Pet skin disease diagnosis chatbox - V1

先做图像分类

然后有时间的话做文字：  
想法1.用图像分类出来的准确结果输入gpt api给treatment

想法2.图像分类出来的几个可能的结果，输入gpt api，让他生成问题，以多轮对话的形式跟用户确认，生成最终的诊断建议和建议

**Overview**

This project aims to develop an AI-powered chatbot that helps pet owners diagnose potential skin diseases in their pets by analyzing both images and text descriptions. By leveraging pre-trained models and APIs, the chatbot provides users with a preliminary diagnosis and actionable recommendations based on veterinary knowledge.

**Key Features**

✅ **Multimodal Input:** Users can upload pet skin images and describe symptoms in text.

✅ **AI-Based Image Analysis:** Uses Google Vision API / Hugging Face models to classify skin diseases.

✅ **Natural Language Understanding:** Uses GPT-4 / Mistral API to understand user descriptions and generate proper suggestions.

✅ **Retrieval-Augmented Generation (RAG):** Enhances AI responses by retrieving relevant veterinary literature/ forum/ textbooks.

**Development Plan**

**Phase 1: Image-Based Disease Classification (1 Week)**

**Expected Output:** Given an input pet image, the system will classify the most likely skin disease (e.g., fungal infection, bacterial infection, allergies) with a confidence score

**Approach**

1. Enhance the dataset: Increase dataset diversity using data augmentation (rotation, flipping, color jitter, random cropping) to improve model generalization.
2. Use API-based solutions: Leverage Google Vision API / Hugging Face models to classify pet skin diseases without training a custom model for faster implementation.

**Datasets:**

1. <https://github.com/OPet-OnlineDiagnosisforYourPet-sDisease/MachineLearning/blob/main/dataset/dogSkinDisease.zip>
2. [Classification of pet dog skin diseases using deep learning with images captured from multispectral imaging device](https://data.mendeley.com/datasets/5dbht54kw7/1)
3. <https://data.mendeley.com/datasets/5dbht54kw7/1>
4. Dog Disease Images & projects: <https://universe.roboflow.com/search?q=dog%2520disease>
5. A small Image dataset: <https://www.kaggle.com/datasets/yashmotiani/dogs-skin-disease-dataset>

**Phase 2: Text-Based Diagnosis + RAG**

**Expected Output:** Given a text input (e.g., “My dog has red, itchy skin”), the system retrieves relevant veterinary articles and combines them with LLM-generated insights to provide a diagnostic suggestion and care recommendation.

**Approach:**

1. Use an LLM API (GPT-4 / Mistral) to process text inputs
2. Retrieve veterinary knowledge from a structured database (FAISS / ChromaDB)
3. Enhance chatbot responses by combining retrieved knowledge with LLM generation

**Datasets:**

1. [9 Common Skin Problems in Dogs: Causes, Symptoms & Treatments](https://www.dogseechew.in/blog/9-common-skin-problems-in-dogs-causes-symptoms-treatments)
2. <https://veterinaryevidence.org/index.php/ve/catalog/category/dogs>
3. <https://www.primarycarevet.com/pet-medical-conditions/> (easy to scrap)
4. <https://www.merckvetmanual.com/> (easy to scrap)

**Phase 3: Multimodal Fusion (Image + Text)**

**Expected Output:** If a user uploads an image AND describes symptoms, the system will cross-analyze both sources to provide a more accurate diagnosis.

**Approach:**

1. Use BLIP-2 / CLIP models to generate image captions
2. Pass both image captions + user descriptions into GPT-4 / Mistral for combined reasoning
3. Enhance RAG retrieval by matching image embeddings + text embeddings

**Phase 4: Deployment**

**Expected Output:** Deploy the multimodal AI model as an API with a web-based frontend.

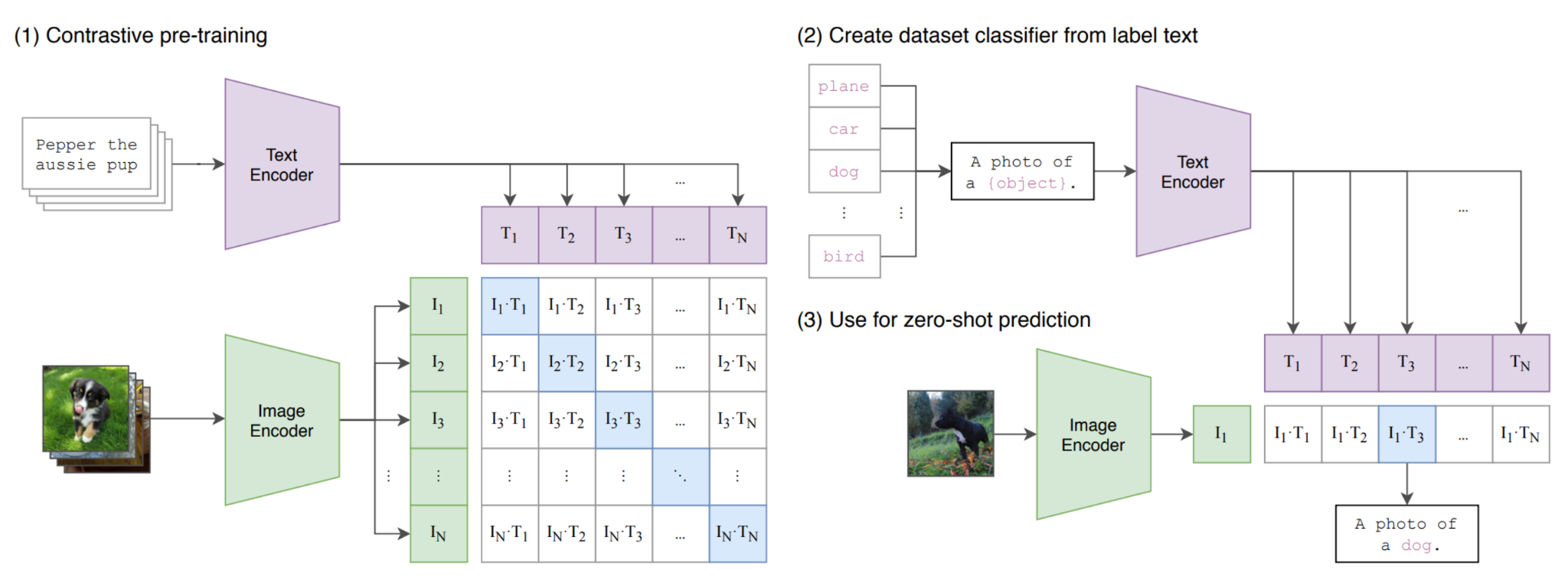
**Approach:**

1. Build API using FastAPI / Flask
2. Deploy models (CV + LLM + RAG) using Hugging Face Spaces / AWS Lambda
3. Create a simple UI using Streamlit / Gradio

**Questions:**

1. Can we first train the image and text models separately and then integrate them into a multimodal system?
   1. How good is the result compared to the baseline
2. Does this course project focus more on model performance, or is the expectation to develop a complete product?
3. Can we use external resources? In addition to open source models or apis, can we use chatbots or front-end frameworks built by others

* Dataset,



Proposal 1: Input Image → Train a Classification Model Using CLIP Multimodal

CLIP:

The text encoder in the CLIP pre-trained model is trained on general datasets, making it highly effective in recognizing objects (e.g., "this is a dog," "this is a cat"). However, the pre-trained CLIP model is not precise enough to distinguish specific "dog diseases."

Using the pre-trained model directly → results would be poor.

What we can do: Fine-tune the model using our own images → significantly improved classification performance.

Text Encoder:

Example text inputs for classification:

* “A photo of Disease 1”
* “A photo of Disease 2”
* “A photo of Disease 3”
* “A photo of Disease 4”

**提案二：input 圖片 + 文字敘述 -> 訓練分類模型**

1. **資料前處理：**

* 資料集不夠：
* data augmentation (saturation, rotation, crop, image color)
* or 會不會類似犬科的皮膚疾病是相近的（狼、狗）

1. Image 轉feature: pre-trained vision model e.g. CLIP
2. Text轉feature: 用api
3. 把image feature & text feature concat
4. 訓練模型 （中間的linear layers層數少一點、因為資料集很少）

-> 輸出一段feature，告訴我們狗是生哪一種病， 最後用機率去判斷是哪種疾病

1. 生成建議：可能要用 RAG 因為目前的LLM可能沒有那麼多醫療文本的訓練（要自己用醫療期刊文本）

### **Proposal 2: Input Image + Text Description → Train a Classification Model**

#### **Data Preprocessing:**

* **Dataset size is insufficient:**
  + Apply data augmentation (e.g., saturation adjustment, rotation, cropping, image color modification).
  + Consider whether similar skin diseases among canines (wolves, dogs) exhibit comparable patterns.

#### **Feature Extraction:**

* **Image to Feature:** Use the pre-trained vision model, such as CLIP.
* **Text to Feature:** Use the pre-trained model
* **Feature Fusion:** Concatenate image and text features.
* **Model Training:**
  + Use a classifier with fewer linear layers (since the dataset is small).
  + Output a feature vector that determines the type of disease affecting the dog.
  + Use the model result (logits) to identify the specific disease.

#### **Generating Recommendations:**

* **Possible need for Retrieval-Augmented Generation (RAG)**
  + Current large language models (LLMs) may lack sufficient medical text training.
  + Additional training on medical journal texts may be required.