

Informatics Tools for Optimized Imaging Biomarkers for Cancer Research & Discovery

Jayashree Kalpathy-Cramer, PhD

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

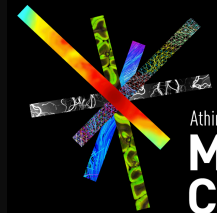
ITCR Annual Meeting, 2017



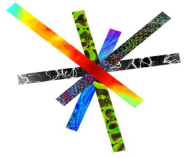
MASSACHUSETTS
GENERAL HOSPITAL



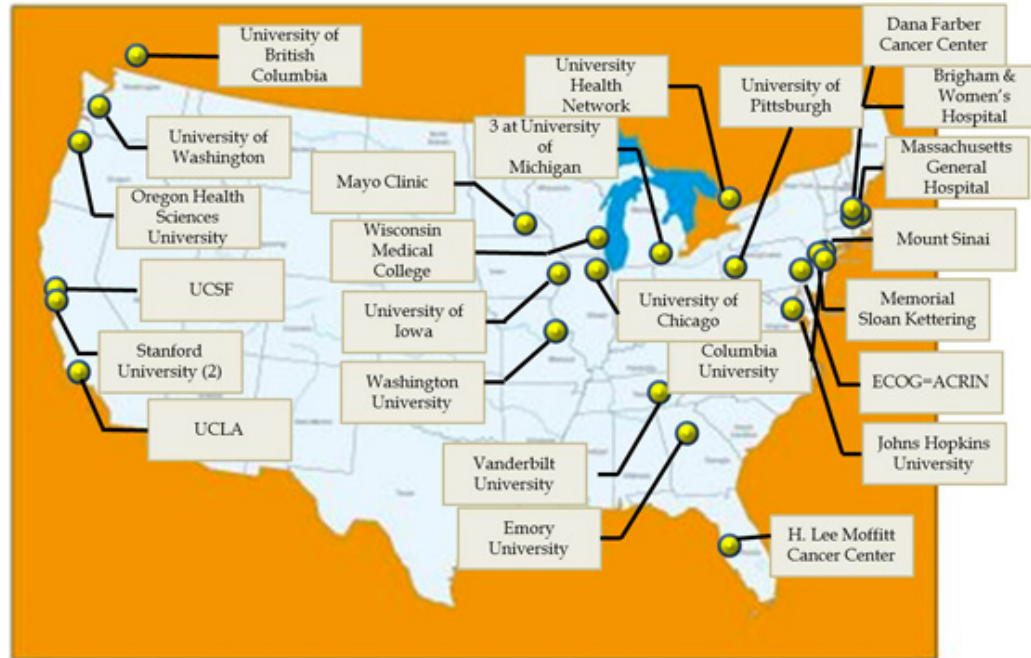
HARVARD
MEDICAL SCHOOL



Athinoula A.
**Martinos
Center**
For Biomedical Imaging



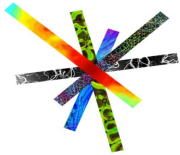
Quantitative Imaging Network



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

Collaborators



Bruce Rosen
Jayashree Kalpathy-Cramer
MGH
QIN U01, Leidos Contract



Binsheng Zhao,
CUMC
QIN U01, R01

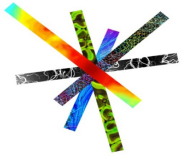


Robert Gillies Dmitry Goldgof
Moffitt Cancer Center/USF
QIN U01/R01

ITCR U24



Sandy Napel
QIN U01 (QIFP)
Stanford University

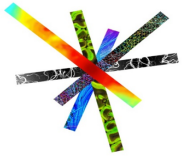


Project Goal:

Make radiology imaging data more accessible to non-imaging scientists

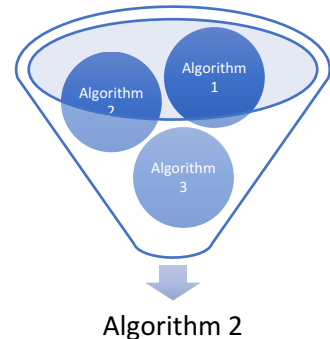
Support reproducible research

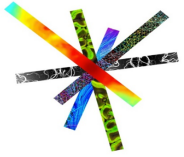
Platform Development	Algorithm Development	Host Challenges
<ul style="list-style-type: none">••Challenge management<ul style="list-style-type: none">••Data••Users••Dockerized Code••Execution engine••Evaluation••Visualization	<ul style="list-style-type: none">••Sites<ul style="list-style-type: none">••CUMC••Moffitt••MGH••Stanford••Image analysis<ul style="list-style-type: none">••Segmentation••Radiomics	<ul style="list-style-type: none">••QIN••QIBA••Community••Societies<ul style="list-style-type: none">••SPIE••AAPM••MICCAI



Why “challenges”?

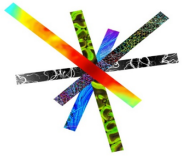
- Reproducibility is an issue in all aspects of medicine
 - Algorithm performance often not replicated by other sites
- Access to clinical data of sufficient variety can be a challenge for (computational) scientists developing algorithms
- Can evaluate the performance of techniques on real, noisy clinical 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





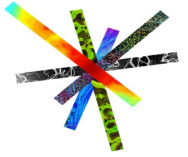
Challenge options

- Traditional model (most Kaggle challenges including recent lung data science bowl)
 - Training and test data available, results uploaded
- Code submission
 - Upload code as executable/python script
- Containerized code submission (e.g. DREAM challenge for digital mammography)
 - Docker submissions (can be too large!)
 - (Private) docker registry



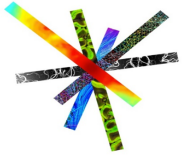
Limitations of traditional model

- Datasets too large to be moved around
- Some data may be too sensitive to share (PHI)
- Limited reproducibility when participants provide just results and not code
 - Easier to “cheat”
- Cannot compare algorithm efficiency (and how should we compare alg A with 95% Dice, 24 hours run time with alg B at 93% Dice, 2.4s run time)
- Cannot compare algorithms performance on unseen, new data



Need to transition to cloud-based evaluation

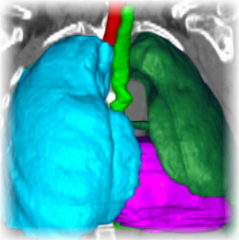
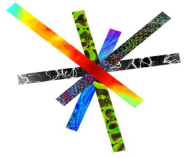
- Algorithms move to data
- Share executables/code/VM/Docker
- Participants never see test data
- Code can be run on new (prospective?) data



System components



Leverages open source projects including R, CodaLab, ePad, Slicer, QIICR, Docker,



AAPM Thoracic Auto-segmentation Challenge

Organized by MarkGooding - Current server time: June 1, 2017, 2:19 p.m. UTC

► Current

Training Phase

May 18, 2017, midnight UTC

Next

Pre-AAPM Challenge

June 19, 2017, midnight UTC

Learn the Details

Phases

Participate

Results

Forums →

Overview

[Contouring Guidelines](#)

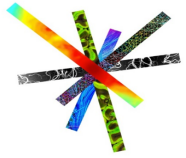
[Evaluation](#)

[Terms And Conditions](#)

[Organizers](#)

Overview

Numerous auto-segmentation methods exist for Organs at Risk in radiotherapy. The overall objective of this auto-segmentation grand challenge is to provide a platform for comparison of various auto-segmentation algorithms when they are used to delineate organs at risk (OARs) from CT images for thoracic patients in radiation treatment planning. The results will provide an indication of the performances achieved by various auto-segmentation algorithms and can be used to guide the selection of these algorithms for clinic use if desirable. The challenge is made up of multiple phases:



Learn the Details

Phases

Participate

Results

Forums →

Overview

Contouring Guidelines

Evaluation

Terms And Conditions

Organizers

Contouring Guidelines

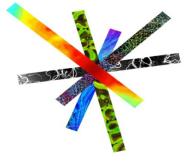
Esophagus

Standard name: Esophagus

RTOG Atlas description: The esophagus should be contoured from the beginning at the level just below the cricoid to its entrance to the stomach at GE junction. The esophagus will be contoured using mediastinal window/level on CT to correspond to the mucosal, submucosa, and all muscular layers out to the fatty adventitia.

Additional notes: The superior-most slice of the esophagus is the slice below the first slice where the lamina of the cricoid cartilage is visible (+/- 1 slice). The inferior-most slice of the esophagus is the first slice (+/- 1 slice) where the esophagus and stomach are joined, and at least 10 square cm of stomach cross section is visible.

Heart

[Overview](#)[Contouring Guidelines](#)[Evaluation](#)[Terms And Conditions](#)[Organizers](#)

Evaluation Criteria

Auto-segmented contours will be compared against the manual contours for all test datasets using the following evaluation metrics

- Dice
- Hausdorff distance (95% Hausdorff distance)
- Mean surface distance

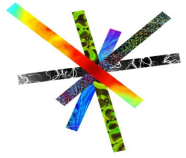
Evaluation slice by slice or in 3D as appropriate to the OAR and the field of view. i.e. OARs fully enclosed within the field of view will be assessed in 3D, whereas OARs that may extend beyond the field of view, e.g. Spinal Cord, will be assessed in 2D.

The winners

Details to follow

[Overview](#)[Contouring Guidelines](#)[Evaluation](#)[Terms And Conditions](#)[Organizers](#)

- Anonymous participation is not allowed
- By entering you give the organizers to publish the results of this study
- Results will not be linked to participants in publications without express permission of the participant to do so



Options of submitting code

- Upload Docker
- Pull Docker from (public/private) registry

Submit Docker

Directories that can be accessed on the server:

- dicom_dir - Directory that contains original DICOM images
- nifti_dir - Directory that contains segmentation objects in NIFTI format
- dso_dir - Directory that contains segmentation objects in DSO format
- output_dir - Directory where you can write output of the docker container
- config_dir - Directory that contains configuration files that can be uploaded with the submission

Docker command (customizable):

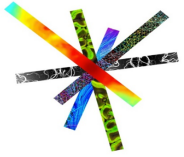
```
docker run --rm  
-v dicom_dir/riipl/data/dicom -v dso_dir/riipl/data/dso -v  
output_dir/riipl/data/output -v config_dir/config.ini:/riipl/data/config.ini
```

● artem/codalab
● cherezov/features
● riipl/3d_qifp
● stanford-features

Upload configuration files

Choose Files No file chosen

Close Send



Feature pipeline output

9	0.028407	submission-stanford.zip	05/23/2017 14:23:01	Finished		—
---	----------	-------------------------	---------------------	----------	--	---

Description:

[update description](#)

[Download your submission](#)

[View scoring output log](#)

[View scoring error log](#)

[View predict output log](#)

[View predict error log](#)

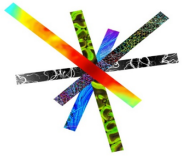
[Download evaluation output from prediction step](#)

[Download evaluation output from scoring step](#)

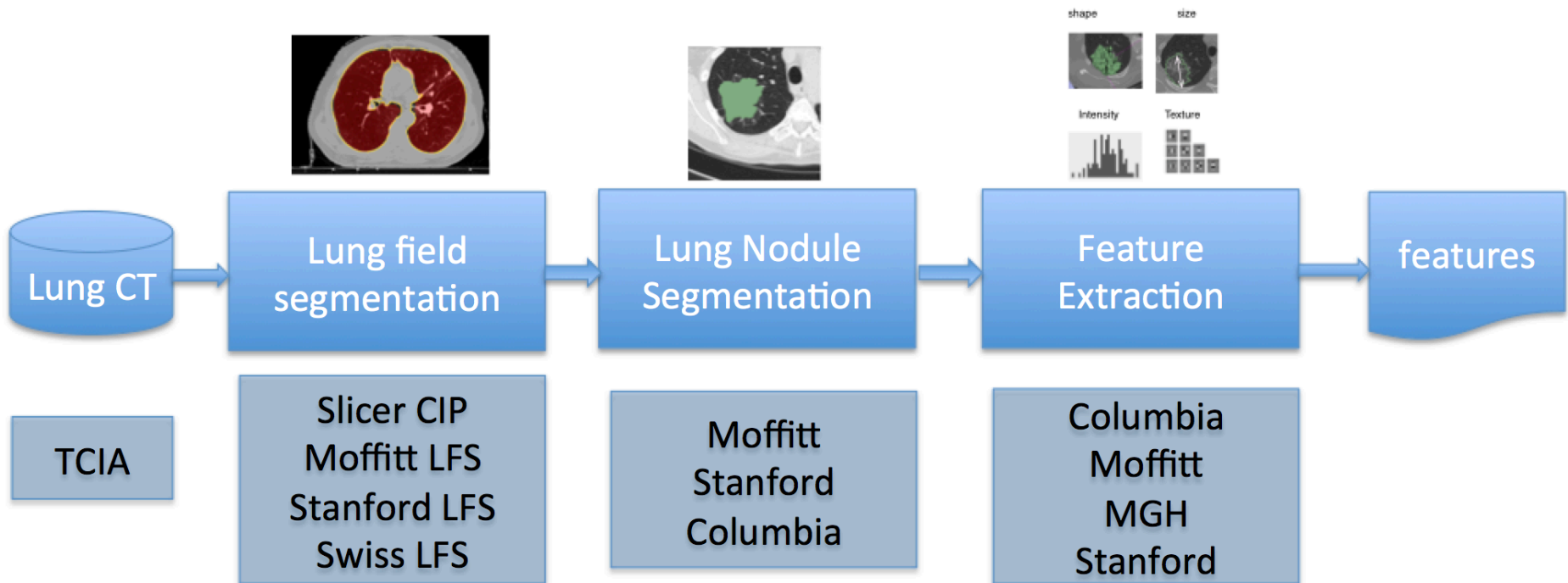
[Download private output from scoring step](#)

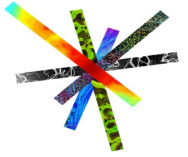
[Submit to Leaderboard](#)

gldm.distance.1mm.correlation.range	0.025925
gldm.distance.1mm.correlation.max	0.25956
gldm.distance.1mm.correlation.min	0.23364
gldm.distance.1mm.correlation.interquartileRange	0.012868
gldm.distance.1mm.correlation.meanAbsoluteDeviation	0.0071221
gldm.distance.1mm.contrast.trimmedMean(90%)	0.72814
gldm.distance.1mm.contrast.mean	0.74018
gldm.distance.1mm.contrast.median	0.74983
gldm.distance.1mm.contrast.kurtosis	3.13
gldm.distance.1mm.contrast.skewness	0.79757



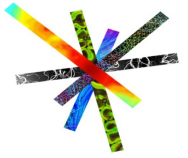
Radiomics Workflow





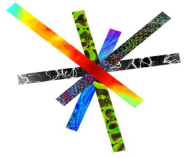
Updates on algorithms

- Dockerized implementations of
 - Lung field segmentations
 - Lung nodule segmentation
 - Radiomics pipelines
 - Deep learning feature extractions
 - Classification
- Evaluation workflow language options



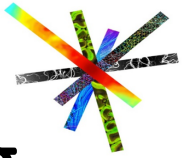
CUMC algorithm

- Lung lesion segmentation:
 - random walk-based algorithm
 - image is modeled as a graph
 - Seed point provided by user
- Quantitative image features
 - tumor size, shape, margin spiculation, sharpness, density distributions (histogram-derived), texture



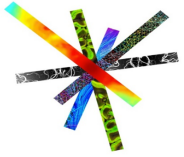
Moffitt Radiomics pipeline

- Implemented in C++ and requires VTK, ITK libraries
- 186 features
 - 181 are texture features and 5 are size features
 - Texture features: Laws, Wavelet, Co-Occurrence, Run length.
 - Size features describe volume
 - All the features are 3D
- The size of the docker container ~6Gb. If repository with the docker is available, then the system will automatically download one at the first call

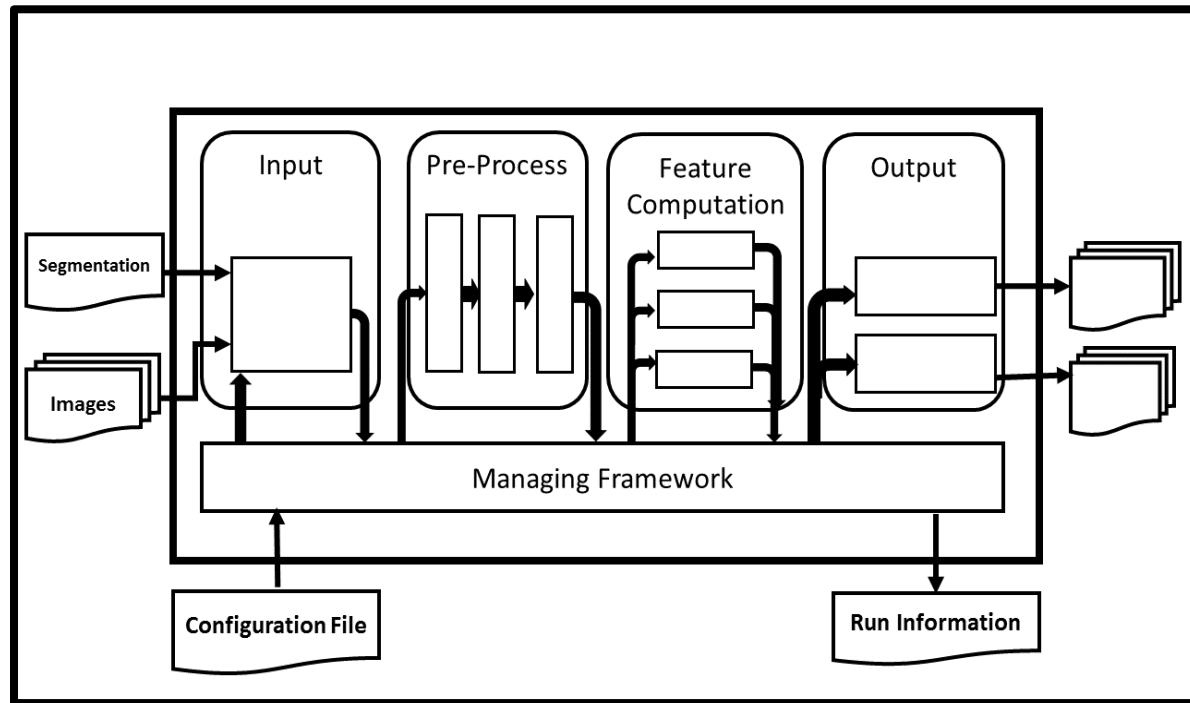


Moffitt Deep Feature – Transfer Learning

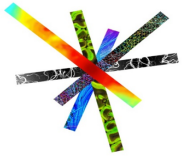
- ImageNet Pre-trained CNN (VGG architectures)
- One slice per patient that has the largest tumor area
- Extracted features from last convolution layer (Before and after applying ReLU)
- Previous result using Quantitative (hand-crafted) features only – Max accuracy 77.5% (AUC -0.713)
- Experimented with Deep features only (5 and 10 features using Relief-f and Symmetric uncertainty) - Max accuracy 82.5% (AUC-0.778) – Decision Tree classifier – 5 features
- Experimented by merging Quantitative (hand-crafted) features with Deep features (5 and 10 features using Relief-f and Symmetric uncertainty) - Max accuracy 90% (AUC- 0.935) – Naïve Bayes Classifier– 5 deep features+ 5 quantitative features



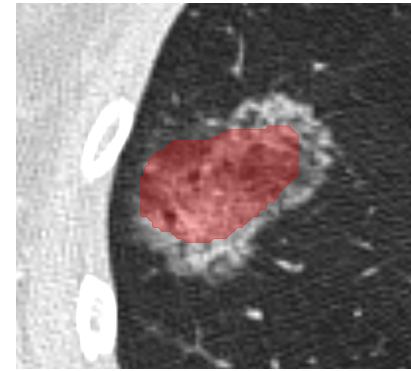
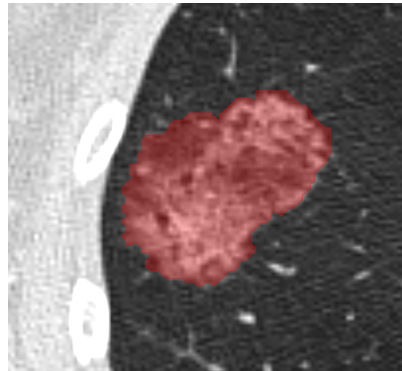
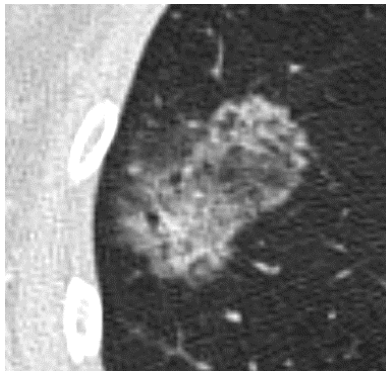
Stanford feature pipeline



Feature extraction pipeline –extracts features given image +segmentation

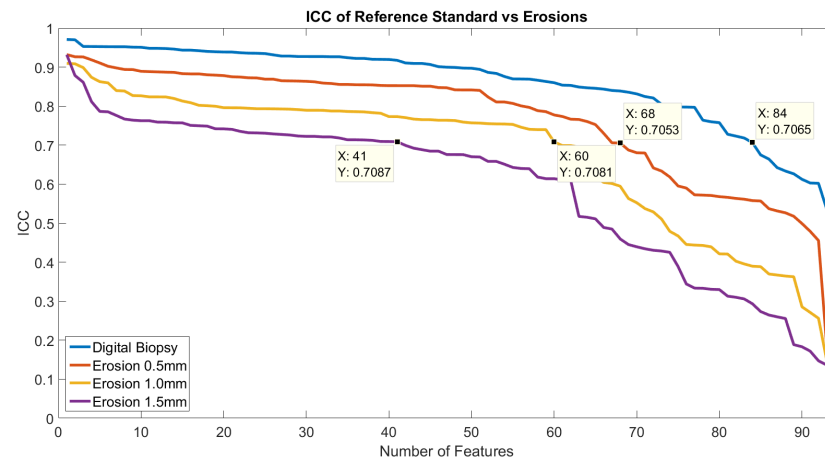


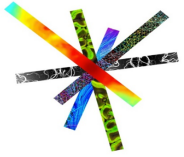
3D Digital Biopsy:



Reference Standard

3D Digital Biopsy

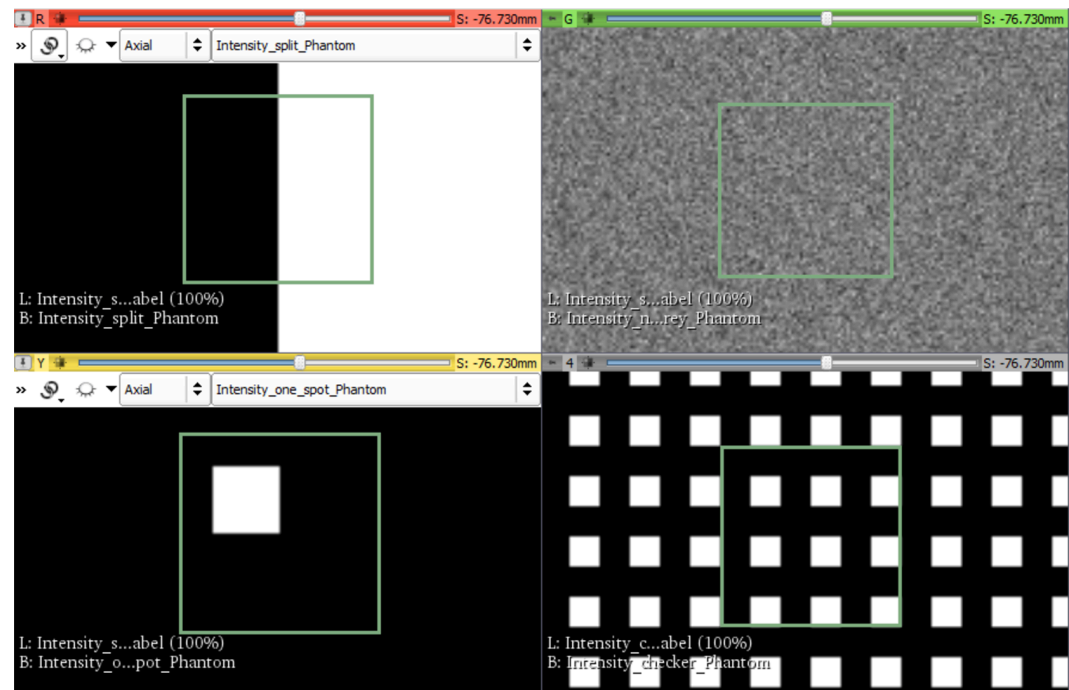


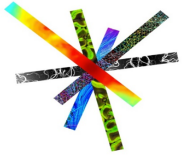


MGH radiomics pipeline

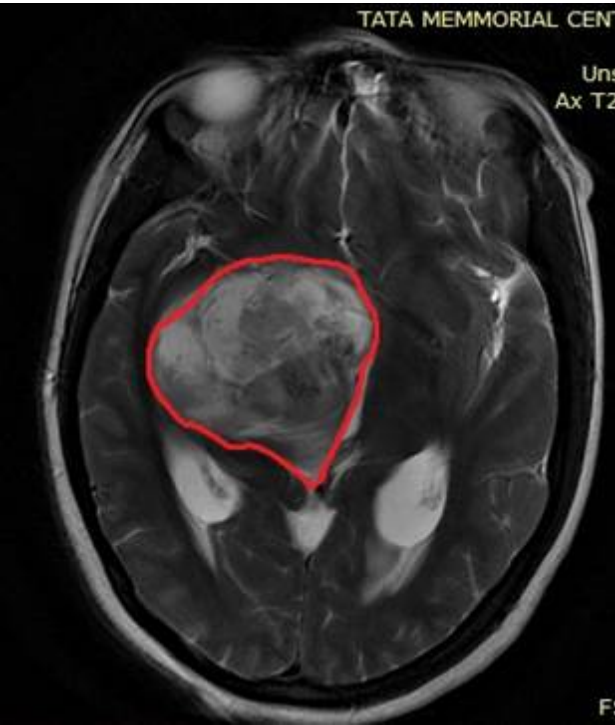
- Python package available as pip install qtim_tools
 - `qtim_tools.qtim_features.extract_features(folder, outfile)`
- Python notebook with tutorial
- Phantoms

mri-phantom from the previous example.

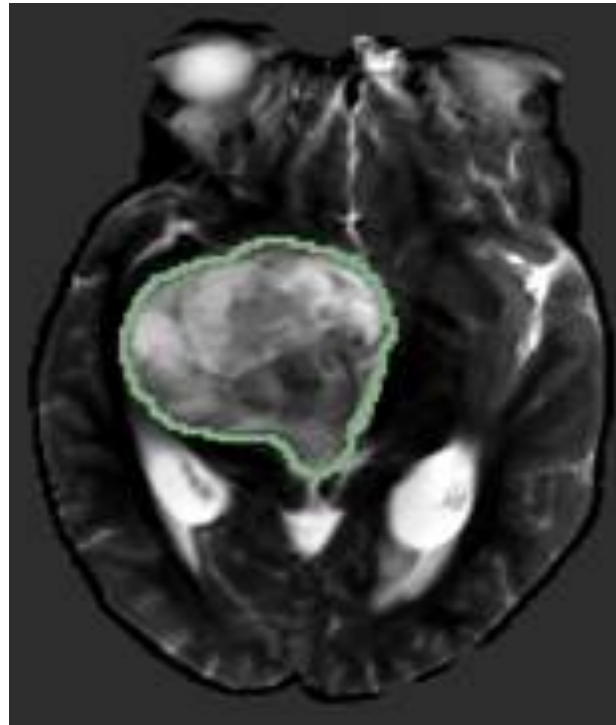




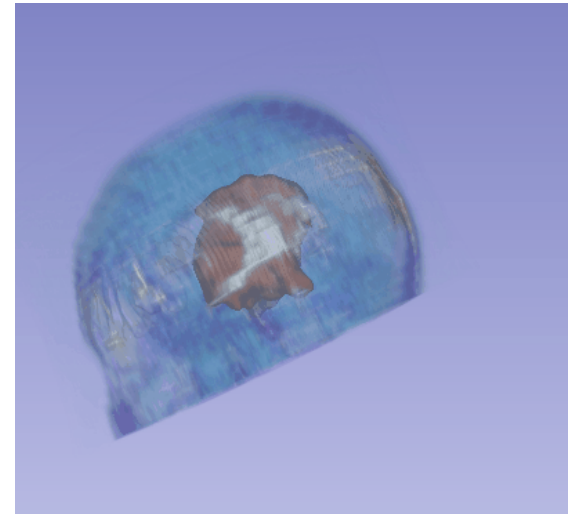
MGH tumor segmentation



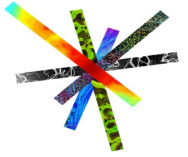
RADIOLOGIST



DEEP LEARNING

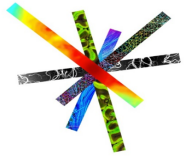


Currently evaluation large retrospective dataset



Deep learning (in progress)

- Predicting status of IDH, 1p19q etc using deep learning
- Survival analysis
- Grade classification

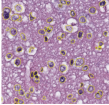
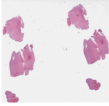
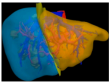


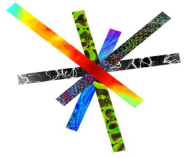
Challenges supported

- Computational Precision Medicine at MICCAI 2016 and 2017
 - Liver challenge
 - Digital pathology challenge

MICCAI 2017 COMPUTATIONAL PRECISION MEDICINE

Worksheets Competitions Help

	<p>Digital Pathology -- Segmentation of Nuclei in Images</p> <p>Organized by cpm.organizingcommittee</p> <p>The goal of this sub-challenge is to evaluate the performance of algorithms for detection and segmentation of nuclear material in ...</p>	<p>Jun 01, 2017-Jul 31, 2017</p> <p>22 participants</p>
	<p>Digital Pathology -- Classification of whole slide tissue images</p> <p>Organized by cpm.organizingcommittee</p> <p>The goal of this challenge is to evaluate the performance of automated classification algorithms.</p>	<p>Jun 01, 2017-Jul 31, 2017</p> <p>18 participants</p>
	<p>Colorectal Liver Metastases Survival Prediction</p> <p>Organized by cpm.organizingcommittee</p> <p>To predict survival based on predictors derived from contrast-enhanced liver CT scans and patient clinical variables.</p>	<p>May 15, 2017-<i>No end date</i></p> <p>15 participants</p>



QIBA CT Virtual Clinical Trial Grand Challenge

Competitions

Competitions I'm In

Competitions I'm Running

My Datasets



CT Virtual Clinical Trial Grand Challenge

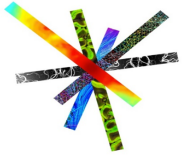
Organized by [organizing.committee](#)

This challenge is to evaluate algorithms for lung nodule segmentation

Jul 26, 2016-Mar 04, 2017

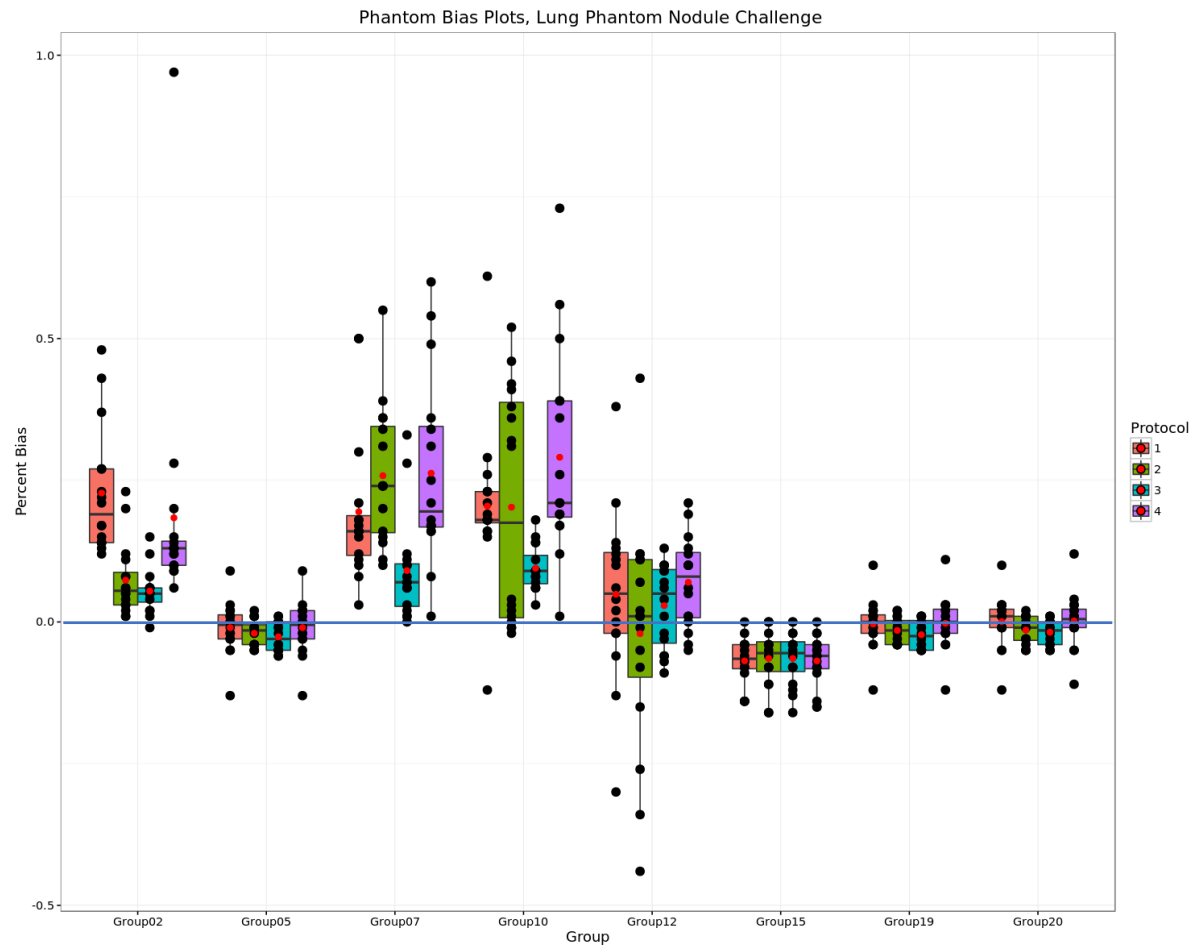
41 participants

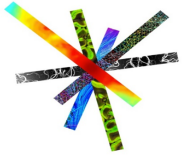
- Compare performance of lung nodule segmentation algorithms
- Compare synthetic lesion insertion methods
- Sponsored by QIBA CT volumetry group
 - Number of commercial participants (who desire anonymity)



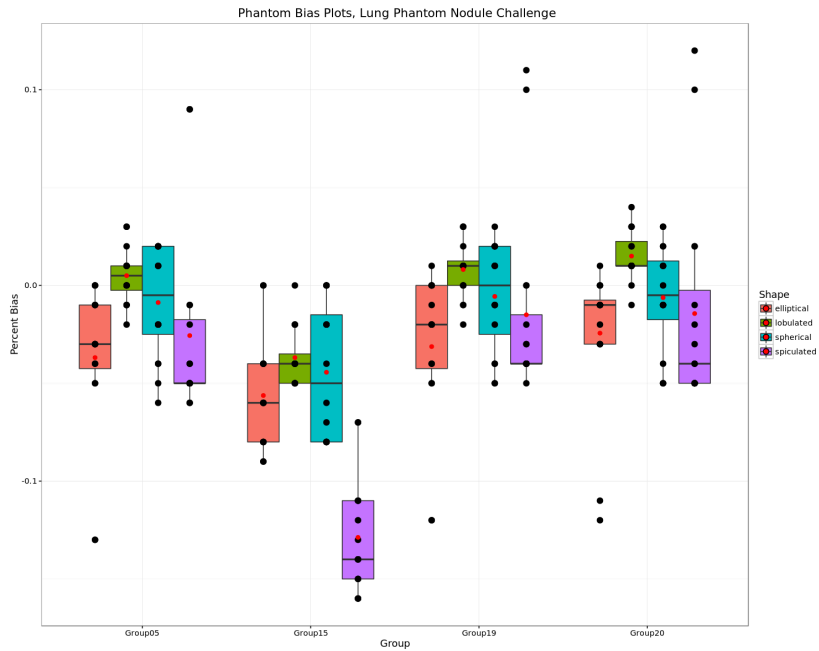
Substantial difference between algorithms

Only 4/8 groups meet QIBA profile specification for bias

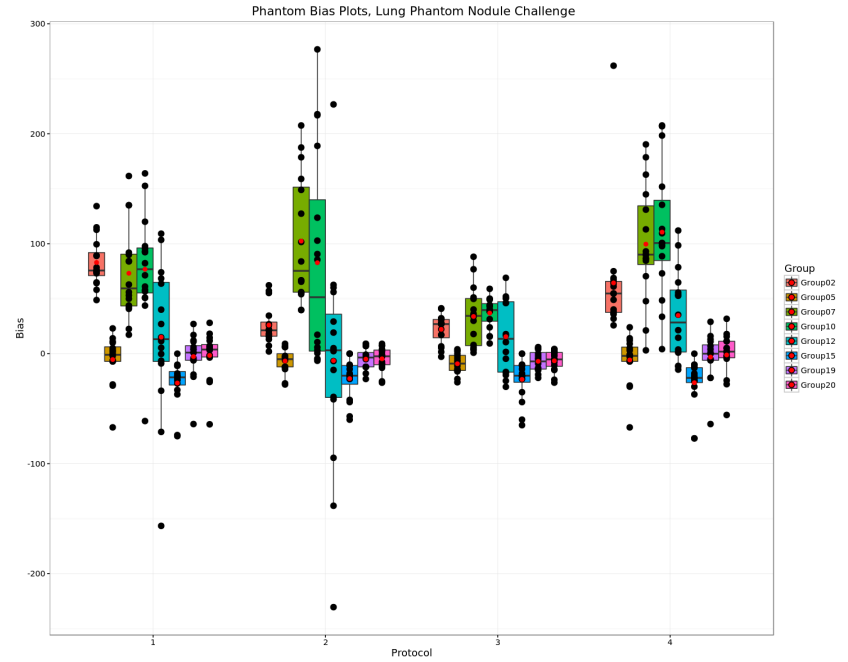




Interactive graphics

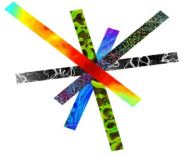


Effect of lesion shape



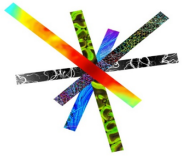
Effect of insertion protocol

Developed using R/shiny/WebGL



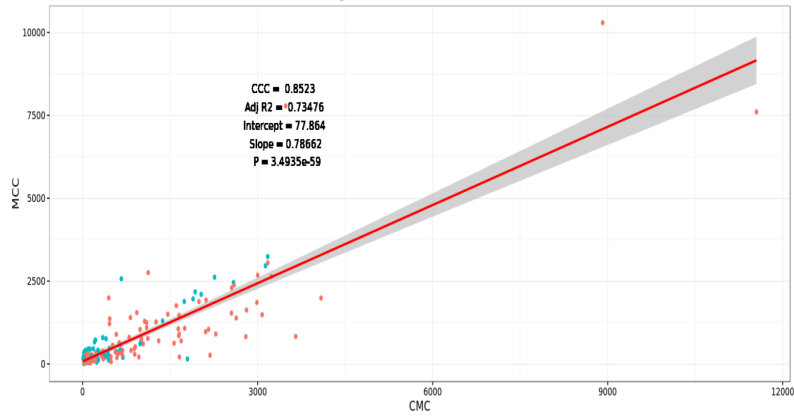
QIN Interval challenge

- 100 patients from NLST (lung cancer)
 - 50 malignant/50 benign
- Two visits each
- 7 algorithms (2 semi-automatic)
- Segmentations provided
- Volume change as a biomarker

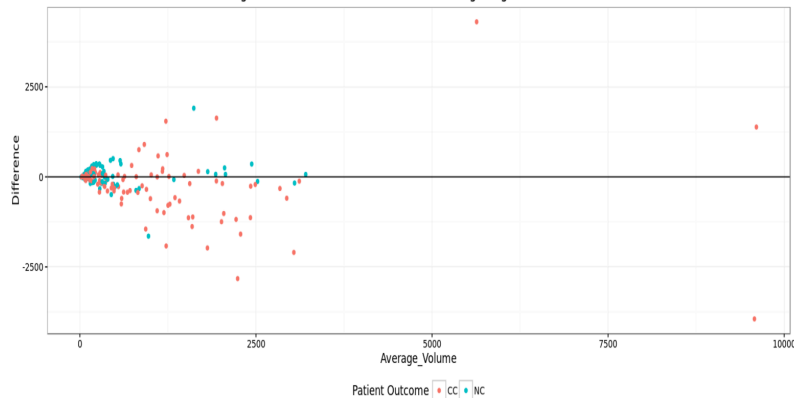


Analysis and visualization

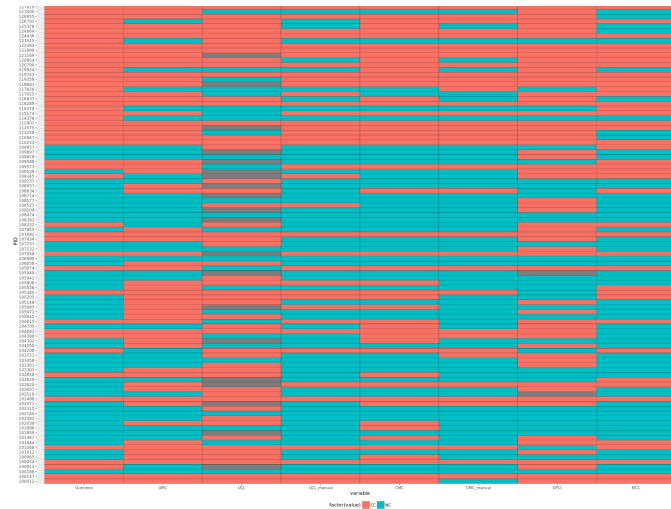
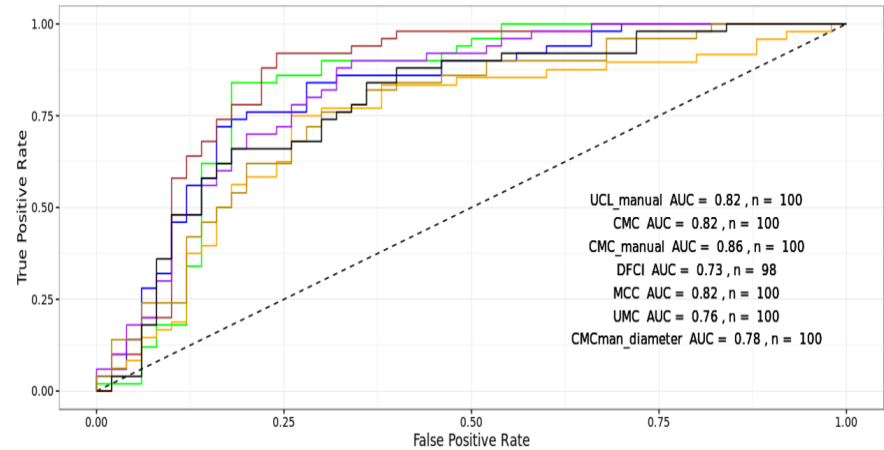
Segmentation Volume, Site vs. Site

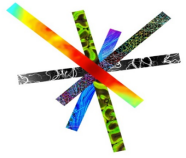


Segmentation Volume Differences vs. Average Segmentation Volumes



ROC Curve, Site 1

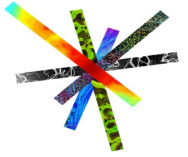




QIN labs – BMMR challenge

The screenshot shows the QIN Labs website interface. At the top left is the QINLABS logo. At the top right are links for 'My Competitions', 'Help', and a user profile for 'artem'. The main content area features a header for 'QIN BMMR - Breast MRI Metrics of Response', organized by 'nci_qin_committee'. Below this is a navigation bar with 'Current' (selected), 'Next', 'Training', and 'Test' buttons. The 'Current' phase is active from April 25, 2016, to May 25, 2016. A secondary navigation bar includes 'Learn the Details', 'Phases', 'Participate', 'Results', and 'Forums'. The 'Overview' section is expanded, showing a 'Welcome!' message and an introduction to the challenge, which is a retrospective analysis of ISPY/ACRIN clinical trial data. The text explains that MRI is effective for monitoring primary breast cancer response to neoadjuvant chemotherapy (NACT) and has the potential to provide prognostic information and serve as a non-invasive biomarker for predicting response. It also mentions that the American College of Radiology Imaging Network (ACRIN) trial 6657 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.

Retrospective analysis of ISPY/ACRIN clinical trial data
Potential to use best methods in prospective trial



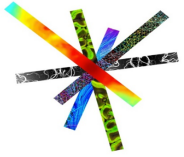
SPIE-AAPM-NCI Prostate challenge

- Challenge to identify quantitative MRI biomarkers for the determination of Gleason Grade Group in prostate cancer

PROSTATIS 2

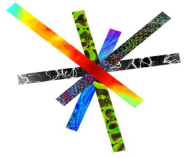
SPIE-AAPM-NCI
Prostate MR
Gleason Grade

May 15, 2017
102 participants



Radiomics “challenge”

- Radiomics pipelines on lung nodules
 - 52 lesions from 41 CT studies
 - Range of shapes and sizes
 - 3 segmentation algorithms, 3 runs each
- 8 radiomics pipelines
 - CUMC, PM, Stanford, UCLA, Iowa, U Mich, USF, MGH
- Participants submitted features and feature dictionary

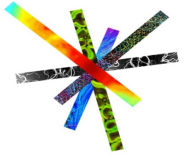


Radiomics terminology

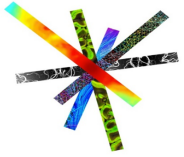
Participant	Size	GSDs	LSDs	Intensity	Margin	Texture					Total
						GLCM	LoG	Law's	Run Length	Wavelet	
CUMC	3	4	8	5		17	6	14		14	71
PM	3	2		5							10
Stanford	2	1	78	17	27	72					197
UCLA	1			4		10					15
Ulowa	2	6		9	151			136			304
UMICH	4	5		6	18				16		49
USF	5					6		125	20	28	184
Total	20	18	86	46	196	104	6	275	36	42	830

Abbreviations: CUMC, Columbia University Medical Center; PM, Princess Margaret Cancer Center; UCLA, University of California Los Angeles; Ulowa, University of Iowa; UMICH, University of Michigan; USF, University of South Florida.

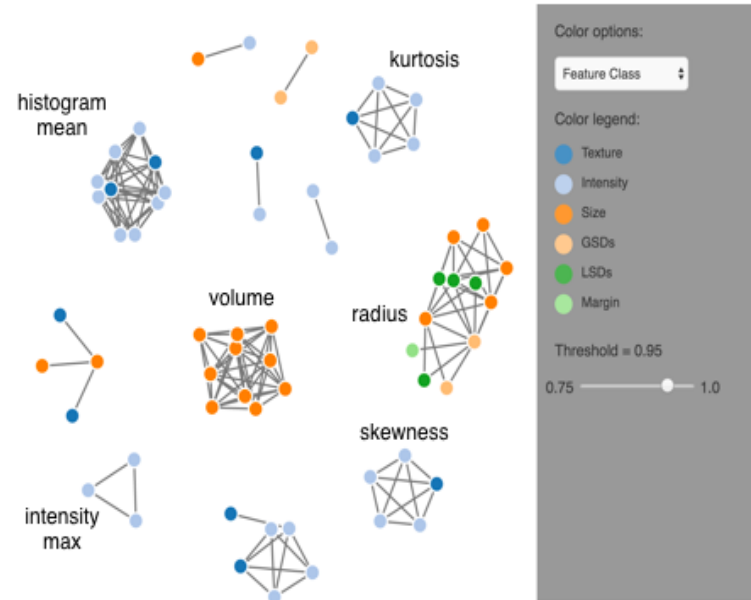
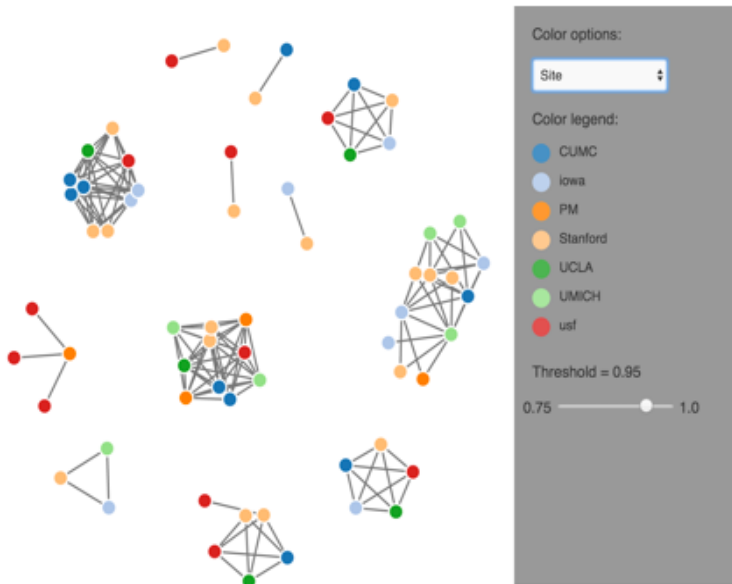
Developed categorization of radiomics features



 This image cannot currently be displayed.

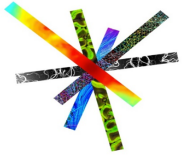


Graphical model of CC between features



Unexpected correlations uncovered

Lack of correlations between supposedly identical features discovered



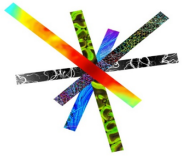
Unexpected correlations

- Texture features were highly correlated with size
 - Should they be?



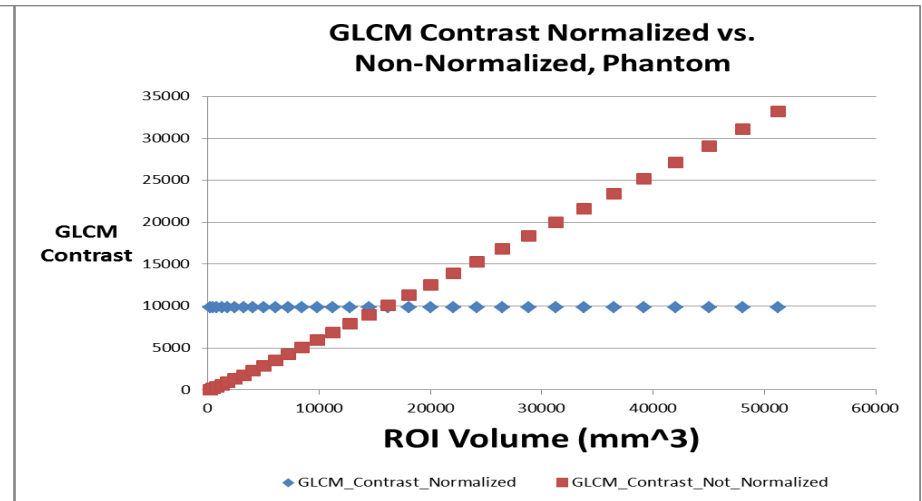
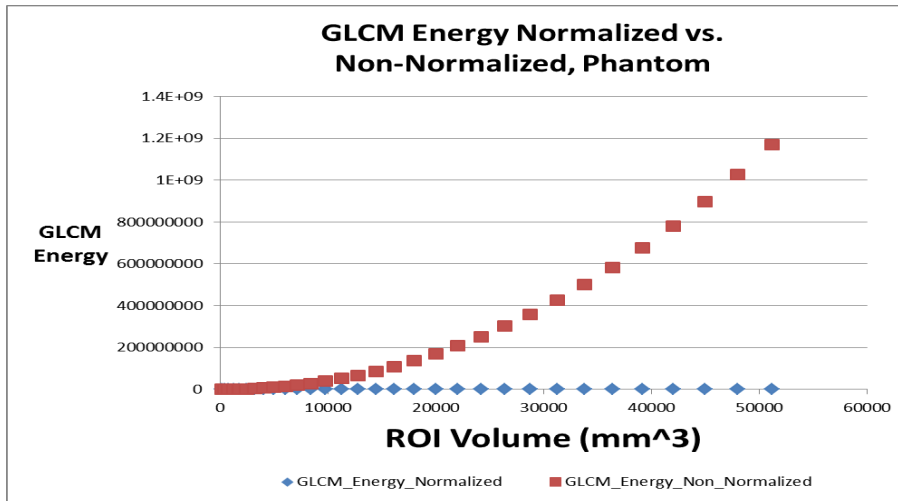
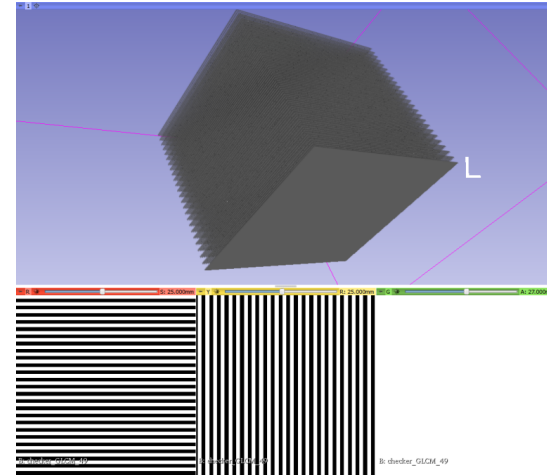
Color legend:

- Texture
- Intensity
- GSDs
- Size
- LSDs
- Margin



Lack of Normalization

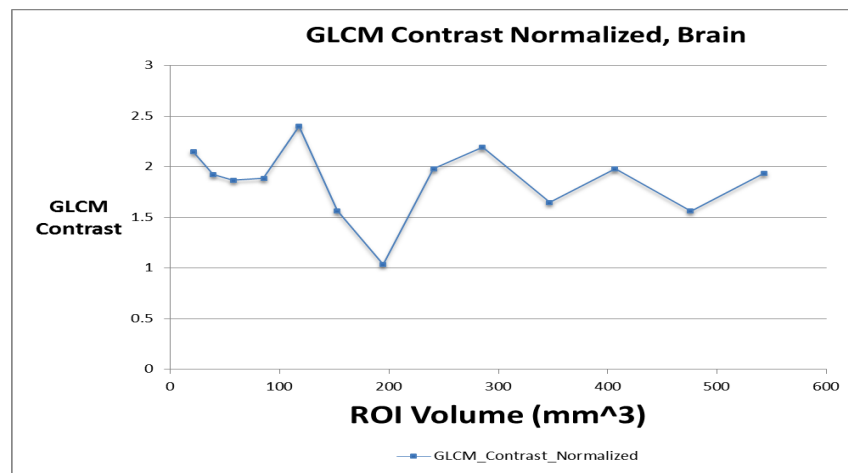
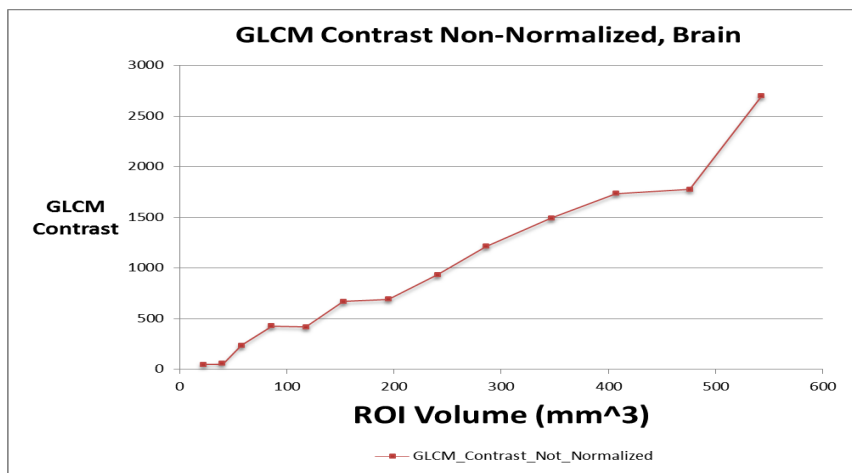
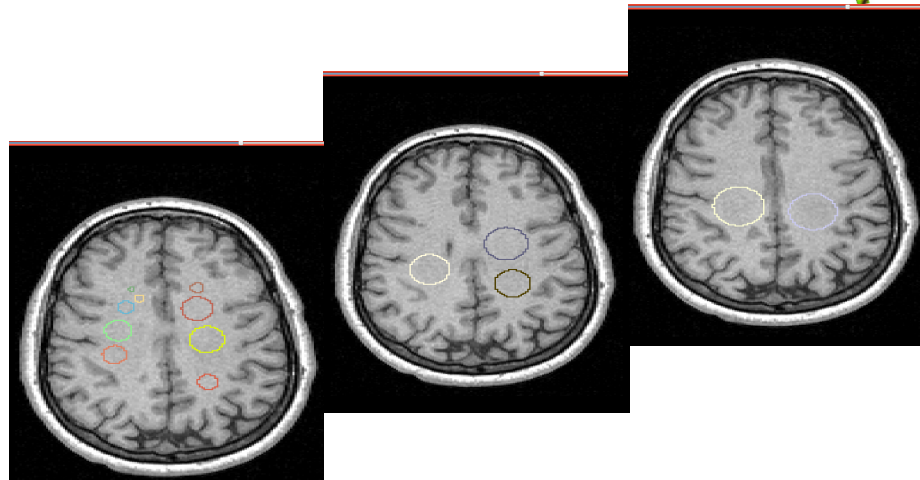
Features suffering from lack of normalization will primarily reflect voxel count and volume.

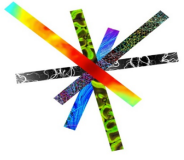




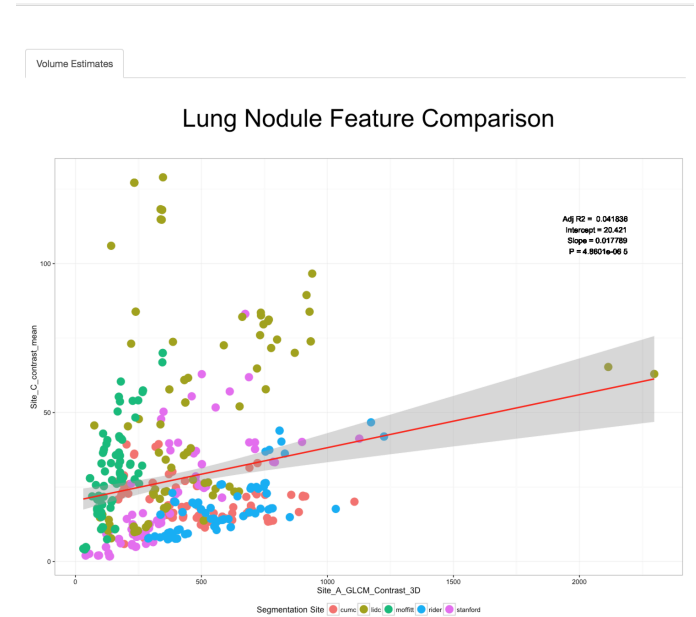
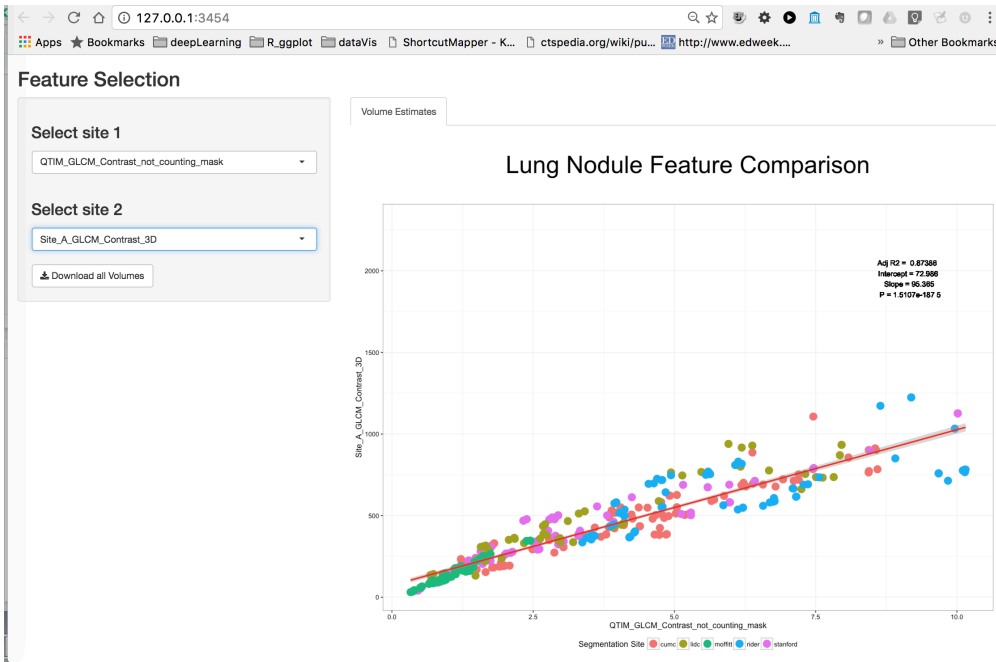
Lack of Normalization

Features suffering from lack of normalization will primarily reflect voxel count and volume.





Implementations can agree (or not)

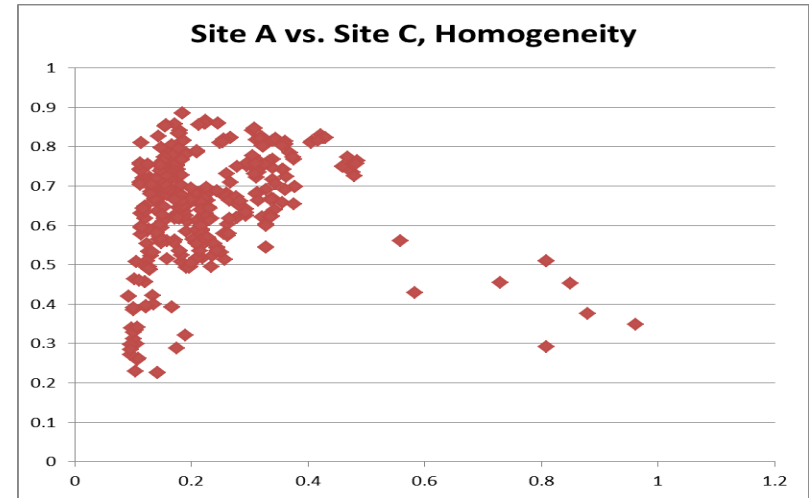
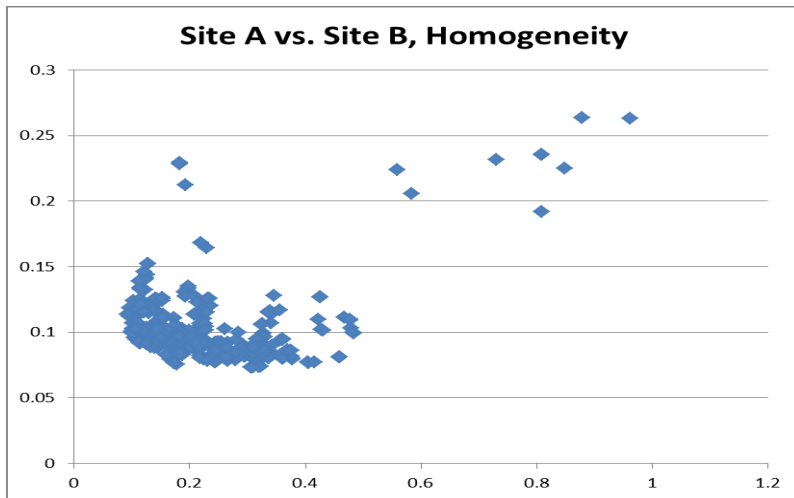
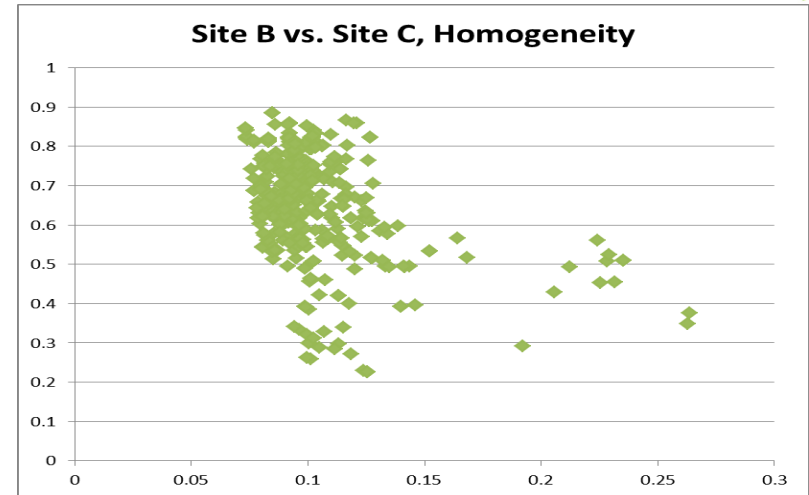


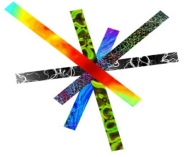
Texture features (contrast) varies significantly by implementation



Lack of Agreement

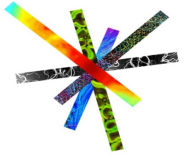
Different sites measuring the same feature are not well-correlated, impeding reproducibility.





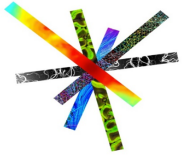
Next steps for radiomics feature standardization

- International effort for texture features underway
- Standardization of feature terminology, mathematical definition
 - Definition based and data-driven
- Prescribing minimal set of information to be for reproducibility
 - Quantization, directions, pre-processing, normalization, treatment of edge pixels



Summary

- Challenges and benchmarks can be important in image analysis, radiomics and radiogenomics.
- The C-BIBOP facilitates conducting of challenges and benchmarks
- Moving algorithms to data is a new paradigm
- Containerization of algorithms facilitates sharing of code and workflows
 - Exploring CWL and WDL



Acknowledgments

- ▣ MGH
 - ▣ QTIM lab
- ▣ Stanford
- ▣ Moffitt/USF
- ▣ CUMC
- ▣ ITCR
- ▣ QIN
- ▣ TCIA
- ▣ BWH
- ▣ Grant support: U24CA180927, U24CA180918, U01CA154601, Leidos contract, U01 CA187947, U01 CA140207, U01 CA143062



Delete

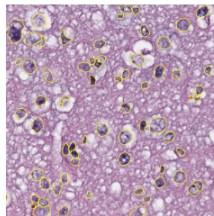
Edit

Publish

Participants

Submissions

Leaderboard



Digital Pathology -- Segmentation of Nuclei in Images

Organized by cpm.organizing.committee

The goal of this challenge is to evaluate the performance of algorithms for detection and segmentation of nuclear material in ...

Jul 22, 2016-Sep 30, 2016

45 participants

Delete

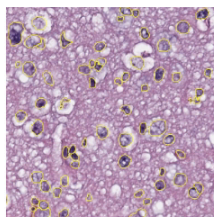
Edit

Publish

Participants

Submissions

Leaderboard



Digital Pathology -- Segmentation of Nuclei in Images

Organized by cpm.organizing.committee

The goal of this challenge is to evaluate the performance of algorithms for detection and segmentation of nuclear material in ...

Jul 22, 2016-Sep 30, 2016

45 participants

Delete

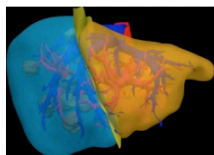
Edit

Publish

Participants

Submissions

Leaderboard



Colorectal Liver Metastases Survival Prediction

Organized by cpm.organizing.committee

To predict survival based on predictors derived from contrast-enhanced liver CT scans and patient clinical variables.

Jul 25, 2016-No end date

10 participants

Delete

Edit

Publish

Participants

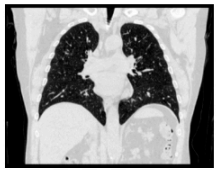
Submissions

Leaderboard

Competitions (My CodaLab)

Competitions Competitions I'm In **Competitions I'm Running** My Datasets

Create Competition



Lung nodule features competition with docker containers

Organized by artem

Participants upload python script with a command for running docker and a file with a name of the docker container ...

Feb 02, 2017 - *No end date*

1 participant

Delete Edit **Publish** Participants Submissions Leaderboard



Site Selection

Select site 1

CMC

Select site 2

MCC

Change Measure:

- Percent Difference
- Absolute Difference

Nodule Size:

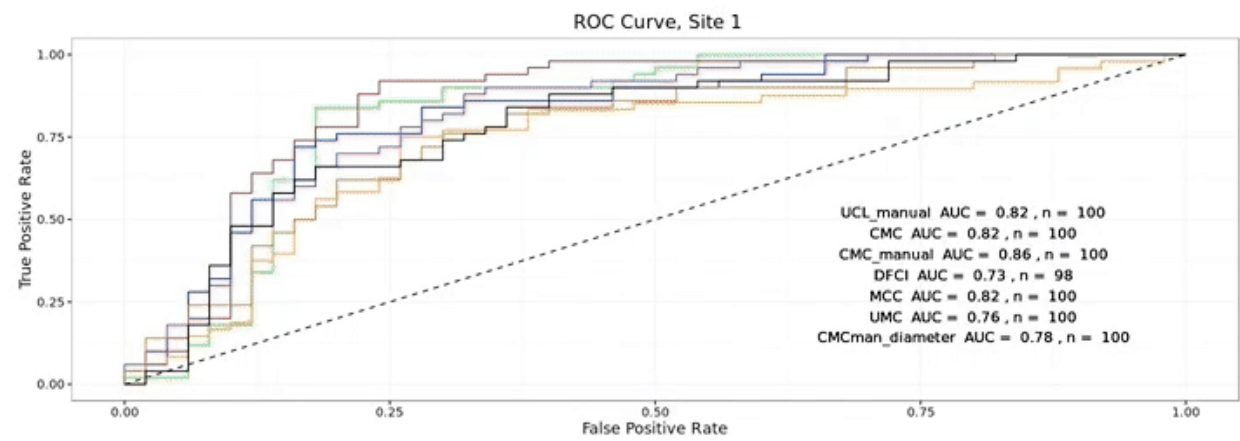
- Above 8mm Diameter (268 mm³)
- Below 8mm Diameter (268 mm³)

Nodule Status:

- Benign
- Malignant

Download all Volumes

- Volume Estimates
- Volume Histogram
- Volume Change Estimates
- ROC Curves
- ROC Curve Comparison**
- DICE Histogram
- CCC Matrix
- Small/Large Category Comparison
- Outcome Prediction
- Confusion Matrices
- All Data



- ### Sites
- UCLA, Red
 - UCLA_Manual, Green
 - CMC, Blue
 - CMC_Manual, Brown
 - DFCI, Orange
 - MCC, Purple
 - UMC, Yellow
 - CMC Manual Diameter, Black