Course locations

15/4	vm+nm L3.3 HC
16/4	vm L3.3 HC
	nm L0 WC
17/4	vm L3.3 HC
	nm L4 WC
18/4	vm L1 WC -> need replacement
	nm L1 HC+WC
22/4	vm L3.3 HC
	nm L0 WC
23/4	vm WC
	nm WC
24/4	vm+nm L4 WC
25/4	vm+nm L4 HC+WC
6/5	vm+nm L0 internship project
7/5	vm+nm L0 internship project
8/5	vm+nm L4 WC
13/5	nm L4 WC
14/5	vm L4 WC CompOmics + ProteomicsML
15/5	vm L0 WC
16/5	vm/nm poster presentation

Course specifications

https://github.com/sdgroeve/Machine_Learning_course_UGent_D012554_2024

Course evaluation: Jupyter notebook analysis (20%)

Machine_Learning_course_UGent_D012554_2024 / Spaceship_Titanic /

Add file

sdgroeve Update README.md 8d0e168 · 3 minutes ago History

Name	Last commit message	Last commit date
..		
README.md	Update README.md	3 minutes ago

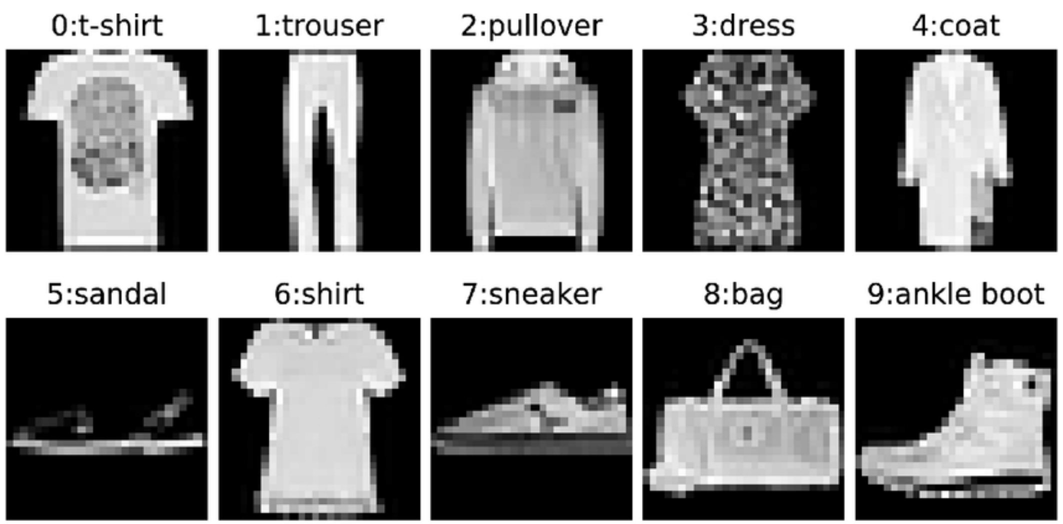
README.md

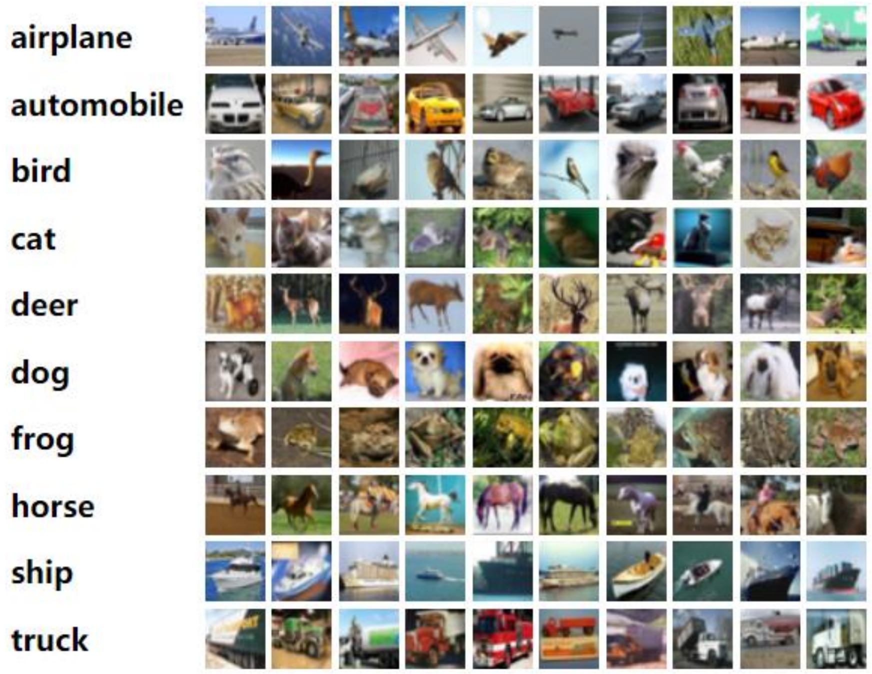
Feature Engineering and Model Building

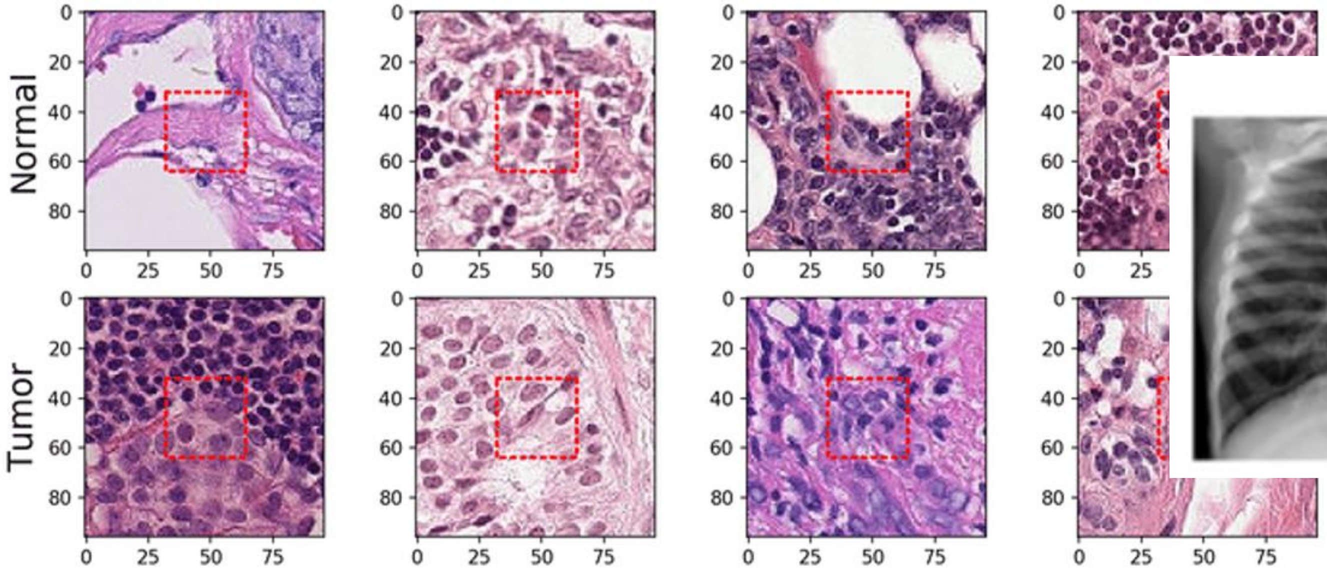
Objective

The goal of this assignment is to build a predictive model from the Spaceship Titanic dataset on Kaggle, which will be

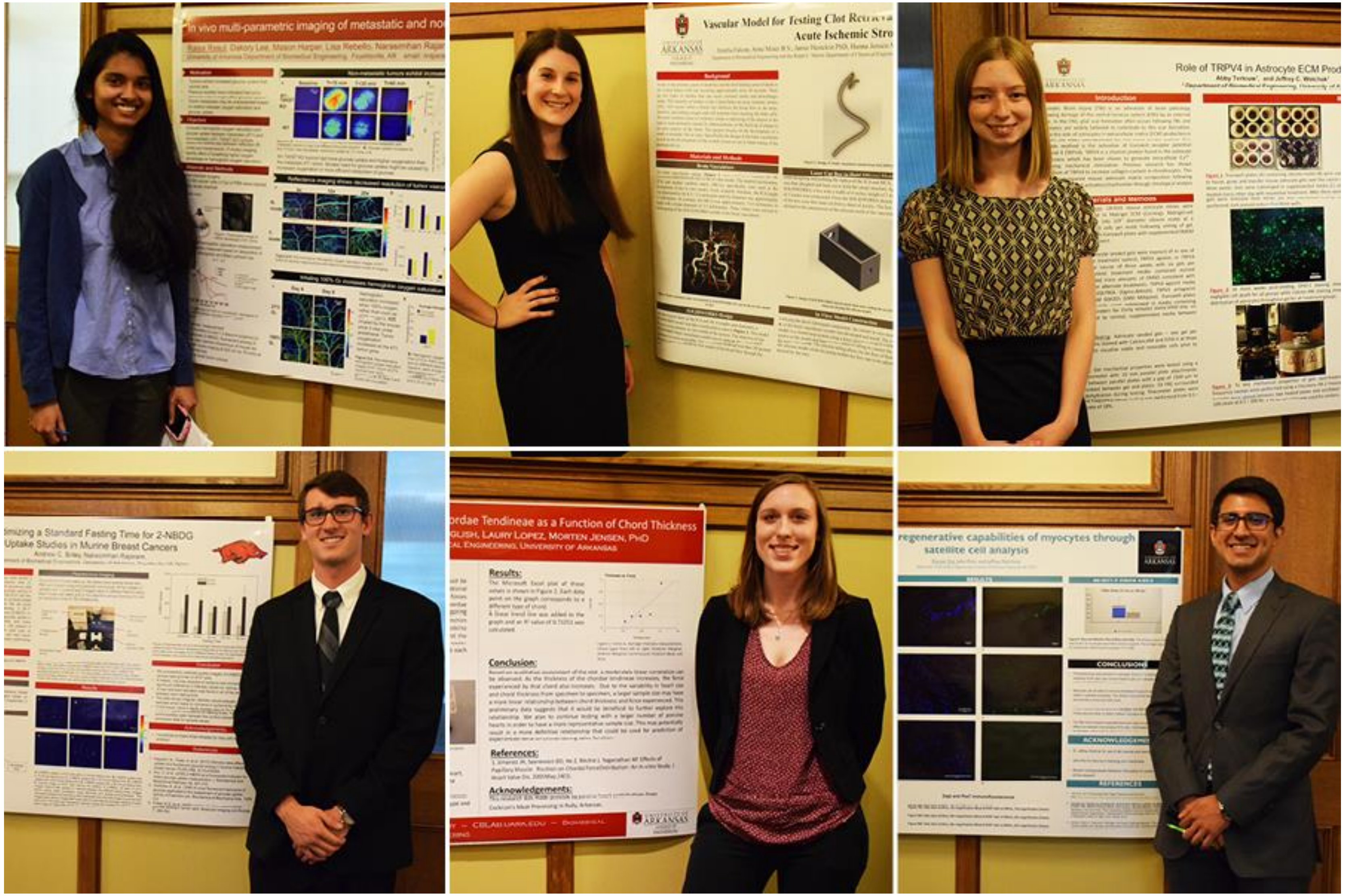
Course evaluation: CNN model building (15%)



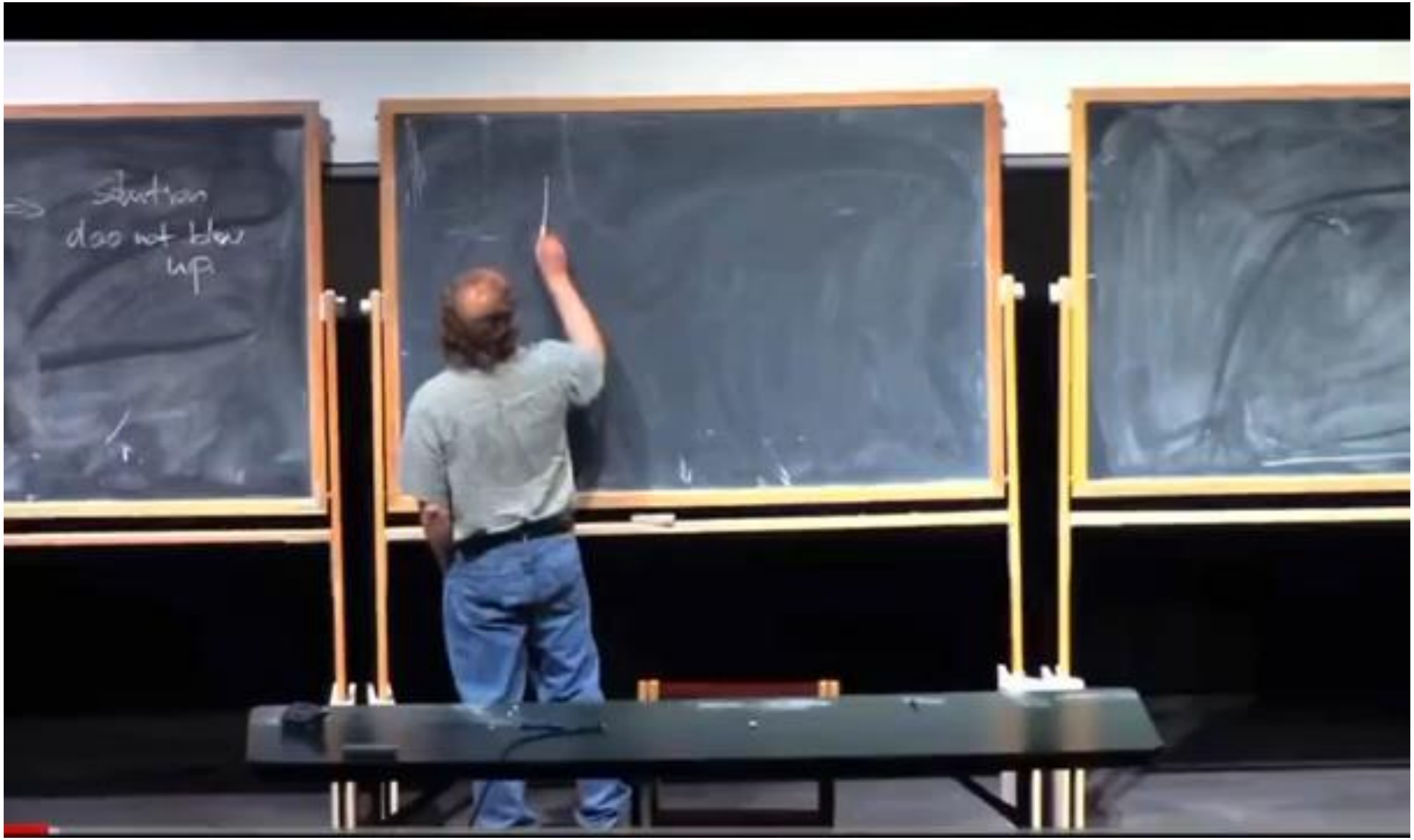




Course evaluation: science poster (15%)

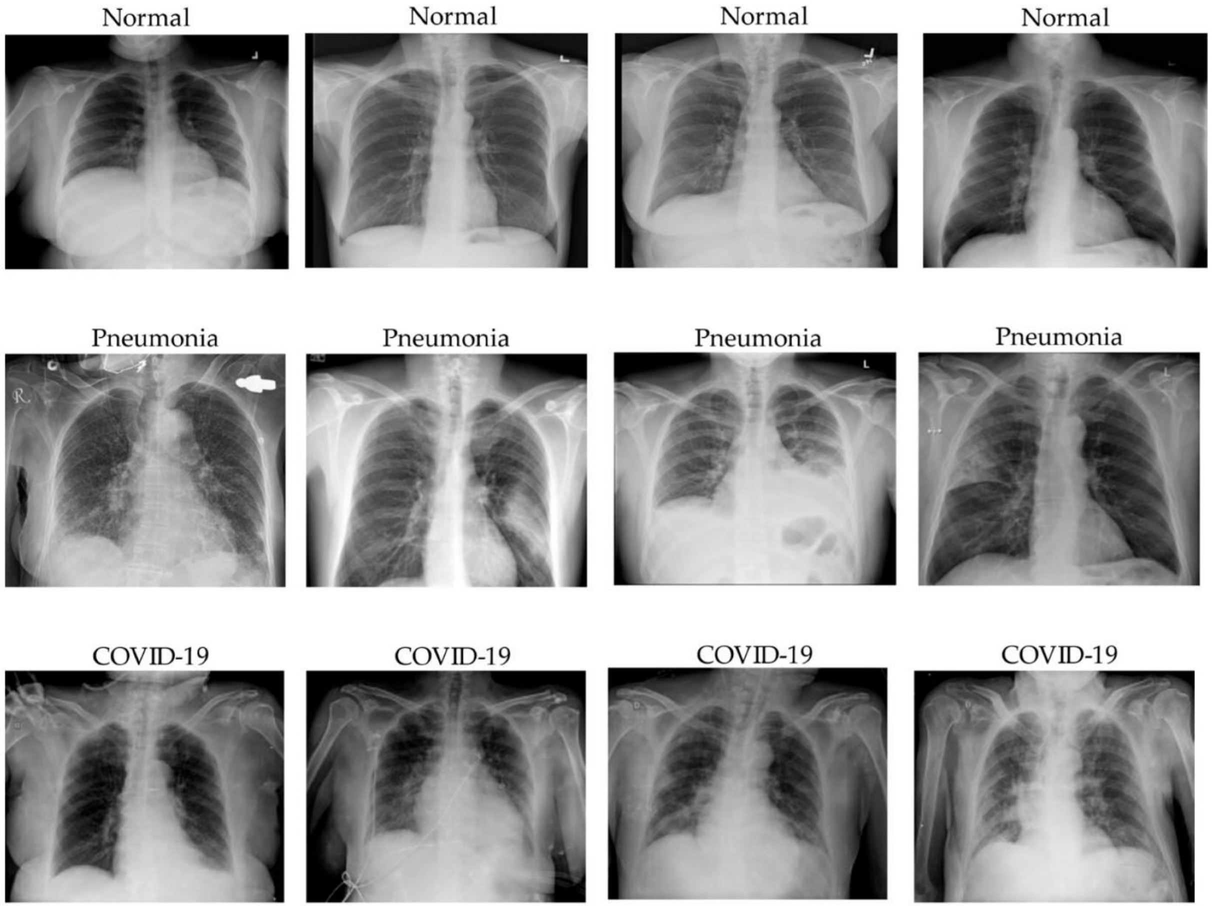


Functions Describe the World (prof. dr. Thomas Garrity)



Applications: Disease Diagnosis and Prediction

Machine learning models are widely used for **diagnosing diseases** from medical imaging, such as detecting cancerous tumours in radiology scans (e.g., breast cancer detection in mammograms) or identifying diabetic retinopathy in eye images.

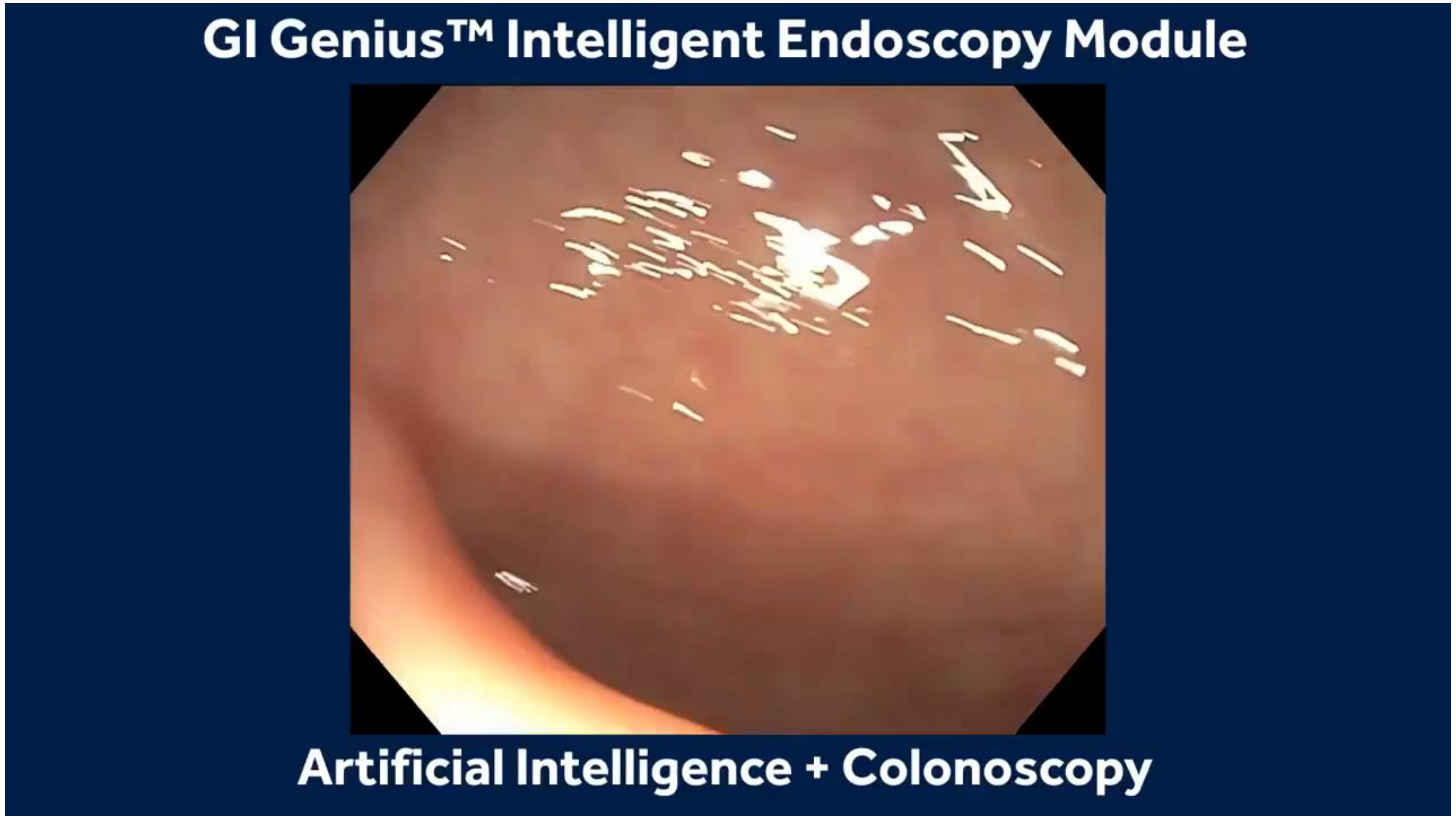


Source: Le Dinh, T, et al. COVID-19 Chest X-ray Classification and Severity Assessment Using Convolutional and Transformer Neural Networks. Appl. Sci. 2022

Applications: Disease Diagnosis and Prediction

Machine learning models are widely used for **diagnosing diseases** from medical imaging (including video), such as detecting cancerous tumours in radiology scans (e.g., breast cancer detection in mammograms) or identifying diabetic retinopathy in eye images.

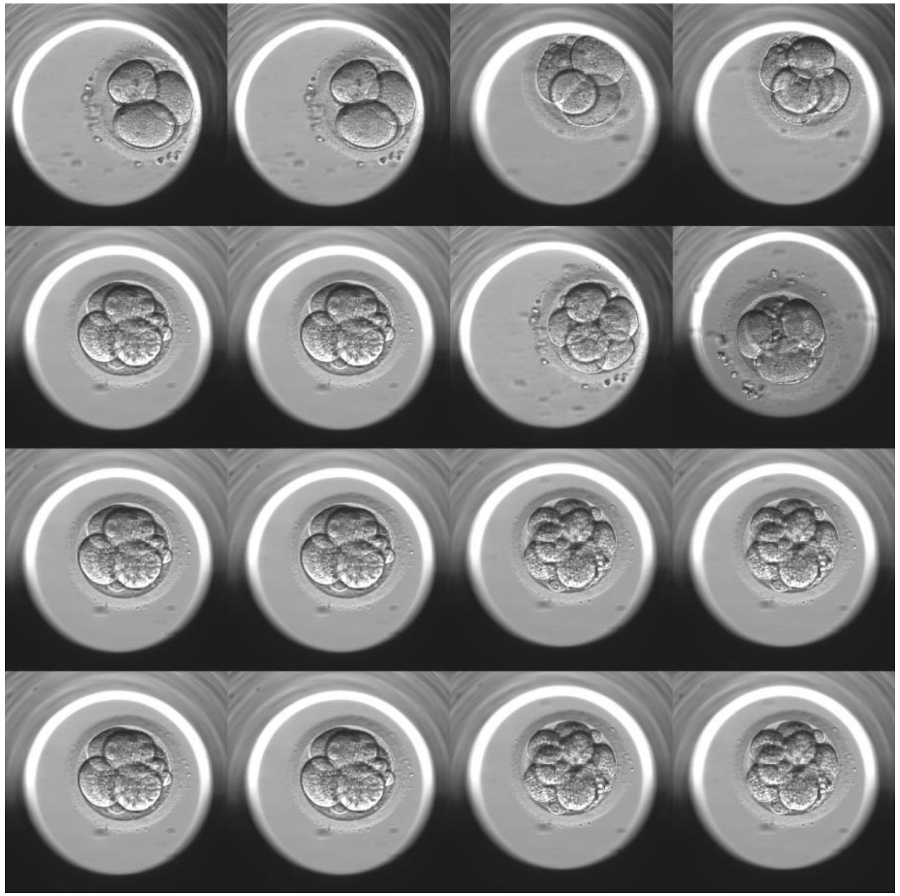
Another example is deep learning applied to gastroenterology for analysing endoscopic images to detect gastrointestinal anomalies, such as polyps, ulcers, and cancers, enhancing the accuracy of endoscopic diagnoses.



Applications: Disease Diagnosis and Prediction

Machine learning models are widely used for **diagnosing diseases** from medical imaging (including video), such as detecting cancerous tumours in radiology scans (e.g., breast cancer detection in mammograms) or identifying diabetic retinopathy in eye images.

Another example is machine learning applied to analyse time-lapse imaging data of embryo development, aiming to predict aneuploidy and select euploid embryos without the need for invasive biopsy procedures.

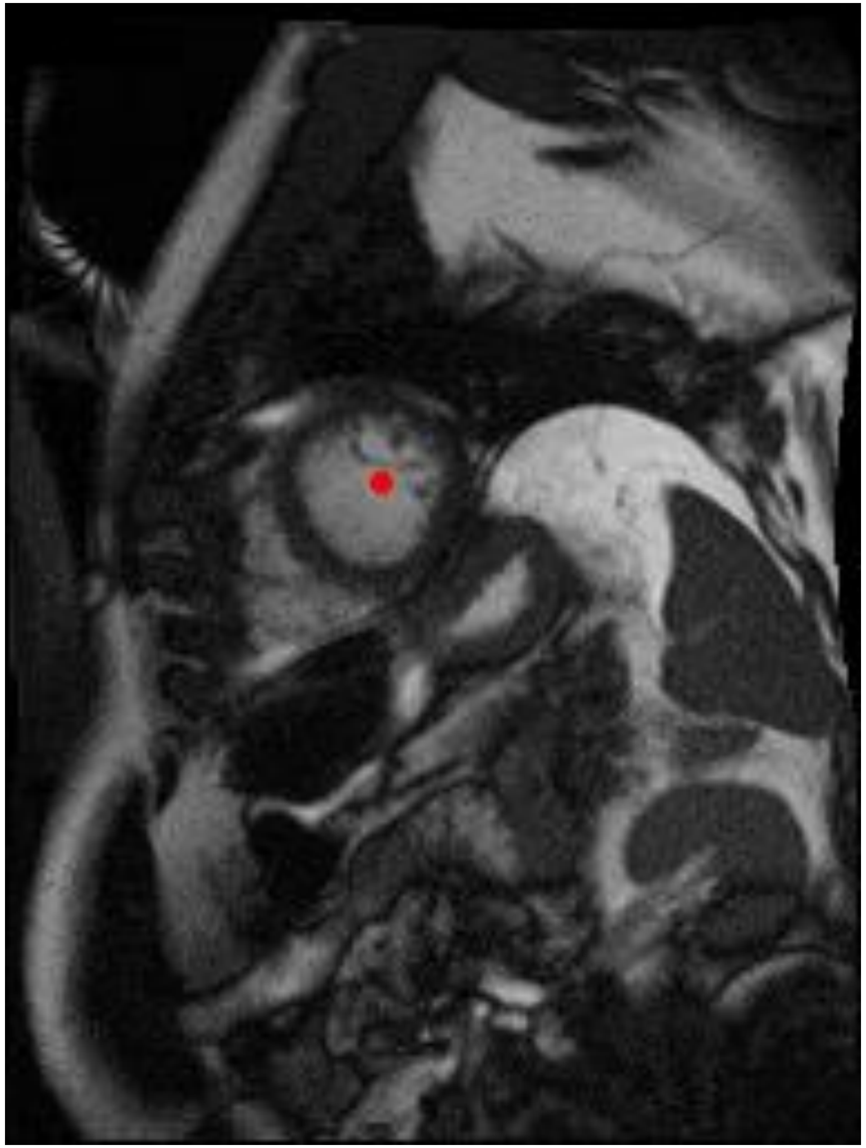


Source: Focus on Reproduction

Applications: Disease Diagnosis and Prediction

Machine learning models assist in the **segmentation** of medical images.

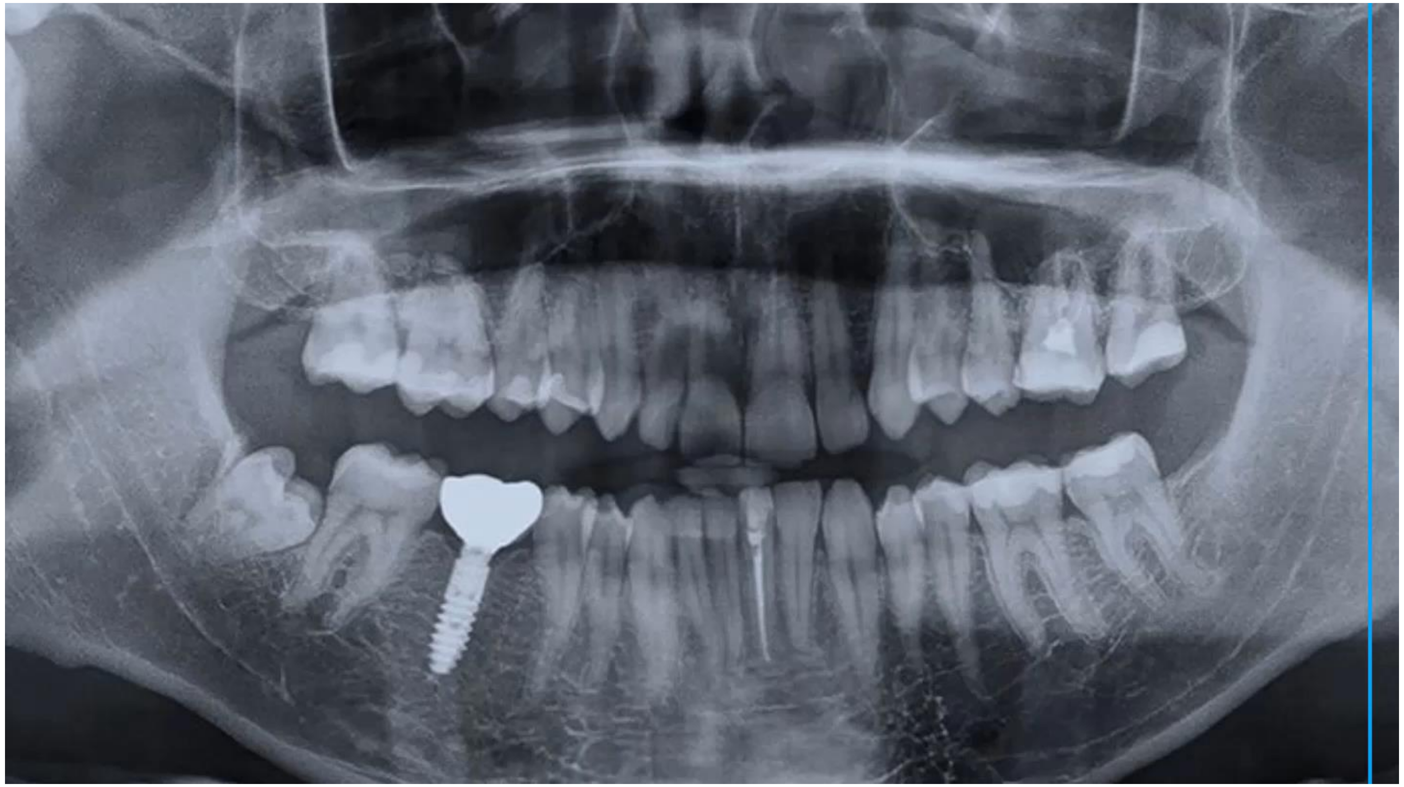
This facilitates the **measurement of volumes** (e.g., tumors, organs), which is crucial for planning treatments and monitoring disease progression.



Source: kaggle.com/second-annual-data-science-bowl (2015)

Applications: Disease Diagnosis and Prediction

Machine learning models assist in the **segmentation** of medical images.

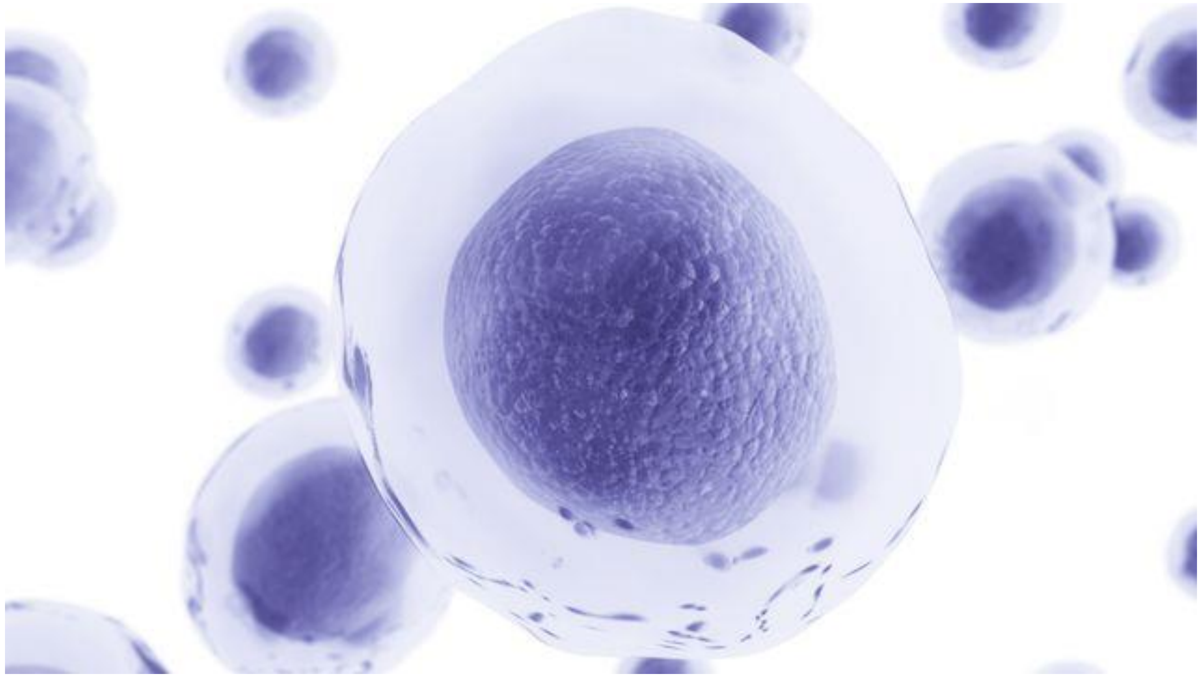


Applications: Personalized Medicine

Machine learning plays a key role in personalized medicine, where models are used to **predict patient response** to different treatments.

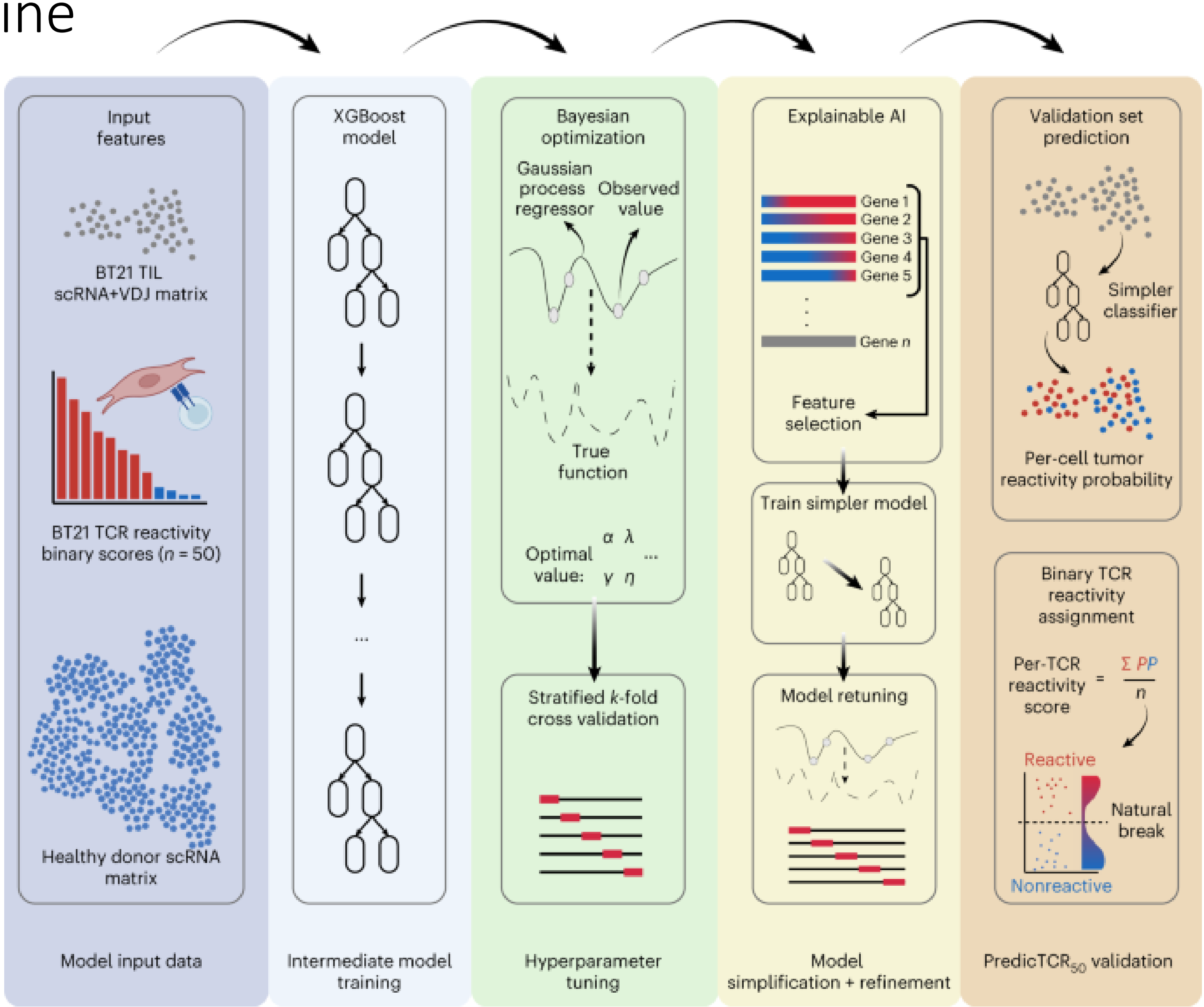
This approach can **optimize treatment plans** based on the individual's genetic makeup, improving effectiveness and reducing side effects.

Examples are:
- predicting drug responses
- cancer treatment personalization
- personalized treatment plans for chronic diseases



Applications: Personalized Medicine

Machine learning quickly and accurately identifies key T-cell receptors for personalized cancer treatment, improving therapy development.



Source: Tan, C.L. *et al.* Prediction of tumor-reactive T cell receptors from scRNA-seq data for personalized T cell therapy. *Nat Biotechnol* (2024)

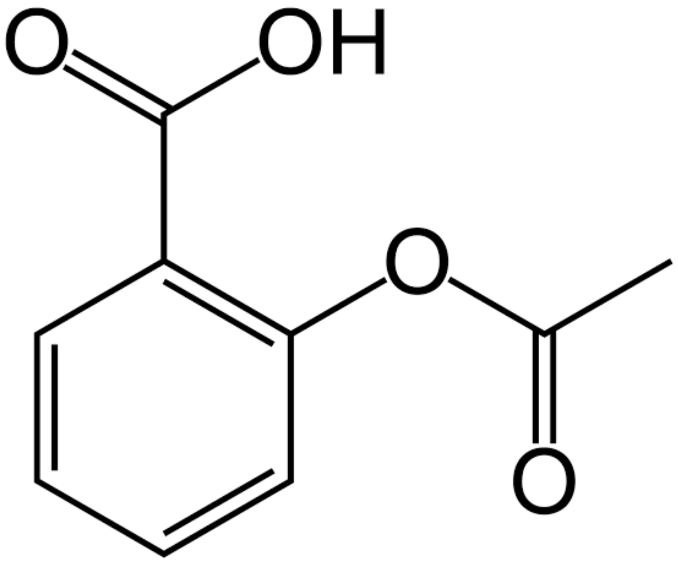
Applications: Drug Discovery and Development

Machine learning algorithms can **analyse vast chemical and biological data** to identify potential drug candidates.

This process **speeds up** the discovery of new drugs and can predict their efficacy and safety profile, significantly reducing the time and cost of drug development.

Applications: Drug Discovery and Development

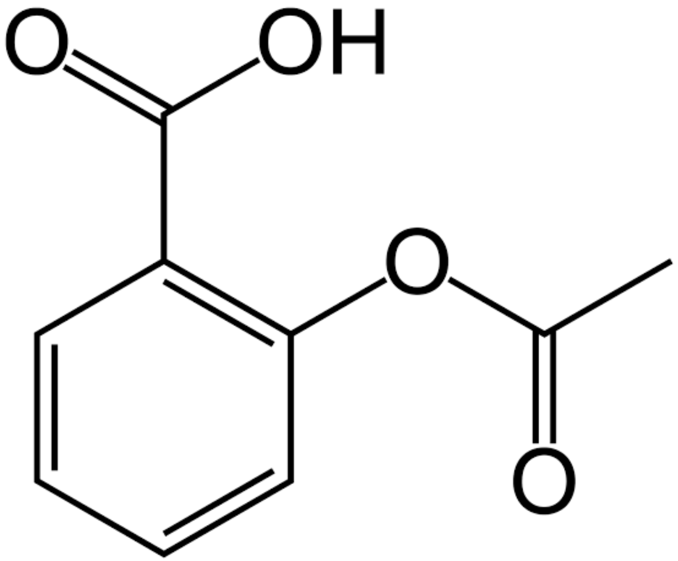
Machine learning algorithms can **analyse vast chemical and biological data** to identify potential drug candidates.



Source: Aspirin

Applications: Drug Discovery and Development

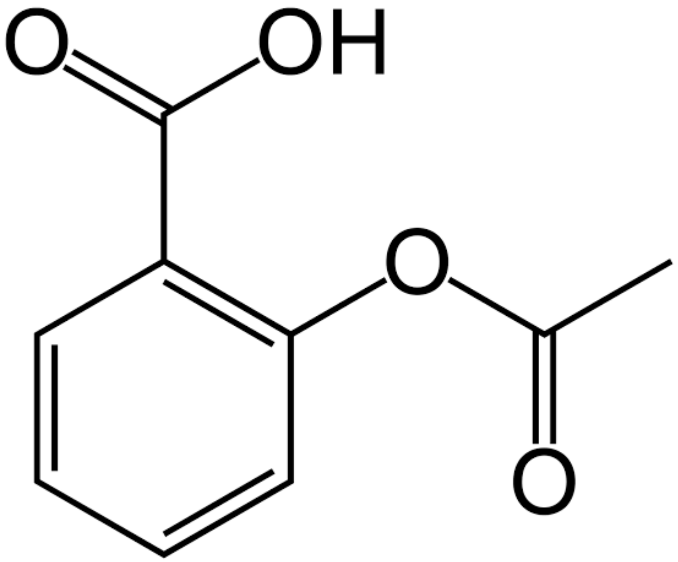
Machine learning algorithms can **analyse vast chemical and biological data** to identify potential drug candidates.



Source: Aspirin

Applications: Drug Discovery and Development

Machine learning algorithms can **analyse vast chemical and biological data** to identify potential drug candidates.

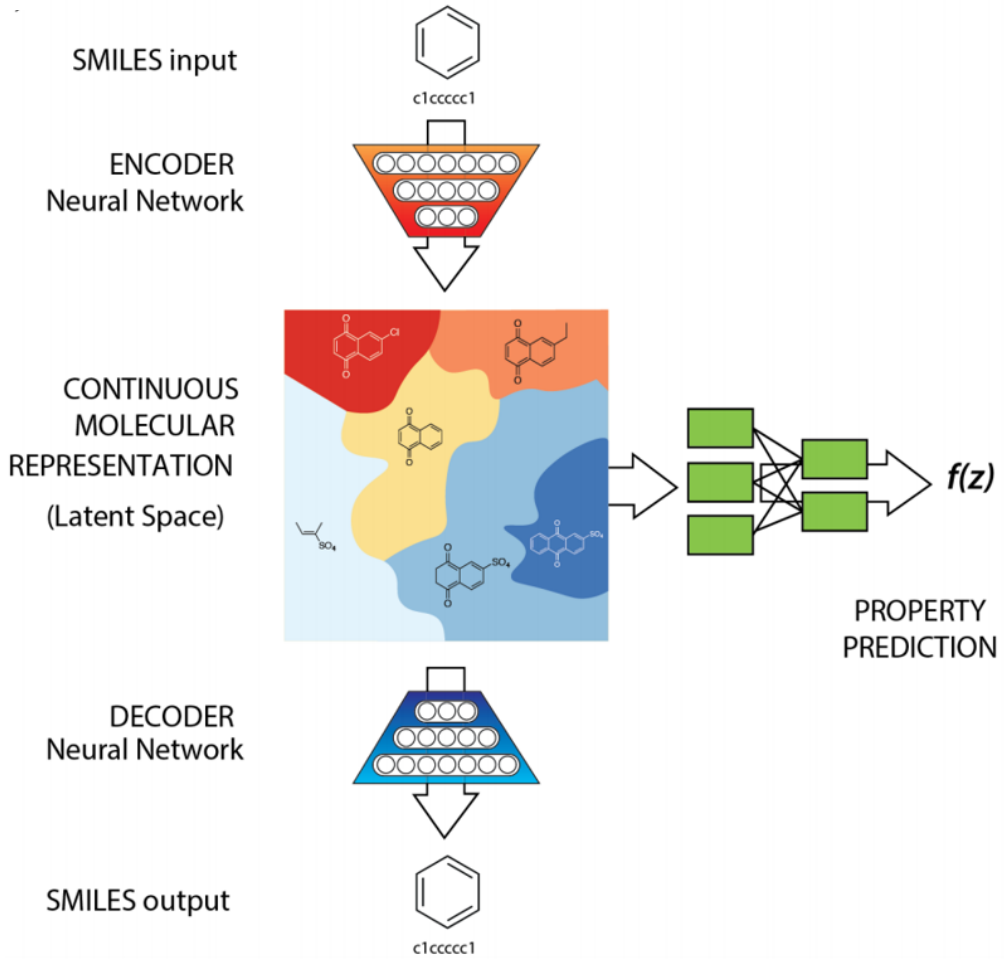


CC(=O)OC1=CC=CC=C1C(=O)O

Source: Aspirin

Applications: Drug Discovery and Development

Machine learning algorithms can **analyse vast chemical and biological data** to identify potential drug candidates.

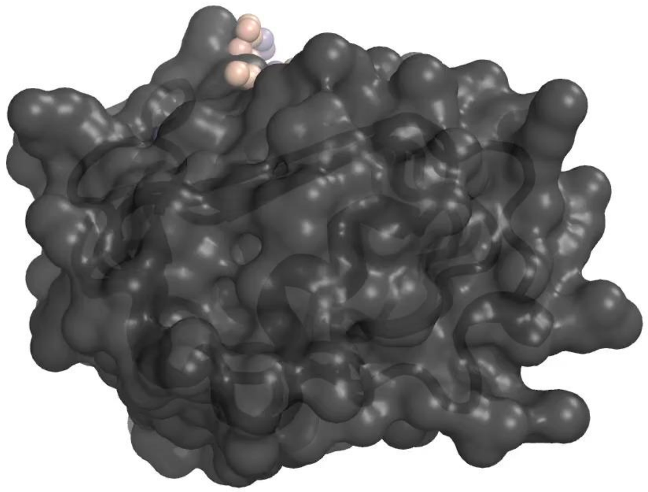


Source: Gómez-Bombarelli et al. ACS Central Science (2019)

Applications: Drug Discovery and Development

Machine learning algorithms can **analyse vast chemical and biological data** to identify potential drug candidates.

Here a diffusion model generates a novel protein that binds to the insulin receptor.



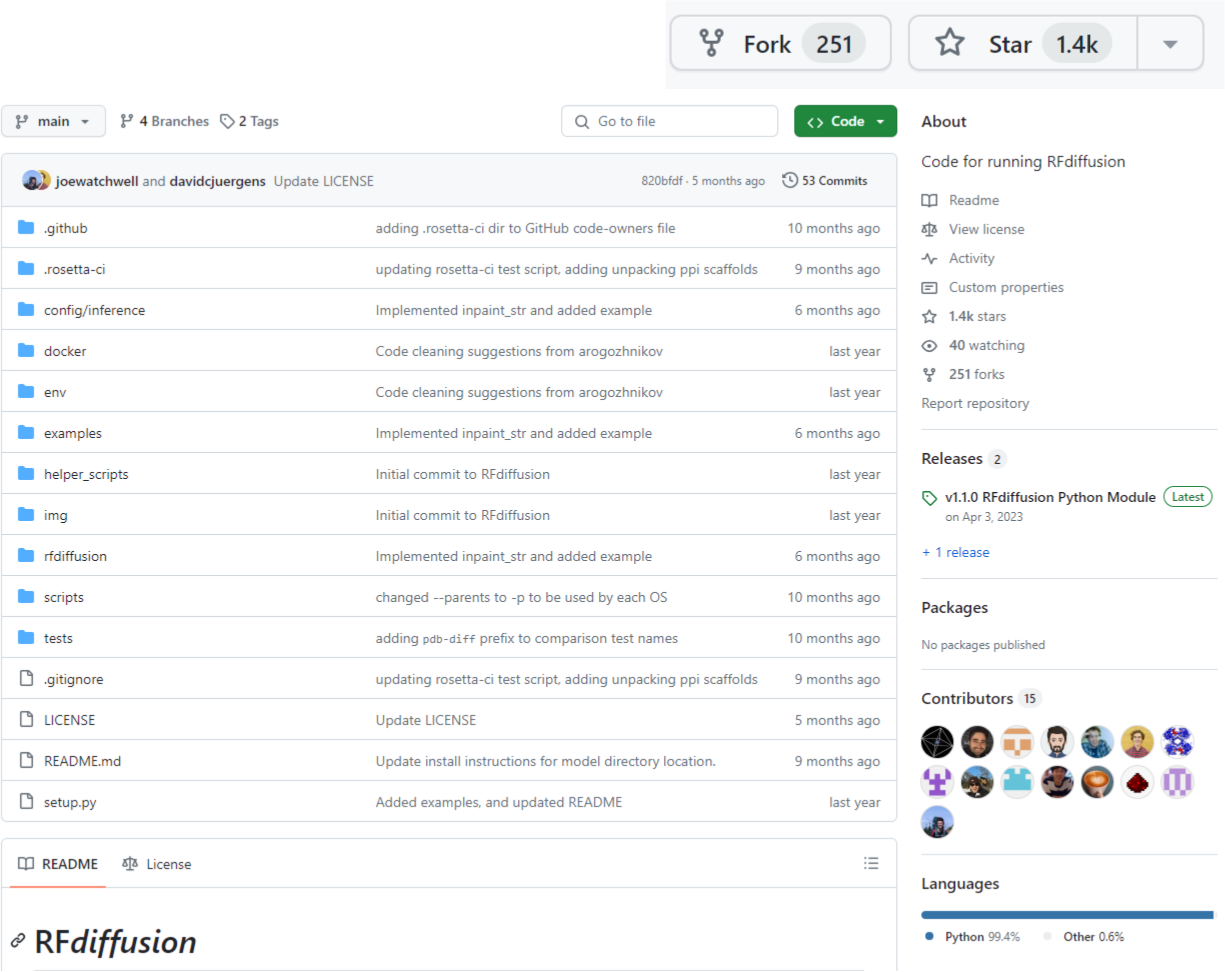
Source: https://www.bakerlab.org/2023/03/30/rf-diffusion-now-free-and-open-source/

Applications: Drug Discovery and Development

Machine learning algorithms can **analyse vast chemical and biological data** to identify potential drug candidates.

Here a diffusion model generates a novel protein that binds to the insulin receptor.

https://github.com/RosettaCommons/RFdiffusion



open source initiative®

Source: https://www.bakerlab.org/2023/03/30/rf-diffusion-now-free-and-open-source/

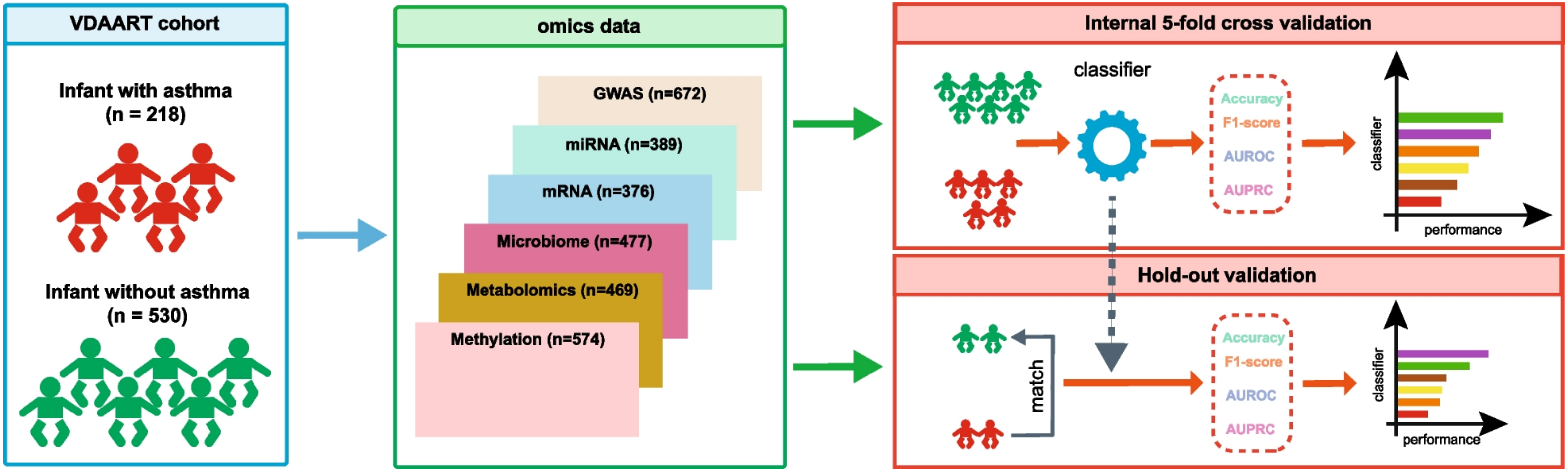
Applications: Genomics, Transcriptomics, Proteomics, … , -omics

Machine learning models are important for the analysis of omics data generated by Genomics, Proteomics, Transcriptomics, Metabolomics, Epigenomics, Pharmacogenomics, Microbiomics, Lipidomics and Glycomics to **understand** variations and their **association** with diseases.

They can **classify patients** based on biomedical (bio)markers, helping in the identification of molecular diseases and the development of therapies.

Applications: Genomics, Transcriptomics, Proteomics, … , -omics

They can **classify patients** based on biomedical (bio)markers, helping in the identification of molecular diseases and the development of therapies.



Source: Wang, XW. *et al.* Benchmarking omics-based prediction of asthma development in children. *Respir Res*

Applications: Wearable Health Monitoring

Machine learning algorithms analyse data from wearable devices to monitor health indicators in real-time, **predict health events**, and provide personalized health advice.

This is key for chronic disease management and **preventive healthcare**.

Applications: Mental Health Analysis

Machine learning models **analyse patterns** in speech, typing speed, and social media usage to detect signs of mental health issues, such as depression or anxiety.

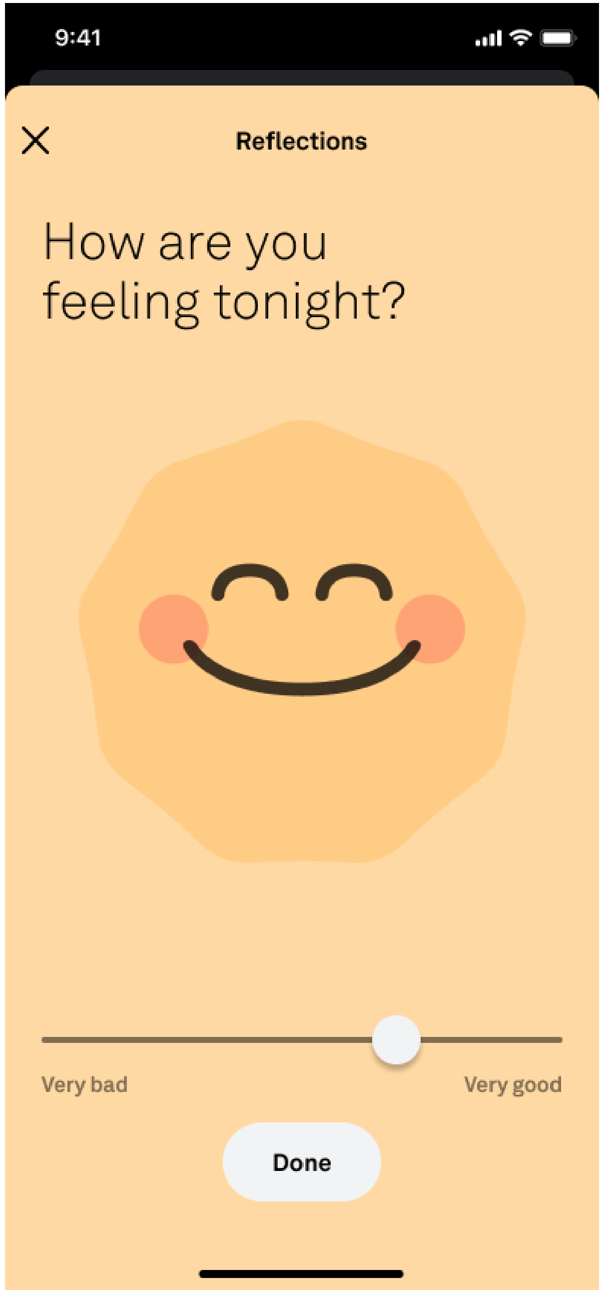
This aids in **early detection** and intervention.

PLoS One. 2022; 17(7): e0272330. PMCID: PMC9337649
Published online 2022 Jul 29. doi: 10.1371/journal.pone.0272330 PMID: 35905087

Machine learning-based predictive modeling of depression in hypertensive populations

Chiyoung Lee, Conceptualization, Data curation, Formal analysis, Writing – original draft, Writing – review & editing[1,*] and Heewon Kim, Conceptualization, Methodology, Supervision, Writing – review & editing[2]

Features in Table 2

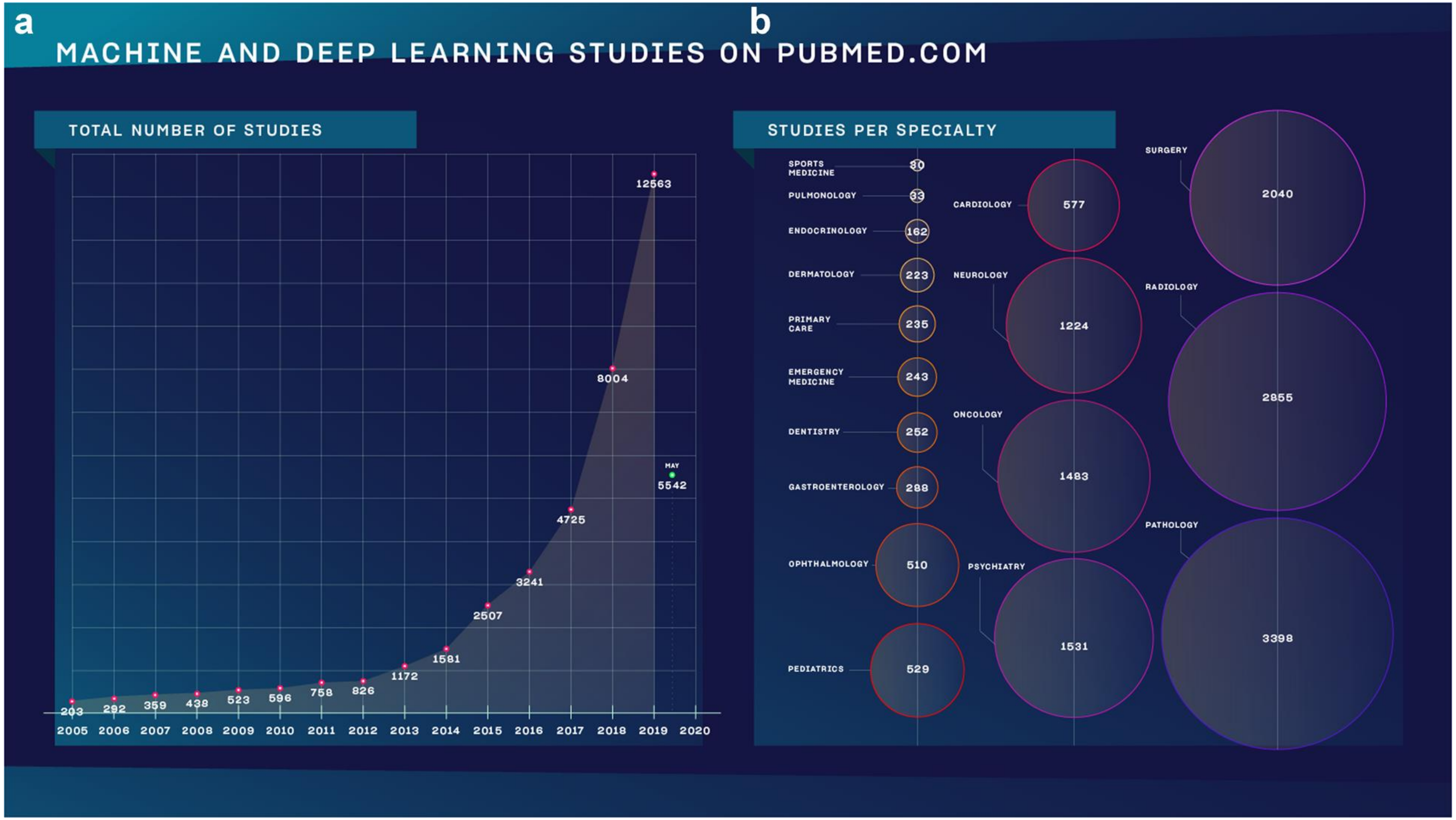


Applications: Healthcare Operations

Machine learning models are used to optimize hospital operations, such as

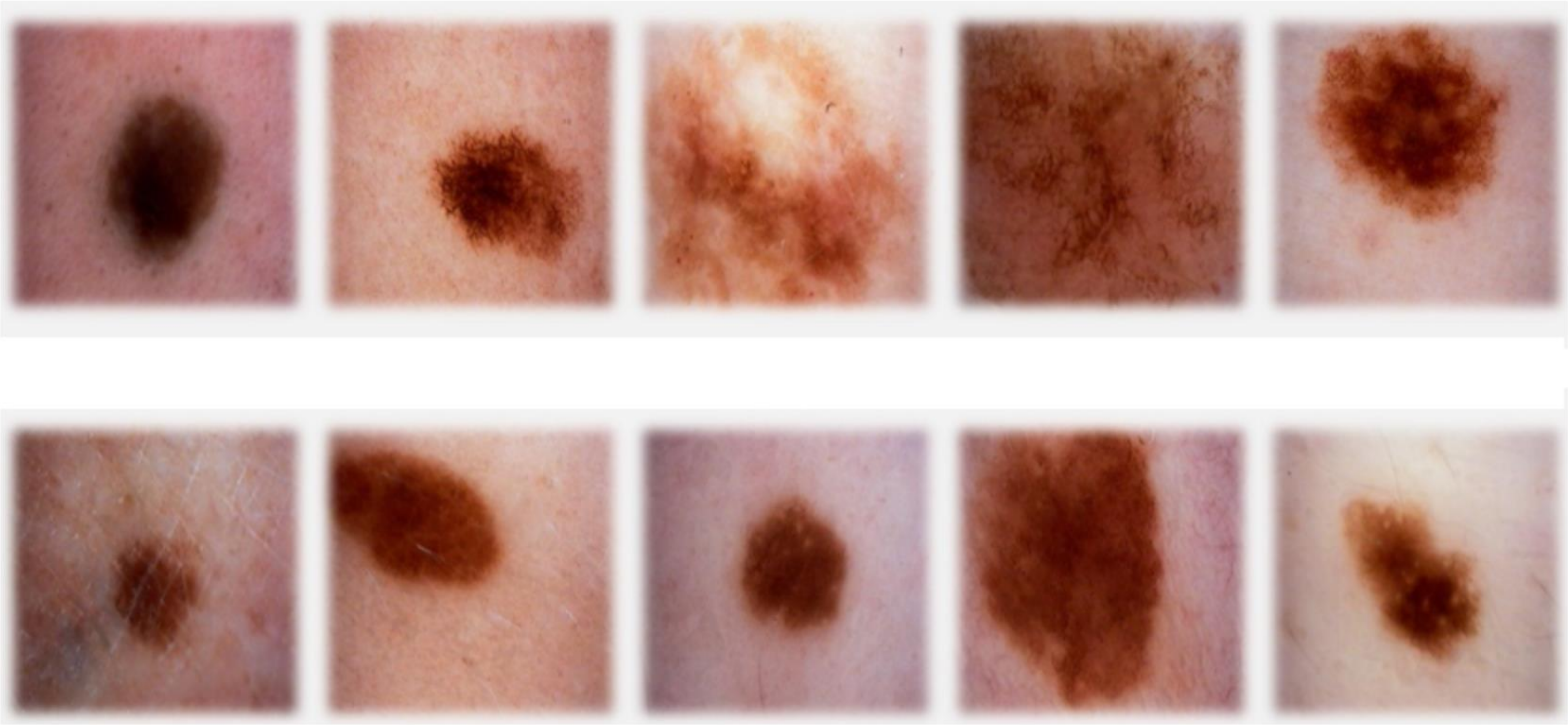
- predicting patient admission rates to manage hospital beds
- scheduling surgeries and medical staff
- reducing wait times for patients
- …

Why follow this course?



Meskó, B., Görög, M. A short guide for medical professionals in the era of artificial intelligence. npj Digit. Med. 3, 126 (2020)

classification

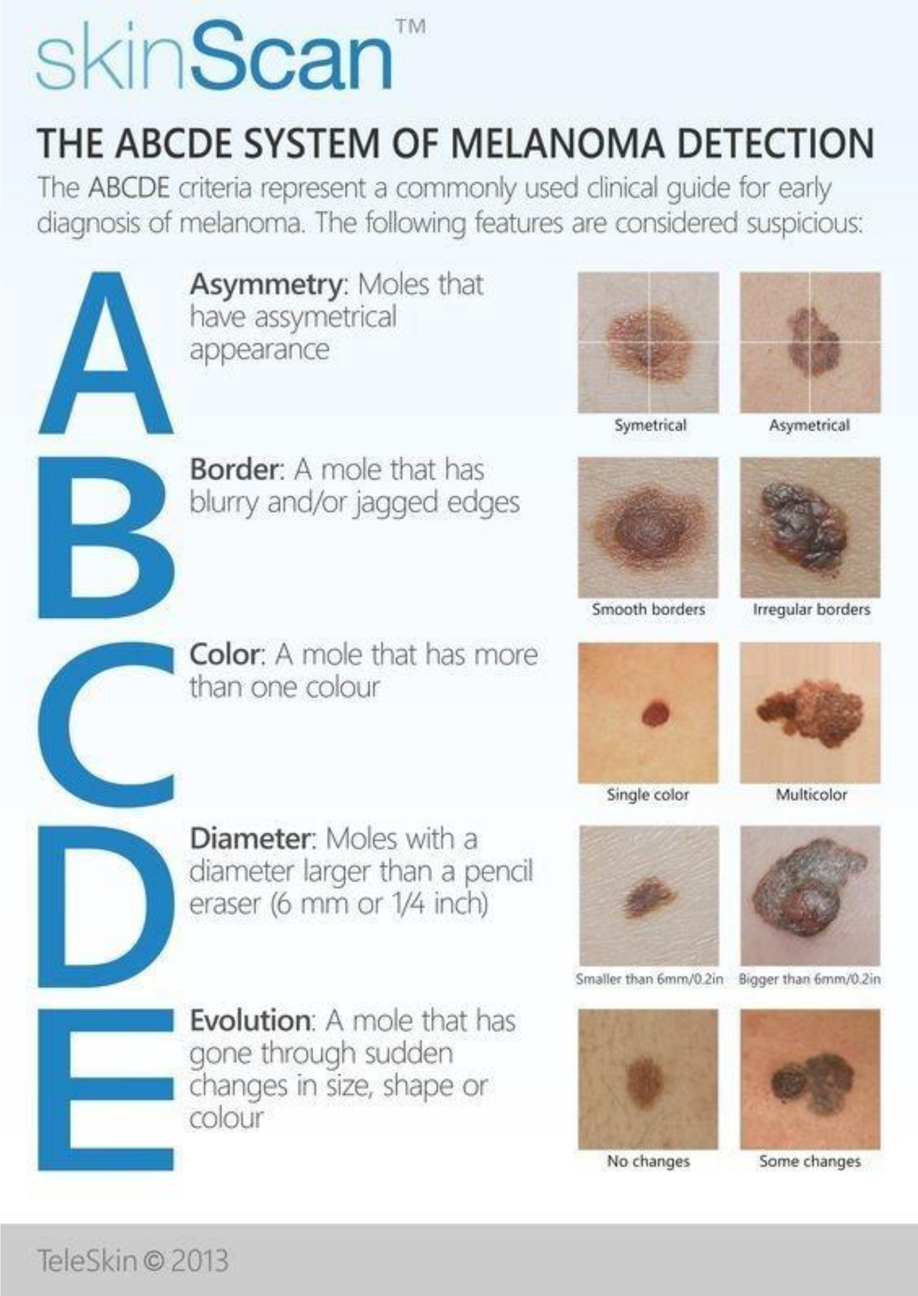


sign of cancer

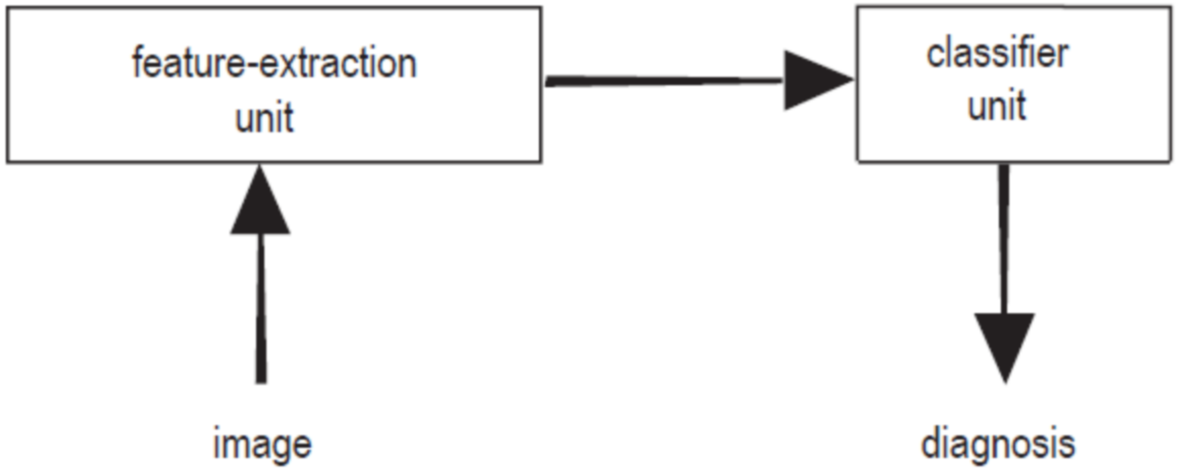
top row malignant

bottom row benign

classification



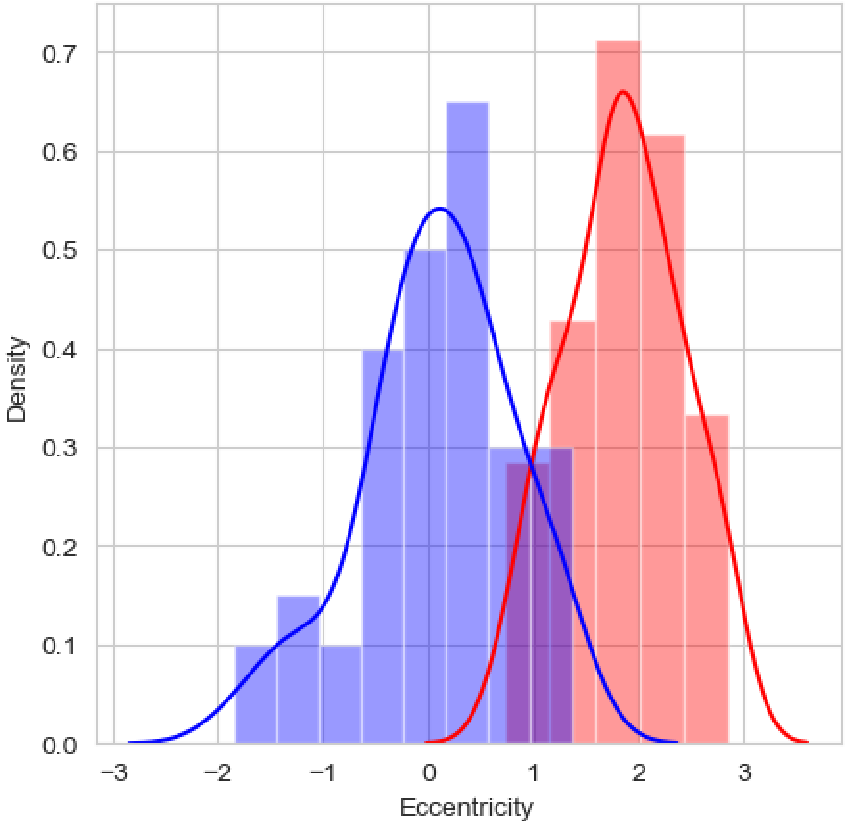
classification: terminology



feature extraction: features (a.k.a. properties or attributes)

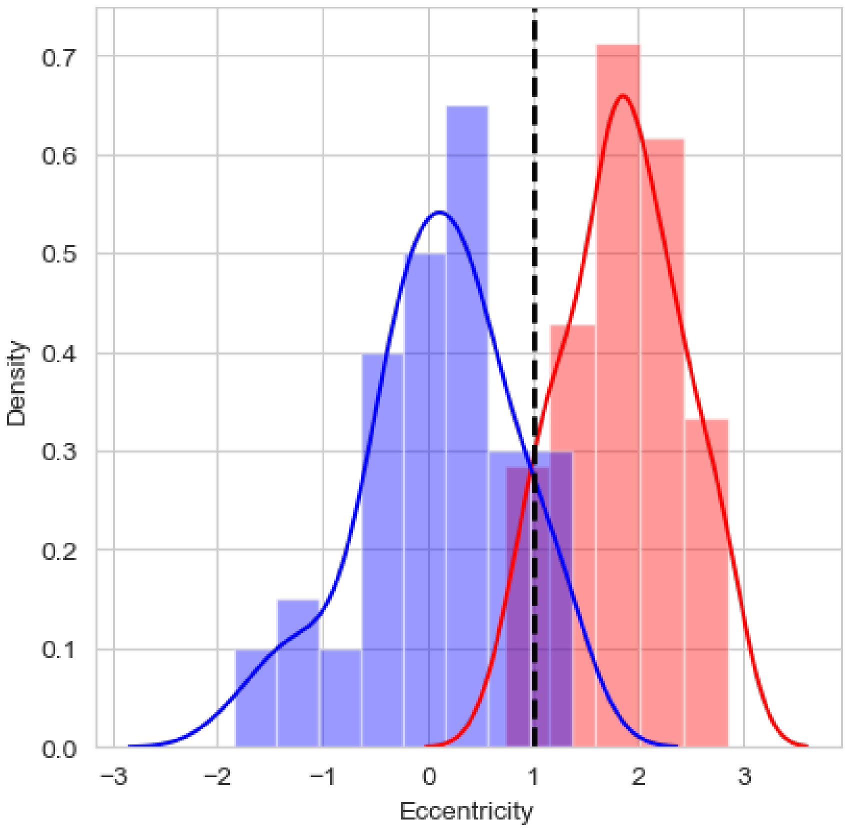
data set, sample (a.k.a. example, instance or data point), label (a.k.a. target)

classification: a feature



feature: eccentricity of lesion (how nearly circular the lesion is)

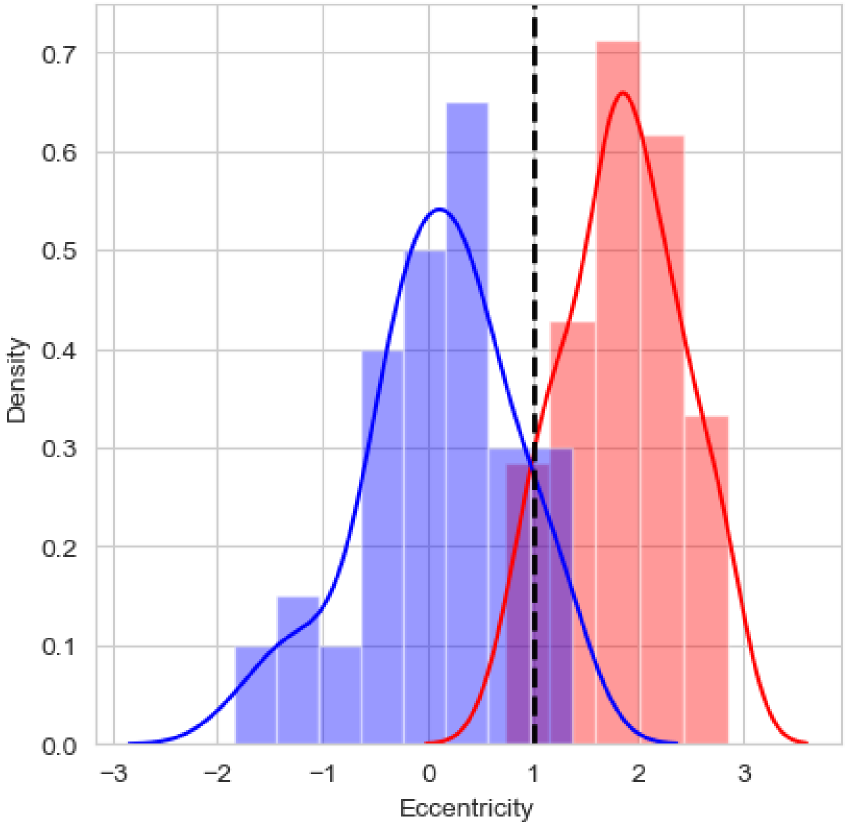
classification: the model

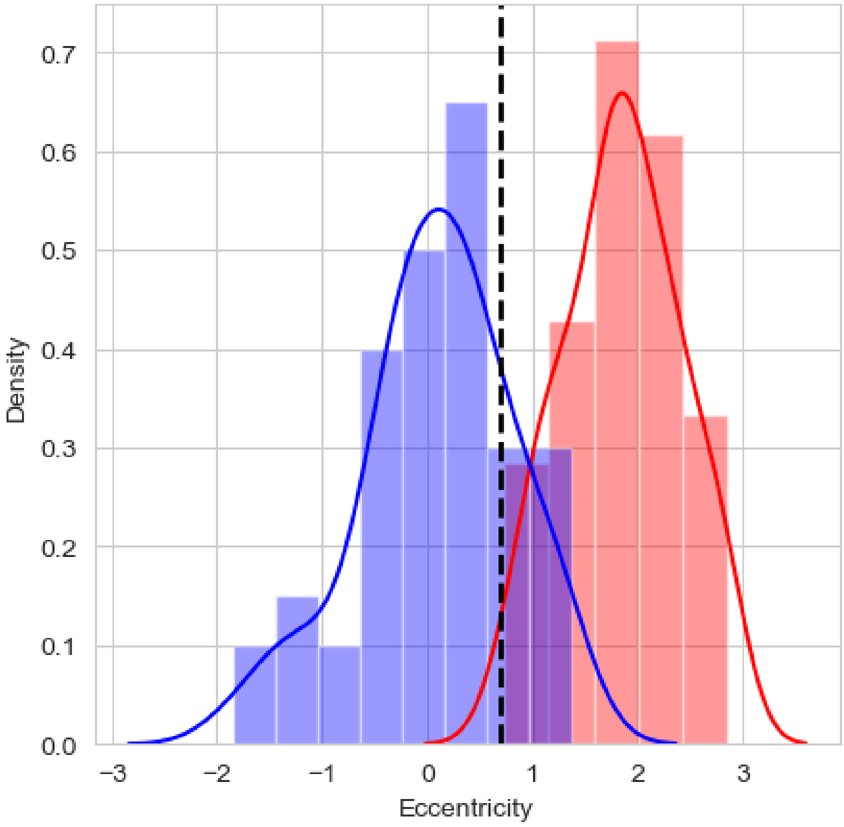


feature: eccentricity of lesion (how nearly circular the lesion is)

model: threshold ***t***

classification: the model





feature: eccentricity of lesion (how nearly circular the lesion is)

model: threshold ***t:*** consequence of the predictions

classification: prediction errors

malignant: **positive** class
benign: **negative** class

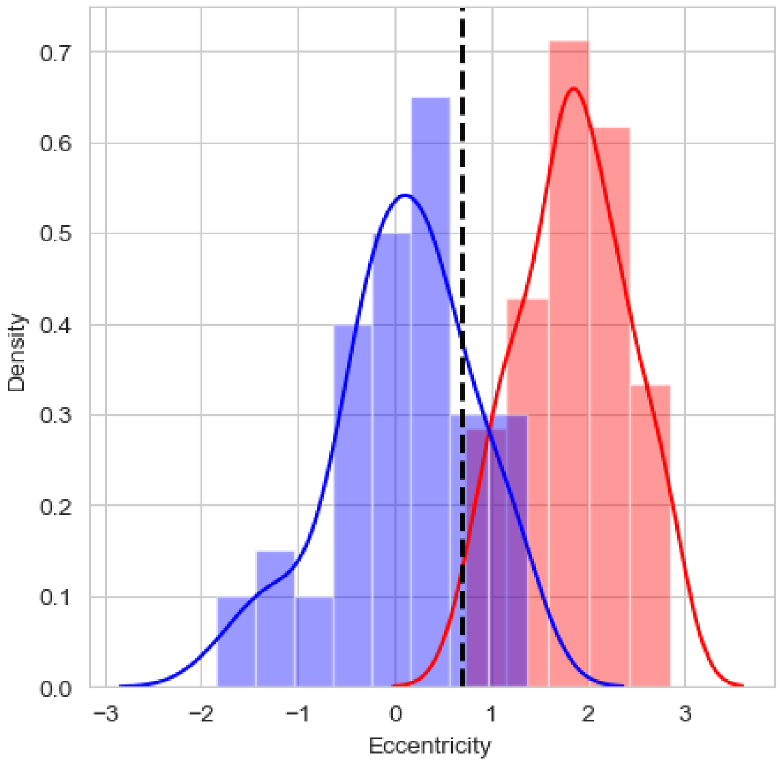
count the number of malignant images with eccentricity value $\geq t$: **true positive** predictions (TP)

count the number of malignant images with eccentricity value $< t$: **false negative** predictions (FN)

count the number of benign images with eccentricity value $\geq t$: **false positive** predictions (FP)

count the number of benign images with eccentricity value $< t$: **true negative** predictions (TN)

Really is \ Classified as	Positive	Negative
Positive	True Positive	False Negative
Negative	False Positive	True Negative



classification: prediction errors

	Classified as: Positive	Classified as: Negative	Really is
	True Positive	False Negative	Positive
	False Positive	True Negative	Negative

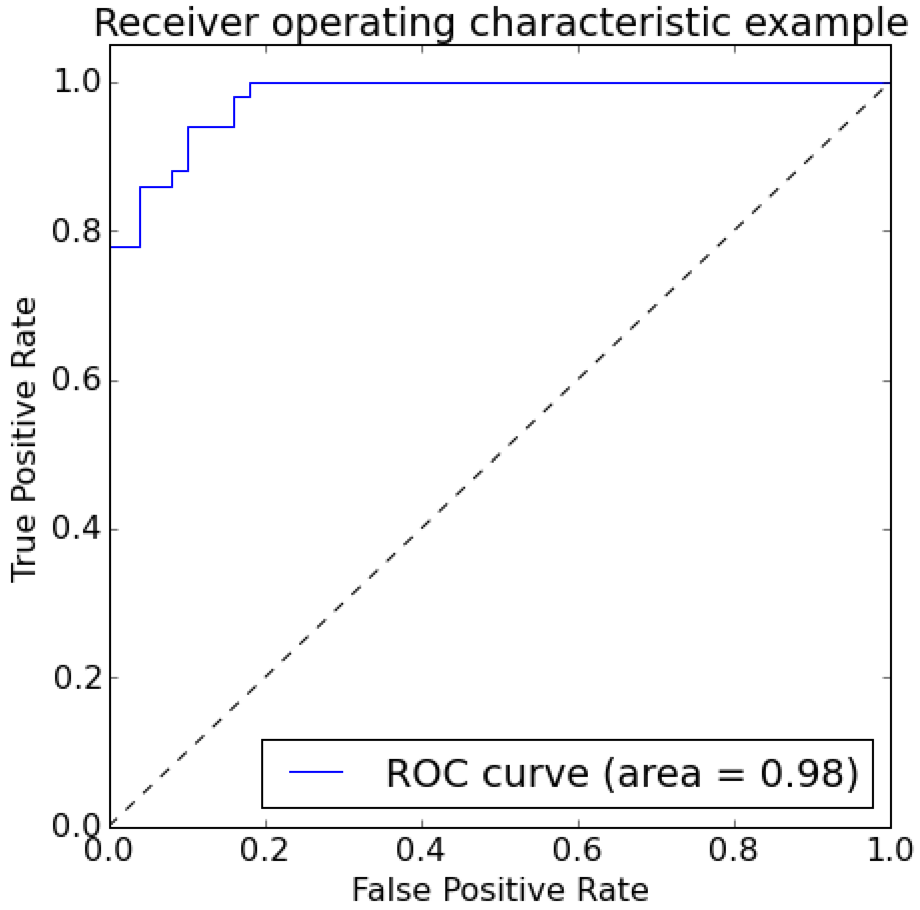
$$\text{accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

$$\text{TPR} = \frac{TP}{TP + FN}$$

$$\text{FPR} = \frac{FP}{FP + TN}$$

classification: prediction errors

$$\mathrm{TPR} = \frac{TP}{TP+FN}$$



$$\mathrm{FPR} = \frac{FP}{FP+TN}$$

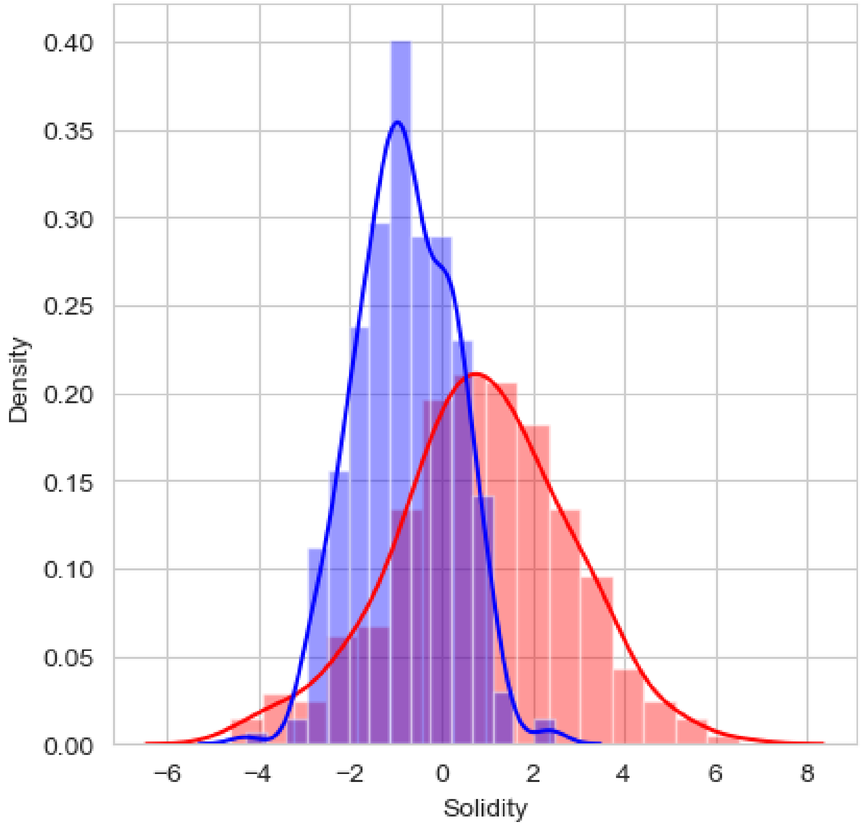
model that classifies all images as malignant:
TPR=1 and FPR=1

model that classifies all images a benign:
TPR=0 and FPR=0

vary threshold *t*

Area Under the Curve (AUC)

classification: multi-dimensional

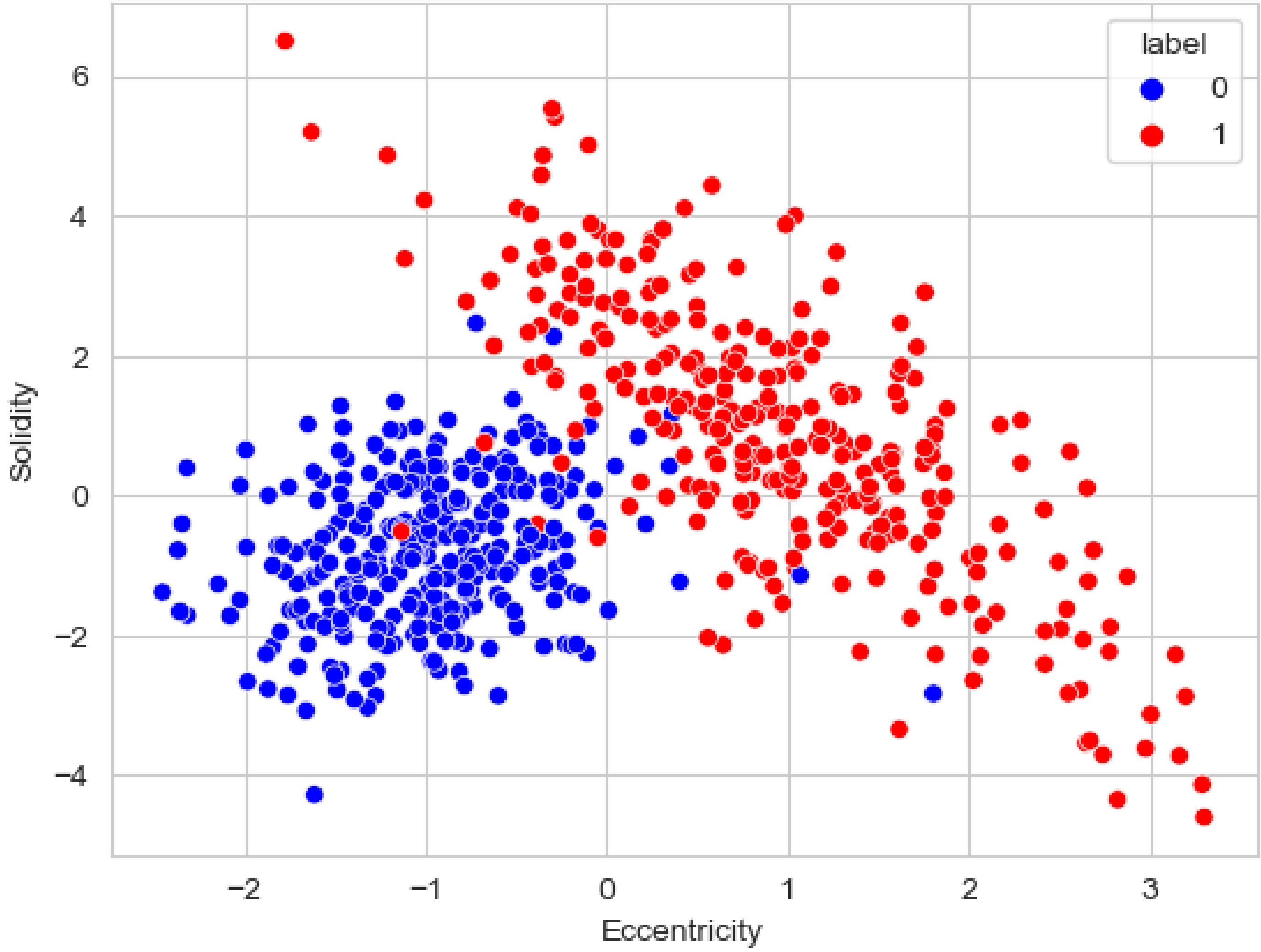


add another feature?

feature vector X

Euclidean vector space

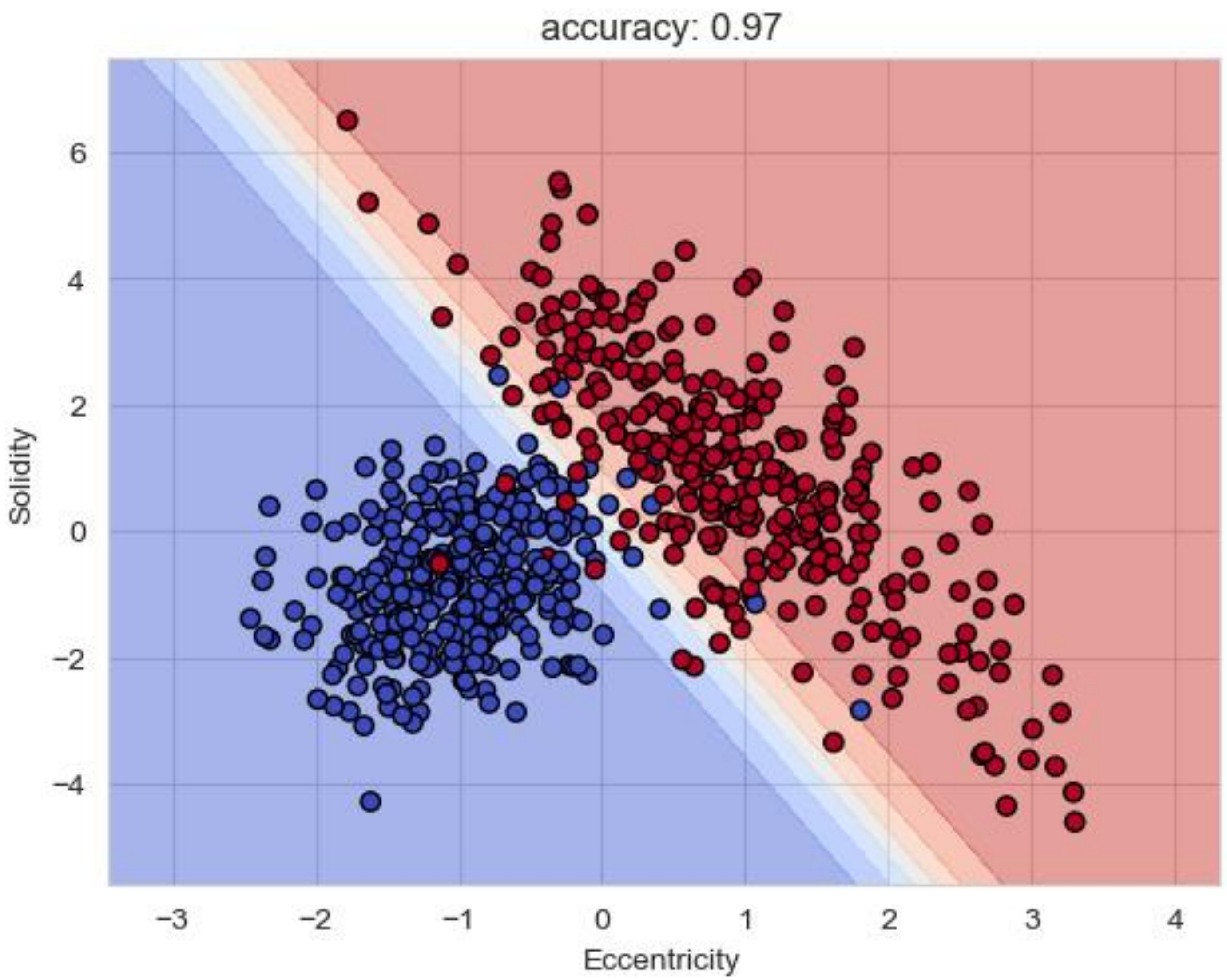
classification: multi-dimensional



feature vector X

Euclidean vector space

classification: multi-dimensional



linear decision boundary

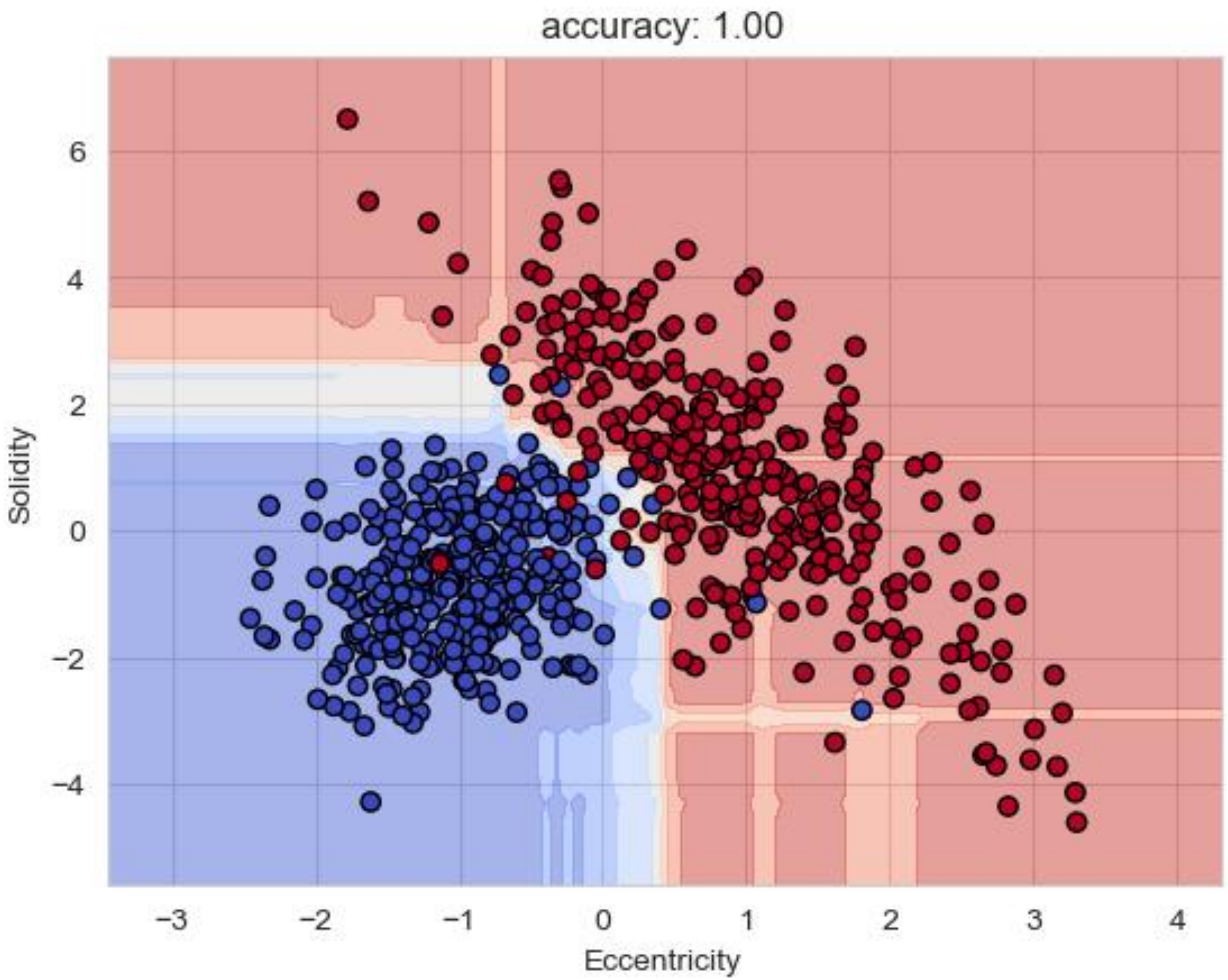
blue region malignant, red region benign

yet more features

can't look at the decision boundary

more complex

classification: model complexity

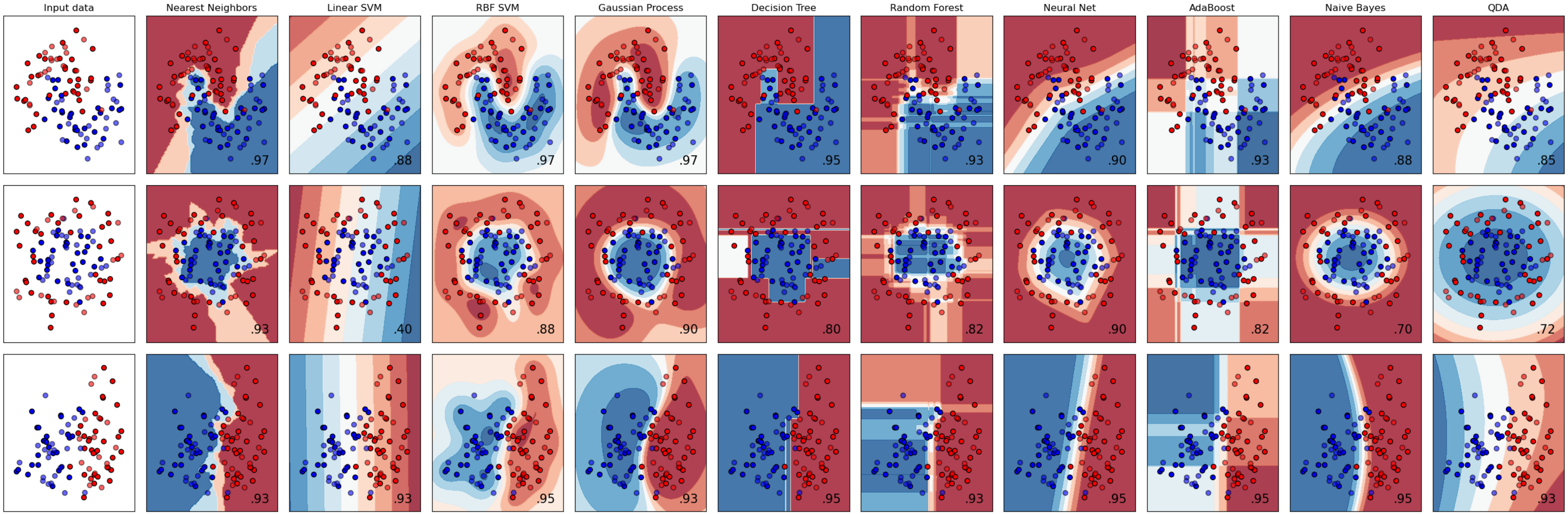


unseen external images

generalization

overfitting

scikit-learn



data normalization

make all features same scale

Eccentricity [0,100], Solidity [-5,7]

weights all features equally in their representation

standardization

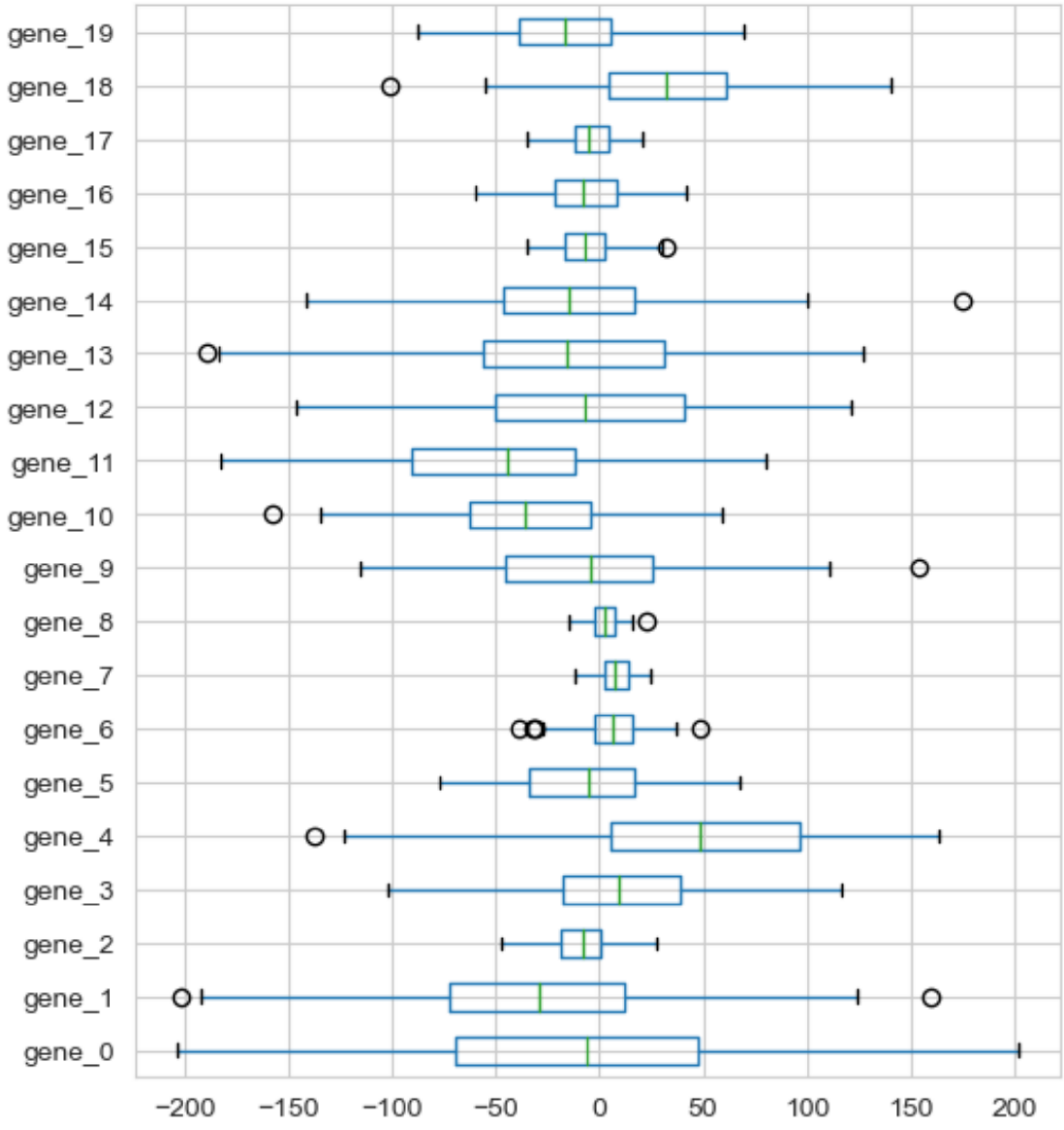
$$\mu = 0 \quad \sigma = 1$$

min-max scaling: scale the features to a fixed range

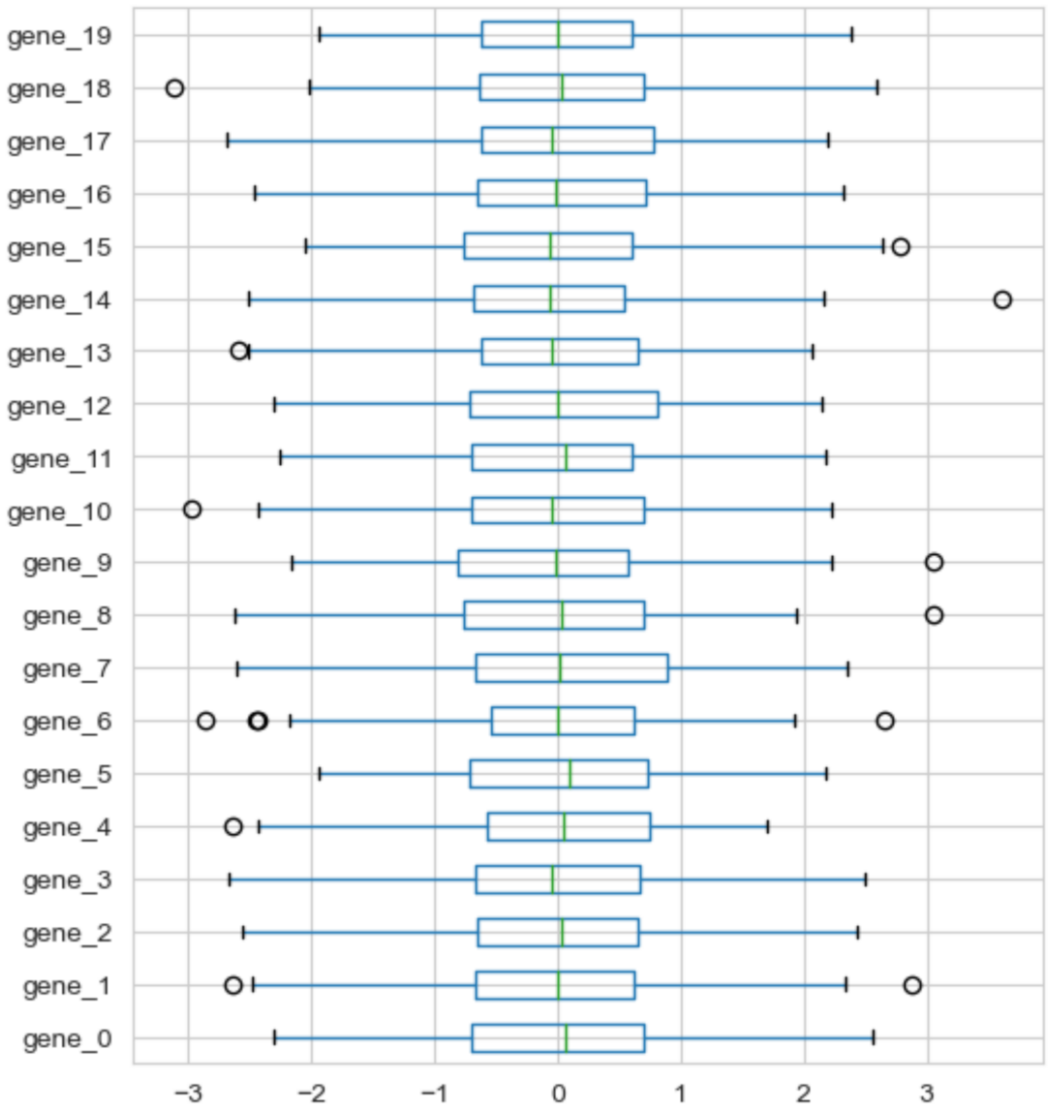
$$x_{norm} = \frac{x - \mu}{\sigma}$$

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

data normalization

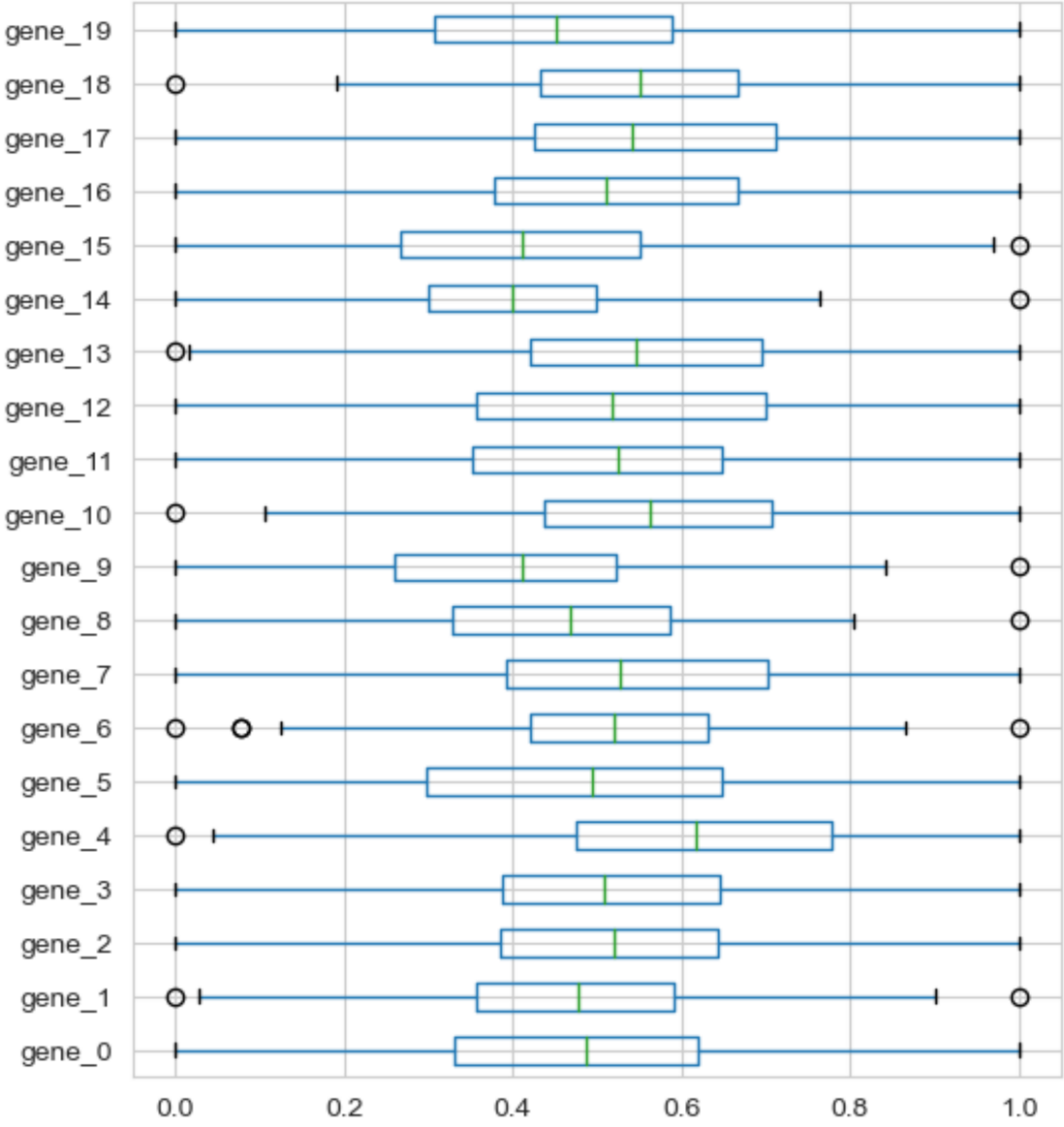


data normalization: standardization



$$x_{norm} = \frac{x - \mu}{\sigma}$$

data normalization: min-max scaling



$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}}$$