Future of AI in healthcare imaging

Tanveer Syeda-Mahmood
IBM Fellow
Chief Scientist
Medical Sieve Radiology Grand Challenge
IBM Research - Almaden
AI over the generations has evolved
Machine learning for medical imaging

• Several players (besides Watson) now for medical imaging
  – RadLogics (web-based CAD)
  – Metamind (now moved to SalesForce)
  – Healthmyne (user-assisted reporting)
  – Enlitic (Lung nodules, fracture detection)
  – Zebra vision (osteoporosis, fatty liver and emphysema)
  – Riverrain (image processing of chest CTs and X-rays)
  – Caradigm (microsoft algorithms in GE)
  – Arteryx (cloud-based image analysis on cardiac MRI)
  – Major PACS vendors (Fuji, Siemens, Phillips, etc.) all have some form of learning
Where do I see AI going in future for radiology

• There is more to AI than machine learning
  – The deep learning buzz is not the only thing
  – Role of knowledge and reasoning

• Large-scale data and validations still needed for machine learning
  – Large-scale ground truthing a problem

• Define a new Man-machine interface
  – Man + Machine > Man, Machine
  – Many applications with machine assistance
A case where deep learning is helpful

• First approach to automatically detect body position from CT imagery.

1) thoracic inlet region, 2) lung apex, 3) origin of great vessels 4) aortic arch, 5) ascending/descending aorta 6) pulmonary trunk, 7) aortic valve/aortic root 8) axial 4-chamber view 9) long axis 2-chamber view.

<table>
<thead>
<tr>
<th>feature group</th>
<th>margin0 acc.</th>
<th>margin1 acc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>HoG</td>
<td>64.7%</td>
<td>90.4%</td>
</tr>
<tr>
<td>LBP</td>
<td>43.9%</td>
<td>77.1%</td>
</tr>
<tr>
<td>Haar</td>
<td>61.3%</td>
<td>89.4%</td>
</tr>
<tr>
<td>convolution layer 1</td>
<td>72.3%</td>
<td>96.2%</td>
</tr>
<tr>
<td>convolution layer 2</td>
<td>53.0%</td>
<td>86.3%</td>
</tr>
<tr>
<td>fully connected layer</td>
<td>60.5%</td>
<td>90.8%</td>
</tr>
<tr>
<td>Regressed Label</td>
<td>91.7%</td>
<td>98.8%</td>
</tr>
</tbody>
</table>

• 595 images from 75 patients, five-fold cross validation
Body position recognition using a combination of clinical and deep learned features

Uses handcrafted and CNN features through a multiple classifier approach, followed by linearly combining the output label of classifiers.

Pre-trained convolutional neural network

Histogram of Gradients (HoG)
Local binary features (LBP)
Haar-like features

One-vs-all SVM 1
One-vs-all SVM 2
One-vs-all SVM 3
One-vs-all SVM 4
One-vs-all SVM 5
One-vs-all SVM 6

Linearly combine all the discrete labels $l_n$ to get a final label $L_n$

$$L_n = \frac{(a_1 \times h_{\text{HoG}} + a_2 \times h_{\text{LBP}} + a_3 \times h_{\text{Haar}} + a_4 \times f_{\text{conv1}} + a_5 \times f_{\text{conv2}} + a_6 \times f_{\text{FC}}) \sum a_k}{\sum a_k}$$

Cross validation on the training data to estimate coefficients $a_i$ that maximize accuracy.
A case where simple learners suffice with better computer vision algorithms

- Automatically identify heart failure
  - Heart failure, Hypertrophy, Normal
Detecting heart failure

Global Features – current clinical measures

<table>
<thead>
<tr>
<th>Diastolic Volume (EDV)</th>
<th>Systolic Volume (ESV)</th>
<th>Stroke Volume (SV)</th>
<th>Ejection Fraction</th>
<th>Cardiac Output</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Diastolic Volume" /></td>
<td><img src="image" alt="Systolic Volume" /></td>
<td><img src="image" alt="Stroke Volume" /></td>
<td>( \frac{SV}{EDV} \times 100 )</td>
<td>( SW \times HR \times HR )</td>
</tr>
</tbody>
</table>

- Global values are insufficient to distinguish diseases that affect only a portion of the myocardium
  - Local functional changes
  - Local wall width changes
- Calculate local strain and wall width in American Heart Association (AHA) defined regions to aid disease discrimination
  - Automatic method for region extraction

- 45 Patients
  - 24 heart failure
  - 10,540 Images analyzed
  - Leave-one-out cross-validation
  - 86.7% Classification Accuracy

Automatically extracted AHA regions
Role of knowledge

• Authentic knowledge sources
  • Over 70 vocabularies
    • UMLS MetaThesaurus, SNOMED-CT, FMA, RadLex, BSPO, NCI, MeSH, RxNorm, NDFRT, ICD9, ICD10, etc.
  • Over 8.5 million concepts
  • Over 10 million relationships
  • Specialty tagging for over 100,000 concepts
  • Measurement ranges for over 5000 measurements
  • Ontological tracing of concepts in several ontologies

• Knowledge assertions
  • Clinician-generated (including probabilistic scores)
    • Breast radiology (1000)
  • Semi-automatically generated Cardiac radiology (3000)
  • NLP + relationship detection for advanced understanding of n-ary assertions

• Procedural knowledge
  • 200 Differential diagnosis flowcharts
  • Patient management flowcharts
Suspected pulmonary embolism

Assess hemodynamic stability

Instable

Assess clinical probability of PE

Low/intermediate

Perform D-Dimer Testing

Elevated

Perform multidetector CT scan

Normal

PE ruled out

High

PE confirmed

Stable

Assess whether patient is critically ill

No

Check availability of multidetector CT

No

Transthoracic or transesophageal echocardiography

Yes

Positive

Right Ventricular Dysfunction

PE Confirmed

No
Putting it all together – Cognitive Assistant

• Advanced AI
  • Machine-learning driven multimodal analysis
  • Large clinical knowledge
  • Advanced multimodal reasoning
  • Knowledge-and-Informatics driven clinical decision support

• Project Avicenna
Avicenna – Clinical Summary

Vitals
- HR: 115
- BP: 125/82
- RR: 30
- O2 Sat: 93%
- Height: 142 cm
- Weight: 56.2 kg
- BMI: 1.6

Summary
- **History of Present Illness**
  1. Pleuritic right lower lobe chest pain
  2. Shortness of breath
- **Chief Complaints**
  - Shortness of breath
- **Physical Examinations**
  1. Physical exam 2015
- **Current Diagnoses**
  - None
- **Current Medications**
  - None
- **Past Hospitalizations**
  - None
- **Past Surgeries**
  - None
- **Allergies**
  - None
- **Family History**
  - None
- **Patient Management**
  - EKG, CXR, Tropo
- **Derived Measurements**
  - Pulmonary Angiogram
- **Laboratory Results**
  - D-dimer: 507
  - HCT: 44
  - Hgb: 14
  - K: 4.3
  - Na: 130
  - Platelets: 356
  - WBC: 8.2

Imaging
- Pulmonary Angiogram
- Chest X-ray

Evolving Differential Diagnosis
- Myocardial Infarction 410.5
- Pulmonary Embolism 415.1
- Chronic Obstructive Pulmonary Disease 518.0
- GERD 530.81
- Esophageal Spasm 530.3

Problem List
- Chronic Plantar Fasciitis
Avicenna – Finding similar cases
Avicenna - Reasoning

Findings from Automatic Analysis of Angiograms

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Embolism Type</td>
<td>Saddle</td>
</tr>
<tr>
<td>Embolism Location</td>
<td>Right and left main pulmonary arteries</td>
</tr>
<tr>
<td>Number of Embolisms</td>
<td>2</td>
</tr>
</tbody>
</table>

Diagram showing relationships among different medical conditions, including Pulmonary Embolism, GERD, Myocardial Infarct, and other related conditions.
A new use of AI in ground truth labeling

• Machine learning needs lots of labeled data
• Getting labeled data in imaging is difficult as regional isolation needed.
• **MedNet** is a platform for crowd-sourcing labeling
• AI can lead to semi-automatic ground truth labeling
  • Using machine learning for training
MedNet
Spot disease vocabulary terms in text reports

Extract visual features from relevant regions within images

Assign ambiguous labels to image

Learn correlation using ambiguous label learning

<table>
<thead>
<tr>
<th>Images attempted for automatic labeling</th>
<th>Original ambiguity label size</th>
<th>Maximum labels per image</th>
<th>Average labels per image before learning</th>
<th>Average labels per image after learning</th>
<th>% of learned labels manually verified to be correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>5641</td>
<td>36</td>
<td>13</td>
<td>3.58</td>
<td>1.03</td>
<td>92.3%</td>
</tr>
</tbody>
</table>
AI applications

• Radiologists
  • Automatic triage/peer review
    • Reduce workload through machine-based triage (e.g. in telemedicine or ER situations)
    • Increase the cases that are peer-reviewed through automation (go to 100% peer review one day)
  • Machine-assisted second opinions
    • Benefit from consensus opinions of other clinicians on similar cases
  • Automated reporting
    • Prefilled from machine generated image-based findings
  • Refined diagnosis from clinical summaries
    • Multimodal data-reduced radiologist-specific summaries (clinical, image summaries)
  • Automatic tumor tracking
    • Longitudinal tracking of studies (e.g. tumor growth/shrinkage)

• Clinicians
  • Snapshot summaries of patient conditions
    • Chief complaint-driven summaries
  • Key performance indicator tracking
    • Longitudinal tracking of key performance indicators (heart rate, blood pressure, cholesterol)
  • Differential diagnosis
    • Based on expert opinions and a priori clinical knowledge
  • Recommendation generation
    • Treatments, outcomes, referrals
AI Applications

• CIO/Back office
  • Risk identification
    • Identify patients at risk for diseases
  • Auditing and compliance
    • Spotting deviations, inconsistencies, fraud (diagnosis, billing)
  • Uncover major trends
    • Uncover major trends in the population for key findings in imaging
  • Peer review
    • Compare performance of radiologists

• Clinical Research and education
  • Compare effectiveness of treatments
    • Correlations of diseases to treatments and outcome
  • Predictive analytics
    • Likelihood of disease development
  • Developing new guidelines for treatment
    • Overall improvement in quality of care
    • Collaborate with standards bodies
  • Enhance radiology curriculums
    • Equip curriculum in future with informatics-derived conclusions
    • Radiology training with big data
Iaso: A case study of missed diagnosis

• Diagnosis propagation errors are rampant in electronic health in radiology/cardiology workflows
  • A report is filed mentioning the disease, it doesn’t make it to the EMR
  • An echocardiographer makes measurements, it doesn’t make it into the report
  • An echocardiographer measures and forgets to save the measurement screens in the video
  • A cardiologists doesn’t see obvious symptoms of a disease and doesn’t order the relevant study.

• AI to the rescue
  • Discover evidence from the data retrospectively and flag those cases
Identifying patients at risk for aortic stenosis study (MICCAI’16)

- Extract disease measurements from text reports
- Extract valve measurements from echo
- Extract valve measurements by echocardiographer
- Learn correlation RF label learning
### Iaso study results

**Table 1.** False discovery rate (FDR) of disease (AS) and measurement (peak velocity and mean gradient) detection.

<table>
<thead>
<tr>
<th>FDR</th>
<th>False positives</th>
</tr>
</thead>
<tbody>
<tr>
<td>AS</td>
<td>2/191</td>
</tr>
<tr>
<td></td>
<td>Indication/Hx: EVAL FOR MS/MR, AS/Al</td>
</tr>
<tr>
<td></td>
<td>De-Identified AS SMOKER</td>
</tr>
<tr>
<td>Peak velocity</td>
<td>1/364 aortic stenosis is present. The aortic valve peak velocity is 2.6 ← 9 m/s, the peak gradient is 28.9 mmHg.</td>
</tr>
<tr>
<td>Mean gradient</td>
<td>0/410</td>
</tr>
</tbody>
</table>

**Table 2.** Accuracy of Doppler envelop extraction and measurement calculation.

<table>
<thead>
<tr>
<th>Measurement made</th>
<th>Images tested</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>( V_{max} )</td>
<td>1054</td>
<td>0.29±/−0.78m/sec</td>
</tr>
<tr>
<td>( M_s )</td>
<td>785</td>
<td>0.08±/−10.05mmHg</td>
</tr>
</tbody>
</table>

**Table 3.** Comparative performance of rule-based baseline and random forest with features extracted from structured information, reports, images, and OCR text. \( \min(I, O) \) refers to the fusion of image and OCR features by taking the minimum of the two for each individual feature/parameter.

<table>
<thead>
<tr>
<th>Features</th>
<th>Structured Report Image OCR min(I,O)</th>
<th>Precision</th>
<th>Recall</th>
<th>F-score</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>0.84</td>
</tr>
<tr>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>1.00</td>
</tr>
<tr>
<td>Random Forest</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>0.96</td>
</tr>
<tr>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>0.94</td>
</tr>
<tr>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>0.94</td>
</tr>
<tr>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>0.78</td>
</tr>
<tr>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>0.93</td>
</tr>
<tr>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>0.82</td>
</tr>
<tr>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>0.96</td>
</tr>
<tr>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>0.87</td>
</tr>
</tbody>
</table>

**Results Summary (547 Aortic Stenosis patients, 23% are new discoveries)***

- Found 547 patients out of 1129 with aortic stenosis, of which 23% were new discoveries from unstructured text and image analysis.
- 136 newly discovered patients, where 97 were identified from reports and 39 from images.
Summary

- Tremendous progress made in AI for medical imaging in recent years
- Main challenges remain:
  - Large scale data for training models and testing
  - Ground truth annotations to drive the training and testing
- Many new problems open up the field for active research