edema (yellow), non-enhancing solid core (red), active core (blue), non-solid core (green).



Example slices with output of segmentation algorithms



Results of computer generated labels



Improvement by combining best methods

BraTS challenge output

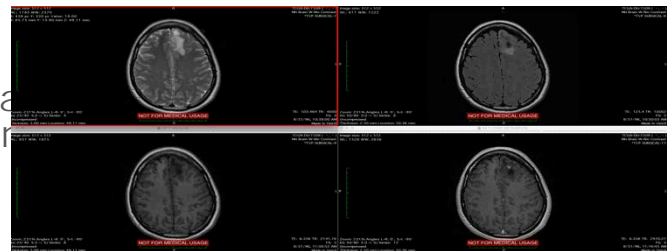
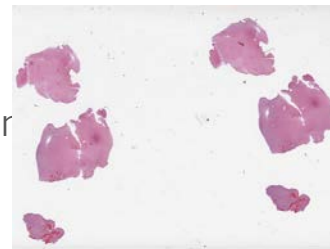
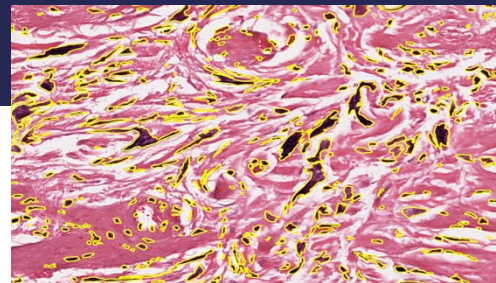
- Over 100 research publications (conference papers, manuscripts)
- Robust open-source segmentation methods for MR brain tumor segmentation
- Expert segmentations of ~20-50 cases in TCIA
- Consensus segmentations of all brain tumor cases back to TCIA (in 2016)
- Continued resource to the community

Challenge organizer perspective

- Potential for being a great resource to the community
 - Benchmark to compare improvements in field
 - Get participation from non-experts in field
 - Opportunity to build community
- However, can be very time consuming
 - Mixed academic value (not necessarily research but potential for large number of citations)
 - Communication is large factor in success of challenge
 - Infrastructure and resources are critical

MICCAI CBTC 2015

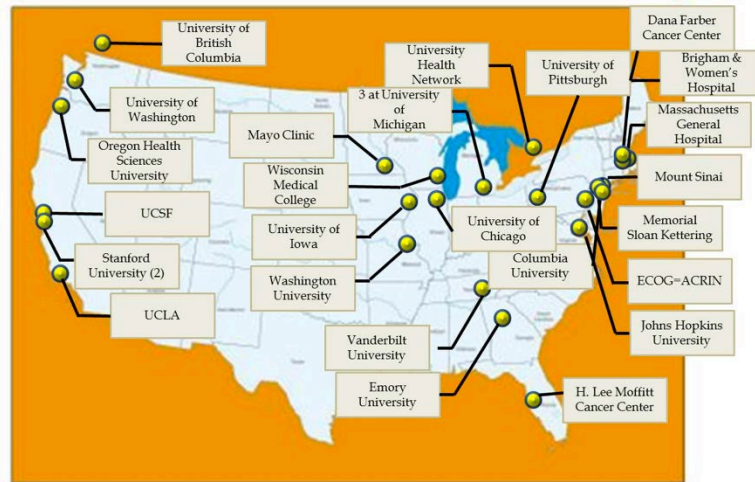
- Nuclear segmentation in pathology images:
 - image tiles from whole slide tissue images
 - For training set nuclei in each tile have been manually segmented
 - <http://miccai.cloudapp.net:8000/competitions/37>
- Joint radiology/pathology classification challenge:
 - Classify Low grade glioma cases from TCIA (LGG) into Oligoder and Astrocytoma
 - Training data consists of “ground truth” classification
 - Training set consists of 32 cases.
 - Test set consists of 20 cases.
 - Combined Radiology and Pathology Classification (<http://miccai.cloudapp.net:8000/competition>)



Ongoing and upcoming challenges

- MICCAI 2016
 - Joint radiology pathology challenge
 - Radiomics challenge (head and neck)
 - Prize: publication!
 - Radio-pathomics challenge (liver mets)
 - Breast CAD (mammography)
- Digital pathology challenge
- QINLabs
 - Breast challenge
 - CT feature challenge
 - ~6-10 new challenges planned for next cycle

Quantitative Imaging Network

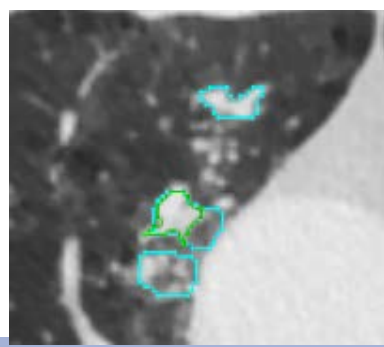
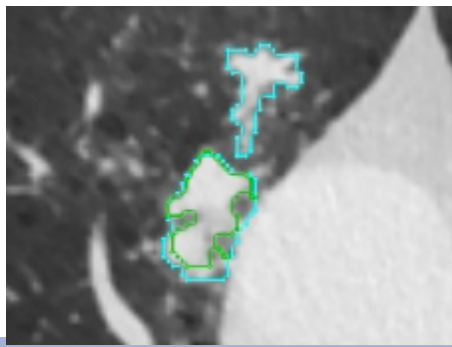
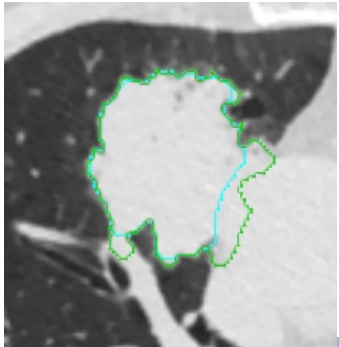
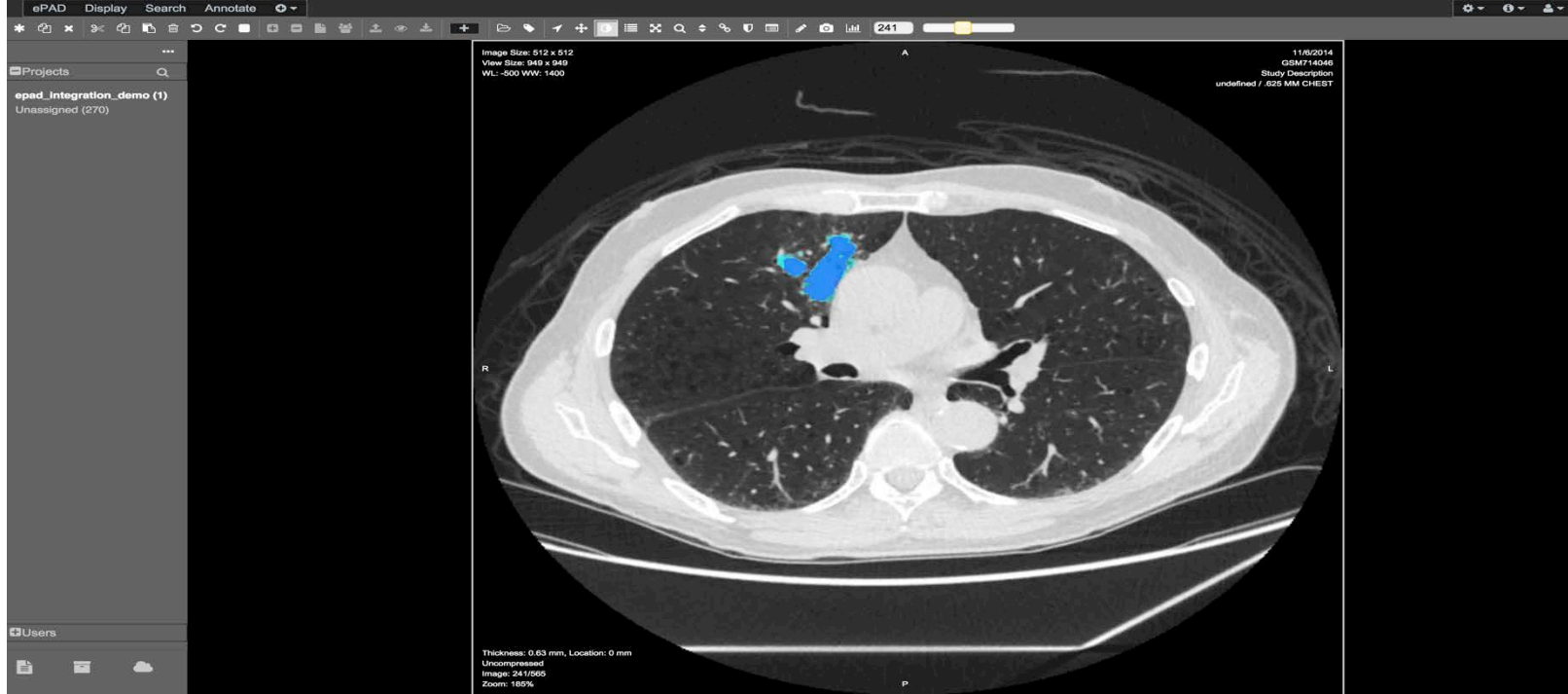


The network is designed to promote research and development of **quantitative imaging** methods for the **measurement of tumor response** to therapies in clinical trial settings, with the overall goal of facilitating clinical decision making.

Grant arose from the collaborative projects as part of QIN

QIN challenge: Lung nodule segmentation

- 52 lesions from 41 CT studies
- 33 to μl to 57 ml and demonstrated a diversity of shapes from round through spiculated.
- Three algorithms, each submitted 3 repeat segmentations per nodule

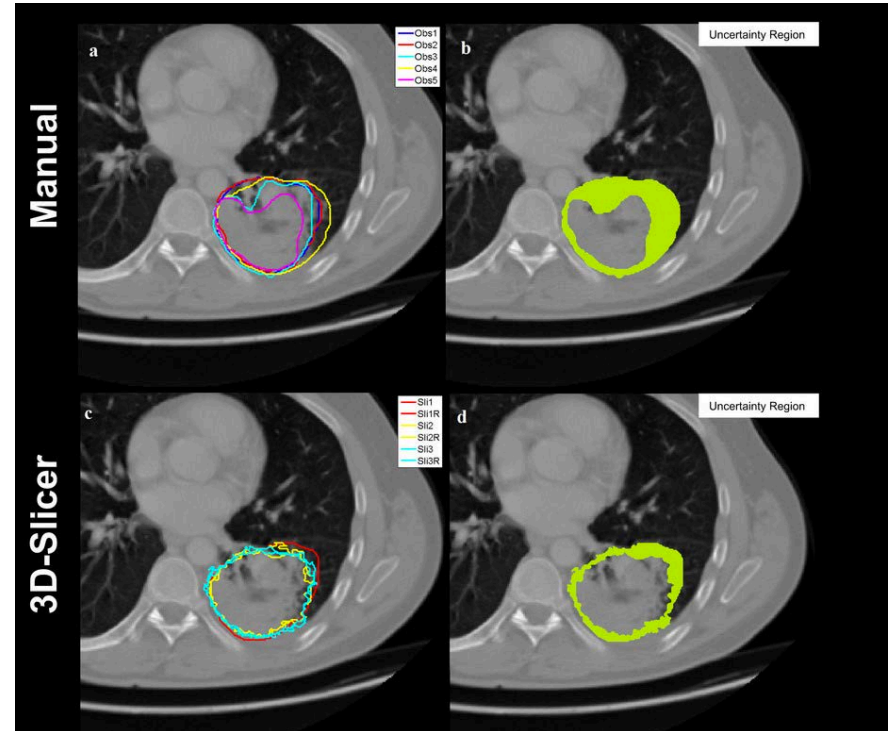


Output of challenge

- CT volumes in TCIA (existing collections)
- Segmentations in TCIA (in DICOM-SEG format)
- Segmentations can be used for radiomics and radiogenomic studies (underway in QIN)
 - Stability of features
 - Correlation between features
 - Identify “habitats” or sub volumes based on features

Features can be sensitive to segmentation


- Reduced uncertainty with machine assisted segmentation



QIN Feature comparison challenge

- “Radiomics” pipelines allow for the quantification of imaging characteristics
 - Can be used in outcomes research
 - Radiogenomics
 - However, features can be sensitive to segmentation
- 8 QIN sites participated
- 10-300 features per site

CT Feature challenge



CT Feature Comparison Study

Organized by artem - Current server time: June 9, 2016, 11:42 p.m. UTC

► Current

End

Evaluate

Competition Ends

Oct. 16, 2015, midnight UTC

Never

Learn the Details

Phases

Participate

Results

Overview

Evaluation

Terms and Conditions

Welcome!


Background: As part of the CT Segmentation Challenge, we evaluated the results of 3 repeated runs of segmentation algorithms from 3 participating institutions. Various measures were developed to compare the results of each run of a specified algorithm, comparison to ground truth (if known) as well as pair-wise comparison of algorithms from each institution.

Goal: The goal of the CT Feature Comparison Study is to evaluate features generated from each of these segmentations for repeatability between repeated runs of each algorithm, and reproducibility across segmentation algorithms.

We will use the 5 collections of DICOM CT images that were used for the "moist run" from the following sources: CUMC_FDA Phantom, Moffitt Cancer Center, Rider, Stanford, and UDC. The collection consists of volumes of 52 nodules and 9 segmentations of each nodule. These collections and full descriptions are available at NCI TCIA under "QIN multi-site collection of Lung CT data with Nodule Segmentations" with DOI <http://dx.doi.org/10.7937/K9/TCIA.2015.1BUVFR7>. (If you have issues reading DICOM-SEG files, you can find segmentations in NIfTI formats at <https://ncipub.org/projects/ctfeature> in new, "nii_segmentations" subdirectory (under Files)).

We will expand on the work of the segmentation challenge as well as the work on feature reproducibility on test-retest data and numerous papers on various measures for feature stability measures, including repeatability and reproducibility. Note that R-scripts will be run on C-BIBOP to compute commonly used metrics. We will coordinate these submissions with separate QIN effort to compare and harmonize features across sites.

Each participating institution would upload a comma separated value file containing their feature computations for each of the 450 segmentations (5*10*3*3) to the NCIP HUB using a standard format (likely csv). Sample output file is provided via NCI-HUB "CT feature



CT Feature Comparison Study

Organized by artem - Current server time: Jan. 12, 2016, 3:03 p.m. UTC

► Current

End

Evaluate

Competition Ends

Oct. 16, 2015, midnight UTC

Never

Learn the Details

Phases

Participate

Results

Evaluate

Phase description

None

Download CSV

Download all submissions on leaderboard

Results				
	User	N features	Avg Repeatability	Avg Reproducibility
1	JohannaUthoff	304.0	0.985 (1)	0.845 (2)
2	ivan.yeung@mmp.uhn.on.ca	10.0	0.983 (2)	0.675 (8)
3	lhadjilski	49.0	0.974 (3)	0.863 (1)
4	mmcnittgray	15.0	0.972 (4)	0.834 (3)
5	LinLu	71.0	0.947 (5)	0.796 (4)
6	sechegaray	198.0	0.928 (6)	0.722 (6)
7	sechegaray-2	774.0	0.894 (7)	0.748 (5)
8	sechegar	762.0	0.861 (8)	0.713 (7)
9	cherezov	186.0	0.842 (9)	0.616 (9)

Results

Site Selection

Select site

UMICH

CUMC

NStanford

PM

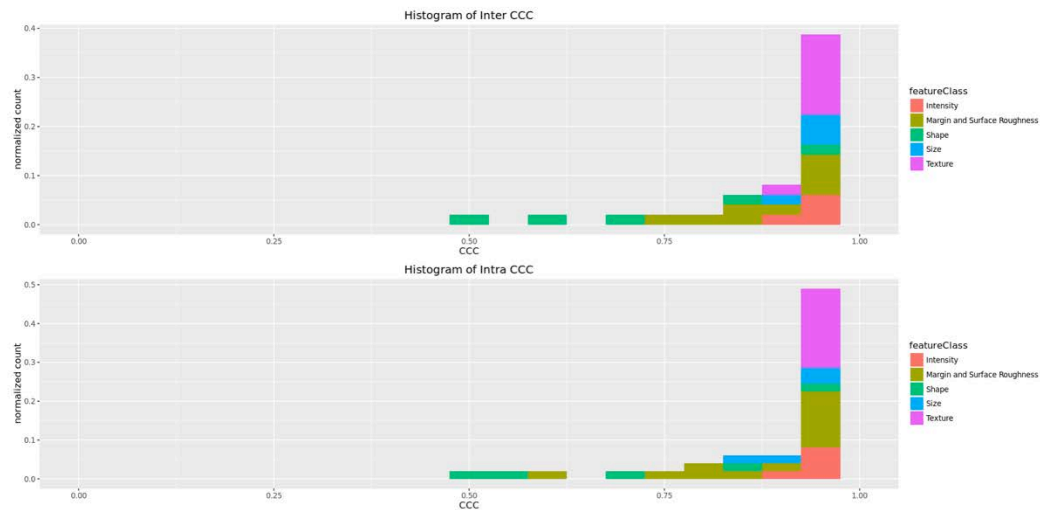
Stanford

UCLA

UMICH

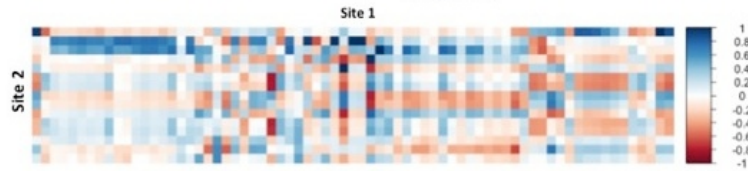
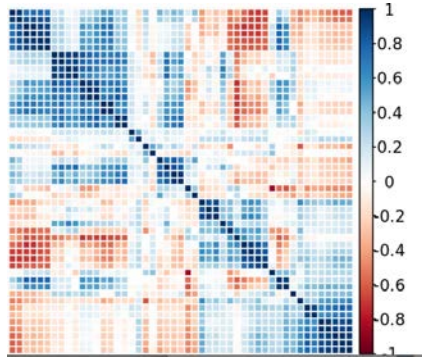
Iowa

and

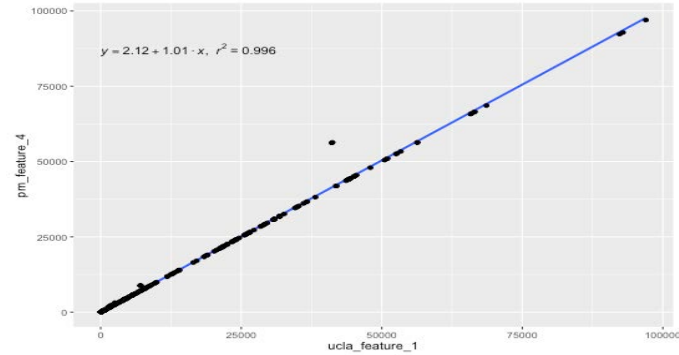


- Feature stability with respect to segmentation

Results



c. Example of heat map of correlation between features submitted between two sites highlighting the high correlation of features across implementation



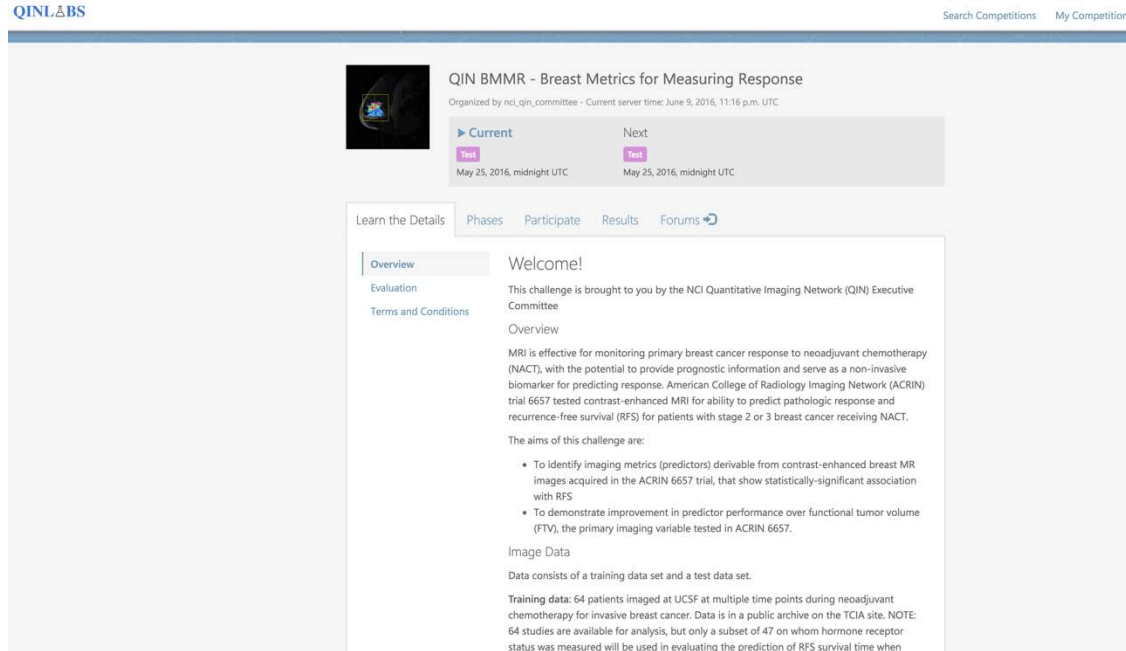
4	feature_4	PM	Size	Volume	3D	no	1
1	feature_1	UCLA	Size	Volume	volume	3D	No

□ Inter and intra-site correlation of features

QIN BMMR challenge (clinical trial data)

- The aims of this challenge are:
 - To identify imaging metrics (predictors) derivable from contrast-enhanced breast MR images acquired in the ACIN 6657 trial, that show statistically-significant association with RFS
 - To demonstrate improvement in predictor performance over functional tumor volume (FTV), the primary imaging variable tested in ACIN 6657.

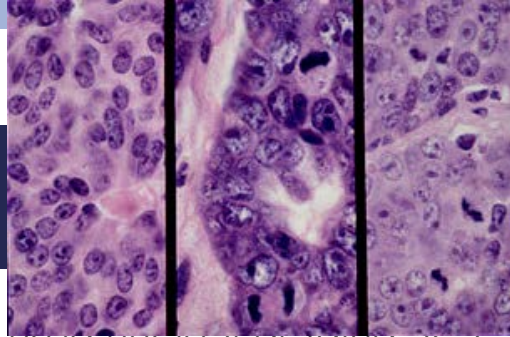
QIN BMMR challenge



The screenshot shows the QIN BMMR challenge website. At the top, the QIN LABS logo is on the left, and links for 'Search Competitions' and 'My Competitions' are on the right. The main header area features a small image of a breast MRI scan on the left. To its right, the title 'QIN BMMR - Breast Metrics for Measuring Response' is displayed, followed by the text 'Organized by nci_qin_committee - Current server time: June 9, 2016, 11:16 p.m. UTC'. Below this, a navigation bar shows 'Current' and 'Next' phases, each with a 'Test' button and a date: 'May 25, 2016, midnight UTC'. A secondary navigation bar includes 'Learn the Details', 'Phases', 'Participate', 'Results', and 'Forums'. On the left side of the main content area, there is a sidebar with links for 'Overview', 'Evaluation', and 'Terms and Conditions'. The main content area begins with a 'Welcome!' message, stating that the challenge is brought to you by the NCI Quantitative Imaging Network (QIN) Executive Committee. It then provides an 'Overview' of the challenge, explaining that MRI is effective for monitoring primary breast cancer response to neoadjuvant chemotherapy (NACT), with the potential to provide prognostic information and serve as a non-invasive biomarker for predicting response. It mentions the American College of Radiology Imaging Network (ACRIN) trial 6657, which tested contrast-enhanced MRI for ability to predict pathologic response and recurrence-free survival (RFS) for patients with stage 2 or 3 breast cancer receiving NACT. The aims of the challenge are listed: to identify imaging metrics (predictors) derivable from contrast-enhanced breast MR images acquired in the ACRIN 6657 trial, that show statistically-significant association with RFS; and to demonstrate improvement in predictor performance over functional tumor volume (FTV), the primary imaging variable tested in ACRIN 6657. The 'Image Data' section states that the data consists of a training data set and a test data set. The 'Training data' section notes that 64 patients imaged at UCSF at multiple time points during neoadjuvant chemotherapy for invasive breast cancer. Data is in a public archive on the TCIA site. NOTE: 64 studies are available for analysis, but only a subset of 47 on whom hormone receptor status was measured will be used in evaluating the prediction of RFS survival time when

- QIN challenge led by Nola Hylton, data from ISPY/ACRIN trial

'tis in the nuclei



- **BACKGROUND:** Pathologists and bioinformaticians have used nuclear grading as a critical part of evaluating the structure of Breast Cancers. Some analysts claim to be able to use nuclear characteristics for prognostication (Axelrod). At the present time, tumor nuclear characteristics are subjectively graded by the pathologists. Many types of GEMM have distinctive nuclear and cytoplasmic characteristics which are readily identified by an experienced pathologist. The challenge is to codify these characteristics and relate them to the biology of the animal.
- **OBJECTIVE:** The objective is to develop and test algorithms capable of distinguishing between the nuclear phenotypes from different mouse genotypes.
- **RESOURCES:** A panel of WSI from four GEMM mouse genotypes
- **CHALLENGE:** Using the nuclear attributes of six GEM models, identify the genotype of unknown HUMAN and MOUSE panels.

Cardiff et al 2001 PMID: 1887859

Lessons Learned

- We need to work together!!
- “our finding **further reinforces the notion that crowd-sourced collaborative competitions are a powerful framework for developing robust predictive models by training an ensemble model aggregated across diverse strategies employed by participants.** ”, Bilal et al, PLOS Computational Biology, 2013
- Ensemble methods (often) outperform individual models
 - “Our experience with the Netflix competition showed that the most successful model is an ensemble of multiple predictors”, Bell et al, SIGKDD Explorations, Vol9
 - “ensemble models trained across multiple user submissions **systematically outperform individual models** within the ensemble”, Bilal et al, PLOS Computational Biology, 2013
- Combining expert knowledge with machine learning can be valuable
 - “**machine learning methods combined with molecular features selected based on expert prior knowledge** can improve survival predictions compared to current best-in-class methodologies”, Bilal et al, PLOS Computational Biology, 2013

Lessons Learned

- Need to agree upon authorship and roles early on.
- Need to make it worth the while to participate in a challenge.
- Incentives (in bioinformatics)
 - Cash Prizes
 - Publications
 - Academic Glory?

Boutros et al, 2014

The best-performing methods will be applied retrospectively to over 10,000 cancer genomes stored in CGHub, and the results will be distributed to the research community. Moreover, the top-scoring methods will be made available as open source tools, allowing users around the world to process their own data with the same pipelines validated and used by the ICGC and TCGA. Nature Publishing Group has stepped up to coordinate publication models stemming from the Somatic Mutation Calling Challenge. Challenge-assisted peer review and early editorial feedback will help identify publishable themes that cut across multiple approaches. The involvement of major journals introduces the possibility of reaching a broad audience and raises the impact and exposure of contestant contributions, thereby increasing incentives and overall morale.

Summary

- Challenges and benchmarks can be important in image analysis for radiology and digital pathology, radiomics and radiogenomics.
- Reviewed challenge infrastructure and requirements to host and participate in challenges.
- Discussed past and upcoming challenges that focus on topics in radiology, radiomics and radiogenomics.

Acknowledgments

- ▣ CBIIT
- ▣ Ed Helton
- ▣ Ulli Wagner
- ▣ MGH
 - ▣ Artem Mamonov
 - ▣ Karl Helmer
- ▣ MICCAI BraTS Org
- ▣ Stanford University
 - ▣ Daniel Rubin
- ▣ Emory University
 - ▣ Ashish Sharma
- ▣ QIN
- ▣ TCIA
- ▣ Microsoft Azure
- ▣ Financial support: Leidos contract, NIH grant U24CA180927

Dedication: Dr. Laurence Clarke

- Dr. Larry Clarke was absolutely instrumental in providing the vision, enthusiasm and drive for this project.



<https://spie.org/about-spie/press-room/spie-member-news/in-memory-laurence-clarke-nci-imaging-cancer>