Overview of the ImageCLEF 2013 medical tasks

Alba G. Seco de Herrera [1], Jayashree Kalpathy–Cramer[2], Dina Demner Fushman[3], Sameer Antani[3], Henning Müller[1,4]

[1]University of Applied Sciences Western Switzerland, Sierre, Switzerland
[2]Harvard University, Cambridge, MA, USA
[3]National Library of Medicine (NLM), USA
[4]Medical Informatics, Univ. Hospitals and University of Geneva, Switzerland
alba.garcia@hevs.ch

Abstract. In 2013, the tenth edition of the medical task of the ImageCLEF benchmark was organized. For the first time, the ImageCLEFmed workshop takes place in the United States of America at the annual AMIA (American Medical Informatics Association) meeting even though the task was organized as in previous years in connection with the other ImageCLEF tasks. Like 2012, a subset of the open access collection of PubMed Central was distributed. This year, there were four subtasks: modality classification, compound figure separation, image–based and case–based retrieval. The compound figure separation task was included due to the large number of multipanel images available in the literature and the importance to separate them for targeted retrieval. More compound figures were also included in the modality classification task to make it correspond to the distribution in the full database. The retrieval tasks remained in the same format as in previous years but a larger number of tasks were available for image–based and case–based tasks. This paper presents an analysis of the techniques applied by the ten groups participating 2013 in ImageCLEFmed.

Keywords: ImageCLEFmed, modality classification, compound figure separation, image–based retrieval, case–based retrieval

1 Introduction

ImageCLEF[1] [1] is the image retrieval track of the Cross Language Evaluation Forum (CLEF). ImageCLEFmed is part of ImageCLEF focusing on medical images [2–7].

In the 10th edition of the medical task, the workshop is for the first time organized outside of Europe at the annual AMIA[2] (American Medical Informatics Association) meeting. The same format as in 2012 was followed and a new task was added, the compound figure separation. Characterisation of compound figures is often difficult, as they can contain features of various image types. Focusing search on the sub figures can lead to better results. In 2013, the modality

[1] http://www.imageclef.org/
[2] http://www.amia.org/amia2013/

classification task also included a larger amount of compound figures to make the task more realistic and correspond to the distribution in the database. The four tasks of 2013 are:

- modality classification;
- compound figure separation;
- image–based retrieval;
- case–based retrieval.

The paper is organized as follows. Section 2 describes the ImageCLEFmed tasks in more detail as well as the participation in each of the tasks. Section 3 presents the main results of the tasks and compares results within the participating groups and the techniques employed. Section 4 concludes the paper.

2 Participation, Data Sets, Tasks, Ground Truth

This section describes the four tasks organized in ImageCLEFmed 2013. The datasets and the ground truth provided for the evaluation campaign are explained in detail.

2.1 Participation

Like 2012, over sixty groups registered for the medical tasks and obtained access to the data sets. Ten of the registered groups submitted results to the medical tasks compared to 17 in 2012 with a total of 166 valid runs submitted, slightly fewer runs than in 2012. The smaller number of participants and submitted runs can be due to a change in the evaluation schedule of CLEF 2013 and may also be due to the fact that the event will be organized outside of Europe.

51 runs were submitted to the modality classification task, 4 runs to the compound figure separation task, 9 runs to the image retrieval task and 45 runs to the case-based retrieval task. As in previous years, the number of runs per group was limited to ten per subtask. The following groups submitted at least one run:

- AAUITEC (Institute of Information Technology, Alpen–Adria University of Klagenfurt, Austria)*;
- CITI (Center of Informatics and Information Technology, Portugal)*;
- DEMIR (Dokuz Eylul University, Turkey);
- FCSE (Faculty of Computer Sciences and Engineering, University of Ss Cyril and Methodius, Macedonia);
- IBM Multimedia Analytics (United States);
- IPL (Athens University of Economics and Business, Greece);
- ITI (Image and Text Integration Project, NLM, United States);
- medGIFT (University of Applied Sciences Western Switzerland, Switzerland);
- MiiLab (Medical Image Information Laboratory, Shanghai Advanced Research Institute,China)*;

– SNUMedInfo (Medical Informatics Laboratory, Seul National University, Republic of Korea)*;

Participants marked with a star had not participated in the medical task in 2012.

2.2 Datasets

In ImageCLEFmed 2013, the same database as in 2012 was supplied to the participants. The database contains over 300,000 images of 75,000 articles of the biomedical open access literature that allow free redistribution of the data. The ImageCLEFmed database is a subset of PubMed Central[3] containing in total over 1.5 million images of over 600,000 articles.

2.3 Modality Classification

The modality classification task was first introduced in 2010. The goal of this task is to classify the images into medical modalities and other images types, such as Computed Tomography, xray or general graphs. A modality hierarchy of 38 classes of which 31 appear in the data was used [8]. Using the modality information, the retrieval results could often be improved in the past by filtering our non–relevant image types [9]. The same hierarchy as in ImageCLEFmed 2012 was used (see Figure 1). In 2013 a larger number of compound figures than in ImageCLEFmed 2012 were provided in the training and test data sets. The current distribution corresponds to that in the PubMed Central data set, much closer to reality than in previous years.

The class codes with descriptions are the following ([Class code] Description):

- $[COMP]$ Compound or multipane images (1 category)
- $[Dxxx]$ 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

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

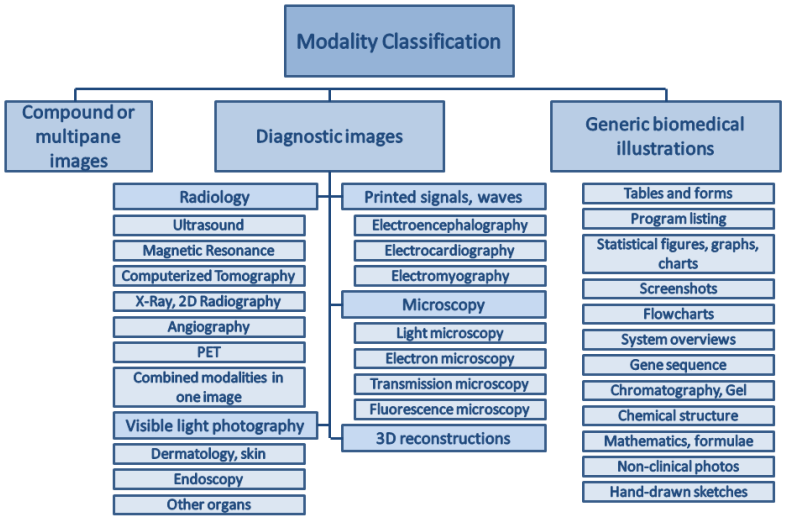
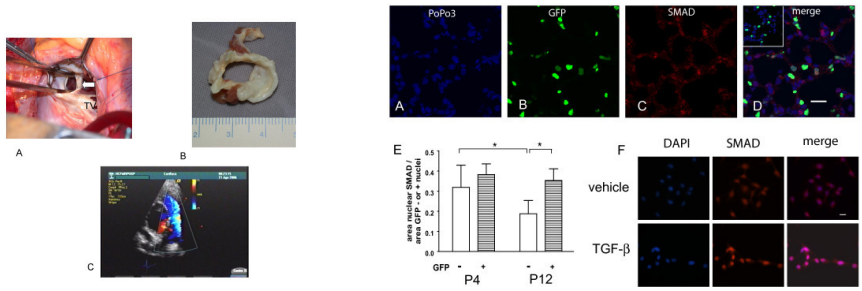


Fig. 1. The image class hierarchy that was development for document images occurring in the biomedical open access literature

- $[DMxx]$ 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, formulae
  - $[GNCP]$ Non–clinical photos
  - $[GHDR]$ Hand–drawn sketches

2.4 Compound Figure Separation

In the ImageCLEFmed 2012 data set [7] between 40% and 60% of the figures are compound or multipanel figures. Making the content of the compound figures accessible for targeted search can improve retrieval accuracy. For this reason the detection of compound figures and their separation into sub figures is considered an important task. Examples for compound figures can be seen in Figure 2.



(a) Mixed modalities in a single figure (b) Graphs and microscopy images in a single figure

Fig. 2. Examples of compound figures found in the biomedical literature

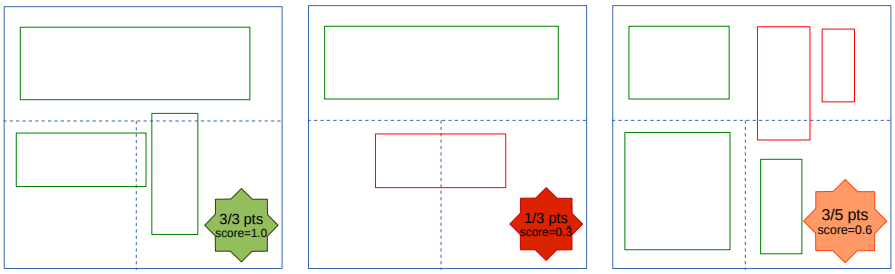
The data set used in the ImageCLEF 2013 compound figure separation task are all figures of the data set from the biomedical litertature. 2,967 compound figures were selected from the complete data set after a manual classification of images into compound and other figures. This subset was randomly split into two parts: a training set containing 1,538 images and a testing set with 1,429 images.

The ground truth for the dataset was generated in a semi–automatic way, using a two–step approach: first, an automated separation process (using the technique described in [10]) was run on both image sets in order to obtain a general overview of the subfigures. The automatic results were then manually corrected. Missing lines were added and incorrect lines removed, whereas often the lines were only slightly changed. Separating lines rather than bounding boxes were used to separate subfigures. The evaluation then required to have a minimum overlap between the ground truth and the data supplied by the groups in their runs.

The terminology used in the evaluation is:

- The term *figure*, F, refers to a compound figure as a whole.
- A *subfigure*, f_i, represents a part (or panel) of a figure. The ground truth for the figure F consists of a set of K^F_{GT} subfigures $f_1, \ldots, f_{K^F_{GT}}$.
- The word *candidate*, c_j, refers to the data being evaluated against the ground truth. Separation of figure F consists of a set of K^F_C candidates $c_1, \ldots, c_{K^F_C}$.

A brief summary of the evaluation algorithm for a given figure F is as follows:



(a) Perfect score (b) Not enough candidates (c) Too many candidates

Fig. 3. Examples for the separation of a compound figure. Dashed blue lines represent the ground truth, while solid lines represent the candidates. Valid candidates are shown in green and invalid candidates in red

- The score S_F is computed based on the number of correct candidates, $C^F_{correct}$.
- For each subfigure f_i defined in the ground truth the best matching candidate subfigure will be determined. Only one candidate is used in case there are several matches.
- The main metric used to compare subfigures is the overlap between a candidate subfigure and the ground truth. To be considered a valid match the overlap between a candidate subfigure and a subfigure from the ground truth must correspond to at least 66% of the candidate's size. If the best candidate is an acceptable match, the number of correctly matched figures $C^F_{correct}$ will be incremented. Since only one candidate subfigure can be assigned to each of the subfigures from the ground truth, $C^F_{correct} \leq K^F_{GT}$.
- The maximum score for the figure is 1 and the normalisation factor used to compute the score will be the maximum between the number of subfigures in the ground truth K^F_{GT} and the number of candidate subfigures K^F_C.

$$S_F = \frac{C^F_{correct}}{max(K^F_{GT}, K^F_C)}.$$

Therefore the maximum score is obtained only when the number of candidates K^F_C is equal to the number of subfigures in the ground truth K^F_{GT} and all of them are correctly matched:

$$C^F_{correct} = K^F_C = K^F_{GT}.$$

Figure 3 contains examples showing different candidates being validated against a reference figure (which contains 3 subfigures), along with their scores.

2.5 Image–Based Retrieval

The image–based retrieval task is the classical medical retrieval task, similar to those organized each year since 2004 with the target unit being the image. In

2013, 35 queries were given to the participants so more than in previous years. The 22 queries used in 2012 [7] were part of the 35 queries that all contain text (in English, Spanish, French and German) with 1–7 sample images for each query. As in previous years, the queries were classified into textual, mixed and semantic, based on the methods that are expected to yield the best results.

2.6 Case–Based Retrieval

The case–based retrieval task has been running since 2009. In this task, a case description, with patient demographics, limited symptoms and test results including imaging studies, is provided (but not the final diagnosis). As in previous years, the goal is to retrieve cases including images that might best suit the provided case description. This year the 26 topics distributed in 2012 were also part of the 35 final topics. Each of the topics was accompanied by one or two images.

3 Results

This section describes the results of ImageCLEF 2013. Runs are ordered based on the tasks (modality classification, compound figure separation, image–based and case–based retrieval) and the techniques used (visual, textual, mixed). In 2013, several groups used the ImageCLEF 2012 [7] database to optimize the parameters [11–13].

3.1 Modality Classification Results

Table 1 shows the classification accuracy obtained by the various runs submitted in the modality classification task. In 2013, the IBM Multimedia Analytics and FCSE [12] obtained the best results in the the three types of runs (visual, textual, mixed).Best results were obtained using multimodal techniques (81.68%) follow by visual techniques (80.79%). The best run using textual methods alone obtained a lower accuracy (64.17%). Only ITI [14] explored hierarchical approaches among the hierarchy distributed and some groups investigated a separation between compound and non–compound images before classifying the remaining categories [11, 15].

Techniques Used for Visual Classification The IBM team achived the best results in the visual classification.FCSE [12] was the second best group (77.14%) using a spatial pyramid in combination with dense sampling using an opponentSIFT descriptor for each image patch. Finally, Support Vector Machines (SVM) with χ^2 kernel were used as a classifier. As in 2012, multiple features were extracted from the images, most frequently color and edge directivity descriptors (CEDD) [11, 13, 14, 16, 17], fuzzy color and texture histogram (FCTH) [11, 13, 14, 16, 17] and scale–invariant feature transform (SIFT) variants [11, 12, 15]. Several classifiers were explored by the participants such as SVM [12, 14, 15, 17], k–nearest neighbour (k–nn) [11, 15] or class–centroid–based techniques [17].

Table 1. Results of the runs of the modality classification task

Run	Group	Run Type	Accuracy
IBM_modality_run8	IBM	Mixed	81.68
results_mixed_finki_run3	FCSE	Mixed	78.04
All	CITI	Mixed	72.92
IBM_modality_run9	IBM	Mixed	69.82
medgift2013_mc_mixed_k8	medGIFT	Mixed	69.63
medgift2013_mc_mixed_sem_k8	medGIFT	Mixed	69.63
nlm_mixed_using_2013_visual_classification_2	ITI	Mixed	69.28
nlm_mixed_using_2013_visual_classification_1	ITI	Mixed	68.74
nlm_mixed_hierarchy	ITI	Mixed	67.31
nlm_mixed_using_2012_visual_classification	ITI	Mixed	67.07
DEMIR_MC_5	DEMIR	Mixed	64.60
DEMIR_MC_3	DEMIR	Mixed	64.48
DEMIR_MC_6	DEMIR	Mixed	64.09
DEMIR_MC_4	DEMIR	Mixed	63.67
medgift2013_mc_mixed_exp_sep_sem_k21	medGIFT	Mixed	62.27
IPL13_mod_cl_mixed_r2	IPL	Mixed	61.03
IBM_modality_run10	IBM	Mixed	60.34
IPL13_mod_cl_mixed_r3	IPL	Mixed	58.98
medgift2013_mc_mixed_exp_k21	medGIFT	Mixed	47.83
medgift2013_mc_mixed_exp_sem_k21	medGIFT	Mixed	47.83
All_NoComb	CITI	Mixed	44.61
IPL13_mod_cl_mixed_r1	IPL	Mixed	09.56
IBM_modality_run1	IBM	Textual	64.17
results_text_finki_run2	FCSE	Textual	63.71
DEMIR_MC_1	DEMIR	Textual	62.70
DEMIR_MC_2	DEMIR	Textual	62.70
words	CITI	Textual	62.35
medgift2013_mc_text_k8.csv	medGIFT	Textual	62.04
nlm_textual_only_flat	ITI	Textual	51.23
IBM_modality_run2	IBM	Textual	39.07
words_noComb	CITI	Textual	32.80
IPL13_mod_cl_textual_r1	IPL	Textual	09.02
IBM_modality_run4	IBM	Visual	80.79
IBM_modality_run5	IBM	Visual	80.01
IBM_modality_run6	IBM	Visual	79.82
IBM_modality_run7	IBM	Visual	78.89
results_visual_finki_run1	FCSE	Visual	77.14
results_visual_compound_finki_run4	FCSE	Visual	76.29
IBM_modality_run3	IBM	Visual	75.94
sari_modality_baseline	MiiLab	Visual	66.46
sari_modality_CCTBB_DRxxDict	MiiLab	Visual	65.60
medgift2013_mc_5f	medGIFT	Visual	63.78
nlm_visual_only_hierarchy	ITI	Visual	61.50
medgift2013_mc_5f_exp_separate_k21	medGIFT	Visual	61.03
medgift2013_mc_5f_separate	medGIFT	Visual	59.25
CEDD_FCTH	CITI	Visual	57.62
IPL13_mod_cl_visual_r2	IPL	Visual	52.05
medgift2013_mc_5f_exp_k8	medGIFT	Visual	45.42
IPL13_mod_cl_visual_r3	IPL	Visual	43.33
CEDD_FCTH_NoComb	CITI	Visual	32.49
IPL13_mod_cl_visual_r1	IPL	Visual	06.19

Techniques Used for Classification Based on Text In 2012, only the ITI team [18] submitted runs for the textual modality classification task. In 2013, seven groups submitted textual results. A variety of techniques was employed using systems as Terrier IR[4] [12, 13], Lucene[5] [11, 16] or Essie [14].

Techniques Used for Multimodal Classification Eight groups submitted multimodal runs, five more than in 2012. The groups fused the techniques described above for visual and textual classification with a variety of fusion techniques, leading to the best results overall with multimodal techniques.

3.2 Compound Figure Separation Results

Three groups participated in the first year of the compound figure separation task (see Table 2). MedGIFT [11] achieved the best result in one of its runs but it simply serves as a point of reference, since it was also used when the separating lines were drawn [10] and thus has an advantage over other techniques. ITI [14] achived 69.27% using a combination of figure caption analysis, panel border detection and panel label recognition. FCSE [12] got 68.59% using an unsupervised algorithm based on a breadth–first search strategy using only visual information. Finally, medGIFT [11] submitted a second run which was not strictly designed for figure separation but provided a point of comparison. The run used a region detection algorithm mainly focused on volumetric medical image retrieval [19] with 46.82% of accuracy showing the possibility to use such techniques.

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

Run	Group	Run Type	Accuracy
HESSO_CFS	medGIFT	Visual	84.64
nlm_multipanel_separation	ITI	Mixed	69.27
fcse-final-noempty	FCSE	Visual	68.59
HESSO_REGIONDETECTOR_SCALE50_STANDARD	medGIFT	Visual	46.82

3.3 Image–Based Retrieval Results

Nine groups submitted image–based runs in 2013. The best results in terms of mean average precision (MAP) were obtained by ITI [14] using multimodal methods. The same group also obtained best results in 2012. The best textual run achieved the same MAP than the best multimodal run (0.3196). As in previous years, visual approaches achieved much lower results than the textual and multimodal techniques. Most of the techniques used in the retrieval task were also used for the modality classification task and are described in Section 3.1.

[4] http://terrier.org/
[5] http://lucene.apache.org/

Visual Retrieval Eight groups submitted 28 visual runs (see Table 3). DEMIR [13] achieved the best position in terms of MAP applying a classification algorithm. In addition to the techniques used in the modality classification task, some participants split and rescaled the images [17, 16]. Borda–fuse methods were also used [20].

Table 3. Results of the **visual** runs for the medical image retrieval task

Run Name	Group	MAP	GM-MAP	bpref	P10	P30
DEMIR4	DEMIR	0.0185	0.0005	0.0361	0.0629	0.0581
medgift_visual_nofilter	medGIFT	0.0133	0.0004	0.0256	0.0571	0.0448
medgift_visual_close	medGIFT	0.0132	0.0004	0.0256	0.0543	0.0438
medgift_visual_prefix	medGIFT	0.0129	0.0004	0.0253	0.0600	0.0467
IPL13_visual_r6	IPL	0.0119	0.0003	0.0229	0.0371	0.0286
image_latefusion_merge	ITI	0.0110	0.0003	0.0207	0.0257	0.0314
DEMIR5	DEMIR	0.0110	0.0004	0.0257	0.0400	0.0448
image_latefusion_merge_filter	ITI	0.0101	0.0003	0.0244	0.0343	0.0324
latefusuon_accuracy_merge	ITI	0.0092	0.0003	0.0179	0.0314	0.0286
IPL13_visual_r3	IPL	0.0087	0.0003	0.0173	0.0286	0.0257
sari_SURFContext_HI_baseline	MiiLab	0.0086	0.0003	0.0181	0.0429	0.0352
IPL13_visual_r8	IPL	0.0086	0.0003	0.0173	0.0286	0.0257
IPL13_visual_r5	IPL	0.0085	0.0003	0.0178	0.0314	0.0257
IPL13_visual_r1	IPL	0.0083	0.0002	0.0176	0.0314	0.0257
IPL13_visual_r4	IPL	0.0081	0.0002	0.0182	0.0400	0.0305
IPL13_visual_r7	IPL	0.0079	0.0003	0.0175	0.0257	0.0267
FCT_SEGHIST_6x6_LBP	CITI	0.0072	0.0001	0.0151	0.0343	0.0267
IPL13_visual_r2	IPL	0.0071	0.0001	0.0162	0.0257	0.0257
IBM_image_run_min_min	IBM	0.0062	0.0002	0.0160	0.0286	0.0267
DEMIR2	DEMIR	0.0044	0.0002	0.0152	0.0229	0.0229
SNUMedinfo13	SNUMedInfo	0.0043	0.0002	0.0126	0.0229	0.0181
SNUMedinfo12	SNUMedInfo	0.0033	0.0001	0.0153	0.0257	0.0219
IBM_image_run_Mnozero17	IBM	0.0030	0.0001	0.0089	0.0200	0.0105
SNUMedinfo14	SNUMedInfo	0.0023	0.0002	0.0090	0.0171	0.0124
SNUMedinfo15	SNUMedInfo	0.0019	0.0002	0.0074	0.0086	0.0114
IBM_image_run_Mavg7	IBM	0.0015	0.0001	0.0082	0.0171	0.0114
IBM_image_run_Mnozero11	IBM	0.0008	0	0.0045	0.0057	0.0095
nlm-se-image-based-visual	ITI	0.0002	0	0.0021	0.0029	0.0010

Textual Retrieval As for visual retrieval, eight groups submitted runs in the textual retrieval task (see Table 4). ITI [14] achieves the best results with a combination of two queries using Essie. The participants explored a variety of retrieval techniques mostly described in Section 3.1. FCSE [12] proposed a concept–scape approach matching the text data to medical concepts.

Multimodal Retrieval Only three groups submitted runs in the multimodal task (see Table 5). As in 2012, ITI [14] submitted the run with the highest MAP. For this run the group used the same method as the best textual run achieving exactly the same results. Mixed approaches combined the above textual and visual approaches using early [11, 14, 17] and late [11, 13, 14, 16] fusion strategies.

Table 4. Results of the **textual** runs for the medical image retrieval task

Run Name	Group	MAP	GM-MAP	bpref	P10	P30
nlm-se-image-based-textual	ITI	0.3196	0.1018	0.2982	0.3886	0.2686
IPL13_textual_r6	IPL	0.2542	0.0422	0.2479	0.3314	0.2333
BM25b1.1	FCSE	0.2507	0.0443	0.2497	0.3200	0.2238
finki	FCSE	0.2479	0.0515	0.2336	0.3057	0.2181
medgift_text_close	medGIFT	0.2478	0.0587	0.2513	0.3114	0.2410
finki	FCSE	0.2464	0.0508	0.2338	0.3114	0.2200
BM25b1.1	FCSE	0.2435	0.0430	0.2424	0.3314	0.2248
BM25b1.1	FCSE	0.2435	0.0430	0.2424	0.3314	0.2248
IPL13_textual_r4	IPL	0.2400	0.0607	0.2373	0.2857	0.2143
IPL13_textual_r1	IPL	0.2355	0.0583	0.2307	0.2771	0.2095
IPL13_textual_r8	IPL	0.2355	0.0579	0.2358	0.2800	0.2171
IPL13_textual_r8b	IPL	0.2355	0.0579	0.2358	0.2800	0.2171
IPL13_textual_r3	IPL	0.2354	0.0604	0.2294	0.2771	0.2124
IPL13_textual_r2	IPL	0.2350	0.0583	0.229	0.2771	0.2105
FCT_SOLR_BM25L_MSH	CITI	0.2305	0.0482	0.2316	0.2971	0.2181
medgift_text_nofilter	medGIFT	0.2281	0.0530	0.2269	0.2857	0.2133
IPL13_textual_r5	IPL	0.2266	0.0431	0.2285	0.2743	0.2086
medgift_text_prefix	medGIFT	0.2226	0.0470	0.2235	0.2943	0.2305
FCT_SOLR_BM25L	CITI	0.2200	0.0476	0.2280	0.2657	0.2114
DEMIR9	DEMIR	0.2003	0.0352	0.2158	0.2943	0.1952
DEMIR1	DEMIR	0.1951	0.0289	0.2036	0.2714	0.1895
DEMIR6	DEMIR	0.1951	0.0289	0.2036	0.2714	0.1895
SNUMedinfo11	SNUMedInfo	0.1800	0.0266	0.1866	0.2657	0.1895
DEMIR8	DEMIR	0.1578	0.0267	0.1712	0.2714	0.1733
finki	FCSE	0.1456	0.0244	0.1480	0.2000	0.1286
IBM_image_run_1	IBM	0.0848	0.0072	0.0876	0.1514	0.1038

Table 5. Results of the **multimodal** runs for the medical image retrieval task

Run Name	Group	MAP	GM-MAP	bpref	P10	P30
nlm-se-image-based-mixed	ITI	0.3196	0.1018	0.2983	0.3886	0.2686
Txt_Img_Wighted_Merge	ITI	0.3124	0.0971	0.3014	0.3886	0.2790
Merge_RankToScore_weighted	ITI	0.3120	0.1001	0.2950	0.3771	0.2686
Txt_Img_Wighted_Merge	ITI	0.3086	0.0942	0.2938	0.3857	0.2590
Merge_RankToScore_weighted	ITI	0.3032	0.0989	0.2872	0.3943	0.2705
medgift_mixed_rerank_close	medGIFT	0.2465	0.0567	0.2497	0.3229	0.2524
medgift_mixed_rerank_nofilter	medGIFT	0.2375	0.0539	0.2307	0.2886	0.2238
medgift_mixed_weighted_nofilter	medGIFT	0.2309	0.0567	0.2197	0.2800	0.2181
medgift_mixed_rerank_prefix	medGIFT	0.2271	0.0470	0.2289	0.2886	0.2362
DEMIR3	DEMIR	0.2168	0.0345	0.2255	0.3143	0.1914
DEMIR10	DEMIR	0.1583	0.0292	0.1775	0.2771	0.1867
DEMIR7	DEMIR	0.0225	0.0003	0.0355	0.0543	0.0543

3.4 Case–Based Retrieval Results

In 2013, the case–based retrieval task became more popular with seven groups submitting 42 runs. More groups than in previous years used visual and multimodal techniques. Textual runs achived the best results and visual runs obtained lower results than the textual and multimodal runs.

Visual Retrieval The results using visual retrieval on the case–based task are shown in Table 6. CITI [16] achived the best result outperforming the second best result by a factor of ten in terms of MAP. This group extracted a set of descriptors for 6×6 image grid.

Table 6. Results of the **visual** runs for the medical case–based retrieval task

Run Name	Group	MAP	GM-MAP	bpref	P10	P30
FCT_SEGHIST_6x6_LBP	CITI	0.0281	0.0009	0.0335	0.0429	0.0238
medgift_visual_nofilter_casebased	medGIFT	0.0029	0.0001	0.0035	0.0086	0.0067
medgift_visual_close_casebased	medGIFT	0.0029	0.0001	0.0036	0.0086	0.0076
medgift_visual_prefix_casebased	medGIFT	0.0029	0.0001	0.0036	0.0086	0.0067
nlm-se-case-based-visual	ITI	0.0008	0.0001	0.0044	0.0057	0.0057

Textual Retrieval Table 7 shows that SNUMedInfo [20] team achieved the best MAP (0.2429) in its first participation. SNUMedInfo used an external corpus (MEDLINE[6]) for robust and effective expansion term inference. CITI [16] achieved close results using MeSH expansion. ITI [14] and FCSE [12] incorporate UMLS (Unified Medical Language System) concepts. In general, the groups used the same techniques or very similar techniques compared to the ad–hoc image retrieval task.

Multimodal Retrieval Three groups submitted multimodal runs, combining of visual and textual techniques. As in the visual case–based task, the CITI [16] team achieved the best results in terms of MAP (see Table 8). A rank–based fusion was applied in their approach improving existing algorithms by a small margin.

4 Conclusions

After one decade of running the ImageCLEF medical task, in 2013 ImageCLEFmed is organized at the annual AMIA meeting in the form of a workshop. The task had 10 groups submitting 166 valid runs to the four subtasks. The main novelty in 2013 was the inclusion of a new task, the compound figure separation

[6] `http://www.nlm.nih.gov/bsd/pmresources.html`

Table 7. Results of the **textual** runs for the medical case–based retrieval task

Run Name	Group	MAP	GM-MAP	bpref	P10	P30
SNUMedinfo9	SNUMedInfo	0.2429	0.1163	0.2417	0.2657	0.1981
SNUMedinfo8	SNUMedInfo	0.2389	0.1279	0.2323	0.2686	0.1933
SNUMedinfo5	SNUMedInfo	0.2388	0.1266	0.2259	0.2543	0.1857
SNUMedinfo6	SNUMedInfo	0.2374	0.1112	0.2304	0.2486	0.1933
FCT_LUCENE_BM25L_MSH_PRF	CITI	0.2233	0.1177	0.2044	0.2600	0.1800
SNUMedinfo4	SNUMedInfo	0.2228	0.1281	0.2175	0.2343	0.1743
SNUMedinfo1	SNUMedInfo	0.2210	0.1208	0.1952	0.2343	0.1619
SNUMedinfo2	SNUMedInfo	0.2197	0.0996	0.1861	0.2257	0.1486
SNUMedinfo7	SNUMedInfo	0.2172	0.1266	0.2116	0.2486	0.1771
FCT_LUCENE_BM25L_PRF	CITI	0.1992	0.0964	0.1874	0.2343	0.1781
SNUMedinfo10	SNUMedInfo	0.1827	0.1146	0.1749	0.2143	0.1581
HES-SO-VS_FULLTEXT_LUCENE	medGIFT	0.1791	0.1107	0.1630	0.2143	0.1581
SNUMedinfo3	SNUMedInfo	0.1751	0.0606	0.1572	0.2114	0.1286
ITEC_FULLTEXT	AAUITEC	0.1689	0.0734	0.1731	0.2229	0.1552
ITEC_FULLPLUS	AAUITEC	0.1688	0.0740	0.1720	0.2171	0.1552
ITEC_FULLPLUSMESH	AAUITEC	0.1663	0.0747	0.1634	0.22	0.1667
ITEC_MESHEXPAND	AAUITEC	0.1581	0.0710	0.1635	0.2229	0.1686
IBM_run_1	IBM	0.1573	0.0296	0.1596	0.1571	0.1057
IBM_run_3	IBM	0.1573	0.0371	0.1390	0.1943	0.1276
IBM_run_3	IBM	0.1482	0.0254	0.1469	0.2000	0.1410
IBM_run_2	IBM	0.1476	0.0308	0.1363	0.2086	0.1295
IBM_run_1	IBM	0.1403	0.0216	0.1380	0.1829	0.1238
IBM_run_2	IBM	0.1306	0.0153	0.1340	0.2000	0.1276
nlm-se-case-based-textual	ITI	0.0885	0.0303	0.0926	0.1457	0.0962
DirichletLM_mu2500.0_Bo1bfree_d_3_t_10	FCSE	0.0632	0.0130	0.0648	0.0857	0.0676
DirichletLM_mu2500.0_Bo1bfree_d_3_t_10	FCSE	0.0632	0.0130	0.0648	0.0857	0.0676
finki	FCSE	0.0448	0.0115	0.0478	0.0714	0.0629
finki	FCSE	0.0448	0.0115	0.0478	0.0714	0.0629
DirichletLM_mu2500.0	FCSE	0.0438	0.0112	0.056	0.0829	0.0581
DirichletLM_mu2500.0	FCSE	0.0438	0.0112	0.056	0.0829	0.0581
finki	FCSE	0.0376	0.0105	0.0504	0.0771	0.0562
BM25b25.0	FCSE	0.0049	0.0005	0.0076	0.0143	0.0105
BM25b25.0_Bo1bfree_d_3_t_10	FCSE	0.0048	0.0005	0.0071	0.0143	0.0105

Table 8. Results of the **multimodal** runs for the medical case retrieval task

Run Name	Group	MAP	GM-MAP	bpref	P10	P30
FCT_CB_MM_rComb	CITI	0.1608	0.0779	0.1426	0.1800	0.1257
medgift_mixed_nofilter_casebased	medGIFT	0.1467	0.0883	0.1318	0.1971	0.1457
nlm-se-case-based-mixed	ITI	0.0886	0.0303	0.0926	0.1457	0.0962
FCT_CB_MM_MNZ	CITI	0.0794	0.0035	0.0850	0.1371	0.0810

task. In its first year three groups joined this complex task. More compound figures were included into the modality classification, so the training and test set are more difficult and correspond to the reality of the database, now. The other two tasks, image and case based retrieval, remained in the same format as in previous years but had a larger number of retrieval topics.

As in previous years, visual, textual or multimodal techniques can all perform best depending on the situation. For the modality classification, a mixed run achieved the best accuracy. For the image–based retrieval task, the highest MAP was achieved by a multimodal run. In the case–based retrieval task, textual techniques obtained the best results. Finally, for the compound figure separation task only visual and mixed techniques were explored, with visual techniques leading to best results.

In 2013, many groups used ImageCLEFmed 2012 database to optimize the parameters. Many of the techniques used had already been employed in previous years. This shows the utility of past campaigns, which provide databases as well as information regarding tools used by other participants. ImageCLEF conducts participative research and experimentation among free and reusable collections and has shown an important impact in visual medical information retrieval.

5 Acknowledgements

We would like to thank the EU FP7 projects Khresmoi (257528) and PROMISE (258191) for their support.

References

1. Caputo, B., Müller, H., Thomee, B., Villegas, M., Paredes, R., Zellhofer, D., Goeau, H., Joly, A., Bonnet, P., Martinez Gomez, J., Garcia Varea, I., Cazorla, C.: Imageclef 2013: the vision, the data and the open challenges. In: Working Notes of CLEF 2013 (Cross Language Evaluation Forum). (September 2013)
2. Hersh, W., Müller, H., Kalpathy-Cramer, J., Kim, E., Zhou, X.: The consolidated ImageCLEFmed medical image retrieval task test collection. Journal of Digital Imaging **22**(6) (2009) 648–655
3. Müller, H., Clough, P., Deselaers, T., Caputo, B., eds.: ImageCLEF – Experimental Evaluation in Visual Information Retrieval. Volume 32 of The Springer International Series On Information Retrieval. Springer, Berlin Heidelberg (2010)
4. Müller, H., Kalpathy-Cramer, J., Jr., C.E.K., Hatt, W., Bedrick, S., Hersh, W.: Overview of the ImageCLEFmed 2008 medical image retrieval task. In Peters, C., Giampiccolo, D., Ferro, N., Petras, V., Gonzalo, J., Peñas, A., Deselaers, T., Mandl, T., Jones, G., Kurimo, M., eds.: Evaluating Systems for Multilingual and Multimodal Information Access – 9th Workshop of the Cross-Language Evaluation Forum. Volume 5706 of Lecture Notes in Computer Science (LNCS)., Aarhus, Denmark (September 2009) 500–510
5. Müller, H., Kalpathy-Cramer, J., Eggel, I., Bedrick, S., Radhouani, S., Bakke, B., Kahn, J.C.E., Hersh, W.: Overview of the clef 2009 medical image retrieval track. In: Proceedings of the 10th international conference on Cross-language

evaluation forum: multimedia experiments. CLEF'09, Berlin, Heidelberg, Springer–Verlag (2010) 72–84
6. Kalpathy-Cramer, J., Müller, H., Bedrick, S., Eggel, I., García Seco de Herrera, A., Tsikrika, T.: The CLEF 2011 medical image retrieval and classification tasks. In: Working Notes of CLEF 2011 (Cross Language Evaluation Forum). (September 2011)
7. 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 classication tasks. In: Working Notes of CLEF 2012 (Cross Language Evaluation Forum). (September 2012)
8. 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)
9. Tirilly, P., Lu, K., Mu, X., Zhao, T., Cao, Y.: On modality classification and its use in text–based image retrieval in medical databases. In: Proceedings of the 9th International Workshop on Content–Based Multimedia Indenxing. CBMi2011 (2011)
10. Chhatkuli, A., Markonis, D., Foncubierta-Rodríguez, A., Meriaudeau, F., Müller, H.: Separating compound figures in journal articles to allow for subfigure classification. In: SPIE, Medical Imaging. (2013)
11. García Seco de Herrera, A., Markonis, D., Schaer, R., Eggel, I., Müller, H.: The medGIFT group in ImageCLEFmed 2013. In: Working Notes of CLEF 2013 (Cross Language Evaluation Forum). (September 2013)
12. Kitanovski, I., Dimitrovski, I., Loskovska, S.: FCSE at medical tasks of ImageCLEF 2013. In: Working Notes of CLEF 2013 (Cross Language Evaluation Forum). (September 2013)
13. Ozturkmenoglu, O., Ceylan, N.M., Alpkocak, A.: DEMIR at ImageCLEFmed 2013: The effects of modality classification to information retrieval. In: Working Notes of CLEF 2013 (Cross Language Evaluation Forum). (September 2013)
14. Simpson, M.S., You, D., Rahman, M.M., Demner-Fushman, D., Antani, S., Thoma, G.: ITI's participation in the 2013 medical track of ImageCLEF. In: Working Notes of CLEF 2013 (Cross Language Evaluation Forum). (September 2013)
15. Zhou, X., Han, M., Song, Y., Li, Q.: Fast filtering techniques in medical image classification and retrieval. In: Working Notes of CLEF 2013 (Cross Language Evaluation Forum). (September 2013)
16. Mourão, A., Martins, F., Magalhães, J.a.: NovaSearch on medical ImageCLEF 2013. In: Working Notes of CLEF 2013 (Cross Language Evaluation Forum). (September 2013)
17. Stathopoulos, S., Lourentzou, I., Kyriakopoulou, A., Kalamboukis, T.: IPL at CLEF 2013 medical retrieval task. In: Working Notes of CLEF 2013 (Cross Language Evaluation Forum). (September 2013)
18. Simpson, M.S., You, D., Rahman, M.M., Demmer-Fushman, D., Antani, S., Thoma, G.: ITI's participation in the ImageCLEF 2012 medical retrieval and classification tasks. In: Working Notes of CLEF 2012. (2012)
19. Foncubierta-Rodríguez, A., Müller, H., Depeursinge, A.: Region–based volumetric medical image retrieval. In: SPIE Medical Imaging: Advanced PACS–based Imaging Informatics and Therapeutic Applications. (2013)
20. Sungbin, C., Lee, J., Cho, J.: SNUMedinfo at ImageCLEF 2013: Medical retrieval task. In: Working Notes of CLEF 2013 (Cross Language Evaluation Forum). (September 2013)