

# Towards the Improvement of Textual Anatomy Image Classification using Image Local Features

Xiao Bing Huang  
Tian Zhao  
Department of Computer  
Science  
University of  
Wisconsin–Milwaukee  
Milwaukee, WI 53201  
{xiaobing,tzhao}@uwm.edu

Yu Cao  
College of Engineering &  
Computer Science  
University of  
Tennessee–Chattanooga  
Chattanooga, TN 37403  
Yu-Cao@utc.edu

Xiangming Mu  
Pierre Tirilly  
School of Information Studies  
University of  
Wisconsin–Milwaukee  
Milwaukee, WI 53201  
{mux,tirilly}@uwm.edu

## ABSTRACT

Image classification methods based on text utilize terms extracted from image annotations (image caption, image-related article, etc.) to achieve classification. For images involving different anatomical structures (chest, spine, etc.), however, the precision of pure textual classification often suffers from highly complex text contents (e.g. text terms extracted out of two MR abdomen images may be quite different from each other: terms from one image may concern gastroenteritis while the other contains terms involving hysteromyoma). This paper tackles the anatomy image classification problem using a hybrid approach. First, a mutual information (MI) based filter is applied to select a set of terms with top MI scores for each anatomical class and help reduce the noise existing in the raw text. Second, local features extracted from the images are transformed as visual descriptors. Last, a hybrid scheme on the results from the textual and visual methods is applied to achieved further improvement of the classification results. Experiments show that this hybrid scheme improves the results over the sole textual or visual method on different anatomical class settings.

## Categories and Subject Descriptors

H.5.1 [Multimedia Information Systems]: Evaluation/methodology; I.1.2 [Computing Methodologies]: Algorithms—*Analysis of algorithms*; I.4.10 [IMAGE PROCESSING AND COMPUTER VISION]: Image Representation—*Multidimensional*

## General Terms

Algorithms, Experimentation, Performance

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MMAR'11, November 30, 2011, Scottsdale, Arizona, USA.  
Copyright 2011 ACM 978-1-4503-0991-2/11/11 ...\$10.00.

## Keywords

Textual feature selection, local image features, clustering, mutual information, anatomy classification

## 1. INTRODUCTION

Advances in sensor technology, high-speed networking, and massive digital storage are being incorporated into today's health care practices and biomedical researches. Tremendous amounts of biomedical image data are captured and recorded in digital format daily. These images can be categorized based on the anatomical parts they depict, such as abdomen, lower-extremity (other categorization methods can also be applied, e.g. the modality under which the images were acquired or the pathology observed in the image). Being able to automatically know the anatomy information of the image provides huge benefits in clinical practices. The radiologists can focus on only images of the anatomy parts that interest them; the internal/online medical image anatomy category can be created/updated in a much faster fashion; the pathological analysis can be more efficient and precise when the anatomical features of the image for a patient are provided; such information will also help greatly when combined with other categorizations (modality, pathology, etc.) to provide an integrated medical image retrieval system.

Despite the significant amount of work that has been put in anatomical part classification, many problems remain unsolved due not only to lack of high-quality anatomy-labeled image collections for analysis, training and evaluation of classification systems, but also to the quality of the available text information coming with these images. Most of the existing medical image classification systems rely mainly on textual information or patient identifications. However, the textual information coming with the images usually contains very few terms related to anatomical parts, making it noisy training data for anatomical part classifiers.

This paper addresses the issues at two different levels. First, the paper presents a highly efficient filter based on mutual information (MI) [26, 27] to select terms that are highly discriminant for anatomical part classification. Only these terms are used to perform the classification. Second, since some "common terms" of low discriminative power may still remain after the first filtering step (even if they only exist in a small subset of the total anatomical classes) and cause some medical images continuously to be classified incorrectly

with the sole textual method, image local features are used to overcome this drawback based on the fact that the visual features (or more specifically, the image local components that form the images’ main structure) of many images from different anatomy categories are quite different against each other while the textual components of these images appear to share a great portion of overlapping. Our experiments show that this hybrid method further improves the textual classification results universally on different combinations of anatomical classes.

## 2. RELATED WORK

Past researches have developed several content-based retrieval systems for images of specific modalities and anatomies. Examples include systems for mammograms [1], high-resolution CT lung images [8] and spine images [13]. There are also a few retrieval and classification systems for general medical images such as MedGIFT [19]. However, visual-based retrieval systems have generally perform poorly compared to textual approaches [18] for general medical images acquired under different modalities and orientations, and depicting different anatomical structures.

Automatic medical image classification can not only improve the performance of image retrieval systems, but also provide image annotations and correct errors in the *Digital Imaging and Communications in Medicine* (DICOM) headers. Starting from a few years ago, the ImageCLEF competition [10] includes tasks to annotate medical images based on their modalities, body orientations, body regions or biological systems. The training and test images provided for these tasks are based on the IRMA project’s X-Ray images annotated by 57 to 196 labels. Some of the best performing runs of these competitions are based on visual bag-of-words (BoW) models [21], which are adapted from the BoW models commonly used in text retrieval [17]. Visual BoW models have been successfully applied to areas such as learning natural scene categories [12] and object matching in videos [21]. The visual words can be based on local image features such as Scale-Invariant Feature Transform (SIFT) descriptors [14] or grayscale patches (Here visual words refer words that represent the visual features in text format). For example, Ginneken et al. [25] achieve good results in the ImageCLEF annotation task with a mixture of global features and SIFT-based visual word histograms. Patch-based BoW models are used to classify endomicroscopic images [2] and breast tissue density in mammograms [4]. They are also effective in X-ray image classification [7, 3]. The results reported by Avni *et al.* [3] suggest that patch-based approaches can achieve similar accuracies in classifying X-ray images as SIFT-based BoW models, at lower computational costs.

While the classification of medical images based on IRMA codes has been successful for X-Ray images (e.g. SVM based schemes in [22, 24]), the results may not be generally applicable to medical images of other modalities or when images cannot be clearly classified with an IRMA code. It may be advantageous to develop classification models separately for modality and anatomy. Modality classification can be used to improve the precision of medical image retrieval systems [11]. Lightweight image descriptors can be combined with textual methods to achieve good precisions for image modality classification [23]. This work follows a similar direction by determining whether integrating visual descrip-

tors with textual descriptors can improve the performance of anatomy-based image classifiers.

## 3. PROPOSED APPROACH

This section describes the MI-based feature selection method employed, then the visual descriptors used in the experiments, and, finally, how textual and visual features are integrated into a hybrid classification scheme.

### 3.1 Textual method

#### 3.1.1 Mutual information

Given two random variables, MI measures how knowing one variable reduces the uncertainty of the other. When applied to text-based classification, MI measures the contribution of term  $t$  (by its presence or its absence) to the description of class  $c$ . Equation (1) gives the formal definition of MI for two binary random variables  $T$  and  $C$ :

$$\text{MI}(T, C) = \sum_{e_t \in \{1,0\}} \sum_{e_c \in \{1,0\}} P(T = e_t, C = e_c) \log_2 \frac{P(T = e_t, C = e_c)}{P(T = e_t)P(C = e_c)} \quad (1)$$

where  $T$  is a random variable that corresponds to the presence ( $e_t = 1$ ) or absence ( $e_t = 0$ ) of term  $t$ , and  $C$  is a random variable that corresponds to the presence ( $e_c = 1$ ) or absence ( $e_c = 0$ ) of class  $c$ , i.e. the fact that the document containing  $t$  belongs or not to  $c$ . With maximum likelihood estimations (MLEs) of the probabilities, Equation (1) is equivalent to Equation (2) [16]:

$$\text{MI}(T, C) = \frac{N_{11}}{N} \log_2 \frac{N N_{11}}{N_{1.} N_{.1}} + \frac{N_{01}}{N} \log_2 \frac{N N_{01}}{N_{0.} N_{.1}} + \frac{N_{10}}{N} \log_2 \frac{N N_{10}}{N_{1.} N_{.0}} + \frac{N_{00}}{N} \log_2 \frac{N N_{00}}{N_{0.} N_{.0}} \quad (2)$$

where  $N_{e_t e_c}$  is the count of documents so that  $T = e_t$  and  $C = e_c$ . For example,  $N_{01}$  is the number of documents that contain  $t(e_t = 0)$  and are not in  $c(e_c = 1)$ ;  $N_{0.} = N_{00} + N_{01}$  is the number of documents that contain  $t(e_t = 0)$  regardless the class membership ( $e_c \in \{0, 1\}$ ); and  $N = N_{00} + N_{01} + N_{10} + N_{11}$  is the total number of documents. By computing the value MI for each term of the collection and each class, it is then possible to obtain, for each class, the links of terms with their contribution to the description of the class.

#### 3.1.2 Selection of top terms

As stated in the introduction, the textual information extracted from image annotations usually contains a large number of noisy/useless terms for anatomical part classification, even after removing all the stop words. MI can help to identify such terms: terms  $t$  that appear multiple times in the annotations of the images from a single class  $c$  will tend to have high  $\text{MI}(T, C)$  values, while terms that appear multiple times in the annotations of the images from multiple classes will have a low  $\text{MI}(T, C)$  values. Therefore, those terms with high MI values (more discriminative) will be more likely to identify an image’s anatomical class label correctly than the terms with low MI values (less discriminative). We can hence greatly reduce the noise by limiting our textual descriptors to the terms that have a top MI value for at least one class. How many top terms we should use depends on the size of the training dataset and the number

of anatomical classes with which we are working. Moreover, since terms in the top term list of one class are likely to be quite different from those of another class, we should select top terms from each of the classes evenly so that the final term set we obtained has an equal discriminative effect on each class. We will study the results of different numbers of top terms in the experiments.

### 3.2 Visual feature-based method

As indicated in [20], if we describe a 2D image based on its entire pixel set, we will end up with too much information. Descriptors based on local features, by opposition, describe only regions of interest, i.e. regions that contain significant information (changes in the 2D signal). This is especially useful for anatomical images since many of them only have partial visibility and contain highly informational content. A good local feature should be highly distinctive and reasonably invariant to image noise, illumination, scale and rotation changes, as well as slight changes in viewpoint. *Scale Invariant Feature Transform* (SIFT) [14] features meet these requirements and are adopted in this paper for the visual classification step. Our visual part of the hybrid method is hinted by the BoW model, which shows excellent performance in image classification [5]. To apply the BoW model, we extract SIFT features for a given set of images, then apply a clustering algorithm on these features. We finally calculate for each image the number of its features falling in each cluster. The resultant histogram contains a set of “visual words” that are used to describe the image. This histogram is deemed as the basis for similarity comparison in later visual classification.

### 3.3 Hybrid scheme

We combine the text-based and visual-based classifier using a linear combination as follows:

$$S_H = W_T * S_T + (1 - W_T) * S_V \quad (3)$$

where  $S_T$  is the classification score returned by the text-based classifier,  $S_V$  is the score returned by the visual-based classifier and  $S_H$  is the hybrid score obtained from the combination.  $W_T$  is a parameter that controls the relative importance of each classifier that must be set up manually or optimized on a validation set. Its value will be discussed in the next section.

## 4. EXPERIMENTS

### 4.1 Data organization

In this section, we describe the data we used and the experimental settings we adopted [6]. The dataset contains 909 anatomy images. It comes from two different resources: 534 images are selected from the ImageCLEF2010 dataset and 375 images come from a semi-automatic selection process, i.e. retrieved from publicly available medical databases, e.g. RSNA and EURORAD, and followed by a manual screening process.<sup>1</sup> All 909 images are labeled by one of 8 anatomical classes. The class names and the number of images in each

<sup>1</sup>This combination process is due to lack of existing anatomical classification datasets in ImageCLEF. Also, we cannot directly perform experiments on the IRMA datasets either. This is mainly because no textual annotations have been provided there and the IRMA datasets do not contain enough images in some categories (e.g. cranium images)

Anatomical class	# of images
abdomen	136
chest	143
cranium	144
lower-extremity	124
pelvis	76
spine	148
upper-extremity	45
whole-body	93

Table 1: The anatomy data summary

Dataset	Anatomical classes
4 classes	chest, cranium, lower-extremity, spine
6 classes	chest, cranium, lower-extremity, spine upper-extremity, whole-body
8 classes	chest, cranium, lower-extremity, spine upper-extremity, whole-body, pelvis abdomen

Table 2: The anatomy types for each experiment.

class is summarized in Table 1. We split the dataset into a training set and a test set on a 7:3 ratio. Since our goal is to examine the improvement effect of the hybrid method with different precisions of the sole textual method, we designed three experiments with three different combinations of classes (see Table 2) where the classification precisions by the sole textual method decrease while more classes are added. This decrease in precision is due to the fact that the newly added anatomical classes contain more noise than others, e.g. the whole-body class contains more noisy terms than other classes.

### 4.2 Textual classification steps

In each experiment, we created a term score file per anatomy type (from the training dataset after removing the stop words) which gives the list of terms in the descending order of their MI scores. We then performed an incremental process which selects the top  $N$  terms (the first  $N$  terms with the highest MI scores) from each score file per anatomy type in each loop where  $N = 5 \dots 100$  with step value of 2<sup>2</sup>. For example, in the case of 4 anatomical classes, we started from  $5 \times 4 = 20$  top terms and went all the way up to  $100 \times 4 = 400$  top terms. Again, terms with high MI values will be more likely to identify an image’s anatomical class label correctly than the terms with low MI values. e.g. from our experiments, the terms with top five MI scores in the cranium anatomical class are: cranial, fossa, maxillary, skull, sinus, while the

that we concern about. Furthermore, IRMA datasets consist mainly of X-ray images and the lack of other modality information (e.g. nuclear imaging) may cause the classification work more restrictive and less practical.

<sup>2</sup>From our experiment observation, the textual feature based method gives low enough precision when more than 100 features are used such that it is not of interest any more whether the visual feature based method can improve its performance (e.g. a pretty impressive improvement of 50% via adding the visual method on a textual method-based precision 40% would be 60% and is still too low to be worth any further investigation).

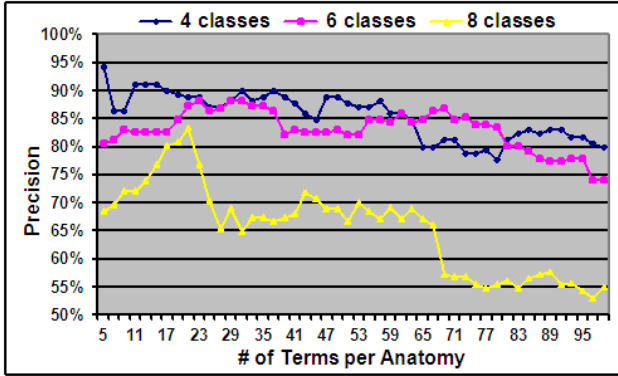


Figure 1: The effect of the number of top terms per anatomy on the classification precision.

terms with top five MI scores in the lower-extremity category are: tibial, extremity, lower, popliteal, lower-extremity. Clearly, How frequent such terms appear in an image’s annotation provides strong indication of the anatomical class of the image. Based on these top terms, we construct two class files, for training set and test set respectively, per anatomy type by removing terms that are not in the top term list. Finally, we count the total number of each term (which gives the frequency of the term appearing in the image’s annotation) from the reduced annotation in the class files and use these numbers as the final text feature values for classification. We applied a SVM-based classifier, the Weka SMO classifier [9], programmatically throughout the classification tasks. Experiment results for the three combinations of anatomical classes from Table 2 are illustrated in Figure 1. We can easily see that, for all three experiments, the total trend of classification precisions starts increasing while more top terms were used due to more strong (highly discriminative) terms work for classification and reaches some peak value; then decreases while more weak terms were added. We also observe that the good number of top terms (to help achieve high precisions) used per anatomy for all experiments are roughly in the range [9,69] and the best precisions were achieved with [15,25] top terms per anatomy.

### 4.3 BoW on image local features

We compute the visual BoW of all images following the method described in Section 3.2. We use a modified version of k-means from the Lire library [15] to group the SIFT features (extracted with the Lire SIFT builder) into 256 clusters. The results of applying the visual method on the different combinations of classes from Table 2 are listed in the visual column in Table 3 (i.e. 87.65%, 79.72%, and 63.77% for 4, 6, and 8 anatomical classes respectively). Now we want to combine the two methods to achieve better classification results, following the method presented in Section 3.3. We apply a weighted average method as shown in Equation-3 to combine the textual and visual methods with the weight slightly favor the textual method since it provides higher precision than what is achieved by the visual method in many cases. The motivation is from our observation that the local features of many images can be used to identify their anatomical class with better precision in the case that top text terms extracted from these images are not so dis-

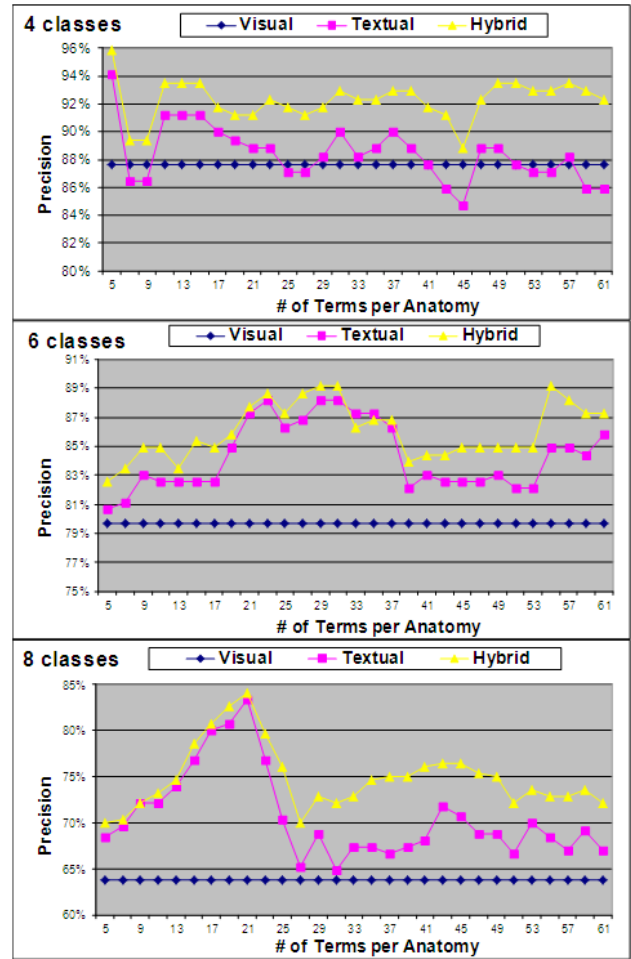


Figure 2: The improvement achieved by the hybrid method comparing to the results of the textual & visual method with different number of top terms per anatomy used.

tinctive for classification. In other words, the information provided by the textual method can be strengthened with the information provided by the visual method.

The WEKA SMO SVM implementation provides the classification scores  $S_T$  and  $S_V$  respectively obtained by the text-based and visual-based classification. Using Equation 3, the value of  $S_H$  is computed for each class and the class that maximizes  $S_H$  is assigned to the current test image.

It must be noted that there is not any ”best-for-all” weight value  $W_T$  that can be used for all datasets. Indeed, the optimal value of  $W_T$  relies on the relative performance of the two classifiers that may change on datasets of different complexities. Ideally, this parameter should then be optimized on a validation set to ensure optimal results. However, this is difficult in practice, given the limited amount of data available. Therefore, we just tested different values of  $W_T$  in the range [0.5, 0.99]. In the remaining of the experiments, we report only the results observed with the optimal value of  $W_T$  that we found: 0.51 for the 4-classes dataset, 0.67 for the 6-classes one and 0.76 for the 8-classes one.

### 4.4 Results and discussion

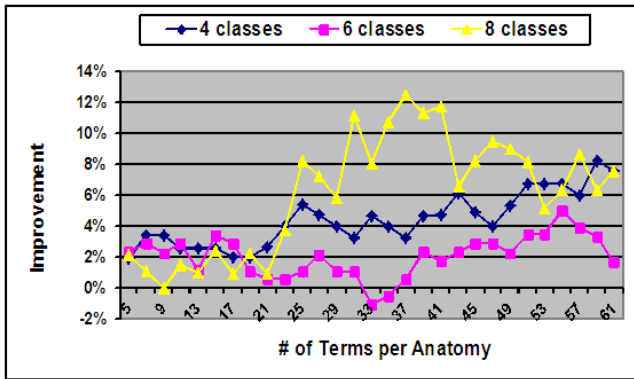


Figure 3: The improvement in % achieved by the hybrid method over the results of the textual method with different number of top terms per anatomy used.

Anatomical classes	Weight	Terms	Textual	Visual	Hybrid
4	0.51	15	91.18%	87.65%	93.53%
6	0.67	23	88.21%	79.72%	88.68%
8	0.76	21	83.33%	63.77%	84.06%

Table 3: Effect of hybrid method on the best results achieved by the textual method solely.

We report the classification results we obtained against different combination of anatomical classes in this section. Before we move forward, there are two things we want to examine when we test the hybrid method.

1. Can the hybrid method improve the best result from the sole textual method?
2. Can the hybrid method greatly improve a relatively low precision achieved when the textual method is applied solely?

The first one concerns the question of whether the visual method can break the bottleneck faced by the textual methods while the second one involves a more important question of whether the visual method can compensate positively the poor results obtained by the sole textual method due to the unavailability or bad quality of the image text information. To answer the above two questions, we designed two different set of experiments. In the first set of experiments, we combined the results of the visual method with the best results from the textual method; in the second set of experiments, we combined the results of the visual method with some poor classification precisions achieved by the textual method. For the first set of experiments, we recorded the new precisions achieved by the hybrid method and summarized the results in Table 3 where the “Weight” column provides the number of top terms per anatomy with which the best result was achieved by the textual method. For the second set of experiments, we kept the same weight values 0.51, 0.67, and 0.76 for 4, 6, and 8 anatomical classes cases, respectively. We then illustrated the improvements (in percentage format) achieved by the hybrid method over the textual method when different number of top terms per

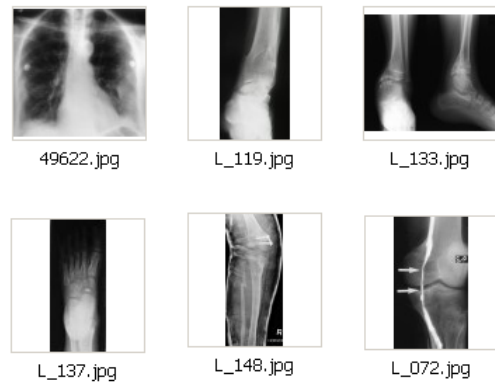


Figure 4: The images from Table 4.

anatomy type are used. The results are summarized in Figure 2 and Figure 3. We can see from the figures that the hybrid method almost universally improved the results from the textual method. In fact, except one zero improvement case in the 8 anatomical classes experiment and two slight drops of -1.08% and -0.54% on the precisions in the 6 anatomical classes case, the rest are all positive improvements. Also, we observe that with some reasonable number of top terms per anatomy (5~61, etc.) selected, the results of the hybrid method outperforms both the textual and visual methods. Another observation from our experiments is that the incorrectly classified images by the textual method are quite different from the ones by the visual method. This is why a good hybrid scheme can help improve the classification precision. Table 4 illustrates how the visual method further improved the best result from the sole textual method in the 4 anatomical classes case (The images are displayed in Figure 4). We can see that the hybrid method successfully corrects the misclassification for the first five images while misclassified only the last image where the sole textual method did correctly. This error occurs for two reasons. First, the image is a close-up of a patient’s kneecap and has some “local features” that also appear in the cranium images frequently; second, the textual method gives a second highest probability to cranium type even the highest probability is given to the correct anatomical class lower-extremity; this causes the weighted average votes cranium as the result.

## 5. CONCLUSIONS

We have presented a new hybrid method for the classification of medical anatomy images in this paper. The textual part of this hybrid method removes poorly distinctive terms from the annotations to achieve a higher classification precision while the visual part further improves the results from the textual method using a SIFT-based visual BoW model. Experiments show strong improvements on datasets of medical images containing up to 8 anatomical classes. Since the textual and visual classifier can be trained separately, we next plan to apply the IRMA images directly to train the visual classifier and examine the classification effect of the resultant classifier on the textual method.

## 6. REFERENCES

- [1] H. Alto, R. Rangayyantt, and J. Desautels.

Image ID	Terms	Textual /Hybrid	True anatomy
49622	radiograph,distal	cranium /chest	chest
L_119	radiograph,left, distal	cranium /l-extremity	l-extremity
L_133	showing	cranium /l-extremity	l-extremity
L_137	left	cranium /l-extremity	l-extremity
L_148	radiograph,distal, femoral,tibial	cranium /l-extremity	l-extremity
L_072	tibial,obtained, images,left, popliteal,femoral	l-extremity /cranium	l-extremity

**Table 4: The improvement of the hybrid over the textual method where “Terms” column contains the top terms left for the image after filtering (l-extremity is short for lower-extremity).**

- Content-based retrieval and analysis of mammographic masses. *Journal Of Electronic Imaging*, 14(2), 2005.
- [2] B. Andr, T. Vercauteren, A. Perchant, A. Buchner, M. Wallace, and N. Ayache. Introducing space and time in local feature-based endomicroscopic image retrieval. *Medical Content-Based Retrieval for Clinical Decision Support*, pages 18–30, 2010.
- [3] U. Avni, H. Greenspan, E. Konen, M. Sharon, and J. Goldberger. X-ray categorization and retrieval on the organ and pathology level, using patch-based visual words. *IEEE Transactions on Medical Imaging*, 30(3), 2011.
- [4] A. Bosch, X. Munoz, A. Oliver, and J. Marti. Modeling and classifying breast tissue density in mammograms. In *CVPR*, volume 2, pages 1552–1558, 2006.
- [5] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray. Visual categorization with bags of keypoints. In *European Conference on Computer Vision (ECCV) Workshop on Statistical Learning in Computer Vision*, 2004.
- [6] CS@UWM. Software and dataset. <http://guangzhou.cs.uwm.edu/med>, 2011.
- [7] T. Deselaers, A. Hegerath, D. Keysers, and H. Ney. Sparse patch histograms for object classification in cluttered images. In *DAGM Symposium*, pages 202–211, 2006.
- [8] J. Dy, C. Brodley, A. Kak, L. Broderick, and A. Aisen. Unsupervised feature selection applied to content-based retrieval of lung images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(3):373–378, 2003.
- [9] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The weka data mining software: An update. *SIGKDD Explorations*, 11(1), 2009.
- [10] ImageCLEF. Image retrieval in clef: Medical retrieval task. <http://www.imageclef.org/2010/medical>, 2010.
- [11] J. Kalpathy-Cramer and W. Hersh. Automatic image modality based classification and annotation to improve medical image retrieval. In *Proceedings of the 12th World Congress on Health (Medical) Informatics (MEDINFO)*, pages 1334–1338, 2007.
- [12] F.-F. Li and P. Perona. A bayesian hierarchical model for learning natural scene categories. In *CVPR*, volume 2, pages 524–531, 2005.
- [13] L. R. Long, S. Antani, D.-J. Lee, D. M. Krainak, and G. R. Thomas. Biomedical information from a national collection of spine x-rays: film to content-based retrieval. In *Proceedings SPIE*, pages 70–84, May 2003.
- [14] D. Lowe. Object recognition from local scale-invariant features. In *ICCV*, volume 2, 1999.
- [15] M. Lux and S. A. Chatzichristofis. Lire: lucene image retrieval: an extensible java chbir library. In *Proceeding of the 16th ACM international conference on Multimedia*, MM ’08, pages 1085–1088, New York, NY, USA, 2008. ACM.
- [16] C. Manning, P. Raghavan, and H. Schuze. *Introduction to Information Retrieval*. Cambridge University Press, 2009.
- [17] C. D. Manning, P. Raghavan, and H. Schutze. *Introduction to Information Retrieval*. Cambridge University Press, July 2008.
- [18] H. Muller, J. Kalpathy-Cramer, I. Eggel, S. Bedrick, J. Reisetter, C. E. K. Jr., and W. Hersh. Overview of the CLEF 2010 medical image retrieval track. In *Working notes of the Image-CLEF 2010 challenge*, 2010.
- [19] H. Muller, C. Lovis, and A. Geissbuhler. The medgift project on medical image retrieval. *Medical Imaging and Telemedicine*, 2005.
- [20] C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 19(5), 1997.
- [21] J. Sivic and A. Zisserman. Video google: a text retrieval approach to object matching in videos. In *ICCV*, volume 2, pages 1470–1477, 2008.
- [22] T. Tatiana, O. Francesco, and C. Barbara. Discriminative cue integration for medical image annotation. *Pattern Recognition Letters*, 29(15):1996–2002, November 2008.
- [23] P. Tirilly, K. Lu, X. Mu, T. Zhao, and Y. Cao. 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 Indexing*, 2011.
- [24] T. Tommasi, F. Orabona, and B. Caputo. An svm confidence-based approach to medical image annotation. In *proceedings of the 9th CLEF workshop 2008*, 2008.
- [25] B. van Ginneken, L. Hogeweg, and M. Prokop. Computer-aided diagnosis in chest radiography: Beyond nodules. *European Journal of Radiology*, 72(2):226–230, 2009.
- [26] I. H. Witten and E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, Amsterdam, 2005.
- [27] Y. Y. Yao. Information-theoretic measures for knowledge discovery and data mining. *Entropy Measures, Maximum Entropy Principle and Emerging Applications*, pages 115–136, 2003.