

# BUAA AUDR at ImageCLEF 2012 Medical Retrieval Task

Wei Song, Danchen Zhang, Junwu Luo

Department of Computer Science and Technology, Beihang University,  
100191 Beijing, China  
songwei@nlsde.buaa.edu.cn  
zhangdanchen@nlsde.buaa.edu.cn  
luojunwu@nlsde.buaa.edu.cn

**Abstract.** This paper presents the participation of the BUAA AUDR group at ImageCLEF 2012 at the Medical Image classification and Retrieval task. We performed two subtasks: modality classification and ad-hoc image-based retrieval. It was our first time to select modality classification task and we concentrated on mono-modal visual-based image classifier. We used LibSVM to train the classifier, and edge histogram feature to represent images. To improve its performance, we tried to extend the training set. However, due to size of the training set and other reasons, its accuracy was even worse. For ad-hoc image-based retrieval, we utilized MeSH as source of query expansion and only textual information were considered. We also explored mixed approaches that combine modality predication and query expansion and our best runs ranked second among all the textual runs.

**Keywords:** ImageCLEF, medical image retrieval, MeSH, query expansion, Modality Classification

## 1 Introduction

This paper presents the second participation of the BUAA AUDR group at ImageCLEF.

ImageCLEF 2012 includes four types of tasks: medical image retrieval, photo annotation, plant identification and robot vision. On the basis of the work in the last year, we continued to focus on medical image retrieval and extended to modality classification subtask. The medical retrieval task of ImageCLEF 2012 uses a subset of PubMed Central containing 305,000 images. For modality classification, previous studies have shown that imaging modality is an important aspect of the image for medical retrieval. In user-studies, clinicians have indicated that modality is one of the most important filters that they would like to be able to limit their search by. Studies have shown that the modality can be extracted from the image itself using visual features. Additionally, using the modality classification, the search results can be improved significantly. In the ad-hoc image-based retrieval, participants will be given a set of 30 textual queries with 2-3 sample images for each query. The queries will be

classified into textual, mixed and semantic, based on the methods that are expected to yield the best results.

This year we performed two subtasks of medical image retrieval: modality classification and ad-hoc image-based retrieval. For modality classification, we concentrated on mono-modal visual-based image classifier. We used LibSVM to train the classifier, and edge histogram feature to represent images. To improve its performance, we tried to extend the training set. In ad-hoc retrieval, we applied a MeSH-based query expansion strategy and added a modality prediction step to further improve the retrieval performance. Our runs were also competitive among textual runs.

The remainder of this paper is organized as follows. In section 2 we describe our approaches in detail. And our submitted runs are discussed in section 3. Then we conclude in section 4.

## 2 Approaches

### 2.1 Modality Classification

In the experiment of medical image modality classification, we only concentrated on mono-modal visual-based image classifier. We performed experiments on three kinds of features: Edge Histogram Feature (EH), Tamura Feature (T) and Gabor Feature (G). We first used Medical Retrieval Images of 2011 as test data and then trained the classifier with data of 2012.

In the experiment, we used LibSVM to train one-versus-all classifiers. At first, we trained three classifiers per feature type and the classifier trained by Edge Histogram Feature had the highest accuracy. To fuse visual features, we simply averaged scores of classification result of different feature types per image. Finally, the combination of three features outperformed Edge Histogram Feature slightly. (see Table 1).

**Table 1.** experiments on Medical Retrieval Images of 2011.

Features	EH	T	G	EH+T	EH+G	G+T	EH+G+T
Accuracy	59.72%	49.93%	49.17%	59.72%	58.16%	53.96%	60.01%

To simplify the training process, we decided to use Edge Histogram Feature only. We tested several LibSVM parameters and in this case, the best result were obtained with a radial basis function kernel, parameter  $\gamma = 0.5$  and  $\text{cost} = 2$ .

To improve accuracy, we decided to extend the image training set. To begin with, we used two approaches to obtain pre-extended images of each modality from the image collection of the medical retrieval task. In the first approach, we trained an initial classifier with provided training images, with which we classified 40,000 images randomly selected from the collection. In the second approach, we selected top 200 images from image collection which were queried by images selected from provided training set[1]. Next, we manually selected appropriate images of each

modality from the pre-extended images and added them to the training set to train the final modality classifier.

## 2.2 Ad-hoc Image-based Retrieval

In ImageCLEF 2011, several groups adopted query expansion as the key techniques with the help of some external sources such as Unified Medical Language System (UMLS) and MeSH and achieved good results. After in-depth research of their approaches we decided to use MeSH as our source of query expansion.

MeSH is the National Library of Medicine's controlled vocabulary thesaurus. It consists of three types of record: Descriptors, Qualifiers and Supplementary Concept Records. Each type of records are organized in tree structure with the same hierarchy and stored in a xml file. As the records in MeSH are always a sequence of terms, so the topics given are pre-processed to get all the combinations of the terms after elimination of the stop words. Then we applied various strategies to explore the related information in MeSH to expand the query.

When we used MeSH to expand the query, we noticed a situation in which two n-grams that both their related child records and entry terms could be used expand the query and one was contained in the other at the same time. Let's take cell and muscle cell for example. Obviously, it will introduce too much noise if we add the child descriptors of the n-gram cell. So under the circumstance only the related child descriptors or entry terms of muscle cell are used to expand the query. Both methods below applied this rule.

1. If the query is a record, its child descriptors are added to the query. No expansion otherwise.
2. If the query is a record, we use its child descriptors to expand the query. If it is an entry term, the child descriptors of the descriptor which the query belongs to are used for expansion.

All our submitted runs applied the second method for query expansion as it gave better results compared with the first one.

To further improve the retrieval performance we combined our retrieval system with modality prediction. Firstly, a standard retrieval step was performed, and 2,000 images were obtained per topic. Then we used the modality classifier to predict modality of these images. Next, we extracted expressed modality (EM) from the images and the query of each topic. Finally, we combined expressed modality and predicted modality (PM) with retrieval result. In this process, the retrieval scores of 2,000 images were modified and top 1,000 images were the final result.

We investigated three fusion strategies with Medical Retrieval Image Collection of 2011 as test data. The first approach is score fusion (SF): the final scores are obtained with the retrieval score and modality classification score following the below equation (d standards for document):

$$s(d) = \alpha S_R(d) + (1 - \alpha) S_M(d) \quad (2.1)$$

The second approach is a variant of SF (SF\_variant): if EM is the same to PM, the final score is the sum of  $S_R$  and  $S_M$ , otherwise, the final score remains the original retrieval score  $S_R$ . The third strategy is Filtering (F): if EM is the same to PM, the

final score is 1.75 times of  $S_R$ , otherwise, the final score remains the original retrieval score  $S_R$  [2].

**Table 2.** experiments on Medical Retrieval Image collection of 2011.

Fusion Type	Map	Bpref	P10	P20
Original retrieval Result	0.175	0.2187	0.3133	0.27
Original retrieval Result +SF	0.1702	0.2133	0.3033	0.2683
Original retrieval Result +SF_variant	0.1854	0.2314	0.3267	0.2833
Original retrieval Result+F	0.1872	0.2361	0.3533	0.3

Finally, we chose Filtering to enhance the performance of retrieval performance.

### 3 Experiments and Results

#### 3.1 Modality Classification

**Table 3.** experiments on Medical Retrieval Images of 2012.

Run	Modality	Accuracy
ModalityClassificationSubmit	visual	39.7%
ModalityClassificationSubmit_Extend	visual	39%

Unfortunately, we submitted a wrong run. Table 3 describes the original run and the second run is the one which should have been submitted.

-V1: only used the provided training set to train classifier.

-V2: we extended the training set with 170 selected images from image collection.

However, the classification result with training set extended decreased slightly compared with the result of the original training set. We think the reason is that the training set is not well balanced, its size is not big enough and images which are added into training set are not classified precisely.

The text retrieval approaches are all based on the bag-of-words model, a text (such as image caption or full article) is represented as an unordered collection of words after tokenization and standard stopword removal. After the pre-processing step, two information models are considered: vector space model and topic model. We also use a query expansion mechanism, pseudo relevance feedback based on information-theoretic.

### 3.2 Ad-hoc Retrieval

Two different indexes were built. One contained caption of the image and the other also added article information. To further improve the performance, we also added a modality prediction step to filter retrieve result.

**Table 4.** Descriptions of ad-hoc image-based retrieval runs

Run	Description
TFIDF_CAPTION_ARTICLE[QE2]_MC	Use image caption and article as indexes with query expansion and modality classification
TFIDF_CAPTION_ARTICLE[QE2]	Use image caption and article as indexes with query expansion
TFIDF_CAPTION_ARTICLE_MC	Use image caption and article as indexes with modality classification
TFIDF_CAPTION_ARTICLE	Use image caption and article as indexes, no query expansion and modality classification
TFIDF_CAPTION[QE2]_MC	Use image caption as indexes with query expansion and modality classification
TFIDF_CAPTION[QE2]	Use image caption as indexes with query expansion
TFIDF_CAPTION_MC	Use image caption as indexes with modality classification
TFIDF_CAPTION	Use image caption as indexes, no modality classification

**Table 5.** retrieval results for ad-hoc retrieval task

Runid	Map	Gm-map	Bpref	P10	P30
TFIDF_CAPTION_ARTICLE[QE2]_MC	0.2081	0.0776	0.2134	0.3091	0.2045
TFIDF_CAPTION_ARTICLE[QE2]	0.2016	0.0601	0.2049	0.3045	0.1939
TFIDF_CAPTION_ARTICLE	0.1891	0.0508	0.1975	0.3318	0.1939
TFIDF_CAPTION_ARTICLE_MC	0.0959	0.0164	0.1075	0.1636	0.1152
TFIDF_CAPTION[QE2]_MC	0.1877	0.0519	0.1997	0.3	0.2045
TFIDF_CAPTION[QE2]	0.1673	0.037	0.1696	0.2955	0.1894
TFIDF_CAPTION_MC	0.1651	0.0467	0.1743	0.3	0.2076
TFIDF_CAPTION	0.1648	0.0441	0.1717	0.3318	0.1909

Our best runs were indexed with image caption and article with both query expansion and modality predication applied and it ranked second among all the textual runs. The results show that both query expansion and modality predication could improve the retrieval performance. Most of our runs which used the captions and article as indexes performed better than those use captions only.

## 4 Conclusion

This article describes the approaches and results of BUAA AUDR group at ImageCLEF 2012. We submitted 9 runs for modality classification and ad-hoc retrieval. For modality classification, the result was poor. We need to do more work to get it better. For ad-hoc retrieval, we adopted a Mesh-based query expansion and enhanced the results by a modality prediction step. The best of our runs ranked second in all textual runs. The results indicate that query expansion and modality predication can improve the retrieval performance. As our modality classification was not satisfied, our retrieval performance still has room for improvement.

## References

1. Guillaume Jacquet, Gabriela Csurka and St'ephane Clinchant. Xrce's participation in Medical Image Modality Classification and Ad-hoc Retrieval Tasks of ImageCLEF 2011. In Working Notes of CLEF 2011, Amsterdam, the Netherlands, 2011.
2. 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.
3. St'ephane Clinchant, Gabriela Csurka, Julien Ah-Pine, Guillaume Jacquet, Florent Perronin, Jorge Sanchez, and Keyvan Minoukadeh. Xrce's participation in wikipedia retrieval, medical image modality classification and ad-hoc retrieval tasks of ImageCLEF 2010. In Working Notes of CLEF 2010, Padova, Italy, 2010.
4. Jacinto Mata, Mariano Crespo, and Manuel J. Maña. LABERINTO at ImageCLEF 2011 Medical Image Retrieval Task In Working Notes of CLEF 2011, Amsterdam, the Netherlands, 2011.
5. Hong Wu and Chengbo Tian UESTC at ImageCLEF 2011 Medical Retrieval Task In Working Notes of CLEF 2011, Amsterdam, the Netherlands, 2011.