Winter In Data Science 5.0

AI-Powered Anomaly Detection in Medical Imagery

Name:

Rushabh Ramesh Bonde

Roll No.:

23b0703

January 28, 2026

Contents

1 Digital Image Processing

Digital Image Processing (DIP) is the study of how images can be represented, analysed, enhanced, and transformed using computers. Instead of treating images as visual objects meant only for humans, DIP treats images as data that can be processed mathematically.

At its core, digital image processing answers three fundamental questions:

- How is an image represented inside a computer?
- How can this representation be modified to improve or analyse the image?
- What information can be extracted from the image?

1.1 Images as Mathematical Objects

A digital image is best understood as a function: $f(x, y)$ where x and y represent spatial coordinates and $f(x, y)$ represents the intensity of light at that location.

- For grayscale images, this intensity is a single value (usually 0–255).
- For colour images, each pixel contains multiple intensity values, one per colour channel.

This viewpoint is essential because it allows image processing to be performed using linear algebra, calculus, probability, and signal processing techniques rather than manual editing.

1.2 Pixels, Resolution, and Sampling

An image is made up of pixels, which are the smallest addressable units of an image.

- **Resolution** refers to the number of pixels in an image (width $\times$ height).
- Higher resolution means more detail, but also more storage and computation.
- Lower resolution reduces detail and may hide important features.

In medical imaging and computer vision, choosing the right resolution is critical because microscopic features (like cells or edges) may disappear if the image is downsampled too much.

1.3 Colour Images and Channels

Most digital images use the RGB colour model, where:

- Each pixel is represented using three values: Red, Green, and Blue.
- Each channel contributes differently to the perceived colour.

Although RGB is intuitive for display, other colour spaces (like HSV, LAB, or YCbCr) are often better for analysis because they separate brightness from colour information. This separation can make tasks like segmentation or feature extraction easier.

1.4 Image Storage in Memory

Conceptually, a colour image looks like a 3D structure: height $\times$ width $\times$ channels

However, computers store everything as 1D sequences of numbers. To make this possible, images are flattened and stored using specific memory layouts:

- HWC (Height $\times$ Width $\times$ Channels)
- CHW (Channels $\times$ Height $\times$ Width)

The choice of layout affects performance, especially in deep learning frameworks and GPU computation.

1.5 Image Transformations and Interpolation

When an image is resized or transformed, new pixel values must be estimated. This is done using interpolation.

- **Nearest Neighbour Interpolation:** Copies the value of the nearest pixel. Fast, but produces blocky images.
- **Bilinear Interpolation:** Averages nearby pixels. Smoother, but slightly blurry.

Interpolation is unavoidable during resizing, but excessive use degrades image quality.

1.6 Information Loss in Images

Whenever an image is reduced in size or heavily processed, information is lost.

- Fine textures, edges, and small objects may disappear.
- Once lost, this information cannot be recovered.

This is why aggressive resizing or compression is dangerous in applications like medical imaging, where small visual details may be clinically important.

2 Types of Healthcare Reports

Healthcare reports are structured documents used to record, communicate, and track patient information across different stages of diagnosis, treatment, and recovery. Each report serves a specific purpose and is created by different healthcare professionals. Together, these reports ensure continuity of care, accurate decision-making, and legal documentation.

2.1 Medical History Report

A medical history report records a patient's past and present health information. It includes details such as previous illnesses, surgeries, allergies, medications, family medical history, and lifestyle factors. This report helps clinicians understand a patient's baseline health and identify risk factors before diagnosis or treatment.

2.2 Laboratory Reports

Laboratory reports present the results of diagnostic tests performed on biological samples such as blood, urine, or tissue. These reports include numerical values, reference ranges, and abnormality indicators. They help clinicians detect infections, metabolic disorders, organ dysfunction, and disease progression.

2.3 Medical Imaging Reports

Medical imaging reports interpret visual data obtained from imaging modalities such as X-rays, CT scans, MRI scans, ultrasound, and PET scans. These reports translate complex images into clinically meaningful findings, allowing physicians to make informed diagnostic and treatment decisions.

2.4 Pathology Reports

Pathology reports analyse tissue, cell, or fluid samples under a microscope. They are crucial in diagnosing diseases such as cancer, identifying tumour type, grade, and stage, and guiding treatment decisions. These reports often form the definitive diagnosis in oncology.

2.5 Radiology Reports

Radiology reports are a specialised subset of medical imaging reports generated by radiologists. They focus on the interpretation of imaging studies and provide detailed descriptions of anatomical structures and abnormalities seen in images.

2.6 Discharge Summary

A discharge summary is prepared when a patient leaves a hospital. It includes the reason for admission, diagnosis, treatment provided, medications prescribed, and follow-up instructions. This report ensures continuity of care after hospitalisation.

2.7 Operative Reports

Operative reports document surgical procedures. They include details such as the indication for surgery, surgical findings, steps performed, complications, and postoperative condition. These reports are critical for legal documentation and postoperative care.

2.8 Progress Notes

Progress notes are written regularly during a patient's hospital stay or treatment course. They track changes in condition, response to treatment, and plans for ongoing care. These notes help healthcare teams monitor patient progress over time.

2.9 Prescription and Medication Reports

These reports list medications prescribed to a patient, including dosage, frequency, and duration. They help ensure safe and accurate medication administration and prevent drug interactions.

2.10 Vital Signs and Monitoring Reports

Vital signs reports record measurements such as heart rate, blood pressure, respiratory rate, temperature, and oxygen saturation. Continuous monitoring reports are especially important in critical care settings to detect early signs of deterioration.

3 Medical Imaging Reports (Detailed Explanation)

Medical imaging reports are formal documents created by radiologists after analyzing medical images. These images include X-rays, CT scans, MRI scans, ultrasound images, and PET scans. While the images themselves contain rich visual information, they require expert interpretation to become clinically useful. The imaging report serves this purpose by converting visual patterns into medical conclusions.

A typical medical imaging report contains several key components:

- **Patient Information:** This section includes essential details such as the patient's age, gender, and relevant clinical history. Context is important because imaging findings can have different meanings depending on the patient's background and symptoms.

- **Imaging Modality and Technique:** Here, the type of imaging study is specified, such as a CT scan with contrast or an MRI using specific sequences (T1, T2, or FLAIR). This information helps clinicians understand how the images were acquired and how to interpret the findings correctly.
- **Findings:** The findings section describes what is observed in the images. This includes normal anatomical structures as well as abnormalities such as lesions, fractures, masses, or fluid accumulation. Radiologists aim to describe findings objectively without immediately drawing conclusions.

3.1 Impression or Conclusion

The impression summarizes the most important findings and provides a clinical interpretation. This section often includes a diagnosis or a differential diagnosis, making it the most critical part of the report for clinicians.

3.2 Recommendations

Based on the findings, the radiologist may suggest follow-up imaging, additional tests, or clinical correlation. These recommendations guide further patient management.

3.3 Importance of Medical Imaging Reports

Medical imaging reports act as a bridge between raw visual data and clinical decision-making. Physicians often rely more on the report than on the images themselves, especially when they lack specialized imaging expertise. In modern healthcare, these reports are essential for diagnosis, treatment planning, monitoring disease progression, and maintaining accurate medical records.

4 Medical Imaging Datasets

Medical imaging datasets are collections of images obtained using clinical imaging modalities such as X-rays, CT scans, MRI, ultrasound, dermoscopy, or histopathology. These datasets are used to train and evaluate machine learning and deep learning models for tasks like disease detection, classification, segmentation, and prognosis.

Unlike general image datasets, medical datasets have several unique characteristics:

- They are often small due to privacy and annotation costs.

- Labels are created by medical experts and may contain uncertainty.
- Images may vary significantly due to different machines, hospitals, and acquisition settings.
- Class imbalance is common, especially for rare diseases.

Understanding these properties is essential before applying any learning algorithm.

4.1 Availability of Medical Datasets (Roboflow and Kaggle)

4.1.1 Kaggle

Kaggle is a widely used platform that hosts many open-source medical imaging datasets. It provides:

- Easy access to datasets uploaded by researchers and institutions
- Metadata describing dataset composition
- Community notebooks and benchmarks

However, datasets on Kaggle are often raw and may require significant preprocessing, cleaning, and reformatting before use.

4.1.2 Roboflow

Roboflow is a platform designed specifically for computer vision workflows. It provides:

- Dataset hosting and versioning
- Annotation tools
- Automatic conversion to formats such as YOLO, COCO, and Pascal VOC
- Built-in dataset analytics (class balance, image resolution, annotation density)

Roboflow is particularly useful for object detection and segmentation tasks and helps standardize datasets for training deep learning models.

4.2 Dataset Analysis

4.2.1 Dataset 1: COVID-19 Radiography Database (Chest X-ray Dataset)

Type of Imaging Data: Chest X-ray images

Medical Use Case: Detection of COVID-19 and related lung infections

Image Format and Size:

- PNG images
- Mostly standardized resolution ($\sim$299 $\times$ 299 pixels)

Dataset Composition:

- COVID-19: $\sim$360 images
- Normal: $\sim$400 images
- Viral Pneumonia: $\sim$400 images
- Total: $\sim$1,200 images

Classes:

- **COVID-19:** X-rays of patients infected with SARS-CoV-2
- **Normal:** Healthy chest X-rays
- **Viral Pneumonia:** Non-COVID viral lung infections

Dataset Imbalance: The dataset shows mild class imbalance, with the COVID-19 class having fewer images than the other classes. While manageable, this imbalance can still introduce bias during training.

Challenges Observed:

1. **Image Quality Variation:** Images come from different hospitals and imaging machines, leading to variations in contrast and noise.
2. **Visual Similarity:** COVID-19 and viral pneumonia X-rays often appear visually similar, making classification difficult.
3. **Limited Dataset Size:** The small size increases the risk of overfitting.
4. **Annotation Reliability:** Labels are based on clinical diagnosis and may not always reflect consensus among radiologists.

This dataset is suitable for rapid experimentation and learning but highlights real-world challenges in medical AI.

4.2.2 Dataset 2: HAM10000 Skin Lesion Dataset (Reduced Size)

Type of Imaging Data: Dermoscopy images of skin lesions

Medical Use Case: Skin cancer and lesion classification

Dataset Composition:

- ∼10,000 images
- 7 lesion classes
- Multi-class classification task

Dataset Imbalance: The dataset is highly imbalanced, with one class dominating the distribution.

Challenges Observed:

- Severe class imbalance
- Visual similarity between different lesion types
- Difficulty in consistent annotation due to subtle differences

This dataset demonstrates the complexity of multi-class medical classification and the importance of addressing imbalance and visual ambiguity.

4.3 Image Processing Experiments

Before applying machine learning models, basic image processing steps were performed to understand how images behave under different transformations.

4.3.1 Grayscale Conversion

Grayscale conversion transforms a color image into a single-channel image by combining the RGB channels into one intensity value.

Why it is done:

- Reduces computational complexity
- Focuses on structural information rather than color
- Useful when color is not diagnostically important (e.g., X-rays)

This step helps simplify analysis and visualization.

4.3.2 Intensity Histogram Visualization

An intensity histogram shows the distribution of pixel values in an image.

What it tells us:

- Whether the image is too dark or too bright
- How contrast is distributed
- Presence of noise or poor illumination

Histograms are useful for deciding whether contrast enhancement or normalization is required.

4.3.3 Image Resizing

Images were resized to a fixed resolution.

Why resizing is necessary:

- Machine learning models require consistent input dimensions
- Reduces memory and computation requirements

Trade-off:

- Excessive resizing can remove fine details
- Interpolation methods may introduce blur or artifacts

4.3.4 Blurring (Gaussian Blur)

Blurring smooths an image by averaging pixel values with their neighbors.

Why blurring is applied:

- Reduces noise
- Removes high-frequency artifacts
- Helps focus on large structures

Limitation:

- Blurring also removes edges and fine details

4.3.5 Image Modification and Noise Handling

Additional modifications such as adding noise or adjusting intensity help simulate real-world imaging conditions.

Purpose:

- Understand how noise affects image quality
- Evaluate robustness of preprocessing methods
- Prepare images for learning-based models

5 Neural Networks and Deep Learning

5.1 Neural Networks (NN)

A Neural Network (NN) is a computational model inspired by the way the human brain processes information. It consists of layers of interconnected units called neurons. Each neuron takes numerical inputs, multiplies them by weights, adds a bias, and applies a nonlinear function to produce an output.

At a basic level:

- Input layer receives raw data (numbers)
- Hidden layers learn patterns
- Output layer produces predictions

Neural networks are powerful because they can learn from data rather than relying on fixed rules. Instead of explicitly programming how to recognize patterns, the network adjusts its internal weights during training to minimize errors.

However, basic neural networks have a major limitation: they treat every input value independently and do not understand spatial structure. This makes them inefficient for images, where neighboring pixels are strongly related.

5.2 Convolutional Neural Networks (CNN)

A Convolutional Neural Network (CNN) is a specialized type of neural network designed specifically for image and visual data.

5.2.1 Why CNNs are needed

Images have structure:

- Nearby pixels are related
- Edges, shapes, and textures matter more than individual pixel values

CNNs exploit this structure using three key ideas:

(a) Convolution Instead of connecting every input pixel to every neuron, CNNs use small filters (kernels) that slide across the image.

Each filter learns to detect a specific pattern, such as:

- Edges
- Corners
- Textures
- Shapes

This reduces the number of parameters and improves learning efficiency.

(b) Weight Sharing The same filter is applied across the entire image. This allows CNNs to detect the same pattern regardless of where it appears.

(c) Pooling Pooling layers reduce spatial size while retaining important information. This makes the network more robust to small shifts and reduces computation.

Because of these properties, CNNs became the foundation of modern computer vision systems.

5.3 Darknet Framework

Darknet is an open-source neural network framework written in C and CUDA, designed for speed and efficiency.

5.3.1 Why Darknet exists

- Early deep learning frameworks were slow for real-time vision
- Object detection needed fast inference
- GPU acceleration was essential

Darknet was built to:

- Run efficiently on GPUs
- Support real-time object detection
- Serve as the original implementation framework for YOLO

Although modern frameworks like PyTorch are more flexible, Darknet laid the foundation for real-time detection systems and influenced many later architectures.

5.4 ImageNet Fundamentals

ImageNet is a massive dataset containing millions of labeled images across thousands of classes.

5.4.1 Why ImageNet is important

Before ImageNet:

- Vision models were shallow
- Datasets were small
- Progress was slow

ImageNet enabled:

- Training very deep CNNs
- Fair benchmarking of architectures
- Discovery of architectural innovations

5.4.2 Transfer Learning

Most modern vision models are pretrained on ImageNet.

Even though ImageNet contains everyday objects, early CNN layers learn:

- Edges
- Color gradients
- Textures

These features are generic and transfer well to medical imaging, surveillance, and industrial vision. This is why ImageNet became the backbone of modern computer vision.

5.5 Common CNN Architectures

The document highlights four major CNN architectures, each solving a specific problem in deep learning.

5.5.1 AlexNet

- First deep CNN to dominate ImageNet
- Proved deep learning works for vision
- Introduced ReLU activations and GPUs

Problem solved: Showed that deep CNNs outperform traditional methods.

5.5.2 VGG

- Very deep architecture
- Uses small (3×3) filters repeatedly
- Simple and uniform design

Problem solved: Demonstrated that depth improves performance, but at the cost of heavy computation.

5.5.3 ResNet

- Introduced skip (residual) connections
- Solved the vanishing gradient problem
- Enabled extremely deep networks (100+ layers)

Problem solved: Allowed deep networks to train reliably.

5.5.4 EfficientNet

- Scales depth, width, and resolution together
- Achieves high accuracy with fewer parameters

Problem solved: Improved efficiency and performance trade-off.

5.6 YOLO (You Only Look Once)

YOLO is an object detection framework designed for speed and simplicity.

5.6.1 What makes YOLO different

Traditional object detection:

- First find regions
- Then classify them

YOLO:

- Looks at the image once
- Predicts bounding boxes and classes in a single pass

This makes YOLO:

- Extremely fast
- Suitable for real-time systems
- Easy to deploy

YOLO is one of the most influential object detection models ever developed.

5.7 Ultralytics Ecosystem

Ultralytics is the modern ecosystem that maintains and extends YOLO models.

5.7.1 Why Ultralytics matters

- Simplifies training and deployment
- Supports multiple tasks with the same framework
- Integrates seamlessly with PyTorch

It allows users to focus on experimentation and intuition rather than low-level implementation details.

5.8 YOLO Modes

YOLO provides different operational modes:

- **Train:** Learn from labeled data
- **Val:** Evaluate performance
- **Predict:** Run inference on new images
- **Export:** Convert model to deployment formats
- **Track:** Track objects across video frames

These modes cover the full lifecycle of a vision model.

5.9 YOLO Tasks

YOLO supports multiple vision tasks:

- **Detect:** Locate objects using bounding boxes
- **Segment:** Pixel-level object boundaries
- **Classify:** Image-level classification

- **Pose:** Detect keypoints (e.g., human joints)

This flexibility allows YOLO to be used across medical imaging, robotics, and surveillance.

6 Medical Skin Image Datasets

6.1 What are medical skin image datasets?

Medical skin datasets consist of 2D colour photographs of human skin, usually captured using dermatoscopes or high-resolution cameras. These images show different skin conditions such as benign lesions, melanoma, basal cell carcinoma, and other dermatological abnormalities.

Unlike natural images:

- The differences between classes are subtle
- Colour, texture, and shape variations are medically significant
- Visual similarity between diseases is very high

This makes skin datasets an excellent testbed for deep learning models.

6.2 Dataset Size, Classes, and Splits

6.2.1 Total number of images

The dataset contains 7206 images in total.

This size is:

- Large enough for CNN training
- Small enough to still risk overfitting if not handled properly

6.2.2 Classes

There are two classes:

- Melanoma
- Basal Cell Carcinoma

These are both serious skin cancers, and distinguishing between them is a clinically important task.

Why two classes matter Binary classification allows:

- Clear evaluation metrics
- Easier interpretation of results
- Focus on feature learning rather than class explosion

6.3 Dataset Splits – Train, Validation, Test

The dataset is split into:

- Training set: 5044 images
- Validation set: 1442 images
- Test set: 720 images

6.3.1 Why splits are necessary

- Training set teaches the model
- Validation set tunes hyperparameters and prevents overfitting
- Test set measures real-world performance

Without proper splits, model evaluation becomes meaningless.

6.4 Class Balance

From the Roboflow analytics:

- Melanoma and Basal Cell Carcinoma have comparable counts in each split
- This suggests low class imbalance

6.4.1 Why class balance matters

If one class dominates:

- Model becomes biased
- Accuracy becomes misleading
- Rare but critical cases are missed

Balanced datasets allow CNNs to learn discriminative features for both classes fairly.

6.5 Image Dimensions and Aspect Ratios

The Dimension Insights section shows:

- Median image size: 600×512 pixels
- Majority of images are square ($\approx 86.5\%$)
- Some images are wide or tall

6.5.1 Why this matters for CNNs

CNNs expect:

- Fixed input size
- Consistent spatial structure

This is why resizing (e.g., to 640×640) is applied before training.

However, resizing can distort lesions if aspect ratios are not handled carefully. Understanding image dimensions prevents unintentional data corruption.

6.6 Preprocessing and Augmentation

In the dataset view, the total image count appears larger after download. This happens because of data augmentation, such as:

- Rotation ($\pm15°$)
- Horizontal and vertical flips
- Auto-orientation
- Resizing

6.6.1 Why augmentation is used

Medical datasets are limited. Augmentation:

- Simulates real-world variability
- Improves generalization
- Reduces overfitting

Augmentation does not create new medical knowledge, but helps models become robust to viewpoint and orientation changes.

6.7 Why CNNs Are Used for This Dataset

Skin lesion images contain:

- Local patterns (edges, borders)
- Texture variations
- Shape irregularities

CNNs are ideal because:

- Convolutions detect local patterns
- Deeper layers capture global lesion structure
- Weight sharing reduces parameter count

Fully connected networks would fail to capture these spatial relationships efficiently.

6.8 From Dataset to Model Training

The Week3_model_train.ipynb notebook operationalizes these concepts:

1. Images are resized to a fixed shape
2. Labels are assigned based on dataset folders
3. CNN / YOLO-based model is initialized (pretrained weights)
4. Training loop learns features distinguishing melanoma vs carcinoma
5. Validation monitors overfitting
6. Test evaluation measures generalization

This notebook is not just code—it is the practical realization of CNN theory.

6.9 Why YOLO Can Be Used Even for Classification

Although YOLO is famous for object detection, modern YOLO versions support:

- Classification
- Detection
- Segmentation

For classification:

- The CNN backbone learns visual features
- The head predicts class probabilities

Using YOLO provides:

- Fast training
- Unified framework
- Easy deployment later if detection is added

7 Histopathology and TIL Detection

7.1 What is Histopathology and Why It Is Different

7.1.1 What is histopathology?

Histopathology is the study of microscopic tissue samples to identify disease. Tissue sections are stained (commonly with H&E stain) and observed under a microscope. These images capture cell-level structures, unlike X-rays or MRIs which capture organs.

7.1.2 Why histopathology is challenging for AI

- Images are extremely high resolution
- Cells are very small objects
- Visual differences are subtle
- Color varies due to staining protocols
- A single slide may contain millions of cells

This makes histopathology one of the hardest computer vision domains in healthcare.

7.2 Tumor-Infiltrating Lymphocytes (TILs)

7.2.1 What are TILs?

Tumor-Infiltrating Lymphocytes (TILs) are immune cells present inside or around tumor tissue. Their presence indicates how strongly the immune system is responding to cancer.

7.2.2 Why TILs matter clinically

High TIL density is associated with:

- Better prognosis
- Improved response to chemotherapy
- Longer survival (especially in breast cancer)

Hence, TILs act as a biological biomarker, not just a visual feature.

7.3 Traditional TIL Assessment vs AI-Based Assessment

7.3.1 Traditional method

- Pathologists manually inspect whole-slide images
- Count or estimate lymphocyte density
- Highly subjective
- Time-consuming
- Difficult to scale

7.3.2 Problems with manual assessment

- Inter-observer variability
- Fatigue and bias
- Whole-slide images are too large to inspect exhaustively

These limitations motivate automation using AI.

7.4 The TIGER Grand Challenge

7.4.1 What is TIGER?

TIGER (Tumor InfiltratinG lymphocytes in breast cancER) is an international challenge designed to:

- Standardize TIL detection
- Encourage robust AI solutions
- Bridge research and clinical practice

The challenge provides curated datasets and evaluation protocols to benchmark AI systems fairly.

7.5 Framing TIL Detection as a Computer Vision Problem

7.5.1 Why not classification?

TILs are small individual cells, not image-level labels. Classifying a whole image as "high TIL" or "low TIL" loses spatial information.

7.5.2 Object detection formulation

The problem is framed as:

- **Input:** Histopathology image patches
- **Output:** Bounding boxes around lymphocytes

This allows:

- Explicit cell counting
- Spatial localization
- Density estimation

This formulation directly supports clinical interpretability.

7.6 Why YOLOv8 Is Used

YOLOv8 is a one-stage object detection model that predicts bounding boxes and class labels in a single forward pass.

7.6.1 Why YOLO is suitable here

- Fast inference (important for large slides)
- Strong performance on small objects
- End-to-end pipeline
- Easy deployment using Ultralytics

The YOLOv8-nano variant is chosen to:

- Reduce overfitting
- Train efficiently on moderate dataset size

7.7 Dataset Characteristics and Why They Matter

7.7.1 Dataset source

The dataset is obtained from Roboflow Universe, curated specifically for lymphocyte detection.

7.7.2 Key properties

- ∼1800+ histopathology patches
- Bounding box annotations
- Single class: lymphocyte
- Train / validation / test split

7.7.3 Why patch-based data is used

Whole-slide images are too large for direct training. Slides are split into smaller patches to:

- Reduce memory usage
- Enable batch training
- Focus on local cellular patterns

7.8 YOLOv8 Annotation Format

YOLOv8 format stores:

- Class ID
- Normalized bounding box coordinates

7.8.1 Advantages

- Lightweight
- Framework-native
- Fast parsing
- Minimal preprocessing

This ensures reproducibility and clean training pipelines.

7.9 Training Strategy

7.9.1 Transfer Learning

The model starts from COCO-pretrained weights.

Why?

- Early layers already understand edges and textures
- Faster convergence
- Better generalization on small datasets

7.9.2 Optimizer: AdamW

AdamW improves stability by:

- Decoupling weight decay
- Preventing over-regularization

7.9.3 Early Stopping

Stops training when validation performance stops improving, preventing overfitting.

7.10 Evaluation Metrics

7.10.1 Precision

Out of all detected lymphocytes, how many are correct?

7.10.2 Recall

Out of all real lymphocytes, how many were detected?

7.10.3 mAP@0.5

Measures detection accuracy with moderate overlap tolerance.

7.10.4 mAP@0.5:0.95

Stricter metric averaging multiple overlap thresholds, preferred in medical AI.

Metrics ensure quantitative trust in the model.

7.11 From Detection to Clinical Meaning

7.11.1 Cell counting

Detected bounding boxes → number of lymphocytes.

7.11.2 Density estimation

Cell count normalized by tissue area.

7.11.3 TIL scoring

Density mapped to heuristic scores (0–3):

- 0: Low immune response
- 3: Strong immune response

This is how raw AI outputs become clinically interpretable biomarkers.

7.12 Challenges Unique to TIL Detection

- **Small object size:** Cells occupy few pixels
- **Visual similarity:** Other nuclei resemble lymphocytes
- **Staining variation:** Color differences across labs
- **Context loss:** Patch-based analysis lacks global tissue view

Understanding these challenges is crucial for realistic evaluation.