Grand Challenges and Public Image Datasets in Medical Imaging Research

Keyvan Farahani
Cancer Imaging Program

Outline
1. Why do a challenge?
2. The Cancer Imaging Archive (TCIA)
3. A brief history of NCI-supported imaging challenges

Why do an imaging challenge?
- Image analysis problems fit well for crowdsourcing
- Good use of reference data sets
- Reproducibility of methods
- Promote open science
- Benchmark algorithms
- Drive consensus on methodologies
- Raise awareness
Challenges as Engines of Innovation

Challenges: Crowdsourcing solutions
Bender E, Nature 533; 562-564; May 12 2016

Crowdsourcing biomedical research: Leveraging communities as innovative engines

www.challenge.gov

ImageNet Challenge

ImageNet – a large database of over 14,000,000 images of objects.

ImageNet Challenge - Large Scale Visual Recognition Challenge (ILSVRC) research teams evaluate their algorithms and compete to achieve higher accuracy on several visual recognition tasks.

http://image-net.org/
TCIA - The Cancer Imaging Archive

- 70 data sets consisting of over 33,000 subjects available for download
- Covers most modalities (CT/MR/PET/RT)
- Wide variety of cancers + phantoms
- Patient populations vary from a handful to >26,000 (NLST)
- Many have associated meta-data
- Demographics/outcomes/therapy
- Pathology imaging
- Radiologist expert and automated computational analysis (segmentations, features)
- 'Omics via TCGA, CPTAC, and GEO

Prostate Segmentation Challenge

- Automated or semi-automated methods to segment the central gland and peripheral zone in prostate MRI
- ISBI 2013
- Artifacts of the challenge, including the ground truth for the leaderboard and test data, are available through the following digital object identifier: http://dx.doi.org/10.7937/K9/TCIA.2015.sIfVdOPv

Other Challenges: Medical Imaging Community

- PROSTATEx Challenges – SPIE-AAPM-NCI
- MICCAI 2016 and 2017 – Computational Precision Medicine
- MICCAI 2013-2017 – Brain Tumor Segmentation (BRATS)
- LUNGx SPIE-AAPM-NCI Lung Nodule Challenge
NCI Quantitative Imaging Network

QIN – A network of 21+ research teams involved in development and optimization of quantitative imaging methods for prediction or evaluation of response to cancer treatments.

QINLabs: A Challenge Evaluation Platform

http://qinlabs.cloudapp.net/

Other Challenges: Global Community

➢ Distinguish patients with cancer from those without cancer, given LDCT scans.
TCIA Stats: Increase in LIDC Data Downloads

DSB Leaderboard

ROC analysis Data Science Bowl test set

<table>
<thead>
<tr>
<th>Lung-RADS category</th>
<th>Non-cancer cases</th>
<th>Cancer cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>66</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>138</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>28</td>
<td>10</td>
</tr>
<tr>
<td>4A</td>
<td>63</td>
<td>68</td>
</tr>
<tr>
<td>4B</td>
<td>54</td>
<td>67</td>
</tr>
</tbody>
</table>

B. van Ginneken, et. al. – unpublished data
ROC analysis Data Science Bowl test set

Lung RADS category | Non-cancer cases | Cancer cases
--- | --- | ---
1 | 66 | 2
2 | 138 | 4
3 | 28 | 10
4A | 63 | 68
4B | 54 | 67

B. van Ginneken, et. al. – unpublished data

Lung Cancer Screening Challenge II

https://lung-cancer.drivendata.org/

Progression of Imaging Tools to Clinical Studies
Conclusions

- Challenges useful in benchmarking and optimization of image analysis methods
- Drive algorithmic excellence and consensus
- Large public datasets essential to robust design and improved outcomes
- Access to challenge data and products may lead to more sustainable impact

Acknowledgments

- SPIE-AAPM-NCI challenge team
- NCI MICCAI challenge teams
- The Cancer Imaging Archive
- NCI Cancer Imaging Program