Artificial intelligence (AI) in personalized oncology

Dr Julien Vibert, MD, PhD
Drug Development Department (DITEP) - Sarcoma Team
INSERM U981 - Computational Oncology
PRISM - National PRecISion Medicine Center in Oncology
Gustave Roussy

IFSBM Module 11

16 Dec 2025

Outline of presentation

http://www.unehistoiredeplumes.fr/a-vol-doiseau/

- Simple intro to AI
- AI in medicine
- AI in oncology
- AI in cancer genomics and transcriptomics
- *Use case*: AI-based classification of cancers of unknown primary
- Conclusions and perspectives

Outline of presentation

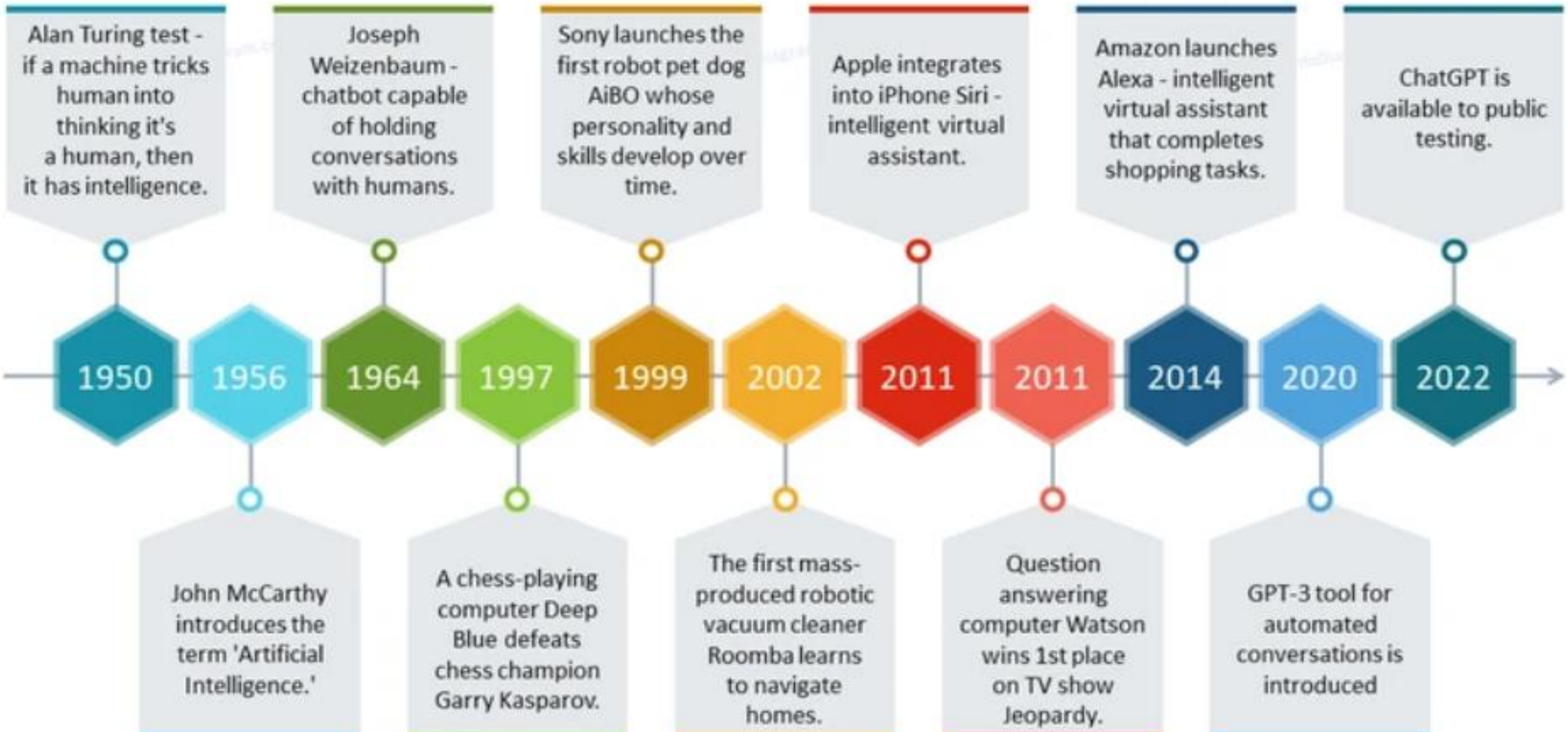
http://www.unehistoiredeplumes.fr/a-vol-doiseau/

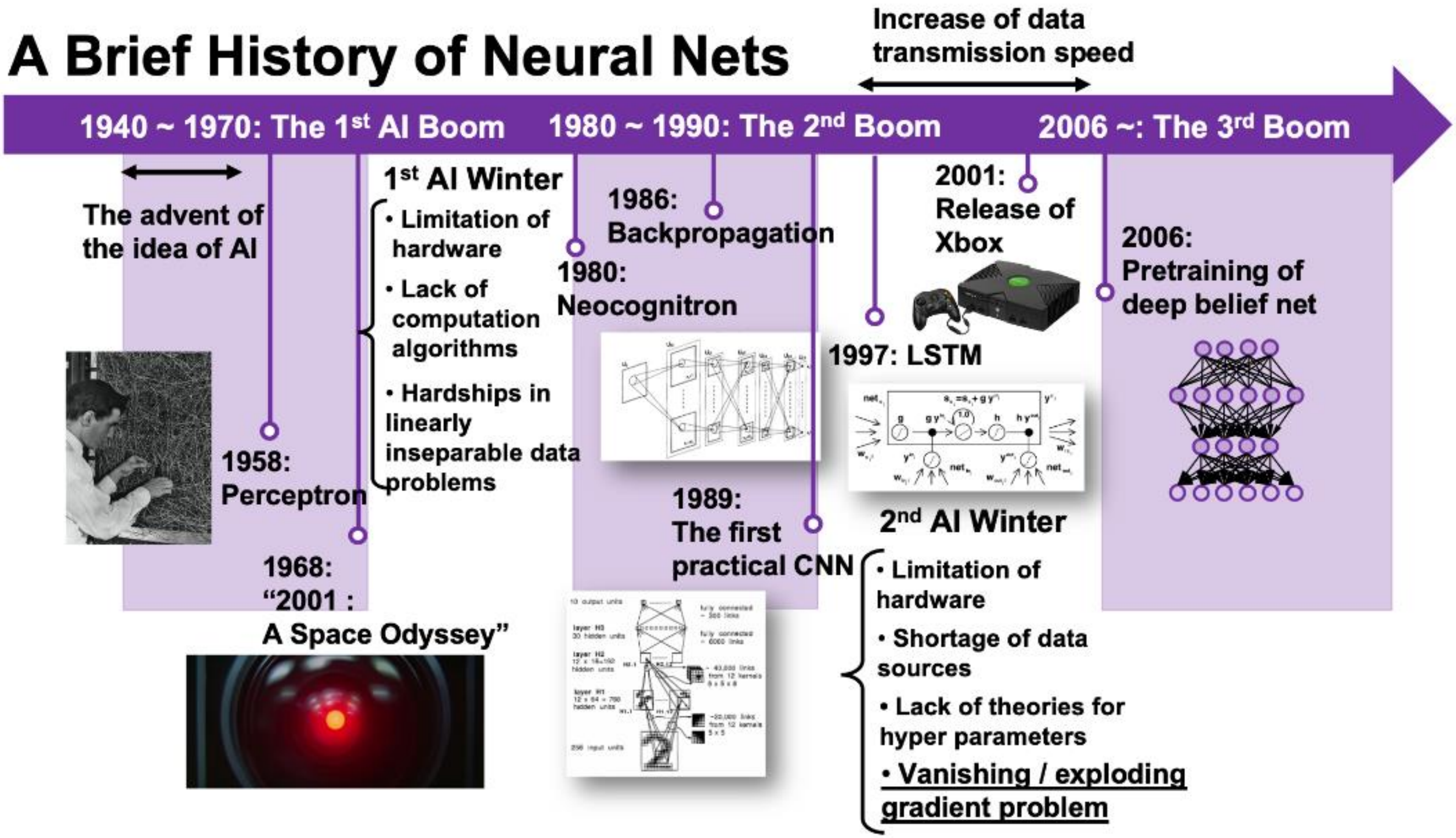
- Simple intro to AI
- AI in medicine
- AI in oncology
- AI in cancer genomics and transcriptomics
- *Use case*: AI-based classification of cancers of unknown primary
- Conclusions and perspectives

Don't hesitate to interrupt and ask questions!

Artificial intelligence (AI): brief history and key concepts

Artificial Intelligence Development History Timeline





https://data-science-blog.com/blog/2020/07/16/a-brief-history-of-neural-nets-everything-you-should-know-before-learning-lstm/

Quiz n°1

Quelles sont les utilisations potentielles de l'IA?

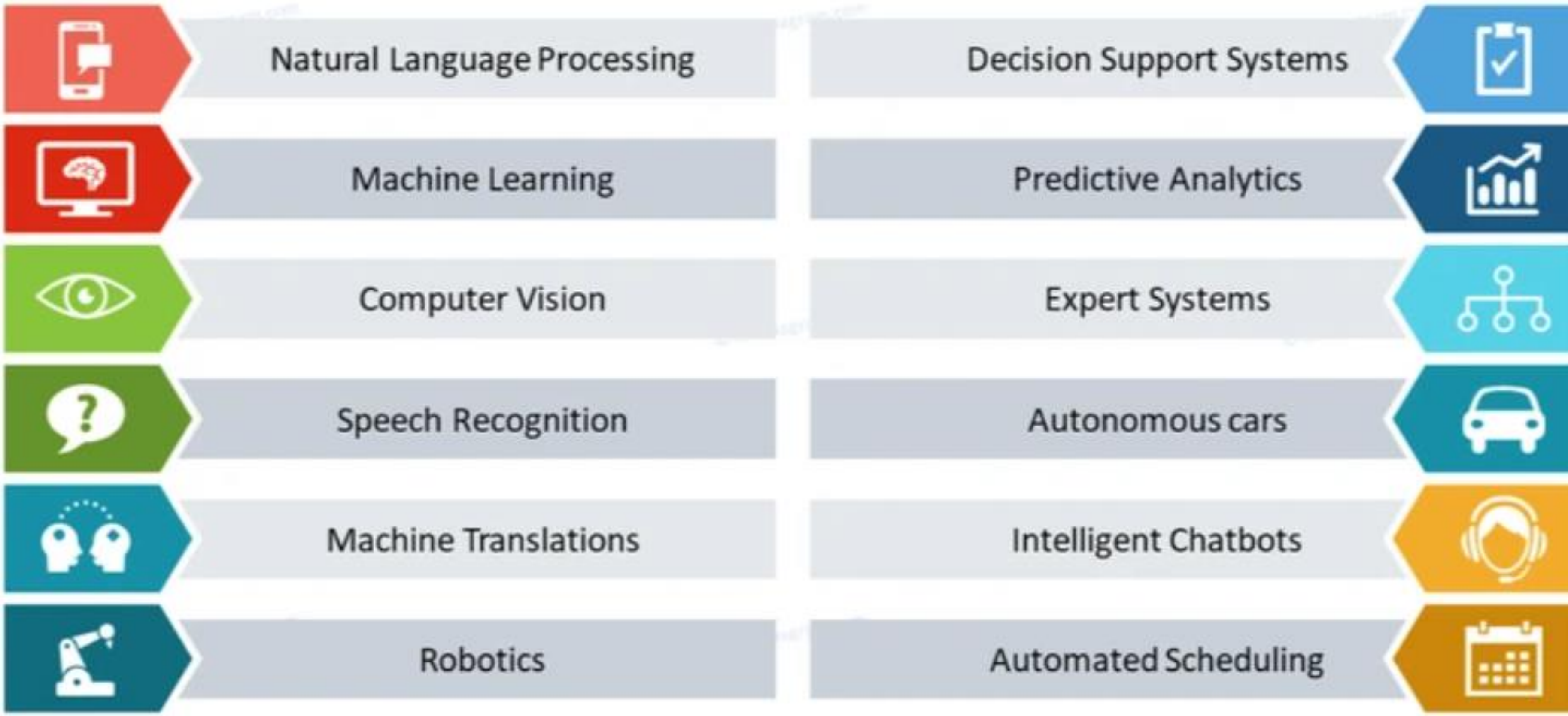
- A) Conduire une voiture
- B) Dépister un cancer du sein sur mammographie
- C) Traduire une conversation en temps réel
- D) Résoudre des problèmes de mathématiques de niveau olympiade

Quiz n°1

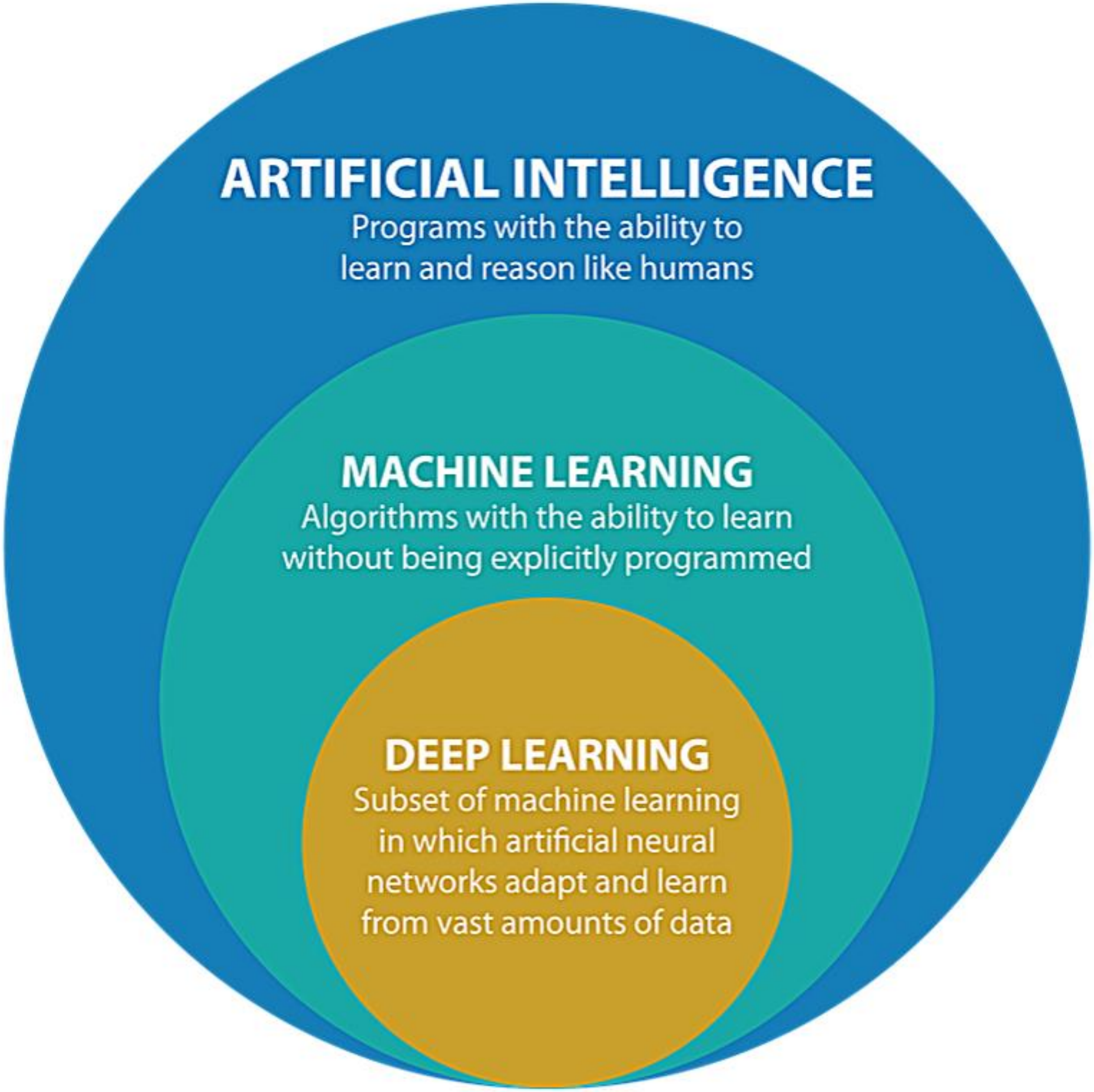
Quelles sont les utilisations potentielles de l'IA?

- A) Conduire une voiture
- B) Dépister un cancer du sein sur mammographie
- C) Traduire une conversation en temps réel
- D) Résoudre des problèmes de mathématiques de niveau olympiade

Some current applications of AI

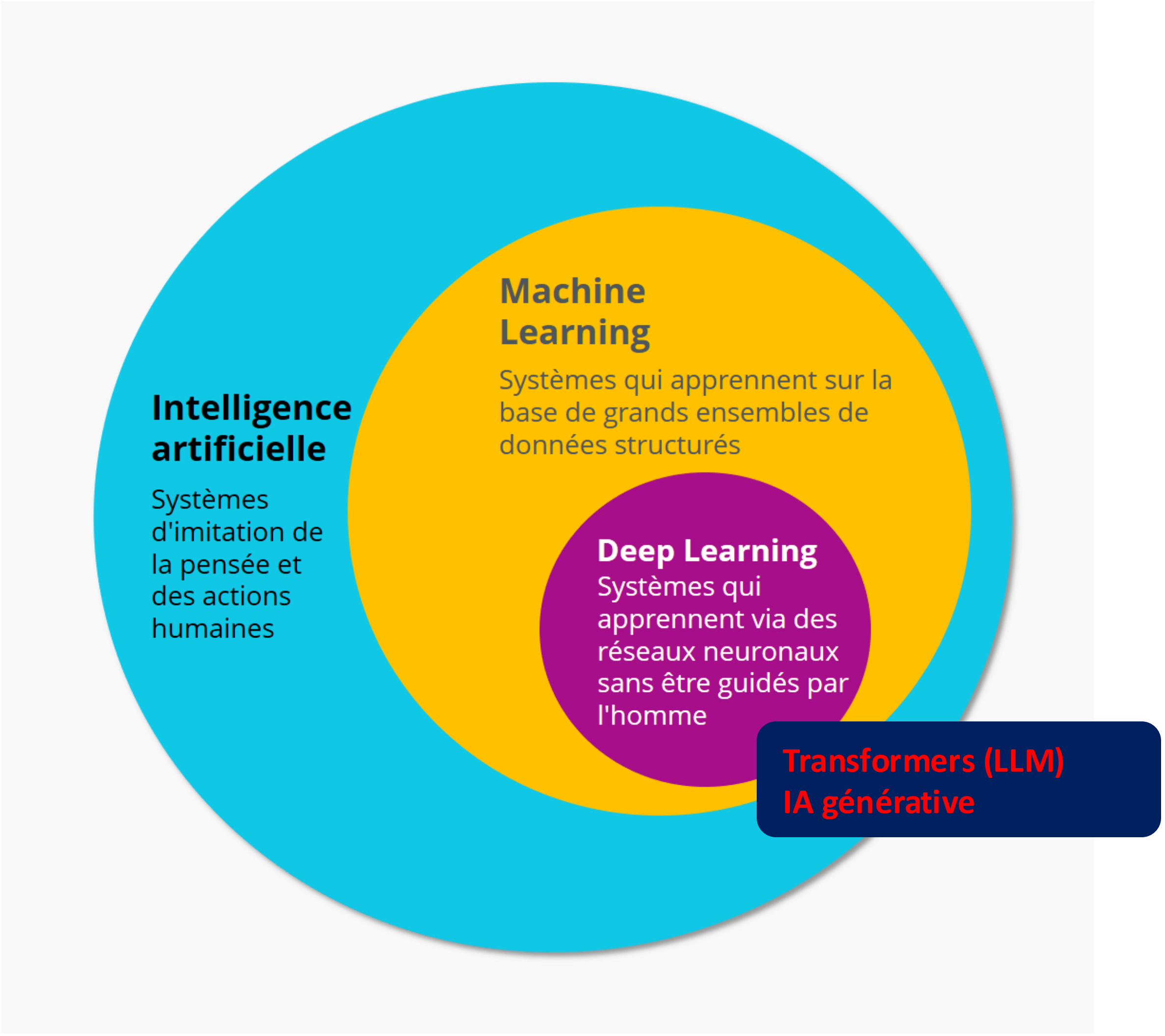


Definitions



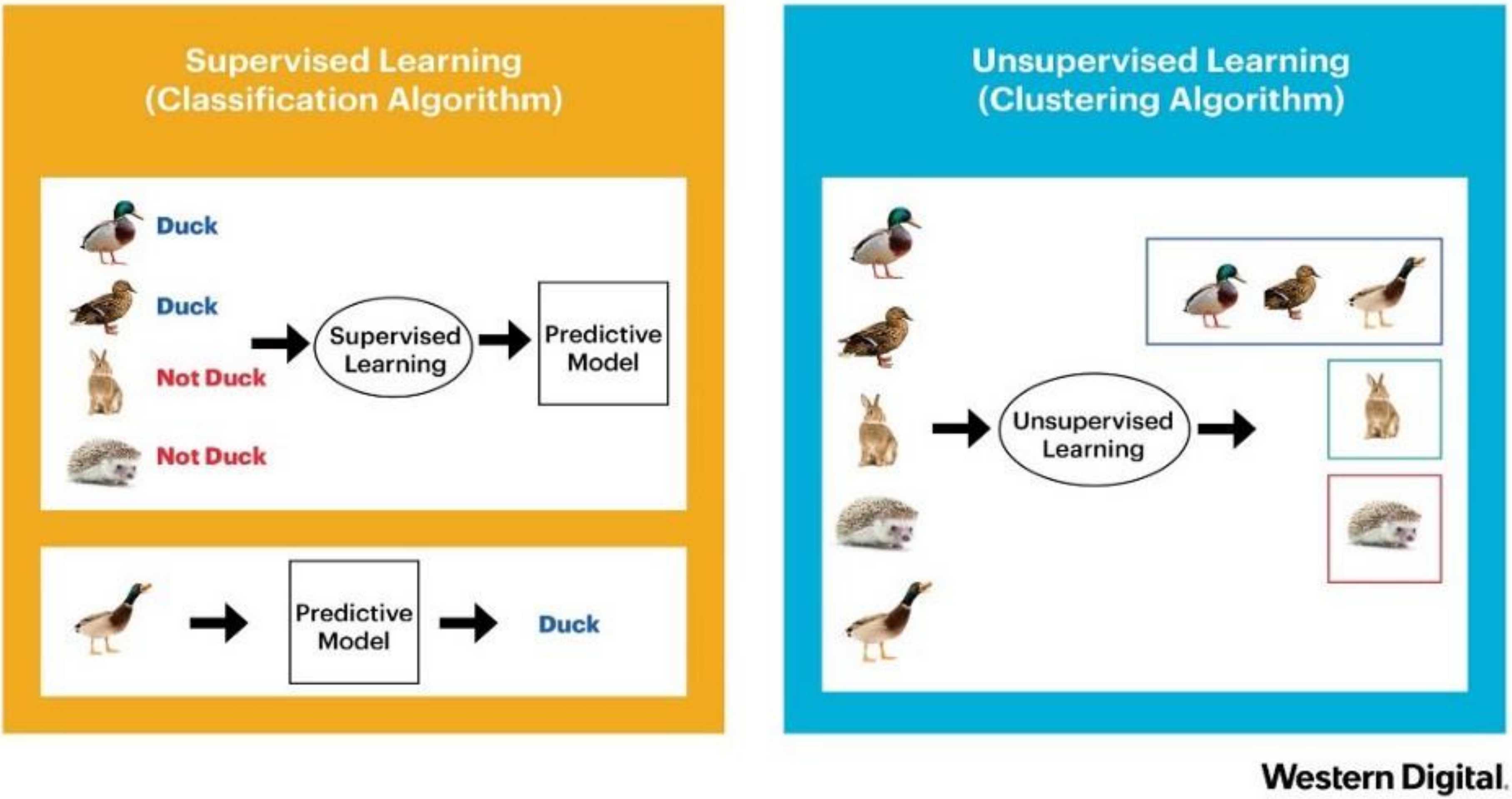
https://blog.digitalogy.co/the-difference-between-artificial-intelligence-machine-learning-and-deep-learning/

Définitions



https://www.ionos.fr/digitalguide/web-marketing/search-engine-marketing/deep-learning-vs-machine-learning/

Supervised/unsupervised learning



https://laptrinhx.com/supervised-vs-unsupervised-learning-1861950659/

Quiz n°2

Je souhaite entraîner un algorithme à discriminer entre lymphome et sarcome à partir du RNA-seq. Quel type d'apprentissage vais-je utiliser ?

- A) Apprentissage supervisé
- B) Apprentissage non supervisé

Quiz n°2

Je souhaite entraîner un algorithme à discriminer entre lymphome et sarcome à partir du RNA-seq. Quel type d'apprentissage vais-je utiliser ?

- A) Apprentissage supervisé
- B) Apprentissage non supervisé

Quiz n°3

Je souhaite entraîner un algorithme à découvrir des sous-groupes de lymphomes à partir du RNA-seq. Quel type d’apprentissage vais-je utiliser ?

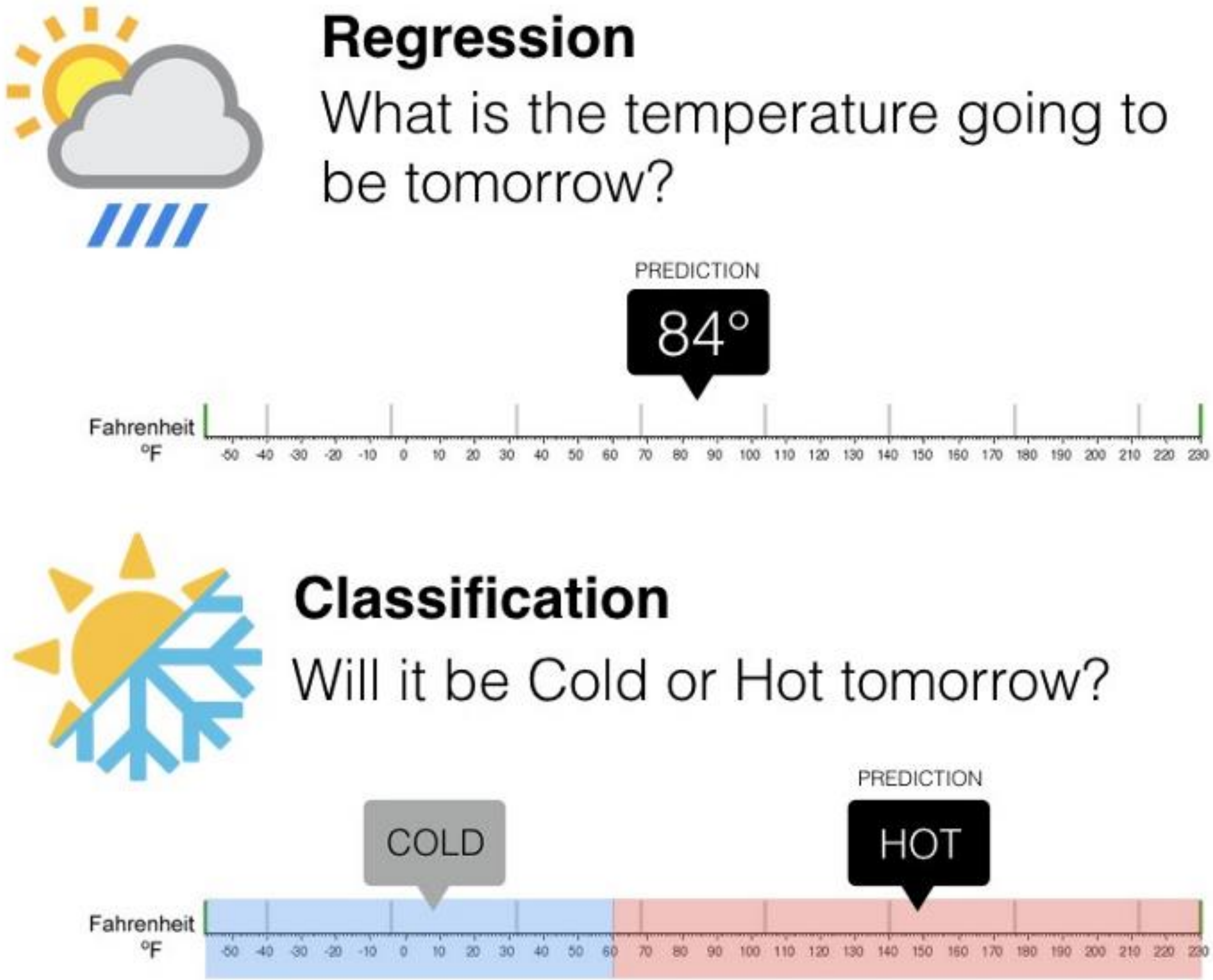
- A) Apprentissage supervisé
- B) Apprentissage non supervisé

Quiz n°3

Je souhaite entraîner un algorithme à découvrir des sous-groupes de lymphomes à partir du RNA-seq. Quel type d'apprentissage vais-je utiliser ?

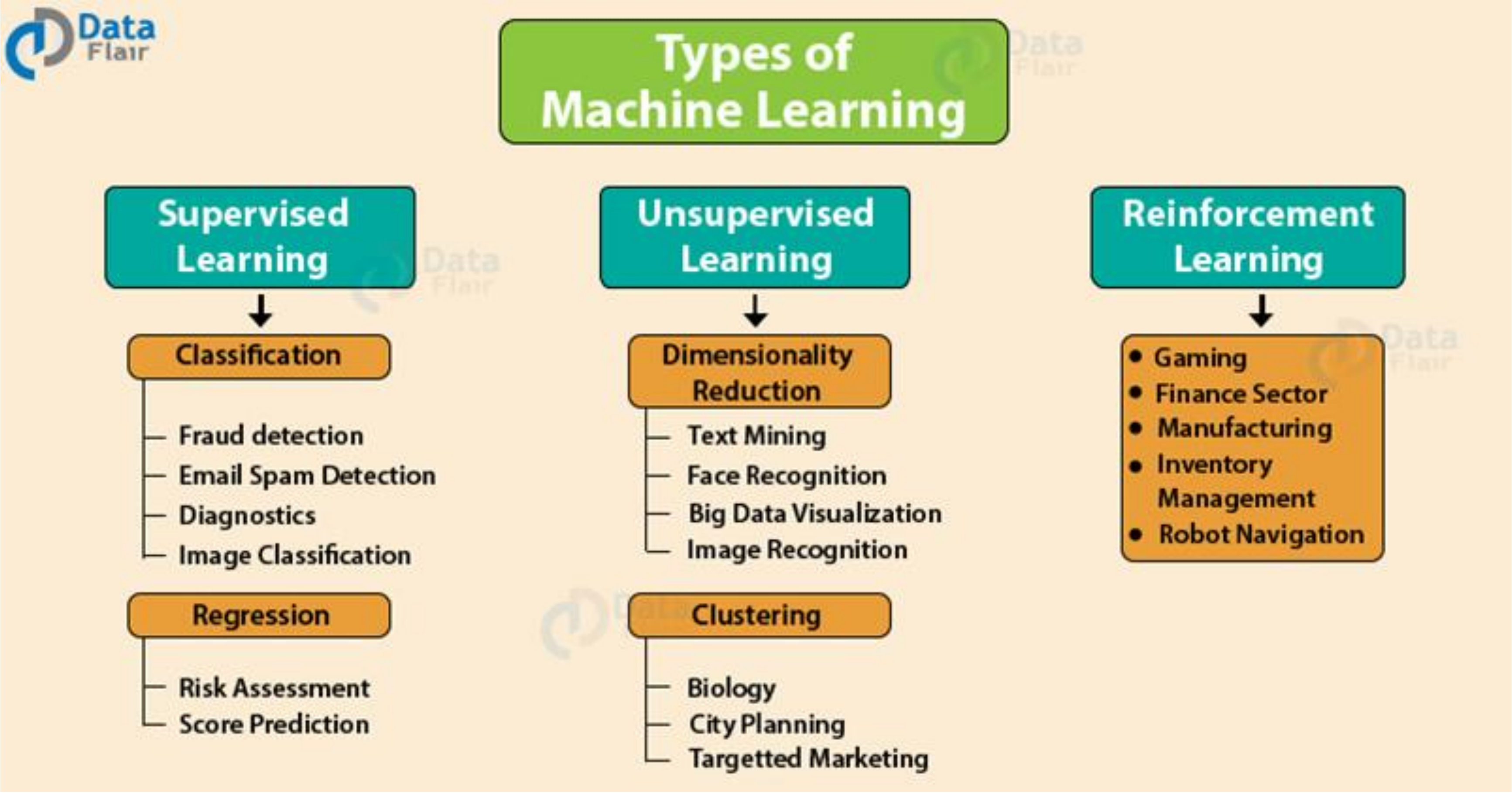
- A) Apprentissage supervisé
- B) Apprentissage non supervisé

Regression or classification



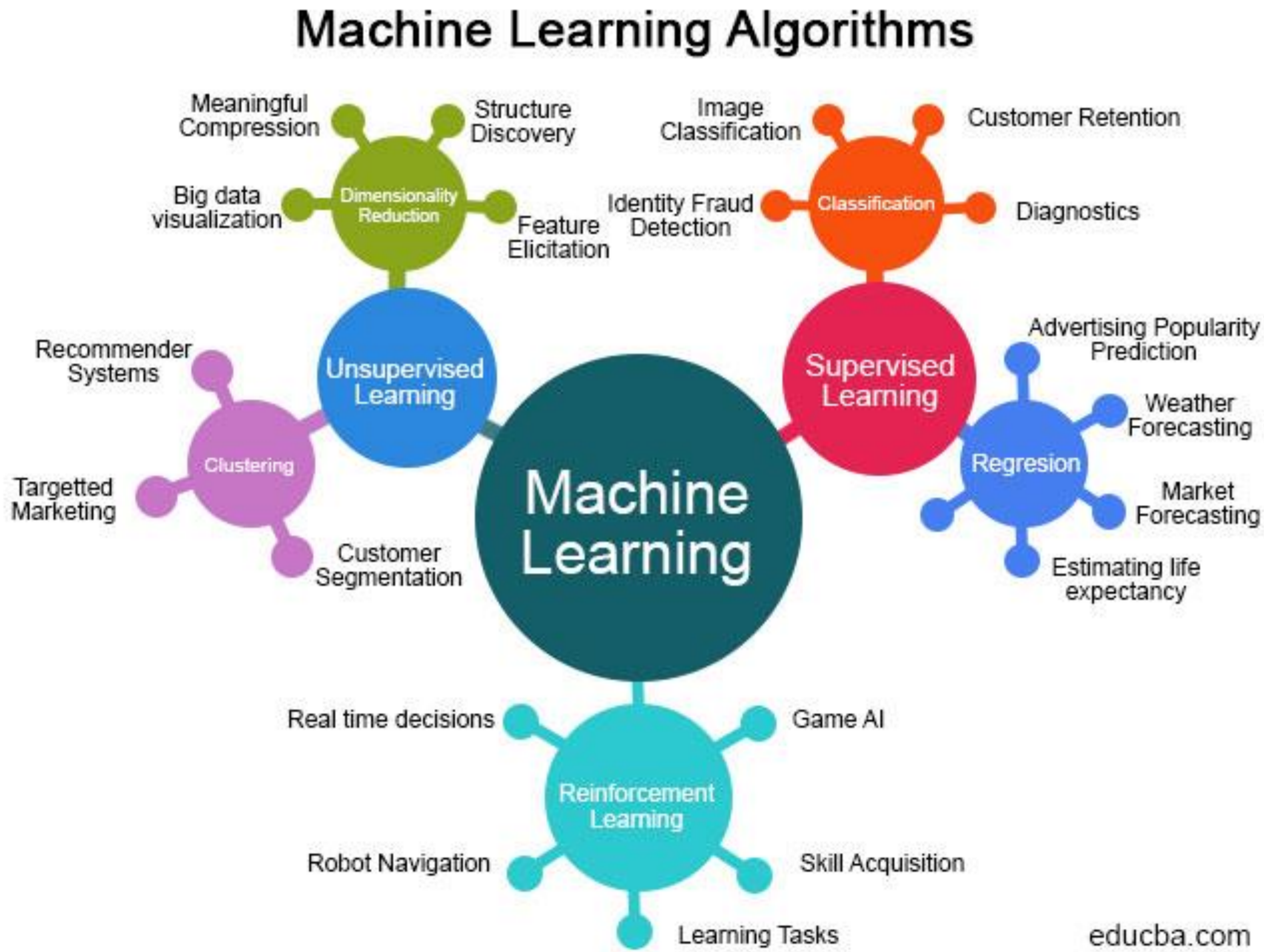
https://towardsdatascience.com/regression-or-classification-linear-or-logistic-f093e8757b9c

Main types of machine learning



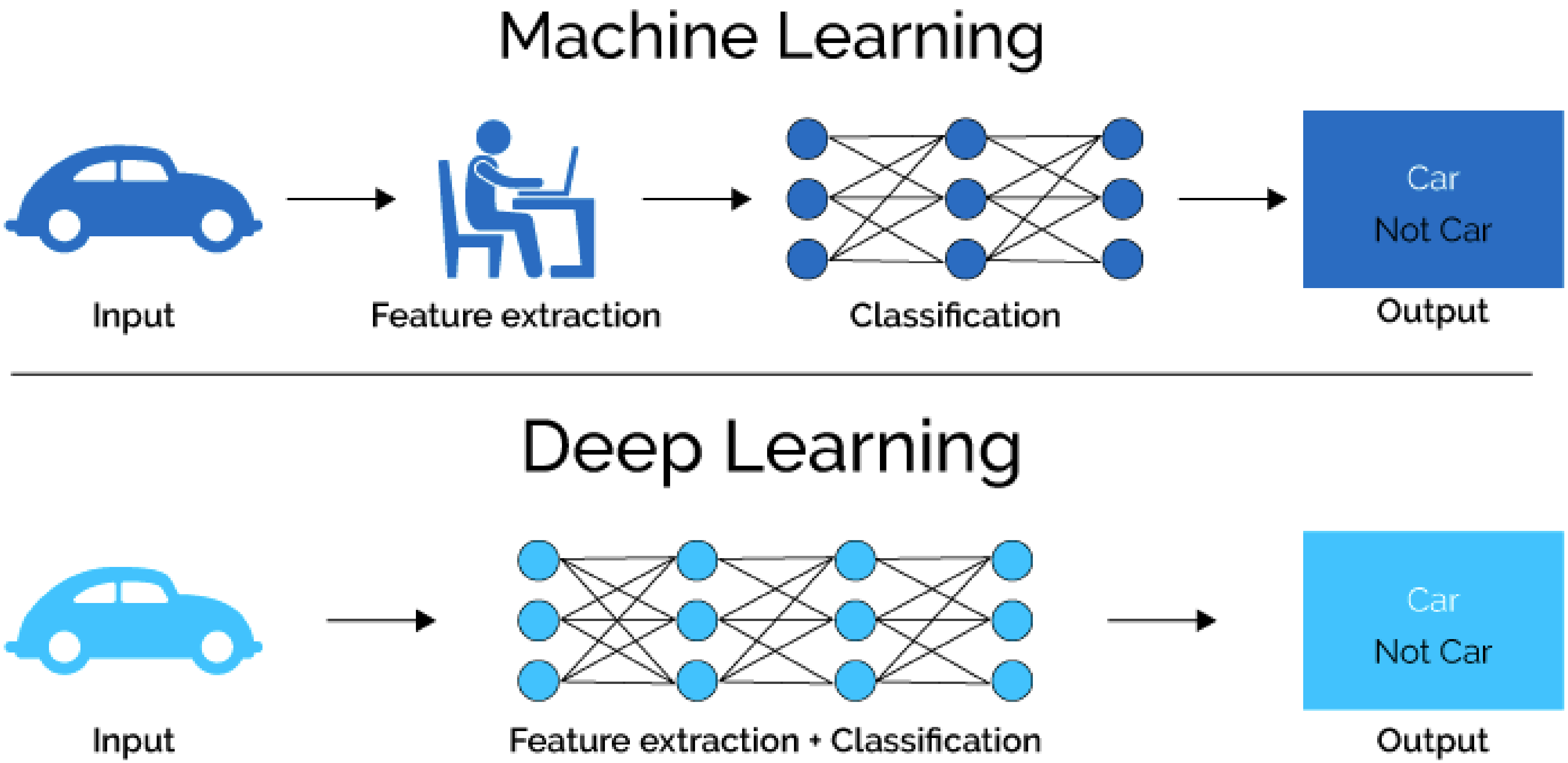
https://data-flair.training/blogs/types-of-machine-learning-algorithms/

Algorithms are proliferating…



https://www.educba.com/machine-learning-algorithms/

Machine learning vs deep learning



https://lawtomated.com/tag/unsupervised-learning/

What algorithms are eating: « Big data »

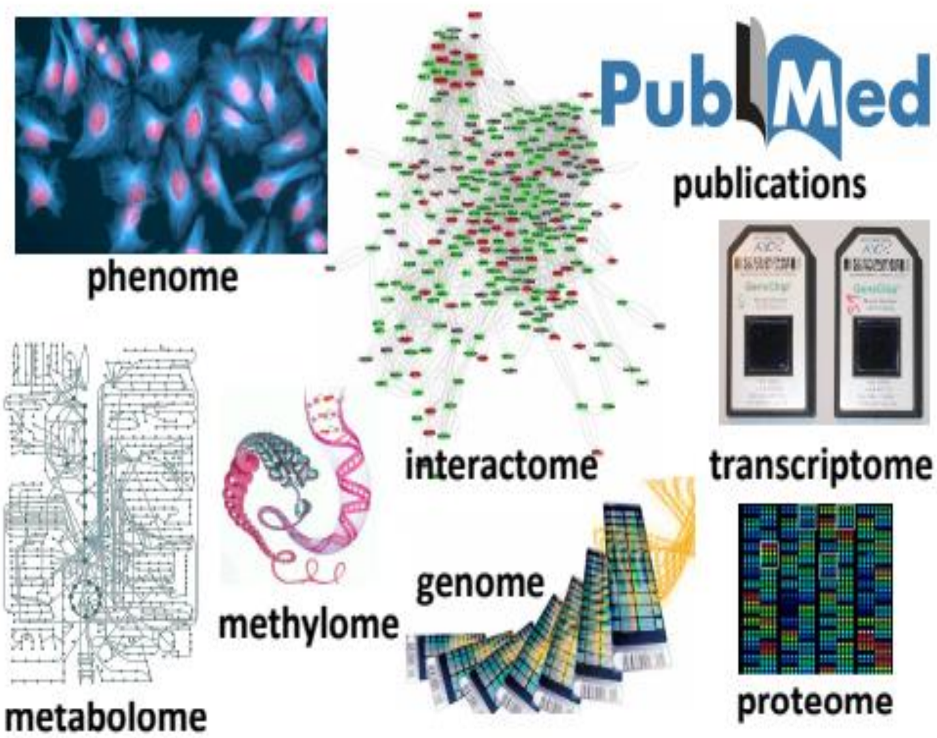
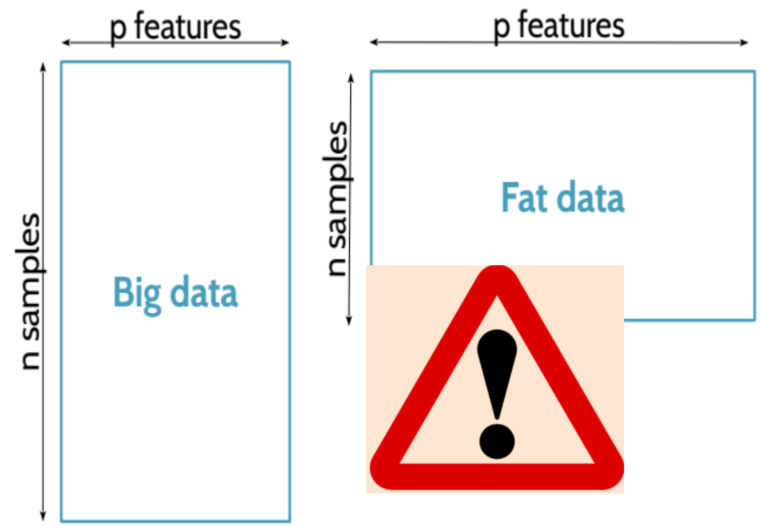


Image sources: ajc1@ flickr; Zlir'a@wikimedia

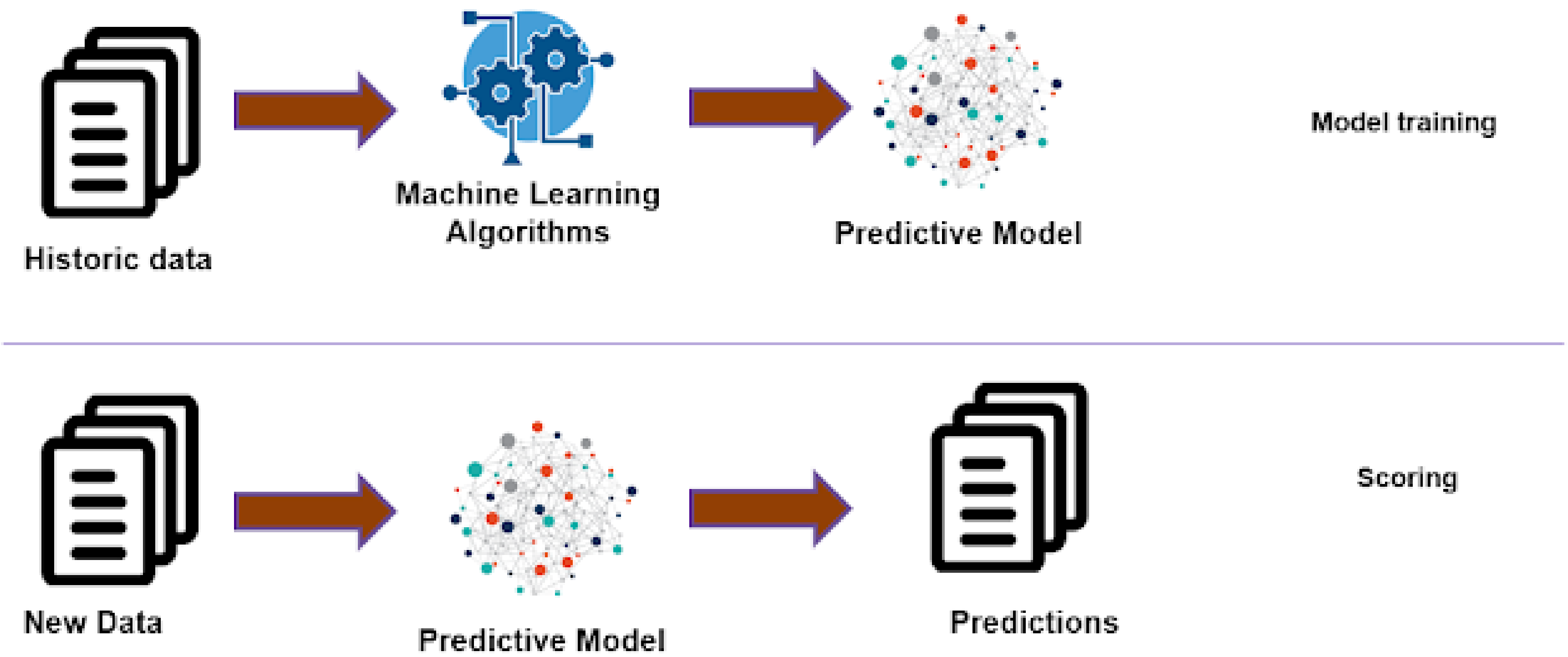


E.g. Genome-Wide Association Studies (GWAS):

- $p = 10^5 - 10^7$ Single Nucleotide Polymorphisms (SNPs)
- $n = 10^2 - 10^4$ samples.

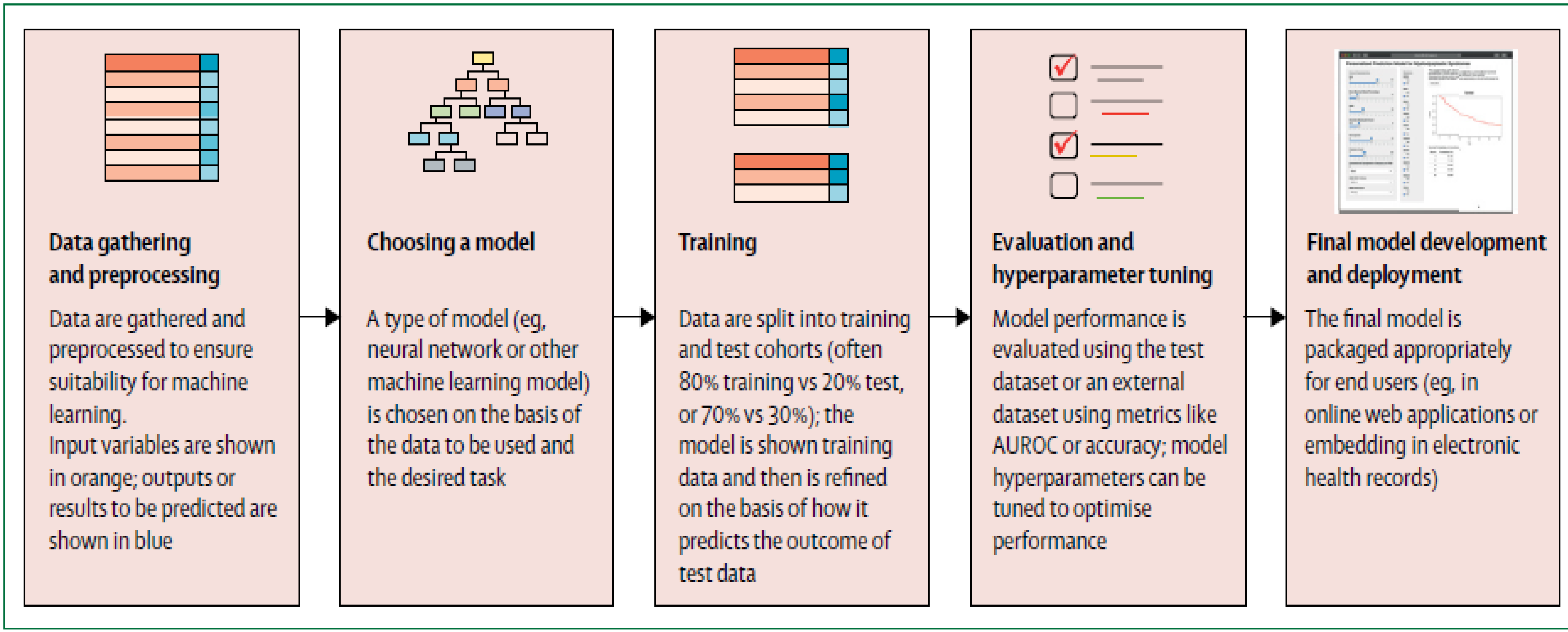
http://cazencott.info/dotclear/public/publications/2016-01-13-azencott.pdf

Typical machine learning workflow



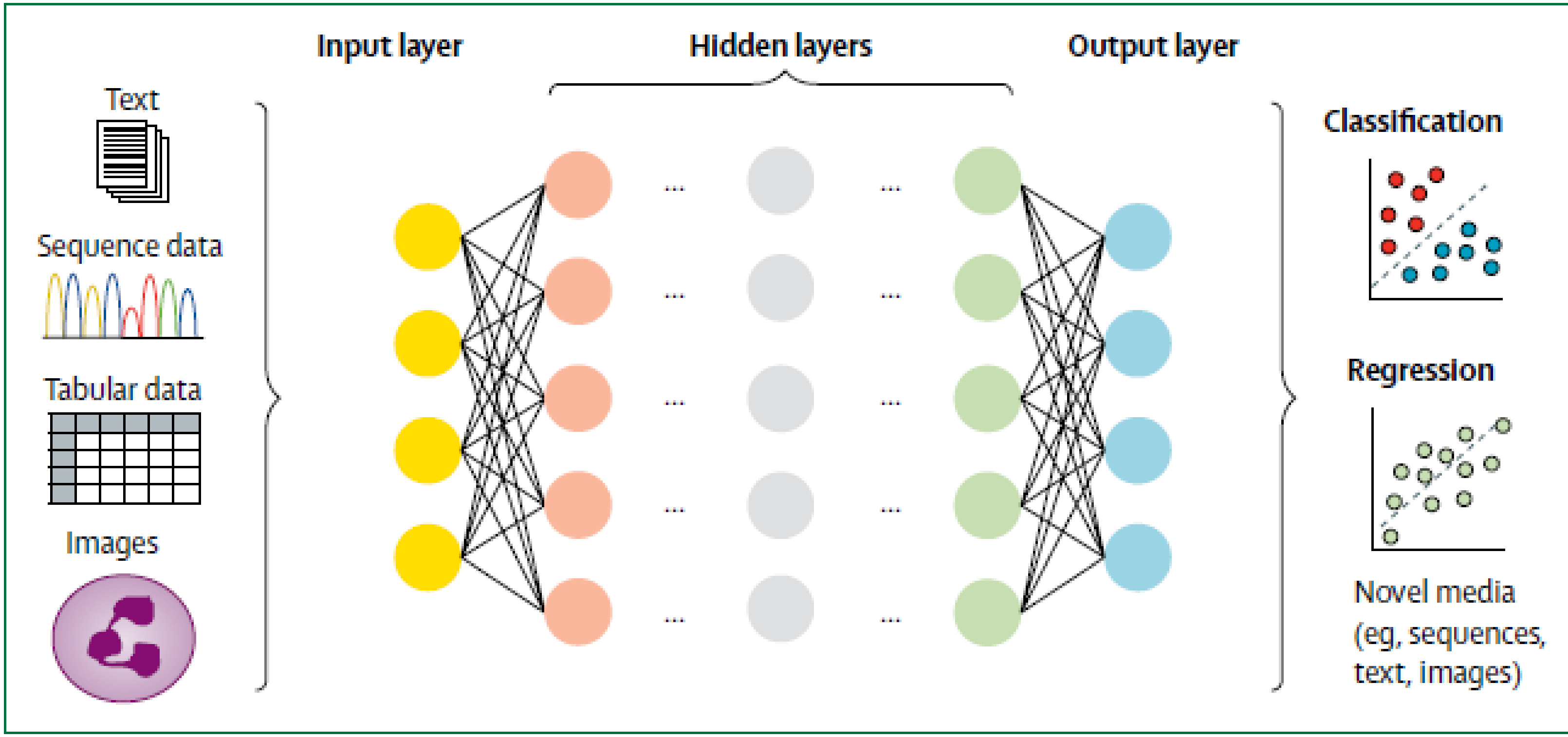
http://singaporebusinessintelligence.blogspot.com/2018/10/what-is-automated-machine-learning.html

Typical machine learning workflow (2)



Radakovich et al., Lancet Haem 2020

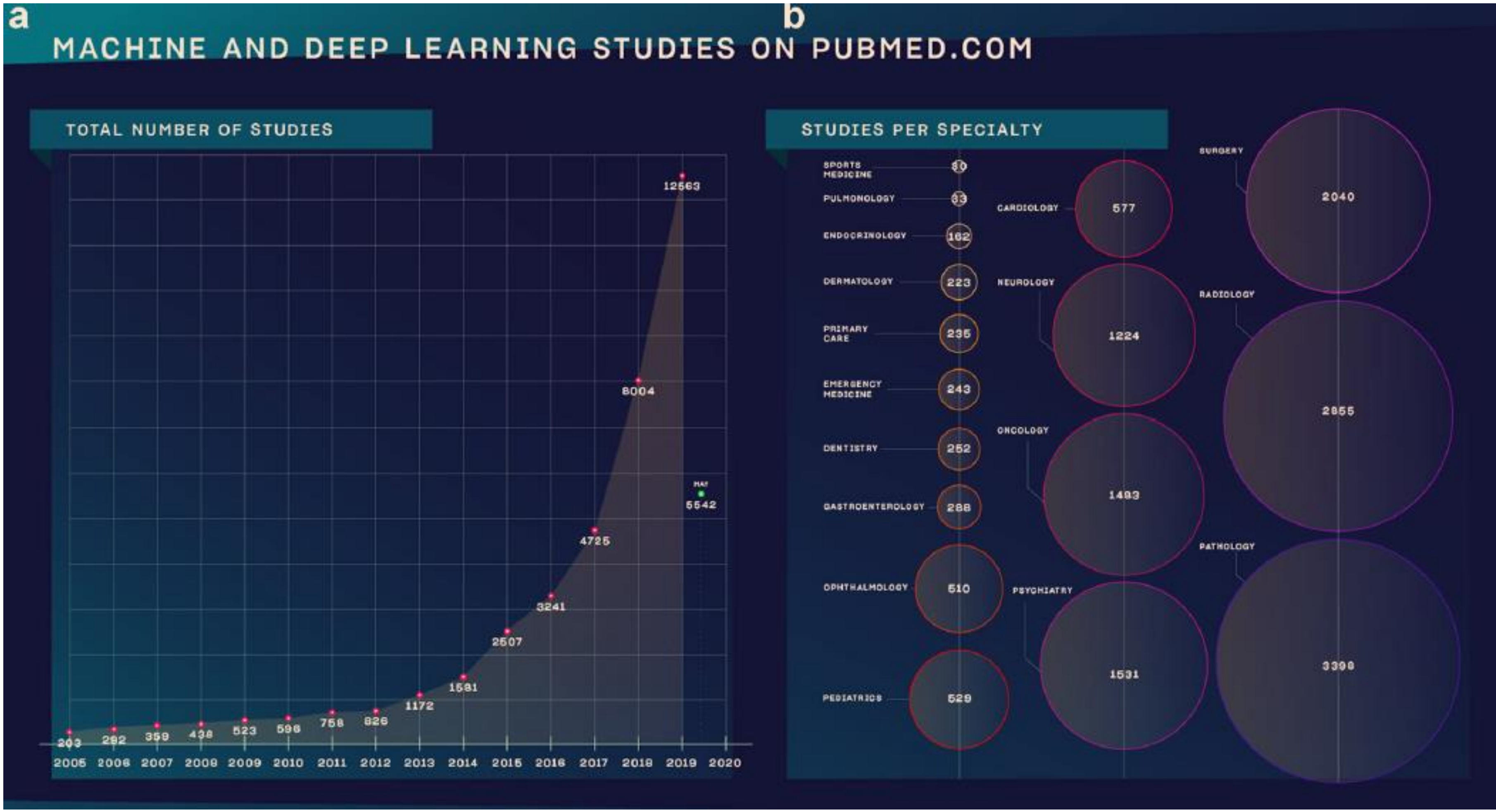
Typical scheme of deep learning



Radakovich et al., Lancet Haem 2020

AI in medicine:
what, why, how?

AI and medicine: articles are proliferating…



Mesko et al., Npj Digital Med, 2020

Potential applications in medicine (non-exhaustive list)

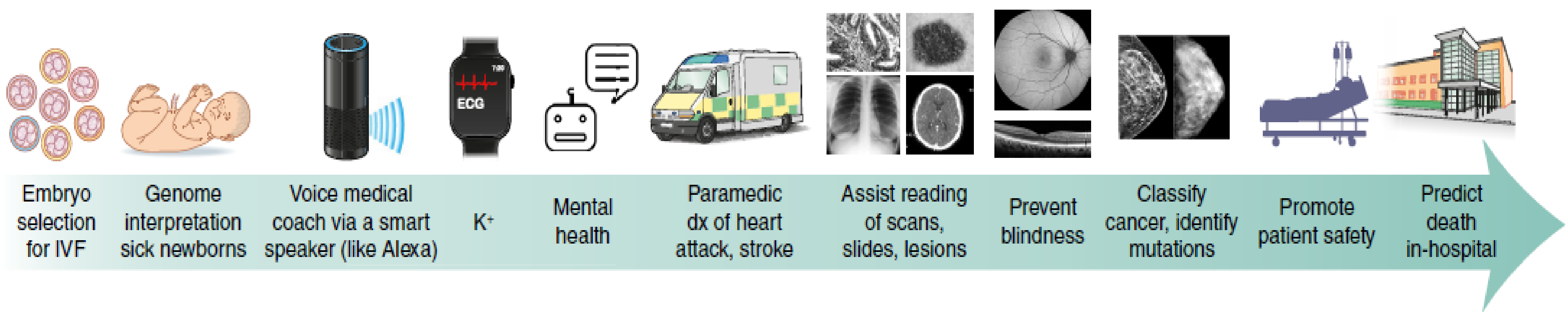


Fig. 2 | Examples of AI applications across the human lifespan. dx, diagnosis; IVF, in vitro fertilization K^+, potassium blood level. Credit: Debbie Maizels/ Springer Nature

Topol, Nat Med 2019

Potential applications in medicine

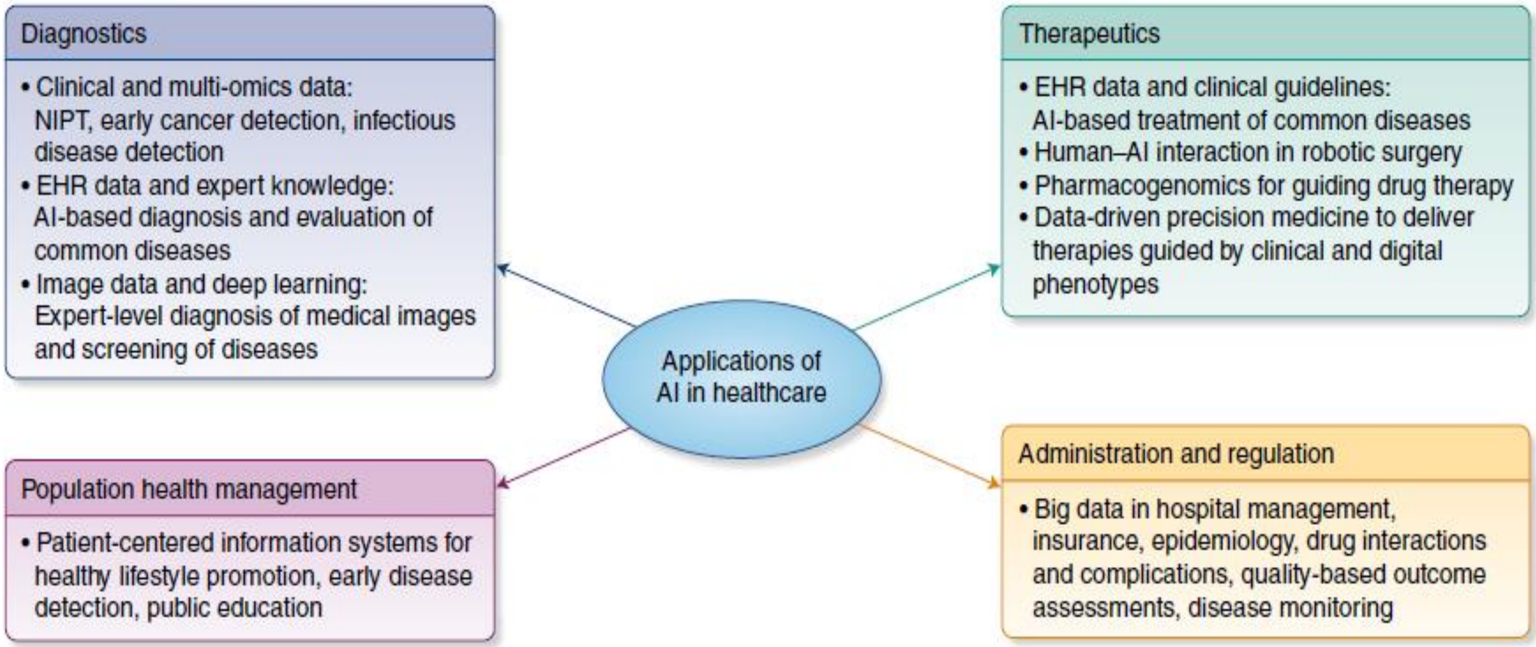
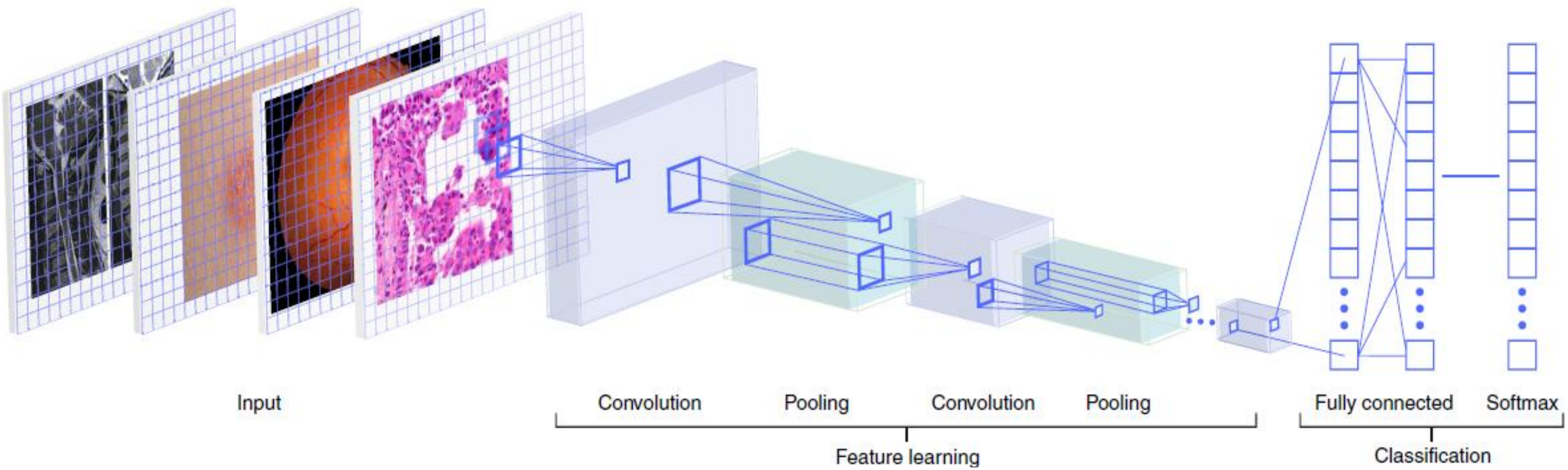
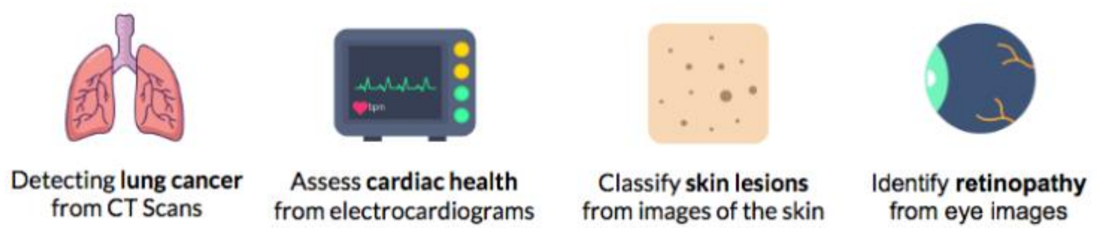


Fig. 1 | Potential roles of AI-based technologies in healthcare. In the healthcare space, AI is poised to play major roles across a spectrum of application domains, including diagnostics, therapeutics, population health management, administration, and regulation. NIPT, noninvasive prenatal test. Credit: Debbie Maizels/Springer Nature

He et al., Nat Med 2019

Pioneer applications: medical image analysis





https://www.datarevenue.com/en-blog/artificial-intelligence-in-medicine

Esteva et al., Nat Med 2019

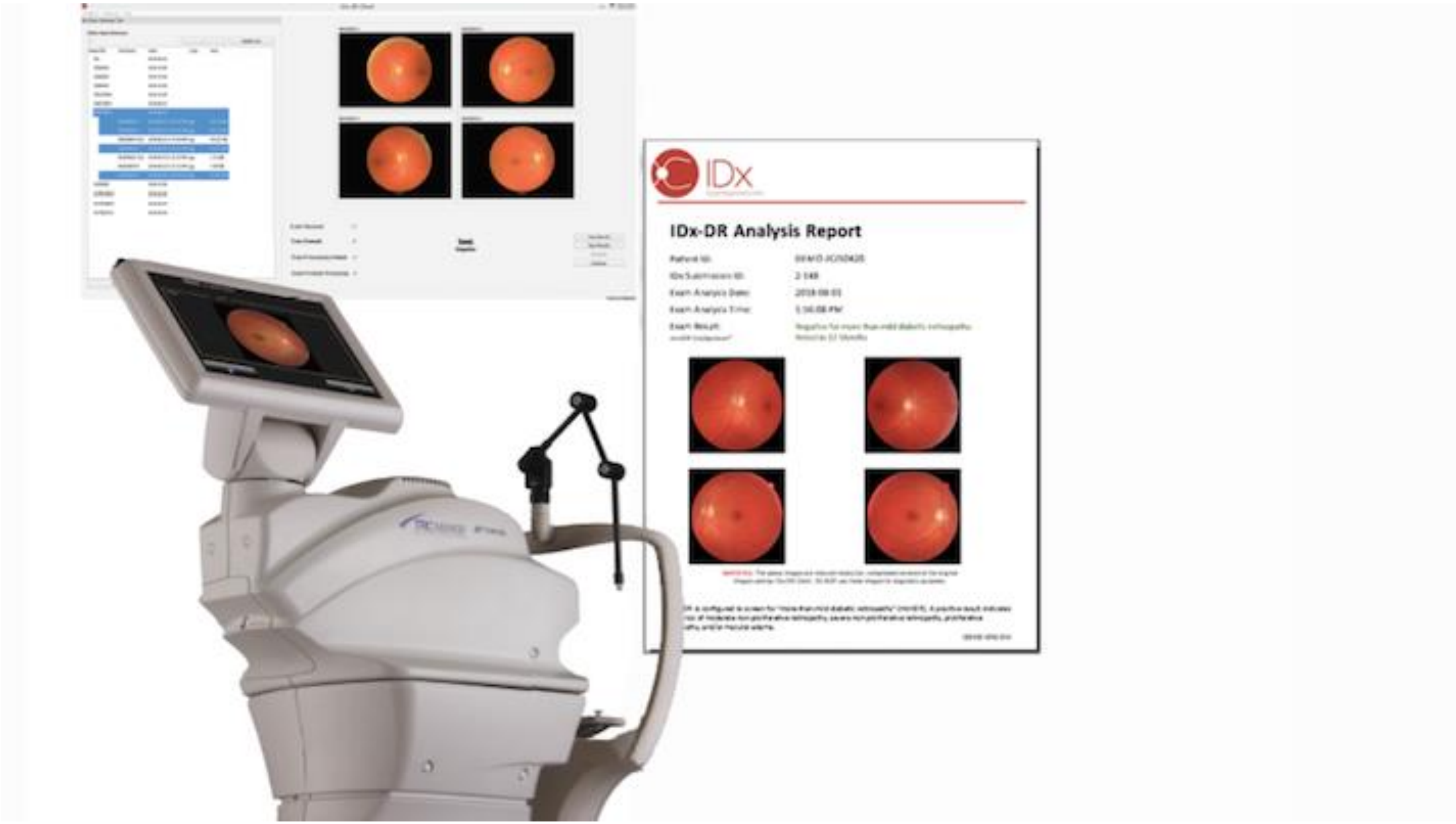
Some AI tools are already used in clinical practice

Table 2 | FDA AI approvals are accelerating

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

Topol, Nat Med 2019

Screening for diabetic retinopathy



https://dxs.ai/

Perspective: the « virtual medical coach »

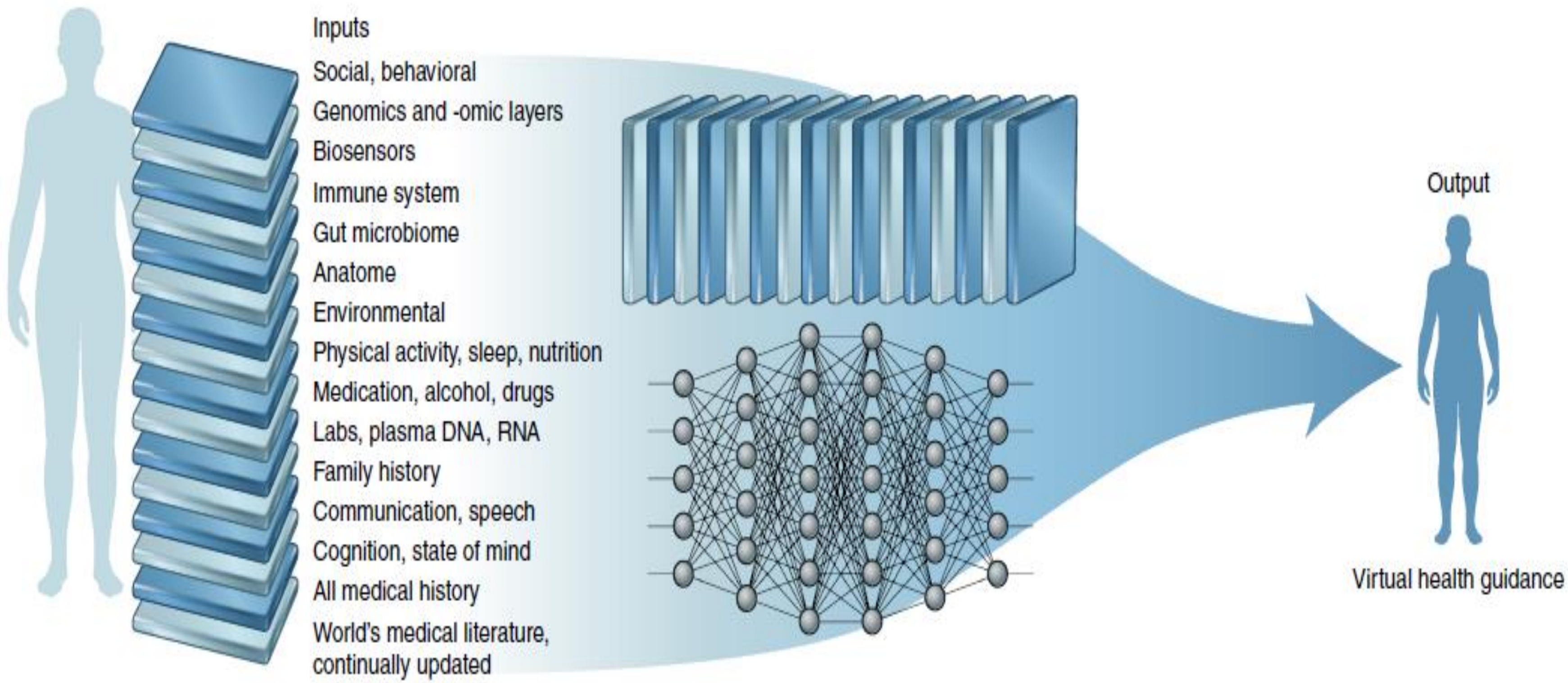


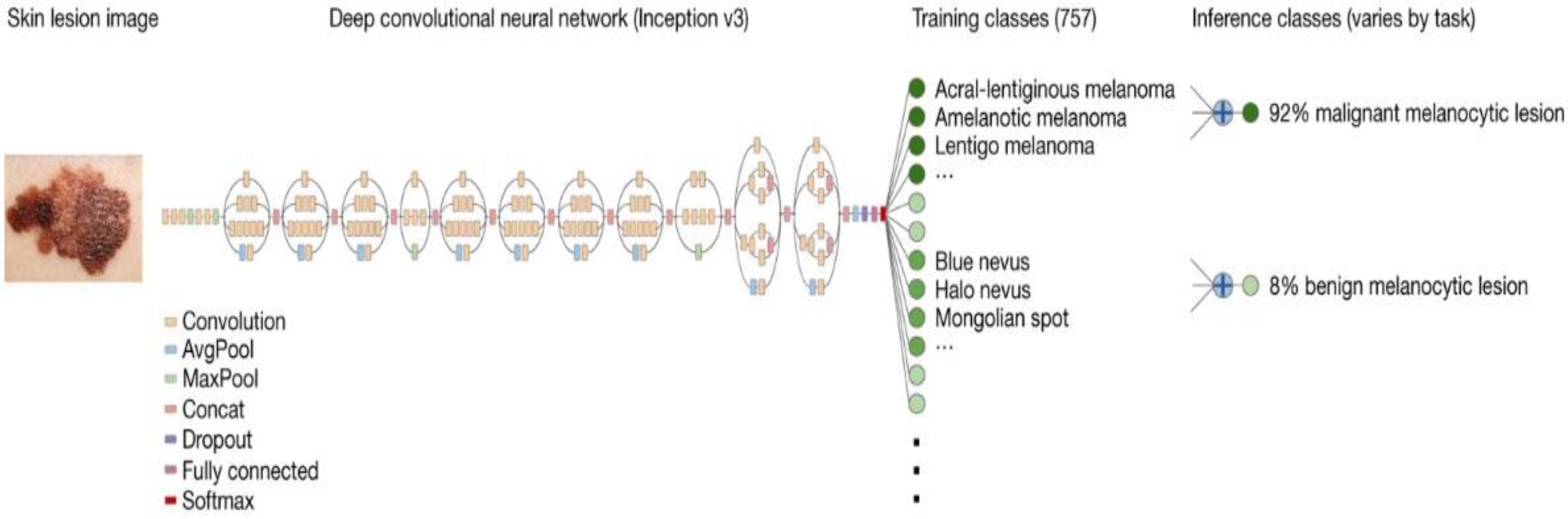
Fig. 3 | The virtual medical coach model with multi-modal data inputs and algorithms to provide individualized guidance. A virtual medical coach that uses comprehensive input from an individual that is deep learned to provide recommendations for preserving the person's health. Credit: Debbie Maizels/ Springer Nature

Topol, Nat Med 2019

AI in oncology:
some applications

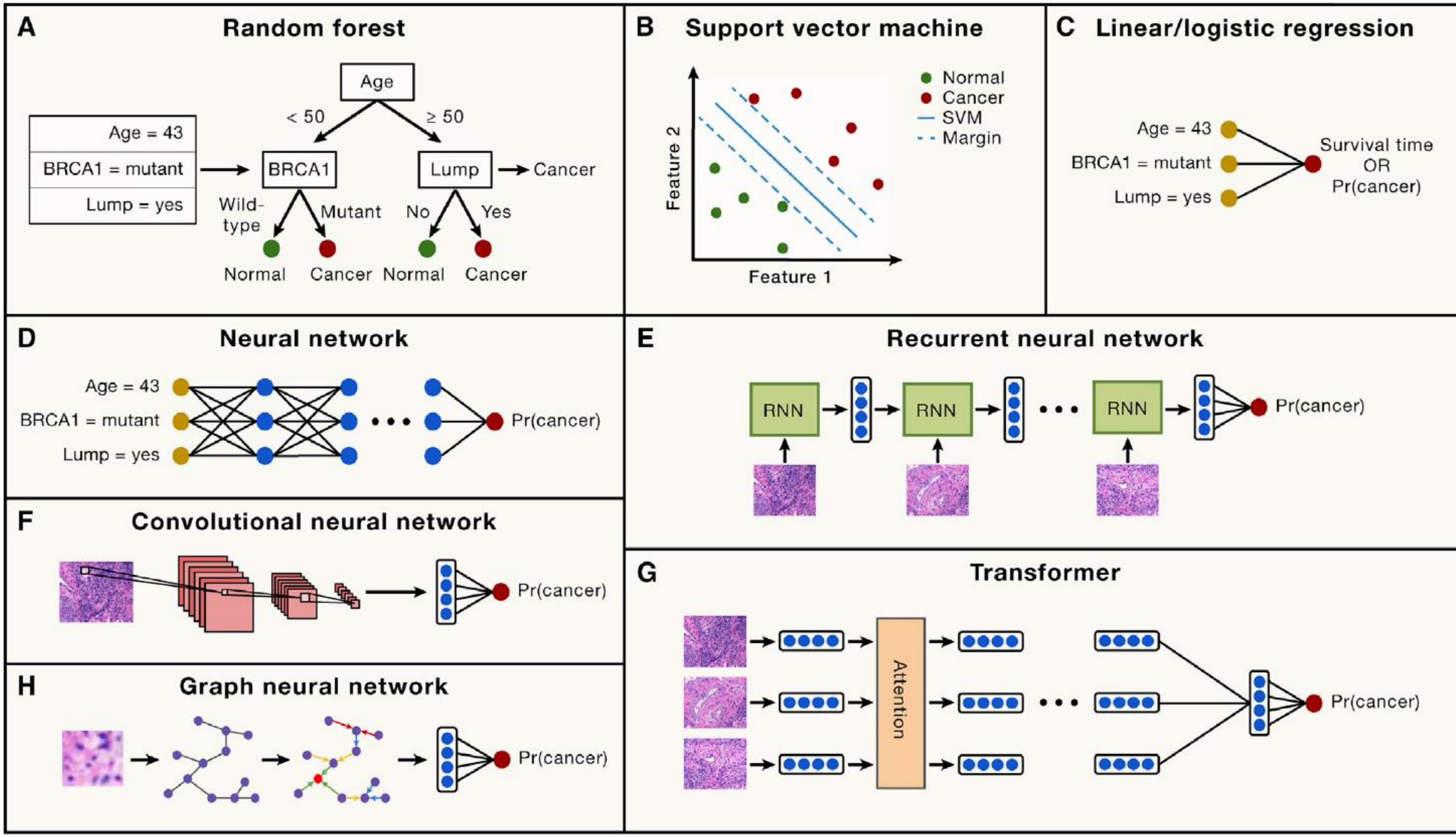
Skin cancer classification

From: Dermatologist-level classification of skin cancer with deep neural networks



Esteva et al., Nature 2017

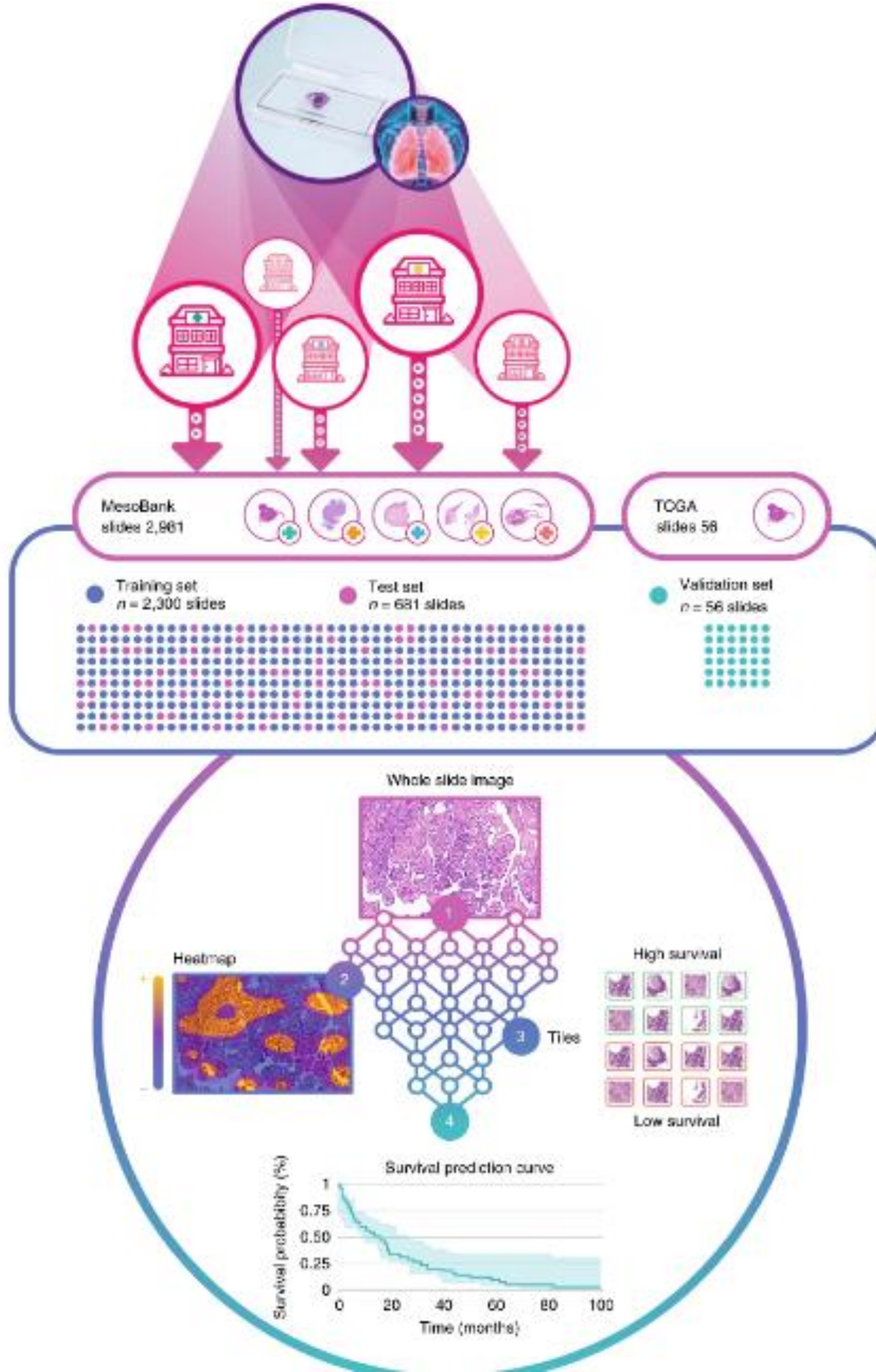
Main types of ML used in oncology



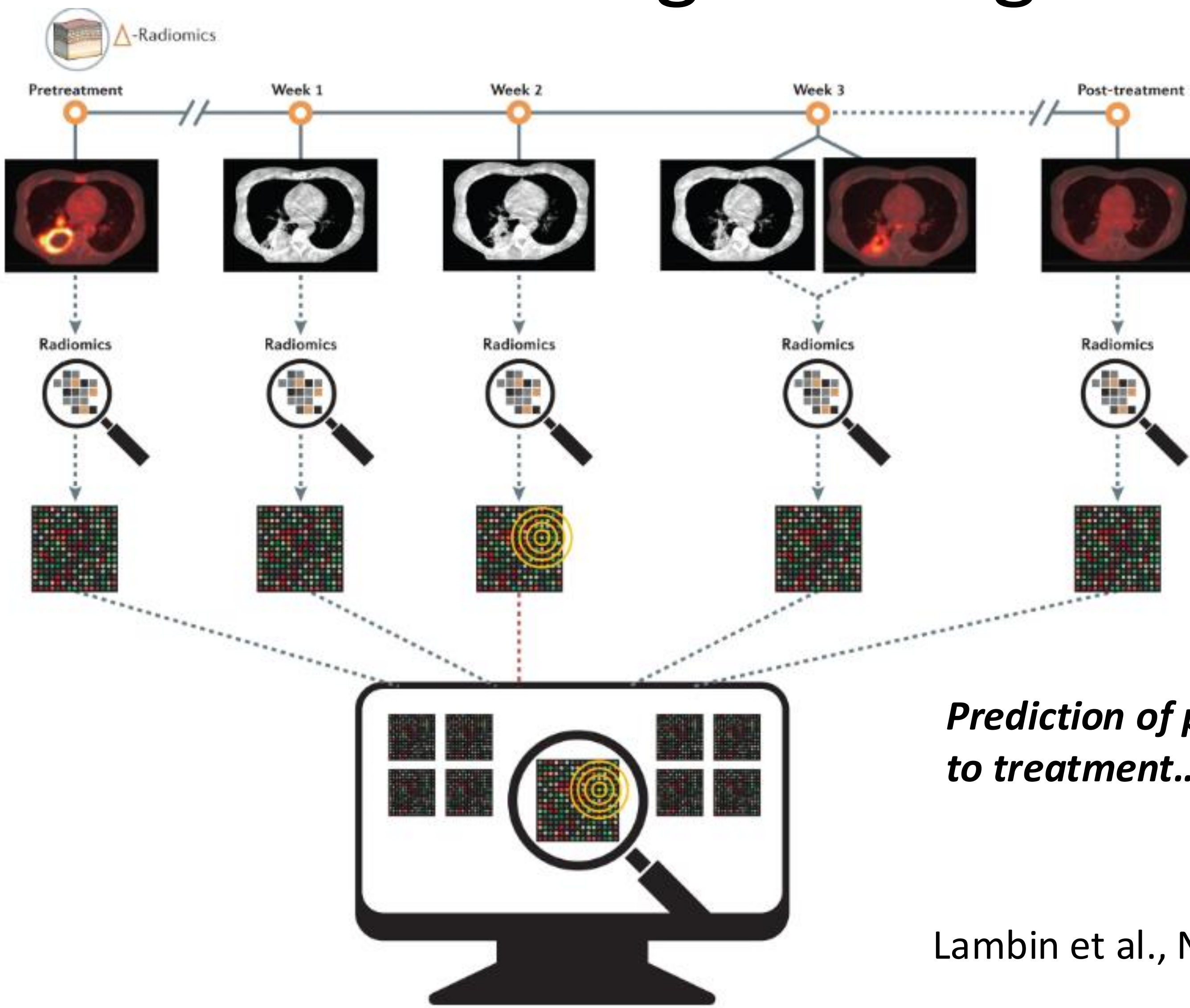
Swanson et al., Cell 2023

Prediction of mesothelioma prognosis based on pathological slides

Courtiol et al., Nature Med 2019

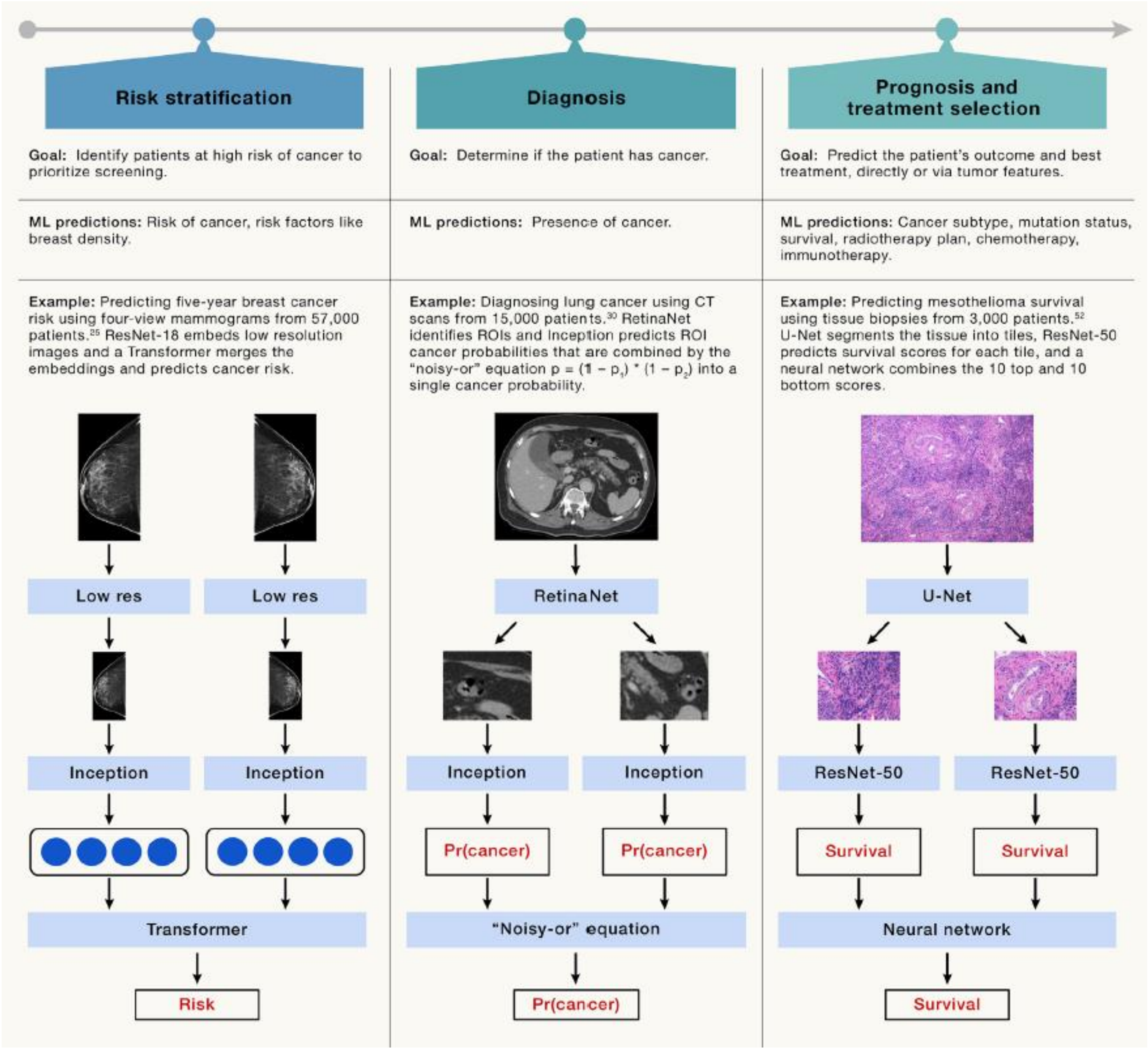


Radiomics: « texture » analysis of radiological images



Prediction of prognosis, response to treatment…

Lambin et al., Nat Rev Clin Onc 2017



Swanson et al., Cell 2023

AI in hematology

	Sample size	Application	Method	Results
Kimura et al (2019)[21]	3261 peripheral smears	Leucocyte classification; distinguishing aplastic anaemia and myelodysplastic syndrome	CNN, gradient boosting	Sensitivity vs specificity was 93·5% vs 96·0% for leucocyte detection, and 96·2% vs 100% for aplastic anaemia vs myelodysplastic syndrome differentiation
Achi et al (2019)[22]	128 patients	Differentiating diffuse large B-cell lymphoma, small lymphocytic lymphoma, Burkitt lymphoma, and normal lymph nodes	CNN	95% accuracy per slide; 100% accuracy per patient
Li (2019)[23]	41 patients	Detection of acute myeloid leukaemia bone marrow involvement via PET-CT	Manual feature engineering (PyRadiomics)	Sensitivity vs specificity was 87·5% vs 89·5% for bone marrow involvement; outperformed visual inspection of scans
Milgrom et al (2019)[24]	251 patients	Predicting refractory Hodgkin lymphoma from PET-CT	CNN	AUROC of 0·95 for model vs 0·78 for tumour volume; 0·65 for standardised uptake value
Moraes et al (2019)[25]	283 patients	Differential diagnosis of chronic lymphocytic leukaemia and B cell lymphomas via flow cytometry	Decision tree	95% inclusion of correct diagnosis in differential diagnosis; 66% definitive diagnosis
Ni et al (2016)[26]	51 patients	Detection of MRD in acute myeloid leukaemia via flow cytometry	Support vector machine	Similar performance with manual flow analysis (concordance=0·986)
Fuse (2019)[27]	217 patients	Prediction of acute leukaemia relapse after allogeneic stem cell transplantation	Decision tree	0·75 AUROC for relapse after transplantation
Goswami et al (2019)[28]	347 patients	Risk stratification for autologous stem cell transplantation in multiple myeloma	Decision tree	Significant risk-stratification and identification of high-risk features
Nazha (2019)[29]	433 and 113 patients	Predicting resistance to hypomethylating agents in patients with myelodysplastic syndrome on the basis of NGS myeloid malignancy panel	Recommender algorithm	Improved stratification of patients by risk of resistance to a hypomethylating agent
Gal et al (2019)[30]	493 patients	Response to induction therapy in paediatric acute myeloid leukaemia	K-nearest neighbours	0·84 AUROC for response to induction
Candia (2015)[31]	60 patients	Unsupervised analysis of microRNA expression signatures in acute myeloid leukaemia and acute lymphocytic leukaemia	Dimension reduction, network analysis	Identification of novel microRNA signatures in acute myeloid leukaemia and acute lymphocytic leukaemia

AUROC=area under the receiver operating characteristic curve. CNN=convolutional neural network. MRD=minimal residual disease. NGS=next-generation sequencing.

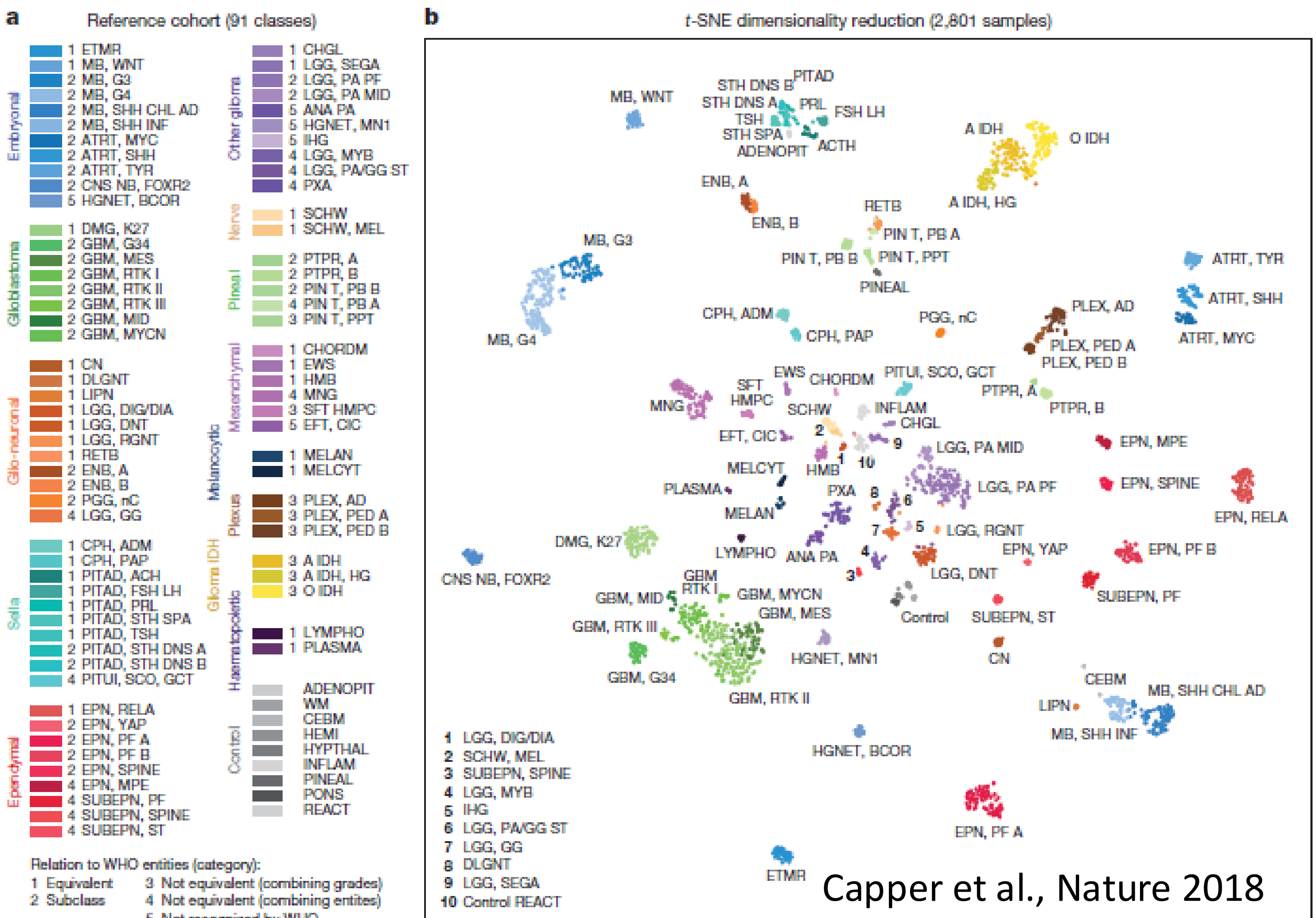
***Table:* Representative artificial intelligence in malignant haematology publications**

AI in cancer genomics and transcriptomics

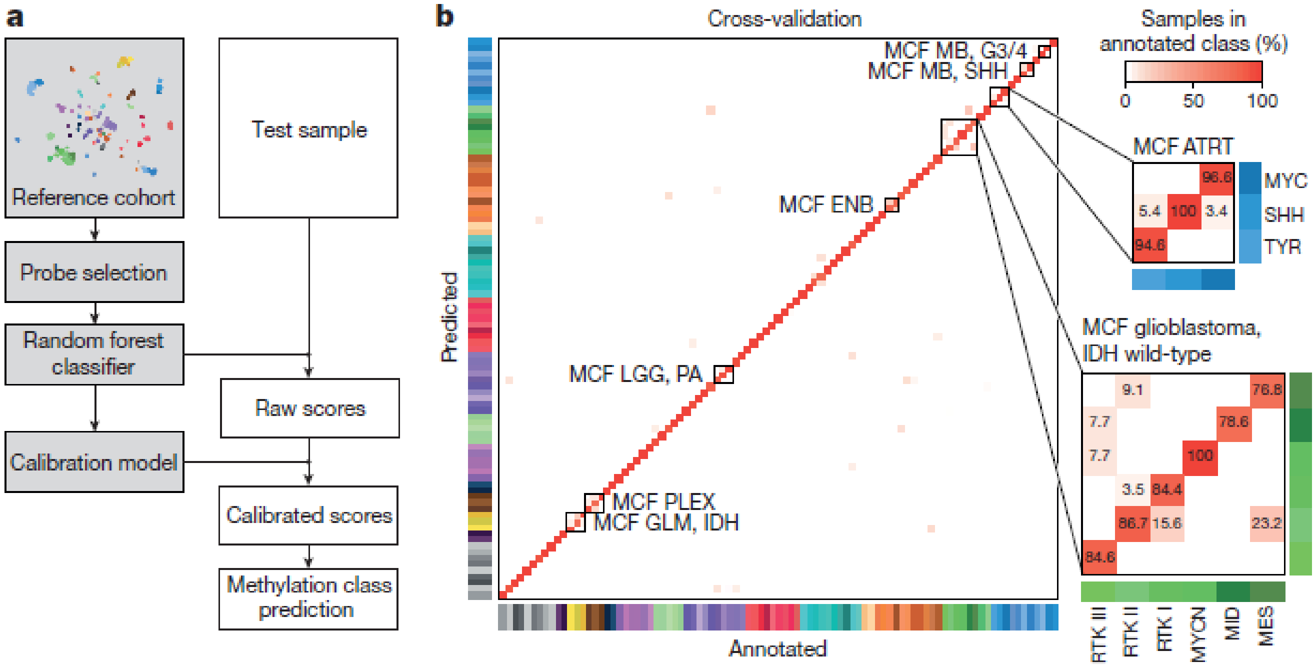
ARTICLE

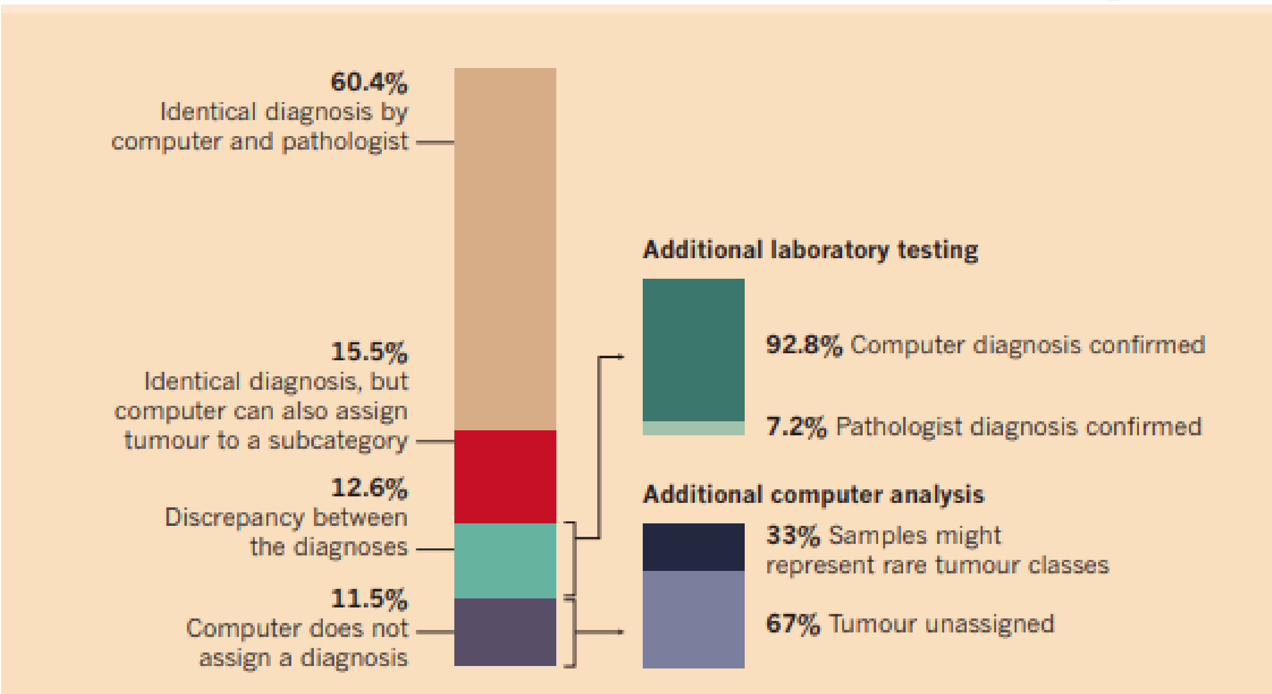
doi:10.1038/nature26000

DNA methylation-based classification of central nervous system tumours

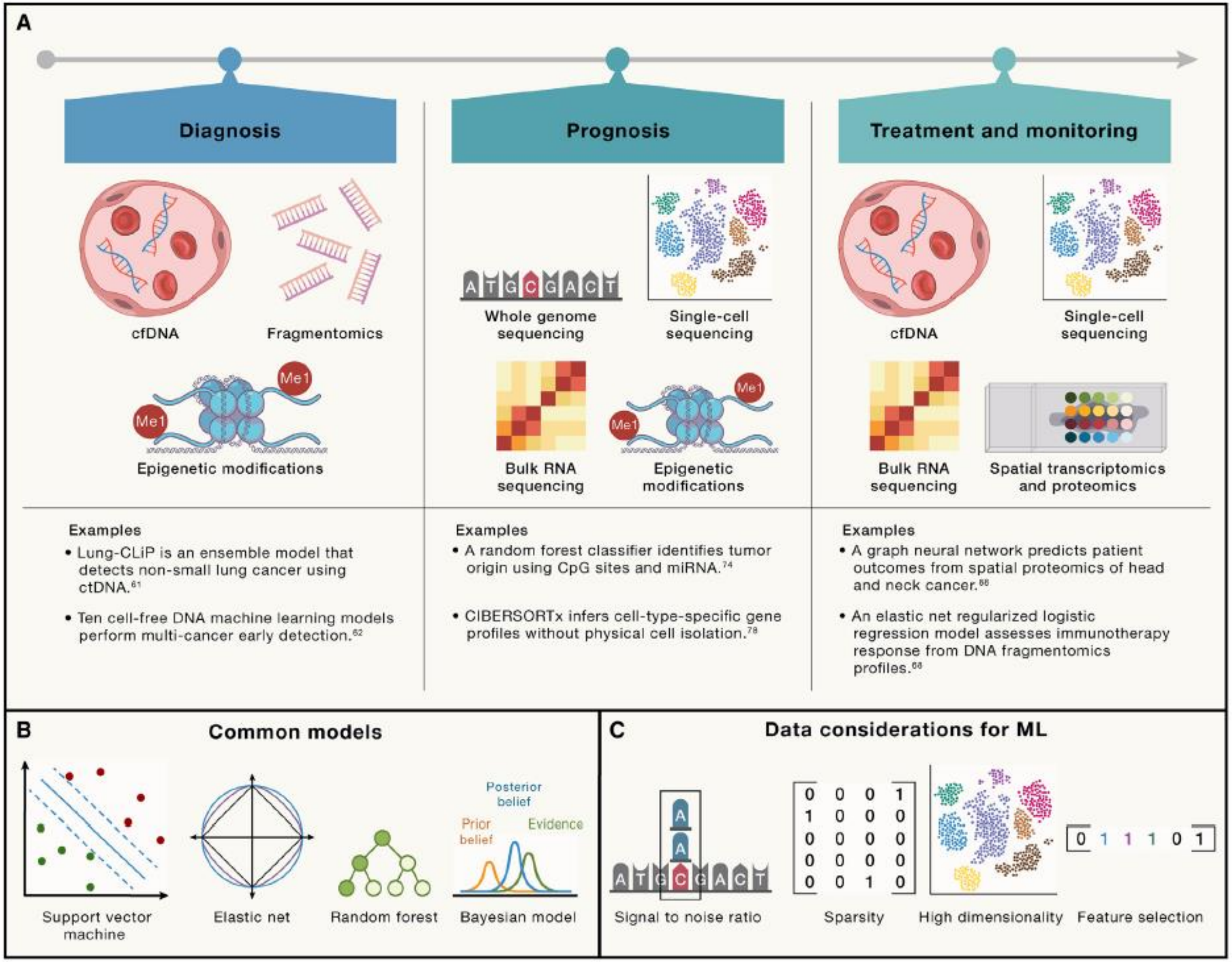


Capper et al., Nature 2018



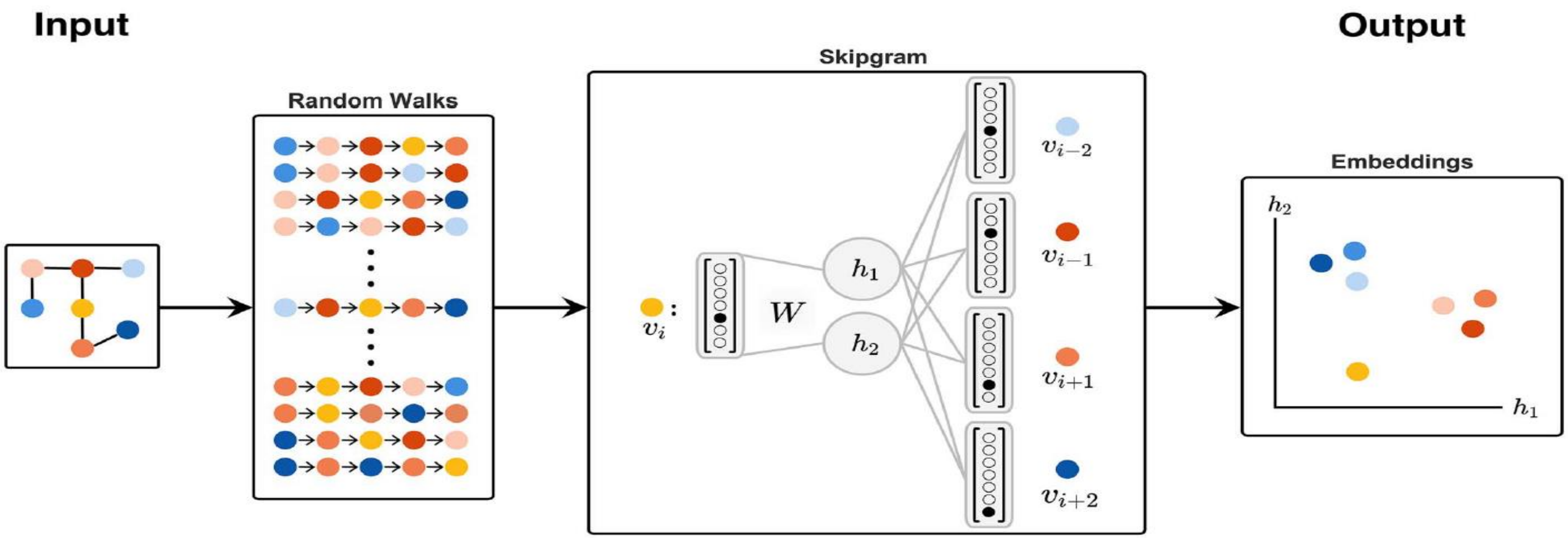


Capper et al., Nature 2018



Swanson et al., Cell 2023

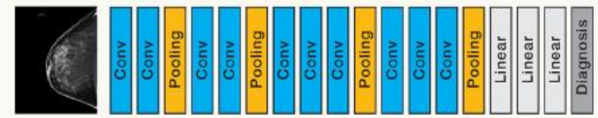
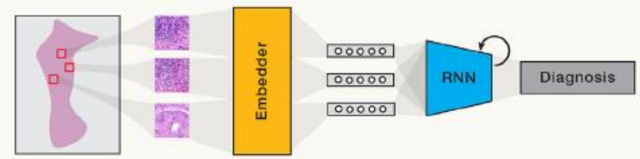
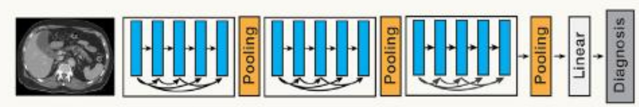
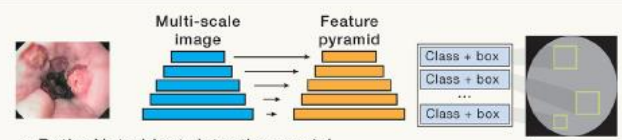
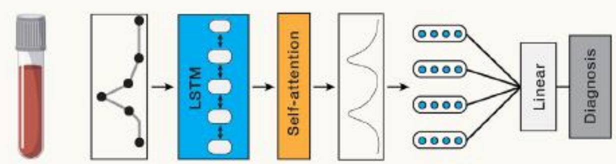
Graph neural networks (GNN)



Prediction of:

- Protein interactions
- Protein function
- Drug target
- Drug-drug interaction
- Metabolic networks
- Gene regulatory networks…

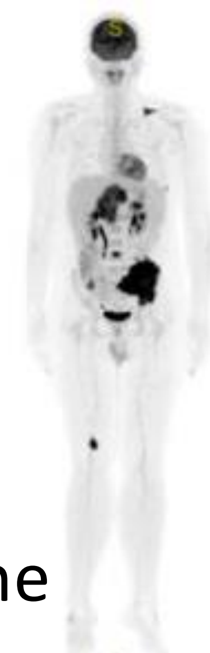
Some AI tools with regulatory approval

Device description	Clinical study details	Model diagram and description (Key: Feature processing; Feature aggregation; Linear layers; Prediction)
Transpara • Breast cancer mammography detection algorithm • FDA approved, 2020	• AI-assisted and standalone studies ◦ 18 readers ◦ 240 exams	 • RetinaNet object detection model • Outputs image and lesion scores
Paige Prostate • Prostate pathology cancer diagnostic algorithm • FDA approved, 2019	• Standalone analytical testing ◦ 847 slides • AI-assisted study ◦ 527 slides ◦ 16 pathologists	 • ResNet-34 CNN feature extractor • RNN for score prediction • Multiple instance learning
Optellum • Lung CT cancer nodule detection algorithm • FDA approved, 2021	• AI-assisted and standalone studies ◦ 300 subjects ◦ 12 readers	 • DenseNet CNN classifier
GI Genius • Lesion detection for endoscopy video • FDA approved, 2021	• Standalone study ◦ 150 videos ◦ 338 lesions	 • RetinaNet object detection model • Video frames are individually processed
InterVenn GLORI • Lab developed test for ovarian cancer diagnosis • CLIA certified, 2021	• Prospective observational study ◦ 1,200 participants	 • LSTM model for signal processing • Regression model for score prediction

Swanson et al., Cell 2023

Use case: AI-based classification of cancers of unknown primary

Clinical case: patient n°1



- 30-year old male patient
- Rapid degradation of health status
- Abdominal mass, diffuse bone lesion
- Addressed to Institut Curie for suspicion of bone sarcoma
- **Unclassified carcinoma** according to pathologists
- Proposed treatment: unspecific cytotoxic chemotherapy

Clinical case: patient n°1

- 30-year old male patient
- Rapid degradation of health status
- Abdominal mass, diffuse bone lesion
- Addressed to Institut Curie for suspicion of bone sarcoma
- **Unclassified carcinoma** according to pathologists
- Proposed treatment: unspecific cytotoxic chemotherapy

RNA-seq available in Curie for diagnosis of sarcomas

Cancers of unknown primary

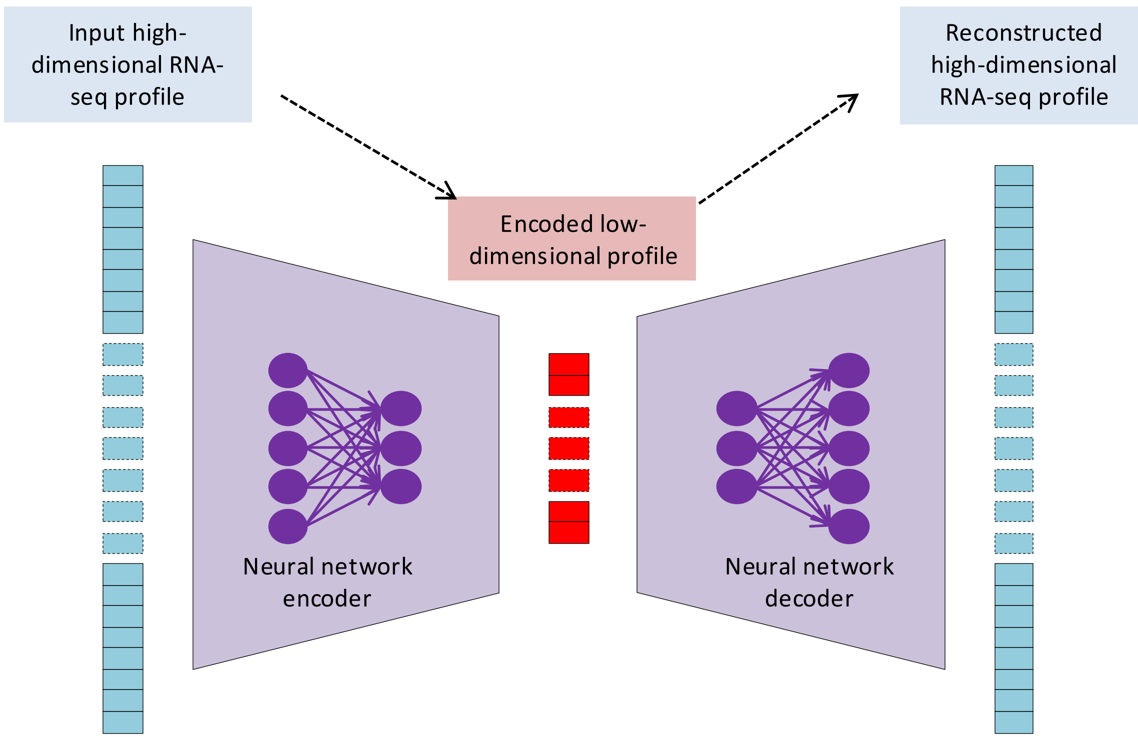
- **Cancers of unknown primary (CUP)** : 2-3 % of metastatic cancers
- The primitive tumor has not been found despite extensive explorations (radiology, pathology…)

Cancers of unknown primary

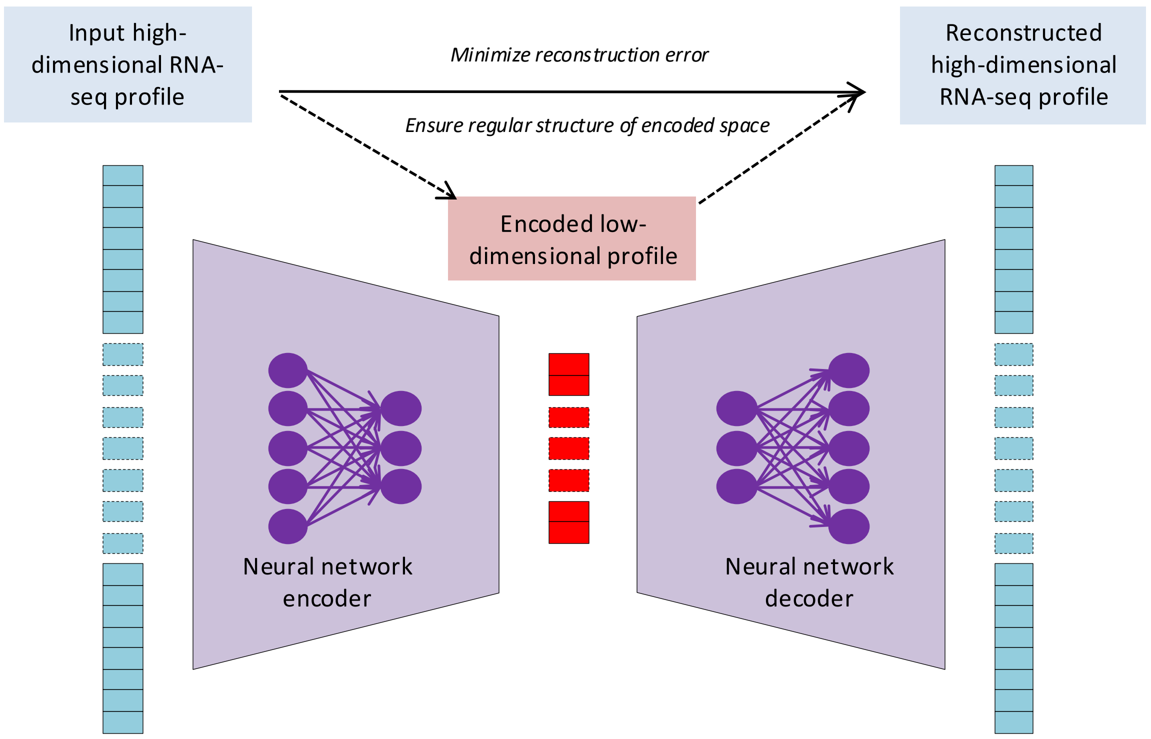
- **Cancers of unknown primary (CUP)** : 2-3 % of metastatic cancers
- The primitive tumor has not been found despite extensive explorations (radiology, pathology…)

-> *Could we identify the tissue of origin based on the transcriptomic profile (> 50,000 features)?*

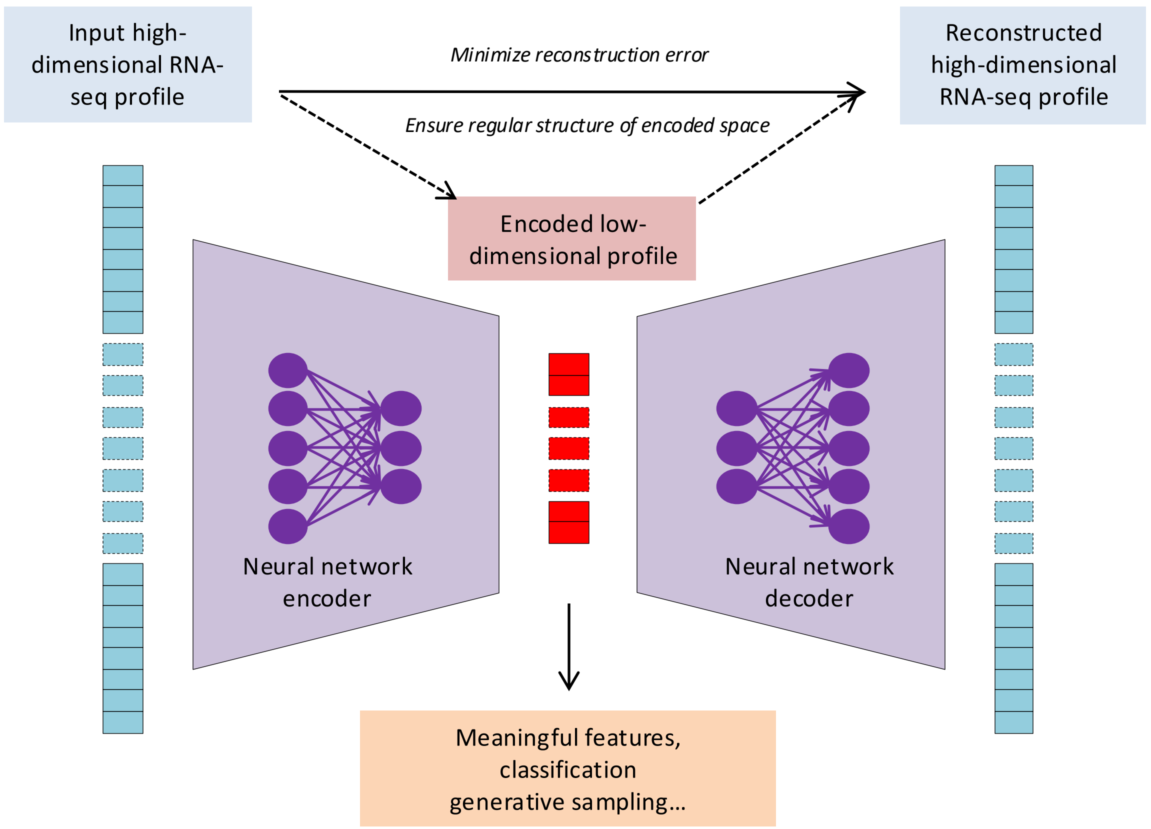
A deep learning algorithm: the variational autoencoder (VAE)



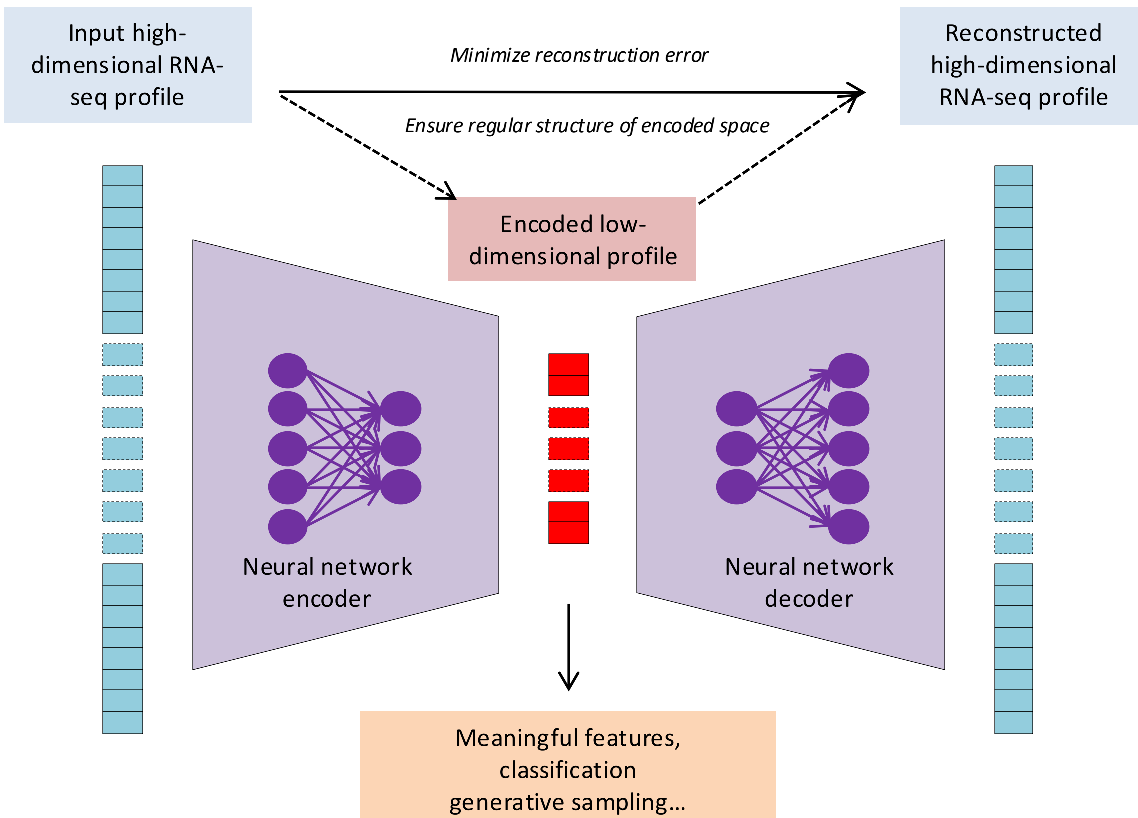
A deep learning algorithm: the variational autoencoder (VAE)



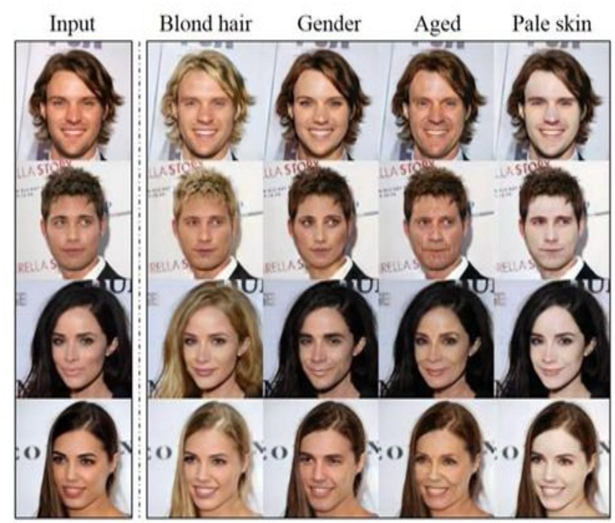
A deep learning algorithm: the variational autoencoder (VAE)



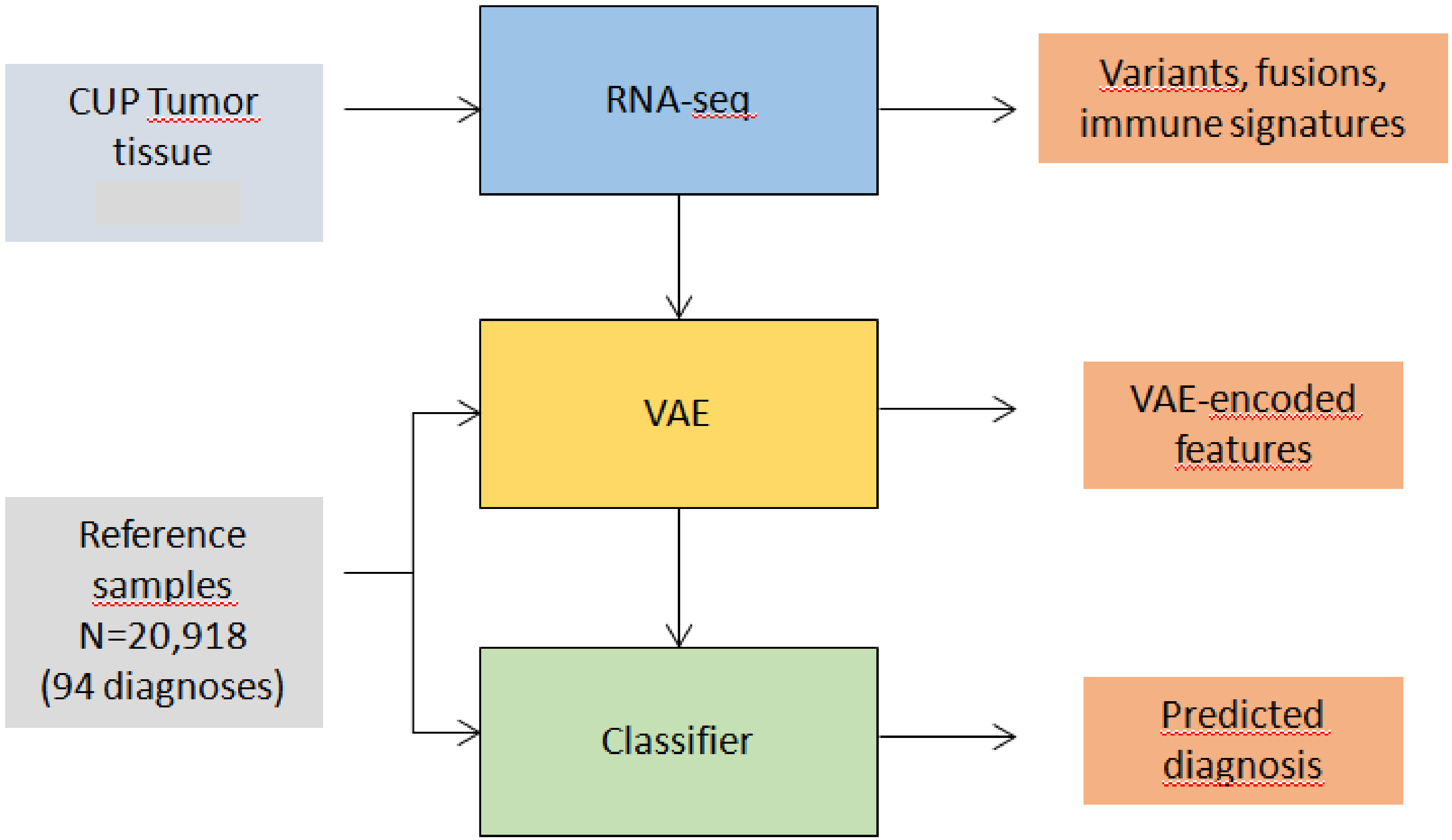
A deep learning algorithm: the variational autoencoder (VAE)



For images



Workflow for classification of CUPs using RNA-seq

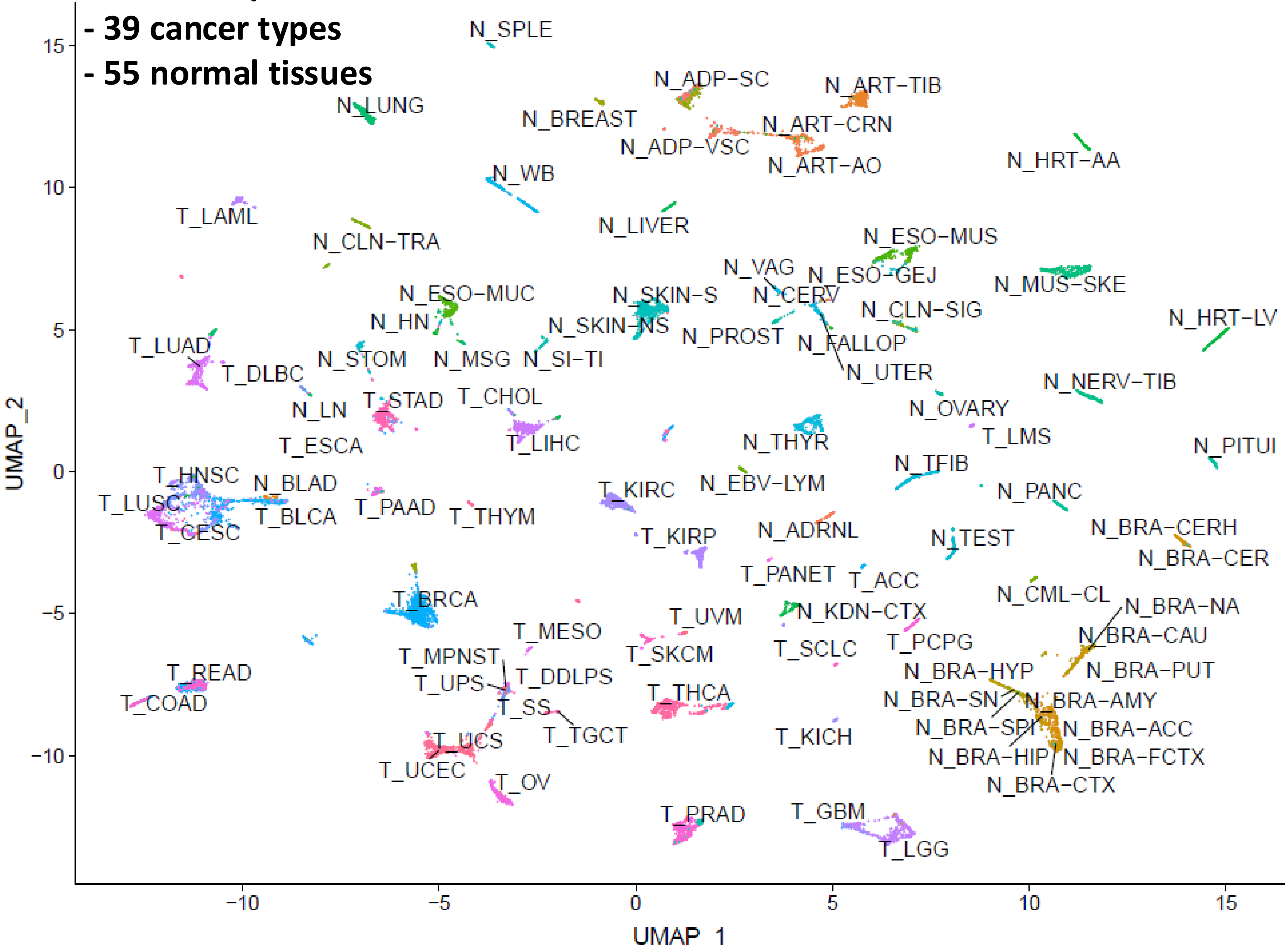


UMAP (2D representation) of all reference samples

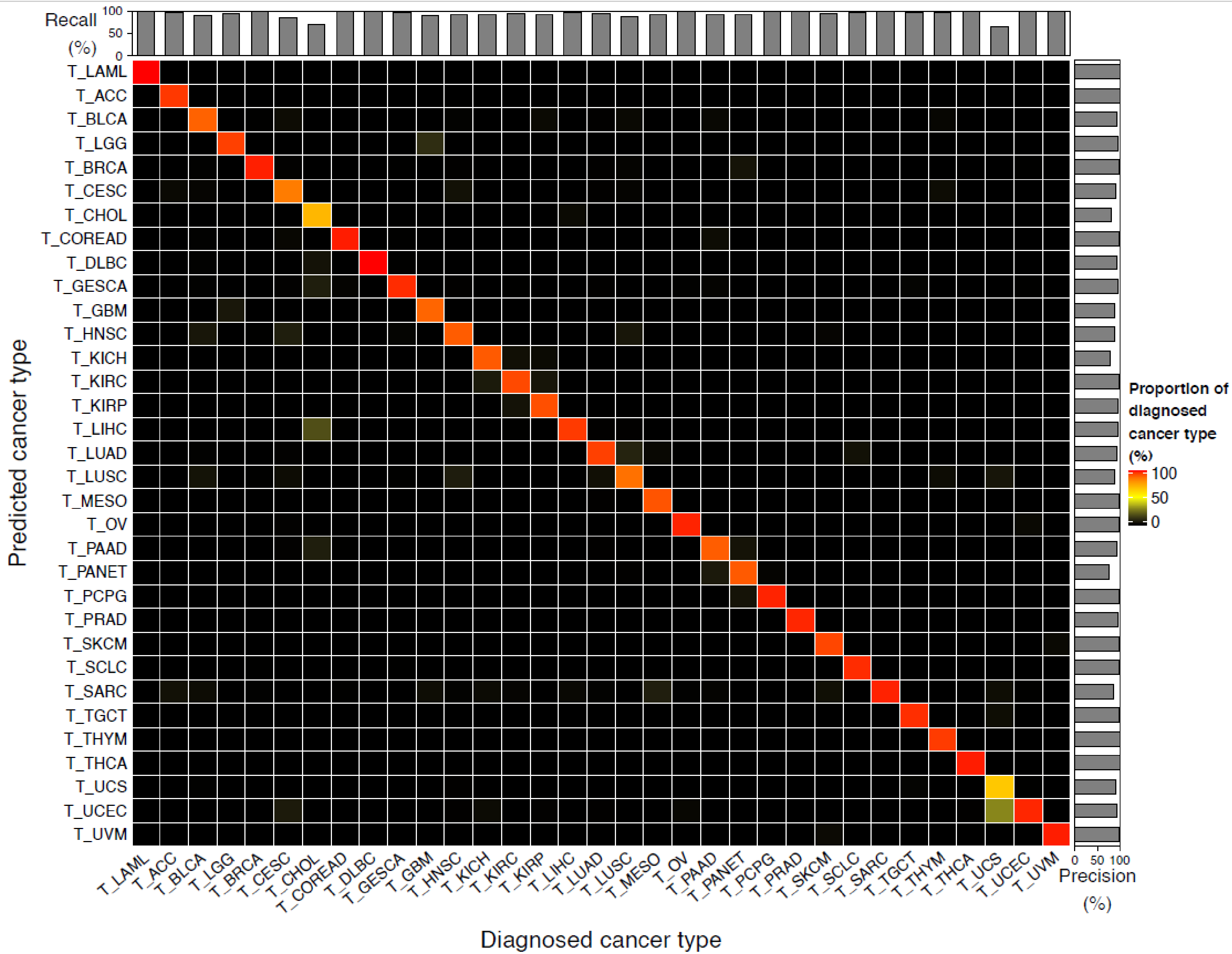
20918 samples

- 39 cancer types

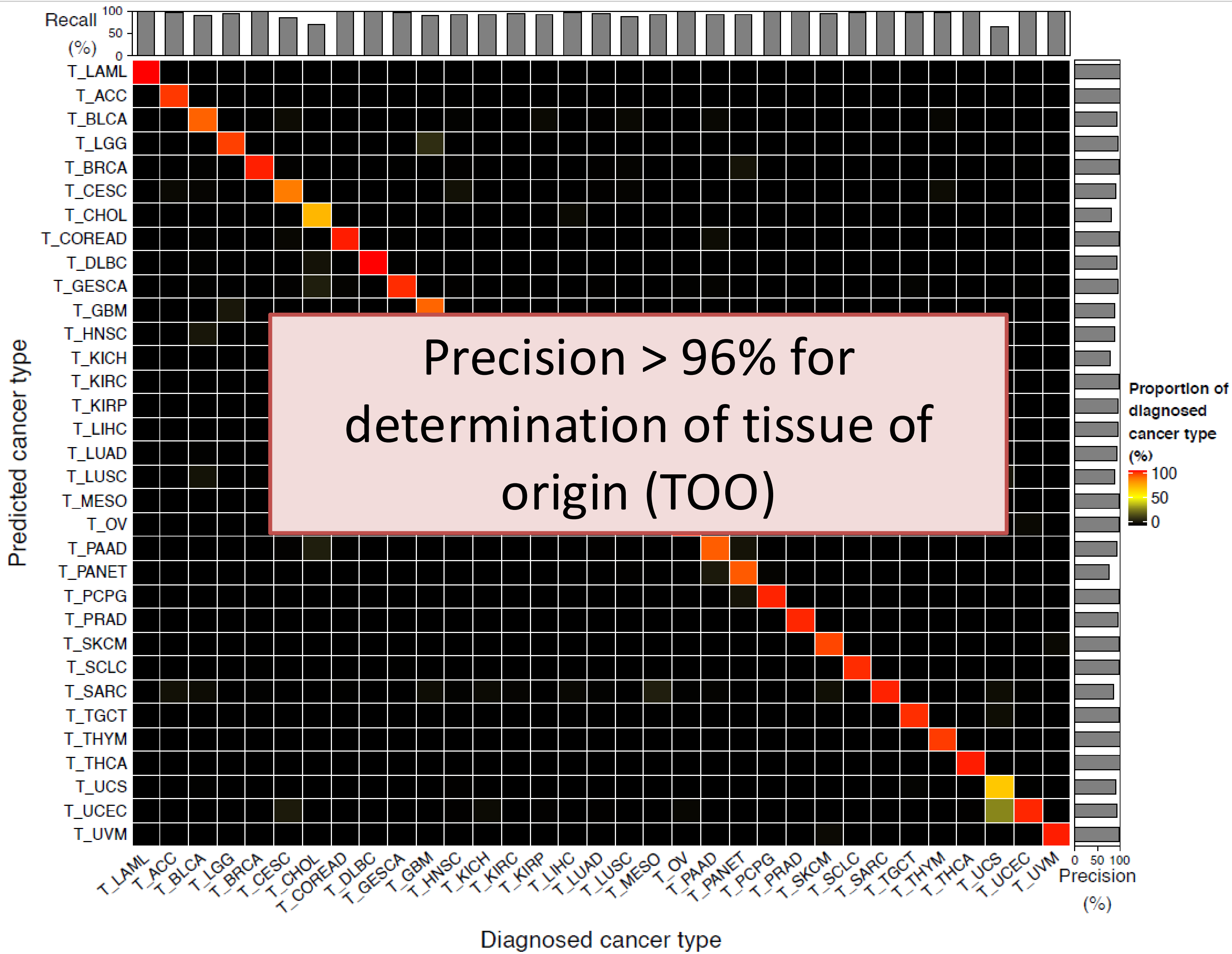
- 55 normal tissues



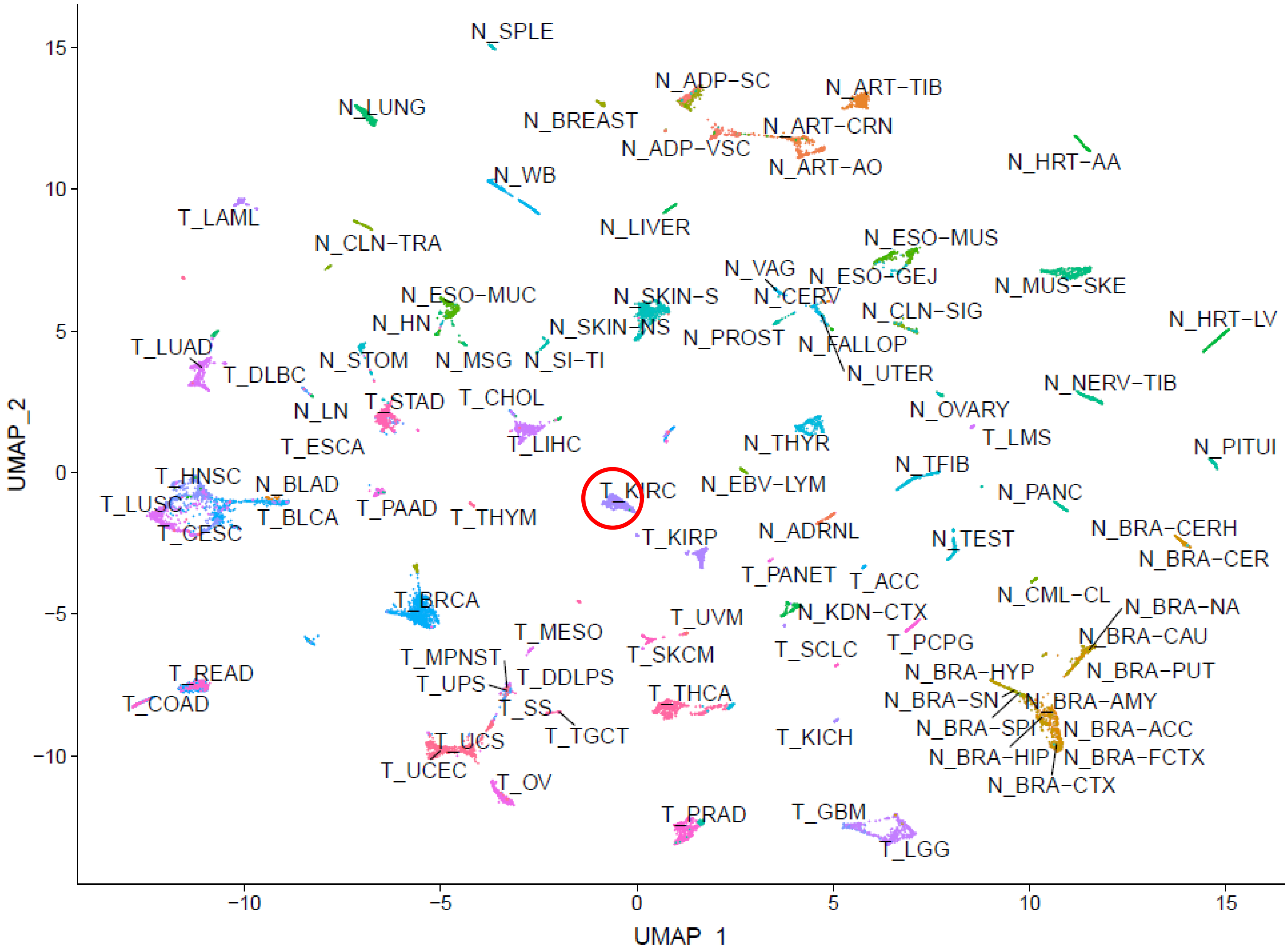
Cross-validation confusion matrix for reference samples



Cross-validation confusion matrix for reference samples



Patient n°1



N_SPLE
N_ADP-SC
N_ART-TIB
N_LUNG
N_BREAST
N_ART-CRN
N_ADP-VSC
N_ART-AO
N_HRT-AA
N_WB
T_LAML
N_LIVER
N_CLN-TRA
N_ESO-MUS
N_VAG
N_ESO-GEJ
N_MUS-SKE
N_ESO-MUC
N_SKIN-S
N_CERV
N_CLN-SIG
N_HRT-LV
N_HN
N_SKIN-NS
N_PROST
T_LUAD
N_FALLOP
N_STOM
N_MSG
N_SI-TI
T_DLBC
N_UTER
N_NERV-TIB
N_LN
T_STAD
T_CHOL
N_OVARY
T_LMS
N_THYR
T_ESCA
T_LIHC
N_PITUI
N_TFIB
T_HNSC
N_BLAD
T_KIRC
N_EBV-LYM
N_PANC
T_LUSC
T_PAAD
T_BLCA
T_THYM
T_CESC
T_KIRP
N_ADRNL
N_BRA-CERH
N_TEST
N_BRA-CER
T_PANET
T_ACC
N_CML-CL
T_BRCA
N_KDN-CTX
N_BRA-NA
T_UVM
T_MESO
N_BRA-CAU
T_PCPG
T_SKCM
T_SCLC
T_MPNST
N_BRA-PUT
T_READ
N_BRA-HYP
T_DDLPS
T_UPS
T_THCA
N_BRA-SN
N_BRA-AMY
T_COAD
T_SS
N_BRA-SPI
N_BRA-ACC
T_TGCT
T_KICH
T_UCS
N_BRA-HIP
N_BRA-FCTX
T_UCEC
N_BRA-CTX
T_OV
T_PRAD
T_GBM
T_LGG
UMAP_2
UMAP_1
15
10
5
0
−5
−10
−10
−5
0
5
10
15

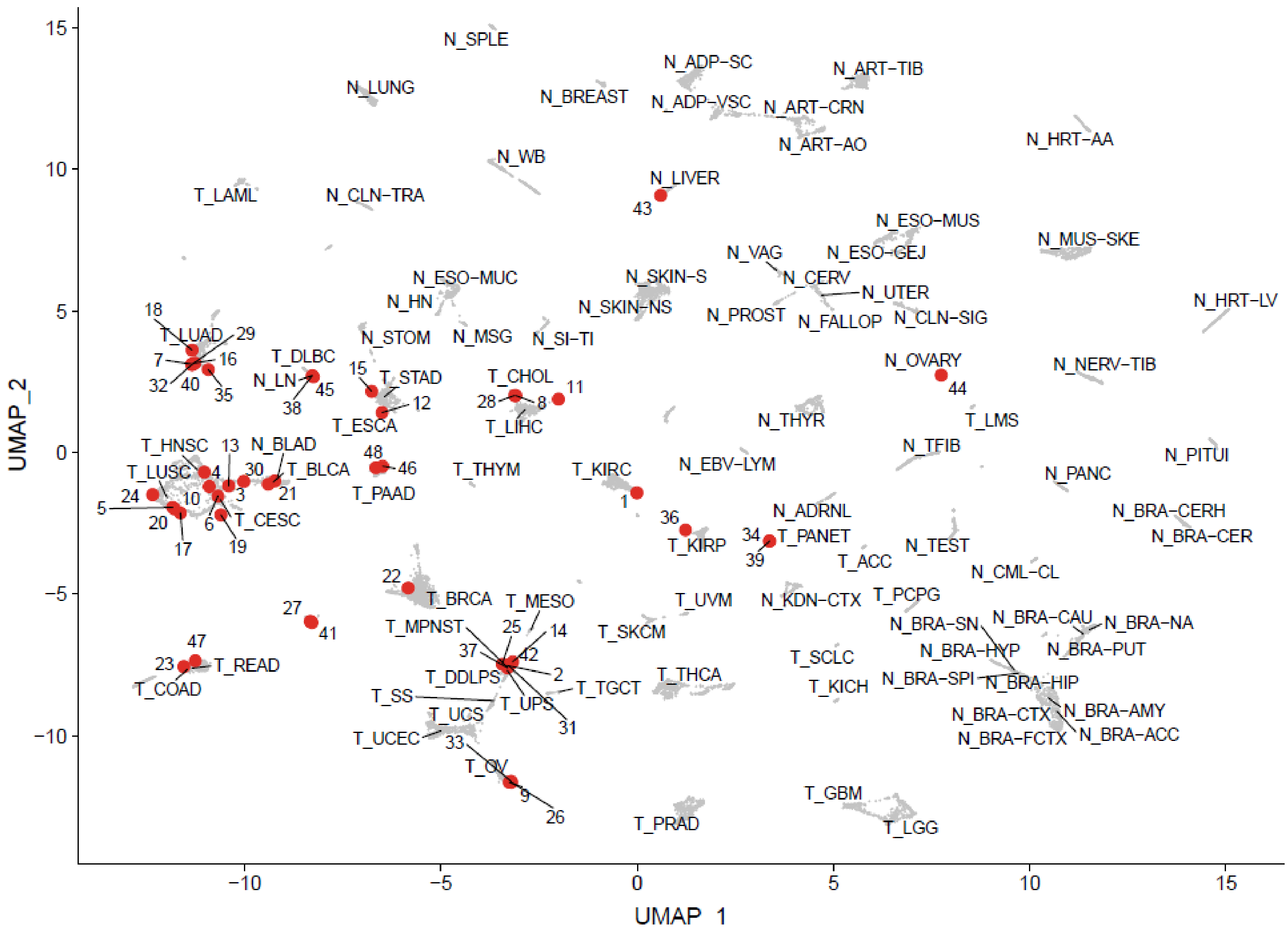
Patient n°1

- **Extrarenal renal cell carcinoma**
- Patient treated accordingly with anti-angiogenic therapy + immunotherapy

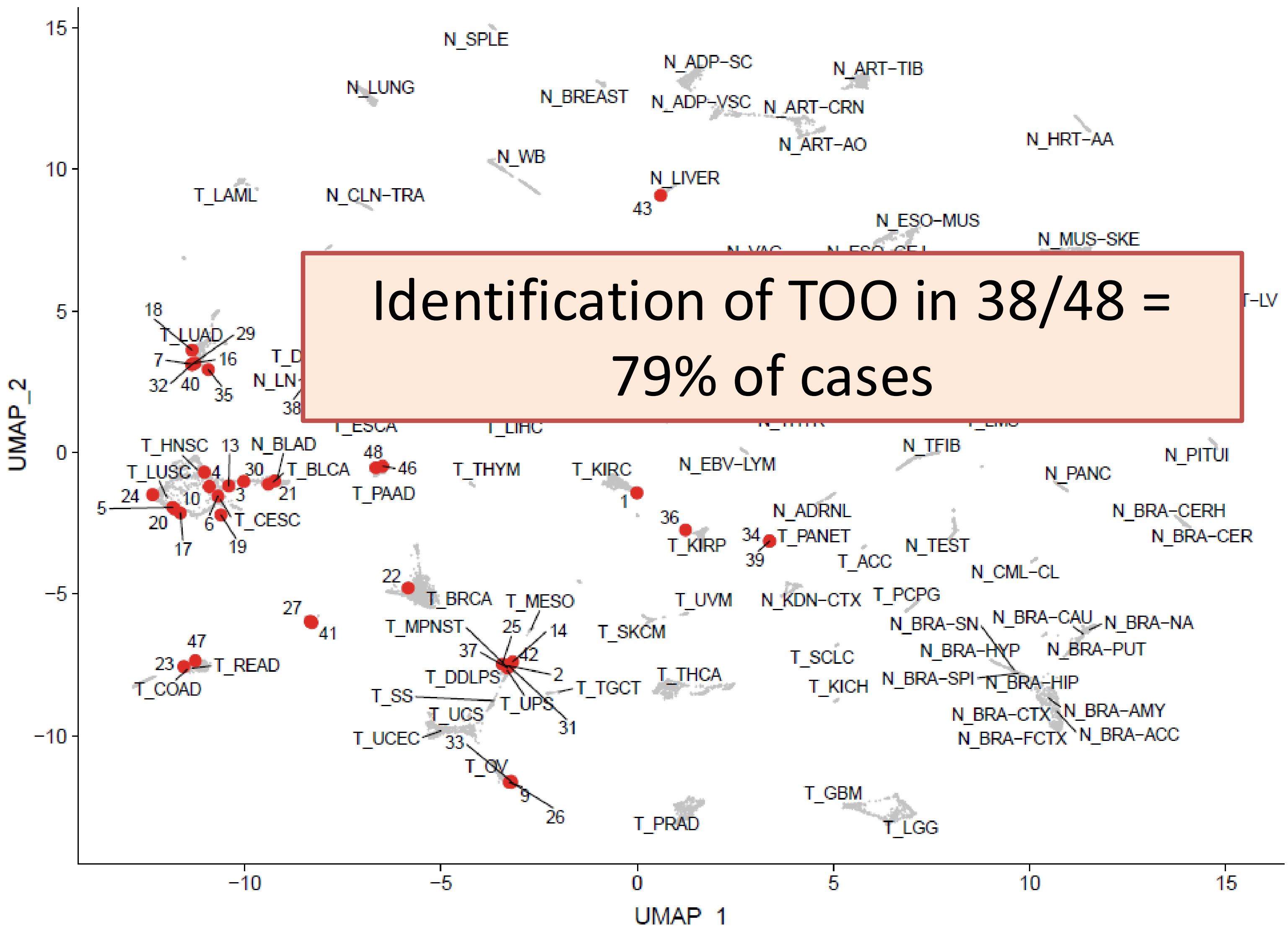
Patient n°1

- **Extrarenal renal cell carcinoma**
- Patient treated accordingly with anti-angiogenic therapy + immunotherapy
- -> *Complete response (disappearance of tumor)*

Diagnostic prediction using RNA-seq on a series of 48 CUPs

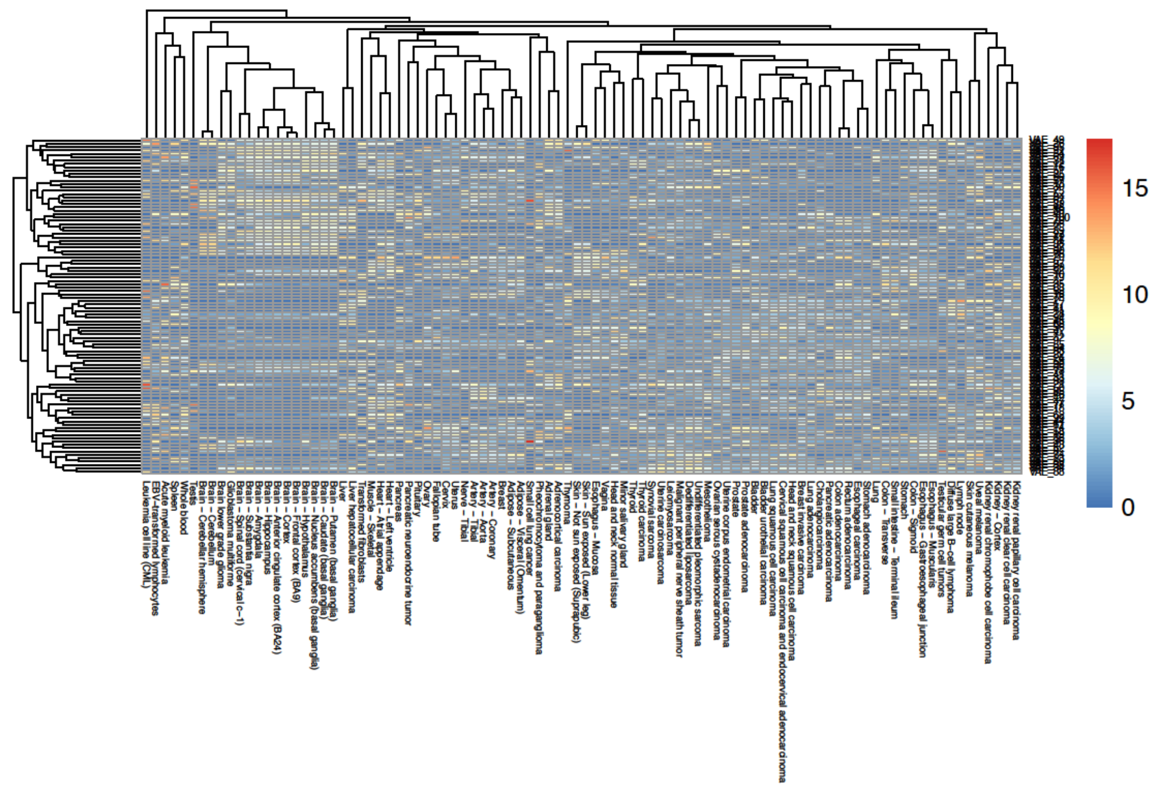


Diagnostic prediction using RNA-seq on a series of 48 CUPs



Interpretability

VAE dimensions associated with biological processes:
- Keratinization
- Neurogenesis
- Activation of immune system…



The Journal of Molecular Diagnostics, Vol. 23, No. 10, October 2021

ELSEVIER

the Journal of Molecular Diagnostics

jmdjournal.org

Identification of Tissue of Origin and Guided Therapeutic Applications in Cancers of Unknown Primary Using Deep Learning and RNA Sequencing (TransCUPtomics)

Check for updates

Julien Vibert,* Gaëlle Pierron,[†] Camille Benoist,[‡] Nadège Gruel,*[§] Delphine Guillemot,[†] Anne Vincent-Salomon,[¶] Christophe Le Tourneau,[‖] Alain Livartowski,** Odette Mariani,[¶] Sylvain Baulande,[††] François-Clément Bidard,**[‡‡] Olivier Delattre,*[†] Joshua J. Waterfall,[§,§§] and Sarah Watson*,**

From the INSERM U830, Équipe Labellisée Ligue Nationale Contre le Cancer, Diversity and Plasticity of Childhood Tumors Lab, PSL Research University, the Department of Translational Research,[§] PSL Research University, the Institut Curie Genomics of Excellence (ICGex) Platform,[††] PSL Research University, and the INSERM U830,[§§] PSL Research University, Institut Curie Research Center, Paris; the Somatic Genetics Unit,[†] Department of Genetics, the Clinical Bioinformatic Unit,[‡] Department of Diagnostic and Theranostic Medecine, the Department of Diagnostic and Theranostic Medecine,[¶] and the Department of Medical Oncology,** Institut Curie Hospital, Paris; the Department of Drug Development and Innovation,[‖] INSERM U900, Paris-Saclay University, Institut Curie Hospital and Research Center, Paris and Saint-Cloud; and the INSERM CIC-BT 1428,[‡‡] UVSQ, Paris-Saclay University, Saint-Cloud, France*

Accepted for publication July 14, 2021.

Address correspondence to Sarah Watson, M.D., Ph.D., INSERM U830, Équipe Labellisée Ligue Nationale Contre le Cancer, Diversity and Plasticity of Childhood Tumors Lab, PSL Research University, Institut Curie Research Center, Paris, France, Department of Medical Oncology, Institut Curie Hospital, Paris, France, 26 rue d'Ulm 75005 Paris, France. E-mail: sarah.watson@curie.fr.

Cancers of unknown primary (CUP) are metastatic cancers for which the primary tumor is not found despite thorough diagnostic investigations. Multiple molecular assays have been proposed to identify the tissue of origin (TOO) and inform clinical care; however, none has been able to combine accuracy, interpretability, and easy access for routine use. We developed a classifier tool based on the training of a variational autoencoder to predict tissue of origin based on RNA-sequencing data. We used as training data 20,918 samples corresponding to 94 different categories, including 39 cancer types and 55 normal tissues. The TransCUPtomics classifier was applied to a retrospective cohort of 37 CUP patients and 11 prospective patients. TransCUPtomics exhibited an overall accuracy of 96% on reference data for TOO prediction. The TOO could be identified in 38 (79%) of 48 CUP patients. Eight of 11 prospective CUP patients (73%) could receive first-line therapy guided by TransCUPtomics prediction, with responses observed in most patients. The variational autoencoder added further utility by enabling prediction interpretability, and diagnostic predictions could be matched to detection of gene fusions and expressed variants. TransCUPtomics confidently predicted TOO for CUP and enabled tailored treatments leading to significant clinical responses. The interpretability of our approach is a powerful addition to improve the management of CUP patients. *(J Mol Diagn 2021, 23: 1380–1392; https://doi.org/10.1016/j.jmoldx.2021.07.009)*

Eight out of 11 prospective CUP patients ***(73 %) received first-line treatment guided by TransCUPtomics****, with tumor responses in most patients*

The Journal of Molecular Diagnostics, Vol. 23, No. 10, October 2021

ELSEVIER

the Journal of Molecular Diagnostics

jmdjournal.org

Identification of Tissue of Origin and Guided Therapeutic Applications in Cancers of Unknown Primary Using Deep Learning and RNA Sequencing (TransCUPtomics)

Check for updates

Julien Vibert,[*] Gaëlle Pierron,[†] Camille Benoist,[‡] Nadège Gruel,[*§] Delphine Guillemot,[†] Anne Vincent-Salomon,[¶] Christophe Le Tourneau,[‖] Alain Livartowski,[**] Odette Mariani,[¶] Sylvain Baulande,[††] François-Clément Bidard,[**‡‡] Olivier Delattre,[*†] Joshua J. Waterfall,[§,§§] and Sarah Watson[*,**]

From the INSERM U830,[*] Équipe Labellisée Ligue Nationale Contre le Cancer, Diversity and Plasticity of Childhood Tumors Lab, PSL Research University, the Department of Translational Research,[§] PSL Research University, the Institut Curie Genomics of Excellence (ICGex) Platform,[††] PSL Research University, and the INSERM U830,[§§] PSL Research University, Institut Curie Research Center, Paris; the Somatic Genetics Unit,[†] Department of Genetics, the Clinical Bioinformatic Unit,[‡] Department of Diagnostic and Theranostic Medecine, the Department of Diagnostic and Theranostic Medecine,[¶] and the Department of Medical Oncology,[**] Institut Curie Hospital, Paris; the Department of Drug Development and Innovation,[‖] INSERM U900, Paris-Saclay University, Institut Curie Hospital and Research Center, Paris and Saint-Cloud; and the INSERM CIC-BT 1428,[‡‡] UVSQ, Paris-Saclay University, Saint-Cloud, France

Accepted for publication July 14, 2021.

Address correspondence to Sarah Watson, M.D., Ph.D., INSERM U830, Équipe Labellisée Ligue Nationale Contre le Cancer, Diversity and Plasticity of Childhood Tumors Lab, PSL Research University, Institut Curie Research Center, Paris, France, Department of Medical Oncology, Institut Curie Hospital, Paris, France, 26 rue d'Ulm 75005 Paris, France. E-mail: sarah.watson@curie.fr.

Cancers of unknown primary (CUP) are metastatic cancers for which the primary tumor is not found despite thorough diagnostic investigations. Multiple molecular assays have been proposed to identify the tissue of origin (TOO) and inform clinical care; however, none has been able to combine accuracy, interpretability, and easy access for routine use. We developed a classifier tool based on the training of a variational autoencoder to predict tissue of origin based on RNA-sequencing data. We used as training data 20,918 samples corresponding to 94 different categories, including 39 cancer types and 55 normal tissues. The TransCUPtomics classifier was applied to a retrospective cohort of 37 CUP patients and 11 prospective patients. TransCUPtomics exhibited an overall accuracy of 96% on reference data for TOO prediction. The TOO could be identified in 38 (79%) of 48 CUP patients. Eight of 11 prospective CUP patients (73%) could receive first-line therapy guided by TransCUPtomics prediction, with responses observed in most patients. The variational autoencoder added further utility by enabling prediction interpretability, and diagnostic predictions could be matched to detection of gene fusions and expressed variants. TransCUPtomics confidently predicted TOO for CUP and enabled tailored treatments leading to significant clinical responses. The interpretability of our approach is a powerful addition to improve the management of CUP patients. *(J Mol Diagn 2021, 23: 1380–1392; https://doi.org/10.1016/j.jmoldx.2021.07.009)*

*Eight out of 11 prospective CUP patients**(73 %) received first-line treatment guided by TransCUPtomics**, with tumor responses in most patients*

TransCUPtomics is currently deployed on **France Médecine Génomique 2025**, a national sequencing program for all CUP patients in France

Consulter le journal

Le Monde

ACTUALITÉS · ÉCONOMIE · VIDÉOS · DÉBATS · CULTURE · LE GOÛT DU MO

SCIENCES • MÉDECINE

Quand l'intelligence artificielle permet d'identifier l'origine inconnue d'un cancer métastasé

Un jeune homme de 30 ans présentant un cancer métastasé d'origine inconnue a été le premier à tester un outil d'intelligence artificielle développé à l'Institut Curie. Le crible a permis d'identifier le rein comme l'organe présentant la tumeur d'origine et le traitement spécifique qui lui a permis de guérir.

Le Parisien

Décryptage Société, Santé

Comment l'intelligence artificielle a identifié le cancer introuvable de Wilfrid

Les médecins ne trouvaient pas l'origine du cancer métastasé de ce quinquagénaire. Une IA, développée par l'Institut Curie, a estimé qu'« à 90 % » la tumeur initiale se trouvait dans les reins. L'outil est mis à l'honneur mardi dans un congrès d'oncologie aux États-Unis.

Lire le journal

Tech & web

Actualités tech · Crypto · Start-up · Tests · Pratique · Jeux vidéos

DOSSIER
Intelligence artificielle: tout ce qu'il faut savoir sur cette nouvelle révolution

Réservé aux abonnés

L'intelligence artificielle, un formidable outil pour sauver des vies

Par Pascal Grandmaison
Publié le 17/02/2023 à 06:01

Le Dr Sarah Watson, de l'Institut Curie. *Christophe LEPETIT pour Le Figaro Magazine*

DÉCRYPTAGE - En révélant des anomalies inaccessibles aux facultés humaines, l'intelligence artificielle s'impose comme l'assistant numéro 1 du médecin.

france inter

Grille des programmes · Podcasts · Info · Culture · Humour

Une intelligence artificielle devine les cancers masqués, le Parisien.

Lundi 17 avril 2023

ÉCOUTER (5 MIN)

Other AI approaches for CUP classification

Deep learning on mutational data (DNA-based)

nature COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-019-13825-8 OPEN

A deep learning system accurately classifies primary and metastatic cancers using passenger mutation patterns

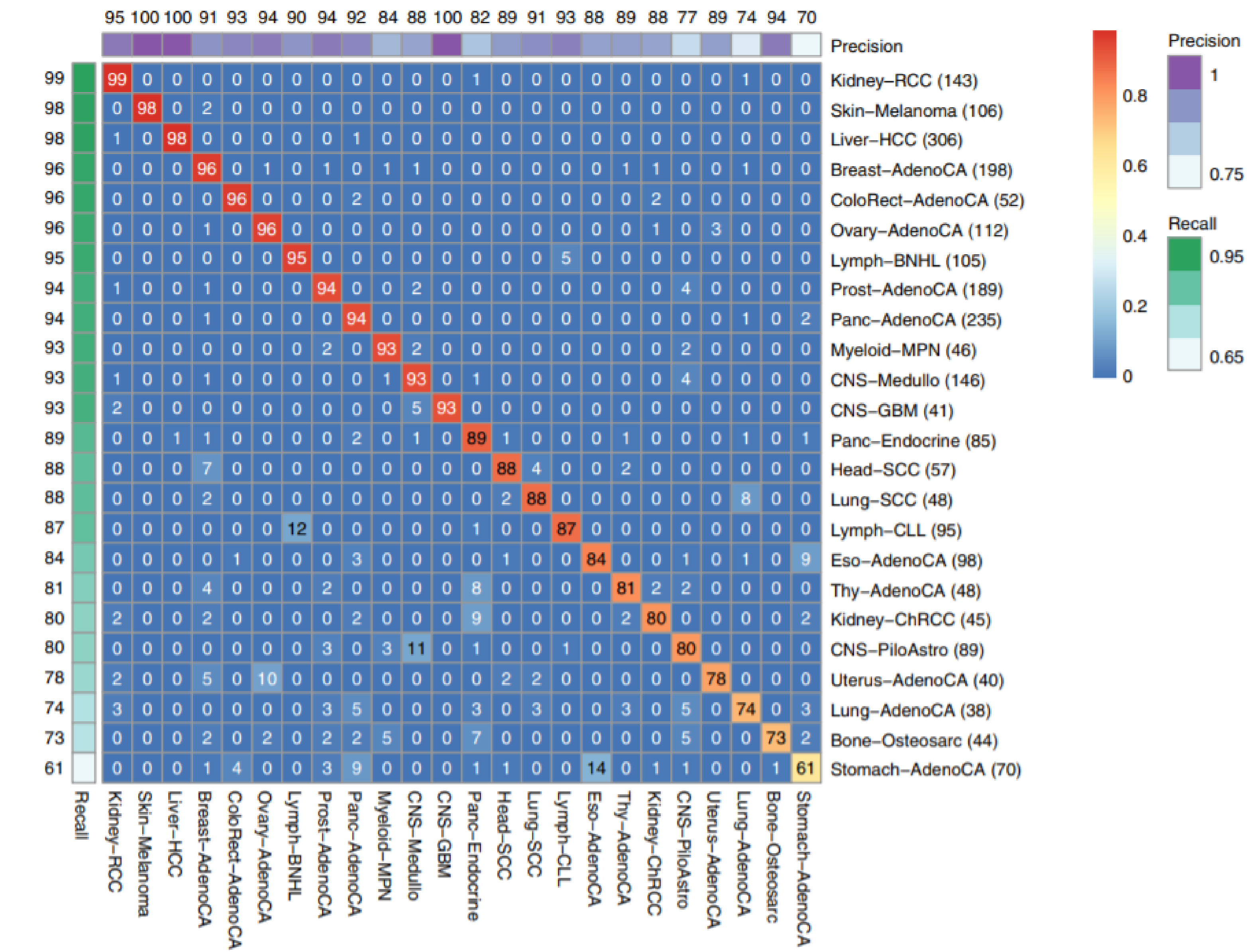
Wei Jiao[1,63], Gurnit Atwal[1,2,3,63], Paz Polak[4,63], Rosa Karlic[5], Edwin Cuppen[6,7], PCAWG Tumor Subtypes and Clinical Translation Working Group, Alexandra Danyi[8], Jeroen de Ridder[8], Carla van Herpen[9], Martijn P. Lolkema[10], Neeltje Steeghs[11], Gad Getz[12], Quaid Morris[1,2,3,13,14,64], Lincoln D. Stein[1,2,64]* & PCAWG Consortium

Methods

- Deep learning classifier
- Trained on whole-genome sequencing (WGS) data
- Uses patterns of mutations to predict tissue of origin

Table 2 WGS feature types used in classifiers.

Feature category	Feature type	Feature count	Description
Mutation distribution	SNV-BIN	2897	Number of SNVs per 1-Mbp bin, and per chromosome, normalised against the total number of SNVs per sample
	CNA-BIN	2826	Number of CNAs per 1-Mbp bin
	SV-BIN	2929	Number of SVs per 1-Mbp bin, and per chromosome, normalised against the total number of SV per sample
	INDEL-BIN	2757	Number of SNVs per 1-Mbp bin, and per chromosome, normalised against the total number of INDEL per sample
Mutation type	MUT-WGS	150	Type of single-nucleotide substitution, double- and triple-nucleotide substitution (plus its adjacent nucleotide neighbours)
Driver gene/pathway	GEN	554	Presence of an impactful mutation in a suspected driver gene
	MOD	1865	Presence of an impactful mutation in a gene belonging to a suspected driver pathway



95 100 100 91 93 94 90 94 92 84 88 100 82 89 91 93 88 89 88 77 89 74 94 70
Precision
Recall
99 Kidney–RCC (143)
98 Skin–Melanoma (106)
98 Liver–HCC (306)
96 Breast–AdenoCA (198)
96 ColoRect–AdenoCA (52)
96 Ovary–AdenoCA (112)
95 Lymph–BNHL (105)
94 Prost–AdenoCA (189)
94 Panc–AdenoCA (235)
93 Myeloid–MPN (46)
93 CNS–Medullo (146)
93 CNS–GBM (41)
89 Panc–Endocrine (85)
88 Head–SCC (57)
88 Lung–SCC (48)
87 Lymph–CLL (95)
84 Eso–AdenoCA (98)
81 Thy–AdenoCA (48)
80 Kidney–ChRCC (45)
80 CNS–PiloAstro (89)
78 Uterus–AdenoCA (40)
74 Lung–AdenoCA (38)
73 Bone–Osteosarc (44)
61 Stomach–AdenoCA (70)
Kidney–RCC
Skin–Melanoma
Liver–HCC
Breast–AdenoCA
ColoRect–AdenoCA
Ovary–AdenoCA
Lymph–BNHL
Prost–AdenoCA
Panc–AdenoCA
Myeloid–MPN
CNS–Medullo
CNS–GBM
Panc–Endocrine
Head–SCC
Lung–SCC
Lymph–CLL
Eso–AdenoCA
Thy–AdenoCA
Kidney–ChRCC
CNS–PiloAstro
Uterus–AdenoCA
Lung–AdenoCA
Bone–Osteosarc
Stomach–AdenoCA
0.8
0.6
0.4
0.2
0
Precision
1
0.75
Recall
0.95
0.65

Deep learning on pathology slides

Article

AI-based pathology predicts origins for cancers of unknown primary

https://doi.org/10.1038/s41586-021-03512-4

Received: 27 June 2020

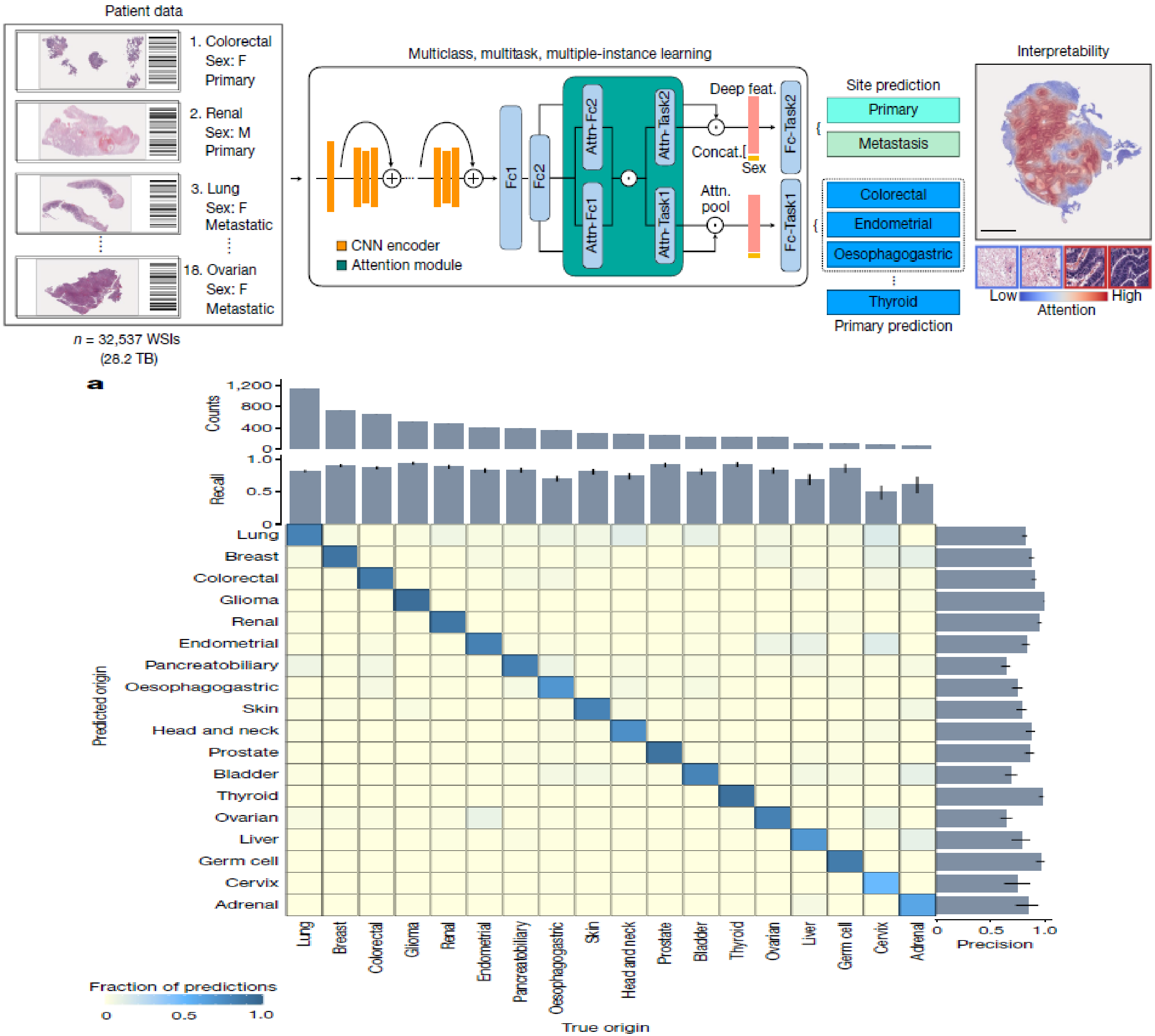
Accepted: 1 April 2021

Published online: 5 May 2021

Check for updates

Ming Y. Lu[1,2,3], Tiffany Y. Chen[1,2,5], Drew F. K. Williamson[1,2,5], Melissa Zhao[1], Maha Shady[1,2,3,4], Jana Lipkova[1,2,3] & Faisal Mahmood[1,2,3]✉

Cancer of unknown primary (CUP) origin is an enigmatic group of diagnoses in which the primary anatomical site of tumour origin cannot be determined[1,2]. This poses a considerable challenge, as modern therapeutics are predominantly specific to the primary tumour[3]. Recent research has focused on using genomics and transcriptomics to identify the origin of a tumour[4–9]. However, genomic testing is not always performed and lacks clinical penetration in low-resource settings. Here, to overcome these challenges, we present a deep-learning-based algorithm—Tumour Origin Assessment via Deep Learning (TOAD)—that can provide a differential diagnosis for the origin of the primary tumour using routinely acquired histology slides. We used whole-slide images of tumours with known primary origins to train a model that simultaneously identifies the tumour as primary or metastatic and predicts its site of origin. On our held-out test set of tumours with known primary origins, the model achieved a top-1 accuracy of 0.83 and a top-3 accuracy of 0.96, whereas on our external test set it achieved top-1 and top-3 accuracies of 0.80 and 0.93, respectively. We further curated a dataset of 317 cases of CUP for which a differential diagnosis was assigned. Our model predictions resulted in concordance for 61% of cases and a top-3 agreement of 82%. TOAD can be used as an assistive tool to assign a differential diagnosis to complicated cases of metastatic tumours and CUPs and could be used in conjunction with or in lieu of ancillary tests and extensive diagnostic work-ups to reduce the occurrence of CUP.



Patient data
1. Colorectal
Sex: F
Primary
2. Renal
Sex: M
Primary
3. Lung
Sex: F
Metastatic
18. Ovarian
Sex: F
Metastatic
n = 32,537 WSIs
(28.2 TB)
Multiclass, multitask, multiple-instance learning
CNN encoder
Attention module
Fc1
Fc2
Attn-Fc2
Attn-Fc1
Attn-Task2
Attn-Task1
Deep feat.
Concat.[
Sex
Attn.
pool
Fc-Task2
Fc-Task1
Site prediction
Primary
Metastasis
Colorectal
Endometrial
Oesophagogastric
Thyroid
Primary prediction
Interpretability
Low
High
Attention
a
Counts
Recall
Predicted origin
Lung
Breast
Colorectal
Glioma
Renal
Endometrial
Pancreatobiliary
Oesophagogastric
Skin
Head and neck
Prostate
Bladder
Thyroid
Ovarian
Liver
Germ cell
Cervix
Adrenal
True origin
Precision
Fraction of predictions
0
0.5
1.0
1,200
800
400

XGBoost on targeted NGS panel

nature medicine

Article https://doi.org/10.1038/s41591-023-02482-6

Machine learning for genetics-based classification and treatment response prediction in cancer of unknown primary

Received: 6 January 2023

Accepted: 30 June 2023

Published online: 7 August 2023

Check for updates

Intae Moon[1,2], Jaclyn LoPiccolo[3], Sylvan C. Baca[3,4], Lynette M. Sholl[5], Kenneth L. Kehl[2], Michael J. Hassett[2], David Liu[2,3,6], Deborah Schrag[7] & Alexander Gusev[2,6,8]

Cancer of unknown primary (CUP) is a type of cancer that cannot be traced back to its primary site and accounts for 3–5% of all cancers. Established targeted therapies are lacking for CUP, leading to generally poor outcomes. We developed OncoNPC, a machine-learning classifier trained on targeted next-generation sequencing (NGS) data from 36,445 tumors across 22 cancer types from three institutions. Oncology NGS-based primary cancer-type classifier (OncoNPC) achieved a weighted F1 score of 0.942 for high confidence predictions (≥ 0.9) on held-out tumor samples, which made up 65.2% of all the held-out samples. When applied to 971 CUP tumors collected at the Dana-Farber Cancer Institute, OncoNPC predicted primary cancer types with high confidence in 41.2% of the tumors. OncoNPC also identified CUP subgroups with significantly higher polygenic germline risk for the predicted cancer types and with significantly different survival outcomes. Notably, patients with CUP who received first palliative intent treatments concordant with their OncoNPC-predicted cancers had significantly better outcomes (hazard ratio (HR) = 0.348; 95% confidence interval (CI) = 0.210–0.570; $P = 2.32 \times 10^{-5}$). Furthermore, OncoNPC enabled a 2.2-fold increase in patients with CUP who could have received genomically guided therapies. OncoNPC thus provides evidence of distinct CUP subgroups and offers the potential for clinical decision support for managing patients with CUP.

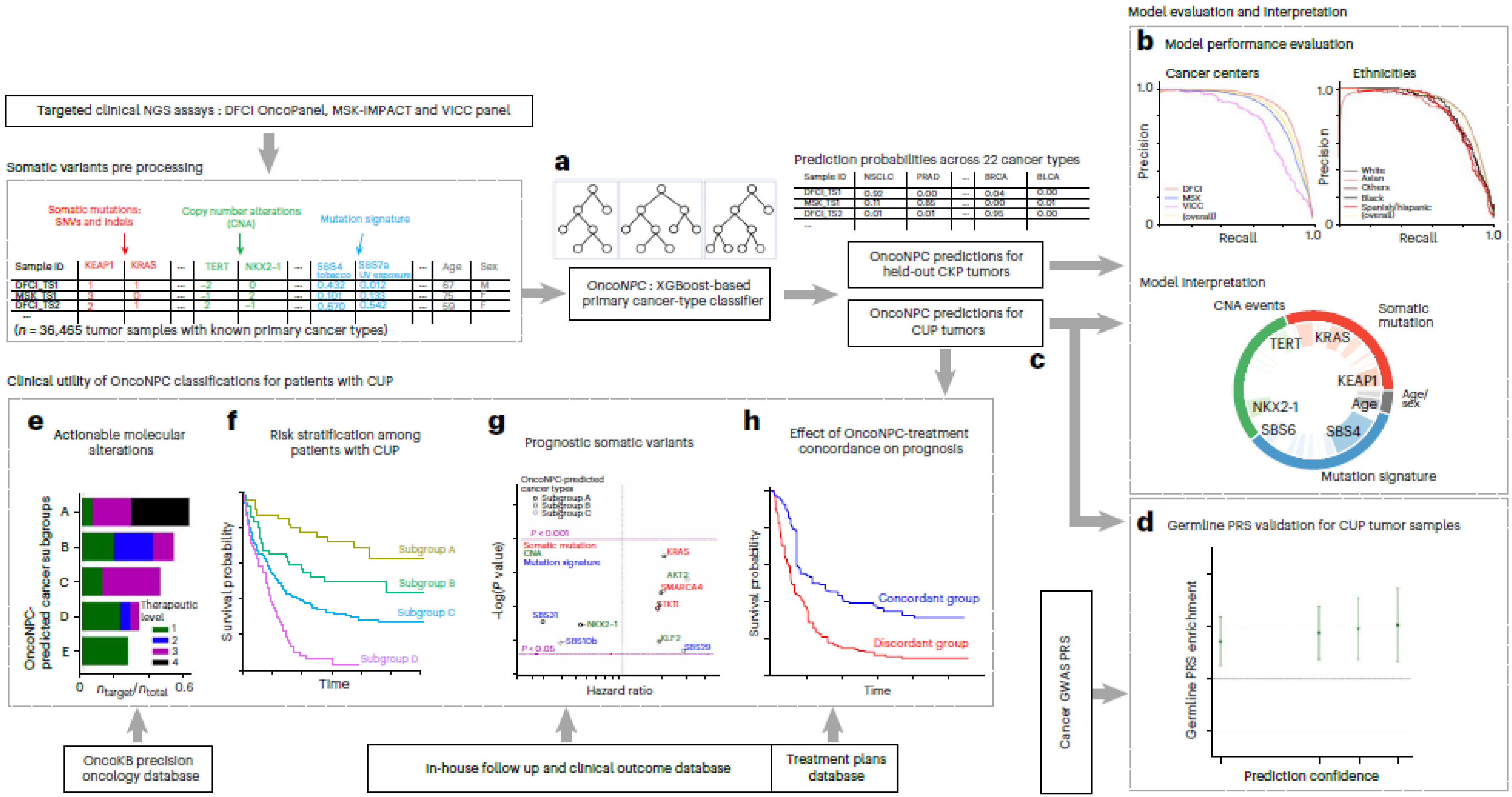


Fig. 1 | Overview of model development and analysis workflow. **a**, OncoNPC, an XGBoost-based classifier, was trained and evaluated using 36,465 cancer with known primary (CKP) tumor samples across 22 cancer types collected from three different cancer centers. **b**, OncoNPC performance was evaluated on the held-out tumor samples ($n = 7,289$). **c**, OncoNPC was applied to 971 CUP tumor samples at a single institution to predict primary cancer types. **d**–**g**, OncoNPC-predicted CUP subgroups were then investigated for association with elevated germline risk (**d**), actionable molecular alterations (**e**), overall survival (**f**) and prognostic somatic features (**g**). **h**, A subset of CUP patients with detailed treatment data was evaluated for treatment-specific outcomes.

Some comments

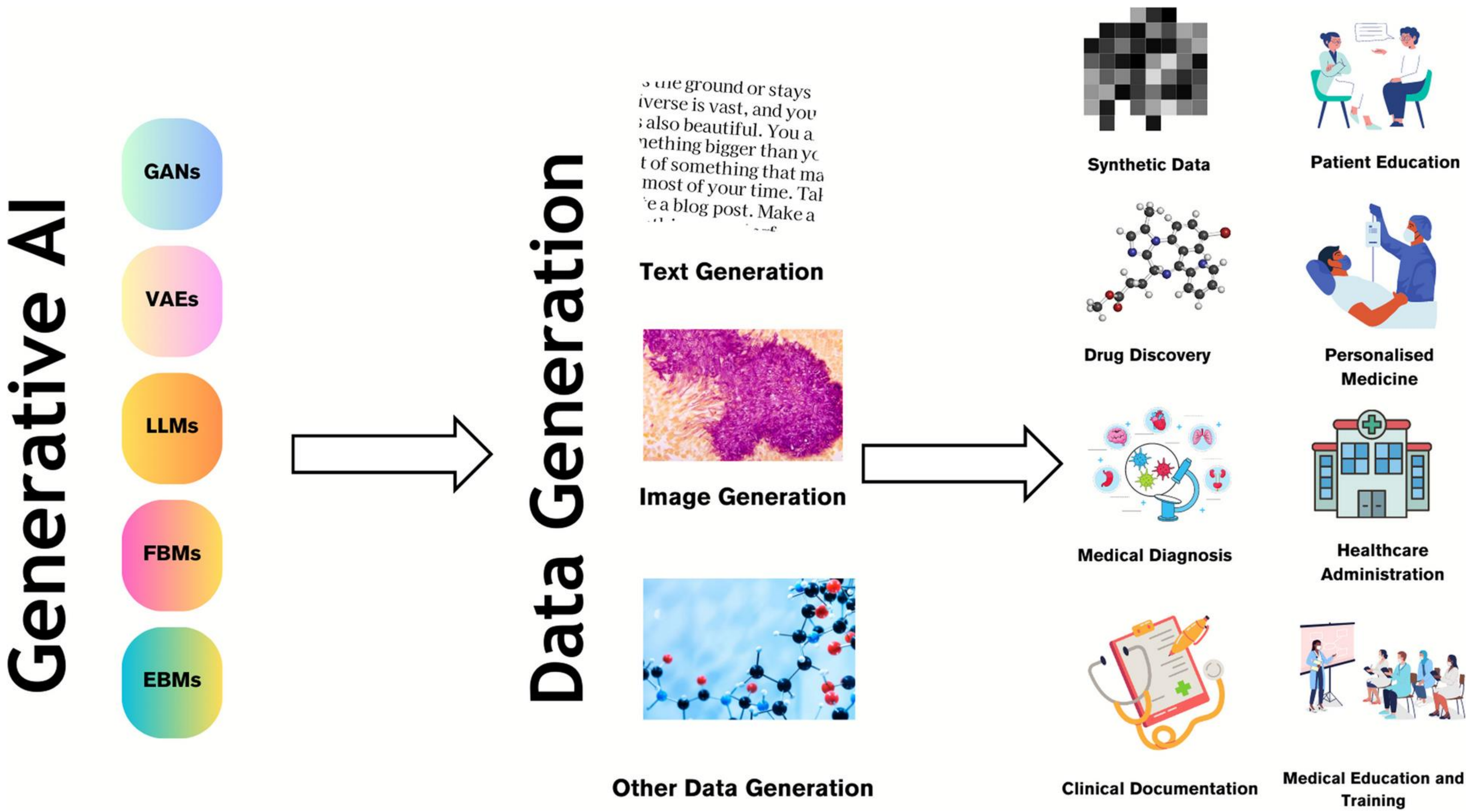
- AI can potentially be used in **every question in medicine that can be answered with data**: prediction of diagnosis, prognosis, response to treatment…
- As soon as there are **enough good data** to train an algorithm
- **Genomics and transcriptomics are well-suited to AI** (high number of features, « big data »), especially with recent techniques (**single-cell**)
- ***But may suffer from low number of samples (« fat data »)***

Conclusions and perspectives

- AI is **essential** to help the human brain in dealing with « big data »
- It has the potential to change practice in almost **all areas of medicine**
- However **AI does not replace** the doctor
- Algorithms have to be validated and show benefit in **real life**
- **Interpretability** issue

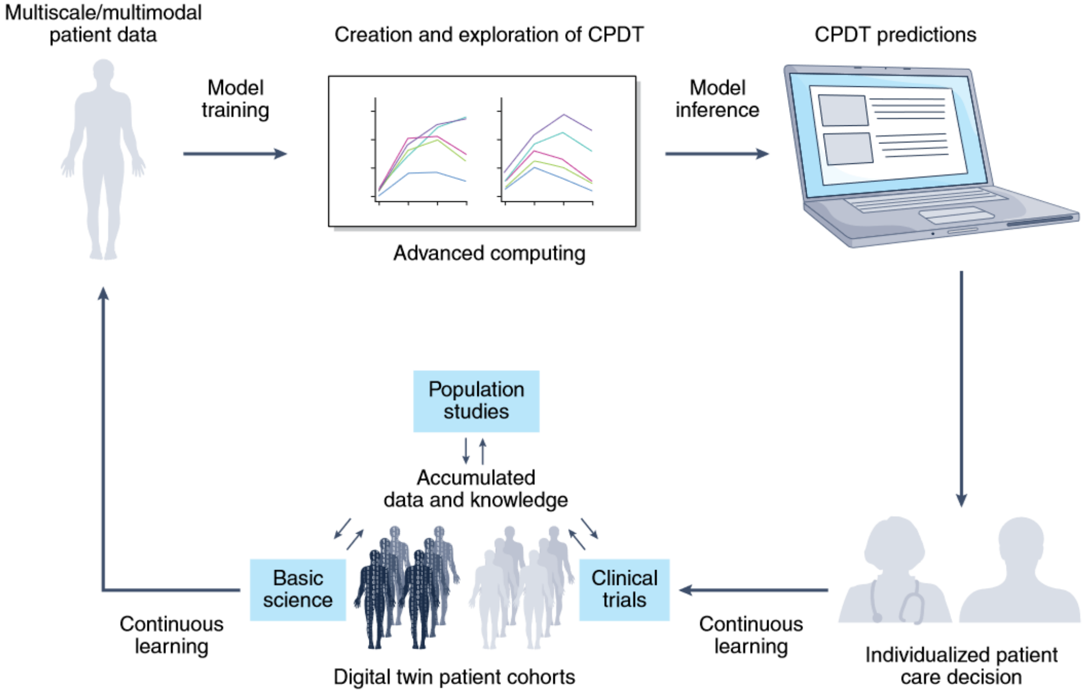
This subject is being updated every day…

Generative AI



Reddy, Impl Sci (2024)

The Cancer Patient Digital Twin (CPDT)

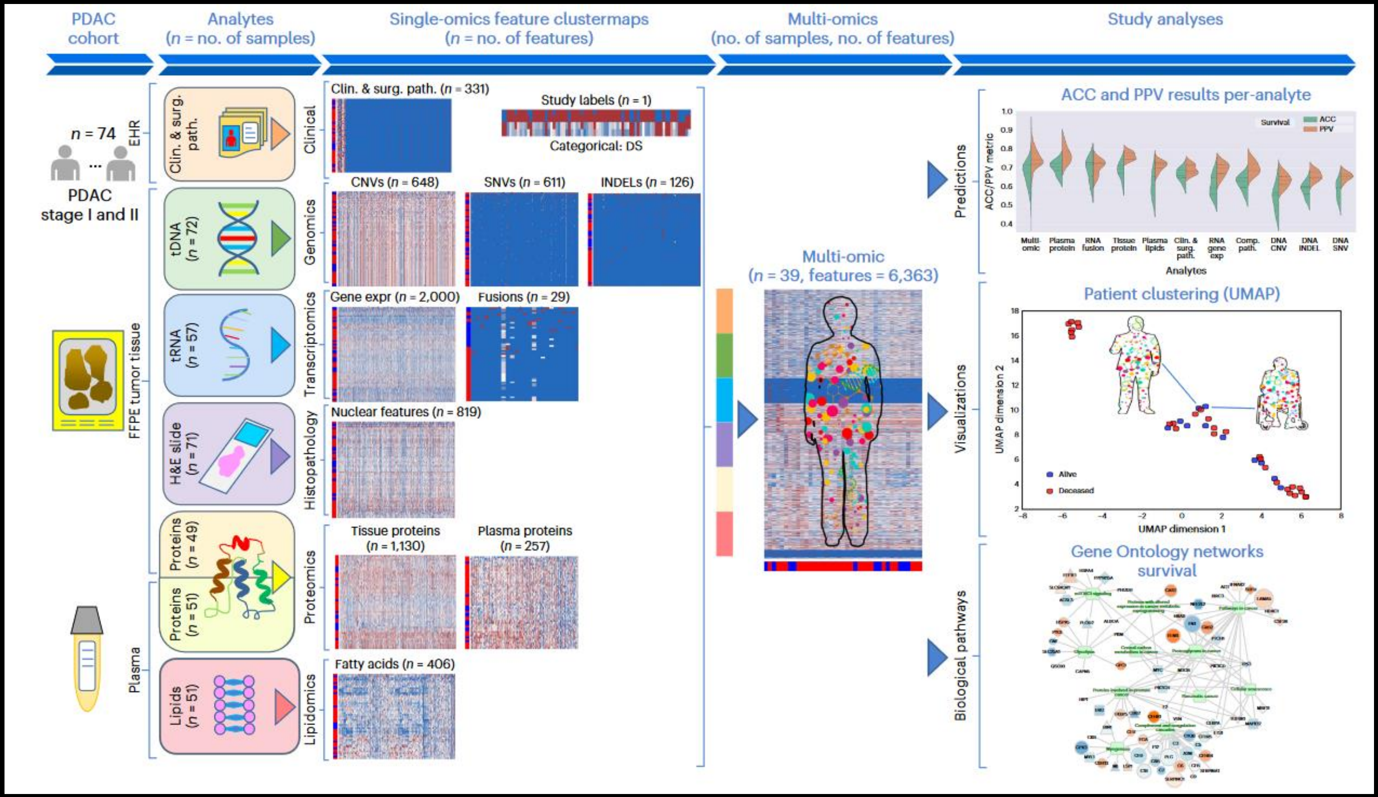


Cycling between:
- Training on patient data
- Learning from accumulated data/knowledge

-> **Individualized patient care for precision medicine**

Hernandez-Boussard et al., Nat Med (2021)

EXAMPLE 1: MOLECULAR TWIN FOR PANCREATIC CA



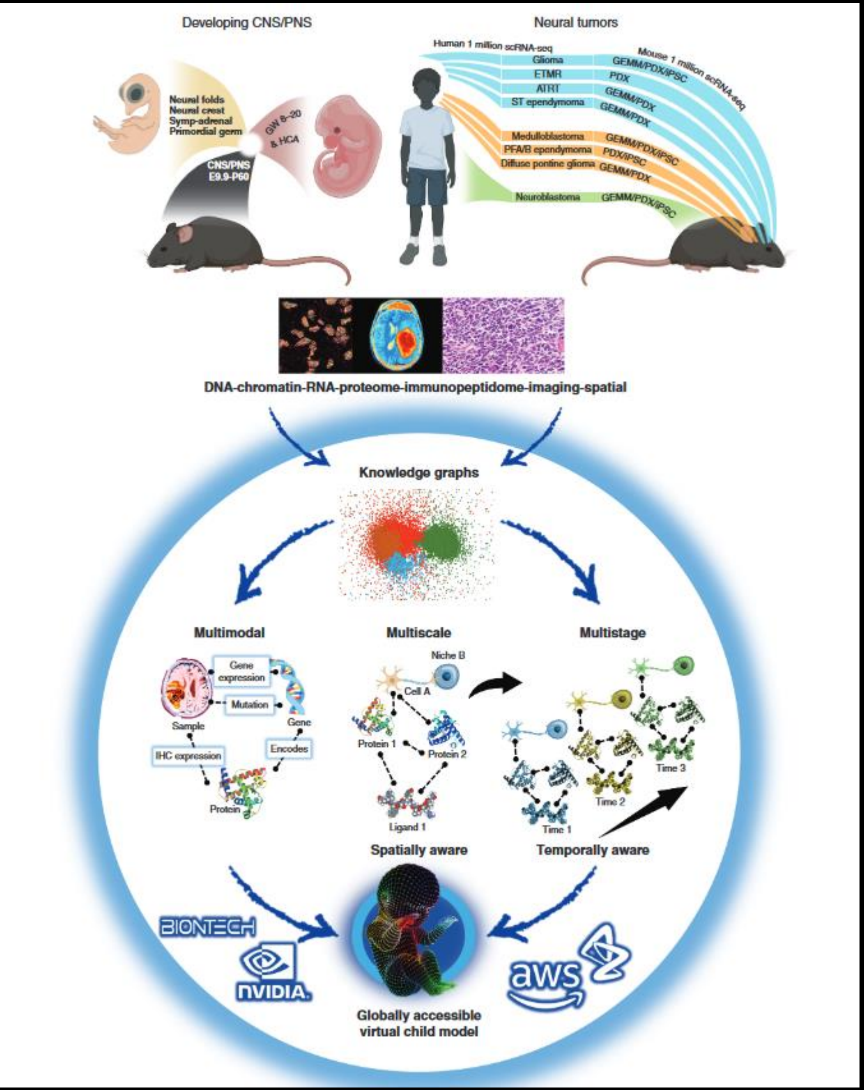
Advanced machine learning for integration of multiomic data
-> **Prediction of disease survival and biomarker discovery**

Osipov et al., Nat Cancer (2024)

© Dassault Systèmes | Confidential Information | 12/18/2025

EXAMPLE 2:
THE VIRTUAL CHILD

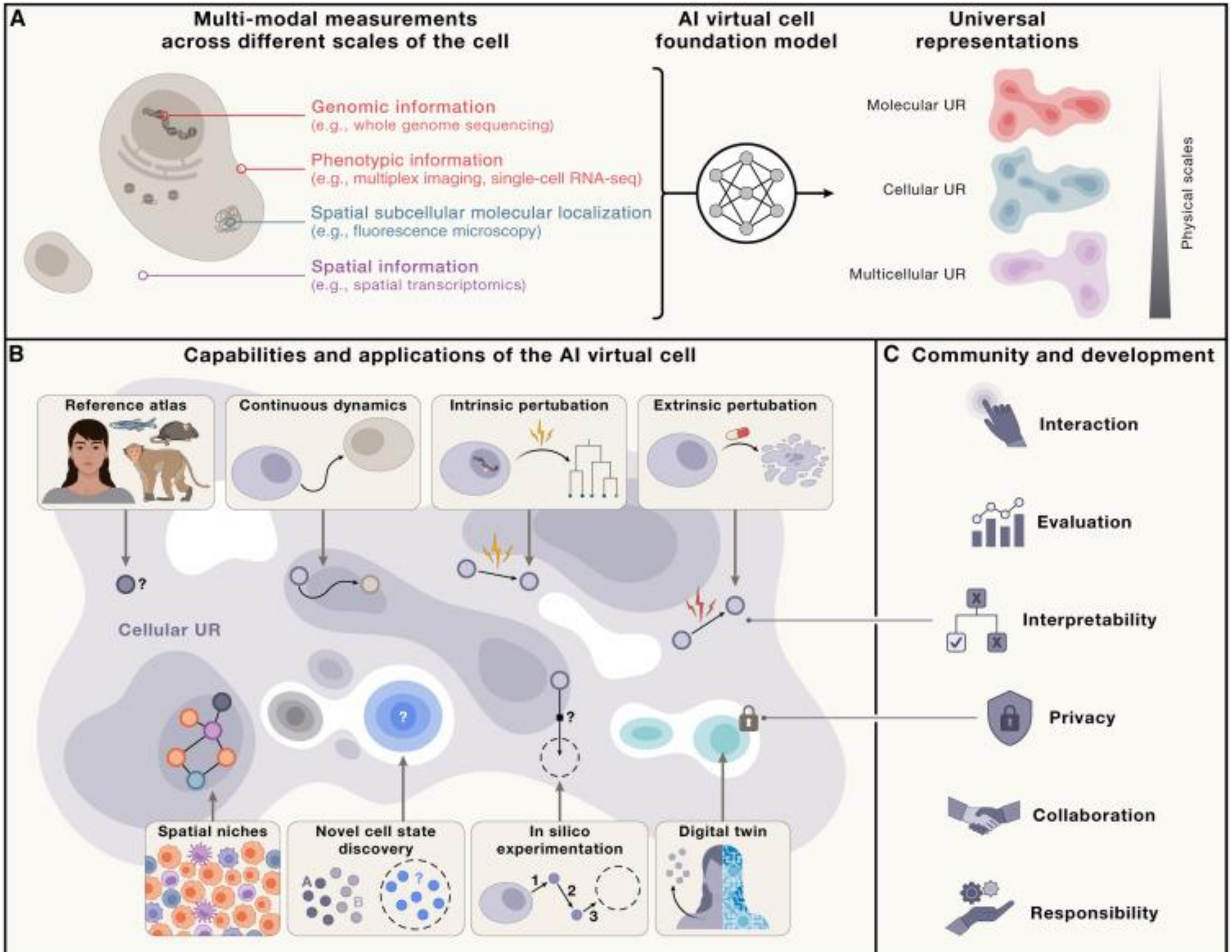
- Large-scale academic-industrial project
- Model at **the cellular level** development of pediatric tumors
- Ability to run **virtual clinical trials**



Gilbertson et al., Cancer Discov (2024)

© Dassault Systèmes | Confidential Information | 12/18/2025

The AI « Virtual Cell » (Cell 2024)



MEDITWIN

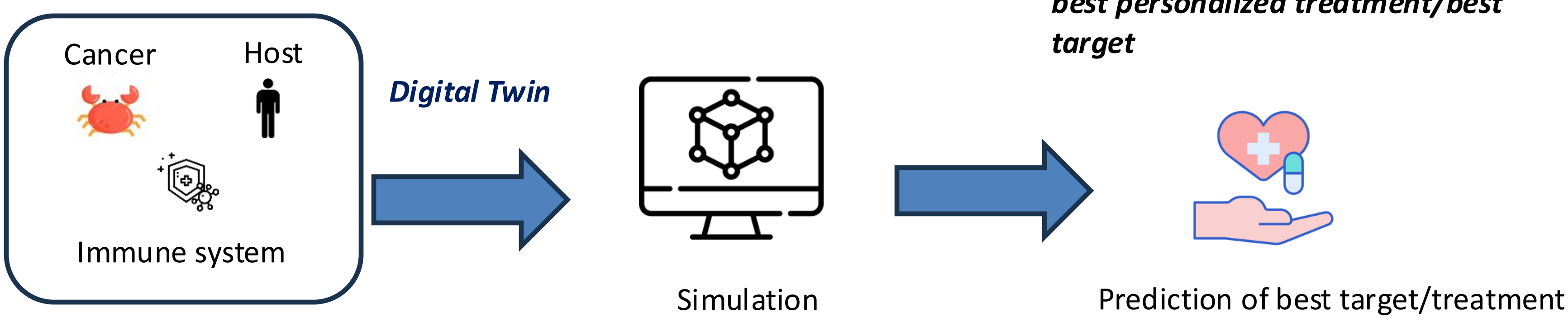
MEDITWIN: large-scale academic-industrial consortium (includin IHU) for creation of digital twins in oncology, cardiology, neurology...

WP3: Precision Medicine for Oncology
(PRISM, Dassault Systèmes)

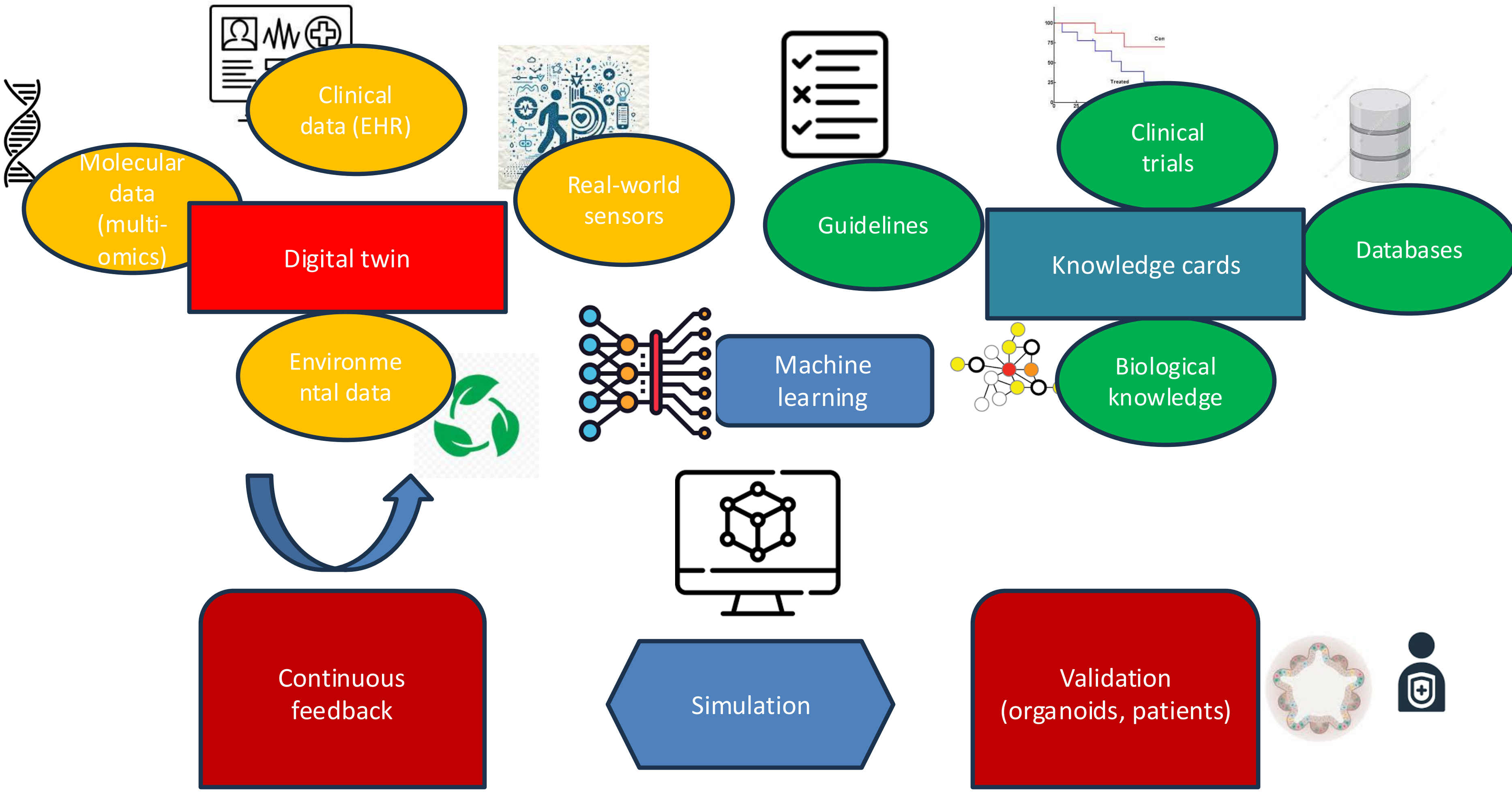
Create digital twins for cancer patients

- ***Multi-scale***
- ***Evolving in time***
- ***Integrating Cancer/Host/Immunity***

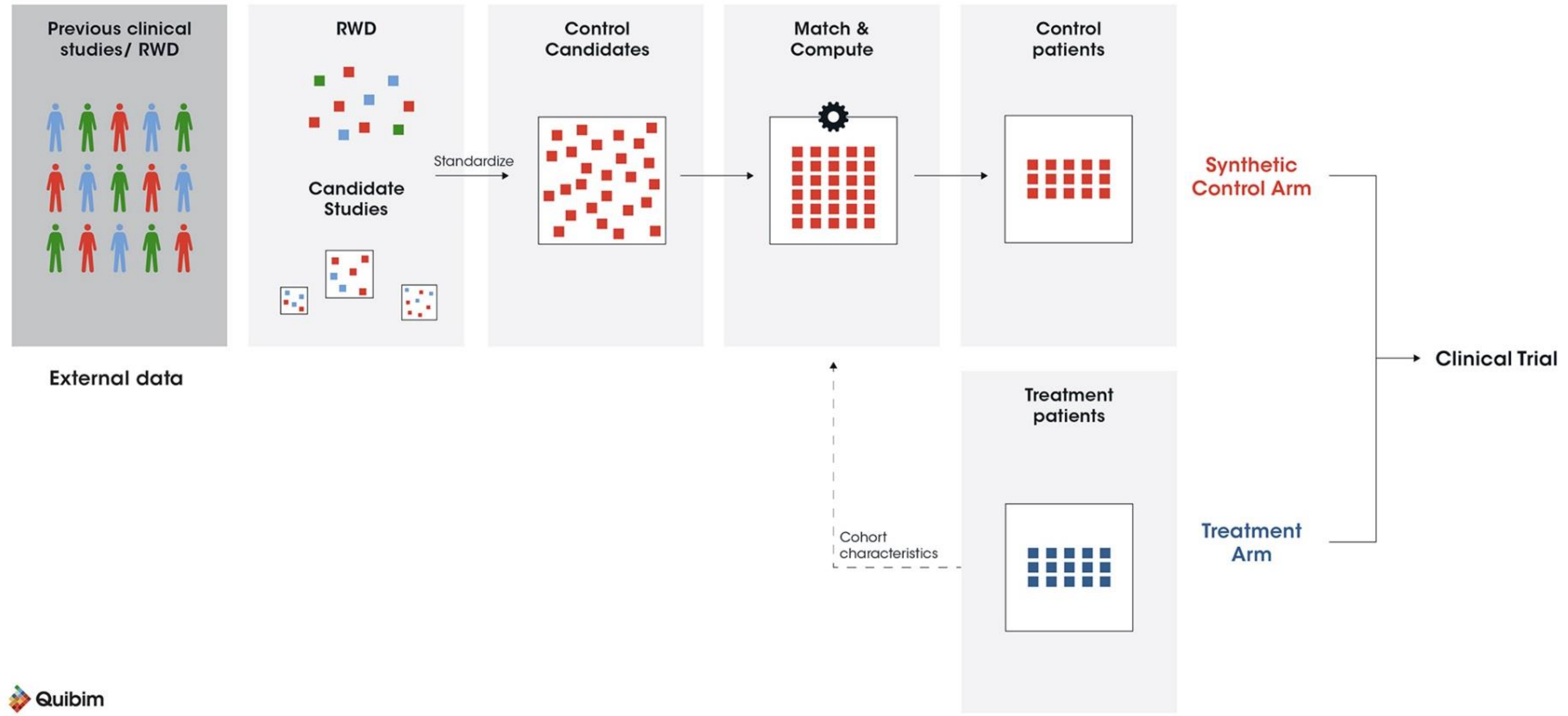
Allow simulation and prediction of best personalized treatment/best target



Idealized Workflow



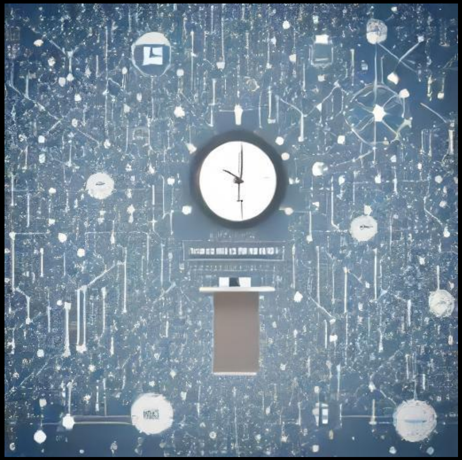
Synthetic control arms for clinical trials



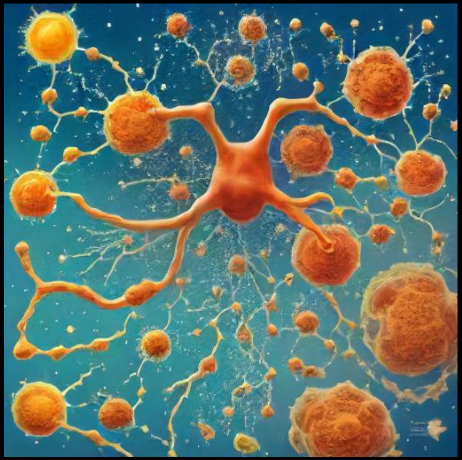
SOME PERSPECTIVES

Consultation with digital twin

Synthetic data for clinical trials



Insights into biology of cancer



AI-generated images

© Dassault Systèmes | Confidential Information | 12/18/2025

DASSAULT SYSTEMES

AI agents

AI agents are autonomous systems that perceive their environment, make decisions, and take actions to achieve specific goals—**often adapting over time through learning**.

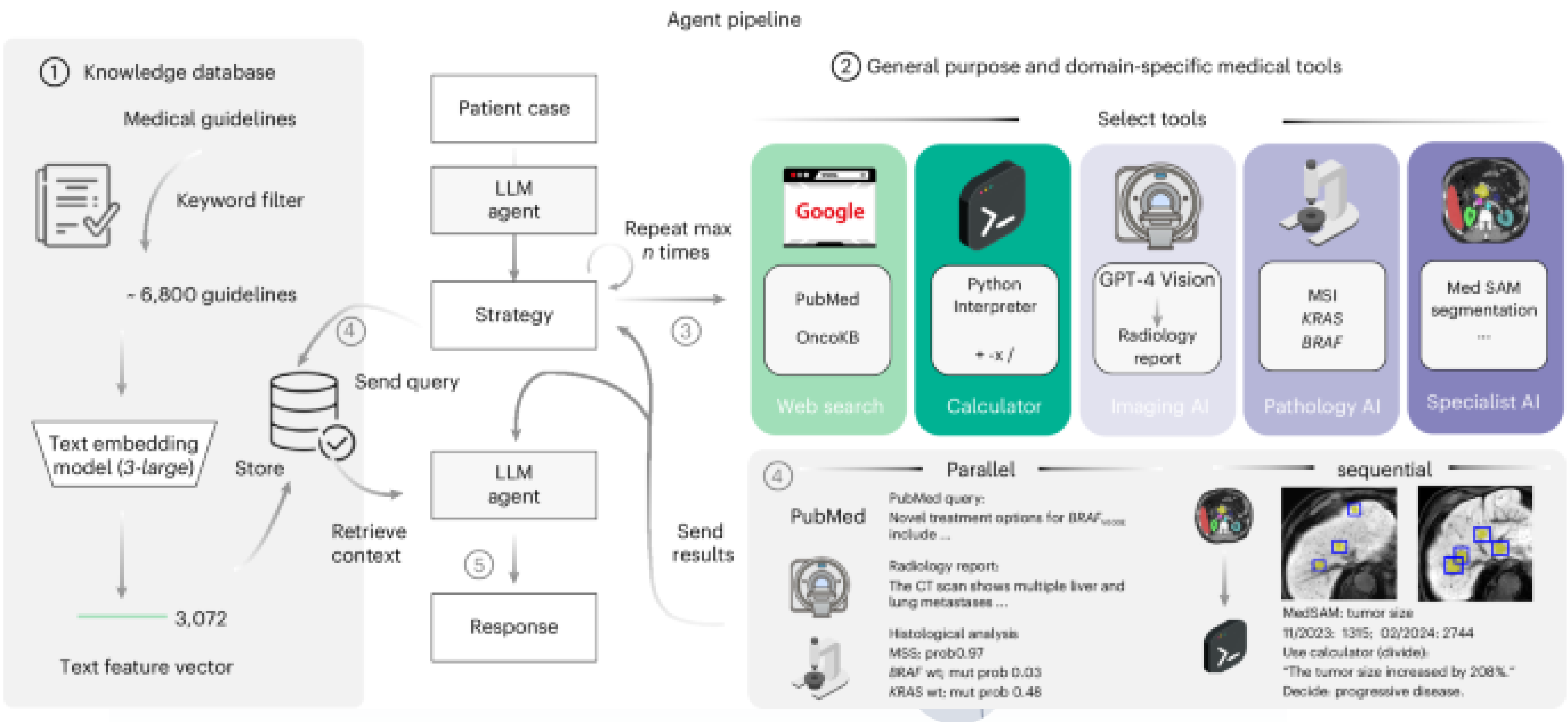
Comment

https://doi.org/10.1038/s43018-024-00861-7

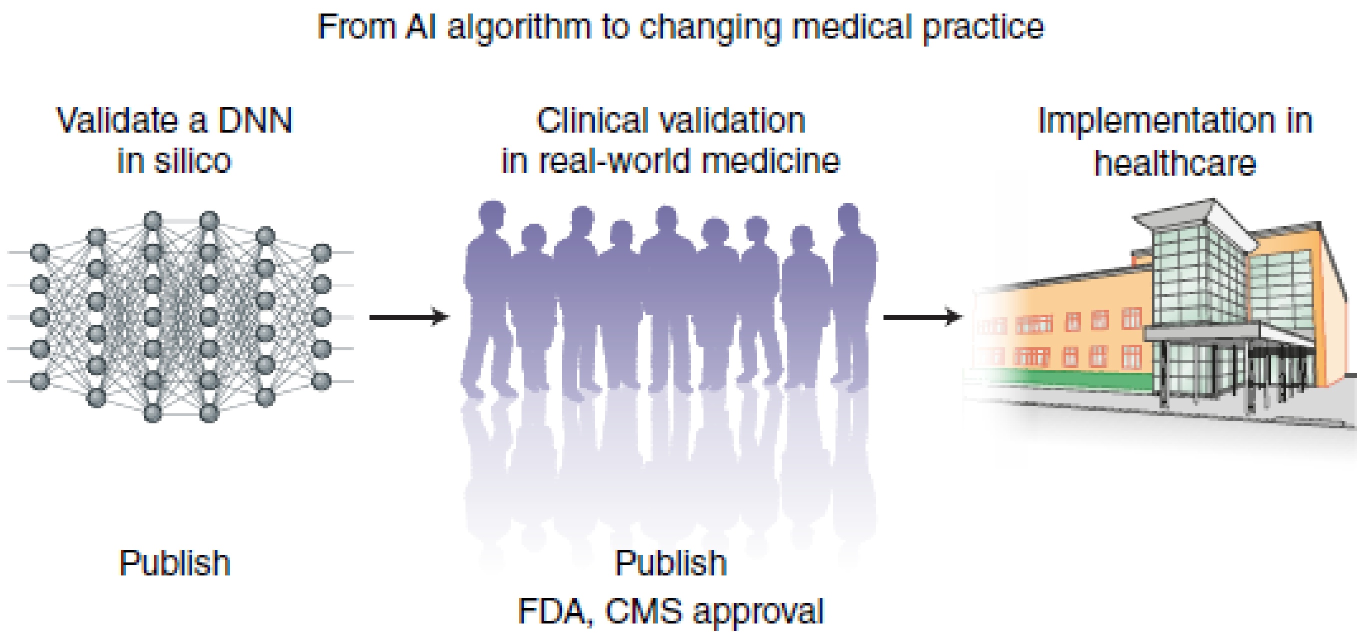
How AI agents will change cancer research and oncology

Yongju Lee, Dyke Ferber, Jennifer E. Rood, Aviv Regev & Jakob Nikolas Kather

Check for updates

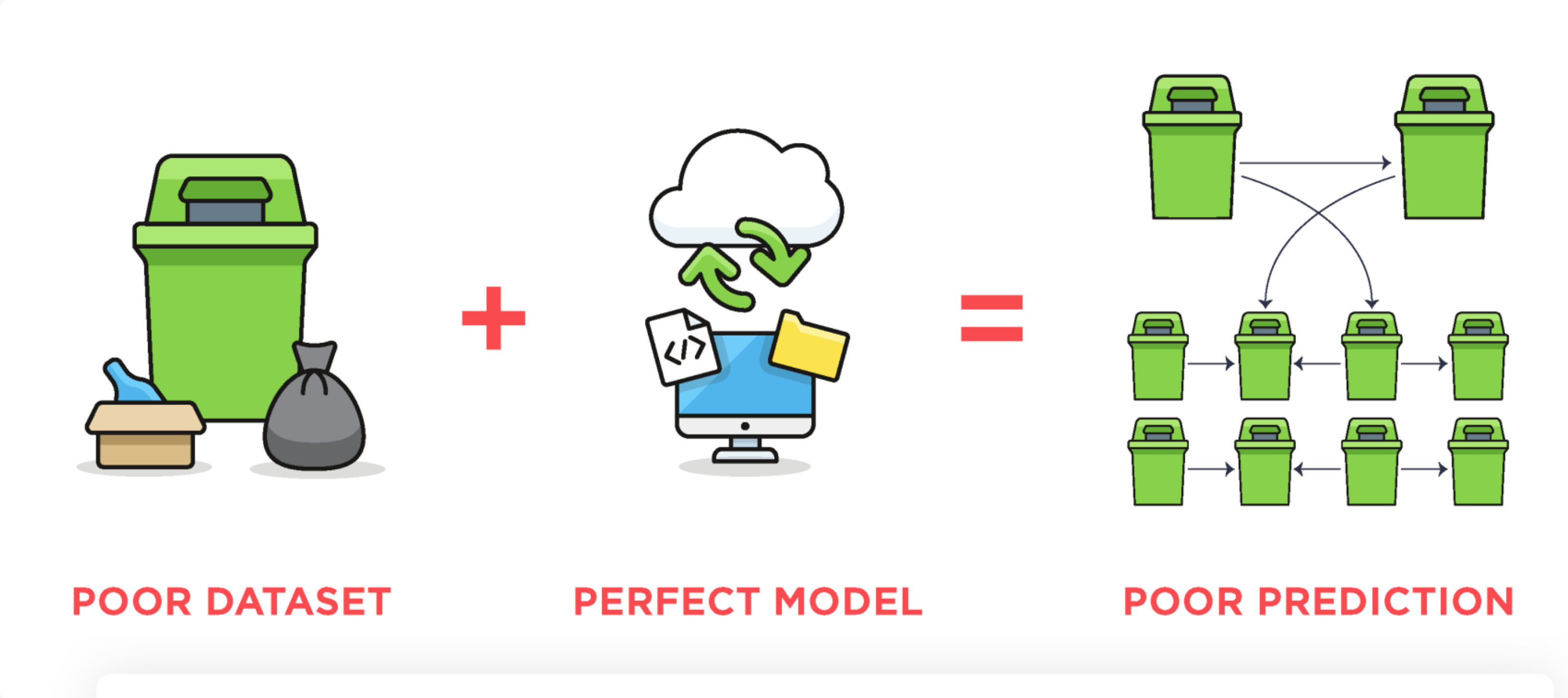


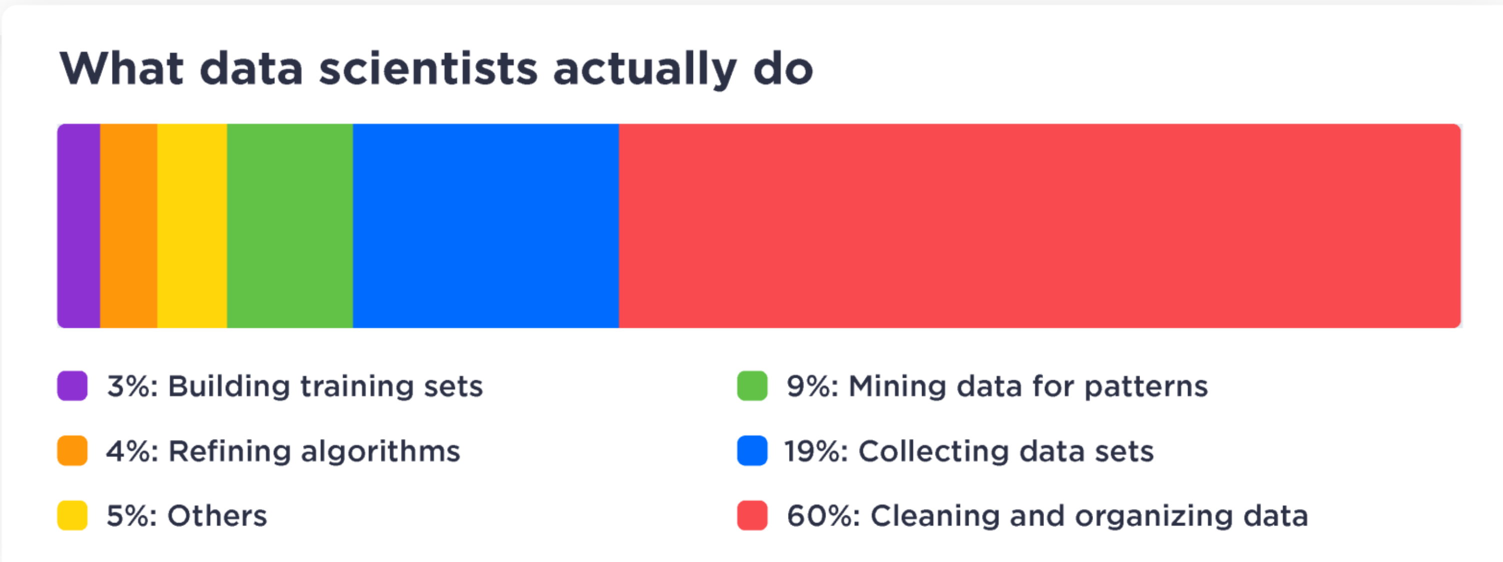
Be careful not to skip steps before use in real life



External validity!

Topol, Nat Med 2019





https://cldcvr.com/news-and-media/blog/clean-data-the-foundation-of-effective-machine-learning/

Acknowledgments

- Fabrice Barlesi, Fabrice André
- DITEP Gustave Roussy
- Inserm U981

- Patients and families
- Healthcare workers

THANK YOU FOR YOUR ATTENTION