MGH & BWH CENTER FOR
CLINICAL DATA SCIENCE

Quantitative Imaging in Artificial Intelligence Applications

Katherine P. Andriole

July 30, 2018

Disclosures

Katherine P. Andriole is the Director of Research Strategy and Operations at the MGH & BWH Center for Clinical Data Science (CCDS).

The CCDS is funded in part by monies and resources from Nvidia Corporation, General Electric Healthcare and Nuance.

OUTLINE

- Machine Learning – Clinical & Research Uses
  - Enabling Factors & Existing Limitations
- Tools Needed for Machine Learning
  - Processing Pipeline – Example Tools
- QI and Machine Learning Examples
- Current State Summary

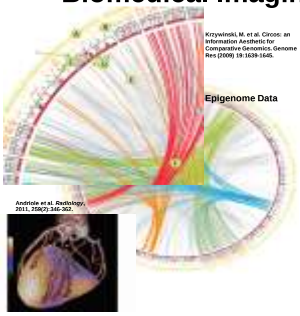
OUTLINE

- MACHINE LEARNING – CLINICAL & RESEARCH USES
  - ENABLING FACTORS & EXISTING LIMITATIONS
- TOOLS NEEDED FOR MACHINE LEARNING
  - PROCESSING PIPELINE – EXAMPLE TOOLS
- QI AND MACHINE LEARNING EXAMPLES
- CURRENT STATE SUMMARY

Deep Learning in Imaging

- **DL ⊆ ML ⊆ AI**
- **CNN based upon the human brain / Neurons**
- **DL algorithms "learn" discriminatory features that best predict an outcome**
  - Detect (Tumor Present / Absent)
  - Classify/Localize/Predict (Benign / Malignant)
  - Data-Driven versus CAD (Human-Defined Features)
- **Requires large amounts of Data**
- **Computationally Intensive**

Biomedical Imaging Big Data



Krzywinski, M. et al. Circos: an Information Aesthetic for Comparative Genomics. Genome Res (2009) 19:1639-1645.

Epigenome Data

Andriole et al. *Radiology*, 2011, 259(2):346-362.

http://www.siam.org/meetings/sdm13/sun.pdf

http://fora.tv/2012/02/13/Sebastian_Seung_Connectome

Healthcare Big Data

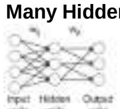
- Structured EHR Data
- Unstructured Clinical Notes & Reports
- Medical Imaging Data
- Genetic Data
- Behavioral & Social Data
- Epidemiological & Evidence-Based Practice Data
- Mobile Transducers

Enabling Factors: Compute Infrastructure

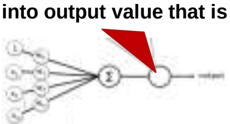
Nvidia 8-GPU DGX1
↑960 TFLOPS and ↑Memory 128GB

Enabling Factors: Algorithms

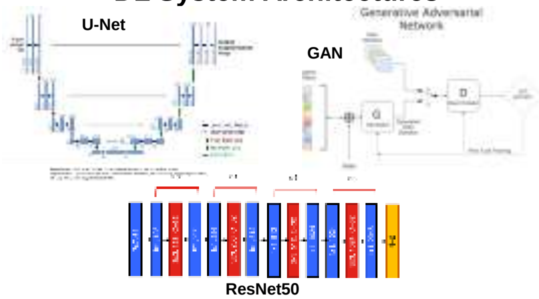
Deep = Many Hidden Layers



Activation Functions converts weighted sum of inputs into output value that is passed to next layer nodes



DL System Architectures



~~You work at a healthcare institution, you have compute, now just build a model, right?~~

Steps Involved in Machine Learning Algorithm Development & Translation into the Clinical Arena

- Clinically Relevant Question
- Data Cohort Definition
- Dataset Collection / Acquisition
- Data Cleaning, Normalization and De-Identification
- Dataset Annotation: Report and Pixel Labeling
- Model Building, Training, Validation and Testing
- ML Result Integration into the Clinical Workflow
- Continuous Learning

Limitations: Data Issues

- Data Access
- Patient Privacy
- Data Security
- Patient Cohort Makeup
  - Dataset Heterogeneity
  - Range of Severity
- Integrity, Curation, Normalization

Limitations: Data Issues

- Missing/Sparse Data
- Unstructured Data
- Multi-Scale Data
- Complex Data
- Longitudinal Data
- Noisy Data

Other Challenges for Imaging

- Lack of
  - Standard Image Acquisition Protocols (e.g., slice thickness, reconstruction kernel, tube current, with contrast)
  - Standard Training Data Sets
  - Uniformity Across Different Algorithms
  - Uniformity Across Vendors, Models, Versions
- Imaging or Device Artifacts
- Is the Data Rendered or Raw, Pre-processed, Compressed, Filtered or Thresholded?

Healthcare Big Data Special Issues

- Data File Size
- Raw Data often discarded FIFO
- Images are Not Labeled / Annotated and this is a difficult task
- Non-image Metadata

Annotating Medical Images

- Even "easy" annotations such as entire organs are subjective
  - Intra- and Inter-reader Variability
- Unclear, non-objective Gold Standard
- Tools often manual & time consuming

Study Annotation via NLP of Report

- Natural Language Processing (NLP)
- Largely Unstructured Free Text
- Variable in Format, Prose
- Qualitative versus Quantitative
- Often ambiguous terms and tone

How Much Data is Required?

It Depends!

- How variable is your data
- Supervised versus Unsupervised ML
- How "good" are your annotations
- What is the task

Mitigation

- Data Augmentation
- Transfer Learning

Things to Watch

Overfitting

- Network learns the specific examples in the training set

Mitigation Methods

- **Data Cohort Selection**
- **Holdout Test Set:** **Train-Validate-Test Sets**
- **Model Development (# features, # layers)**
- **Dropout Regularization:** **Randomly remove subset of network nodes during each training epoch**
- **Batch Normalization and Data Augmentation**: **to boost the size/variability of training set**

BJ Erickson, *et al*. Deep Learning In Radiology: Does One Size Fit All? *JACR* 2017.

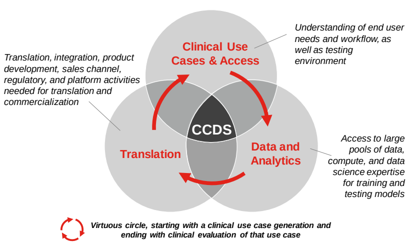
Delivery of ML Output to Clinical Arena

- **Need Standards for**
  - Label Formats (eg, Binary, ROI Masks, Quantitative Metric)
  - Machine Learning Output Formats
  - Result Delivery to Point-of-Care Systems
  - Integration into Clinical Systems
  - Visualization and GUI
  - Machine Learning Output Archival (or regenerate on the fly)
- **DICOM, FHIR, HL7, but…**

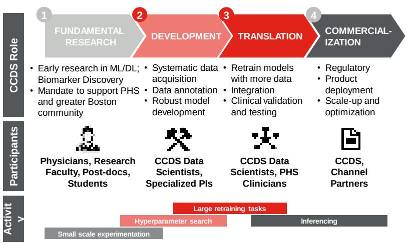
OUTLINE

- Machine Learning – Clinical & Research Uses
  - Enabling Factors & Existing Limitations
- Tools Needed for Machine Learning
  - Processing Pipeline – Example Tools
- QI and Machine Learning Examples
- Current State Summary

Multi-Disciplinary Core Capabilities



The Research & Development Pipeline



Large Datasets Require Infrastructure Investment

State-of-the-art Machine Learning algorithms require massive datasets

Required Tooling

- **Identify Cases –** Queryable Report Storage
- **Label Reports –** NLP, Annotation Pipeline
- **Collect Image Data –** Research VNA
- **Label Images –** Visualization / Annotation Tools
- **Normalize Data –** Normalization Tooling
- **Train Models –** GPU Cluster
- **Deliver Results –** Integration / Visualization Tools

Whitepaper at ccds.io

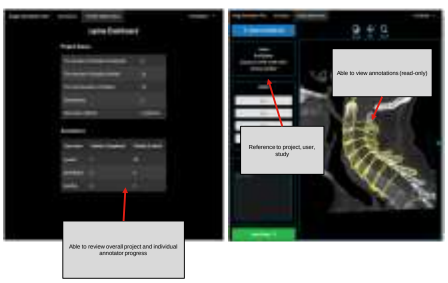
OUTLINE

- Machine Learning – Clinical & Research Uses
  - Enabling Factors & Existing Limitations
- **Tools Needed for Machine Learning**
  - Processing Pipeline – **Example Tools**
- QI and Machine Learning Examples
- Current State Summary

Report Annotation Tool

Image Annotation Tool

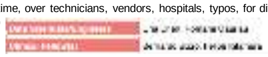
User Management / Project Supervision



Automated MRI Brain Sequence Selection

Why is it challenging? **Today we have to deal with this…**

- Protocol-related description that can be modified as free text by the technologist
- Mixes information of very different nature
- Very variable, Not reliable (changes over time, over technicians, vendors, hospitals, typos, for different MR sequence types, acronyms, abbreviations)



Use Machine Learning to Solve the Problem

Information from DICOM Header
Preliminary Experiment on Dataset of 32,000 Exams

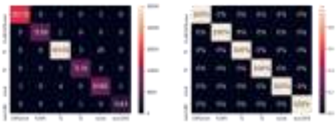
Selection of a small subset of Series Descriptors (~400 – most frequent)

Radiologists Annotation (T1, T2, FLAIR, Diffusion, susceptibility/GRE, scout)

Decision Tree Classifier training (per series)

Results

Confidence Matrices on Test set: 72,957 images from 6,570 exams



- **High detection accuracy 99.94%**
- **Validates the hypothesis: DICOM holds valuable information for MRI sequence selection**

Transferable to other anatomies with minimum adaptation

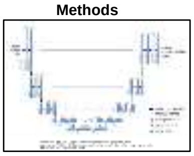
OUTLINE

- Machine Learning – Clinical & Research Uses
  - Enabling Factors & Existing Limitations
- Tools Needed for Machine Learning
  - Processing Pipeline – Example Tools
- **QI and Machine Learning Examples**
- Current State Summary

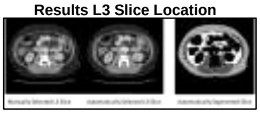
Body Comp: Measuring Muscle and Fat Segmentation in CT

Team: Chris Bridge, Brad Wright, Gopal Kotecha, Michael Rosenthal, Florian Fintelmann, Katherine Andriole
Objective: Develop ML to locate L3 slice in CT CAP; segment muscle and fat; measure volume of each
Background: Amount/Distribution of muscle mass and subcutaneous/visceral fat is a health Biomarker; Enable Population Health Research and Precision

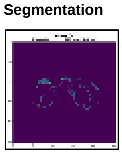
Methods



Results L3 Slice Location



Results Segmentation

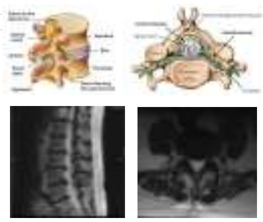


Project Demonstration: DeepSPINE

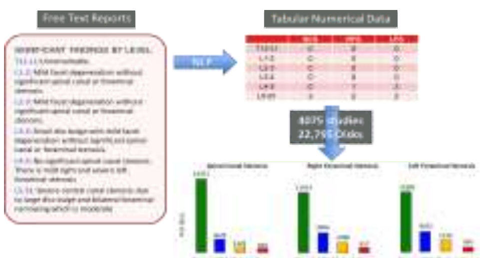
Objective	Automated vertebral segmentation, disk level labeling, and level-by-level stenosis grading for MRI of lumbar spine performed for degenerative disease
Data Scientists	Jen-Tang Lu, Stefano Pedemonte, Brad Wright, Chris Bridge
Software Engineers	Sean Doyle, Mark Walters
Clinical Fellow	Bernardo Bizzo
Clinical Champion	Stuart R. Pomerantz

Background – lumbar spinal stenosis

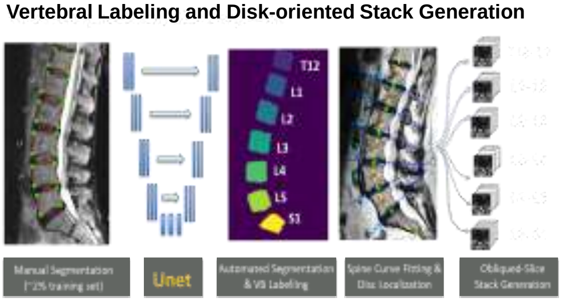
- Major cause of low back pain (global prevalence can be as high as 42% according to WHO).
- Prevalent diagnostic tool: MRI
- Time consuming, costly, and high inter-reader variability



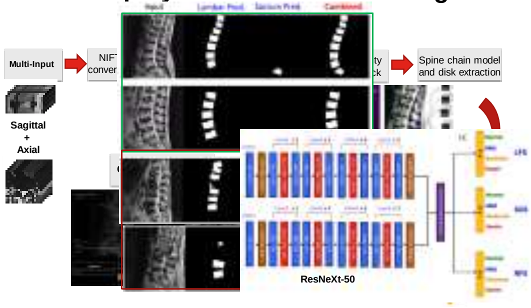
Cohort Creation and Annotation



Vertebral Labeling and Disk-oriented Stack Generation



Model Deployment: End-to-End Program

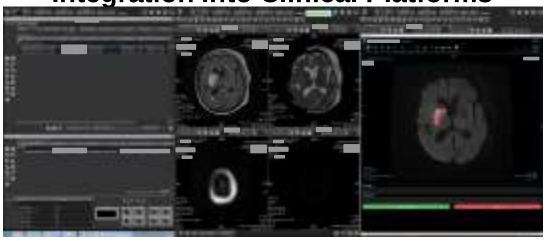


Metrics and Published Work

	Zhang et al. (2017)	Jamaludin et al. (2017a)	CCDS
Type of scan	Axial	Sagittal	Axial + Sagittal
Spinal canal stenosis (%, mean ± std)			
L3-L4	87.2 ± 3.2	94.7	94.5 ± 0.7
L4-L5	85.1 ± 3.4	85.9	95.3 ± 0.2
L5-S1	87.5 ± 3.3	93.7	99.1 ± 0.5
Foraminal stenosis (%, mean ± std)			
L3-L4	84.3 ± 3.9	N/A	94.0 ± 0.7
L4-L5	84.0 ± 4.0	N/A	89.0 ± 1.4
L5-S1	87.1 ± 3.4	N/A	91.2 ± 1.6

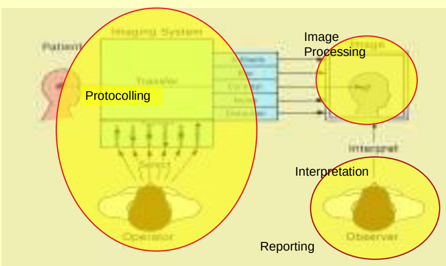
Integration into the Radiologist’s Workstation

Integration Into Clinical Platforms



- Real-time Inference
- GUI Feedback
- Continuous Learning

Medical Imaging Chain



OUTLINE

- Machine Learning – Clinical & Research Uses
  - Enabling Factors & Existing Limitations
- Tools Needed for Machine Learning
  - Processing Pipeline – Example Tools
- QI and Machine Learning Examples
- Current State Summary

Current State

- Data Access, Patient Privacy
- Data Cohort Selection
- Data Cleaning, Preprocessing, Data Annotation
- Clinical Relevance
- Lack of Standards for Data Acquisition, Annotation Labels, ML Output, Clinical Workflow Integration
- Narrow AI