Enjeux de l'Intelligence Artificielle en médecine

Pr Jean-Emmanuel Bibault
Service d'Oncologie Radiothérapie, HEGP, Paris
INSERM UMR 1138, Centre de Recherche des Cordeliers

Email : jean-emmanuel.bibault@aphp.fr
Twitter : @jebibault

ASSISTANCE PUBLIQUE HÔPITAUX DE PARIS

Hôpital européen Georges-Pompidou AP-HP

Université Paris Cité

Inserm

CARPEM CANCER RESEARCH FOR PERSONALIZED MEDICINE

1 - Définitions et concepts

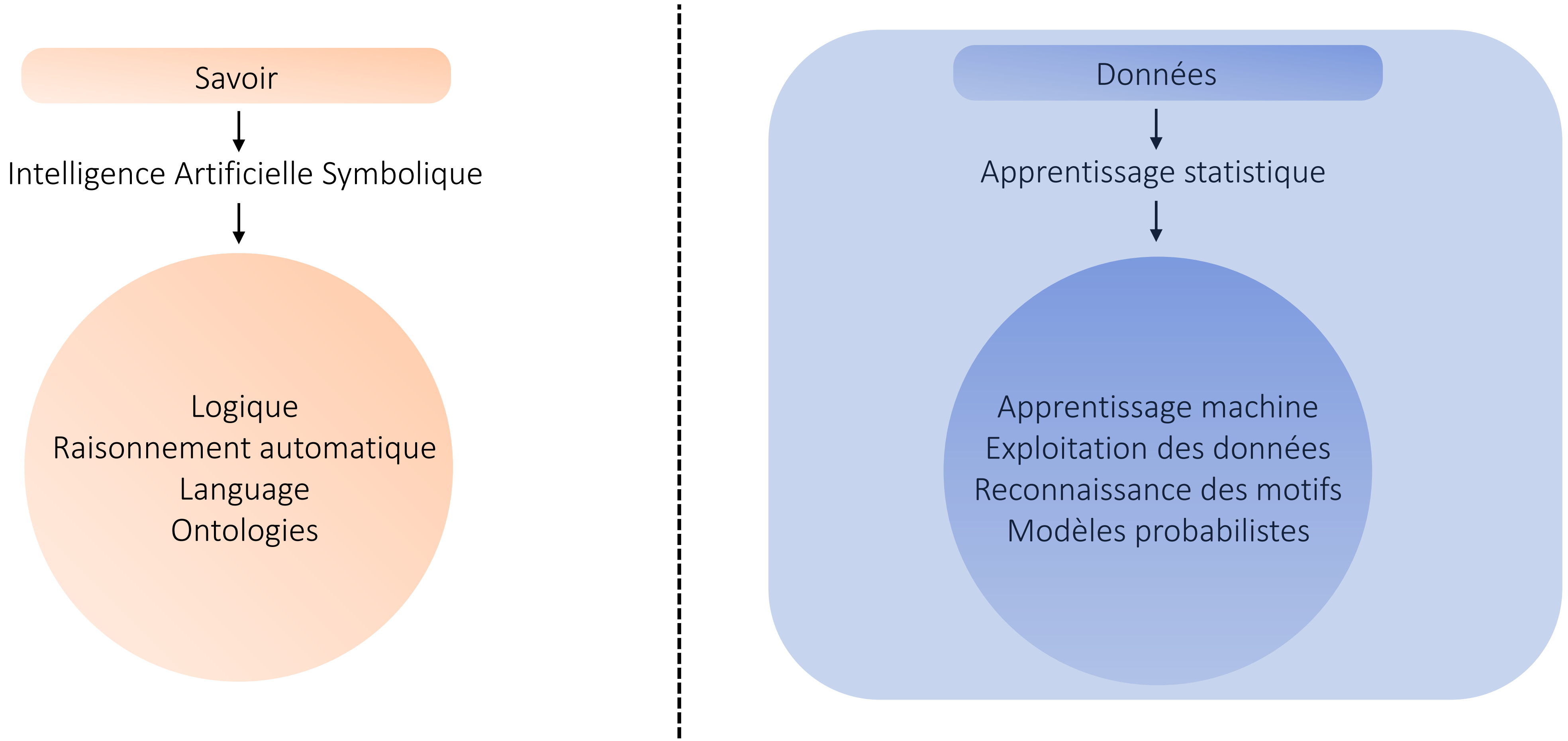
Intelligence Artificielle

- Ensemble de théories et de techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence

- Statistique vs déterministe

- En médecine :
  - Système d’aide à la décision
  - Agent conversationnel
  - Classifieur / Prédicteur

Intelligence Artificielle



Historique

- 1940 : Premier concept de réseaux neuronaux
- Cycles de promesses et de déceptions

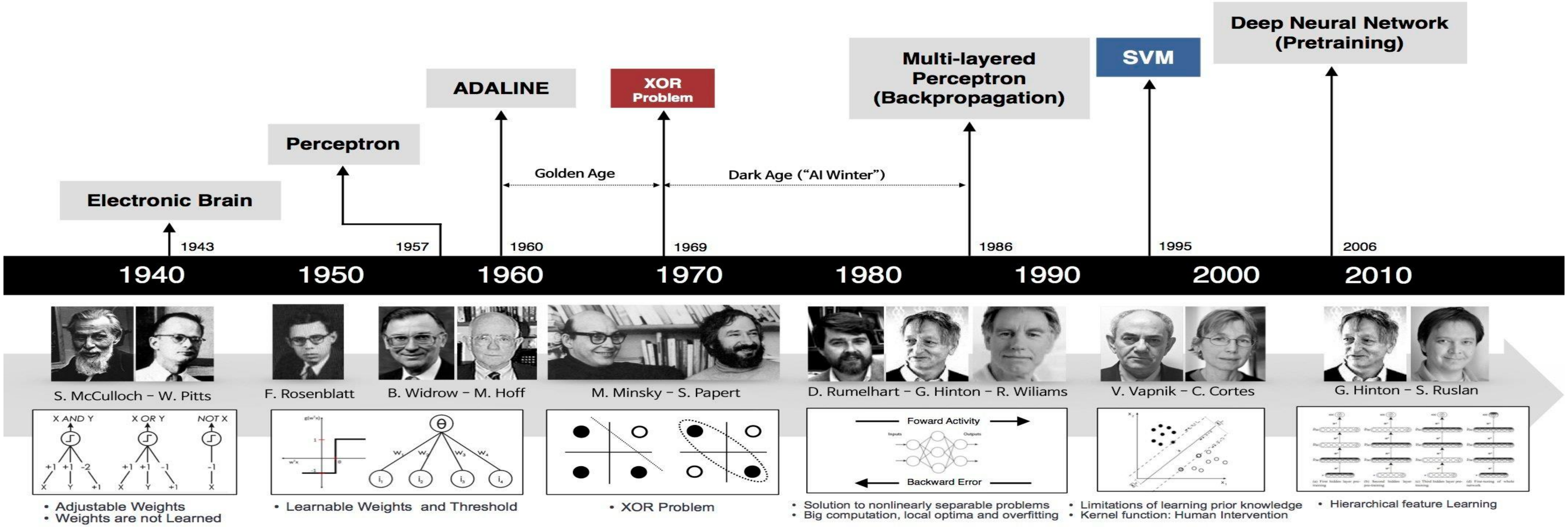
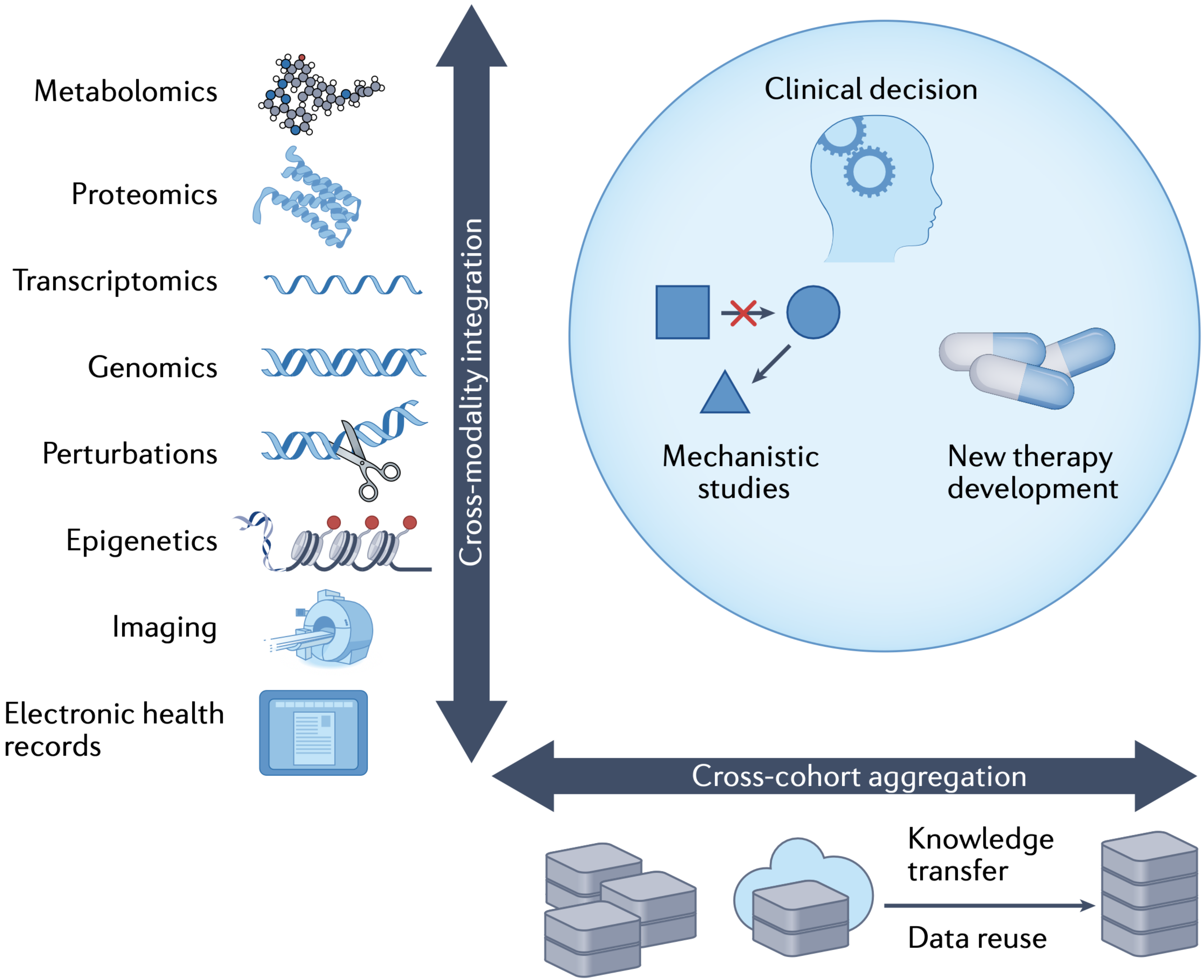
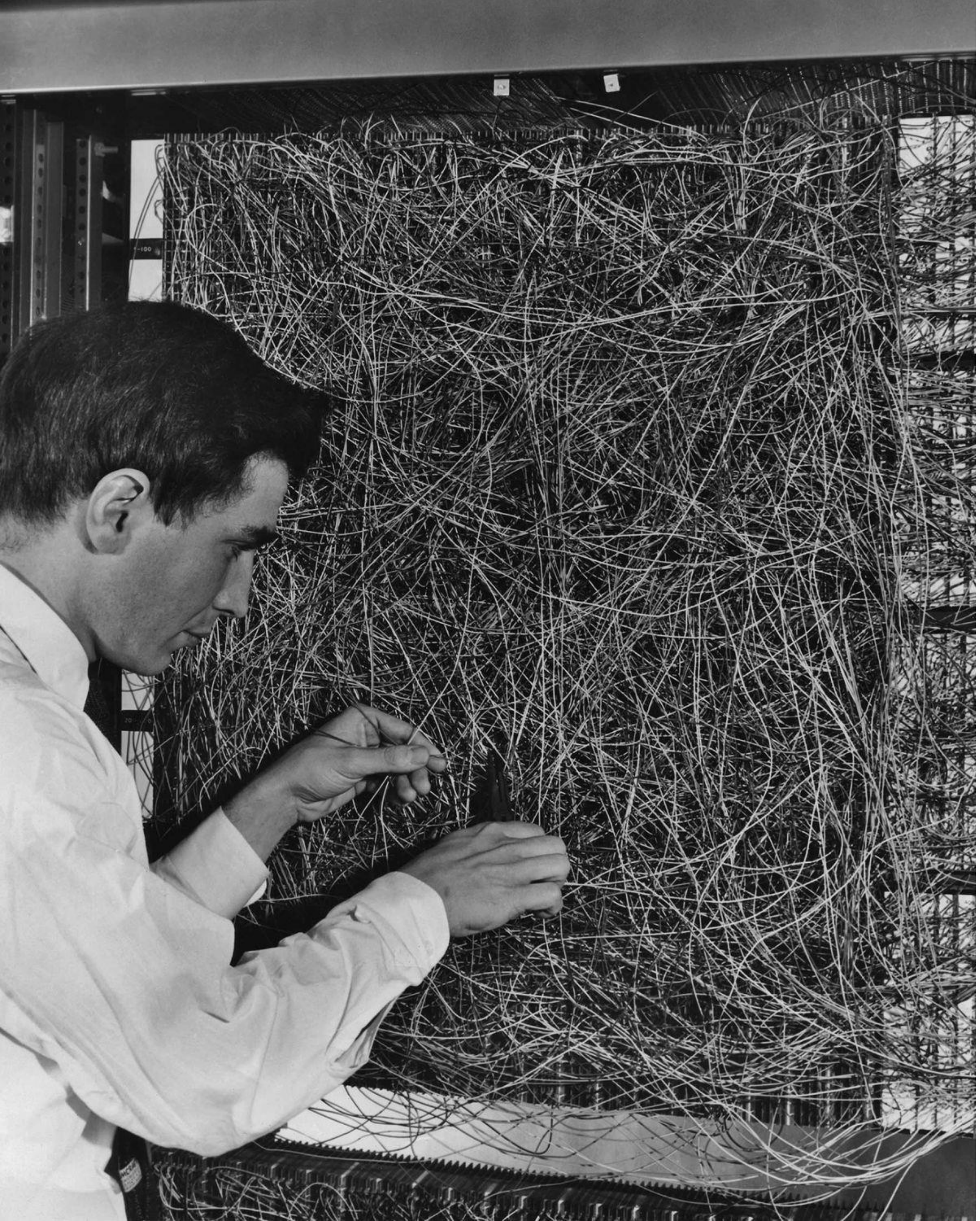


Image source : Andrew Beam



Jiang P. et al, Nature Reviews, 2022



TensorFlow™

PYTORCH

scikit learn

Keras
A deep learning library

Democratising medicine

The Economist

The crowd will see you now
The computer will see you now

Your Smartphone Will See You Now

THE WALL STREET JOURNAL.
WSJ

Dr. Google Will See You Now

The New York Times

IBM'S WATSON IS READY TO SEE YOU NOW

FAST COMPANY

The Robot Will See You Now

the Atlantic

The Avatar Will See You Now

MIT Technology Review

2 - Méthodes de machine learning

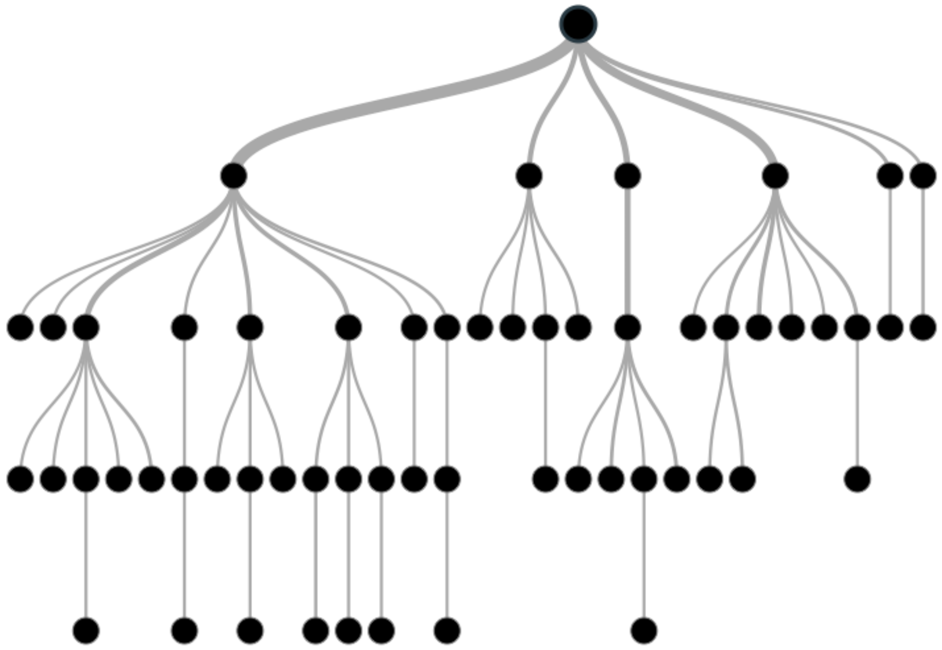
Différentes méthodes

Algo	Type	Tolerance number features	Parametrization	Memory size	Minimal required quantity	Com m	Overfitting Tendency	Difficulty	Time for Learning	Time for predicting
Linear Regression	R	Weak	Weak	Small	Small	++	Low	Weak	Weak	Weak
Logistic Regression	C	Weak	Simple	Small	Small	++	Low	Weak	Weak	Weak
Decision Tree	R & C	Strong	Simple / intuitive	Large	Small	+++	Very high	Weak	Weak	Weak
Random Forest	R & C	Strong	Simple / intuitive	Very Large	Large	++	Average	Average	Costly	Costly
Boosting	R & C	Strong	Simple / intuitive	Very Large	Large	+	Average	Average	Costly	Weak
Naive Bayes	C	Weak	No params.	Small	Small	++	Low	Weak	Weak	Weak
SVM	C	Very strong	Not intuitive	Small	Large	--	Average	High	Costly	Weak
Neural Network (NN)	C	Very strong **	Not intuitive	Inter	Large	---	Average	Very high	Costly	Weak
Deep Neural Network	C	Very strong **	Not intuitive	Very Large	Very Large	---	High	Very high	Very costly	Weak
K-Means	CL*	Strong	Simple / Intuitive		Small	+		High	Weak	
One class SVM	A	Very strong	Not intuitive	Weak	Large	--	Average	High	Costly	Weak

Decision Tree / Random Forest

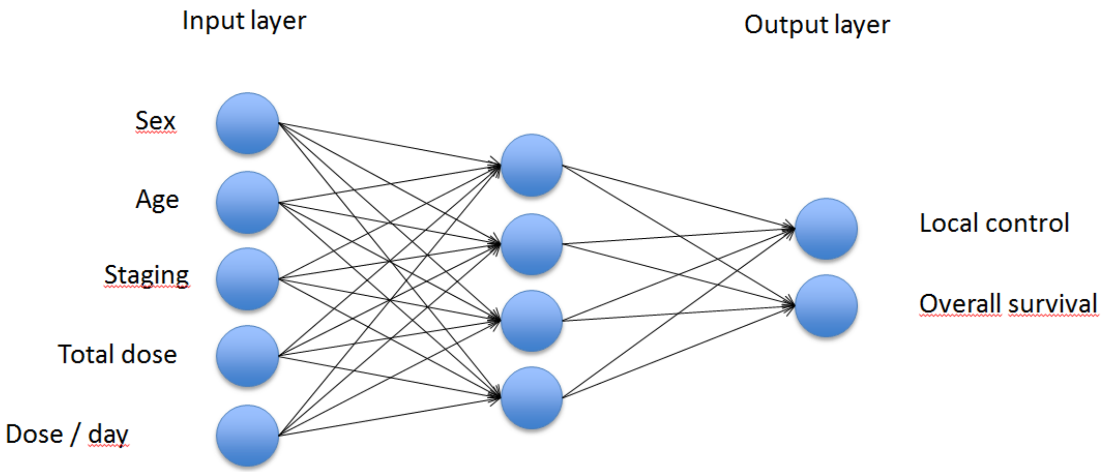
- Chaque embranchement correspond à des questions binaires à réponse unique

- Permet de classifier des sujets à partir de leurs caractéristiques



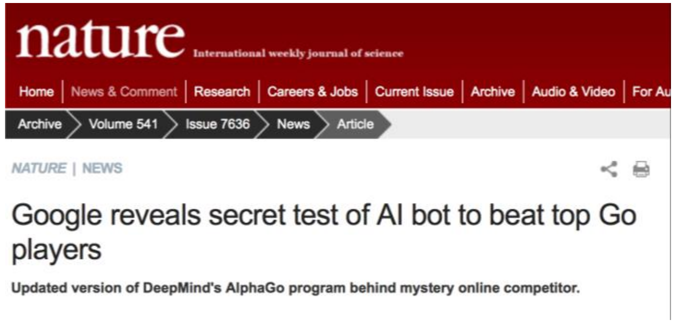
Neural Network

- Chaque nœud d'entrée correspond à une variable, avec un poids et un biais
- Apprentissage : le réseau converge pour prédire les classes de sortie à partir des données connues



Deep Learning

- 2006 : apparition du terme
- 2012 : première percée
  - Large Scale Visual Recognition Challenge (Imagenet database)
  - AlexNet fait passer le taux d’erreur de 28 à 16%
- 2014 : GoogLeNet
- 2015 : Inception
- 2016 : ResNet (56 couches)


nature International weekly journal of science

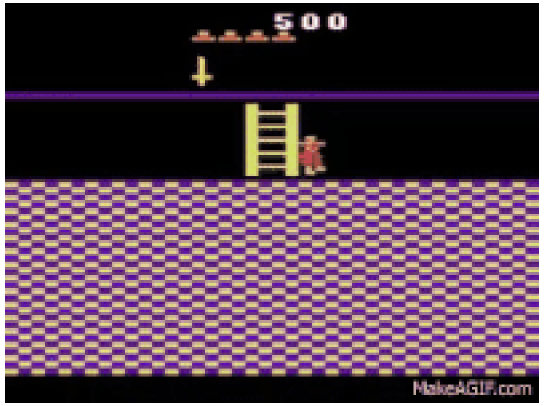
Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For Au

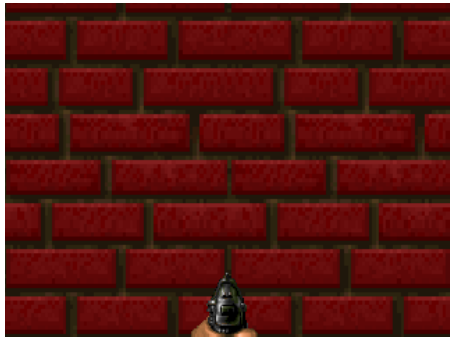
Archive > Volume 541 > Issue 7636 > News > Article

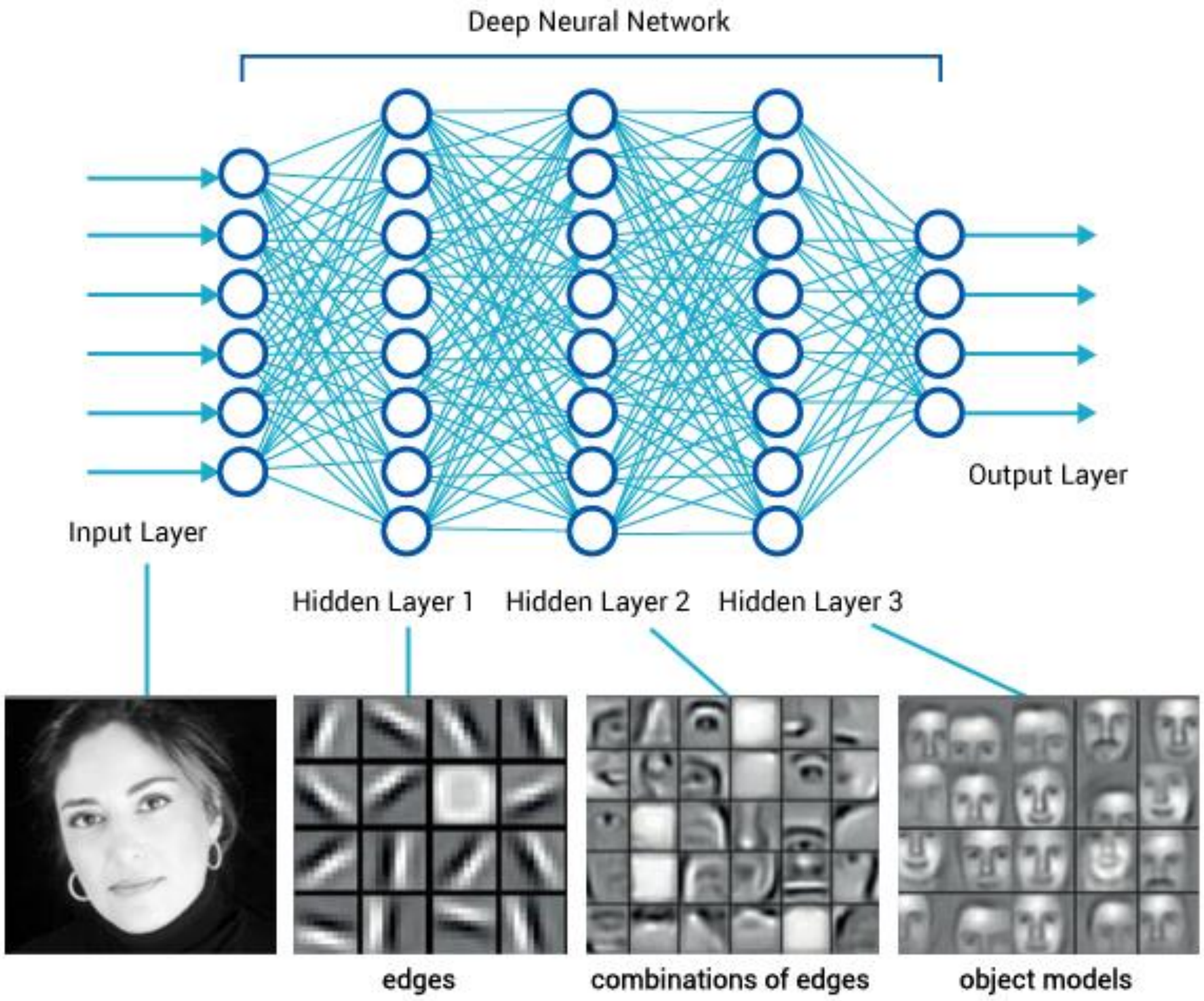
NATURE | NEWS

Google reveals secret test of AI bot to beat top Go players

Updated version of DeepMind's AlphaGo program behind mystery online competitor.





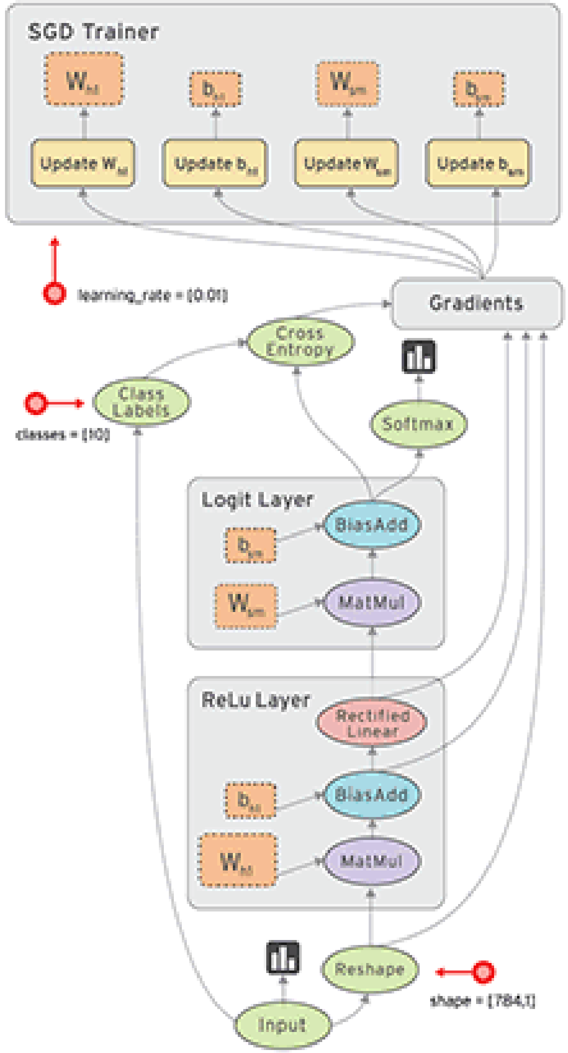


Source image : https://www.amax.com/blog/?p=804

Frameworks

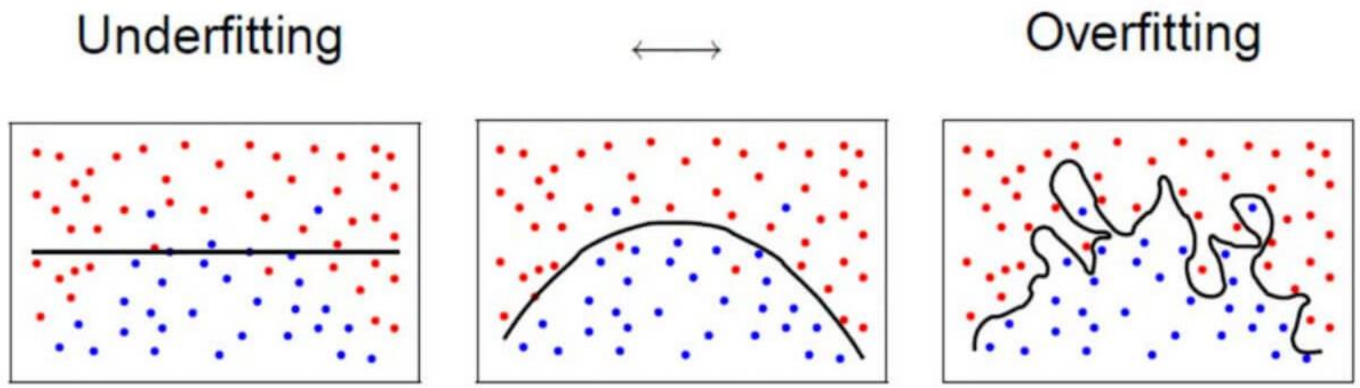
	Languages	Tutorials and training materials	CNN modeling capability	RNN modeling capability	Architecture: easy-to-use and modular front end	Speed	Multiple GPU support	Keras compatible
Theano	Python, C++	++	++	++	+	++	+	+
Tensor-Flow	Python	+++	+++	++	+++	++	++	+
Torch	Lua, Python (new)	+	+++	++	++	+++	++	
Caffe	C++	+	++		+	+	+	
MXNet	R, Python, Julia, Scala	++	++	+	++	++	+++	
Neon	Python	+	++	+	+	++	+	
CNTK	C++	+	+	+++	+	++	+	

TensorFlow




```python
from __future__ import absolute_import
from __future__ import division
from __future__ import print_function

import os
import urllib
import numpy as np
import tensorflow as tf

# Data sets
Train = "Train.csv"
Test = "Test.csv"

def main():
  # Load datasets
  training_set = tf.contrib.learn.datasets.base.load_csv_with_header(
      filename=Train,
      target_dtype=np.int,
      features_dtype=np.float32)
  test_set = tf.contrib.learn.datasets.base.load_csv_with_header(
      filename=Test,
      target_dtype=np.int,
      features_dtype=np.float32)

  # Specify that all features have real-value data
  feature_columns = [tf.feature_column.numeric_column("x", shape=[28])]

  # Build 3 layer DNN with 10, 20, 10 units respectively
  classifier = tf.estimator.DNNClassifier(feature_columns=feature_columns,
                                          hidden_units=[1000,2000,1000],
                                          n_classes=2,
                                          model_dir="/tmp/RadiomicsICC_model3")
  # Define the training inputs
  train_input_fn = tf.estimator.inputs.numpy_input_fn(
      x={"x": np.array(training_set.data)},
      y=np.array(training_set.target),
      num_epochs=None,
      shuffle=True)

  # Train model
  classifier.train(input_fn=train_input_fn, steps=300000)

  # Define the test inputs
  test_input_fn = tf.estimator.inputs.numpy_input_fn(
      x={"x": np.array(test_set.data)},
      y=np.array(test_set.target),
      num_epochs=1000,
      shuffle=False)

  # Evaluate accuracy
  accuracy_score = classifier.evaluate(input_fn=test_input_fn)["accuracy"]

  print("\nTest Accuracy: {0:f}\n".format(accuracy_score))

if __name__ == "__main__":
    main()
```

Difficultés



- Data en médecine :
  - Faible nombre de patients
  - Faible granularité, peu structurée
  - Ou au contraire grand nombre de variables

  →Risque d'overfitting

- Effet « boîte noire » :
  - Création d'un modèle incompréhensible pour un esprit humain

  → Importance de l'interprétabilité

Interprétation des résultats d’un modèle

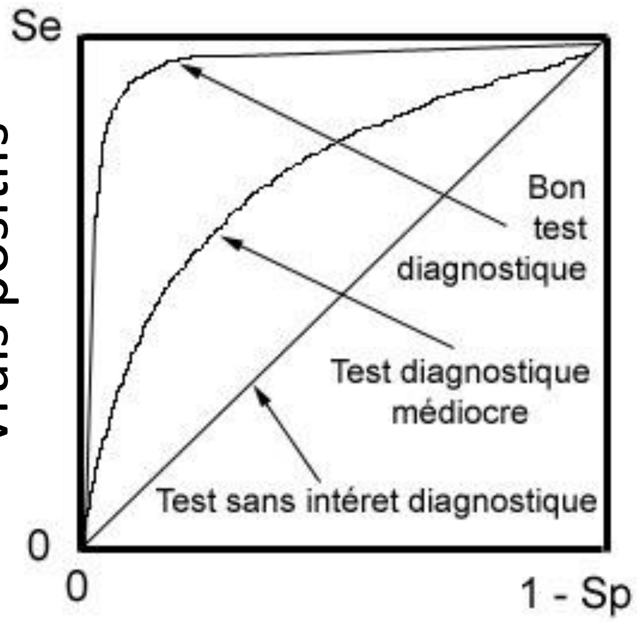
- Precision (accuracy) :

  nombre de classifications correctes / nombre total de classifications

- AUC (Area Under the Curve) :

- Fournir :
  - Intervalle de confiance
  - Matrice de confusion

Vrais positifs



Faux positifs

3 - Exemples d’application

A partir de données textuelles

Electronic Health Records

Prédiction de la survie dans le cancer de prostate

Jean-Emmanuel Bibault, MD, PhD
Lei Xing, PhD

Laboratory of Artificial Intelligence in Medicine and Biomedical Physics

Méthodes

- Essai PLCO:
  - Etude prospective, multicentrique, randomisée
  - 76 693 patients inclus dans 10 centres

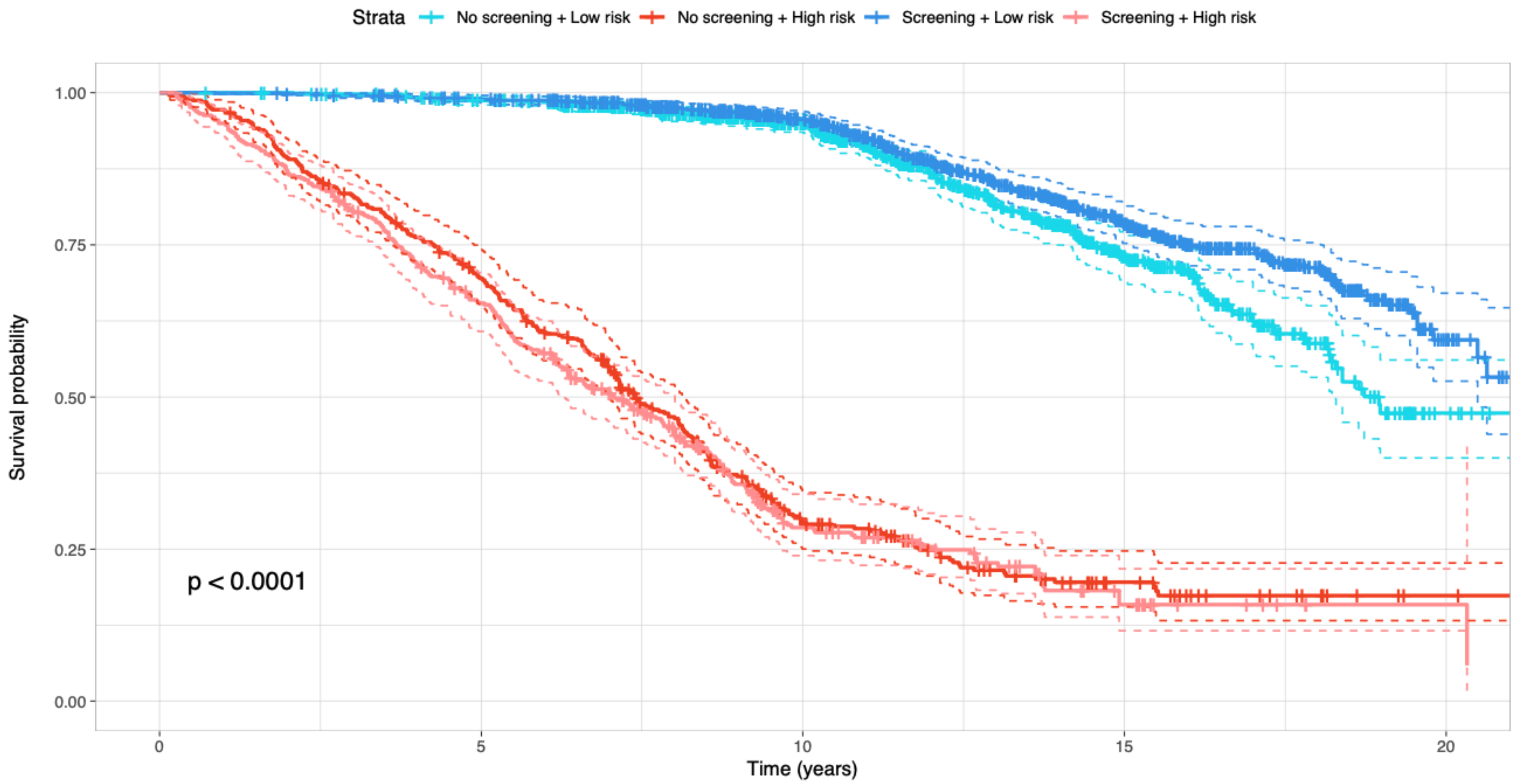
- Sélection des patients avec un cancer de prostate

- Caractéristiques:
  1. Cancer de prostate
  2. Antécédents
  3. Activité physique
  4. Statut socio-économique
  5. Statut hormonal

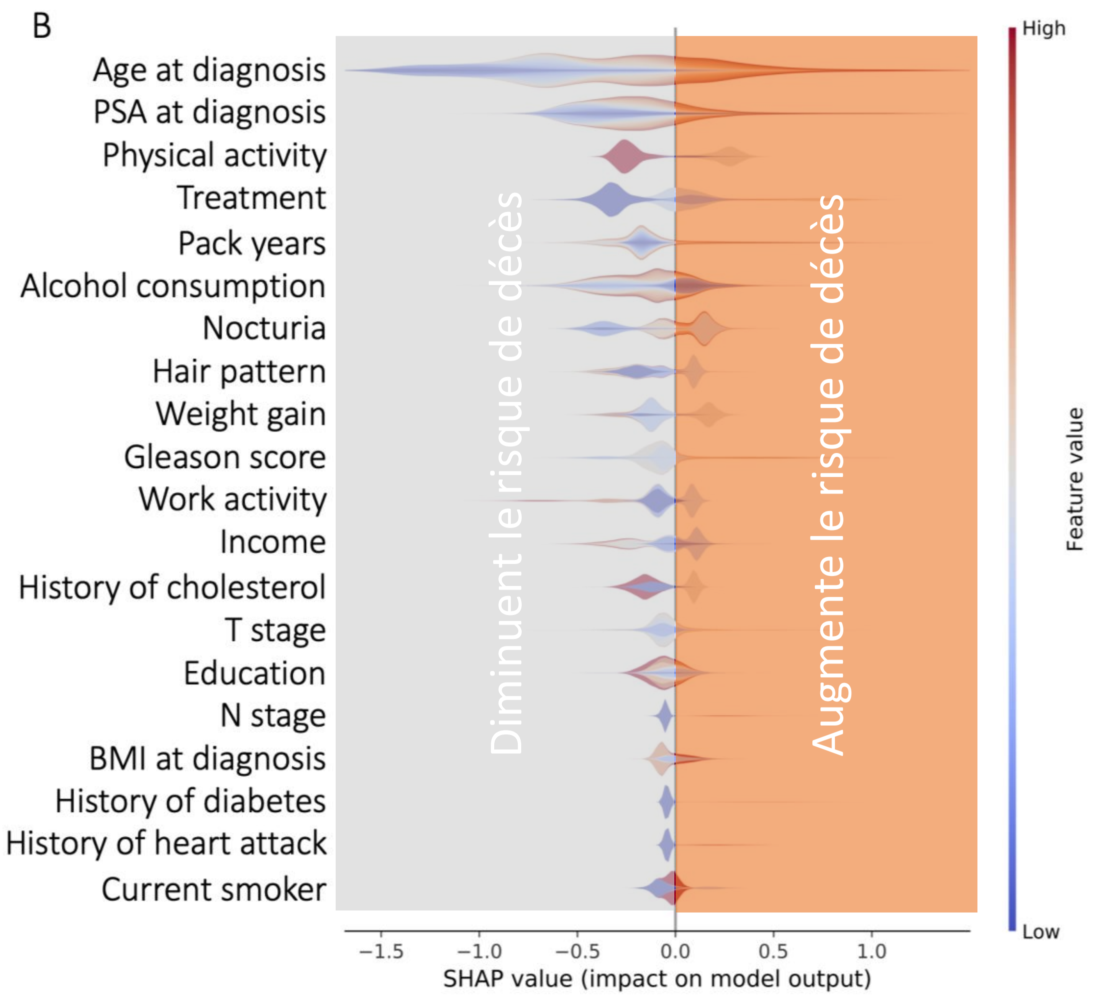
Résultats - Performances des modèles

Metric	Definition	CSS	OS
Accuracy	Number of correct predictions / total number of input samples	0.98 (±0.01)	0.86 (±0.09)
Precision	Number of correct positive predictions / number of positive predictions	0.80 (±0.1)	0.65 (±0.03)
Recall	Number of correct positive predictions / number of all positive samples	0.60 (±0.08)	0.79 (±0.04)
f1-score	Harmonic mean of the precision and the recall	0.66 ((±0.07)	0.72 (±0.03)
ROC AUC	Area under the curve of true positive rate and false positive rate at various thresholds	0.80 (±0.04)	0.84 (±0.02)
PR AUC	Area under the curve of precision and recall at various thresholds	0.54 (±0.07)	0.59 (±0.03)

Dépistage vs dépistage + AI



Caractéristiques les plus importantes



Stanford MEDICINE | Laboratory of Artificial Intelligence in Medicine and Biomedical Physics
Department of Radiation Oncology

Predict prostate cancer survival with AI

This model allows you to predict 10-year cancer-specific and overall survival in patients with prostate cancer.

What are these model for?

In the United States alone, each year, an estimated 180,890 new cases will be diagnosed and 26,120 men will die from the disease. PSA testing has resulted in a significant increase in the diagnosis and treatment of prostate cancer. But the management of prostate cancer that is detected on the basis of prostate-specific antigen (PSA) levels remains controversial.

Many men do not benefit from treatment because the disease is either indolent or disseminated at diagnosis. Because prostate cancer progresses slowly, patients often die of competing causes.

In order to assess whether a patient with prostate cancer could actually benefit from cancer treatment, and not die from another cause, we created models to predict 10-year cancer-specific and overall survival.

The PLCO Trial

The Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial was conducted to assess the role of prostate cancer screening on survival. From 1993 through 2001, 76,693 men at 10 U.S. study centers were randomized to receive either annual screening (38,343 subjects) or usual care as the control (38,350 subjects). Men in the screening group were offered annual PSA testing for 6 years and digital rectal examination for 4 years. The subjects and health care providers received the results and decided on the type of follow-up evaluation. Results of the trial were published in three articles:

Andriole GL et al., New England Journal of Medicine, 2009

Interprétabilité à l'échelle d'un patient

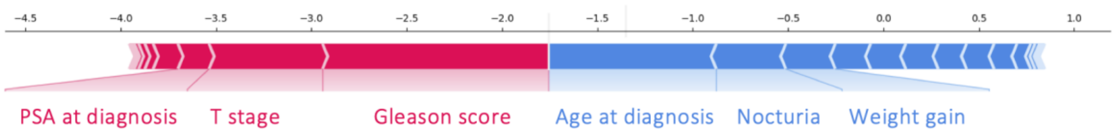
Haut risque:

- Gleason 9
- PSA = 25 ng/ml
- T3bN0M0

Sans comorbidités :

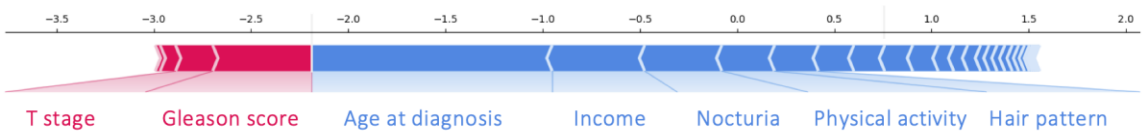
- 55 ans
- Non fumeur
- Pas d'alcool
- Activité physique régulière

A



Probabilité de décéder du cancer de prostate : 19%

B



Probabilité de décéder de toute autre cause : 20%

Interprétabilité à l'échelle d'un patient 2

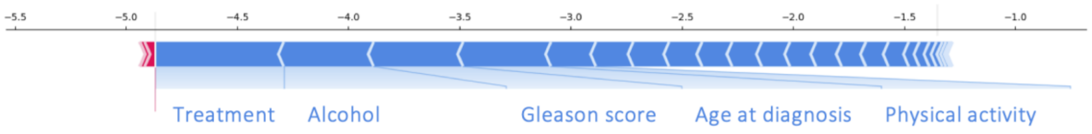
Risque intermédiaire :

- Gleason 7
- PSA = 12 ng/ml
- T2cN0M0 stage

Comorbidités:

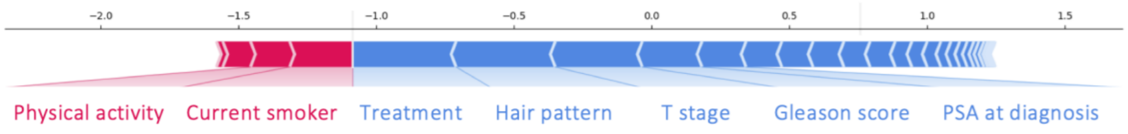
- 70 ans
- Fumeur (50 PA)
- 2 verres d'alcool/j
- Pas d'activité physique

C



Risque de décès (PCa): 1%

D



Risque de décès (toute cause) : 25%

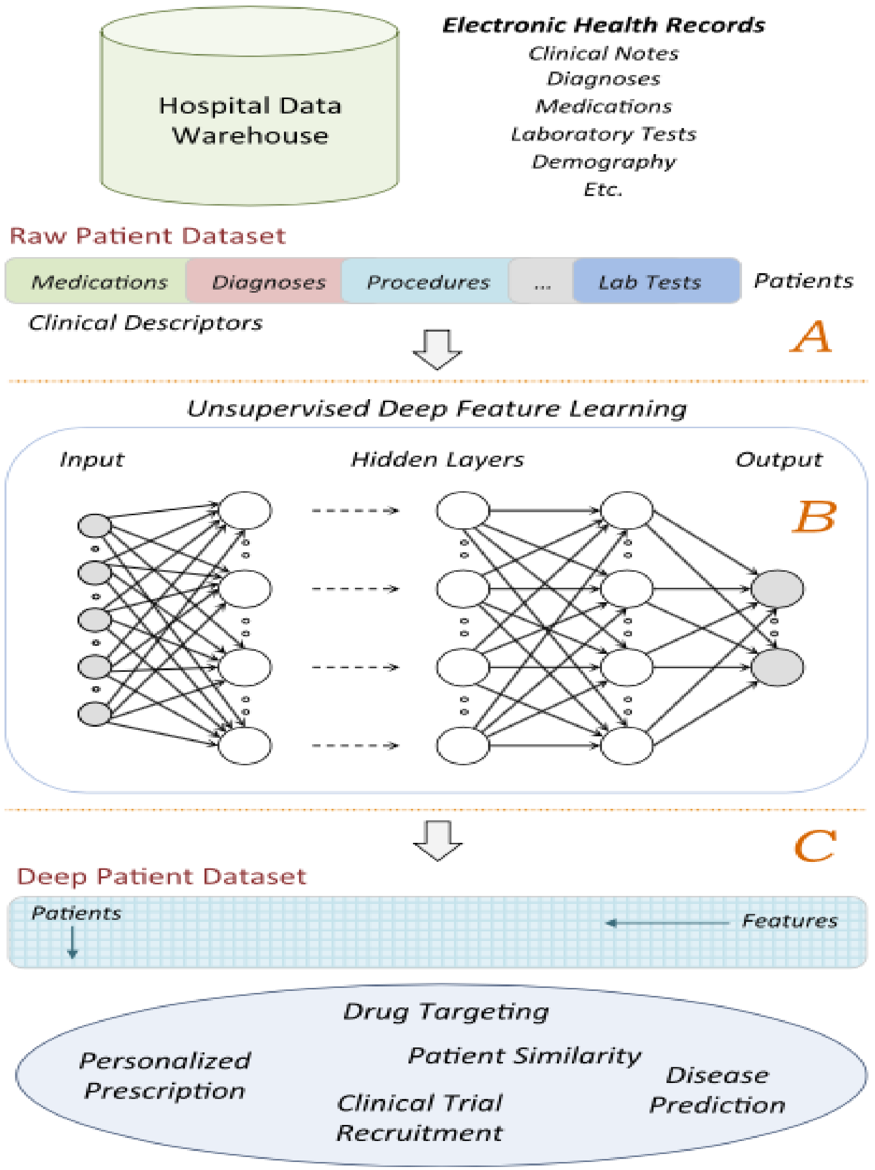
OPEN

Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records

Received: 28 January 2016
Accepted: 27 April 2016
Published: 17 May 2016

Riccardo Miotto[1,2,3], Li Li[1,2,3], Brian A. Kidd[1,2,3], Joel T. Dudley[1,2,3]

- 76214 patients
- Prédiction de 78 maladies



Miotto et al, Scientific Reports, 2016

Time Interval = 1 year (76,214 patients)			
	Area under the ROC curve		
Disease	**RawFeat**	**PCA**	**DeepPatient**
Diabetes mellitus with complications	0.794	0.861	**0.907**
Cancer of rectum and anus	0.863	0.821	**0.887**
Cancer of liver and intrahepatic bile duct	0.830	0.867	**0.886**
Regional enteritis and ulcerative colitis	0.814	0.843	**0.870**
Congestive heart failure (non-hypertensive)	0.808	0.808	**0.865**
Attention-deficit and disruptive behavior disorders	0.730	0.797	**0.863**
Cancer of prostate	0.692	0.820	**0.859**
Schizophrenia	0.791	0.788	**0.853**
Multiple myeloma	0.783	0.739	**0.849**
Acute myocardial infarction	0.771	0.775	**0.847**

Miotto et al, Scientific Reports, 2016

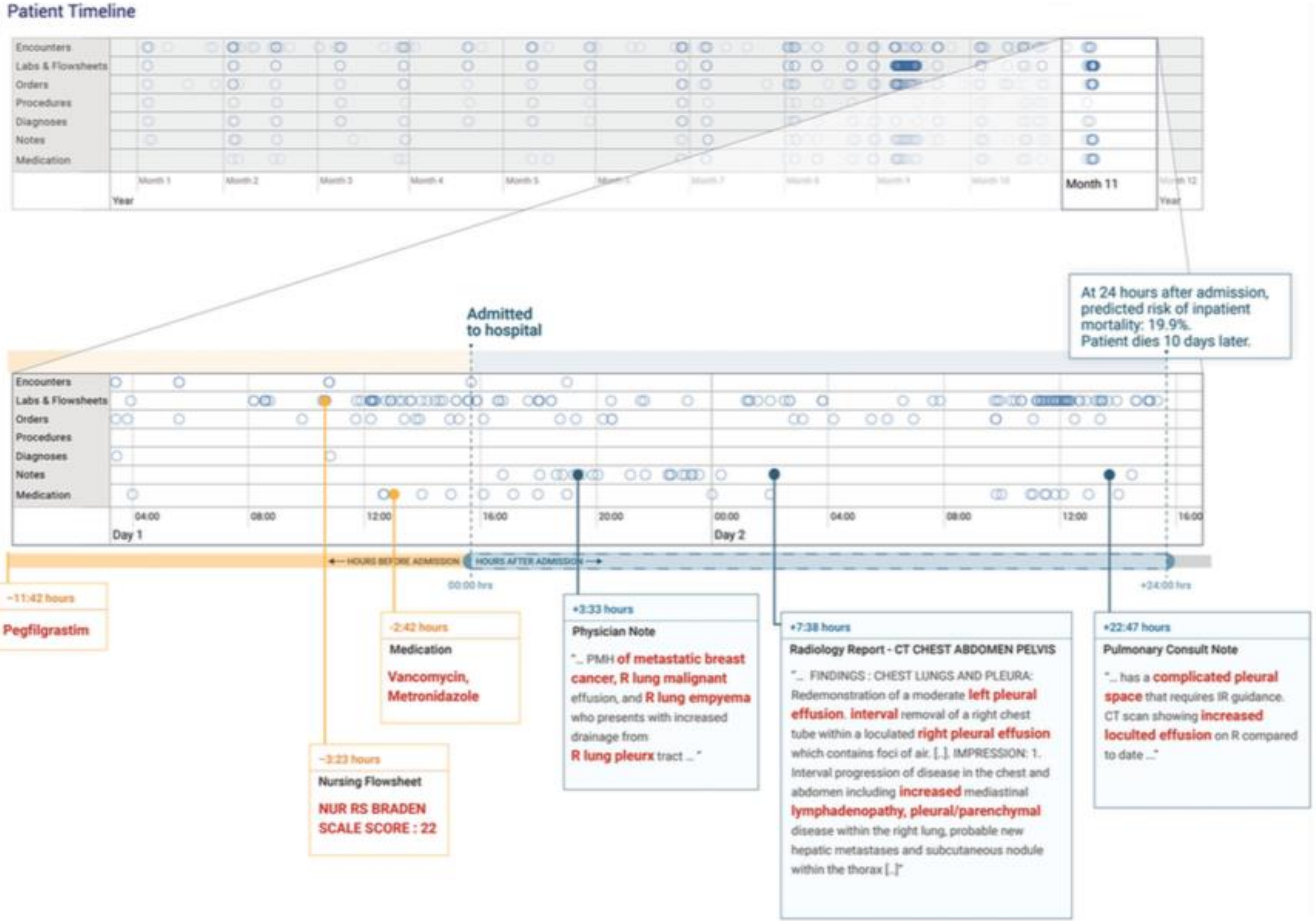
npj | Digital Medicine
www.nature.com/npjdigitalmed

ARTICLE OPEN

Scalable and accurate deep learning with electronic health records

Alvin Rajkomar[1,2], Eyal Oren[1], Kai Chen[1], Andrew M. Dai[1], Nissan Hajaj[1], Michaela Hardt[1], Peter J. Liu[1], Xiaobing Liu[1], Jake Marcus[1], Mimi Sun[1], Patrik Sundberg[1], Hector Yee[1], Kun Zhang[1], Yi Zhang[1], Gerardo Flores[1], Gavin E. Duggan[1], Jamie Irvine[1], Quoc Le[1], Kurt Litsch[1], Alexander Mossin[1], Justin Tansuwan[1], De Wang[1], James Wexler[1], Jimbo Wilson[1], Dana Ludwig[2], Samuel L. Volchenboum[3], Katherine Chou[1], Michael Pearson[1], Srinivasan Madabushi[1], Nigam H. Shah[4], Atul J. Butte[2], Michael D. Howell[1], Claire Cui[1], Greg S. Corrado[1] and Jeffrey Dean[1]

216,221 patients
46 864 534 945 datapoints



Patient Timeline
Encounters
Labs & Flowsheets
Orders
Procedures
Diagnoses
Notes
Medication
Month 1
Month 2
Month 3
Month 4
Month 5
Month 11
Year
Admitted to hospital
At 24 hours after admission, predicted risk of inpatient mortality: 19.9%. Patient dies 10 days later.
04:00
08:00
12:00
16:00
20:00
00:00
04:00
08:00
12:00
16:00
Day 1
Day 2
HOURS BEFORE ADMISSION
HOURS AFTER ADMISSION
00:00 hrs
+24:00 hrs
-11:42 hours
Pegfilgrastim
-2:42 hours
Medication
Vancomycin, Metronidazole
-3:23 hours
Nursing Flowsheet
NUR RS BRADEN SCALE SCORE : 22
+3:33 hours
Physician Note
"... PMH of metastatic breast cancer, R lung malignant effusion, and R lung empyema who presents with increased drainage from R lung pleurx tract ..."
+7:38 hours
Radiology Report - CT CHEST ABDOMEN PELVIS
"... FINDINGS : CHEST LUNGS AND PLEURA: Redemonstration of a moderate left pleural effusion. interval removal of a right chest tube within a loculated right pleural effusion which contains foci of air. [..]. IMPRESSION: 1. Interval progression of disease in the chest and abdomen including increased mediastinal lymphadenopathy, pleural/parenchymal disease within the right lung, probable new hepatic metastases and subcutaneous nodule within the thorax [..]"
+22:47 hours
Pulmonary Consult Note
"... has a complicated pleural space that requires IR guidance. CT scan showing increased loculted effusion on R compared to date ..."

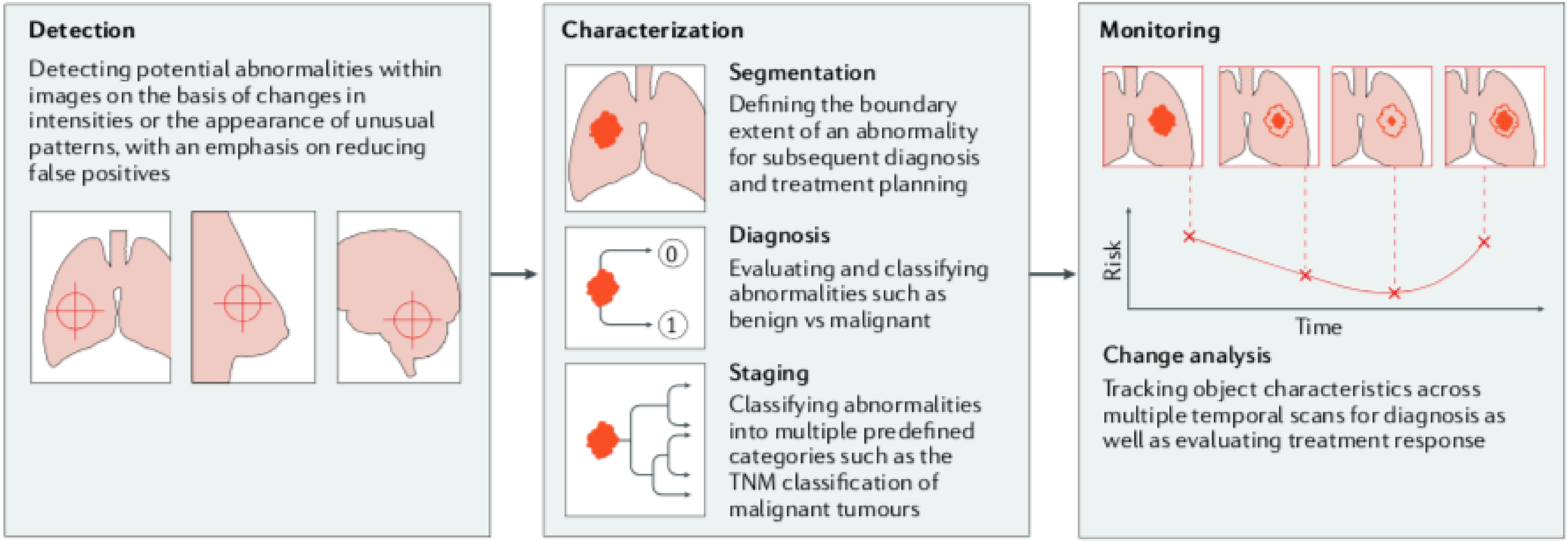
Table 2. Prediction accuracy of each task made at different time points

	Hospital A	Hospital B
Inpatient mortality, AUROC[a] (95% CI)		
24 h before admission	0.87 (0.85–0.89)	0.81 (0.79–0.83)
At admission	0.90 (0.88–0.92)	0.90 (0.86–0.91)
24 h after admission	**0.95 (0.94–0.96)**	**0.93 (0.92–0.94)**
Baseline (aEWS[b]) at 24 h after admission	0.85 (0.81–0.89)	0.86 (0.83–0.88)
30-day readmission, AUROC (95% CI)		
At admission	0.73 (0.71–0.74)	0.72 (0.71–0.73)
At 24 h after admission	0.74 (0.72–0.75)	0.73 (0.72–0.74)
At discharge	**0.77 (0.75–0.78)**	**0.76 (0.75–0.77)**
Baseline (mHOSPITAL[c]) at discharge	0.70 (0.68–0.72)	0.68 (0.67–0.69)
Length of stay at least 7 days, AUROC (95% CI)		
At admission	0.81 (0.80–0.82)	0.80 (0.80–0.81)
At 24 h after admission	**0.86 (0.86–0.87)**	**0.85 (0.85–0.86)**
Baseline (Liu[d]) at 24 h after admission	0.76 (0.75–0.77)	0.74 (0.73–0.75)
Discharge diagnoses (weighted AUROC)		
At admission	0.87	0.86
At 24 h after admission	0.89	0.88
At discharge	**0.90**	**0.90**

[a]Area under the receiver operator curve
[b]Augmented Early Warning System score
[c]Modified HOSPITAL score for readmission
[d]Modified Liu score for long length of stay
The bold values indicate the highest area-under-the-receiver-operator-curve for each prediction task

A partir d’images

CT-Scan, IRM, TEP

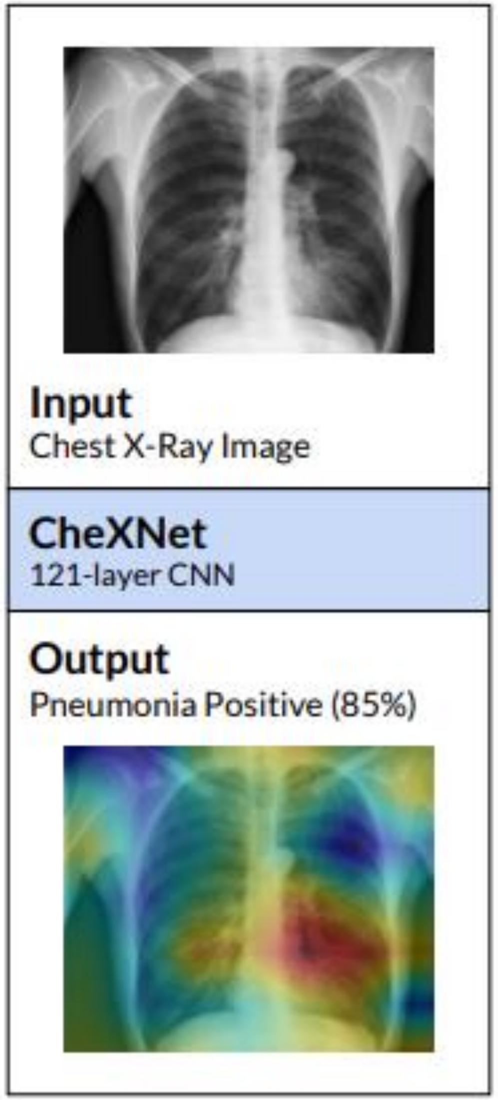


Hosny A et al, Artificial intelligence in radiology, Nature Reviews Cancer 2018

CheXnet

- Détection de pneumopathie sur radiographie du thorax de face
- Réseau neuronal de 121 couches entraîné sur 100 000 radiographies pour détecter 14 anomalies

Rajpurkar P et al, arXiv, Dec 2017



Résultats

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

Pathology	Wang et al. (2017)	Yao et al. (2017)	CheXNet (ours)
Atelectasis	0.716	0.772	**0.8094**
Cardiomegaly	0.807	0.904	**0.9248**
Effusion	0.784	0.859	**0.8638**
Infiltration	0.609	0.695	**0.7345**
Mass	0.706	0.792	**0.8676**
Nodule	0.671	0.717	**0.7802**
Pneumonia	0.633	0.713	**0.7680**
Pneumothorax	0.806	0.841	**0.8887**
Consolidation	0.708	0.788	**0.7901**
Edema	0.835	0.882	**0.8878**
Emphysema	0.815	0.829	**0.9371**
Fibrosis	0.769	0.767	**0.8047**
Pleural Thickening	0.708	0.765	**0.8062**
Hernia	0.767	0.914	**0.9164**

Table 2. CheXNet outperforms the best published results on all 14 pathologies in the ChestX-ray14 dataset. In detecting Mass, Nodule, Pneumonia, and Emphysema, CheXNet has a margin of >0.05 AUROC over previous state of the art results.

Rajpurkar P et al, arXiv, Dec 2017

npj | Digital Medicine

www.nature.com/npjdigitalmed

ARTICLE OPEN

Advanced machine learning in action: identification of intracranial hemorrhage on computed tomography scans of the head with clinical workflow integration

Mohammad R. Arbabshirani[1], Brandon K. Fornwalt[1,2], Gino J. Mongelluzzo[1], Jonathan D. Suever[1,2], Brandon D. Geise[1], Aalpen A. Patel[1,2] and Gregory J. Moore[1]

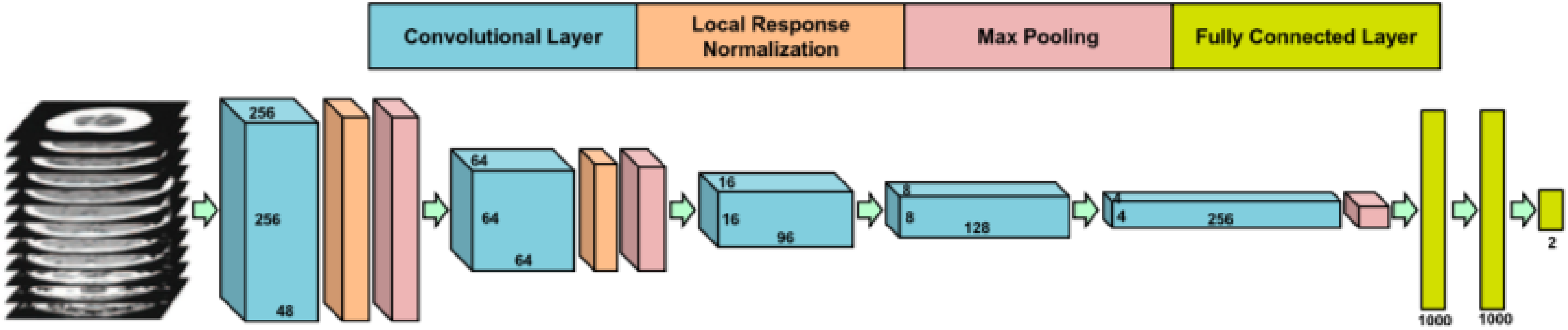


Fig. 1 Proposed CNN architecture for ICH detection. This architecture has five convolutional and two fully connected layers

37,084 CT scan

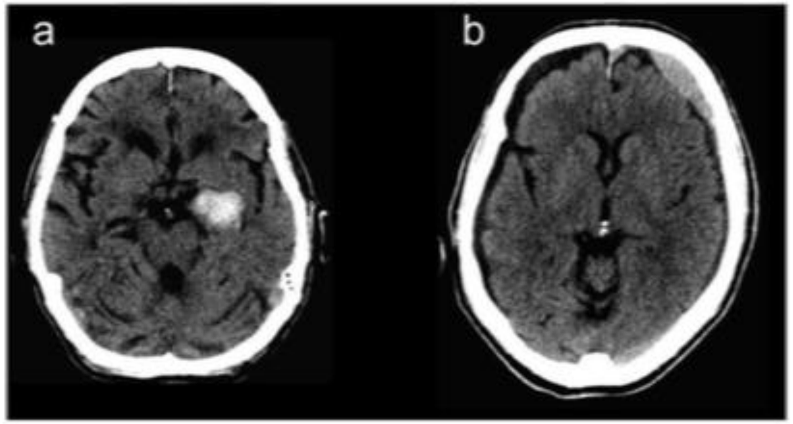


Fig. 5 Representative head CT image slices associated with case #1 (**a**) and case #2 (**b**)

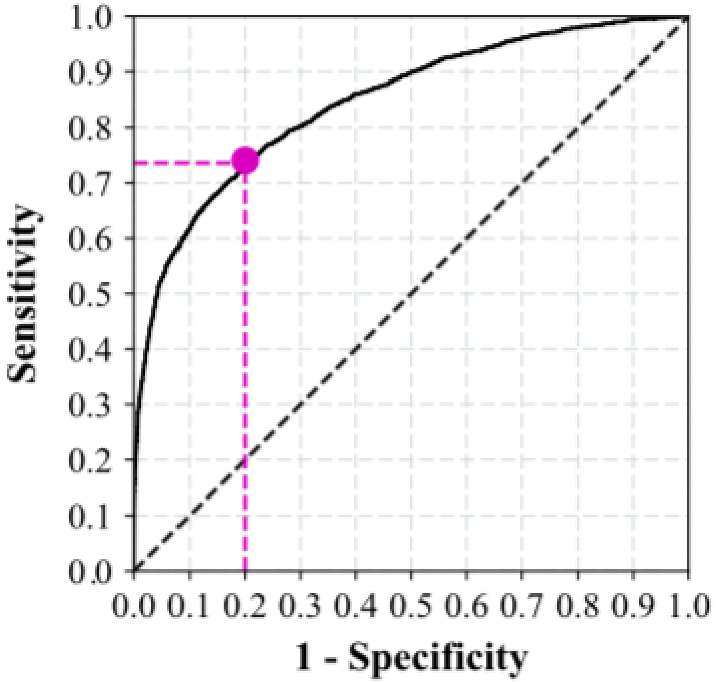


Fig. 2 ROC curve of the ICH detection algorithm on the testing dataset. The magenta circle illustrates the operating point. AUC of the ROC is 0.846

Prédiction de la réponse thérapeutique

www.nature.com/scientificreports

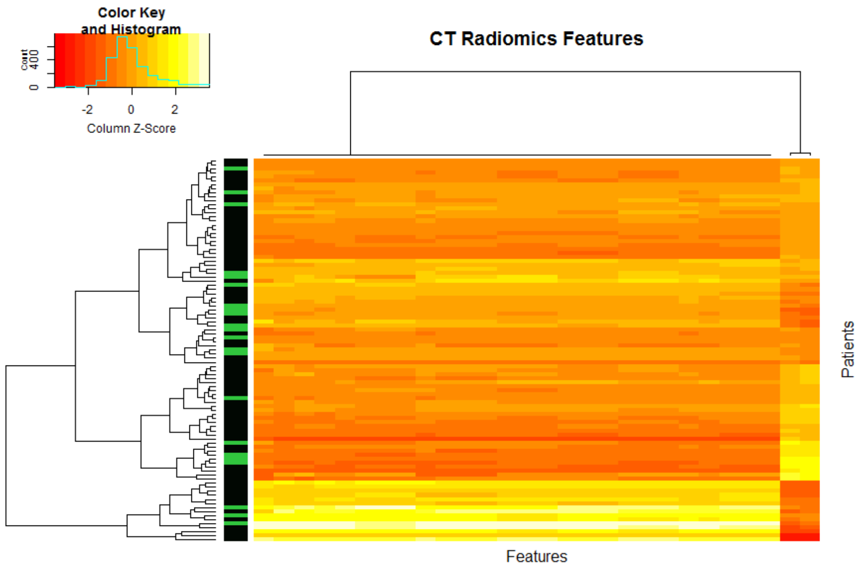
SCIENTIFIC REPORTS

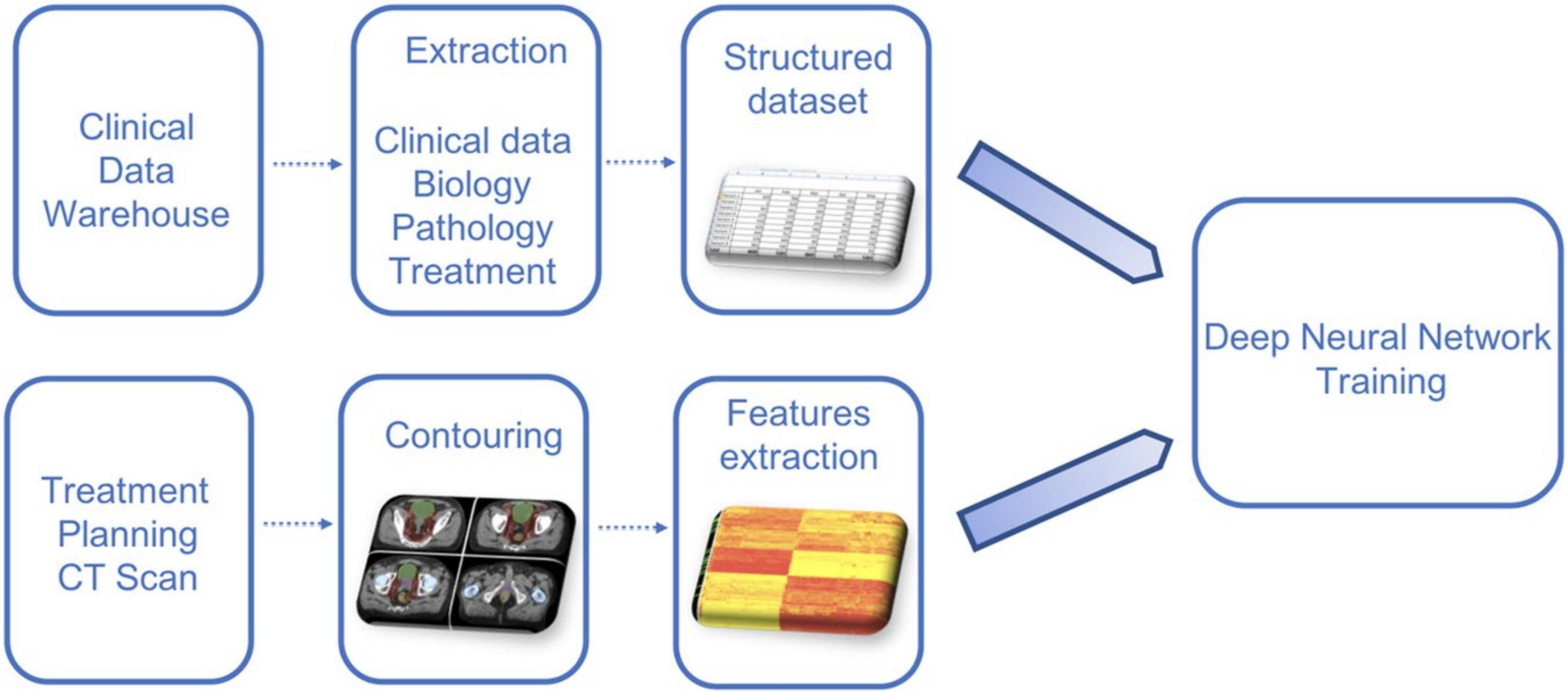
OPEN

Received: 1 March 2018
Accepted: 3 August 2018
Published online: 22 August 2018

Deep Learning and Radiomics predict complete response after neo-adjuvant chemoradiation for locally advanced rectal cancer

Jean-Emmanuel Bibault[1,2], Philippe Giraud[1], Catherine Durdux[1], Julien Taieb[3], Anne Berger[4], Romain Coriat[5,6], Stanislas Chaussade[5,6], Bertrand Dousset[7], Bernard Nordlinger[8] & Anita Burgun[2,9]





Clinical Data Warehouse
Extraction
Clinical data
Biology
Pathology
Treatment
Structured dataset
Deep Neural Network Training
Treatment Planning CT Scan
Contouring
Features extraction

Résultats

	Accuracy (IC 95%)	Sensibilité (IC 95%)	Spécificité (IC 95%)	AUC (IC 95%)
Logistic Regression	69,5% (59,2% - 78,51%)	34,78% (16,38% - 57.27%)	80,56% (69.53% - 88.94%)	0,59 (0,458 – 0,686)
Support Vector Machine	71,58% (61,4%-80,36%)	45,45% (24,39%-67,79%)	79,45% (68,38% - 88,02%)	0,62 (0.51 – 0.74)
Deep Learning	80% (70,54% - 87,51%)	68,2% (45,13% - 86,14%)	83,56% (73,05% - 91,21%)	0,72 (0.65 - 0.87)

MRI		0-40% (0-53%)	92-98% (88-99%)	0,58-0,76 (0,47-0,86)
MRI + DWI		52-64% (85-99%)	89-97% (85-99%)	0,78-0,8 (0,69-0,91)

COMPARAISON PERFORMANCE HUMAINE ?

Lambregts DMJ et al. 2011. « Diffusion-Weighted MRI for Selection of Complete Responders After Chemoradiation for Locally Advanced Rectal Cancer: A Multicenter Study » Annals of Surgical Oncology

OPEN

Precision Radiology: Predicting longevity using feature engineering and deep learning methods in a radiomics framework

Received: 8 December 2016
Accepted: 6 April 2017
Published online: 10 May 2017

Luke Oakden-Rayner[1,2], Gustavo Carneiro[3], Taryn Bessen[1], Jacinto C. Nascimento[4], Andrew P. Bradley[5] & Lyle J. Palmer[2]

- Définir l'état de santé d'un patient à partir de son CT

- Prédire sa longévité

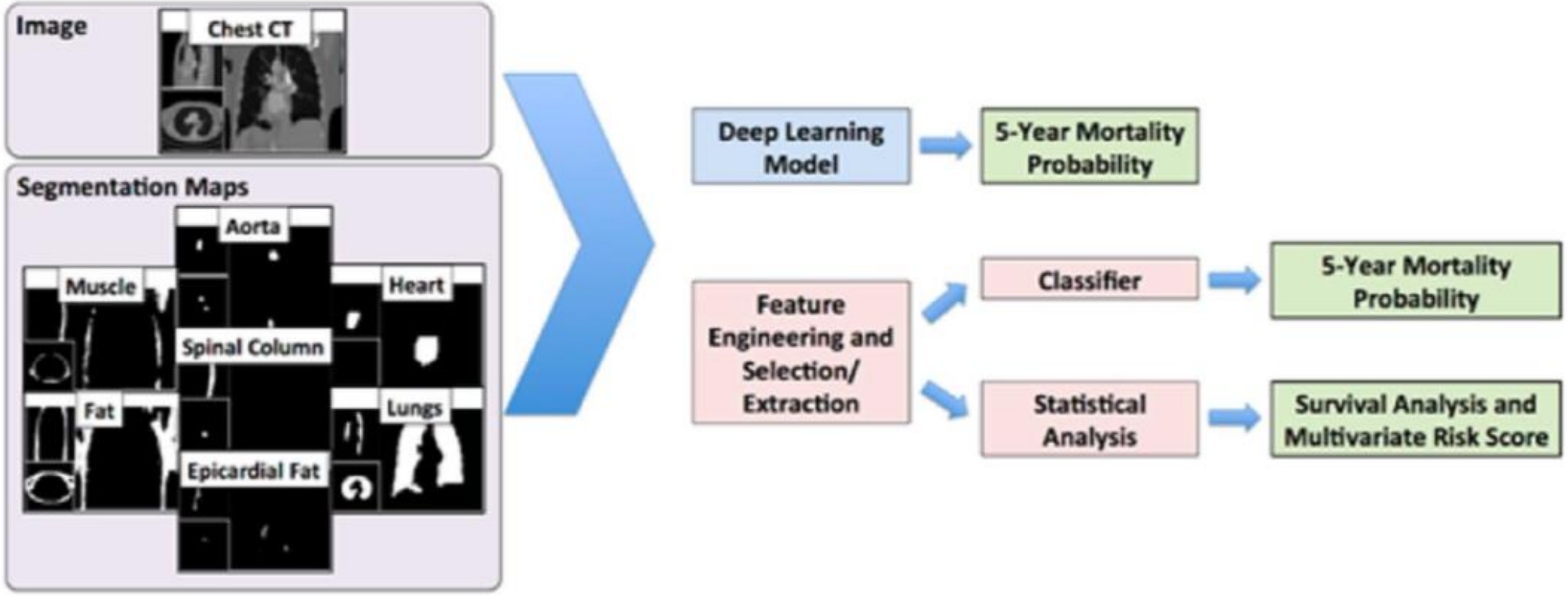
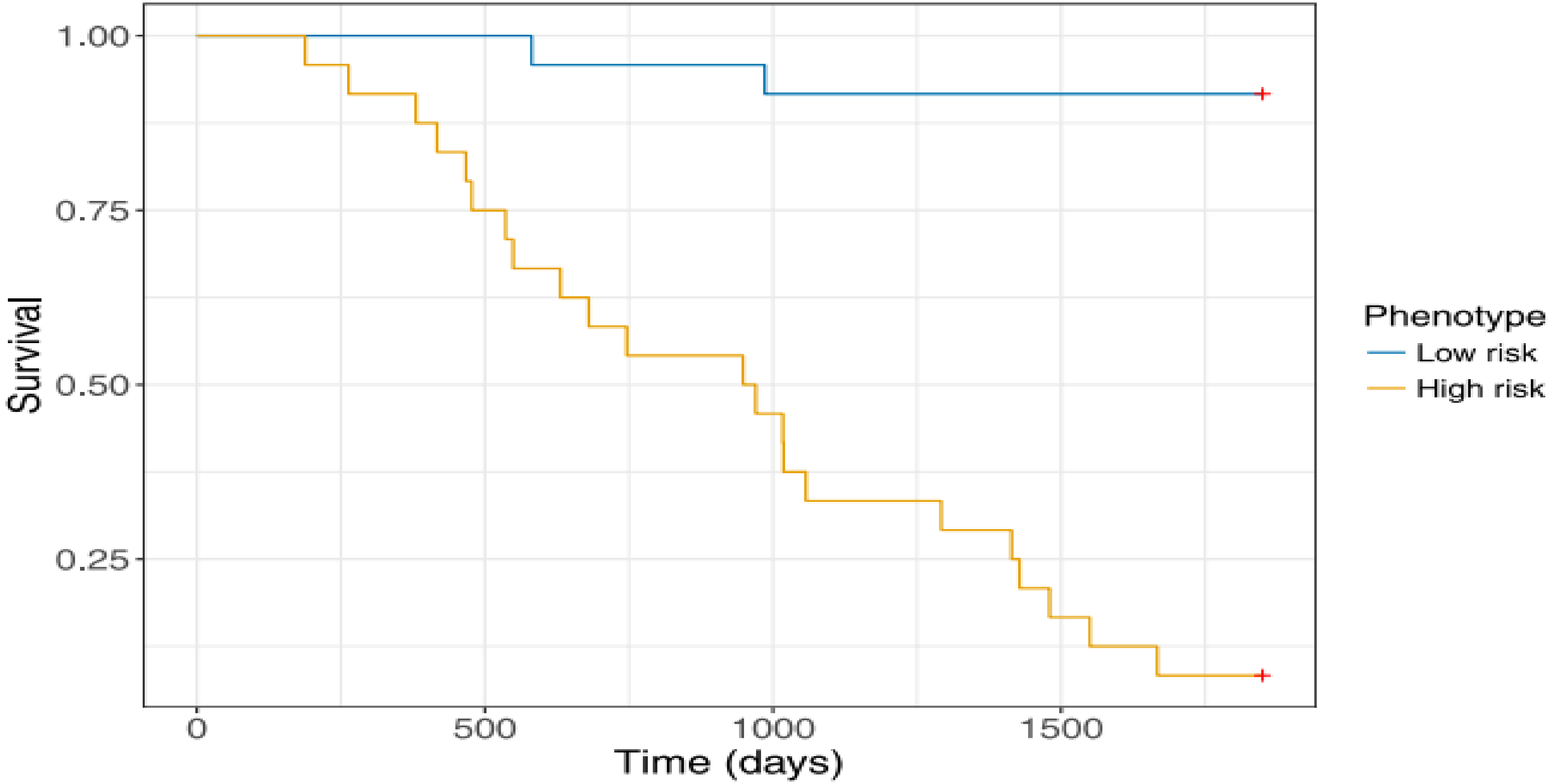


Figure 6. Diagrammatic representation of the analysis pipelines with engineered features and deep learning methods.

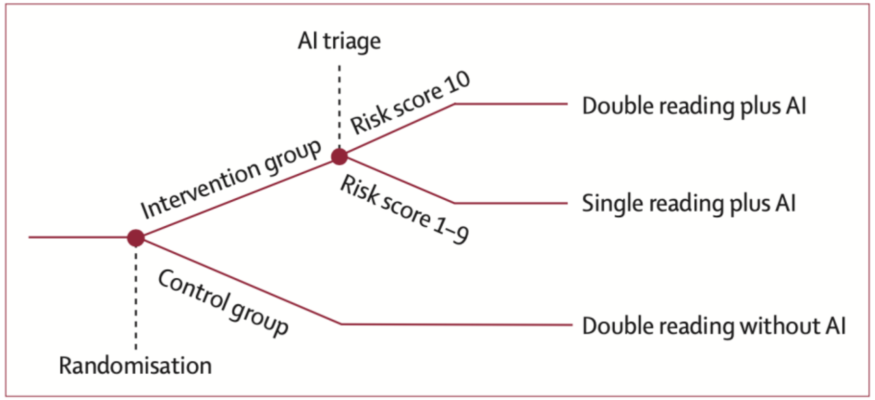


Survival
1.00
0.75
0.50
0.25
0
500
1000
1500
Time (days)
Phenotype
Low risk
High risk

Artificial intelligence-supported screen reading versus standard double reading in the Mammography Screening with Artificial Intelligence trial (MASAI): a clinical safety analysis of a randomised, controlled, non-inferiority, single-blinded, screening accuracy study

Kristina Lång, Viktoria Josefsson, Anna-Maria Larsson, Stefan Larsson, Charlotte Högberg, Hanna Sartor, Solveig Hofvind, Ingvar Andersson, Aldana Rosso

THE LANCET Oncology



Etude prospective randomisée
80 033 femmes
2 radiologues vs 1 radiologue + IA

Taux de détection similaire
Réduction du temps de 44%

	Intervention group (n=39 996)	Control group (n=40 024)
Early screening performance		
Number of recalls	861	817
Recall rate, %	2·2% (2·0–2·3)	2·0% (1·9–2·2)
Number of screen-detected cancers	244	203
Cancer-detection rate, per 1000 participants screened	6·1 (5·4–6·9)	5·1 (4·4–5·8)
False positive rate, %	1·5% (1·4–1·7)	1·5% (1·4–1·7)
Positive predictive value of recall, %	28·3% (25·3–31·5)	24·8% (21·9–28·0)
Workload		
Number of screen readings	46 345	83 231
Number of consensus meetings	1584	1576
Consensus meeting rate	4·0% (3·8–4·2)	3·9% (3·8–4·1)

Data are n or point estimate (95% CI).

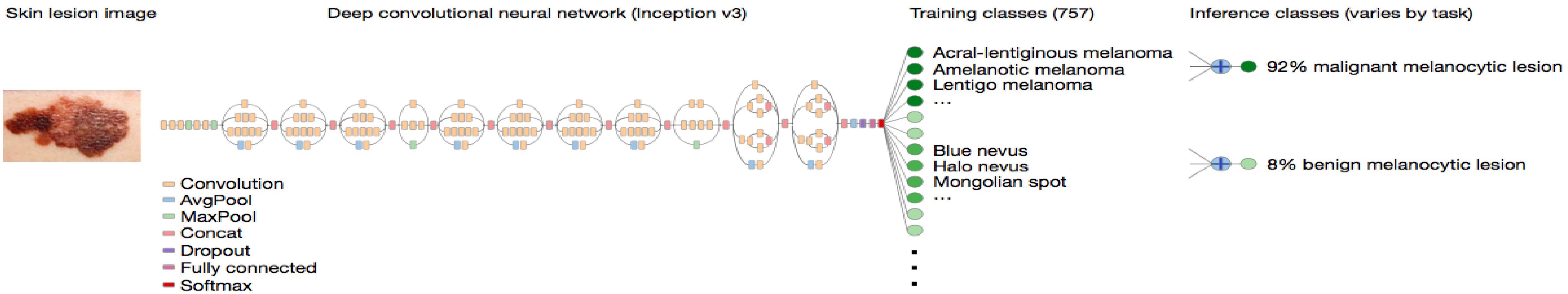
Table 2: **Early screening performance and workload measures, modified intention-to-treat population**

A partir d’autres types d’images

Photos, Satellite, Réseaux sociaux

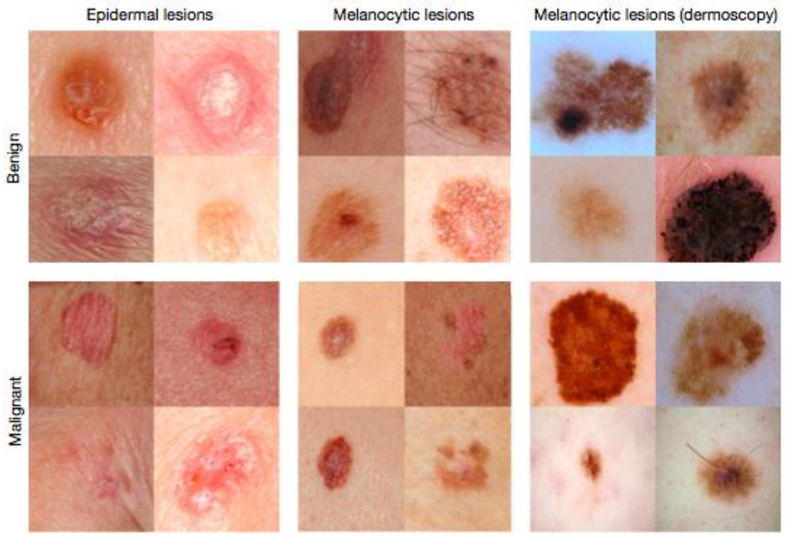
Dermatologist-level classification of skin cancer with deep neural networks

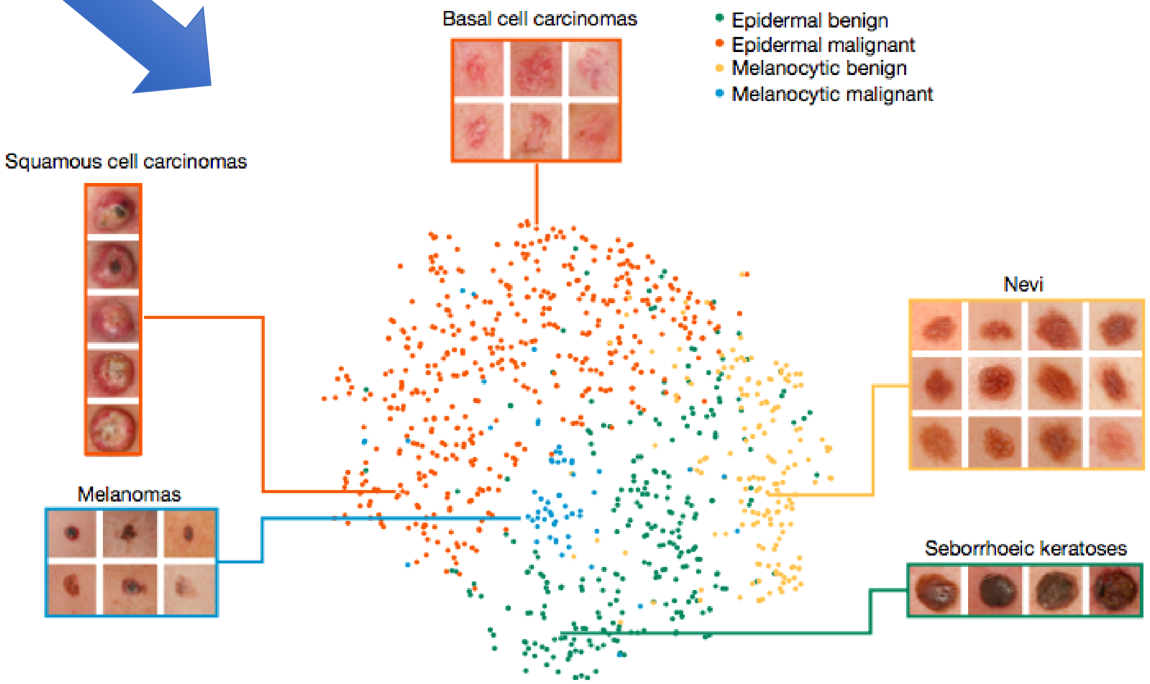
Andre Esteva[1]*, Brett Kuprel[1]*, Roberto A. Novoa[2,3], Justin Ko[2], Susan M. Swetter[2,4], Helen M. Blau[5] & Sebastian Thrun[6]



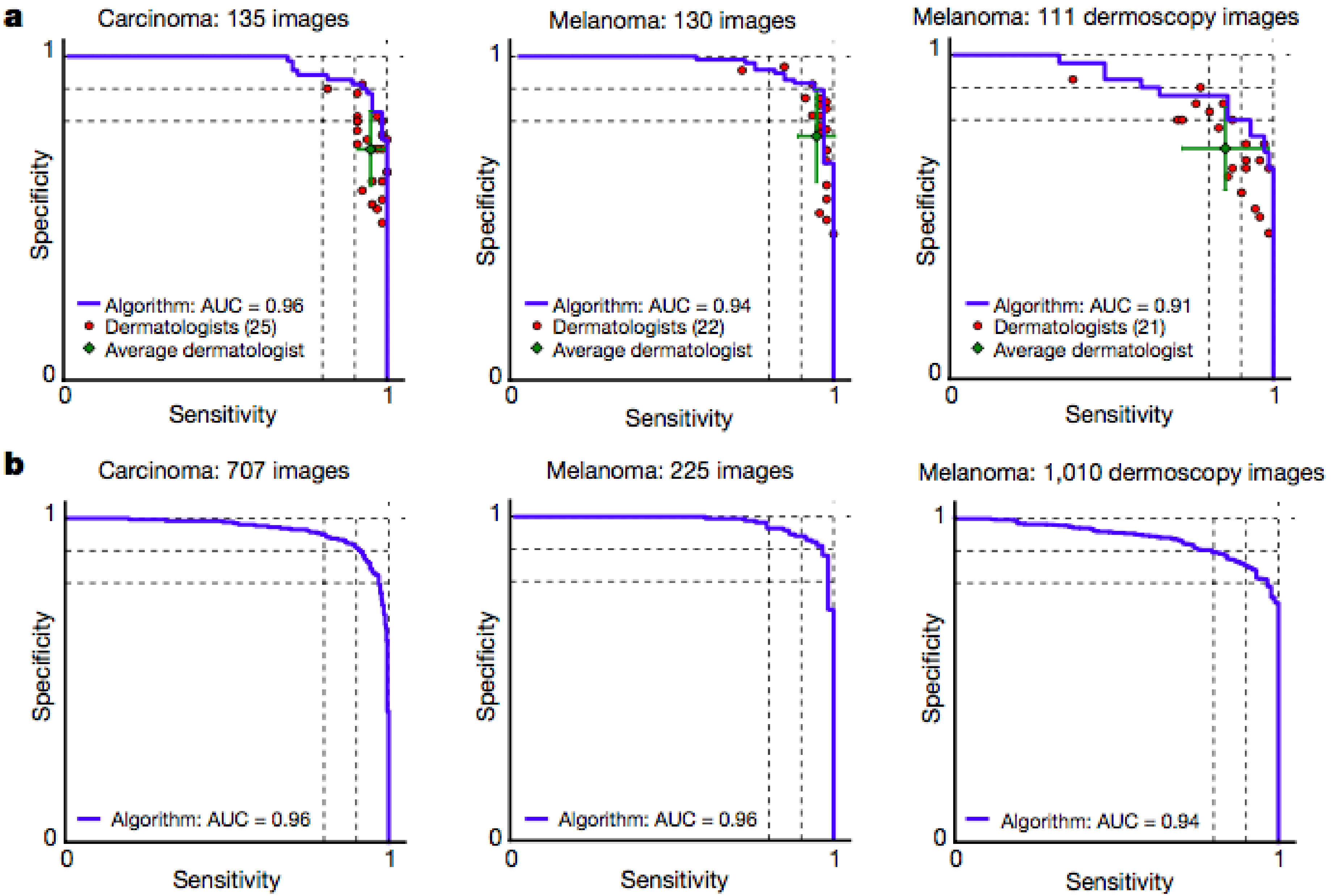
129450 lésions – Inception v3 versus 21 dermatologues

Esteva et al, Nature, 2016





Esteva et al, Nature, 2016



Esteva et al, Nature, 2016

JAMA | Original Investigation | INNOVATIONS IN HEALTH CARE DELIVERY

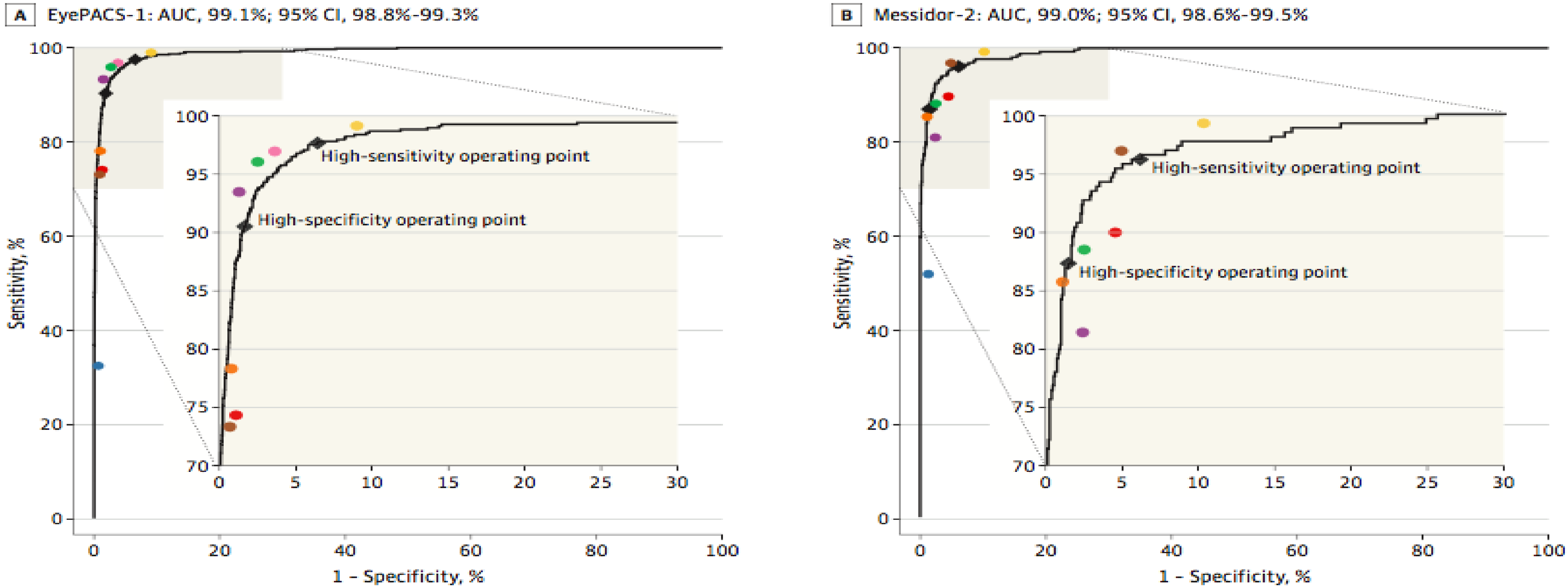
Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs

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

- 128 174 images de rétines
- 54 ophtalmologues
- 1 CNN

- Validé sur deux datasets :
  - EyePACS-1 (9963 images)
  - Messidor-2 (1748 images)

Gulshan et al, JAMA, 2016



Gulshan et al, JAMA, 2016

Tool Detection and Operative Skill Assessment in Surgical Videos Using Region-Based Convolutional Neural Networks

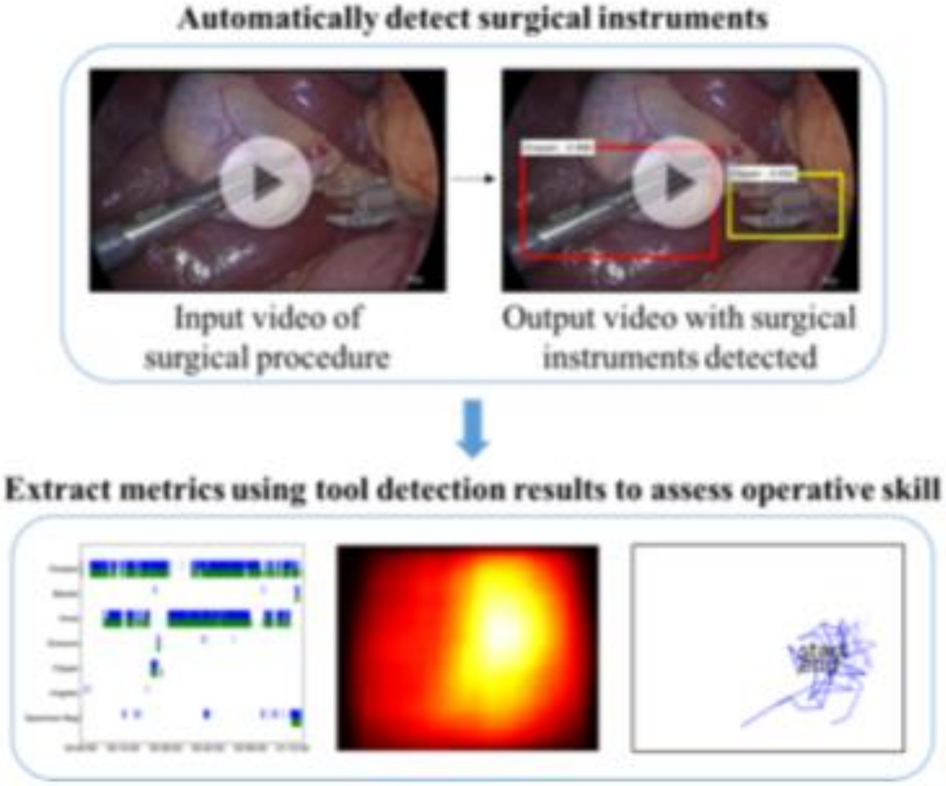
Amy Jin, Serena Yeung, Jeffrey Jopling, Jonathan Krause, Dan Azagury, Arnold Milstein, and Li Fei-Fei
Stanford University
Stanford, CA 94305

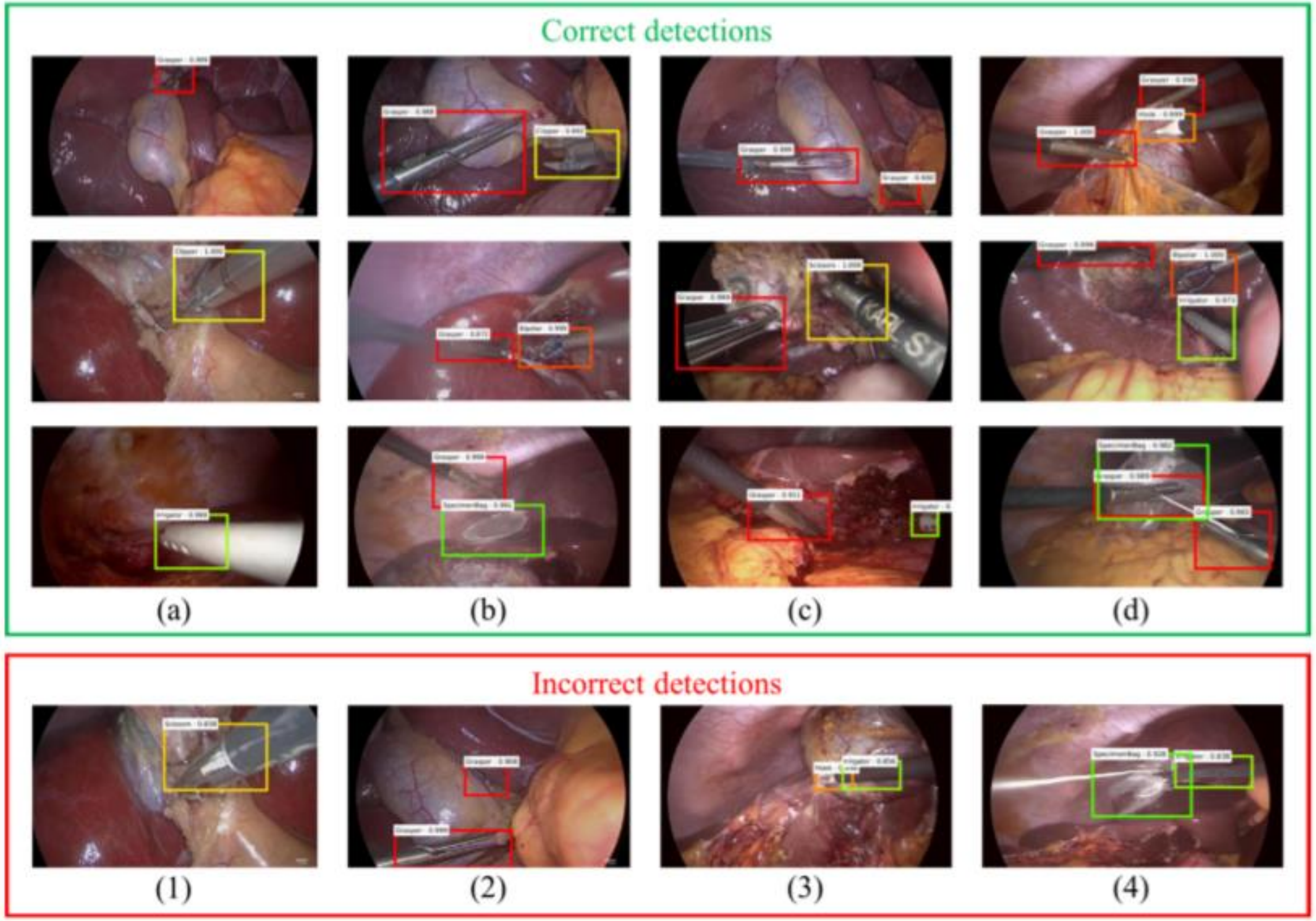
15 vidéos enregistrées à Strasbourg
Annotées

Utilisation d’un CNN

Tool	Number of annotated instances
Grasper	923
Bipolar	350
Hook	308
Scissors	400
Clipper	400
Irrigator	485
Specimen Bag	275
Total	3141
Number of Frames	2532

Table 1: Number of annotated frames for each tool.





Tool	Twinanda *et al.* [9]	Sahu *et al.* [6]	Raju *et al.* [8]	Ours
Grasper	82.2	73.9	NA	**87.2**
Bipolar	50.3	40.8	NA	**75.1**
Hook	89.4	95.1	NA	**95.3**
Scissors	17.0	26.2	NA	**70.8**
Clipper	43.6	35.3	NA	**88.4**
Irrigator	12.5	33.2	NA	**73.5**
Specimen Bag	72.2	76.6	NA	**82.1**
mAP	52.5	54.5	63.7	**81.8**

Figure 4: Example frames of spatial detection results. Bounding box color corresponds to predicted tool identity. Correct predictions are boxed in green (top), and mistakes are boxed in red (bottom). The model is able to successfully detect, classify, and localize surgical instruments despite varying tool positions and angles, and despite some parts of the tools being occluded, as shown in column (c).

Reece and Danforth *EPJ Data Science* (2017) 6:15
DOI 10.1140/epjds/s13688-017-0110-z

EPJ.org

REGULAR ARTICLE Open Access

Instagram photos reveal predictive markers of depression

Andrew G Reece[1*] and Christopher M Danforth[2,3,4*]

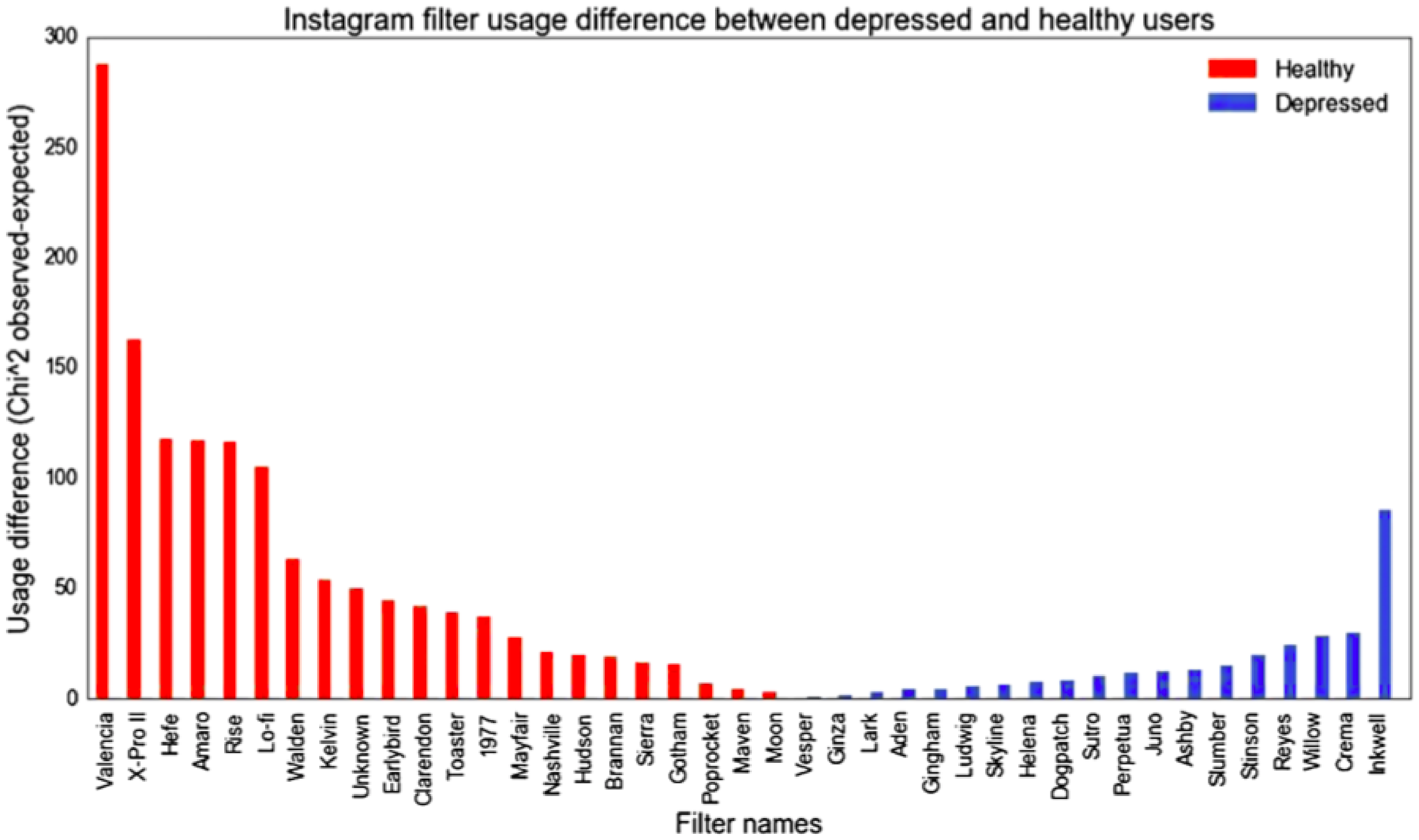


Figure 3 Instagram filter usage among depressed and healthy participants. Bars indicate difference between observed and expected usage frequencies, based on a Chi-squared analysis of independence. Blue bars indicate disproportionate use of a filter by depressed compared to healthy participants, orange bars indicate the reverse. All-data results are displayed, see Additional file 1 for Pre-diagnosis plot.

Stanford MEDICINE

Prédire la prevalence du cancer à partir d'images satellitaires

Jean-Emmanuel Bibault, MD, PhD
Maxime Bassenne, PhD
Hongyi Ren, MSc
Lei Xing, PhD

Laboratory of Artificial Intelligence in Medicine and Biomedical Physics

Br. J. Cancer (1973) **27**, 167.

ENVIRONMENTAL FACTORS AND CANCER OF THE COLON AND BREAST

B. S. DRASAR AND DOREEN IRVING*

From the Department of Bacteriology, Wright-Fleming Institute of Microbiology, St Mary's Hospital Medical School, London W2 1PG

*and the *Department of Medical Statistics and Epidemiology, London School of Hygiene and Tropical Medicine, Keppel Street, Gower Street, London WC1E 7HT*

Int. J. Cancer: **15**, 617-631 (1975)

ENVIRONMENTAL FACTORS AND CANCER INCIDENCE AND MORTALITY IN DIFFERENT COUNTRIES, WITH SPECIAL REFERENCE TO DIETARY PRACTICES

by

Bruce ARMSTRONG and Richard DOLL

Department of the Regius Professor of Medicine, Radcliffe Infirmary, Oxford OX2 6HE, England

The New England
Journal of Medicine

VOLUME 343 JULY 13, 2000 NUMBER 2

ENVIRONMENTAL AND HERITABLE FACTORS IN THE CAUSATION OF CANCER

Analyses of Cohorts of Twins from Sweden, Denmark, and Finland

PAUL LICHTENSTEIN, PH.D., NIELS V. HOLM, M.D., PH.D., PIA K. VERKASALO, M.D., PH.D., ANASTASIA ILIADOU, M.SC., JAAKKO KAPRIO, M.D., PH.D., MARKKU KOSKENVUO, M.D., PH.D., EERO PUKKALA, PH.D., AXEL SKYTTHE, M.SC., AND KARI HEMMINKI, M.D., PH.D.

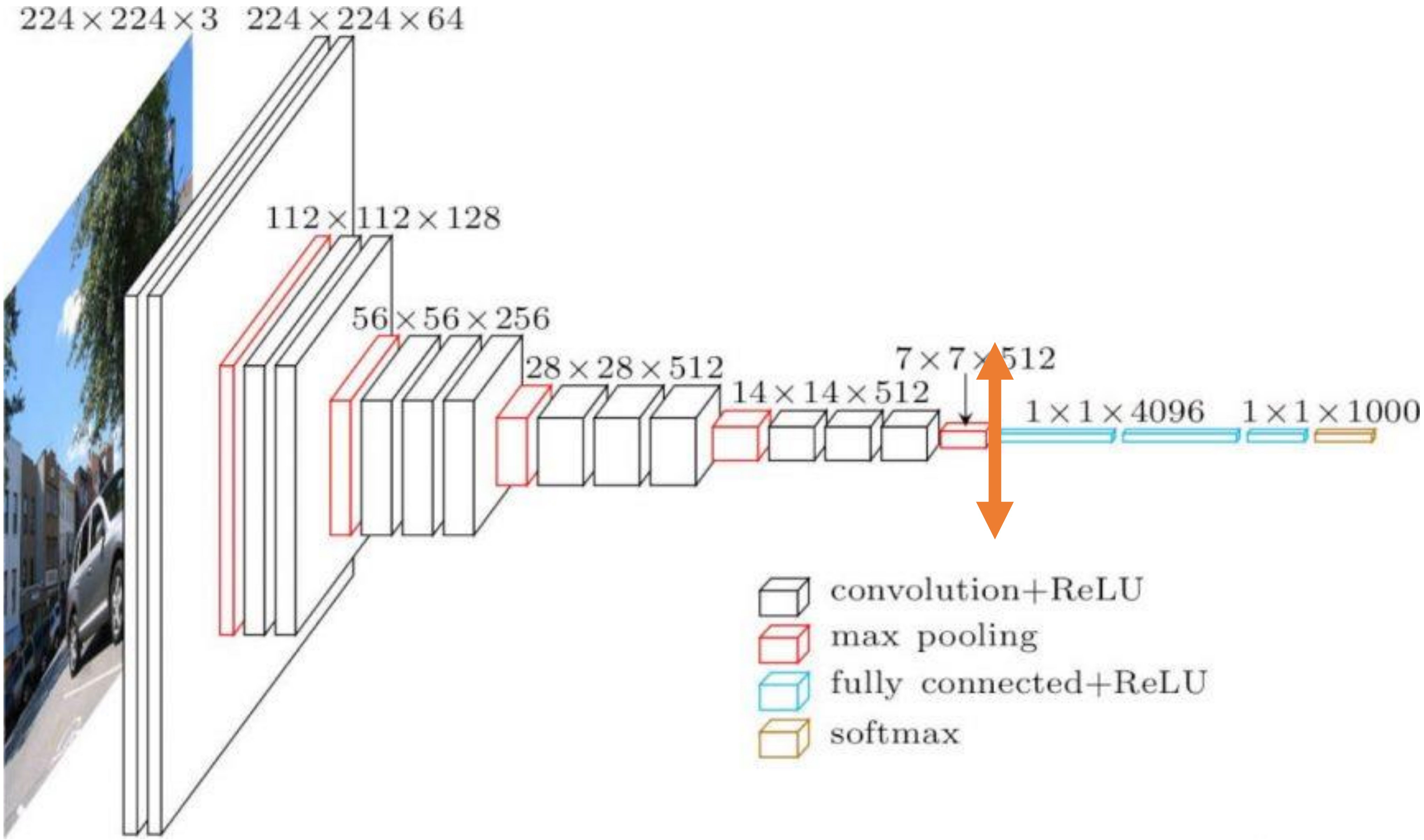
Conclusions Inherited genetic factors make a minor contribution to susceptibility to most types of neoplasms. This finding indicates that the environment has the principal role in causing sporadic cancer. The relatively large effect of heritability in cancer at a few sites (such as prostate and colorectal cancer) suggests major gaps in our knowledge of the genetics of cancer. (N Engl J Med 2000;343:78-85.)

British Journal of Cancer (2011) 105, S77 – S81

www.bjcancer.com

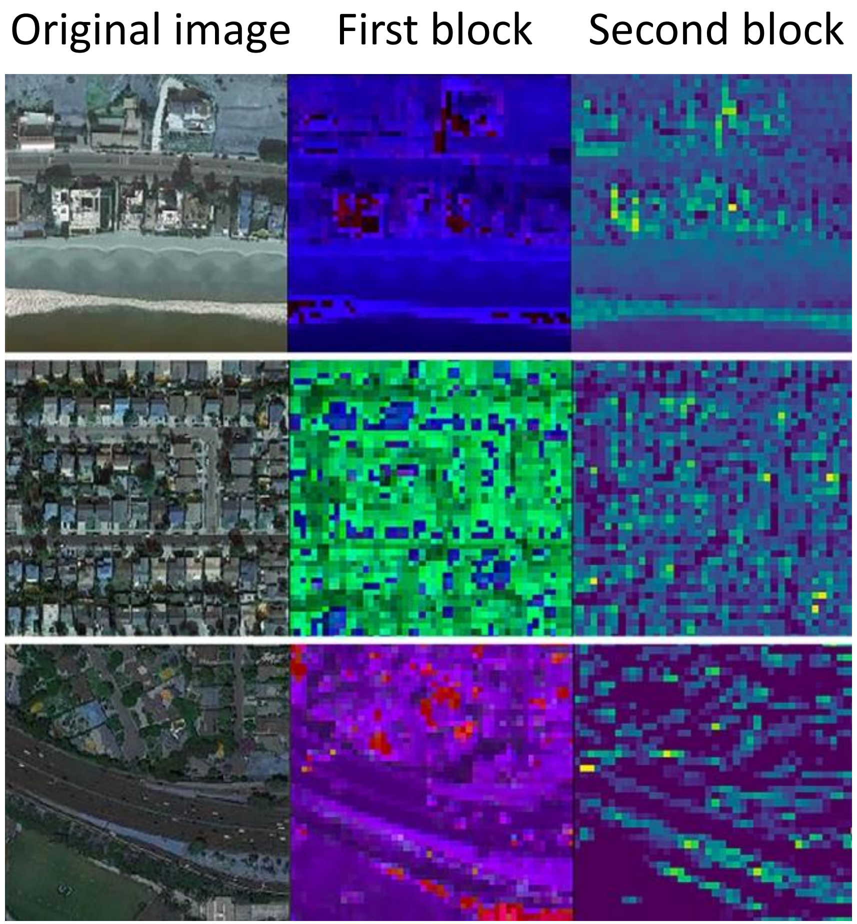
The fraction of cancer attributable to lifestyle and environmental factors in the UK in 2010

Population-attributable fractions provide a valuable quantitative appraisal of the impact of different factors in cancer causation, and are thus helpful in prioritising cancer control strategies. However, quantifying the likely impact of preventive interventions requires rather complex scenario modelling, including specification of realistically achievable population distributions of risk factors, and the timescale of change, as well as the latent periods between exposure and outcome, and the rate of change following modification in exposure level.



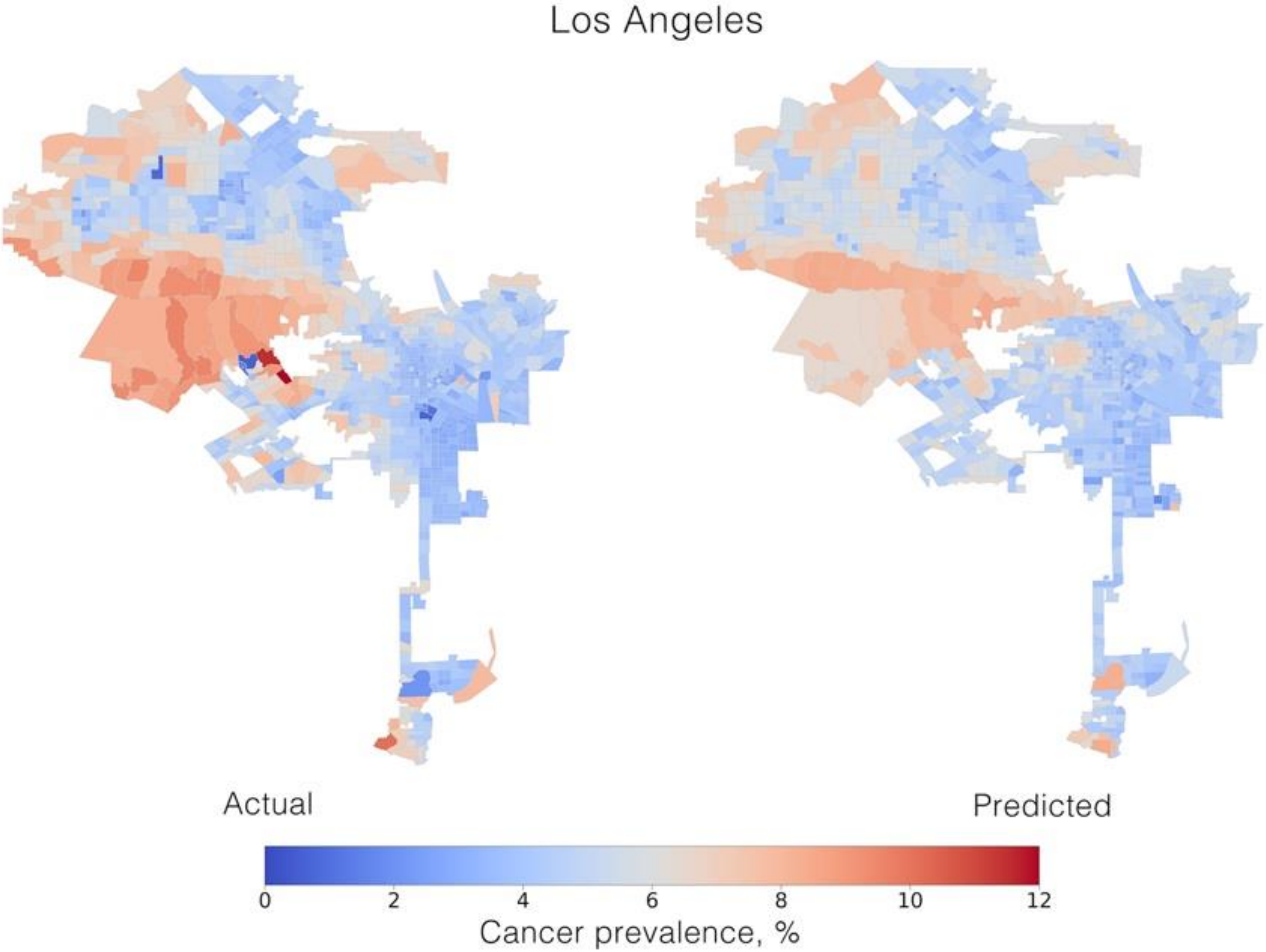
224 × 224 × 3
224 × 224 × 64
112 × 112 × 128
56 × 56 × 256
28 × 28 × 512
14 × 14 × 512
7 × 7 × 512
1 × 1 × 4096
1 × 1 × 1000
convolution+ReLU
max pooling
fully connected+ReLU
softmax

ResNet50

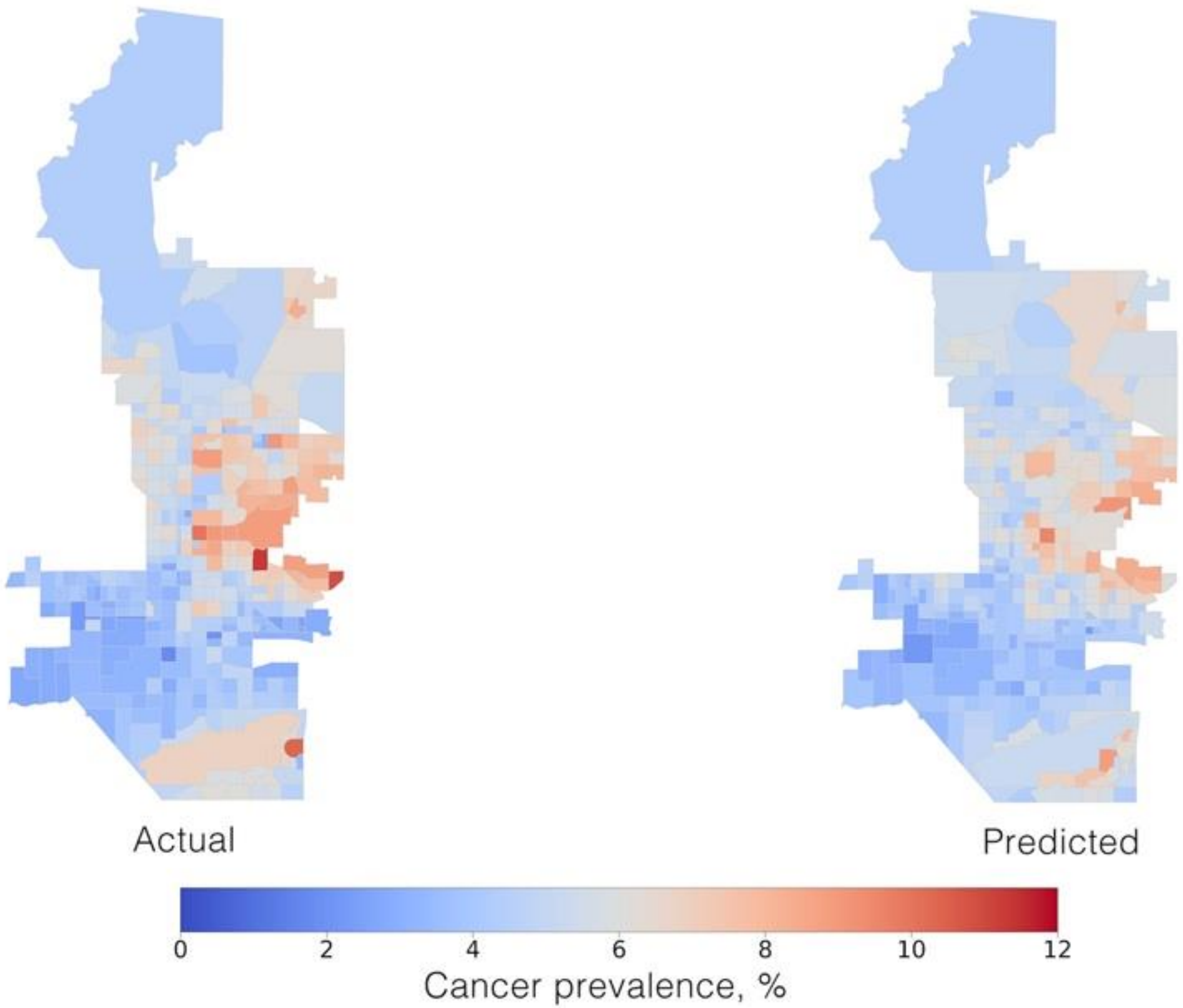


Résultats sur 7 villes

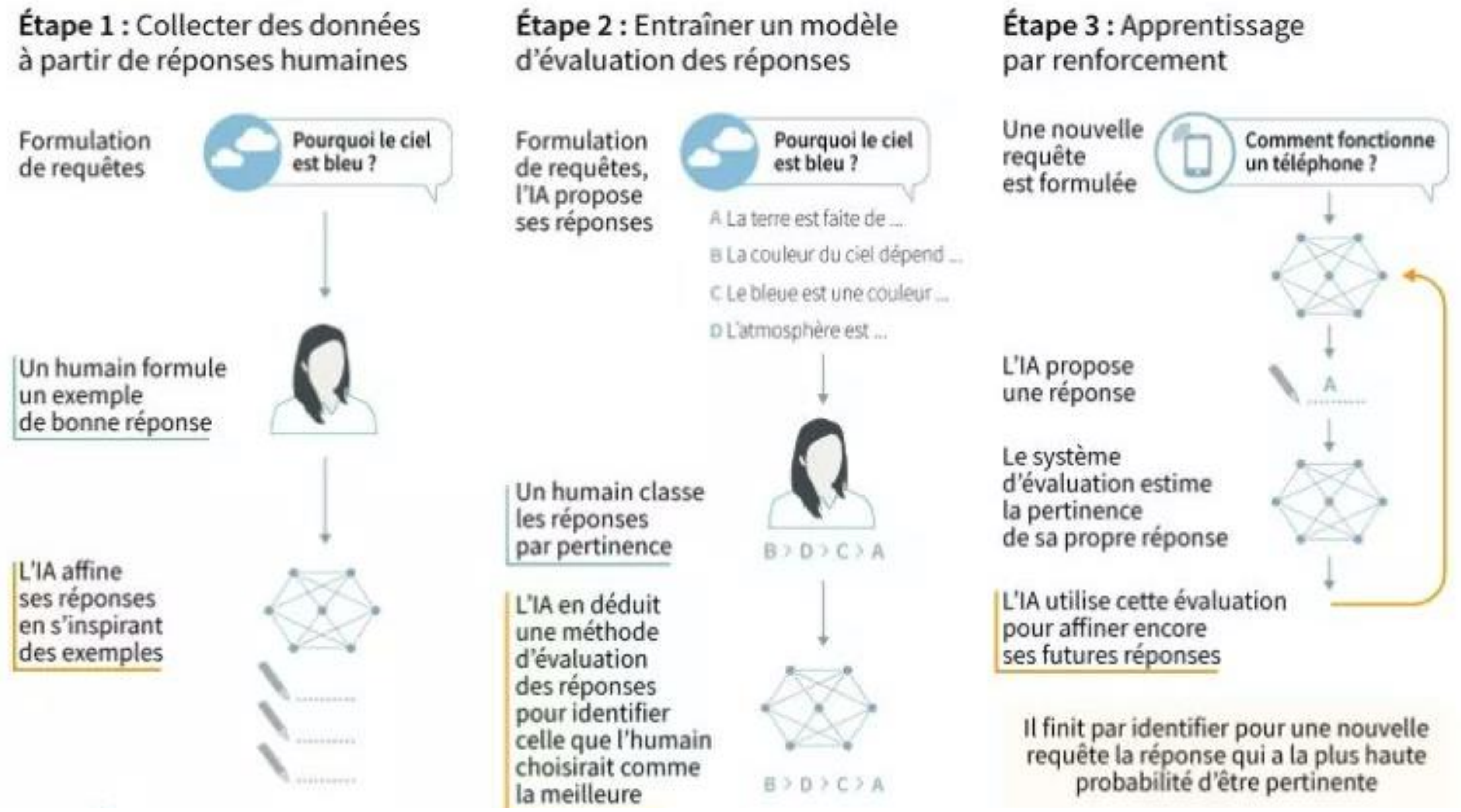
	Chicago	Dallas	Houston	Los Angeles	Phoenix	San Diego	San Jose
Population	2,705,994	1,345,047	2,325,502	3,990,456	1,660,272	1,425,976	1,030,119
Land area (square miles)	588.7	882.90	1,651.10	1,213.90	1,340.60	842.30	459.70
Population density (person / square miles)	4,596.56	1,523.44	1,408.46	3,287.30	1,238.45	1,692.96	2,240.85
Number of census	794	304	553	994	356	286	213
Actual cancer prevalence (%)	5.0	4.8	4.8	4.9	5.2	5.0	4.9
r^2 (%)	40.19 (+/- 5.23)	41.17 (+/- 20.77)	44.39 (+/- 9.60)	55.27 (+/- 5.28)	64.37 (+/- 6.05)	43.53 (+/- 9.74)	41.54 (+/- 15.23)
Mean squared error	1.17 (+/- 0.13)	2.56 (+/- 1.97)	1.36 (+/-0.28)	1.12 (+/-0.25)	0.96 (+/- 0.23)	2.30 (+/- 1.27)	1.16 (+/- 0.52)
Mean absolute error	0.81 (+/- 0.02)	1.06 (+/-0.16)	0.88 (+/-0.10)	0.76 (+/-0.05)	0.73 (+/- 0.07)	1.03 (+/- 0.14)	0.79 (+/- 0.10)


Los Angeles
Actual
Predicted
0
2
4
6
8
10
12
Cancer prevalence, %

Phoenix



4 - IA générative


Étape 1 : Collecter des données à partir de réponses humaines
Formulation de requêtes
Pourquoi le ciel est bleu ?
Un humain formule un exemple de bonne réponse
L'IA affine ses réponses en s'inspirant des exemples
Étape 2 : Entraîner un modèle d'évaluation des réponses
Formulation de requêtes, l'IA propose ses réponses
Pourquoi le ciel est bleu ?
A La terre est faite de ...
B La couleur du ciel dépend ...
C Le bleue est une couleur ...
D L'atmosphère est ...
Un humain classe les réponses par pertinence
B > D > C > A
L'IA en déduit une méthode d'évaluation des réponses pour identifier celle que l'humain choisirait comme la meilleure
B > D > C > A
Étape 3 : Apprentissage par renforcement
Une nouvelle requête est formulée
Comment fonctionne un téléphone ?
L'IA propose une réponse
A
Le système d'évaluation estime la pertinence de sa propre réponse
L'IA utilise cette évaluation pour affiner encore ses futures réponses
Il finit par identifier pour une nouvelle requête la réponse qui a la plus haute probabilité d'être pertinente

GPT4, Med-PaLM 2, etc

- Passe l'USMLE avec 85% de bonnes réponses (Kung et al, PLoS Dig. Health, 2023)

- Pose un diagnostic correct à partir de tableaux cliniques (symptômes)

87% de bonnes réponse
65% pour les médecins humains (Hailu et al, STAT, 2023)

- A plus d'empathie qu'un humain (Ayers et al, JAMA Int. Med., 2023)

→ Patient-facing AI

3 – Fiabilité et sécurité

Generative adversarial networks (GANs)

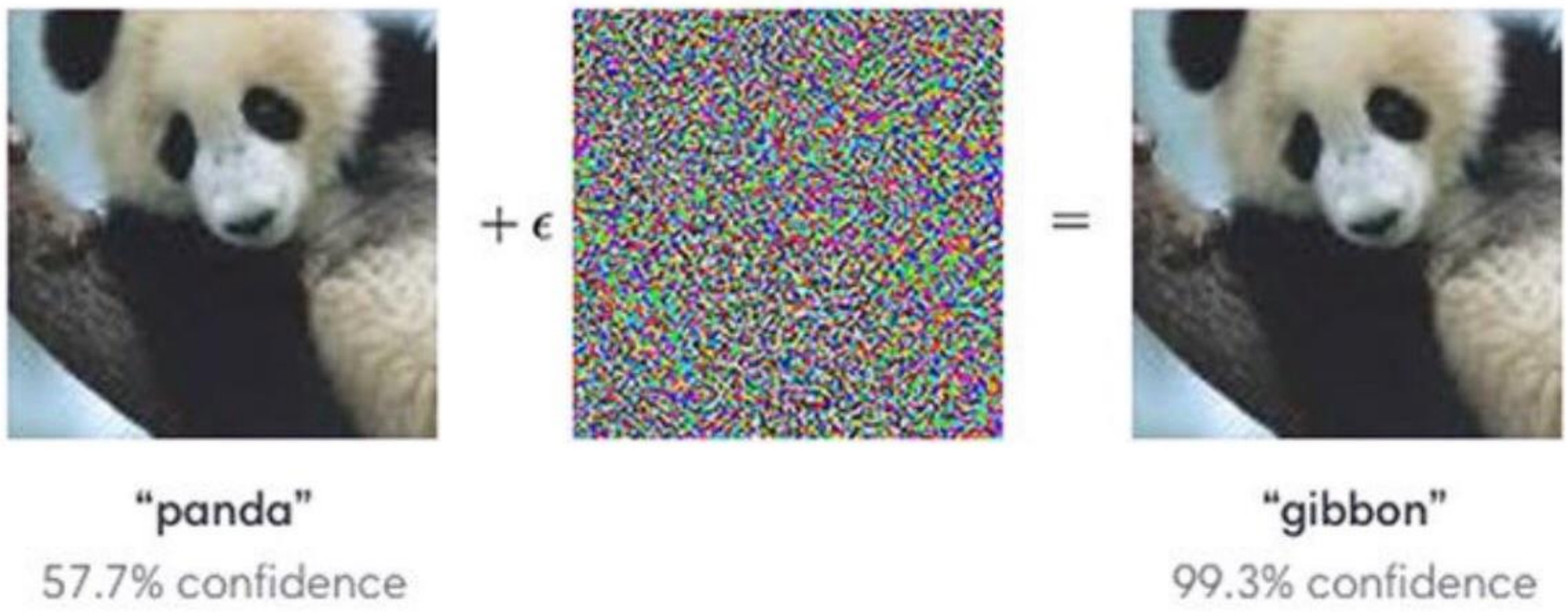


Image: OpenAI

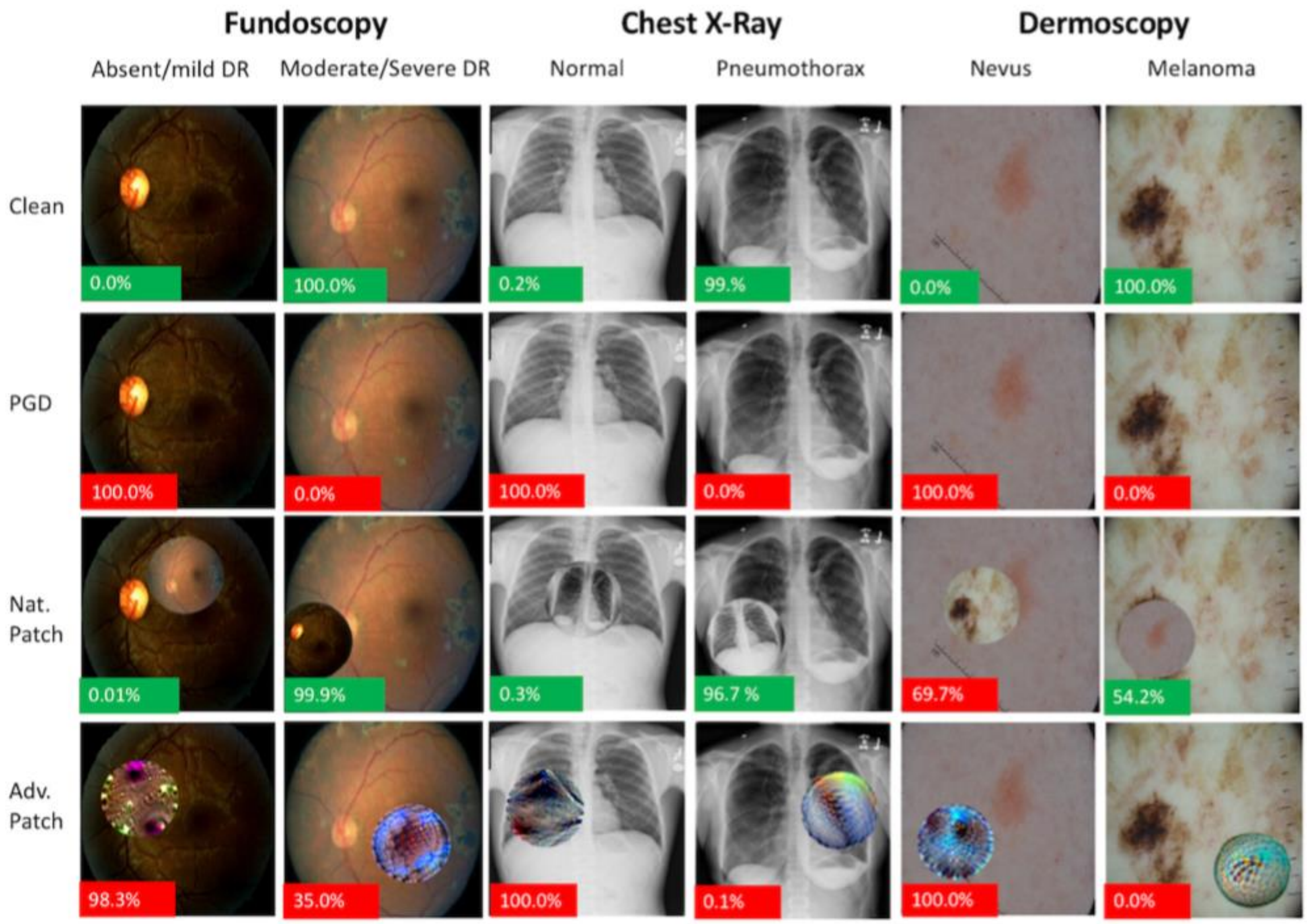
Adversarial Attacks Against Medical Deep Learning Systems

Samuel G. Finlayson
Department of Systems Biology
Harvard Medical School
Boston, MA 02130
samuel_finlayson@hms.harvard.edu

Hyung Won Chung
Massachusetts Institute of Technology
Cambridge, MA 02139
hwc@mit.edu

Isaac S. Kohane
Department of Biomedical Informatics
Harvard Medical School
Boston, MA 02130
isaac_kohane@hms.harvard.edu

Andrew L. Beam
Department of Biomedical Informatics
Harvard Medical School
Boston, MA 02130
andrew_beam@hms.harvard.edu


Fundoscopy
Chest X-Ray
Dermoscopy
Absent/mild DR
Moderate/Severe DR
Normal
Pneumothorax
Nevus
Melanoma
Clean
PGD
Nat. Patch
Adv. Patch
0.0%
100.0%
0.2%
99.%
0.0%
100.0%
100.0%
0.0%
100.0%
0.0%
100.0%
0.0%
0.01%
99.9%
0.3%
96.7 %
69.7%
54.2%
98.3%
35.0%
100.0%
0.1%
100.0%
0.0%

5 - Ethique

Risques

- Outils très puissants
- Risques très importants de dérives et d’usages non appropriés

- Rôle des médecins et de chercheurs de s’approprier ces outils pour en faire bon usage

- Besoin d’une législation ?

DEEP NEURAL NETWORKS CAN DETECT SEXUAL ORIENTATION FROM FACES

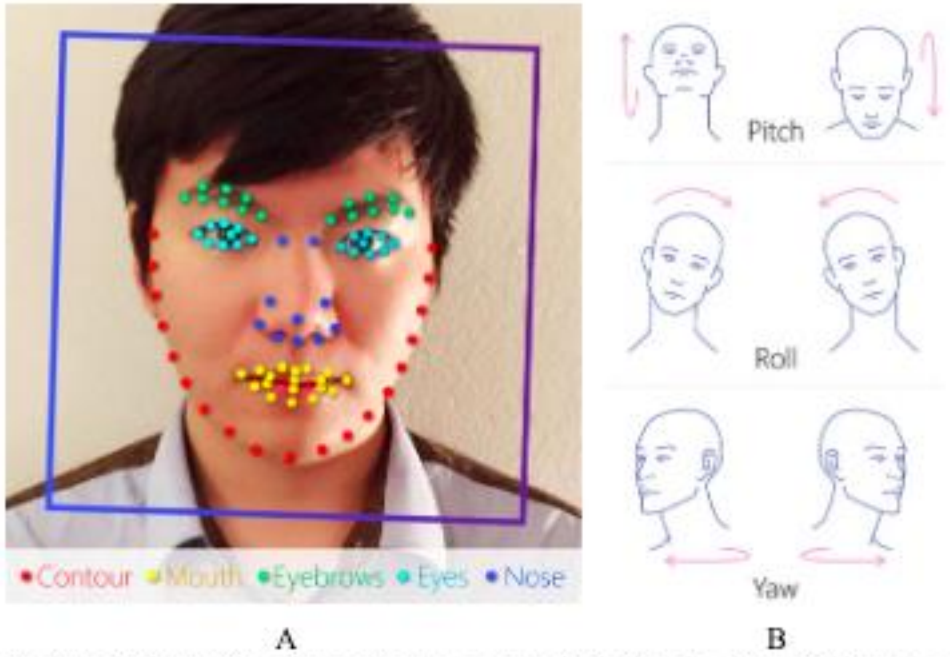
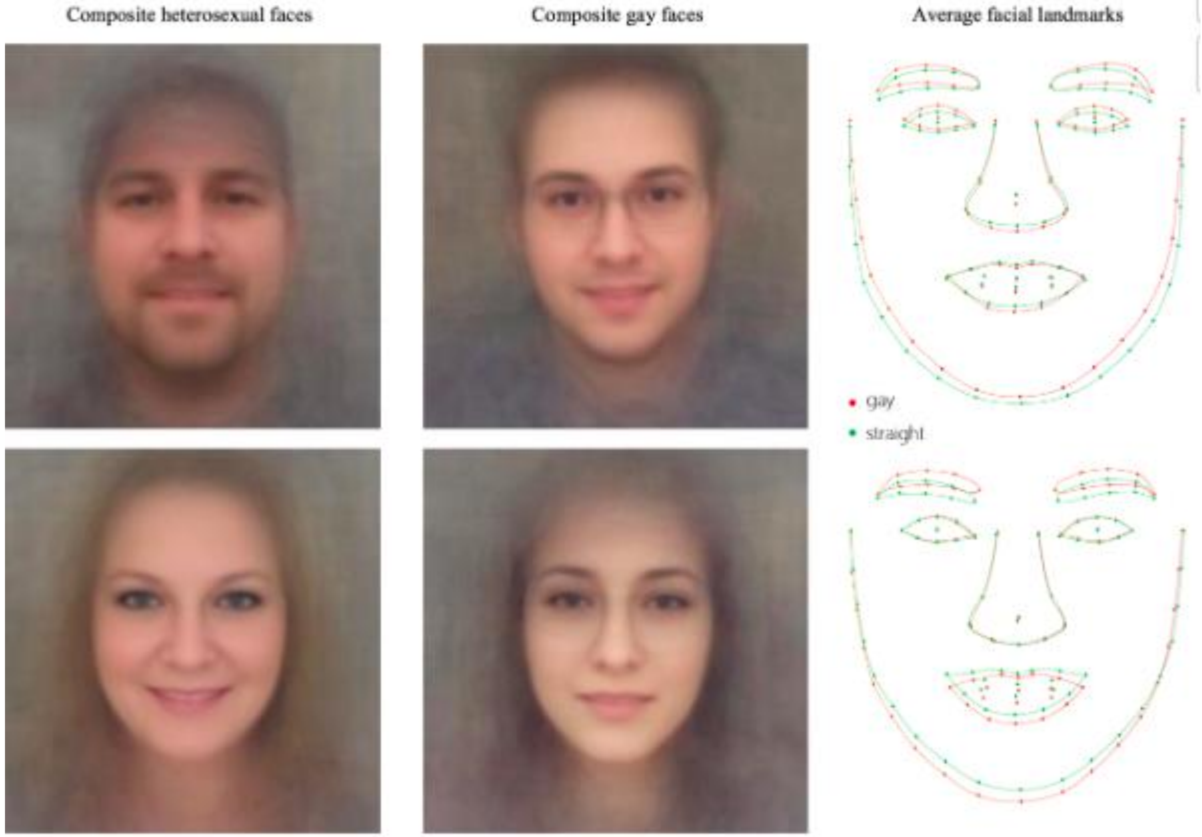


Figure 1. Graphical illustration of the outcome produced by Face++. Panel A illustrates facial landmarks (colored dots, n=83) and facial frame (blue box). Panel B illustrates pitch, roll, and yaw parameters that describe the head's orientation in space.

35,326 facial images

Classifier could correctly distinguish between gay and heterosexual men in 81% of cases, and in 74% of cases for women.

Human judges achieved much lower accuracy: 61% for men and 54% for women



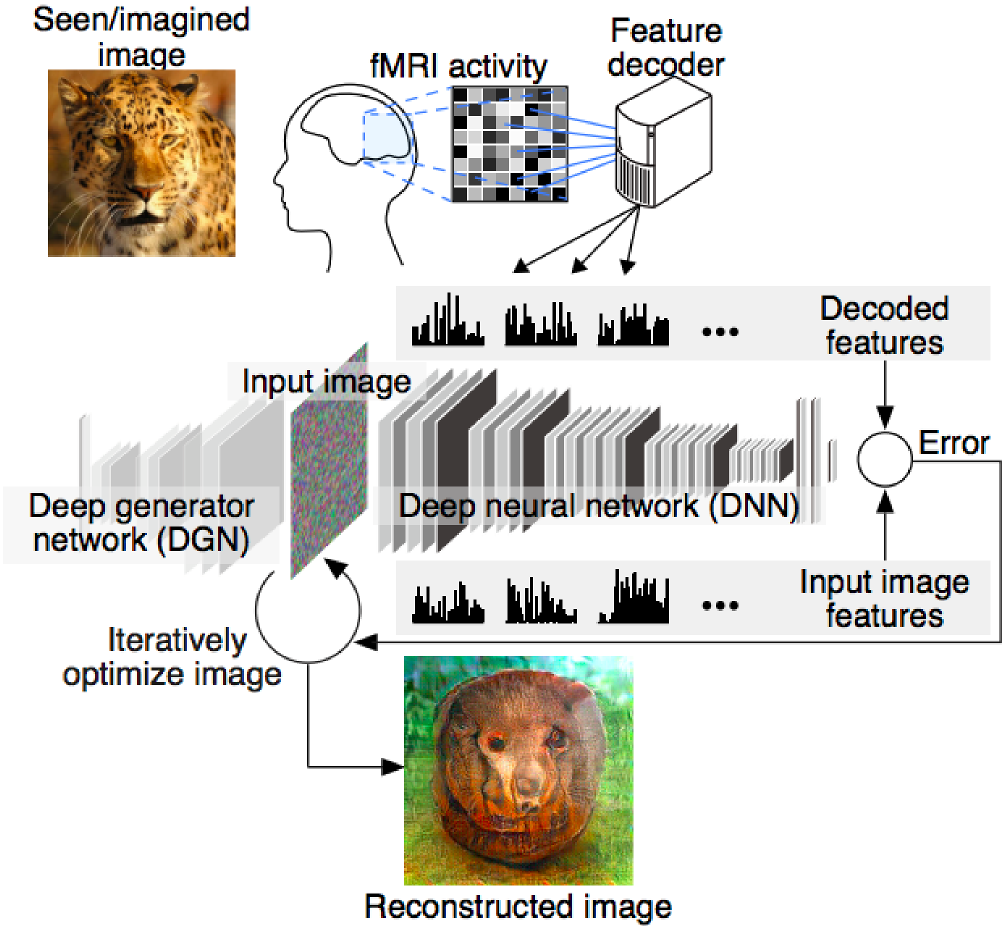
Yilun Wang, Michal Kosinski
Graduate School of Business, Stanford University

Journal of Personality and Social Psychology, 114(2), 246-257

Biais d’entraînement ?

- Etudes faites sur des populations caucasiennes
- Quid d’individus d’autres origines ?

Deep image reconstruction from human brain activity



Guohua Shen, Tomoyasu Horikawa, Kei Majima, and Yukiyasu Kamitani, ATR Computational Neuroscience Laboratories, Kyoto 619-0288, Japan 2 Kyoto University, Kyoto 606-8501, Japan

→ Lire les pensées ?

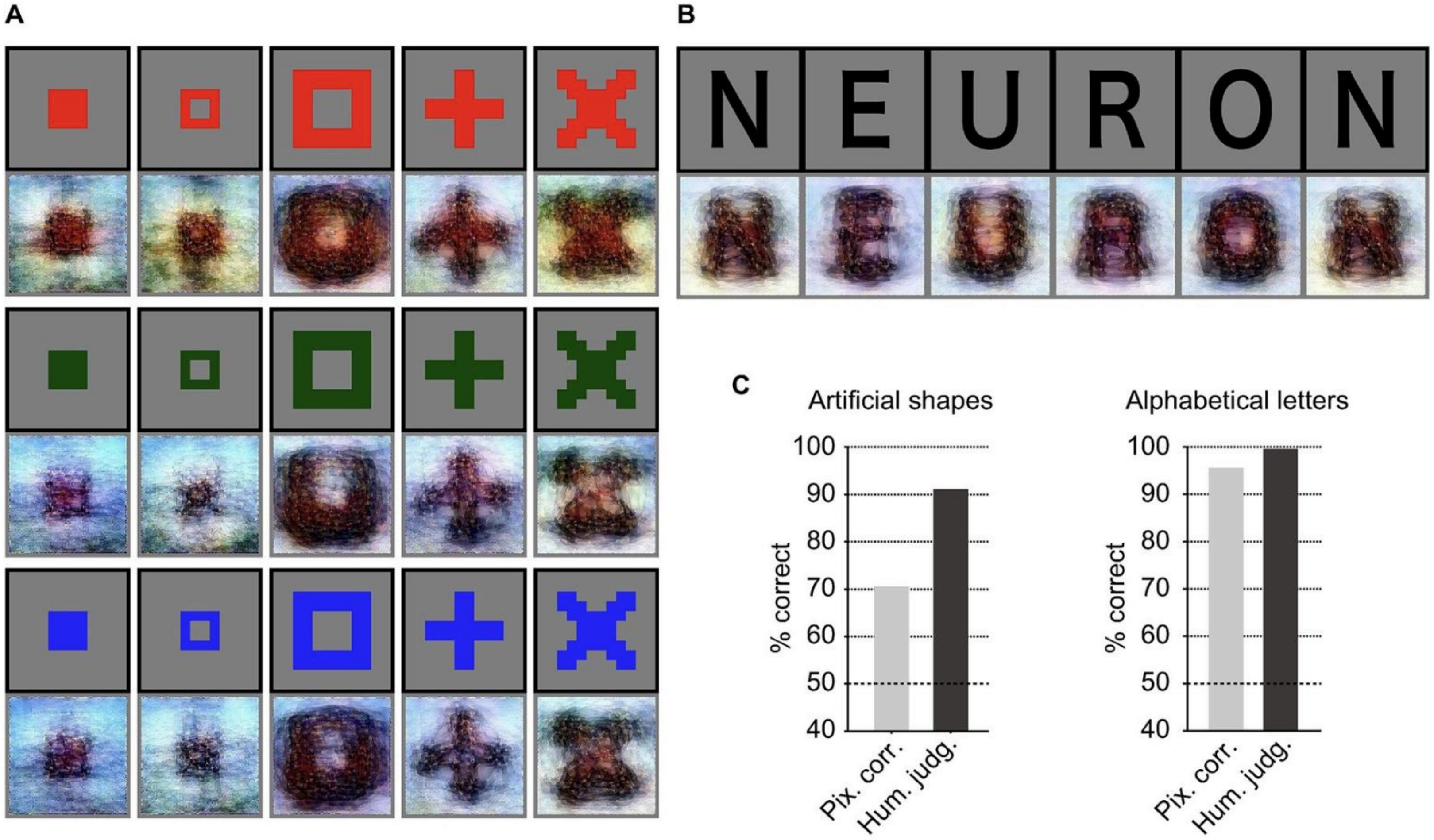


Fig 6. Seen artificial image reconstructions. The black and gray surrounding frames indicate presented and reconstructed images respectively (VC activity, DNN 1–8, without the DGN). (**A**) Reconstructions for seen artificial shapes. (**B**) Reconstructions for seen alphabetical letters. The reconstructed letters were arranged in the word: "NEURON". (**C**) Reconstruction quality of artificial shapes and alphabetical letters (three subjects pooled, $N = 120$ and 30 for artificial shapes and alphabetical letters, respectively; chance level, 50%).

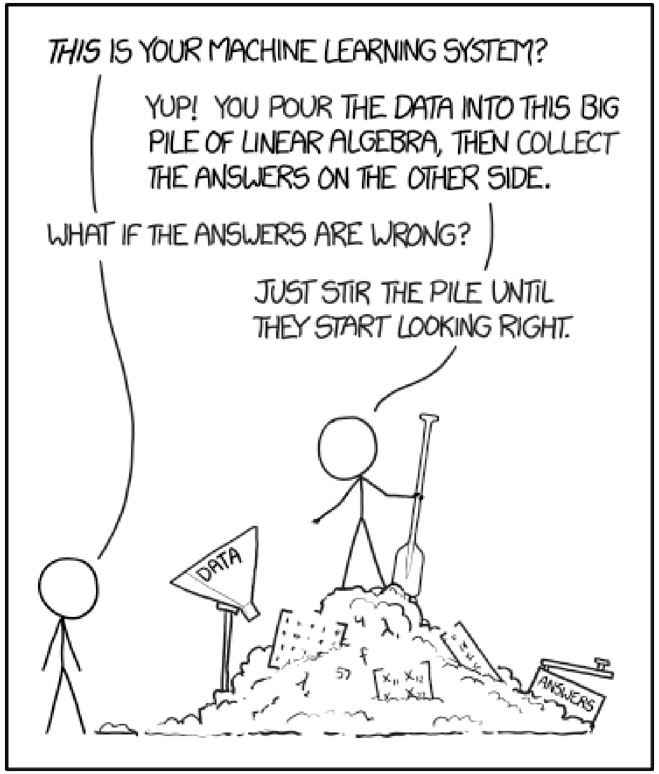
→ Lire les pensées ?

→ Lire les pensées ?

6- Perspectives

Ne pas négliger les limites de ces approches en médecine

- Comment en tester la précision future ?
- Comment valider leur utilisation en routine clinique ?
- Essai clinique randomisé ?
- Comment faire évoluer la médecine si tout repose sur l'entraînement d'algorithmes utilisant des données rétrospectives ?



Company	FDA Approval	Indication
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

Vers une IA « forte »

- Nécessité une IA capable d'accomplir plusieurs tâches différentes
- Problème : « catastrophic forgetting »

PathNet: Evolution Channels Gradient Descent in Super Neural Networks

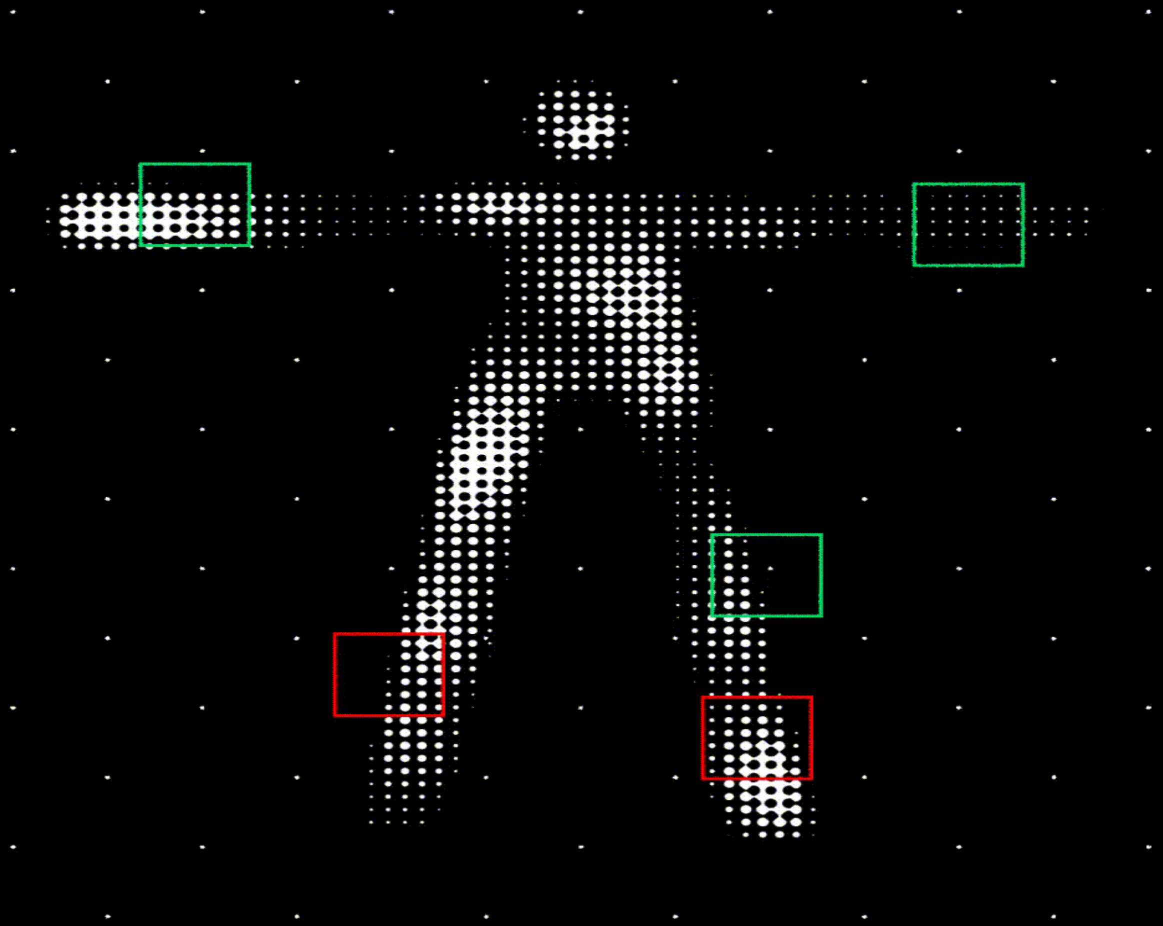
Chrisantha Fernando, Dylan Banarse, Charles Blundell, Yori Zwols, David Ha[†], Andrei A. Rusu, Alexander Pritzel, Daan Wierstra
Google DeepMind, London, UK. [†]Google Brain
chrisantha@google.com

ANNALS OF MEDICINE APRIL 3, 2017 ISSUE

A.I. VERSUS M.D.

What happens when diagnosis is automated?

By Siddhartha Mukherjee



Le médecin artificiel

- Redéfinition du rôle du médecin

- Moins de technique
- Plus d’humain, d’empathie

- Jusqu’à quand ?
  - Robots humanoïdes
  - Imitation des émotions humaines

Simuler (ou avoir ?) des émotions

Intelligent Machines

An AI Poker Bot Has Whipped the Pros

It's another seminal moment for machine learning, and a painful schooling for humans.

by Jamie Condliffe   January 31, 2017

MIT
Technology
Review

jean-emmanuel.bibault@aphp.fr
twitter.com/jebibault