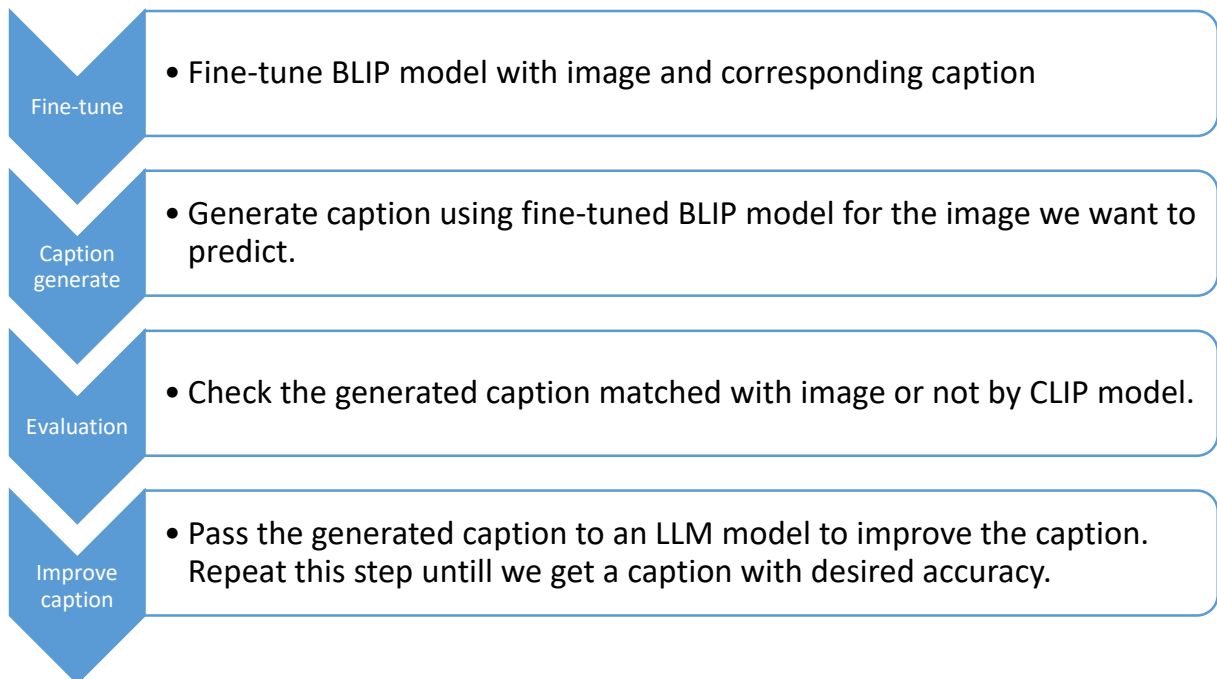


# Cancer prediction by VLM

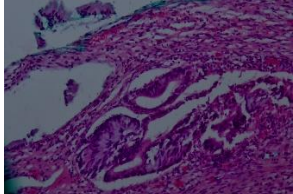
Dr. Debashis Bhowmik

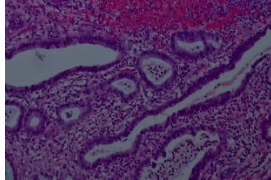
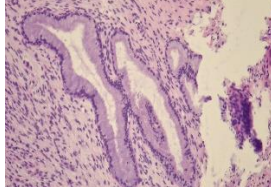
Many people studied cancer prediction to detect malignant patterns in histopathology slides from tissue images. Most have used deep learning models, particularly CNN to extract features from the image. But, CNN will check whether it is malignant. If we have labeled with cancer stage then we can predict the stage at most and nothing more than that. Here I will propose one method that not only extracts the information of malignancy and stage of it but also some other information if it has. For that, we will use the vision language model (VLM).

VLM can be used for many purposes, but here I will use it for image captioning. However, any open-source model for image captioning is trained on general data. As I want in a particular domain, therefore, I need to fine-tune the model. After fine-tuning the model, we can pass an image that I want to detect and it will return a description of the image which is called a caption. Now, I can check how accurate the generated caption is by a text-image model. If the accuracy is not up to the mark, then I can pass the caption to an LLM model to improve the caption. Check the accuracy of the caption with the image, and repeat this method until reach the desired accuracy. Finally, we will get the final caption.



In my case, images and corresponding descriptions look like the following:

Image	Description
	Cervical histopathology image showing abnormal squamous cells

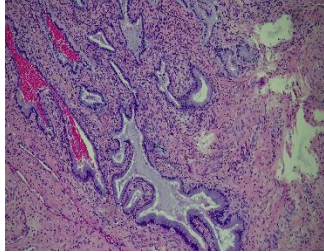
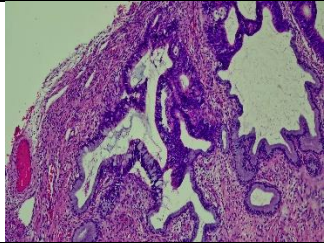
	Histopathology scan revealing early-stage cervical carcinoma
	Cervical biopsy slide showing no malignant cell clusters

Note that, in the original data set it is only known whether a particular image is malignant or not. I randomly wrote the description.

Setup: (Because of the limited resource)

- ❖ Reduce the image shape to (16,16)
- ❖ Fine-tune the BLIP model with
  - 20 images + description data
  - per\_device\_train\_batch\_size=1
  - num\_train\_epochs=1
  - per\_device\_eval\_batch\_size=1
  - gradient\_accumulation\_steps=2,
  - fp16=True
  - save\_total\_limit=1
- ❖ Used **gpt-4o-mini** LLM model to improve caption.

Sample output:

Image	Description
	Histopathological image of cervical tissue demonstrating cellular abnormalities suggestive of potential pre-cancerous changes. Key features highlighted in the image include atypical cell morphology, an elevated nuclear-cytoplasmic ratio, and variations in cellular architecture, all of which are crucial for accurate diagnosis and assessment of risk."   <b>Accuracy: 0.3218</b>
	Detailed overview of cervical histopathology, emphasizing key cellular structures such as squamous epithelial cells, glandular cells, and stromal components, while identifying potential abnormalities such as dysplasia, HPV infection, and neoplastic changes that may indicate specific cervical conditions."   <b>Accuracy: 0.3059</b>

You can find the code in my [GitHub](https://github.com/Debashis9564/Cancer-prediction-by-VLM) page: [Debashis9564/Cancer-prediction-by-VLM](https://github.com/Debashis9564/Cancer-prediction-by-VLM)

I believe if the training dataset size is large and the corresponding description is almost accurate, and if I have all the resources then we don't have to compress the image and we can take a good batch and epoch size. Then the predicted accuracy will be good.