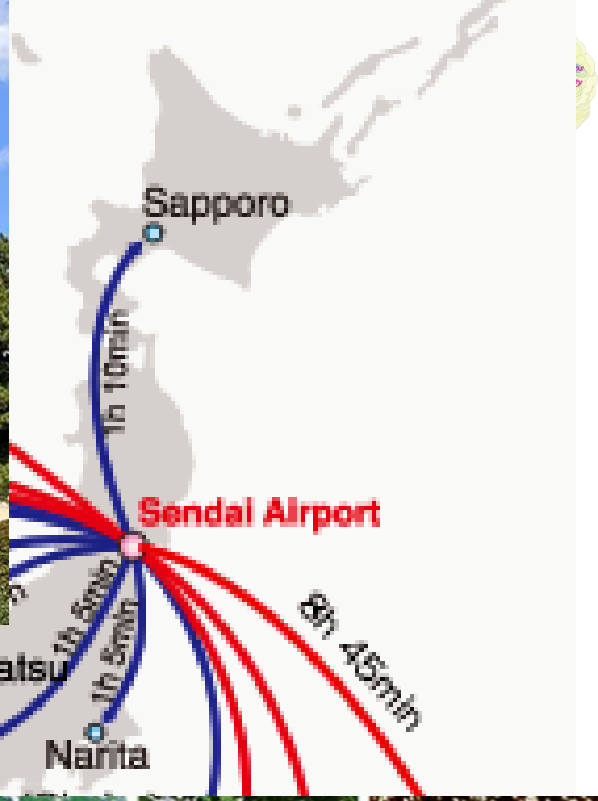
MacKay Int'l. Medical Education Conference

Impact of Artificial Intelligence on Healthcare and Society:
Insights from Japanese Medical Education

Noriyasu Homma
Tohoku University Graduate School of Medicine
homma@med.tohoku.ac.jp
http://www.rii.med.tohoku.ac.jp/

Disclosure

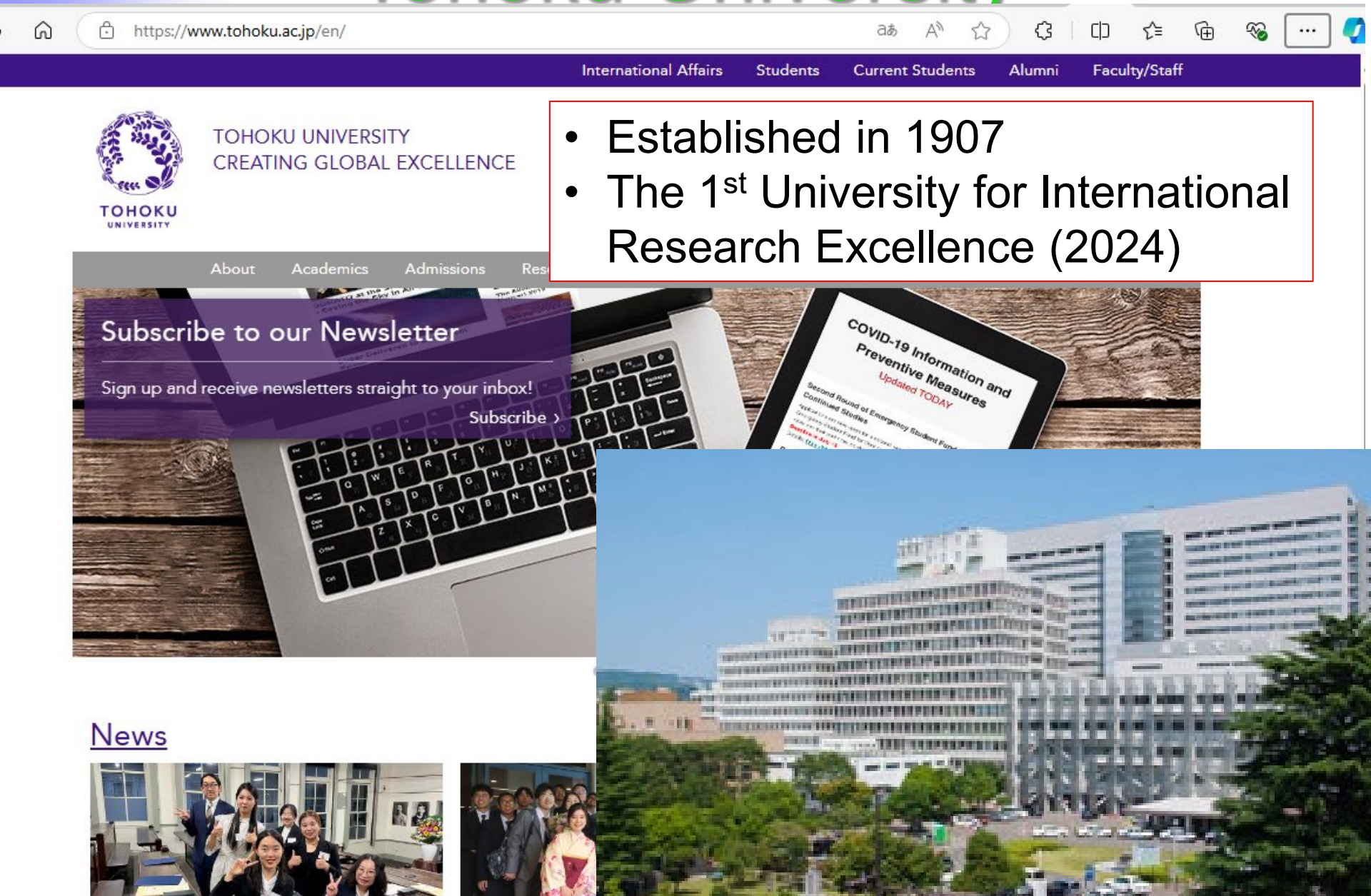
- The following studies were supported in part by research grants from Varian Medical Systems, Palo Alto, California and from RISE Corp., Japan.


Sapporo
1h 10min
Sendai Airport
Komatsu
1h 5min
1h 5min
Narita
8h 45min


2h 20min
Shanghai
Okinawa
Taipei


SENDAI S

Tohoku University



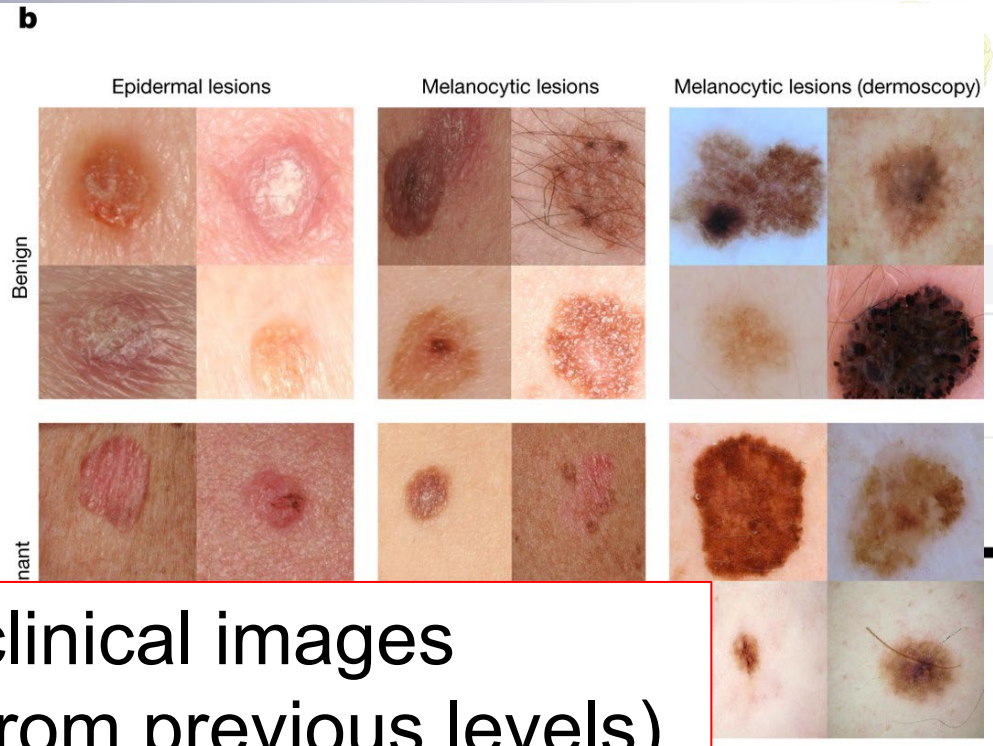
- Established in 1907
- The 1st University for International Research Excellence (2024)

Contents

- Impact: Medical image diagnosis and computer (AI) support
  - The Impact of deep learning: Radiologists will be fired?
  - Example of breast cancer
  - Will deep learning surpass human knowledge and skills: Will it be useful for medicine (in clinic)?

- Challenges and educational solutions

AI for medical



- Training with 130,000 clinical images (double-digit increase from previous levels)
- Achievement of performance comparable to that of specialists

ncer

Andre Esteva, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau & Sebastian Thrun

Nature **542**, 115–118(2017) | Cite this article

A Corrigendum to this article was published on 29 June 2017

Another example

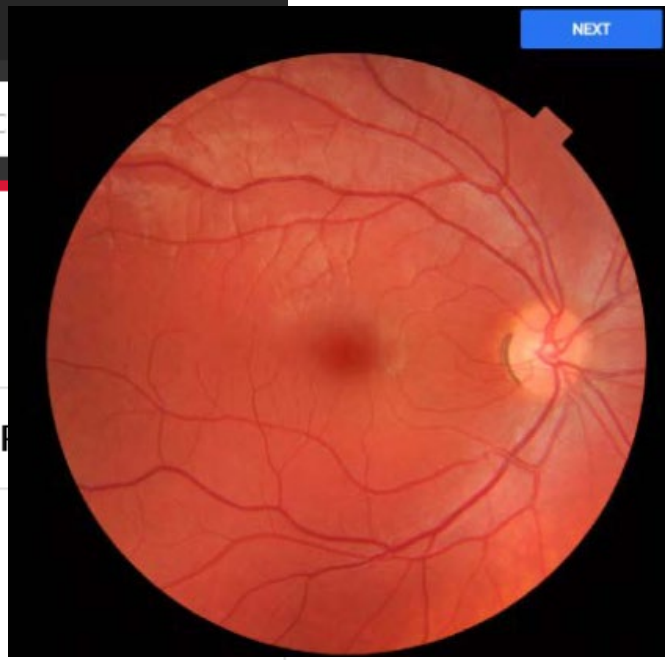
https://jamanetwork.com/journals/jama/fullarticle/2588763

JAMA Network

JAMA

Search All

Enter Searc



This Issue | Views 270,434 | Citations 1,446 | Altmetric 1080

Download PDF | More | Cite This

Original Investigation | Innovations in Health Care Delivery

December 13, 2016

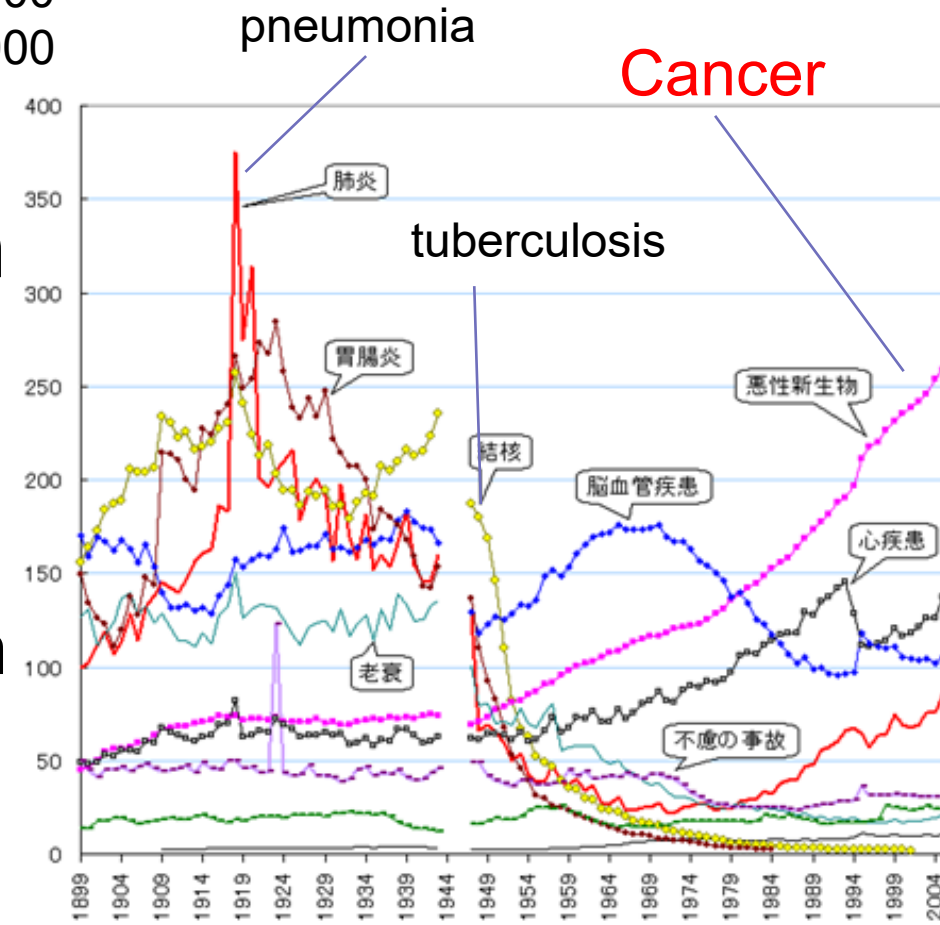
Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs

Varun Gulshan, PhD[1]; Lily Peng, MD, PhD[1]; Marc Coram, PhD[1]; et al

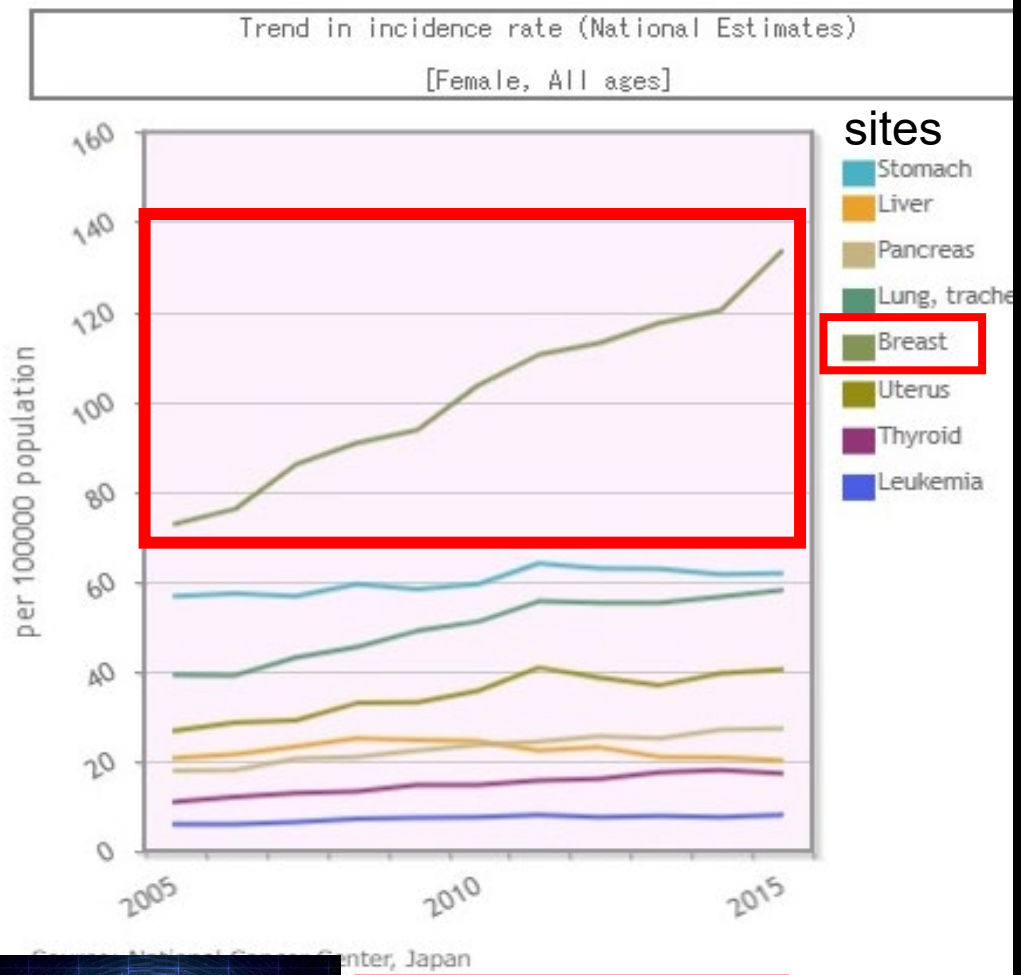
Cancer

1975: 210,000 patients
1985: 330,000
1995: 480,000
2005: 660,000 (320,000 died)
2015: 800,000
2020: 850,000

- Cancer is the worst cause of human death
  - A half of population gets cancer in Japan
  - 1/4 was died in Japan and Taiwan (more than 40 consecutive years)

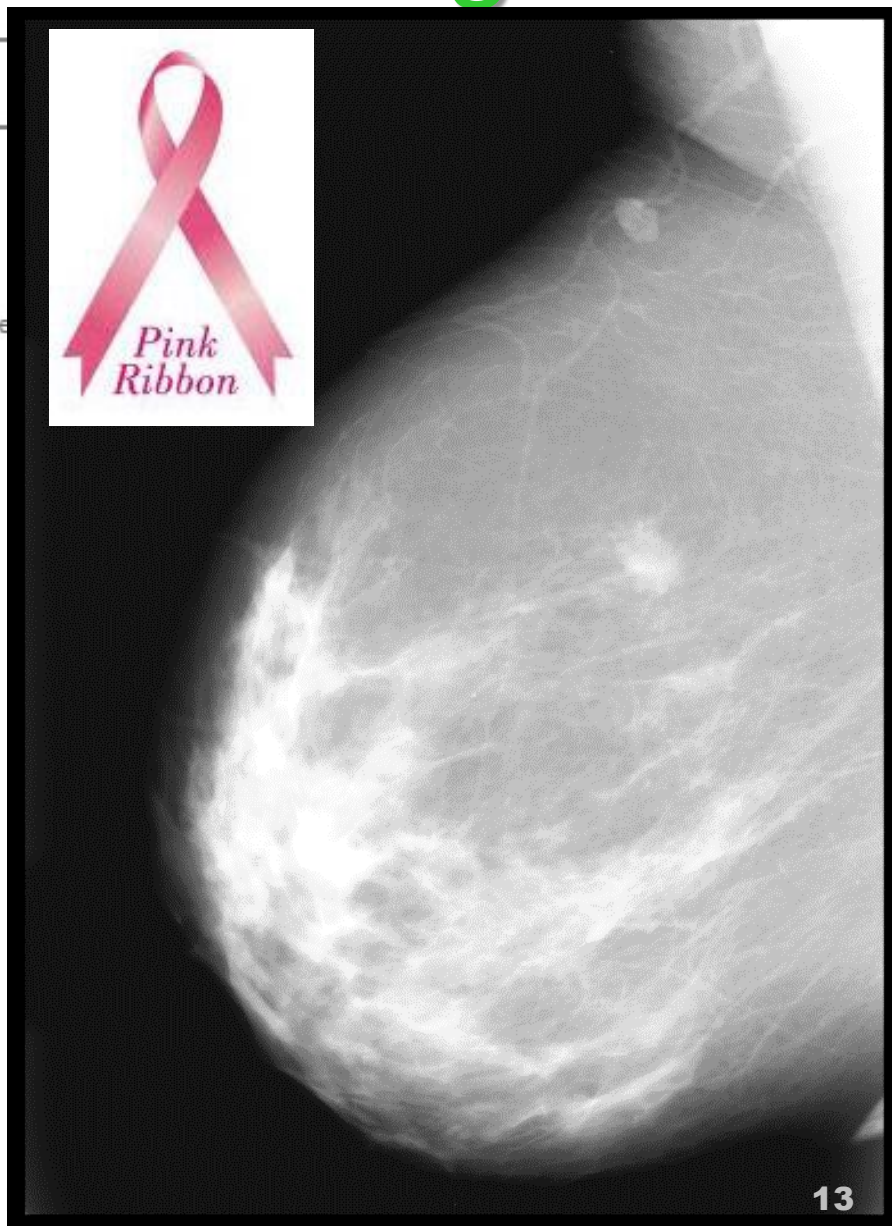


Mammographic screening



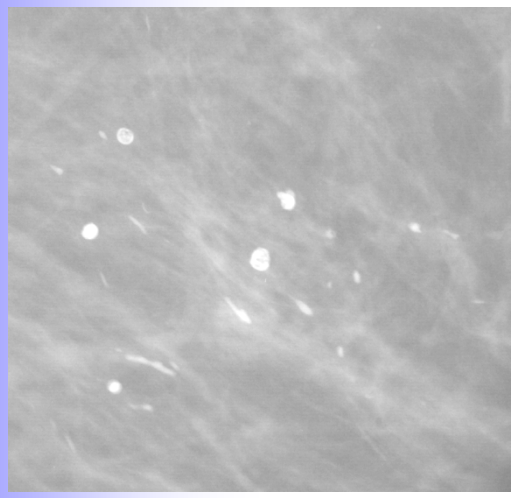
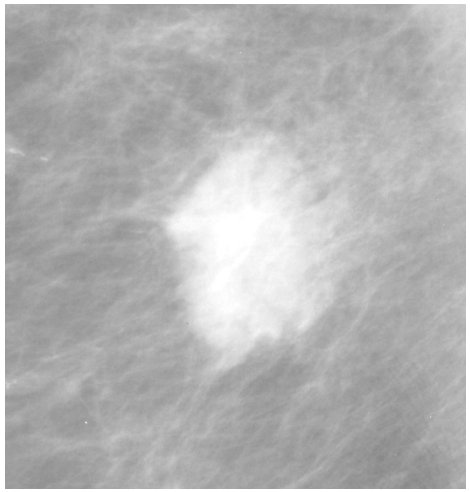
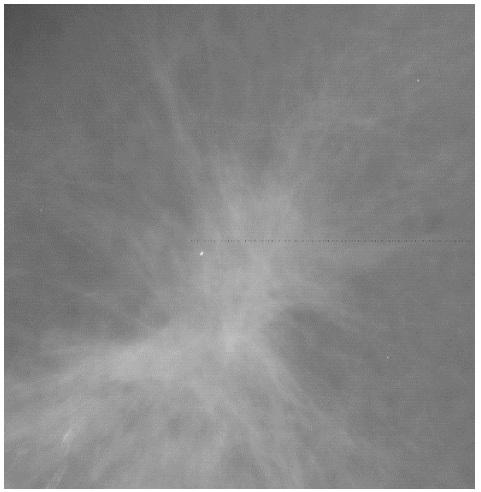
US Pat. 9,808,217

http://honvieew.com/technology/479



Typical findings in mammogram

Micro-calcification (MCC)　Mass　Architectural distortion (AD)

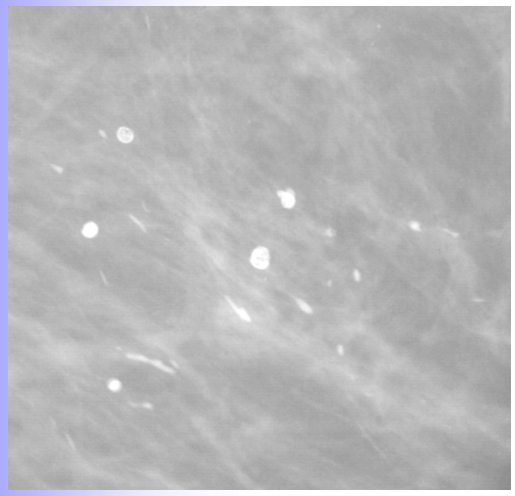




- Detection rate by medical interns [Hatanaka et al., 2002]

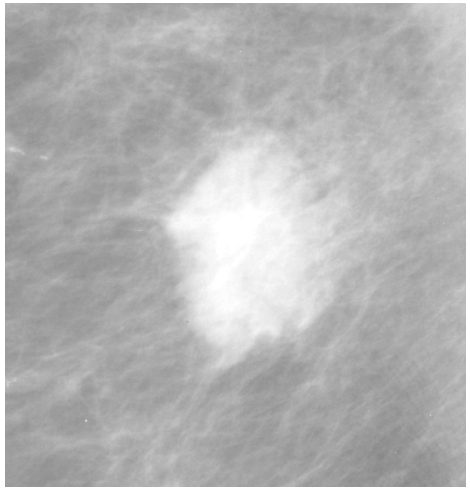
findings	MCC	Mass	AD
rate	94 %	85 %	60 %

Comparison: poor performance

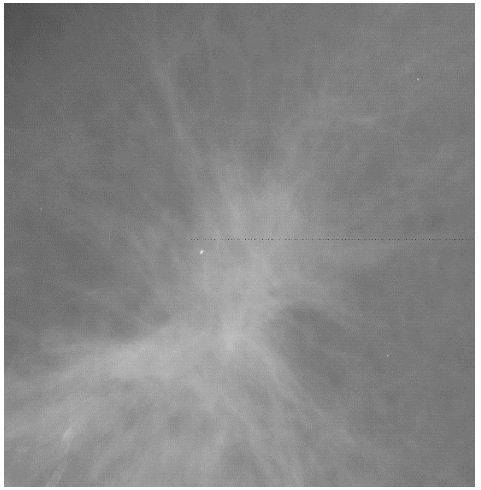
MCC



Mass



AD



findings	MCC	Mass	AD
Interns	94 %	85 %	60 %
Computer	95 %	~70 %	~50 %

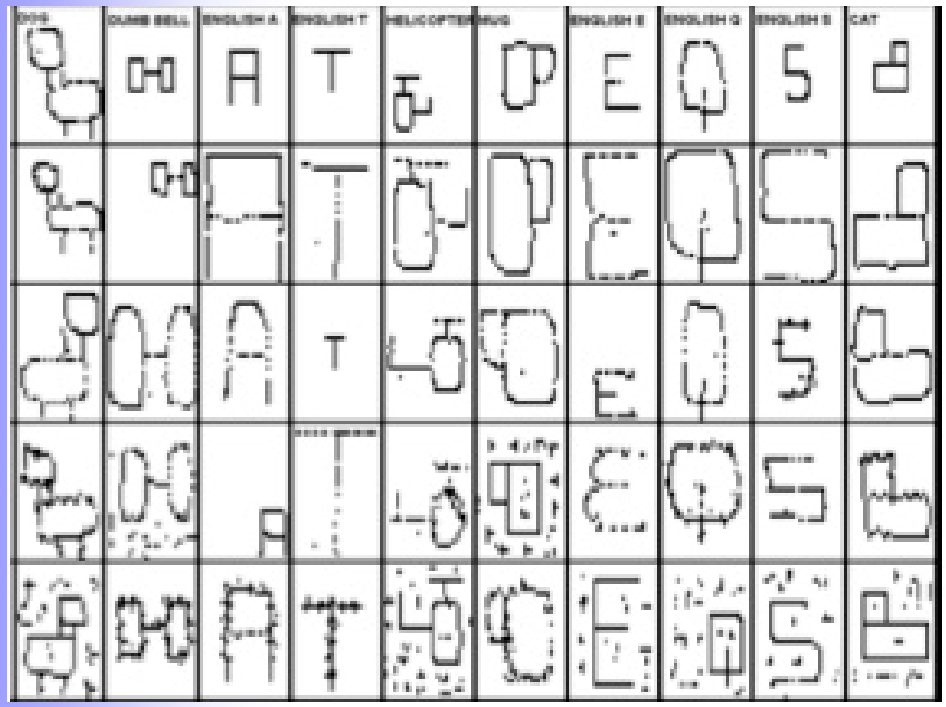
Before deep learning: Not good enough!!

What's the diagnostic mechanism of AI?

- Rough introduction to mathematical background

Pattern recognition

Feature extraction and classification



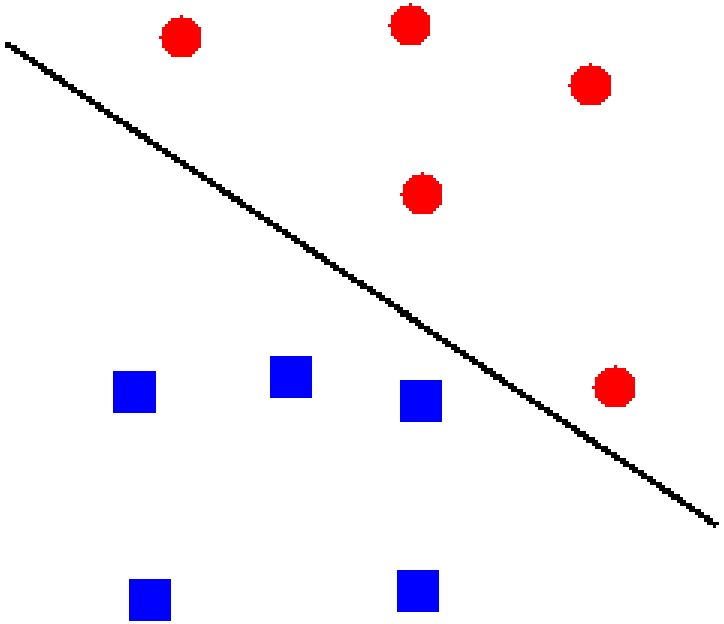


Image (pattern) recognition

Input

→ Feature extraction

- color, shape, size …
- manual design

→ classification

- thresholding, SVM
- manual design or machine learning

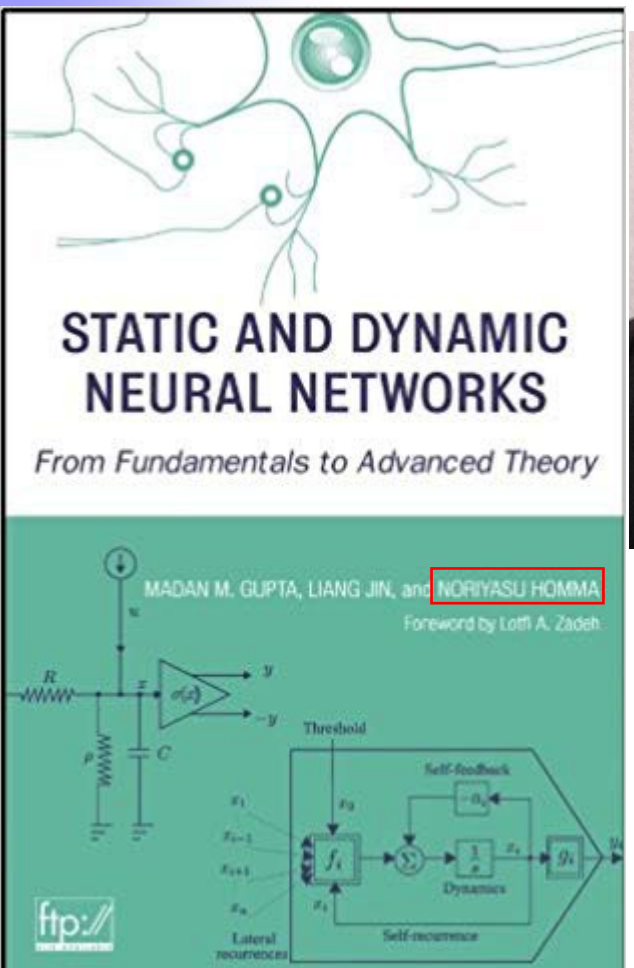
→ ‘ Cock ’
「ニワトリ」

Breakthrough

Deep learning

Congratulations to the pioneers

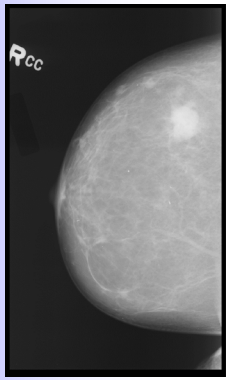
- **2018 Turing Award** **https://awards.acm.org/about/2018-turing**
  - Yoshua Bengio、Geoffrey Hinton、Yann LeCun



日経「AIのゴッドファーザー」雌伏の30年より

https://www.amazon.co.jp/Static-Dynamic-Neural-Networks-Fundamentals/dp/0471219487

Application: Pattern recognition Computer-aided diagnosis (CAD)



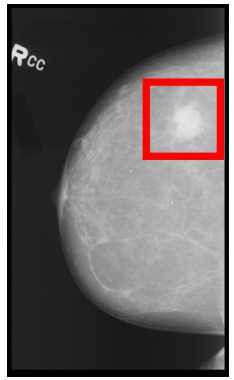
input

Feature extraction

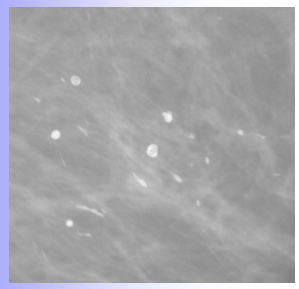
- color, shape, size …
- manual design

classification

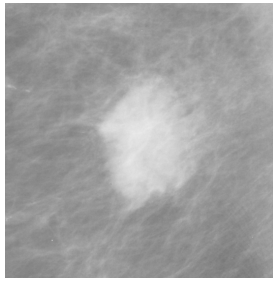
- thresholding, SVM
- manual design or machine learning



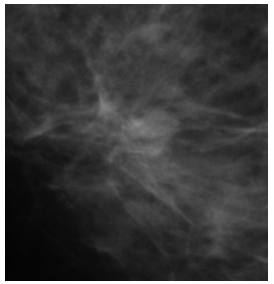
PCT/JP2013/8212



MCC
- tinny white spots
- clusters



Mass
- round shape
- concentric circle –like intensity

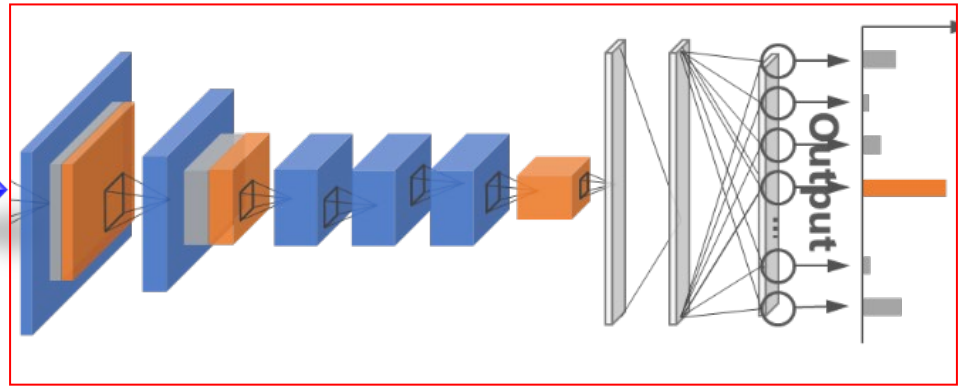


AD
- radial spicular
- unclear boundary

Breakthrough: Deep learning

Artificial Neural Networks

入力画像



‘ Cock ’

「ニワトリ」

- feature extraction can be done by training
- automatic design can achieve very high performance

TOHOKU UNIVERSITY

Results of mass and normal classification

Accuracy: **85.4**%

True Positive: **89.9**%

False Positive: **19.2**%

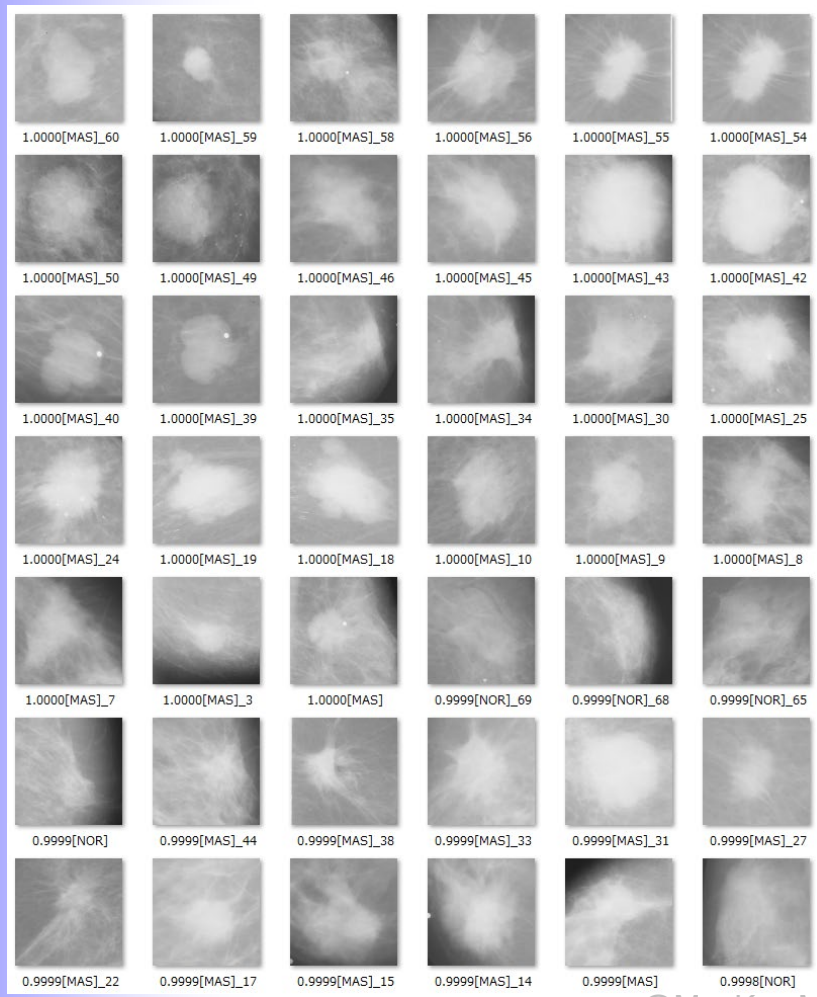
< 1.0 / image

～2015 (10 years ago)

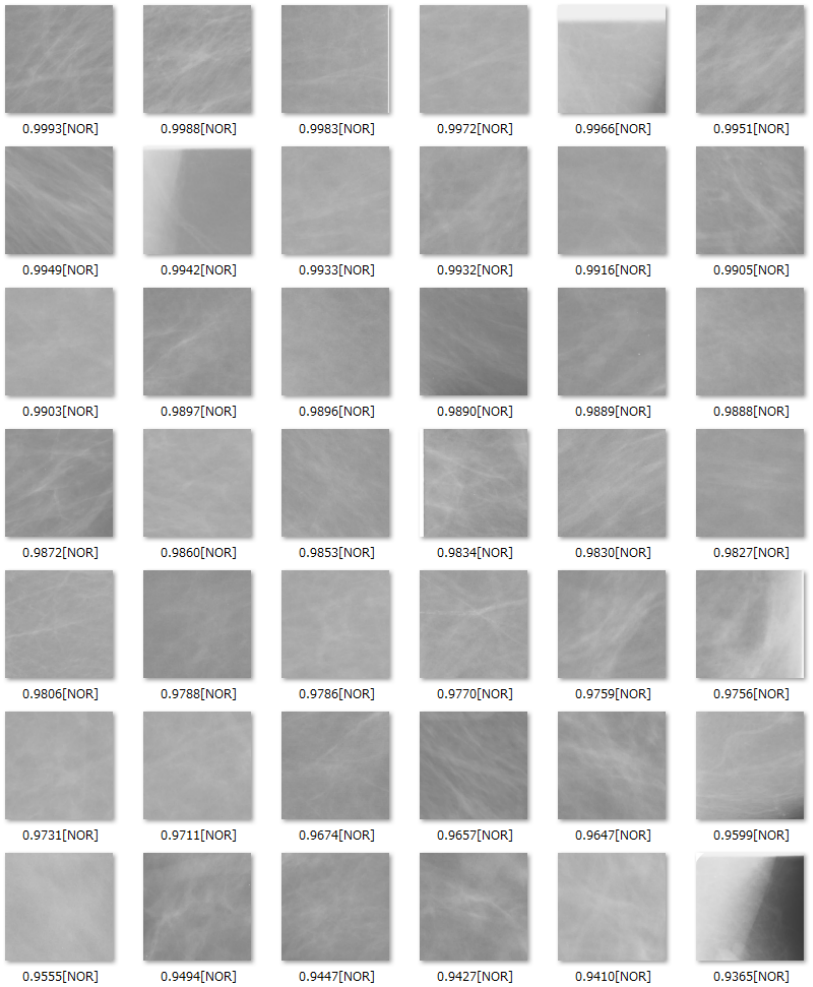
RESULTS

Examples

The MASS classification
with probability ranking



The NORMAL classification
with probability ranking



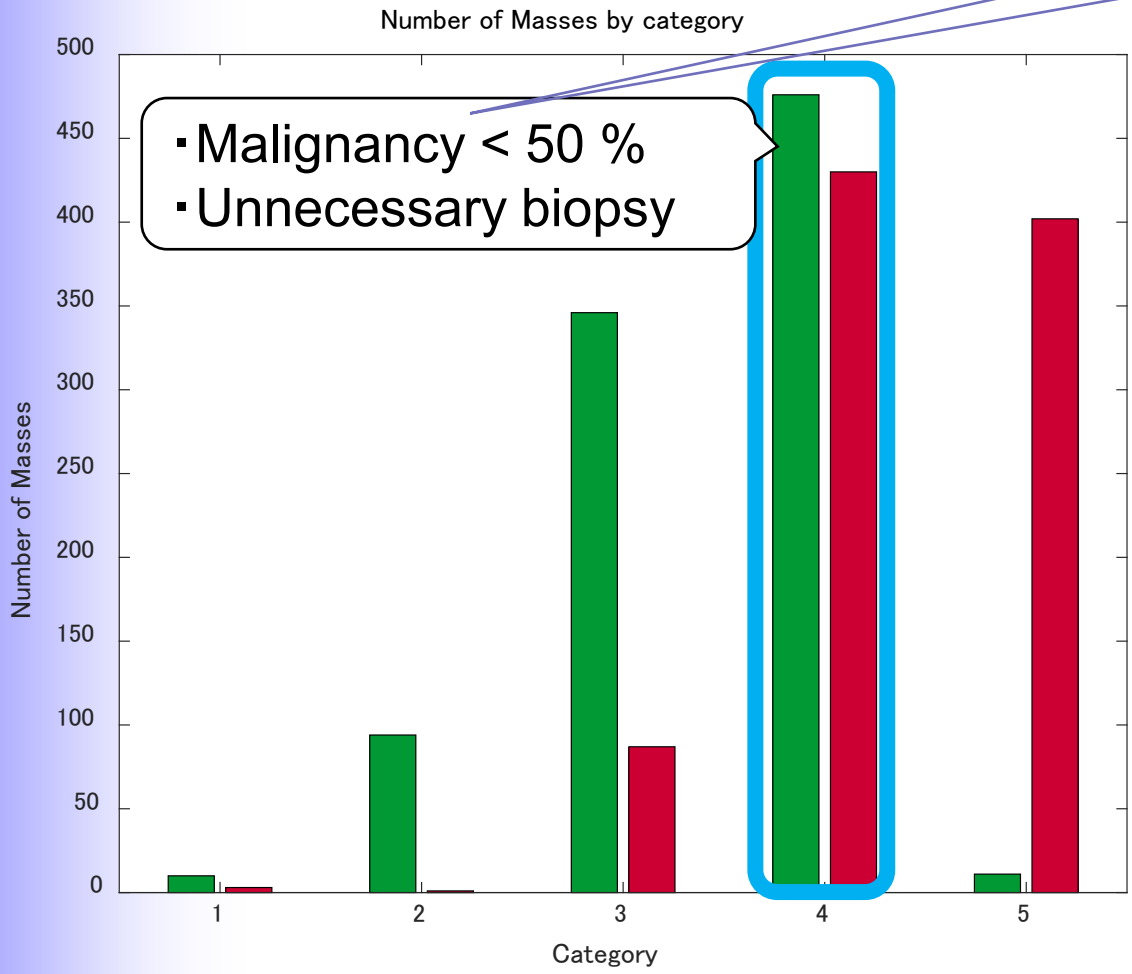
Will deep learning surpass human knowledge and skills?

- Is artificial intelligence the enemy of humanity?

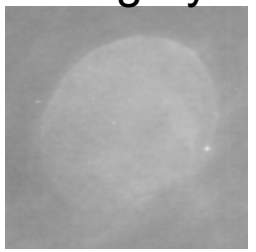
Difficult cases in mammography screening

AUC=0.79 (AI)

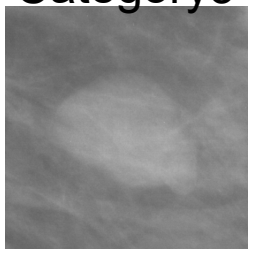
Noro et al. 2018



Category 2

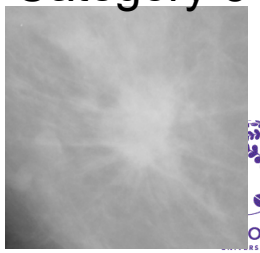


Category3



Category 4

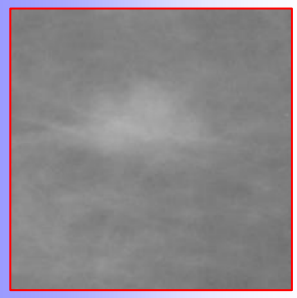
Category 5



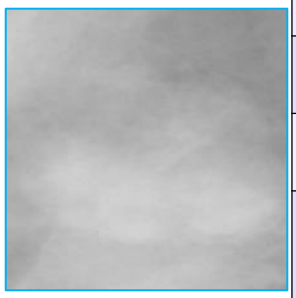
The Digital Database for Screening Mammography, Michael Heath et al. in Proceedings of the Fifth International Workshop on Digital Mammography, M.J. Yaffe, ed., 212-218, Medical Physics Publishing, 2001.

Result

- If the doctors and CAD predictions are the same
  → Doctor's confidence was increased

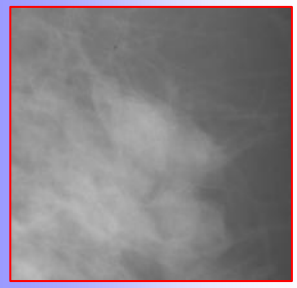


Reader	malignancy
Doctor	67%
CAD	100%
Doctor + CAD	**88%**



Reader	malignancy
Doctor	43%
CAD	0%
Doctor + CAD	**6.7%**

- If the doctors and CAD predictions are different
  → Doctors follow to CAD



Reader	malignancy
Doctor	29%
CAD	100%
Doctor + CAD	**57%**

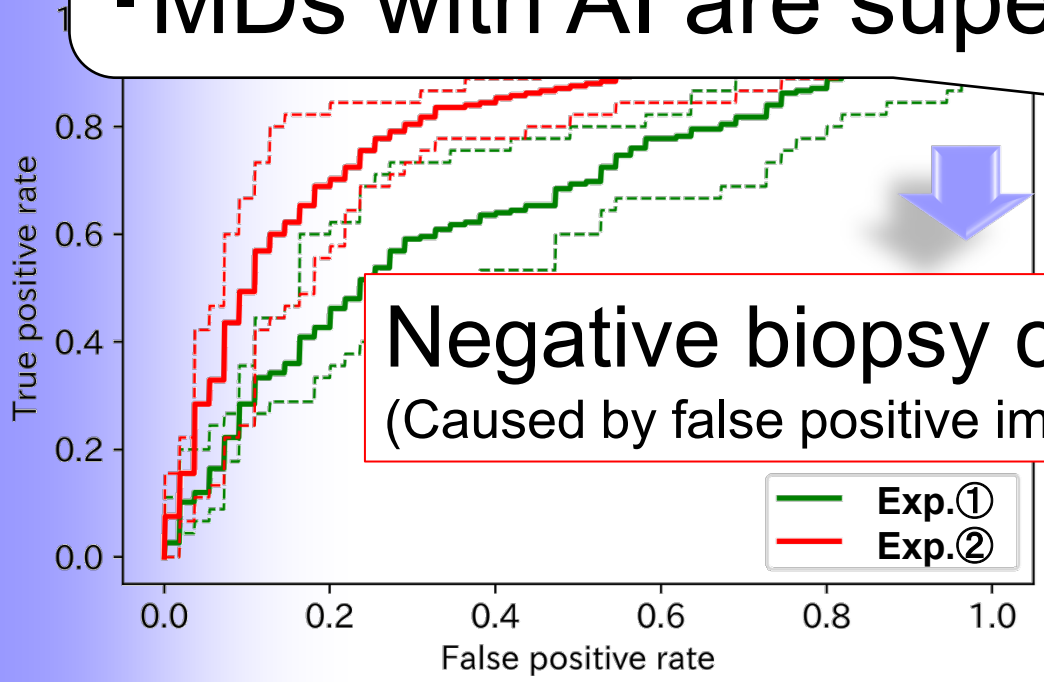
Reader	malignancy
Doctor	63%
CAD	0%
Doctor + CAD	**26%**

Result (2)

- Diagnosis ability was improved by using the CAD

CAD AUC = 0.79

- AI-assist is effective
- MDs with AI are superior to AI



		Doctor + CAD AUC
		0.74
B		**0.82** (> 0.79)
C	0.61	0.79
		(> 0.79)
		(> 0.79)

Negative biopsy can be reduced
(Caused by false positive image diagnosis)

t-test	
P value	Significant difference
6.9×10^{-4}	**Yes**

N. Homma et al. (2020)

Contents

- Impact: Medical image diagnosis and computer (AI) support

- Challenges and educational solutions for clinical applications
  - Differences from familiar techniques (for clinicians)
  - What patients want: accountability
  - Educational solution

What's wrong? - Reality

https://www.technologyreview.com/2020/04/27/1000658/google-medical-ai-accurate-lab-rea

MIT Technology Review

Topics Magaz

WIKIMEDIA COMMONS

Artificial intelligence / Machine learning

Google's medical AI was super accurate in a lab. Real life was a different story.

If AI is really going to make a difference to patients we need to know how it works when real humans get their hands on it, in real situations.

Did you know?

bmj.com/content/368/bmj.m689

thebmj covid-19 Research Education News & Views Campaigns Jobs

Research

Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies

BMJ 2020 ; 368 doi: https://doi.org/10.1136/bmj.m689 (Published 25 March 2020)
Cite this as: *BMJ* 2020;368:m689

Linked Editorial
Artificial intelligence versus clinicians

Article | Related content | Metrics | Responses | Peer review

COVID-19 pandemic

ARTICLES
https://doi.org/10.1038/s42256-021-00338-7
nature machine intelligence

Check for updates

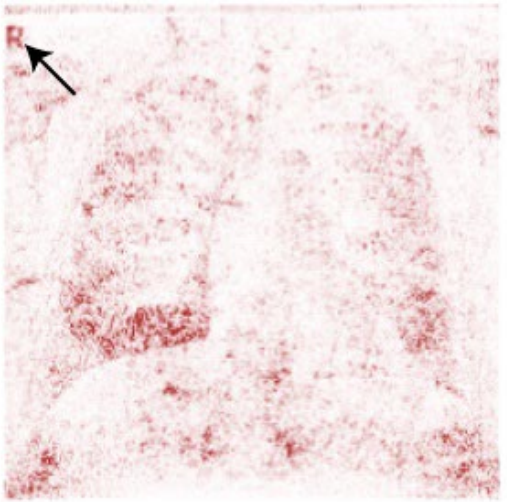
AI for radiographic COVID-19 detection selects shortcuts over signal

Alex J. DeGrave[1,2,3], Joseph D

Artificial intelligence (AI) researchers chest radiographs. However, the robus AI, we demonstrate that recent deep l rather than medical pathology, creati new hospitals. We observe that the ap AI to learn these spurious 'shortcuts'. the detection of COVID-19 in compute reveals a far-reaching problem in medi ficient to ensure AI systems rely on me not impair performance in new hospit clinical deployment of machine-learni

Negative image

Important pixels

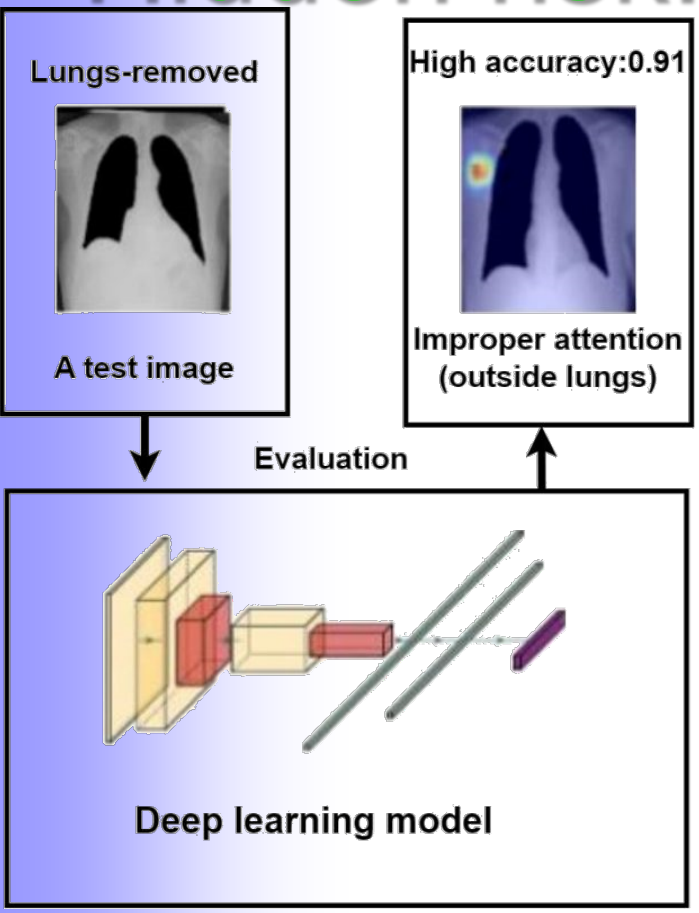


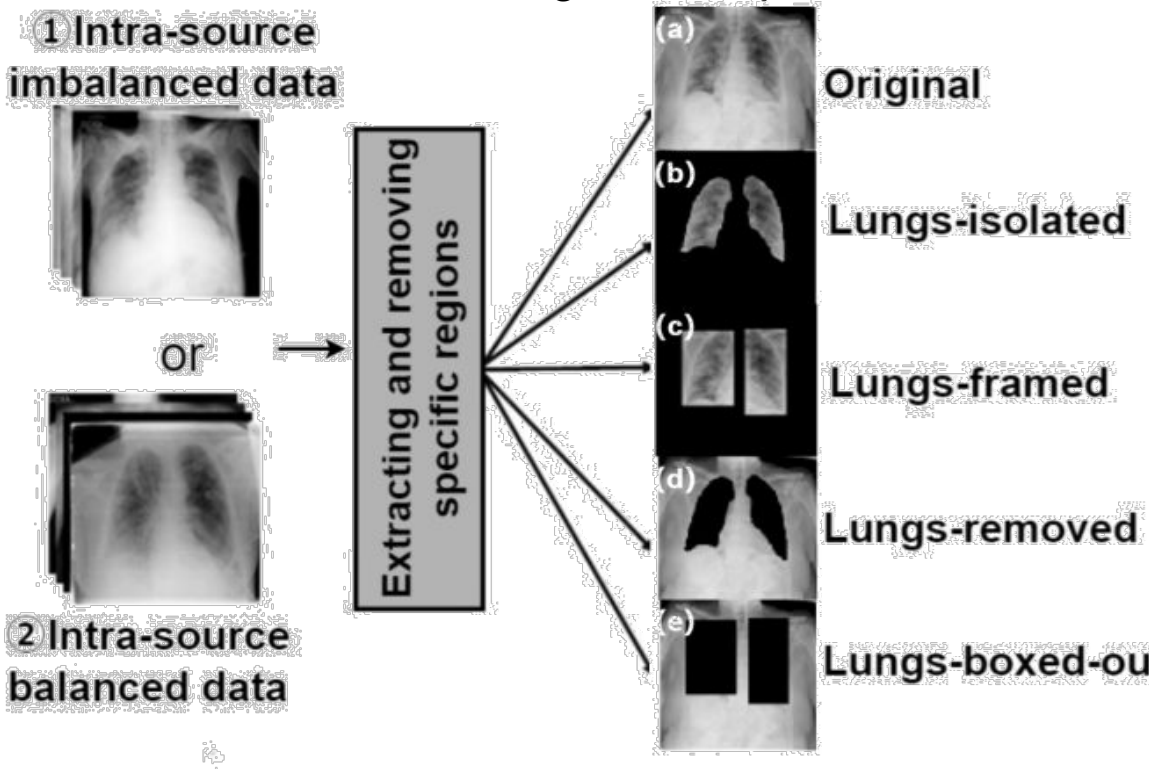
The prospect of applying artificial detection of COVID-19 in chest radiographs has generated interest from machine learning (ML) researchers and radiologists alike, given its potential to (1) help guide management in resource-limited settings that lack sufficient numbers of the

image features using state-of-the-art explainable artificial intelligence (AI) techniques, including both saliency maps and generative adversarial networks (GANs)[16–19]. These enquiries reveal how seemingly high-performance AI systems may derive the majority of their

Hidden risk: Reliable?

Zhang, Sci. Rep.（2023）



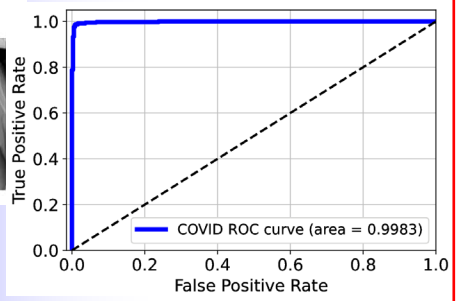


Deep learning model accurately diagnosed COVID-19 infection even in images where the medically significant region was removed.

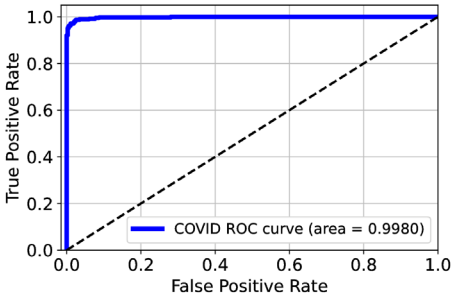
Image processing techniques to extract or remove specific regions such as the lungs from the original images. We introduced two new datasets: one with intra-source balance and the other with intra-source imbalance that can be a hidden risk of AI training.

ROC curves: Qata-COV19(Intra-source imbalanced)

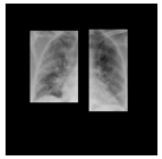
Original CXR

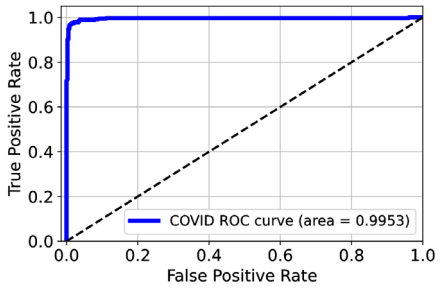


Lung boxed out

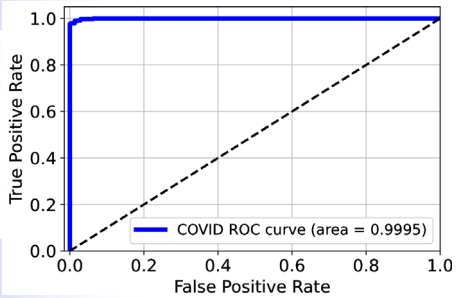


Lung framed

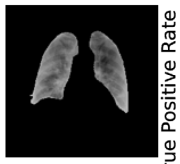


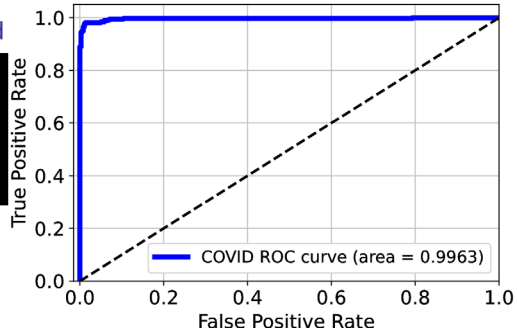


Lung removed



Lung isolated





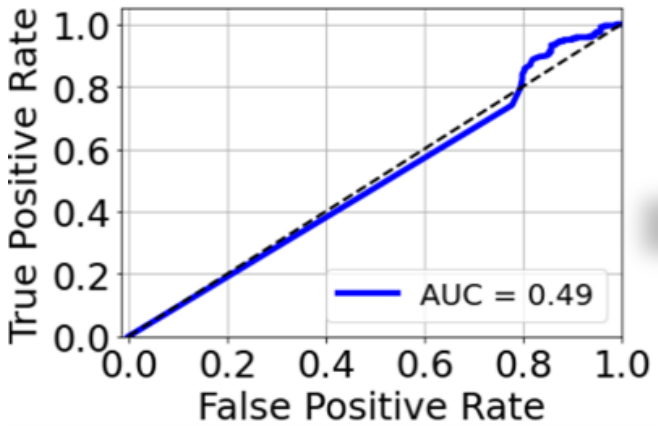
- High but unreliable performance

Test Performance	AUC
Original	0.9983
Lung boxed out	0.9980
Lung framed	0.9953
Lung removed	0.9995
Lung isolated	0.9963

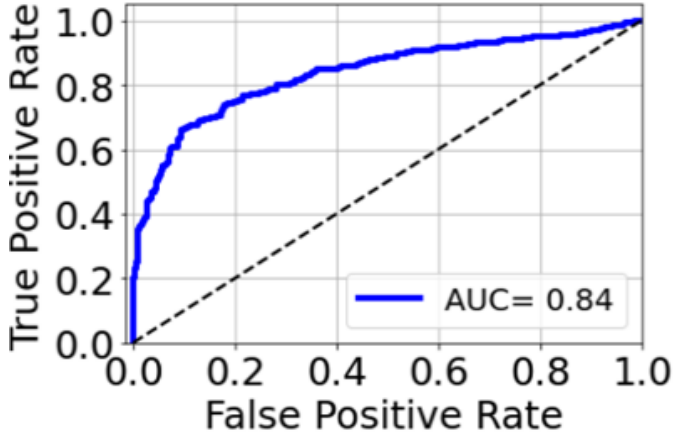
@MacKay Memorial Hospital in Taipei

Improvement: data quality

- Reveal clear evidence that
  - Intra-source balance is vital for training data
  - Deep learning model trained by intra-source imbalanced data might classify images based on the features characterizing data sources



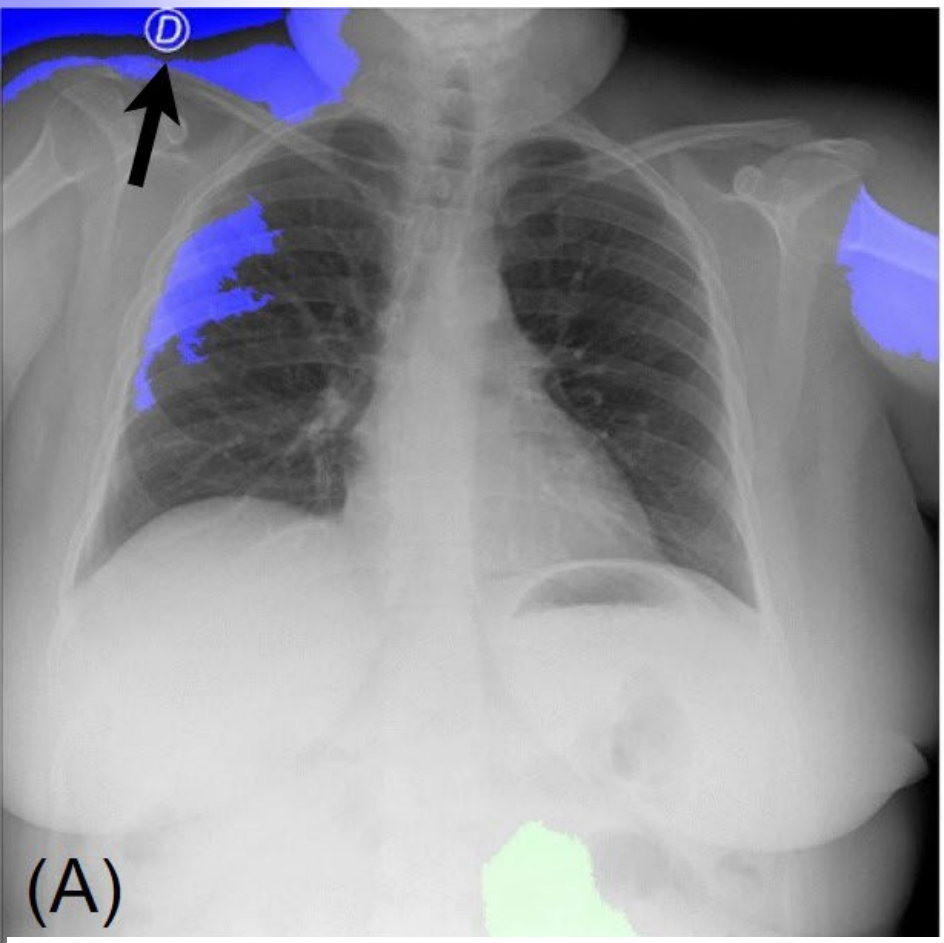
Model trained by intra-source imbalanced data in cross-dataset test

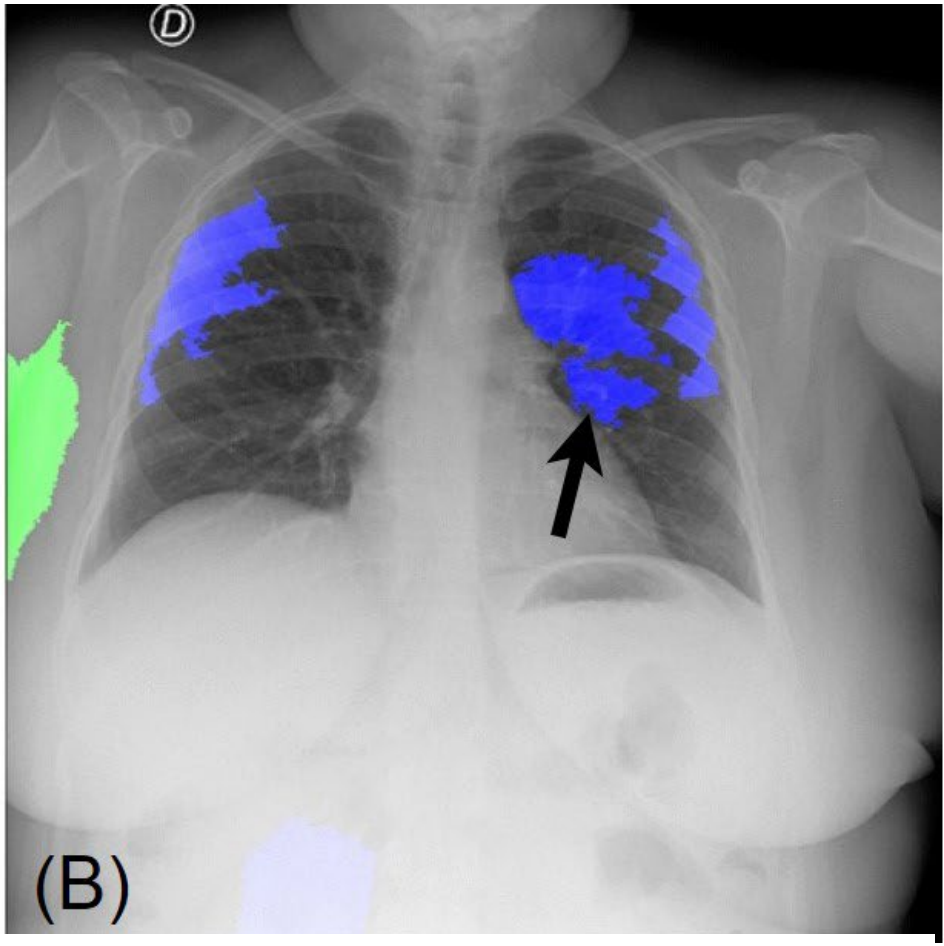


Model trained by intra-source balanced data in cross-dataset test

- Raise concerns about intra-source imbalance

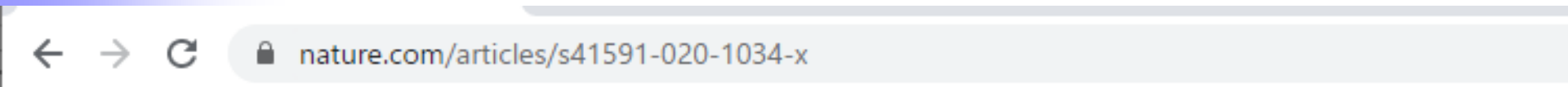
Where did AI focus on?





The regions focused by the two models in diagnosing COVID-19 infection: (A) The model trained on the intrinsically imbalanced dataset is directing attention to text 'D' (improper area). (B) The model trained on the intrinsically balanced dataset is appropriately focusing on the medically relevant lung region, demonstrating higher reliability in the basis of diagnosis.

Guidelines for clinical AI-trials



nature medicine

Consensus Statement | Open Access | Published: 09 September 2020

Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension

Xiaoxuan Liu, Samantha Cruz Rivera, David Moher, Melanie J. Calvert, Alastair K. Denniston ✉ & The SPIRIT-AI and CONSORT-AI Working Group

Patient apprehensions*

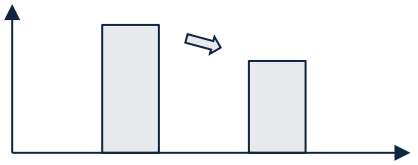
Accountability

1. Participants were excited about healthcare AI but wanted assurances about safety
2. Patients expect their clinicians to ensure AI safety
3. Preservation of patient choice and autonomy
4. Concerns about healthcare costs and insurance coverage
5. Ensuring data integrity
6. Risks of technology-dependent systems

*J. Richardson et al., Patient apprehensions about the use of artificial intelligence in healthcare, *digital medicine*, 140 (Sept. 2021), https://www.nature.com/articles/s41746-021-00509-1

Why AI in Medicine

The Essential Requirements for AI in Medicine



* Increased work speed / Reduced labor effort
  - might be necessary but not enough
  - can be adequately achieved with existing methods without AI

(by reviewing of rules, changing mindset, and improving culture)

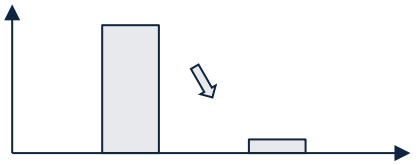
My experience

- required no technique
- delivered dramatic results

However, these initiatives are often not carried out. Why?

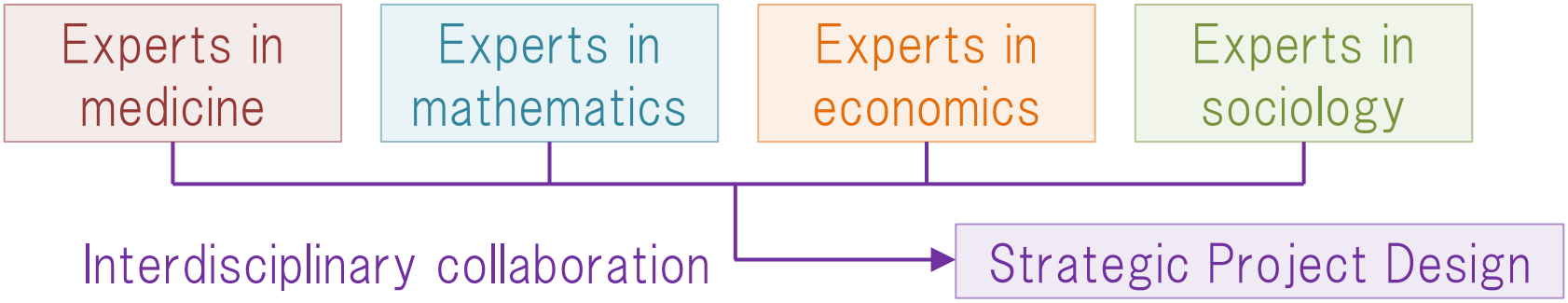
People do not accept changes in values and behavior.

* Encouraging a reassessment of existing values
  - bring about behavioral changes in people



For Creating Innovative AI Solutions

Aggregation of Insights from Various Experts

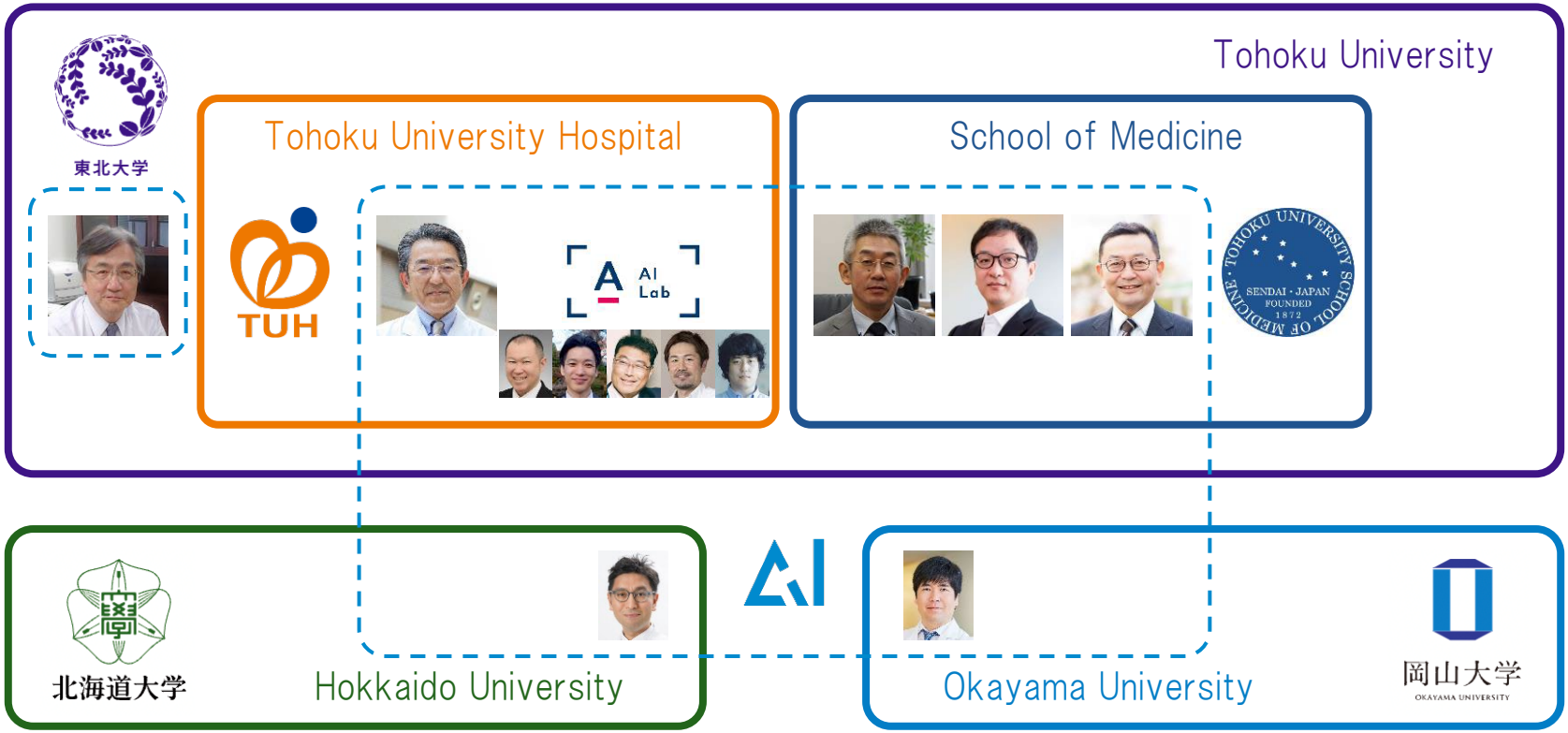


Team Building

* Mere talent pool ⟶✕ Innovation
* For seamless exchange and creation of insights
  - common languages
  - shared culture
  - leaders with understanding in each respective field

AI Lab / Clinical AI

AI Lab for Innovation, Clinical AI for Education

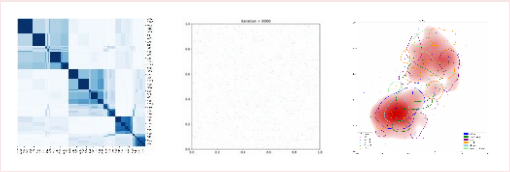


AI Lab – Center for Medical AI Projects

Leading for Joint Research and Development

＊ Support Mathematics and Design Across Departments

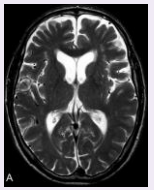
AI Lab
- Refine research design
- Obtain grant funding
- Introduce new technology
- Accumulate know-how
- Establish core algorithm



Technologies / Ideas →

← Needs / Data

Department A
- Survey of clinical needs
- Acquisition of data
- Procedure of documents
- Recruitment of human resources



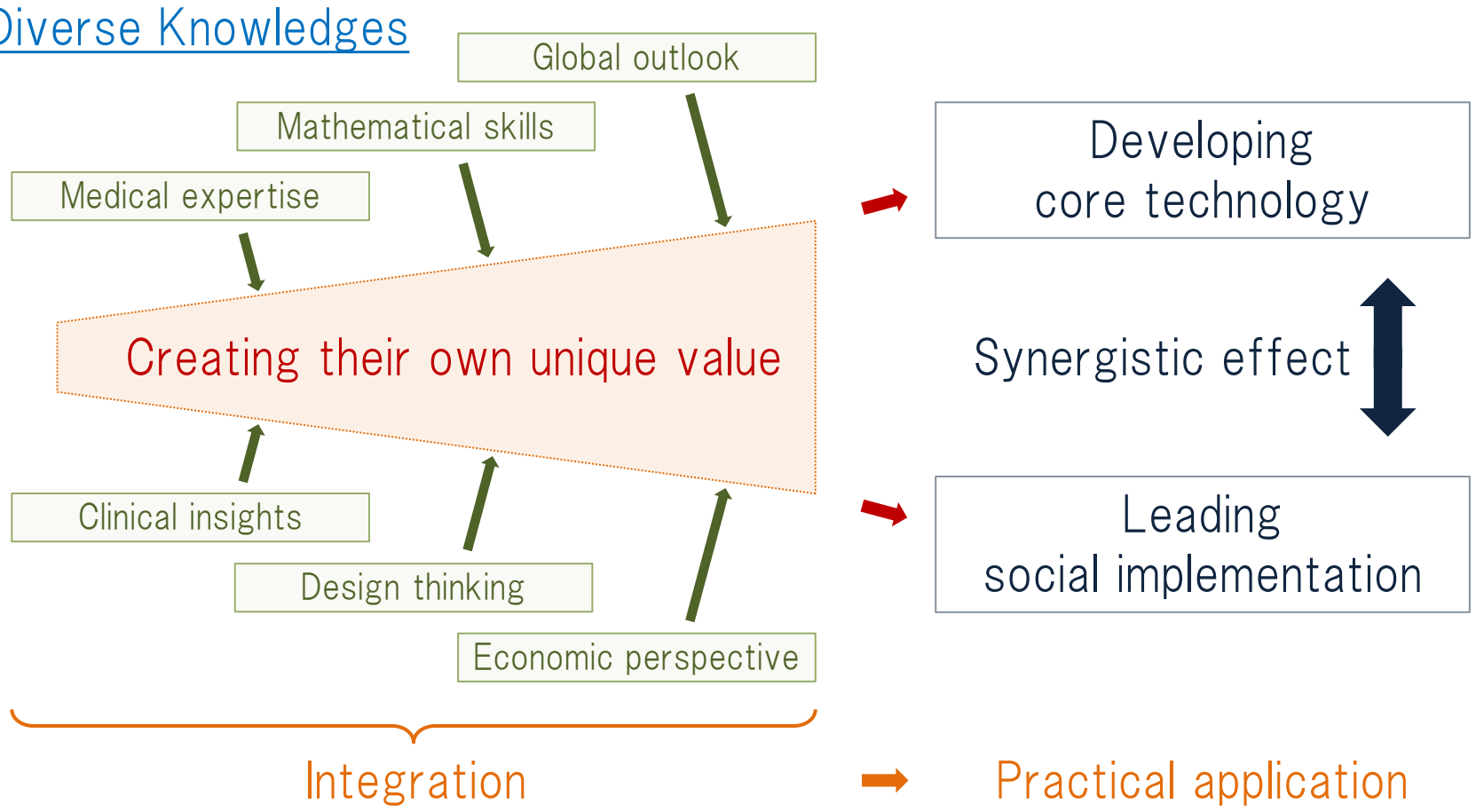
Assist →

Deriving ⇩ results

Social implementation
- Resolving clinical problems
- Medical digital transformation

Department B | Department C | Department D

Clinical AI – Ideal Profile of Trainees



Number of students

As of July 1st, 2024

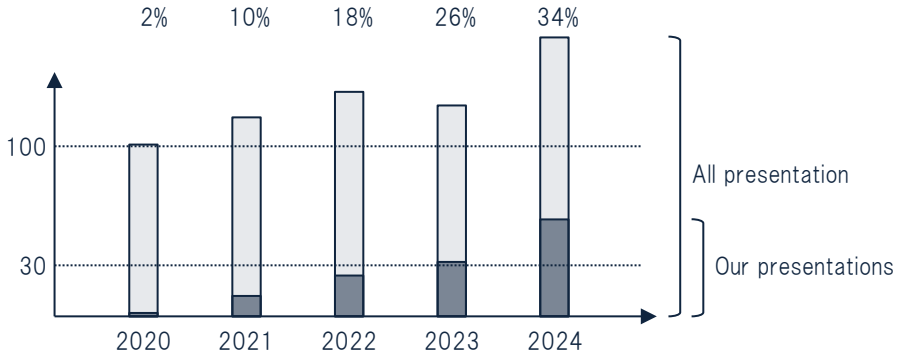
type	university	First batch (Enrolled in 2023)		Second batch (Enrolled in 2022)		Third batch (Enrolled in 2023)		Forth batch (Enrolled in 2024)	
		Reg.	KPI	Reg.	KPI	Reg.	KPI	Reg.	KPI
Regular (4 years)	Tohoku	9	5	7	5	5	5	6	5
	Hokkaido	8	2	17	4	14	4	13	4
	Okayama	7	3	16	3	17	3	12	3
Intensive (1 year)	Tohoku	55	7	223	7	560	7	343	7
	Hokkaido	33	6	100	6	210	6	227	6
	Okayama	54	6	210	6	290	6	197	6

Clinical AI – Activities of Trainees

Academic Output

Until June 2023

Projects	152
Paper submission	28
Research funding	16



Number of presentations at the
Annual Meeting of Japanese Association for Medical AI

Career Path

At university
- advanced research and development
- medical data management
- education

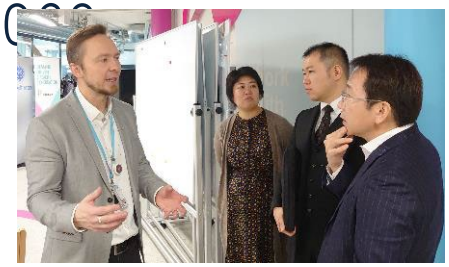
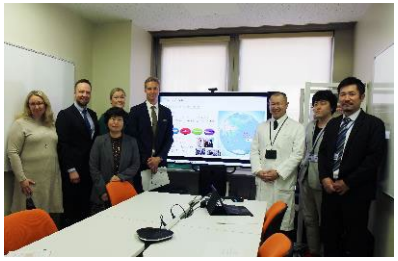
Move to
- other research institutes
- startup companies
- administrative agencies

The next generation of medical digital transformation

AI Lab / Clinical AI – International Exchange

Mutual Site Visit

＊ Finland (Oulu) – 2019, 2023

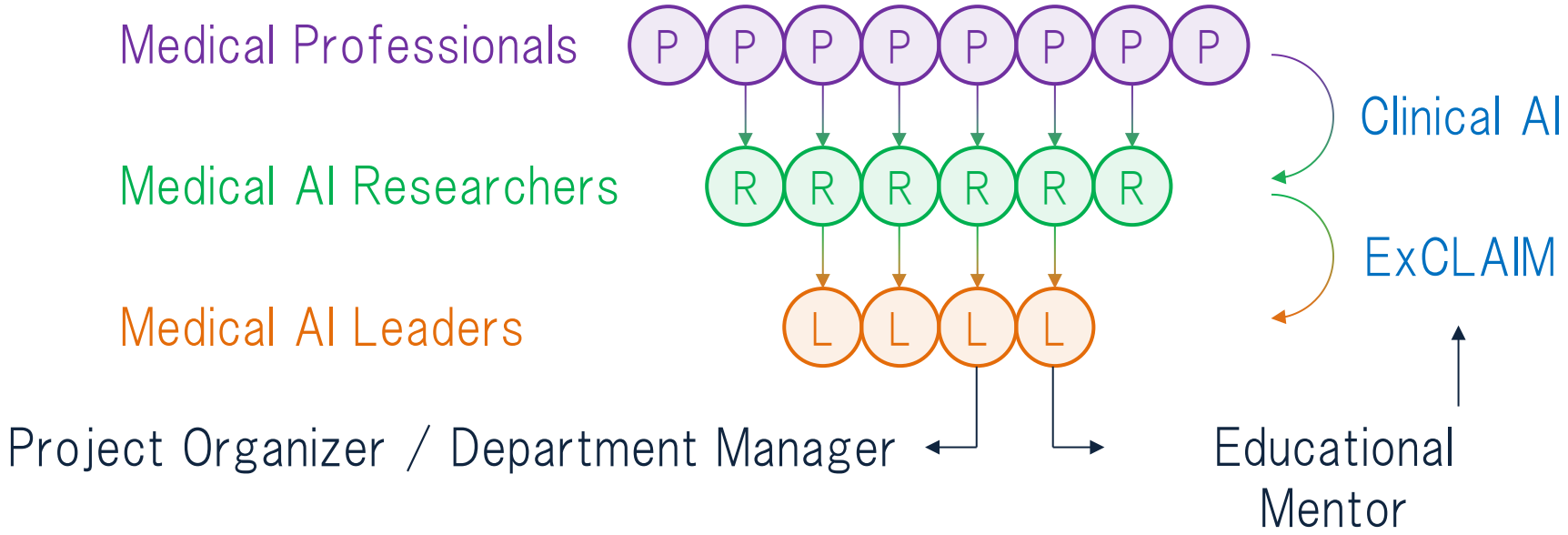
＊ India (Bangalore) – 2019,


＊


For Further Talent Development

ExCLAIM (Experience in Creation for Leaders of AI in Medicine)

* Target: postgraduates / industry professionals
* Method: participate in the activities as a member of the AI Lab
* Objective: talent development for Medical AI Leaders



Summary

医用画像工学分野 検　索

- Impact: Medical image diagnosis and computer (AI) support
- Challenges and educational solutions for clinical applications