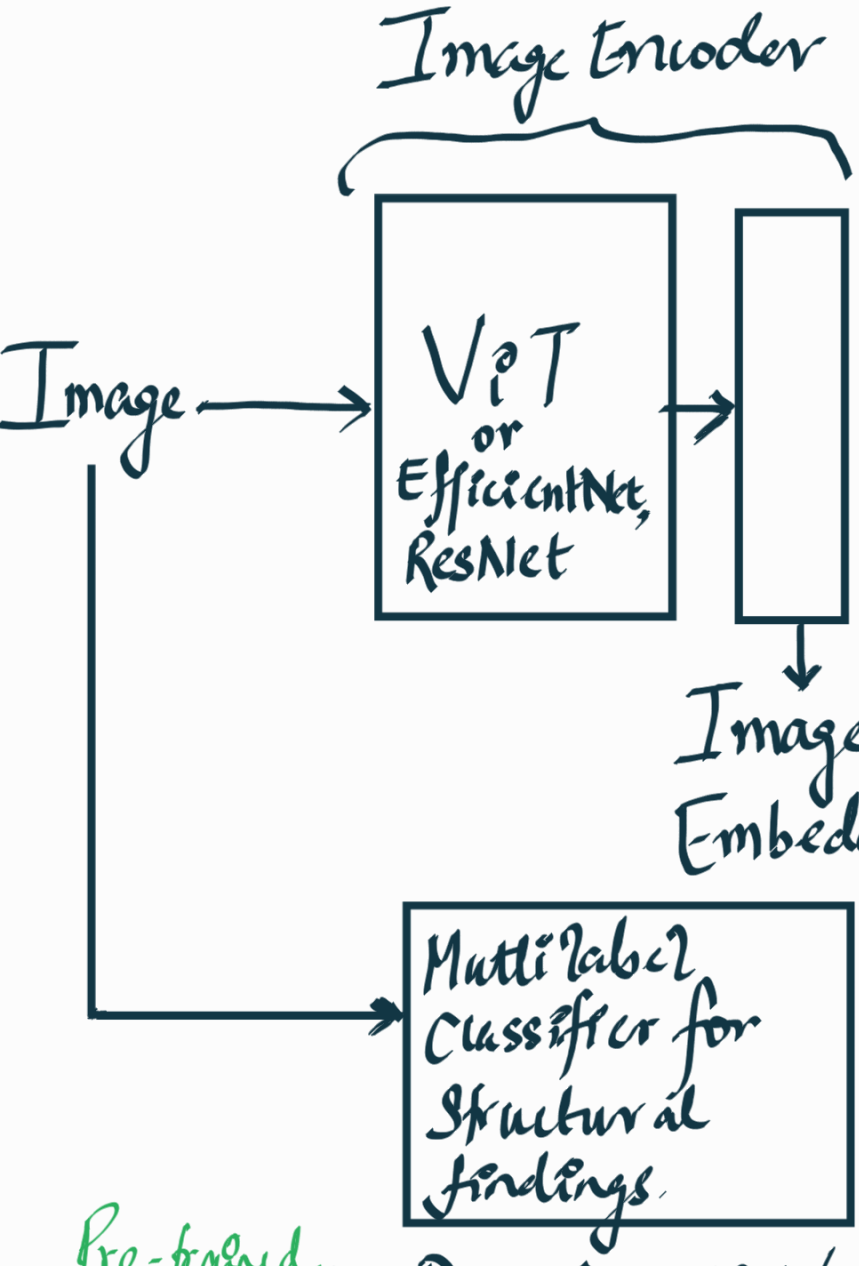


How we use the image embeddings?

(1) Create a shared space for image and text embeggings.

(2) Use a multi-modal which is pre-trained to align images & text.

Aim: To convert Image Embeddings to Text Embeddings.

These text embeddings will be used in LLM along with labels we get from structural findings to generate medical Report.

Needs Transfer Learning on our dataset.

Computationally Expensive & requires clinical Validation.

Vision Transformer

Multi label Classification for each image.

a medical [illegible]

## Case 1: Create a shared space

Use another model like BERT or medical variants of BERT like BioBERT, ClinicalBERT.

## Case 2: Use multimodel which generate text embeddings

We can use multimodels like MedCLIP to directly generate text embeddings.

Other ways: Generate Image embeddings & use a Multimodel that can understand image embeddings directly and generate medical report.

Prompt construction:
Image features + Multi labels

## LLM for report generation

Use LLMs trained on medical data to create medical reports, which can

be fine-tuned on our data.
(Bio-MedTS, Med-PaLM)

## Performance Metric:

We need to choose this carefully,
could be BLUE, ROUGE, BERTScore etc.

## Data (Can use same data)

TCGA-GBM Dataset (Includes labels & Impressions)

CQ500 Dataset

Intracranial Hemorrhage Detection

Vin Dr-CXR and Vin Dr-Mammo.

Datasets on X-rays:

CheXpert Plus
ChesX-ray14
Pad Chest.