ImageCLEF 2026: Multimodal Challenges in Medicine, Science, Agritech, and Security

⋆ Bogdan Ionescu[1], Henning Müller[2], Dan-Cristian Stanciu[1], Ahmedkhan Radzhabov[13], Alba García Seco de Herrera[4], Alexandra-Georgiana Andrei[1], Alexandra Băicoianu[19], Andrea Storås[11], Asma Ben Abacha[5], Benjamin Bracke[3], Lea Reinartz[3], Benjamin Lecouteux[6], Christoph M. Friedrich[3], Cynthia Sabrina Schmidt[8], Corneliu-Nicolae Florea[11,19], Diandra Fabre[6], Didier Schwab[6], Dimitar Dimitrov[9] and Emmanuelle Esperança-Rodier[6], Gabriel Constantin[1], Hendrik Damm[3], Henning Schäfer[8], Ivan Rodkin[12], Ivan Koychev[9], Josiane Mothe[17], Liviu-Daniel Ștefan[1], Maja J. Hjuler[6], Mehmet Kurt[18], Meliha Yetisgen[18], Michael A. Riegler[11], Mihai Dogariu[1], Mihai Ivanovici[19], Ming Shan Hee[12], Mohammad El Sakka[17], Momina Ahsan[12], Obioma Pelka[8], Pål Halvorsen[11], Preslav Nakov[12], Raphael Brüngel[3], Sarfraz Ahmad[12], Steven A. Hicks[11], Sushant Gautam[11], Tabea M. G. Pakull[8], Bahadır Eryılmaz[8], Vajira Thambawita[11], Vassili Kovalev[14,15], Wen-Wai Yim[5], and Zhuohan Xie[12]

[1] National University of Science and Technology Politehnica Bucharest, Romania
[2] University of Applied Sciences Western Switzerland (HES-SO), Switzerland
[3] University of Applied Sciences and Arts Dortmund, Germany
[4] Universidad Nacional de Educación a Distancia (UNED), Spain
[5] Microsoft, USA
[6] Univ. Grenoble Alpes, CNRS, Grenoble INP*, Grenoble, France
[7] Bauhaus-Universität Weimar, Germany
[8] University Hospital Essen, Germany
[9] Sofia University "St. Kliment Ohridski", Bulgaria
[10] University of Kassel, hessian.AI, and ScaDS.AI, Germany
[11] SimulaMet, Norway
[12] Mohamed bin Zayed University of Artificial Intelligence, United Arab Emirates
[13] Belarus National Academy of Sciences, Belarus
[14] Belarus State University, Belarus
[15] Hospital Clinic of Barcelona, Spain
[16] Stanford University, USA
[17] Université de Toulouse, France
[18] University of Washington, USA
[19] Transilvania University of Brasov, Romania

Abstract. Since its inception in 2003, the various ImageCLEF challenges have provided large and complex datasets targeting a wide array of subjects in medicine, argumentation, reasoning, content recommendation, data generation, and question answering. In its 24th edition, ImageCLEF will have five main tasks: (i) a *Medical* task, which aims to promote the synergy between four medical challenges: *Caption*, involving concept

⋆ apart from the general organisers, the authors are listed in alphabetical order.

detection and caption prediction in radiology images, Synthetic Medical Image Generation in the *GANs* task, *Visual Question Answering* for improving the diagnosis and classification of real medical gastrointestinal images, and multimodal dermatology response generation and a new *MEDIQA-CORE* challenge focusing on predicting or correcting tumor type labels and identifying and summarizing major and minor differences between pairs of radiology reports; (ii) the *ToPicto* task, involving text to pictogram translation and prediction, (iii) the *Multimodal Reasoning* task on visual, multi-language, interdisciplinary question answering, and 2 new tasks: (iv) *AI4Agriculture*, involving predicting agricultural potential before planing and crop type identification and (v) the *Deepfake* detection and generation task. In its last edition, 56 teams finished our challenges, continuing to show the impact in the community.

Keywords: Medical AI, Image Captioning, GANs, Visual Question Answering, Pictogram translation, Deepfake, Agricultural AI

1 Introduction

The ImageCLEF evaluation campaign, in its 24th edition, continues to provide a space for research and benchmark activities in the scientific community, providing challenges in various fields, such as medicine, argumentation, reasoning, content recommendation, data generation, and question answering. The 24th edition will take place in Jena, Germany, September 21-24, 2026 [20]. Over the years, both ImageCLEF and the CLEF campaign have demonstrated significant scholarly impact [29, 30], with the term ImageCLEF being mentioned in 7,980 papers since 2003, according to Google Scholar. This paper presents our five proposed tasks for this year's edition.

2 ImageCLEFmedical

The ImageCLEFmedical task is currently in its 22nd edition [20]. The 2026 edition introduces 4 medical tasks, listed below:

The *Caption* [21] task addresses the interpretation of clinical and diagnostic insights from radiology images. In this 10th edition [7, 9, 10, 21–23, 25–27], three subtasks are proposed: *concept detection*, *caption prediction*, and *explainability*. The *concept detection* subtask focuses on developing models capable of predicting Unified Medical Language System (UMLS®) Concept Unique Identifiers (CUIs) directly from medical images. The evaluation is based on the F1-Score [12]. This year, the updated and extended *ROCOv2* [28] dataset is used, incorporating additional figures from PubMed Central publications released after October 2024 and further refined through new annotations and validation. The *caption prediction* subtask evaluates systems designed to automatically generate descriptive and clinically meaningful captions for radiology images. Domain-specific language models such as BioBERT [17] continue to be employed for evaluation using

[20] https://clef2026.clef-initiative.eu/
[21] https://www.imageclef.org/2026/medical/caption

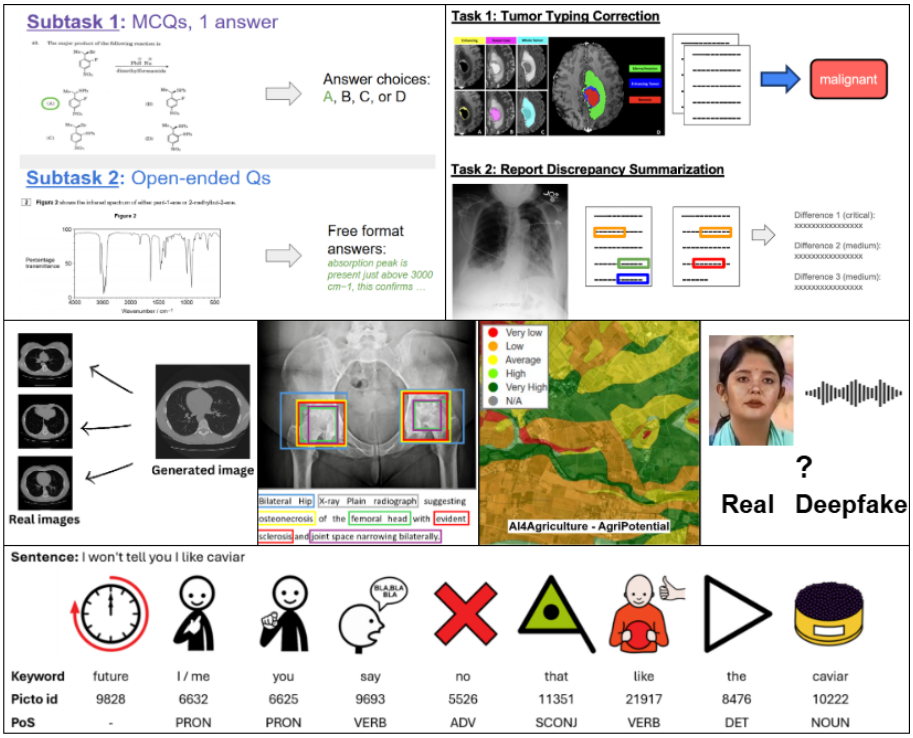


Fig. 1. Sample images from tasks - left to right, top to bottom - MultimodalReasoning, MEDIQA-CORE, GANs, Caption, AI4Agriculture, Deepfake, ToPicto

BERTScore, complemented by alternative metrics including ClinicalBLEURT and MedBERTScore [5]. Readability measures such as LENS [19] and the Coleman–Liau Index [6] (CLI) are used to assess linguistic quality, coherence, and clarity of the generated captions. The *explainability* subtask extends across both concept detection and caption prediction, requiring participants to provide interpretable visual or textual explanations for model predictions. These may include techniques such as heatmaps, attention maps, or Shapley-based attributions. Submitted explanations will be manually reviewed to evaluate interpretability and clinical relevance.

ImageCLEFmedical-GANs[22]. The 4th edition of the GANs task will continue evaluate both the quality and privacy of synthetic medical images generated by a various generative models, including Generative Adversarial Networks (GANs). Building upon the results of the previous editions [1–3], this iteration focuses on the critical challenge of ensuring patient privacy and data confidentiality in medical imaging while maintaining the fidelity, realism, and clinical utility of the

[22] https://www.imageclef.org/2026/medical/gans

generated data. In addition to the analytical components featured in previous editions where participants were asked to assess data, this year's task introduces a generation component, requiring participants to produce synthetic medical images themselves. This extension is designed to encourage a deeper exploration of privacy risks and mitigation strategies associated with generative models. Specifically, participants will investigate whether identifiable information, such as unique anatomical structures, latent image fingerprints, or subtle patterns from real patient scans, can be unintentionally reproduced by generative systems. The provided dataset will include both real and synthetic axial CT slices of tuberculosis patients captured at various stages of treatment.

ImageCLEFmedical-MEDVQA[23]. The 4th MEDVQA-GI challenge advances Visual Question Answering (VQA) for gastrointestinal (GI) imaging with a continued emphasis on explainability, clinical safety, and multimodal reasoning. Building on previous editions [13–15], the 2026 challenge introduces enhanced evaluation criteria that integrate behavioral safety and retrieval-augmented reasoning into the existing subtasks. The first subtask focuses on developing models that accurately answer clinically relevant questions from GI endoscopy images using the expanded Kvasir-VQA-x1 dataset [11] with more than 150,000 question–answer pairs. The second subtask extends the evaluation of explainability by requiring coherent multimodal justifications that combine textual and visual evidence aligned with medical reasoning while ensuring responses remain safe, factual, and consistent across clinical contexts. A dedicated safety layer evaluates whether models exhibit undesirable behaviors such as overconfidence, misleading explanations, or non-compliance with medical best practices during these tasks. To support retrieval-augmented generation methods, the challenge also provides a curated database of endoscopy resources containing verified clinical references. Evaluation integrates quantitative metrics for language accuracy with expert-based assessments of interpretability, safety, and clinical reliability.

ImageCLEFmedical-MEDIQA-CORE[24]. In this edition MEDIQA explores Multimodal Reasoning & Reconciliation in Radiology, a shared task designed to advance clinical AI systems capable of complex medical understanding and decision support. This competition comprises two complementary tasks that evaluate model performance on diagnosis accuracy and clinical report alignment using multimodal data. In the first task, we study tumor type prediction by jointly reasoning over radiology and pathology images, along with their corresponding reports. We assesses models' ability to integrate multimodal clinical evidence and improve diagnostic accuracy. In the second task, we focus on identifying and summarizing major and minor differences between pairs of radiology reports and associated images that describe the same patient but offer conflicting interpretations. It emphasizes clinical language and image understanding, fine-grained discrepancy detection, and reconciliation of divergent expert assess-

[23] https://www.imageclef.org/2026/medical/vqa
[24] https://www.imageclef.org/2026/medical/mediqa

ments. The competition will utilize real, de-identified clinical images and reports collected from multiple U.S. hospitals.

3 ImageCLEFtoPicto

In its third edition, ImageCLEFtoPicto [18] continues to advance research on the automatic conversion of text into meaningful sequences of pictograms designed to support Alternative and Augmentative Communication (AAC).

For the 2026 edition, the challenge introduces two new tasks: The first task, *Text-to-Picto* Translation, extends the previous editions while introducing a major shift: the source language will be English instead of French. Participants will be required to generate coherent sequences of pictogram labels from English text, adapting their systems to account for linguistic, syntactic, and semantic differences between the two modalities. The dataset provided for this task will include aligned pairs of English text and ARASAAC pictograms. This resource will either be obtained by translating the existing ProPicto–Orféo corpus into English or by curating a new parallel corpus from scratch.

The second task, *Next-Pictogram Prediction*, is a newly introduced subtask inspired by Pereira et al. [24]. It focuses on predicting the most probable next pictogram in a given sequence, with performance evaluated using a top-N accuracy metric.

4 MultimodalReasoning

The 2nd edition of **Multimodal Reasoning** will maintain its focus on assessing the reasoning capabilities of Vision-Language Models (VLMs) in the context of Multilingual Visual Question Answering (VQA). Building upon the first edition of the task [8], this iteration will expand the dataset within the multiple-choice question (MCQ) format - **Subtask 1**, preserving the original formulation: given an image of a question accompanied by two to five possible answers, participants must determine the single correct answer.

To further stimulate reasoning and mitigate the risk of models relying solely on pattern recognition or memorized responses, the second edition will additionally introduce open-ended questions in multiple languages. These will form **Subtask 2** within the **Multimodal Reasoning** challenge, requiring participants to generate answers directly from the visual and textual information presented in the image, without predefined alternatives. This extension is designed to promote the development and evaluation of VLMs capable of more robust multimodal reasoning and linguistic comprehension.

5 AI4Agriculture

AI4agriculture task follows AI4AGRI European project. It is composed of two sub-tasks:

AgriPotential. The AgriPotential sub-task is linked to sustainable agriculture. The challenge is to evaluate the capability of artificial intelligence models

to predict the potential of land for cultivating different crop types from multi-spectral and multi-temporal Sentinel-2 imagery. Participants will work with the AgriPotential dataset that covers regions of Southern France. Images are annotated at the pixel level for three crop types - viticulture, market gardening, and field crop -based on traditional in-situ analysis. The objective is to automatically estimate the agricultural potential across five ordinal levels, from « very low » to « very high », by modelling spectral and temporal features derived from the data.

DACIA5 sub-task is divided into two main challenges: *Crop identification: Past_ vs_ present*, and *Early crop identification* using the DACIA5 dataset [4] available at Zenodo [16]) and any additional open data. Submissions will be evaluated per track using specific quality metrics, primarily focusing on accuracy measurements, including overall accuracy and average accuracy across different crop classes, as well as accuracy for specific crop types. The dataset provided for this task contains all the available images from Sentinel-1 SAR and Sentinel-2 MSI from 2020-2024 over a specific area to the north of Brașov city, Romania, together with the 32×32 pixel radar and multi-spectral patches for crop identification. The labels represent 17 types of agricultural crops over 47 parcels.

6 ImageCLEF-Deepfake

The first edition of the Deepfake task aims to evaluate the existing means of deepfake creation and detection in two modalities: **visual and audio**. For each modality, the task is split into 2 sub-tasks: **detection and generation**. The task strives to advance the deepfake detection field by drawing conclusions regarding what makes a good deepfake detector, as well as what makes a deepfake easy or hard to detect.

The **Generation** sub-task challenges participants to generate image or audio deepfakes, based on set of real data. The generated deepfakes will be evaluated using a combination of multiple methods: (1) an identity score, to ensure that the generated deepfake resembles the original identity, (2) a detection score based on how well it can mislead several state-of-the-art deepfake detection methods, (3) a detection score based on how well the other methods in the Detection sub-task can detect the deepfakes.

The **Detection** sub-task is a binary classification problem for deepfakes. The deepfakes will be generated in 2 ways: (1) by the organizers, using state-of-the-art methods, (2) by the other participants, in the Generation sub-task. The final score will combine the individual scores for the two evaluation sets.

7 Conclusions

This paper showcases the tasks for the 24th edition of ImageCLEF. This year's challenges continue to provide opportunities for researchers to assess the performance of their systems in fields like medicine, science, security, agritech. The tasks present rich multimodal challenges with well-defined datasets and evaluation benchmarks, allowing researchers to develop and test their systems under realistic conditions.

Acknowledgement.

The work of Dan-Cristian Stanciu, Alexandra-Georgiana Andrei is funded by UEFISCDI, Solutions Axis, grant SOL4/2024. The work of Gabriel Constantin is funded by UEFISCDI, Solutions Axis, grant SOL12/2024. The work of Bogdan Ionescu, Liviu-Daniel Stefan is funded by the 5G-ENRICH European project, Project ID: 101181353 (23-EU-DIG-5G-ENRICH).

The work of Dimitar Dimitrov and Ivan Koychev is partially funded by the EU NextGenerationEU project, through the National Recovery and Resilience Plan of the Republic of Bulgaria, project SUMMIT, No BG-RRP-2.004-0008.

The work of Benjamin Bracke, Lea Reinartz and Raphael Brüngel was partially funded by a PhD grant from the University of Applied Sciences and Arts Dortmund (FH Dortmund), Germany. The work of Hendrik Damm, Tabea M. G. Pakull, Bahadır Eryılmaz and Henning Schäfer was funded by a PhD grant from the DFG Research Training Group 2535 Knowledge- and data-based personalisation of medicine at the point of care (WisPerMed).

For the AI4Agriculture task, the creation of the DACIAS dataset was funded by the European Union through the AI4AGRI project entitled Romanian Excellence Center on Artificial Intelligence on Earth Observation Data for Agriculture⁉ received funding from the EU's Horizon Europe research and innovation programme under the grant agreement no. 101079136. The Agripotential dataset was also partially funded by that project.

References

1. Andrei, A., Radzhabov, A., Karpenka, D., Prokopchuk, Y., Kovalev, V., Ionescu, B., Müller, H.: Overview of 2024 imageclefmedical gans task–investigating generative models' impact on biomedical synthetic images. In: CLEF2024 Working Notes, CEUR Workshop Proceedings, CEUR-WS. org, Grenoble, France (2024)
2. Andrei, A.G., Constantin, M.G., Dogariu, M., Radzhabov, A., Ştefan, L., Prokopchuk, Y., Kovalev, V., Müller, H., Ionescu, B.: Overview of imageclefmedical 2025 gans task: training data analysis and fingerprint detection. In: CLEF2025 Working Notes, CEUR Workshop Proceedings, CEUR-WS. org, Madrid, Spain (2025)
3. Andrei, A.G., Radzhabov, A., Coman, I., Kovalev, V., Ionescu, B., Müller, H.: Overview of imageclefmedical gans 2023 task: identifying training data "fingerprints" in synthetic biomedical images generated by gans for medical image security. In: Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2023). vol. 3497 (2023)
4. Baicoianu, A., Plajer, I.C., Debu, M., Stefan, M., Ivanovici, M., Florea, C., Cataron, A., Coliban, R.M., Popa, S., Oprisescu, S., Racoviteanu, A., Olteanu, G., Marandskiy, K., Ghinea, A., Kazak, A., Majercsik, L., Manea, A., Dogar, L.: Dacia5: a sentinel-1 and sentinel-2 dataset for agricultural crop identification applications. Big Earth Data **0**(0), 1–32 (2025). https://doi.org/10.1080/20964471.2025.2512685, https://doi.org/10.1080/20964471.2025.2512685
5. Ben Abacha, A., Yim, W.w., Michalopoulos, G., Lin, T.: An investigation of evaluation methods in automatic medical note generation.

In: Findings of the Association for Computational Linguistics: ACL 2023. pp. 2575–2588. Association for Computational Linguistics, Toronto, Canada (Jul 2023). https://doi.org/10.18653/v1/2023.findings-acl.161, https://aclanthology.org/2023.findings-acl.161
6. Coleman, M., Liau, T.L.: A computer readability formula designed for machine scoring. Journal of Applied Psychology **60**(2), 283–284 (1975). https://doi.org/10.1037/h0076540, 515 citations (Crossref/DOI) [2024-10-16] 515 citations (Crossref/DOI) [2024-10-16]
7. Damm, H., Pakull, T.M.G., Becker, H., Bracke, B., Eryilmaz, B., Bloch, L., Brüngel, R., Schmidt, C.S., Rückert, J., Pelka, O., Schäfer, H., Idrissi-Yaghir, A., Abacha, A.B., de Herrera, A.G.S., Müller, H., Friedrich, C.M.: Overview of ImageCLEFmedical 2025 – medical concept detection and interpretable caption generation. In: CLEF 2025 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org, Madrid, Spain (September 9–12 2025)
8. Dimitrov, D., Hee, M.S., Xie, Z., Joyti Das, R., Ahsan, M., Ahmad, S., Paev, N., Koychev, I., Nakov, P.: Overview of ImageCLEF 2025 – Multimodal Reasoning. In: CLEF 2025 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org, Madrid, Spain (September 9–12 2025)
9. Eickhoff, C., Schwall, I., García Seco de Herrera, A., Müller, H.: Overview of ImageCLEFcaption 2017 – the image caption prediction and concept extraction tasks to understand biomedical images. In: Working Notes of Conference and Labs of the Evaluation Forum (CLEF 2017). CEUR Workshop Proceedings, vol. 1866. CEUR-WS.org (2017)
10. García Seco De Herrera, A., Eickhof, C., Andrearczyk, V., Müller, H.: Overview of the ImageCLEF 2018 caption prediction tasks. In: Working Notes of Conference and Labs of the Evaluation Forum (CLEF 2018). CEUR Workshop Proceedings, vol. 2125. CEUR-WS.org (2018)
11. Gautam, S., Riegler, M.A., Halvorsen, P.: Kvasir-VQA-x1: A Multimodal Dataset for Medical Reasoning and Robust MedVQA in Gastrointestinal Endoscopy. arXiv (Jun 2025). https://doi.org/10.48550/arXiv.2506.09958
12. Goutte, C., Gaussier, E.: A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In: Advances in Information Retrieval – 27th European Conference on IR Research (ECIR 2005). pp. 345–359. Springer (2005)
13. Hicks, S.A., Halvorsen, P., Riegler, M.A., Thambawita, V.: Overview of ImageCLEFmedical 2025 – Medical Visual Question Answering for Gastrointestinal Tract. In: CLEF2025 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org, Gernoble, France (September 9-12 2025)
14. Hicks, S.A., Storås, A., Halvorsen, P., de Lange, T., Riegler, M.A., Thambawita, V.: Overview of ImageCLEFmedical 2023 – Medical Visual Question Answering for Gastrointestinal Tract. In: CLEF2023 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org, Thessaloniki, Greece (September 18-21 2023)
15. Hicks, S.A., Storås, A., Halvorsen, P., Riegler, M.A., Thambawita, V.: Overview of ImageCLEFmedical 2024 – Medical Visual Question Answering for Gastrointestinal Tract. In: CLEF2024 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org, Gernoble, France (September 9-12 2024)
16. Ivanovici, M., Baicoianu, A., Plajer, I.C., Debu, M., Ștefan, F.M., Florea, C., Cataron, A., Coliban, R.M., Popa, S., Oprisescu, S., Racoviteanu, A., Olteanu, G., Marandskiy, K., Ghinea, A., Kazak, A., Majercsik, L., Manea, A., Dogaru, L.: Ai4agri sentinel-2 brasov area 2020-2024 multi- spectral dataset for crop monitoring and identification (Dec 2024). https://doi.org/10.5281/zenodo.14283243, https://doi.org/10.5281/zenodo.14283243

17. Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C.H., Kang, J.: BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinform. **36**, 1234 – 1240 (2020). https://doi.org/10.1093/bioinformatics/btz682
18. Macaire, C., Fabre, D., Lecouteux, B., Schwab, D.: Overview of the 2025 imagecleftopicto task – investigating the generation of pictogram sequences from text and speech. In: CLEF 2025 Working Notes. CEUR Workshop Proceedings, Madrid, Spain (2025), https://ceur-ws.org/Vol-4038/paper$_1$73.*pdf*, *creativeCommonsCCBY* 4.0
19. Maddela, M., Dou, Y., Heineman, D., Xu, W.: LENS: A learnable evaluation metric for text simplification. In: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 16383–16408. Association for Computational Linguistics (2023). https://doi.org/10.18653/v1/2023.acl-long.905, https://aclanthology.org/2023.acl-long.905, 4 citations (Crossref/DOI) [2024-10-16] 4 citations (Crossref/DOI) [2024-10-16] 2 citations (Crossref/DOI) [2024-09-11]
20. Müller, H., Kalpathy-Cramer, J., García Seco de Herrera, A.: Experiences from the ImageCLEF medical retrieval and annotation tasks. In: Information Retrieval Evaluation in a Changing World, pp. 231–250. Springer (2019)
21. Pelka, O., Abacha, A.B., García Seco de Herrera, A., Jacutprakart, J., Friedrich, C.M., Müller, H.: Overview of the ImageCLEFmed 2021 concept & caption prediction task. In: Working Notes of Conference and Labs of the Evaluation Forum (CLEF 2021). CEUR Workshop Proceedings, vol. 2936. CEUR-WS.org (2021)
22. Pelka, O., Friedrich, C.M., García Seco de Herrera, A., Müller, H.: Overview of the ImageCLEFmed 2019 concept detection task. In: Working Notes of Conference and Labs of the Evaluation Forum (CLEF 2019). CEUR Workshop Proceedings, vol. 2380. CEUR-WS.org (2019)
23. Pelka, O., Friedrich, C.M., García Seco de Herrera, A., Müller, H.: Overview of the ImageCLEFmed 2020 concept prediction task: Medical image understanding. In: Working Notes of Conference and Labs of the Evaluation Forum (CLEF 2020). CEUR Workshop Proceedings, vol. 2696. CEUR-WS.org (2020)
24. Pereira, J.A., Macêdo, D., Zanchettin, C., de Oliveira, A.L.I., do Nascimento Fidalgo, R.: Pictobert: Transformers for next pictogram prediction. Expert Systems with Applications **202**, 117231 (2022). https://doi.org/https://doi.org/10.1016/j.eswa.2022.117231, https://www.sciencedirect.com/science/article/pii/S095741742200611X
25. Rückert, J., Ben Abacha, A., G. Seco de Herrera, A., Bloch, L., Brüngel, R., Idrissi-Yaghir, A., Schäfer, H., Bracke, B., Damm, H., Pakull, T.M.G., Schmidt, C.S., Müller, H., Friedrich, C.M.: Overview of ImageCLEFmedical 2024 – Caption Prediction and Concept Detection. In: CLEF2024 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org, Grenoble, France (September 9-12 2024)
26. Rückert, J., Ben Abacha, A., García Seco de Herrera, A., Bloch, L., Brüngel, R., Idrissi-Yaghir, A., Schäfer, H., Müller, H., Friedrich, C.M.: Overview of ImageCLEFmedical 2022 – Caption Prediction and Concept Detection. In: CLEF2022 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org, Bologna, Italy (September 5-8 2022)
27. Rückert, J., Ben Abacha, A., G. Seco de Herrera, A., Bloch, L., Brüngel, R., Idrissi-Yaghir, A., Schäfer, H., Müller, H., Friedrich, C.M.: Overview of ImageCLEFmedical 2023 – Caption Prediction and Concept Detection. In: CLEF2023 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org, Thessaloniki, Greece (September 18-21 2023)
28. Rückert, J., Bloch, L., Brüngel, R., Idrissi-Yaghir, A., Schäfer, H., Schmidt, C.S., Koitka, S., Pelka, O., Abacha, A.B., de Herrera, A.G.S., Müller, H., Horn,

P.A., Nensa, F., Friedrich, C.M.: Rocov2: Radiology objects in context version 2, an updated multimodal image dataset. Scientific Data **11**(1), 688 (June 2024). https://doi.org/10.1038/s41597-024-03496-6, https://doi.org/10.1038/s41597-024-03496-6

29. Tsikrika, T., García Seco de Herrera, A., Müller, H.: Assessing the scholarly impact of ImageCLEF. In: CLEF 2011. pp. 95–106. Springer Lecture Notes in Computer Science (LNCS) (2011)
30. Tsikrika, T., Larsen, B., Müller, H., Endrullis, S., Rahm, E.: The scholarly impact of CLEF (2000–2009). In: Information Access Evaluation. Multilinguality, Multimodality, and Visualization, pp. 1–12. Springer (2013)