From Pixels to Patients: Revolutionizing Medical Care with AI

Martin J. McKeown
BEng | MD | FRCP
Professor and Head, Division of Neurology
John Nichol Chair in Parkinson's Research

martin.mckeown@ubc.ca
https://mckeownlab.ca/

Parkinson's Foundation
CENTER OF EXCELLENCE

THE UNIVERSITY OF BRITISH COLUMBIA
a place of mind

Disclosures

- Have current grants from NSERC, CIHR, Focused Ultrasound Foundation, Pacific Parkinson's Research Institute, and the VGH Foundation
- Am a *Honorary Visiting Professor* at **Nanyang Technological University's Lee Kong Chian School of Medicine,** Singapore
- Have received honoraria for speaking and travel support from Abbvie, Merz

- I am not a Computer Scientist, but a clinician-scientist
- I am an *adult* neurologist

Learning Objectives

- Examine the transformative potential of artificial intelligence (AI) in clinical medicine
- Critically assess the challenges in the implementation of AI
- Analyze the current limitations of AI technologies
  - explore future directions

Outline:

Basics of AI and Machine Learning

The Good

- Examples of the Transformative Potential of AI

The Bad

- Limitations
- Risks

The Future

Outline:

Basics of AI and Machine Learning

The Good

- Examples of the Transformative Potential of AI

The Bad

- Limitations
- Risks

The Future

Artificial Intelligence (AI): Why Now?

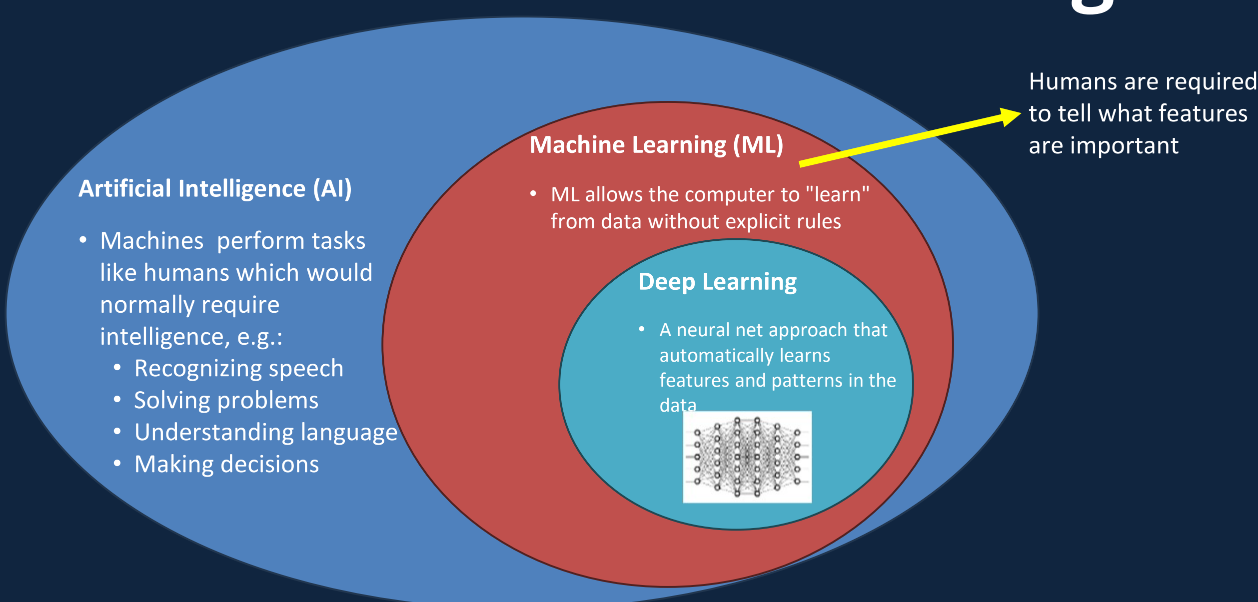
- **MASSIVE increases in:**
  - Computing power
  - Digital data:
    - MRIs and other imaging
    - Genetics
    - Electronic Medical Records
    - *In neonatal intensive care units (NICUs), estimated data generated:* ***terabyte/****bed/yr* (~ 100 HD movies)
- Quantum Improvements in Pattern Recognition
  - Deep Learning (e.g. Speech Recognition)
- Improvements in information transfer
  - 5G

Nomenclature -- ***Features***

Features: a characteristic or clue that helps the computer recognize patterns

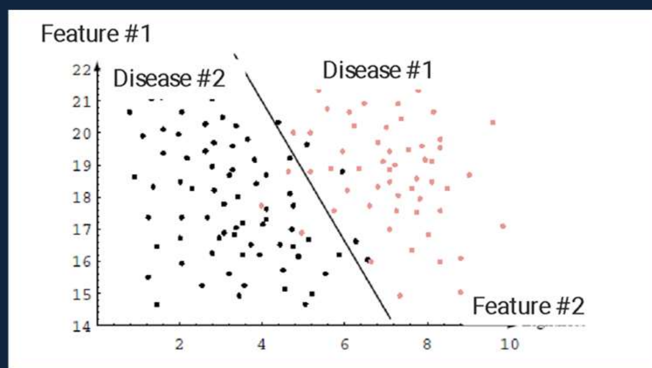
- demographics
- "signs and symptoms"
  - fever
  - headache
  - …

AI and Machine Learning



Nomenclature -- *Classifier*

Classifier : a computational machine that takes in ***features*** as inputs and returns ***labels (e.g. "Disease X")***



Nomenclature -- *Classifier*

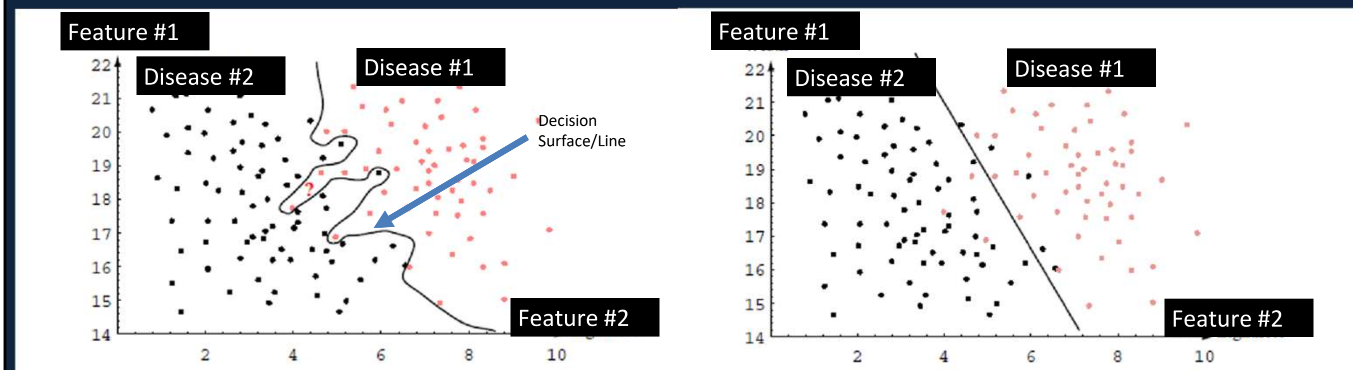
Classifier : a computational machine that takes in ***features*** as inputs and returns ***labels***

Aphasia
R hemiplegia
R extensor plantar response

Normally, these are "hand chosen" through years of experience and "expert knowledge"

L Middle Cerebral Artery stroke

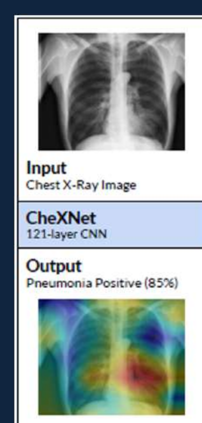
Nomenclature -- *Overfitting*



Need to test model on data different from what it was trained on!

Supervised vs *unsupervised* learning

Supervised: “here are a million examples of CXRs with **labels of** either ‘normal’ or ‘pneumonia’. Try to learn so that if I show you a **new** CXR, you can **classify** it as either ‘normal’ or ‘pneumonia’ ”



Unsupervised: “here are a million examples of CXRs with **no labels**. Determine if there is a natural clustering of CXRs”

Outline:

Basics of AI and Machine Learning

The Good

- Examples of the Transformative Potential of AI

The Bad

- Limitations
- Risks

The Future

Why is AI suitable for medical applications?

Much of clinical medicine is ***pattern recognition*** *("gestalt")*:

- E.g. during residency a radiologist may review 100s-1,000s of chest x-rays.
  - **A recent study trained a model on 165,988 CXRs !**

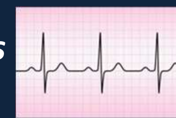
Many traditional words of medical wisdom can be rigorously integrated into AI models:

- "The plural of anecdote is not data"
- "When you hear hoofbeats, think horses, not zebras"
- "Atypical presentations of common diseases is more common than typical presentations of rare diseases"
- "When you have eliminated the impossible, whatever remains, however improbable, must be the truth" (Sherlock Holmes)

Fan, W., Yang, Y., Qi, J. *et al.* A deep-learning-based framework for identifying and localizing multiple abnormalities and assessing cardiomegaly in chest X-ray. *Nat*

Why is AI suitable for medical applications?

- Humans are excellent pattern recognizers in ***low-dimensional spaces*** (e.g. looking at a single EKG trace)



  - However they are poor at **high dimensional spaces**
    - Combining data from many different sources simultaneously
      - Demographic data
      - Past medical history
      - Physical examination findings
      - Vital signs
      - Monitor data
      - Laboratory results
      - Outcome data
      - Genetic data
      - ….

    - Never gets tired, never needs coffee, doesn't mind call …

Examples of AI approaches

Classification of Bilirubin Levels to Detect Jaundice from smartphone images
- smartphone picture with colour calibration card
- estimated bilirubin levels with an accuracy of 85% compared to blood tests

Detection of Necrotizing Enterocolitis
- microbial data from infants' stool samples
- can also be applied longitudinally, to increase accuracy
- can predict NEC 8 days before the disease starts with a specificity of 86% and a sensitivity of 90%

https://doi.org/10.1007/s10916-016-0523-4
https://doi.org/10.1186/s12859-022-04618-w

Examples of AI approaches

Detection of Abnormal Otoscopic Images

- A modified otoscope with an attached smartphone is used to take the images
- Could classify images as normal/abnormal with a sensitivity of 99% and specificity of 95.2%
- Wax plugs detected with 100% sensitivity and 97.7% specificity

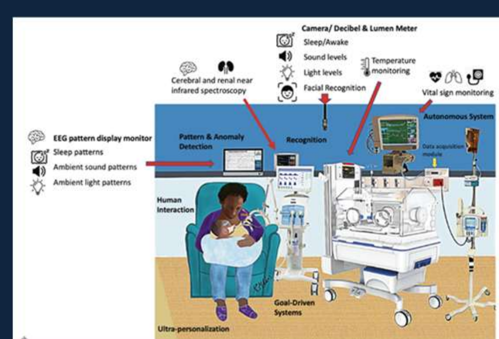
https://doi.org/10.3389/fped.2020.00525

Examples of AI approaches

Prediction of neonatal disease in NICU

Machine learning algorithms have shown breakthrough performance in predicting neonatal disease, such as:

- Sepsis
- Bronchopulmonary dysplasia
- Intraventricular hemorrhage
- Necrotizing enterocolitis
- Retinopathy of prematurity



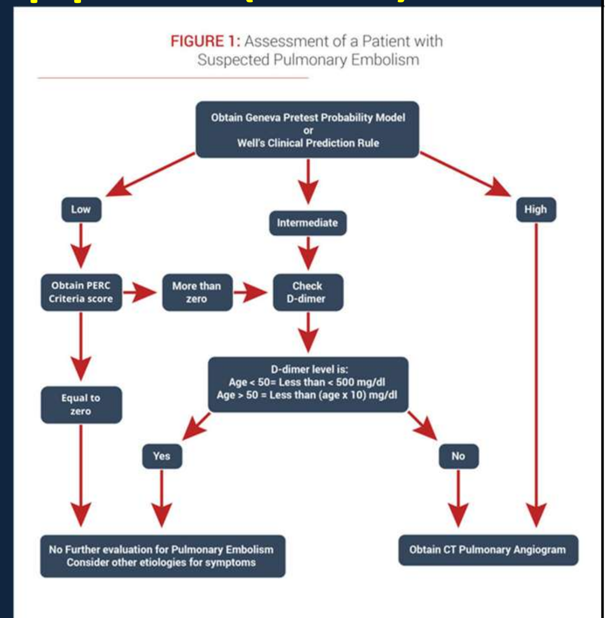
Patterns of AI relevant in the NICU: a focus on bronchopulmonary dysplasia

McAdams, R.M., Kaur, R., Sun, Y. et al. Predicting clinical outcomes using artificial intelligence and machine learning in neonatal intensive care units: a systematic review. J Perinatol 42, 1561–1575 (2022).

Clinical decision support (CDS)

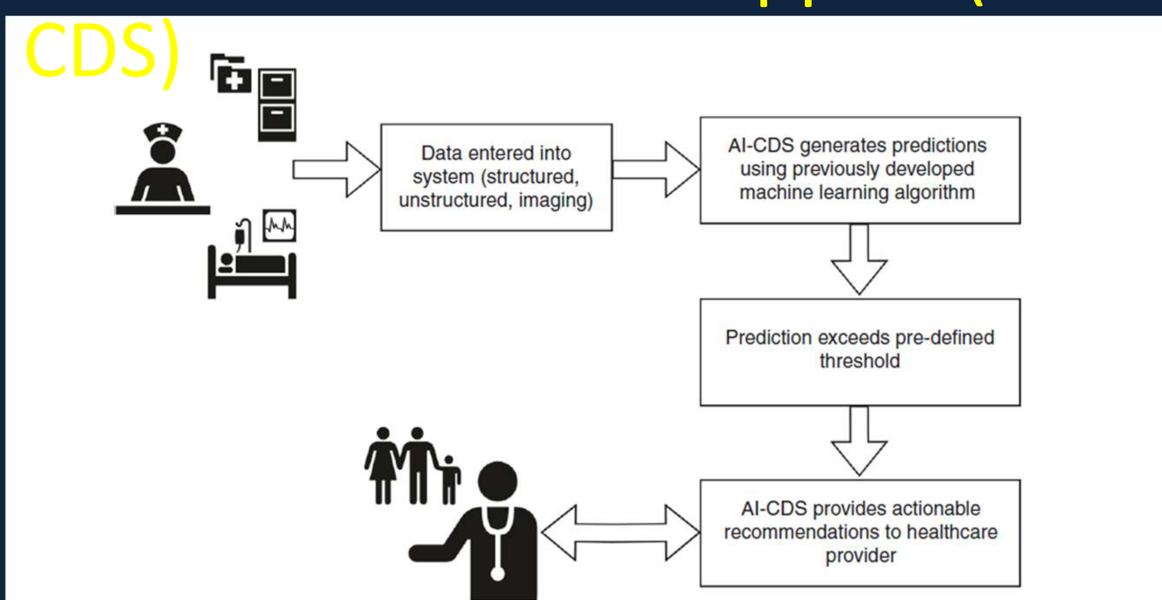
- Traditional, **rule-based** CDS followed distinct pathways
- AI-CDS, or Artificial Intelligence-Clinical Decision Support, differs significantly in the way it generates predictions and recommendations

Doesn't scale well!



https://mypulmonaryembolism.com/flow-chart/

AI-Clinical decision support (AI-CDS)



Ramgopal, Sriram, et al. "Artificial intelligence-based clinical decision support in pediatrics." *Pediatric research* 93.2 (2023): 334-341.

AI-Clinical decision support (AI-CDS)

Potential benefits:

Increased Model Accuracy:
- AI-CDS learn and adapt based on the data they are trained on
  - higher accuracy compared to rule-based systems

Fewer False Alerts and Missed Patients:
- Improved accuracy of AI-CDS models → reduction in false-positive alerts
  - greater physician satisfaction
  - reduced alarm fatigue → enhanced patient safety

Enhanced Handling of High-Dimensional Data:
- AI-CDS well-suited to handle datasets with a large number of variables

Examples of AI-CDS

Prediction of Clinically Important Traumatic Brain Injury (TBI):

- Pediatric Emergency Care Applied Research Network (PECARN) has rules to identify children at ***very low, intermediate***, and ***high risk*** of clinically important traumatic brain injury

- Reanalyzed data from 42,412 children:
  - identified more children at ***very low risk*** without missing more patients with clinically important traumatic brain injury.

https://www.youtube.com/watch?v=fgBI7P-czCY

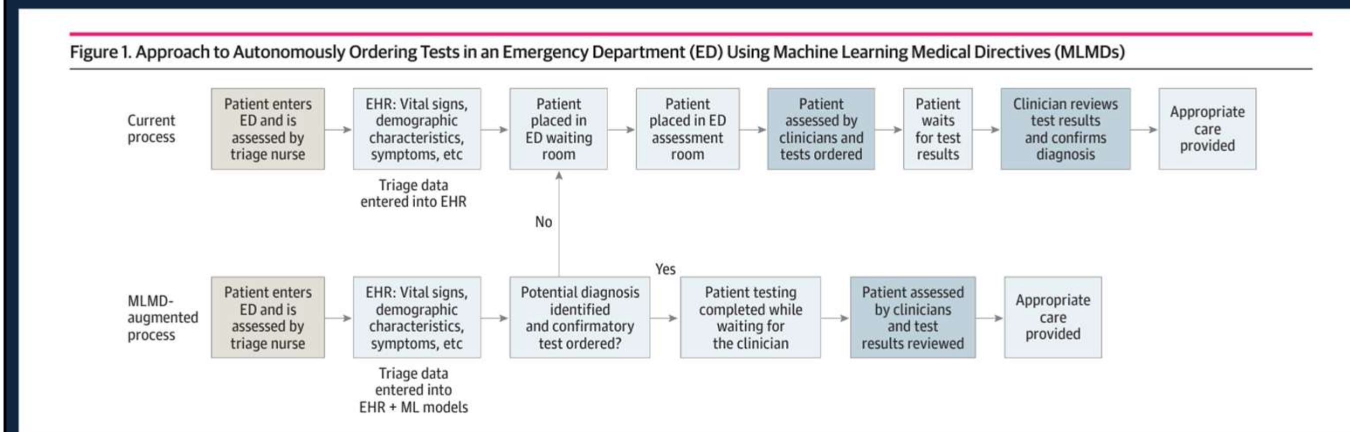
Bertsimas, Dimitris, et al. "Comparison of machine learning optimal classification trees with the pediatric emergency care applied research network head trauma decision rules." *JAMA pediatrics* 173.7 (2019): 648-656.

Examples of AI-CDS

Can Emergency Department Testing be streamlined?

- Urinary dipstick testing
- Electrocardiograms
- Abdominal ultrasonography
- Testicular ultrasonography
- Bilirubin level testing
- Forearm radiographs

- Positive predictive values (0.77-0.94) across each of the use cases



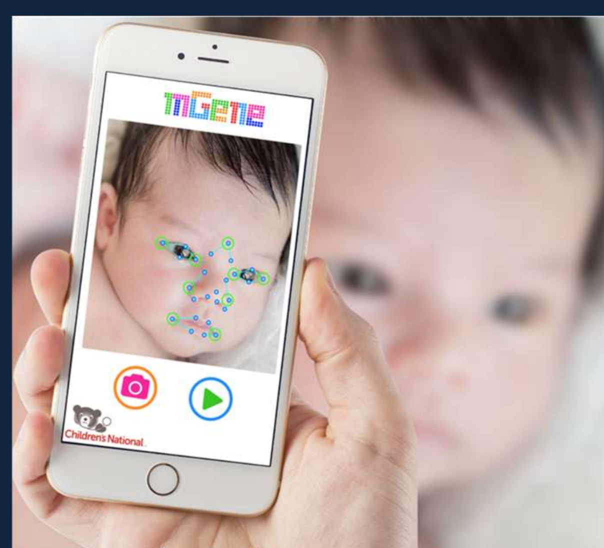
Singh, Devin, et al. "Assessment of machine learning–based medical directives to expedite care in pediatric emergency medicine." *JAMA Network Open* 5.3 (2022): e222599-e222599.

Facial Dysmorphism

Can Detect different syndromes:

- Down
- DiGeorge
- Williams
- Noonan

- > 90% accuracy

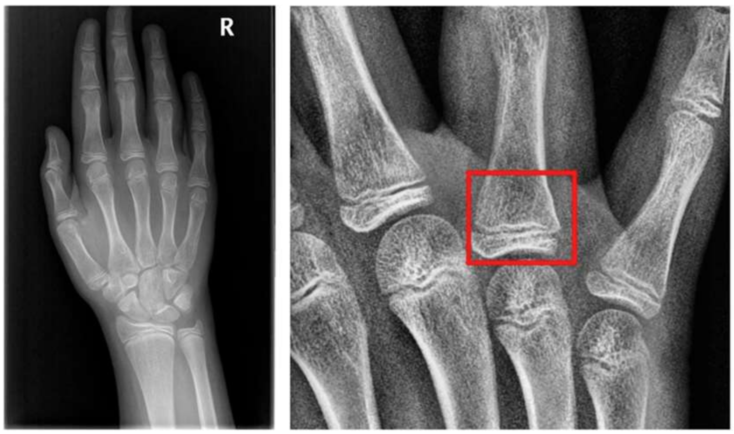


https://www.pbs.org/wgbh/nova/article/how-ai-is-helping-doctors-diagnose-and-treat-patients/

Imaging - Fractures

Fig. 1

From: Artificial intelligence improves resident detection of pediatric and young adult upper extremity fractures



a b

Thirteen-year-old male

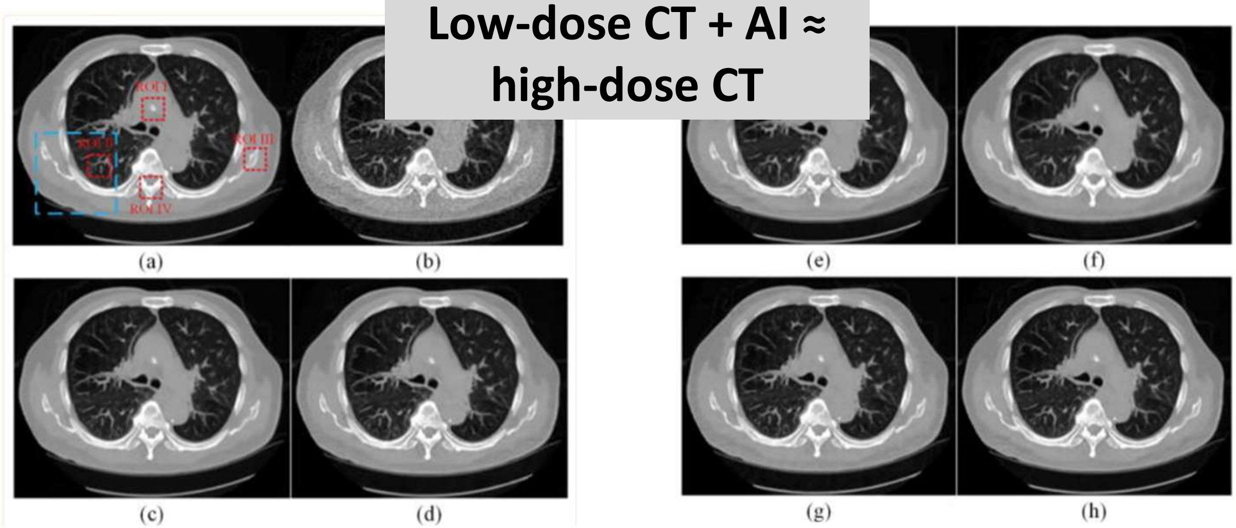
A Original radiograph and
B cropped, magnified, contrast-enhanced, and AI-annotated radiograph positive for incomplete fracture of the **ulnar aspect of the base of the fourth proximal phalanx**. The AI identified this fracture, and access to its prediction increased overall resident accuracy from 0% (n=0/5) to 80% (n=4/5). This patient was clinically diagnosed with an occult fracture and splinted.

Solid red box = AI region of high fracture probability (50-100%)

Zech, J.R., Ezuma, C.O., Patel, S. *et al.* Artificial intelligence improves resident detection of pediatric and young adult upper extremity fractures. *Skeletal Radiol* 53, 2643–2651 (2024)

Imaging – Low-Dose CT

Low-dose CT + AI ≈ high-dose CT



IEEE Trans Med Imaging. 2017 Jun 13;36(12):2524–2535

AI & Voice as a biomarker

Vocalization is a rich source of information:
- Speech
- Cry
- Cough
- Other respiratory sounds

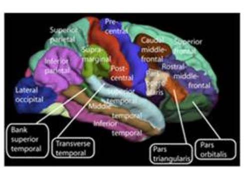
Over 62 studies conducted across 25 countries

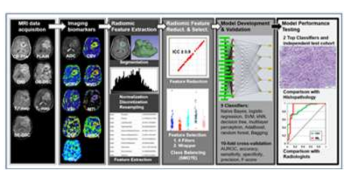
Most Studies so far on:
- Autism spectrum disorder (ASD)
- Intellectual disabilities
- Asthma

Rogers, Hannah Paige, et al. "Voice as a Biomarker of Pediatric Health: A Scoping Review." *Children* 11.6 (2024): 684.

AI & Imaging – Brain Tumours

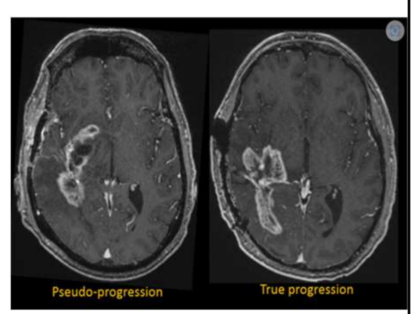
1) Use Multimodal assessment of borders of "eloquent cortex" for surgical planning





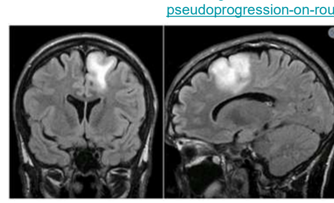
https://www.mdpi.com/2072-6694/14/10/2363

2) Differentiate tumor progression from effects of radiation and chemotherapy ("pseudoprogression")



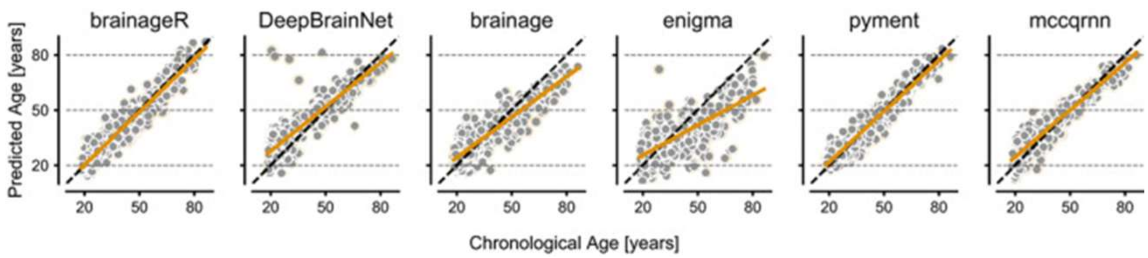
https://consultqd.clevelandclinic.org/using-machine-learning-to-distinguish-brain-tumor-progression-from-pseudoprogression-on-routine-mri/

3) Predict cognitive changes after chemotherapy
- Should we proceed or not with chemo?



https://case.edu/med/neurology/NR/Glioma

AI & Imaging: “Brain Age”



Estimating “brain age” from MRI is now available via several software packages

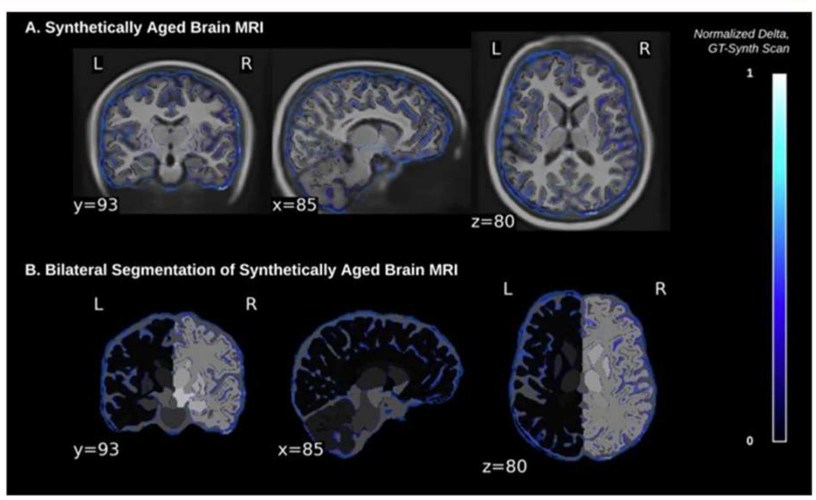
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10619370/

AI & Imaging: “Brain Age”

SynthBrainGrow: Synthetic Diffusion Brain Aging for Longitudinal MRI Data Generation in Young People

“Use these many examples of normal brain development to predict what *this particular MRI* will look like in 2 years time…”

Data from Adolescent Brain Cognitive Development (ABCD) study projected 2yrs into the future



https://arxiv.org/pdf/2405.00682

AI & Epilepsy

JAMA Neurology | Original Investigation

Automated Interpretation of Clinical Electroencephalograms Using Artificial Intelligence

Jesper Tveit, PhD; Harald Aurlien, MD, PhD; Sergey Plis, PhD; Vince D. Calhoun, PhD; William O. Tatum, DO; Donald L. Schomer, MD; Vibeke Arntsen, MD; Fieke Cox, MD, PhD; Firas Fahoum, MD; William B. Gallentine, DO; Elena Gardella, MD, PhD; Cecil D. Hahn, MD; Aatif M. Husain, MD; Sudha Kessler, MD; Mustafa Aykut Kural, MD, PhD; Fábio A. Nascimento, MD; Hatice Tankisi, MD, PhD; Line B. Ulvin, MD; Richard Wennberg, MD, PhD; Sándor Beniczky, MD, PhD

- SCORE-AI achieved human expert-level performance in fully automated interpretation of routine EEGs.
- Validated by:
  - a multicenter data set of 100 representative EEGs evaluated by 11 experts
  - a single-center data set of 9785 EEGs evaluated by 14 experts

- Current trends:
  - Automated EEG analysis
- MRI analysis for cortical dysplasia
- Better data fusion between EEG, MRI, PET scans

Large Language Models (LLMs)

ALL DOCTORS HATE DOCUMENTATION!

- But necessary: old adage:
  "If it isn't documented, it didn't happen!"

Microsoft's Nuance reveals clinician note-taking tool with GPT-4

Nuance Communications said the app creates notes based on patient conversations that clinicians can then review.

By Emily Olsen | March 21, 2023 | 11:35 am

32

Large Language Models (LLMs)

Clinical Documentation

- Drafting letters
- Prior authorizations - quickly generate detailed, structured summaries of medical necessity
- Patient education materials
- Help clinicians draft messages to patients

Research and Data Analysis

- Quality improvement initiatives by reviewing narrative charts and extracting key information

Privacy and Security Applications

- Preventing unintentional disclosures of protected health information by identifying confidential content within adolescents' clinical notes

33

AI in Drug Development

Milestones

2020

- First-ever AI-designed drug molecule (for obsessive-compulsive disorder) entered human clinical trials.

2021

- An AI system predicted the protein structures for 330,000 proteins
  - Expanded to include over 200 million proteins -- all cataloged proteins known to science.

2022

- Start of Phase I clinical trials for the first-ever AI-discovered molecule based on an AI-discovered novel target (Idiopathic Pulmonary Fibrosis)

2023

- Created and validated *de novo* antibodies *in silico* using generative AI

Pharmacogenomics (PGx)

AI is being utilized to optimize drug treatments for pediatric patients based on their genetic profiles

- Drug-gene interactions involving cytochrome P450 (CYP450) enzymes
  - Ondansetron
  - CYP2D6 polymorphisms for opioids
- Optimize dosing of pantoprazole and lansoprazole in children

Outline:

Basics of AI and Machine Learning

The Good

- Examples of the Transformative Potential of AI

The Bad

- Limitations
- Risks

The Future

What data is required for AI

Structured data

- Organized within predefined fields, making it easily searchable and analyzable, e.g.
  - vital signs
  - laboratory results
  - diagnosis codes

What data is required for AI

Structured data

- Organized within predefined fields, making it easily searchable and analyzable, e.g.
  - vital signs
  - laboratory results
  - diagnosis codes.

Unstructured data (less organized and often requires additional processing techniques)

- Clinician notes
  - Natural Language Processing (NLP) can help translate free text
- Imaging reports
  - **Neural networks** and **deep learning models** can be employed to automatically interpret imaging data

What data is required for AI

Structured data
- Organized within predefined fields, making it easily searchable and analyzable, e.g.
  - vital signs
  - laboratory results
  - diagnosis codes.

Unstructured data (less organized and often requires additional processing techniques)
- Clinician notes
  - Natural Language Processing (NLP) can help translate free text
- Imaging reports
  - **Neural networks** and **deep learning models** can be employed to automatically interpret imaging data

In the future: data collection through prospective trials:
- Dedicated research studies specifically designed for ML applications can provide high-quality, detailed data on pediatric populations.

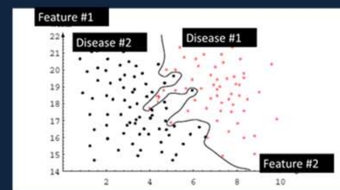
What data is required for AI

- Data driven approaches require LOTS and LOTS of data
  - Relative scarcity of large pediatric datasets compared to adult populations
- Ideal Data is **detailed**, including:
  - demographic data
  - extensive clinical data
    - past medical history, physical examination findings, vital signs , monitor data , laboratory results, outcome data
  - (this granular level of detail is crucial for ML algorithms to effectively learn patterns and make accurate predictions)
- Administrative datasets, (e.g billing codes), often lack this necessary granularity

Getting Labels is Expensive !

- **Supervised learning** requires (lots and lots of) **labelled data**
  - (What happens if some of the labels are incorrect)?

Engineers assume that the "domain experts" – i.e., clinicians --- are providing 100% accurate labels



JAMA Neurology | Original Investigation

Automated Interpretation of Clinical Electroencephalograms Using Artificial Intelligence

Jesper Tveit, PhD; Harald Aurlien, MD, PhD; Sergey Plis, PhD; Vince D. Calhoun, PhD; William O. Tatum, DO; Donald L. Schomer, MD; Vibeke Arntsen, MD; Fieke Cox, MD, PhD; Firas Fahoum, MD; William B. Gallentine, DO; Elena Gardella, MD, PhD; Cecil D. Hahn, MD; Aatif M. Husain, MD; Sudha Kessler, MD; Mustafa Aykut Kural, MD, PhD; Fábio A. Nascimento, MD; Hatice Tankisi, MD, PhD; Line B. Ulvin, MD; Richard Wennberg, MD, PhD; Sándor Beniczky, MD, PhD

Required 11-14 experts to label each EEG segment!

CLINICAL LABELLING - Inter-rater Variability

- In clinical practice, an inter-rater reliability coefficient of at least 0.6 is considered acceptable, with 0.8 as the gold standard

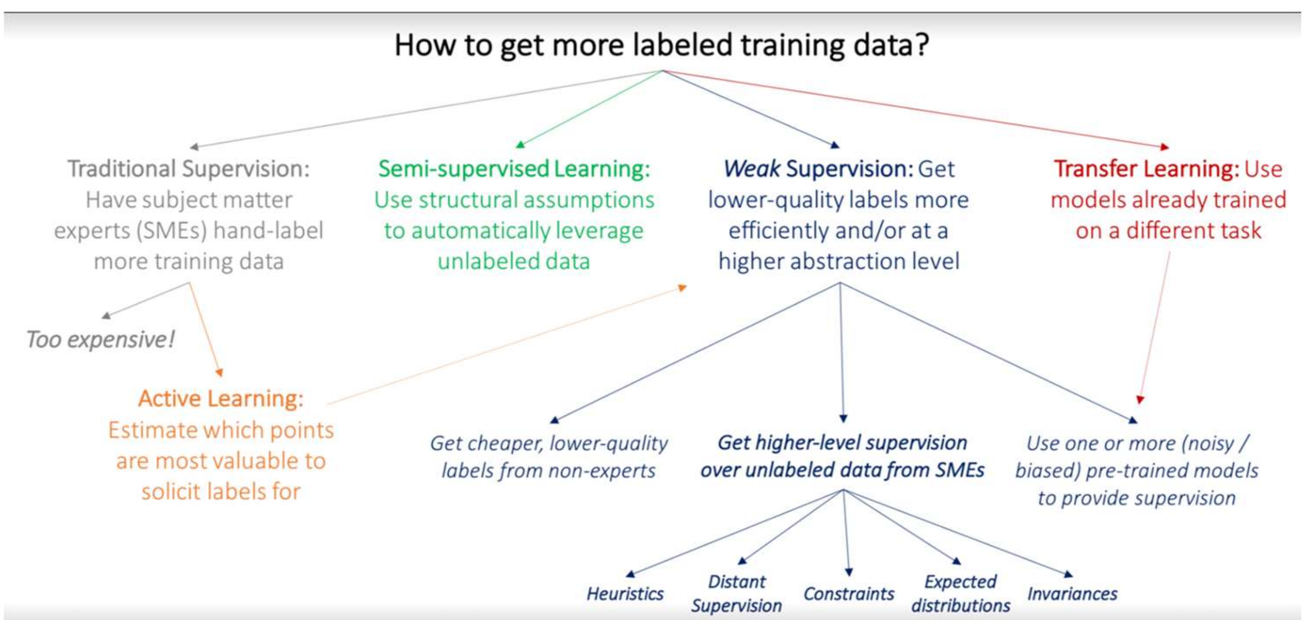
• SOURCES OF VARIABILITY:
  - examiner bias : "hawk" (==strict) vs. "dove" (==lenient) examiners
  - halo effect: an overall first impression influences subsequent judgments
  - differences in interpretation of clinical signs or symptoms
  - varying levels of experience or expertise among examiners

CLINICAL LABELLING - VARIABILITY

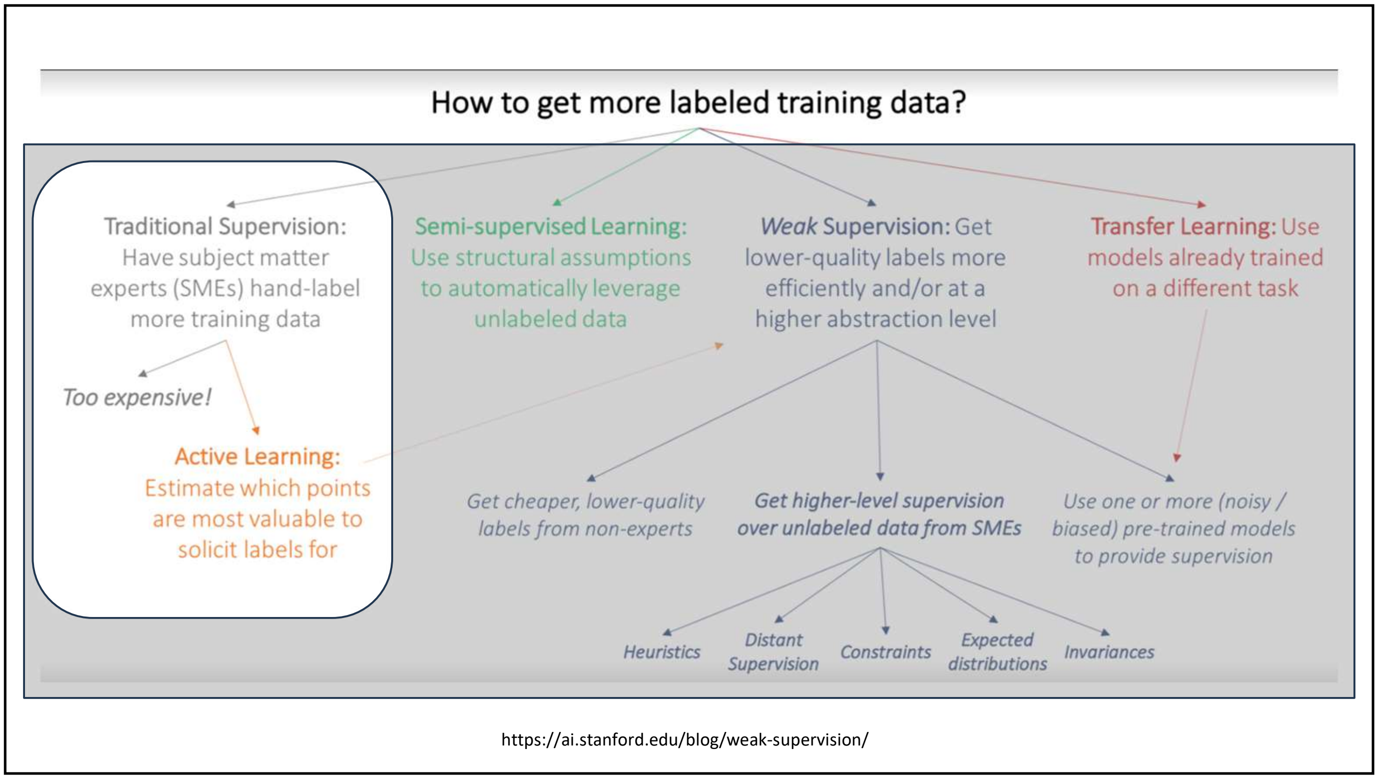
DEALING WITH VARIABILITY

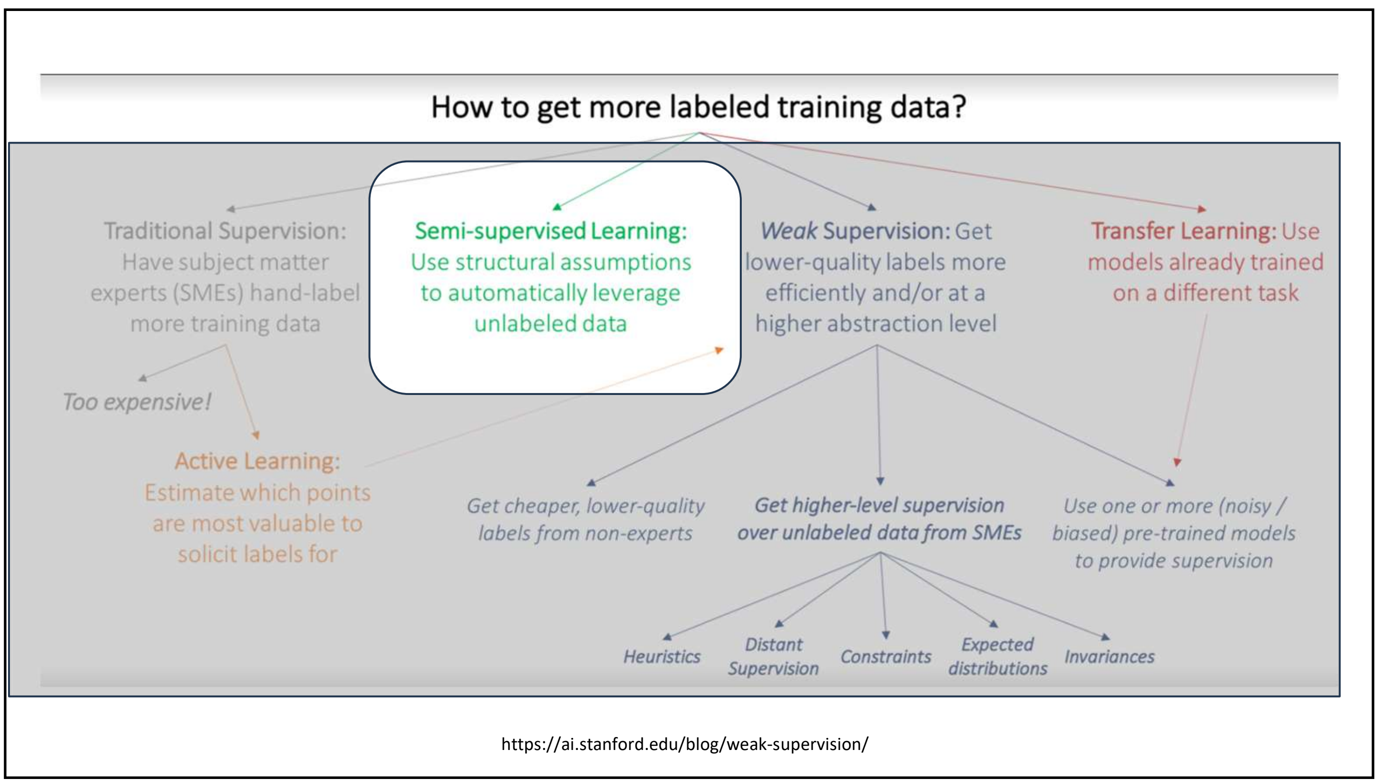
- **Implement training programs for examiners and standardize assessment procedures**
- **Paired examiner approach: reach a consensus score**
  - **personality factors may play a role: more extroverted examiners ma less likely to change their scores when paired with another examine**

- **Panel of experts (very expensive)**

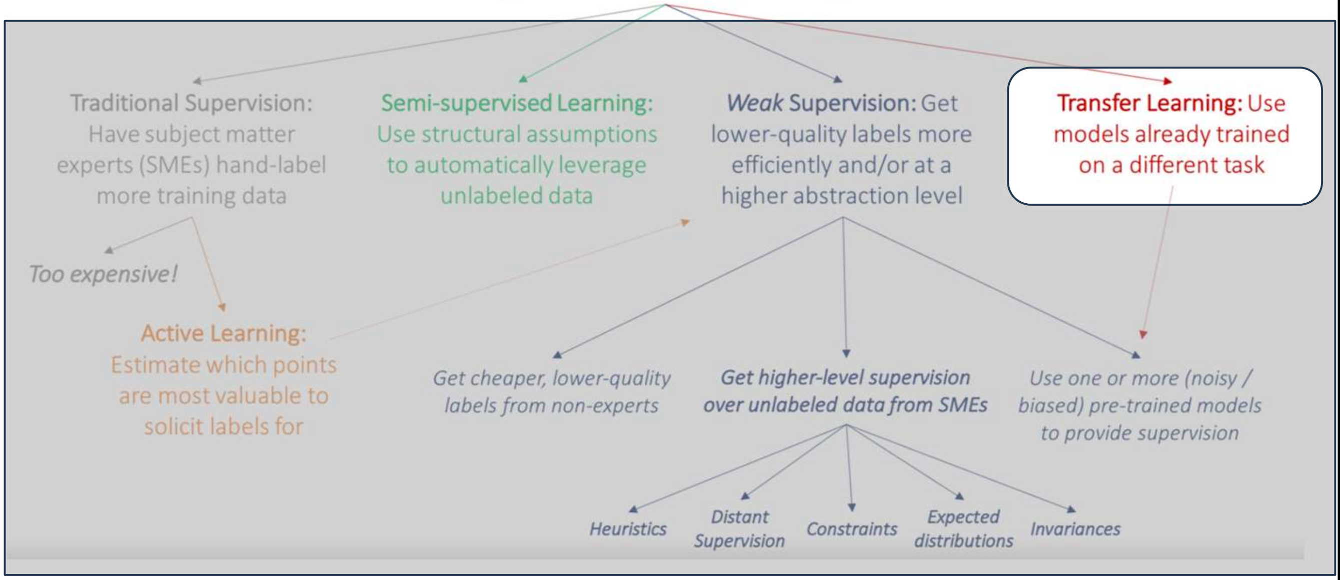


https://ai.stanford.edu/blog/weak-supervision/


How to get more labeled training data?
Traditional Supervision: Have subject matter experts (SMEs) hand-label more training data
Too expensive!
Active Learning: Estimate which points are most valuable to solicit labels for
Semi-supervised Learning: Use structural assumptions to automatically leverage unlabeled data
Weak Supervision: Get lower-quality labels more efficiently and/or at a higher abstraction level
Transfer Learning: Use models already trained on a different task
Get cheaper, lower-quality labels from non-experts
Get higher-level supervision over unlabeled data from SMEs
Use one or more (noisy / biased) pre-trained models to provide supervision
Heuristics
Distant Supervision
Constraints
Expected distributions
Invariances
https://ai.stanford.edu/blog/weak-supervision/

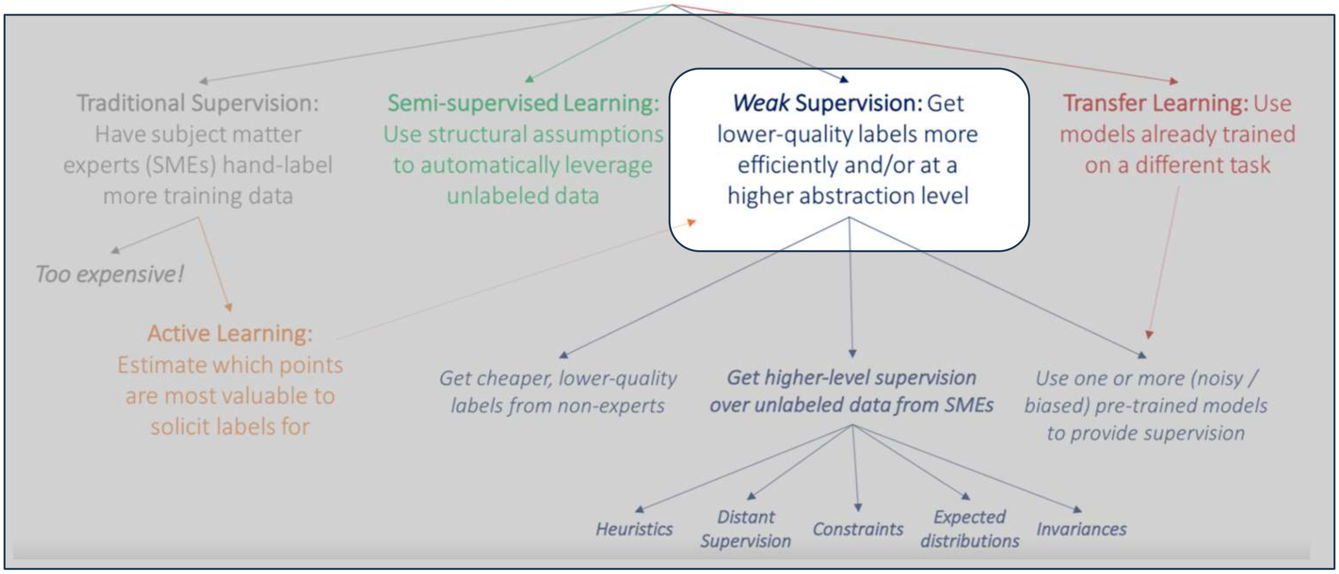

How to get more labeled training data?
Traditional Supervision: Have subject matter experts (SMEs) hand-label more training data
Too expensive!
Active Learning: Estimate which points are most valuable to solicit labels for
Semi-supervised Learning: Use structural assumptions to automatically leverage unlabeled data
Weak Supervision: Get lower-quality labels more efficiently and/or at a higher abstraction level
Transfer Learning: Use models already trained on a different task
Get cheaper, lower-quality labels from non-experts
Get higher-level supervision over unlabeled data from SMEs
Use one or more (noisy / biased) pre-trained models to provide supervision
Heuristics
Distant Supervision
Constraints
Expected distributions
Invariances
https://ai.stanford.edu/blog/weak-supervision/

How to get more labeled training data?



https://ai.stanford.edu/blog/weak-supervision/

How to get more labeled training data?

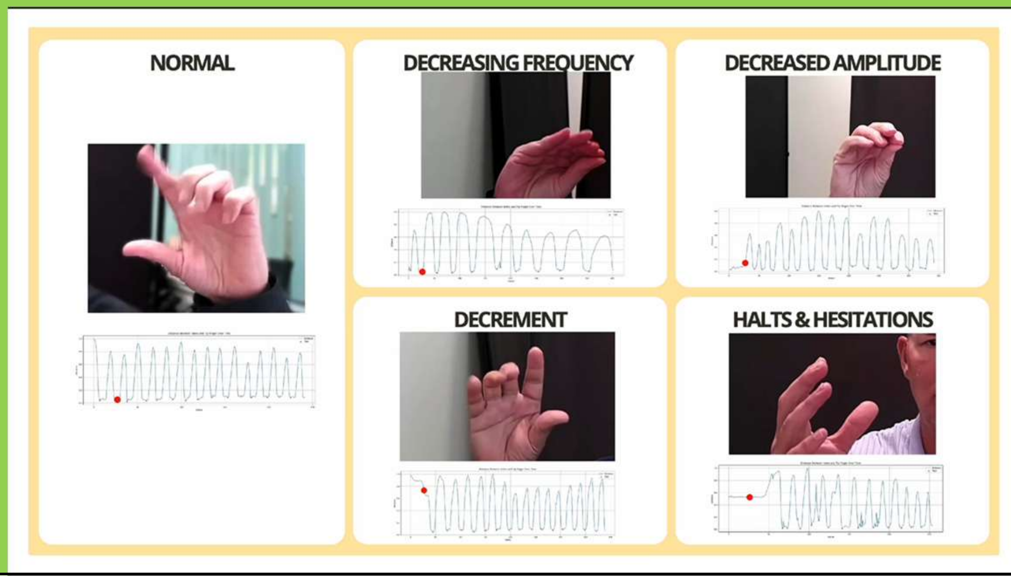


https://ai.stanford.edu/blog/weak-supervision/

EXAMPLE:
ASSESSMENT OF PARKINSON'S DISEASE

Patients are examined with a standard ***finger-tapping task***

clinicians are trained to observe:



EXAMPLE:
ASSESSMENT OF PARKINSON'S DISEASE

We have empirically observed that different raters tend to:

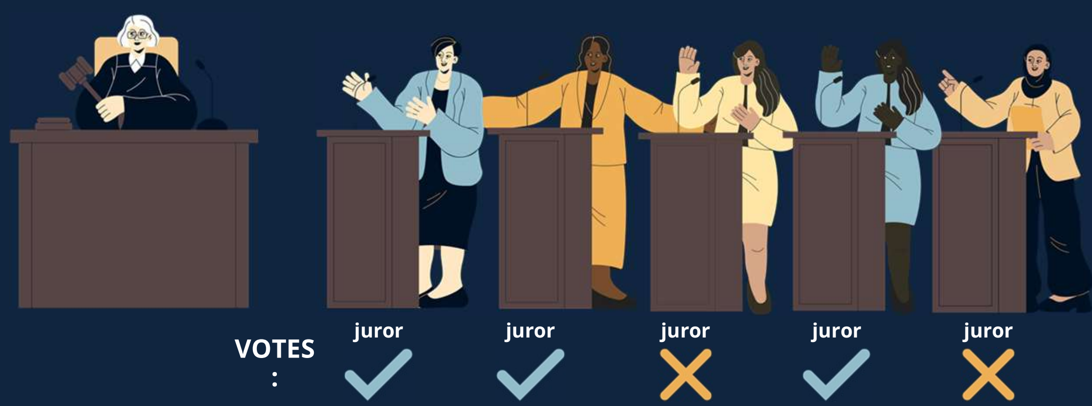
AGREE...
ON THE PRESENCE/ABSENCE OF FEATURES

- normal
- decrement
- halts & hesitations
- decreasing amplitude
- decreasing frequency

DISAGREE...
ON HOW TO INTEGRATE THESE DIMENSIONS INTO OVERAL DISE RITY

How can we 100% sure of labels if we don't know the "ground truth"?

imagine a panel of jurors that must determine if a sequence of defendants are *guilty* or *not guilty*
(WHERE ACTUAL GUILT IS UNKNOWN)

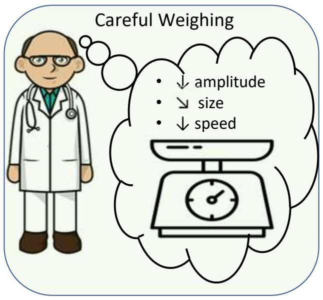


IMPROVING ON MAJORITY VOTE

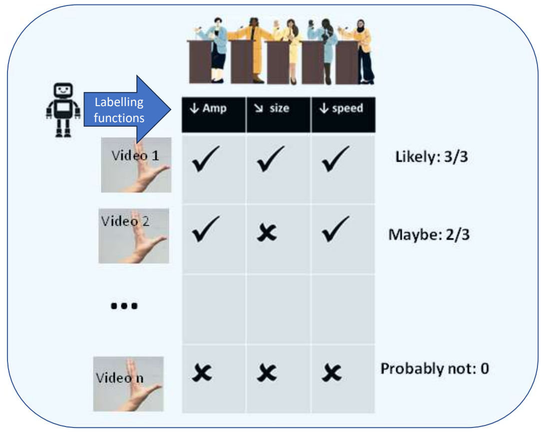
AI – Weak Supervision

The labels provided by clinicians may not be "gold standard" but rather "silver" or "bronze" standard.

Can we aggregate weak labels to create a robust label?



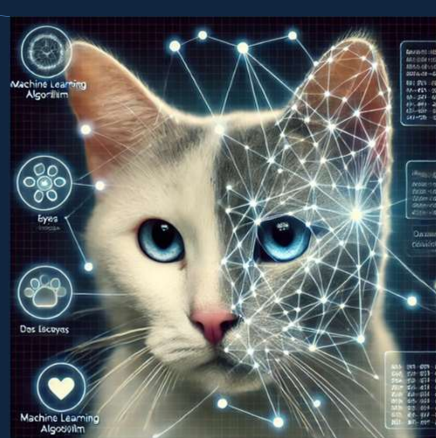
Expert Insight



Solution: Probabilistic Labeling

BIAS in data sets

Training Data



?

Privacy Concerns

AI typically requires Large data sets be transferred across the internet

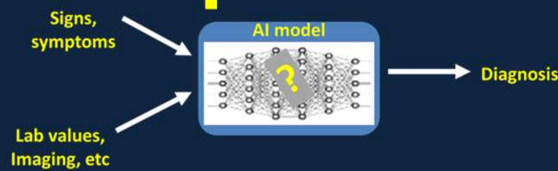
- Some cloud servers may not be in Canada
- All data transfer must be compliant with:
  - Privacy Act and the Personal Information Protection and Electronic Documents Act (**PIPEDA**)
  - Freedom of Information and Protection of Privacy Act (**FIPPA**)
  - Personal Information Protection Act (**PIPA**)
- British Columbia: E-Health (Personal Health Information Access and Protection of Privacy) Act

Privacy Concerns

AI an assist with some anonymization of data, etc.

Avatars

AI & Explainability



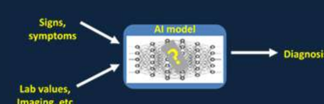
Trust and Adoption Issues

- Hesitancy among clinicians to adopt AI systems
- Patients feeling uncomfortable with AI-assisted decisions

Ethical and Legal Concerns

- **Informed Consent**: Without a clear understanding of how AI reaches its conclusions, challenging to obtain truly informed consent
- **Accountability**: When errors occur, difficult to determine responsibility if decision-making process is opaque
- **Regulatory Compliance**: Many jurisdictions require transparency in automated decision-making

Explainability



Clinical Risks

- **Undetected Biases**: AI systems may inadvertently perpetuate or amplify biases present in training data, leading to unfair or discriminatory outcomes
- **Misaligned Focus**: need to ensure that AI models are focusing on clinically relevant features rather than irrelevant artifacts in the data
- **Overreliance**: Clinicians might over-rely on AI recommendations without understanding their limitations or potential flaws

Impeded Improvement and Discovery

Explainable AI is crucial for:

- Debugging and improving AI models
- Discovering new clinical insights that could advance medical knowledge
- Identifying when models are making decisions based on spurious correlations rather than meaningful medical factors

Patient Safety Concerns

- In critical care situations, inability to quickly understand and verify AI recommendations could potentially compromise patient safety

Outline:

Basics of AI and Machine Learning

The Good

- Examples of the Transformative Potential of AI

The Bad

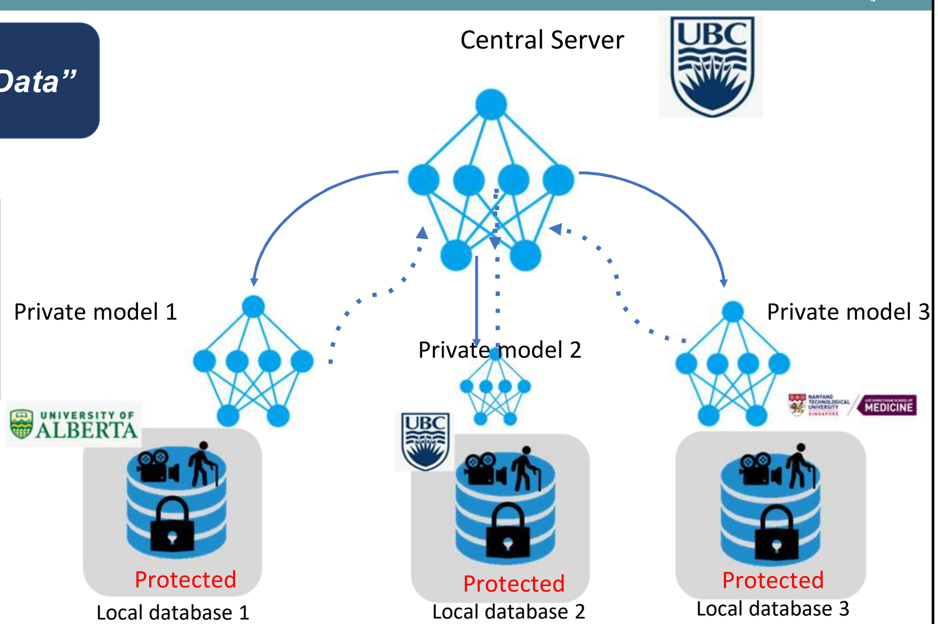
- Limitations
- Risks

The Future

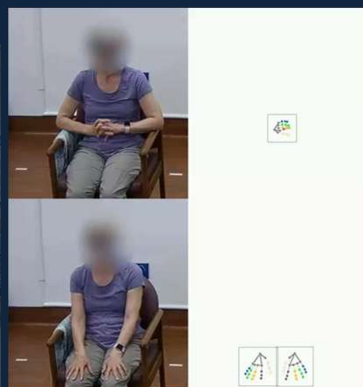
AI for data analytics – Federated Learning

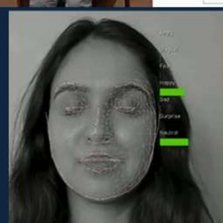
"Share the *Model* not the *Data*"

Vitally important when it is very difficult to get approvals to move data between jurisdictions!



Automation of parts of Physical (and History)





How are you feeling today?

Brain-Computer Interfaces for Rehabilitation

https://www.youtube.com/watch?v=5SrpYZum4Nk

Summary

Basics of AI and Machine Learning

The Good

- Examples of the Transformative Potential of AI

The Bad

- Limitations
- Risks

The Future

ACKNOWLEDGEMENTS

Maryam Mirian
Sepideh Hajipour
Kye Won Park
Atefeh Irani
Michael Grundy
Suzie Adam

Tianze Yu
Juana Ayala Castaneda
Hanaa Diab
Christina Kim
Martin Keung
Varsha Sreenivasan
Ben Quinn

Faculty:

Jane Z. Wang
Xiaoxiao Li

Questions?

martin.mckeown@ubc.ca
https://mckeownlab.ca/

Parkinson's Foundation
CENTER OF EXCELLENCE

THE UNIVERSITY OF BRITISH COLUMBIA
a place of mind