Semantic based medical visual question answering with explainable artificial intelligence

Sheerin Sitara Noor Mohamed, Kavitha Srinivasan, Raghuraman Gopalsamy

Department of Computer Science and Engineering, Sri Sivasubramaniya Nadar College of Engineering, Chennai, India

Article Info

ABSTRACT

The medical visual question answering (MVQA) system takes the advantage of both computer vision (CV) and natural language processing (NLP) to accept the medical image and corresponding question as input and generates the respective answer as output. One step further, the MVQA system capable of generating the answer based on the semantics has a distinct place and hence semantic based medical visual question answering (SMVQA) system is proposed in this research. In SMVQA, the semantics for input image and question are generated using layerwise relevance propagation explainable artificial intelligence (LRP XAI) technique and the answer is derived using deductive reasoning method. For this, seven MVQA datasets are used for model creation, testing and validation. The training phase of the SMVQA system is implemented using VGGNet, long short-term memory (LSTM), LRP XAI, ResNet and bidirectional encoder representations from transformers (BERT) to generate a model file. Then the inference is derived in the testing phase based on the generated model file for the test set. Finally, the answer is derived from the inference using natural language toolkit (NLTK) library, term frequency-inverse document frequency (TF-IDF), cosine similarity, best match25 (BM25) techniques along with deductive reasoning. As a result, the proposed SMVQA system gives improved performance then the existing MVQA system especially for abnormality type samples.

Corresponding Author:

Sheerin Sitara Noor Mohamed
Department of Computer Science and Engineering, Sri Sivasubramaniya Nadar College of Engineering
Chennai, Tamil Nadu, India
Email: sheerinsitaran@ssn.edu.in

1. INTRODUCTION

The recent development in technological advancements over the past few years has been aimed towards simplifying tasks at every level. The medical field is no exception; through the integration of automation and secondary tools, healthcare professionals can allocate their time more efficiently to other productive tasks, while patients gain a clear understanding of their health conditions [1]. This progress has been facilitated by the evolution of artificial intelligence (AI) in healthcare named medical visual question answering (MVQA).

The MVQA aims to bridge the gap between the visual content of medical images and the linguistic context of questions asked by healthcare professionals or users. To attain this, the MVQA system takes the medical image and question with respect to different organ, plane, modality or abnormality as input and generates respective answer as output. The generalized MVQA system able to answer all types of samples (especially abnormality type because it is difficult as compared to other types) has huge scope and hence proposed work concentrates on developing semantic based medical visual question answering (SMVQA)

system. Because it understands the interworking of the model and analyses the reason behind the predicted answer. To implement the SMVQA system, a combination of two techniques namely layerwise relevance propagation explainable artificial intelligence (LRP XAI) [2] and deductive reasoning [3] are chosen. LRP XAI technique highlights the significant region in the input that contributes to answer prediction and, deductive reasoning shows the transformation of the input by solving sub-statements in the inference logically towards deriving the answer. As deductive reasoning with respect to MVQA is a new idea to apply in this domain, the subsequent paragraphs discusses only the collaborative efforts of different researchers in developing MVQA models along with XAI techniques.

The model interpretability achieved through XAI techniques helps in visualizing the significant regions that contribute towards answer prediction. Various XAI techniques, as outlined by Bennetot *et al*. [4], include Shapley additive explanations (SHAP), diverse counterfactual explanation (DiCE), transformer interpret (TI), logic tensor networks, and template system for natural language explanation (TS4NLE), gradient-weighted class activation mapping (Grad-CAM), and LRP XAI. Consequently, Joshi *et al*. [5] used Grad-CAM to identify the regions that contribute to generating answers for the visual question answering medical (VQA-MED) 2020 dataset. Similarly, the authors in [6], [7] utilized Grad-CAM for the VQA-MED 2019, VQA-RAD, and Path-VQA datasets due to its ability to extract rich textual information and demonstrate its visual reasoning capabilities. The XAI for medical VQA was developed by Canepa *et al*. [8] for VQA-MED 2019 dataset but it fails for two closely related disorders. This underscores the significance of XAI-based visualizations in identifying reasons behind incorrect predictions. Augmenting this, leveraging external knowledge base (EKB) can mitigate this challenge. Thus, Huang *et al*. [9] developed the medical knowledge-based VQA network (MKBN) for answering questions based on images in the Patient-oriented VQA dataset. Concurrently, Mohamed and Srinivasan in 2023 [10] devised an EKB derived from medical and linguistic terms from ImageCLEF and linguistic websites to infer answers based on semantic rules for the natural language inference for clinical trial (NLI4CT) dataset.

From the literature review, it's evident that MVQA model generation coupled with Grad-CAM XAI visualization primarily highlights the significant image information but LRP XAI highlights the significant image and text information together. Hence, in the proposed work, LRP XAI is preferred. Along with this, deep learning techniques and deductive reasoning are used because the LRP XAI highlights the significant regions in the input, deep learning techniques generates the inference based on the significant region and deductive reasoning derives the answer by retrieving the sub-statement of the inference that matches the rules. Through this, the proposed SMVQA system improves the performance of the model generated from samples of the abnormality category in the datasets. As the abnormality region is small or corresponds to multiple regions, highlighting the significant region and deriving the answer through generated inferences improves the overall performance.

The rest of the paper is organized as follows: Section 2 gives a brief description about existing MVQA datasets, design of the proposed SMVQA system based on inference obtained from the literature survey. Section 3 explains the experimental setup required, describes the implementation as a sequence of processes with sample input, discusses the results of correctly and wrongly classified samples and validates the results of the proposed SMVQA system using quantitative metrics. Finally, conclusion and future work are summarized in section 4.

2. METHOD

The schematic representation of the proposed SMVQA system for MVQA datasets is shown in Figure 1. In the proposed work, seven MVQA datasets are used to develop the SMVQA system using deep learning techniques and deductive reasoning method. Even though all datasets have abnormality type samples, VQA-MED 2020 and 2021 datasets completely belongs to abnormality type samples. These MVQA datasets are partitioned into training, validation and test sets as mentioned in Table 1. The training and validation set are combined to develop SMVQA model in the training phase, test set is used to generate inference using generated SMVQA model in the testing phase and then answer is derived from the inference using deductive reasoning method in the validation phase.

The proposed SMVQA system is implemented as three phases: Training, testing and validation. In the training phase, sequence of steps is executed using deep learning techniques namely, VGGNet [11], long short-term memory (LSTM) [12], LRP XAI, ResNet [13], and bidirectional encoder representations from transformers (BERT) [14], as shown in Figure 1. Initially. the image and text features are extracted from the training set using VGGNet and LSTM. For the extracted features from image and text, LRP XAI technique is applied to highlight the significant region in the image and text, named as super imposed image (SII) and super imposed QA-Pairs (SIQAP) respectively. Then the relevant features are extracted from the super imposed image and QA pairs using ResNet and BERT techniques and the features are concatenated as a

single vector for each sample. Finally, concatenated image and text features are used to generate a SMVQA model using LSTM. This model file is used for testing phase using text dataset.
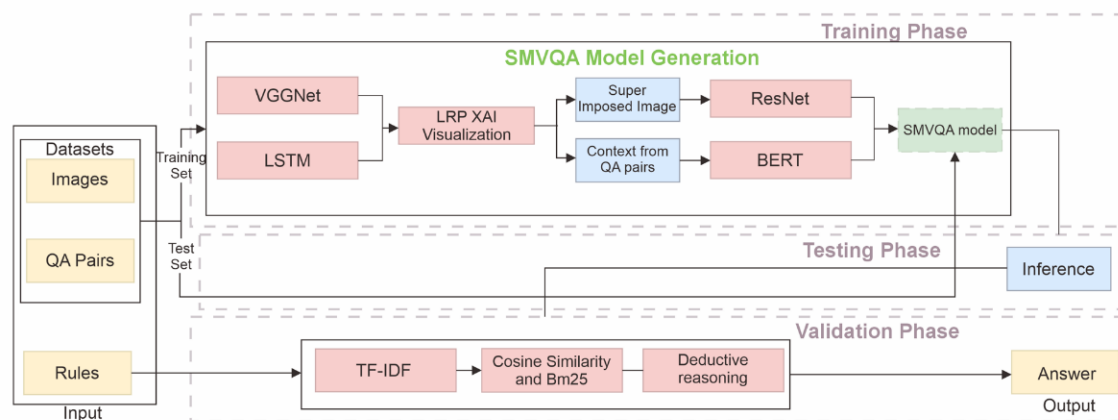

Figure 1. SMVQA system using deep learning techniques and deductive reasoning

Table 1. Datasets used

Datasets	Training set		Validation set		Test set	
	Images	QA-Pairs	Images	QA-Pairs	Images	QA-Pairs
VQA-MED 2018 [15]	2,278	5,413	324	500	264	500
VQA-MED 2019 [16]	3,200	12,792	500	2,000	500	500
VQA-MED 2020 [17]	4,000	4,000	500	500	500	500
VQA-MED 2021 [18]	4,500	4,500	500	500	500	500
VQA-MED 2023 [19]	2,000	36,000	-	-	1,949	35,082
VQA-RAD [20]	275	3,115	-	-	40	400
Path-VQA [21]	3,998	26,239	-	-	1,000	6,560

The developed SMVQA model generates "inferences" for the test set as shown in Figure 1. The "inferences" consist of natural language statements that represent the input image with respect to the question. The number of words in the statements generated for each sample depends on the type of image and question. i.e. variable size. In our observation, for abnormality category, the number of words is between 4 to 21 and for organ, plane and modality categories, the number of words is between 1 to 5 words. This generated inferences is used as an input in the validation phase.

The validation phase of SMVQA system is carried out using the inferences generated for test set along with deductive reasoning method. The suitable library and techniques used in this phase are: NLTK [22], term frequency inverse document frequency (TF-IDF) [23], cosine similarity [24], BM25 [25] techniques and deductive reasoning method. The sequence of steps for deriving one final answer is as follows: i) the inferences obtained from the test set with vocabulary list are given as input to NLTK library for preprocessing and to generate N-equivalent inference (NEI) statements. The preprocessing methods used are tagging, stemming, lemmatization, and tokenization. ii) from the generated NEI statements and defined rules, the suitable M-inference statements that matches the maximum number of sub-statements are identified using TF-IDF technique. iii) for the M-inference statements, the cosine similarity and BM25 techniques are applied for ranking the statements and the top ranked statement is selected. iv) then the answer is derived using defined rules through deductive reasoning from the top ranked statement. The deductive reasoning rules [26] proposed and applied in this research are: modus ponens, conjunction, and hypothetical syllogism. Finally, the results obtained are validated using appropriate quantitative metrics namely accuracy and bilingual evaluation understudy (BLEU) score.

3. RESULTS AND DISCUSSION

The hardware and software used for implementing the SMVQA system includes: i) an Intel i5 processor with NVIDIA GeForce Ti 4800, operating at 4.3 GHz clock speed, 16 GB RAM, graphical processing unit, and 2 TB disk space, and ii) Linux - Ubuntu 20.04 operating system, Python 3.7 package with necessary libraries such as TensorFlow, Torch, Scikit-learn, NLTK, Pickle, and Pandas. In this section,

the implementation steps of the proposed SMVQA system is explained in Figure 2 as a process flow diagram and the output generated at intermediate stages are shown in Figure 3, for the test samples. Also, the performance of the proposed SMVQA system is analysed using quantitative metrics and the results are compared with existing work and MVQA, tabulated in Tables 2 and 3 respectively.
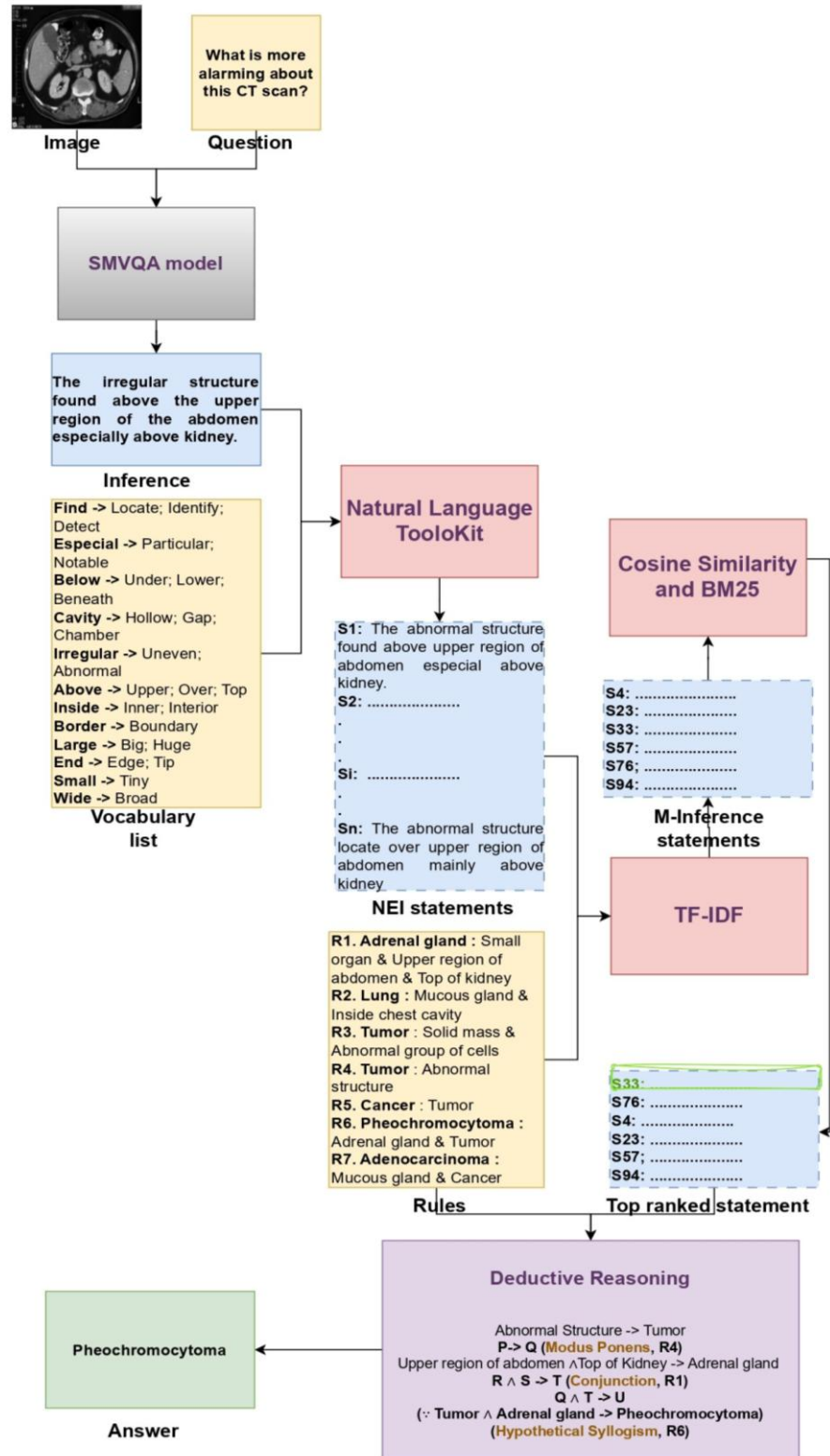

Figure 2. Process flow of deriving answer for a test sample using SMVQA system

The sequence of steps involved in deriving the answer for a test sample is shown in Figure 2 for clear understanding. The sample test image is a CT image with the question as "What is more alarming about this CT scan?". For this sample, the SMVQA training model generates one inference as "The irregular structure found above the upper region of the abdomen especially above kidney". With this inference statement, vocabulary list is combined for NLTK preprocessing and it generates 94 NEI statements. Based on the content of NEI statements, 7 rules are selected from the defined 32 rules. The 32 rules used in the proposed SMVQA system is listed as follows:

R1. Adrenal gland: Small organ & Upper region of abdomen & Top of kidney
R2. Lung: Mucous gland & Inside chest cavity
R3. Tumor: Solid mass & Abnormal group of cells
R4. Tumor: Abnormal structure
R5. Cancer: Tumor
R6. Osteo: Bone
R7. Pulmonary: Lung
R8. Cyst: Irregular & Projection
R9. Lump: Red & Swallon
R10. Embolism: Blood vessel & Blockage
R11. Kidney: Below rib cage & Behind spine
R12. Asymmetric cartilage lesion: Unequal lesion distribution & Size and location variation
R13. Granulomatous colitis: Inflammation & Mucosa
R14. Villous adenoma: Polyp & Colon
R15. Bilateral cleft palate: Cleft lip & Top of mouth
R16. Esophagus: Tube connect mouth throat and mouth
R17. Stroma: Connective tissue
R18. Lymphocytic leukemia: Bone marrow & Cancer
R19. Gastrointestinal stromal tumor: Cancer & Stroma
R20. Esophageal varices: Enlarged vein & Esophagus
R21. Enchondromatosis: Multiple enchondromas & Asymetric cartilage lesion
R22. Ovarian torsion: Fallopian tube & Tissue death & Twist in the tissue
R23. Ovarian torsion: Ovary & Twist in the tissue
R24. Azybos lobe: Right lung & Upper region & Slight deformation
R25. Appenditis: Lower Abdomen & Extra tissue
R26. Horseshoe Kidney: Kidney & Fusion & Lower end
R27. Osteosarcoma: Bone & Cancer
R28. Chondrocalcinosis: Knee joint & Calcium deposit
R29. Pulmonary Embolism: Lung & Embolism
R30. Sarcoidosis: Lung & Lump
R31. Pheochromocytoma: Adrenal gland & Tumor
R32. Adenocarcinoma: Mucous gland & Cancer

Then the TF-IDF technique is applied to select M-inference statements from the 94 NEI statements using the selected rules. For this test sample, six M-inference statements (S4, S23, S33, S57, S76, and S94) are selected based on 7 rules. The count of NEI statements, M-inference statements and rules are not unique across test samples. The selected M-inference statements are ranked as S33, S76, S4, S23, S57, and S94 based on the similarity score calculated by cosine similarity. From the sorted M-inference statements, S33 is selected by BM25 because it is top-ranked as compared to other M-inference statements. Finally, deductive reasoning method are applied to top ranked statement with respect to the rules to generate the sub-answers "Tumor" and "Adrenal glands". From these sub-answers, final answer is derived as "Pheochromocytoma".

The process of deriving answers for two different test samples from the VQA-MED 2021 dataset, along with three intermediate stages, is illustrated in Figure 3 as two subfigures. These subfigures depict the stage-wise results that help explain the reasoning behind the derived answers. The three intermediate stages such as SII and SIQAP, inference, and top-ranked statement play a crucial role in answer prediction. Because, i) the SII and SIQAP highlight the most significant regions of the image and the question that contributes to answer prediction, using blue and pink colors, respectively; ii) the inference is generated based solely on features extracted from these significant regions; and iii) the top-ranked statement is selected from the generated inferences based on the highest similarity score according to defined rules. This selection directly leads to deriving the final answer. For example, in Figure 3(a), the SII and SIQAP correctly identify an abnormality in the adrenal gland region, leading to the correct inference and top-ranked statement, as a result the answer is derived as "Pheochromocytoma," for the CT scan image. However, in Figure 3(b), while the radiology image presents an abnormality in the knee bone ('Osteochondroma'), the SII and SIQAP

incorrectly identifies it as inflammation in the periosteum (a tissue in the thigh bone), resulting in an incorrect derived answer.
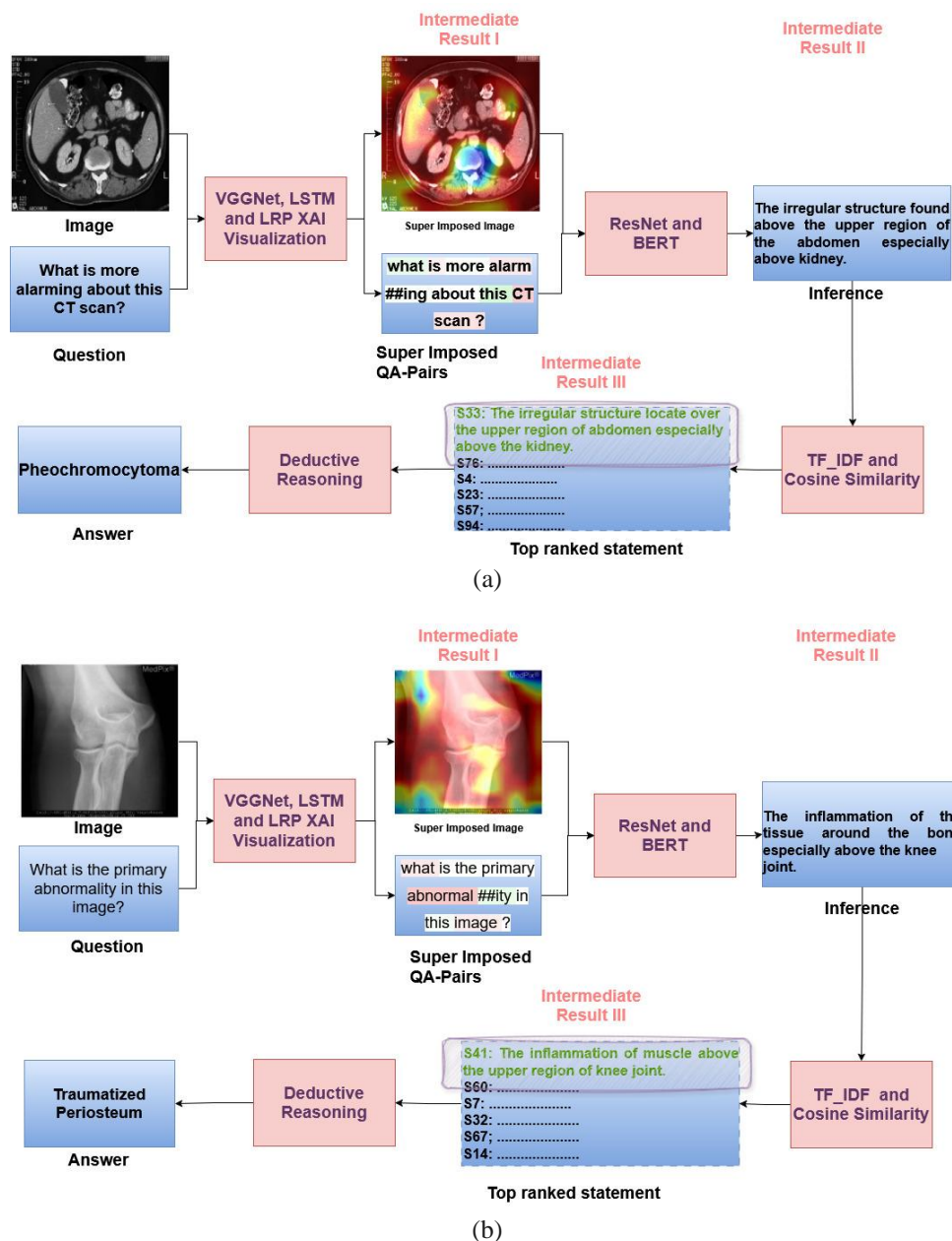

Figure 3. Intermediate results for correctly and wrongly classified samples from test set
(a) Correctly classified sample and (b) Wrongly classified sample

The performance of the proposed work is compared with existing work in terms of quantitative metrics like accuracy and BLEU score for seven MVQA datasets are given in Table 2. The SMVQA system achieves an accuracy and BLEU score ranging from 40.1% to 62.2% and 36.9% to 65.8%, respectively. Notably, the accuracy for VQA-MED 2019, 2023, and Path-VQA datasets exceeds 60.0%, because of prominent images, a sufficient number of answers, and a reasonable number of samples per answer category. Furthermore, the SMVQA system outperforms existing systems for VQA-MED 2019, 2020, 2021, and Path-VQA datasets by the range of 1.4%, 7.0%, 7.6%, and 4.5%, respectively. This improvement is due to the approach of deriving answers through deductive reasoning method based on rules/conditions, rather than predicting/generating answers. However, the proposed work lacks for VQA-MED 2023 and VQA-RAD

datasets because one-third of the samples requires colour or numeric oriented answers. Comparing all seven datasets, the performance of the model generated from abnormality type datasets namely, VQA-MED 2020 and VQA-MED 2021 dataset are improved by 7.0% and 7.6% respectively as compared with existing work. Because the SMVQA system mainly focuses on improving the performance of abnormality type samples using LRP XAI and deductive reasoning.

From Table 3, it has been inferred that SMVQA system improves the overall performance for all MVQA datasets as compared with MVQA system. Most importantly, the proposed SMVQA system gives improved accuracy of 11.9% and 15.5% for VQA-MED 2020 and 2021 datasets which has more abnormality related samples. From this, it has been evidenced that the proposed SMVQA system performs better than MVQA system and, SMVQA system is more suitable for abnormality type dataset or dataset containing more number of abnormality type samples. In addition, for VQA-MED 2018 dataset, the performance is increased by 12.0% using SMVQA because deriving an answer from the inference overcomes the disadvantage of wide combination of distinct samples.

Table 2. Existing work Vs proposed work using quantitative metrics

Datasets	Author and Year	Techniques	Existing work		Proposed work (SMVQA)	
			Accuracy	BLEU score	Accuracy	BLEU score
VQA-MED 2018	Peng et al. 2018 [27]	ResNet 152 and GRU	-	0.188	0.401	0.369
VQA-MED 2019	Al-Sadi et al., 2021 [28]	VGGNet and data augmentation	0.608	0.634	0.622	0.658
VQA-MED 2020	Joshi et al., 2023 [5]	Multi-modal multi-head self-attention based MedVQA	0.361	0.409	0.431	0.411
VQA-MED 2021	Gong et al., 2021 [29]	Mixup and ensemble of 8 pre-trained model	0.382	0.416	0.458	0.498
VQA-MED 2023	Wang et al., 2023 [30]	BLIP – 2(ViT-G and LLM)	0.739	-	0.620	0.614
VQA-RAD	Wang et al., 2023 [30]	Multi-task pre-trained model and LSTM	0.741	-	0.511	0.631
Path-VQA	Naseem et al., 2023 [21]	Stacked attention network	0.574	0.621	0.619	0.640

Table 3. Comparison and analysis of MVQA Vs SMVQA using quantitative metrics

Datasets	MVQA System		SMVQA System	
	Accuracy	BLEU score	Accuracy	BLEU score
VQA-MED 2018	0.281	0.341	0.401	0.369
VQA-MED 2019	0.579	0.616	0.622	0.658
VQA-MED 2020	0.312	0.343	0.431	0.411
VQA-MED 2021	0.303	0.304	0.458	0.498
VQA-MED 2023	0.586	0.594	0.620	0.614
VQA-RAD	0.498	0.603	0.511	0.631
Path-VQA	0.569	0.611	0.619	0.640

4. CONCLUSION

The proposed SMVQA system is developed to answer questions based on the semantics of the input using LRP XAI and deductive reasoning. The performance of the SMVQA system in terms of BLEU score is increased in the range of 0.2% to 18.1% when compared with existing works. Also, accuracy of the SMVQA system is increased in the range of 1.3% to 15.5% than the MVQA system for all datasets. Because the SMVQA system reduces the misclassification error on abnormality type samples by considering more than one information to answer the questions like affected organ, shape, color, size and position. In the future work, the performance of the SMVQA system can be further enhanced by: i) exploring different XAI techniques for medical datasets, ii) expanding the size of EKB by updating the vocabulary list and rules, and iii) incorporating diverse reasoning techniques.

FUNDING INFORMATION

This research received no specific grant from any funding agency or commercial sectors.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Sheerin Sitara Noor Mohamed	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	✓
Kavitha Srinivasan		✓	✓	✓		✓	✓	✓	✓		✓	✓		
Raghuraman Gopalsamy	✓				✓		✓			✓	✓	✓	✓	✓

C : **C**onceptualization	I : **I**nvestigation	Vi : **Vi**sualization
M : **M**ethodology	R : **R**esources	Su : **Su**pervision
So : **So**ftware	D : **D**ata Curation	P : **P**roject administration
Va : **Va**lidation	O : Writing - **O**riginal Draft	Fu : **Fu**nding acquisition
Fo : **Fo**rmal analysis	E : Writing - Review & **E**diting	

CONFLICT OF INTEREST STATEMENT

The authors declare that there is no conflict of interest.

INFORMED CONSENT

This work did not involve human participants.

DATA AVAILABILITY

The medical datasets used are open source and collected from ImageCLEF.

REFERENCES

[1] P. Szolovits, R. S. Patil, and W. B. Schwartz, "Artificial intelligence in medical diagnosis," *Annals of Internal Medicine*, vol. 108, no. 1, pp. 80–87, 1988, doi: 10.7326/0003-4819-108-1-80.
[2] L. Weber, S. Lapuschkin, A. Binder, and W. Samek, "Beyond explaining: opportunities and challenges of XAI-based model improvement," *Information Fusion*, vol. 92, pp. 154–176, 2023, doi: 10.1016/j.inffus.2022.11.013.
[3] P. N. Johnson-Laird, "Deductive reasoning," *Annual Review of Psychology*, vol. 50, pp. 109–135, 1999, doi: 10.1146/annurev.psych.50.1.109.
[4] A. Bennetot *et al.*, "A practical guide on explainable AI techniques applied on biomedical use case applications," *SSRN Electronic Journal*, 2022, doi: 10.2139/ssrn.4229624.
[5] V. Joshi, P. Mitra, and S. Bose, "Multi-modal multi-head self-attention for medical VQA," *Multimedia Tools and Applications*, vol. 83, no. 14, pp. 42585–42608, 2024, doi: 10.1007/s11042-023-17162-3.
[6] P. Li, G. Liu, J. He, Z. Zhao, and S. Zhong, "Masked vision and language pre-training with unimodal and multimodal contrastive losses for medical visual question answering," *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*, pp. 374–383, 2023, doi: 10.1007/978-3-031-43907-0_36.
[7] X. Huang and H. Gong, "A dual-attention learning network with word and sentence embedding for medical visual question answering," *IEEE Transactions on Medical Imaging*, vol. 43, no. 2, pp. 832–845, 2024, doi: 10.1109/TMI.2023.3322868.
[8] L. Canepa, S. Singh, and A. Sowmya, "Visual question answering in the medical domain," *2023 International Conference on Digital Image Computing: Techniques and Applications, DICTA 2023*, pp. 379–386, 2023, doi: 10.1109/DICTA60407.2023.00059.
[9] J. Huang *et al.*, "Medical knowledge-based network for patient-oriented visual question answering," *Information Processing and Management*, vol. 60, no. 2, 2023, doi: 10.1016/j.ipm.2022.103241.
[10] S. S. N. Mohamed and K. Srinivasan, "SSNSheerinKavitha at SemEval-2023 task 7: semantic rule based label prediction using TF-IDF and BM25 techniques," *17th International Workshop on Semantic Evaluation, SemEval 2023-Proceedings of the Workshop*, pp. 950–957, 2023, doi: 10.18653/v1/2023.semeval-1.131.
[11] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *3rd International Conference on Learning Representations, ICLR 2015-Conference Track Proceedings*, 2015.
[12] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.
[13] B. Koonce, "ResNet 50," *Convolutional Neural Networks with Swift for Tensorflow*, pp. 63–72, 2021, doi: 10.1007/978-1-4842-6168-2_6.
[14] M. V. Koroteev, "BERT: a review of applications in natural language processing and understanding," *arXiv-Computer Science*, pp. 1-18, 2021.
[15] S. A. Hasan, Y. Ling, O. Farri, J. Liu, H. Müller, and M. Lungren, "Overview of ImageCLEF 2018 medical domain visual question answering task," in *CEUR Workshop Proceedings*, vol. 2125, pp. 1–7, 2018.
[16] A. B. Abacha, S. A. Hasan, V. V. Datla, J. Liu, D. Demner-Fushman, and H. Müller, "VQA-MED: overview of the medical visual question answering task at ImageClef 2019," in *CEUR Workshop Proceedings*, vol. 2380, pp. 1–11, 2019.

[17] A. B. Abacha, V. V. Datla, S. A. Hasan, D. Demner-Fushman, and H. Müller, "Overview of the VQA-Med task at ImageCLEF 2020: visual question answering and generation in the medical domain," in *CEUR Workshop Proceedings*, vol. 2696, pp. 1–9, 2020.

[18] A. B. Abacha, V. V. Datla, S. A. Hasan, D. Demner-Fushman, and H. Müller, "Overview of the VQA-Med task at ImageCLEF 2021: visual question answering and generation in the medical domain," in *CEUR Workshop Proceedings*, vol. 2696, pp. 1–8, 2021.

[19] S. Hicks, A. Storås, P. Halvorsen, M. Riegler, and V. Thambawita, "Overview of ImageCLEFmedical 2024-medical visual question answering for gastrointestinal tract," in *CEUR Workshop Proceedings*, vol. 3740, pp. 1429–1436, 2024.

[20] J. J. Lau, S. Gayen, A. Ben Abacha, and D. Demner-Fushman, "A dataset of clinically generated visual questions and answers about radiology images," *Scientific Data*, vol. 5, 2018, doi: 10.1038/sdata.2018.251.

[21] U. Naseem, M. Khushi, A. G. Dunn, and J. Kim, "K-PathVQA: knowledge-aware multimodal representation for pathology visual question answering," *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 4, pp. 1886–1895, 2024, doi: 10.1109/JBHI.2023.3294249.

[22] P. Maddigan and T. Susnjak, "Chat2VIS: generating data visualizations via natural language using ChatGPT, codex and GPT-3 large language models," *IEEE Access*, vol. 11, pp. 45181–45193, 2023, doi: 10.1109/ACCESS.2023.3274199.

[23] S. V. Balshetwar, A. Rs, and D. J. R, "Correction to: fake news detection in social media based on sentiment analysis using classifier techniques," *Multimedia Tools and Applications*, vol. 82, no. 23, 2023, doi: 10.1007/s11042-023-15801-3.

[24] H. Steck, C. Ekanadham, and N. Kallus, "Is cosine-similarity of embeddings really about similarity?," *WWW 2024 Companion-Companion Proceedings of the ACM Web Conference*, pp. 887–890, 2024, doi: 10.1145/3589335.3651526.

[25] A. Askari, A. Abolghasemi, G. Pasi, W. Kraaij, and S. Verberne, "Injecting the BM25 score as text improves BERT-based re-rankers," *Advances in Information Retrieval*, pp. 66–83, 2023, doi: 10.1007/978-3-031-28244-7_5.

[26] U. Goswami, "Inductive and deductive reasoning," *The Wiley-Blackwell Handbook of Childhood Cognitive Development, Second edition*, pp. 399–419, 2010, doi: 10.1002/9781444325485.ch15.

[27] Y. Peng, F. Liu, and M. P. Rosen, "UMass at ImageCLEF medical visual question answering (Med-VQA) 2018 task," in *CEUR Workshop Proceedings*, vol. 2125, pp. 1–7. 2018.

[28] A. Al-Sadi, M. Al-Ayyoub, Y. Jararweh, and F. Costen, "Visual question answering in the medical domain based on deep learning approaches: A comprehensive study," *Pattern Recognition Letters*, vol. 150, pp. 57–75, 2021, doi: 10.1016/j.patrec.2021.07.002.

[29] H. Gong, R. Huang, G. Chen, and G. Li, "SYSU-HCP at VQA-Med 2021: a data-centric model with efficient training methodology for medical visual question answering," in *CEUR Workshop Proceedings*, vol. 2936, pp. 1218–1228, 2021.

[30] S. Wang *et al.*, "Adapting pre-trained visual and language models for medical image question answering," in *CEUR Workshop Proceedings*, vol. 3497, pp. 1744–1753, 2023.

BIOGRAPHIES OF AUTHORS

Ms. Sheerin Sitara Noor Mohamed 🆔 🔗 SC ◖ is currently pursuing her Ph.D. in the Department of Computer Science and Engineering at Sri Sivasubramaniya Nadar College of Engineering (SSNCE) in Chennai. As a research assistant, she began her research journey at SSNCE in India in 2019. From 2020 to 2023, she was a Junior Research Fellow (JRF) at the same institution and has published 15+ papers in reputable international journals. She holds membership in IEEE. Her research focuses on medical image processing, natural language processing, machine learning, and deep learning. She can be contacted at email: sheerinsitaran@ssn.edu.in.

Dr. Kavitha Srinivasan 🆔 🔗 SC ◖ is an associate professor in the Department of Computer Science and Engineering at Sri Sivasubramaniya Nadar College of Engineering, Chennai. She has 25 years of teaching experience including 13 years of research experience in different fields namely bioinformatics, data mining, machine learning, and soft computing. She holds membership in IEEE, ACM and the Machine Learning Research Group of SSN, along with lifetime membership in the Computer Society of India (CSI) and the Indian Society for Technical Education (ISTE). She has published more than 70 publications in referred journals and conferences. She can be contacted at email: kavithas@ssn.edu.in.

Dr. Raghuraman Gopalsamy 🆔 🔗 SC ◖ is an associate professor in the Department of Computer Science and Engineering at Sri Sivasubramaniya Nadar College of Engineering, Chennai. He has 15 years of teaching experience including 4.5 years of research experience in image processing, multi agent systems, and cloud computing. He has published more than 25 research publications in referred journals and conferences. He holds membership in the Computer Society of India (CSI), Indian Society for Technical Education (ISTE) and IEEE. He can be contacted at email: raghuramang@ssn.edu.in.