ABSTRACT

The increase of multimedia data in medical field has led to the challenging problem of developing techniques that can provide efficient and accurate search and navigation through the large digital archives. Multi-modality of medical images such as x-ray, CT scan and MRI constitutes an important source of anatomical and functional information to provide valuable teaching and research, effective training and diagnosis of diseases.

The approach of conventional text-based queries and exact matching with database is becoming obsolete. The mismatch of medical term between user query and document has become an issue. This drawback of conventional text-based retrieval motivates researchers towards more effective text-based retrieval and visual content-based image retrieval (CBIR) in medical field which has been an active research area in computer vision for the past few years. However CBIR solely has not yet succeeded in bridging the semantic gap between human concepts and low-level visual features. Information fusion support for human or automated analysis and processing relies on the hypothesis that the combination of multiple information sources allows for better results in information retrieval performance.

This research is focused on information fusion of text and visual content analysis in medical information retrieval system. Initially the design, development and evaluation process are executed separately for text and content-based features. Medical hierarchical conceptual model is applied in both text and content-based frameworks which emphasize on modality, anatomy and pathology concepts. Multi-modality Medical Information Retrieval System (M3IRS) text-based framework consists of four main components which

are document pre-processor, query processor, retrieval process and ranking strategies. Two ranking models are introduced namely Comprehensive and MedHieCon ranking models. Multi-modality Medical Image Classification System (M3ICS) is the content-based framework which is based on extracting visual features of texture, shape and color in global and local descriptors and applying semantic classification using MedHieCon model. Supervised learning technique is heavily used in visual classification to evaluate the performance of the framework. The final stage is the information fusion with the combination of text and content-based information sources. Hierarchical processing in late fusion technique is applied where the output from text-based processing will be the input for content-based system.

Dataset from ImageCLEF 2010 medical task is used which includes 77,500 multi-modality medical images and documents. The performance of text-based, content-based and information fusion of text and content-based processing are evaluated. The effectiveness of text-based performance is better than content-based. However content-based approach is complement to text-based retrieval in order to increase the relevant documents into higher position ranking in medical domain information retrieval. Retrieval based on information fusion achieves the best performance by improving the effectiveness of retrieving relevant documents in medical domain.

ABSTRAK

Peningkatan jumlah data multimedia dalam bidang perubatan telah membawa kepada masalah dalam membangunkan teknik dapatan carian yang cekap dan tepat melalui arkib digital yang bersaiz besar. Imej perubatan yang dihasilkan oleh mesin imbasan yang berbeza memaparkan ciri-ciri yang berbeza. Dengan pelbagai kaedah yang sedia ada dalam imej perubatan seperti X-ray, CT dan MRI adalah sukar bagi penyelidik untuk mewujudkan sistem dapatan yang cekap dengan hanya menggunakan carian berdasarkan teks yang berasaskan manual. Sistem dapatan yang berdasarkan teks sudah tidak boleh digunapakai. Ketidakpadanan istilah perubatan antara pertanyaan pengguna dan dokumen telah menjadi satu isu. Kelemahan capaian berasaskan teks dalam bidang perubatan ini mendorong penyelidik ke arah sistem dapatan semula berasaskan visual, yang juga telah menjadi penyelidikan yang aktif dalam bidang komputer untuk beberapa tahun yang lalu. Walau bagaimanapun sistem dapatan semula berasaskan visual semata-mata tidak lagi berjaya merapatkan jurang semantik antara konsep pandangan manusia dan ciri-ciri visual pada imej. Gabungan pelbagai sumber maklumat seperti teks dan visual membolehkan keputusan yang lebih baik diperolehi dalam sistem dapatan maklumat.

Penyelidikan ini tertumpu kepada gabungan maklumat berdasarkan teks dan kandungan dalam imej sistem dapatan maklumat perubatan. Pada mulanya reka bentuk, pembangunan dan proses penilaian akan dilakukan berasingan untuk teks dan kandungan dalam imej. Model *Medical Hierarchical Conceptual* (MedHieCon) diaplikasikan pada rangka kerja teks dan kandungan dalam imej berdasarkan konsep modaliti, anatomi dan patologi. *Multi-modality Medical Information Retrieval System* (M3IRS) berdasarkan rangka kerja teks berasaskan empat komponen utama iaitu dokumen pra proses, pemproses soalan, strategi

dapatan dan strategi pemeringkatan. Dua model kedudukan diperkenalkan iaitu model kedudukan Comprehensive and MedHieCon. *Multi-modality Medical Image Classification System* (M3ICS) ialah rangka kerja berdasarkan kandungan dalam imej berdasarkan ciri-ciri visual tekstur, bentuk dan warna dalam pemerihal global dan tempatan dan mengaplikasikan semantik berasaskan pengelasan menggunakan model MedHieCon. Teknik pembelajaran terselia digunakan dalam pengelasan visual bagi menilai prestasi rangka kerja. Akhirnya adalah gabungan maklumat lain gabungan sumber maklumat teks dan kandungan dalam imej. Pemprosesan berhierarki dalam teknik gabungan lewat diaplikasikan iaitu output dari berasaskan teks adalah diinput untuk sistem berdasarkan kandungan dalam imej..

Dataset daripada ImageCLEF 2010 bidang perubatan telah digunakan dan ia merangkumi 77,500 imej perubatan pelbagai modaliti dan dokumen. Prestasi sistem berasaskan teks, visual dan dan gabungan teks dan visual akan dinilai berasingan. Keberkesanan prestasi berasaskan teks lebih baik daripada berdasarkan kandungan dalam imej. Bagaimanapun pendekatan berdasarkan kandungan dalam imej ialah pelengkap untuk dapatan berasaskan teks supaya dapat meningkatkan posisi dokumen berkaitan di tempat yang lebih tinggi. Oleh itu gabungan antara teknik berasaskan teks dan kandungan dalam imej dapat meningkatkan keberkesanan di dalam system dapatan maklumat untuk domain perubatan.

ACKNOWLEGMENT

In the name of Allah, the Most Gracious and the Most Merciful

Alhamdulillah, all praises to Allah for the strengths and His blessing in completing this thesis. My first debt of gratitude must go to my advisor, Professor Dr Roziati Zainuddin. She patiently provided the vision, encouragement and advice necessary for me to proceed through the doctoral program and complete my dissertation. Not forgotten, my appreciation to my co-supervisor, Professor Dr Sabirin Mohamed for her support and knowledge regarding this topic. My acknowledgement also goes to all the lecturers, technicians and office staffs of Faculty of Computer Science and Information Technology for their co-operations.

Sincere thanks to all my friends especially Nurfadhlina, Noris, Mas Rina, Aziah Evi, Niza, Shaliza, Mumtaz and others for her unflagging encouragement, kindness and moral support during my study. Thanks for the friendship and memories.

Last but not least, my deepest gratitude goes to my beloved husband; Mohd Faizal Aziz, my parents; Mr. Hj Madzin Majid and Mrs. Hjh Halimah Ahmad and also to my siblings for their endless love, prayers and encouragement. Also not forgetting my dearly loved, Muhamad Rayyan, Muhamad Rizq Hadif and Muhamad Izz Aydin for bring love and joy to my life. To those who indirectly contributed in this research, your kindness means a lot to me. Thank you very much.

Hizmawati Madzin

2013

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

CT	-Computer Tomography
MRI	-Magnetic Resonance Image
US	-Ultrasound
NM	-Nuclear Medicine
CBIR	-Content-Based Image Retrieval
IR	-Information Retrieval
MIR	-Multimedia Information Retrieval
PACS	-Picture Archiving and Communication Systems
IRMA	- Image Retrieval in Medical Applications
SPIRS	-Spine Pathology and Image Retrieval System
MIMS	- Medical Image Management System
DICOM	- Digital Imaging and Communications in Medicine
MedHieCon	- Medical Hierarchical Conceptual
eof	- end-of-file
MAP	- Mean Average Precision
M3IRS	-Multi-Modality Medical Information Retrieval System
M3ICS	- Multi-Modality Medical Image Classification System
IFM3IRS	-Information Fusion of Multi-Modality Medical Information Retrieval System
k-NN	-k-Nearest Neighborhood
SVM	- Support Vector Machine
SNR	-Signal-to-noise ratio
PX	-optical imaging
GX	-graphic imaging
CNR	- contrast-to-noise ratio
PET	- Positron emission tomography
RSV	- retrieval status value
VSM	- Vector space model
IDF	- Inverse Document Frequency
SC	- simple coefficient
PRP	- Probability Ranking Principle
tf	- term frequency
MeSH	- Medical Subject Heading
NLM	- National Library of Medicine
UMLS	- Unified Medical Language System
CLEF	-Cross Language Evaluation Forum
MedHieCon	- Medical Hierarchical Conceptual
RGB	- red, green, blue
HSV	- hue, saturation, value
HLS	- hue, lightness, saturation

CLD	- Color Layout Descriptor
EHD	- Edge Histogram Descriptor
JKEC	- Joint Kernel Equal Contribution
DCT	- Discrete Cosine Transform
RSNA	- Radiological Society of North America
DTD	- Document Type Definition
XTE	- XML Tag-based Extraction
MR	- medical documents
EQ	- expanded query
RMR	- retrieve relevant medical documents
meq	- matched expanded query
MEQHit	- total number of medical terms that matched
GLCM	- Gray-level co-occurrence matrix
SMO	- sequential minimization algorithm
ROC	- receiver operating characteristics
AUC	- area under ROC curve
MMD	- MeSH-Medical Descriptor
MST	- MeSH-Synonym Terms
MC	- Medical Conceptual
RBF	- Radial Basis Function

1.0 Introduction

1.1 Research Inspiration

Medical data are ubiquitous with the increasing amount of medical data captured and recorded. The on-going development of medical instrumentation and techniques has created an enormous growth in the quantity of data including large amount of medical images. In Geneva University Hospital alone the number of images produced by Radiology department has increased up to 70,000 images per day in 2007, and over 117,000 images per day in 2009, (Brant, 2013; Yu Cao et al., 2011; Müller, 2003). Automated information retrieval systems are needed to reduce information overload.

The technology of medical image production has been rapidly changed over the past few years. Medical image was produced from film-based record initially to electronic record and current technology is the integration of multimedia resources (Haux, 2006). Modern computer technology has created the possibility of producing several new imaging modalities that use different radiant energy techniques to elucidate properties of body tissues. The ability to extract significant and accurate information from conventional or tomography radiographic images such as computed tomography (CT), magnet resonance image (MRI), ultrasound (US) and nuclear medicine (NM) image has developed since the discovery of x-rays. Various medical imaging techniques are used heavily to provide spatially resolved medical information (Huang & Davis, 2011). Multi-modality of medical images constitutes an important source of anatomical and functional information which can be very useful.

Content-Based Image Retrieval (CBIR) is based on computational strategies in searching for relevant images using visual content analysis. Activities related to the usage of CBIR in medical application include education, research, diagnosis, annotation and classification of medical images. However medical information is not only based on medical images. Biomedical information can be presented in variety of forms such as text, illustrations and images in journal articles, documents and other collections such as patient cases in electronic health records (Rahman et al., 2012). When there is some text associated to images, the retrieval can be based on multimodal analysis. Having such bulk volume of image data with varying image modalities, it is then important to develop a suitable medical information retrieval system. The capability of multi-modality medical images can be extended to provide valuable teaching and research, effective training and diagnosis of diseases and enhanced image interpretation support, by developing techniques supporting the automated archiving and the retrieval of images by content. The fundamental problem is how to enable or improve the response of medical retrieval systems with multimodal information sources of using textual information associated with various modalities of medical images and visual features representing the visual content of the images. In line with the present era of technology advancement, research on information fusion and multimodalities of medical images is sought after to enhance retrieval of medical information.

1.2 Research Background

Information retrieval (IR) can be defined as a process of searching, recovering, and interpreting information based on question or query given (Baeza-Yates et.al., 1999). Since information can be recorded in various forms, such as tables, images, text and video, IR systems must be able to retrieve information from varying media representations. Multimedia Information Retrieval (MIR) involves extraction of semantic information from multimedia data sources which can be either directly perceivable media such as image and video or indirectly perceivable sources such as text and biosignal (Kludas et.al, 2008). MIR can be defined as the management (storage, retrieval and manipulation) of multiple types of media data (Kludas et.al., 2008). Recently MIR has been used in many particular domains such as medical, economic, agriculture, and automotive where generally, the data collection in specific domain contains multimedia information (Kalpathy-Cramer et.al, 2010; Ustin et.al, 2009). In medical domain, the medical data are inclusive of text (caption, title and descriptions) with embedded images (including medical images, illustrations, charts and visual material) or video clips with streamed image and audio content (Antani, 2010; Hsu et.al, 2009). Different multimedia data have their own different distinctive feature representations. The descriptions of these features reveal the summary of media content and pattern detection of each media (Petite et al., 2010). For example, texture and shape descriptions represent visual domain for content-based retrieval while tokenization and n-grams method are among the methods applied in text-based retrieval.

1.2.1 Multimedia Information Retrieval in Medical Application

With the availability of various medical imaging techniques such as x-ray, CT, MRI and US, the amount of digital data produced in hospitals increases incredibly fast. Therefore, the importance of dissemination of medical knowledge in MIR increases and has the potential to give considerable impact on the quality of care provided by clinicians and the tasks of efficient storing, processing and retrieving medical image data have become important research topics. The aspiration of MIR in medical application can be defined as delivering the needed information at the right time, the right place to the right persons in order to improve the quality and efficiency of care processes (Müller, 2003).

Conventionally, picture archiving and communication systems (PACS) are used in several hospitals. PACS (Berg, 2001; Ghebreab et.al, 2003) are basically computer networks that are used for storage, retrieval, and distribution of medical image data. Other MIR systems that specialize on medical data are as follows:

- IRMA (Lehmann et al., 2003) [a CBIR system for radiologic image archive],
- Medline (U.S. National Library of Medicine, 1998) [a biomedical information system which includes bibliographic information for articles from academic journals, involving medicine, nursing, pharmacy, dentistry, veterinary medicine, and health care]
- MedSearch (Hliaoutakis et al., 2006) [a system that supports retrieval of bibliographic information from Medline].

These medical retrieval systems involve solely image-based retrieval, text-based retrieval or combination of both image and text retrieval to represent medical data.

1.2.2 Information Fusion in Medical Retrieval System

Information fusion is a revision of efficient method to transform information from different sources and different points in time into a representation that leads to effective support for human or automated decision making (Boström et al., 2007). According to Klein (1993) the definition of information fusion is "*multilevel, multifaceted process dealing with automatic detection, association, correlation, estimation, and combination of information from single and multiple sources".* Generally, information fusion in medical domain is mainly concentrated on the combination of text and visual retrieval (Depeursinge & Müller, 2010). Combination of text and visual information potentially allows a reduction in semantic gap for more meaningful search and retrieval, reflecting the modelling representation of human observation for a particular image and visual information (Depeursinge et.al, 2010). This is because by using both information sources of text and visual will lead to more semantic understanding in the image as well as to improve the effectiveness performance in information retrieval system.

Medical data often contain multimodal information sources such as visual information (image) as well as text information. Both types of information are important for medical retrieval system (Muller et al., 2003). Due to the information limitation at different levels of sources, the application of information fusion becomes a real need in medical application.

Text retrieval is a type of information retrieval whereby the retrieved information is stored primarily in the form of text. Text retrieval system deals with the search of relevant documents which describe the information within document or looking for metadata about documents in a collection of dataset based on user-provided queries. User queries can be described as a range of multi-sentence of an information to an ad hoc, few words query. Generally, the process of text retrieval is to find relevant documents to user queries and provide relevant retrieved document based on the evaluation of matching results and sort them according to relevance as illustrated in Figure 1.1.

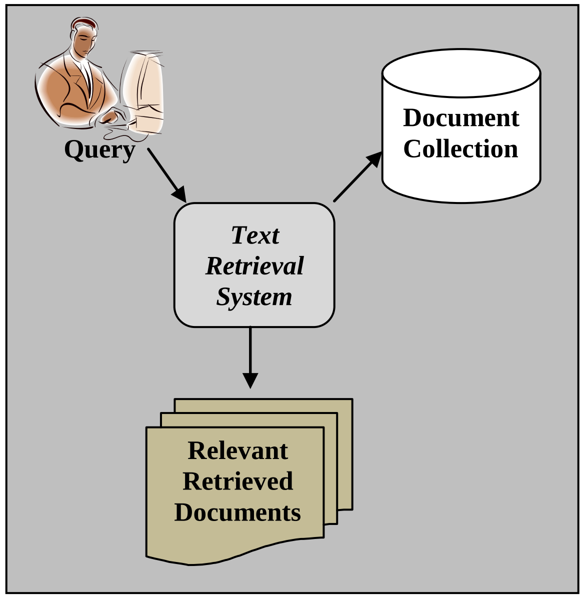


Figure 1.1: Process of Text Retrieval System

Medical domain is often claimed as one of the principal application domains in CBIR research field (Smeulders et al., 2000; Orphanoudakis et al., 1994). Due to the large amount of medical images, CBIR in medical application has become an important research topic. CBIR system is used to search similar medical images or cases in large medical image repositories based on diagnosis or anatomical region and capable to retrieve images with different diagnosis but have visually similar cases. Generally most of CBIR systems have a very similar architecture involving query image and images in the collection and comprising tools for the extraction of visual features, the storage for image features, distance measurements or similarity calculation and finally the output of retrieved images as depicted in Figure 1.2.

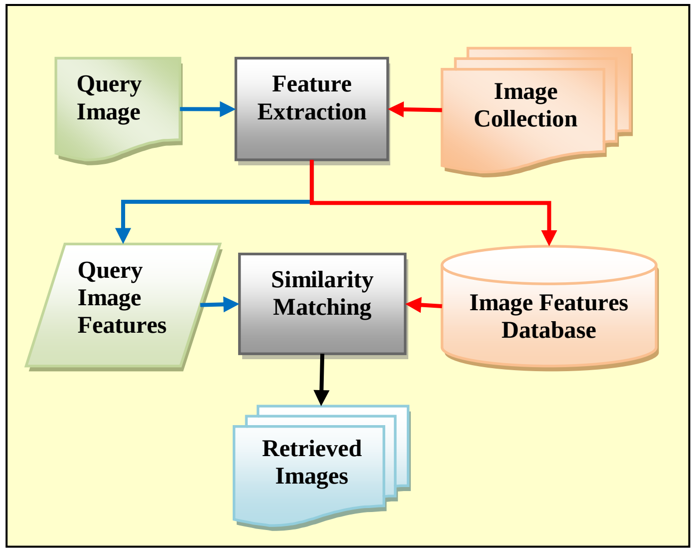


Figure 1.2: Process of CBIR system

1.3 Problem Definition

Thousands of multimedia data in medical field are produced daily which led to the challenging problem of developing techniques that can provide efficient and accurate search through large digital archives. The simple manual text-based query or request based on exact matching with database is becoming obsolete (Mihalcea, 2006). Additionally, in biomedical information term mismatch between user query and document has become an issue (Dinh & Tamine, 2011). The unstructured text in natural language can cause various different interpretations by different people. Text description in medical information is not sufficient for depictive subjective perception whereby medical image may contain several objects which convey important specific information (Prasad et al., 2009). The content of medical image itself may incur difficulties when described in words. Medical images which are produced by different scanning machines may display different features. Although with the existence of multi-modality medical images such as x-ray, CT and MRI, it is possible for researchers to create a retrieval system by using only manual text-based retrieval, it is laborious, time-consuming and prone to errors.

The drawback of conventional text-based retrieval motivates researchers towards more effective text-based retrieval and CBIR in medical field, whereby CBIR has been among the most vivid research areas in the field of computer vision recently. A number of CBIR systems on medical images have been developed such as IRMA which is an automated classification system for radiography medical image (Lehmann et al., 2003); SPIRS which is a retrieval system specifically for spine x-ray image (Hsu et al., 2009); MIMS as a medical image management system which concentrates on x-ray images (Chbeir et al., 2000) and ASSERT which is a content-based retrieval system that is specific on high resolution CT images of lung (Shyu et al., 1999). Nonetheless most of existing systems are task specific which is limited to a particular modality, organ or diagnostic study. As there are various modalities of medical images produced daily in tremendously large quantities on different anatomical structures, this issue needs to be dealt with.

Generally, multi-modality images are used to provide complementary and more information about the image obtained from different imaging systems (e.g., CT, MRI, x-ray) compared to only any single image type (unimodal). As such, there is a significant need for multi-modality medical image retrieval concept for flexibility of the system to improve the overall performance with more exhaustive information. The characteristics of each modality in medical image need to be differentiated. The differentiation of each modality can be measured by using quality characteristics such as blur, contrast and noise and the visual features of texture, shape and color (Schowengerdt, 2006).

Information fusion from heterogeneous sources can lead to better understanding and improved accuracy compared to a single source (Hall & Llinas, 1997). Consequently combination of text and visual content-based information may improve the performance of retrieval systems for multi-modality images since these information sources are able to extract the underlying semantic structure of particular documents (Depeursinge et.al, 2010; Saleem et.al, 2012; Zhao et.al, 2002; Zhou et.al, 2010).

1.3.1 Text Retrieval in Medical Application

Finding relevant information from large domain-specific text collections is an active research in text retrieval field and it is a challenging task to earn high efficiency and effectiveness in the performance (Volk et al., 2002). Domain-specific refers to restricted domain knowledge such as medical domain where terms are much more important than words (Radhouani et al., 2008). A term is defined as an exhaustive list of noun phrases that belong to a terminology; which has unique and significant meaning in a given domain (Missikoff et.al, 2003).

Using conventional text retrieval may not be suitable since this approach directly extracts words and these meaningless words are used for indexing. Alternatively, indexing using term should improve the precision in text retrieval system since the index denotes significance in meaning.

Terms in medical domain are highly synonymous and ambiguous. This leads to the motivation of expanding the original query terms using ontology (Bhogal et al., 2007). Identifying a suitable query expansion technique is required to improve performance in finding relevant documents based on user’s query. The queries inserted by users may be general and ambiguous which can adversely affect efficiency and accuracy. The acronyms and abbreviations might be different from the queries and medical data in the collection (Stevenson & Guo, 2010). There is also a scenario that the medical terms in the queries do not match with those in the collection (Soualmia et.al, 2012). However, there is a possibility that the synonyms of the query terms are in the collection (Chen et.al, 2006).

Various retrieval models such as Boolean model (Serra, 1980) and statistical models (vector space (Lee et.al, 1997) and probabilistic models (Merialdo, 1994)) need to be compared in order to identify a suitable retrieval model for this research. For a medical retrieval system, the performance measurement is based on the greater degree of relevant documents retrieved from queries. Therefore, it is significant to identify or create new ranking model to rank relevant documents according to the nearest degree to query and their importance.

1.3.2 Content-based Image Retrieval in Medical Application

CBIR or alternatively known as query by image content or content-based visual information retrieval deals with the search by analysing the actual contents of images rather than the metadata such as keywords or descriptions associated with the images (Venters et.al, 2000). CBIR has significant application in medical field whereby it can be used to retrieve the known pathology images that are similar to patients' images, thereby assisting doctors in diagnosis or clinical activities.

CBIR is based on automatically extracting features from the image itself. There are various forms in extracting visual features such as color, shape, and texture. Furthermore the extraction of features can be in global, local or combination of both descriptors. Therefore it is important to identify suitable methods and techniques to extract visual features. Unfortunately, in CBIR system the interpretation of low-level features is different from high-level human analysis.

There is the *semantic* gap of CBIR low-level features extracted by computer system and high-level human vision concept (Liu et.al, 2007). As such the information from low-level features is not sufficient to capture the semantic content of the image. Bridging the *semantic* gap between low-level features and high-level semantic concept used by human is still a challenging problem (Wang et.al, 2010). A method to bridge the *semantic* gap is by training the medical data using machine learning technique based on supervised classification. However for this technique, it is important to identify the suitable classifiers for the classification (Zhang et.al, 2012).

Digital Imaging and Communications in Medicine (DICOM) (Pianykh, 2011) is a standard medical imaging format which includes information of modality, anatomy and pathology in the DICOM header. According to Lehmann (2002), 16% error rate has been reported for the field "anatomical region" in the DICOM header which contributes to the problem of using purely text-based methods in medical domain.

For MIR in medical application, it is significant to know the semantic meaning of which category the image should be classified prior to any processing procedure (Mojsilovic et al., 2000). Therefore, the medical semantic conceptualization is essential to support effective retrieval system.

1.3.3 Information Fusion in Medical Application

Although the purely image-based method may not be able to replace text-based methods, it is very useful in complementing text-based method. Fusion of information sources of text and visual content may increase the performance of MIR in medical field. However, the detail components and processes regarding textual and visual features, their extraction and fusion approach need to be identified. Selection of an appropriate MIR model in information fusion for multi-modality medical data collection is important in order to produce a retrieval system that is capable to retrieve multi-modality medical images. The stages involved in the testing process need to be established, taking into account the order and number of stages with either unit level of text and content-based separate execution or integration level of both text and visual sources.

1.4 Main Aim and Objectives

The main aim of this research work is to develop efficient and effective multi-modality medical information retrieval systems based on text, visual content and information fusion. In order to achieve this main aim, specific objectives are as follows.

Objective 1: To design text-based framework for multi-modality medical information retrieval system.

Sub-objectives:

(1a) To manage medical documents in data collection for effective processing in assessing relevant documents in the higher position ranking.

(1b) To manipulate appropriate thesaurus in extracting significant information for efficient performance.

(1c) To formulate query model for medical retrieval system using external knowledge thesaurus in order to ease the medical terms extraction in the query.

(1d) To form the strategies and steps in retrieval and ranking of relevant documents for effective performance.

(1e) To conduct evaluation on effectiveness and efficiency of the framework.

Objective 2: To design visual content-based framework for multi-modality medical image classification system.

Sub-objectives:

(2a) To implement appropriate indexing algorithm in visual features extraction.

(2b) To create a platform for semantic description of visual features classification.

(2c) To identify and apply suitable machine learning technique for multi-modality medical image classification.

(2d) To conduct a set of evaluations for global and local descriptors with visual features.

Objective 3: To design multi-modality medical information retrieval framework based on information fusion of text and visual content.

Sub-objectives:

(3a) To form the steps for an integrated framework of text and content-based information sources.

(3b) To conduct experiments based on information fusion of bimodal text and visual features.

(3c) To perform a set of unit and integration evaluations with different features (text, visual and information fusion) and parameter settings in signifying the strength of our approach.

1.5 Research Focus and Scope

This research is mainly focused on text, content-based and information fusion of multi-modality medical information retrieval system. The testing and evaluation are based on medical data from ImageCLEF 2010 medical task collection. This collection consists of multi-modality medical images namely x-ray, ultrasound, MRI, CT-scan, nuclear medicine, graphical and optical imaging whereby each image is attached to a manually annotated medical text document in XML format and the average number of words for each document is 67 (details in section 4.4). The classification of multi-modality medical image involves unbalanced datasets whereby some classes have huge number of samples compared to other classes with smaller number of samples. Due to the imbalance among the training data, one solution is to define a threshold for each class. Sixteen ad-hoc queries have been provided as in Appendix A and relevant judgment list is used for evaluation. The evaluation for text-based information retrieval involves precision, recall, mean average precision (MAP) and F-measure. These measurements measure the performance of retrieval system based on number of relevant documents retrieved. Meanwhile for medical image classification, the performance measurement is based on percentage of correctness rate in classifying multi-modality medical image. The final results are compared with other experimental setup and results generated by other researchers who used the same dataset.

Initially, the design, development and evaluation process were done separately for text and content-based retrieval. The medical thesaurus is used to extract medical terms from the data. For content-based process, semantic-based classification is applied. Feature extraction of texture, shape and color is based on global and local descriptors and the combination of both. These features were trained with model based on semantic description of medical concepts namely modality, anatomy and pathology using supervised machine learning method. In improving the retrieval system for multimedia data, information fusion technique is applied. Combination of different information sources such as text and visual features may improve the accuracy results in MIR.

1.6 Research Questions

Several research questions have been posed as the questions become the driving force behind this research process and serve as guideline to conduct this research:

Q1. How to manage huge text documents in a data collection?

Q2. How to identify medical terms in a query or text document?

Q3. What is the process involved to extract significant medical terms from medical documents?

Q4. What is the suitable process to rank retrieved relevant documents?

Q5. What is the process involved to extract visual features in multi-modality medical retrieval system?

Q6. Does application of semantic description affect meaningful retrieval system?

Q7. Do combining different sources of text and visual content affect performance of multi-modality medical information retrieval system?

1.7 Research Methodology and Approach

The research methodology comprises several principal stages as follows:

- *Literature review:* The investigation of literature is divided into three main categories of text, content-based and information fusion analysis. The text-based retrieval strategies were studied and compared. Furthermore the query expansion techniques were reviewed and the utilization of medical thesaurus is discussed. Other text-based retrieval systems that used the same ImageCLEF 2010 dataset were reviewed. As for content-based analysis the approaches to extract visual features and classification based on semantic representation were investigated. Multi-modality medical image characteristics are discussed. The description of visual features of shape, texture and color were studied. The information fusion between different sources of text and content-based is discussed and the systems developed by other researchers are outlined.

- *Framework Design and Development:* Text and visual content-based frameworks for multi-modality medical information are designed separately. The text-based framework of the system (text-based M3IRS) concentrates on retrieving the relevant medical documents based on text query. For text-based retrieval, there are several processes involved which are medical document pre-process, query expansion, retrieval and ranking processes as shown in Figure 1.3 (details in chapter 4). The retrieval strategy involves Boolean Model (Salton et. al, 1983) and number of term frequency (Salton et. al, 1975). MedHieCon ranking model is used to rank the results retrieved. Generally the query will be processed and medical terms from the query will be extracted. Prior to that, medical documents will be pre-processed to remove noise and duplication and store in local repository. Later the medical terms from query will be matched with documents in local repository and produced relevant documents. These relevant documents will be ranked using MedHieCon ranking model associate with total of term frequency.

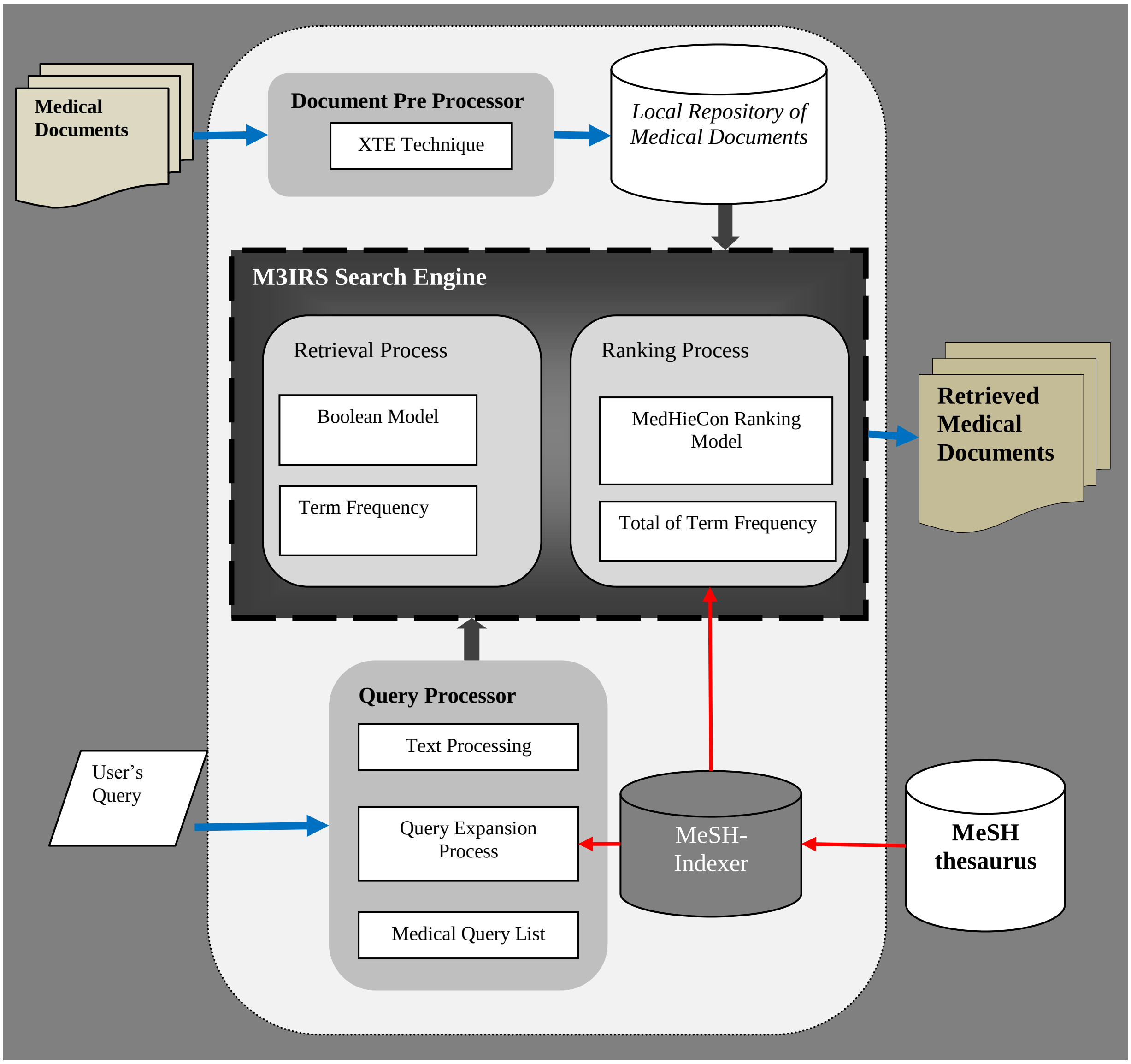


Figure 1.3: Text-based Framework of M3IRS

In contrast the visual content-based framework (content-based M3ICS) is based on multi-modality medical images where the framework is designed to classify the medical images based on semantic-based medical concepts classification as shown in Figure 1.4 (details in Chapter 5). The feature extraction is based on global and local levels where the global features extract the whole image information. Meanwhile the local level the feature extractions are based on 2×2, 4×4, 8×8 patches and also interest blocks of the medical image. The train data are classified based on medical concepts using supervised machine learning method. The train feature vectors are then stored at the train vector storage Later the test data is classified based on comparison function between test data and train data and the outcome produced is the medical image annotation. The comparison function is based on distance measurement calculation in two classifiers which are k-NN and support vector machines (SVM).

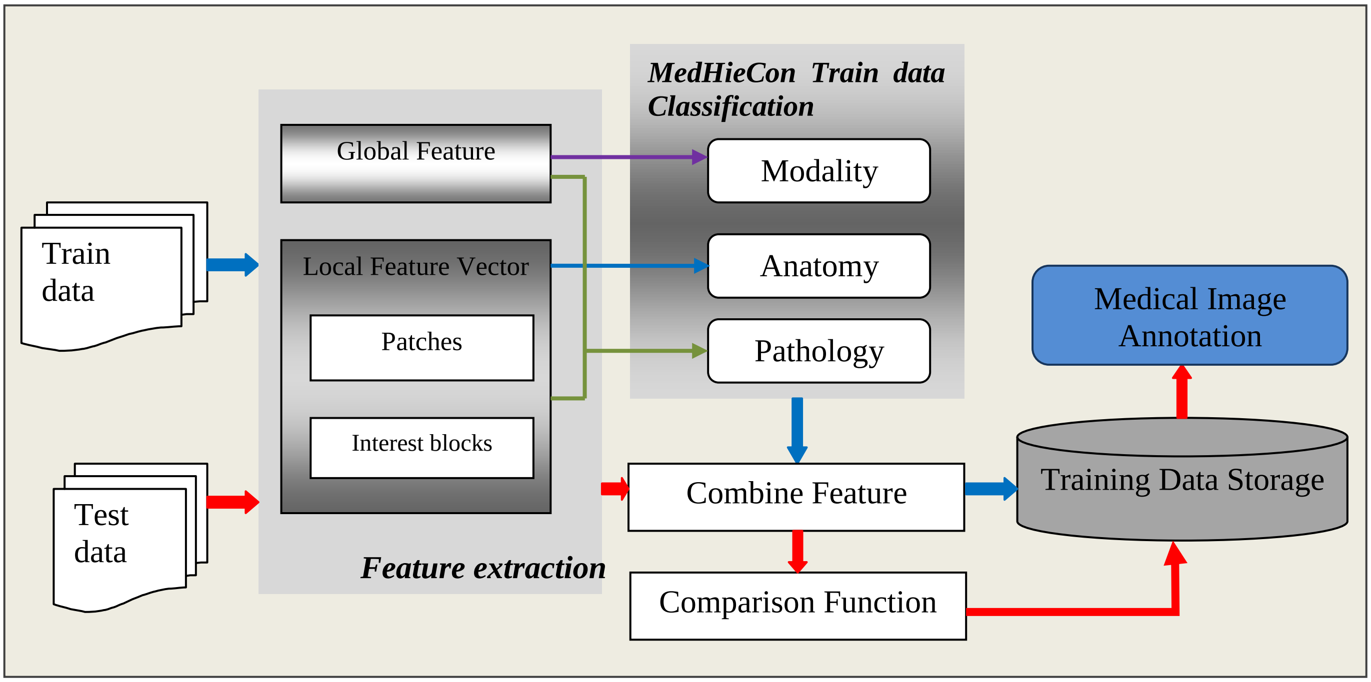


Figure 1.4: Content-based Framework of M3ICS

Both text and content-based frameworks apply the same medical hierarchical conceptual (MedHieCon) model which contemplates on the concept entity of modality, anatomy and pathology. The third aim of this research is to design information fusion of both textual and visual analysis for the multi-modality medical information retrieval framework (IFM3IRS) as depicted in Figure 1.5 (details in Chapter 6). Late fusion technique based on hierarchical processing is used in this research. In this framework the text-based retrieval is executed at the initial level and the results from text-based retrieval will be the input for the next level which is content-based retrieval. The main purpose to extract visual features from the images in text-based results is to filter the modality in order to retrieve accurate modality that is requested from the query (details in section 6.2).

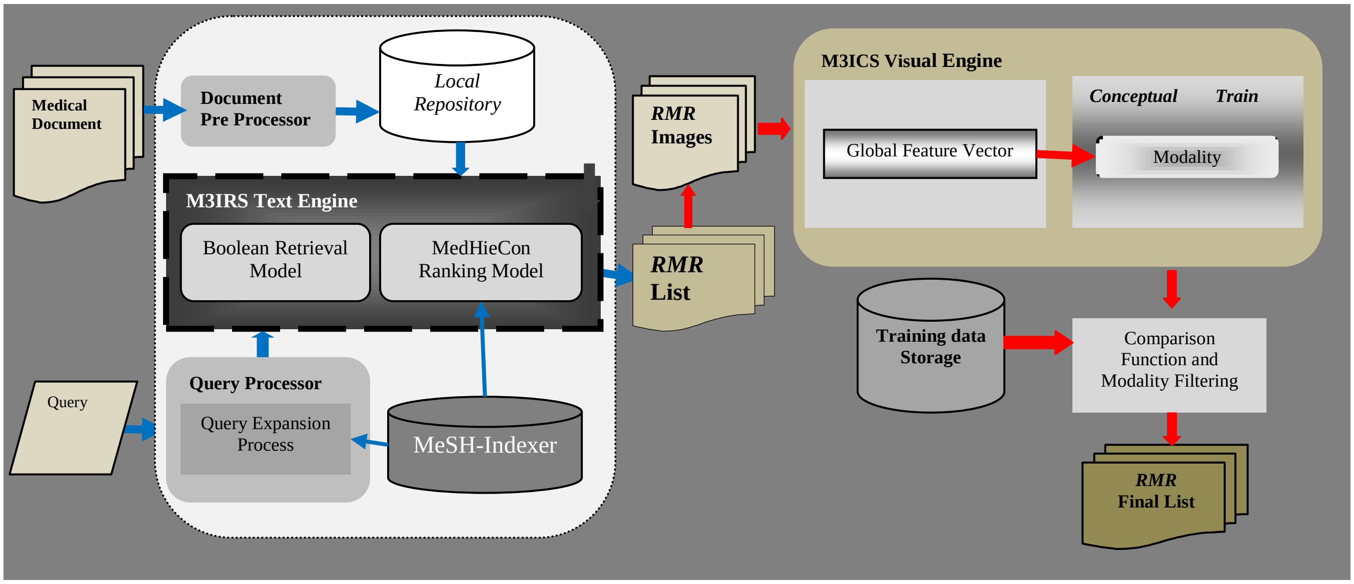


Figure 1.5: Information Fusion of Text and Visual Framework (IFM3IRS)

- *Experiments:* Unit and integration experiments were conducted. The execution of unit experiments is to amend and validate the text and content-based frameworks. The integration experiments were conducted using medical data from ImageCLEF 2010 medical task for both text and content-based features. Details on experimental setup are described in Chapter 7.
- *Evaluation, Benchmarking and Results Analysis:* The evaluation was performed on ImageCLEF 2010 medical task dataset. Performance will be measured based on the effectiveness of the framework which involved the evaluation of MAP, precision, recall and F-measure for M3IRS and IFM3IRS and percentages of correctness rate in medical image classification are analysed for M3ICS. There will be standard relevant judgments provided by ImageCLEF 2010 dataset as a benchmark to evaluate the frameworks

performance. Furthermore, there will be a comparison between these frameworks with other run systems that used the same dataset.

1.8 Research Achievements

The following are the main research achievements produced as research output with the details described in corresponding chapters:

- New frameworks of medical information retrieval system are formulated.
- Local repository of medical documents and MeSH-Indexer are produced after document pre-processing and thesaurus organization.
- Query expansion and enrichment are obtained.
- A novel MedHieCon ranking model is developed.
- Multi-modality Medical Information Retrieval System codenamed M3IRS and Multi-modality Medical Image Classification System codenamed M3ICS are designed and developed for both multi-modality medical documents and medical images.
- IFM3IRS framework based on information fusion is developed.
- Performance comparisons between different features of textual and visual techniques are obtained.
- Research publications (see Appendix B).

1.9 Thesis Outline

The thesis is organized as follows:

- Chapter 1 (Introduction) introduces the research which includes the motivation of undertaking this research and research background. Underlying problems are identified with the main aim and objectives as well as the research focus and pertaining research questions. It also gives an overview of the main stages involved in the research methodology and approaches as well as the main achievements.
- Chapter 2 (Multi-modality Medical Images) describes medical image quality characteristics namely contrast, blur and noise. This chapter also presents review on the eight modalities of medical images used in this research study.
- Chapter 3 (Review on Text, Content-based and Information Fusion Retrieval in Medical Domain Application) describes the state-of-the-art of information retrieval in medical domain. It explains the application of text-based retrieval in medical domain and medical thesaurus in order to extract significant information in medical documents. Retrieval strategy models are compared. It covers query expansion. Furthermore this chapter reviews visual features namely texture, shape and color in global and local descriptors. Machine learning classifiers of k-NN and support vector machine (SVM) are used for performance comparison.
- Chapter 4 (M3IRS Text-based Framework and Module Descriptions) introduces M3IRS text-based framework which consists of four main components, namely document pre-processor, query processor, retrieval strategies and ranking strategies. Two types of ranking models which are Comprehensive and MedHieCon ranking models are introduced.

- Chapter 5 (M3ICS Content-based Framework and Module Descriptions) introduces M3ICS content-based framework which consists of two main components, namely feature extraction phase and classification of conceptual train data. Two types of architectures including global and local descriptors are introduced. Furthermore medical concepts are introduced in semantic annotation for medical images based on modality, sub modality, anatomy and pathology.
- Chapter 6 (IFM3IRS Information Fusion of Text and Content Retrieval) introduces the combination of textual and visual results to increase IFM3IRS performance. Sequence processing in late fusion technique is used where the result from textual framework is evaluated using visual features in order to get better result.
- Chapter 7 (Experimental Setups and Framework Evaluation) describes the experimental setups executed in this research. This includes five main sections namely (i) multi-modality medical image characteristic evaluation, (ii) M3IRS textual-based framework, (iii) M3ICS content-based framework, (iv) IFM3IRS information fusion framework and (v) performance evaluation criteria.
- Chapter 8 (Results and Discussion) presents the results based on the experiments described in chapter 7.
- Chapter 9 (Conclusions and Future Research Implication) concludes this research with various findings and suggestions for future research direction.

2.0 Multi-modality Medical Images

2.1 Introduction

Modern computer technology has created the possibility for human to diagnose abnormal conditions and thus able to guide therapeutic procedures (Yamaguchi et.al, 2013). Nevertheless, there is no such medical image that can completely reveal the human structure. Sprawls (1995) mentioned that the visibility of body structure and the quality of medical image is depend on medical imaging equipment, operator's skill, imaging time and patient radiation exposure.

New technologies in medical imaging methods have led to the existence of multi-modality medical images such as x-ray, US and CT scan. Quality medical images help doctors in good decision making (Covens et.al, 2012). This chapter presents factors that influence the quality of medical images followed by the description of multi-modality medical imaging method.

2.2 Medical Image Characteristics

Different modalities of medical imaging (MRI, CT & etc) generate images either via detection of photons or the use of electromagnetic waves (Khandelwal et.al, 2012). These different medical imaging methods divulge different characteristics of the human body. The quality of these modalities can be measured and compared through contrast, blur, noise, artifacts and distortion (Tsai et. al, 2011) as shown in Figure 2.1. The quality characteristics of contrast, blur and noise concentrates on the affect of visibility objects in imaging methods. Meanwhile artifacts and distortion characteristics do not significantly affect object visibility and interpreted more as an anatomical feature (Iskandrian & Garcia, 2008). In addition to visual features of shape, texture and color that are used for feature extraction; we also emphasize on the measurement of contrast, blur and noise for each modality for further analysis to review the significant affect of object visibility in multi-modality medical images.

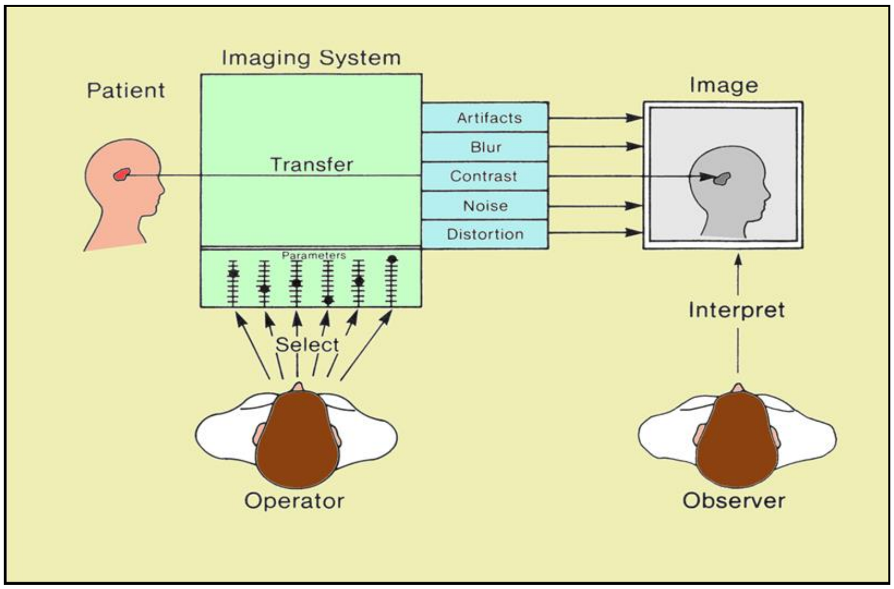


Figure 2.1: Components associated with the medical imaging process.
Adapted from (Kuhn, 1995)

2.2.1 Contrast

Contrast is the most fundamental characteristic of an image. Contrast represents visual properties determined by the difference in color, intensity and brightness of the image. Visibility of anatomy and sign of abnormal conditions in medical images depend on the contrast that is present in the image (Saleem et al., 2012).

In medical image, the contrast sensitivity represents the lowest contrast of the image which is able to visualize more detail objects such as soft tissue. Contrast sensitivity is one of the significant characteristics in imaging methods since this variable relates to the system's ability to translate physical object contrast into image contrast. CT image generally has higher contrast sensitivity than conventional radiography due to its ability of imaging soft tissue objects which are not visible in radiography image (Sprawls, 1995).

Contrast of an image can be measured by quantifying the intensity contrast between a pixel and its neighbours over the whole image (Soille, 2003) which can be mathematically represented as

$$\sum_{i,j} |i-j|^2 \, P(i,j) \tag{2.1}$$

where *i* is the row number; *j* is the column number in the image and *P(i,j)* is the normalized value in the cell *i,j*. For example images may have poor contrast due to glare. In image processing normalization, also known as contrast stretching is a process which changes the range of pixel intensity values. Normalization has the purpose to bring the image into a range that is more familiar or normal to the senses to achieve consistency in dynamic range for a set of images to avoid mental fatigue (Sivakumar, 2004).

2.2.2 Blur

Blurring effect occurs in all imaging processes including medical imaging methods. Different modalities have different effect of blur. This characteristic reduces the contrast and visibility of small objects (Liang, 2008). Nevertheless there are medical imaging methods that produce sharp images. Blur effect can be measured in units of length. The value represents the width of small blurred object in various modalities as depicted in Figure 2.2. The figure shows that radiography image has the most blur effect. This explains why radiography image has limitation in viewing the visibility of small objects and structures.

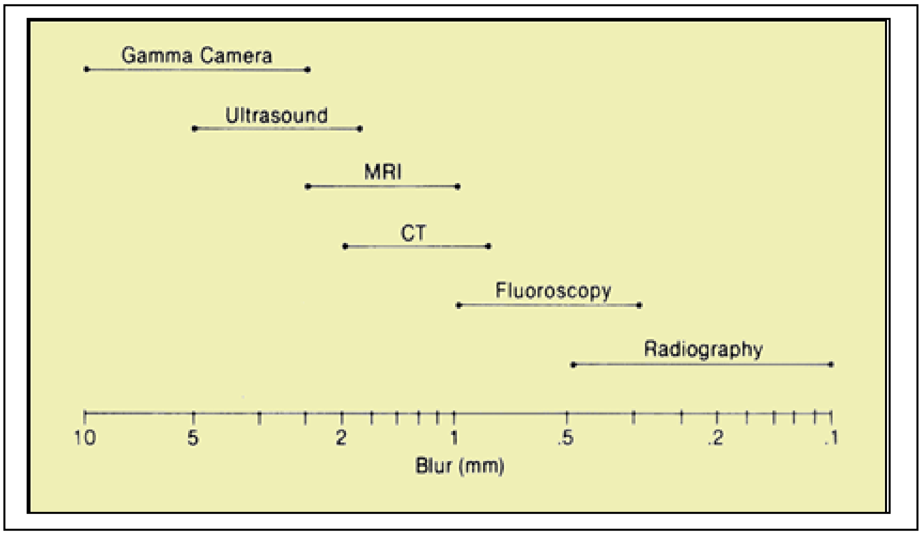


Figure 2.2: Range of blur values in multi-modality of medical image. Adapted from (Kuhn, 1995)

Blur effect can be presented in three descriptions namely blur size, blur shape and blur profile. Blur size represents the amount of blurring which is presented as the dimension of blurred image of a very small object point (Ramakrishnan, 2010). Different modalities may produce different blur shape. For example x-ray images produce round blur patterns. In contrast, CT and MRI images produce square or 3D blur patterns. Finally, the blur profile specifies the manner in which the point image is spread or distributed within the blur area (Susil & Taylor, 2007).

Medical images are complex and different modalities of medical image have different levels of blur. No-reference blur estimation method can be used to measure blur in the image. This method is used to measure the blur in an image by blurring it and comparing the variations between neighbouring pixels before and after applying low-pass filters (Crete et al., 2007) as illustrated in Figure 2.3.

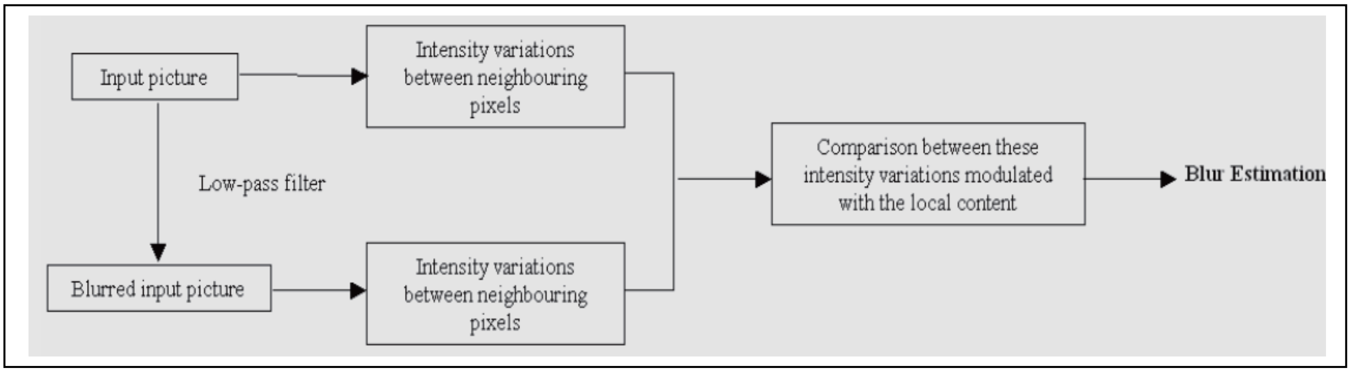


Figure 2.3: Simplified flow-chart of the blur estimation principle
Adapted from (Crete et al., 2007)

2.2.3 Noise

Image noise is a common problem in all imaging processes. The existence of noise leads to mottled, smeared, grainy textures or snowy appearance in images. Sometimes noise is used to cover and reduce the visibility of certain features within the images. This function is significant to use if we need to reduce the visibility of low-contrast objects (Adams Jr, 2010). NM medical imaging has the most noise effect followed by MRI, CT and US. In contrast, modality of x-ray constructs less noise since x-ray has the highest value of blur. The noise in the images can be reduced by blurring the images. This is because blurring technique can reduce the visibility of image details (Wang et.al, 2004). Noise in an image can be measured using spatial frequency distortion weighting (Miyahara, 1998).

Signal-to-noise ratio (SNR) is used in science and engineering field as the measurement to compare the level of a desired signal to the level of background noise (De Boer et al., 2003). It implies the ratio measurement of useful information (signal) to false or irrelevant data (noise) as shown in equation (2.2).

$$SNR = \frac{P_{signal}}{P_{noise}} \tag{2.2}$$

where P is average power. Nevertheless signal has a very wide dynamic range. To reduce this wide range, logarithmic function is used to obtain normal scale of values. For instance SNR is described using logarithmic decibel scale in amplitude ratio as follows

$$SNR_{db} = 20\ log_{10} \frac{A_{signal}}{A_{noise}} \tag{2.3}$$

Measuring SNR in images requires vision contrast sensitivity function which can be presented in spatial frequency (Miyahara, 1998). Spatial frequency can be approximately modeled as

$$S(\omega) = 1.5e^{-\sigma^2\omega^2/2} - e^{-\boldsymbol{2\sigma^2\omega^2}} \tag{2.4}$$

where $= 2$, $\omega = \frac{2\pi f}{60}$, $f = \sqrt{u^2 + v^2}$ and u and v are the horizontal and vertical spatial frequencies in cycles per degree.

2.2.4 Artifacts

The characteristics of contrast sensitivity, blur, and noise in imaging method have caused the visibility of certain body objects. Nevertheless image artifacts create image features that do not represent a body structure or object and can obscure a part of an image or may be interpreted as an anatomical feature. A variety of factors associated with each imaging method can cause image artifacts such as the changes of main magnetic field around certain objects or tissues due to the difference properties of magnetic susceptibility (Toennies, 2012).

2.2.5 Distortion

A medical image should not only make internal body objects visible, but should give an accurate impression of image representation. Image distortion involves imaging procedure that provides size, shape, and relative positions of the image (Maintz & Viergever, 1996).

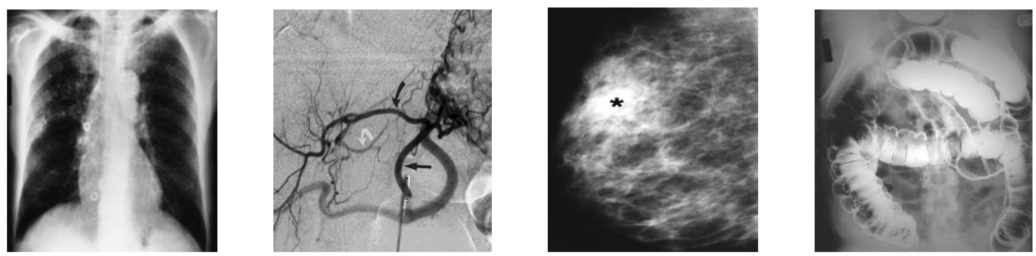
Since in many situations, artifact and distortion characteristics do not significantly affect object visibility and diagnostic accuracy, we will not use these characteristics as quality characteristics measurement.

2.3 Multi-modality Medical Imaging Method

In this research we concentrate on eight different modalities of medical images namely x-ray, computed tomography (CT), ultrasound (US), nuclear medicine (NM), positron emission tomography (PET), magnetic resonance imaging (MRI), optical imaging (PX) and graphic imaging (GX) from ImageClef 2010 medical data collection are used (Müller H., 2010).

2.3.1 X-ray

X-ray imaging is a transmission-based technique and a form of electromagnetic radiation. Different tissues have different value of contrast because of different attenuation of x-ray. Image produced in x-ray radiography is a two-dimensional projection of the tissue lying between x-ray source and the film. There are several different imaging techniques that apply x-ray namely angiography, fluoroscopy and mammography (Ammari, 2008) as depicted in Figure 2.4. Angiography is used on visualizing the inside of blood vessels and organs of the body, such as the arteries, veins and the heart chambers. Fluoroscopy concentrates more on genitourinary and gastrointestinal. Fluoroscopy transmits a continuous x-ray image on a monitor, like an x-ray movie for diagnosis by displaying the movement of a structural body part or of an instrument or dye as contrast agent through the body. Mammography focuses on detecting small lesion in breast (Shambaugh et al., 1995).



(a) (b) (c) (d)

Figure 2.4: X-ray images of (a) chest x-ray radiography, (b) blood vessel angiography (c) breast mammography and (d) colon fluoroscopy. Adapted from (Müller et al., 2010; MedPix, 2009).

The quality characteristic of x-ray imaging can be measured from SNR, non-uniform distribution of signal intensities, spatial resolution and contrast-to-noise ratio (CNR) which is the contrast of image that relates to the difference on intensity between body regions. The effect from SNR and spatial resolution also affect the CNR value. Therefore the ability of a physician to interpret an x-ray image depends on CNR value (Kalender, 2001).

2.3.2 Computed Tomography (CT)

Computed tomography is a medical imaging method that produces three-dimensional radiography image of a body structure that is constructed by a computer from a series of plane cross-sectional images made along an axis (Ammari, 2008). An example is illustrated in Figure 2.5. CT is used for various clinical purposes such as cerebral scans, pulmonary disease detection and abdominal imaging (Kalender, 2001). CT image is the first medical application that utilizes x-rays for forming images of tissues based on their x-ray attenuation coefficient. CT images have a high spatial resolution value and provide reasonable contrast between soft tissues.

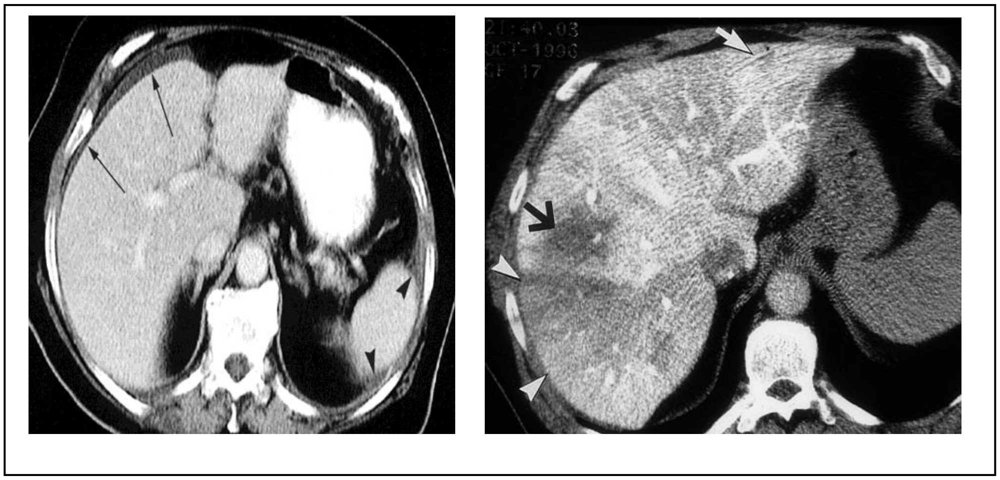


Figure 2.5: CT images of abdomen. Adapted from (MedPix, 2009; Müller et al., 2010)

Fundamentally, tomography imaging deals with reconstructing of a series of one-dimensional projections to obtain image integral in the direction specified by different angles, as illustrated in Figure 2.6.

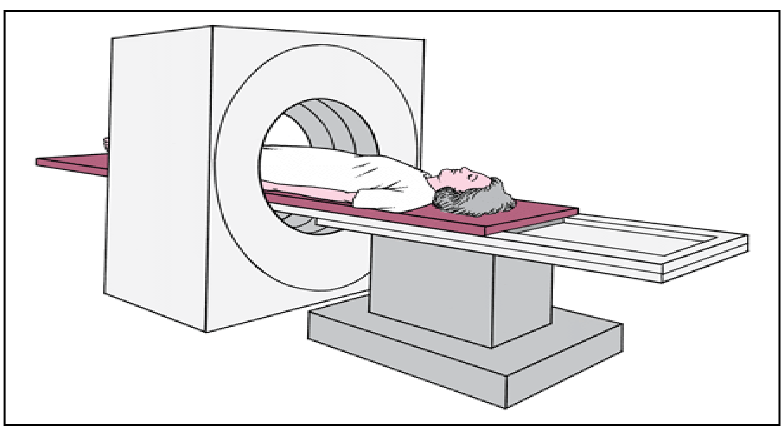


Figure 2.6: A patient receiving a CT scan for disease scanning Adapted from (Macmillan Cancer Support, 2010)

2.3.3 Ultrasound (US)

Ultrasound is a frequency-based, relatively simple, portable and relatively inexpensive diagnostic imaging method (Cosgrove, 2011). Ultrasound image is produced via backscattering of mechanical energy from boundaries between tissues and from small structures within tissue. Different values of frequency are used to view different objects. For example low frequencies of 1-3 MHz are used to view deep-lying structures such as liver and higher frequencies of 5-10MHz are used for objects closer to body surface such as baby fetus (Wieszczycka & Scharf, 2001). Figure 2.7 illustrates an US. The US image characteristics such as contrast, signal intensity and noise are determined by the propagation properties of ultrasound via tissue and the interactions that give rise to backscattered signals.

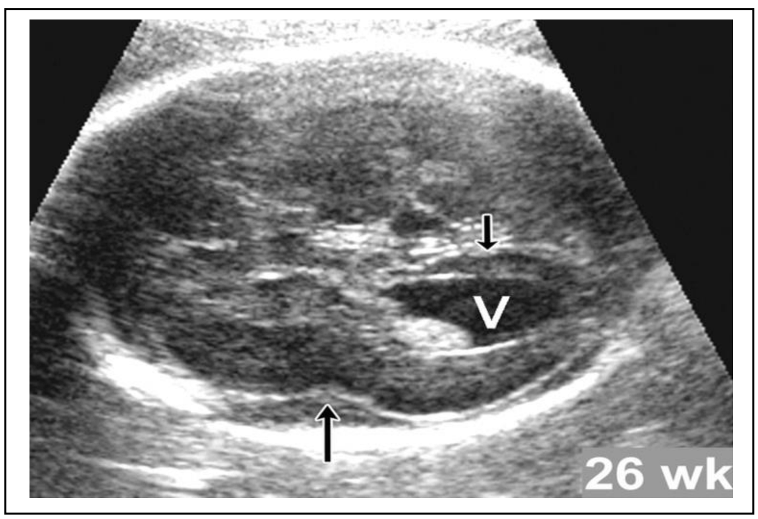


Figure 2.7: Ultrasound image of fetus. Adapted from (Müller et al., 2010)

This imaging method is mostly used in obstetrics and gynaecology involving the assessment of fetal health and detection of compromised blood flow in veins and arteries. However, the disadvantage of ultrasound is that is has relatively poor soft-tissue contrast and not all body organs can easily be imaged. Nevertheless US is a fast imaging technique since it can be performed a full image from one single transmit whatever the size simultaneously without the time delay and has high spatial resolution (Kulandavelu et al., 2006).

2.3.4 Nuclear Medicine (NM)

A unique capability of nuclear medicine is the use of specific radiotracers known as radiopharmaceuticals for imaging organ function and disease condition (Moriguchi et.al, 2013). Nuclear medicine procedures have the ability of mapping physiological function and metabolic activity which contribute to provide more specific information about the organ function and dysfunction (Signore et al., 2010). In contrast with other modalities of medical imaging, NM technique images the spatial distribution of radiopharmaceutical introduced into the body. The role of NM image is as an indicator to monitor the development of patient's disease by imaging the pathological conditions via inhalation into the lungs, direct injection into the bloodstream and oral administration (Cherry, 2003). The chemical structures of particular radiopharmaceuticals reveal the biodistribution in the body and act as strong indicator of disease. Figure 2.8 shows an example of nuclear medicine image. It is significant to have imaging technique that is sensitive to these early biochemical changes for clinical diagnosis (Stoker, 2009). The mapping of the radiopharmaceutical distribution produces images of functional morphology of organs in a non-invasive manner. It is useful for diagnosis of many common diseases associated with organ malfunctioning and in detection of certain type of cancers (Theis & Meyer-Bäse, 2009). The characteristics of NM image are low SNR and poor spatial resolution. However, this imaging method has very high CNR value compared to other modalities. Due to the availability of a vast range of specific radiopharmaceuticals there are widespread usage and growing demands for such techniques (Malvi, 2012).

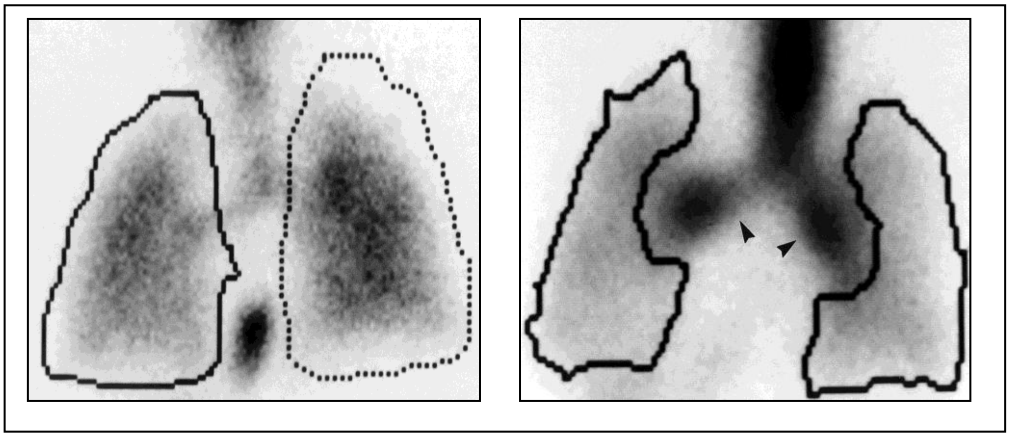


Figure 2.9: Example of nuclear medicine images of pediatric pulmonary and cardiovascular complications. Adapted from (Müller et al., 2010; Kendall, 2010).

2.3.5 Positron Emission Tomography (PET)

Positron emission tomography is a nuclear imaging technique that is used to map the biodistribution of positron-emitting radiopharmaceutical within a body as illustrated in Figure 2.10.

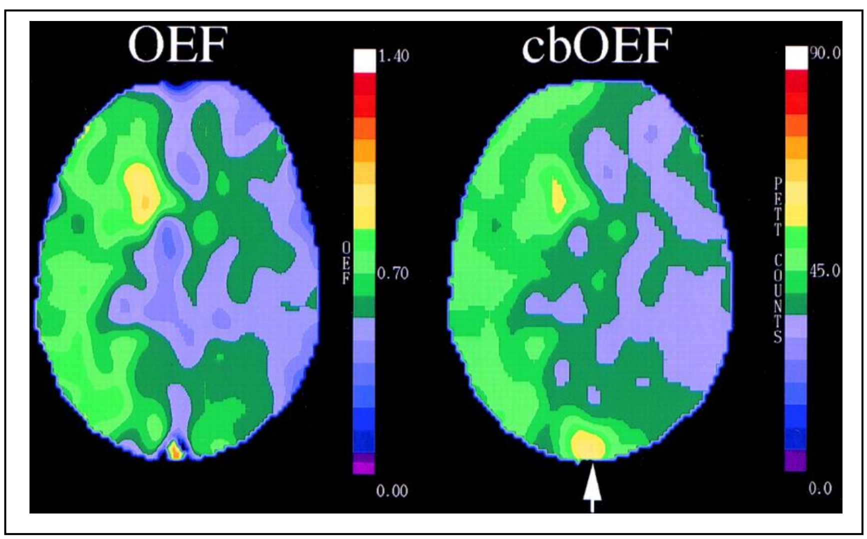
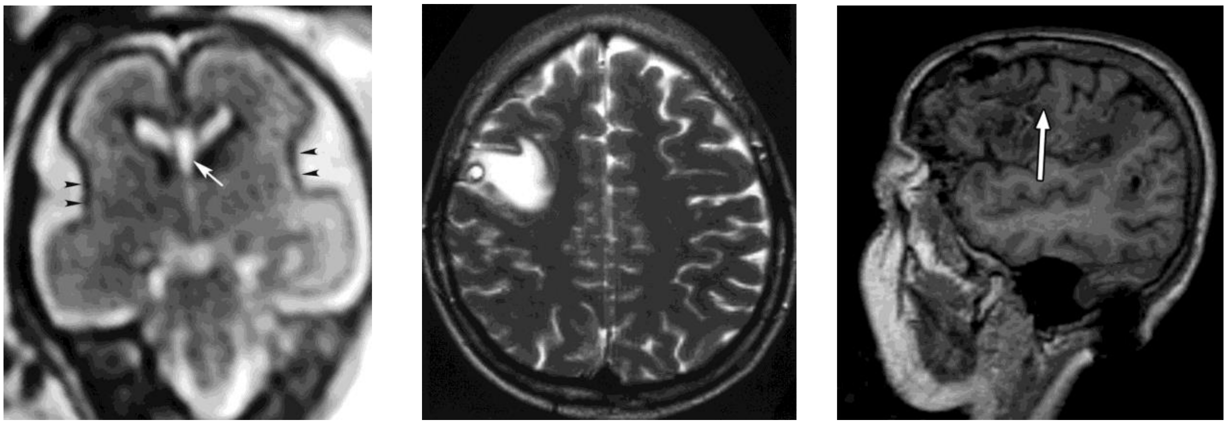


Figure 2.10: PET method for predicting ischemic stroke in patients with symptomatic carotid arterial occlusion. Adapted from (Müller et al., 2010)

Positron-emitting tomography is a tracer system that detects pairs of gamma rays emitted indirectly and thereby produces a three-dimensional image or picture of functional processes in the body (Valk et al., 2006). PET concentrates on measuring the physiology and function rather that imaging the structure of the anatomy. PET is clinically mainly used in oncology, cardiology and neurology for tumors in breast, lung, head and neck. This imaging method can distinguish diseased from healthy tissue (Kitson et al., 2009). However PET is an expensive medical imaging method which typically costs USD 1.5 to 7.5 million for a system.

2.3.6 Magnetic Resonance Imaging *(*MRI)

Magnetic resonance imaging is a three-dimensional imaging method and it's excellent in soft-tissue contrast and high spatial resolution. The drawbacks of MRI are that (i) this technique is slower than US or CT scan, vulnerable to patient motion and (ii) the cost of MRI imaging method is relatively high. In MRI imaging technique, the patient is placed inside a scanner with a very strong magnet (Fass, 2008). Then spatial information is encoded into the image using magnetic field gradients. An MRI technique is mainly used for assessing brain disease, spinal disorder and musculoskeletal damage. In clinical diagnosis, MRI technique can provide a series of slices which include well defined orientation and thickness via anatomical area of interest (Lee & Carroll, 2010). The choice of slice-select direction states the orientation of coronal, axial and sagittal as illustrated in Figure 2.11.



(a) (b) (c)

Figure 2.11: A series of MRI brain images in (a) coronal, (b) axial and (c) sagittal directions Adapted from (Müller et al., 2010; Kendall, 2010).

2.3.7 Optical Imaging (PX)

Optical imaging used in this research is categorized into two categories namely microscopy image and gross medical image. Microscopy image as shown in Figure 2.12 involves inference from the deflection of light emitted from laser or infrared source to structure, texture, anatomy and chemical properties of material such as crystal and cell tissue (Weissleder & Mahmood, 2001).

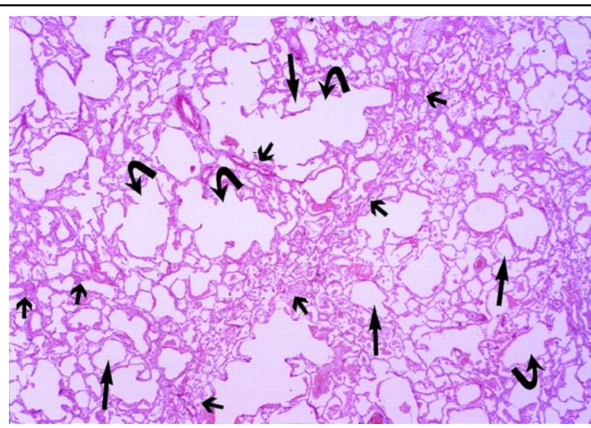


Figure 2.12:Microscopy image of cell tumor of abdomen Adapted from (Müller et al., 2010).

A microscopy image process begins with fundamental techniques and intended to produce the most accurate information contained in a microscopic sample. The process includes adjusting the brightness and contrast of the image, averaging images to reduce image noise and correcting for illumination non-uniformities (Russ, 2006).

Gross medical image is for the study of anatomy at the macroscopic level as depicted in Figure 2.13.

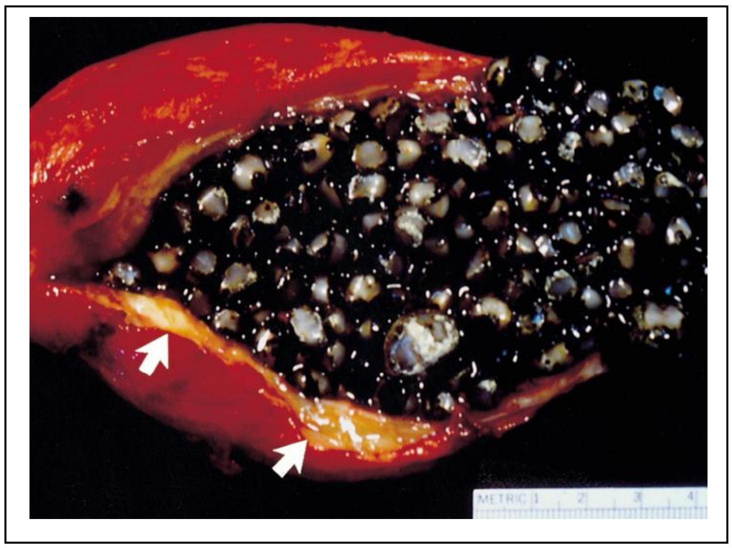


Figure 2.13: Gross medical image of a resected gallbladder Adapted from (Müller et al., 2010).

Gross medical image represents the anatomy structures that are too small to be seen by eye such as internal organs. Endoscopy is used to view internal organs in which a video camera-equipped instrument is inserted via a small incision in the subject to explore the internal organs and other structures of living animals (Gono et al., 2004).

2.3.8 Graphic Imaging (GX)

Graphic imaging represents graphical medical data such as chart and diagram. For example ECG diagram that is used to measure the electrical activity of the heart as shown in Figure 2.14

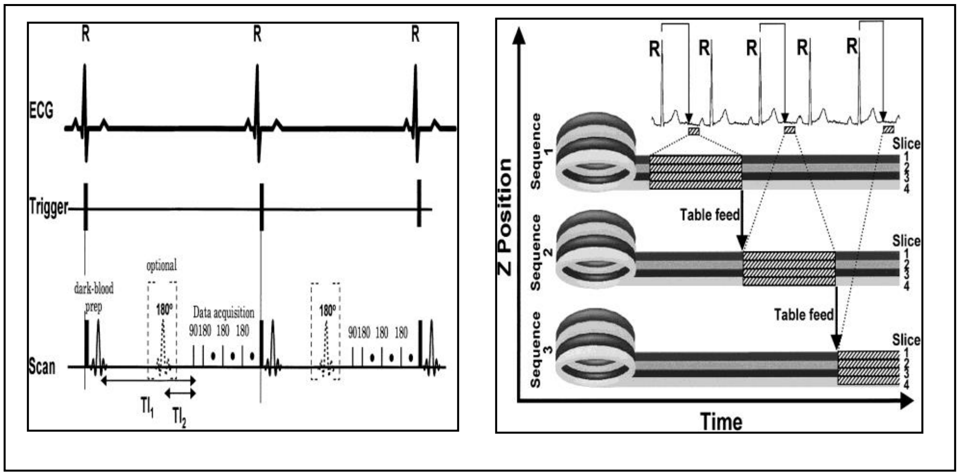


Figure 2.14: Diagrams of different pulse sequences for the visualization of myocardial infarction. Adapted from (MedPix, 2009; Müller et al., 2010).

2.4 Summary

Eight different modalities namely x-ray, CT, US, NM, PET, MRI, PX and GX are discussed in this chapter which also covers contrast blur and noise characteristics. Appendix C shows several multi-modality medical images that will be used in this research. Table 2.1 shows the summary of technique used advantage and disadvantage of each modality.

Table 2.1: Summarize of Imaging Technique, Advantages and Disadvantages for Multi-Modality Medical Image

Modality	**Imaging Technique**	**Advantage**	**Disadvantage**	**Comments**
x-ray	• transmission-based technique • electromagnetic radiation image	x-ray constructs less noise	limited in viewing the visibility of small objects and structures	several x-ray imaging techniques : angiography, fluoroscopy and mammography
CT	• three-dimensional radiography image of a body structure • Provide one-dimensional projections in different angels	ability of imaging soft tissue objects	High noise effect	Suitable for clinical purposes such as cerebral scans, pulmonary disease detection and abdominal imaging
US	• frequency-based technique	Inexpensive, fast imaging technique and high spatial resolution	poor soft-tissue contrast	Used widely in obstetrics and gynaecology field
NM	• spatial distribution-based of radiopharmaceutical	Used to monitor the development of patient's disease	poor spatial resolution	NM imaging act as strong indicator of disease
PET	• map the biodistribution of positron-emitting radiopharmaceutical in body	measuring the physiology and function within body	expensive medical imaging method	PET is clinically mainly used in oncology, cardiology and neurology for tumors in breast, lung, head and neck.
MRI	• three-dimensional imaging method with soft-tissue contrast and high spatial resolution	provide a series of slices via anatomical area of interest	slower than US or CT scan expensive medical imaging method	Mainly used for assessing brain disease, spinal disorder and musculoskeletal damage
PX	• microscopy image • gross medical image.	High-contrast, high-resolution images Ideal for studying and interpreting thin specimens	Thick specimens can appear distorted Shade-off and halo effect	The capacity to observe living cells and, as such, the ability to examine cells in a natural state
GX	• chart, diagram or graph	Non-complex medical data	Not represent and part of body structure	Used in medical reports

3.0 Review on Text, Content-based and Information Fusion Retrieval in Medical Domain Application

3.1 Introduction

Information retrieval plays a significant role in transmitting knowledge in the forms of both text and images. Concerning the tremendous amount of visual information produced due to its fast growing technology and medical necessity, it is obvious that the development of retrieval systems using traditional text-based is becoming obsolete (Ziegler et. al., 2012). The visual information provides significant information in representing medical knowledge. With the advancement in digital multimedia technology, medical information systems are continuously expanding to combine different types of information sources including text, document, image, graphics, video and hypertext. This chapter gives an overview of text and content-based retrieval studies in general and specific to medical-domain and outlines the significance of retrieval based on information fusion of text and content-based information sources.

3.2 Text-based Retrieval

The main idea of IR in text retrieval which involves the process of representation, storage, organization, and access to information items (Baeza-Yates & Ribeiro-Neto, 1999) is to locate documents that contain terms that users specify in queries. The objective of an IR system is to retrieve information which is related to the query and might be useful to the users. User queries can be ad hoc, which are simple and short query or case-based query which represents lengthy description of scenario. The query must be extracted into a set of keywords which summarize the description of the needed information. However the increasing amount of text information in many application domains such as medicine, digital libraries and the Web, brings new challenges to information retrieval.

Apparently, conventional information retrieval system which uses direct match word is not suitable for specific domain such as medical domain. This is because in conventional IR systems process treats specific words of domain as ordinary terms using general statistical methods (Salton et. al, 1983). Furthermore this traditional IR system treats the variations of specific subjects, synonyms and/or abbreviations as different terms and not related with the specific term (Salton et. al, 1984). Extracting meaningful term from the text by applying term-based information retrieval in particular, plays a significant task towards better understanding of the contents of document collections and can be used for improving the accuracy of processes such as document indexing and retrieval (Drymonas et al., 2010). Term can be defined as words or multi-words expression that contrast with general language words (Quirk et. al, 2012). For example the word “card” in general language can be any card such as poker card, credit card or maybe a birthday card. Nevertheless if “card” defined as a term, in a specific knowledge such as computer; the word “card” can be understood as memory card. Terms are deliberately created for scientific or technical linguistic understanding, specialized concept distinction and classification purposes (Hliaoutakis et al., 2006). Terms can also be related to existing knowledge. The notion of term similarity also has been considered in different ways such as terms may have functional, structural, lexical or other similarities. Using the information access between terms from a corpus is indispensable for improving information extraction, document categorization and information retrieval (Meyer & Mackintosh, 1994).

IR system works as an engine to compare the query with documents in the collection and returns the documents that suit with user's query. Initially, data in document collection is indexed prior to any user query. The measurement of similarity between query and the indexed document is known as retrieval strategies (Sy et. al, 2012). Retrieval strategy is based on the common notion that the more terms are found in both document collection and query, it is considered to be more relevant to the query (Jimeno-Yepes et. al, 2010). The basic algorithm in retrieval strategy consists of a query, a set of document and the measurement of similarity or ranking value also known as retrieval status value (RSV). Techniques that exist in retrieval strategies can be of Boolean model or statistical models such as vector space (Salton et.al, 1975) and probabilistic (Merialdo,1994). Query expansion technique can be used to improve the performance of IR by producing significant keywords which lead to the direction of retrieving more relevant data from the collection. The next section gives more descriptions of retrieval strategies followed by query expansion technique.

3.3 Retrieval Strategies

Retrieval strategy which is a major process in IR systems works as a similarity measurement between a query and document. The retrieval strategies are based on the common notion that the higher occurrence of terms found in document and the query, the more 'relevant' the document is deemed to be to the query (Grossman & Frieder, 2004).

Generally, a retrieval strategy represents an algorithm which involves a query Q, and a set of documents $D_x = (d_1, d_2, \ldots d_x)$ and identifies the similarity $S(Q,D_x)$ for each document d_i, $1 \leq i \leq x$. Conventionally, there are two main models of retrieval strategy methods namely Boolean model and statistical model with vector space and probabilistic models in the statistical model which are described in the following sections where the variables and subscripts are written within context with most of them in standard consistent representations.

3.3.1 Boolean Model

Boolean approach is a simplified retrieval model based on set theory and Boolean algebra. It is based on the following mathematical idea and representation.

Document $D_x = (d_1, d_2, \ldots d_j)$ represent sets of terms $\{k_a, k_b, k_c\}$. Queries are Boolean expression terms $[q = k_a \wedge (k_b \vee \neg k_c)]$ where index terms are AND ($\wedge$), OR ($\vee$), and NOT ($\neg$).

The query is equivalently represented in disjunctive normal form (DNF) as depicted in Figure 3.1 where 1 and 0 refer to existence or non-existence of $\{k_a, k_b, k_c\}$ terms respectively. The similarity measurement between the document d_j and query q can be defined as in equation (3.1) where 1 equals to relevant document and 0 is a non-relevant document and c represent the term match.

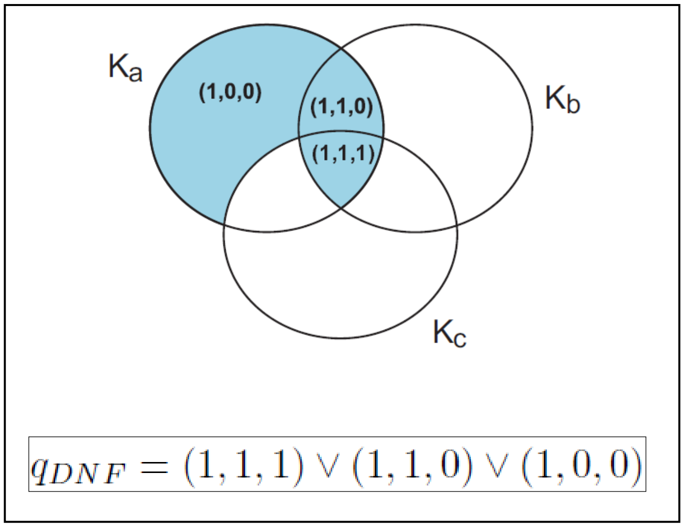


Figure 3. 1: Boolean expression query representation

$$sim\left(d_j, q\right) = \begin{cases} 1\ if\ \exists c(q) \mid c(q) = c\left(d_j\right) \\ \quad 0\ otherwise \end{cases} \tag{3.1}$$

Boolean approach offers several advantages in IR. The algorithm is easy to implement and it is computationally efficient (William, 1992). Also it enables users to express structural and conceptual constraints to describe important linguistic features (Marcus, 1991). Professionals in specific domain such as economic (Alexander, 2003) and biomedical (Darabos, 2011) prefer to use Boolean model since the results are precise which means document either matches the query or not. This offers users greater control and transparency over what is retrieved. Furthermore, the Boolean approach possesses a great expressive power in its representation and clarity in interpretation.

However the drawback of this model is that its retrieval strategy is based on a binary decision criterion (i.e. a document is predicted to be either relevant or non-relevant) without any notion of grading scale. Furthermore, using Boolean model, the result obtained is very rigid and difficult to be ranked according to its relevancy to the query since all matched documents logically satisfy the query (Lucarella, 1988) in the sense of its algorithm without relevance weightage.

3.3.2 Vector Space Model (VSM)

Vector space model (VSM) is an algebraic model for representing text documents (and any objects, in general) as vectors of identifiers, such as, index terms (Salton et al., 1975). VSM involves vector presentation of document and query; which document in $D_x = (d_1, d_2, \dots d_j)$ is conceptually represented by a vector of terms frequencies $d_j = (t_{j1}, t_{j2} \dots . t_{jk})$ where t_{ji} $(1 \leq i \leq k)$ is a non-negative value denoting the single or multiple occurrences of i th term (or can also be written as term i) in documentd_j. The terms extracted from the document, are associated with weights representing the importance of the terms in the document and within the whole document collection. Thus, each unique term in the document collection corresponds to a dimension in the space. Similarly, a query, Q is modelled as a list of terms frequencies $(t_{q1}, t_{q2} \dots . t_{qm})$ where t_{qi} $(1 \leq i \leq m)$ is a non-negative value denoting the number of occurrences of i th term in the query, denoting the weights and reflecting importance of the terms in the query. A common approach to represent terms weight for documents is inverse document frequency (idf) method.

Precisely, the weight, d_{ji} of a term i in document j is depicted as $d_{ji} = tf_{j,i} \times idf_{\ i}$ which also equals to $d_{ji} = tf_{j,i} \times \log \frac{N}{df_{\ i}}$ where $tf_{j,i}$ is term i occurrences in the document j and $df_{\ i}$ is number of documents that contain term i and N is the number of documents in the document collection.

Once the term weights are determined, similarity measurement between query and document vectors is required to rank the document. Precisely, the similarity between D_j and Q is defined as a simple coefficient (SC) represented by the dot product of two vectors as in equation (3.2)

$$SC(Q, D_j) = \sum_{i=1}^{m} t_{qi} \, X \, d_{ji} \tag{3.2}$$

There are also several other techniques used based on VSM. Among the popular approaches for computation of similarity based on the statistical representation of the documents are (i) cosine similarity that measures the similarity between two vectors by measuring the cosine of the angle between them (Tata & Patel, 2007), (ii) Dice coefficient which may be defined as twice the shared information (intersection) over the sum of cardinalities (Mihalcea et al., 2006), (iii) $tfidf$ which is used to evaluate how important a word is to a document in a collection or corpus based on the number of the word's occurrence in the documents (Hiemstra, 2000).

Recently VSM has been used in many medical retrieval models (Alexandros, 2010; Díaz-Galiano, 2010; Hong Wu, 2010) since it allows easy ranking based on weight given for each term in the data. VSM is also a simplified direct mathematical-based approach that provides partial matching and efficient for large document collections.

Nevertheless, there are also several issues related to VSM which include missing semantic and syntactic information such as phrase structure and proximity information (Sharef & Madzin, 2012). At the matching step the method solely depends on the weight of terms (as in its formulation in this model) which does not represent how important that particular terms in the document. Moreover a weight is computed for every term in the document with the possibility to have zero-valued components to increase (Grossman & Frieder, 2004; Göker & Davies, 2009) and this dissipates the available storage space.

3.3.3 Probabilistic Model

The probabilistic model computes the SC between D_x and Q as the probability that the document will be relevant to the query. The probabilistic model uses formal probability theory and statistics to estimate the probability of relevance document. This approach is used to improve the drawback of VSM retrieval model that ranks retrieved items by similarity measure whose values are not directly interpretable as probability (Robertson & Jones, 1976). The probabilistic model is suitable to use when relevant and non-relevant data are available (Göker & Davies, 2009).

Technically, the term's weight estimation is based on how often the term appears or does not appear in relevant documents and non-relevant documents. Probability Ranking Principle (PRP) is used to determine optimal effectiveness which occurs when documents are ranked based on an estimate of the probability of their relevance to a query (Robertson, 1977). The terms in the query are assigned weights which correspond to the probability. Consider a document d_i with list of terms $(w_1, w_2, \dots, w_t)$. Assume that w_i is the term i that results to d_i being relevant. Weight value for each d_i is based on the probability of relevance for each w_i in a document as stated in (3.3).

$$\frac{P(w_i|rel)}{P(w_i|nonrel)} \qquad (3.3)$$

where rel represents relevant document and $nonrel$ represents non relevant document. Then the probability for each term is combined to compute the final probability that a document is relevant to the query as expressed in (3.4).

$$\sum_{i=1}^{t} log \frac{P(w_i|rel)}{P(w_i|nonrel)} \tag{3.4}$$

Probabilistic model does not require any additional term weighting algorithm to be implemented and the ranking algorithms are completely derived from theory. The disadvantage of probabilistic model is the difficulties to access the information of relevant and non-relevant documents. As such this model is inadequate for web search or typically short queries since long sentence queries are needed to distinguish term presence and absence in documents (Göker & Davies, 2009) for probability computation. In this situation, probability of relevance estimation is of theoretical interest only. Moreover, probabilistic model only defines a partial ranking of the documents and does not allow the user to really control the retrieved set of documents (Hiemstra, 2001).

3.3.4 Language Model

Statistical language models were developed as a general natural language processing tool. Language models were first successfully used for automatic speech recognition at the end of the 1970's. The theory behind the speech recognition models is part of hidden Markov model theory. Recently, hidden Markov models are studied as part multivariate probabilistic models used in statistics, systems engineering, information theory and pattern recognition.

There are two models in statistical language model which are basic retrieval model and statistical translation retrieval model. The basic model defines the system's matching process. The statistical translation model adds statistical to the basic retrieval model to model both the matching process and the query formulation process. Matching is modelled by the generation of a random query from a relevant document and query formulation is modelled by translation of the query into the request (Hiemstra and De Jong 1999).

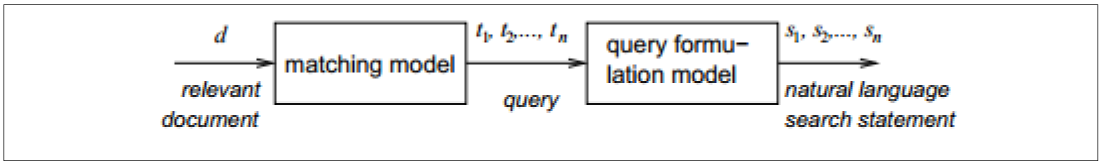


Figure 3. 2: Model of matching and query formulation

Figure 3.2 suggests an information theoretic view of the problem (Miller et al, 1999; Berger and Lafferty, 1999). Information theory was developed by Shannon (1948) to model the problem of decoding a message that is sent over a noisy communication channel. From this viewpoint, a relevant document *d* gets 'corrupted' into a query $t_1 \ldots t_n$ by sending it through a noisy channel, and the query gets again corrupted into a request $s_1 \ldots s_n$ by sending it through a second noisy channel. A natural language information retrieval system can be thought of as a decoding function $f: s_1 \ldots s_n \rightarrow d$ that tries to reproduce the message that was originally sent, that is, to find the document that is relevant to the request. An optimal retrieval system will choose $f(s_1 \ldots s_n)$ such in equation 3.5

$$f(s_1 \ldots s_n) = \underset{d}{\text{argmax}} \; P(D{=}d|S_1{=}s_1, \cdots, S_n{=}s_n) \tag{3.5}$$

By Bayes' rule and because $P(S_1 = s_1, \cdots, S_n = s_n)$ does not depend on *d* as in equations 3.6 and 3.7

$$\underset{d}{\text{argmax}} \; P(S_1 = s_1, \cdots, S_n = s_n, D = d) \tag{3.6}$$

$$\underset{d}{\text{argmax}} \; \sum_{t1, \cdots, tn} P(S_1 = s_1, \cdots, S_n = s_n, T_1 = t_1, \cdots, T_n = t_n, D = d) \tag{3.7}$$

Because there are two independent channels as shown in equation 3.8:

$$\underset{d}{\operatorname{argmax}} \sum_{t1,\cdots,tn} P(S1 = s1,\cdots,Sn = sn \mid T1 = t1,\cdots,Tn = tn, D = d) \\ P(T_1 = t_1,\cdots,T_n = t_n \mid D = d) P(D = d) \tag{3.8}$$

$P(D = d)$ is the prior probability of relevance of the document d, and $P(T_1 = t_1,\cdots,T_n = t_n \mid D = d)$ is the probability of the query given a relevant document. Together $P(D = d)$ and and $P(T_1 = t_1,\cdots,T_n = t_n \mid D = d)$define the matching model. $P(S1 = s1,\cdots,Sn = sn \mid T1 = t1,\cdots,Tn = tn)$ is the probability of the natural language request given the query, which defines the query formulation model. A real life retrieval system does not know these probabilities, but instead defines them by some simple basic principles. A basic principle for the matching model might be that each document has the same probability of being relevant, and that within a document each occurrence of a term has the same probability of ending up in the query. A basic principle for the query formulation model might be that each query term is translated to one and only one word in the request.

3.3.5 Comparison of Retrieval Strategy Models

Boolean model in IR is easy to implement and well-understood due to its clear representation and procedure and since it is exact matching between query and data collection (Salton et. al, 1983). Furthermore, this model provides users control and transparency on the result retrieved (Langer et. al, 2011). However the main disadvantage of Boolean model is that it is too rigid with the result (Das-Gupta, 1988). This leads to many non-relevant documents retrieved. Furthermore, it is difficult to rank the result obtained using Boolean model since it is not numerical weight-based retrieved document (Grossman, 2004).

In statistical model, both probability and vector space models are better at this task because they rank retrieved documents according to the number of occurrences of semantic closeness term to the original query and to the importance of the query terms (Grossman, 2004).. However, these heavy index-based approaches require large indexing storage, demand extra effort and pose sparse data representation (Sharef et. al, 2012). A weight computed for a term in a document vector is non-zero only if the term appears in the document. This is not suitable for a large document collection where the document vectors are likely to contain mostly zeros when there is no term match occurs in the documents (Caudill, 2009). This will lead to generate sparse matrix offline which is time consuming and not necessarily useful when no query regarding the recorded terms are handled (Grossman et. al, 2004). Furthermore, adding a single new document will lead to the changes of document frequencies of term occurrences which changes to vector lengths of every document and it is required to re-rank the document using the new vector value (Göker et. al, 2009).

In probabilistic model, the terms are assumed independent since handling dependencies involves complexities and substantial computation in probability formulation (Heckerman et. al, 1990). Therefore, there is an issue in computing the effectiveness of this method when dependencies are considered and it is required to know the relevant and non-relevant documents of the data collection. Moreover this method is computationally expensive, and difficult to obtain sufficient training data for both relevant and non-relevant data (Grossman et. al, 2004).

3.4 Query Expansion

There is situation where the query from the user is too general or data collection is too big and the user is then faced with information overload (Efron, 2011). In some queries, the best words that describe a relevant document set are too general, and impossible to refine. This will result in a huge list of documents retrieved and even if the documents are ranked, due to practicality, the user will view only the top ranking documents; some relevant results will fall off the bottom of the list. Another limitation of conventional IR methods is that the words provided by the user are often not the same as the one indexed in the document collection. There are still problems with search engines such as word mismatch as the majority of information retrieval systems compare query and document terms with respect to lexical level rather than semantic level (Hazra et. al, 2009). Short, general queries and incompatibility between terms in queries and documents have a great effect on retrieval of relevant documents. Query expansion is an alternative method to solve this issue. In this method, the query is expanded using terms which have similar meaning or bear some relation to those in the query (Goyal et. al, 2012). Applying query expansion approach may increase the chances of matching term in relevant documents. With query expansion, additional terms or phrases with similar meaning or some other relation to the original query are supplemented to improve the retrieval performance.

Typically users will submit queries consisting of two or three words related to the topic of interest (Walker, 2001). However due to huge data collection, it is difficult to find relevant documents just by using short query. Therefore, more robust queries to improve the search features are required. Short queries are poor for recall and precision. This is because they do not consider the variety of terms and tend to be too general to describe a specific topic (Abdelali et. al, 2007). Expanding the original query with additional related term in thesaurus will lead to the increment of recall value (Al-Kabi et. al, 2009). A thesaurus is any “data structure” that defines “semantic relatedness between words.” (Schutze, 1997). There is a significant proof in the literature that reported query expansion as an effective technique in the general information retrieval system (Salton, 1997; Mitra et al., 1998).

Thesaurus offers the most systematic method for offering synonyms, homonyms, and other term relationships. Models that work towards this direction have introduced semantic distance formulas, taxonomy structure information, online thesaurus such as WordNet (Voorhees, 1994) and and Longman’s Subject Codes (Liddy et al., 1993). However, these thesauruses are not typically used in text retrieval systems, since the terms in the thesaurus are too broad to be useful. Domain specific thesaurus constructed by experts in a particular field creates online thesaurus such as Medical Subject Heading (MeSH) and Unified Medical Language System (UMLS) which are described in sections 3.5.1 and 3.5.2. Other studies that adopted ontology and thesaurus in retrieving relevant documents from large volumes of biomedical information are reported (Bodenreider, 2004; Nelson, 2004; Aronson, 2001; Myhre, 2006).

3.4.1 Major Classes of Query Expansion

As query expansion is the process of reformulating the original seed query, the user's input needs to be evaluated and technique in expanding the search query is applied in matching additional documents. There are various ways to expand the original query which involve the type of thesaurus used and the query expansion method. Query expansion either requires user’s involvement or without user’s intervention as described in the following sections.

3.4.1.1 Relevance Feedback

Relevance feedback is the interaction from users to select words or documents that they consider to be relevant for query expansion. Technically the user formulates an initial query, which will be expanded by the system to come out with primary retrieval set. Then the user is required to select from the list which they think relevant to their information need. The result is then used by system to re-weight, expand, and/or reformulate a new query for searching. (Robertson, 1976; Salton & Buckley, 1997) have shown that this approach has significantly improved the retrieval performance in recall and precision. The main drawback using this interactive method is the result produced is not accurate if the user is not from expert background.

Relevance feedback approach in query expansion has been used widely. Generally, relevance feedback approach by Salton's VSM (Salton, 1990) is commonly used although work has also been done using probabilistic model (Van Rijsbergen, 1979; Robertson, 1976) and Boolean system (Salton et al., 1984; Dillon et al., 1983). However, most of these works is more on re-weighting the terms in original query than actually expanding the query. There are several works that used relevance feedback approach in various forms such as using thesaurus in connection with relevance feedback (Efthimiadis, 2001). Although proven to be useful, more research in comparing this method with traditional document-based relevance feedback mechanisms is required. There is also automated relevance feedback approach which avoids user interaction known as pseudo-relevance feedback where the system automatically takes the terms from the primary retrieval list and expands the query (Buckley et al., 1994 ; Xu & Croft, 1996). However, this approach can lead to reduction of effectiveness if non-relevant terms are automatically added to the list of query expansion.

3.4.1.2 Automatic Query Expansion

Automatic query expansion is a fully automated process in query expansion which involves computationally derived thesaurus. Words in query can be ambiguous and development of automatically derived thesaurus in query expansion techniques can improve the retrieval performance.

Semantic relationship approach between words using terms co-occurrence in documents has been developed by Crouch (1990;1992) by grouping together significant terms into categories also known as term co-occurrence measures or term frequency (*tf*). However, this approach does not consider the synonymous terms which generally do not appear within the same document or co-occur between documents. Smeaton (1983) developed an approach of similar perspective but he randomly selected the terms rather than grouping them together. The approach yielded better results than those drawn from term co-occurrence statistics. According to Peat & Willett (1991), term co-occurrence is the sole measures of similarity where terms that share same frequency are considered to be similar. This is true for most co-occurrence analysis formulae, including the cosine and Dice coefficients (Walker, 2001).

Since frequency plays such a significant role in determining the relevance between query and documents; most systems tend to expand queries with terms that already appear frequently in the database. As a result not much improvement has been made in distinguishing the relevant and non-relevant documents, since frequently occurring terms are poor discriminators (Walker, 2001). The limitation from term co-occurrence measures inspired Jing (1994) and Schutze (1997) to consider the proximity of words within a document; which means not simply the frequency of word, but the context in which those words appear.

3.5 Medical-domain Thesaurus Utilization

In recent years, there has been a rapid growth in biomedical literature at a rate up to 7% a year (Lu, 2011). The medical digital collections such as PubMed and Medline Plus (Plus, 2007) provide comprehensive literature and teaching materials for the purpose of functional information for education, medical research and diagnosis of disease. However, the facilities in accessing specialized medical information may create other problem. With the availability of huge medical collection, it is difficult for doctors and other medical professionals in identifying relevant information for specific topic (Pestotnik, 2000).

A solution to this problem is to apply standard IR techniques to the medical domain. However to find relevant information in a large medical collection, specialized knowledge resources from the medical knowledge is required (Zhang et. al, 2002). This clearly seems that combining IR techniques with knowledge from the medical domain using medical knowledge resources is significant to overcome this problem.

Thesaurus or dictionaries have been widely used in textual-based IR systems (Nédellec et. al, 2009). Computational models that work towards this direction have introduced semantic distance formulas (Dar et. al, 1996), taxonomy structure information (Fiedler et. al, 1996), online dictionary and encyclopaedia such as WordNet (Diaz Galiano, 2009), and word sense disambiguation (Stevenson et. al, 2003). The problem with WordNet is that it has a broad coverage which leads to inexactness and ambiguity. Therefore, researchers prefer to use specific domain thesaurus such as MeSH and UMLS. These approaches allow the computation of similarity between conceptually similar which focus on the meaning of the terms but not necessarily lexically similar measurement which concentrates on mathematical technique to identify relationships between the terms.

The utilization of such semantic tool can be an aid in returning more relevant documents and increase the performance of the IR process. This can help to solve the ambiguity from acronym and abbreviation used in the query because sometimes users may use slightly different format in the query which may not exist in the document (Stevenson & Guo, 2010). Acronym and abbreviation may also represent different semantics in different domains.

3.5.1 Medical Subject Heading (MeSH)

MeSH which was created in 1960 is a semantic medical knowledge resource (thesaurus) developed by the National Library of Medicine (NLM) of the United States (Lei Zeng & Mai Chan, 2004). It has been exploited by several researchers in order to bridge the gap between surface linguistic form and meaning. It consists of 26,142 medical subject heading terms naming descriptors in a hierarchical structure that permits searching for various related information at various levels of specificity. MeSH represents subject descriptors or medical terms that usually appear in several sources like the MEDLINE/PubMed or NLM databases.

MeSH is used for indexing and searching for biomedical and health-related information and documents. Each subject descriptor is organized in records. Each record contains MeSHRecords:Descriptors, which consists of more than a single concept element. ‘Concept’ corresponds to a class of terms which are synonymous with each other and

‘Term’ is the synonymous medical term in each concept (Beall & Shephard, 1997). Figure 3.3 shows the MeSH ontology for one subject heading term which starts by <DescriptorRecord>. As a general rule in the MeSH Descriptor structure, each child element inherits the properties of its parent and higher objects.

```
<DescriptorRecord ...><!-- Descriptor  -->
   <DescriptorUI>D000005</DescriptorUI>
   <DescriptorName><String>Abdomen</String></DescriptorName>
   <Annotation> region & abdominal organs...
   </Annotation>
   <ConceptList>
      <Concept PreferredConceptYN="Y"><!-- Concept  -->
          <ConceptUI>M0000005</ConceptUI>
          <ConceptName><String>Abdomen</String></ConceptName>
          <ScopeNote> That portion of the body that lies
          between the thorax and the pelvis.</ScopeNote>
          <TermList>
             <Term ... PrintFlagYN="Y" ... ><!-- Term  -->
                <TermUI>T000012</TermUI>
                <String>Abdomen</String><!-- String = the term itself -->
                <DateCreated>
                   <Year>1999</Year>
                   <Month>01</Month>
                   <Day>01</Day>
                </DateCreated>
             </Term>
             <Term IsPermutedTermYN="Y" LexicalTag="NON">
                 <TermUI>T000012</TermUI>
                 <String>Abdomens</String>
             </Term>
          </TermList>
      </Concept>
   </ConceptList>
</DescriptorRecord>
```

Figure 3.3: Example of MeSH Ontology

Many researchers have used MeSH ontology in their work such as to extract medical terms from biomedical text based on <DescriptorName> (Diaz Galiano, 2009; Jean-pierre Chevallet, 2005; Antani, 2011) and to extract synonymous terms for medical words based on <ConceptName> and <TermList> of each descriptor (Alexandros et al., 2010; Ragia Ibrahim, 2010). More detail on MeSH data structure is on Appendix E.

3.5.2 Unified Medical Language System (UMLS)

MeSH ontology is part of UMLS metathesaurus. UMLS is a standard medical knowledge source developed by the NLM in 1986 (Bodenreider, 2004). This metathesaurus contains of 140 multilingual terminology databases as well as tools for accessing, researching and integrating of biomedical and health information. UMLS has been widely used in many task of information accessing and indexing especially in biomedical domain. In IR, UMLS is used for text-to-concept mapping and indexing, query translation, semantic relations exploration and query expansion (Diem et al., 2007).

UMLS Metathesaurus defines over 800,000 medical terms which contain information concerning concept definitions, their hypernym relations and their context in terminology resources. Semantic relations between concepts in UMLS Metathesaurus can be applied by annotated semantic relation between pairs of concept based on Semantic Network's relation between semantic types (Volk et. al, 2002). The Semantic Network in UMLS provides the essential knowledge structures for deriving classification of concepts into semantic types and establishes relations between them (Lei Zeng et al, 2004). Semantic type is a notion using for the classification of concepts into more general topic. The Semantic Network represents hundreds of *semantic types* such as Disease or Syndrome, Body Part, etc. For example the semantic types of Disease or Syndrome, there are 44,000 concepts available in the UMLS Metathesaurus such as fatty liver, breast cancer, lung cancer, diabetes, etc. Figure 3.4 shows the example of semantic type and its relationship with other entity.

Although UMLS has more biomedical information compared to MeSH and MeSH ontology is a part of UMLS metathesaurus, research done by Díaz-Galiano, (2009) has proved that MeSH provides better result compared to UMLS in medical information retrieval task. Too many information in the UMLS has lead to more general and ambiguous mapping of selecting term in biomedical text. To conclude, it is better to have sufficient amount (although lesser but sufficient) but more specific textual information such as MeSH thesaurus in query expansion technique; having relatively lesser amount but more related and relevant terms.

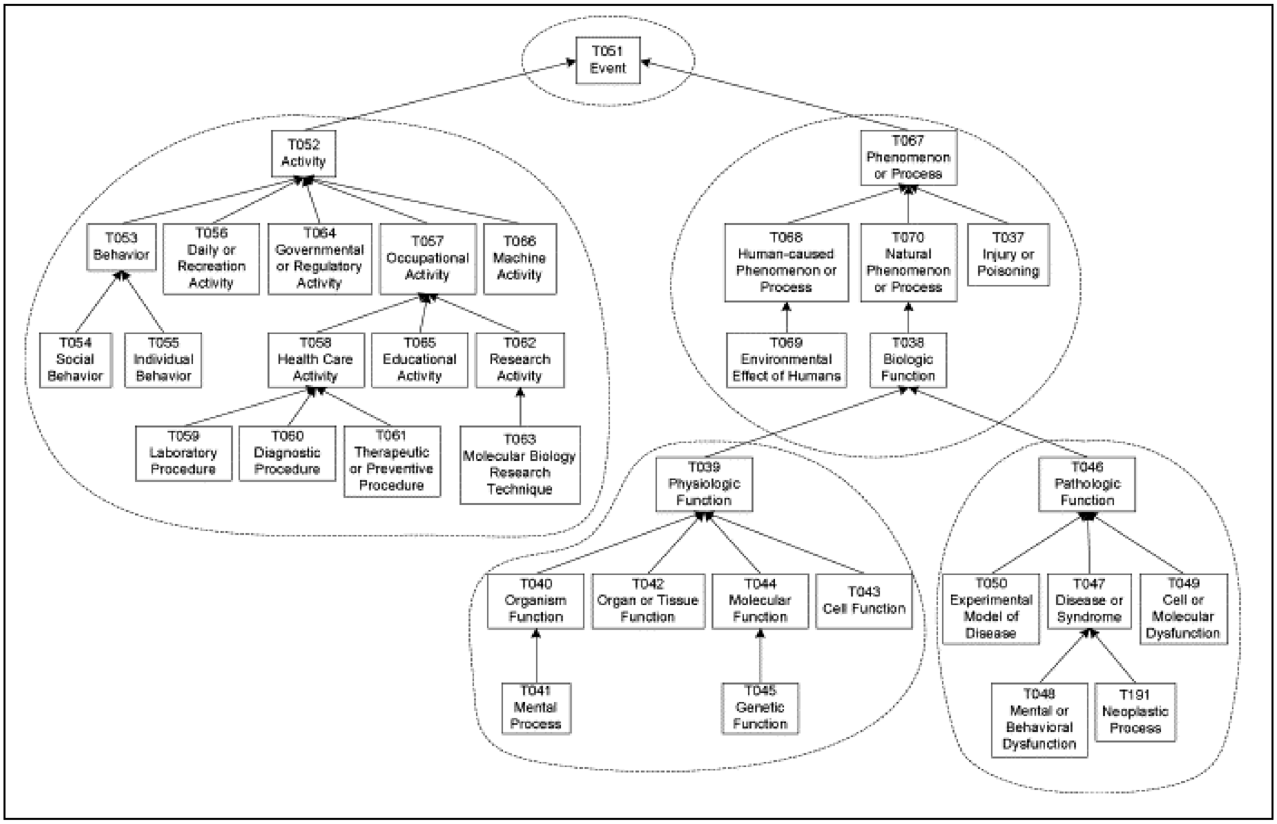


Figure 3.4: Example of Semantic Network in UMLS (Taken from Zhang et al, 2005)

3.6 Text-based Medical Retrieval in ImageCLEF 2010

In this research study, medical data collection from ImageCLEF 2010 (Müller et al., 2010) is used. ImageCLEF is a series of evaluation campaign with four main tasks, namely medical retrieval, photo annotation, robot vision and wikipedia retrieval, involving various areas of interest with some common aspects. It provides an evaluation forum for the cross–language annotation and retrieval of information and images (Clough et. al, 2006). This campaign is motivated by the needs of accessing information from multilingual users

(Peters, 2001). The main goal of ImageCLEF is to support the advancement of the fields of visual media analysis, indexing, classification, and retrieval. ImageCLEF was a part of the Cross Language Evaluation Forum (CLEF) that provides support for the evaluation of (1) language-independent methods for the automatic annotation of images with concepts, (2) multimodal information retrieval methods based on combination of visual and textual features, and (3) multilingual image retrieval methods, so as to compare the effect of retrieval of image annotations and query formulations in several languages. In the process of undertaking this research ImageCLEF2010 was available to implement the theory in various stages and amount. Various research fields such as information retrieval (Baeza-Yates, 1999), cross–lingual information retrieval (Peters, 2001), computer vision and pattern recognition (Zheng et. al, 2007), medical informatics (Pestotnik, 2000), human-computer interaction (Smith, 1993) are involved in ImageCLEF with the participation from both academic and industry research groups worldwide.

The medical retrieval task of ImageCLEF 2010 contains 77,500 medical documents and images where each medical image is associated with short description of the image (medical document) in XML format with 67 average lengths of words. The images in the ImageCLEF 2010 medical task collection are taken from articles published in the important, closely related specific journal, Journal of Radiology and Radiographics including the text of the captions and a link to the html of full text articles. It provides 16 ad-hoc queries with relevance judgment containing relevant documents for each query. This set of relevant documents are assumed as the standard in our experiment and referred to as the relevant set in the remainder of this thesis.

In user-studies, clinicians have indicated that modality (US, x-ray, CT scan, etc) is an important filter that they would like to be able to limit their search by. Therefore it is an advantage if the IR system can support multi-modality for the clinician usage. For example Many image retrieval websites such as Goldminer (Kahn, 2007) and Yottalook (Woojin Kim MD, et al., 2006) allow users to limit the search results to a particular modality. However, this modality of the image which is typically extracted from the caption is often not correct since the result of modality is depend on the text content. The modality of the image can be extracted from the image itself using visual features. Additionally, using modality classification, the search results can be improved significantly (Csurka & Perronnin, 2011).

Various researches have been focusing on the ImageCLEF 2010 medical task. UESTC (Wu et al., 2010) used phrase extraction as indexing term. This is done by extracting the phrases and sub phrase using MetaMap (Aronson, 2006) which is a semantic knowledge representation that maps biomedical text to UMLS Metathesaurus to discover concept and meaning in the query and medical documents. ITI studied by Simpson et al. (2010) also used UMLS to identify UMLS concepts and represent the tf.idf feature vector. The SINAI project (Díaz-Galiano, 2010) utilized MeSH ontology using machine learning for decision making of MeSH term. Ibrahim & Arafa (2010) used sentence selection component to select the most relevant sentence in the caption given. This is done by segmenting the caption based on natural language processing.

The IPL group unified the medical documents with same title and caption into one record tagged by figure index (Stougiannis et.al., 2010). OHSU used Medline term in PubMed website extracted from the PubMed ID in CLEF data and MeSH headings in MetaMap for

MeSH term. OHSU has also adopted modality filter which classifies type of medical images according to categories such as CT scan, x-ray and US (Bedrick & Kalpathy-Cramer, 2010).

The technique developed by ISSR has proved that sentence selection is better than paragraph extraction (Ibrahim & Arafa, 2010). It is also found that paragraph extraction does not improve the MAP score but increases recall. The Bioigenium approached the IR process by translating the query to French using Google Machine Translation System and UMLS thesaurus. However this approach is computational expensive and time consuming and can cause low performance. Finally, Hos-Su VS built new dictionary for identifying XML medical documents and indexed the vector based on normalization value of term frequency (Kalpathy-Cramer & Hersh, 2010).

The main differences between this research study with the above studies as regards to similarity methods with more details stated in subsequent chapters where this research work is described and explained are that (i) this research manipulated MeSH architecture to speed up retrieval process; (ii) a new ranking model called MedHieCon ranking model is developed to associate with Boolean strategy model in overcoming its disadvantage of this model. All previously analyzed techniques mentioned above are mostly based on statistical model such as VSM (Dinh & Tamine, 2010) or probabilistic model such as the Okapi model (Stougiannis et al., 2010) with their model comparisons also covered in earlier sections. Table 3.1 shows the list of retrieval strategies between other studies that use medical task ImageCLEF 2010 dataset with comments given in the last column for further description.

3.7 Content-based Image Retrieval

Generally CBIR is the application of solving problems in searching for digital images in large database. The “content-based” means that the retrieval result is based on the content information of the image themselves rather depend on human-inputted metadata such as captions or keywords. CBIR has been an extremely active research field for the past few years (Antani, Long & Thoma, 2002). The availability of large and steadily growing amounts of visual and multimedia data underline the need to create access method based on image content that offers more than simple text-based query (Müller, 2003). CBIR has been applied in various fields such as personal photos, medical imaging, crime prevention, geographical information, remote sensing systems, education and training (Platt, 2003; Datta et al., 2008; Kekre, 2011; Rahman, 2012). Generally, CBIR is an information retrieval system that extracts features from a query image and compares the features with the information stored in a database to find similar images. It was proposed that an alternative method is invented to overcome the problem of traditional way of text-based retrieval (Rui, 1999). CBIR system makes use of lower-level features such as texture, color and shape to represent the images.

Among the earliest well-known CBIR systems are IBM QBIC (Flickner et al., 1995) which used visual features of color percentages, color layout and texture for image retrieval; VisualSEEK (Smith & Chang, 1997) which employed diagramming spatial arrangements of color regions to form a query and Virage (Gupta, 1996) which was developed for video content based retrieval. In other research, relevance feedback was introduced by Mars (Huang et al., 1999) in CBIR application to improve query formulation by changing the original query. This is done by creating a retrieval loop where retrieval session is divided into few consecutive feedbacks. It was described that the users would give feedback about the retrieval results whether the results are relevant or irrelevant with the queries (Kowalski, 1997).

Table 3.1: Comparison of similarity methods by studies using ImageCLEF 2010 data

Approaches	Similarity and Representation Method	Other Technique/Component Involved	Comments
IPL	TFIDF Okapi BM25 model	MeSH term, Lucene tool	• Okapi BM25 in Lucene tool adds new parameter ∂=0.5 for non-occurrence term which leads to biasedness in result.
OHSU	VSM Boolean model	Index term from MEDLINE article	• Combination of VSM with Boolean increases the complexity in text indexing with inconsistency in weighting. • Other technique that used keywords from MEDLINE may lead to increment of non-relevant documents.
ISSR	Cosine similarity TFIDF	Translate English-French using Google MT, Reverso Dictionary and UMLS thesaurus	• Disadvantage of cosine similarity is biasedness to larger dataset (Vinh, 2010). • Using Google MT and UMLS thesaurus is computational expensive and time consuming with large volume of data.
Bioigenium	TFIDF Latent semantic index (LSI)	–	• Sparse vector representation in LSI leads to increment of storage space (Kontothatis, 2006)
UESTC	VSM	Use phrase and sub-phrase for indexing	• Generate sparse matrix and time consuming • Using phrase indexing lead to precision drop due to irrelevant document retrieved
ITI	TFIDF	UMLS term	• Re-index documents in the collection for every new data insert. • Inefficient and computational expensive.
Hes-So VS	TFIDF	Use bag of words	• Re-index documents in the collection for every new data insert. • The actual meaning of words cannot be captured by word co-occurrence only and high dimensionality of representation space (Elberrichi, 2008)

The "semantic gap" between low-level layout and high-level semantic concepts still remains as the most challenging issue of CBIR (Hiremath et. al, 2007). Semantic gap is the difference between high level language presentation such as natural language and computational language presentation (Gabrilovich, 2009). Therefore there is an attempt to try to close the gap between these two presentations. As to map the semantic information from low-level features for specific domain application is a demanding aspect, therefore, it is significant to apply effective method in indexing and classification in CBIR. At the end of this chapter the application of CBIR in medical domain is described followed by the description of low-level features and classification for multi-modality medical images.

Two main processed involved in the development of CBIR which are feature indexing and image retrieval. In feature indexing, initially the images from the database will be extracted based on low-level features. Later, these features information will be indexed as *m*-dimensional feature vectors in the database. As for image retrieval process, the feature vectors in database are compared with query vector based on similarity measurement (Rubner et al., 2002). Similarity measurement computes the distance between these two feature spaces. The similarity between two images is defined by the shortest distance of feature vector. The results of retrieving relevant images are ranked based on increasing value of distance between query and images in database.

The features of color, texture and shape are used for describing image content. These features can either be at global or local levels. Different combinations have been used for different CBIR systems such as local color and texture features (Carson et al., 2002; Chen 2002; Natsev, 1999; Li, 2000); local properties of texture histograms (Sadineni et al., 2012); color, texture and shape features (Hiremath and Jagadeesh Pujari, 2007); global color and texture features (Stricker, 1995).

3.8 CBIR in Medical Application

Modern computer technology has created the possibility of development of several new medical imaging modalities that use different radiant energy technique to elucidate properties of body tissues. This technology is able to extract significant and accurate information from conventional or tomography radiographic images such as CT, MRI and NM.

The ongoing developments of medical imaging instrumentations and techniques have created an enormous growth in the quantity of data produced including large amount of medical images. Multi-modality of medical images constitutes an important source of medical information such as anatomical and functional information for diagnosis of various diseases, medical research and education. The potential of multi-modality imaging in providing information can be very useful and significant in biomedical research and clinical investigations.

The capabilities of this application field can be extended to provide valuable teaching, training and enhanced image interpretation support by developing techniques supporting the automated archiving and the retrieval images by content. Medical CBIR can also be beneficial for finding other images of the same modality and the same anatomic region of the same disease. There are thousands of multi-modality medical images produced at radiology departments daily (Vaccari & Saccavini, 2006). Hence it is important to extract the structure and content of the medical images in order to increase the effectiveness of image retrieval system for patient care, education and research (Hsu et al., 2009).

Therefore, pertaining to a lot of CBIR systems medical image is cited as a principal domain for content-based retrieval (Müller et al., 2004). Examples of medical CBIR systems are brain MRI retrieval system (Kitson et al., 2009); ASSERT (Shyu et al., 1999) system for lung CT image classification, and IRMA (Lehmann et al., 2003) system for the image classification according to anatomical area. Clinical benefit of medical CBIR is already recognized such as in clinical decision making process (Kulikowski et al., 2002).

There are also several studies using audio features in medical application such as analyzing heart sound using PCG (phonocardiogram) (Khorasani et al., 2011; Babaei & Geranmayeh, 2009) and ECG (electrodiagram) (Vijila et al., 2006; Swarnalatha & Prasad, 2010). However sound feature can only be used in specific type of disease such as heart disease and lung problem. Furthermore the information from sound feature can only be represented in certain modality of medical image such as US. Since this research is focussed on multi-modality medical images, various anatomy and types of diseases; sound feature is not used in feature extraction due to its non-existence or scarcity in medical data compared to image.

3.9 Visual Features

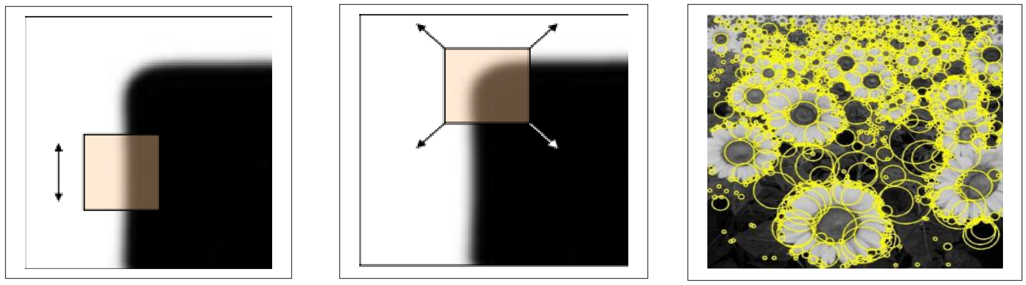
In CBIR systems color, texture and shape information have been the primitive image descriptors with their features used for describing image content. The low-level features are derived from properties and patterns of pixels of images. According to Müller (2003), in order to bridge up the semantic gap, the CBIR application must be more specialized; selection of visual features is very significant to represent images in CBIR index for effective retrieval. Generally, visual features can be extracted in two forms which are global and local.

3.9.1 Global Descriptor

Global presentation is defined as extracting visual features of an image as a whole (Zhao, 2002). For example, extracting color feature of the whole image to measure the percentage of different colors used in the image. Conventionally, global representation is used to extract shape and color features that provide overall information of the particular image but not in detail (Lehman et al., 2005). Tristan (2004) explained that the advantage of global representation is high speed execution of extraction and matching process which means that using global descriptor can improved time performance in IR system. Furthermore it easy to matched the image based on the whole image compared to segment or patches images.

3.9.2 Local Descriptor

A local feature is an image pattern which has different value or characteristic from its instantaneous neighbourhood and associated with a change of image properties (intensity, color and texture) simultaneously (Tuytelaars & Mikolajczyk, 2008). The use of local features instead of global features allows estimating more complicated transformations between images. Invariant feature is a value that remains unchanged when a transformation (object’s position and/or orientation changes) is applied (Madzin, 2009). Therefore, local invariant feature is a new image representation that allows describing the objects/parts in any different transformation (translation, rotation and scale). The features can be extracted from points, edges or small image patches (Tuytelaars & Mikolajczyk, 2008). In local invariant features, the term detector refers to a tool that extracts the features from the image (Tinne, 2007). The detector can be based on corner, blob or edge detector.



(a) "edge" no change along the edge direction

(b) "corner" significant change in all directions

(c) "blob" independently detect corresponding regions in scaled versions of the same image

Figure 3.5: Example of (a) edge, (b) corner and (c) blob. Taken from Mikolajczyk et al (2002)

The approach of local color and texture features segments the image into regions which are close to human perception (Carson et al., 2002; Chen 2002; Natsev, 1999; Li, 2000) and are used as the basic building blocks for feature computation and similarity measurement. Such systems which are known as region based image retrieval (RBIR) systems have proven to be more efficient in retrieval performance. The advantages of local presentation are its robustness with respect to noise, variability in object shape and partial occlusions (Setia et al., 2008). Therefore, it is suitable for selecting interesting location in the image and to speed up analysis since there is no need of segmentation. Furthermore segmentation itself is still faced with difficulties of unsolved problem in medical imaging research field. Local presentations extract the features in more detail compared to global presentation (Datta et al., 2005) and potential to produce a good result in classification (Paredes et al., 2007).

3.9.3 Texture

Texture describes repetition of visual patterns over a region in an image and has the properties of homogeneity. It contains vital information about the surface structure arrangement such as woods, clouds, bricks, fabrics, etc (Acharya & Ray, 2005). Texture feature generally captures the information of image characteristic with respect to the changes in certain direction and scale of the images. This information gives benefit for regions or images with homogeneous texture. Among popular texture descriptor methods that have been used for medical image indexing and retrieval are co-occurrence matrices (Kuo et al., 2002), wavelets (Pizurica, 2003; Chen & Tseng, 2007) and Fourier transform (Sumanaweera & Liu, 2005).

3.9.4 Shape

Shape feature can be described as visual information that is based on two classes: contour and region of an image. Contour shape-based technique in global level has been introduced (Sumanaweera & Liu, 2005). The application is easy to compute and robust to noise but with limited discriminatory power. The study of shape-based spectral descriptor of Fourier and wavelet has been performed by Zhang & Lu (2003). Contour shape-based technique has advantage of robustness to noise and easy normalization, but not inherently in rotations invariance. In a region-based technique, moment-based shape feature provides a numerical shape-preserving representation that is invariant to translation, rotation and scale (Zhu et al., 2002). Moment describes the image content with respect to its axes. The purpose of extracting visual features is to measure the similarity between two images. However, images that depict the same object usually have the same criteria under transformations such as translations, rotations and scaling. Moment-based shape features is an alternative for characterizing an image with arbitrary accuracy.

3.9.5 Color

Color feature can be the most effective feature for systems that employ colored images. The color descriptors consist of histogram descriptor, a dominant color descriptor, and a color layout descriptor (Manjunath et al., 2001). Color is widely used for characterizing image and the color spaces are mostly quantified by using the red, green, blue (RGB) components; the hue, saturation, value (HSV) and the hue, lightness, saturation (HLS). HSV and HLS have been found to be more effective in measuring the color similarity between two images (Sural et al., 2002). The color indexing was initially demonstrated by Swain & Ballard (1991), using color histogram in representing consumer images. This method was commonly used in representing color index. Furthermore color can also be represented in color coherence vector (Pass et al., 1997) and color correlogram (Huang et al., 1999). In medical imaging, involvement of color is very limited since most of medical images are in greyscale (Deng, 2009).

Table 3.2 shows studies of content-based retrieval of medical image application done by other researchers. Not all studies used local descriptor in extracting visual features for medical images. In addition, majority of the studies concentrated on a particular modality. Comments are given in the last column for comparison.

Table 3.2: Visual methods in Medical Application

Researcher	**Descriptor**		**Visual Feature**			**Modality**	**Comments**
	Global	**Local**	**Color**	**Shape**	**Texture**		
Cootes (2001)	√			√		MRI (brain)	• Concentrated only on one modality and on brain • Less detail in extracting features • Used only global shape
Robert (2002)		√			√	X-ray	• Concentrated only on one modality • Extraction based only on one feature **(**texture) at local level • Missing overall information
Aleksandra (2002)	√		√		√	Multimodality	• Less detail in extracting features • Comparatively more variation with 2 features in multimodality
Samee (2003)	√			√		X-ray (cervical spine)	• Only one modality • Less detail in extracting features
Lehman (2005)	√				√	Multimodality	• Less detail in extracting features
Jing Liu (2006)	√	√	√	√		Multimodality	• Less detail in extracting features • More work involved with 2 features in multimodality
Lokesh (2008)		√			√	X-ray	• Only one modality • Only one feature at local level
Thomas (2008)	√	√		√	√	X-ray	• Only one modality • Complementary information of overall and detail; Color is not necessary in X-ray
Mueen (2008)	√	√		√	√	X-ray	• Concentrated only on one modality Both information of overall and detail; Color is not necessary in X-ray}
Zheng (2008)		√		√		CT scan (Neuroimage)	• Only one modality • Only one feature • Less detail in extracting features

3.10 Image Classification

There is a semantic gap between high-level and low-level features (described in section 3.7). In order to minimize this limitation and motivated by the successful use of machine learning in IR, classification-based medical image retrieval method can be applied.

The challenge of image classification is to label an image with appropriate identifiers using visual features (Wang, 2009). These identifiers are determined by the area of interest (McRoberts, 2002), whether it is general classification for arbitrary pictures (for instance from the internet) (Qiu Chen, 1994), or a specific domain (Choi et al, 2009), for instance medical x-ray images or geographical images of terrain. Image classification is also an active area in the field of machine learning, in which it uses algorithms that set map of input (Kumar et. al, 2001); attributes or variables are placed in feature space where *x*-axis is used to set of labeled classes of *y*-axis (Akbani, 2004). These algorithms are known as classifiers (Kumar et. al, 2001).

There are two approaches in image classification; supervised and unsupervised classification. Supervised classification uses training sets of images to create descriptors for each class. Each image's feature vector inclusive of the information of visual features is stored in the database. The training sets are carefully trained to represent a common visual feature of that particular class. The classifier method then analyzes the training set, generating a descriptor for that particular class based on the common features of the training sets. The descriptor is also able to determine class for new unlabeled images based on the training set analysis result. Supervised image classification is a subset of supervised learning.

With the rapid advances of machine intelligence in recent years, sophisticated machine learning techniques have been used in classification for CBIR system. Machine learning is one of the popular approaches to build multi-class classifiers for CBIR in medical application, for example the Naïve Bayes (Soria, 2011), k-NN (Lehmann et al., 2005) decision tree (Vlahou, 2003) and SVM (Takeuchi & Collier, 2005).

Naive Bayes is a probabilistic-based classifier which calculates the probability of a image belonging to a particular class, P(class|image) = P(class)×P(image|class)/P(image). Naive-Bayes classifiers are simple, efficient and robust to noise and irrelevant attributes. However the disadvantage of this classifier is that it can't learn interactions between features and this classifier is not suitable for huge size of features and train set (Colas & Brazdil, 2006). Therefore Naive Bayes classifier will be not used in this research since it involved 460-dimensional of feature vectors and 14,537 medical images as train data.

Decision tree is a method to visually and explicitly represent decisions and decision making. Two types of algorithm that based on decision tree namely Classification and Regression Trees (Loh, 2011) and ID3/C4.5 (Quinlan, 1993). The split at each node is based on the feature that gives the maximum information gain. Each leaf node corresponds to a class label. A new example is classified by following a path from the root node to a leaf node, where at each node a test is performed on some feature of that example. The algorithm can naturally handle binary or multiclass classification problems. However, over-sensitivity to the training set, to irrelevant attributes and to noise (Quinlan, 1993) makes decision trees especially unstable. Which means a minor change in one split close to the root will change the whole subtree below. Due to small variations in the training set, the algorithm may choose an attribute which is not truly the best one (Vlahou et. al, 2003; Rokach et. al, 2008).

k-NN algorithm is considered among the oldest non-parametric classification algorithms and quite straightforward where an image is categorized by only looking at the training images that are most similar to it by using some distance measurement such as Euclidean (Langley et al, 1993). This classifier has been widely used in medical (Korn et. al, 1996), agriculture (Bradbury et. al, 2007) and statistical (Salamov et. al, 1995) studies. The k smallest distances are identified, and the most represented class in these k classes is considered the output class label. The value of k is normally determined using a validation set or using cross validation. k-NN classifier is among popular similarity function used in image classification which generally used well-known Euclidean distance or cosine measurement (Bay, 1998). k-NN classifier is suitable for large training data and its robust to noisy in training data although its time consuming (Bhatia, 2010).

.

Given two sets of medical images instances belonging to two categories, SVM seeks a hyper-plane in the feature space that maximizes the margin between the two sets of instances (Cortes, 1995). Support Vector machine are widely explored in the fields of machine learning and pattern classification (Burges, 1998). SVM also has been demonstrates one of the most promising and successful approaches for medical image classification (Mueen et al., 2010). Especially popular in multi-class classification problems where very high-dimensional spaces are the norm (Medlock, 2008). Comparing classification methods above, we have decided to apply k-NN and SVM classifiers for this research study due to the algorithms support multi-class classification and large size of train data.

Unsupervised image classification used unlabelled data and do not rely on a training set. Unsupervised image classification is based on clustering techniques which measure the distance between images and group the images with common features in the same class. This group can then be labeled with different class-identifiers. Nevertheless, this research study deals with supervised classification. As such unsupervised techniques will not be further explored.

3.11 Information Fusion in Medical Retrieval System

The multimodal feature of MIR data creates the importance of information fusion in analysis, indexing and retrieval. Information fusion is a revision of efficient method to transform information from different sources and different points in time into a representation that lead to effective support for human or automated decision making (Boström et al., 2007). Nevertheless in medical domain retrieval system, information fusion is mainly concentrated on the combination of text and visual retrievals (Zhou et al., 2010). Information fusion can be represent as (i) single modality and multiple sensors, (ii) single modality and multiple features, (iii) single modalities and multiple classifiers and (iv) multi modalities (Kludas et.al, 2008). As for multi modalities fusion the application can be done in parallel or hierarchical processing (Kludas et.al, 2008). Parallel processing is multi modalities information sources will be processed simultaneously in multiple parallel system and hierarchical (also known as sequential) processing means information source is processed first at one level and the output of this level will be processed to another level (Sahoo & Choubisa, 2012).

There are two approaches of information fusion that has been applied in MIR systems which are early fusion and late fusion approaches (Kludas et al., 2008). Early fusion approach also known as feature fusion level is conducted as text and visual feature attributes concatenate in one vector to generate one unique feature space (Depeursinge et al., 2012) This approach represents true multimedia where one vector inclusive of all information sources. Furthermore, early fusion approach is based on weighting scheme technique (Müller et al., 2010). However, the major disadvantage is the large dimension of vector which contributes to scatter the homogeneous clusters of instances of same concept (Kludas et al., 2008).

Late fusion approach also known as decision fusion level is defined as combination of output from different sources of information (Kludas et al., 2008). There are several types of late fusion strategies such as reordering and linear combination. Reordering technique is based on reordering of documents to gain final ranking list. For example the textually-retrieved document denotes the final ranking that based on visual score reordering (Yao et al., 2010) and visually-retrieved is based on textual scores reordering (Depeursinge & Müller, 2010).

The common technique used in the late fusion approach is linear combination which combines text and visual scores in a linear combination to obtain the final score. These score are required to normalized and given extra weight on text score since textual retrieval performed better than content-based retrieval (Deselaers et al., 2006). Jarvelin (2008) defined weight for each modality of medical image based on the corpus of classes and Rahman (2012) updated the weight value based on user's relevance feedback.

However the major drawback of using weighting scheme is the weight value is specific to the data they used (Forbes et al, 2005). Different collection may have different value of weight. Therefore it is not flexible and difficult to maintain when using different data collection.

3.12 Information Fusion Systems using ImageCLEF 2010 Medical Task Collection

Various researches have been focusing on the ImageCLEF 2010 medical task. ITI studied by (Simpson, 2009) used UMLS to identify UMLS concepts and represent the tf.idf feature vector for text retrieval and for content-based retrieval; they applied Color Layout Descriptor (CLD) and Edge Histogram Descriptor (EHD). The feature vectors from both sources are combined and trained using SVM classifier. However this approach is time consuming since each feature from text and visual need to be trained by the classifier and the large dimension of feature space. The Bioigenium text-based approached the IR process by translating the query to French using Google Machine Translation System and UMLS thesaurus and used tf.idf weighted scheme as final representation. As for content-based, they used Discrete Cosine Transform (DCT) features (Hare, 2008) which produced 2,000 to 5,000 features. However this approach is computational expensive and time consuming and can cause slow performance. The IPL group unifies medical documents with same title and caption into one record tagged by figure index which represent using binary histogram that consist of 90-dimensional and gray and color intensity histogram are used to index visual features. Both text and visual features are concatenated in one vector using Joint Kernel Equal Contribution (JKEC) (Han & Chen, 2011). OHSU uses the combination of VSM and Boolean retrieval strategies for text-based retrieval and the modality of medical image taken from the text-based retrieval output will be filtered based on relevant feedback by the user. However the drawback of this approach is that not all users' decision is purely accurate. Finally, MedGIFT used standard setting of Lucene tool for text-based retrieval; which is based on VSM strategy and applied color histogram and Gabor filter to extract texture information. The results from both sources will be combined and ranked using rank fusion technique of combSUM and combMNZ (Zhou et al. 2010).

Table 3.3 illustrate the summary of text, content-based and information fusion techniques that used by the mentioned run systems.

Table 3.3: Comparison of techniques for several run systems that used ImageCLEF 2010 medical task collectio

Run System	Text-based Technique	Content-based Technique	Information Fusion technique	Comments
ITI	tf.idf weighted scheme	CLD and EHD	Early fusion	Produce large vector dimension which lead to scatter the cluster of same concept (Kludas, 2008).
Bioigenium	tf.idf weighted scheme	DCT	Late fusion using linear combination	Linear combination required normalizing and assigning weight for text and visual parameter which is not flexible for different data collection (Deselaers, 2006).
IPL	Binary histogram	Gray and color histogram	Early fusion	Produce large vector dimension which lead to scatter the cluster of same concept.
OHSU	Combine VSM and Boolean strategies	Relevant feedback	Late fusion based on hierarchical processing	Execute retrieval from one level to another lead to time consuming.
MedGIFT	VSM strategy using Lucene tool	Color histogram and Gabor filter	Late fusion using linear combination	Linear combination required to normalize and assign weight for text and visual parameter which is not flexible for different data collection

3.13 Summary

This chapter briefly reviews about information retrieval of text and content-based features in general and specific domain of medical. This chapter also discusses about conventional text retrieval models such as Boolean and statistical models. Traditional Boolean model matches the documents with the query based on exact matching. In contrast VSM and probabilistic model are vector-based indexing.

The advantages and disadvantage of these models are also discussed in this chapter. Since this research study focuses on medical specific domain, medical thesaurus is utilized to extract medical terms from the query. Therefore more specific, relevant and meaningful documents will be retrieved in the system. There are many types of medical thesaurus and they have also been discussed in this chapter.

CBIR is allowable to create an access method based on the image content that offer more than simple text-based query. Recently, CBIR has been an active research field especially in specific domain such as medical domain. The availability of large and steadily growing amounts of visual and multimedia data underline the requirement to have CBIR system. This chapter discuss the application of CBIR in medical domain. The visual features of texture, shape and color also been discussed in this chapter. Finally, a brief description of image classification and machine learning technique are presented. Lastly the significance of information fusion in MIR is discussed and several studies applying information fusion technique in ImageCLEF 2010 medical task data are highlight.

4.0 M3IRS Text-based Framework and Module Descriptions

4.1 Introduction

Application in MIR for specific domain such as medical domain using conventional text retrieval model is obsolete (Mueen et. al, 2010). Association between MIR modules with other knowledge source is needed to improve the performance of MIR system in specific domain. This chapter introduces Multi-modality Medical Information Retrieval System (M3IRS) text-based framework for medical MIR. M3IRS text-based framework consists of four main components which are: i) document pre-processor, ii) query processor, iii) retrieval strategies and iv) ranking strategies as depicted in Figure 1.3. The details of each main module are given in the following sections.

4.2 Multi-modality Medical Information Retrieval System (M3IRS) for Text-based Retrieval

Technically, the process starts with medical data collection which is pre-processed before indexing into local repository. User's query will be processed via query expansion technique and will be filtered via Mesh-indexer to produce medical query list as input of the system. The retrieval strategies and ranking strategies components are the main engines for M3IRS retrieval model whereby medical query list will be compared and matched with the indexed data in local repository and produced relevant list based on matching medical terms followed by the computation of number of term frequency. Later, the relevant list is ranked based on MedHieCon model and total of term frequency. The output from M3IRS will be the ranked medical documents with the top list as the most relevant documents to the query given.

4.3 Document Pre-Processor

M3IRS framework can be applied in any medical domain data collection. The component of document pre-processor is used to construct the document to be more structured and standardized.

In this research ImageClef 2010 medical task data is used. The collection includes title and figure caption from the articles and all images including illustration, graph, chart, figure and medical image. Overall there are nearly 77,500 medical documents and images from over 5,600 articles with average length of 67 words for each document. The description of each image in the collection is represented in medical document with XML format. The tags in medical document are defined using Document Type Definition (DTD) as depicted in Figure 4.1. Generally each record of medical document describes the <figureID> which represents ID for the medical image, <figureURL> and <articleURL> representing URL link from Journal of Continuing Medical Education in Radiology (Jeffrey & Klein, 1981) where the <caption> in medical documents and images are taken from that particular publication. <pmid> which represents PubMed identifier or PubMed unique identifier, is a unique number assigned to each PubMed record. PubMed comprises more than 21 million citations for biomedical literature from MEDLINE, life science journals, and online books. Since there can be many unnecessary information in the medical document, it is necessary to index the annotation files into an easily accessible presentation. More detail on ImageCLEF data structure is on Appendix D.

```
<record>
  <figureID>27986</figureID>
  <figureURL>http://radiology.rsnajnls.org/cgi/content/full/210/1/37/F2A</figureURL>
  <caption> &lt;B&gt;Figure 2a. &lt;/B&gt; &lt;B&gt;&lt;/B&gt; Posteroanterior radiographs in
a patient with burns. &lt;B&gt;(a)&lt;/B&gt; At the time of admission, the patient was
dehydrated and the lungs showed oligemia, with fluid out of physiologic control (ie, "third
spacing"). The azygos vein (arrows) is reduced in size. &lt;B&gt;(b)&lt;/B&gt; After hydration,
the circulating BV is restored, and the pulmonary BV is normal (ie, normovolemic). The size of
the azygos vein (arrows) is increased because of an increase in systemic BV. Pulmonary and
systemic BV usually cannot be dissociated. &lt;B&gt</caption>
  <title>Pulmonary oligemia in aortic valve disease</title>
  <pmid>9885584</pmid>
  <articleURL>http://radiology.rsnajnls.org/cgi/content/full/210/1/37</articleURL>
  <imageLocalName>27986.jpg</imageLocalName>
</record>
```


Figure 4.1: Example of ImageCLEF XML medical document.

The <caption> and <title> in medical document represent the description of the image. There are cases that several images are in sequence order taken from the same article. Therefore there are several medical documents that contain same <title> and information <caption> for different images with different <figureID>. This duplication has contributed to low position of relevant documents since not all documents are relevant. Due to that reason, a new approach namely XML Tag-based Extraction (XTE) has been developed. XTE technique is used to create a new more organised simplified version of medical documents in the collection and to avoid duplication in the dataset. There is no need to use other pre-processing technique such as removing stop word and stemming. This is because this framework used Boolean matching which only extract medical terms that match with the query. Hence these techniques will lead to time consuming. Figure 4.2 shows the flow of processes involved in XTE technique.

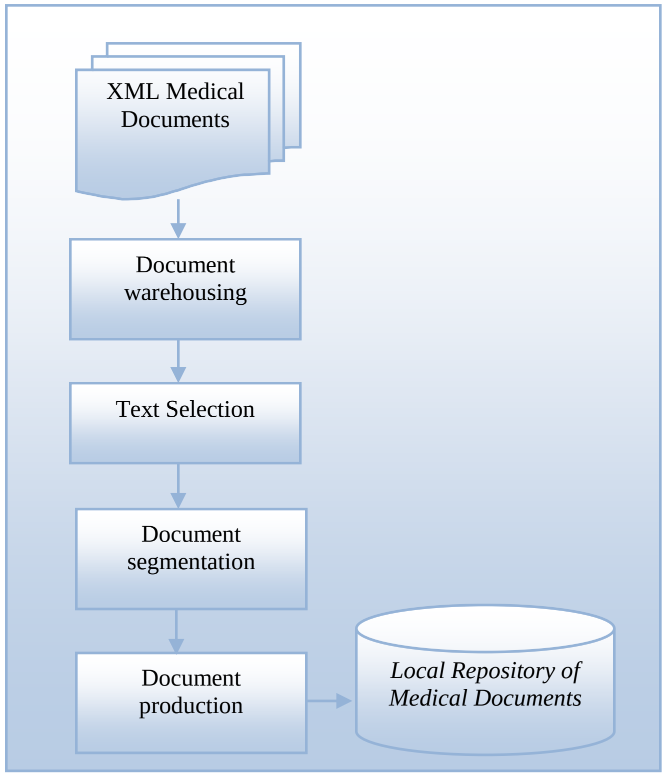


Figure 4.2: Process flow chart of XTE technique

Below are the explanations for each process in XTE technique:

i) **Document warehousing**

In this process the non-relevant tags such as “&lt” and “B&gt” in the <caption> are removed. The text is converted into lower-case.

ii) **Text selection**

The utilization of text selection is to select significant DTD that has meaningful text. In this research paper, DTD of <figureID>, <caption> and <title> are used and to be taken into account as new representation in our new data indexing in local repository.

iii) Document segmentation

Document segmentation process involves filtering sentences in the <caption> based on the figure number. For example in Figure 4.1, the description in caption “<caption> <B>Figure 2a. </B> <B></B> Posteroanterior radiographs in a patient with burns. <B>(a) </B> At the time of admission, the patient was dehydrated and the lungs showed oligemia, with fluid out of physiologic control (ie, "third spacing"). The azygos vein (arrows) is reduced in size. <B>” is dedicated to “Figure 2a”. The (b) description will be discarded.

Therefore, selection is done on (a) description in the caption segment. We also restrict for duplication of description in the <caption> for each medical document. The new description will be tagged as <modifiedCaption>.

iv) Document production

After segmentation, the new XML medical document with DTD of <figureID>, <modifiedCaption> and <title> will be stored in local repository as depicted in Figure 4.3. These new index medical documents are stored in local repository. Storing pre-processed documents using XTE method optimizes search time and increases efficiency of the system.

```
<record>
<figureID>27986</figureID>
      <modifiedCaption> Figure 2a.   Posteroanterior radiographs in a patient with burns. (a)
At the time of admission, the patient was dehydrated and the lungs showed oligemia, with fluid
out of physiologic control (ie, "third spacing"). The azygos vein (arrows) is reduced in size.
</modifiedCaption>
      <title>Pulmonary oligemia in aortic valve disease</title>
</record>
```

Figure 4.3: Example of new XML medical document with new DTD of <modifiedCaption>

4.4 Query Processor

Reformulating the original queries via query expansion may improve retrieval performance by including the terminology that is synonymous. Applying external medical resource (thesaurus) may assist researchers in order to bridge the gap between surface linguistic form and meaning and finding the medical term and synonyms in the query and medical documents. The component of query processor, shown in Figure 1.3 involves text processing, query expansion process and production of new medical query list. Figure 4.4 shows the flow chart for query processor component where the processor involves tokenization for text processing, and processes to extract medical and synonymous terms from the original query. The output from this flow chart will be the query list of Type 2 and Type 3 (details in section 4.4.1). In order to extract medical term from original query, medical-context aware query expansion technique is created to expand the original query. Further detail of this technique is explained in the next section.

4.4.1 Medical Context Aware Query Expansion

We define query expansion as the process of automatically enriching the original query, q_0 submitted by user with medical terms and the synonymous. A set of additional terms suitable to q_0 will then be generated (also called set of expanded queries in this research).

Prior to that, there will be several queries processing such as removing stop words and converting the sentences into lower case alphabets.

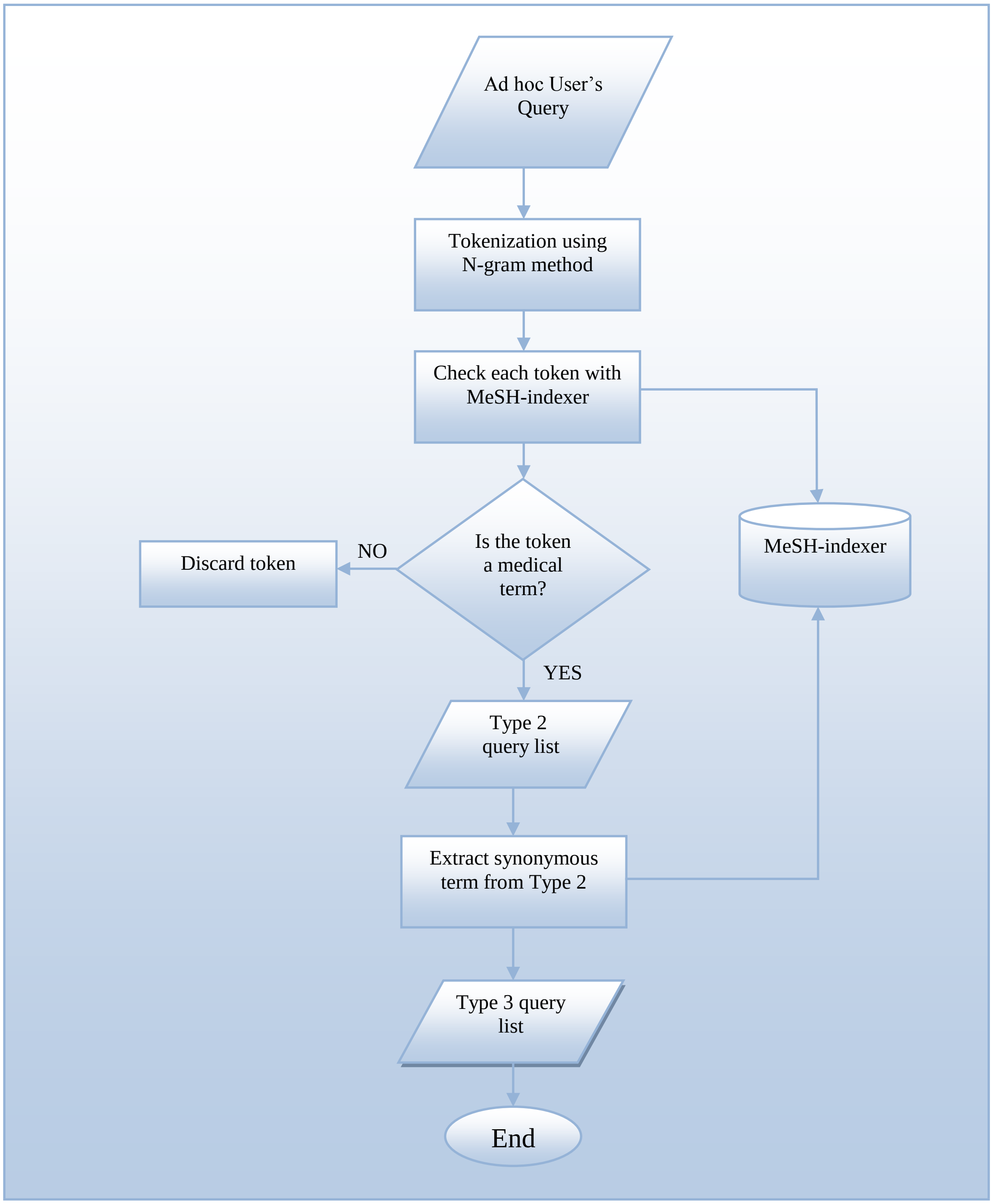


Figure 4.4: Flow chart for Query Processor Component

The query expansion process can be divided into 3 types:

- Type 1: Exact match with the original query.
- Type 2: Expansion of original query.
- Type 3: Combination of expansion and MeSH-enriched query

Retrieval using query generated from query Type 1 is based on standard Boolean matching by checking the occurrence of q_0 in the documents. Typically, it is possible to retrieve relevant medical documents by using only q_0. However this traditional method was found to possess many limitations as it only returns results that contain exact matching.

Meanwhile, Type 2 process is a process that parse the query into tokens which are automatically generated by using n-gram method (Cavnar & Trenkle, 1994) where n=[1, length of q_0] where length of q_0 represents number of words content in q_0. For example; given q_0="CT images containing fatty liver", the length of q_0 = 5 and the set of n-gram is:

1-gram**:** *CT, images, containing, fatty, liver*

2-grams: *CT images, images containing, containing fatty, fatty liver*

3-grams: *CT images containing, images containing fatty, containing fatty liver*

4-grams: *CT images containing fatty, images containing fatty liver*

However, not all words produced from n-gram process can be used and have significant meaning. For example, the words "images containing" and "containing fatty" in 2-grams tokenization are not useful and meaningless. This leads to the issue to identify significant terms in the query. Since M3IRS is a medical-domain retrieval system, mapping the words into medical terms using external medical resource such as medical thesaurus is required. To solve this problem the MeSH-Medical Descriptor (MMD) from Mesh-indexer (explanation in Section 4.4.2) is utilized to filter out words that contain medical terms. In previous example of $q_{0,}$ the set of filtered queries are “CT”, “fatty liver” and “liver”. For retrieval process, queries generated using Type 2 query list will be matched based on Boolean function (see section 4.5).

The query expansion in Type 3 performs query enrichment with the synonymous terms recorded in the MeSH-Synonym Terms (MST) folder (explanation in Section 4.4.2). Based on Type 2 query list, only “liver” and “fatty liver” terms have synonym which are “liver steatosis” and “steatohepatitis”. Type 3 query list will be the combination of expansion (Type 2 query list) and enrichment of synonymous term. Therefore, the expanded queries generated in Type 3 are “ct”, “fatty liver”, “liver”, “liver steatosis” and “steatohepatitis”.

Figure 4.5 shows the system execution of medical-context aware query expansion technique. Based on Mesh-indexer, the system will extract the medical terms. The final step in query processor component is to identify the synonymous terms for each medical term extracted and the output from this component is the query list of Type 3. The new query list will be used as input for M3IRS search engine in searching for relevant documents from the local repository.

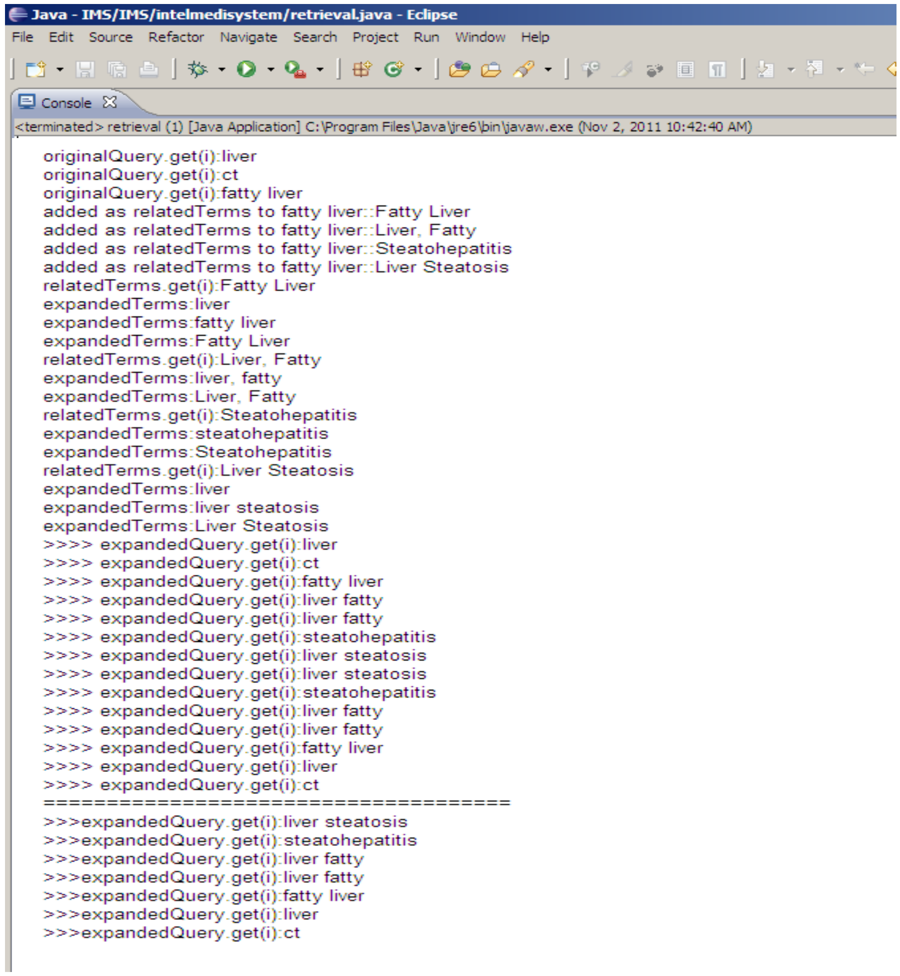


Figure 4.5: Example of running system to execute medical terms expansion and synonymous enrichment from original query

4.4.2 Mesh-Indexer

MeSH thesaurus is heavily used in the query expansion in identifying medical terms and any synonymous terms. Moreover not all information in the MeSH are used for this research. Therefore, MeSH-indexer offline has been created to reduce searching time and increase the performance in terms of extracting and finding the medical terms in query and medical documents. The indexer is created by manipulating and filtering the MeSH thesaurus as an effort for more efficient thesaurus reference. The ontology structure of MeSH thesaurus is as in Figure 3.2.

Mesh-indexer contains three folders:

i. **MeSH-Medical Descriptor (MMD) folder**. Each file contains medical terms that start with the same alphabet. This medical term is taken from <DescriptorName> of each <DescriptorRecord> in MeSH thesaurus.

ii. **MeSH-Synonym Terms (MST) folder**. Each file contains a set of data in the form of medical descriptor <DescriptorName> and list of concepts <ConceptName> and terms <TermList> (synonymous terms) as shown in Figure 4.7.

iii. **Medical Conceptual (MC) folder**. This folder contains semantic type <SemanticType> and conceptual files (details in Table 4.5) for each medical descriptor, concepts and terms as shown in Figure 4.8.

These three folders are organized according to alphabetical order as depicted in Figure 4.6. Therefore it will speed up the search between data and information in MeSH by only looking at the initial letter of the term.

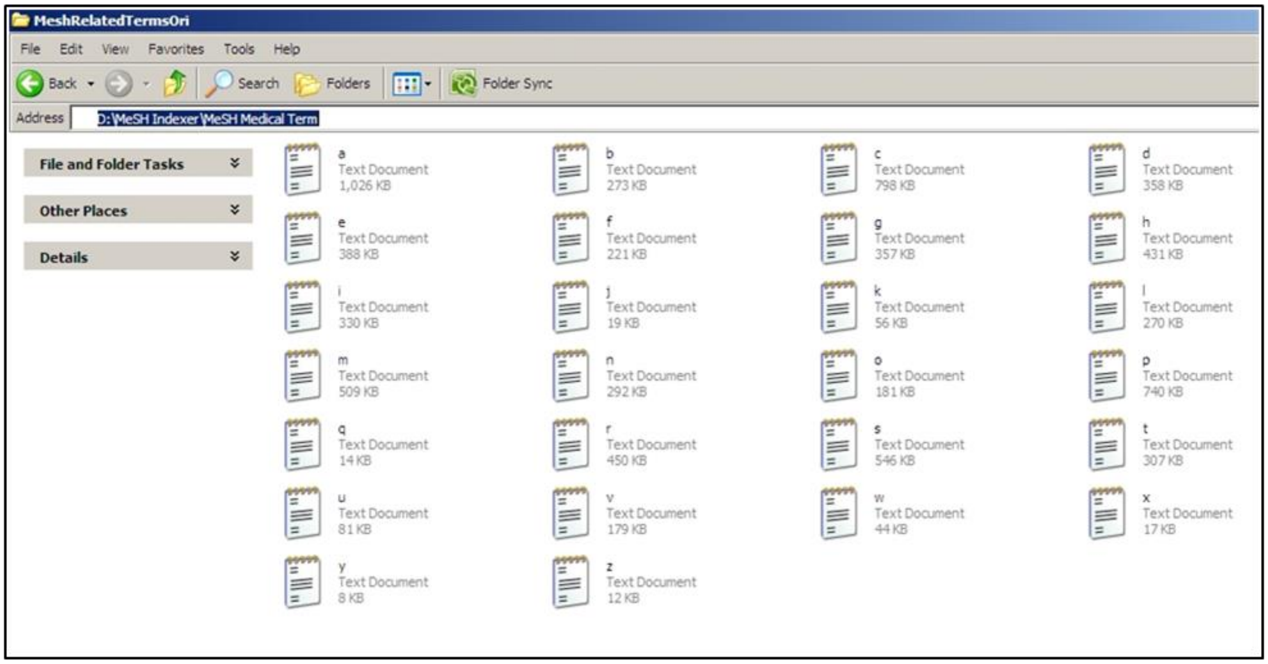


Figure 4.6: MeSH-indexer folders organized in alphabetical order

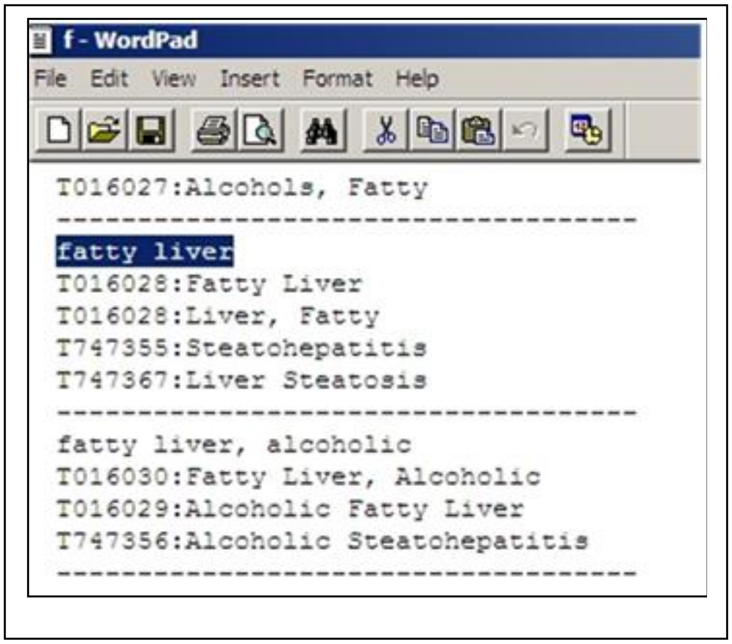


Figure 4.7: MeSH Synonym Terms (MST)

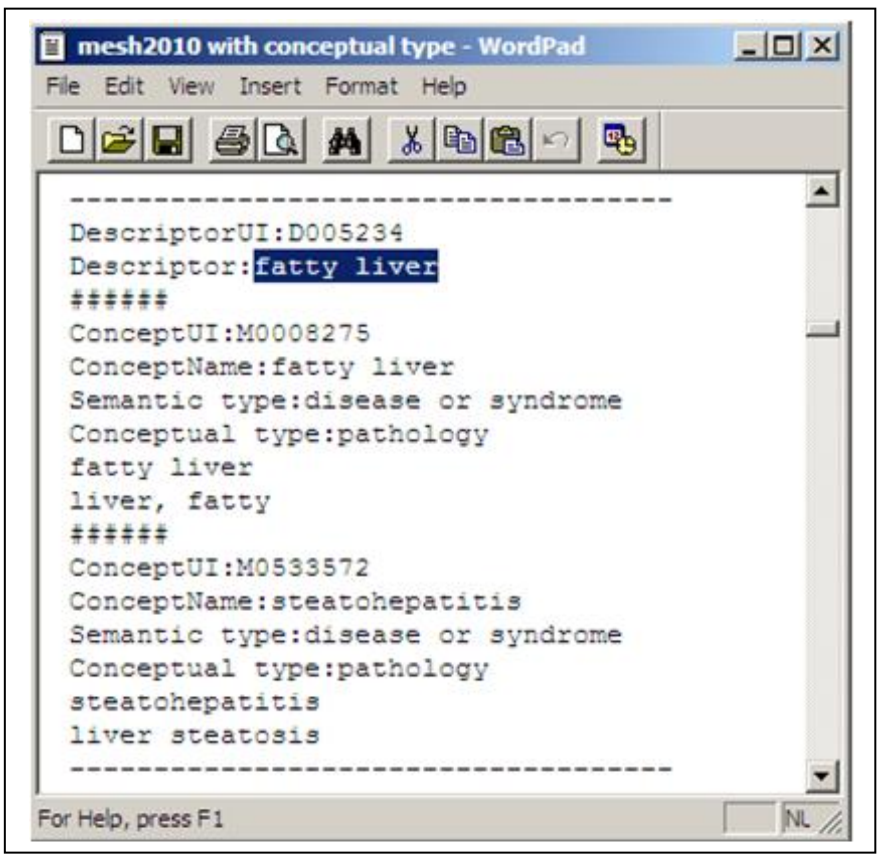


Figure 4.8: MeSH Medical Conceptual (MC) folder

4.5 Retrieval Strategies

Figure 4.9 represent the flow chart for retrieval strategy component and Table 4.1 represent Execute_Matching() algorithm for pre-process medical documents prior matching between query list and medical documents in local repository. Retrieval function is execution of finding matches in medical documents *(MR)* from the expanded query *(EQ)* in order to retrieve relevant medical documents *(RMR)*. *EQ* list is obtained from the output of query processor component and represented as the set of descending-length sorted expanded queries, $EQ= [eq_1, eq_2, ..., eq_j, ..., eq_m]$. The reason to sort these expanded queries are to give more priority matching to longer terms. Using the example from previous section (see section 4.4.1), the *EQ* list for original query (q_0) of “ct images containing fatty liver” will be eq_1 = “liver steatosis”, eq_2= “fatty liver”, eq_3 = “steatohepatitis”, eq_4= “liver” and eq_5= “ct”

Next, the system will look for matches in each of eq_j in every *MR* in the local repository which is represented as mr_i at index *i*. Prior to that *MR* in the local repository are parsed, tokenized and normalized by folding upper-case letters to lower-case.

Every match of expanded queries, eq_j in the medical documents, mr_i is represented as *meq* and the occurrence number of *meq* is counted and labeled as *meq_hit*. A set of *meq* in each medical document, mr_i is labeled as *MEQ*. The total of *meq_hit* is represented as *MEQHit.*

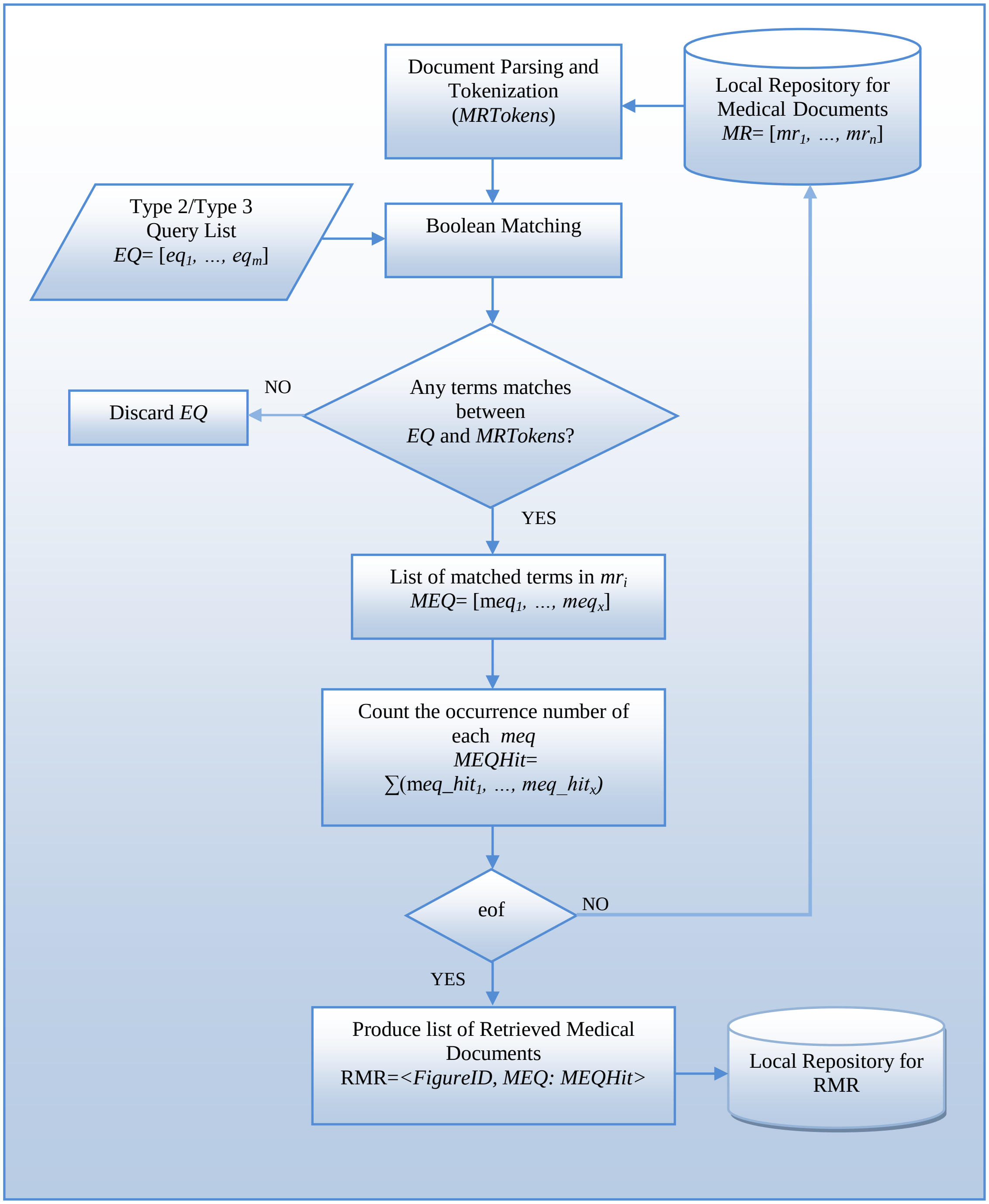


Figure 4.9: Flow chart for Retrieval Strategy Component

The Execute_Matching() function concentrates on the matches of *eq* in the $mr_{i..}$ In this function sentences in mr_i need to be parsed and tokenized using n-gram method which is represented as *MRTokens*. Next the matching is done by comparing eq_j with the *MRTokens* with minimum length equal to 1 and the maximum value is number of words in the query.

The technique of tokenization has been adapted from the fuzzy grammar based text fragment extraction method (Sharef, Martin & Shen, 2008). As example, for multiple matching where meq_1 = "liver" (bolded font) and meq_2 = "fatty liver" (bolded font), the system will take the longest contiguous tokenized *MRTokens* as the best match which is eq_2 as shown in Table 4.2. The output from Execute_Matching() algorithm is the occurrence number of meq_x in mr_i and the results are updated to *MEQ* and *MEQHit* .

Table 4.1: Algorithm to match query list with medical documents using Boolean method

Algorithm 2: Execute_Matching()

Input: Set of medical documents *(mr)*, expanded query *(eq)*

Output: Hit, number of matched *eq* in *mr*

Process:

1. Set hit=0
2. Set QL=number of words in *eq*
3. Set MRTokens=set of tokenized *mr* based [1,*QL*]
4. FOR each item x, in MRTokens
    a. IF MRTokens[x] == eq THEN hit++
    b. END IF
5. END FOR

Table 4.2: Example of best match expanded query based on longest contiguous tokenized

Expanded Queries *(EQ):* Fatty Liver, Liver Retrieved Medical Documents *(RMR): … Focal spared area in the fatty liver along the porta hepatis …*	
eq_1 = "liver"	*… Focal spared area in the fatty **liver** along the porta hepatis …*
eq_2 = "fatty liver"	*… Focal spared area in the **fatty liver** along the porta hepatis …*
Best match: *… Focal spared area in the **fatty liver** along the porta hepatis …*	

The flow chart of retrieval strategy component in Figure 4.9 presents the process of retrieving the relevant medical documents, *RMR* that match with expanded queries, *EQ* in a formal way. The output from retrieval strategies component is a list of <*FigureID, MEQ: MEQHit*> where the output will be printed out in Excel format inclusive of Figure ID, Title, Caption, total number of medical terms that match (*MEQHit*) and the list of medical term (*meq)* followed by number of occurrence of each medical term (*meq_hit)* as depicted in Figure 4.10.

Figure ID	Title	Caption	*MEQHit*			
28055	hepatic lesions	figure 7d hepatocellular carcin	3	liver:3		
28080	obstruive sleep	figure 1 lateral scout view show	1	fatty liver:1		
36507	nondiffuse fatty	figure 2a patient 5 focal spared	6	fatty liver:2	liver:3	ct:1
37377	value of iterative	figure 2a images in a 59-year-ol	13	liver:9	ct:4	
106151	recurrent hepat	figure 1d transverse contrast-en	3	liver:2	ct:1	
106311	littoral cell angi	figure 3a patient 7 littoral cell an	2	steatohepatitis:1	ct:1	
112963	metastases in	figure 7a false-negative lesion or	5	fatty liver:1	ct:3	liver:1
177922	atrophy-hypertr	figure 2a pv in 70-year-old woma	5	liver:2	ct:3	
180378	imaging patterns	figure 14adifferentiation of aden	5	fatty liver:2	liver:3	

Figure 4.10: Example of output from retrieval strategies

The algorithm utilizes standard Boolean model which based on binary matching and associated with *MEQHit.* Therefore this retrieval strategy has the advantage of Boolean model's strength while taking care of its weakness overcome by other components and stages for system efficiency and effectiveness including data cleansing and re-structuring in document pre-processor, medical context aware query expansion and MeSH-indexing. The issue is also handled by the ranking strategy described in the next section.

This justifies why vector space and weighting model is not necessary in this research. Specifically, we argue the performance of the methods which usually generate sparse matrix offline which is time consuming and not necessarily useful when no query regarding the recorded terms are handled. Moreover, these traditional methods have not supported for semantic understanding of the query. Nor do they have easy adaptation when new document is available in the collection as the term frequency (*tf*) vector is changing and needs to be reconstructed. By using M3IRS text-based framework, MeSH-indexer is used in order to supply domain knowledge in the retrieval process to allow semantic understanding and it is flexible and accommodating even with many stages of new document or thesaurus are fed, Therefore, this framework applies practical simplistic process and minimal overhead of the retrieval process is involved.

4.6 Ranking Strategies

The main disadvantage of using Boolean model is that it does not provide a ranking of retrieved documents (Hiemstra, 2009). This requires having a mechanism for determining a document score, which encapsulates how good the match of a document is for a query. For this research a new ranking method is proposed and is discussed in the next section.

The ranking strategy component is a continuing process from the retrieval strategy component. The *RMR* output which consists of medical documents that are matched between medical query list and documents in the local repository is obtained from retrieval strategies component. In order to know which *RMR* in the list is more relevant to the query, ranking mechanism needs to be applied. Therefore, a ranking model is created to rank the *RMR* list in descending order whereby the most relevant document will be on the top of the list. This section introduces two ranking models, which are comprehensive ranking model and medical hierarchical conceptual (MedHieCon) ranking model. Both models are explained in detail in the next section.

4.6.1 Comprehensive Ranking Model

Comprehensive ranking model ranks the documents based on the size of term matched in *RMR*. As previously mentioned, *RMR* list consists of Figure ID, Title, Caption, *MEQHit* and *meq* followed by *meq_hit*. We assume that, the more terms that match between query and document, the more relevant the document is deemed to be to the query. Figure 4.11 shows the flow chart of the comprehensive ranking model.

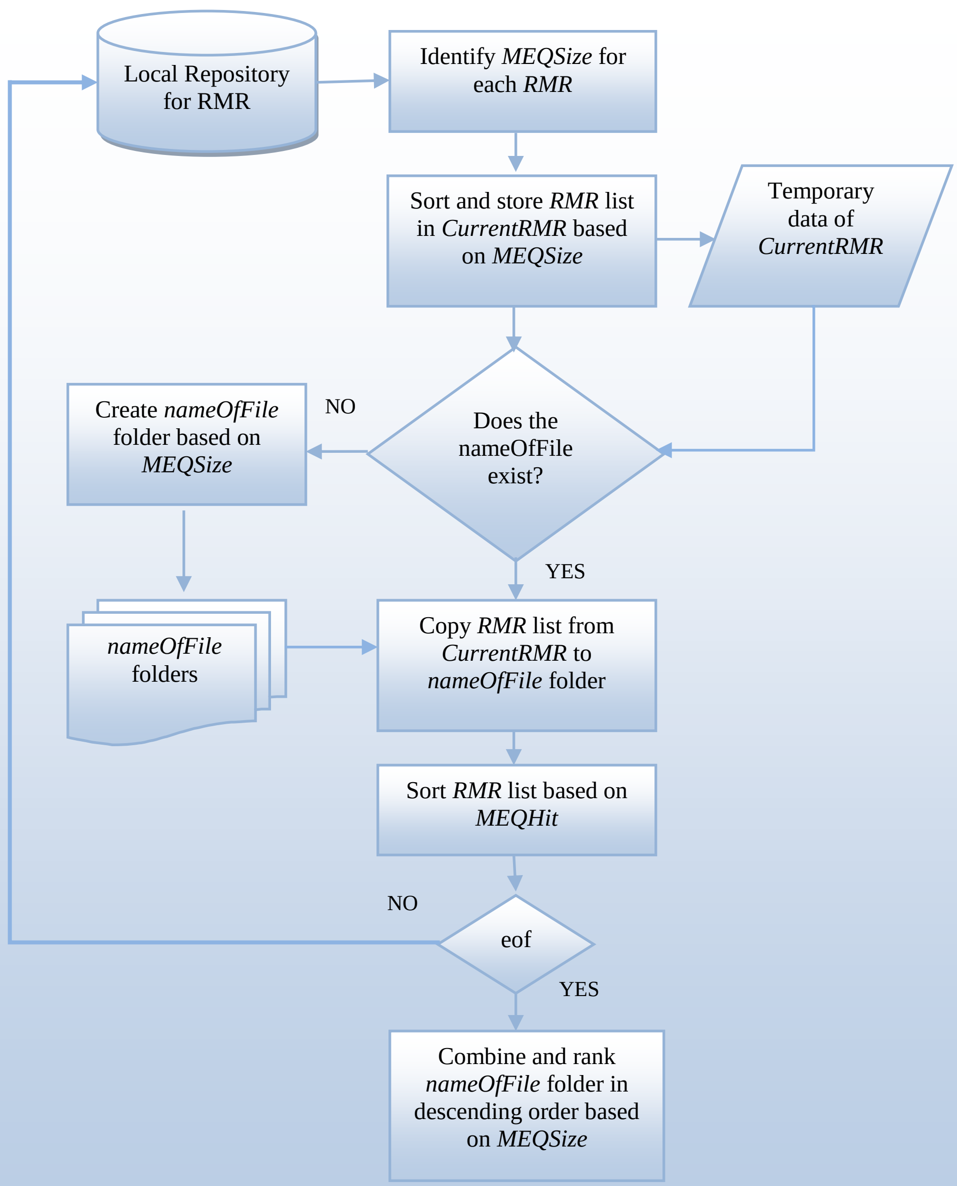


Figure 4.11: Flow chart of Comprehensive Ranking Model

The input of this flow chart is retrieved *RMR* while the output is the result file of ranked *RMR*. Initially the system initializes *MEQSize* as number of item in *MEQ* that exist in each *RMR*. Then the next step is to sort *RMR* based on the largest size of *MEQSize*. To execute *RMR* ranking process, *CurrentRMR* is created as a temporary variable to place the current *RMR*. For each *RMR*, the *MEQSize* is identified and later *nameOfFile* is created based on *MEQSize* and *RMR* is copied by *CurrentRMR* into *nameOfFile* folder. For each folder *CurrentRMR* is sorted based on descending order of *MEQHit*. Finally, *nameOfFile* folders are combine and sorted in descending order based on *MEQSize*.

Figure 4.12, Figure 4.13 and Figure 4.14 show the example for q_0="CT images containing fatty liver" containing three *nameOfFile* folders namely MatchedSize-1, MatchedSize-2 and MatchedSize-3. For each *nameOfFile* folder, *CurrentRMR* is sorted based on descending order of *MEQHit*.

FigureID	Title	Caption	MEQHit	MEQSize=1
28055	ith multiphase breath	the suspected lesion (op	3	liver:3
28080	ance of the soft palat	the soft palate (arrow) a	1	fatty liver:1

Figure 4.12: MatchedSize-1 file

FigureID	Title	Caption	MEQHit	MEQSize =1	MEQSize =2
28045	vith multiphase breath	inium-enhanced mr angio	5	liver:4	ct:1
28056	trast medium on panc	0 ml of nonionic contrast	3	liver:1	ct:2
28057	trast medium on pan	on volume of 150 ml and	2	liver:1	ct:1

Figure 4.13: MatchedSize-2 file

FigureID	Title	Caption	MEQHit	MEQSize =1	MEQSize =2	MEQSize =3
59549	cs and pseudolesions	n obtained during arteria	8	fatty liver:1	liver:2	ct:5
112963	omparison of helical	angle 15&#176) and (c) sp	5	fatty liver:1	ct:3	liver:1
117448	n focal lesions at micr	enous phase us image th	4	fatty liver:1	liver:2	ct:1

Figure 4.14: MatchedSize-3 file

FigureID	Title	Caption	MEQHit	MEQSize=1	MEQSize=2	MEQSize=3
59549	cs and pseudolesions	obtained during arteria	8	fatty liver:1	liver:2	ct:5
112963	mparison of helical	angle 15&#176) and (c) s	5	fatty liver:1	ct:3	liver:1
117448	focal lesions at micr	enous phase us image th	4	fatty liver:1	liver:2	ct:1
28045	ith multiphase breath	ium-enhanced mr angio	5	liver:4	ct:1	
28056	trast medium on pan	ml of nonionic contrast	3	liver:1	ct:2	
28057	trast medium on pan	n volume of 150 ml and	2	liver:1	ct:1	
28055	ith multiphase breat	he suspected lesion (op	3	liver:3		
28080	ance of the soft palat	the soft palate (arrow) a	1	fatty liver:1		

Figure 4.15: Example output in comprehensive ranking model

This process continues for every item in *RMR* to generate the final result of the ranking process. The *nameOfFile* folders will be gathered and ranked and sorted in descending order of *MEQSize*. Finally Figure 4.15 is the example of *RMR* list output from comprehensive ranking model which similar with the output of retrieval strategy component but the order is based on *MEQSize* followed by *MEQHit*.

Note that the figures above show only the extraction of the actual result to represent the example in this section. The actual value for q_0=“CT images containing fatty liver” retrieval list is 34,126 documents.

4.6.2 Medical Hierarchical Conceptual Ranking Model (MedHieCon)

Based on previous section, comprehensive ranking model provides high recall (number of relevant documents retrieved as fraction of all relevant documents) but low precision (proportion of relevant documents in the set of all documents retrieved) by prioritizing *MEQSize*. Figure ID “28080” as shown in Figure 4.15 is a relevant document (according to ImageCLEF relevant judgment data) and supposed to be among those listed at the top in *RMR*. However, due to the *MEQSize* equal to 1; the document is ranked in low position.

Low position will lead to low precision value. Therefore, this will affect effectiveness of M3IRS text-based framework model.

To improve the precision we decided to map medical term into more specific meaning entity. From close analysis of semantic medical terms, there is regularity that appears in every query which involves modality, anatomy and pathology as we assigned it as "concept". Therefore, MedHieCon ranking model is created to rank *RMR* based on these three medical concepts namely modality, anatomy and pathology as Figure 4.16 shows the flow chart of MedHieCon ranking model.

For example, the query is "CT images containing fatty liver", where "ct" is modality concept, "liver" is the anatomy and "fatty liver" is the pathology. According to Chevallet et. al (2005) *"Relevant documents must contain the anatomy and the pathology terms of the query"*. Considering the importance of anatomy and pathology concepts and from perspective of medical and clinical interest, pathology is then the most important concept. These concepts can be arranged according to hierarchy where top level is modality which refers to any type of medical modality (CT, MRI, US, x-ray & etc.), moving down to another level explaining the anatomy (body part or organ component) along with the bottom level which is pathology description (disease or syndrome).

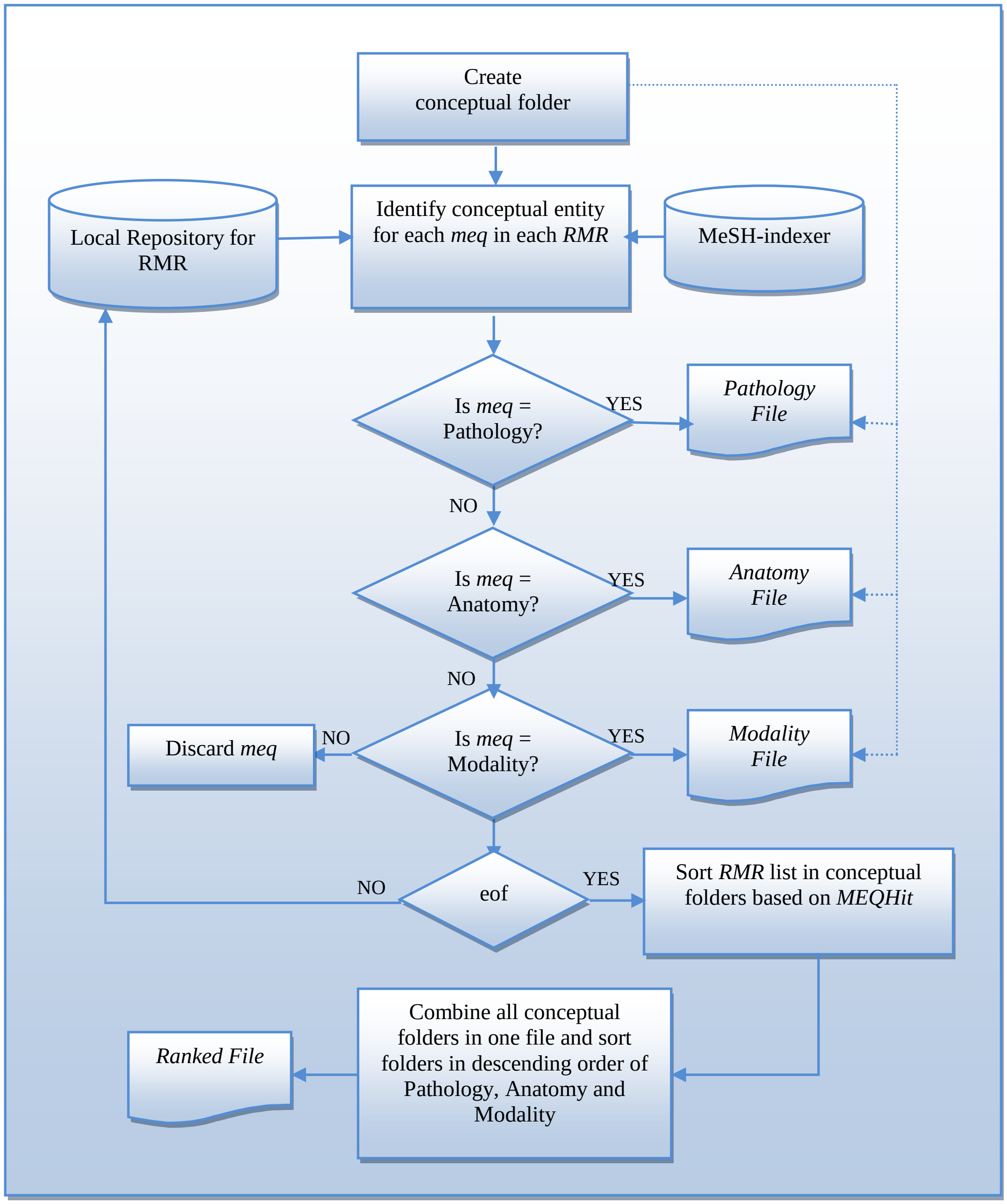


Figure 4.16: Flow chart of MedHieCon Ranking Model

In order to map the medical concepts with medical terms, we utilize the semantic type which is provided in the MeSH-indexer. Therefore, MC folder from MeSH-indexer (see section 4.4.2) is used to map each medical term with medical concepts entity. MC folder consists of medical term with semantic type and medical concepts as depicted in Figure 4.8. Medical concepts entity is assigned based on semantic type which is adapted from UMLS semantic network. Table 4.3 shows the list of semantic types used in this research that are mapped with the medical concepts of “Modality”, “Anatomy” and “Pathology” to provide important implication for interpreting the meaning into our MedHieCon ranking model.

Table 4.3: The mapping of semantic types into concept entities

Concepts	Semantic Types
Modality	“manufactured object”, "diagnostic procedure”, "natural phenomenon or process”, "therapeutic or preventive procedure”
Anatomy	“cell, "body part, organ, or organ component", "body space or junction", "body location or region"
Pathology	"acquired abnormality”, ”congenital abnormality", "anatomical abnormality", "sign or symptom”, “finding", “pathologic function", "injury or poisoning", "disease or syndrome", “neoplastic process", "neoplasms”, "bacterium", "body substance"

MedHieCon ranking model executes the ranking process by ranking *RMR* that prioritizes medical term which maps into pathology concept; and if there is no pathology concept in the query list then the prioritization will be on anatomy concept and so on.

For each *meq* in *RMR*, the system will refer to MC folder to access the semantic type and medical concepts of the term. The system also creates new files according to the concept entities namely *PathologyFile, AnatomyFile* and *ModalityFile* as shown in Figure 4.17. Then the *RMR* will be assigned to one of these folders according to the map between *meq* and medical concepts. As previously mentioned, the priority will be given to “Pathology” concept followed by “Anatomy” and “Modality”. For example, if a *RMR* consists of a list of *meq* that map with these three concepts, therefore the *RMR* will be assigned into *PathologyFile* and if the *RMR* consists list of *meq* with “Anatomy” and “Modality” concepts then the *RMR* will be assigned into *AnatomyFile* and if the *meq* has only "Modality" concept than the *RMR* is assigned to *ModalityFile* as shown in Figure 4.18, Figure 4.19 and Figure 4.20.

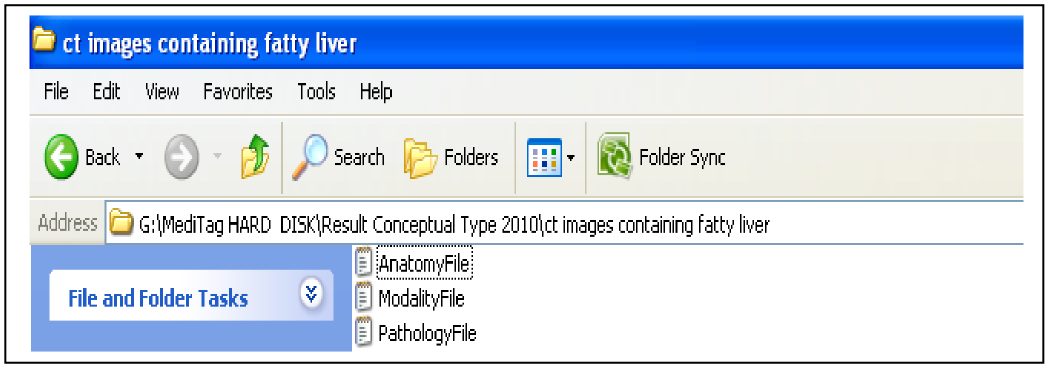


Figure 4.17: Concept files automatic generated by the system

FigureID	Title	Caption	MEQHit	Pathology	Anatomy	Modality
59549	cs and pseudolesions	obtained during arteria	8	fatty liver:1	liver:2	ct:5
112963	omparison of helical	angle 15&#176) and (c) s	5	fatty liver:1	ct:3	liver:1
117448	i focal lesions at mic	renous phase us image th	4	fatty liver:1	liver:2	ct:1
28080	ance of the soft palat	the soft palate (arrow) a	1	fatty liver:1		

Figure 4.18: *PathologyFile* with each *RMR* has *meq* with "Pathology" concept

FigureID	Title	Caption	MEQHit	Pathology	Anatomy	Modality
28045	ith multiphase breat	hium-enhanced mr angio	5		liver:4	ct:1
28056	trast medium on pan	) ml of nonionic contras	3		liver:1	ct:2
28055	ith multiphase breat	he suspected lesion (op	3		liver:3	
28057	trast medium on pan	n volume of 150 ml and	2		liver:1	ct:1

Figure 4.19: *AnatomyFile* which each *RMR* has *meq* with "Anatomy" concept

FigureID	Title	Caption	MEQHit	Pathology	Anatomy	Modality
106151	recurrent hepatocell	figure 1d transverse cor	3			ct:3
106311	littoral cell angioma	figure 3a patient 7 littor	2			ct:2

Figure 4.20: *ModalityFile* which each RMR has meq with "Modality" concept

These files will update the figure ID and *MEQHit* for each new *RMR* that are added on afterwards. Later new file called *RankedFile* is created to extract content of *PathologyFile* followed with *AnatomyFile* and *ModalityFile* files accordingly and ranked *RMR*.

Figure 4.21 shows an example of output result using MedHieCon ranking model. The figure shows that Figure ID "28080" has been ranked higher since "fatty liver" is in pathology concept although the size of matched term is lesser than the others. This method results to high position of potential relevant medical reports which lead to high precision value.

FigureID	Title	Caption	MEQHit	Pathology	Anatomy	Modality
59549	cs and pseudolesions	obtained during arteria	8	fatty liver:1	liver:2	ct:5
112963	omparison of helical	angle 15&#176) and (c) s	5	fatty liver:1	ct:3	liver:1
117448	n focal lesions at micr	enous phase us image th	4	fatty liver:1	liver:2	ct:1
28080	ance of the soft palat	the soft palate (arrow) a	1	fatty liver:1		
28045	ith multiphase breath	ium-enhanced mr angio	5		liver:4	ct:1
28056	trast medium on pan	0 ml of nonionic contrast	3		liver:1	ct:2
28055	ith multiphase breath	he suspected lesion (op	3		liver:3	
28057	trast medium on pan	in volume of 150 ml and	2		liver:1	ct:1
106151	recurrent hepatocell	figure 1d transverse cor	3			ct:3
106311	littoral cell angioma (	figure 3a patient 7 littor	2			ct:2

Figure 4.21: Final result using MedHieCon Ranking Model

4.7 Summary

M3IRS text-based framework introduced in this chapter consists of four main components namely document pre-processor, query processor, retrieval strategies and ranking strategies. Query processor involves three types of queries which are i) exact match query, ii) medical term query expansion and iii) medical and synonymous terms query enrichment. Unlike statistical model, this framework uses Boolean model as retrieval strategy since this framework emphasizes on efficient process with advantages discussed earlier and which is only based on matching between query and documents and no term occurrence indexing required on the documents. Therefore new documents can easily be inserted in the data collection without having to re-index the length vector of the documents. However these new documents are still required to be processed at Document Pre-processor stage before can be stored in local repository. Since Boolean model does not provide result in ranking manner, ranking strategies component is used to rank the results from retrieval strategies component. Two ranking models are introduced as comprehensive and MedHieCon ranking model. Query expansion technique is applied on the original query to provide more relevant and meaningful results in the rank list. XTE method is used to pre-process the medical documents in data collection which are later indexed to local repository.

5.0 M3ICS Content-based Framework and Module Descriptions

5.1 Introduction

This chapter introduces content-based framework for M3ICS. This framework is based on extracting visual features of texture, shape and color in global and local descriptors and applied semantic classification-based medical image retrieval framework. This framework uses supervised machine learning technique to utilise guidance (training) with example dataset to predict the unknown dataset. MedHieCon model which represent three concepts of modality, anatomy and pathology to create models for training data in supervised classification. Supervised learning technique is heavily used in visual classification to evaluate the performance of the framework. M3ICS content-based framework and its breakdown components are diagrammatically presented.

5.2 M3ICS Content-based Retrieval Framework

The M3ICS content-based framework contains two main components namely feature extraction and classification conceptual train data. The feature extraction phase is divided into two sub-phases which extract features in (i) global and (ii) local. Both descriptors extract visual features of texture, shape and color. In classification, a medical concept modelling is created to interpret the semantic of medical image to more specific annotation. These concepts are trained in classification conceptual train data phase as a series of concepts namely modality, anatomy and pathology and accumulated in feature vector storage. The overall organization of M3ICS content-based framework is illustrated in Figure 1.4. This figure shows the organization of the processes, dataset and information flow within the framework. Distinguished symbolical notations are used to denote dataset, execution processes or systems and document collection.

5.3 Feature Extraction Phase

The feature extraction phase of the M3ICS framework is used to extract significant information. The extraction is based on global, which extract features from the whole image and local, which involves a small group of patches. The global feature captures overall characteristics of an image while local features show more details. Most content-based medical image retrieval systems tend to use global image features (Cootes, 2001; Aleksandra, 2002; Samee, 2003; Lehman, 2005), which describes an image as a whole or provide overall structure of an image. On the other hand, local features are more robust to occlusion and clutter. Combination of both global and local features can improve the final result. The more the discriminated features, the better the classification result. Visual features of texture, shape and color are used to represent each medical image. Normally, image recognition rates can be improved by combining different image features such as texture, shape and color (Lehmann et al., 2003).

5.3.1 Texture

In this research we analyse texture features based on contrast, correlation, energy and homogeneity. These texture measures try to capture the characteristics of the image parts with respect to changes in certain directions and the scale of the changes. Gray-level co-occurrence matrix (GLCM) is a statistical method of examining texture that considers the spatial relationship of pixels which also known as the gray-level spatial dependence matrix.

GLCM is well-known texture extraction tool originally introduced by Haralick R.M. (1973).

GLCM is a tabulation of how often different combinations of gray levels co-occur in an image. Texture feature calculations use the contents of the GLCM to give a measure of variations in the image texture at the pixel of interest. There are four co-occurrence matrixes (contrast, energy, homogeneity, and entropy) for four different orientations are obtained. Therefore vector of texture feature extraction may produce the length of 16-dimensional. The texture features information is derived from GLCM by using the following formula.

1) *Contrast:* Measures intensity contrast between a pixel and its neighbour over whole image.

$$\sum_{i,j} |i-j|^2 \; p(i,j) \qquad (5.1)$$

2) *Correlation:* Measures the joint probability occurrence of the specified pixel pairs

$$\sum_{i,j} \frac{(i-\mu_i)(j-\mu_j)\vec{p}(i,j)}{\sigma_i \sigma_j} \qquad (5.2)$$

3) *Energy:* Provides the sum of squared elements in the GLCM. Also known as uniformity or the angular second moment

$$\sum_{i,j} p(i,j)^2 \qquad (5.3)$$

4) *Homogeneity:* Measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal

$$\sum_{i,j} \frac{p(i,j)}{1+|i-j|} \qquad (5.4)$$

where; i is the row number, j is the column number, Pij is the normalized pixel value in the cell i,j and N is the number of rows or columns. μ_i, μ_j, σ_i and σ_i are the means and standard deviations of the marginal distribution with Pij.

5.3.2 Shape

In this research work, we analyse shape feature using Hu moment invariant method. As mentioned in section 3.9.4, moment-based shape features provide a numerical shape-preserving representation that is invariant to translation, rotation and scale. Medical images are usually complex to high variability and have subtle differences from each other in the context of visual appearance (Madzin et. al, 2011). Therefore, Hu moment invariant method is suitable to measure the shape feature for medical images. Hu moment invariant method explanation can be access in (Hu, 1962). In particular, Hu described a set of six moments which are rotation, scaling, translation invariant and the seventh invariant is skew invariant as depicted in (5.10).

For a 2-D continuous function *f*(*x*,*y*) the m_{pq} moment of order (*p* + *q*) is defined as

$$\sum_{x=1}\sum_{y=1} x^p y^p f(x,y) \tag{5.5}$$

For moment representation of area (for binary images) or the sum of grey level (for greytone images) is M_{00} and the centroid is represent as $\{x, y\} = \{M_{10}/M_{00}, M_{01}/M_{00}\}$. To measure the central in moments of digital image, μ_{pq} is used and it represented as in equation (5.6)

$$\mu_{pq} = \sum_x \sum_y (x - x')^p (y - y')^q f(x, y) \tag{5.6}$$

where the centroid of image is defined in terms of the first order moments as $x' = \frac{M_{10}}{M_{00}}$ and $y' = \frac{M_{01}}{M_{00}}$. The centroid moments μ_{pq}computed using the centroid of the image *f(x,y)* is equivalent to the m_{pq}whose center has been shifted to centroid of the image. Therefore, the central moments are invariant to image translations. Scale invariance can be obtained by scaling normalization which defined as in (5.7)

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}\gamma} \tag{5.7}$$

where the normalization factor is $\gamma = (p+q+2)/2$. The central moments are order up to 3 as follows as in (5.8)

$$\begin{aligned}
&\mu_{00} = M_{00},\\
&\mu_{01} = 0,\\
&\mu_{10} = 0,\\
&\mu_{11} = M_{11} - x'M_{01} = M_{11} - y'M_{10},\\
&\mu_{20} = M_{20} - x'M_{10},\\
&\mu_{02} = M_{02} - x'M_{01},\\
&\mu_{21} = M_{21} - 2x'M_{11} - y'M_{20} + 2x'^2M_{01},\\
&\mu_{12} = M_{12} - 2y'M_{11} - x'M_{02} + 2y'^2M_{10},\\
&\mu_{30} = M_{30} - 3x'M_{20} + 2x'^2M_{10},\\
&\mu_{03} = M_{03} - 3y'M_{02} + 2y'^2M_{01}
\end{aligned} \tag{5.8}$$

To measure moments η_{ij}, where $i + j \geq 2$ is by dividing the corresponding central moment with the properly scaled $(00)^{th}$ moment. The formula constructed can be invariant to both translation and changes in scale by following formula:

$$\eta_{ij} = \frac{\mu_{ij}}{\mu_{00}^{\left(1+\frac{i+j}{2}\right)}} \tag{5.9}$$

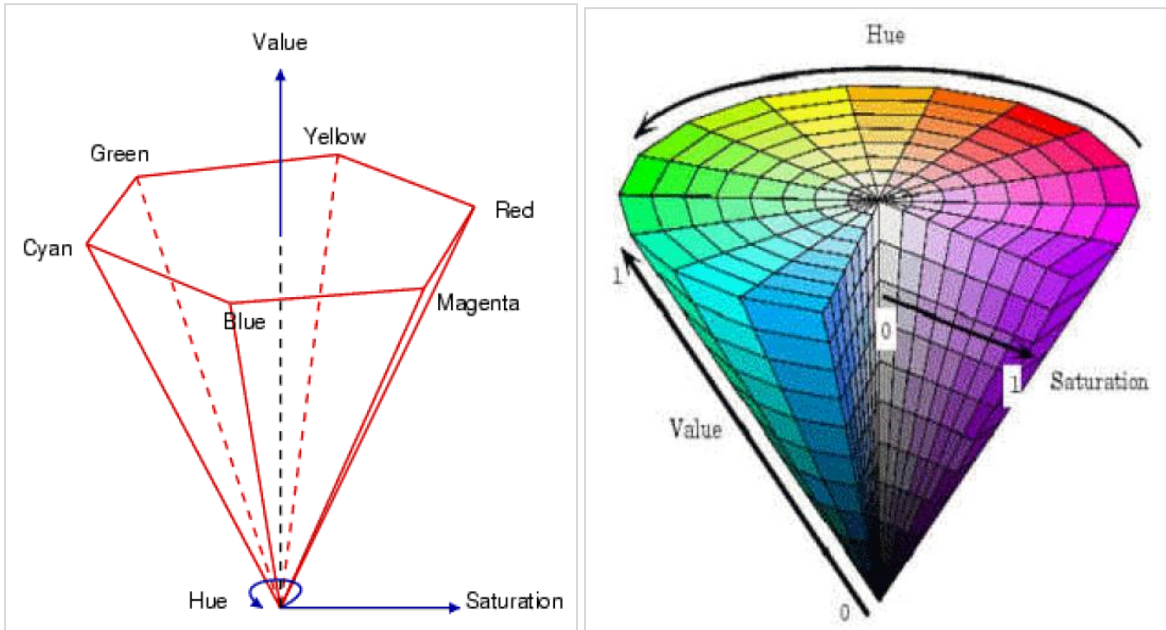
Basically Hu moment invariant is described as a set of moments that are invariant under translation, changes in scale, and also rotation as follows:

$$
\begin{aligned}
I_1 &= \eta_{20} + \eta_{02} \\
I_2 &= (\eta_{20} - \eta_{02})^2 + (2\eta_{11})^2 \\
I_3 &= (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \\
I_4 &= (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \\
I_5 &= (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + \\
&\quad (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\
I_6 &= (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \\
I_7 &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] - \\
&\quad (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2].
\end{aligned}
\tag{5.10}
$$

I_1 represents analogous to the moment of inertia around the image's centroid, where the pixel’s intensities are analogous to physical density and I_7 is a skew invariant moment which enables one to distinguish mirror images of otherwise identical images. Each I_i represent the value in shape feature vector with the length of 7-diemnsional.

5.3.3 Color

The color histogram describes the proportion of pixels of each color in an image with simple and computationally effective manner. The color histogram is obtained by quantizing image colors into discrete levels and then counting the number of times of each discrete color occurs in the image. In this research work, generic color histogram is used (Manjunath, et al., 2001). In order to measure color, greyness and brightness, the image color needs to be in HSV format.



(a) (b)

Figure 5.1:The HSV coordinate system in a hexacone in (a) and (b) a view of the HSV color model. (Adapted from Lei et. al, 1999)

HSV means Hue-Saturation-Value as depicted in Figure 5.1. Hue represents color and it is not easy thing to compare as hue is often represented as a circular angle which provides value between the ranges of 0.0 to 1.0. Being a circular value means that 1.0 is the same as 0.0. Saturation is the greyness, where saturation value near 0 means dull or grey looking whereas saturation value of 0.8 might mean a very strong color. Value represents the brightness of the pixel, so 0.1 is blackish and 0.9 is more to white color. Once the medical image is converted to HSV format, HSV histogram is created. Generic color histogram descriptor was able to capture the color distribution with reasonable accuracy for image search and retrieval applications. Finally normalization of color histogram is executed to produce better result.

Color histogram is created based on the joint probabilities mass function of the intensities of the three color channels (H,S,V) defined as

$$h_{a,b,c} = N.Prob\ (A = a, B = b, C = c) \qquad (5.11)$$

where A , B and C represent the three color channels (H,S,V) and N is the number of pixels in the image. Computationally, the color histogram is formed by discretizing the colors within an image and counting the number of pixels of each color.

Euclidean distance is used to measure the difference of two color histograms where ***h*** and ***g*** represents the histograms and can be computed using the equation (5.12)

$$d^2(h, g) = \sum_A \sum_B \sum_C (h(a, b, c) - g(a, b, c))^2 \qquad (5.12)$$

However equation (5.12) represents only comparison between the identical bins in the respective histograms. Two different bins may represent perceptually similar colors but are not compared obliquely. All bins contribute equally to the distance. Therefore intersection of histogram is used (Swain, 1991).

The histograms ***h*** and ***g*** are given by:

$$d(h, g) = \frac{\sum_A \sum_B \sum_C (h(a,b,c), g(a,b,c))^2}{\min(|h|, |g|)} \qquad (5.13)$$

where |*h*| and |*g*| give the magnitude of each histogram, which is equal to the number of samples. The sum is normalized by the histogram that has the fewest samples.

5.4 Global Descriptor Architecture

Global descriptor represents visual features for entire image which means in this research study M3ICS will extract visual features of each medical image in the collection at global level as illustrated Figure 5.2. This diagram includes two main processors namely image pre-processing and feature extraction. The features will be indexed and stored in a vector form in feature vector database. Then, classifier is used to train the vector for classification purposed.

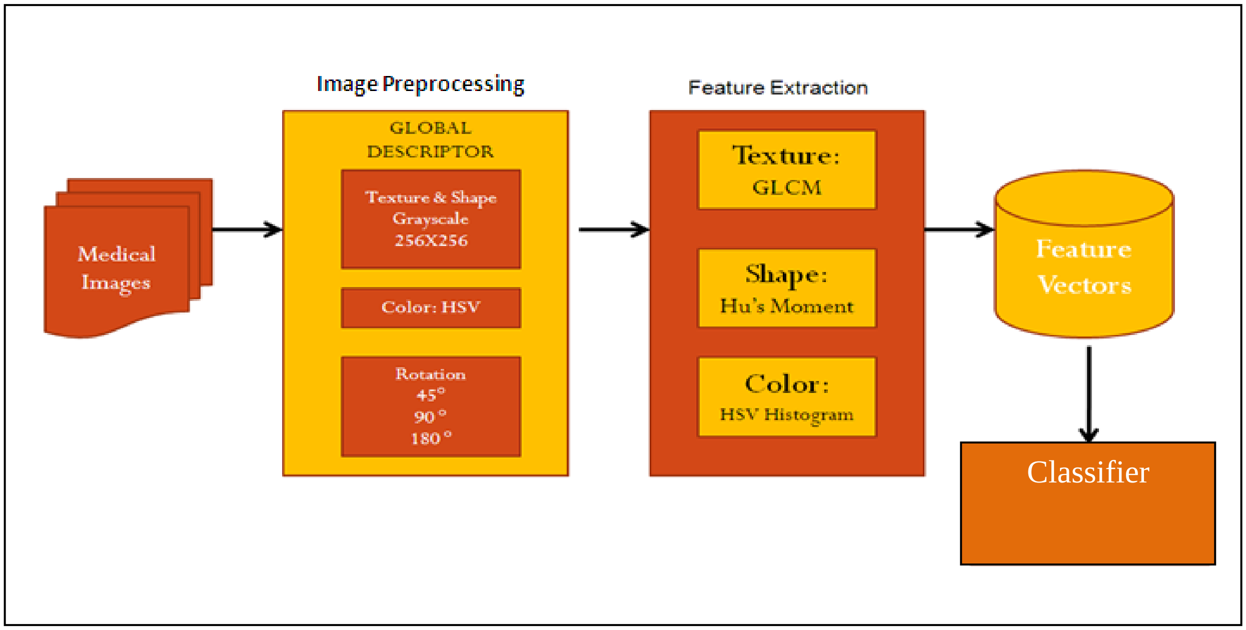


Figure 5.2: Architecture of Global descriptor process

Texture and shape feature extractions require grayscale image. In pre-processing phase, original medical image is converted to grayscale image of 256×256 pixels. For texture features which apply GLCM, it is important to consider a number of co-occurrence matrix, one for each relative location to obtain different texture cues or the same cues at different scales. Therefore the greyscale image will be rotated to four directions (*0^o, 45^o, 90^o and 180^o*) as depicted in Figure 5.3.

As mentioned in section 5.3.3, to extract color feature, the original medical image need to be converted to HSV format in order to produce HSV histogram.

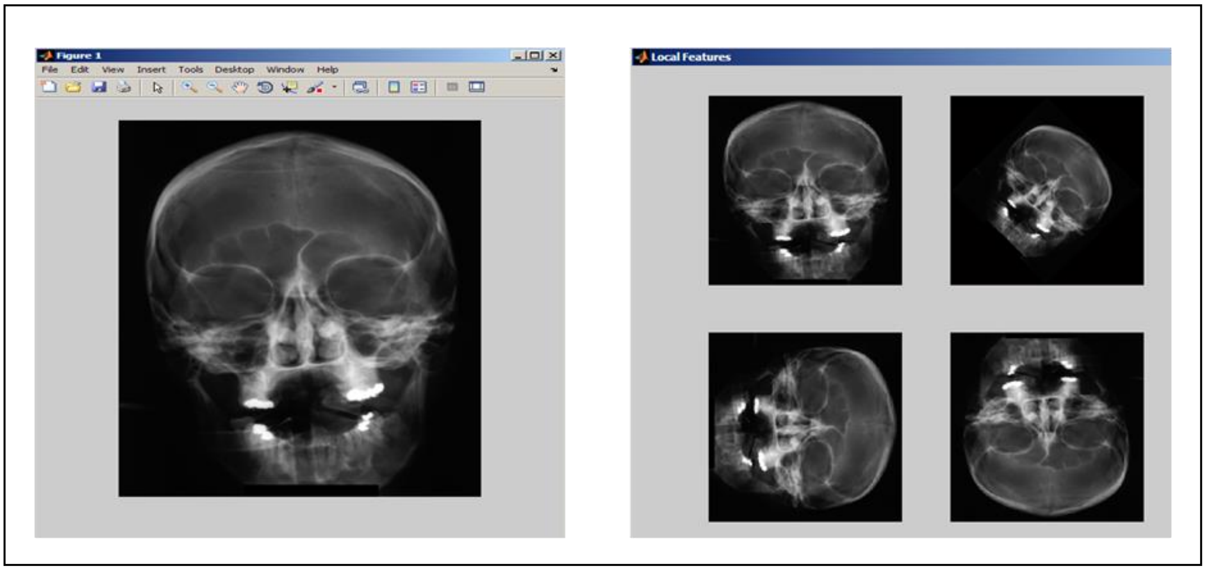


Figure 5.3: Example of original medical image and rotation of 0°, 45 °, 90 ° and 180 ° medical images

After image pre-processing has been executed, the system will extract texture, shape and color features from each medical images. These features will be combined into a single feature vector with 75-dimensional as depicted in Table 5.1. The feature vectors are then stored to feature vectors database before it's been classified.

Table 5.1: Global Visual Features Dimensional

Features	**n-dimensional**
Color	4-dimensional
Moment Shape	7-dimensional
GLCM Texture (*0°, 45°, 90° and 180°*)	64-dimensional
Total	**75-dimensional**

5.5 Local Descriptor

Similar to global descriptor, local descriptor framework consists of two main processors namely image pre-processing and feature extraction as illustrated in Figure 5.4. However the technique used in local descriptor image pre-processing phase is slightly different from that of global descriptor. Initially, the medical image needs to be converted to greyscale for texture and shape features and also in HSV format for color feature extractions. As for shape feature, no rotation is applied since there will be seven moment formulas will be applied to extract shape features. The next procedure in local descriptor is to divide the original medical image into patches. In this research work, we patched the original medical image into 2×2, 4×4 and 8×8 patches. However to get more detailed information we extract interest points from the original medical image. Finally these features will be combined in a single feature vector and stored in feature vector database.

5.5.1 Patches

In MIARS system developed by Mueen A. (2010), 2×2 patches were used. However there is no concrete proof to show that 2×2 patches is suitable in extracting features in local descriptor. Therefore in this research, we executed each size of 2×2, 4×4 and 8×8 patches in separate individual experiments. The evaluation of various sizes was done to determine which size of patches is suitable for extracting visual features in medical image. Figure 5.5 until 5.8 show the examples of original medical images which has been patched into several sizes for modality of x-ray of skull, CT of liver, MRI of heart and microscopy image.

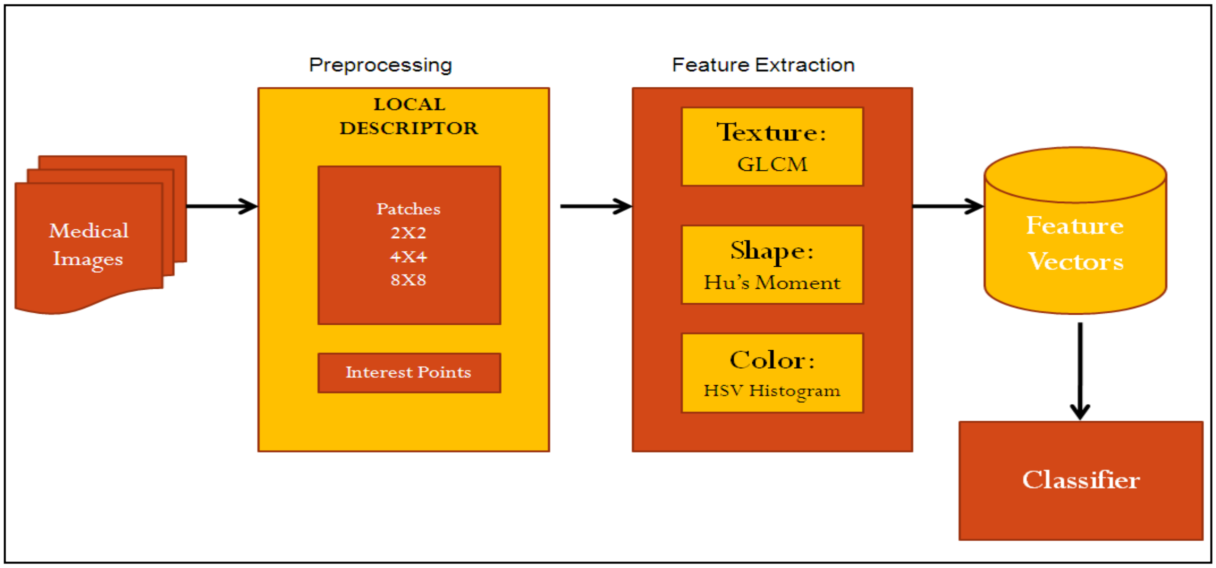


Figure 5.4: Architecture of Local Descriptor Process

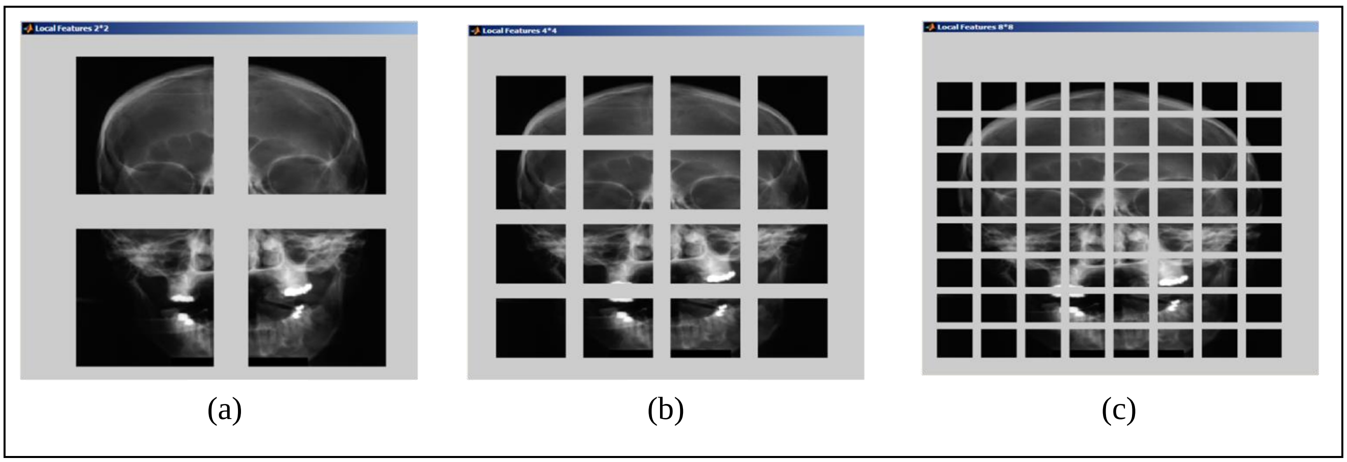


Figure 5.5: X-ray of skull patches in (a) 2×2, (b) 4×4 and (c) 8×8

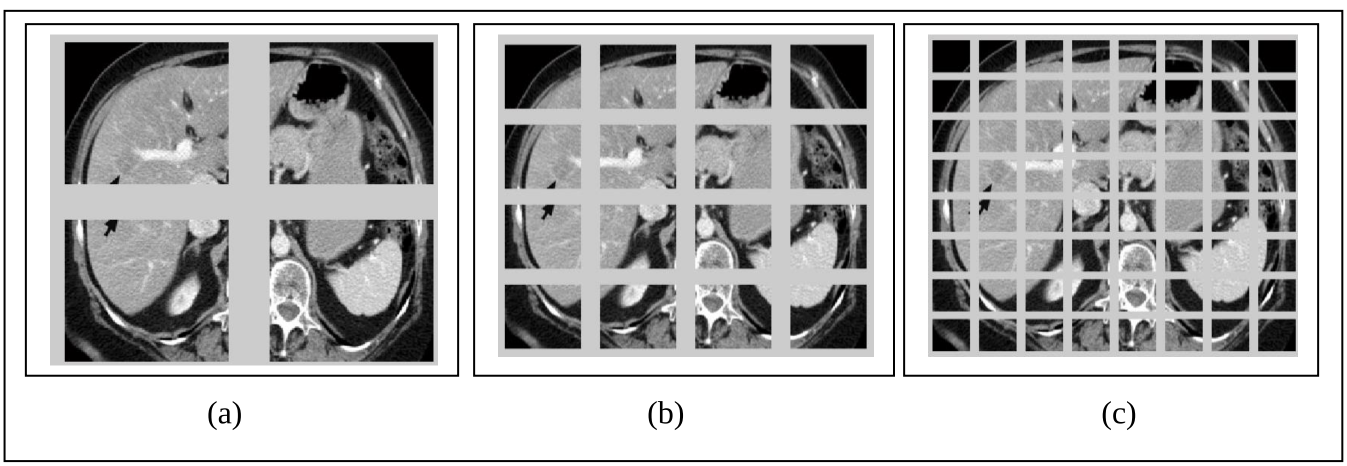
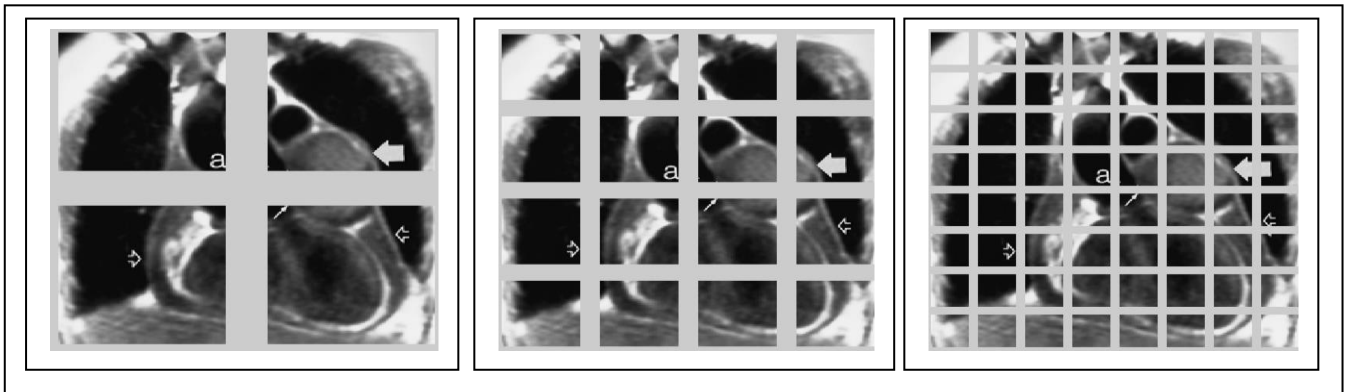
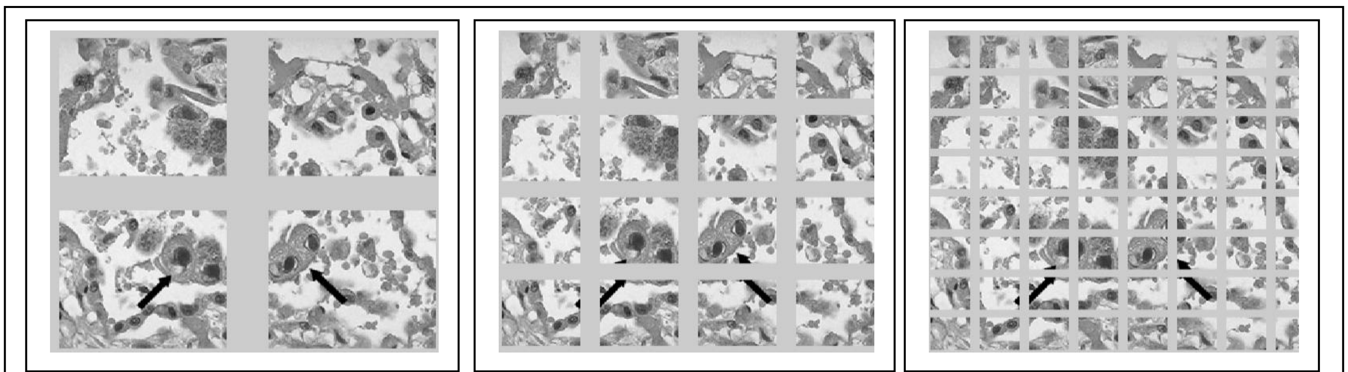


Figure 5. 6: CT of liver patches in (a) 2×2, (b) 4×4 and (c) 8×8



(a) (b) (c)

Figure 5.7: MR of heart patches in (a) 2×2, (b) 4×4 and (c) 8×8



(a) (b) (c)

Figure 5.8: Microscopy image patches in (a) 2×2, (b) 4×4 and (c) 8×8

Table 5.2: Local Patches Visual Features with Dimensional Sizes

Patches Size	**2x2**	**4x4**	**8x8**
Color	8-dimensional	16-diemsional	32-dimensional
Moment Shape	14-diemnsional	28-dimensional	56-dimensional
GLCM Texture (*0°*)	32-dimensional	64-dimensional	128-dimensional
Total	**54-dimensional**	**108-dimensional**	**166-dimensional**

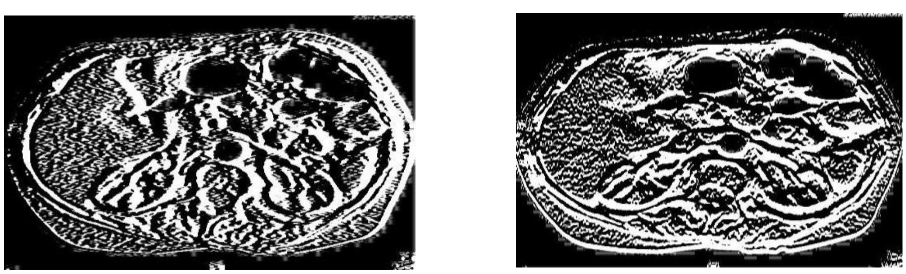
Each patch will be used to extract texture, shape and color features. Different size of patches will produce different dimensional size in feature space as depicted in Table 5.2. The table clearly shows that the more patches used the longest length of dimensional stored in a feature vector.

5.5.2 Interest Blocks

Harris detector is based on second moment matrix, M created by Harris and Stephens (1988). The method was developed to cater image regions with texture and isolated feature. The matrix calculation describes the gradient distribution in the local neighbourhood of a point as shown in equation (5.14).

$$M = \sum w(x,y) \begin{pmatrix} i_{x^2}(x) & i_x i_y(y) \\ i_x i_y(x) & i_{y^2}(y) \end{pmatrix} \quad (5.14)$$

Initially, the first order of local image derivatives, i_x and i_y is computed. Then, the product of these gradient images is taken. Figure 5.9 shows the initial step of Harris Detector.



(a) (b)

Figure 5.9: Figure (a) represent the local derivative of medical image in x-direction and figure (b) represent the local derivative of medical image in y-direction.

The next step is to smooth the image with Gaussian kernels, *w(x,y)* in different scale values (σ) as stated in equation (5.15). The eigenvalues represent the significant signal changes in two orthogonal directions in neighbourhood around the point.

$$w(x,y) = g(x,y,\sigma) = \frac{1}{2\pi\sigma} e^{\frac{-x^2+y^2}{2\sigma}} \qquad (5.15)$$

Different values in Gaussian Kernel may produce different number of interest points as we have experimented in Madzin (2009). Finally, Harris measured the cornerness that is defined as positive local extrema in equation (5.16).

$$\text{cornerness} = \det(M) - \lambda \,\text{trace}(M) \qquad (5.16)$$

where a constant value of λ is 0.04.

Once the interest points are generated, a block of 20×20 pixels is generated for each point and interest point as the centre point. The size of 20×20 pixels is chose since this size is suitable and appropriate to extracts information of texture and shape. In this research, we set the maximum number of generated interest points is 20 as depicted in Figure 5.10. This is to avoid the increment of dimensional value in feature vector.

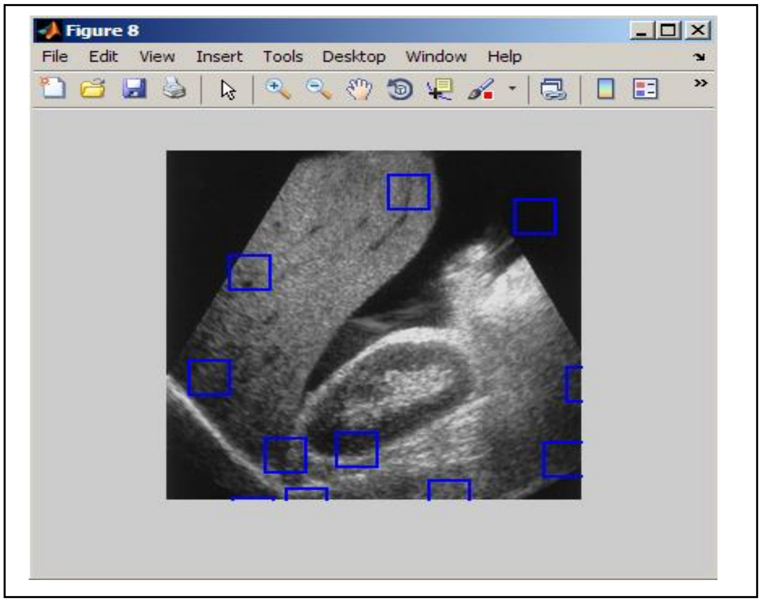


Figure 5.10: Haris interest point feature extraction

Figure 5.11 shows the example of interest blocks take from different modality of CT, MR and microscopy image.

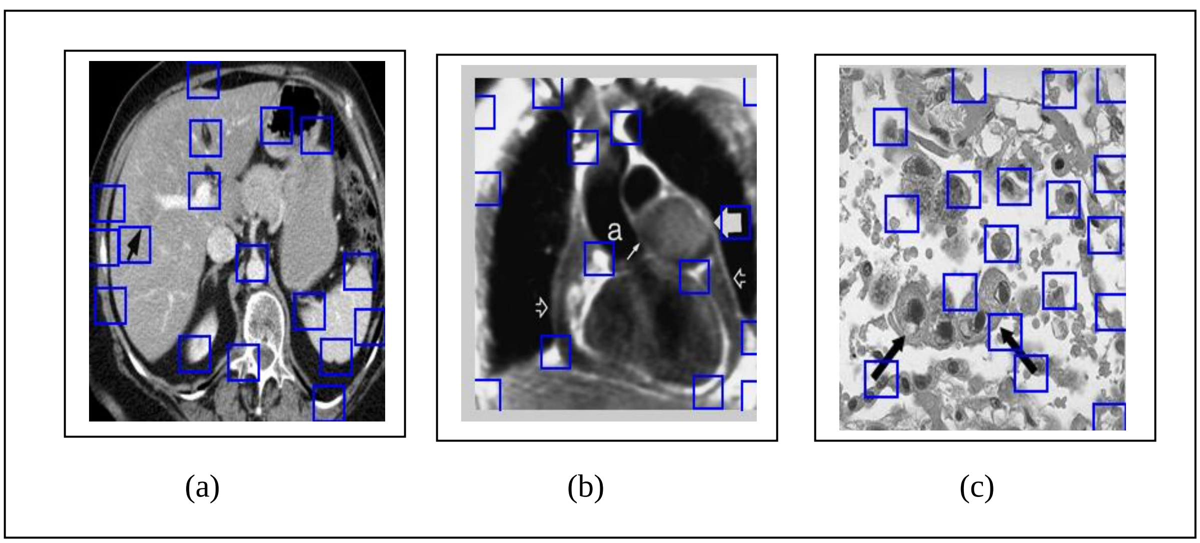


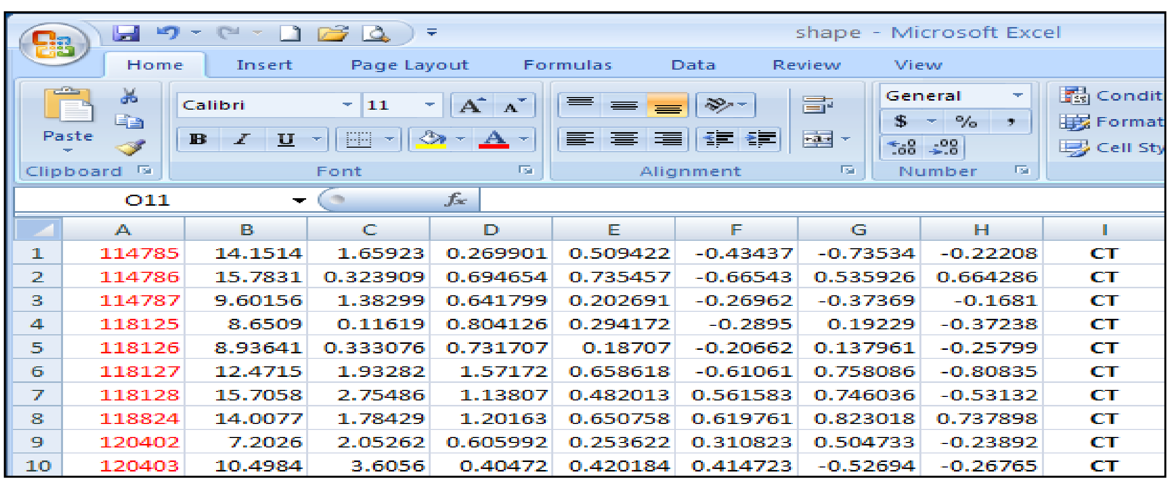
Figure 5.11: Figure 5.11: Examples of interest blocks in (a) CT of liver, (b) MR of heart and (c) microscopy medical image

Therefore there will be 20 blocks in each medical image. Each interest point will be extracted with shape and texture features. Color feature will not be used since HSV color histogram is not suitable in extracting a small size block (Garcia et. al, 1999). Table 5.3 shows the size of dimensional of a feature vector for one medical image.

Table 5.3: Local Interest Points Visual Features Dimensional

Feature	n-dimensional
Moment Shape	20 X 7-dimensional = 140-dimensional
GLCM Texture (*0*°)	20 X 16-dimensional = 320-dimensional
Total	**460-dimensional**

The feature vectors then are stored in excel format and later will be trained based on its class. For example in Figure 5.12, the feature vectors of shape feature extraction where first column represent image ID (red font), followed by seven values of moments and final column is the name of the class (bold font).



	A	B	C	D	E	F	G	H	I
1	114785	14.1514	1.65923	0.269901	0.509422	-0.43437	-0.73534	-0.22208	**CT**
2	114786	15.7831	0.323909	0.694654	0.735457	-0.66543	0.535926	0.664286	**CT**
3	114787	9.60156	1.38299	0.641799	0.202691	-0.26962	-0.37369	-0.1681	**CT**
4	118125	8.6509	0.11619	0.804126	0.294172	-0.2895	0.19229	-0.37238	**CT**
5	118126	8.93641	0.333076	0.731707	0.18707	-0.20662	0.137961	-0.25799	**CT**
6	118127	12.4715	1.93282	1.57172	0.658618	-0.61061	0.758086	-0.80835	**CT**
7	118128	15.7058	2.75486	1.13807	0.482013	0.561583	0.746036	-0.53132	**CT**
8	118824	14.0077	1.78429	1.20163	0.650758	0.619761	0.823018	0.737898	**CT**
9	120402	7.2026	2.05262	0.605992	0.253622	0.310823	0.504733	-0.23892	**CT**
10	120403	10.4984	3.6056	0.40472	0.420184	0.414723	-0.52694	-0.26765	**CT**

Figure 5.12: Example of feature vectors stored in excel format

5.6 Classification Conceptual Train Data Phase

Many research have been conducted to improve CBIR in medical application by relying solely on low-level features in order to identify visually similar images (El-Naqa et. al, 2004). It is common to find medical images that appear similar but not related to each other (Ibrahim et. al, 2010). Applying low-level features without any semantic interpretation may contribute to unsuccessful classification of medical images in different semantic categories due to the *semantic gap* in CBIR (Wang, 2008). This is because each values of feature information that extract from feature extraction are meaningless. To minimize limitations of low-level features classification-based medical image retrieval method is applied. This method is motivated by the successful use of machine learning in IR (Rahman et al., 2012). Building a medical concept model for semantic interpretation of the image can bridge up the semantic gap between low-level and high-level semantic annotation.

Concept model is used as a knowledge representation to visualize the concepts and relationships among the concepts (Christophe, 2010). The purpose of conceptualization in medical retrieval system is to classify the medical image into more specific entity or class in increasing the performance for ranking purpose, and annotate the images with semantic labels in a natural language description of the visual concepts of interests. Initially, number of training images must be assigned in each concept. After the system is trained by using visual features as training samples, it is possible to depict the semantic content of each query image by identifying its class assignment using a classifier (Kurkure et. al, 2010). This methodology allows classification of medical images into semantically meaningful annotation using low-level visual features.

A bottom top MedHieCon model is designed in pyramid structure as illustrated in Figure 5.13 for interpretation of semantic medical concepts in a specific domain of medical images. From this figure, the modality, anatomy and pathology concepts are utilized, and the descending order of the pyramid gives more specific information. For example the medical image of “CT images containing fatty liver”, therefore we trained the image of “CT” as modality, “liver” as anatomy and “fatty liver” as pathology.

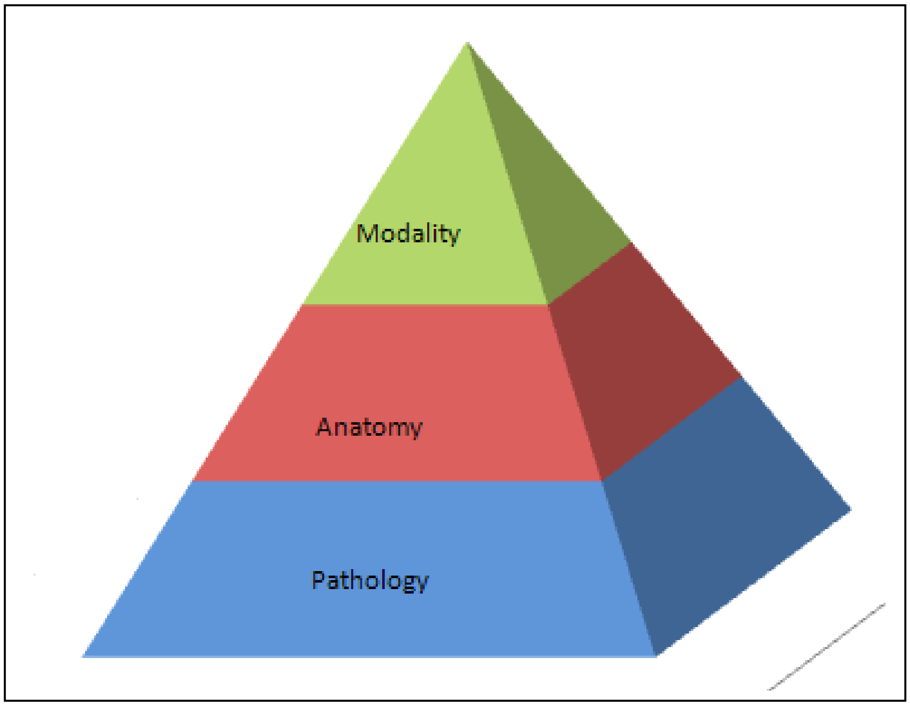


Figure 5.13: MedHieCon Model

5.6.1 Model Generation

In this research multi-class classification method is used by combining pairwise comparison of binary SVM classifier, known as one-against-one or pairwise coupling (Vapnik, 1998). Figures 5.14 to 5.16 show the multi-level learning procedure to generate classification functions. There will be three medical concept models namely modality, anatomy and pathology. The training image will be trained based on these concepts.

5.6.2 Modality Model

Figure 5.14 shows the first step as the modality concept learning stage to generate the first classifier.

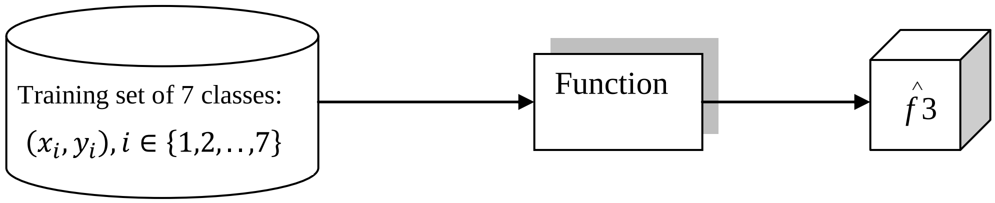


Figure 5.14: Modality model generation

The top hierarchy is modality concept which consists of seven modalities namely x-ray, PET, PX, US, CT, GX and MR. Therefore, the modality classifier $\hat{f}3$ obtained by training set of seven classes, L1 = {(x1, y1), (x2, y2), ….. (x7, y7)} where each *yi* is a label of the class associated with each *xi*.

5.6.3 Anatomy Model

This step is to generate a classifier according to 10 classes of a human anatomy. Figure 5.15 show the second-level learning stage to generate the second classifier.

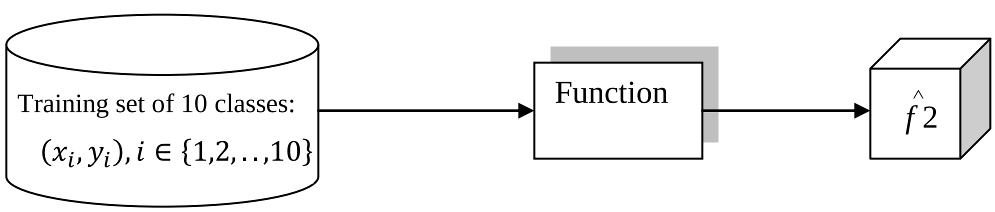


Figure 5.15: Anatomy model generation

The training set of 10 classes, L2 = {(x1, y1), (x2, y2), …, (x10, y10)} where each *yi* is a label of the class associated with each *xi*, are used to train the level 2 classifier $\hat{f}2$. The ten classes of anatomy are thoracic aorta, chest, brachial plexus, liver, bone, eye, heart, blood vessel, dermatome and coronary arteries.

5.6.4 Pathology Model

The final classifier deals with 12 classes. Figure 5.16 shows the pathology concept learning stage to generate first-level classifiers. The training set of 12 classes, L3 = {(x1, y1), (x2, y2), …, (x12, y12)} where each *yi* is a label of the class associated with each *xi*, are used to train the level 1 classifier $\hat{f}1$.

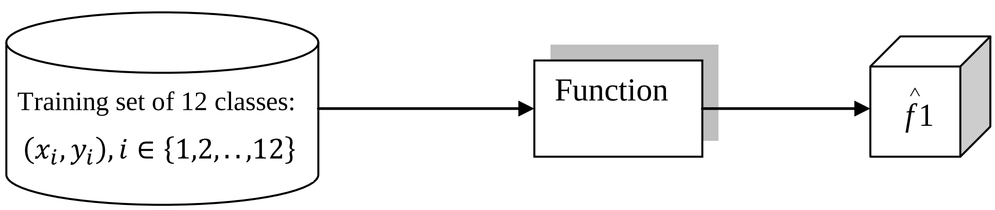


Figure 5.16: Pathology Model Generation

The 12 classes of pathology are thoracic aortic dissection, acute myloid leukemia, heart failure, brachial plexus nerve block, fatty liver, greenstick fracture, streptococcus pneumoniae, papilledema, pericardial effusion, atherosclerosis, sacral fracture and dermato fibroma

Once the $\hat{f}1$, $\hat{f}2$ and $\hat{f}3$ classifiers are generated, the training is assign with the semantic label based on its class name. When unlabelled image is given, first the visual features of texture, shape and color both in global and local descriptors are extracted to form one vector. Then, $\hat{f}1$, $\hat{f}2$, and $\hat{f}3$ classifiers are used to predict classes for each concepts. The evaluation is based on the correctness rate to classify the unlabeled image in the accurate class.

5.7 Classification

Generally there are two types of classification namely supervised and unsupervised classification. In the unsupervised methods, no target labeled is identified in the data and data mining algorithm is used to search for patterns and structures among all the variables. Unlike unsupervised technique, supervised method represents such particular pre-specified target variable and the algorithm is given by many examples where the value of the target variable is provided, so that the algorithm may learn which values of the target variable are associated with which values of the predictor variables (Johnson, 1998). During the testing of a feature in an image, each of the classifier votes for one class and the winning class is the one with the largest number of accumulated votes (Rahman et al., 2012). In this research, supervised classification method is utilized and two types of classifiers are applied namely k-NN and SVM. K-NN and SVM are widely used in medical image classification (Lin et al, 2006; Rahman et al, 2008; Zhang et al, 2009).

Figure 5.17 shows the example of modality classification result. The feature vectors from the storage (example in Figure 5.12) will be trained and modelled based on medical concepts in MedHieCon model. This figure shows seven classes of modality are modelled namely PET, XR, PX, US, CT,GX and MR and the evaluation result is based on TP Rate (true positive), FP Rate (false positive), Precision, Recall, F-measure, ROC area and confusion matrix. However in this research, we will concentrate on precision and recall values.

```
=== Run information ===

Scheme:       weka.classifiers.functions.SMO -C 1.0 -L 0.0010 -P 1.0E-12 -N 0 -V -1 -W 1 -K
Relation:     modality
Instances:    2200
Attributes:   77
Number of kernel evaluations: 10713 (80.705% cached)


=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.85      0.008      0.895     0.85      0.872      0.953    PET
                 0.803     0.096      0.63      0.803     0.706      0.892    XR
                 0.043     0.001      0.667     0.043     0.08       0.76     PX
                 0.766     0.049      0.709     0.766     0.736      0.94     US
                 0.481     0.061      0.587     0.481     0.529      0.856    CT
                 0.977     0.002      0.994     0.977     0.986      0.996    GX
                 0.522     0.118      0.494     0.522     0.508      0.805    MR

=== Confusion Matrix ===

   a   b   c   d   e   f   g   <-- classified as
  85   8   1   2   1   2   1 |   a = PET
   3 187   0   9  15   0  19 |   b = XR
   1   6   2   3  10   0  25 |   c = PX
   0   5   0 141  12   0  26 |   d = US
   1  30   0  16 101   0  62 |   e = CT
   5   1   0   1   1 345   0 |   f = GX
   0  60   0  27  32   0 130 |   g = MR
```

Figure 5.17: Example of multi-modality image classification result for medical concept of modality

5.7.1 k -Nearest Neighbour (k-NN)

k-Nearest Neighbour is one of the oldest and simplest supervised method for pattern classification (Peterson, 2009). Nevertheless k-NN is among preference classifier used by many researchers due to its positive competitive results, and in certain domain it has significantly advanced the state-of the-art (Wang, 2006). k-NN is a distance-based system. Initially, the example images will be labeled and trained and the training dataset are stored, and new unlabeled data is classified by comparing them with the most similar records in the training set (Hinneburg et al., 2000) and k-NN classifier remembers all training data and selects most similar vector at the moment it is asked to make a prediction.

k-NN methodology is used to compute the distance between two data's (the new unlabeled data to be classified and the training dataset). Distance Euclidean measurement technique is the most common measurement metric. The measurement distance between two data's can be defined as

$$d(p,q) = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2} \tag{5.17}$$

Where $p = (p_1, p_2, \ldots p_i)$ and $q = (q_1, q_2, \ldots q_i)$ are two points in Euclidean n-space. The best choice of k depends upon the data; generally, larger values of k reduce the effect of noise (irrelevant feature) on the classification, but make boundaries between classes less distinct. For example in Figure 5.18 consider figure (a) where k = 1, the feature vectors of query image, point X, belongs to class 3 (green circles). However if k = 5 as in (b) the point X best-fit in class 1 (blue circles) according to majority vote of the five nearest points.

The application of parameter optimization, for example, cross validation can be used to increase the performance of k-NN by selecting a suitable value of k (Duch et al., 2008).

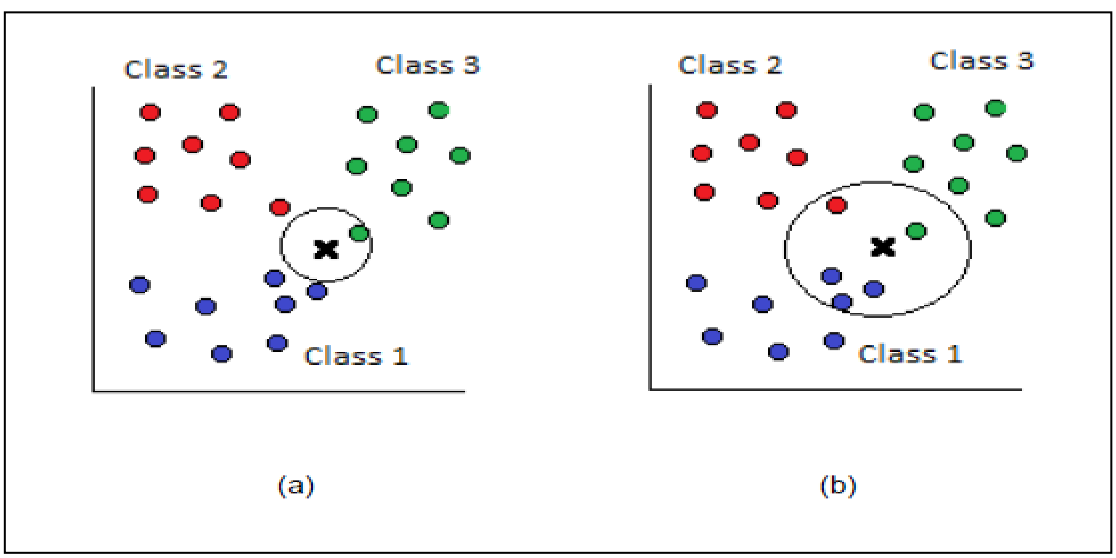


Figure 5.18: Example of (a) 1-NN classifier and (b) 5-NN classifier in k-NN classification

The difference between k-NN and other classifiers is that in the case of k-NN, training points are used during the classification, whereas in other methods, usually the training points are needed only during the training.

5.7.2 Support Vector Machine (SVM)

SVM is a supervised machine learning method in statistical learning theory which is performed to predict the output from a given input (Takeuchi, 2005). Generally SVM is a trainable machine classifier that gave two-class training set and attempts to specify a maximum-margin separating hyperplane between the data points of the two classes. This hyperplane is optimal in the sense that it generalizes well to unseen data. In this research work, we applied supervised learning method which the process is to find a function that suitable to describe the relation between the input data and the output data. For SVM's binary class case output prediction function can be defined as

$$f(x) = sign(\sum_{i=1}^{m} \alpha_i y_i K(x_i, x) + b) \tag{5.18}$$

where $x_i, i = 1, \ldots, m$ support vectors which represent the selected training examples and x are the input vector. $K(x_i, x)$ is a symmetric positive function which known as kernel, y_i the vector label of (1, -1) and α_i represent as support vector's weight determined during the training process and b is the bias of the hyperplane. There are several different types of kernels namely linear, polynomial and Gaussian and they are shown in Figure 5.19 (Chang & Lin, 2011). Polynomial and Gaussian are non linear kernels and suitable for application if there is no linear relation between the labels and the input which support the non-linear problems.

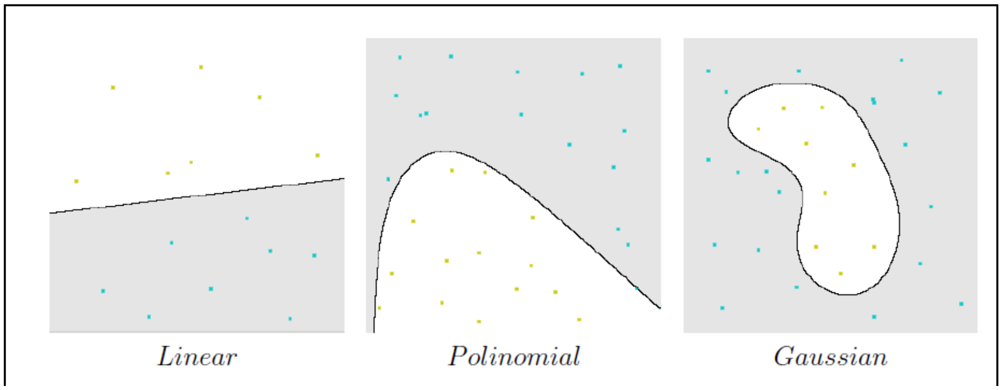


Figure 5.19: Examples for kernel's classifier representation taken from LibSvm (Chang & Lin, 2011)

There are several advantages using SVM as classifier which are (i) it is low computational cost and better generalization due to only few vectors are selected to become the support vectors in the training set during training process and (ii) the solution is not dependent on the starting conditions unlike neural networks. A more detailed description of SVMs are found in Burges, (1998) and Vapnik, (1998). To increase the performance of SVM, fast iterative algorithm namely Platt's sequential minimization algorithm (SMO) is used (Platt, 1999).

Many datasets encountered in medical domain and other areas of application are unbalanced, for example; one class contains a lot more examples than the other. (Ben, A., 2010). SVM is chose to solve the problem of imbalanced data because SVM is based on strong theoretical foundations (Vapnik, 1995) and it performs well with moderately imbalanced data even without any modifications (Akbani, R.,2000). Its unique learning mechanism makes it an interesting candidate for dealing with imbalanced datasets, since SVM only takes into account those instances that are close to the boundary, i.e. the support vectors, for building its model. This means that SVM is unaffected by non-noisy negative instances far away from the boundary even if they are huge in number. In this research SVM is used to classify multiclass. There are several approaches to adopting SVMs to classification problems with multiclass classification which are multiclass ranking, one-against-all classification and pairwise classification. Multiclass ranking SVMs, in which one SVM decision function attempts to classify all classes. One-against-all classification, in which there is one binary SVM for each class to separate members of that class from members of other classes. Pairwise classification, in which there is one binary SVM for each pair of classes to separate members of one class from members of the other. However in this research pairwise classification is used since this method produce more accurate results on the data set (Abe, S., 2003).

5.8 Summary

This chapter explains the application of CBIR in medical domain. M3ICS content-based framework is introduced in this chapter which contains feature extraction and training data phase. For feature extraction component, global and local descriptors are utilized, global level is to extract visual features (texture, shape and color) for the whole medical image. In contrast local level is based on patches and interest points of medical images. Various sizes of patches are applied in order to evaluate which size is the best to extract medical information. Furthermore, 20 interest points are generated to create 20 blocks size 20×20 pixels of interest region and no image segmentation processed needed. These features will be extracted and stored as feature vector. These feature vectors are labelled based on MedHieCon model which include modality, anatomy and pathology. Supervised classifiers are used to train the labelled data in concept models namely modality, anatomy and pathology and classified the new unlabelled data as test data based on the models. The evaluation is based on how accurate the test data are correctly classified. The percentage of correctness rate is used to evaluate the performance of our M3ICS content-based framework.

6.0 IFM3IRS: Information Fusion of Text and Content-based Systems

6.1 Introduction

M3IRS text and M3ICS content-based frameworks are introduced in chapter 4 and 5. Both chapters explain the methodology of extracting features for medical documents and images independently. Although text-based retrieval in MIR outperforms content-based framework, it is significant to determine the overall performance improvements which include both text and content-based (Muller et al., 2010). Moreover, there is disadvantage for each of these feature frameworks. As for the text-based framework, it is difficult to filter out which images in the relevant documents retrieved which do not represent the right modality. For example Figure 6.1 shows the list of images resulted from text-based framework with the query of *"CT images containing fatty liver".* However, it clearly shows that there are several images which are not CT scan images. The list includes PET, microscopy, graph and US images as highlight in red circles. As for content-based framework, it is difficult to differentiate the images based on specific detail such as pathology concept. The results retrieved from content-based framework for the same query might be all CT scan images of liver but not all dedicated to fatty liver disease. This chapter introduces information fusion framework (IFM3IRS) based on the combination of text and content-based information sources. Combination of text and visual information potentially allows a reduction of the semantic gap, which is the modelling representation between human observation for a particular image and visual information.

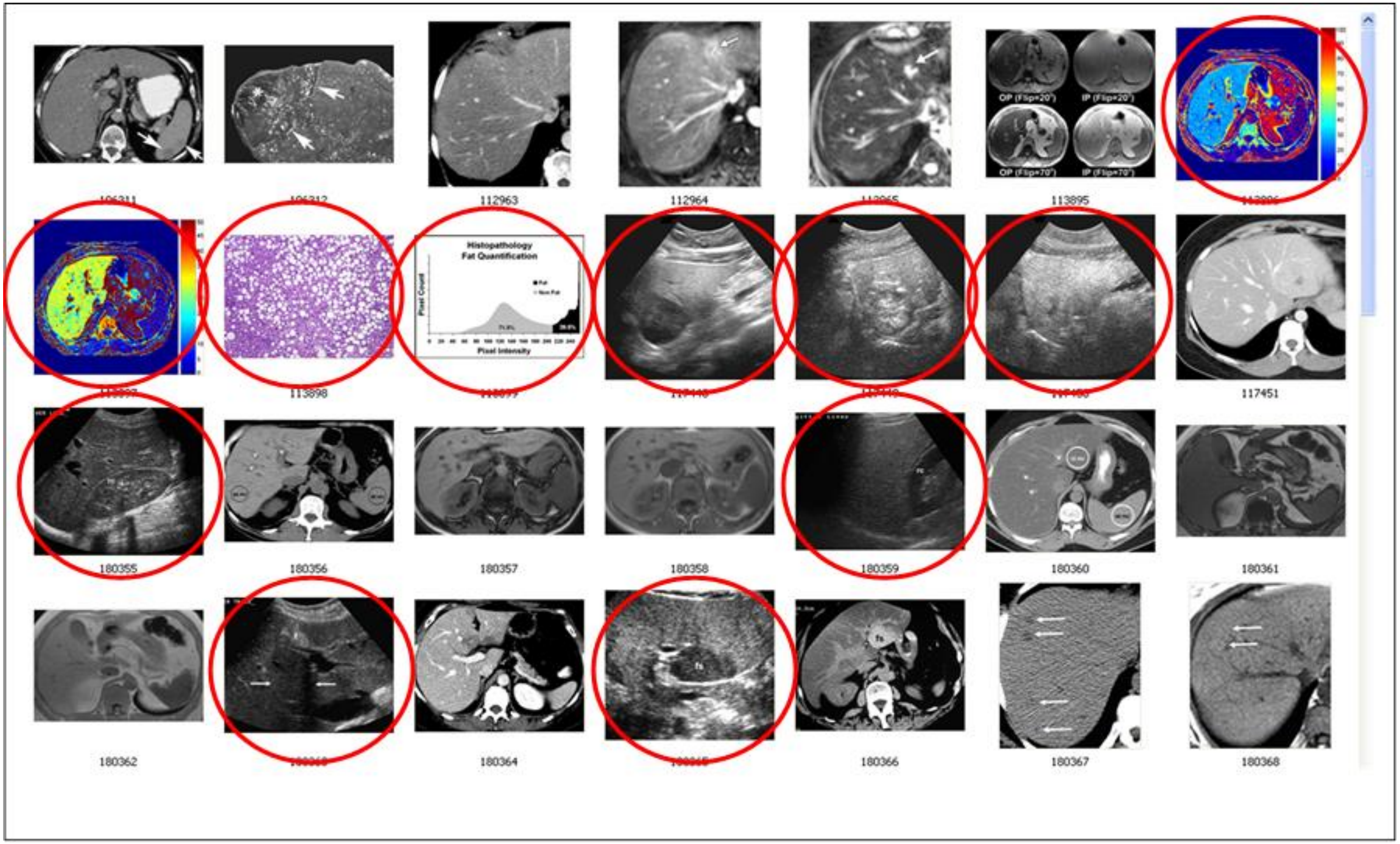


Figure 6.1: Images from list of RMR for "CT images containing fatty liver"

6.2 Information Fusion M3IRS Framework (IFM3IRS)

IFM3IRS framework concentrates on fusion of the information sources of text and visual content using hierarchical processing in late fusion technique. The framework is proposed to fuse both text and content-based information sources in order to increase the accuracy in medical image retrieval. Information fusion is a revision of efficient method to transform information from different sources and different points in time into a representation that leads to effective support for decision making. Combination of text and visual information allows a reduction in semantic gap for more meaningful search and retrieval and to improve the performance effectiveness. Early fusion approach (feature fusion level) is conducted as text and visual feature attributes concatenate in one vector to generate one unique feature space. The major disadvantage is the large dimension of vector which contributes to scatter the homogeneous clusters of instances of same concept. For hierarchical processing in late fusion approach information source is processed first at one level and the output of this level will be processed to another level that is unit level of text and content-based separate execution. As textual retrieval performs better than content-based retrieval, the text-based retrieval is executed at the initial level and the results from this will be the input for the next level which is content-based retrieval. The purpose to extract visual features from the images in text-based results is to filter the modality. Late fusion strategy of reordering technique is used to increase effectiveness whereby it is based on reordering of documents to gain final ranking list. The textually-retrieved documents produce the final ranking based on Medical Hierarchical Conceptual (MedHieCon) Ranking Model for the medical semantic effectiveness and visual retrieval is based on textual scores reordering. Figure 1.5 illustrates the framework which is based on hierarchical processing where the system executes text-based retrieval and the result from the text-based processing will automatically be the input for content-based classification. Figure 6.2 represents the flow chart of IF3MIRS framework whereby the blue shapes represent text-based retrieval and green shapes represent content-based retrieval. Note that the fusion in this research is based on two processes executed separately which combine based on FigureID of medical document and image. Which means it is hierarchical processing where the M3IRS text-based is executed first, and then the result from M3IRS is passed to M3ICS based on FigureID to import the medical image to the M3ICS process.

As previously mentioned in chapter 4, the output of M3IRS text-based retrieval system is the *RMR* list which consists of Figure ID, title, caption, score and attribute list of *MEQSize* as shown in Figure 6.3. From the *RMR* list, the IFM3IRS system will automatically import the medical images from the data collection based on Figure ID as highlighted in red font in the figure as input data for M3ICS visual engine. Note that the figure below shows only the extraction of the actual result to represent the example in this section. The actual value for "CT images containing fatty liver" retrieval list is 34,126 documents.

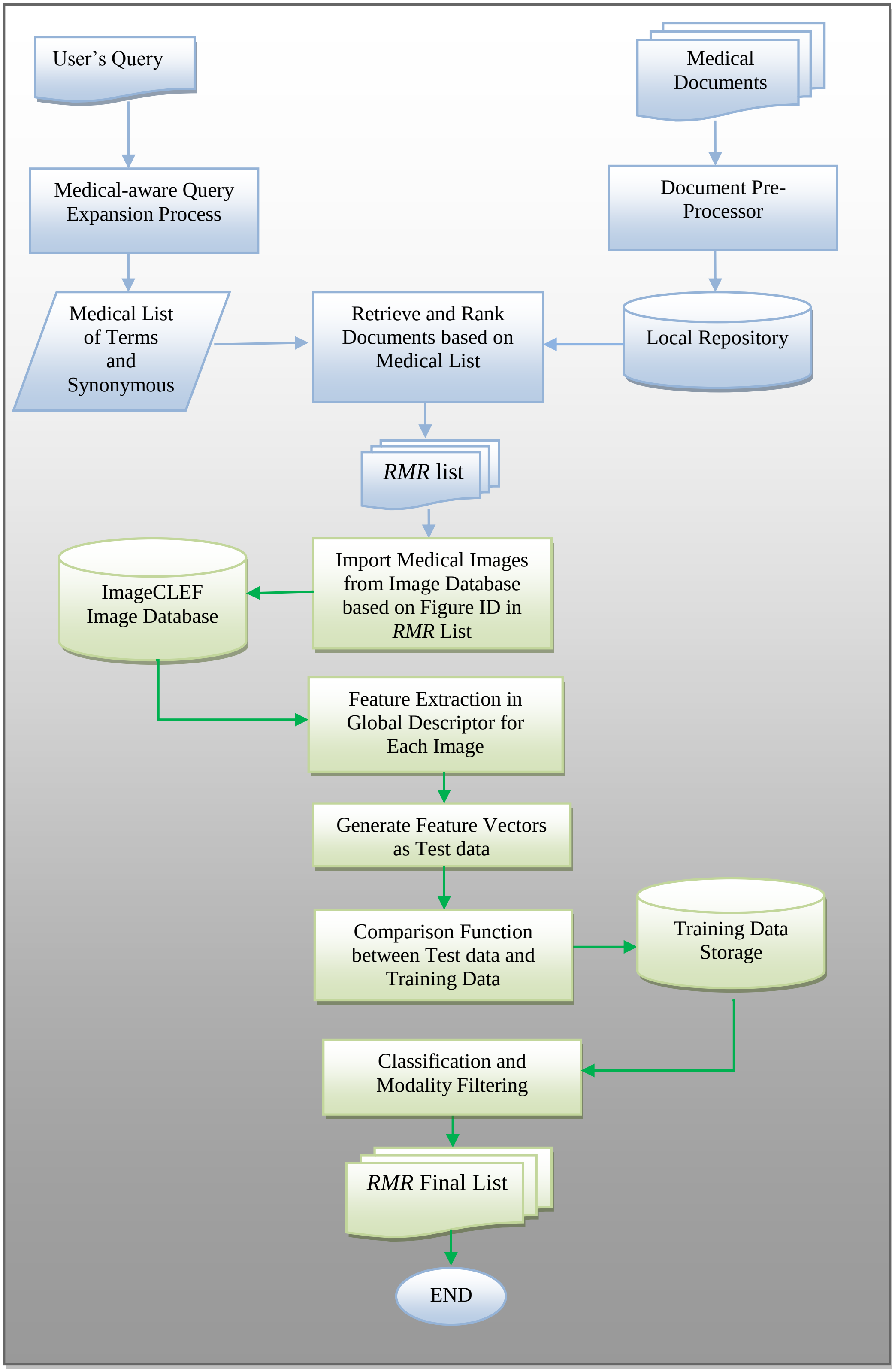


Figure 6.2: Flow chart of IF3MIRS framework

Figure ID	Title	Caption	Score	MEQSize=1	MEQSize=2	MEQSize=3
36507	nondiffuse fatty chang	figure 2a patient 5 foca	6	fatty liver:2	liver:3	ct:1
112963	metastases in candid	figure 7a false-negative	5	fatty liver:1	ct:3	liver:1
180378	imaging patterns and	figure 14adifferentiatio	5	fatty liver:2	liver:3	
106311	littoral cell angioma o	figure 3a patient 7 litto	2	steatohepatitis:1	ct:1	
28080	obstruive sleep apnea	figure 1 lateral scout	1	fatty liver:1		
37377	value of iterative reco	figure 2a images in a 5	13	liver:9	ct:4	
177922	atrophy-hypertrophy	figure 2a pv in 70-year-	5	liver:2	ct:3	
106151	recurrent hepatocellu	figure 1d transverse co	3	liver:2	ct:1	
28055	hepatic lesions morph	figure 7d hepatocellul	3	liver:3		

Figure 6.3: RMR list output from M3IRS text-based framework

This input acts as a series of unlabeled images that will be the testing data for the content-based framework. The main purpose to fuse text-based with content-based frameworks is to filter the modality of medical images. Therefore, for IF3MIRS framework the system will extract the visual features of testing data based on global descriptor which only filters the modality concept of the images. Then the testing data will be classified using the classifier (detail in section 5.7) and compared with training data in the training data storage.

In this process the system will only filter the accurate modality and select the testing data based on the modality request from the query. Prior to that training data of medical images with seven different modalities are trained and modelled and will be stored in training data storage as described in section 5.6.1. The test data are classified based on these models. Test data that belong to the query class will remain in *RMR* list while others will be discarded.

Figure 6.4 shows the medical images imported from the Figure ID listed in Figure 6.3 to be the test data for content-based framework, and Figure 6.5 illustrates the feature vectors extracted from the image.

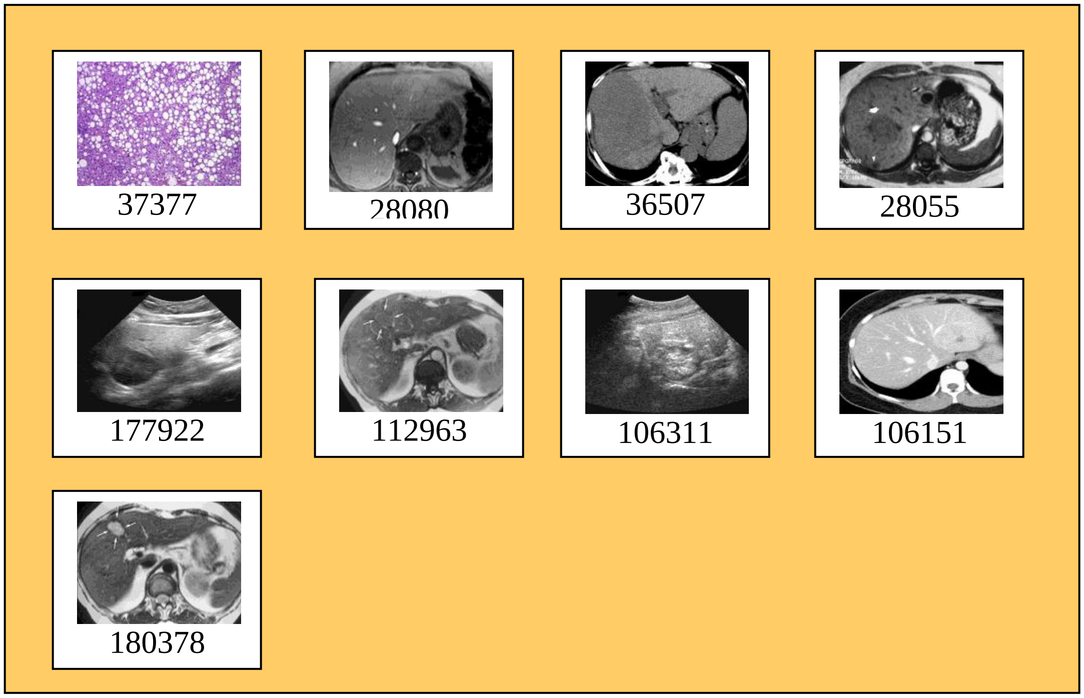


Figure 6.4: RMR list input of images to M3ICS content-based framework

```
@RELATION test_data

@ATTRIBUTE name NUMERIC
@ATTRIBUTE 1 NUMERIC
@ATTRIBUTE 2 NUMERIC
@ATTRIBUTE 3 NUMERIC
@ATTRIBUTE 4 NUMERIC
@ATTRIBUTE 5 NUMERIC
@ATTRIBUTE 6 NUMERIC
@ATTRIBUTE 7 NUMERIC
@ATTRIBUTE 8 NUMERIC
@ATTRIBUTE 9 NUMERIC
@ATTRIBUTE 10 NUMERIC
@ATTRIBUTE 11 NUMERIC
@ATTRIBUTE 12 NUMERIC
@ATTRIBUTE 13 NUMERIC

@ATTRIBUTE class {x-ray, PET, PX, US, CT, GX, MR}

@DATA
36507,0.316617,12.3185,4.0156,0.654379,0.625266,-0.628048,1.15621,0.439081,0.866713,1.11845,0.602911,1.18782,0.9
112963,0.314612,27.1114,0.50801,0.319162,0.334957,-0.330022,-0.34268,0.205243,1.26396,1.68331,0.862423,1.6129,0.
180378,0.0507487,0.01696,0.0476753,0.884616,36.9562,1.19815,0.640459,0.403371,-0.443805,-0.574051,0.356671,1.449
106311,0.341729,0.316617,0.2419,0.0997545,10.5091,0.834909,0.399673,0.524268,-0.435675,-0.474931,0.460401,0.2433
28080,0.197448,0.340455,0.393059,0.0690392,15.7671,0.200237,0.481668,0.319431,-0.251502,-0.229522,-0.351013,0.39
37377,0.443789,0.314612,0.179213,0.0623864,8.92621,1.13852,0.116562,0.387056,-0.280442,-0.545026,0.228468,0.5946
177922,0.299638,0.53689,0.158711,0.00476003,15.8535,2.97031,0.635816,0.737092,-0.684219,-1.15202,-0.600004,0.174
106151,0.200915,0.281554,0.427353,0.0901782,15.8282,1.16173,0.707096,0.365216,0.380566,-0.425868,0.406562,0.2846
```

Figure 6.5: Feature vectors of test data

Each feature vector listed in the Figure 6.4 will be classified based on seven classes of modality namely x-ray, PET, PX, US, CT, GX and MR. In this example, only feature vector that is classified in CT will remain and others will be discarded. Therefore the final output of IF3MIRS framework will be *RMR* final list which inclusive of medical information (image and title of image) that have been classified in CT modality as depicted in Figure 6.6. It shows that most images retrieved are actually in accurate modality which is CT of liver except for image of Figure ID 106311 which is of US modality. In conclusion it shows that there is improvement in combining two different sources of text and visual content in retrieving multi-modality medical images where for text-based framework the system can emphasize more on anatomy and pathology concepts and in addition the content-based framework is to filter the modality of medical images.

6.3 Summary

This chapter introduces IFM3IRS information fusion framework which is based on fusion sources of M3IRS text and M3ICS content-based frameworks that are adapted from Chapter 4 and Chapter 5. Late fusion technique is used in information fusion methodology which is based on hierarchical processing where the output from text-based framework will be the input to content-based framework from which the output from this framework is then the list of medical images with modality according to query. The final output of IFM3IRS is the *RMR* final list of text and image medical information.

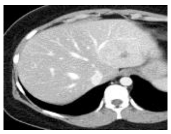
IF3MIRS OUTPUT

112963

Fatty liver: imaging patterns and pitfalls

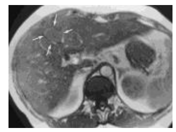
36507

Nondiffuse fatty change of the liver: discerning pseudotumor on MR images enhanced with ferumoxides-initial observations



10151

Blood flow patterns in focal liver lesions at microbubble-enhanced



112963

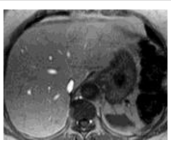
Liver metastases in candidates for hepatic resection

106311

Nonenhanced CT scan shows multiple hyperattenuating lesions in a hypoattenuating fatty liver

180378

Blood flow patterns in focal liver lesions at microbubble-enhanced



28080

False-negative lesion on CT scan in a 65-year-old woman with a diffusely fatty liver and multiple metastasesobservations

Figure 6.6: Final output of medical information retrieved from IF3MIRS framework

7.0 Experimental Setups and Framework Evaluation

7.1 Introduction

There will be two major experimental setups namely text features experiments and visual features experiments. The experiment used ImageCLEF 2010 medical data collection (Müller H., 2010) supplied by the Cross Language Evaluation Forum (CLEF) (Peters Ed., 2000). This data collection consists of 77,500 medical test documents in XML format and images in .jpg format. Sixteen ad-hoc queries with relevance judgment data containing relevant documents for each query are provided for the evaluation. These sets of relevant judgments are assumed as the standard in our experiment and referred to as the relevant set in both text and visual experiments. This chapter describes the experimentation setups for text and visual features followed by the description of evaluation criteria.

7.2 Experimental Setups

Several experimental setups were prepared to conduct experiments on both text and content-based frameworks. Prior to that multi-modality medical image characteristics are evaluated based on contrast, noise and blur. These experimental setups can be divided into four groups namely multi-modality evaluation characteristics, text features, visual features and information fusion experiments.

7.3 Quality Characteristics in Multi-modality Medical Image Evaluation

In chapter 2, the description of medical image characteristics is discussed. Generally the best image quality is defined based on contrast, blur and noise values. Different modality medical image shows different level of quality characteristics. The experiment is to evaluate these characteristics in various modalities of medical images using SVM classifier. The next experiment is to extract texture, shape and color features based on local descriptor. The methods used to extract features as described in section 5.3. The performance will be evaluated based on the formula correctness rate, which is the percentage of correctly classified image divided by total number of images and also the rate of precision and recall of each modality.

Eight modalities are used in this experiment namely x-ray, CT, US, PET, NM, MRI, PX and GX. Medical task of ImageCLEF 2010 dataset is used and 2450 medical images (see Table 7.1) will be classified using SVM linear kernel. Pairwise classification technique is used to classify multiclass classification. Classifier is built for every pair of classes, using only the instances from these two classes. Logistic regression model is applied to produce probability estimates from the different classifiers (Perlich et. al, 2003).

Table 7.1: Image Modalities and number of training images

CT	314
GX	355
MRI	299
PET	285
PX	330
US	307
x-ray	296
NM	250

7.4 Text Features Experiment

In chapter 4, we have explained the framework of M3IRS text-based retrieval system which includes four main components which are (i) document pre-processor, (ii) query processor, (iii) retrieval strategies and (iv) ranking strategies. The experiments for text features in this study are based on these components. The setup begins with the experiment on the performance of document pre-processor component and end up the experiment with the comparison between M3IRS with other study that used the same data collection. We also evaluate the time performance of M3IRS using Mesh-indexer.

7.4.1 Document Pre-processor Experiment with XTE Method

As mentioned earlier, the medical documents used in this research are in XML format. There are 77,500 medical documents available in ImageClef 2010 medical task collection. Some of the documents may have duplication in the caption and title due to the description of a series of medical images in the same article. Therefore XTE method is used to simplify medical documents to be more structured and standardized with no duplication (see section 4.4). In preparation of this experimental setup the ImageCLEF medical documents are pre-process using XTE method. Only several DTD were used in indexing the new medical document namely <figureID>, <title> and <modifiedCaption>. ModifiedCaption is a new section of caption for medical document which represent sentences for particular figure and remove the redundancy from other medical document. All indexed medical documents will be stored in new local repository.

To verify the effectiveness of XTE method, we implement the experiment to evaluate the proportion of relevant documents retrieved between the original ImageCLEF medical documents and our new indexed medical documents using XTE method. The evaluation is based on MAP value for both data collection.

7.4.2 Time Performance Experiment with Mesh-indexer

MeSH-indexer is developed to reduce access time in searching medical terms during query expansion process. MeSH-indexer is organized according to alphabetical order. To prove the efficiency of M3IRS using MeSH-indexer, an experiment is conducted to compare the time performance in accessing medical terms between using the MeSH-indexer and the original MeSH thesaurus. In terms of size, the MeSH thesaurus file is 276MB and for three folders in MeSH-indexer (MMD, MST and MC), the total size is 29.2MB.

The evaluation is based on how fast (in seconds) the system can produce medical query list from medical-context aware query expansion technique in M3IRS.

7.4.3 Query Processor Experiment with Medical Context Aware Query Expansion Technique

The purpose of this experiment is to verify the effectiveness of medical context aware query expansion technique. The evaluation is based on MAP result for each type of query. This means that if the MAP value is high and approaches to value 1 it shows that this method is very effective. This is because MAP value represents the most relevant documents ranked in the higher position. There are three types of query expansion techniques used in this experiment. The techniques are as follows:

i. *Type 1: Direct matching with original query*

   Evaluate proportion of relevant documents retrieved using original query.

ii. *Type 2: Expansion*

   Evaluate proportion of relevant documents retrieved based on medical terms extract from original query. The process of medical term extraction is explained detail in Section 4.3.1.

iii. *Type 3: Expansion and Enrich*

   Query Type 3 includes of the enrichment of medical term which is the synonymous of medical term identified in query Type 2. Evaluate proportion of relevant documents retrieved by combining both terms in expansion and enrichment (synonymous term).

7.4.4 Ranking Strategies Experiment

The purposed of ranking strategies experiment is to evaluate the performance between comprehensive ranking (see section 4.6.1) and MedHieCon ranking (see section 4.6.2) models. Comprehensive ranking method prioritized on the size of term matched between medical query and document followed by number of occurrence terms matched in a document. The documents are ranked in decrement order.

The MedHieCon ranking model is used to rank the retrieved medical documents based on medical concepts entities of pathology, anatomy and modality. The bottom top hierarchy has prioritized the concept of pathology followed by anatomy and modality. The medical concepts are map to identified medical terms for ranking purposes. Each result in the experiments is viewed in order to verify the expected improvement. The evaluation will be based on average precision value of each query and MAP to evaluate overall result for each ranking models.

The results will be compared to other research that used the same data collection. The evaluation is based on relevance judgment obtained from the ImageCLEF 2010 collection. MAP value is used for the comparison since this method is a standardized measurement evaluation in most CLEF researches. Moreover MAP is a more stable measurement compared to other measurement such as precision and R-precision since it can assess more information (Buckley, 2000).

7.5 Visual Features Experimental

In chapter 5, we have explained the M3ICS content-based framework which involves feature extraction in global and local descriptor. Furthermore the data are trained based on bottom top MedHieCon model of pathology, anatomy and modality. The data in the collection will be trained and modelled based on these medical concepts. k-NN and SVM classifiers are used for classification purposed. The experiments for visual features involved the performance of each feature of texture, shape and color followed by global and local descriptors performance.

The evaluation of MedHieCon model performance using classifiers of k-NN and SVM are performed separately. To evaluate the performance of visual features, MedHieCon model and M3ICS content-based framework, we used 14,537 relevant judgement medical images provided by ImageCLEF 2010 medical task as test and train data. The data is based on 16 ad hoc queries (Q-1, Q-2, ..., Q-16) as the queries in Appendix A and example of images are listed in Appendix C. Table 7.2 shows the number of train and test data for each query used in this experiment with the ratio of 75% of training and 25% of testing data. The ratio is chose based on the experiment handled in section 7.5.2. In order to support imbalanced data the majority class can be subsampled (as for Weka used filter SpreadSubsample) and oversampling is for the minority class, creating synthetic examples (as for Weka used SMOTE). Other alternative is to make the classifier cost sensitive (as for Weka used metaclassifier CostSensitiveClassifier). However, each of the methods has its own strengths and weaknesses and need to be experimented to perform better results (Hall et al, 2009).

Table 7.2: List of training and testing data for each query

Query List	Training Set (75%)	Testing Set (25%)
Q-1	644	215
Q-2	464	155
Q-3	594	198
Q-4	735	245
Q-5	617	206
Q-6	704	235
Q-7	785	262
Q-8	683	228
Q-9	819	273
Q-10	644	215
Q-11	875	292
Q-12	631	210
Q-13	657	219
Q-14	557	186
Q-15	619	206
Q-16	873	291

7.5.1 Experiment in Selecting Optimum Value for Training and Testing Data

Prior to evaluate the performance of visual features, the optimum size of training and testing data needs to be determined. In this experiment the optimum percentage of test and train data need to be determined. Cross-validation technique (Browne, 2002) is used to segment the training and testing data. Each image is extracted by combination features of shape, texture and color in one feature vector which consist of 75-dimensional in global descriptor. The feature vectors of all data are trained based on 16 classes are which based on 16 ad hoc queries as listed in Appendix A. Then the data will be evaluated based on two types of classifiers namely k-NN with k=1 and SVM using polykernel. We experimented seven different categories of test data and train data percentage as listed in Table 7.3.

Table 7.3: Classifiers Experiment with k-NN and SVM

Category	Test data (%)	Train data (%)
1	10	90
2	20	80
3	25	75
4	50	50
5	75	25
6	80	20
7	90	10

7.5.2 Classifiers Experiment with k-NN and SVM

This experiment is carried out to evaluate the performance of k-NN and SVM classifiers. Three experiments are executed in this section. The first experiment is done to determine which k value in k-NN method is the best for this research study. The second experiment is performed to evaluate which kernel application is suitable in order to obtain better result in SVM classifier as we compared kernel of polykernel (Burges, 1998) and RBF (Burges, 1998). The final experiment is carried out to make comparison between k-NN and SVM classifier is performed. The same data and setup is used from previous experiment in section 7.5.1.

7.5.3 Primitive Visual Features Experiments with Texture, Shape and Color Features in Global Descriptor

As mentioned in chapter 5, we extracted texture, shape and color features for global descriptor. The combination of these features produced 75-dimensional feature vector for each medical image. For texture feature, we used GLCM approach which concentrates on contrast, correlation, energy and homogeneity of the image and the extraction is based on four directions which are 0^{o}, 45^{o}, 90^{o} *and* 180^{o}. As for shape feature, we employed Hu's moment invariant method, which invariant to translation, rotation and scale while HSV histogram is applied to represent color feature. These features are extracted for the whole image in global descriptor. The evaluation is based on MAP and percentage of correctness rate between two classifiers of k-NN with k=3 and SVM using polykernel.

7.5.4 Local Level Evaluation with 2×2, 4×4 and 8×8 Patches and Interest Blocks

This section describes local descriptor performance which involves patches and interest blocks experimental. In local descriptor feature extraction, no rotation involved. For patches experimental work, there are three experiments. Each experiment is done to evaluate the performance of divided medical image into 2×2, 4×4 and 8×8 patches. The experiments use the combination of texture, shape and color features. The evaluation is based on which size of patches performs the best in this experiment.

The next experiment in local descriptor which is carried to extract visual features based on interest blocks. Prior to that, Harris detector (Mikolajczyk et. al, 2002) method is used to extract interest points in each medical image. For interest blocks, color feature extraction is not applied due to difficulties on converting small patches into HSV format. This is because each patch will turn into brown-scale color and produce inaccuracy in the result. For each interest point, a square block of size 20×20 pixels is created as interest point to be the centre point. Then texture and shape features are extracted in these blocks and then combine all features in one feature vector. The maximum of 20 interest points will be produced for each medical image. This is done to avoid large dimension of feature vector for each medical image.

Finally the performance of patches, interest blocks and combination of patches and interest blocks are compared.

7.5.5 Evaluation on the Performance of MedHieCon Model

In section 5.6, the MedHieCon model for visual features is explained. This model is a bottom top diagram which illustrates general information on top of the diagram down towards lower level of the pyramid is more specific descriptions. The illustration is described as modality on top of diagram, followed by anatomy and the bottom part is pathology. Two classifiers are used namely k-NN with k=3 and SVM using polykernel to train data into concept classes and test the concept based on classification.

In this research, the application of MedHieCon model is based on 16 ad hoc queries provided by ImageCLEF 2010. Table 7.4 express the list of 16 ad hoc queries and concepts assign for each query. For example Q-1 “CT images thoracic aortic dissection”; the concepts involve is “CT” for modality, “thoracic aorta” for anatomy and “thoracic aorta dissection disease” for pathology. For those queries that have information of modality and pathology medical concepts such as Q-7 “X-ray images of a greenstick fracture” which only state “X-ray” for modality and “greenstick fracture” for pathology; we will defined the anatomy concept manually which is categorized the anatomy as “bone”. Other queries that we manually defined the anatomy concept are Q-4, Q-6, Q-9 and Q-10.

There will be 14,573 medical images are used for training data and will be modeled for three concepts namely modality, anatomy and pathology. These images are taken from the relevant judgment data from the ImageCLEF 2010 medical task collection. This is because all of these medical images are label images. Therefore for modality concept will have 7 models, anatomy concept will have 10 models and pathology concept will have 12 models. For modality concept, the extraction will be based on global descriptor process (as shown in Figure 5.3). Meanwhile, for anatomy concept local descriptor is applied (as shown in

Figure 5.5). As for pathology concept, the combination of global and local descriptors will be used.

Table 7.4: List of 16 ad-hoc queries and the medical concepts

	Query ImageCLEF 2010 Medical Task	Modality	Anatomy	Pathology
Q-1	CT images thoracic aortic dissection	CT	Thoracic Aorta	Thoracic Aortic Dissection
Q-2	A microscopic image of Acute Myeloid Leukemia	PX Microscopy		Acute Myeloid Leukemia
Q-3	ECG images	GX ECG		
Q-4	X-ray showing congestive heart failure	XR	Chest	Heart Failure
Q-5	CT images for brachial plexus nerve block	CT	Brachial Plexus	Brachial Plexur Nerve Block
Q-6	CT images containing fatty liver	CT	Liver	Fatty Liver
Q-7	X-ray images of a greenstick fracture	XR	Bone	Greenstick Fracture
Q-8	Microscopic images streptococcus pneumonia	PX Microscopy		Streptococcus Pneumoniae
Q-9	MR images papilledema	MR	Eye	Papilledema
Q-10	MR images pericardial effusion	MR	Heart	Pericardial Effusion
Q-11	All types images with atherosclerosis in blood vessels		Blood Vessel	Atherosclerosis
Q-12	Radiation therapy treatment plans			
Q-13	Images of dermatome		Dermatome	
Q-14	Images showing sacral fracture		Bone	Sacral Fracture
Q-15	Images coronary arteries		Coronary Arteries	
Q-16	Images dermato fibroma		Dermatome	Dermato fibroma

7.5.6 Experiment to compare the performance of M3ICS and MIARS

The performance with our M3ICS content-based framework with MIARS methodology (Mueen, 2010) using ImageCLEF 2010 medical task data collection is compared. MIARS methodology involves of converting the medical image to greyscale and extract features in global and local level descriptors with pixel intensity information. For local descriptor, each medical image is divided into 2×2 patches and extract texture feature using GLCM and edge histogram (Park et al., 2000) for shape feature extraction and produced 53-

dimensional of feature vector for each medical image. The evaluation is based on MAP value.

7.6 Information Fusion Experiment

Hierarchical processing of late fusion technique is applied in this experiment where the result from text retrieval will be the input for content-based retrieval. The input data used in this experiment are explained in section 6.2 where 16 ad hoc queries are used. Initially, results of *RMR* list from experiment 7.4.4 of each query are taken as input. For each *RMR* list, the system import medical images based on figure ID in the *RMR* list and stored them into another folder. All these images are assigned as the test data on M3ICS content-based framework. The evaluation is based on MAP value. The comparison of text, visual and fusion is based on MAP values. The performance of M3ICS with other run system (as shown in table 3.3) is then compared.

7.7 Evaluation Criteria

There are many evaluation methods proposed in IR system to evaluate the effectiveness of query expansion (Hersh, 2003). Precision, recall, and MAP are used as evaluation values. The performance was first evaluated by recall (number of relevant documents in the collection retrieved by the query) followed by precision (number of relevant documents retrieved by the query). Then, MAP value is obtained by taking the mean value of average precision from all the queries. Average precision is considering the order of returned relevant documents in the ranked sequence.

7.7.1 Precision and Recall

In the field of information retrieval, precision is the fraction of retrieved documents that are relevant to the search which takes all retrieved documents into account (Manning et. al, 2008). Precision is also used with recall, which is the fraction of the documents that are relevant to the query that are successfully retrieved.

Let R be the set of relevant documents that have been establish by ImageCLEF collection and RMR denoted as the retrieved documents from M3IRS system. Then the precision and the recall of the system are calculated using the following equations:

$$precision\ (p) = \frac{|R \cap RMR|}{|RMR|} \qquad (7.1)$$

$$recall\ (r) = \frac{|R \cap RMR|}{|R|} \qquad (7.2)$$

It is trivial to achieve recall of 100% by returning all documents in response to any query. Therefore, recall alone is not enough and one needs to measure the number of non-relevant documents also, for example by computing the precision.

7.7.2 F-measure

F-measure is a formula to measure test's accuracy. The calculation involves precision p and recall r value as depicted in equation (7.1) and (7.2). F-measure score also can be represented as weight average for precision and recall where the best score at 1 and worst at 0.

$$F = 2\,\frac{(p.r)}{(p+r)} \qquad (7.3)$$

This is also known as the F_1 measure, because recall and precision are evenly weighted.

7.7.3 Average Precision

Since there are several queries involved in this experiment and the *R* documents for each query might not be the same amount; therefore average precision is the precision values at the positions of each *RMR* is retrieved. This is significant in order to ensure that the most relevant document will be ranked in higher position (Song et. al, 1999). Precision and recall are single-value metrics based on the whole list of documents returned by the system. Average precision computing a precision and recall at every position in the ranked sequence of documents. The calculation involves the position of *RMR* (*RMRpos*) where $RMR \cap R$. The average precision calculation is represented by equation (7.4) where r is the number of position and *rel(r)* is an indicator function equals to 1 if the item at position r is a relevant document and zero if r is a non-relevant document. Therefore the result is more precise where each document is measured based on its position in the list and we will know which medical document is the most relevant with the query. In contrast precision value only concentrates on the overall performance.

$$AveP = \sum_{r=1}^{N} \frac{(Rpos(r) \times rel(r))}{R} \qquad (7.4)$$

Further, equation 7.5 shows the measurement of the MAP for each queries where Q is number of queries.

$$MAP = \frac{\sum_{q=1}^{Q} AveP(q)}{Q} \qquad (7.5)$$

7.7.4 Percentage of Correctness Rate

The correctly and incorrectly classified instances show the percentage of training test instances that were correctly classified. The percentage of correctness rate instances is often called accuracy or sample accuracy which represents the accuracy of the method used in classifying data based on classifier. The calculation can be defined as

$$Percentage\ of\ Correctness\ Rate\ = \frac{Number\ of\ Test\ Data\ that\ Correctly\ Classified}{Total\ of\ Test\ Data} \times 100$$

(7.6)

For example given 100 example data as test data and during classification process only 70 data is correctly classified, therefore the percentage of correctness rate is 70%.

7.7.5 AUC under ROC

Due to imbalanced data in several modalities receiver operating characteristics (ROC) graph is used for result presentation. ROC is a two-dimensional graph for visualizing, organizing and selecting classifiers based on their performance without regard to class distribution or error costs (Zweig et. al, 1993). In signal detection theory, ROC is used to characterize the tradeoffs between hit rate and false alarm rate over a noisy channel (Egan, 1975). Generally in classification there are four outputs in a classifier and an instance. For example; if the instance is positive and it is classified as positive, it is counted as a true positive; and if it is classified as negative, it is counted as a false negative. As for negative instance and it is classified as negative, it is counted as a true negative; if it is classified as positive, it is counted as a false positive.

ROC graphs involve true positive (tp) rate plotted on the *y*-axis and false positive (fp) rate on the *x*-axis. True positive (tp) rate can be defined as

$$tp\ rate \approx \frac{positive\ correctly\ classified}{Total\ positives} \tag{7.7}$$

and false positive rate of classifier is classified as

$$fp\ rate \approx \frac{negatives\ incorrectly\ classified}{Total\ negatives} \tag{7.8}$$

Simple classification accuracy provide poor metric for measuring performance and therefore ROC graphs have been used in machine learning field for performance measurement (Provost, 1997). Area under the ROC curve (AUC) is used to represent ROC graph in single scalar value (Bradley,A.P., 1997). The calculation is based on the area of the unit square whereby the value is between 0 which indicate the worst classification and 1.0 which indicate the best classification performance. AUC performs very well and is often used when a general measure of predictiveness is desired.

7.8 Summary

This chapter describes experimental setups of M3IRS text, M3ICS content-based and information fusion frameworks. The experiment is based on components consist in the frameworks. The ImageCLEF 2010 medical task dataset is used and 16 ad hoc queries and relevant judgments are included. For M3IRS text-based framework, the experiments start with the evaluation of XTE method in document processor followed by time evaluation of the MeSH- indexer in M3IRS. Later, the query processor experiments which include two types of queries namely query expansion and query expansion and enrichment are performed. The final part for textual experiment is the comparison of ranking method between comprehensive and MedHieCon ranking models.

For content-based experimental work, there are five main experiments starting with the experiment to find optimum value of testing and training data. This optimum value is significant to determine the percentage of testing and training data and will be used for following experiments. Next is the comparison between two classifiers namely k-NN and SVM. This experiment is performed to validate which classifier has better performance in multi-modality medical image classification. Visual primitive features inclusive texture, shape and color will be experimented for both global and local descriptors. As for local descriptor the experiments includes the performance evaluation of 2×2, 4×4 and 8×8 patches and interest blocks. We introduced medical hierarchical concept to annotate the training data based on modality, anatomy and pathology.

Finally will be the IF3MIRS framework experiment. In this experiment setup, the output from M3IRS text-based framework will be the input of M3ICS content-based framework. The evaluation is based on the comparison and improvement of IF3MIRS with M3IRS text and M3ICS content-based framework. Furthermore the result will be compared with other researchers that used the same data collection.

8.0 Results and Discussion

8.1 Introduction

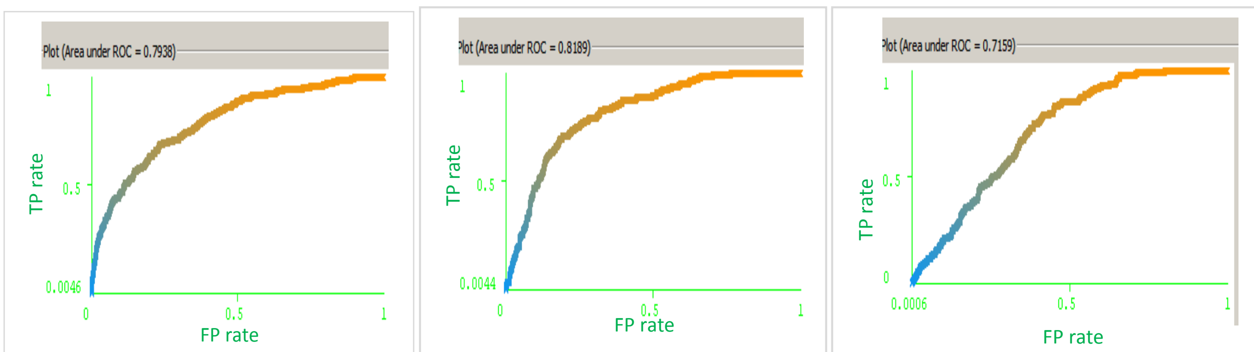
This chapter covers the results and discussions based on experimental setup explained in chapter 7. The experimental results were organised according to four main categories namely experimental for multi-modality medical image characteristic evaluation, text features, visual features and information fusion experimental. Different evaluation criteria is used as described in section 7.7.

8.2 Results of Quality Characteristics in Multi-modality Medical Image Evaluation

In this experiment, we analyzed the image quality characteristics namely contrast, blur and noise for eight modalities of medical image which are x-ray, CT, US, NM, PET, MRI, PX and GX. The methods used to measure these characteristics have been explained in section 2.2. The feature extraction is based on global level which the measurement is for the whole medical image. SVM classifier was used for classification and SMO (Keerthi et al., 2001) is applied for speedy training the instances of a support vector classifier for each modality. The output of an unknown test example was based on which class receives the most votes. The results shown are the ROC graph for individually contrast, blur and noise characteristics for each modality.

8.2.1 X-ray

Figure 8.1 shows the ROC graphs of contrast, blur and noise for x-ray modality. From the AUC value of each graph, blur characteristic show better performance in x-ray modality, followed by contrast and noise. This result is tally with the fact that x-ray modality indicate the highest effect of image blur with low value of noise. As such the visibility of the detail is limited (Sprawls, 1995).

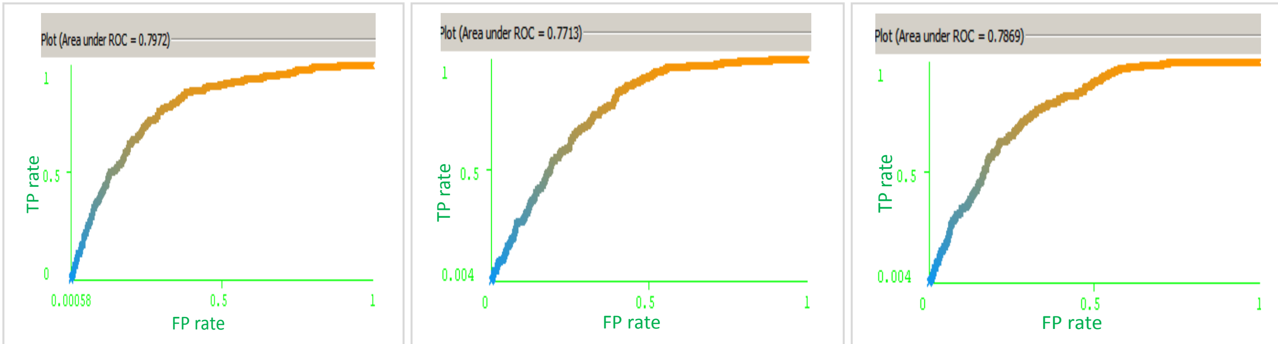


(a) Contrast (b) Blur (c)Noise

Figure 8.1: ROC graph for x-ray modality with AUC value (a) 0.7938, (b) 0.8189 and (c) 0.7159

8.2.2 CT Scan

As for CT modality, the AUC value of contrast, blur and noise are about the same as depicted in Figure 8.2. High contrast and noise values contribute to the good contrast between soft tissue such as kidney, liver and muscle. CT generally has a higher contrast value compare to x-ray (Donath et. al, 2010). This reason why CT is able to image soft tissue objects that cannot be imaged using x-ray.

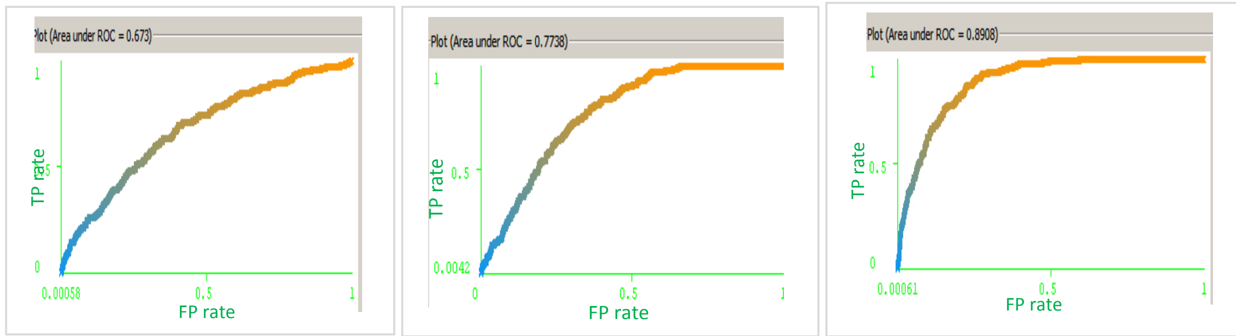


(a) Contrast (b) Blur (c) Noise

Figure 8.2: ROC graph for CT modality with AUC value (a) 0.7972, (b) 0.7713 and (c) 0.7869

8.2.3 US

ROC graphs presented in Figure 8.3 shows that US has a high value of noise and low value of blur and contrast. The high value of noise in US has contributed to well-defined beams and focused to explore between human body and tissue structures. Technically US has high contrast (Harvey et. al, 2002). However, the ROC graph of shows contrasts characteristic result. The visibility of object is based on the operating frequency. This means that at low frequency, US is able to capture deep-lying structures and at high frequency it’s more suitable for imaging object close to body surface. Therefore we assumed that the US images used in this work were taken at high frequency level which caused low value of contrast (Kuhn, 1995).

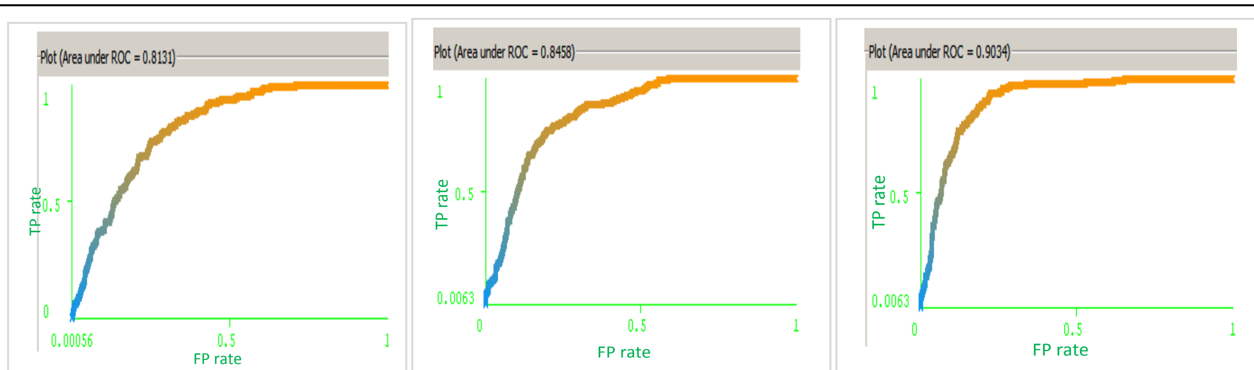


(a) Contrast (b) Blur (c) Noise

Figure 8.3: ROC graph for US modality with AUC value (a) 0.673, (b) 0.7738 and (c) 0.8908

8.2.4 NM

Compared with other modalities, NM images are generally the noisiest image which is proven in Figure 8.4 (as US is the second nosiest with value of 0.8908). The noise AUC value of NM images is 0.9034. Furthermore NM has the highest value among other modalities studied in this research work except GX since graphic image does not represent human body image but more in visualizing result of a disease. High value of noise can cover and reduce the visibility of certain features within the image (Khandelwal, 2012). However, in the case of NM, it only focuses on radiopharmaceuticals objects.

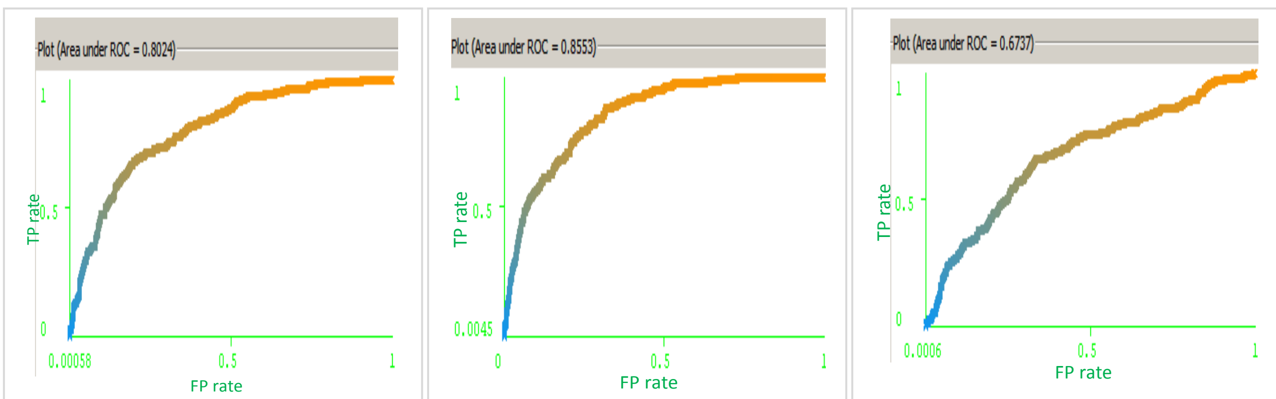


(a) Contrast (b) Blur (c) Noise

Figure 8.4: ROC graph for NM modality with AUC value (a) 0.8131, (b) 0.8458 and (c) 0.9034

8.2.5 PET

PET is biodistribution of positron-emitting radiopharmaceutical within the body by using nuclear medicine technique. Basically PET images should have high value of noise compared to blur characteristic (Worsley et. al, 1996). In contrast, the result shows in Figure 8.5 indicates low value of noise and high value of blur. This may be due the reason that the method used to measure noise and blur is not suitable for PET images. However, the contrast graphs can be accepted. The contrast graph exhibited high AUC value of 0.8024.

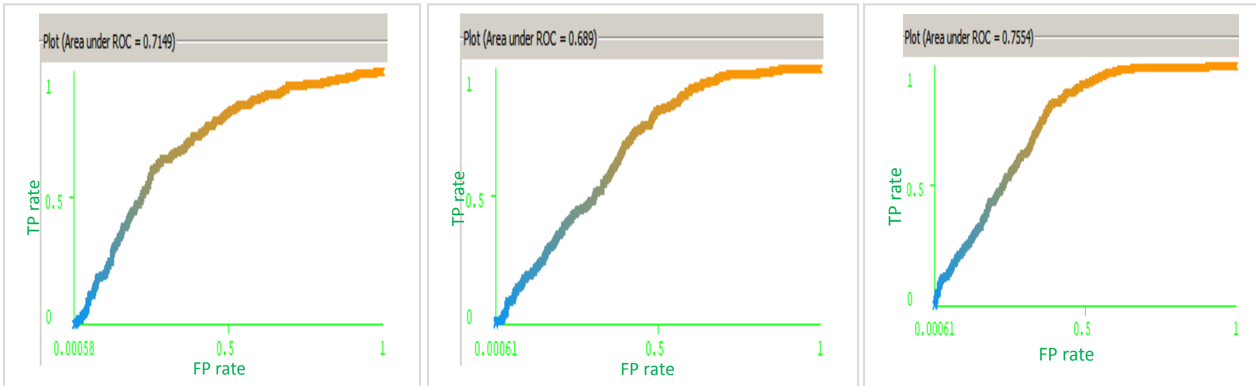


(a) Contrast (b) Blur (c) Noise

Figure 8.5: ROC graph for PET modality with AUC value (a) 0.8024, (b) 0.8553 and (c) 0.6737

8.2.6 MRI

ROC graph for MRI modality is shown in Figure 8.6. This observation is consistent with the fact in that noise is significant for MRI in order to visualize soft-tissue in human body (Sprawls, 1995). In principle, MRI and CT are categorized in the same type of imaging technique namely tomographic imaging which produce images of selected planes or slices of tissue in the human' body. However in this experiment, the values of contrast, blur and noise for MRI are lower than those for CT. This is maybe due to misclassification between CT and MRI since those images have similar texture and subtle difference from each other.

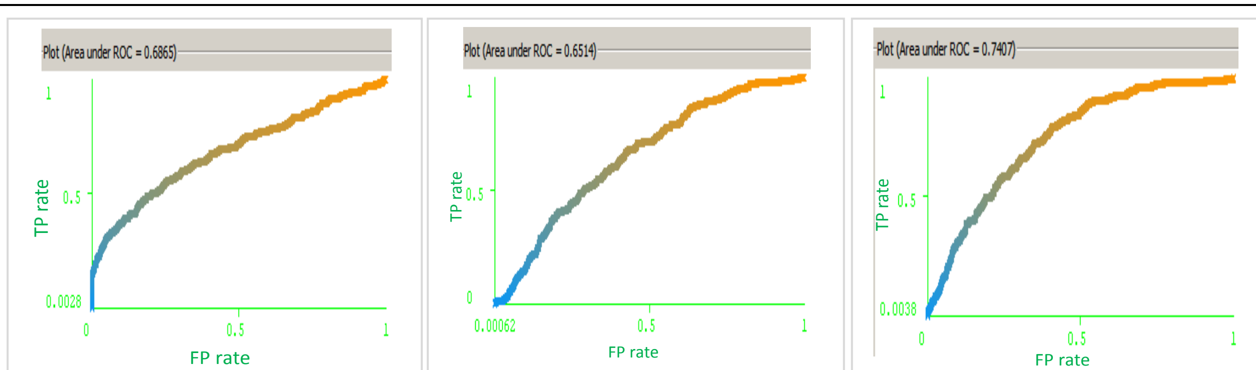


(a) Contrast (b) Blur (c) Noise

Figure 8.6: ROC graph for MRI modality with AUC value (a) 0.7149, (b) 0.689 and (c) 0.7554

8.2.7 PX

PX provided in ImageClef 2010 collection involves microscopy and gross anatomy image. Visually these two images have distinct characteristics of texture and shape. These results contribute to inconsistent in vectors of training data and the course of low value of contrast, blur and noise characteristics as depicted in Figure 8.7.

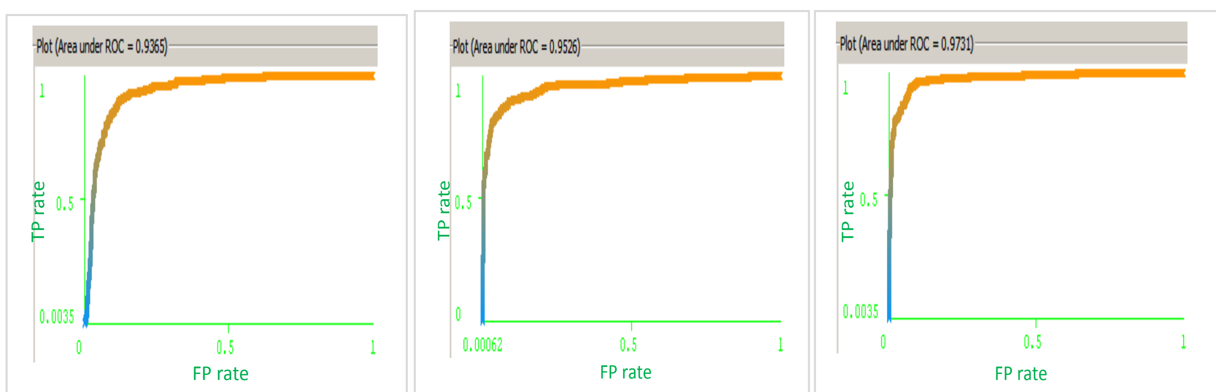


(a) Contrast (b) Blur (c) Noise

Figure 8.7: ROC graph for PX modality with AUC value (a) 0.6865, (b) 0.6514 and (c) 0.7407

8.2.8 GX

GX images provided by ImageCLEF 2010 collection are in the form of a chart, diagram or graph. Theoretically there are no image characteristics of graphic image in the physical principles or theory of medical imaging description. However, graphic image is important in analysing and visualizing medical results. From the ROC graphs in Figure 8.8, it is clear that graphic image exhibited high value of contrast, blur and noise. This shows that graphic image has the best performance in measuring image quality characteristics. The graphic image features such as sharp, clear and non-complex image has led to high results of these characteristics.



(a) Contrast (b) Blur (c) Noise

Figure 8.8: ROC graph for GX modality with AUC value (a) 0.9365, (b) 0.9526 and (c) 0.9730

8.2.9 Visual Features Evaluation in Multi-modality Medical Image

The experiment is carried out to investigate which visual features those are suitable to use in order to extract significant information from multi-modality medical image. Table 8.1 show the percentage of accuracy classified based on each modality for visual features of texture, shape and color. From the table we can explain that GX has the higher value in each visual feature. This is because GX only represent chart and graph of medical image which it is not a complex image. In contrast XR and MR have low value in all features due to the complexity of the image. Nevertheless the difference between the values from each modality is subtle. For NM modality, the color percentage is high due to this modality is a color medical image. Therefore it is easy to classify this modality based on color. The difference in shape for US modality with other modalities has contributed to the high value of correctly classified in shape for US. As for PET it easy to differentiate with other modalities using texture feature which result to high correctly classified values.

Table 8.1: Percentage of Multi-modality Medical Image Correctly Classified

Modality	Texture	Shape	Color
CT	56	49	30
GX	91	89	85
XR (x-ray)	49	13	29
MRI	17	25	42
NM	26	52	69
US	38	64	41
PET	54	27	38
PX	18	46	52

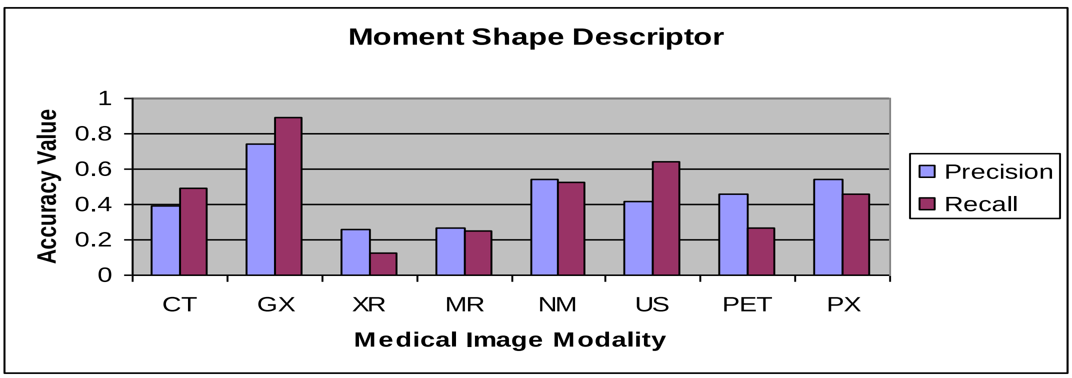


Figure 8.9: Precision and recall for texture feature

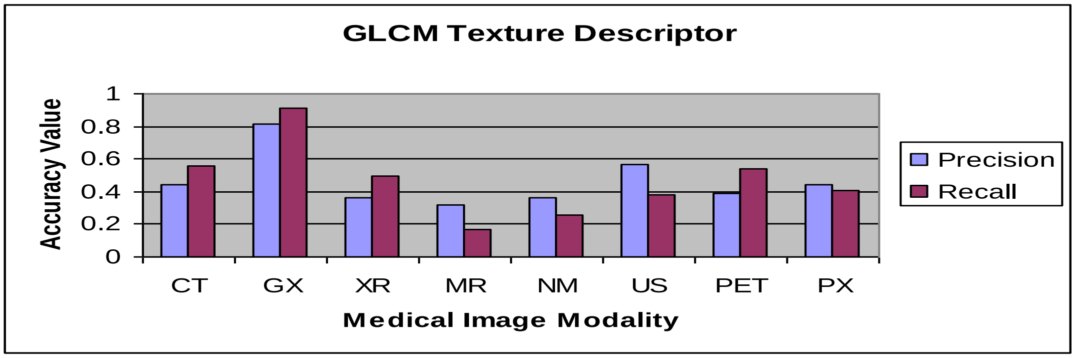


Figure 8.10: Precision and recall for shape feature

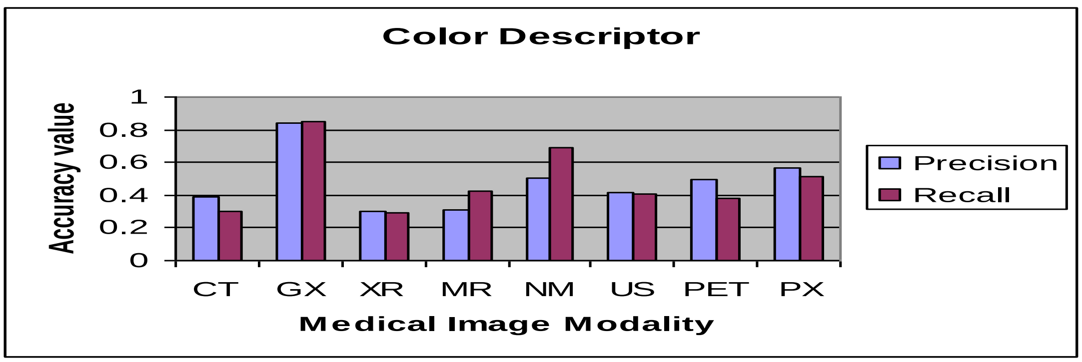


Figure 8.11: Precision and recall for color feature

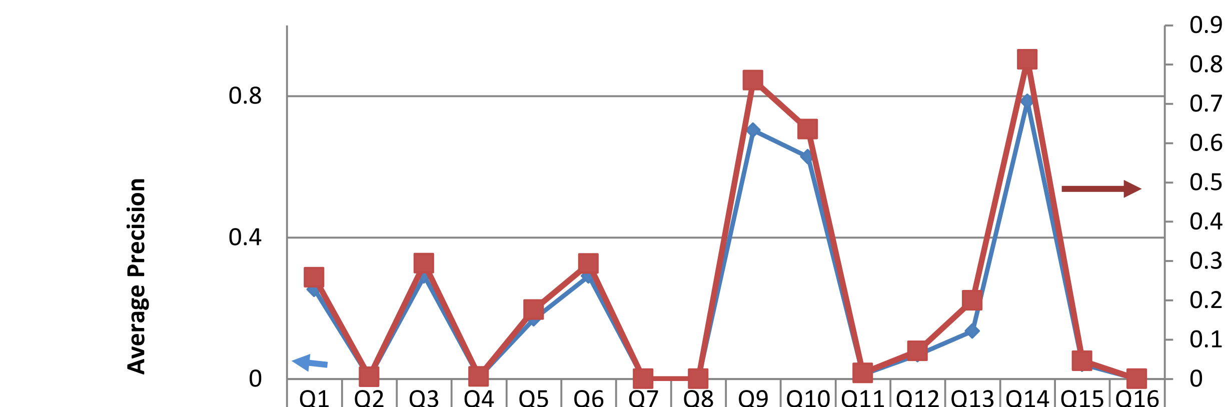
Figure 8.9, Figure 8.10 and Figure 8.11show the accuracy value of precision and recall for, shape, texture and color features. It shows that CT, x-ray, and PET are suitable to be classified using texture descriptor as depicted in Figure 8.10. As for moment shape descriptor, the result almost similar with texture but the value of accuracy is higher compared to texture as shown in Figure 8.9. It explains that shape descriptor using local level has better presentation compare to texture descriptor in analysing multi-modality medical images. Finally the color descriptor in Figure 8.11 shows that GX, PX, NM and PET have higher value. This is due to those modalities have used more colors compare to other modalities which concentrate only on grey-scale image.

8.3 Text Features Results for M3IRS Framework

Previously in section 7.4 we have elucidated experimental setup for text features. These experiments were conducted to evaluate the performance of each component in M3IRS text-based framework.

8.3.1 Results for Document Pre-processor Experiment with XTE Method

The experiment is carried out to compare the performance of M3IRS using original ImageCLEF medical documents and the new indexed and manipulated medical documents using XTE method. The performance evaluation is based on average precision of each query and MAP for overall performance of the queries. Figure 8.12 shows the average precision of each query for both ImageCLEF collection and our new indexed medical documents using XTE method for document pre-processor. The figure clearly shows that XTE method in indexing medical documents produced higher average precision values compared to the original ImageCLEF data collection.



	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16
ImageCLEF	0.25	0	0.29	0.01	0.17	0.29	0	0	0.7	0.63	0.01	0.07	0.14	0.79	0.04	0
XTE-based Index	0.26	0.01	0.29	0.01	0.18	0.29	0	0	0.76	0.64	0.02	0.07	0.2	0.81	0.05	0

Figure 8.12: Comparison of average precision value between original ImageCLEF and new index document using XTE method

Although there is only a slight increment in the result but it proves that our technique can improve the result and performance in IR system. The result also suggests that the technique is able to simplify the XML data and remove the redundance of the caption thereby promoting the relevant document into higher position leading to slightly higher precision. Average precision values for Q7, Q8 and Q16 are equal to zero value because there are no such medical terms in the MeSH thesaurus. However, in the research done by Wu et al (2010), they changed the original Q16 ad hoc query to “images of dermato fibramo” which separates the term dermatofibramo into two words and we followed the guideline for the rest of the experiments. This is because these two words of “dermato” and “fibramo” represent medical terms that are contained in MeSH thesaurus..

Table 8.2: MAP values for original ImagecLEF data and modified data using XTE Method

ImageCLEF collection	**XTE-based Index Documents**
0.212	**0.223**

Table 8.2 presents the MAP value exhibited by M3IRS using the original ImageCLEF data and XTE-based index medical documents. The table clearly shows that using our XTE method to index new medical documents results is 5% higher on MAP compared to that the value obtained when original ImageCLEF data was used. Removing duplication in medical documents results to the increment of relevant document in higher position. Remove the duplication sentences in each medical document definitely help to increase the position of relevant documents.

8.3.2 Result for Time Performance Experiment with Mesh-indexer

As described earlier, MeSH thesaurus is used frequently in retrieval process. We have developed a MeSH-indexer to organize the components of this medical thesaurus which is focused on giving the MeSH heading, synonym terms and semantic type. An experiment was carried out to observe the computational performance between the original MeSH version and MeSH-indexer. The execution time is taken starting from the system received a query and then process the query expansion technique of Type 2 (see section 4.3.1).

Table 8.3: Time evaluation between MeSH-indexer and original MeSH thesaurus

	Expansion (milliseconds)	**Expansion and Enrich (milliseconds)**
Original MeSH	4896	4896
MeSH-Indexer	74	101

Table 8.3 clearly shows that Mesh-indexer takes only 74 milliseconds to execute the task. In contrast using MeSH thesaurus, it takes 4896 millisecond. This means that using MeSH-indexer, the performance is 26 times faster. Therefore using MeSH-indexer for query expansion technique can increase the efficiency of M3IRS system. Furthermore the file size of MeSH-indexer is only 28.2MB while the file size of the original version is 276MB.

8.3.3 Results for Query Processor Experiment with Medical Context Aware Query Expansion Technique

Mentioned in section 4.3.1, there are three types of queries considered for this research. For Type 1 query which is the original query, no result is shown since there is no document retrieved based on exact matching of the original query. Therefore only Type 2 and Type 3 from medical context aware query expansion technique are used in this experiment. Type 2 used only medical terms identified in the query and Type 3 is the enrichment of Type 2 and the synonymous terms.

The experiment was conducted according to the description in section 7.4.3 and comprehensive ranking model is applied (see section 4.6.1). The purpose of this experiment is to evaluate the effectiveness of Type 2 and Type 3. For this purpose, therefore average precision value for each query is used to review the results that illustrate the position of relevant documents in *RMR*. The more relevant documents ranked on top of the *RMR* list, the higher average precision values.

Figure 8.13 shows the average precision values of each query. It clearly shows that there is improvement in the precision value when synonymous term as enrichment are used compared to when only applying medical term in the query list in Q1, Q3, Q16. Synonymous terms provide more varieties of medical terms to find more possibility relevant documents in each query. On the other hand, Q14 and Q15 values increased in Type 3 because of the changes of relevant documents position into higher rank which also explained the reasons of Q4 and Q9 values are decreased due to the position of relevant documents dropped to lower rank position. Note that the Q16 in Type 3 increase to 0.1 due to separate the term “dermatofibroma into two words “dermato fibroma” as mentioned in section 8.3.1

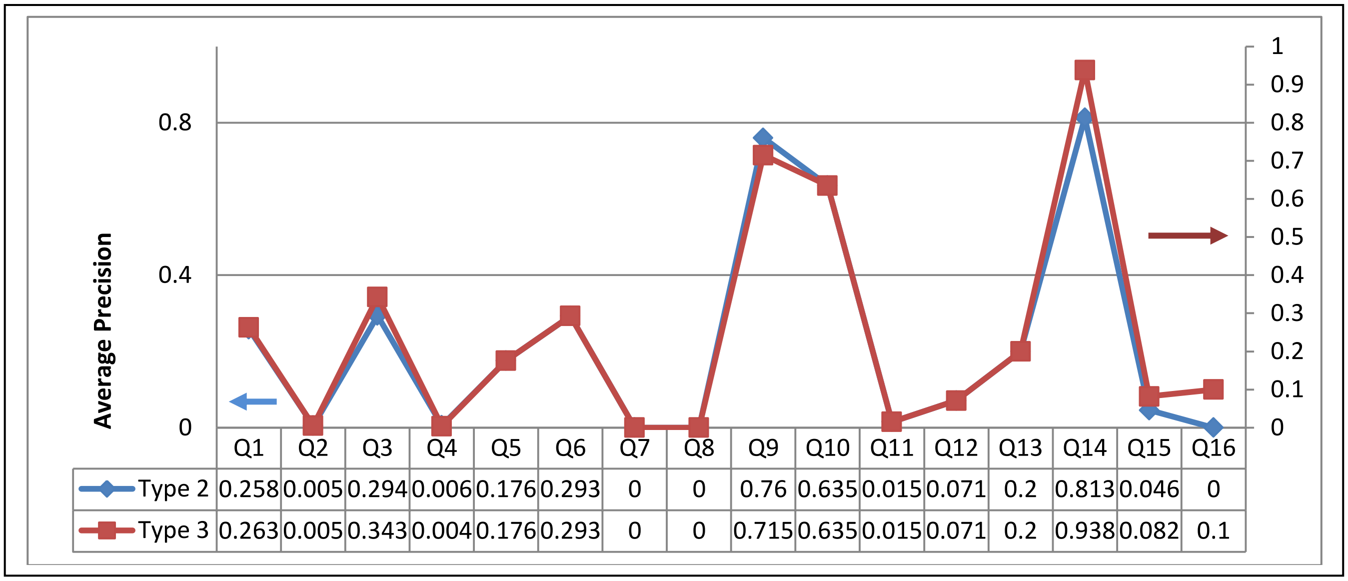


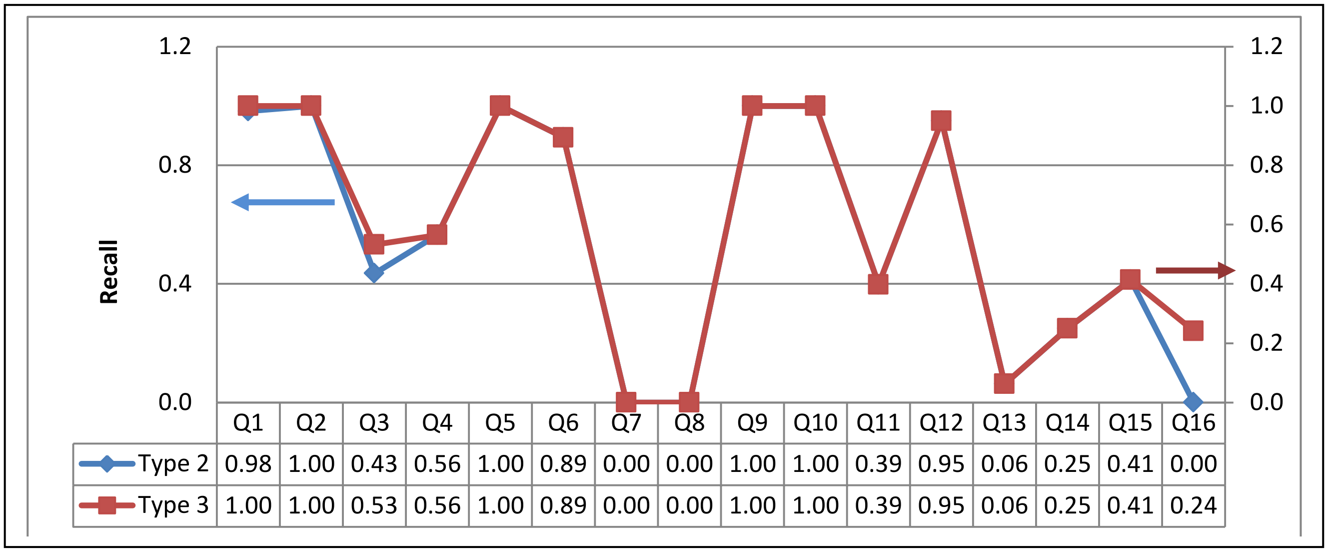
Figure 8.13: Average Precision values of each query for expansion only (Type 2) and expansion and enrich (Type 3) data

Table 8.4: Statistical measurements for Type 2 and type 3 queries

Type 2	**Type 3**
Mean = 0.21838	Mean = 0.25206
Standard Deviation = 0.26877	Standard Deviation = 0.3106
Standard Error = 0.06719	Standard Error = 0.07765
Probability Value: **0.74519**	

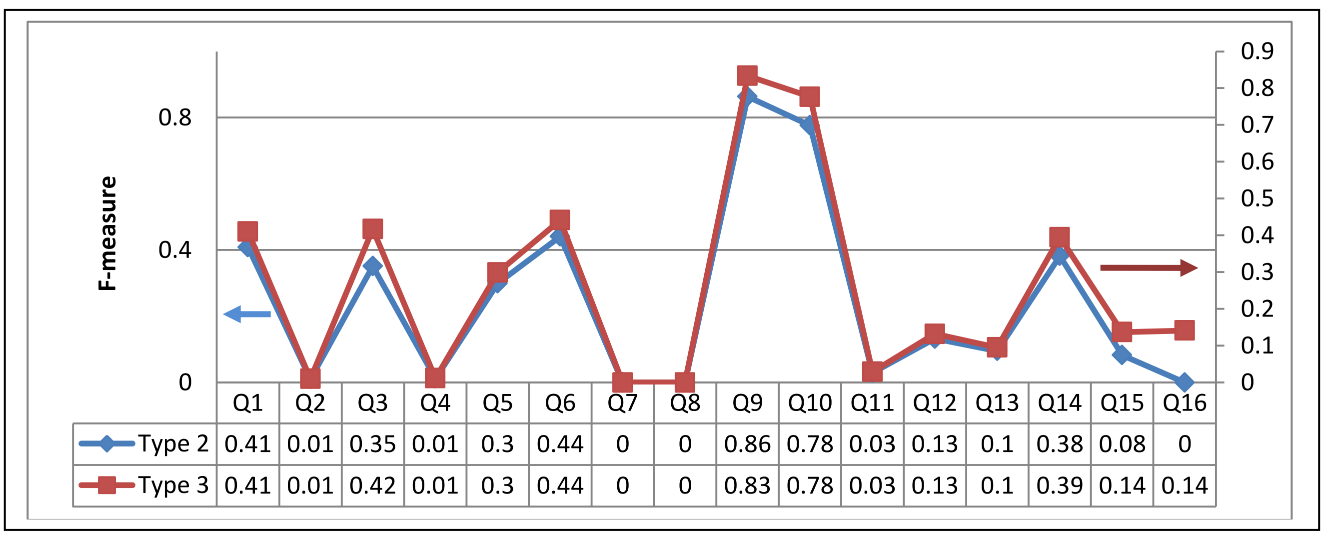
Table 8.4 shows several calculations of statistical measurements in comparing Type 2 and Type 3 queries. The calculation involves standard deviation, standard error and the probability value with 0.74519. Figure 8.14 represents recall values of each query and it shows that only Q1, Q3 and Q16 in Type 3 are higher values than Type 2. It explains that synonymous terms obtained from Q1, Q3 and Q16 contribute to the higher values in Type 3. This means medical documents that contains synonymous terms in Q1, Q3 and Q16 are in the list of relevant judgment. Figure 8.15 clearly shows that F-measure values of Q1, Q3, Q14, Q15 and Q16 from Type 3 are higher than Type 2. F-measure determine the accuracy

of medical context aware expansion technique based on Type 2 and Type 3 results whereby the calculation involves average precision and recall values.



	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16
Type 2	0.98	1.00	0.43	0.56	1.00	0.89	0.00	0.00	1.00	1.00	0.39	0.95	0.06	0.25	0.41	0.00
Type 3	1.00	1.00	0.53	0.56	1.00	0.89	0.00	0.00	1.00	1.00	0.39	0.95	0.06	0.25	0.41	0.24

Figure 8.14: Recall value of each query for expansion only (Type 2) and expansion and enrich (Type 3) data



	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16
Type 2	0.41	0.01	0.35	0.01	0.3	0.44	0	0	0.86	0.78	0.03	0.13	0.1	0.38	0.08	0
Type 3	0.41	0.01	0.42	0.01	0.3	0.44	0	0	0.83	0.78	0.03	0.13	0.1	0.39	0.14	0.14

Figure 8.15: F-measure value of each query for expansion only (Type 2) and expansion and enrich (Type 3) data

Table 8.5: MAP, Recall and F-measure values for Type 2 and Type 3 queries

Query Type	MAP	Recall	F-measure
Type 2	0.223	0.559	0.319
Type 3	0.240	0.582	0.340

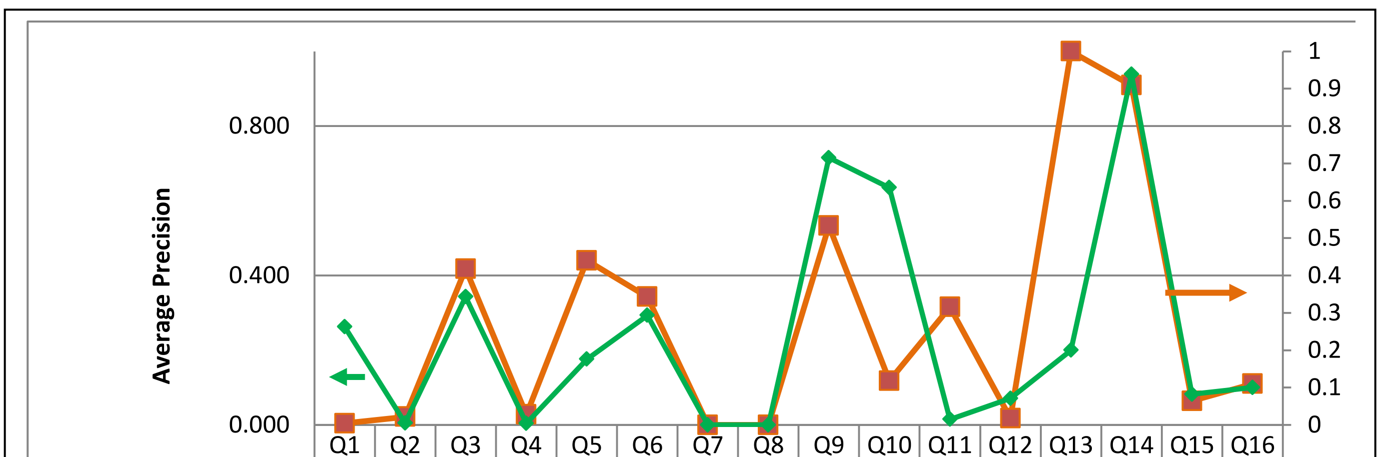
Table 8.5 shows the MAP, recall and F-measure for all queries and it clearly shows that Type 3 values are slightly higher than Type 2 due to the support of applying synonymous terms in query expansion technique. The increment values of MAP, Recall and F-measure in Type 3 prove that using synonymous term in query expansion technique has supported the performance of getting more relevant documents retrieved.

8.3.4 Results for Ranking Strategies Experiment

This experiment was carried out to compare the two types of ranking models used in this research as mentioned in section 4.6. Our retrieval model is based on Boolean matching. The disadvantage of Boolean model is it does not rank the *RMR*. As for retrieval system ranking the relevant documents is important because it can identify the most relevant documents to the query. Therefore ranking model is used in this framework to rank *RMR* list. In this model, the most relevant document is listed on top of the list and vice versa. The experimental setup is described in section 7.4.4 and the experiment was done using Type 3 query expansion technique.

The evaluation is based on average precision values of each query and MAP for theoverall performance in the queries. The recall values for both set of queries in comprehensive and MedHieCon ranking models are the same since the data to evaluate MedHieCon ranking model is taken from the experiment' result in section 8.3.3 which is Type 3 query expansion technique using comprehensive ranking model. The purpose of this experiment is to observe whether using medical concepts of modality, anatomy and pathology can lead to improvement in increasing the *RMR* list into higher position which affects the increment of average precision result.

The average precision values of Q2, Q3, Q4, Q5, Q6, Q11, Q13 and Q16 are shown in Figure 8.16 show that MedHieCon ranking model are higher compared to those of the reference. Eight out of 16 ad-hoc queries are tremendously enhanced by using MedHieCon ranking model. On the other hand, Q1, Q9, Q10, Q12, Q14 and Q15 in MedHieCon model show lower values than Comprehensive model. One of the reasons is due to the difficulty to conceptualize the query. For example in Q12, there is no specific pathology, anatomy or modality to conceptualize medical term in the query which results to the decrement of average precision value.



	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16
MedHieCon	0.004	0.022	0.418	0.028	0.440	0.343	0	0	0.533	0.118	0.316	0.018	1.000	0.909	0.064	0.110
Comprehensive	0.263	0.005	0.343	0.004	0.176	0.293	0	0	0.715	0.635	0.015	0.071	0.200	0.938	0.082	0.100

Figure 8.16: Comparison of average precision values between MedHieCon and comprehensive models

The other reason is that most medical terms that appeared in the query list are ambiguous as whether the term is a subheading or synonymous term in MeSH thesaurus. For example “coronary arteries” is the medical term of Q15 and it is a synonymous term of “coronary artery bypass” where by the system can only conceptualize the heading medical term in MeSH and not the synonymous term.

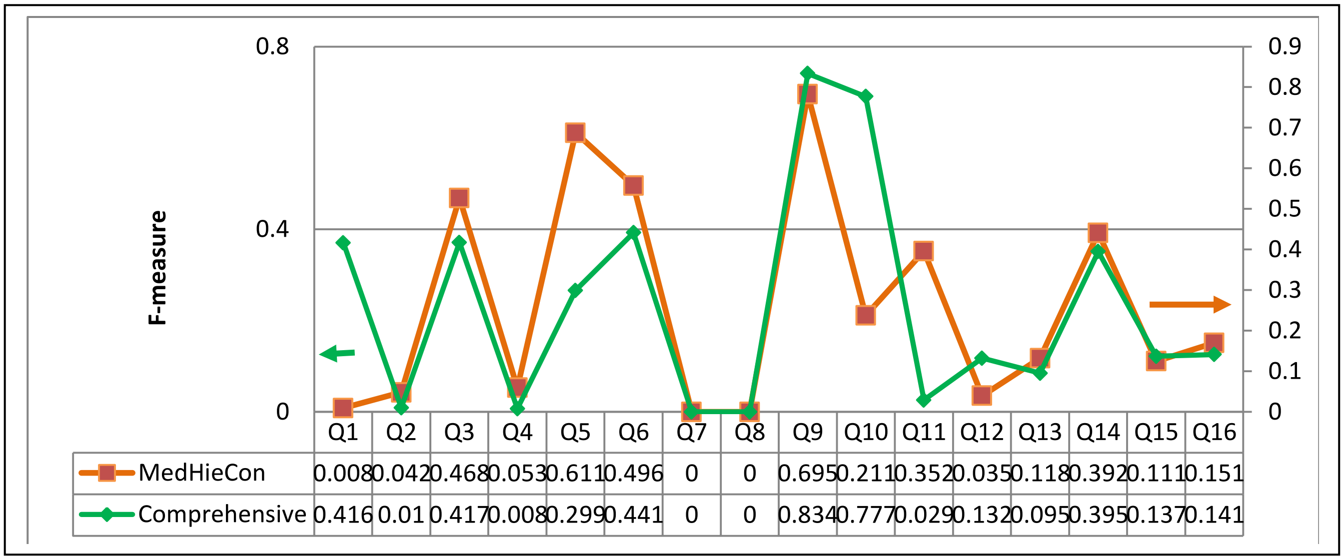


Figure 8.17: Comparison of F-measure values between MedHieCon and comprehensive models

The F-measure results for each query which are presented in Figure 8.17 are consistent with average precision values in Figure 8.13 where those queries that obtained high values of average precision have high values of F-measure. Q2, Q3, Q4, Q5, Q6, Q11, Q13 and Q16 from MedHieCon model have higher F-measure values compared to Comprehensive model. Significance measurement of these two ranking model is presented in table 8.6.

Table 8.6: comprehensive and MedHieCon ranking model

Comprehensive	MedHieCon
Mean = 0.25206	Mean = 0.27012
Standard Deviation = 0.3106	Standard Deviation = 0.32251
Standard Error = 0.07765	Standard Error = 0.08063
Probability Value: 0.87294	

Table 8.7 represents number of documents retrieved and the amount of relevant document retrieved for each query. Furthermore the amounts of relevant judgment document are listed as well. From these values, the measurement percentage of TP (true positive) and FP (false positive) can be calculated. From the table it shows that there are five queries which obtained 100% of relevant document retrieved (TP) which are Q1, Q2, Q5, Q9 and Q10. However there are still percentages of FP value but the values are lower than the percentages of TP for most queries. The results are statistically significant at 95% confidence interval with probability p=0.0177. It proves that our M3IRS framework is effective although the retrieval strategy involved Boolean model and not the conventional statistical model, and mapping the medical term into more meaningful semantic term for ranking purposes.

Table 8.7: List of total amount of document retrieved, relevant judgment document

Query	Document Retrieved	Relevant Document	Relevant Document Retrieved	Percentage of TP	Percentage of FP
Q1	32590	108	108	100%	43.309%
Q2	222	1	1	100%	0.295%
Q3	80	62	27	44%	0.071%
Q4	5684	23	13	57%	7.561%
Q5	31198	32	32	100%	41.555%
Q6	34126	94	86	91%	45.387%
Q7	387	4	0	0%	0.516%
Q8	5851	1	0	0%	7.801%
Q9	28686	12	12	100%	38.232%
Q10	28764	21	21	100%	38.324%
Q11	2912	103	40	39%	3.829%
Q12	4533	40	38	95%	5.993%
Q13	6	32	2	6%	0.005%
Q14	8	28	7	25%	0.001%
Q15	3346	406	168	41%	4.237%
Q16	30	29	7	24%	0.0031%
Probability Value:				0.0177	

Figure 8.18 shows the correlation graph between number of document retrieved and relevant judgement document for all 16 ad-hoc queries. The correlation coefficient value is 0.308. This value shows that there is positive correlation between number of document retrieved from our M3IRS retrieval strategy and the relevant judgement document taken from ImageCLEF 2010 medical task collection. However the low value of correlation coefficient prove the reason why the overall results of all run systems are low. This is because the documents in the collection are very huge which is 75,000 documents but the number of relevant documents retrieved are low.

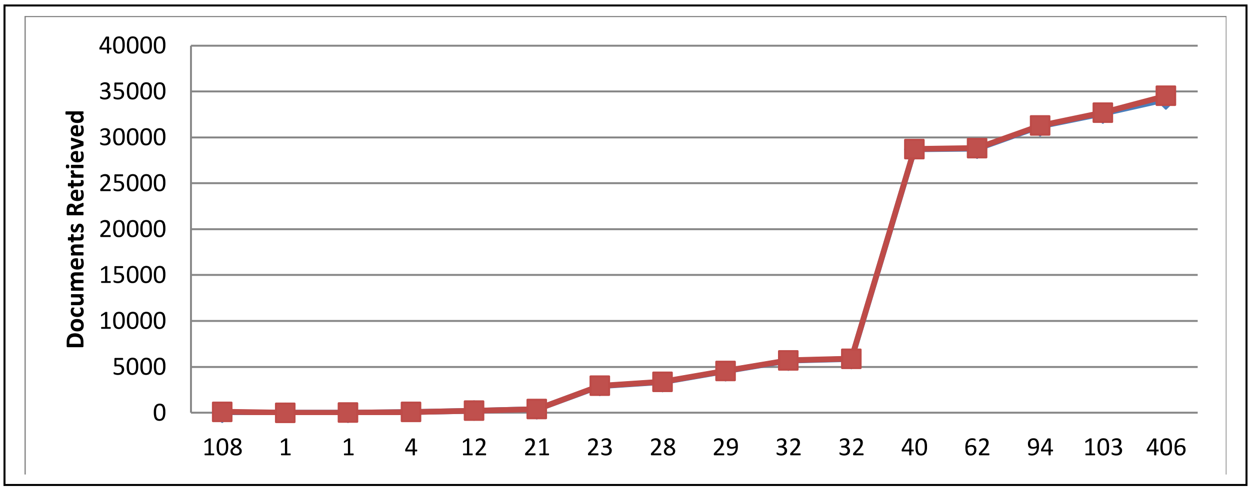
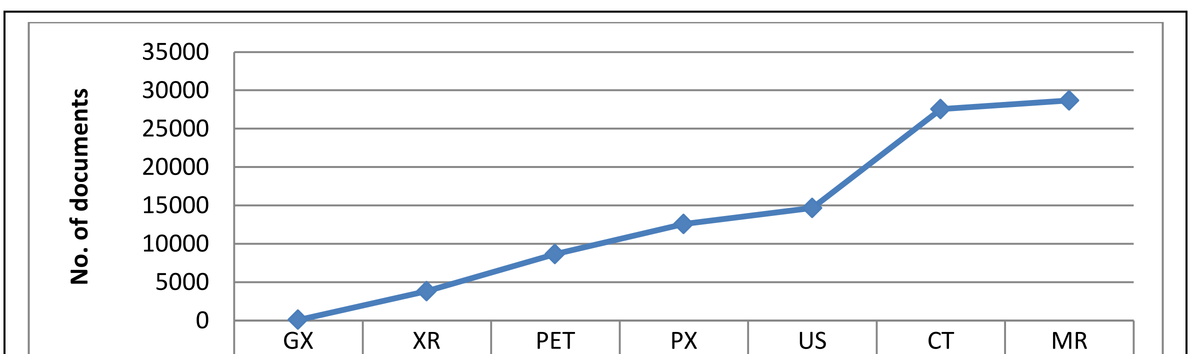


Figure 8.18: Correlation between document retrieved and relevant document retrieved for 16 queries.



	GX	XR	PET	PX	US	CT	MR
Amount	80	3820	8652	12563	14652	27537	28668

Figure 8.19: Number of documents retrieved based on modality

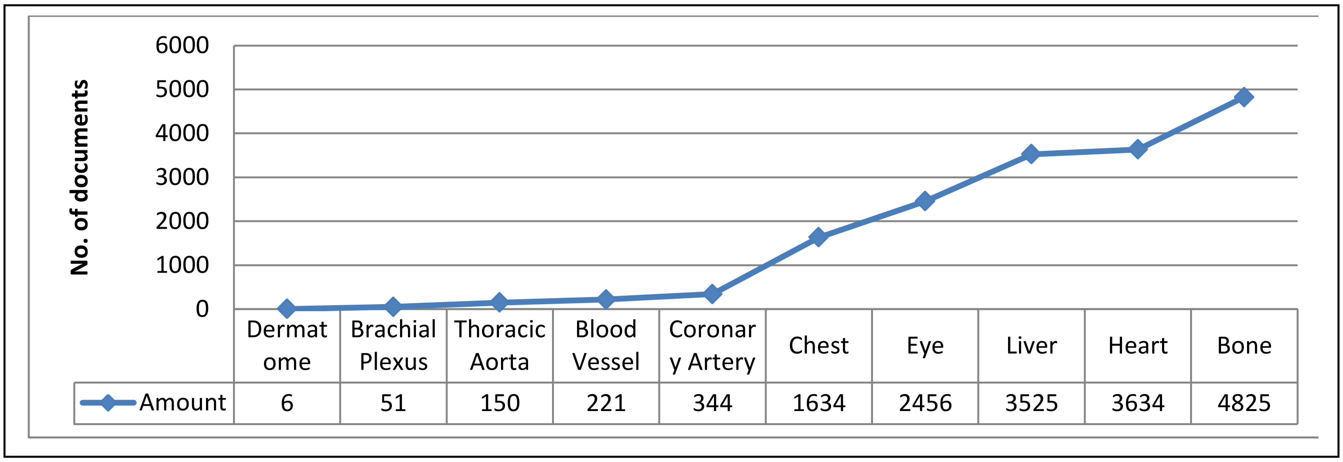


Figure 8.20: Number of document retrieved based on anatomy

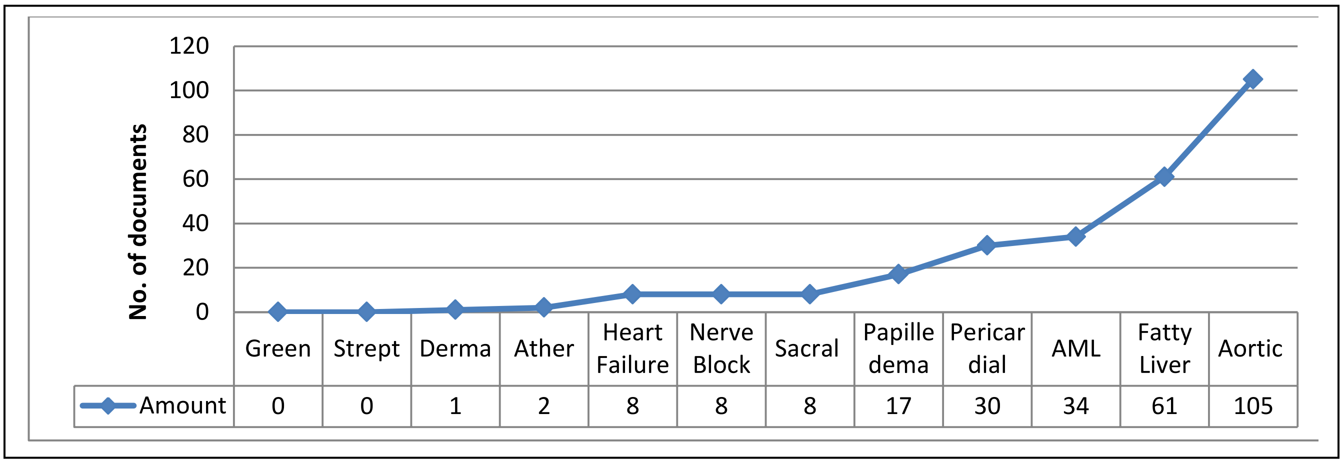


Figure 8.21: Number of document retrieved based on pathology

As previously mentioned in section 5.6.1, there are 7 modalities, 10 anatomies and 12 pathology used in this research based on the 16 ad-hoc queries. Figure 8.19, 8.20 and 8.21 show the number of documents retrieved based on the particular medical term. For modality the document can be retrieved up to 28,000 medical documents out of 75,000 documents given from ImageCLEF2010 which only reach up to 37% of entire documents. As it goes for more specific information to anatomy, the highest documents retrieved is bone with 4825 documents which only reach up to 6.4% of entire documents and for pathology the highest document retrieved is the aortic thoracic dissection which only retrieved 105 documents with 0.14% of entire documents. From these figures it shows that although there are 75,000 medical documents in the ImageCLEF collection, only small percentage of these medical documents are relevant.

Table 8.8: MAP and F-measure values for overall queries in using comprehensive and MedHieCon ranking model

Ranking Model	MAP	F-measure
Comprehensive	0.240	0.340
MedHieCon	0.270	0.369

Tabulated in Table 8.8 are the MAP and F-measure values for MedHieCon and Comprehensive ranking models. Table 8.8 shows that M3IRS framework using MedHieCon ranking model improved 11% of MAP and 8% of F-measure values of M3IRS performance compared to Comprehensive model. This improvement proves that MedHieCon ranking model can be used to support Boolean model in retrieval framework to increase the effectiveness of a retrieval system.

Table 8.9 shows the comparison of MAP and P@10 results obtained from various studies that used the same ImageCLEF 2010 medical task data collection including of 16 ad-hoc queries and relevant judgments. As discussed in section 2.5, most of the run systems used statistical model for retrieval strategy (IPL (Stougiannis et.al., 2010), OHSU (Bedrick & Kalpathy-Cramer, 2010), UESTC (Ibrahim & Arafa, 2010)) and used weighting scheme in order to index the vector for each medical document.

In more specific, certain run system such as IPL, used Lucene search engine (Hatcher, 2004) to automate the text indexing and retrieval process. The default function is Okapi

BM25 and they modified the similarity function by adding new parameter '∂' and set to 0.5. That is the reason why IPL achieved higher rank since they treat the non-occurrence term as it appeared in the text to avoid non-occurrence of a term which can cause probability of zero for this term.

Nevertheless there are also disadvantages such as weighting scheme is somehow biased and difficult to use with the other data collection because different dataset has different priority which requires re-weighting different parameters based on the prioritization. Furthermore, when new data are inserted to the collection, the system needs to re-index the documents which involves the whole set of documents in the data collection and this is time consuming and increase complexity.

Table 8.9: MAP and P@10 results for various research studies using ImageCLEF 2010 data collection

No	Run System	MAP	MAP Absolute Difference	MAP Relative Difference	p@10
1	IPL	0.316	-	-	0.452
2	OHSU	0.299	0.017	5%	0.43
3	UESTC	0.279	0.037	12%	0.313
4	IPL	0.278	0.038	12%	0.375
5	UESTC	0.273	0.043	14%	0.344
6	***M3IRS MedHieCon Ranking Model***	***0.27***	**0.046**	**15%**	***0.354***
7	OHSU	0.261	0.055	17%	0.258
8	ISSR	0.258	0.058	18%	0.319
9	HES-SO VS	0.257	0.059	19%	0.35
10	OHSU	0.256	0.06	19%	0.381
11	***M3IRS Expansion & Enrich***	***0.24***	**0.076**	**24%**	***0.293***
12	ISSR	0.231	0.085	27%	0.281
13	***M3IRS Expansion Only***	***0.223***	**0.093**	**29%**	***0.245***
14	ISSR	0.219	0.097	31%	0.325
15	ITI	0.188	0.128	41%	0.375
16	ITI	0.158	0.158	50%	0.325
17	ISSR	0.147	0.169	53%	0.257
18	HES-SO VS	0.131	0.185	59%	0.181
19	Bioingenium Research Group	0.101	0.215	68%	0.188
20	ISSR	0.098	0.218	69%	0.15

In contrast to traditional Boolean model although it is not a favourite retrieval model but with the support of our pre-process XTE method for indexing the medical documents and MedHieCon ranking model, it can really help the run system to be among the best retrieval model. It shows that ranking based on semantic type which gives more meaning to the medical terms has contribute to rank the relevant documents in higher position. Our M3IRS text-based framework emphasizes on simplistic process where it is suitable for other data collection to use and provides non-complex calculation if there is new data need to be add on.

Table 8.9 clearly shows that our run system outperformed many other systems and is in the 6th place with MAP value of 0.270 and P@10 of 0.354. All top 5 lists are run systems that used weighting-based for ranking the relevant documents. This competitive result proves that by using suitable retrieval and ranking strategies such as M3IRS framework, Boolean model has potential to improve the effectiveness of retrieval system to the same level as statistical model. There is also calculation on MAP absolute difference (difference between the best value, that is IPL MAP and the MAP **for** other systems) and relative difference (MAP absolute difference divided by the IPL MAP). It shows that our M3IRS framework using MedHieCon ranking model has only 0.046 of r MAP absolute difference and 15% of relative difference. This means that the difference from IPL in measuring effectiveness of the performance is low. OHSU and UESTC show relatively big subsequent jump in the difference. The 4th, 5th and *M3IRS* systems show small jump between each other.

8.4 Results of Visual Features for M3ICS Framework

Similar to text-based framework, content-based framework also used ImageCLEF 2010 medical data inclusive of 16 ad hoc queries. The training images used for visual experiments are based on images from relevant judgment list for each 16 ad-hoc queries. The main reason images from relevant judgment data is used because these images are already classified by expertise that associate with ImageCLEF. Therefore these images are accurately labelled and result to accuracy of our result and avoid bias.

This section the results of low-level features (texture, shape and color) experiments based on global and local descriptors are presented. The preliminary experiments involved finding optimum value for training and testing data and optimum value for *k* in k-NN classification. This section also reviews the results of semantic mapping using medical concepts of modality, anatomy and pathology.

Two classifiers namely k-NN and SVM are used to train medical data into classes that based on 16 ad hoc queries. SVM is widely used for statistical learning and classification. Primarily SVM deals with binary classification problem but currently two multiple classification approaches, one-against-one, and one-against-all are also used (Akbani et. al, 2004). One-against-one also known as pairwise classification is chosen for the experiment because it is computationally faster. In addition, polykernel is applied where it is chosen based on empirical study. The second most widely used classification method k-NN is used for further comparison. The comparisons are made based on MAP value and percentage of correctness rate.

8.4.1 Results for Optimum Value for Training and Testing Data

In this experiment we applied seven categories of test and train data. For the train data, the feature vector of each image including of texture, shape and color features are based on global descriptor. These feature vectors are trained into 16 classes using k-NN and SVM classifiers and test data will be classified based on these classes. As shown in Table 8.10 that category 3 which test data is 25% and training data is 75% have the highest MAP value for both k-NN and SVM classifiers. For classifier comparison, it is observed SVM has higher value than k-NN classifier.

Shown in Figure 8.22 are the percentage of correctness rate in different set testing and training data. The figure clearly shows that using SVM classifier produces the highest percentage of correctness rate in category 3 which is 92.28%. Similar to SVM, k-NN classifier obtained highest correctness rate in the same category that is category 3 (83.58%). However, correctness rate is lower than SVM. Therefore, for further experiments category 3 is applied (75% of train data and 25% of test data) as optimum value in classification.

Table 8.10: List of Test and Training Data Percentage and MAP Values for

Category	Test Data (%)	Train Data (%)	MAP Value	
			KNN	SVM
1	10	90	0.0912	0.1189
2	20	80	0.0932	0.1213
3	**25**	**75**	**0.0942**	**0.1268**
4	50	50	0.0748	0.1237
5	75	25	0.0748	0.0763
6	80	20	0.0853	0.0842
7	90	10	0.0796	0.0839

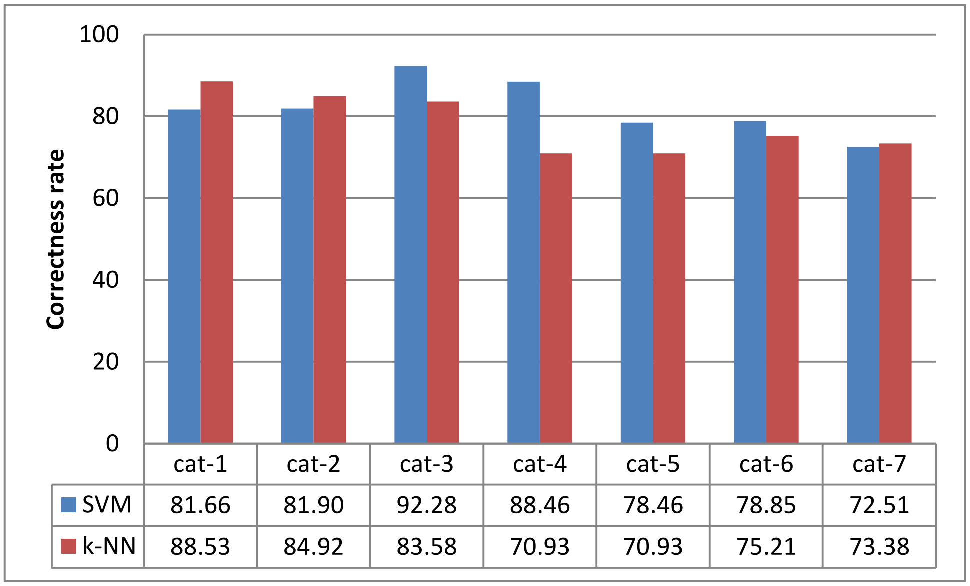
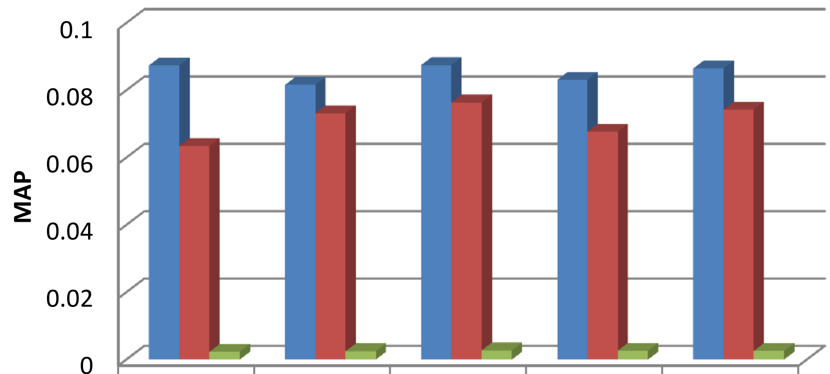


Figure 8.22: Percentage of correctness rate in different set testing and training data

8.4.2 Results for Classifiers Experiment with k-NN and SVM

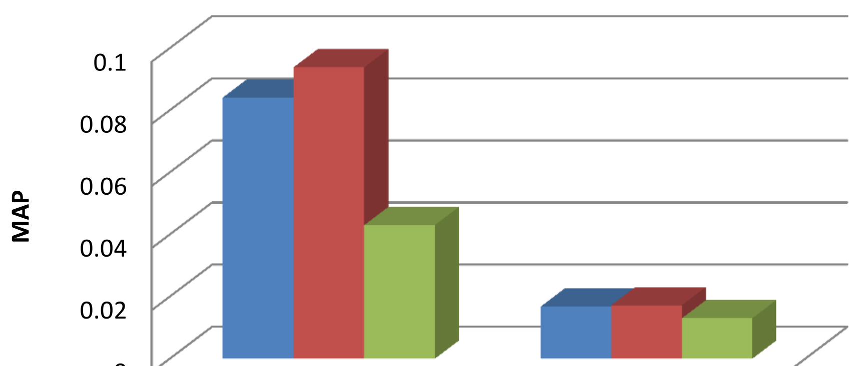
As mentioned earlier, k-NN and SVM classifiers are used to train data into 16 classes and classify test data based on identified classes. Therefore, optimum value *k* for k-NN and suitable kernel for SVM need to be identified. The results of which best *k* value and suitable kernel from this experiment was later used for the following experiments. As mentioned in 7.5.2, we conducted the experiment of k-NN classifier for *k*= {1, 2, 3, 4, 5} and for SVM classifier we compared the performance of polykernel and RBF kernel.



	1-NN	2-NN	3-NN	4-NN	5-NN
TEXTURE	0.0874	0.0816	0.0875	0.0831	0.0865
SHAPE	0.0635	0.0732	0.0764	0.0677	0.0743
COLOR	0.00231	0.00247	0.00275	0.00265	0.00257

Figure 8.23: MAP values for visual features in different k values of k ={1,2,3,4,5}

Figure 8.23 shows MAP values for visual features in different k values of k ={1,2,3,4,5}. It clearly shows that k-NN classifier highest result when *k*=3 for all features of texture, shape and color. Although there are slightly different from other *k* values, we determined to used *k* = 3 for the following experiments.



	POLYKERNEL	RBF KERNEL
TEXTURE	0.0841	0.0168
SHAPE	0.0938	0.0172
COLOR	0.043	0.013

Figure 8.24: MAP value for polykernel and RBF kernel in SVM Classifiers

As for SVM classifier, it is observed that polykernel has outperformed RBF kernel with difference of 80% for texture, 82% for shape and 70% for color as depicted in Figure 8.24. Therefore polykernel was used in SVM classifier in following experiments.

8.4.3 Result of Primitive Visual Features Experiments with Texture, Shape and Color Features in Global Descriptor

As previously mentioned in section 5.4, global descriptor means extract visual features from the whole medical image. Prior to that, the medical image was resized to 256×256 pixels and produced 75-dimensional feature vector (4-dimensions for color, 7-dimensions for shape and 64-dimensions for texture) for each medical image. From the results presented in previous section, we decided to use k =3 for k-NN classifier and also the used of polykernel in SVM classifier. In this section, the comparison among features in global level was performed. The performance of each features namely texture, shape, color and combination of these features are evaluated.

Table 8.11: Comparison of MAP value for texture, shape and color features for 3-NN and SVM Polykernel

	k-NN	**SVM**
TEXTURE	0.0864	0.084
SHAPE	0.0935	0.094
COLOR	0.0028	0.043
COMBINE	0.0987	0.116

Table 8.11 clearly shows that shape feature outperformed texture and color features based on global descriptor. This result is consistent with (Newsam et al., 2005) which also used GLCM for texture and moment method for shape in the simulation data. However, Newsam et al. (2005) used different methodology where the image is represented in 6 tiles inclusive dilated, flipped, eroded, rotation of 36°, 90° and 150°.

Shape feature is significant to represent the whole medical image in order to classify multi-modality medical images. This explains why MAP value of shape feature for both 3-NN and SVM using polykernel classifiers are the highest among the three features. The color feature obtains the lowest values. This is because most the medical images are grayscale images. Only certain modalities are in color format. Nevertheless the results from combination features are 5.2% and 19% better for both k-NN and SVM than independent features as depicted in Figure 8.25. Combination features involving extract texture, shape and color features in a feature vector.

Figure 8.25 also shows an impressive increment in the percentage of correctness rate for both SVM (91.87%) and k-NN (88.63%) when combining all visual features compared to independent features. The percentage of correctness rate of SVM using polykernel is 2.4% better than k-NN with k=3 in global descriptor classification.

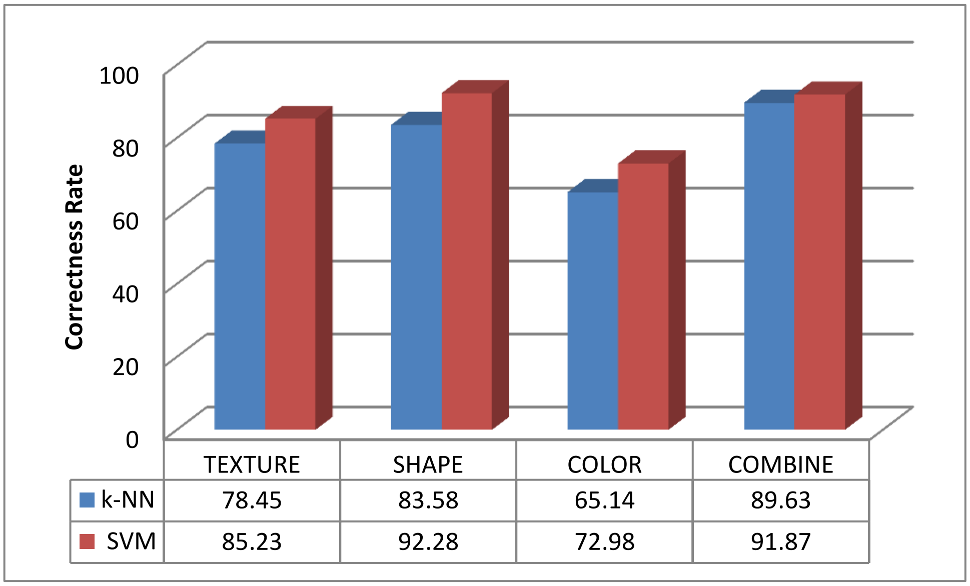


Figure 8.25: Percentage comparison between 1-NN and SVM Polykernel

8.4.4 Results for Local Descriptor Evaluation with 2x2, 4x4 and 8x8 Patches and Interest Points

As mentioned earlier, the purpose experiment described in section 7.5.4 is to evaluate which size of patches has the best performance in extracting features in local patches descriptor. Patches of different sizes produce different dimension of feature vectors as depicted in Table 8.12. The more patches are created the higher the number of dimensions produced in feature vector.

Table 8.12: List of dimensions produced based on patches

Patches	Dimensional
2x2	54
4x4	108
8x8	166

Table 8.13 shows the MAP values produced by each size of patches. It shows that 4×4 patches has obtained the highest MAP value for both k-NN and SVM classifiers with percentage of correctness rate in SVM is approximately 1% higher than k-NN as illustrated in Figure 8.26.

Table 8.13: MAP values for different size of patches for k-NN and SVM

Size of patches	k-NN	SVM
2x2	0.0863	0.0923
4x4	0.119	0.121
8x8	0.0963	0.0987

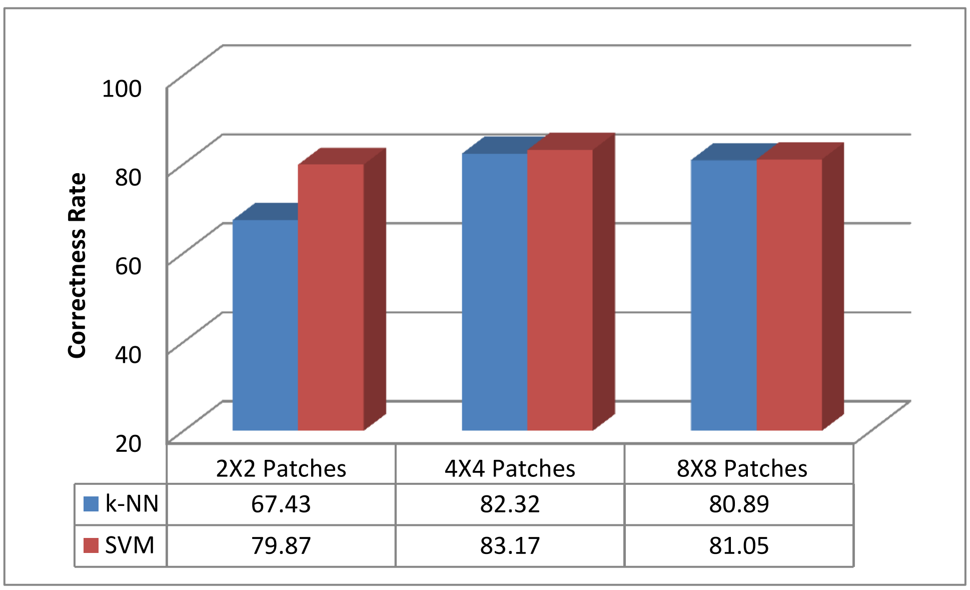


Figure 8.26: Correctness rate of various sizes of patch in k-NN and SVM

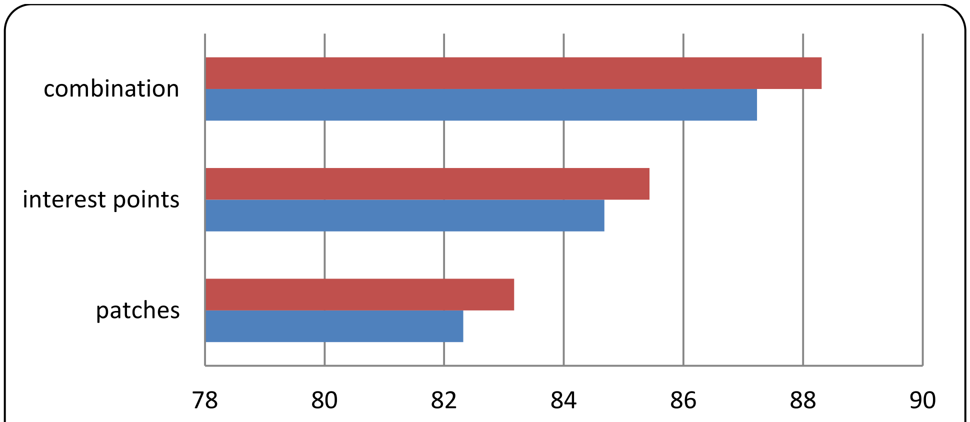
Based on the result, 4×4 patches is determined to be used for further experiments. Applying 2×2 patches in medical image is kindly similar extract the whole image (global level) since each patch is in large size. In contrast to 8×8, the patches are too small and many of the patches contain plain dark pixel surrounding the edges of the medical image.

The results for interest blocks extraction and combination of patches and interest blocks experiments are listed in Table 8.18. The size of interest blocks feature vector of each medical image is 460-dimensional (320-dimensional of texture and 140-dimensional of shape features). The result of 95.43% of correctness rate and MAP value of 0.128 are obtained from the experiment using SVM classifier. This means that the interest point's result is 5.5% better than 4×4 patches.

The final experiment for this section is to combine 4×4 patches and interest blocks which produced 568-dimensional (108-dimensional from patches, 460-dimensional from interest blocks). As depicted in Figure 8.27 the combination result is 3% higher than interest blocks result which is 98.31% of correctness rate with MAP value of 0.135 (Table 8.14).

Table 8.14: MAP values for 4X4 patches, interest points and combination of patches and interest blocks

	patches	**interest blocks**	**combination**
MAP	0.121	0.128	0.135



	patches	interest points	combination
SVM	83.17	85.43	88.31
k-NN	82.32	84.68	87.23

Figure 8.27: Percentage of correctly classified for patches, interest blocks and combination

8.4.5 Results of Performance Evaluation on MedHieCon Model for Visual Features

In this section, the performance results of MedHieCon model are presented. Nevertheless not all queries contain all of these concepts as described in Table 7.5 (section 7.5.5). For example only 10 queries (Q-1 to Q-10) contained modality concept and the rest of the queries did not emphasize on modality. For modality model, the training data is trained based on global descriptor process. As for anatomy model the images were trained using local descriptor process and pathology model the combination of global and local descriptor processes are used to train the train data. The difference of processing type is based on how details the information required to represent the models as mentioned in section 5.6. Overall there are 7 classes for modality, 10 classes for anatomy and 12 classes for pathology as illustrate in Table 8.15. 14,573 images were trained and the models of modality, anatomy and pathology are build and test the performance of each models using classifiers of k-NN with k=3 and SVM using polykernel. The main goal of this experiment is to validate how well the M3ICS system classify and annotate medical images based on these concepts and classes.

Table 8.15: List of classes for Modality, Anatomy and Pathology Concepts

Concept	Class
Modality	x-ray, PET, PX,US, CT, GX, MR
Anatomy	Thoracic Aorta (A-1), Chest (A-2), Brachial Plexus (A-3), Liver (A-4), Bone (A-5), Eye (A-6), Heart (A-7), Blood Vessel (A-8), Dermatome (A-9), Coronary Arteries (A-10)
Pathology	Thoracic Aortic Dissection (P-1), Acute Myeloid Leukemia (P-2), Heart Failure (P-3), Brachial Plexus Nerve Block (P-4), Fatty Liver (P-5), Greenstick Fracture (P-6), Streotococcus Pneumoniae (P-7), Papilledema (P-8), Pericardial Effusion (P-9), Atherosclerosis (P-10), Sacral Fracture (P-11), Dermato Fibroma (P-12)

Table 8.16 represents five types of run systems executed in M3ICS. The first run system is the feature extraction based on global descriptor while local descriptor of patches, interest blocks and combination of patches and interest blocks represent second, third and fourth run systems in M3ICS. The final run system is the combination of global and local descriptor processes. The evaluation is based on MAP values of different run systems of M3ICS in decreasing order of MAP and correctness rate values for k-NN and SVM.

It can be seen that M3ICS run system of global descriptor (texture, shape and color features), local patches (texture, shape and color features), and local interest blocks (texture and shape features), SVM outperforms k-NN classifier in both MAP and correctness rate values. Nevertheless k-NN classifier also obtained the same order of result as SVM which the combination of global and local descriptor processes outperformed other run systems. The combination of all features, namely global, local with patches and interest blocks values achieved the highest level of correctness rate at 92.05% with SVM and 88.67% with k-NN. Different features of an image reflect different attributes; therefore the combination has lead to better results with SVM of approximately 3% better than k-NN classifier.

Table 8.16: MAP value for M3IRS run systems

M3ICS Run System	MAP k-NN	k-NN Correctness Rate	MAP SVM	SVM Correctness Rate
Global +Local (patch) + (interest point)	0.139	88.67	0.142	92.05
Local (patch) + Local (interest point)	0.131	87.23	0.135	88.31
Local interest point features	0.125	84.68	0.128	85.43
Local patch features	0.119	82.32	0.121	83.17
Global features	0.0987	78.63	0.116	81.87

Figure 8.28 shows the result from modality concept. Based on Table 8.3, GX class achieved the highest score of 100% correctness rate in SVM and 93% in k-NN classifier. GX which is based on medical graphs and charts can easily be classified compared to other modalities since these medical images are not complex. Other medical modalities that achieved high correctness rate are PET, x-ray and US. PET modality consists of color feature which is easier to be classified while most of other modalities are in grayscale format. The texture of US is different from other modalities which contribute to the high correctness rate in classification. However, in Figure 8.28 clearly shows that CT and MR obtained low value of correctness rate. This is because these two classes have subtle different from each other and difficult to be differentiated.

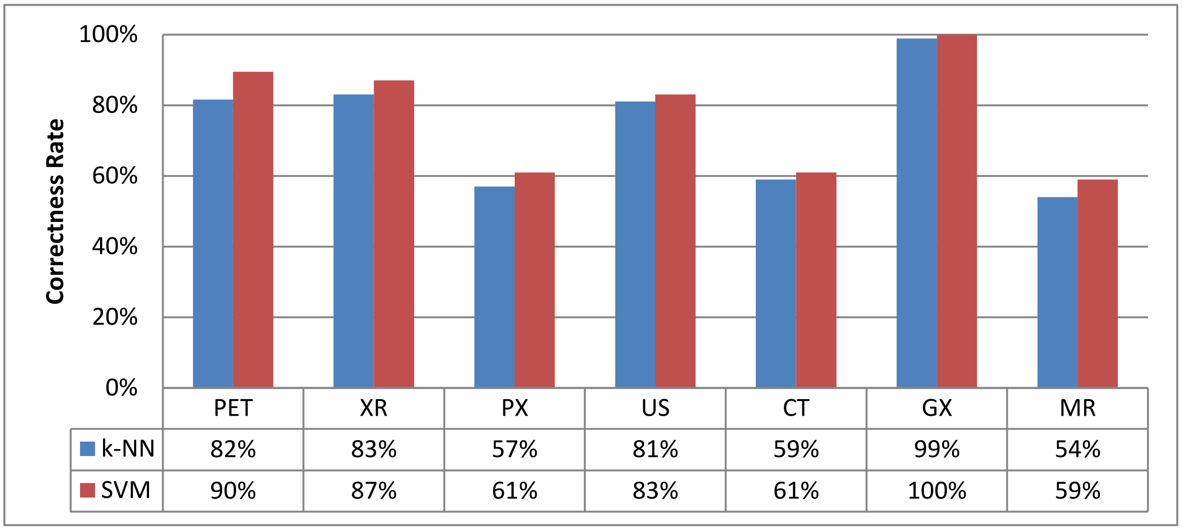


Figure 8.28: Classifier Comparison for Modality

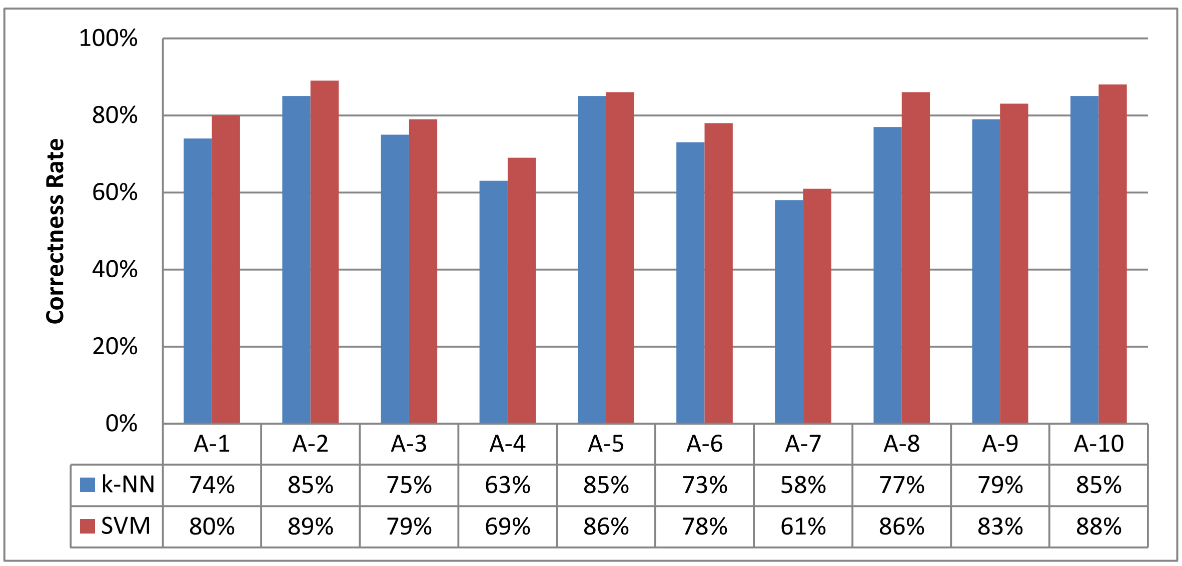
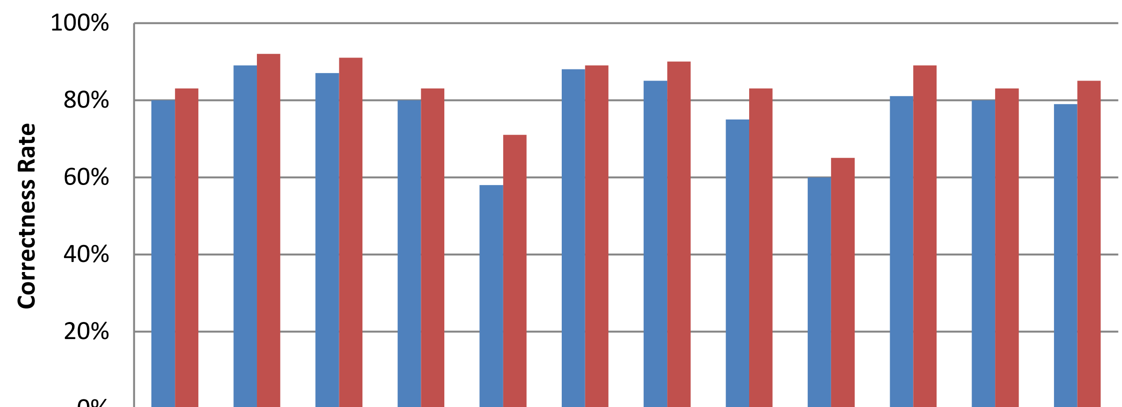


Figure 8.29: Classifier Comparison for Anatomy

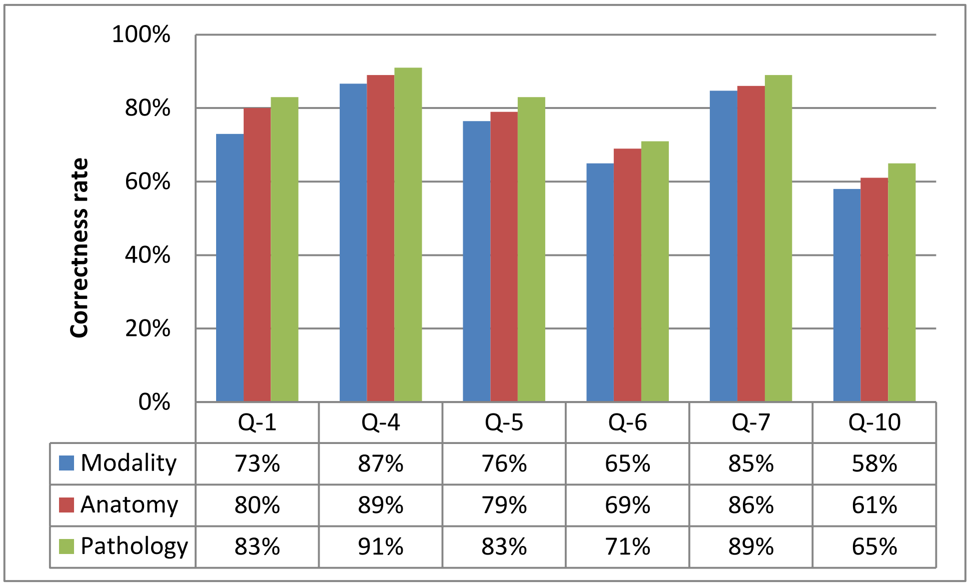
For anatomy concepts, the feature extraction involved local descriptor of patches and interest blocks. As can be seen in Figure 8.29, x-ray chest (A-2), x-ray bone (A-5) blood vessel (A-8), dermatome (A-9) and coronary arteries (A-10) classes achieved the highest percentage of correctness rate in SVM classifier. The results show that any parts of anatomy with x-ray modality are easier to be classified. The medical images with no specific modalities such as blood vessel and coronary arteries achieved also high value of correctness rate. Compared to images in class of A-1, A-4, A-6 and A-7 which is from CT and MR modalities, most of the images are not at the same depth and each image not really concentrate on the same angle and directions which contribute to low correctness rate.

The results of the pathology concept are depicted in Figure 8.30. The feature extraction involved the combination for of global and local descriptors which includes patches and interest blocks. For this experiment there are 12 classes in pathology and P-2 achieved the highest percentage of correctness rate which is 92% in SVM and 89% for k-NN. Followed by P-3 and P-7 which obtained 91% and 90% in SVM classifier. The lowest value is P-8. This may be due to misclassification between MR heart and CT thoracic aorta classes in anatomy concept.



	P-1	P-2	P-3	P-4	P-5	P-6	P-7	P-8	P-9	P-10	P-11	P-12
k-NN	80%	89%	87%	80%	58%	88%	85%	75%	60%	81%	80%	79%
SVM	83%	92%	91%	83%	71%	89%	90%	83%	65%	89%	83%	85%

Figure 8.30: Classifier Comparison for Pathology



	Q-1	Q-4	Q-5	Q-6	Q-7	Q-10
Modality	73%	87%	76%	65%	85%	58%
Anatomy	80%	89%	79%	69%	86%	61%
Pathology	83%	91%	83%	71%	89%	65%

Figure 8.31: Comparison of correctness rate between modality, anatomy and pathology

Q-1, Q-4, Q-5, Q-7 and Q-10 in Figure 8.30 are evaluated since these queries contained all medical concepts of modality, anatomy and pathology in the query list. Figure 8.31 shows that feature attributes in different descriptor contribute to the increment of correctness rate. The queries in the figure represent list of queries that contain modality, anatomy and pathology concepts in one query. As previously mentioned that modality concept used global descriptor as feature extraction, anatomy concept used local patches and pathology used combination of global and local descriptors inclusive of patches and interest blocks. It can be concluded that the more features extracted in the image the more detail the information obtained. This contributes to higher result of MAP and correctness rate which in turn increase the M3ICS performance.

8.4.6 Results of Experiment to compare the performance of M3ICS and MIARS

This section compares the results between M3ICS content-based framework and MIARS (Mueen, 2010). MIARS make use of low-level features such as gray-level co-occurrence matrix (GLCM), texture histogram, and pixel intensity information in both global and local descriptors. However, in local descriptor, MIARS only applied 2×2 patches. MIARS methodology is executed using the same ImageCLEF 2010 dataset. The classifiers for medical image classification are k-NN with k=3 and SVM using polykernel.

Table 8.17: Comparison of MAP value between M3ICS and MIARS

	k-NN	**SVM**
M3ICS	**0.139**	**0.142**
MIARS	0.136	0.138

The results of this experiment are given in Table 8.17. The table clearly shows that M3ICS methodology has achieved higher better result than MIARS for both k-NN and SVM classifiers. This shows that combining more features such as interest blocks can improve classification result.

8.5 Result of Comparison Performance between Text, Visual and Fusion Features

This section reveals the result of the performance of text, visual and information fusion features frameworks. The result is based on fusion of text and content-based information sources based on hierarchical processing in late fusion technique. The comparison evaluation is between information fusion and M3IRS text-based framework. The text-based framework was chosen because it has been reported to show better performance than visual-based retrieval system (Muller et al., 2010). Combining both text and visual is expected to improve the overall performance.

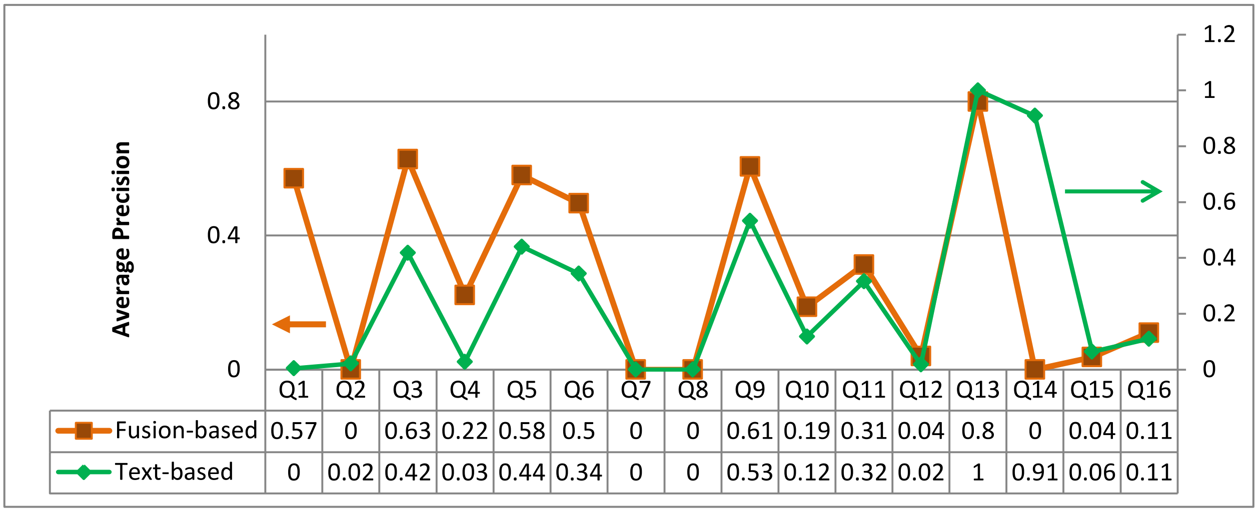


Figure 8.32: Comparison of MAP value between information fusion and text-based features

Figure 8.32 clearly shows that there are improvements in MAP value for several queries Q1, Q3, Q4, Q5, Q6, Q9, Q10 and Q12. However, Q13, Q14 and Q15 show lower MAP values. This may be due to the reason that these three queries do not represent any specific modality of medical image (refer to Table 7.5) and thus the classification is based on majority votes from each class. The images which are not in the majority class were discarded. Tabulated in table 8.18 is the statistical significance measurement of M3IRS and IFM3IRS with probability value of 0.87699

Table 8.18: Statistical significance measurement of M3IRS and IFM3IRS

M3IRS	IFM3IRS
Mean = 0.27012	Mean = 0.28688
Standard Deviation = 0.32251	Standard Deviation = 0.28173
Standard Error = 0.08063	Standard Error = 0.07043
Probability Value: 0.87669	

Table 8.19: MAP value between fusion, text and visual-based

Fusion	Text-based	Content-based
0.287	**0.270**	**0.142**

Table 8.19 compares the results between IFM3IRS fusion, M3IRS text and M3ICS content-based frameworks. It shows that MAP values for fusion of text and content-based has increased by 6% improvement compared to text-based solely. It shows that although text-based is outperformed visual-based in MIR, but visual features can compliment and support text-based to improve their performance.

The performance of other run systems is depicted in Table 8.20. It shows that of IFM3IRS information framework is in the 1st place. Although the result from M3IRS text-based framework is not on the top of the list but overall performance is better than other run system.

Table 8.20: Comparison of MAP values with other run systems

Run System	MAP
IFM3IRS	0.287
IPL	0.279
OHSU	0.256
ITI	0.107
Bioingenium	0.0395
medGIFT	0.0245

8.6 Summary

This chapter presents the results for experimental setups described in Chapter 7. Different modalities provide different image quality characteristics. This can be proven by using AUC value in ROC graph. The higher the AUC value the better performance of the characteristics. However, certain modalities are not suitable in measuring contrast, blur and noise characteristics. For examples GX (which actually represent diagrams and not part of human body) and PX (which represent different type of images; microscopy and gross anatomy image) are within the same modality. To conclude SVM is suitable to classify different modalities based on image quality characteristics since most of the results are consistent with the physical principle of medical imaging (Sprawls, 1995; Webb, 2003).

The following section describes the M3IRS text-based framework which includes query processor, document pre-processor, retrieval and ranking strategies components. Overall, our M3IRS run system using the combination of Boolean model and MedHieCon Ranking model is in 6th place out of 20 run systems. Although our system is not in top rank but the result from this study shows that Boolean model also can be competitive retrieval model with the support of MedHieCon ranking model. Multi-modality Medical Image Classification System (M3ICS) is the content-based framework which is based on extracting visual features of texture, shape and color in global and local descriptors and applying semantic classification using MedHieCon model. Finally are the IFM3IRS framework results which contribute to improvement of the MIR system by fusion of the information sources of M3IRS text and M3ICS content-based frameworks.

9.0 Conclusions and Future Research Implication

9.1 Introduction

This chapter highlights the summary and conclusive findings for the research work reported in this thesis. The first experiment is to evaluate multi-modality medical images in quality characteristics (contrast, blur, noise) and visual features (texture, shape and color). Next is to evaluate the performance of the MedHieCon model (modality, anatomy and pathology) for both text and visual content-based retrieval systems . In addition, the summary of the performance based on information fusion of text and visual features is also presented. Further research direction and possible enhancements in medical information retrieval are given at the end of this chapter.

The availability of multi sources in medical collection such as medical documents and images allows for information fusion in medical-based MIR system. The contribution of this research is the framework models which are based on (i) text (M3IRS), (ii) visual content (M3ICS) and (iii) information fusion (IFM3IRS) for multi-modality medical information retrieval. The effectiveness of text and content-based information fusion for the retrieval has shown good results with better performance compared to text or content-based systems solely.

9.2 Findings in Characteristics of Multi-modality Medical Images

This research involves multi-modality of medical images namely x-ray, CT scan (CT), ultrasound (US), nuclear medicine (NM), positron emission tomography (PET), magnetic resonance image (MRI), optical image (PX) and graphical image (GX). Experiments to evaluate the characteristics of these modalities were executed. These include the evaluation of quality characteristics of contrast, blur and noise for each modality. The results are based on AUC in ROC graph which represents the classification performance of each characteristic in each modality. It shows that GX images obtain high values of AUC for all the three characteristics since graphic image does not represent human body image but it is more on illustrating medical reports which is easy to differentiate with other modalities. The other modality that obtains high AUC value of classification for quality characteristics is NM as NM only focuses on radiopharmaceuticals objects.

The other experiment is to evaluate multi-modality medical images on visual features of texture, shape and color. In this experiment it shows that GX has the higher percentage of correct classification in each visual feature. This is because GX represents chart and graph of medical image whereby it is not a complex image and easy to be classified. In contrast x-ray and MRI modality obtain low percentage of correct classification for all features due to the complexity of the image. From other aspect of precision and recall value, CT, x-ray, US, PET and PX are suitable to be classified using texture descriptor. Nevertheless for overall performance in classifying multi-modality medical images using local descriptor, shape feature has better presentation compared to texture feature. Finally for color feature, GX, PX, NM and PET have higher precision and recall value due to those modalities using colors in the image meanwhile other modalities concentrate only on grey-scale image.

9.3 Findings in M3IRS Text-based Framework

M3IRS text-based framework is introduced to achieve the first objective of this research which is to design text-based framework for multi-modality medical information retrieval system stated in section 1.4. This framework has been implemented in the text-based M3IRS system which consists of four main components namely (i) document pre-processor, (ii) query processor, (iii) retrieval process and (iv) ranking process developed in this research.

9.3.1 Text Documents Management in ImageCLEF Medical Data Collection

In this research, 77,500 text documents from ImageCLEF 2010 medical task data collection are used. Issue of text duplication in the documents has led to low precision value. To meet the first sub-objective (1a pertaining to the first objective) and research question on managing medical documents for effective processing, XML Tag-based Extraction (XTE) technique is developed to organize and simplify information in medical documents, removing the non-related information and indexing new information and hence producing the suitable simplified structured version of the original medical documents. The XTE-based new index documents have achieved 5% improvement in MAP value compared to the original ImageCLEF XML documents, resulting in an increase in the effectiveness of M3IRS performance and reducing storage size.

9.3.2 Automatic Identification of Medical Terms in Text Documents

In this research, MeSH, an external medical thesaurus is used to identify medical terms and their synonyms in the query and text documents. In dealing with the second sub-objective (1b) and research question on manipulating appropriate thesaurus in extracting significant information and identifying medical terms in a query or text document, MeSH-indexer is created as a medical knowledge source adaptation from MeSH thesaurus. MeSH-indexer is used in query processor component to extract medical terms and their synonyms that exist in the query. Creating MeSH-indexer is beneficial to M3IRS system since it performs 26 times faster than the original MeSH thesaurus. The irrelevant information in MeSH thesaurus has contributed to slow searching of medical term and the synonyms in query expansion technique.

By using MeSH-indexer, the computational cost and execution time in M3IRS system are reduced. The query expansion technique involves converting all letters into lower case and performing query tokenization. Medical terms are identified based on tokens generated from the query expansion technique by using MeSH-Medical Descriptor (MMD) folder in MeSH-indexer where the list of medical terms taken from descriptor in MeSH thesaurus are labelled as query expansion Type 2. Later the synonymous of each medical term in Type 2 taken from list terms of each descriptor in MeSH thesaurs are identified from MeSH-Synonym Terms (MST) folder and the combination of the synonyms and Type 2 are labelled as query expansion Type 3. Sub-objective (1c) is met in formulating query model using external knowledge thesaurus.

The query expansion technique described earlier has answered the research question number 3; which is related to the process involved in extracting significant medical terms from medical documents. The enrichment of synonymous terms, Type 3 in query expansion technique has shown 22.3% improvement in MAP value compared to Type 2. This is because adding synonymous terms in the query list provides more option and possibility to find relevant documents. This is consistent with the recall value of 4% for Type 3 (0.582) which is better than that of Type 2 (0.559). Sub-objective (1e) is met in conducting evaluation of the framework.

9.3.3 Ranking Models for M3IRS Text-based Multi-modality Medical Information Retrieval System

Boolean approach is a retrieval model based on set theory and Boolean algebra. It is computationally efficient (William, 1992) and the algorithm is easy to implement. Vector space model (VSM) involves vector presentation of document and query, expressing text documents as vectors of identifiers. Issues related to VSM include missing semantic and syntactic information such as phrase structure and proximity information as reported in our publication (Sharef & Madzin, 2012). At the matching step it solely depends on the weights of terms which do not represent the importance of the particular terms in the document. As a weight is computed for every term in the document with the possibility to have zero-valued components to increase (Grossman & Frieder, 2004; Göker & Davies, 2009), this dissipates the available storage space. As for probabilistic model the term’s weight estimation is based on how often the term appears or does not appear in relevant documents and non-relevant documents, its disadvantage is the difficulties to access the information of relevant and non-relevant documents (Göker & Davies, 2009).

In statistical model, both probability and VSM rank the retrieved documents. However, these heavy index-based approaches require large indexing storage, demand extra effort and pose sparse data representation. A weight for a term in a document vector is non-zero only if there is term match. This is not suitable for a large document collection where the document vectors are likely to contain mostly zeros when there is no term match in the documents. This produces sparse matrix offline which is time consuming and not necessarily useful when no query regarding the recorded terms are handled (Grossman & Frieder, 2004). Adding a new document will also change document frequencies of term occurrences which alter vector lengths of every document and re-ranking is needed using the new vector value (Göker & Davies, 2009).

The retrieval in this research involves execution of finding matches between medical documents *(MR)* and the Type 2 or Type 3 expanded query list *(EQ)* to retrieve relevant medical documents *(RMR)*. The result is difficult to be ranked according to its relevancy to the query since it is not numerical weight-based retrieved document. Although it can provide high recall value, it still remains meaningless if the highly relevant document is in the low position which contributes to low precision value.

The drawback of Boolean model which is based on the decision criterion of a document to be either relevant or non-relevant without any notion of grading scale is overcome by a ranking mechanism which is very important to arrange the retrieved relevant documents in order of relevance and demonstrate which relevant document is the best match for the particular query. Two ranking models are introduced namely Comprehensive and MedHieCon ranking models which only have to operate on the set of relevant retrieved medical documents, *RMR* as output of the earlier retrieval process without being intertwined with the retrieval process at the previous stage involving all the documents in the collection. Comprehensive ranking model ranks the relevant documents based on the size of matched terms between query and documents associated with the total of terms occurrences. In contrast, MedHieCon ranking model is based on semantic type of matched terms taken from Medical Conceptual (MC) folder in MeSH-indexer. The motivation to create MedHieCon ranking model is based on the weakness found during the execution of Comprehensive ranking model.

The term frequency of appearance in the document is not correlated with the importance and the relevance of the document in the ranking position. It is significant to know the semantic representation of the matched terms in order to rank the documents based on the best relevant documents that match the query. The application of semantic representation of modality, anatomy and pathology concepts in MedHieCon ranking model has shown 11% improvement in MAP value which is 0.27 compared to Comprehensive model which only obtains MAP value of 0.24; which answers the fourth research question on the suitable process to rank retrieved relevant documents. We have also achieved the fourth sub-objective in the first research objective (1d) to form the strategies and steps in retrieval and ranking of relevant documents. Again sub-objective (1e) is met in conducting evaluation of the framework.

Table 8.9 lists the MAP value of M3IRS text-based framework and other text-based run systems that used the same ImageCLEF 2010 data collection. M3IRS is in the 6th top rank for MAP value comparison and for p@10 evaluation out of twenty run systems. This observation shows that although there is no relevance weightage in Boolean model for retrieval strategy compared to VSM but using the right and suitable ranking mechanism such as MedHieCon semantic ranking model, it can produce results that are comparable to other run systems that apply VSM. Again objective (1e) is met in conducting evaluation of the framework.

9.4 Findings in M3ICS Content-based Framework

To achieve our second objective, M3ICS content-based framework is developed. M3ICS content-based framework highlights the architecture of visual feature extraction methodology which consists of two main components namely feature extraction and classification of conceptual train data.

9.4.1 Visual Features for M3ICS Content-based Framework

The feature extraction of medical images involve global and local descriptor processes while features of texture, shape and color are used as information attributes for multi-modality medical image for both descriptors. These features are then combined into one feature vector. Two classifiers namely k-NN with k=3 and SVM using polykernel are used for classification and the division of 75% train data and 25% test data is applied in order to achieve our third sub-objective in second research objective (2c) in identifying and applying suitable machine learning technique for multi-modality medical image classification.

In global descriptor process, features are extracted for the whole medical image with four different directions (0°, 45°, 90° and 180°) used to extract texture feature with total of 75 dimensions in a feature vector. For local descriptor process, features are extracted based on patches and blocks of medical images. The results of the experiments done in this research work show that 4×4 patches (108-dimensional) produce better results than 2×2 and 8×8 patches with MAP value of 0.121 and 83.17% of correctness rate in SVM. To extract features in interest blocks, 20 interest points are generated in each medical image. For each interest point, a block of 20×20 pixels is created, and texture and shape features are extracted for each block. From the experiment result, it can be seen that interest blocks in local descriptor are 5.5% better than patches with MAP value of 0.128, and 85.43% of correctness rate. We have achieved our first and fourth sub-objective in second research objective (2a and 2d) to form the methodology of indexing in visual features extraction and conduct the experiment in comparison of global and local descriptors. This is also discussed below.

Nevertheless combination of patches and interest blocks in extracted features for local descriptor outperform both patches and interest blocks solely with MAP value of 0.135. The explanation of extracting features using both processes in global and local descriptors has answered the research question number 5 on the process involved to extract visual features.

9.4.2 Semantic Description of Medical Concepts in M3IRS Image Classification

The second main component of M3IRS content-based framework is conceptual train data. To achieve second sub-objective (2b) in our second objective, to create a platform for semantic description of visual features classification, MedHieCon model is applied by training the medical images into three medical concepts namely modality, anatomy and pathology. 14,573 medical images taken from ImageCLEF 2010 relevant images in the medical task collection are trained based on these medical concepts. For this research, modality contains 7 classes, anatomy contains 10 classes and pathology has 12 classes. These classes are labelled based on the concepts contained in the 16 ad hoc queries in ImageCLEF 2010 medical task. To train modality concept, global descriptor process is used which only involves the extraction of whole image. For anatomy concept, local descriptor of patches and interest blocks are used in order to extract detail information in the medical images. As for pathology concept, the combination of global and local descriptor processes is applied. The result obtained from the experiment shows that pathology concept obtains the highest correctness rate in both k-NN and SVM classifiers compared to other concepts. This indicates that the more information extracted from the image the higher the correctness rate is achieved. This also proves that using semantic description of MedHieCon model in classification of multi-modality medical images improves the M3ICS performance by 5% with MAP value of 0.142 in SVM classifier compared to combination of global and local descriptors without the model with MAP value of 0.135. Hence this answers the research question number 6 on the application of semantic description in affecting meaningful retrieval system. MedHieCon model is also implemented in text-based M3IRS for its semantic significance.

9.5 Findings in IFM3IRS Information Fusion Framework

The third objective of this research is to design multi-modality medical information retrieval framework based on information fusion of text and visual content. IFM3IRS information fusion framework is developed. This framework is a hierarchical processing of late fusion technique of text and content-based framework, which means that the result from text-based framework process is automatically the input for the content-based framework. The drawback of the text-based retrieval is that the result obtained is solely based on document information. Therefore, content-based approach can be used to filter the result based on the requirement in the query. In this research, content-based framework is used to filter the modality requested by the query. The result shows that the support of M3IRS content-based framework has increased the number of relevant data and improved the performance of IFM3IRS up to 6% with MAP value of 0.287. IFM3IRS framework has achieved 1st place from six run systems that used the same ImageCLEF 2010 medical task data as shown in Table 8.19. This result shows that although content-based does not perform as well as the text-based, but it can complement and support text-based to improve the performance. This answers the research question number 7 in combining different sources of text and visual content in affecting performance of multi-modality medical information retrieval system. The descriptions above show that we meet the sub-objectives (3a) in forming the steps for an integrated framework of text and content-based information sources; (3b) conducting experiments on IFM3IRS; and (3c) performing unit and integration evaluations with different features (text, visual and information fusion) and parameter setting in signifying the strength of our approach. It involves finding suitable ranking model for text-based retrieval and medical concept for modality filtering in content-based classification used in IFM3IRS framework. As for IFM3IRS framework, MedHieCon ranking model is used in text-based retrieval and only global descriptor of visual feature extraction is applied to filter the modality of medical image in content-based processing. Furthermore, the automated input from the text-based to content-based framework has made the system easier to use since the user especially a beginner does not have to manually process or execute the system between text and content-based processing.

9.6 Overall Conclusion

Three frameworks have been developed and experiments conducted on text-based M3IRS, content-based M3ICS and IFM3IRS. The experiments executed in this research show that M3IRS text-based framework outperforms M3ICS content-based framework. Representation of medical data in text form is distinguishable and easier with the ability of text features to extract the term based on its specific domain. With the help of external knowledge source such as thesaurus, the term is more meaningful and significant. However, the drawback of text-based features is the difficulty in understanding the content of image. The list of retrieved data in the text-based may be corrected but to have an accurate result which is really based on query requirement, content-based features are required to filter the image based on the request of a particular query. Therefore information fusion of text and visual content is the solution to solve the problem. Combination of text and content-based information using hierarchical processing of late fusion technique can increase the accuracy of result and the performance in MIR system.

In this research, IFM3IRS has been developed for specific domain of medical data. This framework consists of hierarchical process of text and content-based retrieval system. The text-based retrieval is executed first and the result automatically becomes the input for the content-based retrieval system. In order to rank the results in order of relevance based on the best match of query and data will be on the top of the list, medical concepts are applied for both text and content-based retrieval systems. These concepts are modality, anatomy and pathology. The hierarchy is such that modality is on top whereby the information is more general, followed by anatomy and pathology (the top priority) with the information getting more specific towards the bottom of the hierarchy.

9.7 Future Research Direction

This research has set up a platform to fuse both text and content-based in one automated structured process. It also shows that although content-based result is not as good as that of text-based, it can be used to support text-based system to increase the performance in retrieval system. There are many techniques to view and compare the results between these frameworks (M3IRS, M3ICS and IFM3IRS). Therefore for further enhancements of test and measurement within these frameworks more statistical significance experiments need to be executed.

In specific domain such as medical domain, it is important to have external knowledge source such as thesaurus in order to extract significant features and terms in the text document. Some of the medical terms used in this research are not in the MeSH thesaurus. Therefore in future, other thesaurus such as UMLS and GO may be used based on our integrated architecture.

Besides medical domain there are also other specific domains of interest. Other feature extraction methods in visual features and classification methods can also be selected in line with the data.

The concepts used in this research are focused on modality, anatomy and pathology. Future research can consider other concepts such as the gender of patient, age and treatment in order to have solution for case-based query and not just ad hoc query.

BIBLIOGRAPHY

Acharya, T. & Ray, A. K. (2005). Image processing: Principles and applications Wiley-Interscience.

Adams Jr, J. E. (2010). The essential guide to image processing. Journal of Electronic Imaging, 19, 029901.

Adiego, J., Puente, P. & Navarro, G. (2004). Merging prediction by partial matching with structural contexts model. Data Compression Conference, 2004. Proceedings. DCC 2004, pp. 522.

Ah-Pine, J., Cifarelli, C., Clinchant, S., Csurka, G. & Renders, J. (2008). XRCE's participation to ImageCLEF 2008. Working Notes of the 2008 CLEF Workshop.

Alexander, J. M. K. (2002). Random Boolean networks and evolutionary game theory.

Ammari, H. (2008). An introduction to mathematics of emerging biomedical imaging. Springer Verlag.

Antani, S. (2010). Integrating image and text information for biomedical information retrieval. Computer-Based Medical Systems (CBMS), 2010 IEEE 23rd International Symposium on, pp. 3-3.

Antani, S., Long, L.R. & Thomas, G.R., (2002) A biomedical information system for combined content-based retrieval of spine x-ray images and associated text information, Proceedings of the Indian Conference on Computer Vision, Graphics, and Image Processing.

Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B. & Ferro, S. (2004). UniProt: The universal protein knowledgebase. Nucleic Acids Research, 32 (suppl 1), pp. D115-D119.

Aronson, A. R. (2001). Effective mapping of biomedical text to the UMLS metathesaurus: The MetaMap program. Proceedings of the AMIA Symposium, pp. 17.

Aronson, A. R. (2006). MetaMap: Mapping text to the UMLS metathesaurus. Bethesda, MD: NLM, NIH, DHHS.

Aronson, A. R. & Lang, F. M. (2010). An overview of MetaMap: Historical perspective and recent advances. Journal of the American Medical Informatics Association, 17(3), pp. 229-236.

Babaei, S. & Geranmayeh, A. (2009). Heart sound reproduction based on neural network classification of cardiac valve disorders using wavelet transforms of PCG signals. Computers in Biology and Medicine, 39(1), pp. 8-15.

Baeza-Yates, R. & Ribeiro-Neto, B. (1999). Modern information retrieval. Addison-Wesley New York.

Bay, S. D. (1998). Combining nearest neighbor classifiers through multiple feature subsets. Proceedings of the Fifteenth International Conference on Machine Learning, 3. pp. 37-45.

Beall, M. W. & Shephard, M. S. (1997). A general topology-based mesh data structure. International Journal for Numerical Methods in Engineering, 40(9), pp. 1573-1596.

Bedrick, S. & Kalpathy-Cramer, J. (2010). Improving retrieval using external annotations: OHSU at ImageCLEF 2010. CLEF (Notebook Papers/LABs/Workshops)'10.

Berg, M. (2001). Implementing information systems in health care organizations: Myths and challenges. International Journal of Medical Informatics, 64(2), pp. 143-156.

Bhatia, N. (2010). Survey of nearest neighbor techniques. International Journal of Computer Science and Information Security, Vol. 8, No. 2, 2010

Bhogal, J., MacFarlane, A. & Smith, P. (2007). A review of ontology based query expansion. Information Processing & Management, 43(4), pp. 866-886.

Bodenreider, O. (2004). The unified medical language system (UMLS): Integrating biomedical terminology. Available: http://www.nlm.nih.gov/research/umls/. Last accessed 17 March 2012.

Boström, H., Andler, S. F., Brohede, M., Johansson, R., Karlsson, A. & Laere, J. (2007). On the definition of information fusion as a field of research. School of Humanities and Informatics, University of Skövde, Tech.Rep.HS-IKI-TR-07-006.

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognition, 30(7), pp. 1145-1159.

Browne, M. W. (2000). Cross-validation methods. Journal of Mathematical Psychology, 44(1), pp. 108-132.

Buckley, C. & Voorhees, E. M. (2000). Evaluating evaluation measure stability. Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 33-40.

Buckley, C., Salton, G. & Allan, J. (1994). The effect of adding relevance information in a relevance feedback environment. Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 292-300.

Burges, C. J. C. (1998). A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery, 2(2), pp. 121-167.

Carson, C., Belongie, S., Greenspan, H. & Malik, J. (2002) Blobworld: Image Segmentation Using Expectation-Maximization and Its Application to Image Querying in IEEE Trans. On PAMI, vol. 24, No.8, pp. 1026-1038.

Cavnar, W. B. & Trenkle, J. M. (1994). N-gram-based text categorization. Ann Arbor MI, 48113(2), pp. 161-175.

Chang, C. C. & Lin, C. J. (2011). LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST), 2(3), 27.

Chbeir, R., Amghar, Y. & Flory, A. (2000). MIMS: A prototype for medical image retrieval. Recherche d'Informations Assistee Par Ordinateur

Chen, Y. & Wang, J. Z. (2002) A Region-Based Fuzzy Feature Matching Approach to Content-Based Image Retrieval, in IEEE Trans. on PAMI, vol. 24, No.9, pp. 1252-1267.

Chen, Y. T. & Tseng, D. C. (2007). Wavelet-based medical image compression with adaptive prediction. Computerized Medical Imaging and Graphics, 31(1), pp. 1-8.

Cherry, S. R. (2003). Physics in Nuclear Medicine, Journal of Nuclear Medicine

Chevallet, J. P., Lim, J. H., Radhouani, S. & France, C. I. (2005). Using ontology dimensions and negative expansion to solve precise queries in clef medical task. Working Notes of the 2005 CLEF Workshop, (Vienna, Austria)

Colas, F. & Brazdil, P. (2006). Comparison of svm and some older classification algorithms in text classification tasks. Artificial Intelligence in Theory and Practice, pp. 169-178.

Comaniciu, D., Meer, P., Foran, D. & Medl, A. (1998). Bimodal system for interactive indexing and retrieval of pathology images. Proceedings of the 4th IEEE Workshop on Applications of Computer Vision (WACV'98), 76.

Cootes, T.F. and Taylor, C.J., (2001) Statistical models of appearance for medical image analysis and computer vision, Proc. SPIE Medical Imaging

Cox, I. J., Miller, M. L., Minka, T. P., Papathomas, T. V. & Yianilos, P. N. (2000). The bayesian image retrieval system, PicHunter: Theory, implementation, and psychophysical experiments. Image Processing, IEEE Transactions on, 9(1), pp. 20-37.

Crete, F., Dolmiere, T., Ladret, P., & Nicolas, M. (2007). The blur effect: Perception and estimation with a new no-reference perceptual blur metric. Human Vision and Electronic Imaging XII, 6492, 11.

Crouch, C. J. (1990). An approach to the automatic construction of global thesauri. Information Processing & Management, 26(5), 629-640.

Crouch, C. J., & Yang, B. (1992). Experiments in automatic statistical thesaurus construction. Proceedings of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 77-88.

Csurka, G. & Perronnin, F. (2011). Fisher vectors: Beyond bag-of-visual-words image representations. Computer Vision, Imaging and Computer Graphics.Theory and Applications, pp. 28-42.

Darabos, C., Di Cunto, F., Tomassini, M., Moore, J. H., Provero, P. & Giacobini, M. (2011). Additive functions in boolean models of gene regulatory network modules. PloS One, 6(11), e25110.

Datta, R., Joshi, D., Li, J. & Wang, J. Z. (2008). Image retrieval: Ideas, influences, and trends of the new age. ACM Computing Surveys (CSUR), 40(2), 5.

Daumke, P., Paetzold, J. & Marko, K. (2010). Morphosaurus in ImageCLEF 2006: The effect of subwords on biomedical IR. Evaluation of Multilingual and Multi-Modal Information Retrieval, pp. 652-659.

De Boer, J. F., Cense, B., Park, B. H., Pierce, M. C., Tearney, G. J. & Bouma, B. E. (2003). Improved signal-to-noise ratio in spectral-domain compared with time-domain optical coherence tomography. *Optics Letters*, *28*(21), pp. 2067-2069.

Del Bimbo, A. (1999). Visual information retrieval (The Morgan Kaufmann Series in Multimedia Information and Systems), Academic Press

Depeursinge, A. & Müller, H. (2010). Fusion techniques for combining textual and visual information retrieval. ImageCLEF, pp. 95-114.

Depeursinge, A., Greenspan, H., Syeda, T. & Müller, H. (2012). Overview of the second workshop on medical Content–Based retrieval for clinical decision support. Medical Content-Based Retrieval for Clinical Decision Support, pp. 1-11.

Depeursinge, A., Racoceanu, D., Iavindrasana, J., Cohen, G., Platon, A., Poletti, P. A. (2010). Fusing visual and clinical information for lung tissue classification in high-resolution computed tomography. Artificial Intelligence in Medicine, 50(1), pp. 13-21.

Deselaers, T. & Ney, H. (2008) Deformations, patches, and discriminative models for automatic annotation of medical radiographs, Pattern Recognition Letters

Deselaers, T., Weyand, T., Keysers, D., Macherey, W. & Ney, H. (2006). FIRE in ImageCLEF 2005: Combining content-based image retrieval with textual information retrieval. Accessing Multilingual Information Repositories, pp. 652-661.

Díaz-Galiano, M., García-Cumbreras, M., Martín-Valdivia, M., Ureña-López, L. & Montejo-Ráez, A. (2009). Query expansion on medical image retrieval: MeSH vs. UMLS. Evaluating Systems for Multilingual and Multimodal Information Access, pp. 732-735.

Díaz-Galiano, M., Martín-Valdivia, M. & Ureṇa-López, L. (2009). Query expansion with a medical ontology to improve a multimodal information retrieval system. Computers in Biology and Medicine, 39(4), pp. 396-403.

Dıaz-Galiano, M., Martın-Valdivia, M., Montejo-Raez, A., & Garcıa-Cumbreras, M. (2009). SINAI at ImageCLEF 2010 medical task.

Diem, L., Chevallet, J. P. & Thuy, D. T. B. (2007). Thesaurus-based query and document expansion in conceptual indexing with umls.

Dillon, M., Ulmschneider, J. & Desper, J. (1983). A prevalence formula for automatic relevance feedback in boolean systems. Information Processing & Management, 19(1), pp. 27-36.

Dilmanian, F., Zhong, Z., Ren, B., Wu, X., Chapman, L., Orion, I., et al. (2000). Computed tomography of x-ray index of refraction using the diffraction enhanced imaging method. Physics in Medicine and Biology, 45, 933.

Dinh, D., & Tamine, L. (2010) IRIT at ImageCLEF 2010: Medical retrieval track.

Dinh, D., & Tamine, L. (2011). Combining global and local semantic contexts for improving biomedical information retrieval. Advances in Information Retrieval, pp. 375-386.

Dougherty, E. R., & Barrera, J. (2002). Pattern recognition theory in nonlinear signal processing. Journal of Mathematical Imaging and Vision, 16(3), pp. 181-197.

Drymonas, E., Zervanou, K., & Petrakis, E. G. M. (2010). Exploiting multi-word similarity for retrieval in medical document collections: The TSRM approach. Journal of Digital Information Management, 8(5), pp. 315-321.

Duch, W., Jankowski, N., Grabczewski, K., Naud, A., & Adamczak, R.(2008) Ghostminer Data Mining Software, Technical Report, Department of Informatics, Nicolaus Copernicus University. Available at http://www.fqspl.com.pl/ghostminer/. Last accessed 22 March 2012

Duin, A. K. J. R. P. W., & Mao, J. (2000) Statistical pattern recognition: A review. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 22(1), pp. 4–37.

Efthimiadis, E. N. (2000). Interactive query expansion: A user-based evaluation in a relevance feedback environment. Journal of the American Society for Information Science, 51(11), pp. 989-1003.

Egan, J. P. (1975). Signal detection theory and ROC analysis, Academic Press

Elberrichi, Z., Rahmoun, A. & Bentaalah, M. A. (2008). Using wordnet for text categorization. The International Arab Journal of Information Technology, 5(1), pp. 16-24.

Fass, L. (2008). Imaging and cancer: A review. Molecular Oncology, 2(2), pp. 115-152.

Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., et al. (1995). Query by image and video content: The QBIC system. Computer, 28(9), pp. 23-32.

Fox, C. (1992). Information retrieval: Data structures and algorithms. Lexical Analysis and Stoplists, pp. 102-130.

García, M. & Puig, D. (2002) Improving texture pattern recognition by integration of multiple texture feature extraction methods. 16th IAPR int. Conf.on Pattern Recognition, 3, pp. 7-10.

Gass, T., Weyand, T., Deselaers, T. & Ney, H. (2008). FIRE in ImageCLEF 2007: Support vector machines and logistic models to fuse image descriptors for photo retrieval. Advances in Multilingual and Multimodal Information Retrieval, pp. 492-499.

Ghebreab, S., Jaffe, C. & Smeulders, A. (2003). Concept-based retrieval of biomedical images. Proceedings of SPIE Medical Imaging: PACS and Integrated Medical Information Systems: Design and Evaluation

Göker, A., & Davies, J. (2009). Information retrieval: Searching in the 21st century John Wiley & Sons Inc.

Gono, K., Obi, T., Yamaguchi, M., Ohyama, N., Machida, H., Sano, Y., et al. (2004). Appearance of enhanced tissue features in narrow-band endoscopic imaging. Journal of Biomedical Optics, 9, 568.

Grossman, D. A., & Frieder, O. (2004). Information retrieval: Algorithms and heuristics Kluwer Academic Pub.

Gupta, A. (1996). Visual information retrieval technology: A virage perspective. Virage Inc., TR 3A

Hall, D. L. & Llinas, J. (1997). An introduction to multisensor data fusion. Proceedings of the IEEE, 85(1), pp. 6-23.

Han, X. H., & Chen, Y. W. (2011). Biomedical imaging modality classification using combined visual features and textual terms. Journal of Biomedical Imaging, 2011, 5.

Haralick, R. M., Shanmugam, K. & Dinstein, I. H. (1973). Textural features for image classification. Systems, Man and Cybernetics, IEEE Transactions on, 3(6), pp. 610-621.

Hare, J. S., Samangooei, S., Lewis, P. H., & Nixon, M. S. (2008). Semantic spaces revisited: Investigating the performance of auto-annotation and semantic retrieval using semantic spaces. Proceedings of the 2008 International Conference on Content-Based Image and Video Retrieval, pp. 359-368.

Harris, C., & Stephens, M. (1988). A combined corner and edge detector. Alvey Vision Conference, 15. pp. 50.

Hatcher, E., Gospodnetic, O., & McCandless, M. (2004). Lucene in action, Maning Publication

Haux, R. (2006). Health information systems-past, present, future. International Journal of Medical Informatics, 75(3-4), pp. 268-281.

Hazra, I. & Aditi, S. (2009) Thesaurus and Query Expansion, International Journal of Computer science & Information Technology (IJCSIT), Vol 1, No 2, November, p pp. 89-97.

Hearst, M. A., Divoli, A., Guturu, H., Ksikes, A., Nakov, P., Wooldridge, M. A., et al. (2007). BioText search engine: Beyond abstract search. Bioinformatics, 23(16), 2196.

Hersh, W. R., Bhupatiraju, R. T., & Price, S. (2003). Phrases, boosting, and query expansion using external knowledge resources for genomic information retrieval. TREC, pp. 503-509.

Hiemstra, D. (2000). A probabilistic justification for using tf× idf term weighting in information retrieval. International Journal on Digital Libraries, 3(2), pp. 131-139.

Hiemstra, D. (2001). Using language models for information retrieval, PhD Thesis, University of Twente

Hiemstra, D. (2009). Information retrieval models. Information Retrieval, pp. 1-19.

Hinneburg, A., Aggarwal, C. C., & Keim, D. A. (2000) What is the nearest neighbor in high dimensional spaces. Proceedings of the 26th International Conference on Very Large Data Bases

Hiremath, P. S. & Jagadeesh, P. (2007) Content Based Image Retrieval based on Color, Texture and Shape features using Image and its Complement, International Journal of Computer Science and Security, Volume (1) : Issue (4), pp. 25-35.

Hliaoutakis, A., Varelas, G., Petrakis, E., & Milios, E. (2006). MedSearch: A retrieval system for medical information based on semantic similarity. Research and Advanced Technology for Digital Libraries, pp. 512-515

Hliaoutakis, A., Zervanou, K., Petrakis, E. G. M., & Milios, E. E. (2006). Automatic document indexing in large medical collections. Proceedings of the International Workshop on Healthcare Information and Knowledge Management, pp. 1-8.

Hsu, W., Antani, S., Long, L. R., Neve, L., & Thoma, G. R. (2009). SPIRS: A web-based image retrieval system for large biomedical databases. International Journal of Medical Informatics, 78, pp. S13-S24.

Hu, M. K. (1962). Visual pattern recognition by moment invariants. Information Theory, IRE Transactions on, 8(2), pp. 179-187.

Huang, J., Ravi Kumar, S., Mitra, M., Zhu, W. J., & Zabih, R. (1999). Spatial color indexing and applications. International Journal of Computer Vision, 35(3), pp. 245-268.

Huang, W. Y., & Davis, J. J. (2011). Multimodality and nanoparticles in medical imaging. Dalton Trans 40, pp. 6087-6103

Ibrahim, R., & Arafa, W. (2010) Medical image retrieval: ISSR at CLEF 2010. In CLEF (Notebook Papers/LABs/Workshops

IHTSDO, (2007). International Health Terminology Standards Development Organisation. Available: http://www.nlm.nih.gov/research/umls/Snomed/snomed_main.html. Last accessed 17 March 2012

Iskandrian, A. E., & Garcia, E. V. (2008). Nuclear cardiac imaging: Principles and applications Oxford University Press, USA.

Järvelin, A., Wilkins, P., Adamek, T., Airio, E., Jones, G., Smeaton, A., et al. (2008). DCU and UTA at ImageCLEFPhoto 2007. Advances in Multilingual and Multimodal Information Retrieval, pp. 530-537.

Jeffrey S. Klein, MD. (1981). Journal of Continuing Medical Education in Radiology. Available: http://radiographics.rsna.org. Last accessed 21st March 2012

Jing, Y., & Croft, W. B. (1994). An association thesaurus for information retrieval. Proceedings of RIAO, 94. (1994) pp. 146-160.

Johnson, D. E. (1998). Applied multivariate methods for data analysts Duxbury Press Pacific Grove, CA.

Kahn Jr, C. E., & Thao, C. (2007). GoldMiner: A radiology image search engine. American Journal of Roentgenology, 188(6), 1475.

Kalender W. A., (2001). Computed Tomography: Fundamentals, System Technology, Image Quality, Applications, MCD, Munich, Germany

Kalpathy-Cramer, J., & Hersh, W. (2010). Multimodal medical image retrieval: Image categorization to improve search precision. Proceedings of the International Conference on Multimedia Information Retrieval, pp. 165-174.

Kalpathy-Cramer J, & Hersh W, (2008) Effectiveness of global features for automatic medical image classification and retrieval - the experiences of OHSU at ImageCLEFmed, Pattern Recognition Letters, 29: pp. 2032-2038

Keerthi, S. S., Shevade, S. K., Bhattacharyya, C., & Murthy, K. R. K. (2001). Improvements to platt's SMO algorithm for SVM classifier design. Neural Computation, 13(3), pp. 637-649.

Kendall, S. K. (2010). Springer images. The Charleston Advisor, 11(3), pp. 50-52.

Khorasani, E. S., Doraisamy, S., & Azman, A. (2011). Automatic heart diseases detection techniques using musical approaches. Journal of Applied Sciences, 11(17), pp. 3161-3168.

Kitson, S. L., Cuccurullo, V., Ciarmiello, A., Salvo, D., & Mansi, L. (2009). Clinical applications of positron emission tomography (PET) imaging in medicine: Oncology, brain diseases and cardiology. Curr Radiopharm, 2, pp. 224-253.

Kiyan, T. (2011). Breast cancer diagnosis using statistical neural networks. IU-Journal of Electrical & Electronics Engineering, 4(2)

Kludas, J., Bruno, E., & Marchand-Maillet, S. (2008). Information fusion in multimedia information retrieval. Adaptive Multimedia Retrieval: Retrieval, User, and Semantics, pp. 147-159.

Kontostathis, A., & Pottenger, W. M. (2006). A framework for understanding latent semantic indexing (LSI) performance. Information Processing & Management, *42*(1), pp. 56-73.

Kuhn, M. H. (1995). Medical imaging: Analysis of multimodality 2D/3D images. Ios Pr Inc

Kulandavelu, S., Qu, D., Sunn, N., Mu, J., Rennie, M. Y., Whiteley, K. J., et al. (2006). Embryonic and neonatal phenotyping of genetically engineered mice. Ilar Journal, 47(2), 103.

Kuo, W. J., Chang, R. F., Lee, C. C., Moon, W. K. & Chen, D. R. (2002). Retrieval technique for the diagnosis of solid breast tumors on sonogram. Ultrasound in Medicine & Biology, 28(7), pp. 903-909.

Lacoste, C., Lim, J. H., Chevallet, J. P., & Le, D. T. H. (2007). Medical-image retrieval based on knowledge-assisted text and image indexing. Circuits and Systems for Video Technology, IEEE Transactions on, 17(7), pp. 889-900.

Lapeer, R. J., Tan, A. & Aldridge, R. (2002). A combined approach to 3D medical image segmentation using marker-based watersheds and active contours: The active watershed method. Medical Image Understanding and Analysis, 2002

Lee, D. C., & Carroll, T. J. (2010). Magnetic resonance imaging. Practical Signal and Image Processing in Clinical Cardiology, pp. 251-273.

Lehmann, T. M., Gold, M., Thies, C., Fischer, B., Spitzer, K., Keysers, D., et al. (2004). Content-based image retrieval in medical applications. Methods of Information in Medicine, 43(4), pp. 354-361.

Lehmann, T. M., Guld, M. O., Deselaers, T., Keysers, D., Schubert, H., Spitzer, K., et al. (2005). Automatic categorization of medical images for content-based retrieval and data mining. Computerized Medical Imaging and Graphics : The Official Journal of the Computerized Medical Imaging Society, 29(2-3), pp. 143-155.

Lehmann, T. M., Schubert, H., Keysers, D., Kohnen, M., & Wein, B. B. (2003). The IRMA code for unique classification of medical images. Proceedings SPIE, 5033. pp. 109-117.

Lehmann, T. M., Wein, B., Dahmen, J., Bredno, J., Vogelsang, F., & Kohnen, M. (2000). Content-based image retrieval in medical applications: A novel multi-step approach. Proceedings SPIE, 3972(32), pp. 312-320.

Lehmann, T., et al. (2003). The irma project: A state of the art report on content-based image retrieval in medical applications. Available: http://irma project.org/index_en.php. Last accessed 17 March 2012.

Lehmann, T., Güld, M., Thies, C., Fischer, B., Spitzer, K., Keysers, D., et al. (2003). The IRMA project: A state of the art report on content-based image retrieval in medical applications. Korea-Germany Workshop on Advanced Medical Image, pp. 161-171.

Lei Zeng, M., & Mai Chan, L. (2004). Trends and issues in establishing interoperability among knowledge organization systems. Journal of the American Society for Information Science and Technology, 55(5), pp. 377-395.

Li, J., Wang, J.Z.& Wiederhold, G. (2000) IRM: Integrated Region Matching for Image Retrieval in Proc. of the 8th ACM Int. Conf. on Multimedia, pp. 147-156.

Liang, Z. (2008). Introduction to biomedical imaging. Medical Devices and Biosensors, 2008. ISSS-MDBS 2008. 5th International Summer School and Symposium on, pp. 12-12.

Liddy, E. D., Paik, W., & Woelfel, J. K. (1993). Use of subject field codes from a machine-readable dictionary for automatic classification of documents. Advances in Classification Research, Proceedings of the Third ASIS SIG/CR Classification Workshop, pp. 83-100.

Lim, J. H., Chevallet, J. P., Le, D. T. H., & Goh, H. (2008). Bi-modal conceptual indexing for medical image retrieval. Proceedings of the 14th International Conference on Advances in Multimedia Modeling, pp. 456-465.

Liu, J. and Hu, Y. and Li, M. and Ma, S. and Ma, W., (2006) Medical image annotation and retrieval using visual features, Evaluation of Multilingual and Multi-modal Information Retrieval.

Liu, Y., Lazar, N., Rothfus, W., Dellaert, F., Moore, A., Schneider, J., et al. (2004). Semantic based biomedical image indexing and retrieval. Trends and Advances in Content-Based Image and Video Retrieval.Springer,

Liu, Y., Zhang, D., Lu, G., & Ma, W. Y. (2007). A survey of content-based image retrieval with high-level semantics. Pattern Recognition, 40(1), pp. 262-282.

Liu, Z., Wu, Q., Zhang, Y., & Philip Chen, C. (2011). Adaptive least squares support vector machines filter for hand tremor canceling in microsurgery. International Journal of Machine Learning and Cybernetics, 2(1), pp. 37-47.

Loh, W. Y. (2011). Classification and regression trees. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1(1), pp. 14-23.

Long, L. R., Antani, S., Lee, D. J., Krainak, D. M., & Thoma, G. R. (2003). Biomedical information from a national collection of spine x-rays: Film to content-based retrieval. Proceedings SPIE, , 5033. pp. 70-84.

Losa, G. A., & Castelli, C. (2005). Nuclear patterns of human breast cancer cells during apoptosis: Characterisation by fractal dimension and co-occurrence matrix statistics. Cell and Tissue Research, 322(2), pp. 257-267.

Lu, Z. (2011). PubMed and beyond: A survey of web tools for searching biomedical literature. Database: The Journal of Biological Databases and Curation, 2011

Lu, Z., Kim, W., & Wilbur, W. J. (2009). Evaluation of query expansion using MeSH in PubMed. Information Retrieval, 12(1), pp. 69-80.

Lucarella, D. (1988). A search strategy for large document bases. Electronic Publishing, 1(2), pp. 105-116.

Luo, Y., & Celenk, M. (2005). Topographic modeling of cellular images. Conference Proceedings Annual International Conference of the IEEE Engineering in Medicine and Biology Society.IEEE Engineering in Medicine and Biology Society.Conference, 4, 3986-3989.

MacArthur, S. D., Brodley, C. E., & Shyu, C. R. (2000). Relevance feedback decision trees in content-based image retrieval. Content-Based Access of Image and Video Libraries, 2000. Proceedings. IEEE Workshop on, pp. 68-72.

Macmillan Cancer Support (2010). How secondary bone cancer is diagnosed . Available: http://www.macmillan.org.uk/Cancerinformation/Cancertypes/Bonesecondary/Symptomsdiagnosis/Diagnosis.aspx. Last accessed on 18 March 2012.

Madzin, H., & Zainuddin, R. (2009). Feature extraction and image matching of 3D lung cancer cell image. Soft Computing and Pattern Recognition, 2009. SOCPAR'09. International Conference of, pp. 511-515.

Madzin, H., & Zainuddin, R. (2009). Medical image analysis using local invariant features – interest points. International Conference of Software Engineering & Computer Systems 2009 (ICSECS'09), pp. 303-307.

Maintz, J. B. A., & Viergever, M. A. (1996). An overview of medical image registration methods. Symposium of the Belgian Hospital Physicists Association (SBPH/BVZF), 12. pp. 1-22.

Malpica, N., Santos, A., Tejedor, A., Torres, A., Castilla, M., Garcia-Barreno, P., et al. (2003). Automatic quantification of viability in epithelial cell cultures by texture analysis. Journal of Microscopy, 209(1), pp. 34-40.

Manjunath, B. S., Ohm, J. R., Vasudevan, V. V., & Yamada, A. (2001). Color and texture descriptors. Circuits and Systems for Video Technology, IEEE Transactions on, 11(6), pp. 703-715.

Marcoulides, G. A. (2005). Discovering knowledge in data: An introduction to data mining. Journal of the American Statistical Association, 100(472), pp. 1465-1465.

Marcus, R. S. (1991). Computer and human understanding in intelligent retrieval assistance. Proceedings of the ASIS Annual Meeting, 28. pp. 49-59.

Marée, R., Geurts, P., & Wehenkel, L. (2007). Content-based image retrieval by indexing random subwindows with randomized trees. Proceeding of ACCV 2007, LNCS,

Medlock, B. (2008). Investigating classification for natural language processing tasks VDM Verlag Dr. Müller.

MedPix (2009) Free Online Medical Image Database by Department of radiology and Biomedical Informatics, Unifformed Services Universuty, USA. Available at: http://rad.usuhs.mil/medpix/medpix.html. Last accessed 1st July 2012

Meyer, I., & Mackintosh, K. (1994). Phraseme analysis and concept analysis: Exploring a symbiotic relationship in the specialized lexicon. Martin, W.Et al., pp. 39-348.

Mihalcea, R., Corley, C., & Strapparava, C. (2006). Corpus-based and knowledge-based measures of text semantic similarity. Proceedings of the National Conference on Artificial Intelligence, 21. (1) pp. 775.

Milanese, R., & Cherbuliez, M. (1999). A rotation, translation, and scale-invariant approach to content-based image retrieval. Journal of Visual Communication and Image Representation, 10(2), pp. 186-196.

Mineev, A. K. (2007). Morphological analysis and pathological changes in the structure of blood cells in fishes from saratov reservoir. Journal of Ichthyology, 47(1), pp. 87-93.

Mitra, M., Singhal, A., & Buckley, C. (1998). Improving automatic query expansion. Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 206-214.

Miyahara, M., Kotani, K., & Algazi, V. R. (1998). Objective picture quality scale (PQS) for image coding. Communications, IEEE Transactions on, 46(9), pp. 1215-1226.

Mojsilovic, A. and Gomes, J., (2002) Semantic based categorization, browsing and retrieval in medical image databases, International Conference on Image Processing

Mojsilovic, A., Kovacevic, J., Hu, J., Safranek, R. J., & Ganapathy, S. K. (2000). Matching and retrieval based on the vocabulary and grammar of color patterns. Image Processing, IEEE Transactions on, 9(1), pp. 38-54.

Mueen, A., Zainuddin, R., & Sapiyan Baba, M. (2010). MIARS: A medical image retrieval system. Journal of Medical Systems, 34(5), pp. 859-864.

Müller, H., Clough, P., & Deselaers, T. (2010). ImageCLEF: Experimental evaluation in visual information retrieval Springer-Verlag New York Inc.

Müller, H., Deselaers, T., Deserno, T., Kalpathy–Cramer, J., Kim, E., & Hersh, W. (2008). Overview of the ImageCLEFmed 2007 medical retrieval and medical annotation tasks. Advances in Multilingual and Multimodal Information Retrieval, pp. 472-491.

Müller, H., et al. (2010). Overview of the CLEF 2010 medical image retrieval track. CLEF. Available: http://clef2010.org/index.php. Last accessed 17 March 2012.

Müller, H., Kalpathy–Cramer, J., Eggel, I., Bedrick, S., Radhouani, S., Bakke, B., et al. (2011). Overview of the CLEF 2009 medical image retrieval track. Multilingual Information Access Evaluation II.Multimedia Experiments, pp. 72-84.

Müller, H., Kalpathy-Cramer, J., Eggel, I., Bedrick, S., Reisetter, J., Jr., C.E.K., and Hersh, W.R. (2010). Overview of the CLEF 2010 medical image retrieval track. CLEF (Notebook Papers/LABs/Workshops)'10,

Muller, 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), pp. 1-23.

Müller, H., Müller, W., Marchand-Maillet, S., Pun, T., & Squire, D. M. G. (2003). A framework for benchmarking in CBIR. Multimedia Tools and Applications, 21(1), pp. 55-73.

Myhre, S., Tveit, H., Mollestad, T., & Lægreid, A. (2006). Additional gene ontology structure for improved biological reasoning. Bioinformatics, 22(16), pp. 2020-2027.

Natsev, R. Rastogi, & K. Shim (1999) WALRUS: A Similarity Retrieval Algorithm for Image Databases, in Proc. ACM SIGMOD Int. Conf. Management of Data, pp. 395–406.

Neil,G., & Curtis, K. (1997). Shape recognition using fractal geometry. Pattern Recognition, 30(12), pp. 1957-1969.

Nelson, S. J., et al. (2004). The MeSH translation maintenance system: Structure, interface design, and implementation. Available: http://www.nlm.nih.gov/mesh/meshhome.html. Last accessed 17 March

Nelson,S.J., Schopen, M., Savage, A. G., Schulman, J. L., & Arluk, N. (2004). The MeSH translation maintenance system: Structure, interface design, and implementation. Medinfo, 11(Pt 1), pp. 67-69.

Newsam, S., & Kamath, C. (2005). Comparing shape and texture features for pattern recognition in simulation data. IS&T/SPIE's Annual Symposium on Electronic Imaging, pp. 16-20.

Orphanoudakis, S.C., Chronaki, C. E., & Vamvaka, D. (1996). I2Cnet: Content-based similarity search in geographically distributed repositories of medical images. Computerized Medical Imaging and Graphics, 20(4), pp. 193-207.

Pan, C., Yan, X., & Zheng, C. (2006). Recognition of blood and bone marrow cells using kernel-based image retrieval. IJCSNS, 6(10), 29.

Paredes, J. M., Ribó, M., Bosch-Ramon, V., West, J. R., Butt, Y. M., Torres, D. F., et al. (2007). Chandra observations of the gamma-ray binary LS I 61 303: Extended X-ray structure? The Astrophysical Journal Letters, 664, L39.

Paredes, R. and Keysers, D. and Lehmann, T.M. and Wein, B. and Ney, H. and Vidal, E., (2002)Classification of medical images using local representations, Medizin 2002: Algorithmen-Systeme-Anwendungen. Proceedings Des Workshop Vom 10

Parikshit Sondhi, J. S. (2010). Medical case-based retrieval by leveraging medical ontology and physician feedback: UIUC-IBM at ImageCLEF 2010. CLEF (Notebook Papers/LABs/Workshops) 2010.

Park, D. K., Jeon, Y. S., & Won, C. S. (2000). Efficient use of local edge histogram descriptor. Image, 2, 3.

Park, J. S., & Keller, J. (1997). Fuzzy patch label relaxation in bone marrow cell segmentation.'Computational Cybernetics and Simulation', IEEE International Conference on Systems, Man, and Cybernetics, 1997, 2

Pass, G., Zabih, R., & Miller, J. (1997). Comparing images using color coherence vectors. Proceedings of the Fourth ACM International Conference on Multimedia, pp. 65-73.

Peat, H. J., & Willett, P. (1991). The limitations of term co-occurrence data for query expansion in document retrieval systems. Journal of the American Society for Information Science, 42(5), pp. 378-383.

Pestotnik, S. (2000). Medical informatics: Meeting the information challenges of a changing health care system. Journal of Informed Pharmacotherapy, 2(1)

Peters, C. (2001). Cross-language information retrieval and evaluation: Workshop of the cross-language evaluation forum, CLEF 2000, lisbon, portugal, september 21-22, 2000: Revised papers Springer-Verlag New York Inc.

Petite, V., Dinh, H. Q., & Fisher, E. (2010). Towards a transmedia search engine: A user study on perceiving analogies in multimedia data.

Pizurica, A., Philips, W., Lemahieu, I., & Acheroy, M. (2003). A versatile wavelet domain noise filtration technique for medical imaging. Medical Imaging, IEEE Transactions on, 22(3), pp. 323-331.

Platt, J. (1999). Advances in kernel methods-support vector learning, chapter fast training of support vector machines using sequential minimal optimization. N, 12, pp. 185-208.

Plus, M. (2007). A service of the US national library of medicine and the national institutes of health. Ginkgo (Ginkgo Biloba L.).Last accessed on 21st March 2012

Prasad, M., Sowmya, A., & Wilson, P. (2009). Multi-level classification of emphysema in HRCT lung images. Pattern Analysis & Applications, 12(1), pp. 9-20.

Provost, F., & Fawcett, T. (1997). Analysis and visualization of classifier performance: Comparison under imprecise class and cost distributions. Proceedings of the Third International Conference on Knowledge Discovery and Data Mining, pp. 43-48.

Qian, F., Li, M., Zhang, H. J., Ma, W. Y., & Zhang, B. (2003). Alternating feature spaces in relevance feedback. Multimedia Tools and Applications, 21(1), pp. 35-54.

Quellec, G., Lamard, M., Bekri, L., Cazuguel, G., Roux, C., & Cochener, B. (2010). Medical case retrieval from a committee of decision trees. Information Technology in Biomedicine, IEEE Transactions on, 14(5), pp. 1227-1235

Quinlan, J. R. (1993). C4. 5: Programs for machine learning Morgan kaufmann.

Radhouani, S., Falquet, G., & Chevalletinst, J. P. (2008). Description logic to model a domain specific information retrieval system. Database and Expert Systems Applications, pp. 142-149.

Rahman, M., Antani, S., Fushman, D., & Thoma, G. (2012). Biomedical image retrieval using multimodal context and concept feature spaces. Medical Content-Based Retrieval for Clinical Decision Support, pp. 24-35.

Rahman, M., Desai, B. C., & Bhattacharya, P. (2007). Multi-modal interactive approach to image CLEF 2007 photographic and medical retrieval tasks by CINDI. Working Notes of the 2007 CLEF Workshop.

Ramakrishnan, H. N. (2010). Detection and Estimation of Image Blur, Thesis, Missouri University of Science and Technology

Randen, T., & Husoy, J. (1999). Filtering for texture classification: A comparative study. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 21(4), pp. 291-310.

Rangarajan,L., & Nagabhushan, P. (2004). Dimensionality reduction of multidimensional temporal data through regression. Pattern Recognition Letters, 25(8), pp. 899-910.

Robertson, S. E. (1977). The probability ranking principle in IR. Journal of Documentation, 33(4), pp. 294-304.

Robertson, S. E., & Jones, K. S. (1976). Relevance weighting of search terms. Journal of the American Society for Information Science, 27(3), pp. 129-146.

Rokach, L., & Maimon, O. Z. (2008). Data mining with decision trees: Theroy and applications, Chapter 6: Advance Decision Tree, World Scientific Pub Co Inc.

Ruch, P., Baud, R., Geissbühler, A., Hilario, M., & Pellegrini, C. (2002). Text mining and information retrieval in medical records: An inquiry into automatic spelling correction. NLPBA 2002 Proceedings.
Available at http://www.genisis.ch/~natlang/NLPBA02/program.htm. Last accessed 21st March 2012.

Rui, Y., Huang, T. S., & Chang, S. F. (1997). Image retrieval: Past, present, and future. International Symposium on Multimedia Information Processing, 10. pp. 1-23.

Ruiz, M., Chen, J., Pasupathy, K., Chin, P., & Knudson, R. (2010). UNT at ImageCLEF 2010: Clir for wikipedia images. Working Notes of the ImageCLEF 2010 Lab.

Russ, John C. (2006). The Image Processing Handbook (5th edition ed.). CRC Press

Sahoo, S., & Choubisa, T. (2012). Multimodal biometric person authentication: A review. IETE Technical Review, 29(1), 54.

Saleem, A., Beghdadi, A., & Boashash, B. (2012). Image fusion-based contrast enhancement. EURASIP Journal on Image and Video Processing, 2012, pp. 10.1-10.17.

Salton, G. (1991). Developments in automatic text retrieval. Science, 253(5023), 974.

Salton, G., & Buckley, C. (1997). Improving retrieval performance by relevance feedback. Readings in Information Retrieval, pp. 355-364.

Salton, G., Fox, E. A., & Wu, H. (1983). Extended Boolean information retrieval. Communications of the ACM, 26(11), pp. 1022-1036.

Salton, G., Voorhees, E., & Fox, E. A. (1984). A comparison of two methods for Boolean query relevancy feedback. Information Processing & Management, 20(5-6), pp. 637-651.

Salton, G., Wong, A., & Yang, C. S. (1975). A vector space model for automatic indexing. Communications of the ACM, 18(11), pp. 613-620.

Schutze, H. and J. Pederson. (1997). A cooccurance-based thesaurus and two applications to information retrieval. [Information Processing and Management], 33, no. 3, pp. 307-318.

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

Shambaugh, E., Young, J., Zippin, C., Lum, D., Akers, C., & Weiss, M. (1995). Statistics and epidemiology for cancer registrars, self-instructional manual for cancer registrars, book 7: Surveillance epidemiology and end results program.

Sharef, N. M. & Madzin, H. (2012). IMS:An improved medical retrieval model via medical-context aware query expansion and comprehensive ranking. IEEE International Conference on Information Retrieval and Knowledge Management 2012, Kuala Lumpur.

Sharef,N. M., Martin, T., & Shen, Y. (2008). Incremental evolutionary grammar fragments. Proceedings of the UK Workshop on Computational Intelligence (UKCI 2008), pp. 89-94.

Shyu, C. R., Brodley, C. E., Kak, A. C., Kosaka, A., Aisen, A. M., & Broderick, L. S. (1999). ASSERT: A physician-in-the-loop content-based retrieval system for HRCT image databases. Computer Vision and Image Understanding, 75(1), pp. 111-132.

Signore, A., Mather, S., Piaggio, G., Malviya, G., & Dierckx, R. (2010). Molecular imaging of inflammation/infection: Nuclear medicine and optical imaging agents and methods. Chem Rev, *110*(5), pp. 3112-3145.

Simpson, M., Rahman, M. M., Demner-Fushman, D., Antani, S., & Thoma, G. R. (2009). Text-and content-based approaches to image retrieval for the ImageCLEF 2009 medical retrieval track. CLEF2009 Working Notes. CLEF 2009 Workshop, 30.

Simpson, M., Rahman, M. M., Singhal, S., Demner-Fushman, D., Antani, S., & Thoma, G. (2010) Text-and content-based approaches to image modality detection and retrieval for the ImageCLEF 2010 medical retrieval track.

Sivakumar, R., Karnan, M., & Deepa, G. G. (2004) An improved modified tracking algorithm hybrid with fuzzy C means clustering in digital mammograms. International Journal Computer Technology and Application, 3

Smeaton, A. F., & Van Rijsbergen, C. J. (1983). The retrieval effects of query expansion on a feedback document retrieval system. The Computer Journal, 26(3), pp. 239-246.

Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A., & Jain, R. (2000). Content-based image retrieval at the end of the early years. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 22(12), pp. 1349-1380.

Smith, J. R., & Chang, S. F. (1997). VisualSEEk: A fully automated content-based image query system. Proceedings of the Fourth ACM International Conference on Multimedia, pp. 87-98.

Soille, P. (2003). Morphological image analysis: Principles and applications Springer-Verlag New York, Inc.

Soria,D., Garibaldi, J. M., Ambrogi, F., Biganzoli, E. M., & Ellis, I. O. (2011). A non-parametric'version of the naive bayes classifier. Knowledge-Based Systems,

Sprawls, P. (1995). Physical principles of medical imaging. Medical Physics Pub.

Stevens, R., Goble, C. A., & Bechhofer, S. (2000). Ontology-based knowledge representation for bioinformatics. Briefings in Bioinformatics, 1(4), pp. 398-414.

Stevenson, M., & Guo, Y. (2010). Disambiguation of ambiguous biomedical terms using examples generated from the UMLS metathesaurus. Journal of Biomedical Informatics, 43(5), pp. 762-773.

Stoker, H. S. (2009). General, organic, and biological chemistry. Brooks/Cole.

Stougiannis, A., Gkanogiannis, A., & Kalamboukis, T. (2010). IPL at ImageCLEF 2010. CLEF (Notebook Papers/LABs/Workshops),

Stricker M. & Orengo, M. (1995) Similarity of Color Images in Proc. SPIE Storage and Retrieval for Image and Video Databases, pp. 381-392

Sumanaweera, T., & Liu, D. (2005). Medical image reconstruction with the FFT. GPU Gems, 2, pp. 765-784.

Sural, S., Qian, G., & Pramanik, S. (2002). Segmentation and histogram generation using the HSV color space for image retrieval. Image Processing. 2002. Proceedings. 2002 International Conference on, 2. pp. II-589-II-592 vol. 2.

Susil, R. C., & Taylor, R. H. (2007). Methods and Systems for Image-Guided Surgical Interventions

Swain, M. J., & Ballard, D. H. (1991). Color indexing. International Journal of Computer Vision, 7(1), pp. 11-32.

Swarnalatha, R., & Prasad, D. (2010). A novel technique for extraction of FECG using multi stage adaptive filtering. Journal of Applied Sciences, 10(4), pp. 319-324.

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Better DECISIONS through. Scientific American, 83.

Takeuchi, K., & Collier, N. (2005). Bio-medical entity extraction using support vector machines. Artificial Intelligence in Medicine, 33(2), pp. 125-137.

Tata, S., & Patel, J. M. (2007). Estimating the selectivity of tf-idf based cosine similarity predicates. ACM SIGMOD Record, 36(2), pp. 7-12.

Theis, F. J., & Meyer-Bäse, A. (2009). Biomedical signal analysis: Contemporary methods and applications The MIT Press.

Toennies, K. D. (2012). Digital image acquisition. Guide to Medical Image Analysis, 21-82.

Tuytelaars, T. & Mikolajczyk, K. (2008). Local invariant feature detectors: A survey. Foundations and Trends® in Computer Graphics and Vision, 3(3), pp. 177-280.

U.S. National Library of Medicine. (1998). U.S. National Library of Medicine. Available: http://www.nlm.nih.gov/bsd/pmresources.html. Last accessed 17 March 2012.

Unay, D. and Soldea, O. and Ekin, A. and Cetin, M. and Eril, A, (2008) Automatic annotation of X-ray images: a study on attribute selection, Medical Image Computing and Computer Assisted Intervention Society, 2009.

Vaccari, G. & Saccavini, C. (2006). Radiology informatics and work flow redesign. PsychNology Journal, 4(1), pp. 87-101.

Vale, R. F., Ribeiro-Neto, B. A., de Lima, L. R. S., Laender, A. H. F., & Junior, H. R. F. (2003). Improving text retrieval in medical collections through automatic categorization. String Processing and Information Retrieval, pp. 197-210.

Valk, P. E., Bailey, D. L., Townsend, D. W., & Maisey, M. N. (2004). Positron emission tomography: Basic sciences Springer Verlag.

Vapnik, V. (1998). Statistical learning theory. Wiley-Interscience.

Vapnik, V. N. (1999). An overview of statistical learning theory. Neural Networks, IEEE Transactions on, 10(5), pp. 988-999.

Vijila, C. K. S., Kanagasabapathy, P., & Johnson, S. (2006). Fetal ECG extraction using softcomputing technique. Journal of Applied Sciences, 6(2), pp. 251-256.

Vinh, N., & Houle, M. (2010). A set correlation model for partitional clustering. Advances in Knowledge Discovery and Data Mining, pp. 4-15.

Vlahou, A., Schorge, J. O., Gregory, B. W., & Coleman, R. L. (2003). Diagnosis of ovarian cancer using decision tree classification of mass spectral data. Journal of Biomedicine and Biotechnology, 2003, pp. 308-319.

Volk, M., Ripplinger, B., Vintar, S., Buitelaar, P., Raileanu, D., & Sacaleanu, B. (2002). Semantic annotation for concept-based cross-language medical information retrieval. International Journal of Medical Informatics, 67(1-3), pp. 97-112.

Voorhees, E. M. (1994). Query expansion using lexical-semantic relations. Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 61-69.

Walker, D. (2001). Query expansion using thesauri: Previous approaches and possible new directions. University of California, Los Angeles,

Wang, W., & Paliwal, J. (2006). Spectral data compression and analyses techniques to discriminate wheat classes. Trans.ASABE, 49(5), pp. 1607-1612.

Webb, A., & Kagadis, G. C. (2003). Introduction to biomedical imaging. Medical Physics, 30, 2267.

Weissleder, R., Mahmood, U., (2001). Molecular Imaging. Radiology; 219: pp. 316–333

Westerveld, T. (2002) Image retrieval: Content versus context. Content-Based Multimedia Information Access, RIAO 2000 Conference, pp. 276–284.

Wieszczycka, W., & Scharf, W. (2001). Chapter 3, Ultra Son, Proton radiotherapy accelerators World Scientific Pub Co Inc.

William B. Frakes, Ricardo Baeza-Yates. (1992). Information retrieval: Data structures and algorithms. NJ, USA: Prentice-Hall, Inc. Upper Saddle River.

Wirjadi, O., Breuel, T. M., Feiden, W., & Kim, Y. J. (2006). Automated feature selection for the classification of meningioma cell nuclei. Bildverarbeitung Fur Die Medizin,

Woojin Kim MD, Khan M. Siddiqui MD, William Boonn MD and Nabile Safdar MD. (2006). Yottalook. Available: http://www.yottalook.com/index.php. Last accessed 21st March 2012.

Wu, H., Hu, C., & Chen, S. (2010). UESTC at ImageCLEF 2010 medical retrieval task. Working Notes of the ImageCLEF,

Xu, J., & Croft, W. B. (1996). Query expansion using local and global document analysis. Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 4-11.

Xu, S., McCusker, J., & Krauthammer, M. (2008). Yale image finder (YIF): A new search engine for retrieving biomedical images. Bioinformatics, 24(17), 1968.

Yao, T., Mei, T., & Ngo, C. W. (2010). Co-reranking by mutual reinforcement for image search. Proceedings of the ACM International Conference on Image and Video Retrieval, pp. 34-41.

Yu Cao, Henning Muller, Charles E. Kahn, Jr., Ethan Munson. (2011). Multi-modal medical image retrieval. SPIE Medical Imaging,

Yu, H., & Cao, Y. G. (2008). Automatically extracting information needs from ad hoc clinical questions. AMIA Annual Symposium Proceedings, 2008. pp. 96.

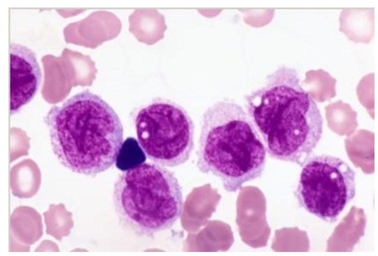
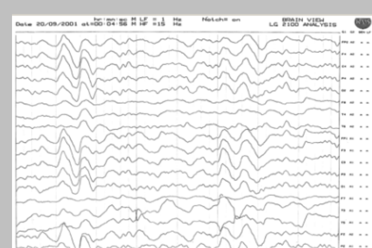
Zhang, D., & Lu, G. (2001). A comparative study on shape retrieval using fourier descriptors with different shape signatures. Proc. of International Conference on Intelligent Multimedia and Distance Education (ICIMADE01), pp. 1-9.

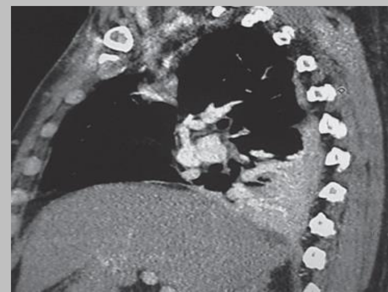
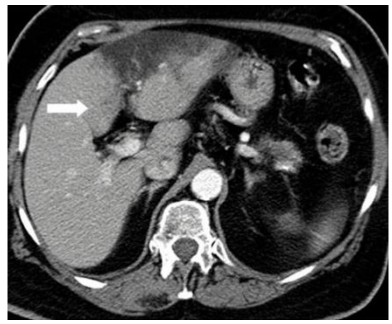
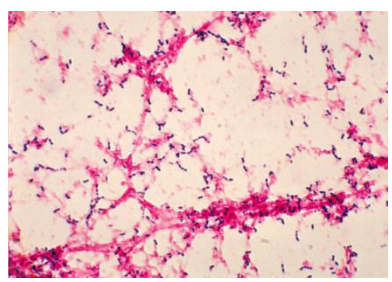
Zheng, X. and Zhou, M.Q. and Wang, X.C., (2008) Interest point based medical image retrieval, Medical Imaging and Informatics

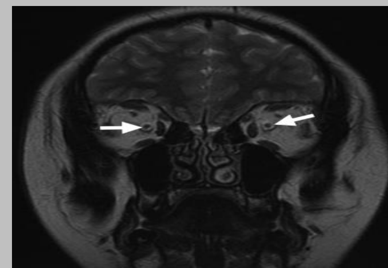
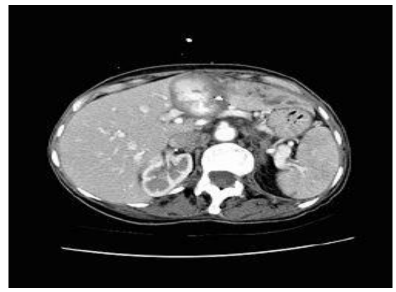
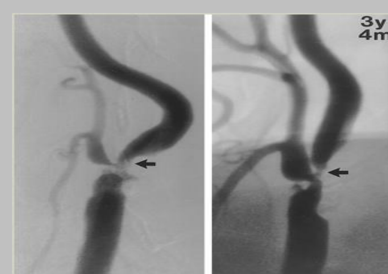
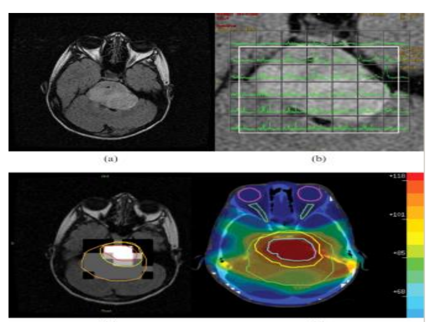
Zhou, X., Depeursinge, A., & Müller, H. (2010). Information fusion for combining visual and textual image retrieval. Proceedings of the 20th International Conference on Recognizing Patterns in Signals, Speech, Images, and Videos, pp. 1590-1593.

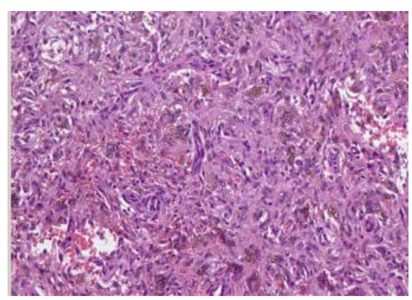
Zhu, Y., De Silva, L., & Ko, C. C. (2002). Using moment invariants and HMM in facial expression recognition. Pattern Recognition Letters, 23(1-3), pp. 83-91.

APPENDIX A: LIST OF 16 AD HOC QUERIES

Q-1	**CT images thoracic aortic dissection**	
Q-2	**A microscopic image of Acute Myeloid Leukemia**	
Q-3	**ECG images**	
Q-4	**X-ray showing congestive heart failure**	

Q-5	**CT images for brachial plexus nerve block**	
Q-6	**CT images containing fatty liver**	
Q-7	**x-ray images of a greenstick fracture**	
Q-8	**Microscopic images streptococcus pneumonia**	

Q-9	**MR images papilledema**	
Q-10	**MR images pericardial effusion**	
Q-11	**All types images with atherosclerosis in blood vessels**	
Q-12	**Radiation therapy treatment plans**	

Q-13	**Images of dermatome**	
Q-14	**Images showing sacral fracture**	
Q-15	**Images coronary arteries**	
Q-16	**Images dermatofibroma**	

APPENDIX B: LIST OF PUBLICATION

Below is the list of publications during this research study is conducted.

Academic Journal

[1] Hizmawati Madzin, Roziati Zainuddin, Nur Sabirin Mohamed (2012) “Analysis of Visual Features in Local Descriptor for Multi-Modality Medical Image “International Arab Journal of Information Technology Volume 11, No. 3, May 2014 (ISI-Cited Publication)

[2] Nurfadhlina Mohd Sharef**,** Hizmawati Madzin (2013) **“**Semantic-based Clinical Records Retrieval via Medical-Context Aware Query Expansion and Comprehensive Ranking” International Journal of Electronic Healthcare (accepted for publication) (Scopus Publication)

[3] Hizmawati Madzin, Roziati Zainuddin (2013) “IF3MIRS: Information Fusion Retrieval System with Knowledge-Assisted Text and Visual Features based on Medical Conceptual Model” Multimedia Tools and Applications (accepted for publication) (ISI-Cited Publication)

Proceedings

[1] Nurfadhlina Mohd Sharef, Hizmawati Madzin (2012) “IMS: An Improved Medical Retrieval Model via Medical-Context Aware Query Expansion and Comprehensive Ranking” Proc of International Conference on Information Retrieval and Knowledge Management, (CAMP’12) Kuala Lumpur, Malaysia

[2] Hizmawati Madzin, Roziati Zainuddin, Nur Sabirin Mohamed (2011) “Multi-modality Medical Images Feature Analysis “Proc. 7th Kuala Lumpur International Conference on Biomedical Engineering (BIOMED 2011) Kuala Lumpur, Malaysia

[3] Hizmawati Madzin, Roziati Zainuddin (2010) “ Shape-based Medical Image Indexing using Local Moment invariant Analysis” Intelligent Systems and Information Technology Symposium (ISITS2010), UPM Serdang, Malaysia

[4] Hizmawati Madzin , Roziati Zainuddin (2009) “ Medical Image Analysis using Local Invariant Features – Interest Points” International Conference of Software Engineering & Computer Systems 2009 (ICSECS’09), pp 303-307, Kuantan Malaysia.

[5] Hizmawati Madzin , Roziati Zainuddin (2009) “ Feature Extraction and Image Matching of 3D Lung Cancer Cell Image” 2009 International Conference of Soft Computing and Pattern Recognition (SoCPAR 2009) pp. 511-515, Malacca,Malaysia

APPENDIX C: MULTI-MODALITY MEDICAL IMAGES

This section illustrates multi-modality medical images from ImageCLEF 2010 medical data collection.

C.1 X-ray

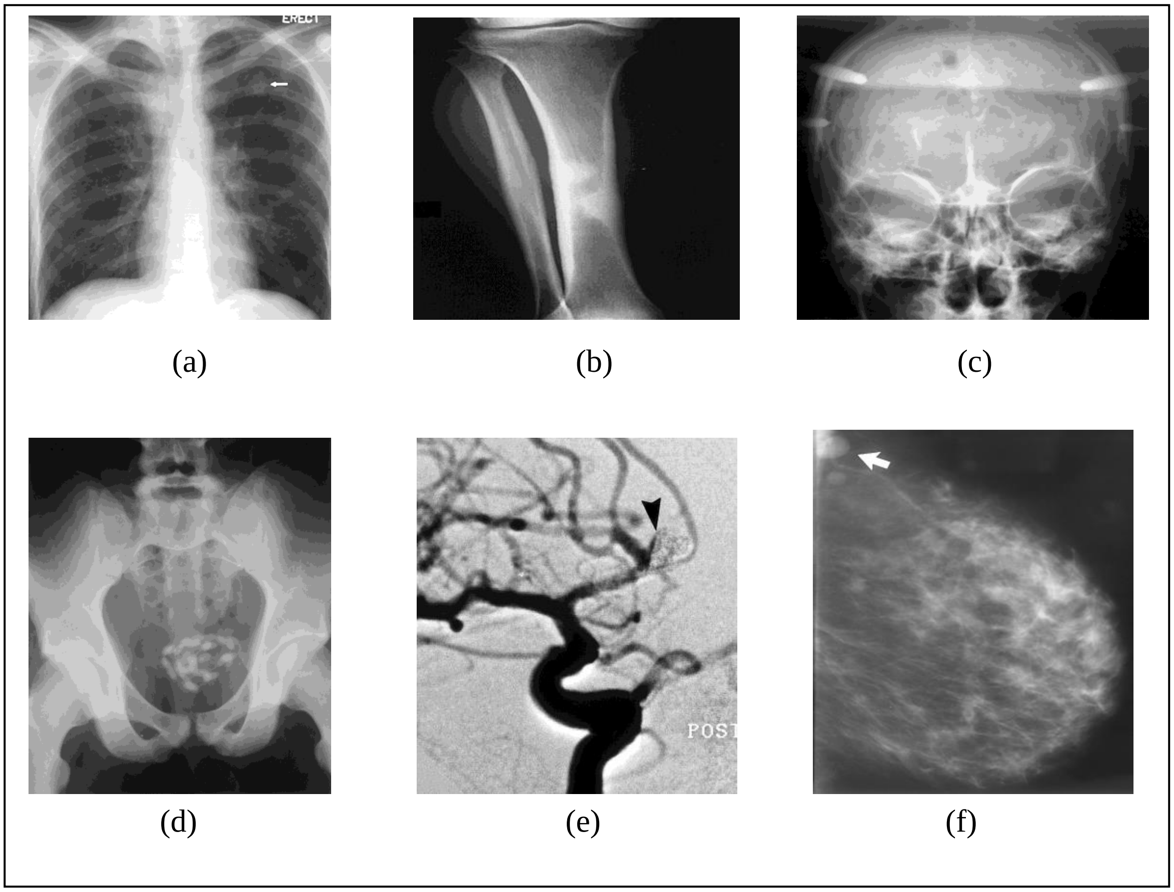
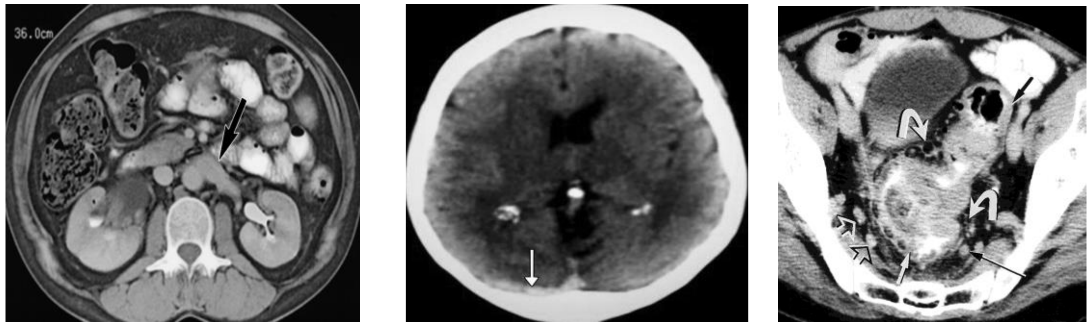


Figure C.1: (a) chest, (b) leg, (c) skull, (d) pelvic, (e) angiography of artery and (f) mammogram of breast

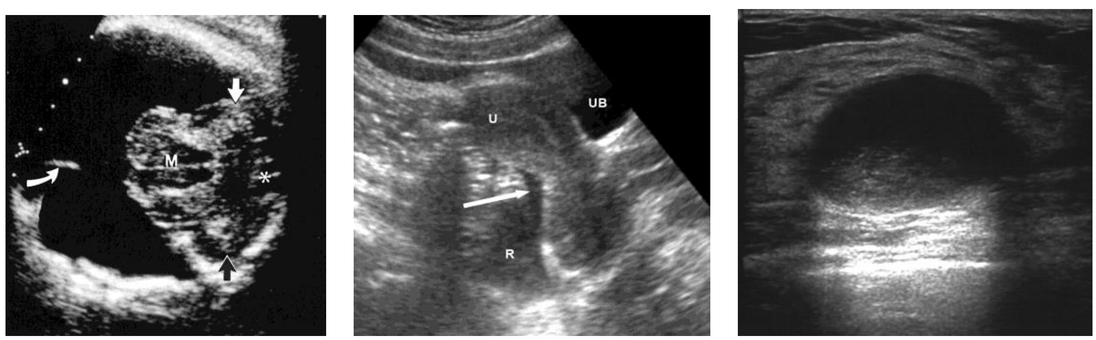
C.2 Computed Tomography (CT)



(a) (b) (c)

Figure C.2: (a) abdomen, (b) brain and (c) pelvis

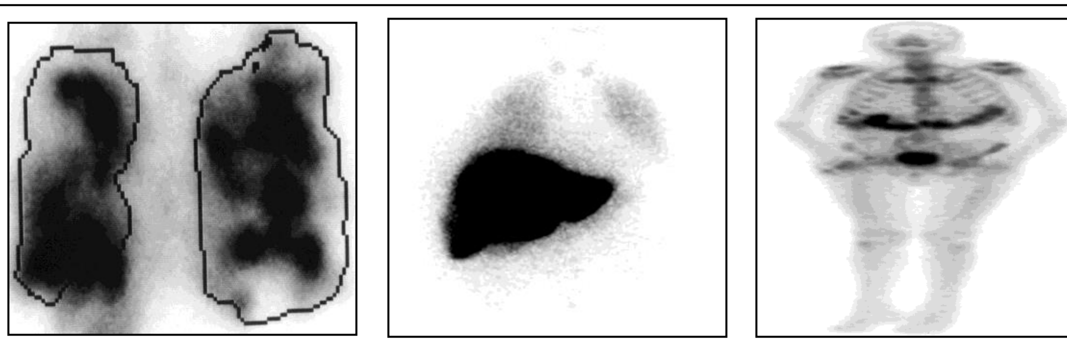
C.3 Ultrasound (US)



(a) (b) (c)

Figure C.3: (a) fetus, (b) kidney and (c) breast

C.4 Nuclear Medicine (NM)



(a) (b) (c)

Figure C.4: (a) lung, (b) liver and (c) whole body

C.5 Positron Emission Tomography (PET)

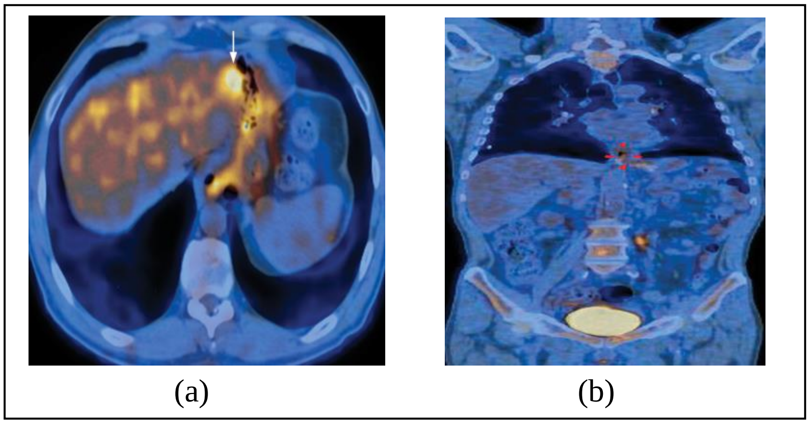


Figure C.5: (a) abdomen and (b) whole body

C.6 Magnetic Resonance Imaging *(*MRI)

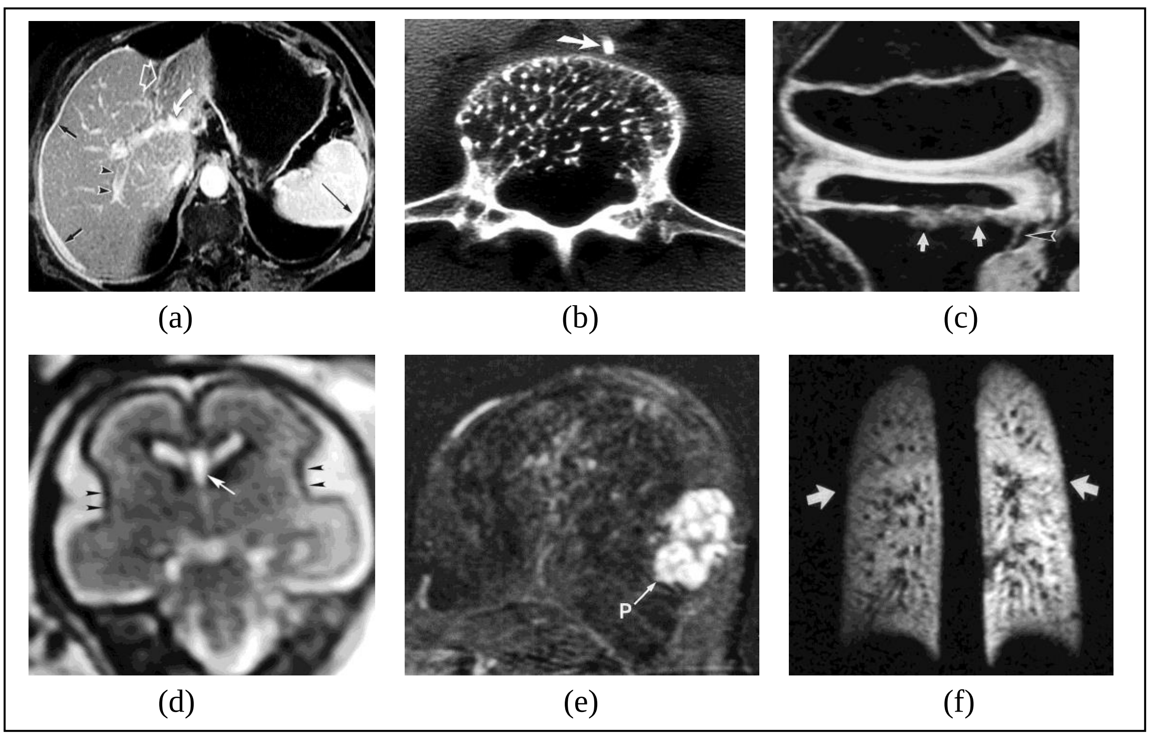
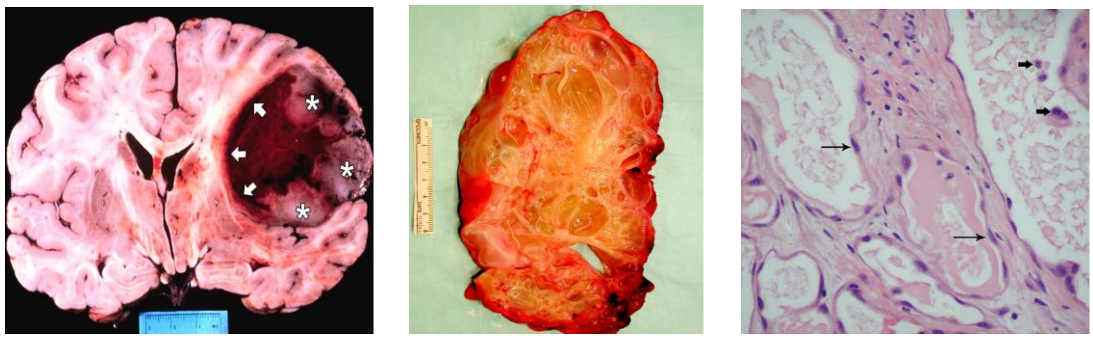


Figure C. 6: (a) abdomen, (b) trabecular bone, (c) joint bone, (d) brain, (e) breast and (f) lung

C.7 Optical Imaging (PX)



(a) (b) (c)

Figure C.7: Gross images of (a) brain and (b) lung. Microscopy image of lung cell in (c)

C.8 Graphic Imaging (GX)

(a) (b) (c)

Figure C.8: Medical graphic data of (a) bar chart, (b) flow diagram and (c) graph

APPENDIX D: ImageCLEF 2010 MEDICAL DOCUMENT DATA STRUCTURE

Example from ImageCLEF 2010 medical text documents used in this research as shown in Figure D.1. Each record of text document represents a medical image which based on <imageLocalName>. The text document is in XML format and used DTD data to represent the information.


```
imageclef2010.xml
<xml version="1.0" encoding="UTF-8"?>
<imageclef>

<record>
<figureID>27977</figureID>
<figureURL>http://radiology.rsnajnls.org/cgi/content/full/210/1/28/F1</figureURL>
<caption> &lt;B&gt;Figure 1. &lt;/B&gt; </caption>
<title>Seymour H. Levitt, MD--President Radiological Society of North America, 1999</title>
<pmid>9885582</pmid>
<articleURL>http://radiology.rsnajnls.org/cgi/content/full/210/1/28</articleURL>
<imageLocalName>27977.jpg</imageLocalName>
</record>

<record>
<figureID>27978</figureID>
<figureURL>http://radiology.rsnajnls.org/cgi/content/full/210/1/36/F1</figureURL>
<caption> &lt;B&gt;Figure 1. &lt;/B&gt; </caption>
<title>William W. Olmsted, MD, New RSNA Education Editor</title>
<pmid></pmid>  <articleURL>http://radiology.rsnajnls.org/cgi/content/full/210/1/36</articleURL>
<imageLocalName>27978.jpg</imageLocalName>
</record>

<record>
<figureID>27979</figureID>
<figureURL>http://radiology.rsnajnls.org/cgi/content/full/210/1/11/F1</figureURL>
<caption> &lt;B&gt;Figure 1. &lt;/B&gt; &lt;B&gt;&lt;/B&gt; Illustration of a neonate at autopsy
whose demise was attributed to "thymic death." The caption drew attention to the "enormous size of
the thymus," which is actually normal in appearance. (Reprinted, with permission, from reference
6.)&lt;P&gt; </caption>
<title>The right place at the wrong time: historical perspective of the relation of the thymus gland and
pediatric radiology</title>
<pmid>9885579</pmid>
<articleURL>http://radiology.rsnajnls.org/cgi/content/full/210/1/11</articleURL>
<imageLocalName>27979.jpg</imageLocalName>
</record>
</imageclef>
```


Figure D.1: Example of medical documents in ImageCLEF 2010 medical task collection

APPENDIX E: MeSH THESAURUS DATA STRUCTURE

Figure E.1 shows a complete descriptor record of *‘Adrenal Cortex Neoplasms’* in MeSH thesaurus. Due to many irrelevant information in this thesaurus, we created MeSH-indexer as described in section 4.3.2 for effectiveness in accessing medical terms and efficiency of reducing time processing.


```
<DescriptorRecord DescriptorClass = "1">
  <DescriptorUI>D000306</DescriptorUI>
  <DescriptorName>
   <String>Adrenal Cortex Neoplasms</String>
  </DescriptorName>
  <DateCreated>
   <Year>1999</Year>
   <Month>01</Month>
   <Day>01</Day>
  </DateCreated>
  <DateRevised>
   <Year>2004</Year>
   <Month>07</Month>
   <Day>07</Day>
  </DateRevised>
  <DateEstablished>
   <Year>1979</Year>
   <Month>01</Month>
   <Day>01</Day>
  </DateEstablished>
  <ActiveMeSHYearList>
   <Year>2005</Year>
   <Year>2006</Year>
   <Year>2007</Year>
   <Year>2008</Year>
   <Year>2009</Year>
   <Year>2010</Year>
  </ActiveMeSHYearList>
  <AllowableQualifiersList>
   <AllowableQualifier>
    <QualifierReferredTo>
     <QualifierUI>Q000097</QualifierUI>
      <QualifierName>
      <String>blood</String>
      </QualifierName>
    </QualifierReferredTo>
    <Abbreviation>BL</Abbreviation>
   </AllowableQualifier>
  </AllowableQualifiersList>
  <Annotation>coord IM with histol type of neopl (IM)
  </Annotation>
<HistoryNote>79(75)63-67; was see under ADRENAL GLAND NEOPLASMS 1975-78
  </HistoryNote>
  <OnlineNote>search ADRENAL GLAND NEOPLASMS 1966-74
  </OnlineNote>
```

```
<PreviousIndexingList>
  <PreviousIndexing>Adrenal Gland Neoplasms (1966-1974)</PreviousIndexing>
 </PreviousIndexingList>
 <TreeNumberList>
  <TreeNumber>C04.588.322.078.265</TreeNumber>
  <TreeNumber>C19.053.098.265</TreeNumber>
  <TreeNumber>C19.053.347.500</TreeNumber>
  <TreeNumber>C19.344.078.265</TreeNumber>
 </TreeNumberList>
 <RecordOriginatorsList>
  <RecordOriginator>NLM</RecordOriginator>
  <RecordMaintainer>lkt</RecordMaintainer>
  <RecordAuthorizer>sjn</RecordAuthorizer>
 </RecordOriginatorsList>
 <ConceptList>
  <Concept PreferredConceptYN="Y">
   <ConceptUI>M0000481</ConceptUI>
   <ConceptName>
    <String>Adrenal Cortex Neoplasms</String>
   </ConceptName>
   <ConceptUMLSUI>C0001618</ConceptUMLSUI>
   <ScopeNote>Tumors or cancers of the ADRENAL CORTEX.
   </ScopeNote>
   <SemanticTypeList>
    <SemanticType>
     <SemanticTypeUI>T191</SemanticTypeUI>
     <SemanticTypeName>Neoplastic Process</SemanticTypeName>
    </SemanticType>
   </SemanticTypeList>
   <ConceptRelationList>
    <ConceptRelation RelationName="NRW">
    <Concept1UI>M0000481</Concept1UI>
    <Concept2UI>M0331834</Concept2UI>
    </ConceptRelation>
   </ConceptRelationList>
   <TermList>
    <Term ConceptPreferredTermYN="Y" IsPermutedTermYN="N" LexicalTag="NON"
PrintFlagYN="Y" RecordPreferredTermYN="Y">
     <TermUI>T001004</TermUI>
     <String>Adrenal Cortex Neoplasms</String>
     <DateCreated>
      <Year>1999</Year>
      <Month>01</Month>
      <Day>01</Day>
     </DateCreated>
     <EntryVersion>ADRENAL CORTEX NEOPL</EntryVersion>
     <ThesaurusIDlist>
      <ThesaurusID>NLM (1975)</ThesaurusID>
     </ThesaurusIDlist>
    </Term>
   </TermList>
  </Concept></ConceptList>

</DescriptorRecord>
```


Figure E.1: An example of descriptor record in MeSH thesaurus