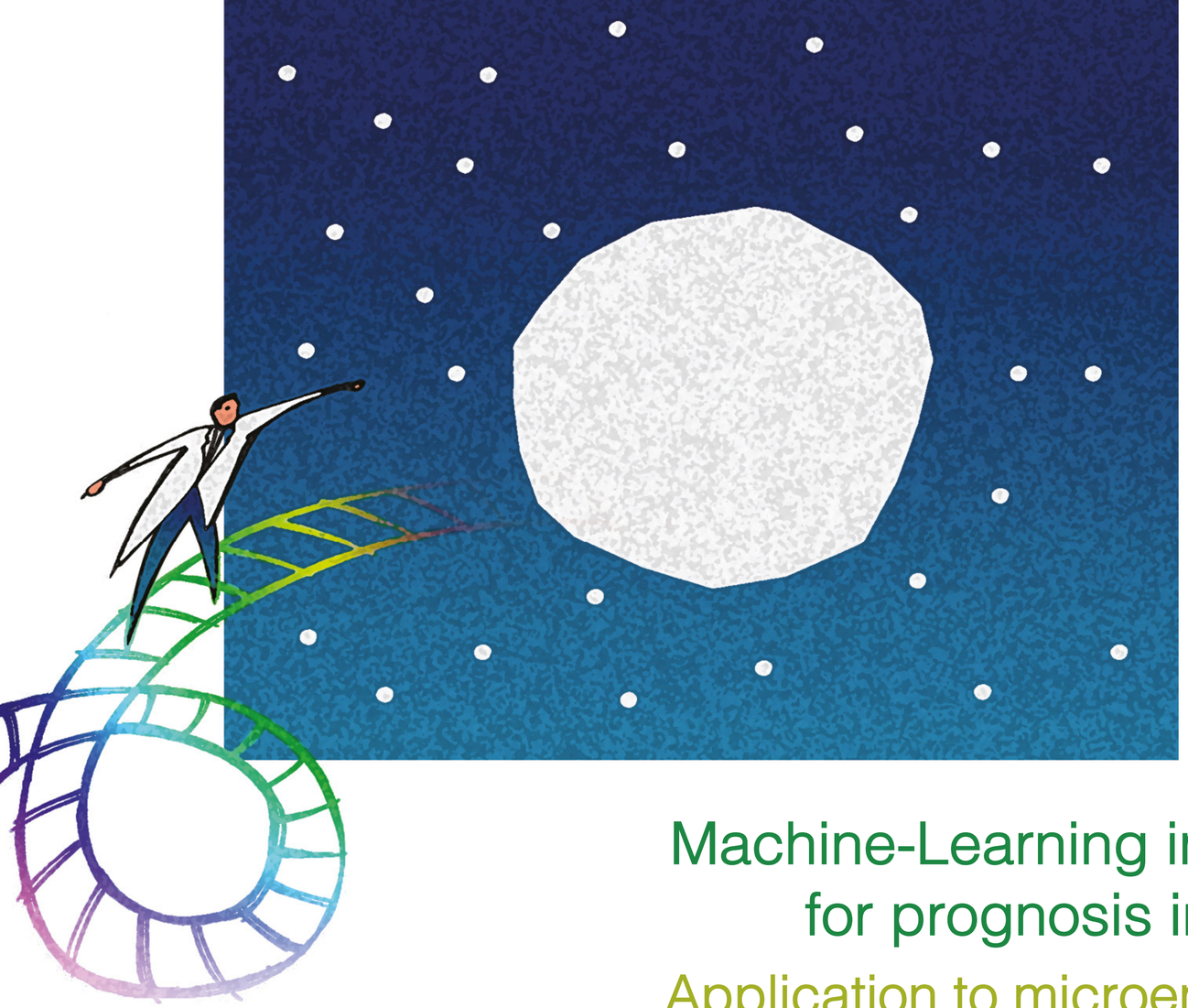


The young side of LYMPHOMA

gli under 40 a confronto

Milano, 14-15 aprile 2023

Machine-Learning in onco-hematology for prognosis in lymphomas:

Application to microenvironment in DLBCL

Gian Maria Zaccaria, PhD

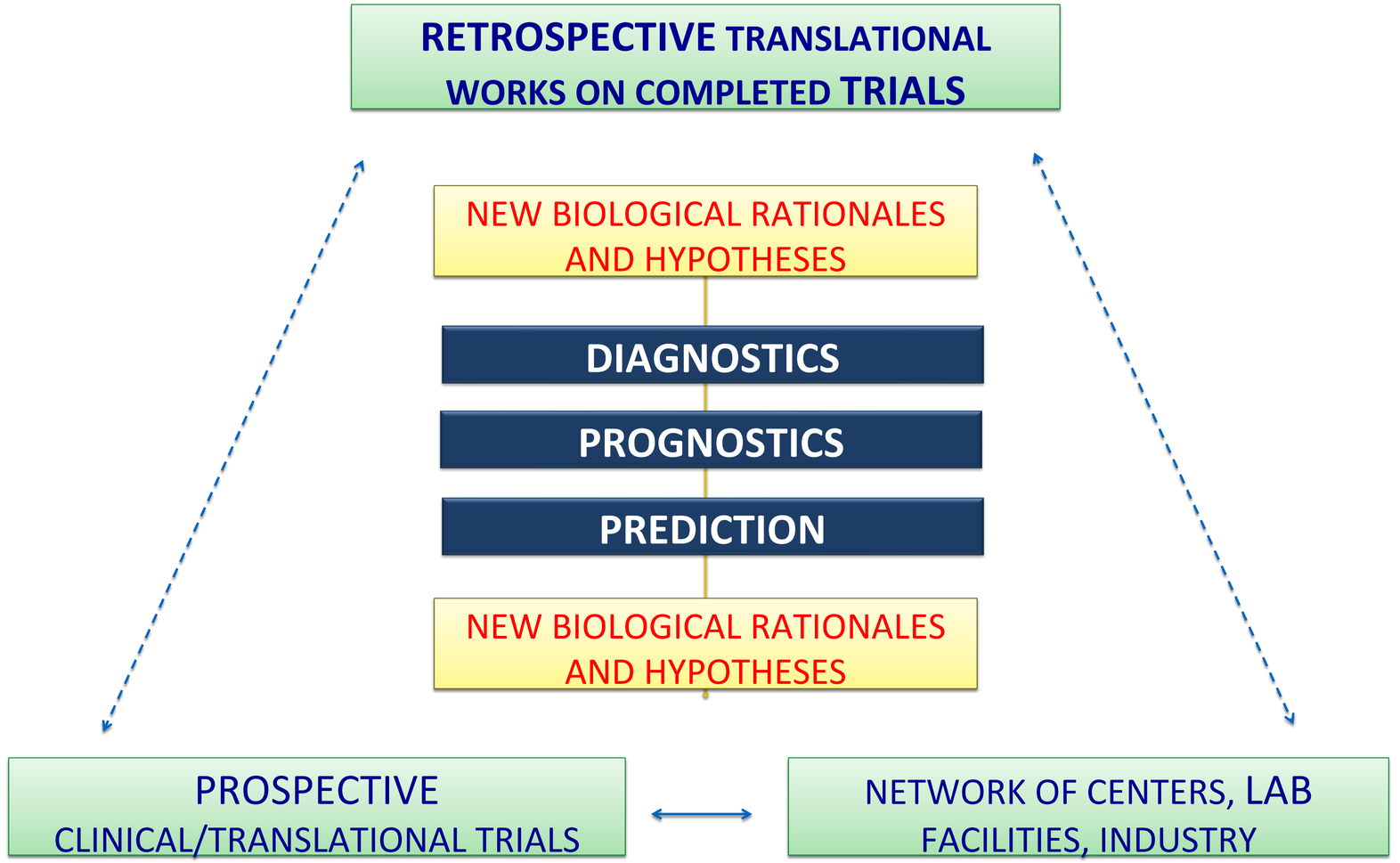
Disclosures of Name Surname

Company name	Research support	Employee	Consultant	Stockholder	Speakers bureau	Advisory board	Other

High-performance Medicine

The convergence of human and artificial intelligence

- New specialists
- Debugging
- Interpretation
- Robust Hypotheses
- Simulation
- Robust validation
- Prospective scrutiny
- Clinical translation



Courtesy of S. Ciavarella

nature methods

Explore content About the journal Publish with us

nature > nature methods > brief communications > article

Published: 24 August 2015

Predicting effects of noncoding variants with deep learning–based sequence model

Jian Zhou & Olga G Troyanskaya

Nature Methods **12**, 931–934 (2015) | Cite this article

67k Accesses | **1051** Citations | **158** Altmetric | Metrics

Abstract

Identifying functional effects of noncoding variants is a major challenge in human genetics. To predict the noncoding-variant effects *de novo* from sequence, we developed a deep learning–based algorithmic framework, DeepSEA (http://deepsea.princeton.edu/), that directly learns a regulatory sequence code from large-scale chromatin-profiling data, enabling prediction of chromatin effects of sequence alterations with single-nucleotide sensitivity. We further used this capability to improve prioritization of functional variants including expression quantitative trait loci (eQTLs) and disease-associated variants.

nature biotechnology

Explore content About the journal Publish with us

nature > nature biotechnology > analyses > article

Published: 27 July 2015

Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning

Babak Alipanahi, Andrew Delong, Matthew T Weirauch & Brendan J Frey

Nature Biotechnology **33**, 831–838 (2015) | Cite this article

185k Accesses | **1443** Citations | **273** Altmetric | Metrics

Abstract

Knowing the sequence specificities of DNA- and RNA-binding proteins is essential for developing models of the regulatory processes in biological systems and for identifying causal disease variants. Here we show that sequence specificities can be ascertained from experimental data with 'deep learning' techniques, which offer a scalable, flexible and unified computational approach for pattern discovery. Using a diverse array of experimental data and evaluation metrics, we find that deep learning outperforms other state-of-the-art methods, even when training on *in vitro* data and testing on *in vivo* data. We call this approach DeepBind and have built a stand-alone software tool that is fully automatic and handles millions of sequences per experiment. Specificities determined by DeepBind are readily visualized as a weighted ensemble of position weight matrices or as a 'mutation map' that indicates how variations affect binding within a specific sequence.

nature methods

Explore content About the journal Publish with us

nature > nature methods > articles > article

Article | Published: 12 October 2020

DeepC: predicting 3D genome folding using megabase-scale transfer learning

Ron Schwessinger, Matthew Gosden, Damien Downes, Richard C. Brown, A. Marieke Oudelaar, Jelena Telenius, Yee Whye Teh, Gerton Lunter & Jim R. Hughes

Nature Methods **17**, 1118–1124 (2020) | Cite this article

10k Accesses | **54** Citations | **43** Altmetric | Metrics

Abstract

Predicting the impact of noncoding genetic variation requires interpreting it in the context of three-dimensional genome architecture. We have developed deepC, a transfer-learning-based deep neural network that accurately predicts genome folding from megabase-scale DNA sequence. DeepC predicts domain boundaries at high resolution, learns the sequence determinants of genome folding and predicts the impact of both large-scale structural and single base-pair variations.

nature methods

Explore content · About the journal · Publish with us

nature > nature methods > brief communications > article

Published: 24 August 2015

Predicting effects of noncoding variants with deep learning–based sequence model

Jian Zhou & Olga G Troyanskaya

Nature Methods **12**, 931–934 (2015)

67k Accesses | **1051** Citations | **1**

Abstract

Identifying functional effects of n
To predict the noncoding-variant
learning–based algorithmic fram
directly learns a regulatory seque
enabling prediction of chromatin
sensitivity. We further used this c
including expression quantitative

nature biotechnology

Explore content · About the journal · Publish with us

nature > nature biotechnology > analyses > article

Published: 27 July 2015

Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning

Babak Alipanahi, Andrew Delong, Matthew T Weirauch & Brendan J Frey

nature methods

Explore content · About the journal · Publish with us

nature > nature methods > articles > article

Article | Published: 12 October 2020

DeepC: predicting 3D genome folding using megabase-scale transfer learning

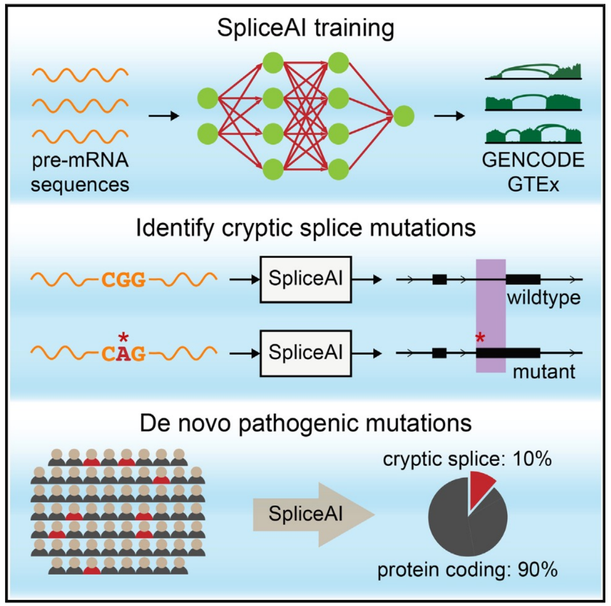
Ron Schwessinger, Matthew Gosden, Damien Downes, Richard C. Brown, A. Marieke Oudelaar, Jelena

Article

Cell

Predicting Splicing from Primary Sequence with Deep Learning

Graphical Abstract



Authors

Kishore Jaganathan, Sofia Kyriazopoulou Panagiotopoulou, Jeremy F. McRae, ..., Serafim Batzoglou, Stephan J. Sanders, Kyle Kai-How Farh

Correspondence

kfarh@illumina.com

In Brief

A deep neural network precisely models mRNA splicing from a genomic sequence and accurately predicts noncoding cryptic splice mutations in patients with rare genetic diseases.

nature protocols

PROTOCOL

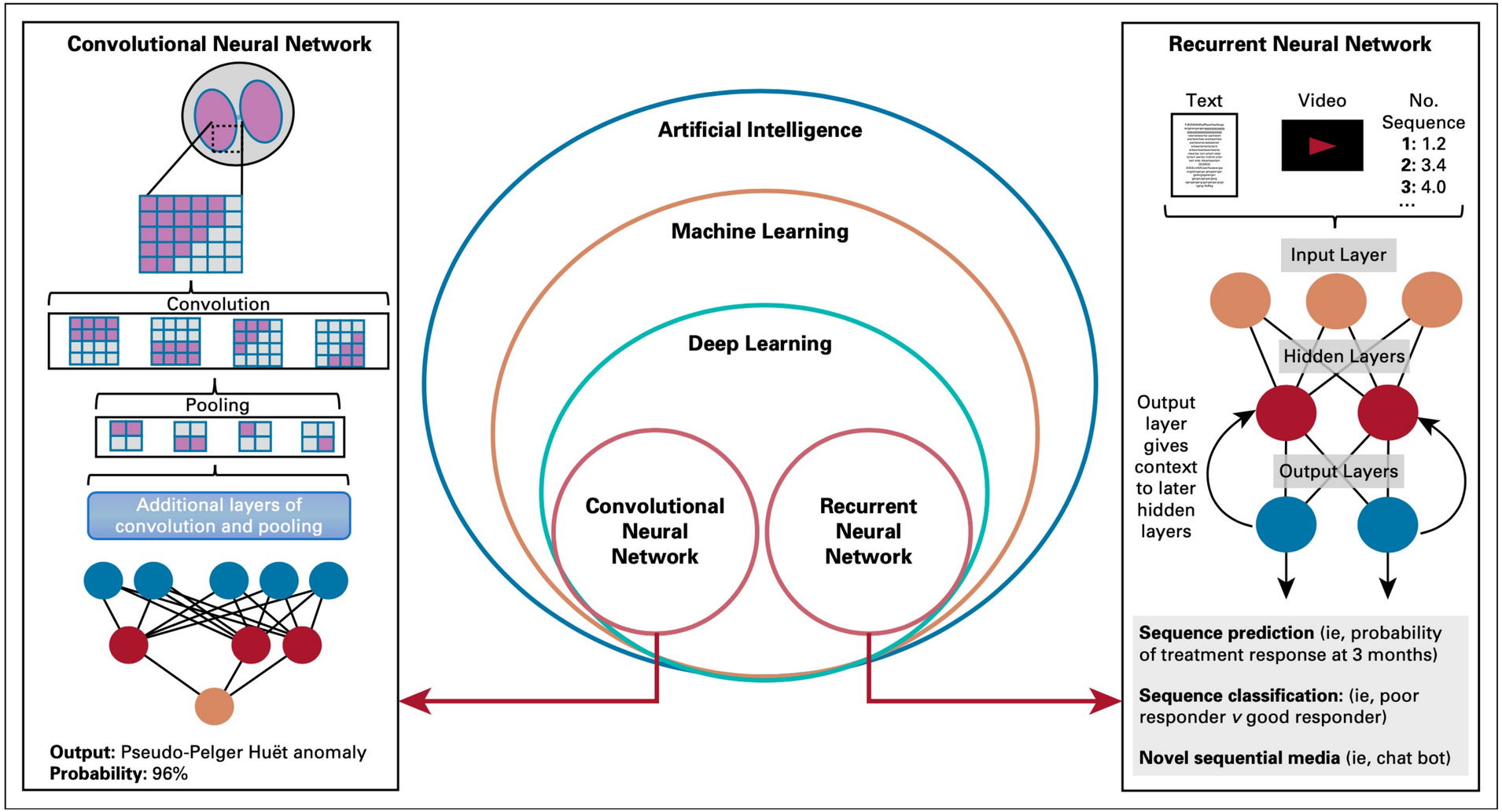
https://doi.org/10.1038/s41596-019-0279-7

Investigating RNA editing in deep transcriptome datasets with REDItools and REDIportal

Claudio Lo Giudice[1], Marco Antonio Tangaro[1], Graziano Pesole[1,2,3] and Ernesto Picardi[1,2,3*]

RNA editing is a widespread post-transcriptional mechanism able to modify transcripts through insertions/deletions or base substitutions. It is prominent in mammals, in which millions of adenosines are deaminated to inosines by members of the ADAR family of enzymes. A-to-I RNA editing has a plethora of biological functions, but its detection in large-scale transcriptome datasets is still an unsolved computational task. To this aim, we developed REDItools, the first software package devoted to the RNA editing profiling in RNA-sequencing (RNAseq) data. It has been successfully used in human transcriptomes, proving the tissue and cell type specificity of RNA editing as well as its pervasive nature. Outcomes from large-scale REDItools analyses on human RNAseq data have been collected in our specialized REDIportal database, containing more than 4.5 million events. Here we describe in detail two bioinformatic procedures based on our computational resources, REDItools and REDIportal. In the first procedure, we outline a workflow to detect RNA editing in the human cell line NA12878, for which transcriptome and whole genome data are available. In the second procedure, we show how to identify dysregulated editing at specific recoding sites in post-mortem brain samples of Huntington disease donors. On a 64-bit computer running Linux with ≥32 GB of random-access memory (RAM), both procedures should take ~76 h, using 4 to 24 cores. Our protocols have been designed to investigate RNA editing in different organisms with available transcriptomic and/or genomic reads. Scripts to complete both procedures and a docker image are available at https://github.com/BioinfoUNIBA/REDItools.

Artificial Intelligence, Machine/Deep-Learning

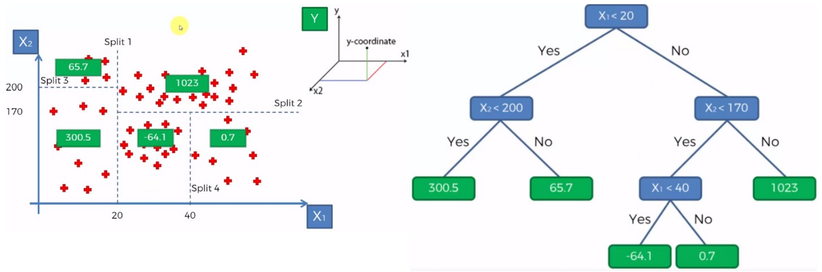


Nagy et al., Machine Learning in Oncology: What should clinicians knows ?, *JCO-Clinical Cancer Informatics*, 2020

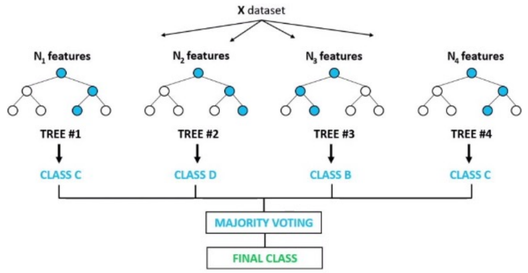
Techniques

- Supervised – clinical outcomes are known and already defined (e.g., clinical response, progression, death)
  - Tree-based vs. Neural Network based algorithms

Decision Tree



Random Forest

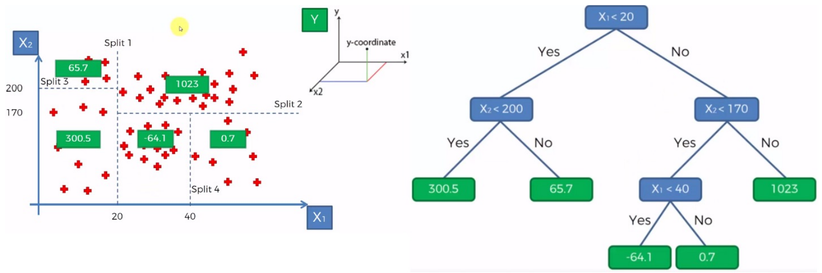


Images top-left from “The Guide to Decision Tree-based Algorithms in Machine Learning”, https://omdena.com/blog/decision-tree-based-algorithms/

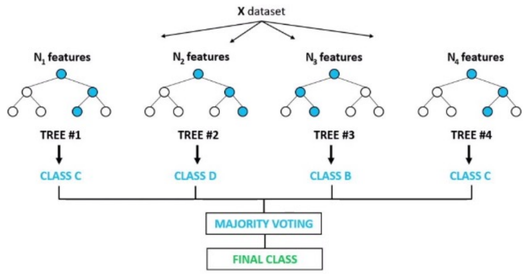
Techniques

- Supervised – clinical outcomes are known and already defined (e.g., clinical response, progression, death)
  - Tree-based vs. Neural Network based algorithms

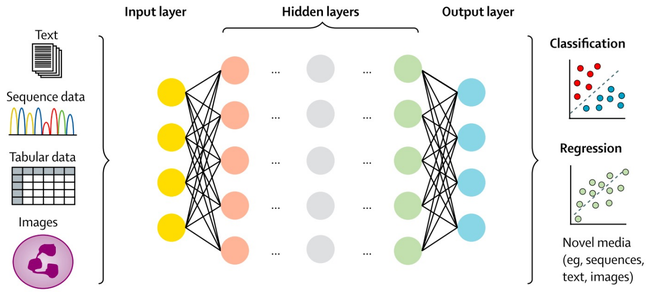
Decision Tree



Random Forest



Neural Networks

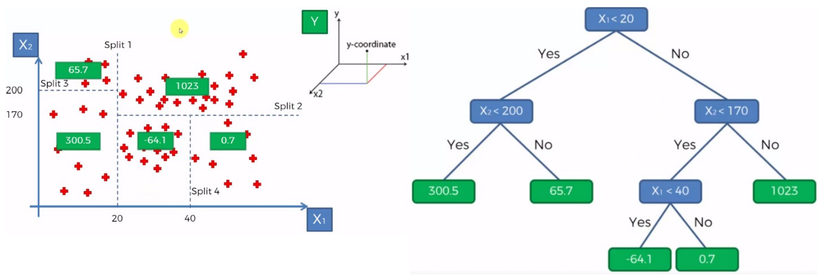


Images top-left from “The Guide to Decision Tree-based Algorithms in Machine Learning”, https://omdena.com/blog/decision-tree-based-algorithms/
Images top-right from Radakovich et al., Machine Learning in hematology malignancies, *The Lancet Hematology*, 2020

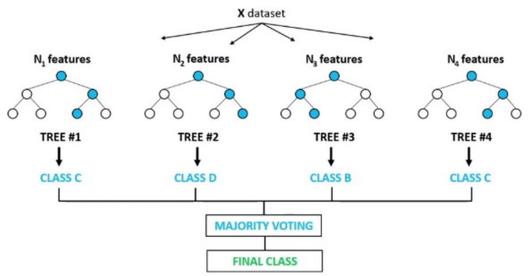
Techniques

- Supervised – clinical outcomes are known and already defined (e.g., clinical response, progression, death)
  - Tree-based vs. Neural Network based algorithms

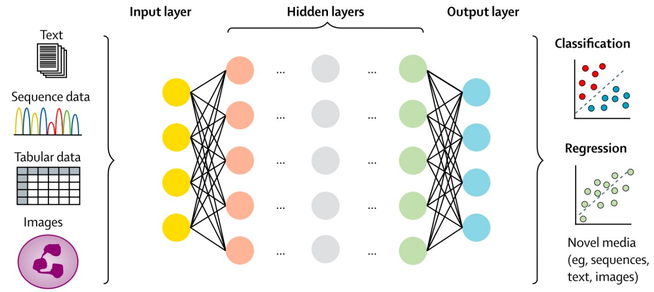
Decision Tree



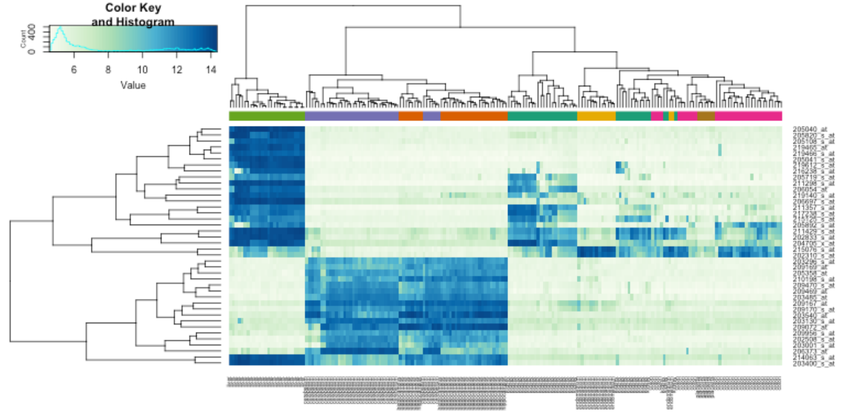
Random Forest



Neural Networks



- Unsupervised – outcomes are not defined (e.g., UMAP, NMF)



Images top-left from “The Guide to Decision Tree-based Algorithms in Machine Learning”, https://omdena.com/blog/decision-tree-based-algorithms/
Images top-right from Radakovich et al., Machine Learning in hematology malignancies, *The Lancet Hematology*, 2020
Image bottom-left from “Hierarchical Clustering Algorith fro Machine-Learning” https://medium.com/geekculture/hierarchical-clustering-simply-explained-f86b9ed96db7
Image bottom-right from “Basic Machine_learning” https://genomicsclass.github.io/book/pages/clustering_and_heatmaps.html

Typical Machine-Learning pipeline

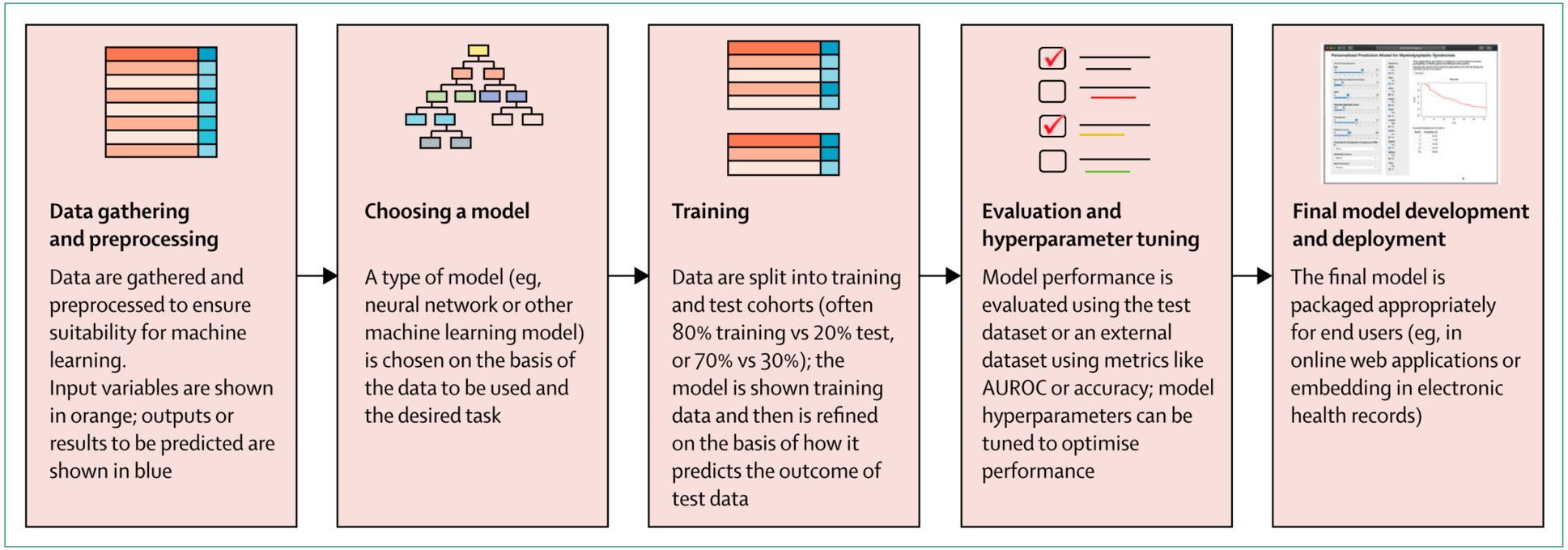


Image from Radakovich et al., Machine Learning in hematology malignancies, *The Lancet Hematology*, 2020

Typical Machine-Learning pipeline

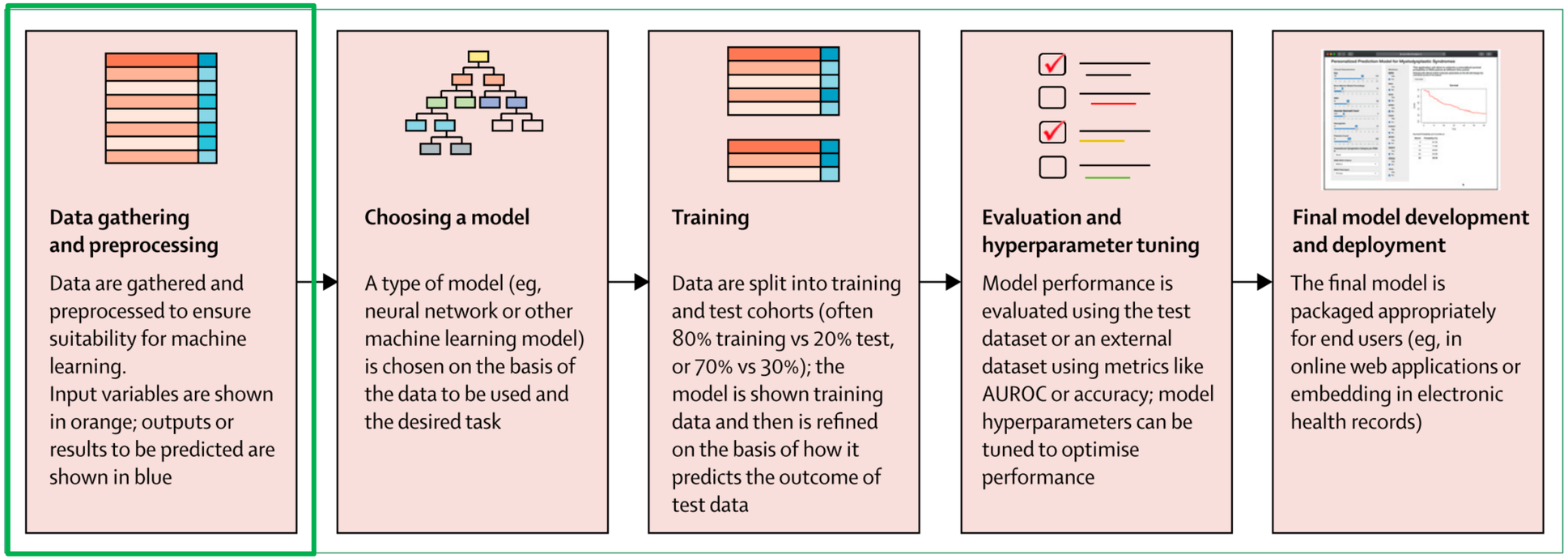


Image from Radakovich et al., Machine Learning in hematology malignancies, *The Lancet Hematology*, 2020

Typical Machine-Learning pipeline

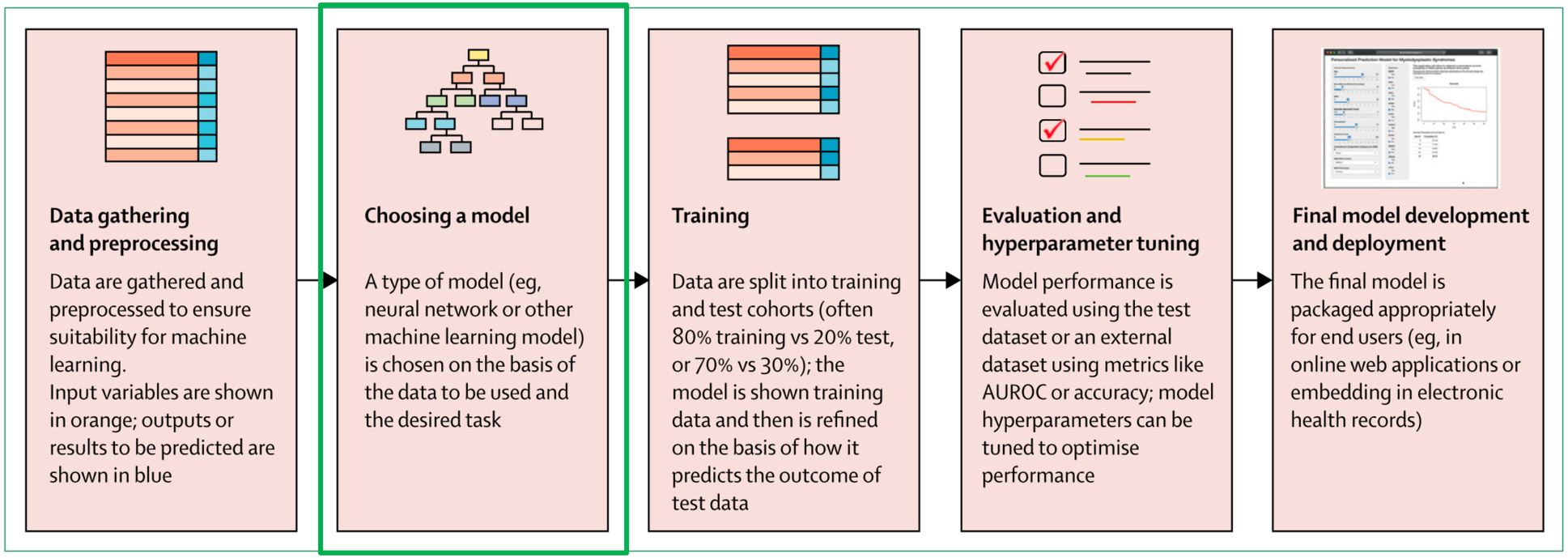


Image from Radakovich et al., Machine Learning in hematology malignancies, *The Lancet Hematology*, 2020

Typical Machine-Learning pipeline

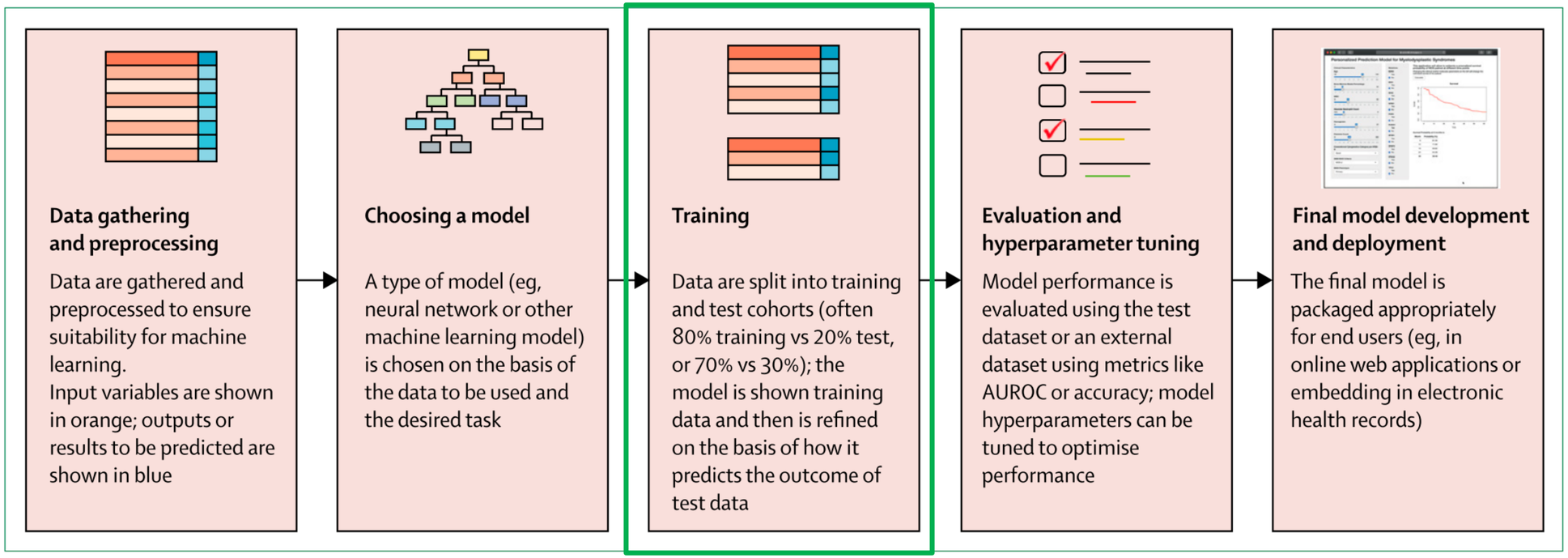


Image from Radakovich et al., Machine Learning in hematology malignancies, *The Lancet Hematology*, 2020

Typical Machine-Learning pipeline

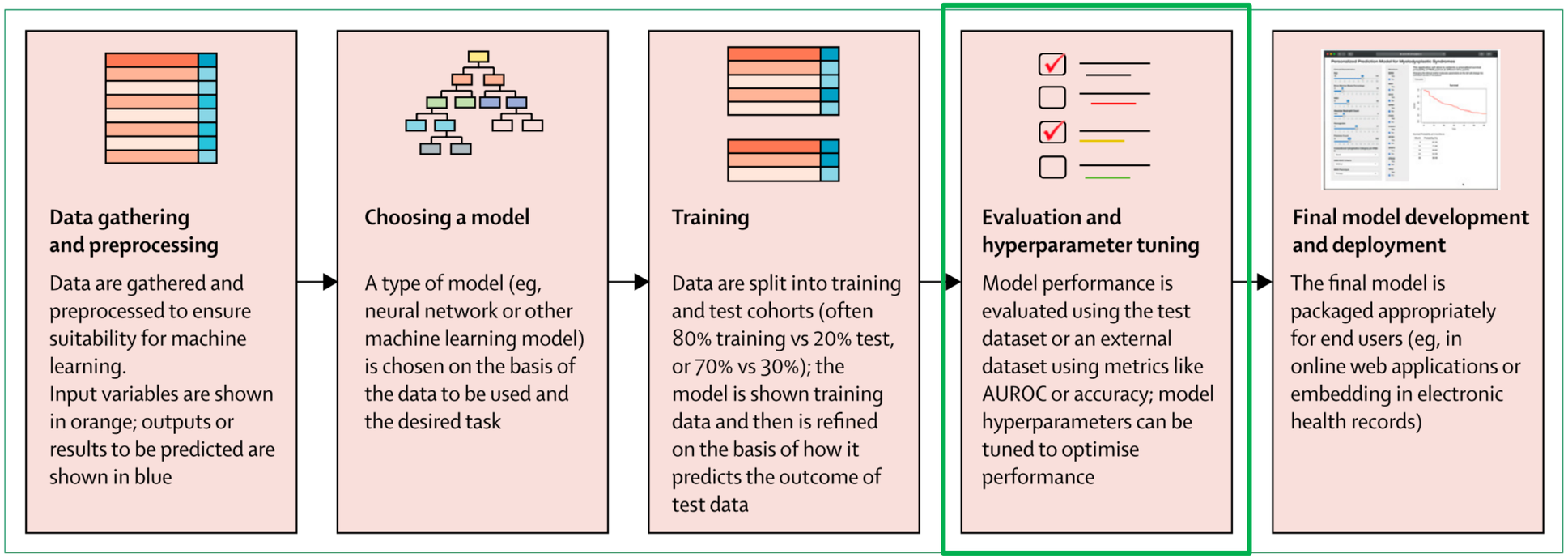


Image from Radakovich et al., Machine Learning in hematology malignancies, *The Lancet Hematology*, 2020

Typical Machine-Learning pipeline

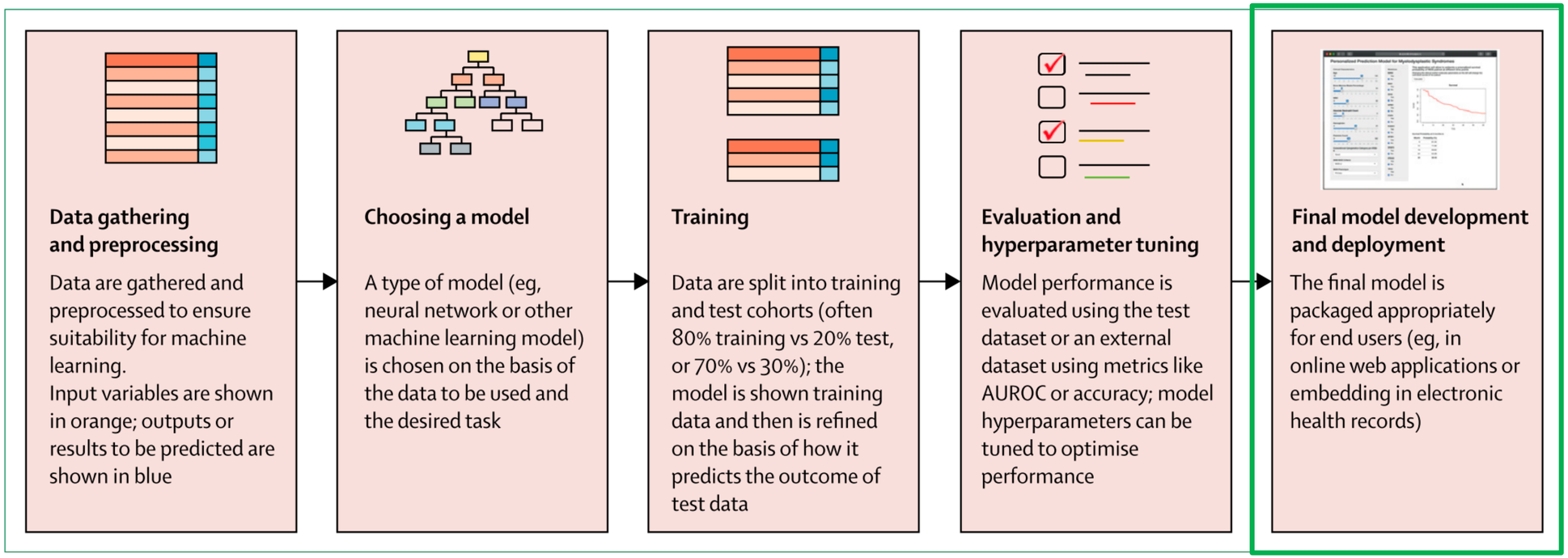
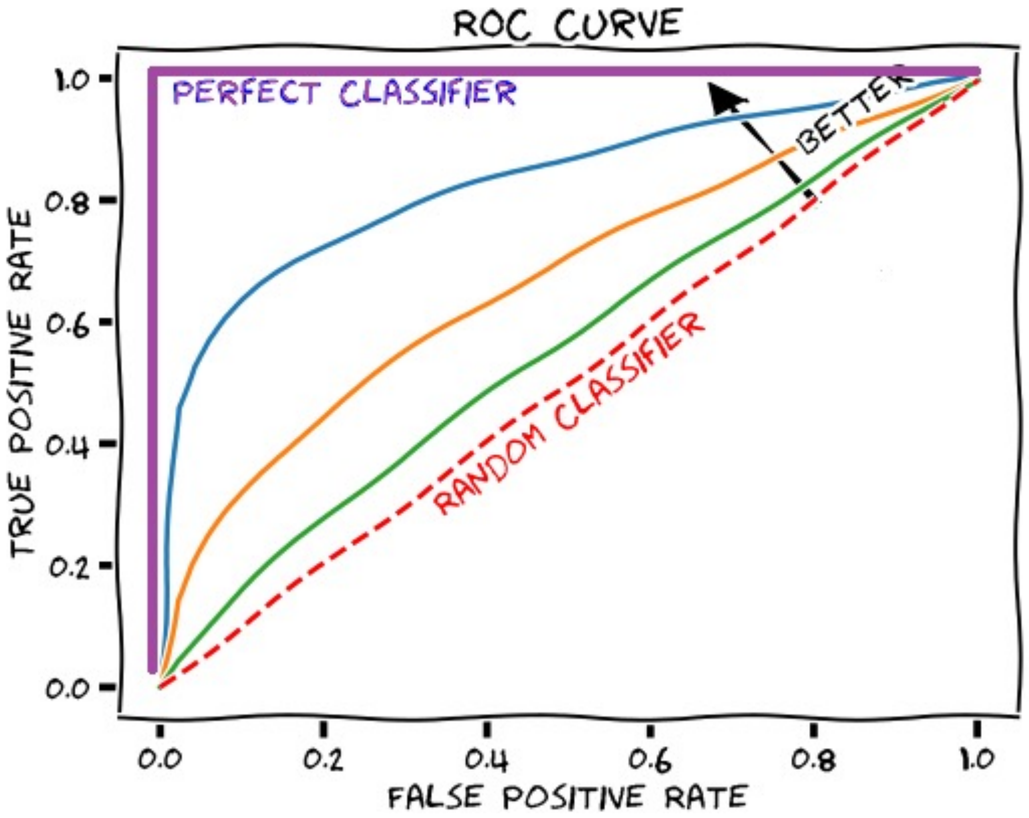


Image from Radakovich et al., Machine Learning in hematology malignancies, *The Lancet Hematology*, 2020

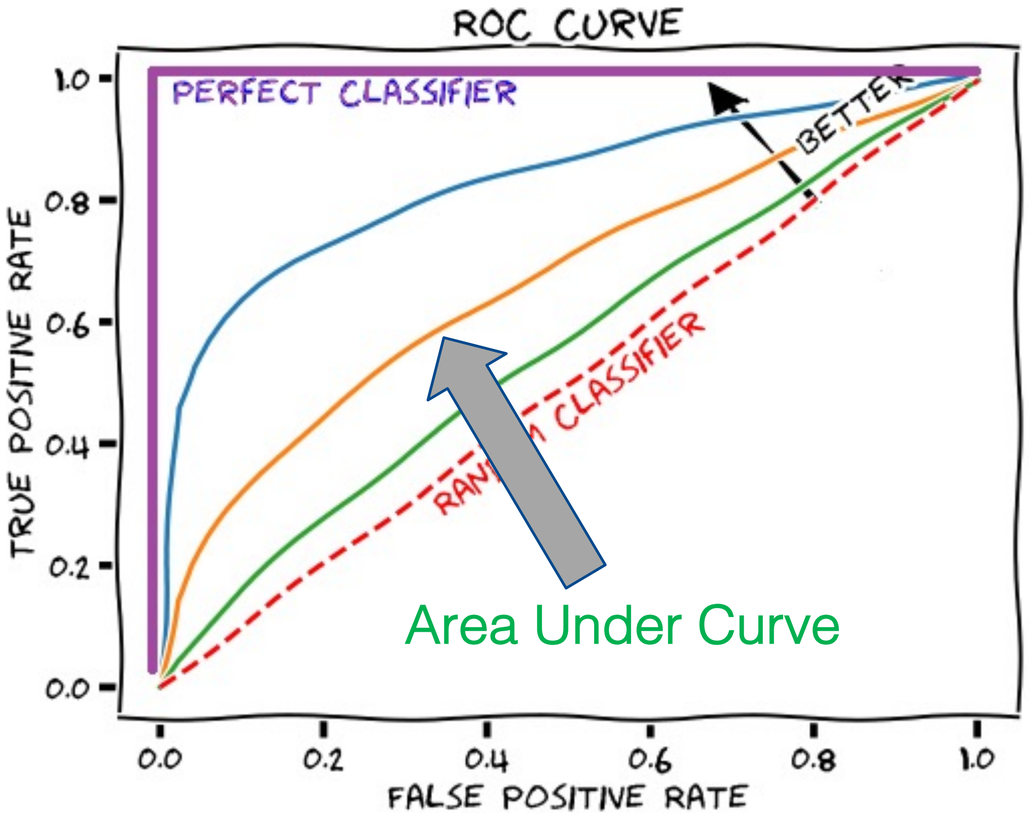
Metric evaluation



Abbreviations. ROC: Receiver Operating Characteristic; AUC: Area under curve

Image on the left from “Glass box, Machine Learning and Medicine”, https://glassboxmedicine.com/

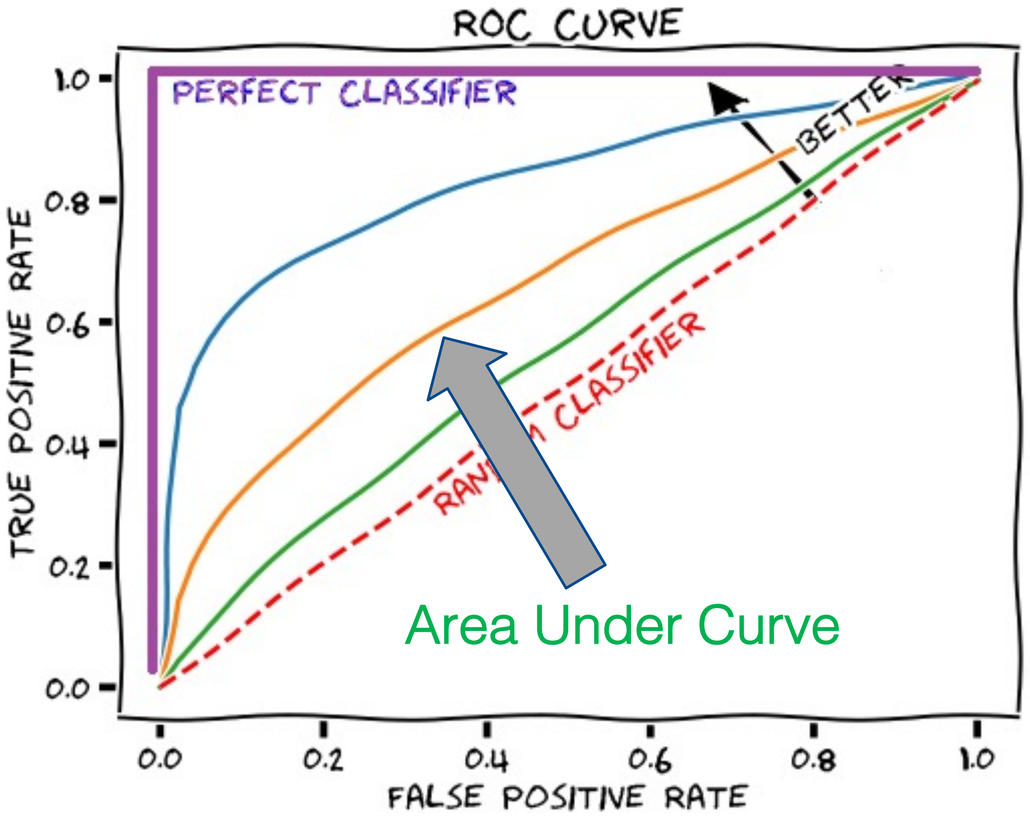
Metric evaluation



Abbreviations. ROC: Receiver Operating Characteristic; AUC: Area under curve

Image on the left from “Glass box, Machine Learning and Medicine”, https://glassboxmedicine.com/

Metric evaluation



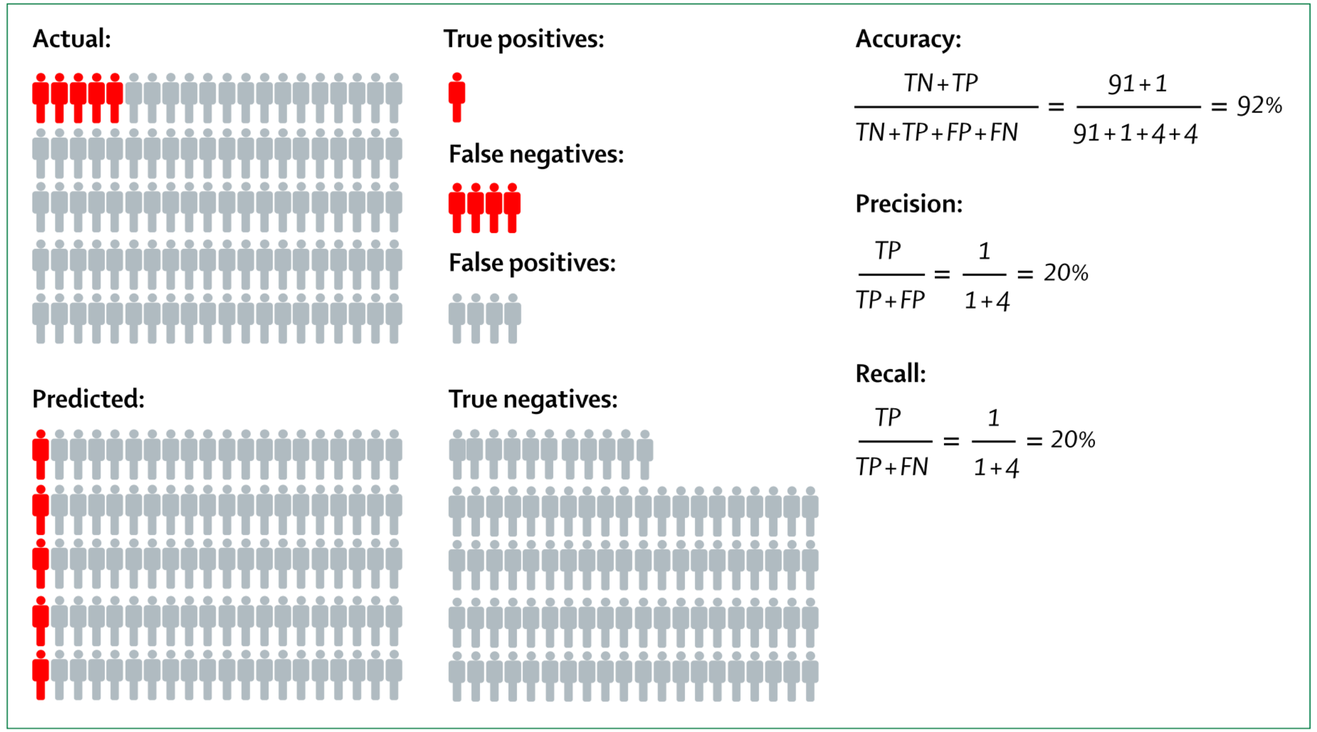
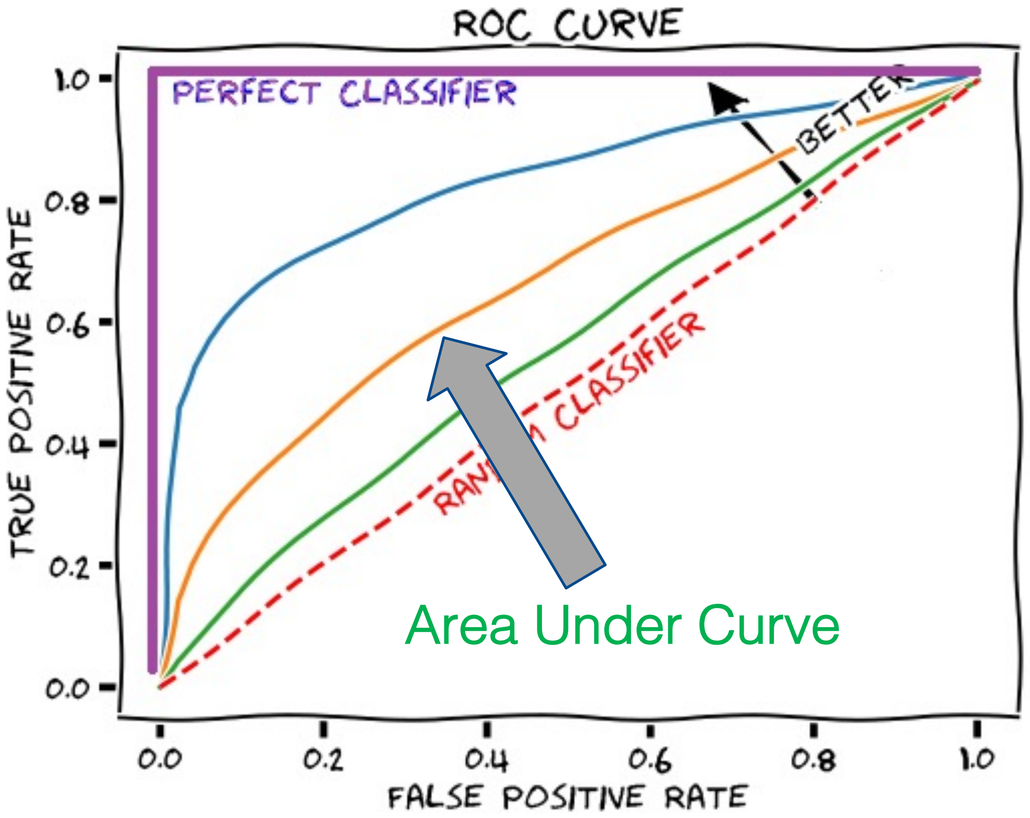


Image on the left from “Glass box, Machine Learning and Medicine”, https://glassboxmedicine.com/
Image on the right from Radakovich et al., Machine Learning in hematology malignancies, *The Lancet Hematology*, 2020

Abbreviations. ROC: Receiver Operating Characteristic; AUC: Area under curve; TN=True Negative; TP=True Positive; FP= False Positive; FN= False Negative

Metric evaluation



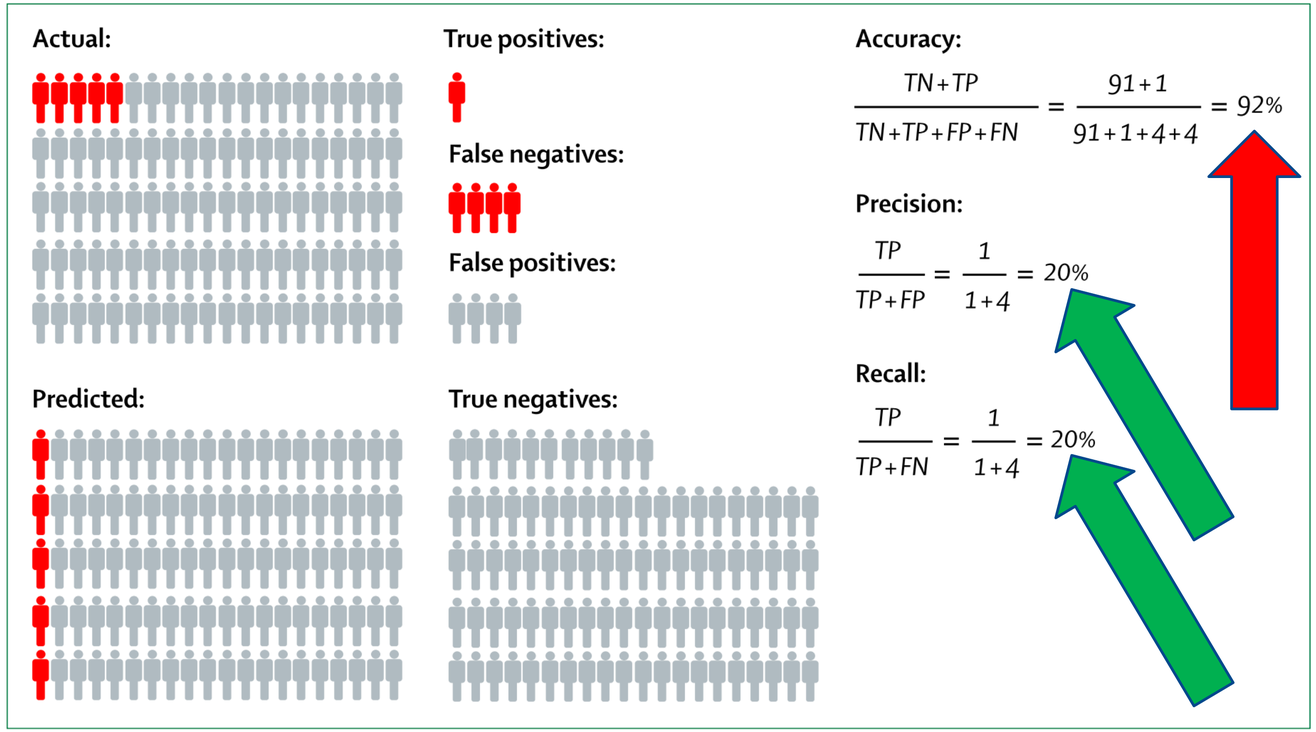


Image on the left from “Glass box, Machine Learning and Medicine”, https://glassboxmedicine.com/
Image on the right from Radakovich et al., Machine Learning in hematology malignancies, *The Lancet Hematology*, 2020

Abbreviations. ROC: Receiver Operating Characteristic; AUC: Area under curve; TN=True Negative; TP=True Positive; FP= False Positive; FN= False Negative

Machine Learning in Hematology

- **Translational Medicine**. Biomarker discovery, Drug-targeted prioritization, Drug discovery, Prediction of chemical toxicity, Genetic variant annotation, Prognostication

- **Clinical Practice**. Disease Diagnosis, Interpretation of patients' genomics, Treatment selection, Automated surgery, Patient Monitoring, Patient risk stratification for primary prevention

P. Becker, Machine Learning in Hematology, 60th ASH Annual Meeting, 7-10 Dec 2019, Orlando, USA

Machine Learning in Hematology

- **Translational Medicine**. Biomarker discovery, Drug-targeted prioritization, Drug discovery, Prediction of chemical toxicity, Genetic variant annotation, Prognostication

- **Clinical Practice**. Disease Diagnosis, Interpretation of patients' genomics, Treatment selection, Automated surgery, Patient Monitoring, Patient risk stratification for primary prevention

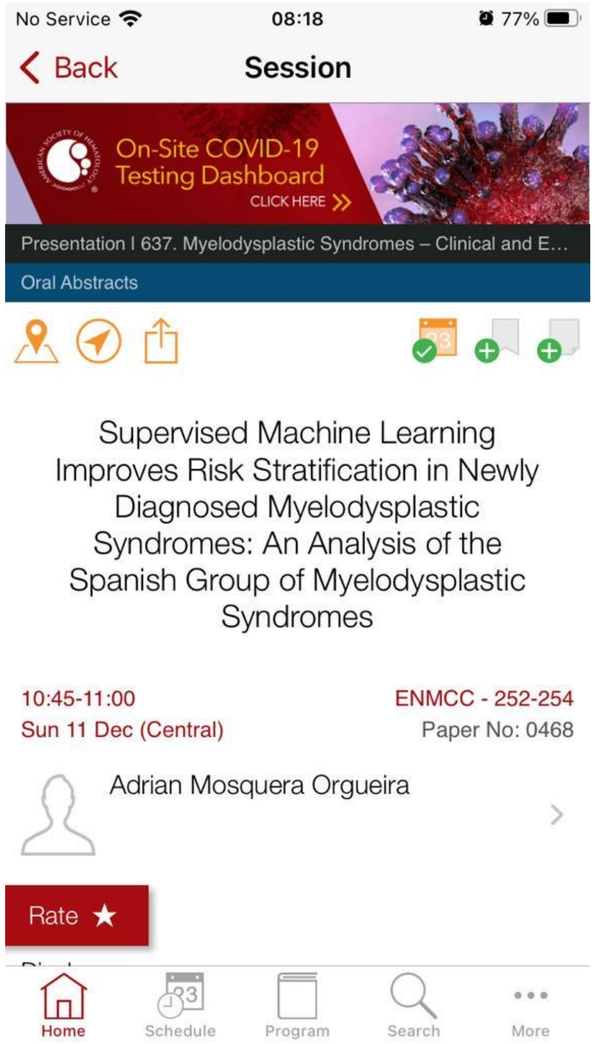
- At the 2021 ASH meeting, **74 abstracts** included within either title or methods the keywords "Machine-Learning", "Deep-Learning", and "Artificial Intelligence", respectively. Of these, 9 were on lymphoma projects

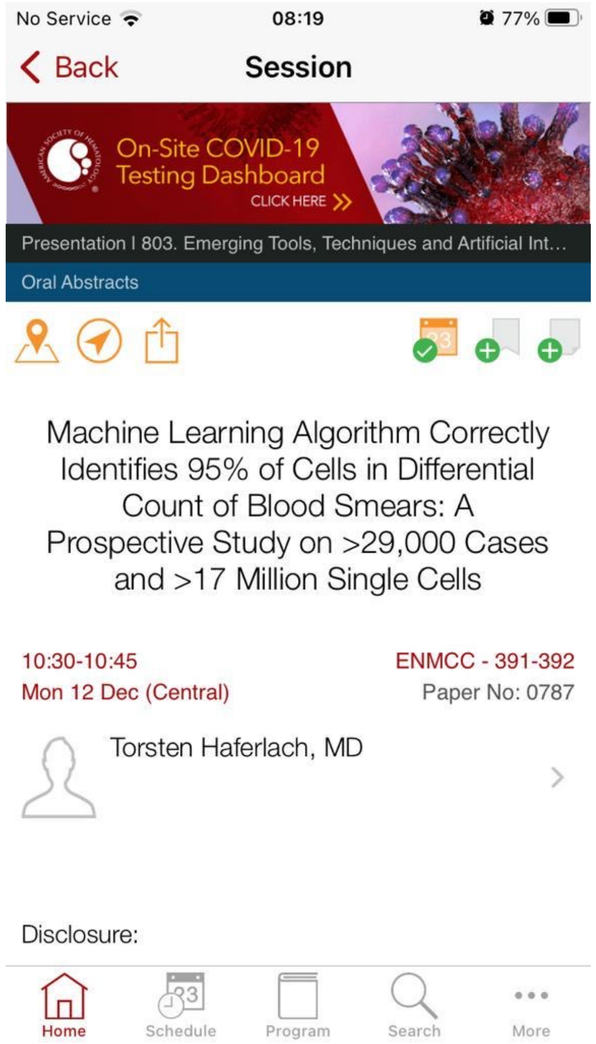
- At the 2022 ASH meeting, **31 abstracts** and a dedicated congress section

P. Becker, Machine Learning in Hematology, 60th ASH Annual Meeting, 7-10 Dec 2019, Orlando, USA


SOS only
08:15
78%
Back
Session
On-Site COVID-19 Testing Dashboard
CLICK HERE
Presentation I 903. Health Services and Quality—Myeloid Malig…
Oral Abstracts
Machine Learning Approaches Incorporating High-Dimensional Longitudinal Data Improve the Prediction of Survival after Allogeneic Hematopoietic Cell Transplantation
10:30-10:45
Sat 10 Dec (Central)
ENMCC - 388-390
Paper No: 0131
Yiwang Zhou, PhD
Rate
Disclosure:
Home
Schedule
Program
Search
More


SOS only
08:15
78%
Back
Session
On-Site COVID-19 Testing Dashboard
CLICK HERE
Presentation I 634. Myeloproliferative Syndromes - Clinical and…
Oral Abstracts
Machine Learning Improves Risk Stratification in Myelofibrosis: An Analysis of the Spanish Registry of Myelofibrosis
16:30-16:45
Sat 10 Dec (Central)
ENMCC - 217-219
Paper No: 0339
Adrian Mosquera Orgueira
Disclosure:
Home
Schedule
Program
Search
More


No Service
08:18
77%
Back
Session
On-Site COVID-19 Testing Dashboard
CLICK HERE
Presentation I 637. Myelodysplastic Syndromes – Clinical and E…
Oral Abstracts
Supervised Machine Learning Improves Risk Stratification in Newly Diagnosed Myelodysplastic Syndromes: An Analysis of the Spanish Group of Myelodysplastic Syndromes
10:45-11:00
Sun 11 Dec (Central)
ENMCC - 252-254
Paper No: 0468
Adrian Mosquera Orgueira
Rate
Home
Schedule
Program
Search
More



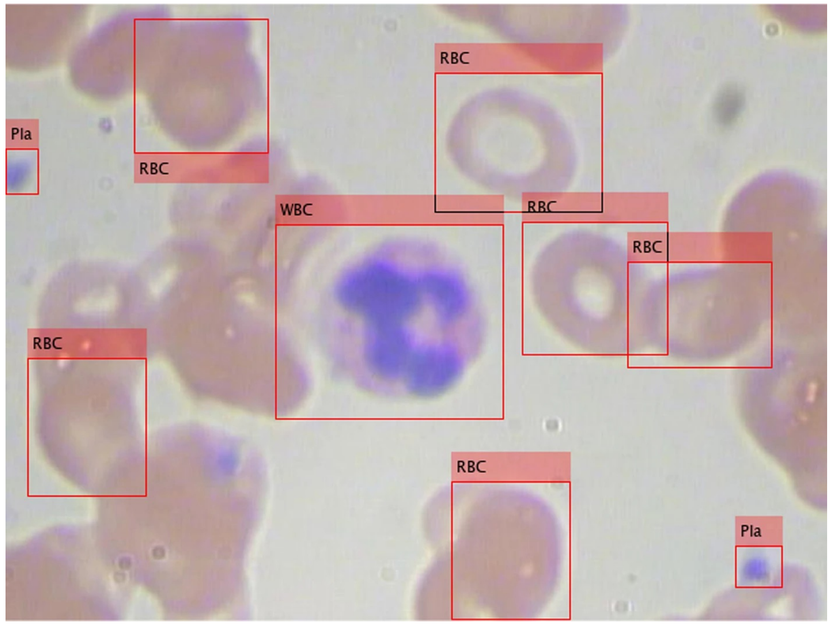


Image from Chen et al., Automatic identifying and counting blood cells in smear images by using single shot detector and Taguchi method, BMC Bioinformatics 22 (Suppl 5), 635 (2021). https://doi.org/10.1186/s12859-022-05074-2

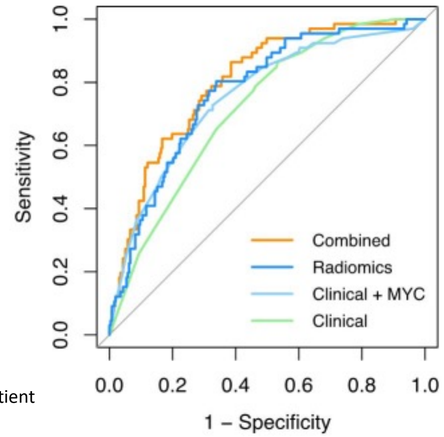
451 Combining PET Radiomic Features with MYC Gene Rearrangement Results in High Prediction of Outcome in Diffuse Large B-Cell Lymphoma

Program: Oral and Poster Abstracts
Type: Oral
Session: 627. Aggressive Lymphomas: Clinical and Epidemiological: Clinical Management of Aggressive B cell NHL
Hematology Disease Topics & Pathways:
Lymphomas, Non-Hodgkin Lymphoma, Clinically Relevant, Diseases, Lymphoid Malignancies, Technology and Procedures, Machine Learning

Sunday, December 12, 2021: 12:00 PM

Jakoba J Eertink, MSc[1*], *Gerben J.C Zwezerijnen, MD*[2*], *Sanne E Wiegers, MSc*[3*], *Martine E.D Chamuleau, MD, PhD*[3], *Pieternella Lugtenburg, MD, PhD*[4], *Daphne De Jong, MD, PhD*[5*], *Bauke Ylstra*[6*], *Matias Mendeville*[5*], *Ulrich Dührsen, Prof, MD, PhD*[7*], *Andreas Hüttmann, MD*[8], *Christine Schmitz, MD*[7*], *Julia Richter, PhD*[9*], *Wolfram Klapper*[9*], *Otto S Hoekstra, Prof, MD, PhD*[2*], *Henrica C.W de Vet, Prof, PhD*[10*], *Ronald Boellaard, PhD*[2*] *and Josée M Zijlstra, MD, PhD*[3]

- **Best clinical model. AUC = 69%**
  - AAStage + LDH + ENinv
- **Best clinical + MYC model. AUC = 71%**
  - MYC + PS + LDH + ENinv
- **Best radiomics model. AUC = 77%**
  - MTV + $Dmax_{bulk}$ + $DSUVpeak_{pat}$ + $Dvol_{pat}$
- **Combined model. AUC = 77%**
  - MYC + LDH + PS + MTV + $Dmax_{bulk}$ + $DSUVpeak_{patient}$ + $DVol_{patient}$



Abbreviations. AUC: Area under curve; AAStage: Ann Arbor Stage; LDH: Lactate Dehydrogenase; EN: Extranodal involvement; PS: WHO Performance Status; MTV: Metabolic Tumor Volume; SUV: standardized uptake value; $Dmax_{bulk}$: maximum distance between the largest lesion and any other lesions; $DSUVpeak_{pat}$: the maximum difference in SUV_{peak} between two lesions; $Dvol_{pat}$: maximum difference in volume between two lesions

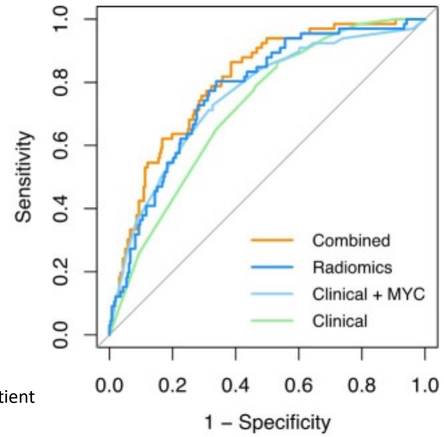
451 Combining PET Radiomic Features with MYC Gene Rearrangement Results in High Prediction of Outcome in Diffuse Large B-Cell Lymphoma

Program: Oral and Poster Abstracts
Type: Oral
Session: 627. Aggressive Lymphomas: Clinical and Epidemiological: Clinical Management of Aggressive B cell NHL
Hematology Disease Topics & Pathways:
Lymphomas, Non-Hodgkin Lymphoma, Clinically Relevant, Diseases, Lymphoid Malignancies, Technology and Procedures, Machine Learning

Sunday, December 12, 2021: 12:00 PM

Jakoba J Eertink, MSc[1], Gerben J.C Zwezerijnen, MD[2*], Sanne E Wiegers, MSc[3*], Martine E.D Chamuleau, MD, PhD[3], Pieternella Lugtenburg, MD, PhD[4], Daphne De Jong, MD, PhD[5*], Bauke Ylstra[6*], Matias Mendeville[5*], Ulrich Dührsen, Prof, MD, PhD[7*], Andreas Hüttmann, MD[8], Christine Schmitz, MD[7*], Julia Richter, PhD[9*], Wolfram Klapper[9*], Otto S Hoekstra, Prof, MD, PhD[2*], Henrica C.W de Vet, Prof, PhD[10*], Ronald Boellaard, PhD[2*] and Josée M Zijlstra, MD, PhD[3]*

- **Best clinical model. AUC = 69%**
  - AAStage + LDH + ENinv
- **Best clinical + MYC model. AUC = 71%**
  - MYC + PS + LDH + ENinv
- **Best radiomics model. AUC = 77%**
  - MTV + $Dmax_{bulk}$ + $DSUVpeak_{pat}$ + $Dvol_{pat}$
- **Combined model. AUC = 77%**
  - MYC + LDH + PS + MTV + $Dmax_{bulk}$ + $DSUVpeak_{patient}$ + $DVol_{patient}$



European Journal of Nuclear Medicine and Molecular Imaging (2022) 49:932–942
https://doi.org/10.1007/s00259-021-05480-3

ORIGINAL ARTICLE

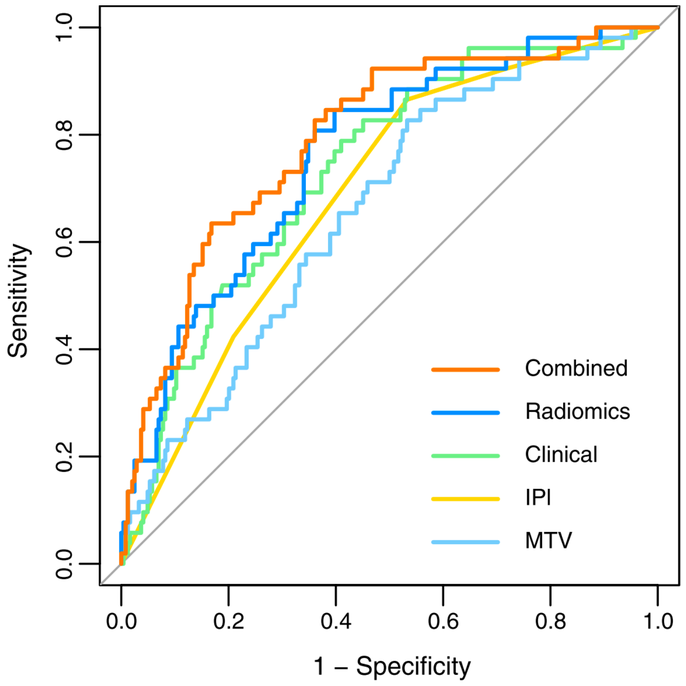
^{18}F-FDG PET baseline radiomics features improve the prediction of treatment outcome in diffuse large B-cell lymphoma

Jakoba J. Eertink[1] · Tim van de Brug[2] · Sanne E. Wiegers[1] · Gerben J. C. Zwezerijnen[3] · Elisabeth A. G. Pfaehler[4] · Pieternella J. Lugtenburg[5] · Bronno van der Holt[6] · Henrica C. W. de Vet[2] · Otto S. Hoekstra[3] · Ronald Boellaard[3] · Josée M. Zijlstra[1]

Table 1 Description of prediction models included in this study

Models	Included features
Model 1: IPI	IPI
Model 2: clinical model	Ann Arbor stage, age, WHO performance status, extranodal involvement, LDH, and bulky disease
Model 3: MTV	MTV
Model 4: limited radiomics model	MTV, SUV_{max}, SUV_{peak}, SUV_{mean}, TLG, number of lesions, $Dmax_{patient}$, $Dmax_{bulk}$, $Spread_{patient}$, $Spread_{bulk}$, and Sphericity
Model 5: all radiomics features (largest and hottest lesion)	485 features for the largest and hottest lesion
Model 6: combined model	Features model 2 and model 4

Abbreviations. AUC: Area under curve; AAStage: Ann Arbor Stage; LDH: Lactate Dehydrogenase; EN: Extranodal involvement; PS: WHO Performance Status; MTV: Metabolic Tumor Volume; SUV: standardized uptake value; $Dmax_{bulk}$: maximum distance between the largest lesion and any other lesions; $DSUVpeak_{pat}$: the maximum difference in SUV_{peak} between two lesions; $Dvol_{pat}$: maximum difference in volume between two lesions



Model	AUC (95%CI)	CV-AUC (95%CI)	Log-likelihood ratio	Specificity	Sensitivity	NPV	PPV
IPI (model 1)	0.68 (0.61–0.75)	0.68 (0.51–0.80)	– 126.11	0.79	0.40	0.86	0.29
Clinical model (model 2)	0.73 (0.66–0.80)	0.71 (0.56–0.86)	– 123.52	0.87	0.38	0.87	0.38
MTV (model 3)	0.66 (0.58–0.74)	0.66 (0.50–0.81)	– 129.96	0.84	0.27	0.84	0.27
Limited radiomics model (model 4)	0.76 (0.69–0.82)	0.75 (0.59–0.88)	– 117.61	0.88	0.44	0.88	0.44
Combined model (model 6)	0.79 (0.73–0.86)	0.77 (0.61–0.90)	– 113.4	0.88	0.44	0.88	0.44

European Journal of Nuclear Medicine and Molecular Imaging (2022) 49:932–942
https://doi.org/10.1007/s00259-021-05480-3

ORIGINAL ARTICLE

^{18}F-FDG PET baseline radiomics features improve the prediction of treatment outcome in diffuse large B-cell lymphoma

Jakoba J. Eertink[1] · Tim van de Brug[2] · Sanne E. Wiegers[1] · Gerben J. C. Zwezerijnen[3] · Elisabeth A. G. Pfaehler[4] · Pieternella J. Lugtenburg[5] · Bronno van der Holt[6] · Henrica C. W. de Vet[2] · Otto S. Hoekstra[3] · Ronald Boellaard[3] · Josée M. Zijlstra[1]

Table 1 Description of prediction models included in this study

Models	Included features
Model 1: IPI	IPI
Model 2: clinical model	Ann Arbor stage, age, WHO performance status, extranodal involvement, LDH, and bulky disease
Model 3: MTV	MTV
Model 4: limited radiomics model	MTV, SUV_{max}, SUV_{peak}, SUV_{mean}, TLG, number of lesions, $Dmax_{patient}$, $Dmax_{bulk}$, $Spread_{patient}$, $Spread_{bulk}$, and Sphericity
Model 5: all radiomics features (largest and hottest lesion)	485 features for the largest and hottest lesion
Model 6: combined model	Features model 2 and model 4

Abbreviations. AUC: Area under curve; AAStage: Ann Arbor Stage; LDH: Lactate Dehydrogenase; EN: Extranodal involvement; PS: WHO Performance Status; MTV: Metabolic Tumor Volume; SUV: standardized uptake value; $Dmax_{bulk}$: maximum distance between the largest lesion and any other lesions; $DSUVpeak_{pat}$: the maximum difference in SUV_{peak} between two lesions; $Dvol_{pat}$: maximum difference in volume between two lesions

original reports

Proposed New Dynamic Prognostic Index for Diffuse Large B-Cell Lymphoma: International Metabolic Prognostic Index

N. George Mikhaeel, MD[1]; Martijn W. Heymans, PhD[2]; Jakoba J. Eertink, PhD[3]; Henrica C.W. de Vet, PhD[2]; Ronald Boellaard, PhD[4] Ulrich Dührsen, MD[5]; Luca Ceriani, MD[6,7]; Christine Schmitz, MD[5]; Sanne E. Wiegers, PhD[3]; Andreas Hüttmann, MD[5]; Pieternella J. Lugtenburg, MD[8]; Emanuele Zucca, MD[6,7]; Gerben J.C. Zwezerijnen, MD[4]; Otto S. Hoekstra, MD[4]; Josée M. Zijlstra, MD[3]; and Sally F. Barrington, MD[9]

European Journal of Nuclear Medicine and Molecular Imaging (2022) 49:932–942
https://doi.org/10.1007/s00259-021-05480-3

ORIGINAL ARTICLE

^{18}F-FDG PET baseline radiomics features improve the prediction of treatment outcome in diffuse large B-cell lymphoma

Jakoba J. Eertink[1] · Tim van de Brug[2] · Sanne E. Wiegers[1] · Gerben J. C. Zwezerijnen[3] · Elisabeth A. G. Pfaehler[4] · Pieternella J. Lugtenburg[5] · Bronno van der Holt[6] · Henrica C. W. de Vet[2] · Otto S. Hoekstra[3] · Ronald Boellaard[3] · Josée M. Zijlstra[1]

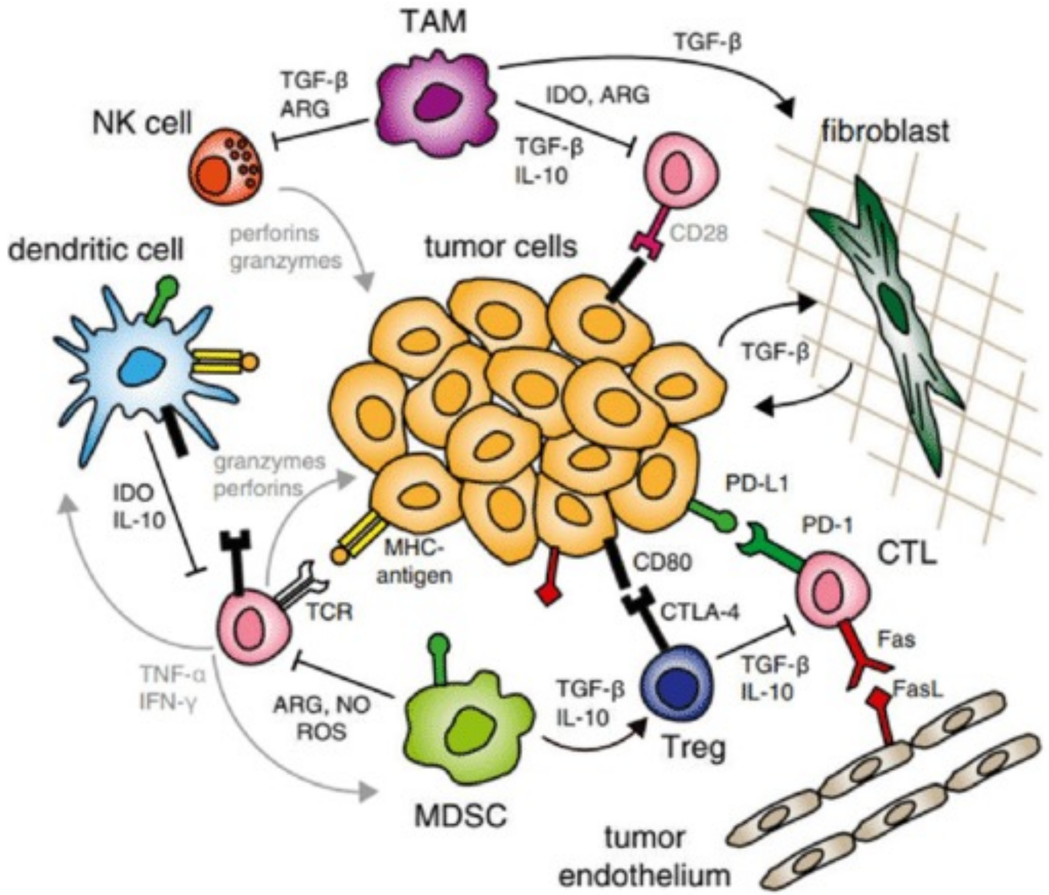
blood

American Society of Hematology
2021 L Street NW, Suite 900,
Washington, DC 20036
Phone: 202-776-0544 | Fax 202-776-0545
editorial@hematology.org

Baseline PET radiomics outperform the IPI risk score for prediction of outcome in diffuse large B-cell lymphoma

Tracking no: BLD-2022-018558R1

TME



TAM: Tumor Associated Macrophage,
CD28, CTL, TCR : CD4/8 T-cell,
Fibroblast,
Treg: regulatory T-Cell,
MDSC: Myeloid Derived Suppressor Cell,
Dendritic Cells,
NK cell: natural killer cells,
....
Adipocytes,
Pericytes,
Artery endothelial cells,
Mast cells,
Neutrophils,
Eosinophils,
Monocytes,
Follicular helper T-cells,
Gamma-delta T-cells,
Plasma cells,
Memory B cells,
Naïve B cells

From Park et al, Biomaterial Research, 2018

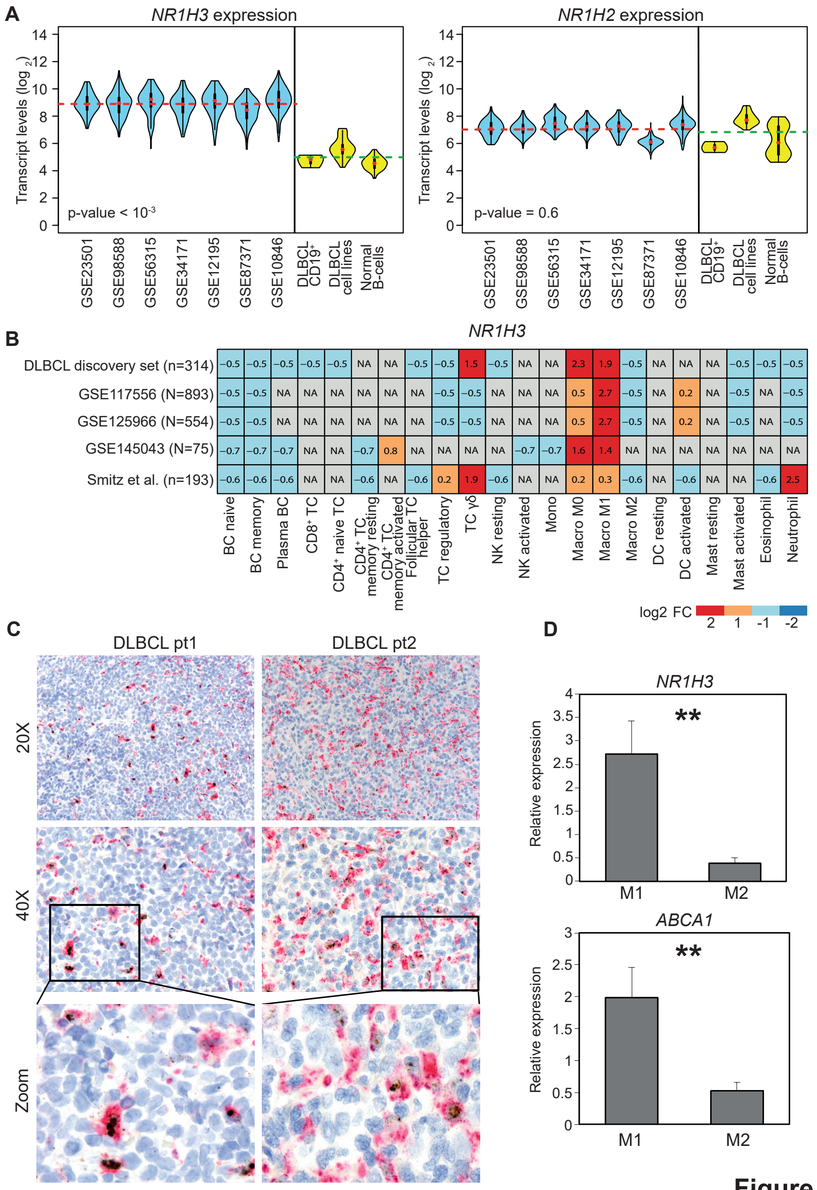


Figure 1

Vegliante et al., Hematological Oncology, 2022

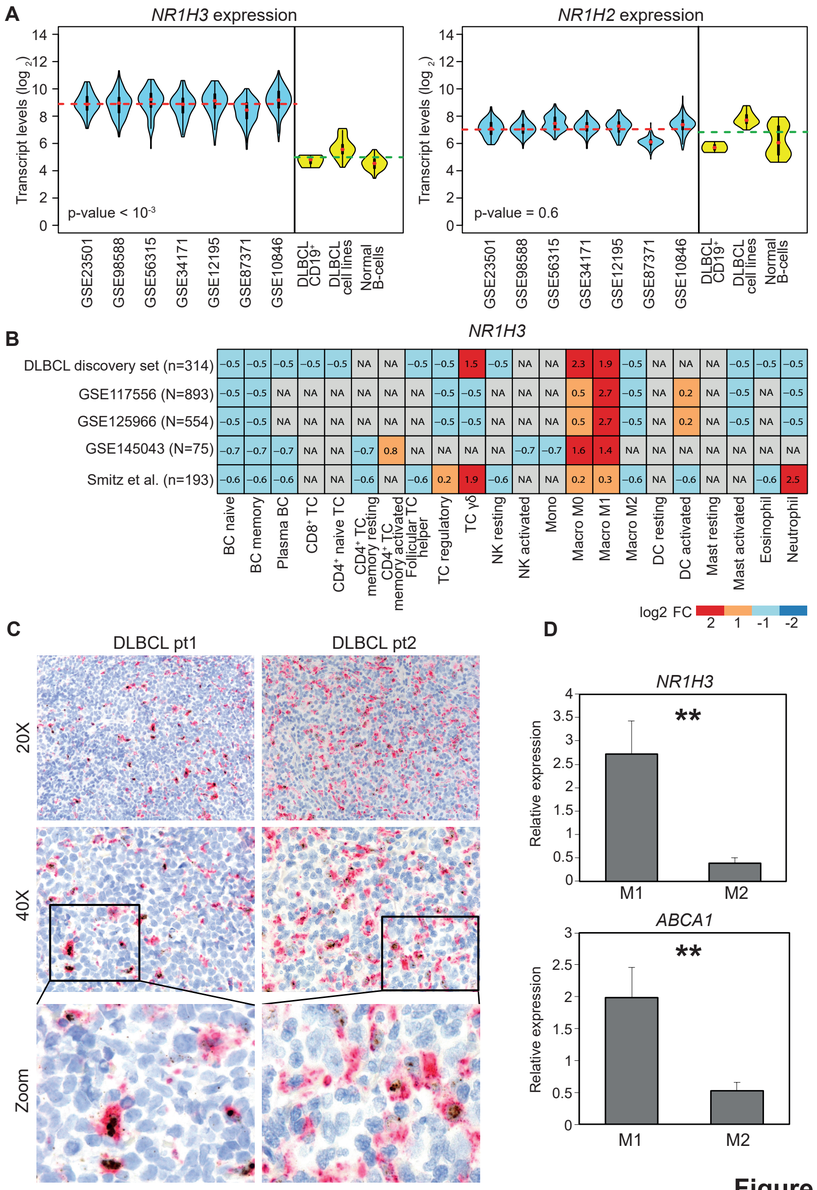
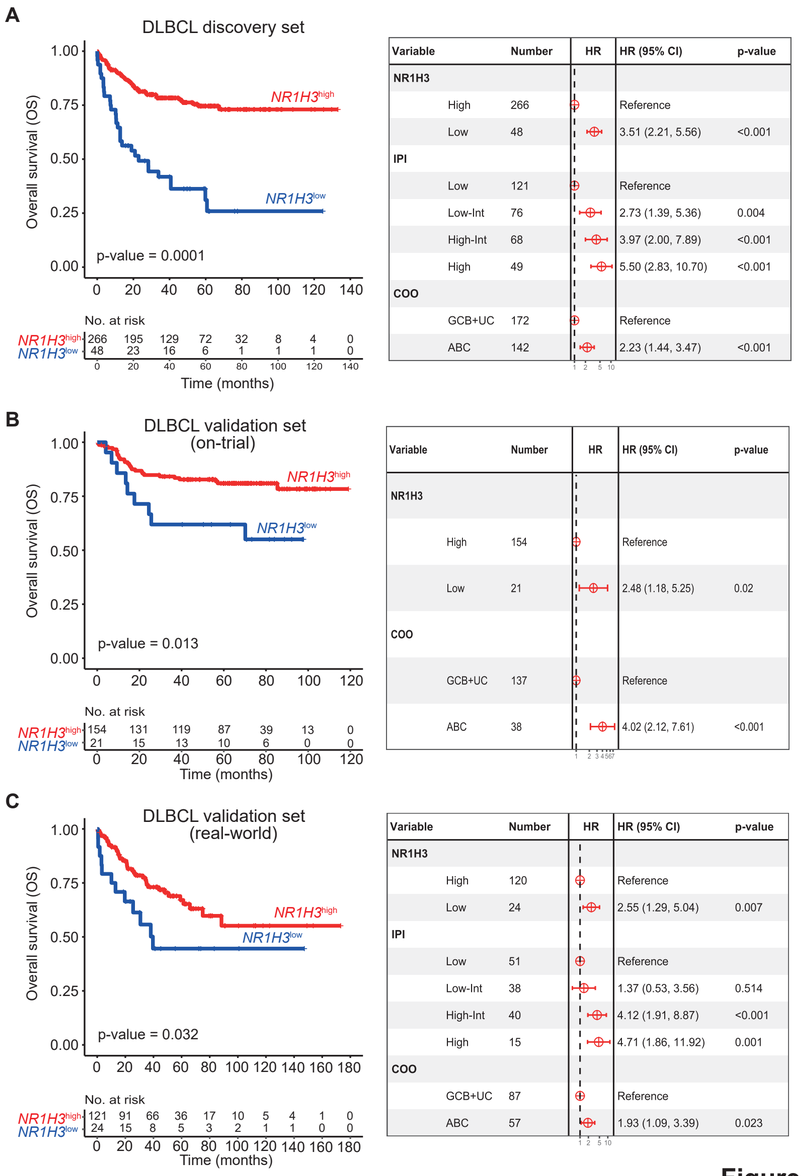


Figure 1



Variable	Number	HR	HR (95% CI)	p-value
NR1H3				
High	266		Reference	
Low	48		3.51 (2.21, 5.56)	<0.001
IPI				
Low	121		Reference	
Low-Int	76		2.73 (1.39, 5.36)	0.004
High-Int	68		3.97 (2.00, 7.89)	<0.001
High	49		5.50 (2.83, 10.70)	<0.001
COO				
GCB+UC	172		Reference	
ABC	142		2.23 (1.44, 3.47)	<0.001

Variable	Number	HR	HR (95% CI)	p-value
NR1H3				
High	154		Reference	
Low	21		2.48 (1.18, 5.25)	0.02
COO				
GCB+UC	137		Reference	
ABC	38		4.02 (2.12, 7.61)	<0.001

Variable	Number	HR	HR (95% CI)	p-value
NR1H3				
High	120		Reference	
Low	24		2.55 (1.29, 5.04)	0.007
IPI				
Low	51		Reference	
Low-Int	38		1.37 (0.53, 3.56)	0.514
High-Int	40		4.12 (1.91, 8.87)	<0.001
High	15		4.71 (1.86, 11.92)	0.001
COO				
GCB+UC	87		Reference	
ABC	57		1.93 (1.09, 3.39)	0.023

Figure 2

Vegliante et al., Hematological Oncology, 2022

Objectives

Study of both clinical and TME variables to combine a novel clinical prognostic index able to define patients' risk

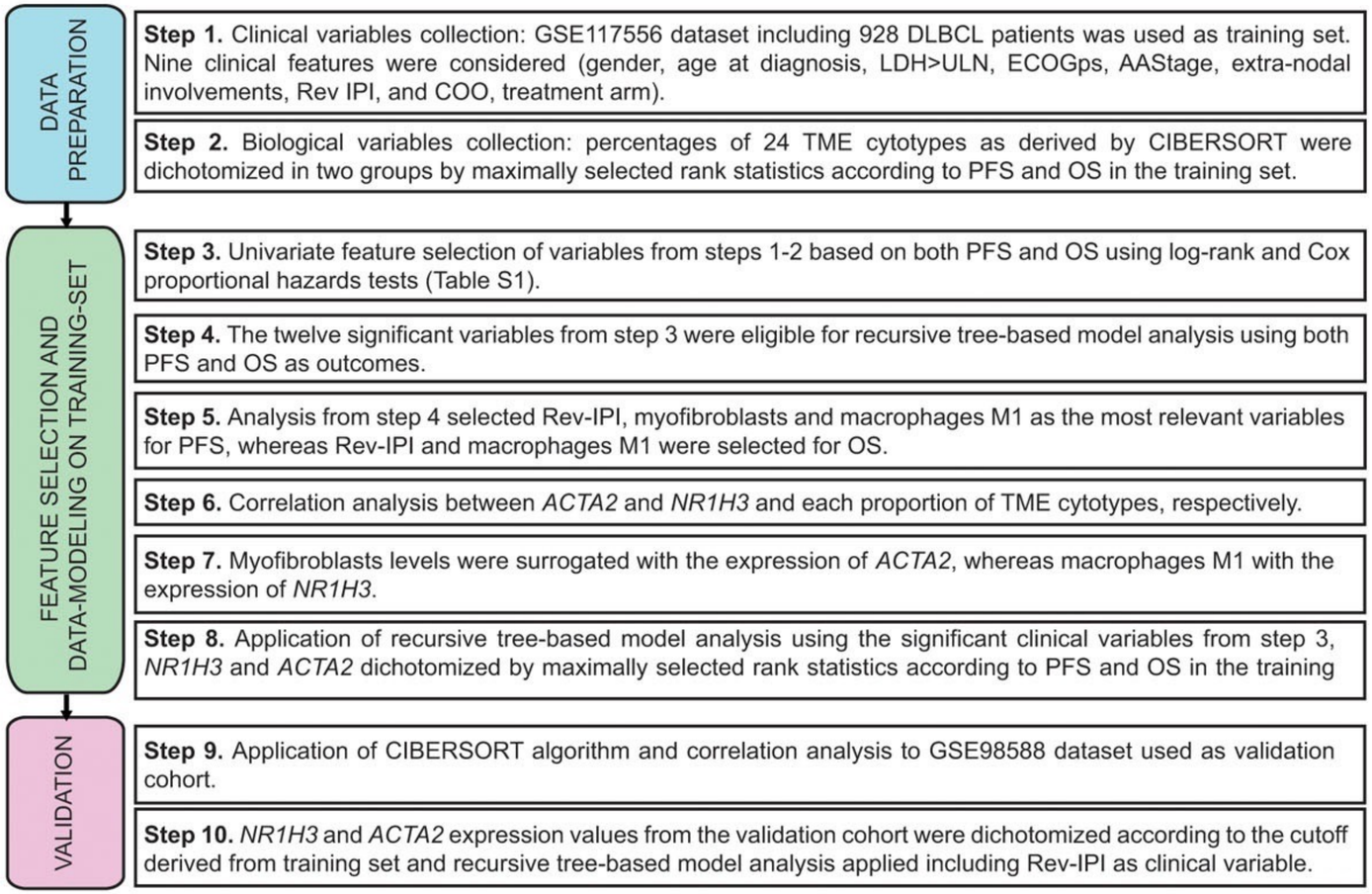
Translational applicability for potential drug discovery

Zaccaria GM, Vegliante MC et al., Hemasphere, 2023

Objectives

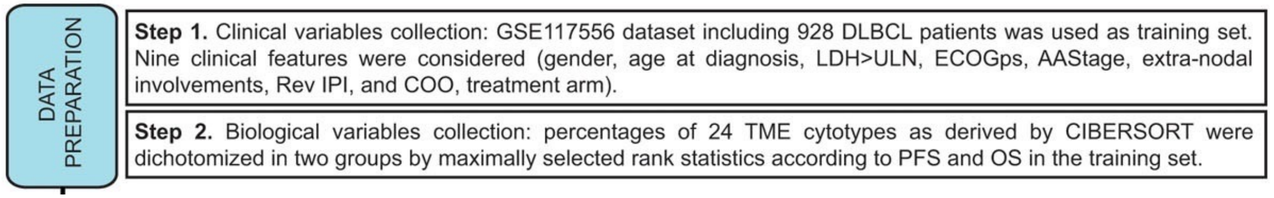
Study of both clinical and TME variables to combine a novel clinical prognostic index able to define patients' risk

Translational applicability for potential drug discovery



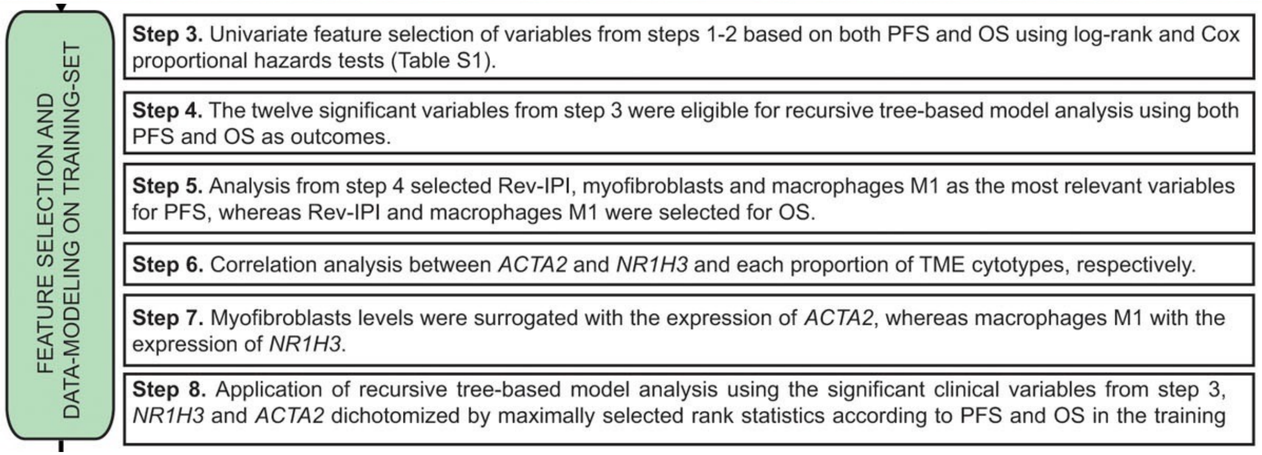
Zaccaria GM, Vegliante MC et al., Hemasphere, 2023

Pipeline



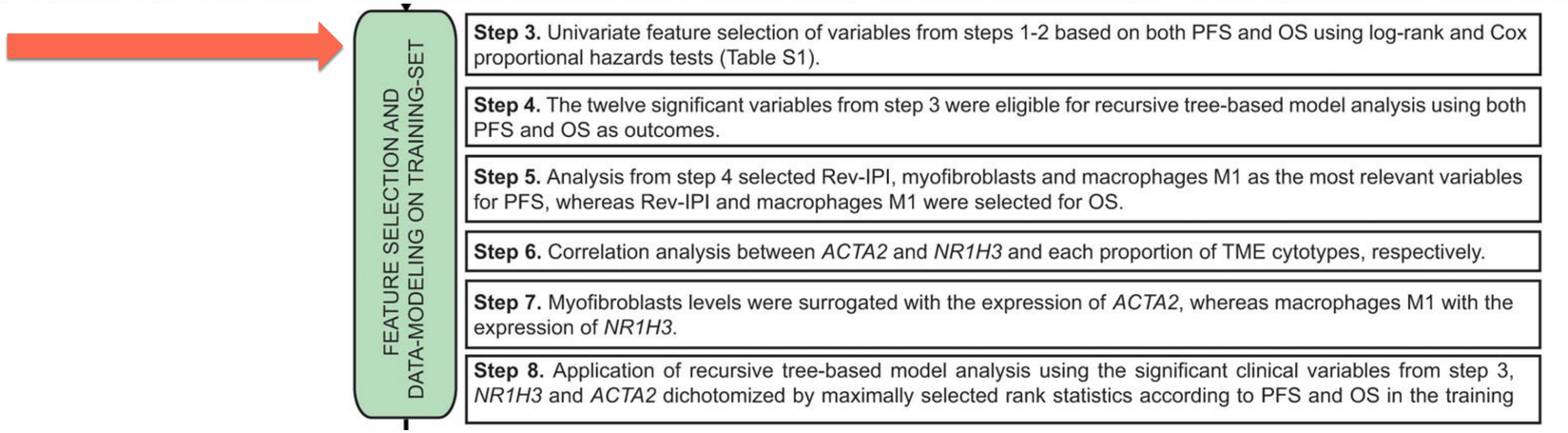
Zaccaria GM, Vegliante MC et al., Hemasphere, 2023

Pipeline



Zaccaria GM, Vegliante MC et al., Hemasphere, 2023

Pipeline



Zaccaria GM, Vegliante MC et al., Hemasphere, 2023

Pipeline

		Log-rank OS	COX proportional hazards OS	Log-rank PFS	COX proportional hazards PFS
		P-value	*P*-value	*P*-value	*P*-value
Clinical variables	Gender	0.6800	0.6300	0.6320	0.6860
	Age	0.0092	0.0010	0.7600	0.7550
	LDH>ULN	0.0001	<0.0001	0.0001	<0.0001
	ECOGps	0.0001	<0.0001	0.0002	0.0002
	AAStage	0.0048	0.0054	0.0018	0.0019
	Extranodal involvements	0.0880	0.0896	0.0100	0.0106
	Rev-IPI	0.0001	<0.0001	0.0001	<0.0001
	COO	0.0360	0.0416	0.0610	0.0681
	Treatment arm	0.6200	0.6180	0.3700	0.3750
Biological variables, including TME	Tumor Cells, GCB	0.0001	<0.0001	0.0063	0.0070
	Tumor Cells, ABC	0.0001	0.0001	0.0050	0.0052
	Naïve B-cells	0.0044	0.0049	0.0009	0.0010
	Memory B-cells	0.0001	<0.0001	<0.0001	<0.0001
	Plasmacells	0.0260	0.0266	0.1600	0.1570
	CD8 T-cells	0.0340	0.0347	0.0360	0.0370
	CD4 T-cells	0.0074	0.0079	0.0150	0.0156
	Gamma-delta T-cells	0.0480	0.0498	0.2700	0.2670
	Follicular helper T-cells	0.0011	0.0017	0.0240	0.0247
	Regulatory T-cells	0.0890	0.0901	0.0360	0.0362
	NK cells resting	0.4600	0.4560	0.1100	0.1100
	NK cells activated	0.0110	0.0119	0.0200	0.0209
	Monocytes	0.0180	0.0185	0.0310	0.0320
	Macrophages M1	0.0001	<0.0001	0.0001	0.0001
	Macrophages M2	0.3800	0.3820	0.3000	0.3060
	Dendritic cells	NA	NA	NA	NA
	Eosinophils	NA	NA	NA	NA
	Neutrophils	0.0052	0.0056	0.0014	0.0015
	Mast cells	0.1500	0.1540	0.0290	0.0298
	Myofibroblasts	0.0003	0.0004	0.0003	0.0004
	Lymphatic endothelial cells	NA	NA	NA	NA
	Artery endothelial cells	0.0001	0.0001	0.0053	0.0056
	Adipocytes	0.0380	0.0392	0.0026	0.0003
	Pericytes	0.0610	0.0644	0.0300	0.0314

Abbreviations. TME: tumor microenvironment; PFS: progression free survival; OS: overall survival; LDH>/≤ULN: lactate dehydrogenase >/□ upper level of normal, ECOGps: eastern cooperative oncology group performance status; AAstage: Ann Arbor Stage; Rev-IPI: Revised International Prognostic Index; COO: Cell of Origin; GCB: Germinal Center B-like; ABC: Activated B-center like; NA: not available.

FEATURE SELECTION AND DATA-MODELING ON TRAINING-SET

Step 3. Univariate feature selection of variables from steps 1-2 based on both PFS and OS using log-rank and Cox proportional hazards tests (Table S1).

Step 4. The twelve significant variables from step 3 were eligible for recursive tree-based model analysis using both PFS and OS as outcomes.

Step 5. Analysis from step 4 selected Rev-IPI, myofibroblasts and macrophages M1 as the most relevant variables for PFS, whereas Rev-IPI and macrophages M1 were selected for OS.

Step 6. Correlation analysis between *ACTA2* and *NR1H3* and each proportion of TME cytotypes, respectively.

Step 7. Myofibroblasts levels were surrogated with the expression of *ACTA2*, whereas macrophages M1 with the expression of *NR1H3*.

Step 8. Application of recursive tree-based model analysis using the significant clinical variables from step 3, *NR1H3* and *ACTA2* dichotomized by maximally selected rank statistics according to PFS and OS in the training

Zaccaria GM, Vegliante MC et al., Hemasphere, 2023

Pipeline

		Log-rank OS	COX proportional hazards OS	Log-rank PFS	COX proportional hazards PFS
		P-value	*P*-value	*P*-value	*P*-value
Clinical variables	Gender	0.6800	0.6300	0.6320	0.6860
	Age	0.0092	0.0010	0.7600	0.7550
	LDH>ULN	0.0001	<0.0001	0.0001	<0.0001
	ECOGps	0.0001	<0.0001	0.0002	0.0002
	AAStage	0.0048	0.0054	0.0018	0.0019
	Extranodal involvements	0.0880	0.0896	0.0100	0.0106
	Rev-IPI	0.0001	<0.0001	0.0001	<0.0001
	COO	0.0360	0.0416	0.0610	0.0681
	Treatment arm	0.6200	0.6180	0.3700	0.3750
Biological variables, including TME	Tumor Cells, GCB	0.0001	<0.0001	0.0063	0.0070
	Tumor Cells, ABC	0.0001	0.0001	0.0050	0.0052
	Naïve B-cells	0.0044	0.0049	0.0009	0.0010
	Memory B-cells	0.0001	<0.0001	<0.0001	<0.0001
	Plasmacells	0.0260	0.0266	0.1600	0.1570
	CD8 T-cells	0.0340	0.0347	0.0360	0.0370
	CD4 T-cells	0.0074	0.0079	0.0150	0.0156
	Gamma-delta T-cells	0.0480	0.0498	0.2700	0.2670
	Follicular helper T-cells	0.0011	0.0017	0.0240	0.0247
	Regulatory T-cells	0.0890	0.0901	0.0360	0.0362
	NK cells resting	0.4600	0.4560	0.1100	0.1100
	NK cells activated	0.0110	0.0119	0.0200	0.0209
	Monocytes	0.0180	0.0185	0.0310	0.0320
	Macrophages M1	0.0001	<0.0001	0.0001	0.0001
	Macrophages M2	0.3800	0.3820	0.3000	0.3060
	Dendritic cells	NA	NA	NA	NA
	Eosinophils	NA	NA	NA	NA
	Neutrophils	0.0052	0.0056	0.0014	0.0015
	Mast cells	0.1500	0.1540	0.0290	0.0298
	Myofibroblasts	0.0003	0.0004	0.0003	0.0004
	Lymphatic endothelial cells	NA	NA	NA	NA
	Artery endothelial cells	0.0001	0.0001	0.0053	0.0056
	Adipocytes	0.0380	0.0392	0.0026	0.0003
	Pericytes	0.0610	0.0644	0.0300	0.0314

Abbreviations. TME: tumor microenvironment; PFS: progression free survival; OS: overall survival; LDH>/≤ULN: lactate dehydrogenase >/□ upper level of normal, ECOGps: eastern cooperative oncology group performance status; AAstage: Ann Arbor Stage; Rev-IPI: Revised International Prognostic Index; COO: Cell of Origin; GCB: Germinal Center B-like; ABC: Activated B-center like; NA: not available.

FEATURE SELECTION AND DATA-MODELING ON TRAINING-SET

Step 3. Univariate feature selection of variables from steps 1-2 based on both PFS and OS using log-rank and Cox proportional hazards tests (Table S1).

Step 4. The twelve significant variables from step 3 were eligible for recursive tree-based model analysis using both PFS and OS as outcomes.

Step 5. Analysis from step 4 selected Rev-IPI, myofibroblasts and macrophages M1 as the most relevant variables for PFS, whereas Rev-IPI and macrophages M1 were selected for OS.

Step 6. Correlation analysis between *ACTA2* and *NR1H3* and each proportion of TME cytotypes, respectively.

Step 7. Myofibroblasts levels were surrogated with the expression of *ACTA2*, whereas macrophages M1 with the expression of *NR1H3*.

Step 8. Application of recursive tree-based model analysis using the significant clinical variables from step 3, *NR1H3* and *ACTA2* dichotomized by maximally selected rank statistics according to PFS and OS in the training

Zaccaria GM, Vegliante MC et al., Hemasphere, 2023

Pipeline

		Log-rank OS	COX proportional hazards OS	Log-rank PFS	COX proportional hazards PFS
		P-value	*P*-value	*P*-value	*P*-value
Clinical variables	Gender	0.6800	0.6300	0.6320	0.6860
	Age	0.0092	0.0010	0.7600	0.7550
	LDH>ULN	0.0001	<0.0001	0.0001	<0.0001
	ECOGps	0.0001	<0.0001	0.0002	0.0002
	AAStage	0.0048	0.0054	0.0018	0.0019
	Extranodal involvements	0.0880	0.0896	0.0100	0.0106
	Rev-IPI	0.0001	<0.0001	0.0001	<0.0001
	COO	0.0360	0.0416	0.0610	0.0681
	Treatment arm	0.6200	0.6180	0.3700	0.3750
Biological variables, including TME	Tumor Cells, GCB	0.0001	<0.0001	0.0063	0.0070
	Tumor Cells, ABC	0.0001	0.0001	0.0050	0.0052
	Naïve B-cells	0.0044	0.0049	0.0009	0.0010
	Memory B-cells	0.0001	<0.0001	<0.0001	<0.0001
	Plasmacells	0.0260	0.0266	0.1600	0.1570
	CD8 T-cells	0.0340	0.0347	0.0360	0.0370
	CD4 T-cells	0.0074	0.0079	0.0150	0.0156
	Gamma-delta T-cells	0.0480	0.0498	0.2700	0.2670
	Follicular helper T-cells	0.0011	0.0017	0.0240	0.0247
	Regulatory T-cells	0.0890	0.0901	0.0360	0.0362
	NK cells resting	0.4600	0.4560	0.1100	0.1100
	NK cells activated	0.0110	0.0119	0.0200	0.0209
	Monocytes	0.0180	0.0185	0.0310	0.0320
	Macrophages M1	0.0001	<0.0001	0.0001	0.0001
	Macrophages M2	0.3800	0.3820	0.3000	0.3060
	Dendritic cells	NA	NA	NA	NA
	Eosinophils	NA	NA	NA	NA
	Neutrophils	0.0052	0.0056	0.0014	0.0015
	Mast cells	0.1500	0.1540	0.0290	0.0298
	Myofibroblasts	0.0003	0.0004	0.0003	0.0004
	Lymphatic endothelial cells	NA	NA	NA	NA
	Artery endothelial cells	0.0001	0.0001	0.0053	0.0056
	Adipocytes	0.0380	0.0392	0.0026	0.0003
	Pericytes	0.0610	0.0644	0.0300	0.0314

Abbreviations. TME: tumor microenvironment; PFS: progression free survival; OS: overall survival; LDH>/≤ULN: lactate dehydrogenase >/□ upper level of normal, ECOGps: eastern cooperative oncology group performance status; AAstage: Ann Arbor Stage; Rev-IPI: Revised International Prognostic Index; COO: Cell of Origin; GCB: Germinal Center B-like; ABC: Activated B-center like; NA: not available.

FEATURE SELECTION AND DATA-MODELING ON TRAINING-SET

Step 3. Univariate feature selection of variables from steps 1-2 based on both PFS and OS using log-rank and Cox proportional hazards tests (Table S1).

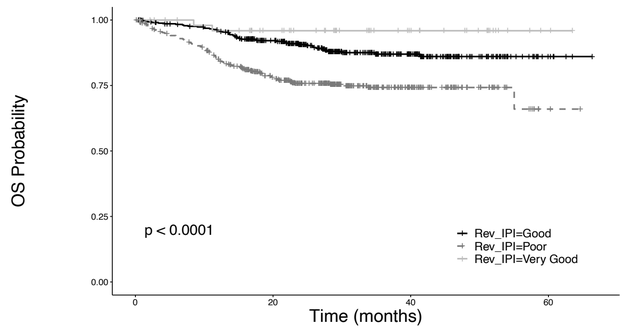
Step 4. The twelve significant variables from step 3 were eligible for recursive tree-based model analysis using both PFS and OS as outcomes.

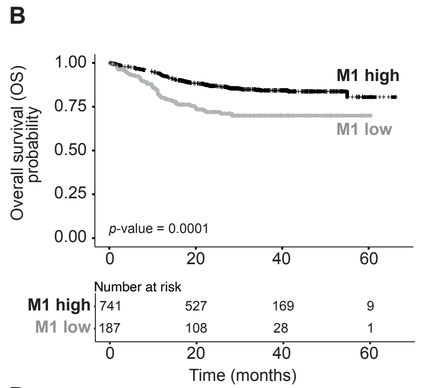
Step 5. Analysis from step 4 selected Rev-IPI, myofibroblasts and macrophages M1 as the most relevant variables for PFS, whereas Rev-IPI and macrophages M1 were selected for OS.

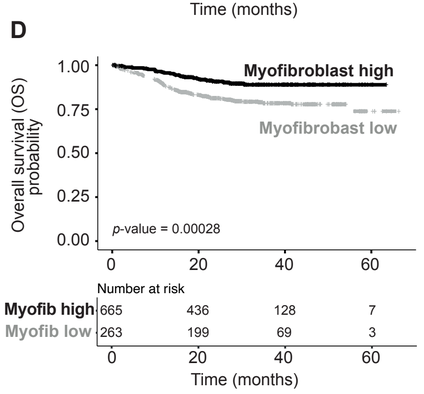
Step 6. Correlation analysis between *ACTA2* and *NR1H3* and each proportion of TME cytotypes, respectively.

Step 7. Myofibroblasts levels were surrogated with the expression of *ACTA2*, whereas macrophages M1 with the expression of *NR1H3*.

Step 8. Application of recursive tree-based model analysis using the significant clinical variables from step 3, *NR1H3* and *ACTA2* dichotomized by maximally selected rank statistics according to PFS and OS in the training

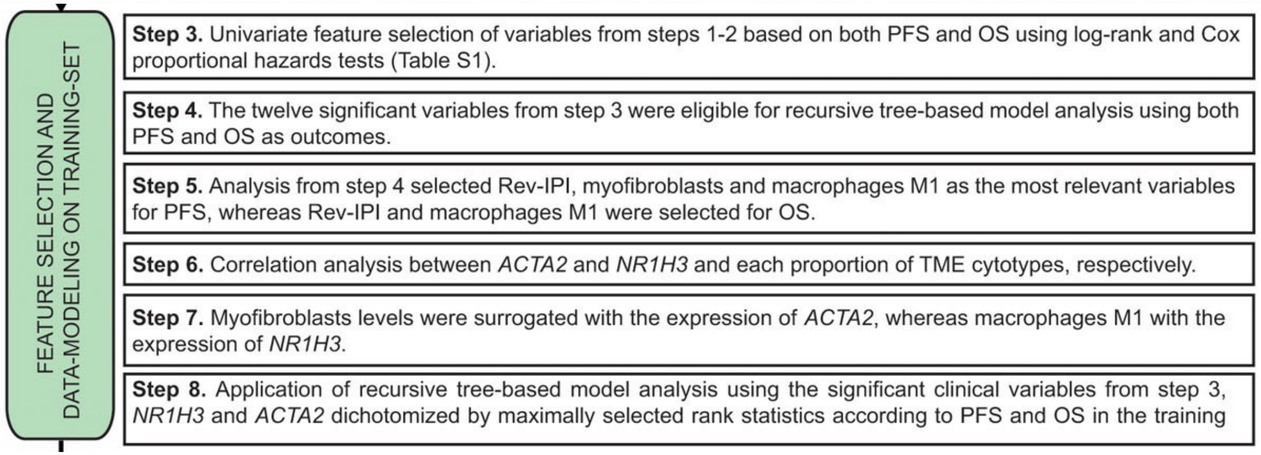






Zaccaria GM, Vegliante MC et al., Hemasphere, 2023

Pipeline



Zaccaria GM, Vegliante MC et al., Hemasphere, 2023

Pipeline

FEATURE SELECTION AND DATA-MODELING ON TRAINING-SET

Step 3. Univariate feature selection of variables from steps 1-2 based on both PFS and OS using log-rank and Cox proportional hazards tests (Table S1).

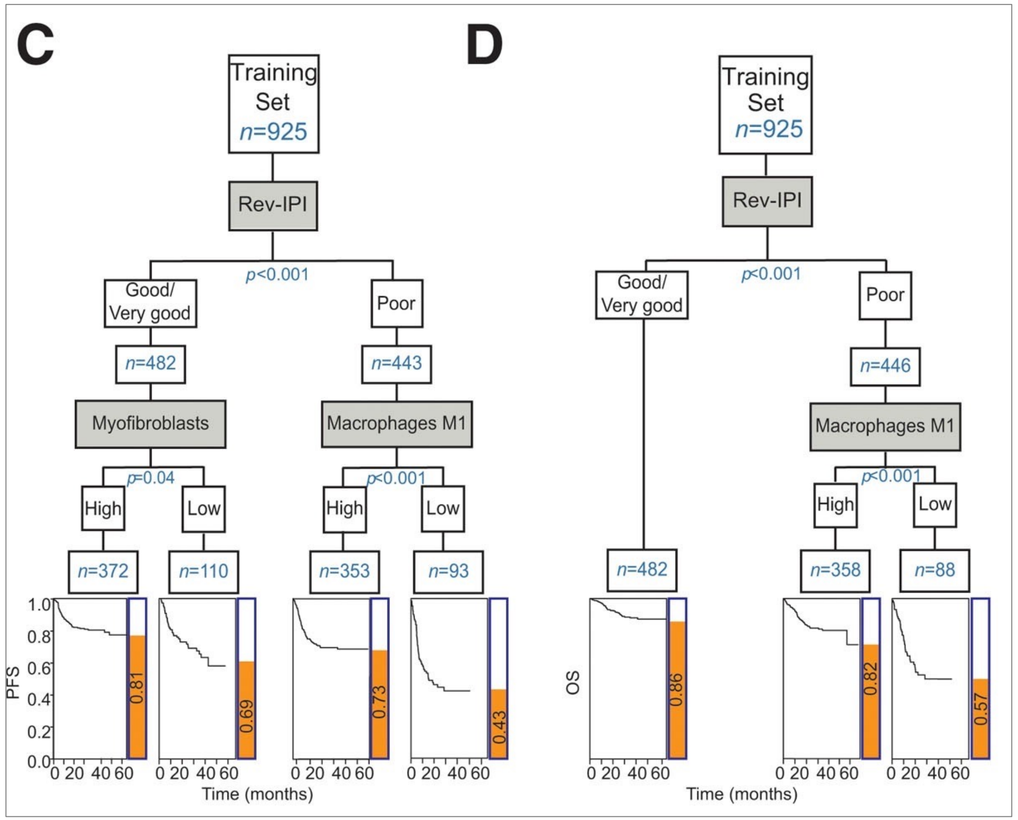
Step 4. The twelve significant variables from step 3 were eligible for recursive tree-based model analysis using both PFS and OS as outcomes.

Step 5. Analysis from step 4 selected Rev-IPI, myofibroblasts and macrophages M1 as the most relevant variables for PFS, whereas Rev-IPI and macrophages M1 were selected for OS.

Step 6. Correlation analysis between *ACTA2* and *NR1H3* and each proportion of TME cytotypes, respectively.

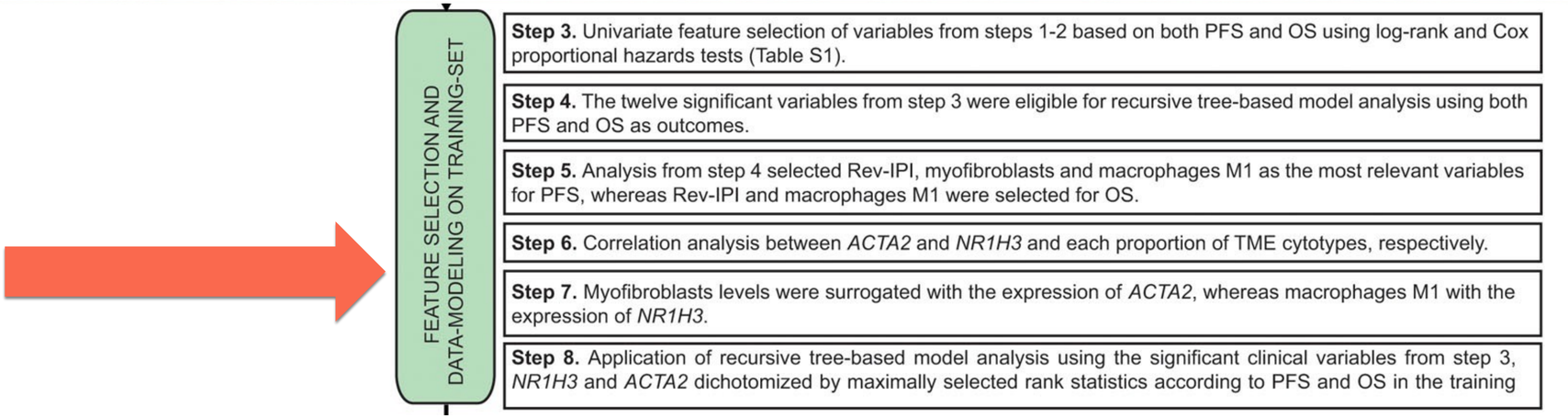
Step 7. Myofibroblasts levels were surrogated with the expression of *ACTA2*, whereas macrophages M1 with the expression of *NR1H3*.

Step 8. Application of recursive tree-based model analysis using the significant clinical variables from step 3, *NR1H3* and *ACTA2* dichotomized by maximally selected rank statistics according to PFS and OS in the training



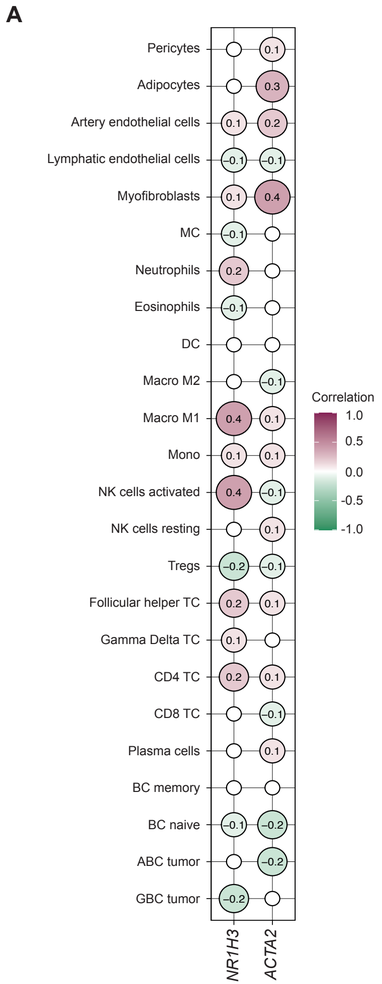
Zaccaria GM, Vegliante MC et al., Hemasphere, 2023

Pipeline



Zaccaria GM, Vegliante MC et al., Hemasphere, 2023

Pipeline



FEATURE SELECTION AND DATA-MODELING ON TRAINING-SET

Step 3. Univariate feature selection of variables from steps 1-2 based on both PFS and OS using log-rank and Cox proportional hazards tests (Table S1).

Step 4. The twelve significant variables from step 3 were eligible for recursive tree-based model analysis using both PFS and OS as outcomes.

Step 5. Analysis from step 4 selected Rev-IPI, myofibroblasts and macrophages M1 as the most relevant variables for PFS, whereas Rev-IPI and macrophages M1 were selected for OS.

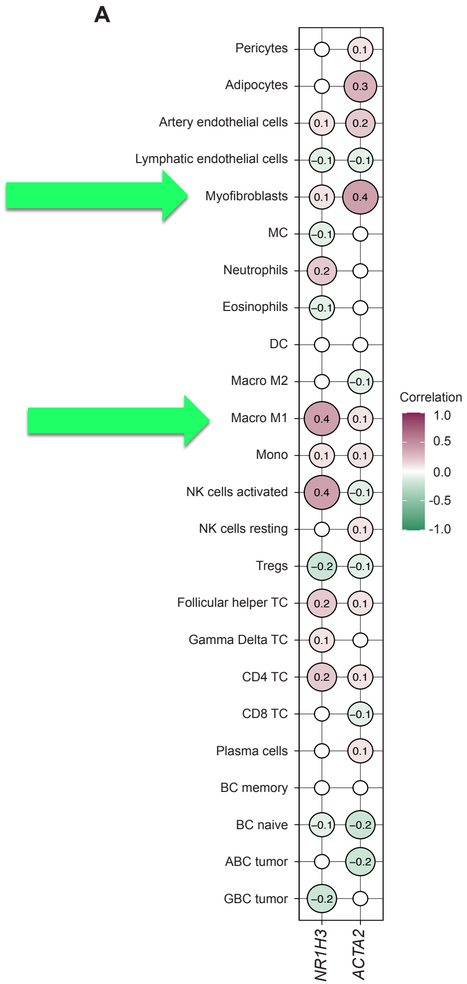
Step 6. Correlation analysis between *ACTA2* and *NR1H3* and each proportion of TME cytotypes, respectively.

Step 7. Myofibroblasts levels were surrogated with the expression of *ACTA2*, whereas macrophages M1 with the expression of *NR1H3*.

Step 8. Application of recursive tree-based model analysis using the significant clinical variables from step 3, *NR1H3* and *ACTA2* dichotomized by maximally selected rank statistics according to PFS and OS in the training

Zaccaria GM, Vegliante MC et al., Hemasphere, 2023

Pipeline



FEATURE SELECTION AND DATA-MODELING ON TRAINING-SET

Step 3. Univariate feature selection of variables from steps 1-2 based on both PFS and OS using log-rank and Cox proportional hazards tests (Table S1).

Step 4. The twelve significant variables from step 3 were eligible for recursive tree-based model analysis using both PFS and OS as outcomes.

Step 5. Analysis from step 4 selected Rev-IPI, myofibroblasts and macrophages M1 as the most relevant variables for PFS, whereas Rev-IPI and macrophages M1 were selected for OS.

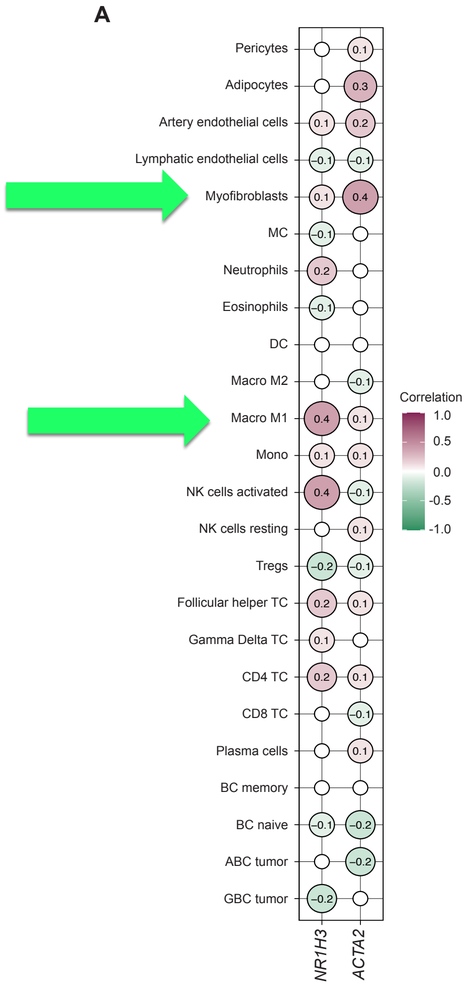
Step 6. Correlation analysis between *ACTA2* and *NR1H3* and each proportion of TME cytotypes, respectively.

Step 7. Myofibroblasts levels were surrogated with the expression of *ACTA2*, whereas macrophages M1 with the expression of *NR1H3*.

Step 8. Application of recursive tree-based model analysis using the significant clinical variables from step 3, *NR1H3* and *ACTA2* dichotomized by maximally selected rank statistics according to PFS and OS in the training

Zaccaria GM, Vegliante MC et al., Hemasphere, 2023

Pipeline



FEATURE SELECTION AND DATA-MODELING ON TRAINING-SET

Step 3. Univariate feature selection of variables from steps 1-2 based on both PFS and OS using log-rank and Cox proportional hazards tests (Table S1).

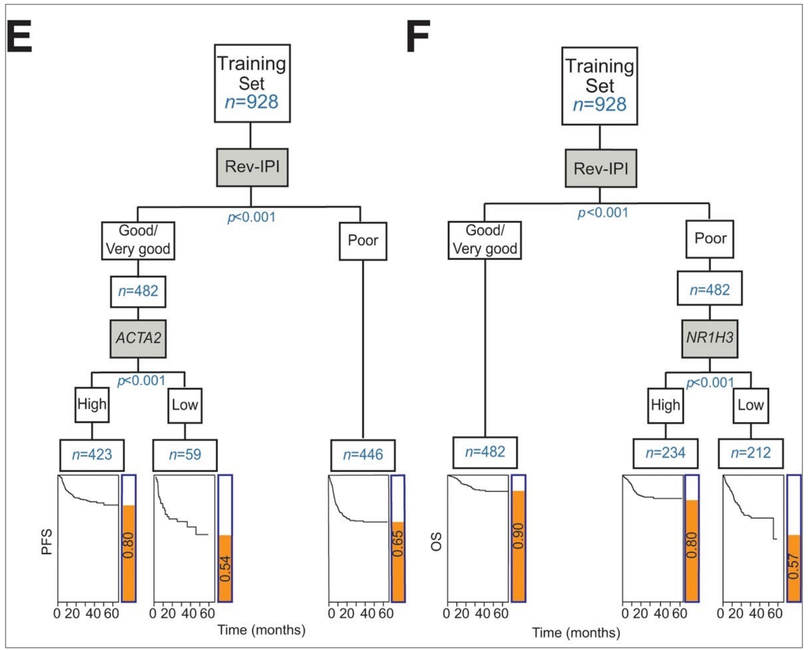
Step 4. The twelve significant variables from step 3 were eligible for recursive tree-based model analysis using both PFS and OS as outcomes.

Step 5. Analysis from step 4 selected Rev-IPI, myofibroblasts and macrophages M1 as the most relevant variables for PFS, whereas Rev-IPI and macrophages M1 were selected for OS.

Step 6. Correlation analysis between *ACTA2* and *NR1H3* and each proportion of TME cytotypes, respectively.

Step 7. Myofibroblasts levels were surrogated with the expression of *ACTA2*, whereas macrophages M1 with the expression of *NR1H3*.

Step 8. Application of recursive tree-based model analysis using the significant clinical variables from step 3, *NR1H3* and *ACTA2* dichotomized by maximally selected rank statistics according to PFS and OS in the training



Zaccaria GM, Vegliante MC et al., Hemasphere, 2023

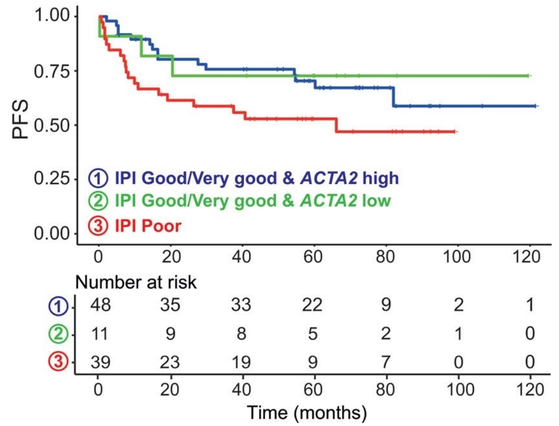
Pipeline

VALIDATION

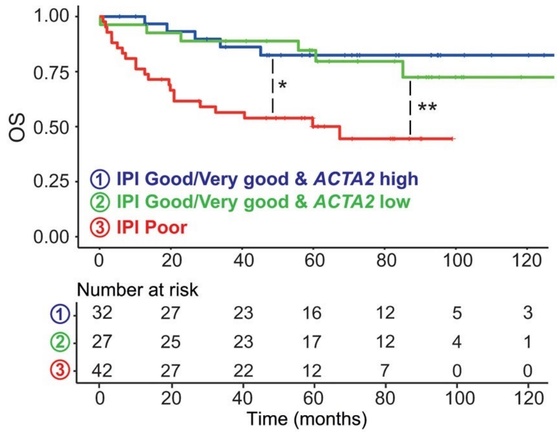
Step 9. Application of CIBERSORT algorithm and correlation analysis to GSE98588 dataset used as validation cohort.

Step 10. *NR1H3* and *ACTA2* expression values from the validation cohort were dichotomized according to the cutoff derived from training set and recursive tree-based model analysis applied including Rev-IPI as clinical variable.

G



I



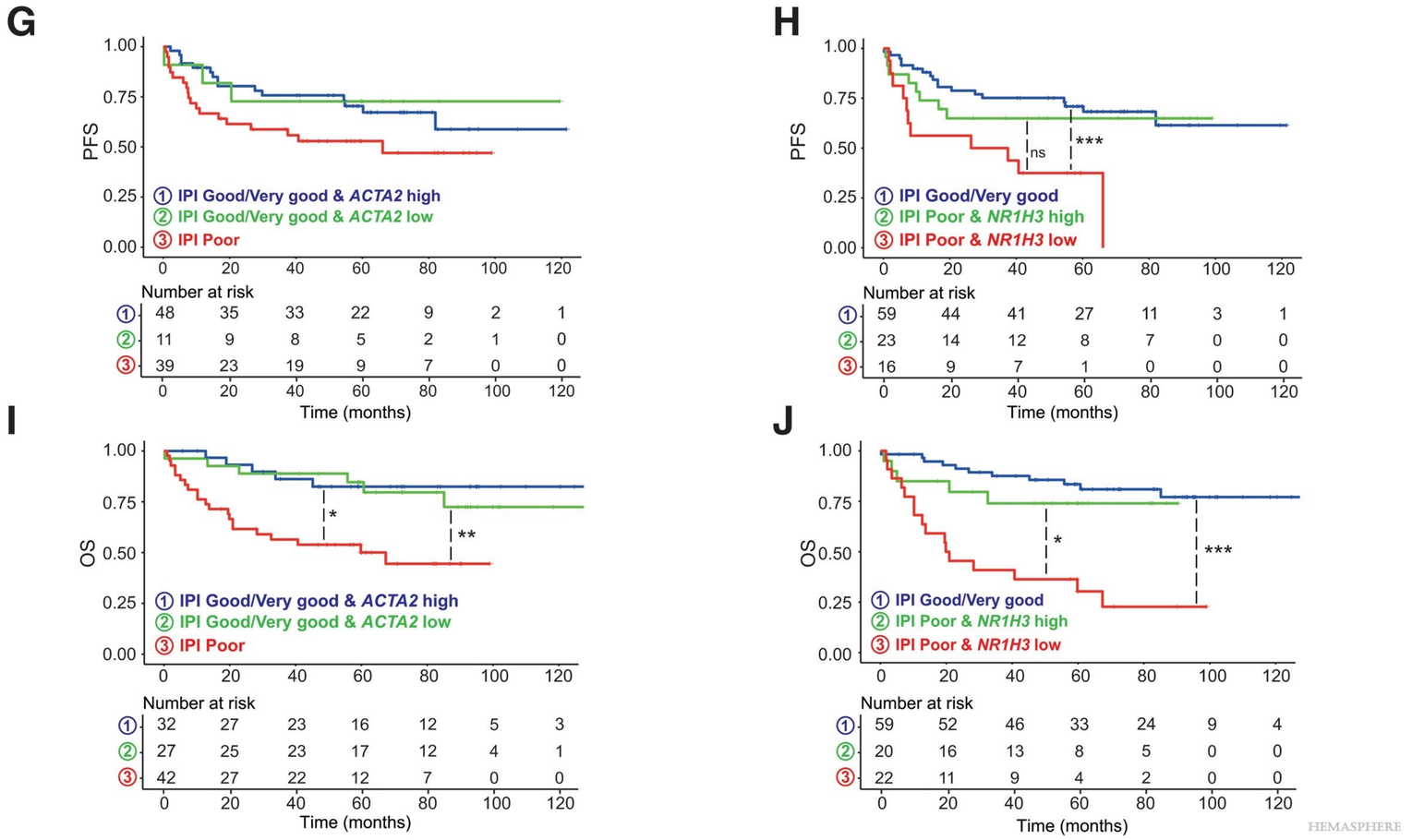
Zaccaria GM, Vegliante MC et al., Hemasphere, 2023

Pipeline

VALIDATION

Step 9. Application of CIBERSORT algorithm and correlation analysis to GSE98588 dataset used as validation cohort.

Step 10. *NR1H3* and *ACTA2* expression values from the validation cohort were dichotomized according to the cutoff derived from training set and recursive tree-based model analysis applied including Rev-IPI as clinical variable.



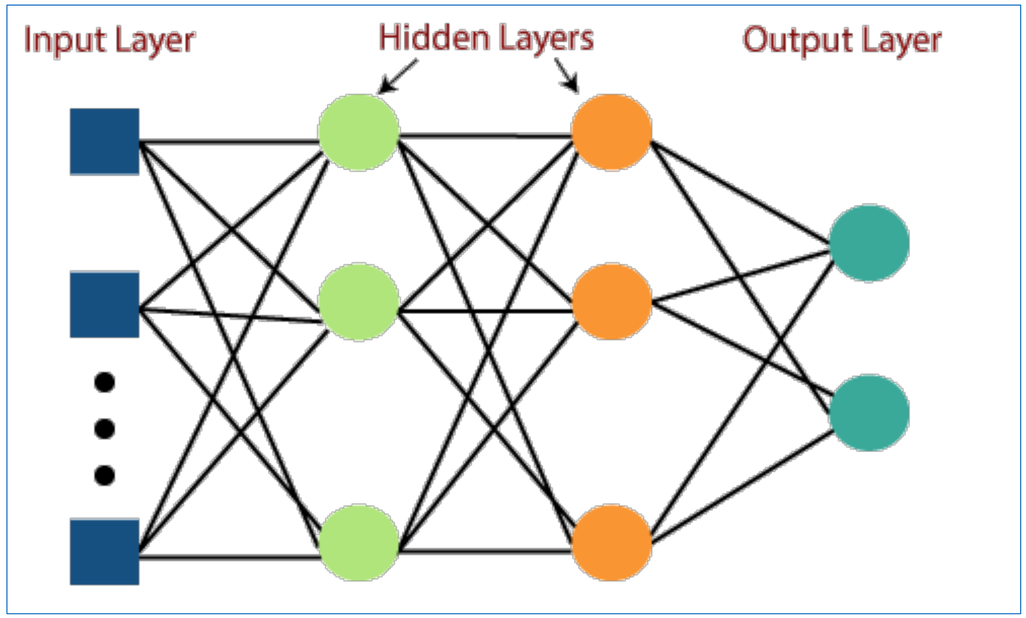
HEMASPHERE

Zaccaria GM, Vegliante MC et al., Hemasphere, 2023

Explainable Artificial Intelligence - XAI

Explainable Artificial Intelligence - XAI

ARTIFICIAL NEURAL NETWORK



PERSPECTIVE
https://doi.org/10.1038/s42256-019-0048-x
nature machine intelligence

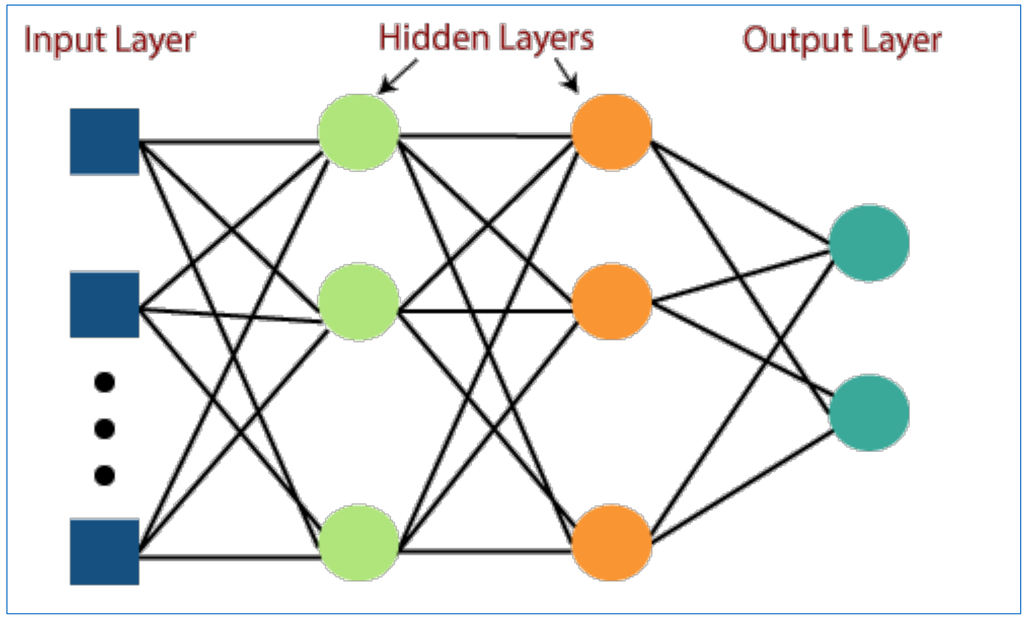
Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead

Cynthia Rudin

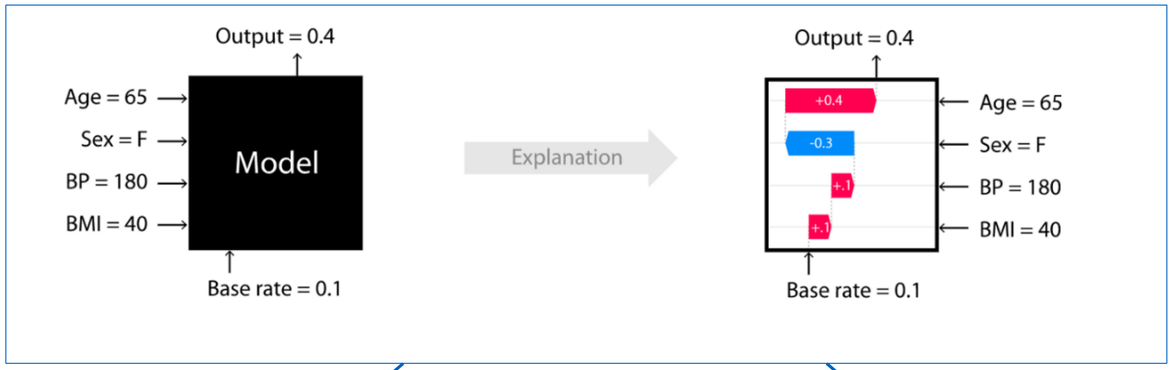
Black box machine learning models are currently being used for high-stakes decision making throughout society, causing problems in healthcare, criminal justice and other domains. Some people hope that creating methods for explaining these black box models will alleviate some of the problems, but trying to explain black box models, rather than creating models that are interpretable in the first place, is likely to perpetuate bad practice and can potentially cause great harm to society. The way forward is to design models that are inherently interpretable. This Perspective clarifies the chasm between explaining black boxes and using inherently interpretable models, outlines several key reasons why explainable black boxes should be avoided in high-stakes decisions, identifies challenges to interpretable machine learning, and provides several example applications where interpretable models could potentially replace black box models in criminal justice, healthcare and computer vision.

Explainable Artificial Intelligence - XAI

ARTIFICIAL NEURAL NETWORK



XAI TECHNIQUES



LIME

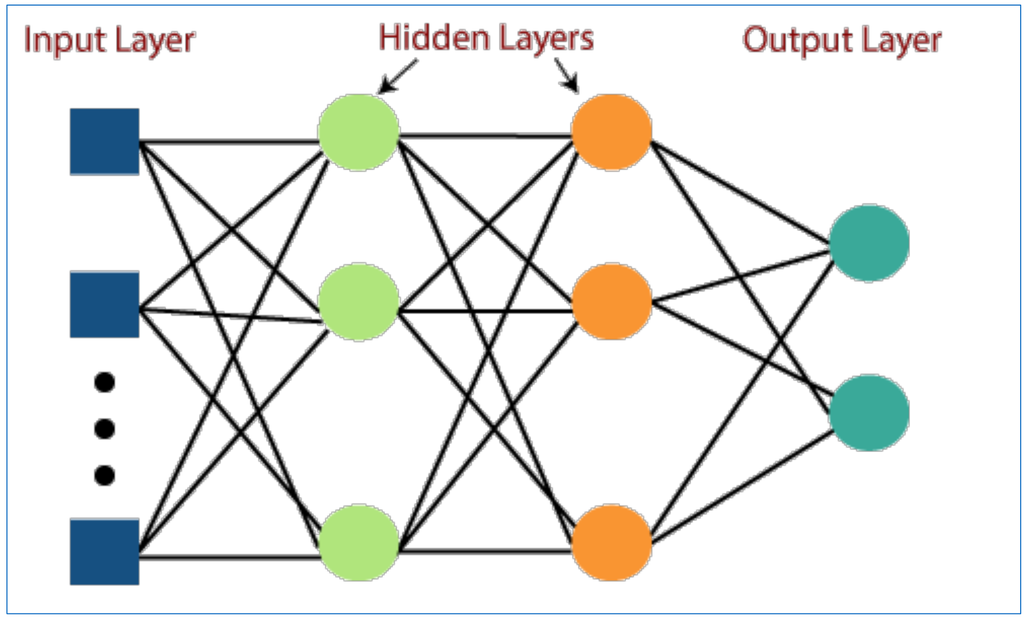
Local Interpretable Model-Agnostic Explanations

SHAP

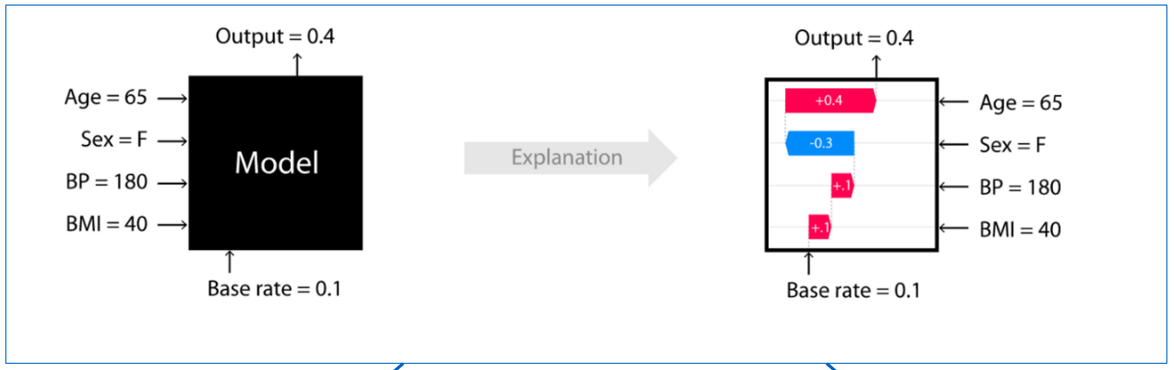
SHapley Additive exPlanations

Explainable Artificial Intelligence - XAI

ARTIFICIAL NEURAL NETWORK



XAI TECHNIQUES



LIME

Local Interpretable Model-Agnostic Explanations

SHAP

SHapley Additive exPlanations

Innovative techniques of gene selection via XAI for prognostic purposes

Conclusions

- Machine Learning can help in prognostication
- ML-based Applications in hematology are increasing
- Supervised and unsupervised ML must be carefully handled to solve a research problem with a robust pipeline
- Simple ML tools as decisional trees might be useful to answer to simple research problems as combination of clinical and TME determinants
- XAI (Explainable Artificial Intelligence) might be potential in boosting the achievement of novel findings

Acknowledgements

Hematology Unit
IRCCS Istituto Tumori
'Giovanni Paolo II'
Dr. Attilio Guarini
Dr. Sabino Ciavarella
Dr. Maria Stella De Candia
Dr. Vincenza De Fazio
Dr. Paolo Ditonno
Dr. Doriana Gramegna
Dr. Angela Iacobazzi
Dr. Filomena Emanuela Laddaga
Dr. Giacomo Loseto
Dr. Pierluigi Masciopinto
Dr. Carla Minoia
Dr. Crescenza Pasciolla
Dr. Angela Maria Quinto
Dr. Antonello Rana
Dr. Bernardo Rossini
Dr. Tetiana Skrypets
Dr. Lorenzo Tonialini

Laboratory Staff
Dr. Antonella Bucci
Dr. Paolo Mondelli
Dr. Antonio Negri
Dr. Susanna Anita Pappagallo
Dr. Antonio Radesco
Dr. Maria Carmela Vegliante
Dr. Antonio Ventura
Dr. Giacomo Volpe

Data-managers
Dr. Felice Clemente
Dr. Alessandro Passiatore
Dr. Fabio Pavone

Department of Pathology
Dr. Alfredo Zito
Dr. Anna Scattone

Polytechnic University of Bari
Prof. Vitoantonio Bevilacqua
Prof. Stefano Mazzoleni
Dr. Nicola Altini
Dr. Antonio Brunetti
Dr. Domenico Buongiorno
Dr. Daniele Di Mascolo

University "Aldo Moro" of Bari
Prof. Nicoletta del Buono
Dr. Grazia Gargano
Dr. Flavia Esposito
Dr. Laura Selicato