AIM 2: Artificial Intelligence in Medicine II

Harvard - BMIF 203 and BMI 702, Spring 2025

Lecture 5: Medical Imaging I

HARVARD
MEDICAL SCHOOL

Today’s Lecture outline

1. Overview of Medical Imaging and Basic AI Tasks
2. Convolutional Neural Networks (CNNs) for Medical Imaging
3. Segmentation in Medical Imaging – Focus on U-Net
4. Applying CNNs to Biomedical Segmentation & Future Directions

Overview of Medical Imaging & Basic AI Tasks

Introduction to medical imaging modalities

- Radiology (X-ray, CT, MRI)
- Oncology imaging (PET scans, specialized MRI for tumor detection)
- Pathology (digital slides)
- Ultrasound, endoscopy, and other modalities

Basic AI tasks in medical imaging

- Classification (detection of disease)
- Regression (e.g., lesion size or tumor volume)
- Segmentation (delineating tumors, organs, or structures)
- Registration (aligning structures between 2 different images)
- Enhancement (denoising, artifact removal, augmentation)

Importance of Medical Imaging in Clinical Practice

High Utilization in Healthcare

- Over **4.2 billion** diagnostic medical imaging procedures performed globally each year (Radiology estimate)
- In the US alone, ~**691 million** exams are performed annually, typically from CT scans, conventional radiology, dental radiography, nuclear medicine

Impact on Diagnosis & Treatment

- Critical for cancer detection, surgical planning, chronic disease management
- Radiology drives ~80% of hospital diagnoses (stat often cited by radiology organizations)

Challenges

- Huge data volume → Necessitates automation and AI
- Variability in acquisition, reconstruction parameters

The Physics of Medical Imaging

How images are formed (broad principle):

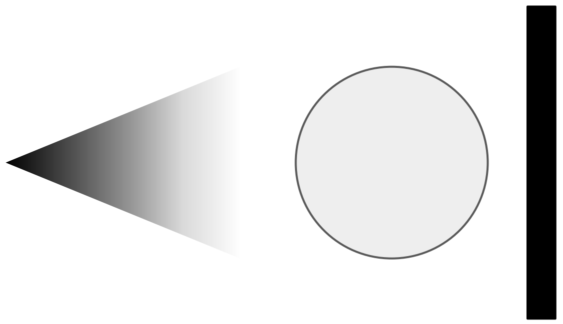
- Emission or transmission of a wave (electromagnetic or acoustic)
- Detectors measure wave attenuation or reflection/scattering to reconstruct an image

Key ideas in mathematics/physics:

- **Inverse problem**: Reconstructing internal structure from measured signals
- **Modalities differ** by type of wave (X-rays, radiofrequency for MRI, sound waves for ultrasound, positrons for PET)

Note: These fundamental physics principles underlie all imaging approaches

$$I_{detected} = I_{emitted} \cdot \exp(-\mu\, d) \text{ (exponential attenuation)}$$



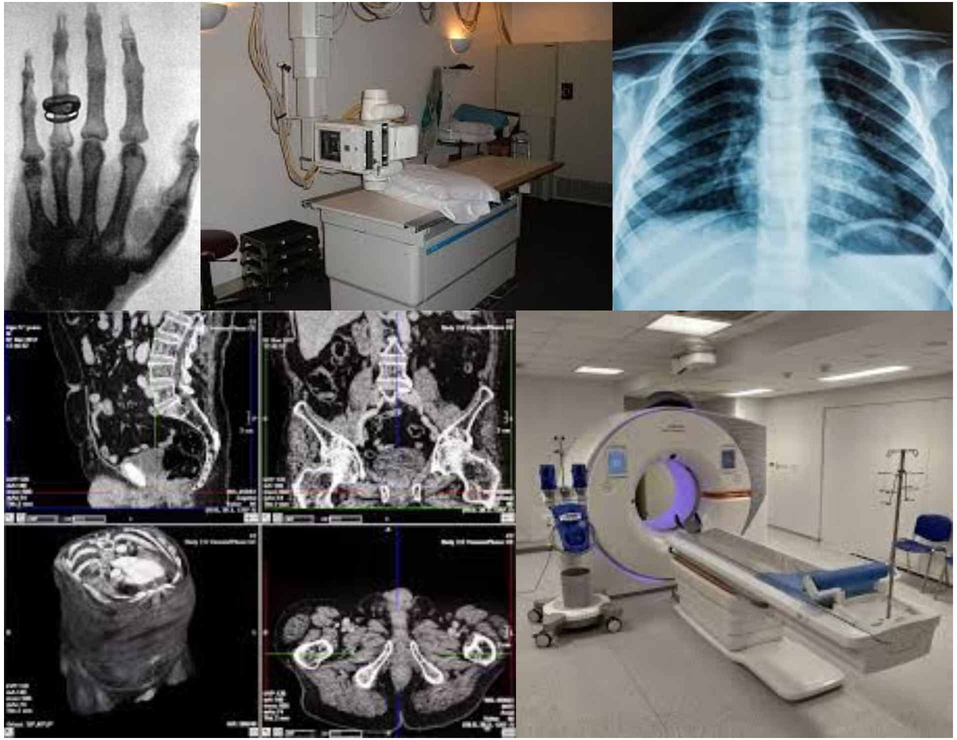
Radiology – X-ray, CT, MRI, Ultrasound

X-ray

- 2D projection imaging using X-ray photons
- Attenuation depends on tissue density
- Applications: Chest radiographs, bone fractures

Computed Tomography (CT)

- Multiple X-ray projections from different angles
- Reconstructed via *Radon transform* or *filtered back-projection*
- Generates 3D volumetric data



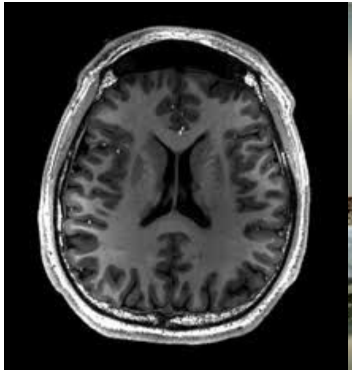
Radiology – X-ray, CT, MRI, Ultrasound

Magnetic Resonance Imaging (MRI)

- Manipulates proton spin alignments via strong magnetic fields & RF pulses
- Signal measured in k-space, reconstructed via *inverse Fourier transform*
- Good soft-tissue contrast

Ultrasound

- Uses high-frequency sound waves, reflection captured by a transducer
- Real-time imaging, widely used for obstetrics, cardiac echo
- Safe (no ionizing radiation), but operator-dependent



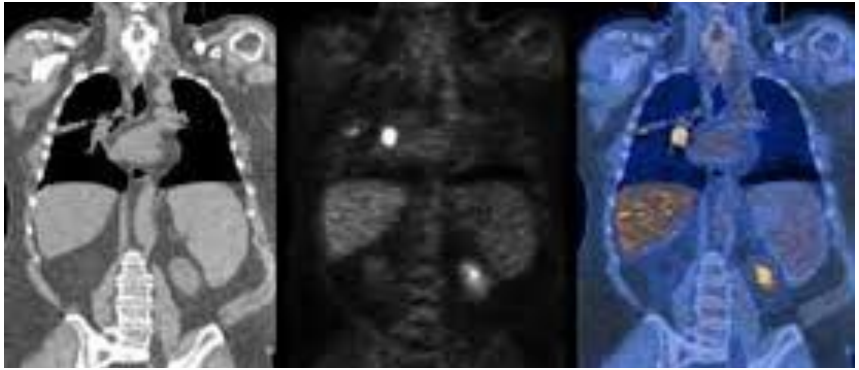
Oncology Imaging – PET & Specialized MRI

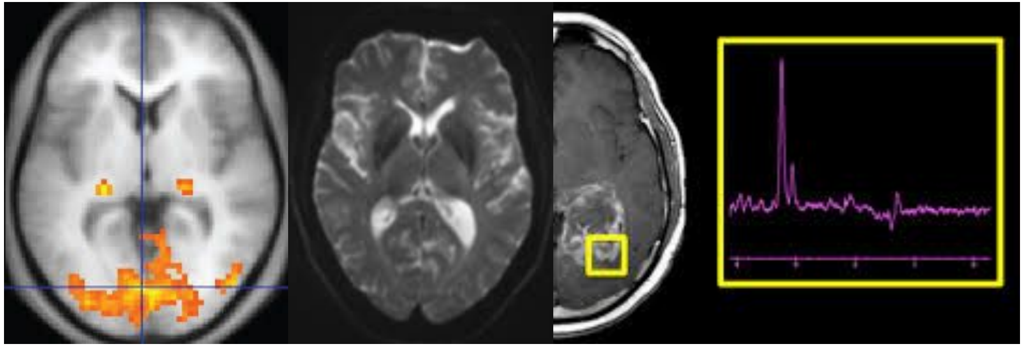
Positron Emission Tomography (PET)

- Inject radioactive tracer (e.g., FDG) that emits positrons
- Detect annihilation photons, reconstruct distribution of tracer uptake
- Highlights metabolic activity, commonly used for tumor detection and staging

Specialized MRI

- *fMRI* for brain function mapping
- *DWI/ADC* for tumor characterization and cellularity
- *MRS* (Magnetic Resonance Spectroscopy) for metabolic profiling





fMRI DWI MRS

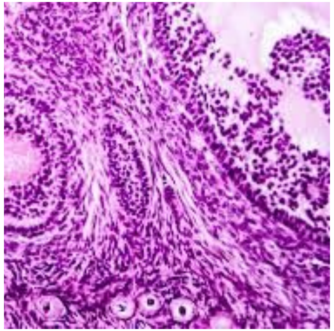
Pathology Imaging – Digital Slides & Advanced Stains

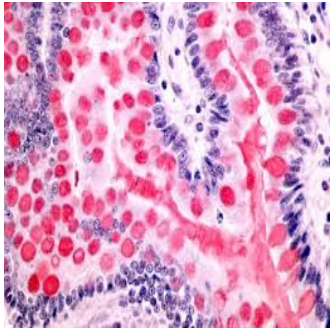
Digital Pathology

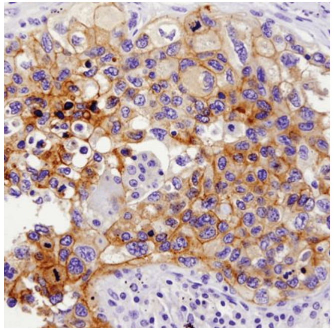
- High-resolution scanning of tissue slides (e.g., 40× magnification)
- Resulting images can be gigapixel-level

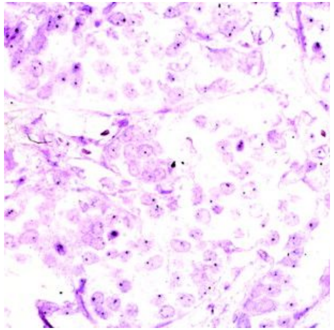
Types of Microscopy & Staining

- **H&E (Hematoxylin & Eosin)**: Standard stain for tissue morphology
- **Histochemical stains**: Highlight specific chemical components
- **Immunohistochemistry (IHC)**: Antibody-based staining for specific proteins
- **In situ hybridization**: Detect specific nucleic acid sequences
- **PCR-based assays**: Tissue-based molecular diagnostics (though not always “imaging,” can produce visually interpretable gels or signals)









Other Modalities & 4D Imaging

Endoscopy

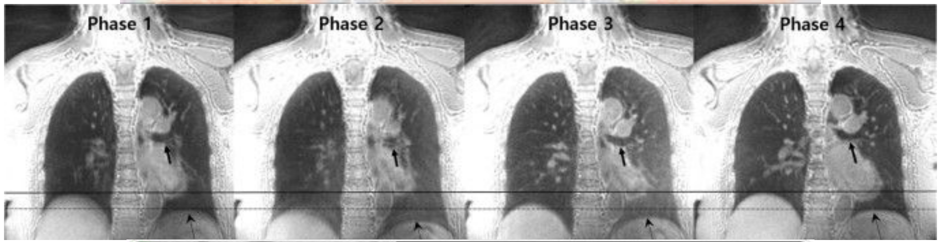
- Direct visualization using cameras inserted into body cavities (GI tract, lungs)
- Often recorded as video (temporal dimension)

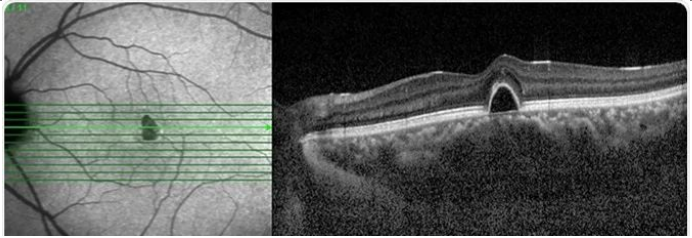
4D Imaging

- 3D + time, e.g., 4D CT in radiotherapy planning for moving organs (lungs)
- Real-time MRI sequences

Emerging or Specialized Modalities

- Optical coherence tomography (OCT), Photoacoustic imaging, etc.





Basic AI Tasks in Medical Imaging

Classification

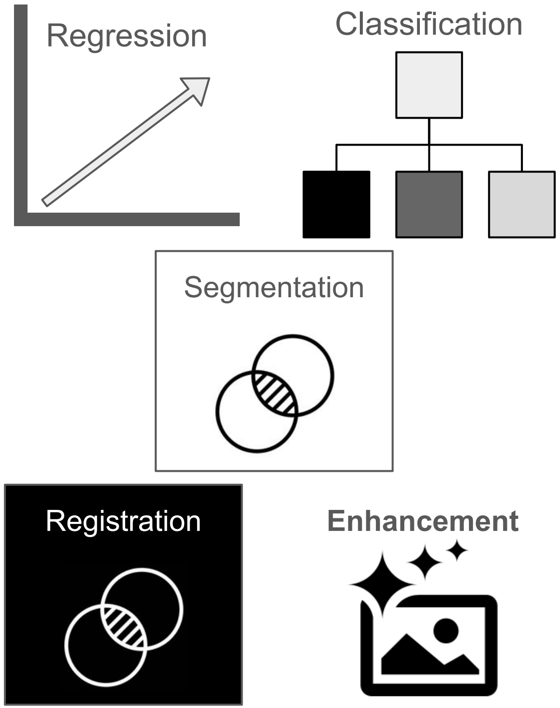
- Detect presence/absence of disease (e.g., tumor vs. normal)
- Multi-class scenarios (e.g., different tumor types)

Regression

- Predict continuous outcomes (e.g., tumor volume, disease progression)
- Often used in quantitative imaging biomarkers

Segmentation

- Delineate structures (tumors, organs) at pixel/voxel-level
- Essential for measuring size, shape, and location
- Forms basis for surgical or radiotherapy planning



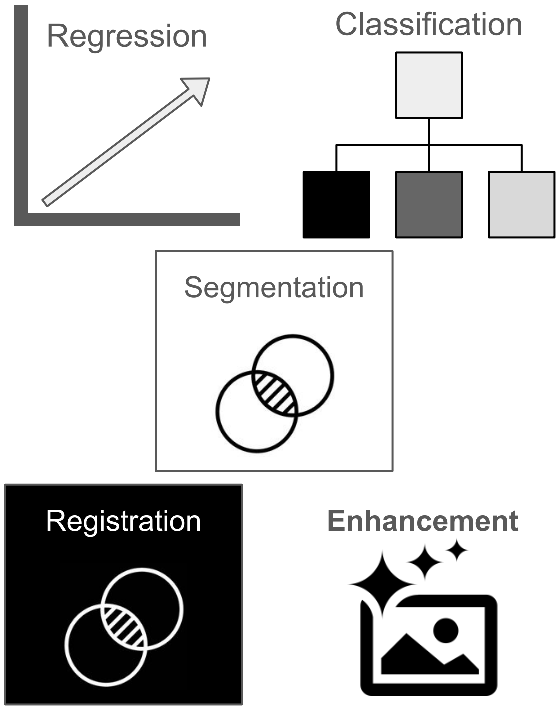
Basic AI Tasks in Medical Imaging

Registration

- Align images from the same or different modalities (e.g., CT-MRI fusion)
- Correct for patient movement and acquisition differences
- Essential for multimodal data fusion and longitudinal studies
- Enables precise anatomical mapping and improved diagnosis

Enhancement

- Improve image quality by reducing noise and artifacts
- Enhance contrast and resolution to reveal fine anatomical details
- Critical for revealing subtle pathologies and aiding diagnosis
- Often used as a preprocessing step for better downstream analysis



Q&A

Convolutional Neural Networks (CNNs) for Medical Imaging

CNN fundamentals

- Convolutional layers, filters/kernels, feature maps
- Pooling layers and their role
- Fully connected layers for classification tasks

Why CNNs are well-suited for medical imaging

- Local receptive fields and translation invariance
- Hierarchical feature extraction for complex patterns

Alternative Architectures

- **Vision Transformers (ViT)**: self-attention instead of convolutions, potential for capturing global context, but often data-intensive
- **Capsule Networks**: preserving spatial hierarchies and orientation information, potential advantages for complex anatomical structures

Convolutional Layers – The Core Operation

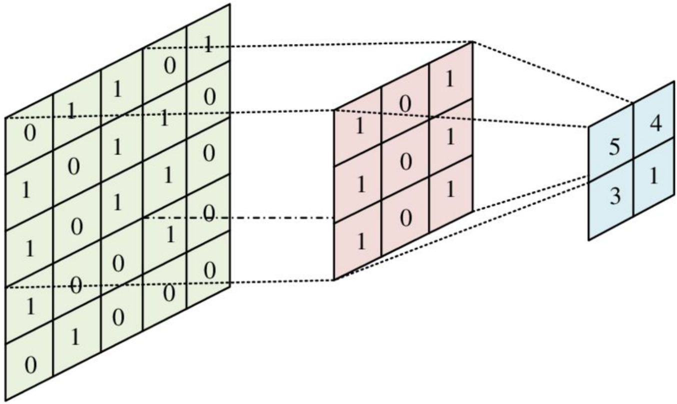
- **2D Convolution**
  - For an input feature map $\mathbf{X} \in \mathbb{R}^{H \times W}$ and a filter/kernel $\mathbf{K} \in \mathbb{R}^{k \times k}$, the convolution output at position (x, y) is:

$$(\mathbf{X} * \mathbf{K})(x, y) = \sum_{i=0}^{k-1} \sum_{j=0}^{k-1} \mathbf{X}(x+i, y+j)\, \mathbf{K}(i, j) + b$$

    - Often implemented as *cross-correlation* in practice, but the concept is similar.
  - **Stride & Padding**:
    - Stride s: controls how the filter steps across the image.
    - Padding (e.g., zero-padding) preserves spatial dimensions.
- **Multiple Channels**
  - In practice, filters have depth matching the input's channel dimension:

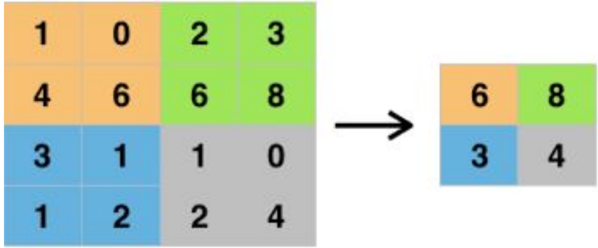
$$\mathbf{K} \in \mathbb{R}^{k \times k \times C_{\text{in}}}$$

  - Produces an output feature map with C_{out} channels, each learned via separate filters.

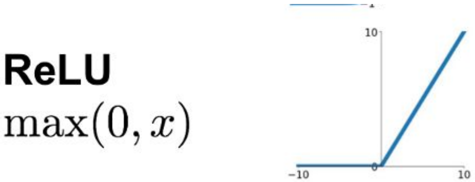


Pooling Layers & Nonlinearities

- **Pooling**
  - Reduces spatial dimensions to achieve translation invariance.
  - Common operations: **max pooling** or **average pooling** with kernel size p.
  - Example: $\mathrm{MaxPool}(2 \times 2)$ halves both height and width.



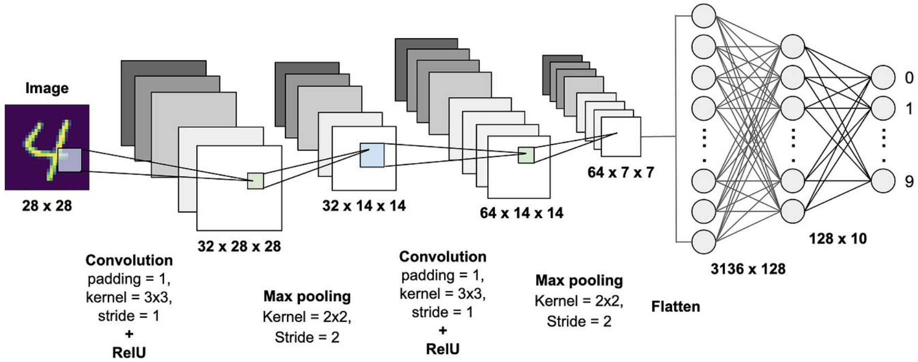
- **Nonlinear Activations**
  - Typically **ReLU**: $\sigma(z) = \max(0, z)$.
  - Other variants: Leaky ReLU, ELU, etc.



Fully Connected Layers & CNN Architectures

Transition to Dense Layers

- After repeated convolution + pooling, feature maps are flattened into a vector.
- Fed into one or more **fully connected (FC)** layers for classification/regression.
- Parameter count in FC layers can be large if feature maps are not sufficiently downsampled.



Example Architectures

- Classic CNNs: LeNet, AlexNet, VGG
- Deeper CNNs: ResNet (skip connections), DenseNet (dense connections)

Why CNNs Are Well-Suited for Medical Imaging

Local Receptive Fields & Translation Equivariance

- Early layers learn low-level edges/textures (helpful for subtle tissue boundaries).
- Convolutions treat local neighborhoods the same across the image (translational invariance).

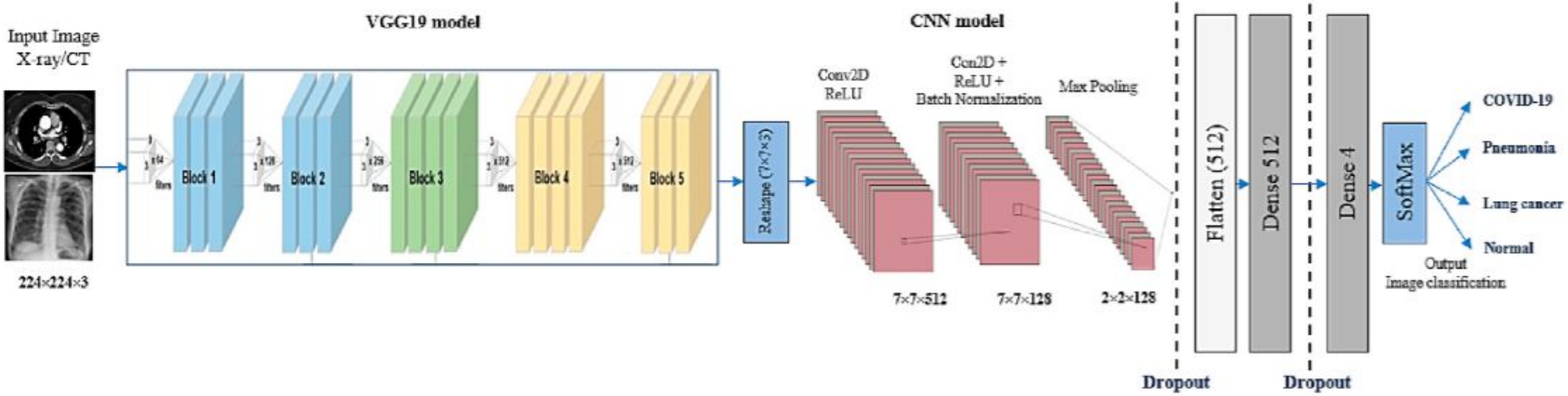
Hierarchical Feature Extraction

- Increasing abstraction: edges → textures → organs/pathologies.
- Large images (e.g., high-resolution scans) can be handled in patches or via downsampling.

Data Efficiency & Transfer Learning

- Pretrained networks on natural images can sometimes be fine-tuned for medical tasks.
- Data augmentation crucial for relatively small medical datasets.

CNNs for Classification

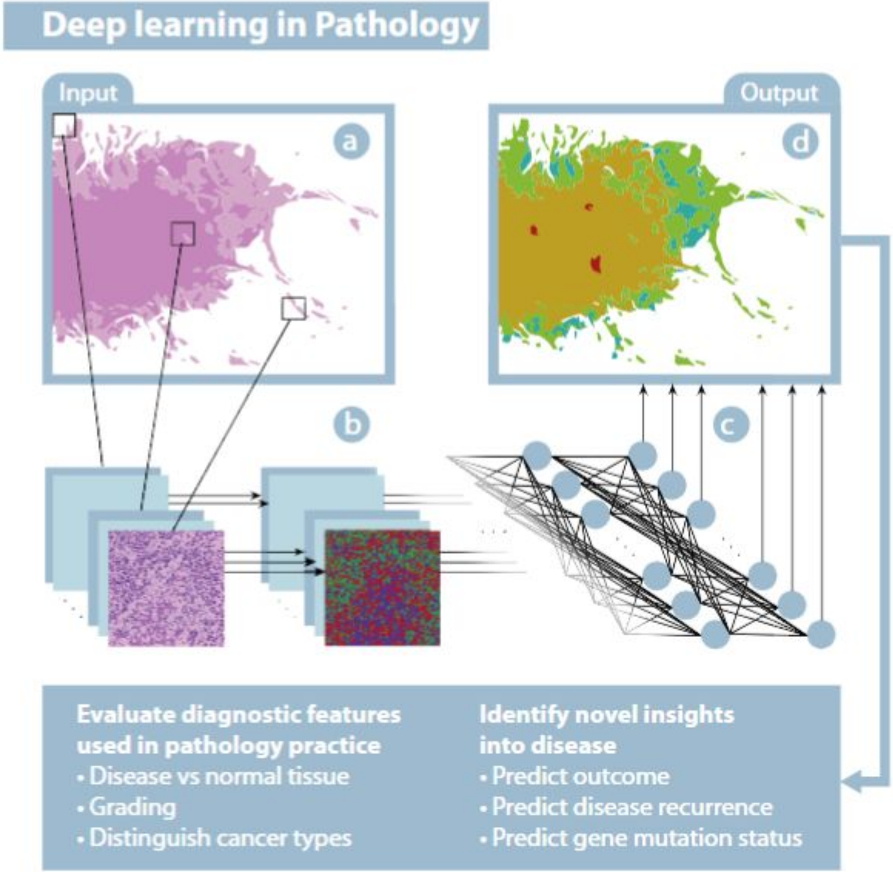


Evaluation metrics for the different models.

Models	Loss	TP	FP	TN	FN	ACC	Recall	PPV	SPC	NPV	F1-Score	MCC	AUC
VGG19+CNN	0.3280	251	4	764	5	**98.05**	**98.05**	**98.43**	**99.5**	**99.3**	**98.24**	**97.7**	**99.66**
ResNet152V2	0.1693	244	12	756	12	95.31	95.31	95.31	98.4	98.4	95.31	93.8	99.17
ResNet152V2+GRU	**0.1350**	246	10	758	10	96.09	96.09	96.06	98.7	98.7	96.09	94.8	99.34
ResNet152V2+Bi-GRU	0.2554	477	34	1502	35	93.36	93.16	93.35	97.8	97.8	93.26	91.1	98.44

Ibrahim, D. M., Elshennawy, N. M., & Sarhan, A. M. (2021). Deep-chest: Multi-classification deep learning model for diagnosing COVID-19, pneumonia, and lung cancer chest diseases. *Computers in biology and medicine*, *132*, 104348.

CNNs for Regression



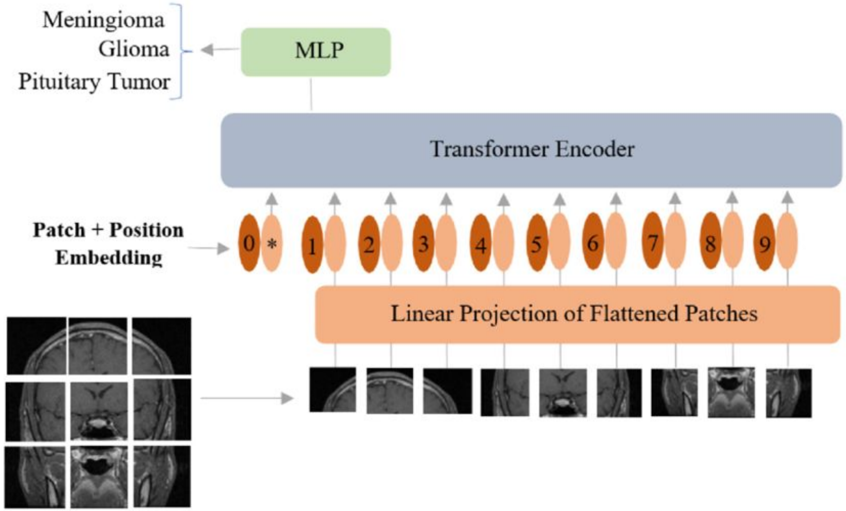
Acs, B., Rantalainen, M., & Hartman, J. (2020). Artificial intelligence as the next step towards precision pathology. *Journal of internal medicine*, *288*(1), 62-81.

Alternative Architectures I – Vision Transformers (ViT)

- **Motivation**: Move from convolution-based local receptive fields to global self-attention.
- **Self-Attention Mechanism**
  - Given Queries ($\mathbf{Q}$), Keys ($\mathbf{K}$), Values ($\mathbf{V}$) of dimension d:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d}}\right)\mathbf{V}$$

  - Captures long-range dependencies without explicit convolution.
- **ViT Basics**
  - Images split into patches, linearly embedded, then processed by transformer blocks.
  - Potential to capture global context better than CNNs.
  - **Drawback**: Often requires large datasets or heavy pretraining; can be data-hungry.



Tummala, S., Kadry, S., Bukhari, S. A. C., & Rauf, H. T. (2022). Classification of brain tumor from magnetic resonance imaging using vision transformers ensembling. *Current Oncology*, *29*(10), 7498-7511.

Alternative Architectures II – Capsule Networks

Concept: Retain spatial hierarchies & object pose in feature vectors.

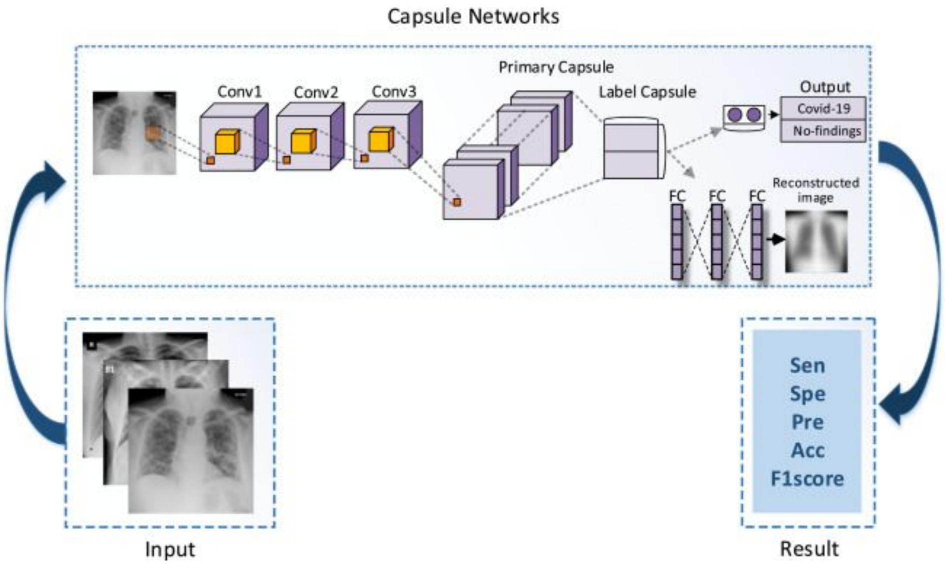
Capsule: A set of neurons whose output is a vector (or small matrix), encoding both the presence and the parameters (pose, orientation) of a feature.

Dynamic Routing (Sabour et al., *NIPS*, 2017)

- Iteratively adjusts "routing coefficients" between lower-level and higher-level capsules.
- Aims to preserve important spatial relationships that might get lost in CNN pooling.

Potential for Medical Imaging

- Detailed structural nuances (organ shape, orientation) are crucial.
- Still less common than CNNs in clinical practice, but a promising research direction.



Toraman, S., Alakus, T. B., & Turkoglu, I. (2020). Convolutional capsnet: A novel artificial neural network approach to detect COVID-19 disease from X-ray images using capsule networks. *Chaos, Solitons & Fractals*, *140*, 110122.

5 min. Break

Segmentation in Medical Imaging – Focus on U-Net

Importance of segmentation tasks

- Common use cases (tumor segmentation, organ delineation)
- Impact on treatment planning, diagnostics, and surgery

U-Net architecture

- Encoder-decoder structure with skip connections
- Advantages for medical image segmentation (handling fewer images, robust feature localization)
- Reference: Ronneberger et al. (2015)

Importance of Segmentation Tasks

Key Role in Diagnostics & Treatment

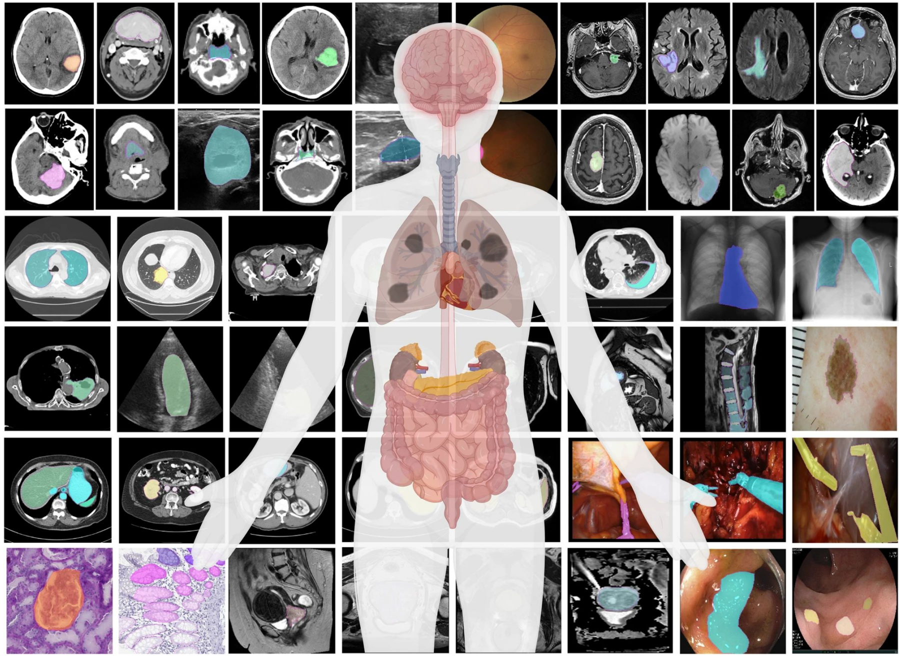
- Tumor boundary detection for radiation therapy
- Organ delineation for surgical planning
- Lesion quantification for disease progression

Granular Analysis

- Pixel/voxel-level detail → more precise than classification or bounding boxes
- Enables volumetric and shape analyses

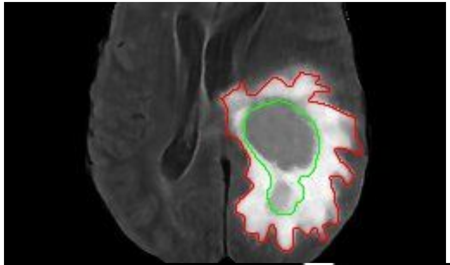
Clinical Impact

- Affects prognosis and treatment strategies (e.g., tumor growth rates)
- Provides consistent, reproducible measurements vs. manual outlining



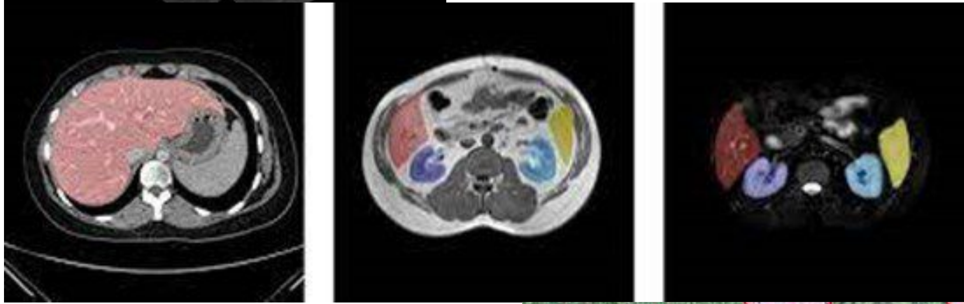
Ma, J., He, Y., Li, F. *et al*. Segment anything in medical images. *Nat Commun* 15, 654 (2024). https://doi.org/10.1038/s41467-024-44824-z

Common Use Cases



Tumor Segmentation

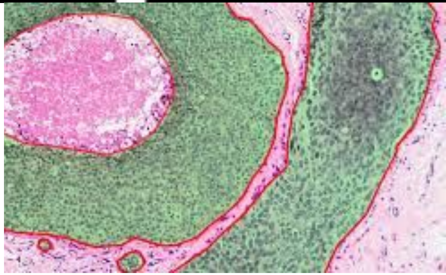
- Brain tumors (gliomas, metastases)
- Lung nodules, liver lesions, breast cancer



Organ Delineation

- Heart chambers in cardiac MRI
- Liver, kidneys, prostate in CT/MRI

Microscopic Pathology Segmentation



- Nuclei, glands, or other histological structures in whole-slide images

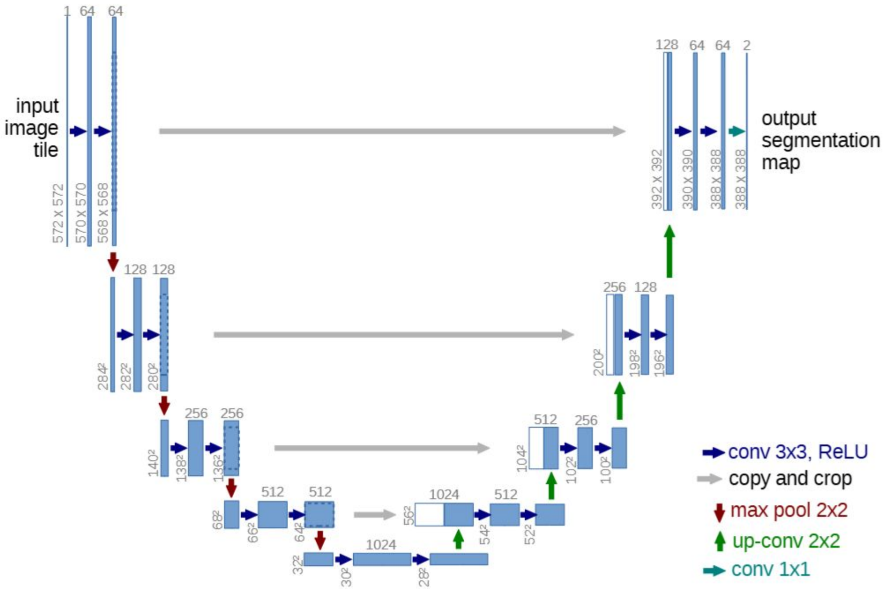
U-Net Architecture Overview

Historical Context

- Proposed by Ronneberger et al. (2015), originally for biomedical microscopy
- Achieved top performance on the ISBI Cell Tracking Challenge

High-Level Structure

- **Encoder**: Downsampling path for context capturing (similar to CNN classification backbones)
- **Decoder**: Upsampling path for precise localization
- **Skip Connections**: Transfer high-resolution features from encoder to decoder



Encoder-Decoder Structure & Skip Connections

Encoder Path (Left side)

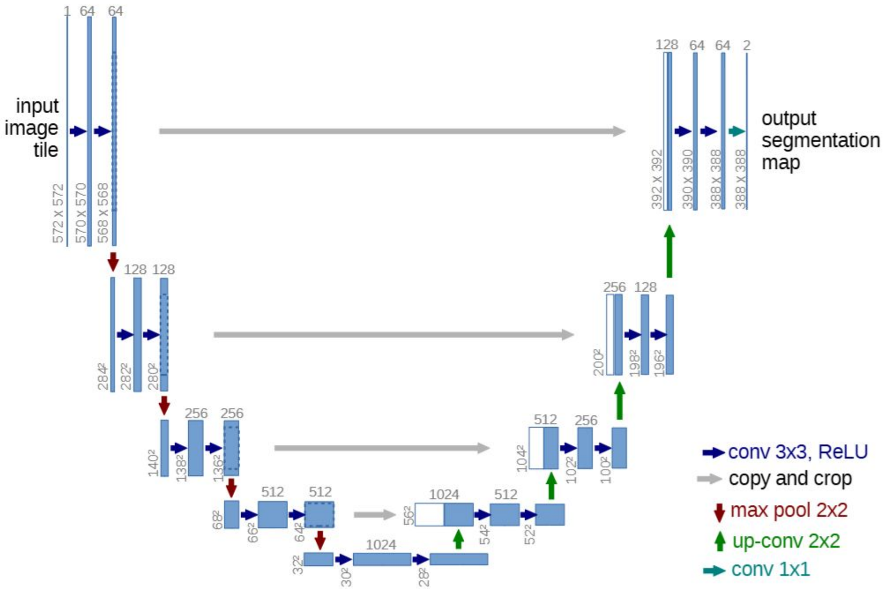
- Series of **convolution + ReLU + pooling**
- Each downsampling roughly **doubles** the number of feature maps, halves the spatial resolution

Bottleneck

- Lowest resolution; deepest features (semantic information)

Decoder Path (Right side)

- **Transposed convolutions** or up-convolutions to increase spatial dimension
- Merges (concatenation) with corresponding features from the encoder via skip connections
- Gradually refines the segmentation mask at higher resolution



Advantages in Medical Image Segmentation

Handling Fewer Images

- U-Net can be trained effectively on relatively small datasets (typical in medical imaging)
- Use of heavy data augmentation is standard

Robust Feature Localization

- Skip connections preserve spatial information lost by pooling
- Helps differentiate fine boundaries (tumor edges, organ interfaces)

2D vs. 3D U-Net Variants

- **2D**: Processes slices independently, good if GPU memory is limited
- **3D**: Captures volumetric context but more memory-intensive

Loss Functions & Evaluation Metrics (U-Net Context)

- **Cross-Entropy (CE) Loss**
  - Pixel-wise classification:

$$\mathcal{L}_{\mathrm{CE}} = -\frac{1}{N}\sum_{i=1}^{N}\left[g_i \ln p_i + (1-g_i)\ln(1-p_i)\right],$$

  where g_i is the ground-truth label (0 or 1), p_i is the predicted probability.
- **Dice Coefficient & Dice Loss**
  - **Dice Coefficient** for predicted mask P and ground-truth mask G:

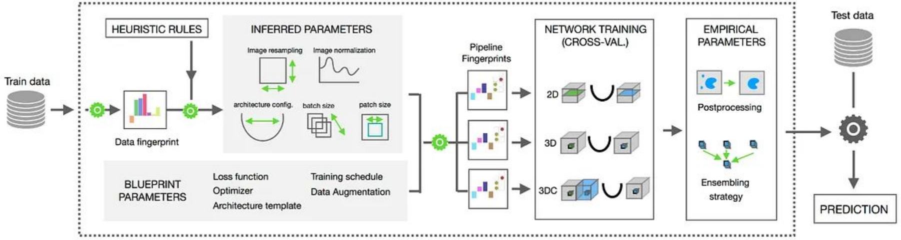
$$\mathrm{Dice}(P,G) = \frac{2\sum_i (p_i g_i)}{\sum_i p_i + \sum_i g_i}.$$

  - **Dice Loss**: $\mathcal{L}_{\mathrm{Dice}} = 1 - \mathrm{Dice}(P,G)$.
  - Robust for class imbalance; used widely in organ/tumor segmentation
- **Composite Loss**
  - Often combine CE + Dice to balance region overlap and pixel-wise accuracy

State-of-the-Art Variants – nnU-Net & MedSAM

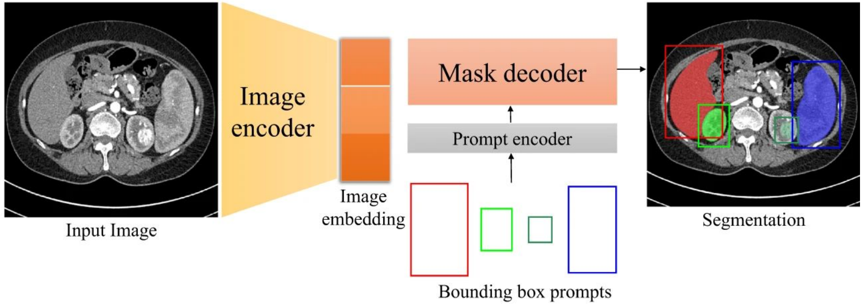
nnU-Net

- Self-configuring U-Net framework
- Automatically adapts architecture and hyperparameters
- Top performance in multiple segmentation challenges



MedSAM

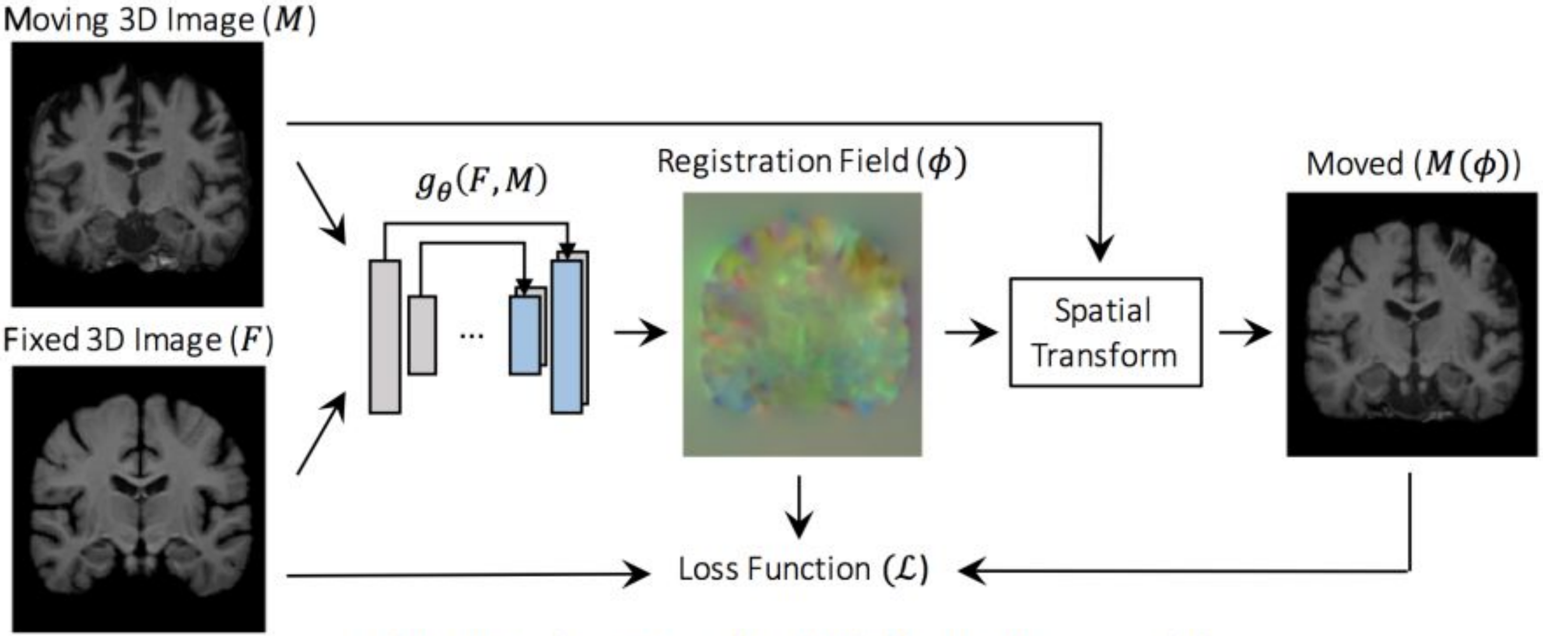
- Adapts “Segment Anything Model” to medical images
- Uses large-scale pretrained embeddings + prompt-based segmentation
- Offers few-shot or zero-shot capabilities for new tasks



Key Takeaway

- Both build on the U-Net paradigm with skip connections
- Ongoing improvements target automated tuning and broad generalizability

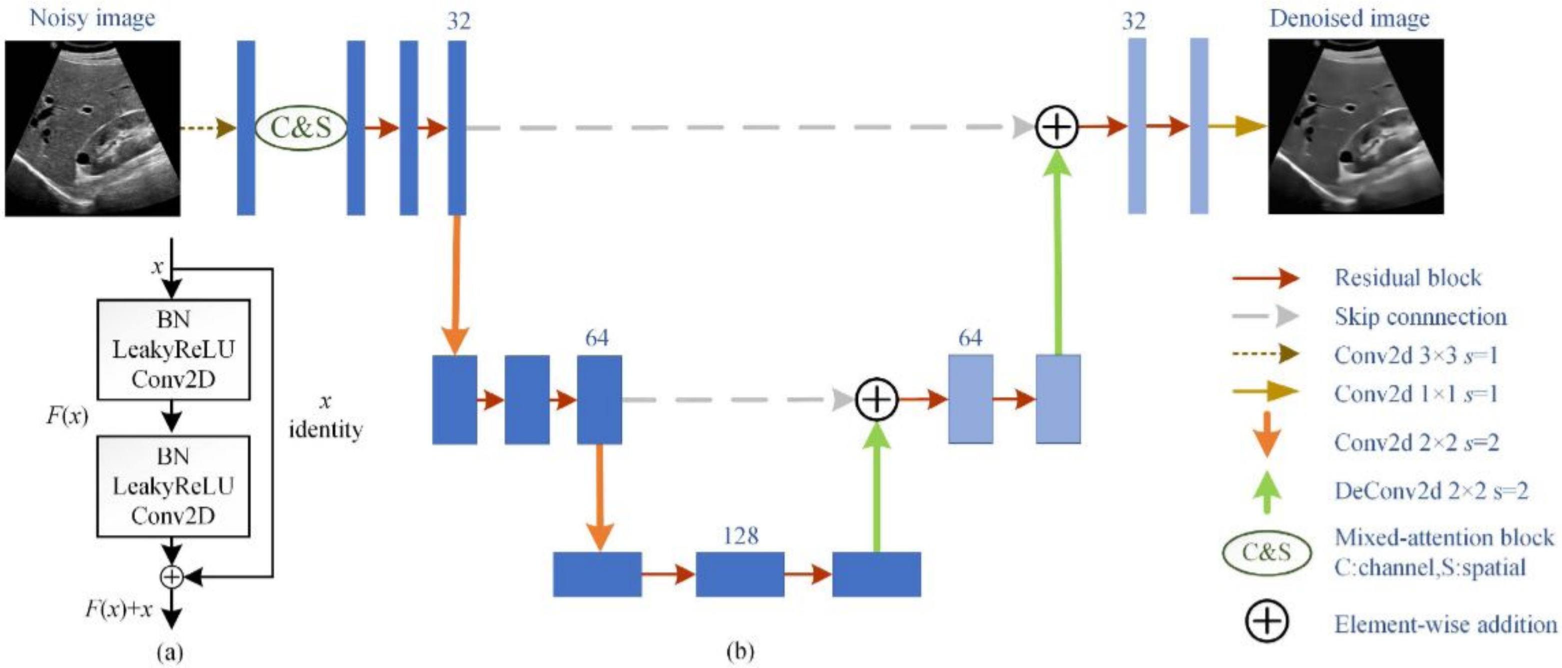
UNet for Image Registration



$$\mathcal{L}(F, M, \phi) = \mathcal{L}_{sim}(F, M(\phi)) + \lambda \mathcal{L}_{smooth}(\phi),$$

Guo, C. K. (2019). *Multi-modal image registration with unsupervised deep learning* (Doctoral dissertation, Massachusetts Institute of Technology).

UNet for Image Enhancement



Lan, Y., & Zhang, X. (2020). Real-time ultrasound image despeckling using mixed-attention mechanism based residual UNet. *IEEE Access*, *8*, 195327-195340.

Q&A

Applying CNNs to Biomedical Segmentation & Future Directions

Preprocessing and data preparation

- Data cleaning (artifact removal, normalization)
- Data augmentation (flips, rotations, intensity shifts)
- Handling class imbalance (sampling strategies, loss functions)

Training and evaluation strategies

- Train/validation/test splits, cross-validation
- Performance metrics (Dice coefficient, IoU, sensitivity, specificity)
- Visual inspection and clinician-in-the-loop for validation

Next Steps & Federated Learning Mention

- **Federated Learning** as a solution for data-sharing barriers among multiple hospitals/institutions.
- High-level benefits (privacy-preserving, larger effective dataset) and challenges (communication overhead, data heterogeneity).
- Encouragement to think about how this could improve generalizability across diverse patient populations.

Preprocessing & Data Preparation

Data Cleaning

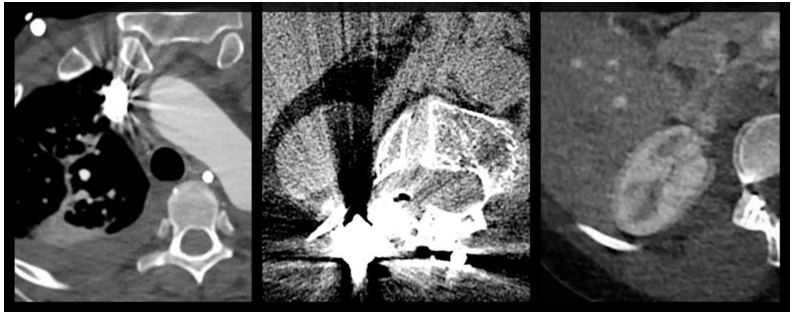
- Artifact removal (e.g., motion, noise)
- Normalization & standardization (intensity scaling)

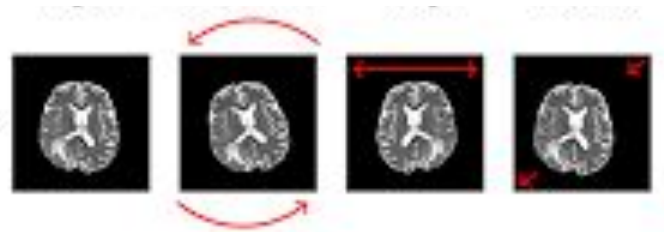
Data Augmentation

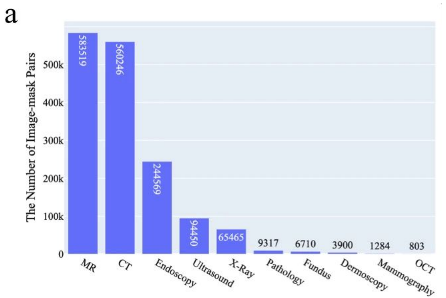
- Flips, rotations, elastic deformations
- Intensity shifts (brightness, contrast)

Handling Class Imbalance

- Oversampling/undersampling methods
- Loss functions (e.g., focal loss)







Training & Evaluation Strategies

Splitting Protocols

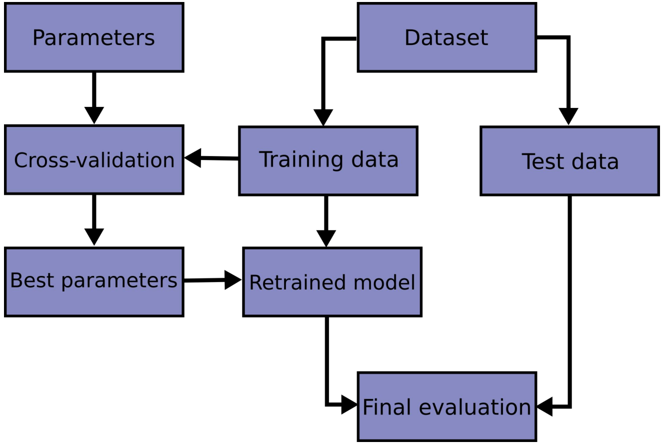
- Train/Val/Test splits
- k-Fold cross-validation for small datasets

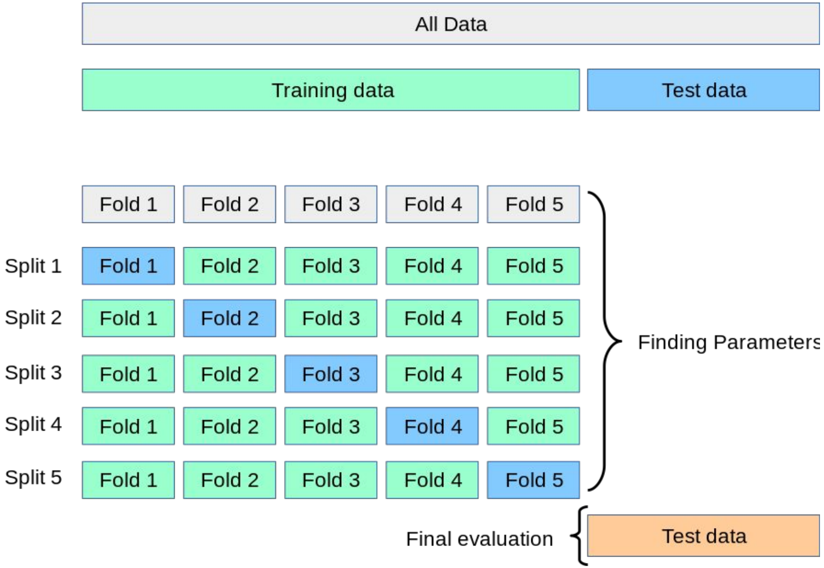
Metrics

- **Dice Coefficient** & **IoU** for segmentation overlap
- **Sensitivity/Specificity** for binary outcomes

Clinician-in-the-Loop

- Importance of visual inspection
- Iterative feedback for model refinement





https://scikit-learn.org/stable/modules/cross_validation.html

Federated Learning

Motivation

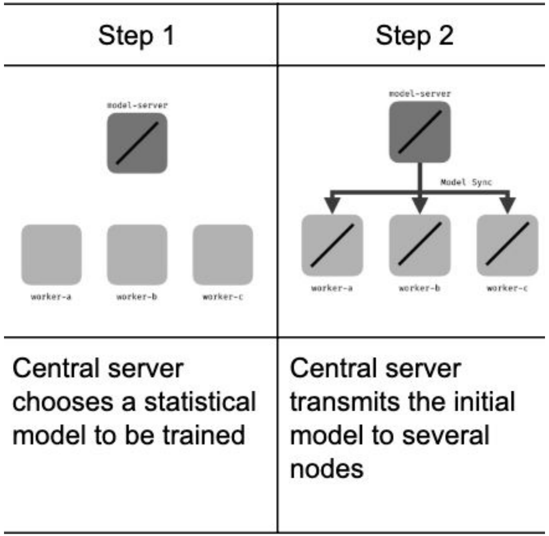
- Train collaboratively across hospitals without sharing raw data
- Larger effective dataset, privacy preserved

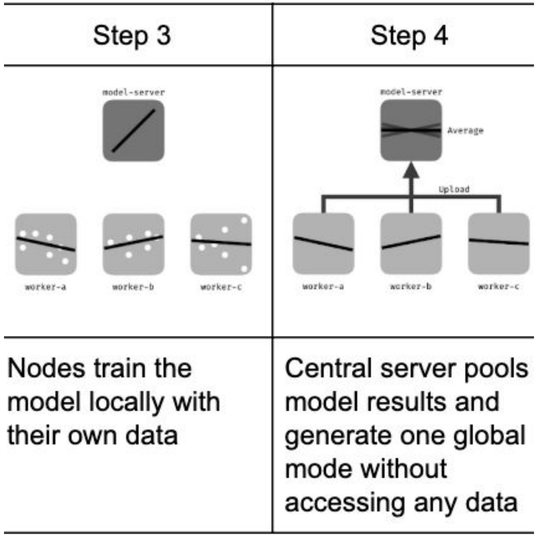
Challenges

- Communication overhead, model synchronization
- Data heterogeneity (different scanners, protocols)

Potential Impact

- Improved model generalizability
- Regulatory compliance (HIPAA, GDPR)





Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., ... & Zhao, S. (2021). Advances and open problems in federated learning. *Foundations and trends® in machine learning*, *14*(1–2), 1-210.

Federated Learning

- **Largest FL study in medical imaging → 71 sites, 6 continents**
- **6,314 cases** → Largest glioblastoma dataset
- **No data sharing** → Privacy-preserving model training

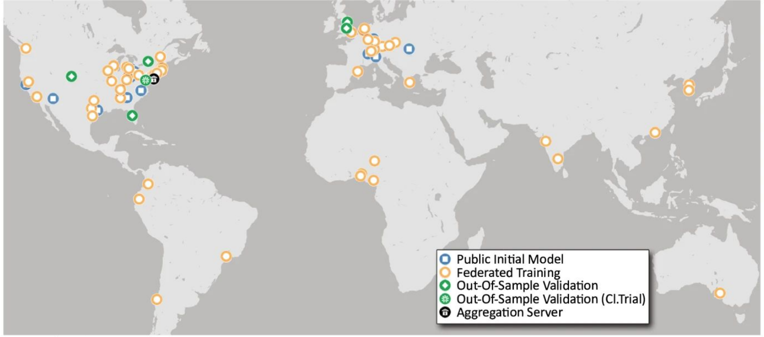
Key Results

- **+33% improvement** in surgically targetable tumor segmentation
- **+23% improvement** in complete tumor segmentation
- **Validated on:**
  - **Local site data (n = 1,043 cases)**
  - **Out-of-sample data (n = 518 cases)**

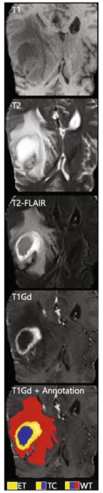
Impact

- **More diverse, generalizable AI models**
- **Public release of the consensus model**
- **New standard for multi-site AI training**

a

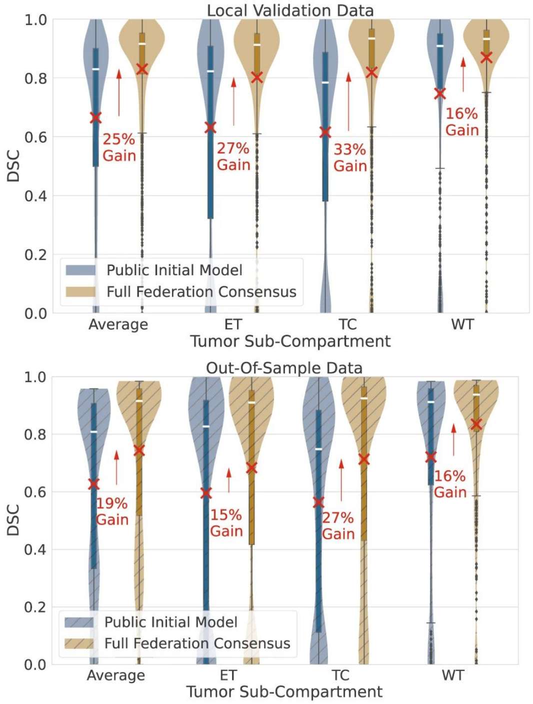


b

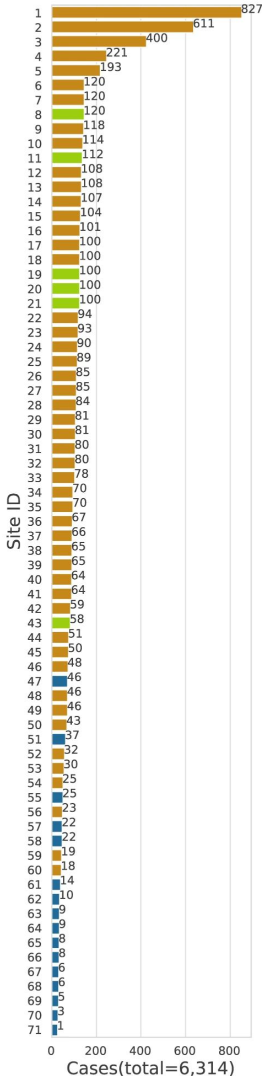


c

d



e



Pati, S., Baid, U., Edwards, B., Sheller, M., Wang, S. H., Reina, G. A., ... & Poisson, L. (2022). Federated learning enables big data for rare cancer boundary detection. *Nature communications*, *13*(1), 7346.

Conclusions & Key Takeaways

Medical Imaging Overview

- Diverse modalities (X-ray, CT, MRI, Ultrasound, Pathology)
- AI tasks: classification, regression, segmentation, registration, enhancement

Deep Learning Foundations

- CNNs: convolution, pooling, encoder-decoder structures (U-Net)
- Advanced methods: Vision Transformers, Capsule Networks

Segmentation & U-Net

- Critical for accurate delineation (tumors, organs)
- Skip connections, specialized loss functions (Dice)

Practical Considerations

- Data preprocessing, augmentation, handling imbalance
- Training strategies, metrics, clinician-in-the-loop validation

Looking Ahead

- Federated learning for distributed data
- Continued progress in robustness, explainability, and generalizability

Q&A