Informatics Tools for Optimized Imaging
Biomarkers for Cancer Research & Discovery

Bruce Rosen/Jayashree Kalpathy-Cramer/Artem Mamonov/Karl Helmer

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

- Robert Gillies, PhD
  - Moffitt Cancer Center

- Dmitry Goldgof, PhD
  - University of South Florida

- Sandy Napel, PhD
  - Stanford University

- Binsheng Zhao, PhD
  - Columbia University
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
ITCR U24 : QIN as test-bed for C-BIBOP

- QIN is engaged in challenges in a fairly substantial manner
- Working on previous QIN challenges highlighted need for resources
  - Infrastructure (platform)
  - Resources to coordinate and conduct challenges
  - Metrology tools
  - Visualization tools
- Paradigm shift
  - Share code, not just results
- Support reproducibility in image analysis
- **Make imaging data more accessible to non-imaging scientists**
  - Note: Also funded by Leidos contract
Cloud-based Image Biomarker Optimization Platform (C-BIBOP)

- Goal to create an open-source platform to support algorithm comparison/benchmarking
  - Supports many use cases including uploading results or running algorithms in the cloud (VMs, Docker)
  - Live at all times, not just at conferences
  - Can be used for benchmarking and algorithm comparison
  - Can be used to share image analysis workflows

- Initial use cases
  - CT volumetry
  - Brain tumor segmentation

- Build on following:
  - QIN challenge infrastructure
  - CodaLab
  - VISCERAL (EU project)
  - TCIA API
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
In the second competition, The National Cancer Institute, part of the National Institutes of Health, and the FNIH will support competitions that will ask coders to create the best computer algorithm that can identify a person as having lung cancer based on two sets of low dose computed tomography images taken 1 to 2 years apart—when the scans may show the growth of a tumor. Another challenge will be more difficult: It will ask coders to create algorithms that can spot lung cancer based on one set of images taken from a single scan or study.
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
  - MICCAI has held “grand challenges” since 2007
  - ISBI
  - SPIE

- On commercial platforms
  - Kaggle
  - TopCoder
  - Sage/Synapse

- Within Organizations such as QIN and QIBA
  - Lung nodule segmentation challenges
  - DCE
  - PET
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
MICCAI Brain tumor segmentation challenge

- Organized by academics, NCI
- 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
Segmentation of brain tumors into 3 (4) regions
- Necrosis
- Enhancing tumor
- Edema

Based on
- T1 (pre and post contrast)
- T2
- FLAIR

Run over 3 years

10-20 groups participated
Example slices with output of segmentation algorithms
Results of computer generated labels

- Improvement by combining best methods
A new paradigm

- Limitations of traditional model:
  - Datasets too large to be moved around (in “Big Data” era)
  - Some data may be too sensitive to share (PHI)
  - Limited reproducibility when participants provide just results and not code
  - 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
Gold Corpus

Test Data

Training Data

Annotation Management System

Test Data

Participant Docker containers

Registration System

Analysis System

Silver Corpus

Participants

Organiser

Adapted from slides courtesy Allan Hanbury/Henning Mueller (VISCERAL)
System components

- CodaLab for challenge management
- Shiny/R integration for statistical analysis
- Integration with TCIA (Rest API)
- ePad for visualization
- caMicroscope for supporting digital pathology challenges
- Containers for sharing code

Also funded through Leidos contract
Docker vs. Virtual Machines

http://www.jayway.com/2015/03/21/a-not-very-short-introduction-to-docker/
QIN challenge: Lung nodule segmentation

- 52 lesions from 41 CT studies
- 33 to \( \mu l \) 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
Radiogenomics
(Radiopathagenomics...)

- Radiomics
- Genomics
- Clinical prediction/outcomes

Gevaert et al, Radiology, 2012
Features can be sensitive to segmentation

- Reduced uncertainty with machine assisted segmentation

“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
C-BIOP used for management
Feature stability with respect to segmentation
Results

- Inter and intra-site correlation of features
The aims of this challenge are:

- 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.

- To demonstrate improvement in predictor performance over functional tumor volume (FTV), the primary imaging variable tested in ACRIN 6657.
QIN BMMR challenge

- QIN challenge led by Nola Hylton, data from ISPY/ACRIN trial
Integrated Challenges (imaging/pathology/omics)

- **MICCAI 2015**
  - Joint radiology pathology brain tumor challenge
  - Classify glioma grade using pathology and radiology images

- **MICCAI 2016**
  - Joint radiology pathology challenge
  - TCIA-TCGA data
Cloud-based imaging workflows

- Tools are shared in Docker containers
- A worksheet can be used to share a pipeline of tools
- Supports reproducibility by allowing sharing of workflows with configurations, data, results
Example workflow

- Lung tumor radiomics

- Lung CT
- Lung field segmentation
  - Slicer CIP
  - Moffitt LFS
  - Stanford LFS
  - Swiss LFS
- Lung Nodule Segmentation
  - Moffitt
  - Stanford
- Feature Extraction
  - Stanford MGH

features
### Stanford tools (dockerized)

<table>
<thead>
<tr>
<th>Repository</th>
<th>Stars</th>
<th>Pulls</th>
</tr>
</thead>
<tbody>
<tr>
<td>ripl/feature</td>
<td>0</td>
<td>38</td>
</tr>
<tr>
<td>ripl/tumor</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>ripl/lung</td>
<td>0</td>
<td>6</td>
</tr>
</tbody>
</table>

- Available on Dockerhub
Stanford tools

- 3D image feature pipeline (DSOs + DICOM series in; features out)
- Lung tumor segmentation (AIM file with seed pixels + DICOM series + DSO for lung field in; DSO of nodule out)
- Lung field segmentation (DICOM series in; DOS for lung field out).
- Stanford features for the feature challenge were computed using the Docker version of our QIFP.
Pathology nuclear segmentation

Docker container from Stony Brook
Intel Personalized Care Platform

- Data organized and registered – TCIA-GBM across 2 “partner” sites
- Tools containerized – AFNI, FSL, DRAMMS
- Workflow converted from shell script to Workflow Description Language
- Job submitted to Execution Engine
Federated Site Arrangement

Partner site 1

- 110 images
- load raw
- unify
- FOV
- Strip
- Register
- Merge

Central site

- Brain tumor atlas
- Atlas calc

Partner site 2

- 274 images
- load raw
- unify
- FOV
- Strip
- Register
- Merge
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
- MICCAI BraTS Org
- Stanford
- Moffitt/USF
- CUMC

- ITCR
- QIN
- TCIA
- BWH
- Stony Brook

Grant support: U01CA154601, U24CA180927, U24CA180918, Leidos contract, ITCR administrative supplement