Published in Proceedings of Conference and Labs of the Evaluation Forum (CLEF) 2019 Working Notes, Lugano, Switzerland, 9-12 September 2019, which should be cited to refer to this work.

VQA-Med: Overview of the Medical Visual Question Answering Task at ImageCLEF 2019

Asma Ben Abacha[1], Sadid A. Hasan[2], Vivek V. Datla[2], Joey Liu[2], Dina Demner-Fushman[1], and Henning Müller[3]

[1] Lister Hill Center, National Library of Medicine, USA
[2] Philips Research Cambridge, USA
[3] University of Applied Sciences Western Switzerland, Sierre, Switzerland
asma.benabacha@nih.gov
sadid.hasan@philips.com

Abstract. This paper presents an overview of the Medical Visual Question Answering task (VQA-Med) at ImageCLEF 2019. Participating systems were tasked with answering medical questions based on the visual content of radiology images. In this second edition of VQA-Med, we focused on four categories of clinical questions: Modality, Plane, Organ System, and Abnormality. These categories are designed with different degrees of difficulty leveraging both classification and text generation approaches. We also ensured that all questions can be answered from the image content without requiring additional medical knowledge or domain-specific inference. We created a new dataset of 4,200 radiology images and 15,292 question-answer pairs following these guidelines. The challenge was well received with 17 participating teams who applied a wide range of approaches such as transfer learning, multi-task learning, and ensemble methods. The best team achieved a BLEU score of 64.4% and an accuracy of 62.4%. In future editions, we will consider designing more goal-oriented datasets and tackling new aspects such as contextual information and domain-specific inference.

Keywords: Visual Question Answering, Data Creation, Deep Learning, Radiology Images, Medical Questions and Answers

1 Introduction

Recent advances in artificial intelligence opened new opportunities in clinical decision support. In particular, relevant solutions for the automatic interpretation of medical images are attracting a growing interest due to their potential applications in image retrieval and in assisted diagnosis. Moreover, systems capable of understanding clinical images and answering questions related to their content could support clinical education, clinical decision, and patient education. From a computational perspective, this Visual Question Answering (VQA) task presents an exciting problem that combines natural language processing and computer vision techniques. In recent years, substantial progress has been made on VQA with new open-domain datasets [3, 8] and approaches [23, 7].

However, there are challenges that need to be addressed when tackling VQA in a specialized domain such as Medicine. Ben Abacha *et al.* [4] analyzed some of the issues facing medical visual question answering and described four key challenges (i) designing goal-oriented VQA systems and datasets, (ii) categorizing the clinical questions, (iii) selecting (clinically) relevant images, and (iv) capturing the context and the medical knowledge.

Inspired by the success of visual question answering in the general domain, we conducted a pilot task (VQA-Med 2018) in ImageCLEF 2018 to focus on visual question answering in the medical domain [9]. Based on the success of the initial edition, we continued the task this year with enhanced focus on a well curated and larger dataset.

In VQA-Med 2019, we selected radiology images and medical questions that (i) asked about only one element and (ii) could be answered from the image content. We targeted four main categories of questions with different difficulty levels: Modality, Plane, Organ system, and Abnormality. For instance, the first three categories can be tackled as a classification task, while the fourth category (abnormality) presents an answer generation problem. We intentionally designed the data in this manner to study the behavior and performance of different approaches on both aspects. This design is more relevant to clinical decision support than the common approach in open-domain VQA datasets [3, 8] where the answers consist of one word or number (e.g. yes, no, 3, stop).

In the following section, we present the task description with more details and examples. We describe the data creation process and the VQA-Med-2019 dataset in section 3. We present the evaluation methodology and discuss the challenge results respectively in sections 4 and 5.

2 Task Description

In the same way as last year, given a medical image accompanied by a clinically relevant question, participating systems in VQA-Med 2019 are tasked with answering the question based on the visual image content. In VQA-Med 2019, we specifically focused on radiology images and four main categories of questions: Modality, Plane, Organ System, and Abnormality. We mainly considered medical questions asking only about one element: e.g., “what is the organ principally shown in this MRI?”, “in what plane is this mammograph taken?”, “is this a t1 weighted, t2 weighted, or flair image?”, “what is most alarming about this ultrasound?”).

All selected questions can be answered from the image content without requiring additional domain-specific inference or context. Other questions including these aspects will be considered in future editions of the challenge, e.g.: ”Is this modality safe for pregnant women?”, ”What is located immediately inferior to the right hemidiaphragm?", "What can be typically visualized in this plane?", "How would you measure the length of the kidneys?"

3 VQA-Med-2019 Dataset

We automatically constructed the training, validation, and test sets, by (i) applying several filters to select relevant images and associated annotations, and (ii) creating patterns to generate the questions and their answers. The test set was manually validated by two medical doctors. The dataset is publicly available[4]. Figure 1 presents examples from the VQA-Med-2019 dataset.

3.1 Medical Images

We selected relevant medical images from the MedPix[5] database with filters based on their captions, modalities, planes, localities, categories, and diagnosis methods. We selected only the cases where the diagnosis was made based on the image. Examples of the selected diagnosis methods: CT/MRI Imaging, Angiography, Characteristic imaging appearance, Radiographs, Imaging features, Ultrasound, Diagnostic Radiology.

3.2 Question Categories and Patterns

We targeted the most frequent question categories: Modality, Plane, Organ system and Abnormality (Ref:VQA-RAD).

1) Modality: Yes/No, WH and closed questions. Examples:

- was gi contrast given to the patient?
- what is the mr weighting in this image?
- what modality was used to take this image?
- is this a t1 weighted, t2 weighted, or flair image?

2) Plane: WH questions. Examples:

- what is the plane of this mri?
- in what plane is this mammograph taken?

3) Organ System: WH questions. Examples:

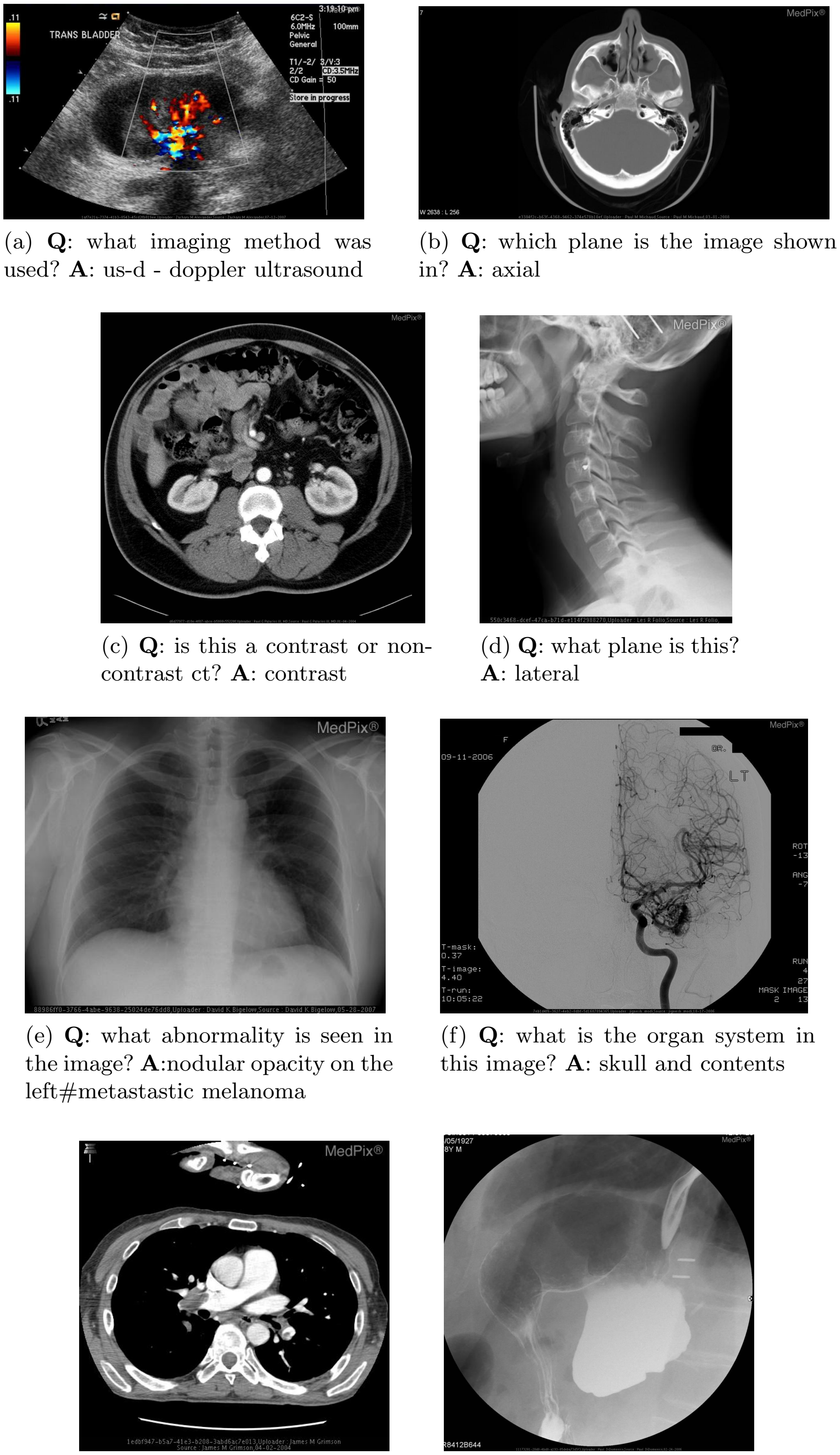
- what organ system is shown in this x-ray?
- what is the organ principally shown in this mri?

4) Abnormality: Yes/No and WH questions. Examples:

- does this image look normal?

[4] github.com/abachaa/VQA-Med-2019

[5] https://medpix.nlm.nih.gov



(a) **Q**: what imaging method was used? **A**: us-d - doppler ultrasound

(b) **Q**: which plane is the image shown in? **A**: axial

(c) **Q**: is this a contrast or non-contrast ct? **A**: contrast

(d) **Q**: what plane is this? **A**: lateral

(e) **Q**: what abnormality is seen in the image? **A**:nodular opacity on the left#metastastic melanoma

(f) **Q**: what is the organ system in this image? **A**: skull and contents

(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)

Fig. 1: Examples from VQA-Med-2019 test set

- are there abnormalities in this gastrointestinal image?
- what is the primary abnormality in the image?
- what is most alarming about this ultrasound?

Planes (16): Axial; Sagittal; Coronal; AP; Lateral; Frontal; PA; Transverse; Oblique; Longitudinal; Decubitus; 3D Reconstruction; Mammo-MLO; Mammo-CC; Mammo-Mag CC; Mammo-XCC.
Organ Systems (10): Breast; Skull and Contents; Face, sinuses, and neck; Spine and contents; Musculoskeletal; Heart and great vessels; Lung, mediastinum, pleura; Gastrointestinal; Genitourinary; Vascular and lymphatic.
Modalities (36):

- [**XR**]: XR-Plain Film
- [**CT**]: CT-noncontrast; CT w/contrast (IV); CT-GI & IV Contrast; CTA-CT Angiography; CT-GI Contrast; CT-Myelogram; Tomography
- [**MR**]: MR-T1W w/Gadolinium; MR-T1W-noncontrast; MR-T2 weighted; MR-FLAIR; MR-T1W w/Gd (fat suppressed); MR T2* gradient,GRE,MPGR, SWAN,SWI; MR-DWI Diffusion Weighted; MRA-MR Angiography/Venography; MR-Other Pulse Seq.; MR-ADC Map (App Diff Coeff); MR-PDW Proton Density; MR-STIR; MR-FIESTA; MR-FLAIR w/Gd; MR-T1W SPGR; MR-T2 FLAIR w/Contrast; MR T2* gradient GRE
- [**US**]: US-Ultrasound; US-D-Doppler Ultrasound
- [**MA**]: Mammograph
- [**GI**]: BAS-Barium Swallow; UGI-Upper GI; BE-Barium Enema; SBFT-Small Bowel
- [**AG**]: AN-Angiogram; Venogram
- [**PT**]: NM-Nuclear Medicine; PET-Positron Emission

Patterns: For each category, we selected question patterns from hundreds of questions naturally asked and validated by medical students from the VQA-RAD dataset [13].

3.3 Training and Validation Sets

The training set includes 3,200 images and 12,792 question-answer (QA) pairs, with 3 to 4 questions per image. Table 1 presents the most frequent answers per category. The validation set includes 500 medical images with 2,000 QA pairs.

3.4 Test Set

A medical doctor and a radiologist performed a manual double validation of the test answers. A total of 33 answers were updated by (i) indicating an optional part (8 answers), (ii) adding other possible answers (10), or (iii) correcting the automatic answer. 15 answers were corrected, which corresponds to 3% of the test answers. The corrected answers correspond to the following categories: Abnormality (8/125), Organ (6/125), and Plane (1/125). For abnormality questions, the correction was mainly changing the diagnosis that is inferred, by the problem

Category	**Most frequent answers (#)**
Modality	no (554), yes (552), xr-plain film (456), t2 (217), us-ultrasound (183), t1 (137), contrast (107), noncontrast (102), ct noncontrast (84), mr-flair (84), an-angiogram (78), mr-t2 weighted (56), flair (53), ct w/contrast (iv) (50), cta-ct angiograph (45)
Plane	axial (1558), sagittal (478), coronal (389), ap (197), lateral (151), frontal (120), pa (92), transverse (76), oblique (50)
Organ System	skull and contents (1216), musculoskeletal (436), gastrointestinal (352), lung, mediastinum, pleura (250), spine and contents (234), genitourinary (214), face, sinuses, and neck (191), vascular and lymphatic (122), heart and great vessels (120), breast (65)
Abnormality	yes (62), no (48), meningioma (30), glioblastoma multiforme (28), pulmonary embolism (16), acute appendicitis (14), arteriovenous malformation (avm) (14), arachnoid cyst (13), schwannoma (13), tuberous sclerosis (13), brain, cerebral abscess (12), ependymoma (12), fibrous dysplasia (12), multiple sclerosis (12), diverticulitis (11), langerhan cell histiocytosis (11), sarcoidosis (11)

Table 1: VQA-Med-2019 Training Set: the Most Frequent Answers Per Category

seen in the image. We expect a similar error rate in the training and validation sets that were generated using the same automatic data creation method. The test set consists of 500 medical images and 500 questions.

4 Evaluation Methodology

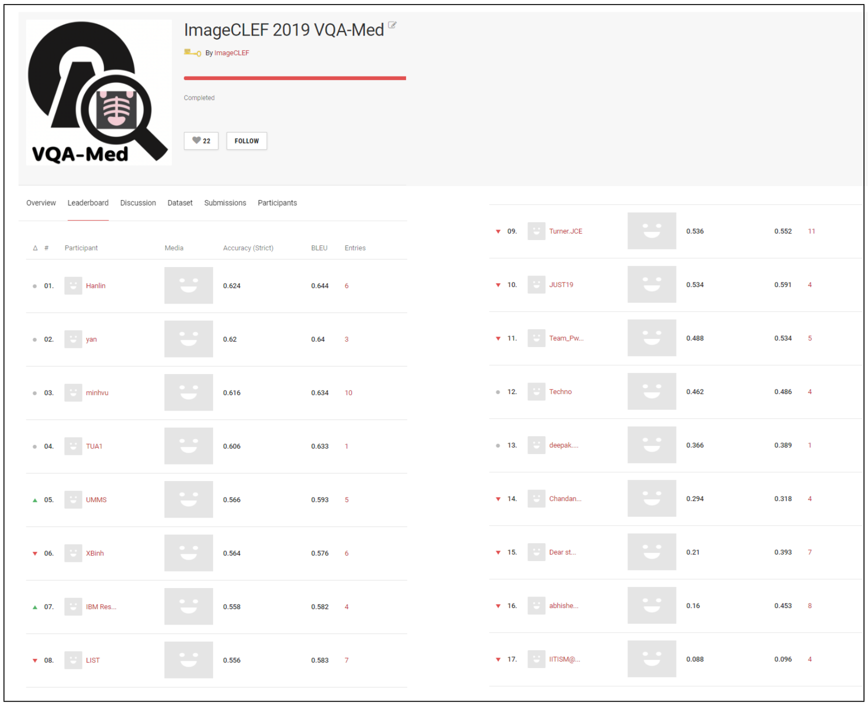
The evaluation of the systems that participated in the VQA-Med 2019 task was conducted based on two primary metrics: Accuracy and BLEU. We use an adapted version of the accuracy metric from the general domain VQA[6] task that strictly considers exact matching of a participant provided answer and the ground truth answer. We calculate the overall accuracy scores as well as the scores for each question category. To compensate for the strictness of the accuracy metric, BLEU [15] is used to capture the word overlap-based similarity between a system-generated answer and the ground truth answer. The overall methodology and resources for the BLEU metric are essentially similar to last year's task [9].

5 Results and Discussion

Out of 104 online registrations, 61 participants submitted signed end-user agreement forms. Finally, 17 groups submitted a total of 90 runs, indicating a notable interest in the VQA-Med 2019 task. Figure 2 presents the results of the 17 participating teams. The best overall result was obtained by the *Hanlin* team,

[6] https://visualqa.org/evaluation.html

achieving 0.624 Accuracy and 0.644 BLEU score. Table 2 gives an overview of all participants and the number of submitted runs[7]. The overall results of the participating systems are presented in Table 3 to Table 4 for the two metrics in a descending order of the scores (the higher the better). Detailed results of each run are described in the ImageCLEF 2019 lab overview paper [11].



Δ	#	Participant	Media	Accuracy (Strict)	BLEU	Entries
●	01.	Hanlin		0.624	0.644	6
●	02.	yan		0.62	0.64	3
●	03.	minhvu		0.616	0.634	10
●	04.	TUA1		0.606	0.633	1
▲	05.	UMMS		0.566	0.593	5
▼	06.	XBinh		0.564	0.576	6
▲	07.	IBM Res...		0.558	0.582	4
▼	08.	LIST		0.556	0.583	7
▼	09.	Turner.JCE		0.536	0.552	11
▼	10.	JUST19		0.534	0.591	4
▼	11.	Team_Pw...		0.488	0.534	5
●	12.	Techno		0.462	0.486	4
●	13.	deepak....		0.366	0.389	1
▼	14.	Chandan...		0.294	0.318	4
▼	15.	Dear st...		0.21	0.393	7
▼	16.	abhishe...		0.16	0.453	8
▼	17.	IITISM@...		0.088	0.096	4

Fig. 2: Results of VQA-Med 2019 on crowdAI

Table 3: VQA-Med 2019: Accuracy scores

Team	Run ID	Modality	Plane	Organ	Abnormality	Overall
Hanlin	26889	0.202	0.192	0.184	0.046	0.624
yan	26853	0.202	0.192	0.184	0.042	0.620
minhvu	26881	0.210	0.194	0.190	0.022	0.616
TUA1	26822	0.186	0.204	0.198	0.018	0.606
UMMS	27306	0.168	0.190	0.184	0.024	0.566
AIOZ	26873	0.182	0.180	0.182	0.020	0.564
IBM Research AI	27199	0.160	0.196	0.192	0.010	0.558
LIST	26908	0.180	0.184	0.178	0.014	0.556

[7] There was a limit of maximum 10 run submissions per team. The table includes only the valid runs that were graded (total# 80 out of 90 submissions)

Turner.JCE	26913	0.164	0.176	0.182	0.014	0.536
JUST19	27142	0.160	0.182	0.176	0.016	0.534
Team_Pwc_Med	26941	0.148	0.150	0.168	0.022	0.488
Techno	27079	0.082	0.184	0.170	0.026	0.462
deepak.gupta651	27232	0.096	0.140	0.124	0.006	0.366
ChandanReddy	26884	0.094	0.126	0.064	0.010	0.294
Dear stranger	26895	0.062	0.140	0	0.008	0.210
abhishekthanki	27307	0.122	0	0.028	0.010	0.160
IITISM@CLEF	26905	0.052	0.004	0.026	0.006	0.088

Table 4: VQA-Med 2019: BLEU scores

Team	**Run ID**	**BLEU**
Hanlin	26889	0.644
yan	26853	0.640
minhvu	26881	0.634
TUA1	26822	0.633
UMMS	27306	0.593
JUST19	27142	0.591
LIST	26908	0.583
IBM Research AI	27199	0.582
AIOZ	26833	0.579
Turner.JCE	26940	0.572
Team_Pwc_Med	26955	0.534
Techno	27079	0.486
abhishekthanki	26824	0.462
Dear stranger	26895	0.393
deepak.gupta651	27232	0.389
ChandanReddy	26946	0.323
IITISM@CLEF	26905	0.096

Similar to last year, participants mainly used deep learning techniques to build their VQA-Med systems. In particular, the best-performing systems leveraged deep convolutional neural networks (CNNs) like VGGNet [18] or ResNet [10] with a variety of pooling strategies e.g., global average pooling to encode image features and transformer-based architectures like BERT [6] or recurrent neural networks (RNN) to extract question features. Then, various types of attention mechanisms are used coupled with different pooling strategies such as multimodal factorized bilinear (MFB) pooling or multi-modal factorized high-order pooling (MFH) in order to combine multimodal features followed by bilinear transformations to finally predict the possible answers.

Analyses of the question category-wise[8] accuracy in Table 3 suggest that in general, participating systems performed well to answer modality questions, followed by plane and organ questions because the possible types of answers for each of these question categories were finite. However, for the abnormality type questions, systems did not perform well in terms of accuracy because of the underlying complexity of open-ended

[8] Note that the question category-wise accuracy scores are normalized (each divided by a factor of 4) so that the summation is equal to the overall accuracy.

Table 2: Participating groups in the VQA-Med 2019 task.

Team	**Institution**	**# Runs**
abhishekthanki [20]	Manipal Institute of Technology (India)	8
AIOZ	AIOZ Pte Ltd (Singapore)	6
ChandanReddy	Virginia Tech (USA)	4
Dear stranger [14]	School of Information Science and Engineering, Kunming (China)	6
deepak.gupta651	Indian Institute of Technology Patna (India)	1
Hanlin	Zhejiang University (China)	5
IBM Research AI [12]	IBM Research, Almaden (USA)	4
IITISM@CLEF	Indian Institute of Technology Dhanbad (India)	3
JUST19 [1]	(Jordan) University of Science and Technology & University of Manchester (UK)	4
LIST [2]	Faculty of Sciences and Technologies, Tangier (Morocco)	7
minhvu [21]	Umeå University (Sweden) & University of Bern (Switzerland)	10
Team_Pwc_Med [16]	Pricewaterhouse Coopers US Advisory (India)	5
Techno [5]	Faculty of Technology Tlemcen (Algeria)	2
TUA1 [24]	Tokushima University (Japan)	1
Turner.JCE [19]	Azrieli College of Engineering Jerusalem (Israel)	10
UMMS [17]	Worcester Polytechnic Institute & University of Massachusetts Medical School (USA)	3
yan [22]	Zhejiang University (China) & National Institute of Informatics (Japan)	1

questions and possibly due to the strictness of the accuracy metric. To compensate for the strictness of the accuracy, we computed the BLEU scores to understand the similarity of the system generated answers and the ground-truth answers. The higher BLEU scores of the systems this year (0.631 best BLEU vs. 0.162 in 2018) further verify the effectiveness of the proposed deep learning-based models for the VQA task. Overall, the results obtained this year clearly denote the robustness of the provided dataset compared to last year's task.

6 Conclusions

We presented the VQA-Med 2019 task, the new dataset, the participating systems, and official results. To ensure that the questions are naturally phrased, we used patterns from question asked by medical students to build clinically relevant questions belonging to our four target categories. We created a new dataset for the challenge[9] following goal-oriented guidelines, and covering questions with varying degrees of difficulty. A wide range of approaches have been applied such as transfer learning, multi-task learning, ensemble methods, and hybrid approaches combining classification models and answer generation methods. The best team achieved 0.644 BLEU score and 0.624 overall accuracy. In future editions we are considering more complex questions that might include contextual information or require domain-specific inference to reach the right answer.

Acknowledgments

This work was supported by the intramural research program at the U.S. National Library of Medicine, National Institutes of Health.
We thank Dr. James G. Smirniotopoulos and Soumya Gayen from the MedPix team for their support.

[9] www.crowdai.org/clef_tasks/13/task_dataset_files?challenge_id=53
github.com/abachaa/VQA-Med-2019

References

1. Al-Sadi, A., Talafha, B., Al-Ayyoub, M., Jararweh, Y., Costen, F.: Just at imageclef 2019 visual question answering in the medical domain. In: Working Notes of CLEF 2019 (2019)
2. Allaouzi, I., Benamrou, B., Ahmed, M.B.: An encoder-decoder model for visual question answering in the medical domain. In: Working Notes of CLEF 2019 (2019)
3. Antol, S., Agrawal, A., Lu, J., Mitchell, M., Batra, D., Zitnick, C.L., Parikh, D.: VQA: visual question answering. In: 2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7-13, 2015. pp. 2425–2433 (2015), https://doi.org/10.1109/ICCV.2015.279
4. Ben Abacha, A., Gayen, S., Lau, J.J., Rajaraman, S., Demner-Fushman, D.: NLM at imageclef 2018 visual question answering in the medical domain. In: Working Notes of CLEF 2018 - Conference and Labs of the Evaluation Forum, Avignon, France, September 10-14, 2018. (2018), http://ceur-ws.org/Vol-2125/paper_165.pdf
5. Bounaama, R., Abderrahim, M.A.: Tlemcen university at imageclef 2019 visual question answering task. In: Working Notes of CLEF 2019 (2019)
6. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of NAACL (2019)
7. Fukui, A., Park, D.H., Yang, D., Rohrbach, A., Darrell, T., Rohrbach, M.: Multimodal compact bilinear pooling for visual question answering and visual grounding. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016. pp. 457–468 (2016), http://aclweb.org/anthology/D/D16/D16-1044.pdf
8. Goyal, Y., Khot, T., Summers-Stay, D., Batra, D., Parikh, D.: Making the V in VQA matter: Elevating the role of image understanding in visual question answering. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017. pp. 6325–6334 (2017), https://doi.org/10.1109/CVPR.2017.670
9. Hasan, S.A., Ling, Y., Farri, O., Liu, J., Müller, H., Lungren, M.: Overview of imageclef 2018 medical domain visual question answering task. In: Working Notes of CLEF 2018 - Conference and Labs of the Evaluation Forum, Avignon, France, September 10-14, 2018. (2018), http://ceur-ws.org/Vol-2125/paper_212.pdf
10. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016. pp. 770–778 (2016), https://doi.org/10.1109/CVPR.2016.90
11. Ionescu, B., Müller, H., Péteri, R., Cid, Y.D., Liauchuk, V., Kovalev, V., Klimuk, D., Tarasau, A., Ben Abacha, A., Hasan, S.A., Datla, V., Liu, J., Demner-Fushman, D., Dang-Nguyen, D.T., Piras, L., Riegler, M., Tran, M.T., Lux, M., Gurrin, C., Pelka, O., Friedrich, C.M., de Herrera, A.G.S., Garcia, N., Kavallieratou, E., del Blanco, C.R., Rodríguez, C.C., Vasillopoulos, N., Karampidis, K., Chamberlain, J., Clark, A., Campello, A.: ImageCLEF 2019: Multimedia retrieval in medicine, lifelogging, security and nature. In: Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the 10th International Conference of the CLEF Association (CLEF 2019), LNCS Lecture Notes in Computer Science, Springer, Lugano, Switzerland (September 9-12 2019)
12. Kornuta, T., Rajan, D., Shivade, C., Asseman, A., Ozcan, A.: Leveraging medical visual question answering with supporting facts. In: Working Notes of CLEF 2019 (2019)

13. Lau, J.J., Gayen, S., Ben Abacha, A., Demner-Fushman, D.: A dataset of clinically generated visual questions and answers about radiology images. Scientific Data **5**(180251) (2018), https://www.nature.com/articles/sdata2018251
14. Liu, S., Ou, X., Che, J.: Vqa-med: An xception-gru model. In: Working Notes of CLEF 2019 (2019)
15. Papineni, K., Roukos, S., Ward, T., Zhu, W.J.: BLEU: a method for automatic evaluation of machine translation. In: Proceedings of the 40th annual meeting on association for computational linguistics. pp. 311–318. Association for Computational Linguistics (2002)
16. Shah, R., Gadgil, T., Bansal, M., Verma, P.: Medical visual question answering at imageclef 2019- vqa med. In: Working Notes of CLEF 2019 (2019)
17. Shi, L., Liu, F., Rosen, M.P.: Deep multimodal learning for medical visual question answering. In: Working Notes of CLEF 2019 (2019)
18. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings (2015), http://arxiv.org/abs/1409.1556
19. Spanier, A., Turner, A.: Lstm in vqa-med, is it really needed? validation study on the imageclef 2019 dataset. In: Working Notes of CLEF 2019 (2019)
20. Thanki, A., Makkithaya, K.: Mit manipal at imageclef 2019 visualquuestion answering in medical domain. In: Working Notes of CLEF 2019 (2019)
21. Vu, M., Sznitman, R., Nyholm, T., Lfstedt, T.: Ensemble of streamlined bilinear visual question answering models for the imageclef 2019 challenge in the medical domain. In: Working Notes of CLEF 2019 (2019)
22. Yan, X., Li, L., Xie, C., Xiao, J., Gu, L.: Zhejiang university at imageclef 2019 visual question answering in the medical domain. In: Working Notes of CLEF 2019 (2019)
23. Yang, Z., He, X., Gao, J., Deng, L., Smola, A.J.: Stacked attention networks for image question answering. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016. pp. 21–29 (2016), https://doi.org/10.1109/CVPR.2016.10
24. Zhou, Y., Kang, X., Ren, F.: Tua1 at imageclef 2019 vqa-med: A classification and generation model based on transfer learning. In: Working Notes of CLEF 2019 (2019)