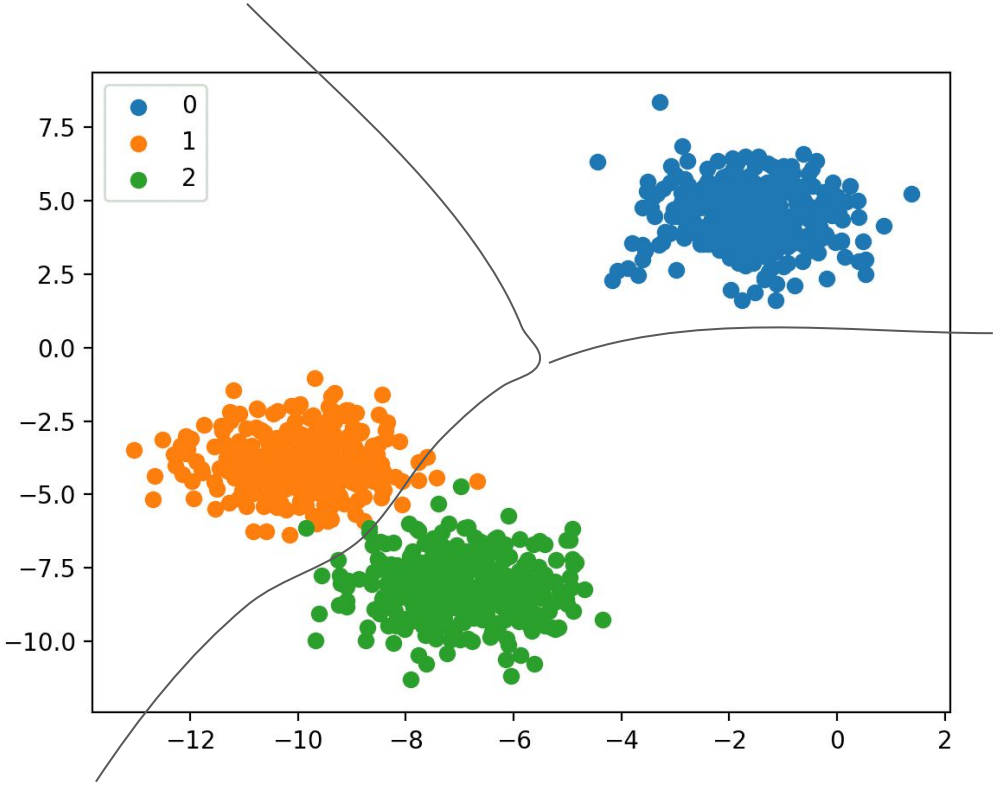
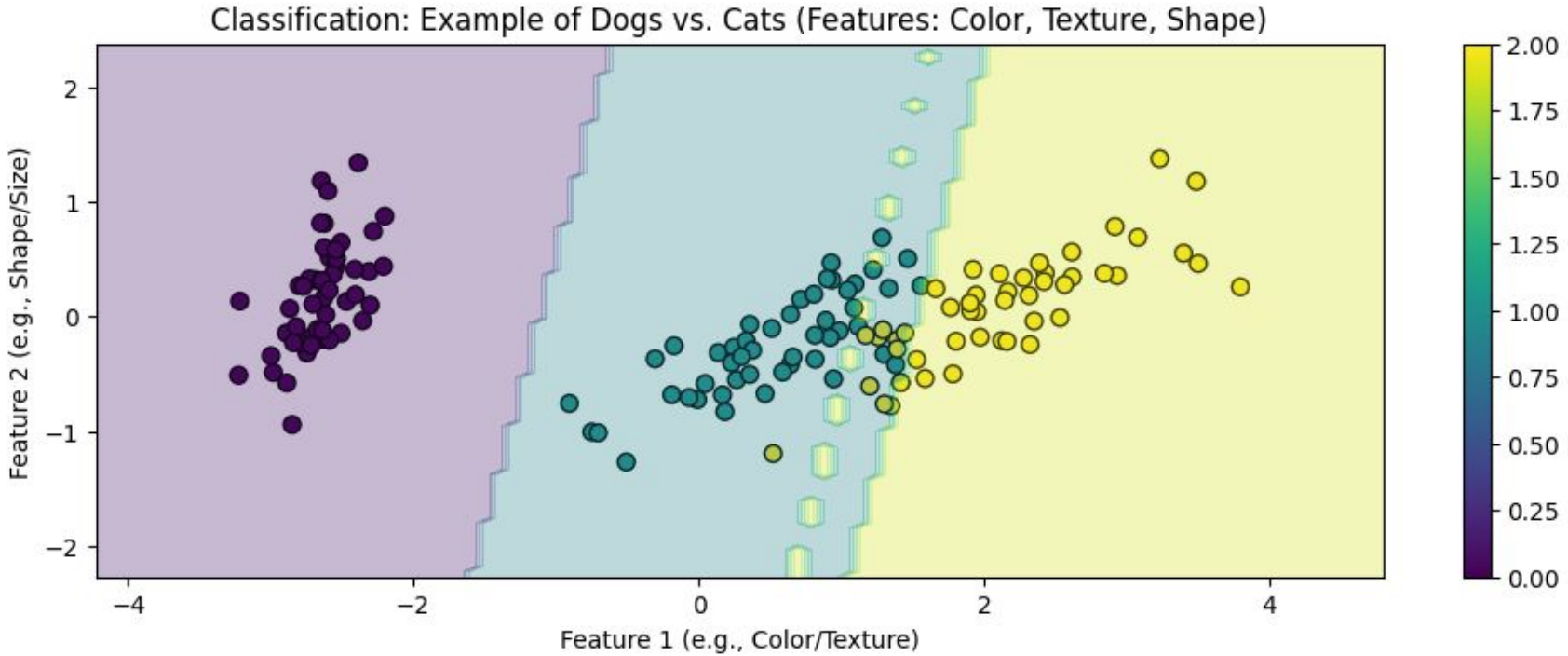
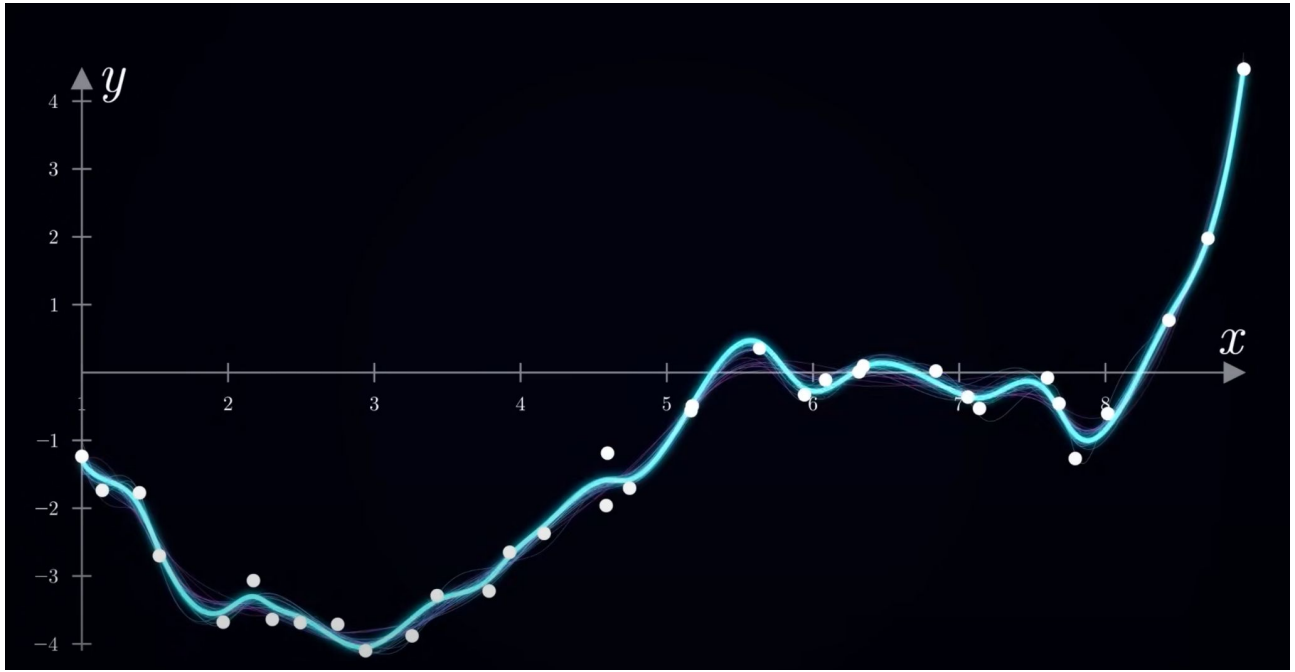
ML Applications

CSC116

Classification Model




Classification: Example of Dogs vs. Cats (Features: Color, Texture, Shape)
Feature 2 (e.g., Shape/Size)
Feature 1 (e.g., Color/Texture)
2
1
0
−1
−2
−4
−2
0
2
4
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00



Regression Model:

The results are around an linear function

$$y(x) = k_0 + k_1 x + k_2 x^2 + k_3 x^3 + k_4 x^4 + k_5 x^5$$

Examples: Classification

Distinguishing between benign and malignant breast tumors (binary classification).

Stratifying diabetic complication risk (low / medium / high risk, three classes).

Detecting whether a CT slice contains bleeding (yes / no).

Examples: Regression

Predicting a patient’s length of hospital stay (in days).

Predicting the future change in clinical score for Parkinson’s disease patients.

Demo: build AI models using PD datasets.

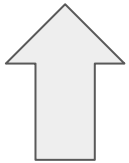
Features

57 features:

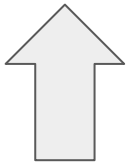
COHORT – Group or study population identifier.
age – Age of the participant.
fampd – Family history of Parkinson's disease (yes/no).
age_datscan – Age when DAT scan was performed (dopamine transporter scan).
👃 **Olfactory (Smell) Tests**
upsit – University of Pennsylvania Smell Identification Test score.
upsit_pctl – Percentile rank of UPSIT score.
upsit_pctl15 – Whether UPSIT is ≤15th percentile (yes/no).
🧠 **Cognitive Assessments**
moca – Montreal Cognitive Assessment (general cognitive screening).
bjlot – Benton Judgment of Line Orientation Test (spatial judgment).
DVS_JLO_MSSA – Digitally scored version of JLO (accuracy score).
DVS_JLO_MSSAE – JLO efficiency score (accuracy per time).
⏰ **Memory Tests (HVLT, DVT)**
clockdraw – Clock Drawing Test (visuospatial and executive function).
hvlt_discrimination – Hopkins Verbal Learning Test (discrimination index).
hvlt_immediaterecall – HVLT immediate recall score.
hvlt_retention – HVLT retention score.
HVLTFPRL – HVLT false positive recognition.
HVLTRDLY – HVLT delayed recall.
HVLTREC – HVLT recognition score.
DVT_TOTAL_RECALL – Digital Verbal Test (total recall).
DVT_DELAYED_RECALL – DVT delayed recall.
DVT_RETENTION – DVT retention rate.
DVT_RECOG_DISC_INDEX – DVT recognition discrimination index.
💬 **Verbal Fluency & Language**
lexical – Lexical fluency (word generation).
DVT_FAS – FAS test (letters-based fluency).
DVS_FAS – Digital version of FAS test.
lns – Letter-Number Sequencing (working memory).
DVS_LNS – Digital version of LNS.
MODBNT – Modified Boston Naming Test (object naming).
DVS_BNT – Digital BNT.
PCTL_BNT – BNT percentile.
🚀 **Processing Speed & Executive Function**
SDMTOTAL – Symbol Digit Modalities Test (processing speed).
DVT_SDM – Digital SDM score.
DVSD_SDM – Digital SDM duration.
TMT_A – Trail Making Test Part A (attention).
TMT_B – Trail Making Test Part B (executive function).
DVZ_TMTA – Digital TMT-A z-score.
DVZ_TMTB – Digital TMT-B z-score.
🐶 **Semantic Fluency**
VLTANIM – Animal fluency (number of animals named).
DVT_SFTANIM – Digital semantic fluency test for animals.
DVS_SFTANIM – Digital semantic fluency efficiency.
⚠️ **Cognitive Diagnosis**
MCI_testscores – Mild Cognitive Impairment (MCI) diagnosis from test scores.
cogstate – Cognitive status category.
🏠 **Daily Living**
MSEADLG – Modified Schwab and England Activities of Daily Living scale.
🏃 **Behavioral Symptoms**
quip_any – Presence of any impulse control disorder.
quip_walk – Walking-based impulse control issues.
💤 **Sleep**
ess – Epworth Sleepiness Scale (daytime sleepiness).
rem – REM sleep behavior disorder status.
🙂 **Mood**
gds – Geriatric Depression Scale.
stai – State-Trait Anxiety Inventory total score.
stai_state – STAI state anxiety (current).
stai_trait – STAI trait anxiety (general tendency).
🩺 **Other Clinical Measures**
orthostasis – Presence of orthostatic hypotension.
NP1DPRS – Depression rating from MDS-UPDRS Part I.
🧪 **Biomarkers**
abeta – Amyloid-beta levels.
tau – Tau protein levels.
ptau – Phosphorylated tau levels.
urate – Uric acid level (sometimes linked to neuroprotection).

2	100889	1	Sporadic PD	
2	100890	2	Healthy Control	
2	100890	2	Healthy Control	
2	100890	2	Healthy Control	
2	100890	2	Healthy Control	
3	100891	1	Sporadic PD	

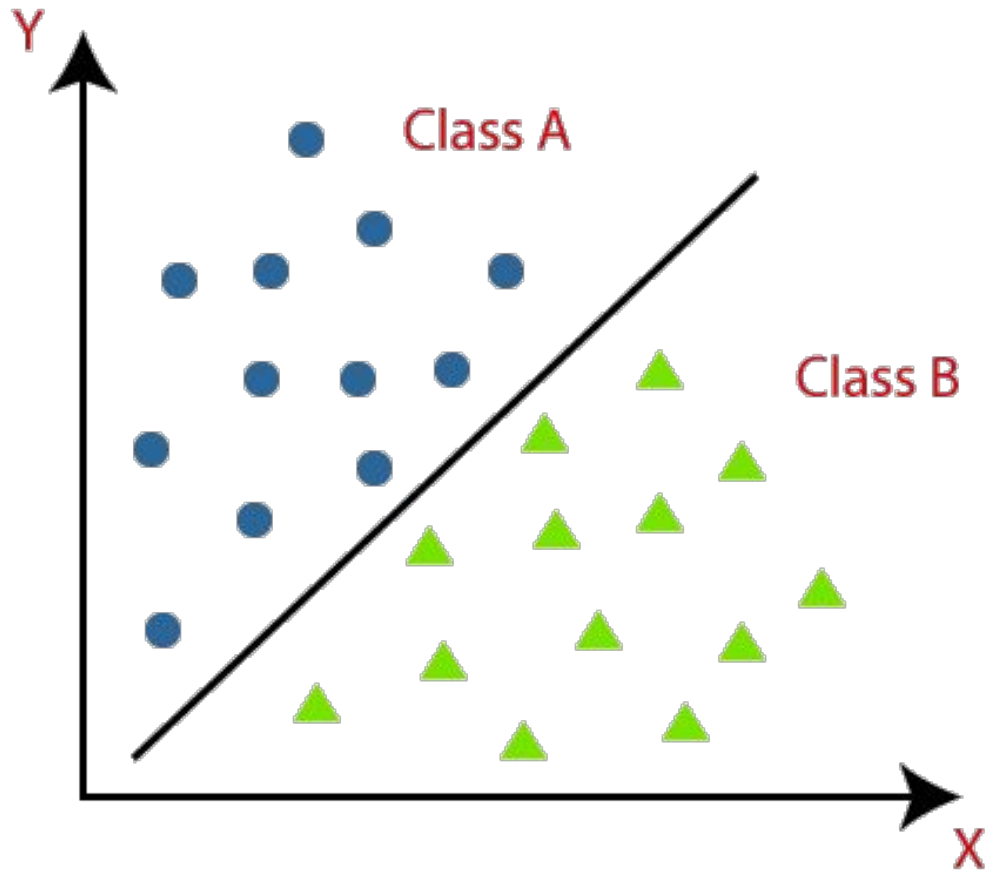
Label: **Parkinson** or **Healthy**



1



2


Y
Class A
Class B
X

4000+ patient records with 57 features

Will you use all the datasets for training??

3200 for training

800 for Testing

How to find your datasets?

Find your datasets in Kaggle, Hugging face or Github, or others.

GPU platform for training

Kaggle

Google Colab

Model **test accuracy** is high in the test datasets.

But it is pretty low in the real cases. What should we do?

Fine-tuning

"Fine-tuning" refers to a **transfer learning** technique where a pre-trained model is further trained on a new, specific dataset to improve its performance on a particular task, rather than training a model from scratch.

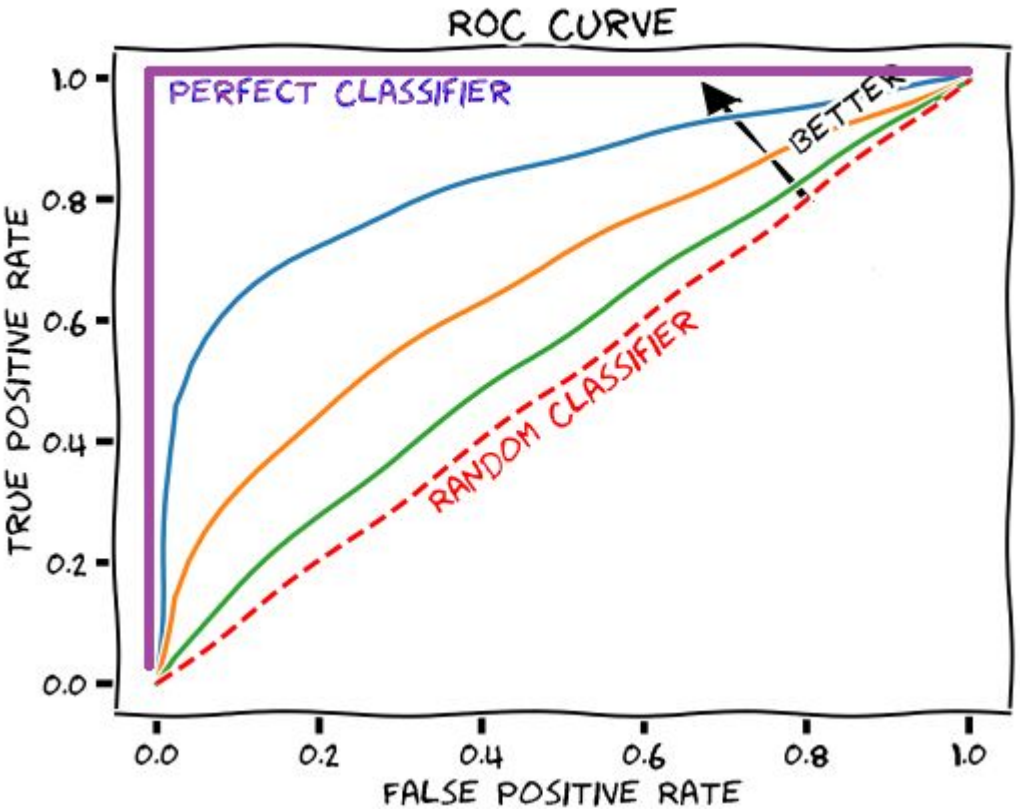
Training Accuracy:

Definition: The percentage of correct predictions made by the model on the **training dataset**, which is the data used to train the model.
Purpose: It tells you how well the model has **learned** from the data it was trained on.

Testing Accuracy:

Definition: The percentage of correct predictions made by the model on the **testing dataset**, which is separate from the training data.
Purpose: It shows how well the model **generalizes** to **new, unseen data**.

AUC/ROC



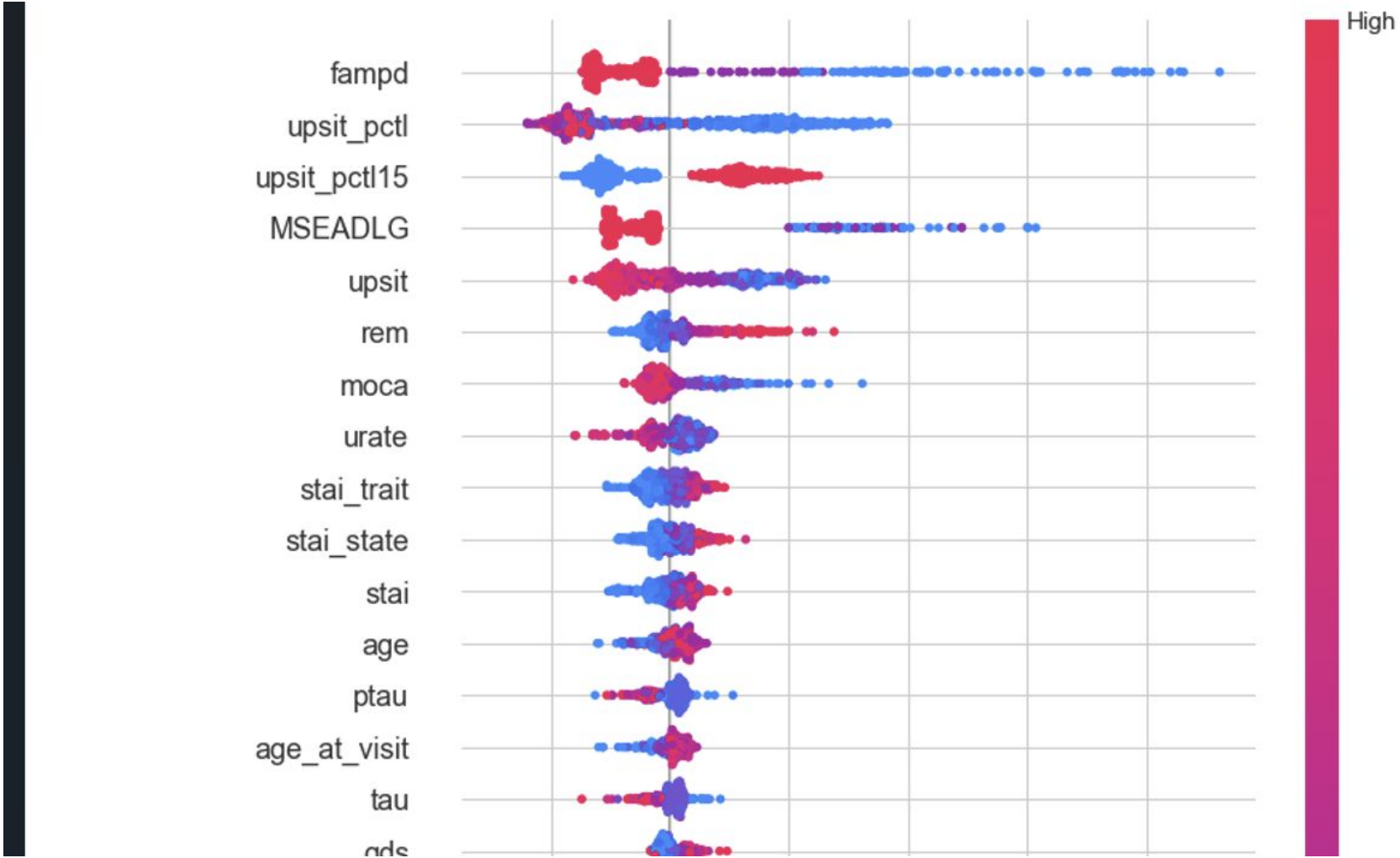
AUC, or Area Under the Curve, is a metric used in machine learning, particularly for evaluating the performance of binary classification models.

How good the model it is?

- **AUC = 1.0**: Perfect classifier
- **AUC = 0.5**: No discriminative power (equivalent to random guessing)
- **AUC < 0.5**: Worse than random (model is misclassifying)

Feature Importances

Explainable AI solutions:
SHAP


fampd
upsit_pctl
upsit_pctl15
MSEADLG
upsit
rem
moca
urate
stai_trait
stai_state
stai
age
ptau
age_at_visit
tau
High

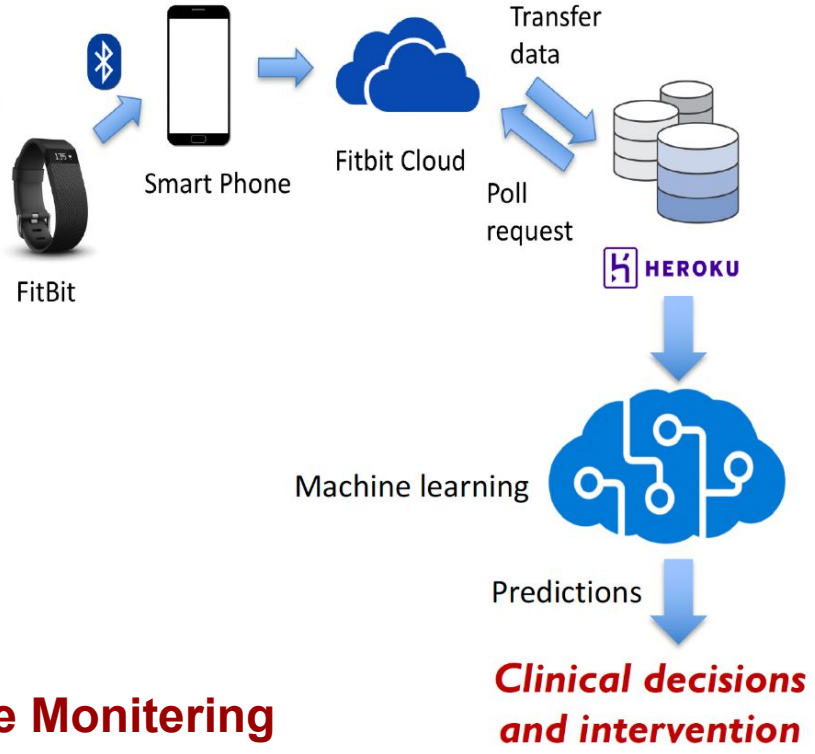
Wearables

- Commonly available: step, heart rate, sleep stages
- More sensing modalities
  - Oxygen saturation (SpO2)
  - Skin temperature
  - Breathing rate
  - Heart rate variability
  - ECG
  - Stress
- **500+ million** wearables sold in 2021

Unprecedented monitoring capability outside hospitals!

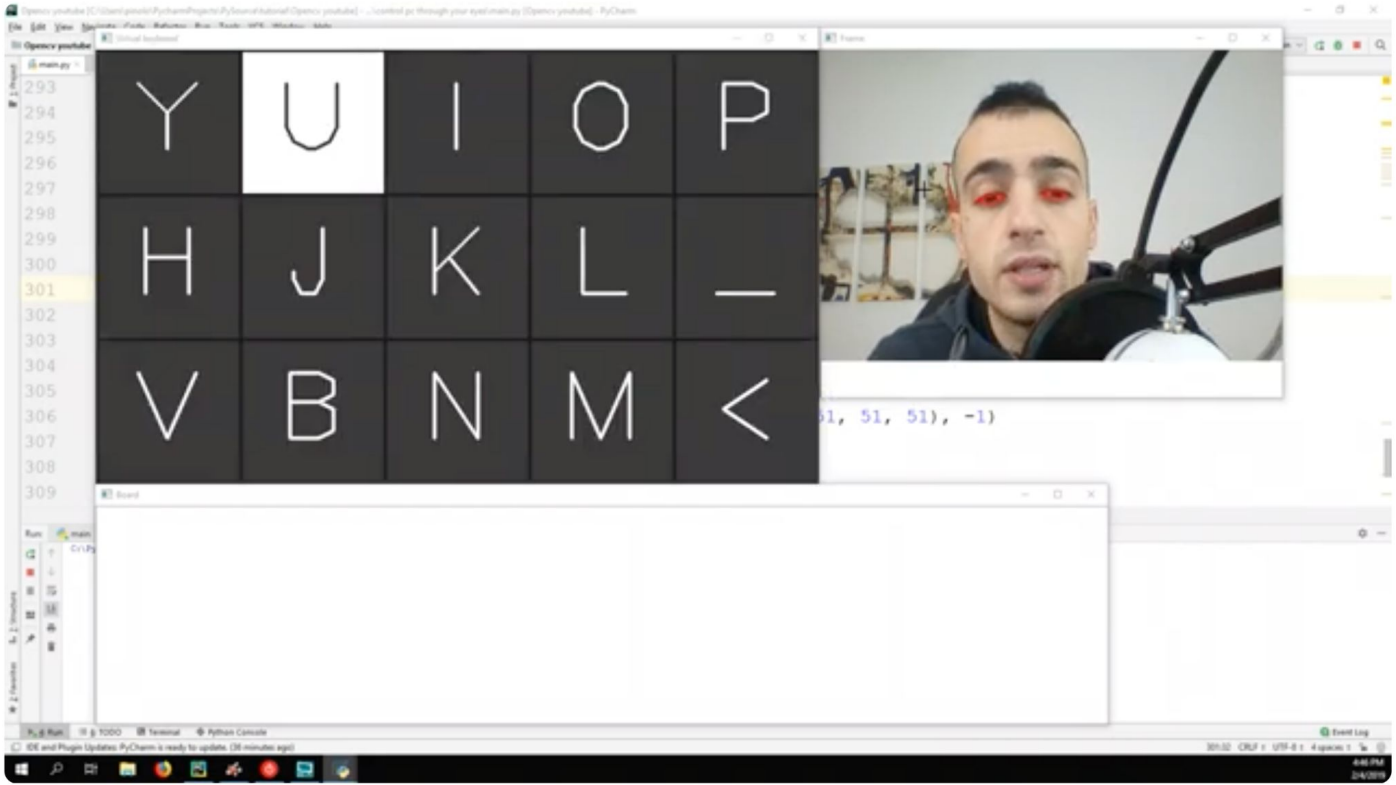
Internet of Medical Things

- **Wearables**: wristband, smartwatch, ring...
  - Long-term, non-obtrusive monitoring
- **Connectivity**: Bluetooth, WiFi, cellular
  - Real-time monitoring and intervention
- **Cloud**: computing and storage
  - Scalable to large population
- **Analytics**: machine learning
  - Predict outcomes and support intervention



Smart Remote Monitering

Use your eyes to print: Classification or Regression??



Why are Hospitals Under Attack?

- Healthcare data is highly **valuable**.
- Example: Ransomware shutting down ICU operations.

A major hospital in Florida had to shut down some of its systems and turn patients away after a ransomware attack disrupted its IT infrastructure. "Hospitals and healthcare organizations are particularly attractive targets for cyber-criminals, and their reliance on technology to manage everything from patient records to surgical equipment makes them uniquely vulnerable. This is compounded by their limited resources to invest in cybersecurity measures," stated Jan Lovmand, BullWall CTO.

Detecting abnormal prescriptions.

GPT /
BERT

How is BERT/GPT used in hospitals?

- **Helps read medical records quickly** and extract key info like symptoms, diagnoses, or medications.
- **Sorts through thousands of documents** to find the right information for doctors and nurses.
- **Understands patient messages** in chatbots or digital forms and sends them to the right department.
- **Assists in clinical decision support systems**, by reading guidelines and matching them to patient data.

Real-world example for nurses:

A patient types: "I feel dizzy after taking the medication."

- A regular computer might just focus on "dizzy."
- BERT understands **the full meaning** and realizes that **the medication could be the cause**, helping systems alert a nurse or doctor.

Why?

Why AI in Hospitals?

- **Faster Diagnosis**: AI helps analyze medical images, lab results, and patient history quickly and accurately.
- **Predictive Analytics**: AI can forecast patient outcomes, readmission risks, and disease progression.
- **Personalized Treatment Plans**: AI tailors care based on patient data, genetics, and treatment history.
- **Virtual Assistants**: AI chatbots provide 24/7 support to patients for queries, follow-ups, or reminders.
- etc.

AI in Action (IT)

- Phishing Email → AI flags → IT reviews → Threat stopped.

Discussion

- Would you trust this AI?

AI may fail but if it is high accuracy, e.g., 99%, it is trusted.

If the model accuracy is 90%, then, the result may be wrong.