Overview of ImageCLEFmedical 2023 – Medical Visual Question Answering for Gastrointestinal Tract

Steven Hicks[1,*], Andrea Storås[1,2], Pål Halvorsen[1,2], Thomas de Lange[3], Michael Riegler[1,2,4] and Vajira Thambawita[1]

[1]SimulaMet - Simula Metropolitan Center for Digital Engineering, Oslo, Norway
[2]OsloMet - Oslo Metropolitan University, Oslo, Norway
[3]Sahlgrenska University Hospital, Mölndal, Gothenburg, Sweden
[4]UiT - The Arctic University of Norway, Tromsø, Norway

Abstract

This paper provides an overview of the Medical Visual Question Answering for Gastrointestinal Tract (MedVQA-GI) challenge held at ImageCLEF 2023, a new challenge that combines visual-text question answering with colonoscopy analysis. The challenge is divided into three tasks, each tackling a different aspect of visual-text question answering. The first task focuses on answer generation based on an image and question, the second task focuses on question generation based on a given set of images and questions, and the last task is segmentation mask generation based on a given image and question. The paper includes details on the data collection and description, task specifics, evaluation methods, participation, and challenge results.

1. Introduction

Identifying lesions within gastrointestinal (GI) images is a popular application of machine learning with much research behind it [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11]. Until recently, the primary focus of GI analysis has been disease detection from videos or images, particularly polyp detection [12, 13, 14, 15, 16, 17, 18]. Plenty of challenges on this topic have been held for several years, showing steady progress in the field [19, 20, 21, 22, 23, 24, 25]. Most of these challenges focused on images or videos and tasks like classification or segmentation, with few focusing on how these solutions could be used in a real-world clinic. Models that analyze images and videos for endoscopy often provide only a number or mask to the user [26, 27]. This may be sufficient in some cases, but a more natural interaction between health professionals and artificial intelligence (AI) systems can lead to a better interpretable and trustworthy system [28]. Therefore, in this challenge, we aim to address the challenge of interaction between a user and a machine learning model using natural language in the form of a Visual Question Answering (VQA) task [29].

The challenge is divided into three tasks, each with its own unique requirements. First,

CLEF 2023: Conference and Labs of the Evaluation Forum, September 18–21, 2023, Thessaloniki, Greece
*Corresponding author.
steven@simula.no (S. Hicks); andrea@simula.no (A. Storås); paalh@simula.no (P. Halvorsen); thomas.de.lange@gu.se (T. d. Lange); michael@simula.no (M. Riegler); vajira@simula.no (V. Thambawita)

the VQA task asks participants to combine image data with textual questions to generate an answer to a question. This requires understanding the image and the textual data, requiring a multimodal approach to produce accurate results. Second, the Visual Question Generation (VQG) task demands participants to generate text-based questions derived from a given image and an associated answer. This task requires an understanding of the image and the answer to produce appropriate questions. Last, the Visual Location Question Answering (VLQA) provides participants with an image and a question about the location of a certain object, like a polyp, which should then be segmented and returned by the system. We see this challenge as a good opportunity to have the medical computer vision community contribute to a relatively new and novel use-case in medical image analysis, and see this as a perfect fit for ImageCLEF 2023 [30].

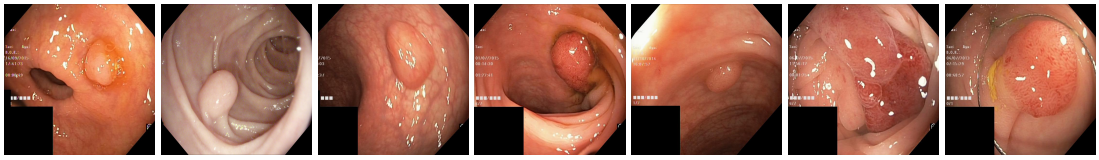
The rest of this paper is organized as follows. We start with an explanation of the dataset creation, it gives insights into how data was collected, validated, and organized. Then, we discuss the specific tasks involved in the MedVQA challenge and the evaluation methods used. In terms of participation, the document outlines key statistics that reveal participants' geographic and institutional diversity. Finally, the paper presents the results submitted by the participants.

2. Dataset Details

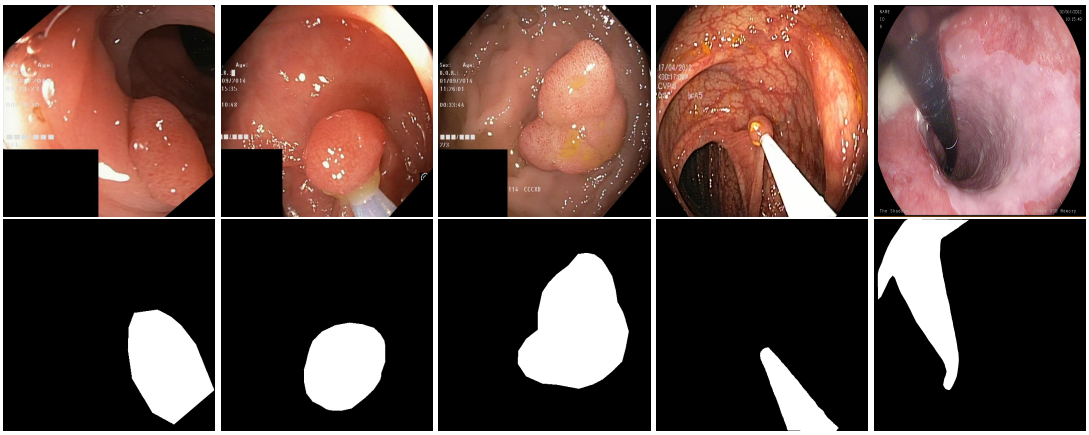
The data used for this challenge is based the HyperKvasir dataset [31] and the Kvasir-Instrument dataset [32], which are publicly accessible at datasets.simula.no/hyper-kvasir and datasets.simula.no/kvasir-instrument, respectively. The dataset for this challenge extends these datasets with question-and-answer ground truth data that we developed and collected in collaboration with our medical partners. The dataset spans the entire gastrointestinal tract, from the mouth to the anus, and encompass a wide array of different normal and abnormal findings. Furthermore, it also includes images of various surgical instruments used in gastrointestinal procedures, like colonoscopies and gastroscopies.

For Task 1 (VQA) and Task 2 (VQG), we provided a set of $2,000$ image samples for the development set and $1,949$ for the testing dataset. It is important to note that not all questions directly correspond to the image's content, meaning that some questions only have relevant answers for some of the images. In such instances, the submissions should be able to handle cases where there is no correct or relevant answer. For Task 3, segmentation masks are included for segmentation training and evaluation. These segmentation masks highlight specific regions of interest within the image. The masks only apply to certain parts of the whole dataset, namely those containing polyps and surgical equipment. An overview of the questions are shown in Table 1, which includes the associated question ID and the expected answer type. Samples from the dataset can be seen in Figure 1, including the ground truth format included in the development dataset.

As mentioned before, the visual parts of the dataset are taken from the HyperKvasir [31] and Kvasir-Instrument [32] datasets. We collected additional ground truth for the visual-text question answering data, for which the labeling was done by a set of computer scientists with assistance from medical professionals with several years experience within GI disease diagnostics. Annotations were created in LabelBox [33], where the computer scientists did the initial run of the annotations and the medical experts then went through and confirmed the



(a) Example images taken from the dataset.



(b) Example images from the dataset that have associated masks.

```
[
    {
        "ImageID": "<Image ID>,
        "Labels": [
            {
                "Question": <Question>,
                "AnswerType": <Answer Type>,
                "Answer": <Ground Truth Answer>
            },
        ]
    }
]
```

(c) The format of the provided JSON ground truth.

Figure 1: Examples from the development dataset that was provided by the challenge organizers.

annotations. It is worth noting that, due to lack of time, not all samples were validated by the medical experts. A more complete version of the challenge dataset will be released in the future, which will contain more samples and complete verification by the domain experts.

3. Task Description and Evaluation

This section details the three tasks that are part of this challenge: VQA, VQG, and VLQA. Each task is designed to assess different aspects handling both textual and visual data. The scripts

Table 1
An overview of the questions included in the dataset for Task 1 and Task 3. Task 2 did not have questions as they were to be generated by the participants.

Task	ID	Question	Answer Type
Task 1	1	What type of procedure is the image taken from?	Text
	2	Have all polyps been removed?	Binary
	3	Is this finding easy to detect?	Binary
	4	Is there a green/black box artifact?	Binary
	5	Is there text?	Binary
	6	What color is the abnormality?	Text
	7	What color is the anatomical landmark?	Text
	8	How many findings are present?	Number
	9	How many polyps are in the image?	Number
	10	How many instruments are in the image?	Number
	11	Where in the image is the abnormality?	Text
	12	Where in the image is the instrument?	Text
	13	Are there any abnormalities in the image?	Text
	14	Are there any anatomical landmarks in the	Text
	15	Are there any instruments in the image?	Text
	16	Where in the image is the anatomical landmark?	Text
	17	What is the size of the polyp?	Text
	18	What type of polyp is present?	Text
Task 3	19	Where exactly in the image is the polyp?	Segmentation
	20	Where exactly in the image is the instrument?	Segmentation

used to verify and evaluate the submissions were provided in our public GitHub repository[1].

3.1. Task 1: Visual Question Answering

The VQA task challenges participants to generate accurate and descriptive text-based answers in response to given text questions and corresponding images. An example scenario might involve an image portraying a colon polyp, accompanied by the question, "Where in the image is the polyp located?" In response, participants should provide a textual description specifying the polyp's location within the image. Such a description could refer to spatial locations like "upper-left" or "center". This task gauges the participants' proficiency in interpreting medical images and translating that interpretation into clear, spatially-referenced, text-based answers.

For submission to this task, participants were instructed to generate a javaScript object notation (JSON) file that encapsulates their task responses corresponding to each image in the designated test dataset. Each entry within this JSON file should contain an entry for each image provided in the testing dataset, where each entry contains 20 question-answer pairs. Here, the questions are pre-defined, while the corresponding answers are by the participant's system. A visualization of the submission format for Task 1 is shown in Figure 2.

Submissions were evaluated by first preforming a set of prepossessing steps on the answers

[1]https://github.com/ImageCLEF/2023_ImageCLEFmed_VQA

```
{
    <Image ID>: [
        {
            "QuestionID": <Question ID>,
            "Question": <Question Text>,
            "Answer": <Predicted Answer>
        },
    ],
}
```

Figure 2: The submission format for Task 1 and Task 3.

and ground truth, then comparing them directly and calculating the metrics. We used standard accuracy as the primary metrics, which was provided on a global-level, question-level, and image-level.

3.2. Task 2: Visual Question Generation

The VQG task inverts the first task's approach by asking participants to generate text questions based on a given text answer and an image pair. For instance, if the provided answer is "The image contains a polyp", and the accompanying image indeed contains a polyp, the participant should generate a question such as "Does the image contain an abnormality?". The complexity of this task lies in its requirement for a deep understanding of the image content and the ability to formulate relevant questions based on that understanding.

Submissions for this task were open, and participants were allowed to submit whatever they pleased. This could include, for example, software, source code, or system-specific documentation. Evaluations for this task were subjective, and performed by the challenge organizers, with no objective score tied to it.

3.3. Task 3: Visual Location Question Answering

The VLQA task extends beyond text-based responses, asking participants to create segmented parts of an image based on a given text question and image pair. This task diverges from the previous two tasks as it necessitates a visual output—a segmentation mask—rather than a textual one. Consider a scenario where the question posed is "Where is the abnormality?" and the provided image contains a polyp. The expected output is a segmentation mask outlining the polyp's location in the image. The VLQA task, therefore, assesses the participants' competence in identifying and visually demarcating areas of interest within the medical imagery, based on text-based inquiries.

Submissions to this task is quite similar to the first but instead of a textual answer, the answer is a segmentation mask corresponding to the posed question. Here, participants were asked to submit a JSON file in the format as described in Task 1, answering the questions "Where exactly in the image is the polyp?" and "Where exactly in the image is the instrument?", where the answers would be the name of the mask file that was also included in the submission.

Table 2
An overview of the submissions to each task availalbe at MedVQA-GI.

Team Name	# Runs for Task 1	# Runs for Task 2	# Runs for Task 3	Working Notes
wsq4747	1		1	[34]
BITM	1			[35]
SSNSheerinKavitha	2	1		[36]
SSN_KDC	1			
utk	1	1		[37]
VisionQAries	1	1	1	[38]
DLNU_CCSE	1			
UIT-Saviors	2	1		[39]
Total	10	4	2	6

Table 3
The results calculated by the challenge organizers based on the submissions for Task 3 delivered by the participants. Note that only the best-performing run is represented in the table and only two of the eight teams submitted to Task 3.

Team Name	Precision	Recall	F1-score / Dice	IoU
wsq4747	0.295	0.258	0.258	0.234
VisionQAries	0.680	0.681	0.677	0.666

This task was evaluated by comparing the participants' outputs with the ground truth segmentation masks, using standard segmentation metrics like precision, recall, Dice, and Intersection over Union (IoU).

4. Participation and Results

This section provides an overview of the participation in the challenge, and discusses the results submitted by those who completed it. An overview of the submissions to each task is shown in Table 2.

4.1. Participation

In total, 26 teams signed up for the task, 8 teams submitted runs to solutions, 6 teams wrote and submitted a working notes paper. Although a drop from 26 down to 6 may seem like a lot, our experience with previous challenges has shown that about 1/4 of registrations end up finishing the challenge [21, 20, 40, 41, 23]. The participants of the challenge represent countries from all over the world, including China, India, Norway, Pakistan, Iran, Italy, Nepal, Poland, Tunisia, Vietnam, The United States, and Saudi Arabia.

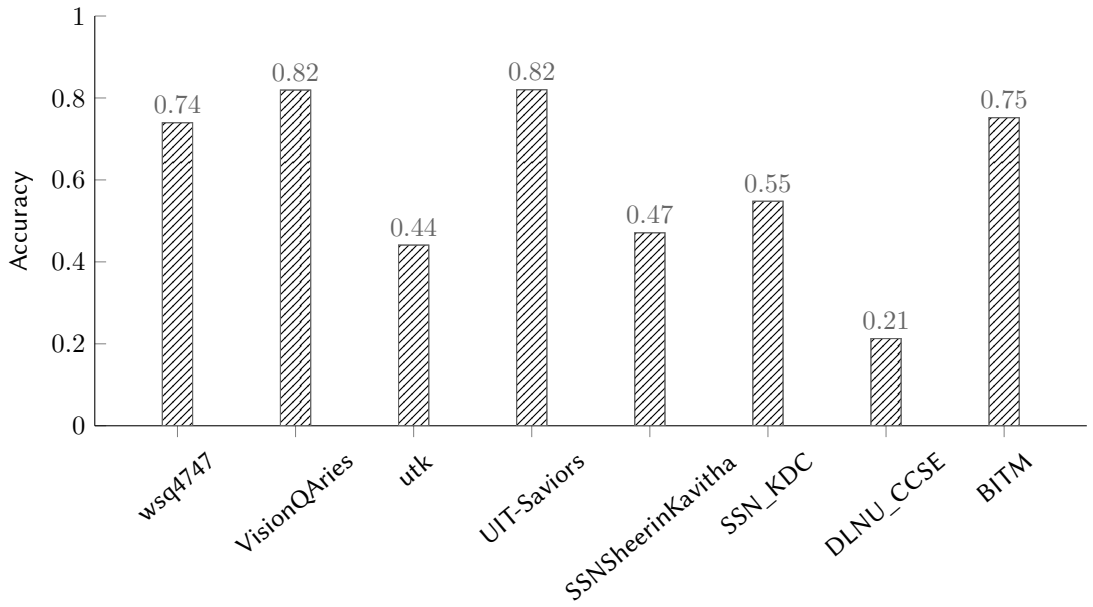


Figure 3: The results calculated by the challenge organizers based on the submissions for Task 1 delivered by the participants. The y-axis represents the accuracy of the submission and the x-axis is the team responsible for the submission. Note that only the best-performing run is represented in the graph.

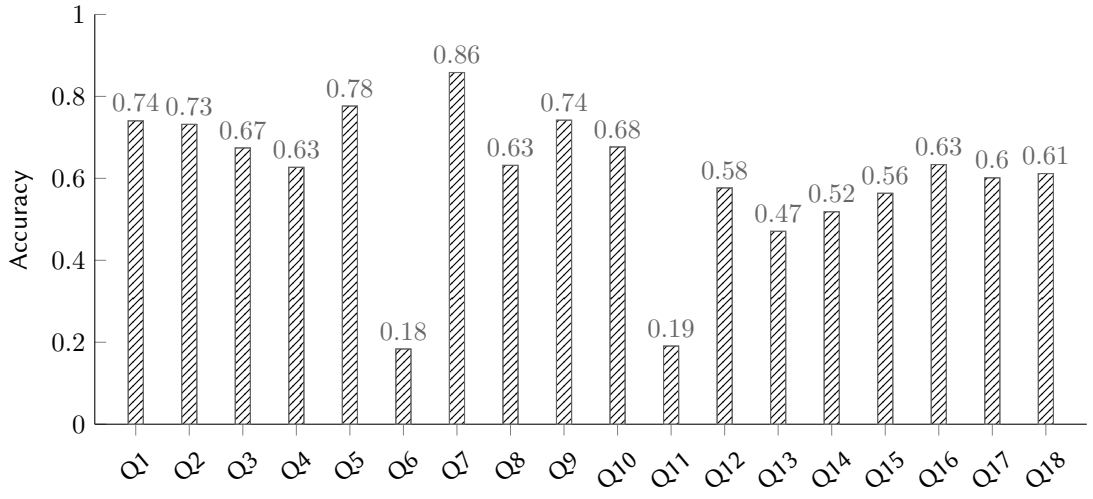


Figure 4: Accuracy scores for each question average across the best runs from all participants.

4.2. Results

Figure 3 and Table 3 show the results for Task 1 and Task 3, respectively. For task 2, we proposed a subjective evaluation of the submissions, however, most teams submitted an inverse of task 1, which does not provide any meaningful information gain. For future versions of this task,

we will develop a separate ground truth and include more strict task requirements. Looking at Figure 3, we see that most teams achieve reasonably good results, with two teams reaching above 80% accuracy. Looking at Figure 4, it looks like "What color is the abnormality?" (Q6) and "Where in the image is the abnormality?" (Q11) were the most difficult questions. We believe this is mostly due to the subjective nature of these questions. For example, the color of the abnormality can vary based on the user making the annotations.

More details about the results of the specific teams can be found in their corresponding working notes paper [35, 36, 37, 38, 39].

5. Conclusion and Future Outlook

This paper presented the MedVQA-GI challenge, which was held for the first time at ImageCLEF 2023. The challenge presented three tasks related to visual-text question answering and had eight participants submit results to at least on of the three available tasks. We believe that this is a promising start for the MedVQA-GI challenge, with several quality submissions. In the future, we plan on expanding the dataset to cover diverse conditions and instruments, refining evaluation metrics, and adding a larger and more diverse question set. Furthermore, we would like to expand on Task 2 to be more robust and include stricter participation criteria.

References

[1] C. Hassan, M. Spadaccini, A. Iannone, R. Maselli, M. Jovani, V. T. Chandrasekar, G. Antonelli, H. Yu, M. Areia, M. Dinis-Ribeiro, et al., Performance of artificial intelligence in colonoscopy for adenoma and polyp detection: a systematic review and meta-analysis, Gastrointestinal endoscopy 93 (2021) 77–85.

[2] A. Alammari, A. R. Islam, J. Oh, W. Tavanapong, J. Wong, P. C. De Groen, Classification of ulcerative colitis severity in colonoscopy videos using cnn, in: Proceedings of the ACM International Conference on Information Management and Engineering (ACM ICIME), 2017, pp. 139–144. doi:https://doi.org/10.1145/3149572.3149613.

[3] D. Bychkov, N. Linder, R. Turkki, S. Nordling, P. E. Kovanen, C. Verrill, M. Walliander, M. Lundin, C. Haglund, J. Lundin, Deep learning based tissue analysis predicts outcome in colorectal cancer, Scientific Reports 8 (2018) 3395. URL: http://dx.doi.org/10.1038/s41598-018-21758-3. doi:https://doi.org/10.1038/s41598-018-21758-3.

[4] Y. Mori, S.-e. Kudo, M. Misawa, Y. Saito, H. Ikematsu, K. Hotta, K. Ohtsuka, F. Urushibara, S. Kataoka, Y. Ogawa, Y. Maeda, K. Takeda, H. Nakamura, K. Ichimasa, T. Kudo, T. Hayashi, K. Wakamura, F. Ishida, H. Inoue, H. Itoh, M. Oda, K. Mori, Real-Time Use of Artificial Intelligence in Identification of Diminutive Polyps During Colonoscopy: A Prospective Study, Annals of Internal Medicine 169 (2018) 357–366. doi:https://doi.org/10.7326/M18-0249.

[5] K. Pogorelov, S. L. Eskeland, T. de Lange, C. Griwodz, K. R. Randel, H. K. Stensland, D.-T. Dang-Nguyen, C. Spampinato, D. Johansen, M. Riegler, P. Halvorsen, A holistic multimedia system for gastrointestinal tract disease detection, in: Proceedings of the ACM

on Multimedia Systems Conference (MMSYS), 2017, pp. 112–123. doi:https://doi.org/10.1145/3193740.

[6] J. Silva, A. Histace, O. Romain, X. Dray, B. Granado, Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer, International Journal of Computer Assisted Radiology and Surgery 9 (2014) 283–293. doi:https://doi.org/10.1007/s11548-013-0926-3.

[7] V. L. Thambawita, D. Jha, H. L. Hammer, H. D. Johansen, D. Johansen, P. Halvorsen, M. Riegler, An extensive study on cross-dataset bias and evaluation metrics interpretation for machine learning applied to gastrointestinal tract abnormality classification, ACM Transactions on Computing for Healthcare (2020).

[8] D. Jha, M. Riegler, D. Johansen, P. Halvorsen, H. Johansen, Doubleu-net: A deep convolutional neural network for medical image segmentation, in: Proceeding of the International Symposium on Computer Based Medical Systems (CBMS), 2020.

[9] Q. Angermann, J. Bernal, C. Sánchez-Montes, M. Hammami, G. Fernández-Esparrach, X. Dray, O. Romain, F. J. Sánchez, A. Histace, Towards real-time polyp detection in colonoscopy videos: Adapting still frame-based methodologies for video sequences analysis, in: Proceedings of Computer Assisted and Robotic Endoscopy and Clinical Image-Based Procedures (CARE CLIP), volume 10550, Springer, 2017, pp. 29–41.

[10] K. Pogorelov, M. Riegler, P. Halvorsen, P. T. Schmidt, C. Griwodz, D. Johansen, S. L. Eskeland, T. de Lange, Gpu-accelerated real-time gastrointestinal diseases detection, in: Proceedings of the International Symposium on Computer-Based Medical Systems (CBMS)„ IEEE, 2016, pp. 185–190. doi:https://doi.org/10.1109/CBMS.2016.63.

[11] M. Riegler, K. Pogorelov, P. Halvorsen, T. de Lange, C. Griwodz, P. T. Schmidt, S. L. Eskeland, D. Johansen, EIR - efficient computer aided diagnosis framework for gastrointestinal endoscopies, in: Proceedings of the IEEE International Workshop on Content-Based Multimedia Indexing (CBMI), 2016, pp. 1–6. doi:https://doi.org/10.1109/CBMI.2016.7500257.

[12] Y. Wang, W. Tavanapong, J. Wong, J. H. Oh, P. C. De Groen, Polyp-alert: Near real-time feedback during colonoscopy, Computer Methods and Programs in Biomedicine 120 (2015) 164–179. doi:https://doi.org/10.1016/j.cmpb.2015.04.002.

[13] D. Jha, P. H. Smedsrud, M. A. Riegler, D. Johansen, T. De Lange, P. Halvorsen, H. D. Johansen, Resunet++: An advanced architecture for medical image segmentation, in: Proceedings of the International Symposium on Multimedia (ISM), 2019, pp. 225–230. doi:https://doi.org/10.1109/ISM46123.2019.00049.

[14] J. Bernal, A. Histace, M. Masana, Q. Angermann, C. Sánchez-Montes, C. Rodriguez, M. Hammami, A. Garcia-Rodriguez, H. Córdova, O. Romain, G. Fernández-Esparrach, X. Dray, J. Sanchez, Polyp detection benchmark in colonoscopy videos using gtcreator: A novel fully configurable tool for easy and fast annotation of image databases, in: Proceedings of Computer Assisted Radiology and Surgery (CARS), 2018. doi:https://hal.archives-ouvertes.fr/hal-01846141.

[15] Y. Guo, J. Bernal, B. J Matuszewski, Polyp segmentation with fully convolutional deep neural networks—extended evaluation study, Journal of Imaging 6 (2020) 69.

[16] M. Min, S. Su, W. He, Y. Bi, Z. Ma, Y. Liu, Computer-aided diagnosis of colorectal polyps using linked color imaging colonoscopy to predict histology, Scientific reports 9 (2019)

2881. doi:https://doi.org/10.1038/s41598-019-39416-7.

[17] N. M. Ghatwary, X. Ye, M. Zolgharni, Esophageal abnormality detection using densenet based faster r-cnn with gabor features, IEEE Access 7 (2019) 84374–84385. doi:https://doi.org/10.1109/ACCESS.2019.2925585.

[18] S. Shah, N. Park, N. E. H. Chehade, A. Chahine, M. Monachese, A. Tiritilli, Z. Moosvi, R. Ortizo, J. Samarasena, Effect of computer-aided colonoscopy on adenoma miss rates and polyp detection: a systematic review and meta-analysis, Journal of Gastroenterology and Hepatology 38 (2023) 162–176.

[19] S. Hicks, M. Riegler, P. Smedsrud, T. B. Haugen, K. R. Randel, K. Pogorelov, H. K. Stensland, D.-T. Dang-Nguyen, M. Lux, A. Petlund, T. de Lange, P. T. Schmidt, P. Halvorsen, Acm multimedia biomedia 2019 grand challenge overview, in: Proceedings of the ACM International Conference on Multimedia (ACM MM), 2019, pp. 2563–2567. doi:https://doi.org/10.1145/3343031.3356058.

[20] K. Pogorelov, M. Riegler, P. Halvorsen, S. A. Hicks, K. R. Randel, D.-T. Dang-Nguyen, M. Lux, O. Ostroukhova, T. De Lange, Medico multimedia task at mediaeval 2018, in: Proceeding of the MediaEval Benchmarking Initiative for Multimedia Evaluation Workshop (MediaEval), 2018.

[21] M. Riegler, K. Pogorelov, P. Halvorsen, K. Randel, S. Eskeland, D.-T. Dang-Nguyen, M. Lux, C. Griwodz, C. Spampinato, T. de Lange, Multimedia for medicine: the medico task at mediaeval 2017, in: Proceeding of the MediaEval Benchmarking Initiative for Multimedia Evaluation Workshop (MediaEval), 2017.

[22] J. Bernal, H. Aymeric, Miccai endoscopic vision challenge polyp detection and segmentation, https://endovissub2017-giana.grand-challenge.org/home/, 2017. Accessed: 2017-12-11.

[23] S. Hicks, M. Riegler, P. Smedsrud, T. B. Haugen, K. R. Randel, K. Pogorelov, H. K. Stensland, D.-T. Dang-Nguyen, M. Lux, A. Petlund, T. de Lange, P. T. Schmidt, P. Halvorsen, Acm multimedia biomedia 2019 grand challenge overview, in: Proceedings of the 27th ACM International Conference on Multimedia, MM '19, Association for Computing Machinery, New York, NY, USA, 2019, p. 2563–2567. doi:10.1145/3343031.3356058.

[24] D. Jha, S. Hicks, K. Emanuelsen, H. Johansen, D. Johansen, T. de Lange, M. Riegler, P. Halvorsen, Medico Multimedia Task at MediaEval 2020: Automatic Polyp Segmentation, in: Proc. of MediaEval 2020 CEUR Workshop, 2020.

[25] D. Jha, S. Ali, S. Hicks, V. Thambawita, H. Borgli, P. H. Smedsrud, T. de Lange, K. Pogorelov, X. Wang, P. Harzig, M.-T. Tran, W. Meng, T.-H. Hoang, D. Dias, T. H. Ko, T. Agrawal, O. Ostroukhova, Z. Khan, M. Atif Tahir, Y. Liu, Y. Chang, M. Kirkerød, D. Johansen, M. Lux, H. D. Johansen, M. A. Riegler, P. Halvorsen, A comprehensive analysis of classification methods in gastrointestinal endoscopy imaging, Medical Image Analysis 70 (2021) 102007. doi:10.1016/j.media.2021.102007.

[26] K. Pogorelov, M. Riegler, S. L. Eskeland, T. de Lange, D. Johansen, C. Griwodz, P. T. Schmidt, P. Halvorsen, Efficient disease detection in gastrointestinal videos–global features versus neural networks, Multimedia Tools and Applications 76 (2017) 22493–22525. doi:https://doi.org/10.1007/s11042-017-4989-y.

[27] V. Thambawita, D. Jha, M. Riegler, P. Halvorsen, H. L. Hammer, H. D. Johansen, D. Johansen, The medico-task 2018: Disease detection in the gastrointestinal tract using global features

and deep learning, in: Proceeding of the MediaEval Benchmarking Initiative for Multimedia Evaluation Workshop (MediaEval), 2018.
[28] E. LaRosa, D. Danks, Impacts on trust of healthcare ai, in: Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, 2018, pp. 210–215.
[29] S. Antol, A. Agrawal, J. Lu, M. Mitchell, D. Batra, C. L. Zitnick, D. Parikh, Vqa: Visual question answering, in: Proceedings of the IEEE international conference on computer vision, 2015, pp. 2425–2433.
[30] B. Ionescu, H. Müller, A. Drăgulinescu, W. Yim, A. Ben Abacha, N. Snider, G. Adams, M. Yetisgen, J. Rückert, A. García Seco de Herrera, C. M. Friedrich, L. Bloch, R. Brüngel, A. Idrissi-Yaghir, H. Schäfer, S. A. Hicks, M. A. Riegler, V. Thambawita, A. Storås, P. Halvorsen, N. Papachrysos, J. Schöler, D. Jha, A. Andrei, A. Radzhabov, I. Coman, V. Kovalev, A. Stan, G. Ioannidis, H. Manguinhas, L. Ştefan, M. G. Constantin, M. Dogariu, J. Deshayes, A. Popescu, Overview of ImageCLEF 2023: Multimedia retrieval in medical, socialmedia and recommender systems applications, in: Experimental IR Meets Multilinguality, Multimodality, and Interaction, Proceedings of the 14th International Conference of the CLEF Association (CLEF 2023), Springer Lecture Notes in Computer Science LNCS, Thessaloniki, Greece, 2023.
[31] H. Borgli, V. Thambawita, P. H. Smedsrud, S. Hicks, D. Jha, S. L. Eskeland, K. R. Randel, K. Pogorelov, M. Lux, D. T. D. Nguyen, et al., Hyperkvasir, a comprehensive multiclass image and video dataset for gastrointestinal endoscopy, Scientific data 7 (2020). doi:10.1038/s41597-020-00622-y.
[32] D. Jha, S. Ali, K. Emanuelsen, S. A. Hicks, V. Thambawita, E. Garcia-Ceja, M. A. Riegler, T. de Lange, P. T. Schmidt, H. D. Johansen, D. Johansen, P. Halvorsen, Kvasir-instrument: Diagnostic and therapeutic tool segmentation dataset in gastrointestinal endoscopy, in: Proceeedings of the International COnference on MultiMedia Modeling (MMM), 2021, pp. 218–229. doi:10.1007/978-3-030-67835-7_19.
[33] Labelbox, Labelbox, 2023.
[34] S. Wang, W. Zhou, Y. Yang, H. Huang, T. Zhang, Z. Ye, D. Yang, Adapting pre-trained visual and language models for medical image question answering, in: CLEF2023 Working Notes, CEUR Workshop Proceedings, CEUR-WS.org, Thessaloniki, Greece, 2023.
[35] S. Upadhyay, S. S. Tripathy, Bit mesra at imageclef 2023: Fusion of blended image and text features for medical vqa, in: CLEF2023 Working Notes, CEUR Workshop Proceedings, CEUR-WS.org, Thessaloniki, Greece, 2023.
[36] S. S. N. Mohamed, K. Srinivasan, R. Gopalsamy, Ssn mlrg at medvqa-gi 2023: Visual question generation and answering using transformer based pre-trained models, in: CLEF2023 Working Notes, CEUR Workshop Proceedings, CEUR-WS.org, Thessaloniki, Greece, 2023.
[37] R. R. Gunti, A. Rorissa, A dual of san's and vgg-16 model-based visual question answering evaluation, in: CLEF2023 Working Notes, CEUR Workshop Proceedings, CEUR-WS.org, Thessaloniki, Greece, 2023.
[38] P. Cieplicka, J. Kłos, M. Morawski, J. Opała, Language-based colonoscopy image analysis with pretrained neural networks, in: CLEF2023 Working Notes, CEUR Workshop Proceedings, CEUR-WS.org, Thessaloniki, Greece, 2023.
[39] T. M. Thai, A. T. Vo, H. K. Tieu, L. N. Bui, T. T. Nguyen, Uit-saviors at medvqa-gi 2023: Improving multimodal learning with image enhancement for gastrointestinal visual question

answering, in: CLEF2023 Working Notes, CEUR Workshop Proceedings, CEUR-WS.org, Thessaloniki, Greece, 2023.

[40] S. Hicks, D. Jha, V. Thambawita, P. Halvorsen, H. Hammer, M. Riegler, An Overview of the EndoTect Challenge at ICPR 2020, in: ICPR2020, 2020. doi:10.1007/978-3-030-68793-9_18.

[41] S. Hicks, V. Thabawita, H. L. Hammer, T. B. Haugen, P. Halvorsen, M. Riegler, ACM MM BioMedia 2020 Grand Challenge Overview, in: ACMMM2020, ACM MM '20, Association for Computing Machinery, New York, NY, USA, 2020.