Visual Question Answering for Medical Images with Explainable AI

Deepananth K 195001027
Jayakrishnan S V 195001040

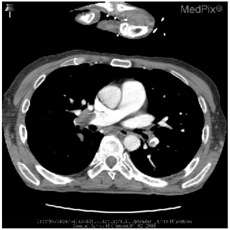
Under the Guidance of
Dr. S. Kavitha

SSN College of Engineering, Chennai

Third Review
April 24, 2023

What is Visual Question Answering and Explainable AI

- Visual Question Answering (VQA) is an emerging approach under the domains of Computer Vision and Natural Language processing, which aims to answer the user's question by analysing the given input image.
- Explainable AI (XAI) is a rapidly emerging research idea which refers to methods and techniques that helps us understand and interpret predictions made by AI models.



(g) **Q**: which organ system is shown in the ct scan? **A**: lung, mediastinum, pleura

(h) **Q**: what is abnormal in the gastrointestinal image? **A**: gastric volvulus (organoaxial)

Figure: Sample Images and Questions with Corresponding Answers from ImageCLEF 2019 VQA-Med Dataset

Problem Statement & Objectives

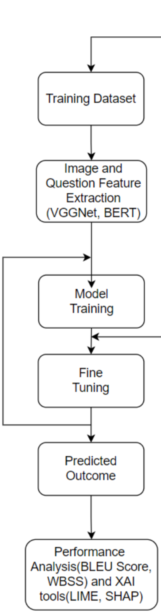
Problem Statement

- The aim of this project is to build an efficient VQA model that generates answers to questions related to medical images using deep learning techniques
- In addition, the reason behind the generated answer has to be analyzed using Explainable AI (XAI) tools such as LIME or SHAP to provide explanations on the outcome.

Objectives

- To collect and analyze the dataset of Visual Question Answering from the CLEF Forum.
- To build an efficient VQA model that generates the answers to the questions related to the given Medical Image using BERT
- To analyze the generated answer by using Explainable AI tools for providing explanations.

System Design & Module Split up



Dataset

In this Project, ImageClef 2019 VQA-Med dataset is used.

Dataset	Images	Question Category	No. of Questions	No. of Classes
Training	3200	Modality	3200	44
		Plane	3200	15
		Organ System	3200	10
		Abnormality	3192	1484
		Total	**12792**	-
Validation	500	Modality	500	-
		Plane	500	-
		Organ System	500	-
		Abnormality	500	-
		Total	**2000**	-
Testing	500	**All**	**500**	-

Table: Dataset Analysis

Performance Analysis

The performance of the VQA Model is analyzed using metrics such as accuracy, BLEU Score and WBSS. Table 2 shows the performance of the VQA model for each categories and overall test data.

Category	No. of Samples	Accuracy	BLEU Score	WBSS
Modality	125	65.6	68.79	71.66
Plane	125	64.8	64.8	65.35
Organ	125	50.4	53.19	54.82
Abnormality	125	6.4	7.65	12.03
Overall	**500**	**46.8**	**48.61**	**50.97**

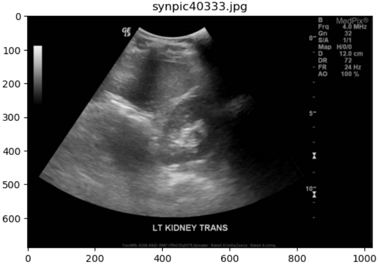
Table: Performance analysis using Accuracy, BLEU Score and WBSS

Model	Accuracy	BLEU Score	WBSS
Task Winner	62.4	64.4	–
Proposed Model	46.8	48.61	50.97

Table: Performance Comparison with task winner

Answer Generation

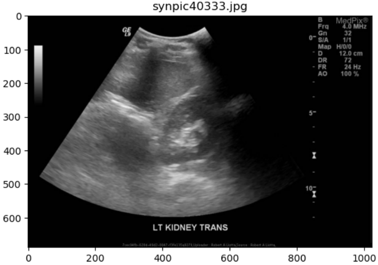
```
generateAnswer('synpic40333','what imaging modality was used to take this image?')
```




```
Generating Answers...
us
ultrasound
[SEP]

'us ultrasound '
```

```
generateAnswer('synpic40333','what organ system is imaged?')
```



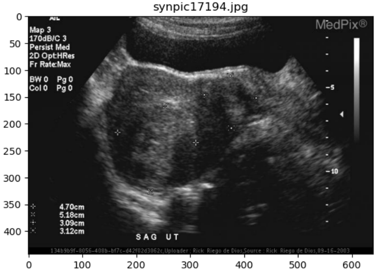

```
Generating Answers...
gastrointestinal
[SEP]

'gastrointestinal '
```

Figure: Sample Answer Generation of images of ID: synpic40333 queried about modality and organ. Both the answers are correct

Answer Generation Contd.

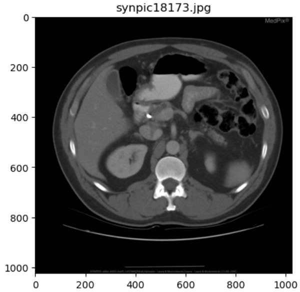
```
generateAnswer('synpic17194','what is the plane shown in this image?')
```




```
Generating Answers...
sagittal
[SEP]

'sagittal '
```

```
generateAnswer('synpic18173','what is the primary abnormality in this image')
```




```
Generating Answers...
pancreatic
adenocarcinoma
[SEP]

'pancreatic adenocarcinoma '
```

Figure: Sample Answer Generation of images of ID: synpic17194 and synpic18173 queried about plane and abnormality. The plane is correct but the abnormality is wrong. Actual answer is **pancreatic duct adenocarcinoma**

XAI - LIME Input & Perturbed Dataset

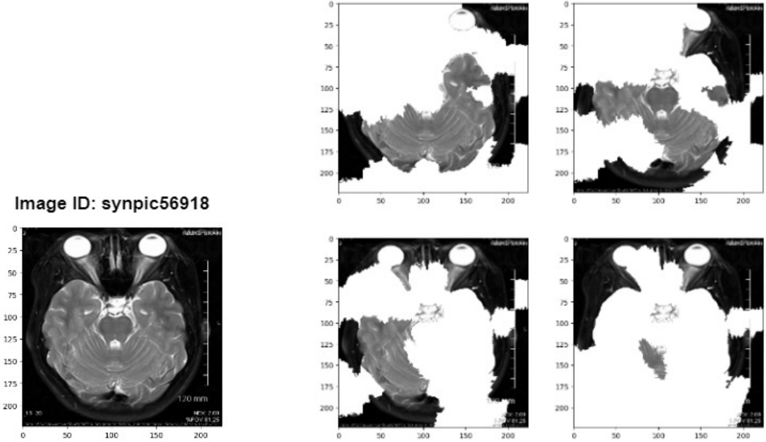


Figure: Sample input for LIME (Image ID: synpic56918) and Perturbed dataset

XAI - LIME Explanations

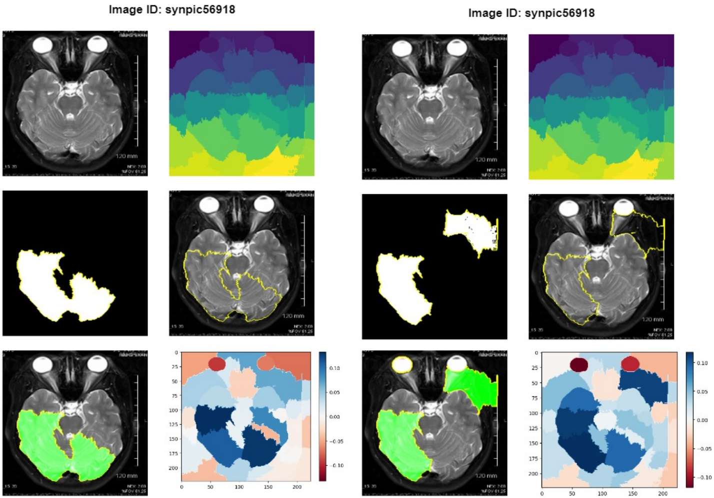


Figure: LIME explanations for a sample image with ID: synpic56918 queried about the organ

XAI - LIME Explanations Contd.

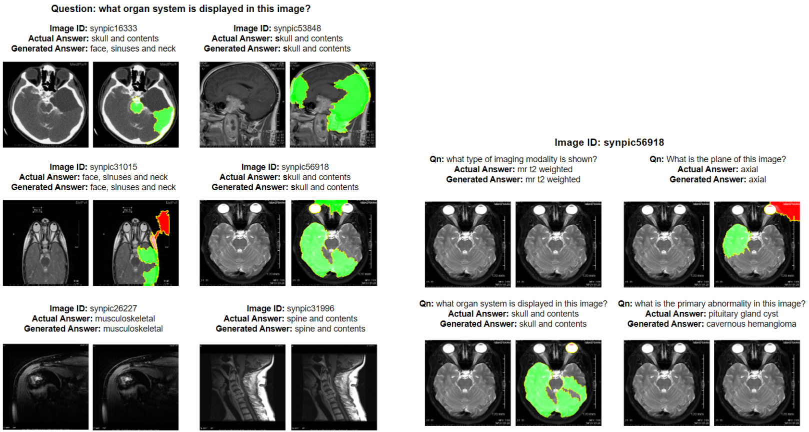


Figure: LIME explanations for a set of sample images queried about the organ & for a single image with 4 different questions

XAI - SHAP Explanations

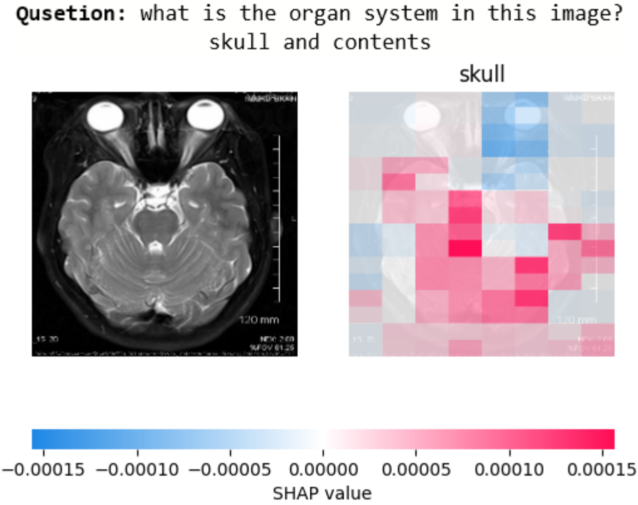


Figure: SHAP Explanations for a sample image with ID: synpic56918 queried about organ

XAI - SHAP Explanations Contd.

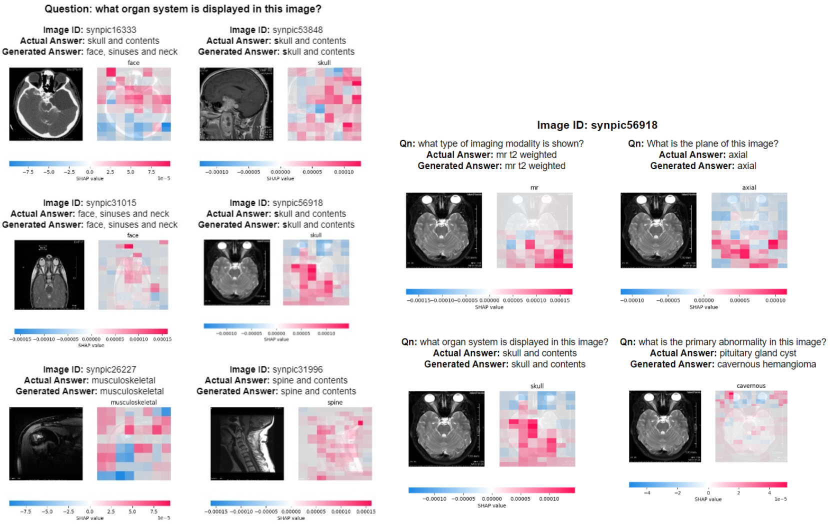


Figure: SHAP Explanations for a set of sample images queried about the organ & for a single image with 4 different questions

Conclusion

- The dataset for the task of Visual Question Answering is collected and it is analyzed
- The image and question features are extracted using VGGNet and BERT. The features are fused by concatenation.
- A BERT model is built and trained for Answer Generation by Masked Language Modeling
- The trained model is used for generating answers for the test data which resulted in 46.8% Accuracy, 48.61 BLEU Score and 50.97 WBSS Score.
- The performance of the trained VQA model is low for Abnormality based question, due to the lack of data enough for 1484 classes of abnormality.
- XAI techniques like LIME and SHAP are applied to generate explanations for the predicted outcome.
- In future, the VQA model can be fine tuned for increasing the accuracy of abnormality based question

References

[1] Lin Z, Zhang D, Tac Q, Shi D, Haffari G, Wu Q, He M, Ge Z. *Medical visual question answering: A survey*. arXiv preprint arXiv:2111.10056. 2021 Nov 19.

[2] Asma Ben Abacha, Sadid A. Hasan, Vivek V. Datla, Joey Liu, Dina Demner-Fushman, Henning Müller. *Overview of the Medical Visual Question Answering Task at ImageCLEF 2019*. CEUR-WS. 2019 Sep 9.

[3] Al-Sadi, Aisha & Al-Ayyoub, Mahmoud & Jararweh, Yaser & Costen, F.. (2021). *Visual Question Answering in the Medical Domain Based on Deep Learning Approaches: A Comprehensive Study. Pattern Recognition Letters.* 150. 10.1016/j.patrec.2021.07.002.

[4] Allaouzi, Imane et al. *An Encoder-Decoder Model for Visual Question Answering in the Medical Domain*, Conference and Labs of the Evaluation Forum (2019).

[5] Sharma, D., Purushotham, S. & Reddy, C.K. *MedFuseNet: An attention-based multimodal deep learning model for visual question answering in the medical domain.* Sci Rep 11, 19826 (2021).

[6] Deepak Gupta, Swati Suman, Asif Ekbal. *Hierarchical deep multi-modal network for medical visual question answering*,Expert Systems with Applications, Volume 164, 2021, 113993, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2020.113993.

References contd.

[7] F. Ren and Y. Zhou, *CGMVQA: A New Classification and Generative Model for Medical Visual Question Answering*, in IEEE Access, vol. 8, pp. 50626-50636, 2020, doi: 10.1109/ACCESS.2020.2980024.

[8] U. Pawar, D. O'Shea, S. Rea and R. O'Reilly, *Explainable AI in Healthcare*, 2020 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA), 2020, pp. 1-2, doi: 10.1109/CyberSA49311.2020.9139655.

[9] Knapič S, Malhi A, Saluja R, Främling K. *Explainable Artificial Intelligence for Human Decision Support System in the Medical Domain*. Machine Learning and Knowledge Extraction. 2021; 3(3):740-770.

[10] Ramprasaath R. Selvaraju, Abhishek Das, Ramakrishna Vedantam, Michael Cogswell, Devi Parikh, Dhruv Batra. *Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization*. International Journal of Computer Vision, vol. 128(2):336–359, Springer Science and Business Media, October 2019.

[11] S. H. P. Abeyagunasekera, Y. Perera, K. Chamara, U. Kaushalya, P. Sumathipala and O. Senaweera, *LISA : Enhance the explainability of medical images unifying current XAI techniques*. 2022 IEEE 7th International conference for Convergence in Technology (I2CT), Mumbai, India, 2022, pp. 1-9, doi: 10.1109/I2CT54291.2022.9824840.

Thank You

Literature Survey

Paper Title	Methodology	Limitations
CGMVQA: A New Classification and Generative Model for Medical Visual Question Answering [**CGMVQA**]	**Dataset**: ImageCLEF 2019 VQA-Med **Image Feature Extraction**: ResNet-152 **Question Feature Extraction**: BERT Tokenizer **Answer Generation**: BERT **Analysis**: Accuracy-62.4, BLEU Score-64.4	The proposed solution for VQA is building different models for different types of question such as Modality, Plane, Organ and Abnormality. But in reality, the exact type of a question may not be known.
Visual question answering in the medical domain based on deep learning approaches: A comprehensive study [**Al-Sadi**]	The Questions are classified into 4 categories and multiple models are trained for each type of question **Dataset**: ImageCLEF 2019 VQA-Med **Image Feature Extraction**: VGGNet16 **Answer Generation**: Ensemble of Classification models **Analysis**: Accuracy-60.8, BLEU Score-63.4	All models built for each question categories are classification models which is completely a black-box approach

Literature Survey Contd.

Paper Title	**Methodology**
Explainable Artificial Intelligence for Human Decision Support System in the Medical Domain [**XAI-CNN**]	**Dataset:** Red Lesion Endoscopy data **XAI tools:** LIME, SHAP, CIU A CNN is trained using the dataset. XAI tools are then used for visualization in terms of heatmap. The result of visualization is then compared
Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization [**Grad-CAM**]	Proposed Gradient-weighted Class Activation Mapping (Grad-CAM) for explaining and understanding CNN based models. This technique helps to visualize the important regions on the input image, that corresponds to the predicted outcome. This technique helps to understand many CNN-based models such as Image-Captioning and VQA models.
LISA : Enhance the explainability of medical images unifying current XAI techniques [**XAI-LISA**]	**Dataset:** COVID-19 Dataset **XAI tools:** LIME, SHAP, Anchors **Other XAI techniques:** Integrated Gradients Transfer Learning is used for the detection of COVID-19. The XAI tools LIME, SHAP Anchor and Integrated Gradient techniques' results are combined to give explanations.