# Classification of MRI Images: A Review

Poonam Sengar
Department of Computer Science
L.N.C.T
Bhopal, India
poonamlikerose@gmail.com

Prof. Alekh Dwivedi
Department of Computer Science
L.N.C.T
Bhopal, India

Prof. Vineet Richhariya
H.O.D, Department of Computer Science
L.N.C.T
Bhopal, India

**Abstract—** With the rapid development of the medical science more and more medical images are generated rapidly like OMR, CT scan, X-ray etc. Due to that an efficient system is essential for the indexing, storing and analyzing such images. The analyzing cost of such images is very high. The analysis quality also differs and highly prone to errors. The classification of such images is a quite harder job. Hence it is essential to develop a strong system for classification of such images. In this paper we are presenting some techniques that are used to classify the X-ray images.

**Keywords—** image classification, medical images, X-ray, medical image databases.

———————————— ◆ ————————————

## I. INTRODUCTION

A data mining is one of the fast growing research field which is used in a wide areas of applications. The data mining consists of classification algorithms, association algorithms and searching algorithms. Different classification and clustering algorithm are used for the synthetic datasets. Clustering and decision tree are two of the mostly used methods of data mining which provide us much more convenience in researching information data. Using the data mining classification technique images is also classified. This concept can be also used to classified medical images. The research on data mining and medical content-based image retrieval has progressed presently. It is possible to develop a system that can capable to handle patient's care system along with successful categorization of information directly extracted from images [1].

When discussing about traditional system the categorization of medical images were restricted to selected classes only. The categorization of images is critical when it deals with real-life constraints of content-based image classification in the medical domain [1]. Content-based image retrieval (CBIR) allows eliminating of the difficulties exist in traditional text-based query for large image database. CBIR as an important corresponding search method has been one of the most active research areas in the field of computer vision for the past years. CBIR deals with the analysis of image content and the development of tools to represent visual content in a way that can be efficiently searched and compared [2].

The performance of CBIR system may be enhanced by choosing a successful classification of image. This classification was essential to eliminate irrelevant Images. For example searching of images to a specific body parts is possible only when image database is classified. Suppose searching criteria is wrist then the search result should be related to it. It is possible only when image database is classified efficiently. All other images should be concealing from the search result. Although classification tasks is very crucial specially when it deals with real-life constraints of content-based image classification in medical domain [2].

In medical retrieval system can also provide diagnostic support to physicians or radiologists by providing proper display of relevant past cases to assist them in the decision making process. The picture archiving and communications system (PACS) is a tool that works on medical image databases. Instead of storing images it also capable to provide general analysis. It also suggest past relevant cases for better analysis and references. But retrieval of similar images or records is difficult just because of textual information search is available. The classification task begins with extracting appropriate features of the image. It is one of the most important factors in design process of such system. Visual features were categorized into primitive features such as color, shape and texture [2].

Automated image categorization and retrieval system required efficient algorithm solution for diagnostic-level categorization, this will help radiologist to search the radiographic medical images. In content-based image retrieval systems images are categorized or accessed by their features.

ISSN 2229-5518

The image features means image color, its texture, shape and etc. some examples of content based image retrieval system are QBIC, Photo book, Virage, Visual SEEK, Netra. Maximum system uses data mining specially clustering technique to classify images [3].

Clustering is a division of data into groups of analogous objects. Every grouping known as cluster consists of objects that are like amongst them and dissimilar compared to object of other groups. Instead of data by fewer clusters essentially loses certain fine details, but achieves generalization. It represents many data objects by only some clusters, and therefore it models data by its clusters. Clustering is the unsupervised classification of patterns into groups known as clusters. Clustering is a difficult problem combinatorial, and dissimilarity in assumptions and contexts in different communities have made the transfer of useful generic concepts and methodologies slow to occur. Cluster analysis aims at identifying groups of related objects and, hence helps to discover distribution of patterns and interesting correlations in large data sets. So that it can be used in wide research since it arises in many application domains. Especially, in the last years the availability of huge transactional and experimental data sets and the arising requirements for data mining created needs for clustering algorithms that scale and can be applied in diverse domains [4].

Clustering is considered an interesting approach for finding similarities in data and putting similar data into dissimilar sets. Clustering partitions a data set into several groups such that the similarity within a group is larger than that among groups. The idea of data grouping, or clustering, is simple to use and in its nature and is very near to the human way of thinking; whenever they are presented with a large amount of data, humans are usually tend to summarize this huge number of data into a small number of groups or categories in order to further facilitate its analysis. Most of the data collected in many problems seem to have some inherent properties that lend themselves to natural grouping. Clustering is a challenged research field which belongs to unsupervised learning. The number of clusters needed is unknown and the formation of clusters is data driven completely [4].

Clustering can be the pretreatment part of other algorithms or an independent tool to obtain data distribution, and also can discover isolated points. Common clustering algorithms are K-MEANS, BIRCH, CURE, DBSCAN etc. Each of them has respective advantages: KMEANS is simple and easy to understand, DBSCAN can filter noises splendidly, CURE is insensitive to input etc. But now there still has no algorithm which can satisfy any condition especially the large-scale high dimensional datasets, so it is important for us to improve and develop clustering methods [4].

K-Means is a mostly used algorithm that can deal with small convex datasets preferably. K-Means Clustering is a method of cluster analysis which aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. The algorithm is called k-means, where k represents the number of clusters required, since a case is allocated to the cluster for which its distance to the cluster mean is the negligible. The achievement in the algorithm centres on finding the k-means [5].

Medical x-ray images are characterized with intensity variations and poor contrast, thus, it is difficult to extract a set of appropriate features according to their characteristics. For this purpose various methods were implemented. But with enhancement of better medical apparatus along with high resolution image it is difficult to work with traditional system. Hence new algorithms are essential to manage this huge image databases. The rest of paper is organized as follows. In Section II describes about background details of medical image classification followed by related work in section III. Section IV describes about conclusion.

## II. BACKGROUND

Presently medical images are explosively used by physicians for diagnosing. With the much development of the medical apparatus report generation and observation is quite easy. This will also lead a production of high quality images. To properly store, manage and categorization of such images are also difficult task. According to general study categorization of medical images are quite difficult due to huge amount of images with their variety of indexing. Due to this an appropriate system is needed to manage these images database and also provide efficient functioning over such systems. For the classification of images clustering is used. There are already work done in this field but efficient algorithms are still required.

## III. RELATED WORK

This section describes some related work to classification of medical image.

A novel approach was proposed to increase the number of classes with a higher accuracy rate by iterative filtering on the training dataset by Zare et al [1]. Filtering is done according to their classification performance. They presented a novel method to achieve classification of class of Image CLEF 2007 medical database [6]. In this scheme they have four iterations steps. These steps hold different classification models. Within the iteration generation process was performed in two steps. The construction of a model from the entire dataset was the first step. This was used to assess filter high accuracy classes (HAC). This will achieve accuracy of 80 % and using this process they can train 20% of the data set. The classes under HAC were only used to construct the classification model under second step. These steps are also continued to next iterations [1].

The classification process consists training phase and testing phase. The training phase was used to selected features that are extracted from all the training images and the classifier is trained on the extracted features to create a classification model. The training phase was also responsible for feature extraction and model generation. In feature extraction they used bag of visual word (BoW). BoW was used for extracting feature using detect and extract local features, codebook construction and represent an image in a histogram using the codebook. Extraction of BoW started with detecting and extracting local features. Feature detection is the process in which a spatially limited image region that is salient or prominent must be identified. The next step in implementation of BoW is codebook construction where it is built using clustering or the vector quantization algorithm. This step usually uses k-means clustering method, and uses cluster centre as visual vocabulary term. The goal of codebook construction is to identify a set of visual patterns that reflects the image collection contents. Upon identification of cluster centres, each image is represented as histograms of these cluster centres by simply counting the frequency of the words appearing in an image [1].

Upon extraction of BoW representation from the training dataset, it was then used as input to support vector machine (SVM) classifier to construct the model. This can evaluate in the training phase itself to ensure that the best possible accuracy rate is attained for every individual class in the database. Eighty percent of it was used to construct the classification models, and the remaining 20% of the training data were taken for test images for evaluation purpose of the generated model. Only one of the classification models was constructed in training phase for testing. This will be applied on the test image in order to classify it into the predefined category. Classification process on the unseen test image starts with identifying the respective classification model for that particular image. If right model was not chosen than it may cause decreasing accuracy rate. For those classes which are visually similar, SVM or any other classification technique would be biased to the classes with a larger number of training images. This examination shows that depending on only one technique to gain high accuracy for every individual class of such a database with the said complexity is unreliable [1].

Automatic classification of medical X-ray images: hybrid generative-discriminative approach was also proposed by Zare et al [2]. Along with rapid progress in the application of local descriptor in pattern recognition, computer vision and image retrieval, the bag of word (BoW) approach has appeared promising for object classification and image retrieval. The classification task begins with extracting appropriate features of the image. It is one of the most important factors in design process of such system. Moreover, the feature extraction step affects all other subsequent processes. The probabilistic latent semantic analysis (PLSA) [7] has been proposed to learn co-occurrence information between elements in the vector space in an unsupervised manner to disambiguate the BoW representation. PLSA can help to disambiguate visual words because of the ability of the PLSA model to generate a robust, high level representation and low-dimensional image representation since PLSA introduces a latent,

In recently Selvi and Kavitha [3] developed an efficient and powerful medical search engine to classify and search the radiographic medical images. They used differnet feature extraction method to develop this system. The images that are used was categorised in two groups one is Low-level image representation and second is local patch-based image representation. In first category Gray Level Co-occurrence Matrix (GLCM), Canny Edge Operator, Local Binary Pattern (LBP) and pixel value type's images were used. While in second category Bag of Words (BoW) was used. They also offered a novel image indexing and retrieval algorithm using local tetra patterns (LTrPs) for content-based image retrieval (CBIR). The standard local binary pattern (LBP) and local ternary pattern (LTP) encode the relationship between the referenced pixel and its surrounding neighbors by computing gray-level difference. This method encodes the relationship between the referenced pixel and its neighbors, based on the directions that are calculated using the first-order derivatives in vertical and horizontal directions. For this they chosen a generic strategy to compute th-order LTrP using th-order horizontal and vertical derivatives for efficient CBIR and analyze the effectiveness of our proposed algorithm by combining it with the Gabor transform [3].

In year 2010, Dan et al [1] presented A Synthesized Data Mining Algorithm Based on Clustering and Decision Tree. They improve the traditional algorithms like CURE and C4.5 appropriately, and present a new synthesized algorithm CA for mining large-scale high dimensional datasets. The basic idea of CA is shown as follows: first introduce PCA to analyze the relevancy between features and replace the whole dataset with several countable composite features; then improve CURE to part the set into several clusters which can be the pretreatment of other algorithms and achieve the reduction of sample scale; finally introduce parallel processing into C4.5 to enhance the efficiency of building decision tree. The decision tree classification part of CA algorithm is improved based on C4.5, and the improvements are mainly embodied in threshold partition and selection of testing features. In traditional C4.5 algorithm, they will divide the datasets dynamically, and select the values with the biggest gain ratio to split continuous features. Introduce three different classifiers to ascertain the correctness of selecting features and avoid bios problems. This synthesized CA algorithm is improved based on traditional CURE and C4.5 methods. It introduces scale reduction, feature reduction and classification analysis to handle large and high dimensional datasets. By applying CA algorithm in maize seed breeding, they can find out the important features which will influence breeding tremendously and obtain the classification model of whole maize samples. The experiments show the efficiency of CA is higher not only in clustering but also in decision tree. There also exist some problems needed to research further more. CA is sensitive to some parameters like the clustering number, shrink factors and the threshold. C4.5 only can covenant with the dataset that has the classification feature. The dataset treated is a little small which will impact the final output of algorithms [4].

Manish Verma et al [5] proposed A Comparative Study of Various Clustering Algorithms in Data Mining. They provide a comparative study among various clustering. They compared six types of clustering techniques- k-Means Clustering, Optics, DBSCAN clustering, Hierarchical Clustering, Density Based Clustering and EM Algorithm. Such clustering methods are implemented and analyzed using a clustering tool WEKA. Performances of the 6 techniques are presented and compared.

Running the clustering algorithm using any software produces almost the same result even when changing any of the factors because most of the clustering software uses the same procedure in implementing any algorithm [5].

In year 2012, Bhaskar N. Patel et al [8] presented an Efficient Classification of Data Using Decision Tree. They survey many techniques related to data mining and data classification techniques. They also select clustering algorithm k-means to improve the training phase of Classification. The Learning classification techniques in data mining can be classified into three fundamental types; first one is supervised second one is unsupervised and finally third one is reinforced. There are at least three techniques which are used to calculate a classifier's accurateness. One method is to divide the training set by using two-thirds for training and the other third for approximation presentation. In another method called cross-validation, the training set is separated into reciprocally exclusive and equal-sized subsets and for each subset the classifier is trained on the union of all the other subsets. The usual rate of the error for each subset is therefore an estimate of the error rate of the classifier. Validation called Leave-one-out is an exceptional case of cross validation. All test subsets consist of a single instance. This validation type is more luxurious computationally, but functional when the most exact estimation of a classifier's error rates requisite. Training an average decision tree escorts to a quadratic optimization problem with bound constraints and one linear equality constraints. Training support vector machines involves a huge optimization problem and many specially designed algorithms have been offered. The algorithm that was used is called "Decision Tree Induction" that accelerates the training process by exploiting the distributional properties of the training data, i.e. the natural clustering of the training data and the overall layout of these clusters relative to the decision boundary of support vector machines [8].

The method called k-means since each of the K clusters is represented by the mean of the objects within it. It is also called the centroid method since at each step the centroid point of each cluster is assumed to be known and each of the remaining points are allocated to the cluster whose centroid is closest to it. The K-means algorithm keeps on as follows. First, it arbitrarily selects k of the objects; each object primarily represents a center. For each of the outstanding objects, an object is allocated to the cluster to which it is the most alike, based on the distance among the object and the cluster. Then it calculates the new mean for all clusters. This procedure repeats in anticipation of the criterion function converges. Decision tree is useful because construction of decision tree classifiers does not require any domain knowledge. It can handle hi-dimensional data. The classification and learning steps of decision tree induction are simple and fast. Their representation of acquired knowledge in tree form is easy to assimilate by users. The decision tree algorithm is a top-down induction algorithm. The mean of this algorithm is to construct a tree that has leaves that are harmonized as potential. The most important step of this algorithm is to carry on dividing leaves that are not homogeneous into leaves that are as homogeneous as possible. Once the result obtained, it can be reused for next research. This research depicts on compares reformulated decision tree with standard decision tree for dataset. This comparison is from threshold (complexity) from low to high with reference to the testing accuracy [8].

In same year Gothai, E. and P. Balasubramanie proposed An Efficient Way for Clustering Using Alternative Decision Tree. Their study proposes a Multi-Level Clustering mechanism using alternative decision tree algorithm that combines the advantage of partition clustering, hierarchical clustering and incremental clustering technique for rearranging the most closely related object. The clustering initiation should happen based on the short name value, each short name pointing to the appropriate whole record object. The proposed MLC algorithm has been experimentally tested on a set of data to find a cluster with closely related object. This method is used to overcome the existing system problem, such as manual intervention, misclassification and difficulties of finding a partition range and so on. MLC forms a tree for the clustering process. In the tree structure, the height of each level of nodes represents the similar degree between clusters. MLC incorporate the futures of ADTree features and overcome the existing hierarchical clustering problem [9].

ADTree divide the data based on short name; if cluster is already available with the short name then insert a record into the same cluster else create a new cluster with the new name of short name then insert into a new cluster. In every cluster, sub-set diminutive name points to the whole record. The cluster formation method mainly focus on form a similarity value in single group, for this purpose they are using different method and result of each method is different cluster based on data and spread condition. The experimental results shows the proposed system has lower computational complexity, reduce time consumption, most optimize way for cluster formulation and better clustering quality compared with the existing hierarchical clustering algorithm. They ran extensive experiments with them to find time consumption and compared them with various versions of existing algorithms in order to show this new system reduces the time consumption. The new method offers more accuracy of cluster data without manual intervention at the time of cluster formation. Compare to existing clustering algorithm either partition or hierarchical, new method is more robust and easy to reach the solution of real world complex business problem [9].

A novel feature extraction framework for medical x-ray images classification is proposed by Ghofrani et al [10]. As per this scheme, extract centre symmetric local binary patterns from local part of shape and directional information extracted from images to achieve a set of capable features after some preprocessing. This method worked in three: preprocessing, feature extraction and classification process. Preprocessing is used to eliminate the effects of noises and also manage grey level variation along with set of capable features. After this feature were extracted local parts of each image. This can be done in three parts. At first, in order to achieve local features, each image is partitioned to 25 sub images. This preserve the information of boundaries of sub images, image partitioning to overlapping sub images is preferable to non-overlapping. Secondly Gabor transform computation was applied before extracting features to achieve more shape and directional information. Finally, in the last stage, CS-CBP features are extracted from filtered images. In last stage after the feature extraction, the images are labeled in to their respective classes using multi-class on-against-one algorithm [10].

Avni et al [11] presented an efficient image categorization and retrieval system for medical image database especially for radiographic images. They presented a patch-based classification system that has demonstrated very strong classification rates while also providing efficiency in the retrieval process. This scheme was composed of a feature extraction phase, a dictionary construction based on the training archive, an image representation phase and a classification phase. In very first step images was represented as collection of small patches. After that sampling techniques were applied. Patches along the border of the image are considered as noise and are ignored. The intensity values within a patch are normalized to have zero mean and unit variance. This provides local contrast enhancement and augments the information within a patch. Patches that have a single intensity value of black are ignored. Based on the representative set on images data dictionary was trained. Each image is represented as set of patches. Using the feature extraction parameters that were learned (PCA, feature weights) and the generated dictionary, each image is represented as a histogram of visual words. In this step images are sampled with a dense grid. As per result obtained they conclude with classification results in a lung pathology detection application [11].

In year 2010, a learning-based algorithm for automatic medical image annotation based on sparse aggregation of learned local appearance cues was suggested by Tao et al [12]. They adopted a hybrid approach based on robust aggregation of learned local appearance findings, followed by the exemplar-based global appearance filtering. This scheme is used to detect multiple focal anatomical structures within the medical image. It detects multiple focal anatomical structures within the medical image. This is achieved via learning-by-example landmark detection algorithm. It performs simultaneous feature selection and classification at several scales. After that inconsistent findings through a robust sparse spatial configuration (SSC) algorithm were eliminated. This was consistent and reliable local detections will be retained while outliers will be removed. A reasoning module assessing the filtered findings were applied at last i.e., remaining landmarks is used to determine the final content/orientation of the image. According to classification task, a post-filtering component using the exemplar-based global appearance check for cases with low classification confidence may also be included to reduce false positive (FP) identifications [12].

Deselaers and Ney presented three methods for the classification and annotation of medical radiographs. Firstly, they describe baseline methods using global image descriptors which can be compared very efficiently. After that a non-linear image deformation model accounting for local image distortions was applied. Finally, a basic approach to make use of the hierarchical class structure in the medical image annotation task of ImageCLEF 2007 is presented. All methods are compared and similarities and differences are analyzed [13]. Selvi and Kavitha presented a system based on content based image retrieval for medical images. Efficient medical image categorization and retrieval system in large radiographic archives (IRMA database) is developed for a medicine physicians and researchers. This method is based on local patch representation of the image content and a bag-of features approach for defining image categories with a kernel based SVM classifier. Firstly they investigate the sensitivity to various parameters that define the system followed by classification and retrieval experiments on large radiograph archives. The system validation was conducted using a database of 2,500 categorized radiographs. This dataset is the basis for the Image CLEF med 2009 medical image classification competition. A set of 2,500 images are used for training, and 500 serve for testing. There are 25 different categories within the archive, differing in either the examined region, the image orientation with respect to the body or the biological system under evaluation [14].

Foncubierta-Rodríguez et al [15] presented a Medical Image Retrieval using a Bag of Meaningful Visual Words. They hypothesize that this variance of the BoVW method is strongly related to the quality of the vocabulary used, understanding quality as the ability of the vocabulary to accurately describe useful concepts for the task. They computed the latent semantic concepts in the dataset in an unsupervised way by analyzing the probability of each word to occur. This allows extracting concepts from a combination of various visual word types, since the concepts are discovered based on the probability of co-occurrence of a set of visual words regardless of their origin. The resulting reduced vocabularies present two benefits over the full ones. Preliminary experiments on the vocabulary pruning technique over the training set were based on removing meaningless visual words from the descriptors but not from the vocabulary [15].

## IV. CONCLUSION

In today's age number of medical images being generated, managed and stored in various hospitals and clinics. This is just because of betterment of medical field and enhancement of variety of medical apparatus along with rapid advancement of software and hardware technology. Effectively and efficiently searching in these large image collections poses significant technical challenges as the characteristics of the medical images differ significantly from other general purpose images. In this paper various related methods are used to classify and indexed these images. Quality and speed of retrieving medical images from large collections of databases are more important and essential for the better functioning of health organizations.

## References

[1] Zare, Mohammad Reza, Ahmed Mueen, and Woo Chaw Seng. "Automatic classification of medical X-ray images using a bag of visual words." IET Computer Vision, vol. 7, no. 2, pp. 105-114, 2013.

[2] Zare, Mohammad Reza, Ahmed Mueen, Mohammad Awedh, and Woo Chaw Seng. "Automatic classification of medical X-ray images: hybrid generative-discriminative approach." IET Image Processing, vol. 7, no. 5, pp. 523-532, 2013.

[3] Selvi, S. Malar, and C. Kavitha. "Radiographic medical image retrieval system for both organ and pathology level using bag of visual words." International Journal of Engineering Sciences & Emerging Technologies, ISSN: 22316604, vol. 6, no. 4, pp. 410 – 416, 2014.

[4] Ji Dan, Qiu Jianlin, Gu Xiang, Chen Li, and He Peng "A Synthesized Data Mining Algorithm Based on Clustering and Decision Tree", IEEE 10th International Conference on Computer and Information Technology (CIT), pp. 2722 – 2728, 2010.

[5] Manish Verma, Mauly Srivastava, Neha Chack, Atul Kumar Diswar, Nidhi Gupta " A Comparative Study of Various Clustering Algorithms in Data Mining", International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622, Vol. 2, pp.1379-1384, Issue 3, May-Jun 2012.

[6] Müller, Henning, Thomas Deselaers, Thomas Deserno, Paul Clough, Eugene Kim, and William Hersh. "Overview of the ImageCLEFmed 2006 medical retrieval and medical annotation tasks." In Evaluation of Multilingual and Multi-modal Information Retrieval, pp. 595-608, Springer Berlin Heidelberg, 2007.

[7] Hofmann, Thomas. "Unsupervised learning by probabilistic latent semantic analysis." Machine learning 42, no. 1-2 (2001): 177-196.

[8] Bhaskar N. Patel, Satish G. Prajapati and Dr. Kamaljit I. Lakhtaria "Efficient Classification of Data Using Decision Tree", Bonfring International Journal of Data Mining, ISSN 2277 – 5048, Vol. 2, No. 1, pp. 6- 12, March 2012.

[9] Gothai, E. and P. Balasubramanie "An Efficient Way for Clustering Using Alternative Decision Tree", American Journal of Applied Sciences, ISSN 1546-9239, vol. 9, no. 4, pp. 531-534, 2012.

[10] Ghofrani, Fatemeh, Mohammad Sadegh Helfroush, Habibollah Danyali, and Kamran Kazemi. "Medical X-ray image classification using Gabor-based CS-local binary patterns." In Proceedings International Conference on Electronics, Biomedical Engineering and its Applications, pp. 7-8, 2012.

[11] Avni, Uri, Hayit Greenspan, Eli Konen, Michal Sharon, and Jacob Goldberger. "X-ray categorization and retrieval on the organ and pathology level, using patch-based visual words." IEEE Transactions on Medical Imaging,vol. 30, no. 3, pp. 733-746, 2011.

[12] Tao, Yimo, Zhigang Peng, Bing Jian, Jianhua Xuan, Arun Krishnan, and Xiang Sean Zhou. "Robust learning-based annotation of medical radiographs." In Medical Content-Based Retrieval for Clinical Decision Support, pp. 77-88, Springer Berlin Heidelberg, 2010.

[13] Deselaers, Thomas, and Hermann Ney. "Deformations, patches, and discriminative models for automatic annotation of medical radiographs." Pattern Recognition Letters, vol. 29, no. 15, pp. 2003-2010, 2008.

[14] Selvi, S. Malar, and Mrs C. Kavitha. "Content Based Medical Image Retrieval System (CBMIRS) Using Patch Based Representation." IOSR Journal of Computer Engineering (IOSR-JCE), ISSN: 2278-0661,pp. 24-36, 2013.

[15] Foncubierta-Rodríguez, Antonio, Alba García Seco de Herrera, and Henning Müller. "Medical image retrieval using bag of meaningful visual words: unsupervised visual vocabulary pruning with PLSA." In Proceedings of the 1st ACM international workshop on Multimedia indexing and information retrieval for healthcare, pp. 75-82, ACM, 2013.