université de BORDEAUX

Introduction to modeling, simulation and data science in oncology

3. Introduction to Statistical Learning

S. Benzekry

Formation « Ecole doctorale Mathématiques et Informatique »

MONC

Modeling in ONCology

Ok Google: What is AI??

Definition: A computer system able to perform tasks that normally require human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages.

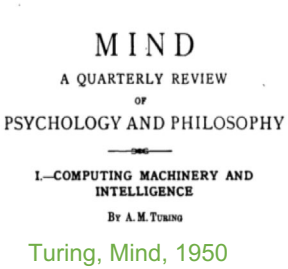
Oxford dictionary

- Exists since decades

- New « hype » since ~2011 mostly thanks to :
  - Computing power
  - Big data

Deep Learning

Enigma



MIND

A QUARTERLY REVIEW

OF

PSYCHOLOGY AND PHILOSOPHY

I.—COMPUTING MACHINERY AND INTELLIGENCE

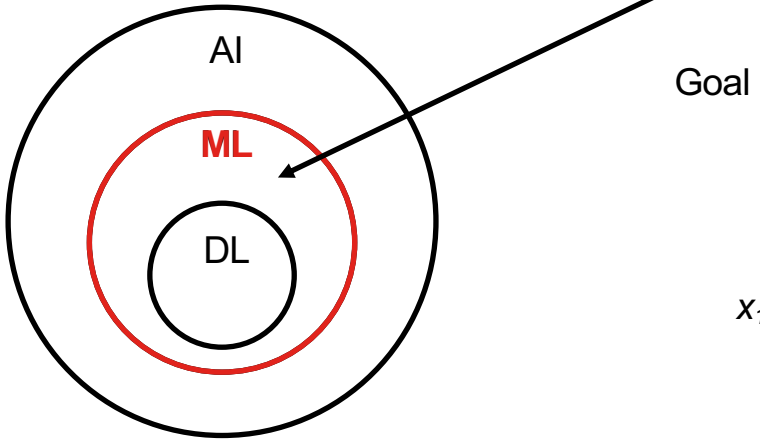
BY A. M. TURING

Turing, Mind, 1950

©Information Age, ACS

Google DeepMind, Nature, 2016

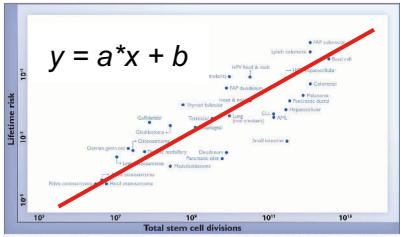
Artificial Intelligence, Machine Learning and Deep Learning



ML = machine (automatic) learning

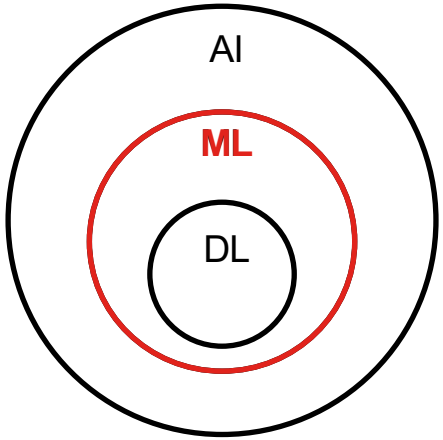
Goal = predict outcome *y* as a function of input/features $x_1, \ldots, x_n$

x_1, x_2, x_3 —



→ *y*

Artificial Intelligence, Machine Learning and Deep Learning



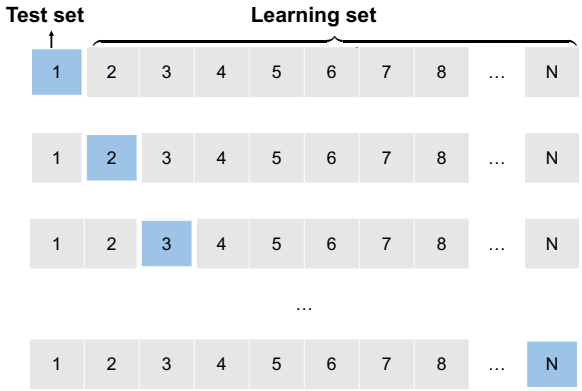
Supervizec

patient

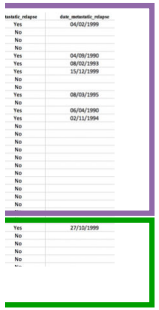
patient

patient

k-fold cross-validation



ıtcome



y^1

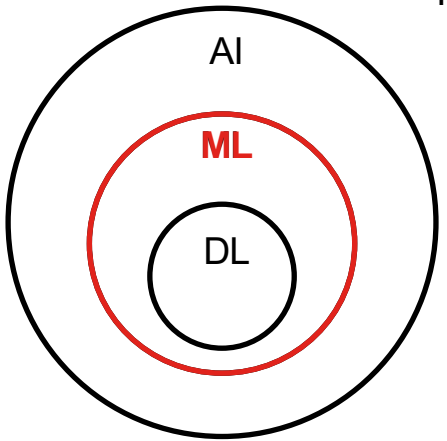
y^2

y^3

Artificial Intelligence, Machine Learning and Deep Learning

Supervized machine learning

Predict outcome *y* as a function of *a lot* of input/features $x_1,\ldots,x_n$ (big data, high dimensional statistics)



patient 1 $x_1^1, x_2^1, \cdots, x_{1000}^1$

patient 2 $x_1^2, x_2^2, \cdots, x_{1000}^2$

$\vdots$

patient 1000 $x_1^{1000}, x_2^{1000}, \cdots, x_{1000}^{1000}$

→ **Model**

$a_1*x_1 + a_2*x_2 + \ldots + a_{1000}*x_{1000}$

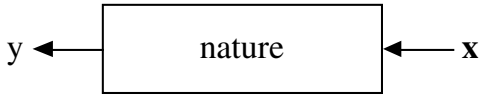
→ y^1, y^2, $\vdots$, y^{1000}

$a_1, a_2, \ldots, a_{1000}$

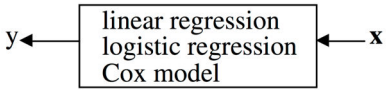
Statistical Science
2001, Vol. 16, No. 3, 199–231

Statistical Modeling: The Two Cultures

Leo Breiman

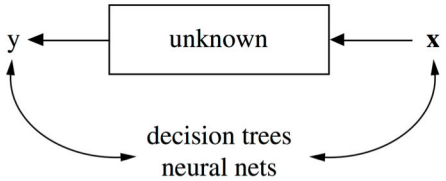


The data modeling culture



Model validation. Yes–no using goodness-of-fit tests and residual examination.
Estimated culture population. 98% of all statisticians.

The algorithmic modeling culture



Model validation. Measured by predictive accuracy.
Estimated culture population. 2% of statisticians, many in other fields.

Statistical (Machine) Learning

$$Y = M(x) + \varepsilon$$

- Goal = predict Y = find M
- Y quantitative ⇒ regression
- Y qualitative ⇒ classification
- Learning = use data
  - only x values ⇒ unsupervised learning
  - (x, y) examples ⇒ supervised learning

3.1 Classification

Confusion matrix

Data $\begin{pmatrix} x^1 \\ \vdots \\ x^N \end{pmatrix}$ $\longrightarrow$ Predictions $\begin{pmatrix} \hat{y}^1 \\ \vdots \\ \hat{y}^N \end{pmatrix} = \begin{pmatrix} \hat{M}(x^1) \\ \vdots \\ \hat{M}(x^N) \end{pmatrix}$ vs reality $\begin{pmatrix} y^1 \\ \vdots \\ y^N \end{pmatrix}$

Model \ Actual	1	0
+	TP (Sensitivity)	FP
-	FN	TN (Specificity)

Accuracy = $\frac{TP+TN}{TP+TN+FP+TN}$

Sensitivity = $SE = \mathbb{P}(+|1) = TPR = \frac{TP}{TP+FN}$

$\beta = \mathbb{P}(-|1) = FNR = 1 - SE =$ proba of type II error

(classify as benign what is cancer)

Specificity = $SP = \mathbb{P}(-|0) = TNR = \frac{TN}{FP+TN}$

$\alpha = \mathbb{P}(+|0) = FPR = 1 - SP =$ proba of type I error

(classify as tumor what is benign)

Positive and negative predictive value

- Sensitivity and specificity are not sufficient to assess a model
- We are often more interested in $\mathbb{P}(1|+)$ (= positive predictive value, PPV) and $\mathbb{P}(0|-)$ (= negative predictive value, NPV)
- From Bayes

$$PPV = \mathbb{P}(1|+) = \frac{\mathbb{P}(+|1)\mathbb{P}(1)}{\mathbb{P}(+)}$$

p prevalence

$$\mathbb{P}(+) = \mathbb{P}(+|0)\mathbb{P}(0) + \mathbb{P}(+|1)\mathbb{P}(1) = \big(1 - \mathbb{P}(-|0)\big)\big(1 - \mathbb{P}(1)\big) + SE \cdot \mathbb{P}(1)$$
$$= (1 - SP) \cdot (1 - p) + SE \cdot p$$

$$PPV = \frac{SE \cdot p}{(1 - SP) \cdot (1 - p) + SE \cdot p}$$

- Other metrics: $F1$ = harmonic mean of PPV (precision) and sensitivity (recall) = $2(PPV^{-1} + SE^{-1})^{-1}$

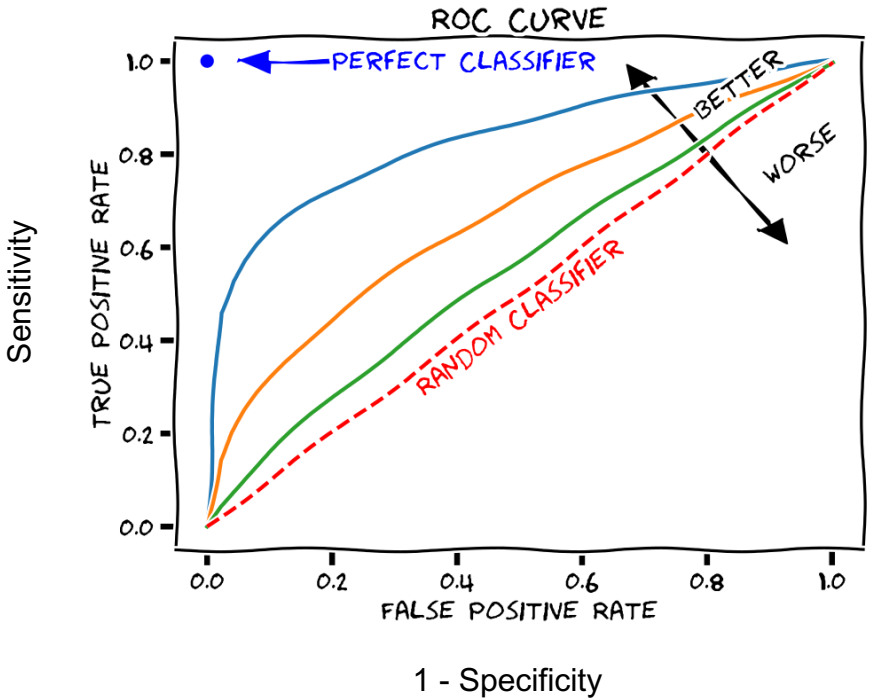
Example: Lung cancer and smoking status

- Percentage of smokers among lung cancer patients = 90%, i.e. SE of a model based on smoking status is 0.9
- Approx. 30% of population is composed of smokers ⇒ $SP (= TNR$, i.e. proportion of people who don't smoke and don't have cancer) is 70%.
- Assuming a lifetime risk of having lung cancer of 7.19% (= prevalence)

$$PPV = \mathbb{P}(\text{lung cancer during lifetime } | \text{smoker}) = 18.9\%$$

ROC curve analysis

- In practical cases a classification model often assigns a score (e.g. proba)
- For each value of a threshold, one SE and one SP value
- Global quantification of performances = area under the curve (AUC)



Interpretation of AUC

$AUC =$ probability that a random pair of predictions $(\hat{y}^1, \hat{y}^2)$ is concordant with the observations i.e that the score of $\hat{y}^1$ is larger than the score of $\hat{y}^2$ if $y^1 > y^2$.

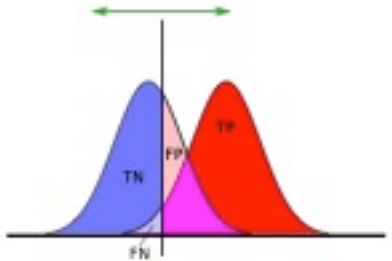
- $S_1 =$ score in class we want to classify as positive (say, malignant), density f_1
- S_0 = score in other class (say, healthy/benign), density f_0
- $T =$ threshold

$$AUC = \int_{T_{max}}^{T_{min}} SE(T)d(FPR(T))$$

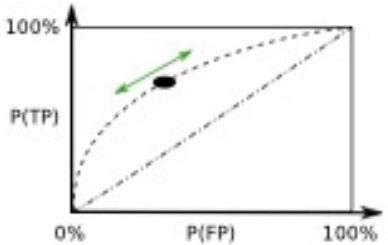
$$SE(T) = \mathbb{P}(S \geq T|1) = \int_{T}^{T_{max}} f_1(x)dx$$

$$FPR(T) = \mathbb{P}(S \geq T|0) = \int_{T}^{T_{max}} f_0(x)dx$$

$$AUC = \int_{T_{min}}^{T_{max}} \int_{T}^{T_{max}} f_1(x)f_0(T)dT$$
$$= \mathbb{P}(S_1 \geq S_0)$$



TP	FP
FN	TN



Logistic regression

$$p = \mathbb{P}(Y = 1) \in (0,1) \xrightarrow{?} \mathbb{R}$$

$$\frac{p}{1-p} \in (0, +\infty) = \frac{\mathbb{P}(Y = 1)}{\mathbb{P}(Y = 0)} = \text{odds} \simeq \text{chance}$$

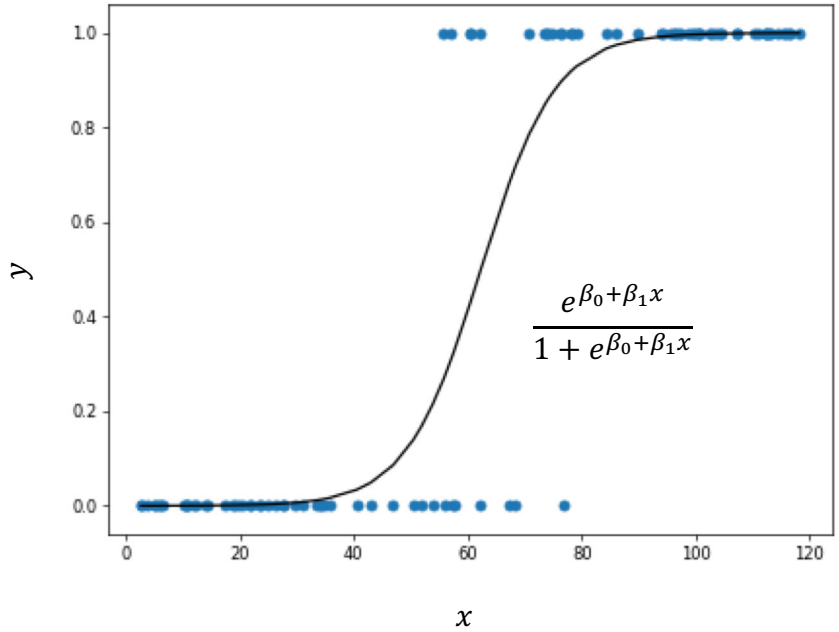
$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_L x_L$$

$$\Leftrightarrow p = \frac{e^{\beta_0+\beta_1 x_1+\cdots+\beta_L x_L}}{1 + e^{\beta_0+\beta_1 x_1+\cdots+\beta_L x_L}} = \pi(x)$$

Estimation: likelihood maximization

Interpretation: for one variable x, $\text{odds}(x) = e^{\beta_0+\beta_1 x}$

$$\Rightarrow \beta = \frac{\text{odds}(x+1)}{\text{odds}(x)} = \text{odds ratio} = OR$$

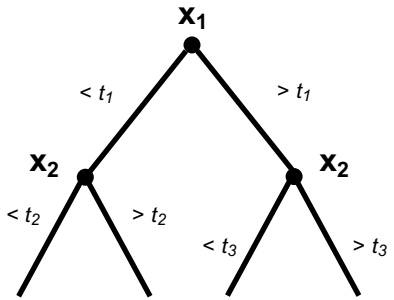


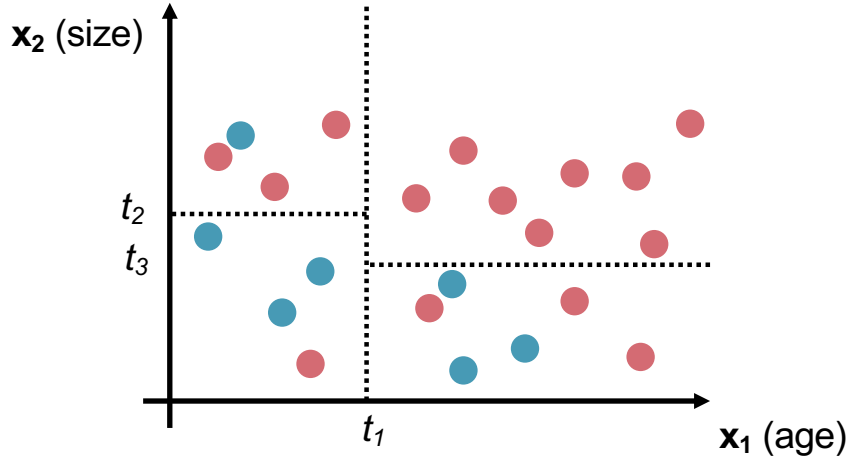
if $OR = 1.5$ there is a 50% increase of chance of having $Y = 1$ for an increase of x of one unit

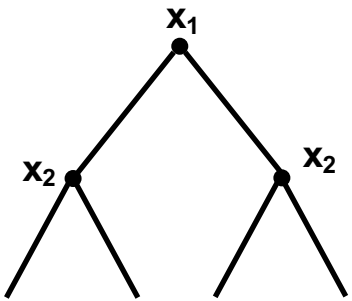
Random Forests

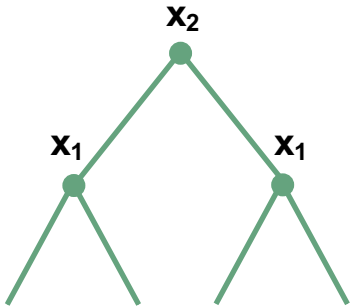
= Tumor
= Benign

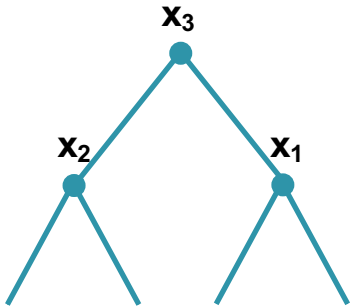
- Binary classification/regression tree
- Sample dataset and choose splitting features randomly to get a forest
- Vote/average over trees

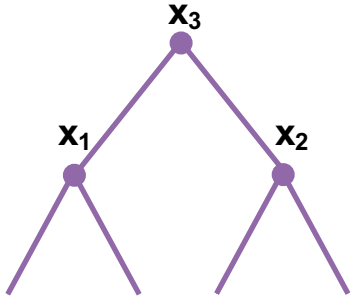






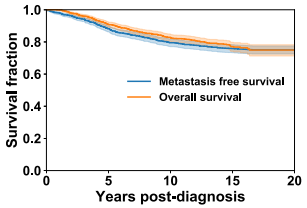






Breiman, Machine Learning, 2001

Example: prediction of 5-years metastatic relapse in early-stage breast cancer



K = 25 features

outcome

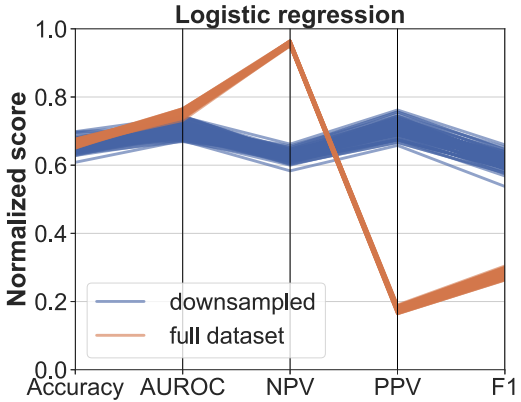
n =642 patients w/o adjuvant

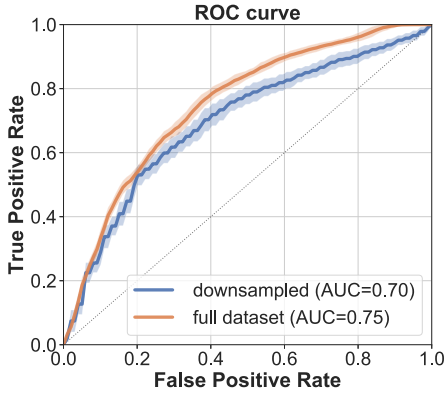
menopausal_status	ER	PR	Ki67	HER2	HER2_intensity	CK56	EGFR	VIM	ALDH1
Post-ménopause	20	0	0	0	0	0	0	0	0
Ménopause	40	95	8	0	0	0	0	0	0
Activité génitale	87	10	26	0	0	0	0	80	0
Post-ménopause	100	100	8	0	0	0	0	0	0
Post-ménopause	0	0	16	82	+++	0	0	0	0
Activité génitale	100	95	12	0	0	0	0	0	1
Activité génitale	56	100	17	0	0	0	0	0	0
Activité génitale	57	85	23	100	+++	0	0	0	0
Post-ménopause	80	5	20	0	0	0	0	0	0
Post-ménopause	0	0	15	100	+++	0	5	0	0
Post-ménopause	100	80	10	0	0	0	0	0	0
Post-ménopause	30	0	5	0	0	0	0	0	0
Post-ménopause	0	0	15	40	+++	0	0	0	0
Ménopause	0	80	8	0	0	0		0	0
Post-ménopause	0	0	27	0	0	0	30	0	1
Post-ménopause	0	0	56	0	0	80	60	100	0
Activité génitale	50	92	2	1	+	0	0	0	0
Post-ménopause	0	47	5	0	0	0	0	80	0
Post-ménopause	65	0	10	0	0	0	0	60	0
Post-ménopause	100	50	11	0	0	0	0	0	0
Ménopause	20	100	0	0	0	0	0	0	0
Activité génitale	90	6	5	0	0	0	0	0	0
Post-ménopause	100	3	5	0	0	0	0	0	0
Activité génitale	0	0	6	0	0	0	0	0	0
Ménopause	80	100	5	0	0	0	0	0	0
Post-ménopause	100	85	25	0	0	0		0	0
Post-ménopause	10	45	11	13	+++	0	0	0	0
Post-ménopause	66	1	2	40	++	0	0	0	0

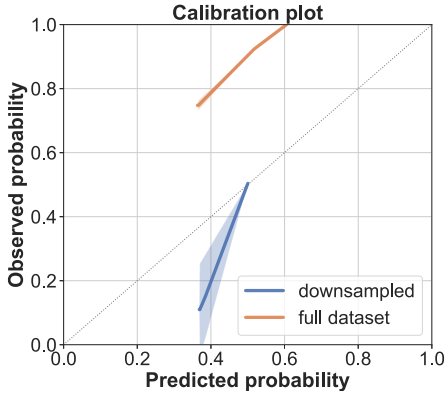
metastatic_relapse	date_metastatic_relapse
Yes	04/02/1999
No	
No	
No	
Yes	04/09/1990
Yes	08/02/1993
Yes	15/12/1999
No	
No	
Yes	08/03/1995
No	
Yes	06/04/1990
Yes	02/11/1994
No	
No	
No	
No	
No	
No	
No	
No	
No	
No	
No	
No	
No	
No	
Yes	27/10/1999
No	
No	
No	
No	

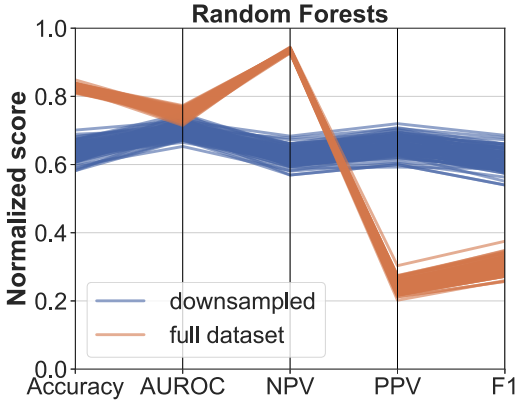
Institut Bergonié, Bordeaux, FR

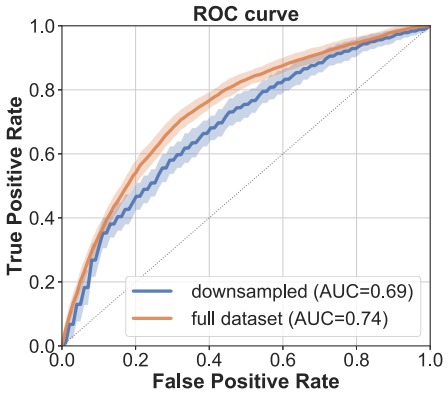
Prediction results

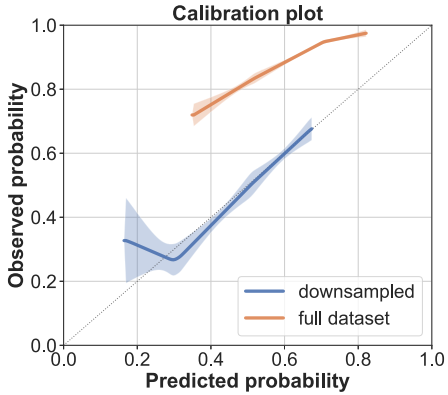




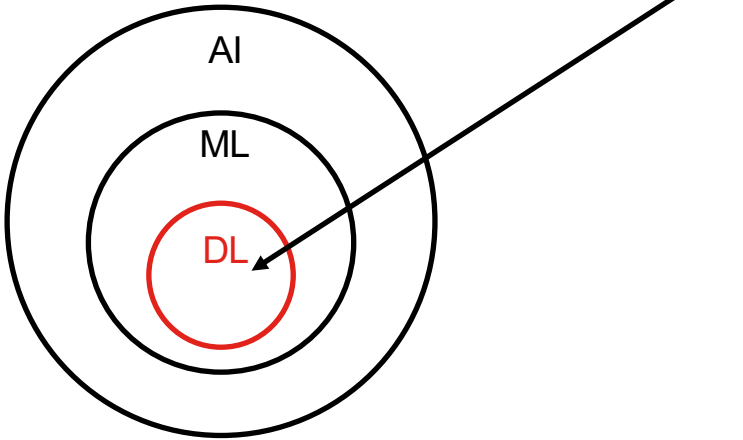




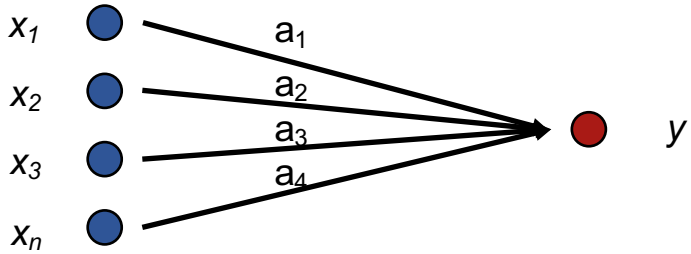




Neural networks

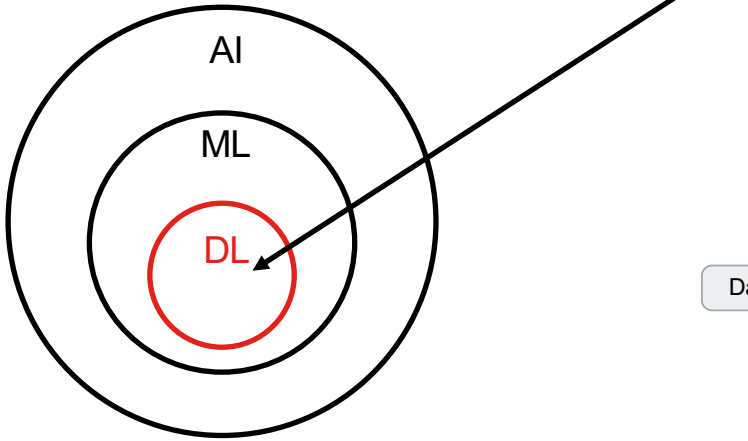


DL: artificial neural networks

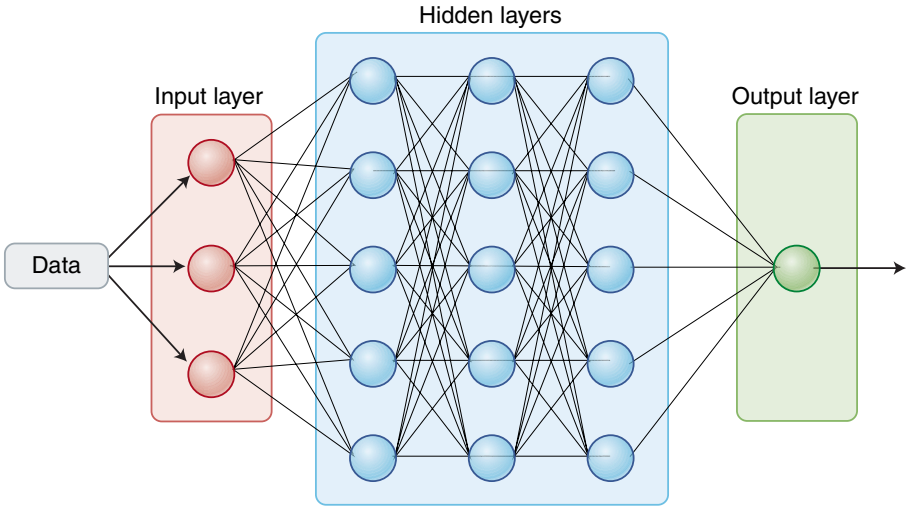


$y = f(b + a_1x_1 + \ldots + a_nx_n)$, f = logistic, softmax, …

Neural networks

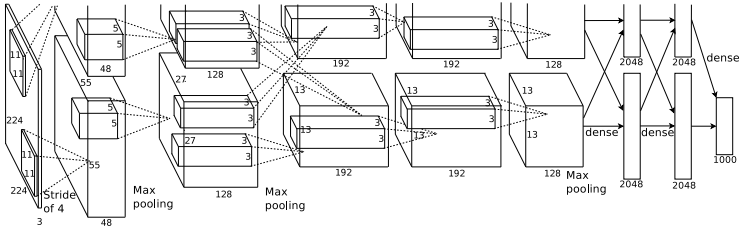


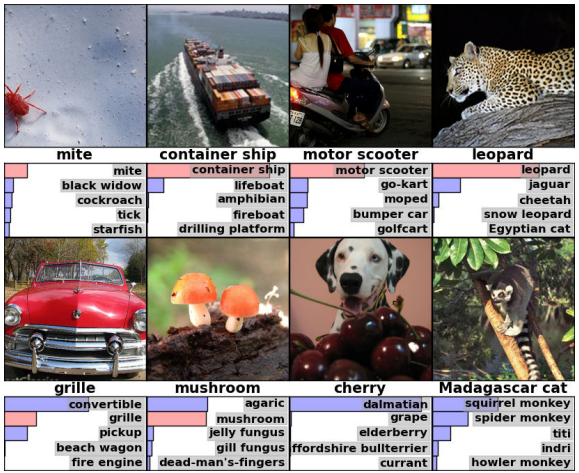
DL: artificial neural networks

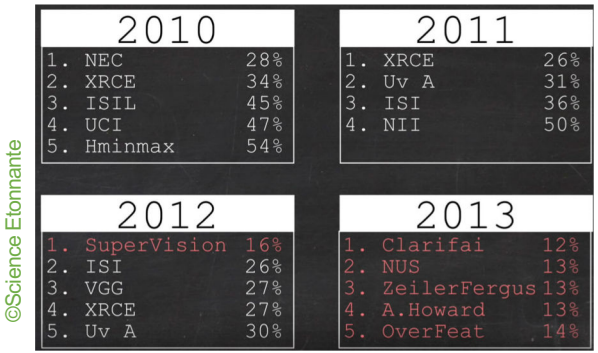


Success example of DL: computer vision

- 1.2 million images (ImageNet, Stanford) used to train a deep convolutional neural network



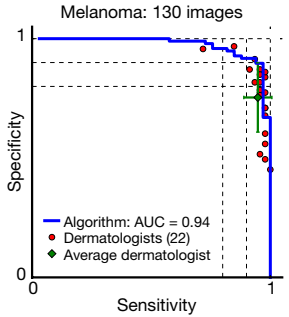


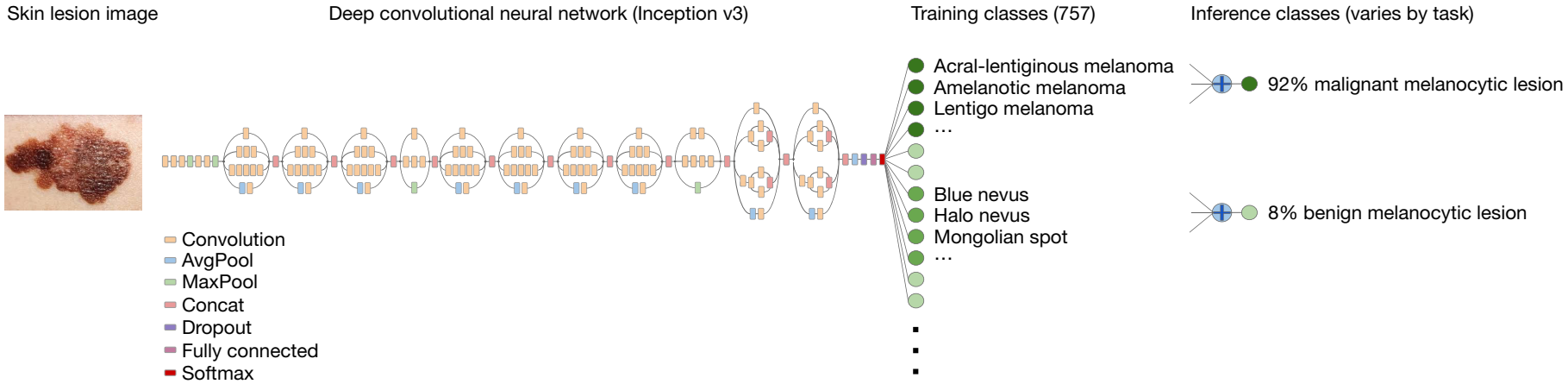


Krizhevsky, A. et al. ImageNet classification with deep convolutional neural networks, NIPS, 2012

Classification of skin lesions

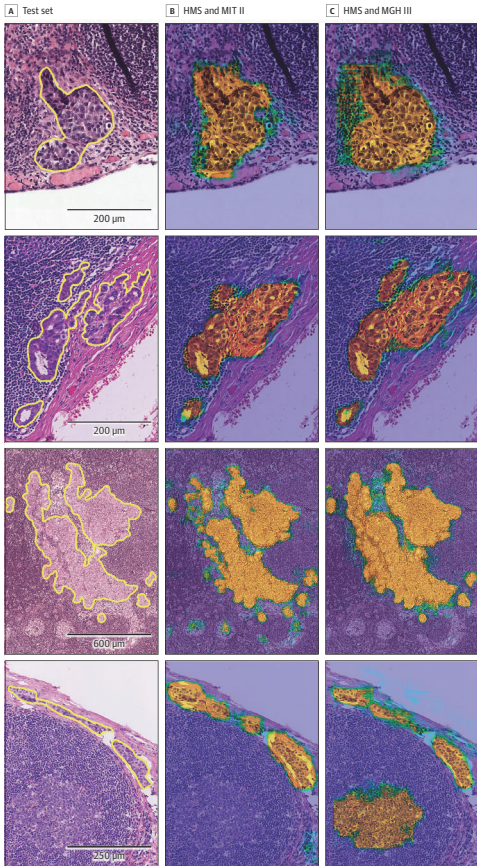
- 129 450 anntotated images
- Task = prediction benign/malignant
- Similar performances as dermatologists



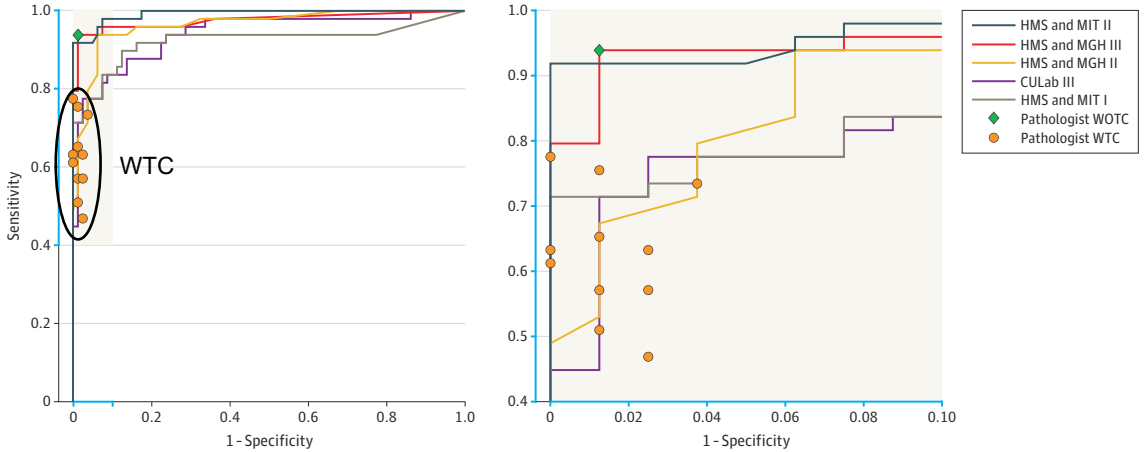


Esteva et al. (Stanford), Dermatologist-level classification of skin cancer with deep neural networks, Nature, 2017

Detection of lymph node metastases from histological images

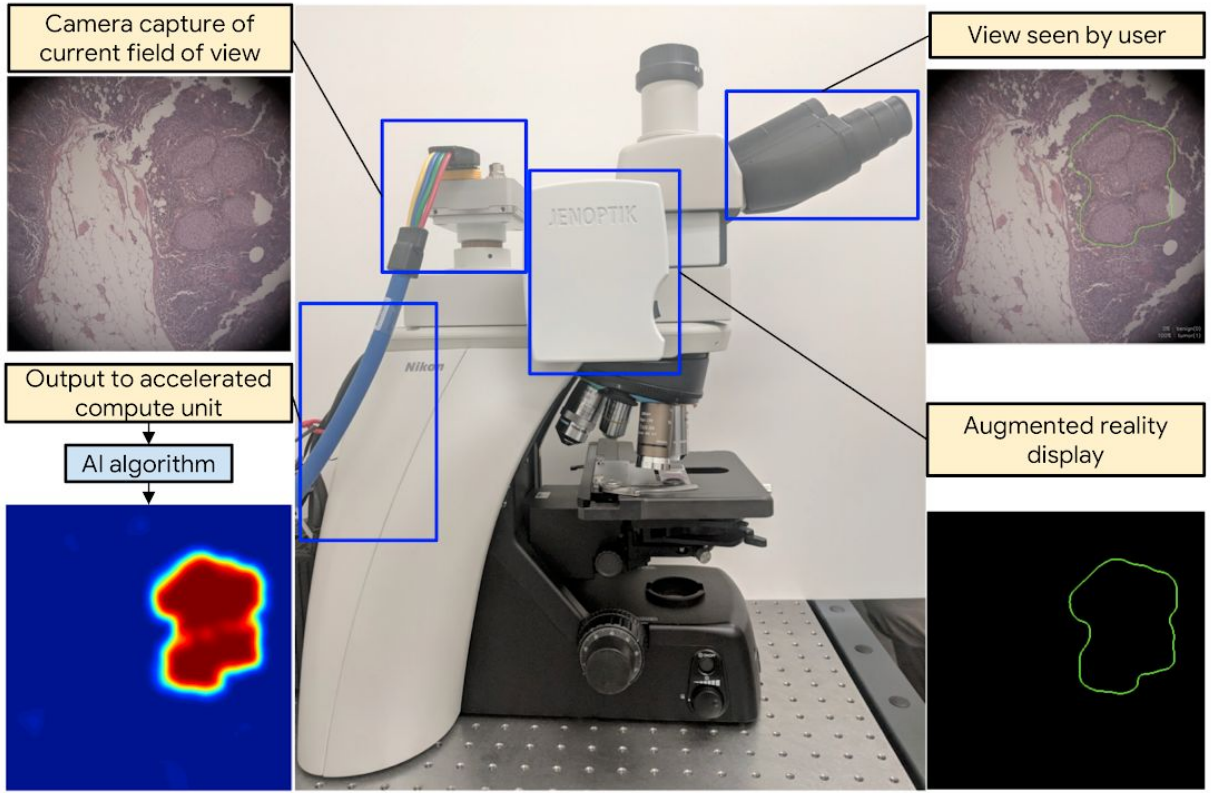


- One pathology slide = several gigapixels
- Best algorithms of the challenge = Deep Learning
- Same performances as pathologists without time constraint, but significatively better than 11 pathologists with constraint (WTC)



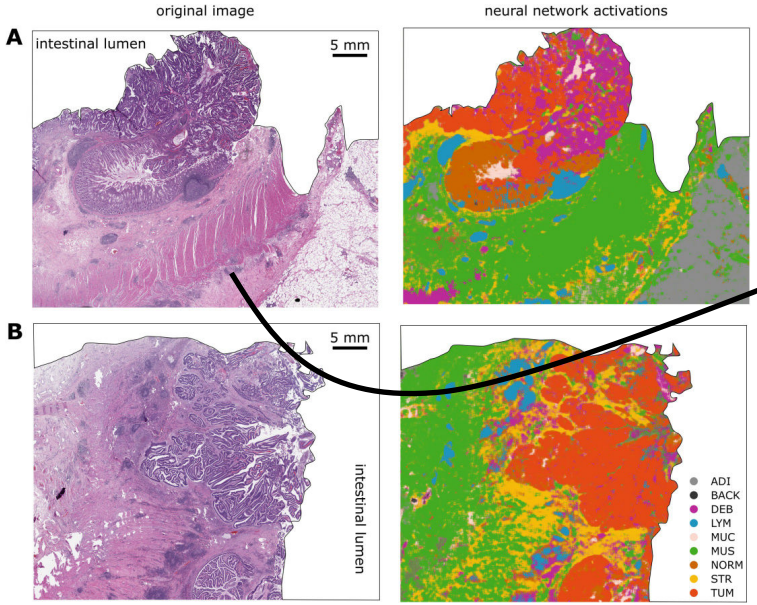
Bejnordi et al., Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer, JAMA, 2017

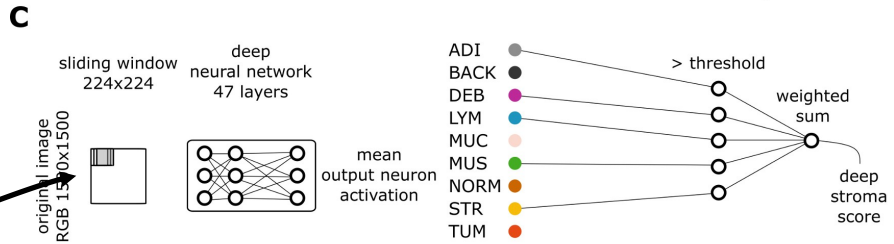
Microscope 2.0



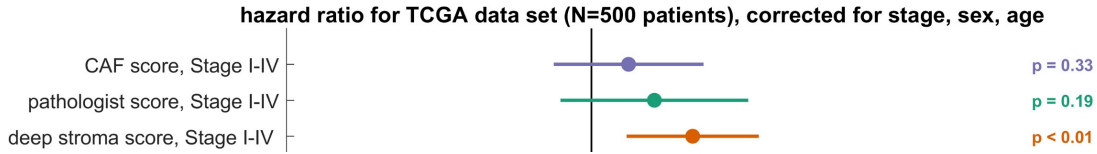
Chen et al. (Google AI Healthcare), Microscope 2.0: An Augmented Reality Microscope with Real-time Artificial Intelligence Integration, arXiv, 2018

Quantitative analysis of histopathological slides in CRC





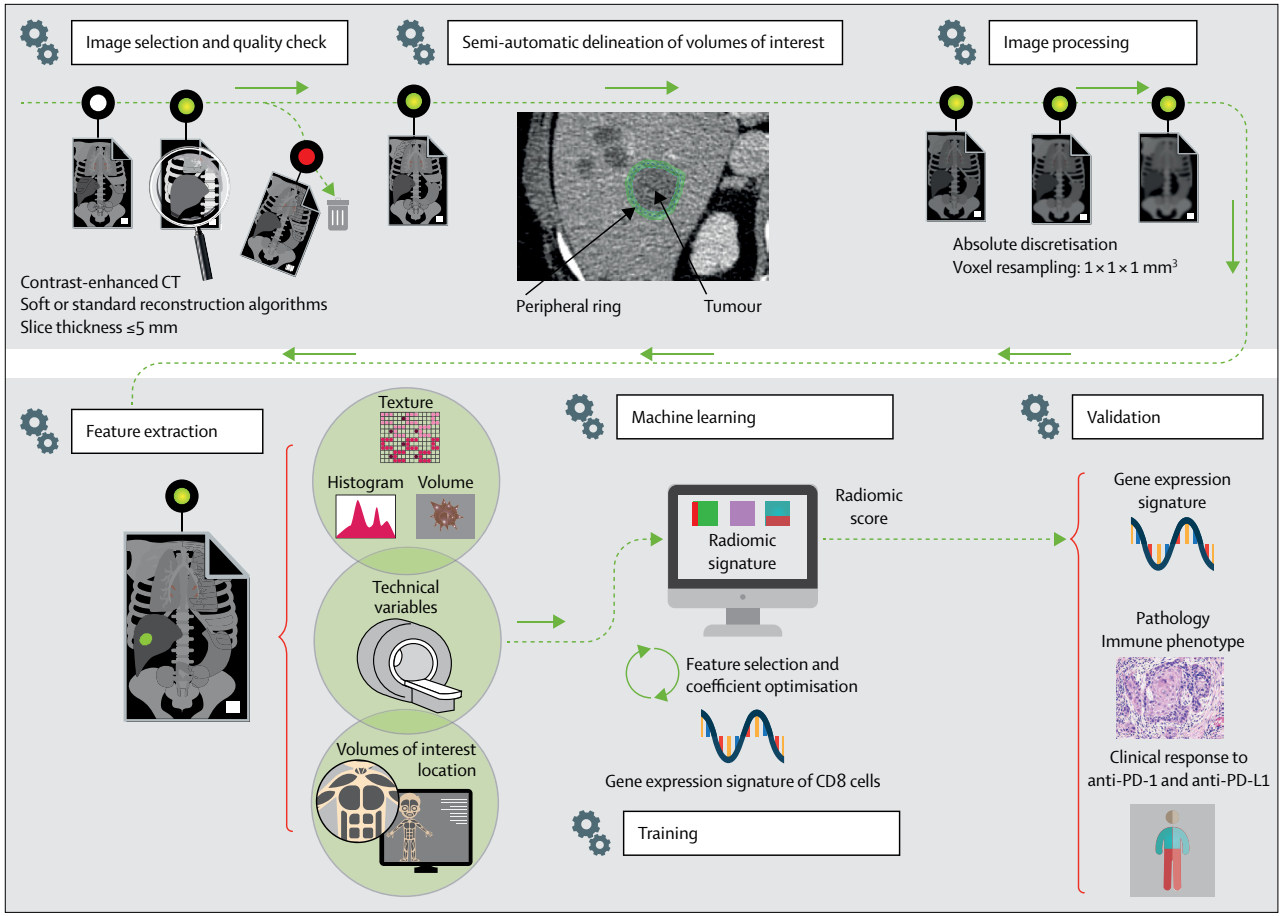
- 100,000 patches of histological slides
- Stroma
- 94% classification accuracy on test data set



- « Deep stroma score » is a predictive factor of survival independent of TNM stage (current state of the art)

Kather et al., Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study, PLoS Med, 2019

Prediction of response to immune-checkpoint inhibition



Sun et al., Lancet Oncol, 2018

Prediction of response to immune-checkpoint inhibition

but…..

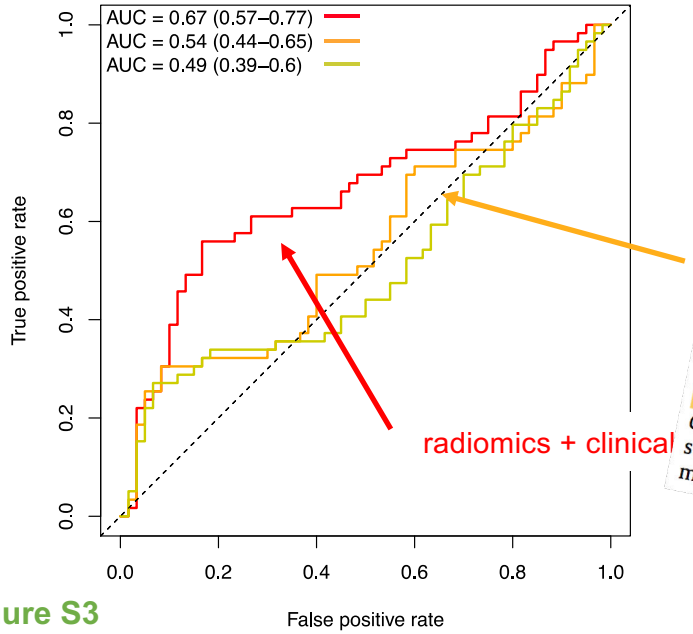


Figure S3

ELSEVIER
2019
Contents lists available at ScienceDirect
Radiotherapy and Oncology
journal homepage: www.thegreenjournal.com

Original article

Vulnerabilities of radiomic signature development: The need for safeguards

Mattea L. Welch [a,f,i], Chris McIntosh [e,f,i], Benjamin Haibe-Kains [a,c,i,j], Michael F. Milosevic [b,e,i], Leonard Wee [g], Andre Dekker [g], Shao Hui Huang [b,i], Thomas G. Purdie [b,e,f,i], Brian O'Sullivan [b,i], Hugo J.W.L. Aerts [h], David A. Jaffray [a,b,d,e,f,i,*]

[a] Department of Medical Biophysics, University of Toronto; [b] Department of Radiation Oncology, University of Toronto; [c] Ontario Institute of Cancer Research, Toronto; [d] IBBME, University of Toronto; [e] Radiation Medicine Program, Princess Margaret Cancer Centre, Toronto; [f] The Techna Institute for the Advancement of Technology for Health, Toronto, Canada; [g] Department of Radiation Oncology (MAASTRO), GROW Research Institute, Maastricht University, the Netherlands; [h] Dana-Farber Cancer Institute, Brigham and Women's Hospital,

A B S T R A C T

Purpose: Refinement of radiomic results and methodologies is required to ensure progression of the field. In this work, we establish a set of safeguards designed to improve and support current radiomic methodologies through detailed analysis of a radiomic signature.

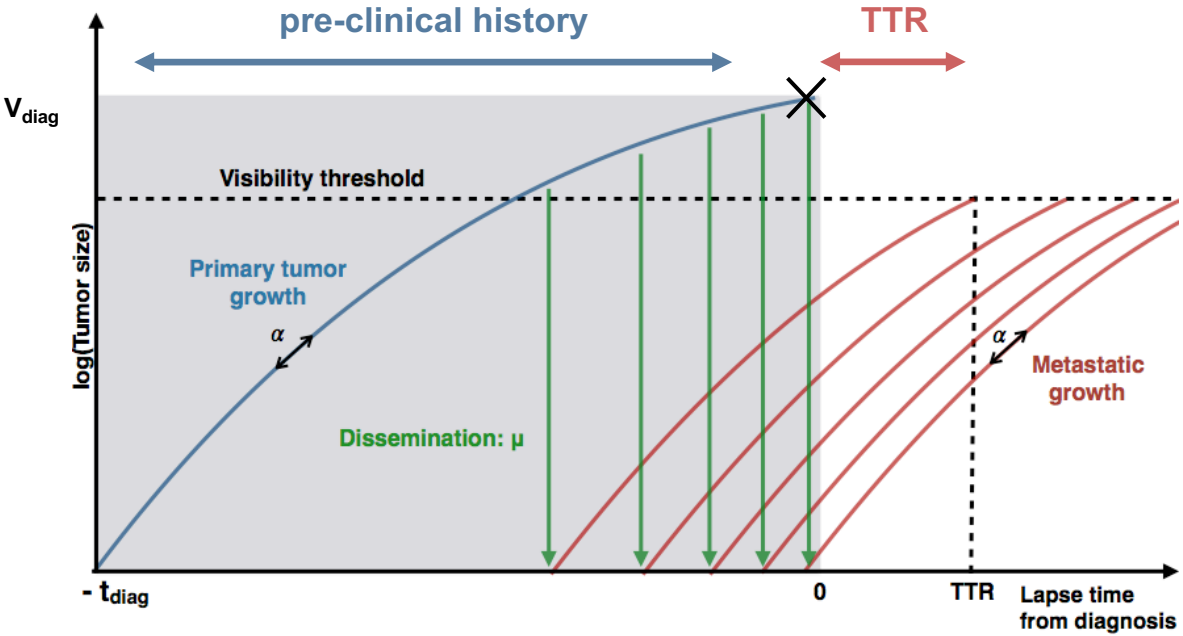
Methods: A radiomic model (MW2018) was fitted and externally validated using features extracted from previously reported lung and head and neck (H&N) cancer datasets using gross-tumour-volume contours, as well as from images with randomly permuted voxel index values; i.e. images without meaningful texture. To determine MW2018's added benefit, the prognostic accuracy of tumour volume alone was calculated as a baseline.

Results: MW2018 had an external validation concordance index (c-index) of 0.64. However, a similar performance was achieved using features extracted from images with randomized signal intensities (c-index = 0.64 and 0.60 for H&N and lung, respectively). Tumour volume had a c-index = 0.64 and correlated strongly with three of the four model features. It was determined that the signature was a surrogate for tumour volume and that intensity and texture values were not pertinent for prognostication.

Conclusion: Our experiments reveal vulnerabilities in radiomic signature development processes and suggest safeguards that can be used to refine methodologies, and ensure productive radiomic development using objective and independent features.

Sun et al.; Eu… 3

Mechanistic modeling of time to relapse



- Number of metastases with size larger than the visible size V_{vis}

$$N_{vis}(t) = \int_{V_{vis}}^{+\infty} \rho(t, v)dv$$
$$= \int_{0}^{t-\tau_{vis}} d(V_p(t))dt$$

τ_{vis} = time to reach V_{vis}

- Time to relapse (TTR) = time elapsed from diagnosis to the appearance of a first visible metastasis

$$TTR = \inf\{t > 0 : N_{vis}(t_{diag} + t) \geq 1\}$$

- Parameter β fixed such that $V_\infty = e^{\frac{\alpha}{\beta}} = 10^{12}$ cells

Mixed-effects statistical model

$$\ln\left(T^i\right) = \ln\left(TTR\left(V^i_{diag};\alpha^i,\mu^i\right)\right) + \varepsilon^i, \quad \varepsilon^i \sim \mathcal{N}(0,\sigma^2)$$ (Observation model)

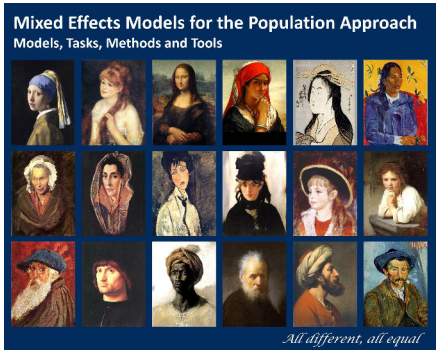
$$S\left(t|\alpha^i,\mu^i\right) = \mathbb{P}\left(T^i > t|\alpha^i,\mu^i\right)$$

Survival function to account for censoring in the likelihood

$$\ln\left(\alpha^i\right) = \ln\left(\alpha_{pop}\right) + \eta^i_\alpha, \quad \eta^i_\alpha \sim \mathcal{N}(0,\omega^2_\alpha)$$
$$\ln\left(\mu^i\right) = \ln\left(\mu_{pop}\right) + \eta^i_\mu, \quad \eta^i_\mu \sim \mathcal{N}(0,\omega^2_\mu)$$

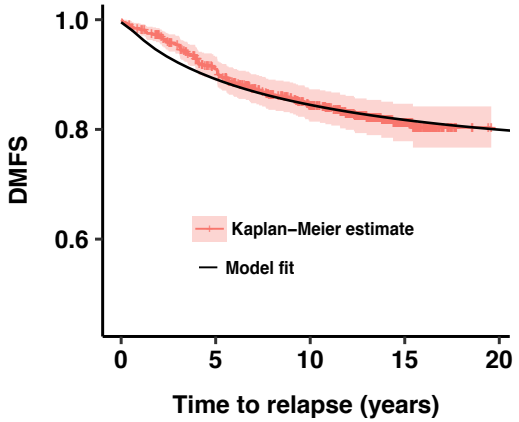
Likelihood maximization performed using the SAEM algorithm implemented in the *saemix* R package

Comets, Lavenu, Lavielle, J Stat Softw, 2017

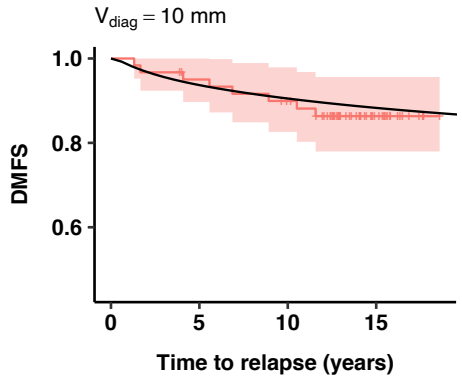


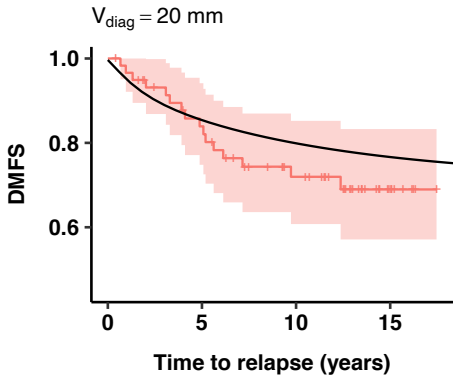
Lavielle, CRC press, 2014

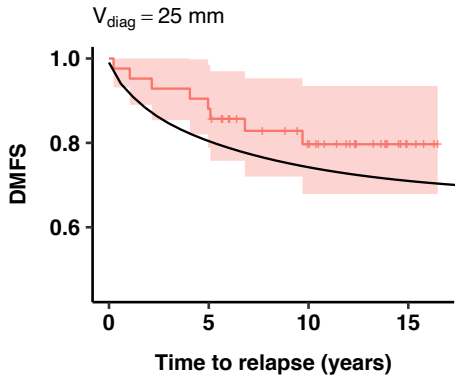
Descriptive power: fit to the data



Parameter	Estimate	r.s.e. (%)
$\log \alpha_{pop}$	-6.34	12.6
$\log \mu_{pop}$	-26.8	3.68
σ	0.542	28.4
ω_α	3.37	36.4
ω_μ	3.78	15.9







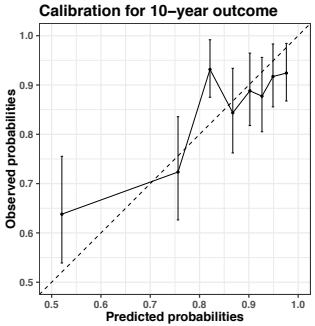
PhD of Chiara Nicolò

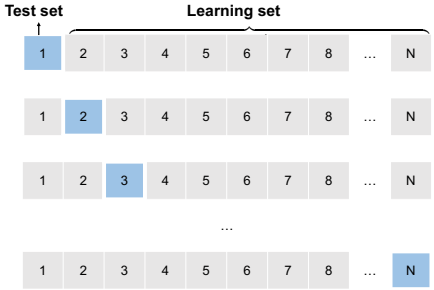
Predictive power: covariates

$$\ln\left(\mu^i\right) = \ln\left(\mu_{pop}\right) + \beta_\mu^T \mathbf{x}_\mu^i + \eta_\mu^i, \quad \eta_\mu^i \sim \mathcal{N}(0, \omega_\mu^2)$$

$$\ln\left(\alpha^i\right) = \ln\left(\alpha_{pop}\right) + \beta_\alpha^T \mathbf{x}_\alpha^i + \eta_\alpha^i, \quad \eta_\alpha^i \sim \mathcal{N}(0, \omega_\alpha^2)$$

Parameter	Estimate	r.s.e. (%)	p-value
$\log \alpha_{pop}$	-8.883	10.151	
$\beta_{\text{Ki67},\alpha}$	0.086	27.376	$2.59 \cdot 10^{-4}$
$\beta_{\text{HER2},\alpha}$	0.029	42.833	0.020
$\beta_{\text{CD44},\alpha}$	0.011	60.816	0.1
$\beta_{\text{TRIO},\alpha}$	0.016	58.119	0.085
$\log \mu_{pop}$	-26.342	3.696	
$\beta_{\text{EGFR},\mu}$	0.039	47.527	0.035
σ	0.606	23.104	
ω_α	2.062	22.715	
ω_μ	3.563	16.759	





c-index = 0.67
(10-folds cross-validation)

Patient ID	Tumor size (mm)	Ki67	HER2	CD44	TRIO	EGFR	Observed TTR (cens)	Predicted TTR	Prediction error (days)
255	25	1	60	90	60	0	1812 (1)	1609	203
47	20	32	100	0	0	50	739 (1)	447	292
143	18	60	0	50	0	0	2798 (1)	434	2364
12	10	20	0	23	0	0	5970 (0)	$+\infty$	-

PhD of Chiara Nicolò

Comparison of predictive metrics

5 years metastatic-free survival

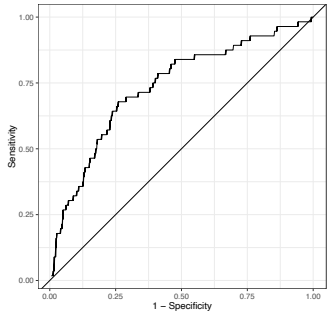
	AUROC	Accuracy	PPV	NPV
RSF	0.75	0.90	0.71	0.71
Mechanistic model	0.73	0.90	0.72	0.70
Cox	0.75	0.91	0.77	0.71

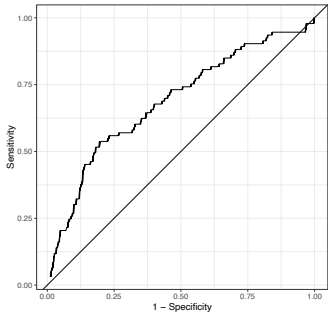
10 years metastatic-free survival

	AUROC	Accuracy	PPV	NPV
RSF	0.69	0.82	0.68	0.66
Mechanistic model	0.69	0.81	0.71	0.64
Cox	0.71	0.82	0.70	0.68

other tested ML models (support vector machine, k-nearest neighbors, gradient boosting) had similar or worse performances

Mechanistic





RSF