Project Report: Applying Pre-trained CLIP Model for Chest X-ray Disease Diagnosis

Introduction

This project applies the pre-trained **CLIP (Contrastive Language–Image Pretraining)** model to diagnose chest diseases using X-ray images. CLIP's ability to link images and text makes it suitable for tasks like **classification, retrieval, and disease detection** in medical imaging. We evaluated the model on the ChestX-ray14 dataset, which includes multi-label annotations for pathologies, using standard performance metrics.

---

1. Dataset Preparation

Dataset Used

- **ChestX-ray14**: Contains chest X-ray images labeled with 14 pathologies, including:
  - **Atelectasis**
  - **Cardiomegaly**
  - **Effusion**
  - **Pneumonia**

Preprocessing

1. **Data Loading and Splitting**:
   - Loaded image paths and labels from a CSV file.
2. **Image Preprocessing**:
   - Converted images to **RGB** format.
   - Resized images to **224x224** (CLIP's standard input size).
   - Applied normalization using predefined mean and standard deviation.
3. **Text Preprocessing**:
   - Converted disease labels to textual descriptions (e.g., *"A chest X-ray showing Cardiomegaly"*).
   - Used **CLIP Processor** to generate text embeddings.

---

2. Model Setup

CLIP Architecture

CLIP is a multimodal model linking images and text via **contrastive learning**. Key components:

1. **Visual Encoder**:
   - **ViT-B/32** (Vision Transformer base model).
2. **Text Encoder**:
   - Transformer-based NLP model.

Fine-tuning for Medical Tasks

1. **Pre-trained Model**:
   - Used OpenAI's openai/clip-vit-base-patch32.
2. **Loss Function**:
   - **BCE With Logits Loss** (for multi-label classification).

---

3. Disease Diagnosis Implementation

Supported Tasks

1. **Classification**:
   - Predict multiple pathologies in a single X-ray image.
2. **Retrieval**:
   - Match new images to disease descriptions.
3. **Zero-shot Prediction**:
   - Diagnose diseases using text prompts without fine-tuning.

Training and Evaluation

1. **Training**:
   - Optimizer: **Adam** (learning rate = 1e-4).
   - Epochs: 15 and 10 images (samples).

2. **Evaluation Metrics**:
    - **Precision, Recall, F1-Score** (per class).
    - **Confusion matrices** for each pathology.

Key Observations

- Fine-tuning significantly improved performance over zero-shot inference.
- Model struggled with **class imbalance** (rare diseases had lower recall).

---

5. Challenges and Future Improvements

Challenges

1. **Data Quality**: Noisy or inaccurate labels in some images.
2. **Computational Cost**: CLIP's large size requires high resources.

Proposed Optimizations

1. **Medical-Specific Models**: Implement **MedCLIP** (a CLIP variant optimized for medical imaging).

---

6. Conclusion

- CLIP achieved promising results in chest X-ray diagnosis after fine-tuning.
- Addressing data quality and class imbalance remains critical for clinical applicability.

Note: MedCLIP, a CLIP variant tailored for medical data, is under exploration as part of future optimizations to enhance diagnostic performance.

Toka AL-Dagamsih 163240
Shahed Rawabdeh 163352