High-performance medicine: the convergence of human and artificial intelligence

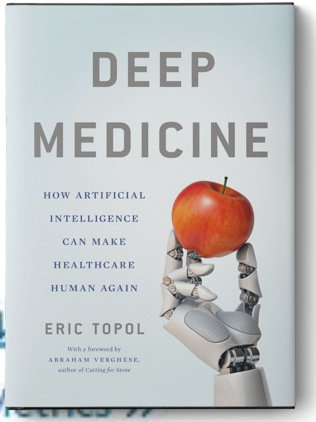
Karl Sousa
kvms@cin.ufpe.br

nature medicine

Review Article | Published: 07 January 2019

High-performance medicine: the convergence of human and artificial intelligence

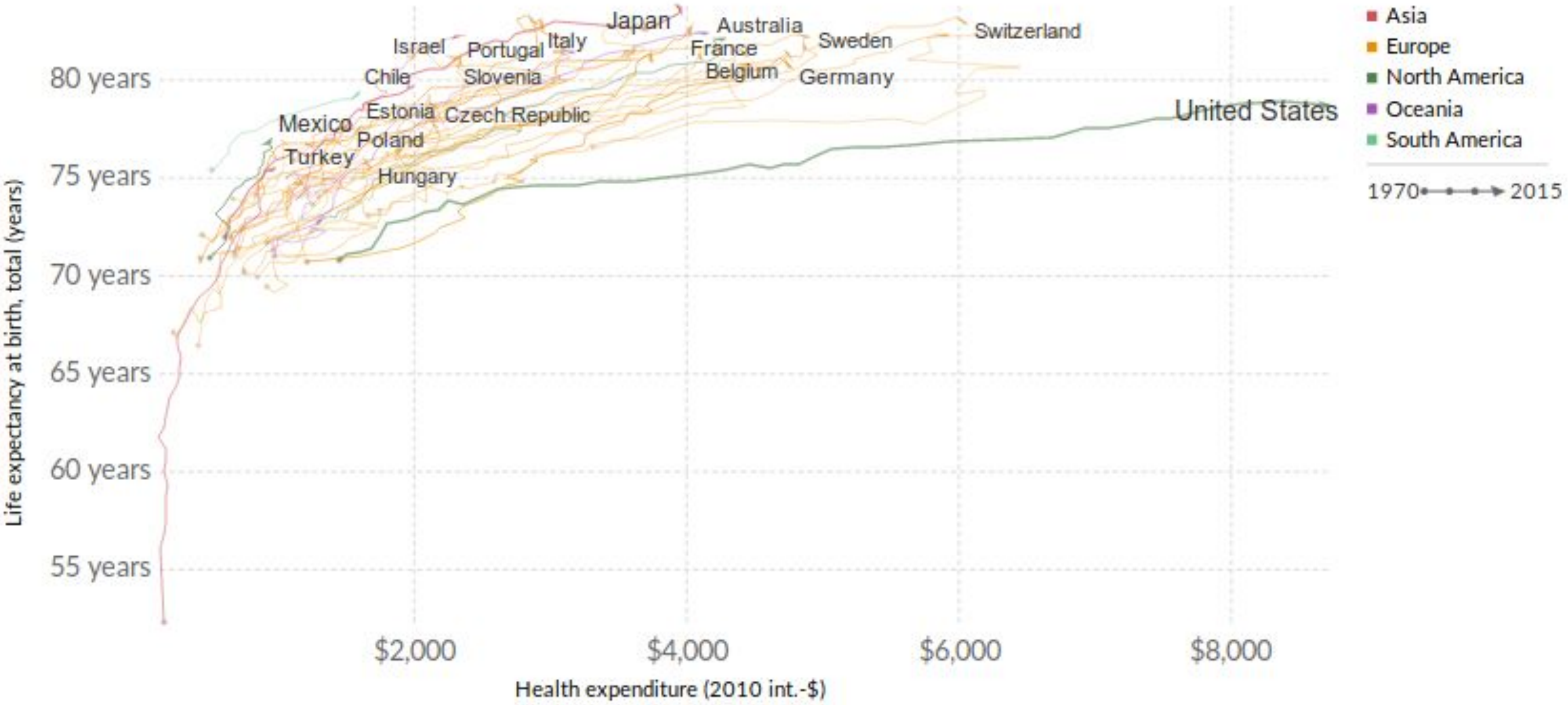
Eric J. Topol



Nature Medicine **25**, 44–56 (2019) | Download Citation

Life expectancy vs. health expenditure, 1970 to 2015

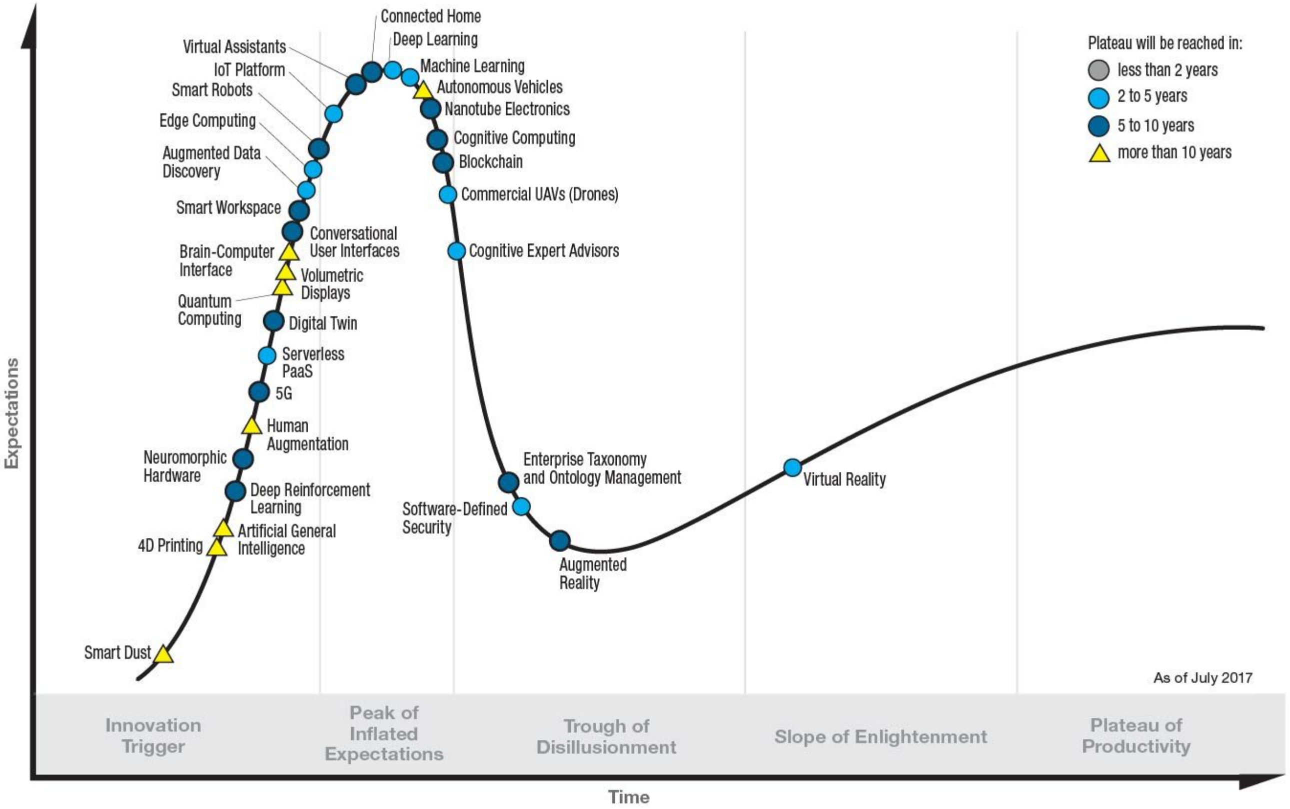
Health financing is reported as the annual per capita health expenditure and is adjusted for inflation and price level differences between countries (measured in 2010 international dollars).



Source: World Bank, Health Expenditure and Financing - OECDstat (2017)


Provider Software
Opt-in Genome Registries
Hospital EHR's
Patient Registries
Non Retail Outlets
Private Payers and Plan Claims
Mobile Data and Wearables
Government Health Plan Claims
Medical Claims
Pharmacy Claims

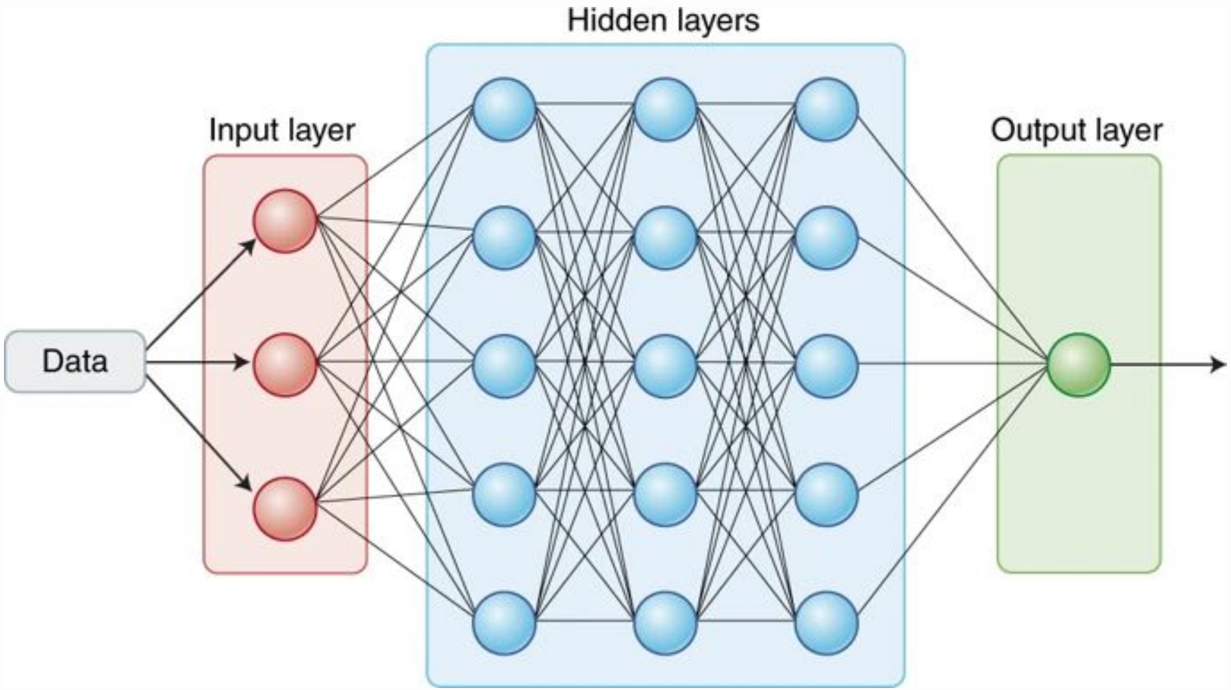
Gartner Hype Cycle for Emerging Technologies, 2017



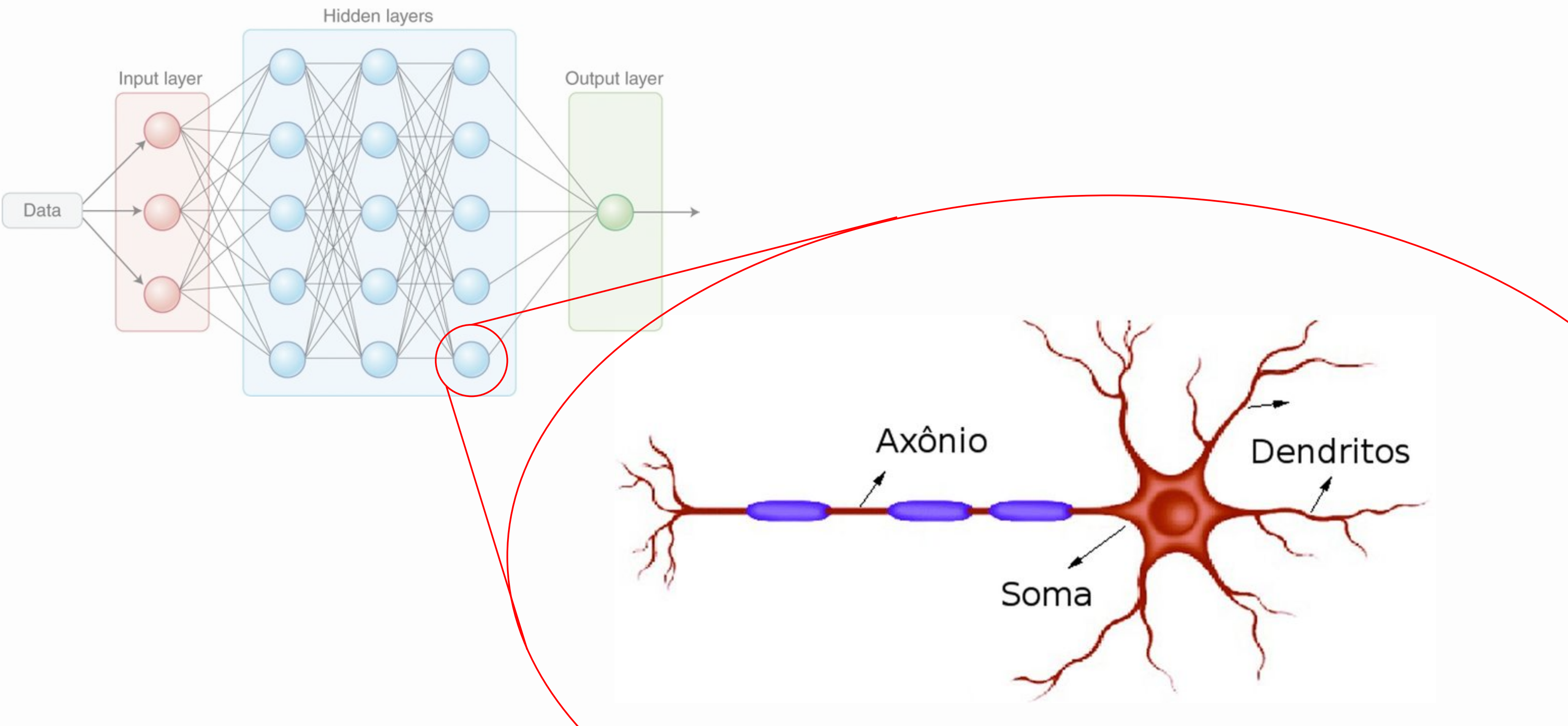
gartner.com/SmarterWithGartner

Source: Gartner (July 2017)

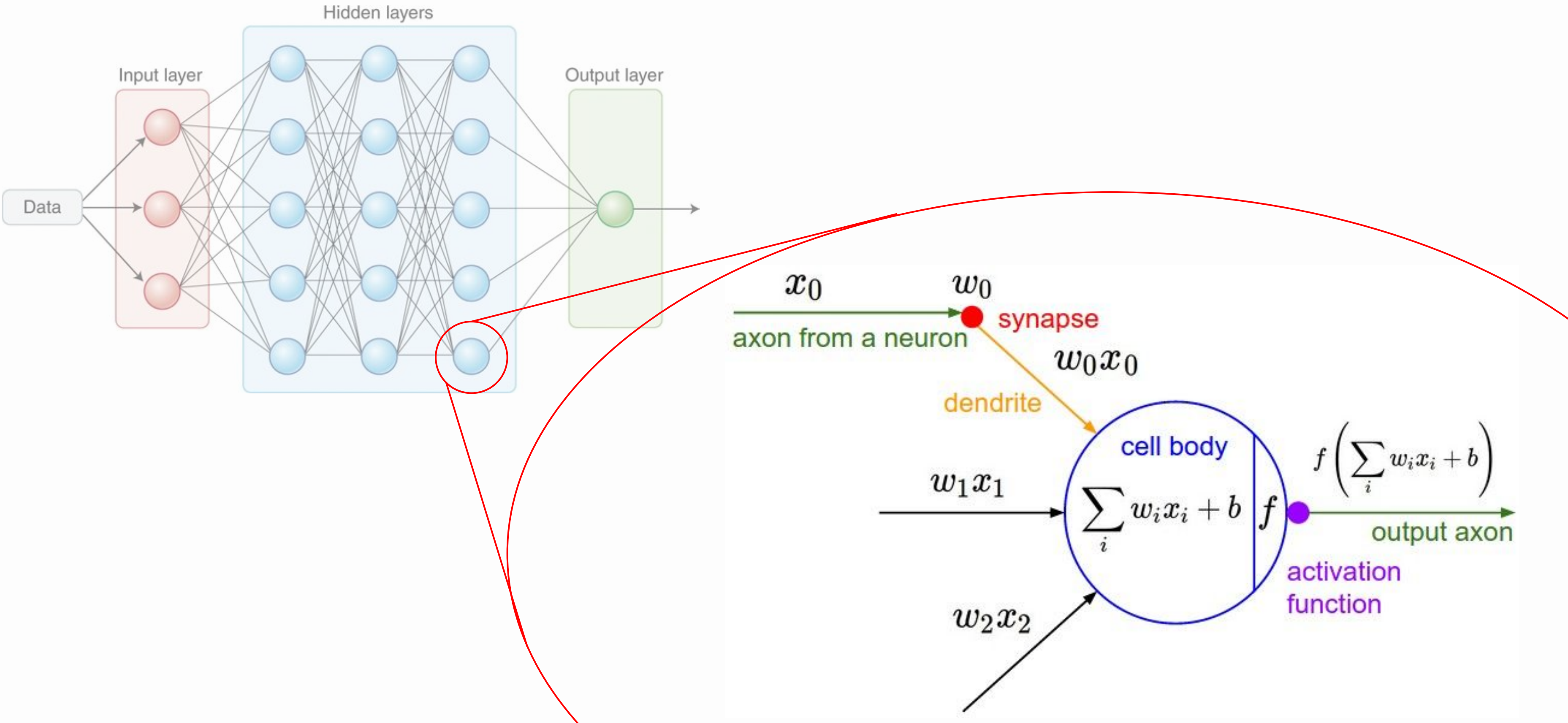
Redes neurais



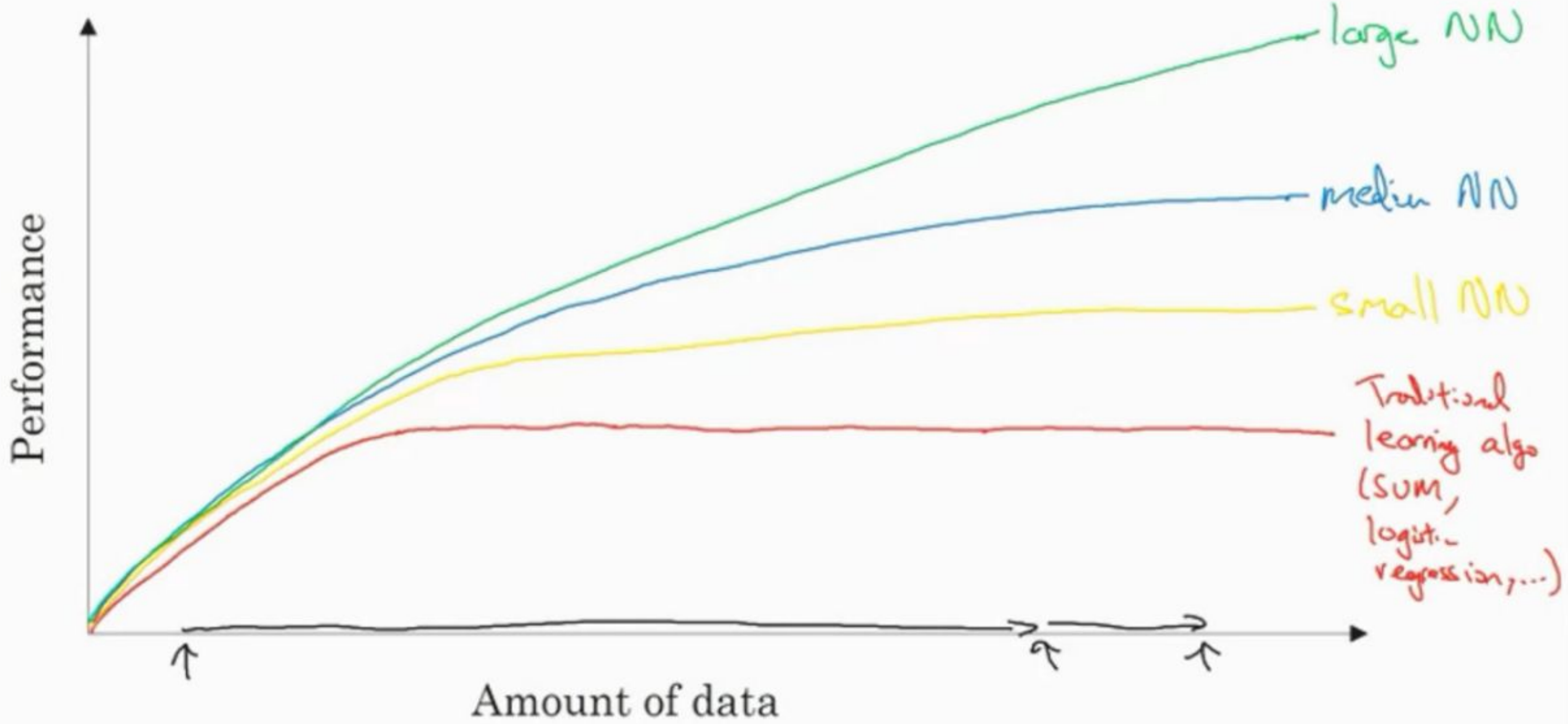
Redes neurais



Redes neurais

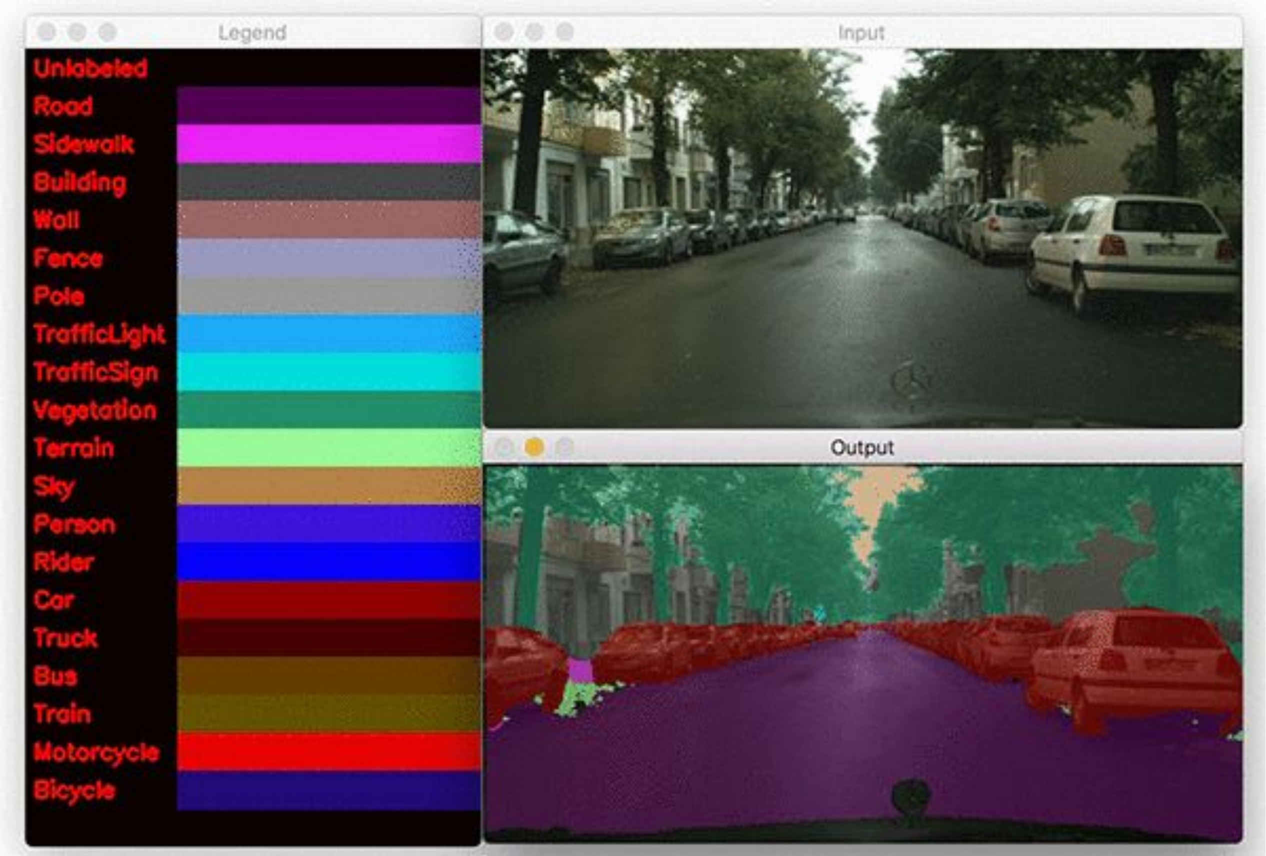


Scale drives deep learning progress

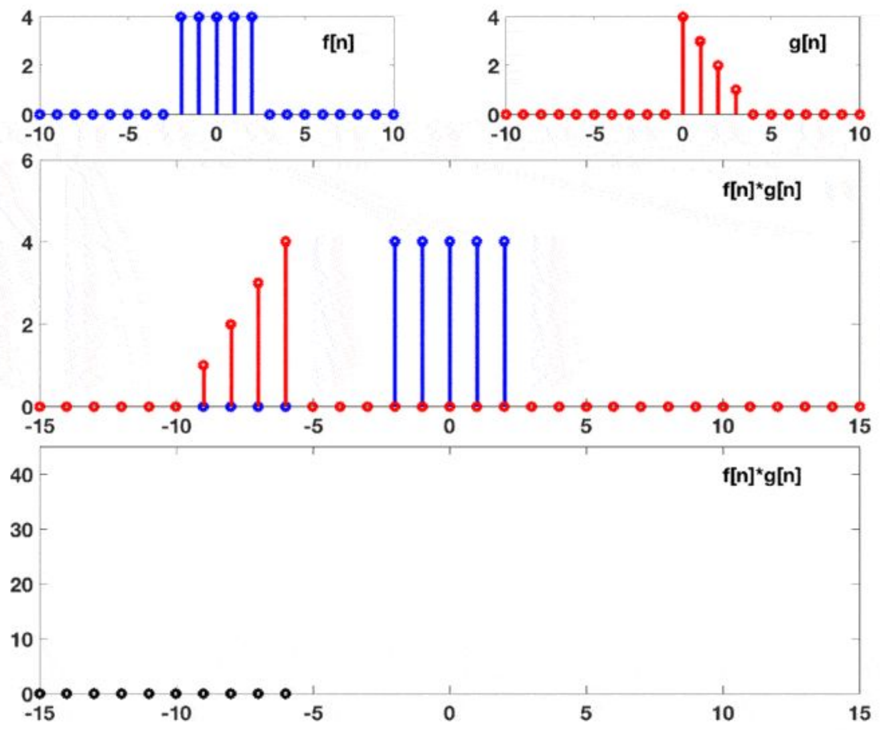


Andrew Ng

Redes neurais convolucionais (CNN)

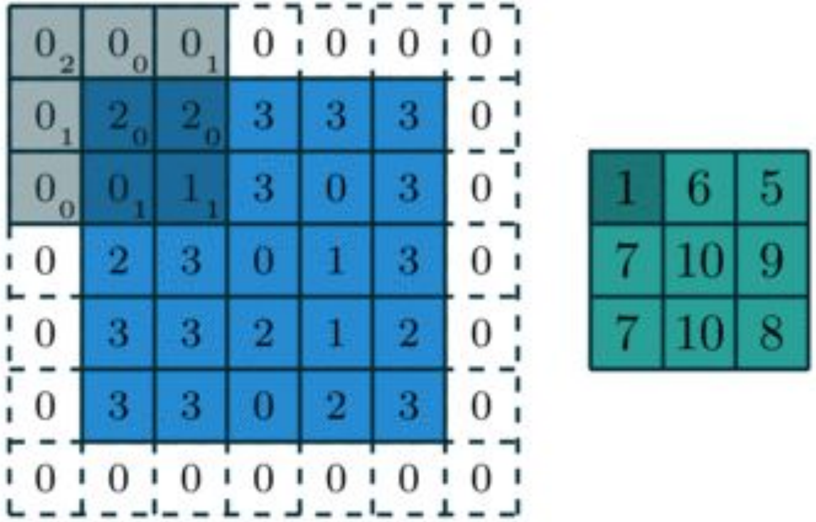


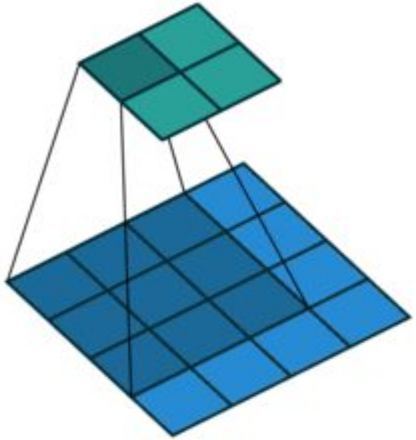
Convolução unidimensional



Bryan Pardo, 2017, Northwestern University EECS 352: Machine Perception of Music and Audio

Convolução bidimensional





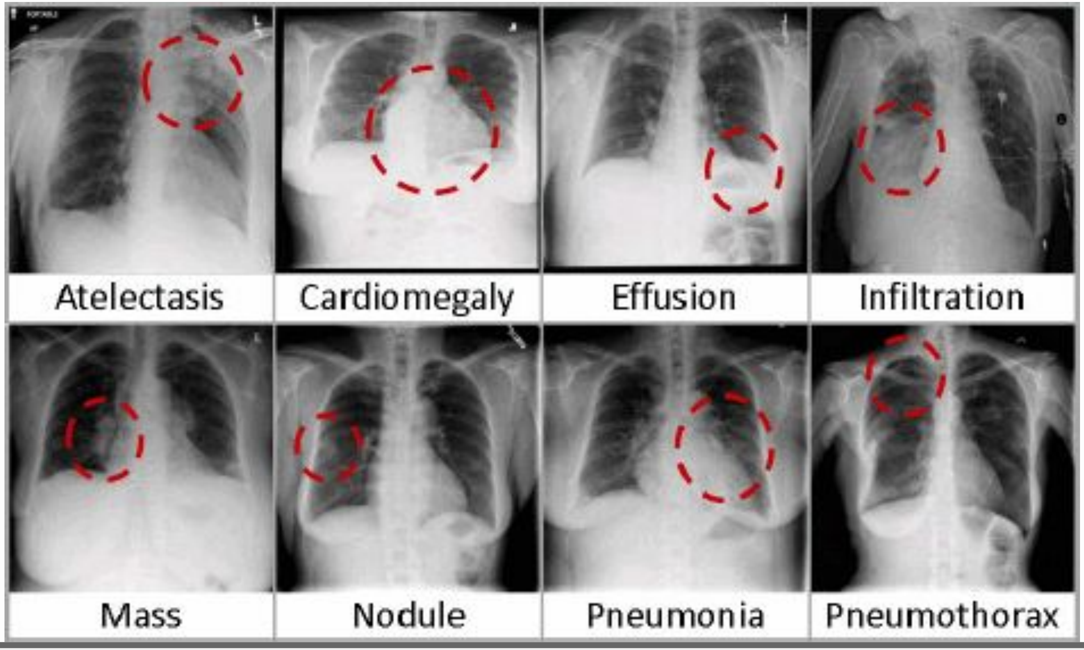
- IA para clínicos
- IA para sistemas de saúde
- IA e pacientes
- IA e análise de dados

Contextos de aplicação

- Exames médicos
- Lâminas de patologia
- Lesões na pele
- Imagens da retina
- Electrocardiogramas
- Endoscopia
- Face
- Sinais vitais

Radiologia

- Tipo mais comum de imagem médica (2 bilhões/ano)



Wang, X. et al. ChestX-ray8: hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases.

Radiologia

Redes Neurais Convolucionais

- 121 camadas
- 112.000 imagens rotuladas

Wang, X. et al. ChestX-ray8: hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases.

Radiologia

Resultados:

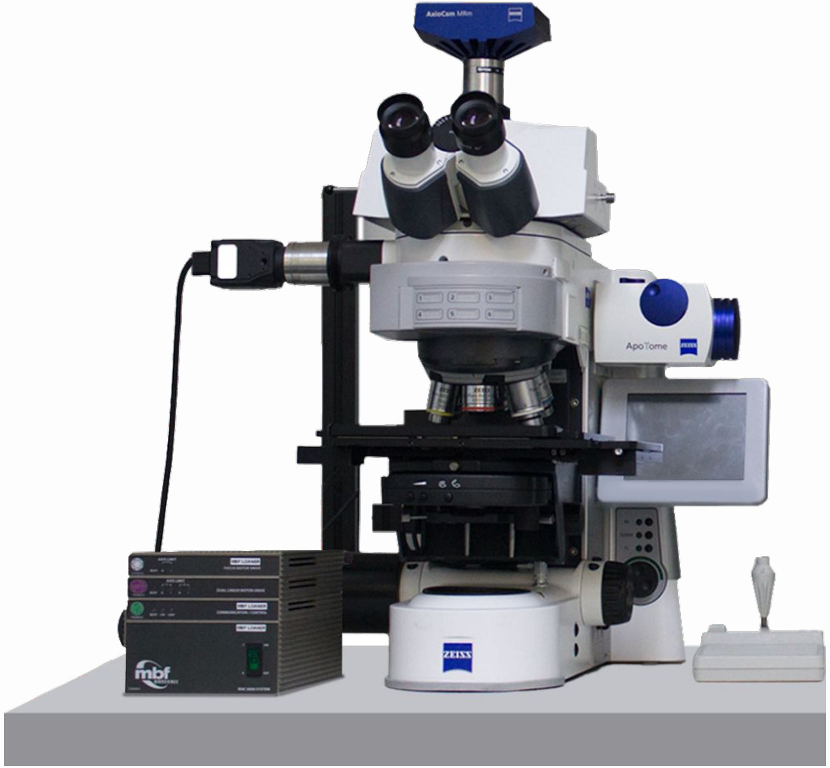
- AUC: 0,76
- Melhor performance que os 4 radiologistas comparados

Desafios:

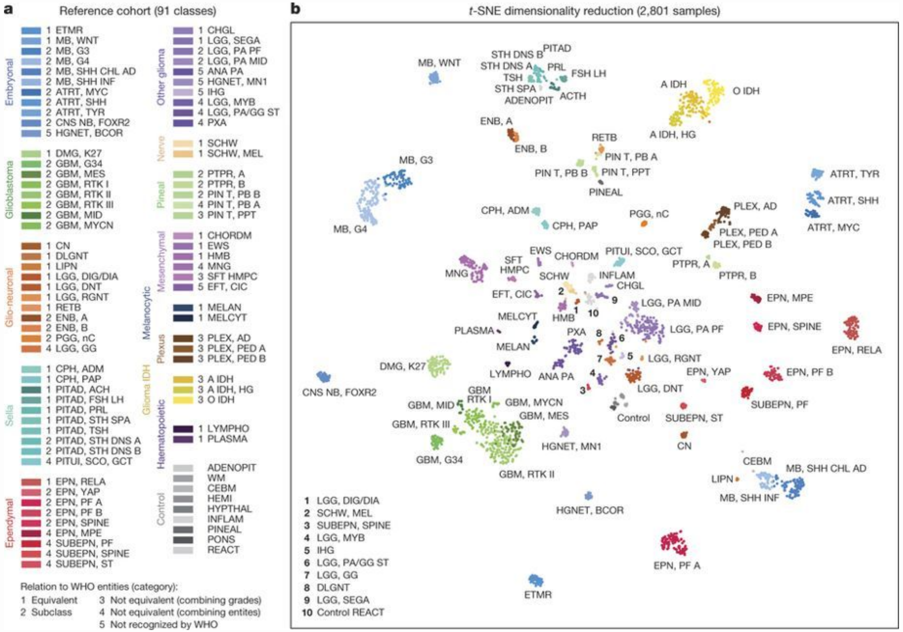
- Analisar múltiplas doenças
- Comparação com mais profissionais

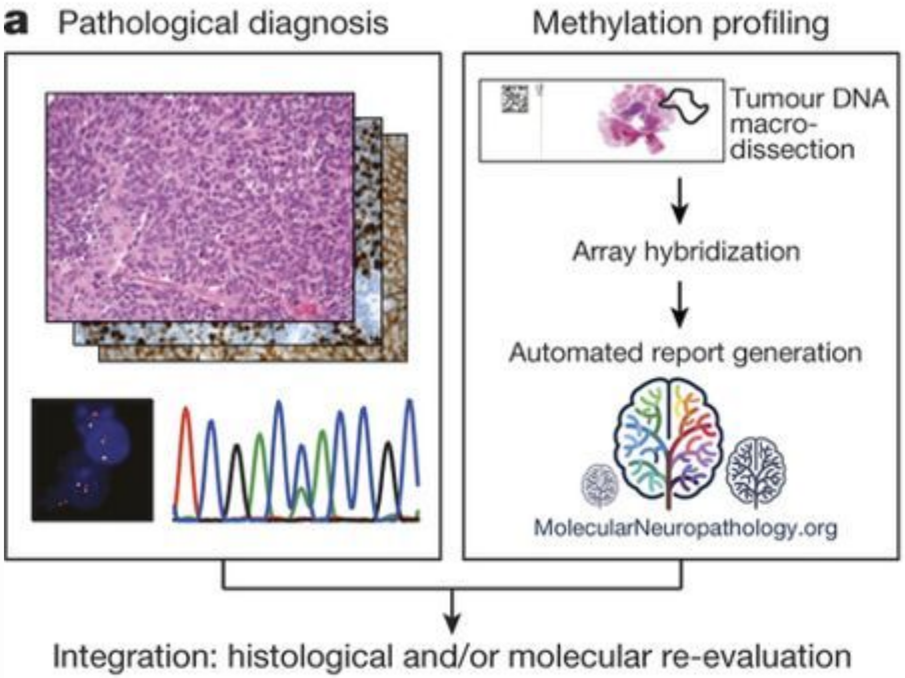
Wang, X. et al. ChestX-ray8: hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases.

Patologia



Patologia





Capper, D. et al. DNA methylation–based classification of central nervous system tumours. *Nature* **555**, 469–474 (2018).

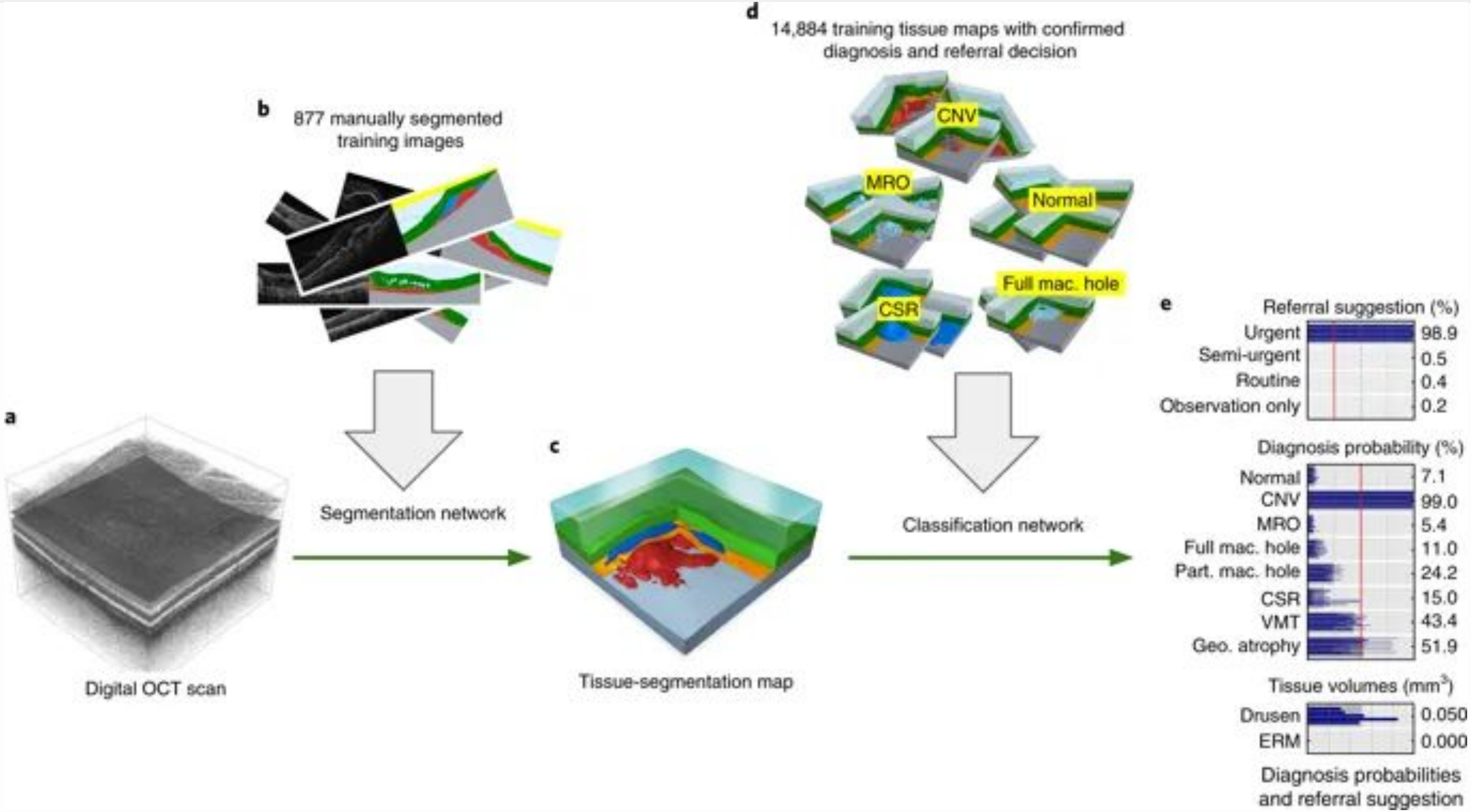
Table 2 | FDA AI approvals are accelerating

Company	FDA Approval	Indication
Apple	September 2018	Atrial fibrillation detection
Aidoc	August 2018	CT brain bleed diagnosis
iCAD	August 2018	Breast density via mammography
Zebra Medical	July 2018	Coronary calcium scoring
Bay Labs	June 2018	Echocardiogram EF determination
Neural Analytics	May 2018	Device for paramedic stroke diagnosis
IDx	April 2018	Diabetic retinopathy diagnosis
Icometrix	April 2018	MRI brain interpretation
Imagen	March 2018	X-ray wrist fracture diagnosis
Viz.ai	February 2018	CT stroke diagnosis
Arterys	February 2018	Liver and lung cancer (MRI, CT) diagnosis
MaxQ-AI	January 2018	CT brain bleed diagnosis
Alivecor	November 2017	Atrial fibrillation detection via Apple Watch
Arterys	January 2017	MRI heart interpretation

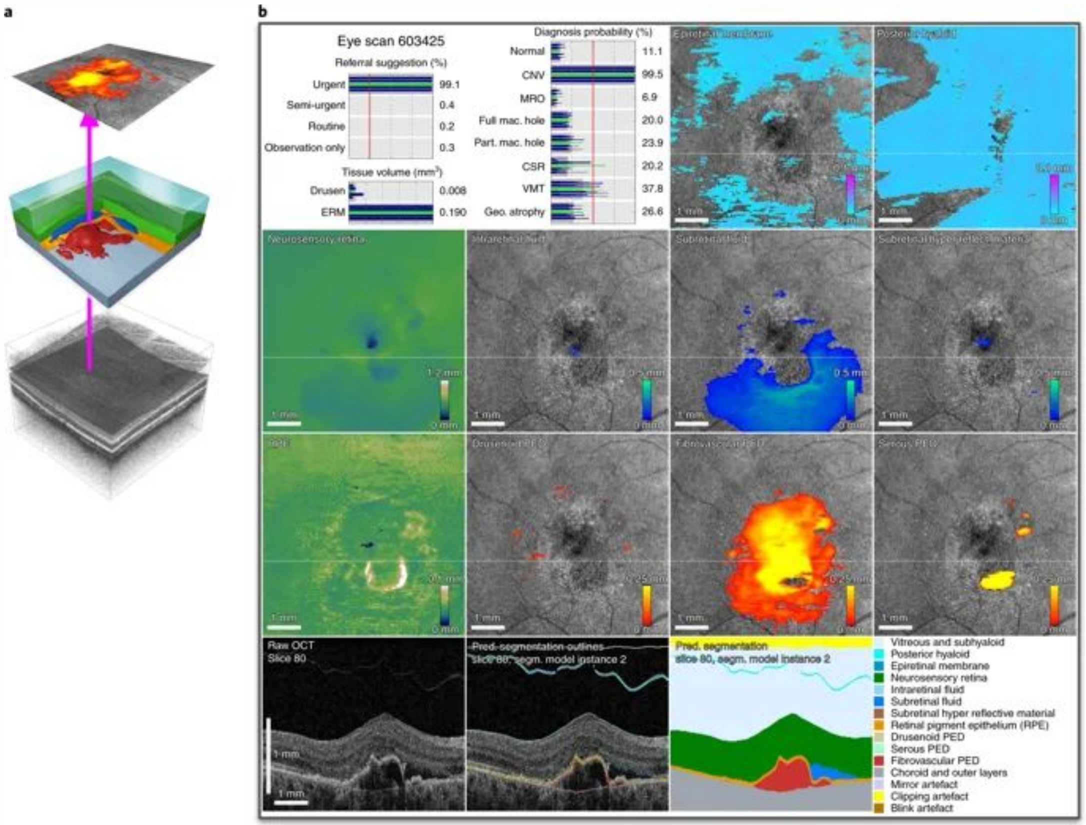
Oftalmologia

- Deep learning aplicado a imagens OCT
- 997 pacientes
- 50 patologias na retina

De Fauw, J. et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat. Med*. **24**, 1342–1350 (2018).



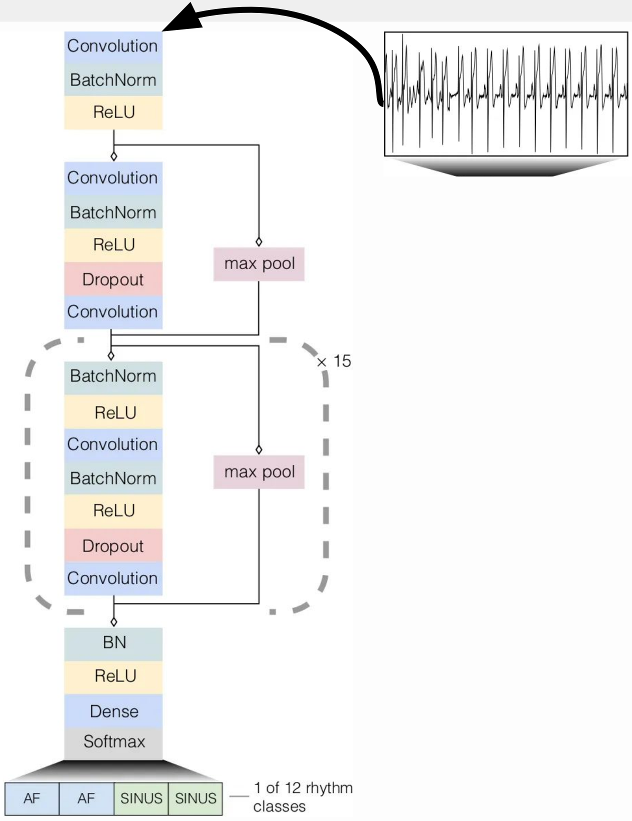
De Fauw, J. et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat. Med*. **24**, 1342–1350 (2018).



De Fauw, J. et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat. Med.* **24**, 1342–1350 (2018).

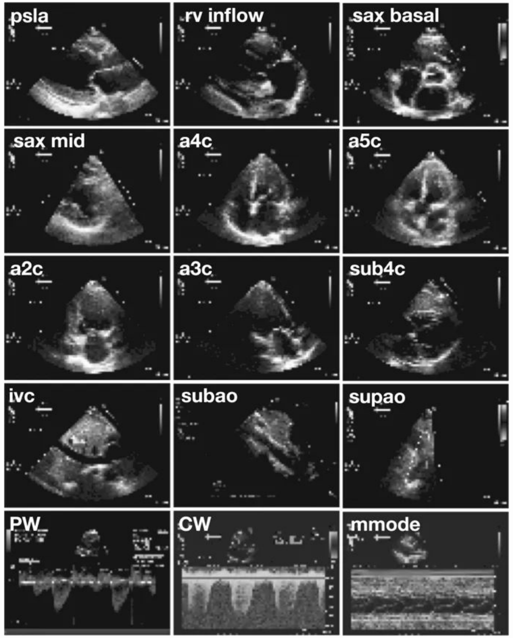
Cardiologia

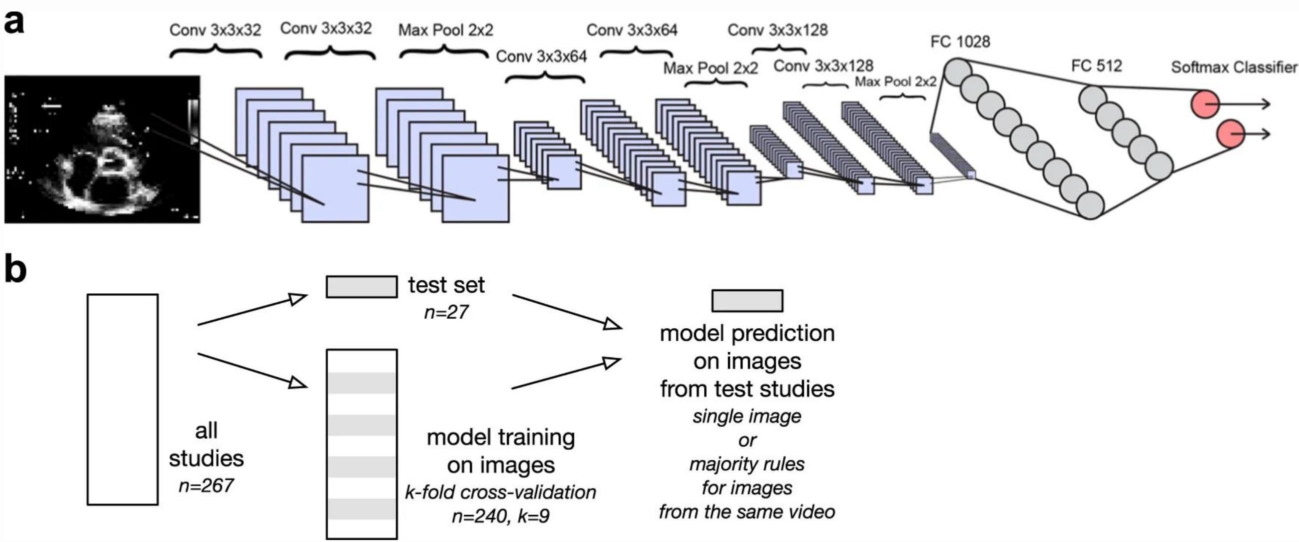
- Detecção e classificação de 12 arritmias
- 91,232 ECGs
- Validação com teste em base independente, com rótulos obtidos por comitê de cardiologistas
- Performance:
  - AUC 0,97
  - F1-score: 0,837 (DNN) x 0,78 (cardiologistas)



Rajpurkar, P. et al. Cardiologist-level arrhythmia detection with convolutional neural networks.

Cardiologia





Madani, A.. et al. Fast and accurate view classification of echocardiograms using deep learning. *NPJ Digit. Med.* **1**, 6 (2018).

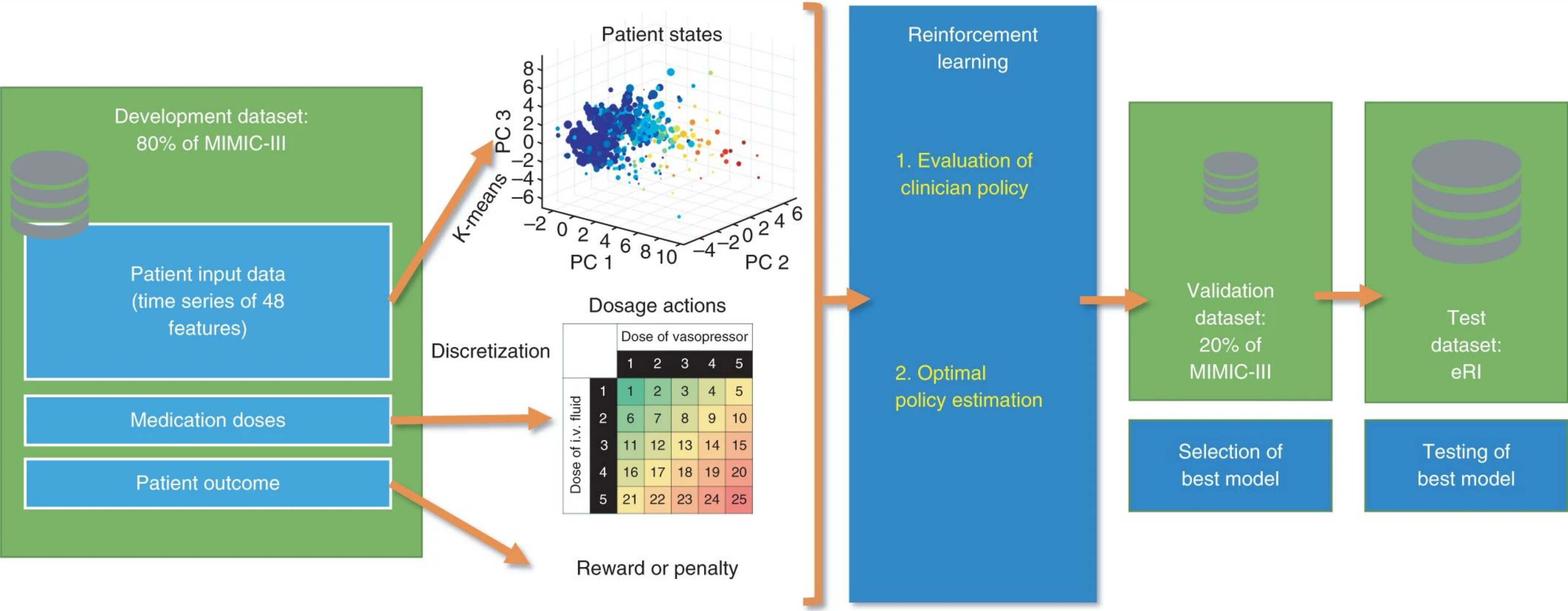
ECG
Embryo selection for IVF
Genome interpretation sick newborns
Voice medical coach via a smart speaker (like Alexa)
K+
Mental health
Paramedic dx of heart attack, stroke
Assist reading of scans, slides, lesions
Prevent blindness
Classify cancer, identify mutations
Promote patient safety
Predict death in-hospital

IA e sistemas de saúde

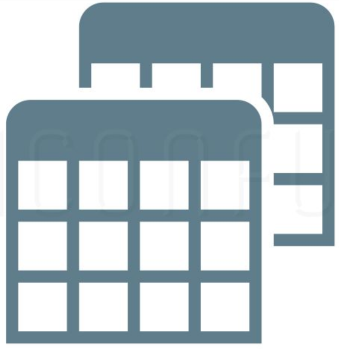
Table 3 | Selected reports of machine- and deep-learning algorithms to predict clinical outcomes and related parameters

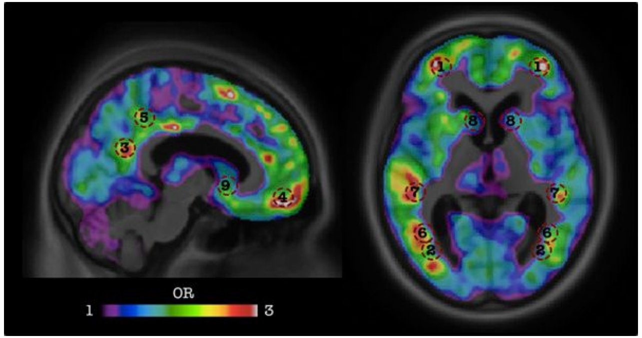
Prediction	*n*	AUC	Publication (Reference number)
In-hospital mortality, unplanned readmission, prolonged LOS, final discharge diagnosis	216,221	0.93*0.75+0.85#	Rajkomar et al.[96]
All-cause 3-12 month mortality	221,284	0.93^	Avati et al.[91]
Readmission	1,068	0.78	Shameer et al.[106]
Sepsis	230,936	0.67	Horng et al.[102]
Septic shock	16,234	0.83	Henry et al.[103]
Severe sepsis	203,000	0.85@	Culliton et al.[104]
Clostridium difficile infection	256,732	0.82++	Oh et al.[93]
Developing diseases	704,587	range	Miotto et al.[97]
Diagnosis	18,590	0.96	Yang et al.[90]
Dementia	76,367	0.91	Cleret de Langavant et al.[92]
Alzheimer's Disease (+ amyloid imaging)	273	0.91	Mathotaarachchi et al.[98]
Mortality after cancer chemotherapy	26,946	0.94	Elfiky et al.[95]
Disease onset for 133 conditions	298,000	range	Razavian et al.[105]
Suicide	5,543	0.84	Walsh et al.[86]
Delirium	18,223	0.68	Wong et al.[100]

LOS, length of stay; *n*, number of patients (training+validation datasets). For AUC values: *, in-hospital mortality; +, unplanned readmission; #, prolonged LOS; ^, all patients; @, structured+unstructured data; ++, for University of Michigan site.



Komorowski, M. et al. The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. *Nat. Med*. **24**, 1716–1720 (2018).

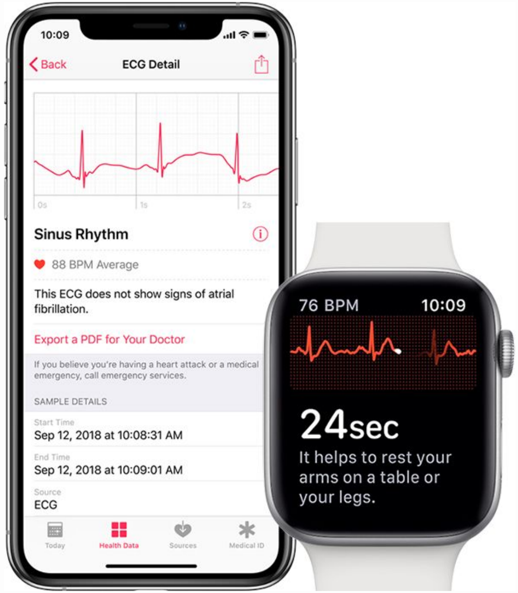




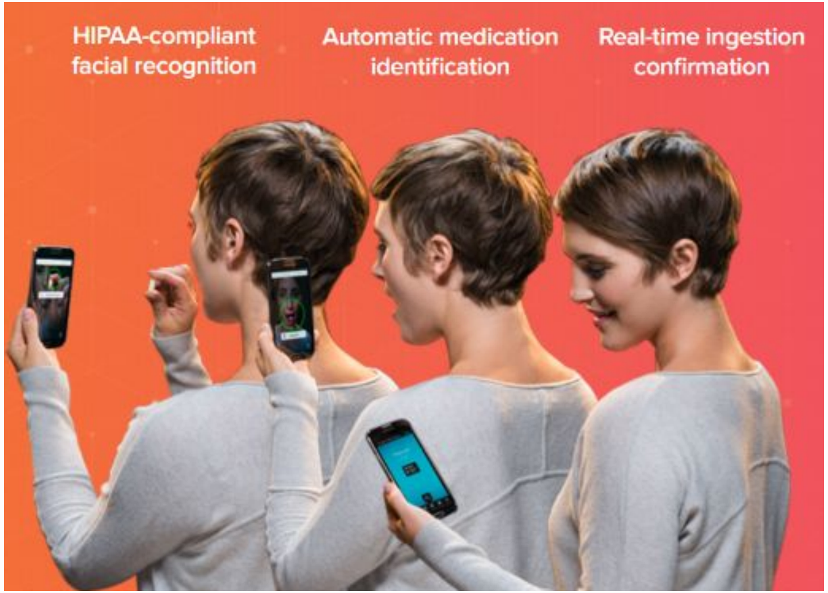
Mathotaarachchi, S. et al. Identifying incipient dementia individuals using machine learning and amyloid imaging. *Neurobiol. Aging.* **59**, 80–90 (2017).

IA e pacientes

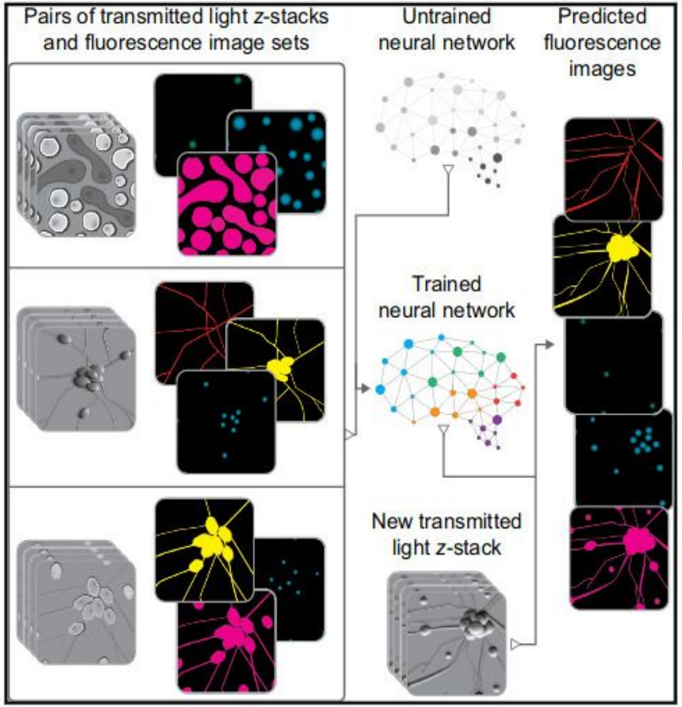
- Apple Watch

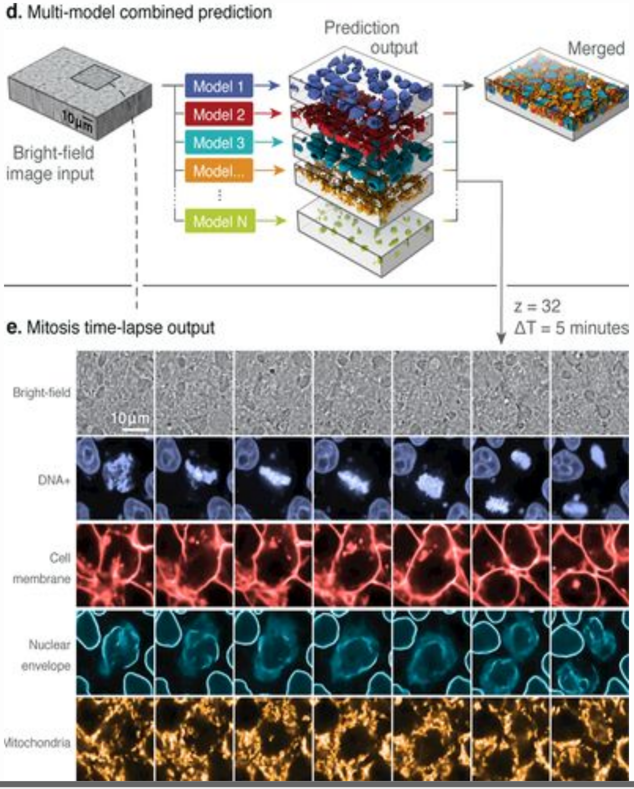


- AiCure



IA e análise de dados





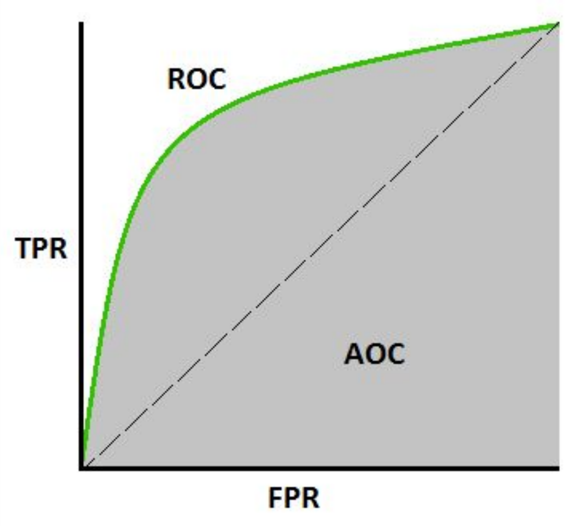
Christiansen, E. M. et al. In silico labeling: predicting fluorescent labels in unlabeled images. *Cell* **173**, 792–803 e719 (2018).

Limitações e desafios

- Dificuldade comparação entre algoritmos
- Uso de métricas de desempenho

Limitações e desafios

- Métrica para avaliação

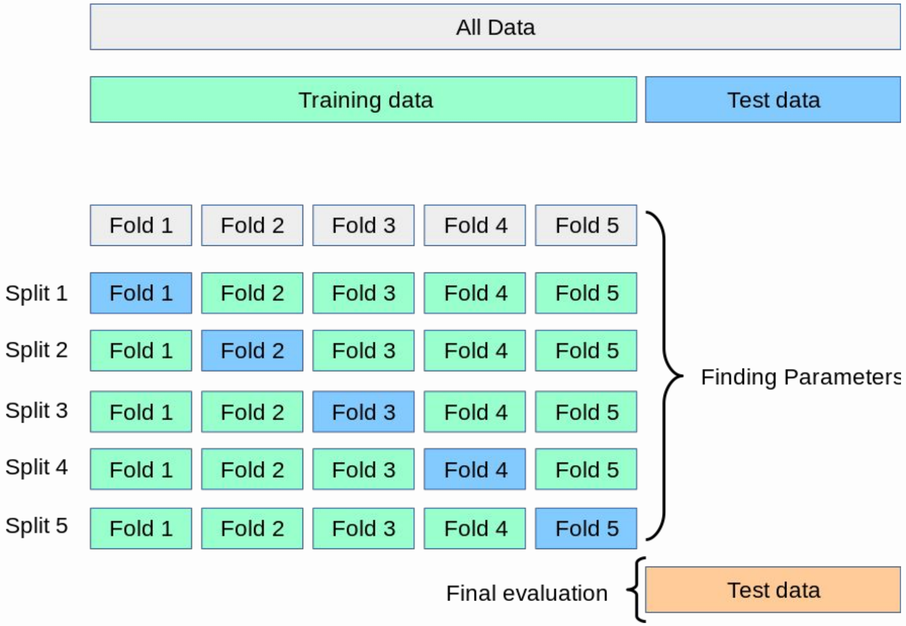


- Validação cruzada k-fold

Limitações e desafios

- Métrica para avaliação

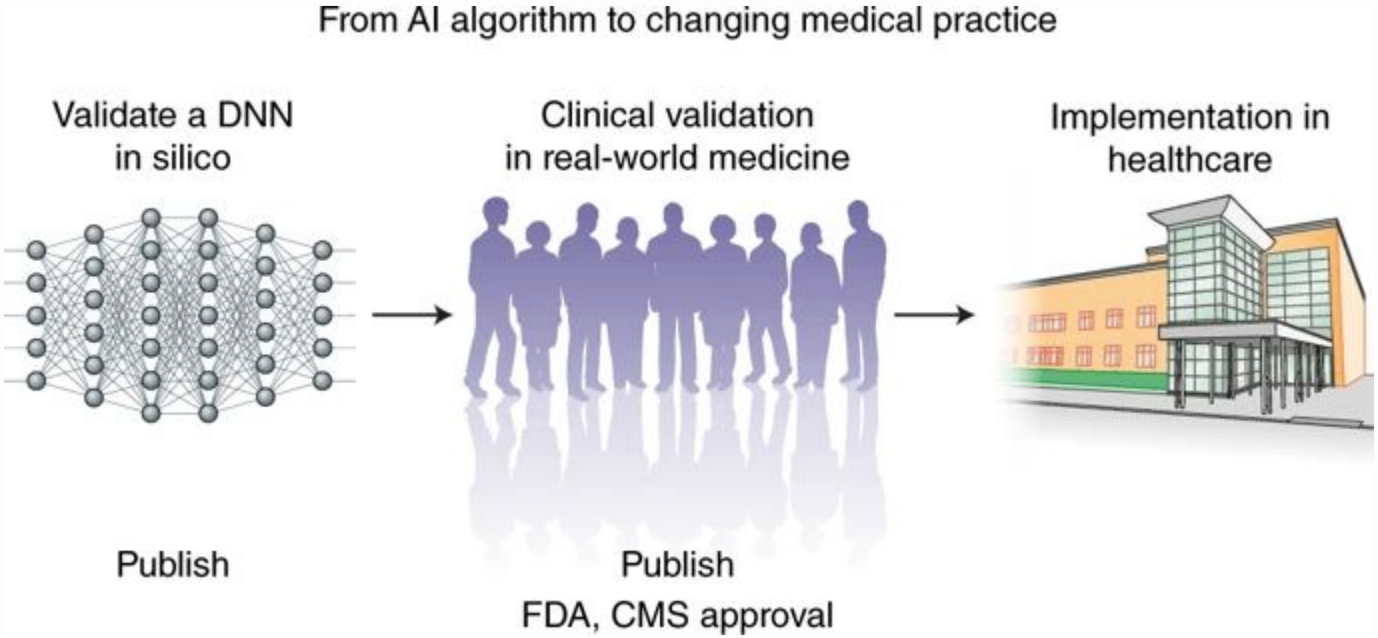
- Validação cruzada k-fold



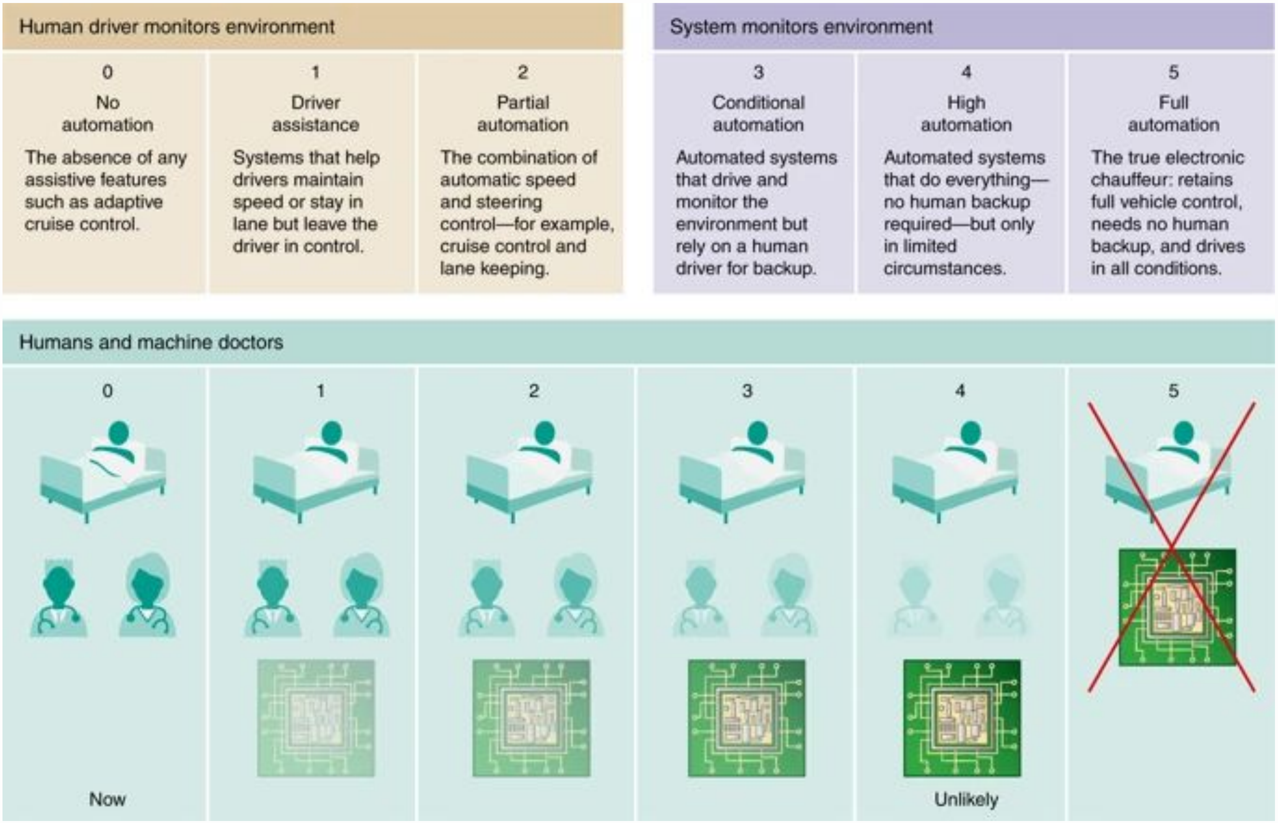
Limitações e desafios

- IBM Watson
- Explicabilidade do modelo

Considerações futuras



Considerações futuras



[Extra] Revisão sistemática de pesquisas

A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis

Xiaoxuan Liu*, Livia Faes*, Aditya U Kale, Siegfried K Wagner, Dun Jack Fu, Alice Bruynseels, Thushika Mahendiran, Gabriella Moraes, Mohith Shamdas, Christoph Kern, Joseph R Ledsam, Martin K Schmid, Konstantinos Balaskas, Eric J Topol, Lucas M Bachmann, Pearse A Keane, Alastair K Denniston

Summary

Background Deep learning offers considerable promise for medical diagnostics. We aimed to evaluate the diagnostic accuracy of deep learning algorithms versus health-care professionals in classifying diseases using medical imaging.

Methods In this systematic review and meta-analysis, we searched Ovid-MEDLINE, Embase, Science Citation Index, and Conference Proceedings Citation Index for studies published from Jan 1, 2012, to June 6, 2019. Studies comparing the diagnostic performance of deep learning models and health-care professionals based on medical imaging, for any disease, were included. We excluded studies that used medical waveform data graphics material or investigated the accuracy of image segmentation rather than disease classification. We extracted binary diagnostic accuracy data and constructed contingency tables to derive the outcomes of interest: sensitivity and specificity. Studies undertaking an out-of-sample external validation were included in a meta-analysis, using a unified hierarchical model. This study is registered with PROSPERO, CRD42018091176.

Lancet Digital Health 2019

Published Online
September 24, 2019
https://doi.org/10.1016/S2589-7500(19)30123-2

See Online/Comment
https://doi.org/10.1016/S2589-7500(19)30124-4

*Joint first authors.

Department of Ophthalmology, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK

[Extra] Revisão sistemática de pesquisas

	Target condition	Reference standard	Same method for assessing reference standard across samples	Type of internal validation	External validation
Abbasi-Sureshjani et al (2018)[24]	Diabetes	Histology	Yes	Random split sample validation	No
Adams et al (2019)[25]	Hip fracture	Surgical confirmation	Yes	Random split sample validation	No
Ardila et al (2019)[19]	Lung cancer	Histology; follow-up	No	NR	Yes
Ariji et al (2019)[26]	Lymph node metastasis	Histology	Yes	Resampling method	No
Ayed et al (2015)[27]	Breast tumour	Histology	Yes	Random split sample validation	No
Becker et al (2017)[28]	Breast tumour	Histology; follow-up	No	Study 1: NA Study 2: temporal split-sample validation	Yes
Becker et al (2018)[20]	Breast tumour	Histology; follow-up	No	Random split sample validation	No

Table 1: Participant demographics for the 82 included studies

Liu, X. et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *The Lancet Digital Health*. (2019).

[Extra] Revisão sistemática de pesquisas

	Subspecialty	Participants				
		Inclusion criteria	Exclusion criteria	Mean age (SD; range), years	Percentage of female participants	Number of participants represented by the training data
(Continued from previous page)						
Wang et al (2018)[93]	Lung cancer	Solitary pulmonary nodule, histologically confirmed pre-invasive lesions and invasive adenocarcinomas	Previous chemotherapy or radiotherapy that can cause texture changes; incomplete CT; patients with ≥2 lesions resected	56 (10·6; NR)	81%	NR
Wang et al (2019)[94]	Thyroid cancer	Ultrasound examination with subsequent histological diagnosis	NR	46 (NR; 20–71)	NR	NR
Wright et al (2014)[95]	Nephrology	NR	Equivocal reports; artefacts; bladder inclusion and residual uptake in the ureters; horseshoe kidney	9 (NR; 0–80)	70%	257
Wu et al (2019)[96]	Gastric cancer	Patients undergoing OGD	Age <18 years; residual stomach content	NR	NR	NR
Ye et al (2019)[97]	Trauma and orthopaedics	Patients with ICH	Missing information or serious imaging artefact	Non-ICH :42 (15; 2–82) ICH: 54 (17; 1–98)	Non-ICH: 55% ICH: 35%	NR
Yu et al (2018)[98]	Dermatological cancer	Benign nevi or acral melanoma with histological diagnosis and dermatoscopic images	NR	NR	NR	NR

Table 2: Model training and validation for the 82 included studies

Liu, X. et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *The Lancet Digital Health*. (2019).

[Extra] Revisão sistemática de pesquisas

	Indicator definition			Algorithm			Data source			
	Method for predictor measurement	Exclusion of poor-quality imaging	Heatmap provided	Algorithm architecture name	Algorithm architecture	Transfer learning applied	Number of images for training/ tuning)	Source of data	Data range	Open-access data
Abbasi-Sureshjani et al (2018)[24]	Fundus image	NR	NR	ResNet	CNN; Residual Network	No	7931/NR	Retrospective cohort; secondary analysis of a subset of the Maastricht study—a population-based cohort (collected in the southern part of the Netherlands), enriched with patients with diabetes	2010-17	No
Adams et al (2019)[25]	X-ray	NR	No	AlexNet	CNN; AlexNet	Yes	512/NR	Retrospective cohort; data from the Royal Melbourne Hospital (Melbourne, VIC, Australia) radiographic archive	NR	No
Ardila et al (2019)[19]	CT	No	Yes	Mask RCNN; RetinaNet; Inception V1	CNN; Inception	Yes	10 396/2228	Retrospective clinical trial data from the National Lung Cancer Screening Trial	2002–04	No
Ariji et al (2019)[26]	Contrast-enhanced CT	NR	No	CNN	CNN	No	NR/NR	Retrospective cohort; data from the Aichi-Gakuin University School of Dentistry (Nagoya, Japan)	2007, 2015	No

Table 3: Indicator, algorithm, and data source for the 82 included studies

Liu, X. et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *The Lancet Digital Health*. (2019).