

---

# Automated Assessment of Tumor-Infiltrating Lymphocytes (TILs)

*Case Study: Breast Cancer Pathology*

---

Jennifer Esbel Mary

January 18, 2026

## Contents

---

<b>1</b>	<b>The Problem: Limitations of Manual TIL Assessment</b>	<b>2</b>
<b>2</b>	<b>Why It Matters: Prognosis and Precision Medicine</b>	<b>2</b>
2.1	Clinical Impact . . . . .	2
<b>3</b>	<b>The Dataset</b>	<b>2</b>
3.1	Data Characteristics . . . . .	2
3.2	Preprocessing Pipeline . . . . .	3
<b>4</b>	<b>The Approach: Deep Learning Pipeline</b>	<b>3</b>
4.1	Model Architecture: U-Net with EfficientNet . . . . .	3
4.2	Optimization Strategy . . . . .	4
<b>5</b>	<b>Experimental Results</b>	<b>4</b>
5.1	Quantitative Metrics . . . . .	4
5.2	Qualitative Performance & TIGER Score . . . . .	4
<b>6</b>	<b>Challenges and Observations</b>	<b>4</b>
6.1	The "Small Object" Dilemma . . . . .	4
6.2	Stain Heterogeneity . . . . .	5
6.3	Confluent Clusters . . . . .	5
<b>7</b>	<b>Conclusion</b>	<b>5</b>

## 1 The Problem: Limitations of Manual TIL Assessment

---

The TIGER (Tumor InfiltratinG lymphocytes in breast cancER) Challenge addresses a critical bottleneck in breast cancer pathology: the lack of a standardized, reproducible, and scalable method for quantifying the immune response to tumors.

While Tumor-Infiltrating Lymphocytes (TILs) are a vital biomarker, the current standard for assessing them relies on visual estimation by pathologists using a microscope. This manual process presents several distinct problems:

- **Subjectivity and Inter-Observer Variability:** Visual assessment is susceptible to high variability. Studies indicate that agreement among pathologists on TIL scores is often only "fair to moderate," leading to inconsistent data that complicates clinical decision-making.
- **Workflow Inefficiency:** Manual scoring is labor-intensive, slow, and requires highly trained experts, making it difficult to scale for large clinical trials or high-volume routine diagnostics.
- **Complexity of Scoring Rules:** Accurately calculating a TIL score requires adherence to complex guidelines (International TILs Working Group). Pathologists must define the tumor area, identify tumor-associated stroma, and exclude regions such as necrosis or artifacts. Replicating this strictly defined procedure consistently is challenging.
- **Quantification Limits:** Manual assessment limits the ability to quantify complex spatial relationships, restricting the prognostic potential of the tumor microenvironment.

The TIGER Challenge seeks to evaluate computer algorithms capable of automatically detecting lymphocytes and plasma cells, segmenting invasive tumor and stroma, and computing a final "TIL score" per slide.

## 2 Why It Matters: Prognosis and Precision Medicine

---

Resolving these measurement challenges is clinically significant because the immune response is a determinative factor in patient survival and treatment strategy.

### 2.1 Clinical Impact

- **Prognostic Value in Aggressive Cancers:** The challenge focuses on Triple Negative (TNBC) and HER2-positive breast cancers. Higher levels of TILs in these subtypes are strongly associated with better long-term survival.
- **Guiding Treatment Decisions:** Accurate TIL measurement helps identify patients who may benefit from immunotherapies or double-blockade treatments versus aggressive chemotherapy.
- **Standardization for Clinical Trials:** Validating AI models on large, independent datasets is essential for introducing automated TIL scoring into routine clinical practice and multi-center trials.

## 3 The Dataset

---

To develop a solution for this challenge, we utilized a specific subset of data provided via the TIGER Challenge repository on Roboflow.

### 3.1 Data Characteristics

The dataset focuses on the cellular level, requiring the differentiation of small lymphocyte nuclei (approx. 5-7  $\mu\text{m}$ ) from larger tumor nuclei and fibroblasts.

Table 1: Dataset Specifications

Attribute	Description
<b>Source</b>	Regions of Interest (ROIs) from Whole Slide Images (WSIs)
<b>Modality</b>	RGB Histopathology images (H&E Stained)
<b>Class 1</b>	Lymphocytes and Plasma cells
<b>Class 0</b>	Background, Tumor cells, Stroma, Necrosis

### 3.2 Preprocessing Pipeline

Deep learning models in pathology require rigorous preprocessing to handle biological variability.

- **Resizing:** ROIs were standardized to  $512 \times 512$  pixels to balance VRAM usage and nuclear detail preservation.
- **Normalization:** Pixel intensity distributions were normalized using ImageNet mean and standard deviation.
- **Augmentation:** To address "Stain Variability," we implemented a robust pipeline using Albumentations:
  - *Geometric:* Random Rotations (Rotate90), Flips (Horizontal/Vertical).
  - *Color:* Random Brightness/Contrast and RGB Shift to prevent overfitting to specific staining shades.

Analysis for: TCGA-AN-A0XU-01Z-00-DX1-6B0DD0FF-A20D-4BA2-8D48-FC357BA5313F\_-13692-10490-13834-10637\_-png.rf.e1ea98c9107d0f255f8468bd85c6bd36.jpg  
TIL Score: 0

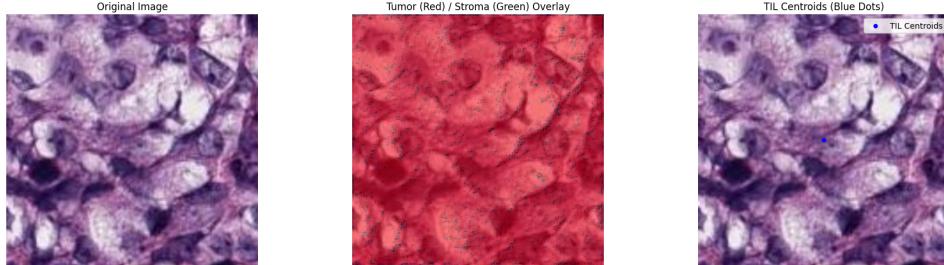


Figure 1: **Dataset Sample.** Left: The original H&E stained tissue ROI. Right: The ground truth binary mask, where white pixels indicate the location of TILs.

## 4 The Approach: Deep Learning Pipeline

Our approach utilizes Semantic Segmentation, treating the TIL assessment as a pixel-wise classification problem.

### 4.1 Model Architecture: U-Net with EfficientNet

We deployed a U-Net architecture, the gold standard for biomedical segmentation.

- **Encoder (Backbone):** EfficientNet-B0 pre-trained on ImageNet. Chosen for its high accuracy with fewer parameters, preventing overfitting on smaller datasets.
- **Decoder:** Uses skip connections to concatenate high-level semantic features with upsampled spatial maps, allowing precise localization of tiny lymphocyte cells.

## 4.2 Optimization Strategy

To address the "Small Object" problem, we utilized a composite loss function:

$$\text{Loss} = \text{DiceLoss} + \text{FocalLoss} \quad (1)$$

**Dice Loss** handles the class imbalance (sparse lymphocytes), while **Focal Loss** forces the model to learn "hard" examples (ambiguous cell boundaries). We optimized using AdamW ( $lr = 1e^{-4}$ ) and a Cosine Annealing scheduler.

## 5 Experimental Results

The training pipeline was executed over 5 epochs to establish a performance baseline.

### 5.1 Quantitative Metrics

Table 2: Performance Baseline (5 Epochs)

Metric	Value
Validation IoU	0.4789
Validation Loss	0.6366

An IoU of  $\sim 0.48$  is a respectable baseline for cell segmentation. Unlike large organ segmentation, cell IoU is strictly penalized for slight boundary misalignments. This score indicates successful detection of lymphocyte clusters.

### 5.2 Qualitative Performance & TIGER Score

The model output was post-processed to generate a clinical metric. The `calculate_tiger_score` function computes the density of TILs to mimic the manual "percentage score."

$$\text{Density} = \frac{\text{Total Lymphocyte Pixels}}{\text{Total Tissue Area}} \times 100 \quad (2)$$

Image: TCGA-AN-A0XU-01Z-00-DX1-6B0DD0FF-A20D-4BA2-8D48-FC357BA5313F\_-13692-10490-13834-10637\_.png.rf.e1ea98c9107d0f255f8468bd85c6bd36.jpg  
TILs Density: 52.96%

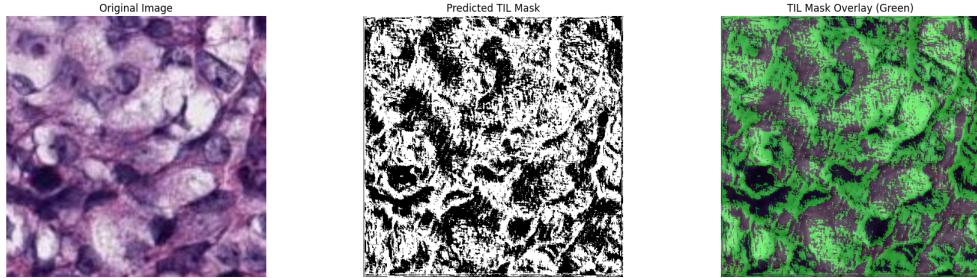


Figure 2: **Model Inference.** The original tissue image with the model's prediction overlaid in green. The model successfully identifies clusters of lymphocytes while ignoring large tumor nuclei.

## 6 Challenges and Observations

### 6.1 The "Small Object" Dilemma

Lymphocytes often occupy  $< 1\%$  of the image pixels. Standard CNN pooling operations tend to lose these features.

- **Observation:** The U-Net’s skip connections were critical. Without them, the decoder failed to recover the location of individual cells.

## 6.2 Stain Heterogeneity

Histopathology slides vary significantly in color (dark purple vs. light pink) due to lab protocols.

- **Solution:** Heavy color augmentation (RGB Shift) was necessary to force the model to ignore color intensity and focus on cell morphology.

## 6.3 Confluent Clusters

In high-grade tumors, lymphocytes clump together. Our semantic segmentation approach treats a clump of 10 cells as one large region. While this yields an accurate Area Density (TIGER Score), it prevents an accurate individual Cell Count. Future work involves transitioning to Instance Segmentation (Mask R-CNN).

## 7 Conclusion

---

This case study demonstrates that Deep Learning can effectively automate TIL scoring. By combining U-Net with EfficientNet and a Focal-Dice loss strategy, we achieved a validation IoU of 0.48. This system addresses the workflow inefficiency and subjectivity inherent in manual pathology, offering a pathway toward reproducible precision medicine.