GE 461: INTRODUCTION TO DATA SCIENCE

Spring 2024-2025

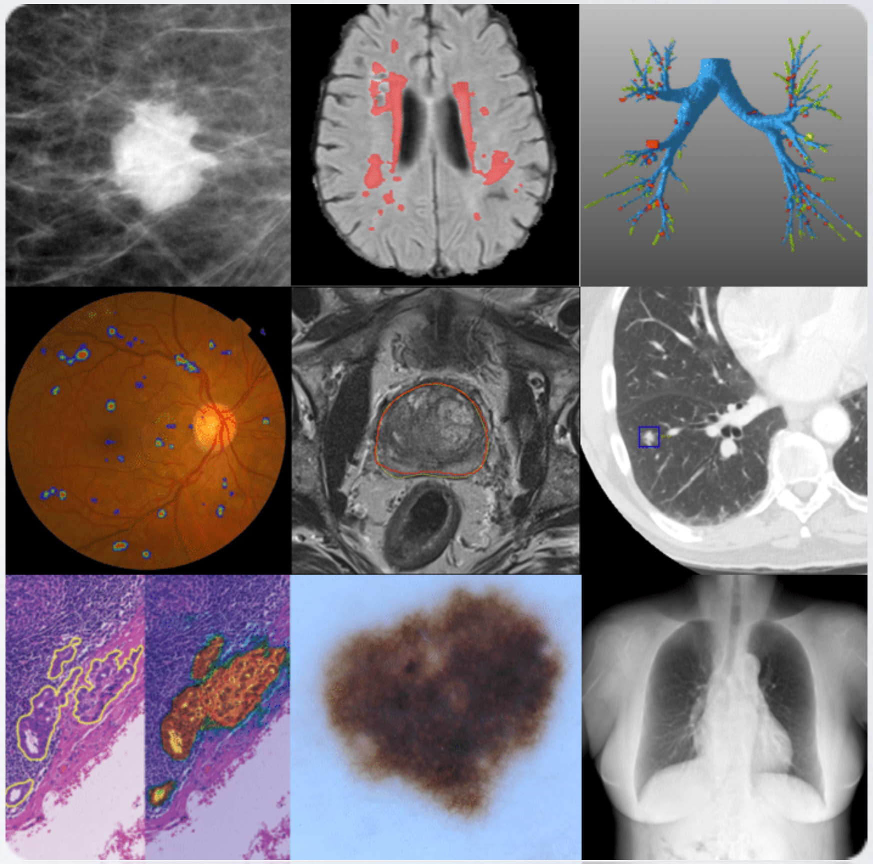
Machine Learning in Healthcare

Tolga Çukur

Machine Learning in Healthcare

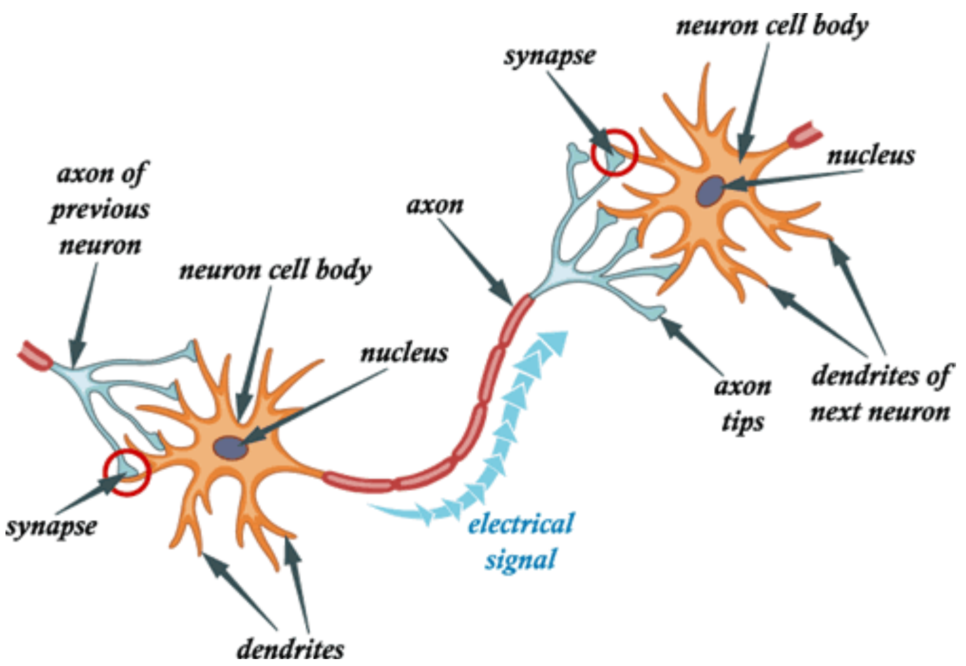
Outline

- Overview of Neural Networks
- History of Machine Learning in Medicine
- Big Data in Medical Applications
- Opportunities/Challenges in Healthcare
- Utility of Machine Learning in Medical Imaging
- Example Applications in Medical Imaging

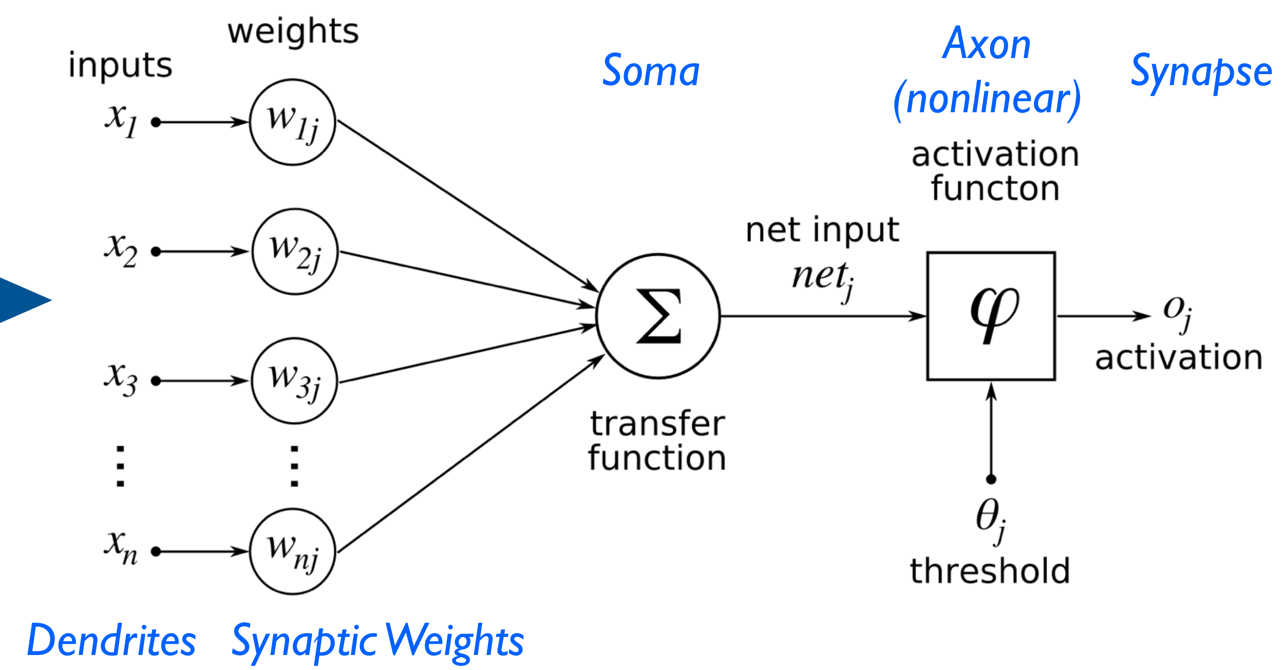


PART I: Overview of Neural Networks

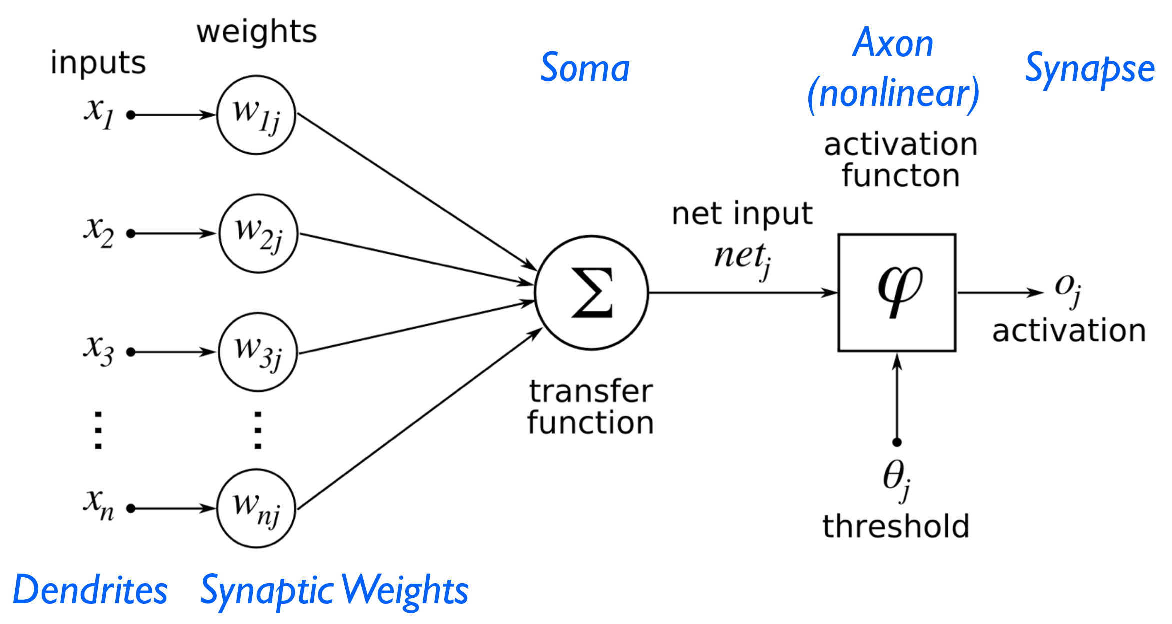
Artificial Neuron: A mathematical abstraction

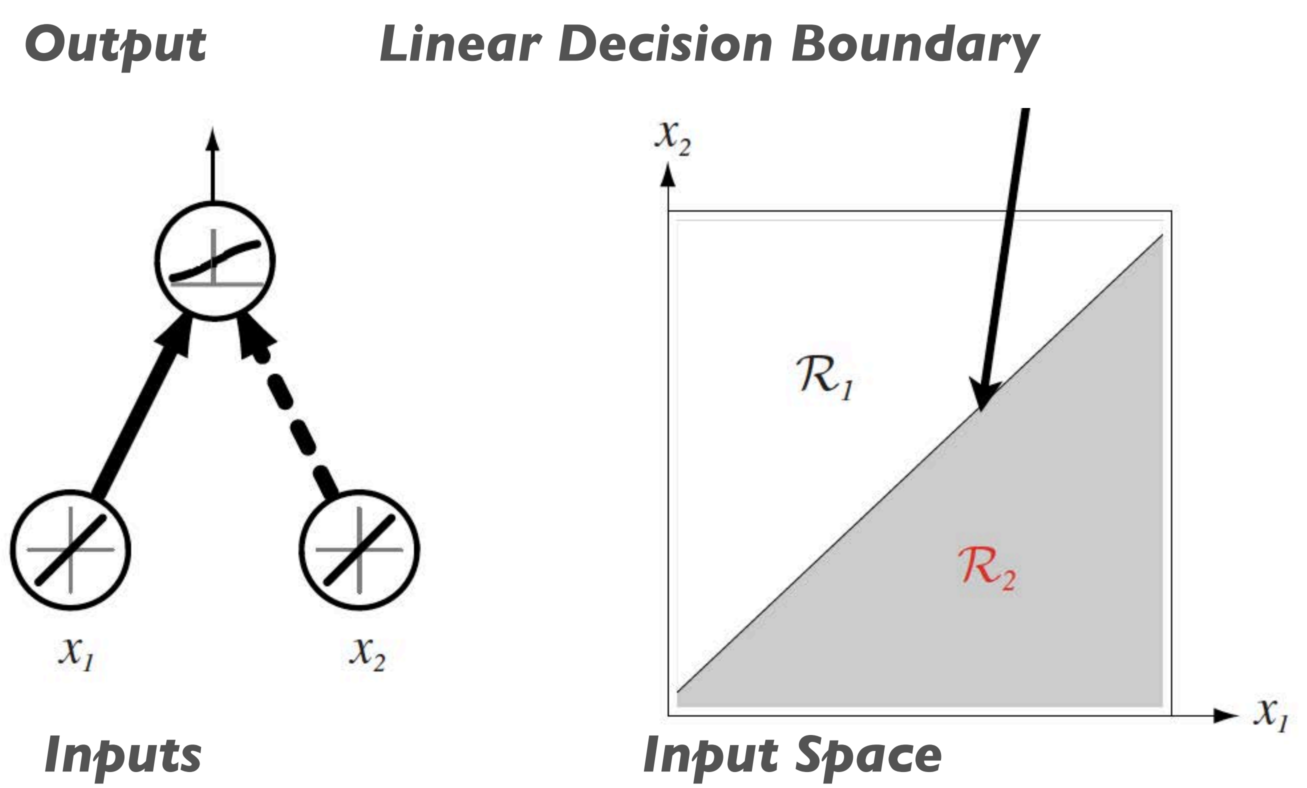


Perceptron Model (McCulloch-Pitts)



Single Neuron: A linear classifier

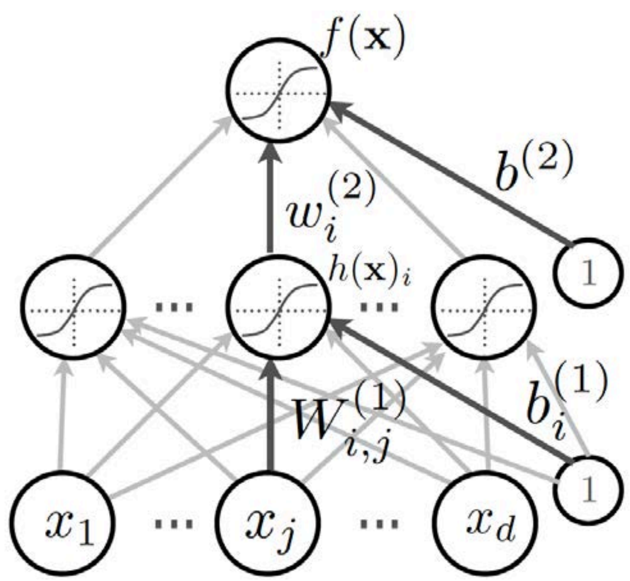




Neural Network: Nonlinear mapping

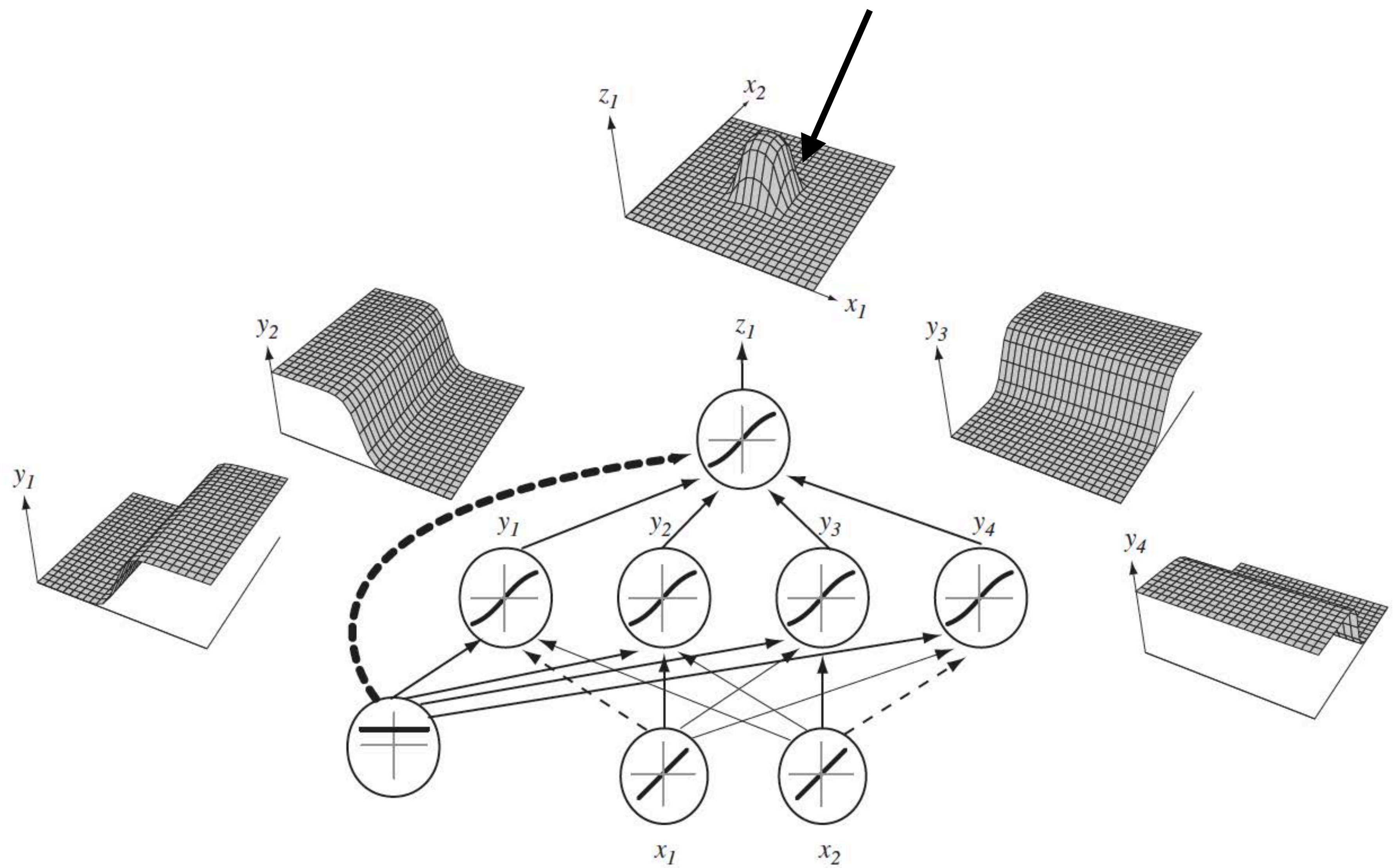
Single Hidden-Layer Network

Output



Inputs

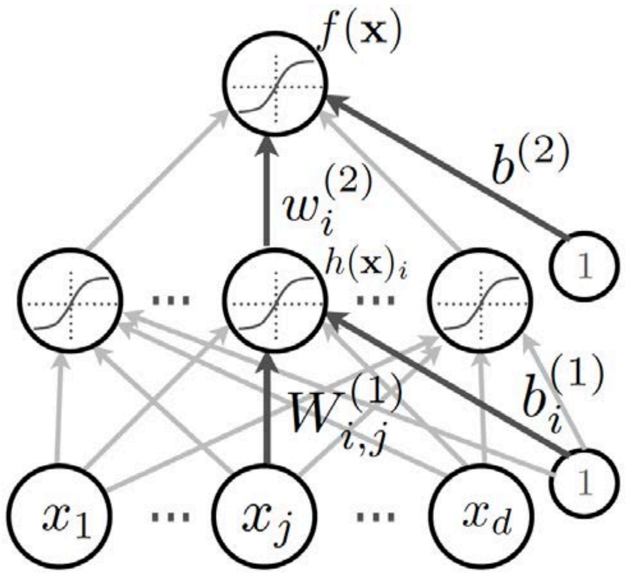
Nonlinear Decision Boundary



Universal Approximation Theorem

Multi-Layer Neural Network

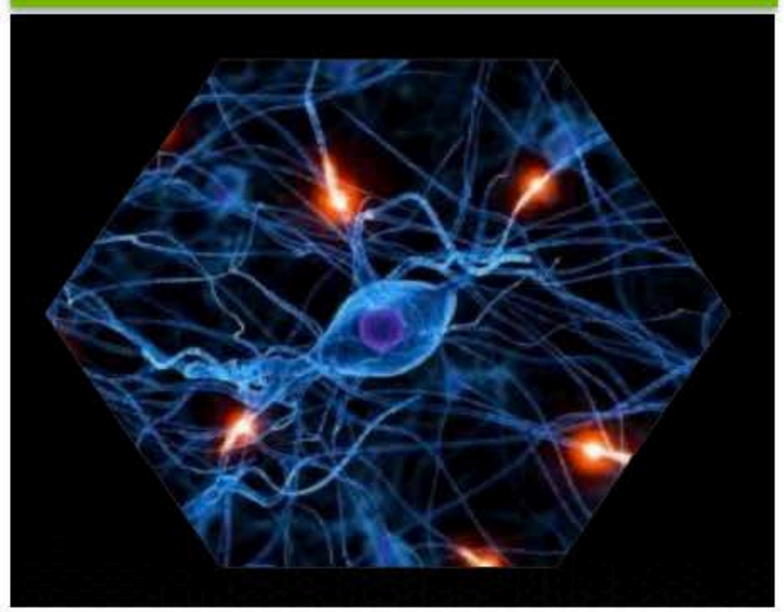
Output



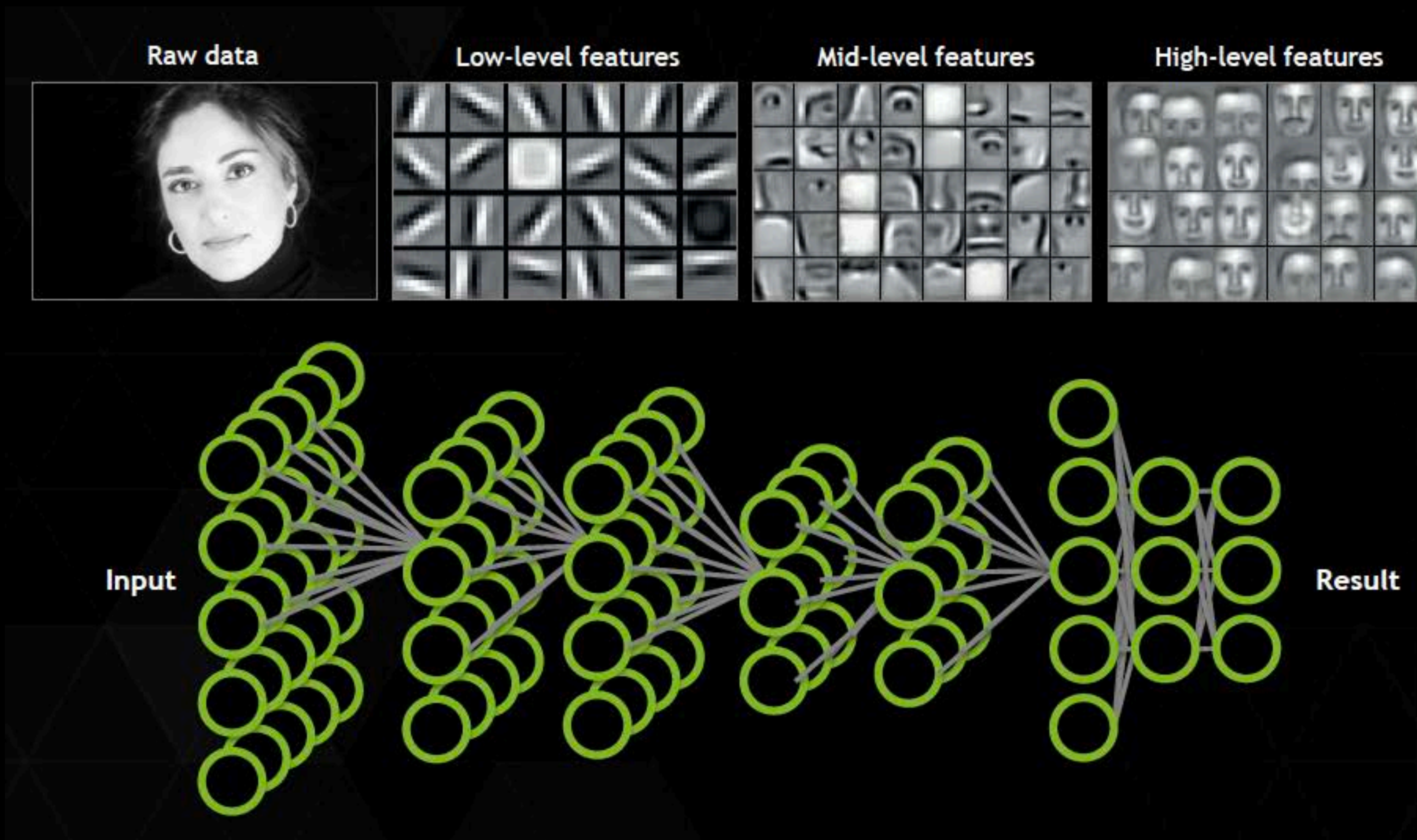
Inputs

- Early 1990s for single hidden-layer networks
- A universal approximator
- Model any continuous nonlinear function (given a sufficient number of neurons)
- <u>No guidance on how to find model parameters…</u>

Why is Deep Learning Hot Today?

Big Data Availability	New DL Techniques	GPU acceleration
350 millions images uploaded per day 2.5 Petabytes of customer data hourly 100 hours of video uploaded every minute		

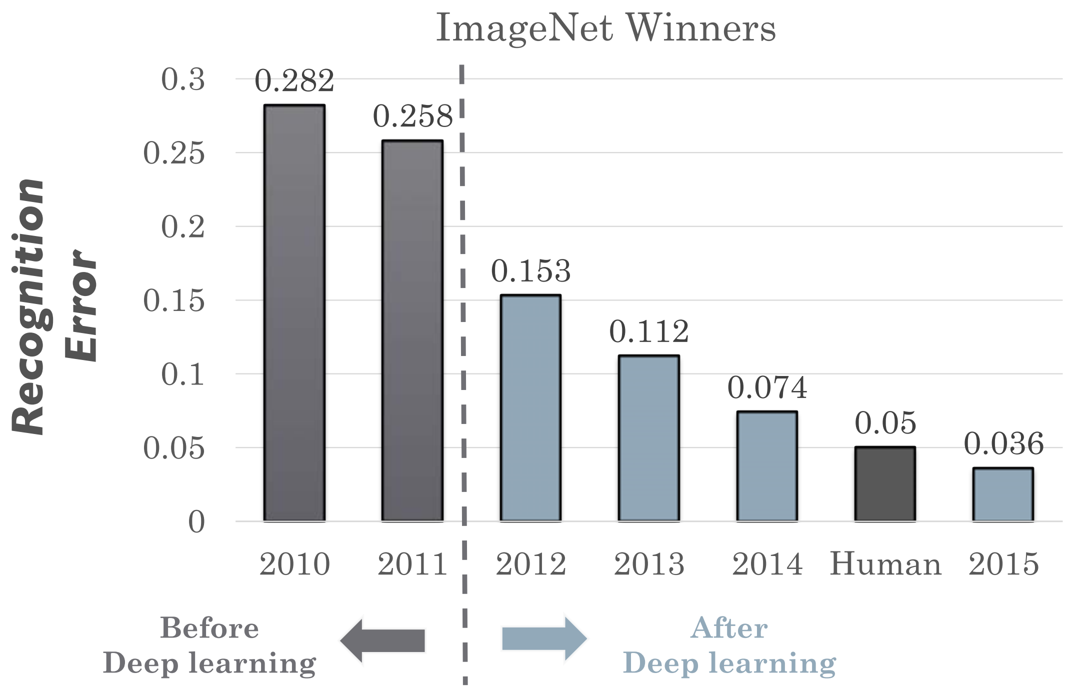
Deep Neural Networks



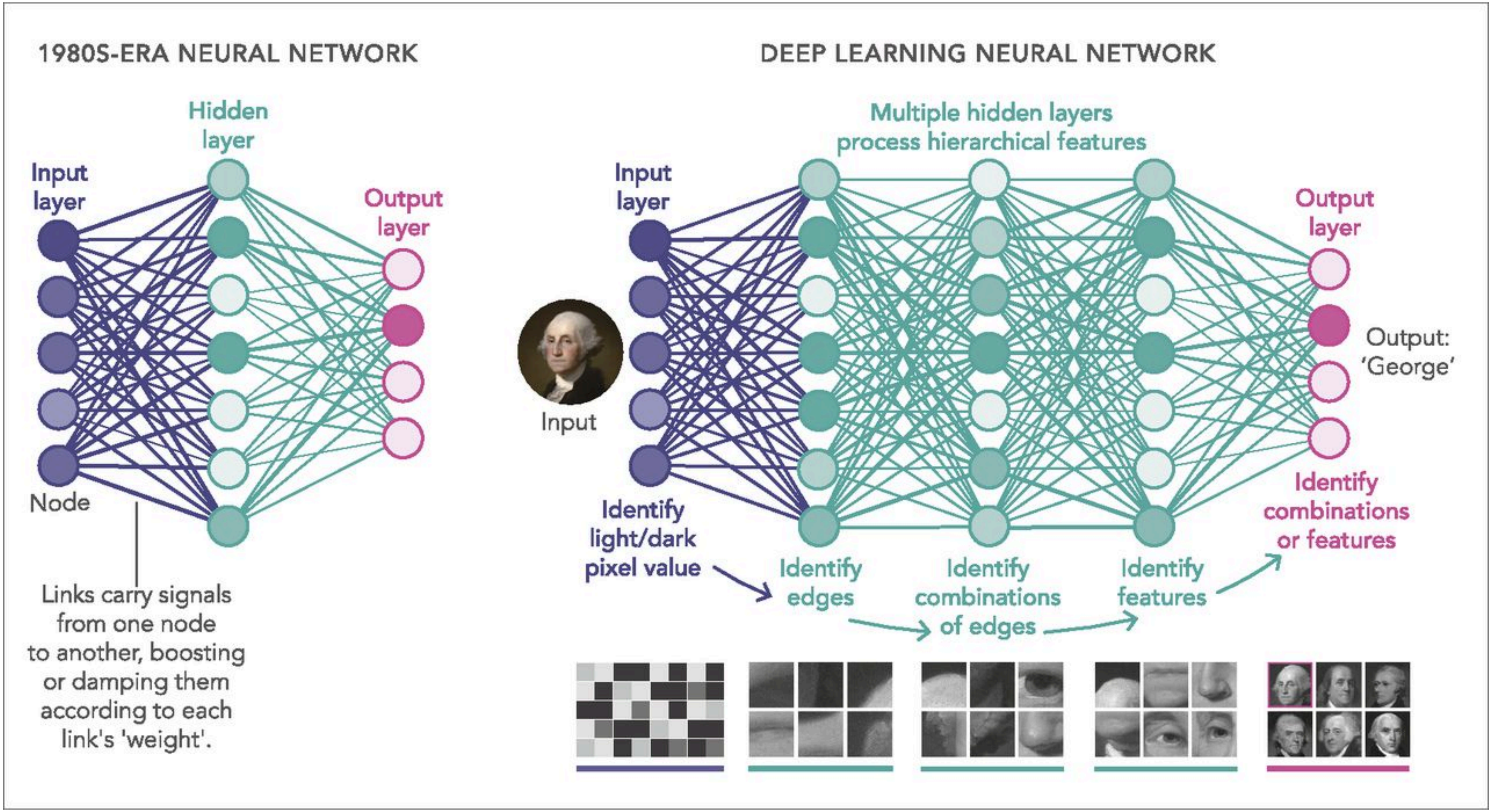
Application components:

Task objective
    e.g. Identify face
Training data
    10-100M images
Network architecture
    ~10 layers
    1B parameters
Learning algorithm
    ~30 Exaflops
    ~30 GPU days

ImageNet Object Recognition Challenge



From Blackbox Models to Dark Magic?



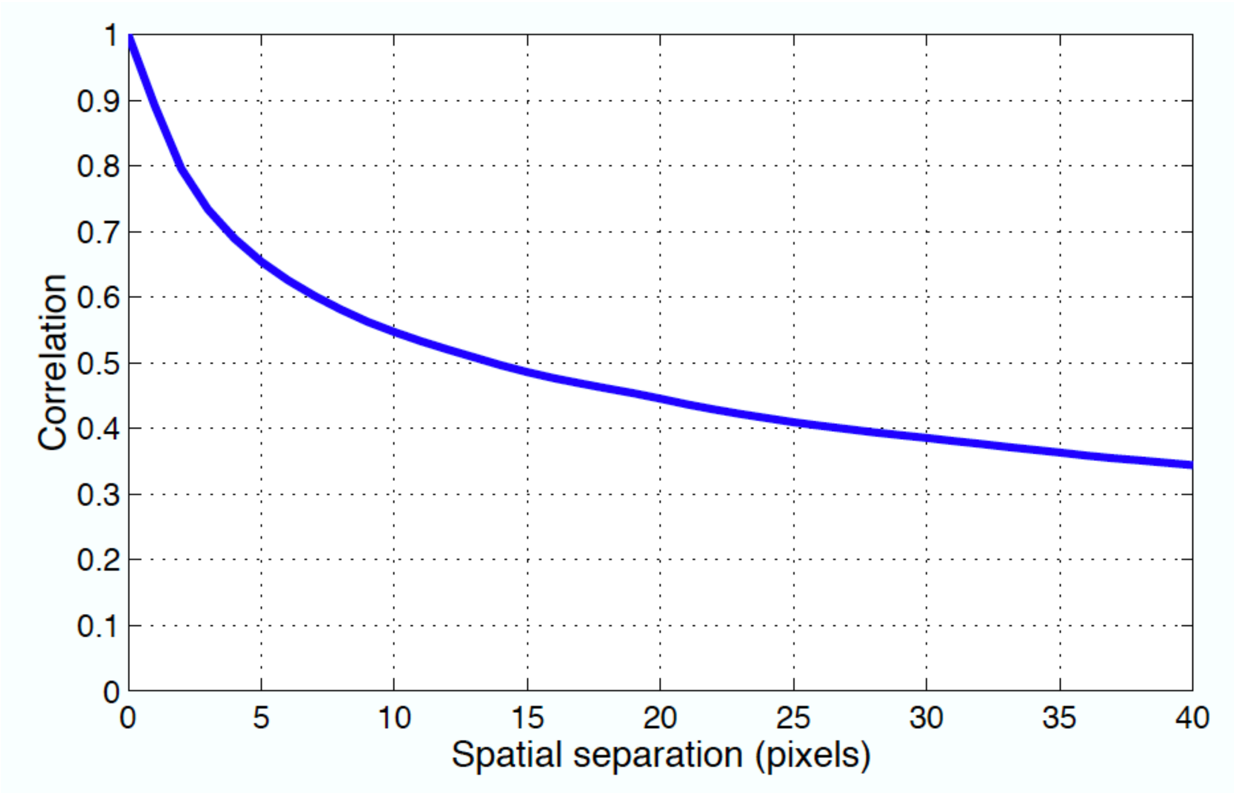
Task-Specific Priors

Task:

Cat?

Priors:

Locally-Coded Features



Task-Specific Priors

Task:

Cat?

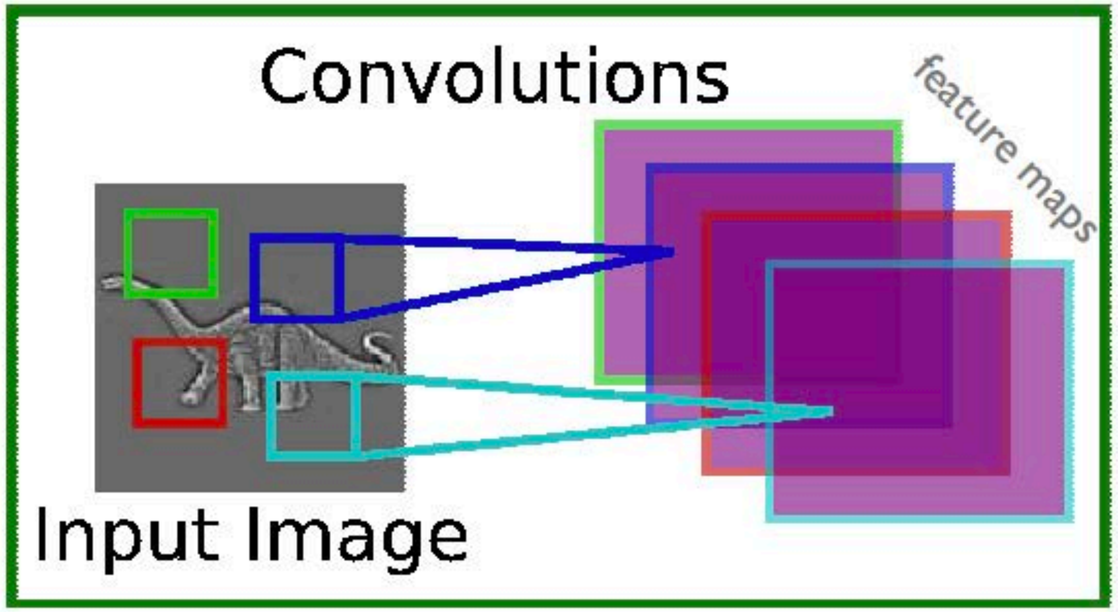
Priors:

Spatially Invariant

Scale Invariant

Ideas: Convolutional Layer

- Filtering with a kernel in small neighborhoods: local processing
- Parameter sharing: translation invariance

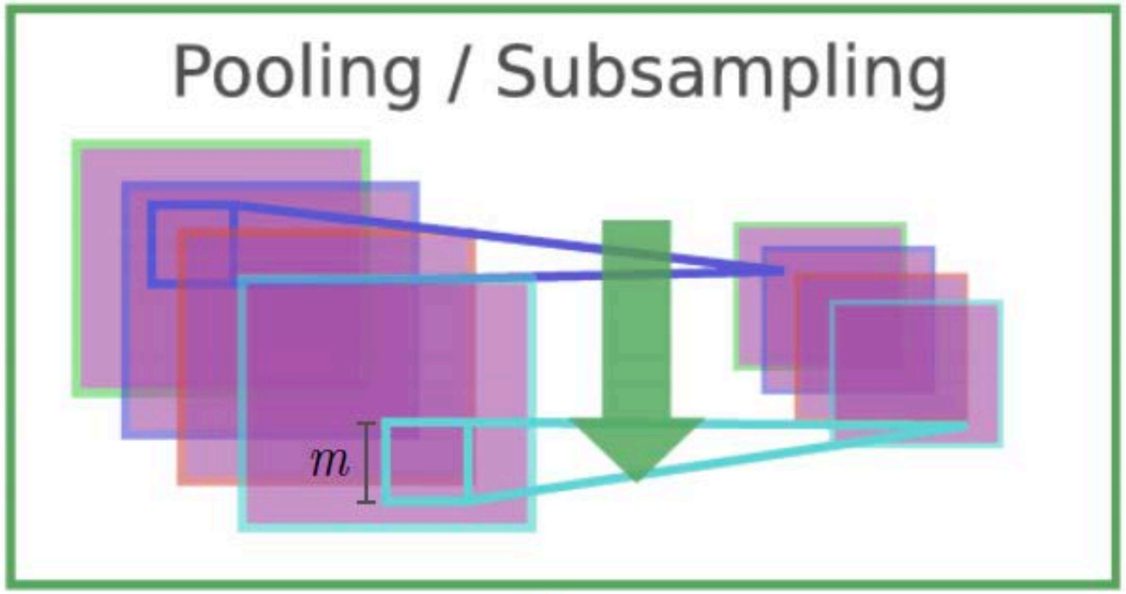


- x_i is the ith channel of input
- k_{ij} is the convolution kernel
- g_j is a learned scaling factor
- y_j is the hidden layer

$$y_j = g_j \tanh(\sum_i k_{ij} * x_i)$$

Ideas: Pooling Layer

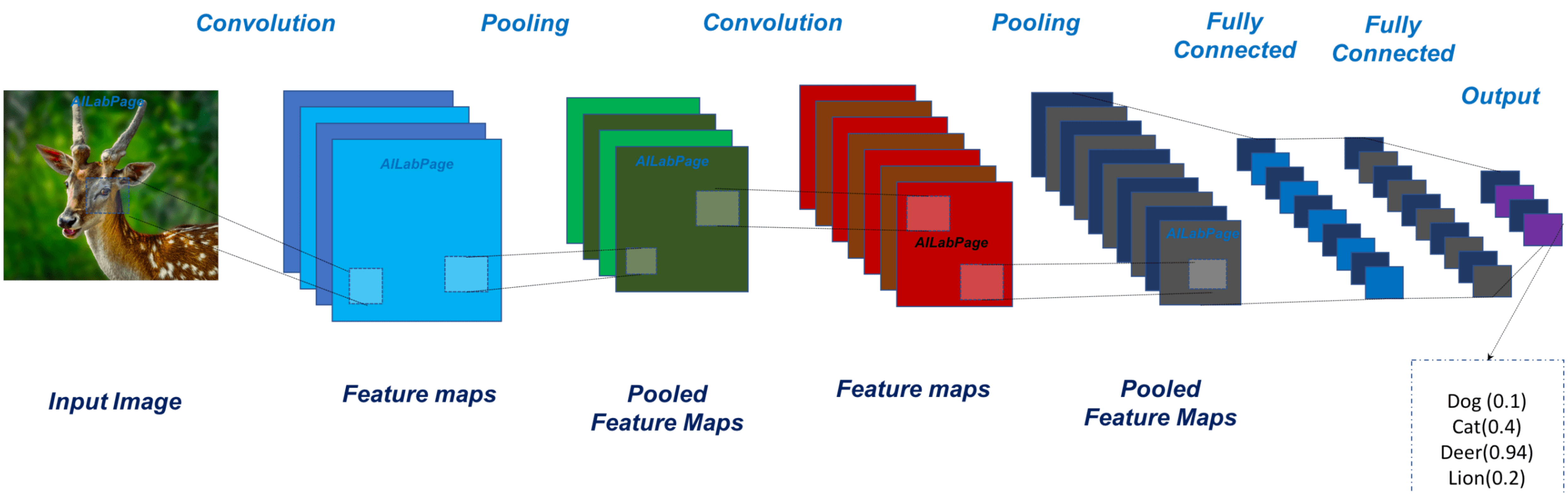
- Pooling hidden-unit responses: translation invariance
- Analysis at multiple-spatial scales: scale invariance



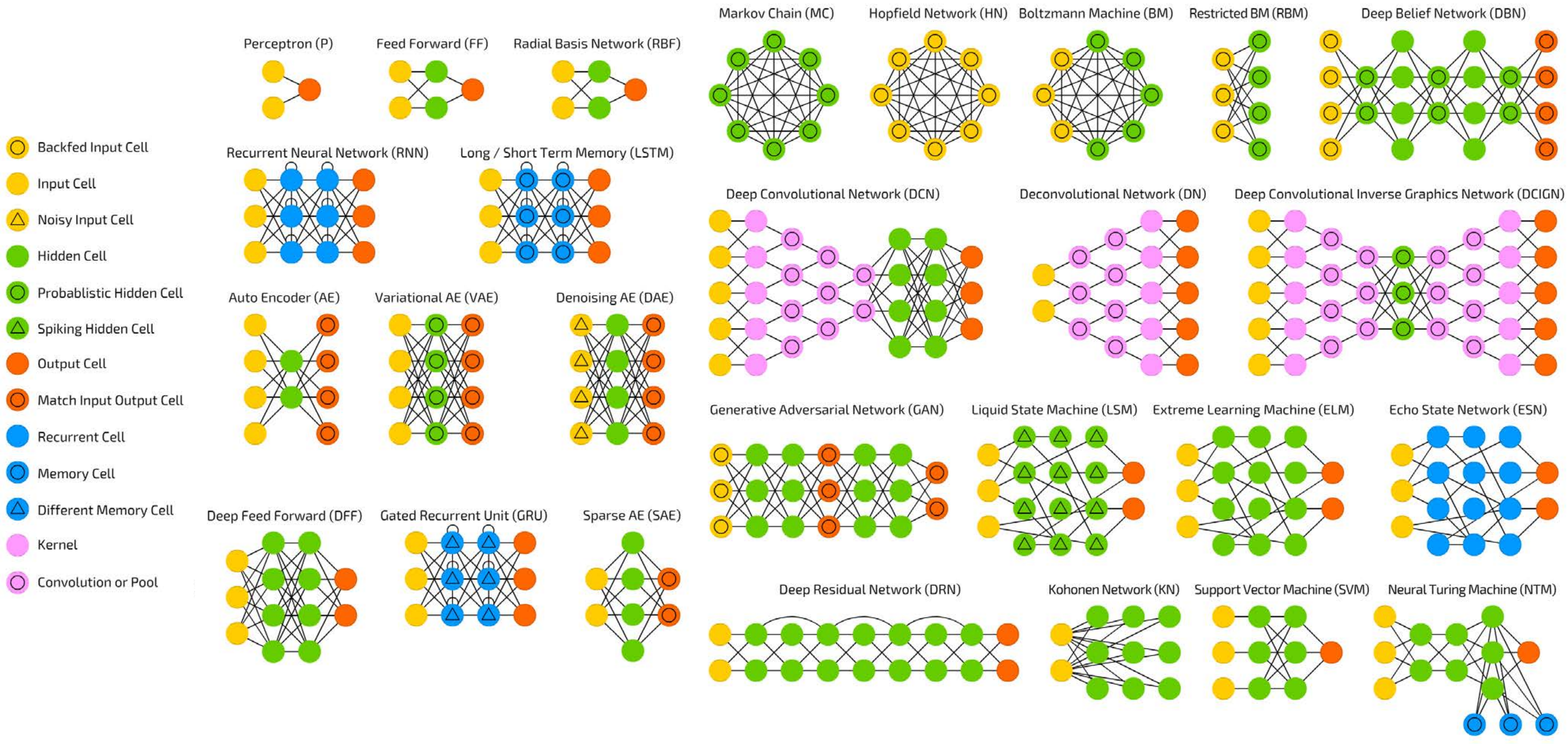
$$y_{ijk} = \frac{1}{m^2} \sum_{p,q} x_{i,j+p,k+q}$$

- $x_{i,j,k}$ is value of the i^th feature map at position j,k
- p is vertical index in local neighborhood
- q is horizontal index in local neighborhood
- y_{ijk} is pooled and subsampled layer
- m is the neighborhood height/width

Convolutional Neural Networks



Network Engineering



PART II: History of ML in Medicine

Expert Systems

1970's: MYCIN expert system

- 1970's (Stanford): MYCIN expert system for identifying bacteria causing severe infections
- Proposed a good therapy in ~69% of cases. Better than infectious disease experts

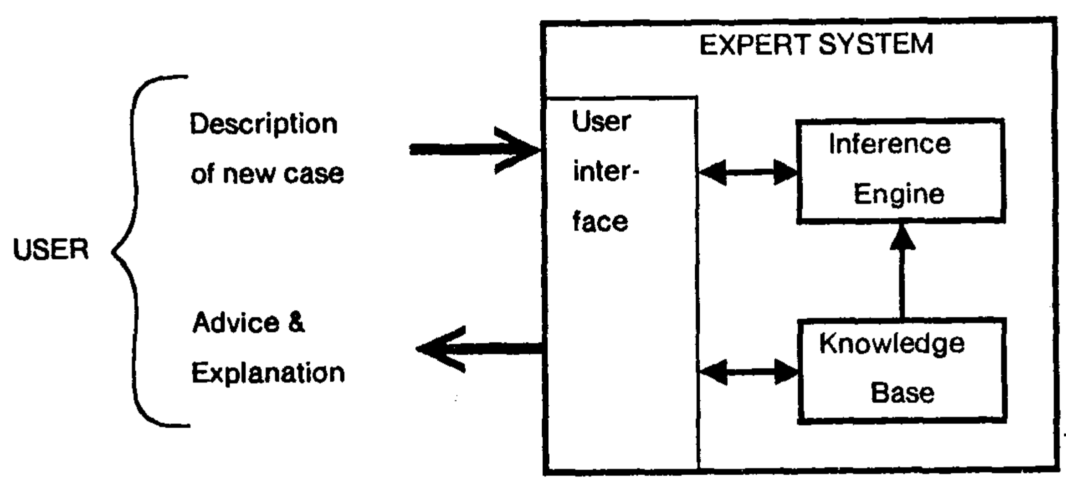


FIGURE 1-1 Major parts of an expert system. Arrows indicate information flow.

Dialogue interface

I am ready

** THIS IS A 26 YEAR OLD MALE PATIENT

My understanding is:
The age of the patient is 26
The sex of the patient is male

** FIVE DAYS AGO, HE HAD RESPIRATORY-TRACT SYMPTOMS

What is his name?

** JO

My understanding is:
The name of the patient is Jo
Respiratory-tract is one of the symptoms that the patient had

** A COUPLE OF DAYS BEFORE THE ADMISSION, HE HAD A MALAISE

Please give me the date of admission

** MARCH 12, 1979

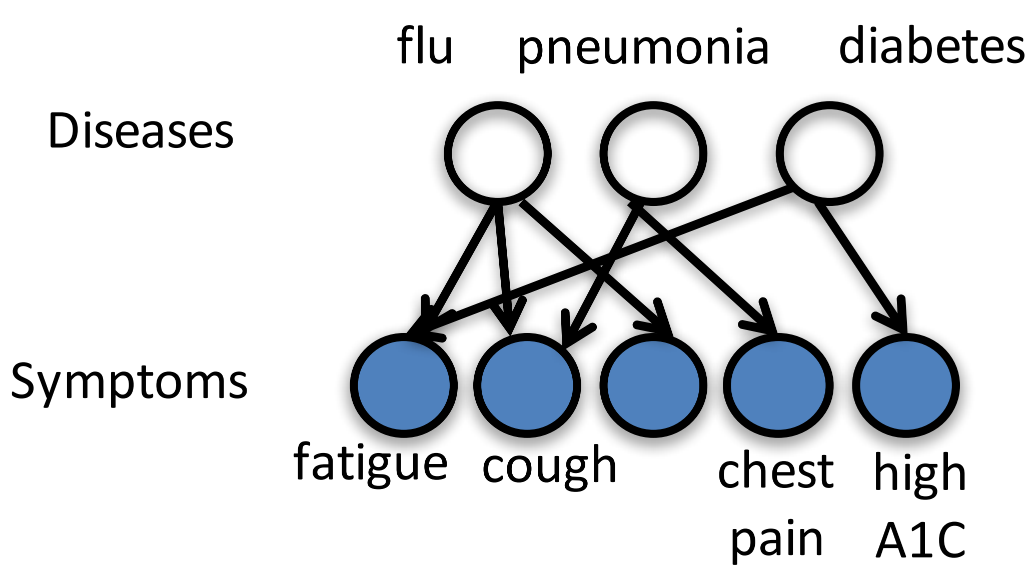
My understanding is:
The patient was admitted at the hospital 3 days ago
Malaise is one of the symptoms that the patient had 5 days ago

FIGURE 33-1 Short sample dialogue. The physician's inputs appear in capital letters after the double asterisks.

QMR Models

1980's: INTERNIST-1/QMR model

- 1980's (Univ. of Pittsburgh): INTERNIST-1/Quick Medical Reference
- Diagnosis for internal medicine



Probabilistic model relating:

570 binary disease variables
4,075 binary symptom variables
45,470 directed edges

Elicited from doctors:
15 person-years of work

Led to advances in ML & AI (Bayesian networks, approximate inference)

Problems:
1. Clinicians entered symptoms *manually*
2. Difficult to maintain, difficult to generalize

Drug Discovery

1980's: automating medical discovery

RX PROJECT: AUTOMATED KNOWLEDGE ACQUISITION

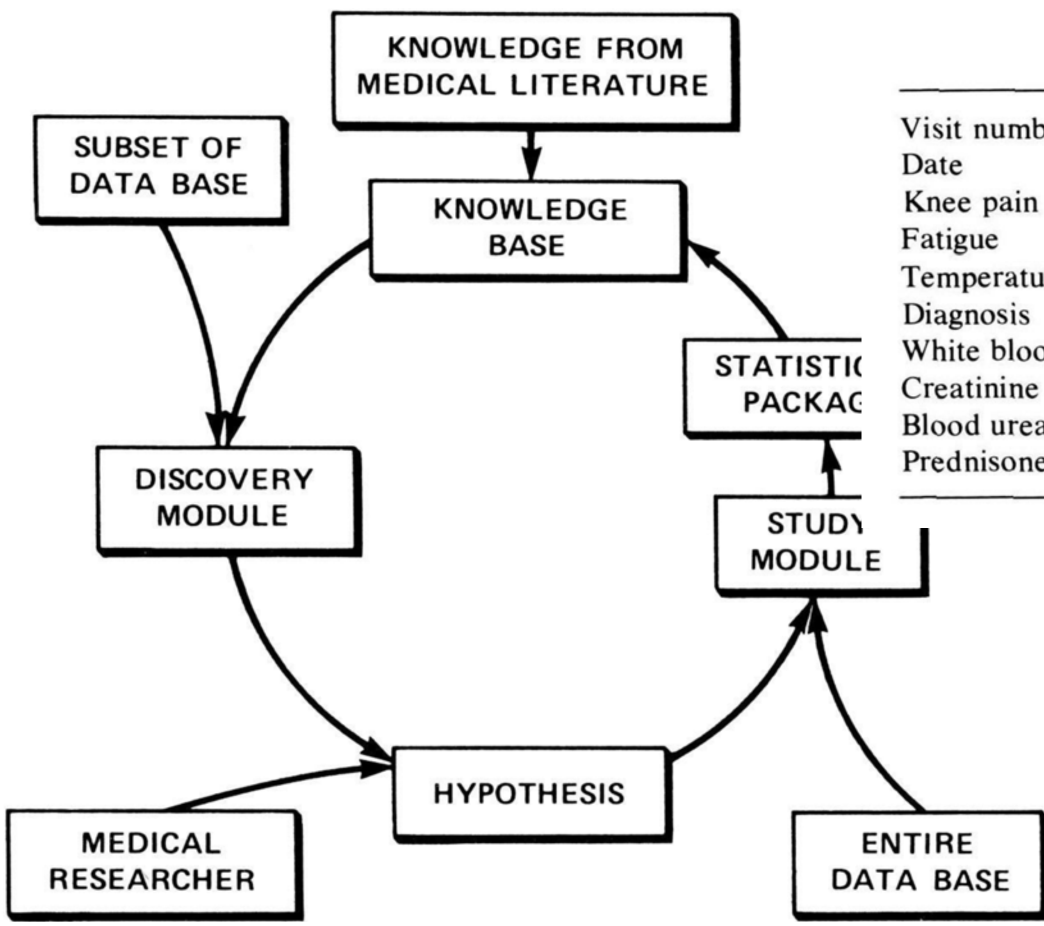


TABLE 1

HYPOTHETICAL TIME-ORIENTED RECORD FOR ONE PATIENT

Visit number	1	2	3
Date	January 17, 79	June 23, 79	July 1, 79
Knee pain	Severe	Mild	Mild
Fatigue	Moderate	—	Moderate
Temperature	38.5	37.5	36.9
Diagnosis	Systemic lupus		
White blood count	3500	4700	4300
Creatinine clearance	45	—	65
Blood urea nitrogen	36	33	—
Prednisone	30	25	20

Discovers that prednisone elevates cholesterol
(Annals of Internal Medicine, '86)

Neural Networks in Medicine

1990's: neural networks in medicine

- Neural networks with clinical data took off in 1990, with 88 new studies that year
- Small number of features (inputs)
- Data often collected by chart review

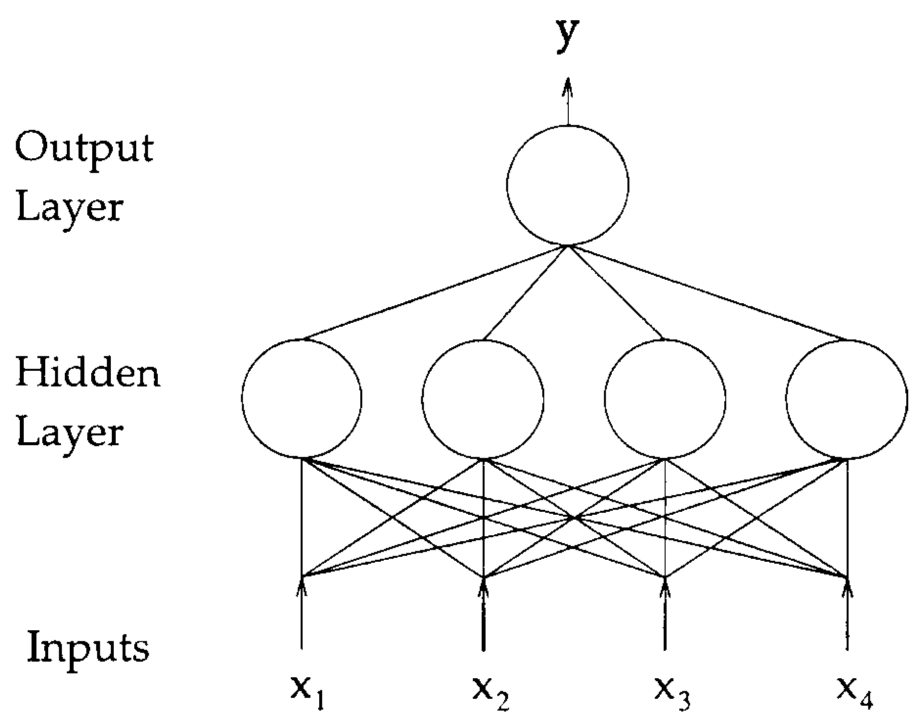


FIGURE 2. A multilayer perceptron. This is a two-layer perceptron with four inputs, four hidden units, and one output unit.

Problems:
1. Did not fit well into clinical workflow
2. Poor generalization to new places

Disease Diagnosis

Table 1 • 25 Neural Network Studies in Medical Decision Making*

Subject	No. of Examples: Training	No. of Examples: Test	P†	Network	D‡	Accuracy§: Neural	Accuracy§: Other
Breast cancer[4]	57	20	60	9–15–2	0.6	80	75
Vasculitis[2]	404	403	73	8–5–1	8.0	94	—
Myocardial infarction[6]	351	331	89	20–10–10–1	1.1	97	84
Myocardial infarction[8]	356	350	87	20–10–10–1	1.1	97	94
Low back pain[11]	100	100	25	50–48–2	0.2	90	90
Cancer outcome[13]	5,169	3,102	—	54–40–1	1.4	0.779	0.776
Psychiatric length of stay[17]	957	106	73	48–400–4	0.2	74	76
Intensive care outcome[23]	284	138	91	27–18–1	0.5	0.82	0.82
Skin tumor[21]	150	100	80	18	—	80	90
Evoked potentials[35]	100	67	52	14–4–3	3.8	77	77
Head injury[47]	500	500	50	6–3–3	20	66	77
Psychiatric outcome[54]	289	92	60	41–10–1	0.7	79	—
Tumor classification[55]	53	6	38	8–9–3	1.4	99	88
Dementia[57]	75	18	19	80–10–7–7	0.6	61	—
Pulmonary embolism[59]	607	606	69	50–4–1	2.9	0.82	0.83
Heart disease[62]	460	230	54	35–16–8–2	3	83	84
Thyroid function[62]	3,600	1,800	93	21–16–8–3	22	98	93
Breast cancer[62]	350	175	66	9–4–4–2	10	97	96
Diabetes[62]	384	192	65	8–4–4–2	12	77	75
Mycardial infarction[63]	2,856	1,429	56	291–1	9.8	85	—
Hepatitis[65]	39	42	38	4–4–3	3.3	74	79
Psychiatric admission[76]	319	339	85	53–1–1	6.0	91	—
Cardiac length of stay[83]	713	696	73	15–12–1	3.5	0.70	—
Anti-cancer agents[89]	127	141	25	60–7–6	1.5	91	86
Ovarian cancer[91]	75	98	—	6–6–2	2.6	84	81
MEDIAN VALUE	350	175	71	20	2.8		

*For reference citations, see the reference list

†P = prior probability of most prevalent category.

‡D = ratio of training examples to weights per output.

§A single integer in the accuracy column denotes percentage overall classification rate and a single real number between 0 and 1 indicates the AUROCC value. Neural = accuracy of neural net, Other = accuracy of best other method.

PART III: Big Data in Medical Applications

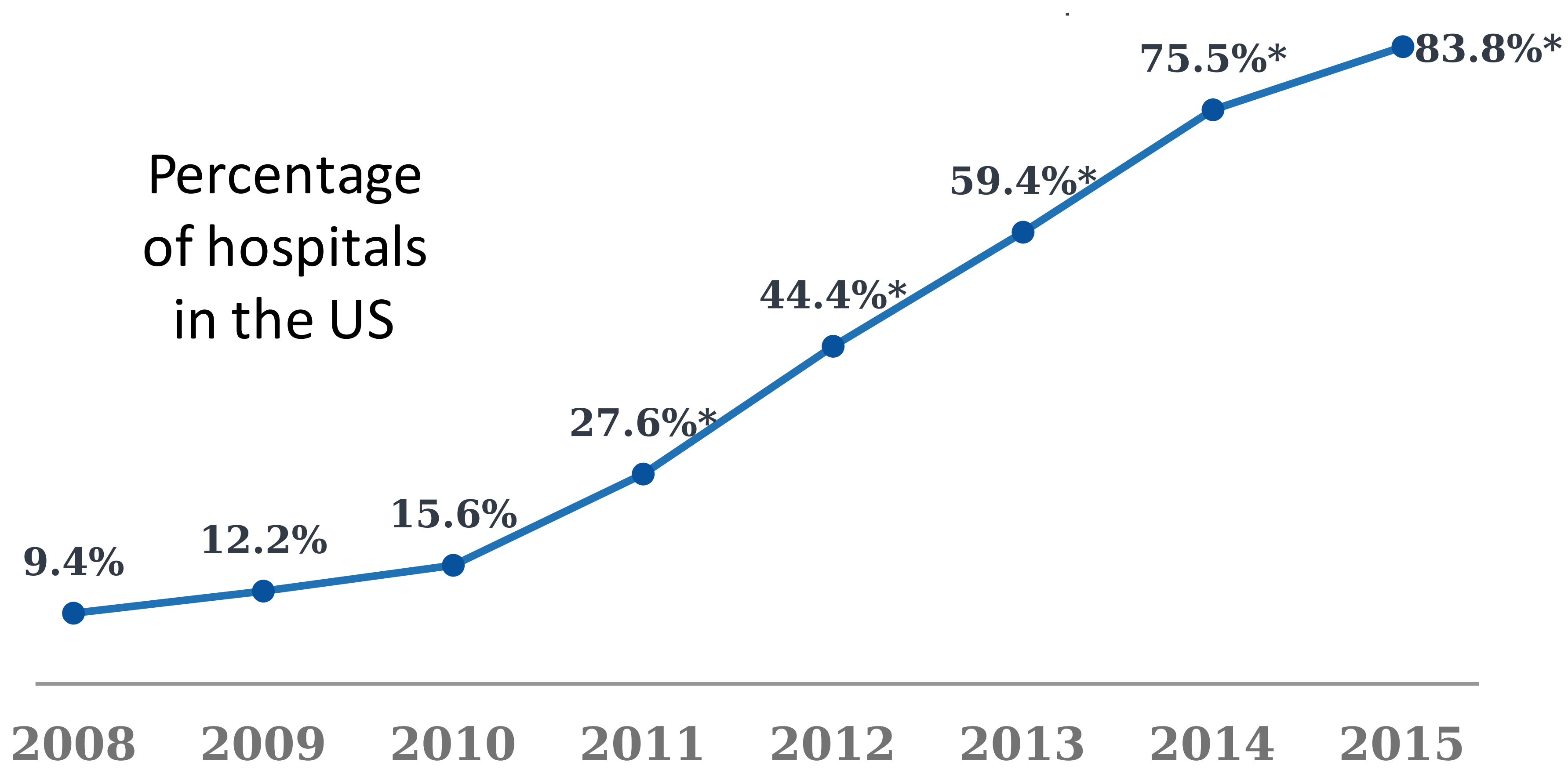
Sources of Medical Data

Sources of Medical Data

Diversity of digital health data

Availability of Medical Data

Adoption of Electronic Health Records (EHR) has increased 9x since 2008



[Henry et al., ONC Data Brief, May 2016]

Standard Classes of Data

PUBLIC

Data that may be freely disclosed to the public

INTERNAL ONLY

Internal data not meant for public disclosure

CONFIDENTIAL

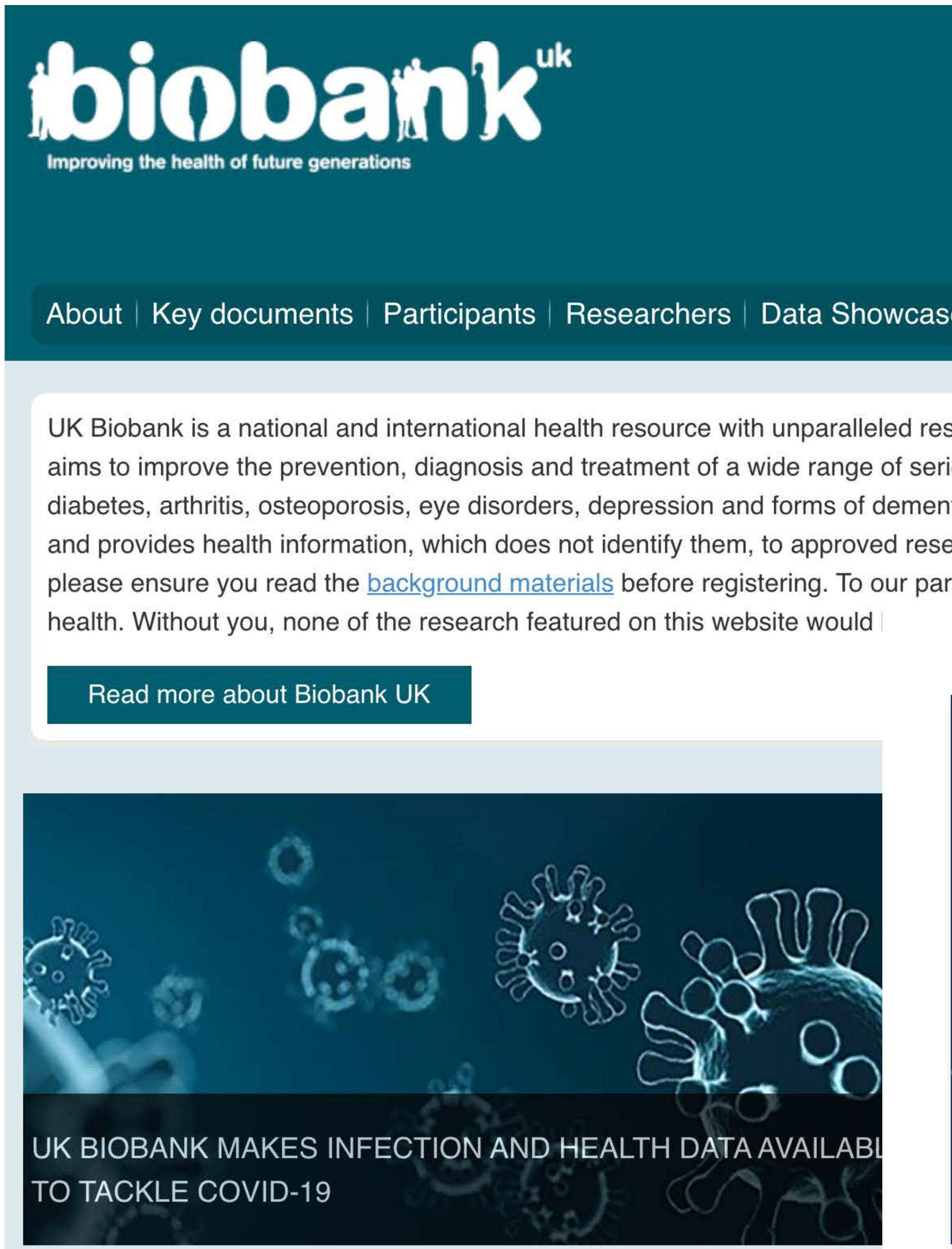
Sensitive data that if compromised could negatively affect operations

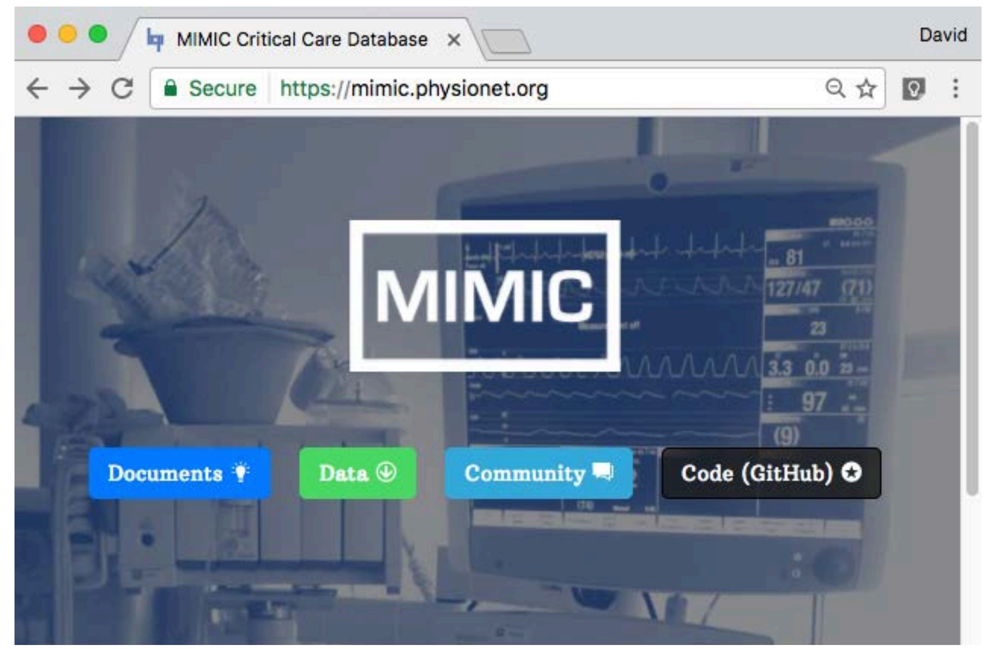
RESTRICTED

Highly sensitive corporate data that if compromised could put the organization financial or legal risk

Medical Data

Large, Public Databases are Emerging

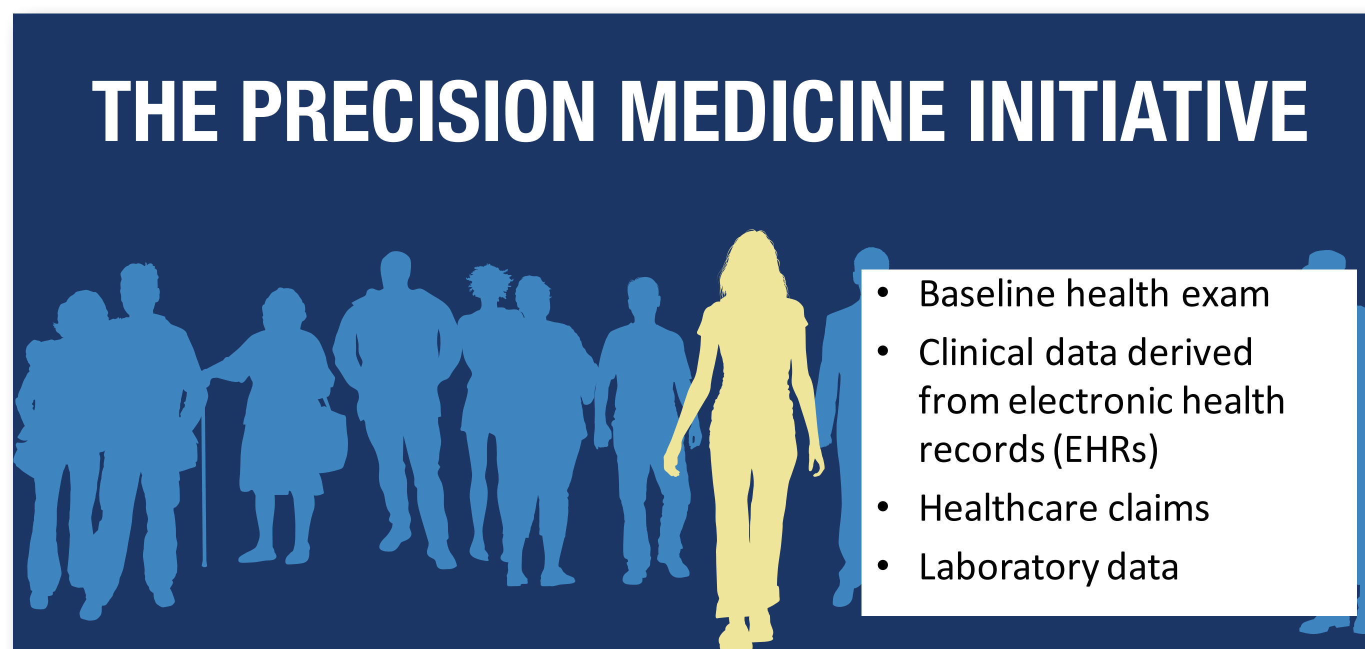




Massachusetts Institute of Technology
Laboratory for Computational Physiology

De-identified health data from ~40K critical care patients

Demographics, vital signs, laboratory tests,



PART IV: Opportunities/Challenges in Healthcare

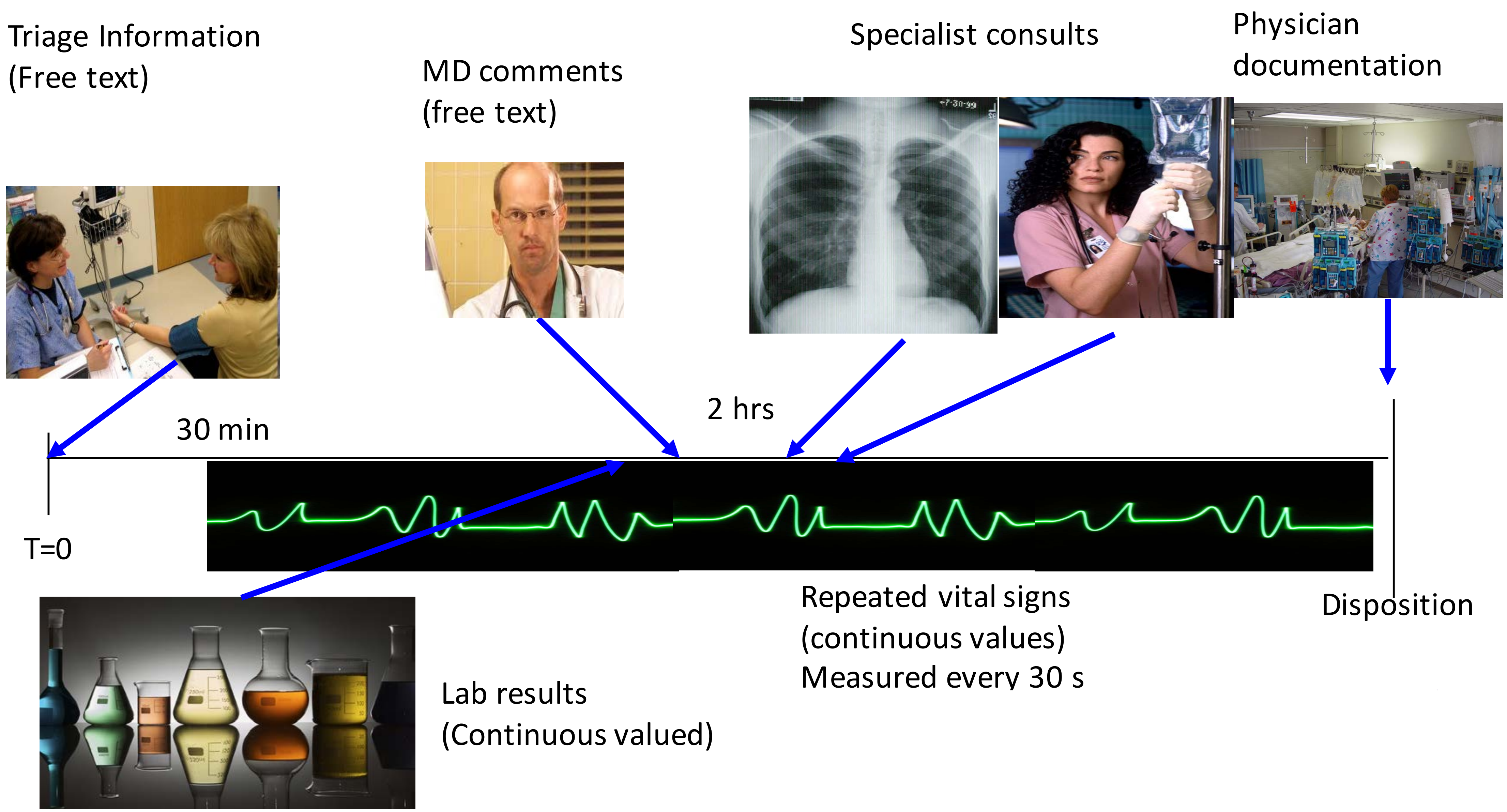
ML/DL in Biomedical Domain

Example: Emergency Departments

Emergency Department:

- **Limited resources**
- **Time sensitive**
- **Critical decisions**

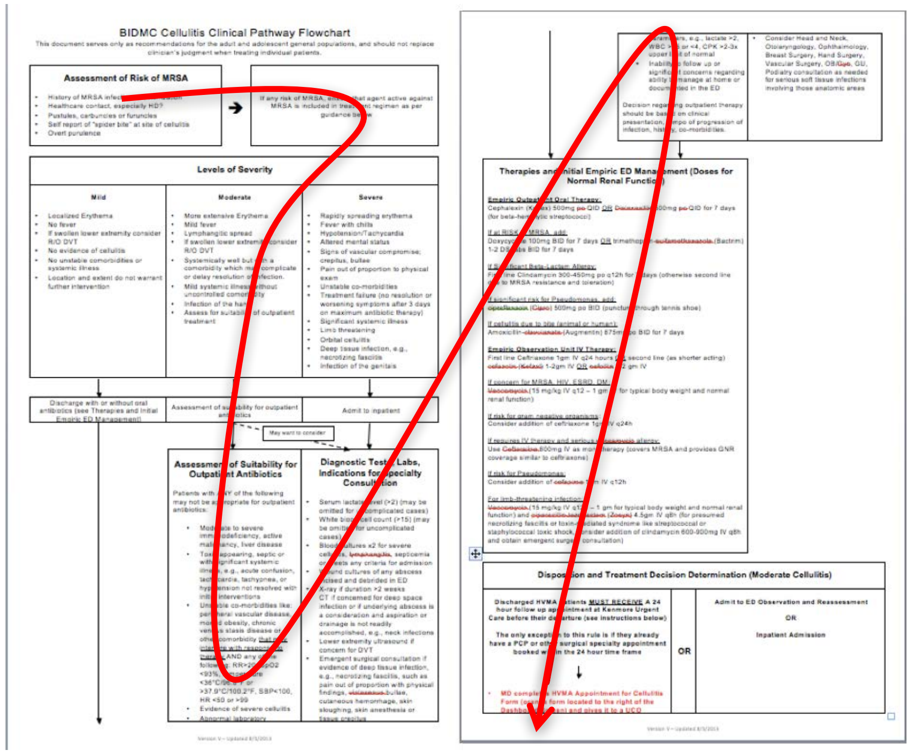
Example: Emergency Departments



How can Machine Learning Help?

- **Triggering clinical pathways**
- Context-specific displays
- Risk stratification
- Improving clinical documentation

BIDMC Cellulitis Clinical Pathway Flowchart

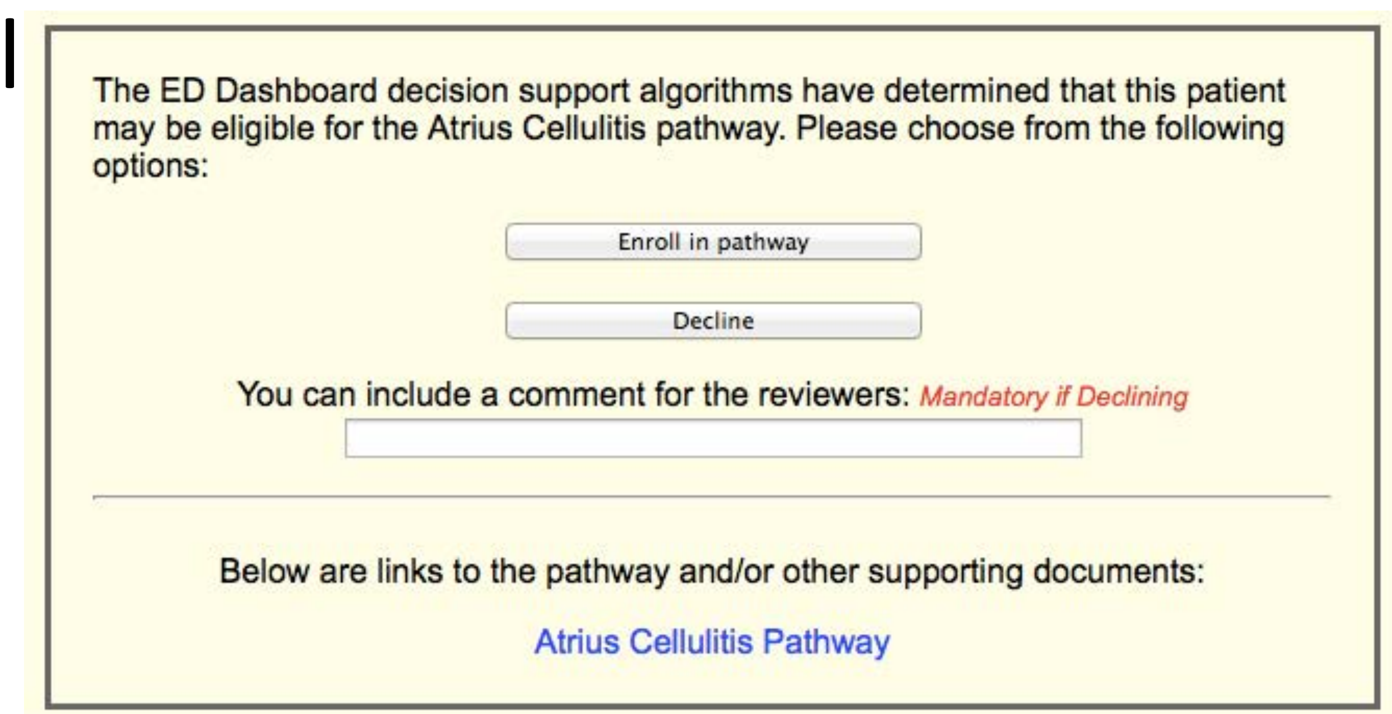


Pathways have been shown to reduce in-hospital complications without increasing costs [Rotter et al 2010]

Automatic Protocol Selection

- **Triggering clinical pathways**
- Context-specific displays
- Risk stratification
- Improving clinical documentation

Automating triggers
Don't rely on the user's knowledge that the pathway exists!


The ED Dashboard decision support algorithms have determined that this patient may be eligible for the Atrius Cellulitis pathway. Please choose from the following options:

Enroll in pathway

Decline

You can include a comment for the reviewers: *Mandatory if Declining*

Below are links to the pathway and/or other supporting documents:

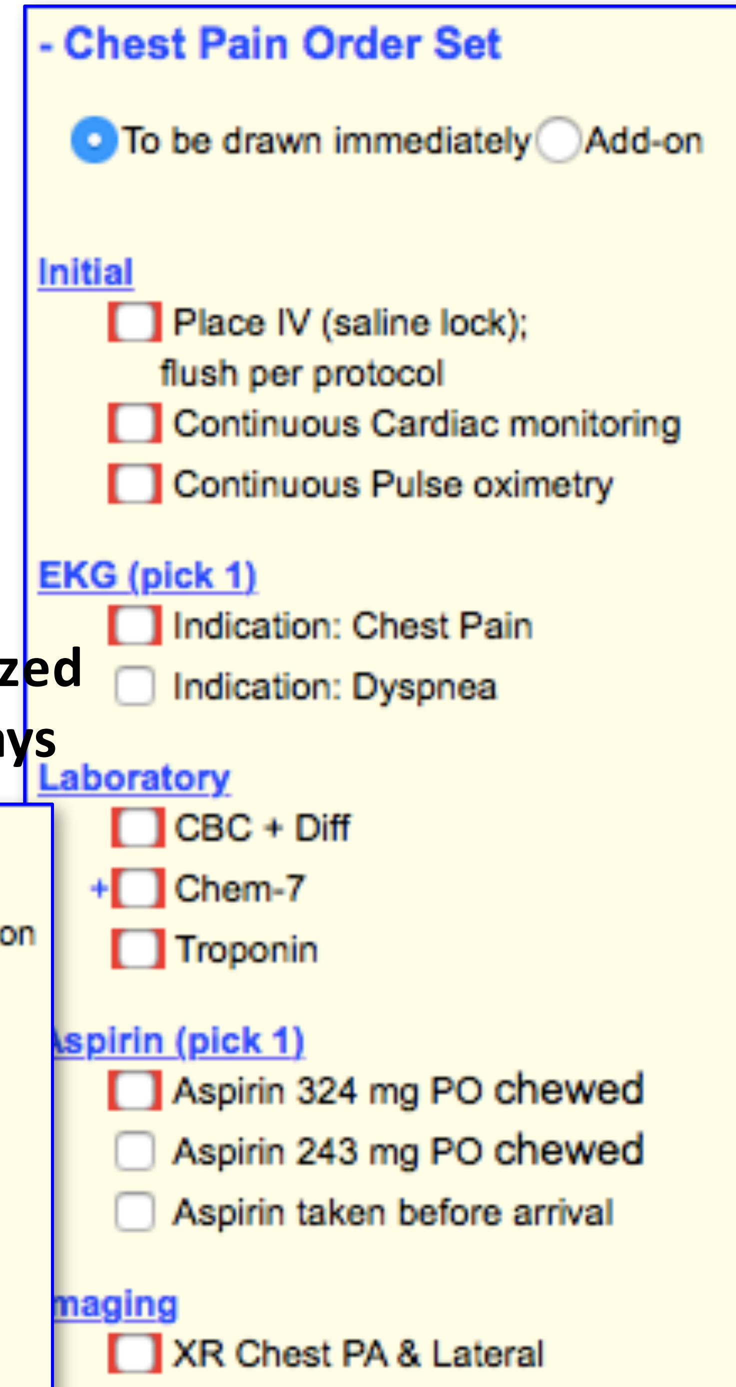
Atrius Cellulitis Pathway

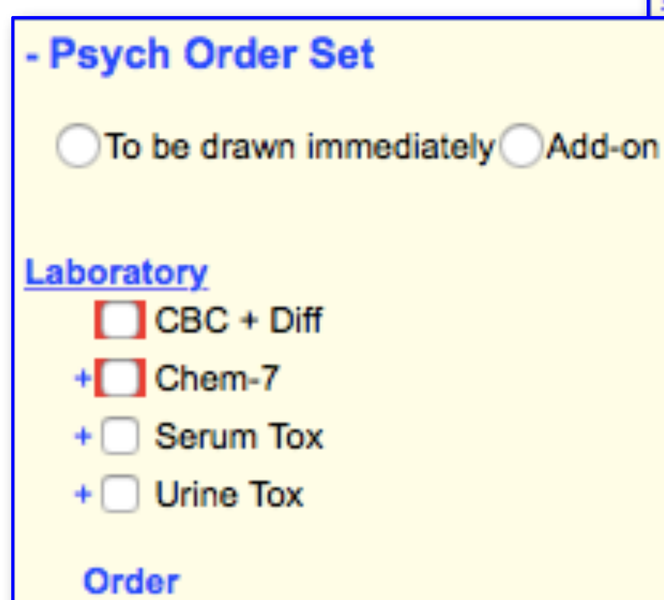
Our task:
Determine whether a patient has or is suspected to have cellulitis

Disease-specific Recommender Systems

- Triggering clinical pathways
- **Context-specific displays**
- Risk stratification
- Improving clinical documentation

Automatically place **specialized order sets on patient displays**





Our task:
Determine whether patient complained of chest pain, or is a psych patient

Minimizing Risk

- Triggering clinical pathways
- Context-specific displays
- **Risk stratification**
- Improving clinical documentation

Ex 1: Likelihood of mortality or admission to ICU

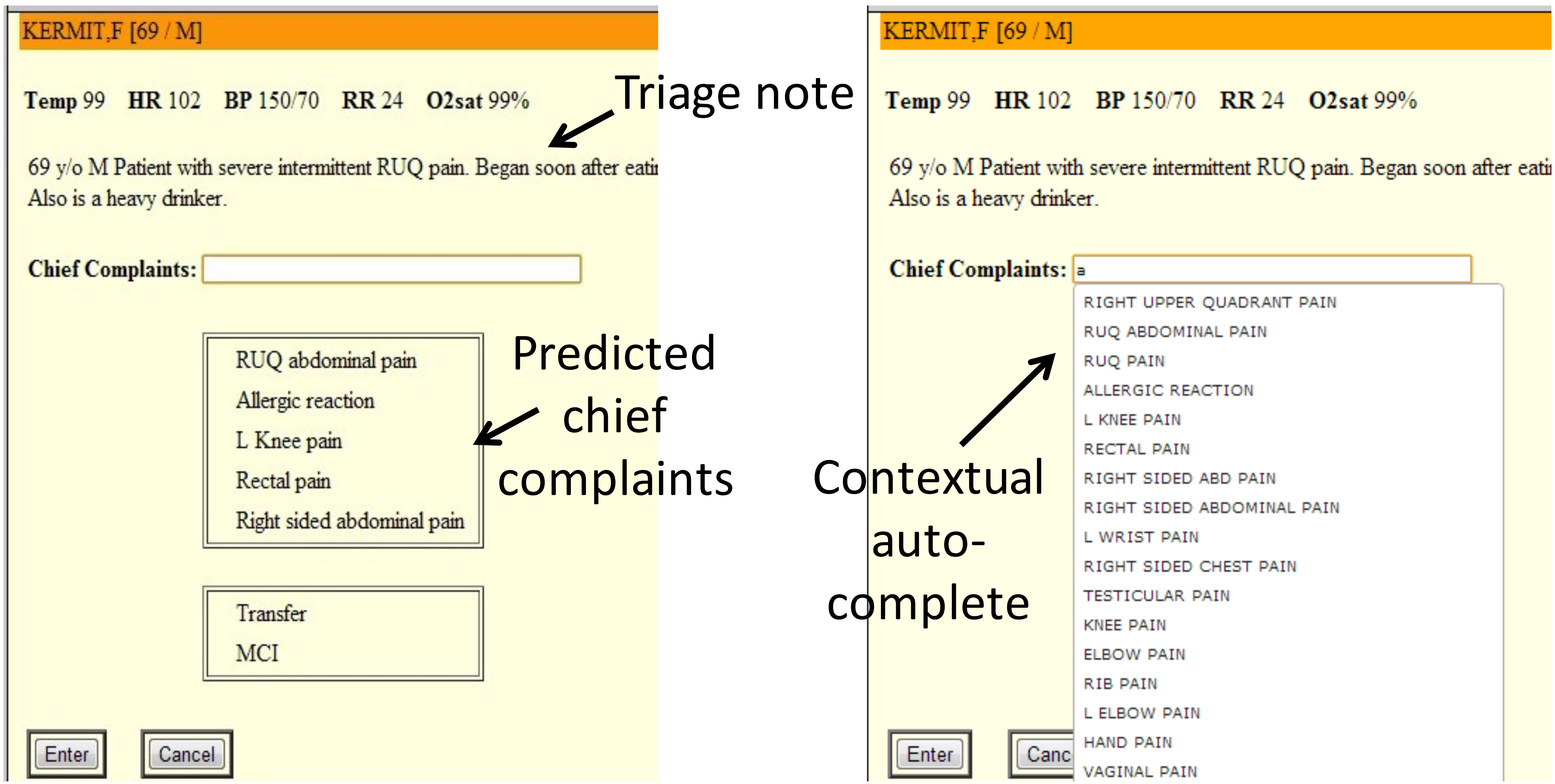
Ex 2: Early detection of severe sepsis

Real-time Disease Prediction

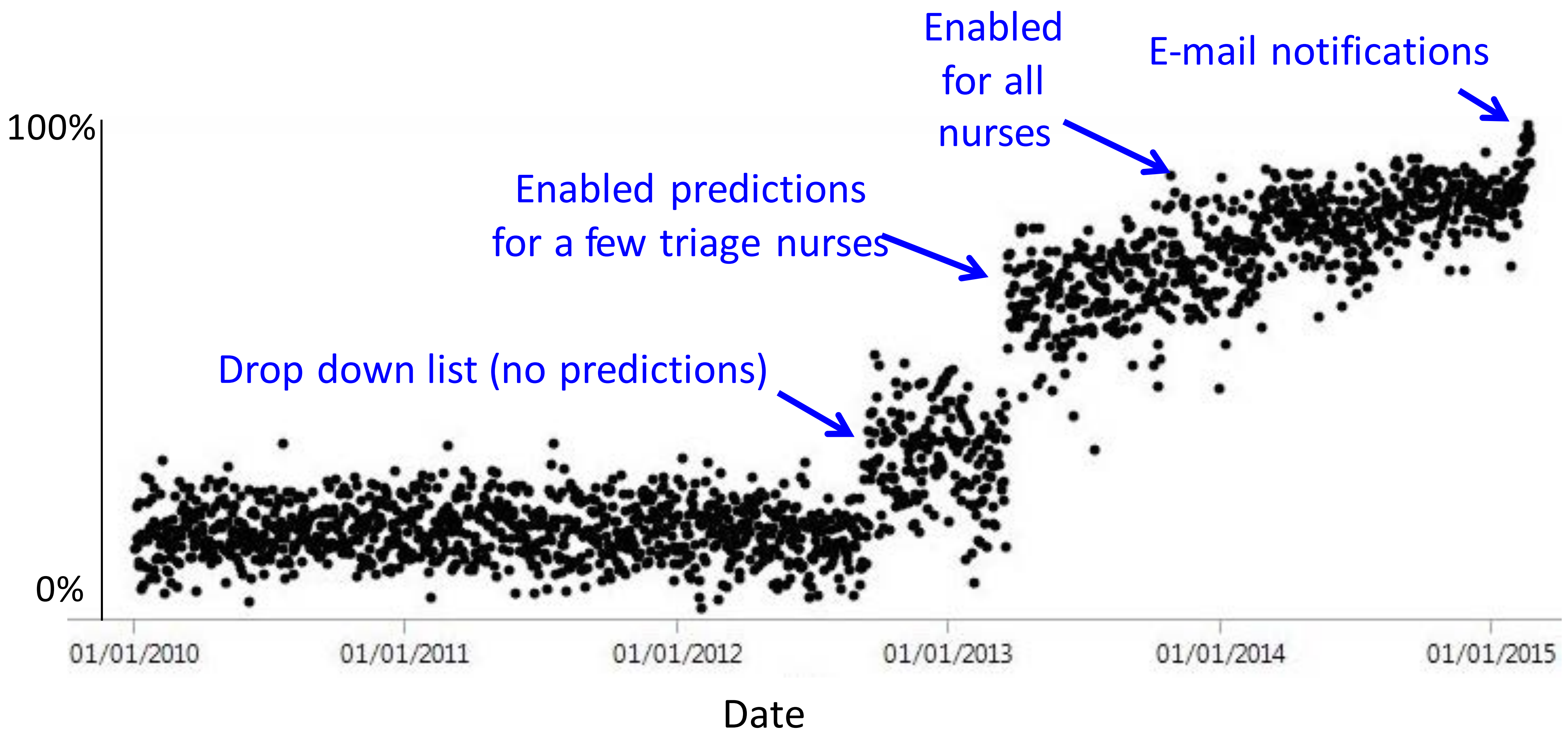
History
Alcoholism
Anticoagulated
Asthma/COPD
Cancer
Congestive heart failure
Diabetes
HIV+
Immunosuppressed
Liver malfunction

Acute		
Abdominal pain	Deep vein thrombosis	Laceration
Allergic reaction	Employee exposure	Motor vehicle accident
Ankle fracture	Epistaxis	Pancreatitis
Back pain	Gastroenteritis	Pneumonia
Bicycle accident	Gastrointestinal bleed	Psych
Cardiac etiology	Geriatric fall	Obstruction
Cellulitis	Headache	Septic shock
Chest pain	Hematuria	Severe sepsis
Cholecystitis	Intracerebral hemorrhage	Sexual assault
Cerebrovascular accident	Infection	Suicidal ideation
	Kidney stone	Syncope
		Urinary tract infection

Improving Clinical Documentation

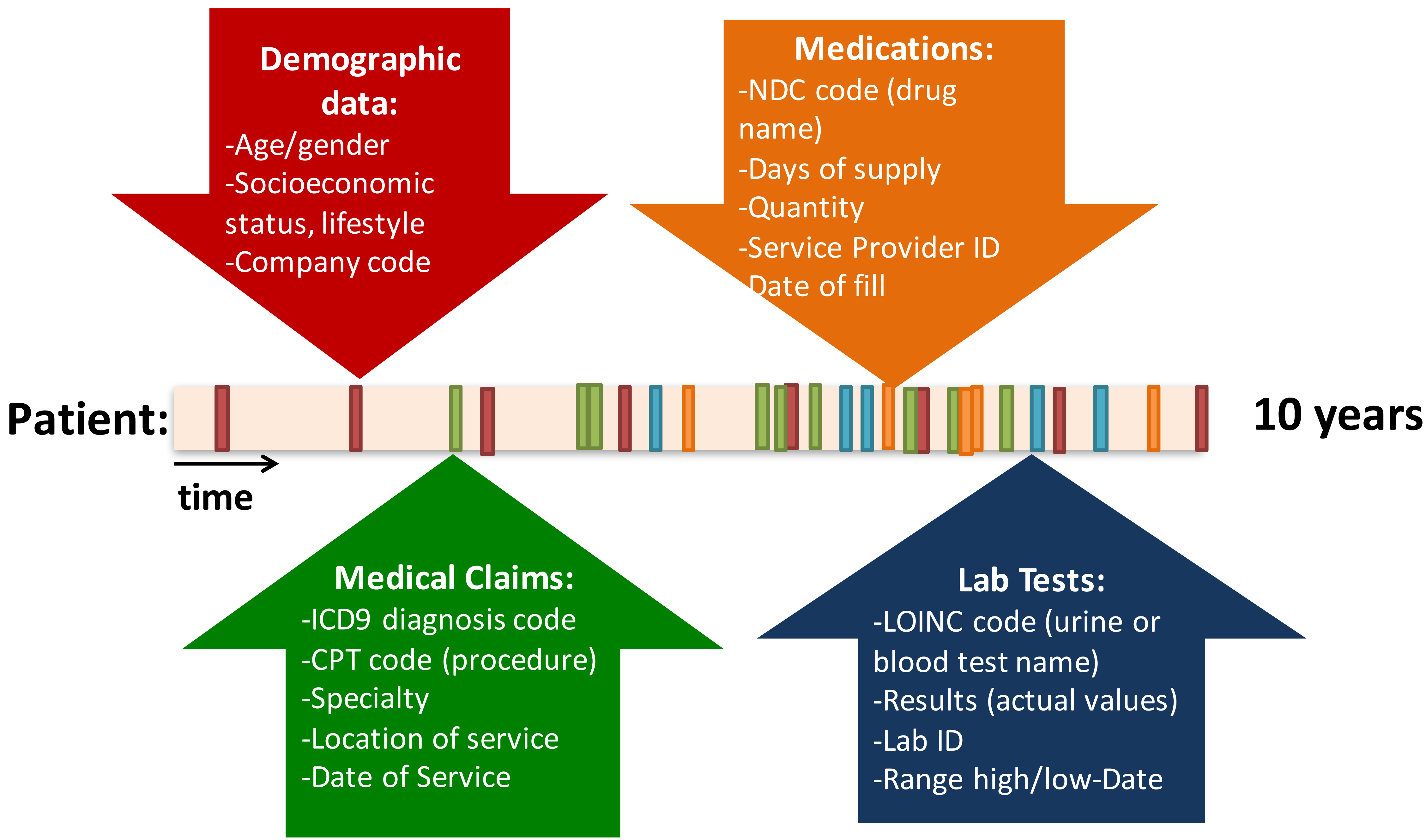


Improving Clinical Documentation



Percentage of *standardized* chief complaints (per week)

At a Broader Time Scale…

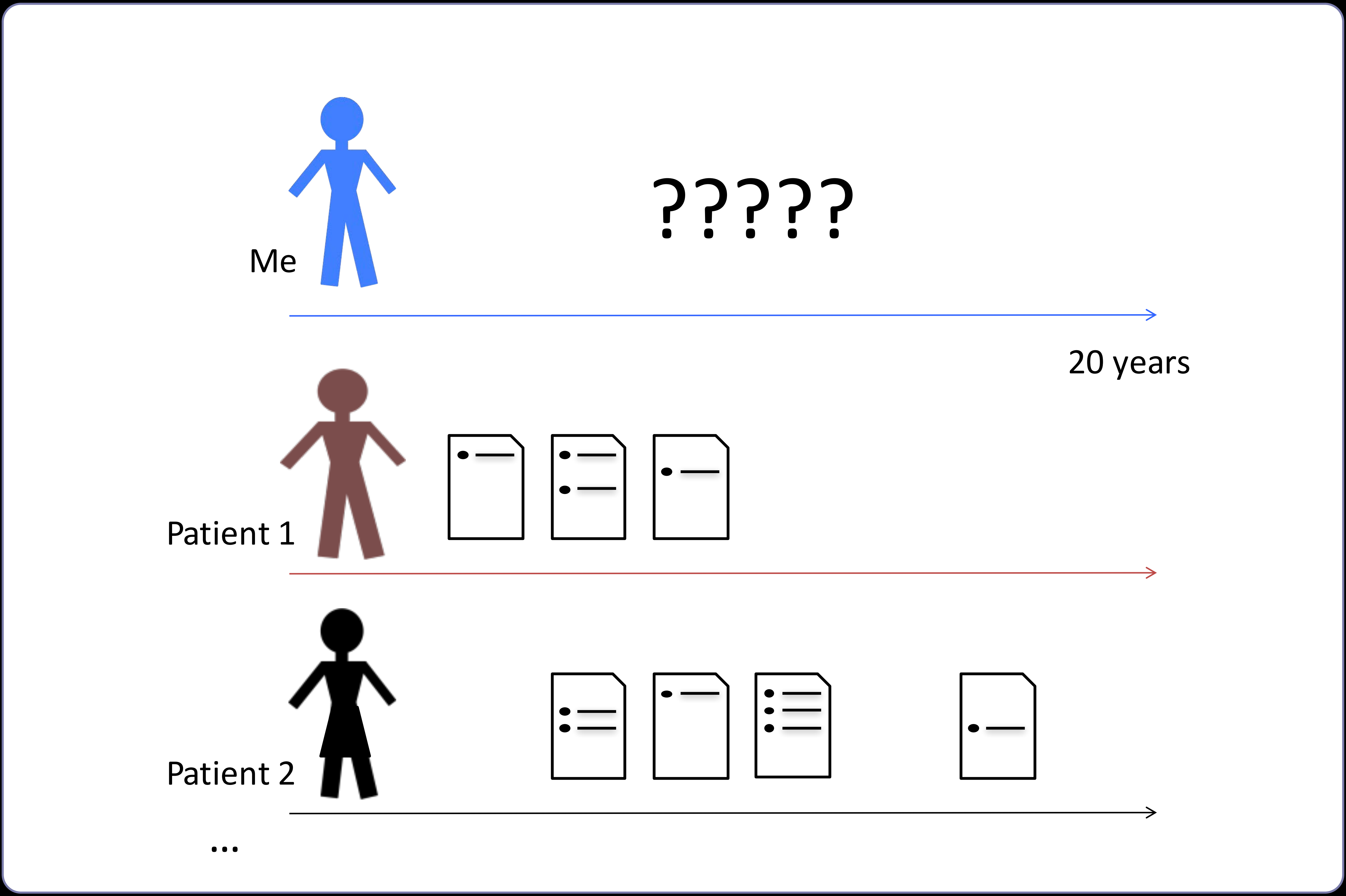


Temporal Modeling of Disease Progression

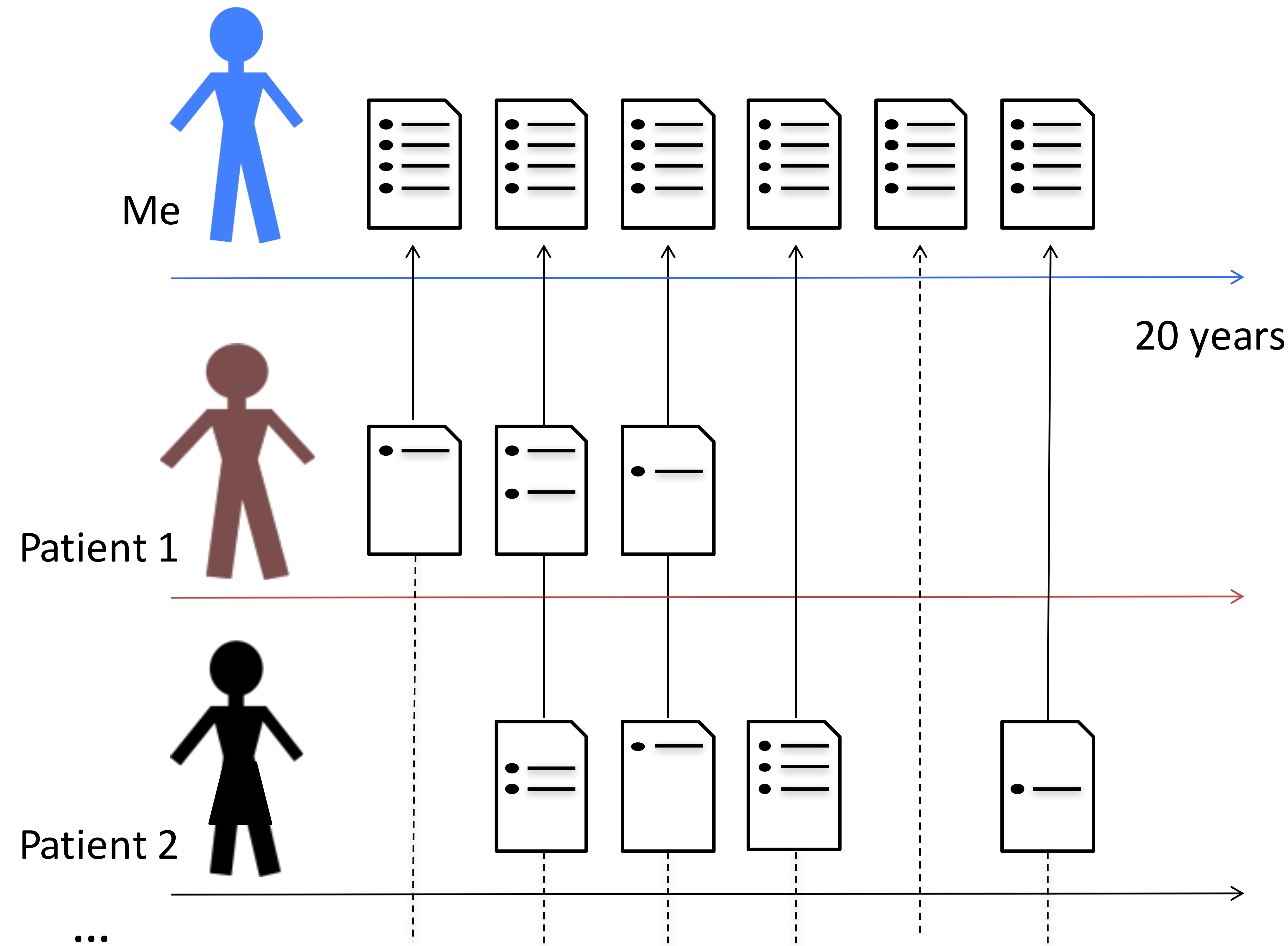
- Find markers of disease stage and progression, statistics of what to expect when
  - *What is the "typical trajectory" of a female diagnosed with Sjögren's syndrome at the age of 19?*

- Estimate a patient's future disease progression
  - *When will a specific individual with smoldering multiple myeloma (a rare blood cancer) transition to full-blown multiple myeloma?*
  - *Which second-line diabetes treatment should we give to a patient?*

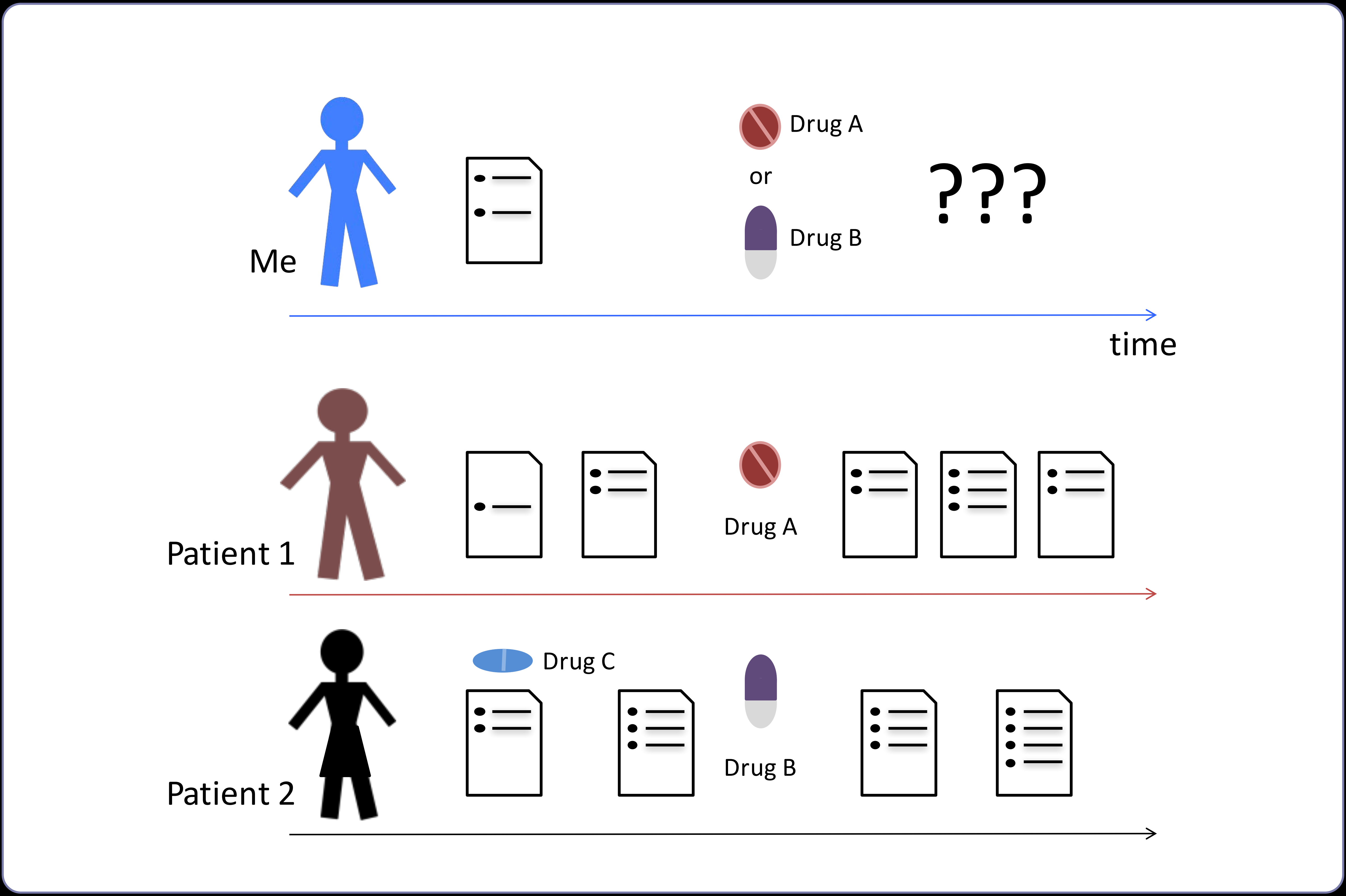
Personalized Medicine



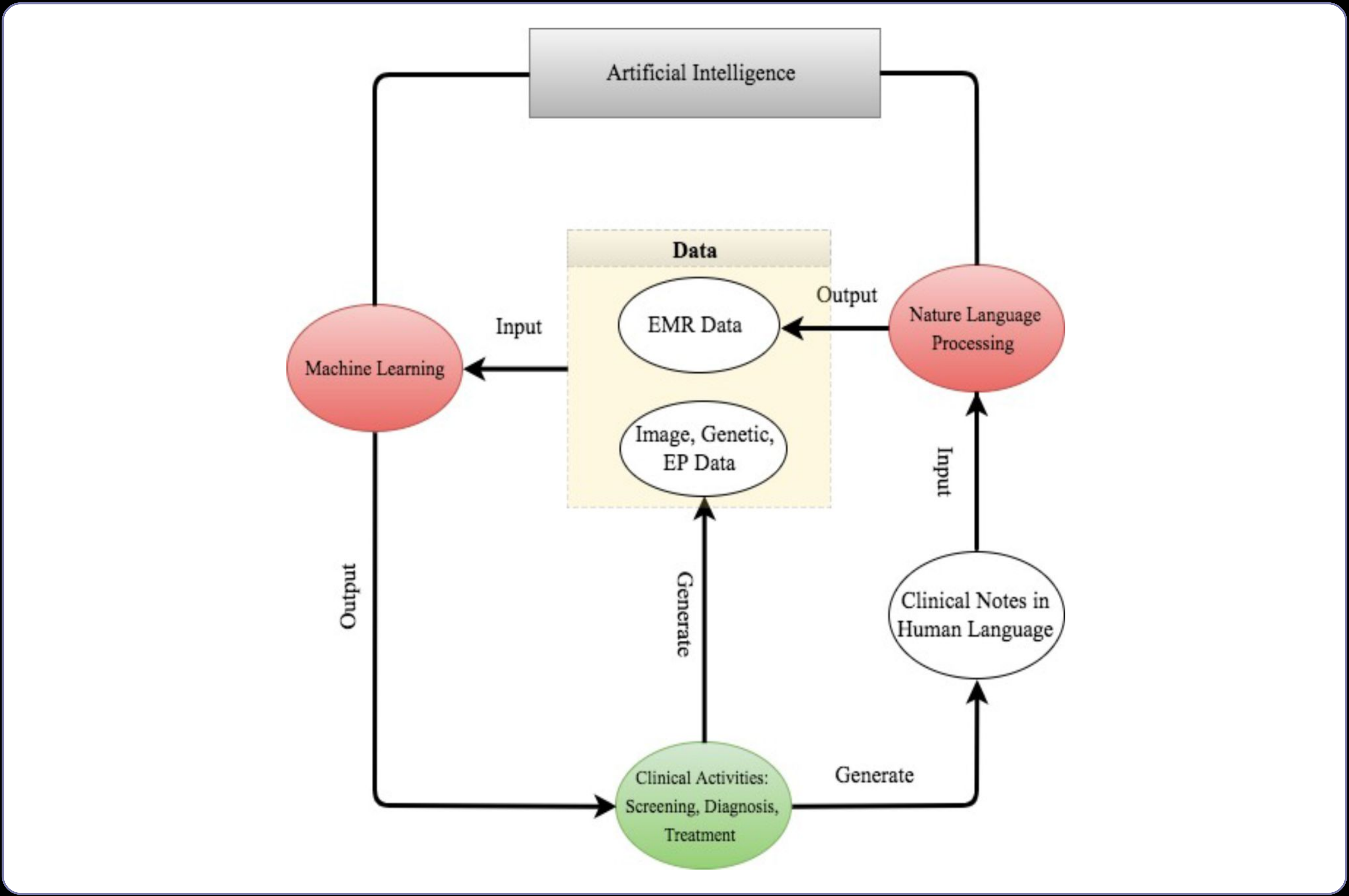
Prediction of Health Status



Personalized Prescriptions



From Data Generation to Decision Making



Many Challenges Unique to Medicine

- Life or death decisions
  - Need **robust** algorithms
  - Checks and balances built into ML deployment
  - (Also arises in other applications of AI such as autonomous driving)
  - Need **fair** and **accountable** algorithms
- Many questions are about unsupervised learning
  - Discovering disease subtypes, or answering question such as “characterize the types of people that are highly likely to be readmitted to the hospital”?
- Many of the questions we want to answer are *causal*
  - Naïve use of supervised machine learning is insufficient

Problems with “Data”

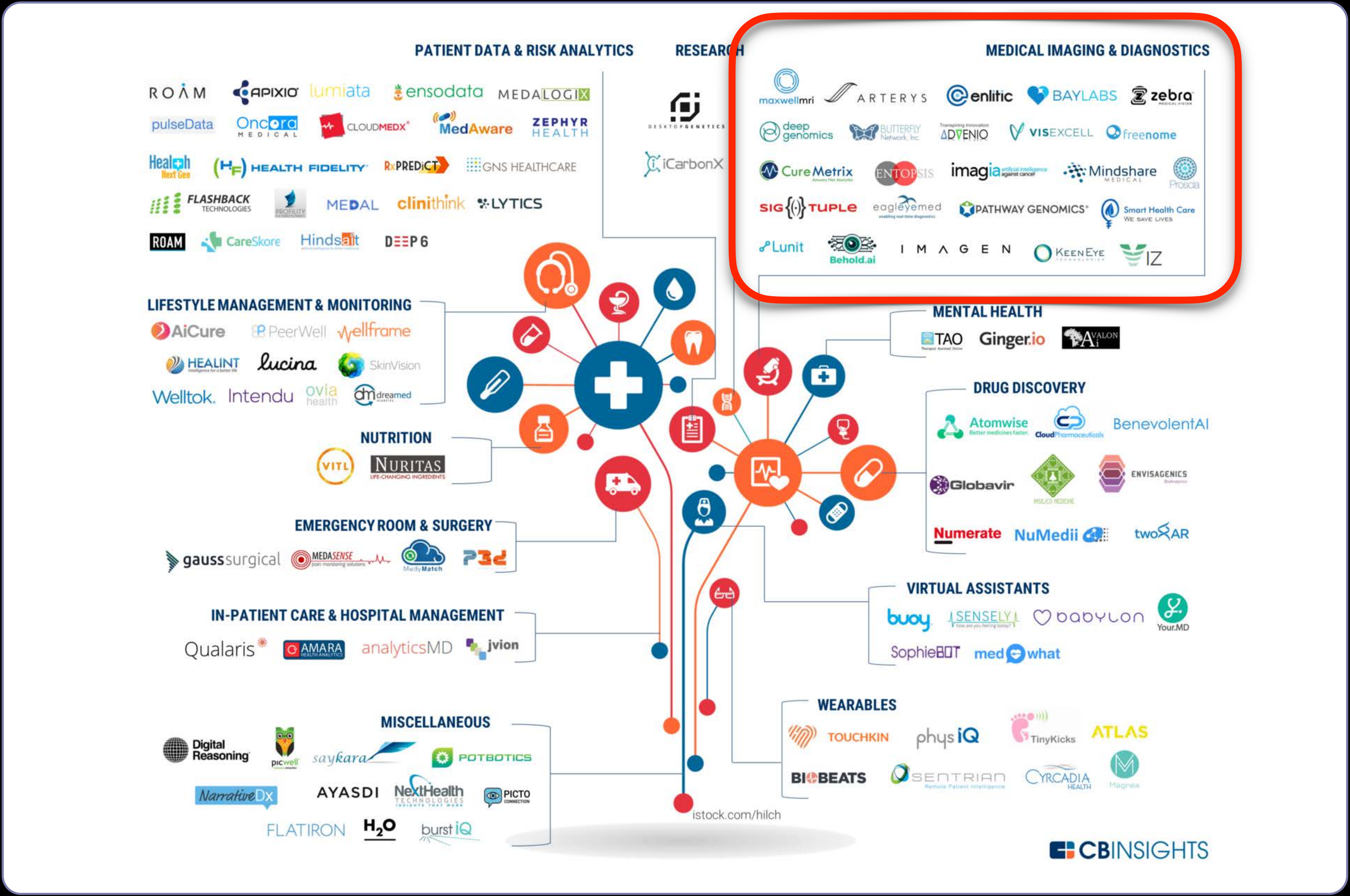
- Often very little labeled data (e.g., for clinical NLP)
  - Motivates semi-supervised learning algorithms
- Sometimes small numbers of samples (e.g., a rare disease)
  - Learn as much as possible from other data (e.g. healthy patients)
  - Model the problem carefully
- Lots of missing data, varying time intervals, censored labels

Problems with Clinical Integration

- Difficulty of de-identifying data
  - Need for data sharing agreements and sensitivity
- Difficulty of deploying ML
  - Commercial electronic health record software is difficult to modify
  - Data is often in silos; everyone recognizes need for interoperability, but slow progress
  - Careful testing and iteration is needed

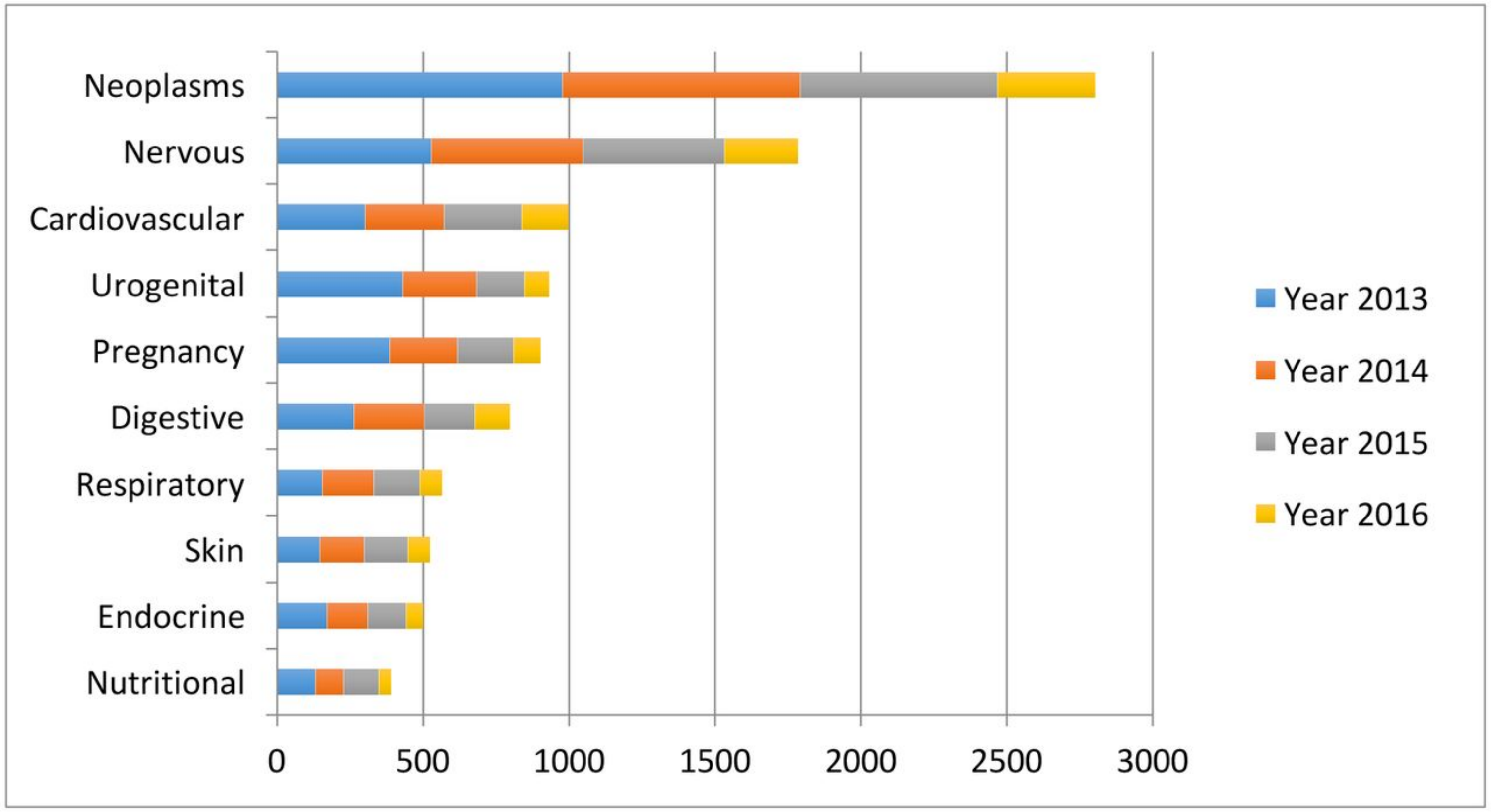
PART V: Utility of ML in Medical Imaging

Medical Uses of Deep Learning

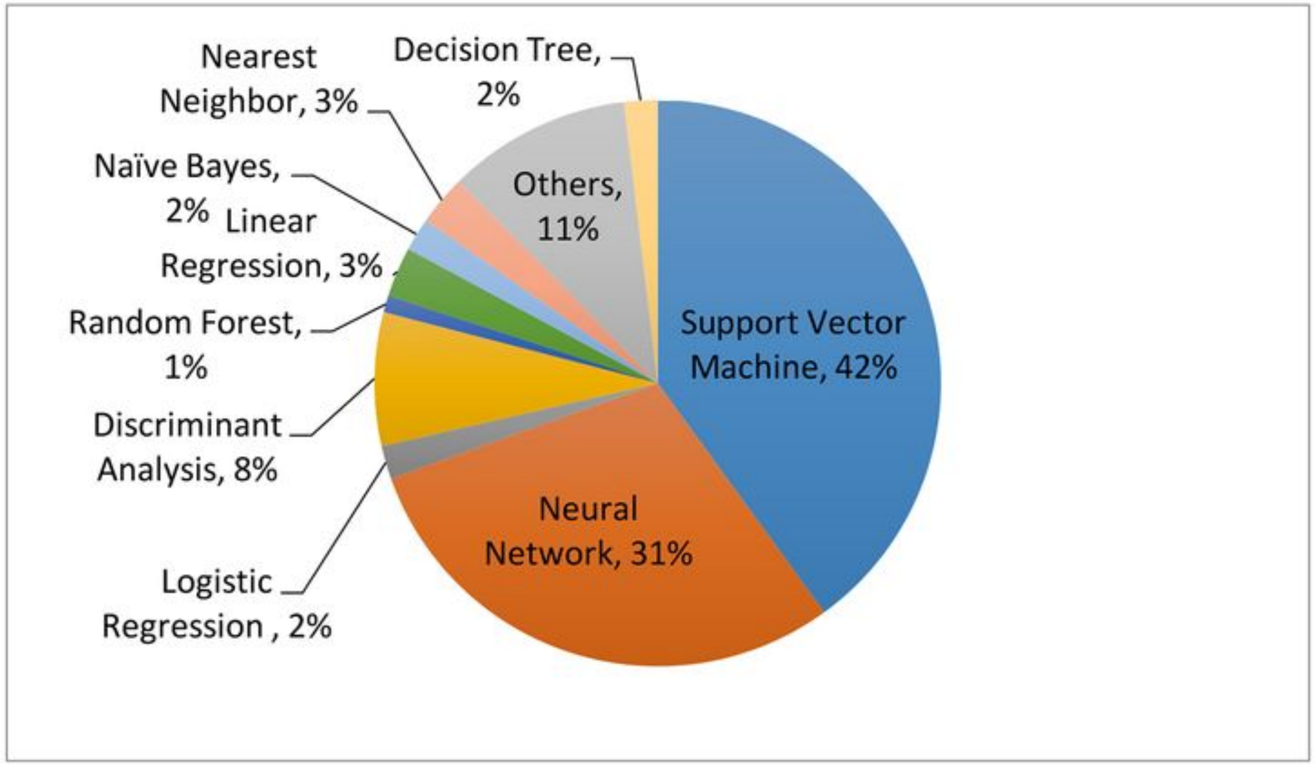


Machine Learning for Diagnosis

Top-10 Diseases

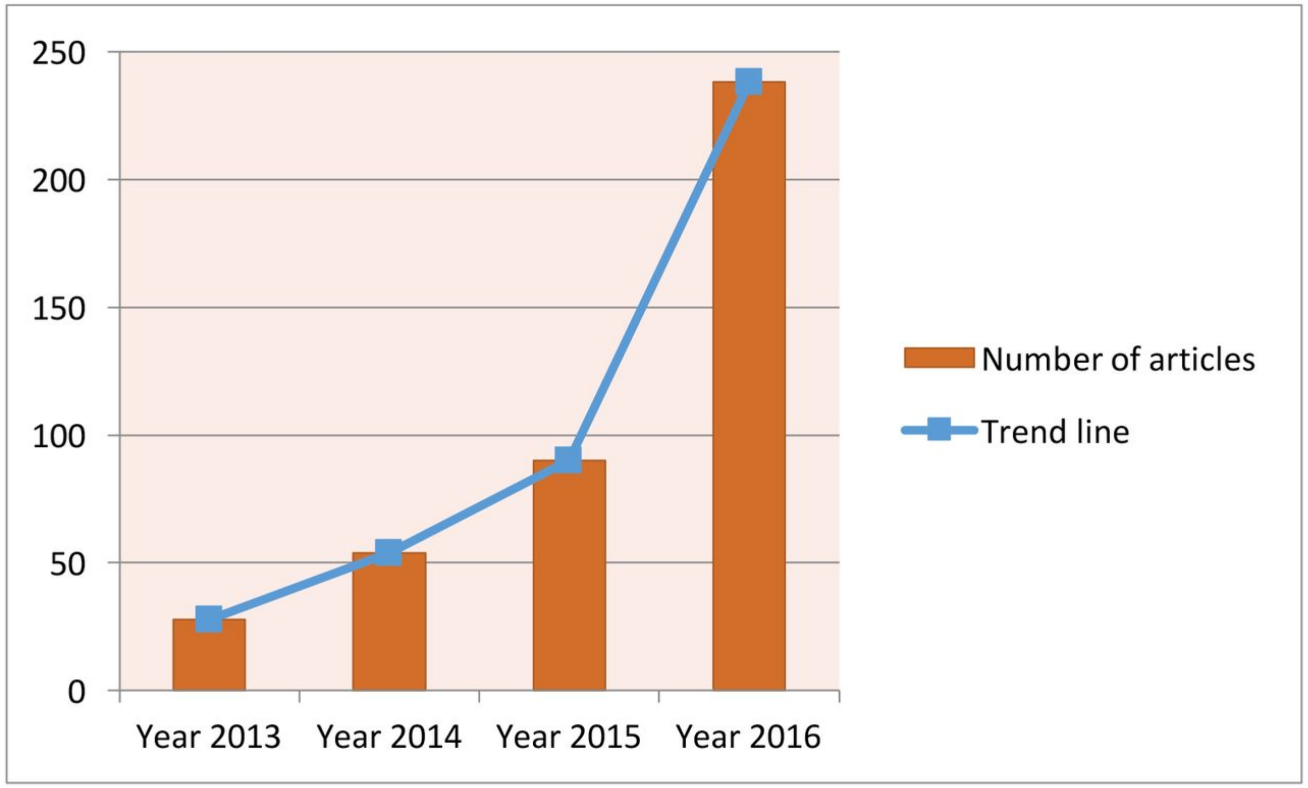


Top-10 Techniques

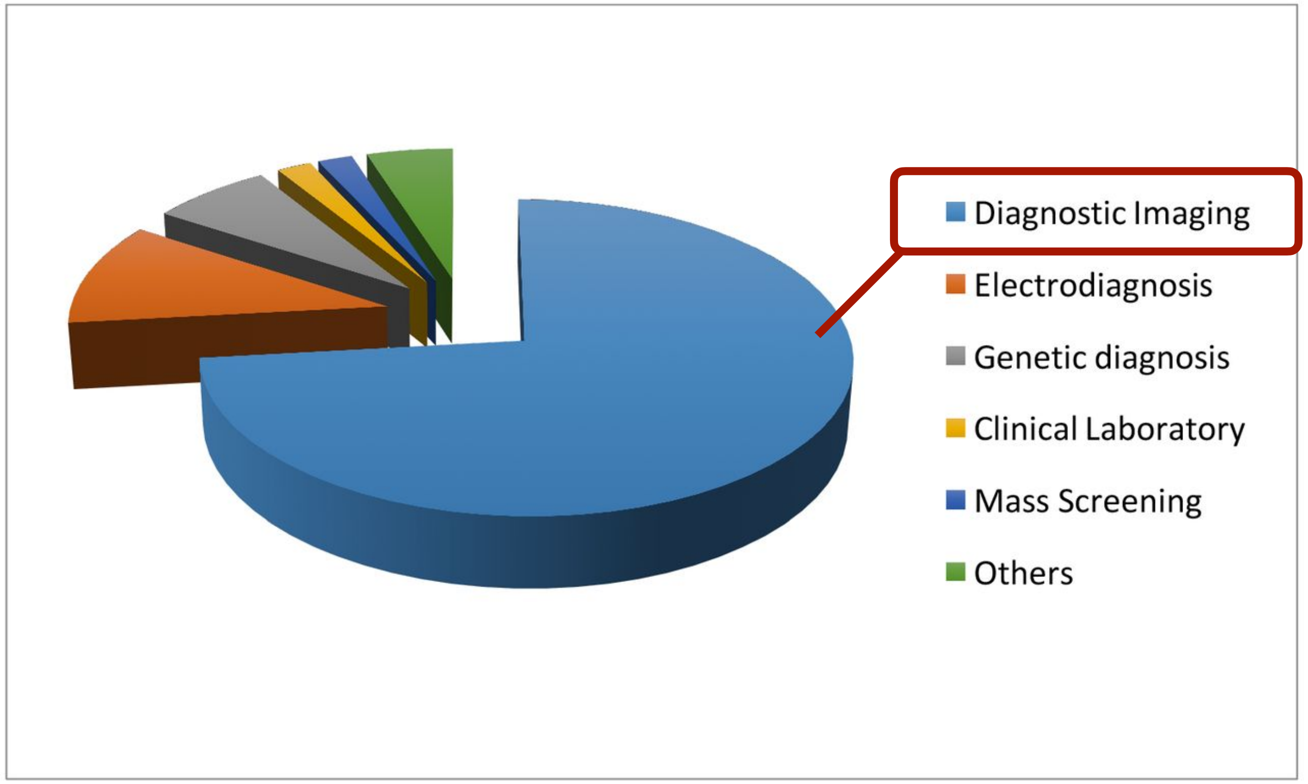


Deep Learning on the Rise

Number of DL Studies

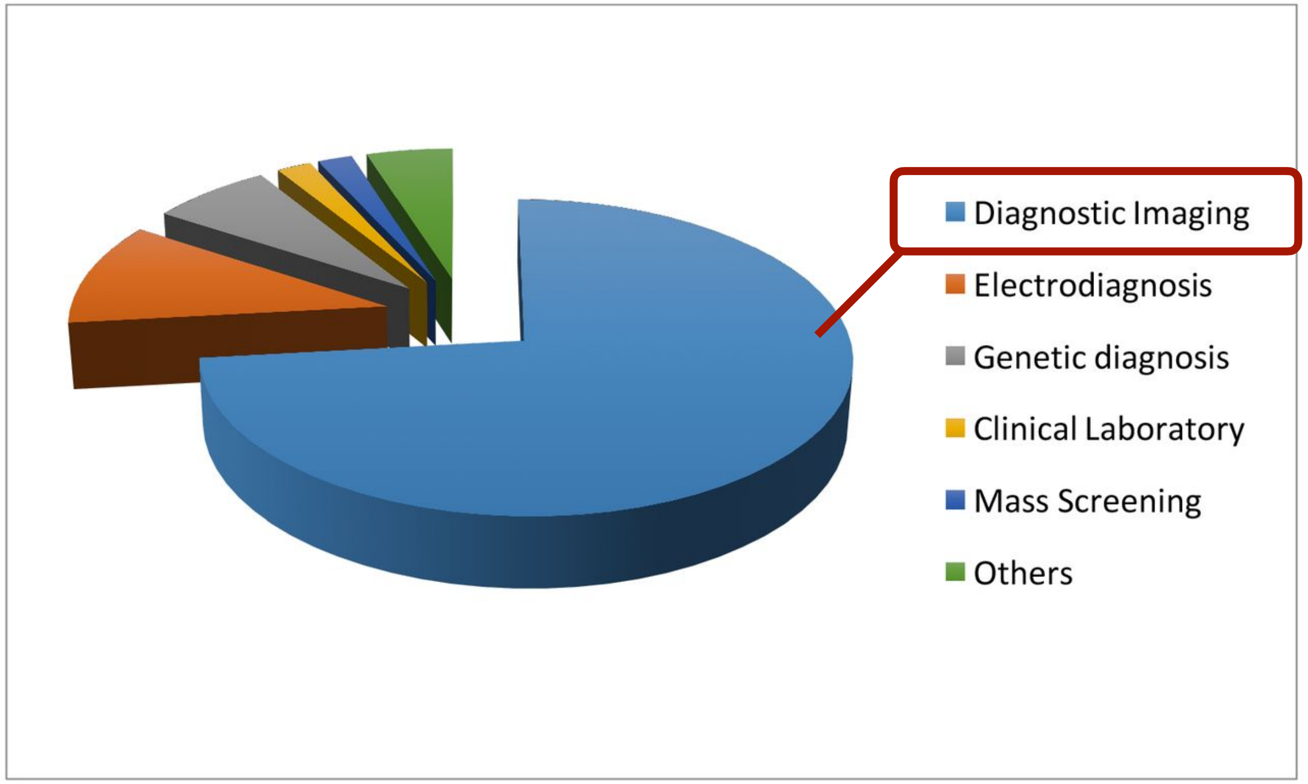


DL Studies Based on Data Type



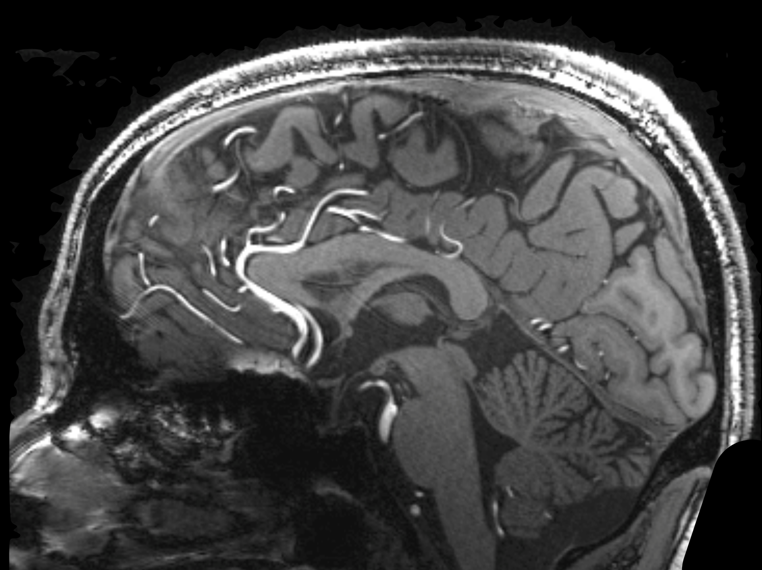
Deep Learning for Medical Imaging

DL Studies Based on Data Type

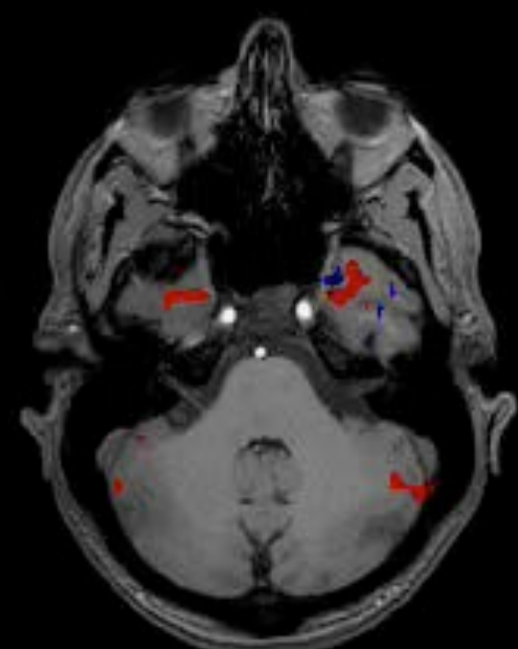


- Medical images are high-dimensional (volumetric and temporal)
- Medical images are mostly interpreted by radiologists (manual labor)
- Humans are quite poor in seeing fine-grained patterns in static images
- Similarities medical–natural images (closely tied to computer vision)

Imaging Morphology and Function



Anatomical

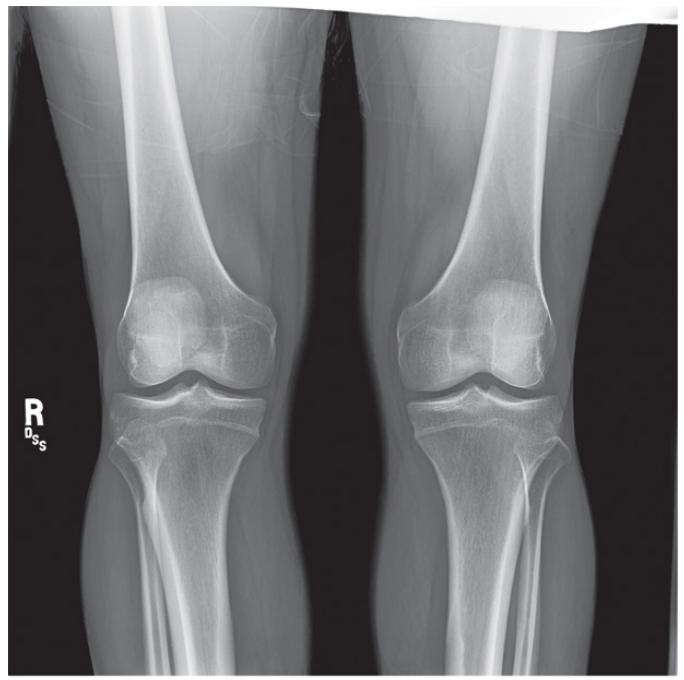


Functional

Diffusion

Modern Imaging Modalities

X-ray

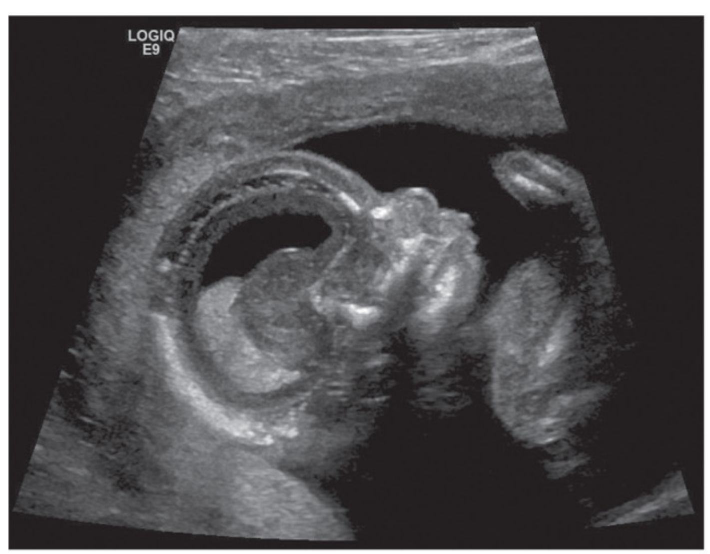


(a)

(b)

Nuclear Medicine

Ultrasound



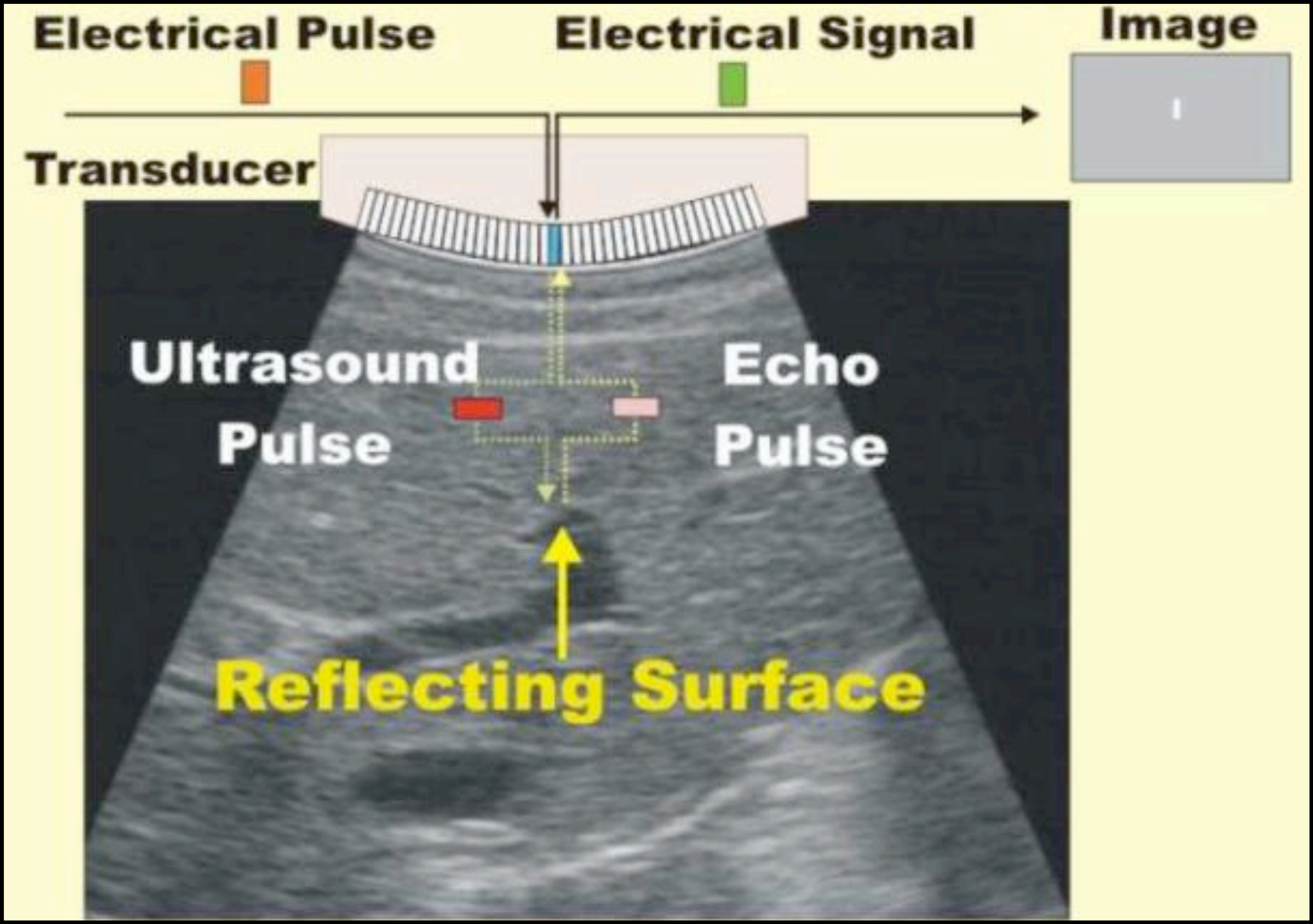
(c)

(d)

MRI

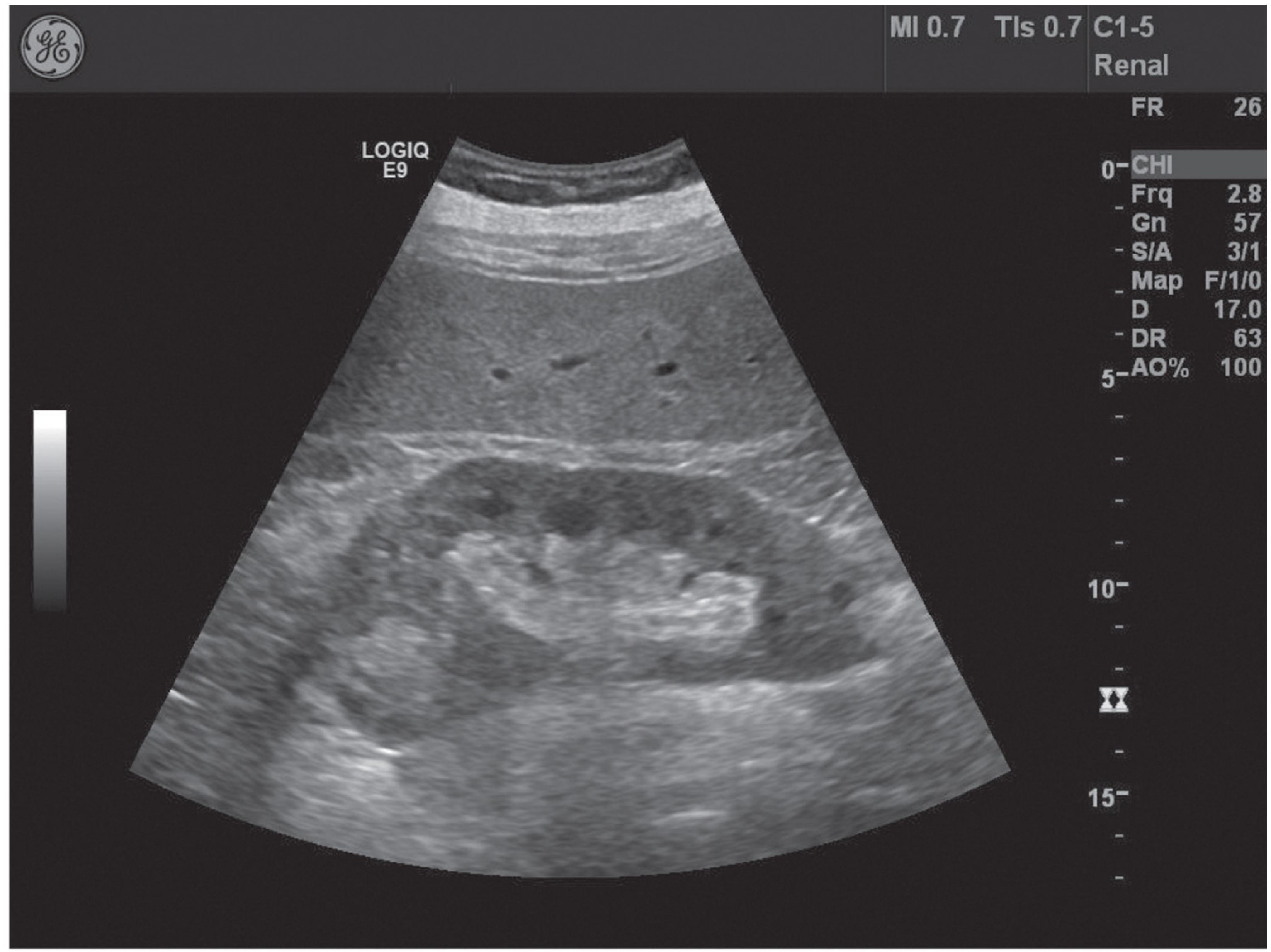
Ultrasound

- Uses sound pressure waves
- We cannot hear these sounds: > 20 kHz
- Typically 2-18 MHz



Ultrasound

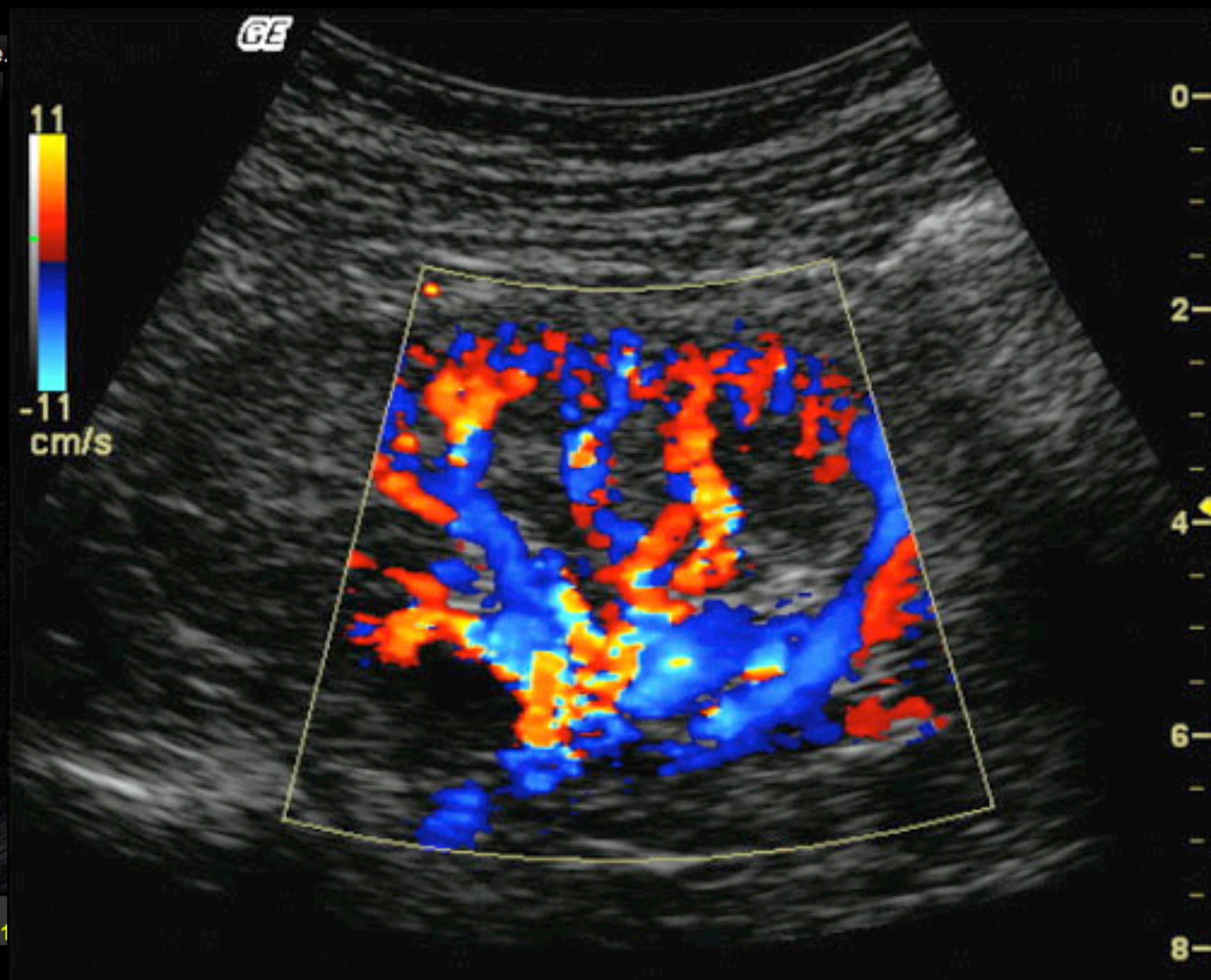
(a)



(b)

Ultrasound

- Higher resolution with increasing frequency
- Cannot image too deep if the frequency is high
- Lots of application areas: cardiology, urology, obstetrics,…

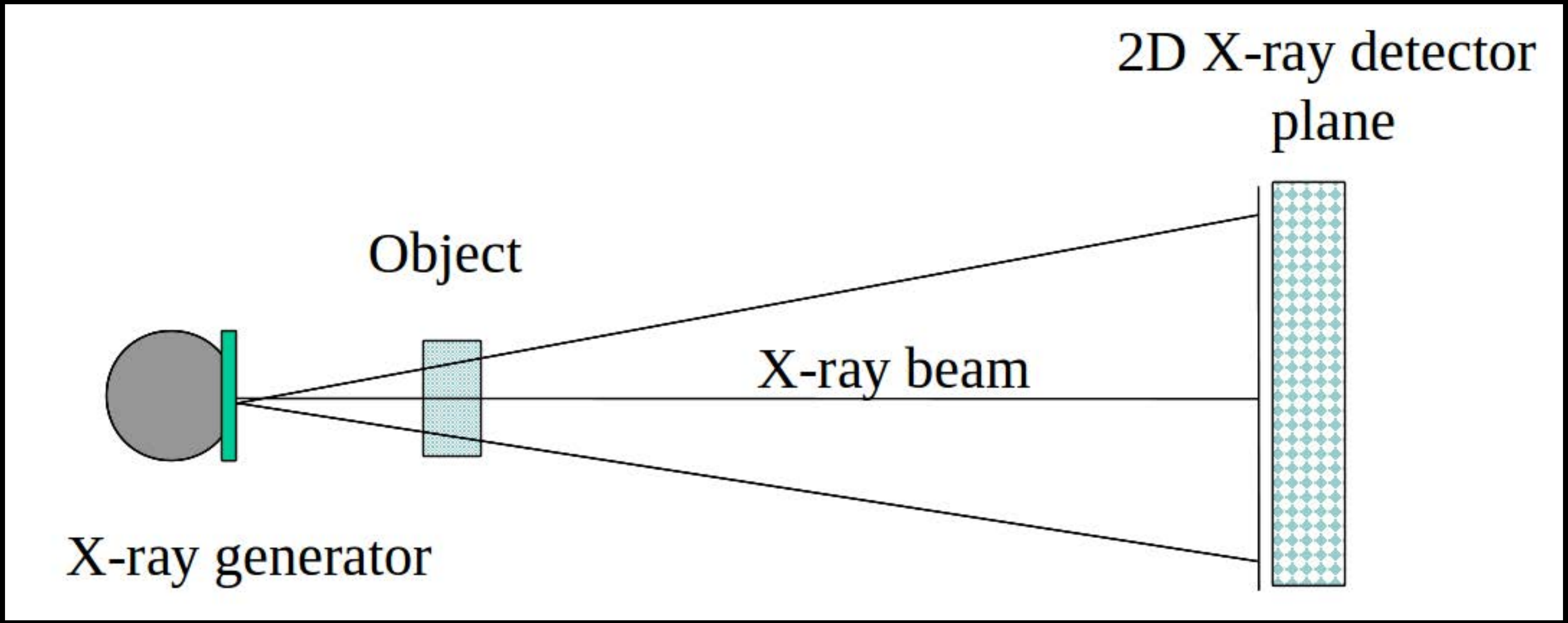


Doppler Ultrasound

X-ray

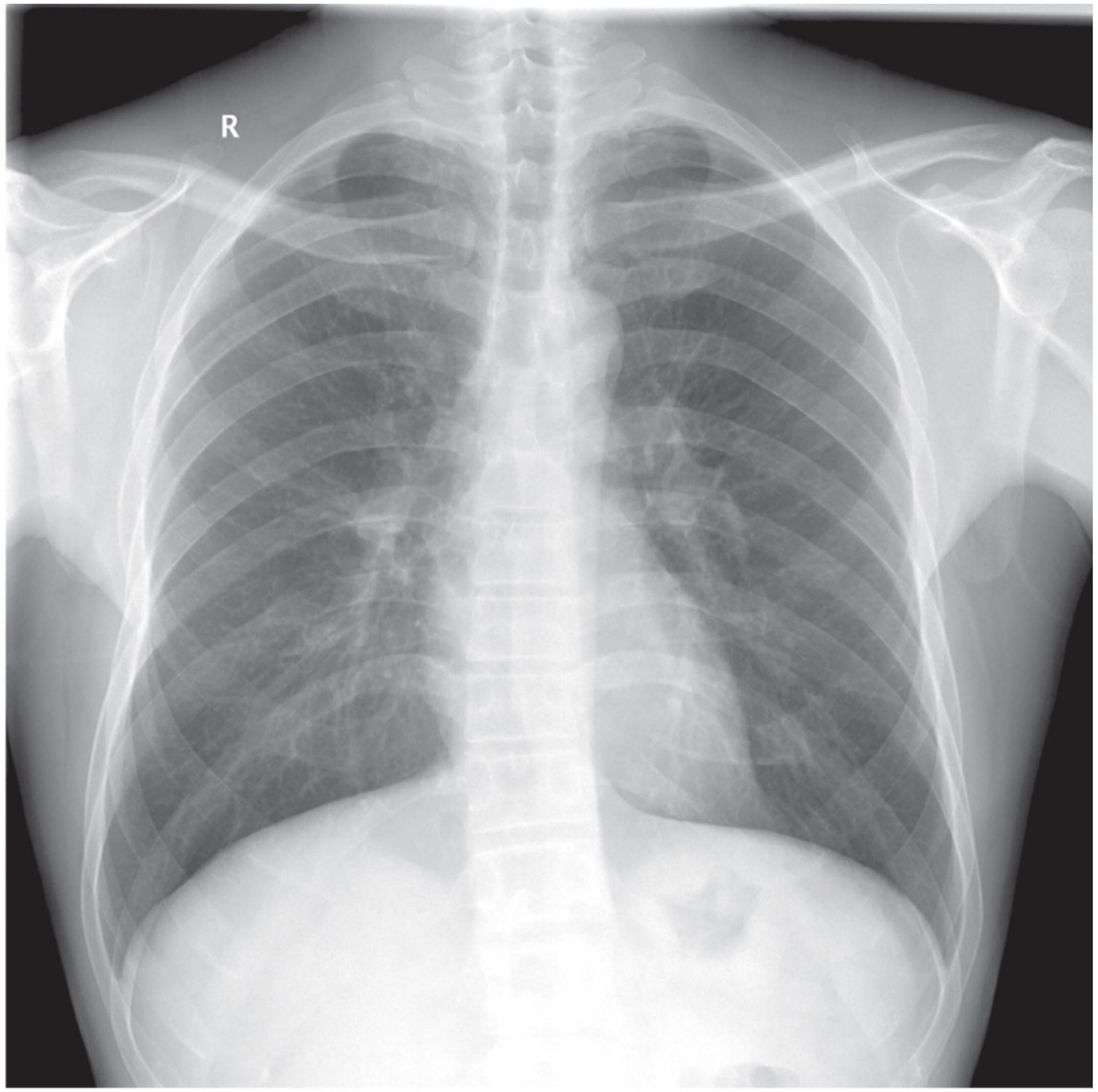
- Uses X-ray photons
- Photons at VERY high frequency: ~10^{16}-10^{19} Hz
- Different tissues attenuate photons differently → contrast
- Very good at detecting bone structure

- Projection images



Chest X-ray

(a)

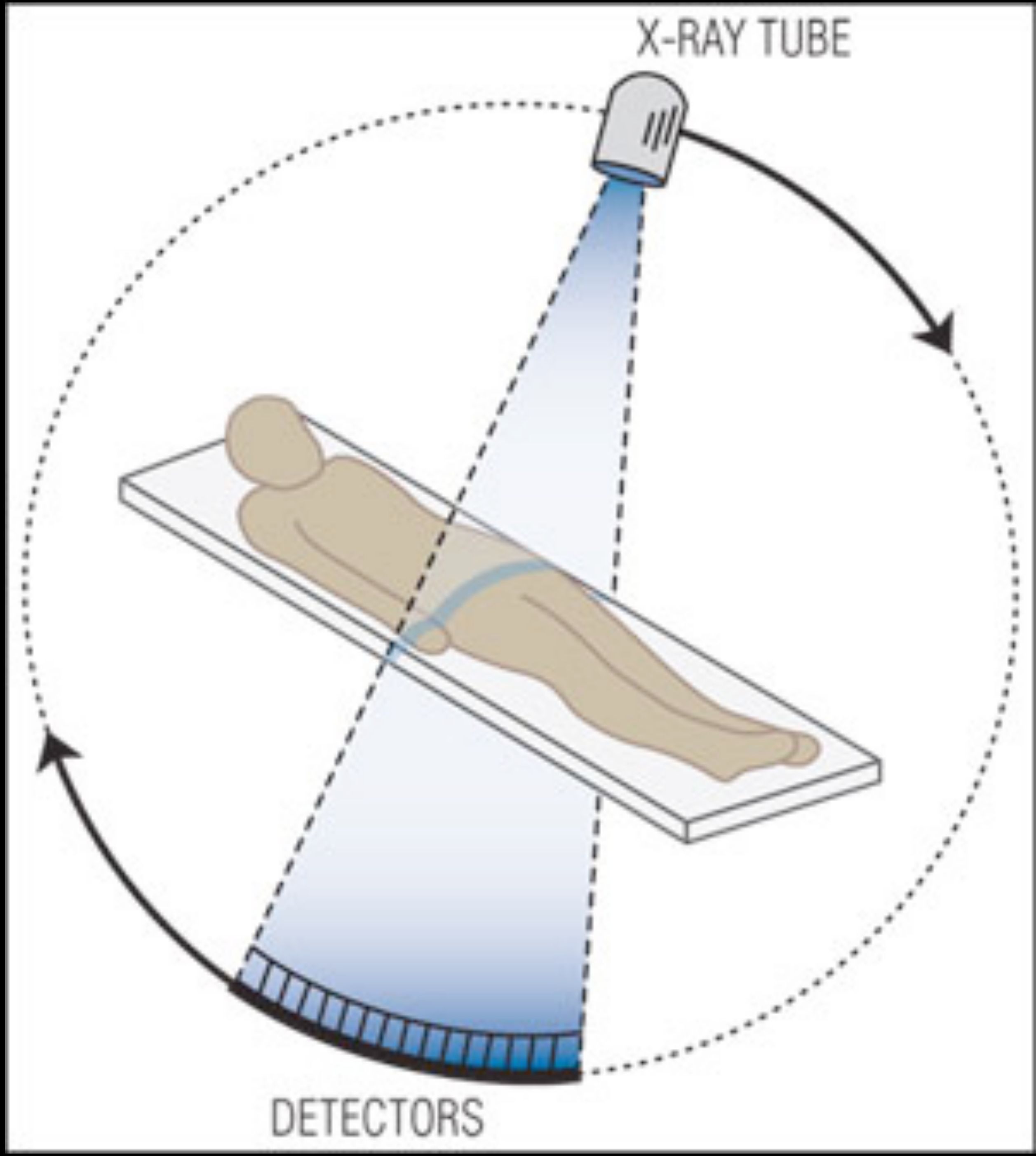


(b)

Projection Through the Body

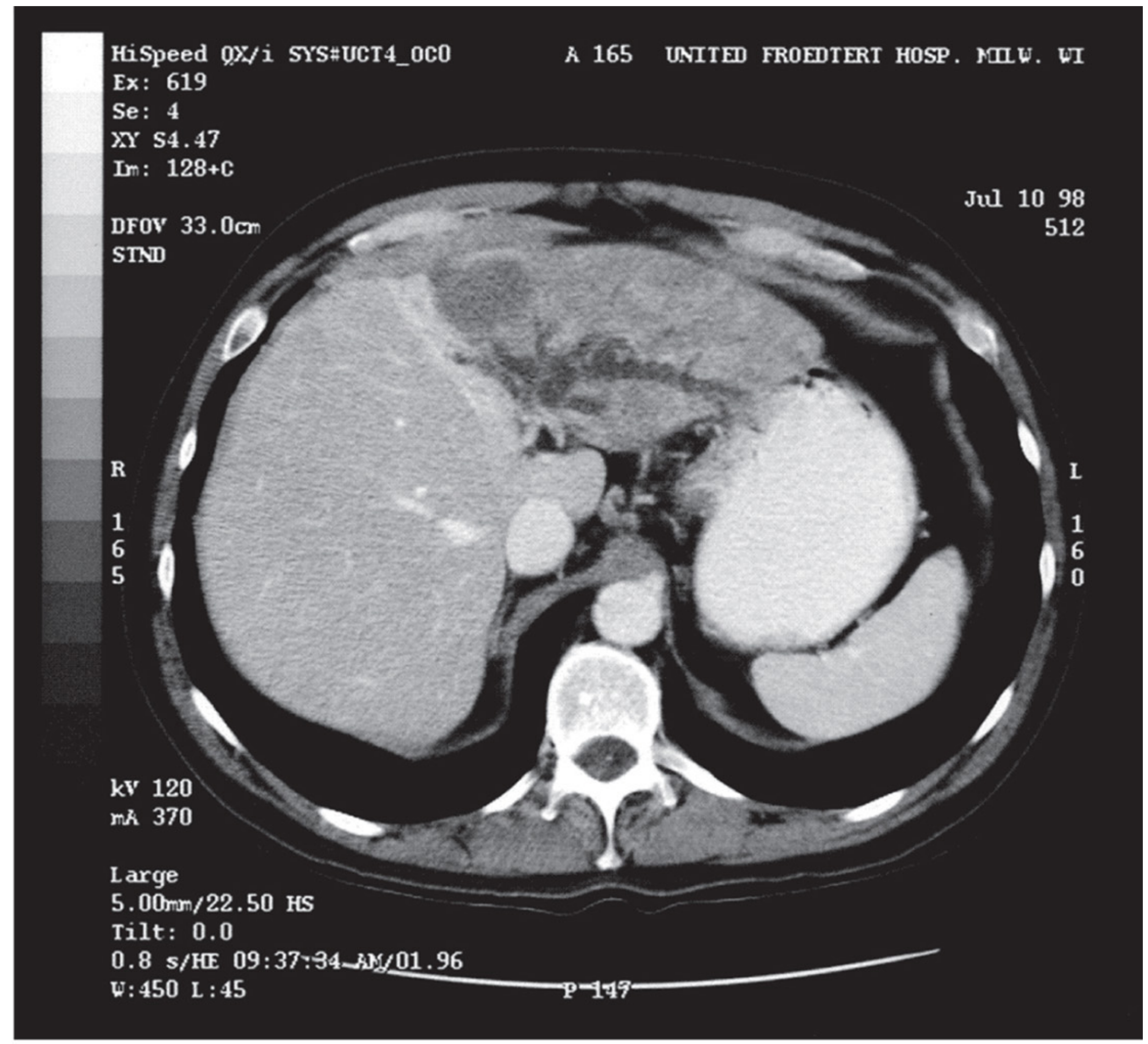
Computerized Tomography

- Extends X-ray imaging to 3D format
- Rotate X-ray source and detectors all together



Computerized Tomography

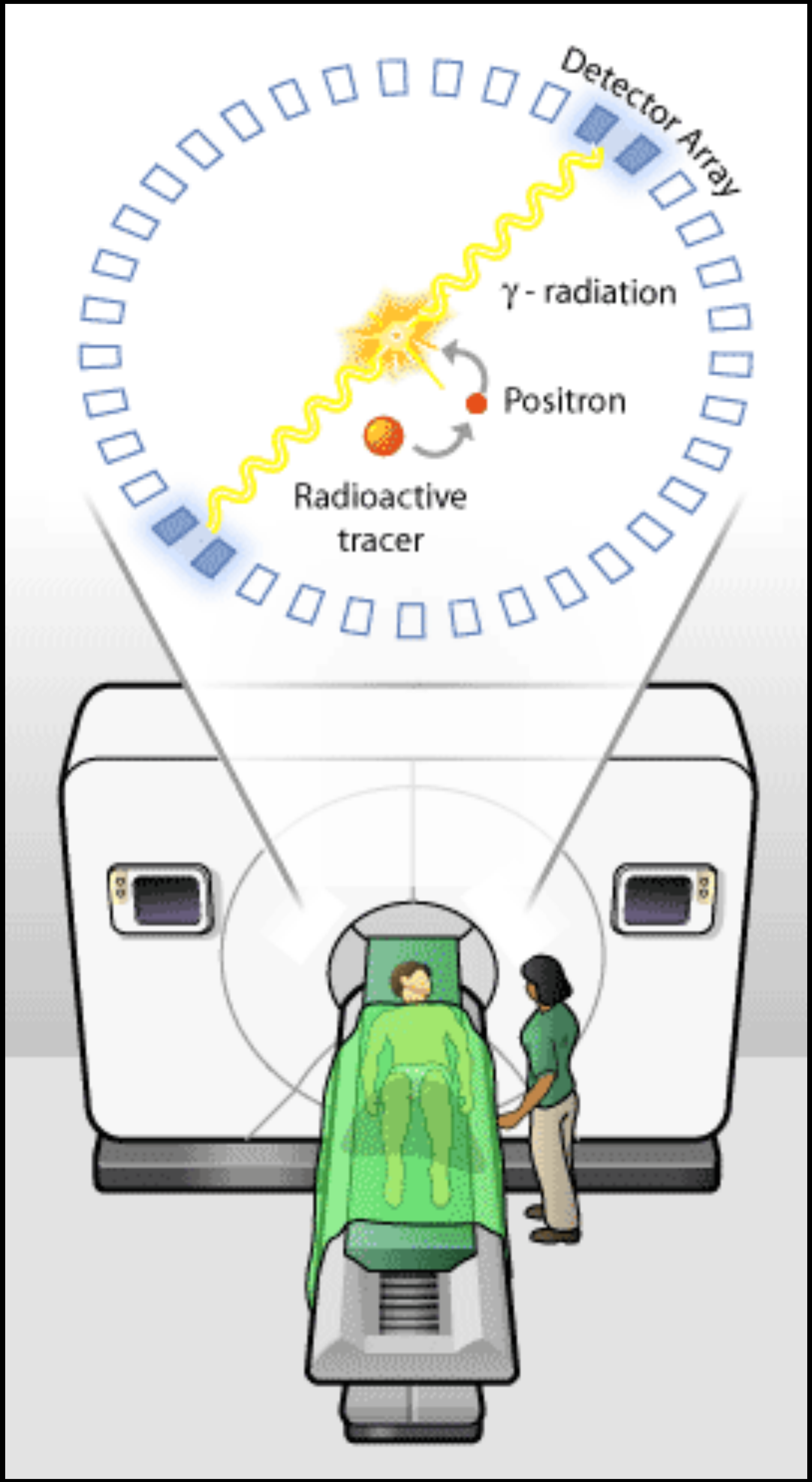
(a)



(b)

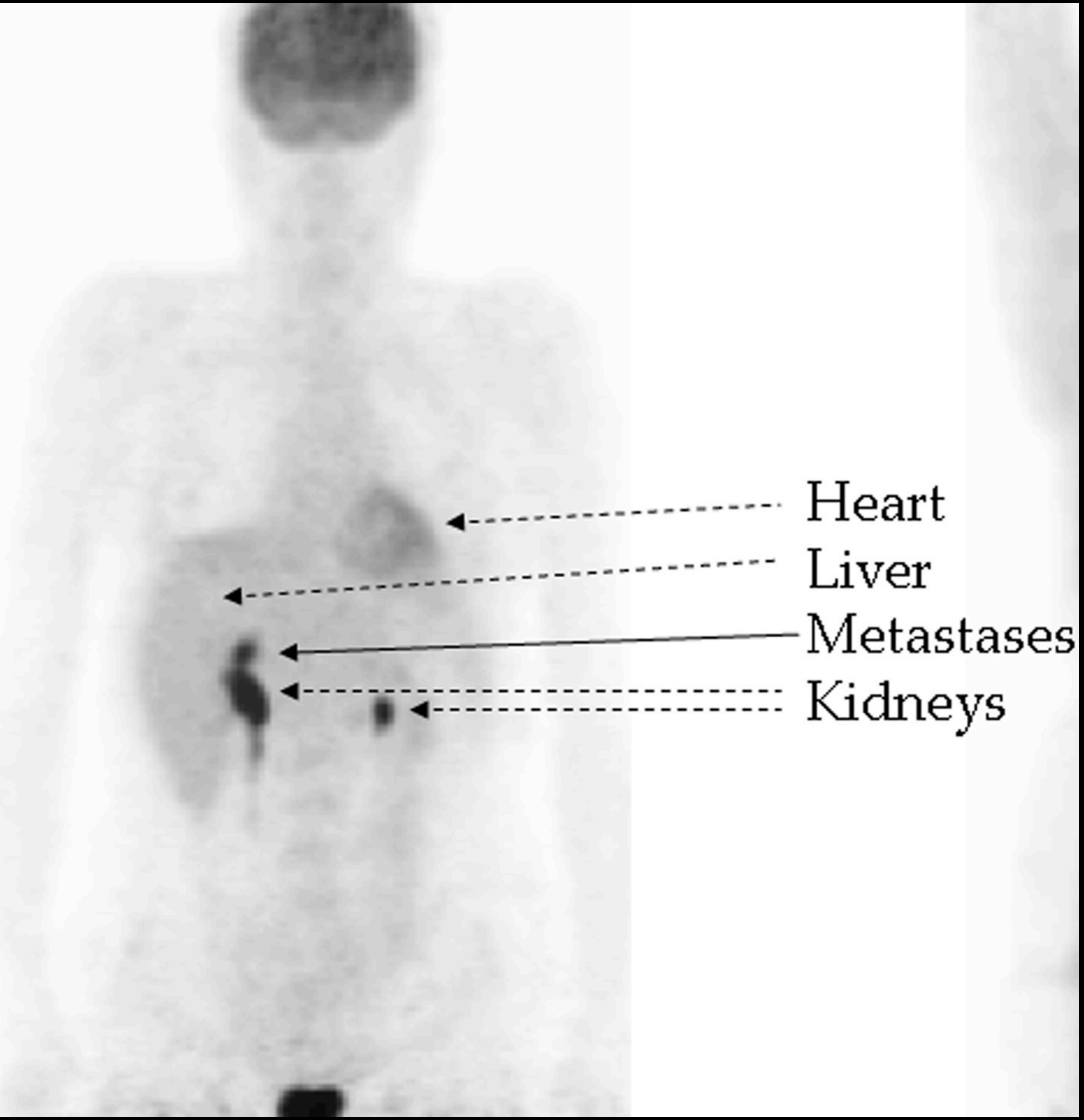
Nuclear Medicine

- A radionuclide is injected into the blood stream
- Goes to cancer locations before cancer cells are hyperactive
- Emitted Gamma rays are detected for imaging



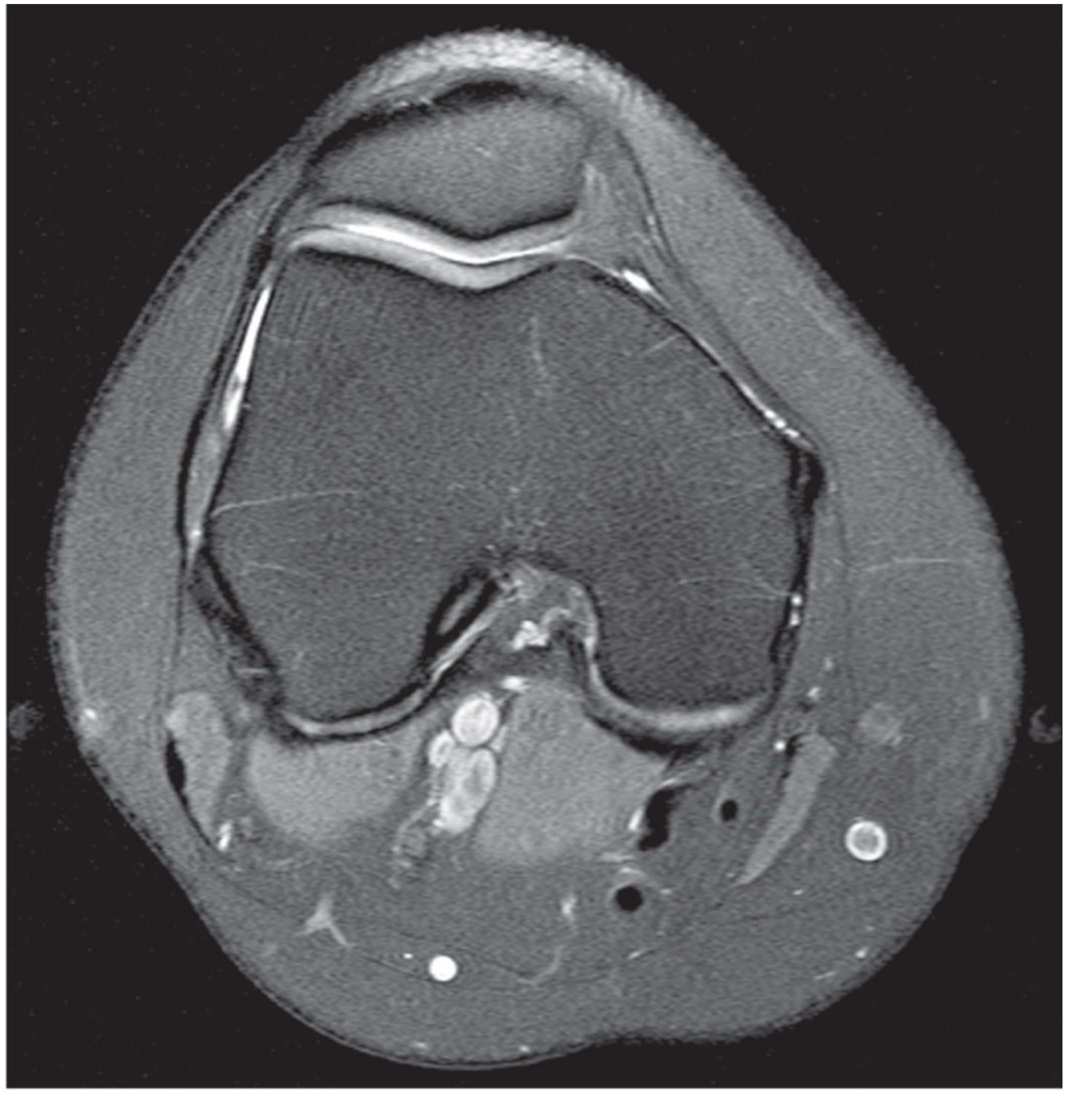
Positron Emission Tomography

- Resolution is not good, but very sensitive to cancer



MRI

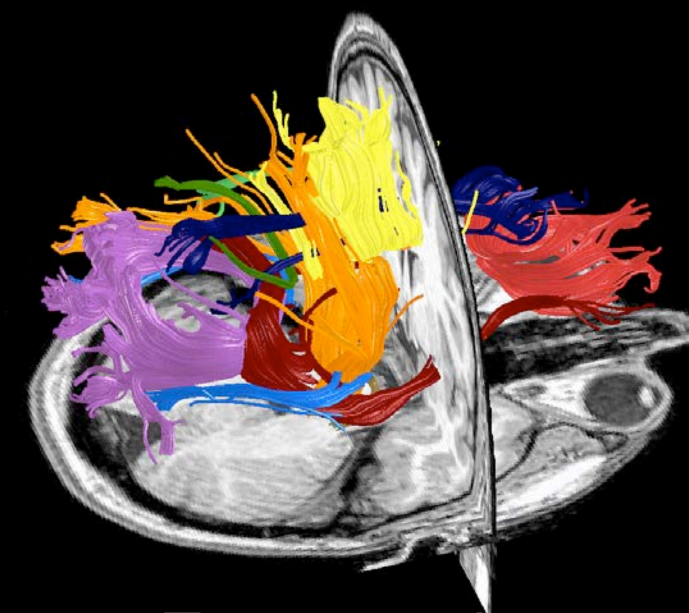
(a)



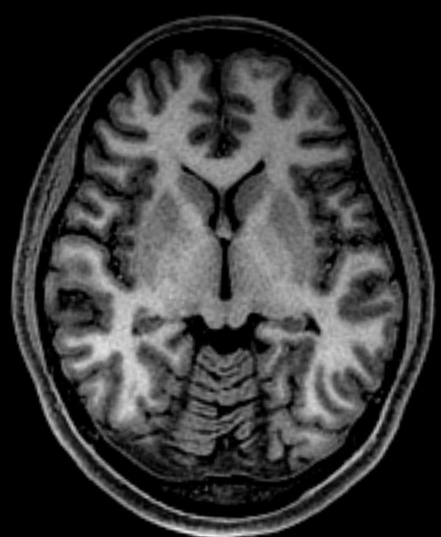
(b)

Magnetic Resonance Imaging (MRI)

- FLEXIBLE CONTRAST
- Arbitrary geometries
- Non-invasive, non-ionizing imaging
- Relatively slow imaging

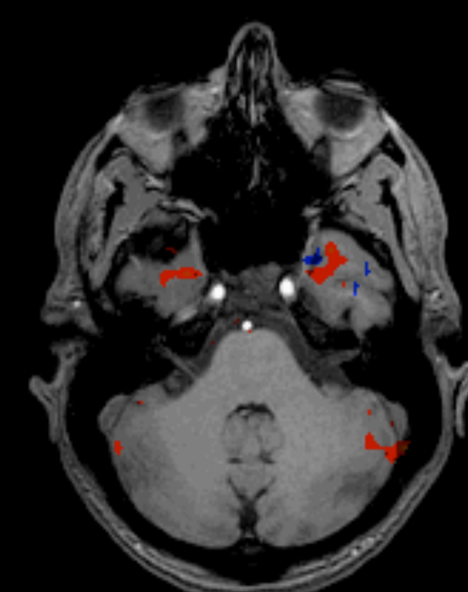


Tractography

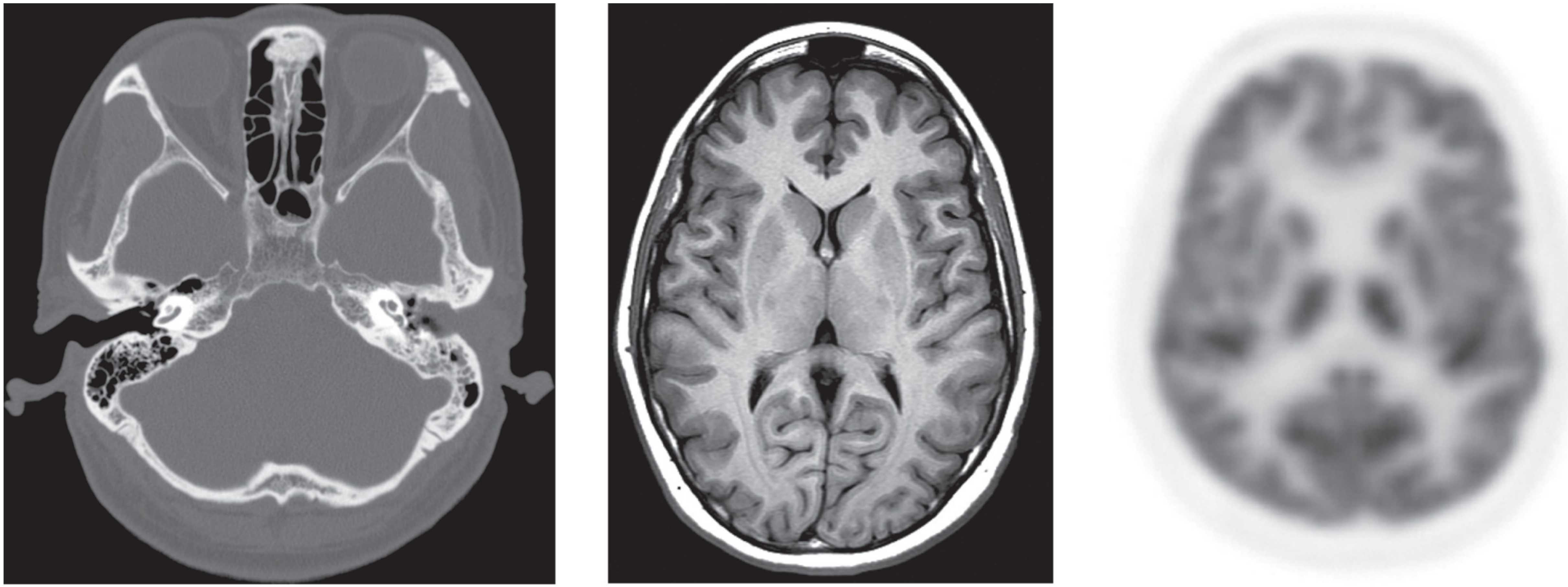


Anatomy

Angiography



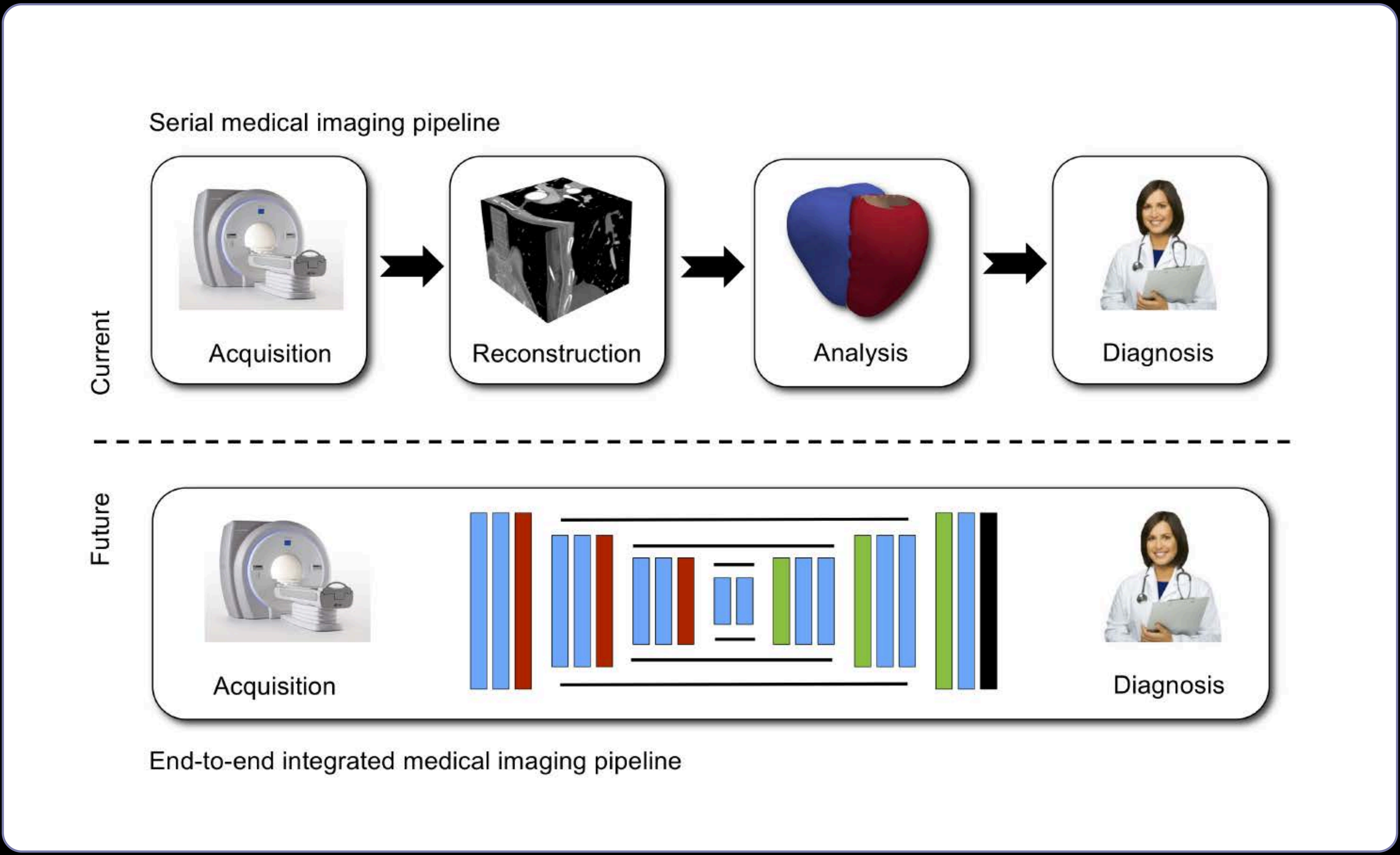
functional MRI



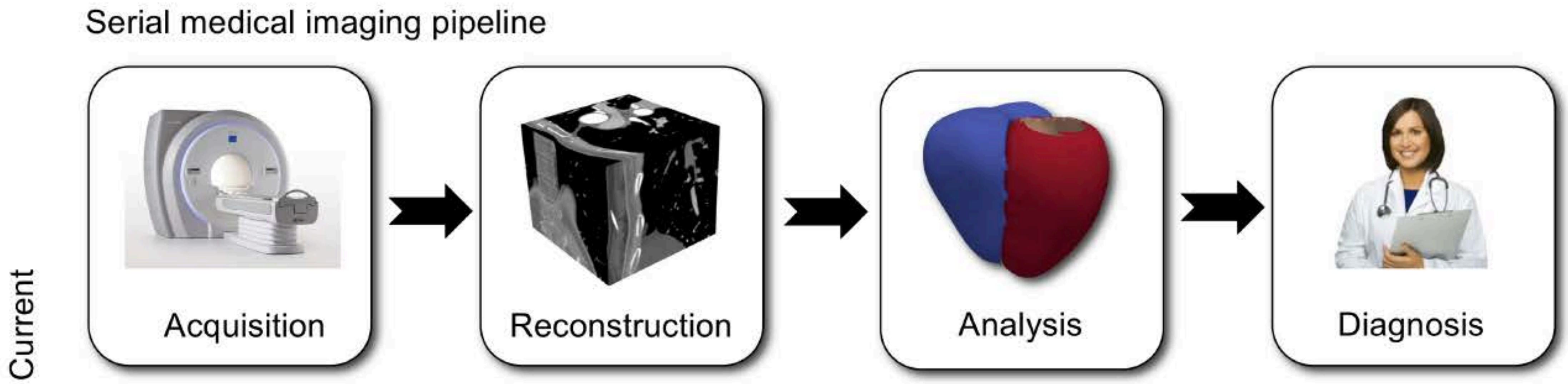
(a) (b) (c)

CT MRI PET

Medical Imaging Pipeline



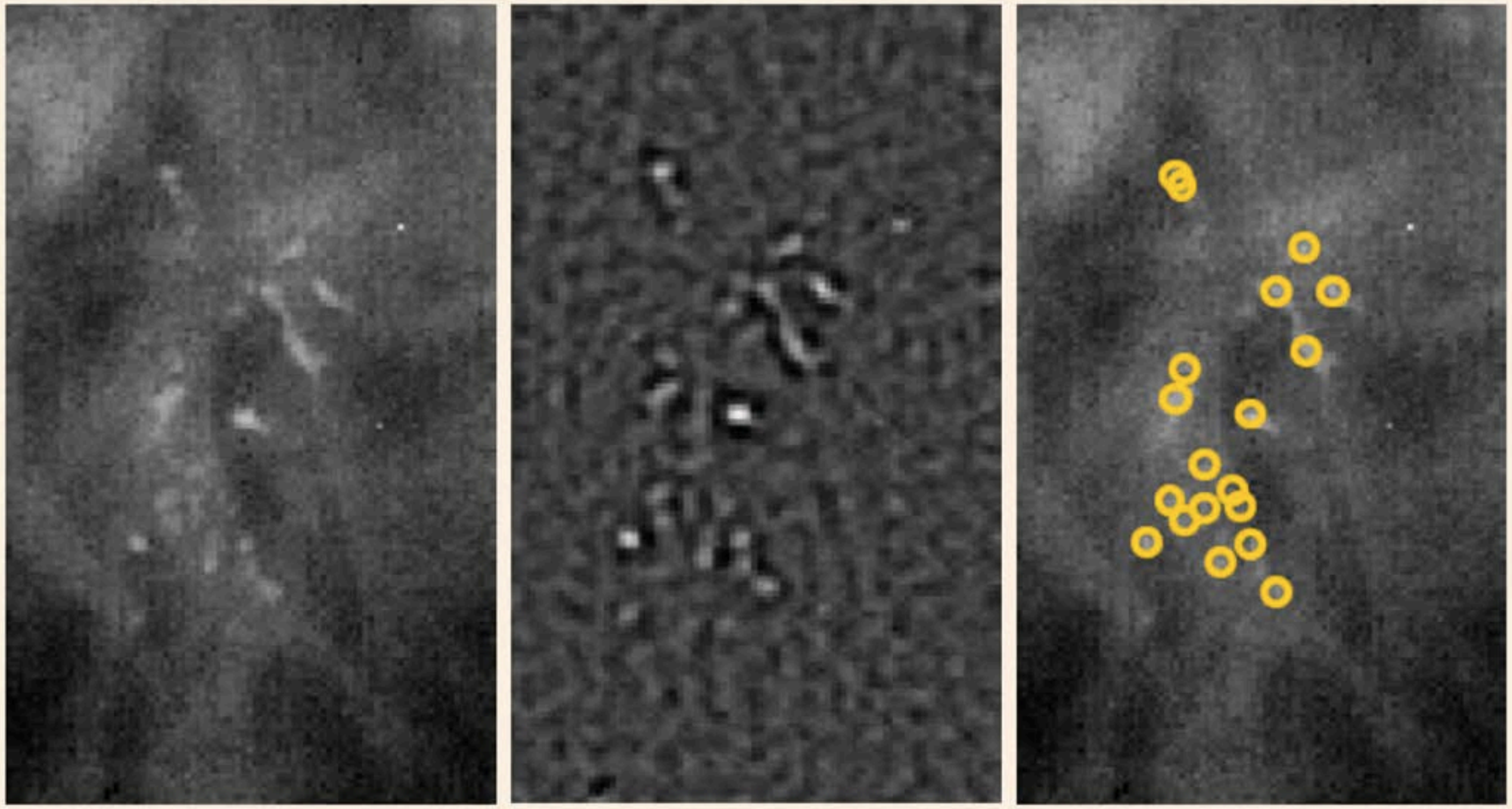
Motivation



- Radiologists need to interpret an excessively large number of images
- Their capacity to correctly interpret images is overwhelmed
- Automated image analysis systems are needed for error reduction
- Machine learning underpins the algorithms for such systems

PART V: Example Applications in Medical Imaging

Examples: Detecting Micro-calcifications

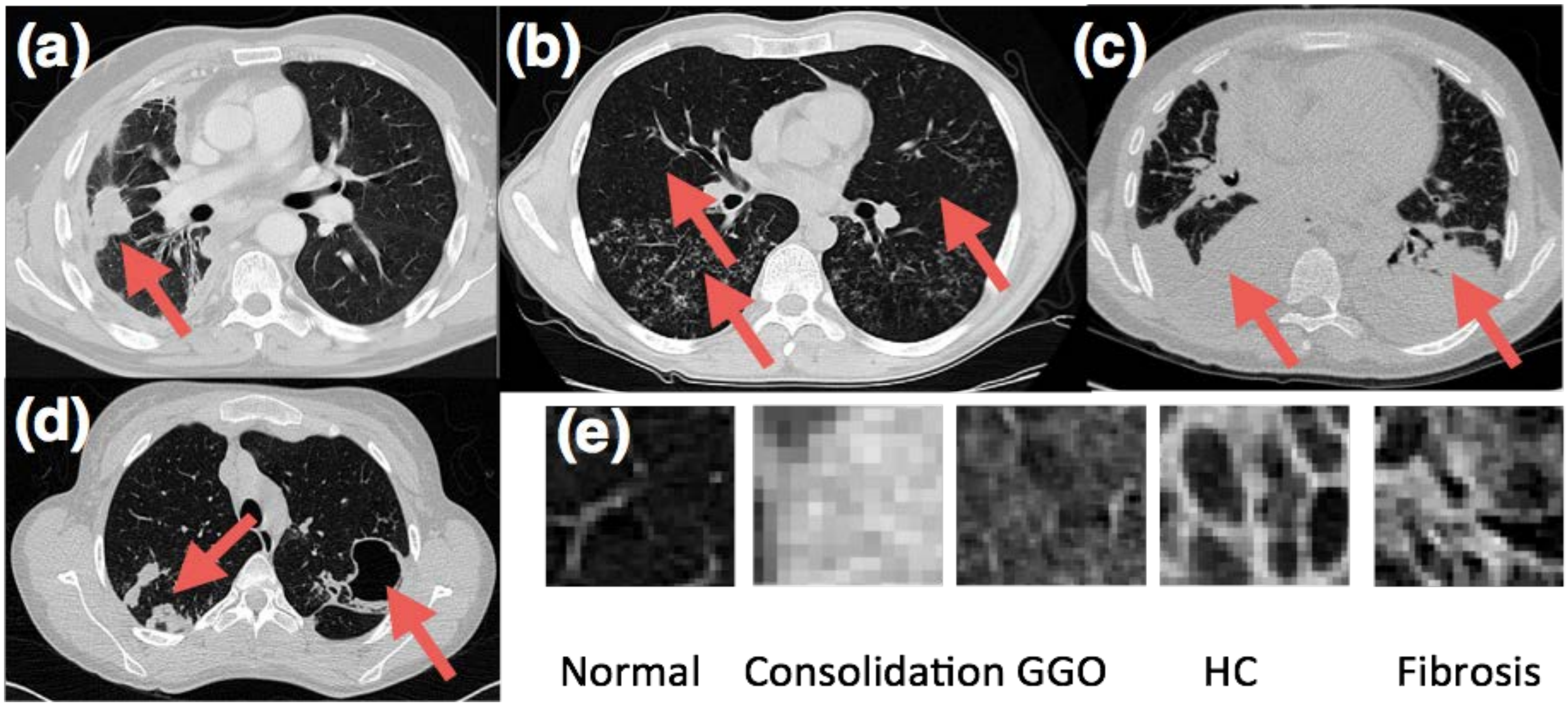


Mammogram Region
(a)

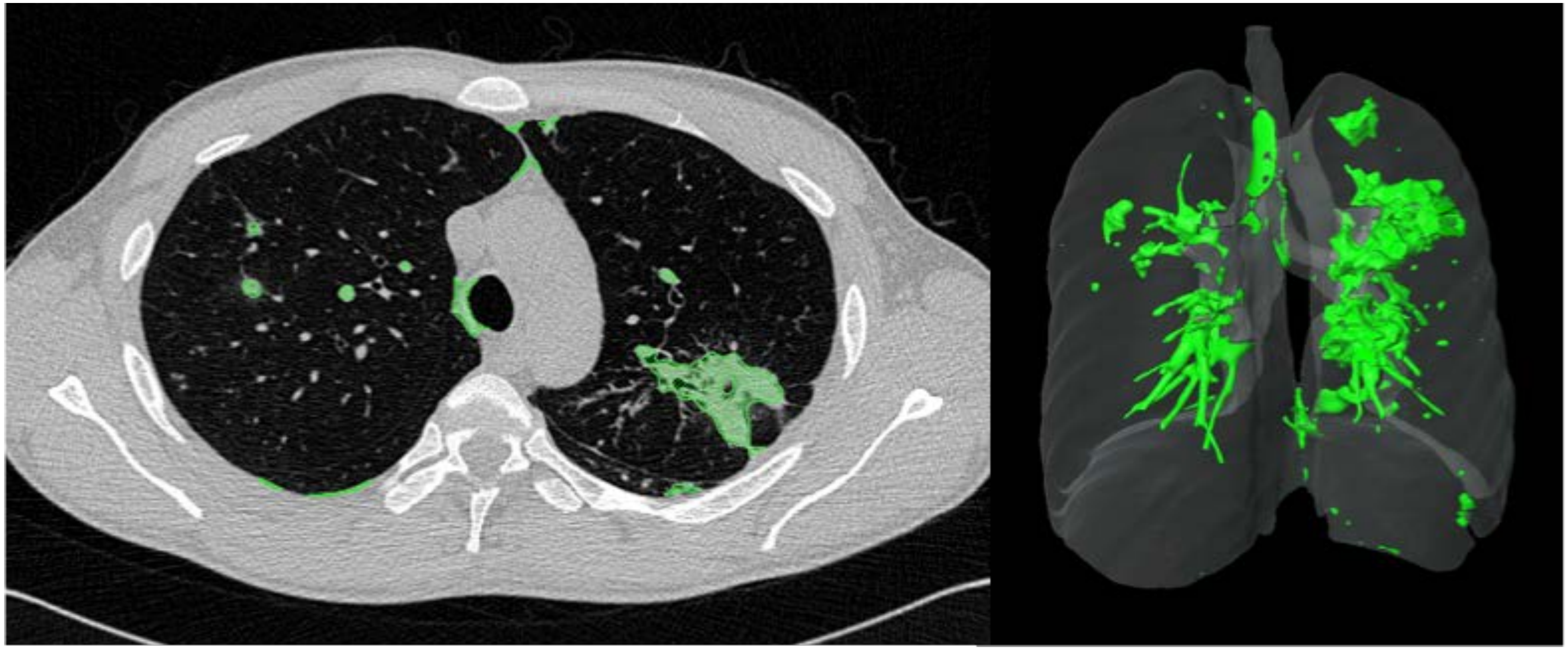
SVM Output
(b)

Detected Lesion Positions
(c)

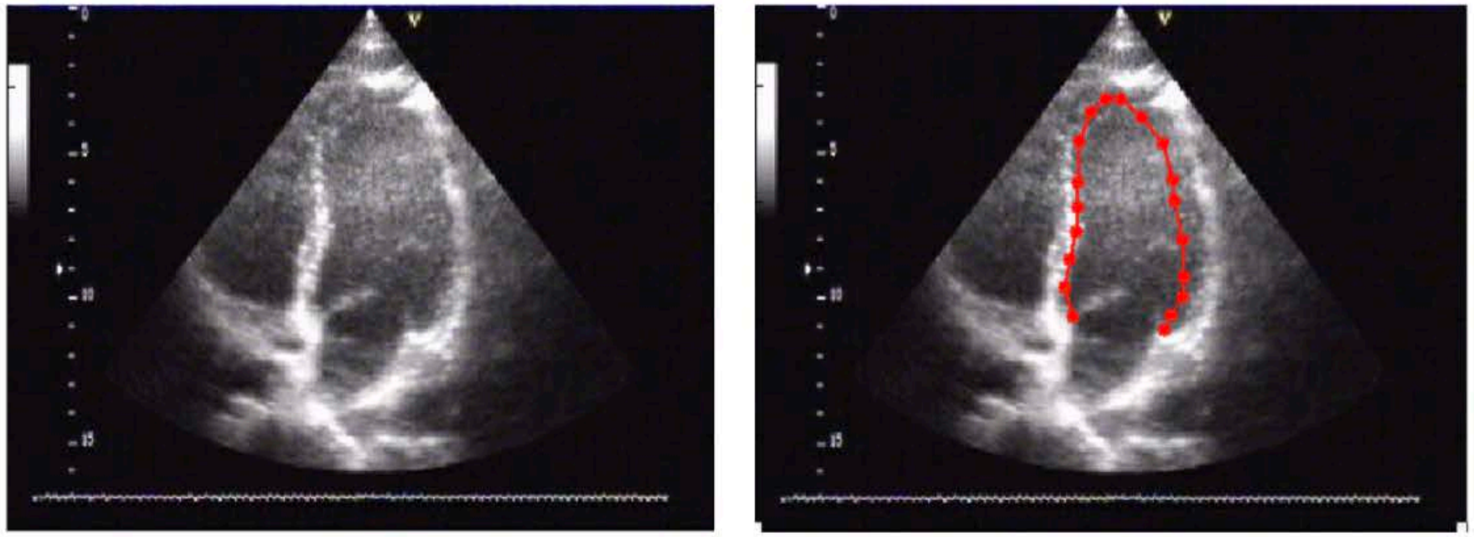
Examples: Detecting Pulmonary Abnormalities



Examples: Detecting Pulmonary Abnormalities



Examples: Segmentation of Ventricles



- Goals
  - Automated functional analysis of the heart
  - Improve workflow, reduce user variability
- Challenges
  - Low signal-to-noise ratio, edge dropout, shadows
  - Training set (machine learning methods need lots of annotated images)

Examples: Segmentation of Ventricles

DEEP BELIEF NETWORK

Topics: deep belief network

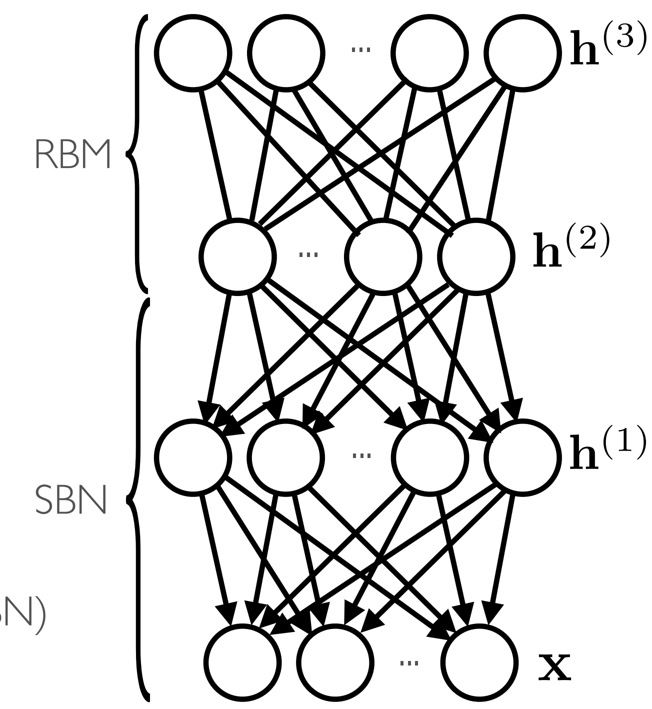
- The idea of pre-training came from work on deep belief networks (DBNs)
  - it is a generative model that mixes undirected and directed connections between variables
  - top 2 layers' distribution $p(\mathbf{h}^{(2)}, \mathbf{h}^{(3)})$ is an RBM
  - other layers form a Bayesian network:
    - the conditional distributions of a layers given the one above it are

$$p(h_j^{(1)} = 1|\mathbf{h}^{(2)}) = \mathrm{sigm}(\mathbf{b}^{(1)} + \mathbf{W}^{(2)\top}\mathbf{h}^{(2)})$$

$$p(x_i = 1|\mathbf{h}^{(1)}) = \mathrm{sigm}(\mathbf{b}^{(0)} + \mathbf{W}^{(1)\top}\mathbf{h}^{(1)})$$

    - this is referred to as a **sigmoid belief network** (SBN)
  - a DBN **is not** a feed-forward network

DBN's graphical model

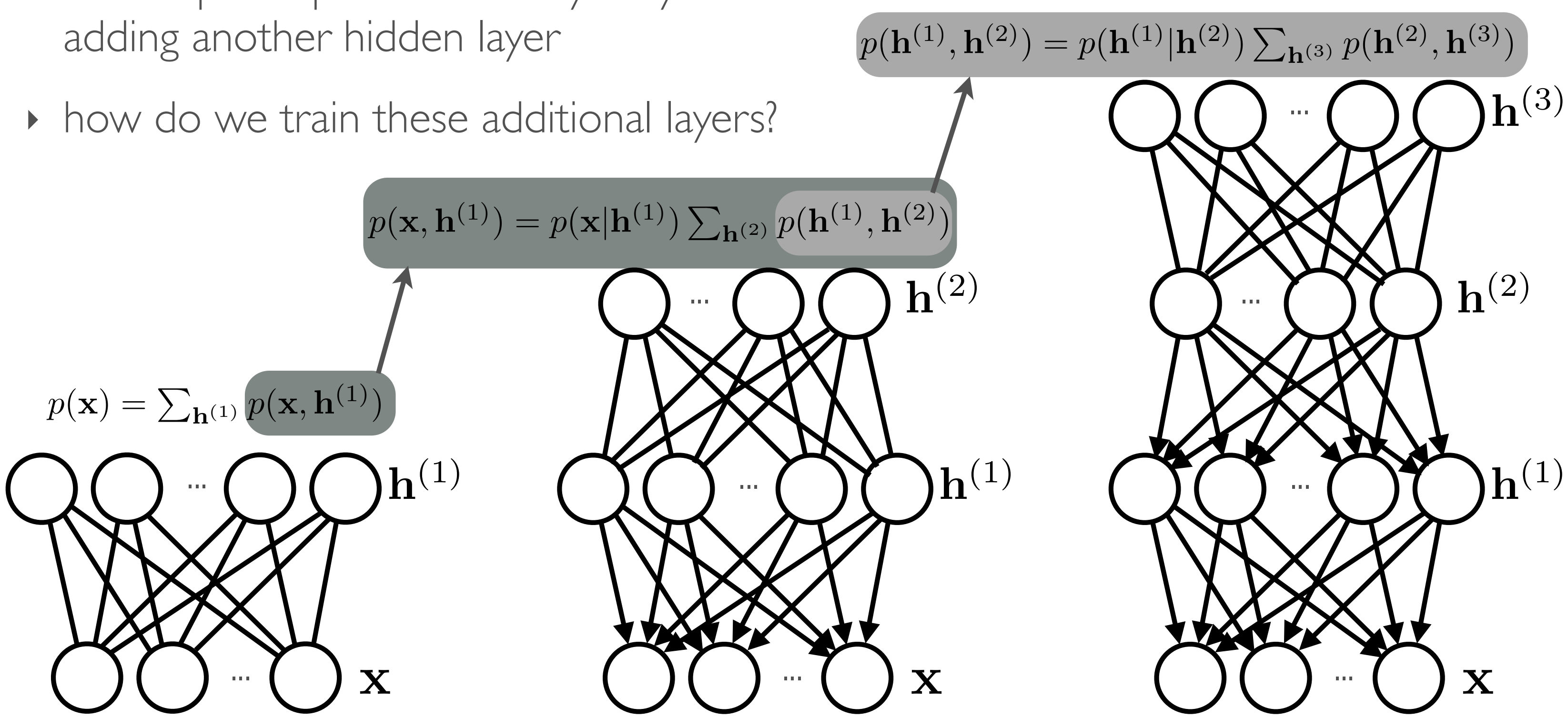


Examples: Segmentation of Ventricles

DEEP BELIEF NETWORK

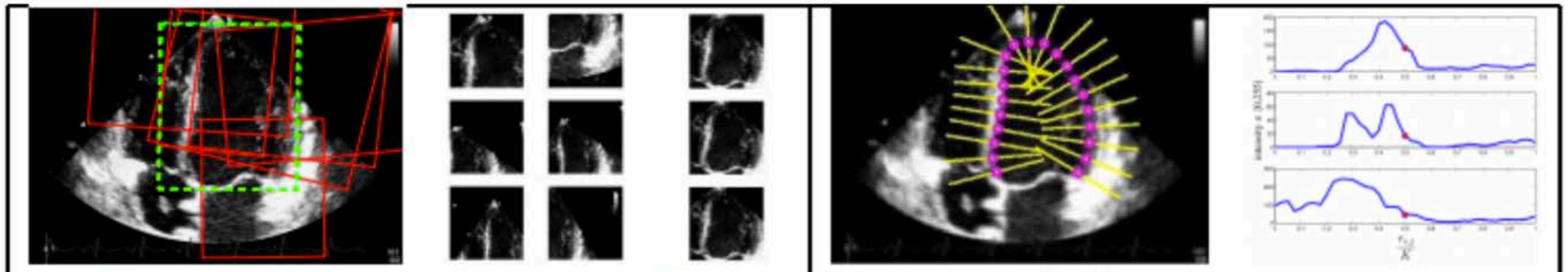
Topics: deep belief network

- This is where the RBM stacking procedure comes from
  - idea: improve prior on last layer by adding another hidden layer
  - how do we train these additional layers?



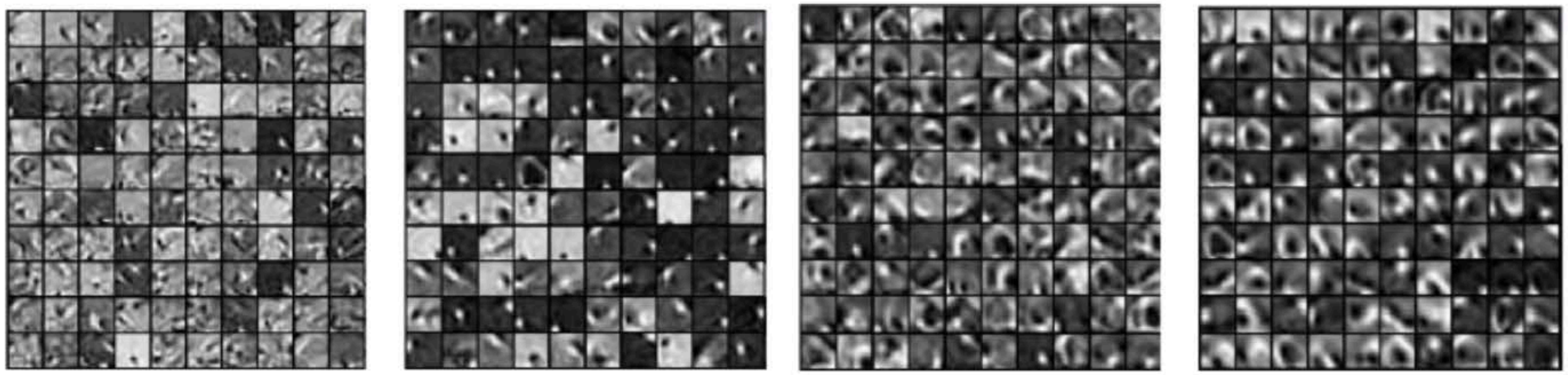
Examples: Segmentation of Ventricles

- Coarse to fine search strategy (3 scales)
- ROI detected from sampling initial distribution (fewer initial points, compared to grid search)
- Gradient-based search in fine stages (less computation than grid-based search)



Training image Negatives Positives
a) Rigid training patches

Perpendicular lines Profiles
b) Non-rigid training profiles

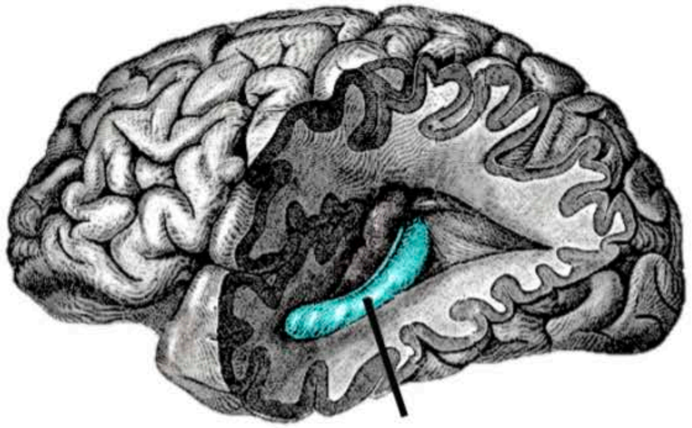


(a) Layer 1 (b) Layer 2 (c) Layer 3 (d) Layer 4

Examples: Segmentation of Hippocampus

- **Importance**

  The volume of the hippocampus is an important trait for early diagnosis of neurological diseases (e.g., Alzheimer's disease)

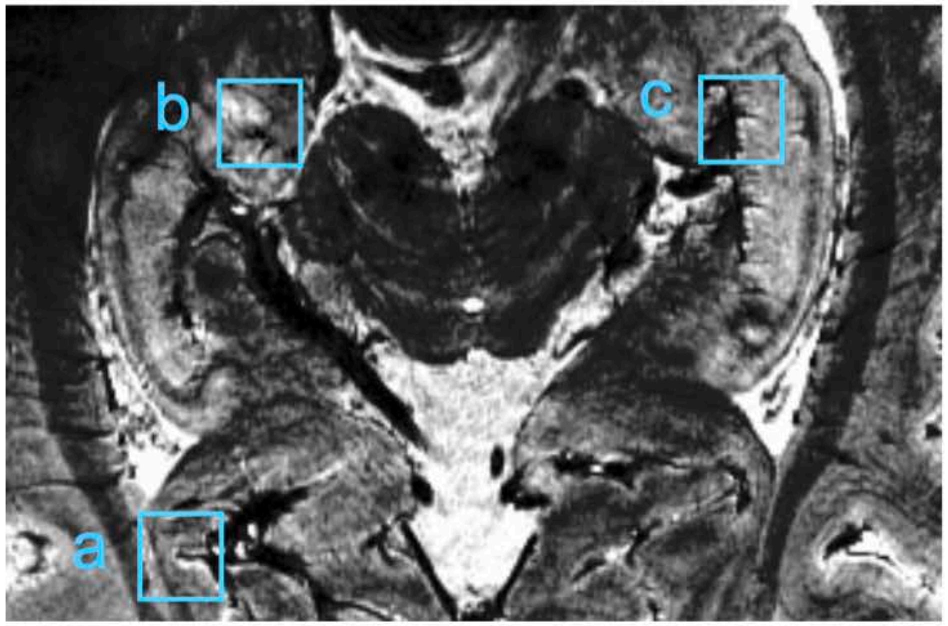


- **Challenges**
  - The hippocampus is small ($\approx 35\times 15\times 7mm^3$)
  - The hippocampus is surrounded by complex structures
  - Low imaging resolution ($\approx 1\times 1\times 1mm^3$) of 1.5T or 3T MRI scanners

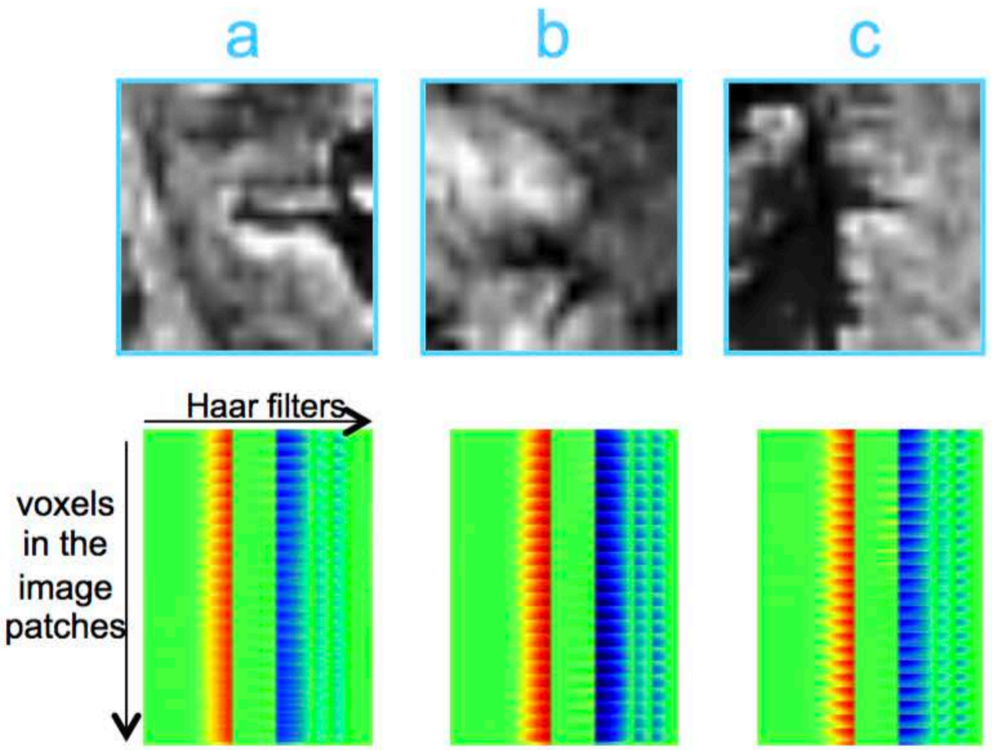
Examples: Segmentation of Hippocampus

Hand-Crafted Features

- **Limited discriminative power**



Extracting patches from a 7T MR image

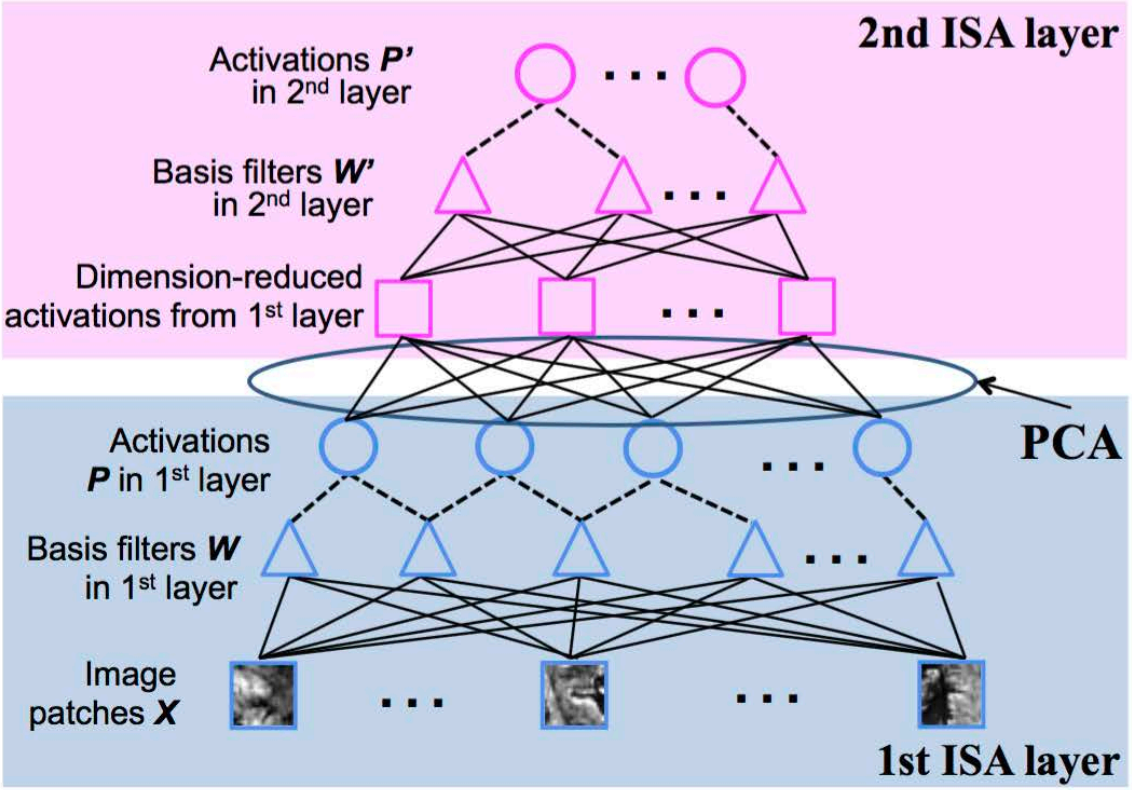


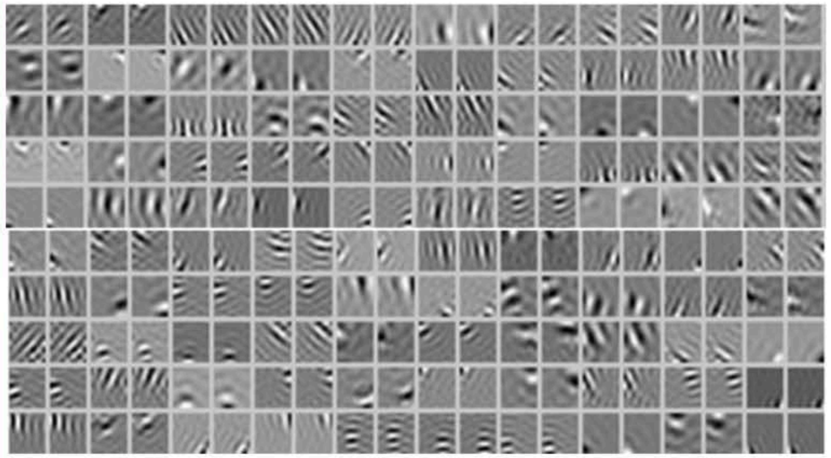
Responses of Haar filters for the image patches in a-c

Examples: Segmentation of Hippocampus

Hierarchical Feature Extraction

Stacked two-layer convolutional ISA (Independent Subspace Analysis)

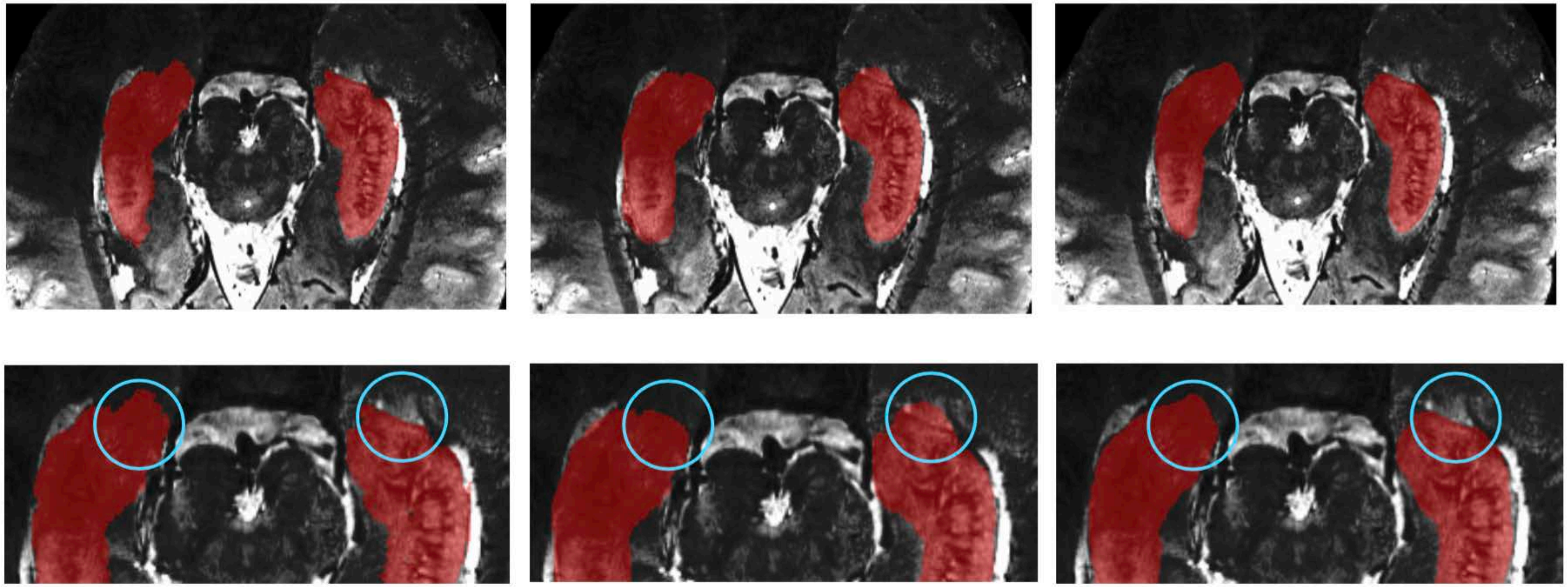




Learned basis filters by the 1st ISA

Examples: Segmentation of Hippocampus

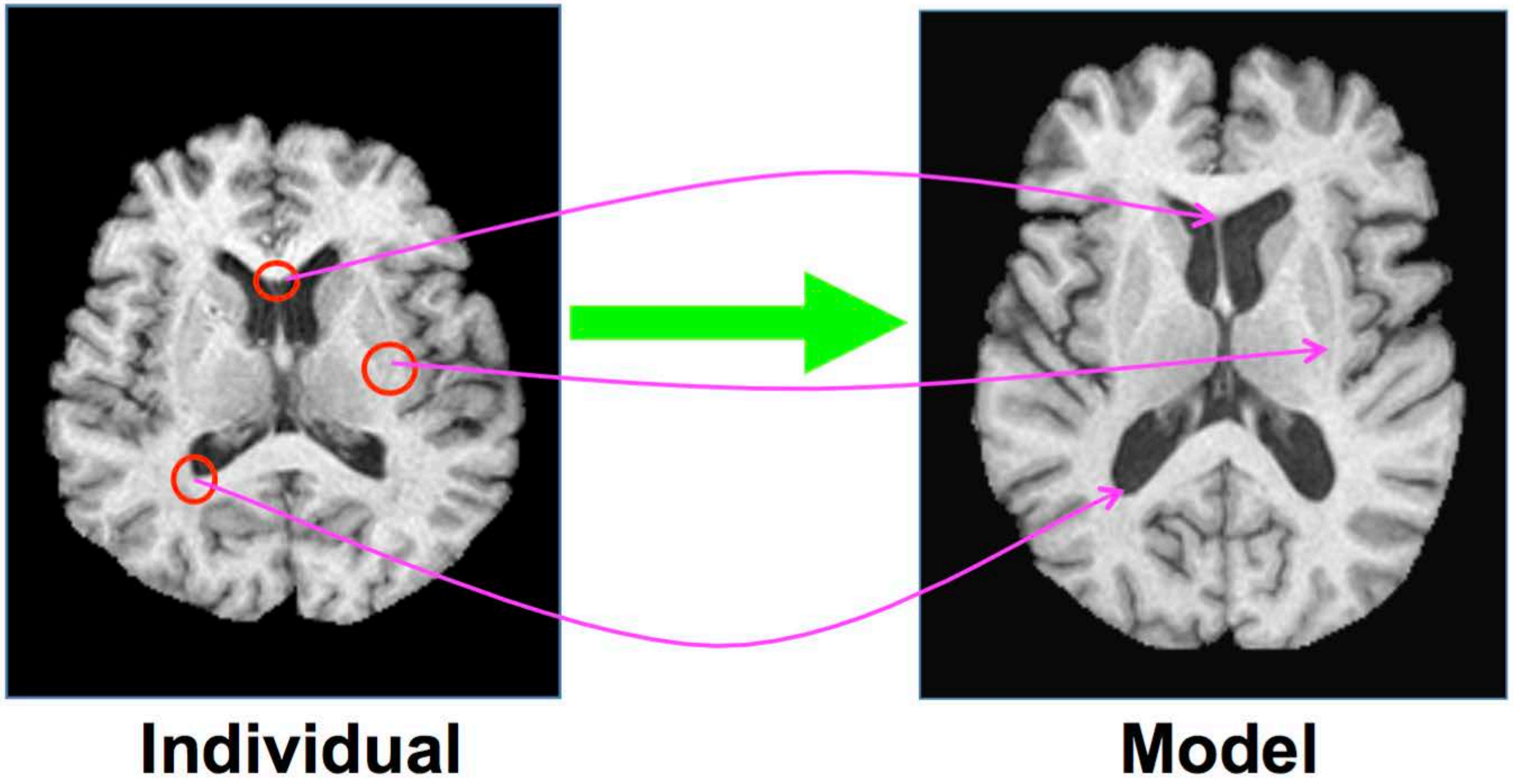
Qualitative Evaluations



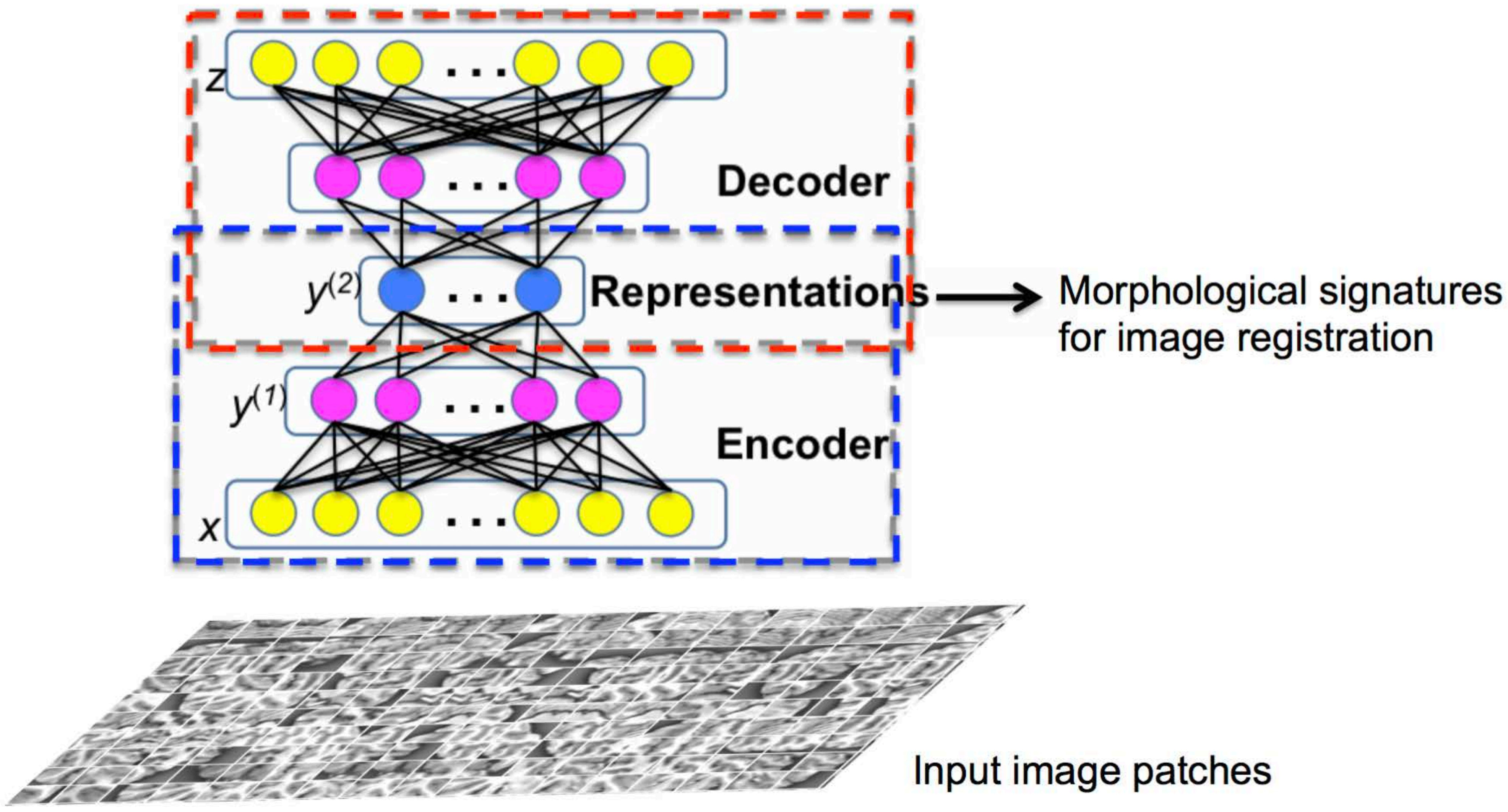
Ground Truth **Haar + Texture Features** **Hierarchical Features**

Examples: Image Registration

Determine accurate correspondences between images



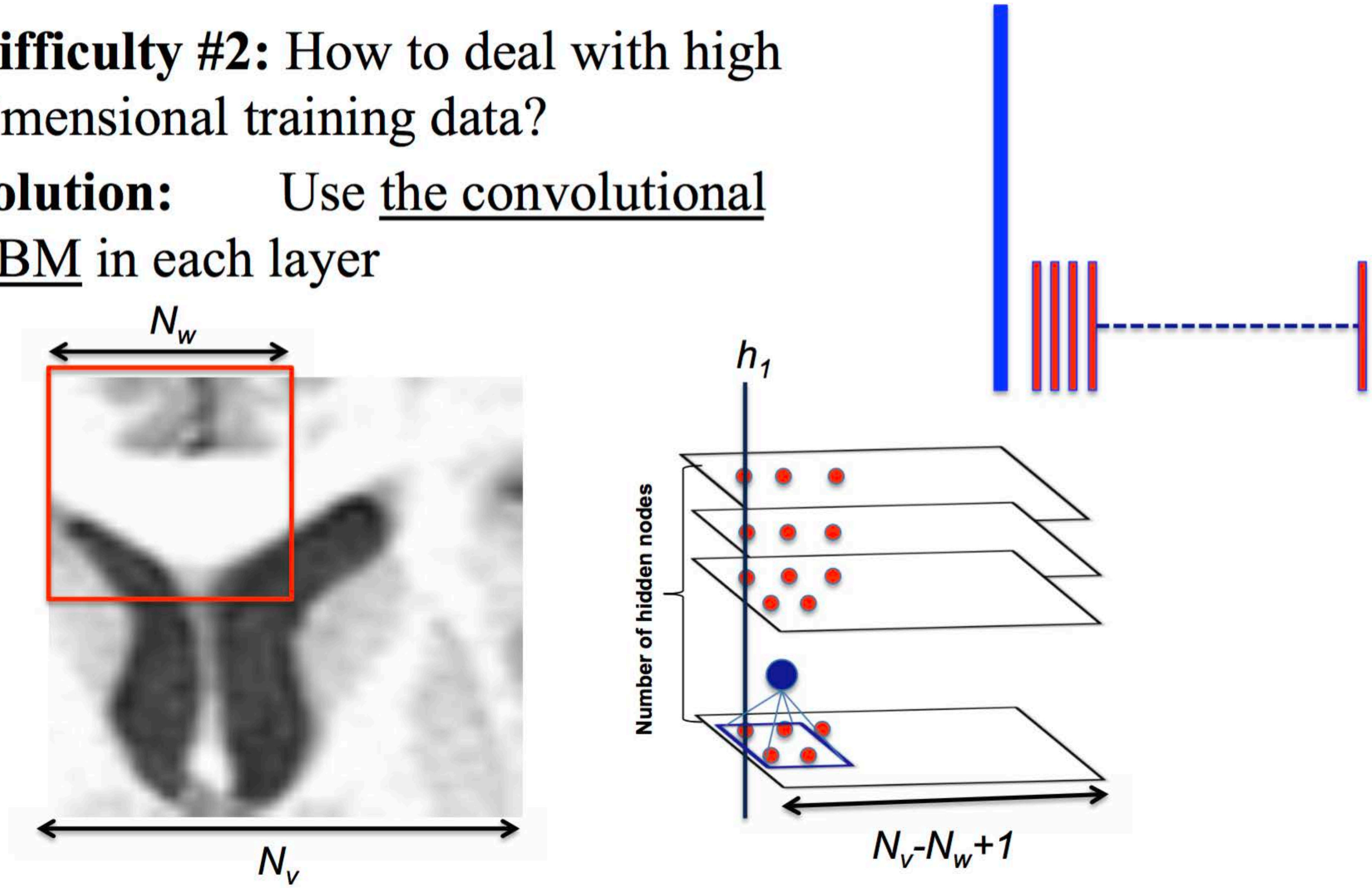
Examples: Image Registration



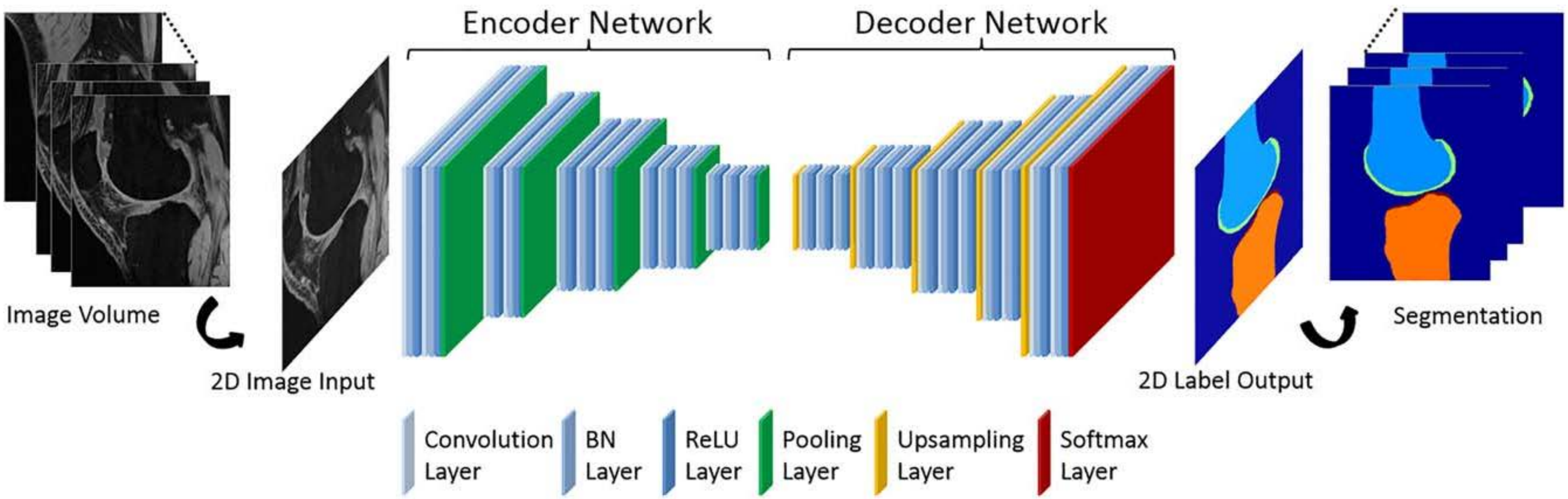
Examples: Image Registration

Difficulty #2: How to deal with high dimensional training data?

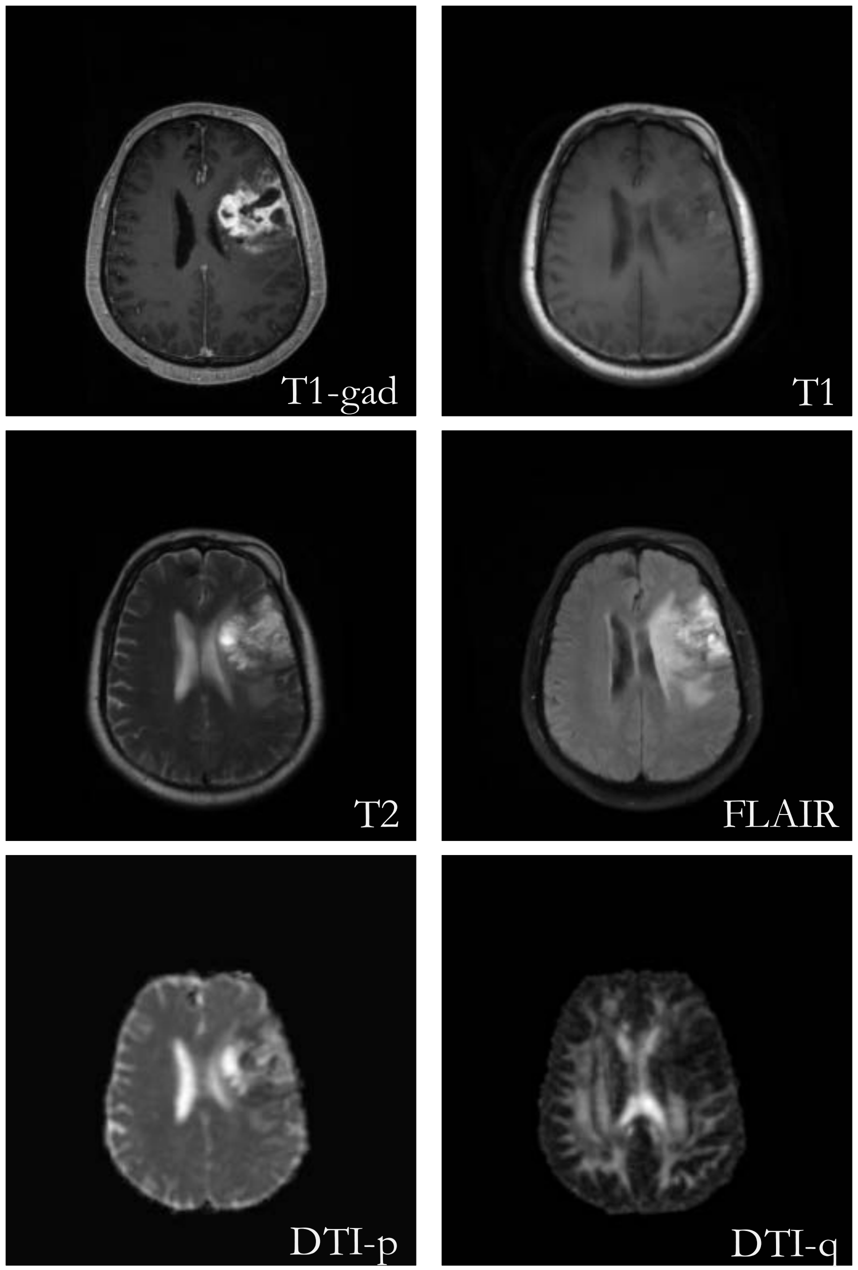
Solution: Use the convolutional RBM in each layer



Examples: Tissue-Specific Segmentation

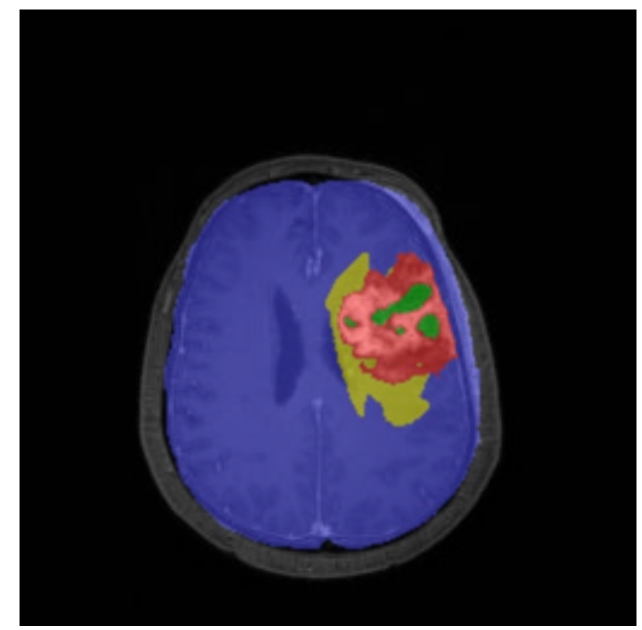


Examples: Tissue-Specific Segmentation



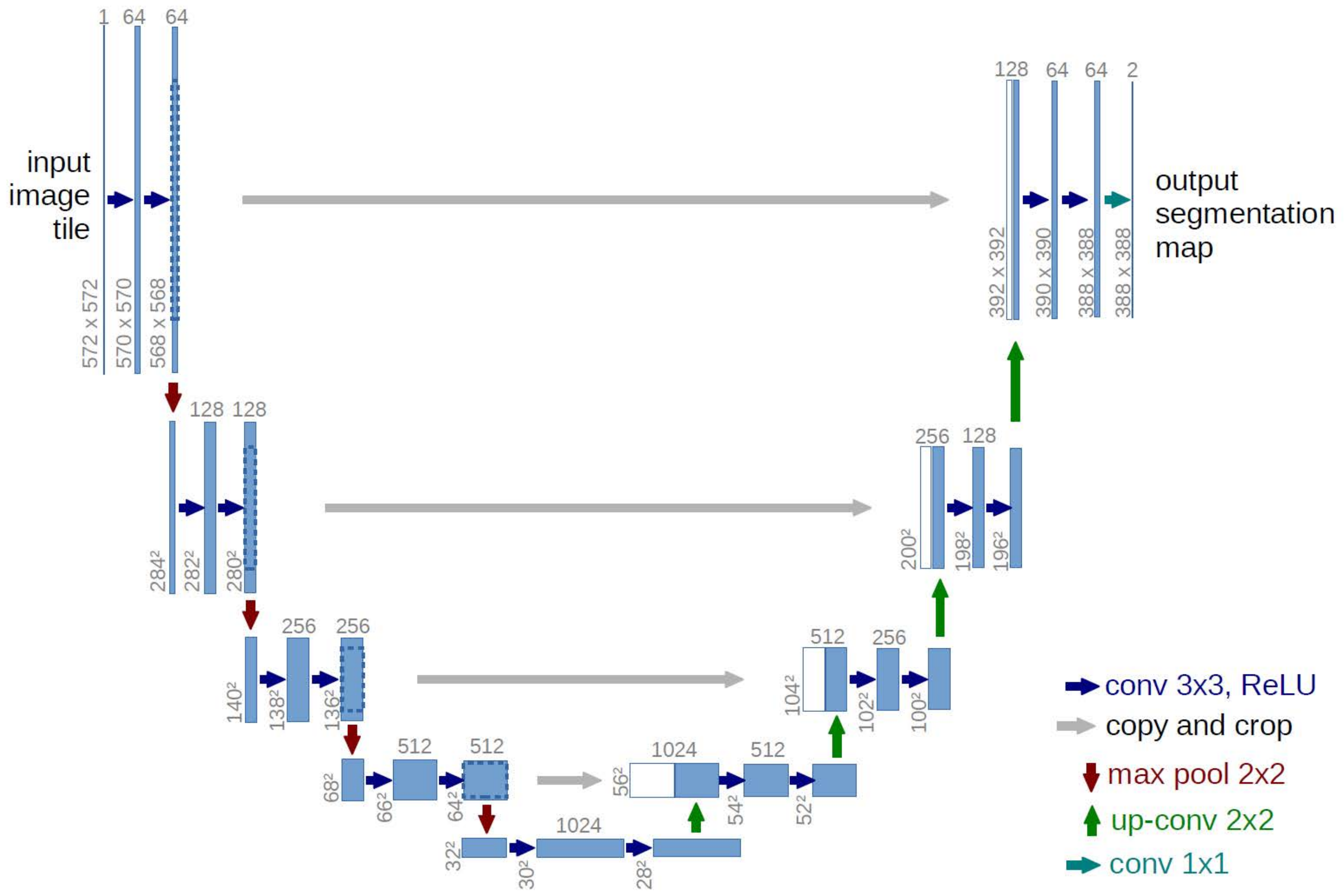
Multi-channel 3D MRI input data

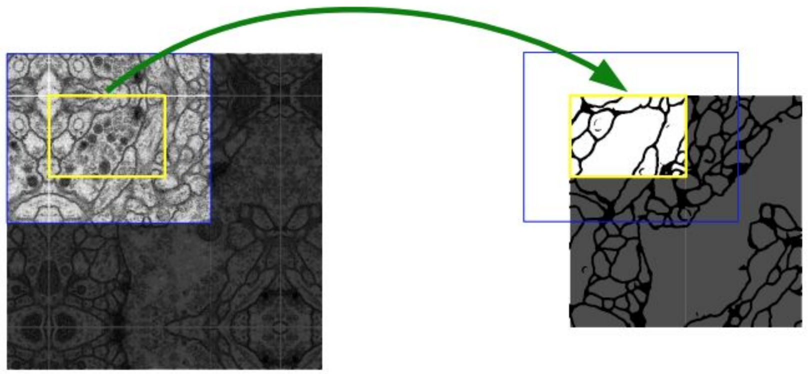
Segmentation of tumorous tissues:



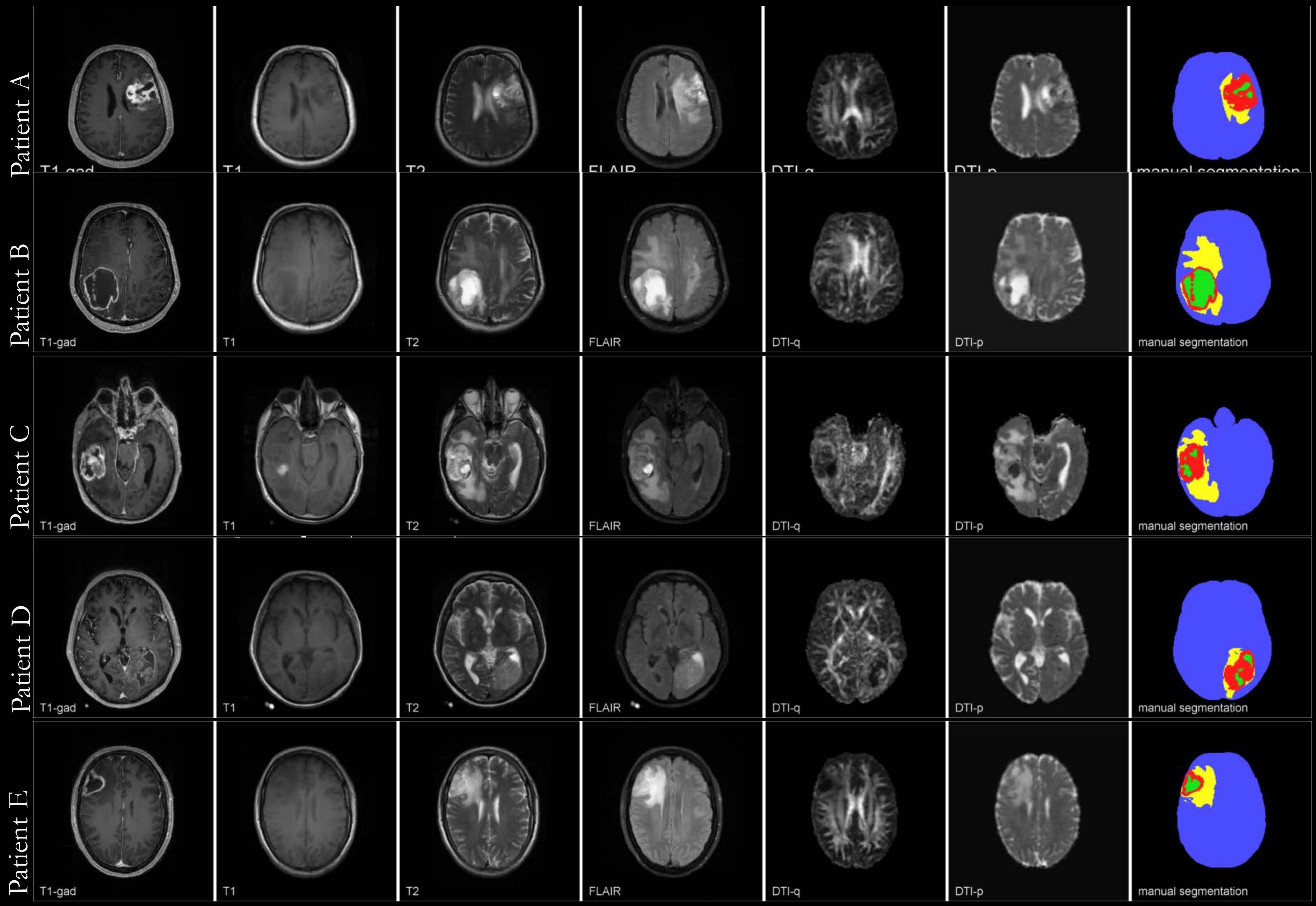
---- Active cells
---- Necrotic core
---- Edema
---- Background

Examples: Tissue-Specific Segmentation

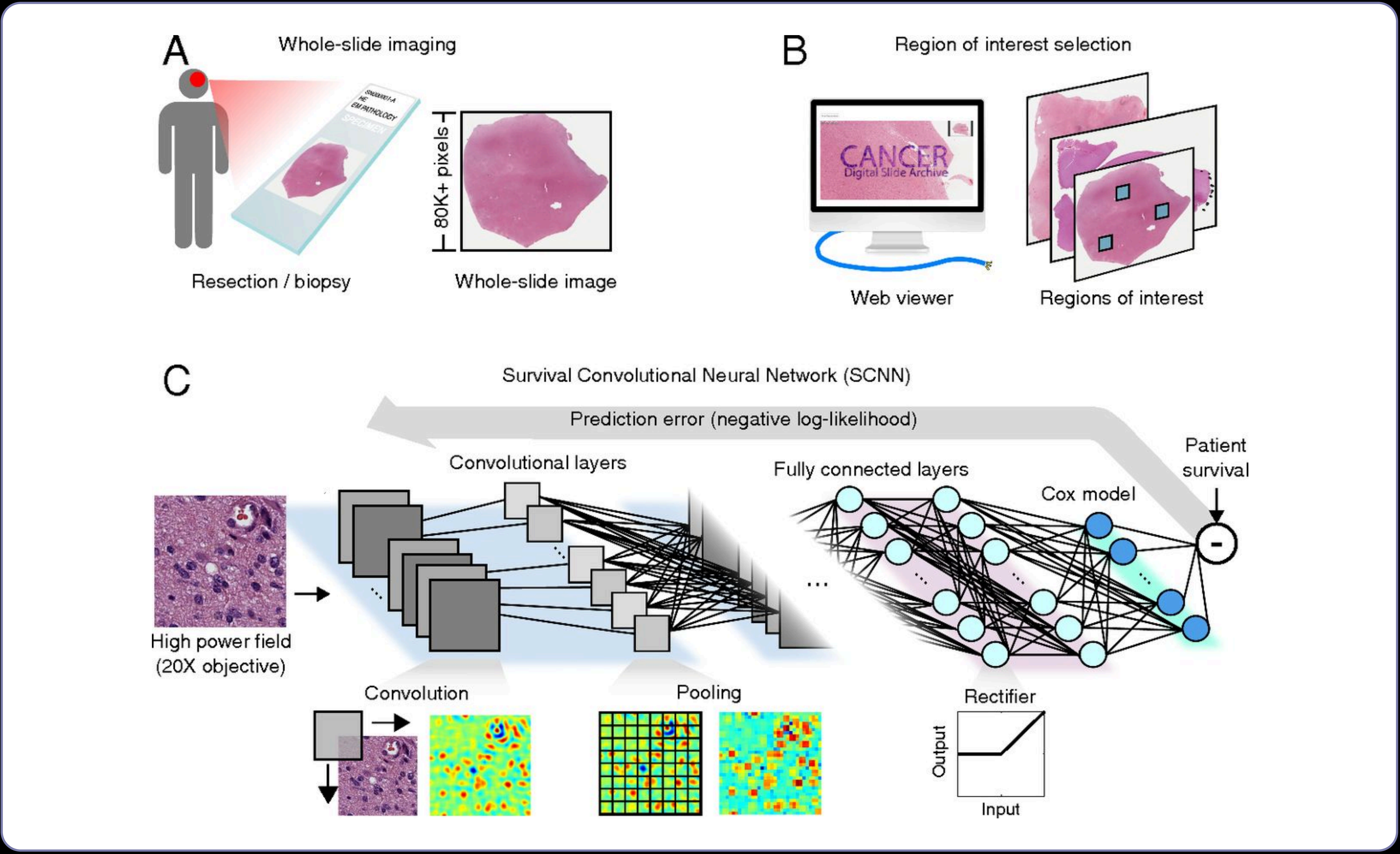




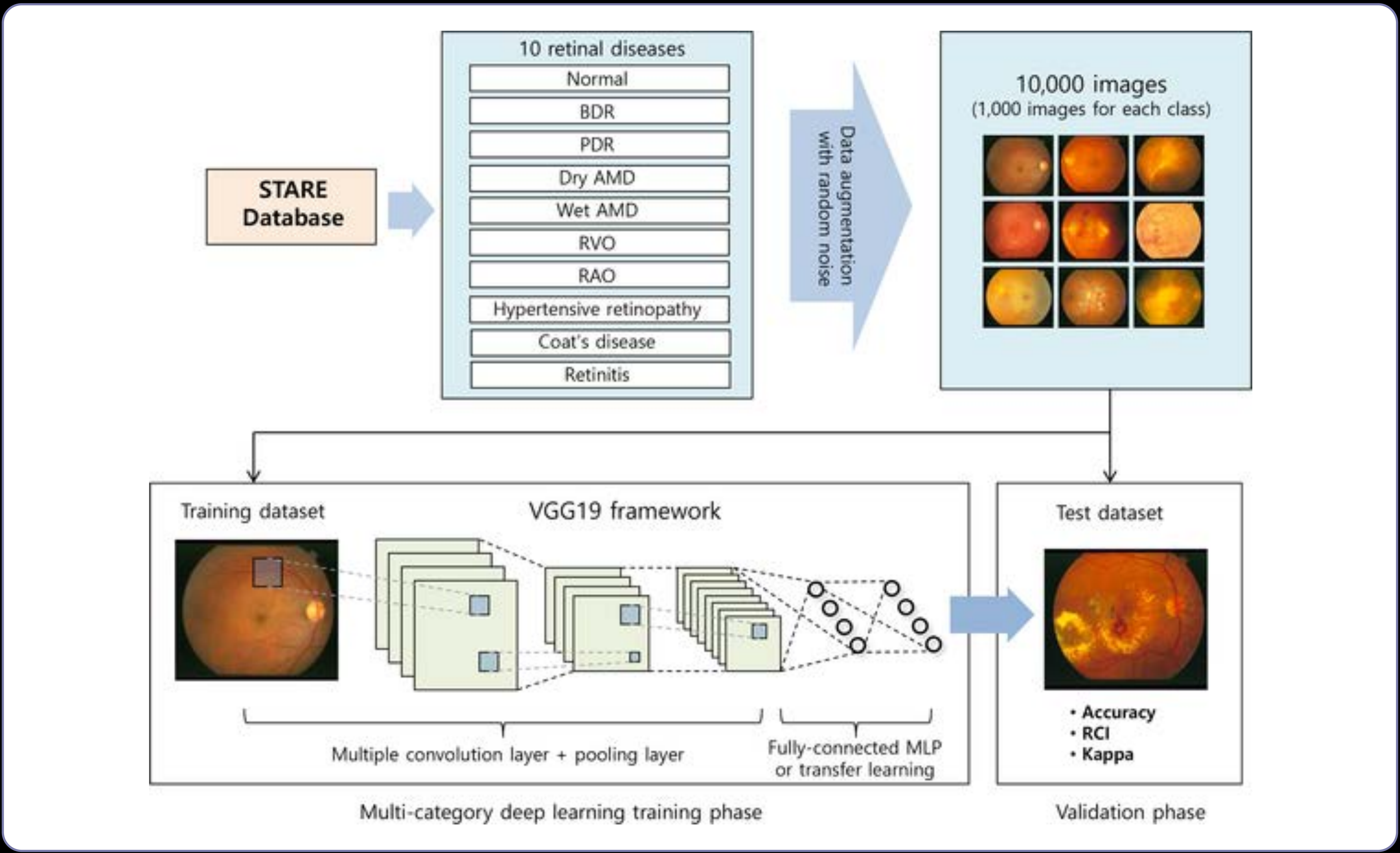
Examples: Tissue-Specific Segmentation



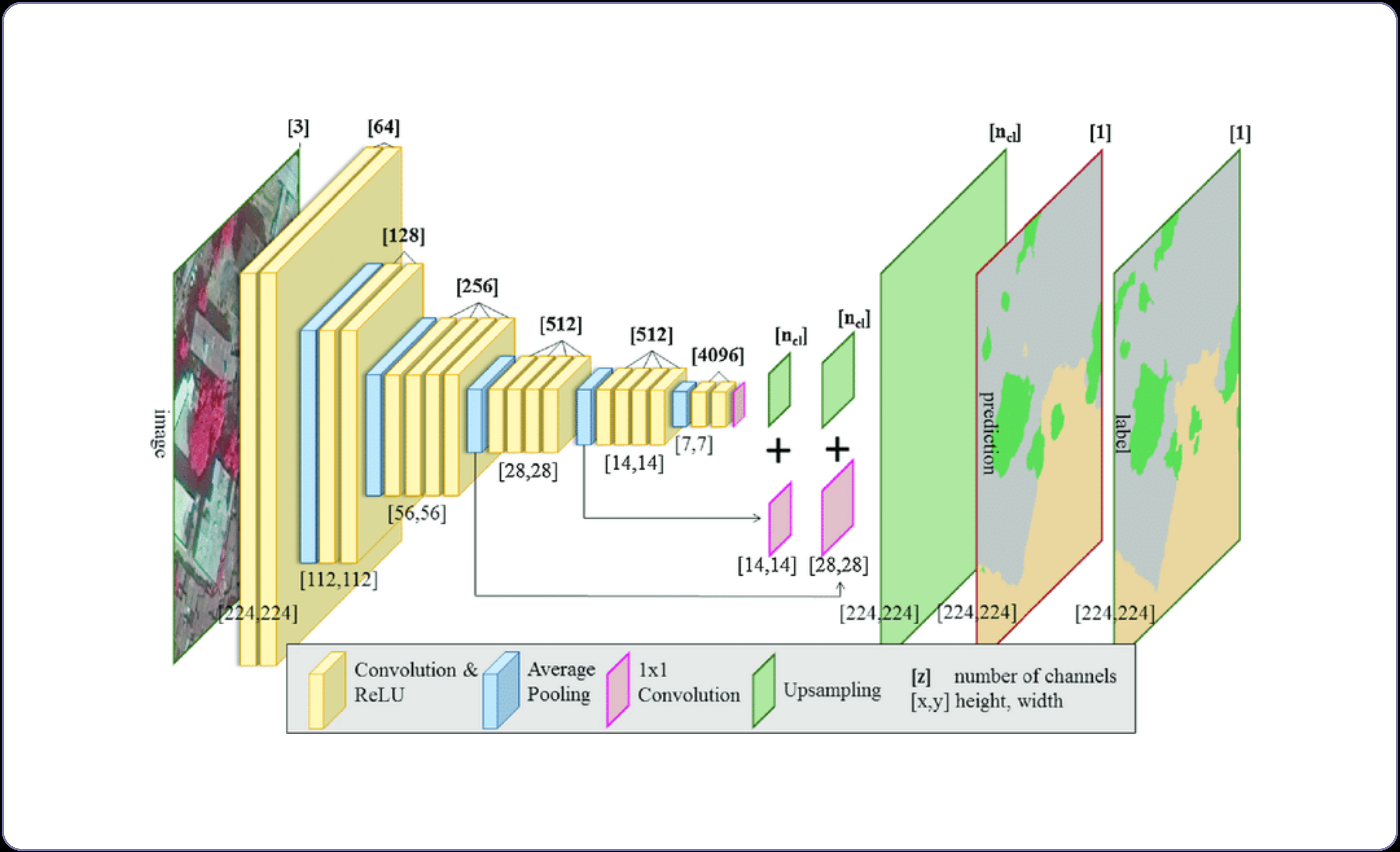
Examples: Predicting Survival from Histopathology



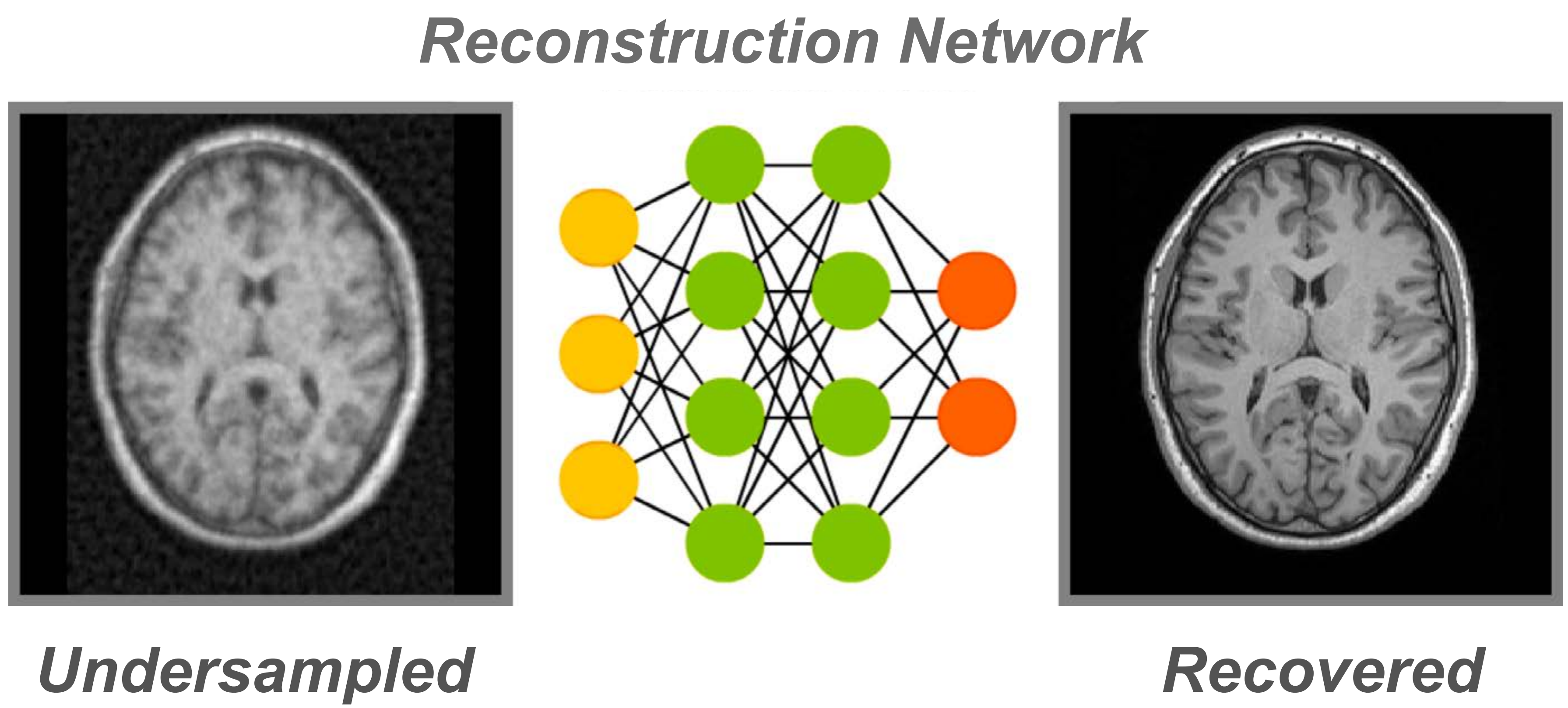
Examples: Classifying Retinal Disease



Examples: Classifying Retinal Disease

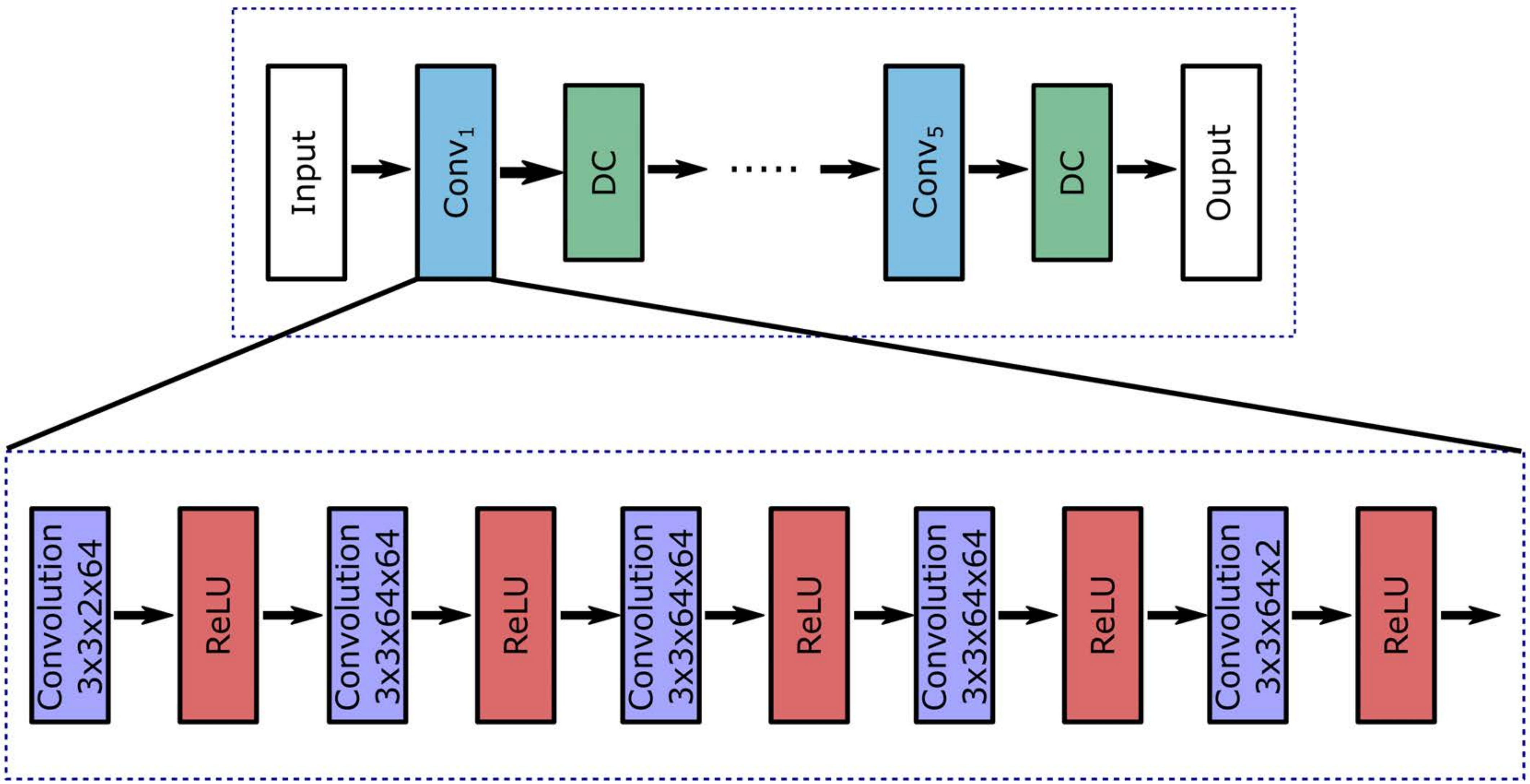


Examples: Denoising/Dealiasing Images

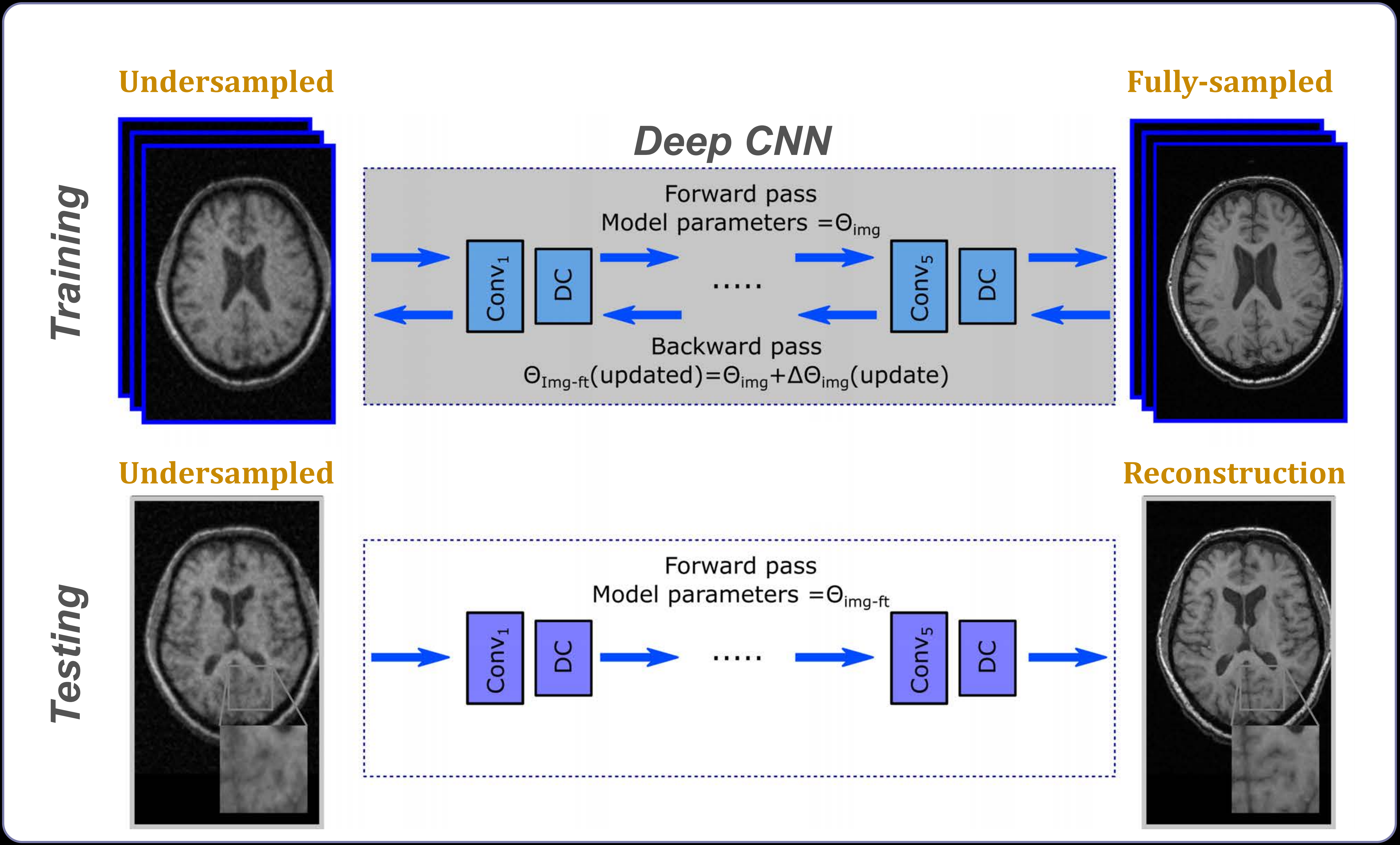


$$\hat{x} = \min_x \underbrace{\left\| F_u x - y_u \right\|_2}_{\text{Data consistency}} + \underbrace{\left\| C(x_u) - x \right\|^2}_{\text{Consistency with network}}$$

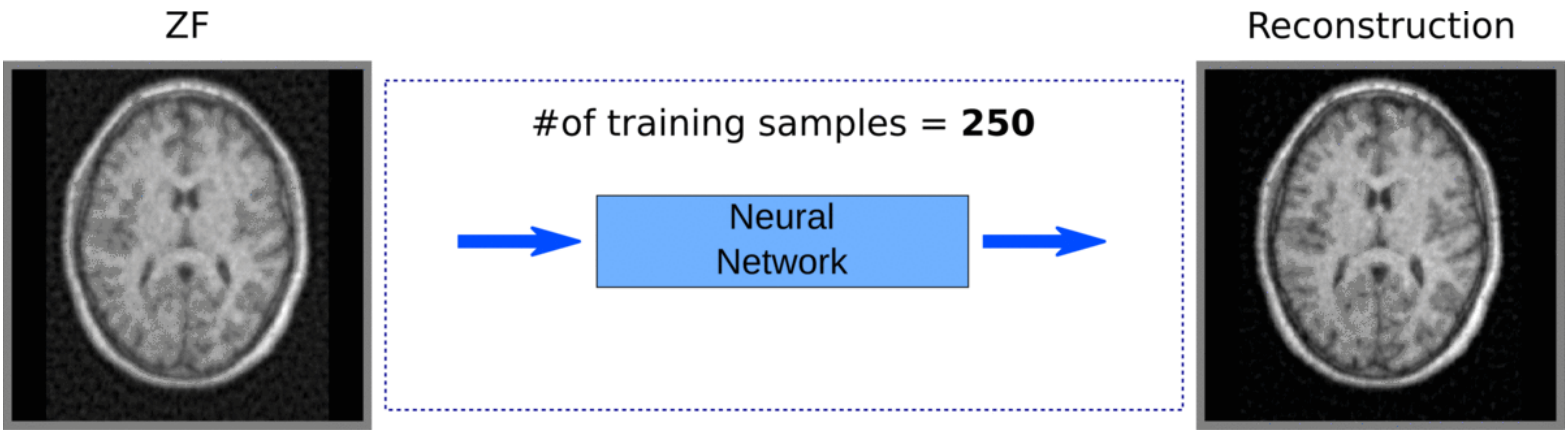
Model-based Deep Learning: Cascaded CNNs



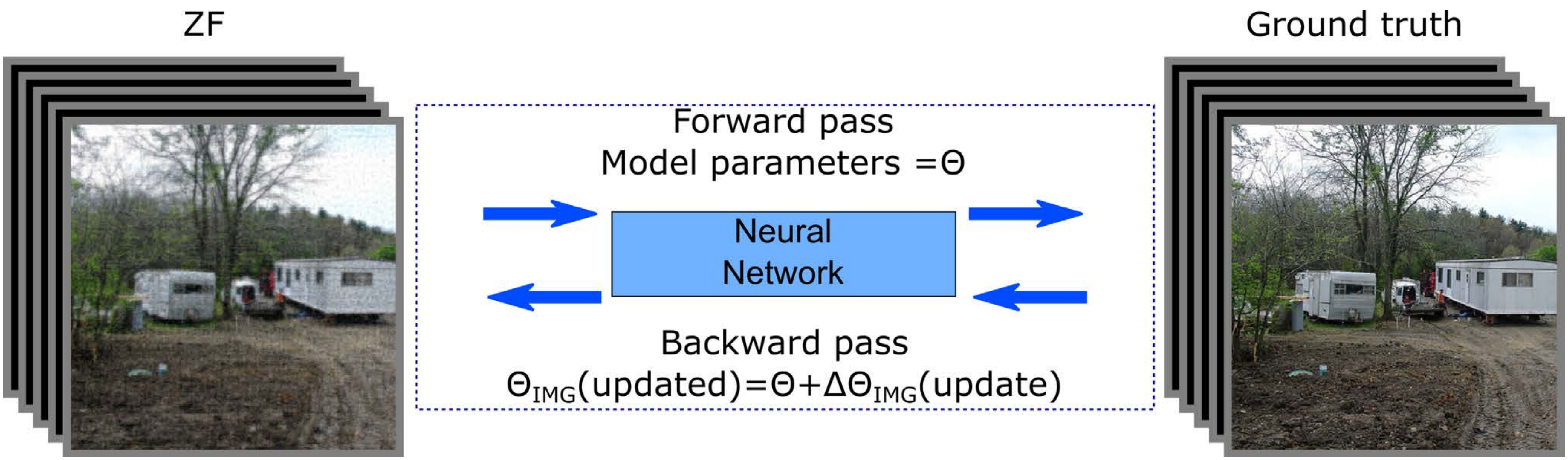
Examples: Denoising/Dealiasing images



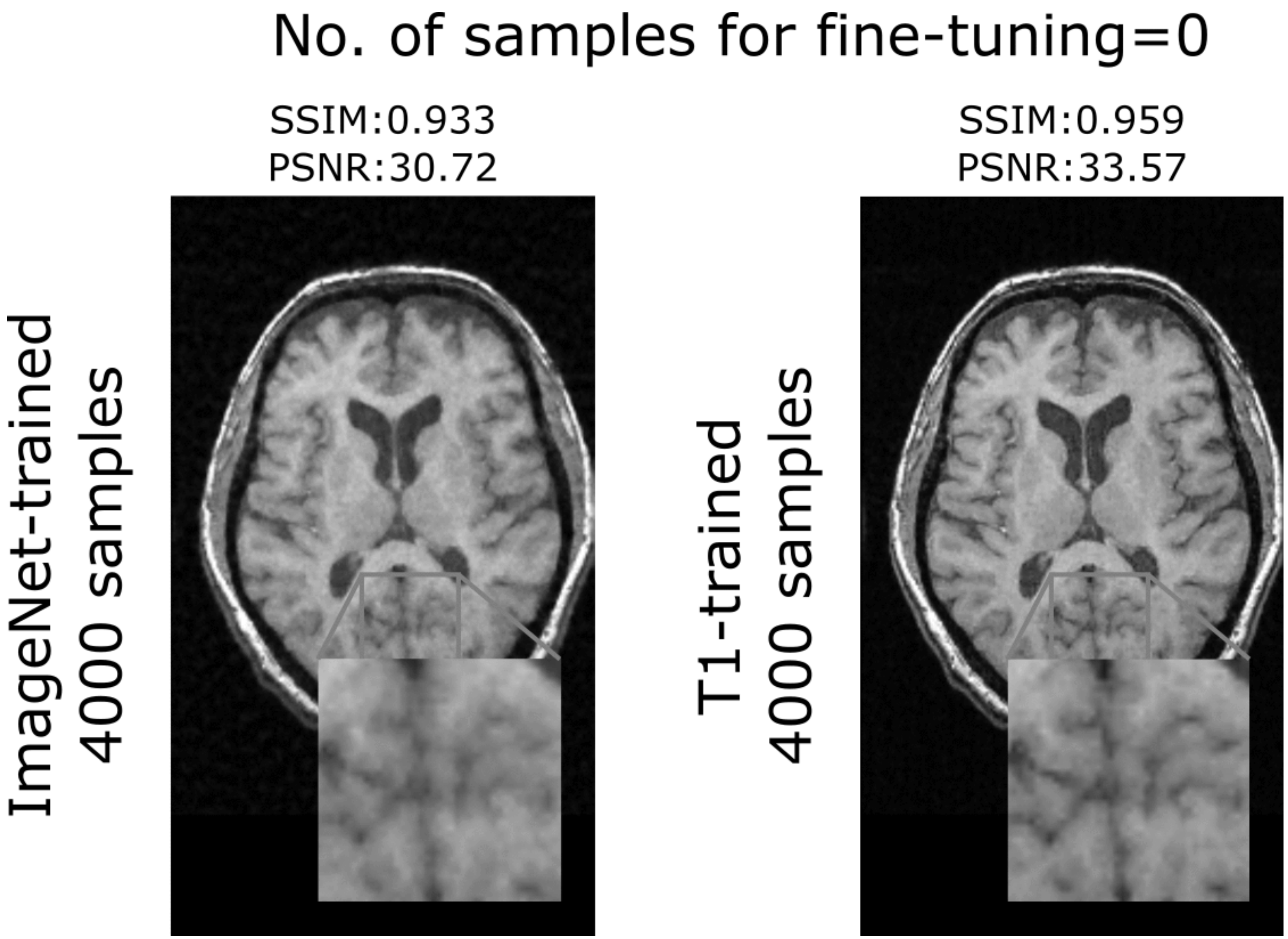
Medical Data Are Scarce



Transfer Learning

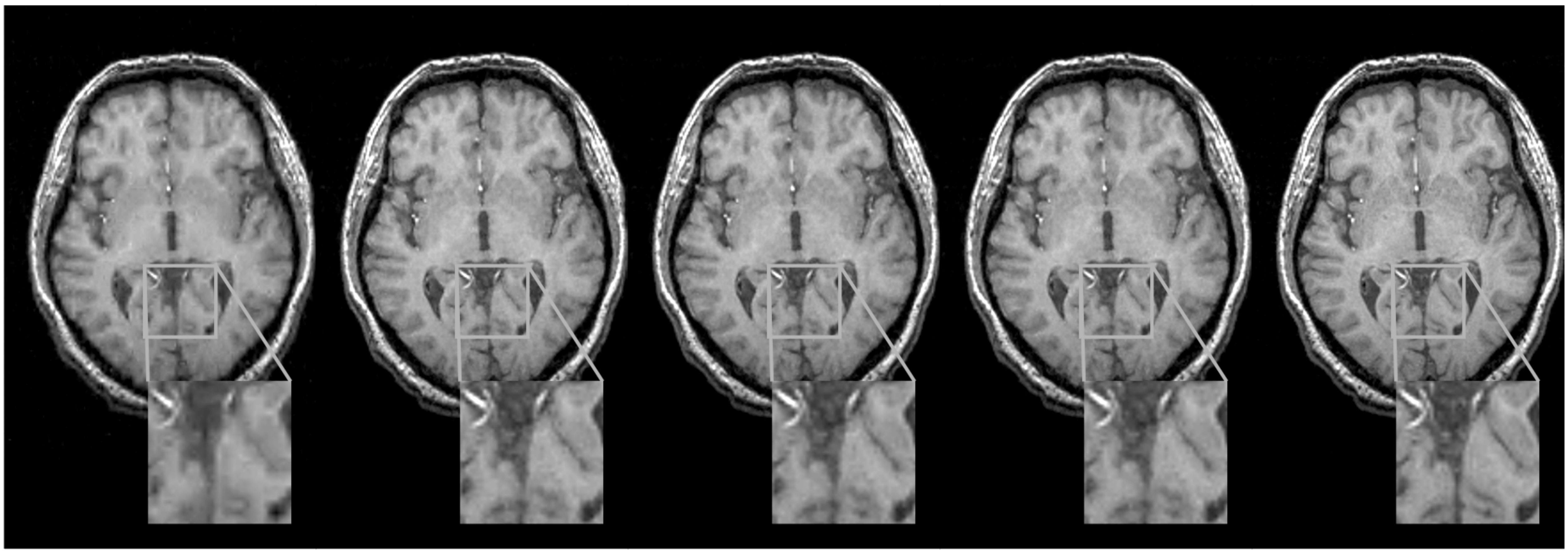


Fine Tuning

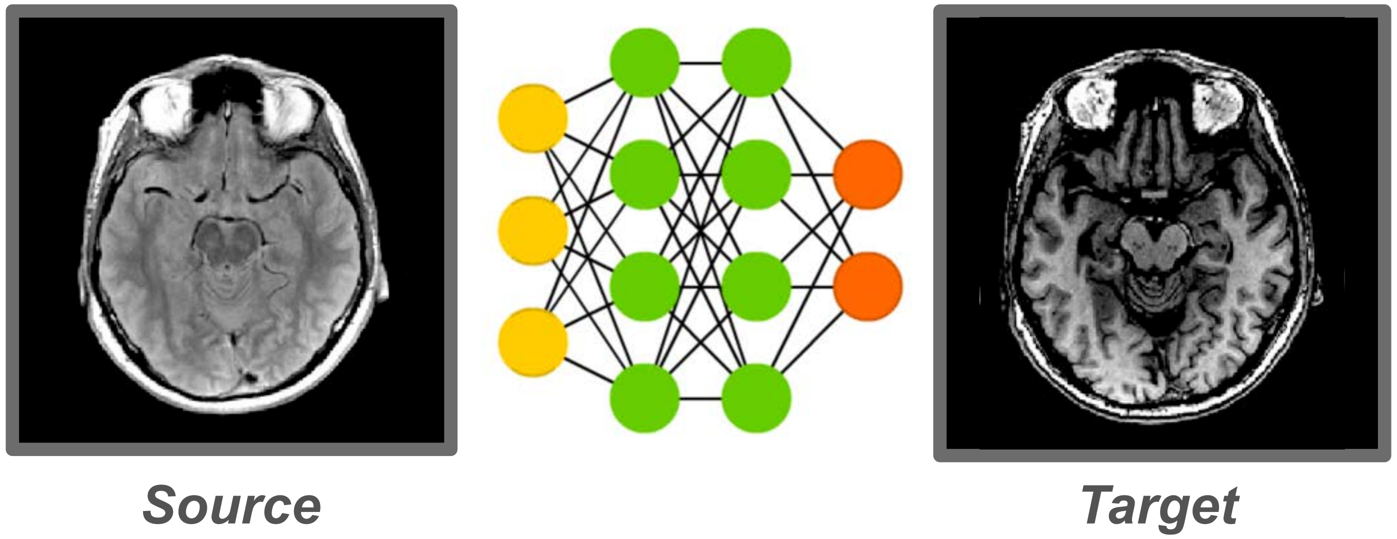

No. of samples for fine-tuning=0
SSIM:0.933
PSNR:30.72
SSIM:0.959
PSNR:33.57
ImageNet-trained
4000 samples
T1-trained
4000 samples

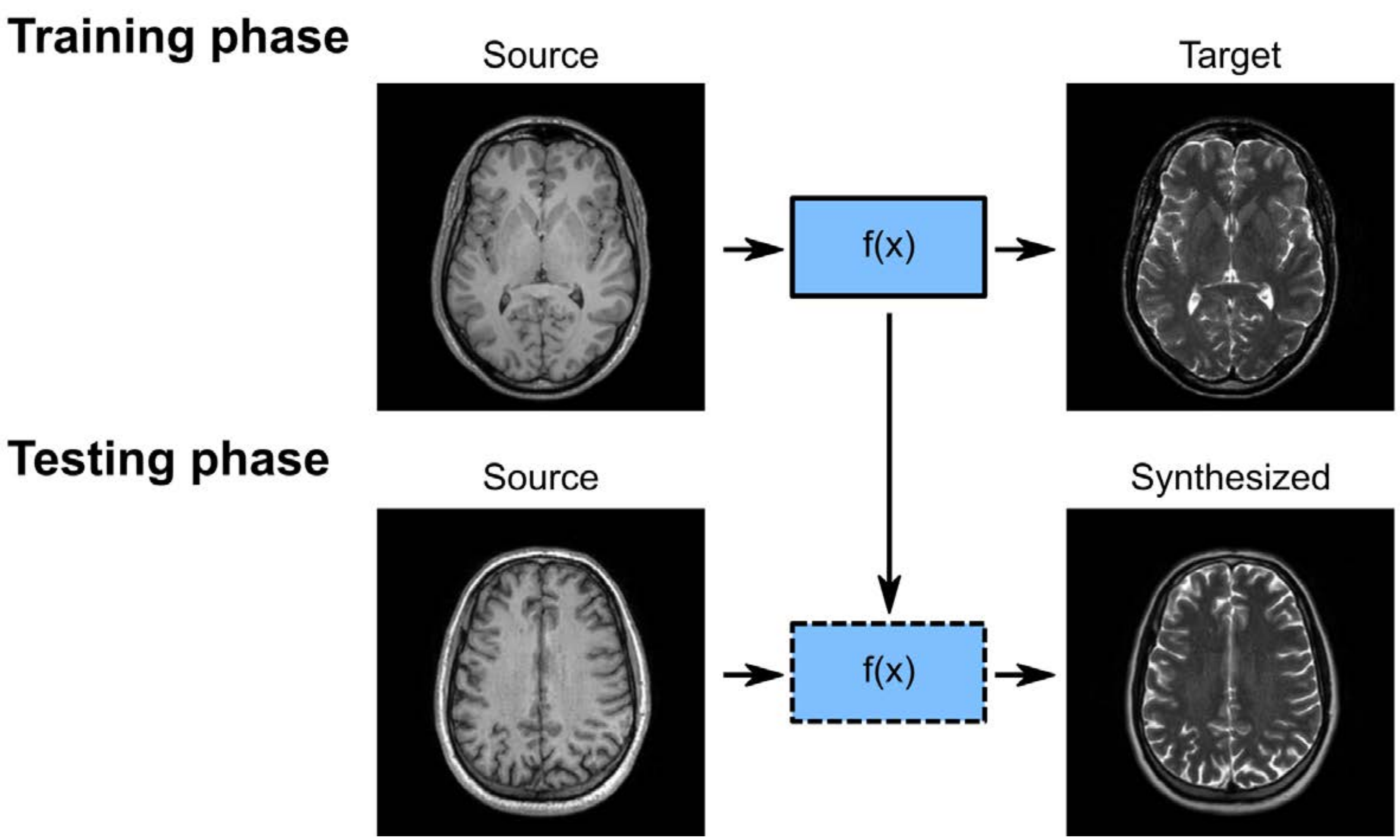
Examples: Denoising/Dealiasing images

CS	ImageNet-trained	T1-trained	T2-trained	Reference
SSIM:0.928 PSNR:30.79	SSIM:0.956 PSNR:33.29	SSIM:0.958 PSNR:33.60	SSIM:0.956 PSNR:33.39	SSIM:1.00 PSNR:Inf

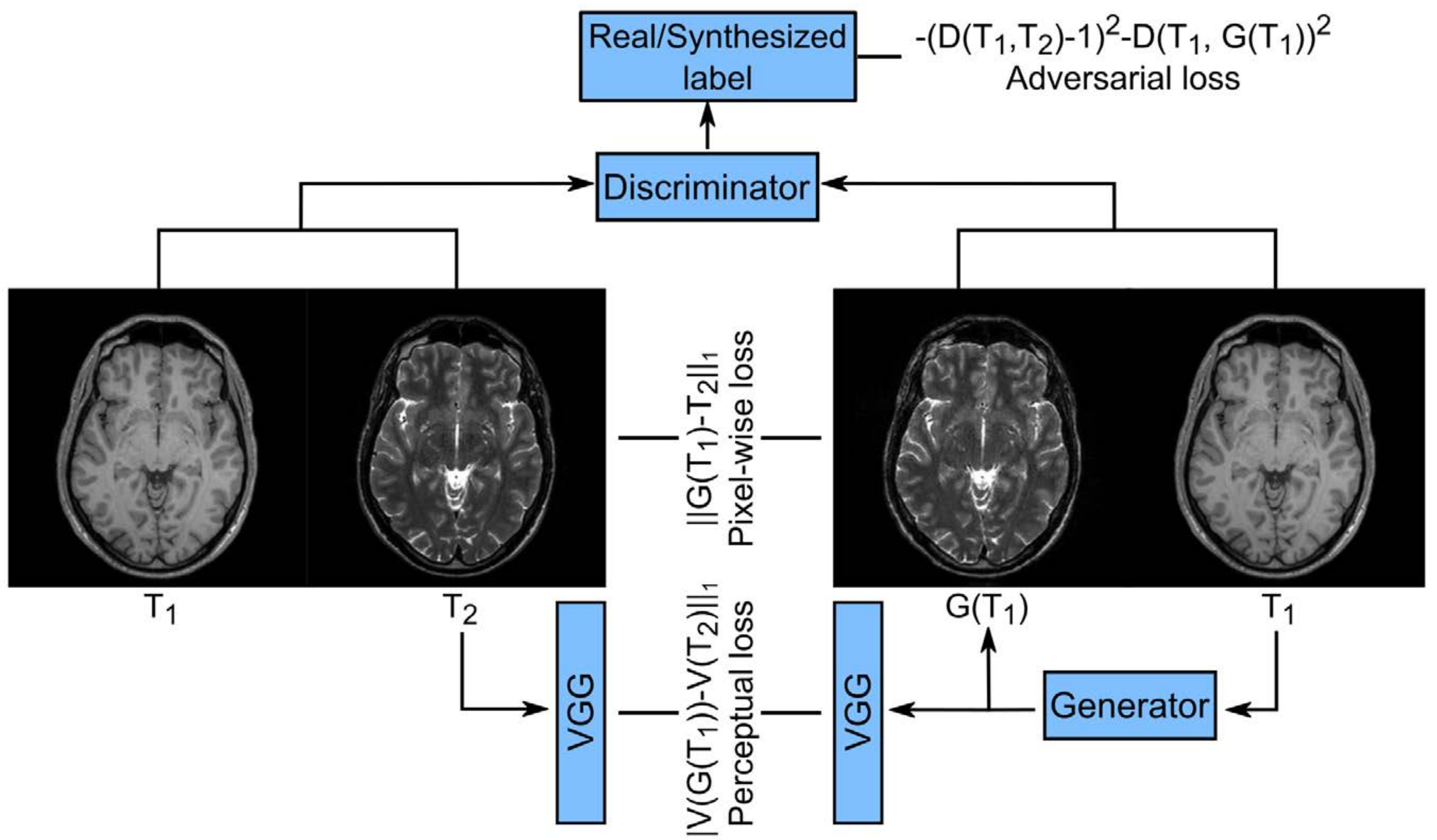


Examples: Synthesizing Missing Images

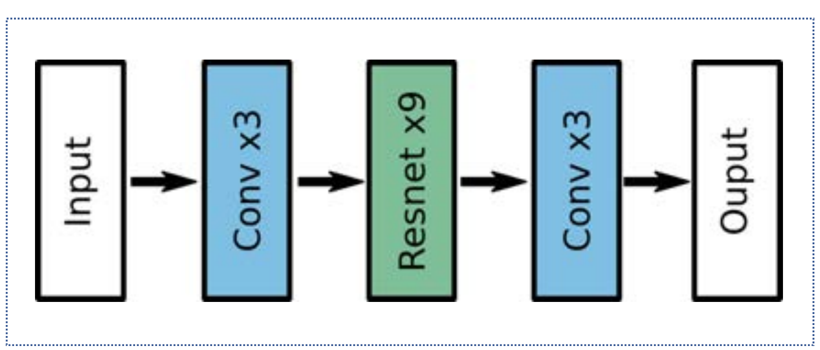




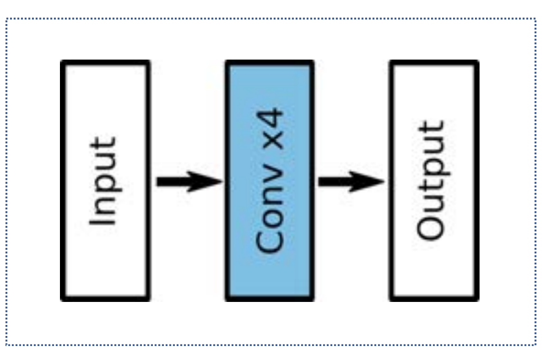
Network Architecture



Generator

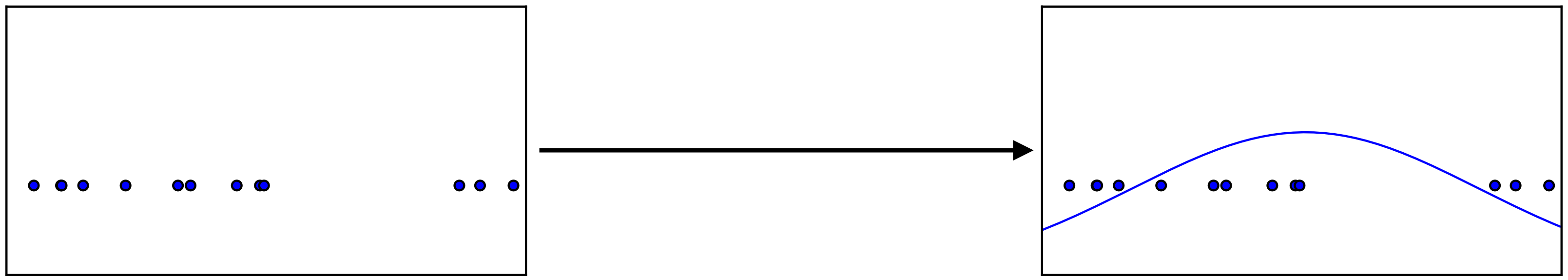


Discriminator

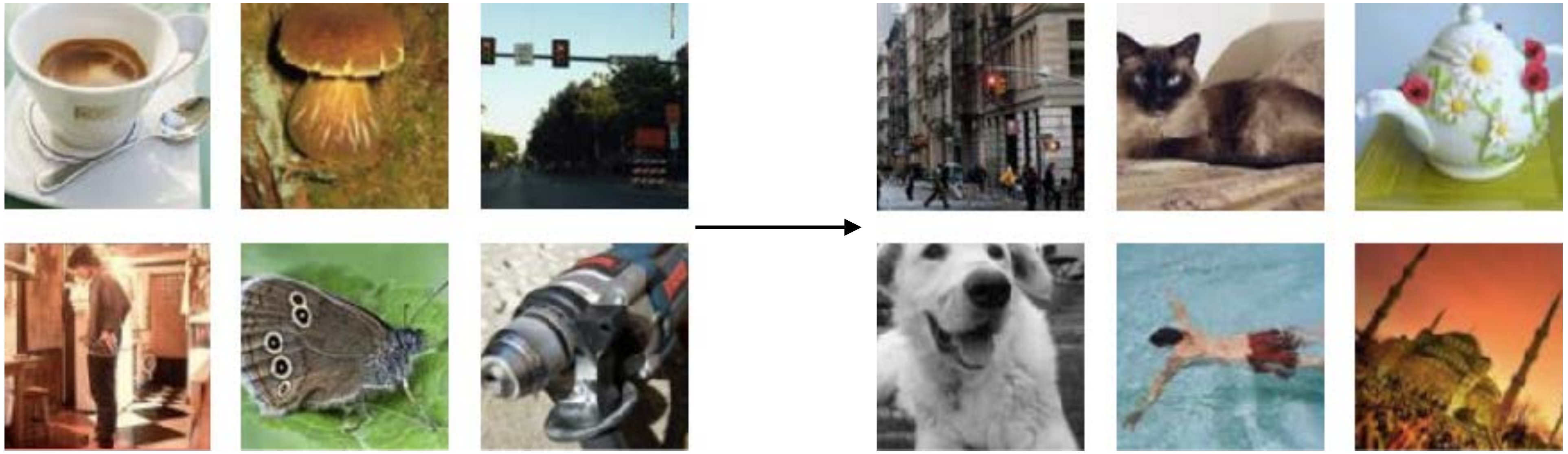


What is a Generative Model?

- Density estimation

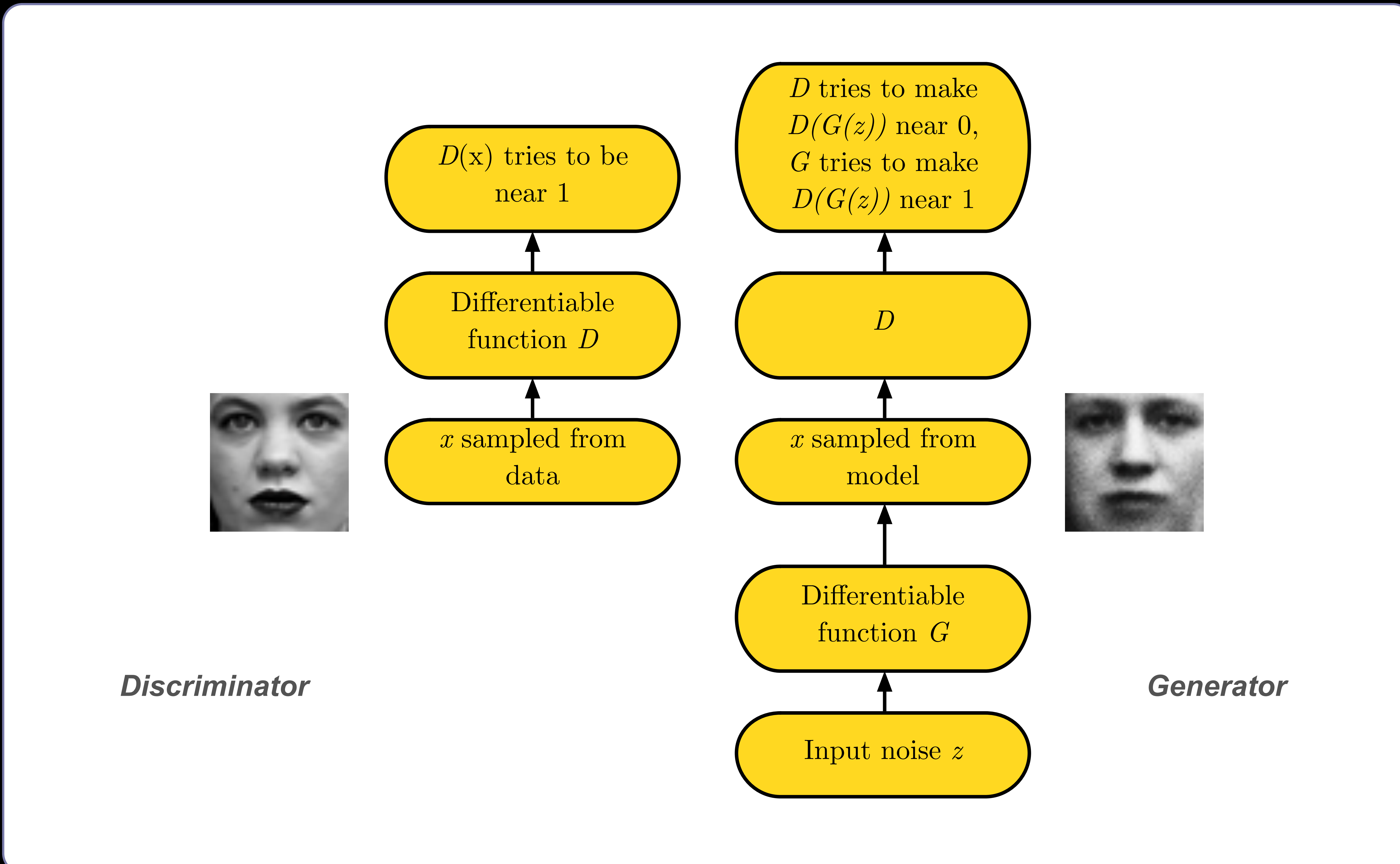


- Sample generation

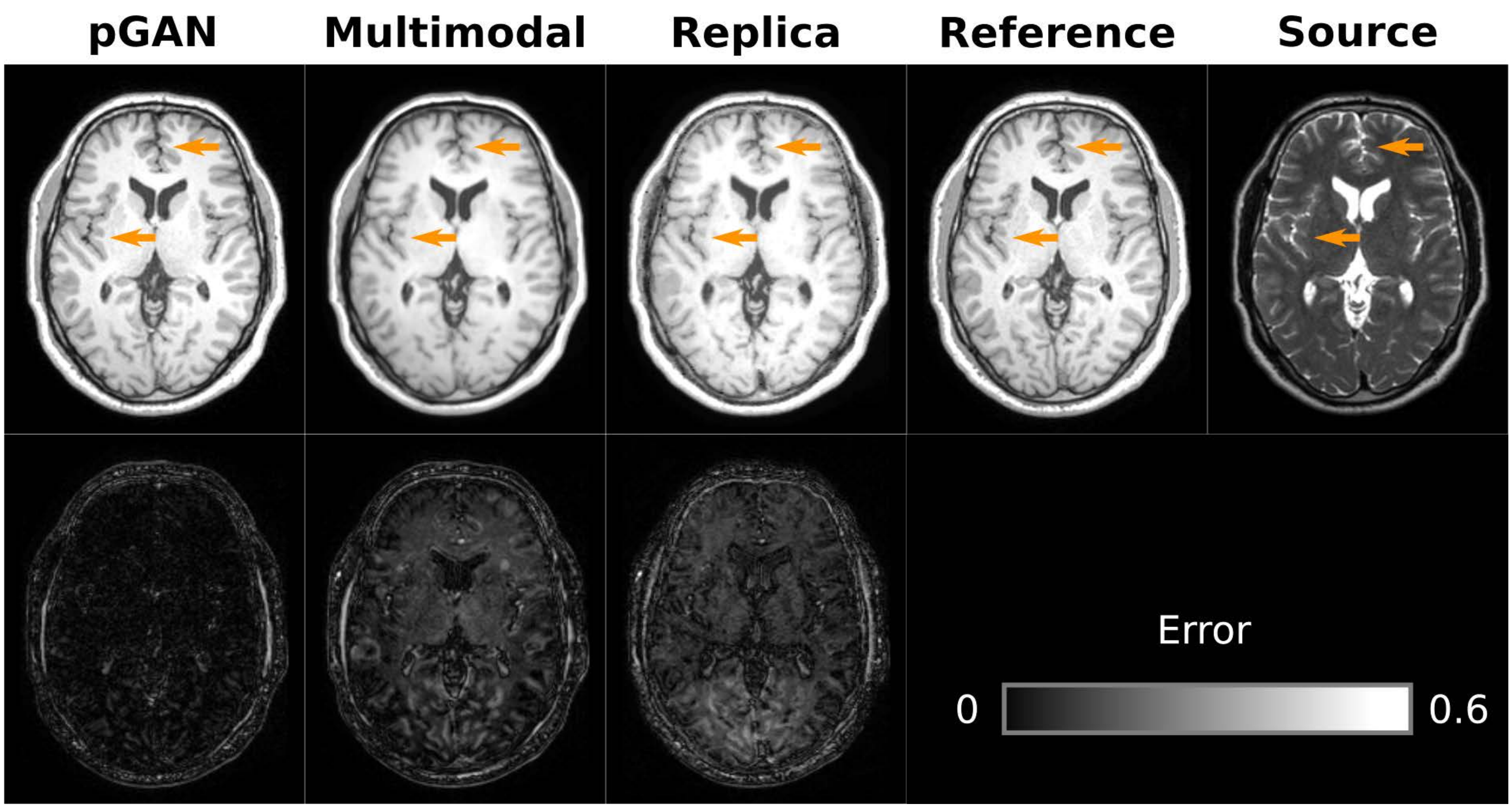


Training examples Model samples

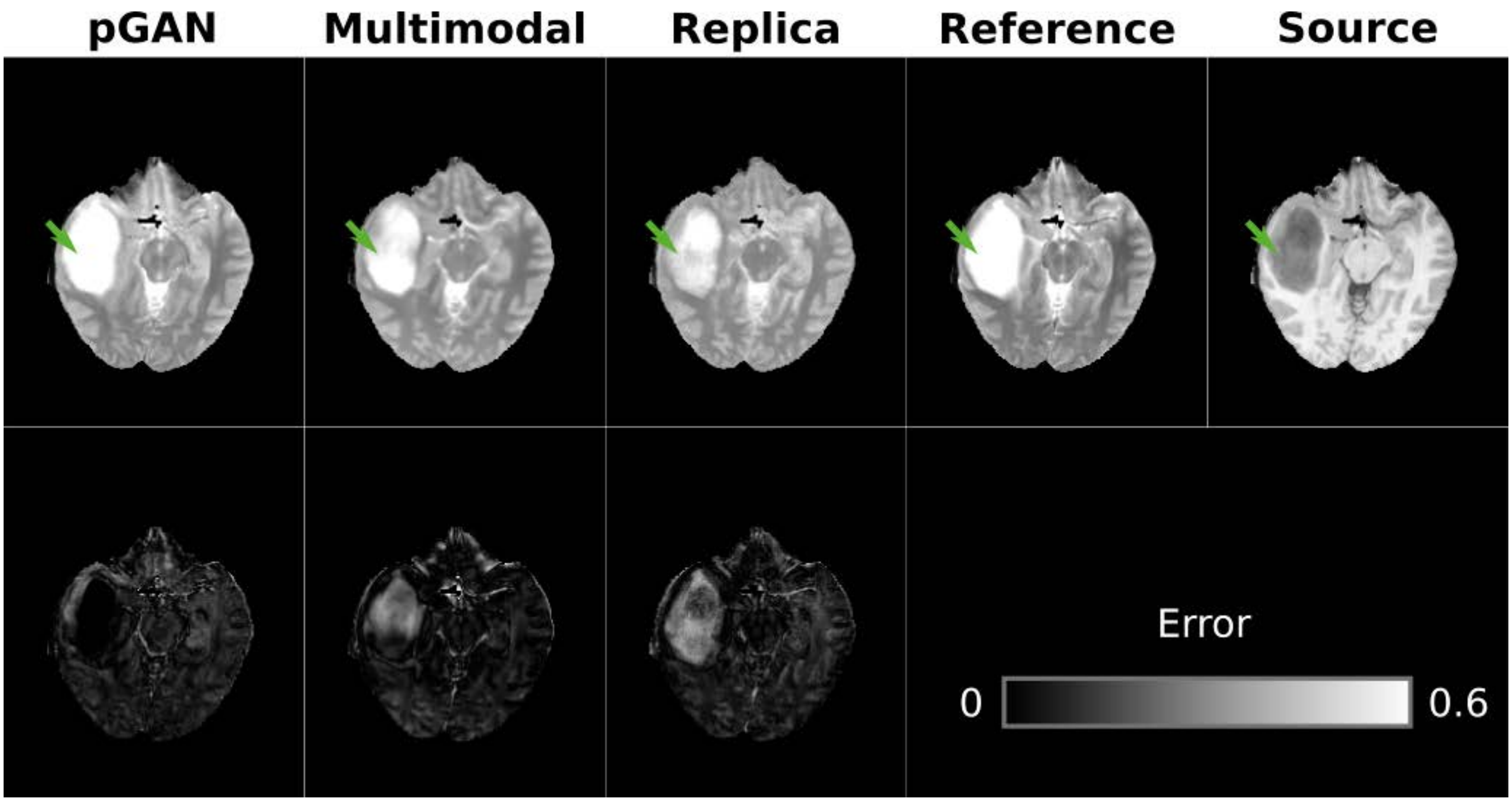
Generative Adversarial Network (GAN)



Examples: Synthesizing Missing Images



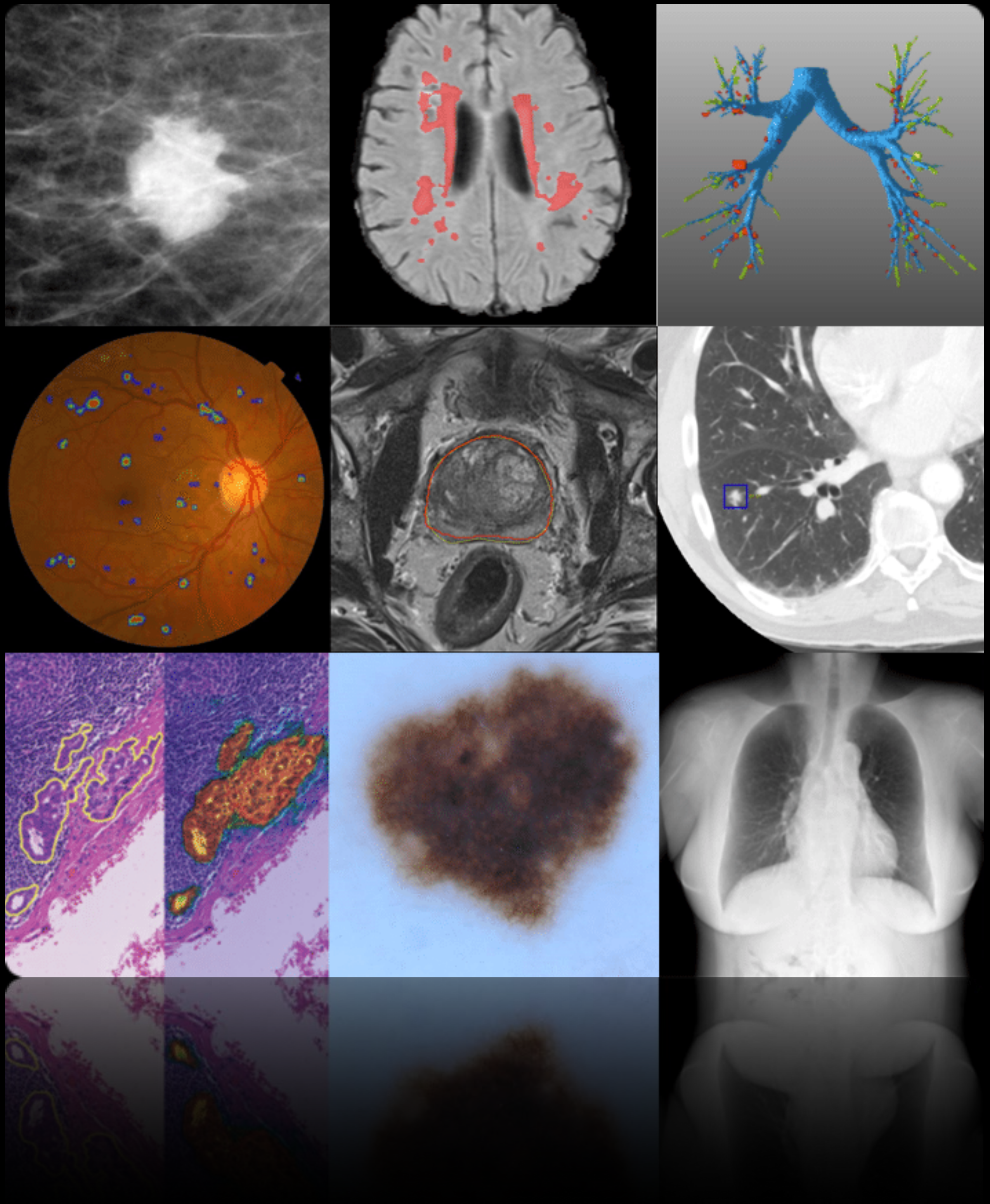
Examples: Synthesizing Missing Images



Future Outlook:

DL Tasks:

- Image Reconstruction
- Image Synthesis
- Data Acquisition
- Image Segmentation
- Tumor/Lesion Detection
- Disease Grading
- Monitoring
- Treatment Planning



Acknowledgment

Slide Credits

- Book: Principles of Neural Science
- Book: Medical Imaging Signals and Systems
- Lecture: Pascal Vincent
- Lecture: David Sontag
- Lecture: Ulas Bagci
- Lecture: Emine U Saritas
- Paper: Jarret, CVPR 2009
- Paper: Rueckert, arXiv 2019
- Paper: Waldrop, PNAS 2019
- Paper: Krizhevsky, NIPS 2012
- Paper: Jiang, SVN 2017
- Site: asimovinstitute.org
- Site: vinodsblog.com
- Site: doi.org/10.1016/j.media.2017.07.005