Artificial Intelligence in Healthcare Analytics

Wei-Hung Weng, MD, MMSc

NUS-MIT Healthcare Analytics Datathon 2017
Jun 30, 2017

CSAIL
Massachusetts Institute of Technology

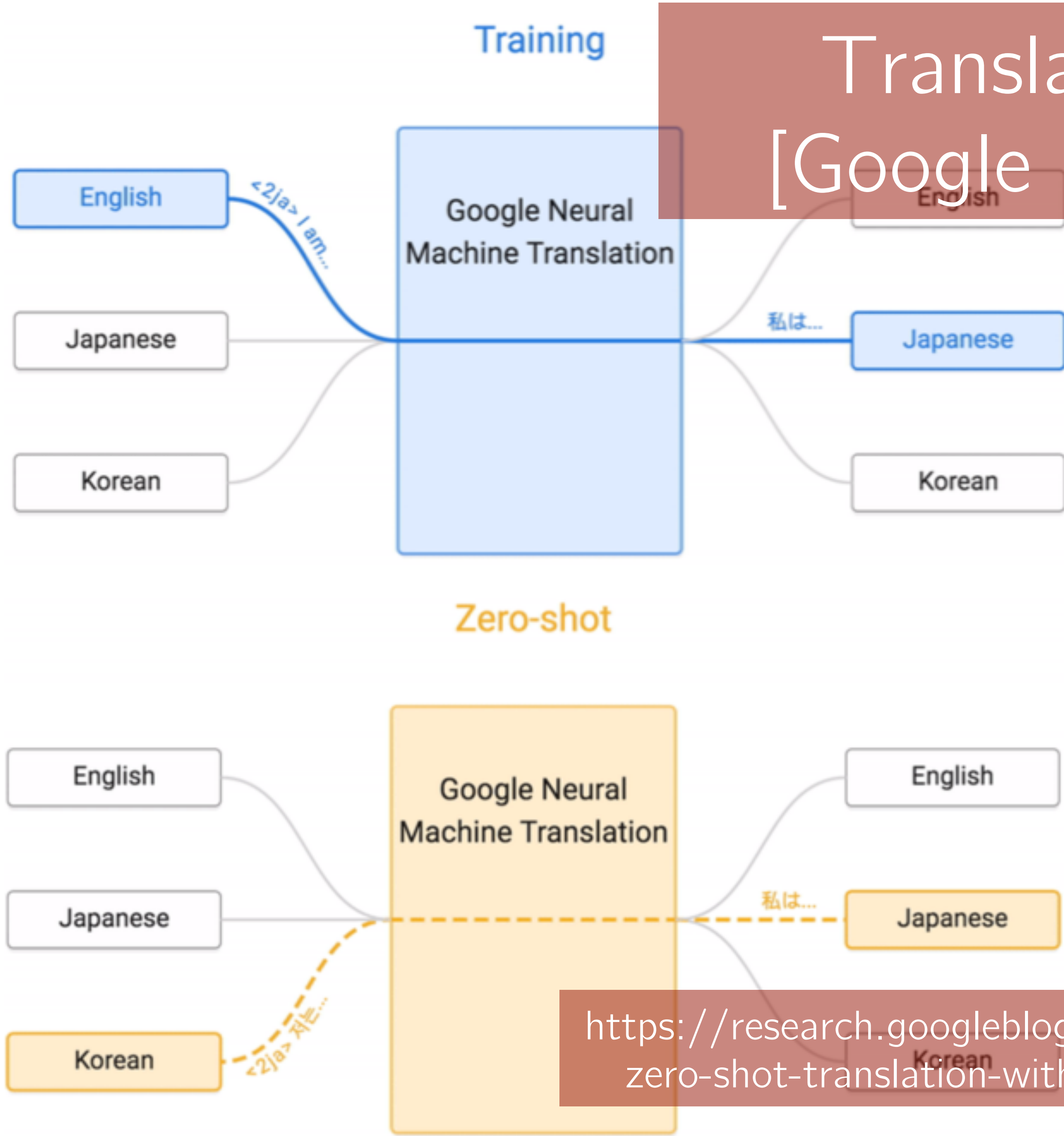
Disclosure

- No conflict of interest

- Artificial intelligence (AI) is defined by the American cognitive scientist Marvin Minsky as *the science of making machines do things that would require intelligence if done by man*.

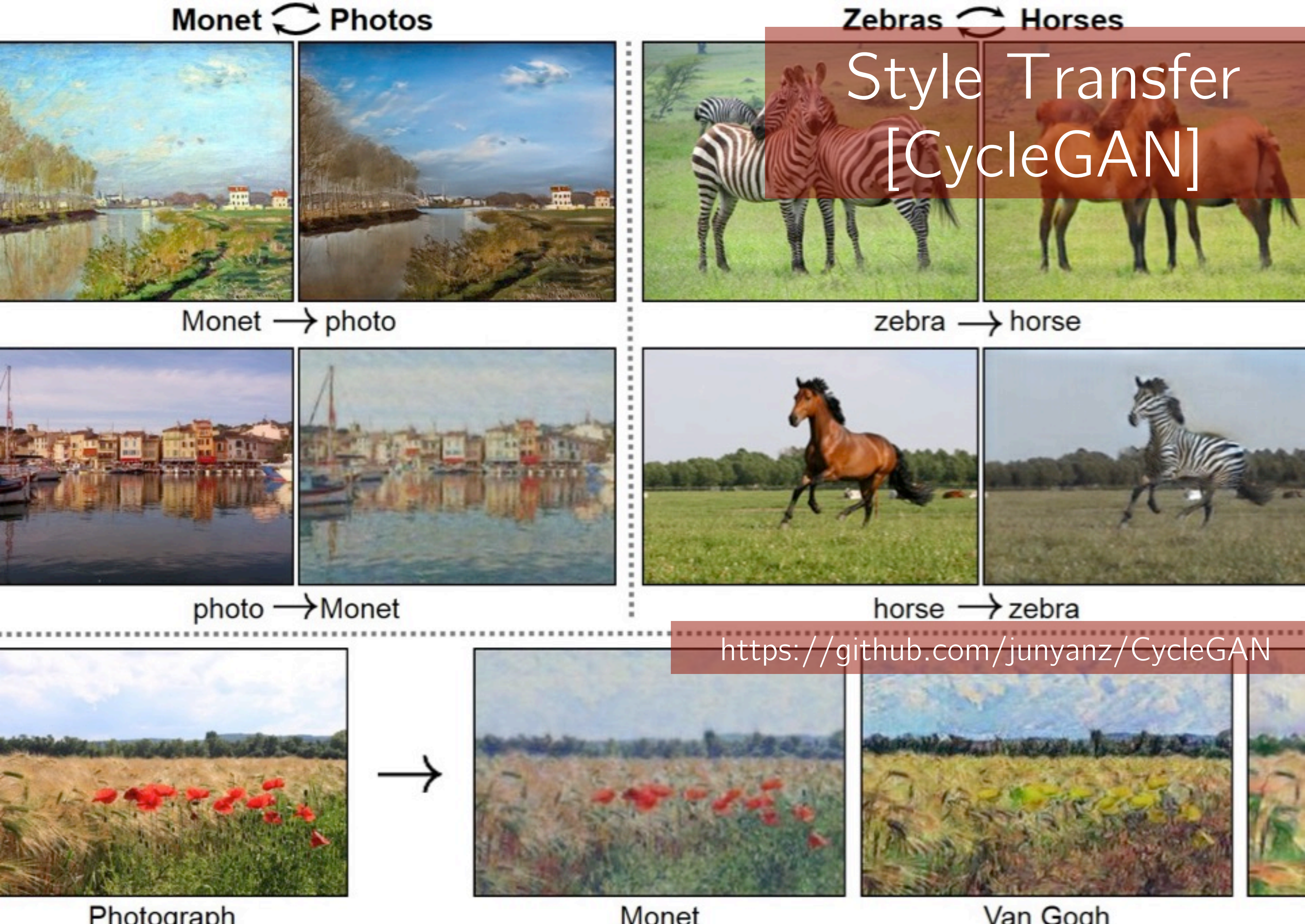

Self-Driving
[Deep Tesla]
Tesla Control
(by Autopilot)
Learned Control
(by Deep Neural Network)
(Ground Truth)
Red = Disagree
Green = Agreee
http://selfdrivingcars.mit.edu/

Translation [Google NMT]



https://research.googleblog.com/2016/11/zero-shot-translation-with-googles.html

Style Transfer [CycleGAN]

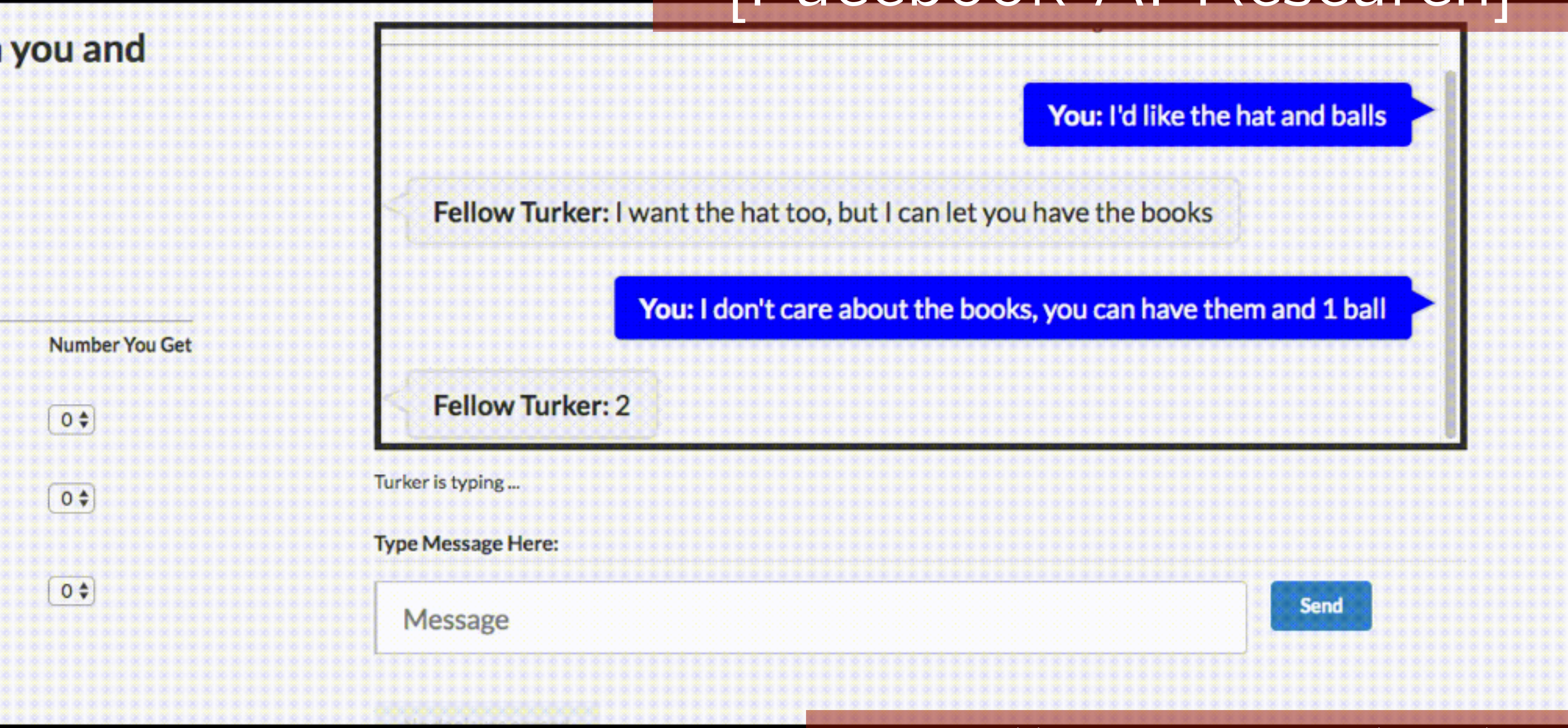


https://github.com/junyanz/CycleGAN

Go Playing [DeepMind AlphaGo]

https://twitter.com/DeepMindAI/status/867996695778410497/photo/1

Negotiating / Bargaining [Facebook AI Research]



https://code.facebook.com/posts/1686672014972296/deal-or-no-deal-training-ai-bots-to-negotiate?pnref=story

Harvard Business Review

TRUE?

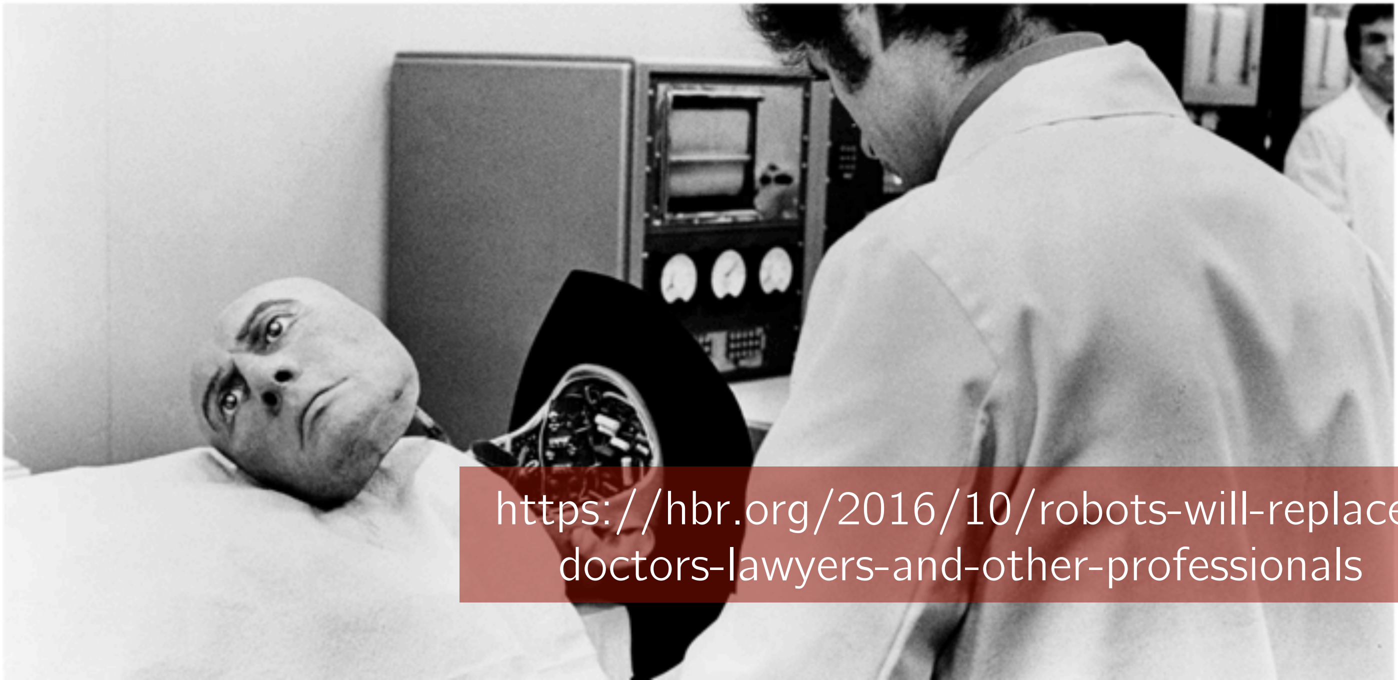
INFORMATION & TECHNOLOGY

Technology Will Replace Many Doctors, Lawyers, and Other Professionals

by Richard Susskind and Daniel Susskind

OCTOBER 11, 2016

SAVE SHARE 26 COMMENT TEXT SIZE PRINT $8.95 BUY COPIES



https://hbr.org/2016/10/robots-will-replace-doctors-lawyers-and-other-professionals

Brief History

- Symbolic AI, rule-based
- Expert system
- Mimicking human decision making process
- Human curated features
- Watson

- MYCIN (Stanford)
- INTERNIST-1 (U Pittsburg)

MYCIN

```
--------PATIENT-1--------
1)  Patient's name: (first-last)
**  FRED SMITH
2)  Sex:
**  MALE
3)  Age:
** 55
4)  Have you been able to obtain positive cultures from a site
    at which Fred Smith has an infection?
**  YES
  --------INFECTION-1--------
  5)  What is the infection?
  **  PRIMARY-BACTEREMIA
  6)  Please give the date and approximate time when signs or symptoms
      of the primary-bacteremia (INFECTION-1) first appeared. (mo/da/yr)
  **  MAY 5, 1975
  The most recent positive culture associated with the
  primary-bacteremia (INFECTION-1) will be referred to as:
    --------CULTURE-1--------
    7)  From what site was the specimen for CULTURE-1 taken?
    ** BLODD
    = BLOOD
    8)  Please give the date and time when this blood culture
        (CULTURE-1) was obtained. (mo/da/yr)
    ** MAY 9, 1975
    The first significant organism from this blood culture
    (CULTURE-1) will be referred to as:
      --------ORGANISM-1--------
      9)  Enter the identity of ORGANISM-1:
      **  UNKNOWN
             {typically the identity is not known and must be inferred}
      10)  Is ORGANISM-1 a rod or coccus (etc.):
      **   ROD
      11)  The gram stain of ORGANISM-1:
      **   GRAMNEG
      12)  Did ORGANISM-1 grow aerobically?
      **   YES (.8)
```

Szolovits, Artificial Intelligence in Medicine 1982

MYCIN

```
My therapy recommendation will be based on the following probable
  infection(s) and potential causative organism(s):
INFECTION-1 is PRIMARY-BACTEREMIA                    {possible identities}
<Item 1> The identity of ORGANISM-1 may be PSEUDOMONAS-AERUGINOSA
<Item 2> The identity of ORGANISM-1 may be KLEBSIELLA-PNEUMONIAE
<Item 3> The identity of ORGANISM-1 may be E.COLI
<Item 4> The identity of ORGANISM-1 may be BACTEROIDES-FRAGILIS
<Item 5> The identity of ORGANISM-1 may be ENTEROBACTER
<Item B> The identity of ORGANISM-1 may be PROTEUS-NON-MIRABILIS

  46)  On a scale of 0 to 4, where higher numbers indicate increasing
       severity how would you rate Fred Smith's degree of sickness?
  **  3
  47)  Does Fred Smith have a clinically significant allergic
       reaction to any antimicrobial agent?
  **  NO
  48)  Fred Smith's weight in kilograms (or <number> POUNDS):
  **  70
                 {note that 2 drugs cover for 6 potential organisms)
[Rec 1] My preferred therapy recommendation is as follows:
  In order to cover for Items <1 2 3 5 6>:
    Give:  GENTAMICIN
    Dose:  119 mg (1.7 mg/kg) q8h IV [or IM] for 10 days
    Comments:  Modify dose in renal failure
  In order to cover for Item <4>:
    Give:  CLINDAMYCIN
    Dose:  595 mg (8.5 mg/kg) q6h IV [or IM] for 14 days
    Comments:  If diarrhea or other GI symptoms develop, patient
      should be evaluated for possible pseudomembranous colitis.

Do you wish to see the next choice therapy?
**  Y                                  {alternative choices are available}
Which, If any, of the above drugs do you wish eliminated from
  consideration?
**  CLINDAMYCIN
[Rec 2] Next best therapy recommendation:
  In order to cover for Items <2 3 4 5 6>:
    Give:  CHLORAMPHENICOL
```

Szolovits, Artificial Intelligence in Medicine 1982

MYCIN

(a) **** HOW DID YOU DECIDE THAT ORGANISM-1 MIGHT BE AN E.COLI?**
I used RULE021 to conclude that the identity of ORGANISM-1 is e.coli. This gave a cumulative CF of (.47).
Then I used RULE084 to conclude that the identity of ORGANISM-1 is e.coli. This gave a cumulative CF of (.55).
Finally, I used RULE003 to conclude that the identity of ORGANISM-1 is e.coli. This gave a cumulative CF of (.74).

(b) **** DID YOU USE RULE 163 TO FIND OUT ANYTHING ABOUT ORGANISM-1?**
RULE163 was tried in the context of ORGANISM-1, but it failed because it is not true that the patient has had a genito-urinary tract manipulative procedure (clause 3).

(c) **** DID RULE 116 TELL YOU ANYTHING ABOUT INFECTION-1?**
No. RULE116 would be relevant to the primary-bacteremia when applied to ORGANISM-1.
But clause 2 ("the gram stain of the organism is not known") was already known to be false for ORGANISM-1, so the rule was never tried.

(d) **** WHY DIDN'T YOU CONSIDER STREPTOCOCCUS AS A POSSIBILITY FOR ORGANISM-1?**
The following rule could have been used to determine that the identity of ORGANISM-1 was streptococcus: RULE033
But clause 2 ("the morphology of the organism is coccus") was already known to be false for ORGANISM-1, so the rule

Szolovits, Artificial Intelligence in Medicine 1982

INTERNIST-I

```
ABDOMEN GUARDING INVOLUNTARY LOCALIZED ?
NO

Please Enter Findings of TRANSFUSION <S> HX
*GO

TRANSFUSION <S> MULTIPLE HX 7
NO

DISREGARDING: SKIN SPIDER ANGIOMATA, CREATININE BLOOD INCREASED,
UREA NITROGEN BLOOD 60 TO 100

CONSIDERING: FECES LIGHT COLORED, JAUNDICE, LIVER ENLARGED
SLIGHT, ALKALINE PHOSPHATASE BLOOD GTR THAN 2 TIMES NORMAL, SGOT
GTR THAN 400, WBC GTR THAN 30000, SGPT 200 TO 600

PURSUING: HEPATIC LEPTOSPIROSIS
CONCLUDE: HEPATIC LEPTOSPIROSIS

DISREGARDING: SKIN SPIDER ANGIOMATA, UREA NITROGEN BLOOD 60 TO 100

CONSIDERING: OLIGURIA HX, CREATININE BLOOD INCREASED

RULEOUT: RENAL LEPTOSPIROSIS, GLOMERULONEPHRITIS ADVANCED CHRONIC,
LEAD NEPHROPATHY, GLOMERULONEPHRITIS ACUTE, PYELONEPHRITIS
CHRONIC, DIABETIC NEPHROPATHY
```

Szolovits, Artificial Intelligence in Medicine 1982

Problems

- Could not fit into clinical workflow
- Incomplete
- Expert-needed
- Can't deal with uncertainty
- Poor generalization

Now?

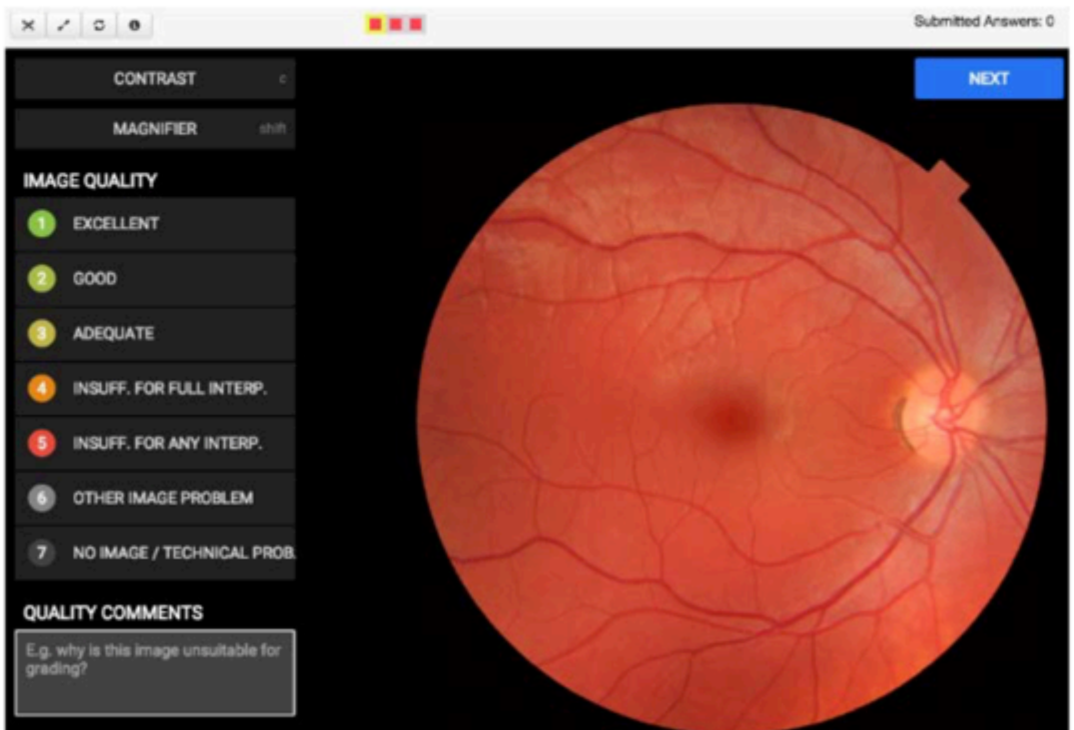
- Symbolic → Data-driven
- Rule-based → Learning-based
- May be able to deal with uncertainty

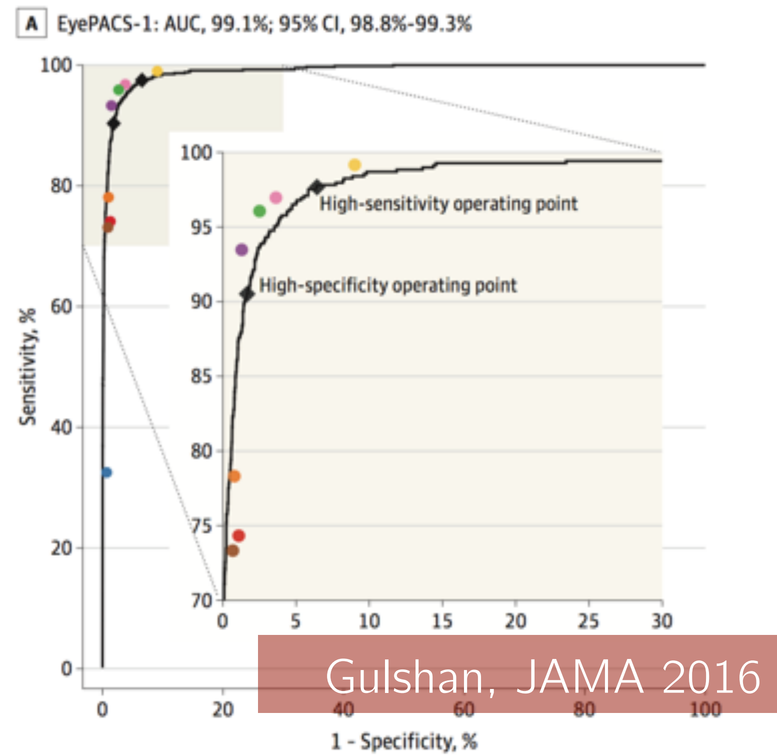
Medical Imaging

JAMA | Original Investigation | INNOVATIONS IN HEALTH CARE DELIVERY

Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs

Varun Gulshan, PhD; Lily Peng, MD, PhD; Marc Coram, PhD; Martin C. Stumpe, PhD; Derek Wu, BS; Arunachalam Narayanaswamy, PhD; Subhashini Venugopalan, MS; Kasumi Widner, MS; Tom Madams, MEng; Jorge Cuadros, OD, PhD; Ramasamy Kim, OD, DNB; Rajiv Raman, MS, DNB; Philip C. Nelson, BS; Jessica L. Mega, MD, MPH; Dale R. Webster, PhD



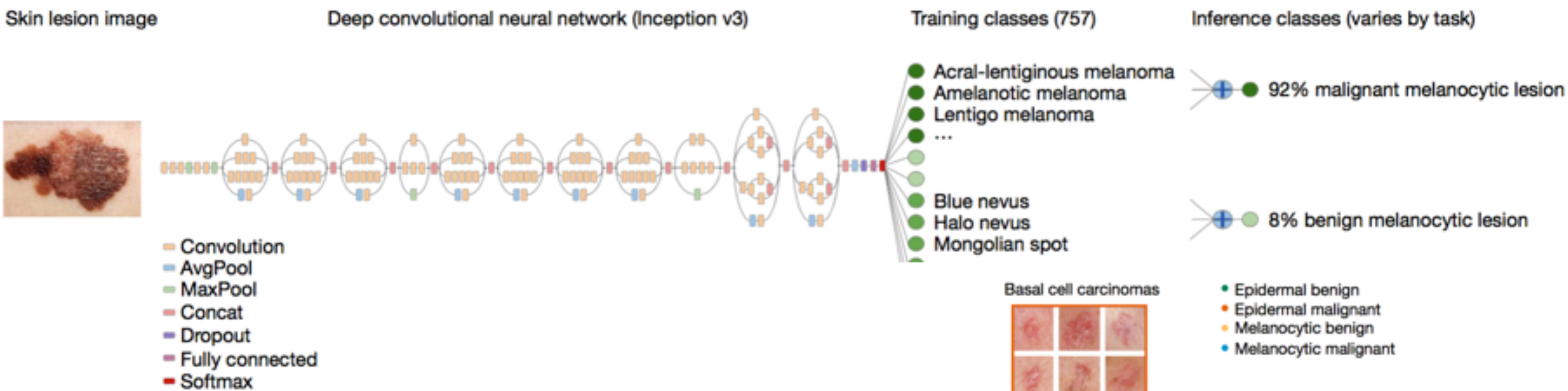


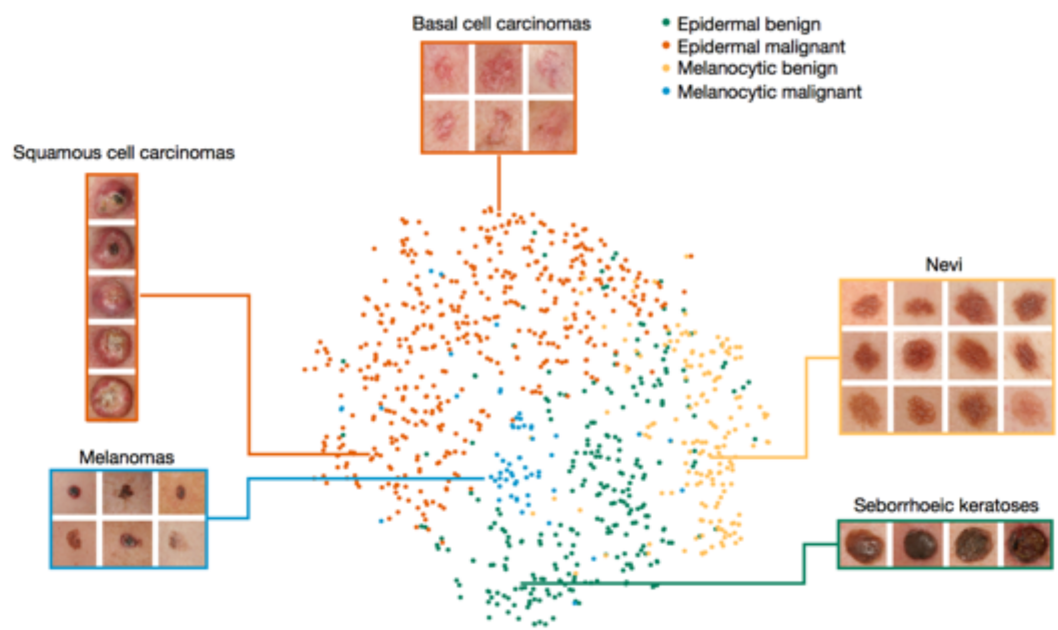
Gulshan, JAMA 2016

Medical Imaging

Dermatologist-level classification of skin cancer with deep neural networks

Andre Esteva[1]*, Brett Kuprel[1]*, Roberto A. Novoa[2,3], Justin Ko[2], Susan M. Swetter[2,4], Helen M. Blau[5] & Sebastian Thrun[6]



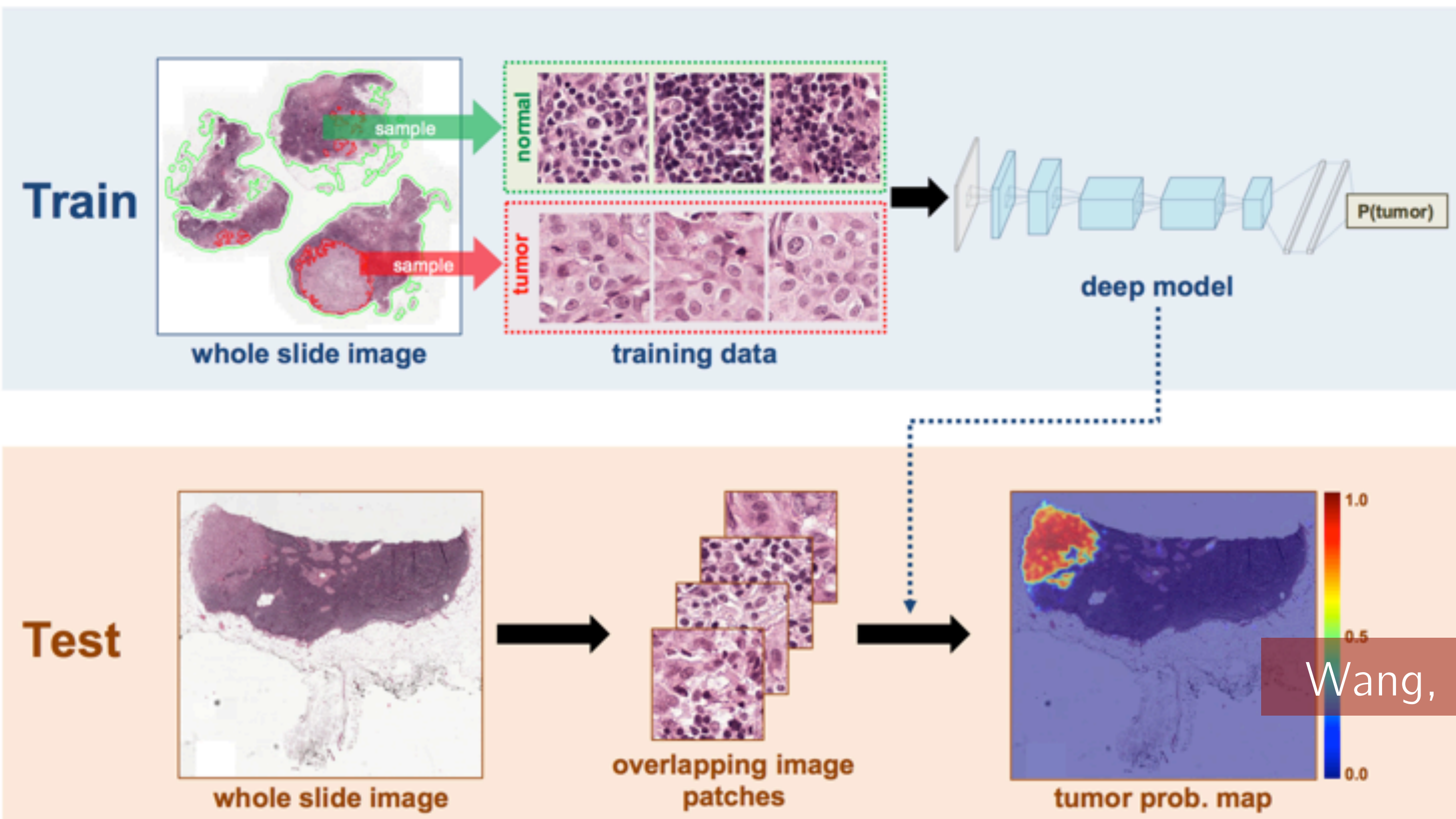


Esteva, Nature 2016

Medical Imaging

Deep Learning for Identifying Metastatic Breast Cancer

Dayong Wang Aditya Khosla* Rishab Gargeya Humayun Irshad Andrew H Beck
Beth Israel Deaconess Medical Center, Harvard Medical School
*CSAIL, Massachusetts Institute of Technology



Wang, arXiv 2016

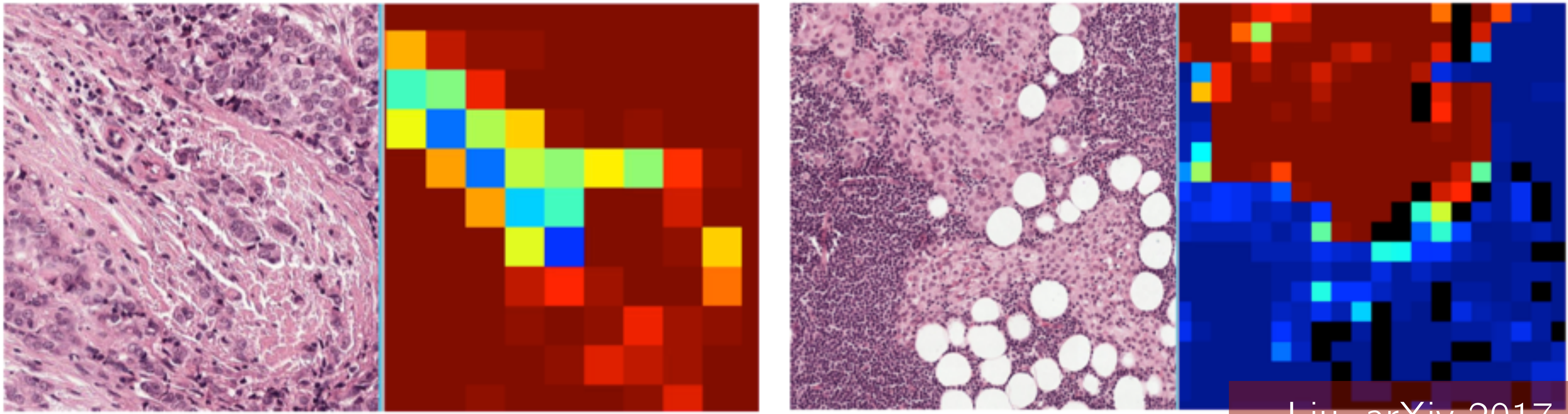
Medical Imaging

Detecting Cancer Metastases on Gigapixel Pathology Images

Yun Liu[1*], Krishna Gadepalli[1], Mohammad Norouzi[1], George E. Dahl[1], Timo Kohlberger[1], Aleksey Boyko[1], Subhashini Venugopalan[2**], Aleksei Timofeev[2], Philip Q. Nelson[2], Greg S. Corrado[1], Jason D. Hipp[3], Lily Peng[1], and Martin C. Stumpe[1]

{liuyun,mnorouzi,gdahl,lhpeng,mstumpe}@google.com

[1]Google Brain, [2]Google Inc, [3]Verily Life Sciences, Mountain View, CA, USA



Liu, arXiv 2017

Medical Text

Using Machine Learning to Parse Breast Pathology Reports

Adam Yala,[1] Regina Barzilay,[1] Laura Salama,[3] Molly Griffin,[2] Grace Sollender,[8] Aditya Bardia,[10] Constance Lehman,[5] Julliette M. Buckley,[2] Suzanne B. Coopey,[2] Fernanda Polubriaginof,[9] Judy E. Garber,[6] Barbara L. Smith,[2] Michele A. Gadd,[2] Michelle C. Specht,[2] Thomas M. Gudewicz,[4] Anthony Guidi,[7] Alphonse Taghian,[3] and Kevin S. Hughes[2]

Pathology Report: REMOVED_ACCESSION_ID
ACCESSIONED ON: REMOVED_DATE
CLINICAL DATA: Carcinoma **right breast**.
*** FINAL DIAGNOSIS ***
LYMPH NODE (SENTINEL), EXCISION
(REMOVED_CASE_ID): METASTATIC CARCINOMA IN 1 OF 1 LYMPH NODE.
NOTE: The metastatic deposit spans 0.19cm and is identified on H&E and cytokeratin immunostains. A second cytokeratin-positive but cauterized focus likely also represents metastatic tumor (<0.1cm). There is **no evidence of extranodal extension**. BREAST (RIGHT), EXCISIONAL BIOPSY (REMOVED_ACCESSION_ID : REMOVED_CASE_ID -B): **INVASIVE DUCTAL CARCINOMA** (SEE TABLE #1). **DUCTAL CARCINOMA IN-SITU**, GRADE 1. **ATYPICAL DUCTAL HYPERPLASIA**. **LOBULAR NEOPLASIA (ATYPICAL LOBULAR HYPERPLASIA)**.
TABLE OF PATHOLOGICAL FINDINGS #1 INVASIVE CARCINOMA
Tumor size: Cannot evaluate. Grade: 1.
Lymphatic vessel invasion: Not identified.
Blood vessel invasion: Not identified.
Margin of invasive carcinoma: Invasive carcinoma extends to less than 0.2cm from the inferior margin of the specimen.
Stains for receptors: Outside **immunohistochemical stains demonstrate that the tumor cells express estrogen and progesterone receptors.**

Name	Extraction
Breast Side	Right
Ductal Carcinoma in Situ	Present
Invasive Lobular Carcinoma	Absent
Invasive Ductal Carcinoma	Present
Cancer	Present
Lobular Carcinoma in Situ	Absent
Atypical Ductal Hyperplasia	Present
Atypical Lobular Hyperplasia	Present
Lobular Neoplasia	Present
Flat Epithelial Atypia	Absent
Blunt Adenosis	Absent
Atypia	Present
Positive Lymph Nodes	Present
Extracapsular Axillary Nodal Extension	Absent
Isolated Cancer Cells in Lymph Nodes	Absent
Lymphovascular Invasion	Absent
Blood Vessel Invasion	Absent
Estrogen Receptor Status	Positive
Progesterone Receptor Status	Positive
HER 2 (FISH) Status	Unknown

Category	Accuracy	F-score
Breast side	1.0	1.0
DCIS	.99	.99
ILC	.99	.99
IDC	1.0	1.0
Carcinoma	.94	.94
LCIS	1.0	.98
ADH	.90	.90
ALH	.98	.98
Lobular Neoplasia	.97	.97
Flat Epithelial Atypia	1.0	1.0
Blunt Adenosis	1.0	1.0
Atypia	.91	.91
Positive LN	.98	.98
ECE	.97	.97
ITC in LN	.96	.96
LVI	.92	.88
BVI	.93	.90
ER Status	.97	.96
PR Status	.97	.95
HER 2 Status	.96	.94
Report-Level	.90	N/A
Average	.97	.96

Yala, arXiv 2016

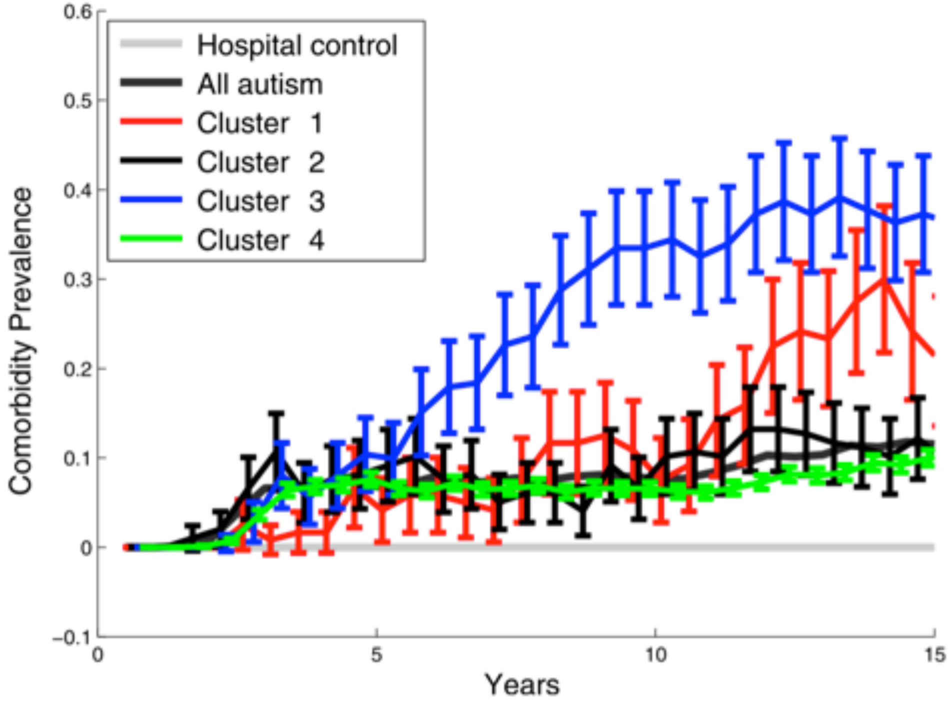
Structured Data

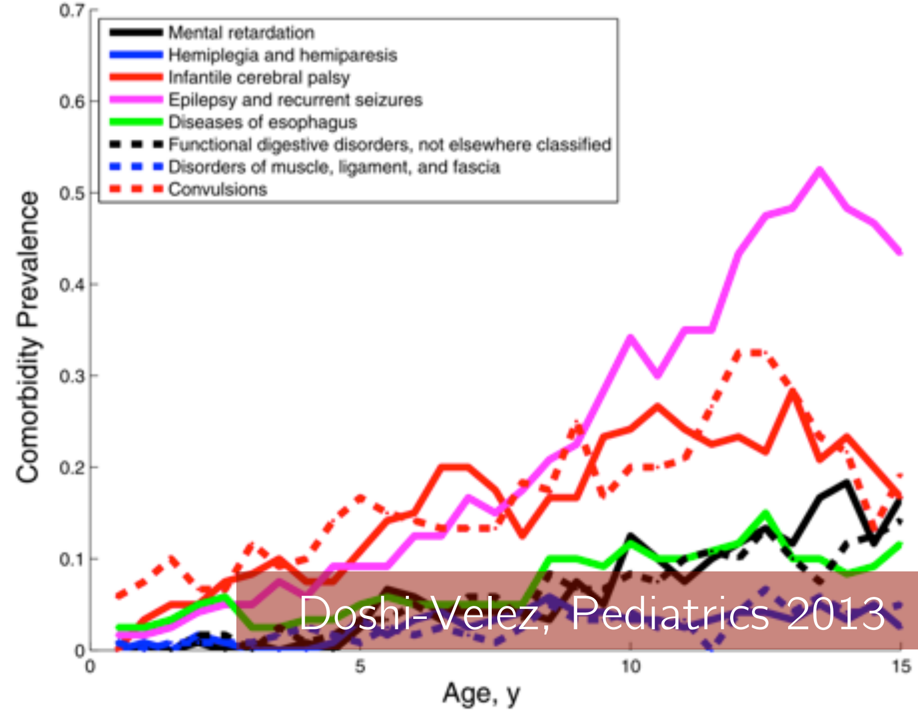
Comorbidity Clusters in Autism Spectrum Disorders: An Electronic Health Record Time-Series Analysis

WHAT'S KNOWN ON THIS SUBJECT: Individuals with autism spectrum disorders have a higher comorbidity burden than the general pediatric population, including higher rates of seizures, psychiatric illness, and gastrointestinal disorders.

AUTHORS: Finale Doshi-Velez, PhD,[a] Yaorong Ge, PhD,[b] and Isaac Kohane, MD, PhD[a]

[a]Center for Biomedical Informatics, Harvard Medical School, Boston, Massachusetts; and [b]Center for Biomedical Informatics, Wake Forest University, Winston-Salem, North Carolina





Doshi-Velez, Pediatrics 2013

Breast FNA Decision Support by a High School Student!

Cloud4Cancer Breast Cancer Detection

...edle Aspirates to determine if a breast mass is malig...
...om the worldwide hospital community. If you would like to contribute samples, please contact Brittany Wenger at cloud4cancerFNA@gmail.com.

...rand Prize Winner! Visit the project site for more informati...

Please enter the attributes of your FNA sample:

Clump Thickness	3, Cells are 80% mono-layered
Uniformity of Cell Size	3, Cells are 80% uniform
Uniformity of Cell Shape	2, Cells are 90% uniform
Marginal Adhesion	6, 50% Stick together
Single Epithelial Cell Size	9, Largest cells appear 90% larger
Bare Nuclei	4, 40% of nuclei have cytoplasm
Bland Chromatin	5, Chromatin is 50% coarse
Normal Nucleoli	4, 40% of nucleoli are abnormal
Mitoses	9, 90% mitotic activity appears abnormal

Send

Remote Procedure Call

Sending scenario to the Breast Cancer Cloud Service:
332694549

Server replies:
MALIGNANT

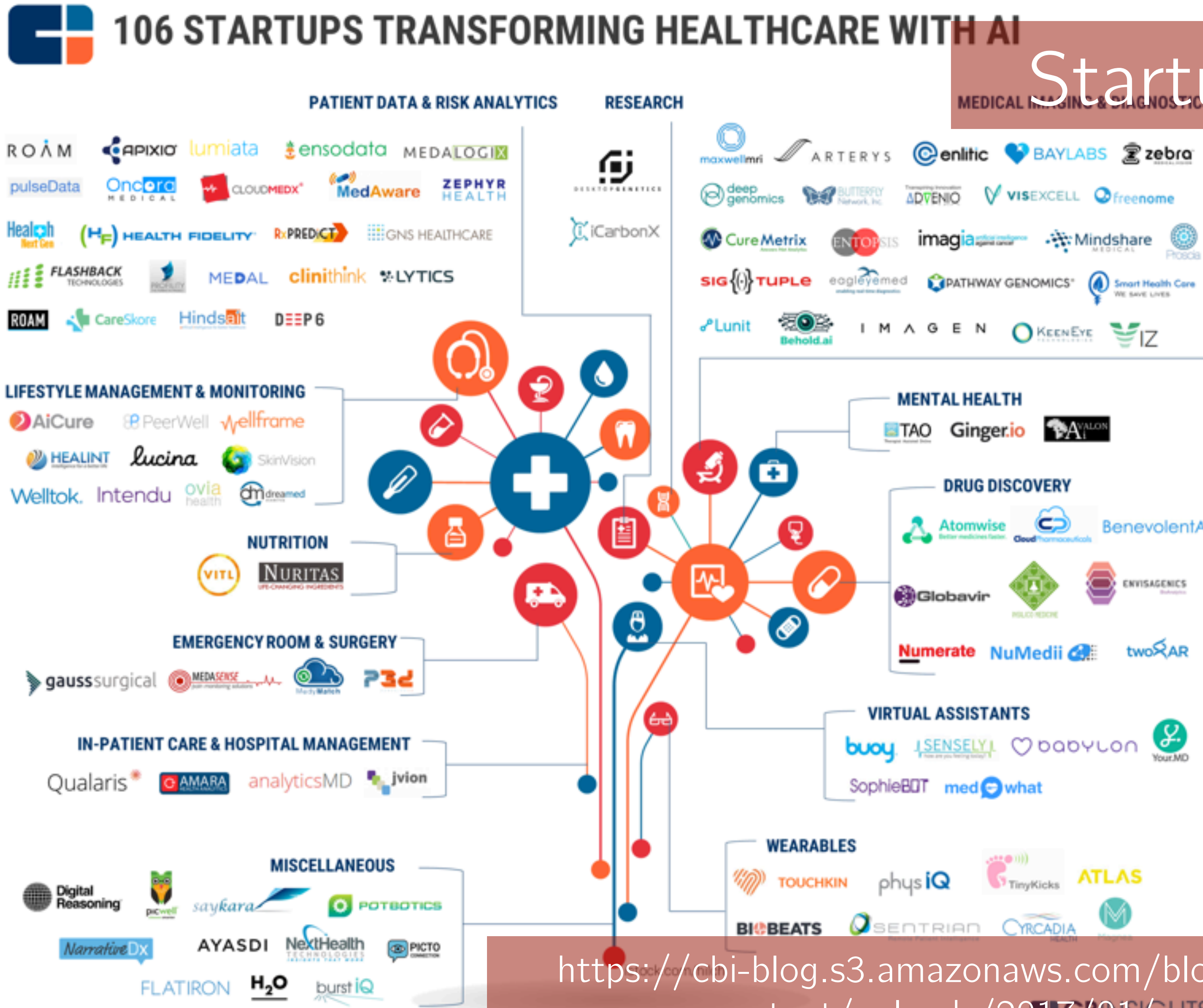
Close

http://cloud4cancer.appspot.com/

Achievement in Industry

- IBM Watson
  - Identify pulmonary embolism on CT
  - Detect abnormal wall motion on echocardiography
  - Watson Oncology in Japan, India, ...
- Enlitic
  - Fracture detection on radiographs
- DeepMind Health
  - High risk blood test detection/notification (Streams)
- Lumiata
- ...

Startups

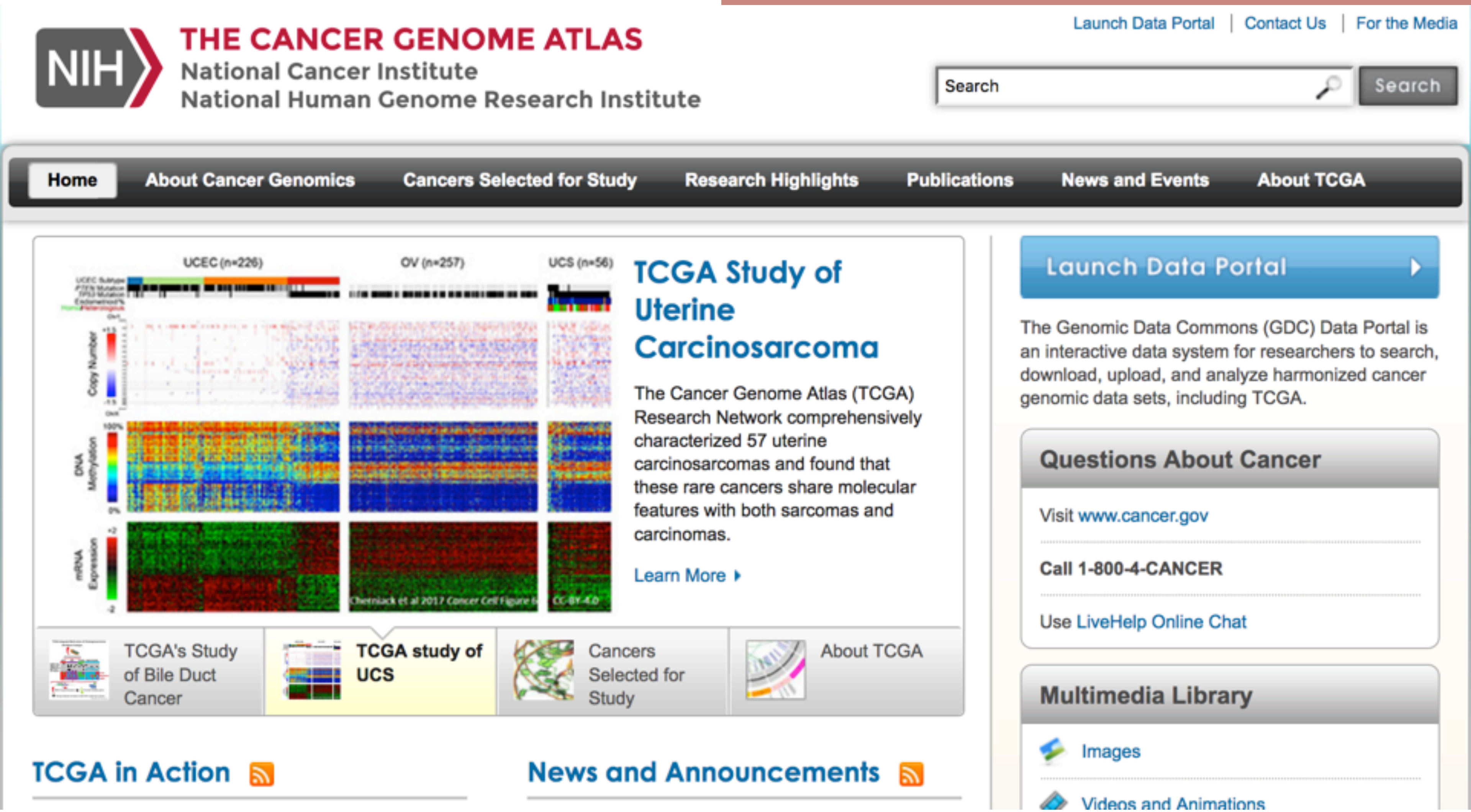


https://cbi-blog.s3.amazonaws.com/blog/wp-content/uploads/2017/01/healthcare_AI_map_2016_1.png

Multimodal Data

https://mimic.physionet.org/
http://eicu-crd.mit.edu/

Multimodal Data



https://cancergenome.nih.gov/

Multimodal Data

NCBI Resources How To

PMC
US National Library of Medicine
National Institutes of Health

PMC

Advanced Journal list

About PMC For Publishers Related Resources

FTP Service

The PMC FTP Service provides access to:

- file lists and files for the articles in the PMC Open Access Subset;
- the Author Manuscript Collection and its file lists; and
- the Historical OCR Collection.

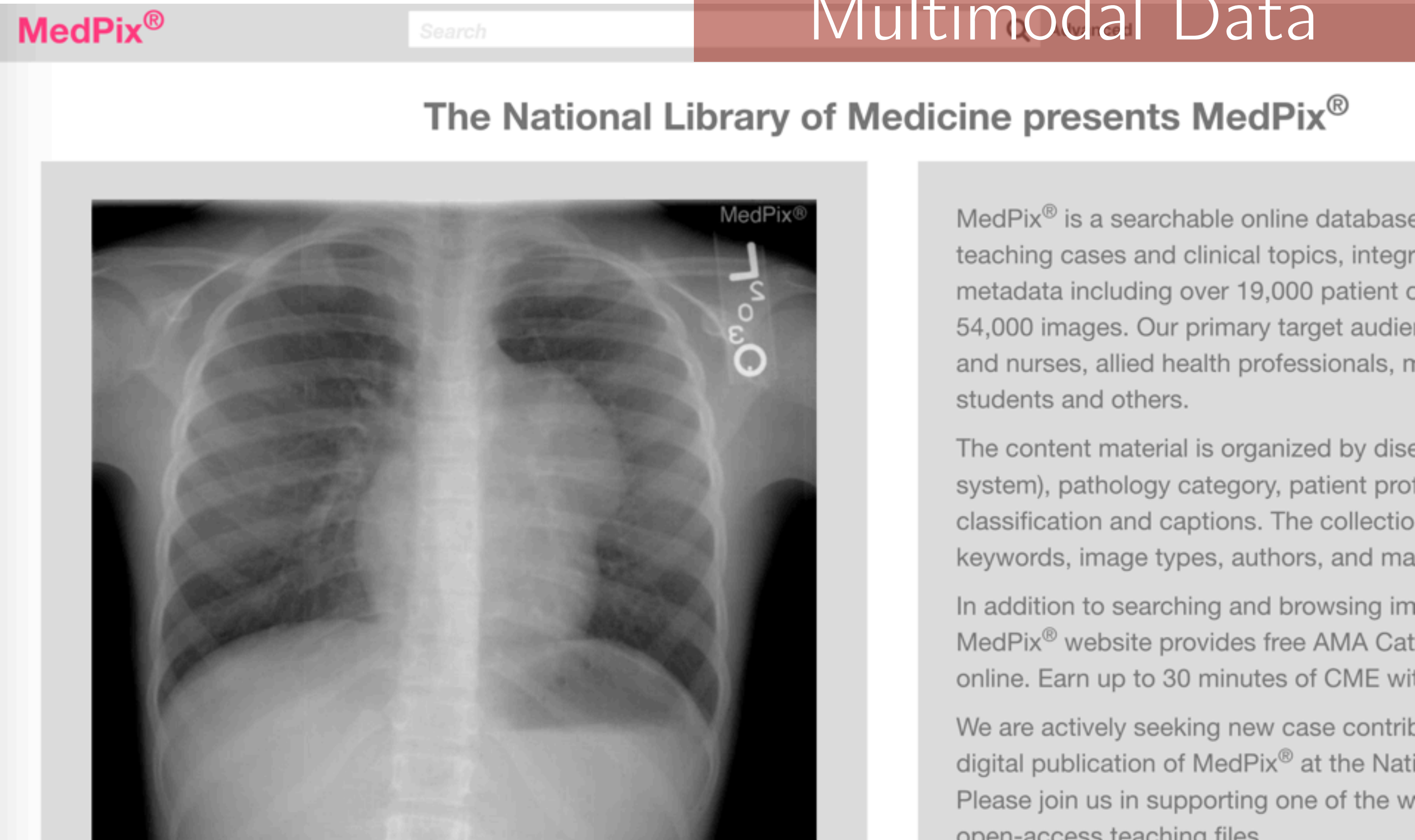
The FTP service also allows users to cross-reference PMC articles with identifiers such as PubMed IDs, DOIs, and Manuscript I
on each of these collections and resources to learn more about the available files.

The directories available via the FTP service include:

- Open access individual article packages (oa_package)
- Open access individual article PDFs (oa_pdf)
- Open access bulk article packages (oa_bulk)

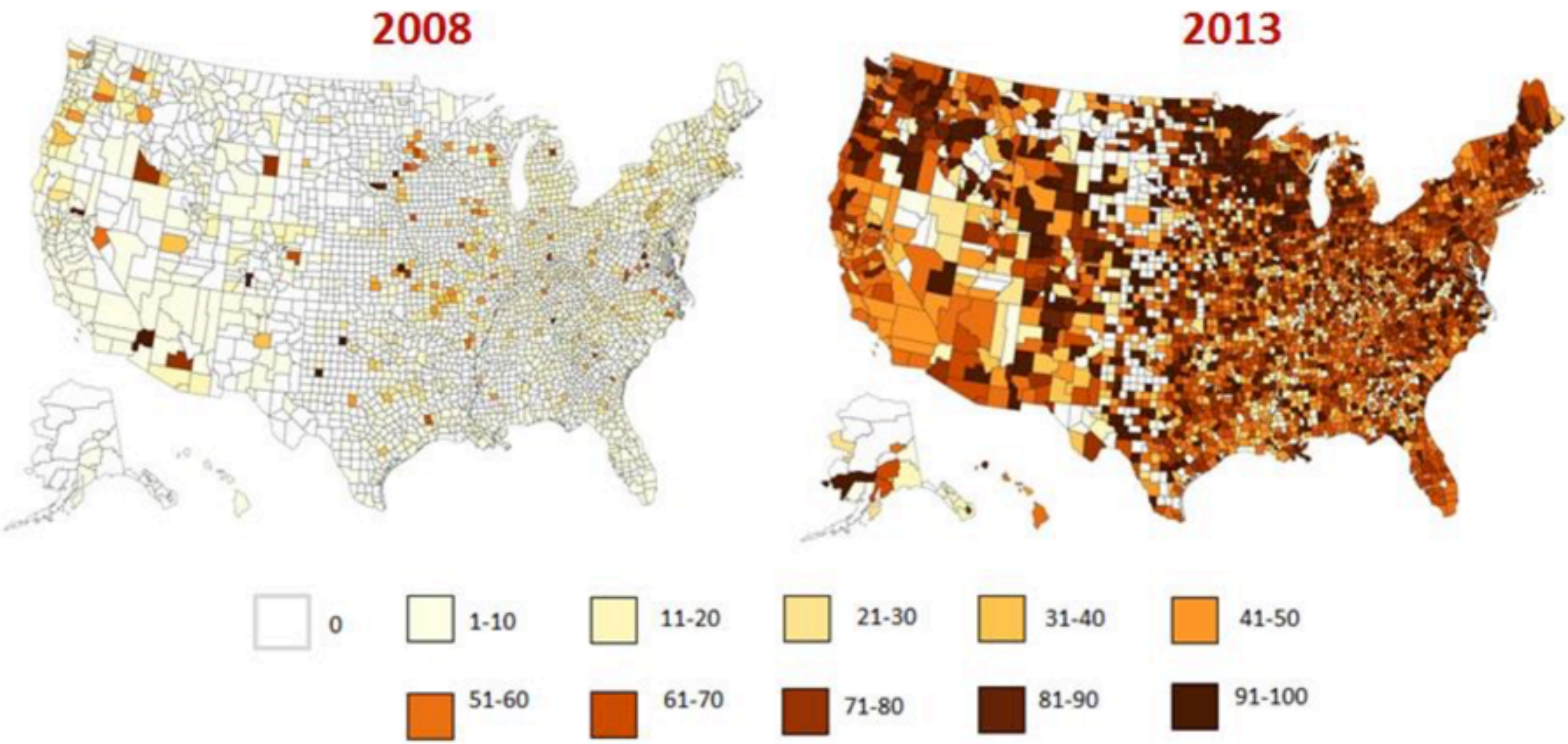
https://www.ncbi.nlm.nih.gov/pmc/tools/ftp/

Multimodal Data



https://medpix.nlm.nih.gov/home

Adoption of EHR



Courtesy by Owen Hsu (Wistron)

Standardization

https://www.linkedin.com/pulse/most-bizarre-icd-10-codes-infographic-nina-keller

Democratization of Knowledge and Resources

UDACITY
coursera
edX
amazon web services™
KHANACADEMY
Google Cloud Platform
IBM Watson
GitHub
APACHE Spark™
scikit learn
TensorFlow
python™

Why Now?

- High quality (?) multimodal data
  - Variety, volume, velocity
  - Social media, sensor, wearable, vital signs, lab data, notes, imaging, -omics
  - Adoption of EHR
- Standardization
  - UMLS, SNOMED, ICD, RxNorm, LOINC, ...
- Advances and democratization in ML
  - Open-source / algorithms & tools
  - Different approaches of knowledge representation

Advantages

- Stability
- Consistency
- Can reduce
  - Medical cost
  - Healthcare inequality
  - Cognitive load
    - Computer vision, natural language processing, ...

Clinical Perspective

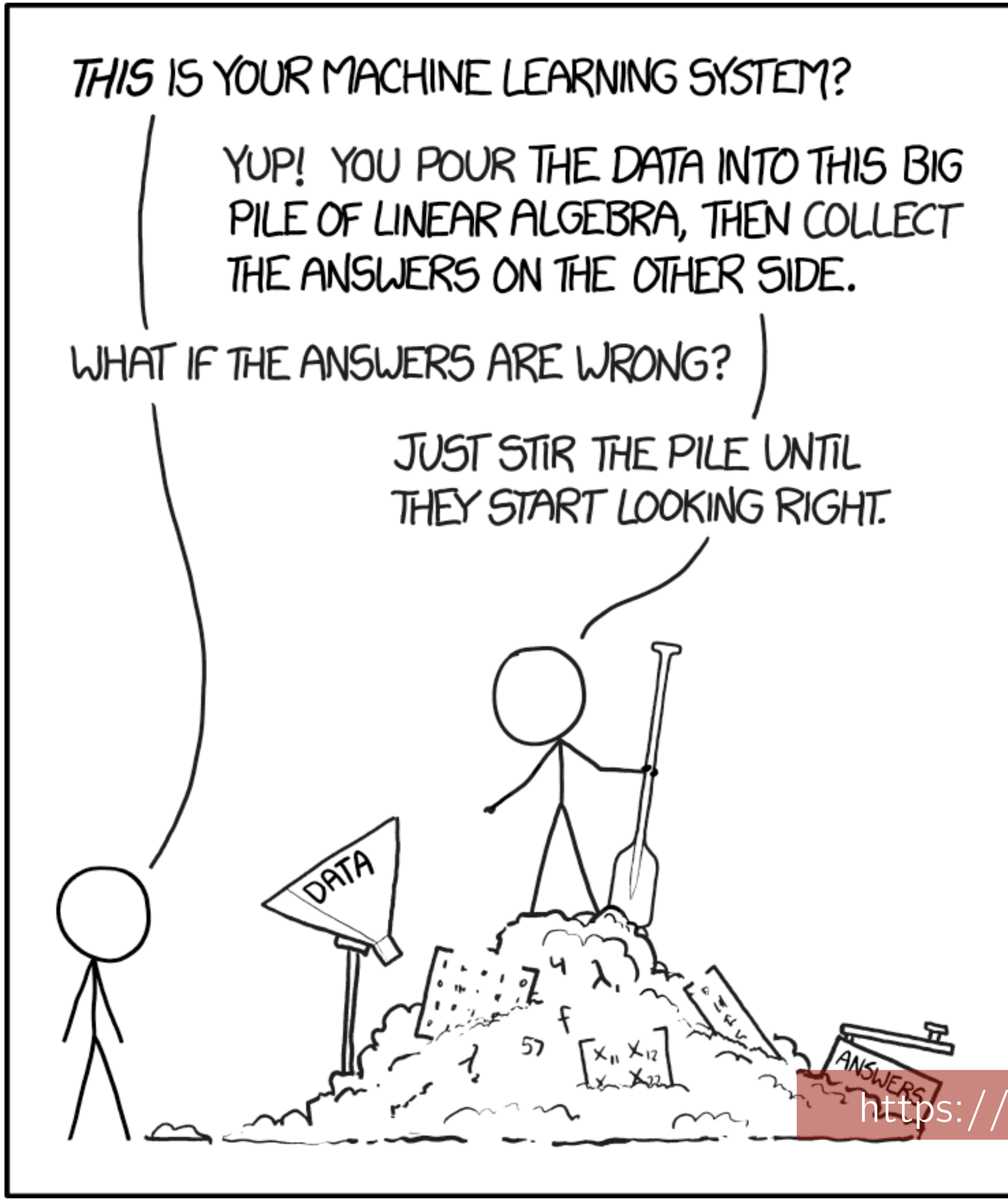
- Cost / Risk assessment and adjustment
  - Insurance
  - Resource redistribution
- Precision / Personalized medicine
  - For
    - Oncology / rare diseases / mental disorders / ...
  - Applications
    - Clinical decision support
    - Drug discovery
    - Outcome prediction
      - Lifespan prediction / Disease progression
    - Chronic disease management
      - Early prediction of blood glucose for self-management

CS/AI/ML Perspective

- Risk stratification
- Causal inference
- Bias
- Time-series
- Unstructured data
- Interpretability
- Disease progression modeling
- Reasoning and decision making

Issues

- Domain specific vs. generalization
- Data quality
- Privacy / Data security
  - Open-source and ubiquitous
  - Innovation, ecosystem
- Transparency



https://xkcd.com/1838/

Replacing Your Job?

The world's first artificially intelligent lawyer was just hired at a law firm

Chris Weller
May 16, 2016, 10:26 AM 44,723

FACEBOOK | LINKEDIN | TWITTER | EMAIL | PRINT

Lawyers can get a bad reputation for being slimy and conniving, but ROSS has neither of those qualities.

Ask ROSS to look up an obscure court ruling from 13 years ago, and ROSS will not only search for

http://www.businessinsider.com/the-worlds-first-artificially-intelligent-lawyer-gets-hired-2016-5

Limitations of AI

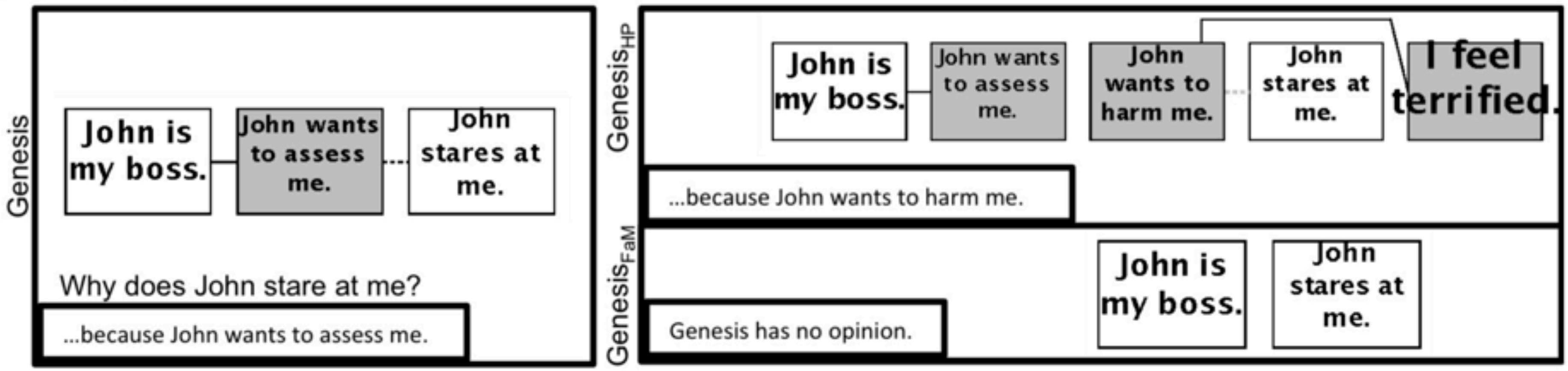
- Weak AI vs. Strong AI / Superintelligence
- Computational intelligence $\neq$ computational consciousness
- Procedure
- Emotion
- Thinking, reasoning and decision making
  - Inference
  - Abstraction
  - Cognition
  - Commonsense knowledge
  - Insight

http://knowyourmeme.com/photos/234739-i-have-no-idea-what-i-m-doing

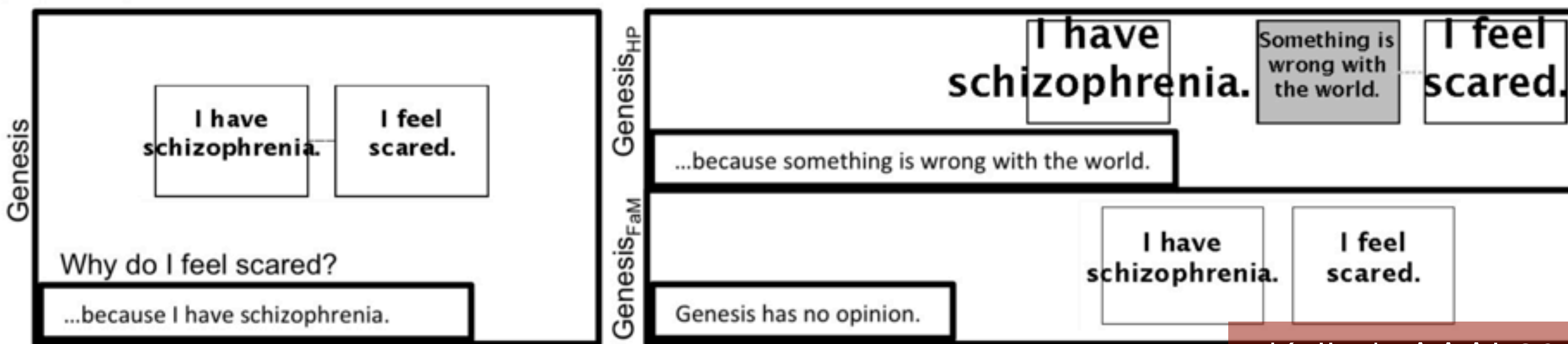
Inducing Schizophrenia in an Artificially Intelligent Story-Understanding System

Pratyusha Kalluri and Patrick Henry Winston

(B) System performance on the Paranoid Delusion Task



(C) System performance on the Persistence of Delusion Task



Kalluri, AAAI 2017

- Not replace, but may change significantly

- *It was the best of times, it was the worst of times*

  A Tale of Two Cities

  Charles Dickens

VIEWPOINT

INNOVATIONS IN HEALTH CARE DELIVERY

Adapting to Artificial Intelligence

Radiologists and Pathologists as Information Specialists

Saurabh Jha, MBBS, MRCS, MS
Department of Radiology, University of Pennsylvania, Philadelphia.

Eric J. Topol, MD
Scripps Research Institute, La Jolla, California.

Artificial intelligence—the mimicking of human cognition by computers—was once a fable in science fiction but is becoming reality in medicine. The combination of big data and artificial intelligence, referred to by some as the fourth industrial revolution,[1] will change radiology and pathology along with other medical specialties. Although reports of radiologists and pathologists being replaced by computers seem exaggerated,[2] these specialties must plan strategically for a future in which artificial intelligence is part of the health care workforce.

Radiologists have always revered machines and tech-

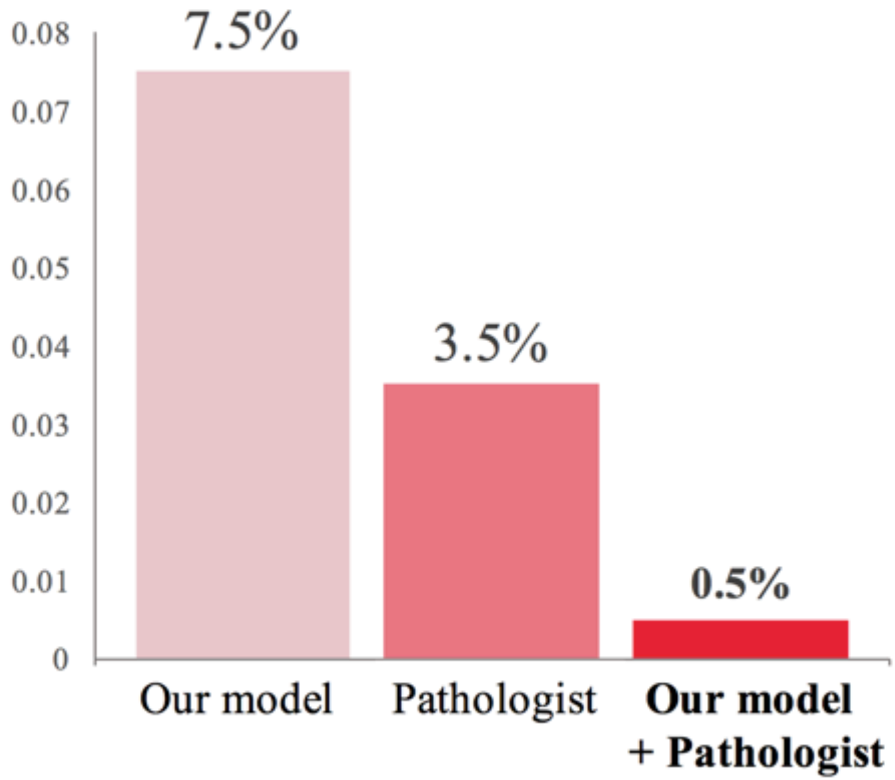
This progress in i
radiologists. Radiolog
ages, such as chest rad
plex and data rich. Cr
and magnetic reson
greater clarity, has m
stances; for example, a
a chest radiograph but
has come at a price—
markedly. For exam
4000 images in a CT

Jha, JAMA 2016

What We Can Do

- Find good problems, collect reliable data > Big data and algorithms
- Collaboration > Expert or AI only
  - Expert: high-level integration, interpretation and decision making (learn Bayesian logic, statistics, and data science and be aware of other sources of information!)
  - AI: pattern recognition and massive repetitive tasks
- Sharing
  - AI researchers open-source platforms and algorithms
  - You can open-source data and knowledge
  - We can do nothing without sharing!
- Experience
  - Learning from AI
  - Communicating with AI

Deep Learning vs Pathologist



The **combination** of a pathologist and the Beck Lab deep learning system **reduces error rate** by 85% **to 0.5%.**

Courtesy by Dr. Andrew Beck (PathAI)

Take Home Message

- Data-driven, machine learning-based approach
- Define good problems and find reliable data sources
  - Cost, risk, precision medicine
  - Causality, bias, unstructured data, interpretability, reasoning and decision making
- Human-machine collaboration
- Share your data and knowledge
- Learn from / communicate with AI

- Acknowledgement
  - Leo Celi (MIT), Mornin Feng (NUS), Alvin Marcelo (UP-AeHIN)
  - Peter Szolovits (MIT), David Sontag (MIT)

- ckbjimmy@mit.edu / Wei-Hung Weng (LinkedIn)