Overview of the ImageCLEFmed 2019 Concept Detection Task

Obioma Pelka[1,2][0000−0001−5156−4429], Christoph M. Friedrich[1,3][0000−0001−7906−0038], Alba G. Seco de Herrera[4][0000−0002−6509−5325], and Henning Müller[5,6][0000−0001−6800−9878]

[1] Department of Computer Science, University of Applied Sciences and Arts Dortmund, Germany
{obioma.pelka,christoph.friedrich}@fh-dortmund.de
[2] Department of Diagnostic and Interventional Radiology and Neuroradiology, University Hospital Essen, Germany
[3] Institute for Medical Informatics, Biometry and Epidemiology (IMIBE), University Hospital Essen, Germany
[4] University of Essex, UK
alba.garcia@essex.ac.uk
[5] University of Applied Sciences Western Switzerland (HES-SO), Switzerland
henning.mueller@hevs.ch
[6] University of Geneva, Switzerland

Abstract. This paper describes the ImageCLEF 2019 Concept Detection Task. This is the 3rd edition of the medical caption task, after it was first proposed in ImageCLEF 2017. Concept detection from medical images remains a challenging task. In 2019, the format changed to a single subtask and it is part of the medical tasks, alongside the tuberculosis and visual question and answering tasks. To reduce noisy labels and limit variety, the data set focuses solely on radiology images rather than biomedical figures, extracted from the biomedical open access literature (PubMed Central). The development data consists of 56,629 training and 14,157 validation images, with corresponding Unified Medical Language System (UMLS®) concepts, extracted from the image captions. In 2019 the participation is higher, regarding the number of participating teams as well as the number of submitted runs. Several approaches were used by the teams, mostly deep learning techniques. Long short-term memory (LSTM) recurrent neural networks (RNN), adversarial auto-encoder, convolutional neural networks (CNN) image encoders and transfer learning-based multi-label classification models were the frequently used approaches. Evaluation uses F1-scores computed per image and averaged across all 10,000 test images.

Keywords: Concept Detection · Computer Vision · ImageCLEF 2019 · Image Understanding · Radiology

1 Introduction

The concept detection task presented in this paper is part of the ImageCLEF[1] benchmarking campaign, that is part of the Cross Language Evaluation Forum[2] (CLEF). ImageCLEF was first held in 2003 and in 2004 a medical task was added that has been held every year [10, 6]. More information regarding other proposed tasks in 2019 can be found in [5].

The caption task was first proposed in 2016 as a caption prediction task. In 2017, the caption task was split into two subtasks: concept detection and caption prediction and ran in that format at ImageCLEFcaption 2017 [1] and 2018 [4]. The format has slightly changed in 2019 with a single task.

The motivation for this task is that an increasing number of images has become available without metadata, so obtaining some metadata is essential to make the content usable. The objective is to develop systems capable of predicting concepts automatically for radiology images, or possibly for other clinical images. These predicted concepts enable order for unlabeled and unstructured radiology images and for data sets lacking metadata, as multi-modal approaches prove to obtain better results regarding image classification [12]. As the interpretation and summarization of knowledge from medical images such as radiology output is time-consuming, there is a considerable need for automatic methods that can approximate this mapping from visual information to condensed textual descriptions. The more image characteristics are known, the more structured are the radiology scans and hence, the more efficient are the radiologists regarding interpretation.

For development data, a subset of the Radiology Object in COntext data set (ROCO) [11] is used. ROCO contains radiology images originating from the PubMed Central (PMC) Open Access Subset[3] [14], with several Unified Medical Language System (UMLS®) Concept Unique Identifiers (CUIs) per image. The test set used for official evaluation was created in the same manner as proposed in Peltka et al. [11].

This paper presents an overview of the ImageCLEFmed Concept Detection Task 2019 with task description and participating teams in Section 2, an exploratory analysis on the data set and ground truth described in Section 3 and the evaluation framework explained in Section 4. The approaches applied by the participating teams are listed in Section 5, which is followed by discussion and conclusions in Section 6.

2 Task and Participation

Succeeding the previous subtasks in ImageCLEFcaption 2017 [1] and ImageCLEFcaption 2018 [4], a concept detection task with the objective of extracting UMLS®CUIs from radiology images was proposed. We work on the basis of

[1] http://imageclef.org/ [last accessed: 06.06.2019]
[2] http://www.clef-initiative.eu/ [last accessed: 06.06.2019]
[3] https://www.ncbi.nlm.nih.gov/pmc/tools/openftlist/[last accessed: 29.05.2019]

a large-scale collection of figures from biomedical open access journal articles (PMC). All images in the training data are accompanied by UMLS®concepts extracted from the original image caption. An example of an image from the training set with the extracted concepts is shown in Figure 1. In comparison to the previous tasks, the following improvements were made:

- To reduce the variety of content and focus the scenario, the images in the distributed collection are limited to radiology images.
- The number of concepts was decreased by preprocessing the captions, prior to concept extraction.

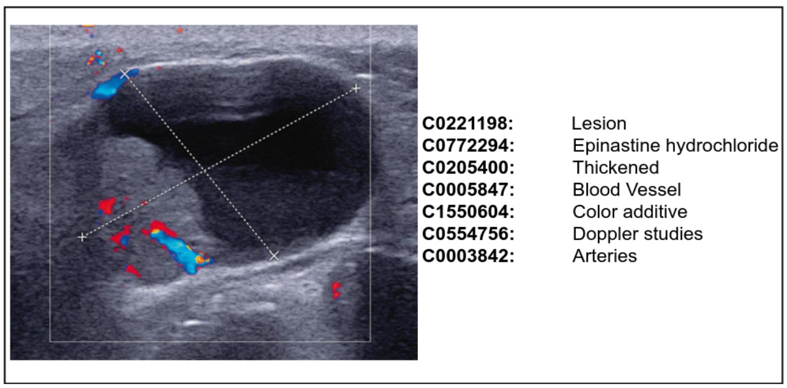

Fig. 1. Example of a radiology image with the corresponding extracted UMLS®CUIs.

The proposed task is the first step towards automatic image captioning and scene understanding, by identifying the presence and location of relevant biomedical concepts (CUIs) in a large corpus of medical images. Based on the visual image content, this task provides the building blocks for the scene understanding step by identifying the individual components of which captions are composed. The concepts can be used for context-based image analysis and for information retrieval. The detected concepts per image are evaluated with precision and recall scores from the ground truth, as described in Section 4.

In Table 2, the 11 participating teams of the ImageCLEFmed Concept Detection task are listed. There were 49 registered participants out of 99 teams, who downloaded the End-User-Agreement. Altogether, 77 runs were submitted for evaluation. Out of the 77 submitted runs, 60 were graded and 17 were faulty submission. The majority of the participating teams are new to the task, as only three groups participated in the previous years.

Table 1. Participating groups of the ImageCLEF 2019 Concept Detection Task. Teams with previous participation in 2018 are marked with an asterix.

Team	Institution	Runs
AUEB NLP Group [8]	Department of Informatics Athens University of Economics and Business	4
Damo [19]	Beihang University, Beijing, China	9
ImageSem* [3]	Institute of Medical Information Chinese Academy of Medical Sciences	10
UA.PT_Bioinformatics* [2]	Biomedical Informatics Research Group Universidade de Aveiro, Portugal	8
richard_ycli	The Hong Kong University of Science and Technology, Kowloon Hong Kong	5
Sam Maksoud [9]	The University of Queensland Brisbane, Australia	2
AI600 [18]	University of International Business and Economics, Beijing, China	7
MacUni-CSIRO [15]	Macquarie University, North Ryde Sydney, Australia	1
pri2si17 [16]	Mentor Graphics LibreHealth Uttar Pradesh, India	3
AILAB*	University of the Aegean Samos, Greece	5
LIST	Faculty of Sciences and Techniques Abdelmalek Essadi University, Morocco	6

3 Data Set

Equivalently to previous editions, the data set distributed for the ImageCLEFmed 2019 Concept Detection task originates from biomedical articles of the PMC Open Access subset.

The training and validation sets containing 56,629 and 14,157 images were subsets of the ROCO data set presented in Peltka et al. [11]. ROCO has two classes: Radiology and Out-Of-Class. The first contains 81,825 radiology images, which was used for the presented work. It includes several medical imaging modalities such as, Computed Tomography (CT), Ultrasound, X-Ray, Fluoroscopy, Positron Emission Tomography (PET), Mammography, Magnetic Resonance Imaging (MRI), Angiography and PET-CT, and can be seen in Figure 2.

From the PMC Open Access subset [14], a total of 6,031,814 image - caption pairs were extracted. Compound figures, which are images with more than one subfigure, were removed using deep learning as proposed in Koitka et al. [7]. The non-compound images were further split into radiology and non-radiology, as the objective was solely on radiology. Semantic knowledge of object interplay present in the images were extracted in the form of UMLS®Concepts using the QuickUMLS library [17]. The image captions from the biomedical articles served as basis for the extraction of the concepts. The text pre-processing steps applied are described in Peltka et al. [11]. Figure 2 displays example images from the training set, containing several radiology imaging modalities.

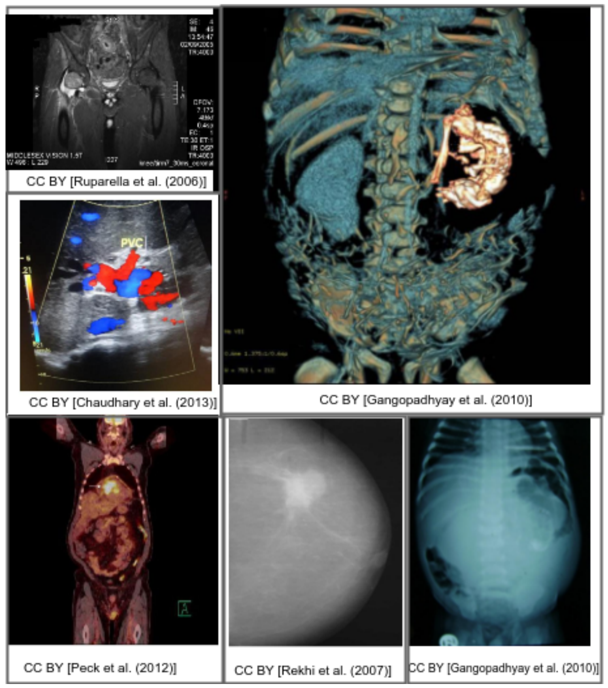

Fig. 2. Examples of a radiology images displaying the broad content of the ROCO data set.

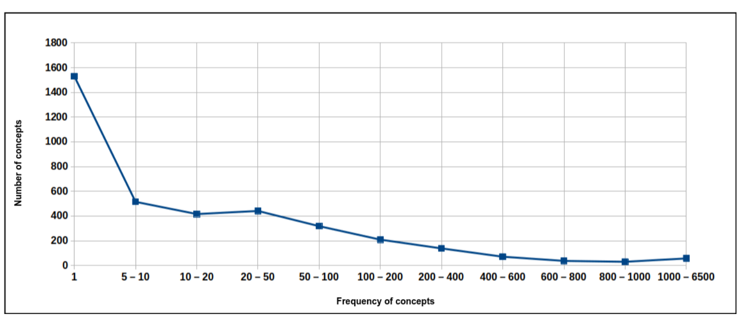

Fig. 3. The frequency versus the number of UMLS®(Unified Medical Language System ®) Concept Unique Identifiers (CUIs) in the development data. For example, 416 concepts occurred 10-20 times in the training images.

Examples of concepts in the training set are listed in descending order of occurence in Table 2. A few concepts were labelled only once, as can be seen in Figure 3.

ROCO contains images from the PMC archive extracted in January 2018, which makes up the training set for the ImageCLEF Concept Detection Task. To avoid an overlap with images distributed at previous ImageCLEF medical tasks, the test set for ImageCLEF 2019 was created with a subset of PMC Open Access (archiving date: 01.02.2018 - 01.02.2019). The same procedures applied for the creation of the ROCO data set were applied for the test set as well.

Concepts with very high frequency (>13,000), such as "Image", as well as redundant synonyms were removed. This lead to reduction of concepts per image in comparison to the previous years. All images in the training, validation and test sets have [1-72], [1-77] and [1-34] concepts, respectively.

Table 2. UMLS®(An excerpt of Unified Medical Language System ®) Concept Unique Identifiers (CUIs) distributed for Tte ImageCLEF Concept Detection Task with their respective number of occurrence. The concepts were randomly chosen in a descending order. All listed concepts were distributed in the training set.

CUI	Concept	Occurrence
C0441633	Scanning	6733
C0043299	Diagnostic radiologic examination	6321
C1962945	Radiographic imaging procedure	6318
C0040395	Tomography	6235
C0034579	Panoramic Radiography	6127
C0817096	Chest	5981
C0040405	X-Ray Computed Tomography	5801
C1548003	Diagnostic Service Section ID - Radiograph	5159
...	...	...
C0000726	Abdomen	2297
...	...	...
C2985765	Enhancement Description	1084
...	...	...
C0228391	Structure of habenulopeduncular tract	672
C0729233	Dissecting aneurysm of the thoracic aorta	652
...	...	...
C0771711	Pancreas extract	456
...	...	...
C1704302	Permanent premolar tooth	177
...	...	...
C0042070	Urography	67
C0085632	Apathy	67
C0267716	Incisional hernia	67
...	...	...
C0081923	Cardiocrome	1
C0193959	Tonsillectomy and adenoidectomy	1

4 Evaluation Methodology

UMLS®CUIs need to be automatically predicted by the participating teams for all 10,000 test images. As in previous editions [1, 4], the balanced precision and recall trade-off in terms of F1-scores was measured. The default implementation of the Python scikit-learn (v0.17.1-2) library was applied to compute the F-scores per image and average them across all test images.

As the training, validation and test set contain a maximum of 72, 77 and 34 concepts per image, the maximum number of concepts allowed in the submission runs was set to 100. Each participating group could submit altogether 10 valid and 7 faulty submission runs. Faulty submissions include:

- Same image id more than once
- Wrong image id
- Too many concepts
- Same concept more than once
- Not all test images included

All submission runs were uploaded by the participating teams and evaluated with CrowdAI[4]. The source code of the evaluation tool is available on the task Web page[5].

5 Results

This section details the results achieved by all 11 participating teams for the concept detection task. The best run per team is shown in Table 3. Table 4 contains the complete list of all graded submission runs. There is an improvement compared to both previous editions, from 0.1583 in ImageCLEF 2017 [1] and 0.1108 in ImageCLEF 2018 [4] to 0.2823 this year in terms of F1-score.

Best results were achieved by the AUEB NLP Group [8] by applying convolutional neural network (CNN) image encoders that were combined either with image retrieval methods or feed-forward neural networks to predict the concepts for images in the test set. On the test set, this CheXNet-based system [13] achieved better results in terms of F1-score, while an ensemble of an k-NN image retrieval system with CheXNet performed better on the development data. AUEB NLP ranked 1st to 3rd place with 3 out of the 4 submitted runs.

[4] https://www.crowdai.org/challenges/imageclef-2019-caption-concept-detection-6812fec9-8c9e-40ad-9fb9-cc1721c94cc1 [last accessed: 02.06.2019]

[5] https://www.imageclef.org/system/files/ImageCLEF-ConceptDetection-Evaluation.zip [last accessed: 02.06.2019]

Table 3. Performance of the participating teams at ImageCLEFmed 2019 Concept Detection Task. The best run per team is selected. Teams with previous participation in 2018 are marked with an asterix.

Team	Institution	F1-Score
AUEB NLP Group [8]	Department of Informatics Athens University of Economics and Business, Greece	0.2823094
Damo [19]	Beihang University, Beijing, China	0.2655099
ImageSem* [3]	Institute of Medical Information Chinese Academy of Medical Sciences, Beijing, China	0.2235690
UA.PT_Bioinformatics* [2]	Biomedical Informatics Research Group Universidade de Aveiro, Portugal	0.2058640
richard_ycli	The Hong Kong University of Science and Technology, Kowloon Hong Kong	0.1952310
Sam Maksoud [9]	The University of Queensland Brisbane, Australia	0.1749349
AI600 [18]	University of International Business and Economics, Beijing, China	0.1656261
MacUni-CSIRO [15]	Macquarie University, North Ryde Sydney, Australia	0.1435435
pri2si17 [16]	Mentor Graphics LibreHealth Uttar Pradesh, India	0.0496821
AILAB*	University of the Aegean Samos, Greece	0.0202243
LIST	Faculty of Sciences and Techniques Abdelmalek Essadi University, Morocco	0.0013269

Damo [19] was the second ranked group with 9 runs and applied two distinct methods to address the concept detection task. The latest deep learning system ResNet-101 was used for a multi-label classification approach, as well as a CNN-RNN model framework with attention mechanisms. Due to the imbalanced concept distribution, the group applied several data filtering methods. This proved to be positive, as the best run was a combination of multi-label classification with a filtered and reduced data set.

A two-stage concept detection approach was presented by the third ranked group: ImageSem [3]. This included a medical image pre-classification and a transfer learning-based multi-label classification model. For the pre-classification step based on body parts, the semantic types of all CUIs from UMLS®were extracted to cluster the images into four body part related categories, including "chest", "abdomen", "head and neck" and "skeletal muscle". Prior to training of a multi-label classifier that was fine-tuned from the ImageNet data set, high frequency concepts were selected. The best run by ImageSem ranked 8 out of all submissions.

The pri2si17 team [16] participated for the first time in the concept detection task. They addressed the task as a multi label classification problem and limited the concepts to the most frequent 25 labels.

Table 4: Concept detection performance in terms of F1-scores

Group Name	Submission Run	F1-Score
AUEB NLP Group	s2_results.csv	0.2823094
AUEB NLP Group	ensemble_avg.csv	0.2792511
AUEB NLP Group	s1_results.csv	0.2740204
Damo	test_cat_xi.txt	0.2655099
AUEB NLP Group	s3_results.csv	0.2639952
Damo	test_results.txt	0.2613895
Damo	first_concepts_detection_result_check.txt	0.2316484
ImageSem	F1TOP1.txt	0.2235690
ImageSem	F1TOP2.txt	0.2227917
ImageSem	F1TOP5_Pmax.txt	0.2216225
ImageSem	F1TOP3.txt	0.2190201
ImageSem	07Comb_F1Top1.txt	0.2187337
ImageSem	F1TOP5_Rmax.txt	0.2147437
Damo	test_tran_all.txt	0.2134523
Damo	test_cat.txt	0.2116252
UA.PT_Bioinformatics	simplenet.csv	0.2058640
richard_ycli	testing_result.txt	0.1952310
ImageSem	08Comb_Pmax.txt	0.1912173
UA.PT_Bioinformatics	simplenet128x128.csv	0.1893430
UA.PT_Bioinformatics	mix-1100-o0-2019-05-06_1311.csv	0.1825418
UA.PT_Bioinformatics	aae-1100-o0-2019-05-02_1509.csv	0.1760092
Sam Maksoud	TRIAL_1.txt	0.1749349
richard_ycli	testing_result.txt	0.1737527
UA.PT_Bioinformatics	ae-1100-o0-2019-05-02_1453.csv	0.1715210
UA.PT_Bioinformatics	cedd-1100-o0-2019-05-03_0937-trim.csv	0.1667884
AI600	ai600_result_weighing_1557061479.txt	0.1656261
Sam Maksoud	TRIAL_18.txt	0.1640647
richard_ycli	testing_result_run4.txt	0.1633958
AI600	ai600_result_weighing_1557059794.txt	0.1628424
richard_ycli	testing_result_run3.txt	0.1605645
AI600	ai600_result_weighing_1557107054.txt	0.1603341
AI600	ai600_result_weighing_1557062212.txt	0.1588862
AI600	ai600_result_weighing_1557062494.txt	0.1562828
AI600	ai600_result_weighing_1557107838.txt	0.1511505
richard_ycli	testing_result_run2.txt	0.1467212
MacUni-CSIRO	run1FinalOutput.txt	0.1435435
AI600	ai600_result_rgb_1556989393.txt	0.1345022
UA.PT_Bioinformatics	simplenet64x64.csv	0.1279909
UA.PT_Bioinformatics	resnet19-cnn.csv	0.1269521
ImageSem	09Comb_Rmax_new.txt	0.1121941
Damo	test_att_3_rl_best.txt	0.0590448
Damo	test_rl_5_result_check.txt	0.0584684
Damo	test_tran_rl_5.txt	0.0567311
Damo	test_tran_10.txt	0.0536554
pri2si17	submission_1.csv	0.0496821
AILAB	results_v3.txt	0.0202243
AILAB	results_v1.txt	0.0198960
AILAB	results_v2.txt	0.0162458
pri2si17	submission_3.csv	0.0141422
AILAB	results_v4.txt	0.0126845
LIST	denseNet_pred_all_0.55.txt	0.0013269
ImageSem	yu_1000_inception_v3_top6.csv	0.0009450
ImageSem	yu_1000_resnet_152_top6.csv	0.0008925
LIST	denseNet_pred_all_0.6.txt	0.0003665
LIST	denseNet_pred_all.txt	0.0003400
LIST	predictionBR(LR).txt	0.0002705
LIST	denseNet_pred_all_0.6_50_0.04(max if null).txt	0.0002514
LIST	predictionCC(LR).txt	0.0002494
AILAB	results_v0.txt	0
pri2si17	submission_2.csv	0

UA.PT_BioInformatics [2] was the overall fourth best team and ranked 16th with their best F1-score of 0.2058 out of all submissions. Two independent approaches were applied to address the concept detection task. Image representations obtained with several feature extraction methods, such as color edge directivity descriptors (CEDD) and adversarial auto-encoder, as well as an end-to-end approach using two deep learning architectures. The best score out of the 8 submitted runs was achieved with a simplenet configuration.

A recurrent neural network (RNN) architecture was proposed by Sam Maksoud [9]. Soft attention and visual gating mechanisms are used to enable the network to dynamically regulate "where" and "when" to extract visual data for concept generation. Two runs were submitted for grading, with the score of 0.1749 ranked 22nd out of all submissions and the group was ranked 6th overall.

The 7th overall ranked group is AI600 with 7 graded submission runs. Multi-label classification based on Bag-of-Visual-Words model with color descriptors and logistic regression, using different SIFT (Scale-Invariant Feature Transform) descriptors as visual features were applied for the concept detection task. The best run with a combination of SIFT, C-SIFT, HSV-SIFT and RGB-SIFT visual descriptors achieved 0.1656261, which is the 26th out of all submissions.

MS-CSIRO [15] submitted 1 run for official evaluation. Relevant concepts were predicted with an approach based on a multi-label classification model using CNN. MS-CSIRO ranked as the 8th best team and their submitted run with the score 0.1435 ranked 36th.

Similar to team Damo, the deep learning system ResNet-101 was utilized as base model. pri2si17 are the ninth best ranked team. Three runs were submitted for grading, of which the best run achieved the score 0.0497 ranking 45th.

6 Discussion and Conclusion

The results of the task in 2019 show that there is an improvement in the F1-scores in this 3rd edition (best score 0.2823) in comparison to ImageCLEF 2017 and ImageCLEF 2018. In the previous years, the best scores were 0.1583 in 2017 and 0.1108 in 2018. There were several new teams participating for the 1st time, as well as 3 teams, who participated in all editions. In addition, an increased number of participating teams and submitted runs was noticed in 2019. This shows the interest in this challenging task.

Most submitted runs are based on deep learning techniques. Several methods such as concept filtering, data augmentation and image normalization were applied to optimize the input for the predicting systems. Long short-term memory (LSTM) recurrent neural networks (RNN), adversarial auto-encoder, CNN image encoders and transfer learning-based multi-label classification models were the frequently used approaches.

The focus this year was reduced from biomedical images to solely radiology images, which led to the reduction of extracted concepts from 111,155 to 5,528. However, there is still an unbalanced distribution of concepts, which shows to be challenging to most teams. This can be due to the different imaging modalities, as well as several body parts included in the data set. Medical data and diseases are also usually unbalanced with a few conditions happening very frequently and most being very rare.

In future work, an extensive review of the clinical relevance for the concepts in the development data should be explored. As the concepts originate from the natural language captions, not all concepts have high clinical utility. Medical journals also have very different policies in terms of checking figure cations. We believe this will assist in creating more efficient systems for automated medical data analysis.

References

1. Eickhoff, C., Schwall, I., de Herrera, A.G.S., Müller, H.: Overview of imageclefcaption 2017 - image caption prediction and concept detection for biomedical images. In: Working Notes of CLEF 2017 - Conference and Labs of the Evaluation Forum, Dublin, Ireland, September 11-14, 2017. (2017), http://ceur-ws.org/Vol-1866/invited_paper_7.pdf
2. Gonalves, A.J., Pinho, E., Costa, C.: Informative and intriguing visual features: Ua.pt bioinformatics in imageclef caption 2019. In: CLEF2019 Working Notes. CEUR Workshop Proceedings, (CEUR-WS.org), ISSN 1613-0073, http://ceur-ws.org/Vol-2380/, Lugano, Switzerland (September 09-12 2019)
3. Guo, Z., Wang, X., Zhang, Y., Li, J.: Imagesem at imageclefmed caption 2019 task: a two-stage medical concept detection strategy. In: CLEF2019 Working Notes. CEUR Workshop Proceedings, (CEUR-WS.org), ISSN 1613-0073, http://ceur-ws.org/Vol-2380/, Lugano, Switzerland (September 09-12 2019)
4. de Herrera, A.G.S., Eickhoff, C., Andrearczyk, V., Müller, H.: Overview of the imageclef 2018 caption prediction tasks. 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/invited_paper_4.pdf
5. Ionescu, B., Müller, H., Péteri, R., Cid, Y.D., Liauchuk, V., Kovalev, V., Klimuk, D., Tarasau, A., Abacha, A.B., 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)
6. Kalpathy-Cramer, J., de Herrera, A.G.S., Demner-Fushman, D., Antani, S.K., Bedrick, S., Müller, H.: Evaluating performance of biomedical image retrieval systems - an overview of the medical image retrieval task at imageclef 2004-2013. Comp. Med. Imag. and Graph. **39**, 55–61 (2015). https://doi.org/10.1016/j.compmedimag.2014.03.004, https://doi.org/10.1016/j.compmedimag.2014.03.004
7. Koitka, S., Friedrich, C.M.: Optimized convolutional neural network ensembles for medical subfigure classification. In: Jones, G.J., Lawless, S., Gonzalo, J., Kelly, L., Goeuriot, L., Mandl, T., Cappellato, L., Ferro, N. (eds.) Experimental IR Meets Multilinguality, Multimodality, and Interaction at the 8th International Conference

of the CLEF Association, Dublin, Ireland, September 11-14, 2017, Lecture Notes in Computer Science (LNCS) 10456. pp. 57–68. Springer International Publishing, Cham (2017). https://doi.org/10.1007/978-3-319-65813-1_5

8. Kougia, V., Pavlopoulos, J., Androutsopoulo, I.: Aueb nlp group at imageclefmed caption 2019. In: CLEF2019 Working Notes. CEUR Workshop Proceedings, (CEUR-WS.org), ISSN 1613-0073, http://ceur-ws.org/Vol-2380/, Lugano, Switzerland (September 09-12 2019)
9. Maksoud, S., Wiliem, A., Lovell, B.: Recurrent attention networks for medical concept prediction. In: CLEF2019 Working Notes. CEUR Workshop Proceedings, (CEUR-WS.org), ISSN 1613-0073, http://ceur-ws.org/Vol-2380/, Lugano, Switzerland (September 09-12 2019)
10. Müller, H., Clough, P.D., Deselaers, T., Caputo, B. (eds.): ImageCLEF, Experimental Evaluation in Visual Information Retrieval. Springer (2010). https://doi.org/10.1007/978-3-642-15181-1
11. Pelka, O., Koitka, S., Rückert, J., Nensa, F., Friedrich, C.M.: Radiology objects in context (ROCO): A multimodal image dataset. In: Intravascular Imaging and Computer Assisted Stenting - and - Large-Scale Annotation of Biomedical Data and Expert Label Synthesis - 7th Joint International Workshop, CVII-STENT 2018 and Third International Workshop, LABELS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Proceedings. pp. 180–189 (2018). https://doi.org/10.1007/978-3-030-01364-6_20, https://doi.org/10.1007/978-3-030-01364-6_20
12. Pelka, O., Nensa, F., Friedrich, C.M.: Adopting semantic information of grayscale radiographs for image classification and retrieval. In: Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2018) - Volume 2: BIOIMAGING, Funchal, Madeira, Portugal, January 19-21, 2018. pp. 179–187 (2018). https://doi.org/10.5220/0006732301790187
13. Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C., Shpanskaya, K., Lungren, M.P., Ng, A.Y.: Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. CoRR **abs/1711.05225** (2017), http://arxiv.org/abs/1711.05225
14. Roberts, R.J.: PubMed Central: The GenBank of the published literature. Proceedings of the National Academy of Sciences of the United States of America **98**(2), 381–382 (Jan 2001). https://doi.org/10.1073/pnas.98.2.381
15. Singh, S., Karimi, S., Ho-Shon, K., Hamey, L.: Biomedical concept detection in medical images: Mq-csiro at 2019 imageclefmed caption task. In: CLEF2019 Working Notes. CEUR Workshop Proceedings, (CEUR-WS.org), ISSN 1613-0073, http://ceur-ws.org/Vol-2380/, Lugano, Switzerland (September 09-12 2019)
16. Sinha, P., Purkayastha, S., Gichoya, J.: Full training versus fine tuning for radiology images concept detection task for the imageclef 2019 challenge. In: CLEF2019 Working Notes. CEUR Workshop Proceedings, (CEUR-WS.org), ISSN 1613-0073, http://ceur-ws.org/Vol-2380/, Lugano, Switzerland (September 09-12 2019)
17. Soldaini, L., Goharian, N.: QuickUMLS: a fast, unsupervised approach for medical concept extraction. In: MedIR Workshop, SIGIR (2016)
18. Wang, X., Liu, N.: Ai600 lab at imageclef 2019 concept detection task. In: CLEF2019 Working Notes. CEUR Workshop Proceedings, (CEUR-WS.org), ISSN 1613-0073, http://ceur-ws.org/Vol-2380/, Lugano, Switzerland (September 09-12 2019)
19. Xu, J., Liu, W., Liu, C., Wang, Y., Chi, Y., Xie, X., Hua, X.: Concept detection based on multi-label classification and image captioning approach - damo

at imageclef 2019. In: CLEF2019 Working Notes. CEUR Workshop Proceedings, (CEUR-WS.org), ISSN 1613-0073, http://ceur-ws.org/Vol-2380/, Lugano, Switzerland (September 09-12 2019)