To share your e-Eso experience use:
#e_ESO

Artificial intelligence for prostate cancer radiotherapy

Expert: **Prof Jean Emmanuel Bibault**, Hôpital Européen Georges Pompidou and & INSERM UMR1138, Paris, France

Discussant: **Prof Pierfrancesco Franco**, University of Eastern Piedmont, Novara, Italy

Artificial Intelligence for prostate cancer radiotherapy

Pr Jean-Emmanuel Bibault
Radiation Oncology Department, Hôpital Européen Georges Pompidou
INSERM UMR 1138, Centre de Recherche des Cordeliers
Paris, France

@jebibault

ASSISTANCE PUBLIQUE HÔPITAUX DE PARIS
HÔPITAUX UNIVERSITAIRES PARIS OUEST
Hôpital européen Georges-Pompidou
Université de Paris
Inserm
CARPEM CANCER RESEARCH FOR PERSONALIZED MEDICINE

Historique

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

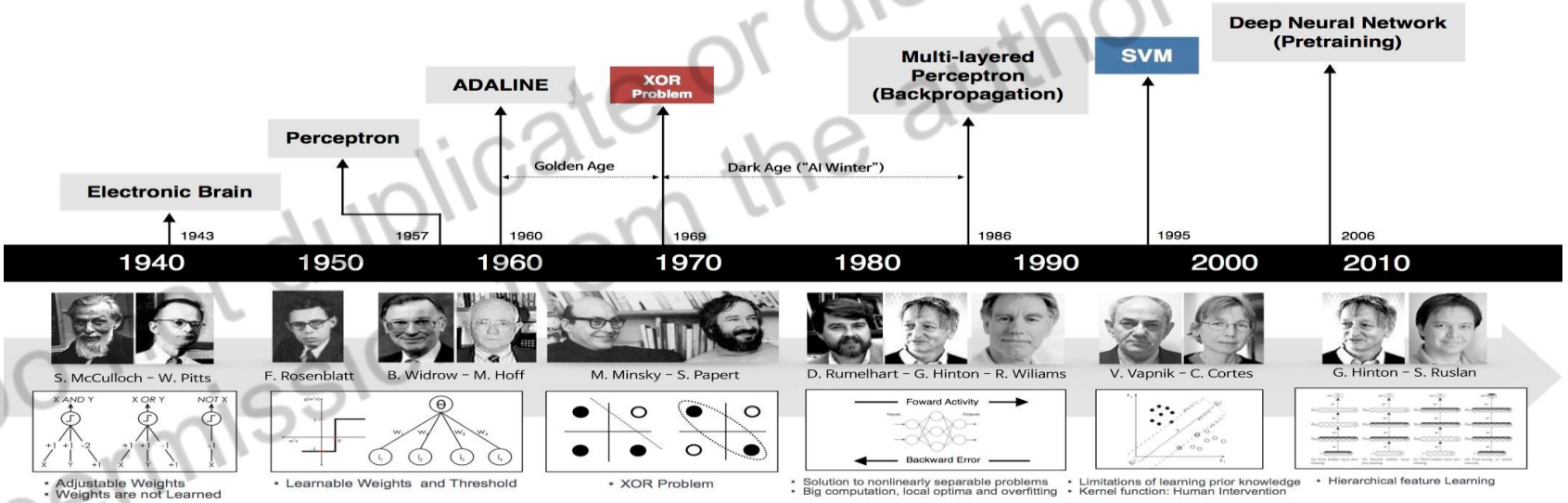
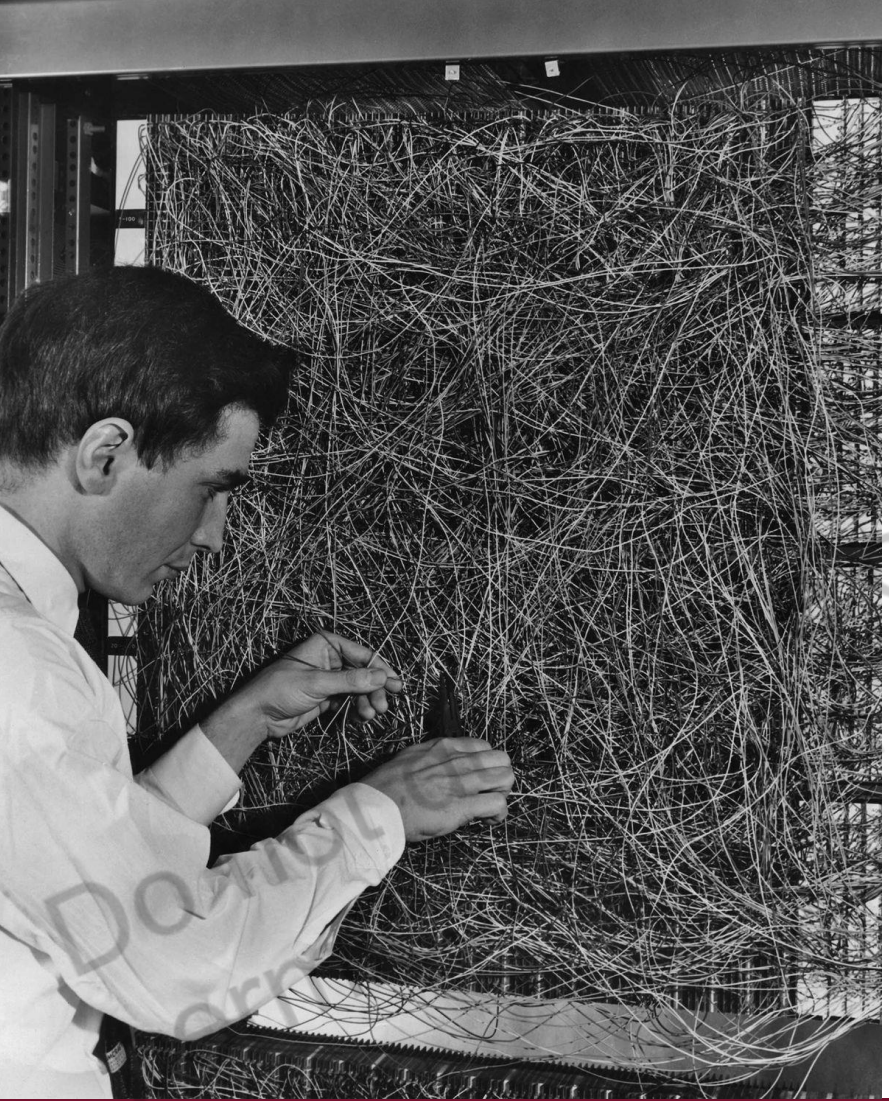


image : Andrew Beam



TensorFlow™

scikit learn

PYTORCH

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

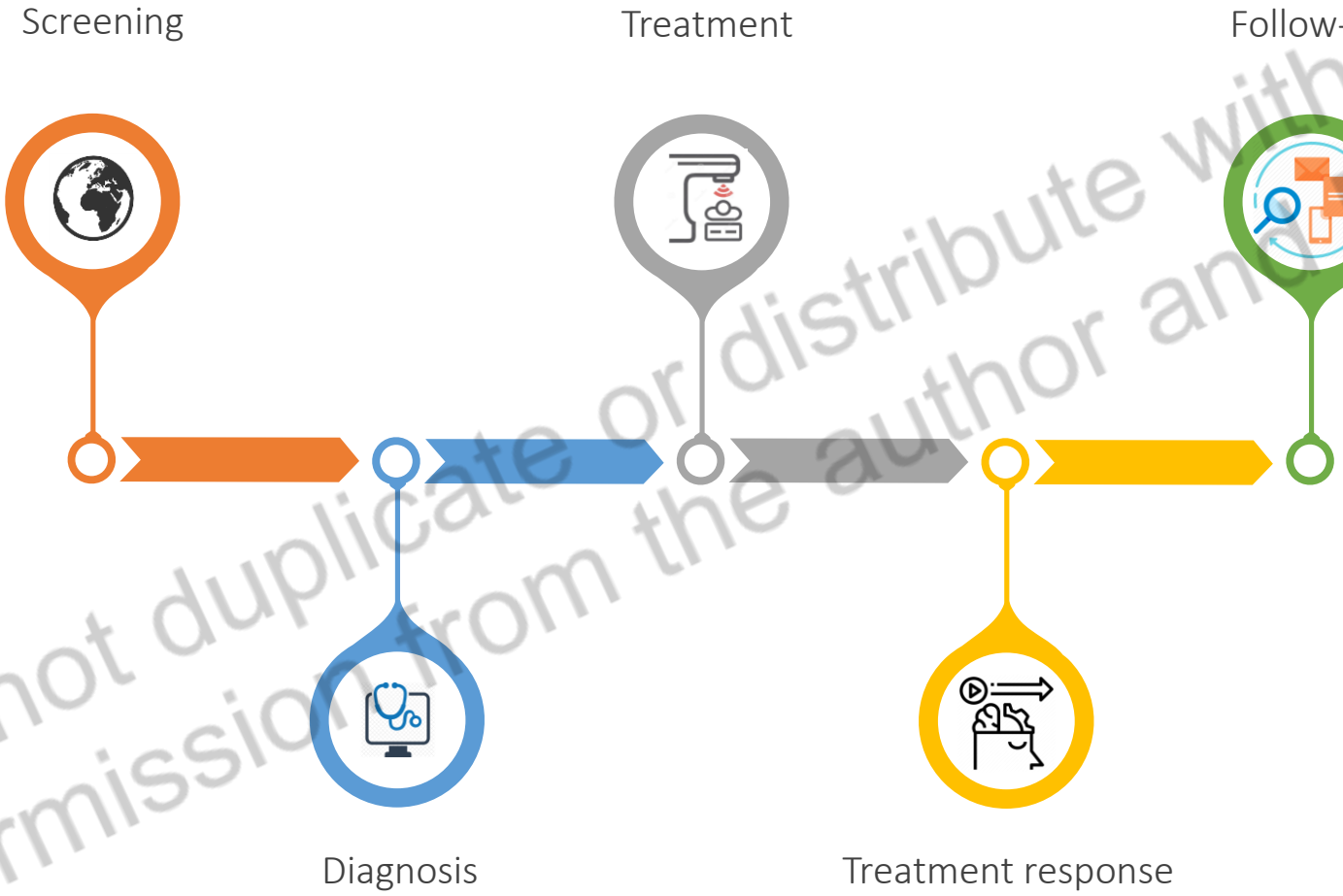
FAST COMPANY

The Robot Will See You Now

the Atlantic

The Avatar Will See You Now

MIT Technology Review



Diagnosis & prediction

In the United States alone, each year:

180,890 new cases of prostate cancer (PCa) will be diagnosed

Diagnosis steps:

- PSA testing
- Prostate biopsy
- Staging: CT, skeletal scintigraphy or PET-CT

Risk stratification

Depends on:

- PSA
- Gleason score (pathology) +/- Genomics (Decipher)
- T (tumor)
- N (lymph nodes)
- M (metastasis)

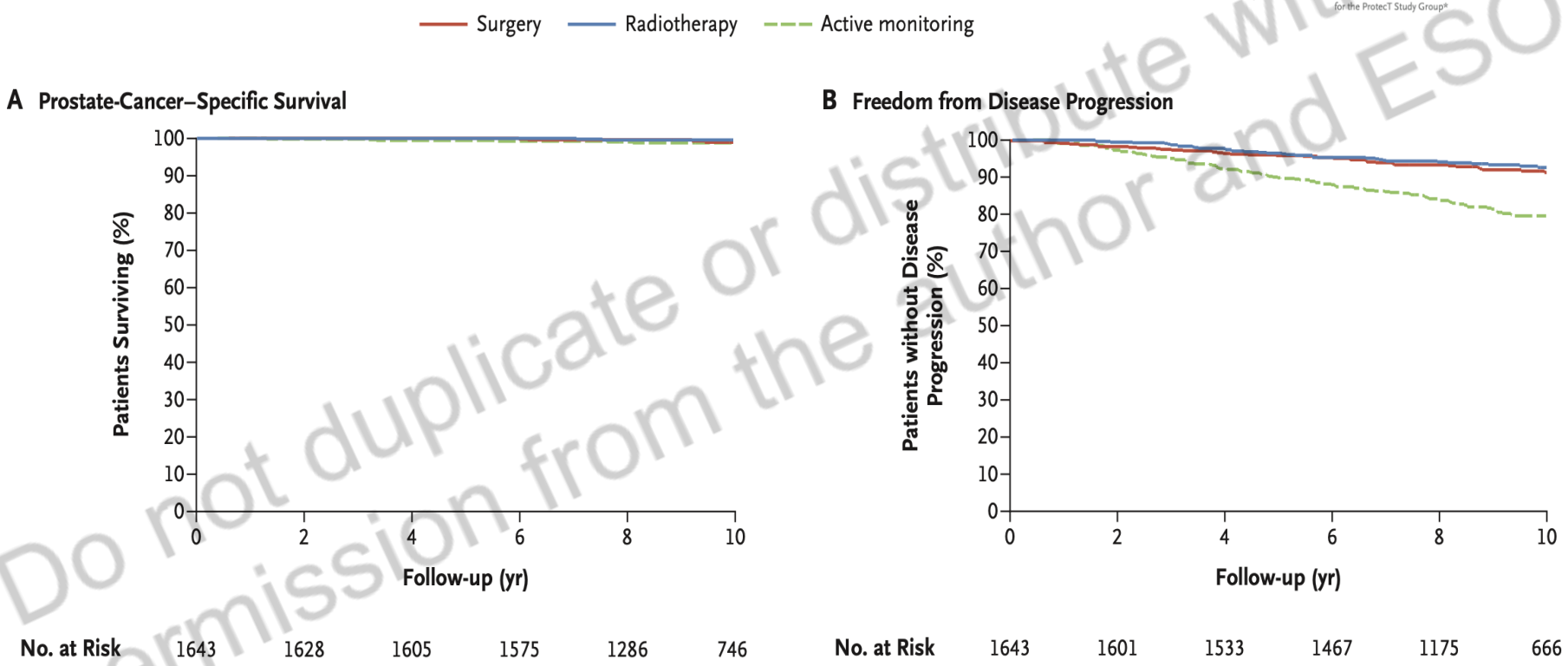
Localized	Locally-advanced	Metastatic	Castration-resistant
- Low risk - Intermediate risk - High risk	Curable ?	Incurable	Incurable

The NEW ENGLAND JOURNAL of MEDICINE
ESTABLISHED IN 1812 OCTOBER 13, 2016 VOL. 375 NO. 15

Does treatment save lives?

10-Year Outcomes after Monitoring, Surgery, or Radiotherapy for Localized Prostate Cancer

F.C. Hamdy, J.L. Donovan, J.A. Lane, M. Mason, C. Metcalfe, P. Holding, M. Davis, T.J. Peters, E.L. Turner, R.M. Martin, J. Oxley, M. Robinson, J. Staffurth, E. Walsh, P. Bollina, J. Catto, A. Doble, A. Doherty, D. Gillatt, R. Kockelbergh, H. Kynaston, A. Paul, P. Powell, S. Prescott, D.J. Rosario, E. Rowe, and D.E. Neal, for the ProtecT Study Group*



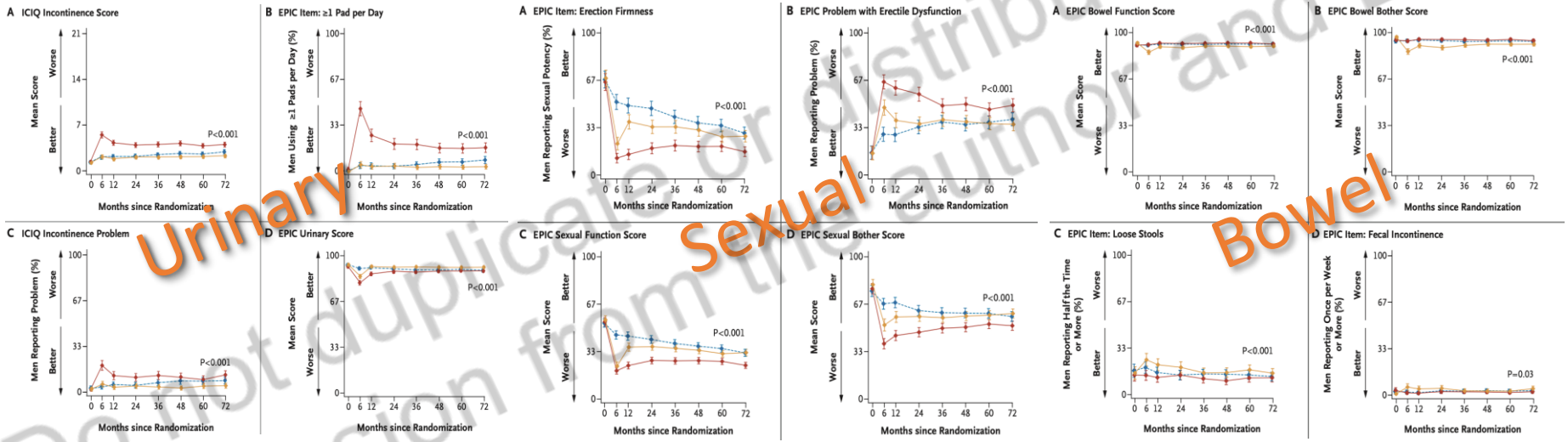
… but treatments are (very) toxic

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Patient-Reported Outcomes after Monitoring, Surgery, or Radiotherapy for Prostate Cancer

J.L. Donovan, F.C. Hamdy, J.A. Lane, M. Mason, C. Metcalfe, E. Walsh, J.M. Blazeby, T.J. Peters, P. Holding, S. Bonnington, T. Lennon, L. Bradshaw, D. Cooper, P. Herbert, J. Howson, A. Jones, N. Lyons, E. Salter, P. Thompson, S. Tidball, J. Blaikie, C. Gray, P. Bollina, J. Catto, A. Doble, A. Doherty, D. Gillatt, R. Kockelbergh, H. Kynaston, A. Paul, P. Powell, S. Prescott, D.J. Rosario, E. Rowe, M. Davis, E.L. Turner, R.M. Martin, and D.E. Neal, for the ProtecT Study Group*



Urinary

Sexual

Bowel

Donovan JL et al., NEJM, 2016

Decision-making in PCa

26,120 men will die from the disease

35,000 men are being overdiagnosed and go through unnecessary treatments, causing complications

How can we determine who will benefit from treatment?

Nomograms

Nomograms exist to predict progression-free survival and cancer-specific survival

Rely on data from one center (→ Not generalizable)

Use Regression models

Do not take into account comorbidities

Primary Treatment Outcomes

PROBABILITY OF CANCER-SPECIFIC SURVIVAL AFTER RADICAL PROSTATECTOMY

10 YR 95 % 15 YR 88 %

10 YEAR | 15 YEAR

95 %

This number shows, as a percentage, your probability of surviving prostate cancer for 10 years following radical prostatectomy. This probability means that for every 100 patients like you, 95 will survive prostate cancer and 5 will have died from prostate cancer.

This prediction addresses survival related specifically to prostate cancer; it does not exclude the possibility of death from other causes, such as heart disease or accident, within this time period.

Methods

PLCO trial

Prospective randomized multicenter trial:

- 76,693 men at 10 U.S. study centers
- Randomly assigned to receive:
  - annual screening (n=38,343)
  - usual care as the control (n=38,350)

Data transfer agreement with the National Cancer Institute (NCI)

Data was downloaded from the NCI Cancer Data Access System

Andriole GL et al, NEJM, 2009
Andriole GL et al, JNCI, 2012
Pinksky PF et al, Cancer 2017

Dataset

Comprehensive dataset:

Contains nearly all the PLCO study data available for prostate cancer screening, incidence, and mortality analyses

One record for each of the participants in the PLCO trial:

- Baseline features
- Screening
- Diagnosis
- Treatment procedures

Population: patients that were diagnosed with prostate cancer during follow-up, irrespective of the arm they were originally included into

Features selection

Assess the predictive power of a simple set of questions as a baseline indicator:

Prostate cancer diagnosis: PSA, T, N, M stage, Gleason score and initial primary treatment (if performed)

Medical history: age, height, weight, current smoking status, smoking pack-years, daily alcohol consumption, history of prostatitis, nocturia, arthritis, bronchitis, diabetes, emphysema, heart attack, hypertension, liver disease, osteoporosis, stroke, cholesterol

Physical activity: activity at least once a month during the last year, physical activity at work

Socio-economic status: family income, education

Hormonal status: hair pattern at 45 y.o., weight gain pattern

Model training

Dataset split in training and testing datasets before any analysis was performed

Classification task

Two separate models:

- 10-year overall survival: patients who died from any cause within ten years of PCa diagnosis
- 10-year cancer-specific survival: patients who died from PCa within ten years of PCa diagnosis

Model training

XGBoost: state-of-the-art for tabular data

Missing values inherently handled by predictor

Hyperparameters selected on training dataset

- Nested, cross validation
- Bayesian Optimization

Class imbalance corrected with positive class weighting

Performance assessed on a test dataset using non-parametric bootstrap .632 procedure (200 splits) to obtain 95 CI

Chen T et al, ACM SIGKDD, 2016
Josse et al, arXiv, 2019
https://github.com/fmfn/BayesianOptimization
James G et al, Springer, 2013

Model interpretation

Need to know whether the prediction relies on the aggressivity of the PCa or on a comorbidity, or a combination of comorbidities

Shapley values: unified approach to interpreting tree models

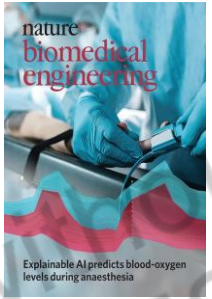
Reflect the importance of every feature for the prediction

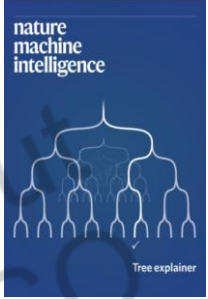
At the population or individual scale

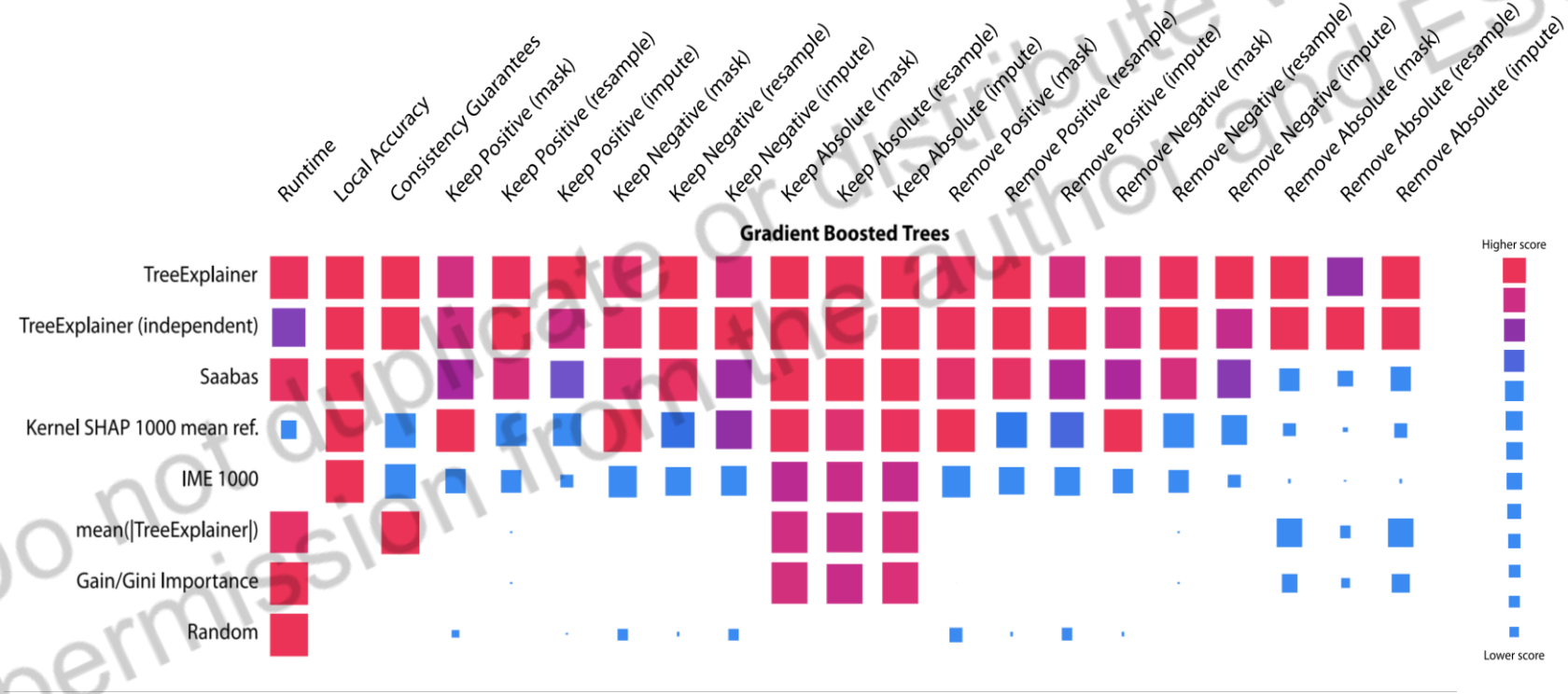
Lundberg S et al, arXiv, 2017
Lundberg S et al, Nat. Bio. Eng, 2018
https://github.com/slundberg/shap

Shapley values

Lundberg S et al, Nat. BE, 2018
Lundberg S et al, Nat. MJ, 2019
https://github.com/slundberg/shap







Model deployment

Deploy the CSS and OS models online

Provide prediction and individual interpretability

Dash framework

GitHub repository for hosting

Heroku for serving

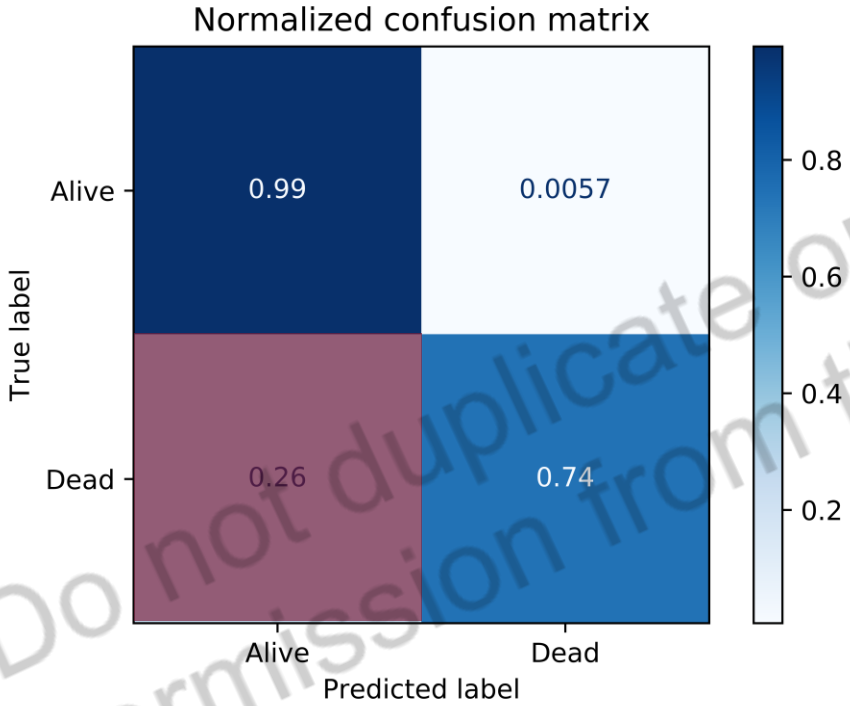
Results – Population (n=8,776)

Characteristic	No. (%)
Localized PCa	7,668 (87.4)
Low-risk	2,940 (33.5)
Intermediate-risk	3,476 (39.6)
High risk	1,252 (14.3)
Locally advanced PCa	913 (10.4)
Metastatic PCa	195 (2.2)
Age	
Under 65 years old	1990 (22.7)
Between 65 and 75 years old	5181 (59)
Over 75 years old	1605 (18.3)
Death from any cause	3,128 (35.6)
Death from prostate cancer	546 (6.2)

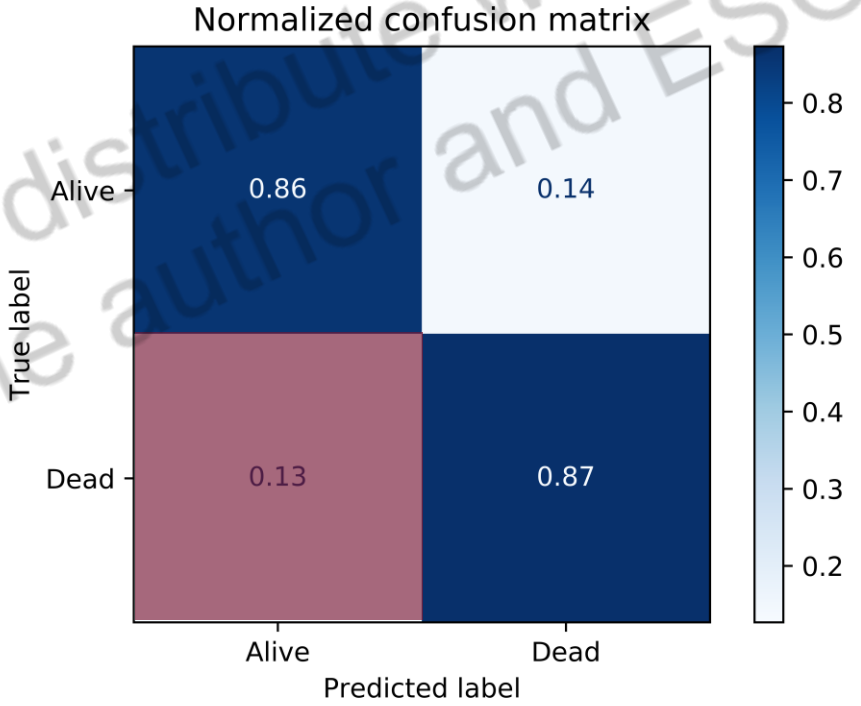
Results - Models' performances

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)

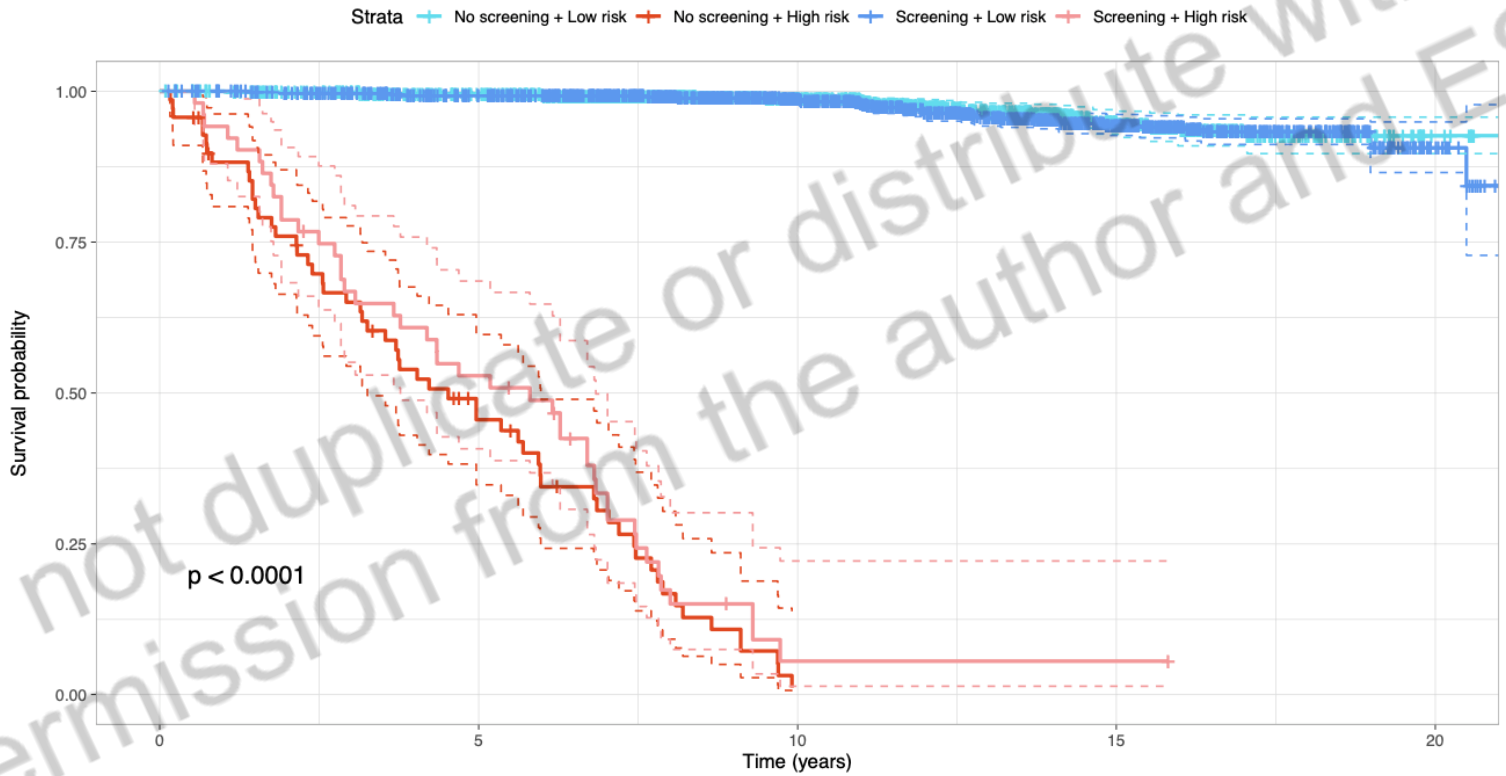
CSS



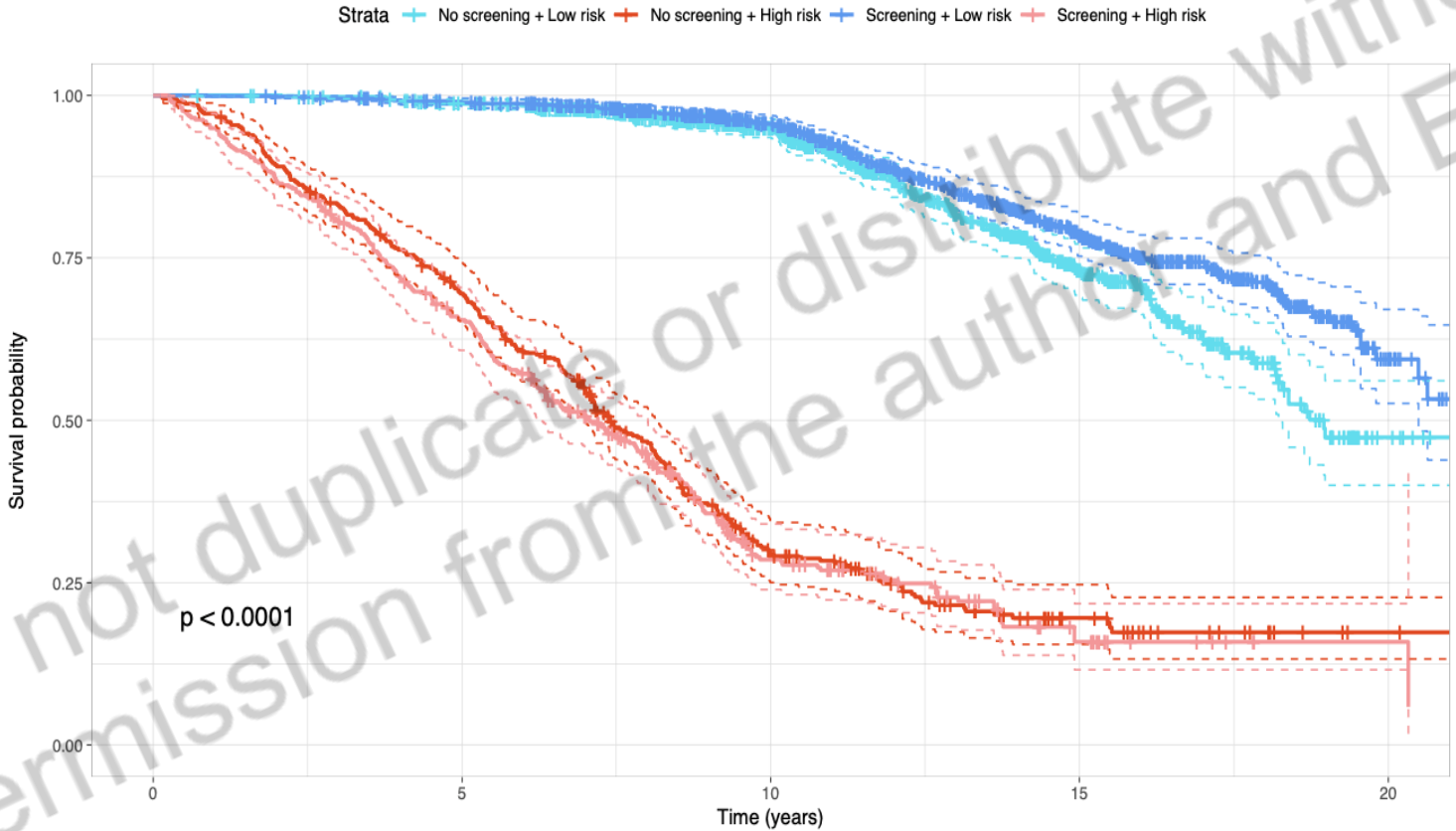
OS



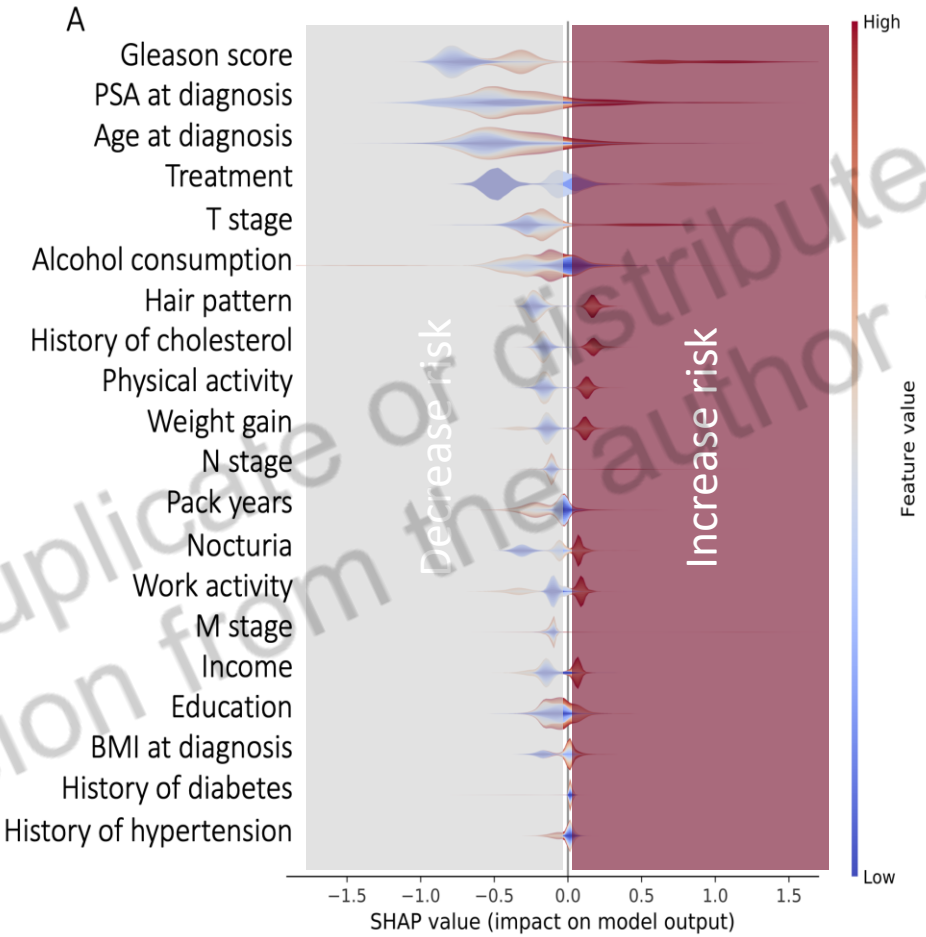
AI vs screening: CSS model



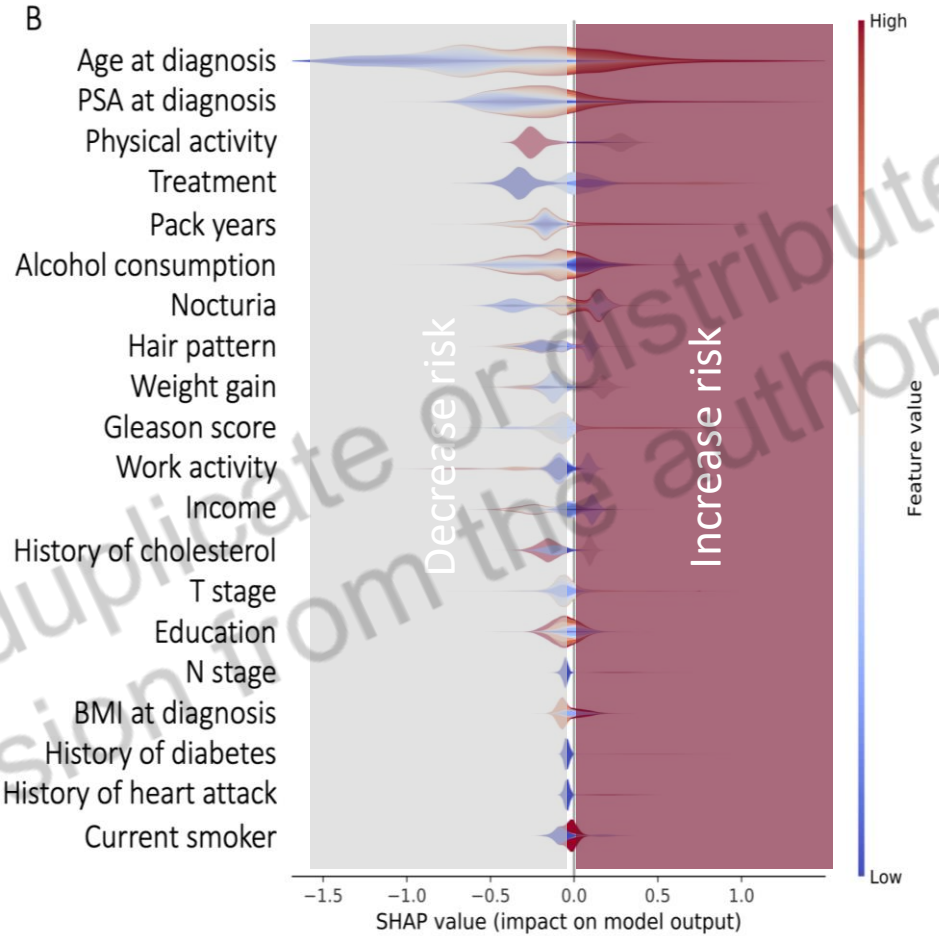
AI vs screening: OS model



Most important features: CSS model



Most important features: OS model



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.

About | Predict cancer-specific survival | Predict overall survival

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

Checking consistency with human intuition

At the individual scale: Virtual patient 1

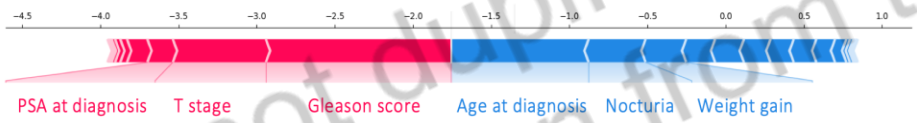
High-risk PCa:

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

Without significant comorbidities:

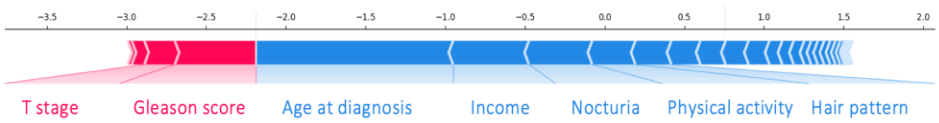
- 55 y.o.
- no smoking
- no alcohol consumption
- regular physical activity

A



Probability of dying from PCa: 18.92%

B



Probability of dying from any cause: 19.57%

At the individual scale: Virtual patient 2

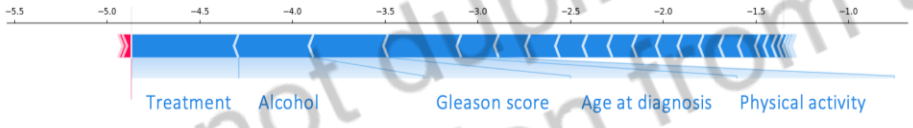
Intermediate risk PCa:

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

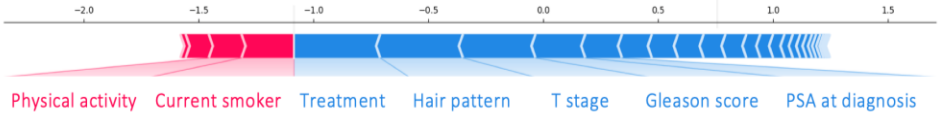
Several comorbidities:

- 70 y.o.
- Smoker (50 pack-years)
- 2 drinks of alcohol per day
- No physical activity

C



D



Probability of dying from PCa: 0.76%

Probability of dying from any cause: 25.2%

Discussion

Models to answer a relevant clinical issue:

- Which patients could benefit from treatment?
- Is a patient at risk of dying from PCa or another cause?
- Is so, why?

Accurate results

First model using machine learning, trained on a large population from 10 different centers

→ generalizability

Cooperberg MR et al, JNCI 2009, Cancer 2011
Goldenberg SL et al, Nat Rev Urol, 2019
Chin J et al, European Urology, 2020

Limits

Trained on data from a prospective trial that was not specifically designed for this

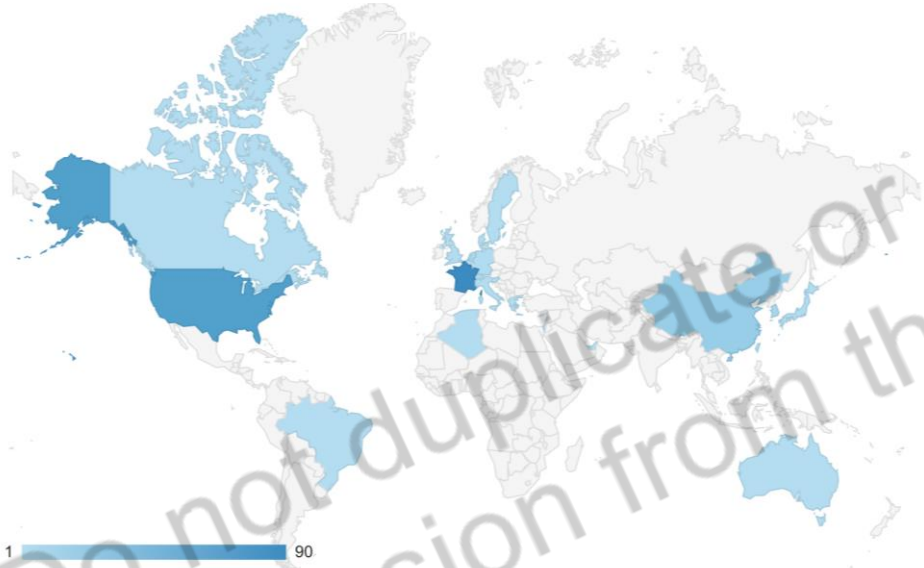
→ Possible biases

Only 195 patients (2.2%) were metastatic in the dataset

→ Caution when using the models for this population

Questionnaire response bias (patient-dependent)

Online use

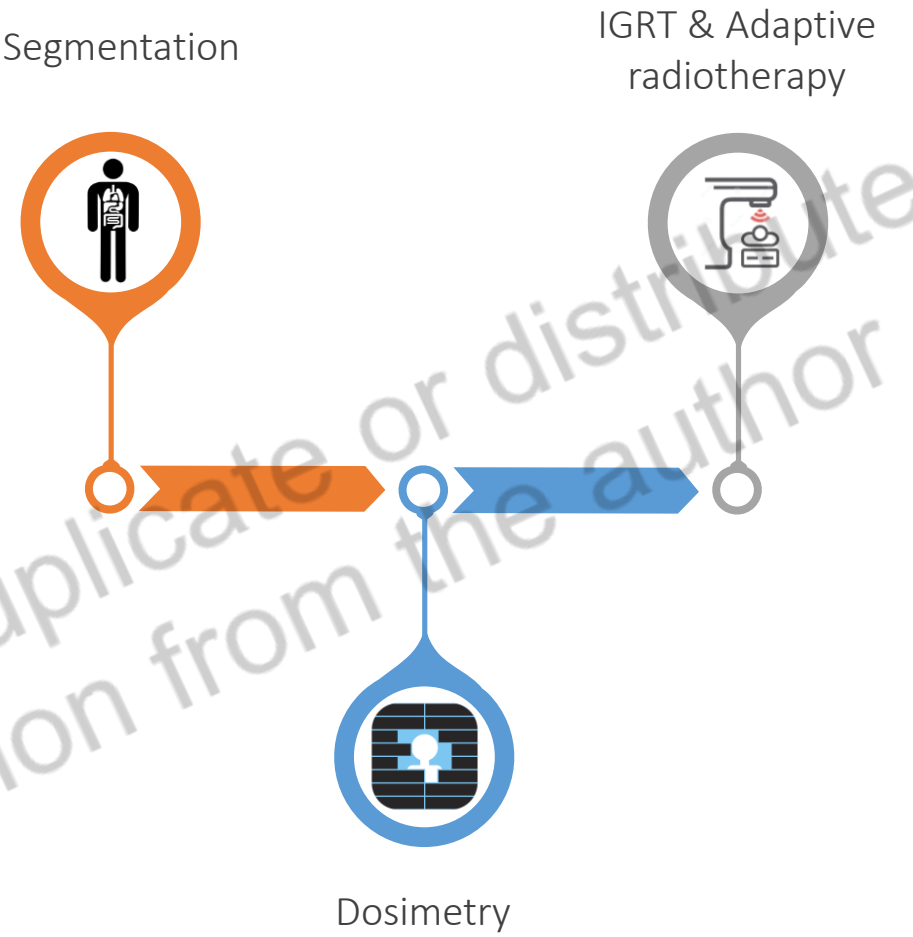


Country	Acquisition		
	Users ↓	New Users	Sessions
	208 % of Total: 100.00% (208)	208 % of Total: 100.00% (208)	228 % of Total: 100.00% (228)
1. France	**90** (43.27%)	90 (43.27%)	93 (40.79%)
2. United States	**67** (32.21%)	67 (32.21%)	72 (31.58%)
3. China	**18** (8.65%)	18 (8.65%)	26 (11.40%)
4. South Korea	**7** (3.37%)	7 (3.37%)	7 (3.07%)
5. Switzerland	**3** (1.44%)	3 (1.44%)	3 (1.32%)
6. Netherlands	**3** (1.44%)	3 (1.44%)	3 (1.32%)
7. Germany	**2** (0.96%)	2 (0.96%)	2 (0.88%)
8. Israel	**2** (0.96%)	2 (0.96%)	2 (0.88%)
9. Japan	**2** (0.96%)	2 (0.96%)	6 (2.63%)
10. United Arab Emirates	**1** (0.48%)	1 (0.48%)	1 (0.44%)

https://prostatecancersurvival.herokuapp.com

Treatment

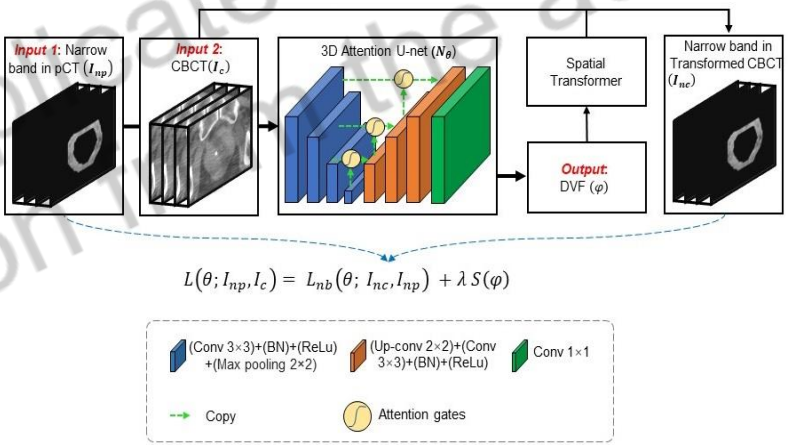
AI for radiotherapy planning



Deep Learning for IGRT/adaptive radiotherapy

DUL (Deep Unsupervised Learning) : U-Net

Automatically propagate prostate segmentation from treatment planning CT scan to daily CBCT



Liang et al, Medical Physics, 2021

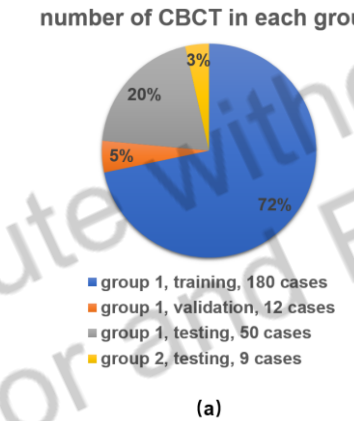
Data

Group 1:

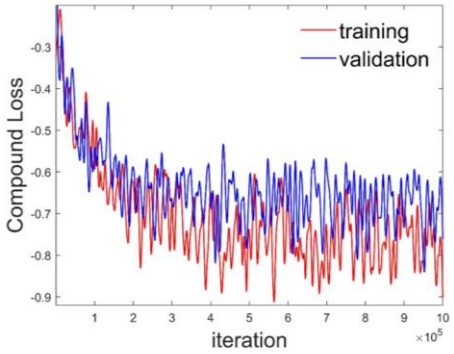
- 180 for training
- 12 for validation
- 50 for testing

Group 2:

- External validation
- 9 CBCT
- 4 different human segmentations
- Consensual segmentation wih STAPLE
- Compared with DICE



(a)

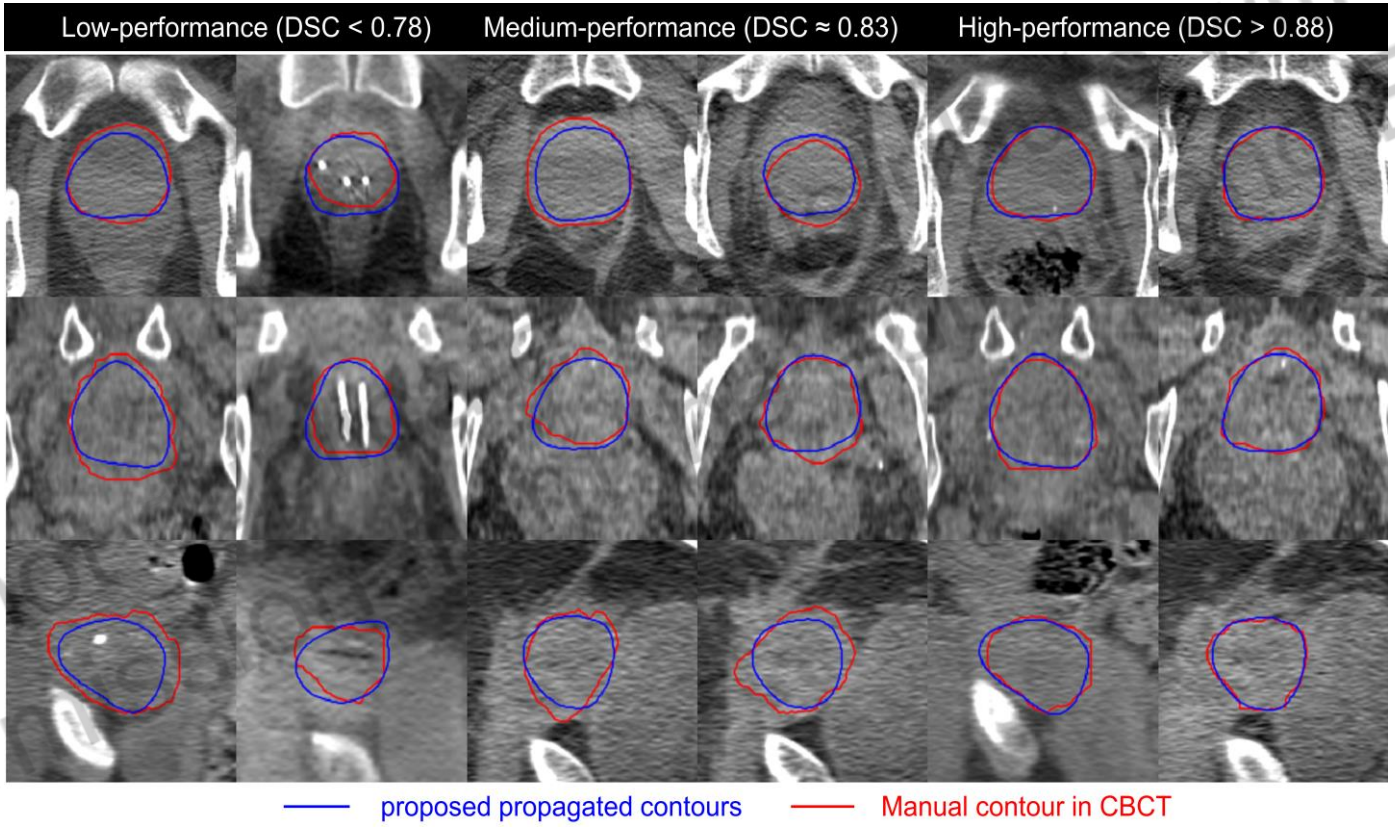


(b)

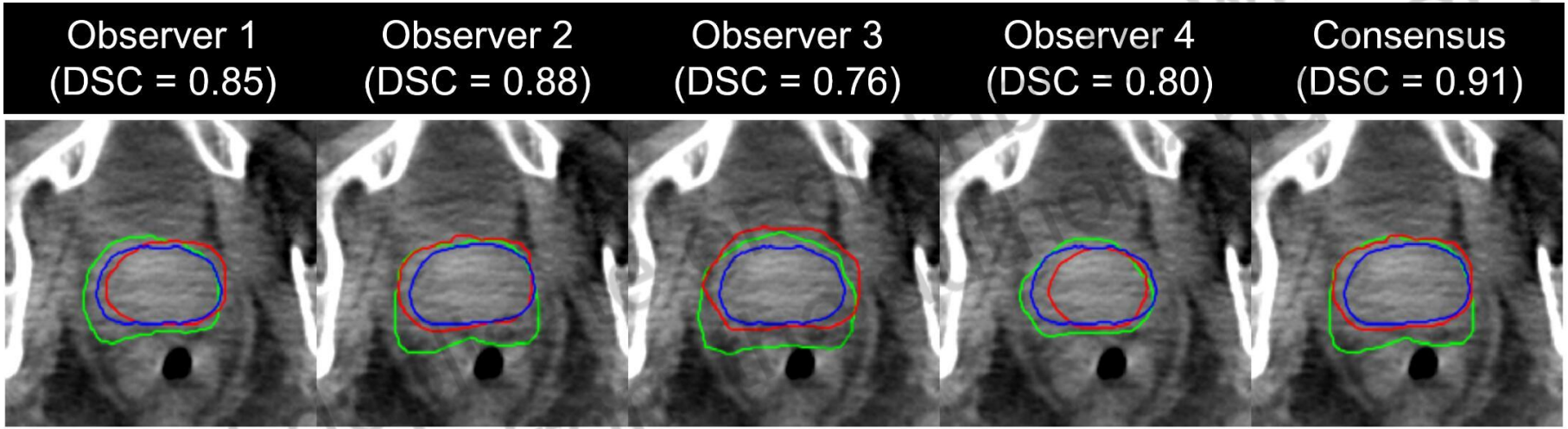
Results

Contour Metric (mean ± SD)	**Group 1 testing**	Group 2				
		Observer 1	Observer 2	Observer 3	Observer 4	Consensus
Dice	0.83 ± 0.04	0.82 ± 0.06	0.83 ± 0.03	0.82 ± 0.04	0.82 ± 0.08	0.85 ± 0.04
Sensitivity	0.87 ± 0.10	0.79 ± 0.11	0.83 ± 0.10	0.79 ± 0.09	0.77 ± 0.12	0.80 ± 0.09
Hausdorff distance	9.10 ± 2.60	7.58 ± 1.98	7.99 ± 2.14	8.21 ± 1.98	10.01 ± 2.12	7.08 ± 2.03
MCC	0.85 ± 0.03	0.84 ± 0.05	0.83 ± 0.03	0.85 ± 0.09	0.83 ± 0.14	0.86 ± 0.05
Contour distance (mm)						
COM (mm)	3.52 ± 1.15	3.04 ± 1.74	3.08 ± 1.57	3.12 ± 1.74	3.50 ± 1.91	2.98 ± 1.42
Superior	1.91 ± 1.29	2.01 ± 1.21	2.15 ± 1.57	2.00 ± 1.29	1.91±2.01	1.65±1.21
Inferior	4.72 ± 2.12	5.45 ± 3.25	3.84 ± 2.75	6.14 ± 2.37	5.81±3.05	4.01±2.09
Anterior	2.01 ± 1.87	1.90 ± 1.10	1.85 ± 1.00	2.27 ± 0.79	2.64±1.07	1.98±0.96
Posterior	1.46 ± 1.17	1.59 ± 0.98	1.89 ± 1.09	2.01 ± 1.85	2.46±1.56	1.40±1.17
Left	1.08 ± 1.17	1.95 ± 0.690	2.08 ± 1.00	1.18 ± 1.28	2.17±1.37	1.01±1.02
Right	1.31 ± 1.04	1.57 ± 0.89	2.11 ± 1.16	1.67 ± 1.14	1.76±1.77	1.28±1.00

Results



Results

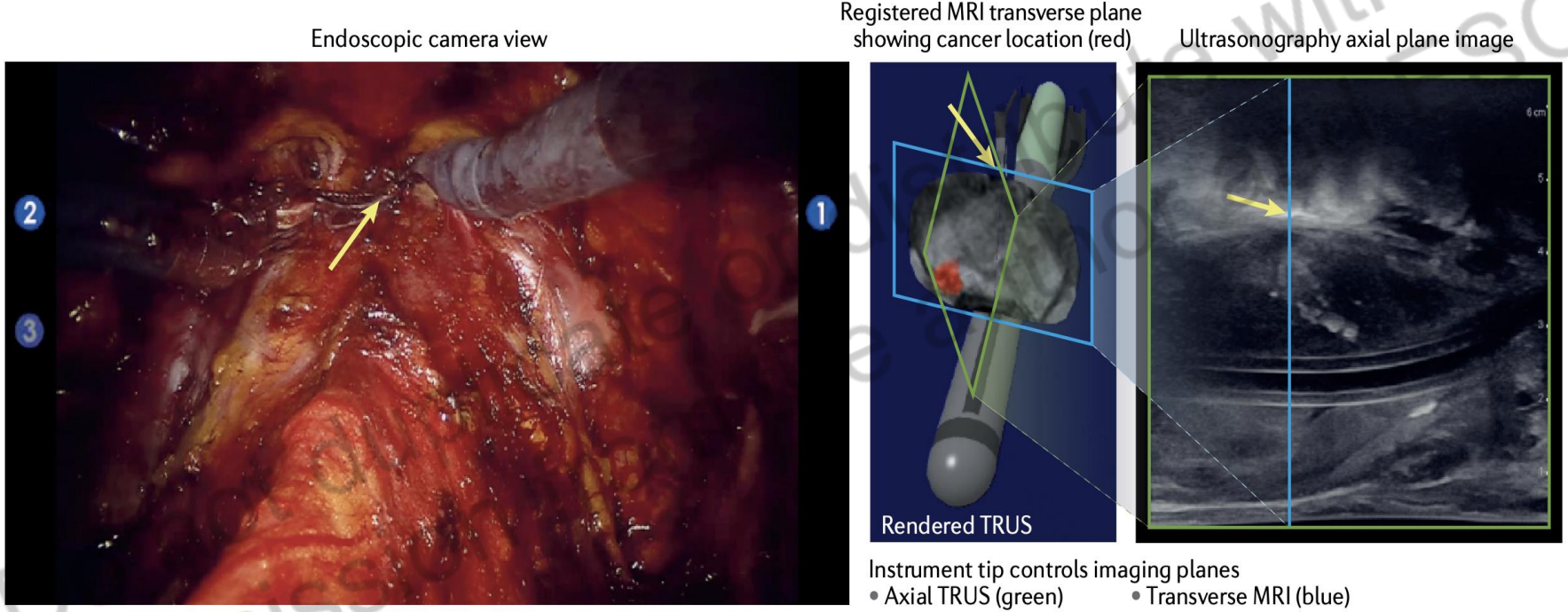


Fast and reliable method

Could be used to better visualize prostate on CBCT for daily IGRT

Perspectives

Surgery?



SL Goldenberg, Nature Review Urology, 2019

Do not forget the limits of these methods

Not a magical wand

Need validation

Should we perform randomized trials?

What are the quality criteria?

→ESTRO-ACROP guidelines

Acknowledgments

Université de Paris
Stanford MEDICINE | Laboratory of Artificial Intelligence in Medicine and Biomedical Physics | Department of Radiation Oncology
Inserm
NIH NATIONAL CANCER INSTITUTE Cancer Data Access System
ASSISTANCE PUBLIQUE HÔPITAUX DE PARIS

- Xing Lab (Stanford, USA)
- Burgun Lab (INSERM, France)
- Stanford Radiation Oncology department
- HEGP Radiation Oncology Department
- National Cancer Institute

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