Artificial Intelligence (AI) in Critical Care Medicine

高國晉 醫師

林口長庚醫院內科部胸腔內科 副部長

長庚大學醫學院 教授

28th, April, 2019

11:30-12:10

Conflict of Interest

- Nothing to disclose for this presentation

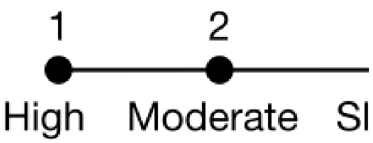
Outline

- AI in Real world
- What is AI ?
- AI in Medicine
  - AI in Critical Care Medicine
- AI : deep thinking
- Summary

緣起

Accounting for Label Uncertainty in Machine Learning for Detection of Acute Respiratory Distress Syndrome

Narathip Reamaroon, Michael W. Sjoding, Kaiwen Lin, Theodore J. Iwashyna, and Kayvan Najarian, *Senior Member, IEEE*



$d_{ab} = 1 - $

where

Using

tion dist

patient, t

Algorithm 1:
time-series dat

Input : All a
for *each patie*
partition d
classifica
if *size of e*
sampl
else
1) sel
samp
2) cal
to ⟨X
3) san
dista
repeat
1)
2)
3)
until
end
end
Output: Partial data $\{X_t, X_{t1}, X_{t2}, \ldots,$
inter-correlation from each patient.

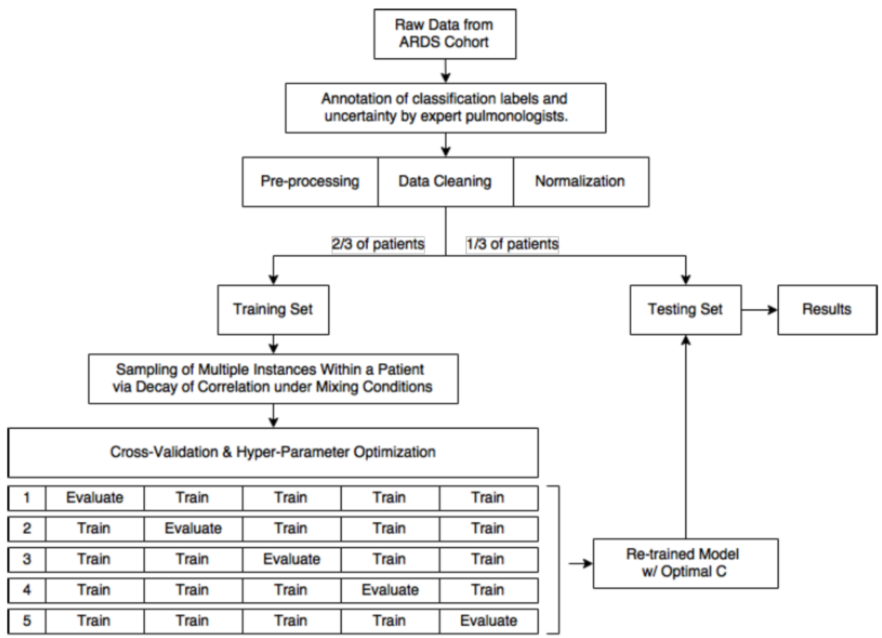
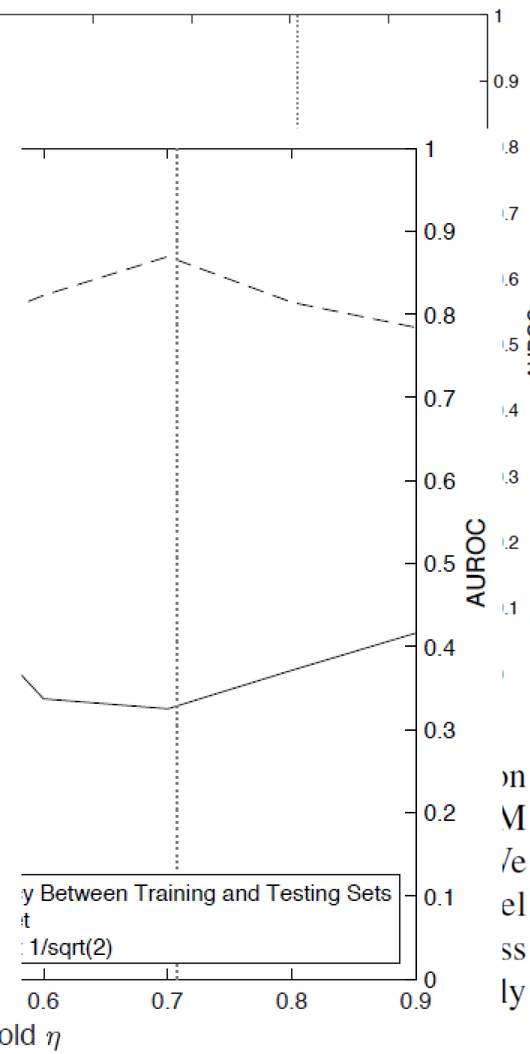


Fig. 4: Flowchart of this study's protocol with 5-fold cross-validation and hyper-parameter optimization using grid search. All samples from the same patient are kept exclusively in either the training or testing set. Hyper-parameter optimization was implemented for separately each model (with and without label uncertainty weight) to give an accurate assessment of performance.



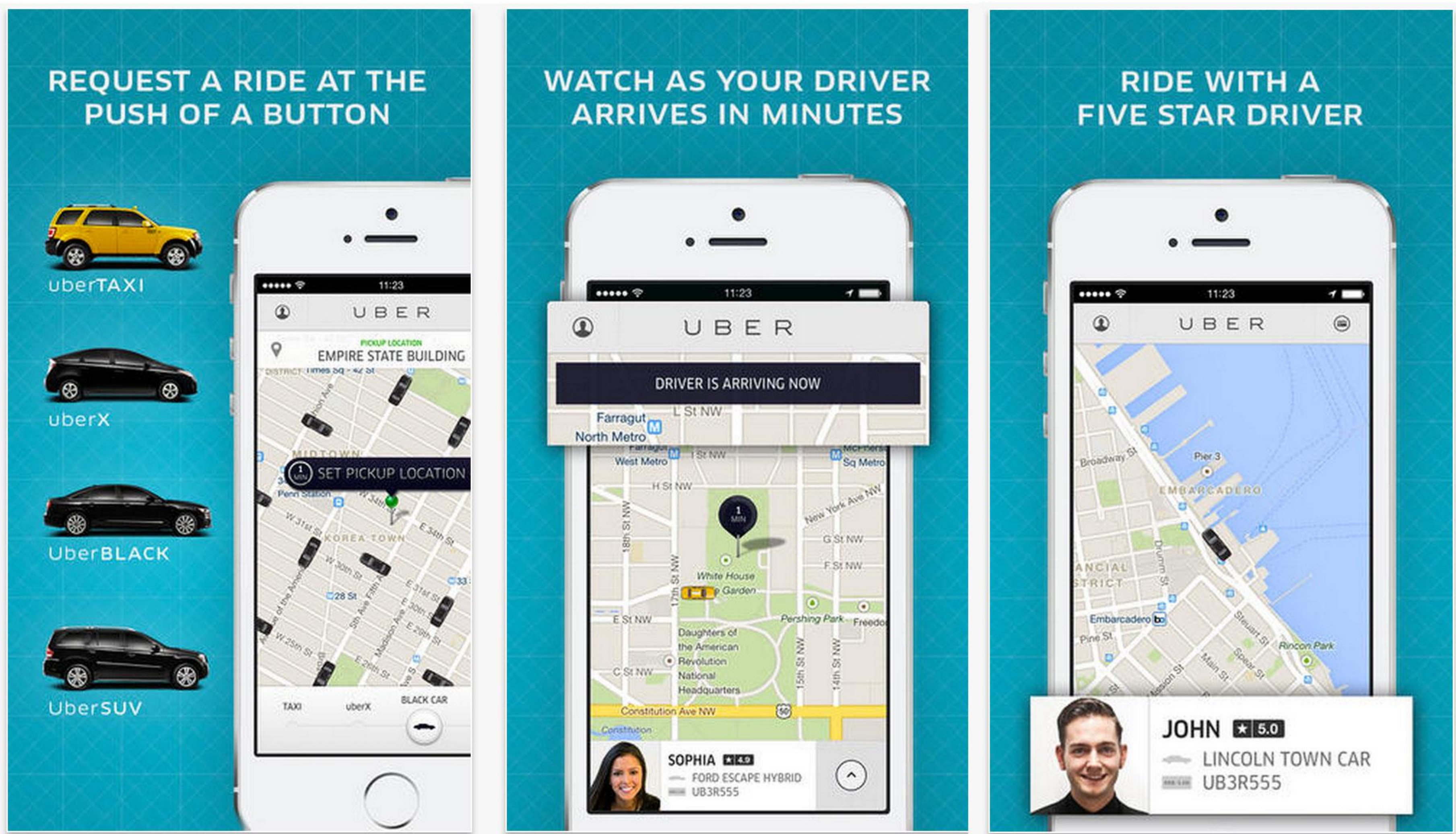
on
M
/e
el
ss
ly

- 看不懂ARDS的paper ，令我驚慌？

AI in Real world

以出國旅遊住飯店為例

UBER


REQUEST A RIDE AT THE PUSH OF A BUTTON
uberTAXI
uberX
UberBLACK
UberSUV
11:23
UBER
PICKUP LOCATION
EMPIRE STATE BUILDING
SET PICKUP LOCATION
TAXI
uberX
BLACK CAR
WATCH AS YOUR DRIVER ARRIVES IN MINUTES
11:23
UBER
DRIVER IS ARRIVING NOW
SOPHIA
FORD ESCAPE HYBRID
UB3R555
RIDE WITH A FIVE STAR DRIVER
11:23
UBER
JOHN 5.0
LINCOLN TOWN CAR
UB3R555

機場櫃台自助check in


HAVE A NICE TRIP!
HELLO KITTY
EVA AIR

自助掛行李

自助通關


歡迎使用自動查驗通關系
Welcome to use E-gate
ROC Citizen & Foreign Resident
with Multiple Re-Entry Only
6
5
4
3
等待線
Waiting Line
#1160577

飛機自動駕駛

自助飯店: 機器人check in


主機自動進行清掃
AUTO自動模式
by水星人


機器人
送餐
全自動飯店
太神啦!!!
SKATE


U148.NET

自助飯店: 機器人check out

What is AI ?

AI = ARTIFICIAL INTELLIGENCE

[人工智能(慧)]

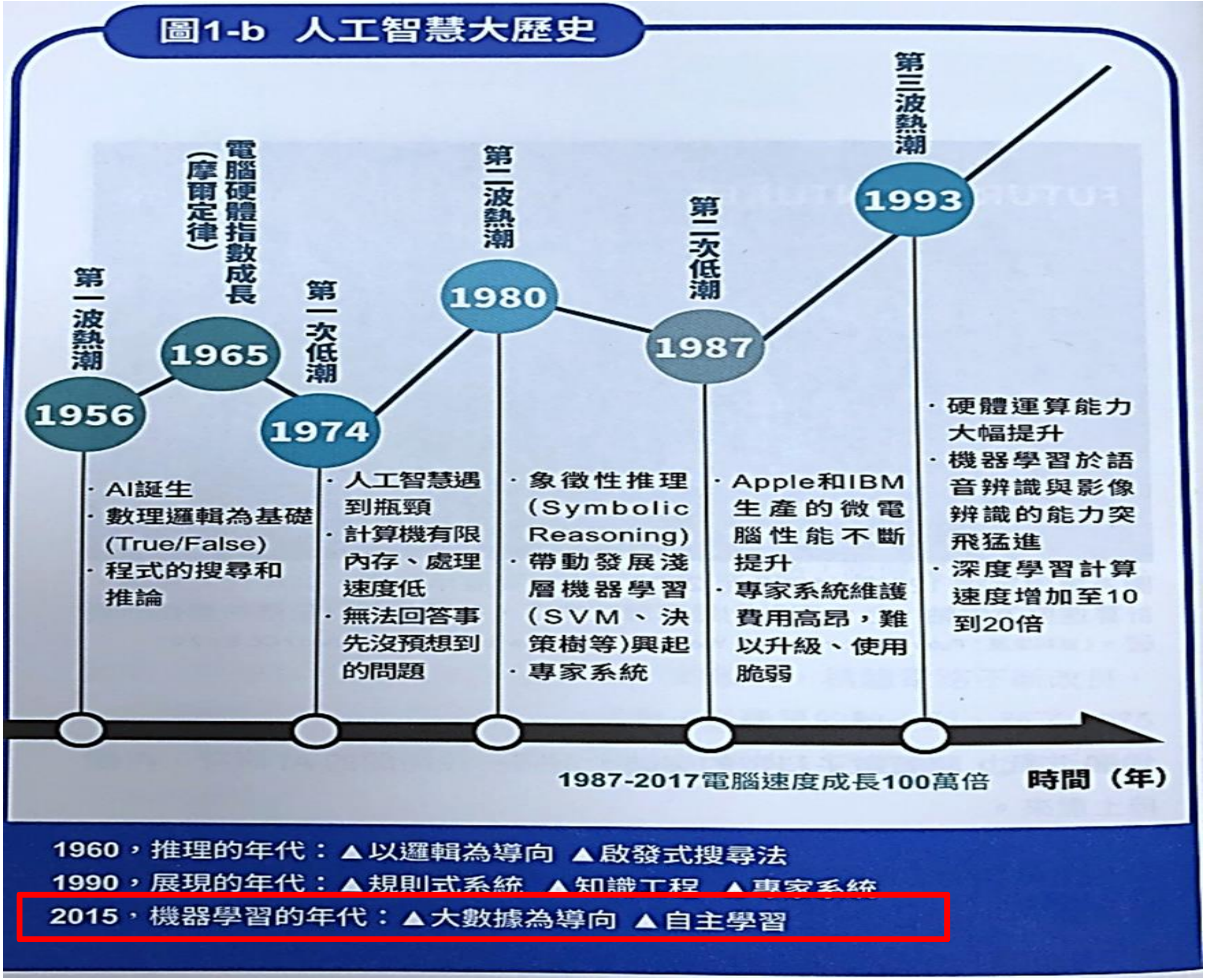
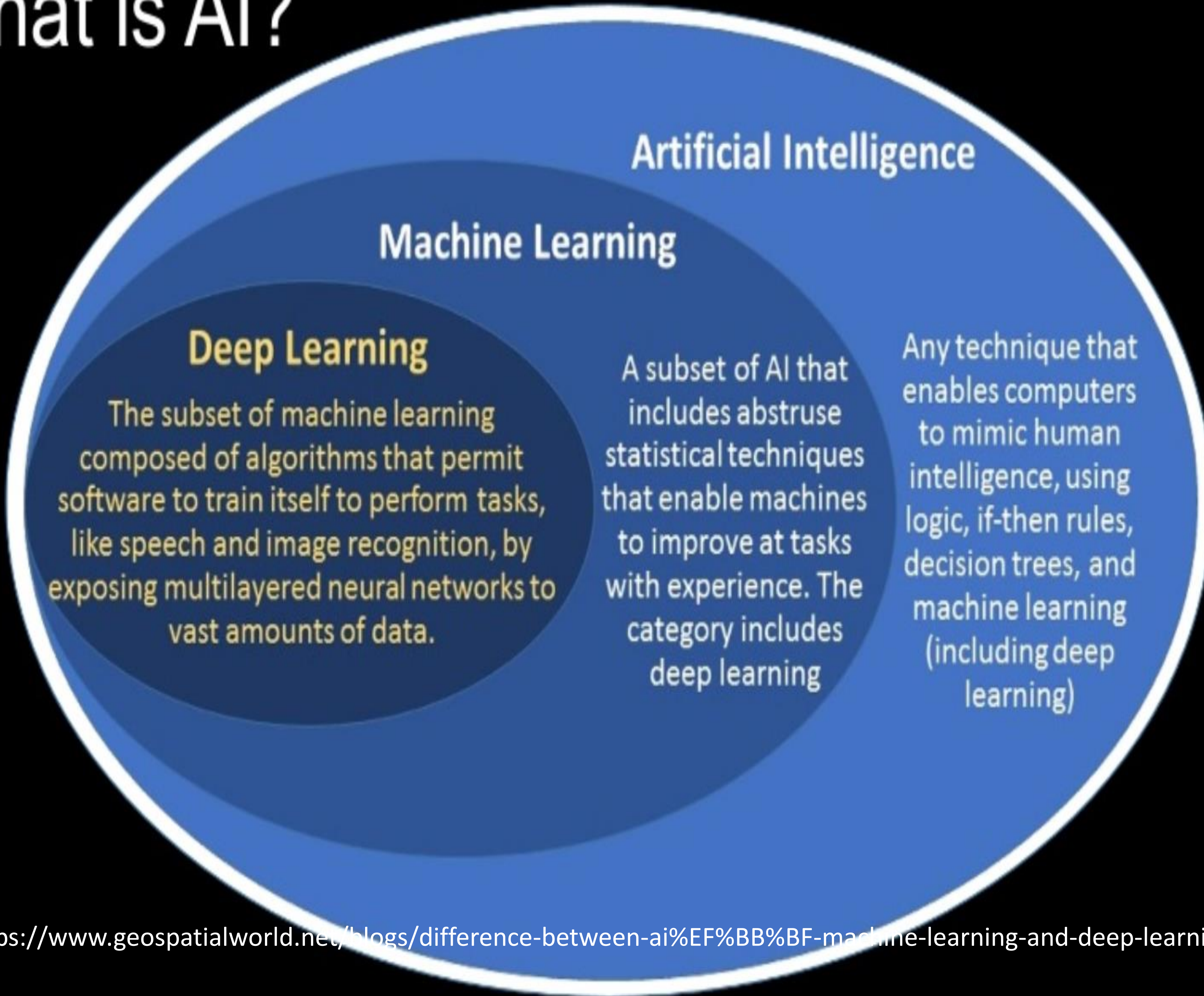


圖1-b 人工智慧大歷史
1956
第一波熱潮
· AI誕生
· 數理邏輯為基礎 (True/False)
· 程式的搜尋和推論
1965
電腦硬體指數成長 (摩爾定律)
1974
第一次低潮
· 人工智慧遇到瓶頸
· 計算機有限內存、處理速度低
· 無法回答事先沒預想到的問題
1980
第二波熱潮
· 象徵性推理 (Symbolic Reasoning)
· 帶動發展淺層機器學習 (SVM、決策樹等)興起
· 專家系統
1987
第二次低潮
· Apple和IBM生產的微電腦性能不斷提升
· 專家系統維護費用高昂，難以升級、使用脆弱
1993
第三波熱潮
· 硬體運算能力大幅提升
· 機器學習於語音辨識與影像辨識的能力突飛猛進
· 深度學習計算速度增加至10到20倍
1987-2017電腦速度成長100萬倍
時間（年）
1960，推理的年代：▲以邏輯為導向 ▲啟發式搜尋法
1990，展現的年代：▲規則式系統 ▲知識工程 ▲專家系統
2015，機器學習的年代：▲大數據為導向 ▲自主學習

李友專：醫療大未來

What is AI?



https://www.geospatialworld.net/blogs/difference-between-ai%EF%BB%BF-machine-learning-and-deep-learning/

- Artificial intelligence:
  - Any technique that enable computer to mimic human intelligence, using logic, if-then rules, decision trees, and machine learning (including deep learning)
- Machine learning:
  - A subset of AI that includes abstruse statistical techniques that enable machines to improve at tasks with experience. The category includes deep learning.
- Deep learning:
  - The subsets of machine learning composed of algorithms that permit software to train itself to perform tasks, like **speech** and **image** recognition, by exposing multilayered neural networks to vast amounts of data.

Artificial intelligence (AI)

- In computer science, **artificial intelligence** (**AI**), sometimes called **machine intelligence**, is intelligence demonstrated by machines, in contrast to the **natural intelligence** displayed by humans and animals.
- The core problems of AI include programming computers for certain traits such as: Knowledge, Reasoning, Problem solving, Perception, Learning, Planning, Ability to manipulate and move objects.

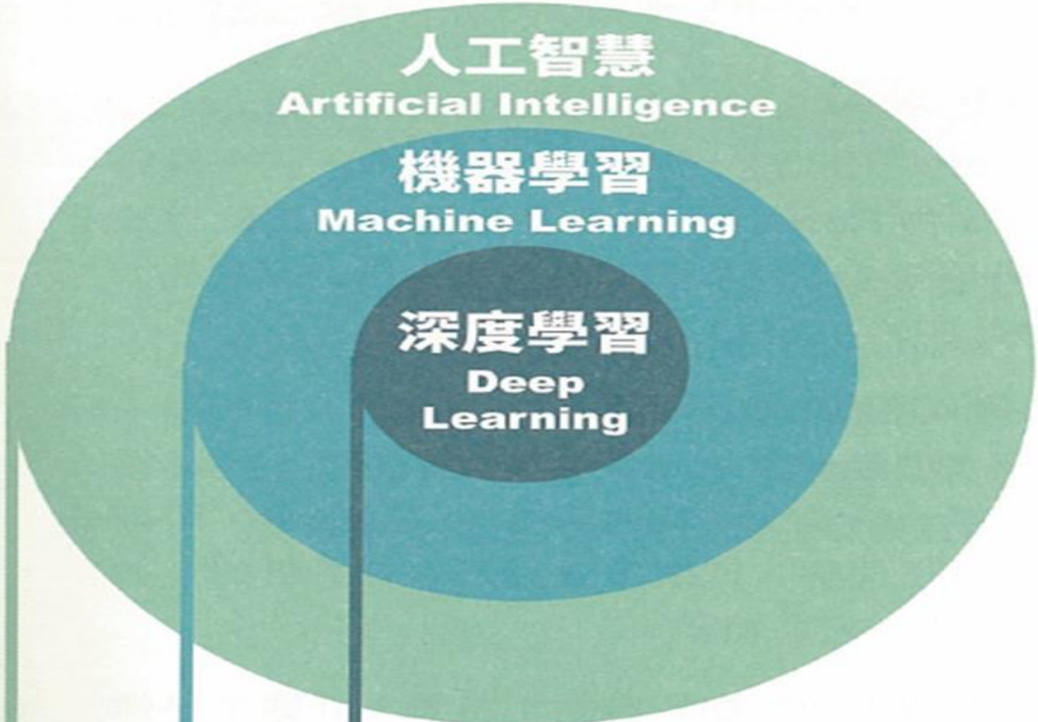
“人工智(能)慧” 定義

「系統正確解釋外部資料，從這些資料中學習並利用這些知識，通過靈活適應實現特定目標和任務的能力」。

by Andreas Kaplan and Michael Haenlein

- Business Horizons, 2019 62(1), 15-25

圖1-d 人工智慧、機器學習和深度學習[6]



實現機器學習的技術之一，讓電腦模擬人類神經網絡運作方式，特別應用在視覺辨識、語音識別、自然語言處理、生物醫學等領域。

實現人工智慧的方法之一，透過使用大量的資料和演算法「訓練」電腦，從中找出規律來「學習」，讓機器像人類一樣，具備學習與判斷能力。

讓機器展現人類的智慧，開始專注於模仿人類推理過程的思考模式。

簡單說，深度學習是機器學習的一種，機器學習是人工智慧方法學的一種。

李友專: 醫療大未來

Machine Learning

- The field of study that focuses on how computers learn from data and the development of **algorithms** that make this learning possible
  - Supervised learning (監督學習):
    - Algorithms that are used to uncover the relationship between a set of **features** and one or more known **outcomes**
  - Unsupervised learning (非監督學習):
    - Algorithms that are used to uncover naturally occurring **patterns or groupings** in the data, without targeting a specific outcome

Machine Learning

- **監督學習**
  - 指事先給定機器一些**訓練樣本**並且告訴樣本的類別，然後根據這些樣本的類別進行訓練，提取出這些樣本的共同屬性或者訓練一個分類器，等**新來一個樣本**，則通過訓練得到的共同屬性或者分類器進行判斷該樣本的類別。

- **非監督學習**
  - 是不給定訓練樣本，直接給定一些樣本和一些規則，讓機器自動根據一些規則進行分類。

Big data

- Digital data that are generated in high volume and high variety and that accumulate at high velocity, resulting in datasets too large for traditional data-processing systems

Algorithms (演算法)

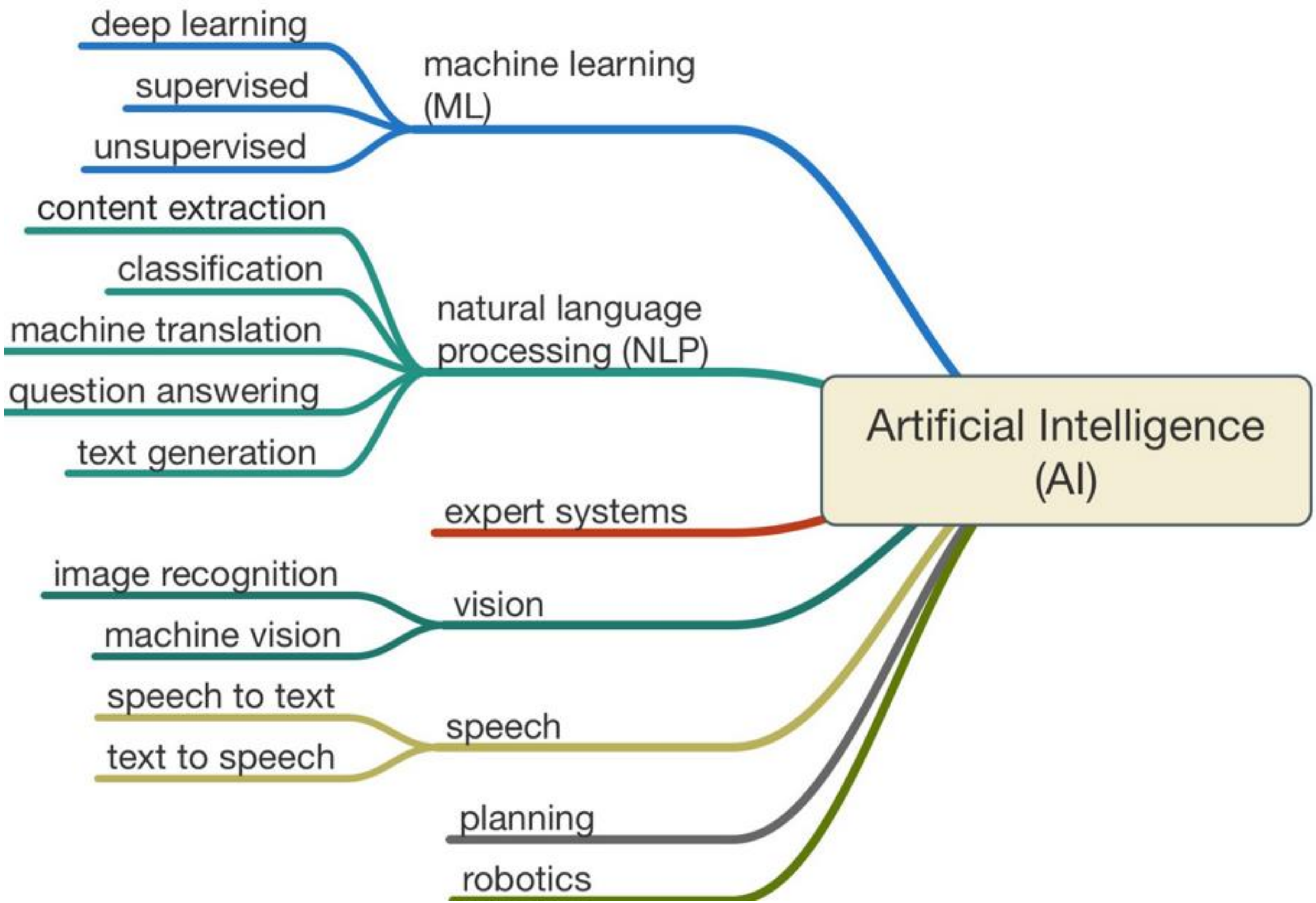
- In mathematics and computer science, an **algorithm** is an unambiguous specification of how to solve a class of problems.
- Algorithms can perform calculation, data processing, automated reasoning, and other tasks.

Models

- Model training
  - The process through which machine learning algorithms develop a model of the data by learning the relationships between features and, in supervised learning, between features and outcomes. This is also referred to as model derivation or data fitting
- Model validation (Model testing)
  - The process of measuring how well a model fits **new**, **independent data**. For example, evaluating the performance of a supervised model at predicting an outcome in new data.
- Predictive model
  - A model generally trained to predict the likelihood of a condition, event, or response. The US FDA specifically considers predictive strategies as those geared toward identifying groups of patients more likely to **respond to an intervention**
- Prognostic model
  - A model specifically trained to predict the likelihood of a condition-related endpoint or outcome such as **mortality**. In general, the goal is to estimate a prognosis given a set of baseline features, regardless of what ultimately leads to the outcome

AI如何運作 –合宜的「模型建構」

- 機器學習 (Machine learning): 操作並建構AI的演算法
  - 機器學習 Deep learning (DL), 強化學習 reinforced learning (RL)
  - 卷積神經網路 Convoluted neural network (CNN)
  - 遞歸神經網路 Recurrent neural network (RNN)
  - 多層感知器 Multilayer perception (MLP)
  - 生成對抗網絡 Generative adversarial network (GAN)
- 多數模型與演算法都被包裝到平台，且開源
  - Google – tensorflow
  - Facebook – Caffe
  - Microsoft Cognitive Toolkit/CNTK
  - Etc.


deep learning
supervised
unsupervised
machine learning (ML)
content extraction
classification
machine translation
question answering
text generation
natural language processing (NLP)
expert systems
image recognition
machine vision
vision
speech to text
text to speech
speech
planning
robotics
Artificial Intelligence (AI)

AI 三要素

- Big data
- Machine learning
- Models

Conceptual Overview of Supervised Machine Learning

A Preparing to Build a Model

Task Definition

Conceptual task: Translate text into another language
More precise task: Convert short snippets of text from English to Spanish

Machine learning starts with a task definition that specifies inputs and corresponding outputs.

Data Collection

Raw data: Transcripts from clinical encounters in which a medical translator participated

After defining the task, a data set from instances in which the task has already been performed is collected.

Data Preparation

Example of raw input:
"I started feeling pain across my chest."

Example of raw output:
"Empecé a sentir un dolor por todo el pecho."

The raw data are preprocessed to produce examples of inputs consisting of a set of features and an output, referred to as a label. In this example, the features are numerical tokens that correspond to words in the raw text (e.g.,"chest" is represented by the token <100>).

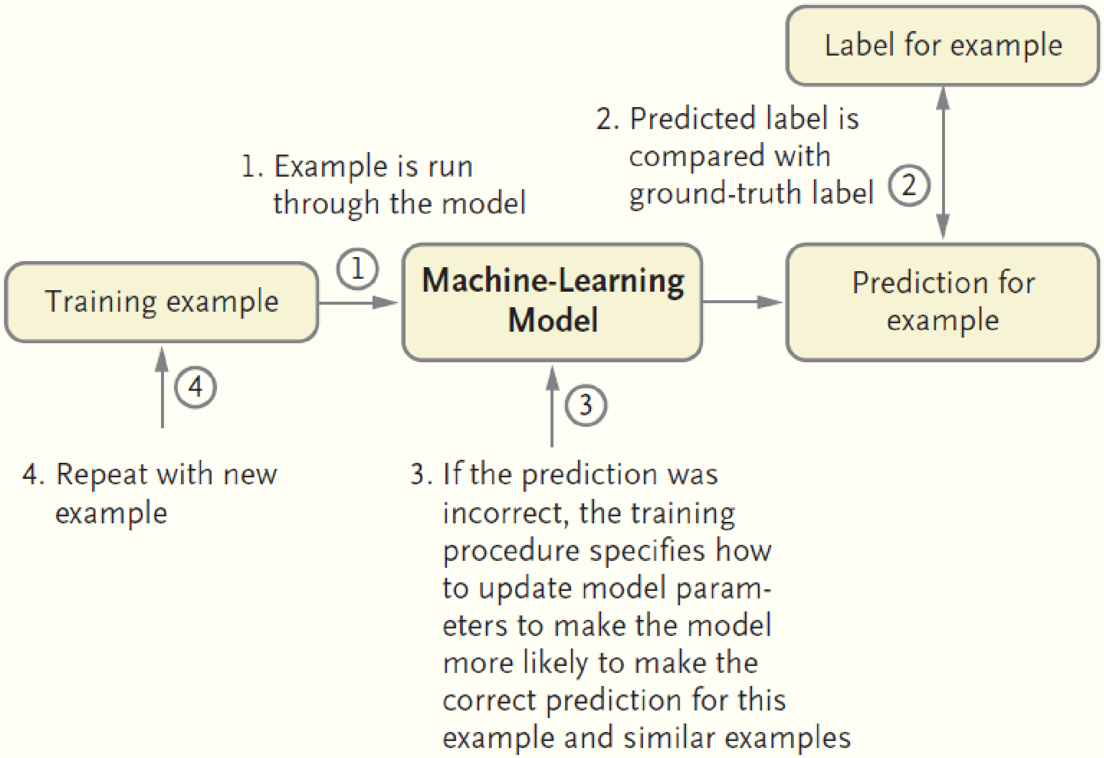
Example of features:
[<1>, <58>, <145>, <3>, <5>, <67>, <22>, <15>, <100>]

Example of label:
[<934>, <1024>, <2014>, <955>, <1001>, <1500>, <1643>, <1923>, <203>]

The set of processed examples is divided into two sets. The first, the training data set, is used to build the model. The second, the test set, is used to assess how well the model performs.

N Engl J Med 2019;380:1347-58.

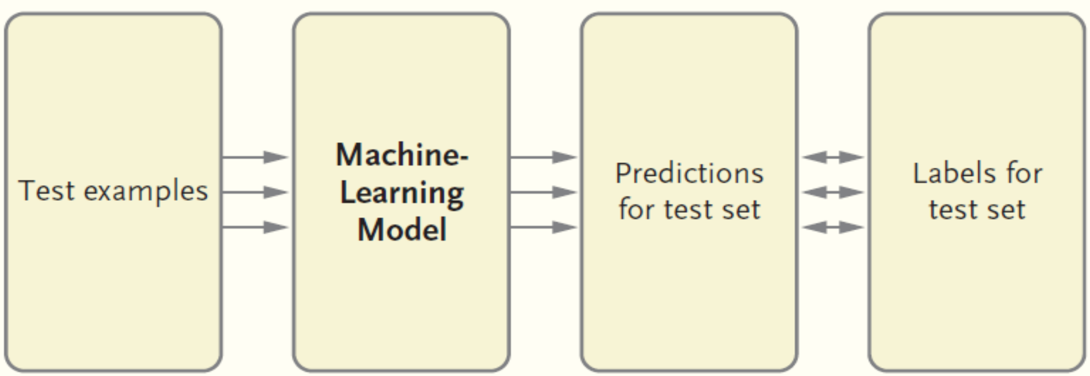
B Training a Model



During model training, an example from the training set is sent through a machine-learning system, which provides a mathematical function that converts features to a predicted label. A simple example is a linear function, $y'=ax_1+bx_2+c$, where y' is the predicted label, x_1 and x_2 are the features, and a, b, and c are parameters. The model parameters are initially randomly assigned, and in the first iteration, the predicted label y' is generally unrelated to the ground-truth label.

In the key step of machine learning (step 3), an algorithm determines how the parameters need to be modified to make the prediction more likely to match the ground truth. The system iterates through all the examples in the training data, potentially multiple times, to complete training.

C Evaluating a Model

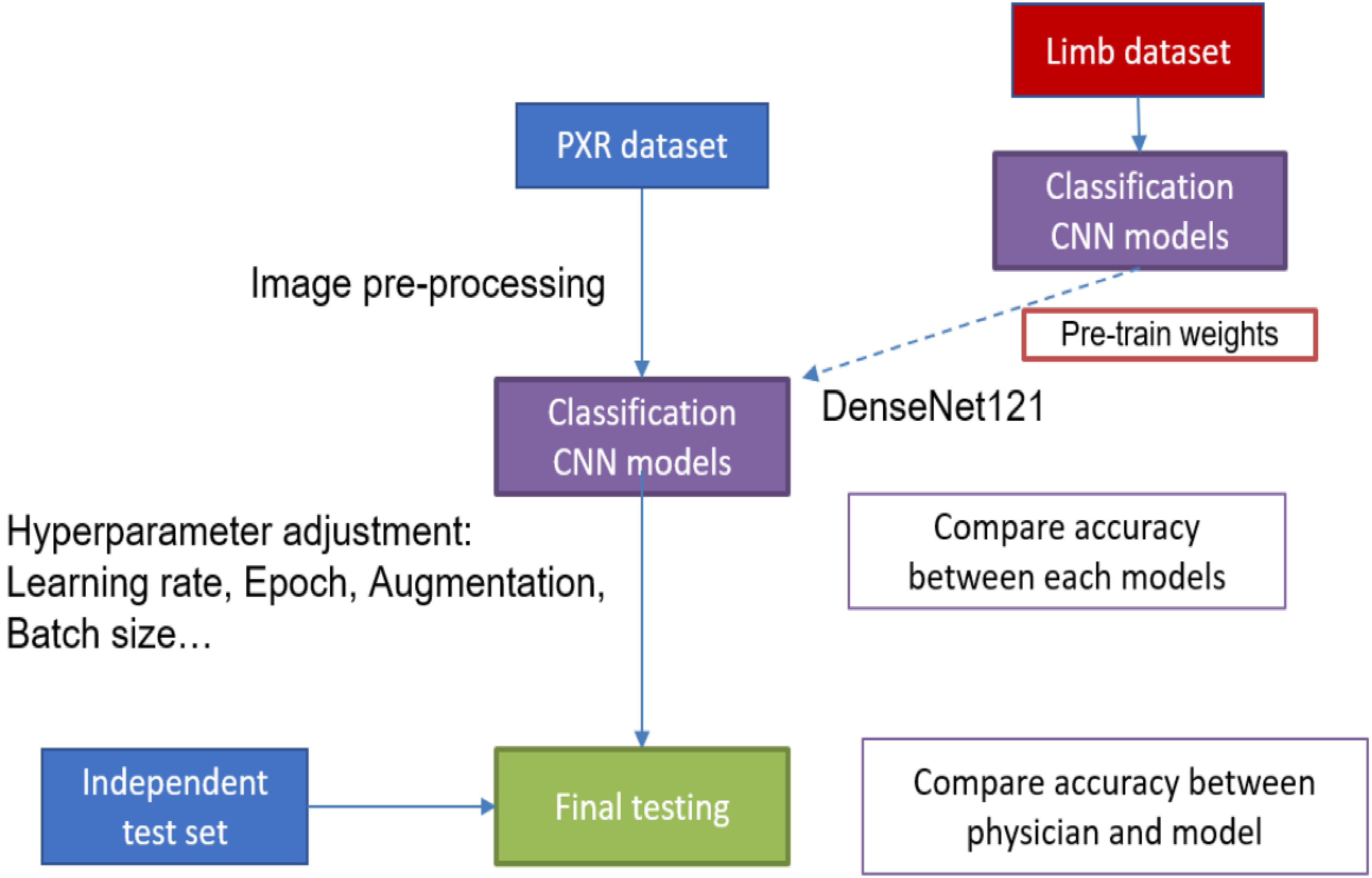


The test set is then run through the final model. Statistics are computed, and the predictions of the test set are compared with the ground-truth labels.

To apply the model, new input examples, which have not been previously labeled, can be run through the model. However, the model learns patterns from data only in the training set, so if new examples are sufficiently different from those in the training data, the model may not produce accurate predictions for them.

N Engl J Med 2019;380:1347-58.

Build CNN models and evaluation



AI in Medicine

Artificial Intelligence (AI) in Medicine

- 第一波:
  - 1954, Eliza DOCTOR
- 第二波:
  - 1980, 專家系統
- 第三波:
  - Machine learning

Artificial Intelligence (AI) in Medicine

- 第一波:
  - 1954, Eliza DOCTOR
- 第二波:
  - 1980, 專家系統
- 第三波:
  - Machine learning

ELIZA-BEGINNING OF ERA OF ARTIFICIAL INTELLIGENCE

etherealcreation (45) 在 science • 去年

Look at a ridiculous conversation between Eliza and Parrry From 1972

Network Working Group V. Cerf
Request for Comments: 439 SU-ERL
NIC: 13771 21 January 1973

PARRY Encounters the DOCTOR

The session recorded below actually happened on 18 September 1972 and is herewith reproduced for your entertainment.

Good Evening.

Tell me your problems. Please terminate input with a period or a question mark.

*People get on my nerves sometimes.

I am not sure I understand you fully.

*You should pay more attention.

Suppose you should pay more attention.

*You're entitled to your own opinion.

Artificial Intelligence (AI) in Medicine

- 第一波:
  - 1954, Eliza DOCTOR
- 第二波:
  - 1980, 專家系統
- 第三波:
  - Machine learning

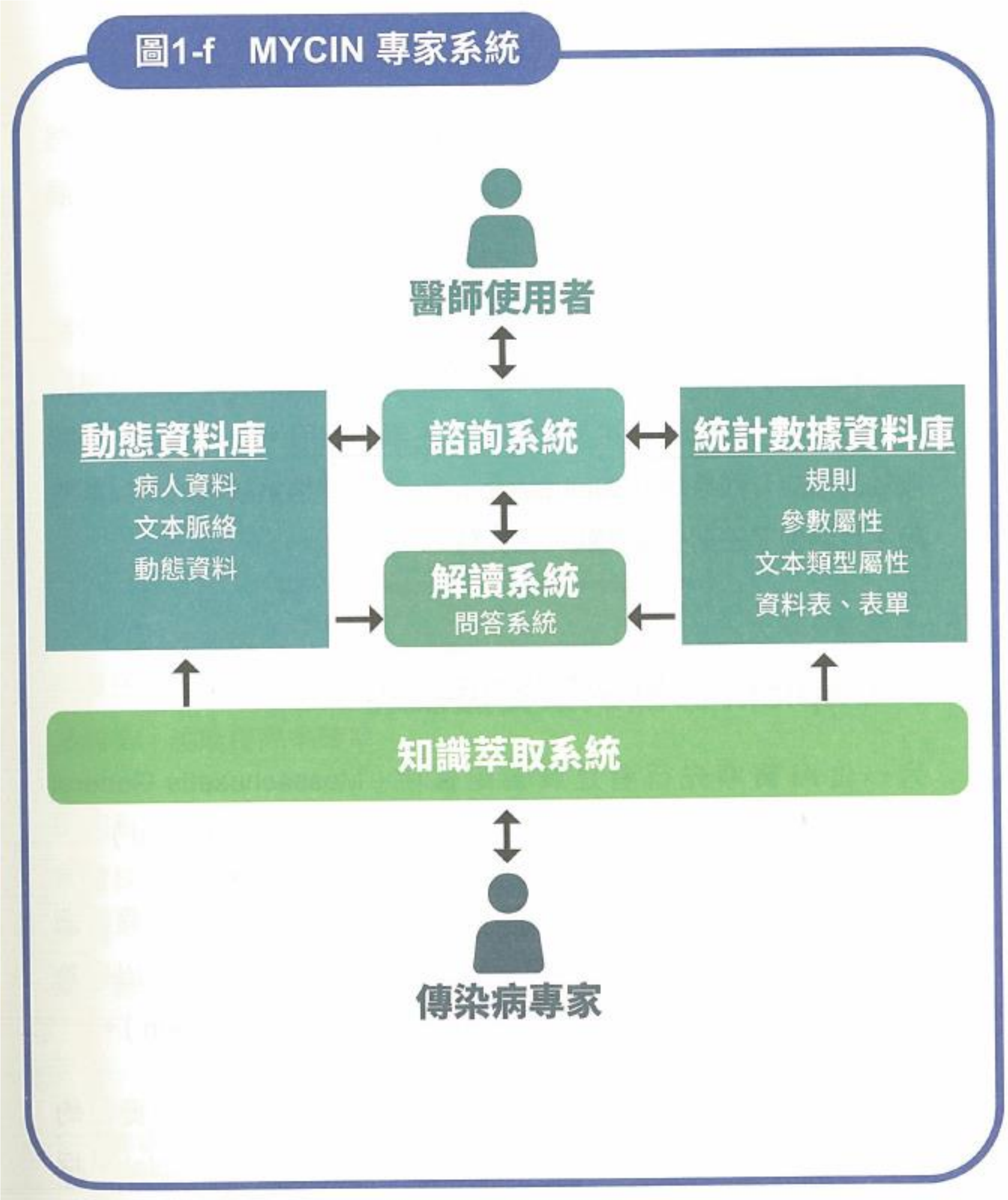
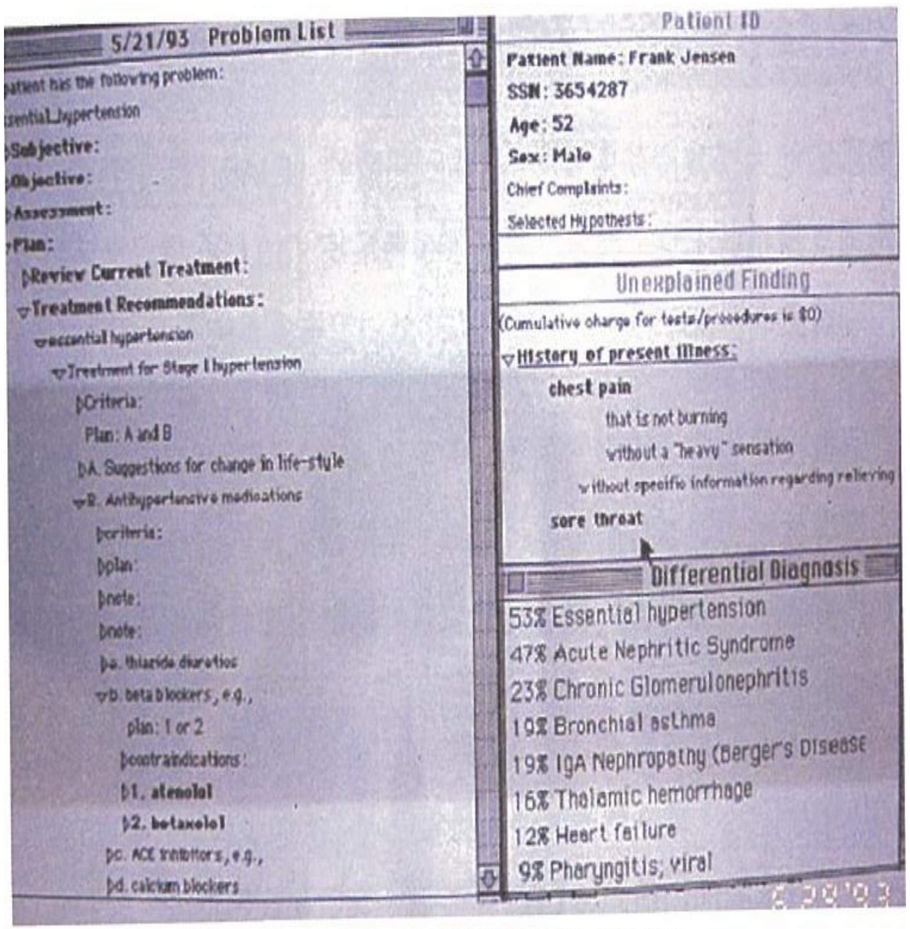


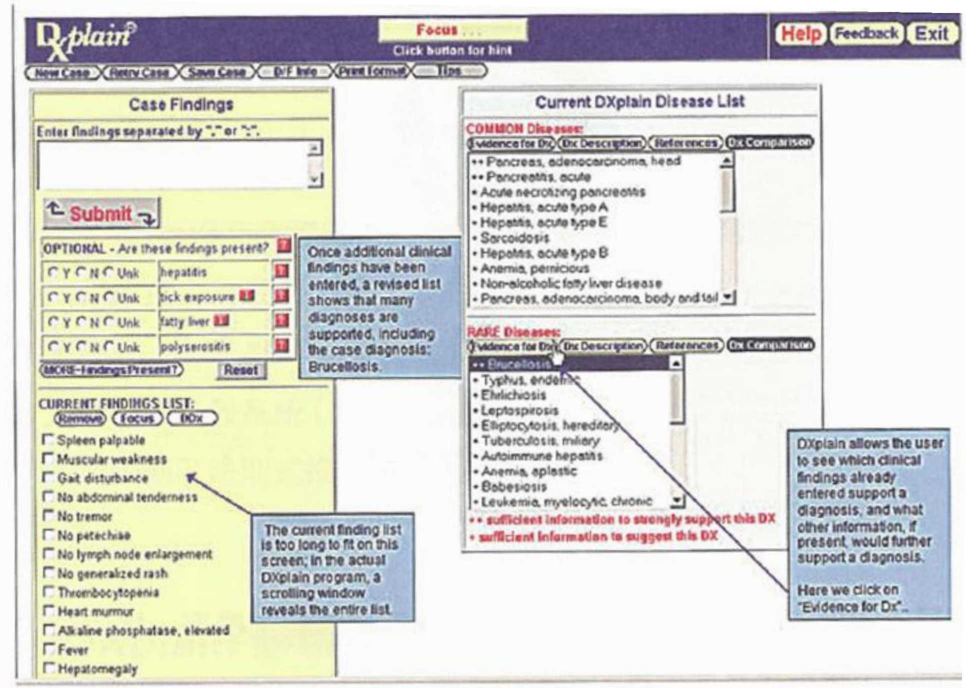
圖1-f MYCIN 專家系統
醫師使用者
動態資料庫
病人資料
文本脈絡
動態資料
諮詢系統
統計數據資料庫
規則
參數屬性
文本類型屬性
資料表、表單
解讀系統
問答系統
知識萃取系統
傳染病專家

李友專: 醫療大未來

Clinical Decision Support System (Program)

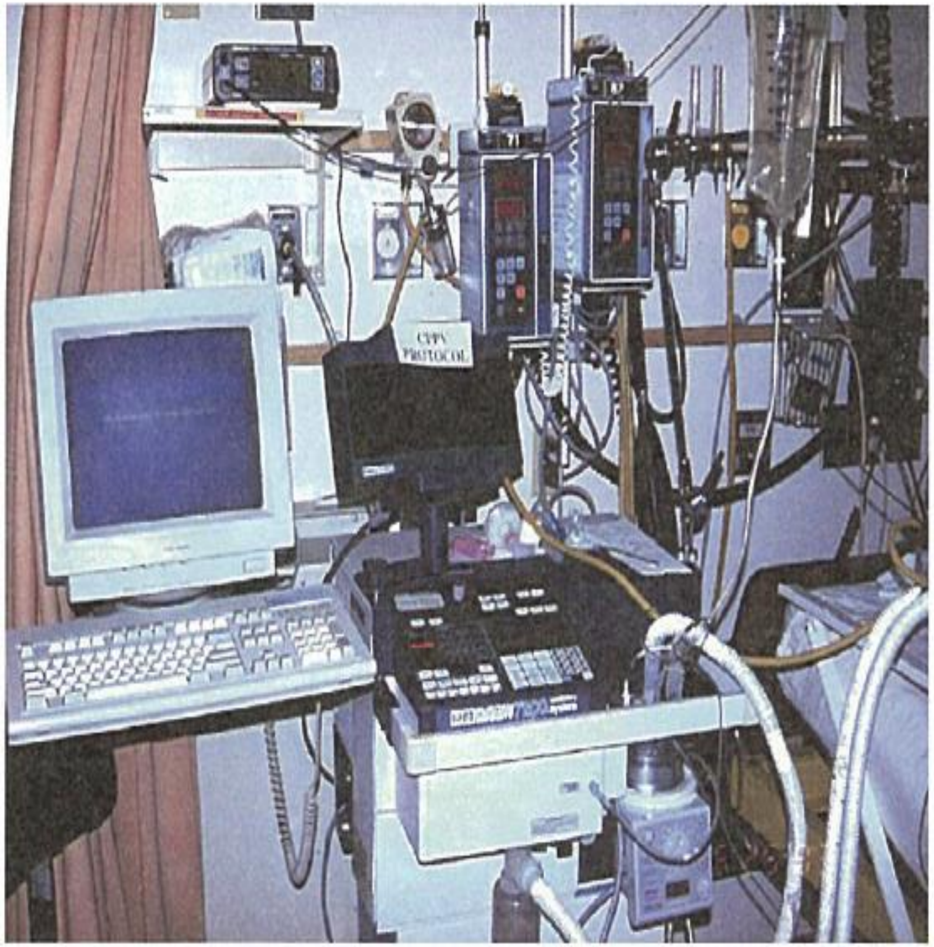


ILIAD 伊里亞德診斷決策支援系統



Dxplain 診斷決策支援系統

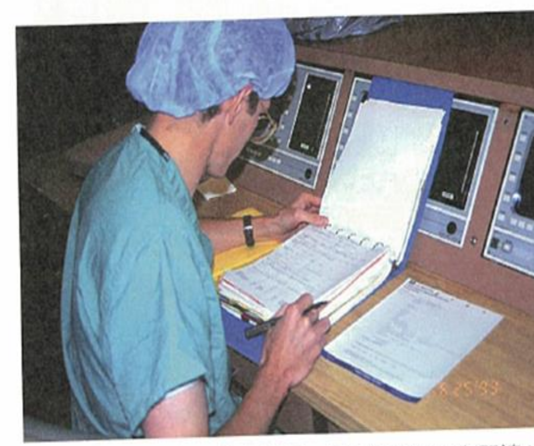
李友專：醫療大未來



與呼吸器密切結合的 AI 決策支援系統。（照片提供：李友專）

1. 左圖紅框中的設備為 Medical Information Bus（MIB），用來整合所有床邊監測設備的生理訊號參數，後來發展為 IEEE 標準之一：IEEE1073。

2. 早期 ICU 感染治療系統。讀取 MIB 資料後，自動判斷 ICU 病人是否發生細菌感染及建議使用抗生素。（該系統之訓練資料為五年 ICU 院內感染病人資料）



早期 ICU 報告必須由人工進行摘要，圖中的醫師正在閱讀 ICU 報告。

（照片提供：李友專）

李友專: 醫療大未來

Artificial Intelligence (AI) in Medicine

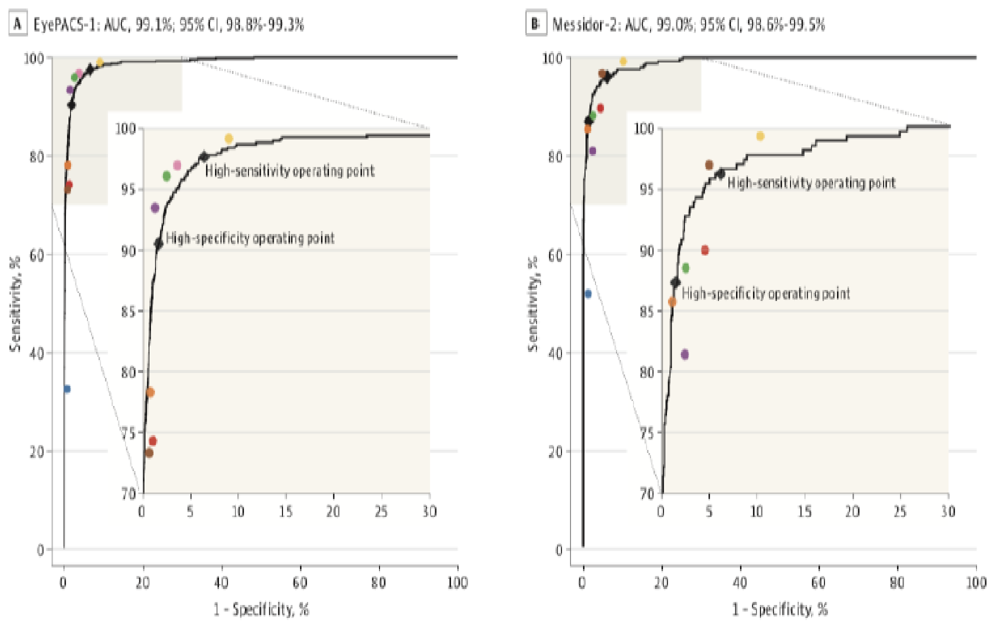
- 第一波:
  - 1954, Eliza DOCTOR
- 第二波:
  - 1980, 專家系統
- 第三波:
  - Machine learning

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

Figure 2. Validation Set Performance for Referable Diabetic Retinopathy

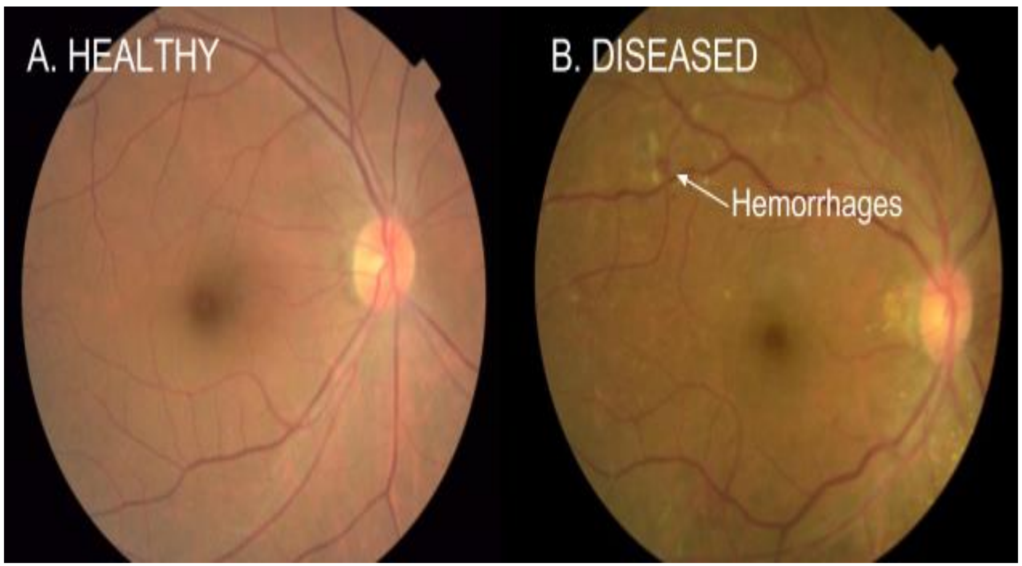


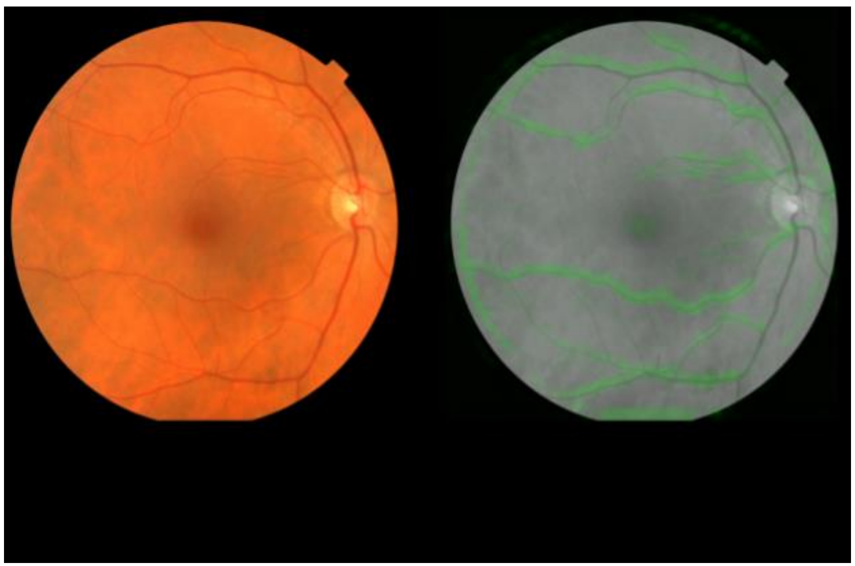
Gulshan V, et al. *JAMA*. 2016;316(22):2402-2410.

JAMA | Original Investigation

Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes

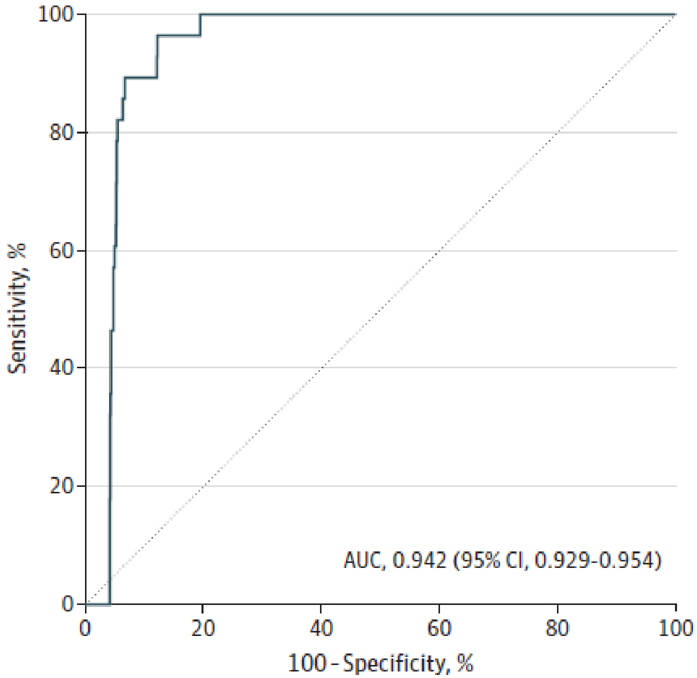
Daniel Shu Wei Ting, MD, PhD; Carol Yim-Lui Cheung, PhD; Gilbert Lim, PhD; Gavin Siew Wei Tan, FRCSEd; Nguyen D. Quang, BEng; Alfred Gan, MSc; Haslina Hamzah, BSc; Renata Garcia-Franco, MD; Ian Yew San Yeo, FRCSEd; Shu Yen Lee, FRCSEd; Edmund Yick Mun Wong, FRCSEd; Charumathi Sabanayagam, MD, PhD; Mani Baskaran, MD, PhD; Farah Ibrahim, MB, BCh, BAO; Ngiap Chuan Tan, MCI, FAMS; Eric A. Finkelstein, MHA, PhD; Ecosse L. Lamoureux, PhD; Ian Y. Wong, FRCOph; Neil M. Bressler, MD; Sobha Sivaprasad, FRCOph; Rohit Varma, MD, MPH; Jost B. Jonas, MD, PhD; Ming Guang He, MD, PhD; Ching-Yu Cheng, MD, PhD; Gemmy Chui Ming Cheung, FRCOph; Tin Aung, MD, PhD; Wynne Hsu, PhD; Mong Li Lee, PhD; Tien Yin Wong, MD, PhD



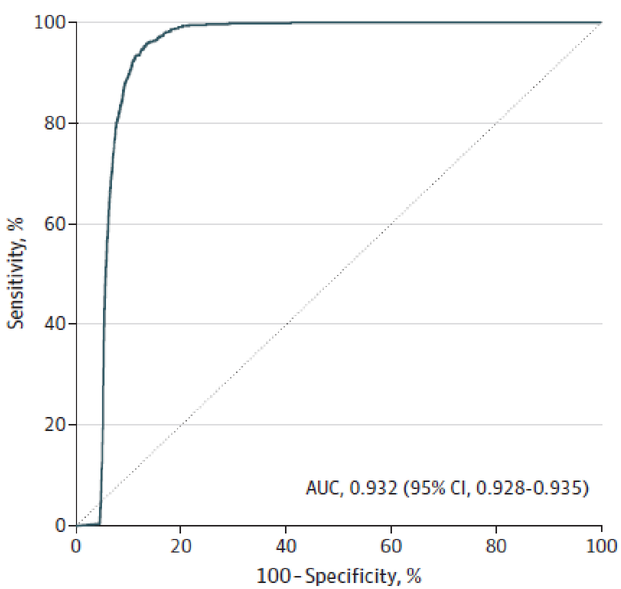


JAMA. 2017;318(22):2211-2223. doi:10.1001/jama.2017.18152

A Deep learning system for detecting possible glaucoma



B Deep learning system for detecting referable AMD



Key Points

Question How does a deep learning system (DLS) using artificial intelligence compare with professional human graders in identifying diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes?

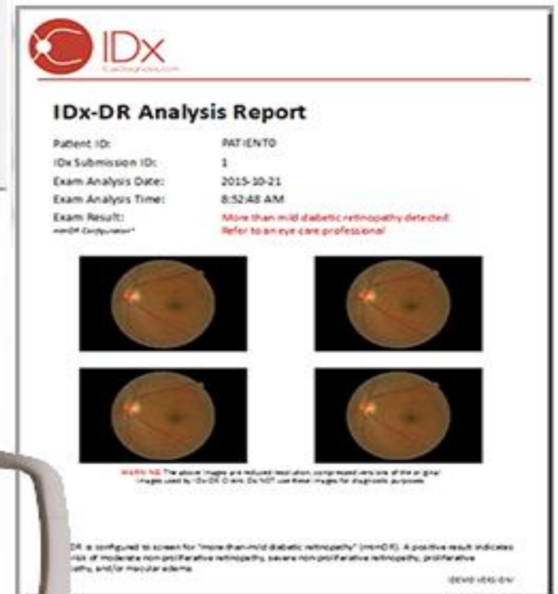
Findings In the primary validation dataset (71 896 images; 14 880 patients), the DLS had a sensitivity of 90.5% and specificity of 91.6% for detecting referable diabetic retinopathy; 100% sensitivity and 91.1% specificity for vision-threatening diabetic retinopathy; 96.4% sensitivity and 87.2% specificity for possible glaucoma; and 93.2% sensitivity and 88.7% specificity for age-related macular degeneration, compared with professional graders.

Meaning The DLS had high sensitivity and specificity for identifying diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes.

JAMA. 2017;318(22):2211-2223. doi:10.1001/jama.2017.18152

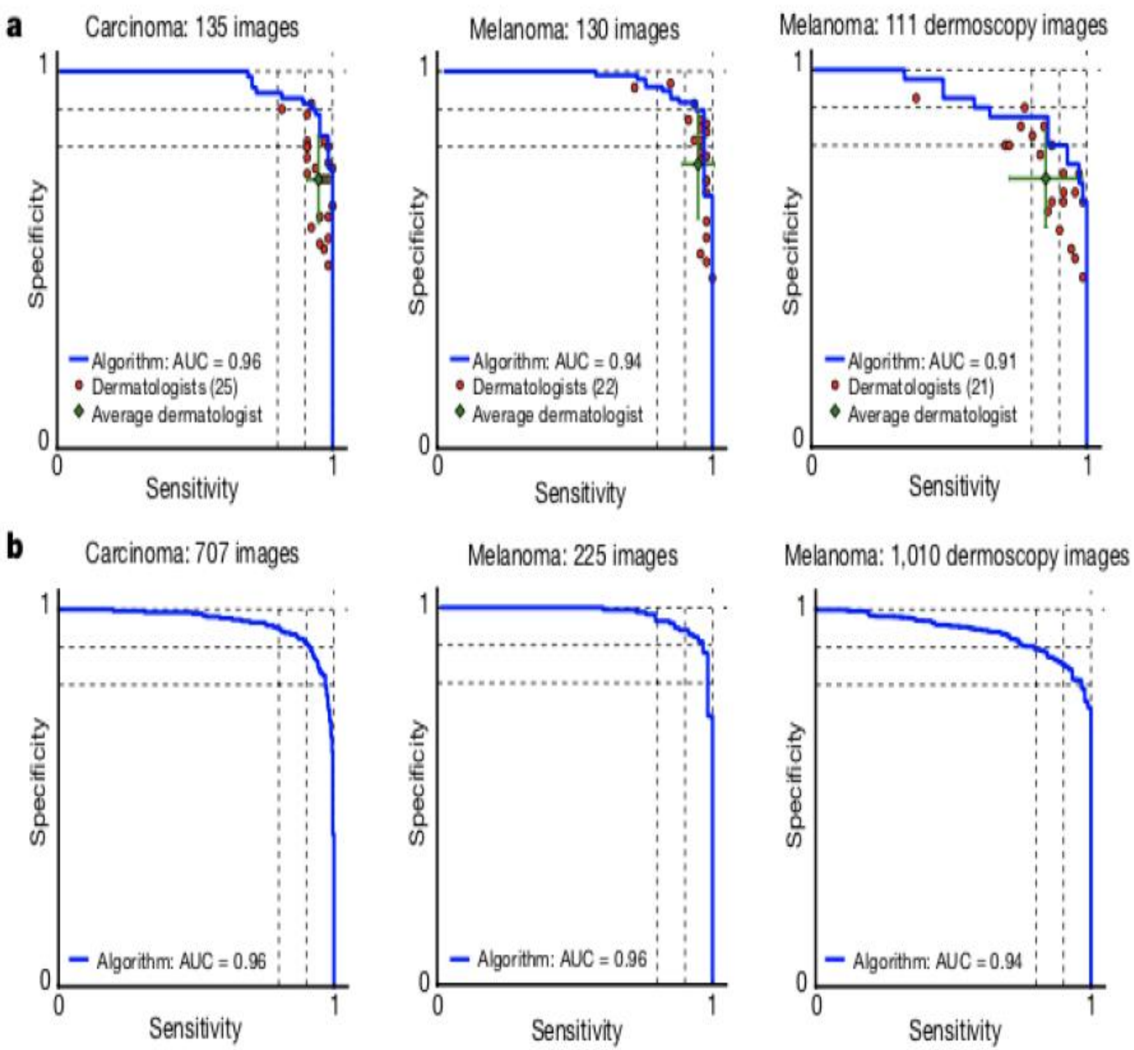
FDA-approved in 11/Apr/2018: IDX-DR

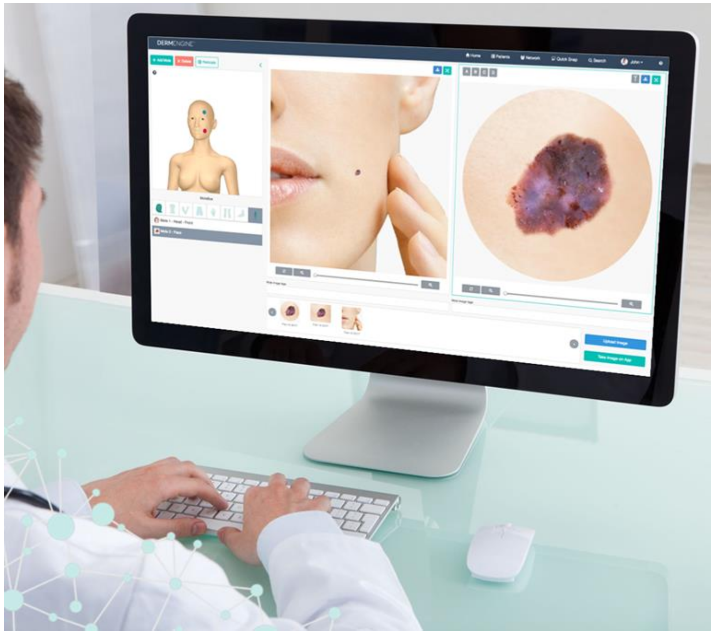
Detect Diabetes Retinopathy

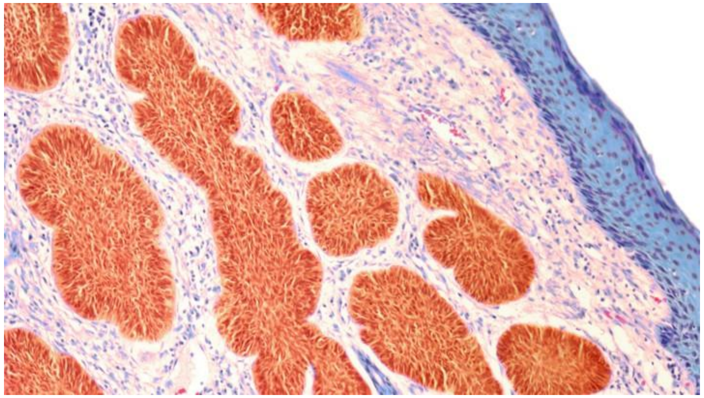


Dermatologist-level classification of skin cancer with deep neural networks

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]



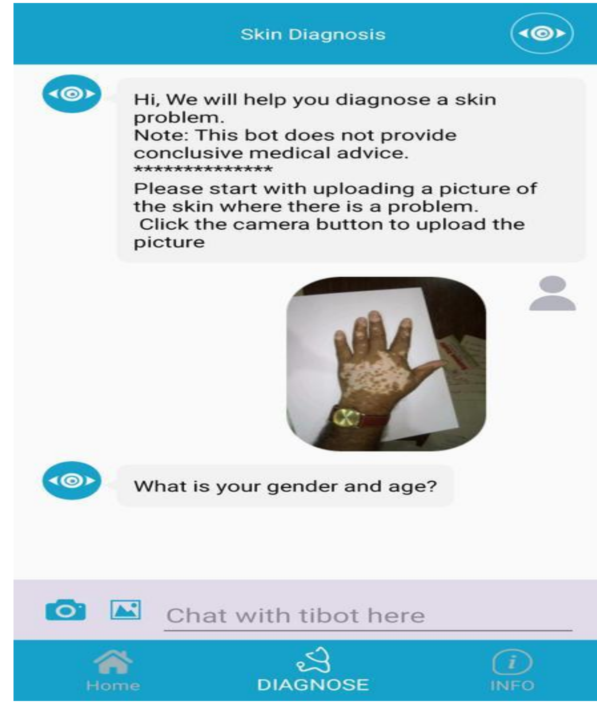




Esteva A, et al. Nature. 2017;542(7639):115-118.

A 49-year-old patient takes a picture of a rash on his shoulder with a smartphone app that recommends an immediate appointment with a dermatologist. His insurance company automatically approves the direct referral, and the app schedules an appointment with an experienced nearby dermatologist in 2 days. This appointment is automatically cross-checked with the patient's personal calendar. The dermatologist performs a biopsy of the lesion, and a pathologist reviews the computer-assisted diagnosis of stage I melanoma, which is then excised by the dermatologist.

N Engl J Med 2019;380:1347-58.



Diagnosis of Idiopathic Pulmonary Fibrosis

An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline

IPF suspected*		Histopathology pattern			
		UIP	Probable UIP	Indeterminate for UIP	Alternative diagnosis
HRCT pattern	UIP	IPF	IPF	IPF	Non-IPF dx
	Probable UIP	IPF	IPF	IPF (Likely)**	Non-IPF dx
	Indeterminate	IPF	IPF (Likely)**	Indeterminate***	Non-IPF dx
	Alternative diagnosis	IPF (Likely)** /non-IPF dx	Non-IPF dx	Non-IPF dx	Non-IPF dx

HRCT Criteria for UIP pattern

UIP	Probable UIP	Indeterminate for UIP	Alternative Diagnosis
Subpleural and basal predominant; distribution is often heterogeneous* Honeycombing with or without peripheral traction bronchiectasis or bronchiolectasis†	Subpleural and basal predominant; distribution is often heterogeneous Reticular pattern with peripheral traction bronchiectasis or bronchiolectasis May have mild GGO	Subpleural and basal predominant Subtle reticulation; may have mild GGO or distortion ("early UIP pattern") CT features and/or distribution of lung fibrosis that do not suggest any specific etiology ("truly indeterminate")	Findings suggestive of another diagnosis, including: • CT features: ◦ Cysts ◦ Marked mosaic attenuation ◦ Predominant GGO ◦ Profuse micronodules ◦ Centrilobular nodules ◦ Nodules ◦ Consolidation

Am J Respir Crit Care Med Vol 198, Iss 5, pp e44–e68, Sep 1, 2018

HRCT Criteria for UIP pattern

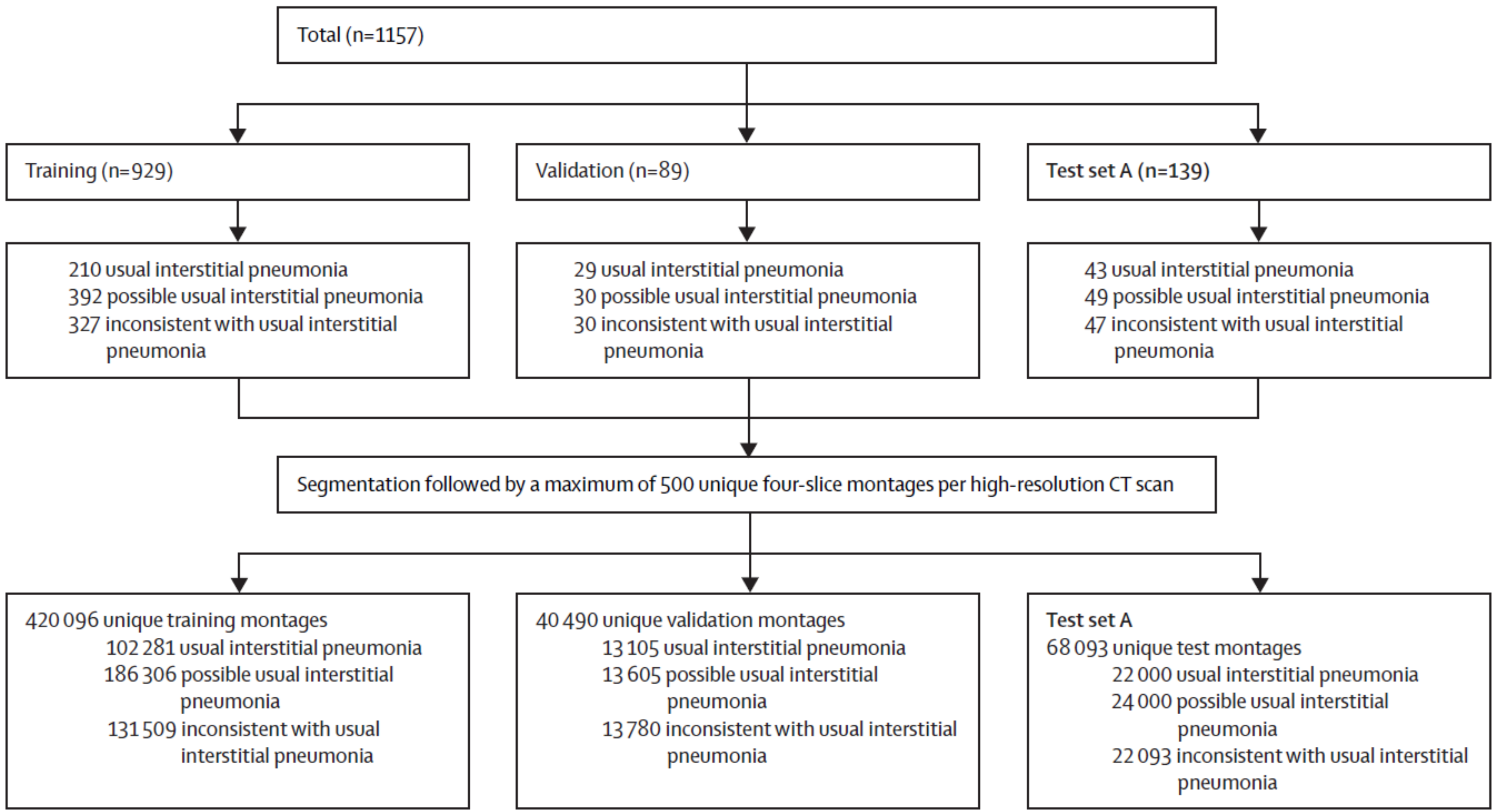
UIP Pattern (All Four Features)	Possible UIP Pattern (All Three Features)	Inconsistent with UIP Pattern (Any of the Seven Features)
• Subpleural, basal predominance • Reticular abnormality • Honeycombing with or without traction bronchiectasis • Absence of features listed as inconsistent with UIP pattern (*see* third column)	• Subpleural, basal predominance • Reticular abnormality • Absence of features listed as inconsistent with UIP pattern (*see* third column)	• Upper or mid-lung predominance • Peribronchovascular predominance • Extensive ground glass abnormality (extent > reticular abnormality) • Profuse micronodules (bilateral, predominantly upper lobes) • Discrete cysts (multiple, bilateral, away from areas of honeycombing) • Diffuse mosaic attenuation/air-trapping (bilateral, in three or more lobes) • Consolidation in bronchopulmonary segment(s)/lobe(s)

Definition of abbreviation: UIP = usual interstitial pneumonia.

Am J Respir Crit Care Med Vol 183. pp 788–824, **2011**

Deep learning for classifying fibrotic lung disease on high-resolution computed tomography: a case-cohort study

Simon L F Walsh, Lucio Calandriello, Mario Silva, Nicola Sverzellati



Lancet Respir Med 2018; 6: 837–45

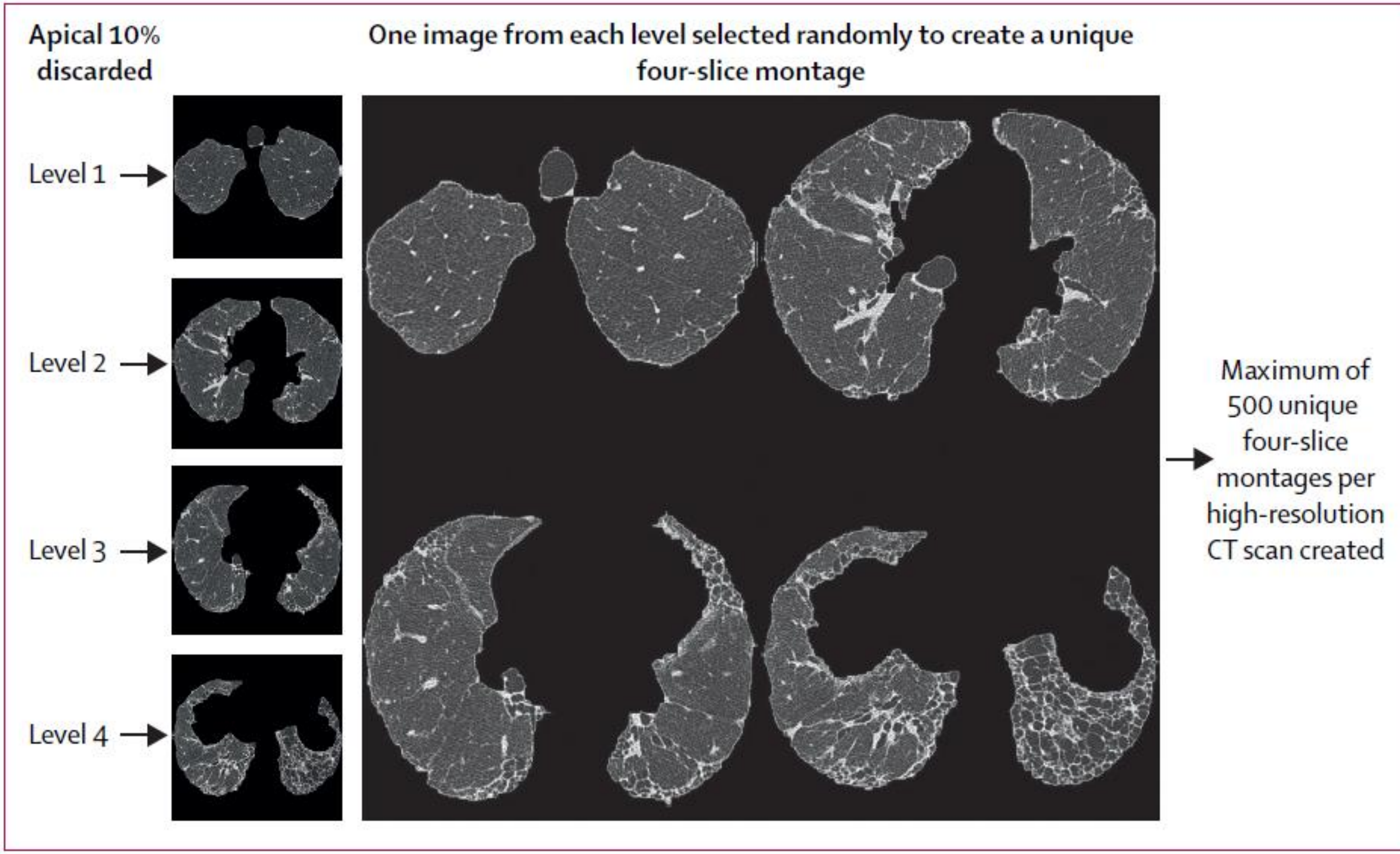


Figure 2: **High-resolution CT preprocessing**

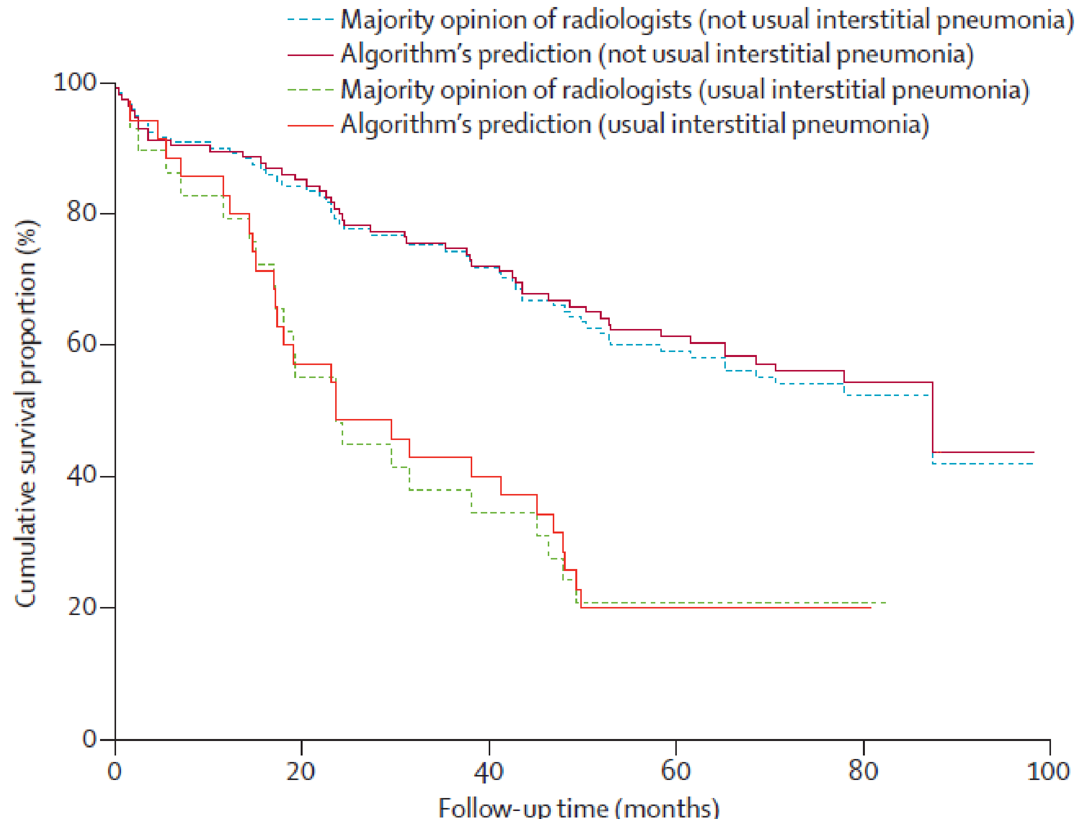
For each high-resolution CT, the lungs were segmented and four axial slice montages were created by randomly selecting a slice from each lung quarter length (excluding the apical 10%). The resampling procedure was programmed to ensure that all montages were unique. A maximum of 500 montages per high-resolution CT were created, generating a total database of 420 096 training montages, 40 490 validation montages, and 68 093 test montages.

Lancet Respir Med 2018; 6: 837–45

		Algorithm			
		Usual interstitial pneumonia	Possible usual interstitial pneumonia	Inconsistent with usual interstitial pneumonia	Total
Majority opinion	Usual interstitial pneumonia	23	5	1	29
	Possible usual interstitial pneumonia	6	10	14	30
	Inconsistent with usual interstitial pneumonia	6	8	77	91
	Total	35	23	92	150

Figure 3: Confusion matrix showing the frequency of deep learning algorithm predictions with respect to the majority opinion of 91 thoracic radiologists on test set B

Lancet Respir Med 2018; 6: 837–45



Number at risk

	0	20	40	60	80	100
Majority opinion of radiologists (not usual interstitial pneumonia)	121	102	87	64	30	0
Algorithm's prediction (not usual interstitial pneumonia)	115	98	83	63	31	0
Majority opinion of radiologists (usual interstitial pneumonia)	29	16	10	5	2	0
Algorithm's prediction (usual interstitial pneumonia)	35	20	14	6	1	0

Figure 4: Kaplan-Meier graphs
Survival differences between high-resolution CTs assigned a radiological diagnosis of usual interstitial pneumonia versus not usual interstitial pneumonia by the algorithm and on the basis of the majority opinion of the thoraci radiologists.

Lancet Respir Med 2018; 6: 837–45

Findings The accuracy of the algorithm on test set A was 76·4%, with 92·7% of diagnoses within one category. The algorithm took 2·31 s to evaluate 150 four slice montages (each montage representing a single case from test set B). The median accuracy of the thoracic radiologists on test set B was 70·7% (IQR 65·3–74·7), and the accuracy of the algorithm was 73·3% (93·3% were within one category), outperforming 60 (66%) of 91 thoracic radiologists. Median interobserver agreement between each of the thoracic radiologists and the radiologist's majority opinion was good (κw=0·67 [IQR 0·58–0·72]). Interobserver agreement between the algorithm and the radiologist's majority opinion was good (κw=0·69), outperforming 56 (62%) of 91 thoracic radiologists. The algorithm provided equally prognostic discrimination between usual interstitial pneumonia and non-usual interstitial pneumonia diagnoses (hazard ratio 2·88, 95% CI 1·79–4·61, p<0·0001) compared with the majority opinion of the thoracic radiologists (2·74, 1·67–4·48, p<0·0001). For Fleischner Society high-resolution CT criteria for usual interstitial pneumonia, median interobserver agreement between the radiologists was moderate (κw=0·56 [IQR 0·55–0·58]), but was good between the algorithm and the radiologists (κw=0·64 [0·55–0·72]).

Interpretation High-resolution CT evaluation by a deep learning algorithm might provide low-cost, reproducible, near-instantaneous classification of fibrotic lung disease with human-level accuracy. These methods could be of benefit to centres at which thoracic imaging expertise is scarce, as well as for stratification of patients in clinical trials.

Funding None.

In conclusion, we have developed a deep learning algorithm with human-level performance for classifying fibrotic lung disease on high-resolution CT on the basis of criteria specified by two international idiopathic pulmonary fibrosis guideline statements. In principle, this algorithm could be deployed anywhere in the world for a low cost and could provide radiological decision support to centres where thoracic imaging expertise is unavailable. Our results warrant further validation in future studies.

Lancet Respir Med 2018; 6: 837–45

AI in chest imaging: CXR

Reference	Application	Remarks
Lo et al. (1995)	Nodule detection	Classifies candidates from small patches with two-layer CNN, each with 12 5 × 5 filters
Anavi et al. (2015)	Image retrieval	Combines classical features with those from pre-trained CNN for image retrieval using SVM
Bar et al. (2015)	Pathology detection	Features from a pre-trained CNN and low level features are used to detect various diseases
Anavi et al. (2016)	Image retrieval	Continuation of Anavi et al. (2015), adding age and gender as features
Bar et al. (2016)	Pathology detection	Continuation of Bar et al. (2015), more experiments and adding feature selection
Cicero et al. (2017)	Pathology detection	GoogLeNet CNN detects five common abnormalities, trained and validated on a large data set
Hwang et al. (2016)	Tuberculosis detection	Processes entire radiographs with a pre-trained fine-tuned network with 6 convolution layers
Kim and Hwang (2016)	Tuberculosis detection	MIL framework produces heat map of suspicious regions via deconvolution
Shin et al. (2016a)	Pathology detection	CNN detects 17 diseases, large data set (7k images), recurrent networks produce short captions
Rajkomar et al. (2017)	Frontal/lateral classification	Pre-trained CNN performs frontal/lateral classification task
Yang et al. (2016c)	Bone suppression	Cascade of CNNs at increasing resolution learns bone images from gradients of radiographs
Wang et al. (2016a)	Nodule classification	Combines classical features with CNN features from pre-trained ImageNet CNN

A survey on deep learning in medical image analysis
Medical Image Analysis, 2018

AI in Critical Care Medicine

Artificial intelligence applications in the intensive care unit

C. William Hanson III, MD, FCCM; Bryan E. Marshall, MD, FRCP, FRCA

Objective: To review the history and current applications of artificial intelligence in the intensive care unit.

Data Sources: The MEDLINE database, bibliographies of selected articles, and current texts on the subject.

Study Selection: The studies that were selected for review used artificial intelligence tools for a variety of intensive care applications, including direct patient care and retrospective database analysis.

Data Extraction: All literature relevant to the topic was reviewed.

Data Synthesis: Although some of the earliest artificial intelligence (AI) applications were medically oriented, AI has not been widely accepted in medicine. Despite this, patient demographic, clinical, and billing data are increasingly available in an electronic format and therefore susceptible to analysis by intelligent software. Individual AI tools are specifically suited to different tasks, such as waveform analysis or device control.

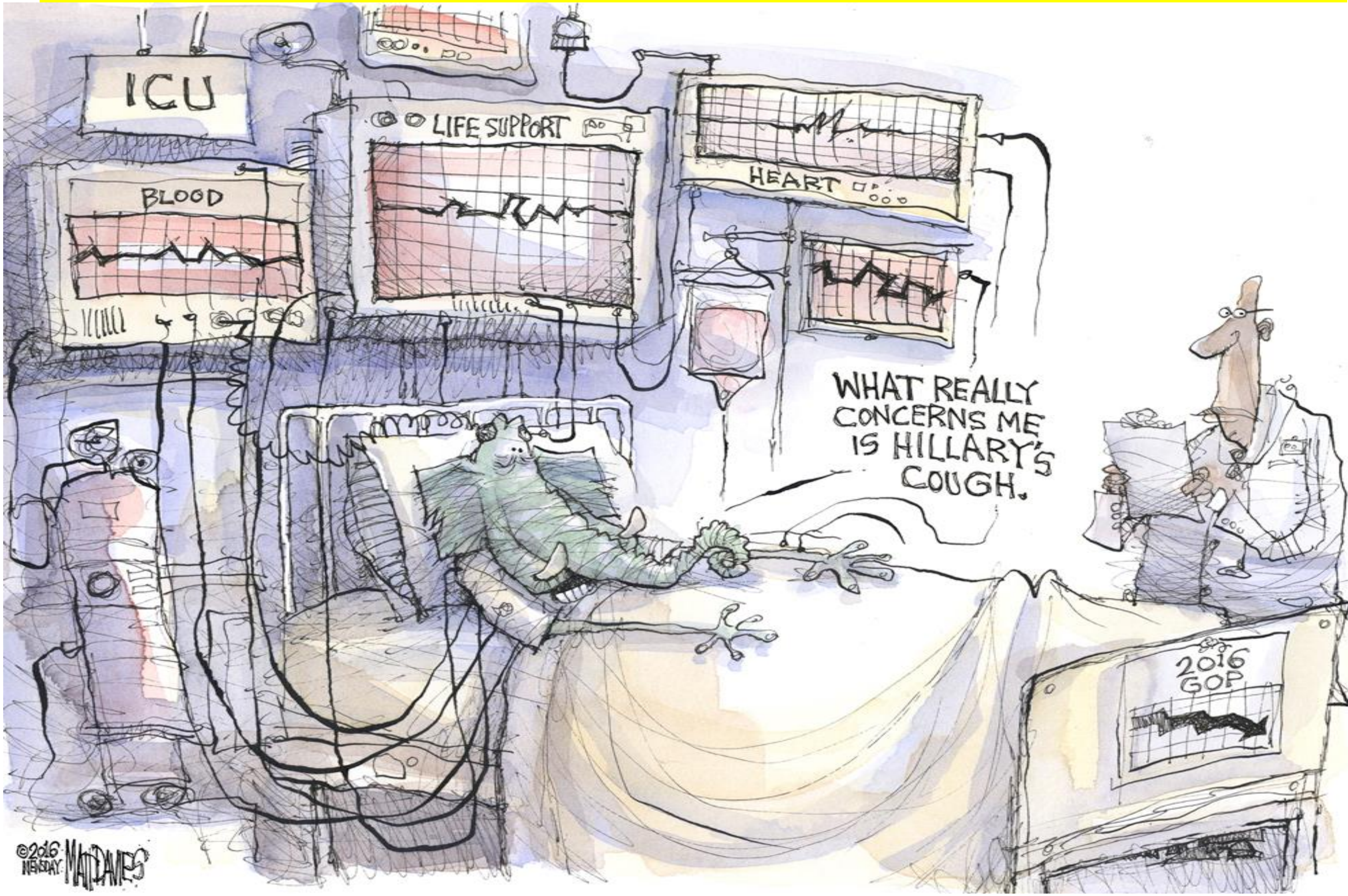
Conclusions: The intensive care environment is particularly suited to the implementation of AI tools because of the wealth of available data and the inherent opportunities for increased efficiency in inpatient care. A variety of new AI tools have become available in recent years that can function as intelligent assistants to clinicians, constantly monitoring electronic data streams for important trends, or adjusting the settings of bedside devices. The integration of these tools into the intensive care unit can be expected to reduce costs and improve patient outcomes. (Crit Care Med 2001; 29:427–435)

Key Words: intensive care unit; artificial intelligence; expert systems; computer-assisted diagnosis; computer-assisted therapy; decision support techniques; neural networks; algorithms; fuzzy logic; data display; computer simulation; clinical decision support systems; management decision support systems

Data driven decision support tools will permit the busy clinician (physician, nurse, RT) to function more efficiently, caring for more patients more safely in much the same way that these same tools have been used to enhance the efficiency of business applications.

Crit Care Med 2001 Vol. 29, No. 2

ICU


ICU
LIFE SUPPORT
BLOOD
HEART
WHAT REALLY CONCERNS ME IS HILLARY'S COUGH.
2016 GOP
©2016 NEWSDAY MATT DAVIES

Big Data and Data Science in Critical Care

L. Nelson Sanchez-Pinto, MD; Yuan Luo, PhD; and Matthew M. Churpek, MD, PhD

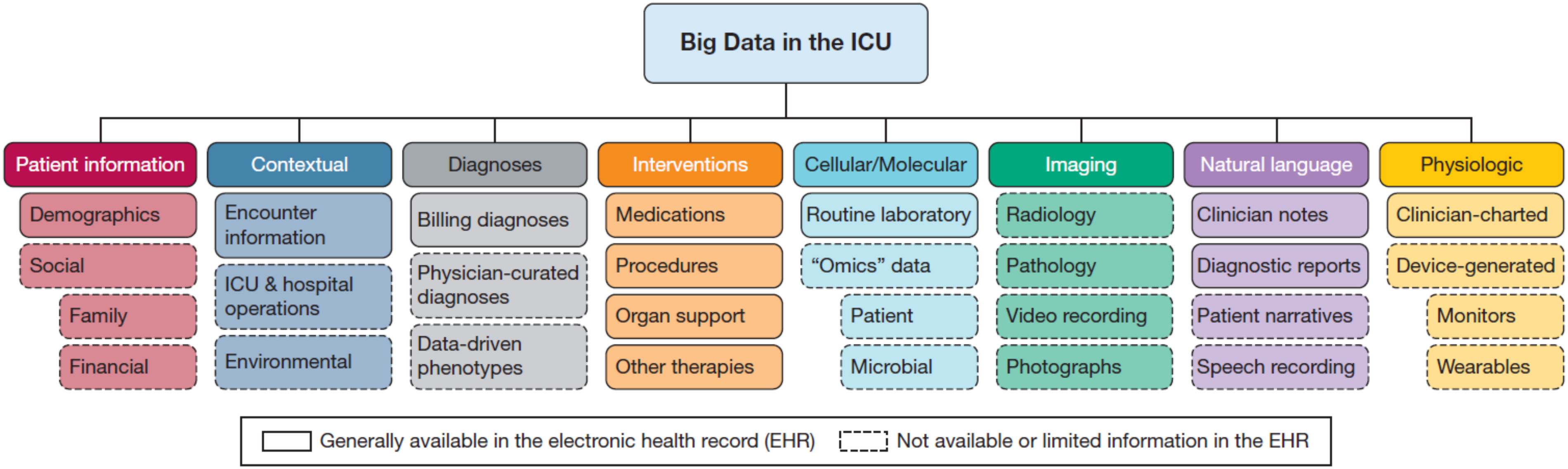
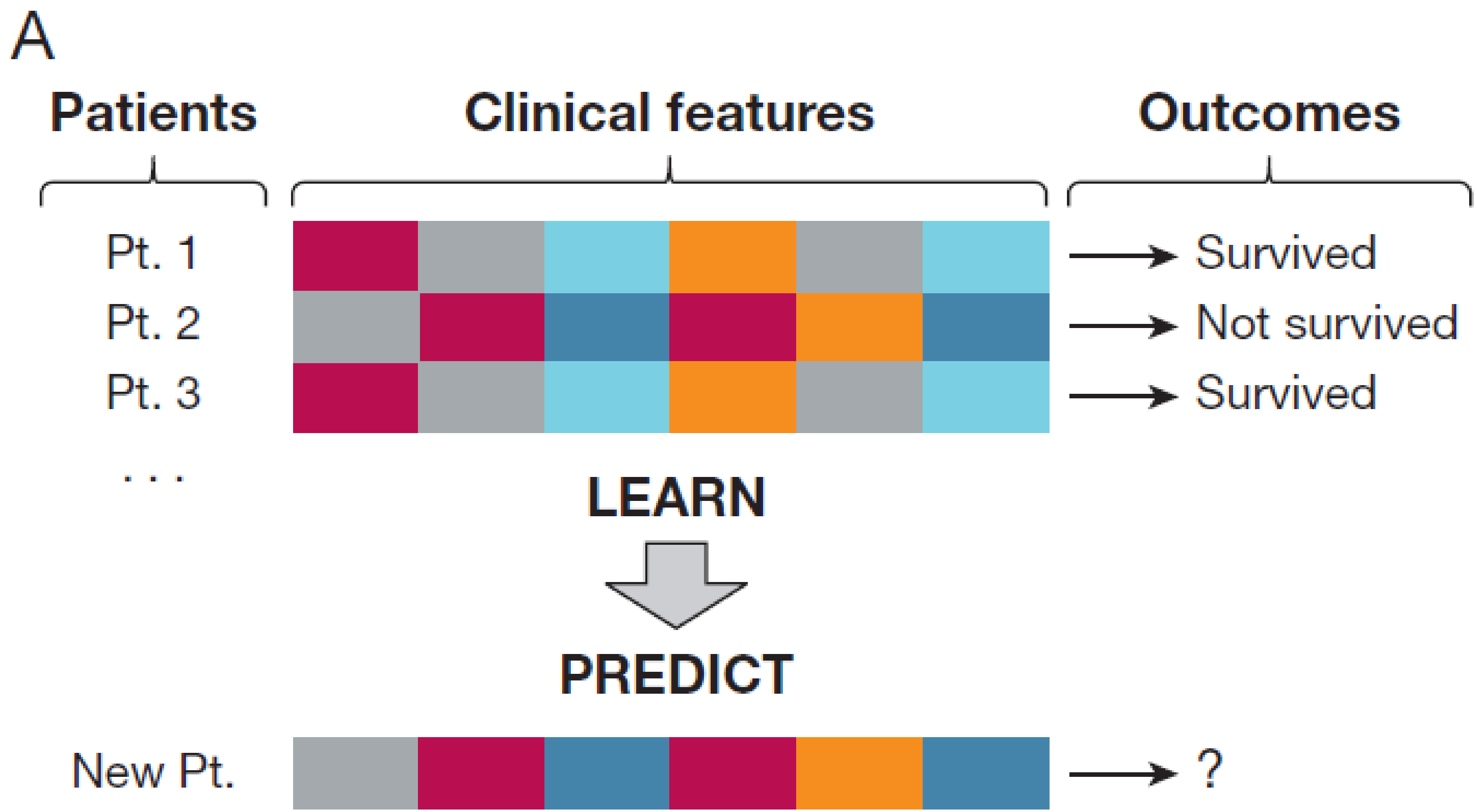


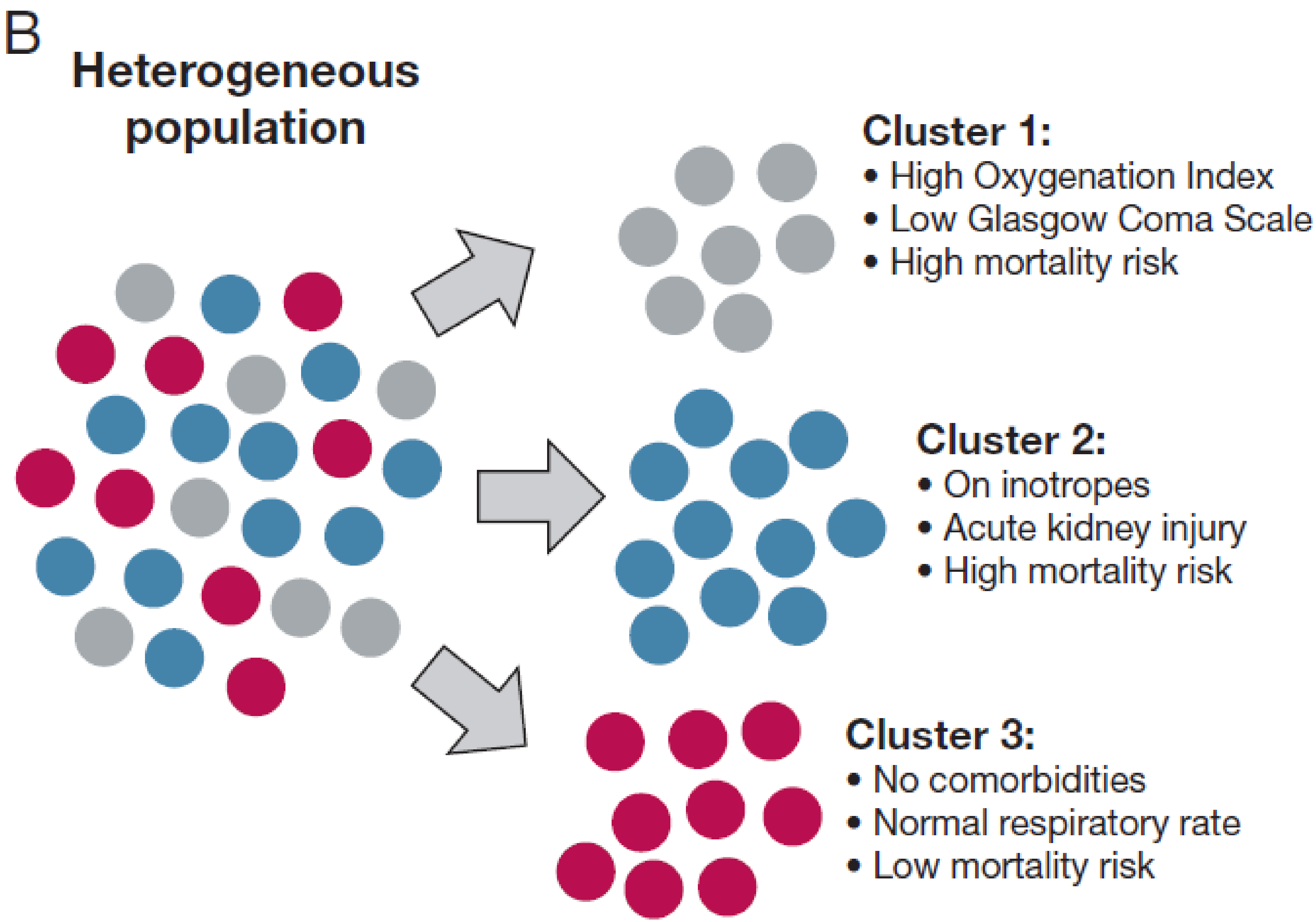
Figure 1 – *Some of the major sources of big data in the ICU. The term "omics" refers to the data derived from modern molecular techniques (eg, genomics, transcriptomics, proteomics, metabolomics, microbiomics). EHR = electronic health record.*

CHEST 2018; 154(5):1239-1248



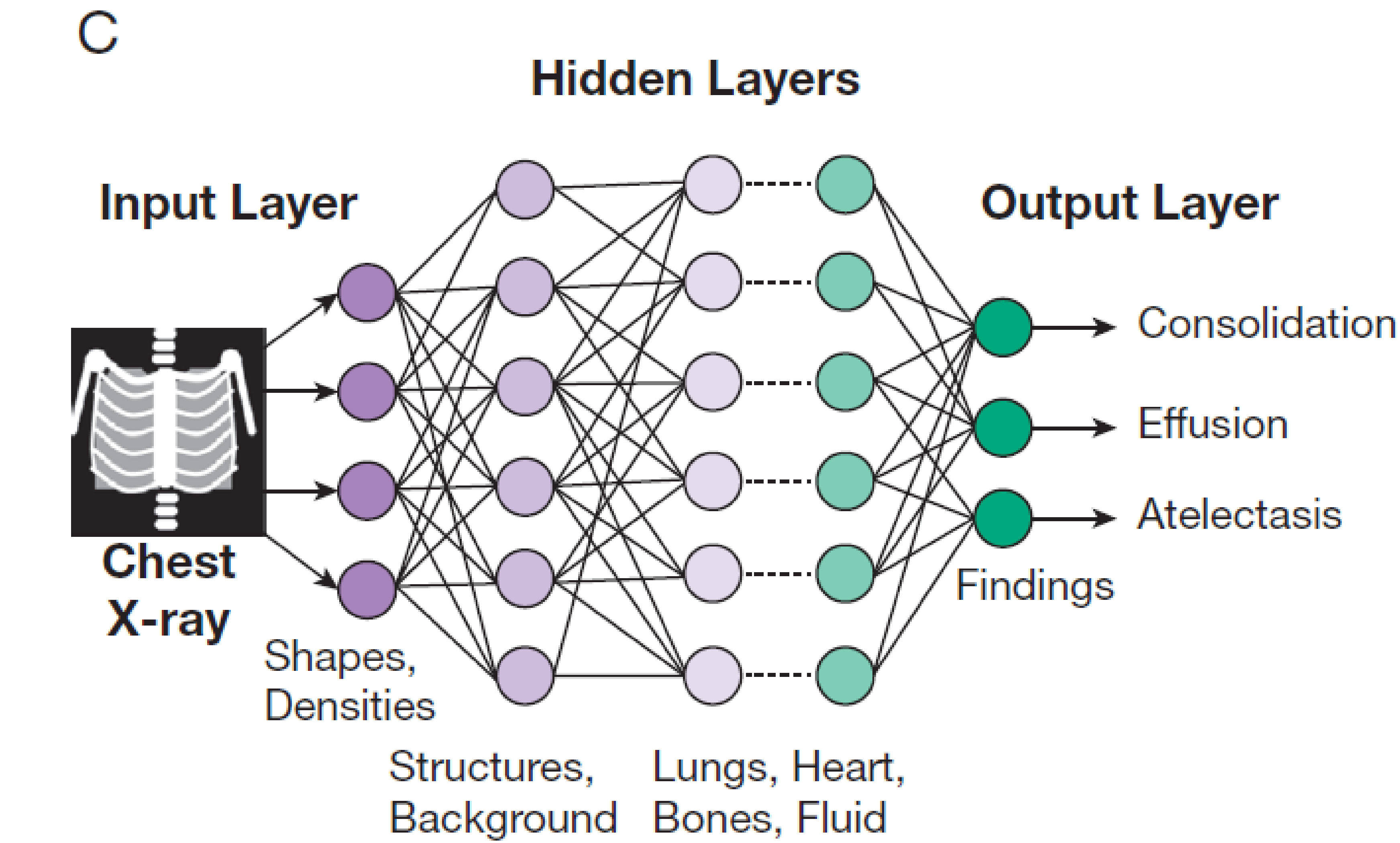
A, Supervised learning algorithms can be used, for example, to uncover the relationship between patient clinical features (eg, labora-tory tests and vital signs) and mortality to predict the outcome in future cases

CHEST 2018; 154(5):1239-1248



B, Unsupervised learning algorithms can be used to uncover naturally occurring groupings or clusters of patients based on their clinical characteristics, without targeting a specific outcome.

CHEST 2018; 154(5):1239-1248



C, Deep learning algorithms can be used, for example, to extract meaningful features from imaging data (eg, chest radiograph) to represent information in an increasingly higher order of hierarchical complexity and be able to make predictions, such as the presence of pathologic findings

CHEST 2018; 154(5):1239-1248

The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care

Matthieu Komorowski[1,2,3], Leo A. Celi[3,4], Omar Badawi[3,5,6], Anthony C. Gordon[1*] and A. Aldo Faisal[2,7,8,9*]

- MIMIC-III: 17,083 admissions (88.4% of eligible patients with sepsis) from five separate ICUs in one tertiary teaching hospital
- eRI database:79,073 admissions (73.6% of eligible patients with sepsis) from 128 different hospitals
- The total volume of **intravenous fluids** and maximum dose of **vasopressors** administered over each 4-h period defined the medical treatments of interest.
- The model aims at optimizing patient mortality, so a reward was associated to survival and a penalty to death.

Komorowski eta l, Nature Medicine, 2018

- an openly available critical care database
- MIT Lab for Computational Physiology, comprising deidentified health data associated with ~40,000 critical care patients.
- demographics, vital signs, laboratory tests, medications, and more.
- MIMIC-III
- 58,000 hospital admissions for 38,645 adults and 7,875 neonates.
- The data spans June 2001 - October 2012

https://mimic.physionet.org/

Table 1 | Description of the datasets

	MIMIC-III	eRI
Unique ICUs (*n*)	5	128
Unique ICU admissions (*n*)	17,083	79,073
Characteristics of hospitals, per number of ICU admissions	Teaching tertiary hospital	Nonteaching: 37,146 (47.0%) Teaching: 29,388 (37.2%) Unknown: 12,539 (15.9%)
Age, years (mean (s.d.))	64.4 (16.9)	65.0 (16.7)
Male gender (*n* (%))	9,604 (56.2%)	40,949 (51.8%)
Premorbid status (*n* (%))		
Hypertension	9,384 (54.9%)	43,365 (54.8%)
Diabetes	4,902 (28.7%)	25,290 (32.0%)
CHF	5,206 (30.5%)	15,023 (19.0%)
Cancer	1,803 (10.5%)	11,807 (14.9%)
COPD or RLD	4,248 (28.7%)	18,406 (23.3%)
CKD	3,087(18.1%)	14,553 (18.4%)
Primary ICD-9 diagnosis (*n* (%))		
Sepsis, including pneumonia	5,824 (34.1%)	41,396 (52.3%)
Cardiovascular	5,270 (30.8%)	11,221 (14.2%)
Respiratory	1,798 (10.5%)	9,127 (11.5%)
Neurological	1,590 (9.3%)	7,127 (9.0%)
Renal	429 (2.5%)	1,454 (1.8%)
Others	2,172 (12.7%)	8,747 (11.1%)
Initial OASIS (mean (s.d.))	33.5 (8.8)	34.8 (12.4)
Initial SOFA (mean (s.d.))	7.2 (3.2)	6.4 (3.5)
Procedures during the 72 h of data collection:		
Mechanical ventilation (*n* (%))	9,362 (54.8%)	39,115 (49.5%)
Vasopressors (*n* (%))	6,023 (35.3%)	23,877 (30.2%)
Renal replacement therapy (*n* (%))	1,488 (8.7%)	6,071 (7.7%)
Length of stay, days (median, (IQR))	3.1 (1.8–7)	2.9 (1.7–5.6)
ICU mortality	7.4%	9.8%
Hospital mortality	11.3%	16.4%
90-d mortality	18.9%	Not available

CHF, congestive heart failure; CKD, chronic kidney disease; COPD, chronic obstructive pulmonary disease; ICD-9, International Classification of Diseases version 9; IQR, interquartile range; OASIS, Oxford Acute Severity of Illness Score; RLD, restrictive lung disease; SOFA, sequential organ failure assessment.

Komorowski eta l, Nature Medicine, 2018

Data flow of the AI clinician

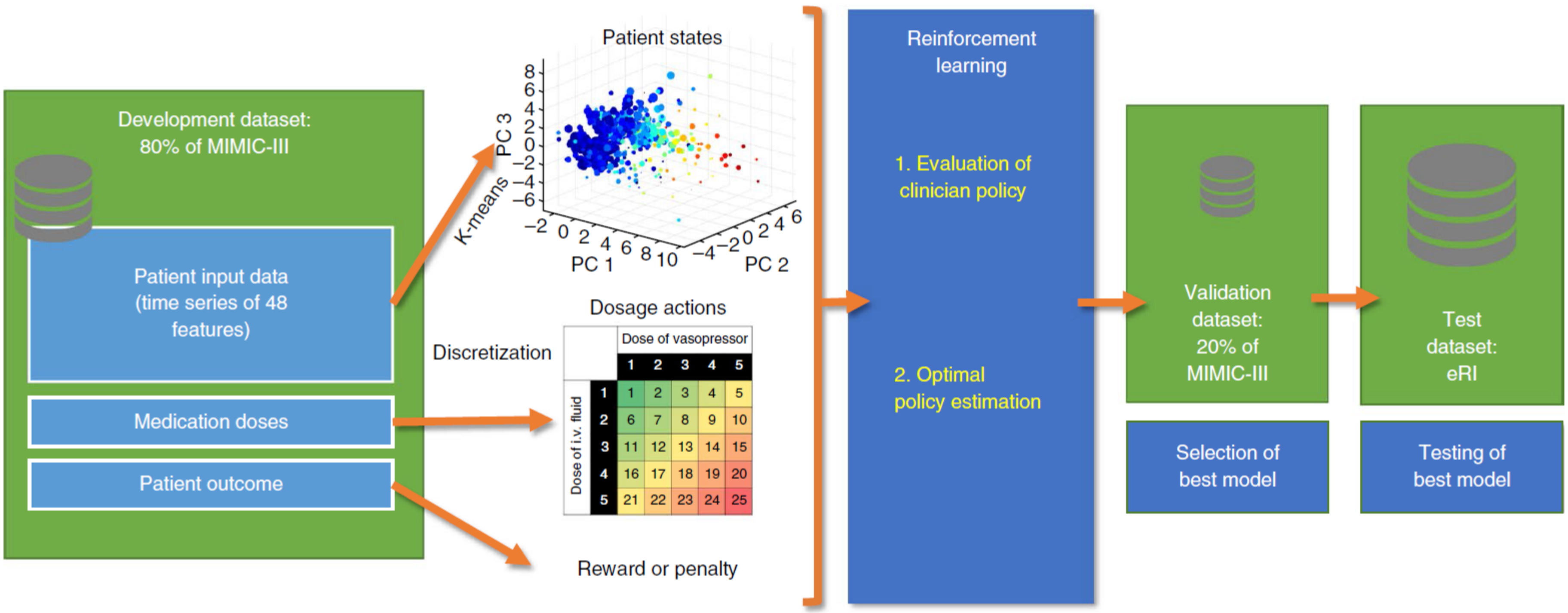


Fig. 1 | Data flow of the AI Clinician. Eighty percent of the MIMIC-III dataset was used to define the elements of the MDP. Time series of patient data were clustered into finite states. The dose of intravenous (i.v.) fluids and vasopressors were discretized into 25 possible actions. Patient survival at 90 d after ICU admission defined reward. Reinforcement learning was used to estimate optimal treatment strategies—the AI policy. The remaining 20% of MIMIC-III data was used to identify the best model among 500 candidates, which was then tested on an independent dataset from the eRI database.

Komorowski eta l, Nature Medicine, 2018

Selection of the best AI policy and model calibration

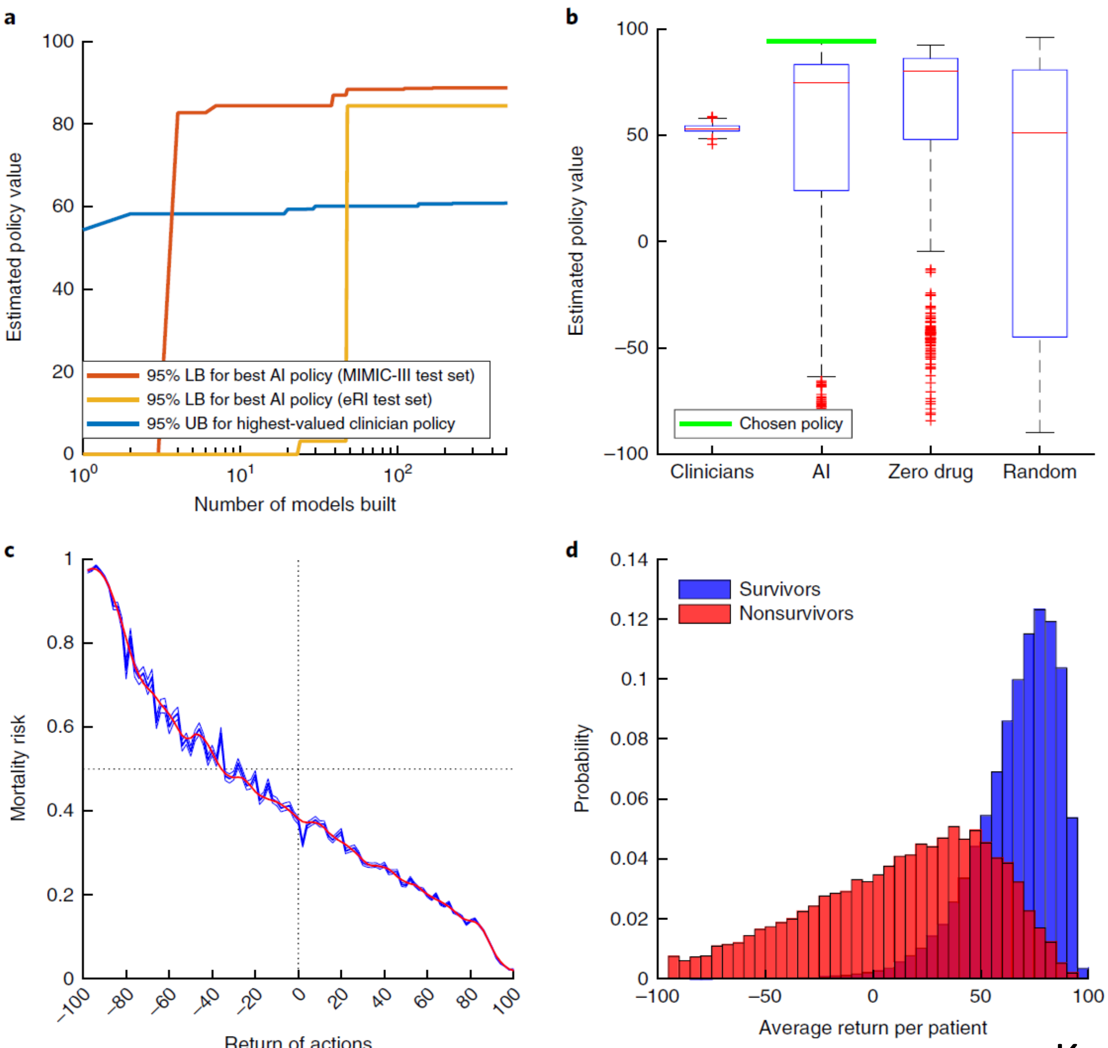


Fig. 2 | Selection of the best AI policy and model calibration. a, Evolution of the 95% lower bound (LB) of the best AI policy and of the 95% upper bound (UB) of highest-valued clinician policy during the building of 500 models. After only a few models, a higher value for the AI policy than the clinician treatment, within the accepted risk, is guaranteed. $n=13{,}666$ patients in the MIMIC-III development dataset, $n=3{,}417$ in the MIMIC-III test set and $n=79{,}073$ in the eRI set. **b**, Distribution of the estimated value of the clinicians' actual treatments, the AI policy, a random policy and a zero-drug policy across the 500 models in the MIMIC-III test set ($n=500$ models in each boxplot). The chosen AI policy maximizes the 95% confidence lower bound. On each boxplot, the central line indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to 1.5 times the interquartile range. Points beyond the whiskers are considered outliers and are plotted individually using the + symbol. **c**, The relationship between the return of clinicians' treatments and patient 90-d mortality in the MIMIC-III training set ($n=13{,}666$ patients). Return of actions were sorted into 100 bins, and the mean observed mortality (blue line for raw, red line for smoothed) was computed in each of these bins. The shaded blue area represents the s.e.m. Treatments with a low return were associated with a high risk of mortality, whereas treatments with a high return led to better survival rates. **d**, Average return in survivors ($n=11{,}031$) and nonsurvivors ($n=2{,}635$) in the MIMIC-III training set. **c** and **d** were generated by bootstrapping in the training data with 2,000 resamplings.

Komorowski eta l, Nature Medicine, 2018

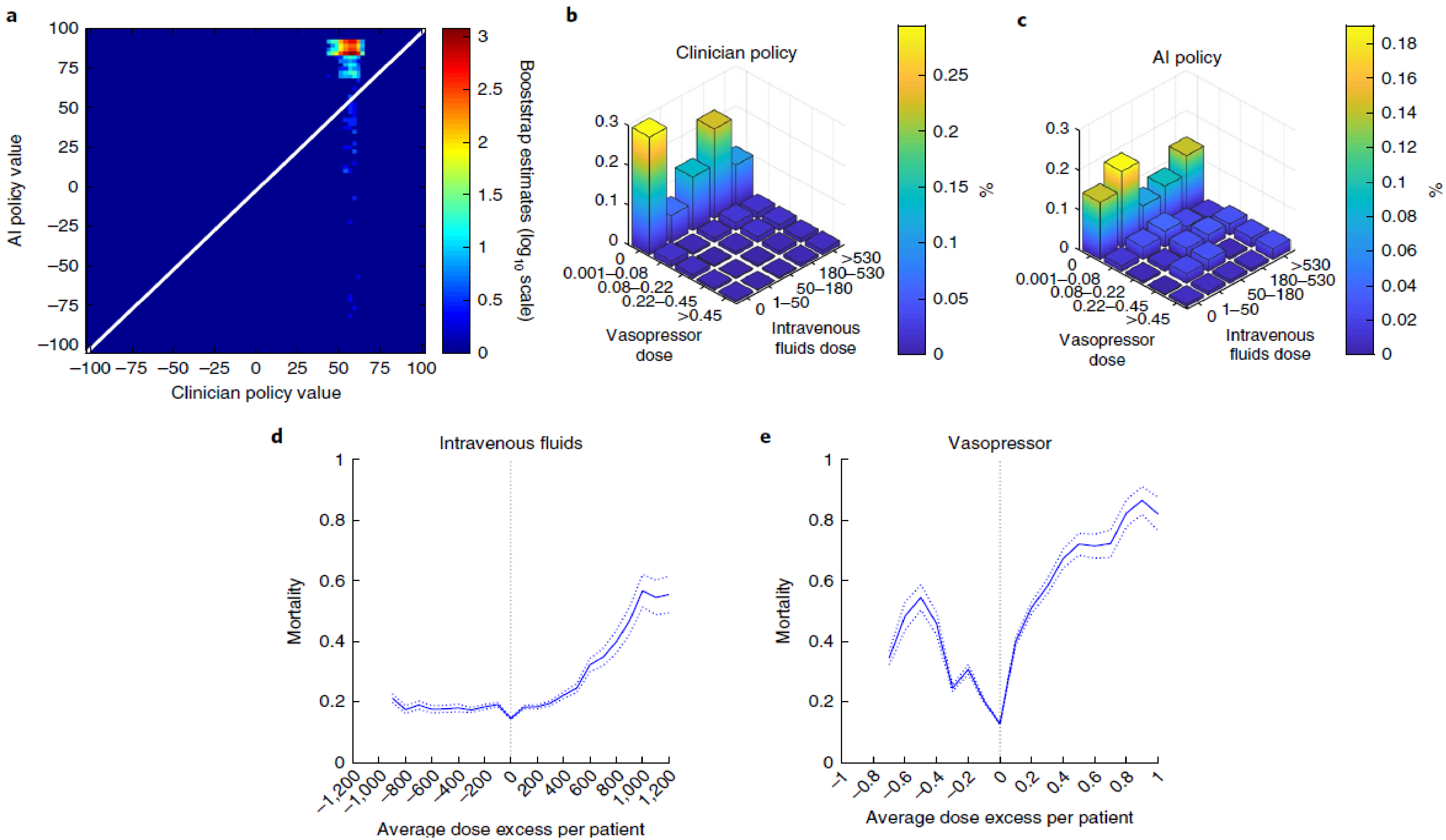
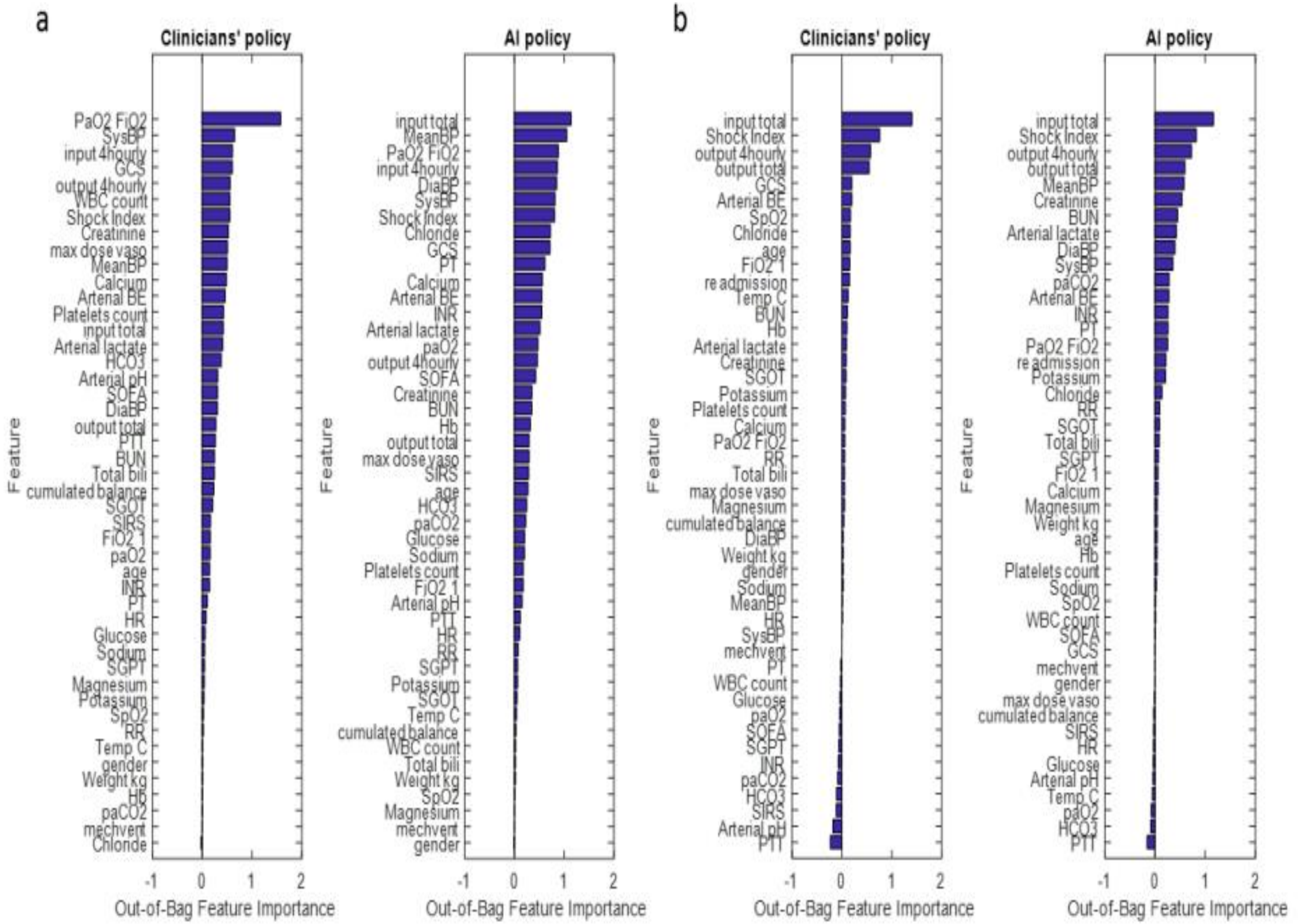


Fig. 3 | Comparison of clinician and AI policies in eRI and average dose excess received per patient of both drugs in eRI with corresponding mortality. a, Distribution of the estimated value of the clinician and AI policies in the selected model, built by bootstrapping with 2,000 resamplings. **b,c**, Visualization of the clinician and AI policies. All actions were aggregated over all time steps for the five dose bins of both medications. On average, patients were administered more intravenous fluid (**b**) and less vasopressor (**c**) medications than recommended by the AI policy. Vasopressor dosage is shown in µg/kg/min of norepinephrine equivalent, and intravenous fluid dosage is shown in mL/4 h. **d**, The dose excess, referring to the difference between the given and suggested dose averaged over all time points per patient, for intravenous fluids (left) and vasopressor (right). The figure was generated by bootstrapping with 2,000 resamplings. The shaded area represents the s.e.m. In both plots, the smallest dose difference was associated with the best survival rates (vertical dotted line). The further away the dose received was from the suggested dose, the worse the outcome.

Komorowski eta l, Nature Medicine, 2018

: Feature importance in the clinicians' policy and the AI policy



Komorowski eta l, Nature Medicine, 2018

Summary

- Administering more or less of either treatment than the AI policy was associated with increasing mortality rates in a dose-dependent fashion
- Avoiding targeting short-term resuscitation goals and instead following trajectories toward longer-term survival
- Aim: real-time AI clinician
- Require prospective evaluation using real-time data and decision-making in clinical trials and also testing in different healthcare settings

Komorowski eta l, Nature Medicine, 2018

AI in (Critical Care) Medicine 的醒思

理論上，不能容許出錯

一旦出錯，可能會要人命

Self-driving Uber kills Arizona woman in first fatal crash involving pedestrian

Tempe police said car was in autonomous mode at the time of the crash and that the vehicle hit a woman who later died at a hospital

2 Dead in 5-Car Crash in Bartlett

By Christian Farr

Published at 9:25 PM CDT on Sep 25, 2018 | Updated at 10:22 PM CDT on Sep 25, 2018

Two people died after five vehicles collided in the northwest suburbs Tuesday morning. NBC 5's Christian Farr reports.
(Published Tuesday, Sept. 25, 2018)

常見 VS 罕見

饅頭 vs. 香菇

Association vs. Casualization

- ARDS的死亡率，在太陽黑子發現的那一天死亡率特別高！

機器人視訊通知病

批醫院沒人性

TVBS 新聞網 2019年3月10日 下午 05:44

美國加州的凱薩醫療中心，醫生透過遠
的病患，他只剩下幾天可活。藉由機器
出面和病人溝通，讓家屬十分不滿，痛

老先生呼吸困難，住進了加護病房，進
機器人。病患孫女：「護士來說醫生要
在這裡。」但螢幕上的醫生，開口就是
治療，但也能停止，更加專注在安寧緩
已經完全衰竭，無法治療，最多只能再
一步是回家臨終安養。對吧？我不確定
塔納身體不適，週日緊急就醫，但在加
端通訊機器人，告訴當時只有孫女陪在

病患孫女：「我開始錄影時，沒想到事
昆塔納耳朵不好，聽不清楚機器人上醫
醫生的話給他聽。病患孫女：「我們知道
但我覺得沒有人，應該這樣聽到這種消

兩天後昆塔納過世，家屬十分不滿，覺
沒人性；院方則向家屬致哀，表示讓家
檢討並改善使用醫療機器人的相關政策

透過機器人視訊宣布

人的視訊畫面，告訴 78 歲病患坤塔
內會過世，引發病患家屬憤怒不滿。
很遺憾，強調未來會更重視病患體驗。
困難，被救護車送往費利蒙市凱薩醫
位護士告訴坤塔納的孫女安娜麗莎·
）醫生等一下會來巡房，沒多久一台
在視訊螢幕。當時病房內只有阿公和
口有這台機器人，醫生看起來像坐在
生表示：「我拿到這些核磁共振成像
肺所剩無幾，我們無能為力。」維哈
我試著不哭，我努力不尖叫出來……
的陪伴下，得知人生最糟的消息。」
的話給阿公聽，因為阿公左耳重聽，
。維哈姆告訴阿公：「醫生在說，也
關懷。對吧？」醫生則回應：「我不
麼到院時向院方抱怨醫生告知病情的
是我們的政策，我們就是這樣辦事。」
世。病患家屬的友人在社群媒體上抱
值與同理心的方式；醫生告訴維哈姆，
點滴直到過世。該友人表示，醫療科
與地點必須黑白分明。南阿拉米達郡
蜜雪兒·加斯基爾-哈米斯（Michelle
情況非常罕見，未能達到病患期待深
曾探視過，後來才有視訊電話，視訊
初步對話，初診時也不會使用。

HAI: Human-centered AI

Stanford
Human-Centered
Artificial Intelligence

THE AGE OF
ARTIFICIAL INTELLIGENCE

@StanfordHAI


UBER
VS.
TAXI

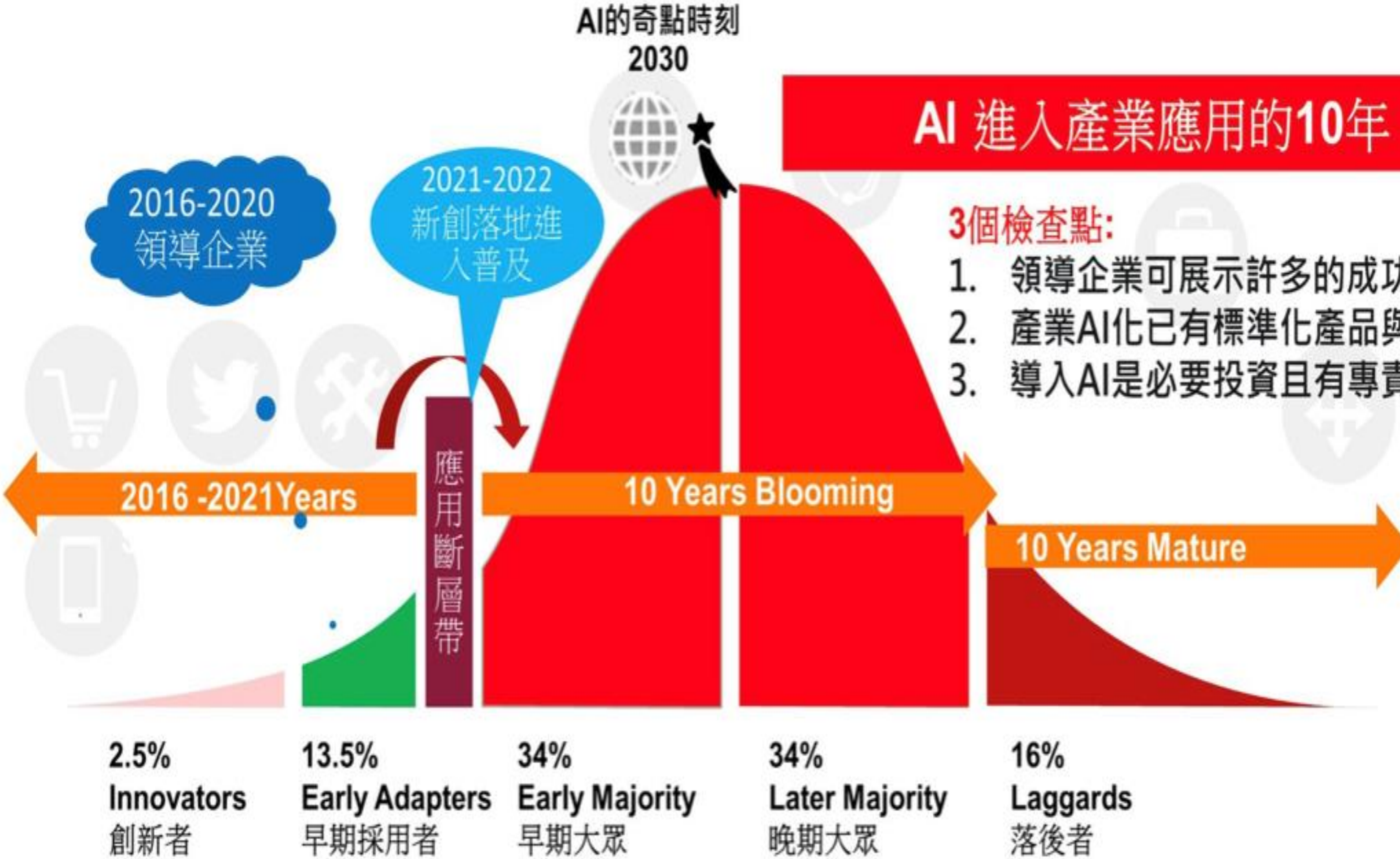
長榮空服第3度勞資協商破裂 工會不排除罷工

Yahoo奇摩（即時新聞） | 23.8k 人追蹤 追蹤

2019年4月17日 下午 04:57

長榮航空跟空服員的勞資爭議越演越烈，今天第3度協商仍破局收場，桃園市空服員職業工會秘書長鄭雅菱說，周五前會召開會員大會，決定是否要進行罷工投票。

產業AI化應用的黃金10年



3個檢查點:

1. 領導企業可展示許多的成功效益
2. 產業AI化已有標準化產品與服務
3. 導入AI是必要投資且有專責部門

Slide credit : Richie Tsai at Taiwan AI Academy

Summary: 個人意見

- AI: Business issue, 創投
- 鳥事給AI做，人事給人做
- 不要為了AI而AI
- 從自己的專長發想如何藉由AI來精進
- 以病人為中心為出發點

Thank you for your attention !

AI研究的條件

- 本院的強項
  - 大量的高品質資料 （Data)
  - 特定領域的知識 (Domain knowledge)
- 本院的弱項 （需建立、已有成熟產品）
  - 龐大的運算能力 (Computing capacity)
  - 適當的運算模式（Modeling)
- 真正的困難
  - 好的PI，需同時具有跨領域能力以及business mind

轉化為黑色素瘤或皮膚癌的風險。

只要輕鬆自拍就能進行初篩，除了可以節省看病時間，也能保有部分個人隱私，達到疾病預防效果。

（圖片提供：臺北醫學大學）

MoleMate 痣拍幫

LINE 聊天機器人

MoleMate
痣拍幫

圖3-b HIMMS EMRAM 定義

7級 最高階，醫院已有完整電子病歷系統，且區域醫療資訊可交換及共用，達到一個病人的連續的健康照護。

6級 是全流程醫療資料閉環管理，高級醫療決策支援，完整的病人安全系統。

5級 醫院的統一資料管理，各部門系統資料集成，基本上已建立以電子病歷為基礎的醫院資訊平台，並有完整的醫療影像系統。

4級 全院資料共用，完整的診間醫令系統，並且進階應用於醫療決策支援。

3級 部門間資料交換，並且初步應用於醫療決策支援。

2級 部門內資料交換，具備臨床數據庫。

1級 部門內初步使用電腦做資料登錄，具備檢驗檢查，放射科，藥局等獨立資訊系統。

0級 尚未完成電子病歷系統。

HIMMS「評價醫療機構實施電子病歷水準」分級，從 0 到 7 級，總共分為 8 級。
資料來源：himssanalytics.orgemr_adoption.asp

圖3-a 多納貝帝恩模型（Avedis Donabedian Model）

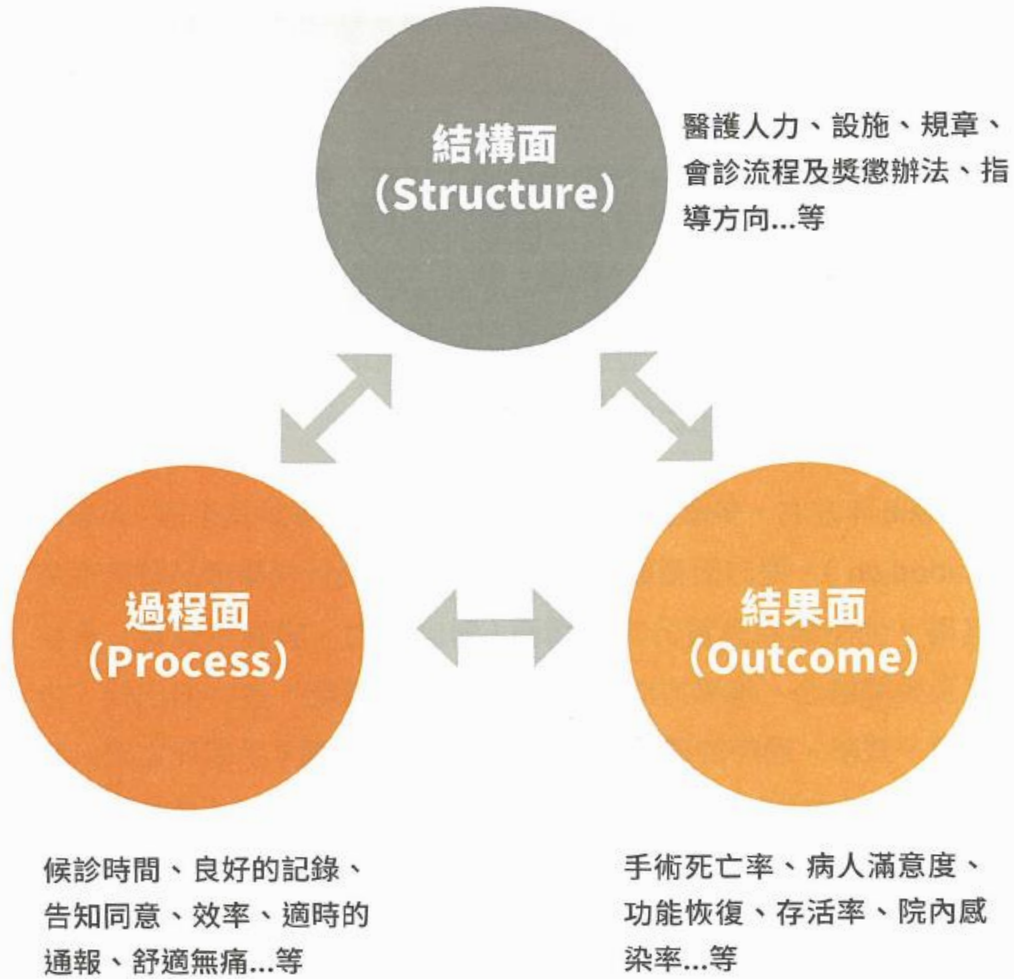
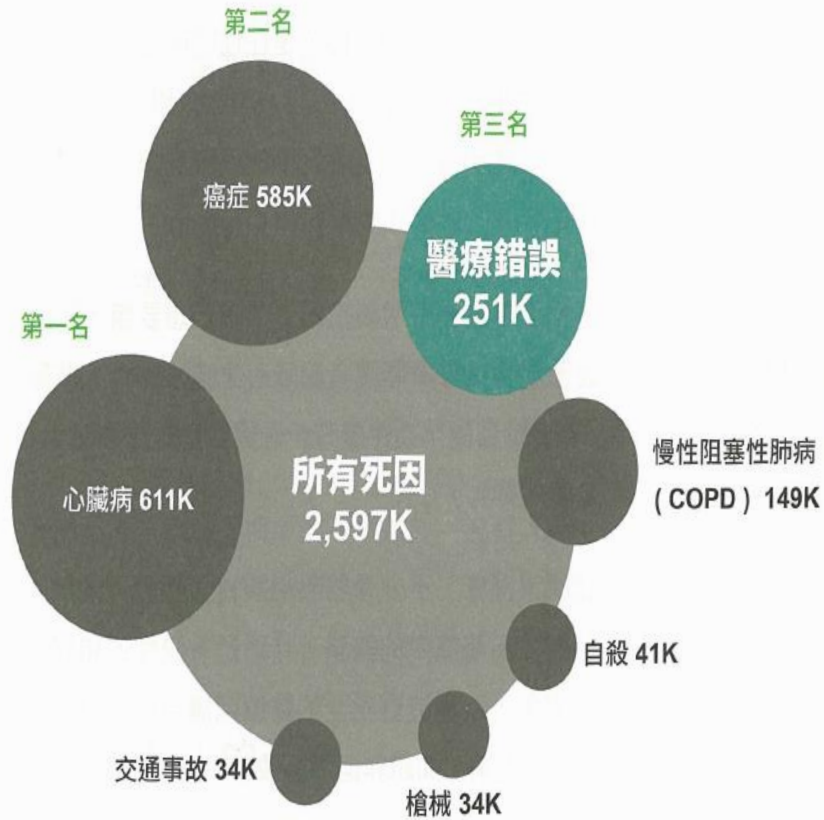


圖2-d 醫療疏失是美國第三大死因 (2016)



註：K=1千人

資料來源：英國醫學期刊（BMJ），2016

圖2-e 智慧病房五大功能

充分知會病人
(Better-informed Patients)

充分知會醫師
(Better-informed Physicians)

充分知會醫院管理人員
(Better-informed Administrators)

強化醫病互動
(Enhanced Patient-Physician Interactions)

偵測、預測、預防不良事件
(Detect, Predict and Prevent Adverse Events)

FDA-approved: AIDOC

- Automatic detection of acute brain hemorrhage
- PACS integration
- FDA approval 15/OCT/2018

Artificial Intelligence (AI)

- Electric Medical Record (EMR)
- Natural Language Processing
- Deep learning; Machine learning

ATS 2018/5

MINI SYMPOSIUM

C15 CRITICAL CARE: BIG DATA AND ARTIFICIAL INTELLIGENCE IN CRITICAL ILLNESS

9:15 a.m. - 11:15 a.m. SAN DIEGO CONVENTION CENTER

Room 29 A-D (Upper Level)

Chairing: M.N. Gong, MS, MD, Bronx, NY
M.W. Sjoding, MD, Ann Arbor, MI
C.R. Cooke, MD, Ann Arbor, MI

Oral Presentations

9:15 **What Comes First in Sepsis- Systemic Inflammatory Response Syndrome (SIRS), the Sequential Organ Failure Assessment (SOFA), or the Quick Sepsis Related Organ Failure Assessment (qSOFA)?**/P.A. Prasad, M.C. Fang, N. Van Groningen, Y. Abe-Jones, C.S. Calfee, M.A. Matthay, K.N. Kangelaris, San Francisco, CA, p.A4441

9:30 **Feasibility of Sepsis Phenotyping Using Electronic Health Record Data During Initial Emergency Department Care**/C.W. Seymour, J. Kennedy, S. Wang, Z. Xu, C.-C.H. Chang, Q. Mi, Y. Vodovotz, G. Clermont, S. Visweswaran, J.C. Weiss, G. Cooper, H Gomez, J.A. Kellum, D.C. Angus, Pittsburgh, PA, p.A4442

9:45 **EMR Clinical Signatures with Unsupervised Topic Modeling in Early Infection and Sepsis**/A. Fohner, J. Chen, J. Greene, P. Kipnis, G. Escobar, V. Liu, Oakland, CA, p.A4443

10:00 **Less Is More: Detecting Clinical Deterioration in the Hospital with Machine Learning Using Only Age, Heart Rate and Respiratory Rate**/D.P. Edelson, K. Carey, C.J. Winslow, M.M. Churpek, Chicago, IL, p.A4444

10:15 **From Pictures to Prediction: Combining Data Visualization with Deep Learning to Predict Clinical Deterioration**/A. Mayampurath, K. Carey, L.-R. Venable, D. Edelson, M.M. Churpek, Chicago, IL, p.A4445

10:30 **Natural Language Processing and Machine Learning for Identification of Acute Respiratory Distress Syndrome**/A. Oakey, D. Dligach, P. Yang, P. Formanek, S. Zelisko, R. Price, C. Joyce, R. Cooper, M. Afshar, Chicago, IL, p.A4446

10:45 **Predicting Intensive Care Unit Readmission with Machine Learning Using Electronic Health Record Data**/J.C. Rojas, K.A. Carey, D.P. Edelson, L.R. Venable, M.D. Howell, M.M. Churpek, Chicago, IL, p.A4447

11:00 **Prediction for Hypotension Episode with Multigranular Data in the Intensive Care Unit**/J.H. Yoon, V. Jeanselme, Y. Chen, A. Dubrawski, M. Hravnak, M. Pinsky, G. Clermont, Pittsburgh, PA, p.A4448

Sepsis ?

Terms (1)

TABLE 1] Definitions of Common Terms in Data Science

Term	Definition
Big data	Digital data that are generated in high volume and high variety and that accumulate at high velocity, resulting in datasets too large for traditional data-processing systems
Data science	The set of fundamental principles that support and guide the principled extraction of information and knowledge from data
Data mining	The extraction of knowledge from data via machine learning algorithms that incorporate data science principles
Domain expertise	The understanding of real-world problems in a given domain (eg, critical care medicine) that helps frame and contextualize the application of data science to solve these problems
Machine learning	The field of study that focuses on how computers learn from data and the development of algorithms that make this learning possible
Features	The data elements, also known as independent variables, used to train a model. Features can be simple transformations of the raw data (eg, average heart rate in the last 24 h) or complex transformation such as the ones performed by neural networks (see Table 2)
Outcomes	The data elements, also known as dependent variables, represent the target for training in a supervised learning model. Outcomes can be categorical (eg, yes/no) or continuous (eg, length of hospital stay). Categorical binary outcomes are the most common in medicine (eg, died or alive by 28 days). Binary outcomes are typically represented as a Boolean logic (ie, true/false or 1/0) but can also be represented using fuzzy logic (ie, a range of probabilities, or degrees of truth, between 0 and 1)

CHEST 2018; 154(5):1239-1248

Terms (2)

Supervised learning	Algorithms that are used to uncover the relationship between a set of features and one or more known outcomes
Unsupervised learning	Algorithms that are used to uncover naturally occurring patterns or groupings in the data, without targeting a specific outcome
Model training	The process through which machine learning algorithms develop a model of the data by learning the relationships between features and, in supervised learning, between features and outcomes. This is also referred to as model derivation or data fitting
Model validation	The process of measuring how well a model fits new, independent data. For example, evaluating the performance of a supervised model at predicting an outcome in new data. This approach is also referred to as model testing.
Predictive model	A model generally trained to predict the likelihood of a condition, event, or response. The US Food and Drug Administration specifically considers predictive strategies as those geared toward identifying groups of patients more likely to respond to an intervention
Prognostic model	A model specifically trained to predict the likelihood of a condition-related endpoint or outcome such as mortality. In general, the goal is to estimate a prognosis given a set of baseline features, regardless of what ultimately leads to the outcome

CHEST 2018; 154(5):1239-1248

Terms (3)

Overfitting	The phenomenon that occurs when an algorithm learns from idiosyncrasies in the training data, usually referred to as noise. Noisy data are data that are randomly present in the training dataset but do not represent the generalizable truth (usually referred to as signal) that explains the relationships between the features and the outcomes. Overfitting will generally lead to poor performance of the model in an independent validation dataset
Digitization	The conversion of something analog or physical (eg, paper documents, printed images) into a digital format (ie, bits or 1s and 0s)
Digitalization	The wide adoption of digital technologies by an organization to leverage their digitized data with the goal of improving operations and performance. The adoption of electronic health records and other digital technologies (eg, picture archiving and communication systems for medical images, pharmacy management systems, billing systems) are examples of digitalization in health care
Data curation	The process of integrating data from different sources, structuring it, authenticating it, and annotating it to ensure its quality, add value, and facilitate its use and reuse
Structured data	Data (usually discrete or numeric) that are easy to search, summarize, sort, and quantify. Examples include vital signs (eg, heart rate) or laboratory test results (eg, CBC)
Unstructured data	Data that do not conform to a prespecified structure, such as a written narrative, images, video, or audio. Unstructured data are generally harder to search, sort, and quantify. Examples include clinician notes, pathology slides, and radiology images

CHEST 2018; 154(5):1239-1248

Algorithms in Data Science

TABLE 2] Examples of Algorithms Use in Data Science

Algorithm Class	Examples	Description
Classic regression	Linear regression, logistic regression	Linear regression is a supervised learning algorithm that models the relationship between one or more features and a continuous outcome by fitting a regression line that minimizes the sum of all the residuals, which are the distances between each feature in the training data and the line being fitted to model them. Logistic regression is a generalization of the linear model that uses the logistic function to estimate the probability of a binary outcome. To do this, the fitted sigmoid-shaped curve of the logistic function maps the feature values into a probability between 0 and 1
Regularized regression	Lasso, ridge regression, elastic net	An extension of the classic regression algorithms in which a penalty is imposed to the fitted model to reduce its complexity and decrease the risk of overfitting (see Table 1).
Tree-based	Classification and regression trees, random forest, gradient boosted trees	A class of supervised learning algorithm based on decision trees. Decision trees are a sequence of "if-then-else" splits that are derived by iteratively separating the data into groups based on the relationship of the features with the outcome. Random forest and gradient boosted trees are example of ensemble tree models. Ensemble models combine the output of many trained models to estimate an outcome

CHEST 2018; 154(5):1239-1248

Support vector machines	Linear, polynomial, radial basis kernel	A class of supervised learning algorithms that represents the data in a multidimensional feature space and then fits a "hyperplane" that best separates the data based on the outcomes of interest
K-nearest neighbor	K-nearest neighbor	A type of supervised learning algorithm that represents data in multidimensional feature space and uses local information about observations closest to a new example to predict the outcome for that example
Bayesian	Naive Bayes, Bayesian network	A class of supervised learning algorithms that use Bayes' theorem of conditional probability, which is the probability that something will happen given that something else has already occurred. In general, Bayesian algorithms work by iteratively updating the probability of an outcome (or posterior belief) given new data
Neural network	Artificial neural network, deep neural network	A class of nonlinear algorithms built using layers of nodes that extract features from the data and perform combinations that best represent the underlying structure, usually to predict an outcome. Neural networks can be shallow (eg, a perceptron with two layers) or deep (multiple layers), which form the basis for the field of deep learning
Dimensionality reduction algorithms	Principal component analysis, linear discriminant analysis	A class of unsupervised learning algorithms that exploit the inherent structure in the data to describe data using less information. Principal components, for example, summarize a large set of correlated features into a smaller number of representative features
Latent class analysis	Latent class analysis	A type of unsupervised learning algorithm that identifies unseen subgroups, or latent classes, in the data. Class membership is unknown for each example so the probability of class membership is indirectly estimated by measuring the patterns in the data
Cluster analysis	K-means, hierarchical cluster analysis	A class of unsupervised learning algorithm that uses the inherent structures in the data to best organize the data into subgroups of maximum commonality based on some distance measure between features

CHEST 2018; 154(5):1239-1248

CONTENTS

- AI in Sepsis
- AI in chest imaging
- AI in ARDS imaging

Five Questions Data Science Answers

- 這是甲，乙、丙還是丁？多元分類（multi-class classification）
- 偵測異狀（anomaly detection）
- 這有多少／這有幾個？涉及數字而非分類, 迴歸（regression）
- （資料）的組成為何？聚類法（clustering）
- 我接下來該做什麼？判斷該採取的行動, 強化學習（reinforcement learning）

Critical care parameters

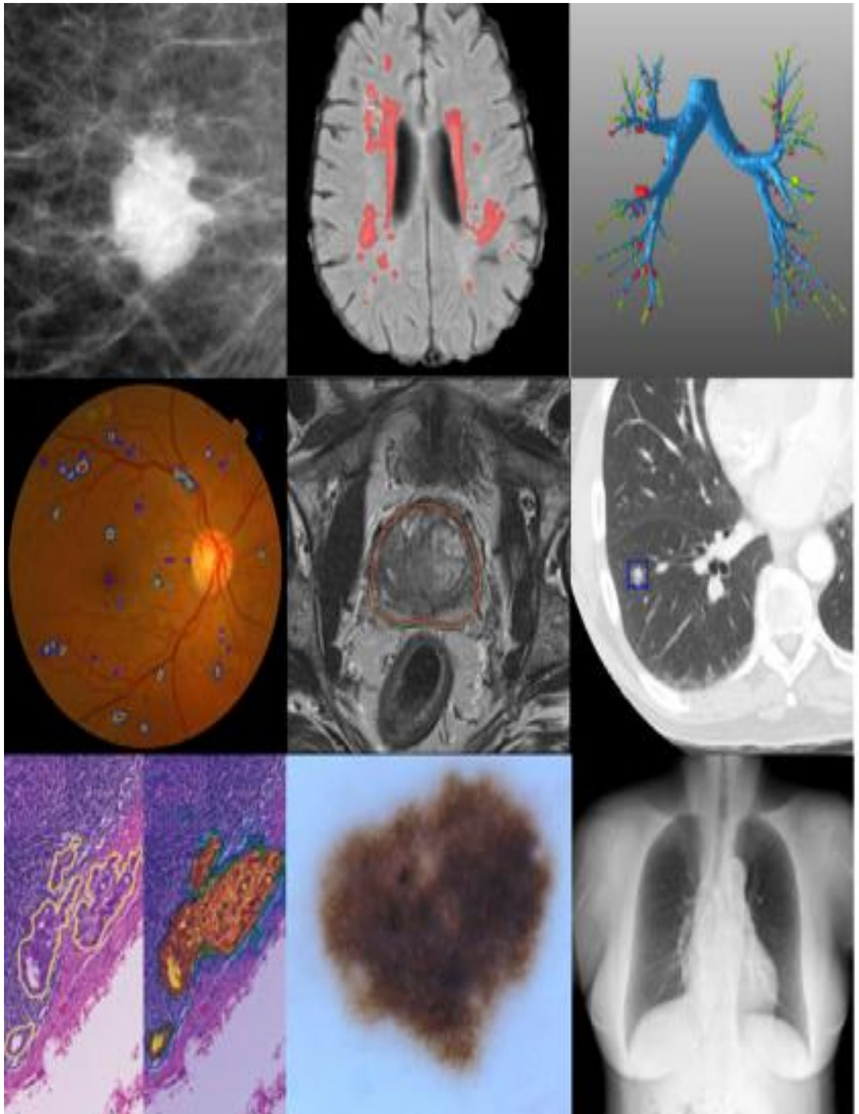
- T. P. R
- Input, Output
- EKG
- Ventilator waves
- A line
- Picco
- Image: CXR, CT, echo…
- ………

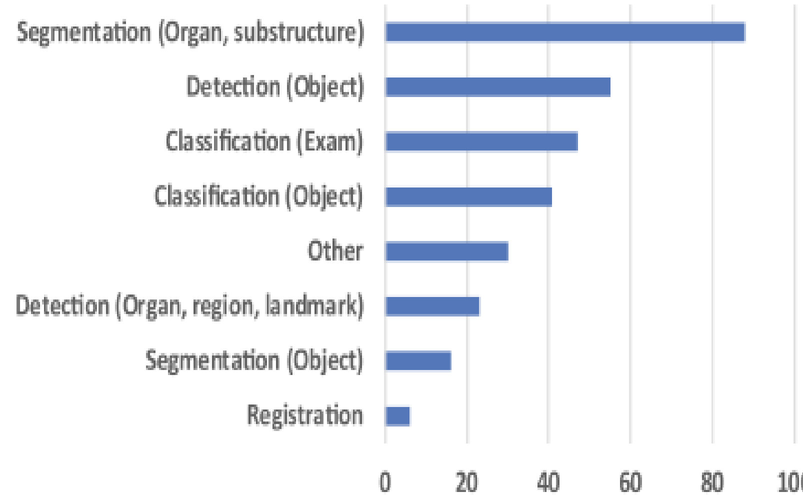
CONTENTS

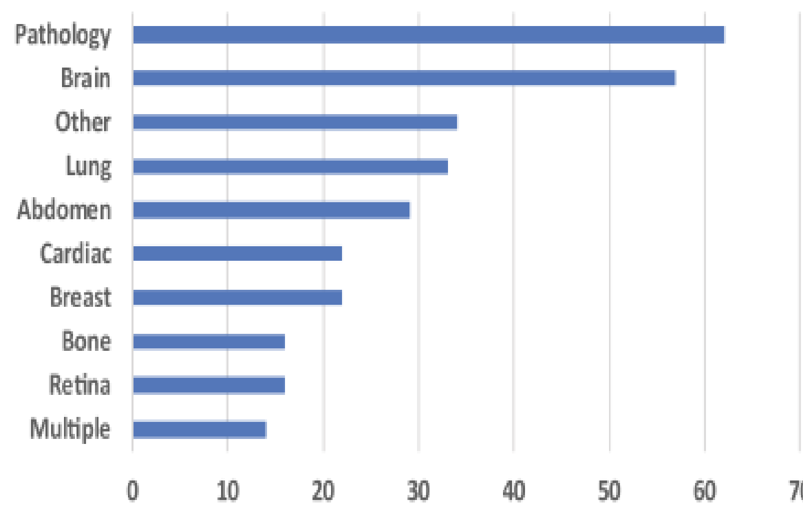
- AI in Sepsis
- AI in chest imaging
- AI in ARDS imaging

A survey on deep learning in medical image analysis

Medical Image Analysis, 2







Medical Image Analysis 42 (2017) 60–88 Number of papers

AI in chest imaging: Lung nodules

- a training data set :42092 chest radiographs

(33467 normal and 8625 nodule CXR) to optimize network weights

Semisupervised learning manner by using all of the image-level labels (normal or nodule), but only part of the pixel-level annotations in the training data set: 37.2% (3213 of 8625) of nodule

- A deep convolutional neural network with 25 layers and eight residual connections
- Brightness, contrast, and image size on input chest radiographs were randomly adjusted to make DLAD irrelevant to the variations

Nam et al, Radiology 20

AI in chest imaging: Lung nodules

- a tuning data set: 600 CXR

(300 normal and 300 nodule CXR) to optimize hyperparameters

- an internal validation data set: 600 CXR

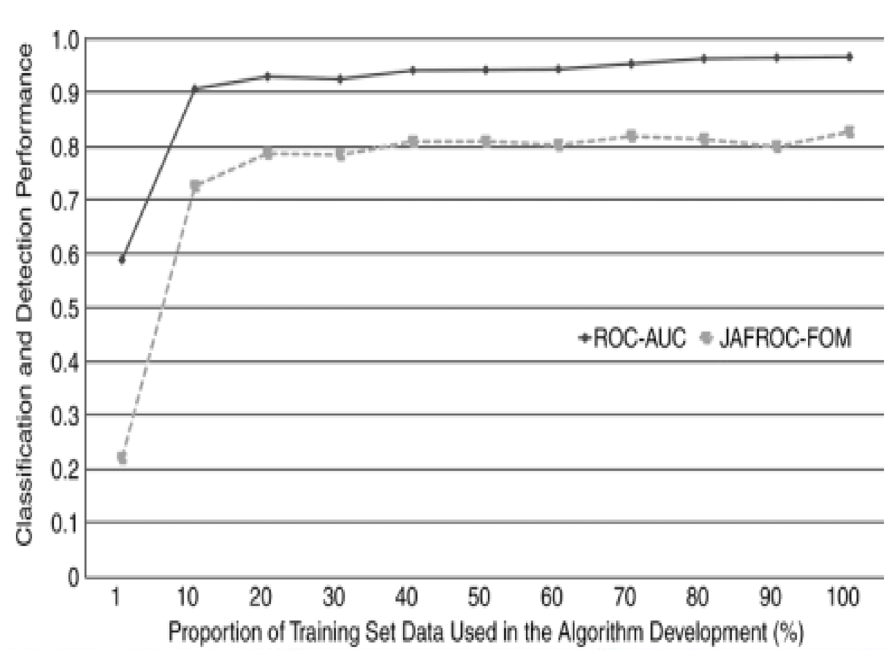
(300 normal and 300 nodule CXR)

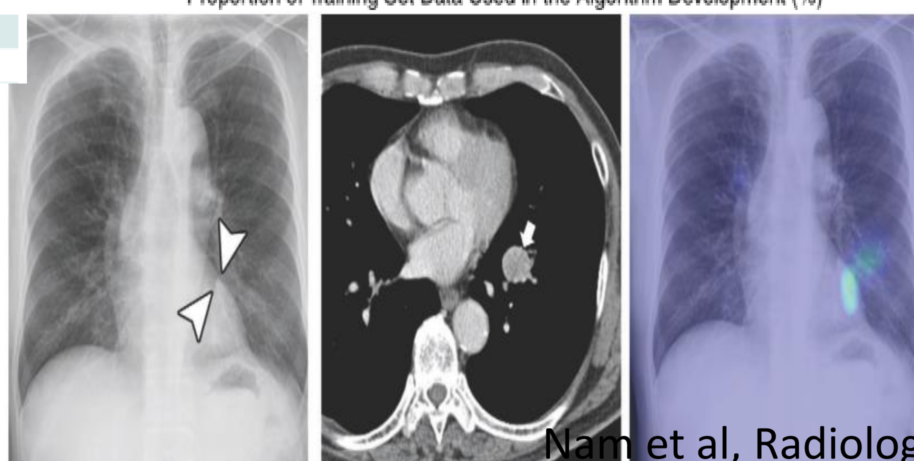
- Four independent data sets for external validation

Nam et al, Radiology 20

AI in chest imaging: Lung nodules

Characteristic	Seoul National University Hospital	Boramae Hospital	National Cancer Center	University of California San Francisco Medical Center
Patient information				
No. of chest radiographs	181	182	181	149
No. of normal radiographs	62	59	70	60
No. of nodule chest radiographs	119	123	111	89
Patients with nodules	119	123	111	89
No. of men	72	89	52	54
No. of women	47	34	59	35
Mean age*	53.1 ± 11.6	71.2 ± 11.2	63.0 ± 9.8	61.1 ± 7.5
No. of healthy patients	62	59	70	60
No. of men	27	26	48	30
No. of women	35	33	22	30
Mean age*	63.0 ± 12.0	53.6 ± 10.5	63.0 ± 9.8	49.2 ± 16.5
Nodule information				
Total no. of nodules	143	139	115	104
No. of nodules per chest radiograph				
One nodule	99	109	107	77
Two nodules	16	12	4	9
Three nodules	4	2	0	3
No. of nodules according to size				
≤1.0 cm	18 (12.6)	4 (2.9)	8 (7.0)	6 (5.8)
1.1–1.5 cm	20 (14.0)	11 (7.9)	26 (22.6)	13 (12.5)
1.6–2.0 cm	25 (17.5)	17 (12.2)	20 (17.4)	17 (16.3)
2.1–3.0 cm	31 (21.7)	33 (23.7)	59 (51.3)	21 (20.2)
>3.0 cm	49 (34.3)	74 (53.2)	2 (1.7)	47 (45.2)
Overlapped or masked by heart	11 (7.7)	13 (9.4)	4 (3.5)	1 (1.0)
Overlapped or masked by diaphragm	9 (6.3)	15 (10.8)	3 (2.6)	3 (2.9)
Overlapped or masked by hilar vessels	21 (14.7)	25 (18.0)	1 (0.9)	14 (13.5)
Overlapped or masked by clavicle/ first rib	26 (18.2)	21 (15.1)	9 (7.8)	8 (7.7)





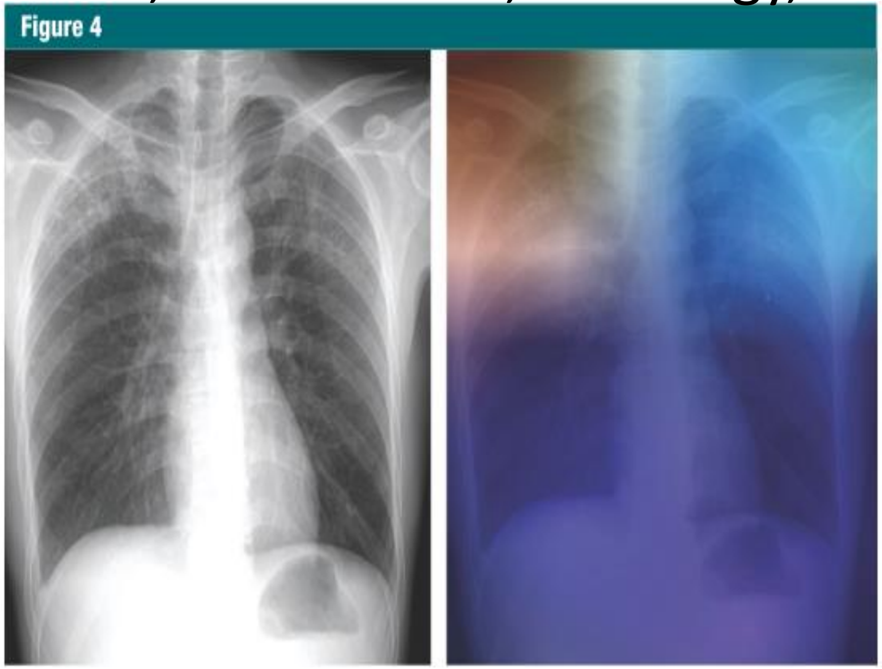
Nam et al, Radiology 2

Table 3: Patient Classification and Nodule Detection at the Observer Performance Test

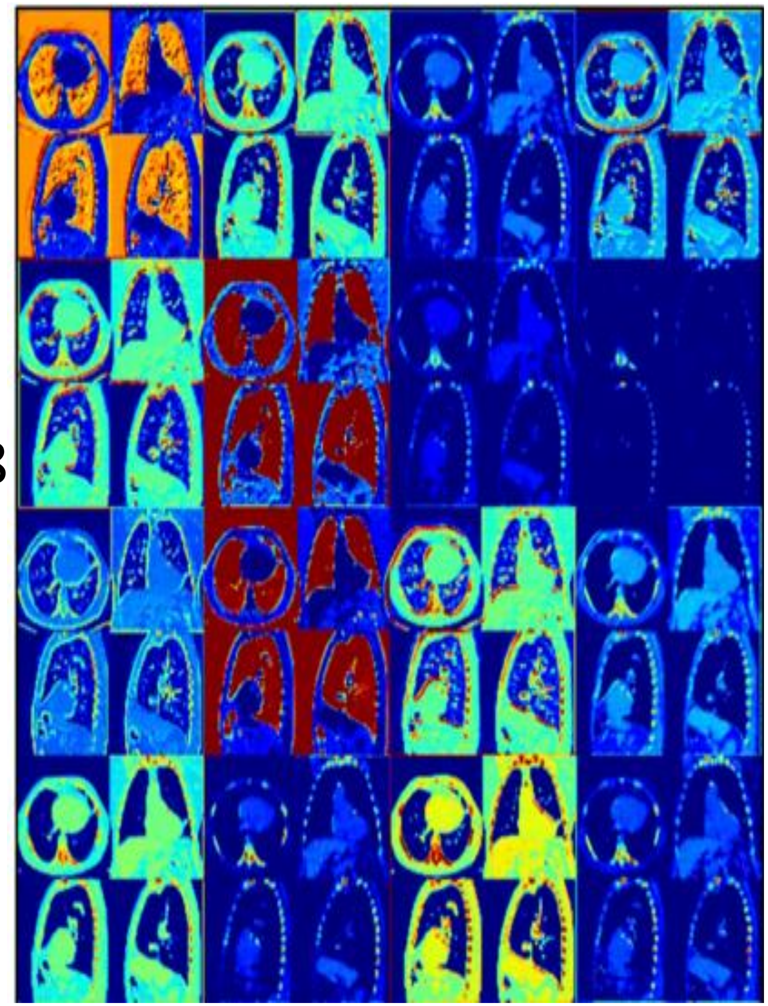
Observer	Test 1		DLAD versus Test 1 (*P* Value)		Test 2		Test 1 versus Test 2 (*P* Value)	
	Radiograph Classification (AUROC)	Nodule Detection (JAFROC FOM)	Radiograph Classification	Nodule Detection	Radiograph Classification (AUROC)	Nodule Detection (JAFROC FOM)	Radiograph Classification	Nodule Detection
Nonradiology physicians								
Observer 1	0.77	0.716	<.001	<.001	0.91	0.853	<.001	<.001
Observer 2	0.78	0.657	<.001	<.001	0.90	0.846	<.001	<.001
Observer 3	0.80	0.700	<.001	<.001	0.88	0.783	<.001	<.001
Group		0.691		<.001*		0.828		<.001*
Radiology residents								
Observer 4	0.78	0.767	<.001	<.001	0.80	0.785	.02	.03
Observer 5	0.86	0.772	.001	<.001	0.91	0.837	.02	<.001
Observer 6	0.86	0.789	.05	.002	0.86	0.799	.08	.54
Observer 7	0.84	0.807	.01	.003	0.91	0.843	.003	.02
Observer 8	0.87	0.797	.10	.003	0.90	0.845	.03	.001
Observer 9	0.90	0.847	.52	.12	0.92	0.867	.04	.03
Group		0.790		<.001*		0.867		<.001*
Board-certified radiologists								
Observer 10	0.87	0.836	.05	.01	0.90	0.865	.004	.002
Observer 11	0.83	0.804	<.001	<.001	0.84	0.817	.03	.04
Observer 12	0.88	0.817	.18	.005	0.91	0.841	.01	.01
Observer 13	0.91	0.824	>.99	.02	0.92	0.836	.51	.24
Observer 14	0.88	0.834	.14	.03	0.88	0.840	.87	.23
Group		0.821		.02*		0.840		.01*
Thoracic radiologists								
Observer 15	0.94	0.856	.15	.21	0.96	0.878	.08	.03
Observer 16	0.92	0.854	.60	.17	0.93	0.872	.34	.02
Observer 17	0.86	0.820	.02	.01	0.88	0.838	.14	.12
Observer 18	0.84	0.800	<.001	<.001	0.87	0.827	.02	.02
Group		0.833		.08*		0.854		<.001*

Nam et al, Radiology

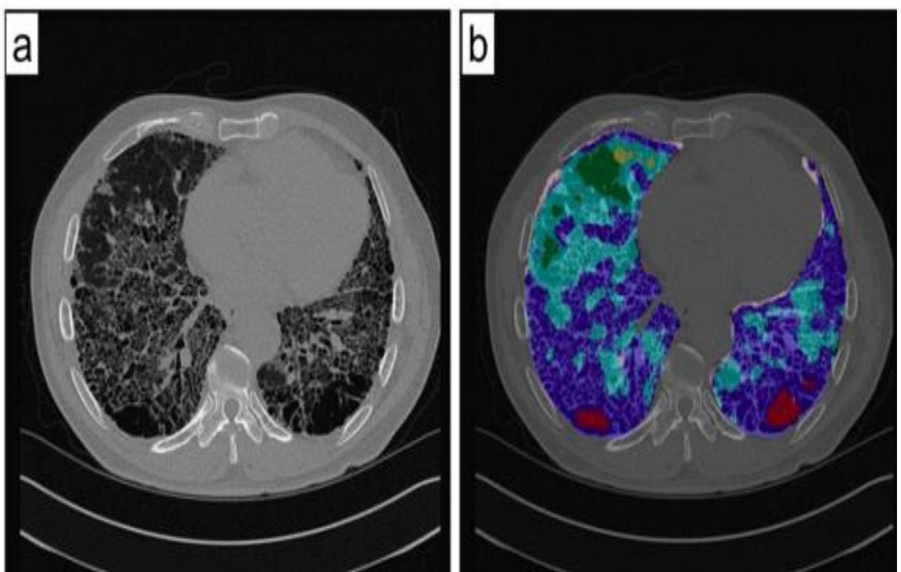
TB, Lakhani et al, Radiology, 2017



COPD, Gonzalez et al, AJRCCM, 20



ILD, Kim et al, J Digit Imaging, 2018



CONTENTS

- AI in Sepsis
- AI in chest imaging
- AI in ARDS imaging

ARDS definition

- Acute onset over 1 week or less
- Bilateral opacities consistent with pulmonary edema must be present and may be detected on CT or chest radiograph
- PF ratio <300mmHg with a minimum of 5 cmH20 PEEP (or CPAP)
- "must not be fully explained by cardiac failure or fluid overload," in the physician's best estimation using available information — an "objective assessment" (e.g. echocardiogram) should be performed in most cases if there is no clear cause such as trauma or sepsis.

ARDS severity

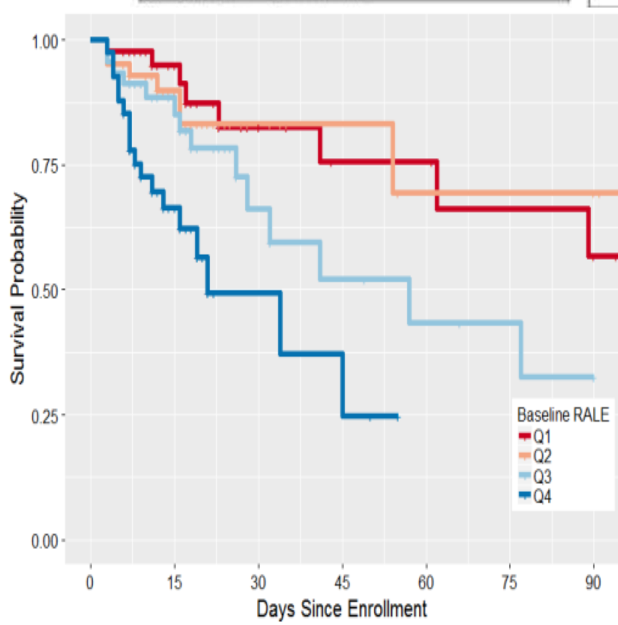
ARDS Severity	PaO2/FiO2 *	Mortality**
Mild	200 – 300	27%
Moderate	100 – 200	32%
Severe	< 100	45%

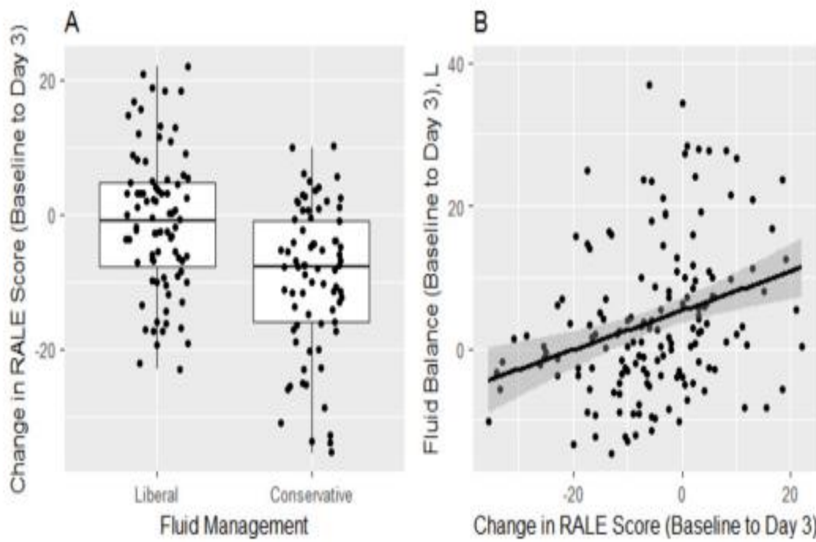
*on PEEP 5+; **observed in cohort

Consolidation[a]	
Consolidation Score	Extent of alveolar opacities
0	None
1	<25%
2	25-50%
3	50-75%
4	>75%
Density[b]	
Density Score	Density of alveolar opacities
1	Hazy
2	Moderate
3	Dense
Final RALE Score[c]	
Right Lung	Left Lung
Upper Quadrant Cons x Den = Q1 score	Upper Quadrant Cons x Den = Q3 score
Lower Quadrant Cons x Den = Q2 score	Lower Quadrant Cons x Den = Q4 score
Total RALE = Q1+ Q2 + Q3 + Q4	

Calculation of the RALE Score for Left Radiograph					
Score	Q1	Q2	Q3	Q4	Total
Consolidation	4	4	1	2	
Density	3	3	3	3	
Quadrant Score	4 x 3 = 12	4 x 3 = 12	1 x 3 = 3	2 x 3 = 6	33

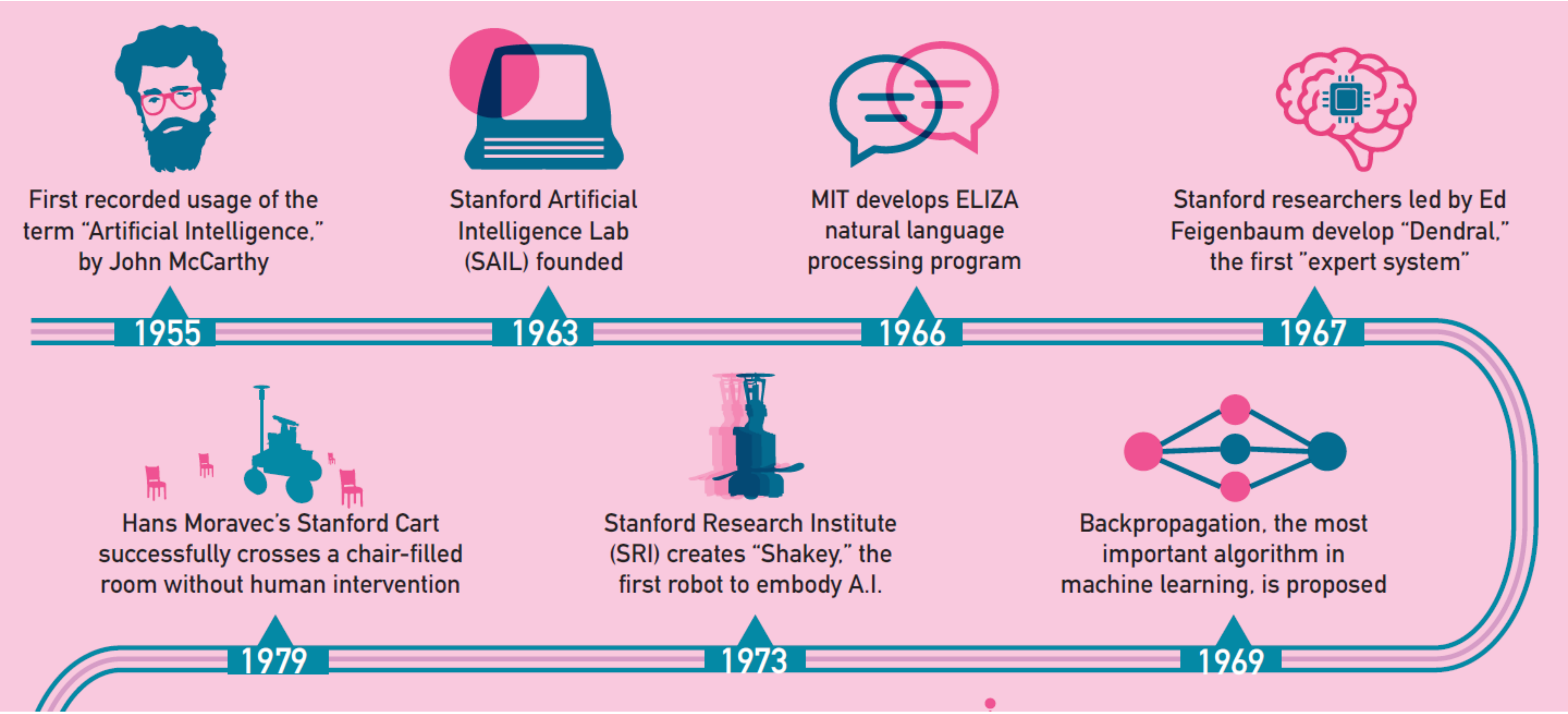
Calculation of the RALE Score for Right Radiograph					
Score	Q1	Q2	Q3	Q4	Total
Consolidation	4	4	4	4	
Density	1	2	1	2	
Quadrant Score	4 x 1 = 4	4 x 2 = 8	4 x 1 = 4	4 x 2 = 8	24



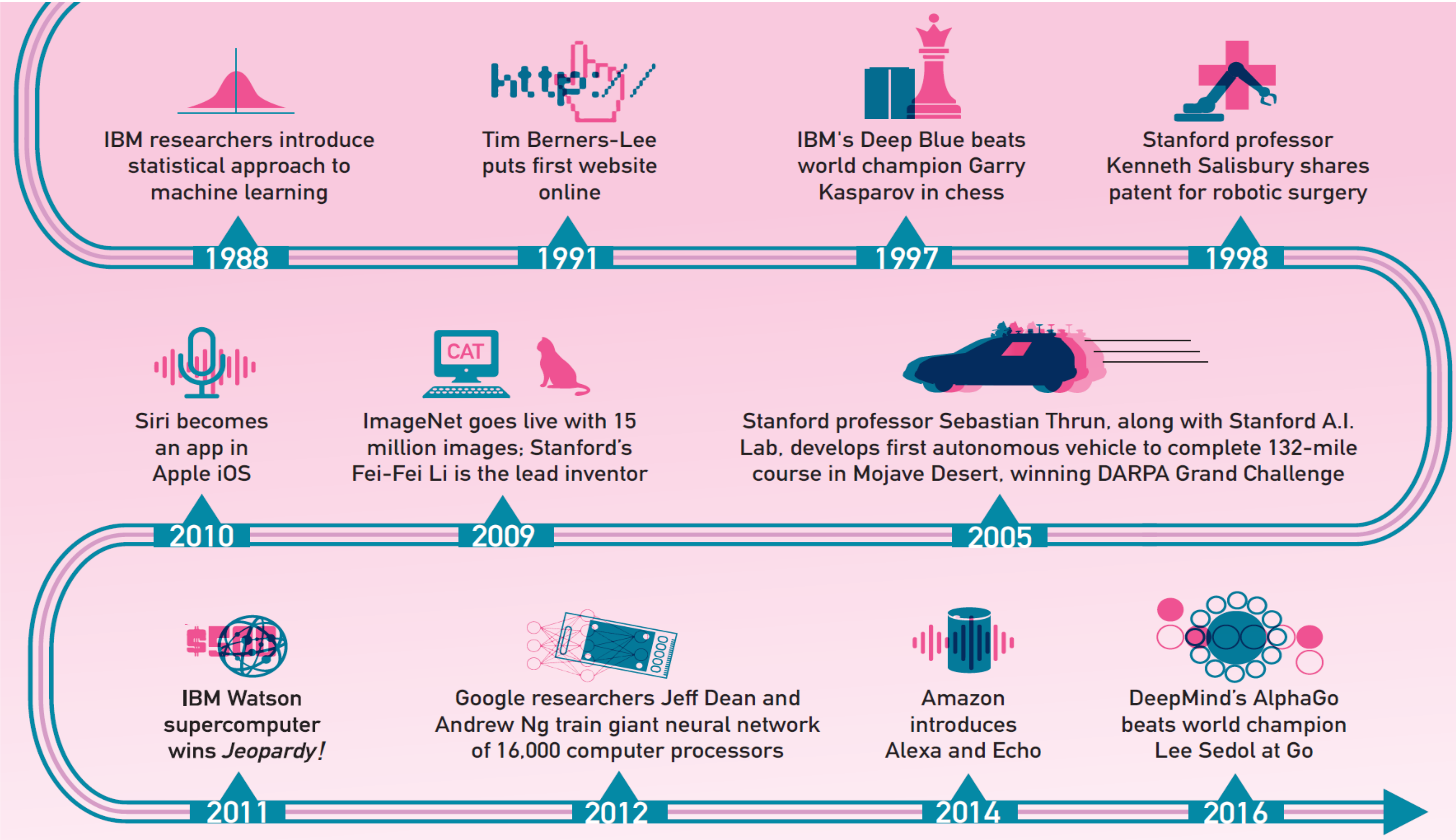


Warren MA, et al. Thora

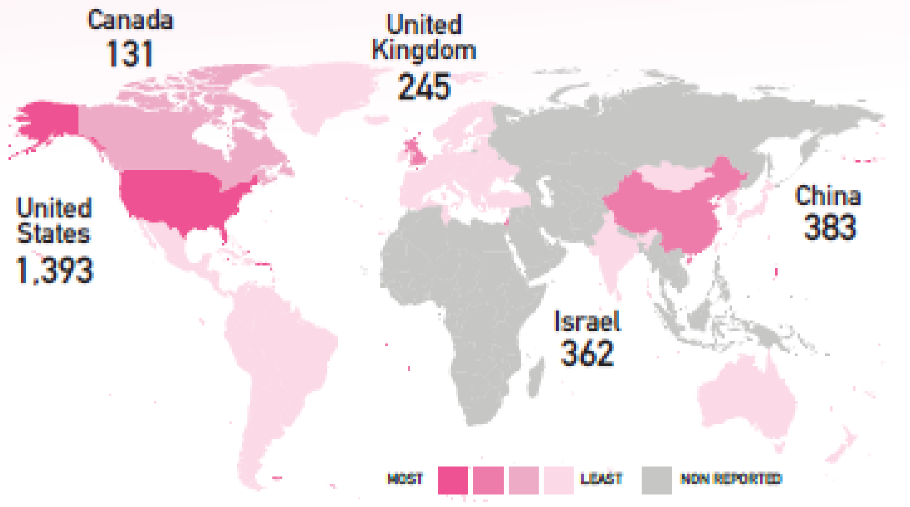
GREAT MOMENTS IN A.I. (1)



GREAT MOMENTS IN A.I. (2)

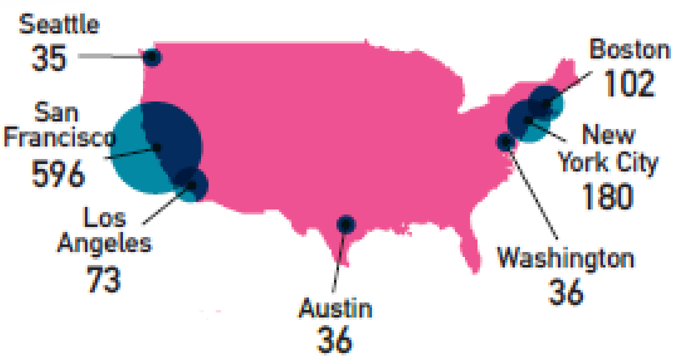


A.I. STARTUPS GLOBALLY



Source Asgard CB Insights

TOP A.I. STARTUP CITIES IN THE U.S.



Source Asgard CB Insights

+$15.7 Trillion

Amount A.I. is estimated
to add to the global economy by 2030

Source https://www.gartner.com/en/newsroom/press-releases/2019-01-21-gartner-survey-shows-37-percent-of-organizations-have

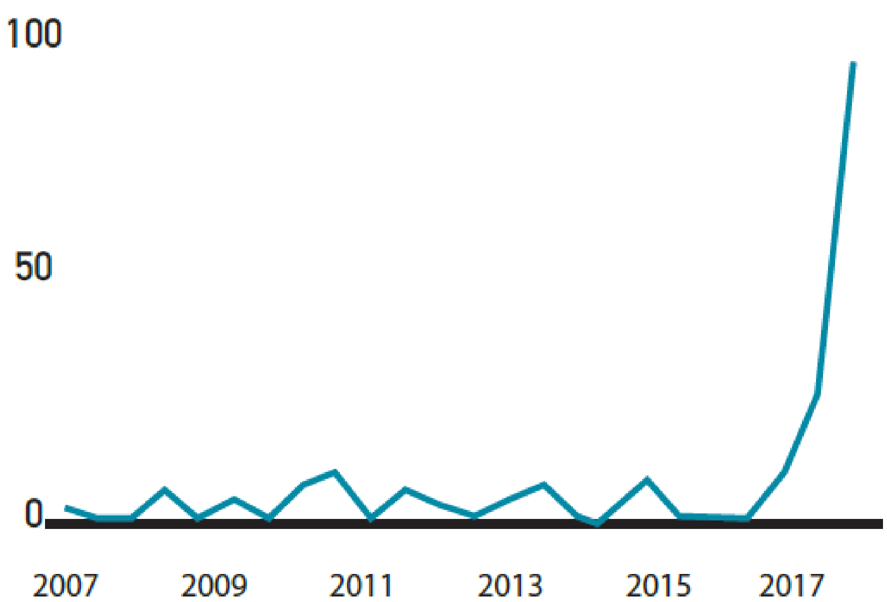
$9.3 Billion

Amount startups raised from
venture capital firms in the U.S. in 2018

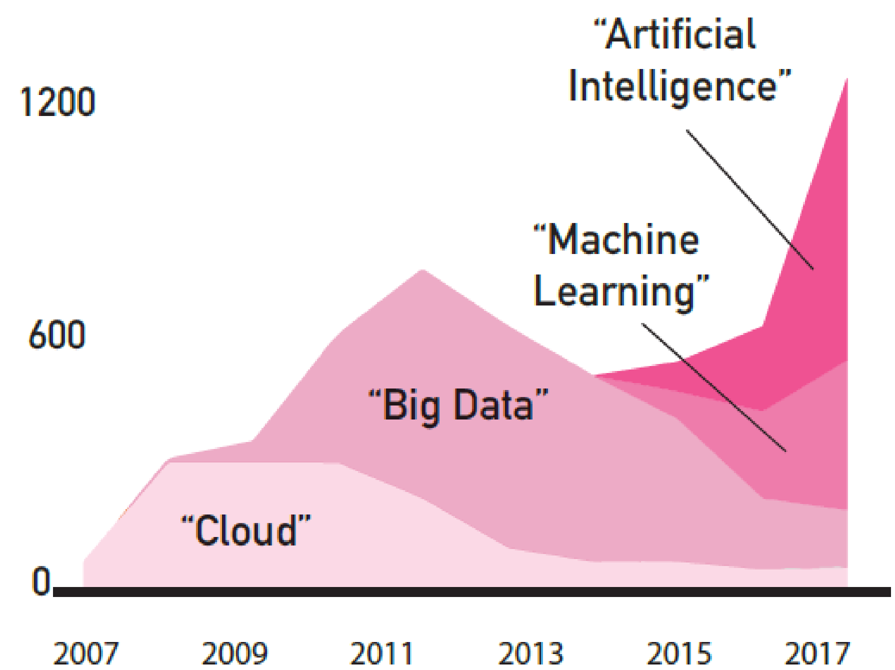
Source PricewaterhouseCoopers Report

A.I.'S INCREASING POPULARITY

Mentions of A.I. in U.S. Congress



Mentions of Tech Terms on Public Company Investor Calls

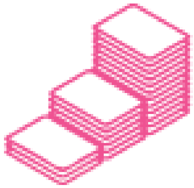


Source Prattle & U.S. Congressional Record Website

A.I. RESEARCH AND EDUCATION

The number of academic papers published on the subject of A.I. has increased by more than 8x since 1996

+8x



Enrollment in introductory A.I. college courses increased 500% from 2012 to 2017

+500x

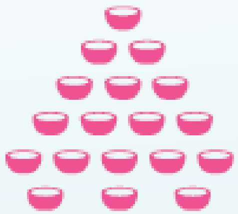
2012 2017

Source A.I. Index

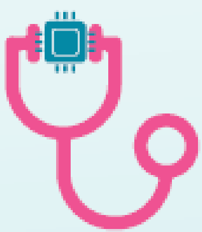
A.I. BREAKTHROUGHS AT STANFORD

Stanley, the self-driving car developed at SAIL, won DARPA's 2005 Grand Challenge

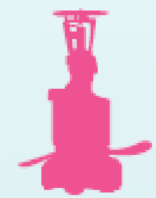
ImageNet changed the course of machine learning/A.I. and ushered in the age of deep learning



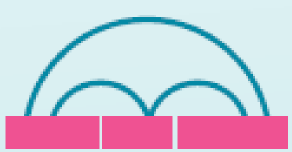
18 Members of the Stanford A.I. Lab have gone on to earn ACM Turing Awards



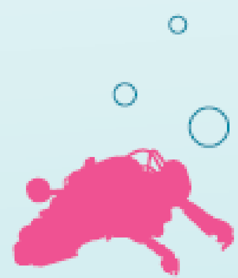
The first super computer for A.I. in medicine was created by and Stanford School of Medicine



Shakey, the first of its kind, led the way for autonomous robots



Stanford CoreNLP, released in 2010, is the leading open source natural language processing toolkit



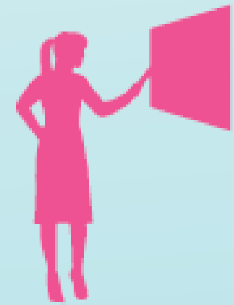
OceanOne, an underwater humanoid robot with haptic feedback, can explore the ocean in high fidelity

A.I. COURSES AT STANFORD

54

A.I. courses taught in 2018

Research topics include computer vision, natural language processing, advanced robotics, and computational genomics



MOST POPULAR
A.I.-RELATED COURSE

CS 221: Artificial Intelligence, Principles and Techniques

Methods of machine learning

- Supervised learning
- Unsupervised learning

- Reinforcement learning

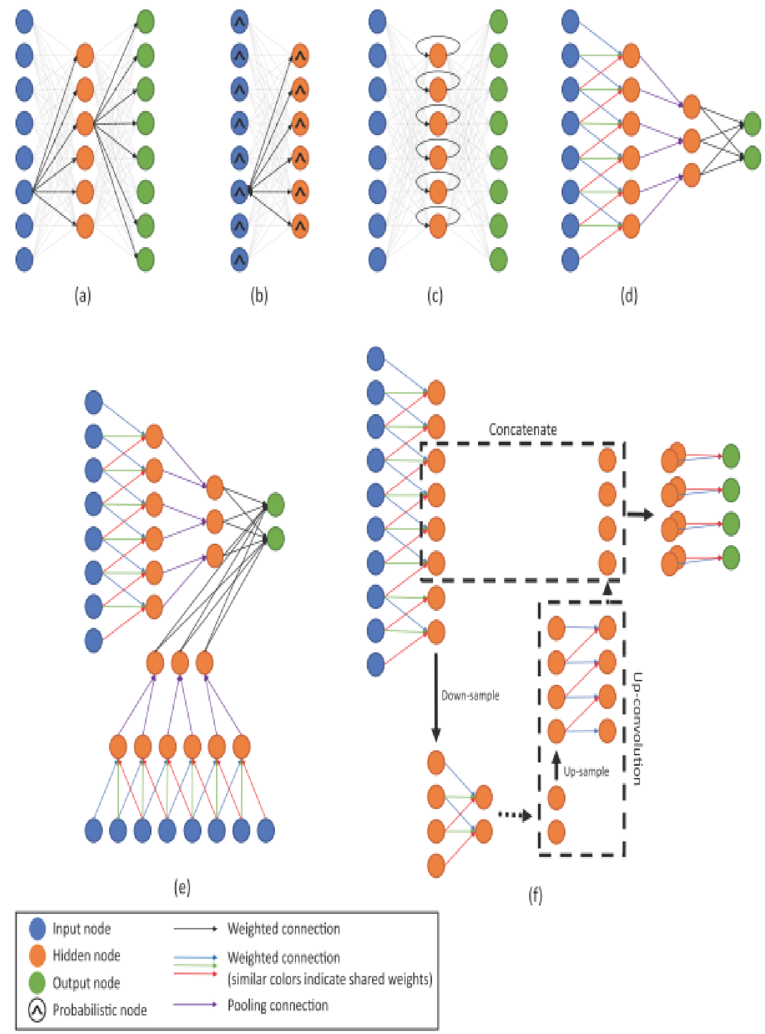
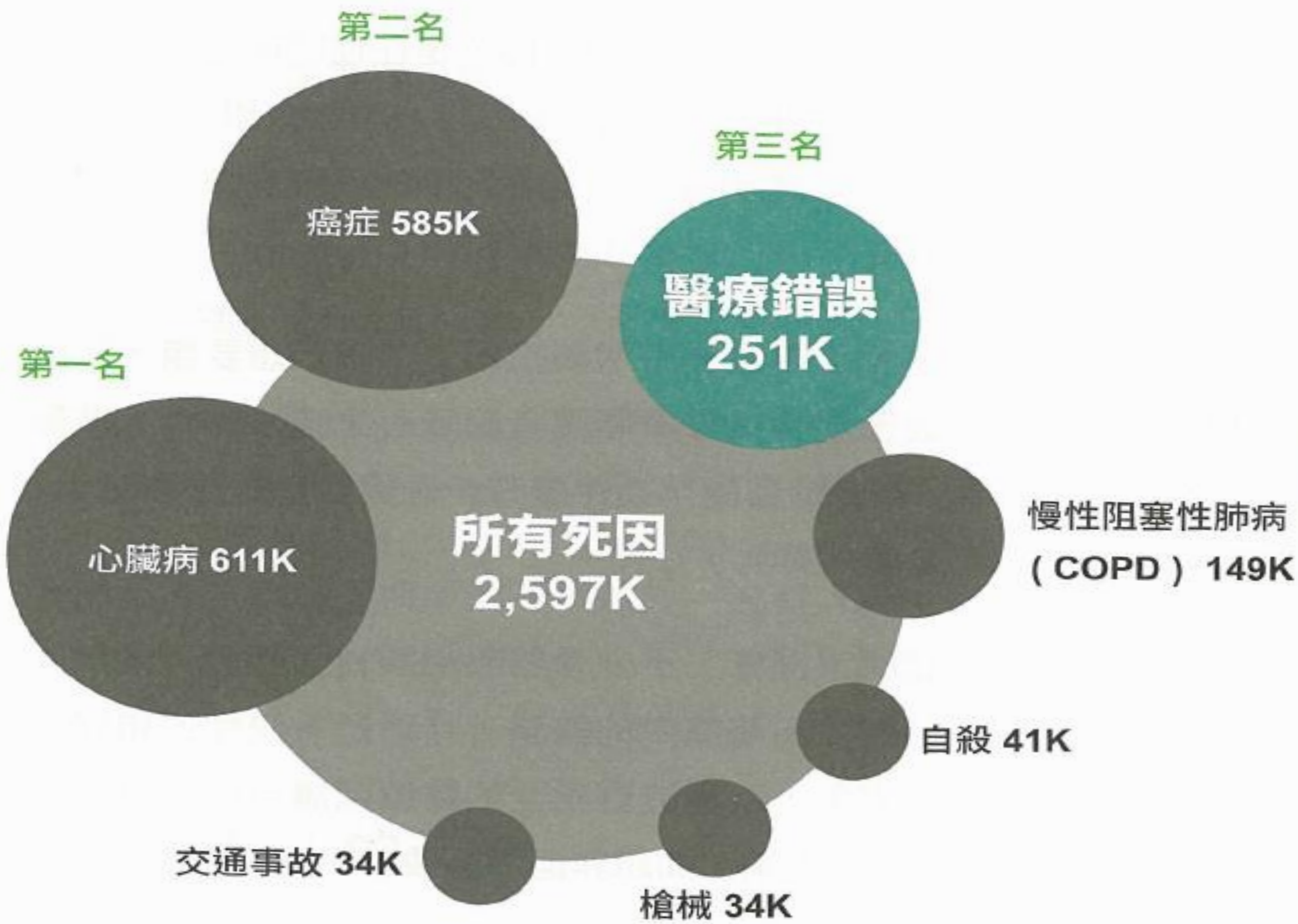
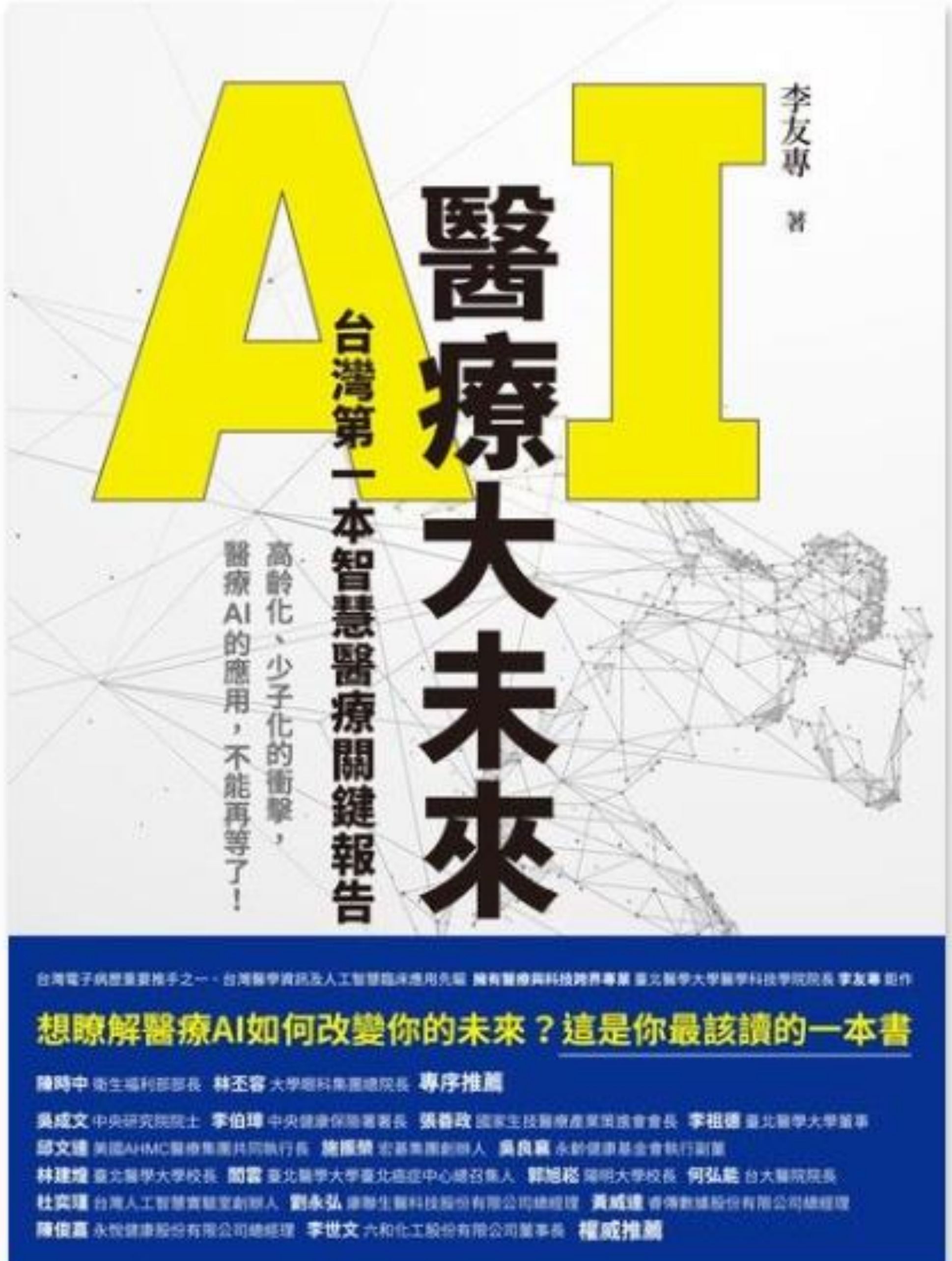


圖2-d 醫療疏失是美國第三大死因 (2016)



註：K=1千人

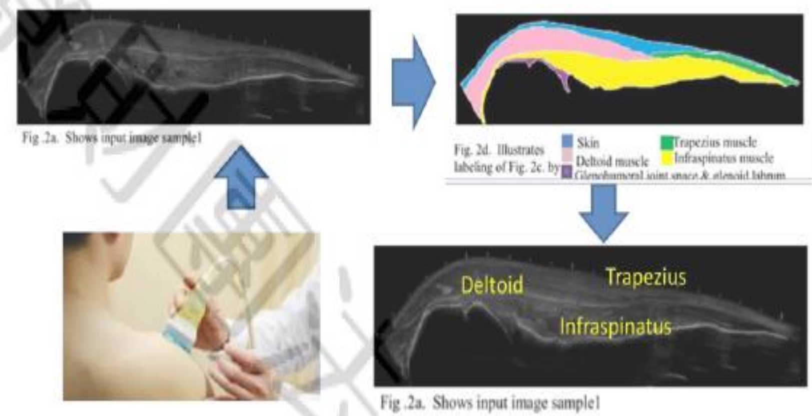
資料來源：英國醫學期刊（BMJ），2016

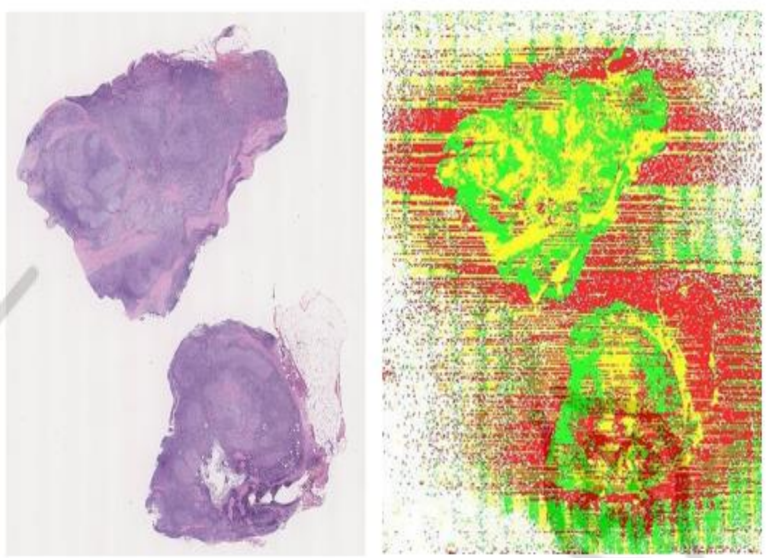


AI醫療大未來
台灣第一本智慧醫療關鍵報告
高齡化、少子化的衝擊，
醫療AI的應用，不能再等了！
李友專 著
想瞭解醫療AI如何改變你的未來？這是你最該讀的一本書
陳時中 衛生福利部部長 林丕容 大學眼科集團總院長 專序推薦
權威推薦

應用

- AI結果需要臨床應用驗證效果，需大量臨床服務支持
- AI主要要解決現有問題，需要臨床醫師提出





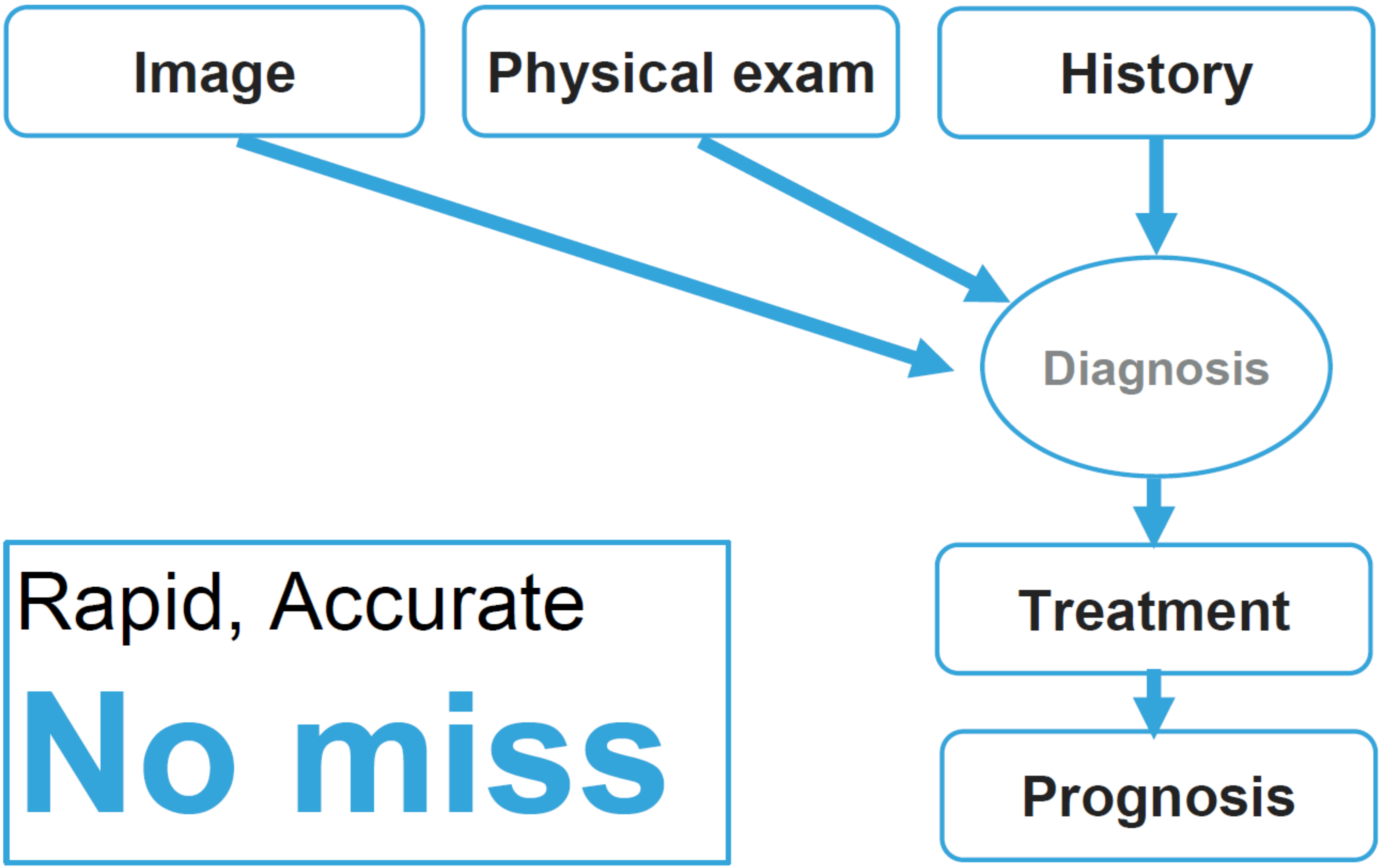


Image
Physical exam
History
Diagnosis
Treatment
Prognosis
Rapid, Accurate
No miss

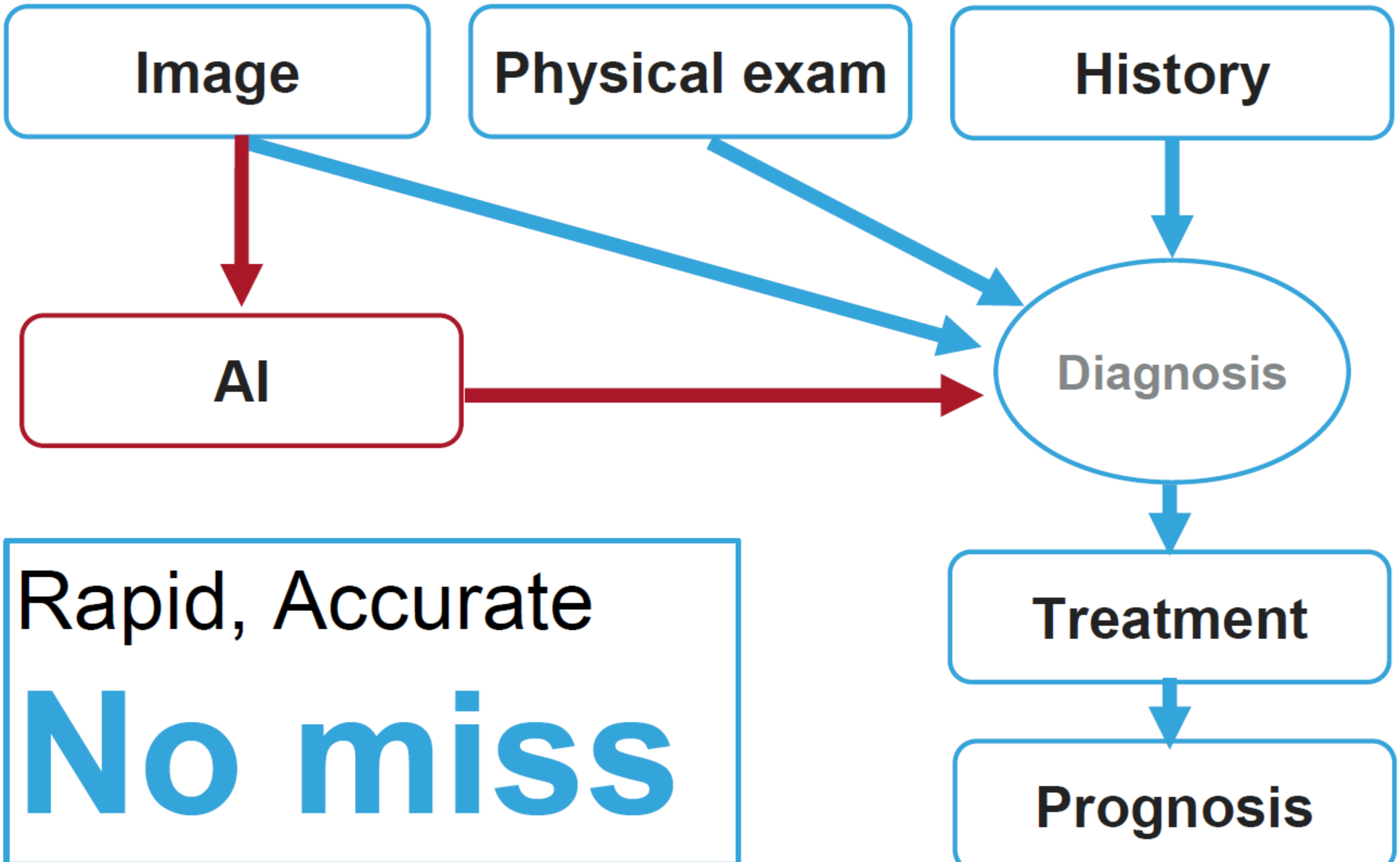


Image
Physical exam
History
AI
Diagnosis
Rapid, Accurate
No miss
Treatment
Prognosis

Comparison of machine learning models for the prediction of mortality of patients with unplanned extubation in intensive care units

Meng Hsuen Hsieh[1], Meng Ju Hsieh[2], Chin-Ming Chen[3,4], Chia-Chang Hsieh[5], Chien-Ming Chao[6] & Chih-Cheng Lai[6]

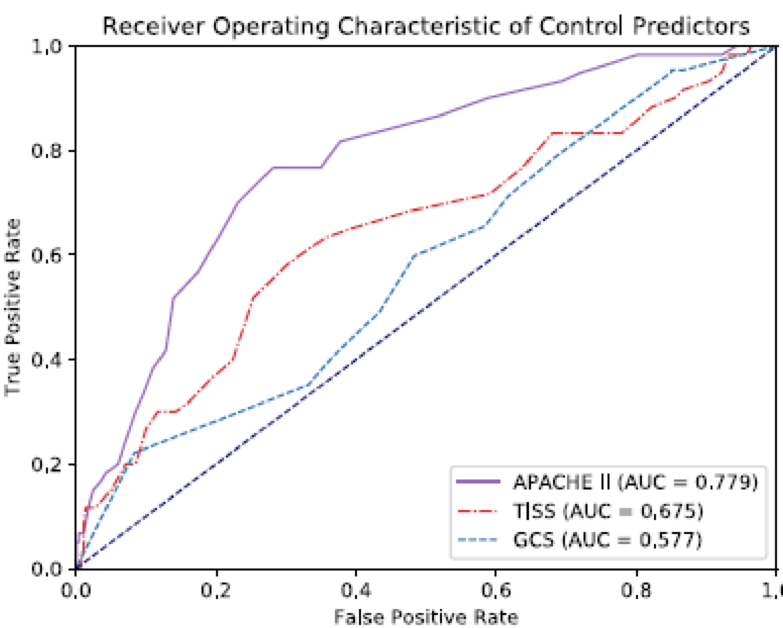


Figure 1. ROC curve of control variables.

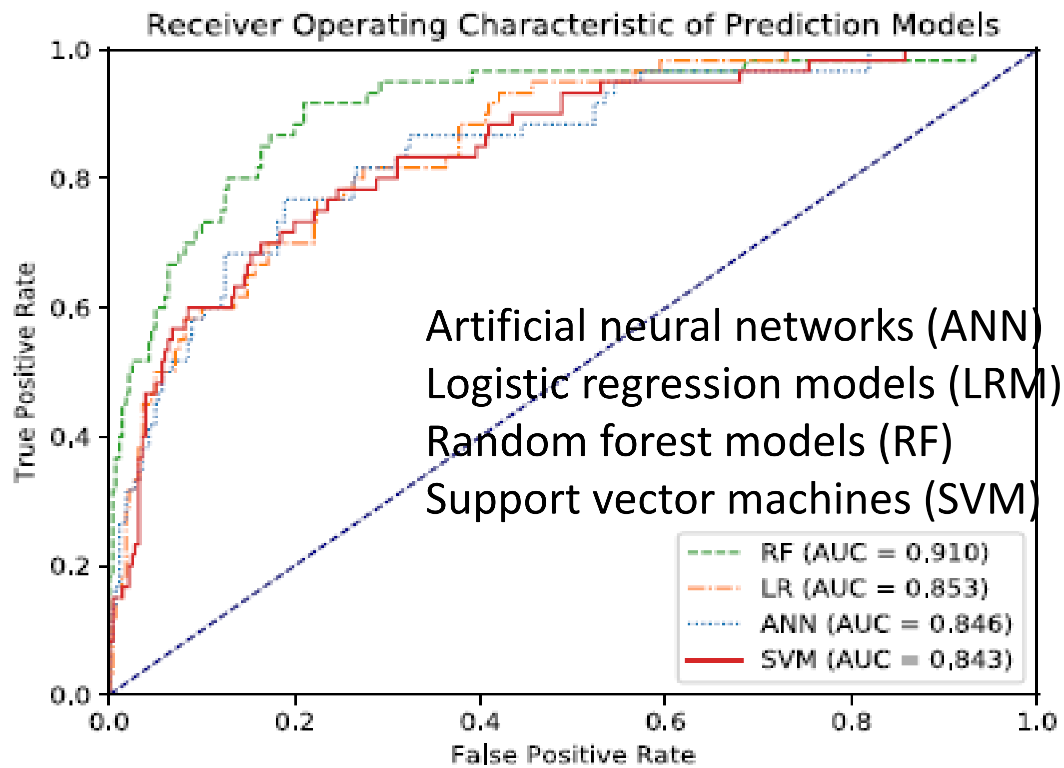


Figure 2. ROC curve of ANN, LR, RF, and SVM models.

第一波是單點效率的提升，像是瑕疵檢測、智能化排程、預防性維修、原料組合最佳化，就是機器取代人，這時新創公司與科技領導廠商主導創新與先期應用。

第二波是企業流程的提升，當所有的單點應用被普及化之後，就會迎來重組作業流程效率的提升，形成線的變化，生產作業、服務程序與人機協作將會帶來流程的梳理與重整，顧問式的專案將興起，挑戰的將是原先的想像力與創造力。

第三波是產業面的提升，這時傳統產業也跟進了，很多成熟與特殊領域的隱形冠軍，在AI應用上會走向產業內的全面運動，在產業應用的廣度與成效有很大提升，成熟標準化產品會推出且開始普及，這會是一個面的變化。

第四波是生態系統的整體，將因為應用者廣且多，因量變帶來質變，嶄新的應用會建構在已有的基礎之上，翻倍的效率提高也會影響整個行業的生態系統，不同生態系統會形成競爭甚至取代了缺乏競爭力的產