LviT: VISION TRANFORMER FOR LUNG CANCER DETECTION

MALIKA HAFIZA PASHA

Presentation Flow:

- Introduction
- Problem Statement
- Proposed Method
- Technological Stack
- Workflow
- Architecture
- Result and Discussion
- Confusion Matrix
- Conclusion
- References

Introduction

Lung Cancer is one of the leading causes of mortality for males and females worldwide with more than 2.29 lakh new cases being identified each year.

The ongoing advancement of technology will aid in identification, treatment, and clinical outcomes which in return lessen the mortality rate and increase the chances of survivability.

Machine Learning techniques are very advantageous for the detection and forecasting of cancer growth and progression.

In the study, We used vision Transformer model using Computer tomography(CT) scans for the identification of cancer.

Problem Statement

- Lung Cancer is the leading cause of death for both men and women. Within five years of receiving their diagnosis, 75% of those people perish away.
- Careful examination and surveillance are required for lung nodules that are still growing. The high intra-tumor heterogeneity (ITH) and complexity of cancer cells that cause drug resistance make cancer therapy more challenging.
- Machine Learning plays crucial in automated detection, segmentation and computer-aided diagnosis of malignant lesions.
- Using Machine Learning methods for detection of cancer will aid doctors make decisions and give remarkably beneficial results.

Proposed Method

Dataset and paring The Training Set – The lung cancer dataset from the National Centre for Cancer Diseases (NCCD) and Iraq-Oncology Teaching Hospital (IQ-OTH/NCCD) was used.

Vision Transformer – The guiding principles of the transformer used in the field of national language processing serve as the foundation for the vision transformer (Vit), a transformer used in the field of computer vision.

Experimental Settings - In our trials, we constructed the ViT model using Python, Google Collab, and TensorFlow extensions. Images used as input are normalized via a batch normalization layer.

Workflow of Vision Transformer

- The working of the Vision Transformer (ViT) can be understood through the following

work flow:-

1. Create patches from an image.
2. Flatten the patches.
3. Make flattened patches into lower-dimensional linear embeddings.
4. Add the positional embeddings.
5. Feed the sequence into a typical transformer encoder as an input.
6. By using image labels, prepare the model (supervised on a huge dataset).
7. Fine-tune the image classification dataset's downstream dataset.

Architecture of Vision Transformer :

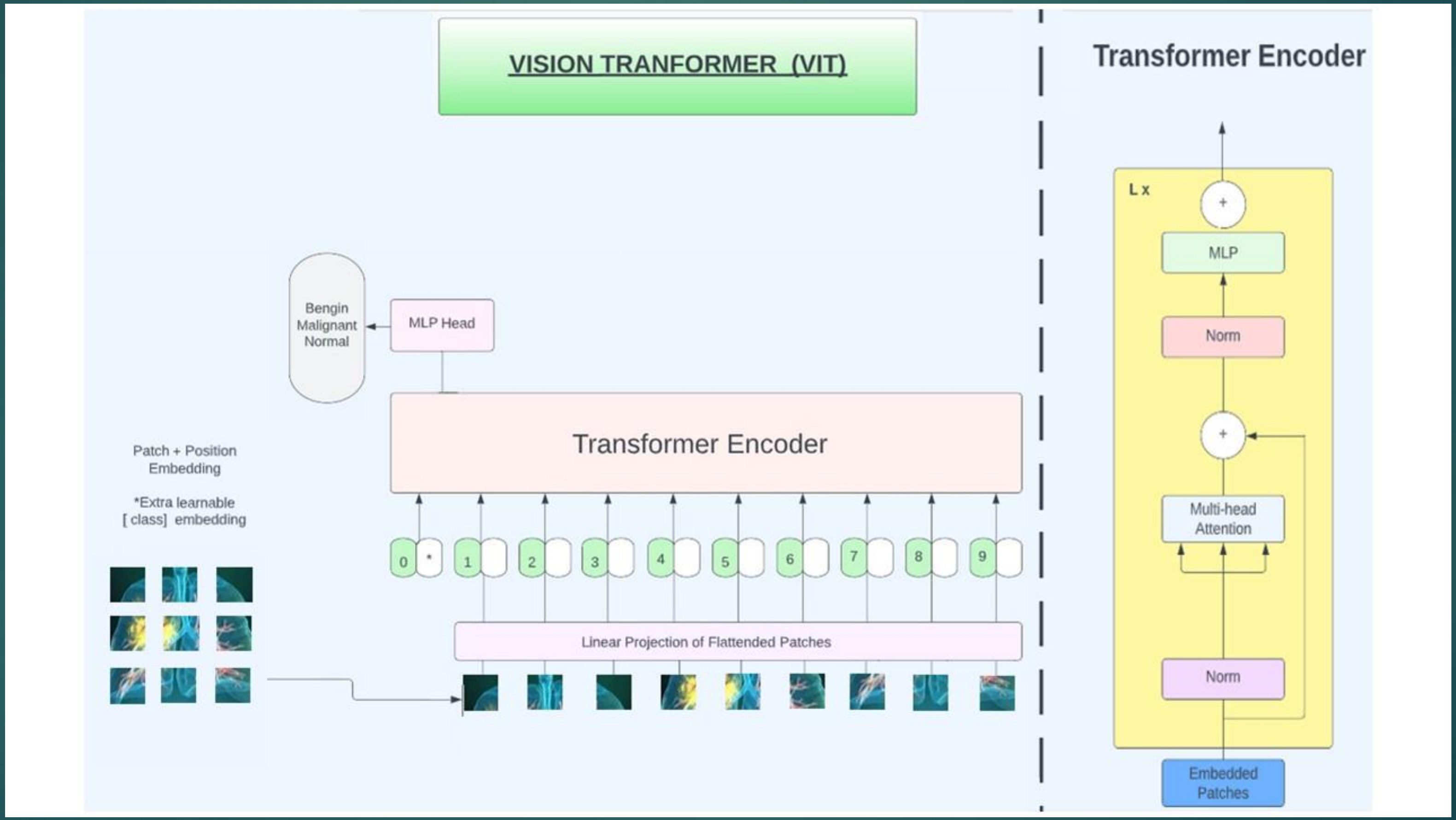

Fig 1. The architecture of LViT

How Multi-Head Attention Works in a vision Transformer?

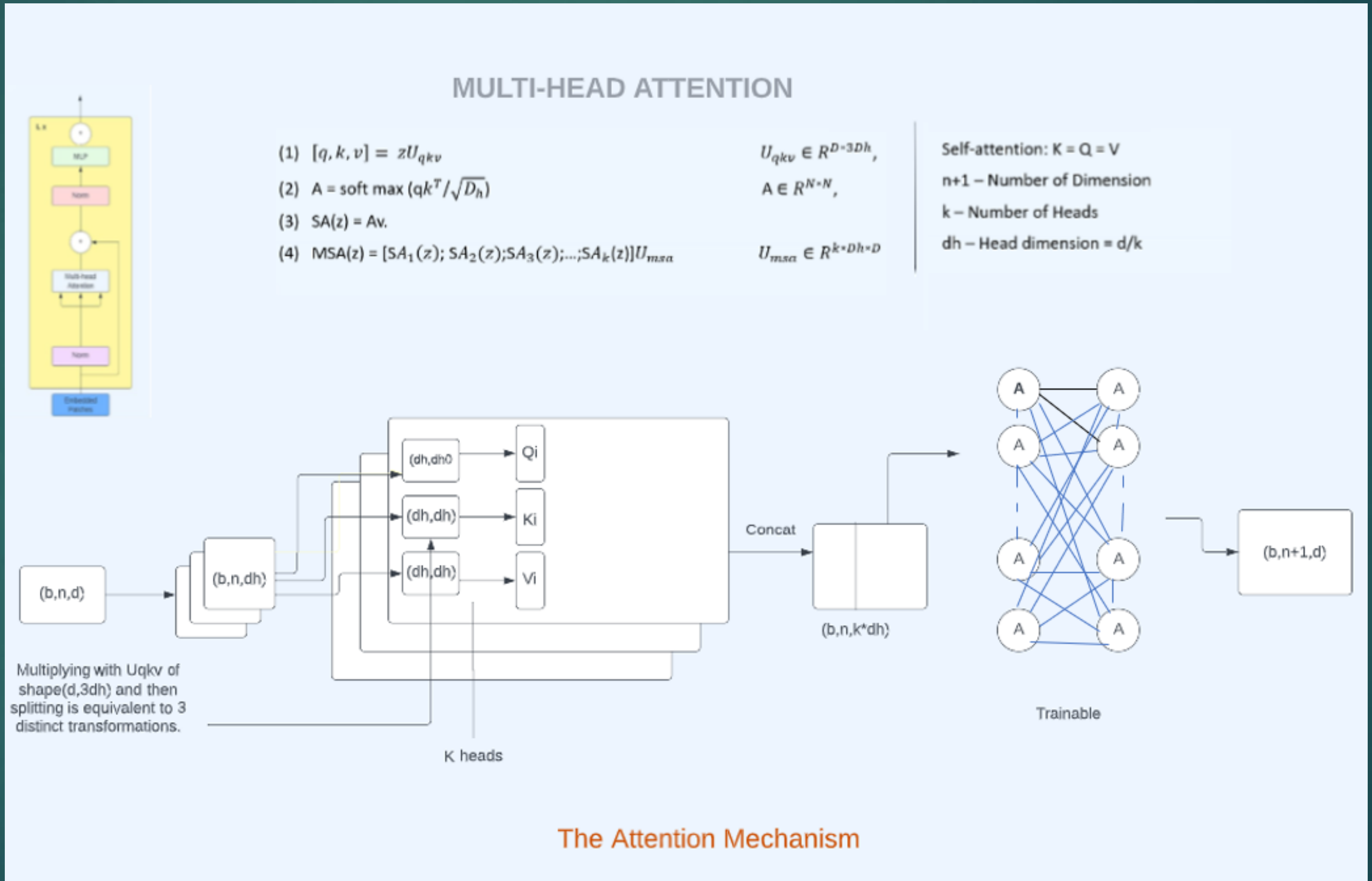

Fig 2. The Attention Mechanism in the MLP

Results And Discussions

MODEL NUMBER	NUMBER OF HEADS	TEST ACCURACY	LOSS
LViT (h=6)	6	78.03%	0.6418
LViT (h=8)	8	81.61%	0.6747
LViT ((h=10)	10	78.92%	0.7077
LViT (h=8)	8	91.93%	0.2816

Fig 3. Maximum Accuracy achieved = 91.93%

RESULTS AND DISCUSSION CONTINUE…

- The very first model trained built using a single epoch and 4 heads resulted in 61.43% test accuracy, which is very low for the prediction models using imaging techniques, especially in the medical sector.
- In the further stages of the research, an increase in the validation accuracy was found when the number of epochs was increased, so for obtaining better results, we increased the number of epochs while training the model, this resulted in the drastic improvement of the prediction models in classifying the type of lung cancer detected through the image segments through the Vision Transformer.
- Furthermore, when the number of heads was increased, the trend of increase in the model accuracy was observed which resulted in improved validation accuracy with the **best accuracy being 91.93%.**

Confusion Matrix

Benign	0.51	0.058	0.43
Malignant	0.0036	0.99	0.0089
Normal	0.039	0.021	0.94
	Benign	**Malignant**	**Normal**

Conclusion

The most frequent cause of mortality and one of the most severe diseases, lung cancer is made more hazardous by the challenge of making an early diagnosis.

The development of lung cancer is foreseen in our work using the Vision Transformer. The suggested method makes optimal use of segmentation concepts for feature extraction from a collection of CT images before applying a vision transformation model.

The proposed approach demonstrates that doctors may effectively use them to help in the detection of lung cancer and give remarkably beneficial results.

If the prediction is true, the doctor may be able to provide a more effective medication and diagnose the patient more quicker.

References

[1] Makaju, S., Prasad, P.W., Alsadoon, A., Singh, A.K., & Elchouemi, A. (2018). Lung Cancer Detection using CT Scan Images. *Procedia Computer Science, 125*, 107-114.

[2] Parthasarathy G., Abirami S., Monica Santhana A.,Nishali C., Pavithrasrisai V., (2020) Prediction of Lung Cancer Using Deep Learning Algorithm, International Journal of Creative Research Thoughts 8(4) 2003-2008

[3] Krishnaiah, V.V., Narsimha, G., & Chandra, N.S. (2013). Diagnosis of Lung Cancer Prediction System Using Data Mining Classification Techniques.

[4] Nasser, Ibrahim M. and Abu-Naser, Samy S., Lung Cancer Detection Using Artificial Neural Network (March 2019). International Journal of Engineering and Information Systems (IJEAIS), 3(3), 17-23, March 2019, Available at SSRN: https://ssrn.com/abstract=3369062

[5] Jakimovski, G., & Davcev, D. (2019). Using Double Convolution Neural Network for Lung Cancer Stage Detection. *Applied Sciences*, *9*(3), 427. https://doi.org/10.3390/app9030427

THANK YOU