

UNIVERSITY OF ALBERTA

gairdner
LES PRIX CANADA GAIRDNER AWARDS

AIMSS
ARTIFICIAL INTELLIGENCE IN MEDICAL SYSTEMS SOCIETY

Gairdner
Alberta Partnered AI + Health Research Symposium

Bridging the Gap: Machine Learning for Medical Professionals

Workshop Instructors



Ying Wan
Research Lead & Acting Education Lead
AI in Medical Systems Society

MSc Student, Biomedical Engineering
Dept. Electrical and Computer Engineering
University of Alberta



Ehsan Misaghi, MSc
Co-Founder & Director of Research and Education
AI in Medical Systems Society

MD/PhD Candidate, Medical Genetics
Dept. Ophthalmology and Visual Sciences
University of Alberta

Workshop Assistants



Rojin Adabdokht, DDS, MSc
PhD Student
Medical Informatics
Dept. Dentistry
University of Alberta



Sacha Davis, MSc
AI Specialist
Dept. Medicine
University of Alberta



Devaanshu Kundu
MSc Student
Cardio-oncology
Dept. Nursing
University of Alberta

Agenda

9 - 9:30	Crash Course on ML and AI
9:30 - 9:40	Quick Break
9:40 - 10:40	Classical ML Activity - ML Triaging
10:40 - 11:30	DL Activity - Brain MRI Segmentation



Gairdner
Alberta Partnered
AI + Health Research
Symposium

UNIVERSITY OF ALBERTA

gairdner

AIMSS

Crash Course: AI and ML



Gairdner
Alberta Partnered
AI + Health Research
Symposium

UNIVERSITY OF ALBERTA

gairdner
LES PRIX CANADA GAIRDNER AWARDS

AIMSS
ARTIFICIAL INTELLIGENCE IN MEDICAL SYSTEMS SOCIETY

AI vs. ML

Artificial Intelligence (AI)
Simulates human intelligence, such as learning, reasoning, problem-solving, and decision-making.



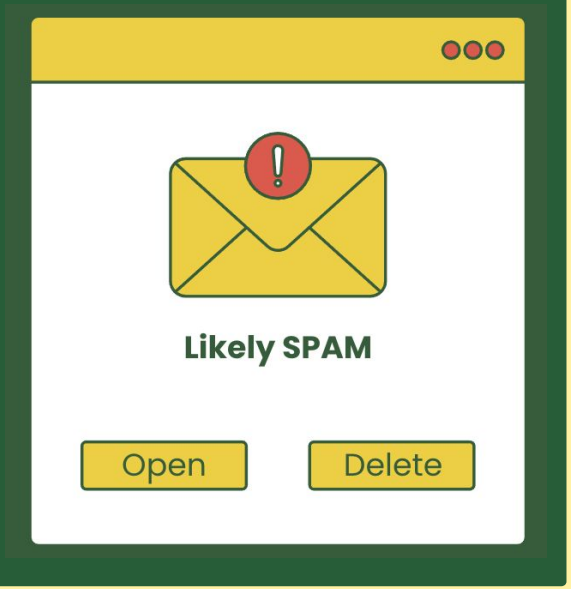
https://learning.tcairem.ai/terminology/

AI vs. ML

Artificial Intelligence
Simulates human intelligence, such as learning, reasoning, problem-solving, and decision-making.

Machine Learning (ML)
Learns patterns from data to make predictions.

Deep Learning (DL)
Subfield of ML that uses neural networks to learn hierarchical representations from data.



https://learning.tcairem.ai/terminology/

AI vs. ML

Artificial Intelligence
Simulates human intelligence, such as learning, reasoning, problem-solving, and decision-making.

Machine Learning
Learns patterns from data to make predictions.

Supervised Learning
Finds patterns in labelled data.
→ Classification
→ Regression

Unsupervised Learning
Finds structure in unlabelled data.
→ Clustering
→ Dimensionality Reduction

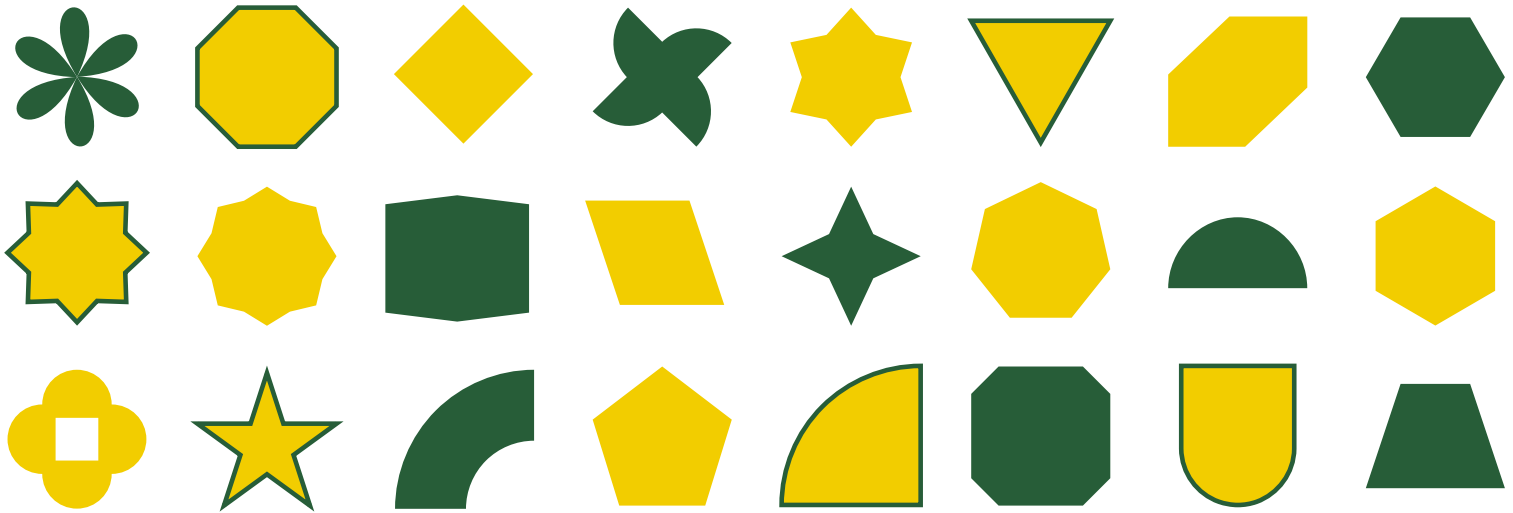
Reinforcement Learning
Learns from trial and error.

Etc. (self-supervised, semi-supervised...)

https://learning.tcairem.ai/terminology/

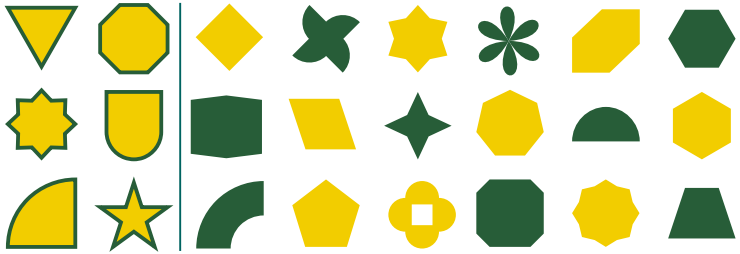
Unsupervised Learning

How would you **cluster** these shapes into 2 groups?

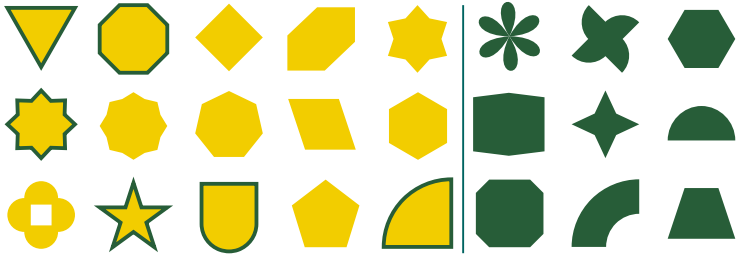


Unsupervised Learning

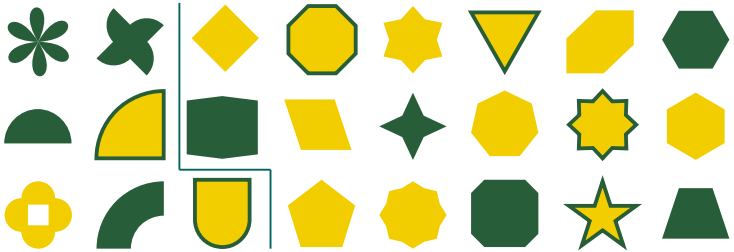
By outline



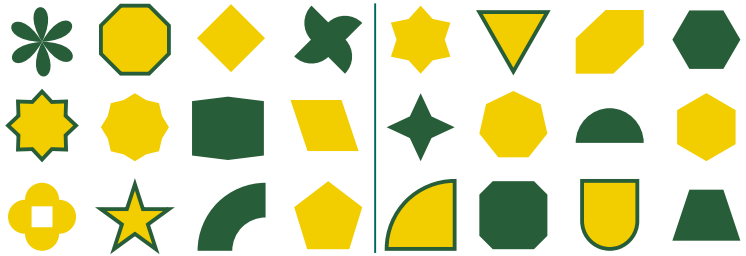
By fill color



By having a curved side



Straight down the middle



Unsupervised Learning

The factors we use for grouping are features and the number of features is dimensionality.
Features are the attributes used by a model to make predictions. They can be numerical, categorical or derived.



Feature	Value
Outline	True
Color	Yellow
Curved Edge	False



Feature	Value
Outline	False
Color	Green
Curved Edge	True

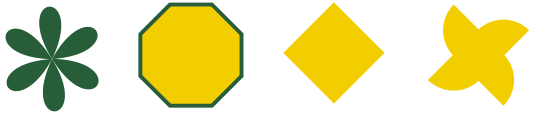
The way we grouped them is through unsupervised clustering
ML models cluster based on patterns or relationships they find in the data. The way the clusters are formed, depends on the algorithm used.



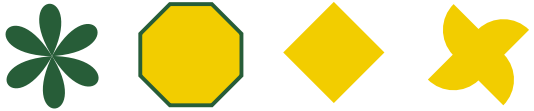
https://learning.tcairem.ai/terminology/

Supervised Learning

What **label** does each shape **classify** as, “A” or “B”?



Since we have no idea what the **labels** mean, let’s just randomly guess…



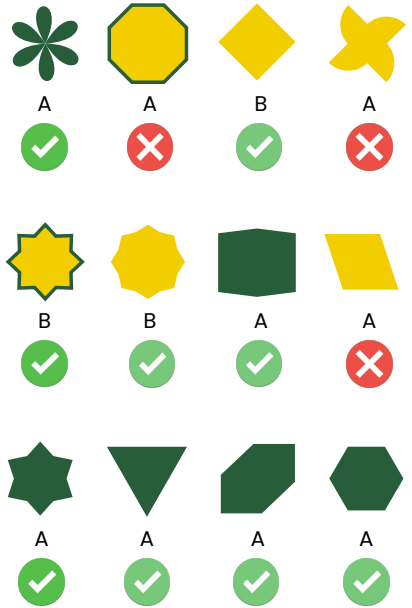
A A B A

And our answers are checked against an answer key (**Ground Truth**).



Supervised Learning

This process repeats for multiple training runs, improving (hopefully) each time, until we have **trained** “enough”.



Supervised Learning

Finally, we **test** if we can **generalize** the modelled relationships to everything by applying what we've learned to unseen data and doing one final scoring against an answer key.



Metrics
Accuracy
Recall (Sensitivity)
Precision

Supervised Learning

The categories “A” and “B” are called target labels/classes and the process of labelling is classification.



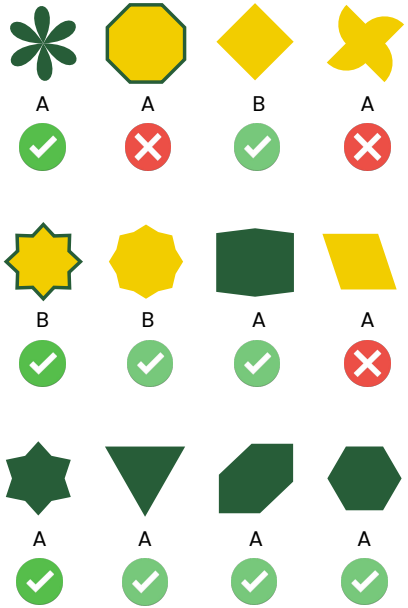
The way we label and check against an answer key is called training.
Teaches the model patterns and relationships between features (attributes) and labels.

The number of training runs is an example of a hyperparameter.
Hyperparameters are settings defined before training that affect model performance.

The final evaluation of model performance is testing.
Check how well the trained model generalizes to unseen data.

A Warning

So based on what our model has observed, what is the unknown shape?



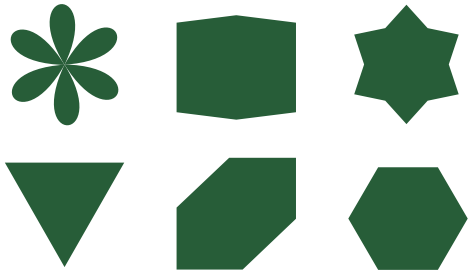


A Warning

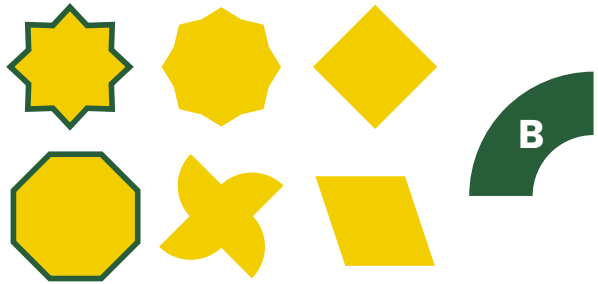
If you said “A”, you are wrong…

The **target** is whether the shape has a number of sides that is a multiple of 3 or not (counting a continuous curve as 1 side)

A (multiple of 3)



B (not a multiple of 3)



A Warning

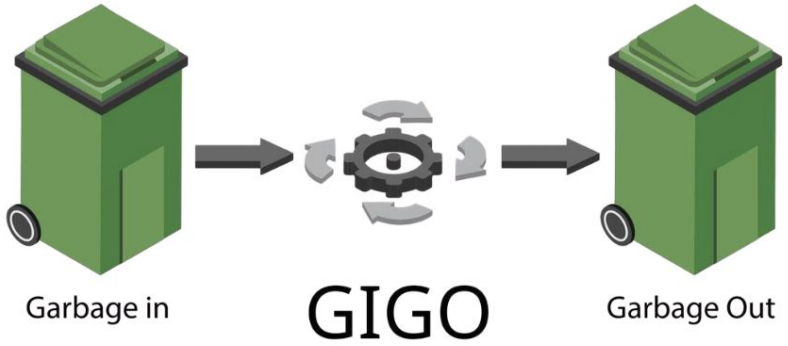
So why didn’t you see the pattern?

- You weren’t given enough data samples and couldn’t accurately model the relationship.
- The data you were given during training didn’t represent what you would see in real life.
- There were too many features, and you couldn’t determine which were important.
- The way you approached the problem was overly simplistic.
- You were exposed to information that unintentionally influenced your decision.
- I gave you a poorly defined problem.

All of these, are reasons models may fail.

A Warning

ML, at its core, is just sophisticated statistics



Break (10 Min)

Please set up your notebook environment if you have not already done so.
If you have any questions, ask any of us for help :)



Gairdner
Alberta Partnered
AI + Health Research
Symposium

UNIVERSITY OF ALBERTA

gairdner
LES PRIX CANADA GAIRDNER AWARDS

AIMSS
ARTIFICIAL INTELLIGENCE IN MEDICAL SYSTEMS SOCIETY

https://drive.google.com/drive/folders/1oykQo8wXf-UEKgYv6sAHyTxYCKlMpbcq?usp=sharing

Activity: ML Triaging



Gairdner
Alberta Partnered
AI + Health Research
Symposium

UNIVERSITY OF ALBERTA

gairdner
LES PRIX CANADA GAIRDNER AWARDS

AIMSS
ARTIFICIAL INTELLIGENCE IN MEDICAL SYSTEMS SOCIETY

https://drive.google.com/drive/folders/12_u-d0pSemTmf-XOVv5B3P_W9OOYIq9T?usp=drive_link

How to Train Your Model

1. Define a problem
2. Collect/Gather Data
3. Clean and Preprocess the Data
   → Exploratory Data Analysis
   → Feature Extraction
   → Data Segregation
4. Choose an ML algorithm
5. Model Evaluation
6. (Deploy and Maintain)

All Throughout: Reiterate



1 Define a Problem

A well defined problem will guide how you obtain and process your data, how to best train and evaluate your models, and how to deploy your pipeline in practice.

- Is the target measurable and clearly defined?
- Can you state the problem as predict/classify/discover/rank?
- Do you have access to meaningful features (labs, vitals, imaging)?
- Do you have a sufficient amount of data (hundreds to thousands of examples)?

Given a patient's presenting vitals, can we predict the Emergency Severity Index (ESI) score they are assigned at triage?

2 Collect/Gather Data

Structured Data

Data of standardized format (ex. table); typically can be modelled with classical ML.

- Clinical Measurements
- Structured Genomics
- Laboratory Results

Semi-Structured Data

Doesn't strictly conform to a table, but still has organizational properties (ex. json, xml, html).

- Electronic Medical Records (EMRs)
- Clinical Notes
- Pathology Reports

Unstructured Data

Data that does not follow a standardized format (ex. images, sound, text); typically requires DL.

- Recordings
- Medical Images
- Physiological Signals

2 Collect/Gather Data

If your model performs poorly even after hyperparameter tuning, it's usually your data.

- Do you need or have access to labelled data?
    - Can you get people to label data for you?
    - Do you need to train people to label data?
- Is there a publicly available dataset you can use?
    - Kaggle, government websites, WHO
- Is your data source reliable?
- Is your data applicable to the problem you are trying to solve?
- Do you need structured or unstructured data?

2 Collect/Gather Data

- Deidentified emergency department (ED) visits between 2013 and 2017 of 3 Yale New Haven hospitals
- Records 972 features of 560,486 patient visits
  - Including demographics, triage evaluation, chief complaint, and past medical history

	dep_name	esi	age	gender	ethnicity	race	lang	religion	maritalstatus	employstatus	...	cc_vaginaldischarge	cc_vaginalpain	cc_weakness	cc_wheezing	cc_withdrawal-alcohol	cc_woundcheck	cc_woundinfection	cc_woundre-evaluation
0	B	4.0	40.0	Male	Hispanic or Latino	White or Caucasian	English	NaN	Single	Full Time	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	B	4.0	66.0	Male	Hispanic or Latino	Native Hawaiian or Other Pacific Islander	English	Pentecostal	Married	Not Employed	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	B	2.0	66.0	Male	Hispanic or Latino	Native Hawaiian or Other Pacific Islander	English	Pentecostal	Married	Not Employed	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	A	2.0	66.0	Male	Hispanic or Latino	Native Hawaiian or Other Pacific Islander	English	Pentecostal	Married	Not Employed	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	A	3.0	84.0	Female	Hispanic or Latino	Other	Other	Pentecostal	Widowed	Retired	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0201016#sec028

3 Clean and Preprocess the Data • EDA

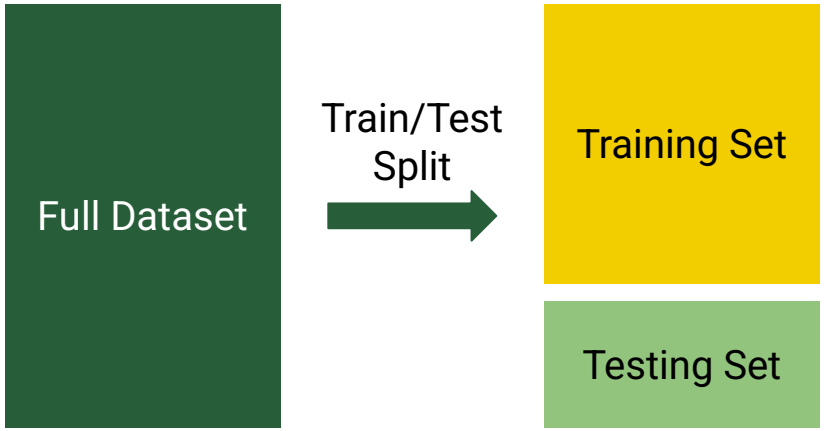
Exploratory data analysis (EDA) is the process of analyzing a dataset for patterns, outliers, and relationships before formal modeling.

See notebook for steps in:
• checking for and handling missing data,
• checking for (and handling) imbalanced data, and
• checking for biases.

Bonus: Plot the linear coefficient map between features. Do the results affect what models you think are appropriate for this problem?

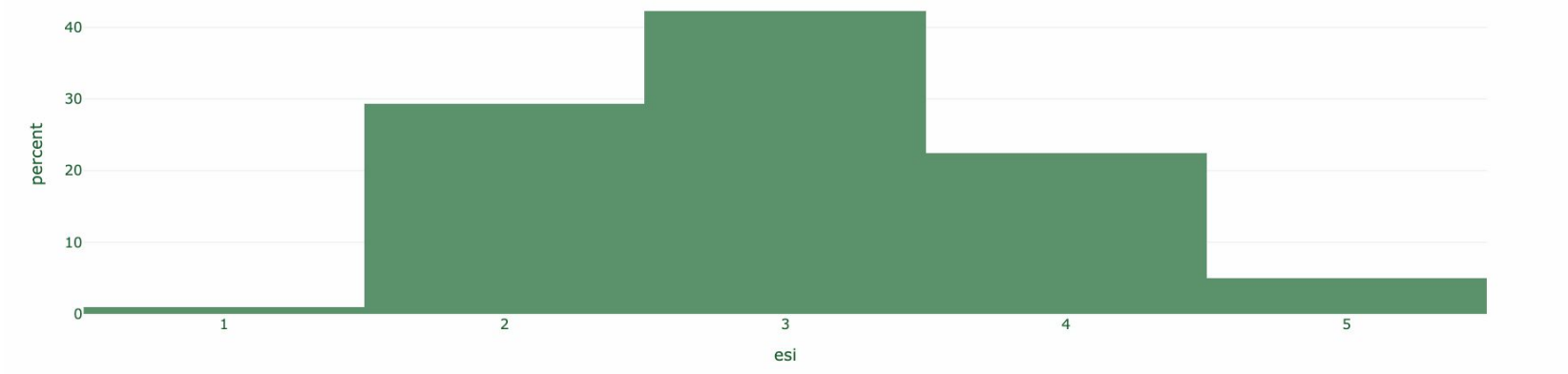
3 Clean and Preprocess the Data • Data Segregation

We need to split the data into training and testing. A typical split is 80% training and 20% testing, but that may not always be true.



3 Clean and Preprocess the Data • Data Segregation

During EDA, we find a class imbalance, where the majority of the data is level 3



What could happen if we were to split all the data into 2 groups (4:1)?

3 Clean and Preprocess the Data • Data Segregation

If we have a group of 16 green and 4 yellow. Statistically, there is 1 yellow for every 4 green.

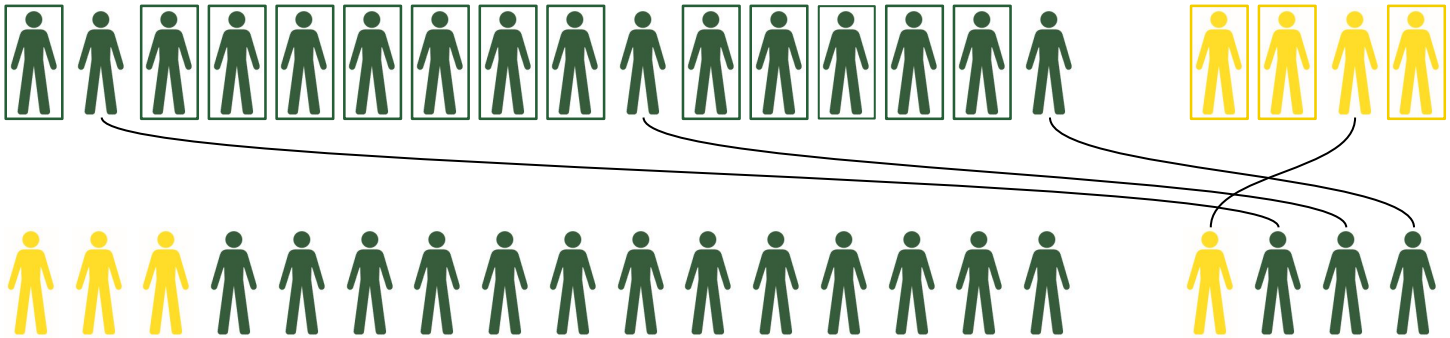


If we randomly split them into 2 groups (4:1), and the classes are not proportionally split, the model may learn an incorrect distribution that would not be true in testing or production.



3 Clean and Preprocess the Data • Data Segregation

Stratification preserves class distribution by randomly sampling within classes.



Bonus: Look into, and possible implement, stratified k-fold cross validation.

4 Choose an ML algorithm

We'll cover:

- Logistic Regression
- K-Nearest Neighbors (k-NN)
- Decision Trees and Random Forests

We won't cover, but are common:

- Support Vector Machine (SVM)
- K-Means Clustering
- Principal Component Analysis (PCA)

4 Choose an ML algorithm • Logistic Regression

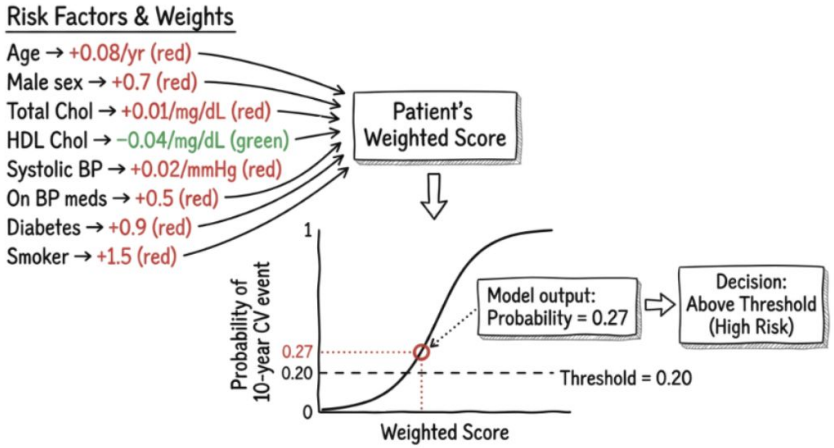
Let’s evaluate the risk of a cardiovascular event occurring in a patient within a decade.

1. While we evaluate all risk factors (**features**), we know that age has a higher correlation to risk than blood pressure, so we **weigh** it more heavily.
2. We then assign them a risk score, a **probability** of a CV event.
3. Depending on the risk score and **thresholds** we set, above which indicates high risk, we can classify them as high risk or not.

4 Choose an ML algorithm • Logistic Regression

Logistic regression learns ideal weights for each input feature and calculates a probability of the sample belonging to a certain class.

A threshold then determines if that sample meets the requirement to be within a class.



4 Choose an ML algorithm • Logistic Regression

Strengths

- Produces probabilities medical staff can interpret
- Lightweight and fast
- Requires few parameters
    - Works well even with limited data
- Good baseline model for risk prediction

Weaknesses

- Only captures linear relationships
    - Poor performance when characterizing complex interactions
- Sensitive to unscaled features (needs preprocessing)
- Can be misled by outliers

4 Choose an ML algorithm • K-Nearest Neighbors

Let’s try to diagnosing a patient with anemia.

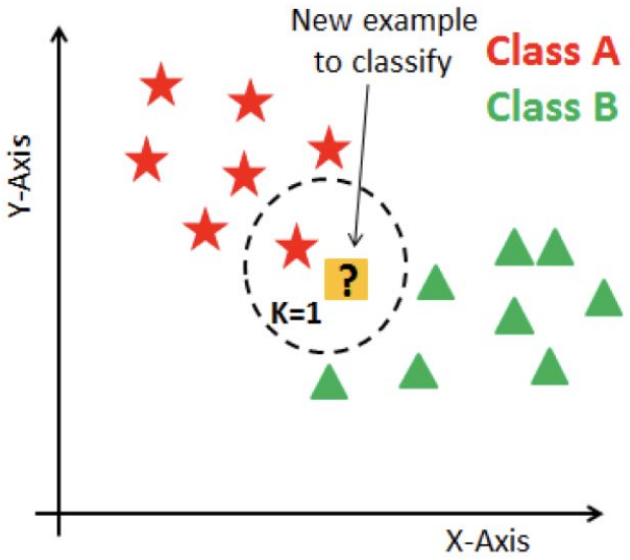
1. We have a database of 1,000 past patients that we can access, many of whom have very similar lab results as our patient.
2. We select ***k*** number of patient charts that are most similar to our patient and from those **neighbors**, we take the majority diagnosis.

4 Choose an ML algorithm • K-Nearest Neighbors

K-NN measures the similarity (distance) between features and predicts the most common label among *k* nearest neighbors.

Small *k*: more sensitive to noise and outliers

Large *k*: may be considering far and irrelevant neighbors



4 Choose an ML algorithm • K-Nearest Neighbors

Strengths

- Good for problems where similarity is related meaning
- Intuitive: “patients similar to this patient”
- No training required
- Works well for small datasets
- Can handle multi-class problems

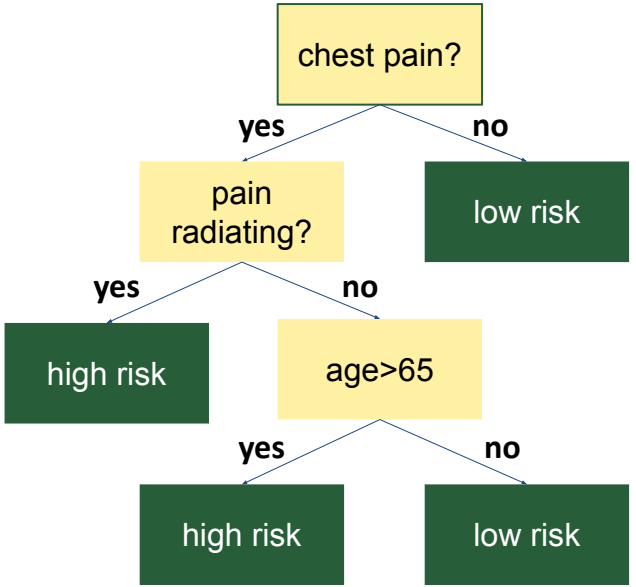
Weaknesses

- Slow for large datasets (must compare to all past data)
- Sensitive to irrelevant features
- Struggles in high dimensions (too much noise)

4 Choose an ML algorithm • Decision Trees

Decision trees learn to split data based on yes/no questions and makes predictions based on the leaf node it ends on.

Deeper trees tend to do better, but if it's too deep, it will specialize to the training data too much and will struggle with generalization.



4 Choose an ML algorithm • Decision Trees

Strengths

- Highly interpretable: mirrors clinical decision flowcharts
- Works with mixed data (numerical, binary, categorical)
- Captures non-linear relations
- Quick to train

Weaknesses

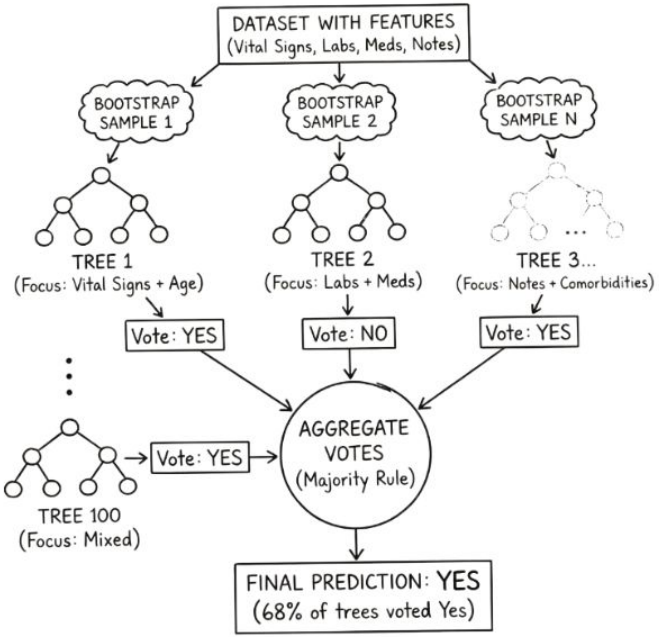
- Prone to overfitting if the tree grows too deep
- Unstable
- Struggles with subtle patterns spread across many variables
- Usually weaker than ensemble methods (Random Forest, XGBoost)

4 Choose an ML algorithm • Decision Trees

A random forest is a collection of decision trees, whereby each tree sees slightly different training data. The prediction is then made by majority vote among all the trees.

Forests are typically more accurate than a single tree, but the results are harder to interpret and training requires more computation.

Bonus: The notebook doesn't currently have a random forest. Try training one yourself!

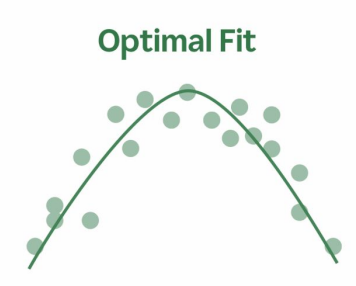


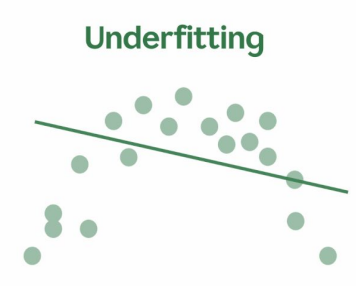
5 Model Evaluation • Fitting

An optimally fitted model captures underlying patterns and generalizes to unseen data.

Underfitting occurs when the model is too simple to capture underlying patterns in data.

- Poor performance on both training and test data
- Can be fixed with:
    - more training iterations, or
    - using a more complex model.



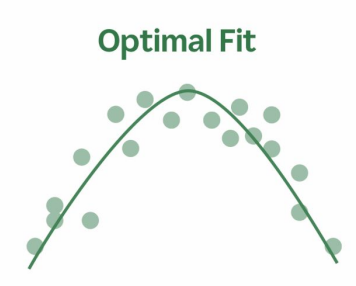


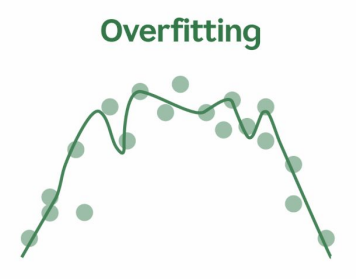
5 Model Evaluation • Fitting

An optimally fitted model captures underlying patterns and generalizes to unseen data.

Overfitting occurs when the model is too complex and memorizes noise in training data.

- Good performance in training but poor performance on test data.
- Can be fixed with:
    - less training iterations,
    - using a less complex model,
    - ensembling (combining) models,
    - or obtaining more data.



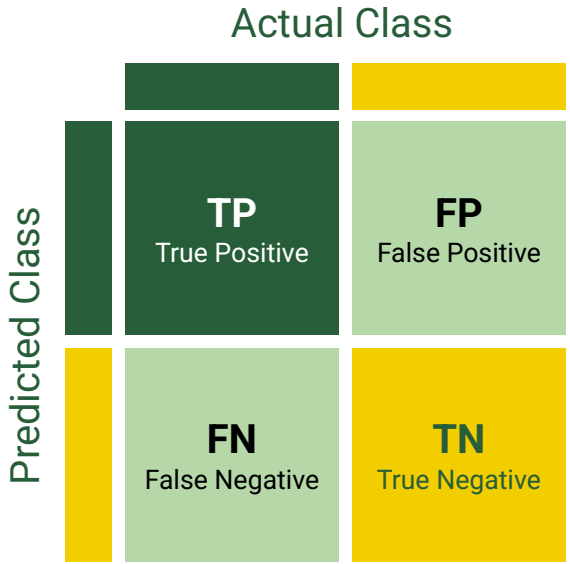


5 Model Evaluation • Confusion Matrix

Let’s use a email spam filter as an example:

TP: filter labelled it as spam and it is spam
FP: filter labelled it as spam but it is not spam
FN: filter labelled it as not spam but it is spam
TN: filter labelled it as not spam and it is not spam

In medicine, we typically prefer being more sensitive to FPs than FNs.
Assuming the risk of a missed diagnosis outweighs the side effects of improper treatment.



5 Model Evaluation • Metrics

Our goal is to maximize TPs and TNs.

Accuracy is the fraction of all samples that the model correctly predicted.

Precision is the fraction, per class, of positive predictions (TP+FP) that are actually positive (TP).

Recall (sensitivity) is the fraction, per class, of positives (TP+FN) that were classified correctly as positives (TP).
Recall is preferred in medicine as it measures the proportion of positives correctly detected.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

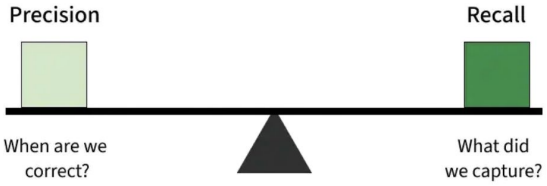
$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

https://developers.google.com/machine-learning/crash-course/classification/accuracy-precision-recall

5 Model Evaluation • Metrics

F1 score combines precision and recall into a single value.
Useful when the dataset has imbalanced classes.



$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

Macro metrics calculate the average metric across classes.

Bonus: There are many more metrics out there, find the one that's best for your needs!

Try It Yourself

You can now try tuning the hyperparameters and answer the consider questions.

If you finish early:
- Try using a different model (ex. random forest)
- Incorporate historical patient data (by changing the features list)
- Consider different metrics or visualizations (ex. AUC/ROC)

ask us if you have questions; we'll resume at 10:40

Activity: DL Segmentation



Gairdner
Alberta Partnered
AI + Health Research
Symposium

UNIVERSITY OF ALBERTA

gairdner
LES PRIX CANADA GAIRDNER AWARDS

AIMSS
ARTIFICIAL INTELLIGENCE IN MEDICAL SYSTEMS SOCIETY

https://drive.google.com/drive/folders/1SM2w8vvcdROr6Z07QohMChLCR_bLOCAQ?usp=drive_link

Why Deep Learning (DL)?

- Your problem is too complex
- You have a huge dataset with too many features
- Your data is unstructured (ex. text, sound, image)

Occam’s Razor

If your problem doesn’t need DL, try a classical ML model first. A lot of problems can still be solved using classical ML at a fraction of the data and computation, with possibly better performance.

Intro to DL • Neurons

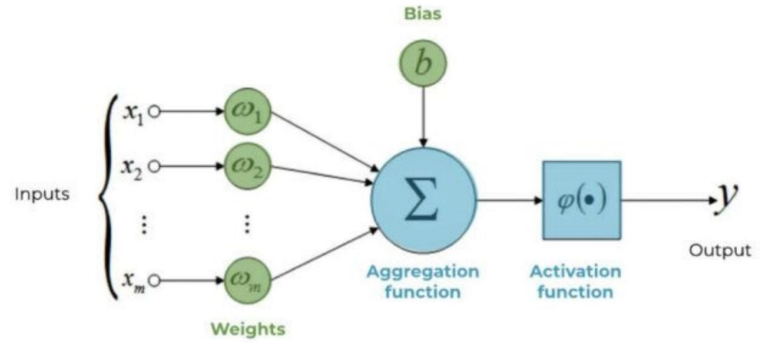
A patient is brought to the ED and it's your job to determine if they are of high risk.

1. While you evaluate all vitals (**inputs**), you know that abnormalities in respiratory and heart rate are higher risk than temperature deviations, so you **weigh** them more heavily.
2. You may also be more cautious and **bias** towards rating higher risk than your colleagues.
3. Then rather than assigning a patient a list of risk factors, you assign them a triage level (**activation function**).

You have just done a forward pass of a neuron!

Intro to DL • Neurons

Neurons calculate a weighted biased aggregate (ex. sum) and pass it to a activation function (ex. threshold) to make a singular prediction.



Learnable through time
Fixed by the AI expert

$$y = \varphi\left(b + \Sigma\, x_i\, \omega_i\right)$$

In our analogy we used pre-existing knowledge to determine weights. Over time the model may learn that the temperature feature is less important, but at the start it has no knowledge and assigns weights at random

Intro to DL • Neural Networks

But maybe you and another doctor disagree, and you need an added **layer** of accountability.

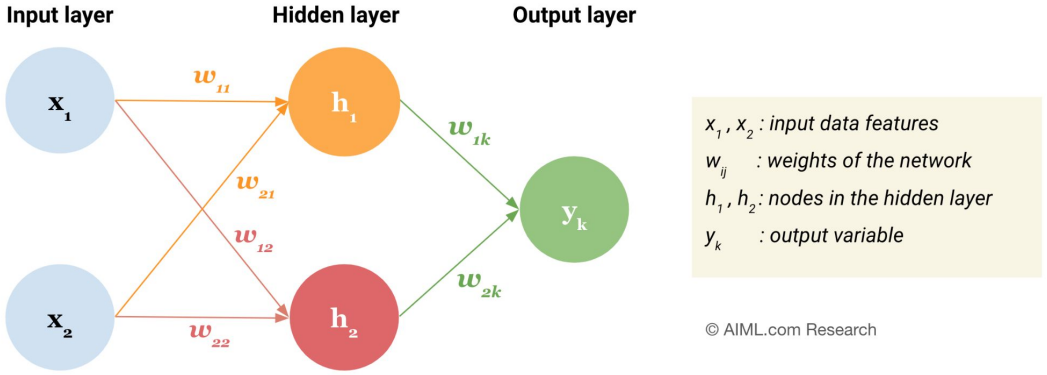
1. An attending takes both your assessments as **input**, **weighs** how much to trust each one, adds their own **bias**, and **outputs** a final triage level.
2. Afterwards, you compare the final score to that patient's outcome (if they crashed or not). The degree of error between **predicted** level and **target** reality is the **loss**.

You have just done a forward pass of a neural network!

Intro to DL • Neural Networks

A neural network (NN) is comprised of multiple layers of neurons.

The feature inputs are passed to neurons in hidden layers, which are fed into an output layer.



Loss is the error between the predicted and expected output.

As such, our goal is to minimize loss.

Intro to DL • Back Propagation

So you now know the **loss** between the attending's score and the actual outcome, but your assessment is **hidden** behind a final **layer** (attending's assessment).

How much of your individual assessment contributed to the final loss?

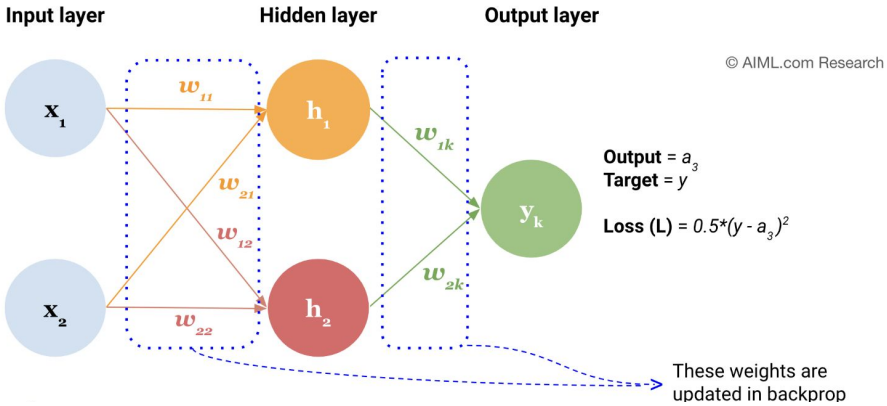
1. Settle how wrong the attending's trust weighting was.
2. Trickle down to how wrong both your underlying vital-sign weightings were.
3. Each person adjusts their own weightings to minimize loss.

You have just done a backwards pass of a network!

This repeats for multiple patients until the final triage level **converges** towards accurate predictions of risk.

Intro to DL • Back Propagation

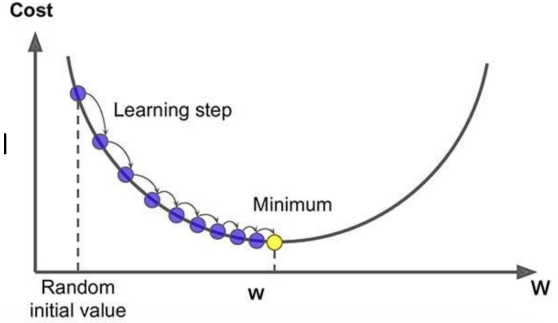
Backpropagation traces back the chain of responsibility so each neuron can adjust their weights to reduce overall loss.

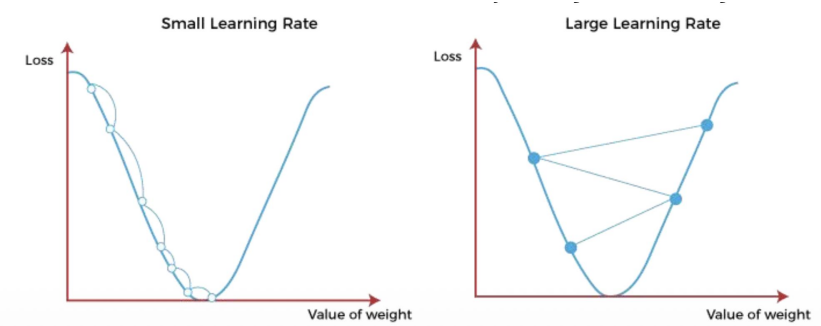


Intro to DL • Loss

A model learns by minimizing loss, but how?

Every time a weight needs to change, we step it by the learning rate (LR) towards the minimum loss (cost). This process is called gradient descent.
A small LR may take longer to converge, but a large LR may miss the minimum entirely.

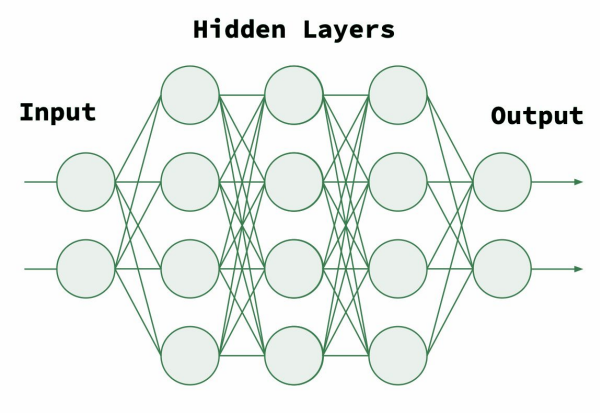




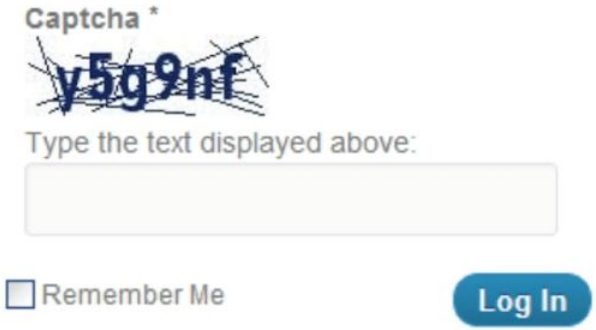
Bonus: LR decays fix the issue of small/large steps by decreasing step size with loss

Intro to DL • Summary

1. Weights initialized at random
2. Train
    a. Input data
    b. Propagate data through the network
    c. Output predictions
    d. Calculate loss
    e. Backpropagate to update weights
3. Repeat for *n* **epochs** (iterations) until loss is at a minimum
    a. If unsuccessful, adjust hyperparameters and retry
4. Your model is trained and ready to use



A Warning



How can you read the captcha?

- What characteristics are you choosing to ignore?
- What are you focusing on?
- Why do you focus on certain pixels?
- What makes a “g” different from the background or any other character?

A Warning

NNs are black boxes



> “What we did was we designed the learning algorithm. That's a bit like designing the principle of evolution. But when this learning algorithm then interacts with data, it produces complicated neural networks that are good at doing things. **But we don't really understand exactly how they do those things.**”
>
> - Geoffrey Hinton (“Godfather of AI”), Aug. 2025

https://www.cbsnews.com/news/geoffrey-hinton-ai-dangers-60-minutes-transcript/

A Note

DL is extremely memory and computation hungry!

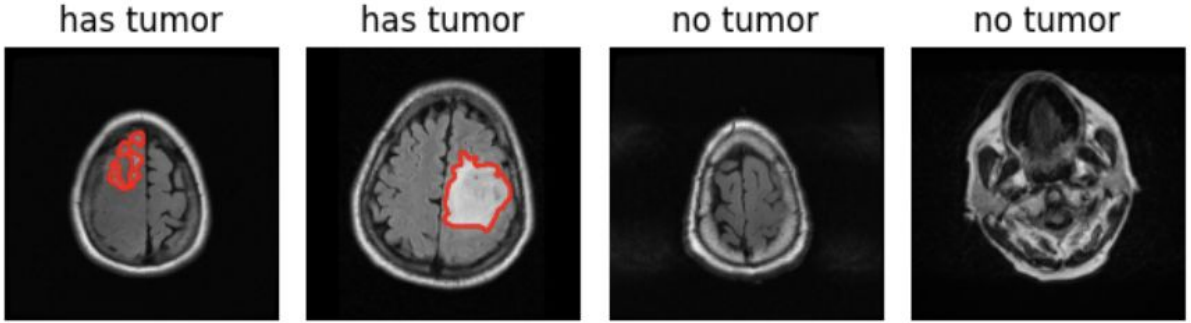
The slides will show results that take much longer to process than what is in the notebook. You are free to change the input dimensions and training epochs, but know that you may be waiting minutes to hours for your model to run.

1 Define a Problem

Given a brain MRI scan from a patient with a lower-grade glioma (LGG), can we automatically segment the tumorous region to match the boundary a radiologist would draw?

2 Collect/Gather Data

- Records ~3900 pre-operative LGG MRI slices (pre-contrast, FLAIR, and post-contrast) of 110 patients.
  - Subset of data collected by The Cancer Genome Atlas (TCGA) at Thomas Jefferson University, Henry Ford Hospital, UNC, Case Western, and Case Western-St. Joseph's.
- Segmentation masks of tumorous regions were marked by a board eligible radiologist.

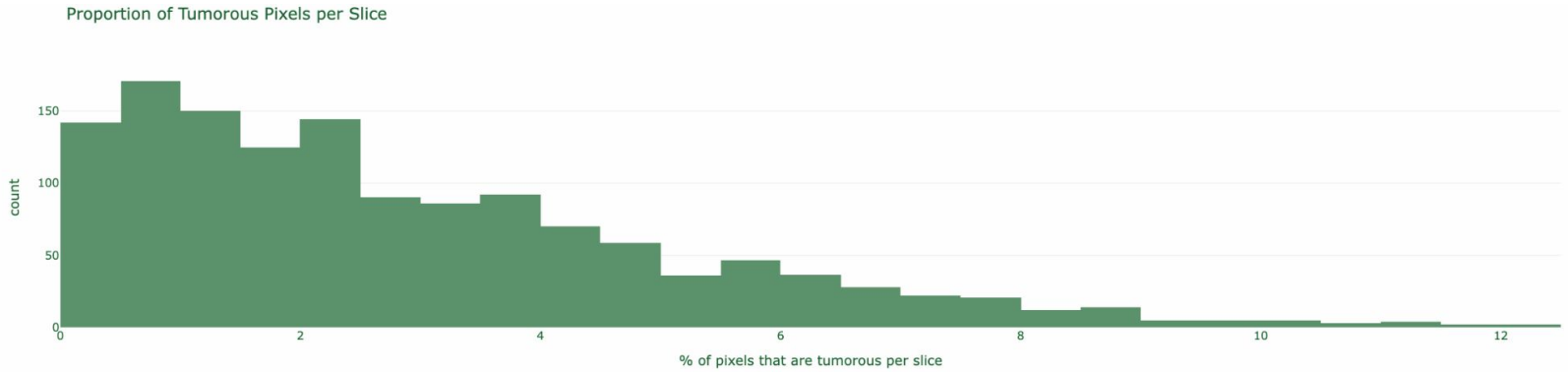


https://www.cancerimagingarchive.net/collection/tcga-lgg/

3 Clean and Preprocess the Data • EDA

34.95% of all slices contain a tumor.

Of the slices with a tumor, many have tumorous regions smaller than 5% of the slice.

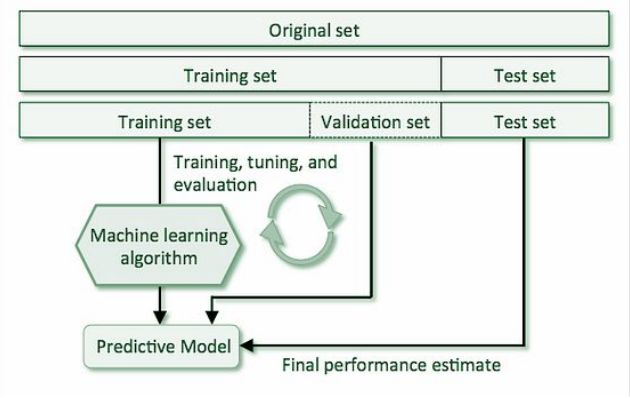


We’re skipping most of the EDA for brevity, but in reality, you would still do a lot of the steps previously mentioned in activity 1.

3 Clean and Preprocess the Data • Data Segregation

We previously split our data into two sets, but now we'll split into three:

1. Training
2. **Validation (Holdout)**: Used during training to fine tune hyperparameters
3. Testing: Used during testing to evaluate model generalization



Bonus: if you haven't already yet, look into cross validation.

4 Choose an Algorithm

We'll cover:

- Convolutional Neural Networks (CNNs)
- Encoder-Decoder Networks
- **Transfer Learning**

We won't cover, but are common:

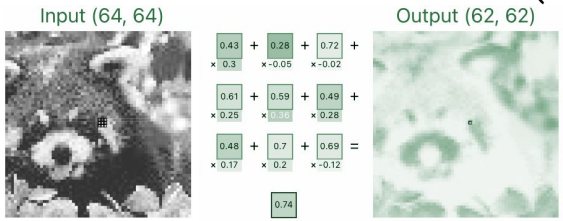
- Recurrent Neural Networks (RNNs)
- Long Short-Term Memory Networks (LSTMs)
- Generative Adversarial Networks (GANs)
- Transformers

Too many to list…

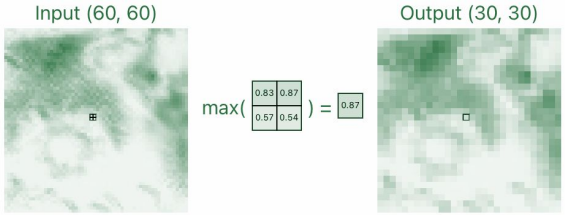
4 Choose an Algorithm • CNNs

Convolutional Neural Networks are made of 3 basic building blocks

Convolutional Layers: Learnable filters that extract features (ex. edges)



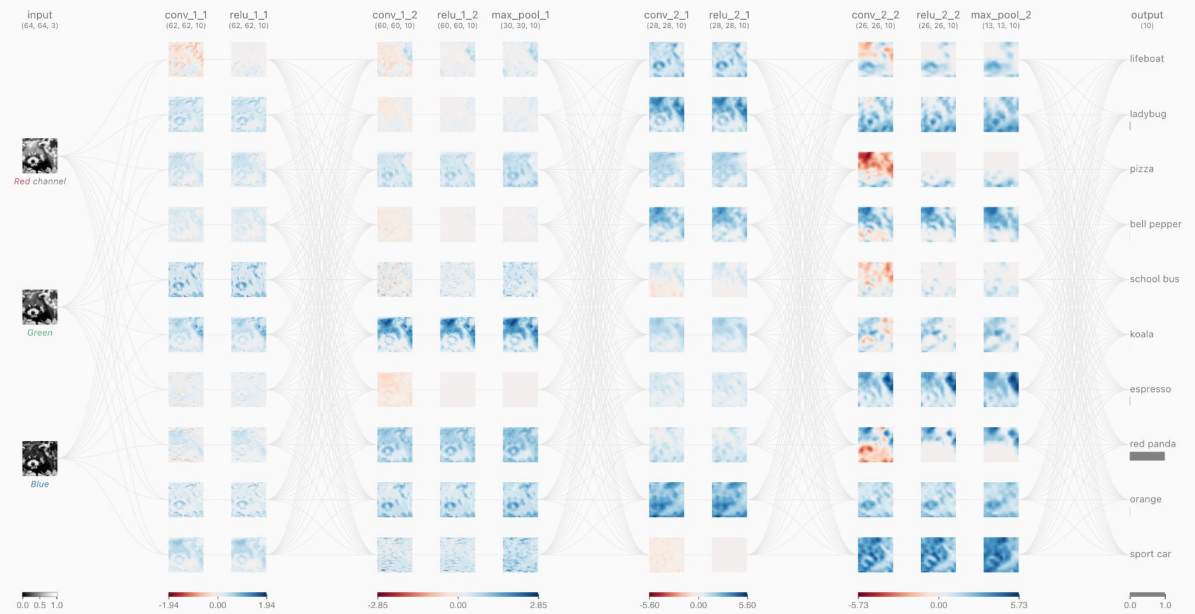
Pooling Layers: Increases the receptive field by downsizing the input



Fully Connected Layers: Makes the prediction

https://poloclub.github.io/cnn-explainer/

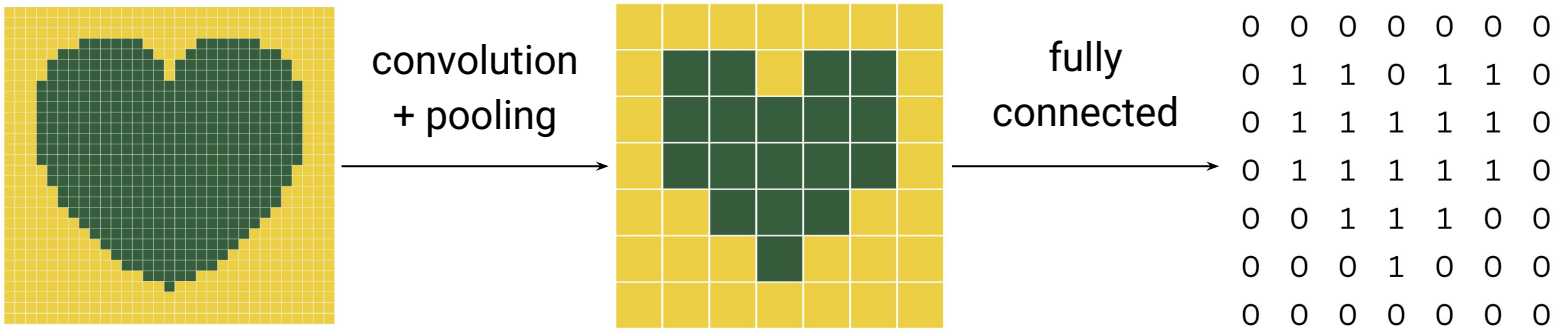
4 Choose an Algorithm • CNNs



https://poloclub.github.io/cnn-explainer/

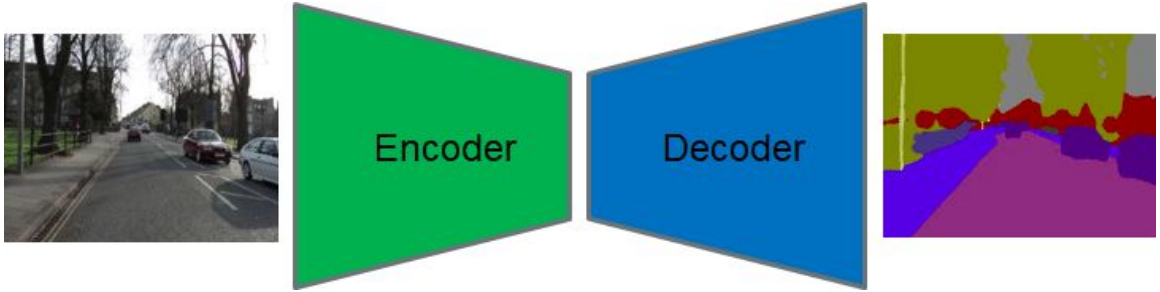
4 Choose an Algorithm • Encoder-Decoder Networks

Segmentation is a classification problem at the pixel level.
So the CNN can get us a mask, but in the process, our image size has been reduced to a fraction of its original dimensions.



4 Choose an Algorithm • Encoder-Decoder Networks

Encoder-Decoder networks add **learnable upsampling** to generate pixel-by-pixel masks of the same dimension as the original image.



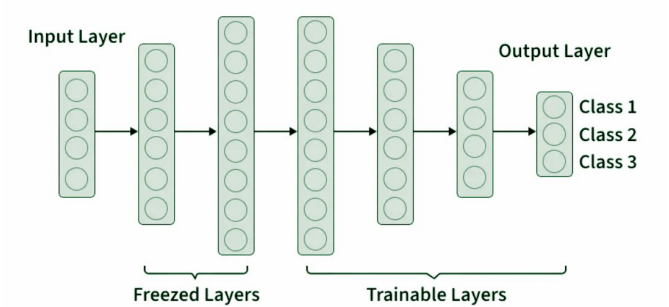
I won’t get into it, but the decoder blocks are effectively the inverse of the encoder blocks.

4 Choose an Algorithm • Transfer Learning

Maybe you don’t have the computational resources, time, or large enough dataset for DL.

Transfer learning allows you to use a model that has been pre-trained on a similar task. Emphasis on similar! A model trained on detecting airplanes, probably won’t do well if you want it to detect melanomas…

Since the model was pre-trained on different data, you have to fine-tune it by further training it on your data.

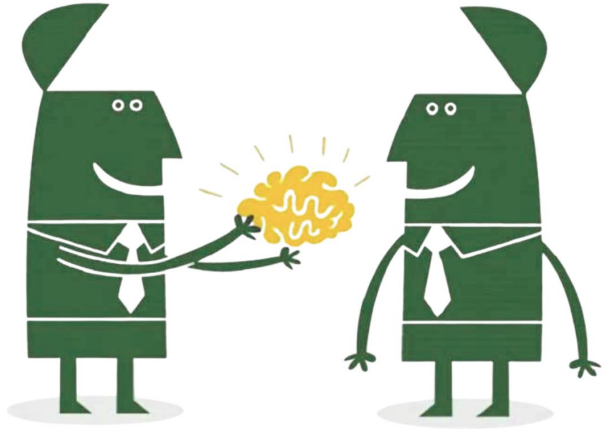


4 Choose an Algorithm • Transfer Learning

Model Cards are short, standardized documentation published alongside a trained model.

They typically include:
- intended use and out-of-scope uses,
- a description of the training data,
- evaluation results
- known limitations and failure modes, and
- ethical considerations.

Use them to judge whether it is reasonable to apply transfer learning for your problem.



https://huggingface.co/blog/model-cards

5 Model Evaluation

Binary Cross Entropy (BCE)

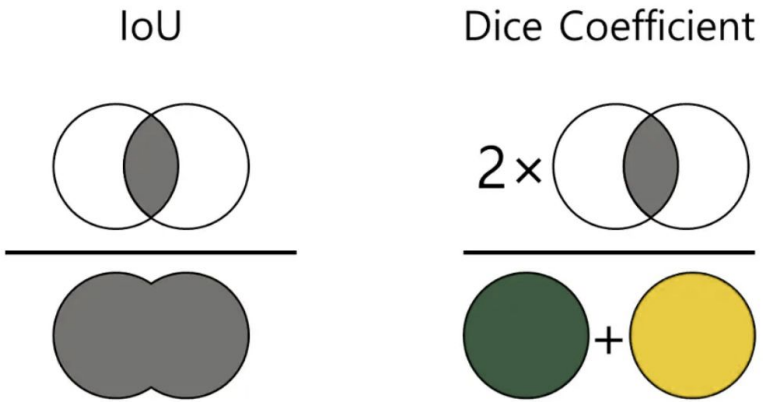
Measures the distance between the true labels and the predicted probabilities.

Intersection over Union (IoU)

Measures the overlap between predicted and ground truth segmentation masks.

Dice

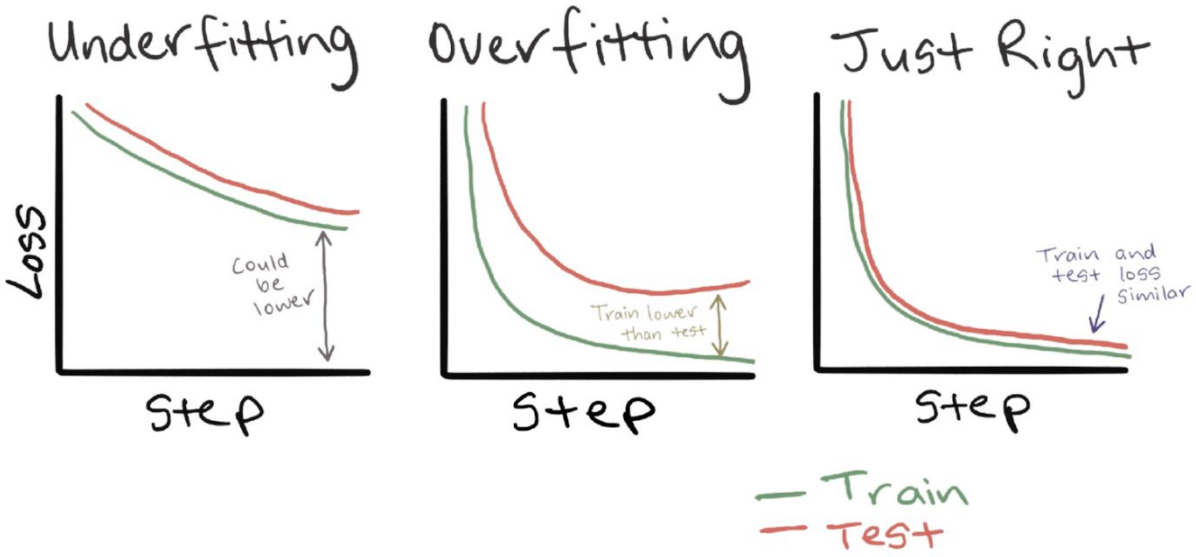
Emphasizes overlap more than IoU. It's commonly used in medical image segmentation where overlap is important.



https://alexkhrystoforov.github.io/ds-cheatsheet/pages/metrics/segmentation-metrics.html

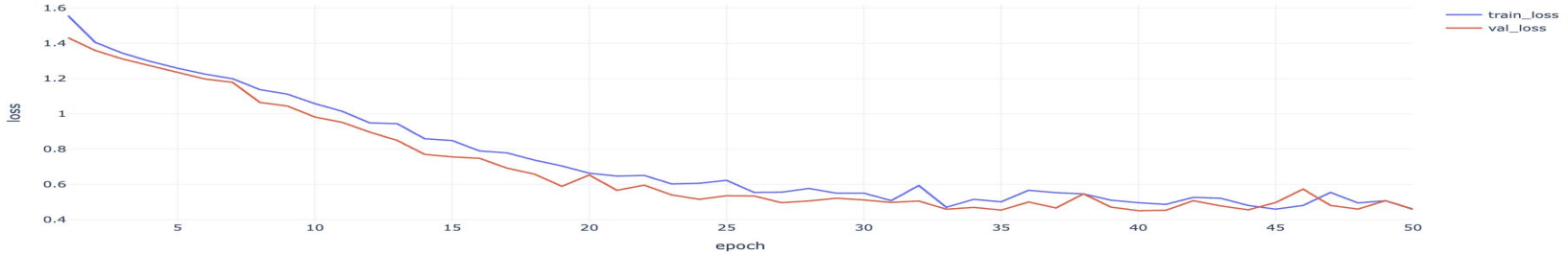
5 Model Evaluation

Learning curves plot the change in loss over epochs

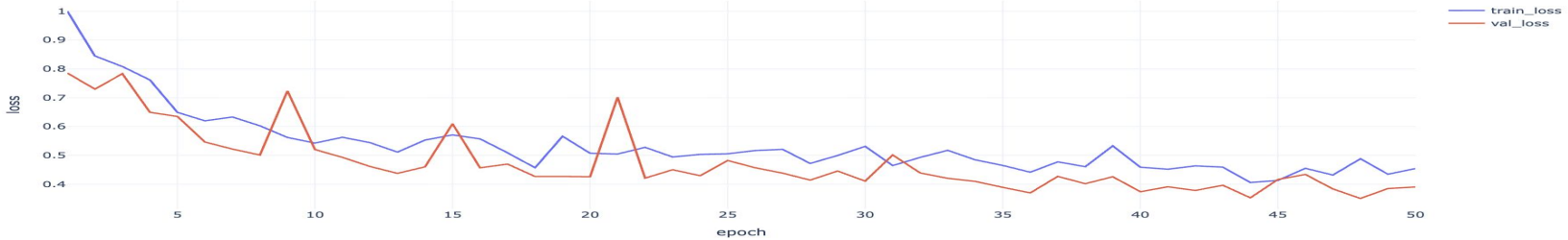


5 Model Evaluation

Learning Curve - Custom Model



Learning Curve - Pretrained Model



Notice that the pretrained model is a lot more unstable (jumpy).

5 Model Evaluation

Base Model

Model	dice	IoU
custom	0.637837	0.494535
pretrained	0.705462	0.579701

After Simple Hyperparameter Tuning

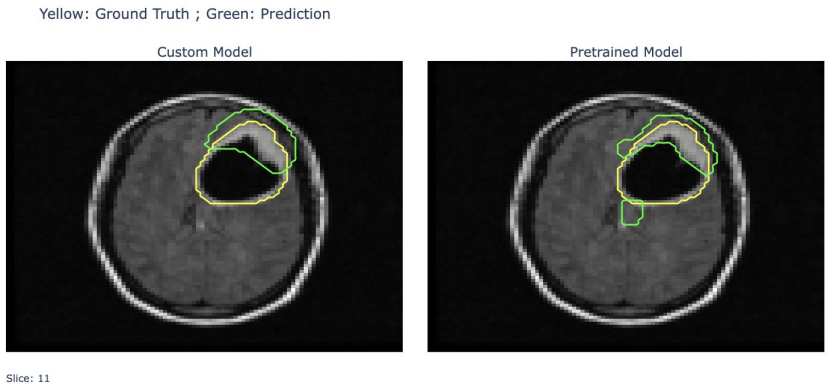
Model	dice	IoU
custom	0.672038	0.539965
pretrained	0.747102	0.622767

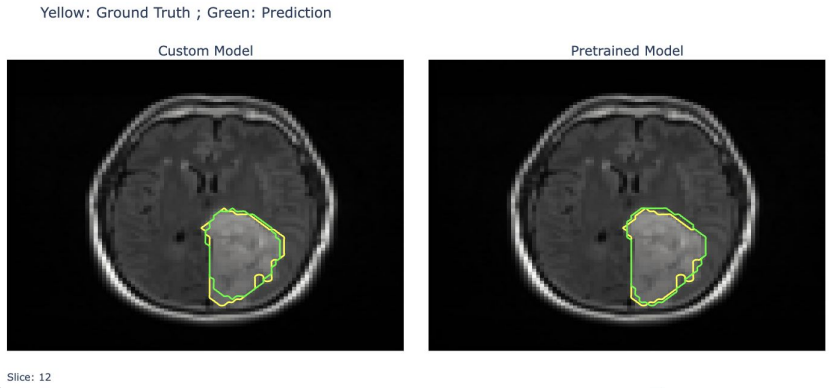
5 Model Evaluation

The fun part of computer vision: you get to look at your samples!

Seems the model may have learned to identify lighter regions as tumorous but ignores darker regions.

- Check if similar data is present in training
- Introduce **data augmentation**



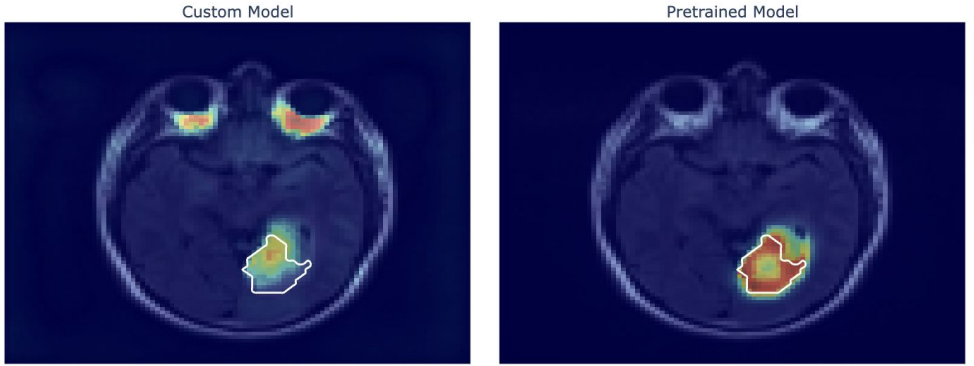


Bonus: Explainable AI (XAI)

XAI attempts to solve the black box problem by pinpointing features that contribute strongly to the output.

BUT:

- different XAI models give different explanations
- knowing what features are important doesn't explain why they are important



Nowadays if you search up XAI, you'll likely end up with something different…
Explainable AI is not related to Elon Musk's companies…

Try It Yourself

You can now try tuning the hyperparameters and answering the consider questions.

If you finish early:
- add data augmentation
- edit the custom model
- try finding a different model to transfer learn from (look into Hugging Face)

If you want to improve your programming practices (more advanced):
1. move the pipeline out of a notebook
2. move hyperparameters into a config (cfg) file
3. add model checkpoint saving
4. add logging

ask us if you have questions; we'll wrap up at 11:30

Thank You



Gairdner
Alberta Partnered
AI + Health Research
Symposium

UNIVERSITY OF ALBERTA

gairdner
LES PRIX CANADA GAIRDNER AWARDS

AIMSS
ARTIFICIAL INTELLIGENCE IN MEDICAL SYSTEMS SOCIETY

AI in Medicine Elective

AIMSS provides a yearly 6-part course on AI in Medicine.
It counts as elective hours for medical students, but anyone is invited to attend.

You can also find previous years' sessions online!

sign up to be notified of this year's sessions

Gairdner
Alberta Partnered
AI + Health Research
Symposium

UNIVERSITY OF ALBERTA

gairdner

AIMSS

https://aimss-ai-in-healthcare.streamlit.app

Connect With Us

www.aimss.ca

hello@aimss.ca



ying@aimss.ca



ehsan@aimss.ca

Gairdner
Alberta Partnered
AI + Health Research
Symposium

https://www.aimss.ca

References + Further Reading



Gairdner
Alberta Partnered
AI + Health Research
Symposium

UNIVERSITY OF ALBERTA

gairdner
LES PRIX CANADA GAIRDNER AWARDS

AIMSS
ARTIFICIAL INTELLIGENCE IN MEDICAL SYSTEMS SOCIETY

References

- https://learning.tcairem.ai/terminology/
- https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0201016#sec028
- https://www.cbsnews.com/news/geoffrey-hinton-ai-dangers-60-minutes-transcript/
- https://www.cancerimagingarchive.net/collection/tcga-lgg/
- https://poloclub.github.io/cnn-explainer/
- https://huggingface.co/blog/model-cards
- https://alexkhrystoforov.github.io/ds-cheatsheet/pages/metrics/segmentation-metrics.html

Further Reading

Terminology
- https://learning.tcairem.ai/terminology/

Learning
- https://aimss-ai-in-healthcare.streamlit.app
- https://poloclub.github.io/cnn-explainer/
- https://alexkhrystoforov.github.io/ds-cheatsheet/index.html
- https://covizdemo.vercel.app

Mathematical Principles
- https://www.youtube.com/@3blue1brown

Programming
- https://pandas.pydata.org/docs/
- https://scikit-learn.org/stable/index.html
- https://docs.pytorch.org/docs/2.14/index.html