AUTOMATED DETECTION OF TUMOR-INFILTRATING LYMPHOCYTES IN BREAST CANCER HISTOPATHOLOGY USING DEEP LEARNING

The Problem:

Tumor-Infiltrating Lymphocytes (TILs) are immune cells present within and around breast tumors. They are a powerful biomarker because their presence reflects how strongly a patient's immune system is responding to the cancer. High level of TILs are associated with better prognosis and improved response to immunotherapy, especially in HR2-positive and Triple-Negative Breast Cancer

Traditionally, TILs are scored manually by pathologists using Hematoxylin and Eosin (H&E) stained histopathology slides This process is slow, subjective and difficult to scale because whole-slide images are extremely large and complex. Different pathologists may also assign different scores for the same slide, leading to variability and reduced reproducibility.

The TIGER Grand Challenge addresses this problem by proposing the use of AI to automatically detect lymphocytes and plasma cells, segment tumor and stroma regions, and compute an automated TIL score. This project focuses on the first task: automatic detention of lymphocytes and plasma cells using deep learning.

Why This Problem Matters:

Breast cancer is the most common cancer worldwide and a leading cause of cancer-related among women. HER2-positive andTriple-Negative brest cancers have particularly poor prognosis and require precise treatment planning.

Accurate TIL assessment help clinicians :
Predict patient survival
Decide whether immunotherapy will be effective
Reduce unnecessary chemotherapy
Personalize treatment strategies

However, manual TIL scoring is limited by human workload, subjectively, and shortage of trained pathologists. AI-based detection systems can provide: Faster diagnosis, Consistent scoring, Scalable solutions for hospitals, Improved patient outcomes.

Projects like TIGER and AIRAT show that AI is moving from research to real clinical deployment, making this a highly impactful healthcare application.

Approach:

Design a complete training pipeline to automatically detect lymphocytes and plasma cells in histopathology image patches using YOLOv8 object detection model.

PIPELINE OVERVIEW:

1. Dataset Preparation
    - Use annotated histopathology patches with bounding boxes for lymphocytes and plasma cells
    - Convert annotations from COCO format to YOLO format
    - Resize images to 512 x 512 for consistent training

2. Preprocessing
    - Image resizing and normalizing
    - Data augmentation
    - Train-validation split

3. Model
    - YOLOv8 object detection network
    - Pretrained backbone for transfer learning

This pipeline directly supports the first TIGER challenge task and forms the foundation for automated TIL scoring

4. Dataset

    I used the TIGER WSIRois dataset provides via Roboflow: https://universe.roboflow.com/xray-u9rf3/wsiroisimages/analytics

This dataset contains: Histo[athology image patches, Bounding box annotations for lymphocytes and plasma cells, COCO formatted labels

The dataset is designed for training detection models and closely reflects real clinical data.

5. Challenges and Observations
    - Small Object Size: Lymphocytes are tiny and densely packed, making detection difficult.
    - Stain Variability: Color differences across slides affect model generalization

- Class Imbalance: Lymphocytes appear much more frequently than plasma cells
- Large Image size: Whole-side images are extremely large, so patch-based learning is required.
- Overfitting Risk: Medical dataset are small, so careful augmentation and transfer learning are necessary