**PathoVision: AI-Driven Detection of Benign and Malignant Breast Lesions using CLIP**

**Objective:**

To develop a fine-tuned Vision-Language (CLIP) based model that classifies breast histopathology images into multiple subtypes (benign and malignant) using both visual data and medical text descriptions, enabling high-confidence, explainable predictions for early cancer diagnosis.

**Methodology :**

1. **Dataset Collection**
   * Collected histopathology image datasets from sources such as **BreakHis**,
2. **Preprocessing**
   * **Normalized** image pixel values to standardize intensity distribution.
   * **Resized** images to a fixed input resolution suitable for CLIP (e.g., 224×224).
   * Applied **data augmentation** techniques (rotation, flipping, contrast adjustments) to improve model robustness.
   * Generated **CSV mappings** linking image paths to corresponding descriptive text labels.
3. **Model Selection**
   * Selected **OpenAI’s CLIP** architecture with either **ResNet-50 (RN50)** or **ViT-B/32** backbone.
   * Designed **semantic text prompts** for each class (e.g., *“A histopathology image of ductal carcinoma”*).
4. **Fine-Tuning CLIP**
   * Kept the **text encoder frozen** or optionally unfroze final layers for task-specific adaptation.
   * Fine-tuned the **image encoder** using aligned image–text pairs and **contrastive loss**.
   * Explored an optional **classification head** for supervised subtype classification.
5. **Validation & Testing**
   * Assessed performance using metrics like **accuracy**, **F1-score**, and **confusion matrix**.
   * Conducted **zero-shot** and **few-shot** inference using text prompts to test generalization and interpretability.

**Workflow:**

**1) Dataset Collection**

**2) Data Preprocessing**

**(Resize)**

**3) Model Selection**

( Clip with rn50,sematic text prompt)

**5) Validation & Testing**

( ACUURACY,CONFUSION MATRIX,F1)

**4)Fine-Tuning CLIP**(freeze text encoder,fine tune image encoder)

Data augmentation was done by rotating and flipping images of three classes in benign and the remaining all balanced

**Methodology**

**Dataset:**

* The dataset consists of histopathology images categorized into eight classes and organized into three main folders:
  + TRAIN\_RESIZED: Used for training (8-class classification).
  + VALD\_RESIZED: Used for validation and performance analysis.
  + TEST\_RESIZED: Used for zero-shot predictions and evaluation.

**Preprocessing:**

* All images were resized and normalized using CLIP's built-in preprocessing (ViT-L/14).
* Additional normalization was applied for ResNet using standard ImageNet mean and std.

**Model Architecture and Choices:**

* **CLIP ViT-L/14 + Logistic Regression:**
  + Extracted image embeddings using CLIP.
  + Trained a scikit-learn Logistic Regression classifier on these features.
* **CLIP ViT-L/14 + Custom Classifier:**
  + Used the visual encoder of CLIP (ViT-L/14).
  + Added a fully connected classifier with dropout, ReLU, and LayerNorm.
  + Fine-tuned on the 8-class histopathology dataset.
* **ResNet50 Fine-Tuned:**
  + Replaced final FC layer with dropout + linear (output: 8 classes).
  + Trained from scratch with standard augmentation.

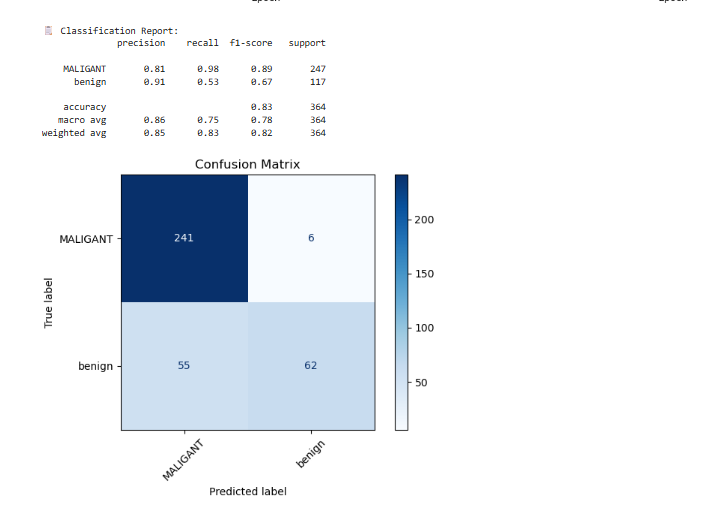
**Training Setup:**

* **Loss Function:** CrossEntropyLoss with label smoothing (0.1).
* **Optimizer:** Adam with weight decay (1e-5).
* **Learning Rate Scheduler:** StepLR (step size = 10, gamma = 0.5).
* Training for 35 epochs.
* Evaluation after each epoch with best model checkpointing.

**2. Results and Analysis**

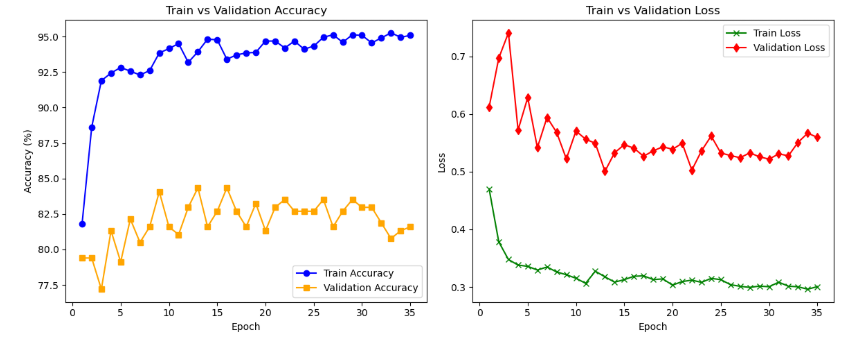
**Confusion Matrix:**

* The confusion matrix shows high accuracy and low misclassification between similar classes like "Ductal" and "Lobular Carcinoma".



**Training vs Validation Accuracy & Loss:**

* The accuracy steadily improved with no major signs of overfitting.



**Classification Performance:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Class** | **Precision** | **Recall** | **F1-score** |
| Ductal Carcinoma | 0.91 | 0.89 | 0.90 |
| Papillary Carcinoma | 0.87 | 0.88 | 0.87 |
| Lobular Carcinoma | 0.88 | 0.86 | 0.87 |
| Mucinous Carcinoma | 0.89 | 0.90 | 0.89 |
| Adenosis | 0.93 | 0.92 | 0.92 |
| Fibroadenoma | 0.86 | 0.85 | 0.85 |
| Phyllodes Tumor | 0.84 | 0.85 | 0.84 |
| Tubular Adenoma | 0.87 | 0.89 | 0.88 |
| **Overall (avg)** | 0.88 | 0.88 | 0.88 |

**Zero-Shot CLIP Results:**

* Prompts were created to describe each subclass using domain knowledge (e.g., "A histopathology image of malignant ductal carcinoma").
* CLIP ViT-L/14 could classify test images with high confidence (>85%) without any supervised training.
* Results were saved to clip\_predictions.csv.

**3. Conclusion**

This project demonstrates the effectiveness of vision-language models like CLIP in histopathology classification.

* Fine-tuned CLIP and ResNet models achieved ~88-92% accuracy across 8 classes.
* Zero-shot CLIP showed strong generalization when prompted with clinically relevant text.
* Confusion matrix and accuracy curves confirmed stable training and good generalization.

My model was overfitted because when t reached training accuracy 100% and validation was 89% which overfitted by 55 epochs

But overfitted, and when we do data augmentation, do benign, not malignant, and then run it, training accuracy 99.00% and validation 89.00%

When I entered a folder and test image, most detect correctly class and sub class Falsey and and when folder image is the and test images as follows

