INTELLIGENT METHODS FOR AUTOMATIC CLASSIFICATION OF MEDICAL IMAGES

MOHAMMAD REZA ZARE

THESIS SUBMITTED IN FULFILLMENT
OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

FACULTY OF COMPUTER SCIENCE AND
INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA

2013

Abstract

The ever increasing number of medical images in hospitals urges on the need for generic image classification systems. These systems are in an area of great importance to the healthcare providers. However, classification of large medical database is not an easy task due to unbalance number of training data, intra class variability and inter-class similarities among them. In this thesis, three classification frameworks are presented to increase the accuracy rate of every individual categories of such database. Bag of Words (BoW) has been explored for image representation techniques in the proposed frameworks.

In the first framework, we proposed an iterative filtering scheme on the database where classes with optimal accuracy rate are filtered out. They are then used to construct a new classification model. These processes are carried out in four iterations. As a result, four classification models are generated from different number of classes. These models are then employed to classify unseen test images.

In continuation of the first framework, another classification framework is proposed for classes which are left with low accuracy rate after the first iteration. These classes are those with high ratio of intra class variability and inter-class similarities. The classification process is carried out by employing three different annotation techniques, i.e. Annotation by binary classification, Annotation by Probabilistic Latent Semantic Analysis (PLSA) and Annotation using top similar images. The annotated keywords produced are integrated by applying ranking similarity. The final annotation keywords were then divided into three levels according to the body region, specific bone structure in body region as well as imaging direction. Different weights were given to each level of the keywords; they are then used to calculate the weightage for each category of medical images based on their ground truth annotation. The weightage computed from the generated annotation of test image was compared with the weightage of each category of medical images, and then the test image would be assigned to the category with closest weightage to the test image.

In the third framework, the unsupervised latent space model is used in feature extraction to discover patterns of visual co-occurrence. In this direction, we employed PLSA to learn the co-occurrence information between elements in the vector space. PLSA model can generate a robust, high level representation and low-dimensional image representation. This would help to disambiguate visual words. Thus, a classification framework based on integration of PLSA and discriminative Support Vector Machine (SVM) classifier is developed. In this framework, both visual features and textual features of the images are incorporated by using multi-modal PLSA model.

The experimental results on all the above frameworks have shown an increment in accuracy rate at the entire database level as well as at class specific level compared with other methods.

Abstrak

Bilangan yang semakin meningkat imej perubatan di hospital-hospital menggesa kepada keperluan untuk sistem klasifikasi generik imej. Sistem ini adalah dalam kawasan kepentingan yang besar kepada pembekal penjagaan kesihatan. Walau bagaimanapun, pengelasan pangkalan data perubatan yang besar bukan satu tugas yang mudah kerana kepada nombor ketidakseimbangan data latihan, kebolehubahan kelas intra dan antara kelas persamaan antara mereka. Dalam tesis ini, tiga rangka kerja klasifikasi dikemukakan untuk meningkatkan kadar ketepatan setiap kategori individu pangkalan data itu. Beg Perkataan (Bow) telah diterokai untuk teknik perwakilan imej dalam rangka kerja yang dicadangkan.

Dalam rangka kerja pertama, kami mencadangkan lelaran penapisan skim pada pangkalan data di mana kelas dengan kadar ketepatan yang optimum akan ditapis keluar. Mereka kemudiannya digunakan untuk membina model pengelasan baru. Proses-proses ini dijalankan dalam empat lelaran. Hasilnya, empat model klasifikasi dijana dari nombor yang berbeza kelas. Model-model ini kemudian digunakan untuk mengelaskan imej ujian ghaib.

Dalam kesinambungan rangka kerja pertama, satu lagi rangka kerja klasifikasi dicadangkan bagi kelas yang ditinggalkan dengan kadar ketepatan yang rendah selepas lelaran pertama. Ini kelas mereka dengan nisbah tinggi kepelbagaian kelas intra dan persamaan antara kelas. Proses pengelasan dijalankan dengan menggunakan tiga anotasi teknik yang berbeza, Anotasi iaitu mengikut klasifikasi binari, Anotasi oleh Analisis Semantik Pendam Kebarangkalian (PLSA) dan Anotasi menggunakan imej atas sama. Kata kunci beranotasi yang dihasilkan bersepadu dengan menggunakan persamaan kedudukan. Kata kunci anotasi akhir itu kemudiannya dibahagikan kepada tiga peringkat mengikut rantau badan, struktur tulang tertentu di rantau badan serta hala pengimejan. Berat yang berbeza telah diberikan kepada setiap peringkat kata kunci; mereka kemudiannya

digunakan untuk mengira pemberat bagi setiap kategori imej perubatan berdasarkan anotasi kebenaran tanah mereka. Pemberat yang dikira dari anotasi dijana imej ujian berbanding dengan wajaran setiap kategori imej perubatan, dan kemudian imej ujian akan diberikan untuk kategori dengan wajaran terdekat kepada imej ujian.

Dalam rangka ketiga, ruang model terpendam tanpa pengawasan digunakan dalam pengekstrakan ciri untuk menemui corak visual bersama-kejadian. Dalam arah ini, kita bekerja PLSA untuk mempelajari maklumat kejadian bersama antara unsur dalam ruang vektor. Model PLSA boleh menjana teguh, perwakilan peringkat tinggi dan rendah dimensi perwakilan imej. Ini akan membantu untuk disambiguate perkataan visual. Oleh itu, rangka kerja klasifikasi berdasarkan integrasi pengelas SVM PLSA dan diskriminatif dibangunkan. Dalam rangka kerja ini, kedua-dua ciri-ciri visual dan ciri-ciri teks imej diperbadankan dengan menggunakan model multi-modal PLSA.

Keputusan eksperimen semua rangka kerja di atas telah menunjukkan kenaikan dalam kadar ketepatan pada peringkat pangkalan data keseluruhan serta di peringkat tertentu kelas berbanding dengan kaedah lain.

Acknowledgment

First and foremost of my entire deep thanks to ALLAH the ALMIGHTY who gave me the ability and patience to complete this work.

This journey would have been impossible without the support, motivation and guidance given to me by my late brother, Dr. Dariush Zare. I would like to dedicate this work to my late brother and his wife, Mrs. Eva Romero and his beautiful daughter, Yasmin Zare who never failed to give me endless courage and strength to achieve my goal.

The completion of this dissertation has definitely been a challenge and could not have been accomplished without the assistance and support from many individuals. In the first place I would like to record my gratitude to my supervisor Dr. Woo Chaw Seng, for his kind supervision, encouragement, patience, guidance and friendship he has given me from the early stage of this research as well as giving me extraordinary experiences throughout the work. I am very grateful for the opportunity to complete this dissertation under Dr. Woo's supervision.

I would also like to thank to my Co-supervisor, Dr. Ahmed Mueen for his commitment to helping see this project through to its completion and his equally generous and wise guidance during its development.

I would like also to thank Dr. Mehdi Ghorbanzadeh, Dr. Mojtaba dashtizad, Dr. Nur Adura Yaakup, Mr. Atif Baloch, Mr. Muhammad Razaullah, Ms. Parveenpal Kaur for their help and motivation during past three years.

Words fail me to express my appreciation to Paria Ebrahimi, my wife, who never stopped supporting and understanding me and had confidence in me to accomplish my goals; and to my son, Dariush Zare, who illuminated my days with his smile. Thank you for your moral support, patience, devotion and patiently sharing these years with me.

Last, but certainly not least, I would like to express my deepest gratitude to my parents, my sisters and brother for their unconditional love. Thank you for your encouragement, prayer, advice and support.

Finally, I would like to thank the people of Malaysia for their hospitality and friendliness extended toward my family and me. We never forget their kindness.

Publication List

- Zare MR, Mueen A, Woo CS, (2013), Automatic Classification of Medical X-ray Images using Bag of Visual Word , IET Computer Vision, Volume 7, Issue 2, p. 105 – 114 (ISI Index Publication)
- Zare MR, Mueen A, Awedh M, Woo CS, (2013), Automatic Classification of Medical X-Ray Images: Hybrid Generative-Discriminative Approach, IET Image Processing, doi: 10.1049/iet-ipr.2013.0049 (ISI Index Publication)
- Zare MR, Woo CS, Mueen A, (2013), Automatic Classification of medical X-ray Images, Malaysian Journal of Computer Science, Volume 26, Issue 1, p. 9-22 (ISI Index Publication)
- Zare MR, Mueen A, Woo CS, (2013), Automatic Medical X-ray Image Classification using Annotation, Journal of Digital Imaging, doi: 10.1007/s10278-013-9637-0 (ISI Index Publication)
- Zare MR, Mueen A, Woo CS, (2011),Combined Feature Extraction of Medical X-ray Images, IEEE Computer Society, Third International Conference on Computational Intelligence, Communication System and Networks
- Zare MR, Mueen A, Woo CS,(2011),Merging Scheme Based Classification of Medical X-ray images, IEEE Computer Society, International Conference on Computational Intelligence, Modelling and simulation

Table of Content

List of Figures

List of Tables

Chapter 1 Introduction

1.1. Motivation

We are living in the age of multimedia information technologies burdened by information overdose. A wide availability of digital devices such as digital cameras, scanners, mobile phones, notebooks at reasonable prices accelerate the growth of multimedia content production. Images are the most popular among the variety of multimedia contents. There is also an increase of digital information in medical domain where medical images of different modalities i.e., X-rays, computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), etc. are produced everyday in massive numbers. For instance, over 640 million medical images were stored in more than 100 National Health Service Trust in UK, as of March 2008 (Khaliq *et al.*, 2008). Medical image databases are the key component in diagnosis and preventive medicine. As medical image databases show a wealth of information, it also creates a problem in terms of retrieving the desired images. As a result, there is an increased demand for a computerized system to manage these valuable resources. In addition, managing such data demands high accuracy since it deals with human life.

Currently, many hospitals and radiography departments are equipped with Picture Archiving and Communications (PACS). In medical domain, such retrieval system can also provide diagnostic support to physicians or radiologists by displaying relevant past cases to assist them in decision making process. Retrieving similar images using PACS is a very challenging task as the searches are carried out according to the textual attributes of image saved in Digital Imaging and COmmunications in Medicine (DICOM) header. Even though

DICOM header contains many important information, it still remains suboptimal due to its high error rate reported in previous study (Güld *et al.*, 2002) which can be an obstruct in an accurate retrieval of the desired images. As such, the research on medical Content Based Image Retrieval (CBIR) has progressed for the past decades. Medical CBIR system enables the elimination of the difficulties that exist in traditional text-based query for large medical image database. It 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 without human assistance. Content-based medical image retrieval can be used in few large domains such as teaching and research. In clinical teaching, lecturers can use a large medical image databases to search for interesting cases to show to students. Medical CBIR system can also be beneficial in research domain where the researchers can find new correlations between the visual nature of the case and its diagnosis or textual description in medical studies. It is also playing an increasing role in a wide range of applications within clinical process. In clinical decision making techniques such as Case-Based Reasoning (CBR) or evidence-based medicine, it can provide cases with similar visual appearance to medical doctors. This can supply a second opinion to the medical doctor to assist them in diagnosis.

However, a CBIR system is unable to understand the data as how human can interpret it. There is a semantic gap between low level visual feature and high level semantic concept used by human to interpret images. Normally, humans recognize objects by using prior knowledge on different objects. This knowledge can be based on personal preferences, previous experiences of similar situations, etc. This kind of information is hard to incorporate in CBIR systems. Another constraint of such CBIR systems is that they are impractical for general users to use as users are required to provide query image in order to retrieve similar images.

In order to reduce the semantic gap in medical CBIR system and also to have an effective and accurate medical image retrieval application, new trends for image retrieval using automatic image classification and annotation has been investigated for the past few years. The automatic medical image classification provides a solution to address some of the issues raised by text-related and DICOM-related systems. It is believed that the quality of such medical systems can be improved by a successful classification of medical images based on modality, body region, orientation, etc. by filtering out the images of irrelevant classes and facilitating the search mechanism. For instance, the process to search images for a query like "Find Anteroposterior (AP) Lumbar Spine X-ray image" starts with pre-filtering the database images according to the imaging modality (X-ray), body region (Spine) and orientation (AP). Then, the search could be performed on the set of filtered images to find specific sub-body region such as "Lumbar Spine". The successful classification can also reduce the computational load as well as the false alarm rate of the medical CBIR system because new image can be inserted to the existing medical archive without user interaction. It can also reduce the cost in medical care significantly, because anatomical features or pathologic appearance can be compared in the image database. This would lead to a faster clinical decision process by a physician. As such, a successful annotation and classification can result in better retrieval performance which can be beneficial in teaching, research and clinical decision making. This motivated us to develop a classification framework for medical image databases with high accuracy rate.

However, unlike earlier years of this research that the classification of medical images was restricted to few classes only, this task is challenged when it deals with large archive of medical database. As this field is rapidly developing, it is still in its infancy and there are many other challenges ahead as discussed in following section.

1.2. Current Challenges

Many studies have been done on automatic medical image classification for the past decade but there is still a lot of room for improvement. Automatic medical image classification is one the most challenging and ambitious problems in computer vision. In natural imagery, the analysis is mainly based on color. In many cases of natural imagery, having color background layout is a dominant weight in characterizing image categories. Whereas in medical images such as X-ray images, the background is consistently black.

Therefore, distinguishing image category is highly depending on the success of characterizing the central object region. Another factor that affects the performance of medical image classification is the content of medical images as they are very noisy and prone to a variety of artifacts. Compared to other classification domain, there are some particular difficulties when working on large medical database as follow:

- Defining an appropriate image representation, transitioning from global-based image representation to localized region-based representation is one of the open challenges in this field. This representation must be robust enough to handle large variability of the medical images. In addition, choosing the classifier techniques is also important in order to achieve the maximum classification accuracy.
- Imbalance number of training images among different classes; this is due to the natural commencement of diseases in different part of the body. Therefore it would reflect on probabilities of the routine diagnosis in a radiological clinic. This is the reason that selected body region like Chest have too many images whereas body regions like forearm have less number of images. This would result in having a large and small number of images in different part of the body.

- Intra class variability refers to the high visual variability among images even though they belong to the same category. Intra class variability in a class makes it difficult to find general features for that particular class. We may have an image representing “wrists joint” which is zoomed in where another image was taken from the same body region as a distance. There are two different images representing “hand, wrist joint” is illustrated in Figure 1.1 with different scale.

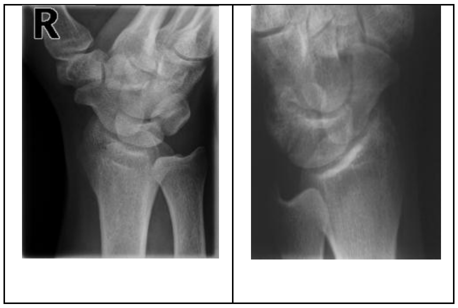


Figure 1.1: Scale Changes in two different images taken from “Hand, Wrist Joint”

Figure 1.2 shows some examples of high intra class variability within the class “(AP, coronal), upper extremity (arm), hand, carpal bones” in ImageCLEF 2007 database.

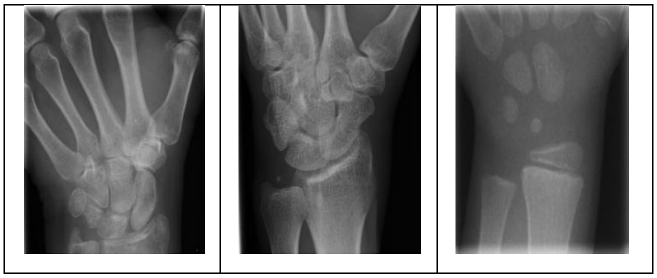


Figure 1.2: Intra-class variability within the class “(AP, coronal), upper extremity (arm), hand, carpal bones”

- Visual similarity or inter-class similarity between images in some classes. This refers to visual similarity among images from different categories. Figure 1.3

represents an example of inter-class similarity. These images belong to two categories which are differing in term of orientation and biological system. The images in upper row belong to class "(AP, coronal), supine-abdomen", whereas the images in lower row refer to class "(PA), upright-abdomen" even though they are visually similar.

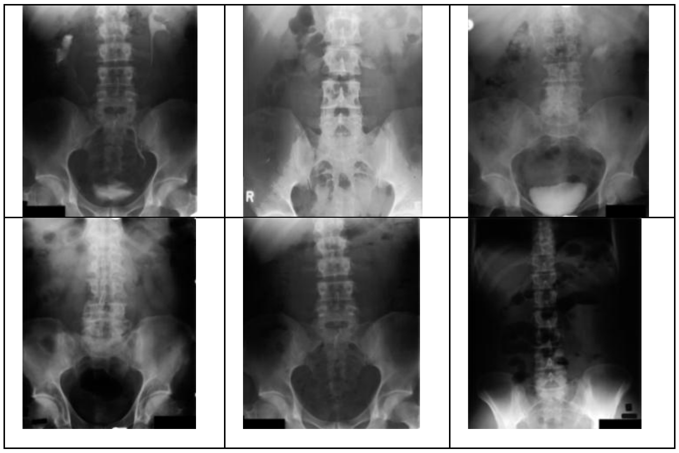


Figure 1.3: Inter-class similarity between two different categories of abdomen

1.3. Aim and Objectives

A set of labeled X-ray images were given from different parts of body and the aim is to construct a classification model. This model is then used to classify any unseen X-ray images into one of the predefined categories.

More specifically, the objectives of this research are as follow:

- To analyze the challenges in classification task of large medical database.
- To explore and analyze various image representation techniques used in medical image classification domain with respect to their performance on the challenge of medical image classification.

- To explore and analyze different machine learning methods exploited in medical domain with regard to their performance on the challenge of medical image classification.
- To design and develop medical classification frameworks to increase the classification accuracy on the entire database as well as every individual class.
- To conduct a set of evaluations in the proposed algorithms with different parameter settings in signifying the strength of the proposed approach to address imbalance number of training data, intra class variability and inter-class similarity

1.4. Contributions

The contribution of this thesis can be divided into four areas as summarized below. Along the thesis, we will present how these contributions resulted in superior classification performance compared with similar works. They allow us to see their impact on classification performance with regard to the challenges of medical image classification systems.

- The most vital component of any classification system is image representation. It is categorized into two main approaches, (i) low-level image representation and (ii) patch based image representation. In this thesis, we have developed an algorithm to conduct different experiments utilizing various image representation techniques to explore their impact on classification performance on the entire database as well as individual categories. The most effective image representation and classification techniques have been identified by analyzing the experimental results.

- We constructed an iterative classification framework that produced four classification models which were constructed from different number of classes. In our approach, we exploited Bag of visual Words (BoW) features, which represents an image by histogram of local patches on the basis of visual vocabulary. As a result, four classification models were generated. The accuracy rate obtained by each generated models outperformed the results obtained by only one model on the entire database.

- We gave a special attention to the classes with high ratio of intra class variability and inter-class similarities. We developed a classification framework via annotation to improve the classification performance of these classes. As such, we utilize three different annotation techniques i.e. Annotation by supervised classification, Annotation by multi-modal Probabilistic Latent Semantic Analysis (PLSA) and Annotation using top similar images. The experimental results showed the significant classification performance on these classes globally and at class specific level.

- We designed a classification framework based on an integration of a generative model such as PLSA with a discriminative classifier such as SVM. In this approach, we are going to show how dimensionality reduction using a latent generative model is beneficial for the task of automatic medical X-ray image classification. The classification process starts with extracting BoW from the entire database. We employed a generative multi-modal PLSA image representation by incorporating both visual features and textual features. A generative PLSA model was applied on the extracted features to help disambiguating the visual words. This is due to the ability of the PLSA model to generate a robust, high level representation and low-dimensional image representation. The classification performance obtained by this framework presents that the number of classes with a higher accuracy rate has been increased.

1.5. List of Publications

During the course of this research, we have published several journal articles and conference papers as listed below:

- Zare MR, Mueen A, Woo CS, (2013), Automatic Classification of Medical X-ray Images using Bag of Visual Word , IET Computer Vision, Volume 7, Issue 2, p. 105 – 114 (ISI Index Publication)
- Zare MR, Mueen A, Awedh M, Woo CS, (2013), Automatic Classification of Medical X-Ray Images: Hybrid Generative-Discriminative Approach, IET Image Processing, doi: 10.1049/iet-ipr.2013.0049 (ISI Index Publication)
- Zare MR, Woo CS, Mueen A, (2013), Automatic Classification of medical X-ray Images, Malaysian Journal of Computer Science, Volume 26, Issue 1, p. 9-22 (ISI Index Publication)
- Zare MR, Mueen A, Woo CS, (2013), Automatic Medical X-ray Image Classification using Annotation, Journal of Digital Imaging doi: 10.1007/s10278-013-9637-0 (ISI Index Publication)
- Zare MR, Mueen A, Woo CS, (2011),Combined Feature Extraction of Medical X-ray Images, IEEE Computer Society, Third International Conference on Computational Intelligence, Communication System and Networks

- Zare MR, Mueen A, Woo CS,(2011),Merging Scheme Based Classification of Medical X-ray images, IEEE Computer Society, International Conference on Computational Intelligence, Modelling and simulation

1.6. Organization of the Thesis

The remainder of the thesis is organized as follow:

- **Chapter 2: Literature Review**

  In this chapter, we review the state of the art in general CBIR and Content-based Medical Image Retrieval. We discussed the challenges of medical image classification and retrieval in a large archive medical database. We presented some ongoing trends and techniques in Semantic based Image Retrieval in medical domain. We discussed different image representation approaches as in general and those image representation approaches have been exploited in medical retrieval domain. We then reviewed and analyzed the impact of different image representation techniques in medical X-ray image classification domain. We explained different techniques for automatic medical image classification such as discriminative and generative techniques. Subsequently, different supervised and unsupervised machine learning techniques are described. To conclude the chapter, we have reviewed and analyzed the classification performance obtained by different researchers with regards to the image representations and classification techniques to explore their impact on the challenges of medical image classification.

- **Chapter 3: Classification Approach using Iterative Filtering**

  This chapter presents an iterative classification framework to improve the classification performance on the entire database as well as on the class specific level. We explain the proposed classification framework step by step. We then evaluate the proposed method's performance by conducting an experiment on a database consisting of 11000 medical x-ray images (training dataset) and 1000 (testing dataset) of 116 classes. The accuracy rate obtained by each generated models outperformed the results obtained by only one model on the entire dataset.

- **Chapter 4: Classification Approach using Annotation**

  In this chapter, we pay a special attention to those classes with high ratio of intra class variability and inter-class similarities. We introduced a classification framework via annotation. The processes involved in classification task of the unseen test image using the proposed framework was given in detail. We then present the experimental results which validate the performance of the proposed approach.

- **Chapter 5: Classification Approach using Hybrid Generative-Discriminative Approach**

  In this chapter, we present a classification algorithm based on an integration of a generative model such as PLSA with a discriminative classifier such as SVM. The classification process starts with extracting BoW from the entire database. A generative PLSA model was applied on the extracted features to help disambiguating the visual words. As a result, high level representation of images is constructed in the generative approach. They are then used as an input to the discriminative classifier in order to construct a classification model.

- **Chapter 6: Conclusion**

Finally, the conclusion of the dissertation is provided in this chapter. We summarized our major contributions and achievements as well as the limitation of our proposed approaches in medical classification domain. Future research direction in this is also recommended.

Chapter 2 Literature Review

2.1. Introduction

In this chapter, the most recent and significant works in the literature on medical image classification will be reviewed. We first discuss the research trends related to Content Based Image Retrieval (CBIR). We then introduce the common approaches in CBIR, followed by some of the existing CBIR system that are developed based on those approaches. We discuss the different ways to represent images in section 2.4.

In section 2.5, approaches of CBIR in medical domains are discussed. In section 2.6, we review and analyze the key methods used as low level image representation in medical classification domain. We pay special attention to the most recent representation method which use bag of word approaches in section 2.7.

We also present the most common classification techniques used in medical domain in section 2.8. The measurement techniques used to evaluate the classification performance are introduced in section 2.9. In section 2.10, we review selected works on image representation and classification techniques with respect to their affect on the classification tasks.

2.2. Content Based Image Retrieval

The first generation of image retrieval systems which was mainly linked to text retrieval system was developed in late 70's (Tamura and Yokaya, 1984). In those systems, all images are annotated manually describing both the content of the image as well as other metadata such as file name, date created, image format, etc. However, manual insertion of these descriptive captions were time consuming and costly. They are also subjective and vary from one person to another. Another problem of such system is that it is impossible to describe some visual properties of the images such as texture or shape by text. As such, the use of more concrete description of visual content is needed that can be related to human perception.

Content Based Image Retrieval (CBIR) enables the elimination of the difficulties that exist in traditional text based query for large image database. The term “content-based” implies that the search will analyze the content of the images rather than the tags and keywords that generally associated with the image. The aim of a CBIR system is to retrieve information that is relevant to the user’s query like text based search engine. In the CBIR approach, retrieval of images is based on features extracted automatically from their visual contents. In following, the common functionalities of CBIR are summarized:

- Image Processing: In this module, certain image enhancement techniques applied on the images to remove noises and increase the contrast. Then low level visual features of the images such as color, shape and texture are extracted locally and globally from all the images.
- Image Representation: upon extraction of visual features of the images, representation of the features in vector form and a notation of similarity are determined, and image is represented as a collection of features.
- Image Retrieval: this function performed based on computing similarity in feature spaces and results are ranked based on the similarity values computed.

2.2.1. Approaches to Content Based Image Retrieval

In general, there are two main approaches in CBIR: (i) *Discrete Approach* and (ii) *Continues Approach* (De Vries and Westerveld, 2004).

i. The discrete approach uses inverted files and text retrieval metrics similar to textual information retrieval. All the features are mapped to discrete features where the presence of a certain image features act like a presence of a word in a text document. One of the popular systems which follow this approach is Visual Information Processing for Enhanced Retrieval (VIPER) system (Squire et al., 1999). This system was developed by University of Geneva Hospitals. This CBIR system has developed a freely available image finding tool like GNU image finding tool.

ii. In continues approach, the images are represented by a feature vector. They are then compared with distance measures. Then those images with lowest distances are retrieved. Most image retrieval systems follow this approach such as CIRES (Iqbal and Aggarwal , 2002), IRMA (Lehmann et al., 2005) and FIRE (Deselaers et al., 2008).

2.3. Difficulties in CBIR Systems

Even after decade of intensive research, the performance of CBIR system is far behind compared to today's search engine. The user expect from an ideal CBIR system to find a meaningful result; this can be the most fundamental challenge that CBIR system faces as there is a mismatch between system's capabilities and user's search requirement. Eakins (2002) proposed three distinct levels of abstraction of search requirement with increasing complexity. First level is based on low-level visual features of images, such as color, texture, shape. For example, find images with a uniform texture pattern and 50% red. In second level, certain degree of logical inference is required. Queries at this level may contain specific object and scenes. For instance, find images containing group of people on the beach. The third level is the highest level of complexity. It requires well understanding of images as well as high degree of reasoning about the meaning and relations of the scenes. This level may contain retrieval of images with emotion. Thus, it is impossible to find images such as “group of happy people on the beach” using low-level visual features without any high-level reasoning. Most of the current CBIR systems mainly operate on first level and partially on second level. Indexing and retrieval at level 3 is only possible currently with help of textual descriptions.

For computers, extracting the semantic content of the image is difficult to understand and process. With the help of prior knowledge on different objects, humans are able to identify objects. This knowledge can be based on personal preferences, previous experiences of similar situations and etc. We can easily identify same objects in an image with different variations based on our previous experiment and reasoning ability. For example, it is easy for physician to identify the semantic similarity between spine images both in coronal view and sagittal view. It is difficult to incorporate this kind of information into CBIR systems. This discrepancy is named as *semantic gap* (Smeulders *et al.*, 2000).

Semantic gap represents the lack of coincidence between the relatively limited descriptive power of low-level visual features and high-level concepts. This causes a constraint in CBIR systems to retrieve images using low-level visual features. It also reduces the efficiency of CBIR systems. In addition, it is not essential for images with similar visual low-level features to have same conceptual meaning. This problem can be rectified by developing a CBIR system which can understand images in the same manner as human perceive the image automatically. However, developing this type of intelligent CBIR system is certainly a challenging task.

2.4. Image Representation

There are many techniques have been used to represent the content of the image in the literature. They are categorized into two main approaches: those techniques that directly extract low level visual features from the images and the other approach represents images by local descriptors. The philosophy of each approach is described as below:

2.4.1. Low Level Image Representation

The first and foremost function in any CBIR system is extracting the visual features of the image (Rui *et al.*, 1999). Since all the images are represented as an unstructured array of pixels, then effective and efficient visual feature are required to be extracted from these pixels. Visual Features are defined as an interesting part of an image. They are referred to any characteristic that describes the content of an image. Appropriate feature representation significantly improves the performance of classification and retrieval systems. Thus, various techniques are developed and studied for feature extractions. Visual features can be extracted either globally or locally. In local approach, the images are divided into a collection of homogenous regions and image description can be obtained from each region.

Images are described in more details using local approaches as compared to global features (Datta *et al.*, 2005) . Global approach use whole image to describe it. (Shen *et al.*, 2005) used the global approach image representation with a special attention on the type of features that must be used. They proposed combined multi-visual features rather than using a single type of feature such as color, shape and texture in order to increase the classification performance. Similarly, combined multi-visual features are also captured using local approach in (Mueen *et al.,* 2010). However, image's foreground cannot be distinguished from background of the image using global approach. Local approach performs better compare with global approach for database with many categories due to existence of intra-class variability on images. The visual features of the images can be extracted and matched with high speed in global approach (Glatard *et al.*, 2004). In both locally and globally approaches, low-level visual features are extracted from image. Low level visual features were categorized into primitive features such as color, shape and texture as explained in following section.

2.4.1.1. Color

Color is one of the fundamental characteristics of the image content and it's one of the most frequently used visual features for content based image retrieval. Color is one of the powerful descriptor that simplifies object recognition.

Histogram is the most commonly used color descriptor technique. Color histogram obtained by quantizing the color space and counting the number of pixels fall in each discrete color. A color histogram is a probability density function. It represents discrete frequency distribution for a grouped data set which has different discrete values. Each image added to the collection is analyzed to compute a color histogram which indicates the proportion of pixels of each color within the image. Then the color histogram for each and every image is stored in the database. During a search, a user can submit a base image where its color histogram is calculated or specify the desired proportion of each color.

Although the most commonly known color space is RGB but HLS (Hue, Lightness, Saturation) and HSV (Hue, Saturation, Value) have been found more effective in order to measure the color similarity between two images.

A RGB model is used to represent all color. It is a 3-Dimentional model and is composed of the primary colors: Red, Green and Blue. They are considered the" additive primaries" since the colors are added to produce the desired color (Muller, *et al.*, 2004). The RGB color space model is represented by a vector with three coordinates. If all the three coordinates are set to one, the corresponding color is white and when all the three coordinates' value is set to zero, the corresponding value will be black. RGB is rarely used for indexing since it does not correspond well to the human color perception (Hengen *et al.*, 2002).

There are many other variants to represent color in CBIR such as color histogram, color coherence vector (Pass *et al.*, 1996),and color correlogram (Jing et al., 1997, Amores et al., 2007, Li et al. 2008). Color is also a key point in morphological diagnosis (Nishibori *et al.*, 2004) but it is not efficient enough for image retrieval when it is used alone as retrieval parameter. (Birgale *et al.* 2006)

2.4.1.2. Shape

Another major image feature is the shape of the object. Shape can represent spatial information that is not obtainable by color and texture histograms. The shape information is determined by the histogram of the edge direction. Edge typically occurs in the boundary between two different regions in an image and has significant local change in intensity. When we consider a large image database, speed reduction can become considerable. There are some notable image retrieval systems where use shape features (Cho & Choi, 2005;

Pawlicki *et al.,* 2007; Zoller & Buhmann, 2007). The effectiveness of using shape features is limited in some CBIR system. However it plays a very important role in simple image database such as trademark retrieval (Zou & Umugwaneza, 2008).

Generally there are two categories of shape descriptors: boundary based and region based (Mehrotra & Gary, 1995). In the boundary based shape descriptor, the focused is on the closed curve that surrounded the shape. There are various models describing this curve such as polygons, Circular arcs, chain codes, boundary fourier descriptor and polynomials. Region based shape descriptor such as Moment Invariants, Morphological Descriptor and etc (Antani *et al.*, 2003; Ye & Androutsos, 2009) , give emphasize to the entire shape region or the materials within the closed boundary.

There are many features that can be used in characterizing shape for object recognition and image retrieval. One of the useful approach is canny edge detector (Canny, 1986) is used to generate edge histogram which were used as edge detection techniques in various studies (Dimitrovski *et al.*, 2011; Mueen *et al.*, 2008; Zare *et al.*, 2011). It is considered to be an optimal edge detector due to its good detection, good localization as well as minimal response.

2.4.1.3. Texture

There is no formal definition for the image texture. Texture is believed to be a rich source of visual information. Texture that inherent all surfaces describe visual patterns, where each contains property of homogeneity. It accommodates important information about the structural arrangement of the surface such as leaves, tree barks, water, clouds, etc. Texture also describes the relationship of the surface to the surrounding environment .It is a feature that describes the distinctive physical composition of a surface. A major characteristic of texture is repetition of pattern or patterns over a region in an image. Typical textural features include contrast, uniformity, coarseness, roughness, frequency, density and directionality.

The existing texture descriptors are classified based on the three principal approaches as described below:

- The first approach is statistical techniques. This approach characterizes textures by using the statistical properties of the grey levels of the pixels containing a surface image. These properties can be computed by the wavelet transformation of the surface as well as the gray level co-occurrence matrix of the surface.
- The second approach is structural techniques that characterize textures as combination of simple primitive structures which is named as texture elements.
- The last approach is spectral techniques which are based on properties of the Fourier spectrum and describe global periodicity of the grey levels of a surface by identifying high-energy peaks in the Fourier spectrum.

2.4.2. Local Feature Extraction

Recently, more promising studies have been focused on local image descriptor within computer vision community. Local image descriptors are getting more attention for the past few years. These approaches have been used successfully in object recognition and classification task (Andre *et al.*, 2010; Bosch *et al.*, 2006a; Dorke, 2006; Fei-Fei and perona, 2006; Nowak *et al.*, 2006; Tommasi et al, 2007). The extraction of local features is the most important steps of any object classification method. In local feature extraction, two decisions have to be made: (i) where are the features extracted and (ii) how is the image regions represented.

(i) Where are the features extracted?

To answer the first question, one of the possible approaches is to extract patches at each position in the image i.e. grid points where local features are extracted on a regularly spaced grid. This would leads to a tremendous amount of data. Other solution is to use randomly sampled points. Extracting local features at randomly sampled position is very simple but they are not reproducible. Another way to decide at which location the local features should be extracted is to use of interest point detectors.

Local interest point detectors are designed to concentrate on points that hold distinctive information in their local surrounding area and whose extraction is stable with respect to geometric transformations and noise. It is essential to obtain local interest areas as invariant as possible to image transformations in order to have stable image representation. The more invariant the local interest area, the more confident we are that the local extracted information will remain the same from image to image. Several interest point detectors exist in the literature. They are defined according to the type of local structure they discover as well as their degree of invariance to image transformations which are explained in section 2.4.3.

(ii) How are the image regions represented?

After detection of the interest points, some kinds of descriptors need to be computed to represent image around those interest points. The most famous local descriptor is Scale Invariant Feature Transform (SIFT) proposed by Lowe (Lowe D., 2004). It is one of the most widely used local descriptor that has been used as one of the major feature extraction techniques in object recognition tasks (Kouskouridas *et al.*, 2012; Morales *et al.*, 2011).

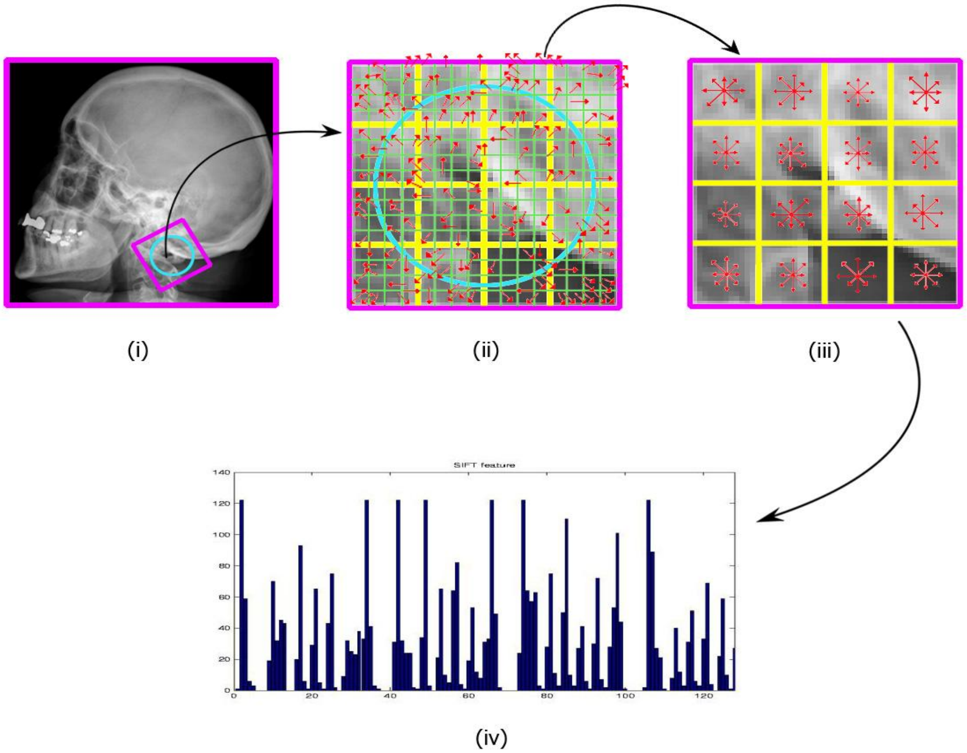


Figure 2.1: Illustration of SIFT extraction process

Each SIFT descriptor is used for extracting distinctive feature that characterizes a set of keypoints for an image. This strategy uses a keypoint detector based on the identification of interesting points in the location-scale space. The keypoints or salient points are the extreme points in the scale space of the image. The keypoints are generated by applying Difference of Gaussian (DoG) (Lowe, 1999) point detector to the image. For each keypoint a distinctive local feature vector is established by computing the gradient magnitude and orientation in a region around the key point location. Each detected region is represented with the SIFT descriptor with the most common parameter configuration: 8 orientations and 4 × 4 blocks, resulting in a descriptor of 128 dimensions. The implementation of SIFT detector is easy and fast to implement due to its low complexity. SIFT detector is also invariance to small errors in the calculation of the position and area.

The extraction process of SIFT feature as shown in Figure 2.1, is summarized as below:

Procedure of SIFT feature extraction

Step 1: Gaussian smoothed image is selected. It is corresponding to the local interest point's characteristic scale.

Step 2: image gradient are sampled based on the scale and orientation of the local interest point. This is done using a regular grid around the local interest point location

Step 3: sampled gradient's orientation are normalized with relation to the local interest point's orientation

Step 4: Gaussian weighting are applied to the gradient's magnitude

Step 5: Gradients orientation is quantized into n orientations where n = 8

Step 6: Grid division orientation histograms is created to accumulate the magnitude of the previously quantized local gradient

Step 7: Then concatenate the Grid histograms into one histogram to form a vector

Step 8: Feature vector are then normalized to further increase the illumination invariance

2.4.3. Example of Interest Point Detector

Several interest points detector are described and analyzed in this section. They can be categorized into three groups: blobs, corners and wavelet-based.

2.4.3.1. Blobs

The notation of “blobs” in the image has formalized by Lindeberg (Lindeberg, 1998).It is a circular image region containing similar intensity values. The blob’s centers are often used as interest points. An example of interest point in this category is Difference of Gaussians (DoG). It was proposed in (Lowe, 1999) to select key location at maxima and minima of the difference of Gaussian operartor in scale space and the result from neighbouring scales are subtracted from each other. As a result, upon identification of local extrema , the position and scale of these extrema are used to determine interest points in the images. The advantage of these points is that they are robust with respect to scale, translation, and rotation. The DoG operator can locates key points at regions and scales of high variation, making these locations particularly stable for characterizing the image.

2.4.3.2. Corner

Another type of image feature that used as interest point is corner. Corners are defined as points of high curvature of the intensity surface of the image (Zhang *et al.*, 1995) . They are well defined in two directions and contain a good amount of edge information. The Harris detector is one the most popular corner detector which is proposed by Harris and Stephens (1998). It is an improved version of another technique proposed by Moravec (Morevec, 1977). Moravec’s corner detector was based on the auto-correlation function of the signal while Harris detector computes the auto-correlation matrix at each pixel in the image, and then those matrixes with high eigenvalues will be taken as interest points. One of the drawbacks of the Harris detector is that it only finds interest points that correspond to the finest scale plane of the image, meaning that it has no inherent notation of scale.

2.4.3.3. Wavelet-based Detector

Another type of interest point detectors uses wavelet which was proposed by Loupias *et al.* (2000). In this method, information about the variations in the image at different scales is given in wavelet representation.

Three wavelets are applied for each scale in 2 Dimensional image signal such as horizontal, vertical, and diagonal. This results in a set of wavelet representation for each scale and each orientation. 2 Dimensional wavelet transform of the image was used to identify interest points in an approach proposed by Sebe *et al.* (2003). The wavelet coefficients are tracked across scales from the coarsest scale to the original pixels and the saliency value of a pixel is set to the sum of all the coefficients in its track. As such, the resulting map is the threshold to obtain interest pints.

As for each pixel, the saliency is computed as the sum of the absolute values of its wavelet coefficients. Those pixels with highest saliencies are chosen as interest points. A multi-scale decomposition of an image is computed using 1 dimensional wavelet in Shokoufandeh *et al.* (1998). As a result, the interest points are the local maxima of the sum of wavelet responses.

2.5. Medical Image Retrieval Techniques

Medical image databases are the key component in diagnosis and preventive medicine. Due to the explosive growth of digital technologies in modern hospital, many medical images in diverse modalities are acquired every day (Müller *et al.*, 2004; M. Rahman *et al.*, 2004). Computed Tomography (CT) images, Magnetic Resonance Images (MRI), ultrasound (US) images, X-ray images are examples these modalities as shown in Figure 2.2.

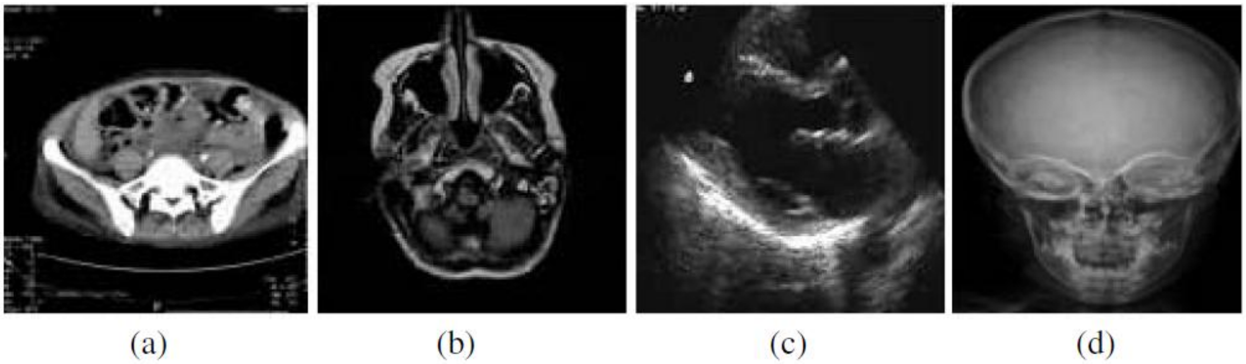


Figure 2.2: Sample images from different modalities (a) CT, (b) MRI, (c) Ultrasound, and (d) X-ray images.

As a result, searching and indexing large medical image databases efficiently and effectively has become a challenging problem. Thus, the need for an efficient retrieval tools is unavoidable. A large amount of research has been carried out on medical image retrieval in the last two decades. Generally there are three types of approaches in medical image retrieval. The first approach is traditional text based image retrieval where textual features such as file names or keywords have been used to retrieve similar images. Currently, many hospitals and radiography departments are equipped with Picture Archiving and

Communications systems (PACs). Such traditional systems have many limitations due to the usage of Digital Imaging and Communications in Medicine (DICOM) header as the searches are carried out according to the textual attributes of image headers. DICOM contains tags to translate the examined body part, patient position and the acquisition modality. Some of this important information are captured and set by digital system automatically, while others are set manually by physician or radiologist. This approach can be unreliable as some of the information might be missing, or entered wrongly. Even though DICOM header contains many important information, but it still remains suboptimal due to its high error rate reported in recent studies (Güld *et al.*, 2002). In such approach, each medical image in the database is annotated manually by physician or medical expert. However, manual annotation of images is not only time consuming, but also error-prone because it is subjective task due to human perception.

The second approach is based on Medical Content-Based Image Retrieval (CBIR). Medical CBIR enables the elimination of the difficulties that exist in traditional text based query for large image database. Medical 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. However, there is a semantic gap between low level visual feature and high level semantic concept used by human to interpret images. Normally humans recognize objects by using prior knowledge on different objects. This knowledge can be based on personal preferences, previous experiences of similar situations and etc. This kind of information is hard to incorporate in Medical CBIR systems. Another constraint of such Medical CBIR systems is that they are impractical for general users to use as they are required to provide query image in order to retrieve similar images.

The third approach for Medical Image Retrieval is Automatic Medical Image Annotation (AMIA) where medical images can be retrieved in similar ways as text documents. The main goal of AMIA approach is to learn the semantic concept from every image in the database and use the concept model accordingly to label the new image. Once medical images are annotated with semantic labels, they can be retrieved by keywords, which is similar to text document retrieval. AMIA can be defined as a supervised medical image classification. This approach is mapping a new image into pre-defined categories and annotates them by propagating the corresponding words of that category. The most important characteristics of AMIA are the searching and retrieval techniques as it employs the advantages of both text-based annotation and CBIR.

As such, the research on medical content-based image retrieval has progressed for the past decades. It is believed that the quality of such medical system can be improved by a successful classification of medical images. Properly classified medical image data can help medical professionals in fast and effective access to data in their teaching, research, training and diagnosis by filtering out irrelevant images. For instance, the process to search images for a query like "Find Anteroposterior (AP) Lumbar Spine X-ray image" starts with pre-filtering the database images according to the imaging modality (X-ray), body region (Spine) and orientation (AP). Then the search could be performed on the set of filtered images to find specific sub-body regions such as "Lumbar Spine".

However, unlike earlier years of this research that the classification of medical images was restricted to few classes only, this task is challenged when it deals with real-life constraints of content-based image classification in the medical domain. This is where the ImageCLEF medical image annotation challenge was born. The goal of this challenge is to categorize the images into pre-defined classes automatically and to assign correct labels to unseen test images. It involves some basic principles such as representation, where visual feature of the images are extracted; and generalization which is training and evaluating the classifier.

In this thesis, special attention was given to the two major aspects of AMIA which are Image Representation and Automatic Classification. In the literature, many techniques have been used to represent content of medical images. The techniques can be categorized into two main approaches: (i) those which directly extract low level visual features from images and (ii) those methods which represent images as Bag of Words. The philosophy of each approach used for medical image representation and automatic classification as well as their related work is described below:

2.6. Low Level Image Representation Techniques

The most crucial component of automatic medical image classification system is extraction of the visual features of the images. Appropriate feature representation significantly improves the performance of semantic learning techniques. This section reviews commonly used low level image representation techniques in medical X-ray images and describes the related works.

2.6.1. Raw Pixel Value

It is the simplest form of image representation technique. In this approach, the images are scaled down to common size and represented by a feature vector that contains image pixel values. It is argued that a good classification and retrieval performance can not be obtained by utilizing image pixel value, because it is not an easy task to find which two pixels should be evaluated when comparing two images (Viitaniemi & Laaksonen, 2006). However, it has been shown that this method serves as a reasonable baseline for classification and retrieval of simple images with few objects such as medical X-Ray images and character recognition (Keysers *et al.*, 2007).

2.6.2. Gray Level Co-occurrence Matrix

Gray-Level Co-occurrence Matrix (GLCM) is one of the well-known statistical tools for extracting texture information of an image which is introduced by (Haralick *et al.*, 1973). It provides information about position of pixels having similar gray level values. GLCM extracts contrast, energy, homogeneity and entropy features of the image along four different directions ($\theta \in$ {0◦, 90◦, 45◦, and 135◦}). For an image *f (i,j)* of size $L_r \times L_c$ with a set of N_g gray levels:

Contrast is the variations of gray level in the image:

$$contrast(d,\theta) = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i-j)^2 p(i,j,d,\theta) \quad (2.1)$$

Energy is a measure of textural uniformity:

$$Energy(d,\theta) = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j,d,\theta)^2 \quad (2.2)$$

homogeneity refers to homogeneity of variance that is similar to the first order statistical variable called standard deviation:

$$Homegeneity(d,\theta) = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{1}{1+|i-j|}p(i,j,d,\theta) \tag{2.3}$$

Correlation measures the disorder or randomness of an image using:

$$Correlation(d,\theta) = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}(i-\mu_x)(j-\mu_y)p(i,j,d,\theta)}{\sigma_x\sigma_y} \tag{2.4}$$

where d and θ denote the distance and the orientation aligning between pixel (X_1, Y_1) and pixel (X_2, Y_2) in the image; μ_x , σ_x μ_y and σ_y are representing mean and standard deviation of pixel value in the row and column direction of the GLCM respectively.

2.6.3. Local Binary Pattern

Local Binary Pattern (LBP) introduced by (Ojala *et al.*, 2002) is a gray-scale invariant local texture descriptor with low computational complexity. It is one of the best performing texture descriptors and it has been used wildly in various applications. Figure 2.3 illustrates the basic LBP operator. The LBP operator assigns a label to every pixel of an image by thresholding the 3 × 3-neighborhood of each pixel with the center pixel value and considering the result as a binary number. Then the histogram of the labels can be used as a texture descriptor.

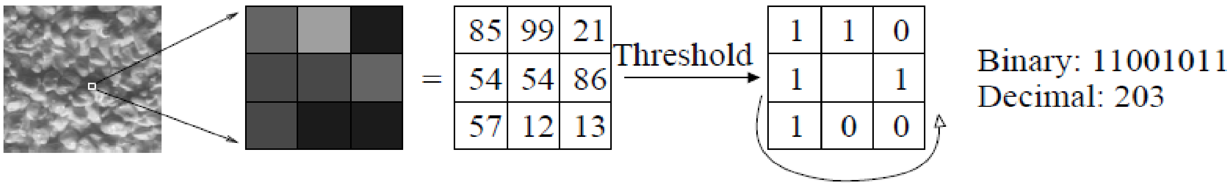


Figure 2.3: Illustration of the Basic LBP Operator

LBP operator can be extended to use neighborhoods of different size in order to deal with textures at different scales as shown in Figure 2.4. The neighborhood is formed by a symmetric neighbor set of *P* pixels on a circle of radius *R*. Formally, given a pixel at (x_c, y_c), the resulting LBP code can be expressed in the decimal form as follow:

$$LBP_{P,R}(x_c, y_c) = \sum_{n=0}^{P-1} s(i_n - i_c)2^n \qquad (2.5)$$

where *n* runs over the *P* neighbors of the central pixel, i_c and i_n are the gray level values of the central pixel and the neighbor pixel, and s(x) is 1 if x$\geq$ 0 and 0 otherwise. After labeling an image with a LBP operator, a histogram of the labeled image $f_1(x, y)$can be defined as:

$$H_i = \sum_{x,y} I(f_1(x, y) = i), \qquad i=0,\ldots,L-1 \qquad (2.6)$$

Where L is the number of different labels produced by the LBP operator, and

$$I(A)=\begin{cases} 1, & \text{if A is true} \\ 0, & \text{if A is false} \end{cases} \qquad (2.7)$$

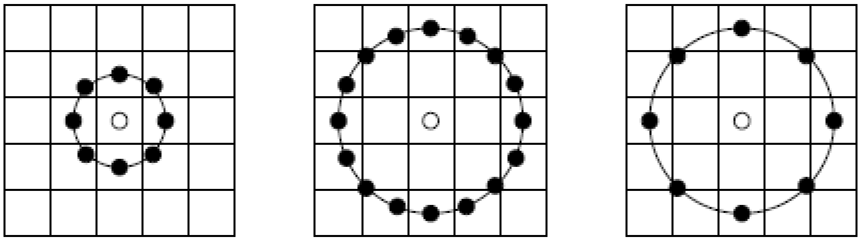


Figure 2.4: Illustration of the LBP Operator with circular neighborhoods

2.6.4. Low Level Image Representation Techniques Applied on Medical Image Classification

This section reviews some of the recent publications focused on low level image representation in medical X-ray images. Analysis of texture centered in X-ray images is necessary because it produces high resolution gray level intensities where textures play an important role. So, feature extraction descriptors are needed with the purpose to identify and select a set of distinguishing and sufficient features to characterize X-ray image.

Majority of medical images of different modalities can be distinguished by their texture characteristics (Lehmann et al., 2005). However, as X-ray images are gray level images and do not contain any color information, the related CBIR systems mostly deal with textures for feature extraction process which were used by many researcher (Bo Q. et al., 2005; Jeanne et al., 2009; Ji-Hyeon et al., 2011; Ko et al., 2011; Muller et al., 2005; Rahman et al., 2011a; Sohail et al., 2010; Zare, et al., 2011) .

Seong-Hoon *et al.* (2010) proposed classification of medical x-ray images where texture information of medical x-ray images are extracted using LBP; the classification accuracy obtained outperforms others who use edge histogram descriptor. Combination of block based local binary pattern with edge histogram was used as medical image representation for the task of automatic medical image annotation in ImageCLEF 2007 (GuangJian, *et al.*, 2008) . Zhy *et al.* (2008) applied segmentation methods on ultrasound medical images based on texture features which are obtained according to GLCM. GLCM were combined with histogram moments as feature extraction to classify ultrasound medical images in (Sohail, *et al.*, 2010). Whereby, Jeanne *et al.* (2009) proposed an Automatic Detection of Body Parts in X-Ray Images. Their experiments on the IRMA database revealed that local binary pattern performs better than average gray descriptor, color layout descriptor and gray-level co-occurrence matrix. Hierarchical classification schemes based on individual SVMs trained on IRMA sub-codes for the task of automatic annotation of ImageCLEF 2007 medical database was proposed in (Unay *et al.*, 2010). The result obtained outperforms the other single classifier schemes.

Bo *et al.* (2005) focused on automatic annotation of ImageCLEF 2005 which can be taken into account as multi-class classification problems. They have started with extracting texture features and regional features such as blobs from x-ray images. These features were then normalized and stacked to form one dimensional feature vector as an input to (SVM) classifier with radial basis function (RBF) kernels. Another widely used strategy is combining different local and global descriptors into a unique feature representation. (Zare *et al.* 2011) combined local binary pattern with gray-level co-occurrence matrix (GLCM), canny edge detector and pixel level information of X-ray image as for feature extraction. They used a 15 × 15 down scaled representation of the images as it was recommended by Mueen *et al.* (2008). Other authors have also combined pixel value as a global image descriptor with other image representation techniques to construct feature vector of the image (Deselaers & Ney, 2008; Dimitrovski*, et al.*, 2011; Tommasi*, et al.*, 2008).

The main advantage of the low-level visual features used in all the above works is that they provide a simple image representation. However, low level image representation are not enough discriminative amongst different classes of images when there is high ratio of intra class variability or remarkable background clutter in images.

2.7. Bag of Words Image Representation

Recently, more promising studies have been focused on local features. SIFT feature has been used as one of the most widely local descriptor in object recognition tasks. With the advances of this local feature, researchers in the field of computer vision have attempted to resolve object classification problems by new approach known as Bag of Words (sometimes also called bag of features or bag of visterns). In recent years, many studies have successfully exploited this feature in general scene and object recognition tasks (Deselaers & Ney, 2008; Jie *et al.*, 2011; Kesorn & Poslad, 2012; Lazebnik *et al.*, 2006; Sui *et al.*, 2012; Teng *et al.*, 2011; Yang *et al.*, 2012; Zhou *et al.*, 2012). The use of Bag of Words (BoW) model can also be found in medical image classification and retrieval tasks (Deselaers et al., 2006; Dimitrovski et al., 2011; Jingyan et al., 2011a, 2011b; Markonis et al., 2012). In the following section, the construction of BoW is explained in detail.

2.7.1. Bag of Words Construction

As illustrated in Figure 2.5, the procedure of implementing BoW generally includes three main steps:

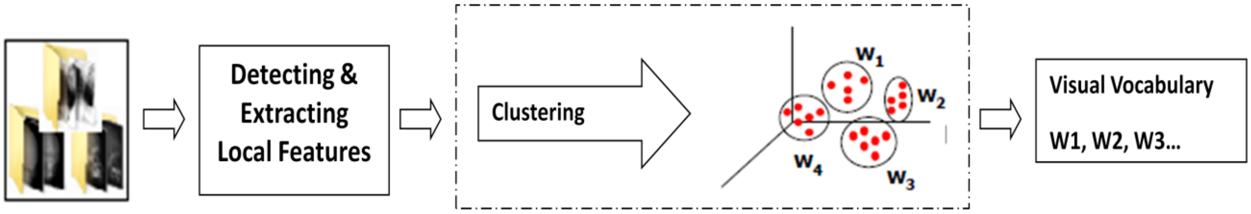


Figure 2.5: Construction of Bag of Visual Words

A) Detect and extract local features;

B) Vector Quantization;

C) Represent an image into a histogram using the codebook.

A) Detect and extract local features: Feature detection is the process in which the relevant components of an image must be identified. Usually, the goal of feature detection is set to identify a spatially limited image region that is salient or prominent. As we discussed in previous section, there are several interest point detectors in the current literature. They vary mostly by the type of image structures they are created to detect and the amount of invariance they theoretically ensure, and achieving invariance by the image property that they exploit. Difference of Gaussians (DoG) is one of the well known point detector used for local interest point detection task. Upon detection of local interest points, SIFT feature is used to represent the description of local interest area detected by DoG.

B) Vector Quantization: The visual dictionary or codebook is built using a clustering or vector quantization algorithm. This step usually uses *k*-means clustering method. It is one of the simplest but well known clustering algorithms. Given an image *d* with a set of features $F(d)$=$\{f_j, j = 1, \dots, N_{f(d)}\}$, *K*-means algorithm performs clustering on these features vector through a set of cluster *k* fixed a priori. The algorithm then randomly chooses *k* points in that feature vector which are used as the initial centers of the clusters. Then the algorithm starts partitioning the feature space into *N* regions. The steps of clustering algorithm can be summarized as below:

1. Randomly place *k* points into the extracted feature vector. These points serve as the initial group centroids.
2. Assign each feature vector to the group that has the closest centroid.
3. When all feature vectors have been assigned, recalculate the positions of the *k* centroids.
4. Repeat Steps 2 and 3 until the centroids no longer move.

The hyper-parameter *k* denotes the size of our vocabulary, which defines the visual coherence and the diversity of our visual words, and consequently of the BoW image representation. Increasing *k* will result in a finer division of the feature space, meaning that the visual words will be more specific. In principle this would produce a more precise representation of our image. However, there is the danger that this may result in an over-segmentation of the feature space, i.e. with several visual words representing the same local content. On the other hand, a small *k* will have visual words representing larger regions of the feature space, making the visual words less specific but making the image representation more stable across similar images.

C) Represent an image into a histogram using the codebook: Once the cluster centers are identified, each feature vector in an image is assigned to a cluster center using nearest neighbor with a Euclidean metric and finally each image is represented as histogram of these cluster centers by simply counting the frequency of the words appear in an image.

2.7.2. Bag of Word Representation Applied on Medical Image Classification

With increasing size of medical X-ray archives, it is important to have simplistic, discrete representations and simple matching measures to preserve computational efficiency. It is argued that BoW paradigm provides efficient means to address the challenge of CBIR system in large in large size databases such as the one in ImageCLEF (Avni, *et al.*, 2011). They proposed X-ray image categorization and retrieval based on local patch representations using a "bag of visual words" approach. They analyzed the effects of various parameters on system performance. The best result was presented using dense sampling of simple features and a nonlinear kernel-based support vector machine. This was an extension of another work where visual words dictionary were generated to represent X-Ray chest images (Avni, *et al.*, 2010). Deselaers *et al*. in (Deselaers *et al*., 2006) extracted features from local patches of different size which were taken at every position and were scaled down to a common size. In that work, rather using dictionary, the feature space was quantized uniformly in every dimension and the image was represented as a sparse histogram in the quantized space. In (Zhi *et al.*, 2009) , authors developed a medical image retrieval method using SIFT features. They proposed three different methods forming visual words from SIFT features. Upon extracting SIFT descriptor from local key points, K-means clustering is used to construct visual vocabulary in three ways; first method is to take features of all images as an input to k-means cluster to get all cluster centers. Second method is based on body part sub-selection where visual words are derived using k-means clustering from each of the body part in every category. The last method is nearby slice subset. In this approach, each of the image category's nearby slices are derived to construct the visual words. Similarity measure is then employed to compute similarity between query image and BOW made from images in database. Experimental result conducted on six body part shows that visual words constructed based on nearby slice subset is suitable for medical image retrieval.

The combination of local and global features was used to address the problem of intra class variability and inter-class similarity for the task of medical image classification in (Tommasi, *et al.*, 2008). They integrated two different local cues that describe structural and textual information of image patches. In another medical image classification work, BoW was combined with other image representation techniques such as LBP, Pixel value and Edge Histogram Descriptor with two different feature fusion schemes; Low Level and High Level (Dimitrovski, *et al.*, 2011). The results obtained by this group clearly show that feature fusion methods outperform the results obtained by using a single feature in classification task. However, they have analyzed the results obtained by different feature extraction techniques and its proven that BoW features perform better than other feature representation used in that work.

2.8. Automatic Medical Image Classification and Annotation

Automatic image annotation is the process of providing a textual annotation which describes the main visual concept represented in the image and automatic image classification is the process of categorizing images according to some concepts. They are inter-dependent on each other. Image annotation is considered as a classification problem. In addition, image annotation can also be considered as a possible solution of semantic gap problem. As such, image annotation has become an increasingly important and active research area in the field of machine learning and pattern recognition. The main advantage of automatic medical image annotation system is that users find it easy to express a query in textual terms.

Therefore, image annotation can be performed by classification of image pattern into predefined classes or categories and then keywords or class labels are automatically assigned to the image. Once image annotation is done, image retrieval with keywords can be an easy task.

The classification process is carried out in two different phases; training/learning phase and testing phase. In training phase, after storing images into the database, low-level visual features of the images are extracted locally and globally.

Generally, there are two models used for training phase: Discriminative Model and Generative Model. In discriminative model, feature vectors or input variables are directly maps to output variables (labels) in order to perform classification whereas in Generative Model, the likelihood of the data is employed for distribution of features and learning process.

In testing phase, visual feature is extracted from the test image locally and globally. Then, a classifier decides on the bases of learning model (Discriminative or Generative) as to which class that feature vector belongs.

2.8.1. Discriminative Model

Discriminative models are class of models used in machine learning for modeling the dependence of an unobserved variable y on an observed variable X. In another word; let's say we have input data X and we want to classify the data into labels y. A discriminative model learns the conditional probability distribution $p\ (y|X)$. There are many discriminative models available such as Support Vector Machine (SVM), K-nearest Neighbour (KNN), Artificial Neural Network (ANN), and Decision Tree. It is very difficult to say which model is better. Their classification performance are varies depending on a specific problem. Among them, SVM and KNN have shown a better generalization performance in medical domain compared with other classification techniques (Pourghasem & Ghasemian, 2008; Rahman *et al.*, 2007; Mueen *et al.*, 2008, Muller *et al.*, 2004):

2.8.1.1. Support Vector Machine

Support Vector Machine (SVM) is a kernel based technique that represents one of the major developments in machine learning algorithm. SVM is a group of supervised learning methods that can be applied for classification and regression. It learns by example to assign labels to objects. Support vector machine has shown its capacities in pattern recognition and a better performance in many domains compared with other machine learning techniques. SVMs have also been successfully applied to an increasingly wide variety of biological applications.

SVM algorithm takes a set of input data points. It then decides that the data point belong to which possible two classes. The aim is to construct a hyperplane or set of hyperplanes in a high or infinite dimensional space that classifies the data more accurately. Therefore, the basic idea is to find a hyperplane that has the greatest distance to the nearest training data points of any class.

A hyperplane can be defined by the following equation:

$$w\,x + b = 0 \tag{2.8}$$

Where *x* is the data point lying in the hyperplane, *w* is normal vector to hyperplane and *b* is the bias. Figure 2.6 shows the basics of SVM.

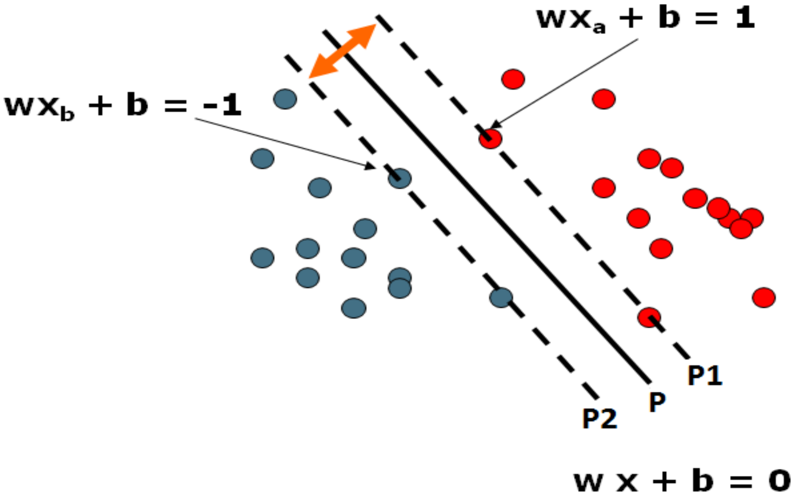


Figure 2.6: The basics of classification by SVM

The idea is to separate two classes (red circle and blue circle, each are labeled as 1 and -1 respectively. Those circles which lie on *p1* and *p2* are support vectors.

For all data points from the hyperplane *p* $(wx + b = 0)$, the distance between origin and the hyperplane p is $\frac{|b|}{||w||}$.

We consider the patterns from the class -1 that satisfy the equality $wx + b = -1$, and determine the hyperplane p1; the distance between origin and the hyperplane *p1* is equal to $\frac{|-1-b|}{||w||}$.

Similarly, the patterns from the class +1 satisfy the equality $wx + b = +1$, and determine the hyperplane *p2*; the distance between origin and the hyperplane *p2* is equal to $\frac{|+1-b|}{||w||}$. Of course, hyperplanes *P, P1,* and *P2* are parallel and no training patterns are located between hyperplanes *P1* and *P2*. Based on the above considerations, the margin between hyperplanes *P1* and *P2* is $\frac{2}{||w||}$.

In order to use the SVM methodology to handle the classes are not linearly separable, then the input vectors such as low level feature vectors are mapped to higher dimensional feature space H via a nonlinear transformation, Φ=$R^d \rightarrow H$. The kernel function $K(x_i, x_j)$ is used then to construct optimal hyperplane in this high dimensional feature space. This kernel function is a products of input vector x_i *and* x_j where $K(x_i, x_j) = \Phi(x_i).\Phi(x_j)$.

Radial Basis Function (RBF) kernel and Polynomials kernel are the most common mappings:

Polynomials kernel: $K(x_i, x_j)$=$(x_i.x_j + 1)^q$ (q is the degree of Polynomial)

Radial Basis Function (RBF) kernel: $K(x_i, x_j) = e^{\frac{||x_i - x_j||^2}{2\sigma}}$ (σ is the Gussian sigma)

The above description of SVM is designed for binary classification which only deals with two class labels +1 and -1. It can be extended to multi-class classification where each data point x can be mapped to a specific label y from the set of available labels. As such, there are two techniques for multi-class classification using SVM such as “one against one” and “one against all”.

One-Against-All: in this case, one SVM is constructed per class, which is trained to distinguish the samples of one class from the samples of all remaining classes. For instance, supposed we have four classes A1, A2, A3 and A4. Accordingly four SVMs are constructed. To classify A1, the respected SVM model compares A1 with all other. The same process applies on classification of A2, A3 and A4.

One-Against-One: In this case, one SVM is constructed for each pair of classes. As such, for N classes, *N (N-1)/2* SVMs are trained to differentiate the samples of one class from the samples of another class. Suppose we have four classes A1, A2, A3 and A4. Then according to this strategy, six SVMs are trained, i.e. the six SVMs classify A1 or A2, A1 or A3, A1 or A4, A2 or A3, A2 or A4, and A3 or A4 respectively.

SVM is an emerging machine learning technology that has already been successfully used for image classification in both general and medical domain (Rahman et al., 2011b; Setia et al., 2008; Wei et al., 2012; Yimo et al., 2011). For example, (Bo et al., 2005) employed SVM with RBF kernel function for classification of medical X-ray images. They combined the blob region feature and three low resolution pixel maps to form a one dimensional feature vector to use as an input to SVM classifier. In (Deselaers & Ney, 2008), Bag of word approach based on local descriptor was extracted from medical X-ray images. The histogram generated using BoW was classified using SVM. A full-body pedestrian detection scheme was proposed by Mohen (Mohen *et al.,* 2001). The first SVM classifier was used to detect the body parts, such as heads, arms and legs. Then, a second SVM classifier integrating those detected parts was used to make the final detection decision.

2.8.1.2. K-Nearest Neighbor

K-nearest-neighbor (KNN) classification is one of the most primary and simple classification methods and should be one of the first choices for a classification study when there is little or no prior knowledge about the distribution of the data. KNN classification technique considers a simplest method conceptually and computationally (Duda et al., 2001). It works based on a majority vote of k-nearest neighbor classes. To understand how the KNN works, let’s consider the task of classifying a new object among a number of known existing objects. In Figure 2.7, existing objects are represented by plus and minus signs, and red circle depicts a new object. Our task is to if the new object (red circle) can be classified as existing objects (plus and minus). To do this, first the distance between the new object and existing objects will be determined. Then, KNN classifier takes only k-nearest neighbor objects which are closer to the new object. For example, in k=1, KNN algorithm classifies the new object with a minus sign since it is a closet object to the new object. In the case where k=2, the classification output is unknown, because the second closest object to the query point (red circle) is plus and so both the plus and the minus signs achieve the same number of votes. In this case of k=5, the KNN classifier algorithm will take majority vote of its 5 nearest neighbor. As shown in Figure 2.7, there are 2 and 3 plus and minus signs respectively which is surrounded by the circle. As such, KNN will assign a minus sign to the outcome of the query point

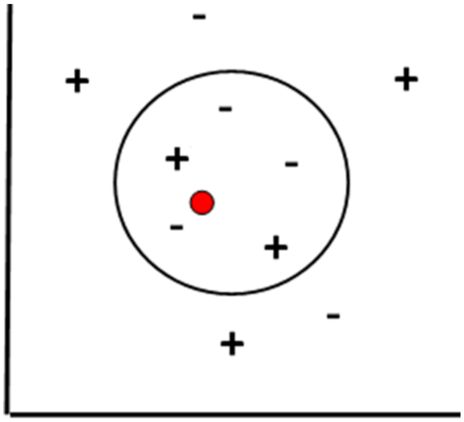


Figure 2.7: K-Nearest Neighbour Classification

A combination of low-level global texture features with low-resolution scaled images and KNN classifier was used for automatic classification of medical X-ray images into 81 categories (Lehmann *et al.,* 2005). In (Inthajak *et al.* 2011), KNN algorithm was used for object detection and applied to distinguish classes of the medical image's blob. The MedGIFT team in (Garcia Seco de Herrera *et al.,*2012) used KNN with Bag of word to classify medical images.

2.8.2. Generative Model

Generative model is a model for randomly generating observable data. It is used in machine learning for either modeling data directly or as an intermediate step to forming a conditional probability density function. Generative models contrast with discriminative models, in that a generative model is a full probabilistic model of all variables, whereas a discriminative model provides a model only for the target variable(s) conditional on the observed variables. Some examples of generative model are Latent Dirichlet Allocation (LDA) (Blei *et al.,* 2003), Naive Bayes classifier, Probabilistic Latent Semantic Analysis (PLSA) and etc.

2.8.2.1. Probabilistic Latent Semantic Analysis

The PLSA was originally proposed by Hofmann (Hofmann, 2001) in the context of text document retrieval. It has also been applied to various computer vision problems such as classification, images retrieval, where we have images as documents and the discovered topics are object categories (e.g. airplane, sky). In this section, PLSA model explained in terms of images, visual words and topics.

The key concept of the PLSA model is to map the high dimensional word distribution vector of a document to a lower dimensional topic vector. Therefore PLSA introduces a topic layer between images and words. Suppose we have a set of images D = $d_1, \dots, d_N$ with words from visual vocabulary X. Each image consists of mixture of multiple topics and thus the occurrence of words is a result of the topic mixture. PLSA assumes the existence of a latent aspect $z_k (k \in 1, \dots, N_z)$ in a generative process of each word $x_j (j \in 1, \dots, N_x)$ in the image $d_i (i \in 1, \dots, N_d)$. The PLSA model is parameterized by $P(z_k|d_i)$ and $P(x_j|z_k)$. $P(x_j|z_k)$ denotes the probability of visual word x_j in topic z_k. $P(z_k|d_i)$ denotes the probability of topic z_k given in document d_i.

Using these definitions, the document is generated as follow:

- Select an image d_i with probability p (d_i).
- For each word in the document, a topic z_k is selected with $P(z_k|d_i)$.
- A word x_j is generated with probability $P(x_j|z_k)$.

Each occurrence x_j is independent from the document it belongs to, given the latent variable z_k, which corresponds to the joint probability expressed by

$$P\left(x_j, z_k, d_i\right) = P(d_i)P(z_k|d_i)P(x_j|z_k) \tag{2.9}$$

The graphical representation of PLSA model is shown in Figure 2.8.

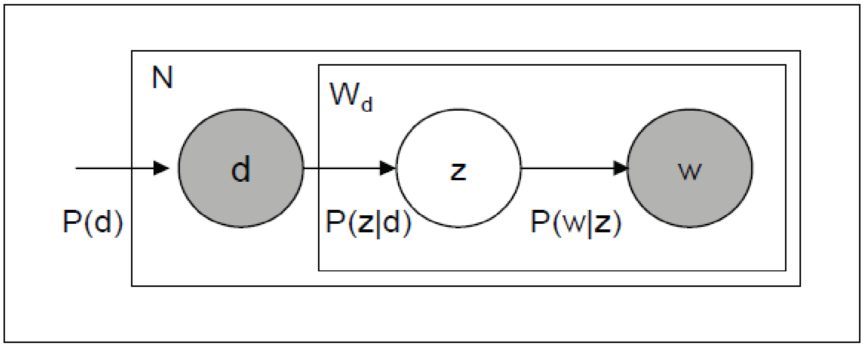


Figure 2.8: Graphical Representation of PLSA Model

The joint probability of the observed variables is the marginalization over the N_z latent aspects z_k as expressed by

$$P(x_j, d_i) = P(d_i) \sum_{k=1}^{N_z} P(z_k|d_i) P(x_j|z_k) \tag{2.10}$$

The unobservable probability distribution $P(z_k|d_i)$ and $P(x_j|z_k)$ are learned from the data using the Expectation –Maximization (EM) algorithm. $P(z_k|d_i)$ denotes the probability of topic z_k given in document d_i. $P(x_j|z_k)$ denotes the probability of visual word x_j in topic z_k. EM algorithm is a standard iterative technique for maximum likelihood estimation, in latent variable models such as Log likelihood. Normally, 100 -150 iterations are needed before converging.

Each iteration is composed of two steps:

1) An Expectation (E) step where, based on the current estimates of the parameters, posterior probabilities are computed for the latent variable z_k.

2) A Maximization (M) step, where parameters are updated for given posterior probabilities computed in the previous E step. It increases the likelihood in every step and converges to a maximum of the likelihood.

With the introduction of Bag of Words methods in computer vision, semantic analysis schemes became popular for tasks like scene classification and segmentation. PLSA works over BoW and has been successfully applied to annotate and retrieve images (Bosch et al., 2006b; Li et al., 2012; Romberg et al., 2012). PLSAWORDS (Monay and Gatica-Perez, 2007) is a representative approach, which achieves the annotation task by constraining the latent space to ensure its consistency in words. (Li *et al.*, 2011) presented a semantic annotation model which employs continuous PLSA and standard PLSA to model visual features and textual words respectively. The model learns the correlation between these two modalities by an asymmetric learning approach and then it can predict semantic annotation for unseen images. Multi-Modal probabilistic Latent Semantic Analysis (MMpLSA) which incorporate visual features and tags by generating simultaneous semantic contexts was proposed in (Pulla & Jawahar, 2010). Lienhart et al (2009) proposed a multi-layer probability Latent Semantic Analysis (PLSA) to solve the multi-modal image retrieval problem.

Apart from above mentioned usability of PLSA, it has dual ability to generate a robust and low-dimensional image representation as well as capturing meaningful image aspects (Quelhas *et al.*, 2005) . This ability will be further explained in chapter five.

2.8.2.2. Bayesian Classifier

The role of a Bayesian classifier is to classify patterns to the classes which are most likely to belong based on prior knowledge. In a Bayesian classifier, the learning agent builds a probabilistic model of the features and uses that model to predict the classification of a new example.

The idea is if an agent knows the class, then it capable of predicting the values of the other features. If it does not know the class, then bayes' rule can be used to predict the class given (some of) the feature values:

$$P(C_j|d) = \frac{P(d|C_j)P(C_j)}{\sum_{i=1}^{N}(P(d|C_i)P(C_i))} \tag{2.11}$$

where $P(C_j|d)$ is a probability of instance d being in class C_j, $P(d|C_j)$ is a probability of generating instance d given class C_j and $P(C_i)$ is a probability of occurrence of class C_j .

The simplest case of Bayesian classifier is naïve bayes classifier which makes the independence assumption that the input features are conditionally independent of each other given the classification. Bayesian classifier was used in Alzheimer disease diagnosis in (Lopez *et al.,* 2009). In another work, Bayesian Classifier has been used to classify mammogram images into benign and malignant (Talha *et al.*, 2012). Bayesian classifier has been used as classification techniques in some other various domains (Huang-Chia *et al.* 2012; Smith and Mobasseri, 2012; Smith *et al.*, 2011).

2.9. Evaluation Measurement

To evaluate the performance of our classification algorithms, average accuracy rate has been used. This measurement technique derived from confusion matrix. Confusion matrix is also used to explain the classification result in detail. This matrix on each column represents the predicted class while on each row represents the actual class as illustrated in Table 2.1.

Table 2.1: Example of Confusion Matrix

		Positive (Predicted)	**Negative (Predicted)**
Actual Value	**Positive (Actual)**	**A1**	**A2**
	Negative (Actual)	**A3**	**A4**
		Classifier Prediction	

where A1 represents the number of exact predictions that an instance is positive; A2 is the number of wrong predictions that an instance is negative; A3 shows the number of wrong predictions that an instance is positive; and A4 are the exact predictions that an instance is negative.

The classification accuracy rate is calculated as follow:

$$\text{Accuracy Rate} = \frac{A1+A4}{A1+A2+A3+A4} \tag{2.11}$$

2.10. Review of Selected Image Representation and Classification Techniques

To date, there have been many studies on ImageCLEF databases by different groups. This studies mainly in image representation space- using local versus global approach, how to define patches and extract features from each patch. In Table 2.2, we summarize selected studies on ImageCLEF medical dataset as regards to the image representation techniques.

In this section, we analyze and discuss the classification performance obtained from the presented studies with respect to their impact on the challenge of medical image classification. As mentioned earlier, an open challenge in automatic medical image classification is inter-class similarities and intra-class variability among images. This problem can be solved with classification algorithms that use the most discriminative information from the available data. Image representation is one of the major aspects of automatic classification algorithms, thus several authors tried to address this challenge using different types of descriptors including global and local features, separately or combination of them in a multi-visual approach.

Avni *et al.* (2011) proposed an X-ray image categorization and retrieval based on local patch representations using a "bag of visual words" approach with a kernel based SVM classifier. Three different feature extraction strategies were used and examined in their work; raw patches, raw patch with normalized variance and SIFT descriptors. The best classification result obtained from normalized patches on entire ImageCLEF 2007 database is 91.29 %.

Hierarchical classification schemes based on individual SVMs trained on IRMA sub-codes (Lehmann *et al.*, 2003) for the task of automatic annotation of ImageCLEF 2007 medical database was proposed by Unay *et al.* (2009). They simplified the classification task by training a separate SVM over each sub-code. The accuracy rates obtained over every sub-code are 96.7 %, 85.6 %, 88.0 % and 96.4 %. The final accuracy rate obtained is 91.7 %. A combination of multi-visual features such as GLCM, Pixel value and Canney edge detector as shape feature was presented in (Mueen, *et al.*, 2008). The accuracy rate obtained by their algorithm on ImageCLEF 2005 with 57 classes was 89 %.

Table 2.2: Image Representation techniques used in various Studies

	Author	**Image Representation Techniques**					**Accuracy Rate (%)**
		Pixel Value	**GLCM**	**LBP**	**Shape**	**BoW**	
1	Tommasi *et al.* 2008	√	×	×	×	√	89.7 %
2	Mueen et al. 2008	√	√	×	√	×	89.0 %
3	Unay *et al.* 2009	×	×	√	×	×	91.7 %
4	Avni *et al.* 2011	×	×	×	×	√	91.29 %

Tommasi *et al.* (2008) represented images using local and global features such as BoW and pixel values. They have integrated these features using multi-cue approaches named as high level, mid level and low level cue integration. They reported an accuracy rate of 89.7 % on ImageCLEF 2007 database.

However, by analyzing the results obtained from the methodologies used in all the above works, it is proven that the results attained are at global level; meaning the performance is obtained on the entire database. This result may not be achieved in every individual class due to the very unbalanced number of images in ImageCLEF database as well as some other complexities such as intra class variability and inter-class similarity existed in certain classes. To prove this, we conducted experiments with various image representation and classifier techniques as presented in Table 2.3. The results of this experiment have been published in (Zare *et al.*, 2013a).

As shown in Table 2.3, there is an average of 30 classes with accuracy below 60 %. By analyzing the results achieved from these 30 classes, it can be seen that most of these classes are suffering from the above mentioned complexity.

Table 2.3: Classification results on 116 classes with various feature extraction and classification techniques

Feature Extraction	**Classification Techniques**	**Accuracy Obtained**	**No. of Classes with Accuracy < 60 %**
GLCM, Canny edge detector, Pixel Value	SVM with RBF kernel	70.45 %	64
GLCM, Canny edge detector, Pixel Value	SVM with Polynominal kernel	66.25 %	77
GLCM, Canny edge detector, Pixel Value	SVM with Linear kernel	68.15 %	70
GLCM, Canny edge detector, Pixel Value	KNN, k=9	65.95 %	85
Local Binary Pattern	SVM with RBF	89.95 %	29
Local Binary Pattern	SVM with Polynominal kernel	85.15 %	40
Local Binary Pattern	SVM with Linear kernel	86.55 %	35
Local Binary Pattern	KNN, k=9	86.0 %	36
Bag of Visual Words	SVM with RBF	90.0 %	32
Bag of Visual Words	SVM with Polynominal kernel	86.10 %	35
Bag of Visual Words	SVM with Linear kernel	87.50 %	33
Bag of Visual Words	KNN, k=9	87.15 %	33

Results from the above table shows that BoW as image representation technique with non-linear multi class SVM with RBF kernel outperformed the other presented approaches.

As explained in section 2.7.1, the main parameter in construction of BoW is *k* which denotes the size of vocabulary. Different value for *k* has been considered in our experiments starting from 100 followed by 200, 300, 400, 500, 600 and 700 to investigate how the classification performance is affected as illustrated in Table 2.4. As presented, the best accuracy rate obtained at *k*=500.

Table 2.4: Classification Rate obtained from Different Vocabulary Size

Vocabulary size	100	200	300	400	500	600	700
Accuracy Rate (SVM with RBF)	82 %	83.9 %	84.5 %	88.0 %	90.0 %	88.5 %	87.9 %

In the following three chapters, different classification frameworks are proposed in order to increase the number of classes with high accuracy rate.

- In chapter three, we applied continuous filtering on the training dataset. Filtering is done according to their classification performance. Therefore every generated model is evaluated and passed the threshold accuracy rate in training phase.
- In chapter four, we continued the work in chapter three by treating the classes with low accuracy rate, separately. They are classified via annotation; as such three different approaches are employed to annotate them.

However, apart from several advantages of bag of words such as its simplicity, discrete representations and simple matching measures to preserve computational efficiency, it still has several drawbacks such as ignorance of spatial information (Lazebnik, *et al.*, 2006).

Another disadvantage of BoW model is ambiguous data representation which was discussed in (Kesorn & Poslad, 2012; Lei *et al.*, 2010) . The ambiguity lies into two areas: *Visual Polysemy* where single visual word occurring on different parts on different object categories and *Visual Synonyms* where two different visual words representing a similar part of an object. To overcome this shortcoming, a generative model such as PLSA has been proposed to learn the co-occurrence between elements in the vector space in an unsupervised manner to disambiguate the BoW representation. Thus, a classification algorithm based on integration of PLSA and discriminative SVM classifier is proposed in chapter five.

Chapter 3 Automatic Medical Image Classification using Bag of Visual Words

3.1. Introduction

We have discussed the challenges in classification of large medical dataset in Section 2.10, i.e. unbalanced number of training data, intra-class variability and inter class similarities among classes. In this chapter and the next two chapters, novel approaches are presented to increase the number of classes with high accuracy rate.

In this chapter, we developed a classification framework to increase the number of classes with higher accuracy rate in large archive medical database. The learning phase consists of four iterations where different classification models were generated as per iteration. For the iterations, model generation process was performed in two steps. The first step starts with construction of model from the entire dataset. This model was then assessed to filter high accuracy classes (HAC). These classes were those predicted with accuracy rate above 80%. This evaluation performed on 20% of the training dataset which was taken as test data. In the second step, classes under HAC were only used to construct the classification model. The same processes will be performed in the next iteration on the classes which were left with the accuracy below 80% from the previous iteration.

The methodology presented is based on Bag of visual Words for feature extraction and the RBF based SVM classifier. The proposed methodology was evaluated on ImageCLEF 2007 medical database (See section 3.2 for more details). As a result, four classification models were generated from 77, 17, 12 and 10 classes, respectively. The accuracy rate obtained by each generated models outperformed the results obtained by only one model on the entire dataset. The results of this experiment have been published in (Zare et al., 2013b). In this chapter, the proposed classification framework is explained and evaluated in detail.

3.2. ImageCLEF 2007 Medical Database

The database used in this study is ImageCLEF 2007 (Muller *et al.,* 2007) which was provided by the IRMA group from RWTH University Hospital of Aachen, Germany. It consists of medical radiographs collected randomly from daily routine work at the Department of Diagnostic Radiology. The quality of radiographs varies considerably and there is a high intra class variability and inter-class similarity among classes. In order to establish a ground truth, the images were manually classified by expert physicians using the IRMA code (Lehmann *et al.,* 2003). The four main facets of IRMA code are image modality also known as Technical (T), body orientation known as Directional (D), body region examined also called Anatomical (A), and biological system called Biological (B). This classification is also now known as ABCD (A=Anatomy, B=Biological, C=Creation of image, and D for Direction).

The ImageCLEF 2007 medical database consisting of 11000 medical x-ray images from 116 categories which differ from each other either on account of image modality, examined region, body orientation and biological system examined. In Table 3.1, the detail of each body regions in this database as well as the number of categories under those body regions are presented.

Table 3.1: Main body region & number of classes per body region

Body Regions	**No. of Classes**	**Body Regions**	**No. of Classes**
Abdomen, unspecified	4	Leg, lower leg	3
Abdomen, upper abdomen,unspecified	1	Leg, foot, toe	2
Chest, unspecified	5	Leg, ankle joint	4
Chest, bones, upper ribs	1	Leg, hip	4
Chest, bones	2	Leg, knee, patella	1
Chest, bones, lower ribs	1	Leg, upper leg, unspecified	2
Arm, Forearm	8	Leg, upper leg	2
Arm, Shoulder	2	Leg, upper leg, distal upper leg	1
Arm, Shoulder , humero-scapular joint	2	Leg, knee	4
Arm, Shoulder , acromio-scapular joint	2	Leg, foot	11
Arm, Upper Arm	1	Leg, lower leg, unspecified	2
Arm, upper arm, proximal upper arm	1	Cranium, facial cranium, temporo mandibular area	1
Arm, upper arm, distal upper arm	1	Cranium, facial cranium, eye area	1
Arm, Elbow	4	Cranium, unspecified	1
Arm, Hand	4	Cranium, facial cranium, mandible	1
Arm, Hand, Finger	2	Cranium, facial cranium, nose area	5
Arm, Hand , Carpal Bones	2	Cranium, neuro cranium, unspecified	2
Arm, Radio Carpal Joint	4	Cranium, neuro cranium, occipital area	1
Spine, lumbar spine, thoraco-lumbar conjuction	2	Breast	4
Spine, cervical spine, unspecified	7	Pelvis	1
Spine, lumbar spine, unspecified	3		
Spine, cervical spine, dens	1		
Spine, lumbar spine, lower lumbar spine	1		
Spine, thoracic spine, unspecified	2		

3.3. Methodology

Classification process is in two steps of training and testing phases. In training phase, the selected features are extracted from all the training images, and classifier is trained on the extracted features to create a model. This model is then used in testing phase to classify the unseen test image into one of the pre-defined categories.

3.3.1. Training Phase

Figure 3.1 illustrates the training phase of the proposed classification framework. Training phase consists of two modules; feature extraction and model generation. In the feature extraction module, Bag of visual Words (BoW) is extracted from every training image.

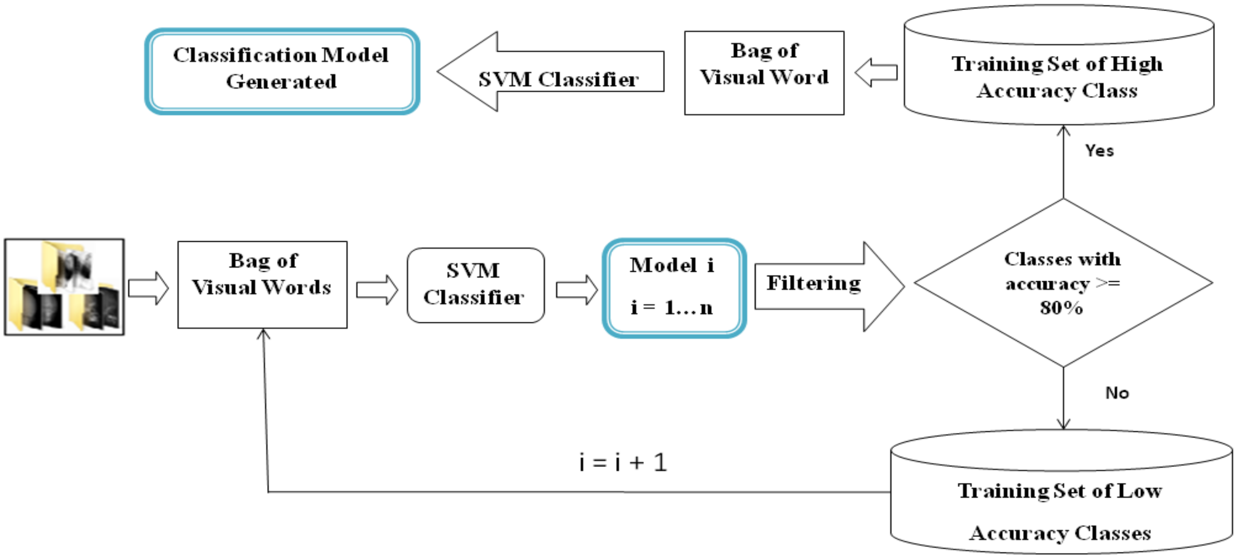

Classification Model Generated
SVM Classifier
Bag of Visual Word
Training Set of High Accuracy Class
Yes
Bag of Visual Words
SVM Classifier
Model i
i = 1…n
Filtering
Classes with accuracy >= 80%
No
i = i + 1
Training Set of Low Accuracy Classes

Figure 3.1: Model Generation using Continues Filtering

3.3.1.1. Construction of Bag of Words (BoW)

The process of BoW started with detecting local interest point. Local interest point detectors have the task of extracting specific points and areas from images which are invariant to some geometric and photometric transformations. One of the popular approaches for the detection of local interest point is Difference of Gaussians (DoG) which is used in this experiment. This detector has been chosen since it was shown to perform well for the task of wide-baseline matching when compared to other detectors. We can observe that the DoG detector is considerably faster since it is based on the subtraction of images. DoG has been built to be invariant to translation, scale, rotation, and illumination changes and samples images at different locations and scales. This technique uses scale-space peaks in the difference of Gaussian operator convolved with the image. The process of Difference of Gaussian detector operator is presented in following:

Local interest point extraction process using DoG Detector

1. Smooth original image $I_{initial}$ with Gaussian G(X, $\sqrt{2}$), to create the first image of the scale space representation.
2. Apply smoothing equal to image $I_{initial}$.
3. Subtract each Gaussian scale-space smoothed image with the image immediately lower in the scale space representation.
4. Perform DoG local interest point detection using a maxima detection procedure on the current Gaussian scale-space representation.
5. Set the initial image $I_{initial}$ to be the last Gaussian smooth image of the current octave.
6. Re-sample the initial image $I_{initial}$ by taking each other pixel, creating an image of half the size of the original.
7. If image size larger than two times the size of the Gaussian kernel used to create the scale space representation, return to step 2.

Next, the detected keypoints are then represented using Scale Invariant Feature Transform (SIFT) as described in section 2.4.2. In short, the image gradient is sampled and its orientation is quantized. Using a grid division of the local interest area, local gradient orientation histograms are created where the gradient magnitude is accumulated. The final feature is the concatenation of all the local gradient orientation histograms. A Gaussian weighting is introduce in the SIFT feature extraction process to give more importance to samples closer to the center of the local interest area. This contributes to a greater invariance of the SIFT descriptor, since samples closer to the center of the local interest areas are more robust to errors in the local interest area estimation.

In Lowe (2004) it was found that the best compromise between performance an speed was obtained by using a 16 × 16 gradient sampling grid and a 4 × 4 sub histogram grouping. The final descriptor proposed in this formulations is 8 orientations and 4 × 4 blocks, resulting in a descriptor of 128 dimensions.

Next step in implementation of bag of visual words is the codebook construction where the 128-dimensional local image features have to be quantized into discrete visual words. This task is performed using clustering or vector quantization algorithm. This step usually uses k-means clustering method, which clusters the keypoint descriptors in their feature space into a large number of clusters and encodes each keypoint by the index of the cluster to which it belongs. We conceive each cluster as a visual word that represents a specific local pattern shared by the keypoints in that cluster. Thus, the clustering process generates a visual-word vocabulary describing different local patterns in images. The number of clusters determines the size of the vocabulary, which can vary from hundreds to over tens of thousands. Mapping the keypoints to visual words, we can represent each image as a "bag of visual words".

3.3.1.2. Model Generation

Upon extraction of BoW representation from the training dataset, it was then used as inputs to SVM classifier to construct the model. This constructed model is then evaluated in training phase itself to assure the best possible classification rate is attained for every individual class in the database. As such, the training dataset is divided into two parts. Eighty percent of it was used to construct the classification models, and the remaining 20% of the training data were taken as for test images for evaluation purpose of the generated model.

As shown in Figure 3.1, the classification process may carry out in several iterations depends on the number of categories. In the first iteration, the extracted BoW from the entire training dataset is fed into SVM classifier to create a model. This model is named as Model *i* where *i=1* since it's a first iteration.

Next, model *i=1* is applied on the BoW representation extracted from the specified test data in the training phase to filter High Accuracy Classes (HAC) from Low Accuracy Classes (LAC). The threshold of 80% has been set for the optimum classification rate. This threshold is chosen because it is very rare to have high percentage of accuracy in large medical database. We had chosen a balanced value here to trade off accuracy with practicality. All the classes with the accuracy rate of 80% and above are labeled as HAC while those below 80% are labeled as LAC.

Figure 3.2 shows the classification accuracy rate obtained by applying Model *i=1* on the test dataset specified in the training phase. The overall classification accuracy rate obtained is 90 % in this stage.

As shown in Figure 3.2, there are 77 classes with the optimum accuracy rate which fall into HAC. These 77 classes are merged to form a new training dataset called *Training Set of High Accuracy Classes*. This new training dataset then go under feature extraction and model generation to construct model one (Based on 77 classes). Similarly, the corresponding classes from test dataset are also separated from the testing dataset. This is where the first iteration of the training phase would end.

Next, if there are any classes which fall into LAC from the first iteration are merged to form Training set of *Low Accuracy Classes*. Thus, the value of '*i*' is increased by 1 which it represents the next iteration. As for the second iteration, BoW is extracted from training set of LAC. Then model *i=2* is generated and evaluated on the remaining set of the test data.

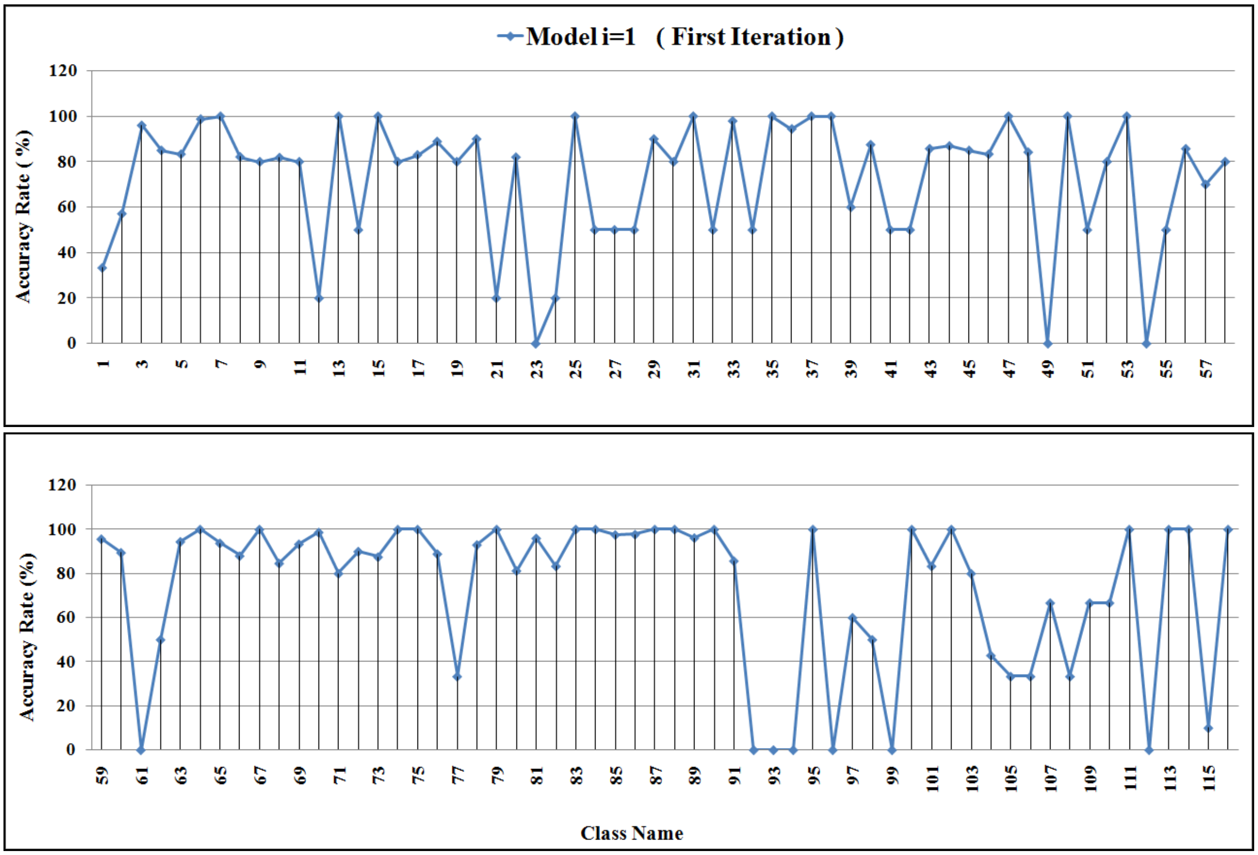


Figure 3.2: Classification Result on 116 Classes (First iteration)

Figure 3.3 illustrates the classification results attained from model *i=2* of training phase. This model was constructed from the remaining 39 classes left in LAC from the first iteration. As shown, there are 17 classes which achieve the optimum accuracy rate.

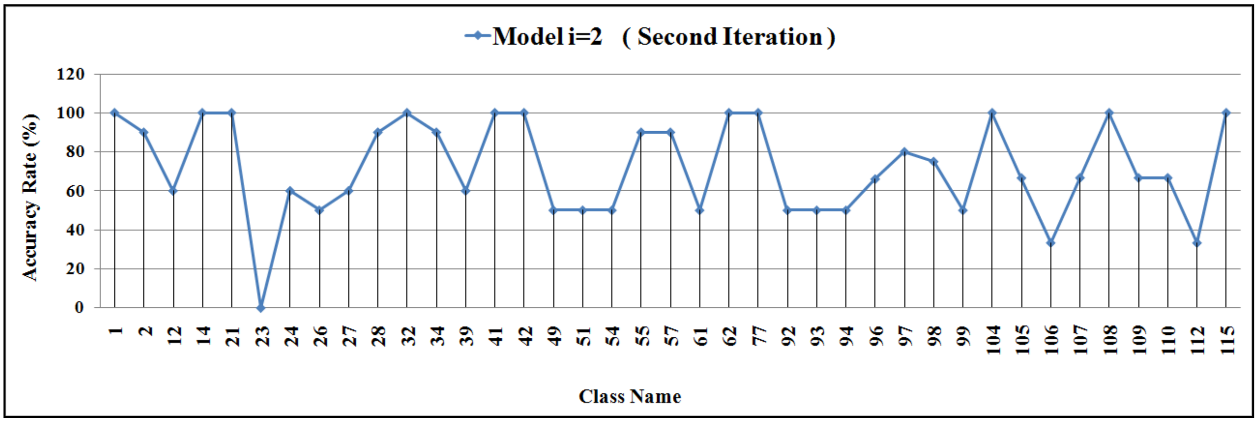


Figure 3.3: Classification Result on 39 Classes (Second Iteration)

Accordingly, these 17 classes also combined to form a new training set. BoW is extracted from the newly training set followed by SVM classifier to create the second classification model (Based on 17 Classes).

Similar to the first two iterations, the remaining 22 classes were used to generate model *i=3*. This model was then evaluated on the corresponding 22 classes of the test dataset. The result obtained for every individual class is shown in Figure 3.4.

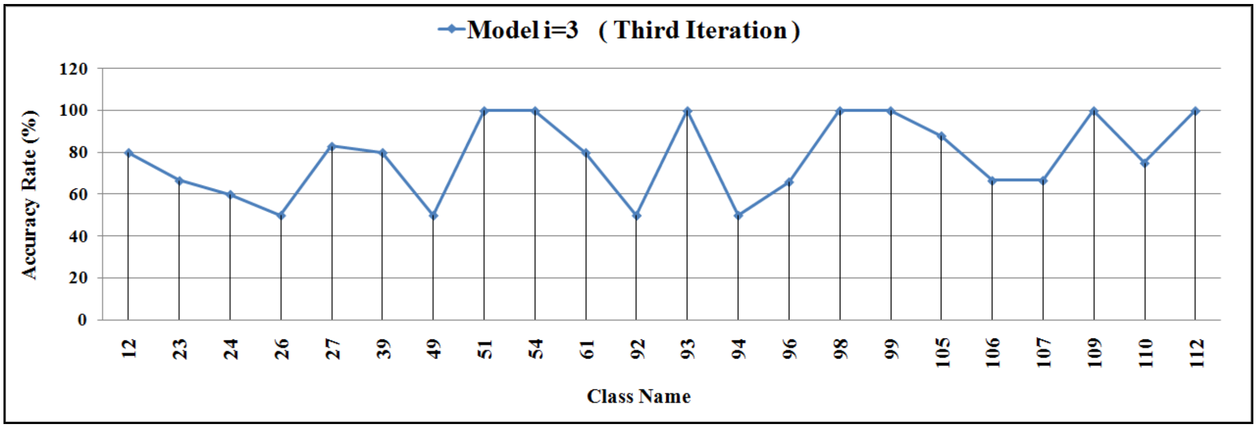


Figure 3.4: Classification Result on 22 Classes (Third Iteration)

As it can be seen in Figure 3.4, twelve classes with the optimum accuracy rate from the third iteration were used to construct the third classification model. The model generation process carried out once again in order to get the optimum classification rate for the remaining 10 classes. Consequently, the fourth classification model is generated based on 10 classes as illustrated in Figure 3.5.

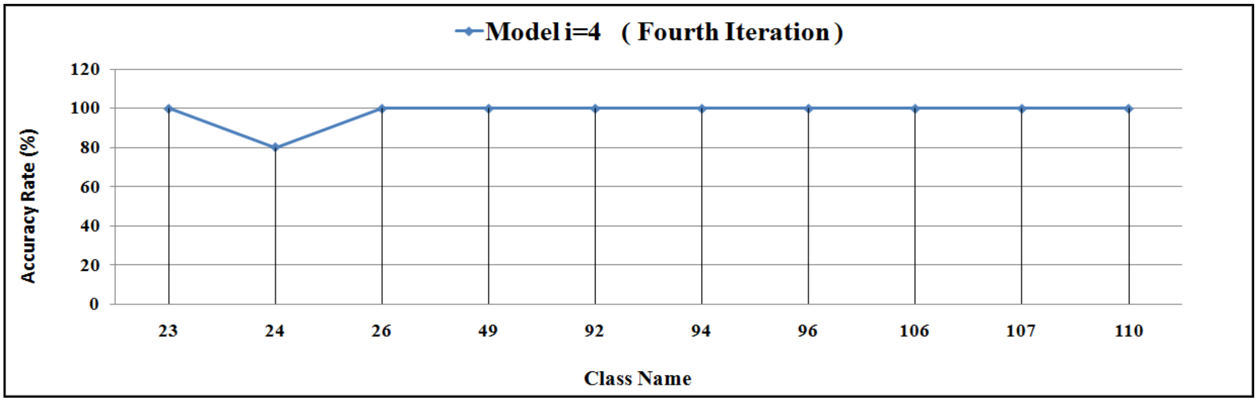


Figure 3.5: Classification Result on 10 Classes (Fourth Iteration)

As a result, four classification models were generated from different number of classes in the training phase. Figure 3.6 depicts the data segments produced by the four iterations described above. As shown, the four generated models are constructed from 77, 22, 12, and 10 classes respectively. These models will be used to classify the unseen test image as explained in the following section.

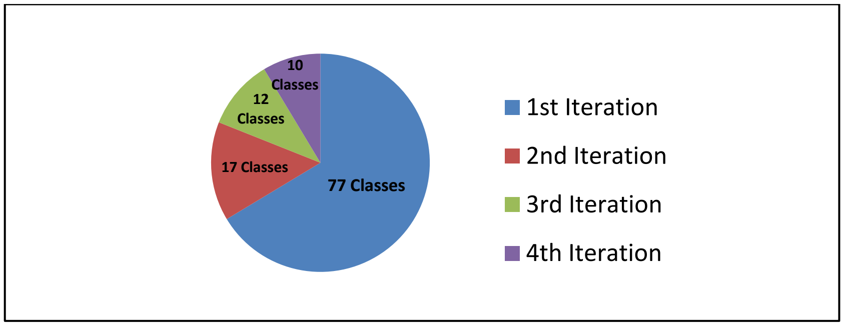


Figure 3.6: No. of Classes with Optimum Accuracy as Per Iteration

3.4. Testing Phase

In testing phase, only one of the classification models constructed in training phase will be applied on the test image in order to classify it into predefined category. Since there are four classification models exist, a question may arise that which model must be applied on the unseen test data? In order to make a decision, we first apply model *i=1* on the extracted BoW from the test image. The model *i=1* was generated from the entire dataset and used to filter out classes with optimum accuracy rate from the first iteration in the training phase. The output of this classification is any number from 1 to 116 because there are 116 possible classes. Based on this output and knowing what classes were used to construct each one of the four classification models, the qualified classification model can be easily determined. Figure 3.7 illustrates the classification process on the test data.

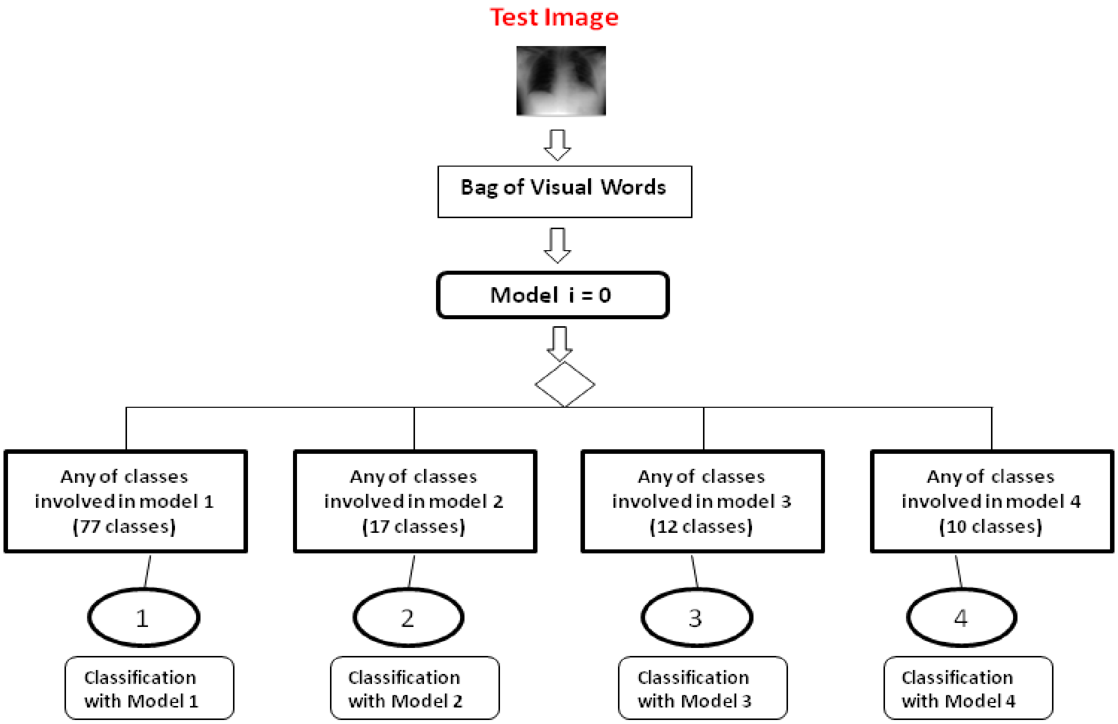

Test Image
Bag of Visual Words
Model i = 0
Any of classes involved in model 1 (77 classes)
Any of classes involved in model 2 (17 classes)
Any of classes involved in model 3 (12 classes)
Any of classes involved in model 4 (10 classes)
1
2
3
4
Classification with Model 1
Classification with Model 2
Classification with Model 3
Classification with Model 4

Figure 3.7: The process of selecting classification model on test image

3.5. Experimental Results

In this section, the performance of the proposed algorithm will be evaluated. As mentioned, this experiment was conducted on ImageCLEF 2007 database; this database consists of 11,000 training images and 1000 test images. As stated in previous section, there was a need for test data in the training phase. Thus, 20% of the training data were taken as for testing images in training phase.

In this stage, the four classification models will be evaluated based on 1000 unseen test images. As shown in Figure 3.8, similar to the training dataset, the numbers of images in every class are not distributed uniformly in the test dataset.

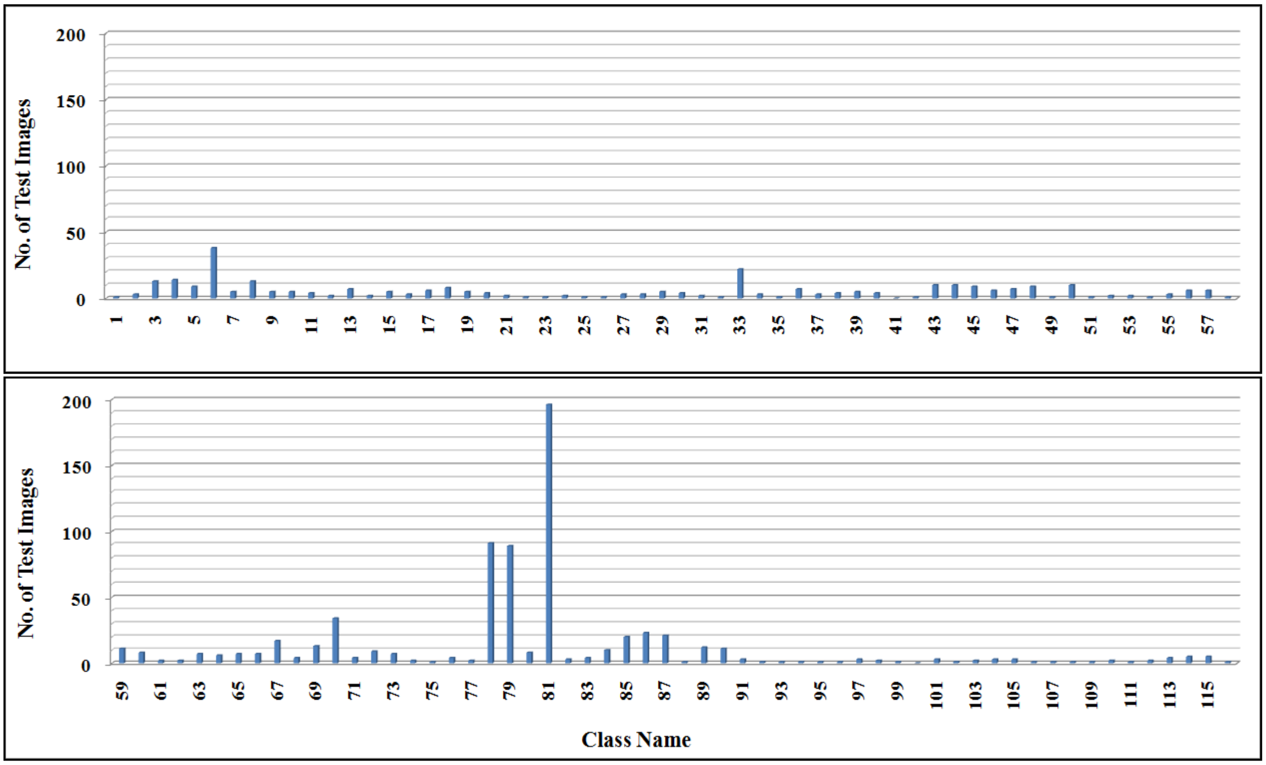


Figure 3.8: Distribution of Test Images in ImageCLEF 2007

3.5.1. Parameter Optimization

In this experiment, images were represented as BoW. The main parameter in construction of BoW is the number of k which represents the vocabulary size. As demonstrated in Table 2.4, various experiments were conducted with different vocabulary size starting from 100 followed by 200, 300, 400, 500, 600 and 700. Based on the experimental results, the best performance is achieved at k=500. LIBSVM software package has been utilized to perform SVM-based classification with RBF kernel functions. We use one-vs-one multi-class extension for SVM. The optimum kernel parameters $\gamma= 0.0001$ and cost $C=250$ have identified empirically with 5-fold cross validation separately for every classification model. The measurement used to compare classification results is the average accuracy. It is also referred to as accuracy rate. It is a ratio of number of images classified correctly over the size of test dataset.

3.5.2. Results Obtained From Each Classification Models

Classification process on the unseen test image starts with identifying the respective classification model for that particular image. As such, if the right model is not assigned to the test image, then the classification rate would drop. As stated earlier, in order to choose the right classification model, we first apply model i=1 on the test image. The total classification rate obtained by applying model i=1 on 1000 test images is 89 %; Meaning that 890 out of 1000 test images were given the right classification model. Based on the analysis that has been done on the results, most of the remaining 110 test images were misclassified within their sub-body region and they are under the same batch of classes which were used to construct each one of the four classification model. Thus, regardless of the output obtained from classification model *i*=1, the right classification model was assigned to the test images. This process is carried out for every unseen test image in order to choose the right classification model. In the following discussions, the results are presented in a class specific level which was obtained from each one of the four classification models.

Figure 3.9 shows the accuracy rate obtained by the first classification model which was generated from 77 classes. The total accuracy rate obtained from the first classification model is 92 %. As we can see, most of the classes are obtaining high accuracy rate except the class 100. This is due to that there is no test image provided for class 100.

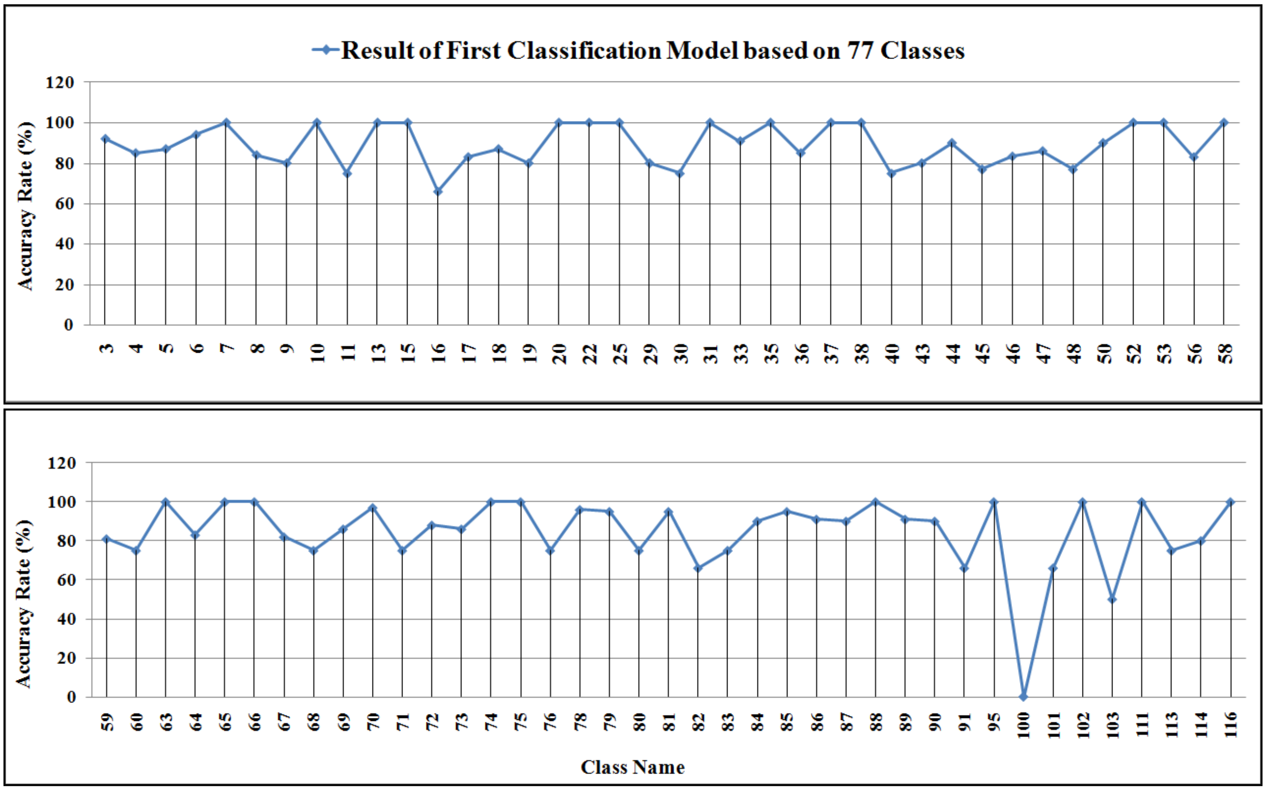


Figure 3.9: Classification Result obtained from the first Model on real dataset

The accuracy rate obtained from the second classification model on the real dataset is illustrated in Figure 3.10. The total accuracy rate is 88.5 %. Similar to the previous model, class 41 has no test image; as such it has zero accuracy rate.

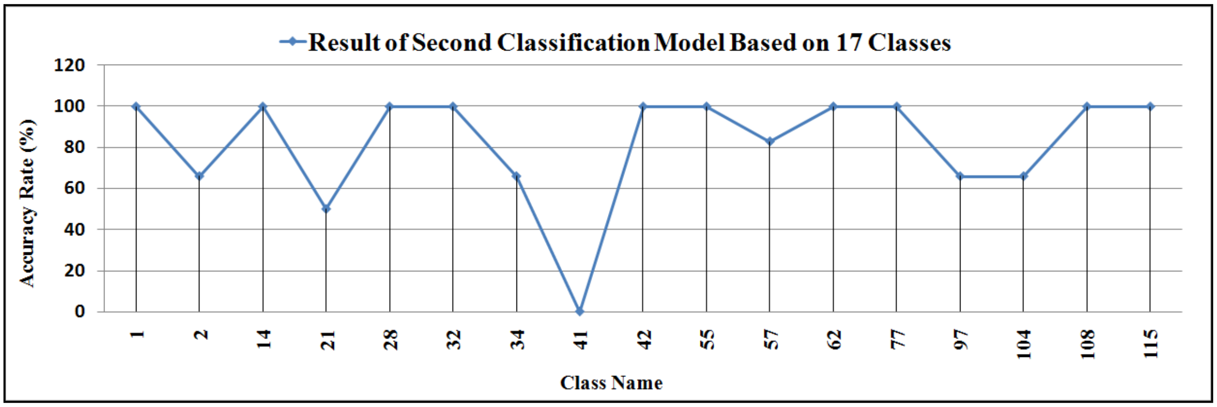


Figure 3.10: Classification Result obtained from the Second Model on real dataset

Similarly, the accuracy rate attained from the third and fourth models are 79 % and 77 %, respectively. The following figures illustrate the classification results obtained from the third and fourth models for every individual class.

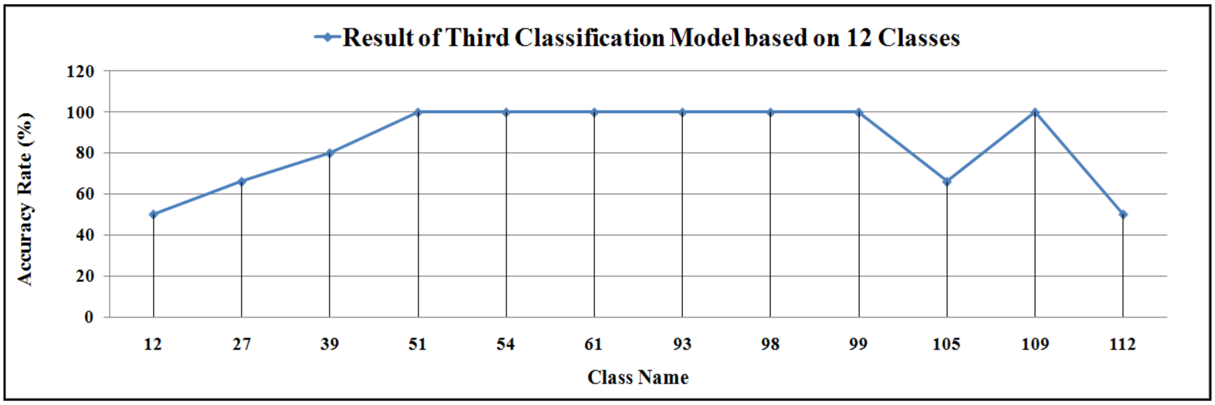


Figure 3.11: Classification Result obtained from the Third Model on real dataset

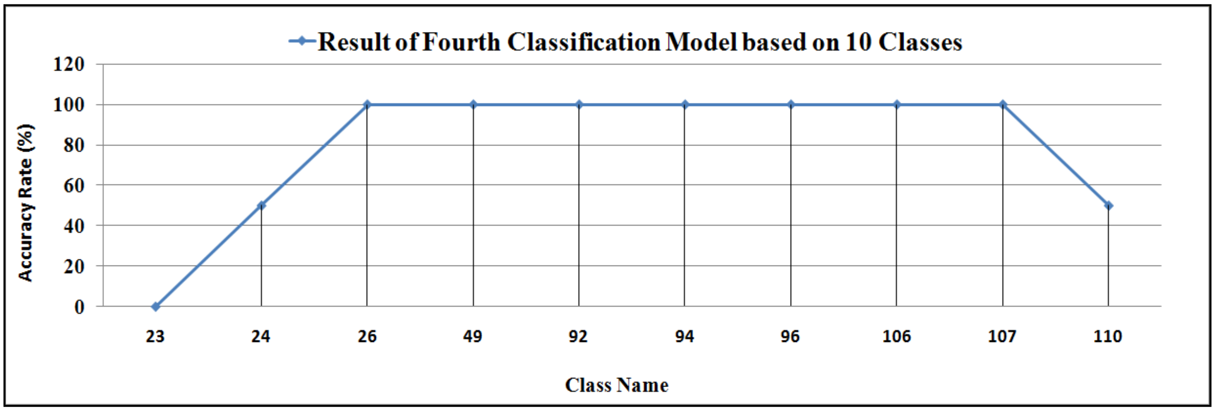


Figure 3.12: Classification Result obtained from the Fourth Model on real dataset

As it can be seen, there are still a number of classes with low accuracy rate obtained from these models. Class 23 has also zero accuracy rate, this is because there is only one test image available for this class, and it was misclassified to class 24. This misclassification is due to the inter-class similarity between these two classes as both of them are referring to the same sub-body region. Drill down analysis has been done on these selected classes and will be explained in discussion section.

3.5.3. Discussion

The aim of this study was to increase the classification performance of large medical database such as ImageCLEF 2007. The experimental result shows that the average accuracy rate obtained on the entire database is 90% which is a good result compared with similar works and smaller number of classes; it is yet to be a satisfactory result to meet the objective. There are still 41 classes with the accuracy below 80% at this stage. By analyzing the classification result of every individual class, it can be seen that almost all large categories have accuracy rates of above 85% whereas images from classes with smaller number of training images are frequently misclassified. Further investigation on misclassified classes shows that they are visually similar to some of the classes with high accuracy rate as they refer to the same sub-body region. The reason of misclassification is that SVM or any other classification techniques would be biased to the category with a bigger number of training images. This examination shows that depending on only one technique to gain high accuracy for every individual class of such database with the said complexity is unreliable. This shortcoming motivated us to perform filtering on the database in several iterations, and consequently a separate model is constructed from each iteration. The concept is to filter out those classes with high accuracy rate from the rest of the database in every iteration. As a result, those classes with low accuracy rate were separated from those with high accuracy rate which are visually similar too. After separation, they will be treated independently by forming a new classification model. These processes were carried out until every class managed to have optimum accuracy rate. In the following session, the classification results obtained by this framework are analyzed.

As we have seen in previous section, four classification models were constructed from different number of image categories. Next, the respective classification model for the test image is chosen and applied on it in order to classify it into one of the predefined classes. A detailed analysis on the classification results of the 1000 unseen test images shows that all the four generated models managed to gain an optimum accuracy for every individual class except seven classes. A drill down analysis has been done to know the reason of low accuracy in these classes. In Table 3.2, the number of training and testing images both in the learning and testing phases as well as the generated model in charge for classification are shown for these seven classes.

Table 3.2: Categories with low Accuracy Rate

	Class	100	41	12	112	23	24	110
Learning Phase	No. of Training Images (In Learning phase)	7	18	24	17	15	25	19
	No. of Test images (In Learning phase)	2	5	5	4	3	5	4
	Accuracy Rate (%)	100	100	80	100	100	80	100
Testing Phase	No. of Test Images (In Testing Phase)	0	0	2	2	1	2	2
	Accuracy Rate (%)	0	0	50	50	0	50	50
	Model in Charge	Model 1	Model 2	Model 3		Model 4		

Except the two classes of 41 and 100 that obtained zero accuracy due to the non-existence of the test image, there are another 5 classes with low accuracy rate compared to the rest of the classes. Table 3.3 is the confusion matrix on these 5 classes.

Table 3.3: Confusion Matrix on Classes with Low Accuracy

Class	12	112	23	24	110	Other classes	Accuracy Rate (%)
12	1					1 (Class 27)	50.0 %
112		1				1 (Class 109)	50.0 %
23				1			0.0 %
24				1		1 (Class 26)	50.0 %
110				1	1		50.0 %

Based on the given IRMA code, class 12 is labeled as “arm, hand, carpal bone” with 2 test images. As illustrated in Table 3.2, based on the result of classification model, one of the test images is misclassified as class 27 which is referring to the same sub-body region as class 12.

Class 112 and class 109 belong to the category of lower leg, with an identical image anatomy code given by IRMA. One of the two test images of class 112 was misclassified as class 109. Similarly, classes 23, 24, 26 and 110 also belong to the sub-body region of “forearm”. As presented in Table 3.2, a few test images are available for these classes too and they were misclassified within the same sub-body region itself. Further investigation on the misclassified images and related classes showed that most of them are suffering from high inter-class similarities and intra-class variabilities.

However, we cannot rely on one or two test images to evaluate the performance of the generated classification models. This is because the probability of getting high accuracy is very low with very small number of test images as the accuracy rates obtained by these models are high in the learning phase with a bigger number of test images.

3.6. Chapter Summary

As we have seen from the results obtained by many works in the literature, it is difficult to obtain high accuracy for every individual class due to the problem of intra-class variability and inter-class similarity in large medical database. All the results obtained are average classification accuracy on the entire dataset. This is not an accuracy that has been achieved in every individual class. To address this issue, we proposed an iterative classification framework that produced four classification models which were constructed from different number of classes. We exploited Bag of visual Words (BoW) features, which represents an image by histogram of local patches on the basis of visual vocabulary. The model generation process was performed in two steps. In the first step, a classification model was generated from the entire dataset. It was then evaluated using the 20% of the training image which were allocated for this purpose. The classes with accuracy rate of 80% and above were filtered out from the rest. These classes were labeled as HAC. In the second step, the classification model is constructed from the classes under HAC. These two model generation steps will be performed on the classes which were left with the accuracy below 80% from the previous iteration. These processes carried out in training phase for four iterations and consequently four classification models were constructed accordingly. Next, all the four generated models are tested with the 1000 unseen test data. Experimental results show that out of 116 classes in ImageCLEF 2007 dataset, 109 classes managed to attain higher accuracy rate. This is a significant classification performance as compared with the results obtained from some relevant work presented in Table 2.2 (Refer to section 2.10 in chapter 2).

In the next chapter, we have developed another classification framework for those classes left with low accuracy rate after the first iteration.

Chapter 4 Automatic Classification of X-Ray Images using Annotation

4.1. Introduction

In this chapter, an approach is presented to gain high accuracy rate for those classes of large medical database with high ratio of intra class variability and inter-class similarities. The classification framework was constructed via annotation using the following three techniques: Annotation by binary classification, Annotation by Probabilistic Latent Semantic Analysis (PLSA) and Annotation using top similar images. Next, final annotation was constructed by applying ranking similarity on annotated keywords made by each technique. The final annotation keywords were then divided into three levels according to the body region, specific bone structure in body region as well as imaging direction. Different weights were given to each level of the keywords; they are then used to calculate the weightage for each category of medical images based on their ground truth annotation. The weightage computed from the generated annotation of test image was compared with the weightage of each category of medical images, and then the test image would be assigned to the category with closest weightage to the query image. The results of this experiment have been published in (Zare et al., 2013c). In the rest of this chapter, the proposed approach is presented in detail followed by experimental results and analysis.

4.2. Methodology

Classification process is in two steps of training and testing phases. In training phase, the selected features are extracted from all the training images, and classifier is trained on the extracted features to create a model. This model is then used in testing phase to classify the unseen test image into one of the pre-defined categories. As stated earlier, the purpose of this study is to improve the classification performance of the classes under Low Accuracy Class (LAC), those classes which are left with low accuracy rate after the first iteration in previous chapter. As such we proposed to classify the unseen test images via three techniques of annotation as described below.

4.2.1. Annotation by Supervised Classification

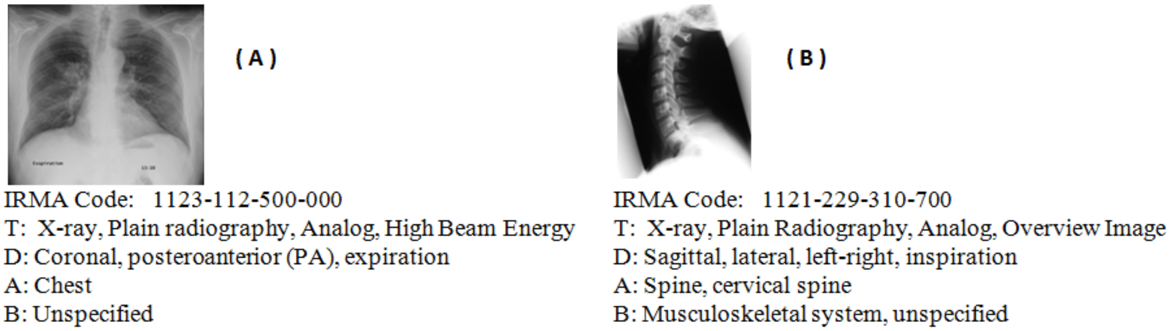
In this approach, supervised learning approach is used to classify images. Classification process consists of two steps of training and testing phases. In the training phase, the selected features are extracted from all the training images, and a classifier is trained on the extracted features to create a model. This model is used to classify the unseen test images into a pre-defined class and then corresponding keywords of that class will be assigned to the unseen test image as an annotation. For instance, if a test image classifies to class 23, it will be annotated by the following keywords: Arm, forearm, wrist joint, elbow joint, radius, ulna, left, AP view.

4.2.2. Annotation using PLSA

In formulation of annotation using PLSA, we incorporate both visual vocabulary and textual vocabulary in construction of multi-modal PLSA model. Each modality (visual vocabulary and textual vocabulary) are treated differently. Based on our empirical studies, we give more importance to textual vocabulary in order to capture meaningful aspects in the data and use them for annotation. This is to ensure the consistent set of textual words is predicted while retaining the ability to jointly model the visual features. To formulate visual vocabulary, we computed a co-occurrence table where an image is represented by BoW as explained in previous chapter. The BoW is represented as 2D matrix with 564 rows and 500 columns. 564 and 500 are the number of training images and visual vocabulary size, respectively. The process of constructing textual vocabulary is described below:

Textual Vocabulary

Based on given IRMA code and comments given by qualified physician, the corresponding annotated keywords for each class of the medical database are identified. After eliminating the duplicate keywords, the unique set of annotated keywords are generated. An average of 5 keywords is specified for every image in the class. For instance, the annotated keywords that were assigned to Figure 4.1(A) are: "*Chest, lung, rib cage, coronal, PA view*". "*Chest, Coronal and PA view"* were taken from the textual labels come with IRMA code; "*lung and rib cage"* were given by physician.

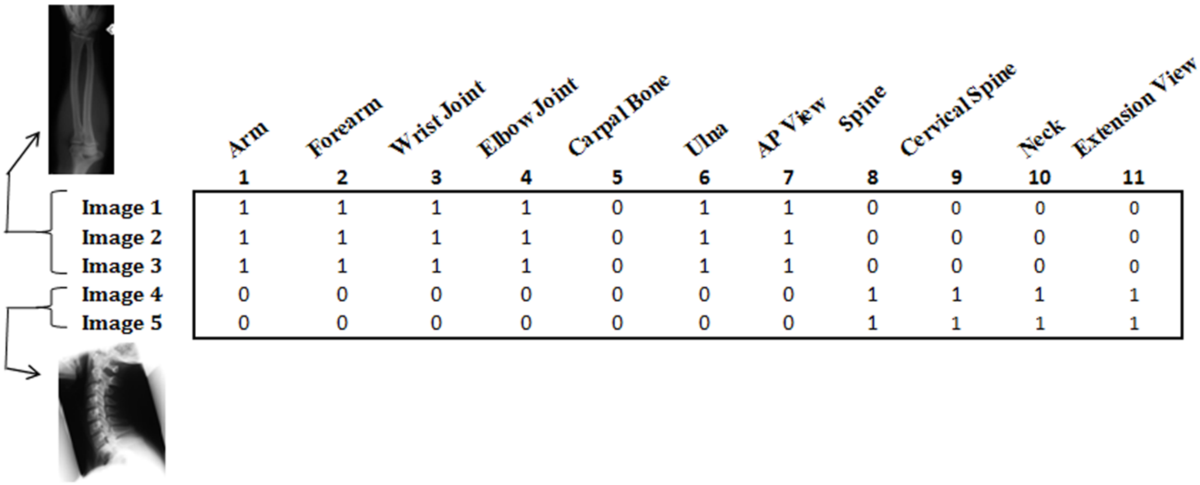


IRMA Code: 1123-112-500-000
T: X-ray, Plain radiography, Analog, High Beam Energy
D: Coronal, posteroanterior (PA), expiration
A: Chest
B: Unspecified

IRMA Code: 1121-229-310-700
T: X-ray, Plain Radiography, Analog, Overview Image
D: Sagittal, lateral, left-right, inspiration
A: Spine, cervical spine
B: Musculoskeletal system, unspecified

Figure 4.1: Sample Images and the corresponding textual labels from IRMA code

To formulate the textual vocabulary, the dataset is then represented as Term-Document Matrix as shown in Figure 4.2 by placing the image names and generated keywords in rows captions and columns captions of the matrix, respectively. Each cell of the matrix is then filled by 1 or 0 where 1 represents the occurrence of the particular keyword and 0 indicates the non-occurrence of that keyword for a specific image. The final Term-Document Matrix is represented as 2 dimensional matrixes with 564 rows and 68 columns. 564 and 68 are the number of training images and number of textual words, respectively.



	Arm	Forearm	Wrist Joint	Elbow Joint	Carpal Bone	Ulna	AP View	Spine	Cervical Spine	Neck	Extension View
	1	2	3	4	5	6	7	8	9	10	11
Image 1	1	1	1	1	0	1	1	0	0	0	0
Image 2	1	1	1	1	0	1	1	0	0	0	0
Image 3	1	1	1	1	0	1	1	0	0	0	0
Image 4	0	0	0	0	0	0	0	1	1	1	1
Image 5	0	0	0	0	0	0	0	1	1	1	1

Figure 4.2: Sample Term-Document Matrix of Textual Vocabulary

Automatic Image Annotation with PLSA

Upon construction of BoW and textual vocabulary, linked pair of PLSA models is trained for the task of automatic image annotation as described below. The flow of learning and annotation on the unseen test image is illustrated in Figure 4.3.

Learning Phase:

1. First PLSA model is completely trained on the set of image captions (textual vocabulary) to learn both $P(x_j|z_k)$ and $P(z_k|d_i)$ parameters. As a result, set of aspects automatically learned on the textual vocabularies and their most probable training images. $P(x_j|z_k)$ is represented as a 2D matrix where x_j is the number of textual words which is 68 and z_k is the number of classes in this training dataset which is 39. $P(z_k|d_i)$ is also represented as a 2D matrix where d_i is the number of training images which is 564.
2. We then consider that the aspects have been observed for these set of images **d** and train a second PLSA on the visual modality (BoW) to compute $P(x_j|z_k)$, keeping $P(z_k|d_i)$ fix which was learnt from step one. The resulting value for $P(x_j|z_k)$ is presented as 2 dimensional matrix XZ where X is the number of textual words and Z is the number of classes in dataset. In this experiment, the value of textual words(X) and number of classes (Z) is 68 and 39, respectively.

Automatic Annotation on the Test Image:

1. Given new visual features from the unseen test image and the previously calculated $P(x_j|z_k)$ parameters, $P(z_k|d_{test})$ is computed for a new image d_{new} using the standard PLSA procedure for a new image. Similar to $P(x_j|z_k)$, the resulting value for $P(z_k|d_{test})$ is shown as 2 dimensional matrix ZD where D (e.x. 5) is the number of test images.
2. The posterior probability of each word in the vocabulary is then computed by

$$P(x_j|d_{new}) = \sum_{k=1}^{k} P(z_k|d_{new})P(x_j|z_k) \tag{4.1}$$

This is performed by multiplying the two matrixes as follows:

C= XZ * ZD (4.2)

The result of multiplication is a 2-D matrix C with 68 rows and 5 columns where each column represents one of the test images. Thus, the top highest five numbers in every column are chosen where each one of them represents a word. As a result, the number of annotated keywords is five.

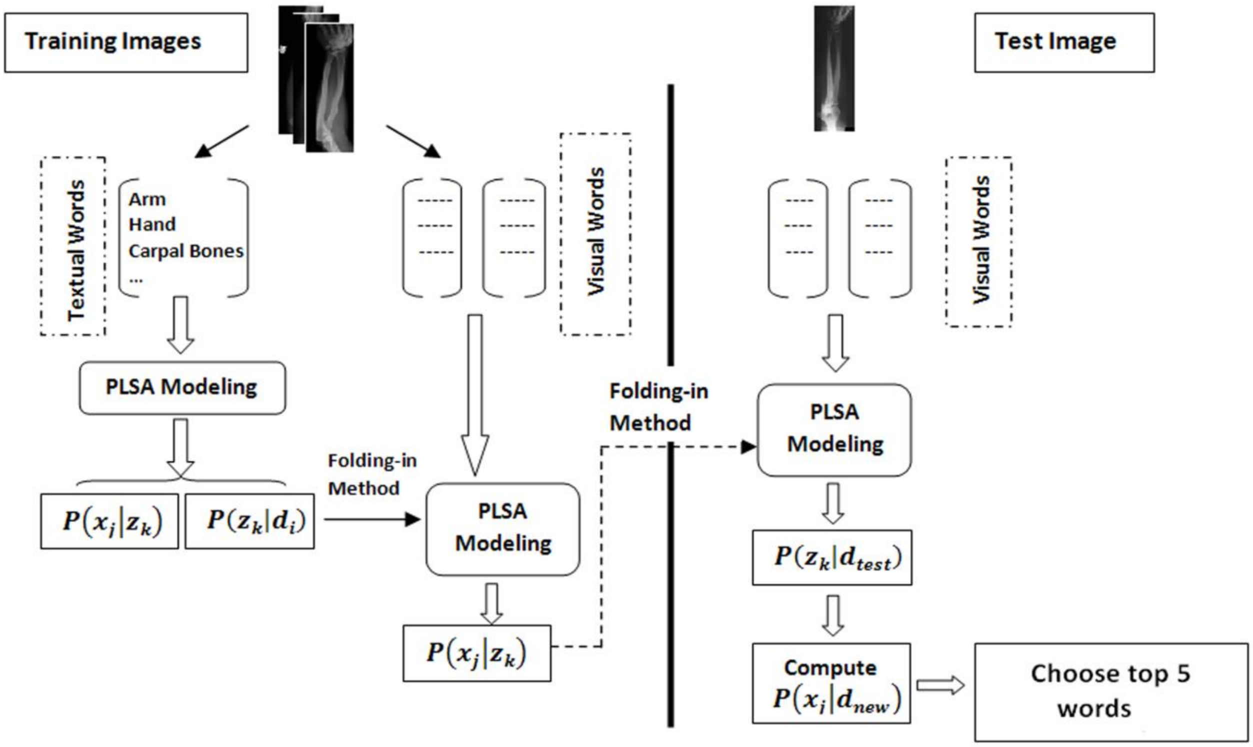


Figure 4.3: The flow of Training and Annotating using PLSA

Pseudocode of the annotation process is demonstrated in following:

```
Start
    1. Input Textual-Vocabulary $[T\_V]_{564 \times 68}$ to PLSA model to learn two matrixes $[X_t Z]_{68 \times 39}$ and $[ZD]_{39 \times 564}$.
    2. Input $[BoW]_{564 \times 500}$ to PLSA model in order to compute/update the matrix$[X_t Z]_{68 \times 39}$, by keeping matrix $[ZD]_{39 \times 564}$ fixed computed from step 1.
    3. The output is matrix$[X_t Z]_{68 \times 39}$.
    4. Insert the visual feature extracted from the unseen test image ($[BoW]_{1 \times 500}$ ) to PLSA model to compute matrix$[ZD]_{39 \times 1}$, by keeping $[X_t Z]_{68 \times 39}$ fixed from step 3.
    5. Multiply two matrixes [XZ] and [ZD] as follow: $[C]_{68 \times 1}$= $[X_t Z]_{68 \times 39}$ $\times$ $[ZD]_{39 \times 1}$.
    6. Output is matrix with 68 rows containing different values, where each rows represents one annotated keywords.
    7. Sort the matrix in descending order; the corresponding annotated keywords of the top highest five values are taken as annotation for the unseen test image.

End
```

4.2.3. Annotation using Top Similar Images

In this approach, the top five training images that are visually similar to the unseen test images would be retrieved followed by identifying the class that they belong to. The corresponding keywords to each class would be then taken as an annotation. These five sets of keywords are then combined to produce distinct set of keywords. The block diagram of retrieving similar images using PLSA is shown in Figure 4.4. The process of retrieving the top five similar training images is as follow:

Learning Phase:

1) Firstly, PLSA model is completely trained on the set of training images with visual words (BoW) as an input to learn both $P(z_k|d_i)$ and $P(x_j|z_k)$.
2) **While** not converge **do**
    a) E-Step: Compute the posterior probabilities $P\left(z_k\middle|d_i, x_j\right)$
    b) M-Step: Parameters $P\left(x_j\middle|z_k\right)$ and $P(z_k|d_i)$ are updated from posterior probabilities computed in the E-Step.

End While

Testing Phase:

1) The E-step and M-step are applied on the extracted BoW of the test image by keeping the probability of $P\left(x_j\middle|z_k\right)$ learnt from the training phase fixed.
2) Calculate the Euclidean distance between $P(z_k|d_i)$ and $(z_k|d_{test})$.
3) Those images with closest distance to $P(z_k|d_{test})$ will be retrieved as similar images.

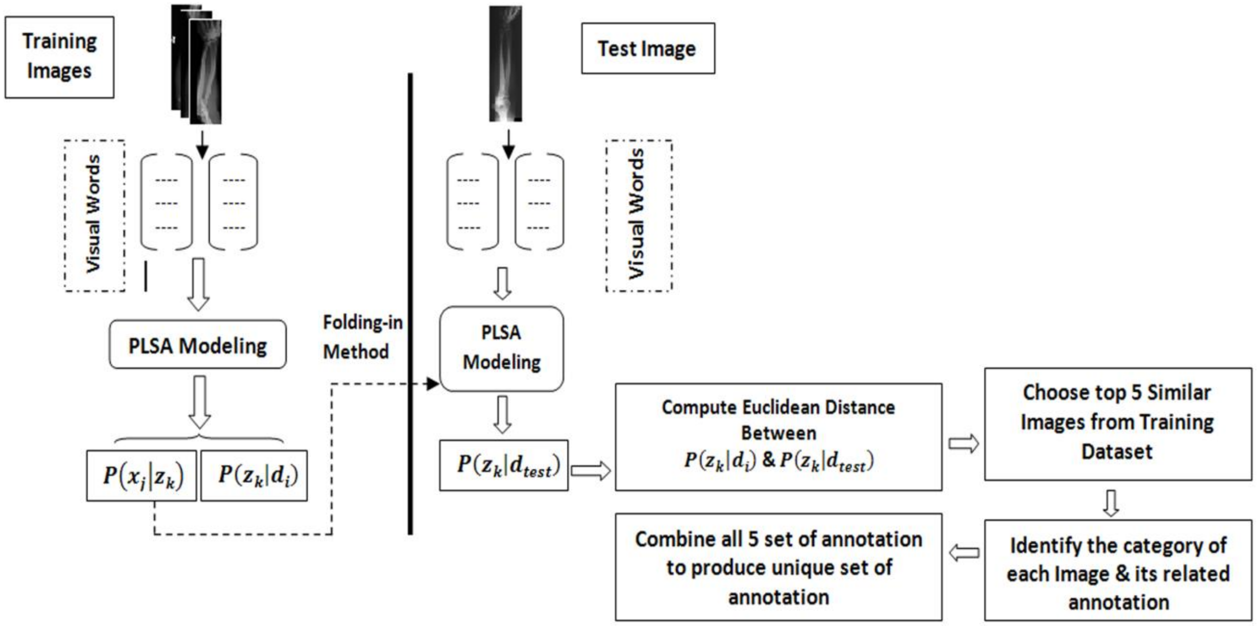

Training Images
Visual Words
PLSA Modeling
$P(x_j|z_k)$
$P(z_k|d_i)$
Folding-in Method
Test Image
Visual Words
PLSA Modeling
$P(z_k|d_{test})$
Compute Euclidean Distance Between $P(z_k|d_i)$ & $P(z_k|d_{test})$
Choose top 5 Similar Images from Training Dataset
Identify the category of each Image & its related annotation
Combine all 5 set of annotation to produce unique set of annotation

Figure 4.4: The flow of Retrieving Similar Images using PLSA

Pseudocode of the retrieval process of 5 similar images to test images is demonstrated in following:

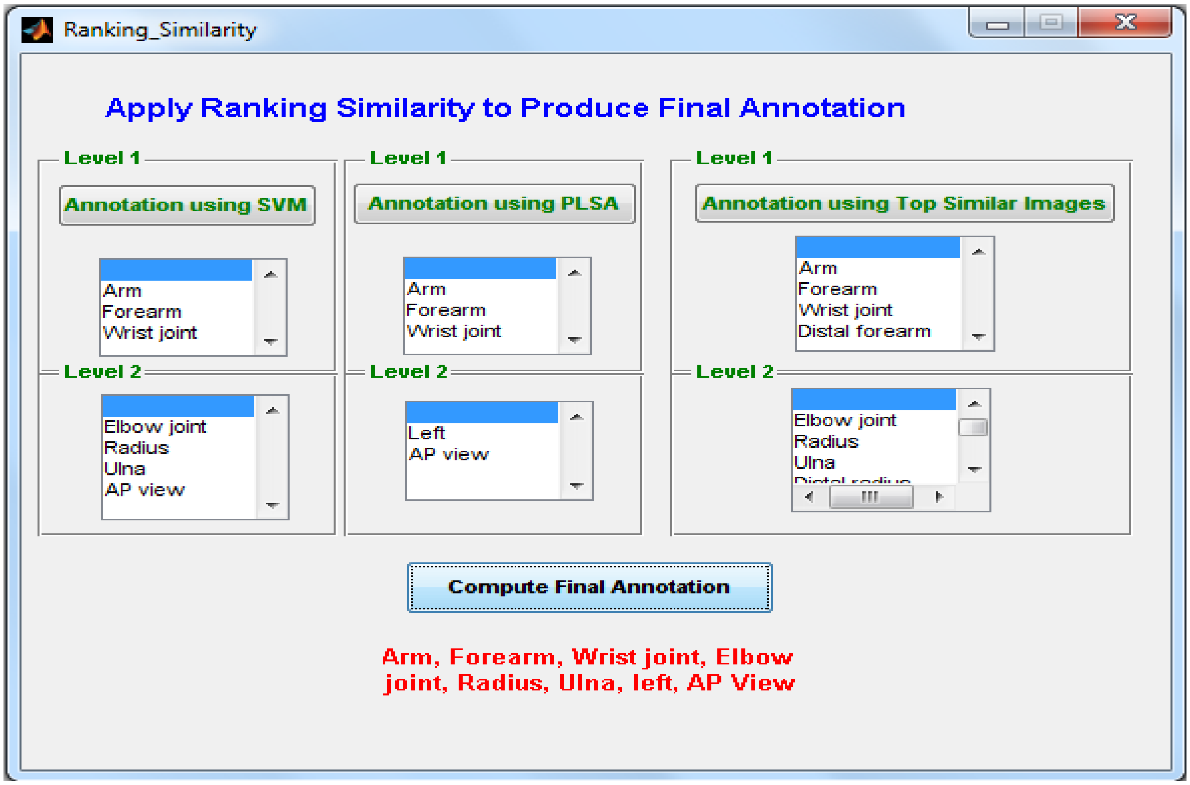
```
Start
    1. Input $[BoW]_{564\times 500}$ to PLSA model in order to compute $[XZ]_{500\times 39}$ and $[ZD]_{39\times 564}$
    2. Insert the visual feature extracted from the unseen test image ($[BoW]_{1\times 500}$ ) to PLSA model to compute matrix$[ZD]_{39\times 1}$, by keeping $[XZ]_{500\times 39}$ fixed from step 1.
    3. The output is matrix$[ZD]_{39\times 1}$.
    4. For i=1 to 564
           Compute Euclidean distance between $[ZD]_{39\times i}$ and $[ZD]_{39\times 1}$ .
           The top five vectors in matrix $[ZD]_{39\times i}$ with closet distance to vector $[ZD]_{39\times 1}$ will be selected. Each vector represents one image.
       End For
    5. Identify the category of each retrieved image.
    6. Get the annotated keyword of each category.
    7. Combine these five sets of keywords to produce final distinct annotated keywords.
End
```

4.2.4. Applying Ranking Similarities To produce Final Annotation

Based on the proposed algorithm, three sets of keywords were generated based on the above three approaches. Each set of the generated keywords was ranked according to their importance to describe the image. Two levels of ranking are applied on the keywords; those keywords which help to distinguish the body region clearly were ranked as the first level, those that describe the objects (specific bone structure) inside the specific body region as well as imaging direction and view are ranked as second level.

Two selection criteria were conducted on these keywords to generate the final annotation. The keywords from the first level which are common in all the three sets were selected to fulfill the first criteria. Those keywords from second level which are generated in any two sets were taken to perform the second selection criteria. The combination of all the selected keywords is the final annotation for the unseen test image.

Figure 4.5 is the sample screenshot that represents this process. The three sets of produced keywords are divided into two levels according to their importance. Each keyword was given a unique number as shown in Figure 4.2. These numbers are used to determine if the annotated words belong to level one or level two. Next the keywords are loaded into the respective list box as shown in Figure 4.5. The selection criteria are applied upon clicking on “Compute Final Annotation”. As can be seen from the screen shot, the first level keywords “arm, forearm, wrist joint” are common in all the three sets; and elbow joint, radius , ulna, left and AP view are appeared in any 2 sets of the keywords were taken as final annotation.



Arm, Forearm, Wrist joint, Elbow joint, Radius, Ulna, left, AP View

Figure 4.5: Interface to represent the process of producing final annotation

4.2.5. Classification

This is the final phase of the proposed classification framework. In this phase, the test images are classified into respected classes through their annotation keywords generated from the previous section. This is done by computing the *Total Weight* (*TW*) of the generated annotation based on the following equation:

$$TW = BR(L_1X + L_2Y + L_3Z) \qquad (4.3)$$

Firstly, the keywords from each body region are divided into three levels; the first level contains those keywords clearly representing the body region. Unlike the annotation phase, the keywords related to imaging direction and view are separated from the second level and formed as level three. *X*, *Y* and *Z* represent the three different levels of keywords.

A different weightage is given to every level of keywords. The weights given to level one and level two are *X=3* and *Y=2*, respectively. Variable Z represents the third level keywords which are imaging view and direction. Examples of such keywords are lateral view, coronal view, left, right and etc. These keywords are not specific to any body-region and some of them may be common in most of the images from different body regions, therefore Z is calculated as follow:

$$Z_{i1\dots i9} = \frac{Total\ \ number\ \ of\ key\,word_{\ i1\dots i9}\ occured\ \ in\ training\ \ set}{Total\ \ number\ \ of\ training\ \ images} \tag{4.4}$$

Variables L_1 , L_2 and L_3 are the number of keywords from level one , level two and level three appeared in its annotation, respectively .

The weight given for each level of the keywords is common for all the body regions. As demonstrated in Table 4.2, the weightage computed for given ground truth annotation for two different body regions is almost similar without multiplying with *BR* (the weight assigned for each body region) value. As such, in order to distinguish the body region from one another, a different value is allocated for each body region as shown in Table 4.1. The resulting value would help to differentiate the body region from one another more clearly. The weight for body region is represented by '*BR'* in equation 4.3.

Table 4.1: Weight assigned to each body region (*BR*)

	Abdomen	**Arm**	**Leg**	**Chest**	**Cranium**	**Spine**
Weight	10	20	30	40	50	60

Thus, the TW_{Test} calculated for each test image will be compared with the TW from each category of the training set, and then the test image classified to the category with closest weightage to TW_{Test}.

Table 4.2 shows the calculation of Total Weight *(TW)* and classification process of sample X-ray image from selected body regions. Every image category carries a different weight calculated based on its ground truth annotation. The same approach is used to compute the weight of the annotation made for the unseen test image. The computed weight of the test image is then compared with all the weight obtained from training dataset, and then the test image is classified to the category with closest weight to the test image's weight.

Table 4.2: Weightage calculation and classification process from selected body region

Ground Truth	**Arm, Distal forearm, distal radius, distal ulna, left, lateral view**	**Cranium, facial cranium, orbits, skull, AP view**
Level 1	Arm, Distal Forearm	Cranium, Facial Cranium
Weight	$L_1X = 2 \times 3$	$L_1X = 2 \times 3$
Level 2	Distal radius, Distal Ulna	orbits, skull,
Weight	$L_2Y = 2 \times 2$	$L_2Y = 2 \times 2$
Level 3	Left, lateral view	AP View
Weight	$Z_{left} = 0.25$ $Z_{lateral\ view} = 0.21$	$Z_{AP\ View} = 0.33$
Weight of Annotation Without Body Region $(L_1X + L_2Y + L_3Z)$	6 + 4 + (0.25+0.21) = **10.46**	6 + 4 + 0.33 = **10.33**
Total Weight including Weight of Body Region: $BR(L_1X + L_2Y + L_3Z)$	20(10.46) = **209.2**	50(10.33) = **516.5**

4.3. Experimental Results

In this section, we evaluate the classification performance of the proposed approach on ImageCLEF 2007 medical database. This experiment specifically conducted on those classes left with lower accuracy rate in previous chapter. These classes were labeled as LAC and contain 39 classes.

4.3.1. Parameter Optimization

LIBSVM software package has been utilized to perform discriminative-based classification with RBF kernel functions. The optimum kernel γ and cost C parameters has identified empirically with 5-fold cross validation. We use one-vs-one multi-class extension for SVM. The main parameter in construction of BoW is the number of k which represents the vocabulary size. As demonstrated in Table 2.4, various experiments were conducted with different vocabulary size starting from 100 followed by 200, 300, 400, 500, 600 and 700. Based on the experimental results, the best performance is achieved at k=500. Another parameter used in this experiment is the number of keywords used to construct the term-document matrix for textual vocabulary. These keywords are from LAC which consists of 39 classes. Totally 68 unique keywords were identified after eliminating the duplicate keywords. The measurement used to evaluate classification performance is average accuracy or also referred as accuracy rate. The measurements used to evaluate the performance of the annotation are recall and precision. Therefore, annotation recall and precision are computed for every word in the testing set. Recall and precision are averaged over the set of testing words. In the case of automatic image annotation, the aim is to get both high recall and precision.

$$Recall = \frac{\text{number of images annotated correctly with a given word}}{\text{number of images that have that word}} \tag{4.5}$$

$$Precision = \frac{\text{number of images annotated correctly with a give n word}}{\text{number of images annotated with that particular word}} \tag{4.6}$$

4.3.2. Classification Results

In this section, we analyze and evaluate the classification performance on the unseen test image. The first section of the proposed classification algorithm is annotation. There were three different techniques used to perform annotation. The first technique was based on the supervised classification. As such, we follow the classification model constructed on LAC in the previous chapter. That model classifies the unseen test image into one of the pre-defined classes. Figure 4.6 shows the classification results obtained by the model generated from these classes. The average accuracy rate reported was 72 %.

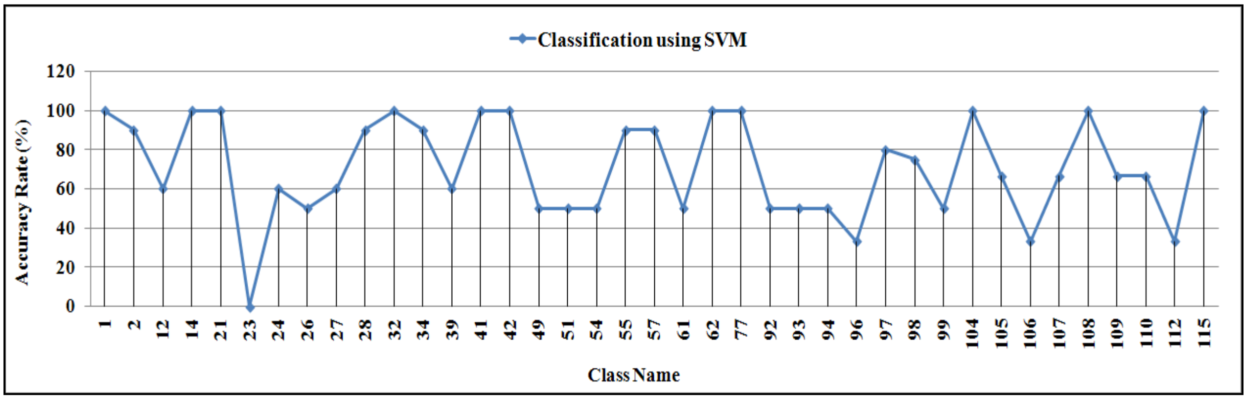


Figure 4.6: Accuracy rate obtained on 39 classes using SVM

Then, the ground truth annotated keywords of every class are assigned to the test image accordingly to produce the first set of annotation. The average recall and average precision of the annotation made by this approach are 0.79 and 0.80, respectively.

For the second annotation techniques, linked pair of PLSA model was applied on the extracted BoW of the test images. As explained in section 4.2, the top five words were selected as annotation keywords for the unseen test images. The average recall and average precision of the annotation made by this approach are 0.77 and 0.78, respectively.

For the third annotation technique, PLSA model was applied on the extracted BoW of the unseen test images. Next, the top five similar images to the test images were selected from the training dataset. The respective class labels for these five images are known because they are from the training dataset. As such, the related keywords of each one of those classes are assigned to the test images. They are then combined to generate the unique annotation for each unseen test image. The average recall and average precision of the annotation made by this approach is 0.85 and 0.86, respectively.

Subsequently, ranking similarity need to be applied on each set of annotation produced by the above three techniques to construct the final annotation. To do this, annotated keywords are divided into two levels based on their importance. Level one consists of those keywords that clearly represents the body region and level two contain those keywords that describe specific bone structure in the body region. Table 4.3 shows two levels of the keywords belonging to category of “arm”. The average recall and average precision of the final annotation made after applying ranking similarity is 0.93 and 0.94, respectively.

Table 4.3: Two levels of keywords belong to "Arm" body region

Level 1	Arm, wrist joint, shoulder joint, distal forearm, forearm, carpal bone
Level 2	Scaphoid, scapuls, distal radious, distal ulna, humerus, elbow joint, radious, ulna, upper humerus, Left, right, oblique view, lateral view, PA view, AP view, axial view

Upon construction of final annotation, the Total Weightage (TW_{Test}) of the annotated keywords is calculated for every unseen test images using equation 4.3. Thus, the computed TW_{Test} will be compared with the TW from each category of the training set, and then the test image classified to the category with closest weightage to TW_{Test} . The average accuracy rate reported by this approach on 39 classes under LAC is 87.5 %. In Figure 4.7, the classification performance obtained by this approach is compared with the result obtained by supervised SVM model for every individual class. The results show an improvement in classification performance as compared with single SVM classifier.

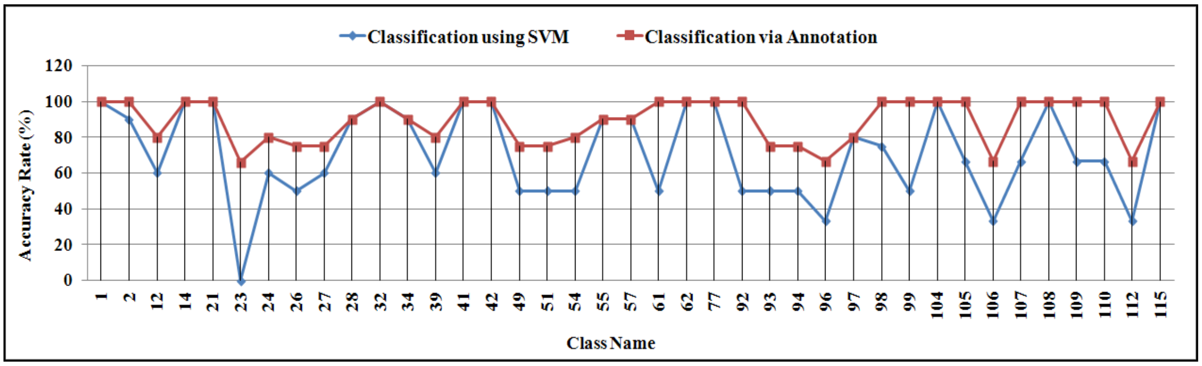
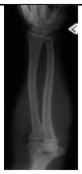
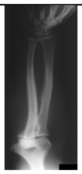
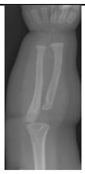


Figure 4.7: Comparison on classification result on LAC obtained by SVM and proposed annotation

The annotation and classification results on the unseen test images from class 23 are further illustrated in Table 4.4. This class contains three test images. As shown in Figure 4.7, we got two images classified correctly with the proposed method whereas the zero accuracy rates were reported using SVM classifier.

Table 4.4: Classification results on selected test image from class 23

Query Test Image			
Ground Truth	**Arm, forearm, wrist joint, elbow joint, radius, ulna, left, AP view**	**Arm, forearm, wrist joint, elbow joint, radius, ulna, left, AP view**	**Arm, forearm, wrist joint, elbow joint, radius, ulna, left, AP view**
Weight: **Class No.:**	**311.60** **23**	**311.60** **23**	**311.60** **23**
Annotation using classification(SVM)	***Arm, forearm, wrist joint, elbow joint, radius, ulna, AP view***	***Arm, forearm, wrist joint, elbow joint, radius, ulna, AP view***	***Arm, forearm, wrist joint, elbow joint, radius, ulna, AP view***
Annotation using PLSA	***Arm, forearm, wrist joint, distal forearm, distal radius, distal ulna, left***	***Arm, forearm, wrist joint, left, AP View***	***Arm, forearm, Wrist Joint, right, AP View***
Top Similar Image 1	Arm, forearm, wrist joint, elbow joint, radius, ulna, left, AP view	Arm, forearm, wrist joint, elbow joint, radius, ulna, left, AP view	Arm, forearm, wrist joint, elbow joint, radius, ulna, left, AP view
Top Similar Image 2	Arm, forearm, wrist joint, elbow joint, radius, ulna, AP view	Arm, Distal forearm, Distal Radius, Distal Ulna, Left, AP View	Arm, Distal forearm, Distal Radius, Distal Ulna, Left, AP View
Top Similar Image 3	Arm, forearm, wrist joint, elbow joint, radius, ulna, AP view	Arm, forearm, wrist joint, elbow joint, radius, ulna, lateral view	Arm, forearm, wrist joint, elbow joint, radius, ulna, lateral view
Top Similar Image 4	Arm, forearm, wrist joint, elbow joint, radius, ulna, lateral view	Arm, forearm, wrist joint, elbow joint, radius, ulna, AP view	Arm, forearm, wrist joint, elbow joint, radius, ulna, AP view
Top Similar Image 5	Arm, forearm, wrist joint, elbow joint, radius, ulna, right, lateral view	Arm, Distal forearm, Distal Radius, Distal Ulna, Left, AP View	Arm, Distal forearm, Distal Radius, Distal Ulna, Left, AP View
Annotation using top similar images	***Arm, forearm, wrist joint, elbow joint, radius, ulna, left, right, Lateral view, AP view***	***Arm, forearm, wrist joint, elbow joint, radius, ulna, Distal Forearm, distal radius, Distal Ulna, left, Lateral view, AP view***	***Arm, forearm, wrist joint, elbow joint, radius, ulna, Distal Forearm, distal radius, Distal Ulna, left, Lateral view, AP view***
Final Annotation	**Arm, forearm, wrist joint, elbow joint, radius, ulna, left, AP View**	**Arm, forearm, wrist joint, elbow joint, radius, ulna, left, AP View**	**Arm, Forearm, Wrist Joint, elbow joint, radius, ulna, AP View**
Weight: **Class No.:**	**311.60** **23**	**311.60** **23**	**306.6** **96**

4.3.3. Discussion

In this chapter, we gave a special attention to those classes with high ratio of intra class variability and inter-class similarities. In order to explain how the proposed annotation framework rectifies the above mentioned problems, drill down analysis has been applied on those classes with high ratio of intra class variability and inter-class similarities. After the first iteration in previous chapter, 39 classes were left with accuracy below 80%. Figure 4.6 shows the classification results obtained by the model generated from these classes. The average accuracy rate reported was 72 %.

Even though the number of classes involved in the model generated from LAC is lesser, yet to obtain a good classification performance. As can be seen in Figure 4.6, there are 13 classes with accuracy below 60%.

We have done detailed investigation on these classes to know the reason of low accuracy rate. Seven of them belong to “arm” body region. In ImageCLEF 2007 dataset, there are 33 classes under “Arm” body region. These classes are distributed into 6 sub-body regions. Based on the result obtained from Figure 4.6, the seven classes with accuracy below 60% belong to three sub body region of “arm” as shown in Table 4.5.

Table 4.5: Number of classes in every sub-body regions of “Arm”

Sub-body region	**Number of classes**
Forearm	4
shoulder	1
hand	2

In Table 4.6, the confusion matrix is created for these three sub-body regions. As can be seen, out of 12 test images in category of “forearm”, 10 of them were classified correctly. One of the test images from the categories of hand and shoulder were misclassified as presented in confusion matrix.

Table 4.6: Confusion matrix on sub-body region of “Arm”

	forearm	**Shoulder**	**Hand**	**Other region**	**Accuracy Rate (%)**
Forearm	10		1	1	83 %
shoulder		4		1	80 %
hand			8	1	89 %

The number of classes for each sub-body region of “Arm” is listed in Table 4.7. We also show the number of classes of this sub-body region with accuracy rate of 60% and above in Table 4.7. As can be seen from Table 4.7, none of the four classes under “forearm” sub-body region could managed to attain accuracy rate of 60 % even though the accuracy rate of this sub-body region was reported 83% in Table 4.6. This analysis represents the high ratio of misclassification among classes under “forearm” body region. This misclassification can be seen in the other two sub-body region of “Arm”.

Table 4.7: Number of classes per each sub-body regions

	Number of classes	**Number of classes with accuracy of above 60%**
Forearm	4	0
shoulder	1	0
hand	2	0

High ratio of inter-class similarities and intra class variability among these classes is the main reason of misclassification. Inspired from this fact, we proposed a classification framework which utilizes the annotated keywords of the images to improve the classification performance.

Annotation module in the proposed framework plays a very important role in the proposed classification algorithm; a good performance in annotation would improve the classification performance. In Table 4.4, annotation and classification results on test images from class 23 are represented. There are three test images provided for class 23 that are misclassified as class 96 using SVM. Both class 23 and 96 are referring to ‘forearm’ sub-body region, they are distinguished from each other only in direction. Class 23 has the direction “left” in its annotation but no direction stated for class 96. Meaning that in the case of annotation using classification, the corresponding annotated keywords from both classes are almost similar.

As for annotation using top similar images, mostly the top five retrieved similar images to the query image are belonging to the same class or same sub-body region. Thus, they are sharing most of the important keywords. In the case of annotation using PLSA, a linked pair of PLSA models is employed to capture semantic information from textual and visual modalities and learn the correlation between them. It is clear that this structure can predict all the important keywords (level one) correctly. Therefore, the combined set of keywords generated from this annotation technique would contain most of the keywords of the respective sub-body region.

The experimental results obtained on the entire database shows an improvement in probability of getting more accurate annotation by fusing the above three techniques which would lead to an increment in classification accuracy. By observing the results obtained from Table 4.4, it is clearly evident that all the three techniques in particular, annotation using PLSA and annotation using top similar images, certain classes can be effectively used to annotate correctly and accurately compared to SVM classification because it incorporate both textual and visual features of the images. Accuracy rate obtained by the proposed annotation algorithm shows tremendous improvement compared to classification rate obtained by SVM as illustrated in Figure 4.7.

4.4. Chapter Summary

In this chapter, a classification framework is proposed to improve the accuracy rate of those classes of medical x-ray images with great intra class variability and inter-class similarities. This classification task carried out by employing three different annotation techniques such as annotation by binary classification, PLSA-based image annotation and annotation using top similar images to the query image. The final annotation is then constructed by utilizing ranking similarity on annotated keywords. Next, the final annotation is used for classification purpose by computing their weightage and comparing with each category’s weightage in database. The experimental result shows that the accuracy rate obtained outperformed the others using conventional supervised classifier such as SVM.

Chapter 5 Automatic Medical X-Ray Image Classification using Hybrid Generative-Discriminative Approach

5.1. Introduction

Bag of Words model has been used as the major image representation approach in this thesis due to its simple and discrete data representation. However, it also introduces the well known synonymy and polysemy ambiguities. In this chapter, a generative Multi-Modal PLSA model is proposed to address this shortcoming by helping to disambiguate visual words. To this end, we proposed a classification framework to improve the classification performance of medical X-ray images based on the combination of generative and discriminative classification approach. The experimental results were based on ImageCLEF 2007 medical database. The classification performance was evaluated on the entire database as well as the class specific level. It was also compared with other classification techniques with various image representations on the same database. The comparison results showed that the superior performance has been achieved especially for classes with less number of training images. The results of this experiment have been published in (Zare et al., 2013d).

In next section, the well known ambiguities of BoW are further explored. We then explain the hybrid classification model followed by the experimental results.

5.2. Ambiguity with BoW

The ambiguity lies into two areas: *Visual Polysemy* where single visual word occurring on different parts on different object categories and *Visual Synonyms* where two different visual words representing a similar part of an object. To illustrate these issues, consider images taken from different body region. For instance, “chest” and its sub-body regions are well defined classes that are dominated by high frequency visual words and thus do not get confused with other classes. On the contrary, “arm” category contain bigger number of sub-body regions as compared to chest, and therefore it presents high intra class variability and inter-class similarity in certain classes within this category which would affect on the classification performance. This ambiguity is due to visual words co-occurrences across images that do not often entail a semantic relationship between them. It is argued that this ambiguity could be occurred during codebook creation. Codebook or visual vocabulary creation plays the important role in the construction of bag of visual words model. This process is carried out based on clustering algorithm such as k-means which is coarse and does not select the most informative descriptors as it tends to ambiguous data representation (Monay et al., 2005; Tirilly et al., 2008)

5.3. Hybrid Generative-Discriminative Classification

Classification process is a two steps process of training and testing phases. In the training phase, the selected features are extracted from all the training images, and a classifier is trained on the extracted features to create a model. The proposed classification framework is named as Hybrid Generative-Discriminative approach for the classification process. It is carried out in two different generative and discriminative approaches as illustrated in Figure 5.1.

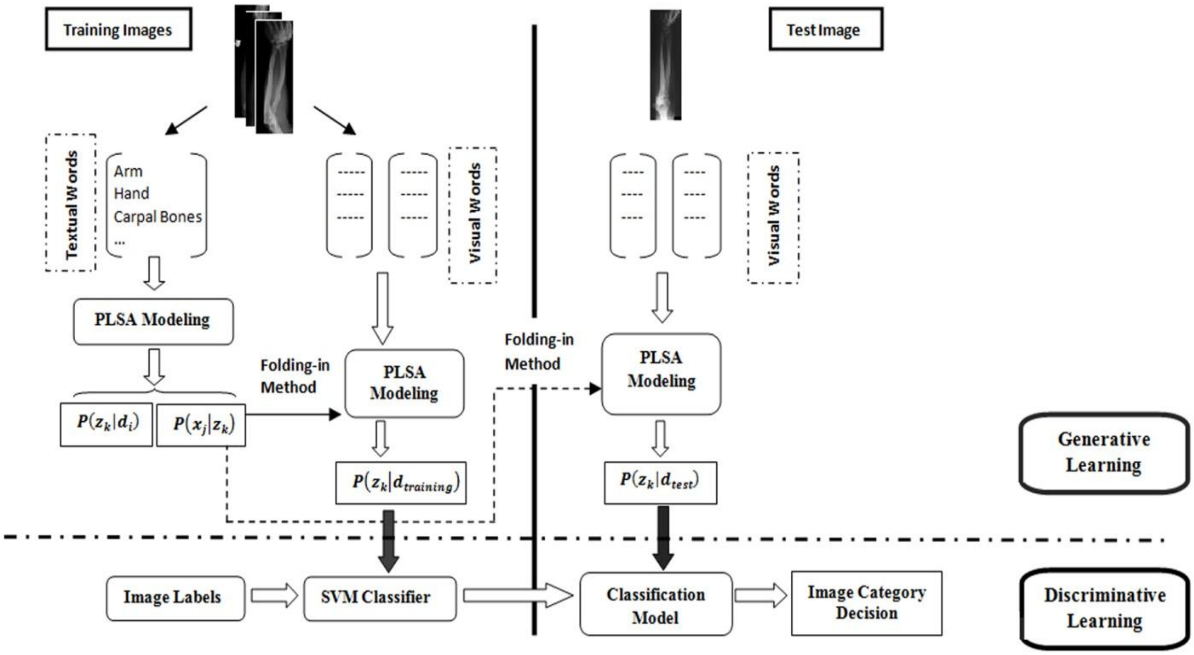


Figure 5.1: Block Diagram of the Proposed Classification Algorithm

Upon extraction of BoW and textual words from the entire training dataset (Refer to sections 3.3.1.1 and 4.2.2 for details) , a linked pair PLSA model was used to associate a latent variable *(Z)* by observing the occurrence of a visual word and textual word in every document/image *(d)*. As a result, high level representation of images is constructed in the generative approach. They are then used as an input to the discriminative classifier in order to construct a classification model. The training and testing of the proposed classification framework are as follow:

Training Phase:

1. First PLSA model is completely trained on the set of image captions (textual vocabulary) to learn both $P(x_j|z_k)$ and $P(z_k|d_i)$ parameters.

2. Next, the second PLSA in trained on the BoW of the same set of documents to compute $P(z_k|d_i)$, keeping $P(x_j|z_k)$ from above fixed.

The unobservable probability distribution $P(z_k|d_i)$ and $P(x_j|z_k)$ are learned from the data using the Expectation –Maximization (EM) algorithm. $P(z_k|d_i)$ denotes the probability of topic z_k given in document d_i. $P(x_j|z_k)$ denotes the probability of visual word/ textual word x_j in topic z_k.

These parameters are used to infer the topic mixture parameter $P(z_k|d_i)$ which denotes the probability of topic z_k for any image based on its given BOW. As a result, each training image is represented by a $P(Z|d_{training}$). where Z is the number of topic learnt. This representation and the respective image labels are later used as input to a discriminative classifier such as Support Vector Machine (SVM) to construct a classification model.

Testing Phase:

Classification of the unseen test image was carried out by running the E-step and M-step (EM algorithm) on extracted BoW of the test image by keeping the probability of $P(x_j|z_k)$ learnt from the training phase fixed. As a result, the test image is represented by a $P(Z|d_{test})$. Then the classification model is used to make a decision on the test image category.

5.4. Experimental Results

In this section, the classification performance of the proposed framework is evaluated. Similar to the previous two chapters, the database used in this experiment was ImageCLEF 2007.

5.4.1. Parameter Optimization

LIBSVM software package has been utilized to perform discriminative-based classification with RBF kernel functions. The optimum kernel γ and cost C parameters has identified empirically with 5-fold cross validation. We use one against one multi-class extension for SVM. The measurement used to compare classification results is average accuracy or also referred as accuracy rate. It is a ratio of number of images classified correctly over the size of test dataset.

The main parameter in construction of BoW is the number of k which represents the vocabulary size. As demonstrated in Table 2.4, various experiments were conducted with different vocabulary size starting from 100 followed by 200, 300, 400, 500, 600 and 700. Based on the experimental results, the best performance is achieved at *k*=500. The term-document matrix constructed for textual vocabulary consists of 110 unique keywords.

5.4.2. Classification Result on Different Vocabulary Size

In this section, we evaluate how the classification performance is affected by the various visual vocabulary sizes both in hybrid generative-discriminative approach and in purely discriminative approach. In discriminative approach, BoW representation of the images is used as inputs to SVM classifier directly. In Hybrid Approach, unsupervised multi-modal PLSA was employed in order to get a more stable representation of the image upon construction of BoW. It was then used as an input to the supervised SVM classifier to build classification model.

Figure 5.2 and Figure 5.3 present the classification performance obtained by both approaches. As it can be seen, the best performance is achieved at V=500 for both approaches. Thus, V=500 was taken as optimal vocabulary size in this study. The best classification accuracy rate obtained by Hybrid Generative-Discriminative approach and Discriminative approach is 92.5 % and 90 %, respectively. Figures 5.2 and 5.3 also demonstrated that the classification result obtained from hybrid approach is less affected by vocabulary size than using pure discriminative approach. The lowest and highest classification rates obtained from both approaches are at these two points, V=100 and V=500, respectively.   In hybrid approach, the difference between classification performances obtained from these two vocabulary size is only 4.5 % where in discriminative approach is 8 %. Therefore, this result supports that the changes in vocabulary size is not affecting the classification performance attained by hybrid approach as compared to discriminative approach.

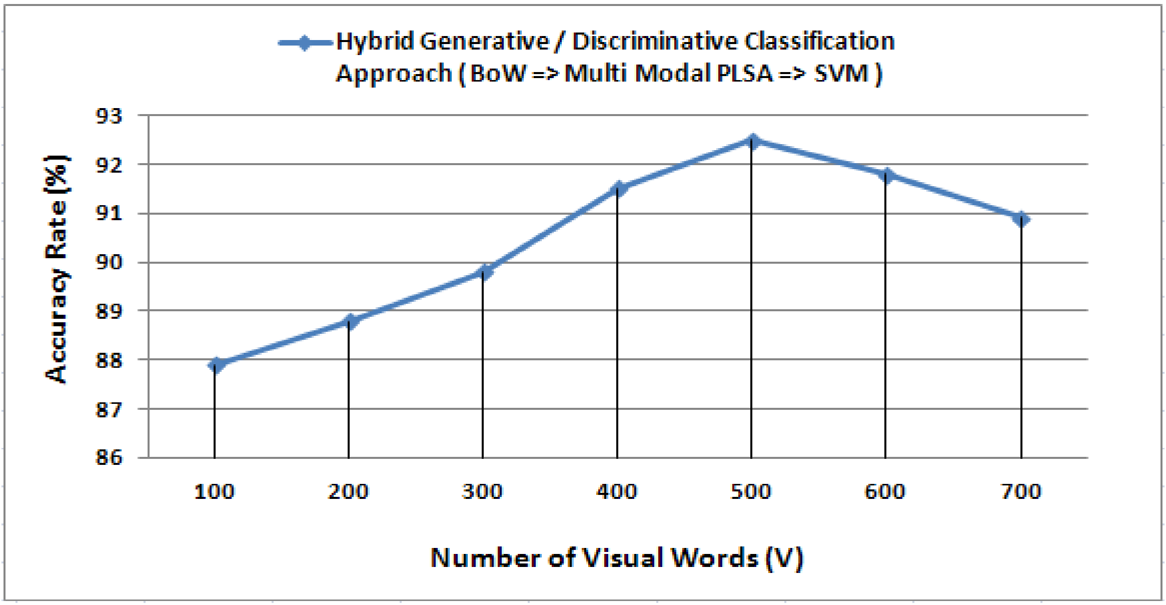


Figure 5.2: Classification Rate Obtained from Different Number of Visual Words using Hybrid Approach

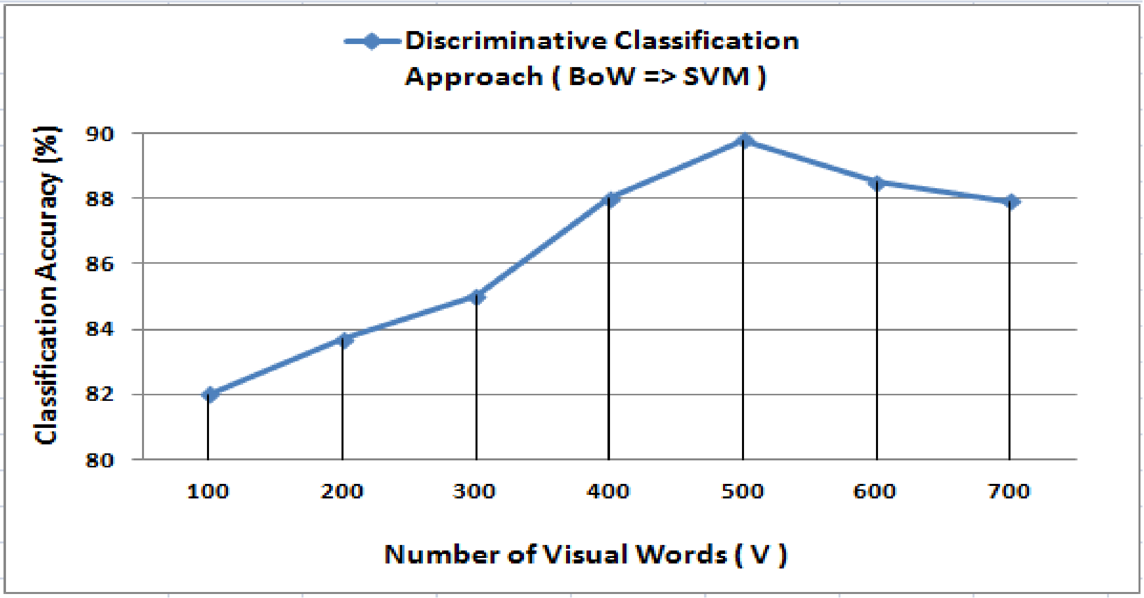


Figure 5.3: Classification Rate Obtained from Different Number of Visual Words using Discriminative Approach

5.4.3. Discriminative Approach Vs. Hybrid Generative-Discriminative Approach

The usage of PLSA in hybrid approach provides more stable representation of data which is less influenced by ambiguity of data. Knowing that PLSA is an unsupervised classification approach where there is no reference to the class label is required during the aspect model learning, the question arise that how discriminative are the information reside on the aspect model which can improve the classification performance. To answer this question, we compare the classification results obtained by both hybrid generative-discriminative approach and purely discriminative approach.

In Figure 5.4, the classification performance is shown in a class specific level. As can be seen, the result obtained from purely discriminative classification approaches shows that there were 30 classes with the accuracy rate of below 60% whereas in hybrid generative/discriminative approach, only 7 classes attained accuracy rate below 60%.

By further analysis on classification performance, we remark that the performance is similar overall in both approaches for those classes with large number of training data. In those classes with less number of labeled training data, the hybrid generative-discriminative classification approach provides lower error rate and produce better classification accuracy than purely discriminative approach. As stated earlier, one of the open challenges of such large archive medical database is the unbalanced number of training data. As we know, discriminative classifier such as SVM is more accurate when there are many labeled training data available whereas images from small classes are frequently misclassified. But the proposed hybrid generative-discriminative approach deals better with classes with limited number of labeled training images because the class label are not required during the aspect model learning in PLSA.

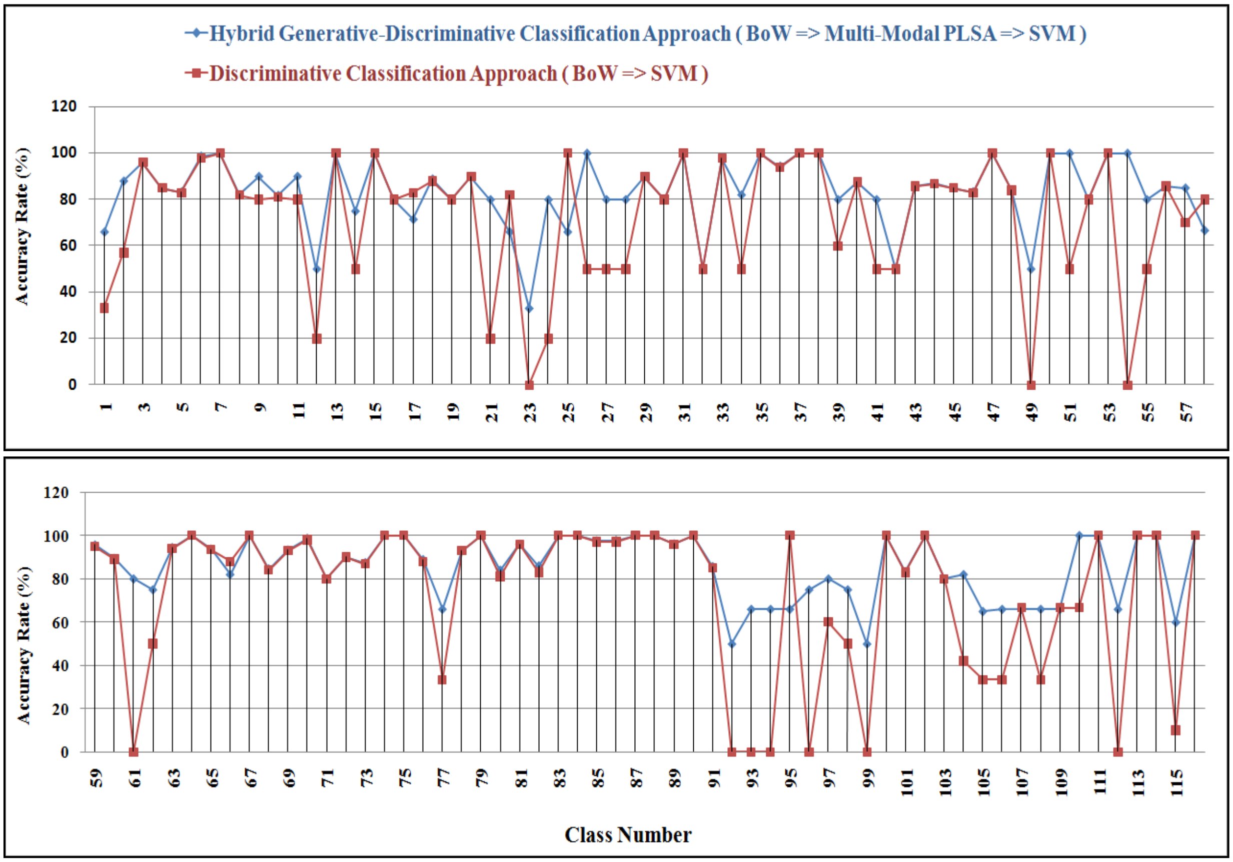


Figure 5.4: Comparison of Classification Performance using Hybrid Generative - Discriminative Approach and Discriminative Approach

To validate this, we take an example of the most complex categories in the database with the highest miss-classification ratio which are under "forearm" sub-body region. Table 5.1 and Table 5.2 demonstrate the confusion matrix between the eight categories in this sub-body region obtained by discriminative approach and Hybrid Generative-Discriminative Approach, respectively. The number of training images for each category is given in the first column.

Table 5.1: Confusion Matrix for Categories under Forearm Sub-body Region Obtained by (BoW ⟶ SVM) approach

No. of Training Images		Class 16	Class 21	Class 23	Class 24	Class 26	Class 96	Class 97	Class 110	Other	Accuracy Rate (%)
32	Class 16	4	1								80.0 %
27	Class 21	2	1					1		1	20.0 %
15	Class 23	2	1								0.0 %
25	Class 24	2	1		1					1	20.0 %
13	Class 26				1	1					50.0 %
14	Class 96				1			2	1		0.0 %
26	Class 97	1			1			3			60.0 %
17	Class 110				1				2		60.0 %

Table 5.2: Confusion Matrix for Categories under Forearm Sub-body Region obtained by (BoW ⟶ Multi-Modal PLSA ⟶ SVM)

No. of Training Images		Class 16	Class 21	Class 23	Class 24	Class 26	Class 96	Class 97	Class 110	Other	Accuracy Rate (%)
32	Class 16	4	1								80.0 %
27	Class 21	1	4								80.0 %
15	Class 23	1	1	1							33.3 %
25	Class 24	1			4						80.0 %
13	Class 26					2					100 %
14	Class 96						3	1			75.0 %
26	Class 97				1			4			80.0 %
17	Class 110								3		100 %

It can be seen that the classification rate has increased for all the eight categories under forearm as compared to the result obtained using BoW ⟶ SVM approach where both experiments are conducted on the same number of training and test data.

Another innovative contribution of this work to generate a better intermediate representation of the images is to exploit the multi-modal PLSA where both visual vocabulary and textual vocabulary are incorporated in formulation of PLSA representation. Based on our empirical studies, we have chosen textual vocabulary first in order to get the mixture of latent in a given image. This is to ensure the consistent set of textual words is predicted while retaining the ability to jointly model the visual features.

5.5. Chapter Summary

A proposed hybrid generative-discriminative classification model is promising based on the presented classification results. In the proposed approach, classification task started with extracting bag of visual words from all the X-Ray images. In discriminative classification approach, the extracted BoW is directly fed into SVM classifier to construct the classification model where in hybrid approach generative-discriminative approach; multi-modal PLSA-based representations of images are computed by just fitting extracted BoW into linked pair of multi-modal PLSA model. In this approach, multi-modal PLSA-based representations of images are used as input for SVM classifier. The experimental results show that multi-modal PLSA-based representation is competitive with Bag of visual words representation of images in terms of performance. It has also shown that using generative model (multi-modal PLSA-based representation) is less affected by vocabulary size and number of training images in classification performance. We have shown that the proposed hybrid generative-discriminative classification framework increases the accuracy at the entire database level as well as at class specific level. Based on the experimental result, a fair claim can be made that proposed classification framework outperforms the results obtained from similar relevant work as presented in Table 2.2 (Refer to section 2.10 in Chapter 2).

Chapter 6 Conclusion

6.1. Overview

Over the last decade, storage of non text based data in database has become an increasingly important trend in information management. Many medical images are acquired everyday in any modern hospital due to the rapid development of digital medical imaging techniques and information technologies. One outcome of this trend is an enormous increase in number of X-ray images. As a result, there is a concomitant demand for a computerized system to manage these valuable resources and aid radiologist in diagnosis process. The use of automatic image classification is demonstrated to be useful in improving the precision of such medical system.

However, obtaining a good performance in classification task of large medical database is a very difficult task. This challenge has to be analyzed from the domain perspective rather than focusing on the universality of the classification method and finding one classification method for all domains. As such, compared with other classification domains there are some particular difficulties when working on large medical database such as unbalance number of images among classes, intra class variability and inter-class similarity.

In this thesis, different classification frameworks were proposed to address these issues. In the following section, we briefly summarize the main achievement of this research.

6.2. Achievements and Contributions

A set of experiments were conducted to identify the suitable image representation and classification techniques in the proposed frameworks. These experiments were carried out using various image representation and classification techniques on the large medical database with the same number of training data and test data. Image representation techniques such as pixel value, GLCM, LBP, BoW were employed locally and globally. SVM and KNN were used as classification techniques in those experiments (Zare *et al.*, 2013a). The classification performance obtained by this experiment is demonstrated in Table 2.3. It can be seen that BoW as image representation techniques with non-linear multi class SVM with RBF kernel outperformed the other presented approaches. We used one-versus-one extension of binary SVM classifier where the classifiers are trained for all pairs of classes in the database.

In this thesis, BoW is used as an image representation technique and SVM as a classification technique to address the first challenge stated in this research. However, the average accuracy rate obtained is 90.0 % by utilizing BoW and multi class SVM. Further analysis on the results show there are 77 classes with accuracy rate above 80 %. The accuracy rate obtained is at global level, meaning the performance is obtained on the entire database. This result may not be achieved in every individual class due to the challenges in classification of such database. As such, we have proposed different classification frameworks. In the following sections, the classification performance achieved by each proposed framework is discussed with regard to how the classification challenges are addressed.

6.2.1. Classification using Iterative Filtering

By observing and analyzing the accuracy rate of every individual class obtained with discriminative approaches, it is clear that almost all large categories of medical images in the database have accuracy rates of above 85% whereas images from small classes are frequently misclassified. This observation shows that discriminatively trained classifiers are usually more accurate when labeled training data is abundant. At the same time, most of the classes with low accuracy rate are those with high ratio of intra class variability and inter-class similarity. To address this issue and increase the number of classes with high accuracy rate, a classification framework was developed to perform filtering on the dataset in several iterations, and consequently a separate model is constructed from each iteration. The idea is to filter out major classes with good accuracy rate in the first iteration. Subsequently, the next iteration only deals with less predominant classes. Indeed the generated model constructed in every iteration consists of those classes with an optimum accuracy rate.

The advantage of this filtering scheme is that those classes with high ratio of intra class variability and inter-class similarity are separated from one another. They are combined with other classes to form a new classification model. We have seen that the classification performance obtained by this framework, which has been successfully published (Zare *et al.*, 2013b), outperformed other techniques presented in Table 2.2.

6.2.2. Classification using Annotation

The classification result obtained from those classes left with low accuracy rate in the first iteration of the previous mentioned experiment in section 6.2.1, showed that most of them are misclassified within their own sub-region due to the high ratio of intra class variability and inter-class similarities. The fact that most of the misclassification are from those classes with the same body region motivated us to use different annotation techniques. Annotation performance of every technique varies from one another depending on the body region in medical database. The idea is to take advantage of the three different approaches in annotation to get to the closest class/category to the test image by considering the weightage of each annotated keywords produced by the annotation approaches. As such, we utilized three different annotation techniques i.e. (i) Annotation by supervised classification, (ii) Annotation by Probabilistic Latent Semantic Analysis (PLSA) and (iii) Annotation using top similar images. Indeed, each annotation technique is a complementary for the other annotation techniques. As a result, the combined set of keywords generated from these annotation techniques would contain most of the keywords of the respective sub-body region. The classification results obtained which has been successfully published (Zare *et al.,* 2013c); show an improvement in probability of getting more accurate annotation and classification accuracy.

6.2.3. Classification using Hybrid Generative-Discriminative Approach

Apart from several advantages of BoW such as its simplicity, discrete representations and simple matching measures to preserve computational efficiency, it still has several drawbacks such as ambiguous data representation which was discussed in detail in chapter 5. To overcome this shortcoming, a generative model such as PLSA has been proposed to learn the 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 due to the ability of the PLSA model to generate a robust, high level representation and low-dimensional image representation. This is because PLSA introduces a latent, for example a topic layer between an image and its visual words. It is assumed that each image consists of multiple topics and the occurrences of visual words in images are a result of topic mixture. So it can be said that high level representation based on the visual features is found once the topic mixture is derived for each image. As such, it is a dimension reduction of the image representation as the number of concept (topic) is smaller than the number of visual words in every image. Thus, a classification framework based on integration of PLSA and discriminative SVM classifier is developed which has been successfully published (Zare *et al.,* 2013d). In this framework, both visual features and textual features of the images are incorporated by using multi-modal PLSA model. The key advantage of the hybrid generative-discriminative classification approach is that the classification performance is better than purely discriminative approach especially for those classes with less number of labeled training images. We have shown that the proposed hybrid generative-discriminative classification framework increases the accuracy at the entire database level as well as at class specific level.

6.3. Future works and Directions

In addition to the contributions presented in this thesis which shows the fulfillment of the objectives, a number of open questions were discovered. The following issues can be considered for further work:

- In construction of BoW, the choice of the quantizer can be an issue. Although K-means is acceptable, it is not the most sufficient choice for this task because it fails to capture information about the distribution of data on other regions of the feature space. That would be possible future research where other methodologies can be used to extract a vocabulary that is more informative and less noisy.

- Another possible research is to extend the proposed hybrid generative-discriminative approach with filtering and annotation scheme presented in chapter three and chapter four.
- Yet another way to improve the classification performance and address the challenges of intra class variability and inter class similarity is to use a merging scheme. Merging scheme is used to combine overlapped classes with each other. Certain criteria can be set to detect overlapped classes such as the accuracy rate, miss-classification ratio and similarity in their body anatomy.
- Since the focus of this research was on medical image classification, one of the potential future works is to incorporate these classification frameworks with any of the existing medical CBIR system. With the integration of our classification frameworks, the respective medical CBIR system would be able to respond correctly to more specific queries.

References

A

Amores J., Sebe N., and Radeva P, (2007) Context-based object-class recognition and retrieval by generalized correlograms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(10):1818–1833,

Andre B, Vercauteren T, Perchant A., Buchner A.,Wallace M., and Ayache N., (2010). Introducing space and time in local feature-based endomicroscopic image retrieval, in Medical Content-Based Retrieval for Clinical Decision Support, B. Caputo, H. Müller, T. Syeda-Mahmood, J. Duncan, F. Wang, and J. Kalpathy-Cramer, Eds. Berlin/Heidelberg: Springer, vol. 5853, *Lecture Notes in Computer Science, pp. 18–30.*

Antani, S., Long, L. R., Thoma, G. R., & Lee, D.-J. (2003). Evaluation of shape indexing methods for content-based retrieval of x-ray images. 405-416. doi: 10.1117/12.476289

Avni, U., Greenspan, H., Konen, E., Sharon, M., & Goldberger, J. (2011). X-ray Categorization and Retrieval on the Organ and Pathology Level, Using Patch-Based Visual Words. *Medical Imaging, IEEE Transactions on, 30*(3), 733-746. doi: 10.1109/tmi.2010.2095026

Avni, U., Jacob , G., Michal , S., Eli , K., & Hayit , G. (2010). Chest x-ray characterization: from organ identification to pathology categorization. *Paper presented at the Proceedings of the international conference on Multimedia information retrieval, Philadelphia*, Pennsylvania, USA.

B

Blei D.M., Ng A.Y., and Jordan M.I., (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993– 1022,

Bo Q., Wei XIONG, Qi TIAN, & XU, C. S. (2005). Report for Annotation task in ImageCLEFmed 2005. *In: working notes of CLEF 2005. (Vienna, Austria, 2005).*

Bosch A., Muñoz X., Oliver A., and Marti J., (2006a). Modeling and classifying breast tissue density in mammograms, *in Proc. CVPR*, pp. 1552–1558.

Bosch, A., Zisserman, A., & Muñoz, X. (2006b). Scene Classification Via pLSA. In A. Leonardis, H. Bischof & A. Pinz (Eds.), *Computer Vision – ECCV 2006* (Vol. 3954, pp. 517-530): Springer Berlin Heidelberg.

C

Cho J. S. and Choi J., (2005). Contour-based partial object recognition using symmetry in image databases. In SAC’05: *Proceedings of the 2005 ACM symposium on Applied computing*, pages 1190–1194, New York, NY, USA,ACM Press.

D

Datta, R., Li, J., & Wang, J. Z. (2005). Content-based image retrieval: approaches and trends of the new age. *Paper presented at the Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval*, Hilton, Singapore.

De Vries A. P. and Westerveld T., (2004). A comparison of continuous vs. discrete image models for probabilistic image and video retrieval. *In International Conference on Image Processing*, pp. 2387-2390, Singapore

Deselaers T., Keysers D., and Ney H., (2008). Features for image retrieval: An experimental comparison. *Information Retrieval*, 11(2):77-107

Deselaers, T., Hegerath, A., Keysers, D., & Ney, H. (2006). Sparse Patch-Histograms for Object Classification in Cluttered Images. In K. Franke, K.-R. Müller, B. Nickolay & R. Schäfer (Eds.), *Pattern Recognition* (Vol. 4174, pp. 202-211): Springer Berlin Heidelberg.

Deselaers, T., & Ney, H. (2008). Deformations, patches, and discriminative models for automatic annotation of medical radiographs. *Pattern Recognition Letters, 29*(15), 2003-2010. doi: http://dx.doi.org/10.1016/j.patrec.2008.03.013

Dimitrovski, I., Kocev, D., Loskovska, S., & Džeroski, S. (2011). Hierarchical annotation of medical images. *Pattern Recognition, 44*(10–11), 2436-2449. doi: http://dx.doi.org/10.1016/j.patcog.2011.03.026

Dorko G. (2006). Selection of Discriminative Regions and Local Descriptors for Generic Object Class Recognition. PhD thesis, Institut National Polytechnique de Grenoble.

F

Fei-Fei L. and Perona P. (2005). A Bayesian hierarchical model for learning natural scene categories. *In IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pp.524-531, San Diego, CA, USA,

Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., Yanker, P. (1995). Query by Image and Video Content: The QBIC System. *Journal of Computer*, *28*(9), 23-32.

G

Garcia Seco de Herrera, A., Markonis, D., Eggel, I., Muller, H. (2012). The medGIFT group in ImageCLEFmed 2012. *In: Working Notes of CLEF 2012.*

Glatard, T., Montagnat, J., & Magnin, I. E. (2004). Texture based medical image indexing and retrieval: application to cardiac imaging. *Paper presented at the Proceedings of the 6th ACM SIGMM international workshop on Multimedia information retrieval*, New York, NY, USA.

Güld, M. O., Michael, K., Daniel, K., Henning, S., Berthold , B. W., Jörg, B., & Thomas M.L. (2002). Quality of DICOM header Information for Image Categorization. *Paper presented at the Intl. Symposium on Medical Imaging.*

GuangJian, T., Hong, F., & Feng, D. D. (2008, 30-31 May 2008). Automatic medical image categorization and annotation using LBP and MPEG-7 edge histograms. *Paper presented at International Conference on the Information Technology and Applications in Biomedicine, 2008. ITAB 2008..*

H

Haralick, R. M., Shanmugam, K., & Dinstein, I. h. (1973). Textural Features for Image Classification. *Systems, Man and Cybernetics, IEEE Transactions on, SMC-3*(6), 610-621. doi: 10.1109/tsmc.1973.4309314

Hengen, H., Spoor, S. L., & Pandit, M. C. (2002). Analysis of blood and bone marrow smears using digital image processing techniques. 624-635. doi: 10.1117/12.467205

Hofmann, T. (2001). Unsupervised Learning by Probabilistic Latent Semantic Analysis. *Mach. Learn., 42*(1-2), 177-196. doi: 10.1023/a:1007617005950

Huang-Chia, S., Che-Yen, C., Chung-Lin, H., & Chi-Hua, L. (2012). Gender classification using bayesian classifier with local binary patch features. *Paper presented at IEEE 4th International Conference on the Nonlinear Science and Complexity* (NSC),

I

Inthajak, K., Duanggate, C., Uyyanonvara, B., Makhanov, S. S., & Barman, S. (2011). Medical image blob detection with feature stability and KNN classification. *Paper presented at the Computer Science and Software Engineering (JCSSE), 2011 Eighth International Joint Conference on.*

Iqbal Q and Aggarwal J., (2002). CIRES: A system for content-based retrieval in digital image libraries. *In International Conference on Control, Automation, Robotics and Vision*, pp 205-210, Singapore

J

Jeanne, V., Unay, D., & Jacquet, V. (2009). Automatic detection of body parts in x-ray images. *Paper presented at the Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on.*

Ji-Hyeon, L., Deok-Yeon, K., Byoung Chul, K., & Jae-Yeal, N. (2011). Keyword Annotation of Medical Image with Random Forest Classifier and Confidence Assigning. *Paper presented at the Computer Graphics, Imaging and Visualization (CGIV), 2011 Eighth International Conference on.*

Jie, F., Jiao, L. C., Xiangrong, Z., & Dongdong, Y. (2011). Bag-of-Visual-Words Based on Clonal Selection Algorithm for SAR Image Classification. *Geoscience and Remote Sensing Letters, IEEE, 8*(4), 691-695.

Jing, H., Kumar, S. R., Mitra, M., Wei-Jing, Z., & Zabih, R. (1997). Image indexing using color correlograms. *Paper presented at the Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on.*

Jingyan, W., Yongping, L., Ying, Z., Honglan, X., & Chao, W. (2011a). Bag-of-Features Based Classification of Breast Parenchymal Tissue in the Mammogram via Jointly Selecting and Weighting Visual Words. *Paper presented at the Image and Graphics (ICIG), 2011 Sixth International Conference on.*

Jingyan, W., Yongping, L., Ying, Z., Honglan, X., & Chao, W. (2011b). Boosted Learning of Visual Word Weighting Factors for Bag-of-Features Based Medical Image Retrieval. *Paper presented at the Image and Graphics (ICIG), 2011 Sixth International Conference on.*

K

Kesorn, K., & Poslad, S. (2012). An Enhanced Bag-of-Visual Word Vector Space Model to Represent Visual Content in Athletics Images. *Multimedia, IEEE Transactions on, 14*(1), 211-222.

Keysers, D., Deselaers, T., Gollan, C., & Ney, H. (2007). Deformation Models for Image Recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, 29*(8), 1422-1435. doi: 10.1109/tpami.2007.1153

Khaliq, W., Blakeley, C. J., Maheshwaran, S., Hashemi, K., & Redman, P. (2008). Comparison of a PACS workstation with laser hard copies for detecting scaphoid fractures in the emergency department Vol. 23. (pp. 100-103).

Ko, B., Kim, S., & Nam, J.-Y. (2011). X-ray Image Classification Using Random Forests with Local Wavelet-Based CS-Local Binary Patterns. *Journal of Digital Imaging, 24*(6), 1141-1151. doi: 10.1007/s10278-011-9380-3

Kouskouridas, R., Gasteratos, A., & Badekas, E. (2012). Evaluation of two-part algorithms for objects' depth estimation. *Computer Vision, IET, 6*(1), 70-78.

L

Lazebnik, S., Schmid, C., & Ponce, J. (2006, 2006). Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. *Paper presented at the Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on.*

Lei, W., Hoi, S. C. H., & Nenghai, Y. (2010). Semantics-Preserving Bag-of-Words Models and Applications. *Image Processing, IEEE Transactions on, 19*(7), 1908-1920.

Lehmann T. M., Guld M. O., Deselaers T. , Keysers D. , Schubert H. , Spitzer K., Ney H. , and Wein B.,(2005). Automatic categorization of medical images for content-based retrieval and data mining. *Computerized Medical Imaging and Graphics*, 29(2):143-155

Lehmann TM, Schubert H, Keysers D, Kohnen M, Wein BB,(2003). The IRMA code for unique classification of medical images. *In: Proceedings SPIE* 5033, pp. 440–451

Li, D.-x., Fan, J.-l., Wang, D.-w., & Liu, Y. (2012). Latent topic based multi-instance learning method for localized content-based image retrieval. *Computers & Mathematics with Applications, 64(4), 500-510.* doi: http://dx.doi.org/10.1016/j.camwa.2011.12.030

Li J., Wu W., Wang T., and Zhang Y.,(2008). One step beyond histograms: Image representation using markov stationary features. *In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, volume 0, pages 1–8, Los Alamitos, CA, USA. IEEE Computer Society.

Li, Z., Shi, Z., Liu, X., & Shi, Z. (2011). Modeling continuous visual features for semantic image annotation and retrieval. *Pattern Recognition Letters, 32*(3), 516-523.

Lienhart R., Romberg S., and Hörster E, (2009). Multilayer pLSA for multimodal image retrieval, *in Proc. of the ACM International Conference on Image and Video Retrieval (CIVR)*, Island of Santorini, Greece, pp. 1-8.

Lindeberg, T. (1998). Feature Detection with Automatic Scale Selection. *Int. J. Comput. Vision, 30*(2), 79-116. doi: 10.1023/a:1008045108935

Lowe D. (2004). Distinctive image features from scale invariant key points. *International Journal of Computer Vision, 60*(2), 91-110.

Lowe, D. G. (1999, 1999). Object recognition from local scale-invariant features. *Paper presented at the Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on.*

M

Ma, W. Y., & Manjunath, B. S. (1997). NeTra: a toolbox for navigating large image databases. *Paper presented at the Image Processing, 1997. Proceedings., International Conference on.*

Markonis, D., Seco de Herrera, A. G., Eggel, I., & Müller, H. (2012). Multi-scale visual words for hierarchical medical image categorisation. 83190F-83190F. doi: 10.1117/12.911550

Mehrotra, R., & Gary, J. E. (1995). Similar-shape retrieval in shape data management. *Computer, 28*(9), 57-62. doi: 10.1109/2.410154

Mohan A., Papageorgiou C., and Poggio T., (2001). Example-based object detection in images by components, *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 23, no. 4, pp. 349–361,

Monay, F., Quelhas, P., Gatica-Perez, D., & Odobez, J.-M. (2006). Constructing Visual Models with a Latent Space Approach. In C. Saunders, M. Grobelnik, S. Gunn & J. Shawe-Taylor (Eds.), *Subspace, Latent Structure and Feature Selection* (Vol. 3940, pp. 115-126): Springer Berlin Heidelberg.

Morales, A., Ferrer, M. A., & Kumar, A. (2011). Towards contactless palmprint authentication. *Computer Vision, IET, 5*(6), 407-416.

Morevec, H. P. (1977). Towards automatic visual obstacle avoidance. *Paper presented at the Proceedings of the 5th international joint conference on Artificial intelligence* - Volume 2, Cambridge, USA.

Mueen, A., Zainuddin, R., & Baba, M. S. (2008). Automatic Multilevel Medical Image Annotation and Retrieval. *Journal of Digital Imaging, 21*(3), 290-295. doi: 10.1007/s10278-007-9070-3

Mueen, A., Zainuddin, R., & Sapiyan Baba, M. (2010). MIARS: A Medical Image Retrieval System. *Journal of Medical Systems, 34*(5), 859-864. doi: 10.1007/s10916-009-9300-y

Muller, H., Geissbuhler, A., Marty, J., Lovis, C., & Ruch, P. (2005). The Use of MedGIFT and EasyIR for ImageCLEF 2005. *In: CLEF 2005 Proceedings. Lecture Notes in Computer Science (LNCS),, 4022*, 724-732.

Müller, H., Michoux, N., Bandon, D., & Geissbuhler, A. (2004). A review of content-based image retrieval systems in medical applications—clinical benefits and future directions. *International Journal of Medical Informatics, 73*(1), 1-23. doi: http://dx.doi.org/10.1016/j.ijmedinf.2003.11.024

Müller H, Deselaers T, Kim E, Kalpathy-Cramer J, Deserno TM, Clough P, Hersh W ,(2007). Overview of the ImageCLEFmed 2007 medical retrieval and annotation tasks. *In: Working Notes of the 2007 CLEF Workshop,*

N

Nishibori, M., Norimichi, T., & Yoichi, M. (2004). Why multispectral imaging in medicine? *Journal of Imaging Science and Technology, 48*, 125-129.

Nowak E, Jurie F., and Triggs B., (2006). Sampling strategies for bag-offeatures image classification. *in Proc. ECCV,* pp. 490–503.

O

Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, 24*(7), 971-987. doi: 10.1109/tpami.2002.1017623

P

Pass, G., Zabih, R., & Miller, J. (1996). Comparing images using color coherence vectors. *Paper presented at the Proceedings of the fourth ACM international conference on Multimedia*, Boston, Massachusetts, United States.

Pawlicki R., Kokai I, Finger J, Smith R., and Vetter T., (2007). Navigating in a shape space of registered models. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1552–1559,

Pourghasem H., Ghasemian H. (2008).Content-based medical image classification using a new hierarchical merging scheme, Computerized Medical Imaging and Graphics, 32 , 651–661

Pulla, C., & Jawahar, C. V. (2010). Multi modal semantic indexing for image retrieval. *Paper presented at the Proceedings of the ACM International Conference on Image and Video Retrieval*, Xi'an, China.

Q

Quelhas, P., Monay, F., Odobez, J.-M., Gatica-Perez, D., Tuytelaars, T., & Gool, L. V. (2005). Modeling Scenes with Local Descriptors and Latent Aspects. *Paper presented at the Proceedings of the Tenth IEEE International Conference on Computer Vision* (ICCV'05) Volume 1 - Volume 01.

R

Rahman, M., Tongyuan, W., & Desai, B. C. (2004, 1-3 Sept. 2004). Medical image retrieval and registration: towards computer assisted diagnostic approach. *Paper* *presented at the Design of Reliable Communication Networks*, 2003. (DRCN 2003). Proceedings. Fourth International Workshop on.

Rahman M, Desai BC, Bhattacharya P. (2007). Medical Image Retrieval with probabilistic multi-class support vector machine classifiers and adaptive similarity fusion. Computerized Medical Imaging and Graphics, 95-108

Rahman, M. M., Antani, S. K., & Thoma, G. R. (2011a). A query expansion framework in image retrieval domain based on local and global analysis. *Information Processing & Management, 47*(5), 676-691. doi: http://dx.doi.org/10.1016/j.ipm.2010.12.001

Rahman, M., Antani, S., & Thoma, G. (2011b). A Learning-Based Similarity Fusion and Filtering Approach for Biomedical Image Retrieval Using SVM Classification and Relevance Feedback. *Information Technology in Biomedicine, IEEE Transactions on*, 15(4), 640-646. doi: 10.1109/titb.2011.2151258

Romberg, S., Lienhart, R., & Hörster, E. (2012). Multimodal Image Retrieval. *International Journal of Multimedia Information Retrieval*, 1(1), 31-44. doi: 10.1007/s13735-012-0006-4.

Rui, Y., Huang, T. S., & Chang, S.-F. (1999). Image Retrieval: Current Techniques, Promising Directions, and Open Issues. *Journal of Visual Communication and Image Representation, 10*(1), 39-62. doi: http://dx.doi.org/10.1006/jvci.1999.0413

S

Seong-Hoon, K., Ji-Hyun, L., ByoungChul, K., & Jae-Yeal, N. (2010). X-ray image classification using Random Forests with Local Binary Patterns. *Paper presented at the Machine Learning and Cybernetics* (ICMLC), 2010 International Conference on.

Setia L, Teynor A, Halawani A, Burkhardt H (2008). Grayscale medical image annotation using local relational features. *Pattern Recognition Letters* 29(15):2039–2045

Shen, J., Shepherd, J., & Ngu, A. H. H. (2005). Semantic-Sensitive Classification for Large Image Libraries. *Paper presented at the Multimedia Modelling Conference, 2005. MMM 2005. Proceedings of the 11th International.*

Smith, G. E., Setlur, P., & Mobasseri, B. G. (2011). Multiple hypothesis tests For robust radar target recognition. *Paper presented at the Radar Conference (RADAR),* 2011 IEEE.

Smith, G. E., & Mobasseri, B. G. (2012). Robust Through-the-Wall Radar Image Classification Using a Target-Model Alignment Procedure. *Image Processing, IEEE Transactions on,* 21(2), 754-767. doi: 10.1109/tip.2011.2166967

Sohail, A. S. M., Rahman, M. M., Bhattacharya, P., Krishnamurthy, S., & Mudur, S. P. (2010, 14-17 April 2010). Retrieval and classification of ultrasound images of ovarian cysts combining texture features and histogram moments. *Paper presented at the Biomedical Imaging: From Nano to Macro, 2010 IEEE International Symposium on.*

Squire D. M. , Muller W. , Muller H. and Raki J. (1999). Content-based query of image database: inspirations from text retrieval: Inverted files, frequency-based weights and relevance feedback. *In Scandinavian Conference on Image Analysis*, pp143-149, Kangerlussuaq, Greenland.

Sui, L., Zhang, J., Zhuo, L., & Yang, Y. C. (2012). Research on pornographic images recognition method based on visual words in a compressed domain. *Image Processing, IET, 6*(1), 87-93.

T

Tamura H. and Yokoya N., (1984). Image database system: A survey. *Pattern Recognition*, 17(1):19-43

Teng, L., Tao, M., In-So, K., & Xian-Sheng, H. (2011). Contextual Bag-of-Words for Visual Categorization. *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, 21*(4), 381-392.

Tommasi T, Orabona F, Caputo B (2007). CLEF2007 Image Annotation Task: an SVM–based Cue Integration Approach. *In: Working Notes of CLEF 2007*, Budapest, Hungary

Tommasi, T., Orabona, F., & Caputo, B. (2008). Discriminative cue integration for medical image annotation. *Pattern Recognition Letters, 29*(15), 1996-2002. doi: http://dx.doi.org/10.1016/j.patrec.2008.03.009

Tirilly P, Claveau V, and Gros P,(2008). Language modeling for bag of visual words image categorization, In Proc. *Int. Conf. Content-Based Image and Video Retrieval*, 2008, pp. 249–258.

U

Unay, D., Soldea, O., Ekin, A., Cetin, M., & Ercil, A. (2010). Automatic Annotation of X-Ray Images: A Study on Attribute Selection. In B. Caputo, H. Müller, T. Syeda-Mahmood, J. Duncan, F. Wang & J. Kalpathy-Cramer (Eds.), Medical Content-Based Retrieval for Clinical Decision Support (Vol. 5853, pp. 97-109): Springer Berlin Heidelberg.

V

Viitaniemi, V., & Laaksonen, J. (2006). Techniques for Still Image Scene Classification and Object Detection. In S. Kollias, A. Stafylopatis, W. Duch & E. Oja (Eds.), *Artificial Neural Networks – ICANN 2006* (Vol. 4132, pp. 35-44): Springer Berlin Heidelberg.

W

Wei, Y., Zhentai, L., Mei, Y., Meiyan, H., Qianjin, F., & Wufan, C. (2012). Content-Based Retrieval of Focal Liver Lesions Using Bag of Visual Words Representations of Single and Multiphase Contrast Enhanced CT Images. *Journal of Digital Imaging, 25*, 708-719.

Y

Yang, F., Lu, H., Zhang, W., & Yang, G. (2012). Visual tracking via bag of features. *Image Processing, IET, 6*(2), 115-128.

Ye, M., & Androutsos, D. (2009). Robust Affine Invariant Region-Based Shape Descriptors: The ICA Zernike Moment Shape Descriptor and the Whitening Zernike Moment Shape Descriptor. *Signal Processing Letters, IEEE, 16*(10), 877-880. doi: 10.1109/lsp.2009.2026119

Yimo, T., Zhigang, P., Krishnan, A., & Zhou, X. S. (2011). Robust Learning-Based Parsing and Annotation of Medical Radiographs. *Medical Imaging, IEEE Transactions on, 30*(2), 338-350.

Z

Zare, M. R., Mueen, A., Woo Chaw, S., & Awedh, M. H. (2011). Combined Feature Extraction on Medical X-ray Images. *Paper presented at the Computational Intelligence, Communication Systems and Networks* (CICSyN), 2011 Third International Conference on.

Zare, M.R., Woo, C.S., Mueen A., (2013a). Automatic Classification of medical X-ray images, *Malaysian Journal of Computer Science.* Vol. 26(1), pp: 9-22

Zare, M.R., Mueen, A., Woo, C.S., (2013b). Automatic Classification of Medical X-ray Images using Bag of Visual Words. *Computer Vision, IET.* Vol. 7(2), pp:105-114

Zare MR, Mueen A, Woo CS, (2013c), Automatic Medical X-ray Image Classification using Annotation, *Journal of Digital Imaging*, DOI: 10.1007/s10278-013-9637-0 (ISI Index Publication)

Zare, M.R., Mueen, A., Awedh, M. H., Woo, C.S.,(2013d). Automatic Classification of Medical X-Ray Images: A Hybrid Generative-Discriminative Approach. *Image Processing, IET.* DOI: 10.1049/iet-ipr.2013.0049

Zhang, Z., Deriche, R., Faugeras, O., & Luong, Q.-T. (1995). A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artif. Intell., 78*(1-2), 87-119. doi: 10.1016/0004-3702(95)00022-4

Zhi, L.-j., Zhang, S.-m., Zhao, D.-z., Zhao, H., & Lin, S.-k. (2009, 17-19 Oct. 2009). Medical Image Retrieval Using SIFT Feature. *Paper presented at the Image and Signal Processing, 2009. CISP '09. 2nd International Congress on*.

Zhou, W., Li, H., Lu, Y., & Tian, Q. (2012). Principal Visual Word Discovery for Automatic License Plate Detection. *Image Processing, IEEE Transactions on, 21*(9), 4269-4279.

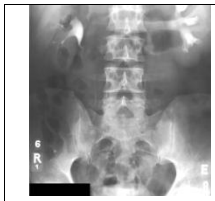
Zhy, C.-m., Gu, G.-c., Liu, H.-b., Shen, J., & Yu, H. (2008, 12-14 Dec. 2008). Segmentation of Ultrasound Image Based on Texture Feature and Graph Cut. *Paper presented at the Computer Science and Software Engineering, 2008 International Conference on.*

Zoller T and Buhmann J. M. (2007). Robust image segmentation using resampling and shape constraints. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(7):1147–1164,

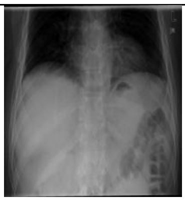
Zou B.J. and Umugwaneza M. P., (2008). Shape-based trademark retrieval using cosine distance method. *In Proceedings of International Conference on Intelligent Systems Design and Applications*, volume 2, pages 498–504, Los Alamitos, CA, USA. IEEE Computer Society.

Appendix A

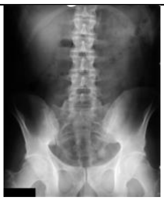
Sample Image from Every Class of ImageCLEF 2007 Medical Database



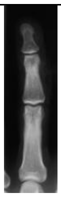
1



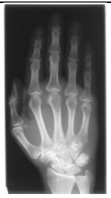
2



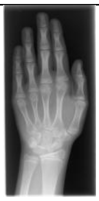
3



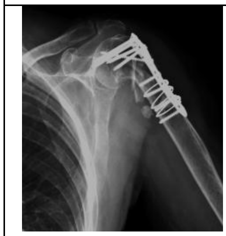
4



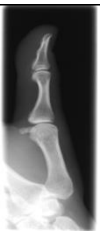
5



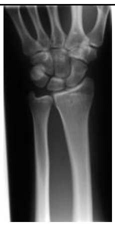
6



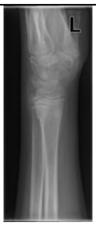
7



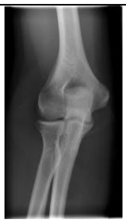
8



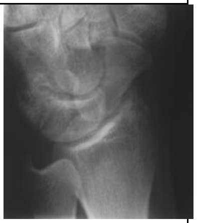
9



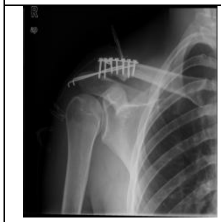
10



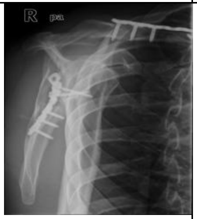
11



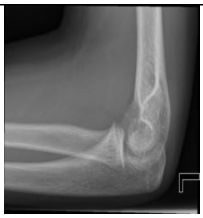
12



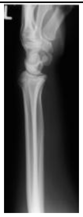
13



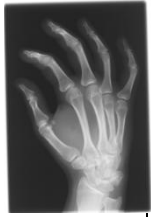
14



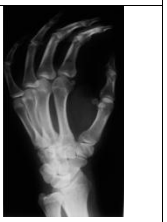
15



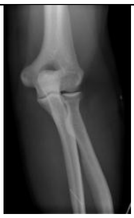
16



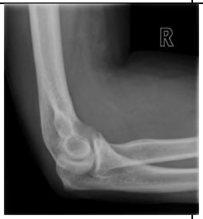
17



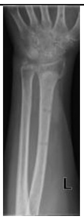
18



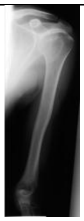
19



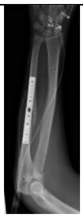
20



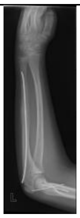
21



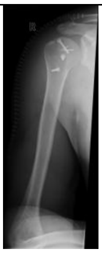
22



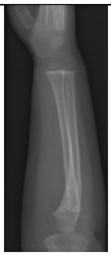
23



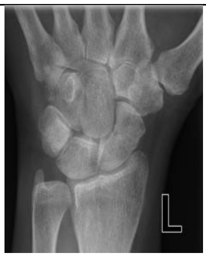
24



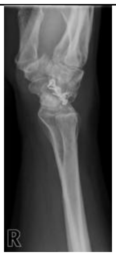
25



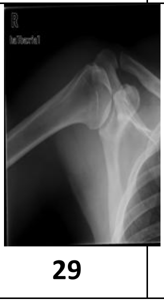
26



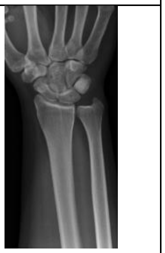
27



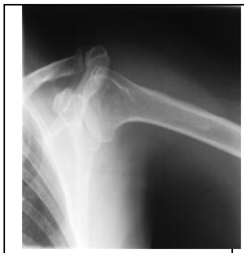
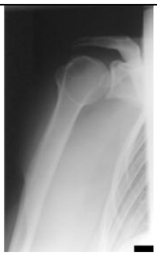
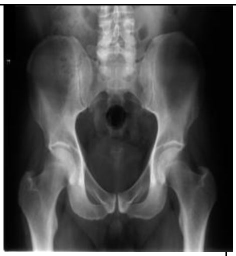
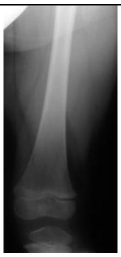
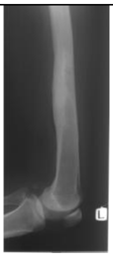
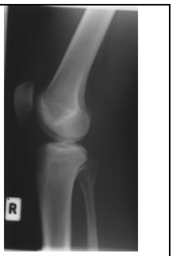
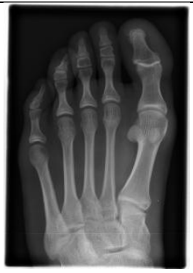
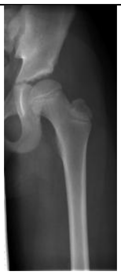
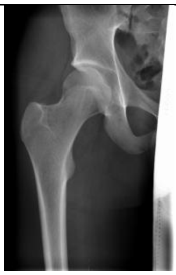
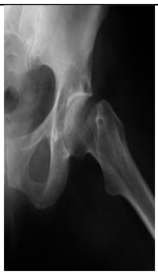
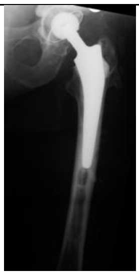
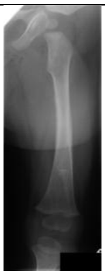
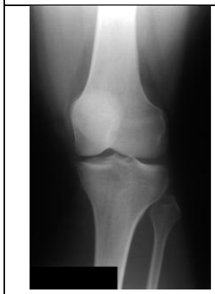
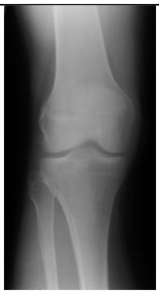
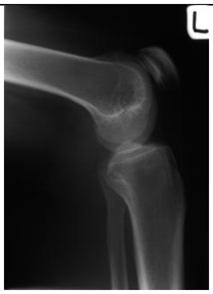
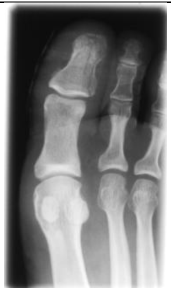
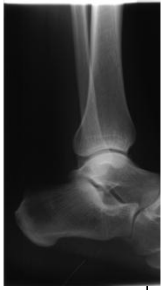
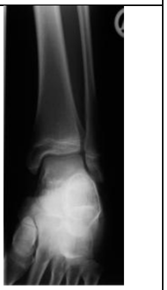
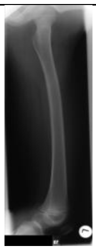
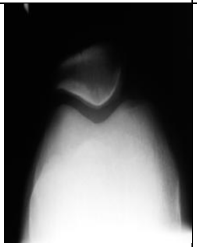
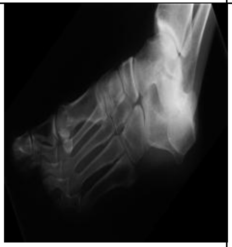
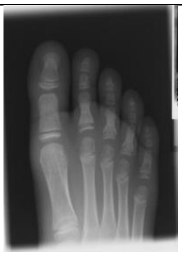
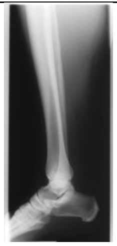
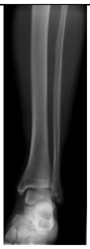
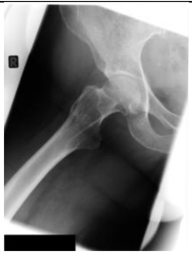
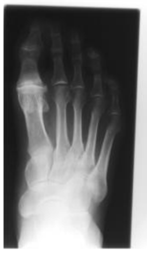
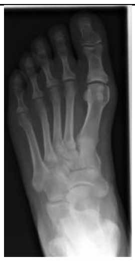
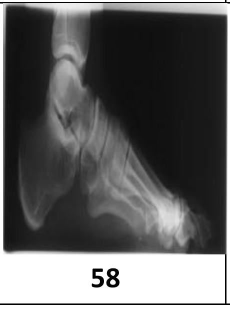
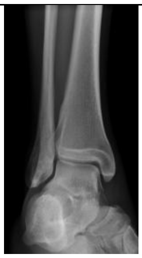
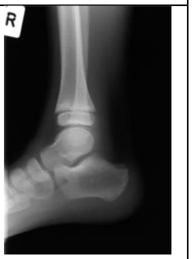
28

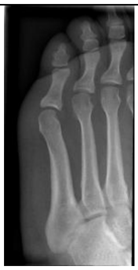
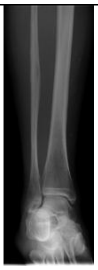
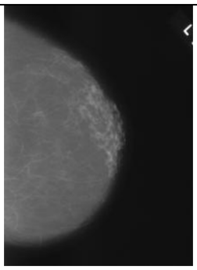
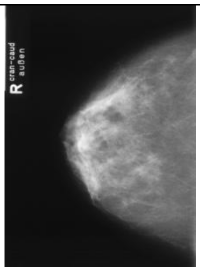
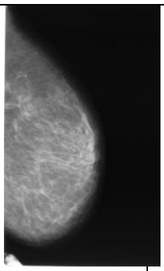
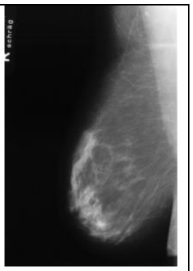
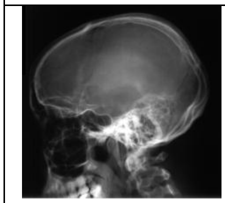
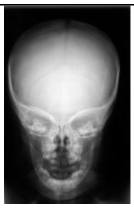
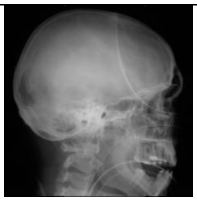
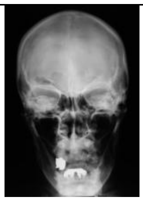
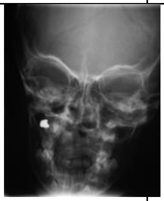
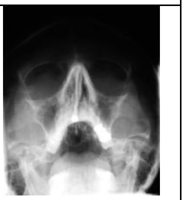
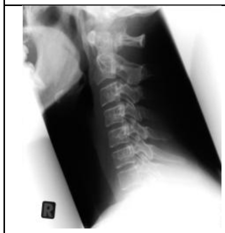
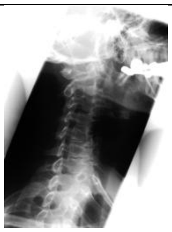
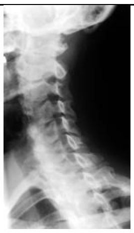
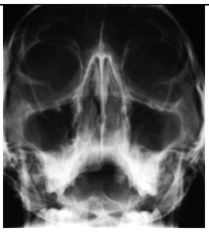
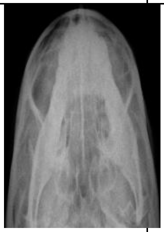
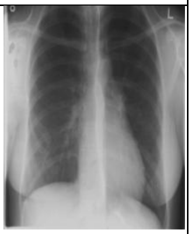
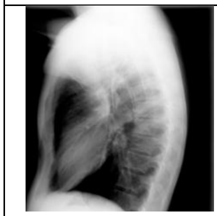
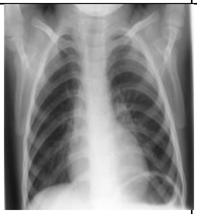
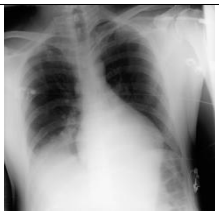
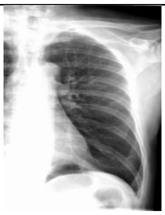
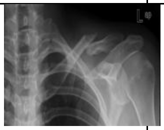
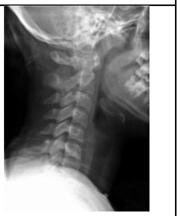
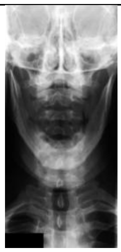
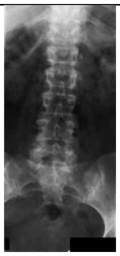
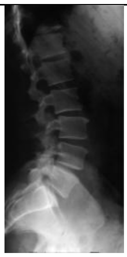
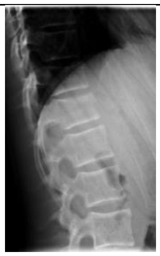
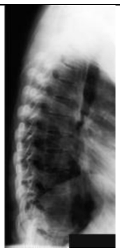
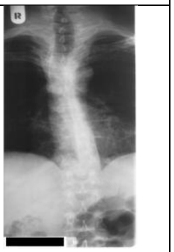
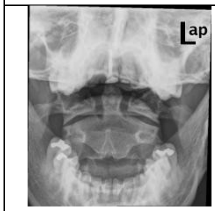
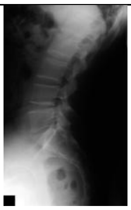
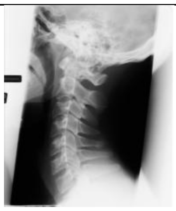
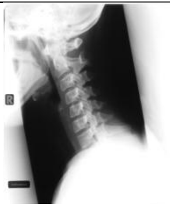
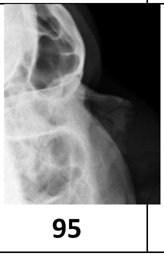
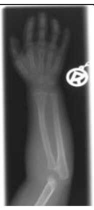


29



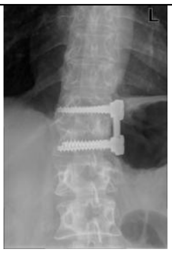
30

 31	 [redacted] 32	 33	 34	 35	 36
 37	 38	 39	 40	 41	 [redacted] 42
 [redacted] 43	 44	 45	 46	 47	 48
 [redacted] 49	 50	 51	 52	 53	 54
 [redacted] 55	 56	 57	 58	 59	 60

 61	 62	 63	 64	 65	 66
 67	 68	 69	 70	 71	 72
 73	 74	 75	 76	 77	 78
 79	 80	 81	 82	 83	 84
 85	 86	 87	 88	 89	 90
 91	 92	 93	 94	 95	 96

97

98


99

100

101

102

103

104

105

106

107

108

109

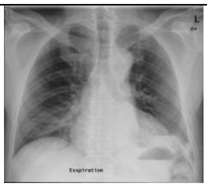
110

111

112

113

114


115

116

Appendix B

Annotated keywords for Every Class

Class 01: Abdomen, IVU,

Class 02: Abdomen, Upper abdomen,

Class 03: Abdomen,

Class 04: Arm, hand, finger, phalanx, interphalangeal joint, metacarpal, PA view

Class 05: Arm, hand, wrist, carpal, Right, PA View

Class 06: Arm, hand, wrist, carpal, Left, PA View

Class 07: Arm, shoulder, humero-scapular joint, Left

Class 08: Arm, hand, finger, phalanx, lateral view

Class 09: Arm, wrist, radio carpal joint, distal radius, distal ulna, Left, AP View

Class 10: Arm, wrist, radio carpal joint, distal radius, distal ulna, Left, Lateral View

Class 11: Arm, elbow joint, Right, AP View

Class 12: Arm, hand, carpal bone, wrist joint, scaphoid, lateral view,

Class 13: Arm, shoulder , acromio-scapular joint, Right, AP View

Class 14: Arm, shoulder, Scapula, Left

Class 15: Arm, Elbow, Lateral view, Left

Class 16: Arm, distal forearm, distal radius, distal ulna, lateral view, left

Class 17: Arm, hand, wrist, Right

Class 18: Arm, hand, wrist, Left

Class 19: Arm, elbow, AP View , Left

Class 20: Arm, elbow joint, lateral view, right

Class 21: Arm, distal forearm, distal radius, distal ulna, AP view, left

Class 22: Arm, upper arm, distal upper arm, humerus, shoulder joint, elbow joint, Left

Class 23: Arm, forearm, radius, ulna, wrist joint, elbow joint, AP view, Left

Class 24: Arm, forearm, radius, ulna, wrist joint, elbow joint, lateral view, Left

Class 25: Arm, upper arm, proximal upper arm, AP View, Right

Class 26: Arm, forearm, radius, ulna, wrist joint, elbow joint, lateral view

Class 27: Arm, hand, carpal bone, wrist joint, scaphoid, PA view, left

Class 28: Arm, distal forearm, wrist joint, lateral view

Class 29: Arm, shoulder, humero scapular joint, Right

Class 30: Arm, wrist, radio carpal joint, distal radius, distal ulna, PA view, Right

Class 31: Arm, shoulder , humero-scapular joint, Left

Class 32: Arm, upper arm, upper humerus, AP View, Right

Class 33: Pelvis, hip joint, sacrum, AP View

Class 34: Leg, upper leg, distal upper leg, distal femur, Left, AP View

Class 35: Leg, upper leg, Distal femur, lateral view

Class 36: Leg, knee, patella, lateral view, right

Class 37: Leg, foot, forefoot, toes, AP View, Left

Class 38: Leg, hip joint, proximal femur, AP View, Left

Class 39: Leg, hip joint, AP View, Right

Class 40: Leg, hip joint, Lateral View, left

Class 41: Leg, hip joint, proximal femur, Left

Class 42: Leg, upper leg, femur, thigh, left , AP View

Class 43: Leg, knee, AP View, left

Class 44: Leg, knee, AP View, Right

Class 45: Leg, knee, patella, lateral view, left

Class 46: Leg, foot, toes, phalanx

Class 47: Leg, ankle joint, lateral view, left

Class 48: Leg, ankle joint, lateral view, right

Class 49: Leg, upper leg, femur, thigh, hip joint, knee joint,

Class 50: Leg, knee, patella

Class 51: Leg, foot, lateral view, right

Class 52: Leg, forefoot, toes, AP View, right

Class 53: Leg, lower leg, distal tibia, lateral view

Class 54: Leg, lower leg, tibia, ankle joint, AP view, left

Class 55: Leg, hip joint, femoral head, femoral neck, right

Class 56: Leg, foot, AP view, right

Class 57: Leg, foot, AP view, left

Class 58: Leg, foot, lateral view

Class 59: Leg, ankle joint, AP View, right

Class 60: Leg, ankle joint, lateral view, right

Class 61: Leg, foot, toes, phalanx, AP View

Class 62: Leg, lower leg, tibia, AP view, right

Class 63: Breast, Left, craniocaudal view

Class 64: Breast, right, craniocaudal view

Class 65: Breast , Left

Class 66: Breast, right

Class 67: Cranium, neuro cranium, skull, lateral view, right

Class 68: Cranium, neuro cranium, skull, occipital area

Class 69: Cranium, neuro cranium, skull, lateral view, left

Class 70: Cranium, skull, frontal sinus, frontal view,

Class 71: Cranium, facial cranium, nose area, paranasal sinus, AP view

Class 72: Cranium, facial cranium, temporo mandibular area, maxillary sinus

Class 73: Spine, cervical spine, lateral view

Class 74: Spine, cervical spine, cervical foramina ,

Class 75: Spine, cervical spine, cervical foramina

Class 76: Cranium, facial cranium, eye area, orbits,

Class 77: Cranium, facial cranium, nose area, mandible

Class 78: Chest, ribs, heart

Class 79: Chest, lateral view

Class 80: Chest, pediatric chest

Class 81: Chest, portable

Class 82: Chest, left

Class 83: Chest, bones, upper ribs,

Class 84: Spine, cervical spine, lateral view

Class 85: Spine, cervical spine, AP view

Class 86: Spine, lumbar spine, AP view

Class 87: Spine, lumbar spine, lateral view

Class 88: Spine, lumbar spine, thoraco-lumbar conjuction, lateral view

Class 89: Spine, thoracic spine, lateral view

Class 90: Spine, thoracic spine, AP view

Class 91: Spine, cervical spine, dens, open mouth view, AP View

Class 92: Spine, lumbar spine, lumbosacral spine, lateral view

Class 93: Spine, cervical spine, neck

Class 94: Spine, cervical spine, lateral view, neck

Class 95: Cranium, facial cranium, nose area, lateral view, left

Class 96: Arm, forearm, radius, ulna, AP view, right

Class 97: Arm, distal forearm, distal radius, distal ulna, wrist joint, lateral view

Class 98: Cranium, facial cranium, nose area, lateral view, right

Class 99: Spine, lumbar spine, thoraco-lumbar conjuction, AP view

Class 100: Spine, lumbar spine, lower lumbar spine, lateral view

Class 101: Chest, bones, upper ribs

Class 102: Leg, foot, calcaneus, heel

Class 103: Chest, right

Class 104: Cranium, facial cranium, nose area, skull, AP view, paranasal sinus

Class 105: Abdomen, barium study

Class 106: Arm, shoulder, humerus head, scapuls, ribs, left

Class 107: Leg, foot, tarsal bone, intertarsal joint, AP view

Class 108: Cranium, facial cranium, mandible

Class 109: Leg, lower leg, knee joint, ankle joint, AP view

Class 110: Leg, distal forearm, distal radius, distal ulna, wrist joint, AP view, right

Class 111: Leg, foot, forefote, toes

Class 112: Leg, lower leg, tibia, knee joint, ankle joint, AP view

Class 113: Leg, foot, Left

Class 114: Leg, foot, right

Class 115: Chest, lung, rib cage, heart

Class 116: Abdomen, decubitus