Introduction to Deep Learning

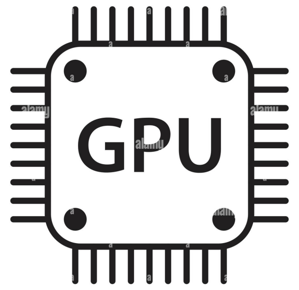
Outset AI for Health Session

Franny Dean, July 8, 2026

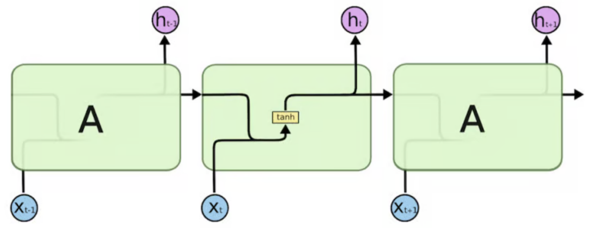
What is 'AI'?

Rules based AI

Machine Learning



Deep Learning

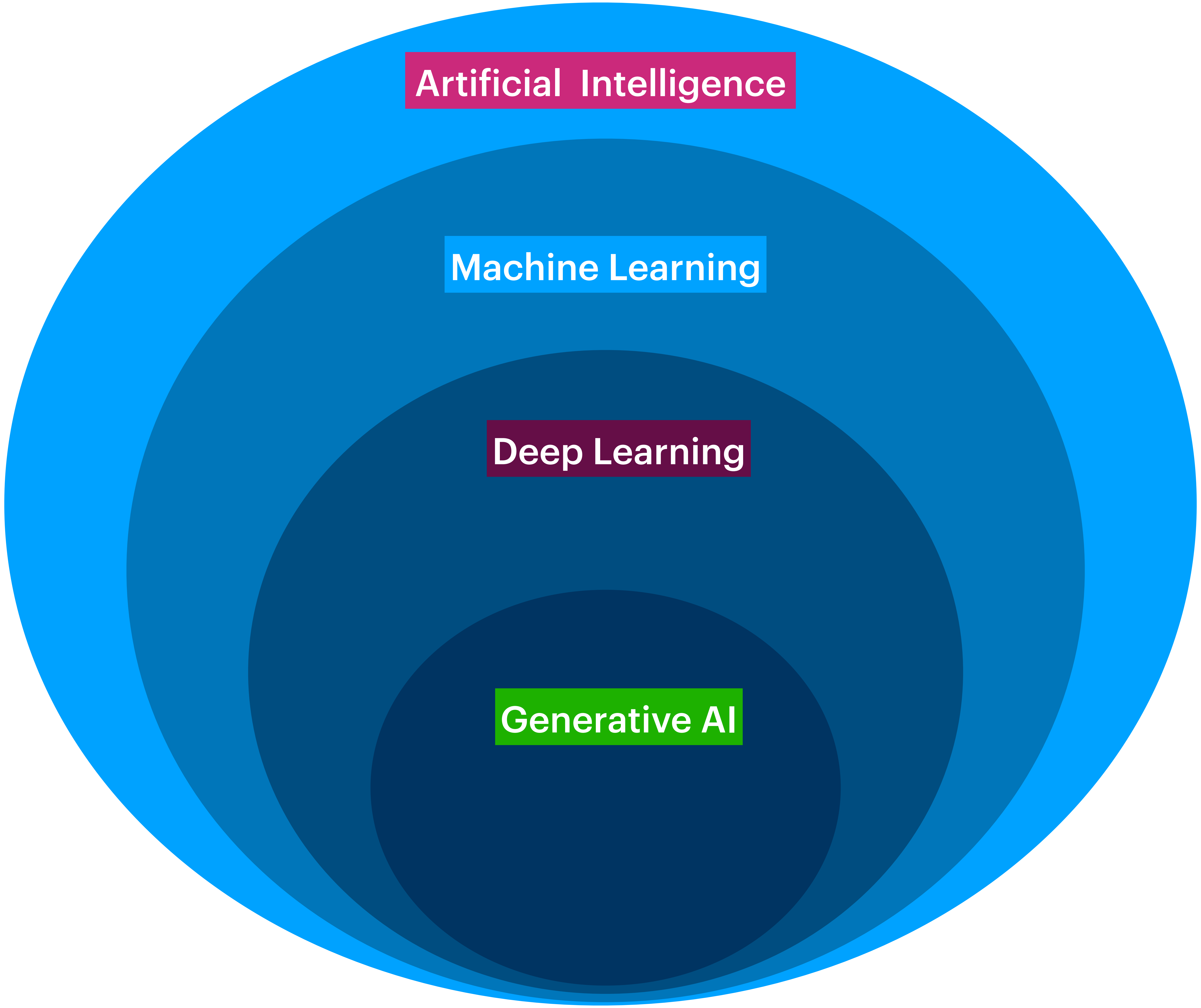


Generative AI

AI Agents

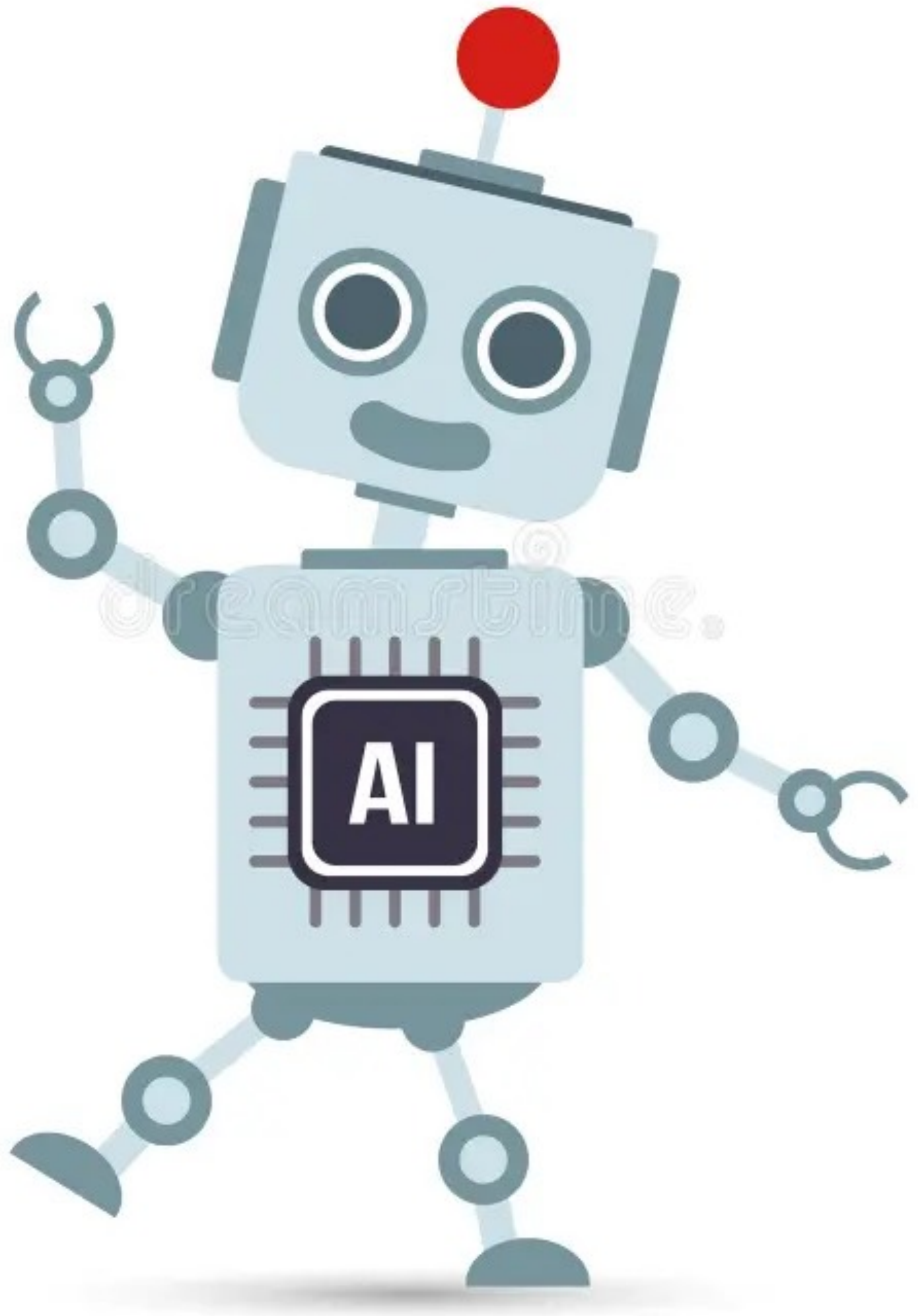
1956

2026


Artificial Intelligence
Machine Learning
Deep Learning
Generative AI

What is 'AI'?

"AI is what hasn't been done yet!" - Tesler's Theorem



Its really the broad set of ways to teach computers to do what humans do!

What is 'AI'?

Traditional rules based approach

Definition: Computers do what humans can do from programmed rules!

Task

Explicit Rules

```
King: Moves exaclty one
square in any direction.

Queen: Moves in any one
straight direction.

…

Perform a tree search for the
best next move.
```

Learning

As complexity scales, humans need to be able to explicitly define the rules.

What is 'AI'?

Machine Learning

Definition: Give a computer a bunch of examples to learn from.

Task: predict heart disease

Examples of positive and negative

Cholesterol	Age	Blood Pressure	Ejection Fraction
190mg/dL	80	130/100	45%
150mg/dL	20	110/80	70%
250mg/dL	50	150/100	35%
130mg/dL	45	90/50	85%

Learning

As complexity scales, computers learn relationships between variables.

How does Machine Learning (ML) work?

Example: Imagine we want one predictor of the outcome.

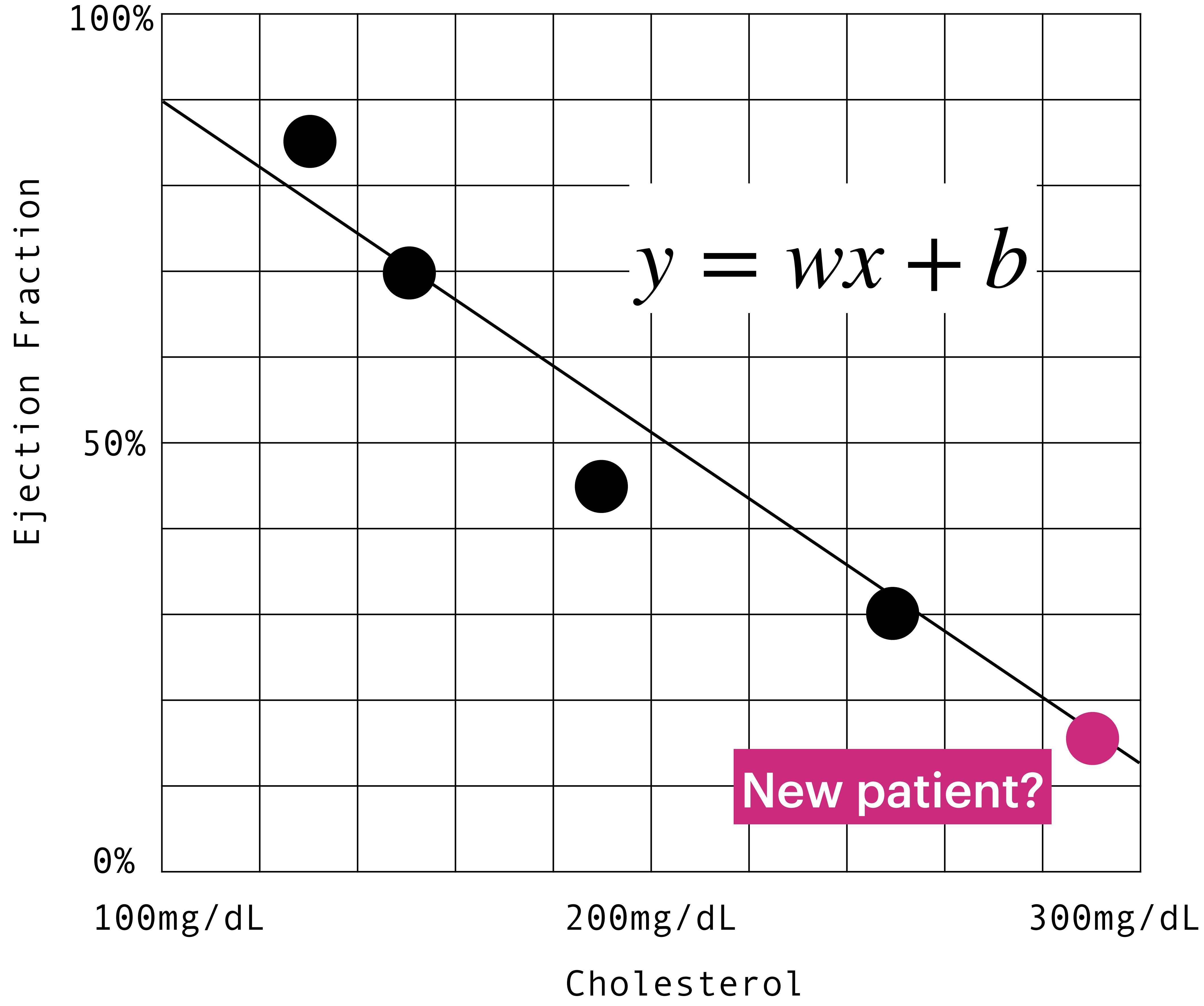
Cholesterol	Age	Blood Pressure	Ejection Fraction
190mg/dL	80	130/100	45%
150mg/dL	20	110/80	70%
250mg/dL	50	150/100	30%
130mg/dL	45	90/50	85%

Predictor

Outcome

Learn best function from the relationship between variables and outcome

i.e. what should w and b be?



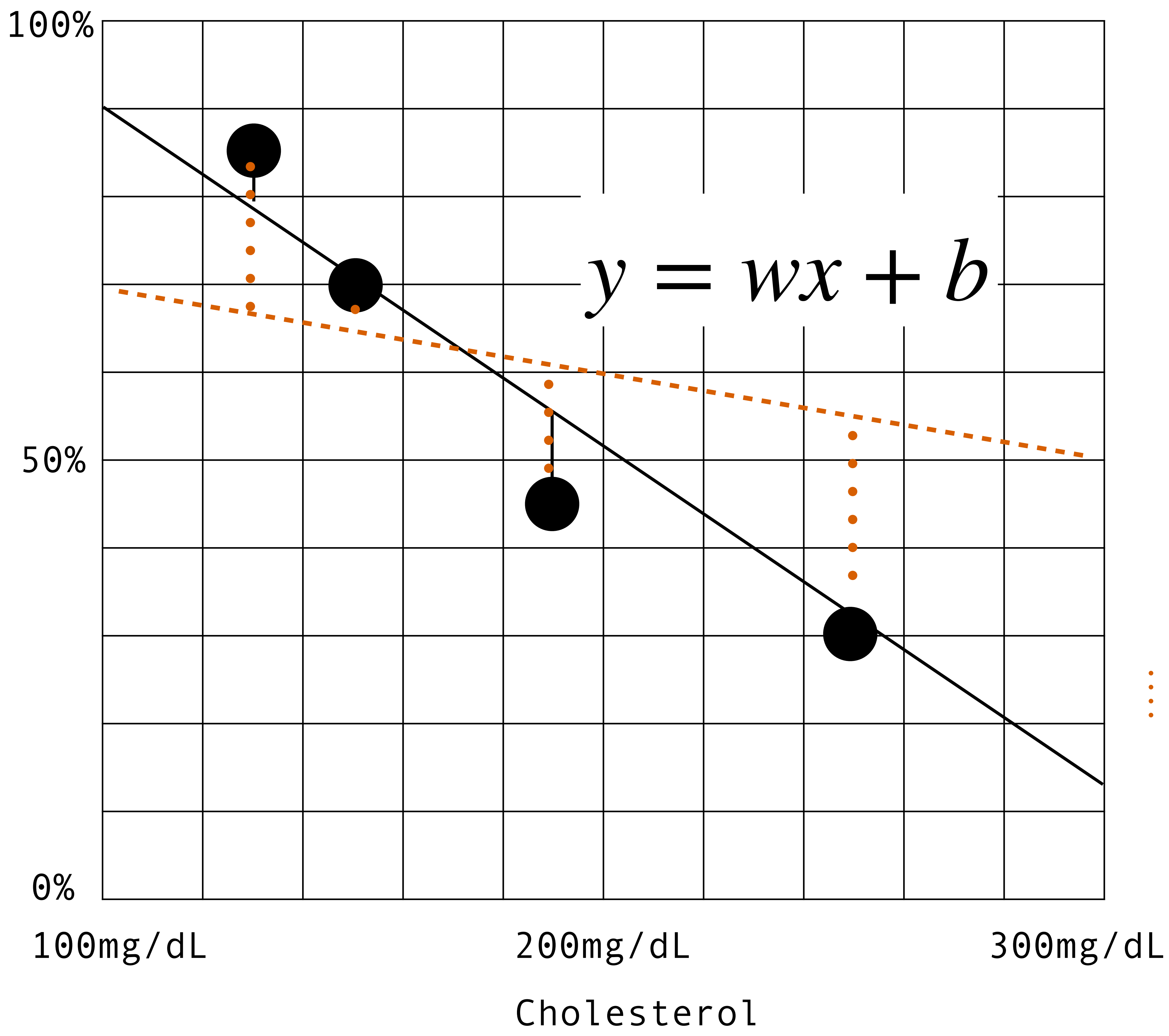
How does Machine Learning (ML) work?

$\texttt{Ejection Fraction} = w \cdot \texttt{Cholesterol} + b$

How do we learn w and b?

We invent a rule for the computer to learn to find the best line.

Rule: Minimize $L(w, b) =$ how far the data is from the line



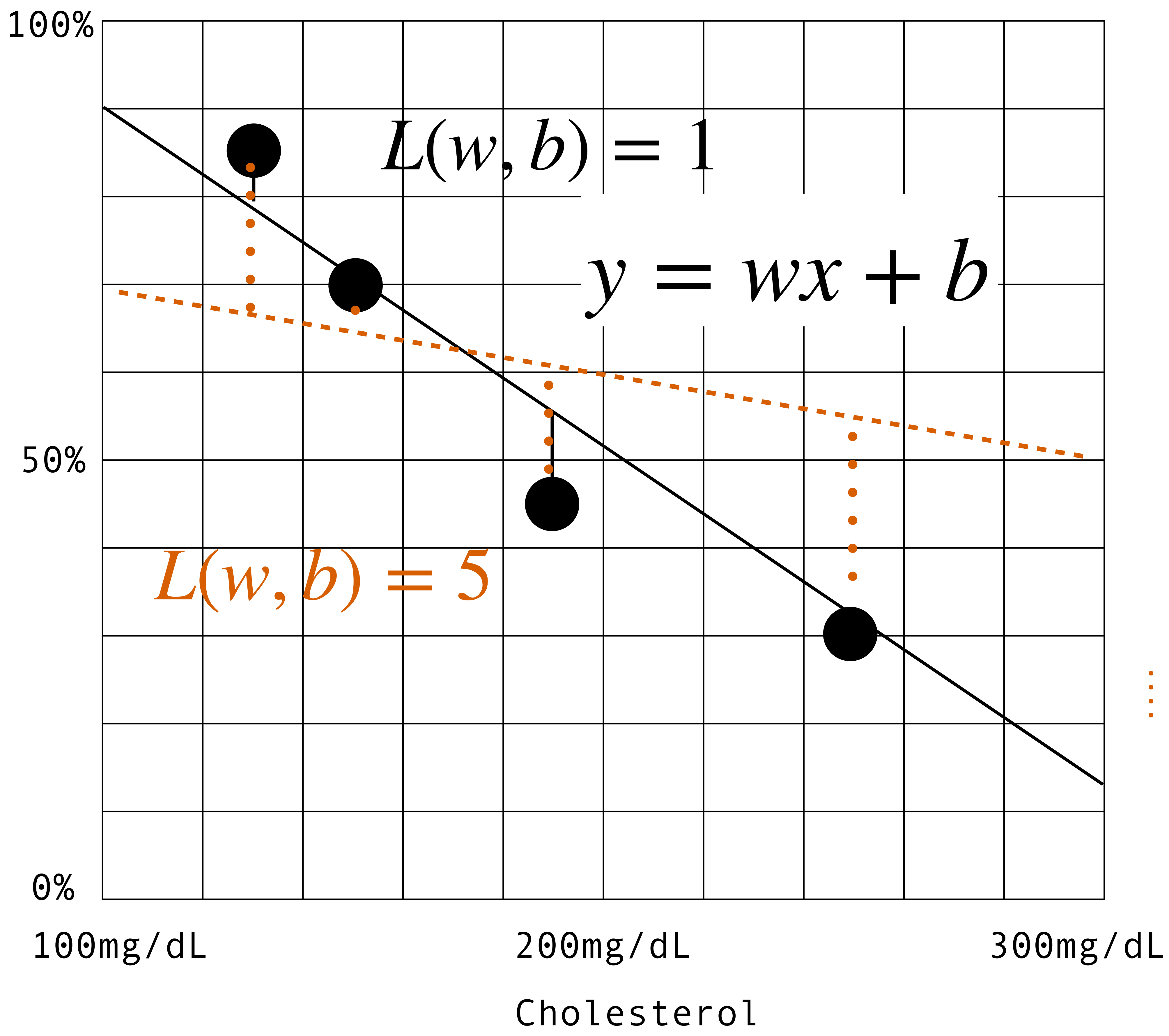
How does Machine Learning (ML) work?

How does the computer use the rule to learn the best w and b?

Rule: Minimize how far the data is from the line

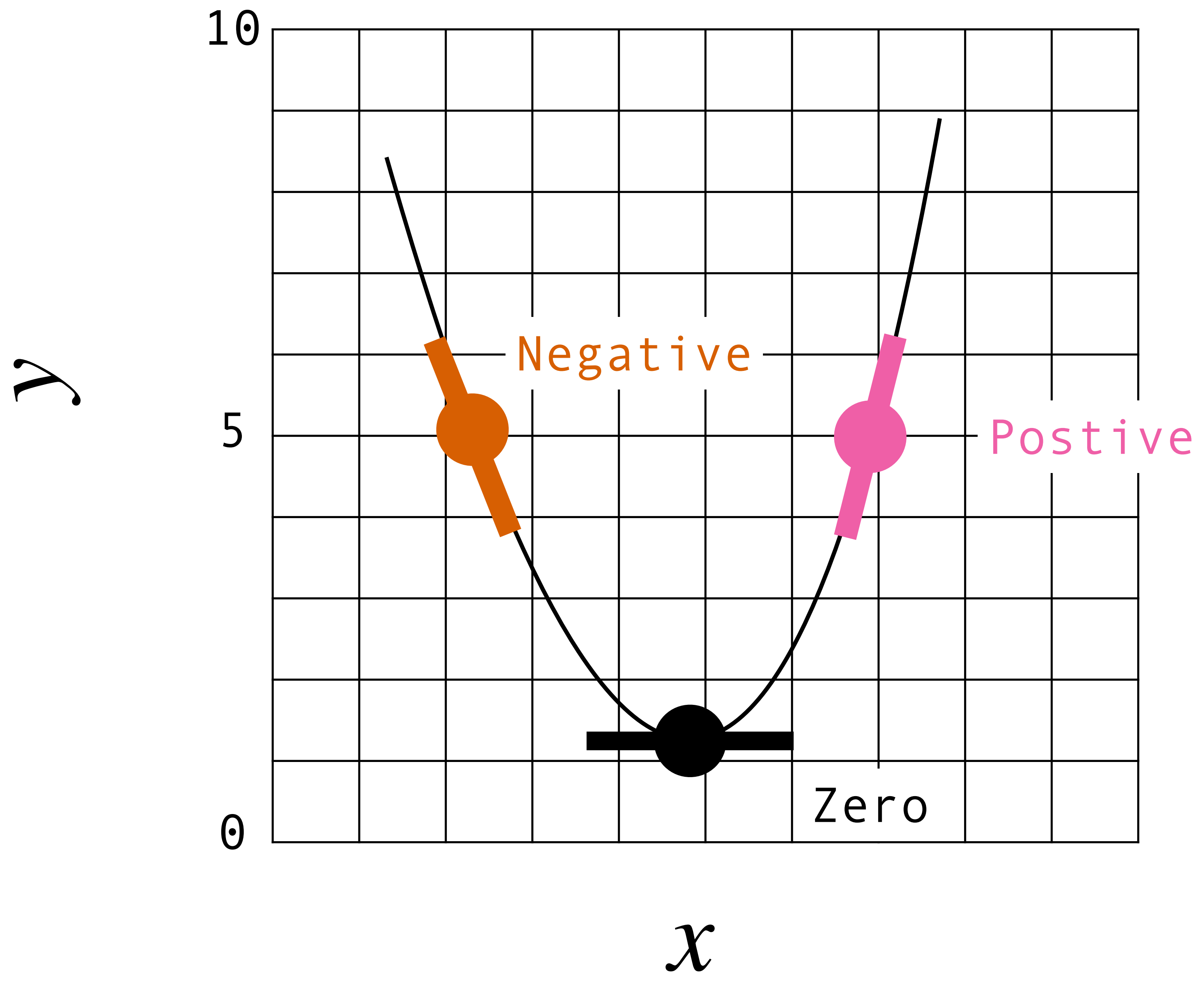
$$L(w, b) = \frac{1}{N} \sum_i (y_i - (wx_i + b))^2$$

We find the right w, b using **gradient descent**.



Derivatives

How fast is a function changing and what direction?

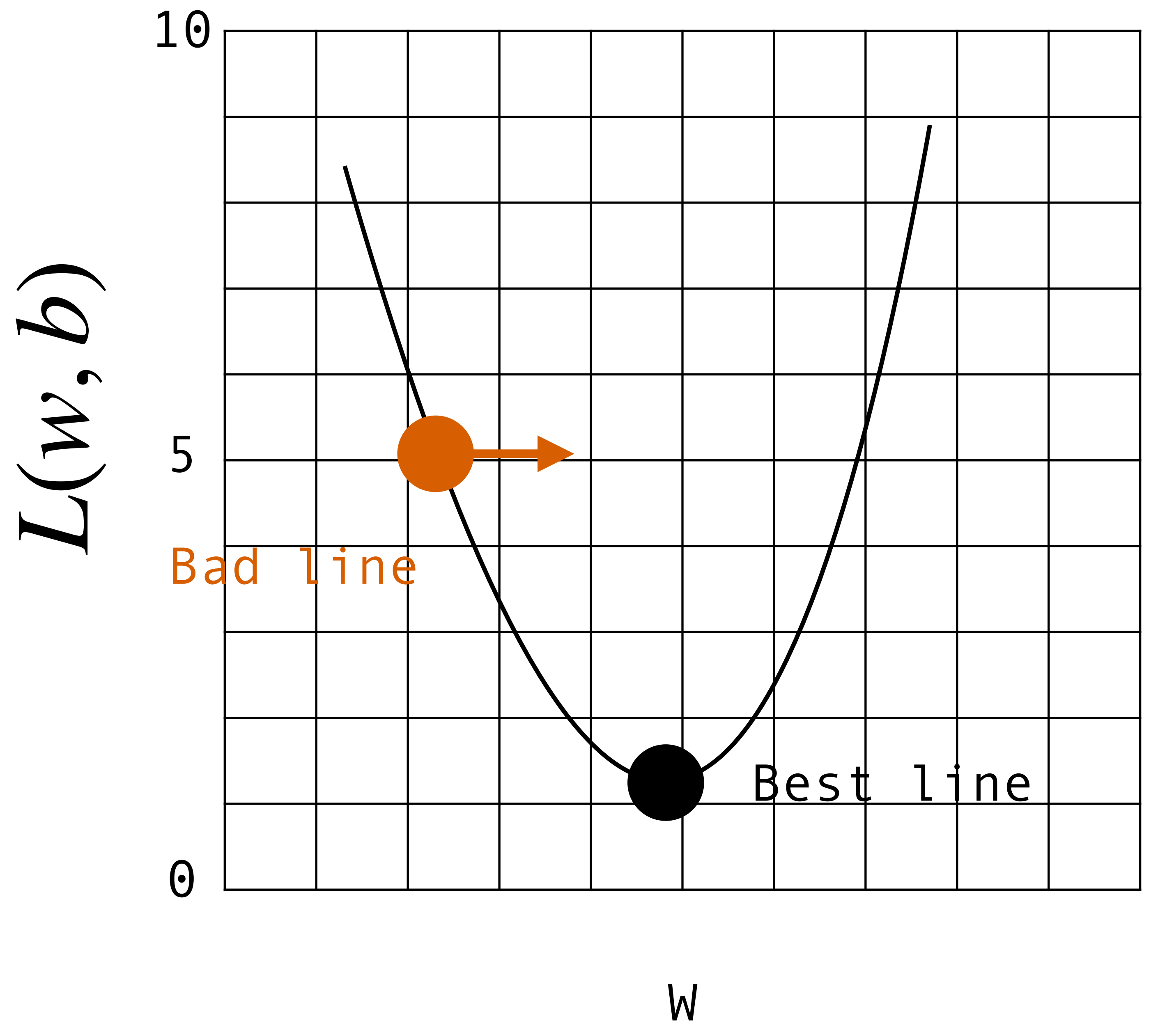


$$\frac{dy}{dx}$$

$$y = wx + b$$

Derivatives

How fast is the loss rule changing and what direction?



We move in the negative direction of the derivative!

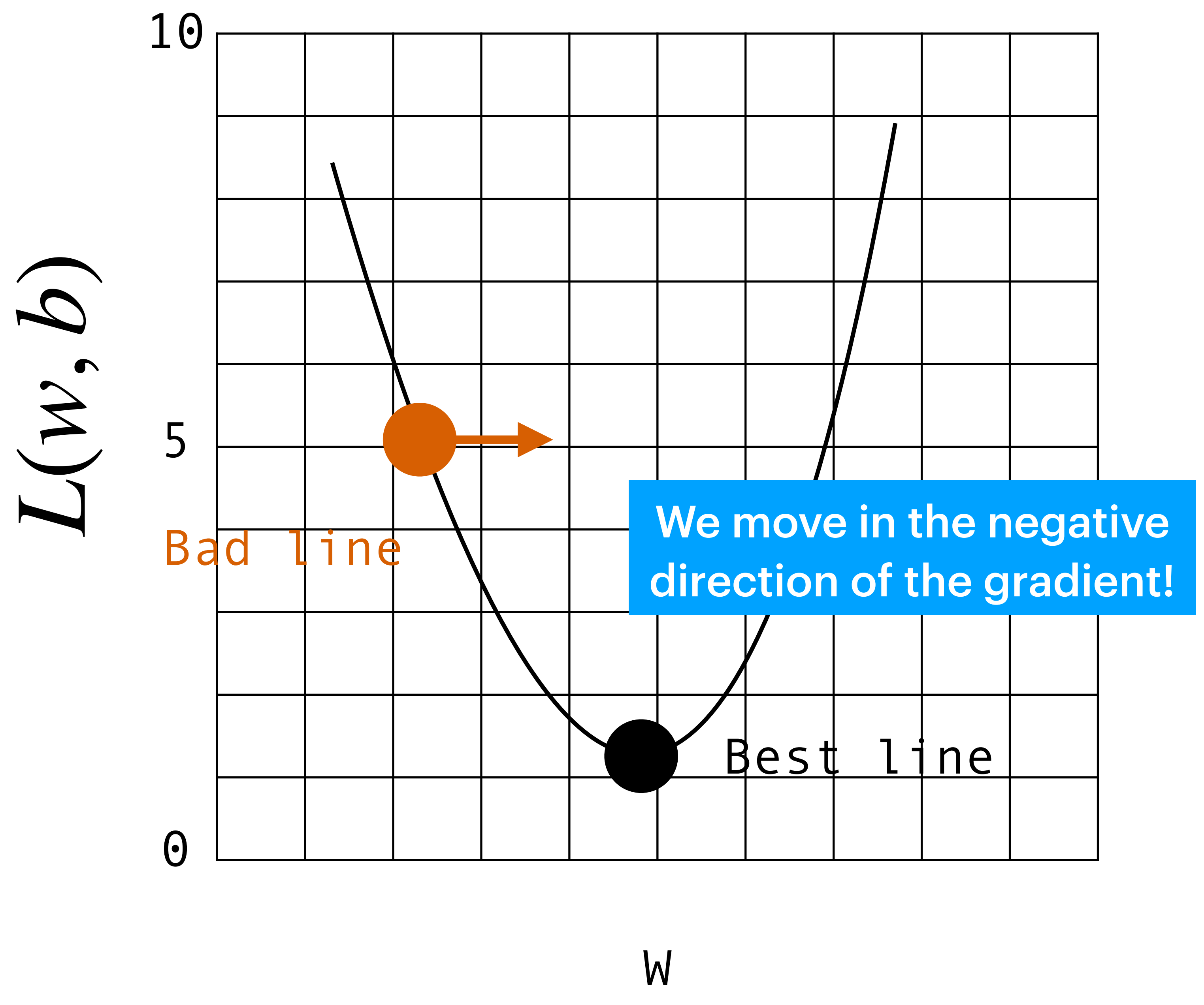
How does Machine Learning (ML) work?

Gradient descent is a process for getting closer to the best w, b according to the rule $L(w, b)$.

1. Compute derivatives to know the direction that can make it smaller.

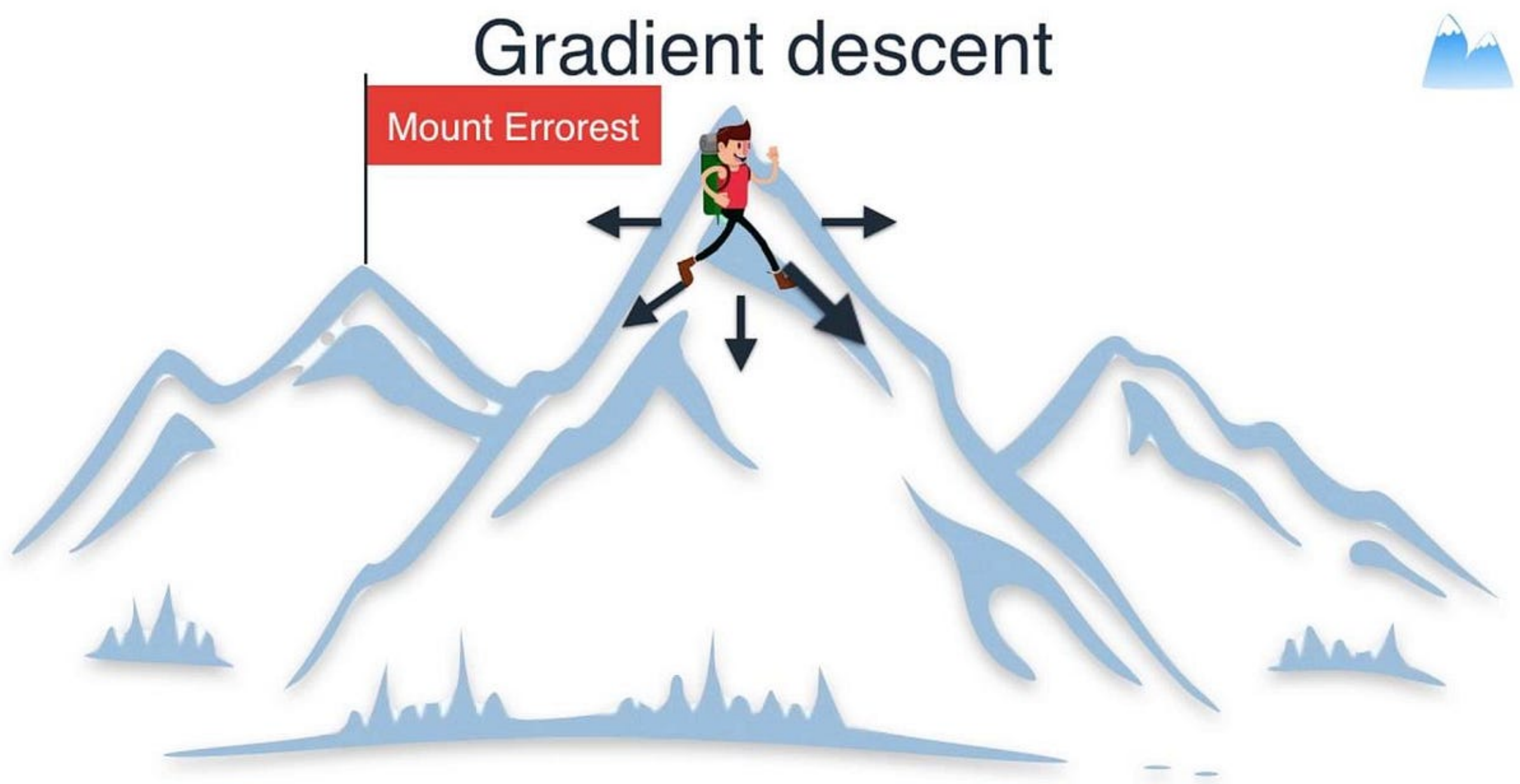
   Calculate $\frac{dL}{dw}$ for direction to change w.

2. Step the model parameters w, b toward the better direction.

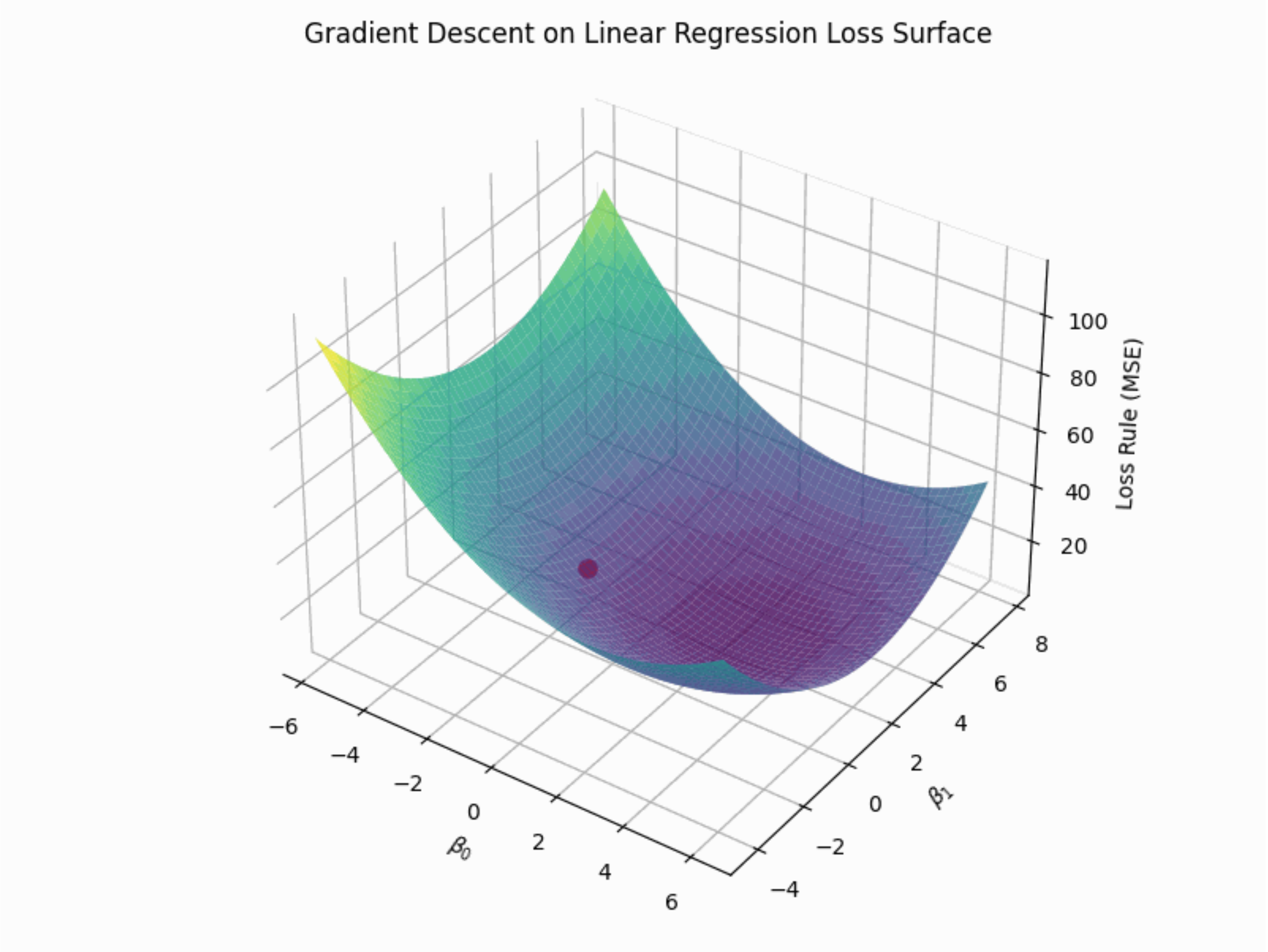


How does Machine Learning (ML) work?

In higher dimensions, its like choosing the best way to hike down a mountain to a valley.



Gradient Descent



Gradient Descent

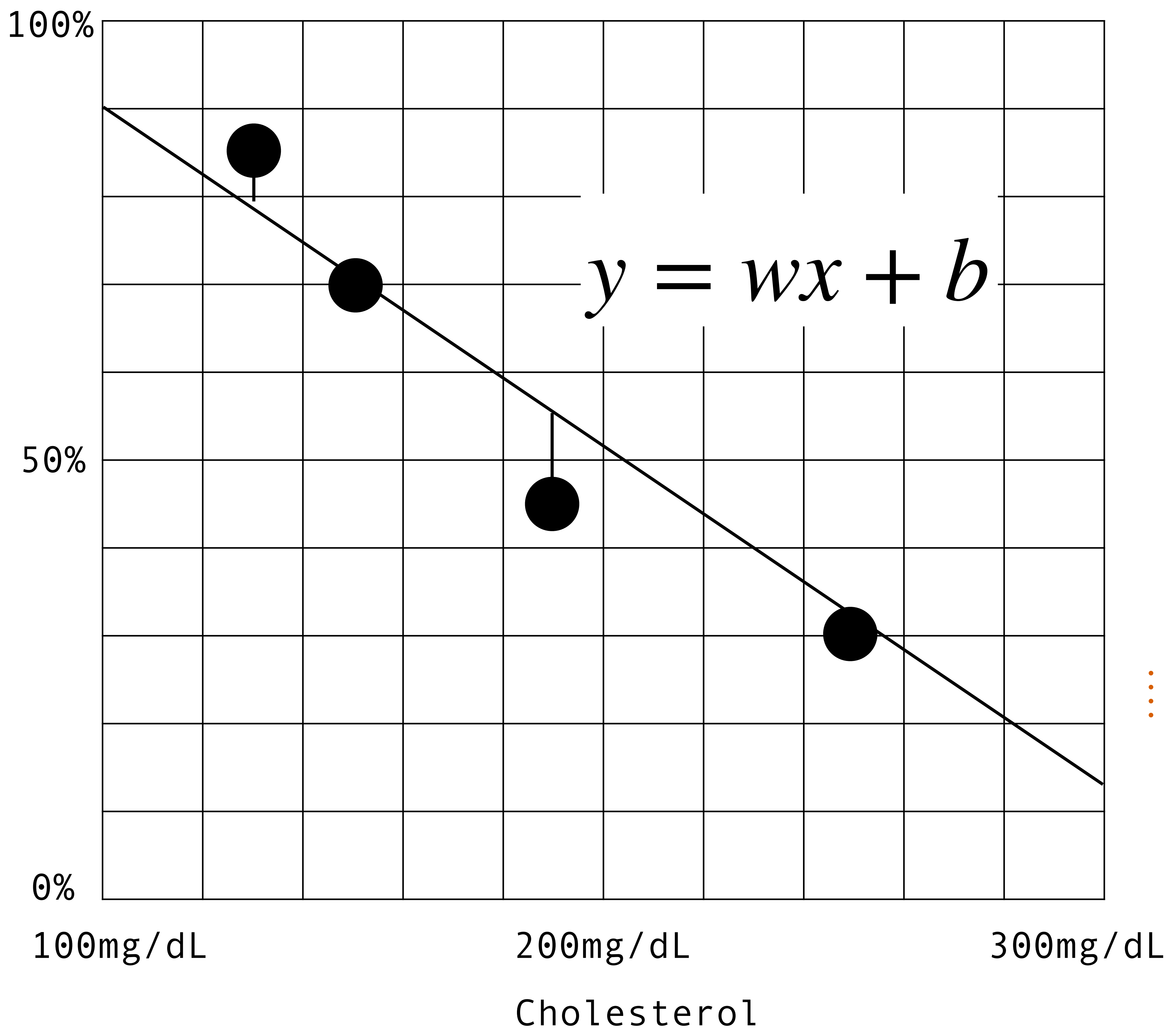
Activity:

https://developers.google.com/machine-learning/crash-course/linear-regression/gradient-descent-exercise

How does Machine Learning (ML) work?

Once we find the correct line with gradient descent, we have “trained” our linear regression.

This is what AI people call **learning**.



How does Machine Learning (ML) work?

Linear regression can predict with any number of variables.

$$\texttt{Ejection Fraction} = w \cdot \texttt{Cholesterol} + w_2 \cdot \texttt{Age} + w_3 \cdot \texttt{Blood pressure} + b$$

Cholesterol	Age	Blood Pressure	Ejection Fraction
190mg/dL	80	130/100	45%
150mg/dL	20	110/80	70%
250mg/dL	50	150/100	30%
130mg/dL	45	90/50	85%

Predictors

Outcome

How does Machine Learning (ML) work?

Ingredients we needed:

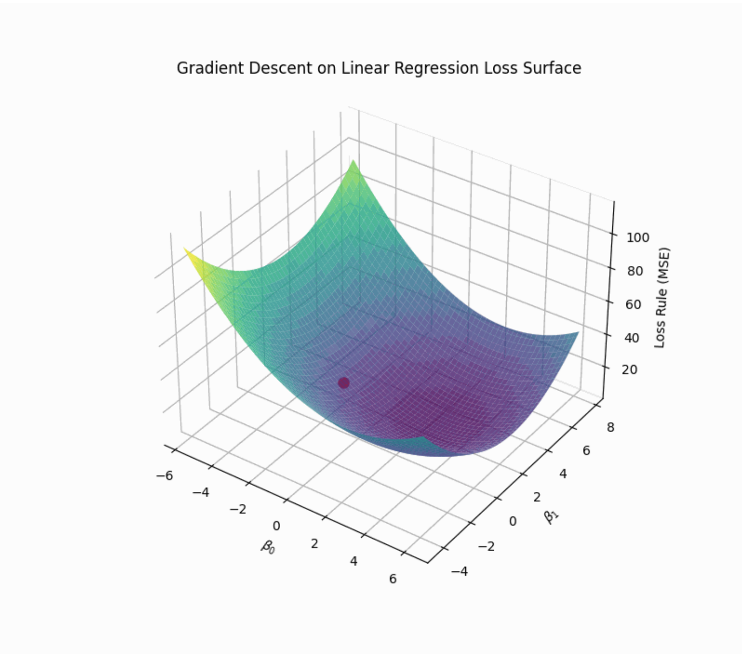
Examples (Data)

Cholesterol	Age	Blood	Ejection
190mg/dL	80	130/100	45%
150mg/dL	20	110/80	70%
250mg/dL	50	150/100	30%
130mg/dL	45	90/50	85%

Model Structure $y = wx + b$

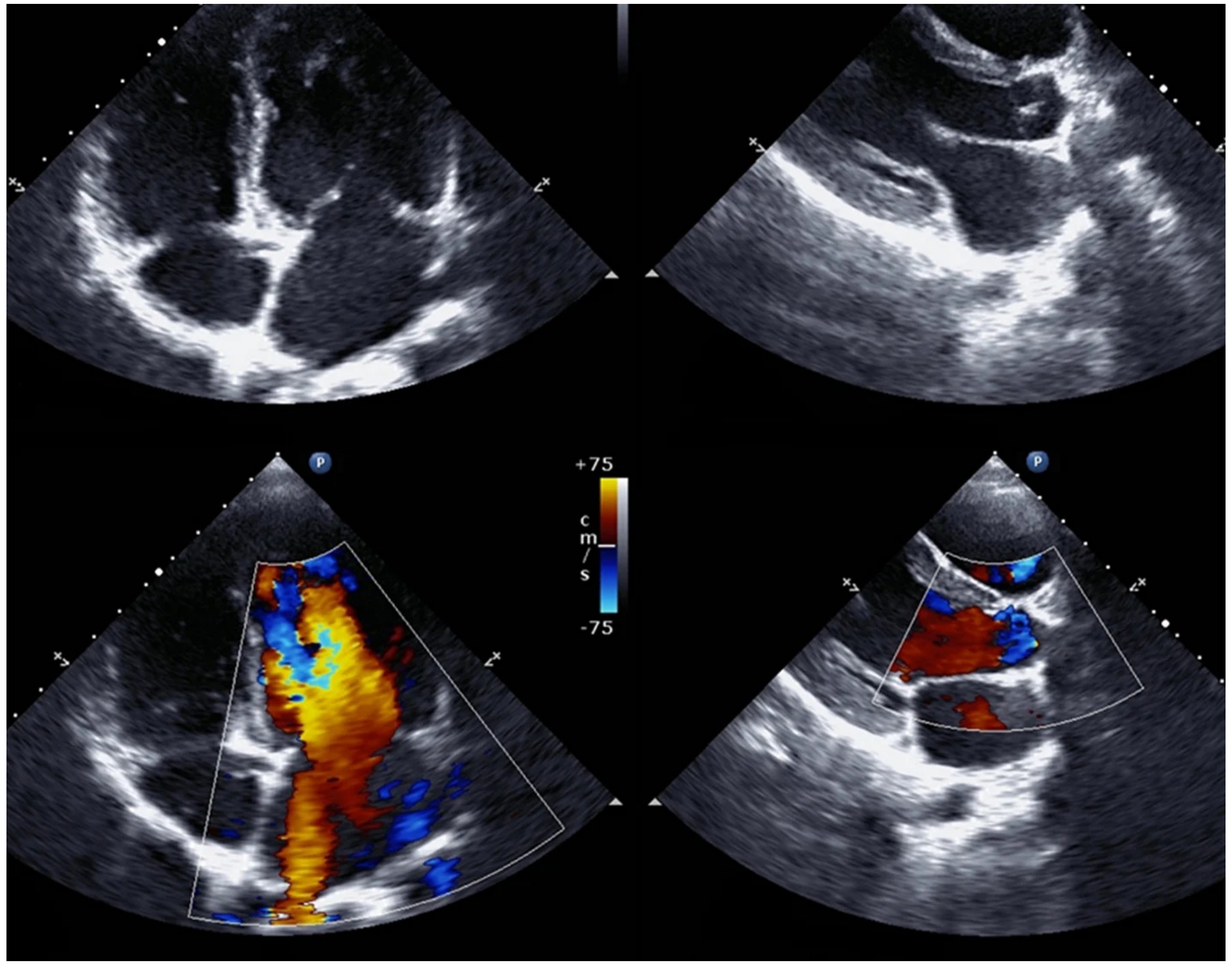
Loss Rule $L(w, b) = \frac{1}{N} \sum_i (y_i - (wx_i + b))^2$

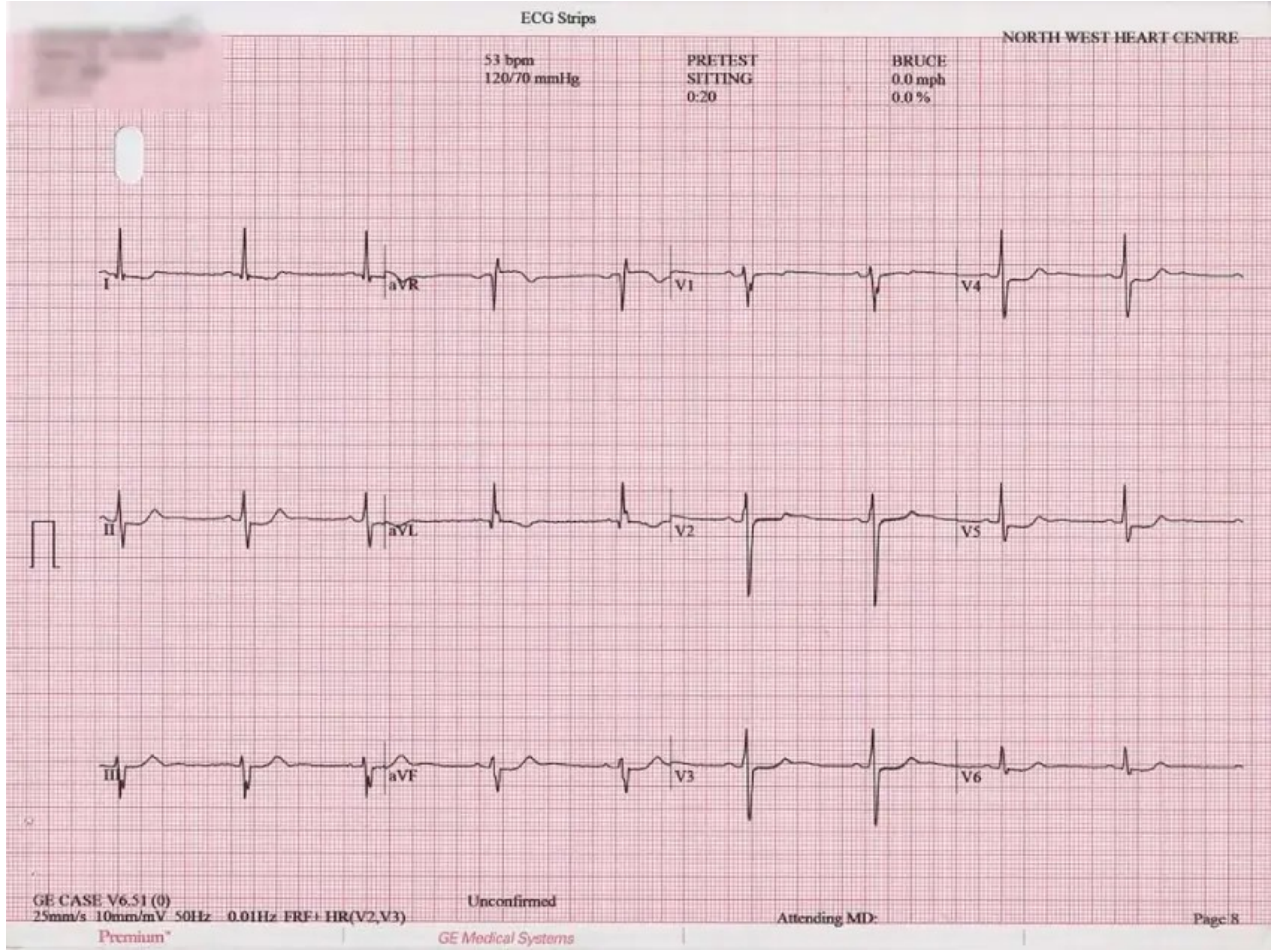
Learning procedure (Gradient Descent)



What is Deep Learning?

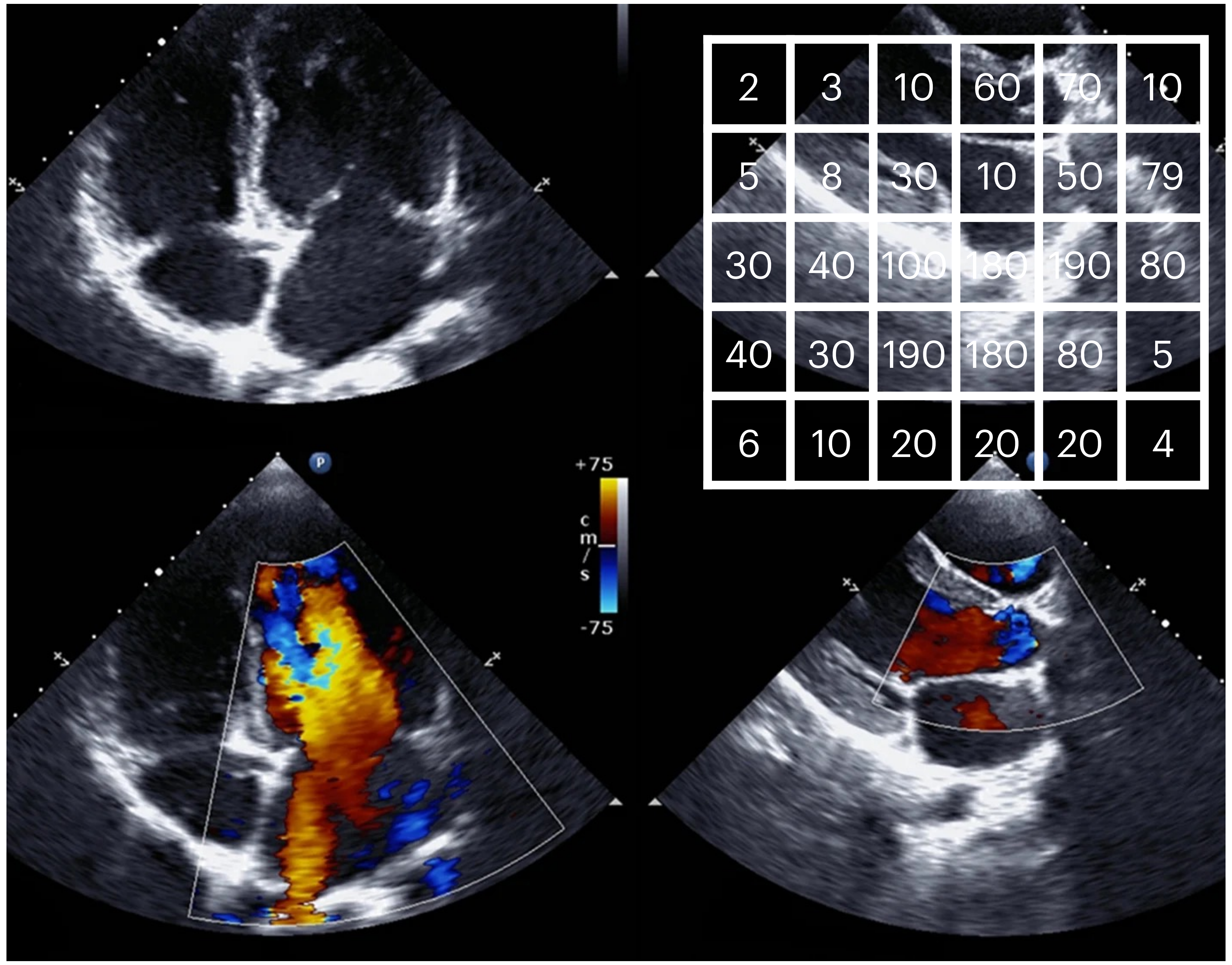
What if our input data are high dimensional and do not have linear relationships with outcomes of interest?



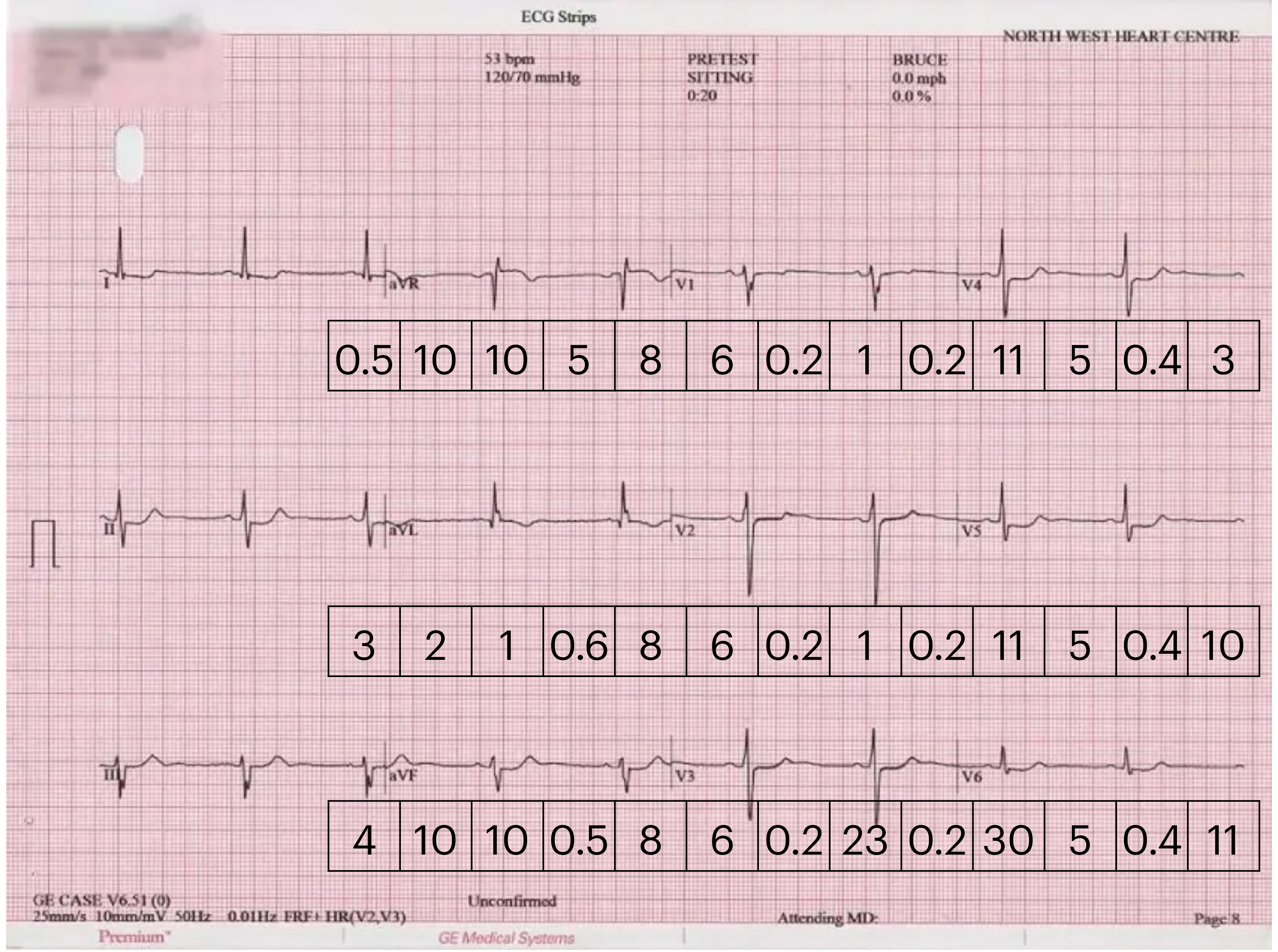


What is Deep Learning?

Images can still be written as numbers, but our function is more complex!



Divide into pixels



Write as amplitudes

What is Deep Learning?

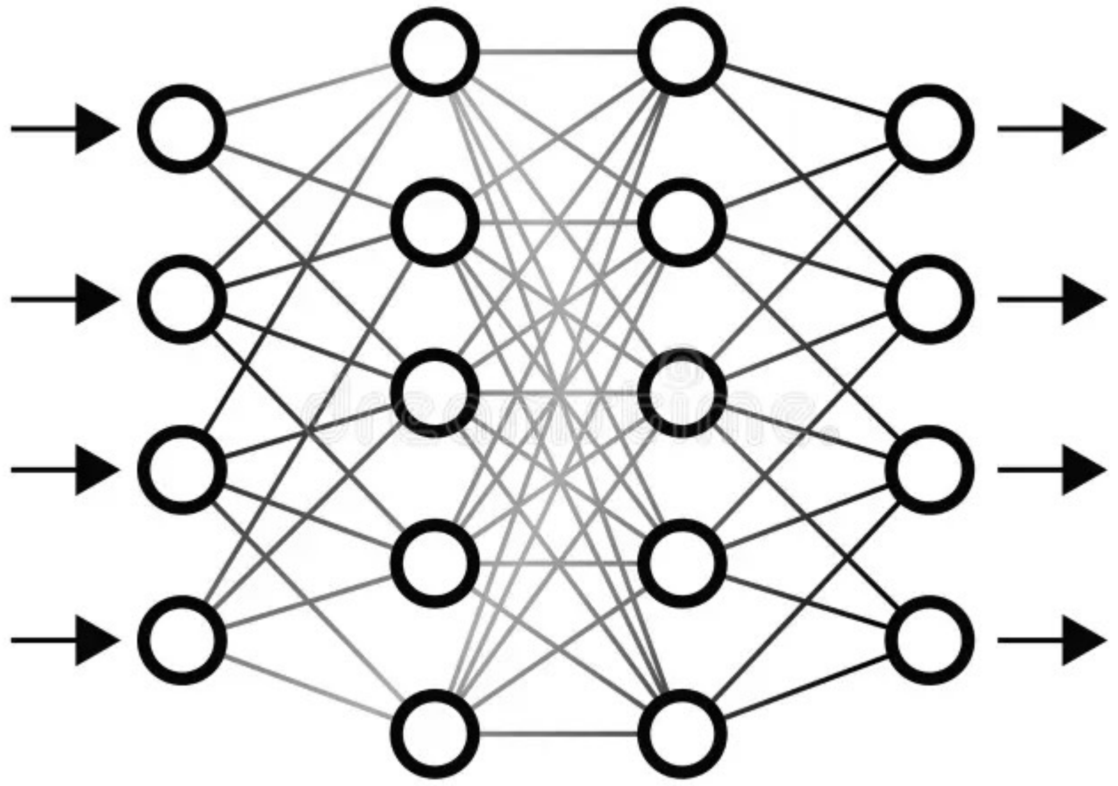
Pixels

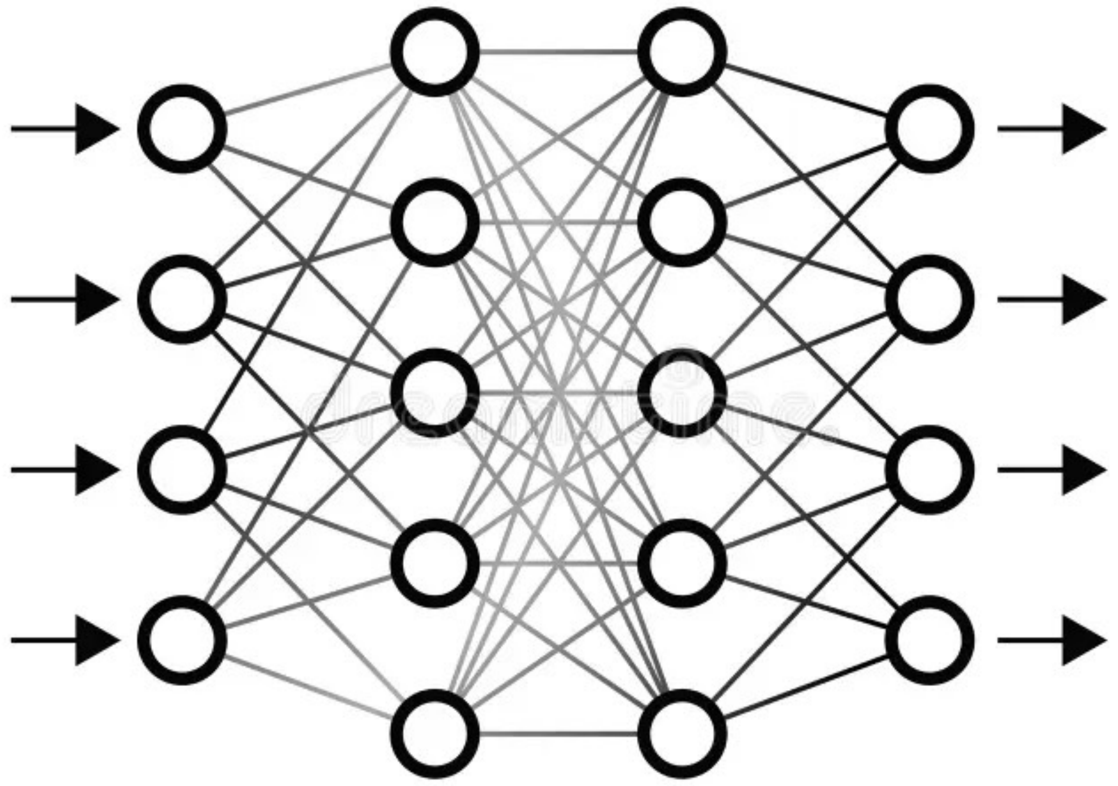
2	3	10	60	70	10
5	8	30	10	50	79
30	40	100	180	190	80
40	30	190	180	80	5
6	10	20	20	20	4

Predictors

Amplitudes

0.5	10	10	5	8	6	0.2	1	0.2	11	5	0.4	3
3	2	1	0.6	8	6	0.2	1	0.2	11	5	0.4	10
4	10	10	0.5	8	6	0.2	23	0.2	30	5	0.4	11





Ejection Fraction
45%
70%
30%
85%

Outcome

Ejection Fraction
45%
70%
30%
85%

Neural Networks

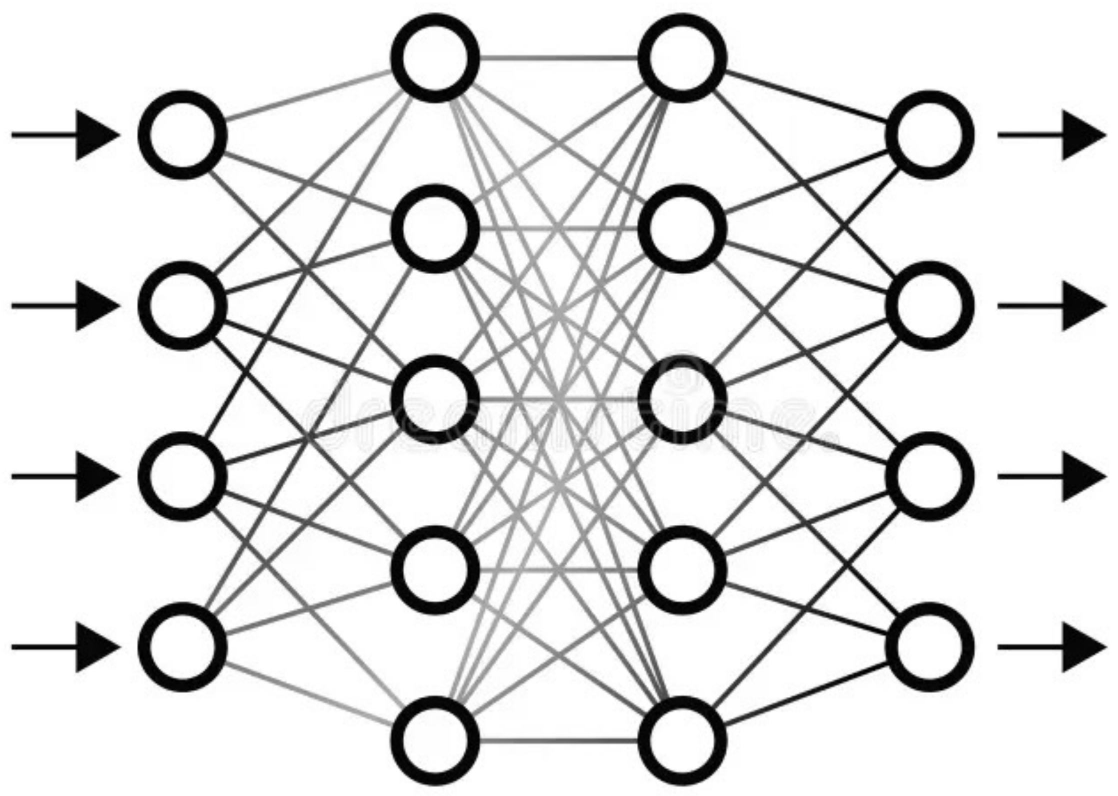
Linear regression has a formula: $y = wx + b$. We learned numbers w, b.

Deep learning uses **neural networks**, which extend this formula.

Pixels

2	3	10	60	70	10
5	8	30	10	50	79
30	40	100	180	190	80
40	30	190	180	80	5
6	10	20	20	20	4

Predictors



Repeated application of:

$$y = \sigma(wx + b)$$

Ejection Fraction
45%
70%
30%
85%

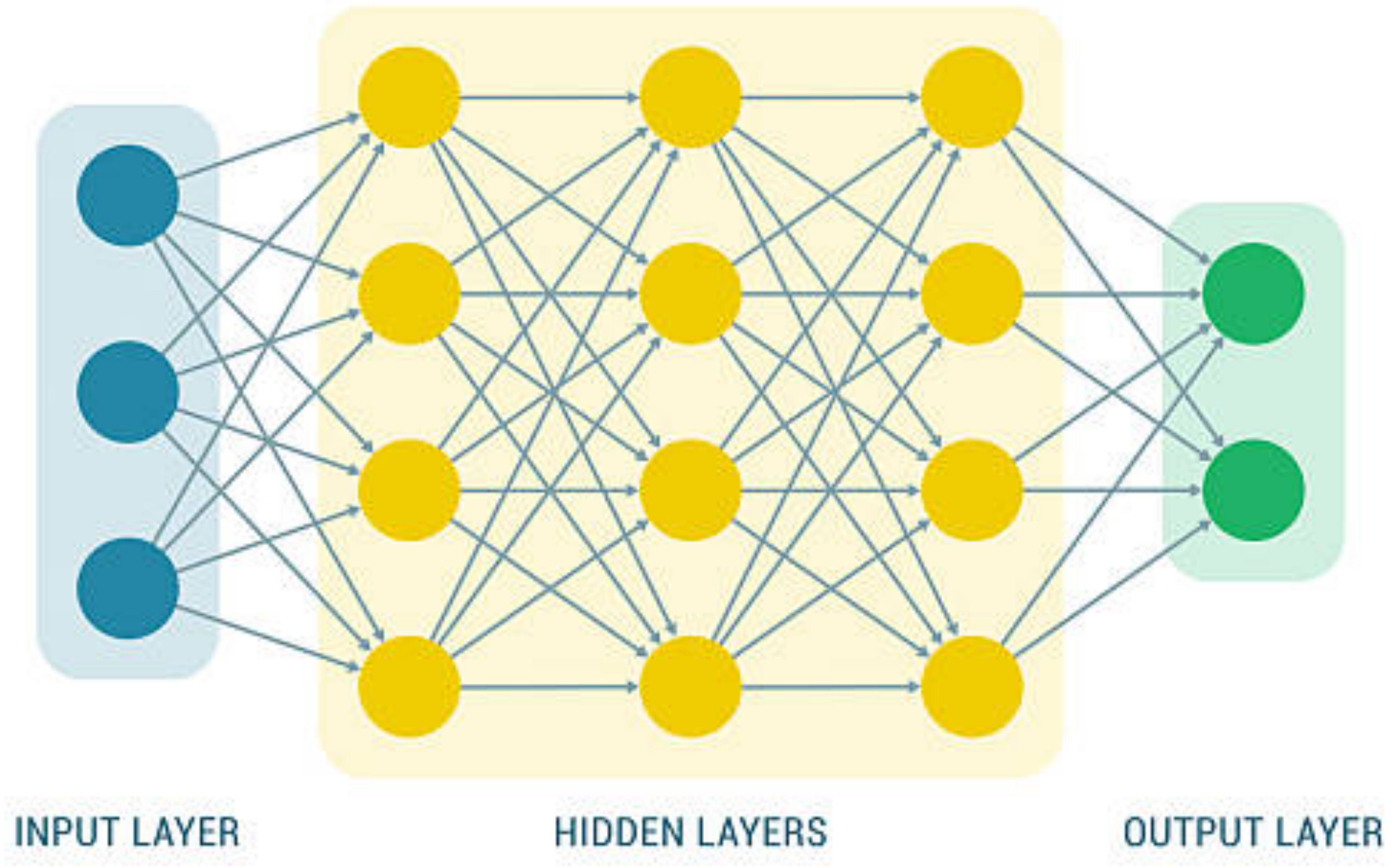
Outcome

Neural Networks

x

3-dimensional.

Predictors



2-dimensional.

Outcome

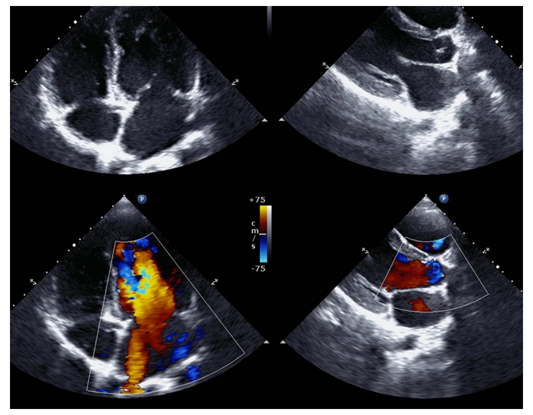
4-dimensional.

But can be any size!

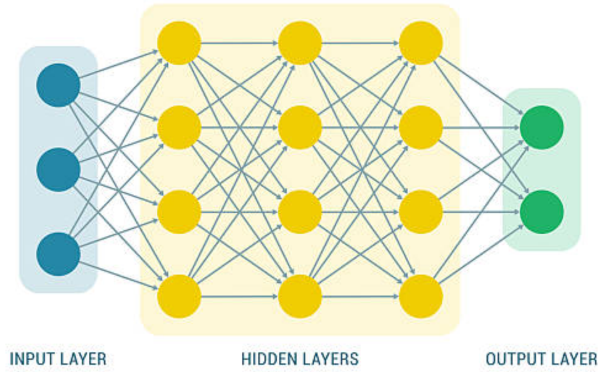
How does Deep Learning work?

Ingredients we need:

Examples (Data)



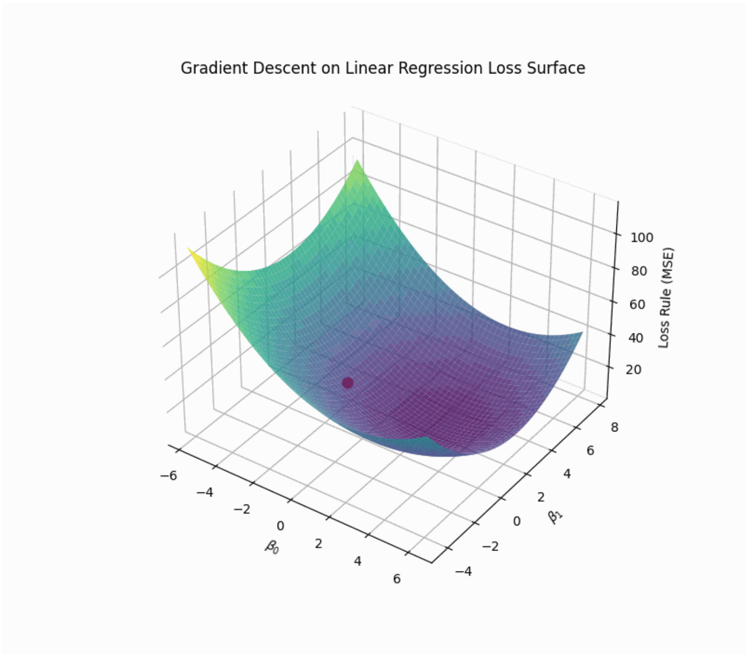
Model Structure



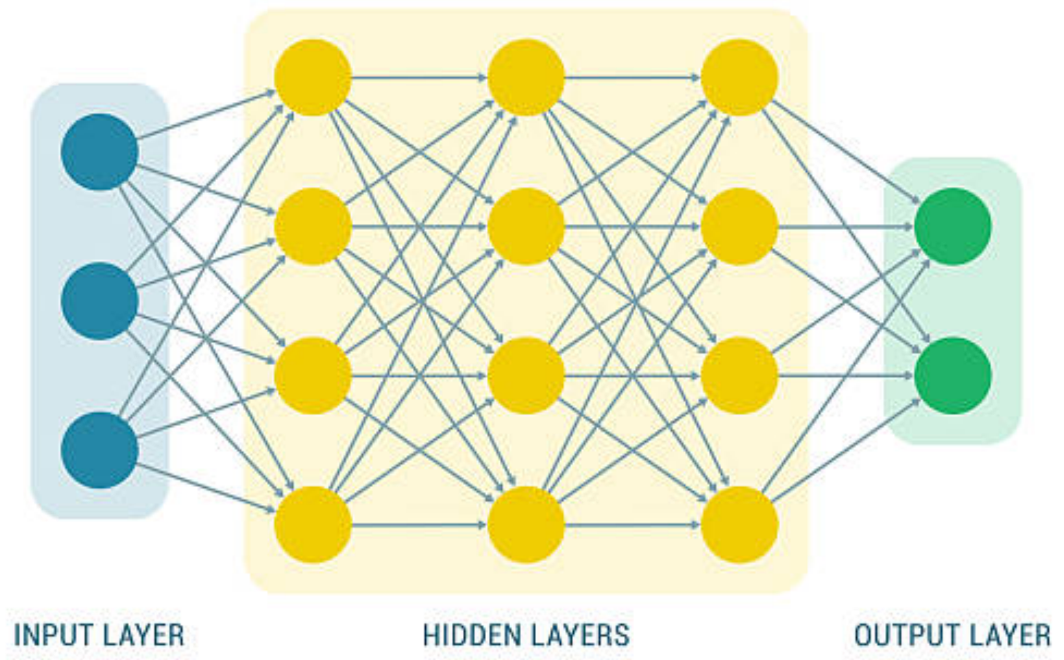
Loss Rule

$$L(w, b) = \frac{1}{N} \sum_i (y_i - (wx_i + b))^2$$

Learning procedure (Gradient Descent)



Neural Networks



At each step:

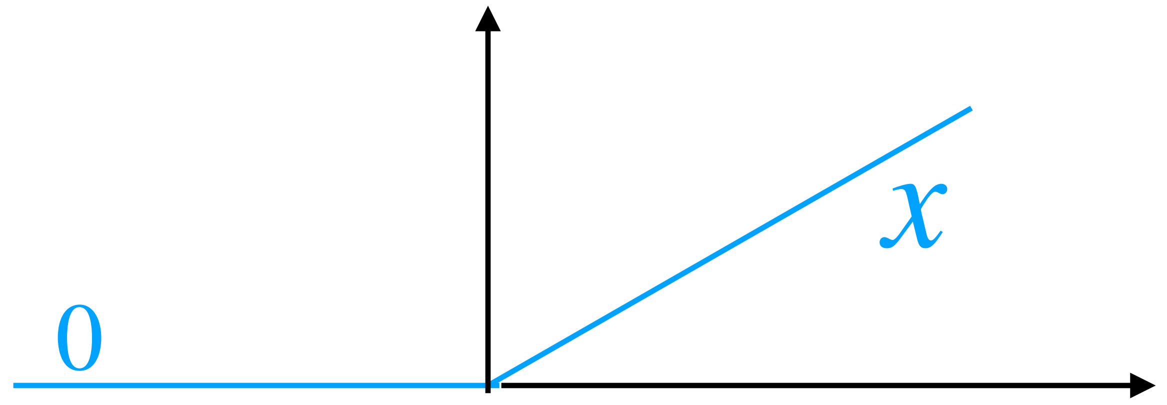
$$x \to w\ x + b \to \sigma(w\ x + b)$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

3 by 4 matrix

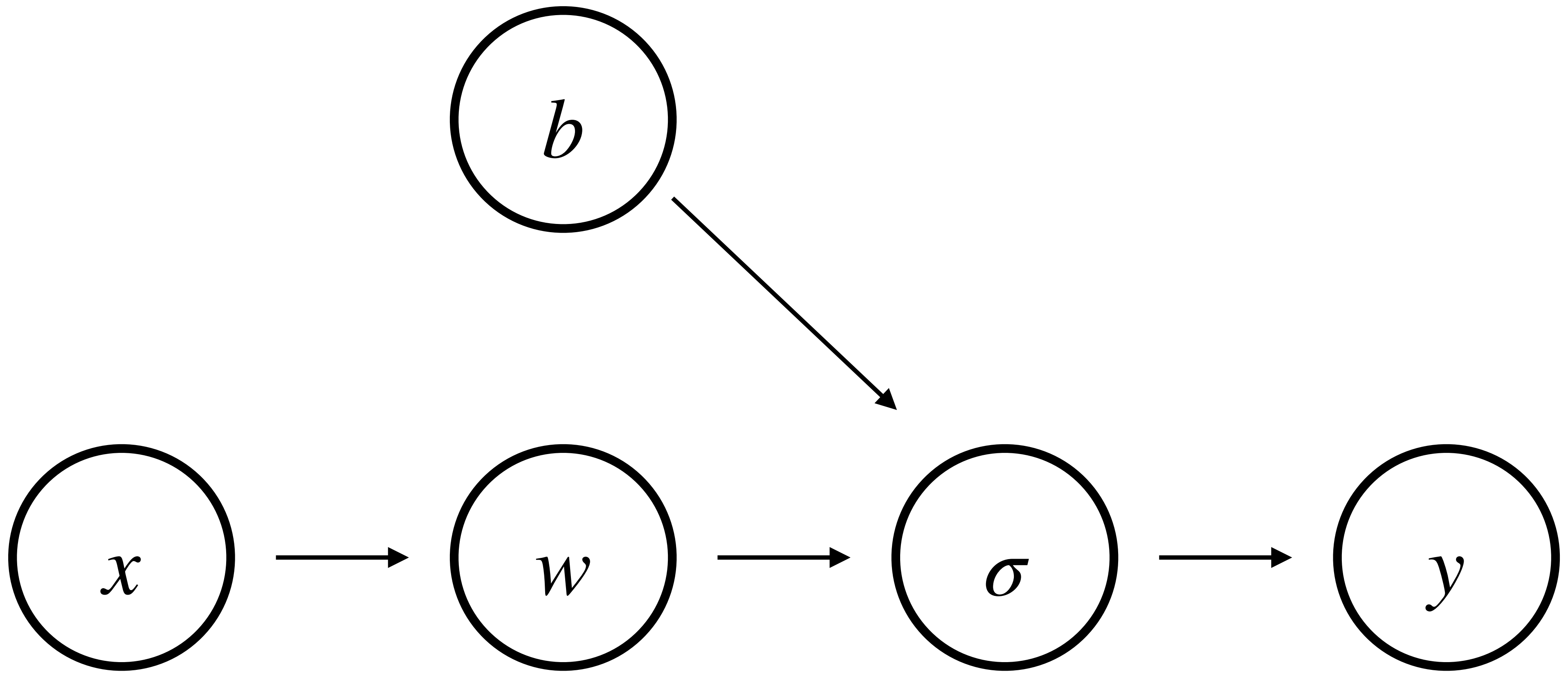
$$\begin{bmatrix} w_{11} & w_{12} & w_{13} & w_{14} \\ w_{21} & w_{22} & w_{23} & w_{24} \\ w_{31} & w_{32} & w_{33} & w_{34} \end{bmatrix}$$

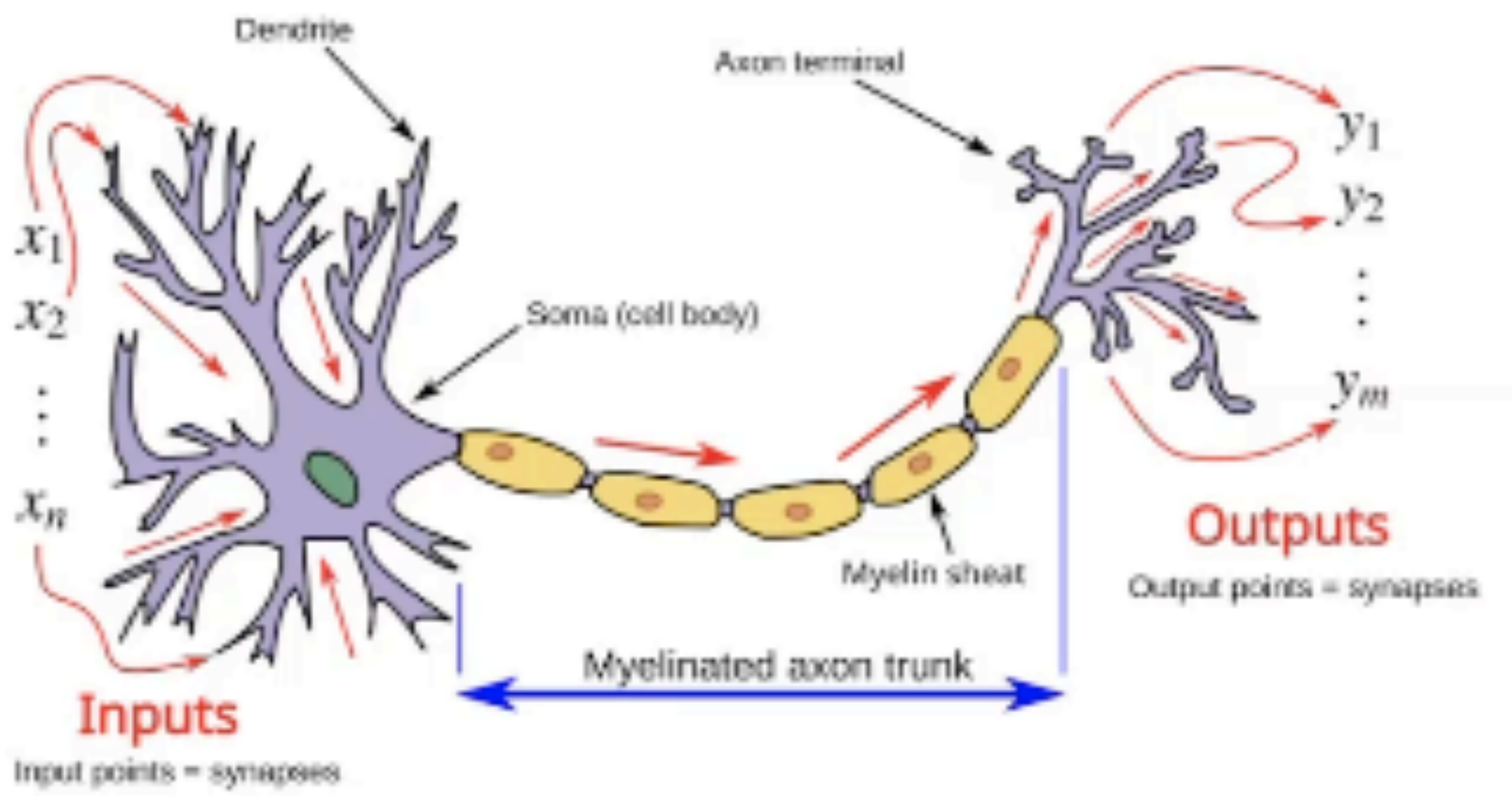
$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix}$$



σ is ReLU function

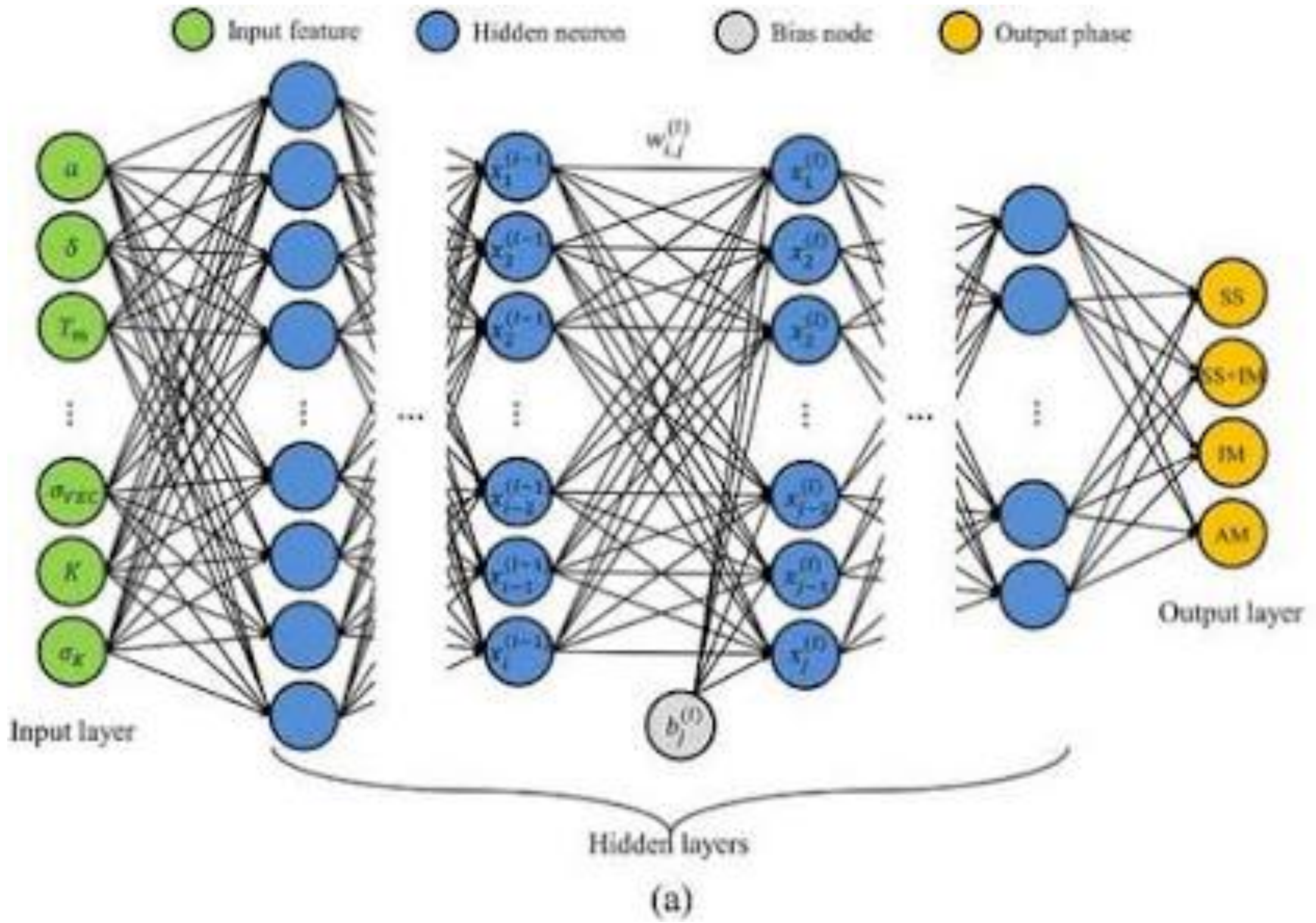
Neural Networks





Neural Networks

Bigger networks are just the same thing!



Neural Networks are mathematically special!

A spate of papers in the 1980s—1990s, from George Cybenko and Kurt Hornik [de] etc, established several universal approximation theorems for arbitrary width and bounded depth.[44][1][7][8] See[45][46][10] for reviews. The following is the most often quoted:

Universal approximation theorem—Let $C(X, \mathbb{R}^m)$ denote the set of continuous functions from a subset X of a Euclidean $\mathbb{R}^n$ space to a Euclidean space $\mathbb{R}^m$. Let $\sigma \in C(\mathbb{R}, \mathbb{R})$. Note that $(\sigma \circ x)_i = \sigma(x_i)$, so $\sigma \circ x$ denotes σ applied to each component of x.

Then σ is not polynomial if and only if for every $n \in \mathbb{N}$, $m \in \mathbb{N}$, compact $K \subseteq \mathbb{R}^n$, $f \in C(K, \mathbb{R}^m), \varepsilon > 0$ there exist $k \in \mathbb{N}$, $A \in \mathbb{R}^{k \times n}$, $b \in \mathbb{R}^k$, $C \in \mathbb{R}^{m \times k}$ such that

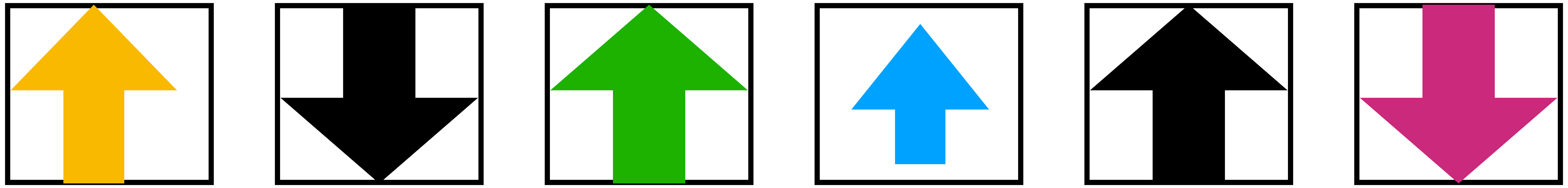
$$\sup_{x \in K} \|f(x) - g(x)\| < \varepsilon$$

where $g(x) = C \cdot (\sigma \circ (A \cdot x + b))$

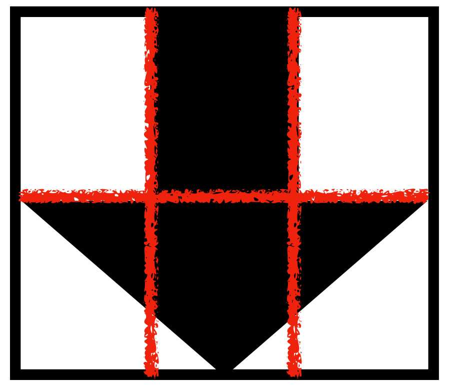
Math research from the 1980s-90s shows that neural networks can approximate any mathematical function as long as they have non-linearities.

How does Deep Learning work?

Task: Is the arrow up or down?



Image



Pixels

0	1	0
0.5	0.8	0.5

Matrix

$$\begin{bmatrix} 0 & 1 & 0 \\ 0.5 & 0.8 & 0.5 \end{bmatrix}$$

Neural Networks

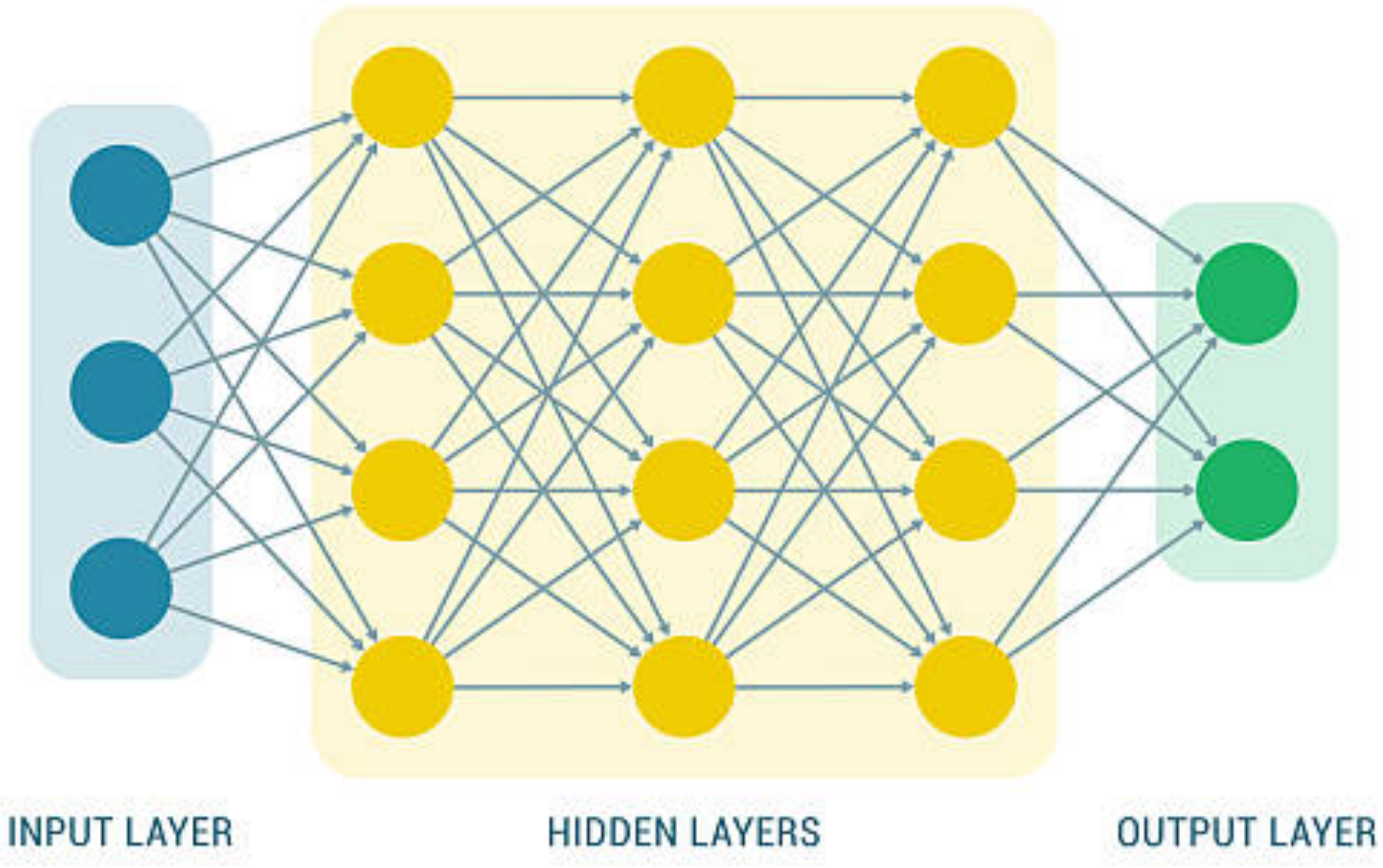
$$x$$

Pixels

$$\begin{bmatrix} 0 & 1 & 0 \\ 0.5 & 0.8 & 0.5 \end{bmatrix}$$

6-dimensional.

Predictors

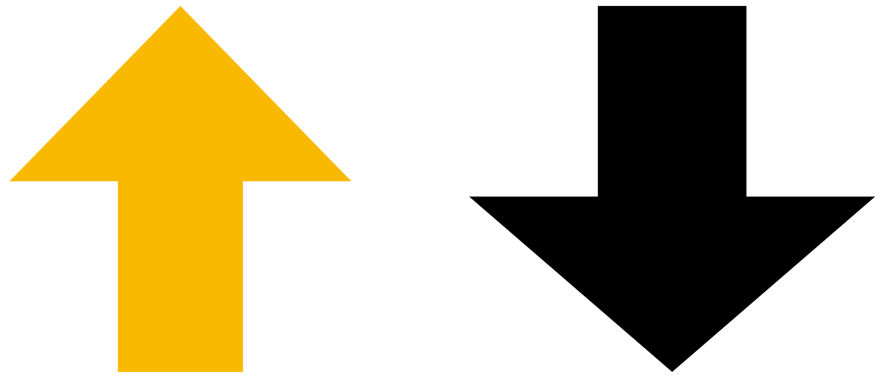


4-dimensional.

But can be any size!

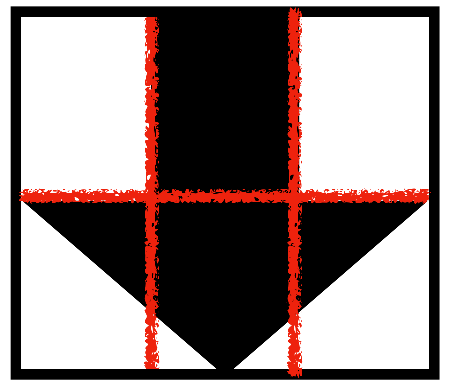
2-dimensional.

Outcome



How does Deep Learning work?

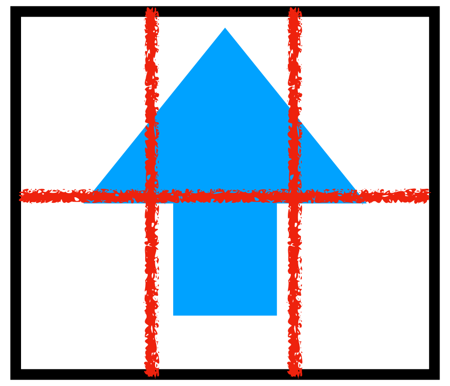
Image



Matrix

$$\begin{bmatrix} 0 & 1 & 0 \\ 0.5 & 0.8 & 0.5 \end{bmatrix}$$

Image



Matrix

$$\begin{bmatrix} 0.2 & 0.8 & 0.2 \\ 0 & 0.65 & 0 \end{bmatrix}$$

Need a function:

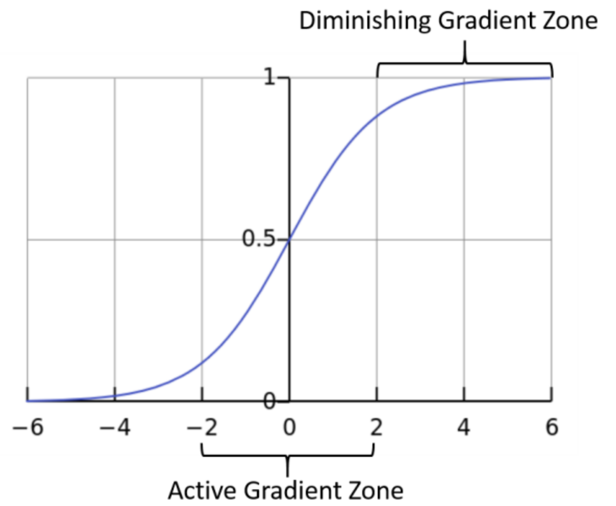
$$x \rightarrow w \; x \rightarrow \sigma(w \; x) \rightarrow [p_{up} \quad p_{down}]$$

6 by 1 vector

2 by 6 matrix

$$\begin{bmatrix} w_{11} & w_{12} & \dots & w_{16} \\ w_{21} & w_{22} & \dots & w_{26} \end{bmatrix} \quad \begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_6 \end{bmatrix}$$

$$\frac{1}{1+e^{-x}}$$

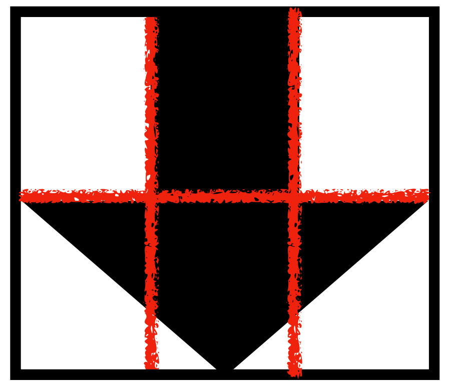


σ is sigmoid function

How does Deep Learning work?

What do we notice? What values of x should be weighted highly for each?

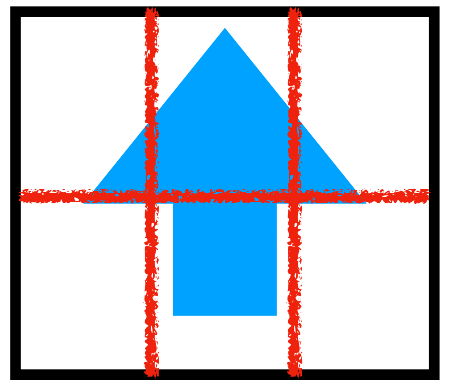
Image



Matrix

$$\begin{bmatrix} 0 & 1 & 0 \\ 0.5 & 0.8 & 0.5 \end{bmatrix}$$

Image

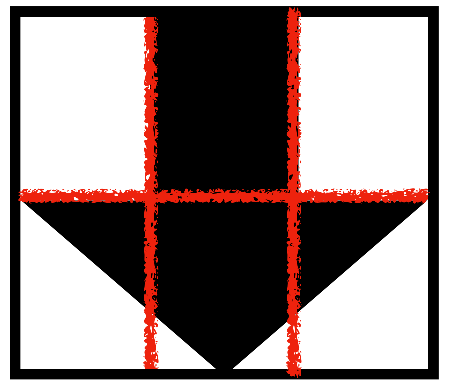


Matrix

$$\begin{bmatrix} 0.2 & 0.8 & 0.2 \\ 0 & 0.65 & 0 \end{bmatrix}$$

How does Deep Learning work?

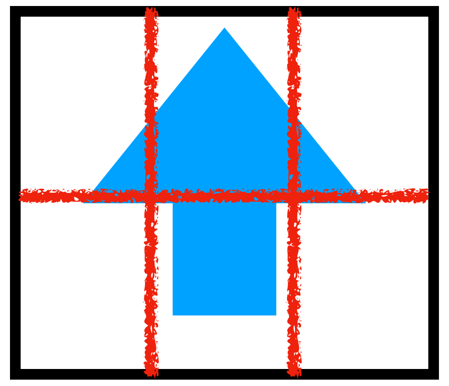
Image



Matrix

$$\begin{bmatrix} 0 & 1 & 0 \\ 0.5 & 0.8 & 0.5 \end{bmatrix}$$

Image



Matrix

$$\begin{bmatrix} 0.2 & 0.8 & 0.2 \\ 0 & 0.65 & 0 \end{bmatrix}$$

$$\sigma\left(\begin{bmatrix} w_{11} & w_{12} & \dots & w_{16} \\ w_{21} & w_{22} & \dots & w_{26} \end{bmatrix}\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0.5 \\ 0.8 \\ 0.5 \end{bmatrix}\right) = \begin{bmatrix} p_{down} & p_{up} \end{bmatrix}$$

$$\sigma\left(\begin{bmatrix} w_{11} & w_{12} & \dots & w_{16} \\ w_{21} & w_{22} & \dots & w_{26} \end{bmatrix}\begin{bmatrix} 0.2 \\ 0.8 \\ 0.2 \\ 0.0 \\ 0.65 \\ 0.0 \end{bmatrix}\right) = \begin{bmatrix} p_{down} & p_{up} \end{bmatrix}$$

Matrix Multiplication

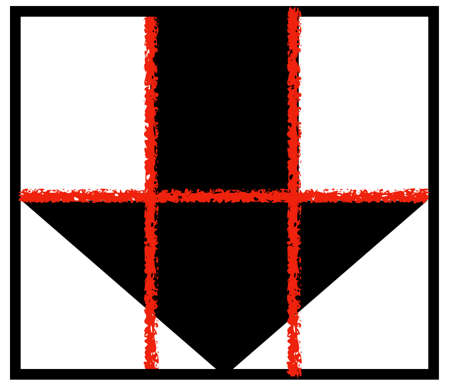
$$\sigma\left(\begin{bmatrix} w_{11} & w_{12} & \dots & w_{16} \\ w_{21} & w_{22} & \dots & w_{26} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0.5 \\ 0.8 \\ 0.5 \end{bmatrix}\right) = \begin{bmatrix} p_{down} & p_{up} \end{bmatrix}$$

$$p_{up} = 0 \cdot w_{11} + 1 \cdot w_{12} + 0 \cdot w_{13} + 0.5 \cdot w_{14} + 0.8 \cdot w_{15} + 0.5 \cdot w_{16}$$

$$p_{down} = 0 \cdot w_{21} + 1 \cdot w_{22} + 0 \cdot w_{23} + 0.5 \cdot w_{24} + 0.8 \cdot w_{25} + 0.5 \cdot w_{26}$$

How does Deep Learning work?

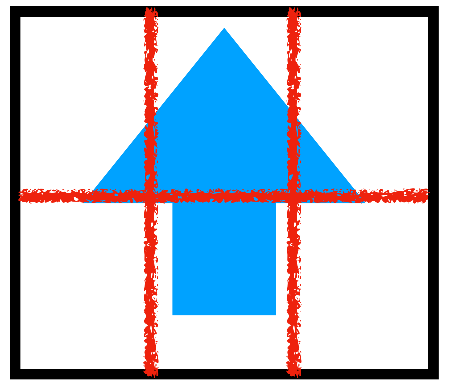
Image



Matrix

$$\begin{bmatrix} 0 & 1 & 0 \\ 0.5 & 0.8 & 0.5 \end{bmatrix}$$

Image



Matrix

$$\begin{bmatrix} 0.2 & 0.8 & 0.2 \\ 0 & 0.65 & 0 \end{bmatrix}$$

$$\sigma\left(\begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0.5 \\ 0.8 \\ 0.5 \end{bmatrix} \right) = \begin{bmatrix} p_{down} & p_{up} \end{bmatrix}$$

$$\sigma\left(\begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.2 \\ 0.8 \\ 0.2 \\ 0.0 \\ 0.65 \\ 0.0 \end{bmatrix} \right) = \begin{bmatrix} p_{down} & p_{up} \end{bmatrix}$$

$$\sigma\left([1 \quad 0] \right) = [0.731 \quad 0.269]$$

$$\sigma\left([0 \quad 0.4] \right) = [0.401 \quad 0.599]$$

Really using softmax... think of as multiclass sigmoid!

How does Deep Learning work?

How does the computer learn this?

Define a loss rule $L(w) = \frac{1}{N} \sum_i (1 - y^i_{down}) \cdot (1 - \log(p_{down})) + y^i_{down} \cdot \log p_{down}$

Binary Cross Entropy Loss

Guess w: $\begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$

Compute on examples:

$$\sigma\left([1 \quad 0] \right) = [0.731 \quad 0.269]$$

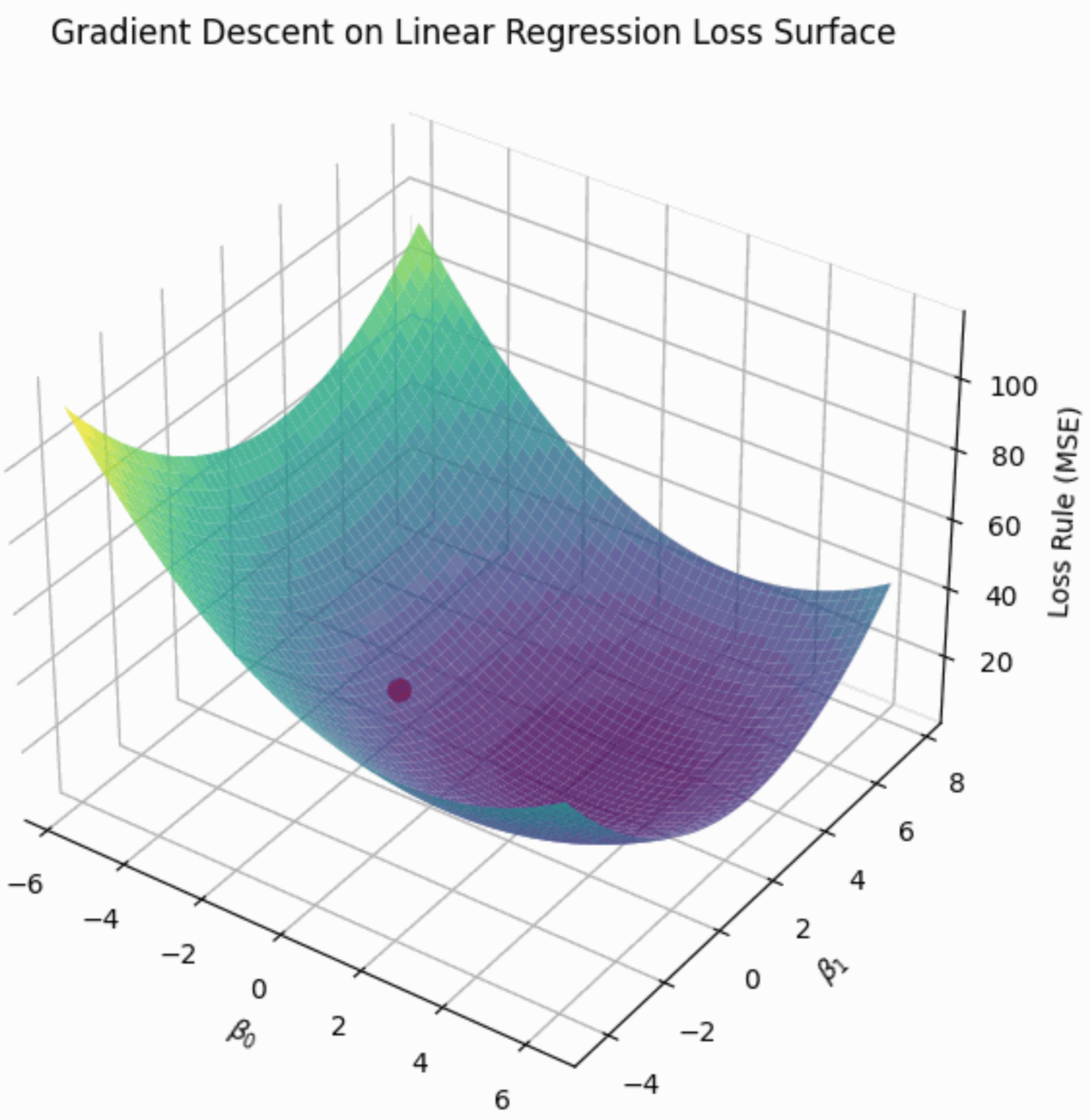
$$L(w) = \log(0.731)$$

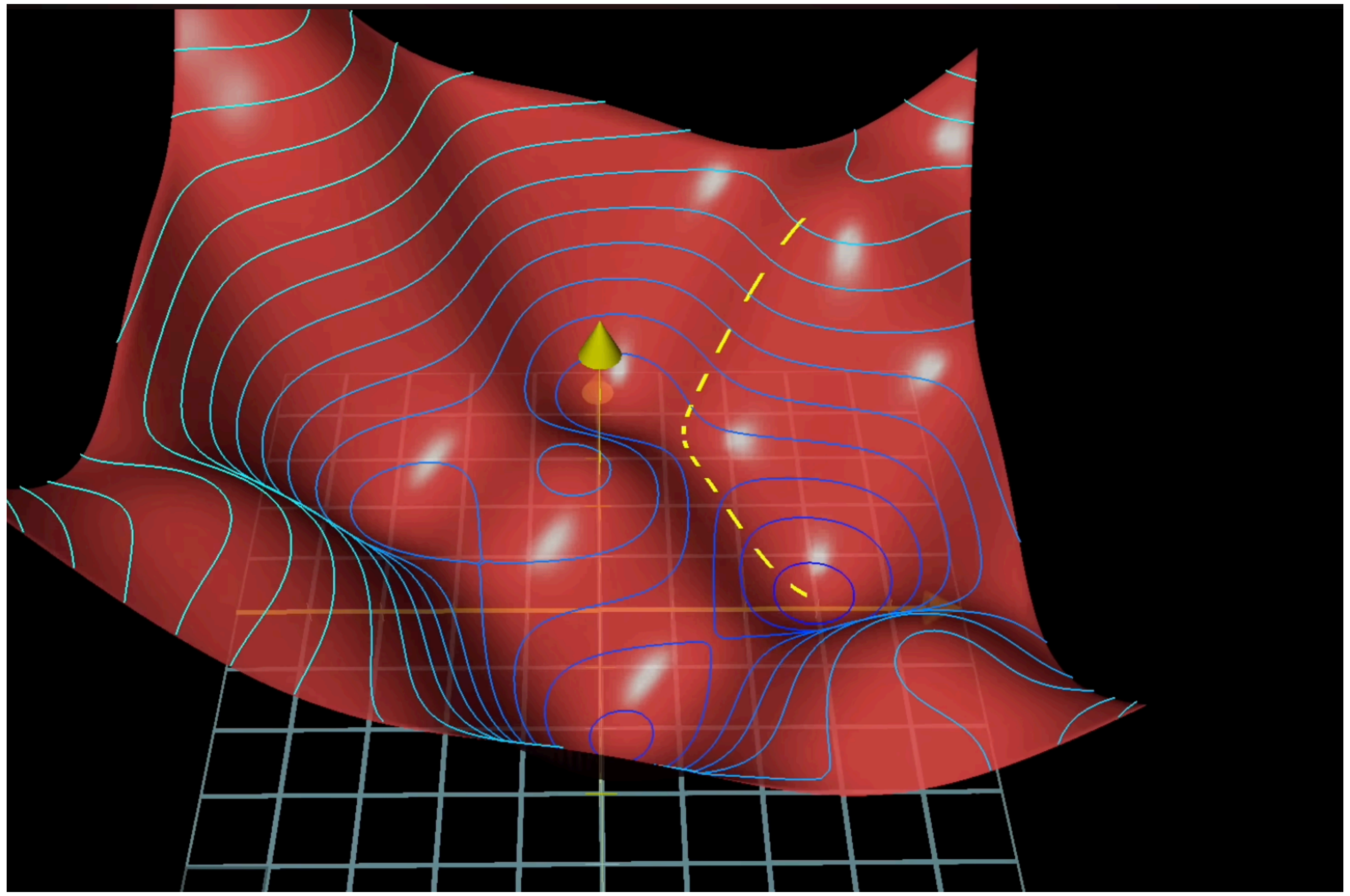
$$\sigma\left([0 \quad 0.4] \right) = [0.401 \quad 0.599]$$

$$L(w) = \log(1 - 0.401)$$

Compute gradient, move, and repeat!

How does Deep Learning work?

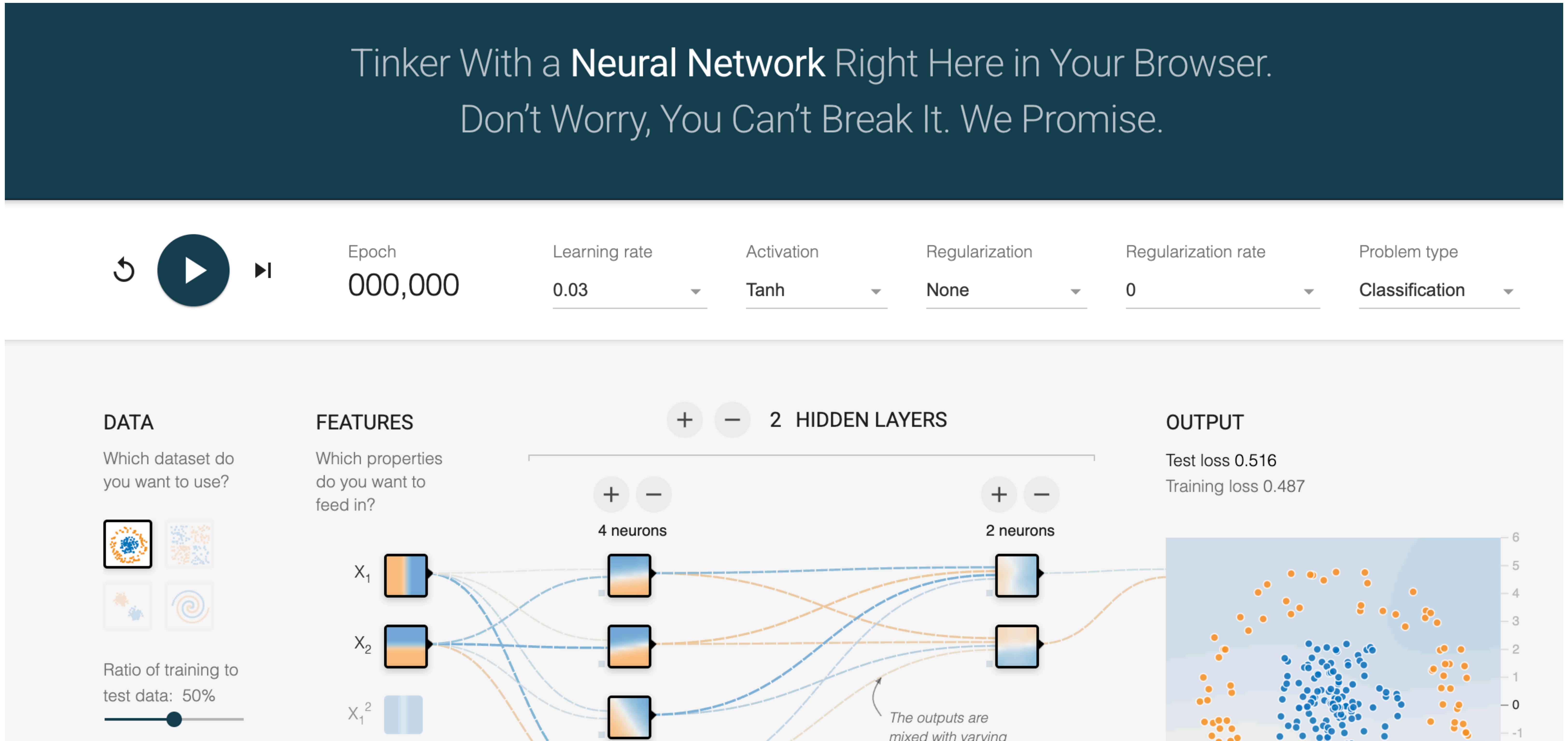




Training Our Own Neural Network

Visit

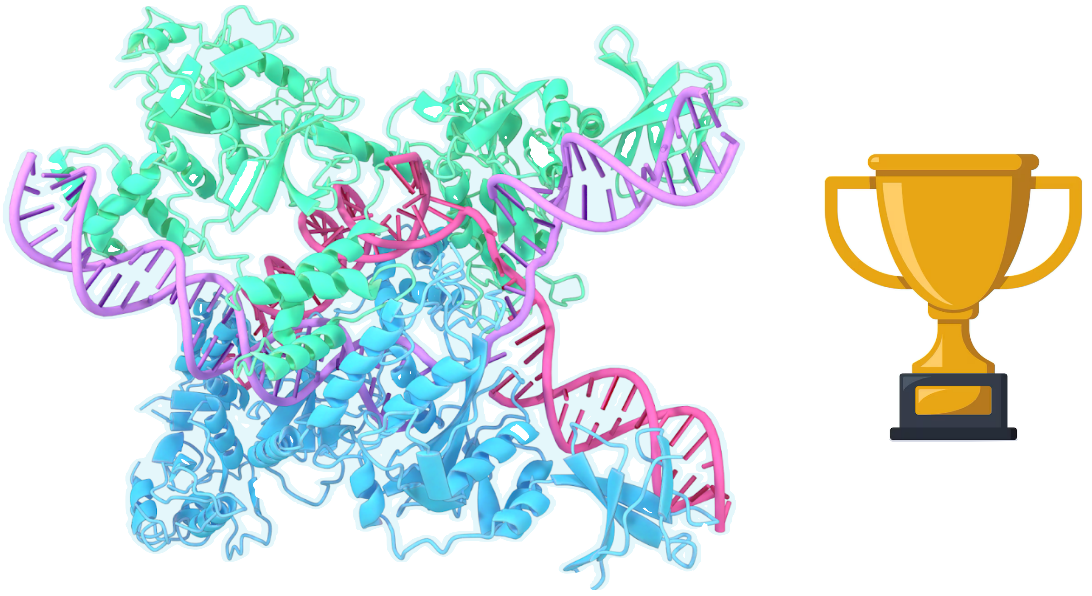
https://playground.tensorflow.org/



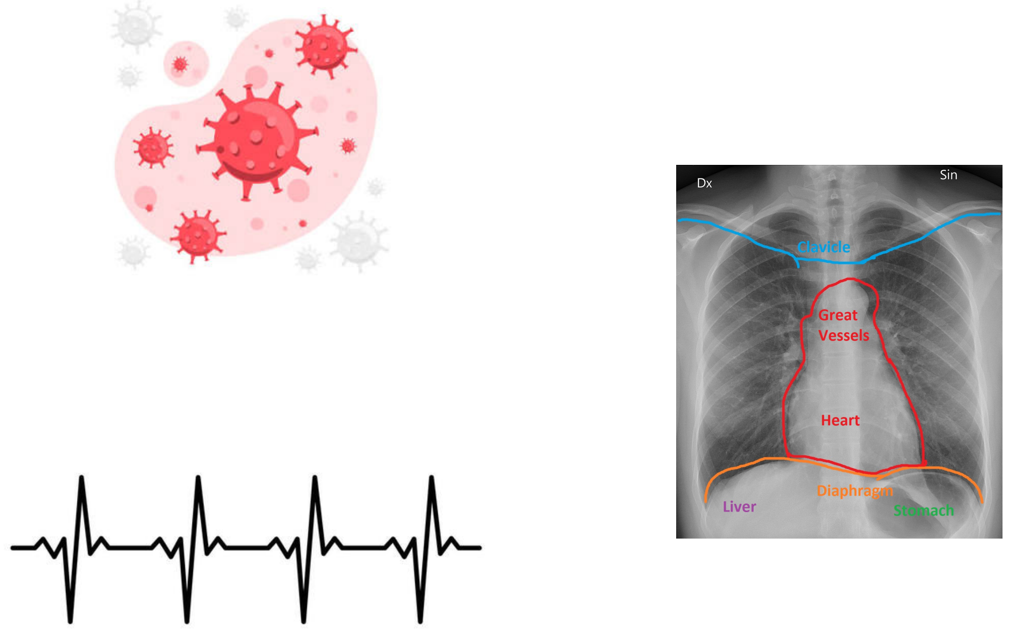
AI for Healthcare

Outset AI for Health Session

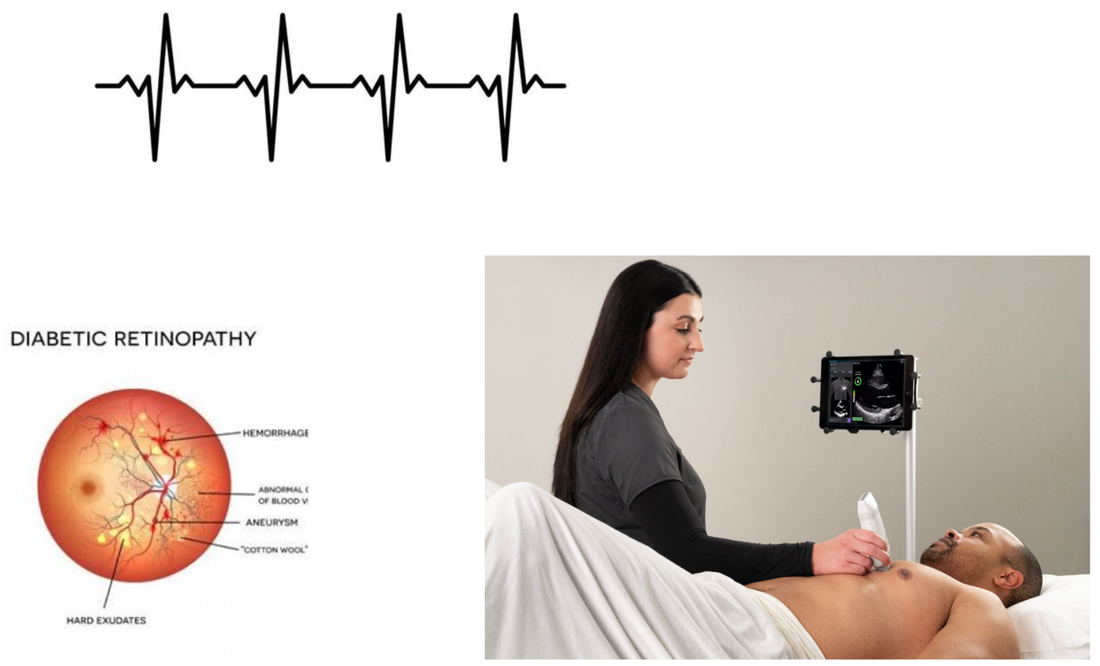
Franny Dean, July 8, 2026



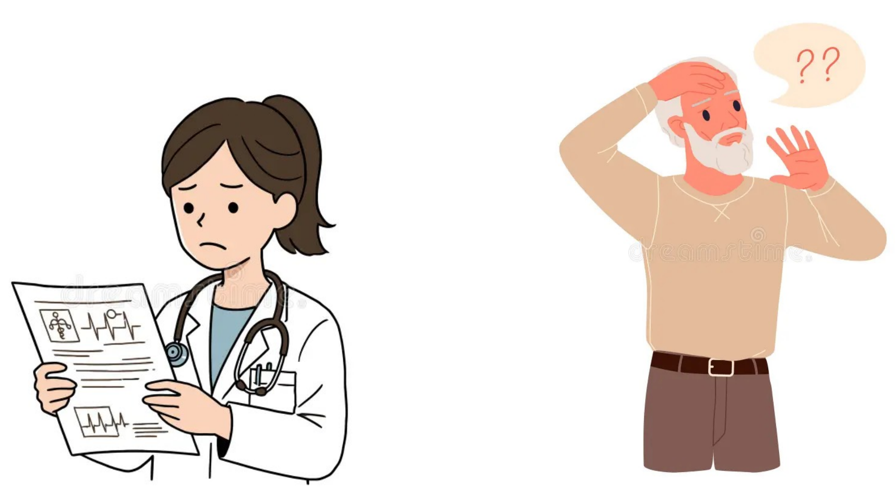
Drug Development



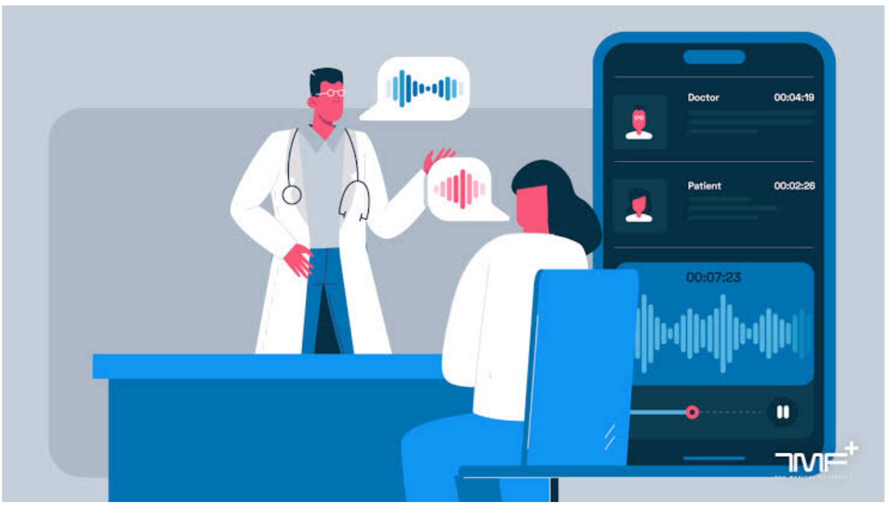
Disease Diagnosis



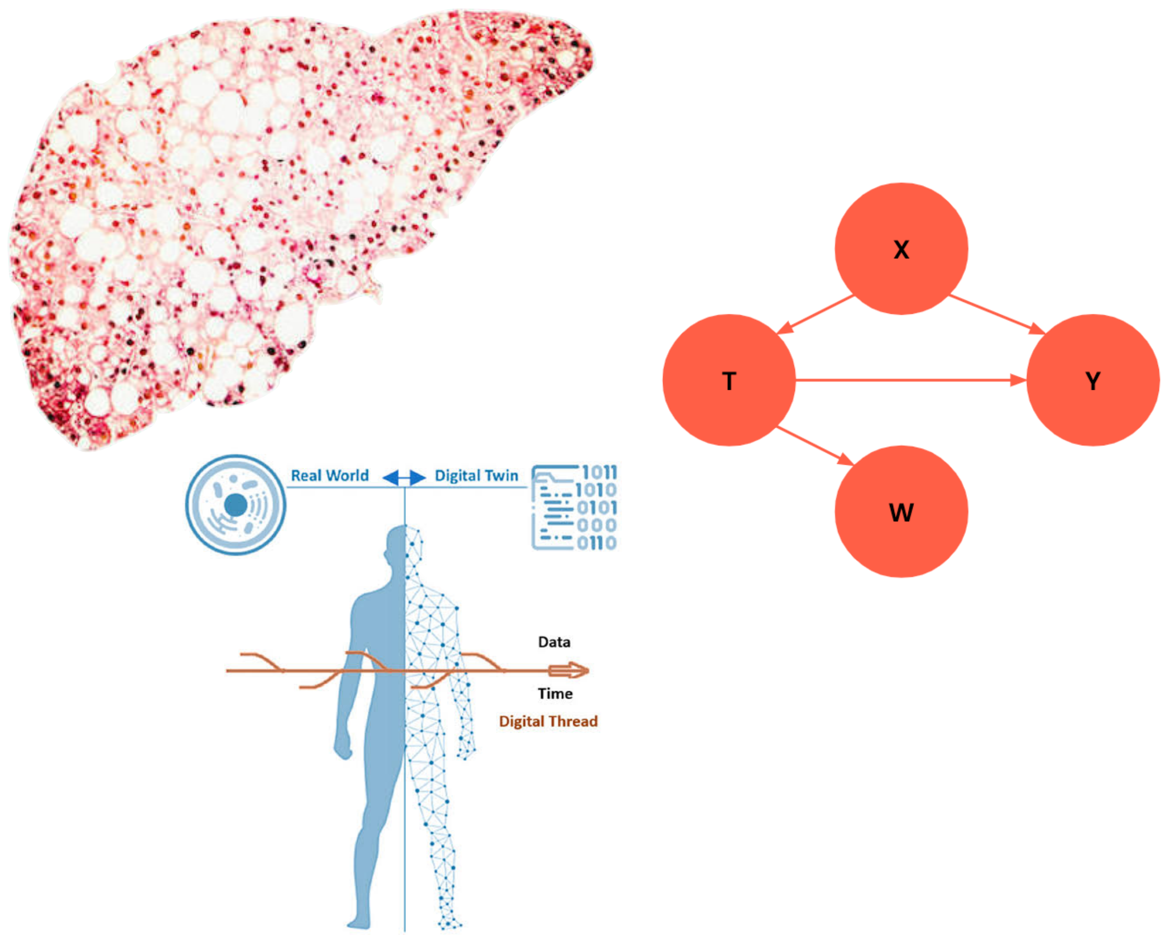
Low Resource Care



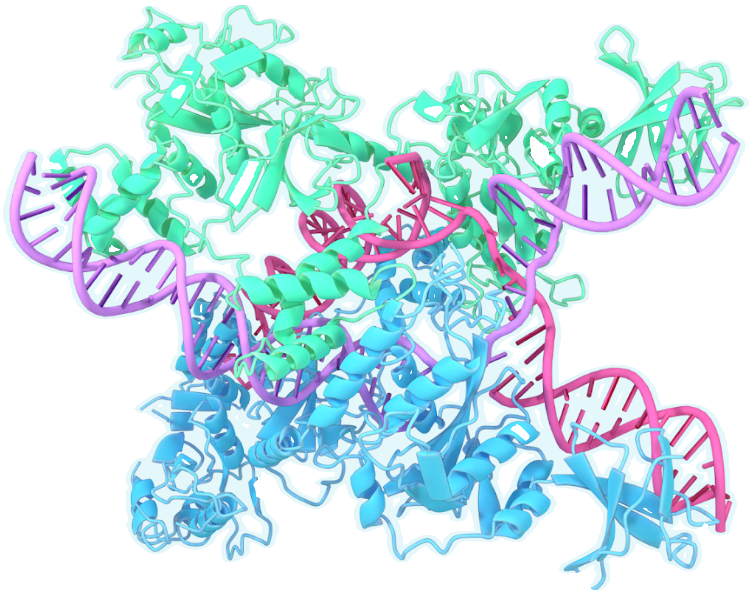
Information Retrieval



Clinical Documentation



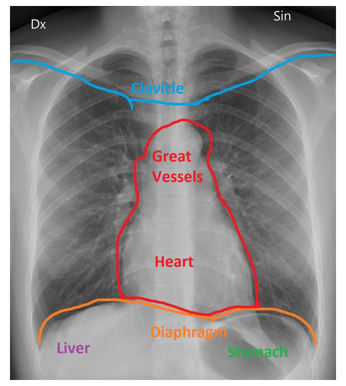
Clinical Trials




Drug Development

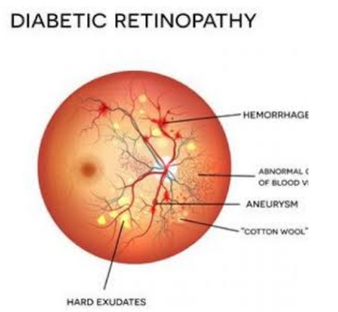
Ongoing, First Drugs





Disease Diagnosis

Early Products

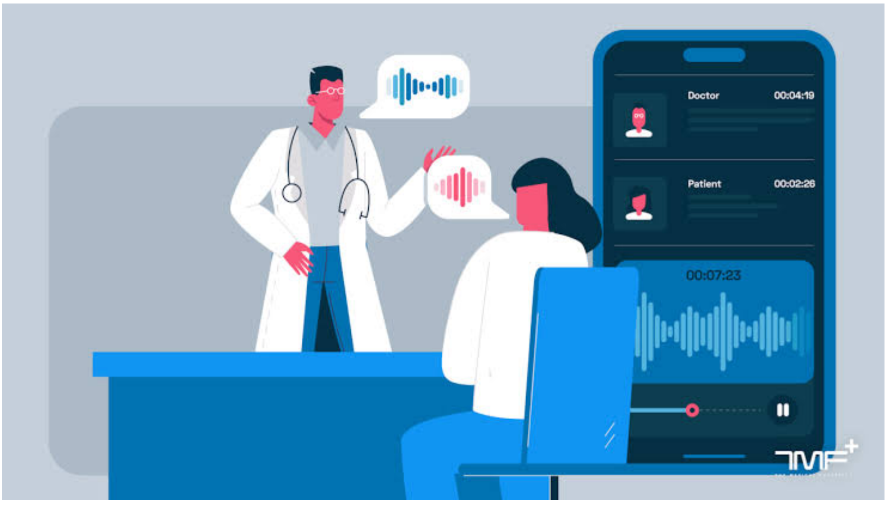



Low Resource Care

Some Studies/Tools

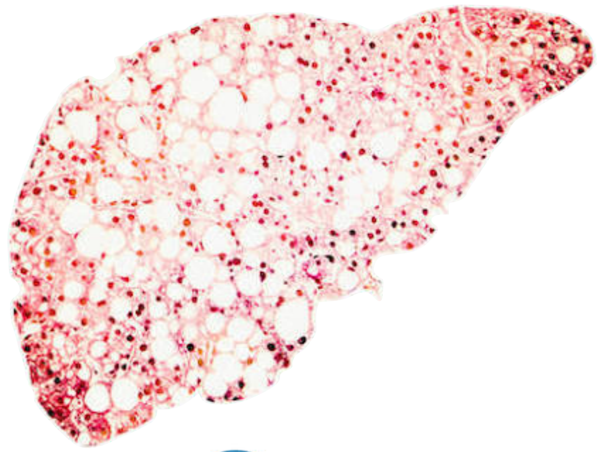
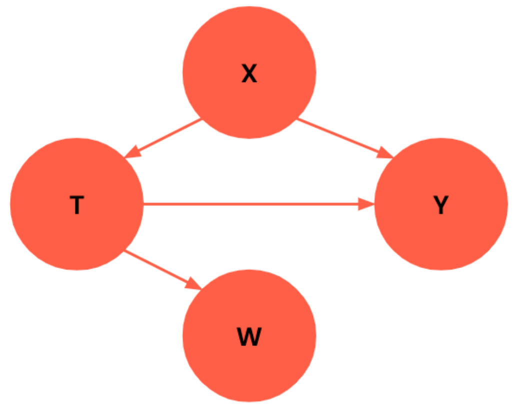

Information Retrieval

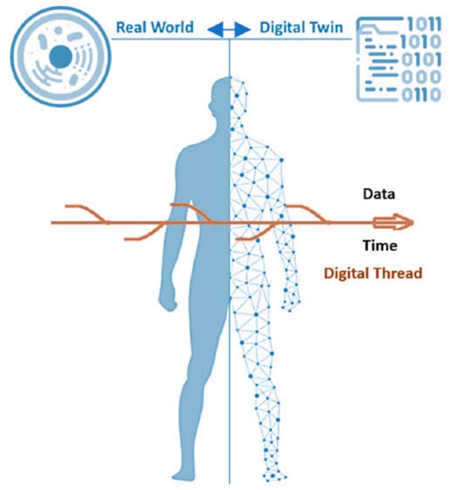
Active Products



Clinical Documentation

Active Products



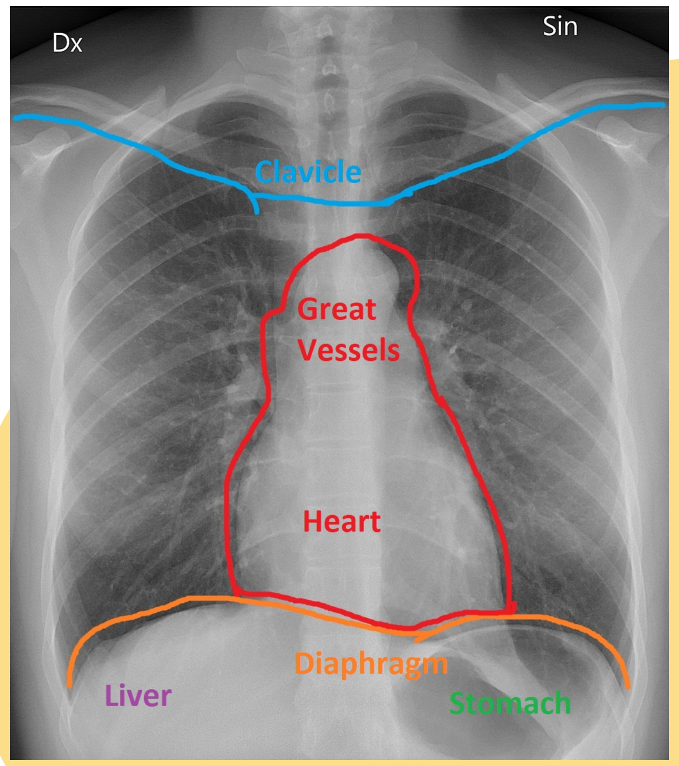
Clinical Trials

Research and Starting

Disease Diagnosis

BAYESIAN
HEALTH

Early Sepsis Diagnosis

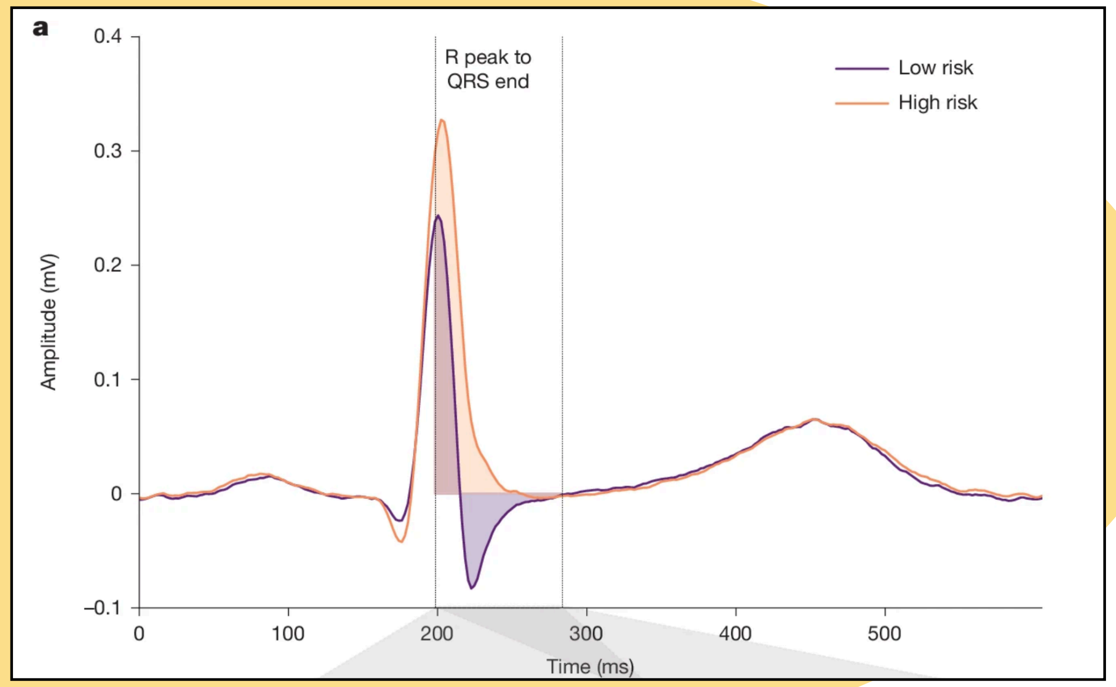


pneumonia

qure.ai

Qure.ai receives 510(k) FDA clearance for qXR, under two critical findings.

Automating Chest X-Ray Reading

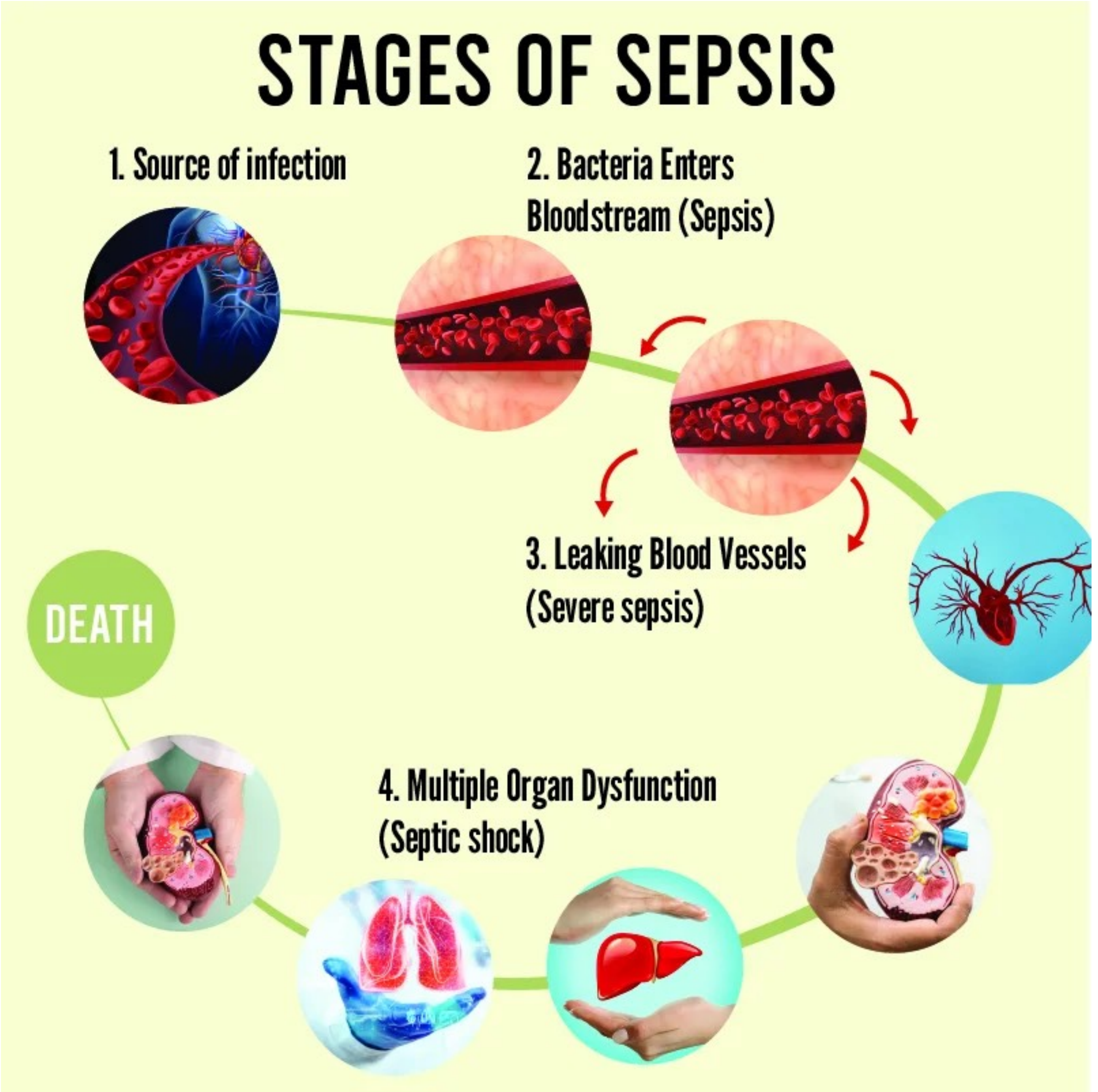


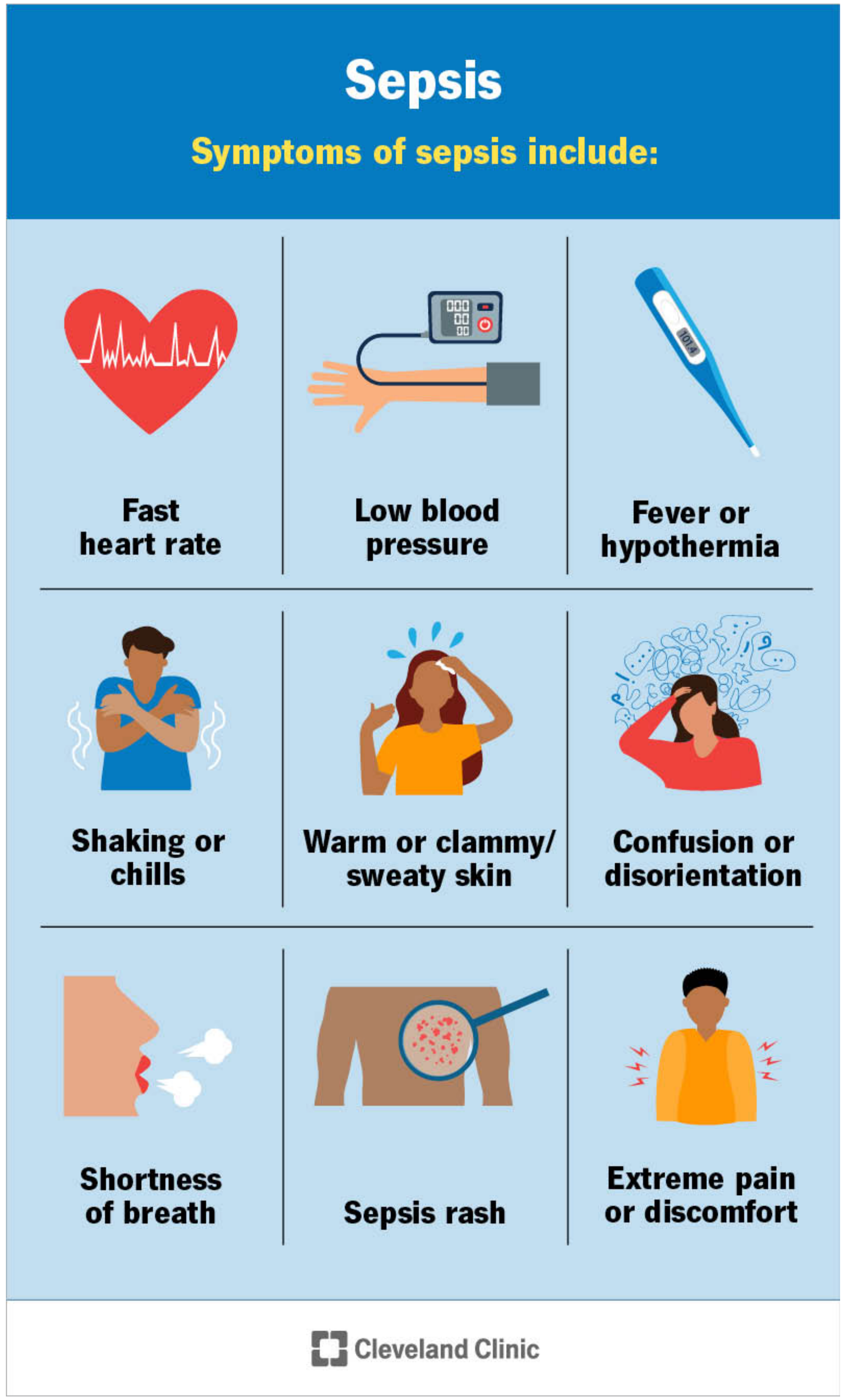
Learning New ECG Signals

And many many more...

Early Sepsis Diagnosis

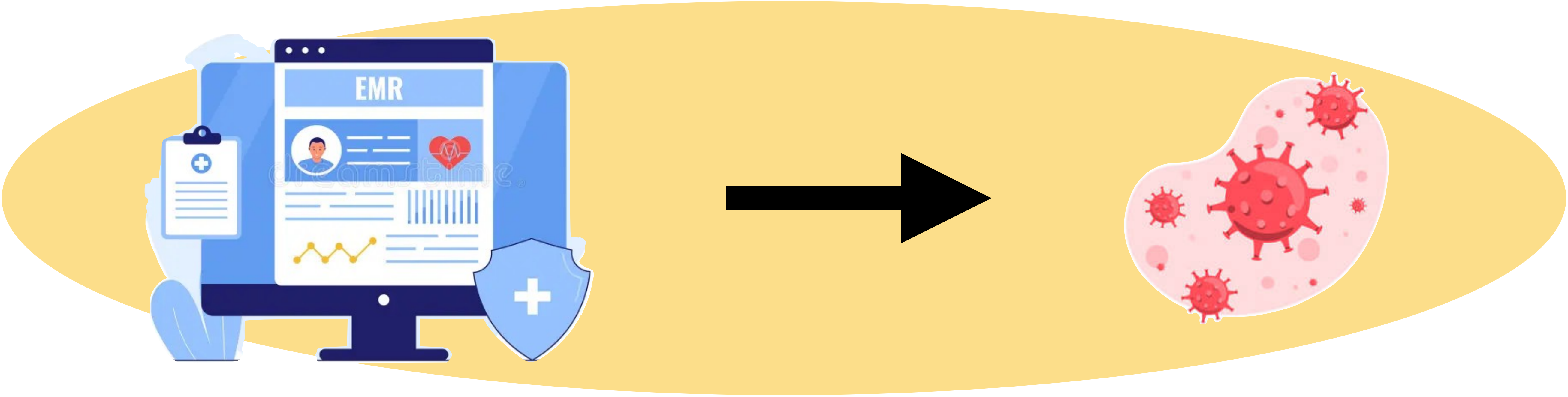
What is sepsis?





"…1.7 million adults develop sepsis every year in the United States, and more than 250,000 of them die. Sepsis is easy to miss because symptoms such as fever and confusion are also common in other conditions…"

Early Sepsis Diagnosis



How do we turn this into a machine learning model?

Early Sepsis Diagnosis

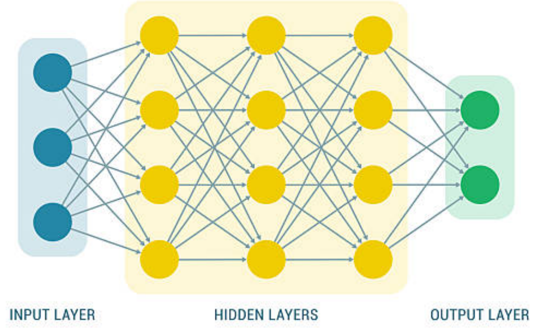
Ingredients we need:

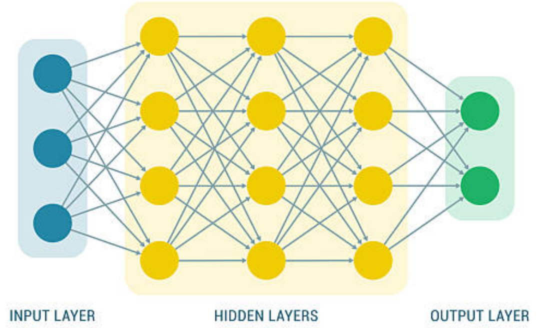
Examples (Data)

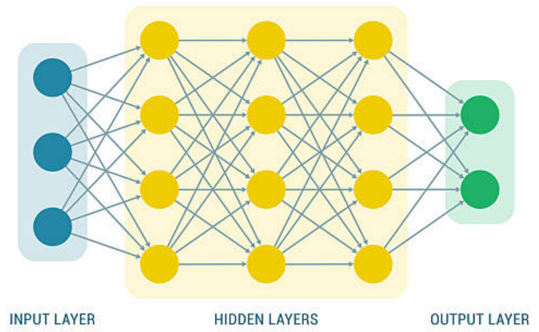
Past records

Labs over time

Model Structure



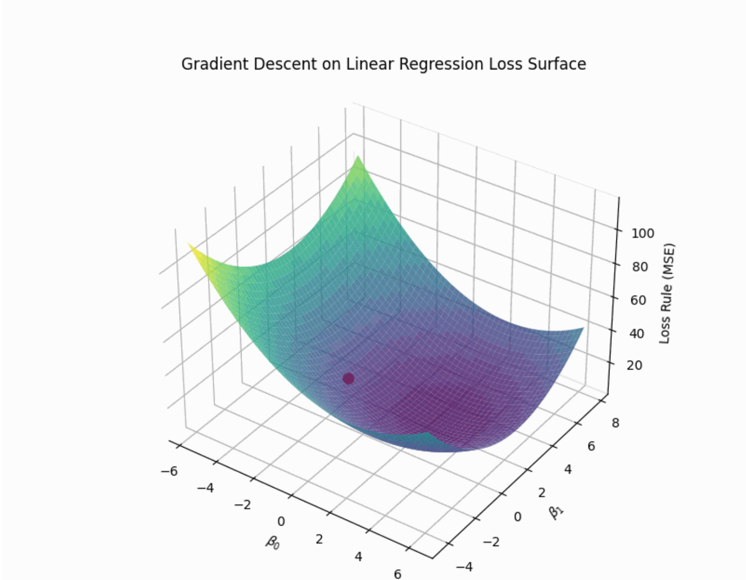




Loss Rule

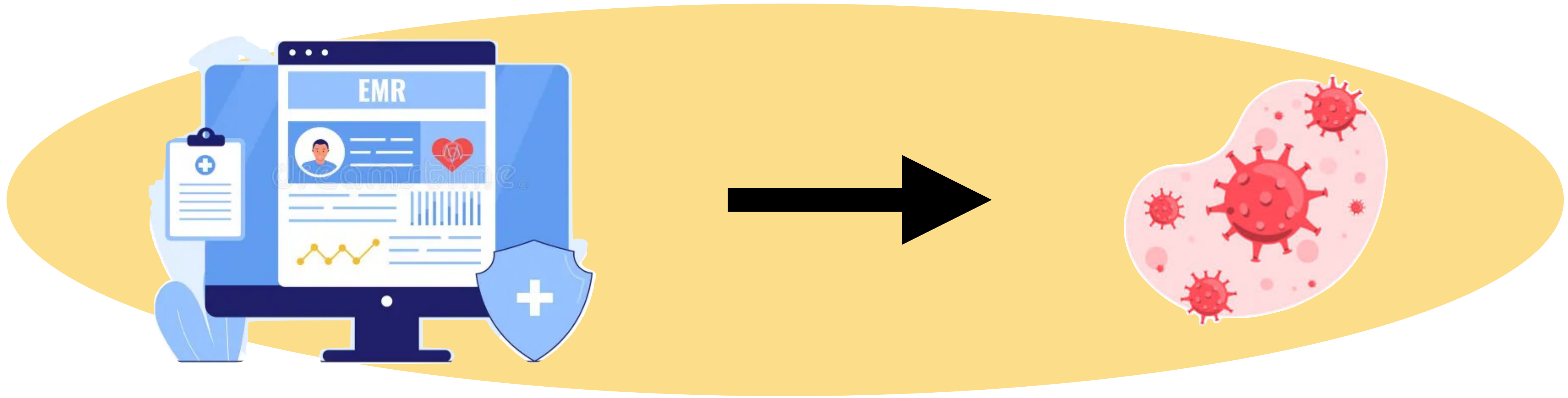
$$L(w) = \frac{1}{N}\sum_i (1 - y^i_{down}) \cdot (1 - \log(p_{down})) + y^i_{down} \cdot \log p_{down}$$

Learning procedure (Gradient Descent)



Early Sepsis Diagnosis

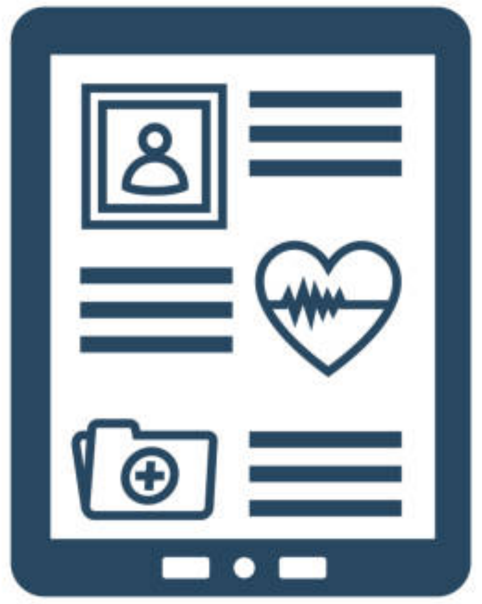
JOHNS HOPKINS UNIVERSITY



BAYESIAN HEALTH

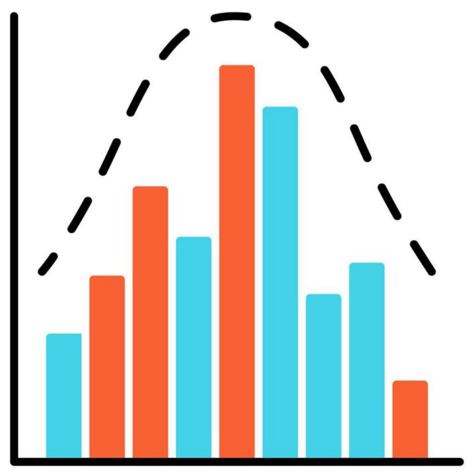
Predictors

Outcome



\+

Sepsis risk score

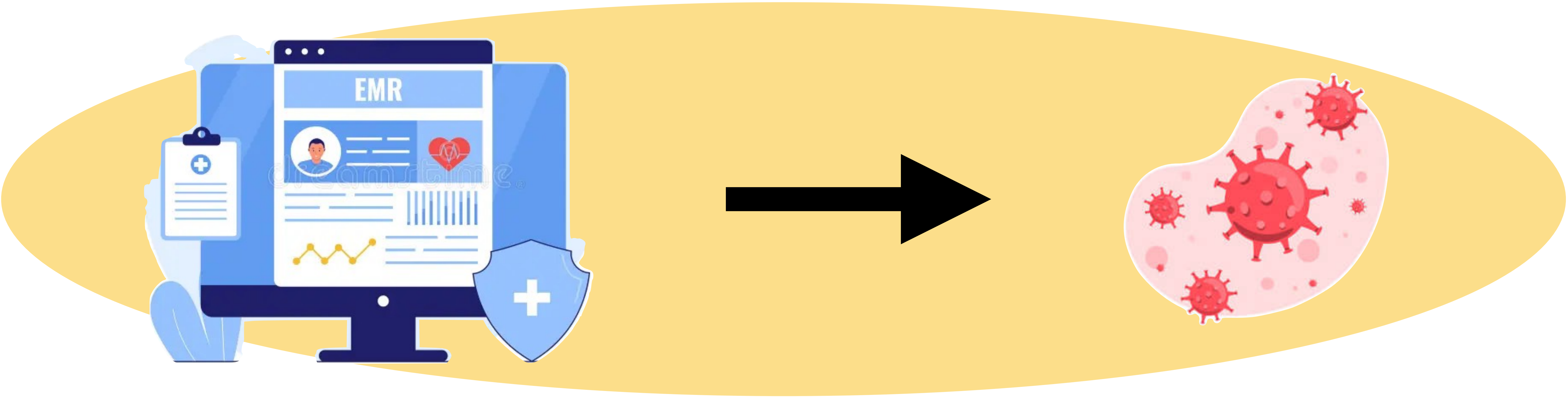


Past records

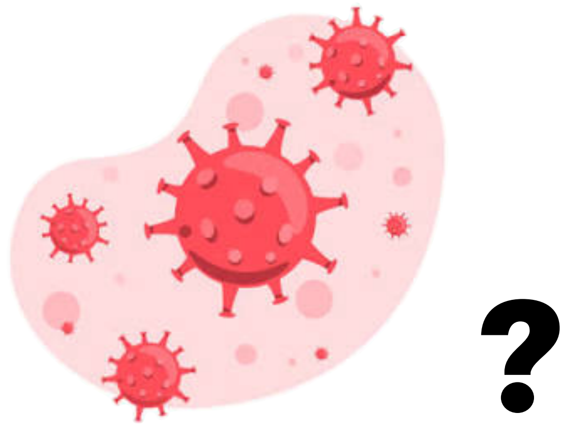
Time series lab modeling

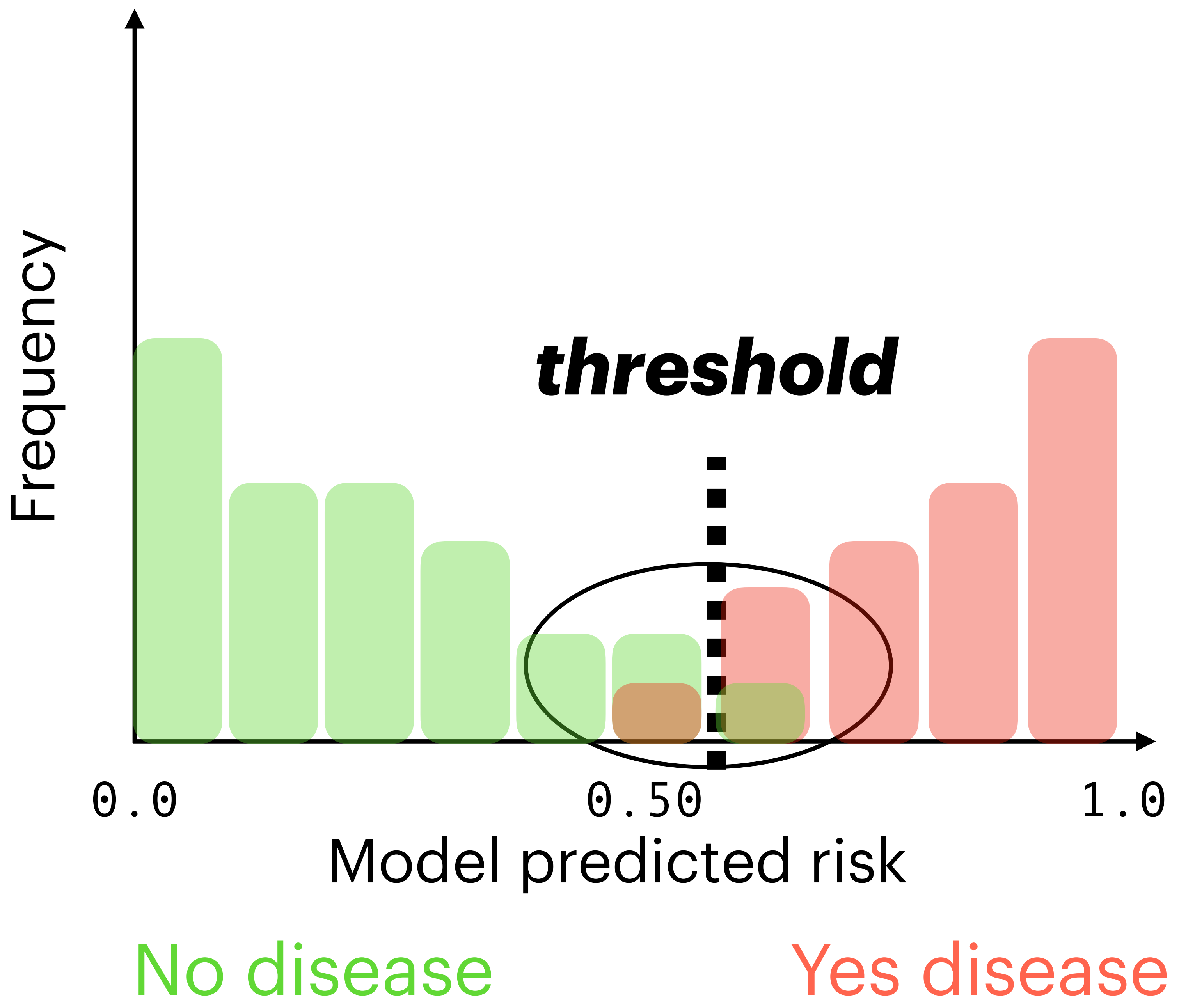
Representation — Not real interface.

How do we assess how accurate AI for sepsis is?



But...





We choose a ***threshold*** to say likely has or does not have.

Then we count how many positives and negatives we got wrong each.

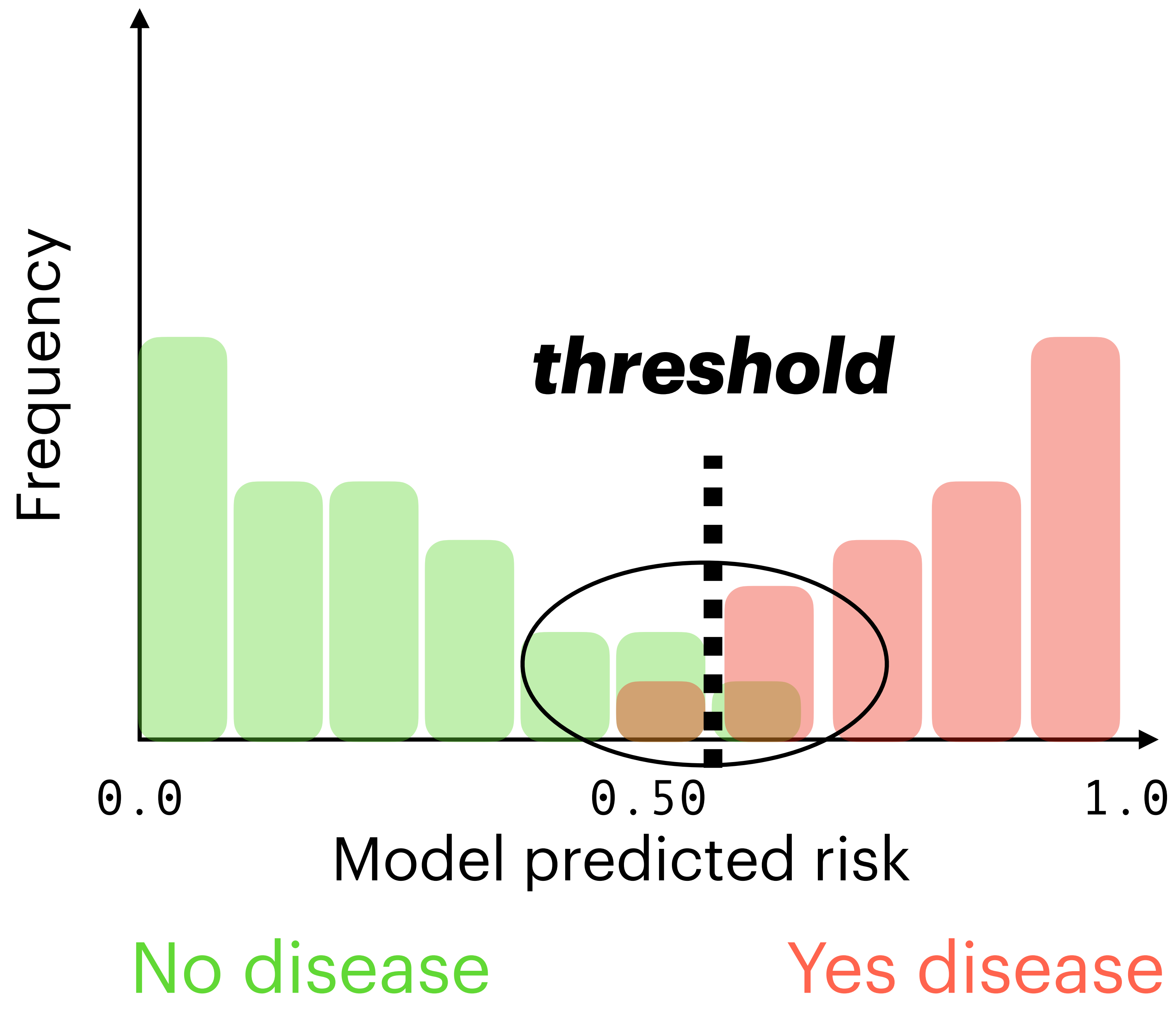
False positive rate (FPR) = number of negatives we got wrong

Specificity or **True negative rate** = 1-FPR

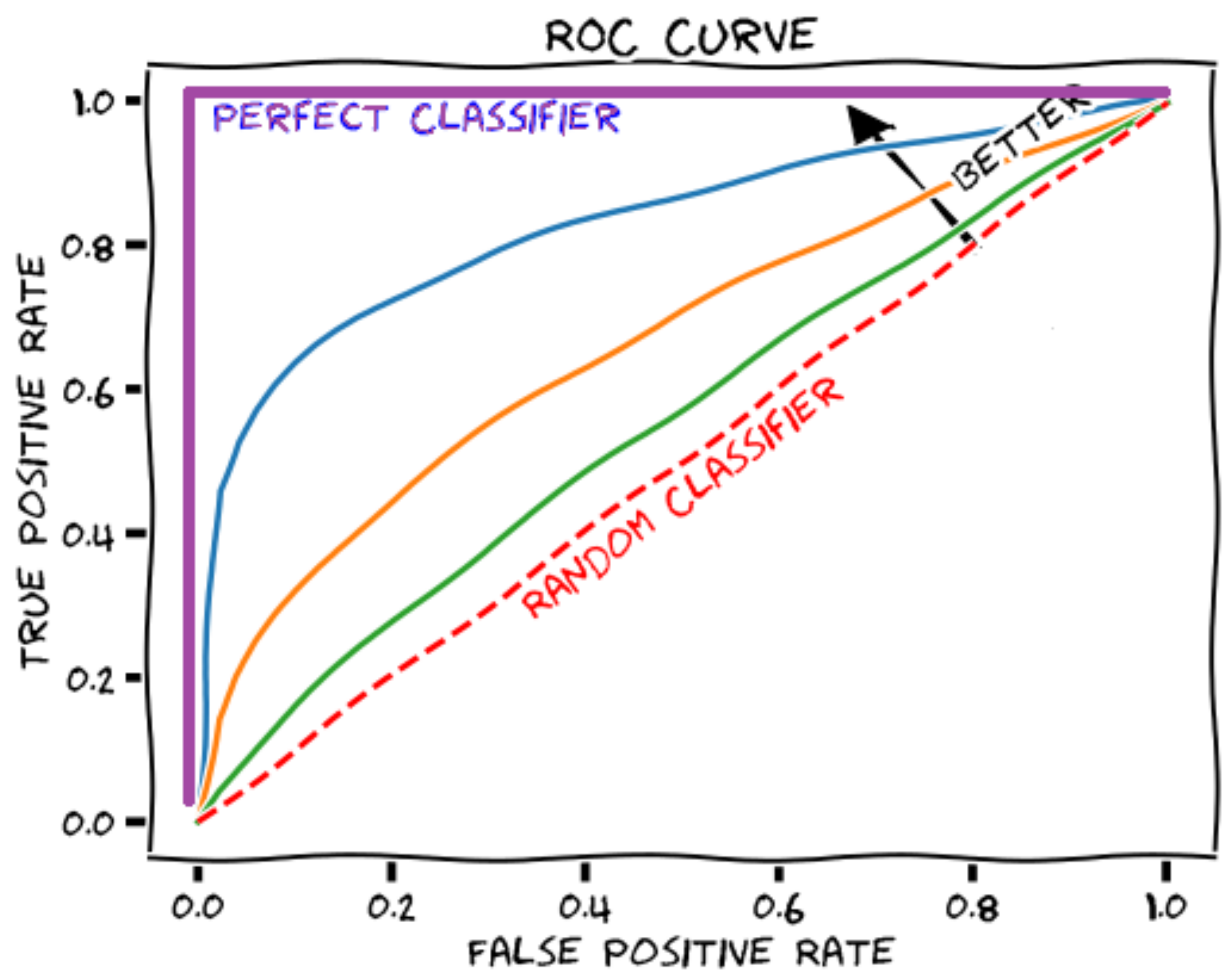
False negative rate (FNR) = number of positives we got wrong

Sensitivity or **True positive rate** = 1-FNR

How do we assess how accurate these tools are?



We compare all posible thresholds using the metric called AUROC/AUC which is the area under the following curve:

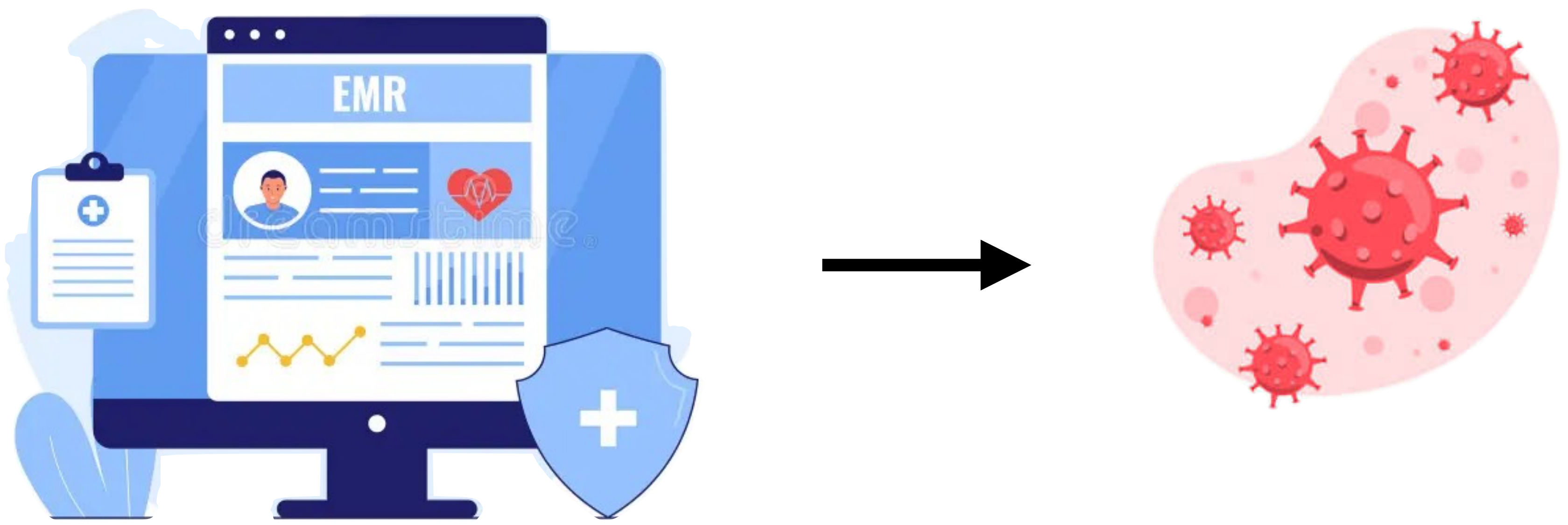


Good AUC: 0.70+

Great AUC: 0.80+

Accuracy in one hospital setting does not necessarily mean accuracy in another.

Epic



Epic deployed a model without evidence that it worked.

University of Michigan researchers showed that the algorithm missed 67% of sepsis patients and 88% of its alerts were false alarms.

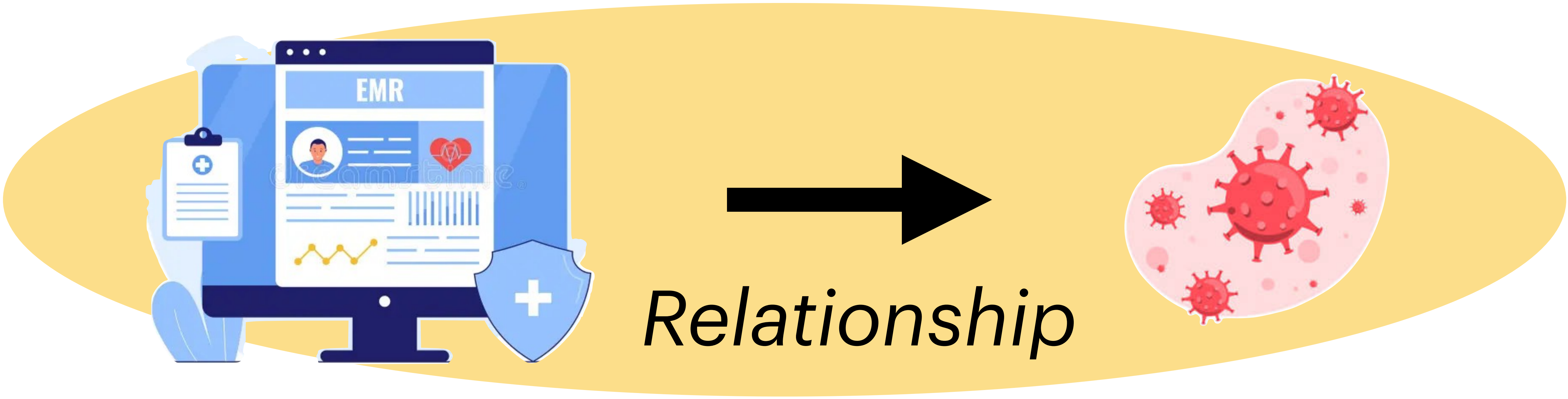
False positive rate (FPR) = 88%

False negative rate (FNR) = 67%

Why?

Distribution shift, lack of external validation, poor model.

What does data distribution shift mean?



Example 3: We get a new monitoring device that gives different frequency reading.

Other examples?

Data distribution has shifted!

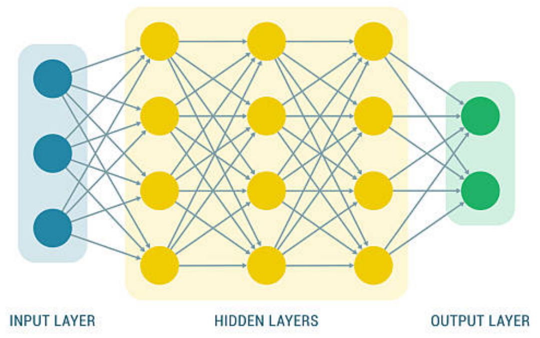
Healthcare AI Ingredients

Ingredients we need in the applied setting:

Understanding and defining the task

Examples (Data)

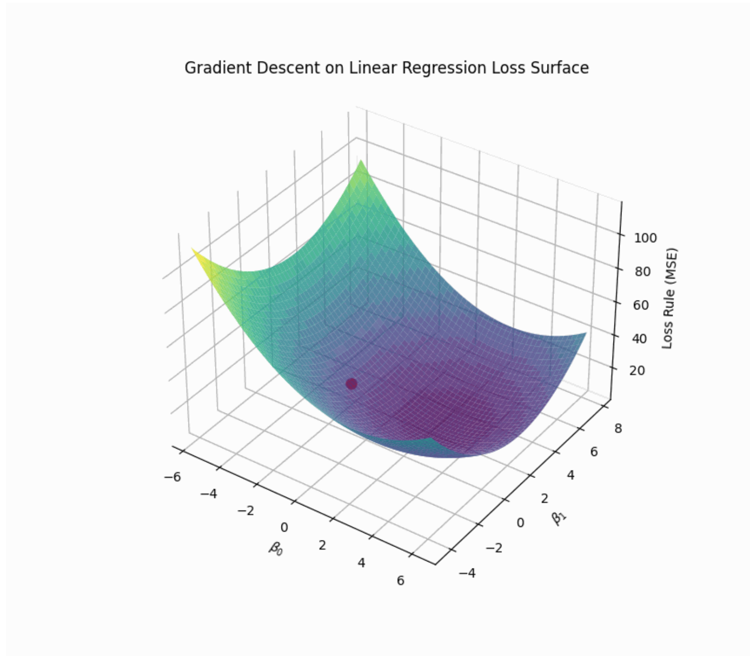
Model Structure



Loss Rule

$$L(w) = \frac{1}{N}\sum_{i}(1 - y^{i}_{down}) \cdot (1 - \log(p_{down})) + y^{i}_{down} \cdot \log p_{down}$$

Learning procedure (Gradient Descent)



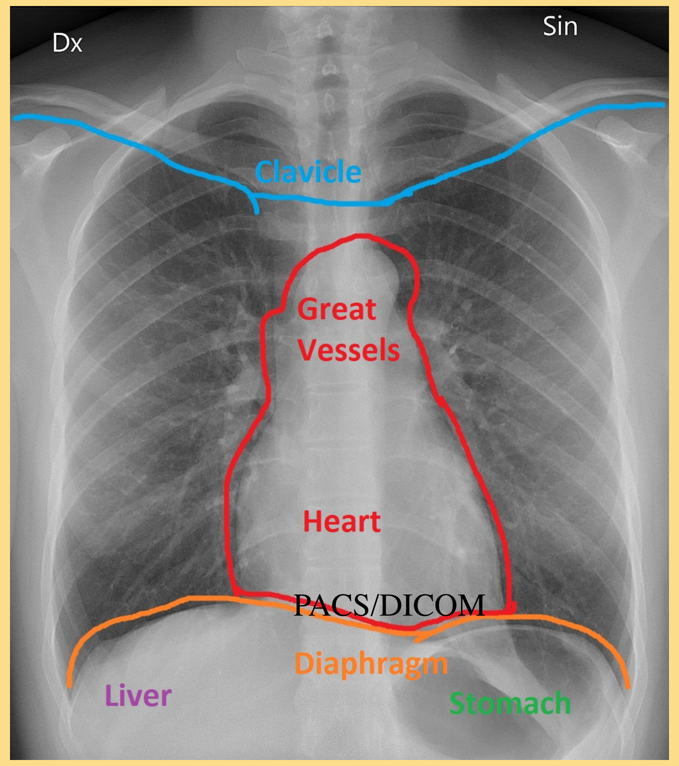
Evaluating accuracy

Implementing into a workflow

Automating Chest X-Ray Reading

Most advanced diagnostic area?

Multi-disease foundation models



→ pneumonia

→ lung cancer

→ clear

qure.ai

azmed

annalise.ai

Qure.ai receives 510(k) FDA clearance for qXR, under two critical findings.

1000+ FDA radiology approvals to date

aidoc Always On AI

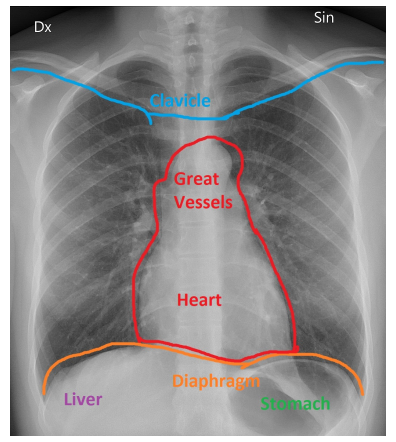
Cognita

Healthcare AI Ingredients

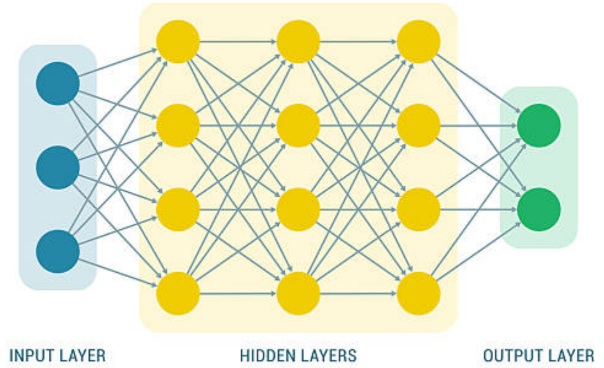
Ingredients we need in the applied setting:

Understanding and defining the task

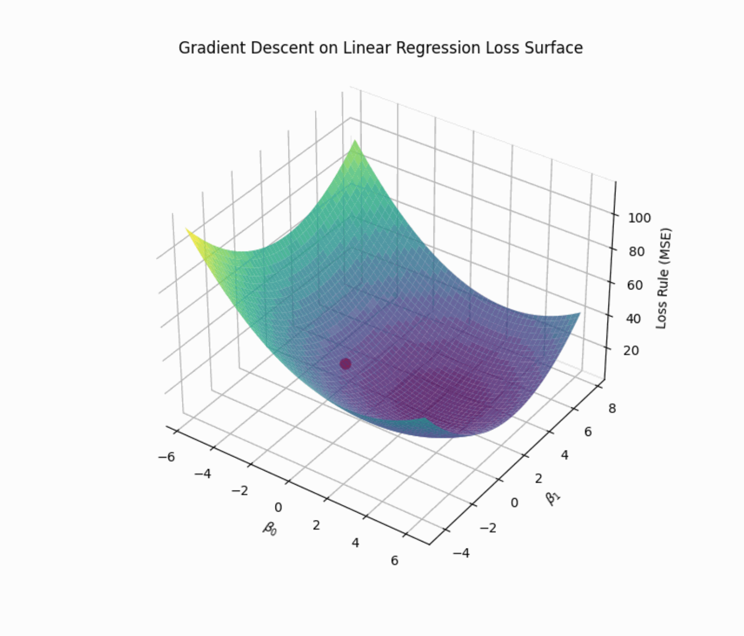
Examples (Data)



Model Structure



Loss Rule $L(w) = \frac{1}{N}\sum_i (1 - y^i_{down}) \cdot (1 - \log(p_{down})) + y^i_{down} \cdot \log p_{down}$



Learning procedure (Gradient Descent)

Evaluating accuracy

Repeat

Repeat

Implementing into a workflow

How did chest x-rays AI lead the charge?

Most advanced diagnostic area?

Super common diagnostic imaging used globally

Advancements in computer vision

Large public datasets

MIMIC-IV

PhysioNet

Share About FAQs Explore

Database | Credentialed Access

MIMIC-CXR-JPG - chest radiographs with structured labels

Alistair Johnson, Matt Lungren, Yifan Peng, Zhiyong Lu, Roger Mark, Seth Berkowitz, Steven Horng

Published: Nov. 14, 2019. Version: 2.0.0 <View latest version>

Beth Israel Deaconess Medical Center

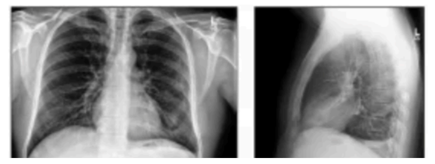
CheXpert: Chest X-rays

Download here ↗

When citing this dataset in your research or other publications, please reference the following DOI: https://doi.org/10.71718/y7pj-4v93 ↗. Proper attribution ensures the continued accessibility and credibility of the dataset for the scientific community.

Dataset Description

CheXpert is a dataset consisting of 224,316 chest radiographs of 65,240 patients who underwent a radiographic examination from Stanford Health Care between October 2002 and July 2017, in both inpatient and outpatient



NIH

kaggle

Create

Home

Competitions

Benchmarks

Search

Sign In

NATIONAL INSTITUTES OF HEALTH CHEST X-RAY DATASET AND 1 COLLABORATOR · UPDATED 8 YEARS AGO

1624 Code Down

NIH Chest X-rays

Over 112,000 Chest X-ray images from more than 30,000 unique patients

Key Takeaway #1:

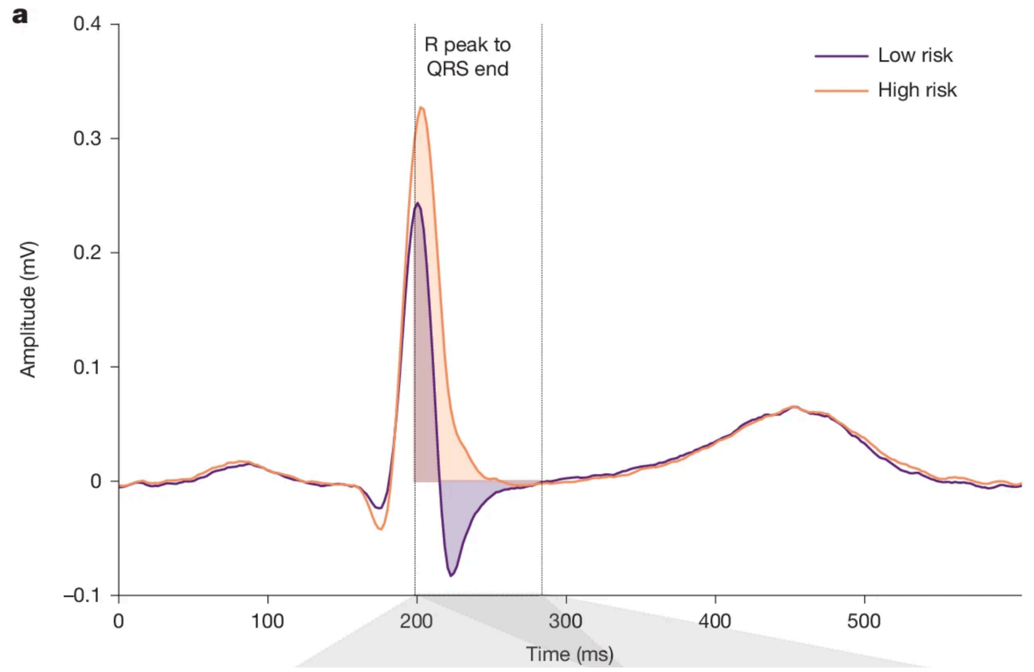
Having data from mulitple institutions is critical to build accurate robust AI for healthcare!

Electrocardiography and AI

New company predicting heart failure signals.

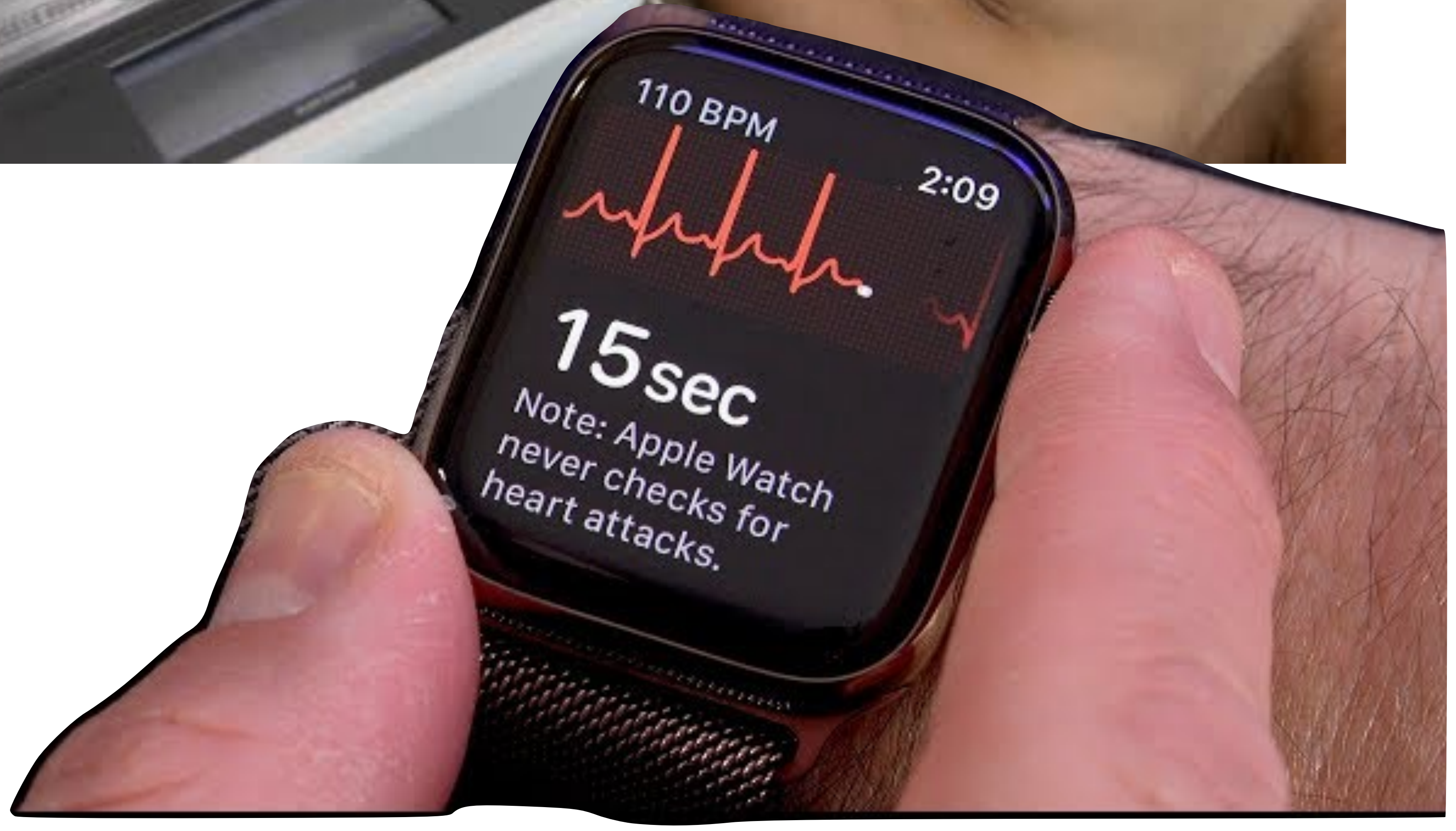
Pathway Labs

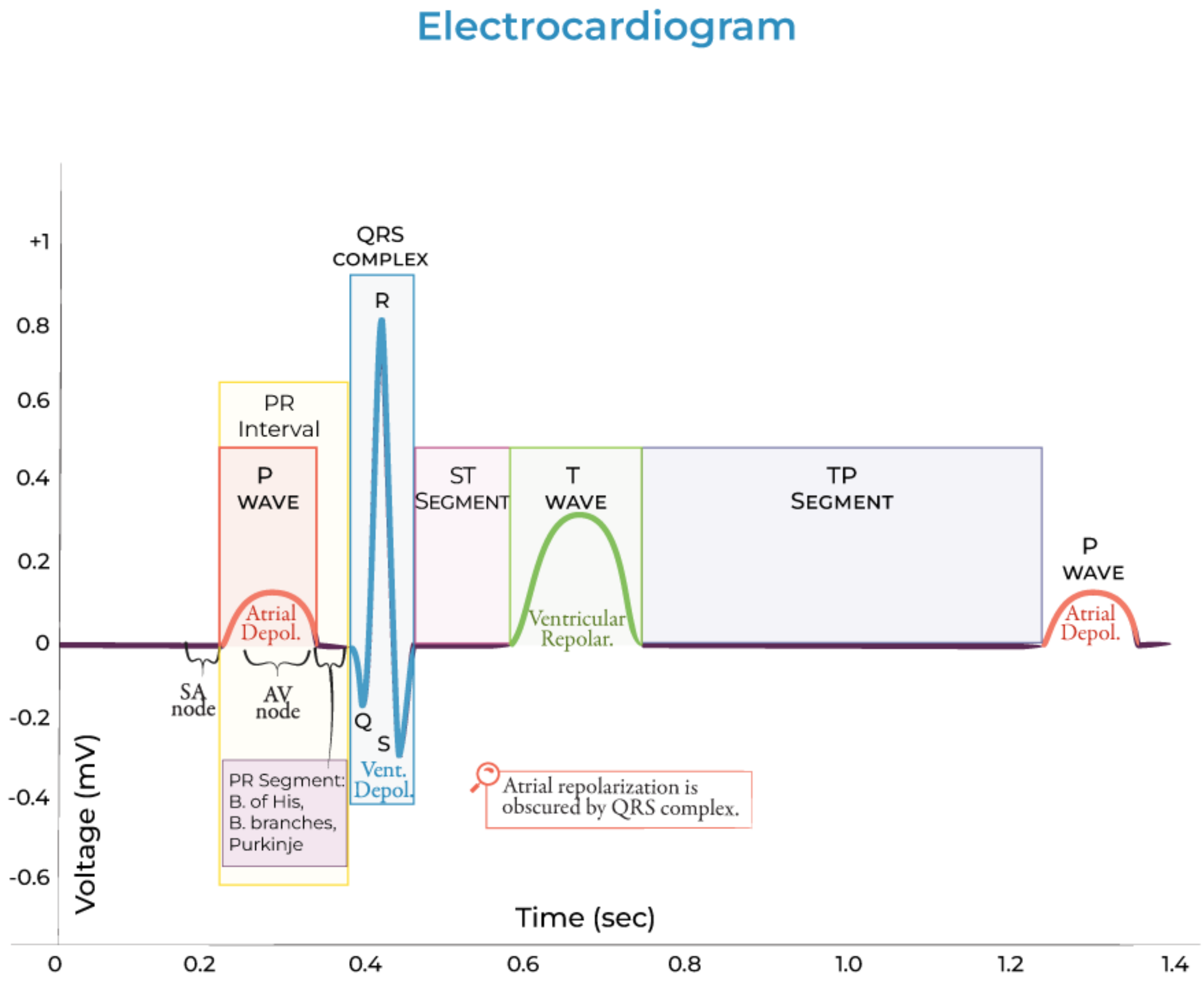
New research finding a visual signal of sudden cardiac death.



Why ECGs?

What's an ECG?





Cheap, non-invasive

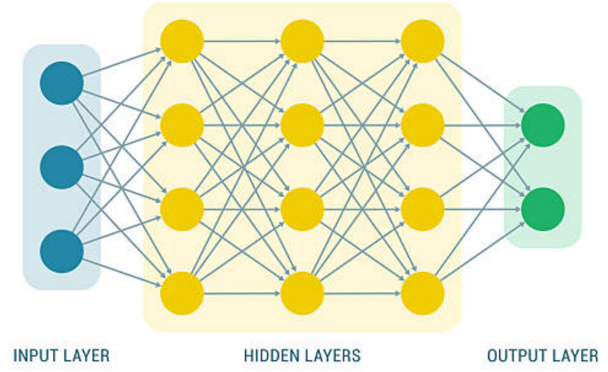
Healthcare AI Ingredients

What ingredients of the process are ECGs good for?

Understanding and defining the task

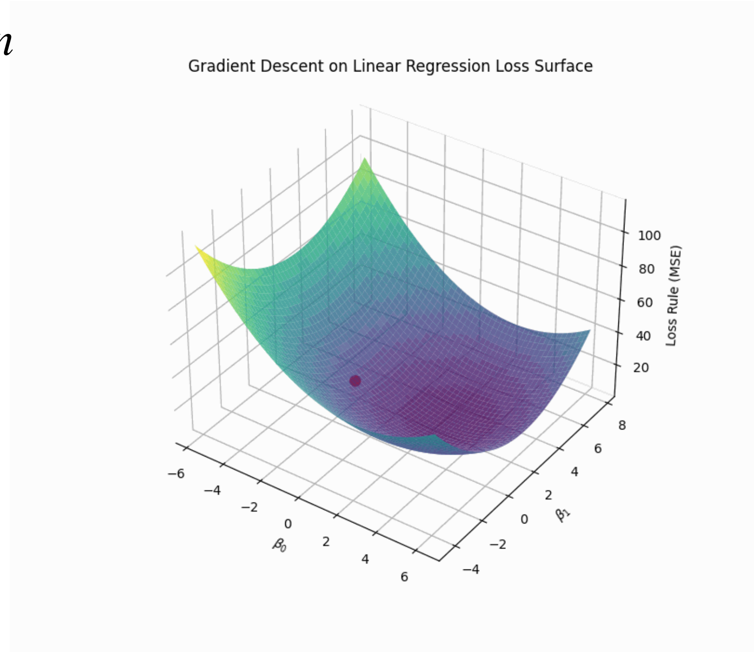
Examples (Data)

Model Structure



Loss Rule $L(w) = \frac{1}{N}\sum_{i} (1 - y^{i}_{down}) \cdot (1 - \log(p_{down})) + y^{i}_{down} \cdot \log p_{down}$

Learning procedure (Gradient Descent)

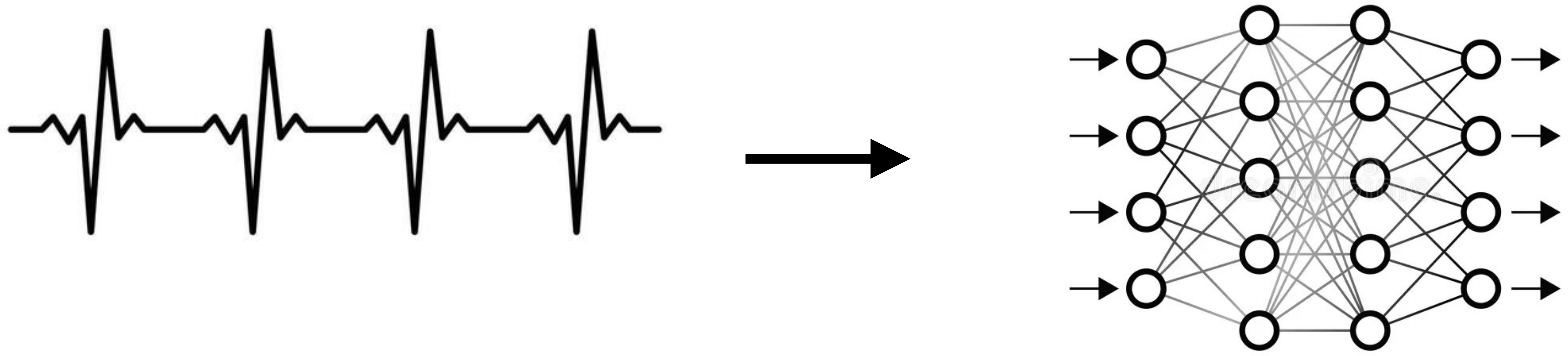


Evaluating accuracy

Implementing into a workflow

Learning New ECG Signals

There are many models predicting cardiovascular disease from ECGs.



Doctors can see

atrial fibrillation

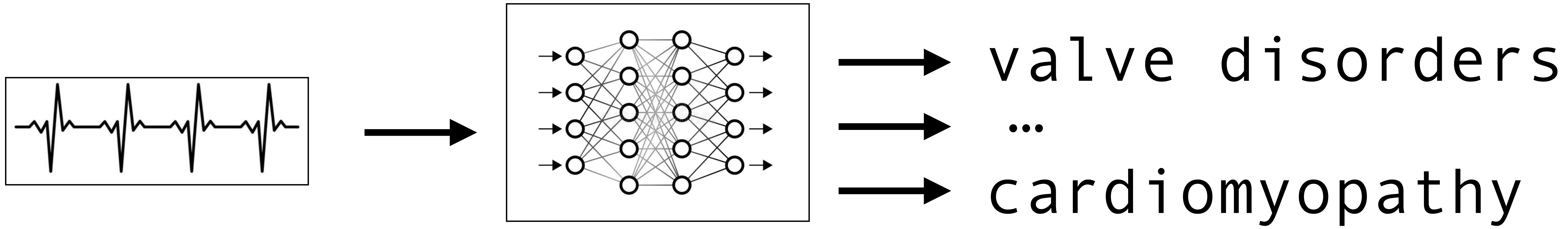
heart attack

heart failure

Doctors can't see

AI ECG models Predict Heart Failure

Trained a deep learning model on ECGs with high accuracy.



NewYork-Presbyterian

Columbia University

New non-invasive, cheap signal for structural heart diseases!

Launch of company!

Pathway Labs

Pathway to PathwayLabs

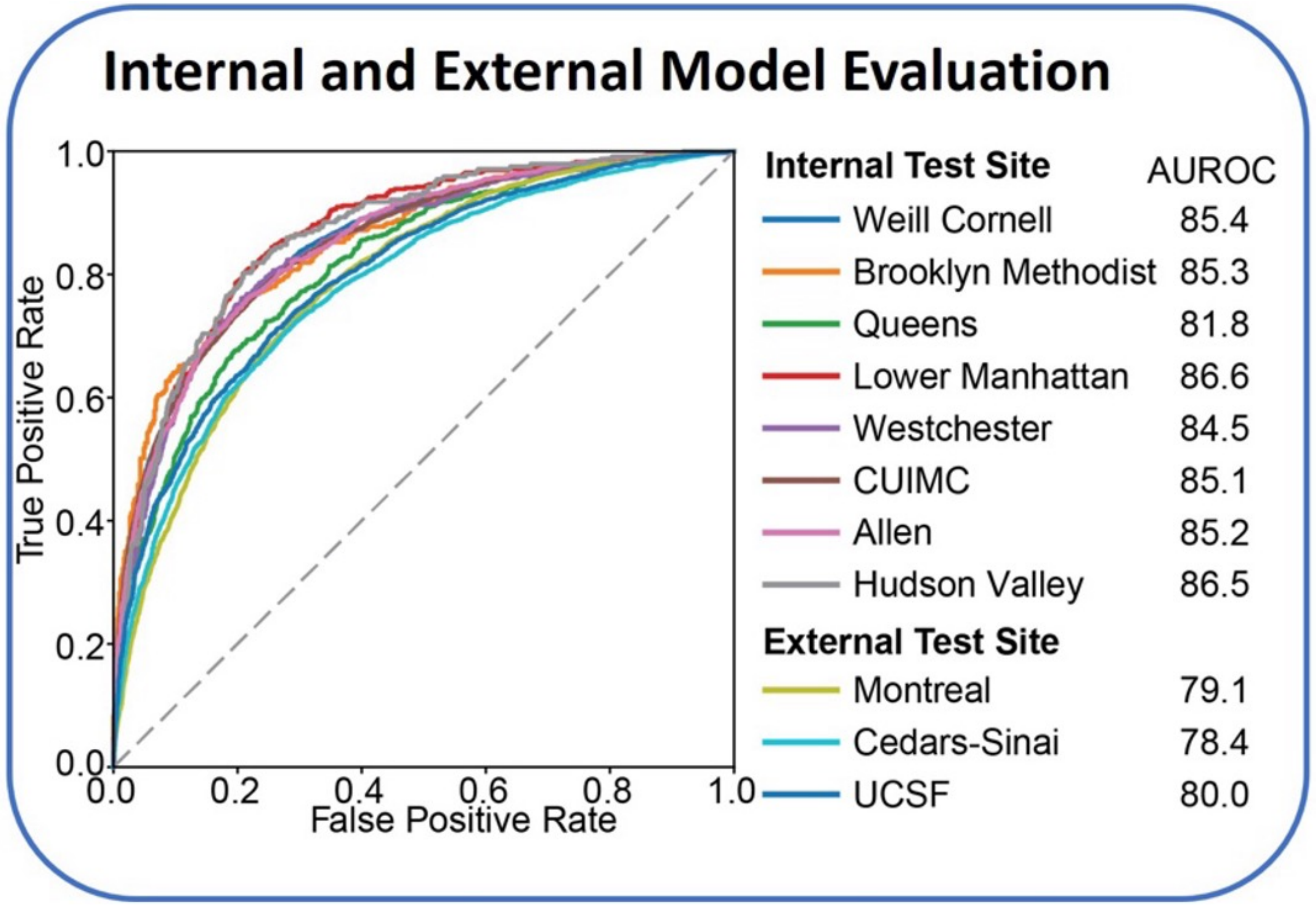
1. Train model

2. Validate model in past data

3. Validate model in past data from external sources

"How well does model predict past outcomes?"

Retrospective



Pathway to PathwayLabs

COLUMBIA UNIVERSITY

NewYork-Presbyterian

1. Train model
2. Validate model in past data
3. Validate model in past data from external sources
4. Prospective test of AI model

Without model

"How well does model predict future outcomes?"

With model

Pathway to PathwayLabs

NewYork-Presbyterian

Columbia University

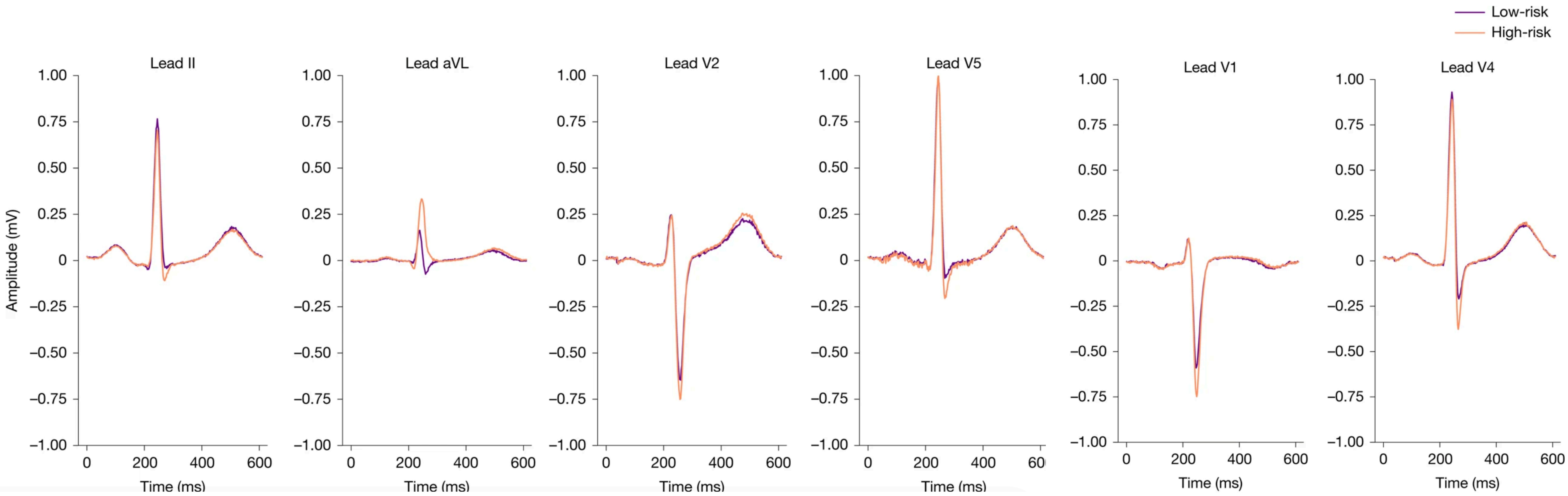
1. Train model
2. Validate model in past data
3. Validate model in past data from external sources
4. Prospective test of AI model
5. FDA approval

AI is being regulated as a medical device

FDA U.S. FOOD & DRUG ADMINISTRATION

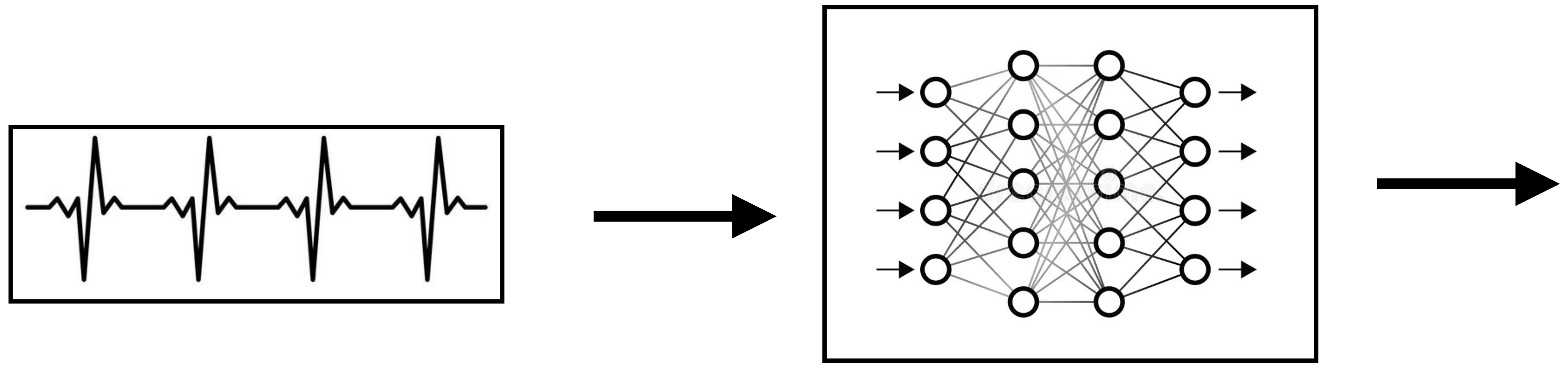
Pathway Labs

Can you see an ECG feature?



AI ECG Signal for Cardiac Arrest

Trained risk prediction model.

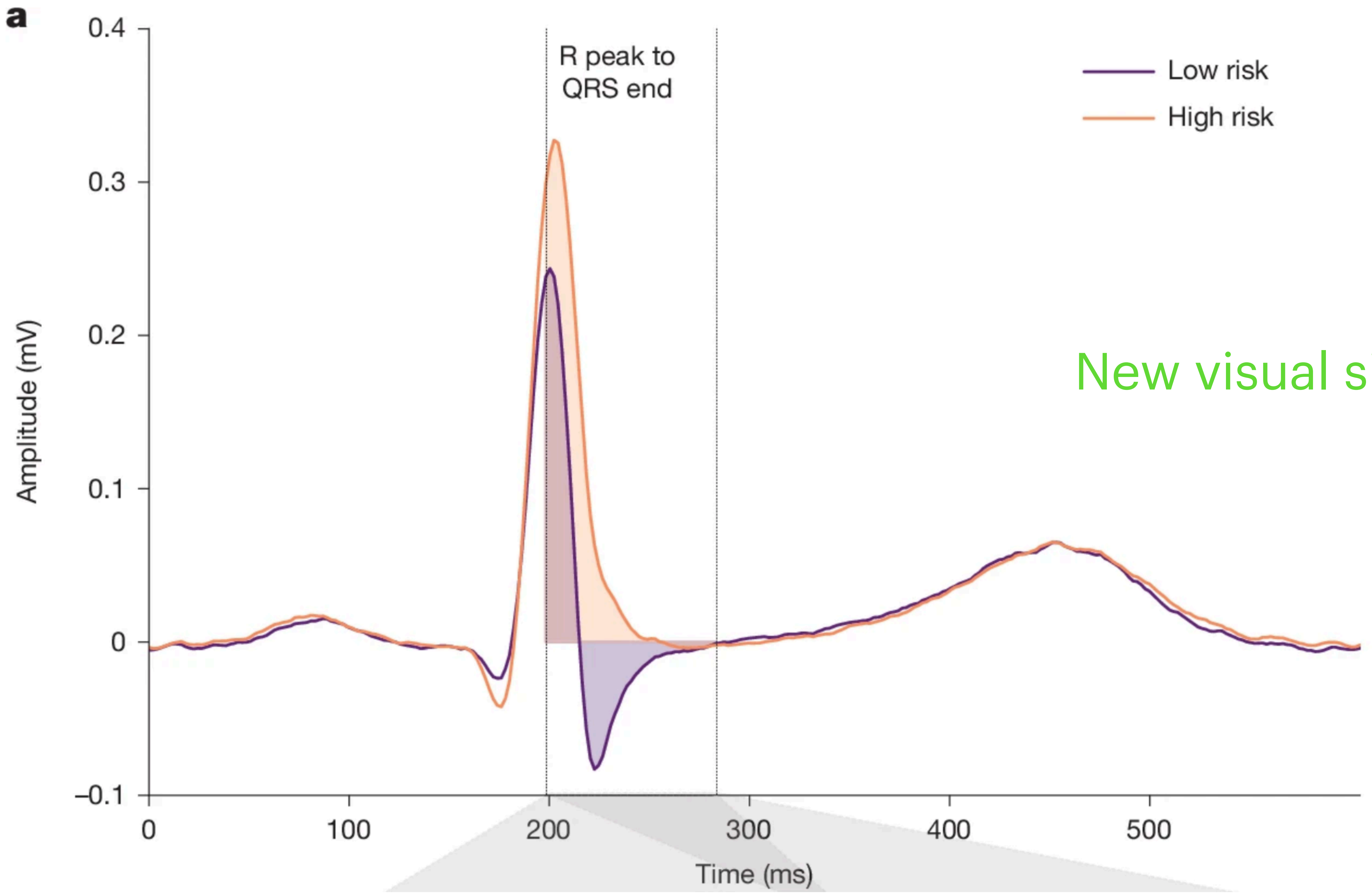


sudden cardiac death

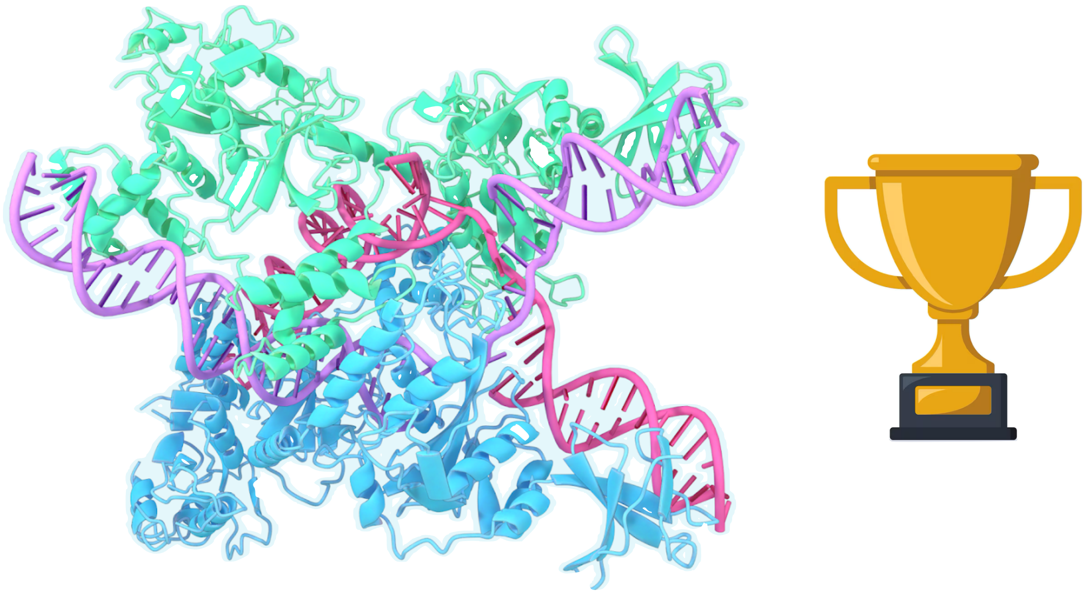
Train a model to generate the waveforms from high risk versus low risk.

Low risk

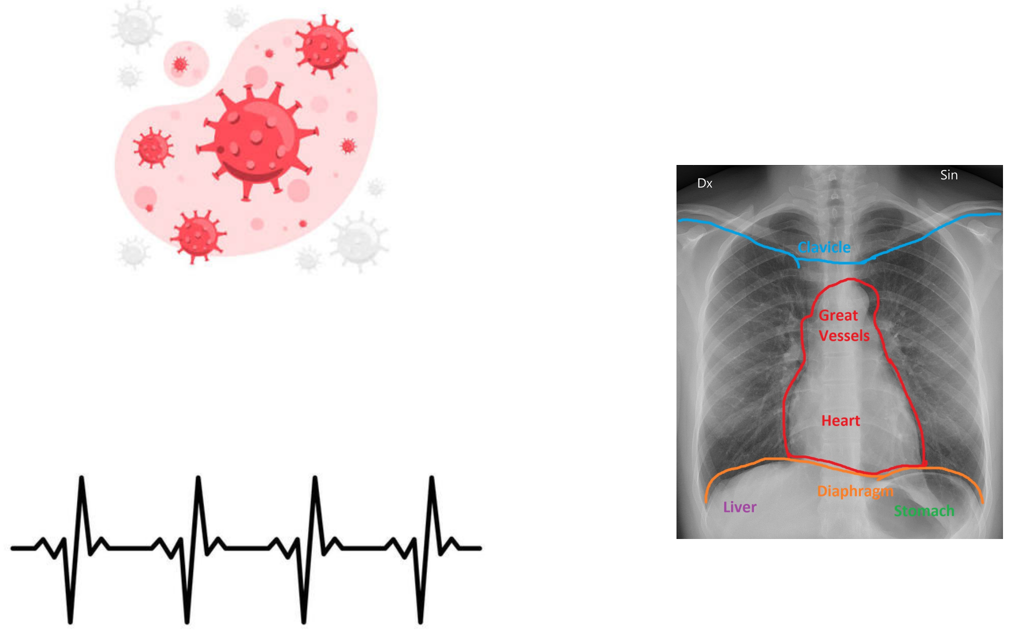
High risk



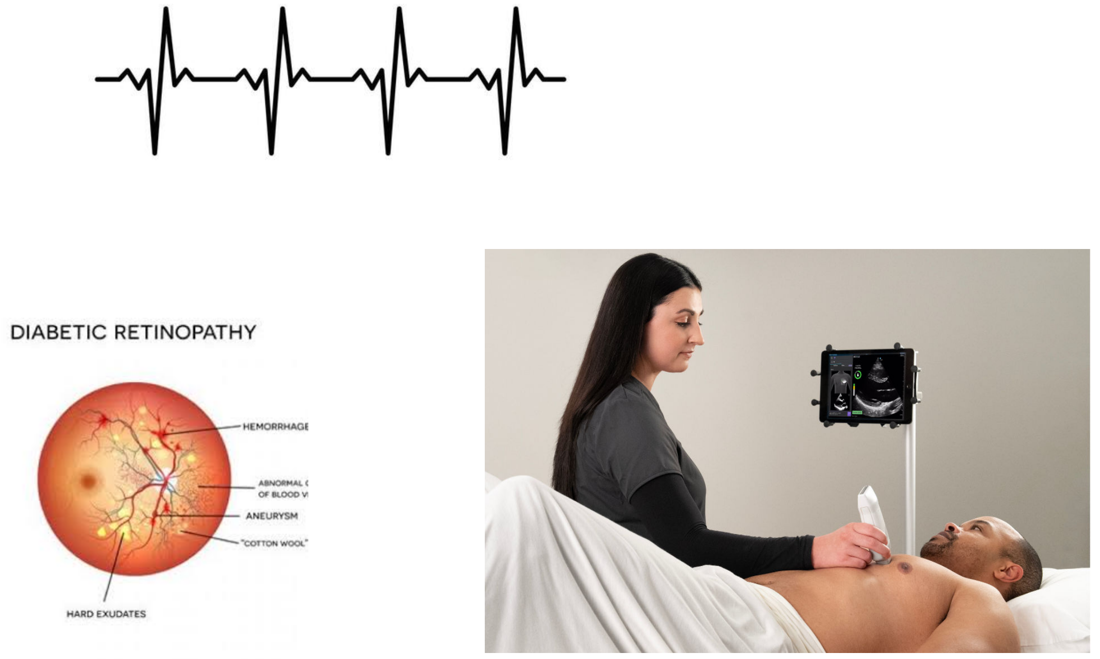
New visual signal!



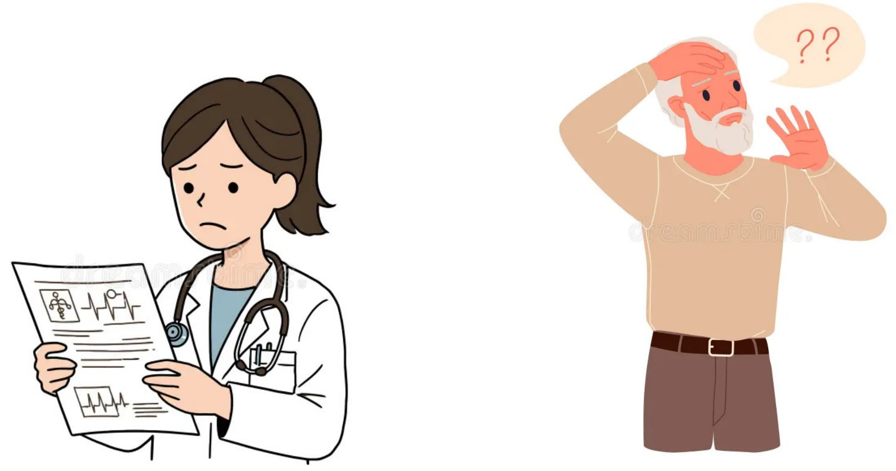
Drug Development



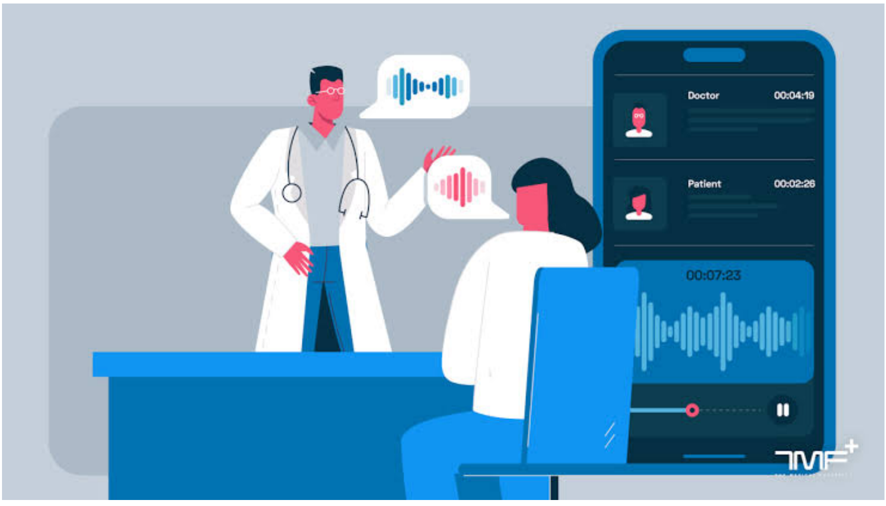
Disease Diagnosis



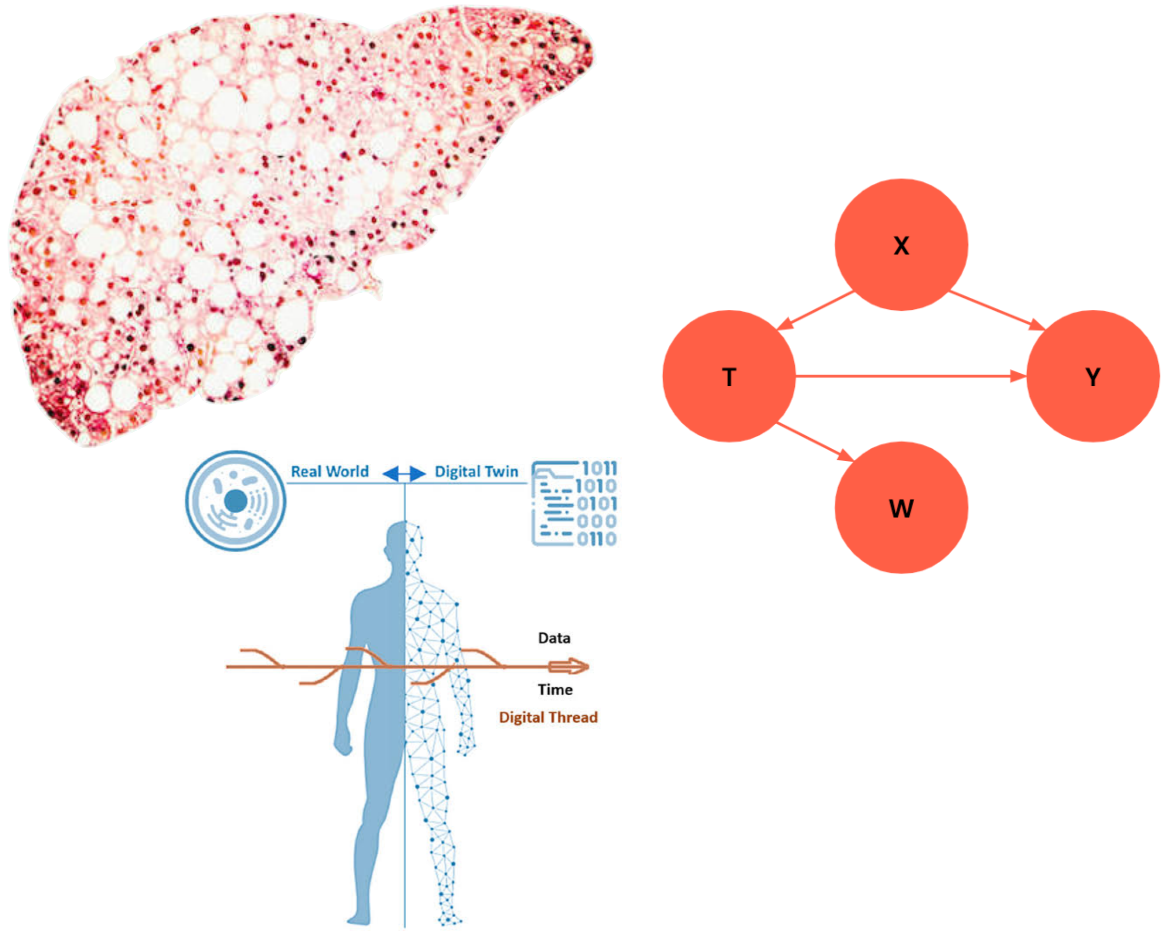
Low Resource Care



Information Retrieval



Clinical Documentation



Clinical Trials

Information Retrieval

Doctors need to stay informed of latest advancements and protocols.

UpToDate®

OpenEvidence®

$12B Company

Patients want to be informed and empowered when traditional medicine has failed or while waiting for care.

Google

ChatGPT Health

UpToDate®

University of California San Francisco

Register

Sign in

UpToDate®

Search UpToDate

Register Now!

Earn and redeem CME/CPD with every clinical question you research. Make sure to register to track your credits.

Already registered? Log in!

Practice Changing Updates   View All

NEPHROLOGY AND HYPERTENSION: A trial of finerenone in nondiabetic chronic kidney disease (June 2026)

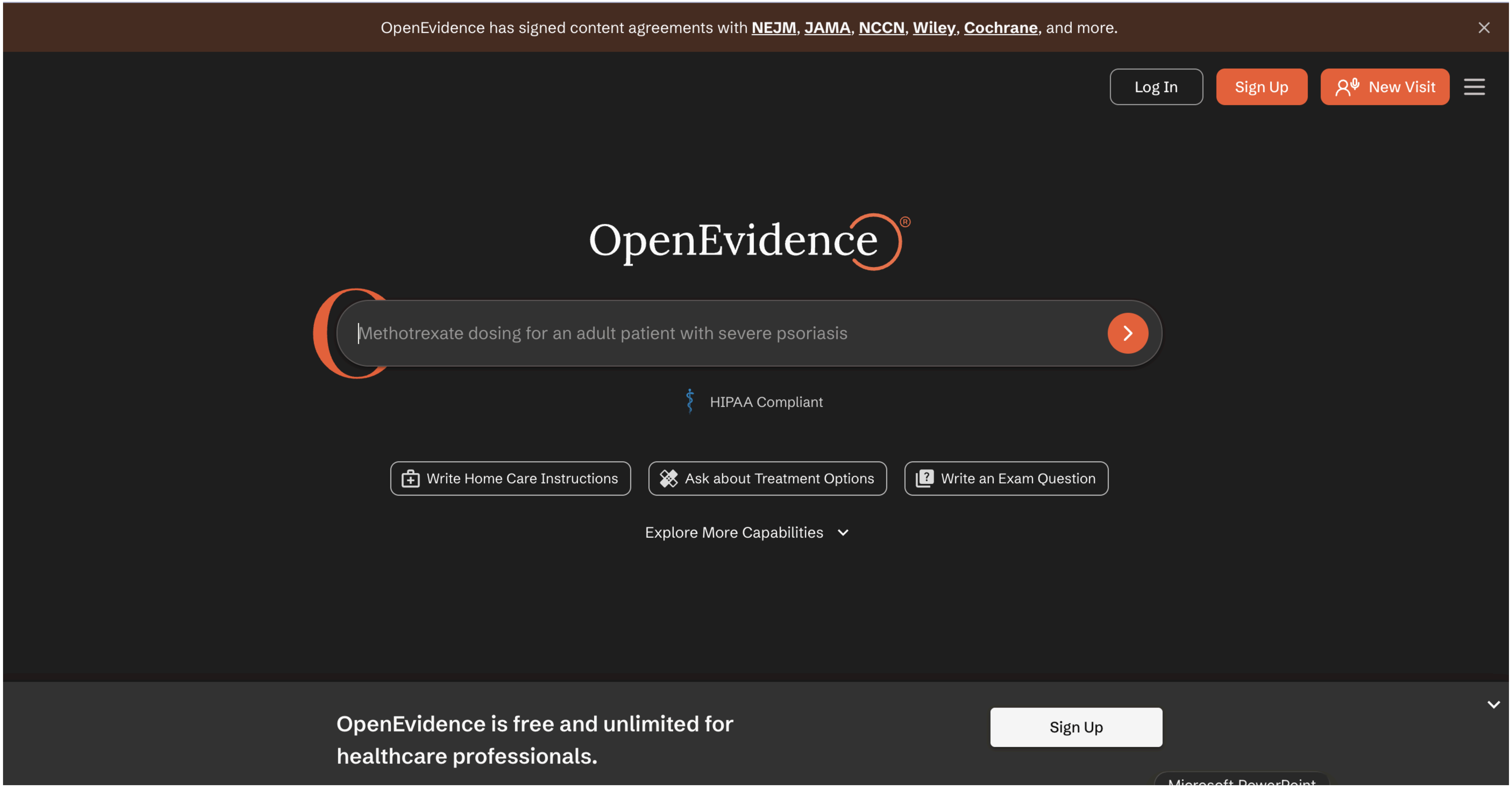
INFECTIOUS DISEASES: Ensitrelvir for post-exposure prophylaxis after household COVID-19 exposure (May 2026)

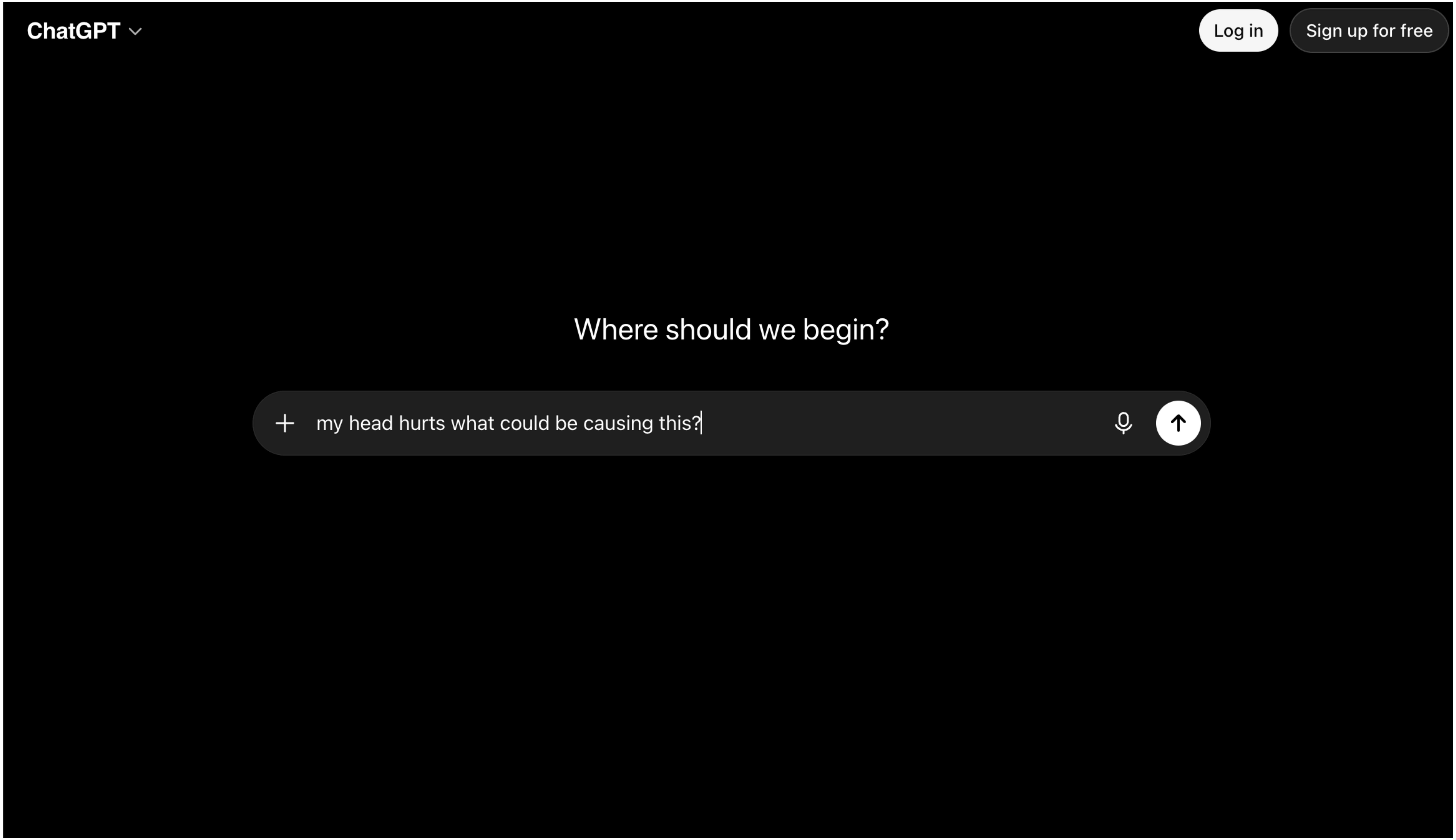
OBSTETRICS, GYNECOLOGY AND WOMEN'S HEALTH:

Drug Interactions

Patient Education

Calculators


OpenEvidence has signed content agreements with NEJM, JAMA, NCCN, Wiley, Cochrane, and more.
Log In
Sign Up
New Visit
OpenEvidence
Methotrexate dosing for an adult patient with severe psoriasis
HIPAA Compliant
Write Home Care Instructions
Ask about Treatment Options
Write an Exam Question
Explore More Capabilities
OpenEvidence is free and unlimited for healthcare professionals.
Sign Up


ChatGPT
Log in
Sign up for free
Where should we begin?
my head hurts what could be causing this?

Let's Try Doctor GPT Together

- Type in:
  - "A 45 year old woman comes in complaining of fatigue, chest pain, and shortness of breath. She has never had a heart attack. What could be the diagnosis?"
  - "A 99 year old man comes in complaining of fatigue, chest pain, and shortness of breath. He has never had a heart attack. What could be the diagnosis?"
  - "A 45 year old woman comes in complaining of fatigue, chest pain, and shortness of breath. She has a history of a bone tumor. What could be the diagnosis?"

Clinical Documentation

Today, clinical scribes may be taking notes at your visit.

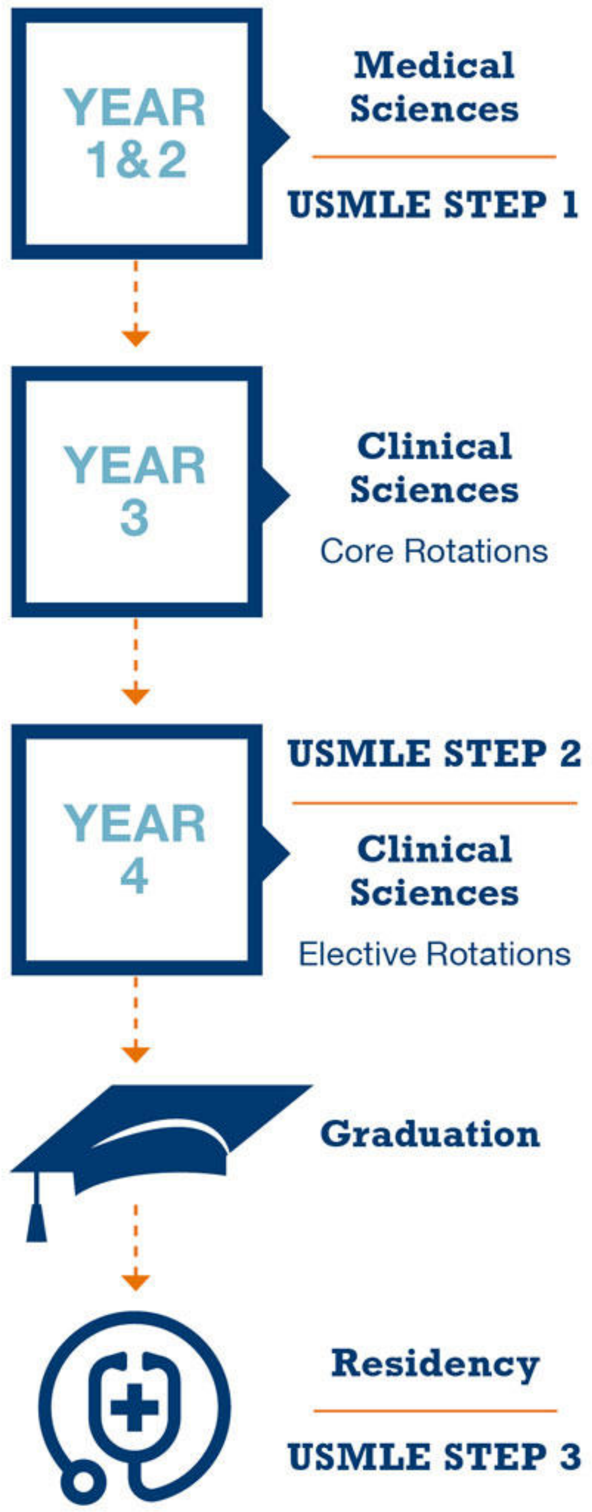
ABRIDGE

Ambience

How accurate are these LLM tools?

We need new ways to assess accuracy with text based generation.

First idea: Medical student exams!



Problems:

- Exams aren't real clinical practice.
- Exam questions are straightforward without clinical nuance.
- Do not evaluate ability to apply knowledge.
- This doesnt evaluate generated text and hallucinations.
- AI doesn't "know" things! It predicts.

How accurate are these LLM tools?

Researchers are creating new benchmarks!

GPBench: A Comprehensive and Fine-Grained Benchmark for Evaluating Large Language Models as General Practitioners

MedHELM

Holistic Evaluation of Large Language Models for Medical Tasks

LIVECLIN: A *Live* CLINICAL BENCHMARK WITHOUT LEAKAGE

Xidong Wang[1,2,*], Shuqi Guo[1], Yue Shen[2], Junying Chen[1], Jian Wang[2], Jinjie Gu[2]
Ping Zhang[4], Lei Liu[2,3,†], Benyou Wang[1,†]
[1] The Chinese University of Hong Kong, ShenZhen [2] Ant Healthcare, Ant Group
[3] Zhejiang University [4] The Ohio State University

ER-REASON: A Benchmark Dataset for LLM-Based Clinical Reasoning in the Emergency Room

Mel Molina, Nikita Mehandru, Niloufar Golchini, Ahmed Alaa

But are benchmarks the right way to evaluate these models?

Key Takeaway #2:

Language models and generative AI in healthcare need new creative modes of evaluation to ensure safety.

What is hard about healthcare AI?

Utility: Doctors already good at diagnosis!

Alert fatigue...

Accuracy rates and hallucinations!

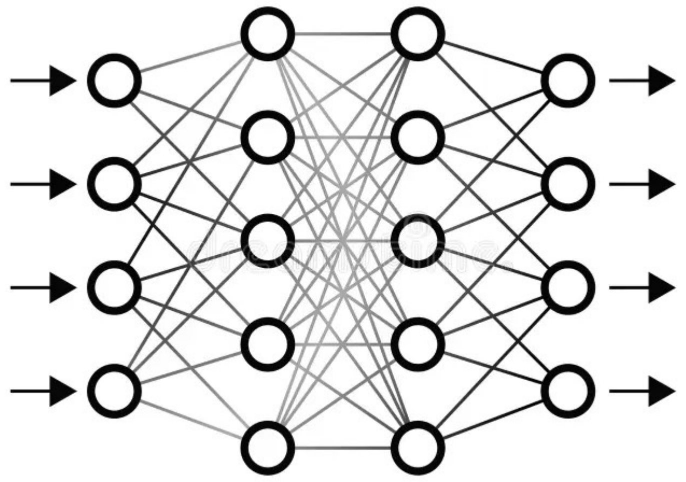
Translation: Models must work across hospitals, prevent data drift, or be interpretable.

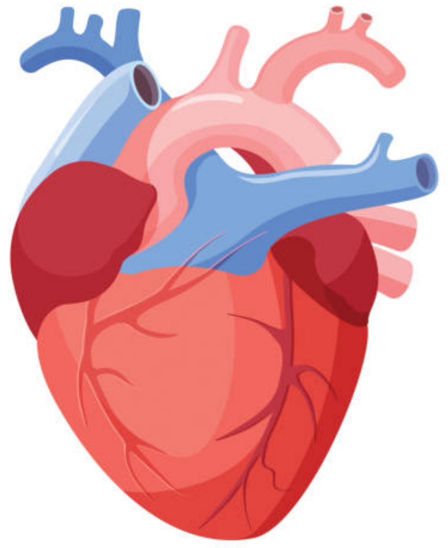
Algorithm Bias...

Algorithm Bias

Predictors

- Crushing or squeezing chest pain/pressure, often described as an "elephant on the chest"
- Pain radiating down the left arm
- Sudden, intense onset





Heart attack

Outcome

Algorithm Bias

Most common symptoms

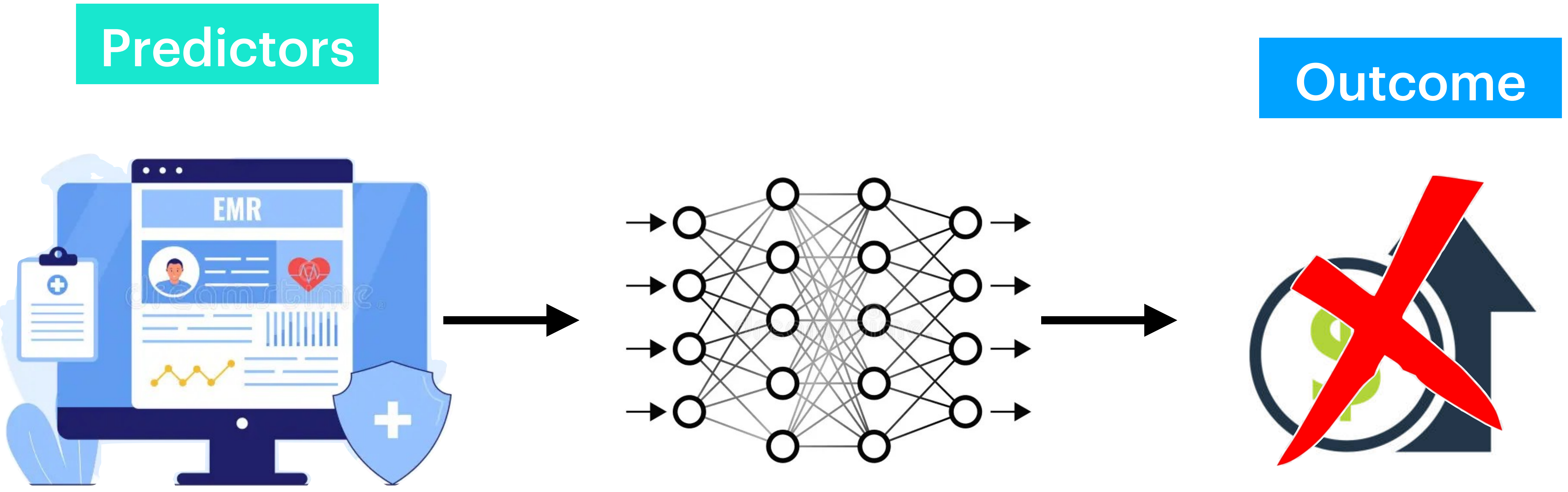
- Crushing or squeezing chest pain/pressure, often described as an "elephant on the chest"
- Pain radiating down the left arm
- Sudden, intense onset

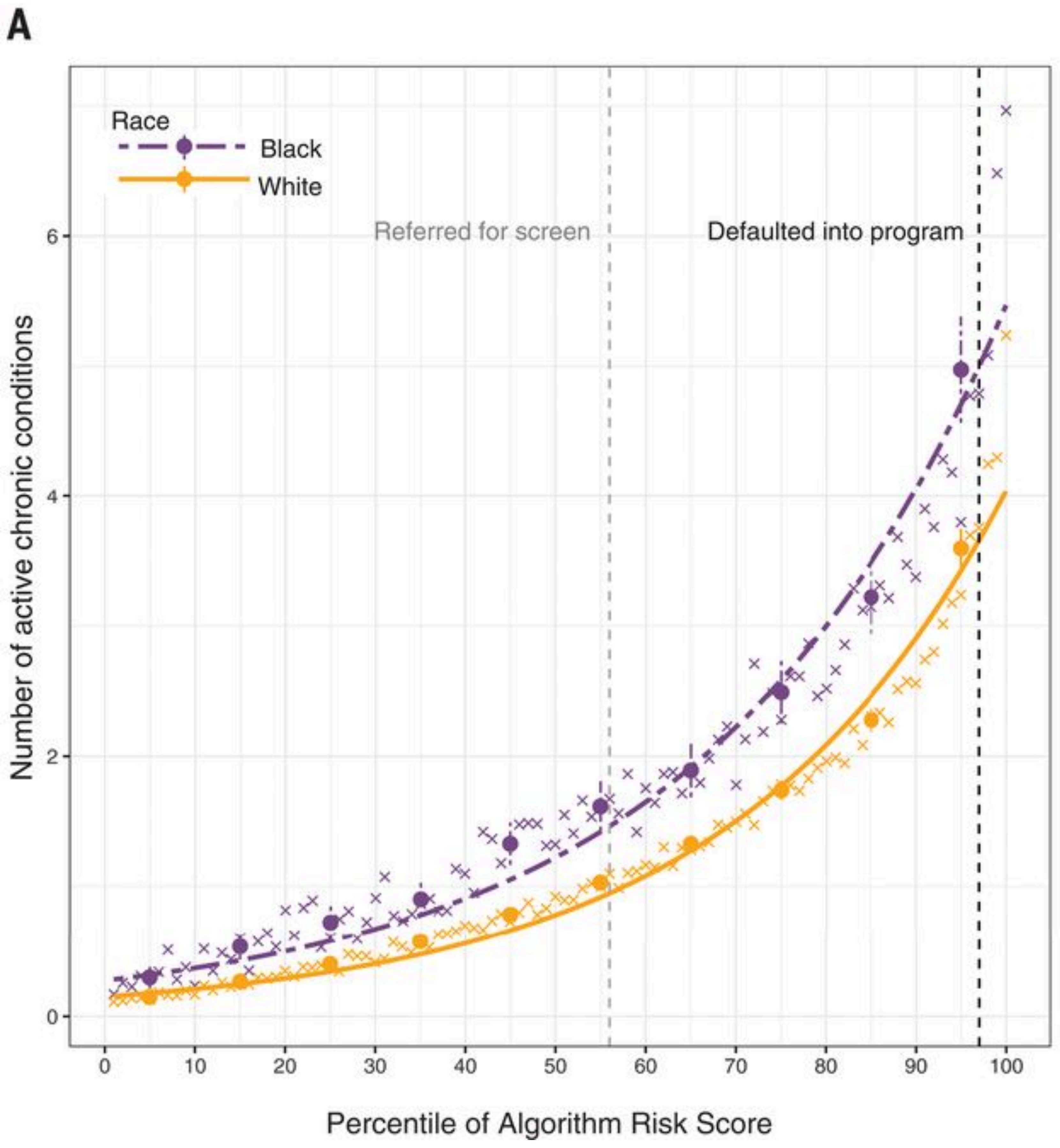
Symptoms that frequently appear in women

- Shortness of breath
- Unusual fatigue
- Nausea or vomiting
- Back or jaw pain
- Pain in the neck, throat, or upper abdomen
- Lightheadedness or dizziness
- Cold sweats
- Indigestion-like discomfort

Not in your model if you don’t include women with atypical presentation!

Algorithm Bias





Cost does not equal need and led to a biased algorithm.

Key Takeaway #3:

Always be questioning models, understand how they were built, and who is served or harmed by their use.

What is hard about healthcare AI?

Utility: Doctors already good at diagnosis!

Alert fatigue...

Accuracy rates and hallucinations!

Translation: Models must work across hospitals, prevent data drift, or be interpretable.

Algorithm Bias...

System Integration...

System Integration

Epic has large share control of the electronic medical records, can be challenging to integrate with.

Globally workflows differ greatly and solutions must be designed for the users: physicians, hospital systems, patients, or administration.

Summary Takeaways

AI for Healthcare

1. Having data from mulitple institutions is critical to build accurate robust AI for healthcare!

2. Language models and generative AI in healthcare need new creative modes of evaluation to ensure safety.

3. Always be questioning models, understand how they were built, and who is served or harmed by their use.

Let’s define a training task!

Visit

https://teachablemachine.withgoogle.com/train

Step 1: Most important step!

Decide what you want the model to do and gather data!

Step 2:

Upload and press train!

Teachable Machine
Class 1
2 Image Samples
Webcam
Upload
Class 2
4 Image Samples
Webcam
Upload
Add a class
Training
Model Trained
Advanced
Epochs: 50
Batch Size: 16
Learning Rate:
0.001
Reset Defaults
Under the hood
Preview
Export Model
Choose images from your files, or drag & drop here
Import images from Google Drive
Output
Class 1
100%
Class 2

Activities

Let's train something ourselves!

Epoch	How many times does the training run through the data?
Learning rate	How large are the gradient updates?
Batch size	How many data points do we calculate gradient descent on at once?
Train/Test	Split into data model sees versus validation data it does not.