Clinical Decision Making Using Deep Learning

Bradley J Erickson, MD PhD
Mayo Clinic

Disclosures

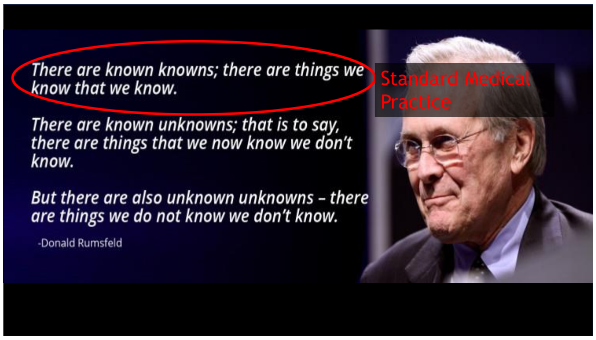
- Grant Funding: DK-90728
- Commercial: FlowSIGMA, VoiceIT, OneMedNet

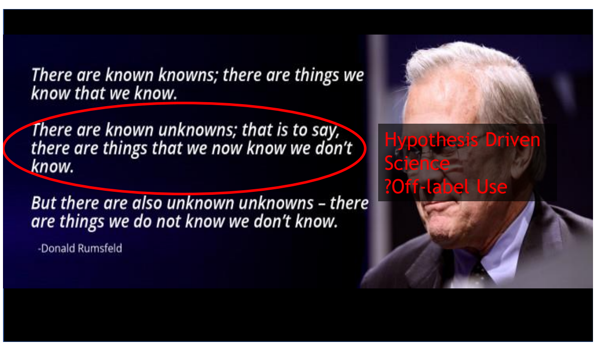
There are known knowns; there are things we know that we know.

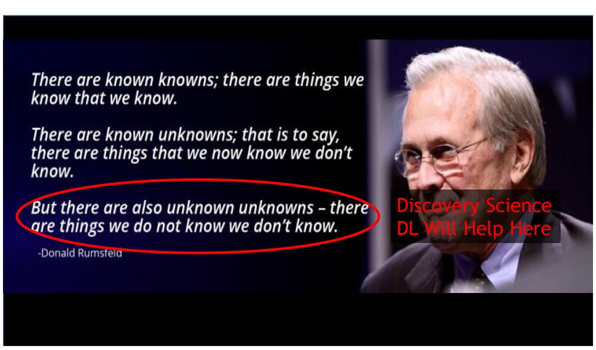
There are known unknowns; that is to say, there are things that we now know we don't know.

But there are also unknown unknowns – there are things we do not know we don't know.

-Donald Rumsfeld


There are known knowns; there are things we know that we know.
Standard Medical Practice
There are known unknowns; that is to say, there are things that we now know we don't know.
But there are also unknown unknowns – there are things we do not know we don't know.
-Donald Rumsfeld


There are known knowns; there are things we know that we know.
There are known unknowns; that is to say, there are things that we now know we don't know.
Hypothesis Driven Science
?Off-label Use
But there are also unknown unknowns – there are things we do not know we don't know.
-Donald Rumsfeld


There are known knowns; there are things we know that we know.
There are known unknowns; that is to say, there are things that we now know we don't know.
But there are also unknown unknowns – there are things we do not know we don't know.
Discovery Science
DL Will Help Here
-Donald Rumsfeld

Clinical Decision Making

- General Decision Making Requires
  - Data
  - Knowledge: meaning of data
  - Judgement: meaning of data where knowledge is not decisively clear
  - Trust: experience that judgements are defendable

Clinical Decision Making

- Decision Making Requires
  - Data
  - Knowledge: meaning of data
  - Judgement: meaning of data where knowledge is not decisively clear
  - Trust: experience that judgements are defendable
- Clinical Decision Making *also* Requires
  - Data Interpolation: **This** patient versus trial group
  - Relevant research is often gray (no trial has 100% response rate)
  - Relationship between research outcomes and causality can be difficult to understand

Anything you can do,
AI can do better

Clinical Decision Making and DL

- DL requires much data

Clinical Decision Making and DL

- DL requires much data
- DL requires much ***annotated*** data

Clinical Decision Making and DL

- DL requires much data
- DL requires much ***annotated*** data
- DL algorithms find associations, not proof

Clinical Decision Making and DL

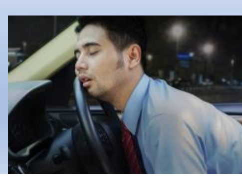
- DL requires much data
- DL requires much *annotated* data
- DL algorithms find associations, not proof
- DL algorithms have poor explain-ability (so far)

Clinical Decision Making and DL

- DL requires much data
- DL requires much *annotated* data
- DL algorithms find associations, not proof
- DL algorithms have poor explain-ability (so far)
- Proposed solutions must have clinical value

DL and Clinical Diagnosis

- DL tools that increase 'attentiveness'
  - Traditional role for screening CAD
  - Usually most helpful to trainees or generalists



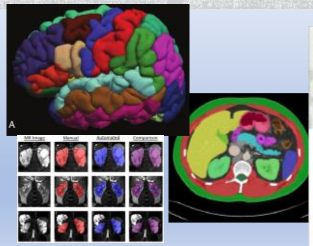
DL and Clinical Diagnosis

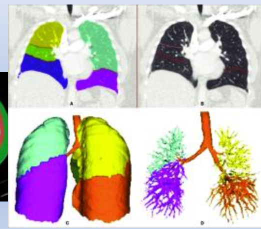
- DL tools can ‘see’ what isn’t seen today

DL and Image Reconstruction

- Several papers have shown DL can reconstruct high quality (‘normal dose’) images with significantly reduced dose
  - PET tracer dose
  - CT dose
  - MR acquisition time
- This is likely to be readily accepted in medical imaging

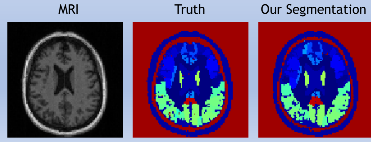
DL and Quantitative Imaging





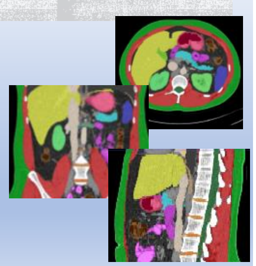
Current Results: Brain

- Fully automatic segmentation of Brain
  - 127 parts of the brain semi-automatically labeled with human expert verification
  - 2000 train, 350 test
  - 30 seconds to segment!
- Dice scores:
  - Mean: 0.954
  - Range: 0.852 – 0.989



Current Results: CT Abdomen

- 100 CT abdomens for range of diseases
- 80 traced by 1 human, used for training
- 20 were used as test/truth
  - Hand labeled by 5
  - STAPLE used to create ‘truth’
- Dice scores range from 0.98 for liver to 0.7 for adrenals and renal veins
- Now using machine labeled as starting point for human correction



DL Enables Radiogenomics

N=498 subjects suing T2-weighted images from Mayo, UCSF, TCIA
398 for training, 100 for testing
50 layer ResNet, VGGNet, Inception

Marker	Sens	Spec	Accuracy
IDH1	0.95	0.95	0.95
1p19q Co-Del	0.91	0.85	0.87
ATRX	0.93	0.89	0.91
MGMT Methylation	0.95	0.95	0.95

*Korfiatis, submitted

That is Exciting, But…

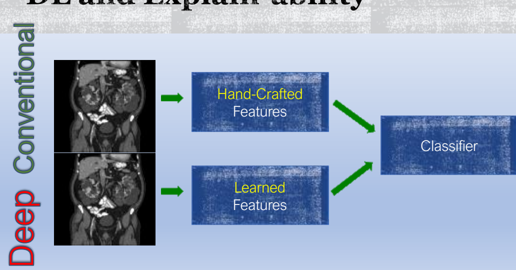
Intelligent Machines

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

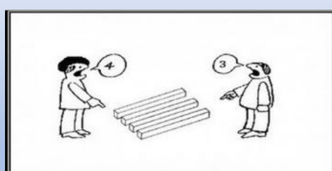
DL and Explain-ability



How are Features Determined?

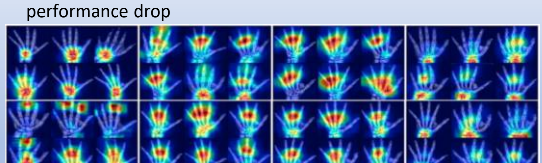
- Many traditional radiology features start with little physical basis
  - Calcification
    - Breast cancers have micro-calcifications (most calcification is not cancer)
    - Gliomas usually have no calcifications, but if they do, they are coarse
    - Lymphomas don't calcify until they are treated
  - The pathological correlate can be seen, but not always fully explained.
  - Expected imaging methods based on pathology sometimes don't turn out due to complexity of the living human environment or because we don't fully understand the biology (which can lead to new insights)

What does Deep Learning see?



Methods to Get Image Features

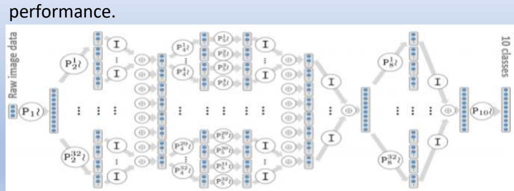
- One can 'blank out' features of image and measure performance drop



* Do, C-MIMI, 2016

Methods to Understand Feature Weights

- Can convert connection weights to decision trees. Slight loss in performance.



*Ioannou, Arxiv, 2016

Implementation Issues

- What is the impact of acquisition technique?
  - Some diseases will be impacted more
  - Device makers are best positioned to address this

Implementation Issues

- Who is responsible for results / reliability
  - Device manufacturers
  - Medical Doctor

Implementation Issues

- Informatics Challenges
  - Must assure all and only the correct image(s) are sent to analytic
    - May need combinations of images (e.g. prior exam)
  - How to capture, represent, transmit, store the output in an informative and computable fashion
  - Clinical workflow: much like 3D lab renderings, these results must be produced reliably on a recognizable set of images, and results must either be completed before scanning is done, or process for notifying radiologist that results are ready must be put in place

Conclusions

- DL has tremendous potential to increase value of medical imaging
  - Quantitative results
  - New diagnostic capabilities
- This will demand
  - Careful assessment of imaging technique impact on DL
  - Increased use of DICOM SR/AIM to convey results efficiently
  - New workflow methods to implement these tools in efficient way