

# ITI's Participation in the ImageCLEF 2012 Medical Retrieval and Classification Tasks

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 2012 ImageCLEF medical retrieval and classification tasks. We present our methods for each of the three tasks and discuss our submitted textual, visual, and mixed runs as well as their results. While our methods generally perform well for each task, our best ad-hoc image retrieval submission was 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 2012 medical retrieval and classification tasks. Our group is from the Communications Engineering Branch of the Lister Hill National Center for Biomedical Communications, which is a division of the U. S. National Library of Medicine.

The medical track [15] of ImageCLEF 2012 consists of an image modality classification task and two retrieval tasks. For the classification task, the goal is to classify a given set of medical 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.” 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 the textual and visual features that comprise our image and article representations (Sections 2–3) and our methods for the modality classification (Section 4) and medical retrieval tasks (Sections 5–6). Our textual approaches primarily utilize the Unified Medical Language System<sup>®</sup>(UMLS<sup>®</sup>) [11] synonymy to identify concepts in topic descriptions and article text, and our visual approaches rely on computed distances between descriptors of various low-level visual features. In developing mixed approaches, we explore the use of clustered visual features that can be represented using text, attribute selection, and ranked list merging strategies.

In Section 7 we describe our submitted runs, and in Section 8 we present our results. For the modality classification task, our best submission achieved a classification accuracy of 63.2% and was ranked within the submissions from the top three participating groups. Our best submission for the ad-hoc image retrieval task was ranked first overall, achieving a mean average precision of 0.2377, which is a statistically significant improvement over the second ranked submission. For the case-based article retrieval task, our best submission achieved a mean average precision of 0.1035 and was ranked within the submissions from the top four participating groups, but this submission is statistically indistinguishable from our other case-based submissions.

## 2 Image Representation for Ad-hoc Retrieval

We represent the images contained in biomedical articles using a combination of the textual and visual features described below.

### 2.1 Textual Features

We represent each image in the collection as a structured document of image-related text called an *enriched citation*. Our representation includes the title, abstract, and MeSH<sup>®</sup> terms<sup>1</sup> of the article in which the image appears as well as the image’s caption and “mentions” (snippets of text within the body of an article that discuss an image). These features can be indexed with a traditional text-based information retrieval system, or they may be exposed as term vectors and combined with the visual feature vectors described below.

### 2.2 Visual Features

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.

**Cluster Words.** To avoid the computational complexity of computing distances between the above visual descriptors, we create a textual representation of visual features that is easily integrated with our existing textual features. For each visual descriptor listed in Table 1, we cluster the vectors assigned to all images using the *k*-means++ [3] algorithm. We then assign each cluster a unique “cluster word” and represent each image as a sequence of these words. We add an image’s cluster words to its enriched citation as a “global image feature” field, which can be searched using a traditional text-based information retrieval system.

**Attribute Selection.** An orthogonal approach to transforming our visual descriptors into a computationally manageable representation is attribute selection. By eliminating unneeded or redundant information, these techniques can also improve our modality classification and image retrieval methods. We perform attribute selection using the WEKA [8] data mining software. First, we group all our visual descriptors into a single combined vector, and we then perform attribute selection to reduce the dimensionality of this combined feature.

---

<sup>1</sup> MeSH is a controlled vocabulary created by U. S. National Library of Medicine to index biomedical articles.

**Table 1:** Extracted visual descriptors.

Descriptor	Dimensionality
Autocorrelation	25
Color and edge directivity* (CEDD) [5]	144
Color layout* (CLD) [4]	16
Color moment	3
Edge frequency	25
Edge histogram* (EHD) [4]	80
Fuzzy color and texture histogram* (FCTH) [6]	192
Gabor moment*	60
Gray-level co-occurrence matrix moment (GLCM) [19]	20
Local binary pattern (LBP <sub>1</sub> ) [14]	256
Local binary pattern (LBP <sub>2</sub> ) [14]	256
Local color histogram (LCH)	1024
Primitive length	5
Scale-invariant feature transformation* (SIFT) [12]	256
Semantic concept (SCONCEPT) [16]	30
Shape moment	5
Tamura moment* [20]	18
Combined	2415

\* Feature computed using the Lucene Image Retrieval library [13].

### 3 Article Representation for Case-based Retrieval

We represent articles using the textual features of each image appearing in the article. Thus, each article enriched citation consists of its title, abstract, and MeSH terms as well as the caption and mention of each contained image.

### 4 Modality Classification Task

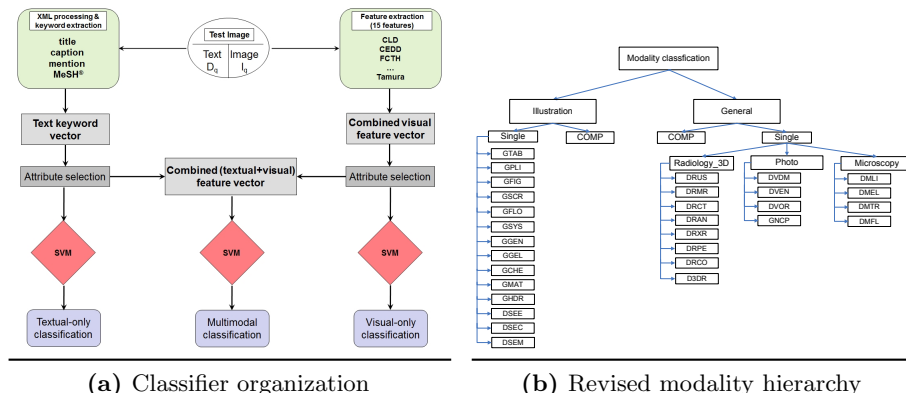
We experimented with both flat and hierarchical modality classification methods. Below we describe our flat classification strategy, an extension of this approach that exploits the hierarchical structure of the classes, and a post-processing method for improving the classification accuracy of illustrations.

#### 4.1 Flat Classification

Figure 1a provides an overview of our basic classification approach. We utilize multi-class support vector machines (SVMs) as our flat modality classifiers. 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.

#### 4.2 Hierarchical Classification

Unlike the flat classification strategy described above, it is possible to exploit the hierarchical organization of the modality classes in order to decompose the task



**Fig. 1:** Overview of our flat and hierarchical modality classification strategies.

into several smaller classification problems that can be sequentially applied. Based on our visual observation of the training samples and our initial experiments, we modified the original modality hierarchy [15] proposed for the task. The hierarchy we used for our experiments is shown in Figure 1b.

We train flat multi-class SVMs, as shown in Figure 1a, for each meta-class. For recognizing compound images, we utilize the algorithm proposed by Apostilova et al. [1], which detects sub-figure labels and the border of each sub-figure within a compound image. To arrive at a final class label, an image is sequentially classified beginning at the root of the hierarchy until a leaf class can be determined.

### 4.3 Illustration Post-processing

Because our initial classification experiments resulted in only modest accuracy for the fourteen “Illustration” classes shown in Figure 1b, we concluded that our current textual and visual features may not be sufficient for representing these figures. Therefore, in addition to the aforementioned machine learning modality classification methods, we also developed several complimentary rule-based strategies for increasing the classification accuracy of “Illustration” classes.

A majority of the training samples contained in the “Illustration” meta-class, unlike other images in the collection, consist of line drawings or text superimposed on a white background. For example, program listings mostly consist of text; thus, the use of text and line detection methods may increase the classification accuracy of Class GPLI. Similarly, polygons (e.g., rectangles, hexagons, etc.) contained in flowcharts (GFLO), tables (GTAB), system overviews (GSYS), and chemical structures (GCHE) are a distinctive feature of these modalities. We utilize the methods of Jung et al. [10] and OpenCV<sup>2</sup> functions to assess the presence of text and polygons, respectively.

## 5 Ad-Hoc Image Retrieval Task

In this section we describe our textual, visual and mixed approaches to the ad-hoc image retrieval task. Descriptions of the submitted runs that utilize these methods are presented in Section 7.

<sup>2</sup> <http://opencv.willowgarage.com/wiki/>

### 5.1 Textual Approaches

To allow for efficient retrieval and to compare their relative performance, we index our enriched citations with the Essie [9] and Lucene/SOLR<sup>3</sup> search engines. Essie is a search engine developed by the U.S. National Library of Medicine and is particularly well-suited for the medical retrieval task due to its ability to automatically expand query terms using the UMLS synonymy. Lucene/SOLR is a popular search engine developed by the Apache Software Foundation that employs the well-known vector space model of information retrieval. We have extracted the UMLS synonymy from Essie and use it for term expansion when indexing enriched citations with Lucene/SOLR.

We organize each topic description into a frame-based (e.g., PICO<sup>4</sup>) representation following the method similar to that described by Demner-Fushman and Lin [7]. Extractors identify concepts related to problems, interventions, age, anatomy, drugs, and modality. We also identify modifiers of the extracted concepts and a limited number of relationships among them. We then transform the extracted concepts into queries appropriate for either Essie or Lucene/SOLR.

### 5.2 Visual Approaches

Our visual approaches to image retrieval 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 after attribute selection.

### 5.3 Mixed Approaches

We explore several methods of combining our textual and visual approaches. One such approach involves the use of our image cluster words. For performing multimodal retrieval using cluster words, we first extract the visual descriptors listed in Table 1 from each example image of a given topic. We then locate the clusters to which the extracted descriptors are nearest in order to determine their corresponding cluster words. Finally, we combine these cluster words with words taken from the topic description to form a multimodal query appropriate for either Essie or Lucene/SOLR.

While the use of cluster words allows us to create multimodal queries, we can instead directly combine the independent outputs of our textual and visual approaches. In a *score merging* approach, we apply a min-max normalization to the ranked lists of scores produced by our textual and visual retrieval strategies. We then linearly combine the normalized scores given to each image to produce a final ranking. Similarly, a *rank merging* approach combines the results of our textual and visual approaches using the ranks of the retrieved images instead of their normalized scores. To produce the final image ranking using this strategy, we re-score each retrieved image as the reciprocal of its rank and then repeat the above procedure for combining scores.

---

<sup>3</sup> <http://lucene.apache.org/>

<sup>4</sup> PICO is a mnemonic for structuring clinical questions in evidence-based practice and represents Patient/Population/Problem, Intervention, Comparison, and Outcome.

Another means of incorporating visual information with our retrieval approaches is through the use of a modality classifier. Using our hierarchical modality classification approach, we can first determine the most probable modalities for a topic’s example images. After retrieving a set of images using either our textual or visual methods, we can eliminate retrieved images that are not of the same modality as the topic images. An advantage of performing hierarchical classification is that we can filter the retrieved results using the meta-classes within the hierarchy (e.g., “Radiology”).

Finally, we often combine the retrieval results produced by several queries into a single ranked list of images. We perform this query combination, or padding, by simply appending the ranked list of images retrieved by a subsequent query to the end of the ranked list produced by the preceding query.

## 6 Case-Based Retrieval Task

Our method for performing case-based retrieval is analogous to our textual approaches for ad-hoc image retrieval. Here, we index the enriched citations described in Section 3 using the Essie and Lucene/SOLR search engines (for performance comparison). We generate textual and mixed queries appropriate for both search engines according to the approaches described in Sections 5.1.

As a form of query expansion for case-based topics, we also explore the possibility of determining relevant disease names to correspond with signs and symptoms found in a topic case. To determine a set of potential diseases, we first use the Google Search API<sup>5</sup> to search the World Wide Web using a topic case as a query. We then process the top five documents with MetaMap [2] to extract terms having the UMLS semantic type “Disease or Syndrome.” Finally, we select the top three most frequent diseases for query expansion.

## 7 Submitted Runs

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

### 7.1 Modality Classification Task

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

- M1. *Visual\_only\_Flat.txt (visual)*: A flat multi-class SVM classification using selected attributes from a combined visual descriptor of 15 features (all descriptors in Table 1 except LCH and SCONCEPT).
- M2. *Visual\_only\_Hierarchy.txt (visual)*: Like Run M1 but classification is performed hierarchically.
- M3. *Text\_only\_Flat.txt (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).
- M4. *Text\_only\_Hierarchy.txt (textual)*: Like Run M3 but classification is performed hierarchically.

---

<sup>5</sup> <https://developers.google.com/custom-search/v1/overview>

- M5. *Visual\_Text\_Flat.txt (mixed)*: A flat multi-class SVM classification combining the feature representations used in Runs M1–3.
- M6. *Visual\_Text\_Hierarchy.txt (mixed)*: Like Run M5 but classification is performed hierarchically.
- M7. *Visual\_Text\_Flat\_w\_Postprocessing\_4\_Illustration.txt (mixed)*: Like Run M5 but additional post-processing is applied for “Illustration” classes.
- M8. *Visual\_Text\_Hierarchy\_w\_Postprocessing\_4\_Illustration.txt (mixed)*: Like Run M7 but classification is performed hierarchically.
- M9. *Image\_Text\_Hierarchy\_Entire\_set.txt (mixed)*: Like Run M6 but applied to all the images contained in the retrieval collection.

## 7.2 Ad-hoc Image Retrieval Task

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

- A1. *nlm-se (mixed)*: A combination of three 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 enriched citation; 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. (A1.Q3) A disjunction of the query images’ cluster words must occur within the global image feature field.
- A2. *nlm-se-cw-mf (mixed)*: A combination of Query A1.Q1 with the additional query below using Essie. (A2.Q2) A lossy expansion of the verbatim topic is allowed occur in any field of an image’s enriched citation and a disjunction of the query images’ cluster words can optionally occur within the global image feature field. Additionally, the retrieved images are filtered so that they share a least common ancestor modality with the query images, as determined by the modality classifier used in Run M9. Query A2.Q2 is distinct from Queries 1.Q2–3 in that the occurrence of a lossy expansion of the topic is not necessarily weighted more heavily than the occurrence of image cluster words.
- A3. *nlm-se-scw-mf (mixed)*: Like Run A2 but image cluster words are only considered if the modality classifier used in Run M9 identically labels all the example images of a topic.
- A4. *nlm-lc (mixed)*: A combination of three queries using Lucene with BM25 similarity and UMLS synonymy. (A4.Q1) A fuzzy phrase-based occurrence of the verbatim topic is allowed in any field of an image’s enriched citation. (A4.Q2) A disjunction of the topic words is allowed to occur in any field. (A4.Q3) A disjunction of the query images’ cluster words must occur within the global image feature field.
- A5. *nlm-lc-cw-mf (mixed)*: A combination of Query A4.Q1 with the additional query below using Lucene with BM25 similarity and UMLS synonymy. (A5.Q2) A disjunction of the topic words is allowed occur in any field of an image’s enriched citation and a disjunction of the query images’ cluster words can optionally occur within the global image feature field. Additionally, the retrieved images are filtered so that they share a least common ancestor modality with the query images, as determined by the modality classifier used in Run M9.

- A6. *nlm-lc-scw-mf (mixed)*: Like Run A5 but image cluster words are only considered if the modality classifier used in Run M9 identically labels all the example images of a topic.
- A7. *Combined\_Selected\_Fileterd\_Merge (visual)*: Similarity matching using 62 min-max normalized attributes selected from a combined visual descriptor of 15 features (all descriptors in Table 1 except LCH and SCONCEPT). Retrieval is performed separately for each query image, and the retrieved results are filtered, according to the modality classifier used in Run M9, so that they share the top two modality levels with the query. Images are scored according to the query image resulting in the maximum score.
- A8. *Combined\_LateFusion\_Fileterd\_Merge (visual)*: Like Run A8 but similarity matching is performed separately for seven features (CLD, GLCM, SCONCEPT, and the color, Gabor, shape, and Tamura moments from Table 1) whose scores are linearly combined with predefined weights.
- A9. *Txt\_Img\_Wighted\_Merge (mixed)*: A combination of visual Run A7 with a textual run consisting solely of Query A1.Q2 using score merging.
- A10. *Merge\_RankToScore\_weighted (mixed)*: A combination of visual Run A8 with a textual run consisting solely of Query A1.Q2 using rank mering.

### 7.3 Case-based Article Retrieval Task

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

- C1. *nlm-se-max (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 enriched citation; 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 and discovered diseases in the sentence is allowed to occur in any field. Articles are scored according to the sentence resulting in the maximum score.
- C2. *nlm-se-sum (textual)*: Like Run C1 but articles are scored according to the sum of the scores produced for each sentence.
- C3. *nlm-se-frames-max (textual)*: A combination of the query below with Query C1.Q2 for each topic sentence using Essie. (C3.Q1) An expansion of the frame-based representation of the sentence is allowed to occur in any field of an article’s enriched citation. Articles are scored according to the sentence resulting in the maximum score.
- C4. *nlm-se-frames-sum (textual)*: Like Run C3 but articles are scored according to the sum of the scores produced for each sentence.
- C5. *nlm-lc-max (textual)*: A combination of two queries for each topic sentence using Lucene with language model similarity, Jelinek-Mercer smoothing, and UMLS synonymy. (C5.Q1) A fuzzy phrase-based occurrence of the verbatim sentence is allowed in any field of an article’s enriched citation. (C5.Q2) A disjunction of all words and discovered diseases in the sentence is allowed to occur in any field. Articles are scored according to the sentence resulting in the maximum score.
- C6. *nlm-lc-sum (textual)*: Like Run C5 but articles are scored according to the sum of the scores produced for each sentence.



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

File Name	Mode	Accuracy (%)
<i>Visual_Text_Hierarchy_w_Postprocessing_4_Illustration.txt</i>	Mixed	63.2
<i>Visual_Text_Flat_w_Postprocessing_4_Illustration.txt</i>	Mixed	61.7
<i>Visual_Text_Hierarchy.txt</i>	Mixed	60.1
<i>Visual_Text_Flat.txt</i>	Mixed	59.1
<i>Visual_only_Hierarchy.txt</i>	Visual	51.6
<i>Visual_only_Flat.txt</i>	Visual	50.3
<i>Image_Text_Hierarchy_Entire_set.txt</i>	Mixed	44.2
<i>Text_only_Hierarchy.txt</i>	Textual	41.3
<i>Text_only_Flat.txt</i>	Textual	39.4

- C7. *nlm-lc-total-max (textual)*: A combination of the query below with Queries C5.Q1–2 (as C7.Q2–3) using Lucene with language model similarity, Jelinek-Mercer smoothing, and UMLS synonymy. (C7.Q1) A fuzzy phrase-based occurrence of the entire verbatim topic is allowed in any field of an article’s enriched citation. Articles are scored according to the sentence resulting in the maximum score.
- C8. *nlm-lc-total-sum (textual)*: Like Run C7 but articles are scored according to the sum of the scores produced for each sentence.

## 8 Results

We present and discuss the results of our modality classification, ad-hoc image retrieval, and case-based article retrieval task submissions below.

### 8.1 Modality Classification Task

Table 2 presents the classification accuracy of our submitted runs for the modality classification task. *Visual\_Text\_Hierarchy\_w\_Postprocessing\_4\_Illustration.txt*, a mixed approach, achieved the highest accuracy (63.2%) of our submitted runs and was ranked fifth overall. However, it ranked within the submissions from the top three participating groups. This result validates our post-processing method used to improve the recognition of “Illustration” classes, and provides, with our previous experience [17], further evidence that hierarchical classification is a successful strategy. Each of our hierarchical classification methods outperforms the corresponding flat approach having the same feature representation.

While our submitted runs were only judged on their ability to identify each of the thirty-one modality classes [15], Table 3 presents the classification accuracy of the intermediate classifiers we used for our hierarchical approaches. For each meta-class in the hierarchy shown in Figure 1b, Table 3 gives the number of classes they contain; the classification accuracy associated with the textual, visual, and mixed feature representations; and the dimensionality of the mixed feature representation after attribute selection. These results demonstrate that the accuracies of the intermediate classifiers generally improve as the number of class labels decreases. Given the limited amount of training data in relation to the number of total modalities, the smaller number of labels per classifier likely is significant for explaining why our hierarchical classification approaches consistently outperform their corresponding flat approaches.

**Table 3:** Accuracy results for our intermediate modality classifiers.

ID	Number of Classes	Feature Accuracy (%)			Dims.*
		Mixed	Visual	Textual	
1	2 (Illustration, General)	96.3	95.6	78.6	159
2	3 (Radiology_3D, Microscopy, Photo)	93.5	87.4	83.8	112
3	8 (DRUS, DRMR, . . . , D3DR)	75.9	64.4	71.3	89
4	4 (DMLI, DMEL, . . . , DMFL)	85.0	83.6	69.4	58
5	4 (DVDM, DVEN, . . . , GNCP)	77.6	62.3	89.1	108
6	14 (GTAB, GPLI, . . . , DSEM)	63.5	53.0	41.2	69

\* Feature dimensionality is given for mixed mode classifiers only.

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

ID	Mode	MAP	bpref	P@10
<i>nlm-se</i>	Mixed	0.2377	0.2542	0.3682
<i>Merge_RankToScore_weighted</i>	Mixed	0.2166	0.2198	0.3682
<i>nlm-lc</i>	Mixed	0.1941	0.1871	0.2727
<i>nlm-lc-cw-mf</i>	Mixed	0.1938	0.1924	0.2636
<i>nlm-lc-scw-mf</i>	Mixed	0.1927	0.1940	0.2636
<i>nlm-se-scw-mf</i>	Mixed	0.1914	0.2062	0.2864
<i>Txt_Img_Wighted_Merge</i>	Mixed	0.1846	0.2039	0.3091
<i>nlm-se-cw-mf</i>	Mixed	0.1774	0.1868	0.2909
<i>Combined_LateFusion_Fileterd_Merge</i>	Visual	0.0046	0.0107	0.0318
<i>Combined_Selected_Fileterd_Merge</i>	Visual	0.0009	0.0028	0.0227

## 8.2 Ad-hoc Image Retrieval Task

Table 4 presents the mean average precision (MAP), binary preference (bpref), and early precision (P@10) of our submitted runs for the ad-hoc image retrieval task. *nlm-se* achieved the highest MAP (0.2377) among our submitted runs and was ranked first overall. *Merge\_RankToScore\_weighted*, the run achieving our second highest MAP (0.2166), was ranked second overall. Comparing these two runs using Fisher’s paired randomization test [18], a recommended statistical test for evaluating information retrieval systems, we find that *nlm-se* achieved a statistically significant increase (9.7%,  $p = 0.0016$ ) over the performance of *Merge\_RankToScore\_weighted*.

That the two highest ranked runs were multimodal, as apposed to textual, is an encouraging result, and provides evidence that our ongoing efforts at integrating textual and visual information will be successful. In particular, the use by *nlm-se* and other runs of cluster words, which are indexed and retrieved using a traditional text-based information retrieval system, is an effective way, not only of incorporating visual information with text, but of avoiding the computational expense common among content-based retrieval methods. Furthermore, *Merge\_RankToScore\_weighted* demonstrates the value of rank merging when combining textual and visual retrieval results. Some of our other mixed runs, in utilizing the results of our modality classifiers, may have been weakened due to the modest performance of our classification methods.

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

ID	Mode	MAP	bpref	P@10
<i>nlm-lc-total-sum</i>	Textual	0.1035	0.1053	0.1000
<i>nlm-lc-total-max</i>	Textual	0.1027	0.1055	0.0923
<i>nlm-se-sum</i>	Textual	0.0929	0.0738	0.0769
<i>nlm-se-max</i>	Textual	0.0914	0.0736	0.0769
<i>nlm-lc-sum</i>	Textual	0.0909	0.0933	0.1231
<i>nlm-lc-max</i>	Textual	0.0840	0.0886	0.0923
<i>nlm-se-frames-sum</i>	Textual	0.0771	0.0693	0.0692
<i>nlm-se-frames-max</i>	Textual	0.0672	0.0574	0.0538

### 8.3 Case-based Article Retrieval Task

Table 5 presents the MAP, bpref, and P@10 of our submitted runs for the case-based article retrieval task. *nlm-lc-total-sum*, a textual approach using language model similarity, achieved the highest MAP (0.1035) among our submitted runs and was ranked seventh overall. However, it ranked within the submissions from the top four participating groups. Using Fisher’s paired randomization test, we find that there is no statistically significant difference ( $p < 0.05$ ) in MAP among any of our submitted runs. The relatively low performance of most of the ImageCLEF 2012 case-based submissions may be due, in part, to the existence in the collection of only a small number of case reports, clinical trials, or other types of documents relevant for case-based topics.

## 9 Conclusion

This article describes the methods and results of the Image and Text Integration (ITI) group in the ImageCLEF 2012 medical retrieval and classification tasks. For the modality classification task, our best submission was ranked within the submissions from the top three participating groups. Our best submission for the ad-hoc image retrieval task was ranked first overall. Finally, for the case-based article retrieval task, our best submission was ranked within the submissions from the top four participating groups, though we found no statistical significance between this run and our other case-based submissions. The effectiveness of our multimodal 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 Antonio Jimeno-Yepes for assisting in expanding case-based topics with disease names, Russell Loane for proving source code for converting frame-based topics to Essie queries, and Srinivas Phadnis for constructing enriched citations and extracting visual features.

## References

1. Apostolova, E., You, D., Xue, Z., Antani, S., Demner-Fushman, D., Thoma, G.: Image retrieval from scientific publications: Text and image content processing to separate multi-panel figures. *Journal of the American Society for Information Science and Technology* (To appear)

2. Aronson, A.R.: Effective mapping of biomedical text to the UMLS metathesaurus: The MetaMap program. In: Proc. of the Annual Symp. of the American Medical Informatics Association (AMIA). pp. 17–21 (2001)
3. Arthur, D., Vassilvitskii, S.: k-means++: The advantages of careful seeding. In: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms. pp. 1027–1035. SODA '07 (2007)
4. 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)
5. 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)
6. 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)
7. Demner-Fushman, D., Lin, J.: Answering clinical questions with knowledge-based and statistical techniques. *Computational Linguistics* 33(1), 63–103 (Mar 2007)
8. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA data mining software: An update. *SIGKDD Explorations* 11(1) (2009)
9. 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)
10. Jung, K., Kim, K.I., Jain, A.K.: Text information extraction in images and video: A survey. *Pattern Recognition* 37(5), 977–997 (2004)
11. Lindberg, D., Humphreys, B., McCray, A.: The unified medical language system. *Methods of Information in Medicine* 32(4), 281–291 (1993)
12. 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)
13. 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)
14. Mäenpää, T.: The Local Binary Pattern Approach to Texture Analysis—Extensions and Applications. Ph.D. thesis, University of Oulu (2003)
15. Müller, H., de Herrera, A.G.S., Kalpathy-Cramer, J., Demner-Fushman, D., Antani, S., Eggel, I.: Overview of the ImageCLEF 2012 medical image retrieval and classification tasks. In: CLEF 2012 Working Notes (2012)
16. 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)
17. Simpson, M., Rahman, M.M., Phadmis, S., Apostolova, E., Demner-Fushman, D., Antani, S., Thoma, G.: Text- and content-based approaches to image modality classification and retrieval for the ImageCLEF 2011 medical retrieval track (2011)
18. Smucker, M.D., Allan, J., Carterette, B.: A comparison of statistical significance tests for information retrieval evaluation. In: Proceedings of the Sixteenth ACM Conference on Information and Knowledge Management. pp. 623–632 (2007)
19. Srinivasan, G.N., G., S.: Statistical texture analysis. In: Proceedings of World Academy of Science, Engineering and Technology. vol. 36, pp. 1264–9 (2008)
20. Tamura, H., Mori, S., Yamawaki, T.: Textural features corresponding to visual perception. *IEEE Transactions on Systems, Man, and Cybernetics* 8(6), 460–73 (1978)