

# ITI's Participation in the 2013 Medical Track of ImageCLEF

Matthew S. Simpson, Daekeun You, Md Mahmudur Rahman, Dina Demner-Fushman, Sameer Antani, and George Thoma

Lister Hill National Center for Biomedical Communications, U.S. National Library of Medicine, NIH, Bethesda, MD, USA

**Abstract.** This article describes the participation of the Image and Text Integration (ITI) group in the ImageCLEF medical retrieval, classification, and segmentation tasks. Although our methods are similar to those we have explored at past ImageCLEF evaluations, we describe in this paper the results of our methods on the 2013 collection and set of topics. In doing so, we present our submitted textual, visual, and mixed runs and our results for each of the four tasks. Like our participation in previous evaluations, we found our methods to generally perform well for each task. In particular, our best ad-hoc retrieval submission was again ranked first among all the submissions from the participating groups.

**Keywords:** Image Retrieval, Case-based Retrieval, Image Modality

## 1 Introduction

This article describes the participation of the Image and Text Integration (ITI) group in the ImageCLEF 2013 medical retrieval, classification, and segmentation tasks. Our group is from the Communications Engineering Branch of the Lister Hill National Center for Biomedical Communications, which is a research division of the U.S. National Library of Medicine.

The medical track [4] of ImageCLEF 2013 consists of an image modality classification task, a compound figure separation task, and two retrieval tasks. For the classification task, the goal is to classify a given set of images according to thirty-one modalities (e.g., “Computerized Tomography,” “Electron Microscopy,” etc.). The modalities are organized hierarchically into meta-classes such as “Radiology” and “Microscopy,” which are themselves types of “Diagnostic Images.” For the compound figure separation task, the goal is to segment the panels of multi-panel figures. Figures contained in biomedical articles are often composed of multiple panels (e.g., commonly labeled “a,” “b,” etc.) and segmenting them can result in improved retrieval performance. In the first retrieval task, a set of ad-hoc information requests is given, and the goal is to retrieve the most relevant images from a collection of biomedical articles for each topic. Finally, in the second retrieval task, a set of case-based information requests is given, and the goal is to retrieve the most relevant articles describing similar cases.

In the following sections, we describe our methods and results. In Section 2, we briefly outline our approach to each of the four tasks. In Section 3, we describe each of our submitted runs, and in Section 4 we present our results. For the modality classification task, our best submission achieved a classification accuracy of 69.28%, which is better than what we achieved in the previous ImageCLEF evaluation. Our submission for the compound figure separation task achieved a similar accuracy of 69.27%. Our best submission for the ad-hoc image retrieval task was a mixed approach that achieved a mean average precision of 0.3196. This result is comparable to what we achieved in the previous evaluation and is again ranked first among all submissions for the participating groups. Finally, for the case-based article retrieval task, our best submission achieved a mean average precision of 0.0886, which is significantly lower than the top-ranked run. In each of the above tasks, we obtained our best results using mixed approaches, indicating the importance of both textual and visual features for these tasks.

## 2 Methods

The methods we used in participating in the 2013 medical track of ImageCLEF are identical to the approaches we explored in the 2012 evaluation [11]. We briefly summarize these methods below.

**Table 1:** Extracted visual descriptors.

No. Descriptor	Dimensionality
1. Autocorrelation	25
2. Edge frequency	25
3. Fuzzy color and texture histogram* (FCTH) [3]	192
4. Gabor moment*	60
5. Gray-level co-occurrence matrix moment (GLCM) [12]	20
6. Local binary pattern (LBP <sub>1</sub> ) [9]	256
7. Local binary pattern (LBP <sub>2</sub> ) [9]	256
8. Scale-invariant feature transformation* (SIFT) [7]	256
9. Shape moment	5
10. Tamura moment* [13]	18
11. Edge histogram* (EHD) [1]	80
12. Color and edge directivity* (CEDD) [2]	144
13. Primitive length	5
14. Color layout* (CLD) [1]	16
15. Color moment	3
16. Semantic concept (SCONCEPT) [10]	30
Combined	1391

\* Feature computed using the Lucene Image Retrieval library [8].

We represent images and the articles in which they are contained using a combination of textual and visual features. Our textual features include the title, abstract, and Medical Subject Headings (MeSH<sup>®</sup> terms) of the articles in which the images appear as well as the images’ captions and “mentions” (snippets of text within the body of an article that discuss the images). In addition to the above textual features, we also represent the visual content of images using various low-level visual descriptors. Table 1 summarizes the descriptors we extract and their dimensionality. Due to the large number of these features, we forego describing them in any detail. However, they are all well-known and discussed extensively in existing literature.

For the modality classification task, we experimented with both flat and hierarchical classification strategies using support vector machines (SVMs). First, we extract our visual and textual image features from the training images (representing the textual features as term vectors). Then, we perform attribute selection to reduce the dimensionality of the features. We construct the lower-dimensional vectors independently for each feature type (textual or visual) and combine the resulting attributes into a single, compound vector. Finally, we use the lower-dimensional feature vectors to train multi-class SVMs for producing textual, visual, or mixed modality predictions. Our flat classifiers attempt to classify images into one of the thirty-one modality classes whereas our hierarchical classifiers attempt to classify images following the structured organization of modalities provided by the ImageCLEF organizers.

For the compound figure separation task, our method incorporates both natural language and image processing techniques. Our method first seeks to determine the number of image panels comprising a compound figure by identifying textual panel labels in the figure’s caption and visual panel labels overlain on the figure. A border detection method combines this information to determine the appropriate borders and segment the figure.

For the ad-hoc image retrieval task, we explored a variety of textual, visual, and mixed strategies. Our textual approaches utilize the Essie [5] retrieval system. Essie is a biomedical search engine developed by the U.S. National Library of Medicine, and it incorporates the synonymy relationships encoded in the Unified Medical Language System<sup>®</sup> (UMLS<sup>®</sup>) Metathesaurus<sup>®</sup> [6]. Our visual approaches are based on retrieving images that appear visually similar to the given topic images. We compute the visual similarity between two images as the Euclidean distance between their visual descriptors. For the purposes of computing this distance, we represent each image as a combined feature vector composed of a subset of the visual descriptors listed in Table 1. We also explored methods involving the clustering of visual descriptors and attribute selection. Finally, our mixed approaches combine the above textual and visual approaches in both early and late fusion strategies.

Our method for performing case-based article retrieval is analogous to our approaches for the ad-hoc image retrieval task. The only substantive difference is that we represent articles by a combination of the textual and visual features of each image they contain.

### 3 Submitted Runs

In this section we describe each of our submitted runs for the modality classification, compound figure separation, ad-hoc image retrieval, and case-based article retrieval tasks. Each run is identified by its file name or `trec_eval` run ID and mode (textual, visual or mixed). All submitted runs are automatic.

#### 3.1 Modality Classification Runs

We submitted the following six runs for the modality classification task:

- M1. *nlm-textual-only-flat (textual)*: A flat multi-class SVM classification using selected attributes from a combined term vector created from four textual features (article title, MeSH terms, and image caption and mention).
- M2. *nlm-visual-only-hierarchy (visual)*: A hierarchical multi-class SVM classification using selected attributes from a combined visual descriptor of features 1–15 of Table 1.
- M3. *nlm-mixed-hierarchy (mixed)*: A hierarchical multi-class SVM classification combining Runs 1 and 2. Textual and visual features are combined into a single feature vector for each image.
- M4. *nlm-mixed-using-2012-visual-classification (mixed)*: A combination of Runs 1 and 2 but using models trained on the 2012 ImageCLEF medical modality classification data set. Images are first classified according to Run 1. Images having no textual features are classified according to Run 2. We use our compound figure separation method to improve the classification accuracy of some classes.
- M5. *nlm-mixed-using-2013-visual-classification\_1 (mixed)*: Like Run 4 but using the 2013 ImageCLEF medical modality classification data set.
- M6. *nlm-mixed-using-2013-visual-classification\_2 (mixed)*: Like Run 5 but using all visual features from Table 1.

#### 3.2 Compound Figure Separation Runs

We submitted the following run for the compound figure separation task:

- S1. *nlm-multipanel-separation (mixed)*: A combination of figure caption analysis, panel border detection, and panel label recognition.

#### 3.3 Ad-hoc Image Retrieval Runs

We submitted the following ten runs for the ad-hoc image retrieval task:

- A1. *nlm-image-based-textual (textual)*: A combination of two queries using Essie. (A1.Q1) A disjunction of modality terms extracted from the query topic must occur within the caption or mention fields of an image’s textual features; a disjunction of the remaining terms is allowed to occur in any field. (A1.Q2) A lossy expansion of the verbatim topic is allowed to occur in any field.

- A2. *nlm-image-based-visual (visual)*: A disjunction of the query images' clustered visual descriptors must occur within the global image feature field.
- A3. *nlm-image-based-mixed (mixed)*: A combination of Queries A1.Q1–Q2 with Run A2.
- A4. *image-latefusion-merge (visual)*: An automatic content-based image retrieval approach. In this approach, features 10–16 of Table 1 are used, and their individual similarity scores are linearly combined with predefined weights based on modality classification results of the query and collection images. All images in each topic are considered and result lists for each topic are combined to produce a single list of retrieved images.
- A5. *image-latefusion-merge-filter (visual)*: Like Run A4 but the search is performed after filtering the collection of images based on modality classification results of the query images.
- A6. *latefusion-accuracy-merge (visual)*: Like Run A4 but the feature weights are based on their normalized accuracy in classifying images in the 2012 ImageCLEF medical modality classification test set.
- A7. *Txt\_Img\_Wighted\_Merge (mixed)*: A score-based combination of Runs A1 and A5.
- A8. *Merge\_RankToScore\_weighted (mixed)*: A rank-based combination of Runs A1 and A5.
- A9. *Txt\_Img\_Wighted\_Merge\_A (mixed)*: A score-based combination of Runs A1 and A6.
- A10. *Merge\_RankToScore\_weighted\_A (mixed)*: A rank-based combination of Runs A1 and A6.

### 3.4 Case-based Article Retrieval Runs

We submitted the following three runs for the case-based article retrieval task:

- C1. *nlm-case-based-textual (textual)*: A combination of three queries for each topic sentence using Essie. (C1.Q1) A disjunction of modality terms extracted from the sentence must occur within the caption or mention fields of an article's textual features; a disjunction of the remaining terms is allowed to occur in any field. (C1.Q2) A lossy expansion of the verbatim sentence is allowed to occur in any field. (C1.Q3) A disjunction of all extracted words in the sentence is allowed to occur in any field. Articles are scored according to the sentence resulting in the maximum score.
- C2. *nlm-case-based-visual (visual)*: A disjunction of the query images' clustered visual descriptors must occur within the global image feature field.
- C3. *nlm-case-based-mixed (mixed)*: A combination of Queries A1.Q1–Q3 with Run A2.

## 4 Results

Tables 2–5 summarize the results of our modality classification, compound figure separation, ad-hoc image retrieval, and case-based article retrieval runs. In Table 2

**Table 2:** Accuracy results for the modality classification task.

<b>ID</b>	<b>Mode</b>	<b>Accuracy (%)</b>
<i>nlm-mixed-using-2013-visual-classification-2</i>	Mixed	69.28
<i>nlm-mixed-using-2013-visual-classification-1</i>	Mixed	68.74
<i>nlm-mixed-hierarchy</i>	Mixed	67.31
<i>nlm-mixed-using-2012-visual-classification</i>	Mixed	67.07
<i>nlm-visual-only-hierarchy</i>	Visual	61.50
<i>nlm-textual-only-flat</i>	Textual	51.23

**Table 3:** Accuracy results for the compound figure separation task.

<b>ID</b>	<b>Mode</b>	<b>Accuracy (%)</b>
<i>nlm-multipanel-separation</i>	Mixed	69.27

and Table 3, we give the accuracy of our figure classification and separation methods. In Table 4 and Table 5, we give the mean average precision (MAP), binary preference (bpref) and precision-at-ten (P@10) of our retrieval methods.

## 5 Conclusion

This article describes the methods and results of the Image and Text Integration (ITI) group in the ImageCLEF 2013 medical classification, segmentation, and retrieval tasks. Our methods are similar to those we have developed for previous ImageCLEF evaluations, and they include a variety of textual, visual, and mixed approaches. For the modality classification task, our best submission was a mixed approach that achieved an accuracy of 69.28% and was ranked within the submissions from the top five participating groups. For the compound figure

**Table 4:** Retrieval results for the ad-hoc image retrieval task

<b>ID</b>	<b>Mode</b>	<b>MAP</b>	<b>bpref</b>	<b>P@10</b>
<i>nlm-se-image-based-mixed</i>	Mixed	0.3196	0.2983	0.3886
<i>nlm-se-image-based-textual</i>	Textual	0.3196	0.2982	0.3886
<i>Txt_Img_Wighted_Merge_A</i>	Mixed	0.3124	0.3014	0.3886
<i>Merge_RankToScore_weighted_A</i>	Mixed	0.3120	0.2950	0.3771
<i>Txt_Img_Wighted_Merge</i>	Mixed	0.3086	0.2938	0.3857
<i>Merge_RankToScore_weighted</i>	Mixed	0.3032	0.2872	0.3943
<i>image.latefusion.merge</i>	Visual	0.0110	0.0207	0.0257
<i>image.latefusion.merge-filter</i>	Visual	0.0101	0.0244	0.0343
<i>latefusuon_accuracy-merge</i>	Visual	0.0092	0.0179	0.0314
<i>nlm-se-image-based-visual</i>	Visual	0.0002	0.0021	0.0029

**Table 5:** Retrieval results for the case-based article retrieval task

<b>ID</b>	<b>Mode</b>	<b>MAP</b>	<b>bpref</b>	<b>P@10</b>
<i>nlm-se-case-based-mixed</i>	Mixed	0.0886	0.0926	0.1457
<i>nlm-se-case-based-textual</i>	Textual	0.0885	0.0926	0.1457
<i>nlm-se-case-based-visual</i>	Visual	0.0008	0.0044	0.0057

separation task, our mixed approach resulted in an accuracy of 69.27% and was ranked second among four submissions from three groups participating in this task. Similar to our experience in previous years, our best submission for the ad-hoc image retrieval task was also a mixed approach, achieving a mean average precision of 0.3106 and ranking first overall. Finally, for the case-based article retrieval task, our best submission obtained a mean average precision of 0.0886. This result is much lower than what we have achieved in previous ImageCLEF evaluations. Despite our performance on the case-based task, the effectiveness of our mixed approaches are encouraging and provide evidence that our ongoing efforts at integrating textual and visual information will be successful.

**Acknowledgments.** We would like to thank Suchet Chandra for preparing our collection and extracting the textual and visual features used by our methods.

## References

1. Chang, S.F., Sikora, T., Puri, A.: Overview of the MPEG-7 standard. *IEEE Transactions on Circuits and Systems for Video Technology* 11(6), 688–695 (2001)
2. Chatzichristofis, S.A., Boutalis, Y.S.: CEDD: Color and edge directivity descriptor: A compact descriptor for image indexing and retrieval. In: Gasteratos, A., Vincze, M., Tsotsos, J.K. (eds.) *Proceedings of the 6th International Conference on Computer Vision Systems*. *Lecture Notes in Computer Science*, vol. 5008, pp. 312–322. Springer (2008)
3. Chatzichristofis, S.A., Boutalis, Y.S.: FCTH: Fuzzy color and texture histogram: A low level feature for accurate image retrieval. In: *Proceedings of the 9th International Workshop on Image Analysis for Multimedia Interactive Services*. pp. 191–196 (2008)
4. de Herrera, G.S., Kalpathy-Cramer, J., Demner-Fushman, D., Antani, S., Müller, H.: Overview of the ImageCLEF 2013 medical tasks. In: *Working notes of CLEF 2013* (2013)
5. Ide, N.C., Loane, R.F., Demner-Fushman, D.: Essie: A concept-based search engine for structured biomedical text. *Journal of the American Medical Informatics Association* 1(3), 253–263 (2007)
6. Lindberg, D., Humphreys, B., McCray, A.: The unified medical language system. *Methods of Information in Medicine* 32(4), 281–291 (1993)
7. Lowe, D.: Object recognition from local scale-invariant features. In: *Proceedings of the Seventh IEEE International Conference on Computer Vision*. vol. 2, pp. 1150–1157 (1999)

8. Lux, M., Chatzichristofis, S.A.: LIRe: Lucene image retrieval—an extensible java CBIR library. In: Proceedings of the 16th ACM International Conference on Multimedia. pp. 1085–1088 (2008)
9. Mäenpää, T.: The Local Binary Pattern Approach to Texture Analysis—Extensions and Applications. Ph.D. thesis, University of Oulu (2003)
10. Rahman, M.M., Antani, S., Thoma, G.: A medical image retrieval framework in correlation enhanced visual concept feature space. In: Proceedings of the 22nd IEEE International Symposium on Computer-Based Medical Systems (2009)
11. Simpson, M.S., You, D., Rahman, M.M., Demner-Fushman, D., Antani, S., Thoma, G.: ITI’s participation in the ImageCLEF 2012 medical retrieval and classification tasks. In: Working Notes for the Conference on Multilingual and Multimodal Information Access Evaluation (CLEF). September 17–20 (Rome, Italy 2012)
12. Srinivasan, G.N., Shobha, G.: Statistical texture analysis. In: Proceedings of World Academy of Science, Engineering and Technology. vol. 36, pp. 1264–9 (2008)
13. Tamura, H., Mori, S., Yamawaki, T.: Textural features corresponding to visual perception. IEEE Transactions on Systems, Man, and Cybernetics 8(6), 460–73 (1978)