Advances and Innovations in Endoscopic Oncology and Multidisciplinary Gastrointestinal Cancer Care

Artificial Intelligence in Oncology: Prime Time

Jacob Shreve, MD, MS

Fellow, Medical Oncology and Hematology

Mayo Clinic, Rochester

Disclosures

- I do not have any relevant financial relationships.

This presentation and/or comments will provide a balanced, non-promotional, and evidence-based approach to all diagnostic, therapeutic and/or research related content.

©2024 Mayo Foundation for Medical Education and Research

Cultural Linguistic Competency (CLC) & Implicit Bias (IB)

STATE LAW:

The California legislature has passed Assembly Bill (AB) 1195, which states that as of July 1, 2006, all Category 1 CME activities that relate to patient care must include a cultural diversity/linguistics component. It has also passed AB 241, which states that as of January 1, 2022, all continuing education courses for a physician and surgeon **must** contain curriculum that includes specified instruction in the understanding of implicit bias in medical treatment.

The cultural and linguistic competency (CLC) and implicit bias (IB) definitions reiterate how patients' diverse backgrounds may impact their access to care.

EXEMPTION:

Business and Professions Code 2190.1 exempts activities which are dedicated solely to research or other issues that do not contain a direct patient care component.

The following CLC & IB components will be addressed in this presentation:

- *Will discuss the bias in artificial intelligence caused by poor sampling of ethnic minorities*
- *Implicit bias of models built upon non-representative cohorts of patients*

©2024 Mayo Foundation for Medical Education and Research | slide-3

WHO AM I

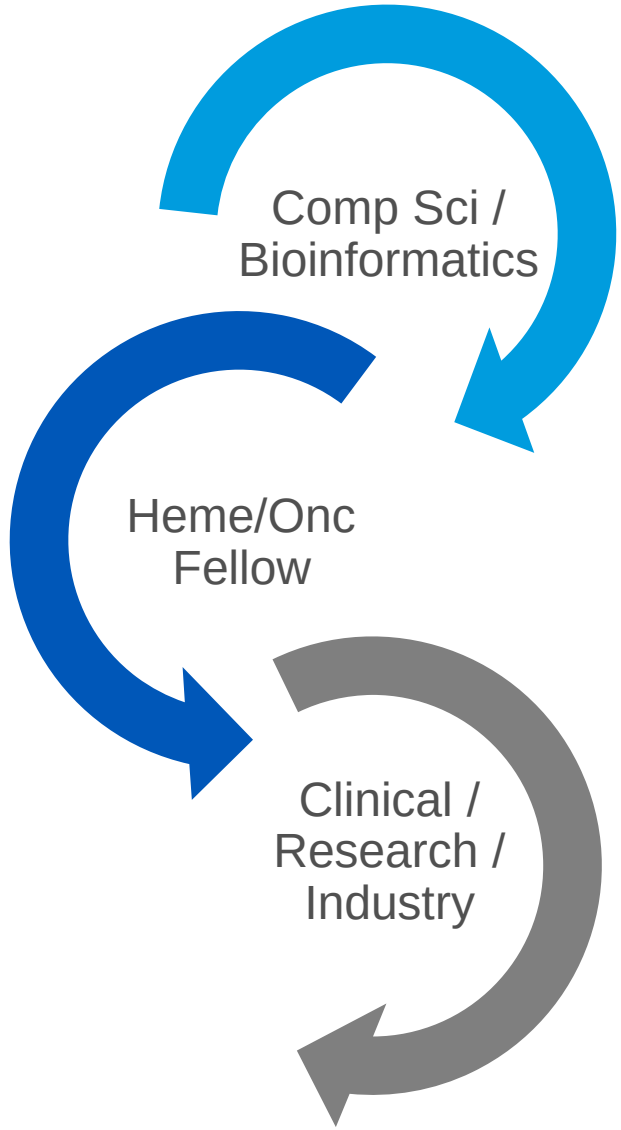
JACOB SHREVE MD

- Clinician, senior heme/onc fellow, Mayo Clinic
- Specialist in bioinformatics at a core facility
- Computational scientist in academia
- Software engineer for biotech, personal exomics
- Cofounder of a healthcare AI tech startup
- Transitioned to medicine to explore how tech can personalize care

I AM NOT:
An expert in all things artificial intelligence

I AM:
Continuously working to expand my knowledge base and skill set

I WANT:
To bring together like-minded people to benefit from our shared experience



©2024 Mayo Foundation for Medical Education and Research

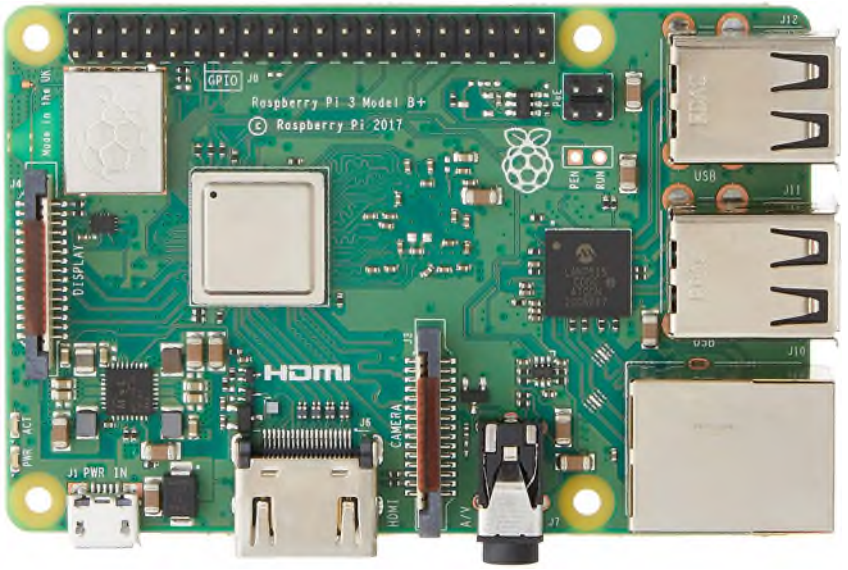
ARTIFICIAL INTELLIGENCE

AGENDA

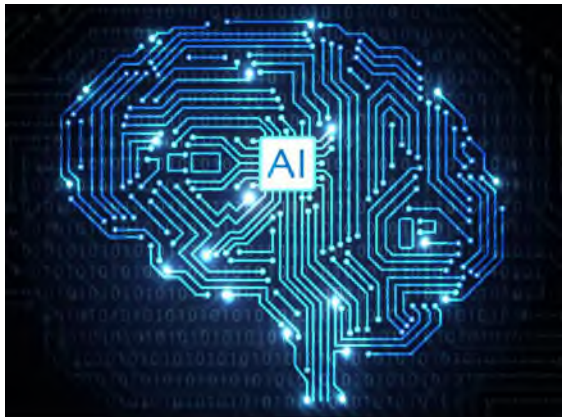
- Why use AI & how does it work
- AI in oncology examples
- Where is the AI revolution?
- How to evaluate AI research

©2024 Mayo Foundation for Medical Education and Research

ARTIFICIAL INTELLIGENCE



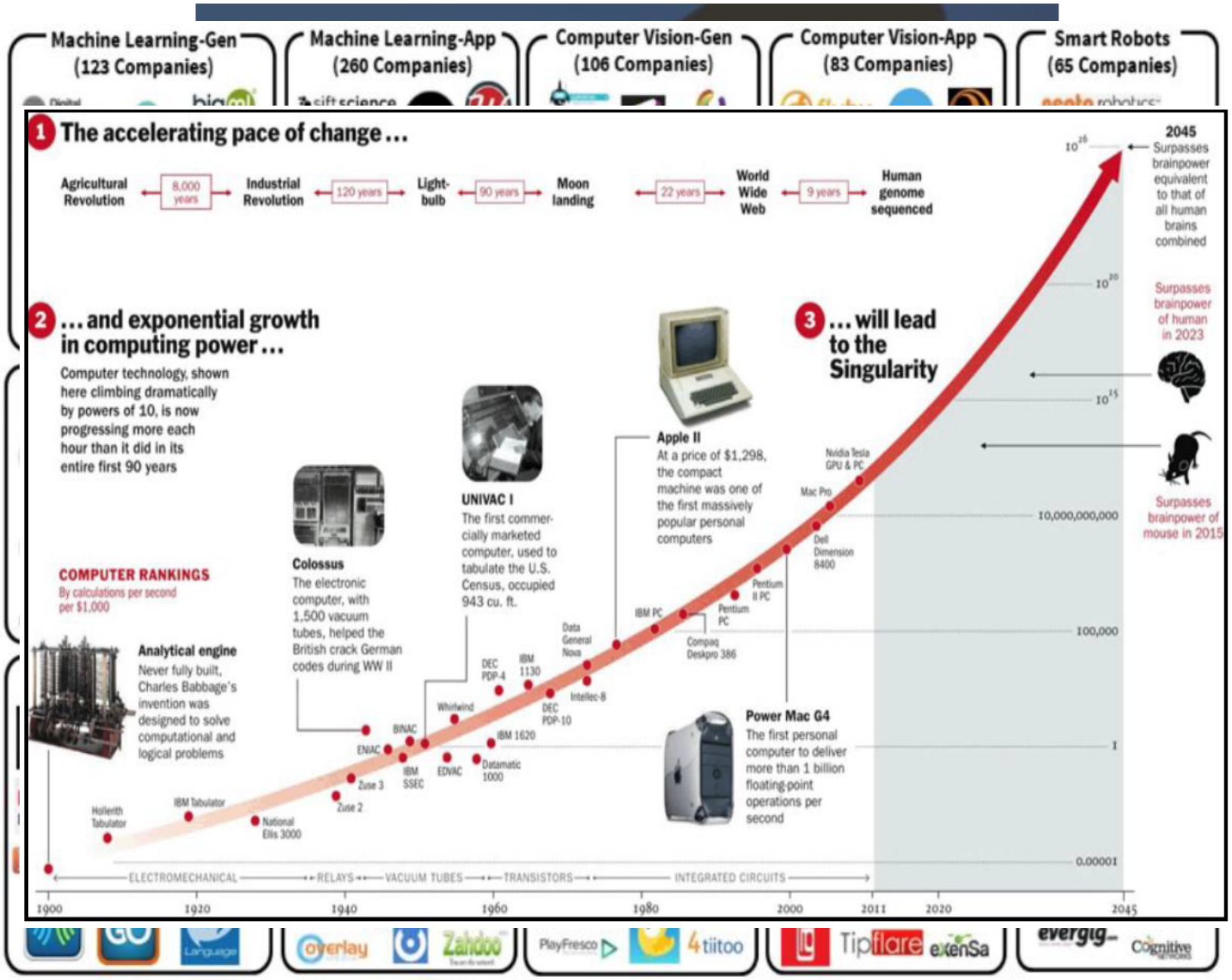
- Decades of progress, starting in 1950s
- Golden Age: currently have numerous open-source packages
- Difference from traditional statistics:
  - Goal of statistics is to assess relationships between variables and provide hypothesis testing
  - Goal of AI is to model a system and provide accurate predictions “end justifies the means”



©2024 Mayo Foundation for Medical Education and Research

WHY AI

TECH TODAY



©2024 Mayo Foundation for Medical Education and Research

AI IN HEALTHCARE

IBM Watson

- Organizational performance, interpretation

Outcomes

- Prevent complications, make better choices

Radiology

- Automated radiograph interpretation

Drug Discovery

- Molecule efficacy prediction

Pathology

- Automated microscopy interpretation

©2024 Mayo Foundation for Medical Education and Research

News & Views | Published: 07 January 2019

Cardiovascular diseases

Artificial intelligence for the electrocardiogram

Ana Minchol

Nature Medi

6659 Acces

Deep-lear
capable of
aid health

Artificial in
ability to o
innovation
impact. In
healthcare
systematic
asymptom
performance computing and cloud computing have become available through hardware
improvements, particularly in graphics processing units.
independent set of 52,870 patients, the network model yi

019

dicine: the convergence of
ntelligence

e Clinician learns optimal
sepsis in intensive care

i, Anthony C. Gordon & A. Aldo Faisal

is article

| Metrics

worldwide and the main cause of mortality in
tegy remains uncertain. In particular, evidence
inistration of intravenous fluids and vasopressors
a proportion of patients[1,4,5,6]. To tackle this
oblem, we developed a reinforcement learning agent, the
cian, which extracted implicit knowledge from an amount of
patient data that exceeds by many-fold the life-time experience of human clinicians and

Perspective | Published: 07 January 2019

The practical implementation of artificial intelligence technologies in medicine

Jianxing He, Sally L. Baxter, Jie Xu, Jiming Xu, Xingtao Zhou & Kang Zhang

Nature Medicine **25**, 30–36 (2019) | Cite this article

28k Accesses | **287** Citations | **57** Altmetric | Metrics

Abstract

The development of artificial intelligence (AI)-based technologies in medicine is advancing rapidly, but real-world clinical implementation has not yet become a reality. Here we review some of the key practical issues surrounding the implementation of AI into existing clinical workflows, including data sharing and privacy, transparency of algorithms, data standardization, and interoperability across multiple platforms, and concern for patient safety. We summarize the current regulatory environment in the United States and highlight comparisons with other regions in the world, notably Europe and China.

NEJM:
Artificial Intelligence in Medicine

“After hearing for several decades that computers will soon be able to assist with difficult diagnoses, the practicing physician may well wonder why the revolution has not occurred…”

1987

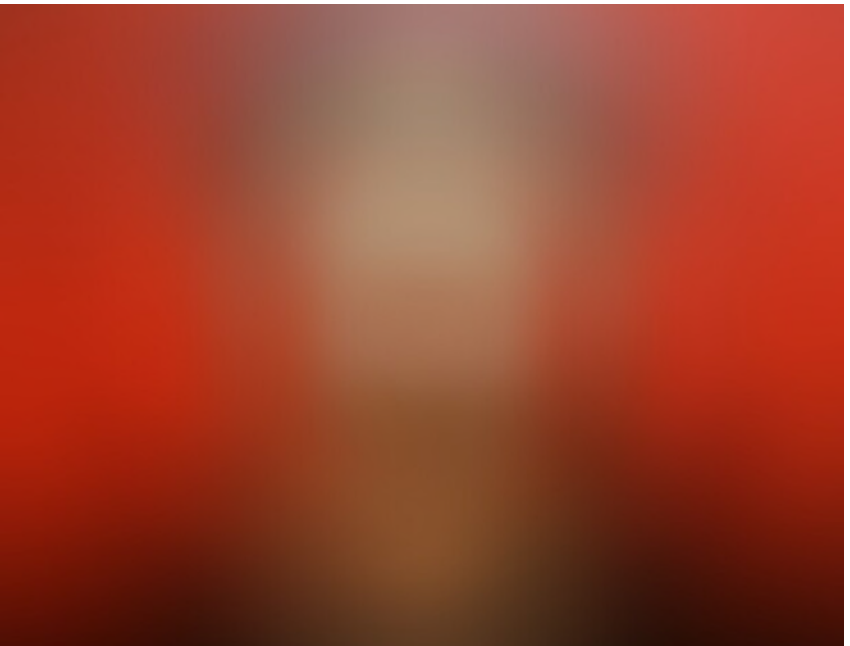
©2024 Mayo Foundation for Medical Education and Research

RATIONAL

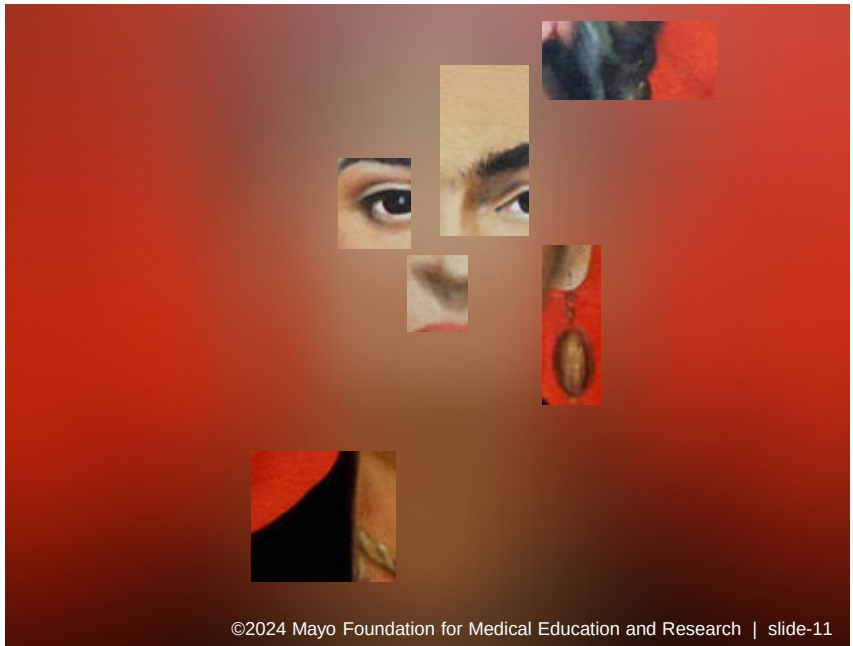
TRADITIONAL MEDICINE

Targeted hypothesis testing, pathophysiology driven, incremental large advancements in the field

Biological System



Our Understanding



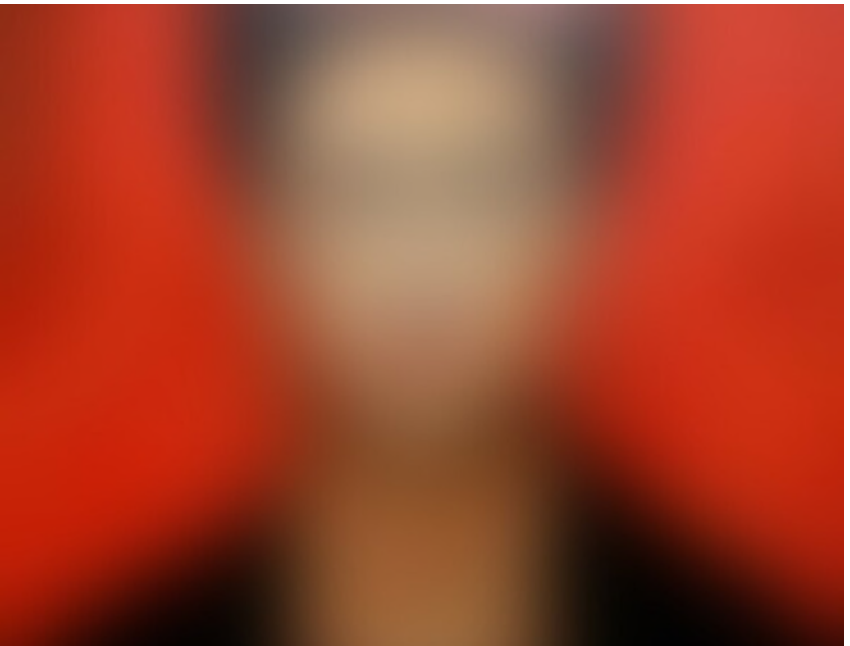
©2024 Mayo Foundation for Medical Education and Research

RATIONAL

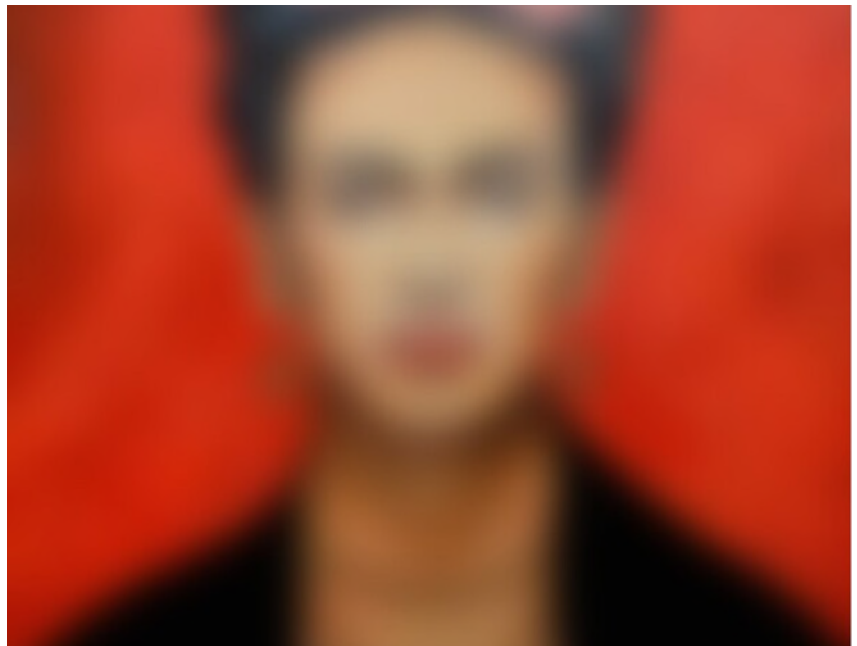
AI / ML APPROACH

Data type integration, larger sample sizes, more sophisticated modeling

EMR



EMR + radiomics



EMR, radiomics,
Proteomics, digital pathology,
Wearable devices, etc

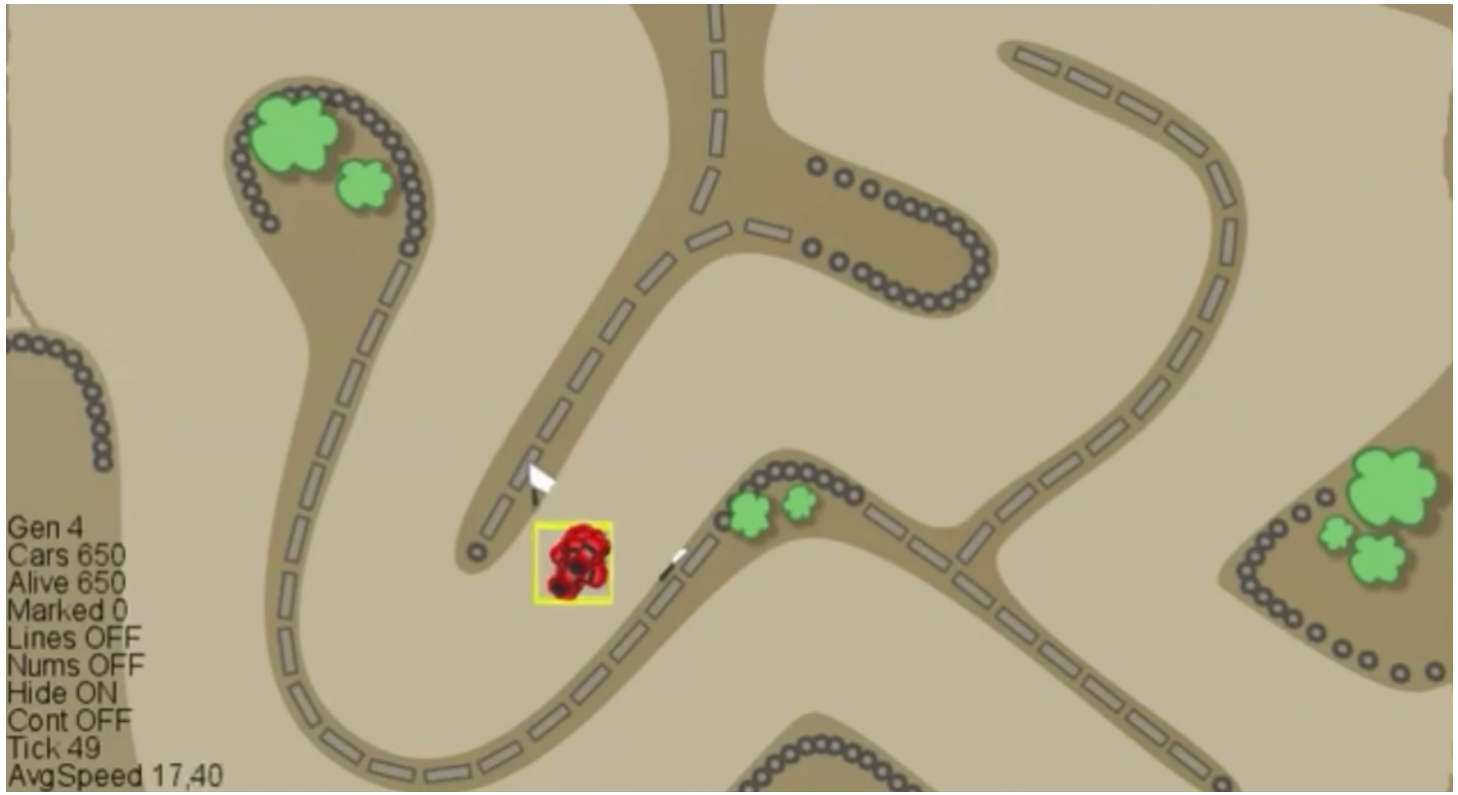
©2024 Mayo Foundation for Medical Education and Research

Demonstration

- The algorithm starts with hundreds / thousands of attempts
- Each attempt has random variables within specific boundaries

©2024 Mayo Foundation for Medical Education and Research

Demonstration



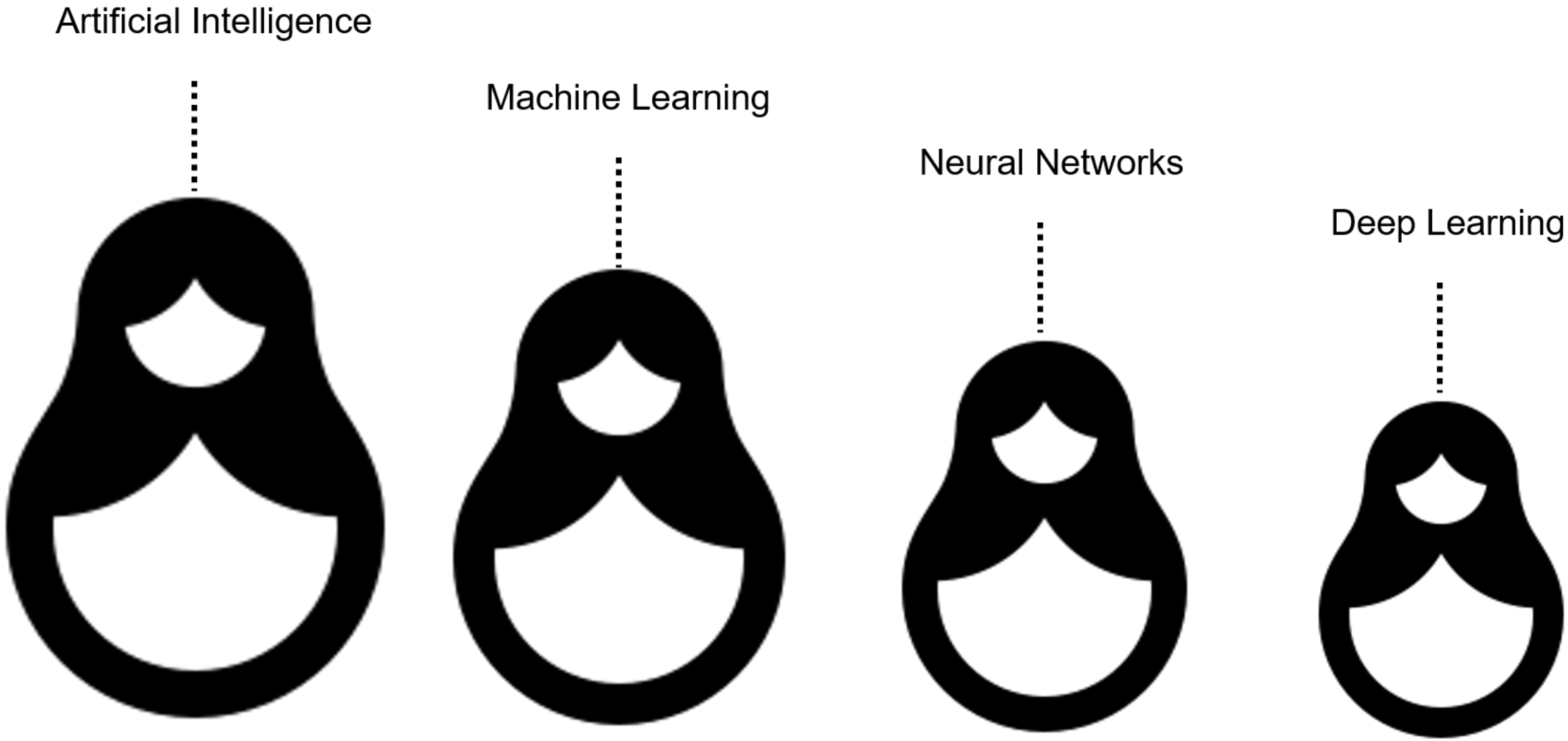
- Another round begins with new variable boundaries

©2024 Mayo Foundation for Medical Education and Research

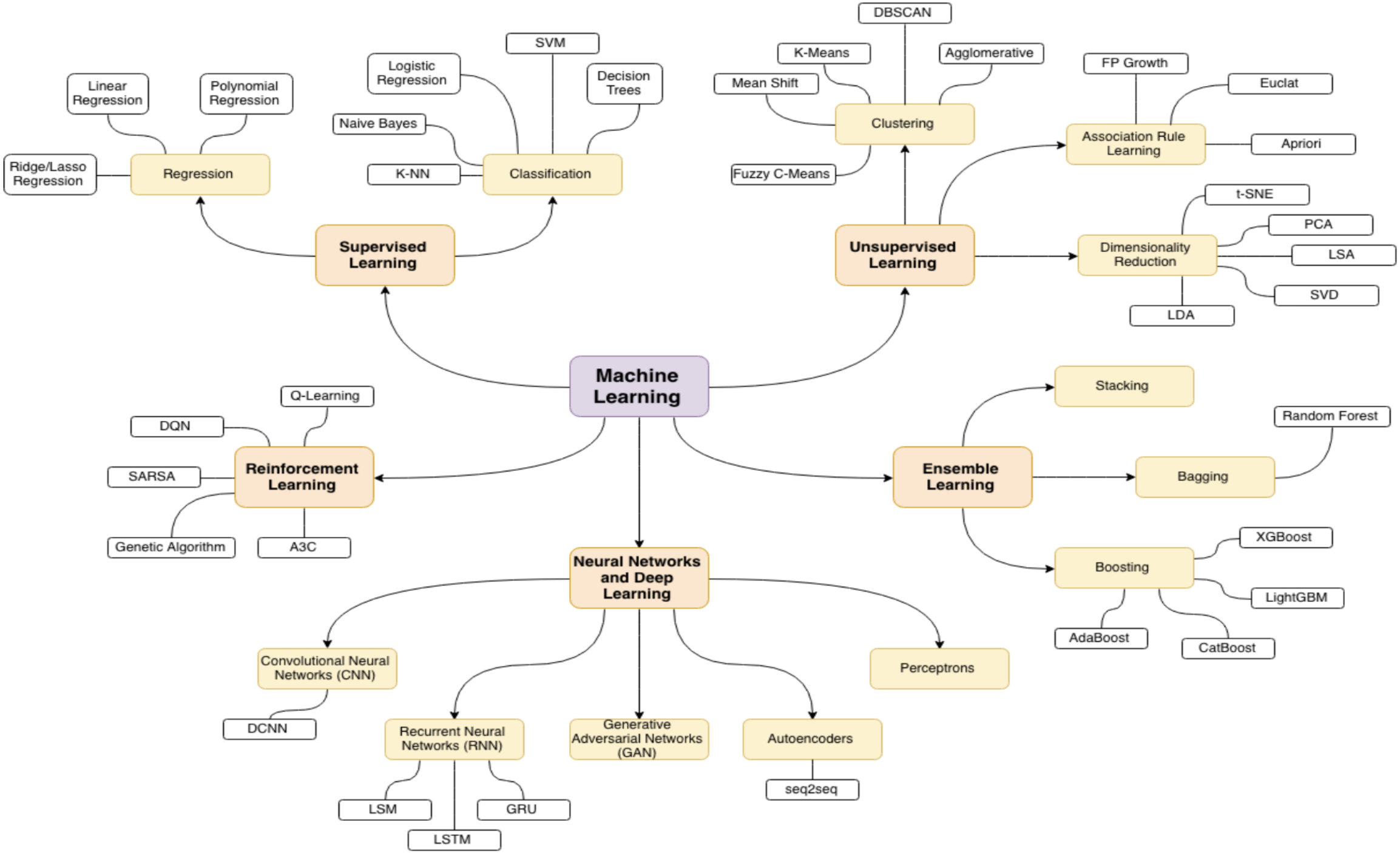
Demonstration

- After many rounds, the optimal variable settings have been determined

©2024 Mayo Foundation for Medical Education and Research



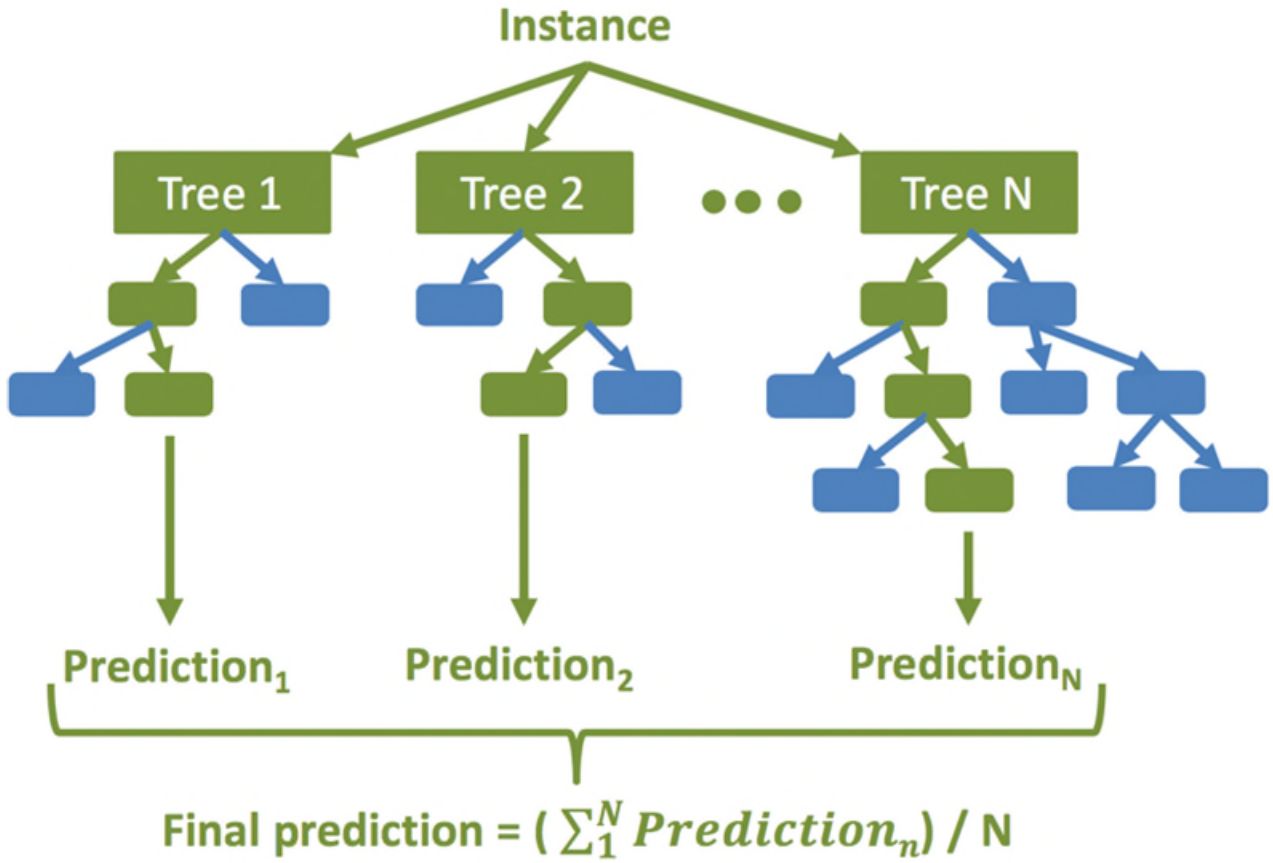
©2024 Mayo Foundation for Medical Education and Research


Machine Learning
Supervised Learning
Regression
Linear Regression
Polynomial Regression
Ridge/Lasso Regression
Classification
Logistic Regression
SVM
Decision Trees
Naive Bayes
K-NN
Unsupervised Learning
Clustering
K-Means
Mean Shift
DBSCAN
Agglomerative
Fuzzy C-Means
Association Rule Learning
FP Growth
Euclat
Apriori
Dimensionality Reduction
t-SNE
PCA
LSA
SVD
LDA
Reinforcement Learning
Q-Learning
DQN
SARSA
Genetic Algorithm
A3C
Ensemble Learning
Stacking
Bagging
Random Forest
Boosting
XGBoost
LightGBM
AdaBoost
CatBoost
Neural Networks and Deep Learning
Convolutional Neural Networks (CNN)
DCNN
Recurrent Neural Networks (RNN)
LSM
GRU
LSTM
Generative Adversarial Networks (GAN)
Autoencoders
seq2seq
Perceptrons

COMMON ALGORITHM GOALS

RANDOM FOREST
Combine weighted decision trees to make prediction

- Generates many randomly created decision trees, assigns weights, sums a prediction path

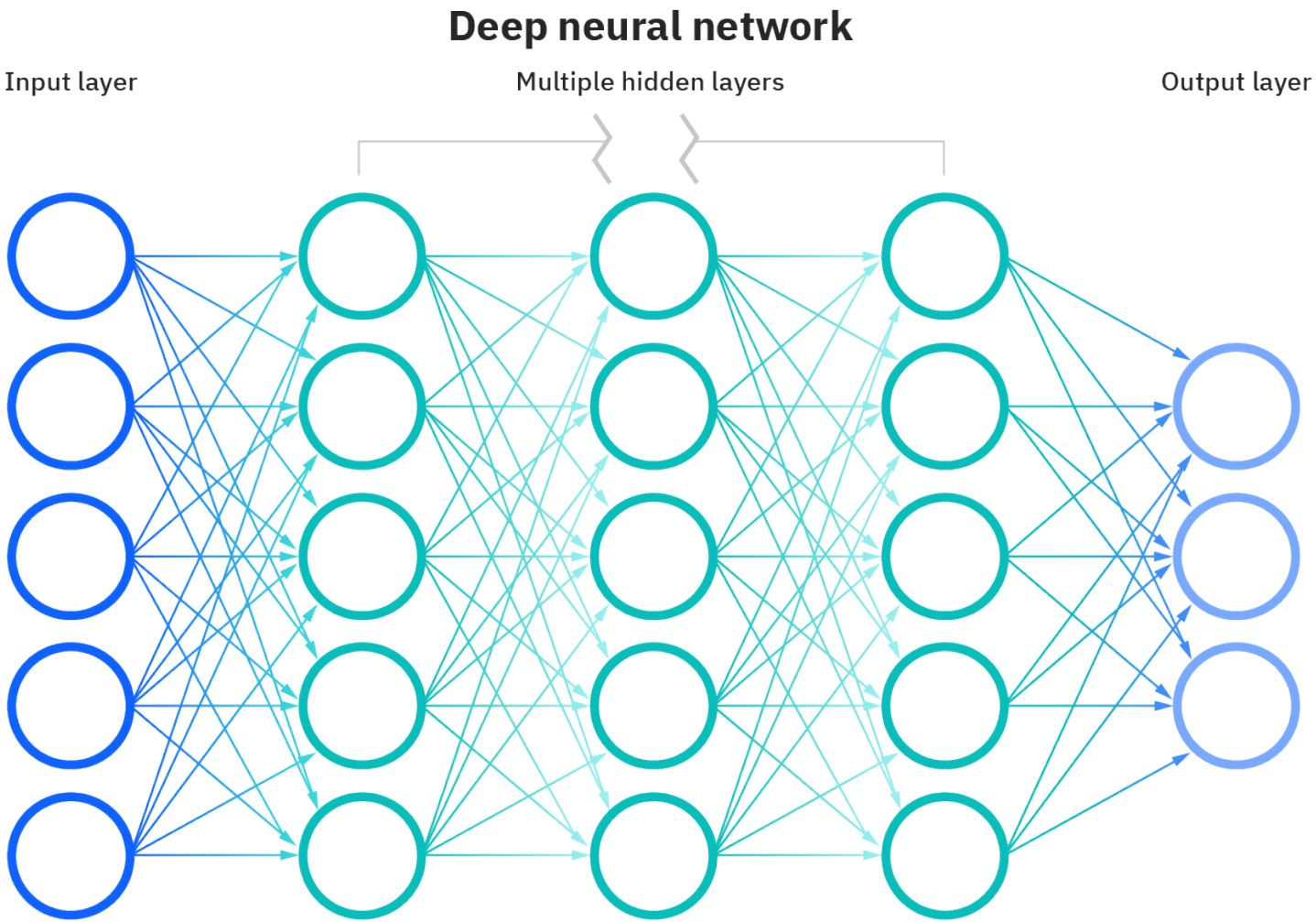


©2024 Mayo Foundation for Medical Education and Research

HOW DOES IT WORK

DEEP LEARNING

- It is a type of machine learning that uses neural networks with 3 or more (many more) hidden layers

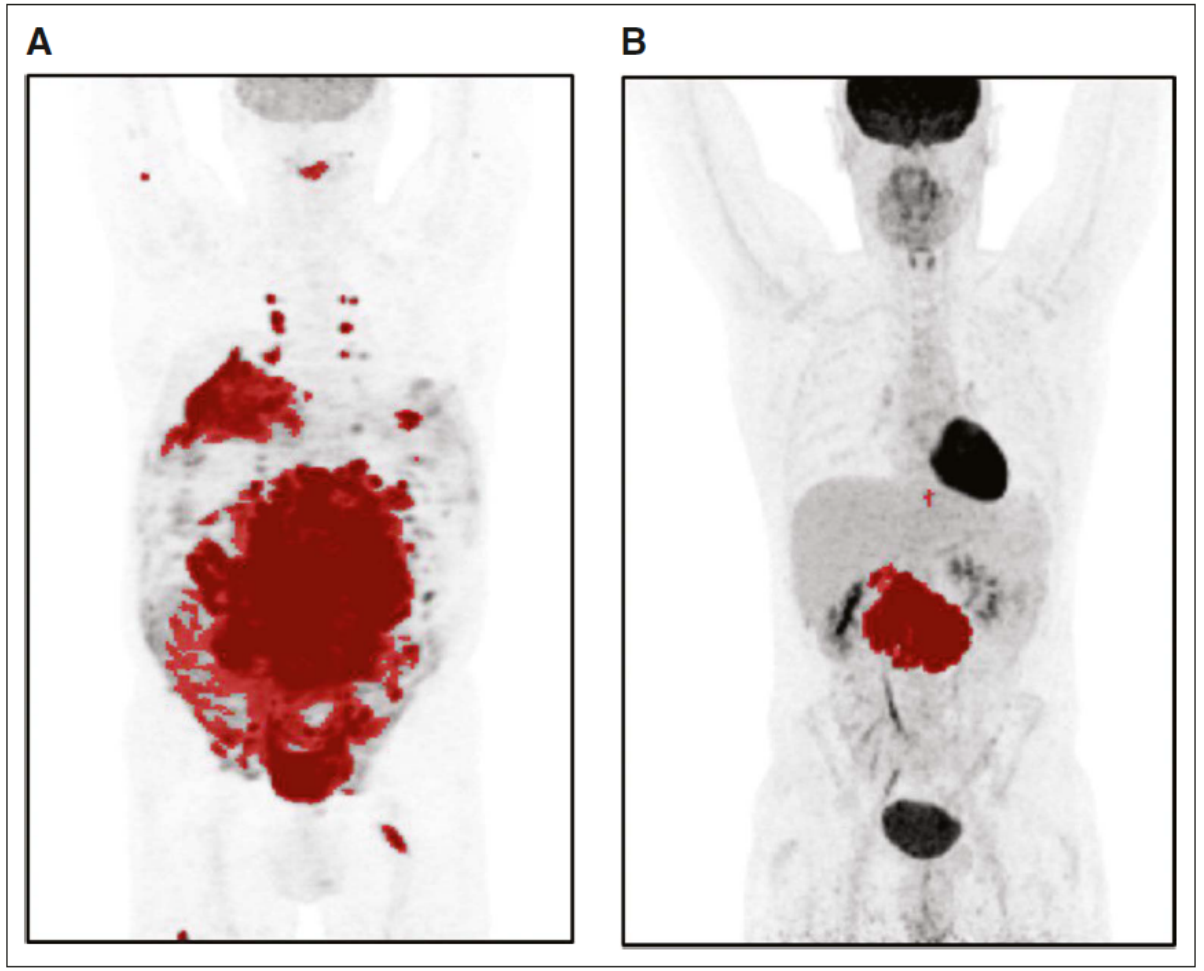


©2024 Mayo Foundation for Medical Education and Research

HOW DOES IT WORK

COMPUTER VISION

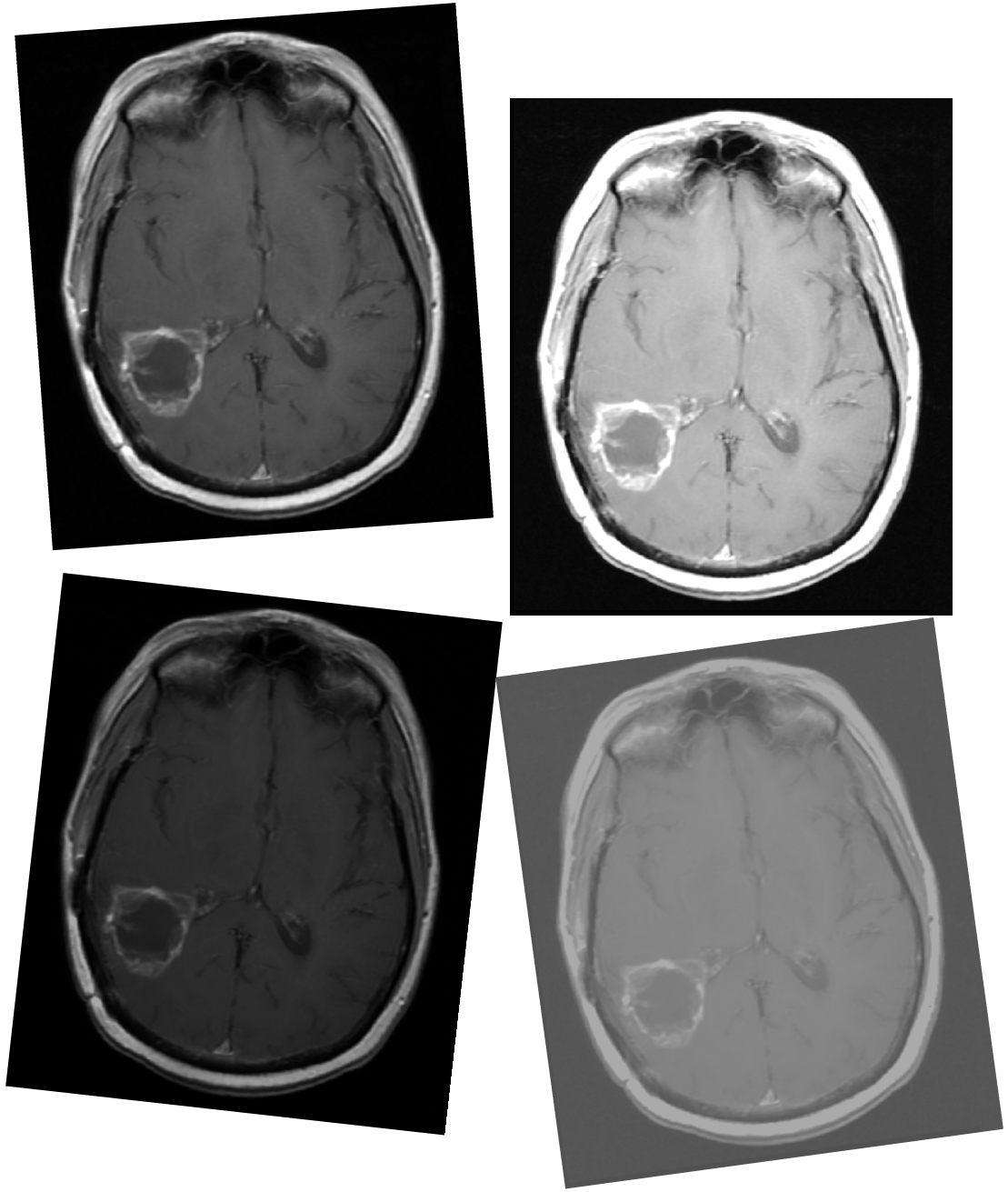
- Show medical imaging to an AI algorithm and have it understand where tumor is located



©2024 Mayo Foundation for Medical Education and Research

Computer vision: neuroimaging feature extraction using deep learning

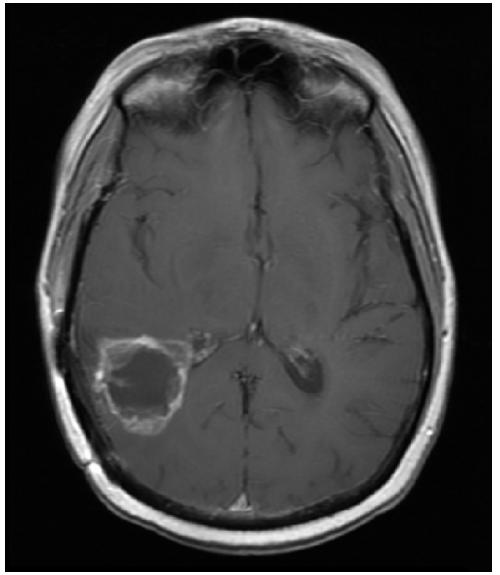
- Image acquisition (MR brain)
- Image registration (spatial alignment)
- Preprocessing, including skull stripping (FSL brain extraction tool / Robust brain extraction)
- Intensity normalization
- Noise reduction (using Gaussian convolution)
- Bias field correction (N4 bias field correction)
- Segmentation
- Feature extraction (1st order, 2nd order, high-order features)

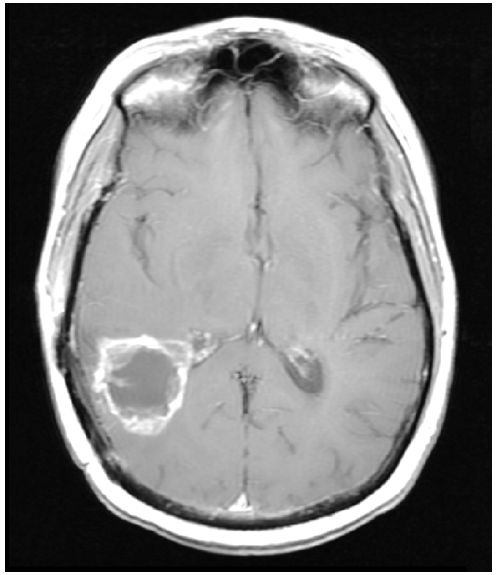


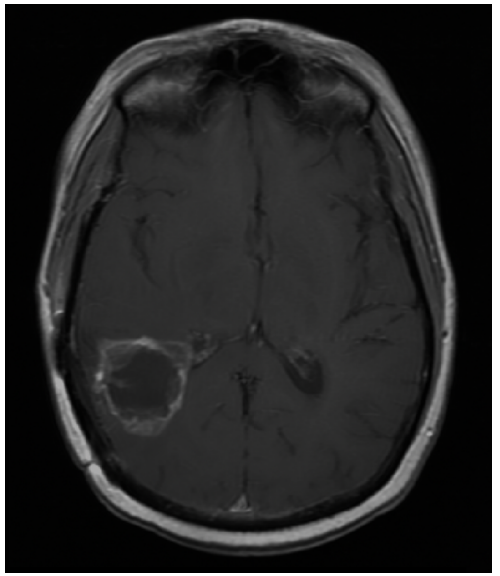
©2024 Mayo Foundation for Medical Education and Research

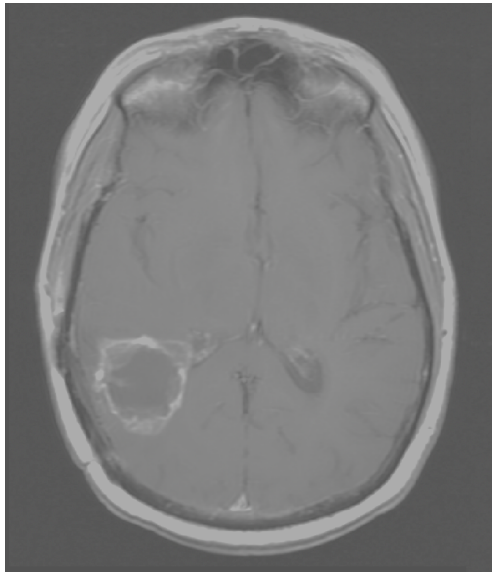
First model: neuroimaging feature extraction using deep learning

- Image acquisition (MR brain)
- Image registration (spatial alignment)
- Preprocessing, including skull stripping (FSL brain extraction tool / Robust brain extraction)
- Intensity normalization
- Noise reduction (using Gaussian convolution)
- Bias field correction (N4 bias field correction)
- Segmentation
- Feature extraction (1st order, 2nd order, high-order features)





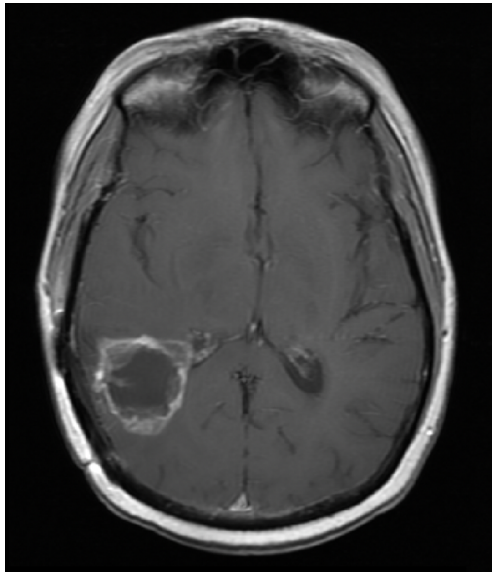


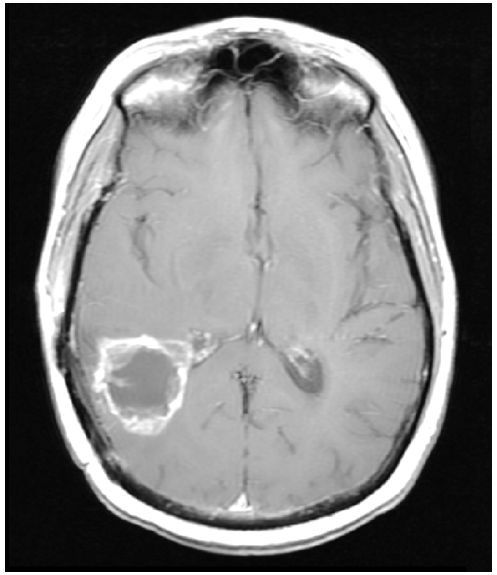


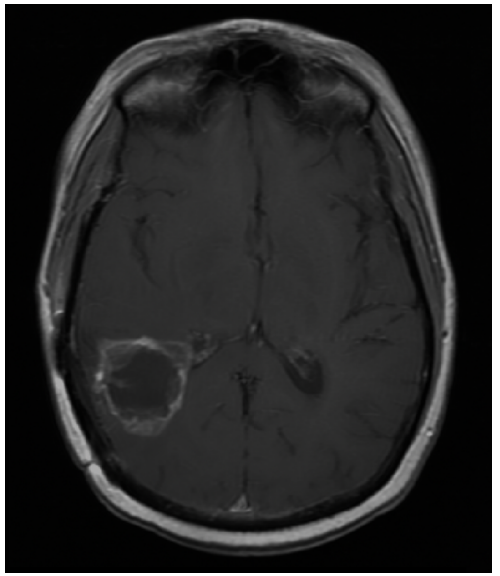
©2024 Mayo Foundation for Medical Education and Research

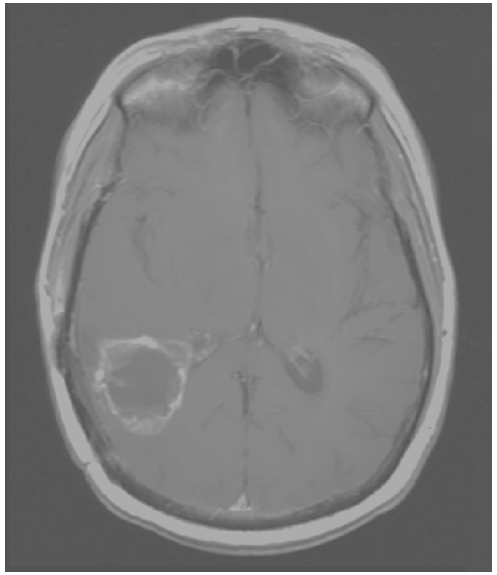
First model: neuroimaging feature extraction using deep learning

- Image acquisition (MR brain)
- Image registration (spatial alignment)
- Preprocessing, including skull stripping (FSL brain extraction tool / Robust brain extraction)
- Intensity normalization
- Noise reduction (using Gaussian convolution)
- Bias field correction (N4 bias field correction)
- Segmentation
- Feature extraction (1st order, 2nd order, high-order features)





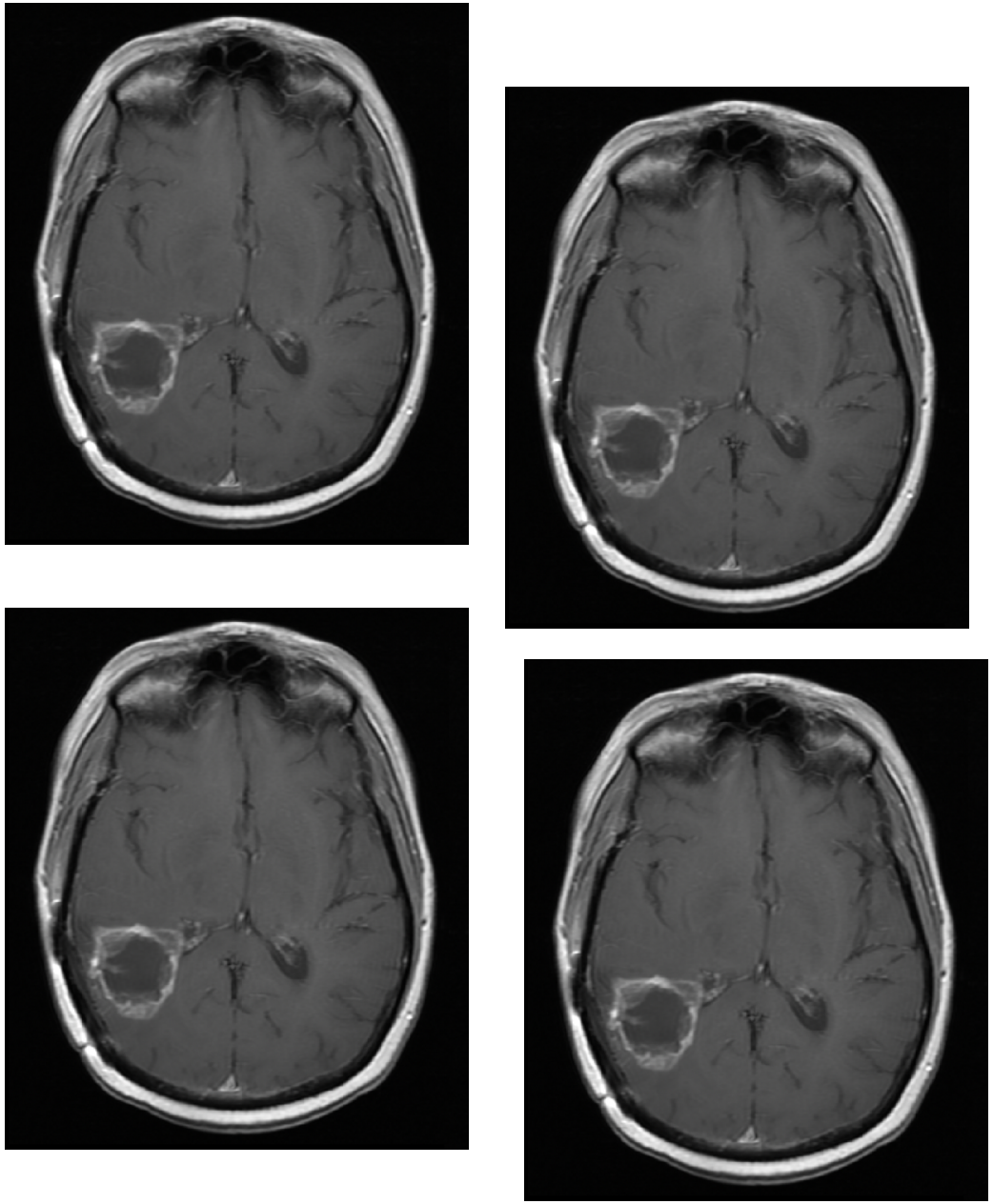




©2024 Mayo Foundation for Medical Education and Research

First model: neuroimaging feature extraction using deep learning

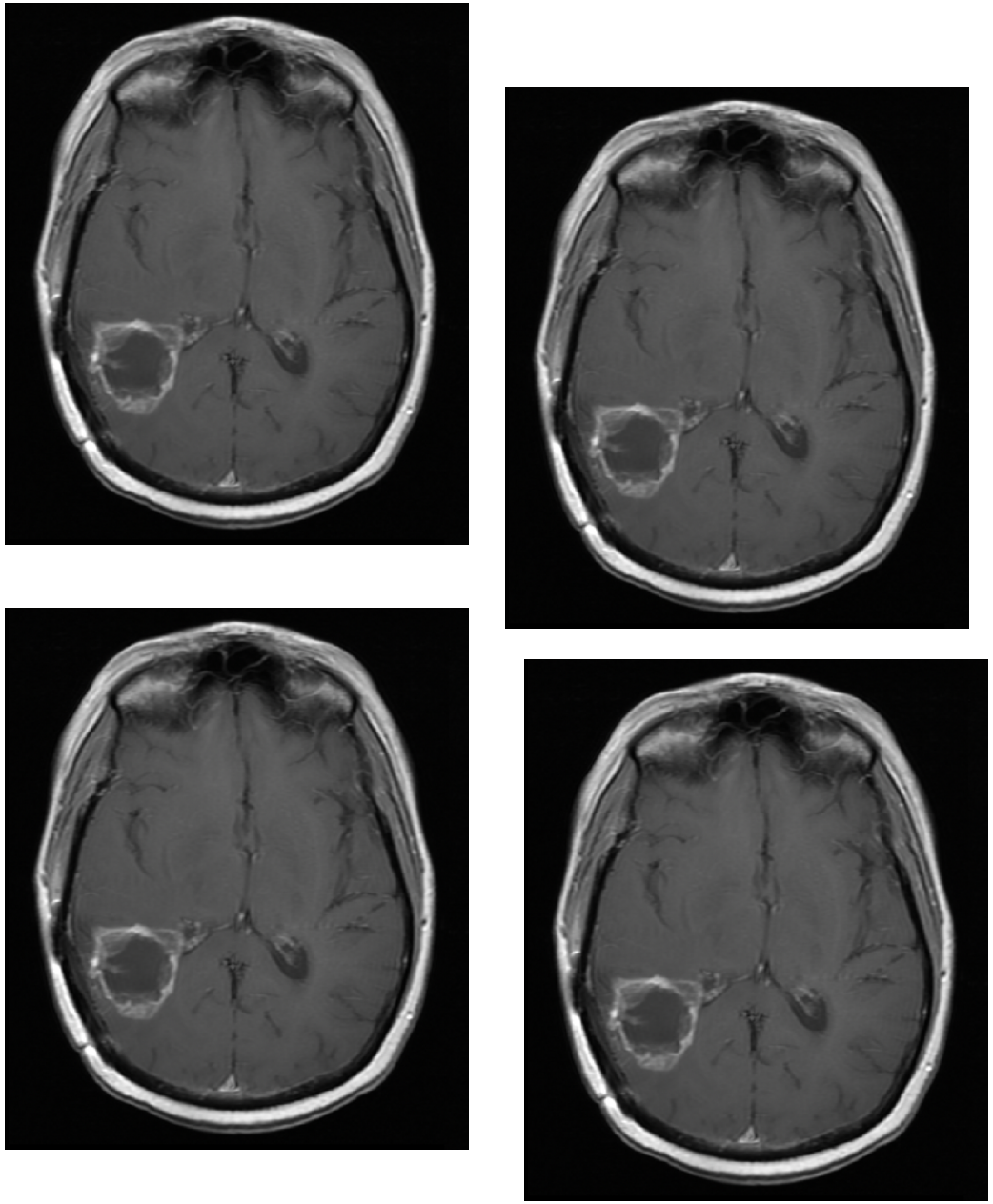
- Image acquisition (MR brain)
- Image registration (spatial alignment)
- Preprocessing, including skull stripping (FSL brain extraction tool / Robust brain extraction)
- Intensity normalization
- Noise reduction (using Gaussian convolution)
- Bias field correction (N4 bias field correction)
- Segmentation
- Feature extraction (1st order, 2nd order, high-order features)



©2024 Mayo Foundation for Medical Education and Research

First model: neuroimaging feature extraction using deep learning

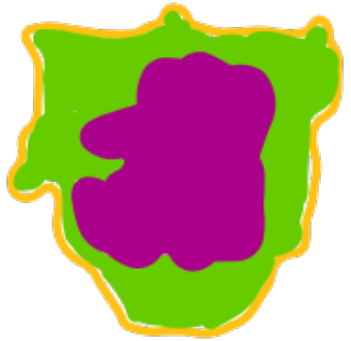
- Image acquisition (MR brain)
- Image registration (spatial alignment)
- Preprocessing, including skull stripping (FSL brain extraction tool / Robust brain extraction)
- Intensity normalization
- Noise reduction (using Gaussian convolution)
- Bias field correction (N4 bias field correction)
- Segmentation
- Feature extraction (1st order, 2nd order, high-order features)



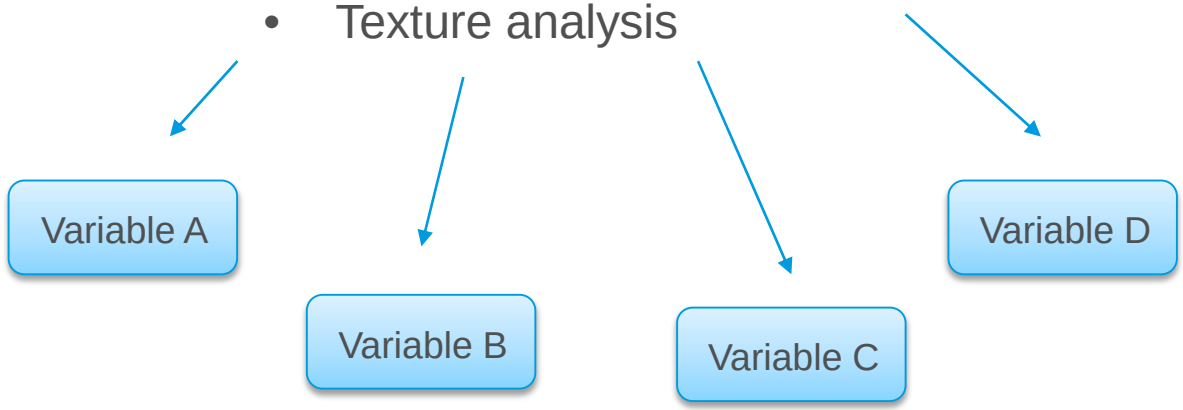
©2024 Mayo Foundation for Medical Education and Research

First model: neuroimaging feature extraction using deep learning

- Image acquisition (MR brain)
- Image registration (spatial alignment)
- Preprocessing, including skull stripping (FSL brain extraction tool / Robust brain extraction)
- Intensity normalization
- Noise reduction (using Gaussian convolution)
- Bias field correction (N4 bias field correction)
- Segmentation
- Feature extraction (1st order, 2nd order, high-order features)



- Tumor volume
- Max tumor signal intensity
- Average tumor signal intensity
- Variance of tumor signal intensity
- Distribution of voxel intensities
- 2D shape characteristics
- 3D shape characteristics
- Texture analysis

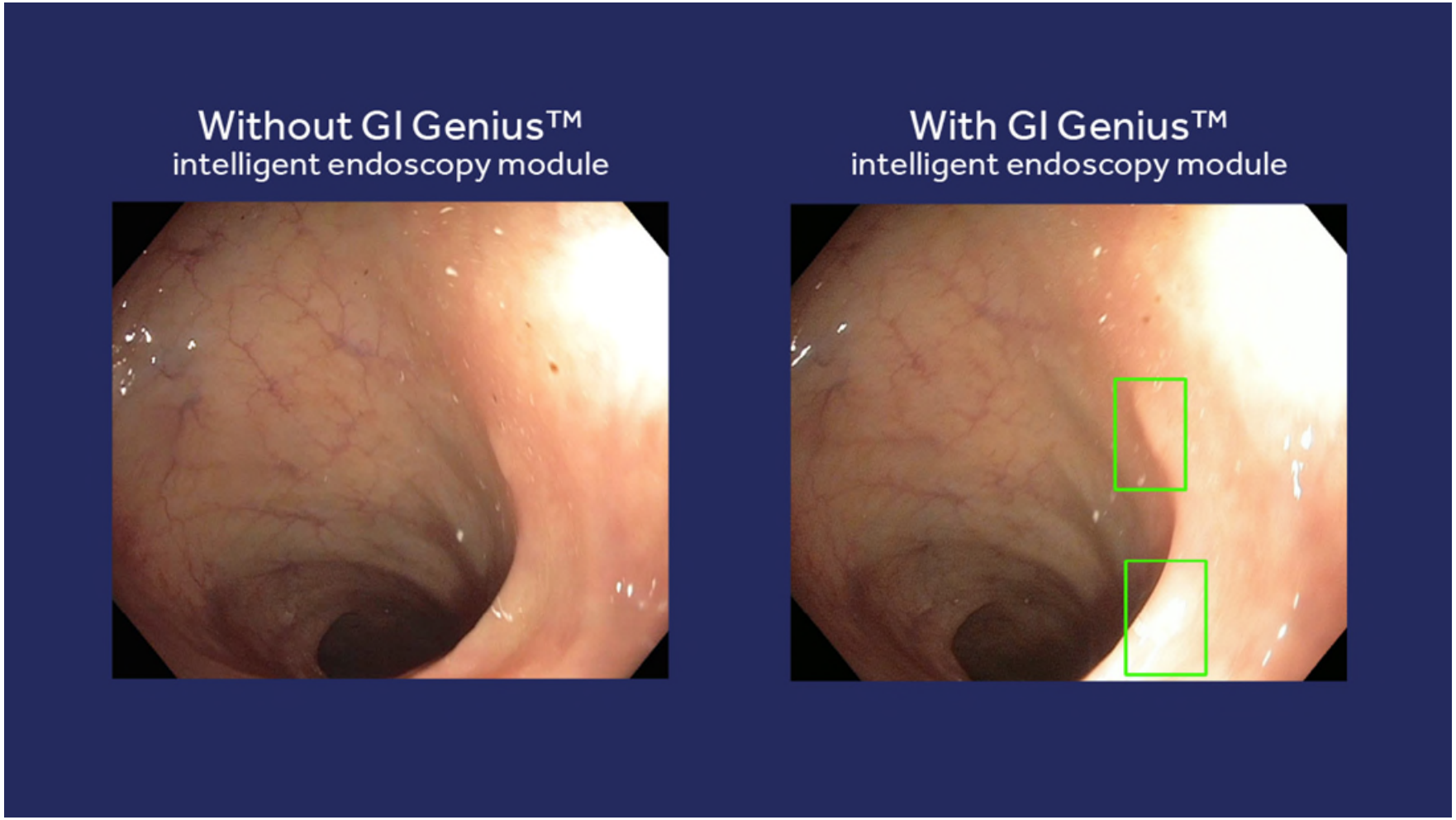


©2024 Mayo Foundation for Medical Education and Research

FDA-approved applications of deep learning: GI Genius™, by Medtronic

FDA SaMD

- Software as a Medical Device
  - 510(k) clearance
  - De Novo request
  - Premarket approval
- 692 devices approved as of late 2023
- 79% computer vision
- 30% increase in approvals over the previous year



©2024 Mayo Foundation for Medical Education and Research

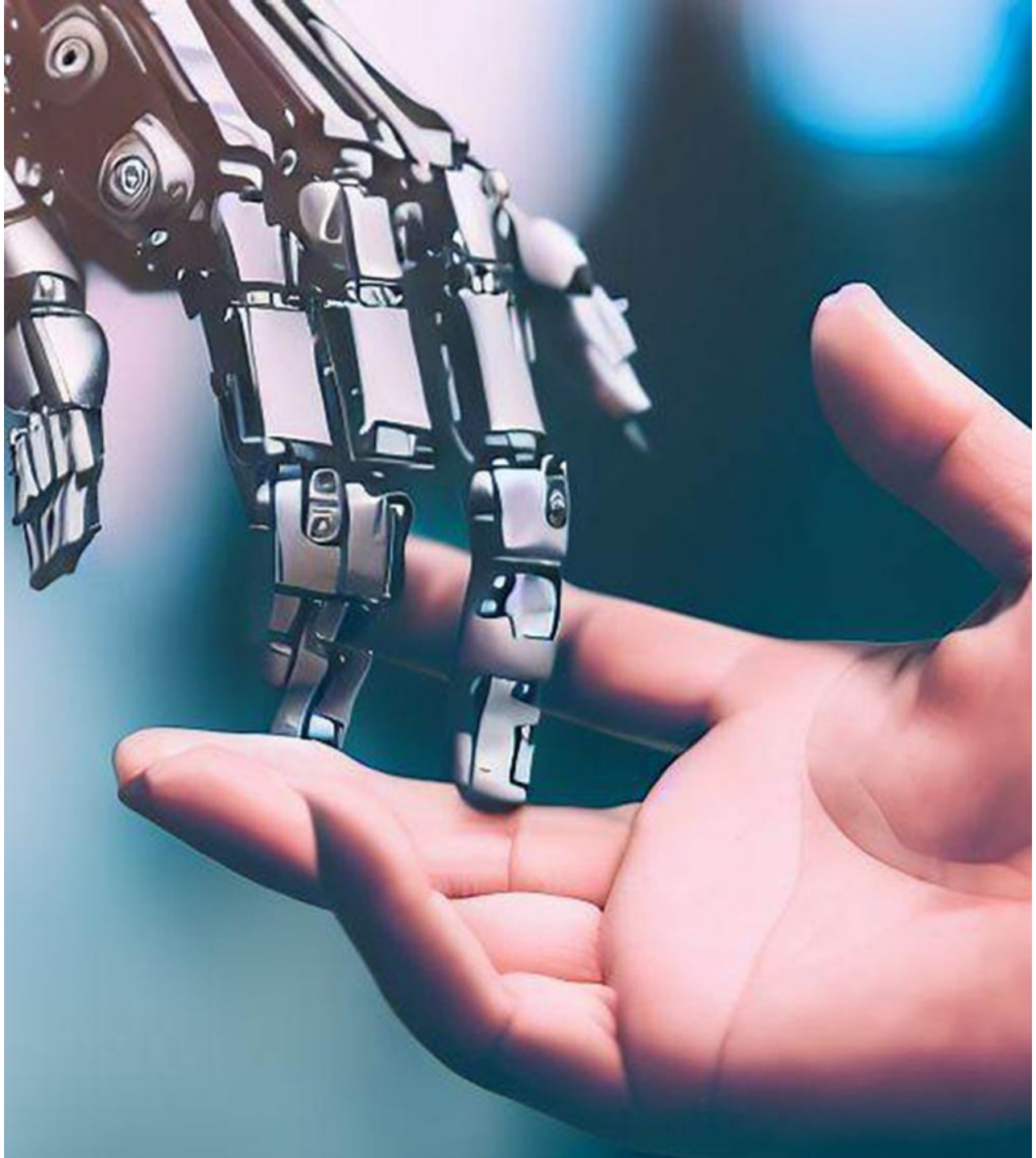
TYPES OF AI

GENERATIVE AI & LLMS

- ChatGPT
- Sentiment analysis
- Paraphrasing / semantic equivalence
- Text summarization
- Information extraction

- Scheduling / in-basket relief
- EMR mining
- Medical education
- Patient summaries / note stems

- No FDA approvals out of 692 devices!



ARTIFICIAL INTELLIGENCE

AGENDA

- Why use AI & how does it work
- AI in oncology examples
- Where is the AI revolution?
- How to evaluate AI research

©2024 Mayo Foundation for Medical Education and Research

Acute Myeloid Leukemia

- Outcomes associated with complex interaction of clinical features (1/3) and cytogenetic/molecular features (2/3)
- Prognostication determines choice of consolidative therapy (hematopoietic cell transplantation vs chemotherapy)
- The European LeukemiaNet (ELN) risk stratifies patients:

Favorable

Intermediate

Adverse

Shreve, J., Meggendorfer, M., Awada, H., Mukherjee, S., Walter, W., Hutter, S., Makhoul, A., Hilton, C.B., Radakovich, N., Nagata, Y. and Rouphail, Y., 2019. A personalized prediction model to risk stratify patients with acute myeloid leukemia (AML) using artificial intelligence. *Blood*, *134*, p.2091.

©2024 Mayo Foundation for Medical Education and Research

Acute Myeloid Leukemia

- Outcomes associated with complex interaction of clinical features (1/3) and cytogenetic/molecular features (2/3)
- Prognostication determines choice of consolidative therapy (hematopoietic cell transplantation vs chemotherapy)
- The European LeukemiaNet (ELN) risk stratifies patients:

Favorable	t(8;21)(q22;q22.1); *RUNX1-RUNX1T1*
	inv(16)(p13.1;q22) or t(16;16)(p13.1;q22); *CBFB-MYH11*
	Mutated *NPM1* without *FLT3-ITD* or with *FLT3-ITD*^{low}
Intermediate	
Adverse	Biallelic mutated *CEBPA*

Shreve, J., Meggendorfer, M., Awada, H., Mukherjee, S., Walter, W., Hutter, S., Makhoul, A., Hilton, C.B., Radakovich, N., Nagata, Y. and Rouphail, Y., 2019. A personalized prediction model to risk stratify patients with acute myeloid leukemia (AML) using artificial intelligence. *Blood*, *134*, p.2091.

©2024 Mayo Foundation for Medical Education and Research

Acute Myeloid Leukemia

- Outcomes associated with complex interaction of clinical features (1/3) and cytogenetic/molecular features (2/3)
- Prognostication determines choice of consolidative therapy (hematopoietic cell transplantation vs chemotherapy)
- The European LeukemiaNet (ELN) risk stratifies patients:

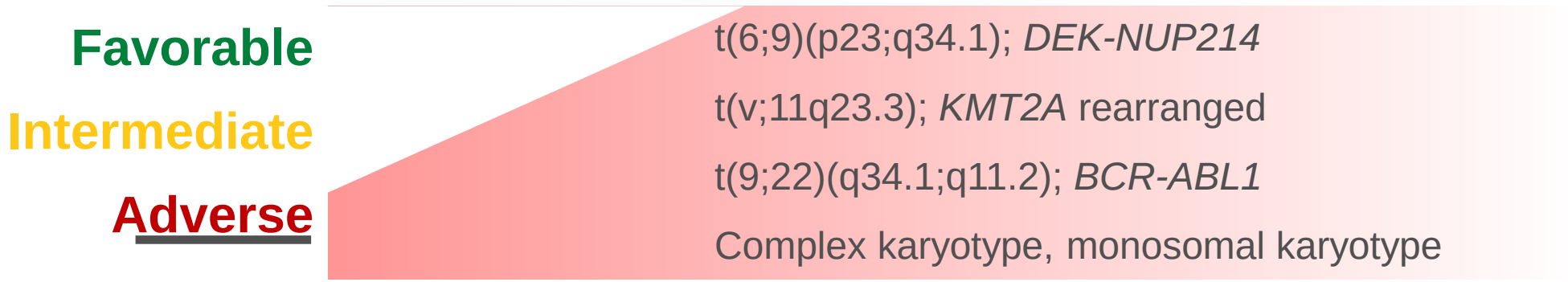
Favorable	Mutated *NPM1* and *FLT3-ITD*^{high}
Intermediate	Wild type *NPM1* without *FLT3-ITD* or with *FLT3-ITD*^{low}
Adverse	t(9;11)(p21.3;q23.3); *MLLT3-KMT2A* Cytogenetic abnormalities not classified

Shreve, J., Meggendorfer, M., Awada, H., Mukherjee, S., Walter, W., Hutter, S., Makhoul, A., Hilton, C.B., Radakovich, N., Nagata, Y. and Rouphail, Y., 2019. A personalized prediction model to risk stratify patients with acute myeloid leukemia (AML) using artificial intelligence. *Blood*, *134*, p.2091.

©2024 Mayo Foundation for Medical Education and Research

Acute Myeloid Leukemia

- Outcomes associated with complex interaction of clinical features (1/3) and cytogenetic/molecular features (2/3)
- Prognostication determines choice of consolidative therapy (hematopoietic cell transplantation vs chemotherapy)
- The European LeukemiaNet (ELN) risk stratifies patients:



Shreve, J., Meggendorfer, M., Awada, H., Mukherjee, S., Walter, W., Hutter, S., Makhoul, A., Hilton, C.B., Radakovich, N., Nagata, Y. and Rouphail, Y., 2019. A personalized prediction model to risk stratify patients with acute myeloid leukemia (AML) using artificial intelligence. *Blood*, *134*, p.2091.

©2024 Mayo Foundation for Medical Education and Research

Acute Myeloid Leukemia

- Current best-practice ELN classification only has 75% accuracy
- How can the numerous prognostic features in AML be better characterized?

Translocations

Chimeric oncogenes

Bone marrow blasts %

WBCs

Rearrangements

Age

Epigenetic regulators

Regulators of apoptosis

Deletions

Tumor suppressors

Duplications

Sex

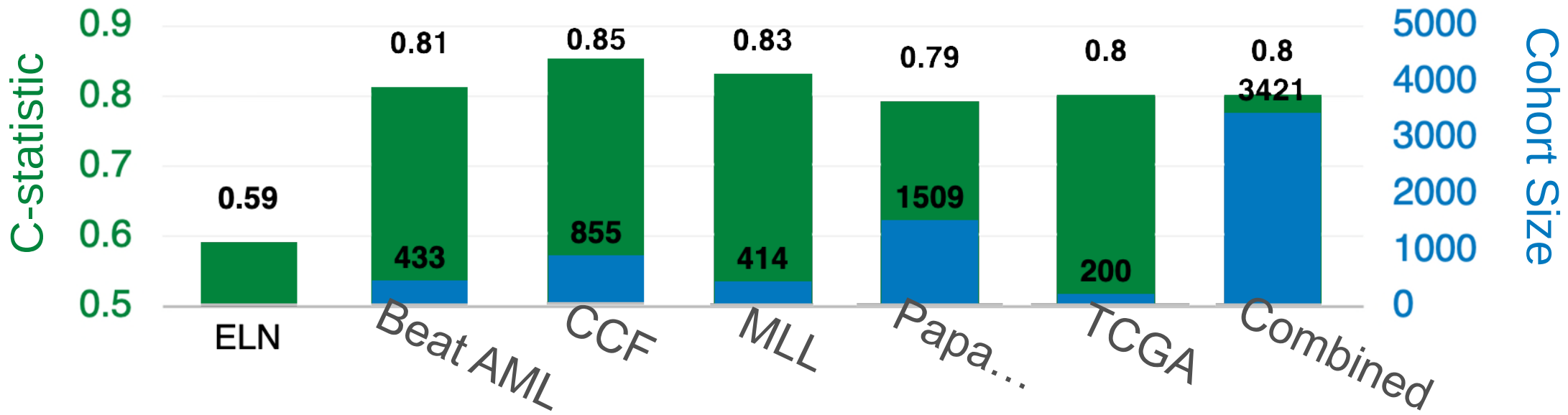
Transplant status

DNA repair

©2024 Mayo Foundation for Medical Education and Research

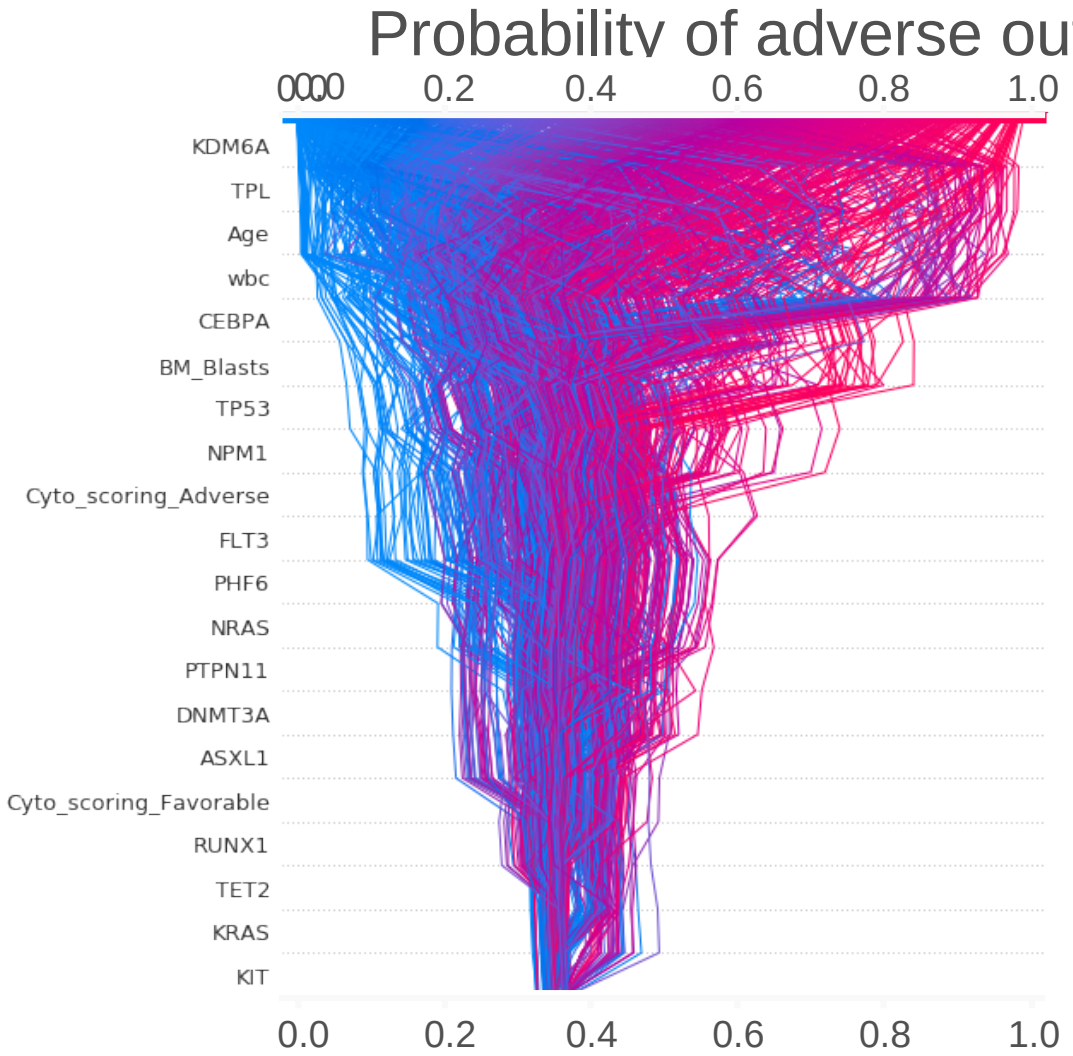
Results

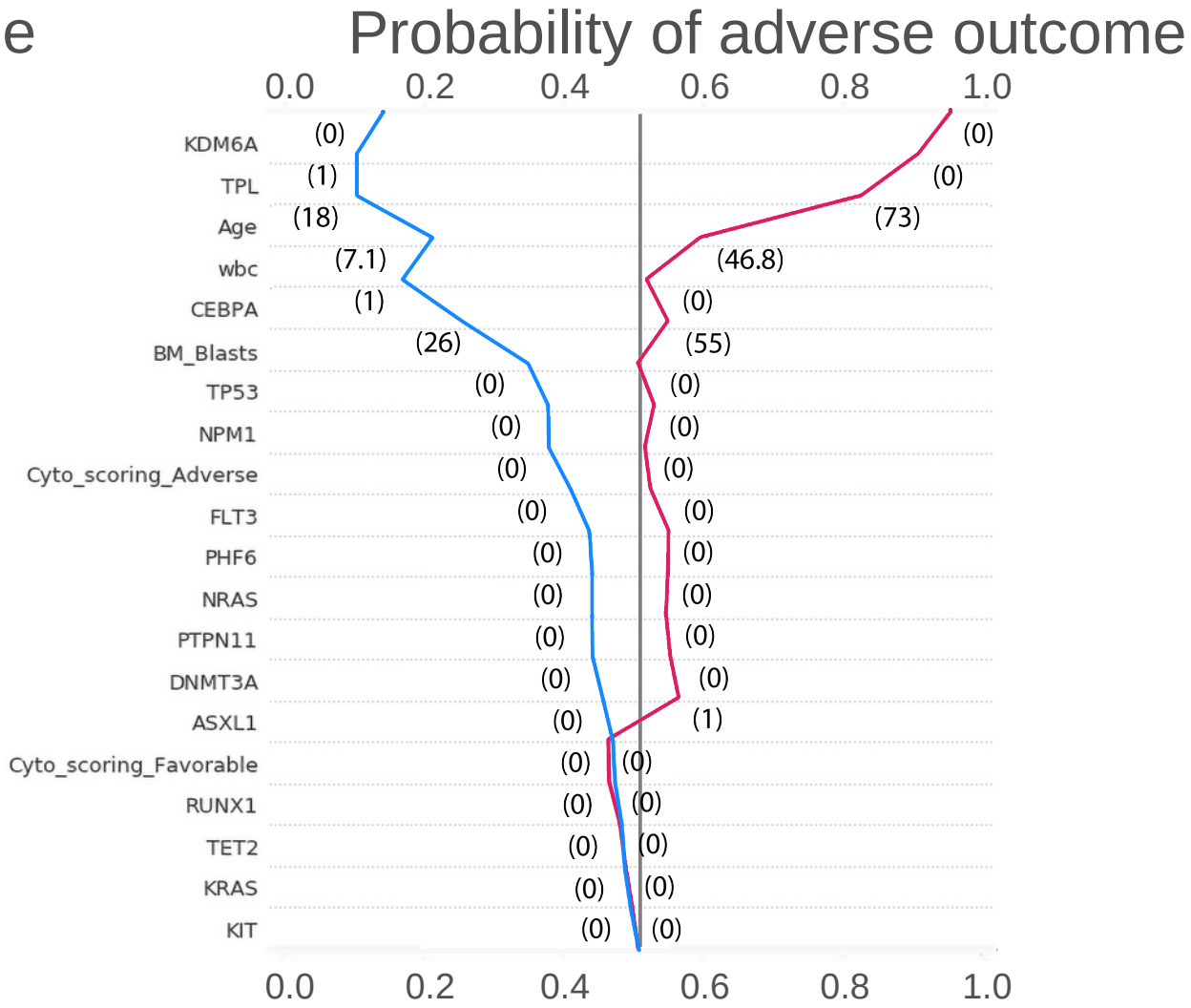
- Features found to significantly impact the prognostic model:
  - Age, transplant status, WBC, bone marrow blast %, cytogenetics
  - ASXL1, CEBPA, DNMT3A, FLT3, KDM6A, KIT, KRAS, NPM1, NRAS, PHF6, PTPN11, RUNX1, TET2, TP53
- The C-index for this new clinical-genomic model was 0.80, significantly outperforming ELN classification (0.59)



©2024 Mayo Foundation for Medical Education and Research

Feature contribution – AML prognostication



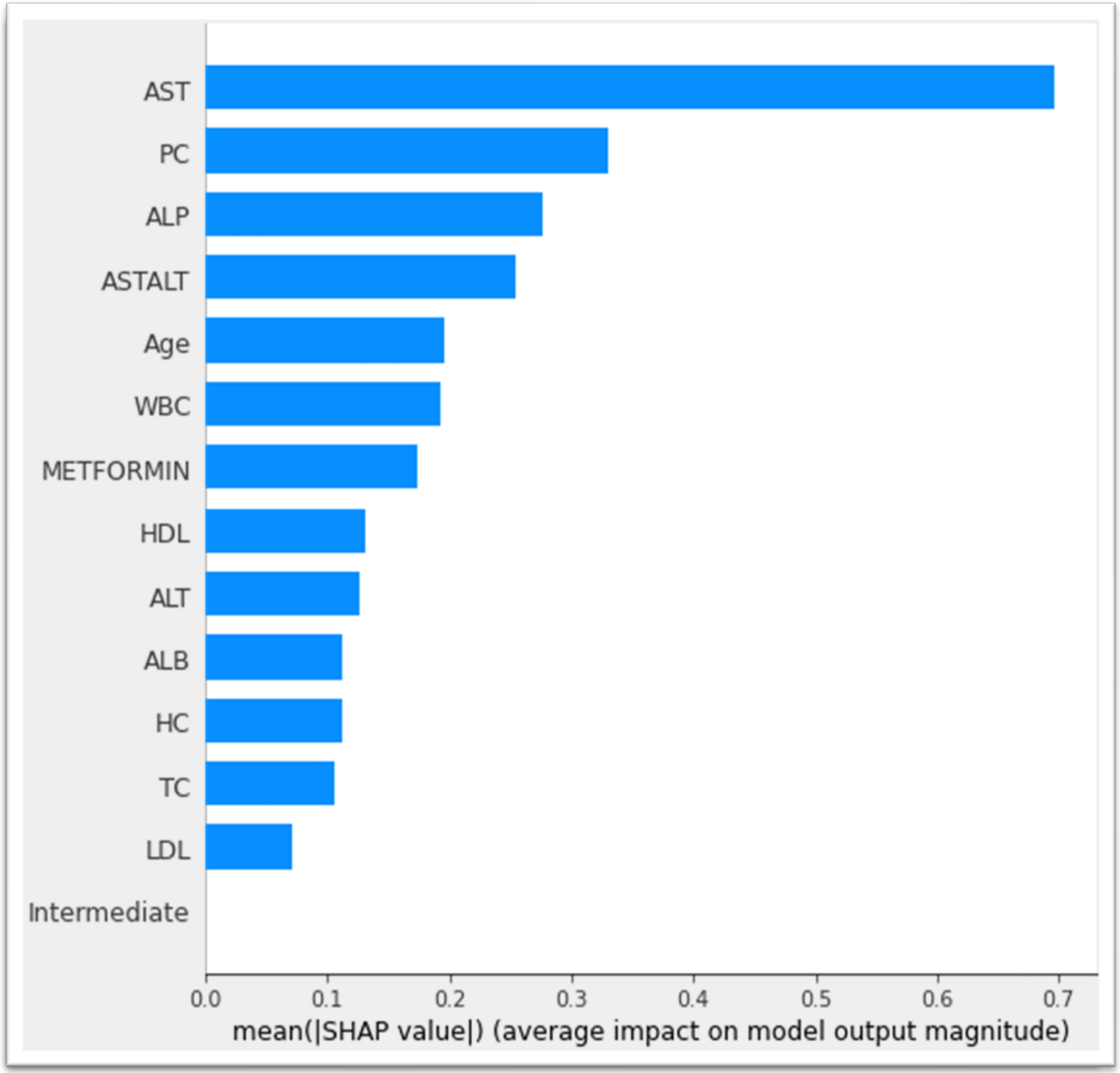


©2024 Mayo Foundation for Medical Education and Research

USE CASE EXAMPLE

LIVER FIBROSIS

- Machine learning model to identify patients with liver fibrosis who were indeterminant by FIB-4 criteria
- 960 patients in cohort, divided into training and test cohort
- The machine learning model correctly classified 80% of the indeterminant subgroup



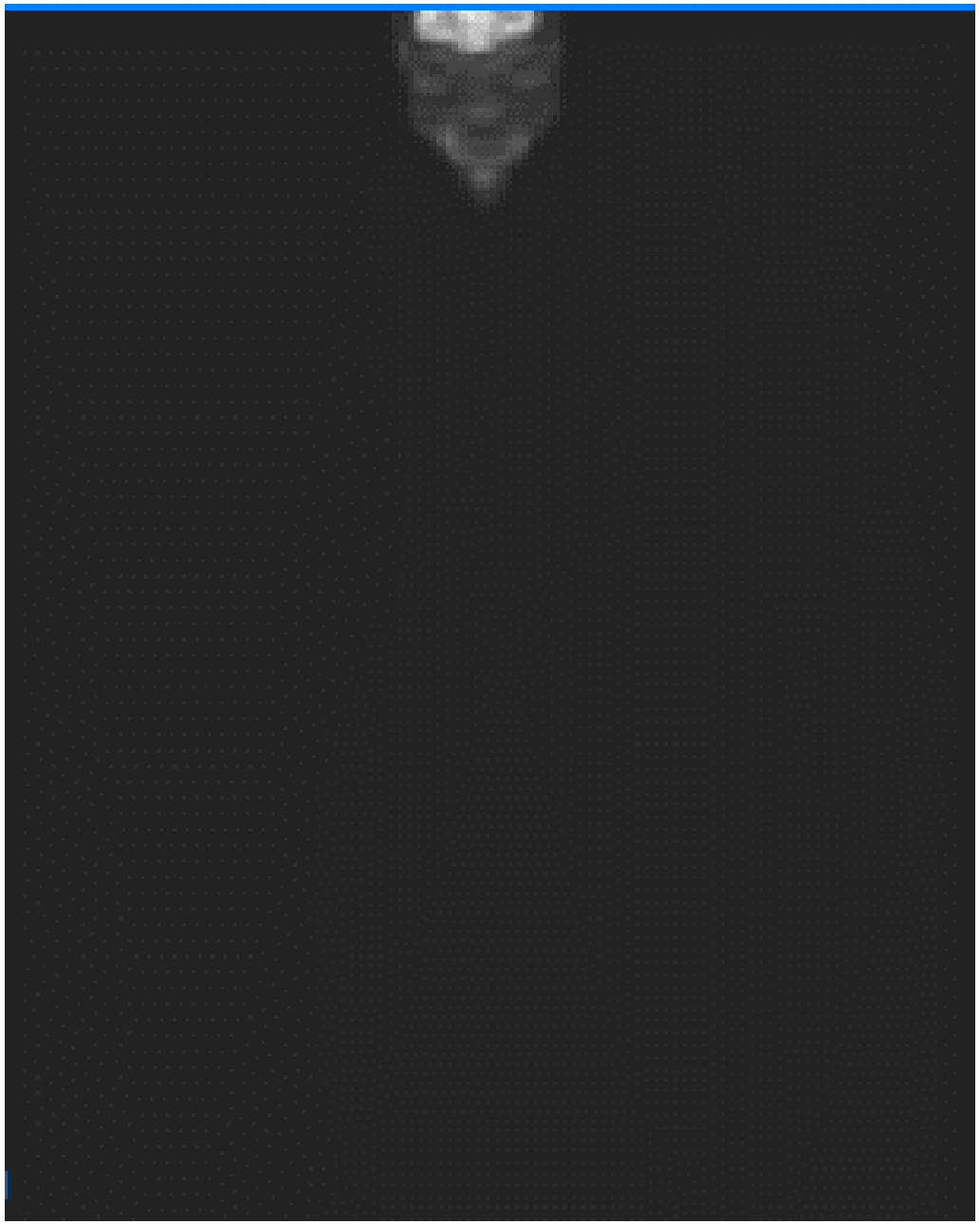
Aggarwal, M., **Shreve, J.** and McCullough, A., 2021. Machine learning model correctly identifies patients with advanced liver fibrosis which are indeterminate by FIB-4 index in non-alcoholic fatty liver disease. *Gastroenterology*, *160*(6), pp.S-114.

©2024 Mayo Foundation for Medical Education and Research

USE CASE EXAMPLE

MYELOMA

- Computer vision prognostication based on baseline PET scan at diagnosis
- PET CT scans are analyzed using the nnU-Net deep learning architecture
- AI-segmented tumor then is used to generate radiomics features
- 3 year OS predictive modeling



Shreve, J. et al., 2023 Predicting high-risk disease biology using artificial intelligence based FDG PET/CT radiomics in newly diagnosed multiple myeloma. *HemaSphere*, *7*(S3), p.e0586106.

©2024 Mayo Foundation for Medical Education and Research

```
Case-1_original_shape_Elongation: 0.7251793505114508
Case-1_original_shape_Flatness: 0.046593588668369 47
Case-1_original_shape_LeastAxisLength: 20.26093754180156
Case-1_original_shape_MajorAxisLength: 434.8438942106192
Case-1_original_shape_Maximum2DDiameterColumn: 182.23109850108727
Case-1_original_shape_Maximum2DDiameterRow: 31.396716244387804
Case-1_original_shape_Maximum2DDiameterSlice: 26.542066726097403
Case-1_original_shape_Maximum3DDiameter: 478.0683577252242
Case-1_original_shape_MeshVolume: 7624.514732716314
Case-1_original_shape_MinorAxisLength: 315.3398127775269
Case-1_original_shape_Sphericity: 0.6032930130850189
Case-1_original_shape_SurfaceArea: 3105.255222004601
Case-1_original_shape_SurfaceVolumeRatio: 0.40727250597079256
Case-1_original_shape_VoxelVolume: 8345.312087496128
Case-1_original_firstorder_10Percentile: 97.1384765625
Case-1_original_firstorder_90Percentile: 431.7506713867188
Case-1_original_firstorder_Energy: 14103991.212084094
Case-1_original_firstorder_Entropy: 4.05925183703205
Case-1_original_firstorder_InterquartileRange: 154.24715042114258
Case-1_original_firstorder_Kurtosis: 3.980125858002287
Case-1_original_firstorder_Maximum: 654.6848754882812
Case-1_original_firstorder_MeanAbsoluteDeviation: 105.44676520427068
Case-1_original_firstorder_Mean: 235.3647518157959
Case-1_original_firstorder_Median: 197.91791534423828
Case-1_original_firstorder_Minimum: 51.024681091308594
Case-1_original_firstorder_Range: 603.6601943969727
Case-1_original_firstorder_RobustMeanAbsoluteDeviation: 69.61179186894952
Case-1_original_firstorder_RootMeanSquared: 271.0318939957768
Case-1_original_firstorder_Skewness: 1.1927087373948912
Case-1_original_firstorder_TotalEnergy: 613032335.1257529
Case-1_original_firstorder_Uniformity: 0.07194010416666666
```

Radiomics feature extraction produces 120 discrete machine-readable variables from a single segmented imaging scan

USE CASE EXAMPLE

MYELOMA

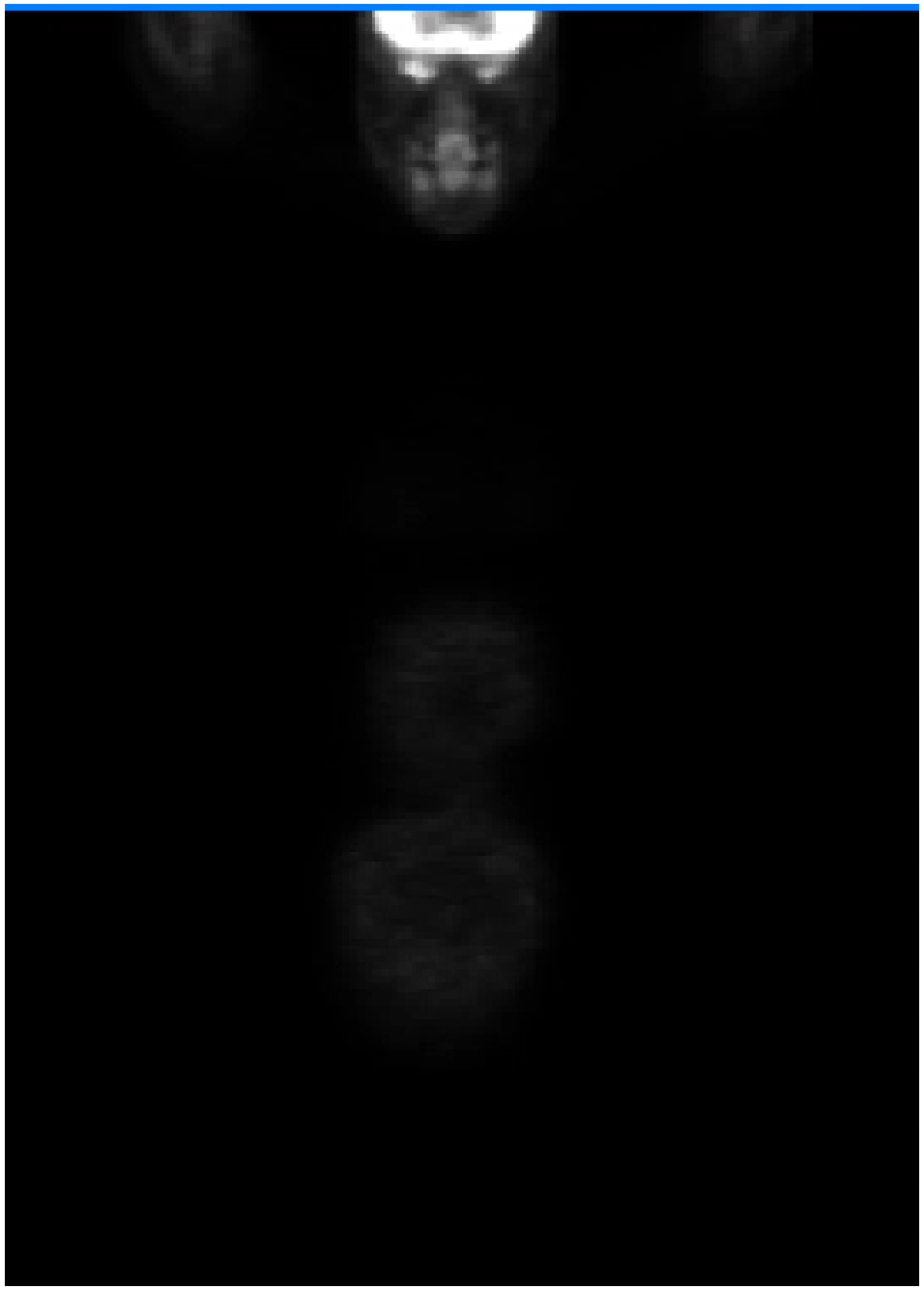
- A total of 506 myeloma patients were processing using the computer vision algorithm
- Failure to achieve 3-year overall survival was strongly associated with maximum 2D tumor diameter (OR 2.26, 95%CI 1.45-3.54, p<0.001).
- Able to predict certain high-risk translocations based solely on automated PET CT radiomics

©2024 Mayo Foundation for Medical Education and Research

USE CASE EXAMPLE

LYMPHOMA

- Computer vision prognostication based on baseline PET scan at diagnosis
- Compared PET scan computer vision prognostication to current gold standard, the NCCN-IPI
- Radiomics features were combined with clinical features to create an integrated model



Shreve, J. et al., 2023. Artificial Intelligence Derived Changes between Baseline and Interim FDG-PET/CT Radiomics Features Are Associated with Survival Outcomes in Diffuse Large B-Cell Lymphoma (DLBCL). *Blood*, *142*, p.5026.

©2024 Mayo Foundation for Medical Education and Research

BACKGROUND

- Manual chart review identified PET CTs within 1 month prior to 1st line therapy, n=861
- Segmentation was accomplished using the nnU-Net architecture loaded with a pretrained lymphoma-specific segmentation model [1]
- Area of interest was determined using a deep learning convolutional neural network (CNN)

[1] Blanc-Durand et al. 2021. Fully automatic segmentation of diffuse large B cell lymphoma lesions on 3D FDG-PET/CT for total metabolic tumour volume prediction using a convolutional neural network. *European Journal of Nuclear Medicine and Molecular Imaging*, *48*, pp.1362-1370.

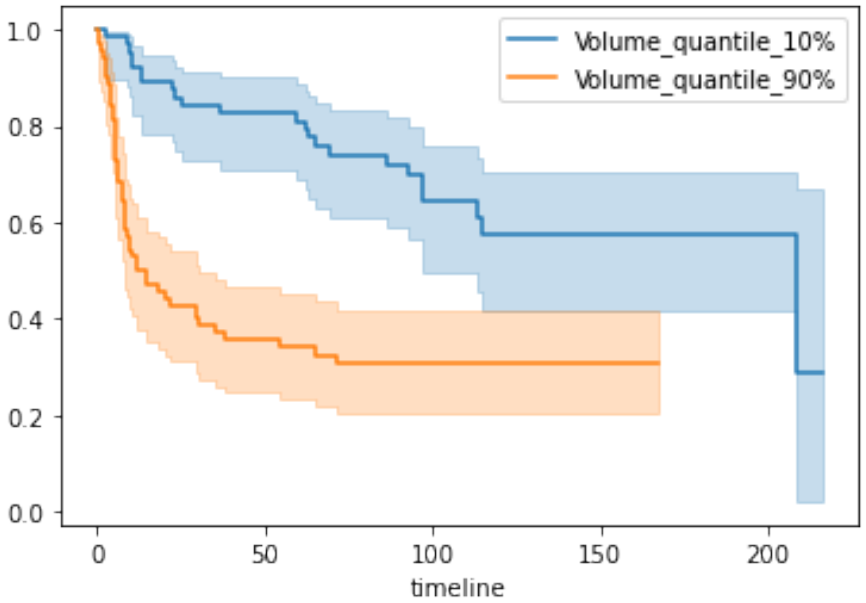
©2024 Mayo Foundation for Medical Education and Research

RESULTS

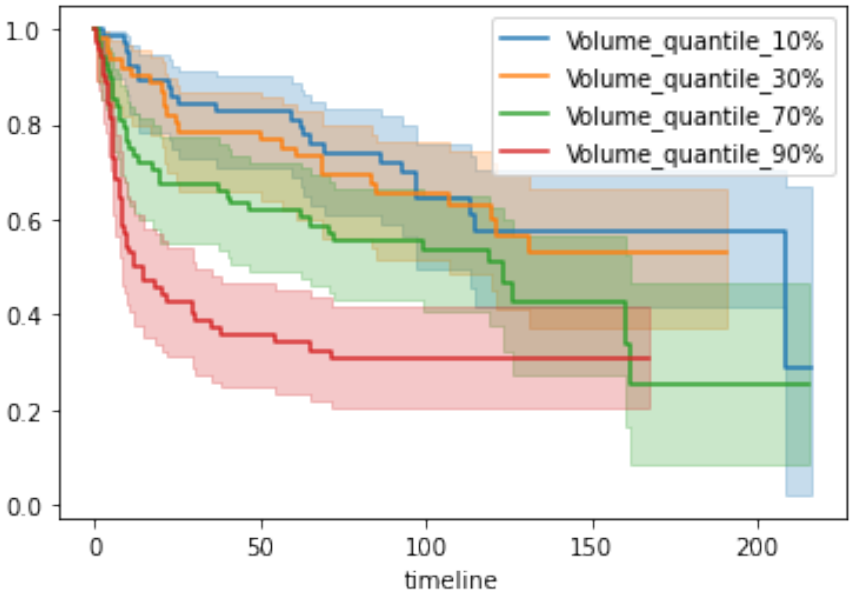
- Evaluation of single radiomics features: **TMTV**

test_statistic	p	-log2(p)
23.87	<0.005	19.89

Kaplan Meier plot of TMTV quantiles (EFS)



Kaplan Meier plot of TMTV quantiles (EFS)



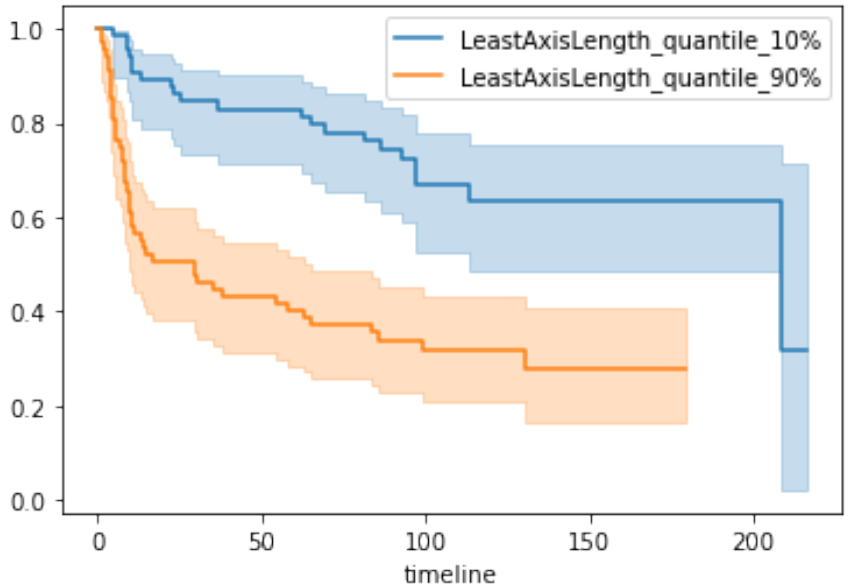
©2024 Mayo Foundation for Medical Education and Research

RESULTS

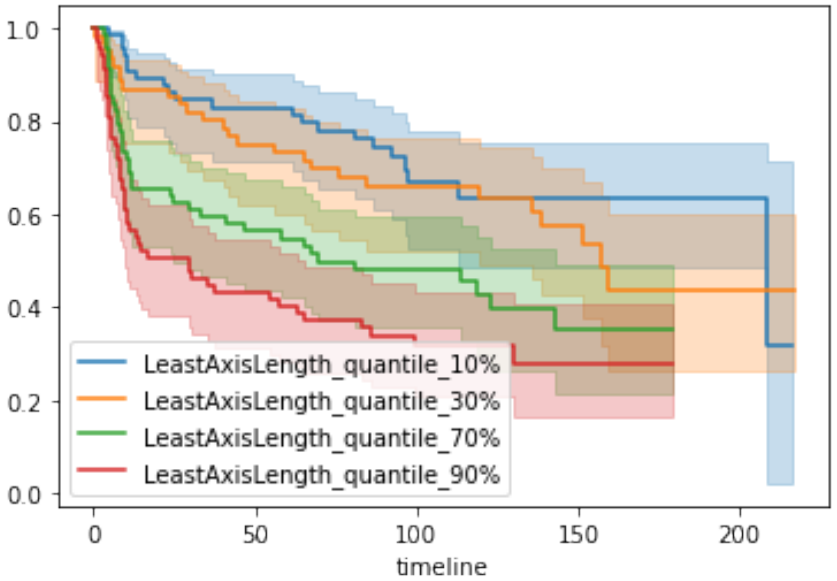
- Evaluation of single radiomics features: **LeastAxisLength**

test_statistic	p	-log2(p)
23.20	<0.005	19.39

Kaplan Meier plot of quantiles (EFS)



Kaplan Meier plot of quantiles (EFS)



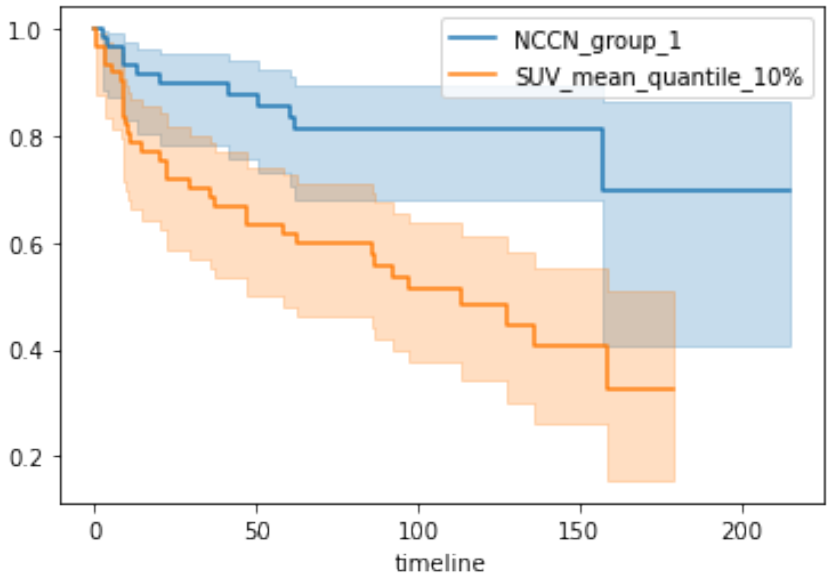
©2024 Mayo Foundation for Medical Education and Research

RESULTS

- Evaluation of single radiomics features: **SUV_mean**

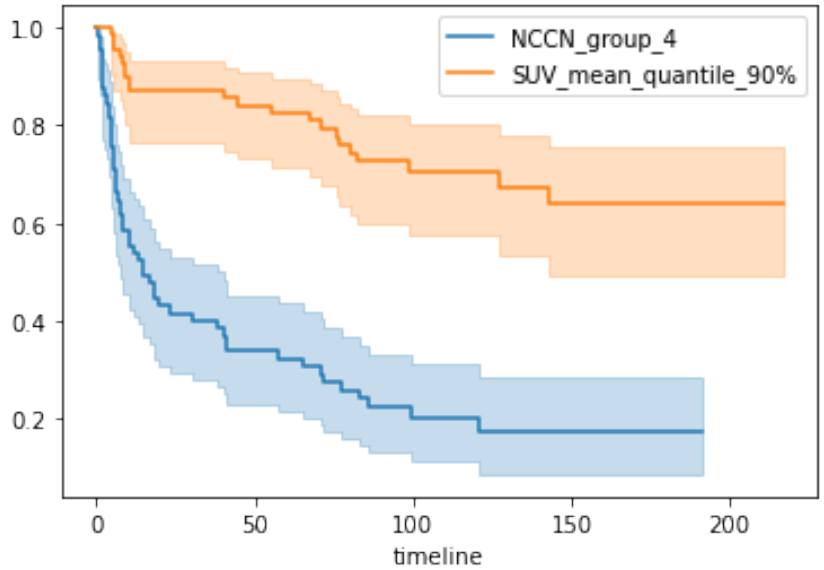
```
---
 test_statistic      p  -log2(p)
          12.29 <0.005     11.10
```

Kaplan Meier plot of quantiles (EFS)




```
---
 test_statistic      p  -log2(p)
          41.80 <0.005     33.20
```

Kaplan Meier plot of quantiles (EFS)



©2024 Mayo Foundation for Medical Education and Research

RESULTS

- Cox Regression of single radiomics features against EFS, right-sided censored (expanded data, expunged missing rows)
  - NCCN-IPI: 64.51% (adjusted to missing data without radiomics)

Partial AIC	2857.46
log-likelihood ratio test	70.36 on 1 df

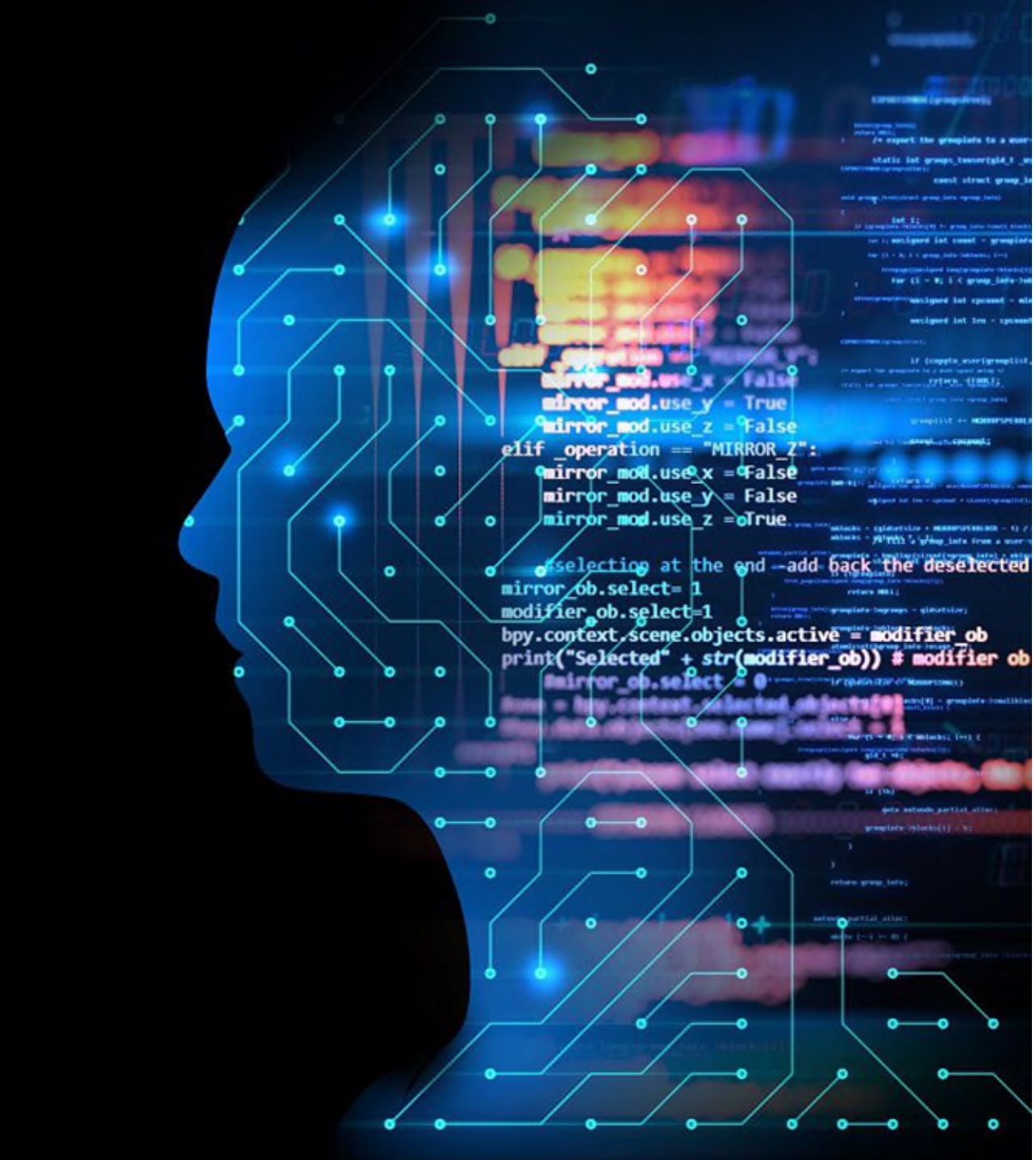
  - 6 radiomics features + 2 labs: 68.16%

Concordance	0.68
Partial AIC	2821.82
log-likelihood ratio test	120.00 on 8 df

  - 5 radiomics features + 8 clinical features: 70.04%

Concordance	0.70
Partial AIC	2808.02
log-likelihood ratio test	147.79 on 15 df
-log2(p) of ll-ratio test	77.00

©2024 Mayo Foundation for Medical Education and Research



ARTIFICIAL INTELLIGENCE

AGENDA

- Why use AI & how does it work
- AI in oncology examples
- Where is the AI revolution?
- How to evaluate AI research

©2024 Mayo Foundation for Medical Education and Research

THE REVOLUTION

- "After hearing for several decades that computers will soon be able to assist with difficult diagnoses, the practicing physician may well wonder why the revolution has not occurred…" 1987

- Predictive modeling using AI is the modern correlate of evidence-based medicine (EBM) scoring algorithms
- Simple stochiometric EBM scoring mechanisms are still standard of care
  - Wells' Criteria for PE risk
  - CHA_2DS_2-VASc for atrial fibrillation stroke risk
- A large part of the "revolution" is using AI to improve upon these methods – medical predictions

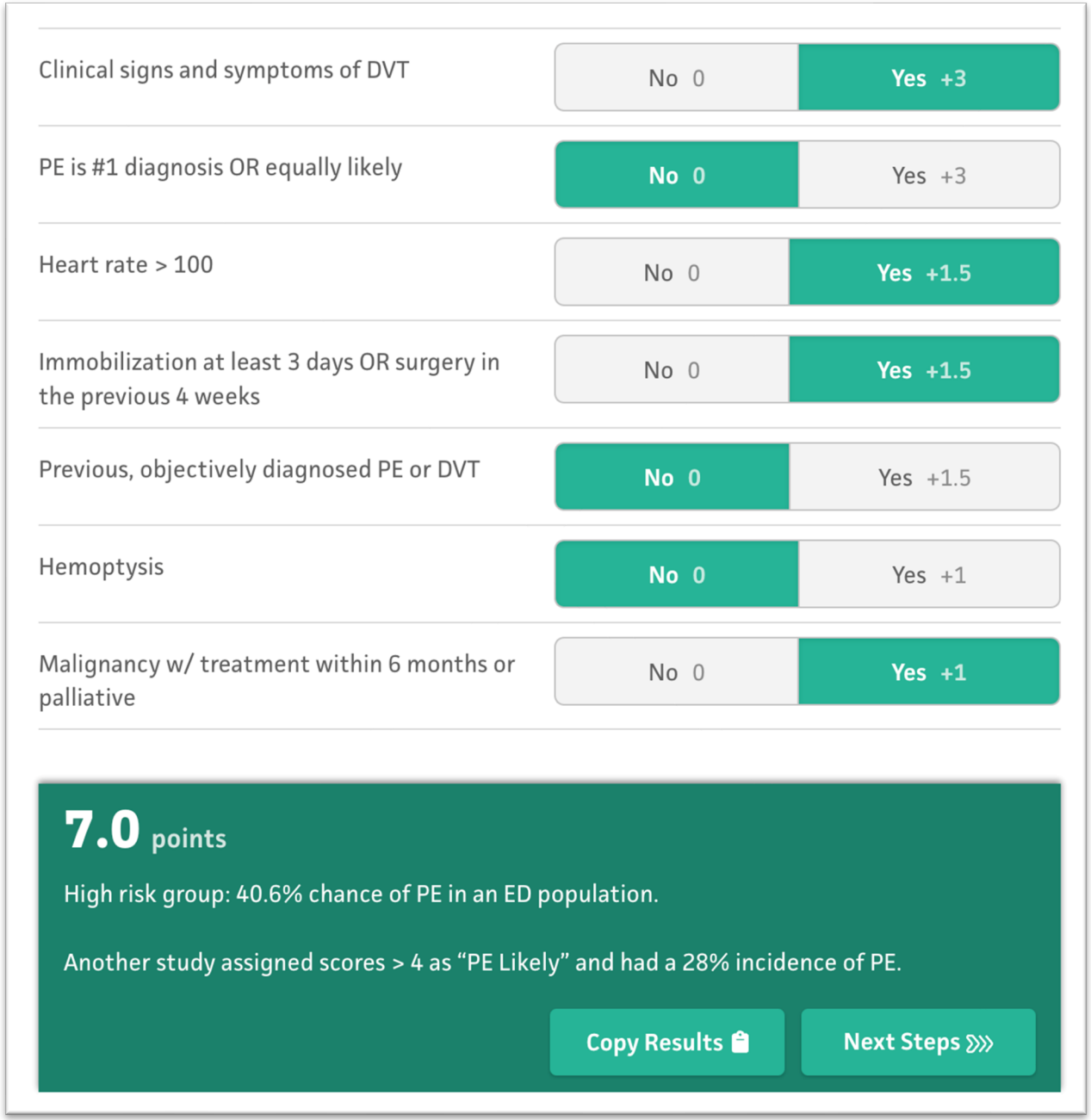
©2024 Mayo Foundation for Medical Education and Research

WHERE IS THE REVOLUTION

WELLS' CRITERIA

- 1995: inception of Wells' rules based on "expert opinion and literature review", pilot study of 91 patients
- 1998: expansion of study to 529 patients
- 2001: seminal study with 1,239 patients the cemented Wells' Criteria into clinical practice

- Since that time many AI-powered approaches have attempted to dethrone Wells' Criteria



©2024 Mayo Foundation for Medical Education and Research

WHERE IS THE REVOLUTION

WELLS' ALTERNATIVE

- EMR-based natural language processing approach
- 3,214 patients used for the model
- 240 patients used as an external control
- AUROC 0.71 when validated on external data
- Dramatically more accurate than Wells' Criteria

Multicenter Study > JAMA Netw Open. 2019 Aug 2;2(8):e198719.
doi: 10.1001/jamanetworkopen.2019.8719.

Development and Performance of the Pulmonary Embolism Result Forecast Model (PERFORM) for Computed Tomography Clinical Decision Support

Imon Banerjee [1 2], Miji Sofela [3], Jaden Yang [4], Jonathan H Chen [5], Nigam H Shah [5], Robyn Ball [4], Alvin I Mushlin [6], Manisha Desai [4], Joseph Bledsoe [7], Timothy Amrhein [8], Daniel L Rubin [1 2], Roham Zamanian [9], Matthew P Lungren [2]

Affiliations + expand

PMID: 31390040 PMCID: PMC6686780 DOI: 10.1001/jamanetworkopen.2019.8719
Free PMC article

Abstract

Importance: Pulmonary embolism (PE) is a life-threatening clinical problem, and computed tomographic imaging is the standard for diagnosis. Clinical decision support rules based on PE risk-scoring models have been developed to compute pretest probability but are underused and tend to underperform in practice, leading to persistent overuse of CT imaging for PE.

Objective: To develop a machine learning model to generate a patient-specific risk score for PE by analyzing longitudinal clinical data as clinical decision support for patients referred for CT imaging for PE.

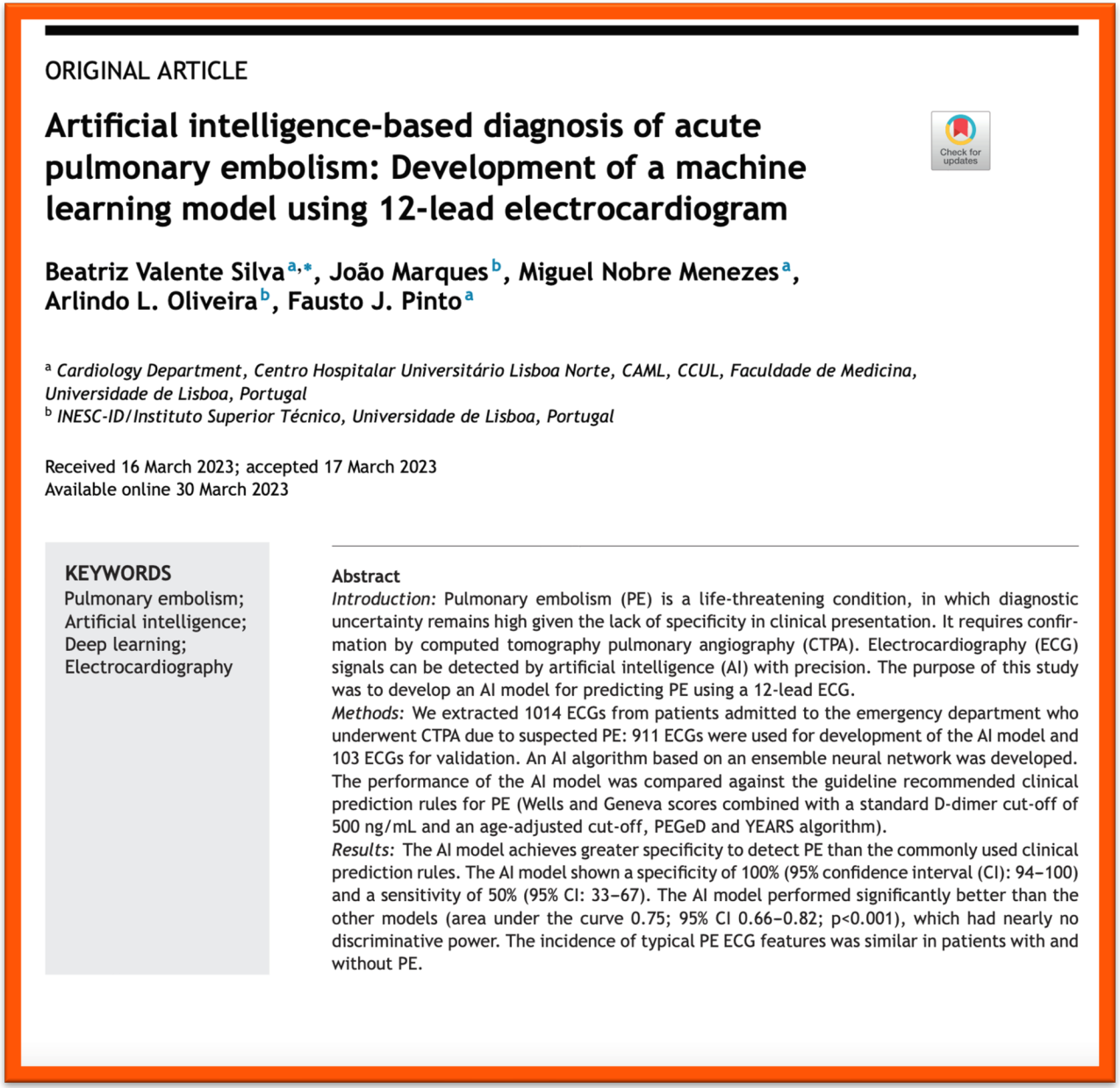
Design, setting, and participants: In this diagnostic study, the proposed workflow for the machine learning model, the Pulmonary Embolism Result Forecast Model (PERFORM), transforms raw electronic medical record (EMR) data into temporal feature vectors and develops a decision analytical model targeted toward adult patients referred for CT imaging for PE. The model was tested on holdout patient EMR data from 2 large, academic medical practices. A total of 3397

©2024 Mayo Foundation for Medical Education and Research

WHERE IS THE REVOLUTION

WELLS' ALTERNATIVE

- ECG deep learning approach without any other features
- 1,014 patients used for the model
- Only internal validation done, no external data
- AUROC 0.75 when validated on hold-out data
- Dramatically more accurate than Wells' Criteria



ORIGINAL ARTICLE

Artificial intelligence-based diagnosis of acute pulmonary embolism: Development of a machine learning model using 12-lead electrocardiogram

Beatriz Valente Silva[a,*], João Marques[b], Miguel Nobre Menezes[a], Arlindo L. Oliveira[b], Fausto J. Pinto[a]

[a] *Cardiology Department, Centro Hospitalar Universitário Lisboa Norte, CAML, CCUL, Faculdade de Medicina, Universidade de Lisboa, Portugal*
[b] *INESC-ID/Instituto Superior Técnico, Universidade de Lisboa, Portugal*

Received 16 March 2023; accepted 17 March 2023
Available online 30 March 2023

KEYWORDS
Pulmonary embolism;
Artificial intelligence;
Deep learning;
Electrocardiography

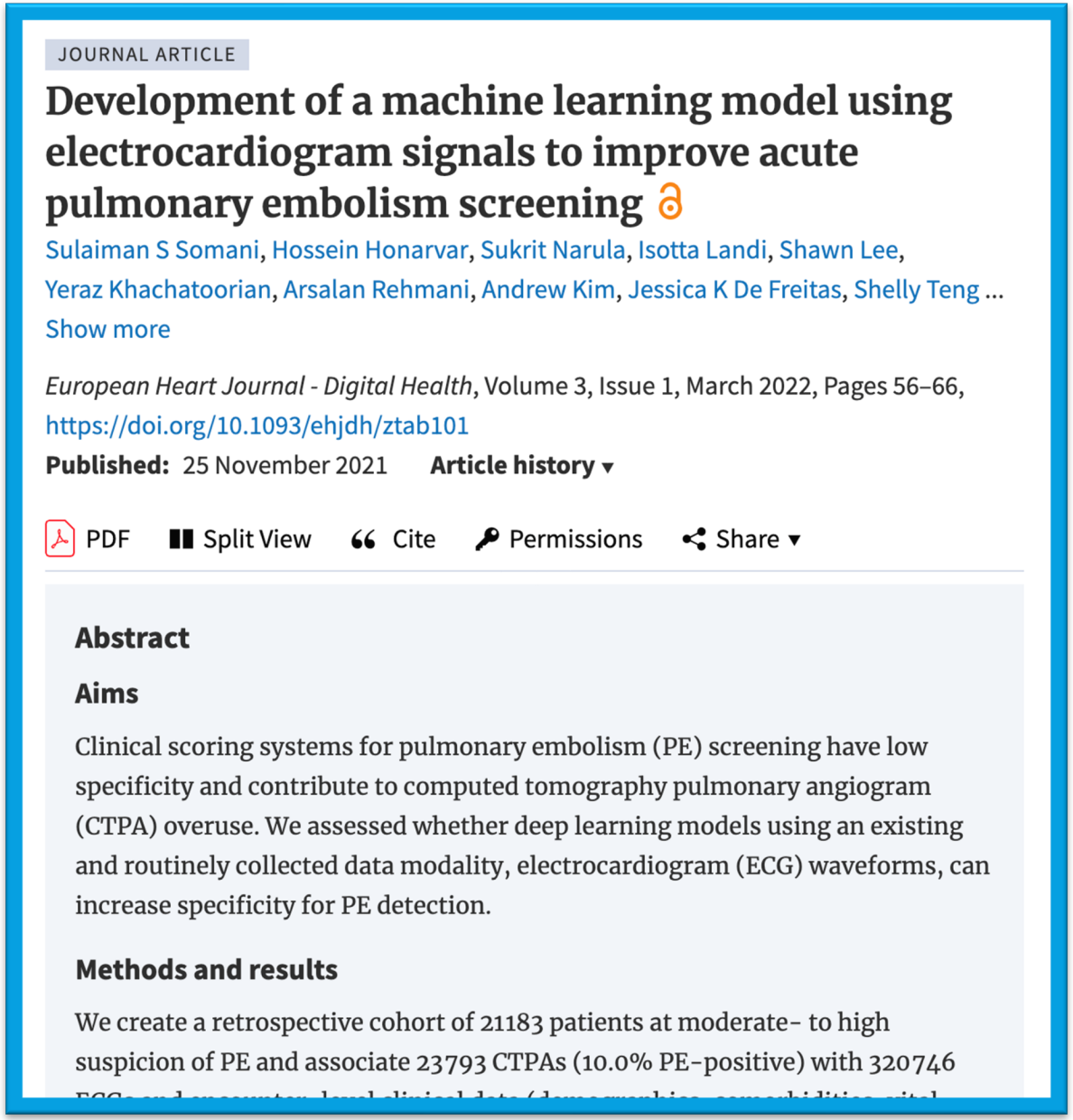
Abstract
Introduction: Pulmonary embolism (PE) is a life-threatening condition, in which diagnostic uncertainty remains high given the lack of specificity in clinical presentation. It requires confirmation by computed tomography pulmonary angiography (CTPA). Electrocardiography (ECG) signals can be detected by artificial intelligence (AI) with precision. The purpose of this study was to develop an AI model for predicting PE using a 12-lead ECG.
Methods: We extracted 1014 ECGs from patients admitted to the emergency department who underwent CTPA due to suspected PE: 911 ECGs were used for development of the AI model and 103 ECGs for validation. An AI algorithm based on an ensemble neural network was developed. The performance of the AI model was compared against the guideline recommended clinical prediction rules for PE (Wells and Geneva scores combined with a standard D-dimer cut-off of 500 ng/mL and an age-adjusted cut-off, PEGeD and YEARS algorithm).
Results: The AI model achieves greater specificity to detect PE than the commonly used clinical prediction rules. The AI model shown a specificity of 100% (95% confidence interval (CI): 94–100) and a sensitivity of 50% (95% CI: 33–67). The AI model performed significantly better than the other models (area under the curve 0.75; 95% CI 0.66–0.82; p<0.001), which had nearly no discriminative power. The incidence of typical PE ECG features was similar in patients with and without PE.

©2024 Mayo Foundation for Medical Education and Research

WHERE IS THE REVOLUTION

WELLS' ALTERNATIVE

- ECG, EMR, NLP data used to train the model in a multimodal approach
- 21,183 patients used for the model
- Only internal validation done, no external data
- AUROC 0.84 when validated on hold-out data
- Dramatically more accurate than Wells' Criteria



JOURNAL ARTICLE

Development of a machine learning model using electrocardiogram signals to improve acute pulmonary embolism screening

Sulaiman S Somani, Hossein Honarvar, Sukrit Narula, Isotta Landi, Shawn Lee, Yeraz Khachatoorian, Arsalan Rehmani, Andrew Kim, Jessica K De Freitas, Shelly Teng ...
Show more

European Heart Journal - Digital Health, Volume 3, Issue 1, March 2022, Pages 56–66, https://doi.org/10.1093/ehjdh/ztab101

Published: 25 November 2021 **Article history** ▾

PDF Split View Cite Permissions Share ▾

Abstract

Aims

Clinical scoring systems for pulmonary embolism (PE) screening have low specificity and contribute to computed tomography pulmonary angiogram (CTPA) overuse. We assessed whether deep learning models using an existing and routinely collected data modality, electrocardiogram (ECG) waveforms, can increase specificity for PE detection.

Methods and results

We create a retrospective cohort of 21183 patients at moderate- to high suspicion of PE and associate 23793 CTPAs (10.0% PE-positive) with 320746

©2024 Mayo Foundation for Medical Education and Research

WELLS' CRITERIA

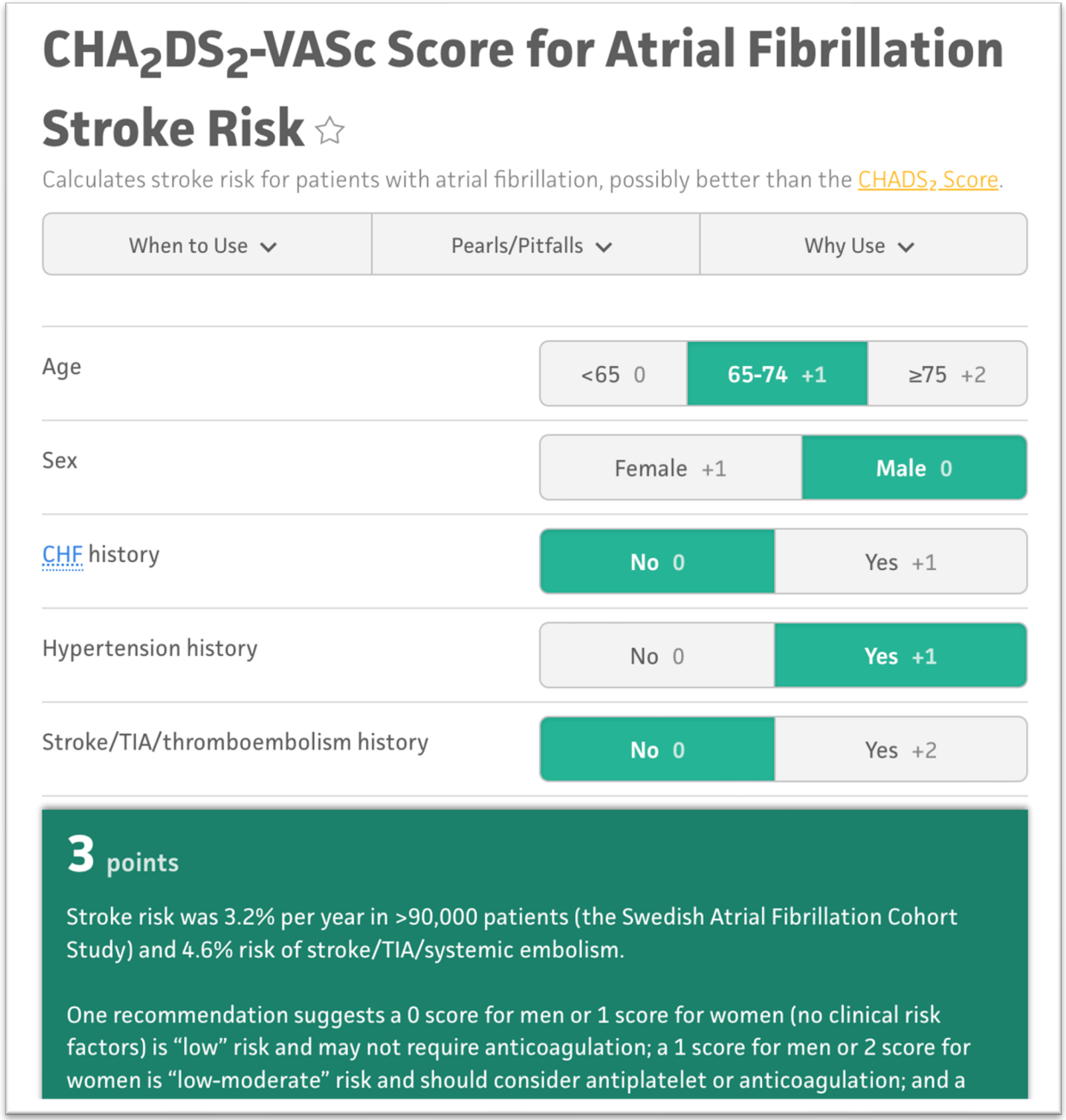
- Understanding why Wells' has been so successful also reveals why the revolution hasn't happened yet

©2024 Mayo Foundation for Medical Education and Research

WHERE IS THE REVOLUTION

CHA_2DS_2-VASC

- Standard of care for atrial fibrillation risk and need for anticoagulation
- Has at least 147 validation studies supporting its use
- Numerous AI-powered alternatives with better AUROC scores
  - Lack external validation
  - Lack calibration studies and other QC metrics of robustness
  - Fall by the wayside

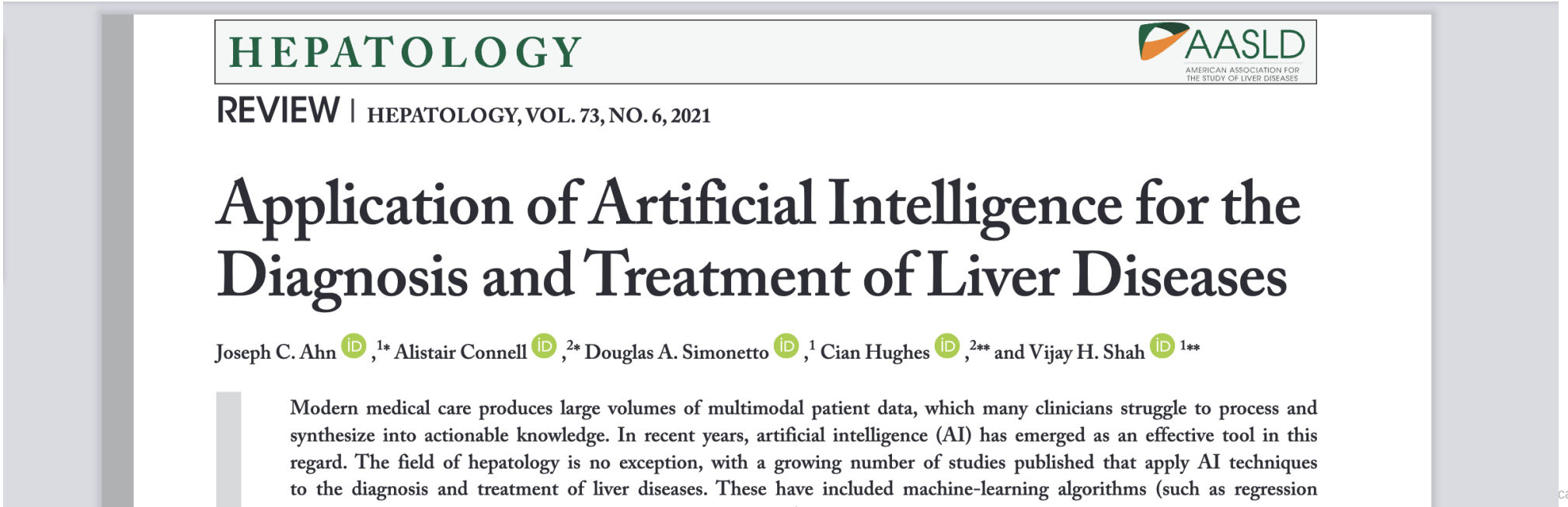


©2024 Mayo Foundation for Medical Education and Research

WHERE IS THE REVOLUTION

LIVER DISEASES

- This review article lists in detail 75 different high-quality AI/ML liver disease models
- Topics range from viral hepatitis, NAFLD, NASH, cirrhosis, acute liver failure, liver transplant
- Very few of these are ever found in clinical practice



HEPATOLOGY

AASLD
AMERICAN ASSOCIATION FOR THE STUDY OF LIVER DISEASES

REVIEW | HEPATOLOGY, VOL. 73, NO. 6, 2021

Application of Artificial Intelligence for the Diagnosis and Treatment of Liver Diseases

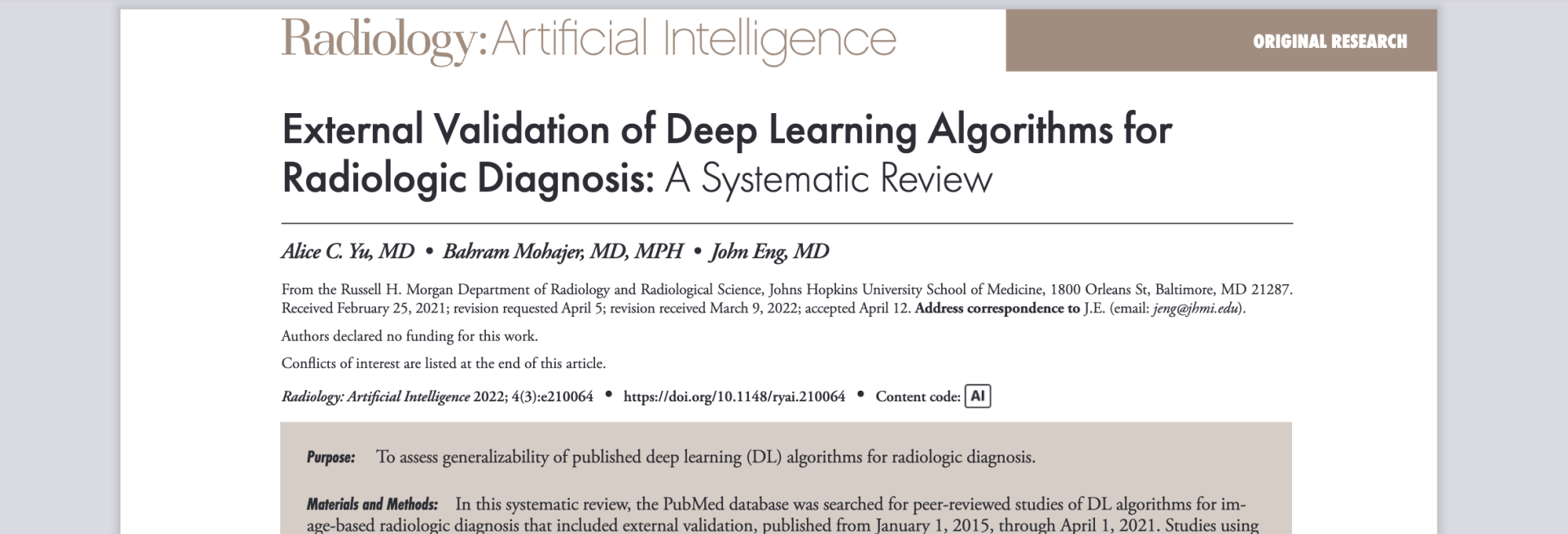
Joseph C. Ahn ,[1*] Alistair Connell ,[2*] Douglas A. Simonetto ,[1] Cian Hughes ,[2**] and Vijay H. Shah [1**]

Modern medical care produces large volumes of multimodal patient data, which many clinicians struggle to process and synthesize into actionable knowledge. In recent years, artificial intelligence (AI) has emerged as an effective tool in this regard. The field of hepatology is no exception, with a growing number of studies published that apply AI techniques to the diagnosis and treatment of liver diseases. These have included machine-learning algorithms (such as regression

WHERE IS THE REVOLUTION

REPRODUCIBILITY

- A review that evaluated 86 radiologic diagnostic models found that 70 such models had decreased performance when applied to external data
- 21 of which produced significantly incongruent results



Radiology: Artificial Intelligence

ORIGINAL RESEARCH

External Validation of Deep Learning Algorithms for Radiologic Diagnosis: A Systematic Review

Alice C. Yu, MD • Bahram Mohajer, MD, MPH • John Eng, MD

From the Russell H. Morgan Department of Radiology and Radiological Science, Johns Hopkins University School of Medicine, 1800 Orleans St, Baltimore, MD 21287.
Received February 25, 2021; revision requested April 5; revision received March 9, 2022; accepted April 12. **Address correspondence to** J.E. (email: *jeng@jhmi.edu*).

Authors declared no funding for this work.

Conflicts of interest are listed at the end of this article.

Radiology: Artificial Intelligence 2022; 4(3):e210064 • https://doi.org/10.1148/ryai.210064 • Content code: AI

Purpose: To assess generalizability of published deep learning (DL) algorithms for radiologic diagnosis.

Materials and Methods: In this systematic review, the PubMed database was searched for peer-reviewed studies of DL algorithms for image-based radiologic diagnosis that included external validation, published from January 1, 2015, through April 1, 2021. Studies using

ENABLE THE REVOLUTION

- Necessary components of a predictive medical model:
  - Training cohort demographics are representative of the population; special attention for minorities and marginalized groups
  - Modeling algorithm should be the lowest complexity possible
  - Should include clinically-relevant features guided by expert consensus; infuses domain knowledge
  - Rigorous quality control metrics
    - k-fold cross validation, bootstrapping, variance estimation, calibration studies
  - Iterative feature reduction to decreases model complexity
    - Ease of adoptability
    - Remaining features should have relative weights reported
  - External validation must be completed at the time of publication

©2024 Mayo Foundation for Medical Education and Research

ARTIFICIAL INTELLIGENCE

AGENDA

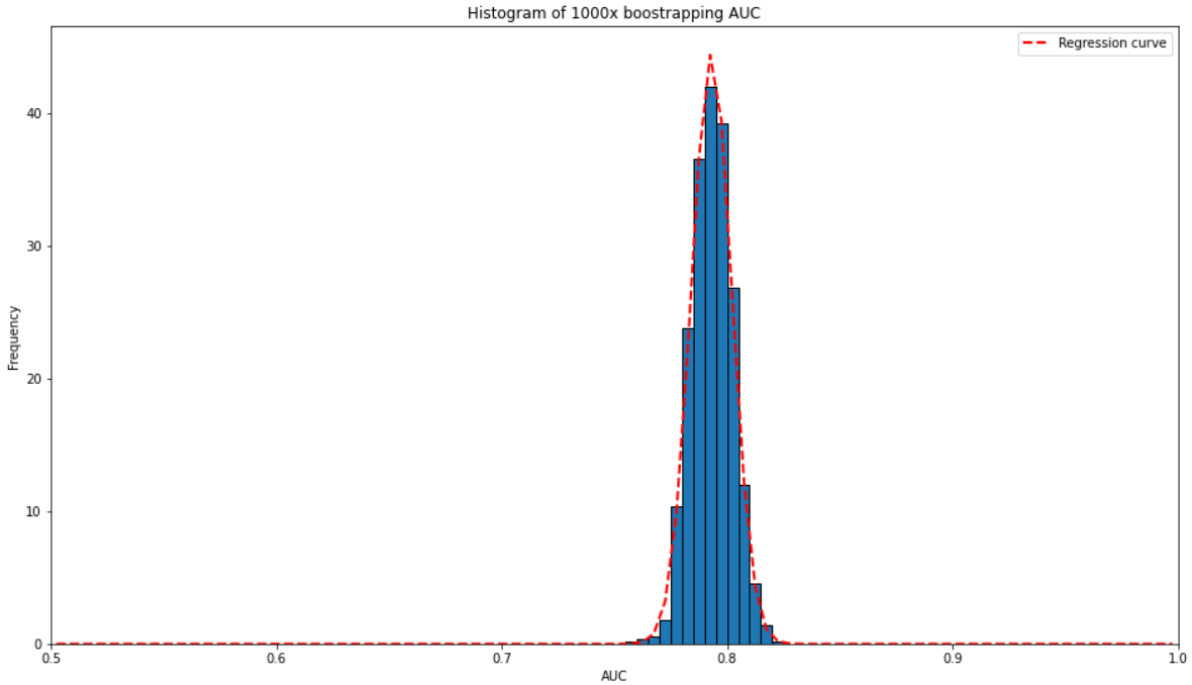
- Why use AI & how does it work
- AI in oncology examples
- Where is the AI revolution?
- How to evaluate AI research

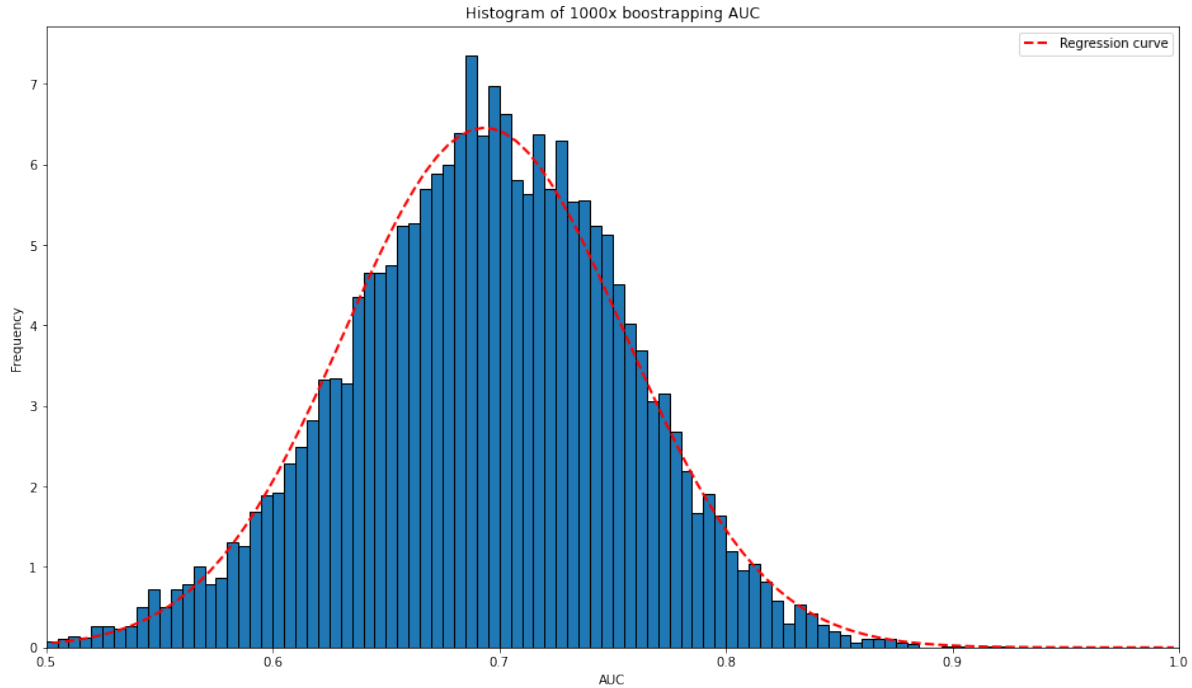
©2024 Mayo Foundation for Medical Education and Research

MODEL EVALUATION

STANDARD QC

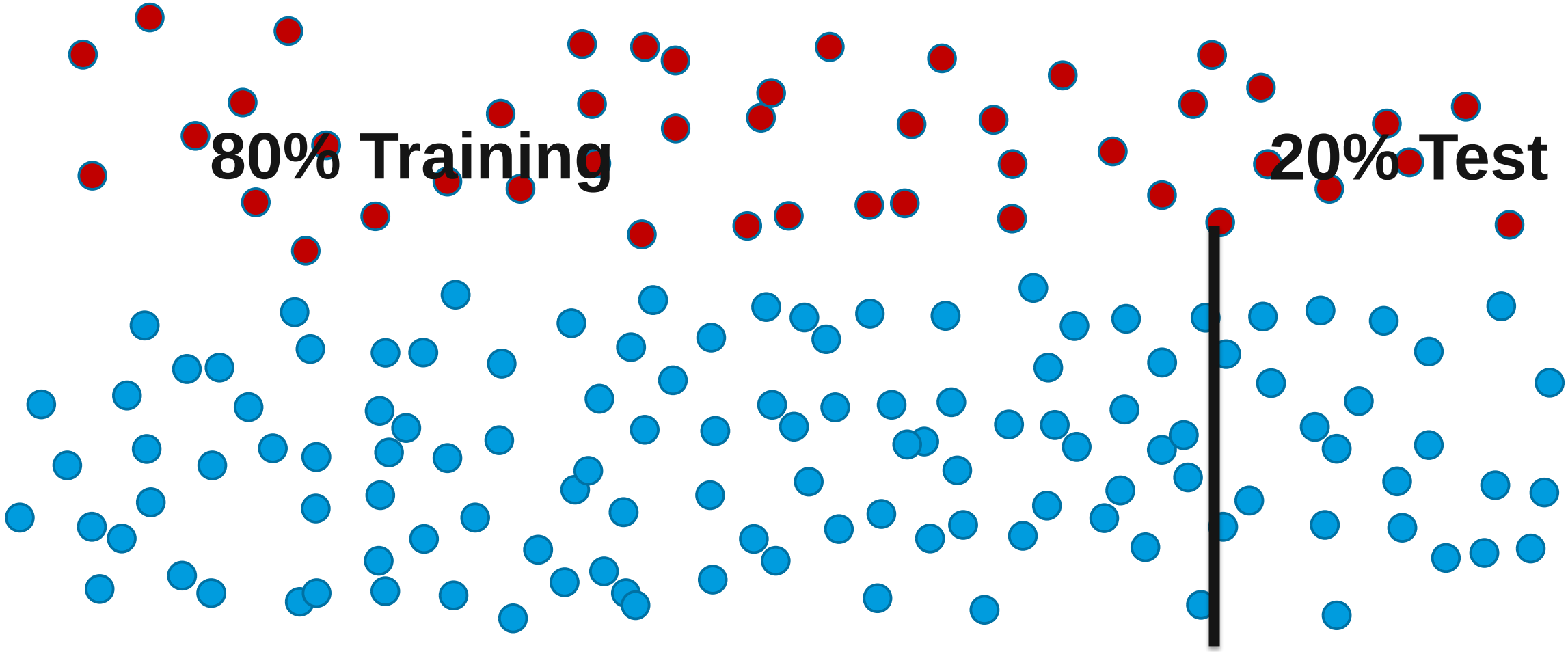
- k-fold cross validation
- AUROC reporting
- Confusion matrix
- Bootstrapping with confidence intervals
- Variance estimation
- Calibration studies
- Feature importance / Shapley statistics
- External validation
- Etc.



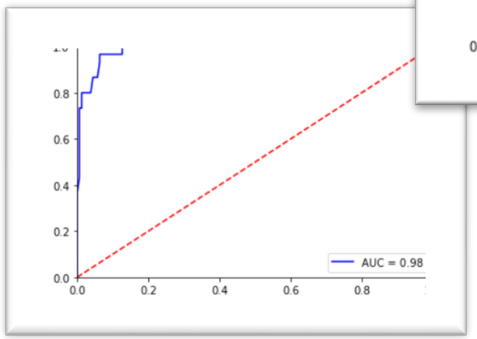


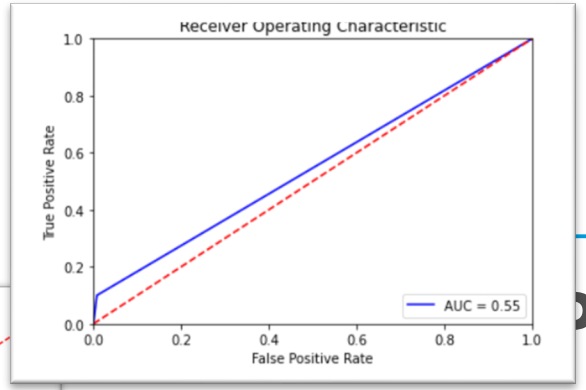
©2024 Mayo Foundation for Medical Education and Research

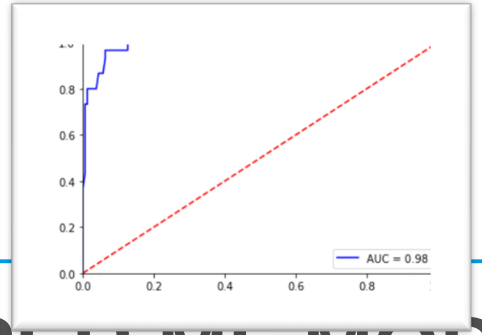
PROPER ML MODEL QC

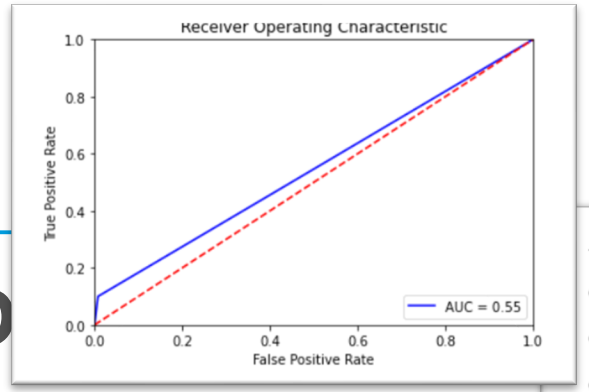


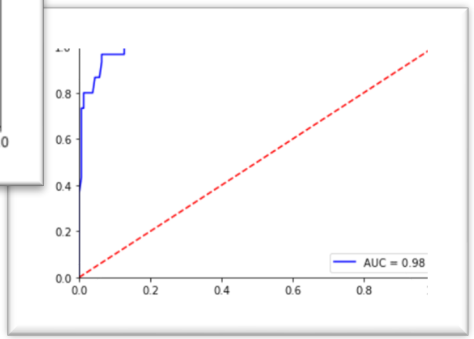
©2024 Mayo Foundation for Medical Education and Research





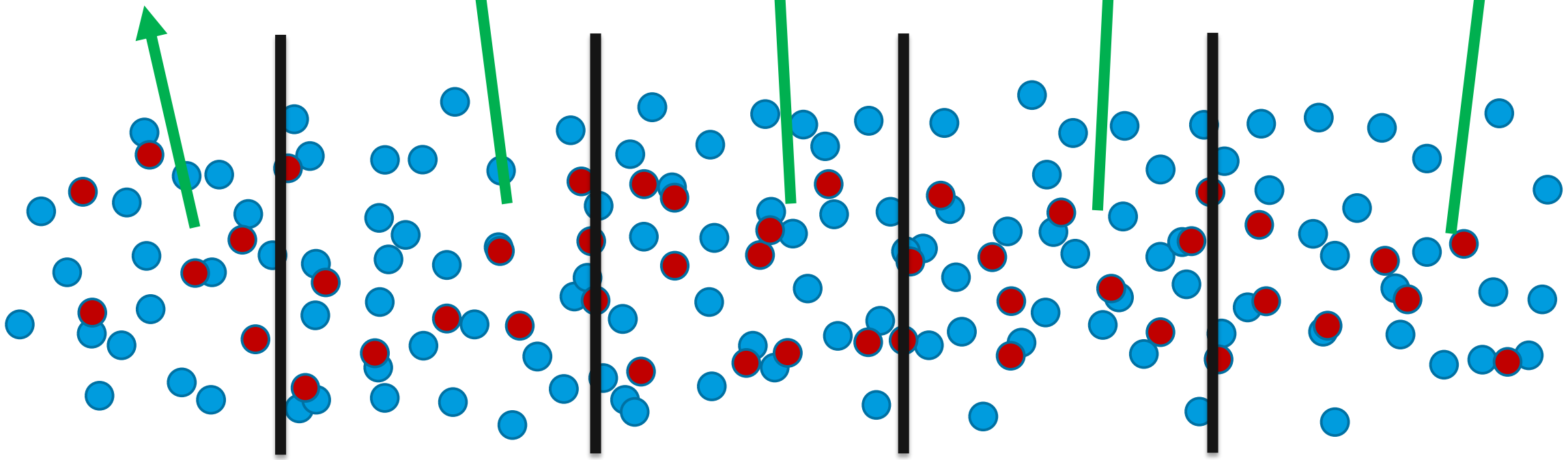






PER ML MOD

Balanced K-fold cross validation



©2024 Mayo Foundation for Medical Education and Research | slide-61

ARTIFICIAL INTELLIGENCE

AGENDA

- Why use AI & how does it work
- AI in oncology examples
- Where is the AI revolution?
- How to evaluate AI research
- Final thoughts

©2024 Mayo Foundation for Medical Education and Research

FINAL THOUGHTS

MULTI-MODAL DATA INTEGRATION

RADIOLOGY

Computer vision for CT / MRI / PET

EMR DATA

Clinical information including PMH, medications, disease course, labs, vitals

INTEGRATED MODEL

GENOMICS

Primary sequencing data, bioinformatics

DIGITAL PATHOLOGY

Computer vision for tumor slides

LANGUAGE ANALYSIS

Natural language processing (NLP) for chart interpretation

©2024 Mayo Foundation for Medical Education and Research

FINAL THOUGHTS

NEAR FUTURE

- **Living databases** that undergo nightly updates directly from EMR data streams, ever increasing study population and the features being evaluated
- **Automatic modeling** using those living databases to continuously improve upon a model as new information is added to the database
- **Multimodal data integration** to begin to capture the actual complexity of the system, edging closer to predictive capabilities

- **Uphill battle**

Must out-perform AND out-validate existing tools and systems if the revolution is to occur.
Must have the maturity, rigor, and reproducibility that is expected in medicine

©2024 Mayo Foundation for Medical Education and Research

CONCLUSIONS

- AI is poised to make explosive changes throughout medicine
- Oncology has a great need for personalized care
- The individual tools already exist, it's now a matter of when, not if

THANK YOU

Jacob Shreve
shreve.jacob@mayo.edu

©2024 Mayo Foundation for Medical Education and Research