Overview of the ImageCLEF 2015 Medical Classification Task

Alba G. Seco de Herrera⋆, Henning Müller, and Stefano Bromuri

University of Applied Sciences Western Switzerland (HES–SO), Switzerland
`alba.garciasecodeherrera@nih.gov`

Abstract. This articles describes the ImageCLEF 2015 Medical Classification task. The task contains several subtasks that all use a data set of figures from the biomedical open access literature (PubMed Central). Particularly compound figures are targeted that are frequent in the literature. For more detailed information analysis and retrieval it is important to extract targeted information from the compound figures. The proposed tasks include compound figure detection (separating compound from other figures), multi–label classification (define all sub types present), figure separation (find boundaries of the subfigures) and modality classification (detecting the figure type of each subfigure). The tasks are described with the participation of international research groups in the tasks. The results of the participants are then described and analysed to identify promising techniques.

Keywords: ImageCLEFmed, compound figure detection, multi–label classification, figure separation, modality classification

1 Introduction

The amount and availability of biomedical literature has increased considerably due to the advent of the Internet [1]. The task of medical doctors has on the other hand not become simpler as the amount of information to review for taking decisions has become overwhelming. Despite this growing complexity, physicians would use services that improve their understanding of an illness even if these involve more cognitive effort than in standard practice [2]. Images in biomedical articles can contain highly relevant information for a specific information need and can accelerate the search by filtering out irrelevant documents [3]. As a consequence image–based retrieval has been proposed as a way of improving access to the medical literature and complement text search [4, 5].

Image classification can play an important role in improving the image–based retrieval of the biomedical literature, as this helps to filter out irrelevant information from the retrieval process. Many images in the biomedical literature (around 40% [6]) are compound figures (see Figure 1), so determining the figure type is not clear as several different types of figures can be present in a single compound figure.

⋆ Alba G. Seco de Herrera is currently working at the National Library of Medicine (NLM/NIH), Bethesda, MD, USA

Fig. 1. Examples of compound figures in the biomedical literature.

Information retrieval systems for images should be capable of distinguishing the parts of compound figures that are relevant to a given query, as usually queries are limited to a single modality [7]. Compound figure detection and multi–label classification are therefore a required first step to focus retrieval of images. Some file formats, such as DICOM (Digital Imaging and Communications in Medicine), contain metadata that can be used to filter images by modality, but this information is lost when using images from the biomedical literature where images are stored as JPG, GIF or PNG files. In this case caption text and visual appearance are key to understanding the content of the image and whether or not it is a compound figure. Both types of information, text and visual, are complementary to each other and can help managing the multi–label classification [8]. From the standpoint of information retrieval and classification of compound images and associated text, the current systems could greatly benefit from the use of multi–label classification approaches [9] by a) defining models that can use the dependencies between the extracted images; b) defining models that can express the importance of a label in a compound figure. In addition, compound figures are naturally redundant sources of information with natural dependencies occurring between the different regions of the image.

Retrieval systems can fail if they are not specifically designed to work with compound figures and partial relevance. Identification of each subpart of the figures can improve retrieval accuracy by enabling comparison of figures with lower noise levels [10].

To promote research on this field a medical classification task is proposed in the context of the ImageCLEF 2015 lab [11]. This paper describes this benchmark in detail.

This article is structured as follows: Section 2 presents an overview of the participants and of the datasets used in the competition. Section 3 discusses the results with respect to the selected datasets. Finally, Section 4 concludes the paper and presents relevant future work for the next edition of ImageCLEF.

2   Tasks, Data Sets, Ground Truth, Participation

This section describes the main scenario of the benchmark including the data used, the tasks, ground truthing and participation.

2.1   The Tasks in 2015

There were four subtasks in 2015:

– compound figure detection;
– compound figure separation;
– multi–label classification;
– subfigure classification.

This section gives an overview of each of the four subtasks.

Compound Figure Detection   Compound figure identification is a required first step to make compound figures from the literature accessible for further analysis. Therefore, the goal of this subtask is to identify whether a figure is a compound figure or not. The task makes training data available containing compound and non–compound figures from the biomedical literature. Figure 2 shows an example of a compound and a non–compound figure.
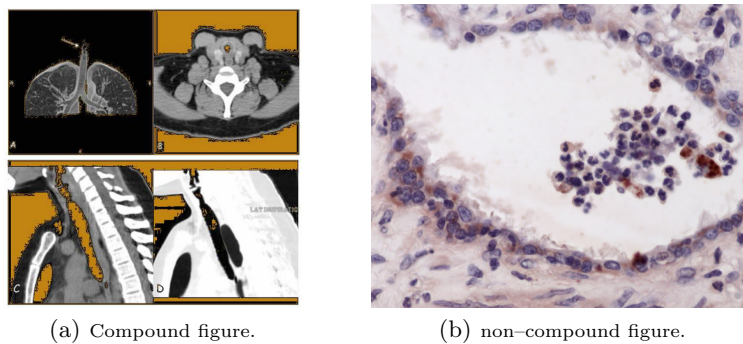

(a) Compound figure.   (b) non–compound figure.

Fig. 2. Examples of compound and non–compound figures.

Figure Separation   This task was first introduced in 2013 and the same evaluation methodology is used in 2015 [6]. The goal of this task is to separate the compound figures into subfigures using separation lines. Figure 3 shows a compound figure which is separated into subfigures by blue lines. In 2015, a larger number of compound figures was distributed compared to the previous years.
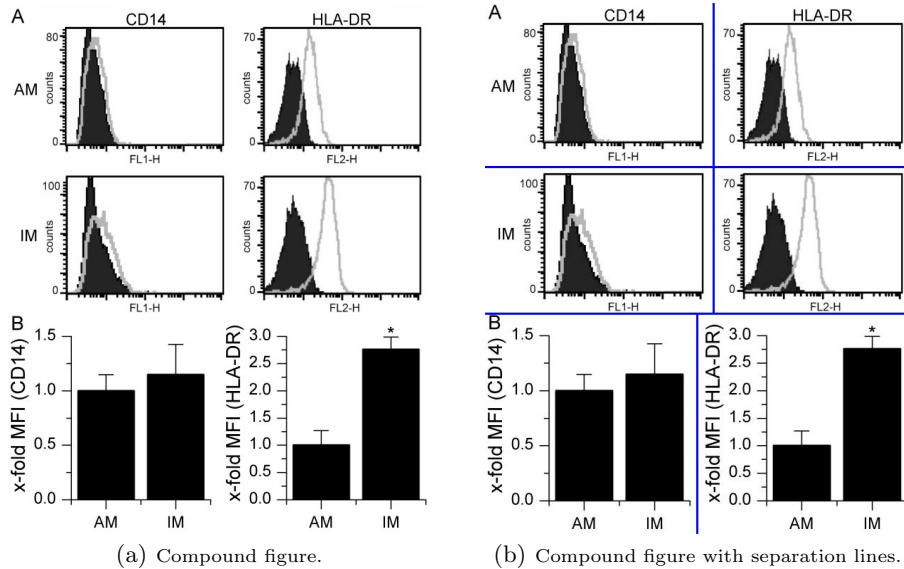
(a) Compound figure. (b) Compound figure with separation lines.

Fig. 3. Example a compound figure and its separation into subfigures by blue lines.

Multi–label Classification The fundamental difference with respect to compound figure separation resides in the fact that the compound figure is not separated into subfigures, but it is rather used entirely to perform a scene classification task. The intuition behind this approach resides in the fact that subfigures in medical papers are usually assembled because they add complementary information concerning the aeticle topic (see Figure 4). In this sense, much work was performed in the multi–label classification community [12] and many algorithms already exist to classify multi–label problems. A multi–label dataset in medical imaging was never considered before to the best of our knowledge. More formally, this problem can be expressed as follows:

Let X be the domain of observations and let L be the finite set of labels. Given a training set $T = \{(x_1, Y_1), (x_2, Y_2), ..., (x_n, Y_n)\}$ $(x_i \in X, Y_i \subseteq L)$ i.i.d. drawn from an unknown distribution D, the goal is to learn a multi–label classifier $h : X \to 2^L$. However, it is often more convenient to learn a real–valued scoring function of the form $f : X \times L \to \mathbb{R}$. Given an instance x_i and its associated label set Y_i, a working system will attempt to produce larger values for labels in Y_i than those that are not in Y_i, i.e. $f(x_i, y_1) > f(x_i, y_2)$ for any $y_1 \in Y_i$ and $y_2 \notin Y_i$. By the use of the function $f(\cdot, \cdot)$, a multi–label classifier can be obtained: $h(x_i) = \{y | f(x_i, y) > \delta, y \in L\}$, where δ is a threshold to infer from the training set. The function $f(\cdot, \cdot)$ can also be adapted to a ranking function $rank_f(\cdot, \cdot)$, which maps the outputs of $f(x_i, y)$ for any $y \in L$ to $\{1, 2, ..., |L|\}$ such that if $f(x_i, y_1) > f(x_i, y_2)$ then $rank_f(x_i, y_1) < rank_f(x_i, y_2)$.
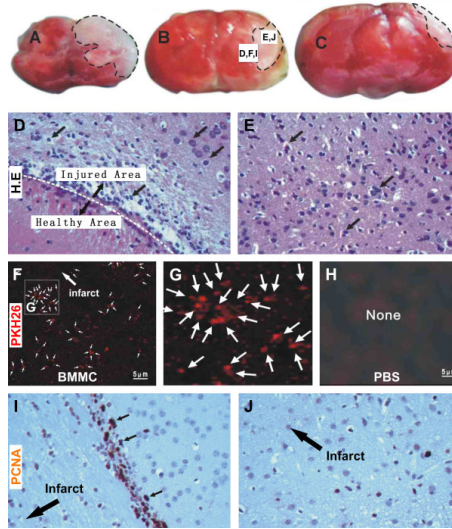
Fig. 4. Example a compound figure containing images of multiple classes which are all related to each other with respect to the localization of transplanted cells and in situ proliferation at the infarct border in an experimental model.

Multi–label performance measures differ from single label ones. Following the same approach presented in [12], the Hamming Loss is proposed as the evaluation measure for multi–label learning in ImageCLEF.

More formally let a testing set $S = \{(x_1, Y_1), (x_2, Y_2), ..., (x_m, Y_m)\}$.

Hamming loss: evaluates how many times an observation–label pair is misclassified. The score is normalized between 0 and 1, where 0 is the best:

$$hloss_S(h) = \frac{1}{m} \sum_{i=1}^{m} \frac{|h(x_i) \triangle Y_i|}{|L|}. \tag{1}$$

where $\triangle$ represents the symmetric difference.

Subfigure Classification The subfigure classification task is a variation of the multi–label classification task in which the subfigures contained in the multi–label figures are provided separately for classification. The main reason to proceed in this way is to provide two matched dataset that researchers can use to compare multi–label classification of the full compound image versus taking each single image in the compound image and classify it.

2.2 Datasets

In 2015, the dataset was a subset of the of the full ImageCLEF 2013 dataset [6], which is a part of PubMed Central[1] containing in total over 1,700,000 images

---

[1] `http://www.ncbi.nlm.nih.gov/pmc/`

in 2014. The distributed subset contains a total of 20,867 figures. The training set contains 10,433 figures and the test set 10,434 figures. Each of these two sets contains 6,144 compound figures and 4289–4290 non–compound figures. The entire dataset is used for the compound figure detection task.

6,784 of the compound figures are used for the figure separation task. 3,403 figures are distributed in the training set and 3,381 in the test set.

A subset of these images containing 1,568 images are labelled for the multi–label learning task. These images are also distributed as a training set (containing 1,071 figures) and a test set (containing 497 figures). The labels were assigned using the same class hierarchy as the one used for the ImageCLEF 2012 [13] and 2013 [6] modality classification task. A slight difference is that in 2015 the class "compound" is not included because only the non–compound parts can be labels with all compound images being split.

Figure 5 shows the ImageCLEF 2015 class hierarchy where the class codes with descriptions are the following ([Class code] Description):

- [$Dxxx$] Diagnostic images:
  - [$DRxx$] Radiology (7 categories):
  - [$DRUS$] Ultrasound
  - [$DRMR$] Magnetic Resonance
  - [$DRCT$] Computerized Tomography
  - [$DRXR$] X–Ray, 2D Radiography
  - [$DRAN$] Angiography
  - [$DRPE$] PET
  - [$DRCO$] Combined modalities in one image
- [$DVxx$] Visible light photography (3 categories):
  - [$DVDM$] Dermatology, skin
  - [$DVEN$] Endoscopy
  - [$DVOR$] Other organs
- [$DSxx$] Printed signals, waves (3 categories):
  - [$DSEE$] Electroencephalography
  - [$DSEC$] Electrocardiography
  - [$DSEM$] Electromyography
- [$DMxx$] Microscopy (4 categories):
  - [$DMLI$] Light microscopy
  - [$DMEL$] Electron microscopy
  - [$DMTR$] Transmission microscopy
  - [$DMFL$] Fluorescence microscopy
- [$D3DR$] 3D reconstructions (1 category)
- [$Gxxx$] Generic biomedical illustrations (12 categories):
  - [$GTAB$] Tables and forms
  - [$GPLI$] Program listing
  - [$GFIG$] Statistical figures, graphs, charts
  - [$GSCR$] Screenshots
  - [$GFLO$] Flowcharts
  - [$GSYS$] System overviews

- [$GGEN$] Gene sequence
- [$GGEL$] Chromatography, Gel
- [$GCHE$] Chemical structure
- [$GMAT$] Mathematics, formula
- [$GNCP$] Non–clinical photos
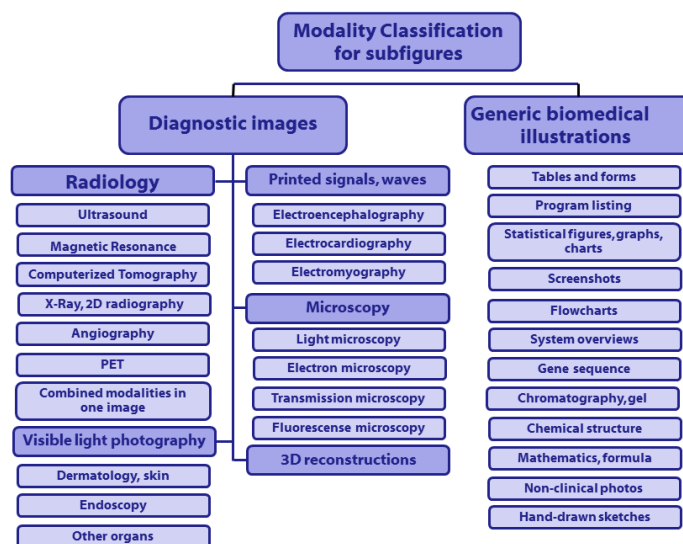- [$GHDR$] Hand–drawn sketches

Fig. 5. The image class hierarchy that was developed for document images occurring in the biomedical open access literature.

Finally, each figure from the multi–label classification task is separated into subfigures and each of the subfigures is labelled. As a result, 4,532 subfigures were released in the training set and 2,244 in the test set. To link the multi–label classification and the subfigure separation tasks, the figure IDs were related. If the figure ID is "1297-9686-42-10-3", then the corresponding subfigure IDs are "1297-9686-42-10-3-**1**", "1297-9686-42-10-3-**2**", "1297-9686-42-10-3-**3**" and "1297-9686-42-10-3-**4**".

In addition to the figures, the articles of the figures are provided to allow for the use of textual information.

2.3 Participation

Over seventy groups registered for the medical classification tasks and obtained access to the data sets. Eight of the registered groups submitted results to the medical classification tasks.

7 runs were submitted to the compound figure detection task, 12 runs to the multi–label classification task, 5 runs to the figure separation task and 16 runs to the subfigure separation task.

The following groups submitted at least one run:

– AAUITEC (Institute of Information Technology, Alpen–Adria University of Klagenfurt, Austria);
– FHDO BCSG (FHDO Biomedical Computer Science Group, University of Applied Science and Arts, Germany);
– BMET (Institute of Biomedical Engineering and Technology, University of Sydney, Australia)
– CIS UDEL (Computer & Information Sciences, University of Delaware Newark, USA)
– CMTECH (Cognitive Media Technologies Research Group, Pompeu Fabra University, Spain)
– IIS (Institute of Computer Science, University of Innsbruck, Austria)
– MindLab (Machine Learning, Perception and Discovery Lab, National University of Colombia, Colombia);
– NLM (National Library of Medicine, USA).

3 Results

This section describes the results obtained by the participants for each of the subtasks.

3.1 Compound Figure Detection

Very good results were obtained for the compound figure detection task, reaching up to 85% for HDO BCSG as seen in Table 1. Table 1 contains the results obtained by the two participants of the compound figure detection task.

Table 1. Results of the runs of the compound figure detection task.

Group	Run	Run Type	Accuracy
FHDO BCSG	task1_run2_mixed_sparse1	mixed	85.39
FHDO BCSG	task1_run1_mixed_stemDict	mixed	83.88
FHDO BCSG	task1_run3_mixed_sparse2	mixed	80.07
FHDO BCSG	task1_run4_mixed_bestComb	mixed	78.32
FHDO BCSG	task1_run6_textual_sparseDict	textual	78.34
CIS UDEL	exp1	visual	82.82
FHDO BCSG	task1_run5_visual_sparseSift	visual	72.51

FHDO BCSG [14] achieved best results with an accuracy of 85.39% using a multi–modal approach. FHDO BCSG applied a combination of visual features and text. With respect to visual features they focused on features detecting the border of the figures and a bag–of–keypoints. The bag–of–words approach is used

for text classification using the provided figure caption. They also proposed two runs applying only either visual or text information obtaining in general lower results that applying multi–modal approaches.

CIS UDEL [15] obtained best results when using only visual information achieving an accuracy of 82.82%. A combination of connected component analysis of subfigures and peak region detection is used.

3.2 Figure Separation

In 2015, two groups participated in the figure separation task. Table 2 shows the results achieved. Best results were obtained by NLM [16]. NLM distinguished two

Table 2. Results of the runs of the figure separation task.

Group	Run	Accuracy	
NLM	run2_whole	visual	84.64
NLM	run1_whole	visual	79.85
AAUITEC	aauitec_figsep_combined	visual	49.40
AAUITEC	aauitec_figsep_edge	visual	35.48
AAUITEC	aauitec_figsep_band	visual	30.22

types of compound images: stitched multipanel figure and multipanel figures with a gap. A manual selection of stitched multipanel figures from the whole dataset is first carried out. Then, two approaches are used. Best results are obtained by "run2_whole" where stitched multipanel figure separation is combined with both image panel separation and label extraction. "run2_whole" achieved an accuracy of 79.85% by combining stitched multipanel figure separation with panel separation.

AAUITEC [17] submitted three runs. Each run used a specific separator line detection based on "bands" (run "aauitec_figsep_band") or "edges" (run "aauitec_figsep_edge"). Best results achieved an accuracy of 49.40% when using a combination of both detection types. A recursive algorithm is used starting by classifying the images as illustrations or not. Depending on the type of image a specific separator line detection is used, based on "bands" or "edges", respectively.

3.3 Multi–label Classification

With respect to the multi–label classification task, there were two participating groups, IIS [18] and Mindlab[2]. Quite interestingly, none of the two participants decided to apply standard multi–label classification algorithms [12] such ask Multi–Label K–nearest neighbours (MLKNN) or Binary Relevance Support

---

[2] https://sites.google.com/a/unal.edu.co/mindlab/

Vector Machines (BR–SVM), but rather decided to come up with two new solutions to the problem. Table 3 presents the results of the runs submitted by the two groups.

ISS applied Khronker decomposition to find a set of filters of the figure as features for a maximum margin layer classifier in which the multi–label task is mapped on a dual problem of the standard margin optimization with SVMs. To achieve this the authors consider the possibility of modelling the problem by introducing an additional kernel matrix calculated starting from the vector of labels associated to the compound figures.

The Mindlab approach is based on building a visual representation by means of deep convolutional neural networks, by relying on the theory of transfer learning which is based in the ability of a system to recognize and apply knowledge learned in previous domains to novel domains, which share some commonality. For this task, Mindlab used the Yangqing Jia et al. (Caffe) [19] pretrained network to represent figures. Caffe is an open source implementation of the winning convolutional network architecture of the ImageNet challenge. For the label assignment the authors proceeded as follows: Once the prediction is made a distribution of the classes is obtained and used to annotate only those with a score above 0.5. Furthermore, in the second run when a sample concept that scores above 0.5 does not exist, the two top labels are assumed as relevant.

The scores of the two presented approaches are quite close in the result, but the best result was achieved by Mindlab with a Hamming Loss of 0.5 as seen in Table 3.

Table 3. Results of the runs of the multi–label classification task.

Run	Group	Hamming Loss
IIS	output_6	0.0817
IIS	output_8	0.0785
IIS	output_9	0.0710
IIS	output_7	0.0700
IIS	output_10	0.0696
IIS	output_5	0.0680
IIS	output_1	0.0678
IIS	output_3	0.0675
MindLAB	predictions_Mindlab_ImageclefMed_multi–label_test_comb2lbl	0.0674
IIS	output_4	0.0674
IIS	output_2	0.0671
MindLAB	predictions_Mindlab_ImageclefMed_multi–label_test_comb1lbl	0.0500

3.4 Subfigure Classification

Three groups participated in the subfigure classification task and the results can be seen in Table 4. The FHDO BCSG group achieved the best classification accuracy (67.60%) by using textual and visual features as described in [14]. FHDO BCSG also achieved the best result when only using visual features (60.91%). The method reuses existing techniques and fuses visual and textual features.

Given the high dimensionality of the data, principal component analysis (PCA) is used to reduce the dimensionality to a subset of components explaining the variance of the dataset. SVMs and Random Forests are used together for the classification.

The CMTECH group used a descriptor based on covariance of the visual features associated with the subfigures [20]. The advantage of the proposed descriptor, being based on covariance, is to be robust with respect to noise. In this sense, the feature vector provided has 11 features. As claimed by the authors, this is particularly interesting because the matrix defined with this approach lies within the Riemann manifold, where images providing similar features are close in the Riemann Geometry. From this standpoint the authors also specified the conditions under which two images can be considered close.

The paper proposes a new approach to classifying images and the approach performs well without using a complicated classifier, which demonstrates the need to define good features.

Table 4. Results of the runs of the subfigure classification task.

Run	Group	Run Type	Accuracy
FHDO BCSG	task4_run5_train_20152013.txt	mixed	67.60
FHDO BCSG	task4_run4_clean_rf.txt	mixed	67.24
FHDO BCSG	task4_run1_combination.txt	mixed	66.48
FHDO BCSG	task4_run8_clean_short_rf.txt	mixed	66.44
FHDO BCSG	task4_run7_clean_comb_librf.txt	mixed	65.99
FHDO BCSG	task4_run6_clean_libnorm.txt	mixed	64.34
FHDO BCSG	task4_run3_textual.txt	textual	60.91
FHDO BCSG	task4_run2_visual.txt	visual	60.91
CMTECH	resultsSubfigureRunWholeCov.txt	visual	52.98
CMTECH	resultsSubfigure.txt	visual	48.61
BMET	sf_run_3.txt	visual	45.63
BMET	sf_run_6.txt	visual	45.00
BMET	sf_run_4.txt	visual	44.34
BMET	sf_run_2.txt	visual	43.62
BMET	sf_run_1.txt	visual	37.56
BMET	sf_run_5.txt	visual	37.56

The last approach was proposed by the BMET group [21]. The article present a convolutional neural networks (CNN) used for the subfigure classification. As specified by the authors, the CNN selected is a simplified version of the CNN used in LeNet–5. The major claim of the paper concerns the ability of the network to extract features in an unsupervised way and without the aid of domain knowledge. The main drawback of the paper is that the method used for the evaluation takes a long time to converge and the results were calculated on only partly optimized models. The classification results reflect this fact. Despite the relatively low results, the BMET contribution presents an interesting experimentation concerning the task proposed in ImageCLEF and it would certainly be interesting to see how these methods can be extended to improve the preliminary results obtained.

4   Conclusions

In this paper the results of the medical classification task ImageCLEF 2015 competition are presented. As in 2013, the challenge involved a subtask on the separation of compound figures from the biomedical literature. This year, three new subtasks were introduced: a subtask on detection of compound figures to identify whether a figure is compound or not; a multi–label subtask in which the subfigure labels of a compound figure need to be determined without separating the subfigures; and a subfigure classification challenge in which the separated subfigures are classified, following a traditional modality classification approach.

In the first year of this task, eight groups participated submitting forty runs. The participants present a variety of techniques for the problems on compound figure analysis. This is a wide problem than can make a large number of subfigures available for image search applications. The large number of different techniques leading to good results shows that many techniques can be used for the problem and the detailed optimization is most often responsible for obtaining a very good performance.

Acknowledgements

References

1. Lu, Z.: PubMed and beyond: a survey of web tools for searching biomedical literature. Database (Oxford) **2011** (2011)
2. Kannampallil, T.G., Franklin, A., Mishra, R., Almoosa, Khalid F.and Cohen, T., Patel, V.L.: Understanding the nature of information seeking behavior in critical care: Implications for the design of health information technology. Artificial Intelligence in Medicine **57**(1) (2013) 21–29
3. Müller, H., Foncubierta-Rodríguez, A., Lin, C., Eggel, I.: Determining the importance of figures in journal articles to find representative images. In: SPIE Medical Imaging. Volume 8674. (2013) 86740I–86740I–9
4. Müller, H., Michoux, N., Bandon, D., Geissbuhler, A.: A review of content–based image retrieval systems in medicine–clinical benefits and future directions. International Journal of Medical Informatics **73**(1) (2004) 1–23
5. Akgül, C., Rubin, D., Napel, S., Beaulieu, C., Greenspan, H., Acar, B.: Content–based image retrieval in radiology: Current status and future directions. Journal of Digital Imaging **24**(2) (2011) 208–222
6. García Seco de Herrera, A., Kalpathy-Cramer, J., Demner Fushman, D., Antani, S., Müller, H.: Overview of the ImageCLEF 2013 medical tasks. In: Working Notes of CLEF 2013 (Cross Language Evaluation Forum). (September 2013)
7. Markonis, D., Holzer, M., Dungs, S., Vargas, A., Langs, G., Kriewel, S., Müller, H.: A survey on visual information search behavior and requirements of radiologists. Methods of Information in Medicine **51**(6) (2012) 539–548

8. Clinchant, S., Ah-Pine, J., Csurka, G.: Semantic combination of textual and visual information in multimedia retrieval. In: Proceedings of the 1st ACM International Conference on Multimedia Retrieval. ICMR '11, ACM (2011) 44:1–44:8

9. Zhang, M.L., Zhou, Z.H.: A review on multi–label learning algorithms. IEEE Transactions on Knowledge and Data Engineering **99** (2013)

10. Müller, H., Kalpathy-Cramer, J., Demner-Fushman, D., Antani, S.: Creating a classification of image types in the medical literature for visual categorization. In: SPIE Medical Imaging. (2012)

11. Villegas, M., Müller, H., Gilbert, A., Piras, L., Wang, J., Mikolajczyk, K., García Seco de Herrera, A., Bromuri, S., Amin, M.A., Kazi Mohammed, M., Acar, B., Uskudarli, S., Marvasti, N.B., Aldana, J.F., Roldán García, M.d.M.: General overview of ImageCLEF at the CLEF 2015 labs. In: Working Notes of CLEF 2015. Lecture Notes in Computer Science. Springer International Publishing (2015)

12. Zhang, M.L., Zhou, Z.H.: A review on multi–label learning algorithms. IEEE Transactions on Knowledge and Data Engineering **26**(8) (2014) 1819–1837

13. Müller, H., García Seco de Herrera, A., Kalpathy-Cramer, J., Demner Fushman, D., Antani, S., Eggel, I.: Overview of the ImageCLEF 2012 medical image retrieval and classification tasks. In: Working Notes of CLEF 2012 (Cross Language Evaluation Forum). (September 2012)

14. Pelka, O., Friederich, C.M.: FHDO Biomedical Computer Science Group at Medical Classification Task of ImageCLEF 2015. In: Working Notes of CLEF 2015. (2015)

15. Wang, X., Shatkay, H., Kambhamettu, C.: CIS UDEL working notes on ImageCLEF 2015: Compound figure detection task. In: Working Notes of CLEF 2015. (2015)

16. Santosh, K., Xue, Z., Antani, S., Thoma, G.: NLM at ImageCLEF2015: Biomedical multipanel figure separation. In: Working Notes of CLEF 2015. (2015)

17. Taschwer, M., Marques, O.: AAUITEC at ImageCLEF 2015: Cmpound figure separation. In: Working Notes of CLEF 2015. (2015)

18. Rodríguez-Sánchez, A., Fontanella, S., Piater, J., Szedmak, S.: IIS at ImageCLEF 2015: Multi–label classification task. In: Working Notes of CLEF 2015. (2015)

19. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R.B., Guadarrama, S., Darrell, T.: Caffe: Convolutional architecture for fast feature embedding. In: Proceedings of the ACM International Conference on Multimedia, MM'14. (2014) 675–678

20. Cirujeda, P., Binefa, X.: Medical image classification via 2D color feature based covariance descriptors. In: Working Notes of CLEF 2015. (2015)

21. Lyndon, D., Kumar, A., Kim, J., Leong, P.H.W., Feng, D.: Convolutional neural networks for subfigure classification. In: Working Notes of CLEF 2015. (2015)