CS184A/284A
AI in Biology and Medicine

Course Introduction, Logistics

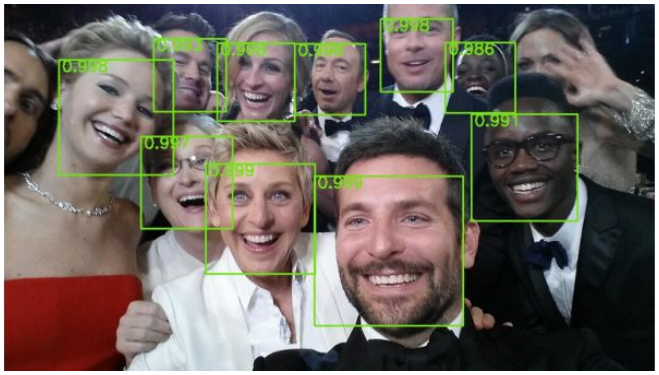
CS184A/284A Artificial Intelligence in Biology and Medicine

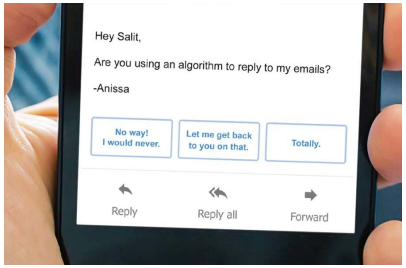
Introduction to **artificial intelligence and machine learning** with a focus on how to apply them to solve problems in biology and medicine. This class will familiarize you with a broad cross-section of models and algorithms from AI and machine learning. Applications will focus on problems from bioinformatics, genomics, medicine and healthcare.

Tentative Topics

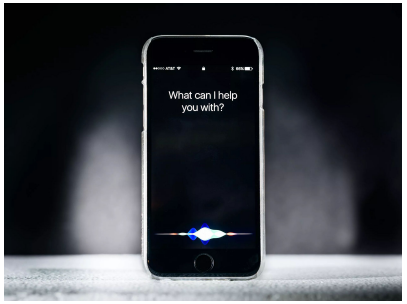
- Introduction. Course Framework
- Nearest neighbor methods, Linear regression
- Perceptron, Logistic regression, Support vector machines, Decision trees
- Application 1: Gene expression analysis, Biomarker discovery, Precision medicine
- Unsupervised learning, Principal Component Analysis, Clustering
- Application 2: Single cell RNA-seq analysis, other genomic applications
- Probabilistic models, Markov models, EM algorithm
- Application 3: Gene discovery, Regulatory motif discovery, CpG islands
- Neural networks, Deep learning
- Application 4: Biomedical image analysis

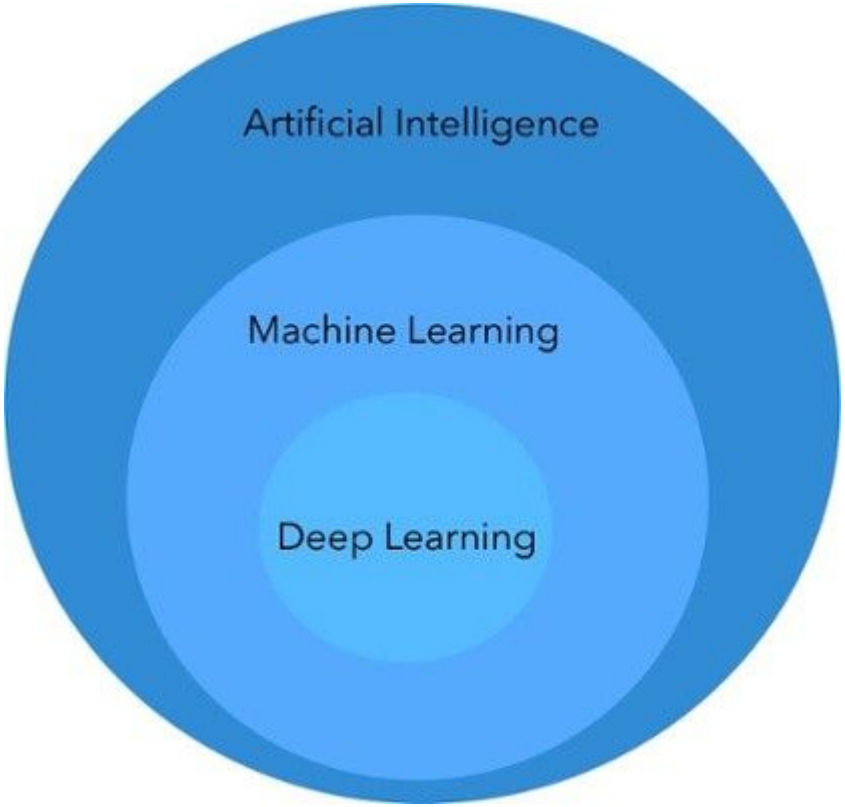
AI in everyday products





Google Translate



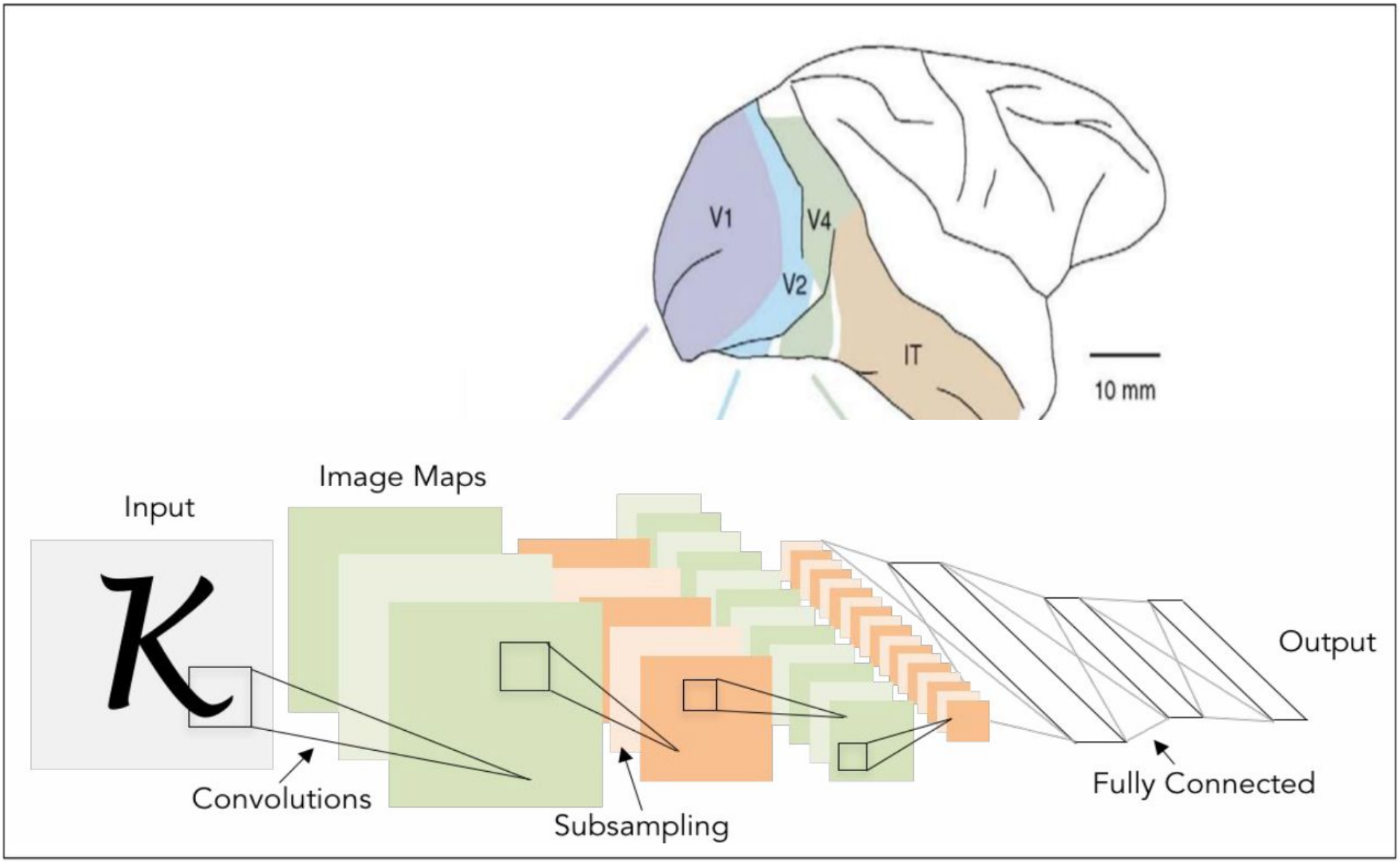


Artificial Intelligence:
Any technique that enables computers to mimic human intelligence, using logic, if-then rules, decision trees, and machine learning including deep learning

Machine Learning:
A subset of AI that includes statistical techniques that enable machines to improve at talks with experience.

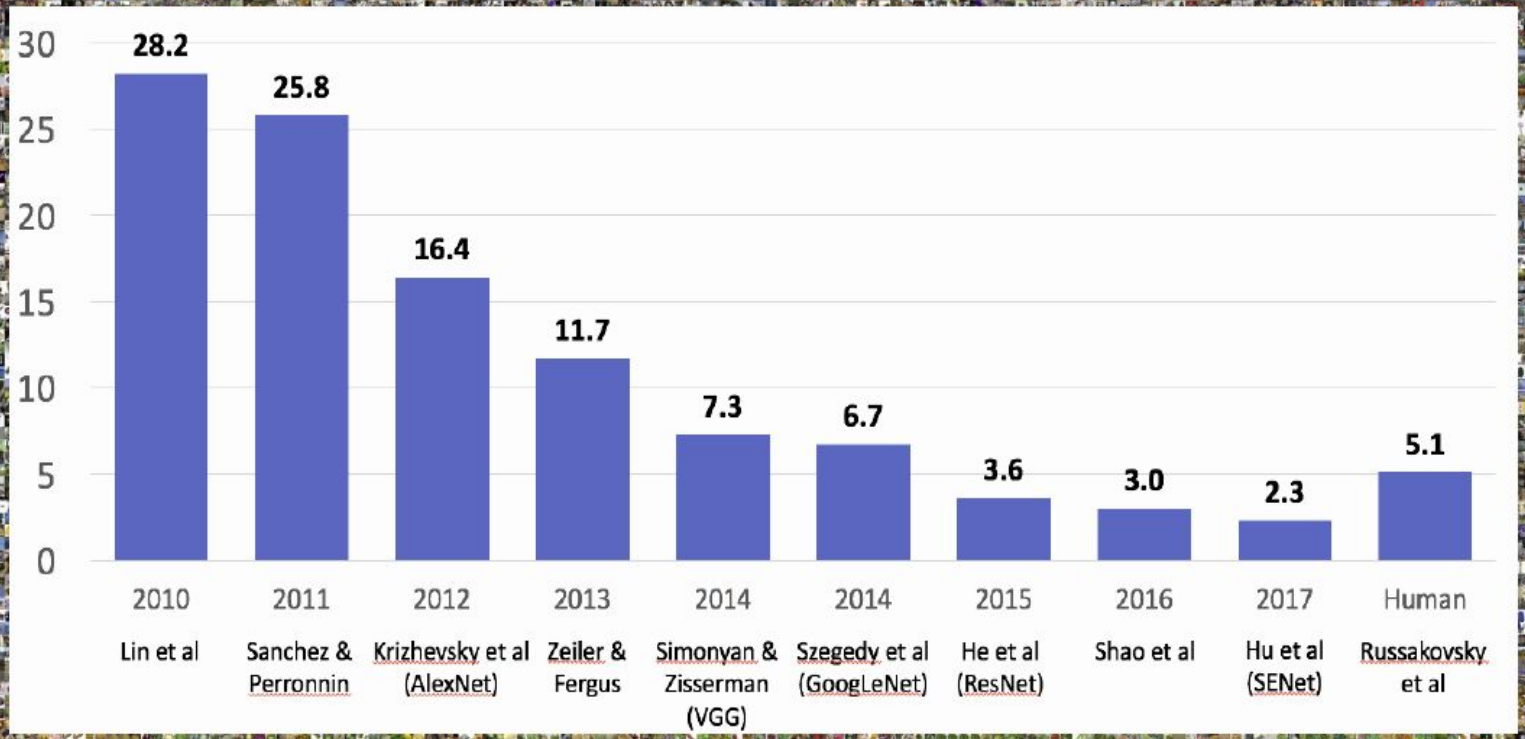
Deep Learning:
A subset of machine learning with models and algorithms built on deep neural networks.

Deep learning works by imitating the brain



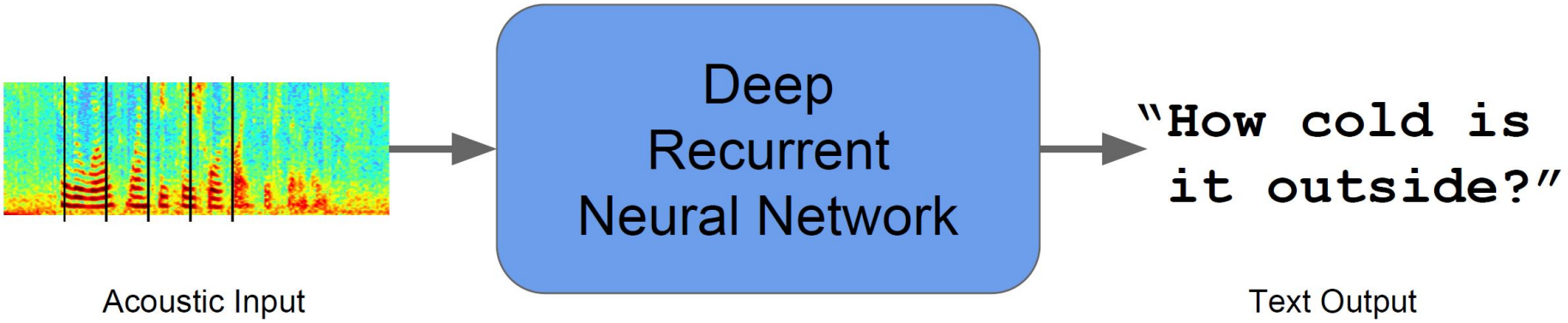
IMAGENET Large Scale Visual Recognition Challenge

The Image Classification Challenge:
1,000 object classes
1,431,167 images



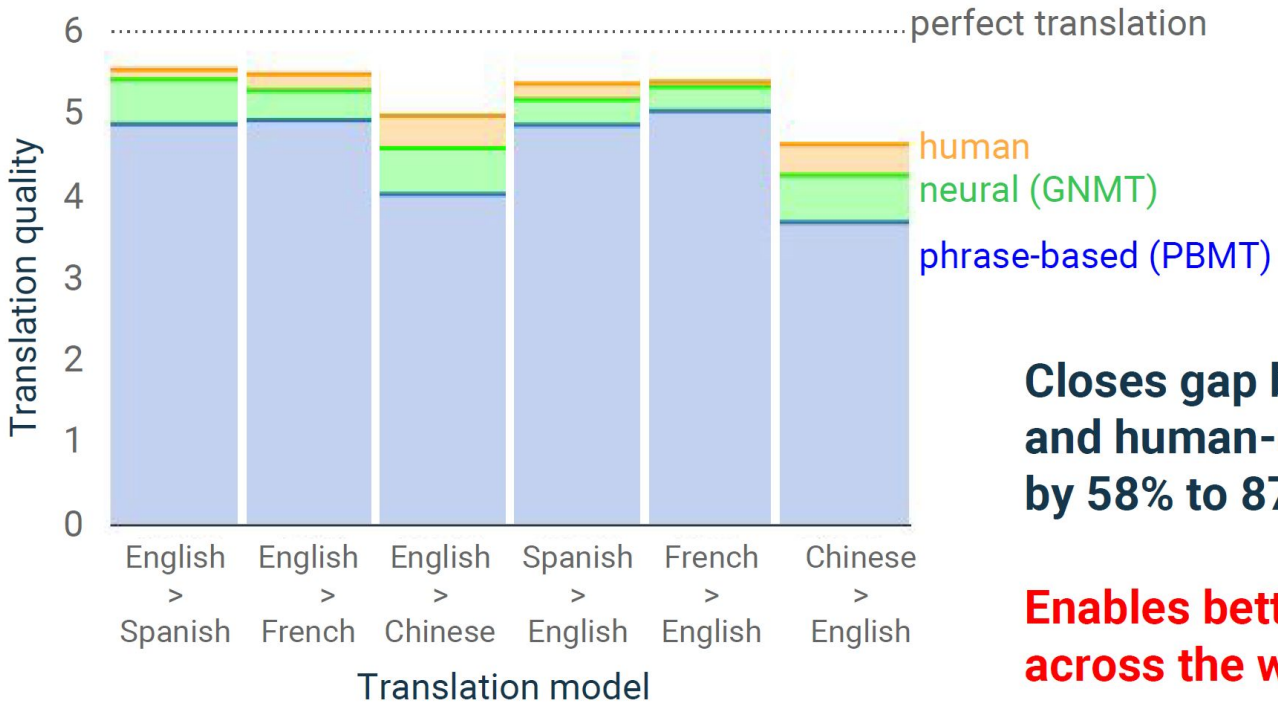
Russakovsky et al. IJCV 2015

Speech Recognition



Reduce error rate by over 30%

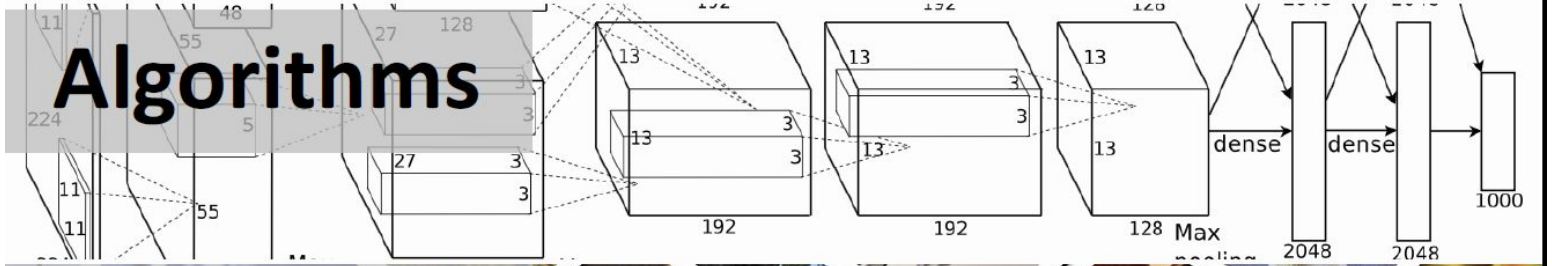
Machine Translation



Closes gap between old system and human-quality translation by 58% to 87%

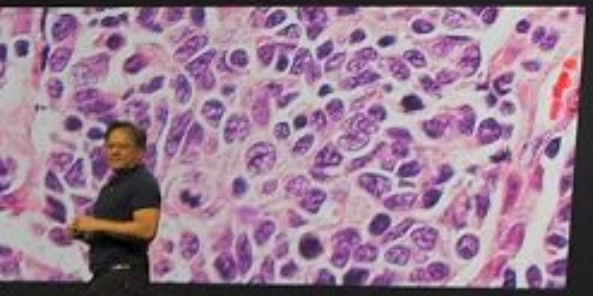
Enables better communication across the world

Ingredients for Deep Learning

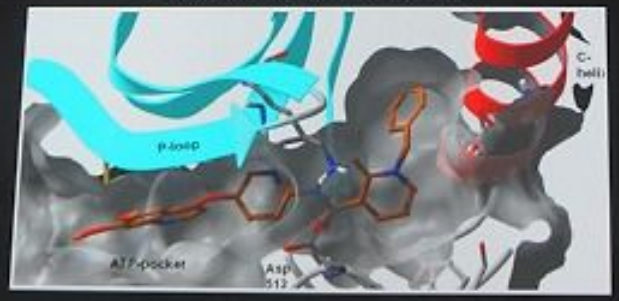


DEEP LEARNING REVOLUTIONIZING MEDICAL RESEARCH

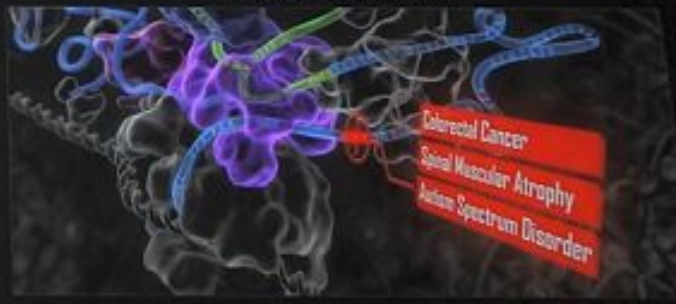
Detecting Mitosis in Breast Cancer Cells
— IDSIA



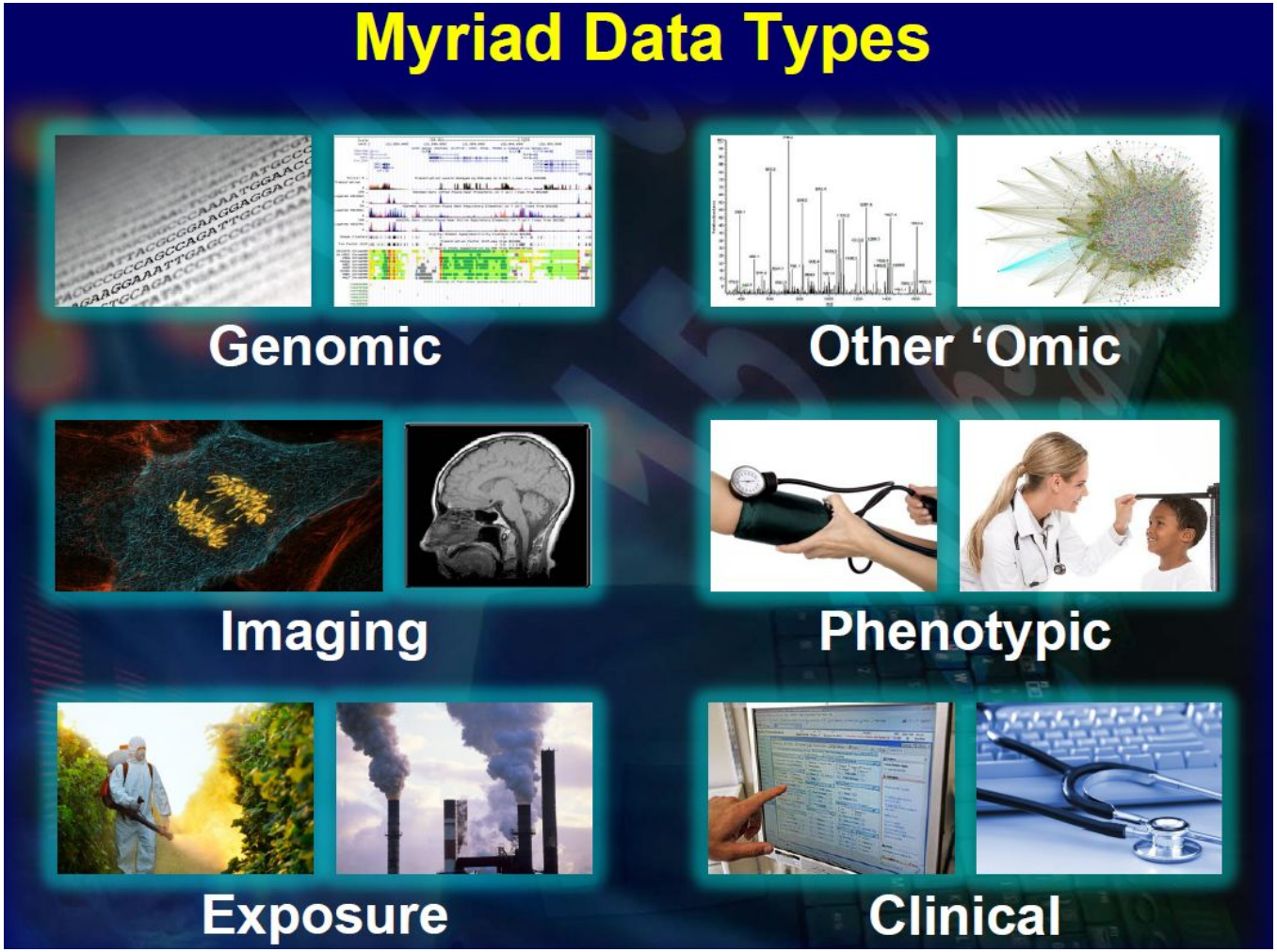
Predicting the Toxicity of New Drugs
— Johannes Kepler University



Understanding Gene Mutation to Prevent Disease
— University of Toronto

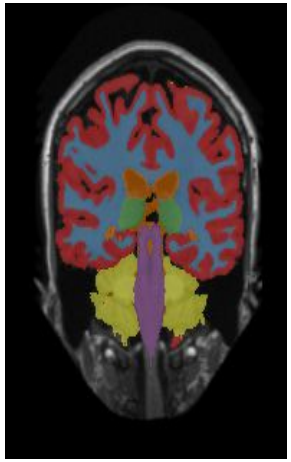
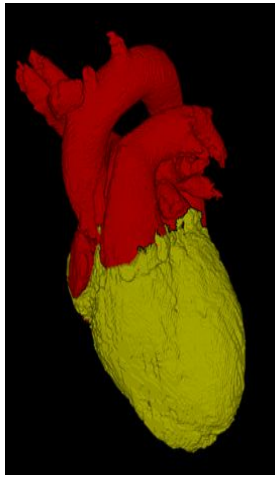


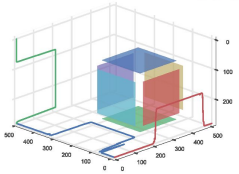
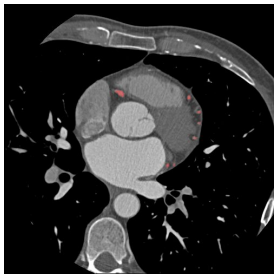
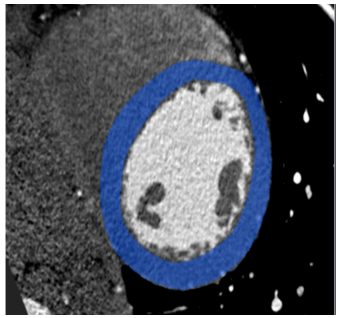
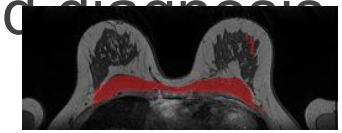
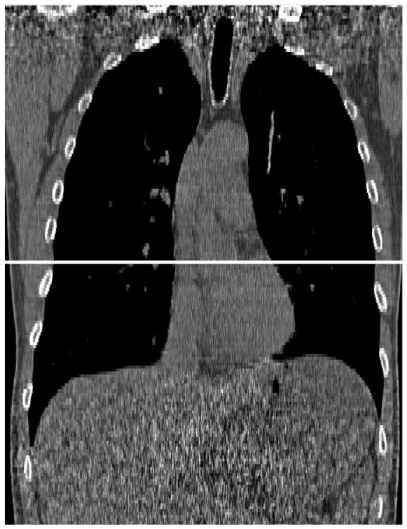
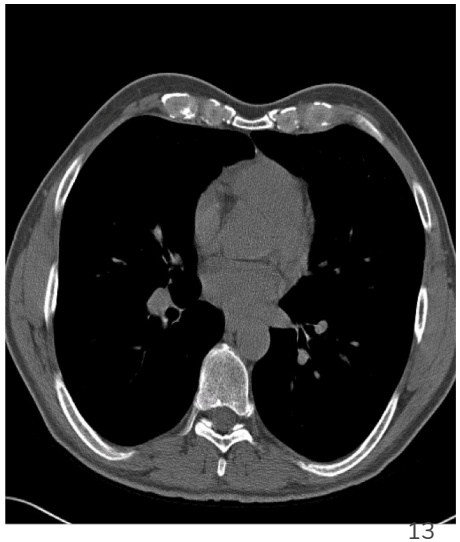
Big data in biomedicine



Deep learning in image analysis

- Application
  - Segmentation
  - Localization
  - Quantification
  - Computer-aided diagnosis

Machine Learning

Class Introduction

Course Logistics

Some example applications

Online resources

Course website

https://www.ics.uci.edu/~xhx/courses/CS284A/

Piazza:

https://piazza.com/uci/fall2020/cs184a/home

Grading

Grading policy:

- Course Project: 50%
  - Project Proposal
  - Presentation
  - Final report
- Final: 40% (online exam)
- Participation: 10%

Programming Assignments

2-3 Programming Assignments

Objective

- **Learn to apply ML techniques**
-

Source Code (Python)

Project

Groups for the Project

- **Team size should be 3 or less**
  - **Larger teams not allowed**
- **More details coming later**
- **Short report due at the end of the quarter**

Participation

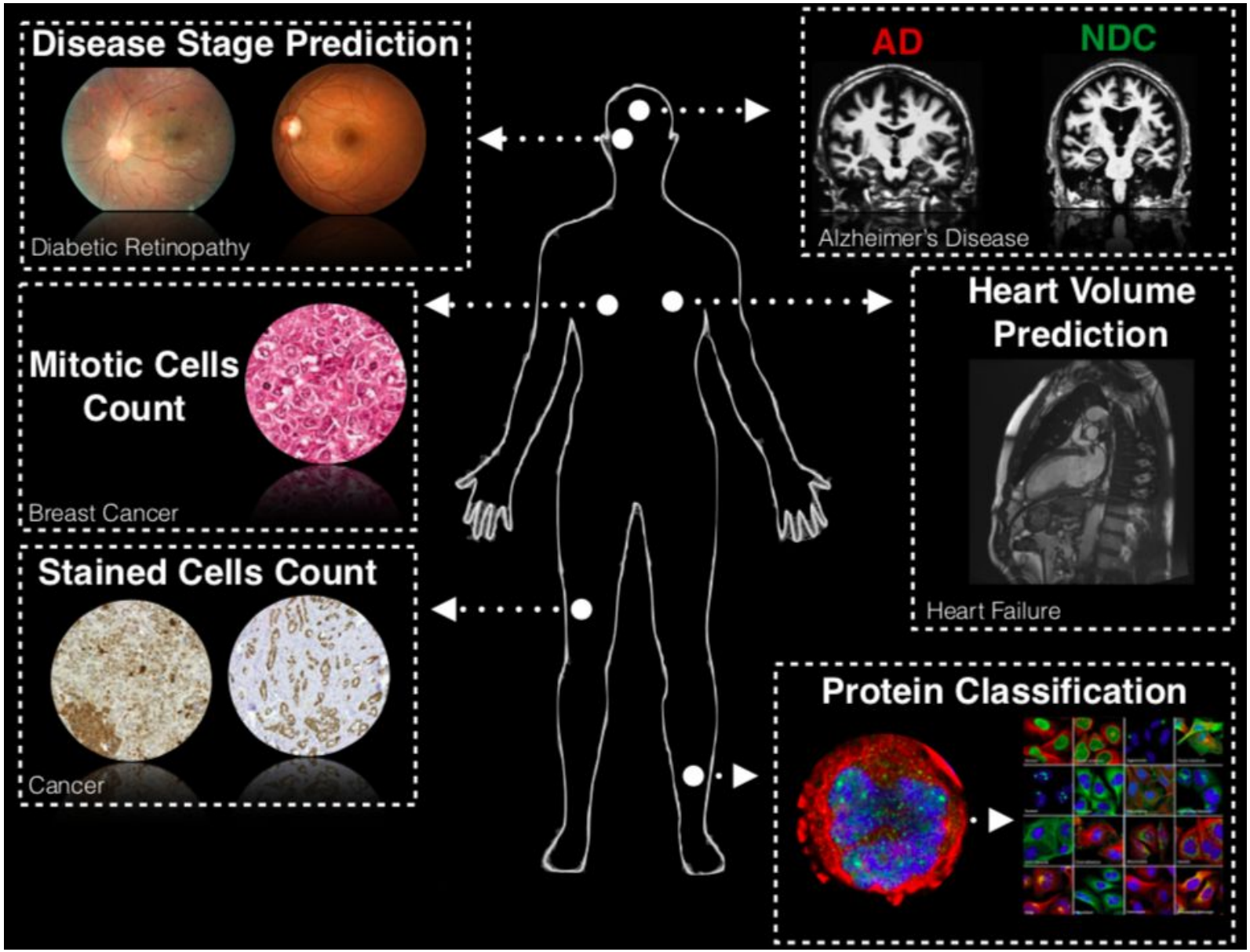
Surveys and Course Evaluation

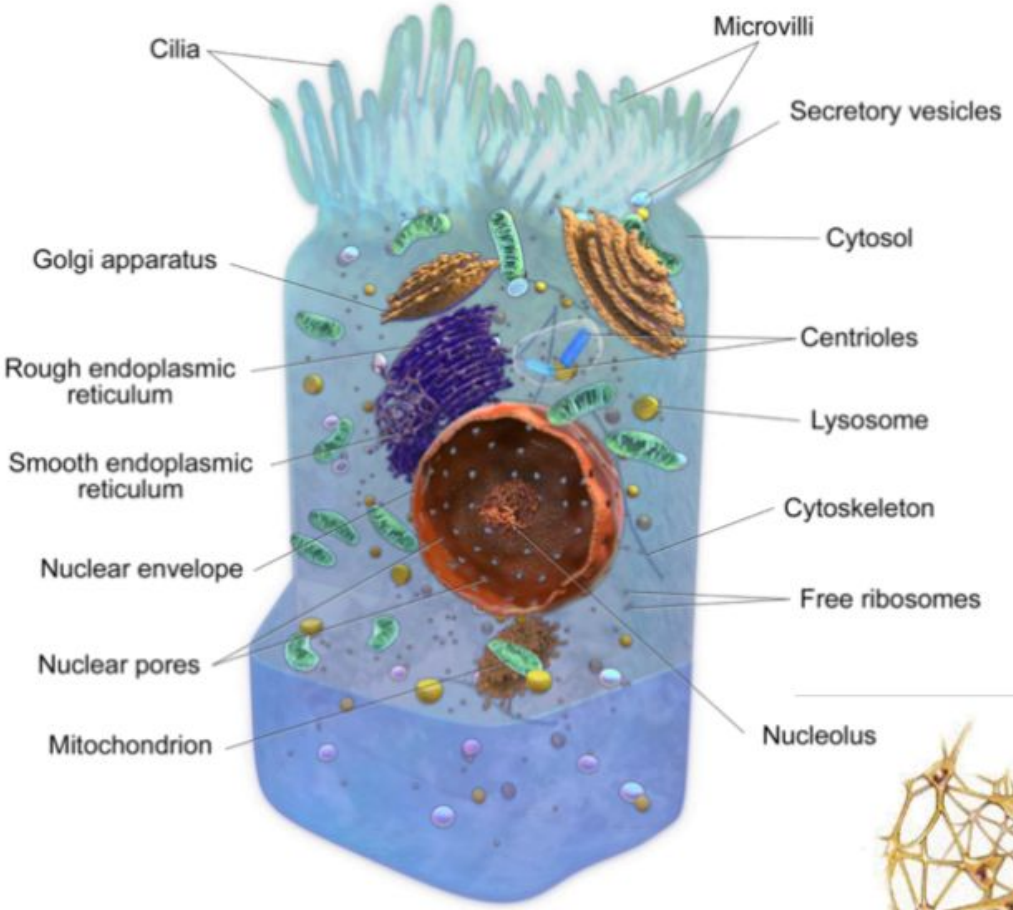
- Occasional polls on Canvas
- Participate in Course Evals

Discussions on Piazza

- Ask questions about material
- Answer posted questions
- Up vote helpful resources
- Post useful links related to the course

Example applications of AI/ML in Biology and Medicine


Disease Stage Prediction
Diabetic Retinopathy
AD
NDC
Alzheimer's Disease
Mitotic Cells Count
Breast Cancer
Heart Volume Prediction
Heart Failure
Stained Cells Count
Cancer
Protein Classification

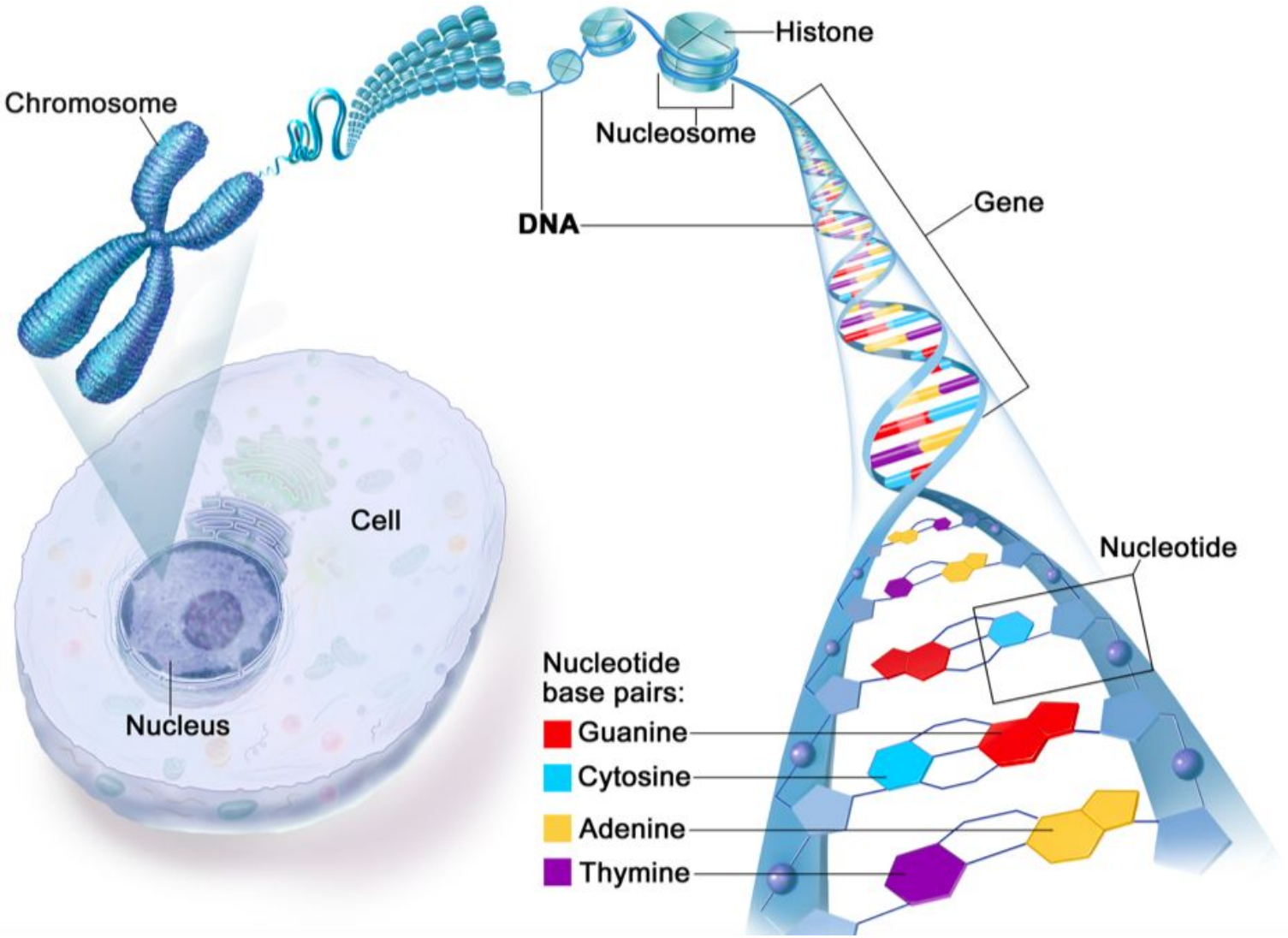


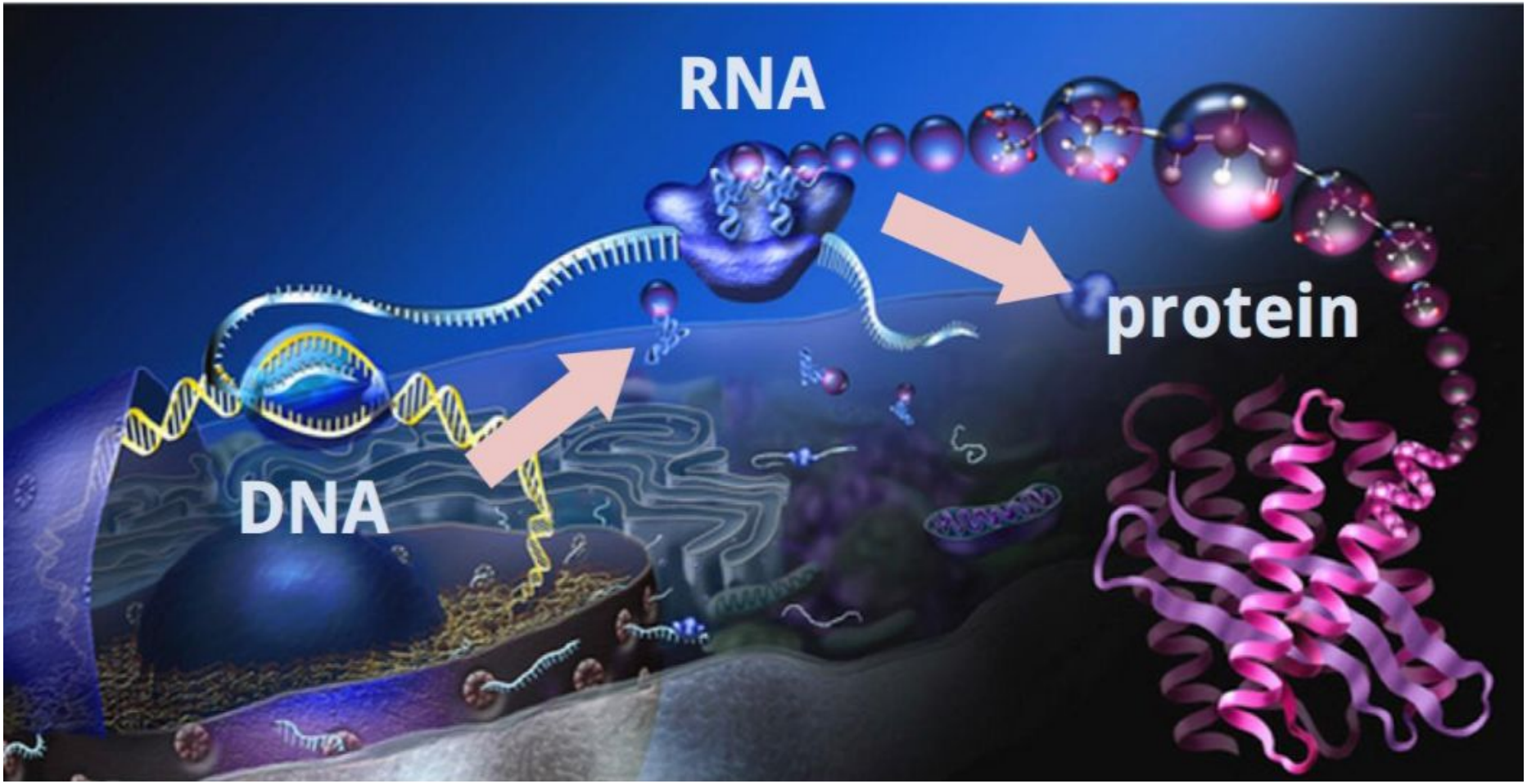
Anatomy of a Cell

Neural cells

Cardiac muscle

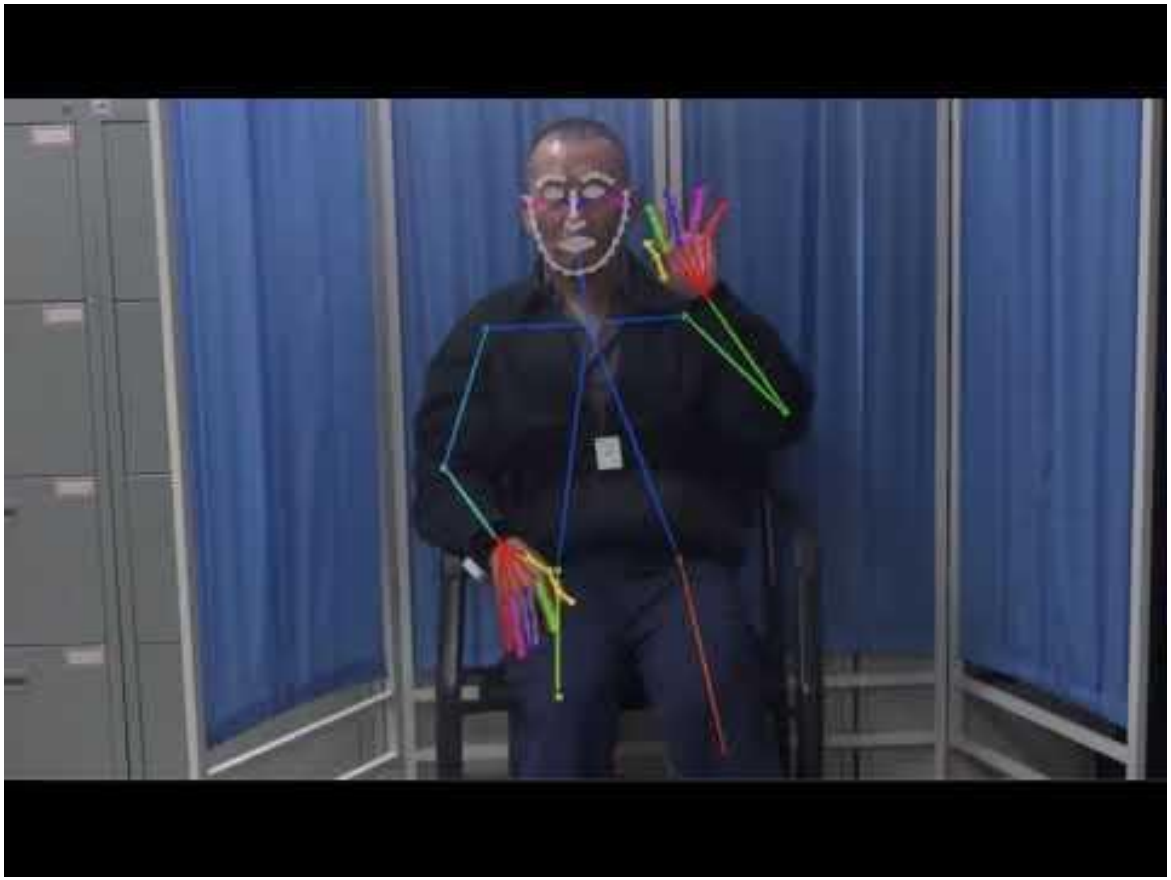
Blood cells


Chromosome
Histone
Nucleosome
DNA
Gene
Cell
Nucleus
Nucleotide
Nucleotide base pairs:
Guanine
Cytosine
Adenine
Thymine


RNA
protein
DNA

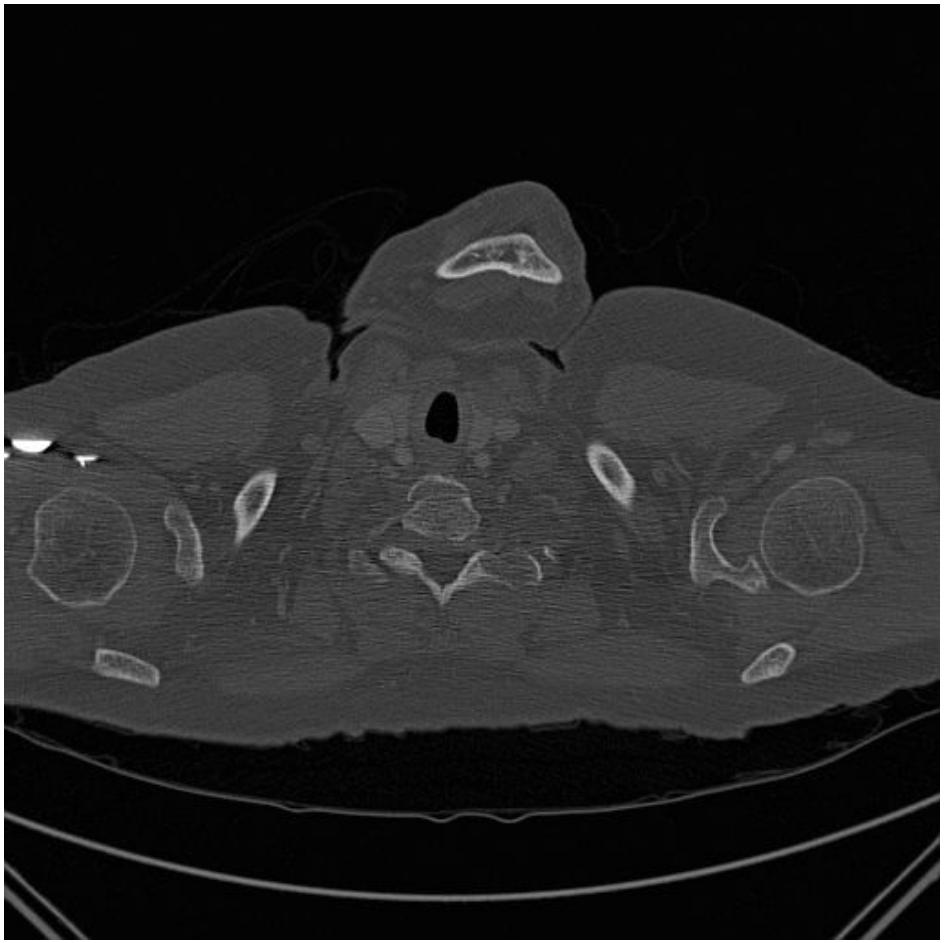
Some example applications

1. **Gesture recognition** - hand pose estimation for detecting and monitoring movement disorders, such as Parkinson's disease.
2. **Medical image analysis** - deep learning methods for detecting lung lesions
3. **Cell-level analysis** - cell nuclei segmentation
4. **Gene expression analysis**
5. **RNA structure prediction**
6. **Protein domain detection**



Detecting and monitoring movement disorders via video analysis

Pulmonary nodule detection



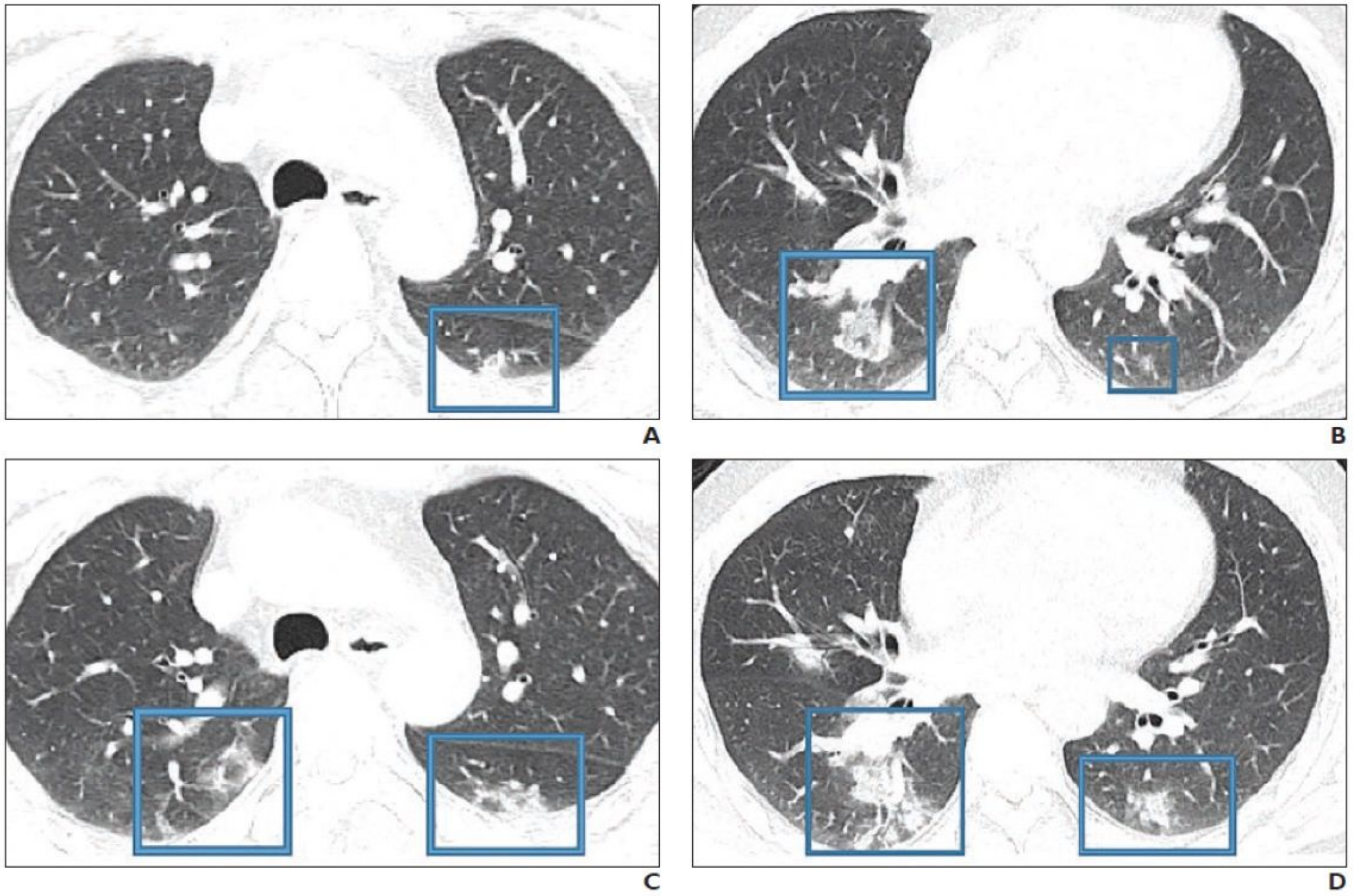
Input:
3D volumetric CT images
Typical CT image size:
512 x 512 x 400

Output:
Bounding box of
detected nodules

Application 2: Analysis of COVID-19 CT Scans

- Deep learning can be applied to automatically extract **image features** from CT scans, which can be used for patient stratification, disease progression monitoring, and studying treatment effect.

- Image features associated with COVID-19
  - **Ground-glass opacities (GGOs) and consolidation**
  - Idiopathic Pulmonary Fibrosis (IPF)
  - Interlobular septal thickening
  - Air bronchogram sign

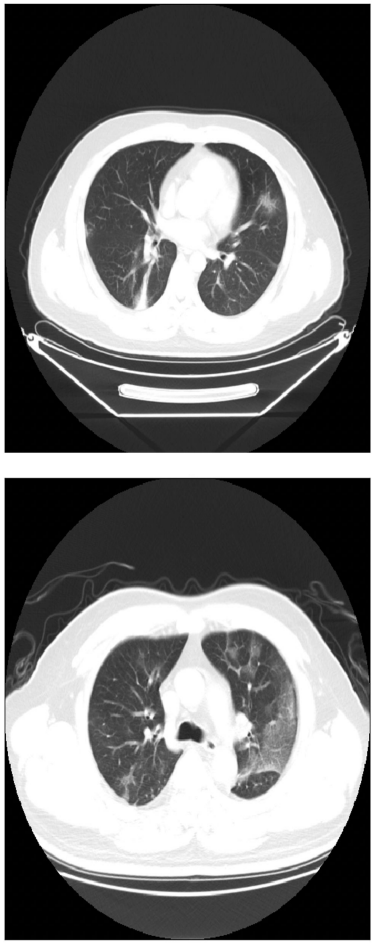


CT scans of a male patient with COVID-19. A and B: Initial CT images indicate GGO at level of aortic arch (A) and ventricles (B). C and D: Follow-up CT images obtained 2 days later show progression of abnormalities (rectangles). Image courtesy of AJR

COVID-19 Vulnerability Score System

CT Scan

AI Analysis



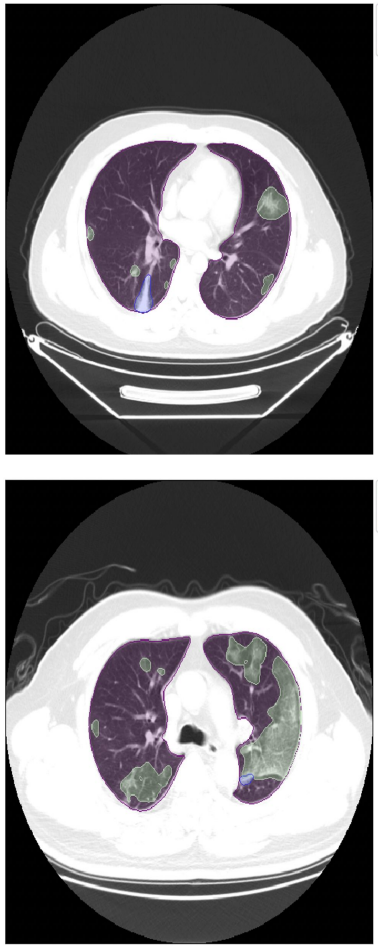


Image Features

- Age (years)
- Gender (0=F, 1=M)
- Comorbidities (HTN, DM, Obesity / BMI > 30, Asthma/COPD, CKD/ESRD, CVD)
- BMI (kg/m2)
- Respiratory Rate
- WBC (1000/mcL)
- Lymphocyte (%)
- Creatinine (mg/dL)
- LDH (U/L)
- Troponin-HS (ng/L)
- Ferritin (ng/mL)
- Procalcitonin (ng/mL)
- CRP-HS (mg/dL)

Model → Vulnerability Score

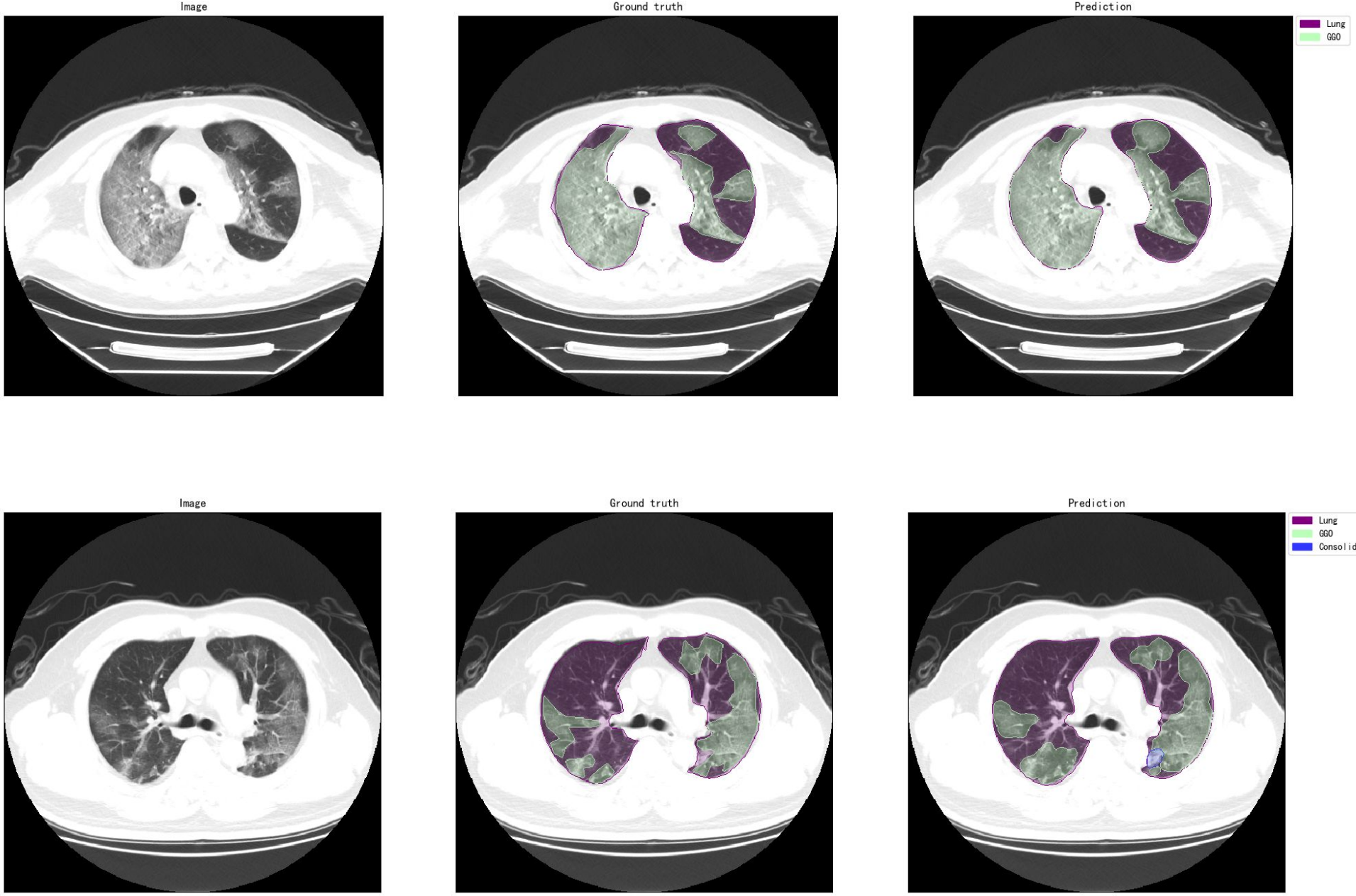
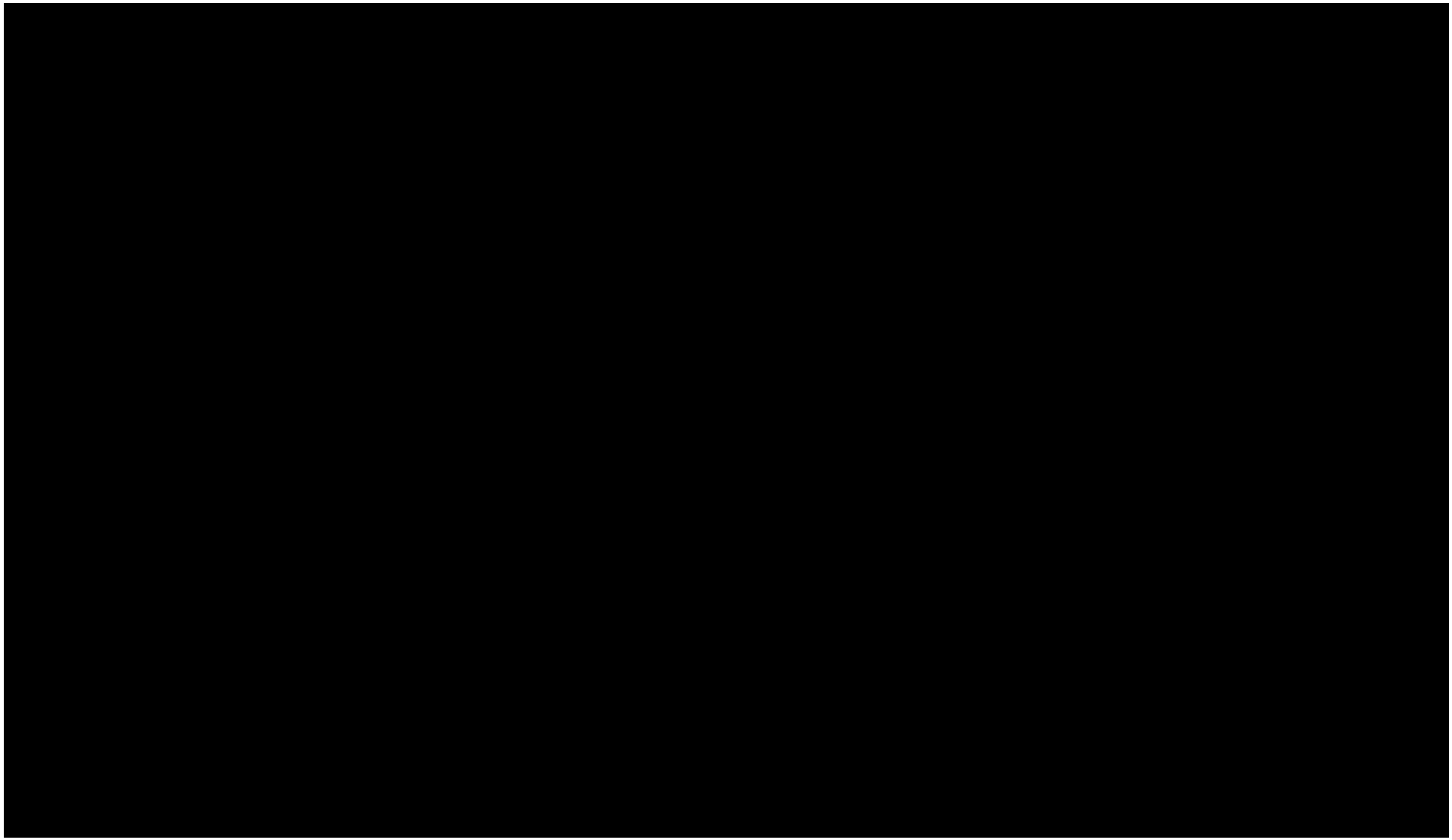
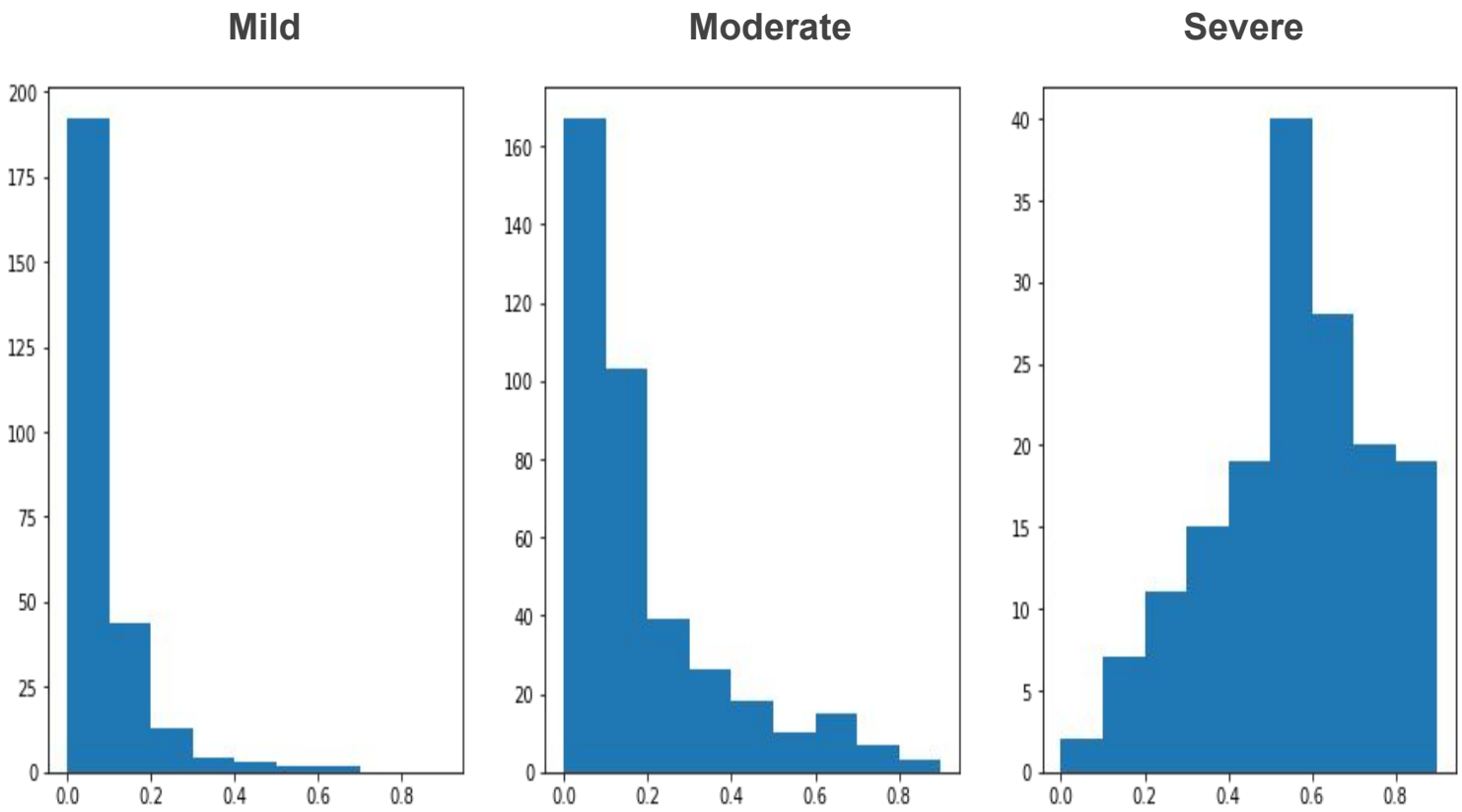


Image
Ground truth
Prediction
Lung
GGO
Image
Ground truth
Prediction
Lung
GGO
Consolid

CT scan of a covid-19 case

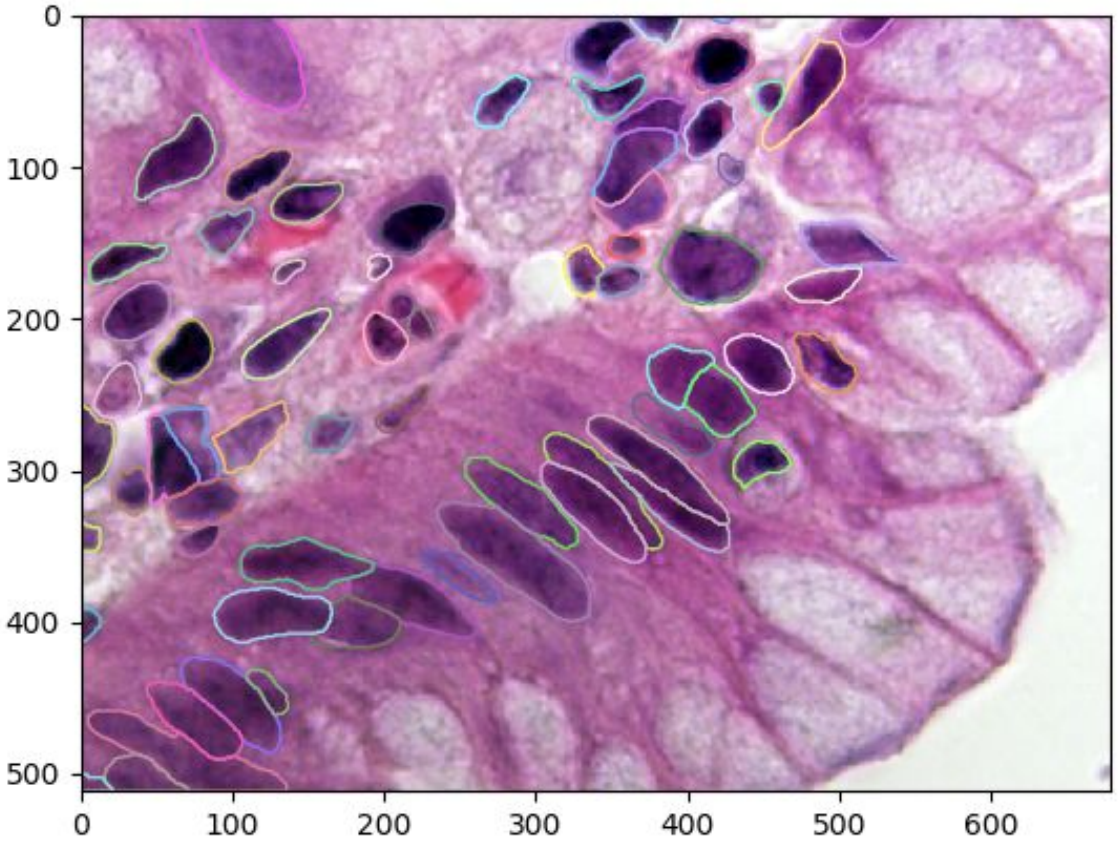


Separation of patients based on GGO scores



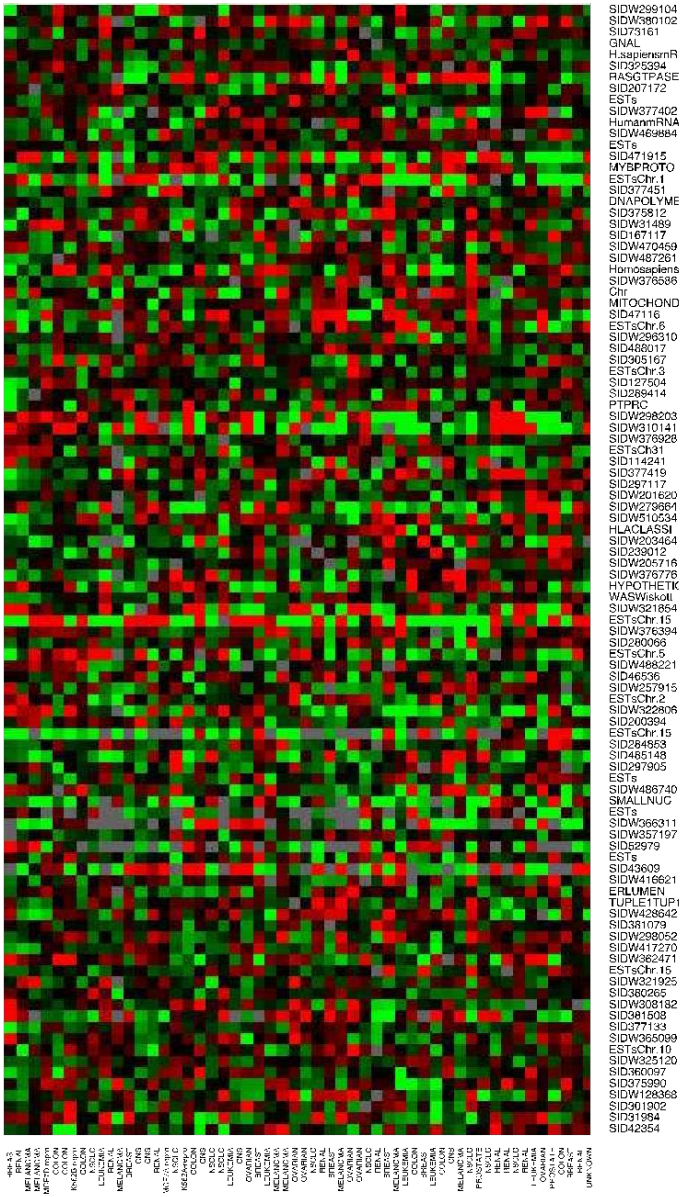
GGO ratio (area of GGO / area of lung)

Cell nuclei segmentation



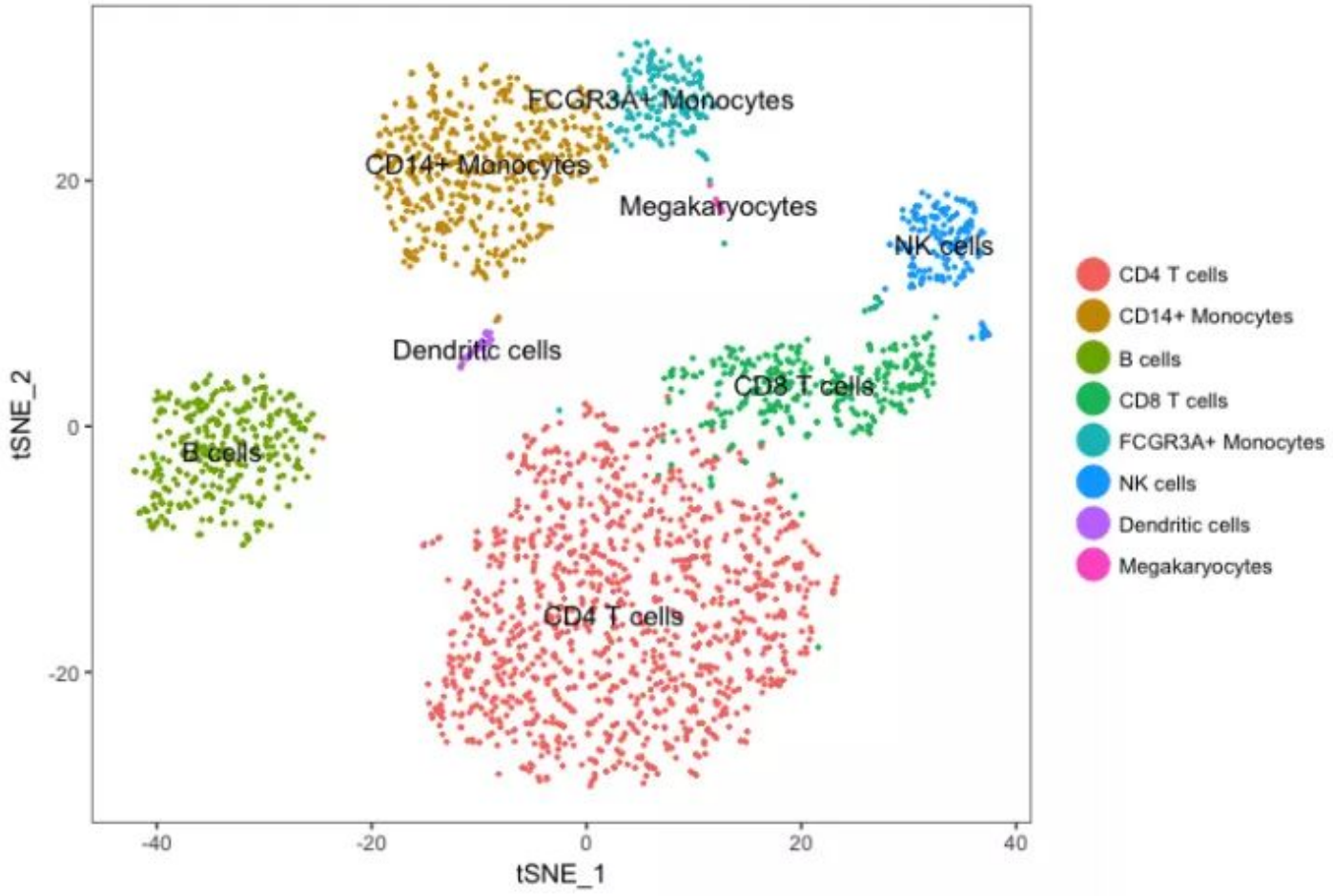
Human Gene Expression Data

- **6830x64 matrix of real numbers**
- **Rows correspond to genes, columns to tissue samples**
- **Cluster rows (genes) to deduce function of unknown genes from experimentally known genes with similar profiles**
- **Cluster columns (samples) to hypothesize disease profiles**



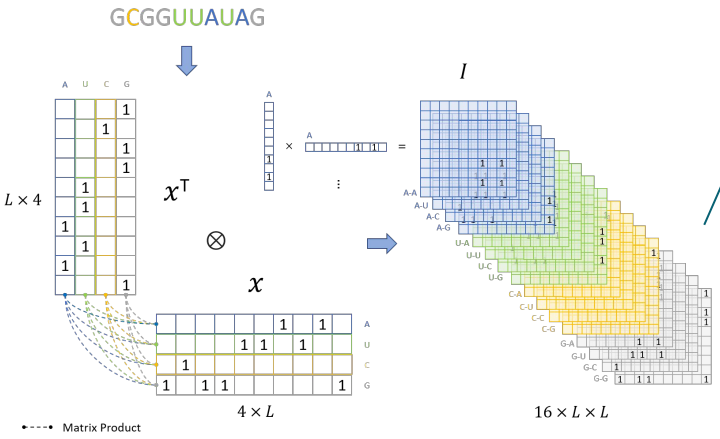
Hastie, Tibshirani, & Friedman 2009

Unsupervised learning for discovering cell types

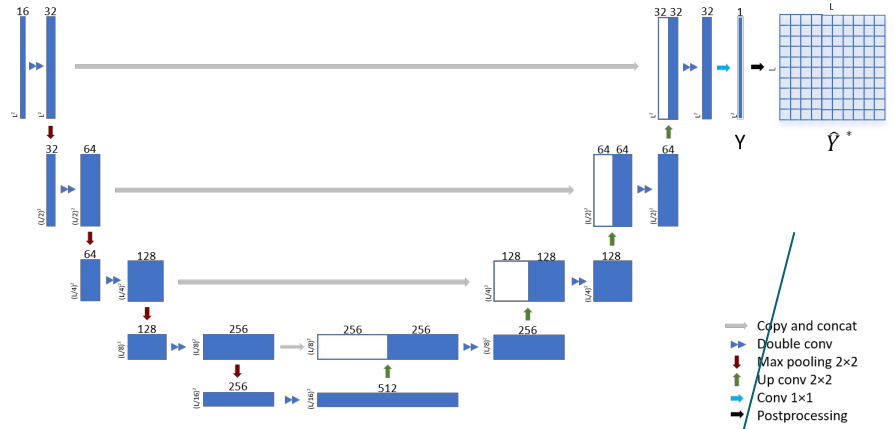


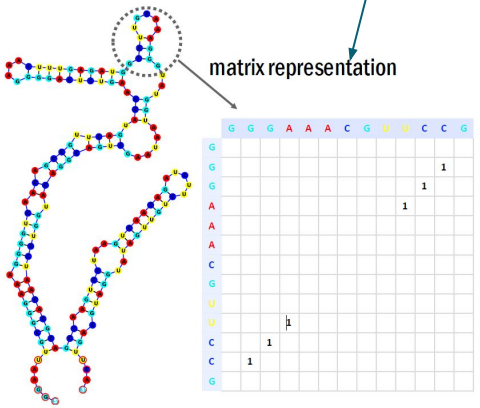
UFold - Deep Learning for RNA Secondary Structure Prediction

New representation of sequence;
convert sequence to “image”



U-Net model





Predict contact map

Demo: RNA secondary structure prediction

https://ufold.ics.uci.edu/

Aligning protein sequences and discovering functional domains

Human Ubiquitin Conjugating Enzymes