Big Data Seminar on Imaging

February 18, 2020

Sponsored by the Biostatistics, Epidemiology, and Research Design (BERD) core of the CCTSI
Overview

• Introduction to Medical Imaging
  • Sarah M. Ryan, PhD Candidate

• Deep Learning and Its Applications in Medical Image Analysis
  • Fuyong Xing, PhD

• Application of Radiomics to Lung Disease to Develop a Novel Biomarker of Sarcoidosis
  • Lisa A. Maier, MD, MSPH, FCCP
Introduction to Medical Imaging
Imaging Basics

• An image is a visual representation of a mathematical function, say $f(x,y)$, where $f$ may give the intensity or color at point $(x,y)$

• Each 2D image is comprised of picture elements, called pixels, which form a matrix. In 3D, images are comprised of volumetric elements, called voxels
Medical Imaging Modalities

• Technologies that are used to view the anatomical areas in order to diagnose, monitor, or treat medical conditions

• Common imaging modalities include:
  • Microscopy
  • Radiography
  • Computed Tomography (CT)
  • Magnetic Resonance Imaging (MRI)

• Each modality uses different techniques to represent what we see
Microscopy

- A technique using properties of light to illuminate different cellular structures
- Results in a 2D image
  - Grayscale -> one number per pixel
  - RGB -> 3 numbers per pixel

https://www.niaid.nih.gov/diseases-conditions/coronaviruses

https://medium.com/@raycad.seedotech/convolutional-neural-network-cnn-8d1908c010ab
Radiography

• An imaging technique using x-rays, gamma rays, or similar ionizing (or non-ionizing) radiation to visualize internal anatomical structures
• Commonly used to detect pathology in bones, lungs, and breasts
• Produces a 2D grayscale image
Computed Tomography (CT)

- A computerized x-ray imaging procedure which generates cross-sectional images of the body at various angles
- Results in a 3D image with each voxel corresponding to the radiodensity of tissue, measured in Hounsfield units
Magnetic Resonance Imaging (MRI)

• An imaging procedure which measures the response of atomic nuclei to high-frequency radio waves when placed in a strong magnetic field

• Produces 3D images of internal anatomical structures, commonly brain
  • Structural -> Produces T1 and T2 3D images, corresponding to magnetization states
  • Functional -> Produces multiple 3D images over time
Image Processing

- Image processing techniques are necessary to enhance image features prior to analysis

- This may include:
  - Removal of scanner effect
  - Removal of physiological effects (e.g., breathing, heartbeat, movement, etc.)
  - Identification and segmentation of region of interest
  - Alignment of spatial coordinates across images (aka registration)

https://github.com/ryansar/lungct
Image Feature Extraction

- **Radiomics** - An emerging field in which large numbers of quantitative features are computed from medical images, providing a rapid, objective, and sensitive quantification of abnormalities.

Radiomic measures from chest high-resolution computed tomography associated with lung function in sarcoidosis
Deep Learning and Its Applications in Medical Image Analysis

Fuyong Xing
Department of Biostatistics and Informatics
Colorado School of Public Health
University of Colorado Anschutz Medical Campus
Big Image Data

ImageNet

Big Image Data

- ImageNet: a large-scale image dataset for computer vision
- 10,000,000+ labeled images, 20,000+ object categories
- Annual ILSVRC Challenges (up to 2017):
  - 1000 object categories
  - 1.2M training, 50k validation, and 100k testing images

How much data is generated every minute?

- Instagram: over 55000 photos shared

Biomedical Images Are Everywhere
Grand Challenges in Biomedical Image Analysis

https://grand-challenge.org/
What Methods Used to Analyze Large-Scale Image Data?

Image credit: Andrew Ng
Deep learning is a class of machine learning techniques that exploit multiple layers of non-linear information processing for supervised or unsupervised feature extraction and transformation, and for pattern analysis and classification.

Deep Fully-connected Neural Network

Input layer 1st hidden layer 2nd hidden layer Output layer

Cat
Convolutional Neural network (CNN)

Input

Convolution
Pooling
Convolution
Pooling
Fully-connected

Class probabilities

Image credit: Mathworks
Stacked Autoencoders

- Stacked autoencoders: stack multiple autoencoders to form a multi-layer neural network
- Typically trained in a layer-wise fashion: train one layer at a time
- Can be further fine-tuned in a supervised learning manner

An autoencoder

Figure credit: F. Xing et al. "Deep Learning in Microscopy Image Analysis: A Survey", IEEE TNNLS, 2018
Recurrent Neural Network (RNN)

\[ s_t = f(Ux_t + Ws_{t-1}) \]
\[ \sigma_t = g(Vs_t) \]

Generative Adversarial Networks (GANs)

Generative Adversarial Networks (GANs)

- **CycleGAN**

  ![CycleGAN Diagram](image)

  - **GAN** $G_{AB}$
  - **GAN** $G_{BA}$
  - **Discriminator** $D_B$
  - **L2 Loss**

  - Real Image in domain A
  - Fake Image in domain B
  - Reconstructed Image

  - $G_{BA}$ generates a reconstructed image of domain A.
  - This makes the shape to be maintained when $G_{AB}$ generates a horse image from the zebra.

- II. [Towards Data Science](https://towardsdatascience.com/image-to-image-translation-using-cyclegan-model-d58cfff04755)
Deep Learning Frameworks

- PyTorch
- MXNet
- Deeplearning4j
- Theano
- Caffe
- TensorFlow
Deep Learning in Medical Image Analysis

- Automated detection of objects, regions, landmarks, etc.
- Automated segmentation of organs, substructures, etc.
- Automated classification of objects, diseases, etc.
- Image registration
- Image retrieval
- Biomarker discovery
- Medical Imaging report generation
- More ...
G. Litjens et al., “A survey on deep learning in medical image analysis”, *Medical Image Analysis*, 2017
Image Classification

- CNN-based classification of HEp-2 cell images

**Object Detection**

- CNN-based individual nucleus/cell localization in pancreatic neuroendocrine tumor images

C: convolutional layers; M: max-pooling layers; F: fully-connected layers. The last layer has two units for binary classification.

Object Detection

- CNN-based mitosis detection in breast cancer histology images

Object Detection

- Nucleus/cell detection with fully convolutional networks

Y. Xie et al., “Efficient and robust cell detection: A structured regression approach”, Medical Image Analysis, 2018
Object Detection

- Nucleus/cell detection generative adversarial networks (GANs)

Image Segmentation

- CNN-based neuronal membrane segmentation electron microscopy images

Image Segmentation

- Cell segmentation with U-Net (an encoder-decoder network)

Image Segmentation

- Nucleus segmentation in histopathology images using stacked denoising autoencoders

H. Su et al., “Robust cell detection and segmentation in histopathological images using sparse reconstruction and stacked denoising autoencoders”, MICCAI, 2015
Object Recognition

- Nucleus recognition with fully convolutional networks

F. Xing et al., "Pixel-to-Pixel Learning With Weak Supervision for Single-Stage Nucleus Recognition in Ki67 Images", IEEE TBME, 2019
Nucleus classification with fully convolutional networks

Y. Zhou et al., “SFCN-OPI: Detection and fine-grained classification of nuclei using sibling FCN with objectness prior interaction,” AAAI, 2018
Image Retrieval

- Skeletal muscle image retrieval with CNNs
Image Retrieval

- Skeletal muscle and lung cancer image retrieval with CNNs

X. Shi et al., "Pairwise based Deep Ranking Hashing For Histopathology Image Classification and Retrieval," *Pattern Recognition*, 2018
Stain/Color Normalization

- Stain/color normalization with generative adversarial networks (GANs)

Stain/Color Normalization

- Stain/color normalization with generative adversarial networks (GANs)

A. BenTaieb and G. Hamarneh, “Adversarial Stain Transfer for Histopathology Image Analysis,” IEEE TMI, 2018
Text Generation

Text generation from pathology images
Link: https://www.youtube.com/watch?v=yy7NUrc3KI0

Thank You

- Web: https://fuyongxing.github.io
- Email: fuyong.xing@cuanschutz.edu
- Address:
  Department of Biostatistics and Informatics
  Colorado School of Public Health
  University of Colorado Anschutz Medical Campus
Application of Radiomics to Lung Disease to Develop a Novel Biomarker of Sarcoidosis

Lisa Maier, MD MSPH, Maierl@njhealth.org
Professor and Chief, Division of Environmental and Occupational Health Sciences
National Jewish Health, University of Colorado
**Sarcoidosis**

- Burden of disease unknown, but may be increasing
  - Increasing hospitalizations
  - Higher costs of hospitalizations vs non-sarc (Ungprasert Lung 2019)
- NHLBI Symposium “Levaraging Disparities..” (Maier AATS 2017)
  - Conduct “Omics” and systems biology research
  - Develop biomarkers of severe disease
  - Pulmonary, cardiac, neurologic disease

Gerke et al. BMC Pulm
Pulmonary Sarcoidosis

• Primary cause of death
• Impacts individuals in the prime of their lives (mean age at death = 55 years)
• Response to therapy variable
• No definitive prognosticators—genetic, genomics proposed
• Designing treatment and longitudinal studies is problematic

Swigris AJRCCM 2011
## Risk Factors for Disease Persistence

- CXR at presentation
  - Absence of lymphadenopathy
  - Ascending Scadding Stage
- Architectural distortion of the airways or cystic changes
- Multiple organ involvement
- Splenomegaly
- Need for systemic therapy
- Older age
- Female gender
- Black race

## Risk Factors for Clinically Bothersome Disease

- Worse CXR
- Dyspnea at time of diagnosis
- Need for treatment in 1st 6 months
- Multiple organ involvement
- Lower socioeconomic status
- Black race
Biomarkers Diagnosis and Prognosis

• No biomarkers for diagnosis
  • Based on biopsy
  • Diagnosis currently one of exclusion

• No biomarkers to prognosticate patients
  • No predictors severe phenotype
    • Remitting versus progressive types
  • Who will or will not respond to Rx
## HLA Class II: Ethnicity, Phenotypes and Prognosis

### HLA DRB1 Risk Groups

<table>
<thead>
<tr>
<th>HLA DRB1</th>
<th>Risk</th>
<th>Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>*0301</td>
<td>Lofgren’s</td>
<td>Swedish, Finnish, Dutch</td>
</tr>
<tr>
<td></td>
<td>Protective</td>
<td>AA</td>
</tr>
<tr>
<td></td>
<td>Resolving</td>
<td>Swedish</td>
</tr>
<tr>
<td>*1101</td>
<td>Overall Disease</td>
<td>US White, AA, Chinese</td>
</tr>
<tr>
<td></td>
<td>Persistent Disease</td>
<td>AA</td>
</tr>
<tr>
<td>*1201</td>
<td>Overall Disease</td>
<td>European, US, AA</td>
</tr>
<tr>
<td></td>
<td>Pulmonary</td>
<td>European, Japanese</td>
</tr>
<tr>
<td>*1401</td>
<td>Disease Risk</td>
<td>European, US, AA</td>
</tr>
<tr>
<td></td>
<td>Progressive pulm</td>
<td>European, Scandinavian</td>
</tr>
<tr>
<td>*1501</td>
<td>Disease Risk</td>
<td>European, US, AA</td>
</tr>
<tr>
<td></td>
<td>Progressive pulm</td>
<td>Scandinavian</td>
</tr>
</tbody>
</table>

### Lofgren’s Syndrome

![Bar chart showing comparison between DRB1*03 positive and negative groups](Grunewald AJRCCM 2009)

Grunewald AJRCCM 2009

Fingerlin, Hamzeh, Maier Clin Chest Med 2015
Blood Gene Transcripts Predict Remitting vs Progressive Sarcoidosis

- Su et al identified pathways associated with DE genes IFN, TCR, and CXCL9 in Sarc blood vs controls
- These gene factors act as predictors of disease course in a longitudinal Sarc cohort
- Initial high CXCL9 expression associated w/ more severe disease longitudinally

Su et al. Eur Resp J 2014
Scadding Staging

- Prognostic value for remission
- Easy to use

- Not linear
- Interobserver variability
- Low sensitivity/specificity for fibrosis
- Correlation with lung function variable
Variable Manifestations of Sarcoidosis on HRCT
Can information from HRCT better inform disease classification and prognosis?

• Radiomics – extracts large quantities of texture and related measures from medical images
  
• Good for detecting differences in CT patterns; lung cancer, COPD

Hypothesis: Detailed radiomic analysis of lung CT images in sarcoidosis will identify new more refined subtypes of pulmonary disease

Study Design:
• Exploratory case-control study using NJH GRADS sarcoidosis cases (N=73) and COPD Non-smoking healthy controls (N=78)
• Classification study using only GRADS sarcoidosis cases (N=321)
Normal

Nodules Coalescing

Stage 4

Scarring Mosaic changes

(A)

(B)

(C)
### Differences in Radiomic Measures between Sarcoidosis & Controls and Scadding Stage

<table>
<thead>
<tr>
<th>Radiomic Feature</th>
<th>Control (N = 78)</th>
<th>Sarcoidosis (N = 73)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skewness</td>
<td>3.615 (0.037)</td>
<td>3.252 (0.064)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>16.12 (0.299)</td>
<td>13.0 (0.451)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Fractal D</td>
<td>2.269 (0.005)</td>
<td>2.236 (0.005)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Moran's I</td>
<td>0.815 (0.002)</td>
<td>0.838 (0.003)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Geary's C</td>
<td>0.135 (0.002)</td>
<td>0.120 (0.002)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

*adj for age, gender, BMI

Ryan ERJ 2019
### Differences in Radiomic Measures by Scadding Stage

<table>
<thead>
<tr>
<th>Radiomic Feature</th>
<th>Stage 0 (N = 9)</th>
<th>Stage I (N = 8)</th>
<th>Stage II (N = 28)</th>
<th>Stage III (N = 11)</th>
<th>Stage IV (N = 17)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skewness</td>
<td>3.356 (0.099)*</td>
<td>3.631 (0.072)*</td>
<td>3.301 (0.103)*</td>
<td>3.479 (0.095)*</td>
<td>2.791 (0.150)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>13.71 (0.785)*</td>
<td>16.06 (0.521)*</td>
<td>13.33 (0.769)*</td>
<td>14.65 (0.850)*</td>
<td>9.57 (0.793)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Fractal D</td>
<td>2.259 (0.011)*</td>
<td>2.246 (0.008)*</td>
<td>2.234 (0.006)*</td>
<td>2.254 (0.015)*</td>
<td>2.210 (0.010)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Moran's I</td>
<td>0.827 (0.008)*</td>
<td>0.816 (0.003)*</td>
<td>0.840 (0.005)*</td>
<td>0.819 (0.007)*</td>
<td>0.863 (0.007)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Geary's C</td>
<td>0.131 (0.005)*</td>
<td>0.126 (0.004)*</td>
<td>0.117 (0.003)**†</td>
<td>0.136 (0.007)*</td>
<td>0.105 (0.005)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

* vs Stage IV; † vs Stage III
Differences in Spatial Radiomic Measures Between Sarcoidosis and Controls
Spirometry is Associated with Radiomic Measures

- Associations stronger and statistically associated with sarcoidosis compared to controls

- Radiomics explained 70.3% of the variability in FVC vs 51.4% for Scadding staging alone

Ryan ERJ 2019
Developing New Radiomic Phenotypes

- Entire GRADS Cohort
- Used first and second order radiomic features from each lung
- Corrected for Scanner Effect
- K-means variable selection clustering- for high dimensional, dependent data
- Gap statistic to determine the optimal number of clusters n=5
- Describe clinical aspects of clusters: Lung Function, Qs, Biomarkers
- Compare to Scadding Stage
## Cluster Analysis of Radiomic Features: 5 clusters

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Overall</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>321</td>
<td>52.91 (9.91)</td>
<td>53.28 (9.96)</td>
<td>49.76 (10.97)</td>
<td>52.49 (10.36)</td>
<td>52.56 (8.86)</td>
<td>55.12 (8.92)</td>
<td>0.043</td>
</tr>
<tr>
<td>Age</td>
<td>147 (45.8)</td>
<td>44 (53.0)</td>
<td>20 (37.0)</td>
<td>26 (44.8)</td>
<td>17 (37.8)</td>
<td>40 (49.4)</td>
<td>0.286</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.005</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>233 (73.0)</td>
<td>72 (87.8)</td>
<td>42 (77.8)</td>
<td>38 (66.7)</td>
<td>34 (75.6)</td>
<td>47 (58.0)</td>
<td>0.006</td>
</tr>
<tr>
<td>Black</td>
<td>77 (24.1)</td>
<td>8 (9.8)</td>
<td>12 (22.2)</td>
<td>16 (28.1)</td>
<td>10 (22.2)</td>
<td>31 (38.3)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>9 (2.8)</td>
<td>2 (2.4)</td>
<td>0 (0.0)</td>
<td>3 (5.3)</td>
<td>1 (2.2)</td>
<td>3 (3.7)</td>
<td></td>
</tr>
<tr>
<td>Height</td>
<td>66.98 (4.16)</td>
<td>67.83 (4.55)</td>
<td>66.72 (4.04)</td>
<td>66.43 (3.93)</td>
<td>66.76 (4.09)</td>
<td>66.81 (3.98)</td>
<td>0.293</td>
</tr>
<tr>
<td>BMI</td>
<td>30.62 (6.49)</td>
<td>32.89 (6.04)</td>
<td>29.36 (6.84)</td>
<td>32.51 (6.33)</td>
<td>26.88 (5.98)</td>
<td>29.84 (5.89)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Scadding</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Demographics ordered from least severe (0) to most severe (4) based on FVC percent predicted.
Clinical Aspects of “Radiomic Biomarker” Class

• Scadding Stage similar associations as to Radiomic Phenotypes
• But Scadding not distributed across Clusters equally
• It may be a surrogate for visual scoring of CT
Future Directions: Integrated Biomarker of Sarcoidosis

- GRADS Cohort
  - Visual Scores - 2 readers
  - BAL transcriptome
  - GWAS data
- Validation cohort NJH and CCF
- Determine function over time
Conclusions:

• Radiomic analyses of Sarcoidosis Chest CT:
  • Differentiates cases and controls
  • Associate with lung function and PROs
  • Classify sarcoidosis subtypes and may be useful as a biomarker

• Potential as a biomarker to augment and or standardize use of CT in clinic

• Integrative Omics approaches show promise

• There is now publicly-available, open-source software to perform many of these analyses for your own data
Processing Chest CTs

Software: `lungct` R package
https://github.com/ryansar/lungct

Functionality
1. Convert DICOM to NIfTI
2. Resample to 1x1x1 mm³
3. Segment the left and right lungs
4. Calculate radiomic features on 2D slice or 3D lung
5. Register lungs to the standard lung template
6. Create study-specific templates
7. Calculate the amount of lung shrinkage
8. And more!
Acknowledgements

National Jewish Health
• Tasha Fingerlin, PhD
• Clara Restrepo, MD
• Li Li, MD PhD
• Nabeel Hamzeh, MD
• Shu-Yi Liao MD, ScD
• Andy Levy, MD
• Peggy Mroz, MSPH
• Bri Barkes, MPH
• Christina Riley
• May Gillespie
• Kristyn MacPhail
• Jill Elliott

UCD
• Nichole Carlson, PhD
• Sarah Ryan, MS, PhD Student

CCF
• Dan Culver, MD

NIH:
1R01 HL114587
1R01 HL142049
UL 1TR002535
U01 HL112695, U01 HL112707

Our patients!