

# Disease Detection and Diagnosis with ML

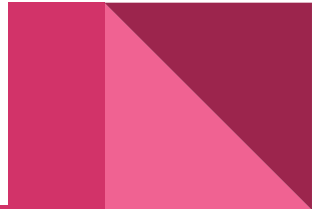
[ter.ps/389iweek3](https://ter.ps/389iweek3)

# GOOGLE'S AI EYE DOCTOR GETS READY TO GO TO WORK IN INDIA

AI can spot signs of  
Alzheimer's before your  
family does

**Chinese startup Infervision  
emerges from stealth with an  
AI tool for diagnosing lung  
cancer**

AI defeats elite doctors in diagnosis  
competition



# What is early detection?

*Diagnosing disease before symptoms appear*



# Early detection → early intervention

- **Better patient outcomes**

- survival, ability to treat and reverse disease, ...

- **Saves doctors valuable time**

- “We simply do not have the number of qualified doctors,”

<https://techcrunch.com/2017/05/08/chinese-startup-infervision-emerges-from-stealth-with-an-ai-tool-for-diagnosing-lung-cancer/>

- **Preparation**

- **Lower costs**



# Common Diseases

- Cancer
- Diabetes
- Dementia
- Heart Disease
- Stroke



# Machine learning

- Feed in samples
- Learn which samples are disease vs. non-disease
- Goal: automatic and accurate diagnosis of disease



# Data Types

# Image data

- Data points extracted from images
- Various methods, e.g., biopsies, ultrasounds, x-rays



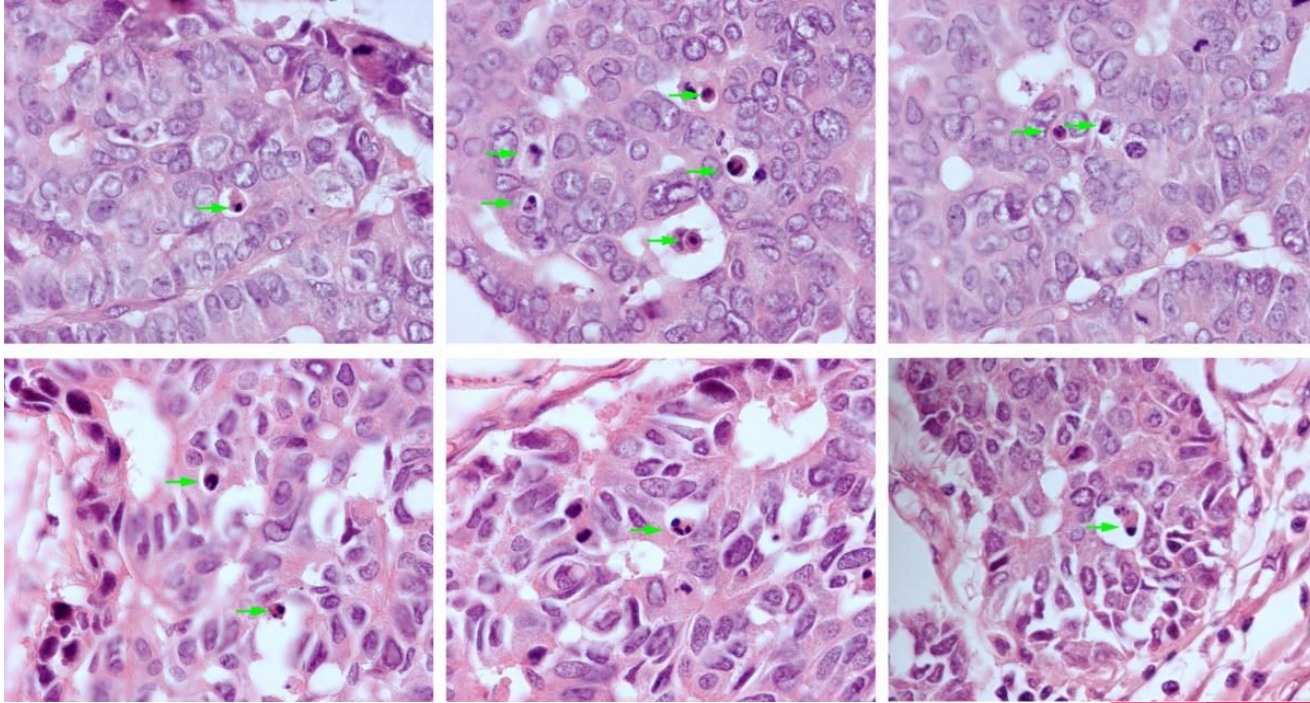


## Images: retinal images



<https://www.nature.com/articles/d41586-018-03067-x>

# Images: biopsies



<https://www.extremetech.com/extreme/233746-ai-beats-doctors-at-visual-diagnosis-observes-times-more-lung-cancer-signals>

# Images: ultrasound



<https://www.medgadget.com/2015/06/philips-heartmodel-brings-anatomical-intelligence-to-ultrasound-video>

# Images: radiography



aorta\_thoracic / tortuous / mild

aorta\_thoracic / tortuous



opacity / lung / middle\_lobe /  
right / aorta\_thoracic / tortuous

opacity / lung / base / left

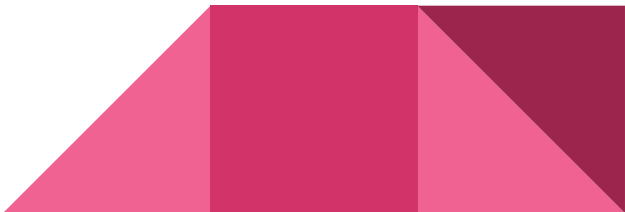


calcified\_granuloma / lung /  
middle\_lobe / right / multiple

calcified\_granuloma / lung /  
hilum / right

<https://news.developer.nvidia.com/detecting-and-labeling-diseases-in-chest-x-rays-with-deep-learning/>

# Other data types

- **Genomic data**
  - **Liquid biopsy**
    - Detect cfDNA and ctDNA / tumor mutations in blood!
  - **Wearable device data (e.g., Apple watch)**
    - Movement: accelerometer, gyroscope
    - Heart measurements: EKG's
- 

# Challenges

- Scarcity of data
- Batch effects
- Measurement error



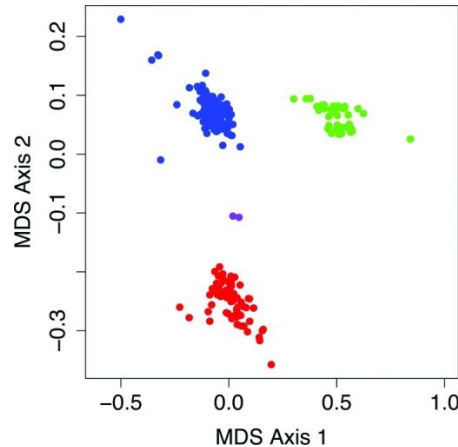
# Amount of Data

- **Few** but **large** samples
  - genomics, liquid biopsies
- Tons of noise, difficult to separate from signal
- Acquiring new samples can be expensive



# Batch effects

- Datasets from different sources can't just be lumped together
- Differences in data collection methods must be accounted for



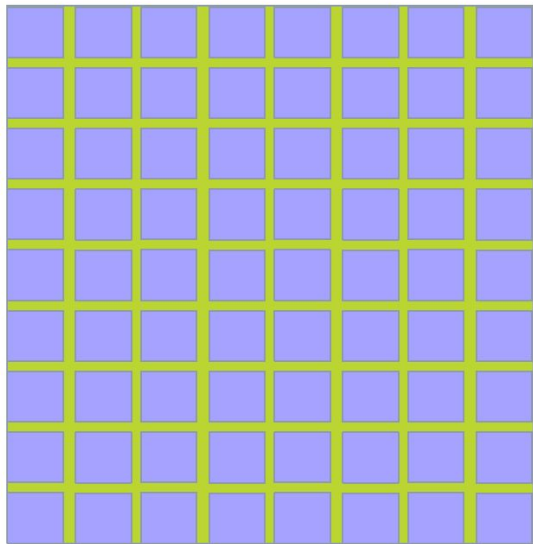
[https://openi.nlm.nih.gov/detailedresult.php?i=C4081054\\_gkuf&req=4](https://openi.nlm.nih.gov/detailedresult.php?i=C4081054_gkuf&req=4)



# Measurement error

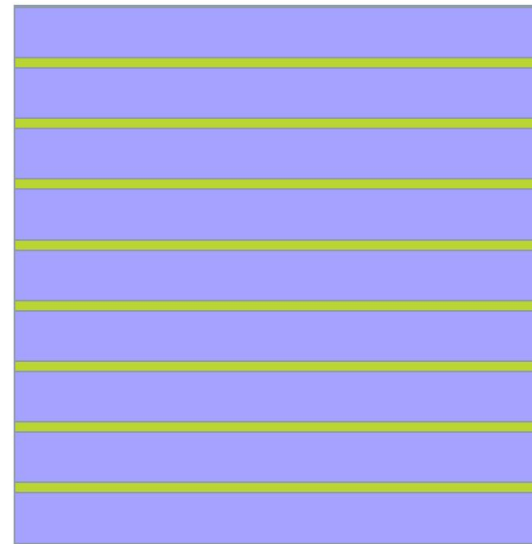
- Biological measurement devices are not perfect
- Sample-, source-, and batch-specific errors
- This makes machine learning hard





### Traditional ML

- **Many, small** samples (kB-MB)
- Cheap(ish) to acquire new instances (~\$0.01)
- Easy(ish) to sample over measurement error



### Biological ML

- **Few, large** samples (GB)
- Expensive to acquire new samples (\$1000)
- Pervasive measurement error

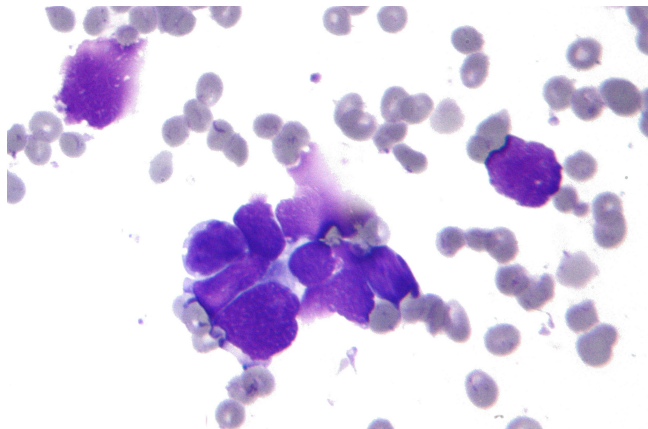
# Ethics - key considerations:

- What if the condition is untreatable?
- Mis-diagnosis
  - Unnecessary surgeries?
- Complement, not replace, physician
  - **diagnose vs. classify**



# Wisconsin Breast Cancer Dataset

- Digitized Fine Needle Aspirate (FNA) images
- Features: cell nuclei characteristics
  - radius, texture, perimeter, area, smoothness, compactness, concavity, concave points, symmetry, fractal dimension



# Further Reading

- <https://www.nature.com/articles/d41586-018-03067-x>
- <https://www.wired.com/2017/06/googles-ai-eye-doctor-gets-ready-go-work-india/>
- <https://techcrunch.com/2017/05/08/chinese-startup-infervision-emerges-from-stealth-with-an-ai-tool-for-diagnosing-lung-cancer/>
- <https://www.technologyreview.com/s/609236/ai-can-spot-signs-of-alzheimers-before-your-family-does/>

