The Importance of QI in Data Science and Radiomics Initiatives

Bradley J Erickson, MD PhD

Disclosures

• Stockholder, CMO, FlowSIGMA, Inc
Radiomics...

- Is the extraction of quantitative data from images --wikipedia
- Can be thought of as the ‘features’ that have traditionally been used in Machine Learning

What is Data Science vis-à-vis Radiology?

- Multidisciplinary: Imaging, Medicine, Engineering, Physics, Computer Science, Statistics
- Focused on extracting and validating information and knowledge from images
- Radiomics is focused on the data extraction component
- Big data and data mining are very closely connected
What is the trajectory of QI (QIBA)?

- Consistent acquisition methods
- Good agreement in measured values for a phantom or tissue type
- Reproducible metrics of biologic interest
- Reproducible, repeatable, low noise, biologically relevant, clinically relevant
What is the trajectory of AI?

Audience Response Question: Are AI and QI:

A. Competitive
B. Coexistent
C. Complementary
Deep Learning: Why the Hype?

Performance in ImageNet Challenge

- Human: 100%
- Deep Learning: 74% (2010), 90% (2016)
- Traditional Machine Learning: 74% (2010), 80% (2016)
Benefit of DL vs Conventional ML

- Deep Learning **finds features** and connections vs just connections
  
  "Computers Programming Computers"

Deep Learning (like Machine Learning) Can Do:

- Regression (e.g. Bone Age)
- Segmentation (i.e. label pixels in an image)
- Classification (i.e. “This is pneumonia” vs normal or vs cancer)
- Reconstruction (e.g. MRI k-space to image)
Segmentation: Body Comp

- CT slice at L3 traced by 2 human experts. >1200 train, 200 test, then applied to all slices in abdomen

<table>
<thead>
<tr>
<th></th>
<th>Dice</th>
<th>Jaccard</th>
<th>TPF</th>
<th>FPF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deep-learning vs gold-standard</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAT</td>
<td>0.98 (0.01)</td>
<td>0.96 (0.02)</td>
<td>0.98 (0.01)</td>
<td>0.02 (0.01)</td>
</tr>
<tr>
<td>Muscle</td>
<td>0.95 (0.02)</td>
<td>0.91 (0.03)</td>
<td>0.95 (0.02)</td>
<td>0.05 (0.02)</td>
</tr>
<tr>
<td>Viscera</td>
<td>0.99 (0.01)</td>
<td>0.98 (0.01)</td>
<td>0.99 (0.01)</td>
<td>0.01 (0.01)</td>
</tr>
<tr>
<td>Bone</td>
<td>0.95 (0.01)</td>
<td>0.91 (0.01)</td>
<td>0.99 (0.01)</td>
<td>0.10 (0.01)</td>
</tr>
</tbody>
</table>

*Weston, Radiology, 2019*
### Individual Patient Body Comp Example

<table>
<thead>
<tr>
<th>Tissue Type</th>
<th>Volume (cm³)</th>
<th>Mass (kg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subcutaneous Fat</td>
<td>6,400</td>
<td>5.8</td>
</tr>
<tr>
<td>Muscle</td>
<td>8,010</td>
<td>8.5</td>
</tr>
<tr>
<td>Visceral Fat</td>
<td>3,500</td>
<td>3.2</td>
</tr>
<tr>
<td>Lean Tissue</td>
<td>7,900</td>
<td>8.4</td>
</tr>
<tr>
<td><strong>Total Abdominal</strong></td>
<td><strong>25,810</strong></td>
<td><strong>25.9</strong></td>
</tr>
</tbody>
</table>

- **34% Fat**  
- Fat is 35% visceral  
- **3.1% Skel Muscle**  

### Segmentation: CT Abdominal Organs & Tissues

- 200 CT abdomens for range of diseases (not normal!)  
- 180 traced by 1 human  
- *Weston, in preparation*
Segmentation: CT Abdominal Organs & Tissues

4-fold cross validation on 236 subjects

©2014 MFMER  |  3449756-17

*Wang, Med Image Anal, 2019
Segmentation: Brain

• Fully automatic segmentation of Brain
  • 132 parts of the brain semi-automatically labeled with human expert verification
  • 2000 train, 350 test

• Dice scores:
  • Mean: 0.954
  • Range: 0.852 – 0.989

*Korfiatis, CMIMI 2018

Segmentation: Polycystic Kidney Disease

• TKV is only FDA qualified image-based biomarker (9/15/2016)

*Kline, J Digit Im, 2017
In The (Near) Future, Medical Images Will Be **Routinely** Segmented/Quantified

• Within 2 years, many/ most MR images will be automatically segmented by deep learning
• Likely will be implemented on the scanner soon thereafter (3-5 years)
• This will lead to a revolution in imaging-based biomarkers
  • Volumes and ratios of volumes as biomarkers of disease or disease risk
  • Densities, textures, other properties of those organs that could not be easily and reliably measured

**DL for Classification**

• Is a disease/finding present
• Is this disease X or Y
• Is this progression or response
• Molecular/Genomic Properties
• ...
Traditional Biomarkers

- Intensity: If mildly hyperdense on non-contrast CT, iso/dark on T2, bright on contrast, then likely lymphoma
- ‘Texture’: If uniform intensity, then more likely lymphoma.

- ‘Textures’ can be challenging to agree on: Which is really ‘Ground Glass’?
Textures Can Be Important Imaging Biomarkers

Kline, Kidney Intl, 2017

Prediction of 4 Key Molecular Markers for Glioma

N=498 subjects using T2-weighted images from Mayo, UCSF, TCIA
No human input/segmentation—fully automatic!
398 for training, 100 for testing
50 layer ResNet

<table>
<thead>
<tr>
<th>Marker</th>
<th>Sens</th>
<th>Spec</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>IDH1</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
</tr>
<tr>
<td>1p19q Co-Del</td>
<td>0.91</td>
<td>0.85</td>
<td>0.87</td>
</tr>
<tr>
<td>ATRX</td>
<td>0.93</td>
<td>0.89</td>
<td>0.91</td>
</tr>
<tr>
<td>MGMT Methylation</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
</tr>
</tbody>
</table>

*Korfiatis, Submitted*
Image-Omics Beats GenOmics

- Tissue-based –omics doesn’t reveal:
  - other parts of the tumor
  - the phenotype
  - gene-gene interactions
  - host response

Textures: Caution needed

"Investigations of feature repeatability and reproducibility are currently limited to a small number of cancer types."

The image biomarker standardisation initiative
Deep Learning Based Reconstruction

- Acquire full dose/full time data set
- Recon the ‘normal’ image
- Give DL a limited dose/time dataset and that normal image

*Magic Happens*

- DL produces something ‘close’ to normal image

---

Deep Learning to Improve Images

[Zaharchuk, AJNR, 2018]
CNN vs Traditional Noise Reduction Methods

Deep Learning: Caution needed

The Dark Secret at the Heart of AI
No one really knows how the most advanced algorithms do what they do. That could be a problem.
by Will Knight  April 11, 2017
“Black Boxes”

Can I Trust It? eXplainable AI

• Deep Learning is NOT a black box anymore

*Saliency map

*Philbrick, AJR, 2018

*CAM / GradCAM
The Real Challenges to Deep Learning in Medicine

• High Quality Annotated data sets
  • Large numbers may not be as important as high quality
    • But both is best and that will happen
  • Privacy concerns limit sharing
    • But technology developments will address this
  • Annotation is a significant effort
    • But AI-based tools will make this ever more efficient
The Real Challenges to Deep Learning in Medicine

• High Quality Annotated data sets
• Diversity in data sets
  • There are known racial differences in disease appearance and response
  • But ‘bias’ can also be viewed as information, allowing more precise medicine
The Real Challenges to Deep Learning in Medicine

- High Quality Annotated data sets
- Diversity in data sets
- Tools
  - Tools are very available.
  - Much published without adequate rigor

The Real Challenges to Deep Learning in Medicine

- High Quality Annotated data sets
- Diversity in data sets
- Tools
- Imaging Community
  - “There’s no hypothesis”
Conclusions

• Deep Learning is and will continue to change the trajectory of Radiomics and Data Science
  • There is great opportunity if we accept ‘discovery science’
  • We must be careful about what features are used
• The demands on QI are more subtle and indirect
  • Need better tools to ‘mine’ DL networks
  • May need new techniques to implement textures into practice