THE UNIVERSITY OF TEXAS
MD Anderson ~~Cancer~~ Center
Making Cancer History®

Artificial Intelligence and Machine Learning in Diagnostic Imaging

Kevin W. McEnery, MD
Professor of Radiology
Director Innovation Imaging Informatics
kmcenery@mdanderson.org

Objectives

- Discuss opportunity for machine learning in diagnostic imaging
- Review potential roadblocks for fully leveraging machine learning processes
- Appreciate machine learning integration opportunities beyond imaging interpretation

AI and Imaging: Overview

- Will AI Replace Radiologists?
- Overview of machine learning in image analysis
- Barriers to machine learning
- AI imaging use cases beyond image interpretation
- Potential AI impact on radiologist-provider interactions
- Conclusion

Will AI Replace Radiologists?

Fig 1. Word cloud of titles of the top 25 nonscientific results from a Google (Menlo Park, California) search of the terms "artificial intelligence radiology."

Yi, Paul H. et al., Artificial Intelligence and Radiology: Collaboration Is Key
Journal of the American College of Radiology , Volume 15 , Issue 5 , 781 - 783

NOVEMBER 15, 2017

Stanford algorithm can diagnose pneumonia better than radiologists

Stanford researchers have developed a deep learning algorithm that evaluates chest X-rays for signs of disease. In just over a month of development, their algorithm outperformed expert radiologists at diagnosing pneumonia.

BY TAYLOR KUBOTA

Stanford researchers have developed an algorithm that offers diagnoses based off chest X-ray images. It can diagnose up to 14 types of medical conditions and is able to diagnose pneumonia better than expert radiologists working alone. A paper about the algorithm, called CheXNet, was published Nov. 14 on the open-access, scientific preprint website arXiv.

"Interpreting X-ray images to diagnose pathologies like pneumonia is very challenging, and we know that there's a lot of variability in the diagnoses radiologists arrive at," said Pranav Rajpurkar, a graduate student in the Stanford Machine Learning Group and co-lead author of the paper. "We became interested in developing machine learning algorithms that could learn from hundreds of thousands of chest X-ray diagnoses and make accurate diagnoses."

Radiologist Matthew Lungren, left, meets with graduate students Jeremy Irvin and Pranav Rajpurkar to discuss the results of detections made by the algorithm. A tool the researchers developed along with the algorithm produced these images, which are similar to heat maps and show the areas of the X-ray most indicative of pneumonia. *(Image credit: L.A. Cicero)*

The work uses a public dataset initially released by the National Institutes of Health Clinical Center on Sept. 26. That dataset contains 112,120 frontal-view chest X-ray images labeled with up to 14 possible pathologies. It was released in tandem with an algorithm that could diagnose many of those 14 pathologies with some success, designed to encourage others to advance that work. As soon as they saw these materials, the Machine Learning Group – a group led by Andrew Ng, adjunct professor of computer science – knew it had found its next research direction.

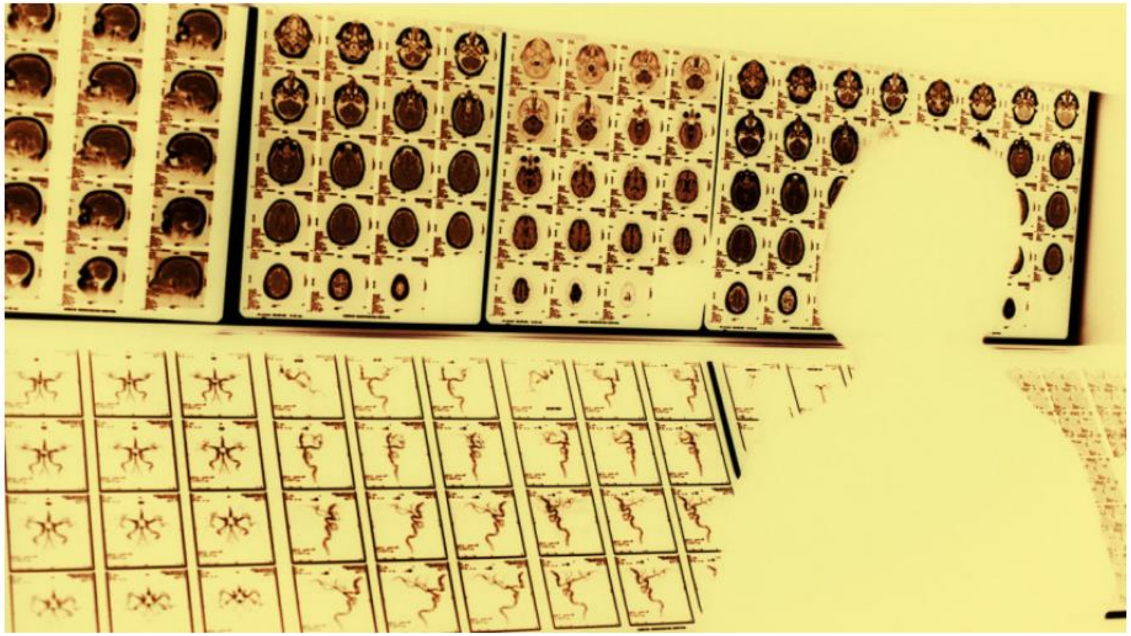
Harvard Business Review

INNOVATION

AI Will Change Radiology, but It Won’t Replace Radiologists

by Thomas H. Davenport and Keith J. Dreyer, DO

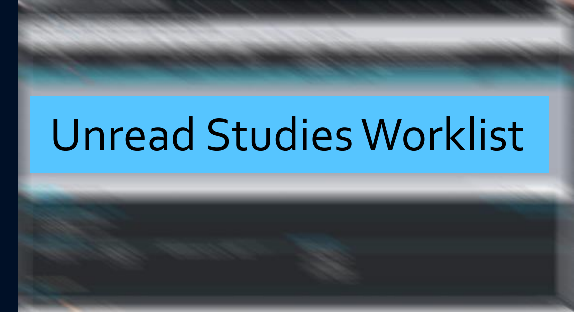
MARCH 27, 2018



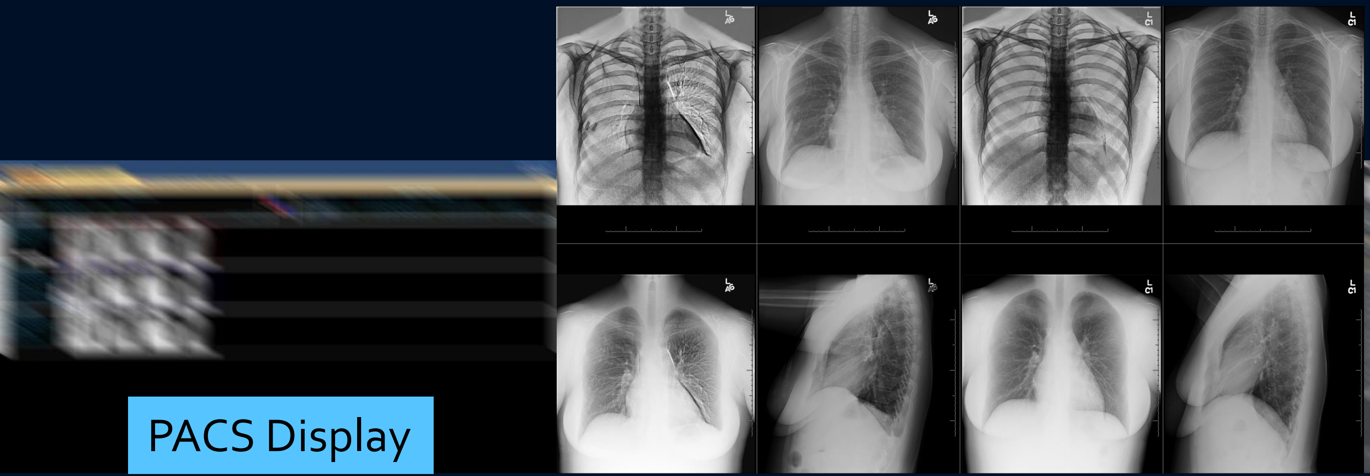
CORBIS/VCG/GETTY IMAGES

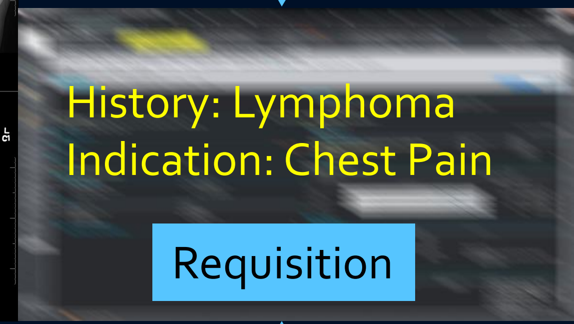
Recent advances in artificial intelligence have led to speculation that AI might one day replace human radiologists. Researchers have developed deep learning neural networks that can identify pathologies in radiological images such as bone fractures and potentially cancerous lesions, in some cases more reliably than an average radiologist. For the most part, though, the best systems are currently on par with human performance and are used only in research settings.

Current MDACC radiologist's workspace…



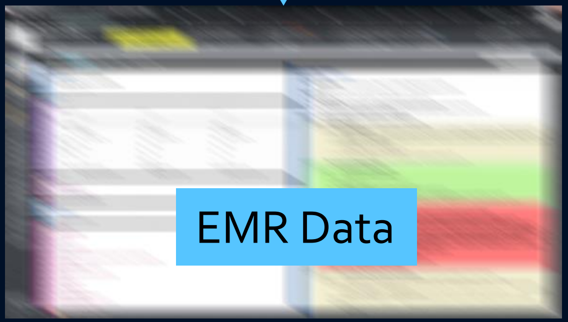
Implements PACS-EMR Integration





-18 years of prior study images available for immediate retrieval
-16,000 studies/month originating from outside imaging centers.

Radiology leverages Same Snapshot as providers



70 inch display span

Automatic Display of current and relevant prior images

History: Lymphoma
Indication: Chest Pain

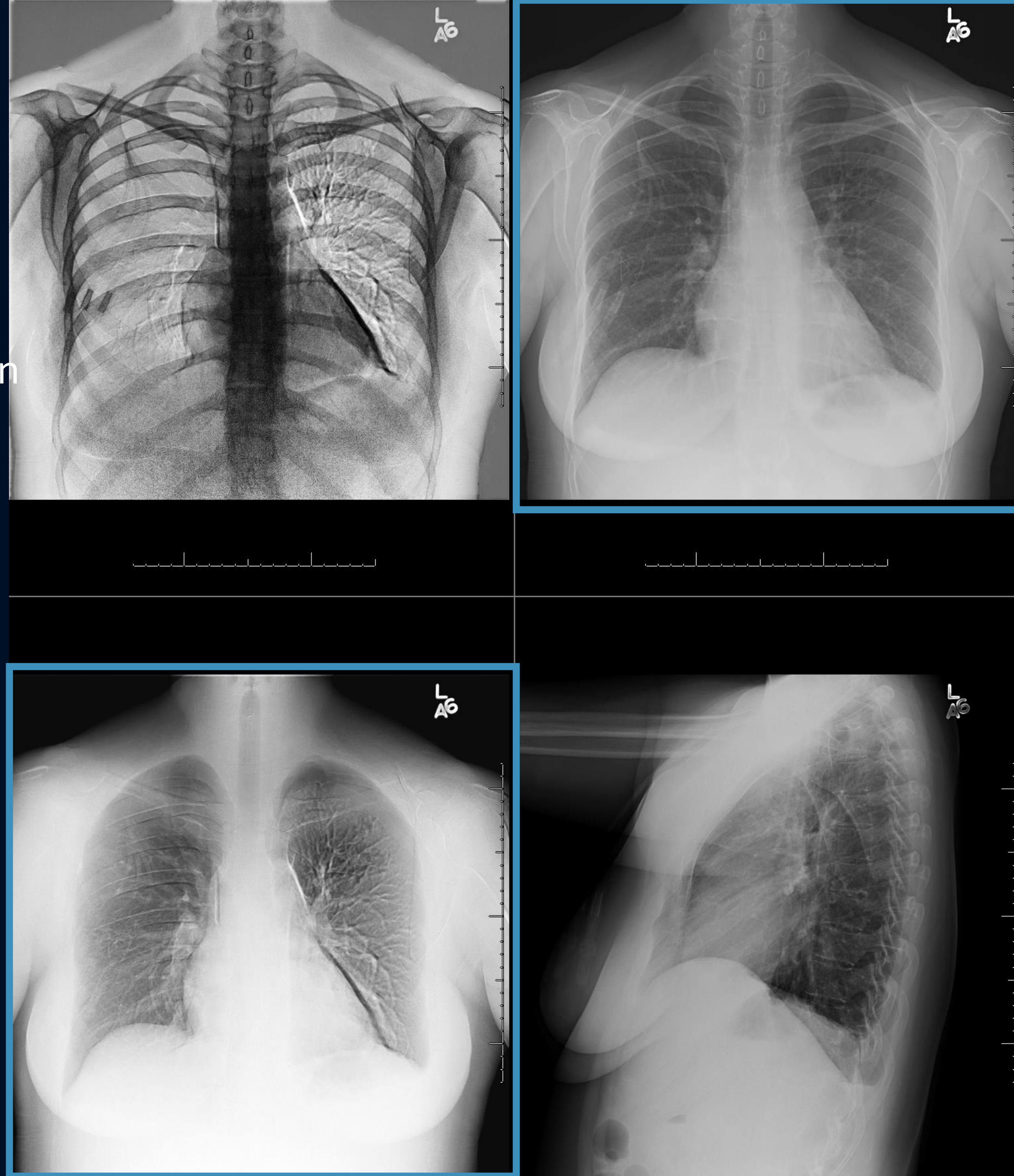
History: Lymphoma
Indication: Chest Pain

Diagnosis: Left PTX, < 15%

RVU: 0.2

Patient Outcome: Observation

Time Spent: < 2 minutes



History: Lymphoma
Indication: Chest Pain

Diagnosis: Left PTX, < 15%

RVU: 0.2

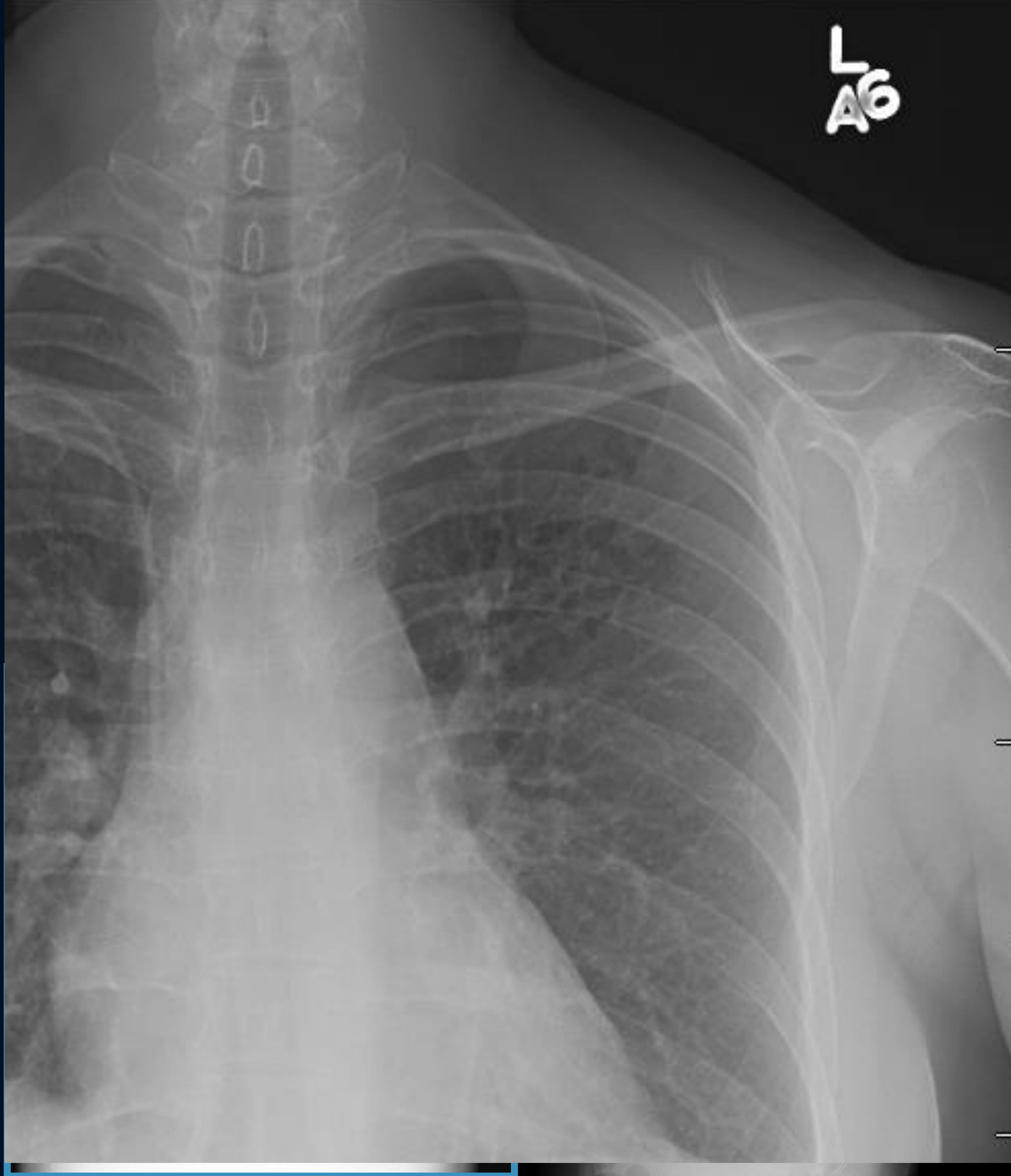
Patient Outcome: Observation

Time Spent: < 2 minutes

No need to include:

"Clinical Correlation is Recommended"

No annotations placed.



Stanford | News

Search Stanford news...

Home Find Stories For Journalists Contact

NOVEMBER 15, 2017

Stanford algorithm can diagnose pneumonia better than radiologists

Stanford researchers have developed a deep learning algorithm that evaluates chest X-rays for signs of disease. In just over a month of development, their algorithm outperformed expert radiologists at diagnosing pneumonia.

BY TAYLOR KUBOTA

Stanford researchers have developed an algorithm that offers diagnoses based off chest X-ray images. It can diagnose up to 14 types of medical conditions and is able to diagnose pneumonia better than expert radiologists working alone. A paper about the algorithm, called CheXNet, was published Nov. 14 on the open-access, scientific preprint website arXiv.

"Interpreting X-ray images to diagnose pathologies like pneumonia is very challenging, and we know that there's a lot of variability in the diagnoses radiologists arrive at," said Pranav Rajpurkar, a graduate student in the Stanford Machine Learning Group and co-lead author of the paper. "We became interested in developing machine learning algorithms that could learn from hundreds of thousands of chest X-ray diagnoses and make accurate diagnoses."

Radiologist Matthew Lungren, left, meets with graduate students Jeremy Irvin and Pranav Rajpurkar to discuss the results of detections made by the algorithm. A tool the researchers developed along with the algorithm produced these images, which are similar to heat maps and show the areas of the X-ray most indicative of pneumonia. *(Image credit: L.A. Cicero)*

The work uses a public dataset initially released by the National Institutes of Health Clinical Center on Sept. 26. That dataset contains 112,120 frontal-view chest X-ray images labeled with up to 14 possible pathologies. It was released in tandem with an algorithm that could diagnose many of those 14 pathologies with some success, designed to encourage others to advance that work. As soon as they saw these materials, the Machine Learning Group – a group led by Andrew Ng, adjunct professor of computer science – knew it had found its next research direction.

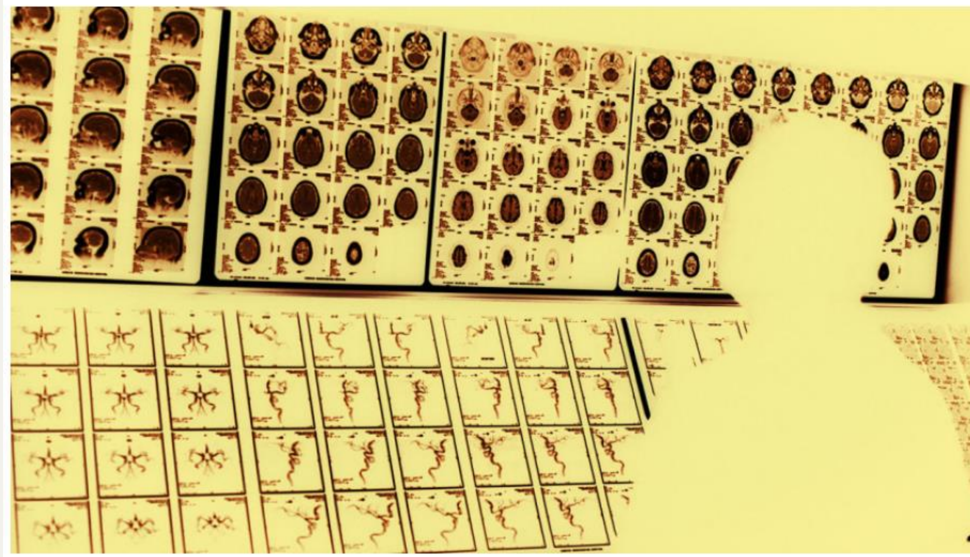
Harvard Business Review

INNOVATION

AI Will Change Radiology, but It Won't Replace Radiologists

by Thomas H. Davenport and Keith J. Dreyer, DO

MARCH 27, 2018



CORBIS/VCG/GETTY IMAGES

Recent advances in artificial intelligence have led to speculation that AI might one day replace human radiologists. Researchers have developed deep learning neural networks that can identify pathologies in radiological images such as bone fractures and potentially cancerous lesions, in some cases more reliably than an average radiologist. For the most part, though, the best systems are currently on par with human performance and are used only in research settings.

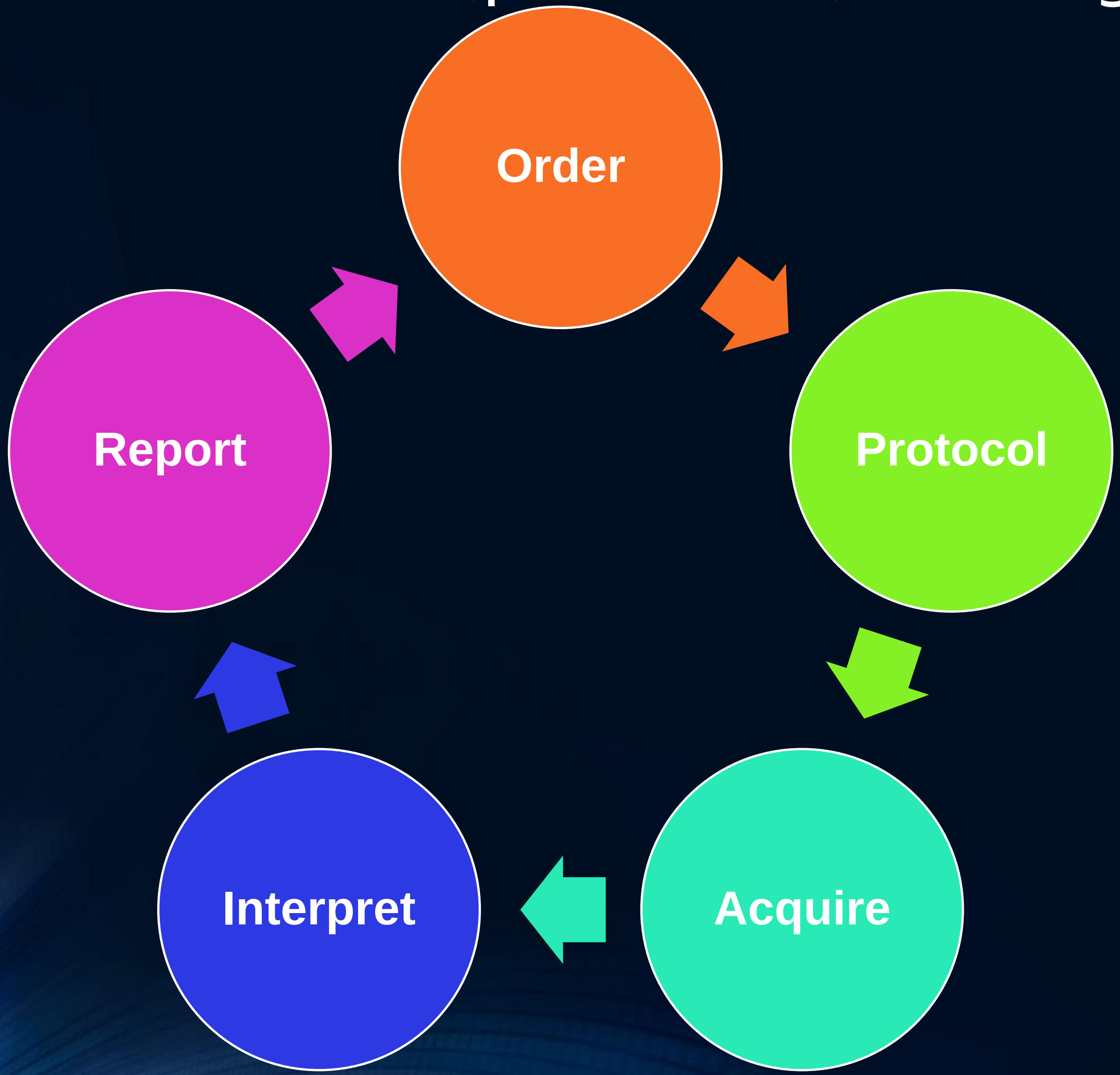
Will AI Replace Radiologists?
Answer: Red Pill or Blue Pill?

HYPE

REALITY

This is your last Chance…. After this, there is no turning back.
You take the blue pill, the story ends. You wake up in your bed and believe whatever you want to believe.
You take the red pill, you stay in Wonderland and I'll show you how deep the rabbit-hole goes..... Remember, all I'm offering you it's the truth, nothing more…

Imaging Process: More than identification of (pertinent) findings



Imaging Value Process: Patient Context

- Orders
  - Appropriate for the patient's complete presentation
- Protocols
  - Optimized to inform the clinical decision process
- Acquisition
  - Optimized to inform at safest level, greatest clinical data
- Interpretation
  - Focus on findings pertinent to patient
- Reports
  - Optimized to efficiently communicate and inform the care process as well as the patient

Machine Learning

- Subfield of computer science that gives computers the ability to learn without being explicitly programmed
- Simply, ML is the science of teaching computers how to learn, in an effort to glean information from data that more conventional statistical approaches may not be able to achieve

Machine Learning

- Arises at the intersection of statistics, which seeks to learn relationships from data, and computer science, with its emphasis on efficient computing algorithms.
- Evolved from the study of pattern recognition and computational learning theory in artificial intelligence.
- Explores the study and construction of algorithms that can learn from and make predictions on data.
- Goal of ML algorithm is to develop a mathematical model that fits the data.

Image-net.org

- ImageNet collaboration maintains a large dataset – (approximately 14 million images) labeled with nouns related to the content of each image
- Sponsors annual competition to test accuracy of image recognition algorithms

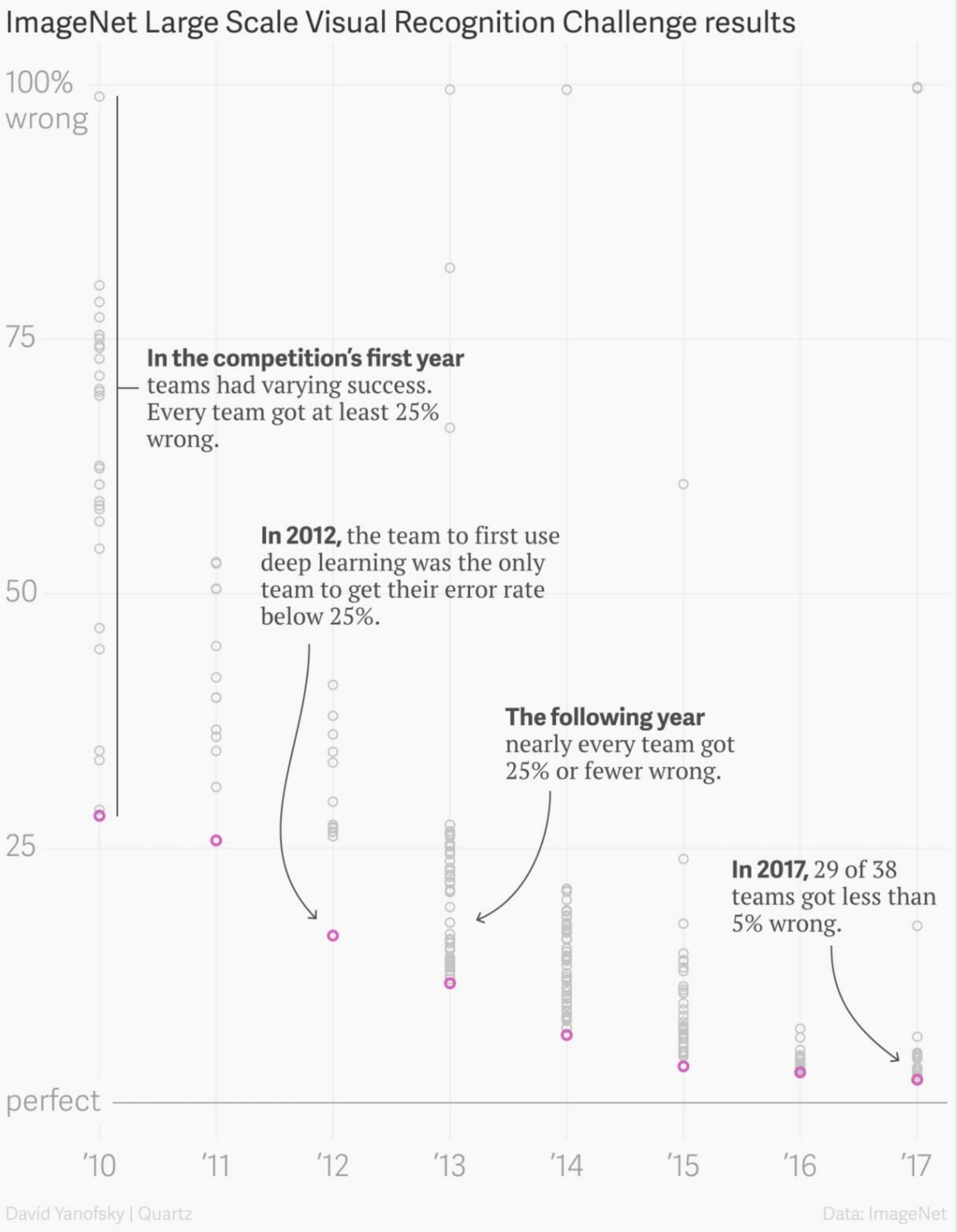


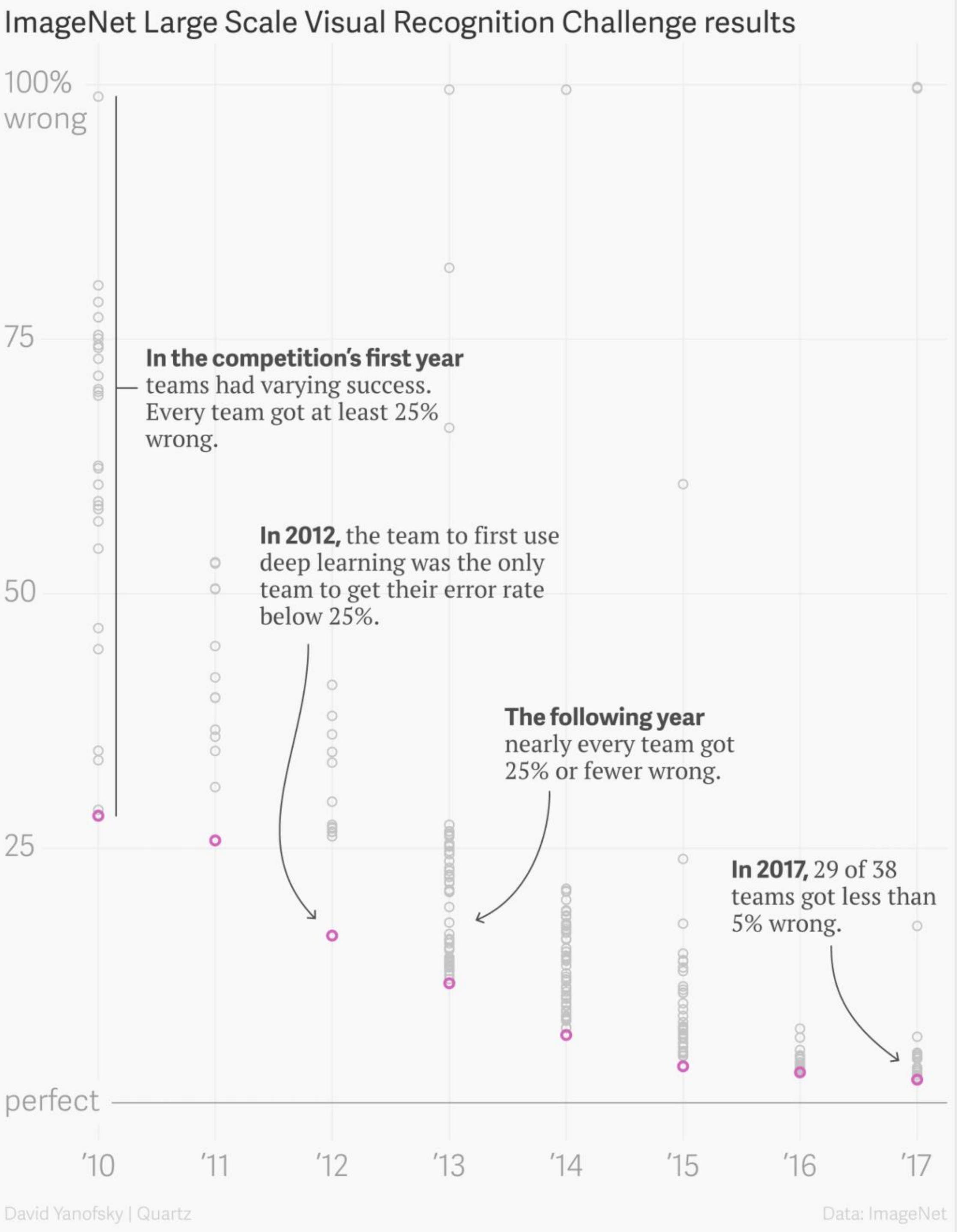
Image Classification…

Blueberry muffin or chihuahua?

Fried chicken or labradoodle?

Image-net.org

- ImageNet collaboration maintains a large dataset – (approximately 14 million images) labeled with nouns related to the content of each image
- Sponsors annual competition to test accuracy of image recognition algorithms



CAD vs. Machine Learning

- Computer-aided diagnosis
  - Assessment of pre-determined image characteristics
  - Prior knowledge of association to disease
    - Example: Mammography - Micro-calcifications
- Machine Learning
  - Image analysis without pre-determined characteristics
  - Analysis process groups images through “identified characteristics”
    - Identified features may or may not be determined
  - Outcome: observe if the grouping properly classifies known disease processes

CAD vs. Machine Learning

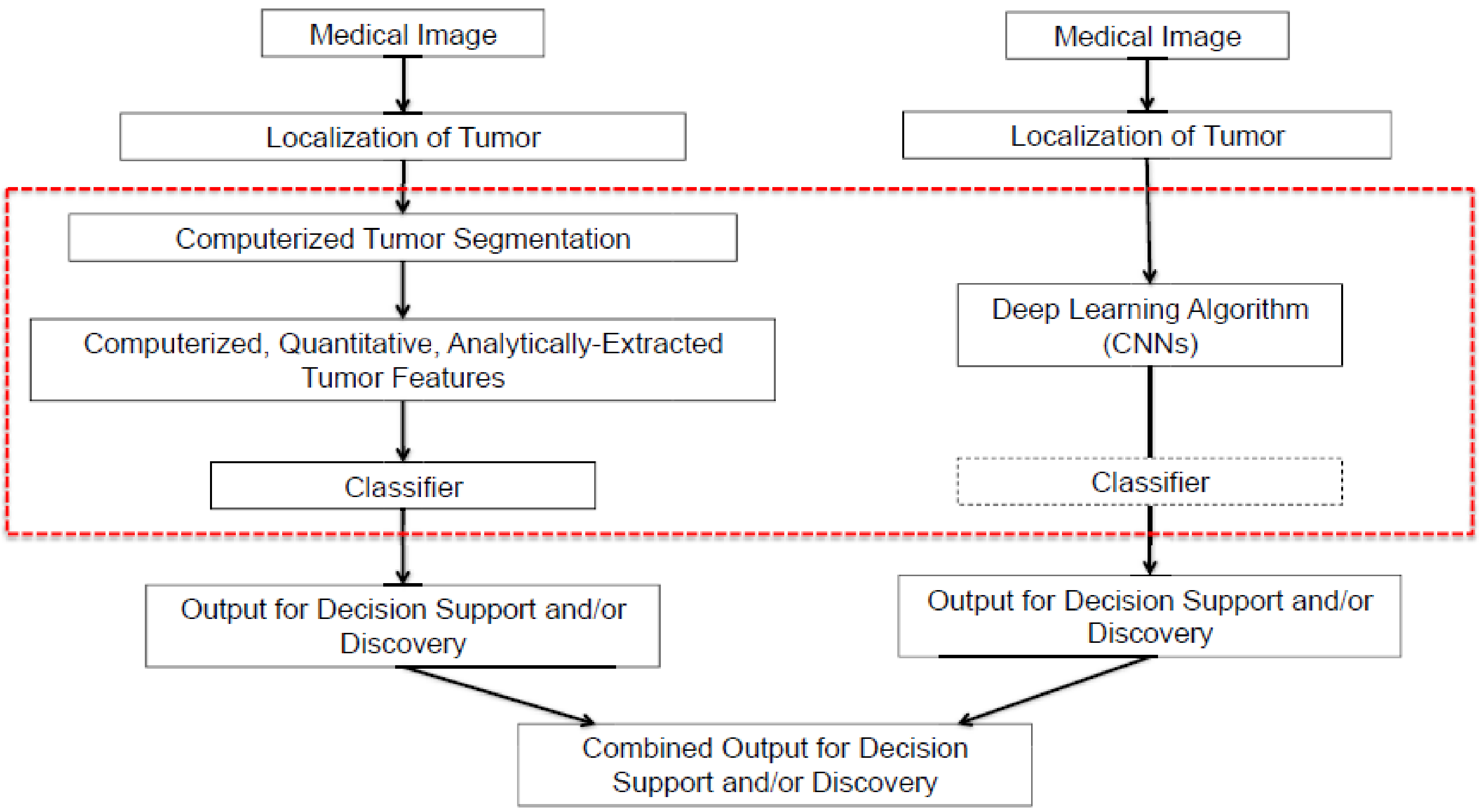
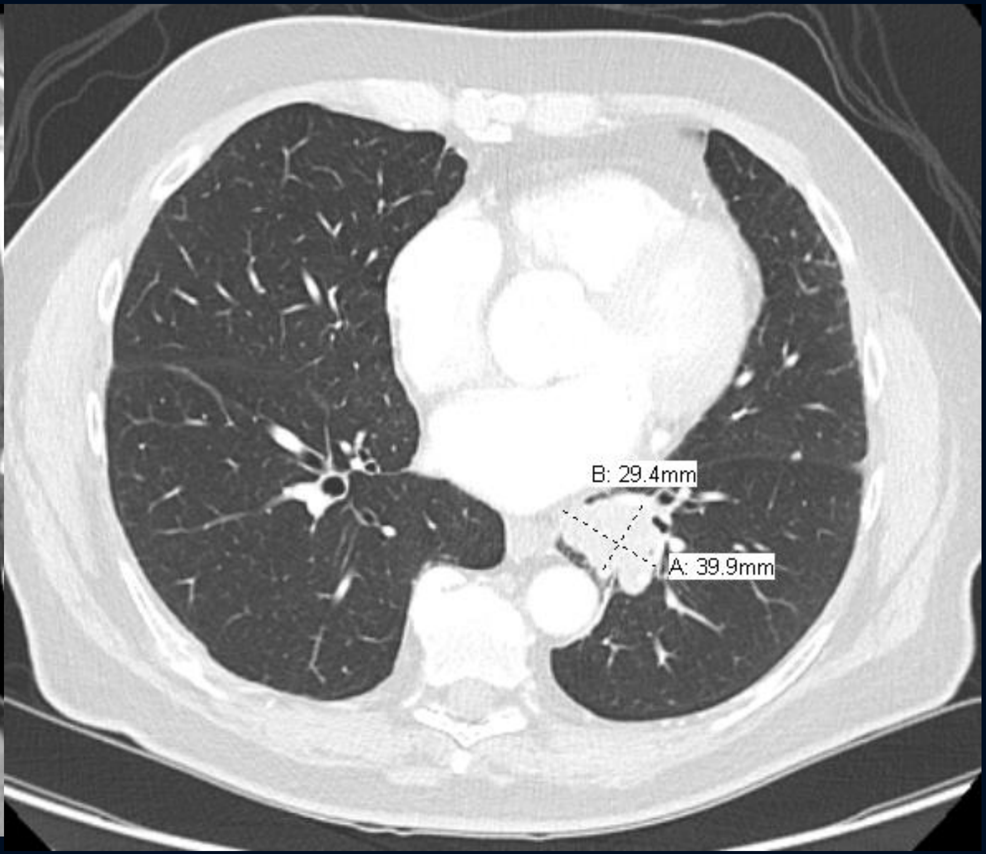


Fig 5. Schematic demonstrating the comparison of conventional hand-crafted computer-aided diagnosis and radiomic features, convolutional neural network (CNN)-extracted features, and an ensemble technique in the task of distinguishing between lesion type as used in Antropova et al [37] and Huynh et al [42].

Giger, ML , Machine Learning in Medical Imaging.
J Am Coll Radiol 2018;15:512-520

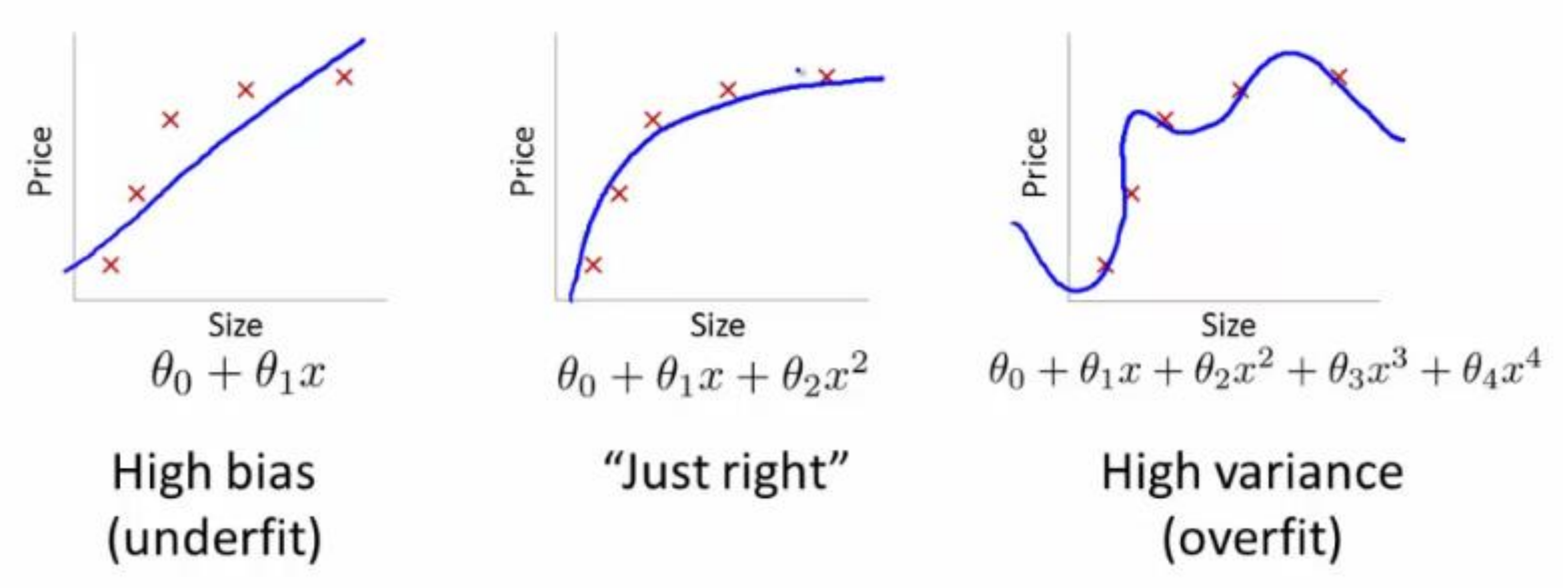
Radiology image classification…



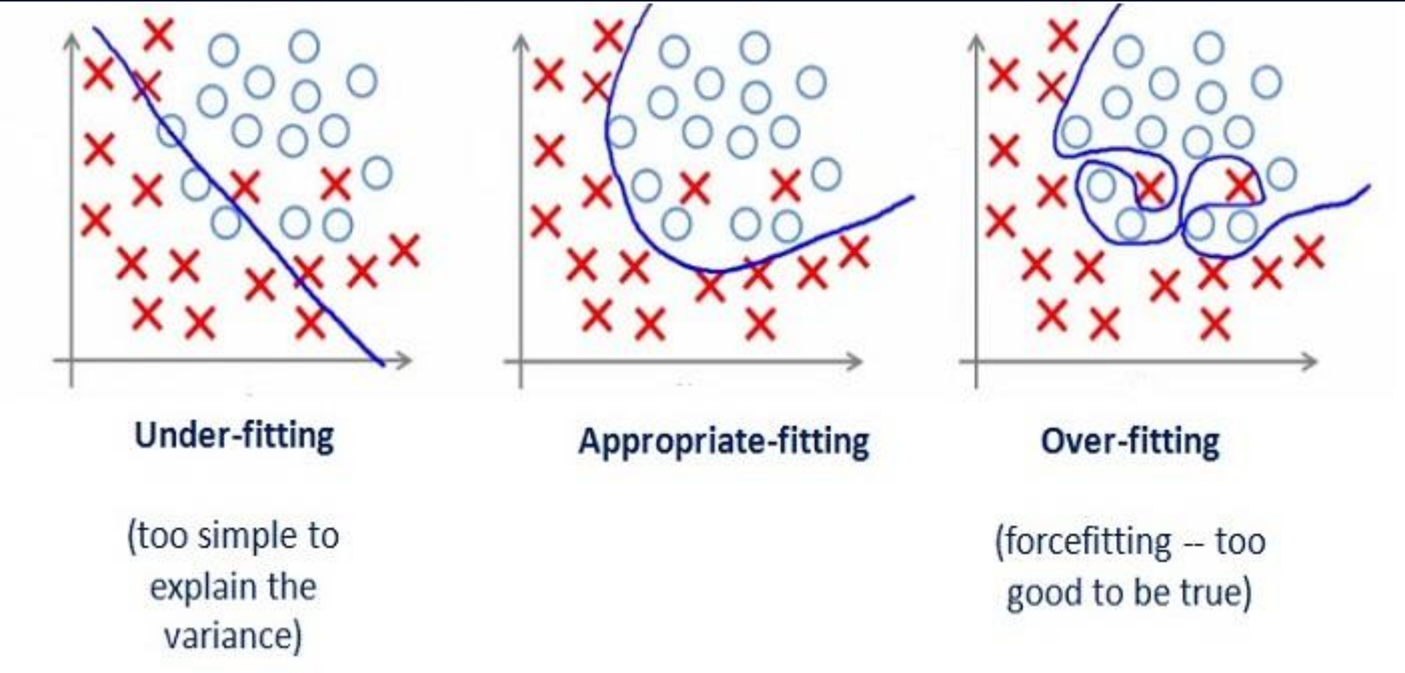
- Abnormality present (y/n)
- Infection, lung cancer, metastasis?
- Staging of detected lesion – location, size, characteristics

Machine Learning Data Sets

- Training
  - Image set to initially establish “hyper-parameters”
- Test
  - Dataset to determine outcome of training
- Validation
  - Final test of Algorithm
    - Dataset not previously “known” to algorithm
  - Determination accuracy (sensitivity/specificity)
  - Avoid “overfitting” of the algorithm
    - Validation cases should be accounted for in the basis of algorithm
    - “Retesting” validation set leads to overfitting



https://i.stack.imgur.com/tozit.png



https://vitalflux.com/wp-content/uploads/2015/02/fittings.jpg

Development and Validation of Deep Learning Algorithms for Detection of Critical Findings in Head CT Scans

Sasank Chilamkurthy[1], Rohit Ghosh[1], Swetha Tanamala[1], Mustafa Biviji[2], Norbert G. Campeau[3], Vasantha Kumar Venugopal[4], Vidur Mahajan[4], Pooja Rao[1], and Prashant Warier[1]

[1]Qure.ai, Mumbai, IN
[2]CT & MRI Center, Nagpur, IN
[3]Department of Radiology, Mayo Clinic, Rochester, MN
[4]Centre for Advanced Research in Imaging, Neurosciences and Genomics, New Delhi, IN

Abstract

Importance Non-contrast head CT scan is the current standard for initial imaging of patients with head trauma or stroke symptoms.

Objective To develop and validate a set of deep learning algorithms for automated detection of following key findings from non-contrast head CT scans: intracranial hemorrhage (ICH) and its types, intraparenchymal (IPH), intraventricular (IVH), subdural (SDH), extradural (EDH) and subarachnoid (SAH) hemorrhages, calvarial fractures, midline shift and mass effect.

Design and Settings We retrospectively collected a dataset containing 313,318 head CT scans along with their clinical reports from various centers. A part of this dataset (Qure25k dataset) was used to validate and the rest to develop algorithms. Additionally, a dataset (CQ500 dataset) was collected from different centers in two batches B1 & B2 to clinically validate the algorithms.

Main Outcomes and Measures Original clinical radiology report and consensus of three independent radiologists were considered as gold standard for Qure25k and CQ500 datasets respectively. Area under receiver operating characteristics curve (AUC) for each finding was primarily used to evaluate the algorithms.

Results Qure25k dataset contained 21,095 scans (mean age 43.31; 42.87% female) while batches B1 and B2 of CQ500 dataset consisted of 214 (mean age 43.40; 43.92% female) and 277 (mean age 51.70; 30.31% female) scans respectively. On Qure25k dataset, the algorithms achieved AUCs of 0.9194, 0.8977, 0.9559, 0.9161, 0.9288 and 0.9044 for detecting ICH, IPH, IVH, SDH, EDH and SAH respectively. AUCs for the same on CQ500 dataset were 0.9419, 0.9544, 0.9310, 0.9521, 0.9731 and 0.9574 respectively. For detecting calvarial fractures, midline shift and mass effect, AUCs on Qure25k dataset were 0.9244, 0.9276 and 0.8583 respectively, while AUCs on CQ500 dataset were 0.9624, 0.9697 and 0.9216 respectively.

Conclusions and Relevance This study demonstrates that deep learning algorithms can accurately identify head CT scan abnormalities requiring urgent attention. This opens up the possibility to use these algorithms to automate the triage process. They may also provide a lower bound for quality and consistency of radiological interpretation.

- Non-contrast head CT
- 313,318 scans/reports
  - 21,000 validate "Cure 25"
  - 292,000 test
- Clinical validation "CQ500
  - 214 studies "B1"
  - 277 studies "B2"

Training >
Test >
Validation>

https://arxiv.org/abs/1803.05854

Development and Validation of Deep Learning Algorithms for Detection of Critical Findings in Head CT Scans

Sasank Chilamkurthy[1], Rohit Ghosh[1], Swetha Tanamala[1], Mustafa Biviji[2], Norbert G. Campeau[3],

Results Qure25k dataset contained 21,095 scans (mean age 43.31; 42.87% female) while batches B1 and B2 of CQ500 dataset consisted of 214 (mean age 43.40; 43.92% female) and 277 (mean age 51.70; 30.31% female) scans respectively. On Qure25k dataset, the algorithms achieved AUCs of 0.9194, 0.8977, 0.9559, 0.9161, 0.9288 and 0.9044 for detecting ICH, IPH, IVH, SDH, EDH and SAH respectively. AUCs for the same on CQ500 dataset were 0.9419, 0.9544, 0.9310, 0.9521, 0.9731 and 0.9574 respectively. For detecting calvarial fractures, midline shift and mass effect, AUCs on Qure25k dataset were 0.9244, 0.9276 and 0.8583 respectively, while AUCs on CQ500 dataset were 0.9624, 0.9697 and 0.9216 respectively.

Conclusions and Relevance This study demonstrates that deep learning algorithms can accurately identify head CT scan abnormalities requiring urgent attention. This opens up the possibility to use these algorithms to automate the triage process. They may also provide a lower bound for quality and consistency of radiological interpretation.

and 277 (mean age 51.70; 30.31% female) scans respectively. On Qure25k dataset, the algorithms achieved AUCs of 0.9194, 0.8977, 0.9559, 0.9161, 0.9288 and 0.9044 for detecting ICH, IPH, IVH, SDH, EDH and SAH respectively. AUCs for the same on CQ500 dataset were 0.9419, 0.9544, 0.9310, 0.9521, 0.9731 and 0.9574 respectively. For detecting calvarial fractures, midline shift and mass effect, AUCs on Qure25k dataset were 0.9244, 0.9276 and 0.8583 respectively, while AUCs on CQ500 dataset were 0.9624, 0.9697 and 0.9216 respectively.

Conclusions and Relevance This study demonstrates that deep learning algorithms can accurately identify head CT scan abnormalities requiring urgent attention. This opens up the possibility to use these algorithms to automate the triage process. They may also provide a lower bound for quality and consistency of radiological interpretation.

Scope of Algorithm

- Non-contrast head CT scans
- Pathology
  - Intracranial hemorrhage (ICH)
    - Intraparenchymal (IPH)
    - Intraventricular (IVH)
    - Subdural (SDH)
    - Extradural (EDH)
    - Subarachnoid (SAH) hemorrhages
  - Calvarial fractures
  - Midline shift
  - Mass effect

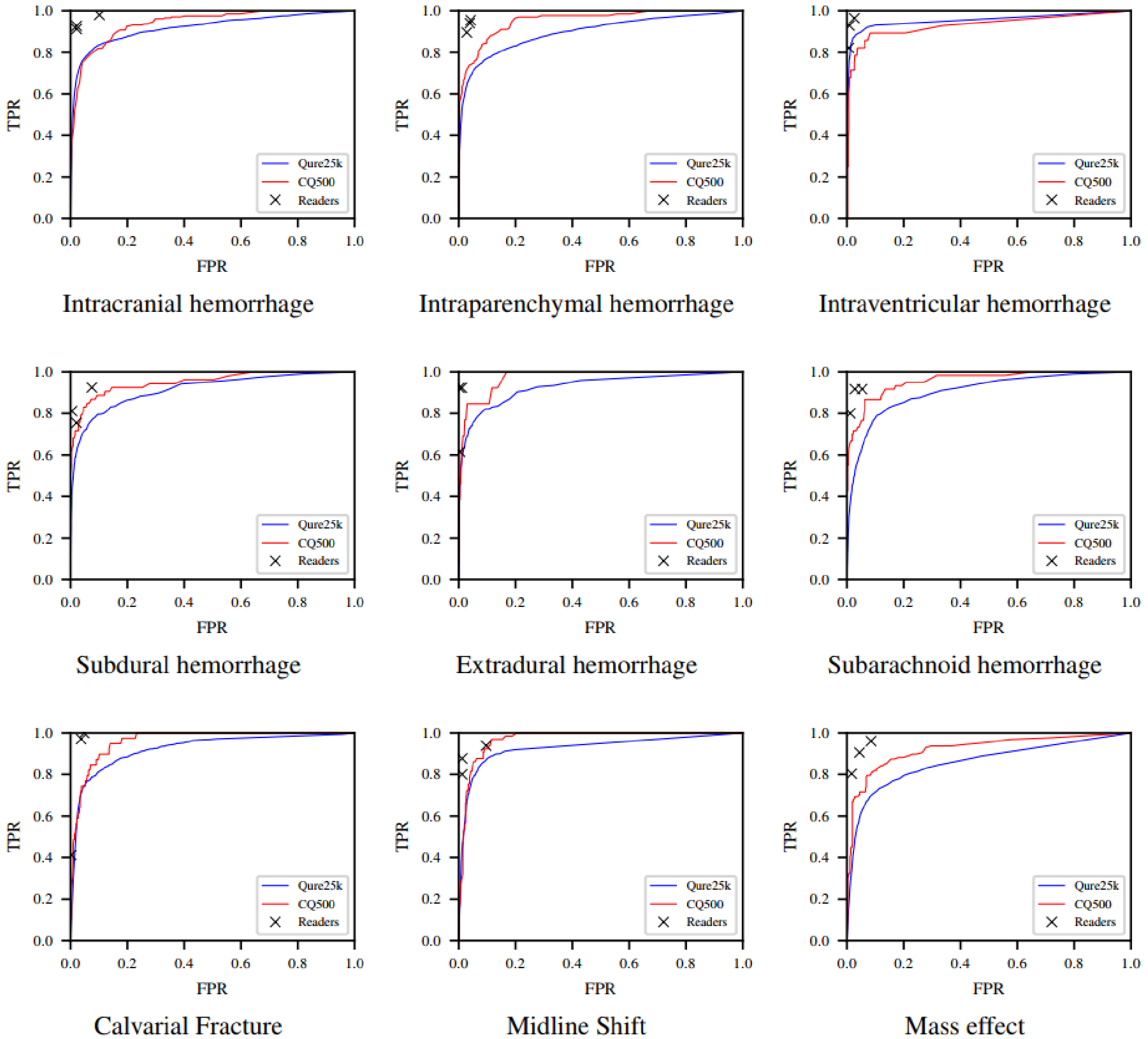
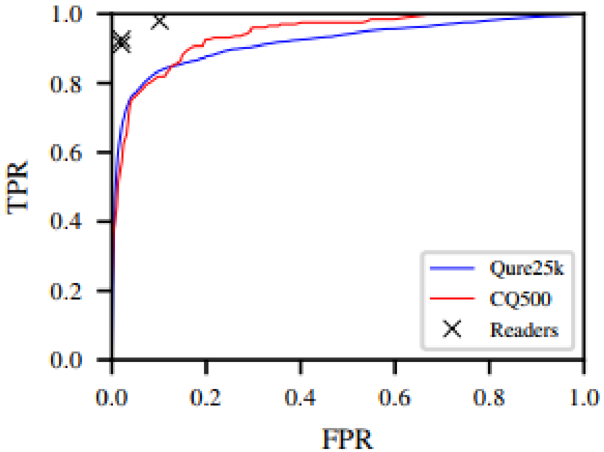
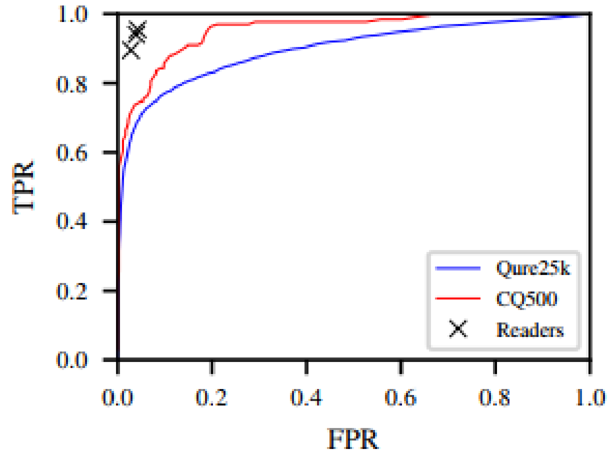


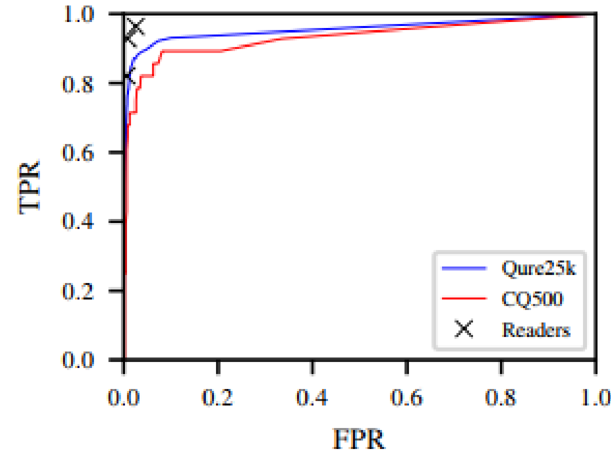
Figure 3: Receiver operating characteristic (ROC) curves for the algorithms on Qure25k and CQ500 datasets. Blue lines are for the Qure25k dataset and Red lines are for the CQ500 dataset. Readers' TPR and FPR against consensus on CQ500 dataset are plotted along with the ROCs for comparison



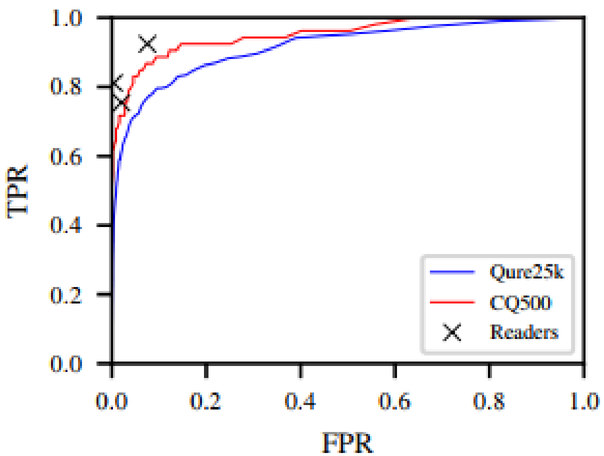
Intracranial hemorrhage



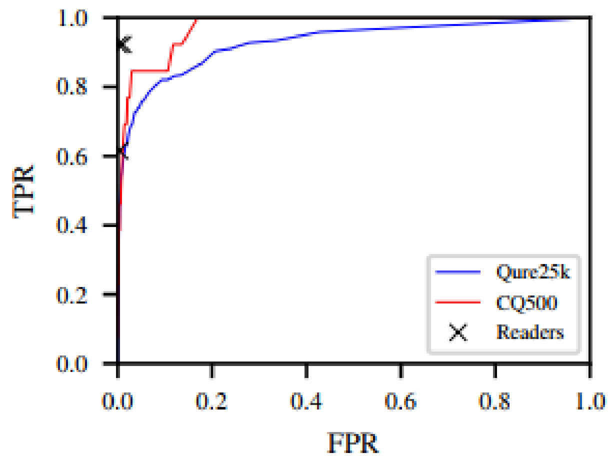
Intraparenchymal hemorrhage



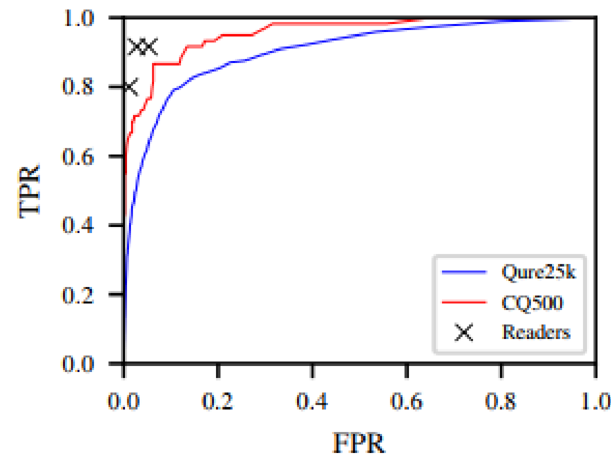
Intraventricular hemorrhage



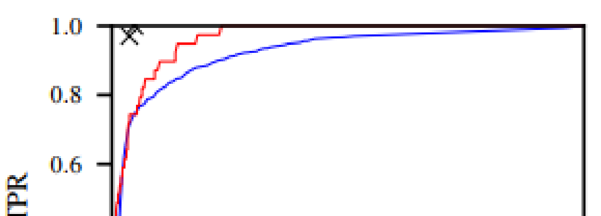
Subdural hemorrhage

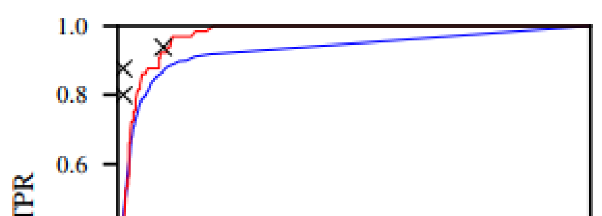


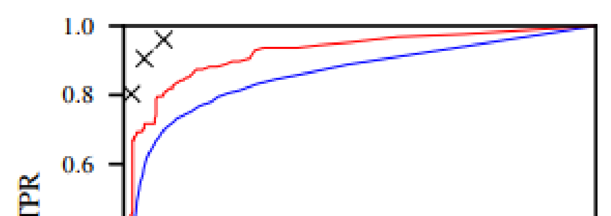
Extradural hemorrhage



Subarachnoid hemorrhage







AI: Pneumonia Detection

arXiv:1711.05225v3 [cs.CV] 25 Dec 2017

CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning

Pranav Rajpurkar [*1] Jeremy Irvin [*1] Kaylie Zhu [1] Brandon Yang [1] Hershel Mehta [1]
Tony Duan [1] Daisy Ding [1] Aarti Bagul [1] Robyn L. Ball [2] Curtis Langlotz [3] Katie Shpanskaya [3]
Matthew P. Lungren [3] Andrew Y. Ng [1]

Abstract

We develop an algorithm that can detect pneumonia from chest X-rays at a level exceeding practicing radiologists. Our algorithm, CheXNet, is a 121-layer convolutional neural network trained on ChestX-ray14, currently the largest publicly available chest X-ray dataset, containing over 100,000 frontal-view X-ray images with 14 diseases. Four practicing academic radiologists annotate a test set, on which we compare the performance of CheXNet to that of radiologists. We find that CheXNet exceeds average radiologist performance on the F1 metric. We extend CheXNet to detect all 14 diseases in ChestX-ray14 and achieve state of the art results on all 14 diseases.

"100,000 image publically available dataset"

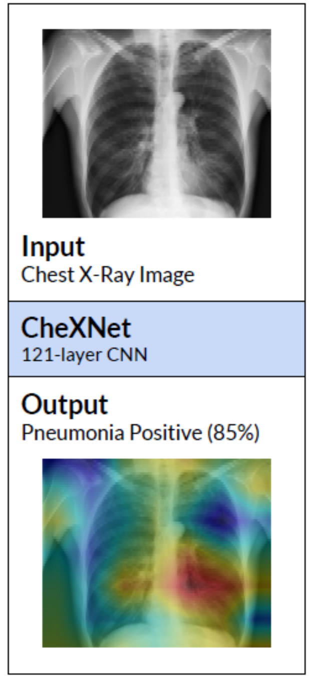
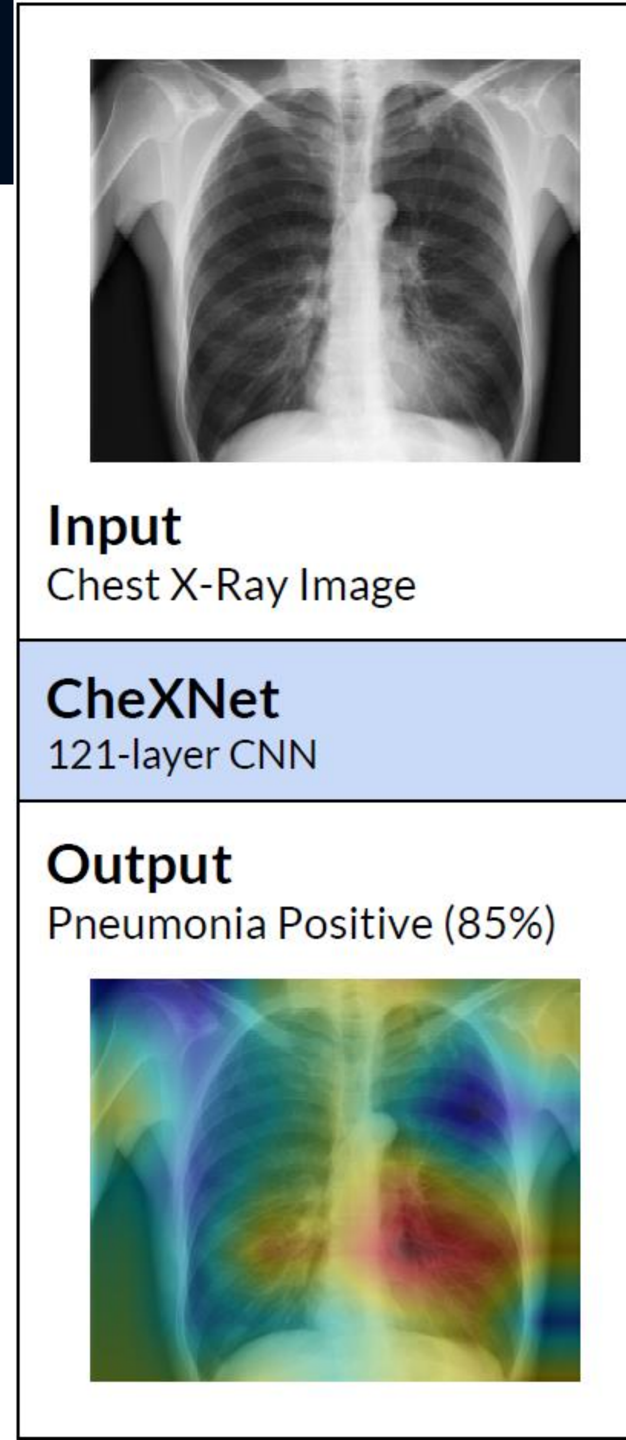


Figure 1. CheXNet is a 121-layer convolutional neural network that takes a chest X-ray image as input, and outputs the probability of a pathology. On this example, CheXnet correctly detects pneumonia and also localizes areas in the image most indicative of the pathology.

1. Introduction

More than 1 million adults are hospitalized with pneumonia and around 50,000 die from the disease every year in the US alone (CDC, 2017). Chest X-rays are currently the best available method for diagnosing pneumonia (WHO, 2001), playing a crucial role in clinical care (Franquet, 2001) and epidemiological studies (Cherian et al., 2005). However, detecting pneumonia in chest X-rays is a challenging task that relies on the availability of expert radiologists. In this work, we present a model that can automatically detect pneumonia from chest X-rays at a level exceeding practicing radiologists.

Our model, ChexNet (shown in Figure 1), is a 121-layer convolutional neural network that inputs a chest X-ray image and outputs the probability of pneumonia



Computer Science > Computer Vision and Pattern Recognition

CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning

Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, Matthew P. Lungren, Andrew Y. Ng

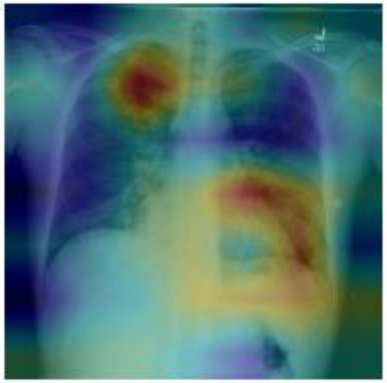
(Submitted on 14 Nov 2017 (v1), last revised 25 Dec 2017 (this version, v3))

We develop an algorithm that can detect pneumonia from chest X-rays at a level exceeding practicing radiologists. Our algorithm, CheXNet, is a 121-layer convolutional neural network trained on ChestX-ray14, currently the largest publicly available chest X-ray dataset, containing over 100,000 frontal-view X-ray images with 14 diseases. Four practicing academic radiologists annotate a test set, on which we compare the performance of CheXNet to that of radiologists. We find that CheXNet exceeds average radiologist performance on the F1 metric. We extend CheXNet to detect all 14 diseases in ChestX-ray14 and achieve state of the art results on all 14 diseases.

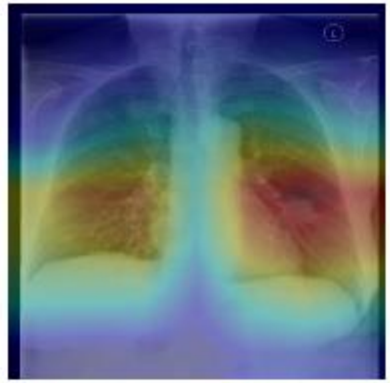
Subjects: **Computer Vision and Pattern Recognition (cs.CV)**; Learning (cs.LG); Machine Learning (stat.ML)
Cite as: **arXiv:1711.05225 [cs.CV]**
(or **arXiv:1711.05225v3 [cs.CV]** for this version)

Submission history

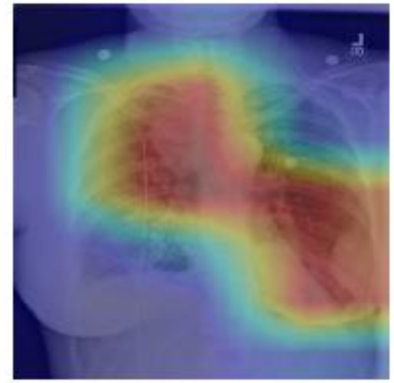
From: Pranav Rajpurkar [view email]
[v1] Tue, 14 Nov 2017 17:58:50 GMT (16273kb,D)
[v2] Sat, 25 Nov 2017 04:21:27 GMT (321kb,D)
[v3] Mon, 25 Dec 2017 11:09:06 GMT (321kb,D)



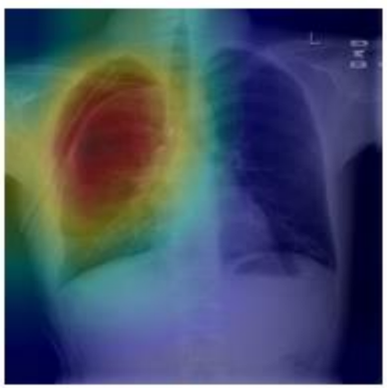
(a) Patient with multifocal community acquired pneumonia. The model correctly detects the airspace disease in the left lower and right upper lobes to arrive at the pneumonia diagnosis.



(b) Patient with a left lung nodule. The model identifies the left lower lobe lung nodule and correctly classifies the pathology.

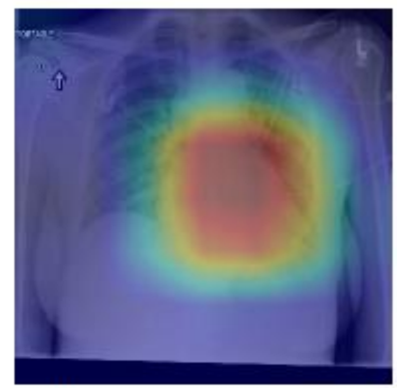


(c) Patient with primary lung malignancy and two large masses, one in the left lower lobe and one in the right upper lobe adjacent to the mediastinum. The model correctly identifies both masses in the X-ray.



(d) Patient with a right-sided pneumothroax and chest tube. The model detects the abnormal lung to correctly predict the presence of pneumothorax (collapsed lung).

(e) Patient with a large right pleural effusion (fluid in the pleural space). The model correctly labels the effusion and focuses on the right lower chest.



(f) Patient with congestive heart failure and cardiomegaly (enlarged heart). The model correctly identifies the enlarged cardiac silhouette.

Figure 2. CheXNet localizes pathologies it identifies using Class Activation Maps, which highlight the areas of the X-ray that are most important for making a particular pathology classification. The captions for each image are provided by one of the practicing radiologists.

Radiology

Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks[1]

Paras Lakhani, MD
Baskaran Sundaram, MD

Purpose: To evaluate the efficacy of deep convolutional neural networks (DCNNs) for detecting tuberculosis (TB) on chest radiographs.

Materials and Methods: Four deidentified HIPAA-compliant datasets were used in this study that were exempted from review by the institutional review board, which consisted of 1007 posteroanterior chest radiographs. The datasets were split into training (68.0%), validation (17.1%), and test (14.9%). Two different DCNNs, AlexNet and GoogLeNet, were used to classify the images as having manifestations of pulmonary TB or as healthy. Both untrained and pretrained networks on ImageNet were used, and augmentation with multiple preprocessing techniques. Ensembles were performed on the best-performing algorithms. For cases where the classifiers were in disagreement, an independent board-certified cardiothoracic radiologist blindly interpreted the images to evaluate a potential radiologist-augmented workflow. Receiver operating characteristic curves and areas under the curve (AUCs) were used to assess model performance by using the DeLong method for statistical comparison of receiver operating characteristic curves.

Results: The best-performing classifier had an AUC of 0.99, which was an ensemble of the AlexNet and GoogLeNet DCNNs. The AUCs of the pretrained models were greater than that of the untrained models ($P < .001$). Augmenting the dataset further increased accuracy (P values for AlexNet and GoogLeNet were .03 and .02, respectively). The DCNNs had disagreement in 13 of the 150 test cases, which were blindly reviewed by a cardiothoracic radiologist, who correctly interpreted all 13 cases (100%). This radiologist-augmented approach resulted in a sensitivity of 97.3% and specificity 100%.

Conclusion: Deep learning with DCNNs can accurately classify TB at chest radiography with an AUC of 0.99. A radiologist-augmented approach for cases where there was disagreement among the classifiers further improved accuracy.

[1] From the Department of Radiology, Thomas Jefferson University Hospital, Sidney Kimmel Jefferson Medical College, 132 S 10th St, Room 1080A, Main Building, Philadelphia, PA 19107-5244. Received October 5, 2016; revision requested November 23; revision received December 12; accepted January 9, 2017; final version accepted January 19. **Address correspondence to** P.L. (e-mail: *paras.lakhani@jefferson.edu*).

Table 3

AUC Test Dataset

Parameter	Untrained	Pretrained	Untrained with Augmentation*	Pretrained with Augmentation*
AlexNet	0.90 (0.84, 0.95)	0.98 (0.95, 1.00)	0.95 (0.90, 0.98)	0.98 (0.94, 0.99)
GoogLeNet	0.88 (0.81, 0.92)	0.97 (0.93, 0.99)	0.94 (0.89, 0.97)	0.98 (0.94, 1.00)
Ensemble				0.99 (0.96, 1.00)

Note.—Data in parentheses are 95% confidence interval.

* Additional augmentation of 90, 180, 270 rotations, and Contrast Limited Adaptive Histogram Equalization processing.

Figure 2

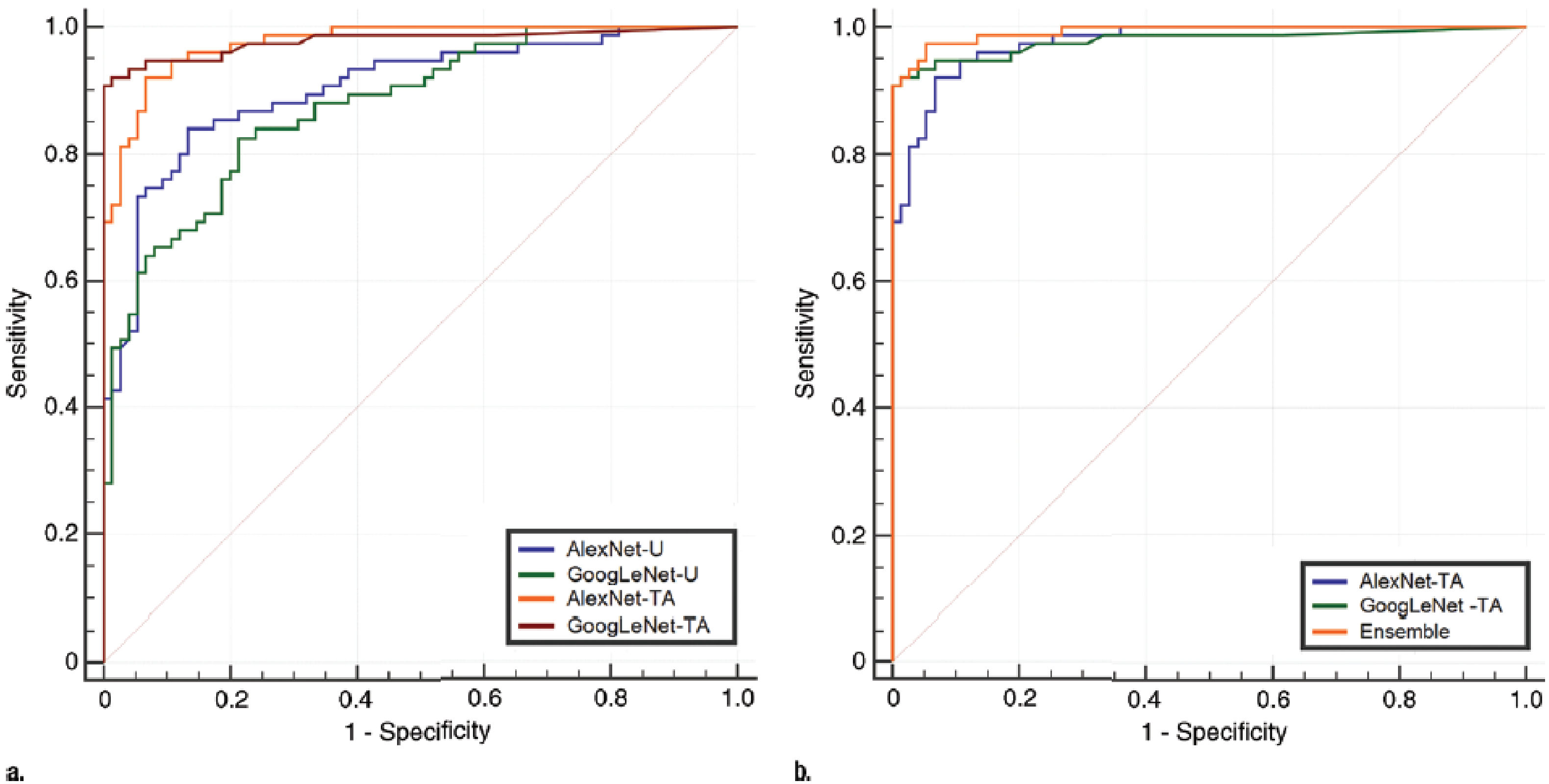


Figure 2: **(a)** Comparison of receiver operating characteristic curves for the untrained AlexNet-U and GoogLeNet-U models and pretrained with augmentation AlexNet-TA and GoogLeNet-TA models. The receiver operating characteristic curves for the AlexNet-TA and GoogLeNet-TA models had an AUC that was significantly greater than that for the untrained AlexNet-U and GoogLeNet-U models ($P < .001$) (Table 3). **(b)** Comparison of receiver operating characteristic curves for the AlexNet-TA, GoogLeNet-TA, and ensemble of the two models. The ensemble provided the best AUC (Table 3).

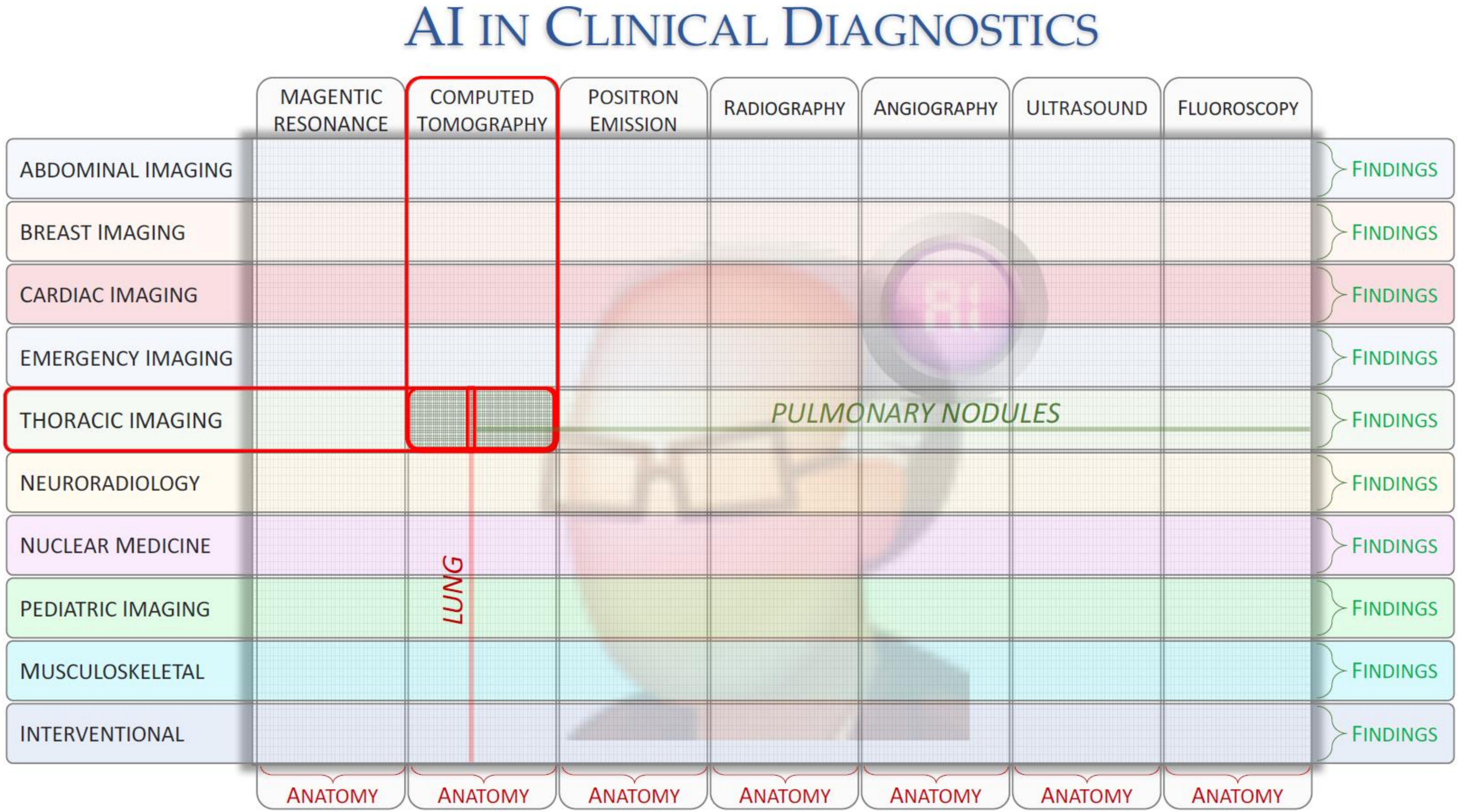
Barriers to Machine Learning in Radiology

- Limited publically available image training datasets
  - Current annotation processes FTE intensive
  - Radiology reporting not aligned to dataset annotation
- Lack of universal standard for image annotation
  - Several standard available (DICOM SR)
- Thousands of imaging use cases

AI IN CLINICAL DIAGNOSTICS

	MAGNETIC RESONANCE	COMPUTED TOMOGRAPHY	POSITRON EMISSION	RADIOGRAPHY	ANGIOGRAPHY	ULTRASOUND	FLUOROSCOPY	
ABDOMINAL IMAGING								FINDINGS
BREAST IMAGING								FINDINGS
CARDIAC IMAGING								FINDINGS
EMERGENCY IMAGING								FINDINGS
THROACIC IMAGING								FINDINGS
NEURORADIOLOGY								FINDINGS
NUCLEAR MEDICINE								FINDINGS
PEDIATRIC IMAGING								FINDINGS
MUSCULOSKELETAL								FINDINGS
INTERVENTIONAL								FINDINGS
	ANATOMY	ANATOMY	ANATOMY	ANATOMY	ANATOMY	ANATOMY	ANATOMY	

AI: Defining High Value Use Cases



Lung-RADS™ Version 1.0 Assessment Categories Release date: April 28, 2014

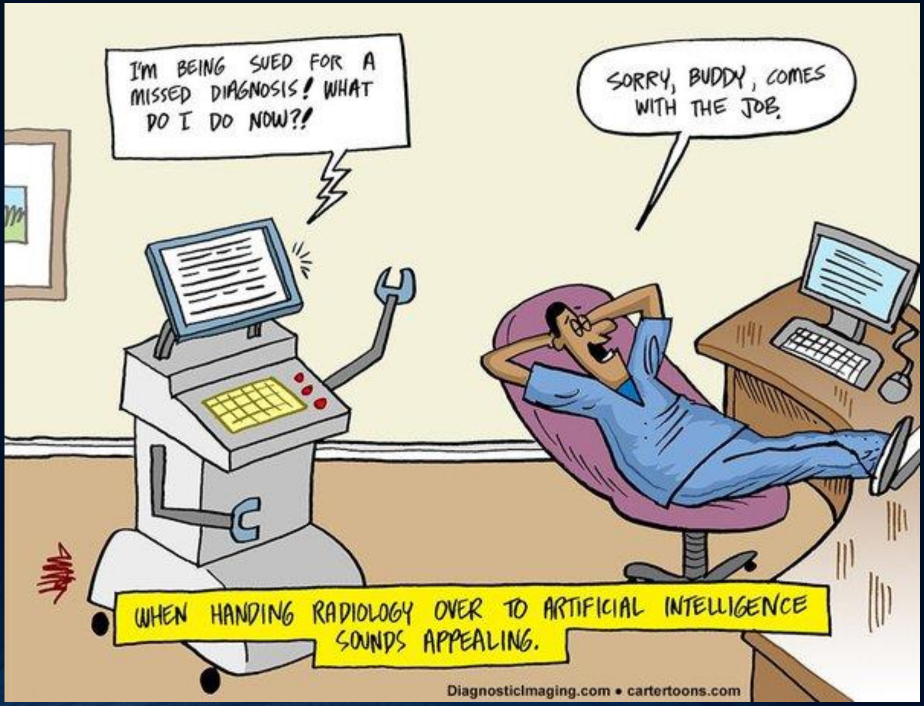
Category	Category Descriptor	Category	Findings	Management	Probability of Malignancy	Estimated Population Prevalence
Incomplete	-	0	prior chest CT examination(s) being located for comparison part or all of lungs cannot be evaluated	Additional lung cancer screening CT images and/or comparison to prior chest CT examinations is needed	n/a	1%
Negative	No nodules and definitely benign nodules	1	no lung nodules nodule(s) with specific calcifications: complete, central, popcorn, concentric rings and fat containing nodules	Continue annual screening with LDCT in 12 months	< 1%	90%
Benign Appearance or Behavior	Nodules with a very low likelihood of becoming a clinically active cancer due to size or lack of growth	2	solid nodule(s): < 6 mm new < 4 mm part solid nodule(s): < 6 mm total diameter on baseline screening non solid nodule(s) (GGN): < 20 mm OR ≥ 20 mm and unchanged or slowly growing category 3 or 4 nodules unchanged for ≥ 3 months			
Probably Benign	Probably benign finding(s) - short term follow up suggested; includes nodules with a low likelihood of becoming a clinically active cancer	3	solid nodule(s): ≥ 6 to < 8 mm at baseline OR new 4 mm to < 6 mm part solid nodule(s) ≥ 6 mm total diameter with solid component < 6 mm OR new < 6 mm total diameter non solid nodule(s) (GGN) ≥ 20 mm on baseline CT or new	6 month LDCT	1-2%	5%
Suspicious	Findings for which additional diagnostic testing and/or tissue sampling is recommended	4A	solid nodule(s): ≥ 8 to < 15 mm at baseline OR growing < 8 mm OR new 6 to < 8 mm part solid nodule(s: ≥ 6 mm with solid component ≥ 6 mm to < 8 mm OR with a new or growing < 4 mm solid component endobronchial nodule	3 month LDCT; PET/CT may be used when there is a ≥ 8 mm solid component	5-15%	2%
		4B	solid nodule(s) ≥ 15 mm OR new or growing, and ≥ 8 mm part solid nodule(s) with: a solid component ≥ 8 mm OR a new or growing ≥ 4 mm solid component	chest CT with or without contrast, PET/CT and/or tissue sampling depending on the *probability of malignancy and comorbidities. PET/CT may be used when there is a ≥ 8 mm solid component.	> 15%	2%
		4X	Category 3 or 4 nodules with additional features or imaging findings that increases the suspicion of malignancy			
	Clinically Significant or					

AI/ML Roadblocks

- Data Integrity
  - Collection and curation of vetted datasets
  - Costly data annotation (retrospective)
- Sharing and Pooling of Datasets
  - Models need large datasets to create, test and validate
  - Lots of ready consumers…
- Network Infrastructure
  - Image datasets and ancillary information
  - Cloud-based processes likely
- Privacy and security of pooled data
  - Anonymization processes

AI/ML Roadblocks

- FDA
  - Algorithm validation
    - Likely more formalized infrastructure
- Medical-legal
  - "Replace radiologist" – who assumes liability?



ML: Patient Scheduling

- MDACC unique challenge is large number of patients from outside of Houston area who on follow-up visits require imaging prior to clinical appointment.
- Machine Learning opportunity: Create patient schedules, based upon patient preference, optimize patient's schedule to reduce time in Houston on return visit
- Status: in pilot testing

Show PVD Schedule | Prior Patient | Next Patient | Next Conflict | Prv Img Conf | Next Img Conf

THE VILLAGES, Florida 3216

APPTtime	Tue MAIN	4/04 MAYS	ROC	Wed MAIN	4/05 MAYS	ROC	Thu MAIN	4/06 MAYS	ROC	Fri MAIN	4/07 MAYS	ROC
6:00 AM												
6:15 AM												
6:30 AM												
6:45 AM												
7:00 AM												
7:15 AM												
7:30 AM												
7:45 AM												
8:00 AM		DIAG CTR LAB~										
8:15 AM		LAB										
8:30 AM		CT~ IMAYS										
8:45 AM		CT CHEST WO CON										
9:00 AM		MAYS CT 15										
9:15 AM		85										
9:30 AM		--25--										
9:45 AM		--25--										
10:00 AM		PET CT~ IM										
10:15 AM		PET CT										
10:30 AM		MAYS PETCT A										
10:45 AM		150										
11:00 AM		150										
11:15 AM		150		ENDOCR CTR~								
11:30 AM		150		ENDOCRINE FOLLO								
11:45 AM		150										
12:00 PM		--30--										
12:15 PM		--30--										
12:30 PM												
12:45 PM												
1:00 PM												
1:15 PM												
1:30 PM												
1:45 PM												
2:00 PM												
2:15 PM												
2:30 PM		NEURO US~										
2:45 PM		US THYROID NIR										
3:00 PM		MAYS NEURO US B										
3:15 PM		--60--										
3:30 PM		--60--										
3:45 PM												
4:00 PM												
4:15 PM			MRI~ IROC									
4:30 PM			MRI ABDOMEN W									
4:45 PM			ROC MR 21									
5:00 PM			--60--									
5:15 PM			--60--									
5:30 PM			--60--									
5:45 PM			--60--									
6:00 PM			MRI~ IROC									
6:15 PM			MRI PELVIS W WO									
6:30 PM			ROC MR 21									
6:45 PM			60									
7:00 PM												
7:15 PM												
7:30 PM												
7:45 PM												
8:00 PM												
8:15 PM												
8:30 PM												
8:45 PM												
9:00 PM												
9:15 PM												
9:30 PM												
9:45 PM												
10:00 PM												

Show PVD Schedule | Prior Patient | Next Patient | Next Conflict | Prv Img Conf | Next Img Conf

THE VILLAGES, Florida 3216

Wed MAIN	4/05 MAYS	ROC	Thu MAIN	4/06 MAYS	ROC	Fri MAIN	4/07 MAYS	ROC
ENDOCR CTR~								
ENDOCRINE FOLLO								

6:15 am → 7:53 am **Nonstop**
Orlando, FL, US (MCO) — Houston, TX, US (IAH - Intercontinental) — 2h 38m total
UA 1768 Boeing 737-900
Revise flight · Details · Seats

7:45 AM			
8:00 AM		DIAG CTR LAB~	
8:15 AM		LAB	
8:30 AM		CT~ |MAYS	
8:45 AM		CT CHEST WO CON	
9:00 AM		MAYS CT 15	
9:15 AM		85	
9:30 AM		--25--	
9:45 AM		--25--	
10:00 AM		PET CT~ |M	
10:15 AM		PET CT	
10:30 AM		MAYS PETCT A	
10:45 AM		150	
11:00 AM		150	
11:15 AM		150	
11:30 AM		150	
11:45 AM		150	
12:00 PM		--30--	
12:15 PM		--30--	
12:30 PM			
12:45 PM			
1:00 PM			
1:15 PM			
1:30 PM			
1:45 PM			
2:00 PM			
2:15 PM			
2:30 PM		NEURO US~	
2:45 PM		US THYROID NIR	
3:00 PM		MAYS NEURO US R	
3:15 PM		--60--	
3:30 PM		--60--	
3:45 PM			
4:00 PM			
4:15 PM			MRI~ |ROC
4:30 PM			MRI ABDOMEN W
4:45 PM			ROC MR 21
5:00 PM			--60--
5:15 PM			--60--
5:30 PM			--60--
5:45 PM			--60--
6:00 PM			MRI~ |ROC
6:15 PM			MRI PELVIS W WO
6:30 PM			ROC MR 21
6:45 PM			60
7:00 PM			
7:15 PM			

2. The Breakfast Klub
2514 reviews
$$ · Breakfast & Brunch, Southern
✓ Good for Breakfast
✓ Good for Brunch
Fourth Ward, Midtown
3711 Travis St
Houston, TX 77002
(713) 528-8561

4:57 pm → 6:34 pm **Nonstop**
Orlando, FL, US (MCO) — Houston, TX, US (IAH - Intercontinental) — 2h 37m total
UA 1527 Boeing 737-900
Revise flight · Details · Seats

4:06 pm → 7:26 pm **Nonstop**
Houston, TX, US (IAH - Intercontinental) — Orlando, FL, US (MCO) — 2h 20m total
UA 1644 Boeing 737-900
Revise flight · Details · Seats

SpringHill Suites Houston Medical Center/NRG Park
1400 Old Spanish Trail Houston, Texas 77054
5.2 miles from Houston
4.3/5
Houston hotel ideally located minutes away from Texas Medical Center, VA Hospital and NRG Park
Free high speed Internet · Free breakfast
Fitness center · Pool
SPRINGHILL SUITES MARRIOTT
From **166** USD/night
View Rates

SpringHill Suites Houston Medical Center/NRG Park
1400 Old Spanish Trail Houston, Texas 77054
5.2 miles from Houston
4.3/5
Houston hotel ideally located minutes away from Texas Medical Center, VA Hospital and NRG Park
Free high speed Internet · Free breakfast
Fitness center · Pool
SPRINGHILL SUITES MARRIOTT
From **166** USD/night
View Rates

Show PVD Schedule | Prior Patient | Next Patient | Next Conflict | Prv Img Conf | Next Img Conf

THE VILLAGES, Florida 32162 United S

APPTtime	Tue MAIN	4/04 MAYS	ROC	Wed MAIN	4/05 MAYS	ROC	Thu MAIN	4/06 MAYS	ROC	Fri MAIN	4/07 MAYS	ROC
		DIAG CTR LAB~										
		LAB										
		CT~ [MAYS										
		CT CHEST WO CON										
		MAYS CT 15										
		85										
		--25--										
		--25--										
10:00 AM		PET CT~ [M										
10:15 AM		PET CT										
10:30 AM		MAYS PETCT A										
10:45 AM		150										
11:00 AM		150										
11:15 AM		150		ENDOCR CTR~								
11:30 AM		150		ENDOCRINE FOLLO								
11:45 AM		150										
12:00 PM		--30--										
12:15 PM		--30--										
12:30 PM												
12:45 PM												
1:00 PM												
1:15 PM												
1:30 PM												
1:45 PM												
2:00 PM												
2:15 PM												
2:30 PM		NEURO US~										
2:45 PM		US THYROID NIR										
3:00 PM		MAYS NEURO US R										
3:15 PM		--60--										
3:30 PM		--60--										
3:45 PM												
4:00 PM												
4:15 PM			MRI~ [ROC									
4:30 PM			MRI ABDOMEN W									
4:45 PM			ROC MR 21									
5:00 PM			--60--									
5:15 PM			--60--									
5:30 PM			--60--									
5:45 PM			--60--									
6:00 PM			MRI~ [ROC									
6:15 PM			MRI PELVIS W WO									
6:30 PM			21									
6:45 PM												
7:00 PM												
7:15 PM												
7:30 PM												
7:45 PM												
8:00 PM												
8:15 PM												
8:30 PM												
8:45 PM												
9:00 PM												
9:15 PM												
9:30 PM												
9:45 PM												
10:00 PM												

6:15 am → 7:53 am Nonstop
Orlando, FL, US (MCO) — Houston, TX, US (IAH - Intercontinental) — 2h 38m total
UA 1768 Boeing 737-900
Revise flight | Details | Seats

8:00 am → 9:39 am Nonstop
Orlando, FL, US (MCO) — Houston, TX, US (IAH - Intercontinental) — 2h 39m total
UA 788 Boeing 737-900
Revise flight | Details | Seats

4:06 pm → 7:26 pm Nonstop
Houston, TX, US (IAH - Intercontinental) — Orlando, FL, US (MCO) — 2h 20m total
UA 1644 Boeing 737-900
Revise flight | Details | Seats

1. Oxheart
270 reviews
$$$$ · American (New)

I feel like I am so late to the game when it comes closes to revamp and I am so glad we did. Truth

SpringHill Suites Houston Medical Center/NRG Park
1400 Old Spanish Trail Houston, Texas 77054
5.2 miles from Houston
4.3/5

Houston hotel ideally located minutes away from Texas Medical Center, VA Hospital and NRG Park
Free high speed Internet | Free breakfast
Fitness center | Pool

From 166 USD/night
View Rates

ML: Imaging Protocols

Journal of Digital Imaging
https://doi.org/10.1007/s10278-018-0066-y

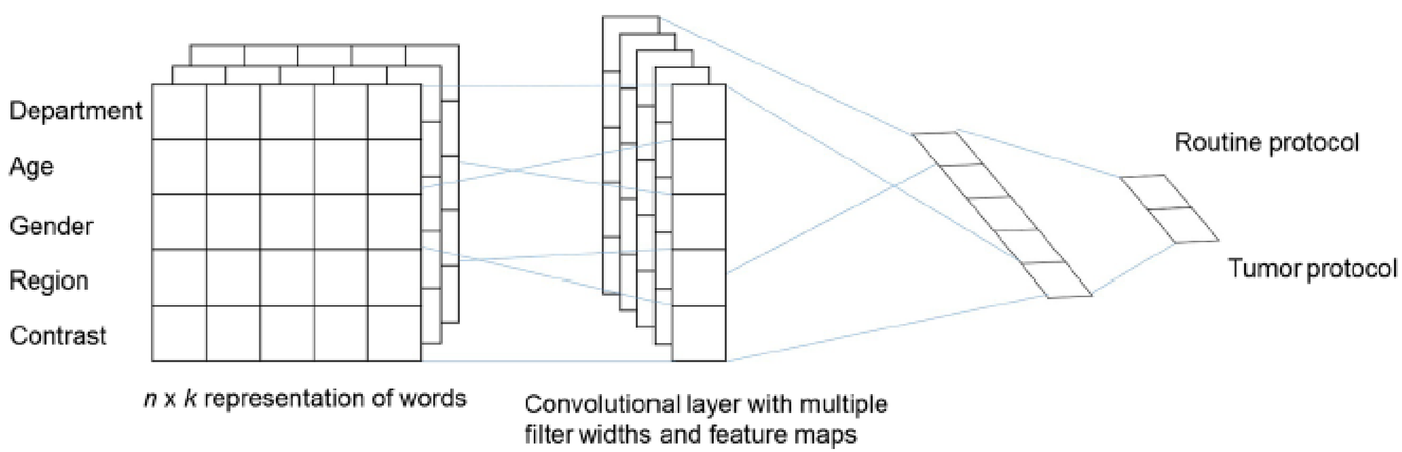
CrossMark

Efficiency Improvement in a Busy Radiology Practice: Determination of Musculoskeletal Magnetic Resonance Imaging Protocol Using Deep-Learning Convolutional Neural Networks

Young Han Lee[1]

Abstract
The purposes of this study are to evaluate the feasibility of protocol determination with a convolutional neural networks (CNN) classifier based on short-text classification and to evaluate the agreements by comparing protocols determined by CNN with those determined by musculoskeletal radiologists. Following institutional review board approval, the database of a hospital information system (HIS) was queried for lists of MRI examinations, referring department, patient age, and patient gender. These were exported to a local workstation for analyses: 5258 and 1018 consecutive musculoskeletal MRI examinations were used for the training and test datasets, respectively. The subjects for pre-processing were routine or tumor protocols and the contents were word combinations of the referring department, region, contrast media (or not), gender, and age. The CNN Embedded vector classifier was used with Word2Vec Google news vectors. The test set was tested with each classification model and results were output as routine or tumor protocols. The CNN determinations were evaluated using the receiver operating characteristic (ROC) curves. The accuracies were evaluated by a radiologist-confirmed protocol as the reference protocols. The optimal cut-off values for protocol determination between routine protocols and tumor protocols was 0.5067 with a sensitivity of 92.10%, a specificity of 95.76%, and an area under curve (AUC) of 0.977. The overall accuracy was 94.2% for the ConvNet model. All MRI protocols were correct in the pelvic bone, upper arm, wrist, and lower leg MRIs. Deep-learning-based convolutional neural networks were clinically utilized to determine

Fig. 2 Model architecture with two channels for the routine or tumor protocols of the musculoskeletal MRI



Structured reporting will necessarily evolve to discrete data reporting…

ORIGINAL ARTICLE

Evaluating Report Text Variation and Informativeness: Natural Language Processing of CT Chest Imaging for Pulmonary Embolism

Marco D. Huesch, MBBS, PhD, Rekha Cherian, MD, Sam Labib, MD, Rickhesvar Mahraj, MD

Abstract

Objective: The aim of this study was to quantify the variability of language in free text reports of pulmonary embolus (PE) studies and to gauge the informativeness of free text to predict PE diagnosis using machine learning as proxy for human understanding.

Materials and Methods: All 1,133 consecutive chest CTs with contrast studies performed under a PE protocol and ordered in the emergency department in 2016 were selected from our departmental electronic workflow system. We used commercial text-mining and predictive analytics software to parse and describe all report text and to generate a suite of machine learning rules that sought to predict the "gold standard" radiological diagnosis of PE.

Results: There was extensive variation in the length of Findings section and Impression section texts across the reports, only marginally associated with a positive PE diagnosis. A marked concentration of terms was found: for example, 20 words were used in the Findings section of 93% of the reports, and 896 of 2,296 distinct words were each used in only one report's Impression section. In the validation set, machine learning rules had perfect sensitivity but imperfect specificity, a low positive predictive value of 73%, and a misclassification rate of 3%.

Conclusion: Use of free text reporting was associated with extensive variability in report length and report terms used. Interpretation of the free text was a difficult machine learning task and suggests potential difficulty for human recipients in fully understanding such reports. These results support the prospective assessment of the impact of a fully structured report template with at least some mandatory discrete fields on ease of use of reports and their understanding.

Key Words: Structured reporting, text analysis, pulmonary embolus, machine learning, variability, prediction, natural language processing, NLP

J Am Coll Radiol 2018;15:554-562.

INTRODUCTION

It is accepted that structured templates represent the future of radiology reporting [1]. Despite early reports of no difference in information transfer efficiency [2], it is now clear that structured templates improve diagnostic yield within radiology [3] and are more user-friendly for clinical partners [4-6].

Beyond these benefits, structured reporting may enhance billing [7], lead to shorter reports that have been found easier for patients to understand [8], and be more useful for population health analytics and other research and operational data mining.

However, a large burden in productivity is placed on the

Department of Radiology, Milton S. Hershey Medical Center, Hershey,

Structured reporting will necessarily evolve to discrete data reporting…

ORIGINAL ARTICLE

CrossMark

Evaluating Report Text Variation and Informativeness: Natural Language Processing of CT Chest Imaging for Pulmonary Embolism

Results: There was extensive variation in the length of Findings section and Impression section texts across the reports, only marginally associated with a positive PE diagnosis. A marked concentration of terms was found: for example, 20 words were used in the Findings section of 93% of the reports, and 896 of 2,296 distinct words were each used in only one report's Impression section. In the validation set, machine learning rules had perfect sensitivity but imperfect specificity, a low positive predictive value of 73%, and a misclassification rate of 3%.

Conclusion: Use of free text reporting was associated with extensive variability in report length and report terms used. Interpretation of the free text was a difficult machine learning task and suggests potential difficulty for human recipients in fully understanding such reports. These results support the prospective assessment of the impact of a fully structured report template with at least some mandatory discrete fields on ease of use of reports and their understanding.

Key Words: Structured reporting, text analysis, pulmonary embolus, machine learning, variability, prediction, natural language processing, NLP

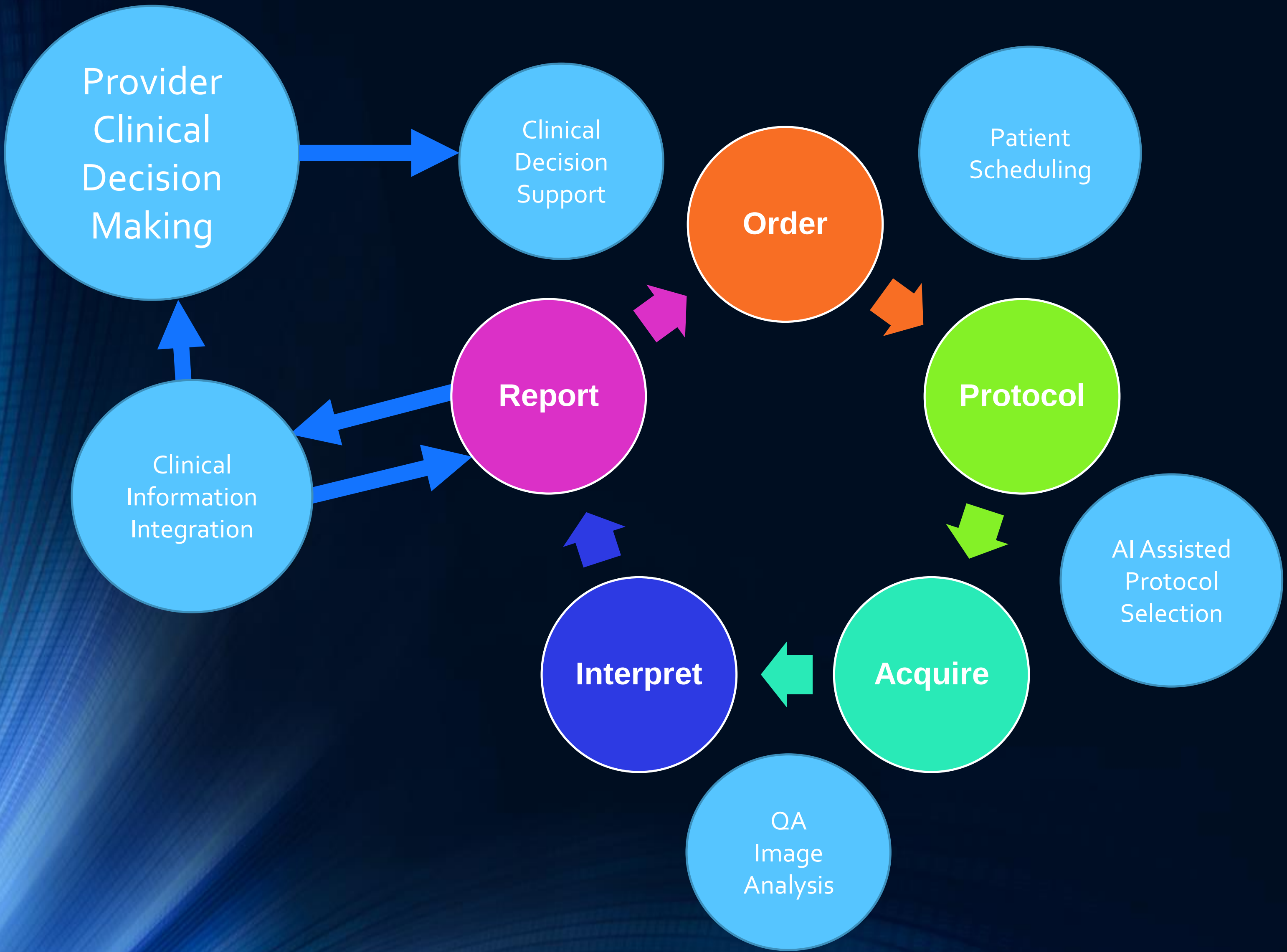
J Am Coll Radiol 2018;15:554-562.

INTRODUCTION

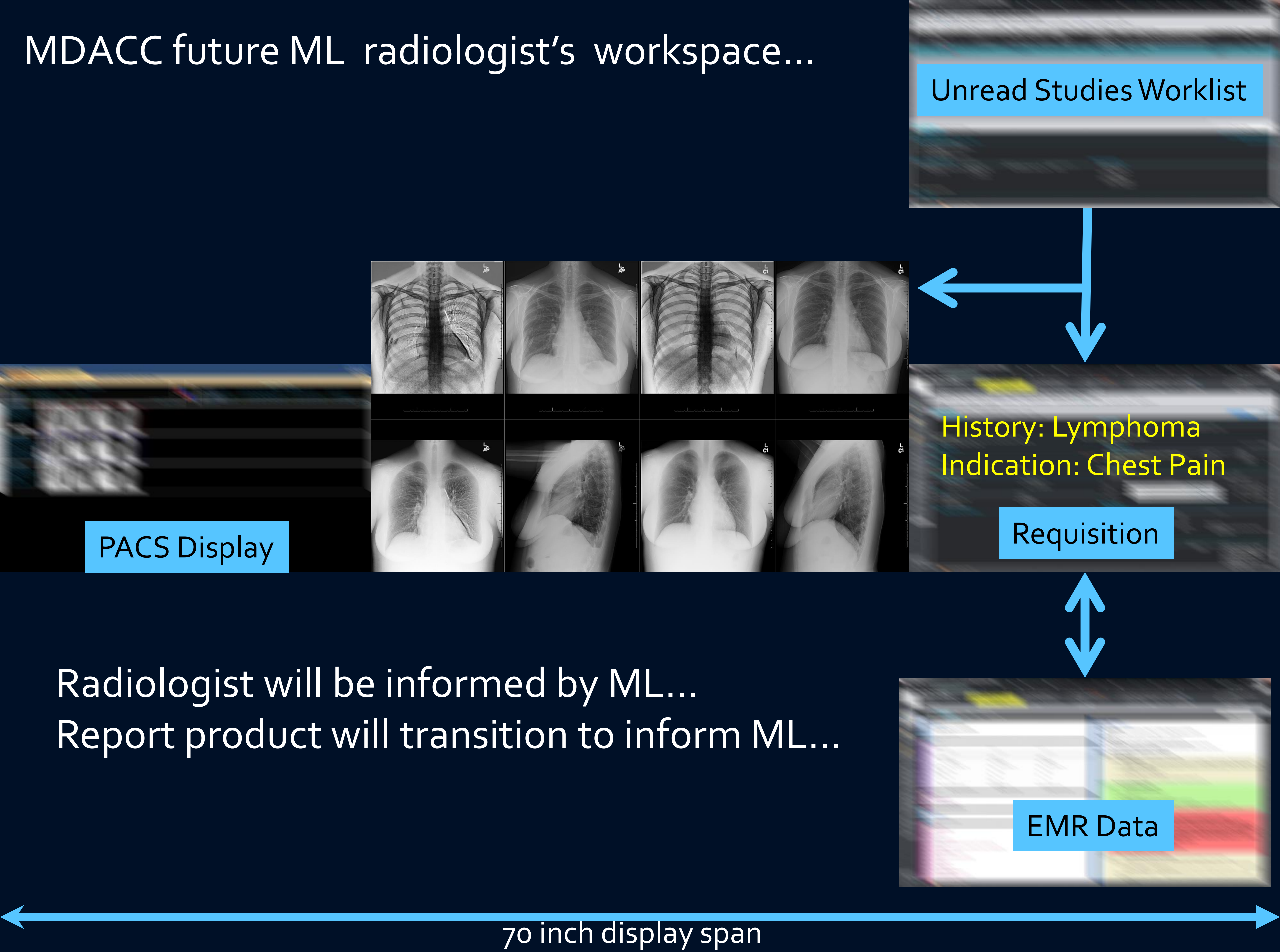
It is accepted that structured templates represent the future of radiology reporting [1]. Despite early reports of no difference in information transfer efficiency [2], it is now clear that structured templates improve diagnostic yield within radiology [3] and are more user-friendly for clinical partners [4-6].

Beyond these benefits, structured reporting may enhance billing [7], lead to shorter reports that have been found easier for patients to understand [8], and be more useful for population health analytics and other research and operational data mining

ML: Imaging Use Cases Beyond Interpretation



MDACC future ML radiologist's workspace…



Radiologist will be informed by ML…
Report product will transition to inform ML…

Why AI Will Not Replace Radiologists…

Radiologists will also be empowered to become more ‘doctor’ than ever before, with productivity gains allowing more time communicating results to both clinicians and patients. I can certainly envisage radiologists as data communicators, both directly to clinical teams on their rounds and tumour boards, and even direct-to-patient information-giving.

Why AI will not replace radiologists, Hugh Harvey

https://towardsdatascience.com/why-ai-will-not-replace-radiologists-c7736f2c7d80

Conclusions

- Initial reports of AI in imaging are promising
- AI and ML will have tremendous impact on imaging in the coming years.
- Application of AI will impact all processes within imaging, including the interpretation process.
  - Likely transition to increased quantitate reporting
  - Align reporting with data needed for algorithms
- Impacts in clinical decision support, scheduling, scanner operations, results delivery and review will be impacted with AI-based processes

Conclusions

Radiologists who refuse to incorporate AI into daily clinical practice will be replaced by radiologists who do…

although the generation of radiologists affected by ML transition TBD…