CAP 5516
Medical Image Computing
(Spring 2025)

Dr. Chen Chen
Associate Professor
Center for Research in Computer Vision (CRCV)
University of Central Florida
Office: HEC 221
Email: chen.chen@crcv.ucf.edu
Web: https://www.crcv.ucf.edu/chenchen/

Lecture 8: Medical Image Classification

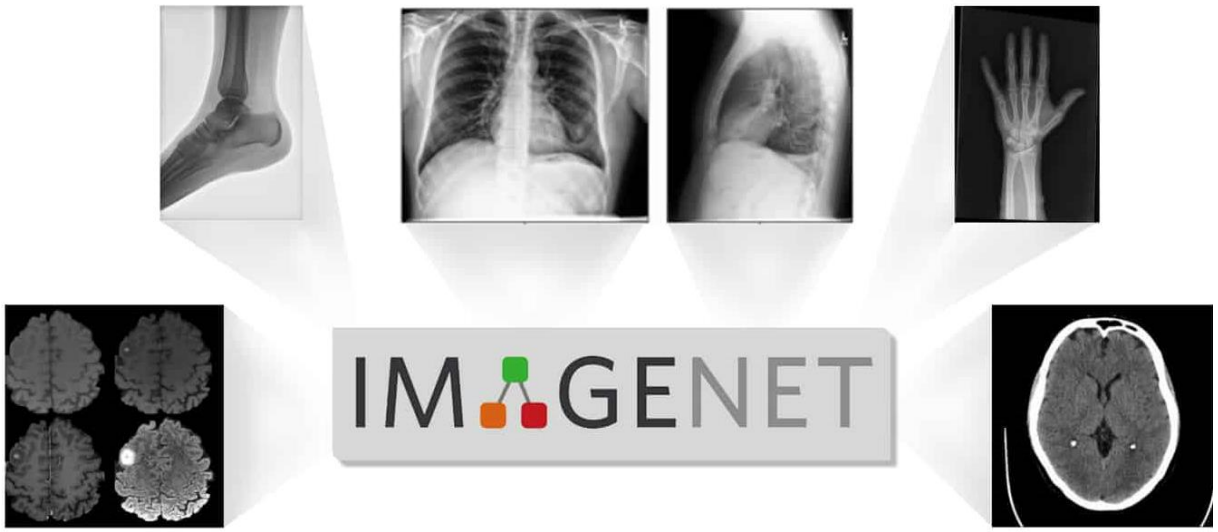


Image source: https://learnopencv.com/transfer-learning-for-medical-images/

Medical Image Classification

- Classification = assigning a label or category to an image (e.g., ‘normal’ vs. ‘abnormal’).”

- Classification is a fundamental building block in medical image analysis. While segmentation or object detection might be more complex tasks, classification remains critical for “yes/no” decisions (e.g., disease present/absent) or “which category” decisions (e.g., tumor type A vs. B).

Medical Image Classification: Why It Matters

- **Clinical Relevance**:
  - Aids in **early detection** of diseases (e.g., cancers, lung nodules).
  - Supports **clinical decision-making** (tumor subtype, treatment planning).

- **Broad Impact**:
  - Millions of imaging exams performed annually. For instance, an algorithm could flag suspicious chest X-rays for priority review in resource-limited settings.
  - High stakes: early intervention can **save lives**.

Data Pre-processing

- Image Normalization & Intensity Scaling

Why Normalize?

- Medical images have varying intensity distributions due to different acquisition settings.

Techniques:

- **Min-Max Scaling**: $x' = \frac{x - x_{min}}{x_{max} - x_{min}}$
- **Z-score Normalization**: $x' = \frac{x - \mu}{\sigma}$
- **Histogram Equalization** for contrast enhancement

Data Pre-processing

- Data augmentation

Why Augment?

- Overcomes data scarcity in medical imaging
- Reduces overfitting in deep learning models

Common Augmentations:

- **Affine Transformations** (Rotation, Scaling, Flipping)
- **Elastic Deformations** (Mimics real anatomical variations)
- **Contrast Adjustments** (Adaptive Histogram Equalization)
- **Noise Injection** (Gaussian Noise)
- **Mixup & CutMix** (blend images to generate synthetic variations)

Data Pre-processing

- MONAI

- https://docs.monai.io/en/0.5.2/highlights.html
- https://docs.monai.io/en/0.5.2/highlights.html#medical-image-data-i-o-processing-and-augmentation

Data Imbalance in Medical Image Classification

Data imbalance occurs when certain classes are underrepresented compared to others.

Why is this a problem?

- Models become biased toward majority classes.
- Poor generalization on minority classes.

Example:

- **Chest X-ray dataset**: 90% normal cases, only 10% pneumonia cases.
- **Brain tumor classification**: Glioblastoma (GBM) is much rarer than benign tumors.

Strategies to Handle Data Imbalance

1. Data-Level Solutions (Resampling & Augmentation)

- **Oversampling** (e.g., SMOTE for synthetic sample generation)
- **Undersampling** (reducing majority class examples)

- **Medical Image-Specific Augmentation**
  - Rotation, flipping, elastic deformation, intensity shifts
  - GAN-based synthetic image generation (e.g., StyleGAN, Diffusion Models)

Strategies to Handle Data Imbalance

2. Algorithm-Level Solutions

- **Cost-sensitive learning**: Assign higher loss weights to minority classes.
- **Class-balanced loss functions**: Focal Loss, Balanced Cross-Entropy

3. Hybrid Solutions

- **Self-Supervised Learning (SSL)**: Pretrain models on large unannotated datasets to learn general features before fine-tuning on small datasets.

Strategies to Handle Data Imbalance

- To make it clear, let's take an example of **pneumonia classification from chest X-rays**. We assume we have a dataset with **two classes**:

- **Class 0** (Normal cases) → **9000 samples**
- **Class 1** (Pneumonia cases) → **1000 samples**
- The dataset is highly imbalanced with a **9:1 ratio**.

Strategies to Handle Data Imbalance

- Issue with Standard Cross-Entropy Loss

Standard **Cross-Entropy (CE) Loss** is defined as:

$$CE = -\sum_{c} y_c \log(p_c)$$

where y_c is the ground truth label (0 or 1), and p_c is the predicted probability for class c.

Problem in an Imbalanced Setting:

- The model learns **bias toward the majority class (Normal cases)** since minimizing CE loss is easier when **predicting most samples as Normal**.
- The model achieves **high accuracy but poor recall on Pneumonia cases**, meaning it **fails to detect pneumonia properly**.

Strategies to Handle Data Imbalance

- Balanced Cross-Entropy (Class-Weighted CE)

Modified Loss Function

$$L = -w_0 y_0 \log(p_0) - w_1 y_1 \log(p_1)$$

where

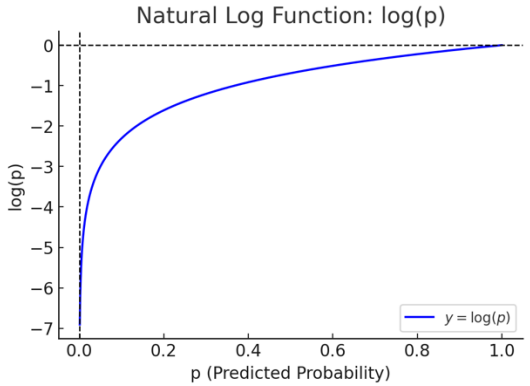
- $w_0 = \frac{1}{\text{number of normal samples}} = \frac{1}{9000}$
- $w_1 = \frac{1}{\text{number of pneumonia samples}} = \frac{1}{1000}$

Total samples / Class c samples

These weights make the **underrepresented class contribute more** to the total loss, preventing the model from ignoring it.

Strategies to Handle Data Imbalance

- Balanced Cross-Entropy works well, but **hard-to-classify samples** (e.g., subtle pneumonia cases) still **don't contribute much to training**.



Focal Loss **modifies CE loss to focus on these hard examples:**

$$FL = -\alpha_t (1 - p_t)^{\gamma} \log(p_t)$$

where: p_t is the predicted probability of the true class.

- $(1 - p_t)^{\gamma}$ **downweights easy examples** (where p_t is high) and **amplifies loss for misclassified (hard) examples.**
- α_t (like weighted CE) assigns a different weight to each class.
- γ (focusing parameter) **controls how much to focus on hard samples** (typically set to 2).

Strategies to Handle Data Imbalance

A Concrete Example:

- **Class 0** (Normal cases) → **9000 samples**
- **Class 1** (Pneumonia cases) → **1000 samples**

- Imagine we have **pneumonia cases (y = 1)**:

Prediction A: p_1 = 0.9 (easy example, high confidence)

Prediction B: p_1 = 0.2 (misclassified, hard example)

Strategies to Handle Data Imbalance

- **True class**: $y = 1$ (Pneumonia)
- **Class distribution**:
  - Class 0 (Normal): 9,000 samples
  - Class 1 (Pneumonia): 1,000 samples
- **Class weights** (for Balanced CE):
  - $w_1 = \frac{\text{Total samples}}{\text{Class 1 samples}} = \frac{10{,}000}{1{,}000} = 10$
  - $w_0 = \frac{\text{Total samples}}{\text{Class 0 samples}} = \frac{10{,}000}{9{,}000} \approx 1.11$
- **Focal Loss hyperparameter**: $\gamma = 2$

Strategies to Handle Data Imbalance

1. Cross-Entropy (CE) Loss

The CE loss for a single example is:

$$\mathrm{CE} = -\left[y \ln(p_1) + (1-y)\ln(1-p_1)\right]$$

Since $y = 1$, this simplifies to:

$$\mathrm{CE} = -\ln(p_1)$$

Prediction A ($p_1 = 0.9$):

$$\mathrm{CE}_A = -\ln(0.9) \approx -(-0.10536) = 0.105$$

Prediction B ($p_1 = 0.2$):

$$\mathrm{CE}_B = -\ln(0.2) \approx -(-1.6094) = 1.609$$

Strategies to Handle Data Imbalance

2. Balanced CE Loss

Balanced CE scales the loss by the inverse class frequency. For $y = 1$:

$$\text{Balanced CE} = w_1 \times \text{CE}$$

Prediction A ($p_1 = 0.9$):

$$\text{Balanced CE}_A = 10 \times 0.105 = 1.05$$

Prediction B ($p_1 = 0.2$):

$$\text{Balanced CE}_B = 10 \times 1.609 = 16.09$$

Strategies to Handle Data Imbalance

3. Focal Loss (γ = 2)

Focal Loss modifies CE by downweighting "easy" examples. For $y = 1$:

$$\text{Focal Loss} = -(1 - p_1)^{\gamma} \ln(p_1)$$

Prediction A ($p_1 = 0.9$):

$$\text{Focal Loss}_A = -(1 - 0.9)^2 \ln(0.9) = -(0.1)^2 \times (-0.10536) = 0.00105$$

Prediction B ($p_1 = 0.2$):

$$\text{Focal Loss}_B = -(1 - 0.2)^2 \ln(0.2) = -(0.8)^2 \times (-1.6094) = 0.64 \times 1.6094 = 1.030$$

Strategies to Handle Data Imbalance

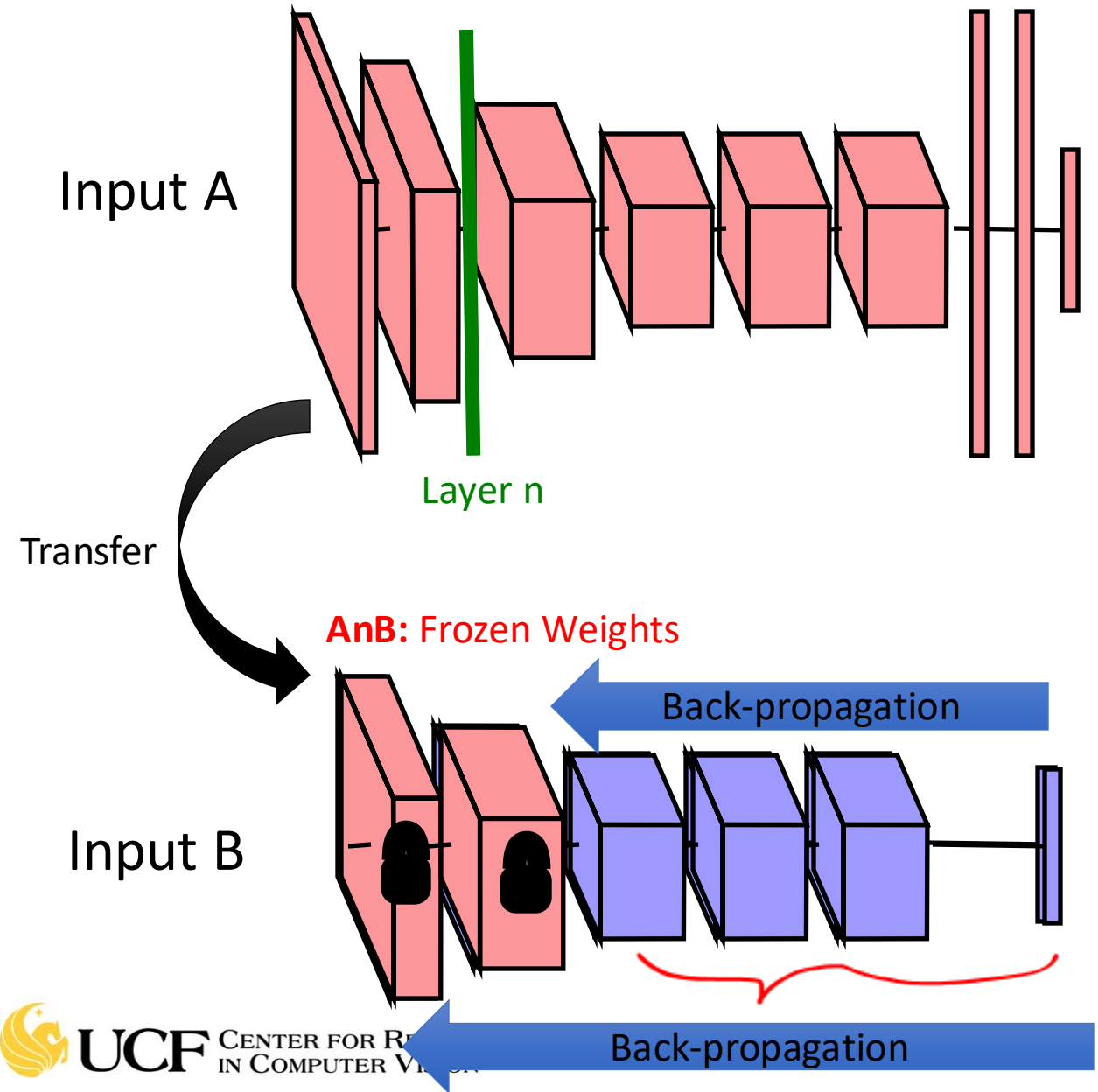
Final Results

Loss Type	Prediction A ($p_1 = 0.9$)	Prediction B ($p_1 = 0.2$)
CE Loss	0.105	1.609
Balanced CE	1.05	16.09
Focal Loss (γ=2)	0.00105	1.030

Focal loss

- Suppresses loss for "easy" examples (e.g., Prediction A: 0.00105 vs. CE's 0.105).
- Focuses on "hard" examples

Transfer Learning



(image: Aghamirzaie & Salomon)

AnB+: Fine-tuning

When and how to fine-tune?

- Suppose we have model A, trained on dataset A
- Q: How do we apply transfer learning to dataset B to create model B?

When and how to fine-tune?

- New dataset is small and similar to original dataset.
  - train a linear classifier on the CNN codes
- New dataset is large and similar to the original dataset
  - fine-tune through the full network
- New dataset is small but very different from the original dataset
  - SVM classifier from activations somewhere earlier in the network
- New dataset is large and very different from the original dataset
  - fine-tune through the entire network

Dataset size	Dataset similarity	Recommendation
Large	Very different	Train model B from scratch, initialize weights from model A
Large	Similar	OK to fine-tune (less likely to overfit)
Small	Very different	Train classifier using the earlier layers (later layers won't help much)
Small	Similar	Don't fine-tune (overfitting). Train a linear classifier

https://cs231n.github.io/transfer-learning/

Examples

- https://pytorch.org/tutorials/beginner/transfer_learning_tutorial.html
- https://blog.keras.io/building-powerful-image-classification-models-using-very-little-data.html

Evaluating Medical Image Classification Models

- Classification accuracy (on test set)

- Can accuracy always reflect the model performance?

Evaluating Medical Image Classification Models

Why Accuracy Isn't Enough

- **Class imbalance example**:
  - 9,000 normal vs. 1,000 pneumonia cases.
  - A model predicting "normal" for all cases achieves 90% accuracy but 0% pneumonia detection.

- **Key Takeaway**:
  - Accuracy is misleading in imbalanced datasets.
  - Clinical impact demands metrics aligned with domain priorities.

Evaluating Medical Image Classification Models

Confusion Matrix Basics

Visual:

	Predicted Normal	Predicted Pneumonia
Actual Normal	TN	FP
Actual Pneumonia	FN	TP

- **Definitions**:
  - **TP** (True Positive): Correctly identified pneumonia.
  - **FP** (False Positive): Normal case misclassified as pneumonia.
  - **FN** (False Negative): Pneumonia missed by the model.
  - **TN** (True Negative): Correctly identified normal.

Evaluating Medical Image Classification Models

1. **Precision (Positive Predictive Value)**:

$$\text{Precision} = \frac{TP}{TP + FP}$$

   - *Interpretation*: How many predicted pneumonia cases are real?
   - *Clinical Use*: Critical when FP costs are high (e.g., unnecessary biopsies).

2. **Recall/Sensitivity (True Positive Rate)**:

$$\text{Recall} = \frac{TP}{TP + FN}$$

   - *Interpretation*: How many actual pneumonia cases are detected?
   - *Clinical Use*: Prioritized when missing a case is dangerous (e.g., cancer screening).

Evaluating Medical Image Classification Models

3. **Specificity (True Negative Rate)**:

$$\text{Specificity} = \frac{TN}{TN + FP}$$

   - *Interpretation*: How well does the model rule out normal cases?

4. **F1-Score**:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

   - Balances precision and recall.

Evaluating Medical Image Classification Models

- Multi-class case

Confusion Matrix (Highlight critical misclassifications)

Actual \ Predicted	Class 1 (Glioma)	Class 2 (Meningioma)	Class 3 (Pituitary)
Class 1 (Glioma)	45	3	2
Class 2 (Meningioma)	5	50	0
Class 3 (Pituitary)	1	0	60

Key Observations:

- Diagonal = **correct predictions** (e.g., 45 gliomas correctly classified).
- Off-diagonal = **confusion between classes** (e.g., 5 meningiomas misclassified as gliomas).

Evaluating Medical Image Classification Models

- Multi-class case

1. Per-Class Metrics

For each class C_i, compute:

- **Precision**$_i$: $\frac{TP_i}{TP_i+FP_i}$
- **Recall**$_i$ **(Sensitivity)**: $\frac{TP_i}{TP_i+FN_i}$
- **Specificity**$_i$: $\frac{TN_i}{TN_i+FP_i}$
- **F1**$_i$: $2 \times \frac{\text{Precision}_i \times \text{Recall}_i}{\text{Precision}_i+\text{Recall}_i}$

Evaluating Medical Image Classification Models

- Multi-class case

2. Aggregated Metrics

- **Macro Average**:
  - Compute metric for each class, then average.
  - Treats all classes equally (good for balanced datasets).

$$\text{Macro F1} = \frac{F1_1 + F1_2 + F1_3}{3}$$

- **Weighted Average**:
  - Compute metric for each class, then average weighted by class support (number of samples).
  - Accounts for class imbalance (common in medical data).

$$\text{Weighted F1} = \frac{(F1_1 \times n_1) + (F1_2 \times n_2) + (F1_3 \times n_3)}{n_1 + n_2 + n_3}$$

Evaluating Medical Image Classification Models

- **Micro Average**:
  - Pool all classes' TP, FP, FN, TN globally.
  - Reflects overall performance (similar to accuracy in balanced cases).

$$\text{Micro F1} = \frac{\sum TP}{\sum TP + \frac{1}{2}(\sum FP + \sum FN)}$$

Evaluating Medical Image Classification Models

Key takeaways:

1. **Confusion Matrix**: Visualizes class-specific errors (critical for clinical auditing).
2. **Macro vs. Weighted Metrics**:
    1. Use **macro** if all classes are equally important (e.g., rare diseases).
    2. Use **weighted** if class distribution reflects clinical prevalence.

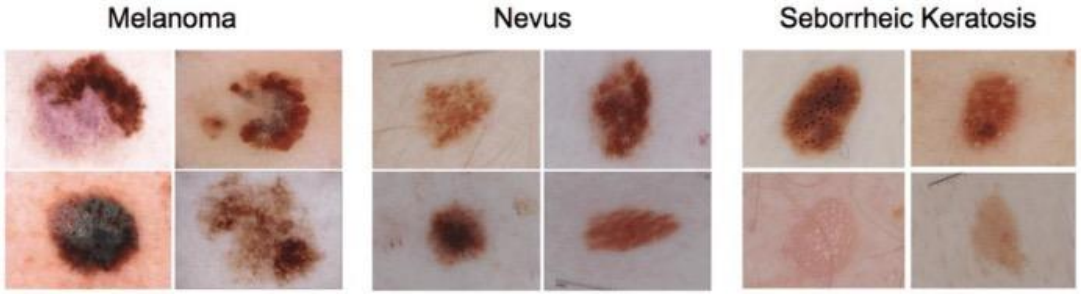
Clinical prevalence refers to the **proportion of individuals in a specific population who have a particular disease or condition at a given time**. It represents how common a disease is in a medical setting and is often expressed as a percentage or ratio.

For example, if **200 out of 10,000** patients in a hospital have **lung cancer**, the clinical prevalence is 2%.

Evaluating Medical Image Classification Models

Question:

"In a 3-class skin lesion dataset (Melanoma, Nevus, Seborrheic Keratosis), which metric would you prioritize if early detection of melanoma is critical?"



Answer:

- **Recall for Melanoma** (minimize false negatives).
- **Macro F1** to ensure rare classes (e.g., melanoma) aren't drowned out by prevalent ones (e.g., nevus).

Chest X-ray Datasets

CheXpert

- **Description**: Large dataset of chest X-rays labeled for 14 pathologies (e.g., pneumonia, edema).
- **Key Features**:
  - 224,316 images from 65,240 patients.
  - Labels include uncertainty flags (e.g., "uncertain pneumonia").
- **Classes**: Multi-label classification (14 conditions).
- **Link**: https://stanfordmlgroup.github.io/competitions/chexpert/

Chest X-ray Datasets

ChestX-ray14 (NIH)

- **Description**: 112,120 frontal-view chest X-rays with 14 disease labels.
- **Key Features**:
  - Labels mined from radiology reports.
  - Popular baseline for pneumonia and cardiomegaly detection.
- **Classes**: Multi-label classification.
- **Link**: https://nihcc.app.box.com/v/ChestXray-NIHCC

Chest X-ray Datasets

RSNA Pneumonia Detection Challenge

- **Description**: Chest X-rays with bounding boxes for pneumonia lesions.
- **Key Features**:
  - 26,684 images (pediatric and adult cases).
  - Designed for classification + localization.
- **Link**: https://www.kaggle.com/c/rsna-pneumonia-detection-challenge

Brain Imaging Datasets

Alzheimer's Disease Neuroimaging Initiative (ADNI)

- **Description**: MRI/PET scans for Alzheimer's disease classification.
- **Key Features**:
  - Longitudinal data (mild cognitive impairment → Alzheimer's progression).
  - Multi-modal (MRI, PET, CSF biomarkers).
- **Classes**: CN (normal), MCI, Alzheimer's.
- **Link**: https://adni.loni.usc.edu/

Brain Imaging Datasets

BRATS (Brain Tumor Segmentation)

- **Description**: MRI scans for brain tumor classification/segmentation.
- **Key Features**:
  - Gliomas (HGG/LGG) with segmentation labels.
  - Multi-modal MRI (T1, T1ce, T2, FLAIR).
- **Classes**: Tumor sub-regions (edema, enhancing tumor, necrosis).
- **Link**: https://www.med.upenn.edu/cbica/brats/

Dermatology & Histopathology Datasets

ISIC Archive

- **Description**: Largest public skin lesion dataset for melanoma classification.
- **Key Features**:
  - Over 70,000 dermoscopy images.
  - Annotations: Benign, malignant, lesion segmentation masks.
- **Link**: https://www.isic-archive.com/

Dermatology & Histopathology Datasets

PatchCamelyon

- **Description**: Histopathology dataset for metastatic cancer detection.
- **Key Features**:
  - 327,680 patches from H&E-stained lymph node slides.
  - Binary classification: Metastatic vs. normal tissue.
- **Link**: https://www.tensorflow.org/datasets/catalog/patch_camelyon

UCF-MultiOrgan-Path

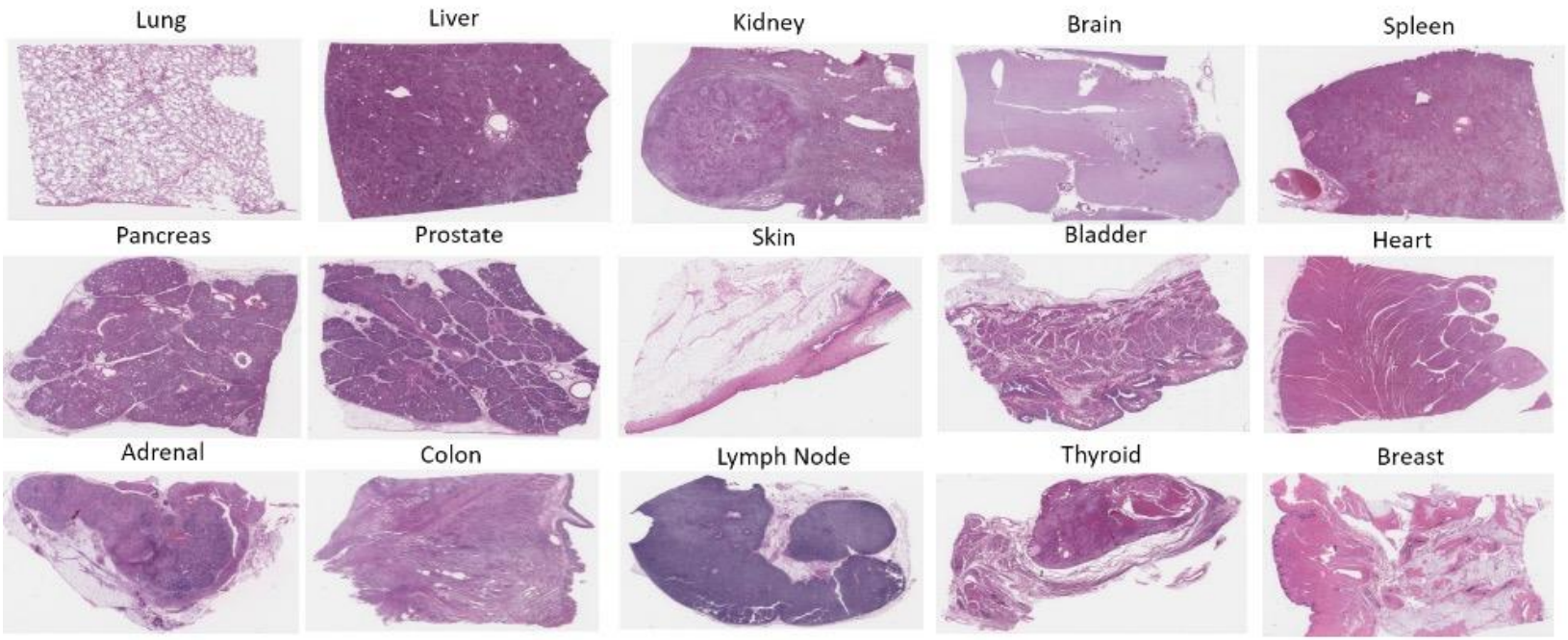


Figure 2. Sample WSI image for each organ class

Hossain, Md Sanzid Bin, Yelena Piazza, Jacob Braun, Anthony Bilic, Michael Hsieh, Samir Fouissi, Alexander Borowsky et al. "UCF-MultiOrgan-Path: A Public Benchmark Dataset of Histopathologic Images for Deep Learning Model Based Organ Classification." *medRxiv* (2024): 2024-11.

Ophthalmology

Diabetic Retinopathy (Kaggle/EyePACS)

- **Description**: Retinal fundus images graded for diabetic retinopathy severity.
- **Key Features**:
  - 88,702 images with 5-class labels (no DR → proliferative DR).
  - Challenges: Class imbalance, subtle lesions.
- **Link**: https://www.kaggle.com/c/diabetic-retinopathy-detection/data

Ophthalmology

OCT (Optical Coherence Tomography) Dataset

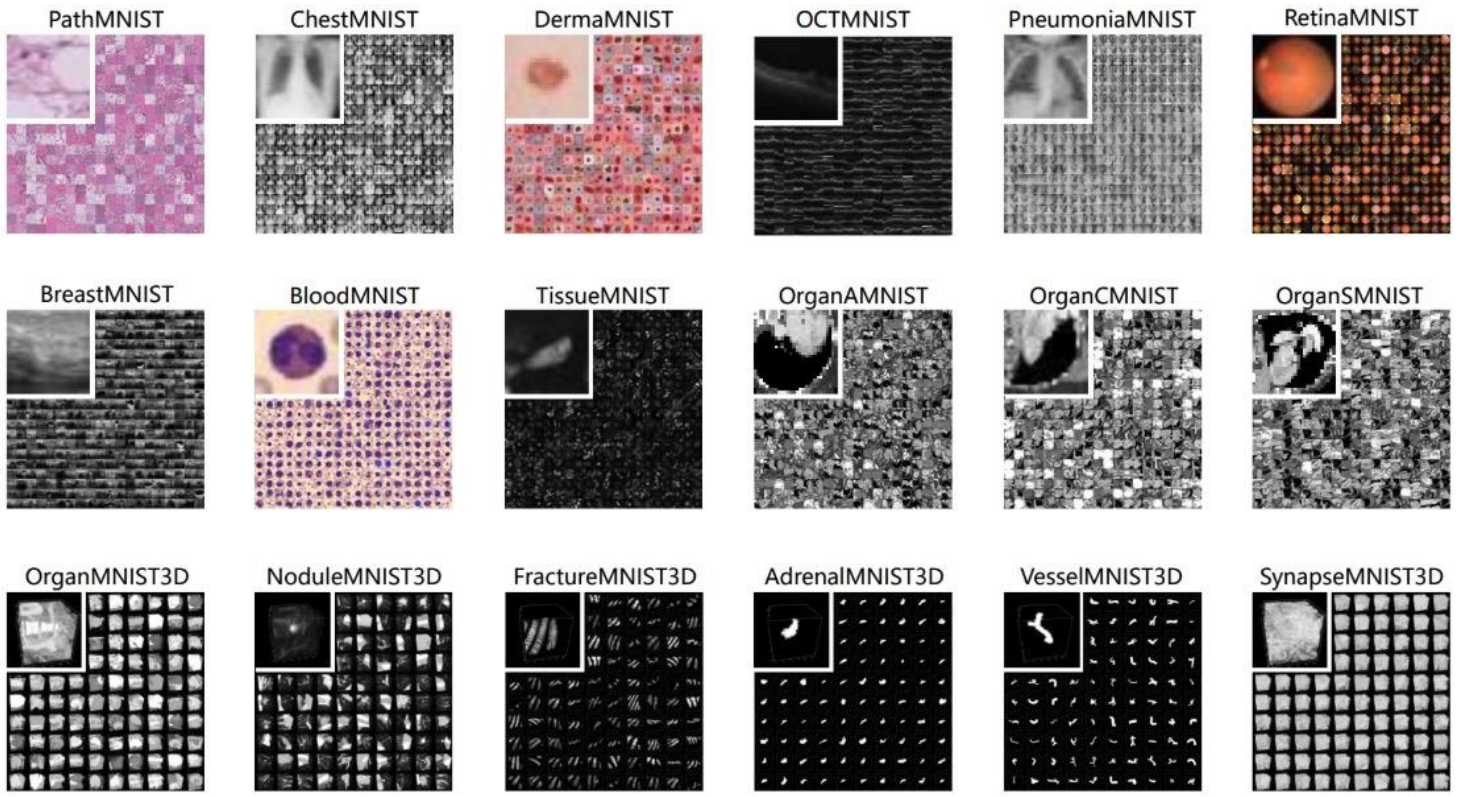
- **Description**: Retinal OCT scans for AMD (age-related macular degeneration) and DME.
- **Key Features**:
  - 84,495 images (CNV, DME, Drusen, Normal).
- **Link**: https://data.mendeley.com/datasets/rscbjbr9sj/3

PhysioNet

- **PhysioNet**
- **Description**: Diverse datasets (ECG, EEG, imaging).
- **Example**:
  - MIMIC-CXR (chest X-rays with free-text reports).
- **Link**: https://physionet.org/about/database/

MedMNIST

- https://medmnist.com/



More Datasets …

- https://grand-challenge.org/

Example: WSI Classification

Whole Slide Image Classification in Histopathology

- **Definition**: High-resolution digital scans of entire tissue slides (gigapixel scale).
- **Modality**: Hematoxylin & Eosin (H&E), immunohistochemistry (IHC).

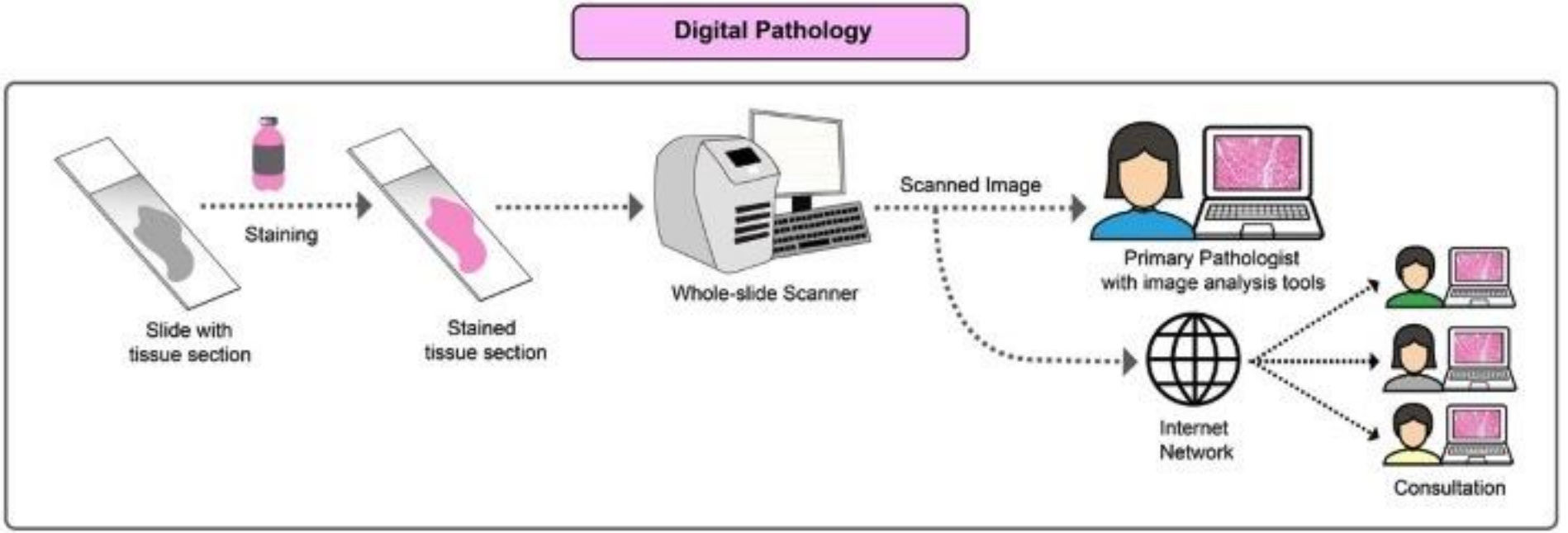


Image source: https://www.precisiononcology.ie/newsevents/blogs/items/text,502846,en.html

Whole Slide Image Classification in Histopathology

- **Typical Setting for Histopathology Image Classification**

Whole Slide Images (WSIs) are digitized high-resolution scans of tissue samples, often exceeding **100,000 × 100,000 pixels** (gigapixel scale).

Labeling Context

- Often, only **slide-level labels** are available (e.g., “benign” vs. “malignant”).
- **No pixel-wise or region-level annotations** (weak supervision).

Example

- **Breast Cancer Diagnosis** from hematoxylin and eosin (H&E) stained slides.
- The entire WSI is labeled “cancer” if any region shows malignancy, and “benign” otherwise.

Whole Slide Image Classification in Histopathology

- **Challenges of WSI Classification**

Computational Intractability:

Directly processing a full WSI is infeasible due to memory and computational constraints.

Weak Supervision:

Only **slide-level labels** (e.g., "cancer" or "normal") are typically available, while diagnostic regions may occupy <1% of the WSI.

High Intra-Slide Variability:

A single WSI contains diverse tissue regions (e.g., tumor, stroma, necrosis, artifacts).

Annotation Cost:

Pixel-/patch-level annotations require expert pathologists and are labor-intensive.

Whole Slide Image Classification in Histopathology

Potential Solutions for WSI Classification

- **Patch-Based Classification + Aggregation**
  - **Patch Extraction**: Subdivide the WSI into smaller tiles (e.g., 256×256).
  - **Patch Classification**: Train a model on labeled patches, if patch-level labels are somehow available.
  - **Aggregating Predictions**: Average or majority-vote patch outputs to get a slide-level prediction.
  - **Limitation**: Requires patch-level labels (rare in real clinical workflows).

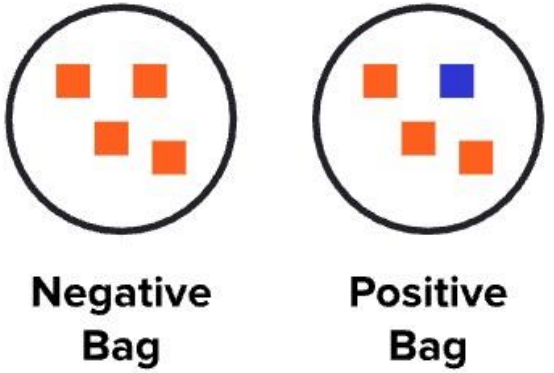
Whole Slide Image Classification in Histopathology

Potential Solutions for WSI Classification

- **Multiple Instance Learning (MIL) based approach**

Whole Slide Image Classification in Histopathology

Multiple Instance Learning (MIL) based approach



Conceptual Overview:

- In MIL, each **sample** is a "bag" containing multiple "instances."
- For WSIs, the **bag** = entire slide, **instances** = patches extracted from the slide.
- A slide-level label indicates whether **at least one patch** in the WSI is positive (e.g., contains cancer cells).
- A bag is positive if **at least one instance** is positive (e.g., tumor).

Whole Slide Image Classification in Histopathology

MIL for WSI – Step-by-Step Workflow

Step 1: Patch Extraction

- **Tiling the WSI** into manageable patches (e.g., 256×256 or 512×512 pixels).
- Often remove background using **tissue segmentation** (reduces noise and file size).
- **Bag Construction**: Each WSI → "bag" of N patches (instances).

Whole Slide Image Classification in Histopathology

MIL for WSI – Step-by-Step Workflow

Step 2: Patch-Level Feature Extraction

- **Pretrained CNN** (e.g., ResNet, EfficientNet) used to embed each patch into a **feature vector**. $\{h_1, h_2, ..., h_N\}$
- This step is often performed offline, storing patch embeddings to disk.
- Reduces training overhead for the subsequent steps (don't repeatedly process raw pixel data).

Whole Slide Image Classification in Histopathology

MIL for WSI – Step-by-Step Workflow

Step 3: MIL Pooling / Aggregation

- The **core** of MIL is combining patch embeddings to form a **slide-level representation**.
- **Common Pooling Strategies**:
  - **Max Pooling**: The patch with the highest score dictates the slide label.
  - **Mean Pooling**: Averages all patch features.
  - **Attention-Based Pooling**: Learns patch-specific attention weights, providing interpretability (highlighting critical patches).
  - **Transformer-Based Approaches**: Model pairwise interactions among patches.

Whole Slide Image Classification in Histopathology

MIL for WSI – Step-by-Step Workflow

Step 4: Slide-Level Classification

- The **aggregated** (pooled) feature vector is passed to a **classifier** (e.g., MLP or fully connected layer).
- The classifier predicts the **slide-level** label (e.g., cancer vs. benign).
- **Loss** is computed at the bag (slide) level (e.g., using cross-entropy).

Whole Slide Image Classification in Histopathology

MIL for WSI – Step-by-Step Workflow

Step 5: Model Training

- **Forward Pass**: For each slide (bag), get patch features → apply MIL pooling → get prediction.
- **Backward Pass**: Compute **slide-level loss** vs. ground-truth label → backpropagate to update both feature extractor and MIL pooling parameters (if trained end-to-end).

Whole Slide Image Classification in Histopathology

A Close Look of Step 3 - MIL Pooling / Aggregation

Step 3: MIL Pooling / Aggregation

- The **core** of MIL is combining patch embeddings to form a **slide-level representation**.
- **Common Pooling Strategies**:
  - **Max Pooling**: The patch with the highest score dictates the slide label.
  - **Mean Pooling**: Averages all patch features.
  - **Attention-Based Pooling**: Learns patch-specific attention weights, providing interpretability (highlighting critical patches).
  - **Transformer-Based Approaches**: Model pairwise interactions among patches.

Whole Slide Image Classification in Histopathology

MIL Pooling / Aggregation (max-pooling)

- **Binary classification**: benign (0) vs. malignant (1).
- **Input**: A **bag** (WSI) consisting of N patches, each patch i has a feature vector $\mathbf{f}_i \in \mathbb{R}^D$.
- **Patch Logits**: A small neural network (often a fully connected layer) produces a **scalar logit** s_i for each patch. Example:

$$s_i = \mathbf{w}^T \mathbf{f}_i + b.$$

- **Max-Pooling**: The final **slide-level logit** s_{slide} is given by

$$s_{\text{slide}} = \max_{1 \le i \le N} s_i.$$

- **Slide Probability**: For binary classification, apply **sigmoid**:

$$\hat{y} = \sigma(s_{\text{slide}}) = \frac{1}{1 + e^{-s_{\text{slide}}}}.$$

Whole Slide Image Classification in Histopathology

MIL Pooling / Aggregation (mean-pooling)

- **Binary classification**: benign (0) vs. malignant (1).
- **Input**: A **bag** (WSI) consisting of N patches, each patch i has a feature vector $\mathbf{f}_i \in \mathbb{R}^D$.
- **Compute the average of all patch features** in a bag.
- Slide-level representation $\mathbf{z}$ is simply:

$$\mathbf{z} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{f}_i.$$

- A fully connected layer transforms $\mathbf{z}$ into a logit s_{slide}, which then goes through a sigmoid for a binary prediction.

Cons: **Averaging** can dilute critical signals if only a small subset of patches is malignant.

Whole Slide Image Classification in Histopathology

MIL Pooling / Aggregation (attention-based pooling)

- **Binary classification**: benign (0) vs. malignant (1).
- **Input**: A **bag** (WSI) consisting of N patches, each patch i has a feature vector $\mathbf{f}_i \in \mathbb{R}^D$.
- Learn **patch-specific attention weights** that highlight the importance of each patch.
- The slide-level representation is a **weighted sum** of patch features:

$$\mathbf{z} = \sum_{i=1}^{N} a_i \, \mathbf{f}_i, \quad \text{where} \quad a_i = \frac{\exp(\alpha_i)}{\sum_{j=1}^{N} \exp(\alpha_j)}.$$

- α_i is computed by a **learnable** function of each patch's feature $\mathbf{f}_i$.

Whole Slide Image Classification in Histopathology

MIL Pooling / Aggregation (attention-based pooling)

Attention Score Calculation

- A small neural network (often MLP) takes $\mathbf{f}_i$ and outputs a *scalar* attention logit α_i.

$$\alpha_i = \mathbf{w}^T \tanh(\mathbf{V}\,\mathbf{f}_i^T) \quad \text{(example formulation)},$$

  with $\mathbf{V}$ and $\mathbf{w}$ as learnable parameters.

Attention Normalization

- Convert $\{\alpha_i\}$ into attention weights $\{a_i\}$ via **softmax**:

$$a_i = \frac{\exp(\alpha_i)}{\sum_{j=1}^{N} \exp(\alpha_j)}.$$

Weighted Summation

- $\mathbf{z} = \sum_i a_i\,\mathbf{f}_i.$

Whole Slide Image Classification in Histopathology

MIL Pooling / Aggregation (Transformer-based approach)

- **Binary classification**: benign (0) vs. malignant (1).
- **Input**: A **bag** (WSI) consisting of N patches, each patch i has a feature vector $\mathbf{f}_i \in \mathbb{R}^D$.

- **Treat each patch embedding as a "token"** in a transformer model.
- A special **[CLS] token** (or similar approach) learns a **bag-level embedding** through **self-attention** across patches.
- The final hidden state of [CLS] represents the **entire slide**.

- **Pros**: Captures **pairwise and global** relationships among patches.
- **Cons**: **High computational cost** and memory usage, especially if N is large.

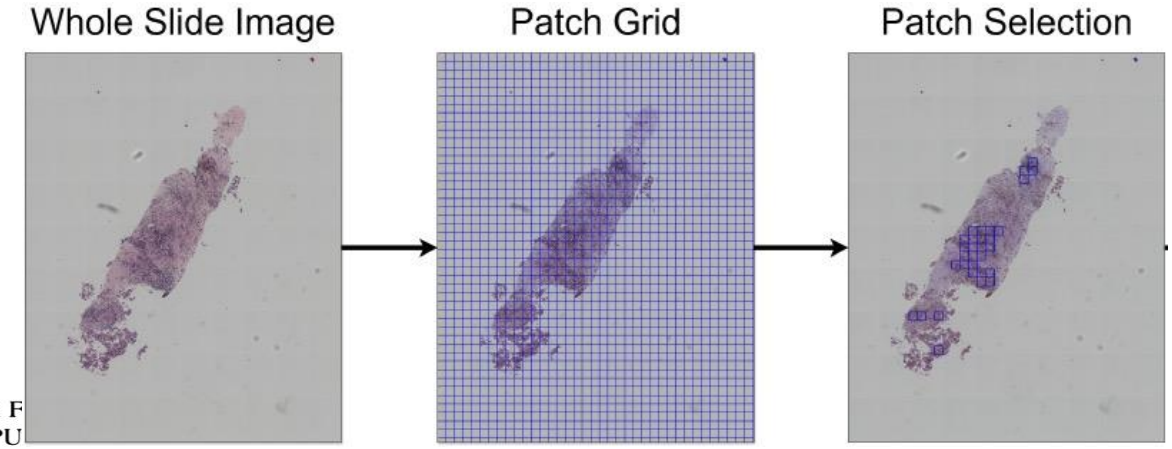
Whole Slide Image Classification in Histopathology

- Patch Selection

Goal: Remove non-informative regions (e.g., blank background).

Methods:

- **Otsu Thresholding**: Separate tissue from background using pixel intensity.
- **Deep Learning Segmentation**: Use U-Net or similar models to segment tissue regions.
  Example: Only select patches with >50% tissue area.



Whole Slide Image Classification in Histopathology

Embedding-Based Clustering or Representative Selection

1. **Extract embeddings** for all patches with a CNN (e.g., ResNet).
2. **Cluster** embeddings (e.g., k-means, hierarchical clustering) in the feature space.
3. **Select representative patches** from each cluster or from clusters that appear “suspicious” if partial labels or heuristics are available.

- **Example**:
- Cluster N=10,000 patch embeddings into 50 clusters.
- Pick the **centroid** or top 1–2 patches per cluster.
- Run MIL on these ~100 selected patches.

Whole Slide Image Classification in Histopathology

- Any ideas on patch selection?

Thank you!

Question?