PhD Thesis

Automatic Annotation, Classification and Retrieval

of Traumatic Brain Injury CT Images

Gong Tianxia

Supervisor: Prof. Tan Chew Lim

Department of Computer Science

School of Computing

National University of Singapore

December, 2011

Abstract

Due to the advances in medical imaging technology and wider adoption of electronic medical record systems in recent years, medical imaging has become a major tool in clinical trials and a huge amount of medical images are proliferated in hospitals and medical institutions every day. Current research works mainly focus on modality/anatomy classification, or simple abnormality detection. However, the needs to efficiently retrieve the images by pathology classes are great. The lack of large training data makes it difficult for pathology based image classification. To solve problems, we propose two approaches to use both the medical images and associated radiology reports to automatically generate a large training corpus and classify brain CT image into different pathological classes. In the first approach, we extract the pathology terms from the text and annotate the images associated with the text with the extracted pathology terms. The resulting annotated images are used as training data set. We use probabilistic models to derive the correlations between the hematoma regions and the annotations. We also propose a semantic similarity language model to use the intra-annotation correlation to enhance the performance. In testing, we use both the trained probabilistic model

1

and language model to automatically assign pathological annotations to the new cases. In the second approach, we use deeper semantics from both images and text and map the hematoma regions in the images and pathology terms from the text explicitly by extracting and matching anatomical information from both resource. We explore hematoma visual features in both 3D and 2D and classify the images into different classes of pathological changes, so that the images can be searched and retrieved by pathological annotation.

Acknowledgements

I would like to thank my supervisor, Prof. Tan Chew Lim, who has stimulated me to be interested in this research area and given me invaluable advice on my research topic. In addition to academic research, I felt indebted to him in many other aspects. I would not have progressed so far without him inspiring me all the time.

I also want to thank my project group leader, Dr. Li Shimiao, who has helped me a lot on many aspects of my research work.

I thank Sun Jun and Chen Bin for their kind help on machine translation and other natural language processing work, and Wang Jie and Liu Ruizhe for their help on image processing.

Finally, I wish to thank my family members for their support over the years. I want to thank my husband Liu Keyao who has supported me with all heart unconditionally, and my parents and parents-in-law for their understanding and encouragement. Last but not least, I want to thank my little baby girl Liu Tingxuan, who has given so much joy and motivation in my PhD studies.

Contents

4

List of Figures

9

11

List of Tables

Chapter 1

Introduction

1.1   Background

Due to the advances in medical imaging technology and wider adoption of electronic medical record systems in recent years, medical imaging has become a major tool in clinical trials. As a result, a huge amount of medical images are proliferated in hospitals and medical institutions every day. A medical examination such as X-Ray, CT or MRI usually consists of one image or a series of images and a radiology report. For example, a CT image series of a traumatic brain injury (TBI) case is shown in Figure 1.1.

Medical findings in medical reports associated with the images mainly refer to pathological changes, i.e. disorders, diseases, and other abnormalities. For example, "hematoma" and "midline shift" are medical findings in the example report shown in Figure 1.2. Apart from the findings, radiologists also note down
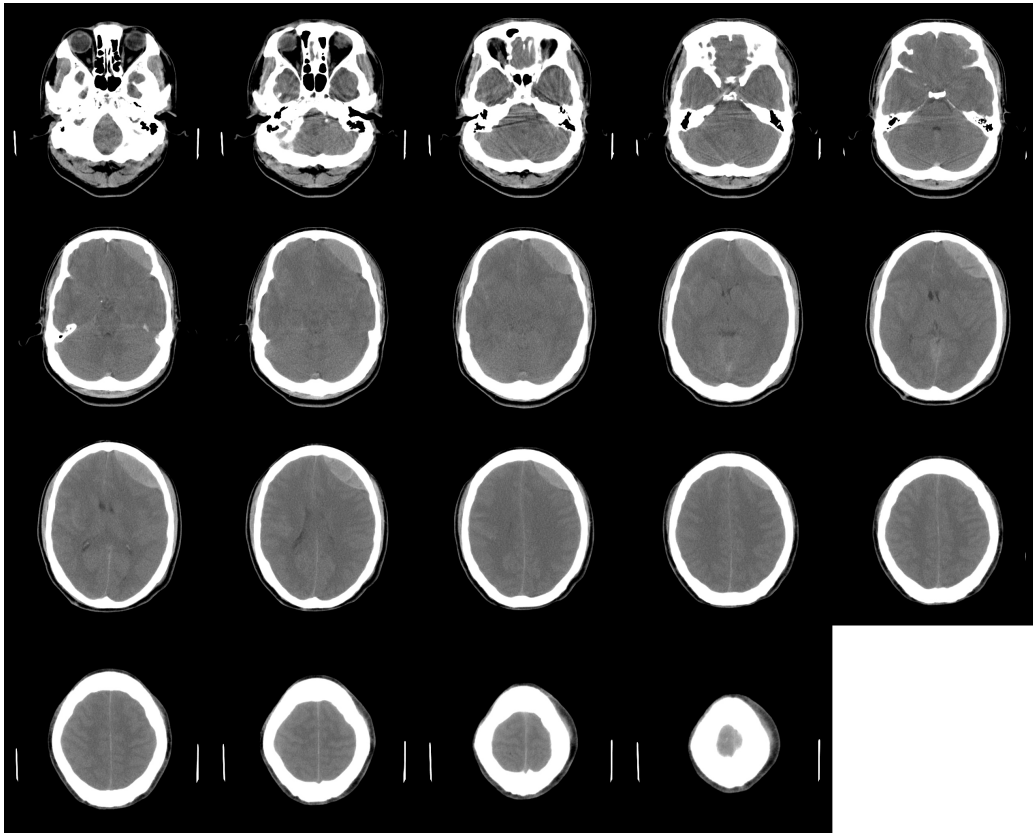
Figure 1.1: The image series of a traumatic brain injury case

more specific details of the findings in the reports. They can be considered as attributes or modifiers of the findings, which include anatomical location (body part), amount or size, direction, probability (how likely the radiologist think the observation is indeed abnormality of the brain), seriousness, and etc..

> Unenhanced axial scans of the brain were obtained. There is large extradural hematoma in the left frontal lobe. This is compressing the underlying brain and distorting the left lateral ventricle. ...

Figure 1.2: The radiology report associated with the CT image series

For traumatic brain injury, CT is a vital tool for the assessment and remains the investigation of choice even following the advent of MRI, due both to the ease of monitoring of injured patients and the better demonstration of fresh bleeding and bony injury [32]. A blow to the skull results in compression injury to the adjacent brain (coup) and stretching on the opposite side (contrecoup). This may result in contusion, shearing injuries and rupture of intra-axial or extra-axial vessels, leading to hemorrhage. There are several types of hemorrhages ("hematoma" is often used interchangeably with "hemorrhage"): extradural hematoma (EDH), subdural hematoma (SDH), intracerebral hemorrhage (ICH), subarachnoid hemorrhage (SAH), and intraventricular hematoma (IVH). In this thesis, we focus on the analysis of the CT images with the presence of various types of the hemorrhages; therefore, we give a brief introduction to these types of hemorrhages according to [32].

An EDH occurs when there is a rupture of a blood vessel, usually an artery, which then bleeds into the space between the "dura mater"" and the skull. The affected vessels are often torn by skull fractures. The expanding hematoma strips the dura from the skull; this attachment is quite strong such that the hematoma is confined, giving rise to its characteristic biconvex shape, with a well defined margin.

An SDH arises between the dura and arachnoid, often from ruptured veins crossing this potential space. The space enlarges as the brain atrophies and so subdural hematomas are more common in the elderly. The blood is of high attenuation, but may spread more widely in the subdural space, with a crescentic

16

appearance and a more irregular inner margin.

An ICH occurs due to stretching and shearing injury, often due to impaction of the brain against the skull on the side opposite to the injury. Thus they may be seen
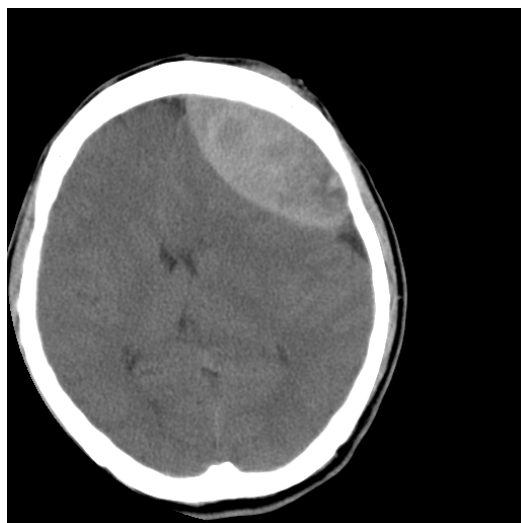
Figure 1.3: A CT image with EDH
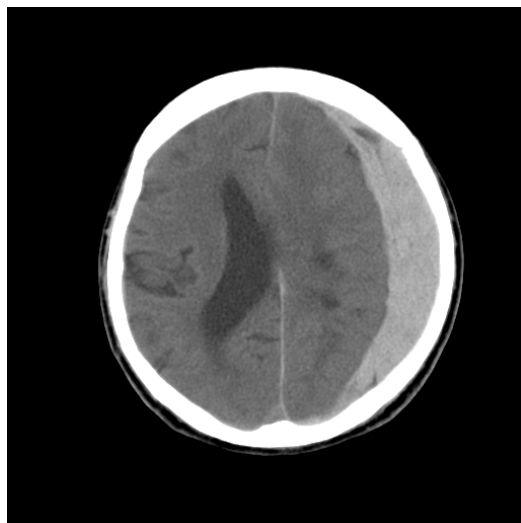
Figure 1.4: A CT image with SDH

directly opposite the impact site, subcutaneous hematoma, fracture, or extradural hematoma (contre coup injury). The inferior frontal lobes and anterior temporal lobes are common sites after a blow to the back of the head. Multiple contusions may be present throughout the cerebral hemispheres. They are often very small and visible at the grey/white matter interface. They are due to a shearing injury with rupture of small intracerebral vessels, and in a comatose patient with no other obvious cause they imply a severe diffuse brain injury with a poor prognosis.
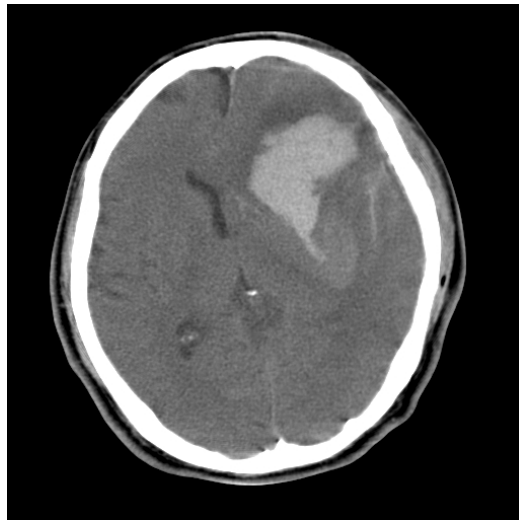
Figure 1.5: A CT image with ICH

SAH may occur alone or in association with other intracerebral or extracerebral hematomas. Increased attenuation is seen in the CSF spaces, over the cerebral hemispheres (look closely at the Sylvian fissure), in the basal cisterns or in the ventricular system. SAH may be complicated by hydrocephalus.

An IVH is a bleeding into the brain's ventricular system, where the cerebrospinal fluid is produced and circulates through towards the subarachnoid space.

It can result from physical trauma or from hemorrhaging in stroke. The injury requires a great deal of force to cause in adults. Thus the hemorrhage usually does not occur without extensive associated damage, and so the outcome is rarely good. Prognosis is also dismal when IVH results from intracerebral hemorrhage related to high blood pressure and is even worse when hydrocephalus follows. It can result in dangerous increases in intracranial pressure and can cause potentially fatal brain herniation

1.2   Current research problems

Most medical images are in the standard DICOM (Digital Imaging and Communications in Medicine) format, and the display and retrieval of CT scan images are mostly via PACS (Picture Archives and Communication System) hardware

Figure 1.6: A CT image with SAH

[57]. However with such standards and hardware, the medical images currently can only be retrieved using patient names or identity card numbers. To retrieve an image pertaining to a particular anomaly without the patient name is literally like looking for a needle in a haystack. In the domain of CT brain images, very often doctors already overloaded with day-to-day medical consultation simply could not remember patients names when they need to refer to cases of certain type of brain trauma seen before and as such valuable information are lost in the sea of raw image pixels.

In addition to medical images, free text medical reports are also produced in large amount daily. These medical texts include the patient's medical history, medical encounters, orders, progress notes, test results, etc. Although these textual data contain valuable information, most are just archived and not referred to again. These are valuable data that are not used to full advantage. A similar

Figure 1.7: A CT image with IVH

situation occurs in the field of radiology. As the reports are in free text format and usually unprocessed, there exists a great barrier between the radiology reports and the medical professionals (radiologists, physicians, and researchers), making it difficult for them to retrieve and use the information and knowledge from the reports.

Text-based image retrieval is friendly to users as only text query is required; it can retrieve images fast as images are indexed by text. However, it can only index and retrieve images with accompanying text. For medical images, those without associated textual information cannot be indexed or retrieved. Content-based medical image retrieval provides an alternative to text-based retrieval by indexing images with visual features so that medical images without accompanying text can still be indexed and retrieved. However, content-based image retrieval poses a limitation on the query format–the query must be an image example. Moreover, it suffers from the semantic gap problem as the visual features are mostly low level and are not directly linked to the understanding of the medical images. Auto-annotation based medical image retrieval seems to have the advantages of both text-based and content-based image retrieval by automatically annotating images with their semantic content and offering users the ease of search images based on the textual annotations. Hence, automatic medical image annotation/classification and annotation-based medical image retrieval have gained popularity in recent years. Numerous tasks have been proposed in CLEF medical image annotation tracks [119]. Most research works focused on automatically generating annotations of acquisition modality (CT, X-ray, MR, etc.), body orientation, body region,

and biological system. Some works also focus on the detection of abnormalities in medical images.

However, while most research works focus on the analysis of the images, very few works put effort into analysis of the radiology report associated with the medical images and the correspondence between the descriptions in the report and the region of interest in the images. For image classification/annotation, most works classify the medical images according to their modality, anatomical body part, or the presence of abnormality, whereas only a few works classify the images according to their pathology classes. As it is often the case that a doctor wants to retrieve all images pertaining to one pathology class, current work cannot satisfy the doctor's need. Indexing and retrieving medical images by their pathological annotations can help to satisfy this need; however, a large labeled training data set is needed for automatic medical image annotation/classification, as manual labeling requires domain expertise and is thus expensive and slow.

In summary, current problems in this research area include:

- There are large amount of images, but mostly are difficult to retrieve.

- Current research works mainly focus on modality/anatomy classification, or simple abnormality detection.

- The lack of large training data makes it difficult for pathology based image classification

1.3   Our solutions and contributions

In this thesis, we propose some solutions to the problems stated above. We process and analyze the medical images and reports to extract deeper semantics and useful information. We provide several modes to suit user's needs to search and retrieval medical images accurately, fast, and conveniently.

Firstly, we apply natural language processing methods and use domain knowledge resource to the free text medical report to extract useful information such as medical findings and the specific descriptors of pathological changes. We use the extracted information to index the reports as well as the images the reports are associated with. In this way, the users can search and retrieve medical images that have accompanying reports by typing text queries into the system, and the system will return medical images that fulfill the text queries. In addition to text-based indexing and retrieval of medical images, another use of the information extracted from the free text medical reports is to use medical findings and their specific descriptors such as anatomical locations to help with image processing in region of interest recognition, classification, and annotation.

Secondly, to cater the need of search and retrieval of visually similar medical images, we provide a content-based mode for medical image retrieval. We process the images, segment the region of interest and convert it into a binary visual feature vector which is used to index the image. When a user submit an image query, we process it in the same way as we process the images in the database. We partitioned the brain image into bins and obtain a binary feature vector of the

query image and compare it with the binary features vectors representing other images in our database, then we return the images according to their similarity to the query binary feature vector. For TBI cases, we are the first to use such method to preserve both shape and location of the ROI in the feature representation. In this way, the users are able to find visually similar images to their query images. This function of our system comes handy for users who are not equipped with much domain knowledge and could not form proper text queries. It can also serve as a good teaching tool for junior doctors.

On top of text-based and content-based image retrieval, we also develop novel frameworks that automatically classify the medical images into pathology change categories and provide annotation-based image retrieval to cater user's needs. While most research works focus on the analysis of the images or text separately, very few works put effort into the analysis of the medical text (e.g. radiology reports) associated with the medical images and the correspondence between the descriptions in the text and the respective regions or findings in the images. We propose two approaches to utilize both the medical images and text to generate a training corpus for pathology based automatic medical image annotation/classification. In the first approach, we extract the pathology terms from the text and annotate the images associated with the text with the extracted pathology terms. The resulting annotated images are used as training data set. We use probabilistic models to derive the correlations between the regions of interest in the images and the annotations since the annotations are mapped to the whole image, not the specific regions. Then we use the trained models to automatically assign

pathological annotations to the images without accompanying text so that these images can be retrieved as well. In the second approach, we explore deeper semantics from both images and text and map the ROIs in the images and pathology terms from the text explicitly by extracting anatomical information from both resource. From the image series, we segment the regions of interest, i.e. the area of pathological changes, and obtain their anatomical location information by registering the image series to a referenced brain atlas. We extract the anatomical terms in addition to pathological terms from the textual report associated with the images, match and label the ROIs and pathological class in images and text, and thus create a region-based labeled data set for training. We explored the features for hematomas in both 3D and 2D so that we could classify the images according to the pathological changes.

In summary, our contributions to this research area include:

- We use both text and images to automatically generate a large training corpus.

- We propose two novel frameworks to classify medical images according to pathological changes.

- We develop text-based, content-based, and annotation-based image retrieval methods for brain CT images.

1.4   Organization of the thesis

We organize the thesis as follows. We review the research works related to our research problems in Chapter 2. Chapter 3 describes our methods and experimental result for free text radiology report processing. Chapter 4 describes our methods and experimental results for medical image processing, binary feature vector generation, and content based medical image retrieval. Then we describe our methods for pathology based automatic medical image annotation, classification, and annotation based medical image retrieval in Chapters 5 and 6. Chapter 5 discusses a probabilistic model approach for unsupervised medical image annotation. Chapter 6 describes a supervised approach for region-based medical image classification. Finally, we discuss possible future research directions and conclude this thesis in Chapter 7.

Chapter 2

Literature review

2.1 Information Extraction from Medical Text

In the medical domain, with the advances in medical technology and wider adoption of electronic medical record systems, medical text data have proliferated at rapid speed and in huge amount in hospitals and other health institutions daily. However, the narrative form of these medical texts is difficult for searching, retrieval, or statistical analysis. Information extraction (IE) from these raw free text data, as a sub topic of information retrieval [81, 122], is needed in order to use these valuable textual data effectively and efficiently.

The goal of IE is to automatically extract structured information from unstructured and/or semi-structured documents. [114] summarized research works in IE in medical domain prior to 1995. [24] and [128] reviewed systems of IE for biomedical text. [88] surveyed recent research works on IE from textual

documents in the electronic health record with more focus on clinical data. The main tasks of IE for medical documents include Natural Language Processing (NLP), Named Entity Recognition (NER) and text mining. An IE system usually is comprised up of a combination of the following components [54]: tokenizer, document decomposer, part-of-speech (POS) tagger, morphological analyzer, shallow/deep parser, gazetteer, named entity recognizer. Some systems have higher level components like discourse module, template extractor, and template combiner. Main approaches to IE in medical domain include pattern matching, shallow/full syntactic parsing, syntactic and semantic parsing approaches. Most systems have a pre-processing step, which could include spelling checking, word sense disambiguation (WSD), POS tagging, and parsing. We will review some mostly cited systems in the following sections.

2.1.1 LSP-MLP

The Linguistic String Project-Medical Language Processor (LSP-MLP) [105] is the earliest NLP system for medical information extraction. LSP-MLP is a large scale project focusing on the extraction and summarization of signs, symptoms, drug information, and identification of possible medication side effects. The program flow of the system consists six steps:

1. Syntactic parsing

2. Semantic selection

3. Transformation

4. Regularization

5. Information formatting

6. Normalization

The parsing module structures the sentences of a medical document and represents the dependencies by means of parse trees. The parser can handle conjunctions. The semantic selection module uses the co-occurrence patterns to improve the parsing tree by resolving cases of structural ambiguity. Also, semantic characterization of parts of the parse tree is done on this level. The transformation module fills in gaps due to conjunction ellipses, reduces all sentence types to the affirmative type, completes relative sentences and regroups verbal splits. The regularization module transforms the semantically augmented parse tree into a canonical tree consisting of elementary sentences that correspond to the basic sub-language sentence types. The inflected forms are replaced by their canonical form and the semantic host and modifiers are identified. The formatting module maps the words of the elementary sentences into the appropriate fields of a format tree and constructs a binary connective- format tree for each sentence with the connectives as parent, and the phrases on which it operates as left and right children. The normalization module recovers implicit knowledge when possible and maps the format trees into the relational database structure.

The information extracted by the LSP-MLP system is stored in a relational database and can be retrieved by SQL queries. Information retrieval task is performed to evaluation the information extraction performance. The system

achieves 98.6% in precision and 92.5% in recall on test data set. LSP-MLP has inspired many research works in this area in later years.

2.1.2 MedLEE: Medical Language Extraction and Encoding System

The Columbia University of New York (together with the Columbia Presbyterian Medical Center) has developed an NLP system MedLEE (MEDical Language Extraction and Encoding System) [39, 114]. MedLEE identifies clinical information in narrative reports and transforms the textual information into a structured and conceptual representation. The main goal is to represent the knowledge of chest X-ray radiology reports, store it in a database and allow physicians to query the knowledge base by means of controlled vocabulary. MedLEE is mainly semantically driven and the semantic grammar consist of 350 DCG rules, specifying well-defined semantic patterns, the interpretations and the target structures into which they should be mapped. MedLEE system processes medical texts in three phases (also as shown in Figure 2.1):

1. Parsing

2. Phrase regularization

3. Encoding

In parsing phrase, the system determines the structure of the text and generates the preliminary structured output form for the clinical information. The parser us-

es a (semantic) grammar and a lexicon. Then in phrase-regularization phase, the system combines the structured outputs of noncontiguous expressions and standardizes them so that they correspond to the appropriate regular form, using a mapping knowledge base (consisting of the structural output forms of multi-word phrases that can be decomposed). Finally in the encoding phase, it maps the standard forms into unique concepts associated with the controlled vocabulary using a synonym knowledge base that consists of standard forms and their corresponding concepts in the controlled vocabulary, the Medical Entities Dictionary [23]. MED was developed at Columbia Presbyterian Medical Center (CPMC) was first served as a knowledge base of medical concepts that consist of taxonomic relations in addition to other relevant semantic relations. At later stage, MedLEE also experimented to use UMLS [61] as the knowledge base and had different evalu-
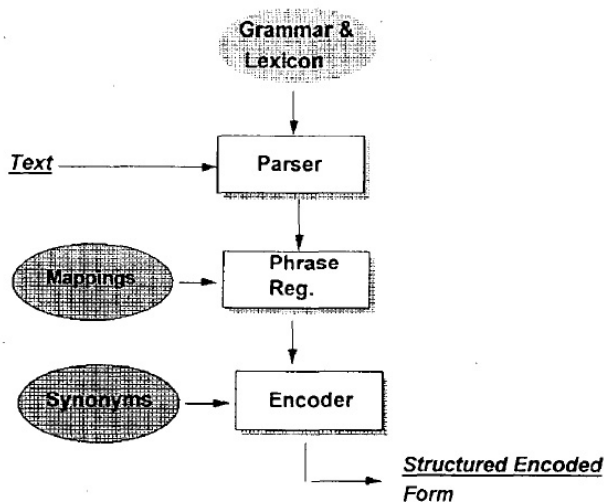
Figure 2.1: The three phases of MedLEE

31

ation results as in [40, 41, 87]. MedLEE had been improved and more features had been added over the ten year, and it remains one of the most cited and popular methods to process text for radiology reports.

2.1.3   RADA: RADiology Analysis Tool

RADA, the Radiology analysis tool as described in [65] provides a method to index findings and associated information described in free text thoracic radiology reports. The system extracts mass lesion and lymph node findings, and links specific information associated with the findings such as size and location. Each glossary entry for RADA is represented by a concept, the smallest fragment of knowledge defined by RADA. A concept encodes both semantic and syntactic knowledge. RADA's glossaries originate from two main sources, the Unified Medical Language Sources (UMLS) [61] and a specialized thoracic glossary. The specialized glossary augments the data found in the UMLS thus providing additional information necessary for the system.

Each concept in RADA system has three attributes as shown in Figure 2.2. The semantic code defines the semantic class to which the concept belongs. Likewise, the syntactic code defines which syntactic class to which the concept belongs. The text string defines the word or phrase to which the concept corresponds. The lexical analyzer uses the text string to match the concept with the text. To provide a human readable form of the concept, the text string is maintained throughout the rest of the system.

Lexical knowledge is encoded in knowledge hierarchies similar to type ab-

straction hierarchies [22]. Type abstraction hierarchies are multilevel knowledge structures that emphasize the abstract representation of information. The meaning of a word or phrase is defined by a hierarchy of related concepts. A concept's semantic code encodes its position in the hierarchy. Different hierarchies exist for different classes of concepts. For example, anatomy concepts form one hierarchy and finding concepts form another. For each concept class they developed a hierarchy of terms and meanings as shown in Figure 2.3. The entries of the glossary are grouped by their meanings.

RADA uses entities to structure the details of extracted radiology findings and anatomy. Entities structure knowledge through a well-defined set of attributes. For example, an entity encoding a radiology finding will have attributes describing the size, location and architecture of the finding. Findings found in a report are stored in instances of the finding entity. RADA creates instances of the entities during
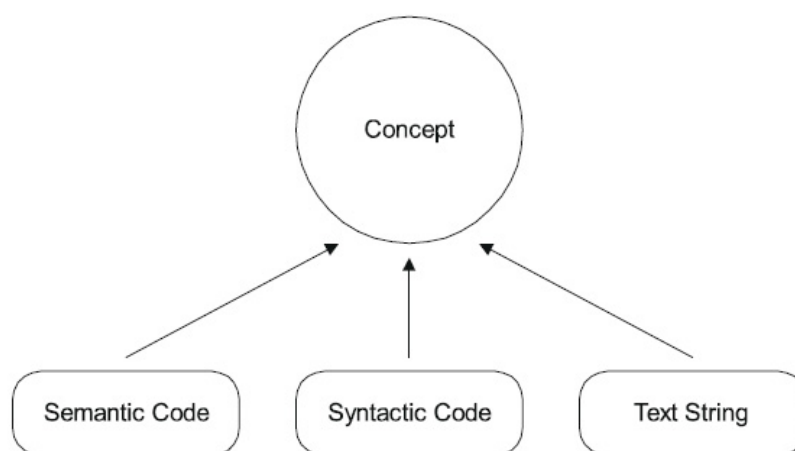
Figure 2.2: The concept representation in RADA

finding analysis.

As shown in the figure 2.4, RADA consists of four parts: Lexical Analyzer, Finding Analyzer, Joiner and Reference Resolver.

The Lexical Analyzer first decomposes sentences into words and phrases. It matches words and groups of words to a specialized glossary of terms, and decomposes the sentence into glossary entries. The Finding Analyzer scans the sentence for articles (the, a, an) and pronouns. When an article is found, a finding entity is created and the sentence is parsed for phrases describing anatomy or findings. Parsing experts process fragments and recognize phrases that can be combined. Different parsing experts process the sentence, insuring that the sentence fragment matches one of several known forms. The Joiner uses several semantic/syntactic parsers to link concepts into the slots of the finding entity. Each parser is a context
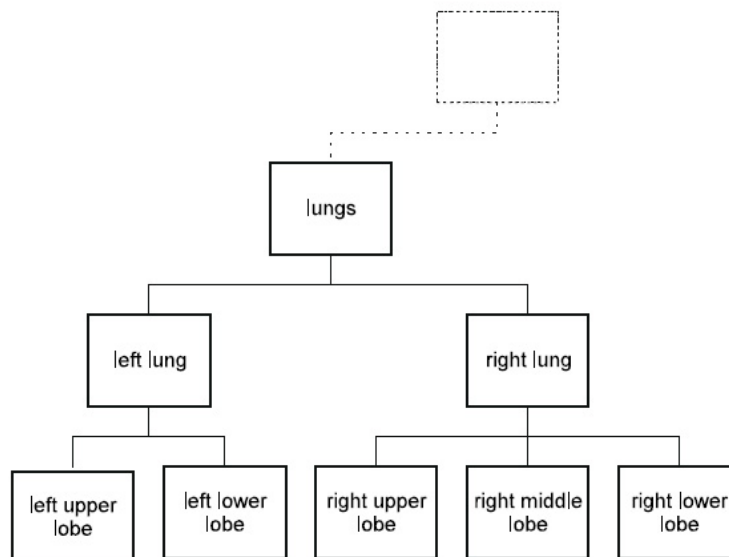
Figure 2.3: The type abstraction hierarchy in RADA

34

free grammar. Joiner iteratively parses the sentence, until no more changes are made to the sentence or any extracted findings. Each grammar parses the sentence in turn, adding concepts to any findings in the sentence and compressing the sentence into a simpler representation. This joiner phase also removes unnecessary information from the sentence and insures that negative findings are accurately modeled. Concepts that represent a finding are combined into an entity. Within the sentence, the entity replaces the concepts it supersedes. Removing the extraneous concepts simplifies the structure of the sentence.
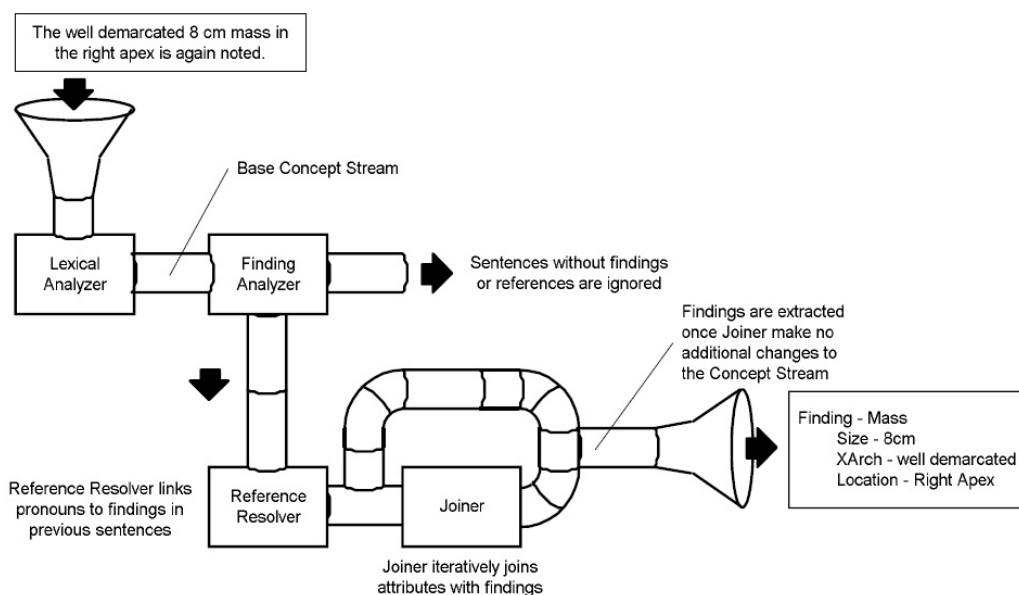
Figure 2.4: The general architecture of RADA
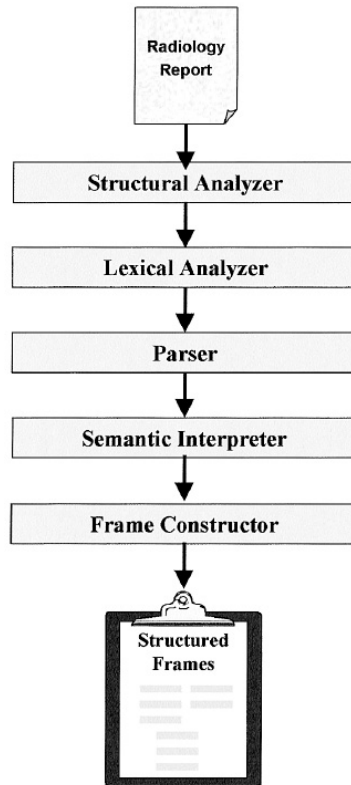
Figure 2.5: The general architecture of Taira et al's statistical NLP system

2.1.4 Statistical Natural Language Processor for Medical Reports

Taira et al [117, 116] developed statistical natural language processor for radiology reports since most tasks in medical information extraction are classification problems. They focus on the specific sub-problems of sentence parsing and semantic interpretation [101]. The statistical NLP system consists five components as shown in Figure 2.5: structural analyzer, lexical analyzer, parser, semantic interpreter and frame constructor.

The Structural Analyzer isolates sections of medical reports (e.g., "Procedure Description", "History", "Findings", "Impressions") and individual sentences within sections. It is implemented based on a conversion from a rule-based system to one that uses a maximum entropy classifier. The Lexical Analyzer looks up semantic and syntactic features of words in a medical lexicon [65], normalizes dates and numerical expressions, and tokenizes punctuation. The Parser creates a dependency diagram between words in an input sentence by adding arcs that indicate a modifier relationship between pairs of words. An arc from word A to word B indicates A modifies B. The mechanism of parsing is conceptualized as a dynamics problem similar to how atoms aggregate to form complex molecules. Words initially have no dependencies with other words. They each exist in a free state. As the parsing step proceeds, each word attempts to configure itself into a more favorable steady state of existence. The final state of the parse reflects the configuration of the words that minimizes the overall energy of the system. Words are modeled as active entities characterized by their signal processing behavior. This includes its emission spectrum, its absorption spectrum, and its response function to resonance conditions. The Semantic Interpreter interprets the links of the parser's dependency diagram and outputs a set of logical relations that form a semantic network for the sentence. The dependency graph that the parser produces has unlabeled arcs between words to show modifier relations. The semantic interpreter applies rules based on semantic features of the words and the direction of the arc between them in the surface structure parse to translate these arcs into the logical relations. Then it bundles logical relations together into output frames that list

attributes of a finding, of a therapeutic or diagnostic procedure, or of an anatomic structure. The Discourse Processor determines whether a finding from a sentence is new or a referent to a finding from previous sentences.

2.1.5 Challenges

Compared to text in general domain, medical text is usually more difficult for NLP and information extraction. The medical texts generated from clinical practices are often ungrammatical and contain many shorthand writing such as abbreviations, acronyms, and telegraphic phrases, which need to be resolved prior to NLP. Misspelling and and spelling variations are also common in free text medical reports and need to be addressed. Ambiguous words, phrases, and sentence structure pose another challenge for medical text processing as well. Medical texts contain a lot of negative expressions and need to be processed so that correct information can be extracted.

2.2 Content based medical image retrieval

As digital images are produced in ever larger amount in the medical field, the need for content-based access to medical images is also on the rapid rise. Content-based image retrieval (CBIR) has become a hot research topic in recent decade and CBIR systems have been developed in different domains. [103, 104] gave reviews on current techniques, promising directions, and open issues for CBIR. [112] gave a very detailed review of CBIR systems before 2000. Medical images,
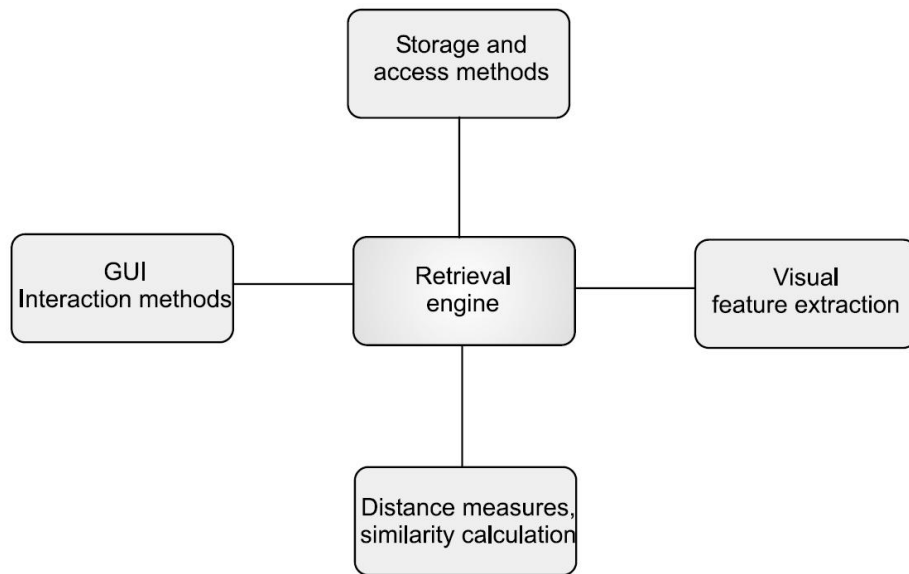
Figure 2.6: The general architecture of a typical CBIR system

especially radiology images, are of the greatest need of such system. Compared to the traditional clinical management systems which access medical images using meta-data index, CBIR system allows medical professionals to access medical image data in an easy and direct way. [97, 96, 10] gave vigorous reviews on the development of CBIR systems in medical domain in recent years. A number of frameworks have been proposed for retrieving medical images using CBIR. As summarized by [97], most CBIR systems have architecture as shown in Figure 2.6. We will give an introduction to the two mostly cited CBIR systems–ASSERT [110] and IRMA (designed for anatomy and modality classification) [70] in the following section.
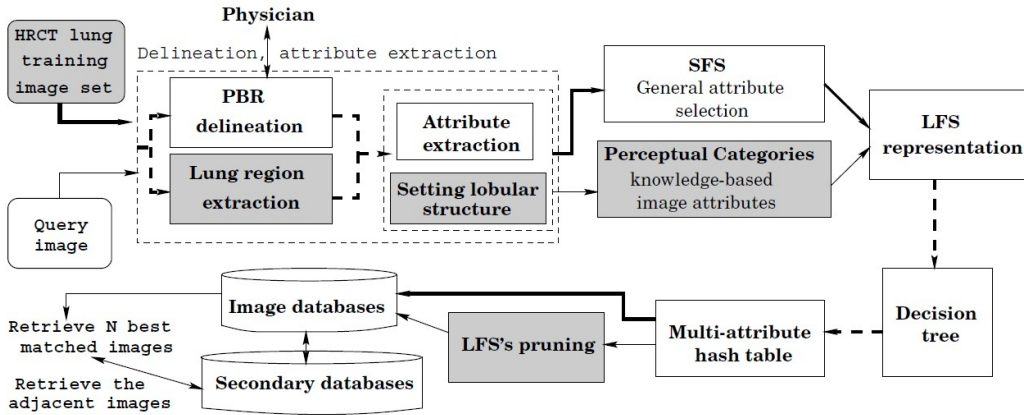
Figure 2.7: The general architecture of ASSERT

2.2.1 ASSERT

ASSERT [110] is designed for content based retrieval of high resolution computed tomography (HRCT) scans of the lung. Figure 2.7 illustrates the general architecture of ASSERT system. In the preliminary step, a human expert (physician) delineates the pathology bearing regions (PBRs) and any relevant anatomical landmarks. The system uses image processing algorithms to detect the boundaries of the lung and extract the lung region. Then The systems extract features that describes each individual PBR and store the features in a vector. To reduce the dimensionality and yet achieve a good PBR classification result, Sequential forward selection (SFS) is applied to the attribute vector for feature selection.

The features of each PBR and the location information of the PBR are represented by a lobular feature set (LFS) in ASSERT as shown in figure 2.8. The location of PBR is either in the lobular region or adjacent to the boundary of a lobular region. LFS is a grouping of a specific region of the lung, usually a lobe

LFS :

```
lobular_region_index,
attributes: {
            adjacent_PBR_attribute1,
            adjacent_PBR_attribute2,
                        ⋮                       for the PBR
                                                 adjacent to
            adjacent_PBR_attributeN,            lobular boundary

            interior_PBR_attribute1,
            interior_PBR_attribute2,
                        ⋮                       for the PBR
                                                 interior to
            interior_PBR_attributeN.            lobular boundary
}
```

Figure 2.8: The lobular feature set (LFS) of ASSERT

or a combination of adjacent lobes, and the PBRs found therein. ASSERT retrain only the largest PBRs for image archiving and retrieval as the case that more than one PBR is inside the lobular region at the same time or more than one PBR is adjacent to the lobular region at the same time is rare.

Each LFS is represented by a single point in a $2N$ dimensional attribute space and is classified to a LFS class using a Gaussian approximation. When a new query image is submitted to the system, its LFS is translated into a hash index that directly points to the appropriate bin of the attribute space. This bin contains pointers to the relevant LFS classes whose distributions cover that bin. The class pointers link to a set of images that are similar to the query image on the basis of LFS similarity. Using a Euclidean metric, the database images thus retrieved are then tested more directly for similarity to the query image and ranked on that basis.

41

2.2.2 IRMA

Keysers et al. [70] developed a content based image retrieval system in medical applications (IRMA) for medical images of various modality, body orientation, anatomic region, or biological system. As image features for general domain images based on color, texture, or shape do not supply sufficient semantics for medical applications, the IRMA concept is based on a separation of seven steps to enable complex image content understanding as shown in Figure 2.9:

1. image categorization

2. image registration

3. feature extraction

4. feature selection

5. indexing

6. identification

7. retrieval

For image categorization, global features are used in the IRMA to distinguishes four major categories: image modality, body orientation, anatomic region, and biological system. Then prototype images, which are defined for each category by an expert based on prior medical knowledge or by statistical analysis, are used for determination of parameters for rotation, scaling, and translation as well as contrast adjustment in the image registration step of IRMA. Then local features are
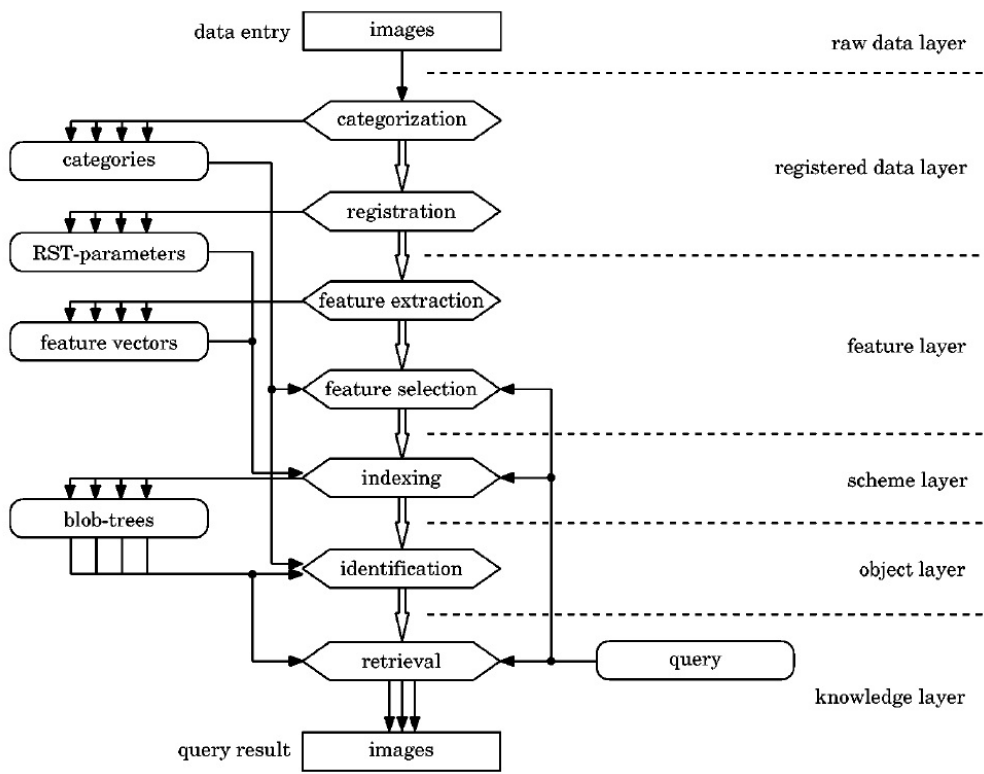
Figure 2.9: The program flow for IRMA system

extracted in the pixel level. Category-specific local features include segmentation by active contours or active shapes, which enable the use of prior shape knowledge from the categories. In next step, prior knowledge about the image category as well as medical knowledge incorporated into the query is used to select a pre-computed set of adequate feature images. Based on feature sets, the image is segmented hierarchically into relevant regions as "blobs" to reduce the amount of information the system has to process for image query. Thereafter, the blob representation of the image is adjusted with respect to the parameters determined in the registration step. The hierarchical blob tree has been registered with respect to a certain category, where categories are represented by prototypes. Since the local features reflect characteristic and discriminant properties of tissue, certain blobs correspond to well defined morphological structures in the image. Vice versa, prior medical knowledge on the content and structure of category prototypes can be used to build a prototype blob structure with characteristic properties for blob identification and labels for semantic queries. In image retrieval step, IRMA performs searches in the hierarchical blob structures and built from a list of possible categories of the recall images, a query by example blob-structure on the optimal scale to process the query, or the set of local features that best describe significant properties for the query.

2.2.3 Challenges

Most CBIR systems in general domain used visual features like color, texture, and shape. However, these primitive visual features do not give enough information

to describe an image in the medical domain. There exists a semantic gap due to this loss of information from an image to such representation by low level visual features. The interpretation of medial images using higher level features are yet to be investigated by most medical CBIR systems.

The lack of publicly available large medical image data set is another obstacle in this research area. Due to many reasons such as non-disclosure agreement between the medical image provider and research institute, the amount of medical images used for research is very limited though the absolute volume of medical images is huge. Even if the research institutes have successfully acquired enough images for research, they cannot make the data public for others to use and compare the results with.

2.3 Automatic image annotation using unsupervised methods

Unlike supervised learning which uses class labels in training to reinforce the learning, unsupervised learning seeks to determine how the unlabeled data are organized. For most annotated image corpus, users only annotate the images but not object by object, i.e. they do not label each object in the images with annotations. Therefore, to recognize which part of the image is which object, we need to resort to unsupervised learning approaches.

Compared to supervised learning in image annotation tasks, unsupervised learning based methods have their advantages: they make an assumption of a model

which can express explicitly the complex relationships between textual words and visual features by incorporating available a prior information. What is more, some approaches, such as the co-occurrence model [95] and the translation model [38], can even associate a word to each region in an image. This annotation-by-region strategy is more informative than annotating an image as a whole.

However, most of the unsupervised learning based methods rely on an EM algorithm for training. The EM algorithm is sensitive to the initial parameters and with its complex objective function it can only produce a local optimum solution, which in turn leads to inferior performance of the model to unseen data. For the non-parametric models, such as CMRM [63], they need to store the whole training data in the annotation system, which is not desirable for large database. Also, non-parametric models assumes that a perfect set of data are available to be used as the reference set, which is not usually the case. Wang Yong [125] has given a detailed review on both parametric and non-parametric models. Our literature survey for automatic image annotation for general domain images will follow their categorization and summary of the existing models.

2.3.1 Parametric Models

In statistics, a parametric model is a collection of probability distributions such that each member of this collection, P_θ, is described by a finite-dimensional parameter θ. In image annotation, parametric models are used to model the co-occurrence between the image regions and textual words. The parameter estimation process, i.e. the training of the parametric model, is usually done by EM

algorithm [30]. The advantage of using a parametric model is that in the annotation result, image regions are explicitly associated with the textual words, making users to know which part of image is labeled with which word. However, EM algorithm can only produce a local optimum and is sensitive to initial parameter settings.

Mori et al. [95] are the first ones to use a parametric model to solve image annotation problem. They proposed a co-occurrence model to represent the relationship between keywords and visual features. Each image is converted into a bag of rectangular image regions obtained by a regular grid. The image regions from the training data are clustered into a number of region clusters. For each training image, they propagate its keywords to each image region in this image. The conditional distribution of keywords of each region cluster can be estimated from the empirical distribution on the training data. Given a new image, the conditional keyword distribution of each individual image region are aggregated to generate the conditional keywords distribution of the test image. The major drawback of the above co-occurrence model is that it assumes that if some keywords are annotated to an image, they are propagated to each region in this image with equal probabilities. This assumption is violated in many real situations because many keywords are object names such as sky,sun and water. The appearance of this kind of concept in an image is usually a small portion of an image instead of the whole image.

Duygulu et al. [38] proposed a machine translation model for image annotation, which is essentially an improvement of the co-occurrence model of Mori et

47

al. [95]. They represent an image as a bag of image regions obtained by image segmentation and performed vector quantization on each of these region features. The vector quantized image regions are treated as visual words and the relationship between these and the textual keywords can be thought as that between one language, such as French, to another language, such as German. The training set is analogous to a set of aligned bitexts, i.e. texts in two languages. Given a test image, the annotation process is similar to translating the visual words to textural keywords using a lexicon learned from the aligned bitexts. They found that a relatively simpler translation model used in the language translation, i.e. the model of Brown et al. [8] produced better performances than other available language translation models. Similar to the co-occurrence model [95], the learned parameters of the translation model are also the conditional distribution probability table, but the translation model does not propagate the keywords of an image to each region with equal probability. Instead, the association probability of a textual keyword to a visual word is taken as a hidden variable and estimated by an Expectation-Maximization (EM) algorithm [30].

Ghoshal et al. [45] proposed a hidden Markov model (HMM) approach, which is similar to but simpler than the machine translation approach. A hidden Markov model (HMM) is a statistical model in which the system being modeled is assumed to be a Markov process with unobserved state. In Ghoshal et al.'s approach, each textual keyword is represented by a hidden state, which can generate visual features following a per state probability distribution. The training process aims to find the best correspondence of image regions and textual keywords and estimate

the parameters for each state. The annotation process of a new image is equivalent to recovering the most likely hidden state of each image region. A major difference between the HMM approach and the machine translation model is that the HMM approach models the continuous distribution of visual features, whereas the translation model represents the keyword distribution of each vector quantized image region. However, the HMM model assumes a transition process between different states (textual keywords) which is not necessarily supported by real data.

Instead of modeling the conditional distribution of textual keywords based on visual features, Barnard and Forsyth [4] proposed methods to model the joint distribution of textural features and visual features. They define a document as a combination of visual features and textual features. A hierarchical factor model is proposed to model the joint distribution of textual features and visual features. The model assumes that a document belongs to a cluster, which is denoted by the leaf nodes in the tree hierarchy. Given the document and the cluster it belongs to, the document is generated by the aspect nodes on the path from the root node to the leaf node following the hierarchical structure. Each aspect on the path can generate image regions and textual features following a per aspect probability distribution. Since different clusters have distinct traversing path, each has a separate joint models of the aspects for each other. Moreover, since all the aspects are organized in a hierarchical structure, the aspects are very compact and it can model the commonalities between clusters in different degrees between. However, this model is optimized for image clustering instead of linking textual words to image regions.

Blei and Jordan [6] proposed the correspondence latent Dirichlet allocation (Corr-LDA) model to find a conditional relationship between image features and textual features. In their work, the dependence of the textual words on the image regions are modeled explicitly.

Monay and Gatica-Perez [94] explored latent semantic analysis (LSA) [29] and probabilistic latent semantic analysis (PLSA) [55] for automatic image annotation. In short, a document of image and texts can be represented as a bag of words, which includes the visual words (vector quantized image regions) and textual words. Then LSA and PLSA can be deployed to project a document into a latent semantic space. Annotating images is achieved by keywords propagation in this latent semantic space.

2.3.2 Non-Parametric Models

Non-parametric models differ from parametric models in that the model structure is not specified a priori but is instead determined from data. They make no assumptions about the probability distributions of the variables being assessed. The advantage of using non-parametric models in automatic image annotation is that they are easy to implement as they do not have a training process as parametric methods do. However, also due to the lacking of learning process, they have to store all the data instead of just the model parameters, which is not desirable for large databases.

Joen et al. [63] proposed one of the best performing non-parametric models. They formulated the problem of automatic image annotation as cross-lingual in-

formation retrieval and have applied the cross-media relevance model (CMRM) to image annotation. Although CMRM also tries to model the joint distribution of visual features and textual words, it is a non-parametric model, like the k-NN [37] approach for pattern classification. The essential idea is that of finding the training images which are similar to the test image and propagate their annotations to the test image. CMRM does not assume any form of joint probability distribution on the visual features and textual features so that it does not have a training stage to estimate model parameters. For this reason, CMRM is much more efficient in implementation than the above mentioned parametric models. A drawback of the CMRM model is that it vector quantized the image regions into image blobs and this can reduce discriminative capability of the whole model. So Lavrenko et al. [76] have proposed an improved model, i.e. the continuous cross-media relevance model (CRM). CRM preserves the continuous feature vector of each region and this offers more discriminative power.

Jin et al. [64] proposed a coherent language model which is extended from CMRM to model the correlation between two textual words. The model defines a language model as a multinomial distribution of words. Instead of estimating the conditional distribution of a single word, they estimate the conditional distribution of the language model. The correlation between words can be explained by a constraint on the multinomial distribution that the summation of the individual words distribution is equal to one. Thus the prediction of one word has an effect on the prediction of another word.

2.4 Automatic image classification using supervised methods

Recently various supervised machine learning methods, i.e. discriminative training methods or classification methods, are used to automatically annotate images. Image classification itself has a much longer history. Research in this field mainly focused on some special image domains until recent years. Due to the rapid progress in the quality of imaging device and digital image databases (online and off-line), huge amount of images of better quality are made available, and more attention has been attracted to general domain image classification, rather than special domain alone.

Image annotation by image classification has obvious advantages: we can adopt existing well-studied statistical classifiers as long as we have a suitable representation of images; it offers a sophisticated classifier for those specific concepts with enough training data. However, there is disadvantage of the classification approaches. Since they view each textual word as an independent class label, they ignore the correlation between these words, which is rather helpful because humans usually annotate an image with a set of words with coherent meaning.

Supervised learning approaches view image annotation as classification problem – each textual word in the annotation is considered as an independent class label; the images or image regions are classified according extracted visual features into these classes; the classification result is the annotation result. As summarized by Wang Yong [125], existing image classification approaches to image

annotation are mainly based on global features, local features, or multi-level classifications.

2.4.1 Global Feature Based Image Classification

Global scene-oriented classification methods which extract a global feature descriptor from an image and then deploy a statistical classifier for image classification. Examples of this kind of class label include "countryside", "landscape", "outdoor" and so on. The task is usually classifying the image as a whole.

Chapelle et al. [16] use support vector machine (SVM) to classify images by global features. They used an enhanced heavy-tailed RBF kernel for high dimensional image features. Fung and Leo [42] decompose the semantics of a scene image into two levels: (1) the primitive semantics at the patch level, and (2) the scene semantics at the image level. The learning of primitive semantics is based on a supervised clustering of the patch features. Their scene classification is achieved by using the distribution of each primitive in an image.

Scene semantics are made more explicitly by Vailaya et al. [121] who proposed a method for hierarchical classification of vocational images: at the highest level, images are classified as "indoor" or "outdoor"; "outdoor" images are further classified as "city" or "landscape"; finally, a subset of "landscape" images is classified into "sunset", "forest", and "mountain" classes. They model the probability density of each scene class through vector quantization and classify images based on the maximum a posterior criterion. Chang et al. [14] proposed a soft categorization method of images based on the Bayes point machines (BPM) [53], which

is another advanced kernel based classifier.

2.4.2 Regional Feature Based Image Classification

Instead of global visual features, some approaches use regional visual features to categorize the whole image. Wang and Li [123] proposed an image categorization method using the 2D multi-resolution hidden Markov model (2D-HMM). Images are segmented into regions by employing a multi-resolution regular grid. 2D-HMM can model the dependency between regions in the same resolution and the regions across different resolutions.

Carneiro and Vasconcelos [12] proposed an image annotation framework based on hierarchical mixture modeling of the probability density estimation of each class. Each image is represented as a set of patch features. The distribution of these patch features for each concept is modeled as a Gaussian mixture model and all the concepts are modeled by a hierarchical Gaussian mixture model (Hier-GMM). Their experimental results show that the Hier-GMM is efficient for large database.

Maree et al. [82] proposed an image classification method by combining the random sampling of subwindow images and an ensemble of extremely randomized tress. Since they have added various transformations in the process of abstracting random sub windows, their approach is robust to both scale and rotation, however they have not tested their approaches on a more complex image dataset for image annotation. Some methods are based on sophisticated probabilistic models.

54

Li and Wang [77] represent each image as a probabilistic distribution of color and texture features. Each image category is modeled as probabilistic distribution of probabilistic distributions. Taking advantage of the fast optimization algorithm, their approach can achieve real time annotation performance on a large scale dataset. However, it is not clear how well their method can perform on individual object concepts.

These above mentioned image classification approaches have been proved to be effective in classifying many scene categories, such as "sunset", "landscape" and "countryside", but they have not shown any advantage in classifying object names, such as "sky", "tiger", "horse" etc.

2.4.3 Regional Feature Based Object Classification

Local object-oriented classification methods classify images by object names. The image content assigned to the labels is usually a part of the image. Examples of these class labels include "sky", "water", "people" and so on.

For individual objects, the corresponding visual appearance in the image is usually a segment of the image instead of the whole image. Sometimes, even collectively, these object segments may only make up a small part of an image. This makes a global visual feature not always an appropriate solution, especially in the case of heavy background clutter or when a number of different objects exist in the image. Therefore, treating an image as a bag of image regions and annotating image by these regions is helpful for the object-based classification of images.

Yang et al. [126] formulated image annotation as a a multiple instance learning (MIL) problem. In the MIL setting, the object to be classified is a bag of instances instead of a single instance. The training data is a set of positive bags and negative bags. A bag is labeled as positive if at least one of the instances in the bag is labeled as positive. A bag is labeled as negative if none of the instances in the bag is labeled as positive. The labels on the training data are only provided for each bag, not for each instance. Given a new unlabeled bag, we need to classify it as positive or negative. This kind of problem can not be solved by traditional statistical classifiers where each training example or test sample is represented as a single feature vector instead of a bag of feature vectors.

2.5 Automatic Medical Image Annotation and Classification

Numerous tasks of automatic medical image annotation have been proposed in the tracks in[119]. The tasks are focused on classifying medical images to different categories of acquisition modality (CT, X-ray, MR, etc.), body orientation, body region, and biological system. Some research works aim to identify abnormal medical images, i.e. classifying images into the categories of normal and abnormal. According to Doi's review on computer-aided diagnosis (CAD) in medical imaging [31], large scale and systematic research and development of various CAD schemes are of growing interest in recent three decades, especially for chest and breast images. A typical medical image analysis system consists of four major

parts as shown in Figure 2.10: image preprocessing, image segmentation, feature extraction, classification. The image preprocess step usually includes noise removal and image normalization. Then in segmentation step, various anatomical entities are identified. Some research works use image registration after the segmentation step. In feature extraction step, global or regional image features, as well as features regarding to a certain region of interest, are extracted and selected. Finally, various classification methods are applied to categorize the image or image regions.
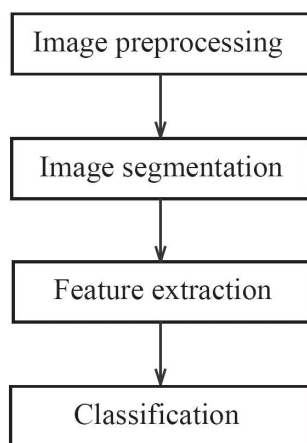
Figure 2.10: A typical medical image analysis system architecture

Sluimer et al. [111] gave a detailed review on computer analysis of lung CT images, in which a large amount of research work is concentrated. Systems on automated detection, quantification, and classification of pulmonary disorders can be grouped by clinical application areas of emphysema, lung cancer, signs of airway diseases, and differential diagnosis of lung disease. [120, 99, 5] detect emphysema and classify lung images into normal and emphysema categories. [89] classifies

lung images into normal and chronic obstructive pulmonary disease. [51, 92] classify the images into different levels of severity. For lung cancer detection, the main focus has been in the detection of lung nodules. Research works measure nodule size and characterize the nodule appearance to determine the probability of the nodule being benign or malignant [52, 85, 86, 67, 68, 69, 83, 80, 3, 2]. Different machine learning methods and algorithms are used to build the classifier, including rule-based methods, linear classifier, LDA, template matching, clustering algorithms, Markov random field, neural network, and Bayesian classifier.

Cheng et al. [19, 21, 20] conducted detailed surveys on computer analysis of breast images from mammography and ultrasound, which is another major category of medical images that many research work focus on. As early detection remain the best opportunity to reduce the mortality caused by breast cancer, the main task for computer systems in this research area is to detect tumors and distinguish the benign and malignant ones. Most CAD systems for breast cancer detection and classification follow the general architecture shown in Figure 2.10. In image preprocessing phase, the main tasks are image enhancement and speckle reduction. Image segmentation partitions the image into non-overlapping regions and separates the objects from the background. The regions of interest (ROIs) are labeled for feature extraction. Then in next step, the key is to find a feature set of breast cancer lesions that can accurately distinguish lesions and non-lesions, or benign and malignant tumors. The features used for breast images can be categorized into the groups of texture features [15, 58, 59, 60, 18, 102, 73, 74, 43], morphologic features [106, 17, 34, 113, 66, 102, 56, 115, 35], descriptor features (such

as shape, presence of calcification) [106, 118, 115], and model-based features [44, 108]. The feature space could be very large and complex, feature selection is usually included in this phase. Finally, based on the selected features, the candidate regions are classified using various classification methods such as template matching [58, 59, 74], linear classifiers [106, 56, 93, 44, 108, 43], neural networks [17, 59, 113, 66], decision tree [18, 73], Bayesian belief network [34, 36, 33, 35] and support vector machines [15, 60, 102].

2.5.1 Brain CT image annotation and classification

As an earlier research work in the area of brain CT image annotation and classification, Cosic and Longaric [25] proposed a rule-based approach to the labeling of computed tomography (CT) head images containing intracerebral brain hemorrhage (ICH). They partitioned the original image into a number of spatially localized regions of same color intensity using fuzzy clustering algorithm. Then they crafted a list of rules to label the image regions as background, brain, skull, hematoma, and calcification. Because the rules are manually created, the system they developed lacks flexibility when adapting to future needs. Experimental results are also lacking from the paper.

Liao et al.'s work [79] is on pathology based brain CT image classification, and we consider it most related to our work. They obtained 48 brain CT images and classify them based on three hematoma types: epidural, subdural, and intracerebral. After they segmented the hematoma region from the image, they extracted the shape features of the hematoma region and constructed a decision tree based

59

on the features. The features included long axe (LA) and short axe (SH) of the hematoma region, the depth of points LA1 and LA2, D(LA1), D(LA2), their sum D(LA1)+D(LA2), the number of blocks in the larger and smaller halves on each side of the long axis, the percentage of the smaller half. Figure 2.11 illustrates the skull recognition and long/short axes labeling of hematoma regions. Then C4.5 algorithm is applied to generate the decision tree as shown in Figure2.12. As they used all 48 images for training, they achieved 100% precision and recall for training data classification; however, the classification result of any testing image is lacking.

Figure 2.11: Liao et al.'s measurement for hematoma axis

Zhang and Wang [127] used mainly global image features to detect abnormal brain CT images without explicit hematoma segmentation. They extracted intensity, shape, texture, and symmetry features of the image and classify the images into normal and abnormal categories. The color intensity features included mean, variance, skewness, and kurtosis values of the whole image. They selected the lateral ventricles as the region of interest (ROI), computed its distortion and
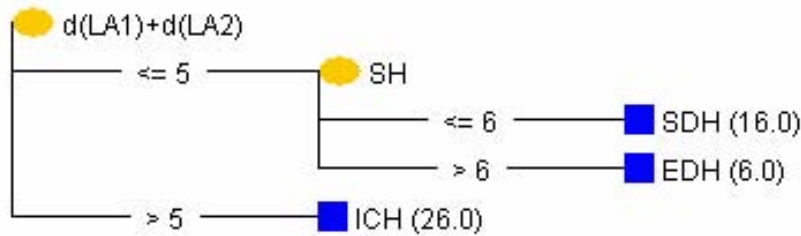
Figure 2.12: The hematoma classification decision tree generated by Liao et al's method

treated it as a shape feature, because with the presence of cerebral hemorrhage, the lateral ventricles will usually be misshaped to some extent. Energy (Angular Second Moment), Contrast, Inverse Difference Moment, and Entropy are used as texture feature for the brain image. They also extracted symmetry feature from the image by comparing the pixels on each side of the brain midline. After feature extraction, they used See5 which is based on C4.5 decision tree algorithm and Radial Basis Function Neural Networks (RBFNN) for image classification. They obtained 212 images in total. 103 of which are normal and the remaining 109 are abnormal. They used 80% of the images for training, and the rest for testing. The results are as follows in Figure 2.13 and 2.14.

Peng et al. [100] used regional features to classification stroke and tumor brain CT images. They first preprocess the image and partitioned the brain content part into four regions of the same size in either way illustrated in Figure 2.15. Then they generated the gradient (edge) of the original partition and the X and Y gradient images. Next they extracted the mean or standard deviation values of the generated images, and the percentage of the area above certain threshold. They

61

Feature	See5	
	Train-rate	Test-rate
Histogram	85%-90%	76%-80%
Gray scale	80%-83%	72%-76%
Shape	93%-96%	85%-90%
Texture	75%-80%	75%-78%
Symmetry	93%-95%	85%-93%
All features	96%-98%	90%-94%

Figure 2.13: The classification result by Zhang and Wang's method using See5

used SVM to classify the brain CT images into stroke cases and tumor cases. They obtained 25 stroke and 25 tumor cases and each case consists 9 images. They compared the results with a baseline classification implemented using Gabor features. The classification results for training is significantly better than the classification using Gabor features as shown in Figure 2.16; however, as they used all images for training, the classification results for testing images are lacking.

Feature	RBFNN		
	Hidden node	Train-rate	Test-rate
Histogram	10	83%-84%	80%-82%
	30	85%-86%	79%-82%
Gray scale	10	77%-78%	72%-75%
	30	82%-85%	73%-77%
Shape	10	85%-87%	84%-87%
	30	87%-89%	84%-87%
Texture	10	68%-70%	64%-65%
	30	77%-78%	67%-71%
Symmetry	10	88%-89%	85%-89%
	30	91%-93%	85%-89%
All features	10	80-%82%	77%-82%
	30	87%-89%	83%-86%

Figure 2.14: The classification result by Zhang and Wang's method using RBFNN
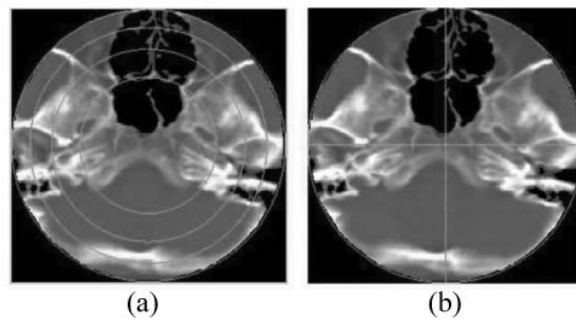
Figure 2.15: Brain CT image partitioning
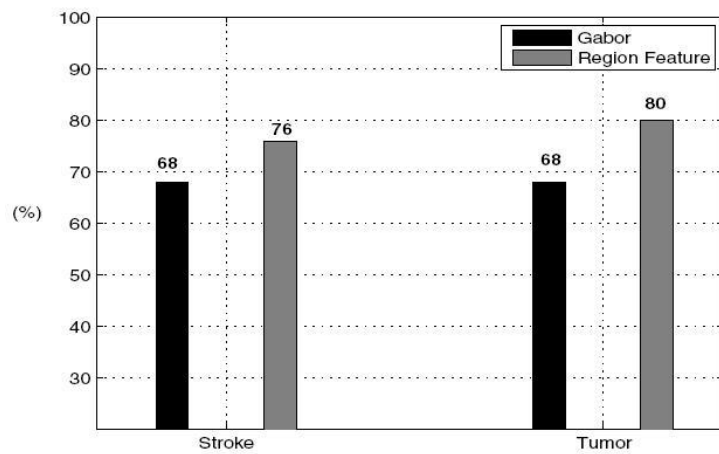
Figure 2.16: The image classification results by Peng et al's method

Chapter 3

Text processing in radiology reports

With the advances in medical technology and wider adoption of electronic medical record systems, large amounts of medical text data are produced in hospitals and other health institutions daily. These medical texts include the patient's medical history, medical encounters, orders, progress notes, test results, etc. Although these text data contain valuable information, most are just filed and not referred to again. These are valuable data that are not used to full advantage.

A similar situation occurs in the field of radiology. As the reports are in free text format and usually unprocessed, there exists a great barrier between the radiology reports and the medical professionals (radiologists, physicians, and researchers), making it difficult for them to retrieve and use useful information and knowledge from the reports. As the information is not accessible, it cannot be used for other related applications such as automatic image annotation. Therefore, to provide the needed information to the medical professionals as well as to

make use of the information, we need to process the unstructured text and extract structured information from it. We have also described the following framework in our paper in [50].

3.1 The medical text processing framework

The medical reports are usually written in natural language, and often contain short hand writing and acronyms. The goal of medical text processing is to extract the medical findings in the medical texts. The general architecture of our medical text processing system follows the program flow of most such systems in this research field, but we also emphasize on attribute extraction besides the main medical finding extraction. The attributes or modifiers of the medical finding such as "location", "duration", and "probability", describe the properties of the medical findings. They are valuable information that could be of important use of other applications developed upon the text processing system. For example, we will use the location information extracted in automatic image annotation training corpus generation described in Chapter in this thesis.

We take the free text medical reports as input, use natural language processing techniques and domain knowledge sources to extract medical findings and their modifiers, and outputs them in a structured form so that the extracted information can be easily accessed again. We use a semantic approach to achieve our text mining task. The system consists of the following components: report chunker, term mapper, parser, finding recognizer, and report constructor. The overall framework

is illustrated in Figure 3.1.
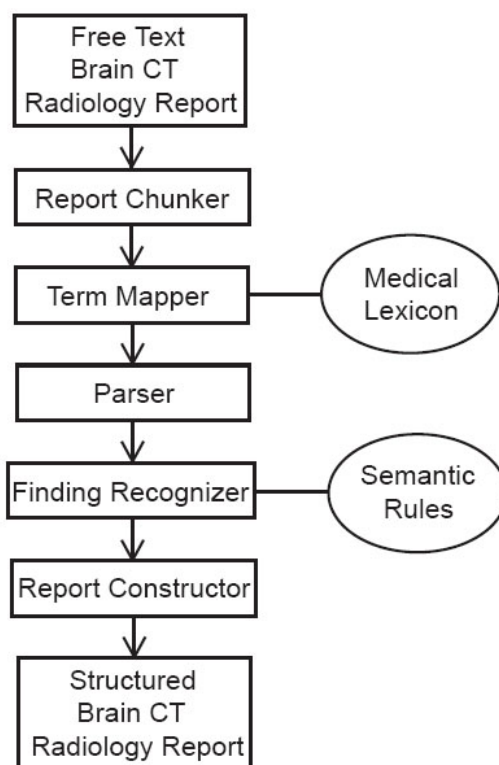
Figure 3.1: Program flow of radiology report processing

3.2 Report normalization and term mapping

To protect patients' privacy, all personal information of the patients have been anonymized prior to text mining process. Excluding patient's bio information and examination details, a radiology report contains a description section (main body of the report), and comment or impression section. We decompose the report into different sections and keep only the main body part and impression part of the

report for medical finding extraction, as these two parts contain detailed finding descriptions.

Then we do spelling correction and spelling variation reduction. From the reports used for training, we summarized a list of mapping of frequently misspelt words and the correct forms of the words. For example, "hemo**rr**hage" is often misspelt as "hemo**r**hage". During spelling checking, we replace the misspelt words with the correct ones if they are in the list. For other unrecognized words that are not present in the list of frequently misspelt words, which may be radiologist's short hand writing, we temporarily leave them in original form in the report and note them down for future improvement. We have also created a list of mapping of words frequently written in non-standard forms and their standard forms. The words in non-standard forms (e.g. "h**ae**matoma", "isch**ae**mia") are not considered spelt wrongly because they are recognized in the community; however, as they are usually not included in the medical lexicons, we replace them to standard forms in the reports so that the information contained in these words can be extracted. We also add a negation filter to detect negative expressions, so that we will only extract the terms referring to pathologies found in the images.

Then we map single-word and multiple-word terms to our medical lexicon and normalizes the terms to standard forms. Medical Subject Headings (MeSH) is a large controlled vocabulary developed by National Library of Medicine used for medical texts indexing [1, 107, 84]. We found it useful for building our brain radiology lexicon, especially the Anatomy (A) and Disease (C) sections. However, as MeSH does not cover the entire set of vocabulary for brain CT radiology reports,

it does not reach the degree of specificity required in the reports. Therefore, our lexicon resorts to other sources as well, including other radiology and anatomy thesaurus and actual brain CT radiology reports.

3.3 Parsing and relation extraction

Besides pathological and anatomical concepts, we also need to extract the relation between them. In order to obtain the sematic relation, we first extract the syntactic relation among them using a parser.

Our parser is developed based on the Stanford Parser [71] [72]. The parser parses each sentence and outputs the typed dependency tree [28], which shows the syntactic relations between the words and phrases in the sentence. Dependency grammars (DG) is preferred for medical text because most narrative medical documents contains many ungrammatical sentences [11]. In DG, each word has only one parent, i.e. a tree structure with the dependencies represents the sentence. For example, the typed dependency graph of the example sentence "There is large extradural haemorrhage in the left frontal lobe." is shown in Figure 3.2.

The finding extractor selects the findings and their modifiers according to a set of semantic rules. A medical finding in the brain CT report refers to the abnormality of the patient's medical condition. For example, "hematoma", "fracture", "midline shift" are common findings in brain CT radiology reports. Each finding may have several modifiers that describe the properties of the finding, such as "location", "duration", "probability" etc. The finding extractor makes use of the
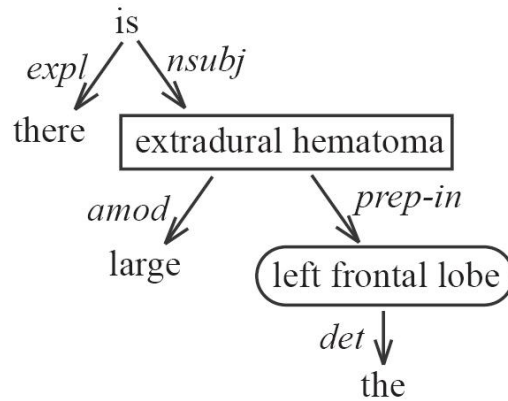
Figure 3.2: The typed dependency tree of example sentence.

intermediate result from term mapper and parser to locate medical findings and the modifiers, and uses a set of semantic rules to translate the syntactic relations in the result from the parser to logical relations between findings and their modifiers. The semantic rules were manually constructed and cover frequent patterns how the findings appear in the sentences.

Negations are also detected by the finding recognizer. Findings associated with negative expressions are outputted as negative findings. It is necessary to extract negative findings and include them in the structured report explicitly as both negative and positive findings are significant to the medical personnel accessing the report.

3.4 Constructing structured report

The report constructor then outputs the result from the finding extractor in XML format. We chose XML to represent the findings and their modifiers, because

XML documents are self-structured, self-defined, easy to understand and retrieve. The hierarchical structure in XML naturally shows the relationship between findings and their corresponding modifiers. Therefore, it satisfies the requirement to represent the medical findings and provides potential convenience for our future radiology report retrieval system. The output result of findings and modifiers of the example sentence in Figure 3.2 is shown in Figure 3.3.

```
<finding>
<pathological change>extradural hematoma</pathological change>
<size>large</size>
<anatomical location>left frontal lobe</anatomical location>
</finding>
```

Figure 3.3: The structured result of the example sentence: "There is large extradural haemorrhage in the left frontal lobe."

In the output, the tags <finding> and </finding> indicate that the content between them is a medical finding extracted from the input radiology report. The tag <pathological change>indicates the disease/abnormality the finding refers to, and other tags indicate different types of attributes that modify the finding.

3.5 Experiment and results

For experiments, we used 467 traumatic brain injury CT radiology reports from National Neuroscience Institute, Tan Tock Seng Hospital. 367 of which were set for training, and 100 were set for evaluation of our system. The average length of the reports is 12 sentences, or 157 words. 753 positive findings, 167 negative findings, and 1520 modifiers were labeled in the testing reports. There are 25 different

71

pathological changes of the medical findings in the testing reports, distribution of which is shown in Table 3.1 (percentages may not sum to 1 due to rounding).

Table 3.1: Pathological change distribution in testing reports

	Percentage
hematoma	33.7%
fracture	17.5%
midline shift	9.3%
effacement	7.3%
hydrocephalus	6.1%
contusion	5.8%
infarct	3.7%
fluid	2.9%
edema	2.2%
herniation	1.7%
swelling	1.5%
pnneumocephalus	1.4%
dilatation	1.2%
hemaantrum	1.0%
ischemia	0.8%
opacification	0.8%
dislocation	0.5%
scarring	0.5%
atrophy	0.3%
emphysema	0.3%
granuloma	0.3%
sinusitis	0.3%
inflammation	0.2%

The overall weighted average precision and recall for pathological change extraction are 95.5% and 87.9% respectively. The detailed evaluation result is shown in Table 3.2. Lower percentage of negative findings were correctly extracted compared to positive findings due to the additional task to recognize patterns

indicating negation in the sentence. Nevertheless, we listed negative findings as a separate category opposed to positive findings, for radiologists and physicians often want to find past reports with explicit presence or absence of certain disorder/disease. If negative findings are not explicitly labeled, the system may mix up positive and negative findings, which does not satisfy the user's needs. Therefore, the explicit labeling of negative findings is a key step prior to build our report and image retrieval system.

For those findings with correctly extracted pathological changes, we calculated the precision and recall for modifiers. The weighted averaged precision and recall for modifiers are 88.2% and 82.8% respectively. The detailed evaluation result for each type of modifier is shown in Table 3.3.

Table 3.2: Evaluation result for medical findings in brain CT radiology reports

	Precision	Recall
positive findings	96.0%	89.5%
negative findings	93.1%	81.0%

Table 3.3: Evaluation result for modifiers of the medical findings in brain CT radiology reports

	Precision	Recall
duration	96.2%	93.8%
location	86.1%	81.2%
amount	81.6%	77.5%
size	85.0%	81.9%
direction	94.7%	90.0%
probability	91.7%	82.5%
seriousness	83.3%	75.0%

3.6 Text-based query and retrieval

After the reports are structured in XML format, the reports and associated images can now be retrieved conveniently using text queries. As shown in Figure 3.4, the retrieval model consists of query analyzer and medical record retriever.
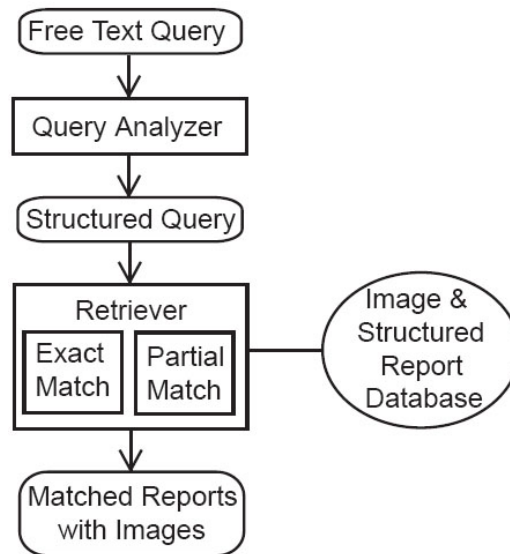
Figure 3.4: The components of report and image retrieval module

The query analyzer is essentially the same as the text processing module we described previously and illustrated in Figure 3.1. Instead of taking a full radiology report as input, the query analyzer takes the user's text query as input, which usually consists of a phrase or a few words. When a query from the user is entered to the system, for example, "acute subdural hematoma, no skull vault fracture", the query analyzer extracts the medical finding from the query and structure it as in Figure 3.5.

The retriever then searches the structured reports and images in the database

```
<finding>
<pathological change>subdural hematoma</pathological change>
<duration>acute</duration>
</finding>

<negative finding>
<pathological change>fracture</pathological change>
<anatomical location>skull vault</anatomical location>
</negative finding>
```

Figure 3.5: An example query in structured format

and return the ones that match the structured query. There are two modes of retrieval: exact match and partial match. Exact match returns results that are mostly needed by the user, whereas partial match returns results similar to what the user queries and facilitates the user to compare similar cases.

Under the exact match mode, only the reports containing exactly the same pathological changes and modifiers as in the query are returned. Take the same query for example, in the mode of exact match, reports with "acute subdural hematoma with fracture in skull vault" (see Figure 3.6) are not returned, as the "skull vault fracture" finding in the query is negative, whereas in the report it is positive. The explicit labeling of positive and negative findings during medical finding extraction is also for more accurate retrieval here. Reports with "chronic subdural hematoma"(see Figure 3.7) or "longitudinal fracture through right temporal bone"(see Figure 3.8) are also not returned in the retrieval results, as their modifiers (duration, type, location) of the findings do not match with the query's. However, reports with more information than necessary can also be returned as result as long as the medical findings and modifiers specified in the query is a sub-

set of the reports' full list of findings. For example, reports with "acute subdural hematoma in left temporal lobe with size of 1.4cm × 1.2cm, no evidence of skull vault fracture" (see Figure 3.9) are returned.

```
<finding>
<pathological change>subdural hematoma</pathological change>
<duration>acute</duration>
</finding>

<finding>
<pathological change>fracture</pathological change>
<anatomical location>skull vault</anatomical location>
</finding>
```

Figure 3.6: Brain CT radiology report in structured format: fragment example 1

```
<finding>
<pathological change>subdural hematoma</pathological change>
<duration>chronic</duration>
</finding>
```

Figure 3.7: Brain CT radiology report in structured format: fragment example 2

```
<finding>
<pathological change>fracture</pathological change>
<type>longitudinal</type>
<anatomical location>right temporal bone</anatomical location>
</finding>
```

Figure 3.8: Brain CT radiology report in structured format: fragment example 3

On the other hand, if the user chooses to use partial match, then reports with findings and modifiers that match part of the query are returned as well as the exactly matched ones. For example, the reports with "acute subdural hematoma with fracture in skull vault", "chronic subdural hematoma", "longitudinal fracture

```
<finding>
<pathological change>subdural hematoma</pathological change>
<duration>acute</duration>
<anatomical location>left temporal lobe</anatomical location>
<size>1.4cm * 1.2cm</size>
</finding>

<negative finding>
<pathological change>fracture</pathological change>
<anatomical location>skull vault</anatomical location>
</negative finding>
```

Figure 3.9: Brain CT radiology report in structured format: fragment example 4

through right temporal bone", which are rejected in exact match mode are returned under the partial match mode.

By indexing using structured text, medical images associated with the reports can be retrieved too. For example, by entering text query "intracerebral hemorrhage" into our system, images in Figure 3.10 will be retrieved.

3.7 Discussion

The system performed well on medical finding types and modifier types that are more frequent. For example, the extraction of "hematoma" and "midline shift" have higher accuracy compared to "inflammation". Abbreviations, misspellings and short-hand writing affect the recall too, as they are more difficult to map to medical lexicon. The presence of ambiguous sentence structure often confuses the parser and may cause the parser to build the wrong dependency tree and thus creating wrong association between medical findings and modifiers or negation

indicators. For example, "No midline shift, hydrocephalus or effacement of basal cisterns is seen." It is difficult for the system to tell whether "no" is associated with only "midline shift" or all three medical findings.

As we index the reports with detailed modifiers of each medical finding, users are able to search reports or images with more specific request such as a certain pathology instead of a more general query such as "CT head") compared to the system we surveyed. We also have implemented a web interface for the end users to access and evaluate. The location modifier describing the anatomical location of the medical findings we extract can be used to identify and label the corresponding region of interest in the medical images associated with the text. We will discuss the details of this aspect in Chapter 6 on automatical training data set generation.

The overall result is satisfactory to the radiologists in the hospital where the reports were obtained. The accuracy and recall of text based medical image retrieval are good enough to cater the need of our collaborators from the hospitals to search and retrieve images pertaining to certain pathology described in the reports. However, more medical professionals from other medical institutions are yet to form a panel to evaluate the system independently. To improve the performance of the system in the future, we can look into problems like abbreviation mapping, term normalization (including misspellings), and co-reference resolution.
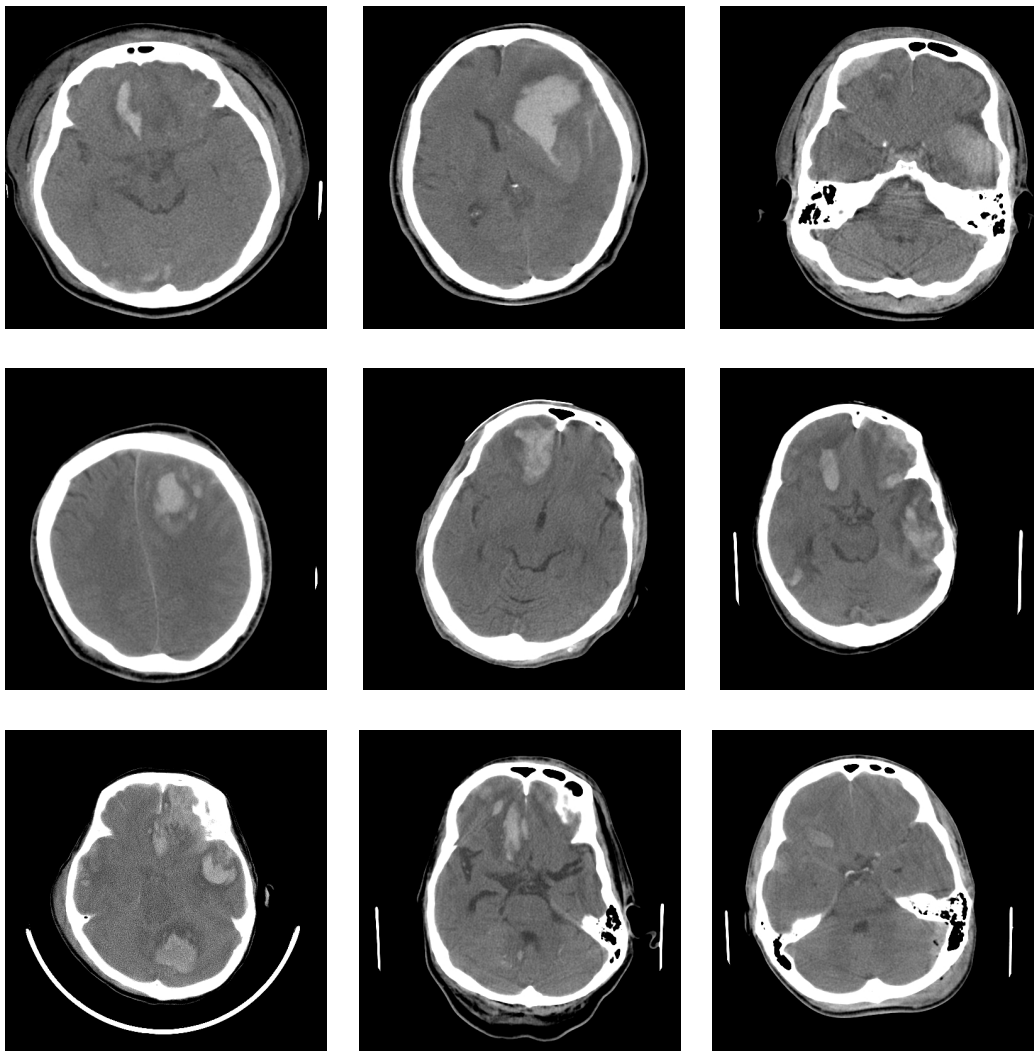
Figure 3.10: Image retrieval results of text query: intracerebral hemorrhage

Chapter 4

TBI CT image processing and visual content based retrieval

In Chapter 3 we proposed a framework for text-based medical image indexing and retrieval by extracting structured information from accompanying medical texts. However, medical images without associated reports cannot be indexed and retrieved by the text-based approach. In this chapter, we propose another framework to index and retrieve medical images by their visual content[1]. We first process the medical images using image processing methods and tools and segment the regions of interest (ROIs). Then we convert the ROIs to binary visual feature vectors and use these vectors to represent the images in database. For traumatic brain images, we propose a novel representation of the brain CT ROIs – we partition the brain image into circular bins which can naturally preserve both

---

[1]This part is done jointly with Dr. Li Shimiao [78].

the ROI intensity and location information. When new image query is submitted by the user, we process it in the same way to get its binary visual feature vector and calculate the similarities of it to the vectors in the database. We return a list of medical images ranked by the visual similarity to the query image. With the use of circular bin binary representation of the brain hematoma, the image processing is fast and can be built into real time online applications in contrast to most content based image retrieval systems, in which the slow image processing speed hinders the real time application development. The circular bins also fit well with the brain image in shape and preserve the location information of the segmented hematoma, which makes the image binary representation as well as image retrieval more accurate.

The image processing and segmentation is also a first step to further image semantic analysis which we will describe in more details in Chapters 5 and 6.

4.1 Intracranial region segmentation

Figure 4.1 shows an example brain CT image of traumatic brain injury. The aim of intracranial region segmentation is to remove the skull and the regions outside the skull, so that only regions inside the brain is left in the image. We first segment the skull by thresholding. If the scanning device is also captured in the image, we remove those regions by thresholding as well. We fill the "hole" of the skull and construct a intracranial region map. We overlap it with the original image to get the actual intracranial region. The image after this step is shown in Figure 4.2.
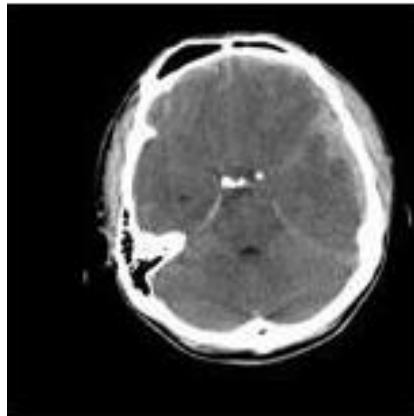
Figure 4.1: The original image
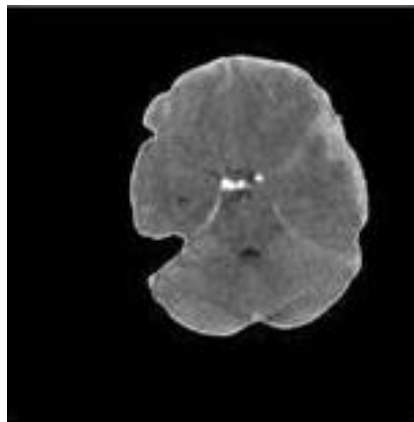
Figure 4.2: Step 1: skull removal

Then we normalize the positioning and intensity. In different cases, the placing of patient's head in the CT machine may be different, results in different center and symmetry axes in the images. We find the centroid and the rotation angle and reposition the intracranial region so that it is in the center of the image with a vertical midline. Due to different setting of CT imaging equipment, the intensity level varies in different cases. To make the intensity standard constant through

out all images, we set the peak of the intensity histogram to be 100 and adjust the gray level accordingly. Figure 4.3 shows the result.
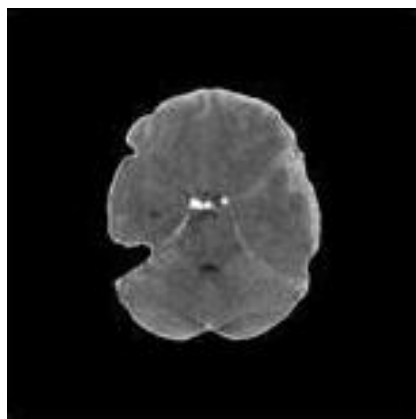
Figure 4.3: Step 2: position adjusting

However, the gray level of the brain content adjacent to the skull may still be affected by the gray level of the skull. In most cases, they appear whiter than they ought to be. This is called "cupping effect" in CT imaging. We use the method described in [13] to remove the cupping artifacts. Figure 4.4 shows the image after cupping effect removal.

As the intracranial regions have various sizes in different images, in order for the low level features comparable, we resize the intracranial region so that it spreads over the square bounding box as shown in Figure 4.5.

4.2 Low level visual feature extraction

Then we set a threshold to segment the possible hematoma regions in the image. To preserve the visual feature as well as to represent the hematoma regions effi-
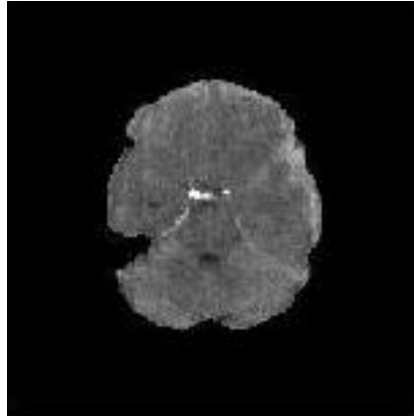
Figure 4.4: Step 3: cupping artifacts removal
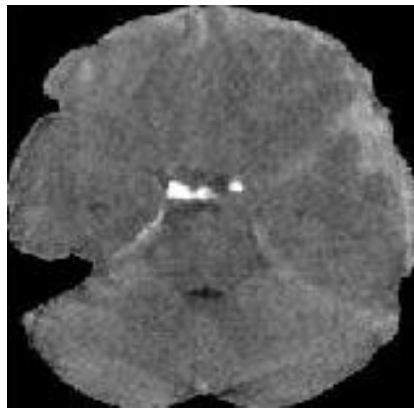
Figure 4.5: Step 4: resizing

ciently, we exploit the fact that intracranial regions are mostly in disk shape and we binarize the segmented hematoma regions according to the circular bins as shown in Figure 4.6 rather than bins of other shapes.

We partition the intracranial region into $R \times T$ equal-sized bins, where R is the number of partitions along the radius and T is the number of angle partitions. We represent the intracranial region segmentation result using a vector v of size
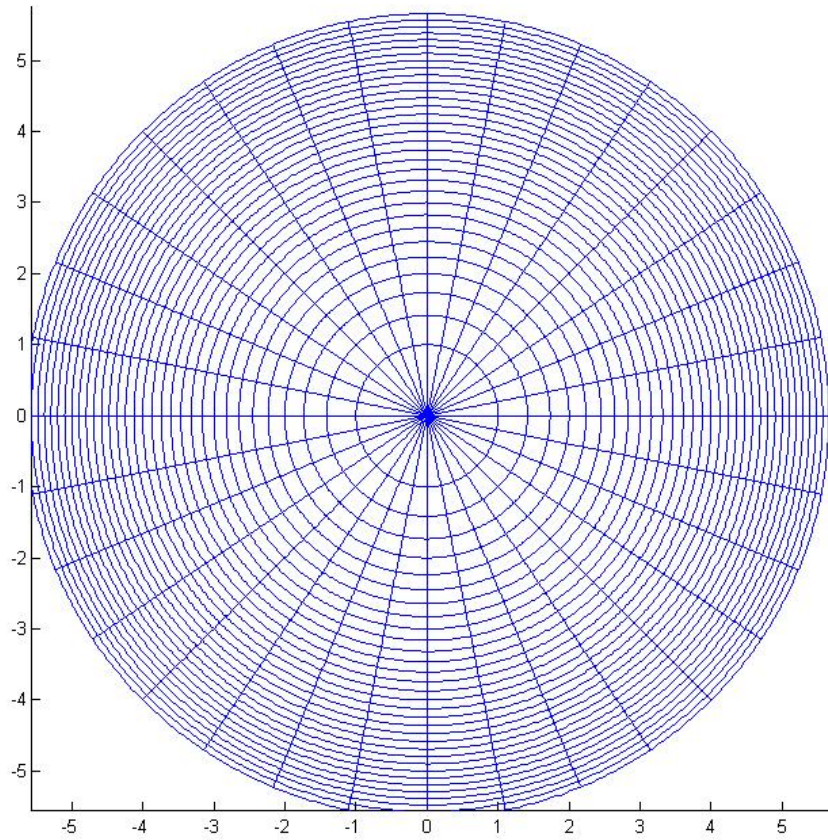
Figure 4.6: Circular bins used for binary feature vector extraction

$R \times T$: $v_{i,j} = 1$ if average intensity in $bin_{i,j} \geq 0.5$; and $v_{i,j} = 0$ if average intensity in $bin_{i,j} < 0.5$, where $1 \leq i \leq R$ and $1 \leq j \leq T$. Figure 4.7 and 4.8 demonstrate the complete process of visual feature extraction for TBI CT images.
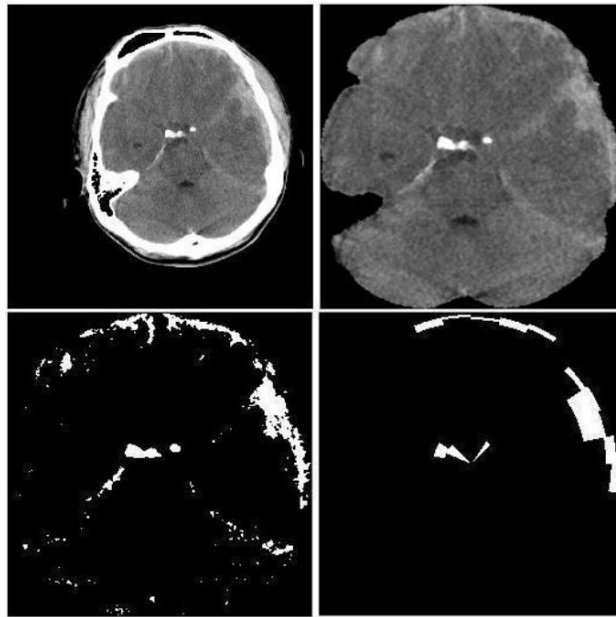
Figure 4.7: Constructing binary feature vector from TBI CT image: example 1

4.3 Medical image retrieval based on low level visual features

We process the TBI images in our database and represent each image using a binary vector as described in the previous section. When we have an image query, we process the query image in the same way as we process the images in the database, then compute the similarity of the query image to each image in the database and retrieve the images according to the similarity measure.

Given the query image Q and reference image I, we measure the similarity of the two images by computing the Jaccard similarity coefficient of the u and v, the two binary vectors that represent Q and I respectively:
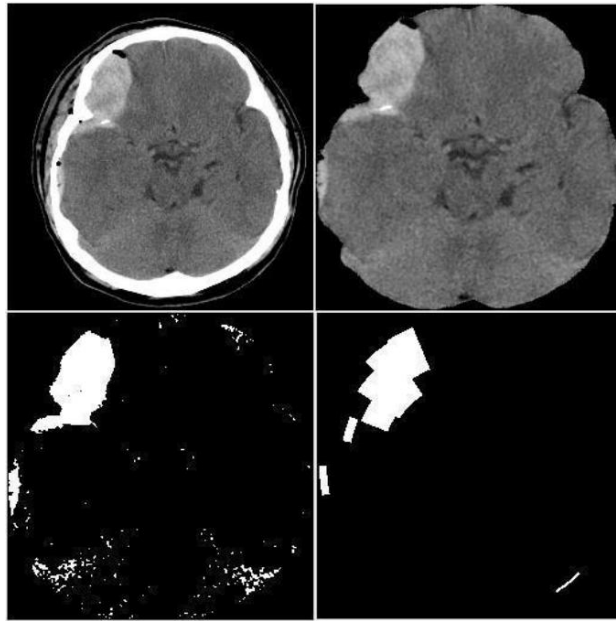
Figure 4.8: Constructing binary feature vector from TBI CT image: example 2

$$Sim(Q, I) = J(u, v). \tag{4.1}$$

The Jaccard similarity coefficient is a useful measure of the overlap of u and v:

$$J(u, v) = \frac{|u \cap v|}{|u \cup v|}. \tag{4.2}$$

Each element (or bin value in this case) of u and v can either be 0 or 1. The total numbers of different combinations of elements for both u and v are defined as follows:

- s_{11} represents the total number of elements where u and v both have a value

of 1.

- s_{01} represents the total number of elements where the element of u is 0 and the element of v is 1.

- s_{10} represents the total number of elements where the element of u is 1 and the element of v is 0.

- s_{00} represents the total number of elements where u and v both have a value of 0.

Each element must fall into one of these four categories, so:

$$s_{11} + s_{01} + s_{10} + s_{00} = n. \tag{4.3}$$

The Jaccard similarity coefficient then can be computed as

$$J(u, v) = \frac{s_{11}}{s_{01} + s_{10} + s_{11}}. \tag{4.4}$$

For 3D volumetric cases such as CT image series, we measure the similarity between each pair of corresponding images and use a weighted sum to represent the similarity value of the two cases. Given a query image series $S_Q = \{Q_1, ..., Q_N\}$ and a reference image series $S_I = \{I_1, ..., I_N\}$, we propose a 3D similarity measure as follows:

$$Sim(S_Q, S_I) = \frac{\sum_{i=1}^{N} w_i Sim(Q_i, I_i)}{\sum_{i=1}^{N} w_i} \tag{4.5}$$

where w_i is the weight given to the ith slice, and $0 \leq w_i \leq 1$. In TBI cases, pathological features are contained mainly in the middle slices. For example, in a 21-slice scan series, slices from 7 to 16 contain the most features. Therefore, we give these middle slices higher weights than the beginning and ending slices.

In clinical practice, The user of the medical image retrieval system may consider pathology features in the symmetrical positions up to certain rotational angles in the images as equally relevant. For example, an image with a SDH in left frontal lobe may be considered very relevant to a query image with a SDH in the right frontal lobe, though the similarity computed may be 0 as the two SDH regions in respective images do not overlap. In this case, we need to perform a flipping transformation on the query image and compare the flipped image with the ones in the indexed database. The horizontal flipping transformation of a query image Q_i is defined as:

$$g_{flip}(Q_i) = \forall P_{xy} \in Q_i, g_{flip-point}(P_{xy}), \tag{4.6}$$

where P_{xy} is a point in Q_i with cartesian coordinates of x (along vertical axis) and y (along horizontal axis). The point flipping function $g_{flip-point}$ is defined as:

$$g_{flip-point}(P_{xy}) = P_{(H-x)(W-y)}, \tag{4.7}$$

where H and W are the height and width of image Q_i respectively.

For another instance, an image with a SDH in right temporal lobe may be considered moderately relevant to a query image with a SDH in the right frontal

89

lobe as right temporal lobe is below the right frontal lobe in a 2D head CT scan. In this case, we may need to do a rotation transformation on the query image first so that indexed images with hematomas in nearby anatomical regions are compared to the query image properly. The rotation transformation of a query image Q_i is defined as:

$$g_{rotate}(Q_i) = \forall P_{rt} \in Q_i, g_{rotate-point}(P_{rt}), \tag{4.8}$$

where P_{rt} is a point in Q_i with polar coordinates of r (radius) and t (angle). The point rotation function $g_{rotate-point}$ is defined as:

$$g_{rotate-point}(P_{rt}) = P_{r(t+a)}, \tag{4.9}$$

where a is the rotation angle applied to image Q_i.

We generalize the geometric transformation of query image Q_i as $g(.)$, and propose the transformed 3D similarity measure for a series of images as in Equation 4.10. We leave it to the user to choose to enable or disable the flipping/rotation search function.

$$Sim(S_Q, S_I) = \frac{\sum_{i=1}^{N} w_i Sim(g(Q_i), I_i)}{\sum_{i=1}^{N} w_i} \tag{4.10}$$

After computing the similarity between the query image series and all image series in the database, we retrieve the ones whose similarity measure to S_Q is above a certain threshold. We also rank the retrieved images according the similarity values–image series that are more similar to S_Q will be put at the beginning

of the retrieved list.

4.4 Experiment

4.4.1 Data set

We build our image database from 500 TBI study cases obtained from National Neuroscience Institute, Tan Tock Seng Hospital. The major pathological classes in TBI include extradural hematoma (EDH), subdural hematoma (SDH), subarachnoid hemorrhage (SAH), intracerebral hemorrhage (ICH), and intraventricular hemorrhage (IVH). We evaluate the system performance based on 30 test queries by CT case example. The system responses a list of relevant cases ranked according to the 3D similarity measure for each query and shows the summary information of the retrieved relevant cases (with 3D similarity measure threshold setting to 0.08).

4.4.2 Evaluation metric

For each of images series S_i retrieved from the database for each query case S_Q, a relevance grade from 0-3 is given by a human expert:

- 0-not relevant.

- 1-slightly relevant.

- 2-moderately relevant.

- 3-very relevant.

We use the Normalized Discounted Cumulative Gain (NDCG) [62] to measure the effectiveness of our content based medical image retrieval system. Cumulative Gain (CG) is the sum of the graded relevance values of all results in a search result list. The CG at a particular rank position p is defined as:

$$CG_p = \sum_{i=1}^{p} rel_i,$$

(4.11)

where rel_i is the graded relevance of the result at position i. As it is assumed that highly relevant documents are more useful when appearing earlier in a search engine result list (have higher ranks) and highly relevant documents are more useful than marginally relevant documents, which are in turn more useful than irrelevant documents; Discounted Cumulative Gain (DCG) is used so that highly relevant images appearing lower in a search result list are penalized as the graded relevance value is reduced logarithmically proportional to the position of the result. The discounted CG accumulated at a particular rank position p is defined as:

$$DCG_p = \sum_{i=1}^{p} \frac{2^{rel_i} - 1}{\log_2(1 + i)}$$

(4.12)

As search result lists vary in length depending on the query, a search engine's performance for retrieval results of different list lengths cannot be compared fairly using only DCG, so the cumulative gain at each position for a chosen value of p should be normalized across queries of different retrieval result list lengths. The image cases of a result list are sorted by relevance to produce an ideal DCG at

position p; and for a query, the normalized discounted cumulative gain, or NDCG, is computed as:

$$NDCG_p = \frac{DCG_p}{IDCGp} \qquad (4.13)$$

The NDCG values for all queries can be averaged to obtain a measure of the average performance of a search engine's ranking algorithm. NDCG is 1 if the retrieval results are perfectly ranked by relevance.

4.4.3 Result

NDCG evaluation result is shown in table 4.1. Figure 4.10 and 4.12 show two image query and retrieval cases. The retrieved images are ranked by similarity to the query image.

Table 4.1: NDCG evaluation result for content based medical image retrieval

Average NDCG for 30 test queries	0.83
Lowest NDCG among 30 test queries	0.66
Highest NDCG among 30 test queries	0.98

4.5 Discussion

The content-based medical image retrieval framework we proposed in this chapter has the advantage of being able to retrieve images according to their visual appearance without accompanying texts. This mode of image retrieval enables users to input an image and query visually similar images. The method we propose is fast
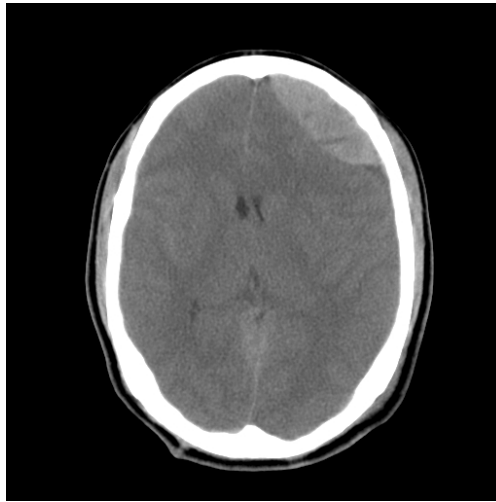
Figure 4.9: Query example 1

and can be built into real time online CBIR systems. This function could be particular useful as a teaching tool in junior medical staff training. The instructor could input a medical image to retrieval all visually similar images, then the instructor could compare the images and tell the students the subtle differences among them. The CBIR is also useful when a medical professional acquires a new image and wants to reference similar cases in the database before making the final judgment of the type and severity of the disease revealed by the newly acquired image.

However, as only low level visual features are used, they give certain image semantics but these semantics are not deep enough. It is well known that such an approach suffers from the semantic gap problem [27]. Similar images retrieved by low level features may be different in high level image semantics. For example, two hematoma segments having similar circular bin binary representation may be of different pathologies. Specifically in the case of traumatic brain injury, a CT

image contains EDH may have similar feature vector with a CT image contains SDH if the hemorrhage regions are in the similar location, even though EDH and SDH are two different pathology concepts and have different impacts on patients. Due to this problem, the retrieved cases may be semantically different from the query case. To solve this problem, we will propose an annotation-based medical image retrieval framework in the following chapters that investigates deeper image semantics.
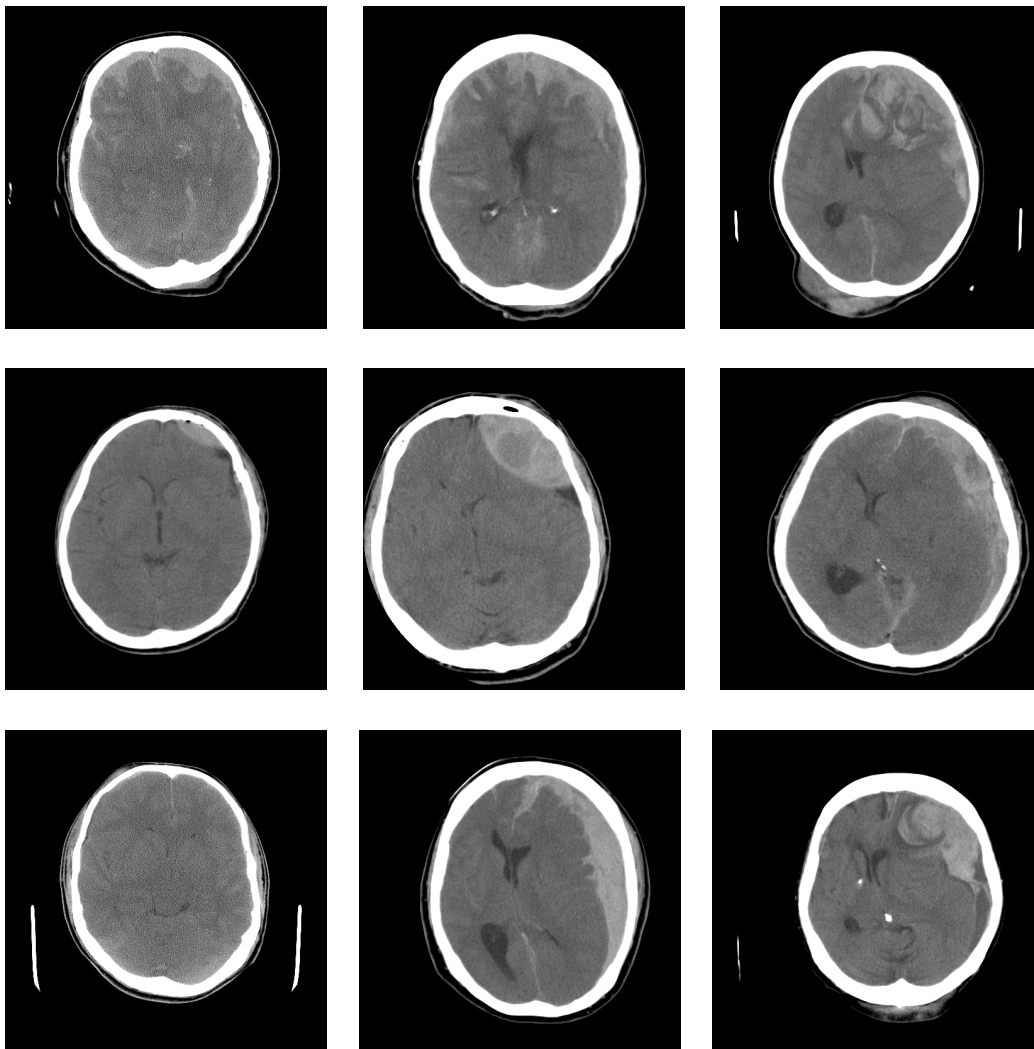
Figure 4.10: Image retrieval results of query example 1
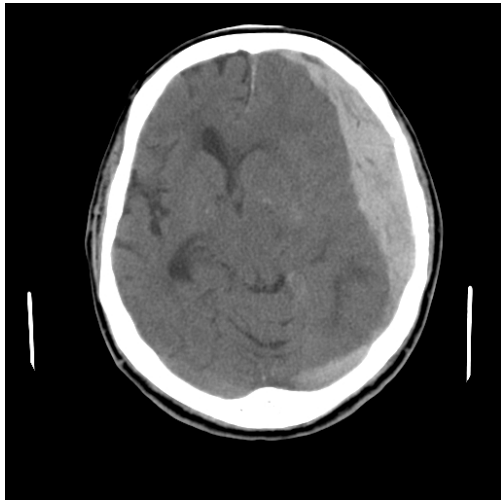
Figure 4.11: Query example 2
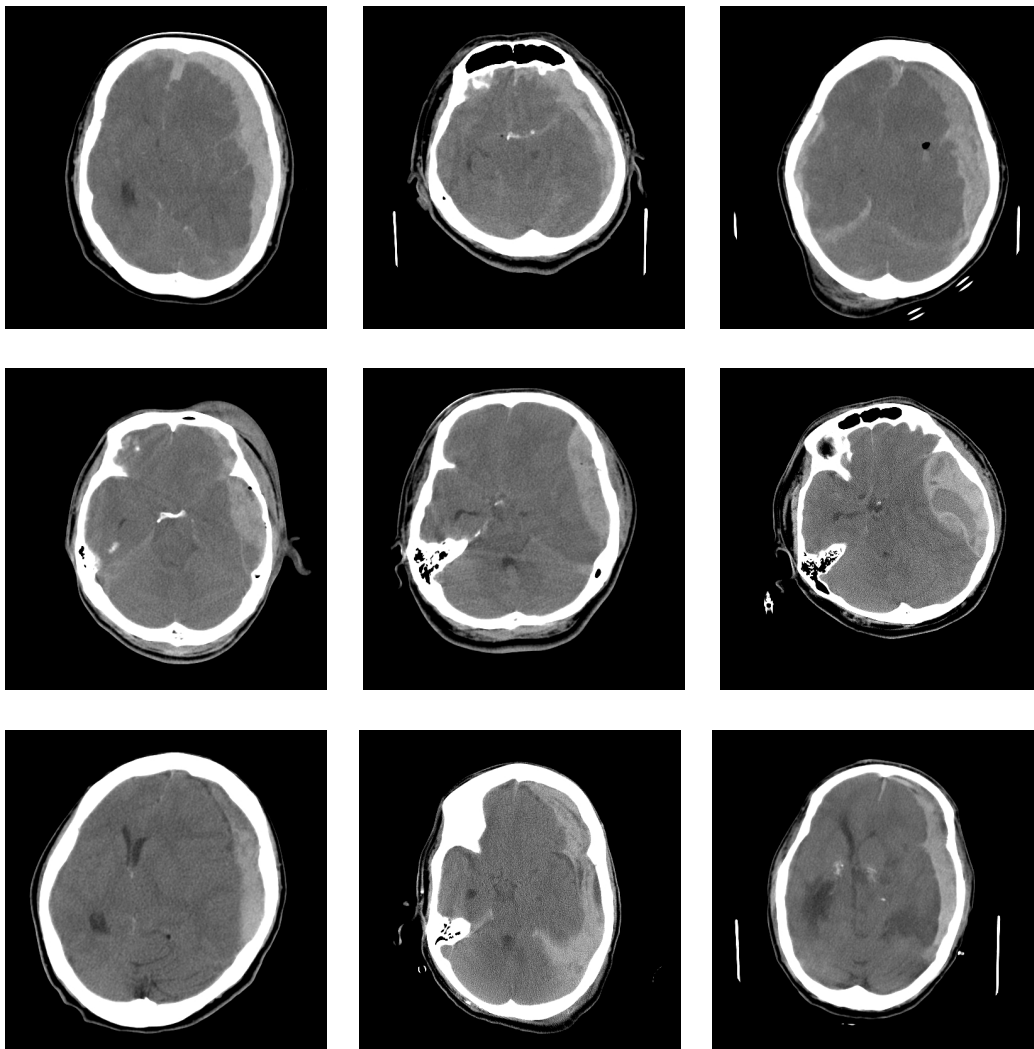
Figure 4.12: Image retrieval results of query example 2

Chapter 5

Automatic medical image annotation framework using probabilistic models

Numerous tasks have been proposed in CLEF medical image annotation tracks [119] in recent years. The tasks focus on automatically generating annotations of acquisition modality (CT, X-ray, MR, etc.), body orientation, body region, and biological system. These works adopt various approaches to provide a solution to the medical image retrieval problem.

The traditional approach – text-based image retrieval systems (TBIR) index the images with associated keywords. It is an early approach with the advantage of easy implementation and fast retrieval, but it requires a large amount of manual work to label the images and suffers from human subjectivity. In Chapter 3 we

developed a method to extract keywords from the radiology reports associated with the medical images and can use these keywords to index and retrieve the images. However, for the vast amount of images without such accompanying textual reports, text-based image retrieval cannot be applied.

Content-based image retrieval (CBIR) methods such as the one we developed in Chapter 4 provides a solution to overcome the disadvantage of TBIR by retrieving medical images that has similar low level visual features to the query image's. CBIR provides a practical solution for medical image retrieval; however, it poses a limitation on the query format – the query must be an image example. Moreover, it suffers from the semantic gap problem.

Auto-annotation based image retrieval (ABIR) seems to have the advantages of both TBIR and CBIR by having a preprocess of automatically annotating images with their semantic content and offering users the ease of using text to search images. Images and their corresponding semantic labels are associated by machine learning methods. However, a large labeled training data set is needed in the process.

For anatomical annotation, such training set is easy to obtain since the labels can be extracted from DICOM header[1] and global visual features can be used for the task. However, for pathology annotation, it is not easy to obtain semantic labels for the training data. Another common difficulty researchers often encounter is the lack of large number of labeled data for training purpose, as manually pro-

---

[1]A DICOM file contains both a header storing meta information (e.g. patient's data, modality of scan, image dimensions, etc), and the image data.

cessing and labeling medical images requires expertise and is thus expensive and slow.

To overcome this limitation, in this chapter, we propose a novel framework for automatic pathology annotation using both medical images and their associated radiology reports. Under our new framework, training images can be automatically labeled by text mining from associated radiology reports. We use an existing probabilistic model to learn the image-annotation relation as we have described in our paper [47], and we propose a new semantic similarity language model to learn the intra-annotation probability to improve the overall image annotation accuracy as described in our paper [46].

5.1 The framework

Unlike image annotation in general domain, the medical images are usually not explicitly labeled with keywords but only accompanied by a radiology report in some cases. Therefore, we build our training corpus from those images with reports and extract the pathology terms from the report as annotation keywords. As shown in Figure 5.1, in training phase, we use probabilistic models to learn the correlations between the regions of interest (ROIs) and the pathology terms extracted from reports associated with the images and generate a ROI-annotation conditional or joint probability table. The ROIs are segmented from the images using the same method as we described in Chapter 4. The free text reports are processed and pathology terms are extracted using the method described in Chapter

3. In testing phase, we segment the ROIs from unannotated images, then we use the probability table generated from the training phase to find out corresponding pathology annotations that maximizes the probability.
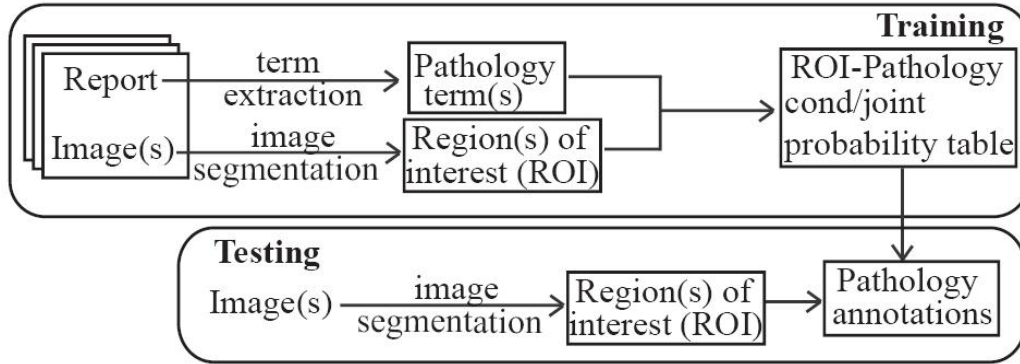
Figure 5.1: The framework of automatic medical image annotation using probabilistic models

5.2 Probabilistic models

The problem of automatic image annotation can be defined as: given a training corpus T consisting of some already annotated images, which annotations A will be used to label a new image I? For most image annotation approaches using probabilistic models, the goal is to find A that maximizes the conditional probability $p(A|I)$:

$$A = \arg\max_{A} p(A|I) \tag{5.1}$$

where A is a set of words $\{w_1, \ldots, w_n\}$ used for annotation and image I is usually

represented by a set of features $\{f_1, \ldots, f_k\}$ or a set of blobs $\{b_1, \ldots, b_m\}$.

$p(A|I)$ can also be rewritten as:

$$p(A|I) = \frac{p(A, I)}{p(I)} \tag{5.2}$$

Since the prior probability of a given image I is usually considered as of uniform distribution, instead of estimating the conditional probability $p(A|I)$ directly, some image annotation approaches use the joint probability $p(A, I)$ to find the best annotation set A:

$$A = \arg\max_A p(A, I) \tag{5.3}$$

In the following sections, we will use the above equations as basis for automatic medical image annotation.

5.2.1 Statistical machine translation model

Statistical Machine Translation Model (SMT) [8] was first used in automatic image annotation in general domain by Duygulu et. al. in [38]. Automatic medical image annotation can also be viewed as a statistical machine translation problem due to many analogies between images and languages. Medical images and the corresponding pathology annotations are two different media referring to the same semantic content–the region of interest. We take medical images as the source language and the pathology annotations as the target language, and use SMT methods to "translate" the abnormal regions in medical images to pathology

annotations. For medical images taken in the radiographic examinations, we need to preprocess the images to extract out the regions of interest (ROIs), before using the ROIs as the translation input. We use radiology reports accompanying the scans as annotation result to train the automatic annotation system. We extract the pathology terms describing the ROIs from the reports and use these terms as target language. Then we use IBM Model 1 [9] to train the ROI-pathology translation table using the training corpus. IBM Models [9] are classic translation models for SMT. Since we use this model for the translation of a set of image segments (the ROIs) and a set of annotation words instead of sentences in natural languages, we assume the ordering or the positions of the individual elements in the ROI set and the annotation word set do not affect the probablity for them to be aligned. Thus, we use IBM Model 1 for the training instead of the higher versions of the IBM Models. In testing phase, we first process the testing images to segment out ROIs, then use the trained ROI-pathology translation table to output the pathology terms associated with each ROI.

In training, we collect all the ROIs (hematoma regions) segmented from all scans in the training corpus, extract features of the ROIs, and cluster them into groups. Unlike image regions in general domain which mainly use color and texture as features, the intensity and texture for hematoma regions are very similar. After segmentation, shape and location features are particularly important to distinguish the hematomas among themselves. We use the following features for hematoma clustering:

- eccentricity–the ratio of the distance between the foci of the ellipse and its

major axis length;

- solidity–the proportion of the pixels in the convex hull that are also in the region;

- extent–the proportion of the pixels in the bounding box that are also in the region;

- skull–whether the ROI is adjacent to skull or not.

After the features of the ROIs segmented from training images are extracted, we use K-means clustering to partition the ROIs into different clusters according to their features. Each cluster consists of visually similar ROIs and is considered as a "word" in the "dictionary" of ROIs. Each ROI categorized in that cluster will be treated as the same "word" that cluster represents. After the images and texts are preprocessed and basic elements for translation are defined and extracted (ROI clusters and pathology terms), we can start the training process. We use the translation model of IBM Model 1 [9] to get the word ROI-to-pathology alignments for training data as shown in Figure 5.2, and thus build the ROI-to-pathology translation table. The translation table contains conditional probability for each pathology annotation word w given each ROI cluster r:

$$p(w|r) = \sum_a p(a, w|r), \tag{5.4}$$

by calculating the probability of possible alignments a between w and r:

$$p(a, w|r) = \prod_{j=1} p(w_j|r_{a_j}), \tag{5.5}$$

where r_{a_j} is the ROI corresponding to annotation word w through alignment a. As we can see from the two equations, finding out the best alignment and the best translation is a chicken-and-egg problem; therefore, EM algorithm is a natural choice to solve it. We use the parameter estimation method in [9] to obtain the best alignments as well as the translation table. The alignments between ROI and pathology annotations are initially of equal weight as the translation table is initialized uniformly. As we iteratively collect fractional counts from the training data and update the translation probabilities and the alignments, we can get the best alignments as the values in the translation table converge as illustrated in Figure 5.2.
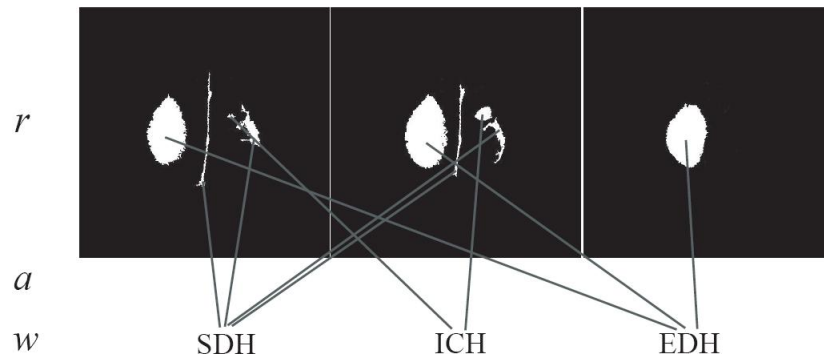
Figure 5.2: Alignments between ROIs and pathology annotations

1:initialize the translation table uniformly.

2:calculate $p(a|w, r)$ using chain rule: $p(a|w, r) = p(a, w|r)/p(w|r)$.

3:use the updated values to collect counts and re-estimate the model.

4:check if the values converge–if not, go to Step 2 and Step 3 to repeat the iteration.

For single medical image in testing phase, we use the trained conditional probability ROI-to-pathology translation table to find the pathology annotation word with the highest translation probability for each individual 2D ROI segmented from the testing image:

$$w = \arg\max_{w} p(w|r) \tag{5.6}$$

For 3D volumetric case (such as CT and MRI) which consists of a series of 2D scan sequences, one 3D volumetric region of interest may be captured in several consecutive 2D scans and form different 2D ROIs in the individual scans. For example in Figure 5.2, the same clot of epidural hematoma (EDH) spreads over the consecutive images and form one 2D ROI in each of the three images. However, it is possible that due to different features of each 2D ROI, they may be clustered into different clusters and may be annotated differently. To solve this conflict of different annotations for the same 3D ROI, we use equation 5.7 to annotate the 3D volumetric ROI captured in n consecutive 2D scans. In Equation 5.7, the probability of annotation word w given a 3D ROI is the sum of the the probability of w given n individual 2D ROI r_i that comprises the 3D ROI r.

$$w = \arg\max_{w} \sum_{i=1}^{n} \frac{1}{n} p(w|r_i) \qquad (5.7)$$

5.2.2 Cross-media relevance model

We use the same text and image preprocessing methods as in the previous section to extract the pathology terms from the radiology report and ROIs from the medical images. We then use the cross-medial relevance model (CMRM) to learn the relation between them. CMRM was first proposed by Jeon et. al. in [63] for automatic image annotation in general domain. CMRM defines each training instance J in the training corpus T as $J = \{b_1 \ldots b_m; w_1 \ldots w_n\}$, where $b_1 \ldots b_m$ represent the blobs corresponding to regions of the image and $w_1 \ldots w_m$ represent the words in the image annotation. As CMRM uses a relevance language model [75], the joint probability of observing the word w_i and the blobs $b_1 \ldots b_m$ in the same image is estimated as the expectation over the images J in the training set:

$$p(w_i, b_1, \ldots, b_m) = \sum_{J \in T} p(J) p(w_i, b_1, \ldots, b_m | J) \qquad (5.8)$$

It is assumed that the events of observing w_i and $b_1 \ldots b_m$ are mutually independent and identically distributed for an image. Equation 5.8 can be rewritten as follows:

$$p(w_i, b_1, \ldots, b_m) = \sum_{J \in T} p(J) p(w_i | J) \prod_{j=1}^{m} p(b_j | J) \qquad (5.9)$$

The prior probabilities $p(J)$ is set to be uniform over all images in T. S-

moothed maximum-likelihood estimates is used for the probabilities in Equation 5.9:

$$p(w_i|J) = (1 - \alpha_J)\frac{|w_i \, in \, J|}{|J|} + \alpha_J\frac{|w_i \, in \, T|}{|T|} \tag{5.10}$$

$$p(b_j|J) = (1 - \beta_J)\frac{|b_j \, in \, J|}{|J|} + \beta_J\frac{|b_j \, in \, T|}{|T|} \tag{5.11}$$

where $|w_i \, in \, J|$ is the number of times the word w_i occurs in the annotation of image J and $|w_i \, in \, T|$ is the total number of times w_i occurs in the annotations of all images in the training set T. Similarly, $|b_j \, in \, J|$ is the number of times the blob b_j occurs in image J and $|b_j \, in \, T|$ is the total number of times b_j occurs in all images in the training set T. $|J|$ is the total count of all words and blobs occurring in image J, and $|T|$ is the total size of the training set. The smoothing parameters α_J and β_J determine the degree of interpolation between the maximum likelihood estimates and the background probabilities for the words and the blobs. From Jeon et. al. [63]'s experiment, the model gives best annotation result when $\alpha_J = 0.1$ and $\beta_J = 0.9$.

5.3 Language model enhancement

5.3.1 A semantic similarity language model

Under probabilistic modeling, the best word set to annotate a given image can be found by maximizing the conditional probability $p(A|I)$ or the joint probability

$p(A, I)$ (recall Equation 5.1 and Equation 5.3 in Section 5.2). As many existing probabilistic modeling approaches for automatic image annotation emphasize on the image-word correlation and directly maximize the conditional or joint probability, the contextual information among the annotation words themselves are usually ignored.

Inspired by the noisy channel model, we use an indirect approach to maximize the desired probability, make use of the word contextual information by incorporating a language model as we have described in our paper in [46]. We view the image annotation problem using the noisy channel model as in Figure 5.3.

Figure 5.3: Noisy channel model

The original signal A generated by the transmitter passes through a noisy channel and changes to a noisy signal I which will be received by the receiver. Using the noisy channel model, we can interpret the image annotation problem as: a person wants to express a few objects in words, but the output of the expression is a picture because the tool he uses (maybe a camera); we have to use the image to predict what were the original words the person was trying the say in his picture using:

$$p(A|I) = \frac{p(I|A)P(A)}{p(I)} \tag{5.12}$$

$$A = \arg\max_A p(I|A)p(A) \qquad (5.13)$$

We then find the best set of annotation words by maximizing the combination of two probabilities, $p(I|A)$ and $p(A)$. $p(I|A)$ can be derived from the original image annotation methods; and $p(A)$ is the annotation set probability derived from the language model. Inspired by the statistical machine translation model (SMT) which maximizes the posterior probability and sets different weights to the probability factors, we also add different weights to the two probabilities to give different emphasis on the original annotation model and the language model to achieve best overall annotation result:

$$A = \arg\max_A p(I|A)^{\lambda_1} p(A)^{\lambda_2}, \qquad (5.14)$$

which can also be expressed in log linear form:

$$A = \arg\max_A (\lambda_1 \log p(I|A) + \lambda_2 \log p(A)) \qquad (5.15)$$

Language modeling is widely used in many natural language processing applications such as speech recognition, machine translation, part-of-speech tagging, parsing and information retrieval. A statistical language model assigns a probability to a sequence of words. Using a language model in statistical machine translation boosts the probability of translating sentence in source language into well-formed sentence in target language. Similarly, we use a language model in image annotation to boosts the probability of annotating semantically coherent

words.

As the set of annotation words in automatic image annotation task is not a sequence as the annotation is not a sentence; therefore, the commonly used bi-gram or tri-gram model is not suitable to model the the probability of a set of words $p(A)$ in image annotation task. Alternatively, we make use of the word co-occurrence information in the annotations in the training corpus to model the probability of the keyword given other keywords in the annotations and the probability of the annotation. In this section, we choose semantic vector model to represent each word and define the language model to be the average pairwise similarity of the semantic vector of each word in the annotation set.

In the semantic vector model, the meaning of each word is represented in terms of vectors of other context words. We first choose a set of words as context words to be included in the semantic vector to represent the meaning of any word. For a small image annotation corpus such as COREL 5K dataset [38], the vocabulary size (the total number of distinct words used in the annotations) is usually small as well (in a few hundred); we could use all the words in the vocabulary as context words. A word w is represented by $\mathbf{v}$, a semantic vector $< v_1, v_2, \ldots, v_i, \ldots, v_m >$, where there are m context words in the vocabulary. In [90, 91], the semantic vectors are based on components defined as the ratio of the conditional probability of a context word given the target word to the overall probability of the context word. We follow the definition in [90, 91] to calculate each component v_i in the semantic vector that represents w:

Table 5.1: Example of annotation words represented by semantic vectors of context words

	city	sky	sun	water	clouds	tree	...
grass	<0.53	2.86	0.70	2.16	0.84	4.24	...>
buildings	<20.29	5.15	0.77	2.71	2.14	3.41	...>
bridge	<5.77	2.77	2.55	8.21	1.01	2.41	...>
mountain	<2.31	7.71	3.07	4.27	12.20	5.80	...>

$$v_i = \frac{p(context_i|w)}{p(context_i)} \tag{5.16}$$

And the the conditional probability is just the relative frequency of the count of the co-occurrences of context word $context_i$ and word w in all annotations over the total number of occurrences of w in the annotations:

$$p(context_i|w) = \frac{count(context_i, w)}{count(w)} \tag{5.17}$$

The semantic vector v represents the distributional properties of a word w in terms of the strength of its co-occurrence with a set of context words. Dividing by the overall probability of each context word prevents the vectors being dominated by the most frequent context words, which will often also have the highest conditional probabilities. Table 5.1 shows some examples of annotation words represented by semantic vectors of context words. In this section, we use data from general domain to illustrate the language model, as it is easier and more intuitive for the readers to understand than medical images and terms.

Assume that words should be semantically similar with the set of contex-

t words in the same annotation, the probability of a set of annotation words $A = \{w_1, \ldots, w_n\}$ can be measured by the similarities of each annotation word to all other words:

$$p(A) \propto \frac{1}{n(n-1)} \sum_{w_i \in A} \sum_{w_j \in A, j \neq i} sim(w_i, w_j) \qquad (5.18)$$

where each word w_i from the annotation set A is represented by its corresponding semantic vector. Similarity can be measured using cosine:

$$sim(w_1, w_2) = \frac{w_1 \cdot w_2}{\|w_1\| \|w_2\|} \qquad (5.19)$$

Using the semantic vector representation stated in Equation 5.16, the dot product in similarity measure is calculated as:

$$w_1 \cdot w_2 = \sum_{i=1}^{m} v_{w_1,i} v_{w_2,i} = \sum_{i=1}^{m} \frac{p(c_i|w_1)}{p(c_i)} \frac{p(c_i|w_2)}{p(c_i)} \qquad (5.20)$$

Table 5.2 shows the pairwise semantic similarity between some example annotation words.

We found that Equation 5.13 is very similar to the fundamental equation statistical machine translation [8] as shown below (translating a foreign sentence f to an English sentence e):

$$e = \arg\max_e p(f|e)p(e) \qquad (5.21)$$

As the search space for the optimal solution is huge, beam search is usually

Table 5.2: Examples of pairwise semantic similarity

	people	sun	street	sky	forest	tree	...
people	1	0.0219	0.4131	0.1513	0.0174	0.2118	...
sun	0.0219	1	0.0120	0.2121	0.0001	0.0827	...
street	0.4131	0.0120	1	0.1833	0.0012	0.1624	...
sky	0.1513	0.2121	0.1833	1	0.0283	0.4211	...
forest	0.0174	0.0001	0.0012	0.0283	1	0.2140	...
tree	0.2118	0.0827	0.1624	0.4211	0.2140	1	...
...	...	...	...	...	...	...	...

used for decoding in statistical machine translation. Similarly, we also use k-best beam search to find the k-best set of annotation words.

5.3.2 Improved statistical machine translation model

To apply our proposed language model to the translation model, we first rebuild the translation model by reversing the translation direction - we treat words as the source language and blobs as the target language. We construct the translation table–the word-to-blob translation table instead of the original blob-to-word translation table. In the new table, each entry $p(b_j|w_i)$ indicates the translation probability from word w_i to blob b_j. The word-to-blob translation probability and alignment probability can be calculated in the equations below respectively.

$$P(b|w) = \sum_a p(a, w|b), \tag{5.22}$$

$$P(a, b|w) = \sum_{j=1} t(b_j|w_{a_j}), \qquad (5.23)$$

where w is the word, b is the blob, and a is a possible alignment between w and b. Finding out the best alignment and the best translation is a chicken-and-egg problem; EM algorithm is a natural choice to solve this problem. Parameter estimation method in [9] is used to obtain the the best alignments as well as the translation table:

Step 1: initialize the blob-to-word translation probability uniformly.

Step 2: apply the model to the training data to calculate the alignment probability $p(a|w, b)$ using chain rule: $p(a|w, b) = P(a, b|w)/p(b|w)$.

Step 3: use the updated values to collect counts and re-estimate the model.

Step 4: check if the values converge–if not, repeat Step 2 and Step 3.

Figure 5.4: EM algorithm to estimate word-to-blob translation and alignment probabilities

After the reverse translation table is built, we find the set of annotation words A for a new image I by Equation 5.1, where $p(A|I)$ can be estimated as:

$$p(A|I) \propto p(A) \prod_{w_i \in A} p(b_i|w_i) \qquad (5.24)$$

where w_i is one annotation word in A, and b_i is the blob translated from the word w_i. Applying the semantic similarity language model defined in Equation 5.15, we find the set of annotation words by:

$$A = \arg\max_A (\lambda_1 \log \prod_{w_k \in A} p(b_k|w_k)$$
$$+ \lambda_2 \log \frac{1}{n(n-1)} \sum_{w_i \in A} \sum_{w_j \in A, j \neq i} sim(w_i, w_j)) \quad (5.25)$$

5.3.3 Improved cross-media relevance model

After generating joint probability from the original CMRM, in order to apply the language model to improve the original model, we first find the conditional probability $p(I|w_i)$ (where $I = \{b_1, \ldots, b_m\}$ from the joint probability $p(w_i, I)$:

$$p(I|w_i) = \frac{p(w_i, I)}{p(w_i)} = \frac{p(w_i, b_1, \ldots, b_m)}{p(w_i)} \quad (5.26)$$

We calculate the prior probability of a word $p(w_i)$ as the count of w_i in training corpus T over the total count of all words used in annotation in T:

$$p(w_i) = \frac{|w_i|}{\sum_{w_k \in T} |w_k|} \quad (5.27)$$

Then we find the set of annotation words A for a new image I by Equation 5.1, where $p(A|I)$ can be estimated as:

$$p(A|I) \propto p(A) \prod_{w_i \in A} p(I|w_i) \quad (5.28)$$

Applying the semantic similarity language model defined in Equation 5.18, we can find the set of annotation words by:

$$A = \arg\max_{A}(\lambda_1 \log \prod_{w_k \in A} p(I|w_k)$$
$$+ \lambda_2 \log \frac{1}{n(n-1)} \sum_{w_i \in A} \sum_{w_j \in A, j \neq i} sim(w_i, w_j)) \qquad (5.29)$$

5.4 Experiments

5.4.1 Data set

Using TBI CT image data set

For experiment, we obtained 500 brain CT examination cases from National Neuroscience Institute (NNI), Singapore. Each examination consists of a series of CT scans and a radiology report. We used 450 cases to build the translation table in training and 50 cases for annotation testing. As the scans are too sparse (with 5mm or 7mm in-between scan distance), we did not reconstruct the images in 3D but used 2D ROIs and clustered them according to their 2D features. For testing images, we use Equation 5.7 to find the best annotation for each ROI.

Using general domain image data set

The COREL 5K image corpus [38] is a publicly available and widely used dataset in evaluating image annotation methods. It contains 5000 images from 50 themes with 100 images from each theme. Each image is segmented into 1 to 10 regions using Normalized Cut [109]. 36 visual features including color, texture, and shape

are extracted for each image region. All image regions are grouped into 500 visual blobs using K-Means clustering on the 36 features. Each image is annotated with 1 to 5 words. A total number of 374 words are used to annotate the entire dataset. The dataset is partitioned into training set with 4500 images and testing set with 500 images. We use 4000 images in the training set to train the improved models and use 500 images in the training set as a validation set to tune the weight parameters λ_1 and λ_2. After the best parameter setting is determined, we use all 4500 training images to train the improved models again and test the models on the 500 testing images. There are 263 distinct words for the testing set. In order to provide a valid comparison with related work, we conducted the experiments on the COREL 5K dataset using the same visual features and visual blob clustering.

5.4.2 Evaluation Metrics

As the main motivation for automatic image annotation is for annotation-based image retrieval, it is natural to use the retrieval metrics to reflect the performance of the image annotation system. In most research works on automatic image annotation, the precision and recall is measured through the process of retrieving testing images with single keyword as shown below:

$$precision(w) = \frac{tp(w)}{tp(w) + fp(w)} \tag{5.30}$$

$$recall(w) = \frac{tp(w)}{tp(w) + fn(w)} \tag{5.31}$$

119

where $tp(w)$ is the number of correctly retrieved images, $fp(w)$ is the number of incorrectly retrieved images, and $fn(w)$ is the number of relevant images not retrieved. The $precision(w)$ measures the correctness in annotating images with word w and the $recall(w)$ measures the completeness in annotating images with word w.

In addition, the number of words with non-zero recall, i.e. the number of single-word queries for which at least one relevant image can be retrieved using the automatic annotation, is also an important metric, because it indicates the range of words that contribute to the average precision and recall and a biased model can achieve high precision and recall by performing well only on a small number of words commonly used in annotation.

5.4.3 Results

Results of experiments on TBI CT image data set

One of the main motivations for automatic image annotation is for annotation-based image retrieval; therefore, it is natural to use the retrieval result to reflect the performance of the image annotation system. In most research works on automatic image annotation, the accuracy and recall is measured through the process of retrieving testing images with individual keyword. For our system performance evaluation, we used the trained translation table to annotate the 50 testing images and then used pathology annotations as keywords to retrieve the testing images. The evaluation result is shown in Table 5.3. The average retrieval precision and

recall for all pathology terms are 76.2% and 72.0% respectively. The retrieval results of pathologies that are more frequent in the training corpus such as "EDH", are better than those appear less frequent such as "IVH". Figure 5.5 shows some brain CT image annotation result. We reported our method the result in [47] as well. The performance of the system measured by medical image retrieval precision and recall is encouraging to the doctors in NNI.

Table 5.3: Evaluation results (in %)

	SDH	EDH	SAH	ICH	IVH
precision	74.1	91.7	72.0	68.4	75.0
recall	74.1	84.6	69.2	72.2	60.0
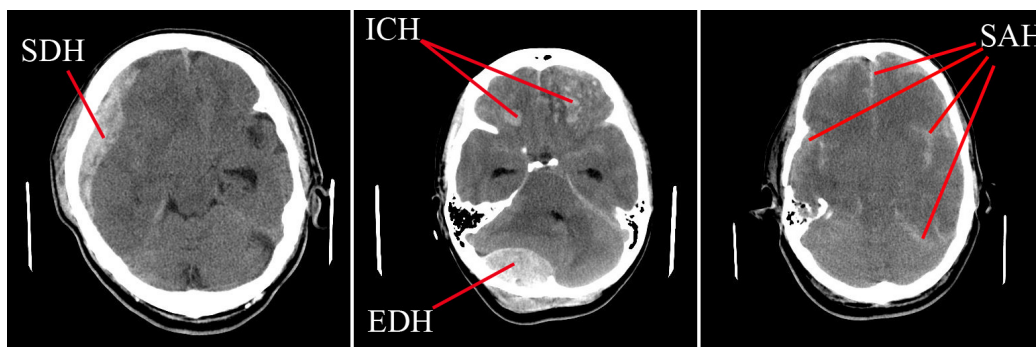
Figure 5.5: Annotation results of some brain CT images

Results of experiments on general domain image data set

For valid comparison with related work, we fixed the number of words to annotate each image to be 5, as the experiment results for related work are available for number of words fixed at 5. From the result of parameter tuning on validation set, we found that the best annotation result for improved statistical machine

Table 5.4: Evaluation results on single keyword retrieval

	on all 263 testing words			on 98 testing words	
	# words	precision	recall	precision	recall
SMT	49	4.0%	6.0%	9.9%	12.9%
SMT + SSLM	65	6.5%	8.6%	16.1%	18.5%
CMRM	66	9.0%	10.0%	22.0%	25.1%
CMRM + SSLM	90	10.5%	13.1%	25.7%	33.0%

translation model (SMT) [38] is achieved when $\lambda_1 = \lambda_2$. For improved Cross Media Relevance Model (CMRM) [63], the system performs best when $\lambda_1 = 4\lambda_2$. We performed single-word queries to retrieve images using auto annotations for all 263 words in the testing set. The number of words with non-zero recall ("# words" for short) for statistical machine translation model is 49. With the semantic similarity language model enhancement (SMT+SSLM), the number increases to 65. For Cross Media Relevance Model, with the semantic similarity language model enhancement (CMRM+SSLM), the number increases from 66 to 90. We union the four query sets to get a new 98 query set. As the 263 testing words are unevenly distributed, we also use the 98 query set from the union to show the performance. The detailed results of the semantic similarity language model improved image annotation models are shown in Table 5.4 in comparison with the original models.

Table 5.5: Automatic annotation examples (fixed length of 5 words) of statistical machine translation model (SMT) and SMT with semantic similarity language model (SMT+SSLM)

Images				
SMT	flowers people mountain tree water	water tree snow buildings rocks	people buildings street cars plants	forest mare flowers tree street
SMT + SSLM	flowers needles blooms cactus grass	plain snow forest coyote wolf	buildings shops street sign writing	forest horse mare foals flowers

Table 5.6: Automatic annotation examples (fixed length of 5 words) of cross media relevance model (CMRM) and CMRM with semantic similarity language model (CMRM+SSLM)

Images				
CMRM	water sky tree people snows	water tree sky people ocean	sky water tree people plane	stone pillar tree sculpture people
CMRM + SSLM	snow fox pagoda railroad locomotive	ocean coral pool fish reefs	plane sky jet runway art	pillar shadows road stone temple
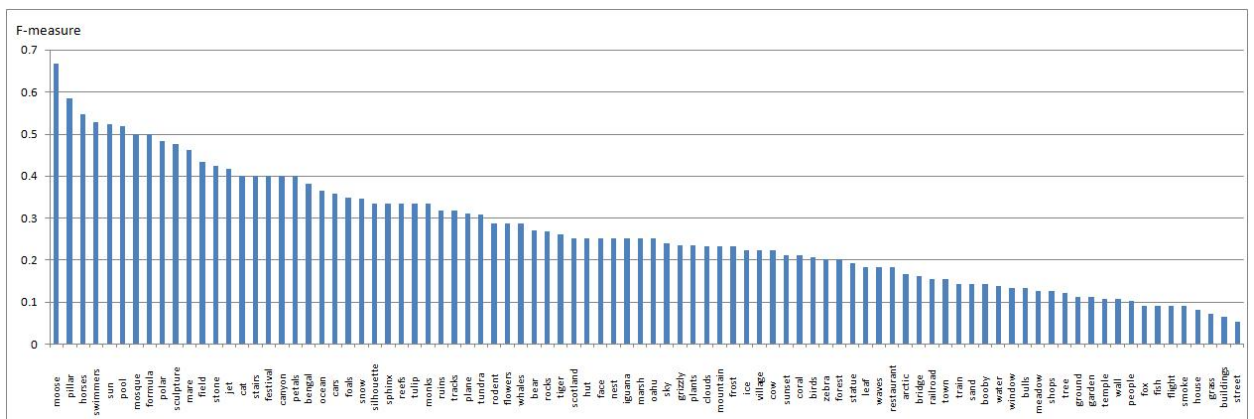
Figure 5.6: 90 non-zero recall words in CMRM+SSLM annotation result, ordered by F-measure

Chapter 6

Region-based medical image classification using auto-generated large training set

In Chapter 5 we proposed a novel framework to automatically annotate medical images with pathology labels via the implicit learning of the image-annotation relation as well as the intra-annotation relation. In the training data set, in case of an image of several regions of interest (ROIs) with multiple pathology labels, it is not clear which label points to which ROI, and the mapping/alignment of the labels and ROIs have to be learned implicitly through probabilistic models.

In this chapter, we address this problem by proposing another new framework that further extracts the image and text semantics and automatically links the pathology labels and ROIs explicitly in the training data. In this way, we gen-

erate a weakly labeled data set so that the training will be region specific. Then we classify the medical images through the visual features of its ROIs so that they can be annotated according to the pathological labels, to provide convenience for users to search and retrieve cases. We have also described our method in the paper [48].

6.1 Automatic generation of large training data set

In this section, we propose a framework to automatically map the information extracted from both text and images, and auto-create a region-based labeled image corpus for training. Figure 6.1 shows the architecture of our system. We take a series of images as input and reconstruct it to 3D. For image series with large distance between scans, we interpolate the intermediate images before 3D reconstruction. Then we set a threshold automatically and segment the 3D region(s) of interest. We register the 3D image to the reference 3D brain CT model we build and label the ROI(s) with the anatomy region(s) it overlaps with in the 3D brain atlas. For free text radiology report, we first extract the pathology and anatomy expressions and map them to standard terms; then we extract the logical relation between the anatomy terms and the pathology terms from their syntactic relation in the sentence. Finally we compare the anatomy labels for ROI from image processing result and the anatomy terms modifying the pathology term in text processing result. If they match, we set the class of the ROI with the pathology term its anatomy term modifies in the report.
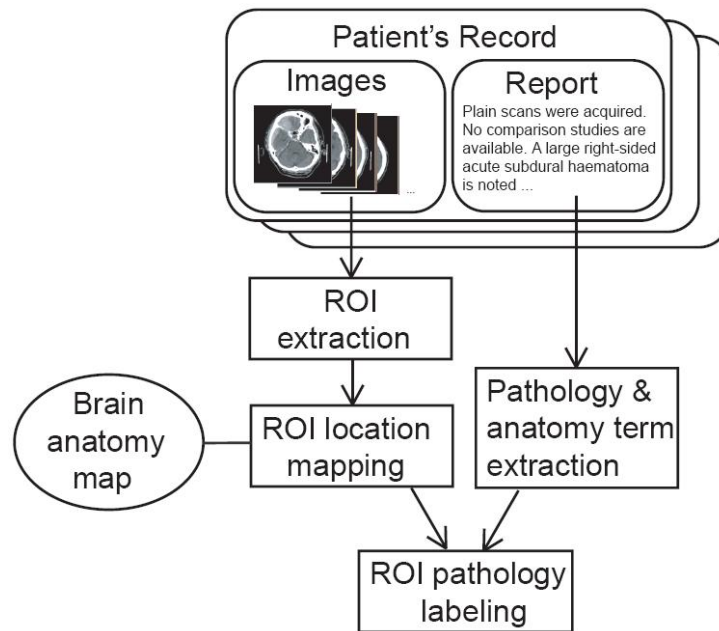
Figure 6.1: The framework of automatic ROI labeling

6.1.1 Anatomical location mapping of ROI

In order to link the report and the images, we need to map the ROIs segment-ed from the images to respective anatomical locations in the brain, so that when comparing with the associated radiology report, we know which ROI(s) the report describes. We consider two methods to map the anatomical location of the ROIs. We use the data provided in [124] as a brain model for an average brain and mark the different anatomical parts [26]. Figure 6.2 shows some slices of the brain anatomical atlas with different colors referring to different anatomical parts.

127

🟥 frontal lobe	🟦 temporal lobe	🟩 parietal lobe
🟨 occipital lobe	🟪 ventricle	
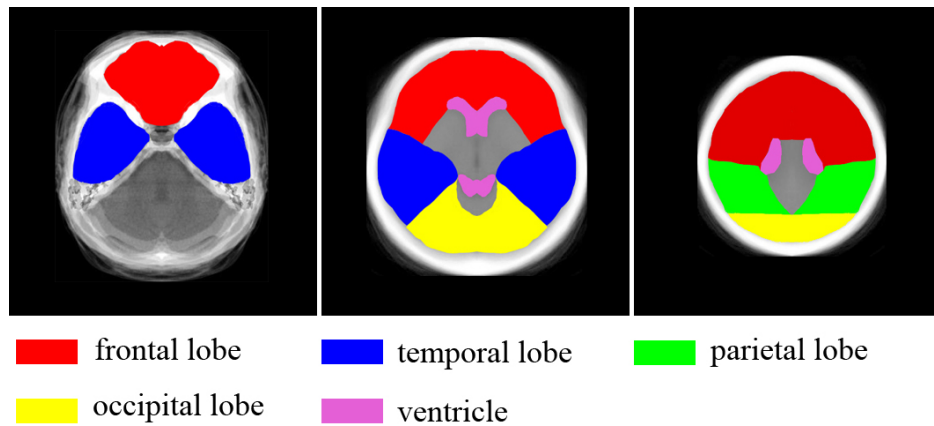
Figure 6.2: Some sample slices from the brain anatomy map

3D mapping

The first method uses reconstructed 3D ROIs and 3D brain model. In some cases, the images are too scarce for the ROIs to be directly reconstructed to 3D, we need to interpolate the intermediate slices prior to 3D reconstruction. Two categories of interpolation techniques are popular for reconstructing 3D objects from sparse sets: grey-level and shape-based interpolation. Grey-level interpolation approaches include nearest-neighbor, splines, linear, or polynomial interpolation. Shape-based interpolation approaches are often used on binary images and consider shape features extracted from the object sets. As the purpose of 3D reconstruction in this step is to register the actual brain to the ideal brain model, we use the cubic method, a grey-level approach to interpolate the intermediate slices of the actual brain image series and the brain model. After 3D reconstruction, we then use the method in [98] to do 3D image registration. We segment and label the ROI in each case with the anatomy region(s) it overlaps with in the referenced

128

3D brain CT model.

(a) view 1   (b) view 2
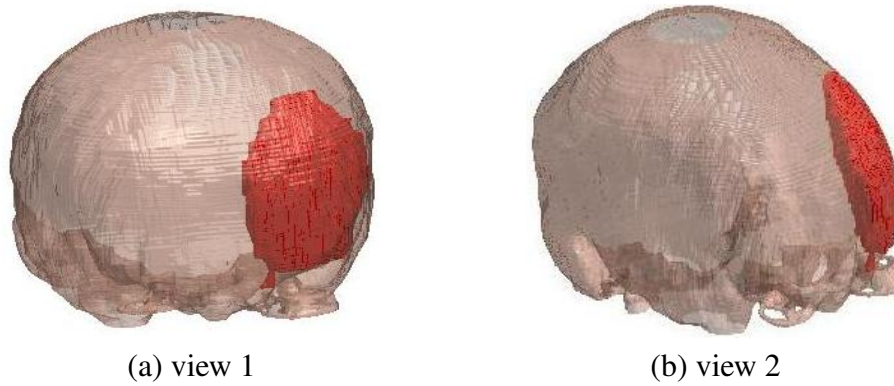
Figure 6.3: Reconstruction result: 3D hematoma in 3D brain

For example, after image registration and ROI segmentation, the 3D hematoma region shown in Figure 6.3 overlaps with the frontal lobe of the left side of the brain, so its anatomical location will be labeled as "left frontal lobe".

2D mapping

The second method uses the 2D brain maps. Since the number of slices of the brain maps is fixed in the model and the slice number of the actual case may be different from that, for each individual slice S_k in the series of image of CT examination $S = \{S_1, S_2, \ldots, S_k, \ldots, S_n\}$, we need to identify its corresponding brain map B_j in the series of brain maps $B = \{B_1, B_2, \ldots, B_k, \ldots, B_m\}$. As the starting and ending positions of CT examination are standard and the distance between two adjacent slices is constant, we first map the first and last images of the series S_1 and S_n to the first and last images of the brain map series B_1 and B_m respectively. Then sequence number j the corresponding brain map B_j of a given

129

slice S_k from an actual image series can be computed by:

$$j = round(k \cdot \frac{m}{n}) \tag{6.1}$$

Then similar to the first method, we use 2D image registration to find out the anatomical location of the 2D ROIs in the images. The anatomical location label of the 3D ROI is the union of all anatomical locations mapped to its 2D ROI components.

Both methods give almost the same results but the second method is faster using 2D image registration without the many interpolated intermediate slices; therefore, we use the 2D method for our experiment. However, should any 3D brain map be obtainable or the actual scans are dense enough, the 3D method could be considered.

6.1.2   ROI class label matching

For each case, after the ROI in the images is mapped to the anatomical location(s), we search the medical finding extracted from its report that also has the same anatomical description. Then we label the ROI with the pathology concept in that finding. For example, the hematoma region in Figure 6.4 is mapped to anatomical locations "left frontal lobe"; we search the medical finding with this anatomy concept, and if found in the structured term extraction result, we take its pathology concept and label the hematoma with it. Figure 6.4 shows the process of the ROI class label matching of the EDH example. After we automatically match and

label the ROIs in all the images with the pathology concepts from all associated radiology reports, we obtain a large size labeled data set for classifier training in the next phase.
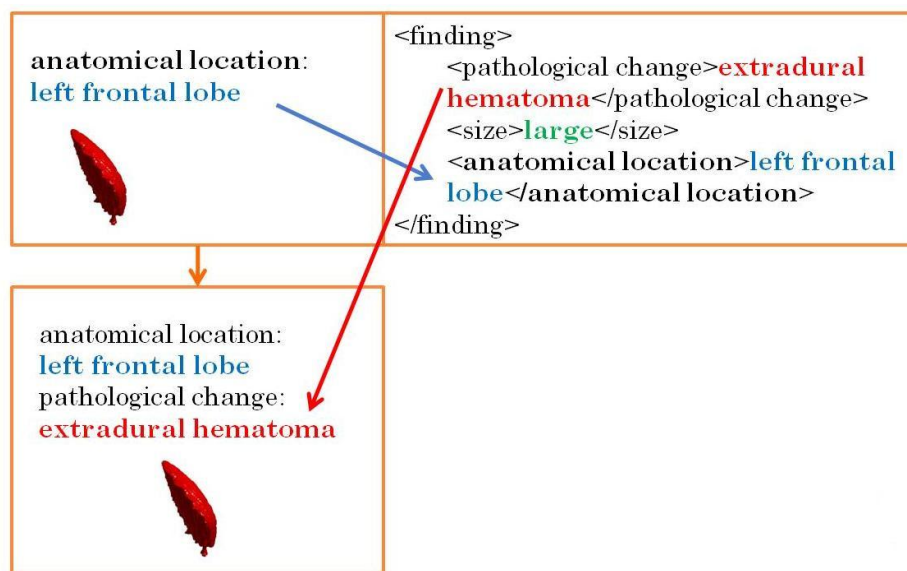
Figure 6.4: An example of ROI class label matching process

6.2 CT Image Classification

We classify the medical images according to the ROIs identified in the images. As it is possible for an image series to have multiple ROIs, the image series can have multiple classes. We provide two ways to classify the ROIs–3D ROI classification and 2D ROI classification.

131

6.2.1 ROI classification using 3D features

We need to extract the color, size, location, and shape features for hematoma classification. As we are interested in the ROIs only, all other details of the brain can be ignored for computation efficiency, we first segment the ROIs from the 2D image series and then use the shape based method described in [7], which was originally used for 3D tooth reconstruction, to do the interpolation to better precision than described in Section 6.1.1.

We use erosion and dilation, the mathematical morphologic transformation operations, for binary ROI shape modeling and interpolation. The morphologic operations transform two adjacent slices by combinations of dilations and erosions. The transformation is iteratively performed in such a way that the resulting ROIs from the adjacent slices become more similar to each other with respect to both shape and dimension.

The dilation of binary image P by using a structuring element B is given by:

$$P \oplus B = \cup_{b \in B} P_b, \tag{6.2}$$

where $\oplus$ denotes dilation and P_b is a structuring element centered onto an element of P. The erosion of P by using B is given by:

$$P \ominus B = \cap_{b \in B} P_b \tag{6.3}$$

For two adjacent CT slices P and Q, we apply the following morphological transformation on P so that the resulting image P' will grow from the shape of P

towards the shape of Q:

$$P' = [(P \ominus B) \cup ((P \cap Q) \oplus B)] \cap (P \cup Q). \qquad (6.4)$$

Equation 6.4 can be interpreted as: new image is to intersect the union of P and Q with the union of eroded P and dilated P and Q intersection. On the other hand, we apply a similar morphological transformation on Q so that the resulting image Q' will grow from the shape of Q towards the shape of P:

$$Q' = [(Q \ominus B) \cup ((Q \cap P) \oplus B)] \cap (Q \cup P). \qquad (6.5)$$

The morphological operations defined in Equation 6.4 and 6.5 are applied iteratively to the binary images resulted from previous morphological operations until the two binary images from respective morphological operations converge to be idempotent.

We use the method described above to interpolate the middle slice between adjacent slices iteratively until the desired density of the slices is obtained. In this way, the transition from one slice to another slice is smoother, thus it gives more accurate shape and volume features for ROI classification. Figure 6.5 demonstrates the 3D reconstruction examples of traumatic brain injury: (a) subdural hematoma (SDH), (b) extradural hematoma (EDH), and (c) intracerebral hemorrhage (ICH). The EDH is reconstructed from the images shown in Figure 1.1.

For the 3D ROIs reconstructed, we fit a ellipsoid to the region so that the ax-

(a) SDH   (b) EDH   (c) ICH
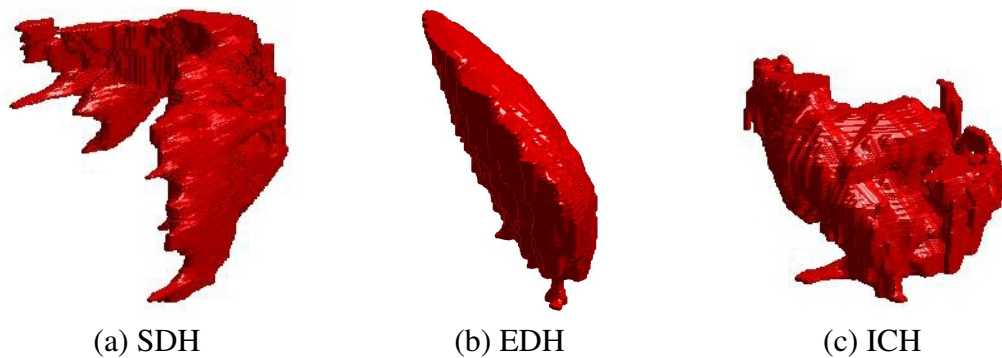
Figure 6.5: 3D hematoma reconstruction result

es of the ellipse have the same normalized second central moments[1] as the 3D
ROI. We use the lengths of the ROI along the three axes of the ellipsoid as shape
descriptors. As demonstrated in Figure 6.5, different types of hematoma usually
have different shapes, we choose extent and solidity as shape features to distin-
guish the hematomas. Table 6.1 shows the details of each feature.

Table 6.1: Features for 3D hematoma regions

Name	Description
Intensity	The average intensity of the hematoma
Volume	The volume of the hematoma
Location	The anatomical location of the hematoma
3D Axes	Lengths of the three axes of the fitting ellipsoid
3D Extent	hematoma volume/bounding box volume
3D Solidity	hematoma volume/convex hull volume

---

[1]In mathematics, a moment is a quantitative measure of the shape of a set of points. The
"second moment", known as the variance, is widely used and measures the "width" of a set of
points in one dimension or in higher dimensions measures the shape of a cloud of points as it
could be fit by an ellipsoid.

6.2.2 ROI classification using 2D features

Some image series are too sparse to get a good 3D reconstruction result, 3D features for ROIs in these cases may not be good enough. Alternatively, we extract 2D features of the ROIs as shown in Table 6.2.

Table 6.2: Features for 2D hematoma regions

Name	Description
Intensity	The average intensity of the hematoma
Area	The Area of the hematoma
Location	The anatomical location of the hematoma
2D Axes	Lengths of the two axes of the fitting ellipse
2D Extent	hematoma area/bounding box area
2D Solidity	hematoma area/convex hull area

Similar to our approach described in Chapter 5, for 3D volumetric case (such as CT and MRI) which consists of a series of 2D scan sequences, one 3D volumetric region of interest may be captured in several consecutive 2D scans and form different 2D ROIs in the individual scans. For example in Figure 1.1, the same clot of extradural hematoma (EDH) spreads over the consecutive images and form one 2D ROI in each of those images. However, it is possible that due to different features of each 2D ROI, they may be classified into different categories respectively. To solve this conflict of different classification for the same volumetric ROI, we use the method described in [47] to annotate the 3D volumetric ROI captured in n consecutive 2D scans. In Equation 6.6, the probability of pathology class c given a 3D ROI is the sum of the the probability of c given n individual 2D ROI r_i that comprises the 3D ROI r.

$$c = \arg\max_{c} \sum_{i=1}^{n} \frac{1}{n} p(c|r_i) \qquad (6.6)$$

6.3 Experiments

We obtained 429 CT image series of severe traumatic brain injury with associated radiology reports from National Neuroscience Institute, Tan Tock Seng Hospital. The pathology terms of hematoma types extracted from the radiology reports are: subdural hematoma (SDH), extradural hematoma (EDH), intracerebral hemorrhage (ICH), intraventricular hemorrhage (IVH), and subarachnoid hemorrhage (SAH). These hematoma types extracted served as class labels. The anatomy terms extracted from the text correspond to the anatomy part in the brain atlas. They include ventricles and left/right sides of the four lobes: frontal lobe, parietal lobe, temporal lobe, and occipital lobe. We partitioned the data set into a training set of 400 image series and a testing set of 29 image series. The 38 ROIs in 29 testing image series are manually labeled with pathology classes. We assigned pathology class labels to the 400 training cases and generated a training corpus. During testing, 34 of the 38 ROIs are assigned with the correct class label by the automatical process as described in Section 6.1.2, i.e. the overall precision of the automatical class label assigning process during creating training instances process is 89.5%. We built an SVM classifier using the auto-generated training corpus. Then we use the classifier to test on the manually labeled 38 ROIs from the 29 testing cases. Among the 38 testing ROIs, there are 13 SDH, 9 EDH, 8 ICH,

3 IVH, and 5 SAH. The pathology based classification results using 3D features and 2D features are shown in Table 6.3 and Table 6.4 respectively.

Apart from the class labeling error produced in automatic training corpus creation process, other factors that affect classification accuracy include the errors produced at different phases: interpolation error, segmentation error, image registration/anatomy labeling error, and term extraction error. For ROIs of relatively smaller size, they appear in very few scans or even just one scan in the image series, the interpolation, 3D reconstruction and feature extraction for such ROIs are poorer. Segmentation error affected the SAH cases more than others, as SAH is not seen as obvious as others in intensity. Image registration error, ROI anatomy labeling error, and term extraction error are minimal and have the least effect on the system performance. The classification result is also affected due to unbalanced data set. SDH is the most frequent class label, but only a few cases are labeled with IVH. Therefore, the classification is biased towards SDH and results in higher recall; whereas the classification result for IVH is poorer than most of the other classes.

Table 6.3: Hematoma classification result using 3D features

	SDH	EDH	ICH	IVH	SAH
precision	84.6	100.0	77.8	66.7	80.0
recall	84.6	88.9	87.5	66.7	80.0

Table 6.4: Hematoma classification result using 2D features

	SDH	EDH	ICH	IVH	SAH
precision	90.9	88.9	77.8	66.7	66.7
recall	76.9	88.9	87.5	66.7	80.0

137

As our work is the first in the area of image automatic classification, there is no other research work for us to compare the results with. However, since the data sets used in this chapter and in previous chapters are of the same type and from the same source, though the specific images series used for training and testing are different, we consider the experiment results comparable. Though the ROIs are manually labeled in [49], the ROI classification result using the method described in this chapter is better than the results from [49] because the automatic region based class labeling process in this chapter provides a much larger data set for training; whereas the data set in [49] is very small due to high cost in manual medical image labeling. The result is also slightly better than the result from Chapter 5, because the region specific labels along with the ROI features provide more insights of the ROIs, and the ROI-annotation alignment error produced by the method described in Chapter 5 is greater than the automatic ROI labeling error using the method described in this chapter.

We use the automatically generated class labels to index the images. Figure 6.6 and 6.7 show some examples of annotation based image retrieval results. Figure 6.6 shows retrieval results to query: pathology class = EDH. Figure 6.7 shows retrieval results to query: pathology class = SDH, anatomical label = left frontal lobe.

6.4 Discussion

We propose a novel framework to utilize deeper semantics of the medical images and text to automatically create a training corpus for region based medical image classification. The accuracy of assigning correct class labels to respective regions of interest in the process of training data set creation is promising to medical professionals. The ROI classification result is better than the results from our approaches in previous chapters due to the novel automatic region based class labeling process, as it provides more insights of the ROIs and create a larger data set for training and testing. Our proposed pathology based automatic medical image classification can provide reference or "second opinion" to radiologist and other medical professionals. We can use the classification result generated by our system to index the medical images according to pathology changes, and provide users convenience in searching and retrieving medical images. The novel framework to label the images automatically can be used to create labeled data for training, so that expensive and time-consuming manual labeling will no longer be necessary. Due to automatic training image labeling process, noise is introduced to the training data inevitably. In future work, we will further study this problem and may find a way to effectively detect the noise and remove the noisy labels from the training data, so that the training images will be less "weakly" labeled and annotation performance can be further improved. Although we experiment on hematomas in traumatic brain injury CT images; however, the framework and methods we describe in this chapter can be further extended and applied to

medical images of other modality or anatomical focus with adaptations on image

processing.
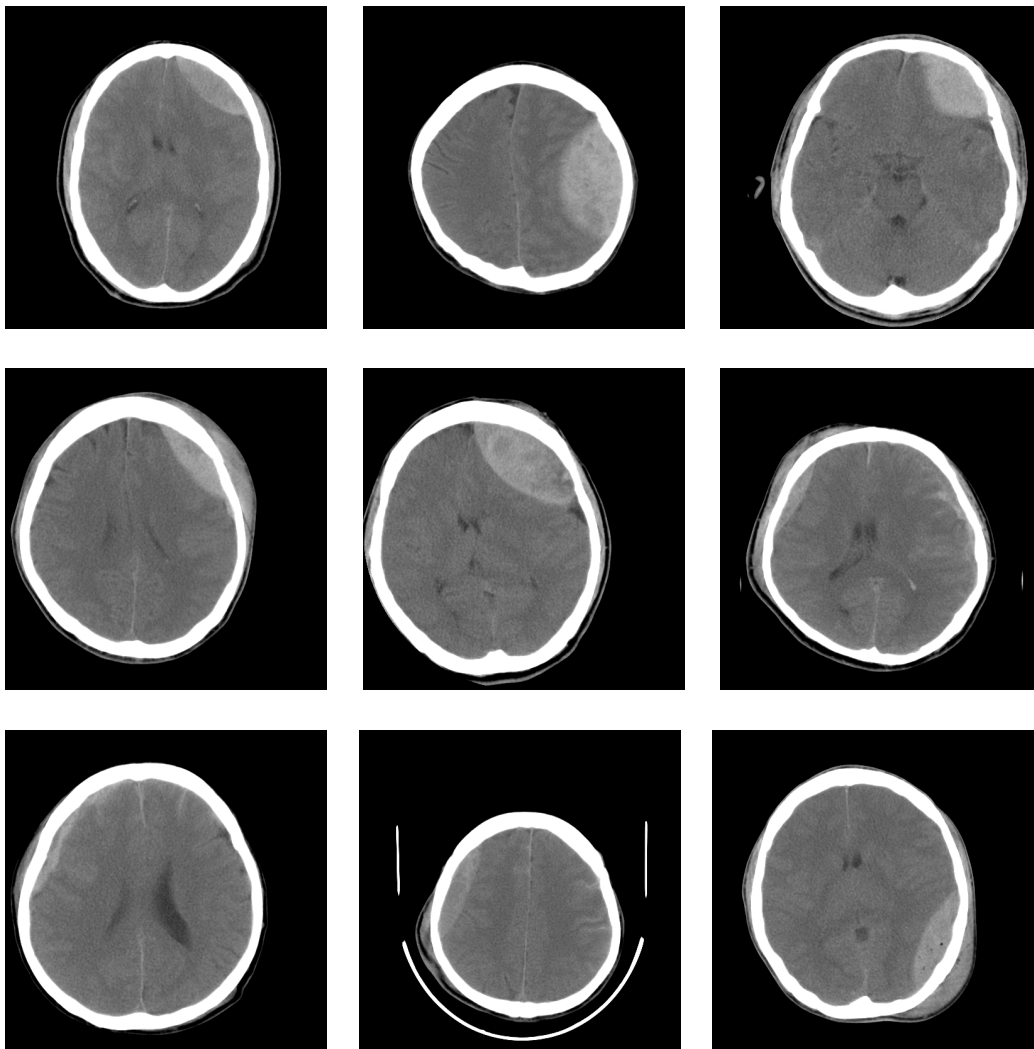
Figure 6.6: Image retrieval results of query example 3: pathology class = EDH
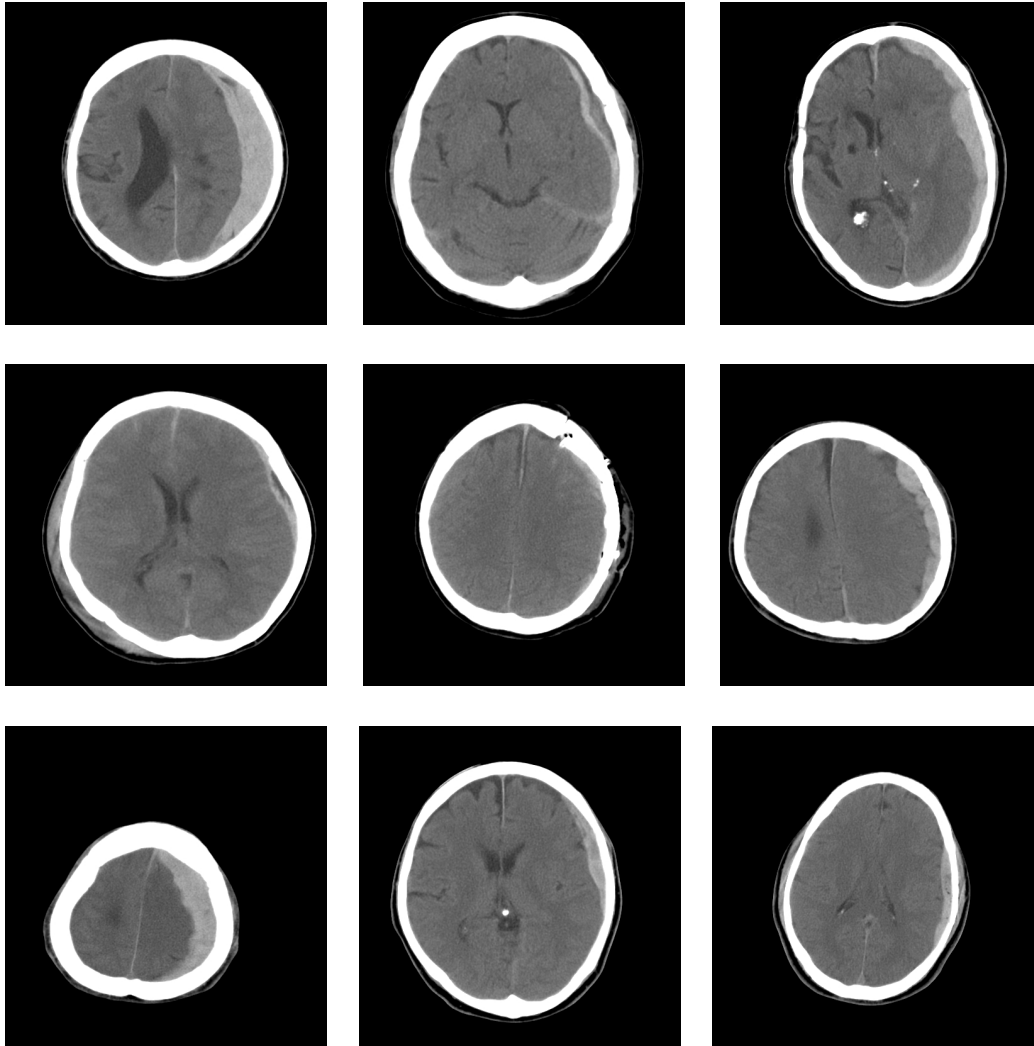
Figure 6.7: Image retrieval results of query example 4: pathology class = SDH, anatomical label = left frontal lobe

Chapter 7

Conclusion

In this thesis, we proposed novel frameworks to solve the problems and difficulties in automatic medical image annotation, classification, and retrieval.

The first framework we proposed in Chapter 3 is for text-based medical image retrieval by extracting structured information from unstructured medical text to index and retrieve associated medical images. We used natural language processing techniques and medical knowledge resource to extract useful information from free text medical report and used the extracted information to index the reports as well as the images the reports are associated with, so that the users can search and retrieve medical images that have accompanying reports by typing text queries into the system, and the system will return medical images that fulfill the text queries. In addition to text-based indexing and retrieval of medical images, the medical findings and their specific descriptors such as anatomical locations extracted from the free text medical reports can also be used to help with image

processing in region of interest recognition, classification, and annotation.

The second framework we proposed in Chapter 4 is for content-based medical image retrieval by processing medical images and indexing them with binary low level visual feature vector. By comparing the binary feature vector of the query image submitted by user and the binary feature vectors of the images in the database, we are able to return a list of visually similar images ranked by similarity to the input image. With this approach, users can retrieve medical images without accompanying textual reports.

The third framework we proposed is for annotation-based medical image retrieval. In contrast to most research works that classify the medical images according to their modality, anatomical body part, or the presence of abnormality, we put more effort in analyzing the types of pathology changes present in the images. Indexing and retrieving medical images by their pathological annotations can help to satisfy doctors' need to search and retrieve medical images pertaining to one pathology class for work, research, or teaching purposes. Different from other approaches, we proposed the new idea of making use of both text and image semantics for automatic image annotation. We extract information from the radiology reports associated with the medical images and thus generating a automatically labeled data set; therefore, the much expense and troubles of requiring human experts to manually label the data can be saved.

We included two different approaches for the framework of annotation-based medical image retrieval. In the first approach described in Chapter 5, the annotation words from the textual reports are mapped to the whole image, thus we used

probabilistic models to derive the correlations between the regions of interest in the images and the annotations. In this second approach described in Chapter 6, we mapped the annotations explicitly to regions of interest in the images using anatomical information extracted from the reports and registered images. In both approaches, after we built annotation/classification models in training phase, then in testing phase, we first process the testing image and assign annotation words to the images according to the trained models.

The frameworks we proposed in this thesis provided solutions to the problems and difficulties we mentioned in Chapter 1. Our systems satisfied the needs of searching and retrieving medical images accurately, conveniently, and efficiently. The results from the experiments we conducted are satisfactory to doctors that tested the systems. We processed the texts and images to obtain deeper information that can be used for further applications in the medical domain. By building the automatic medical image annotation/classification models, we are able to predict the pathology changes in newly acquired medical images and thus give reference or second opinions to doctors or radiologists. The systems and platforms we developed can be used for training junior medical professionals. Some novel methods we developed along with the frameworks can benefit the medical informatics community as well as the computer science community in general. For examples, the semantic similarity language model we developed in Chapter 5 and the improved automatic image annotation methods can not only apply to medical image annotation, but also to automatic image annotation in the general domain. The 3D visualization result from Chapter 6 can also be used to help surgeons to

visualize the brain before the actual surgeries.

For future work, we can further extend our work to other medical image modalities like MRI and X-ray or medical images of other anatomical part such as lung and abdomen. For automatic medical image annotation and classification, the medical findings we focus on are hematomas in traumatic brain injury images. However, other medical findings are also important. We can extend our work to include fracture, midline shift, edema, hydrocephalus, and other pathological changes found in traumatic brain injury cases.

List of Abbreviations

Abbreviation	Details
ABIR	Annotation based image retrieval systems
ASSERT	Automatic Search and Selection Engine with Retrieval Tools
CAD	Computer-aided diagnosis
CBIR	Content based image retrieval
CG	Cumulative gain
CMRM	Cross-media relevance model
CSF	Cerebrospinal fluid
CT	Computed tomography
DCG	Discounted cumulative gain
DICOM	Digital Imaging and Communications in Medicine
EDH	Epidural hematoma or Extradural hematoma
EM	Expectation-maximization
GMM	Gaussian mixture model
HMM	Hidden Markov model
ICH	Intracerebral hemorrhage
IRMA	Image retrieval in medical applications
IVH	Intraventricular hemorrhage
k-NN	K-nearest neighbor
LDA	Latent Dirichlet allocation
LFS	Lobular feature set
LSA	Latent semantic analysis
LSP-MLP	Linguistic String Project-Medical Language Processor

MedLEE	MEDical Language Extraction and Encoding System
MeSH	Medical Subject Headings
MIL	Multiple instance learning
MRI	Magnetic resonance imaging
NDCG	Normalized discounted cumulative gain
NER	Named entity recognition
NLP	Natural language processing
PACS	Picture Archiving and Communication System
PBR	Pathology bearing region
PLSA	Probabilistic latent semantic analysis
POS	Part of speech
RADA	Radiology Analysis Tool
RBF	Radial basis function
RBFNN	Radial basis function neural networks
ROI	Region of interest
SAH	Subarachnoid hemorrhage
SDH	Subdural hematoma
SMT	Statistical machine translation
SQL	Structured Query Language
SSLM	Semantic similarity language model
SVM	Support vector machine
TBI	Traumatic brain injury
TBIR	Text based image retrieval systems
UMLS	Unified Medical Language System
XML	Extensible Markup Language

Bibliography

[1] Fact sheet: Medical subject headings. `http://www.nlm.nih.gov/pubs/factsheets/mesh.html`.

[2] Masahito Aoyama, Qiang Li, Shigehiko Katsuragawa, Feng Li, Shusuke Sone, and Kunio Doi. Computerized scheme for determination of the likelihood measure of malignancy for pulmonary nodules on low-dose ct images. *Medical Physics*, 30(3):387–394, 2003.

[3] S.G. Armato and H. MacMahon. Automated lung segmentation and computer-aided diagnosis for thoracic ct scans. In *Proceedings of Computer Assisted Radiology and Surgery 2003*, pages 977–982, 2003.

[4] Kobus Barnard and David Forsyth. Learning the semantics of words and pictures. In *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, pages 408–415, 2001.

[5] R. A. Blechschmidt, R. Werthschutzky, and U. Lorcher. Automated ct image evaluation of the lung: a morphology-based concept. *IEEE Transactions on Medical Imaging*, 20(5):434–442, May 2001.

[6] David M. Blei and Michael I. Jordan. Modeling annotated data. In *Proceedings of ACM SIGIR International Conference on Research and Development in Informaion Retrieval*, pages 127–134, 2003.

[7] Adrian G. Bors, Lefteris Kechagias, and Ioannis Pitas. Binary morphological shape-based interpolation applied to 3-d tooth reconstruction. *IEEE Transactions on Medical Imaging*, 21(2):100–108, 2002.

[8] Peter F. Brown, John Cocke, Stephen A. Della Pietra, Vincent J. Della Pietra, Fredrick Jelinek, John D. Lafferty, Robert L. Mercer, and Paul S. Roossin. A statistical approach to machine translation. *Computer Linguistics*, 16(2):79–85, 1990.

[9] Peter F. Brown, Vincent J.Della Pietra, Stephen A. Della Pietra, and Robert. L. Mercer. The mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics*, 19:263–311, 1993.

[10] Tom Weidong Cai, Jinman Kim, and David Dagan Feng. *Content-Based Medical Image Retrieval*, pages 83–114. Elsevier, 2008.

[11] David A. Campbell and Stephen B. Johnson. A transformational-based learner for dependency grammars in discharge summaries. In *Proceedings of the ACL-02 Workshop on Natural Language Processing in the Biomedical Domain*, volume 3, 2002.

[12] Gustavo Carneiro and Nuno Vasconcelos. Formulating semantic image annotation as a supervised learning problem. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 163–168, 2005.

[13] Tao Chan. Computer aided detection of small acute intracranial hemorrhage on computer tomography of brain. *Computerized Medical Imaging and Graphics*, 31(4-5):285–298, June-July 2007.

[14] Edward Y. Chang, Kingshy Goh, Gerard Sychay, and Gang Wu. Cbsa: content-based soft annotation for multimodal image retrieval using bayes point machines. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(1):26–38, January 2003.

[15] Ruey-Feng Chang, Wen-Jie Wu, Woo Kyung Moon, , and Dar-Ren Chen. Improvement in breast tumor discrimination by support vector machines and speckle-emphasis texture analysis. *Ultrasound in Medicine and Biology*, 29(5):679–686, 2003.

[16] Olivier Chapelle, Patrick Haffner, and Vladimir Vapnik. Support vector machines for histogram-based image classification. *IEEE Transactions on Neural Networks*, 10(5):1055–1064, September 1999.

[17] Chung-Ming Chen, Yi-Hong Chou, Ko-Chung Han, Guo-Shian Hung, Chui-Mei Tiu, Hong-Jen Chiou, and See-Ying Chiou. Breast lesions on sonograms: Computer-aided diagnosis with nearly setting-independent features and artificial neural networks. *Radiology*, 226:504–514, 2003.

[18] Dar-Ren Chen, Wen-Jia Kuo, Ruey-Feng Chang, Woo Kyung Moon, and Cheng Chun Lee. Use of the bootstrap technique with small training sets for computer-aided diagnosis in breast ultrasound. *Ultrasound in Medicine and Biology*, 28(7):897–902, 2002.

[19] Deng-Da Chenga, Xiaopeng Cai, Xiaowei Chen, Liming Hu, and Xueling Lou. Computer-aided detection and classification of microcalcifications in mammograms: a survey. *Pattern Recognition*, 36(12):2967–2991, 2003.

[20] Deng-Da Chenga, Juan Shan, Wen Jua, Yanhui Guo, and Ling Zhang. Automated breast cancer detection and classification using ultrasound images: A survey. *Pattern Recognition*, 43(1):299–317, 2010.

[21] Deng-Da Chenga, Xiangjun Shi, Rui Min, Liming Hu, Xiaopeng Cai, and Haining Du. Approaches for automated detection and classification of masses in mammograms. *Pattern Recognition*, 39(4):646–668, 2006.

[22] Wesley W. Chu and Qiming Chen. A structured approach for cooperative query answering. *IEEE Transactions on Knowledge and Data Engineering*, 6(5):738–749, October 1994.

[23] J. Cimino, P. Clayton, G. Hripcrsk, and S. Johnson. Knowledge based approaches to the maintenance of a large controlled medical terminology. *Journal of the American Medical Informatics Association*, 1(1):35–40, 1994.

[24] Aaron M. Cohen and William R. Hersh. A survey of current work in biomedical text mining. *Briefings in Bioinformatics*, 6(1):57–71, 2005.

[25] Dubravko Cosic and Sven Loncaric. Rule-based labeling of ct head image. In *Proceedings of the sixth Conference on Artificial Intelligence in Medicine in Europe (AIME 1997)*, pages 453–456, 1997.

[26] Hanna Damasio. *Human Brain Anatomy in Computerized Images*. Oxford University Press, 2005.

[27] Ritendra Datta, Dhiraj Joshi, Jia Li, and James Z. Wang. Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys*, 40:5:1–5:60, May 2008.

[28] Marie-Catherine de Marneffe, Bill MacCartney, and Christopher D. Manning. Generating typed dependency parses from phrase structure parses. In *Proceedings of the fifth international conference on Language Resources and Evaluation (LREC2006)*, Genoa, Italy, 2006.

[29] Scott Deerwester, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, and Richard Harshman. Indexing by latent semantic analysis. *Journal of the Society for Information Science*, 41(6):391–407, 1990.

[30] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1):1–38, 1977.

[31] Kunio Doi. Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. *Computerized Medical Imaging and Graphics*, 31(4–5):198–211, 2007.

[32] Andrew Downie. Tutorial: Ct in head trauma, 2001.

[33] Karen Drukker, Darrin C. Edwards, Maryellen L. Giger, Robert M. Nishikawa, and Charles E. Metz. Computerized detection and 3-way classification breast lesions ultrasound images. In *Proceedings of SPIE Medical Imaging 2004*, pages 1034–1041, 2004.

[34] Karen Drukker, Maryellen L. Giger, Karla Horsch, Matthew A. Kupinski, and Carl J. Vyborny. Computerized lesion detection on breast ultrasound. *Medical Physics*, 29(7):1438–1446, 2002.

[35] Karen Drukker, Maryellen L. Giger, and Charles E. Metz. Robustness of computerized lesion detection and classification scheme across different breast us platforms. *Radiology*, 238(1):834–840, 2006.

[36] Karen Drukker, Maryellen L Giger, Carl J Vyborny, and Ellen B Mendelson. Computerized detection and classification of cancer on breast ultrasound. *Academic Radiology*, 11(5):526–535, 2004.

[37] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification*. Wiley-Interscience, second edition, 2000.

[38] Pinar Duygulu, Kobus Barnard, Joo F. G. de Freitas, and David A. Forsyth. Object recognition as machine translation: Learning a lexicon for a fixed

image vocabulary. In *Proceedings of European Conference on Computer Vision (ECCV)*, pages 97–112, 2002.

[39] Carol Friedman, Philip O. Alderson, John H. M. Austin, James J. Cimino, and Stephen B. Johnson. A general natural language text processor for clinical radiology. *Journal of the American Medical Informatics Association*, 1(2):161–174, March April 1994.

[40] Carol Friedman and George Hripcsak. Evaluating natural language processors in clinical domain. In *Proceedings of the Conference on Natural Language and Medical Concept Representation*, pages 41–52, 1997.

[41] Carol Friedman, Lyudmila Shagina, Yves Lussier, and George Hripcsak. Automated encoding of clinical documents based on natural language processing. *Journal of the American Medical Informatics Association*, 11(5):392–402, September October 1994.

[42] Cheong Yiu Fung and Kai Fock Loe. Learning primitive and scene semantics of images for classification and retrieval. In *Proceedings of the seventh ACM international conference on Multimedia*, pages 9–12, 1999.

[43] B. S. Garra, B. H. Krasner, S. C. Horii, S. Ascher, S. K. Mun, and R. K. Zeman. Improving the distinction between benign and malignant breast lesions: the value of sonographic texture analysis. *Ultrasonic Imaging*, 15(4):267–285, 1993.

[44] Smadar Gefen, Oleh J. Tretiak, Catherine W. Piccoli, Kevin D. Dono-hue, Athina P. Petropulu, P. Mohana Shankar, Vishruta A. Dumane, Lexun Huang, M. Alper Kutay, Vladimir Genis, Flemming Forsberg, John M. Reid, and Barry B. Goldberg. Roc analysis of ultrasound tissue characterization classifiers for breast cancer diagnosis. *IEEE Transactions on Medical Imaging*, 22(2):170–177, 2003.

[45] Arnab Ghoshal, Pavel Ircing, and Sanjeev Khudanpur. Hidden markov models for automatic annotation and content-based retrieval of images and video. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 544–551, 2005.

[46] Tianxia Gong, Shimiao Li, and Chew Lim Tan. A semantic similarity language model to improve automatic image annotation. In *Proceedings of the 22nd International Conference on Tools with Artificial Intelligence (IC-TAI2010)*, pages 197–203, 2010.

[47] Tianxia Gong, Shimiao Li, Chew Lim Tan, Boon Chuan Pang, C.C. Tchoyoson Lim, Cheng Kiang Lee, Qi Tian, and Zhuo Zhang. Automatic pathology annotation on medical images: A statistical machine translation framework. In *Proceedings of the 20th International Conference on Pattern Recognition*, pages 2504–2507, 2010.

[48] Tianxia Gong, Shimiao Li, Jie Wang, Chew Lim Tan, Boon Chuan Pang, C.C. Tchoyoson Lim, Cheng Kiang Lee, Qi Tian, and Zhuo Zhang. Automatic labeling and classification of brain ct images. In *Proceedings of*

156

the 18th International Conference on Image Processing (ICIP 2011) (to appear), 2011.

[49] Tianxia Gong, Ruizhe Liu, Chew Lim Tan, Neda Farzad, Cheng Kiang Lee, Boon Chuan Pang, Qi Tian, Suisheng Tang, and Zhuo Zhang. Classification of ct brain images of head trauma. In *Proceedings of the second IAPR International Workshop on Pattern Recognition in Bioinformatics*, pages 401–408, 2007.

[50] Tianxia Gong, Chew Lim Tan, Tze Yun Leong, Cheng Kiang Lee, Boon Chuan Pang, Tchoyoson Lim, Qi Tian, Suisheng Tang, and Zhuo Zhang. Text mining in radiology reports. In *Proceedings of the eighth IEEE International Conference on Data Mining*, pages 815–820, 2008.

[51] J. Guo, J. M. Reinhardt, H. Kitaoka, L. Zhang, M. Sonka, G. McLennan, and E. A. Hoffman. Integrated system for ct-based assessment of parenchymal lung disease. In *Proceedings of the 2002 IEEE International Symposium on Biomedical Imaging*, pages 871–874, 2002.

[52] Claudia I. Henschke, David F. Yankelevitz, Ion Mateescu, Dean W. Brettle, Timothy G. Rainey, and Fred S. Weingard. Neural networks for the analysis of small pulmonary nodules. *Clinical Imaging*, 21(6):390–399, November-December 1997.

[53] Ralf Herbrich, Thore Graepel, and Colin Campbell. Bayes point machines. *The Journal of Machine Learning Research*, 1:245–279, September 2001.

[54] Jerry R. Hobbs. Information extraction from biomedical text. *Journal of Biomedical Informatics*, 35(4):260–264, 2003.

[55] Thomas Hofmann. Unsupervised learning by probabilistic latent semantic analysis. In *Machine Learning*, volume 42, pages 177–196, 2001.

[56] Karla Horsch, Maryellen L. Giger, and Carl J. Vyborny. Computerized diagnosis of breast lesions on ultrasound. *Medical Physics*, 29(2):157–164, 2002.

[57] H. K. Huang. *PACS: Basic principles and applications*. Wiley, 1999.

[58] Yu-Len Huang, Dar-Ren Chen, and Ya-Kuang Liu. Breast cancer diagnosis using image retrieval for different ultrasonic systems. In *Proceedings of the 11th International Conference on Image Processing (ICIP 2004)*, pages 2957–2960, 2004.

[59] Yu-Len Huang, Sheng-Hsiung Lin, and Dar-Ren Chen. Computer-aided diagnosis applied to 3-d us of solid breast nodules by using principal component analysis and image retrieval. In *Proceedings of the 27th Annual International Conference on Engineering in Medicine and Biology Society (IEEE-EMBS 2005)*, pages 1802–1805, 2005.

[60] Yu-Len Huang, Kao-Lun Wang, and Dar-Ren Chen. Diagnosis of breast tumors with ultrasound texture analysis using support vector machines. *Neural Computing and Applications*, 15(2):164–169, 2006.

[61] Betsy L. Humphreys and Donald A. B. Lindberg. The umls project: making the conceptual connection between users and the information they need. *Bulletin of the Medical Library Association*, 81(2):170–177, April 1993.

[62] Kalervo Jarvelin and Jaana Kekalainen. Cumulated gain-based evaluation of ir techniques. *ACM Transactions on Information Systems*, 20(4):422–446, 2002.

[63] J. Jeon, V. Lavrenko, and R. Manmatha. Automatic image annotation and retrieval using cross-media relevance models. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, 2003.

[64] Rong Jin, Joyce Y. Chai, and Luo Si. Effective automatic image annotation via a coherent language model and active learning. In *Proceedings of the 12th annual ACM International Conference on Multimedia*, pages 892–899, 2004.

[65] David B. Johnson, Ricky K. Taira, Alfonso F. Cardenas, and Denise R. Aberle. Extracting information from free text radiology reports. *International Journal on Digital Libraries*, 1(3):297–308, December 1997.

[66] Segyeong Joo, Yoon Seok Yang, Woo Kyung Moon, and Hee Chan Kim. Computer-aided diagnosis of solid breast nodules:use of artificial neural network based on multiple sonographic features. *IEEE Transactions on Medical Imaging*, 23(10):1292–1300, 2004.

[67] Yoshiki Kawata, Noboru Niki, Hironobu Ohmatsu, Masahiko Kusumo-to, Ryutaro Kakinuma, Kiyoshi Mori, Hiroyuki Nishiyama, Kenji Eguchi, Masahiro Kaneko, and N. Moriyama. Computer-aided diagnosis of pulmonary nodules using three-dimensional thoracic ct images. In *Proceedings of 4th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 1393–1394, 2001.

[68] Yoshiki Kawata, Noboru Niki, Hironobu Ohmatsu, Masahiko Kusumo-to, Ryutaro Kakinuma, Kiyoshi Mori, Hiroyuki Nishiyama, Kenji Eguchi, Masahiro Kaneko, and N. Moriyama. Searching similar images for classification of pulmonary nodules in three-dimensional ct images. In *Proceedings of 2002 IEEE International Symposium on Biomedical Imaging*, pages 189–192, 2002.

[69] Yoshiki Kawata, Noboru Niki, Hironobu Ohmatsu, Masahiko Kusumoto, Ryutaro Kakinuma, Kouzo Yamada, Kiyoshi Mori, Hiroyuki Nishiyama, Kenji Eguchi, Masahiro Kaneko, and N. Moriyama. Pulmonary nodule classification based on nodule retrieval from 3-d thoracic ct image database. In *Proceedings of 7th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 838–846, 2004.

[70] Daniel Keysers, Jorg Dahmen, and Hermann Ney. Statistical framework for model-based image retrieval in medical applications. *Journal of Electronic Imaging*, 12(1):59–68, 2003.

160

[71] Dan Klein and Christopher D. Manning. Fast exact inference with a factored model for natural language parsing. In *In Advances in Neural Information Processing Systems*, pages 3–10. MIT Press, 2002.

[72] Dan Klein and Christopher D. Manning. Accurate unlexicalized parsing. In *In Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pages 423–430, 2003.

[73] Wen-Jia Kuo, Ruey-Feng Chang, Dar-Ren Chen, and Cheng Chun Lee. Data mining with decision trees for diagnosis of breast tumor in medical ultrasonic images. *Breast Cancer Research and Treatment*, 66(1):51–57, 2001.

[74] Wen-Jia Kuo, Ruey-Feng Chang, Cheng Chun Lee, Woo Kyung Moon, and Dar-Ren Chen. Retrieval technique for the diagnosis of solid breast tumors on sonogram. *Ultrasound in Medicine and Biology*, 28(7):903–909, 2002.

[75] V. Lavrenko and W. Croft. Relevance-based language models. In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, 2001.

[76] V. Lavrenko, R. Manmatha, and J. Jeon. A model for learning the semantics of pictures. In *Proceedings of Advances in Neural Information Processing Systems*, 2003.

[77] Jia Li and James Z. Wang. Real-time computerized annotation of pictures. In *Proceedings of ACM International Conference on Multimedia*, pages 911–920, 2006.

[78] Shimiao Li, Tianxia Gong, Jie Wang, Ruizhe Liu, Chew Lim Tan, Tze Yun Leong, Boon Chuan Pang, C. C. Tchoyoson Lim, Cheng Kiang Lee, Qi Tian, and Zhuo Zhang. Tbidoc:3d content-based ct image retrieval system for traumatic brain injury. In *Proceedings of SPIE Medical Imaging 2010*, 2010.

[79] Chun-Chih Liao, Furen Xiao, Jau-Min Wong, , and I-Jen Chiang. A knowledge discovery approach to diagnosing intracranial hematomas on brain ct: recognition, measurement and classification. In *Proceedings of the first International Conference on Medical Biometrics*, pages 73–82, 2008.

[80] Shih-Chung B. Lo, Li-Yueh Hsu, Matthew T. Freedman, Yuan Ming F. Lure, and Hui Zhao. Classification of lung nodules in diagnostic ct: an approach based on 3-d vascular features, nodule density distributions, and shape features. In *Proceedings of SPIE Medical Imaging 2003*, pages 183–189, 2003.

[81] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schtze. *Introduction to Information Retrieval*. Cambridge University Press, New York, NY, USA, 2008.

[82] Raphael Maree, Pierre Geurts, Justus Piater, and Louis Wehenkel. Random subwindows for robust image classification. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 34–40, 2005.

[83] Yuichi Matsuki1, Katsumi Nakamura, Hideyuki Watanabe, Takatoshi Aoki, Hajime Nakata, Shigehiko Katsuragawa, and Kunio Doi. Usefulness of an artificial neural network for differentiating benign from malignant pulmonary nodules on high-resolution ct: evaluation with receiver operating characteristic analysis. *American Journal of Roentgenology*, 178(3):657–663, 2002.

[84] A. T. McCray, A. Burgun, and O.z Bodenreider. Aggregating umls semantic types for reducing conceptual complexity. *Studies in Health Technology and Informatics*, 84((Pt 1)):216–220, 2001.

[85] Michael F. McNitt-Gray, Eric M. Hart, Nathaniel Wyckoff, James W. Sayre, Jonathan G. Goldin, , and Denise R. Aberle. A pattern classification approach to characterizing solitary pulmonary nodules imaged on high resolution ct: preliminary results. *Medical Physics*, 26(6):880–888, 1999.

[86] Michael F. McNitt-Gray, Nathaniel Wyckoff, James W. Sayre, Jonathan G. Goldin, , and Denise R. Aberle. The effects of co-occurrence matrix based texture parameters on the classification of solitary pulmonary nodules imaged on computed tomography. *Computerized Medical Imaging and Graphics*, 23(6):339–348, 1999.

[87] Eneida A. Mendonca, Janet Haas, Lyudmila Shagina, Elaine Larson, and Carol Friedman. Extracting information on pneumonia in infants using natural lanuage processing of radiology reports. *Journal of Biomedical Informatics*, 38:314–321, 2005.

[88] S. M. Meystre, G. K. Savova, K. C. Kipper-Schuler, and J. F. Hurdle. Extracting information from textual documents in the electronic health record: a review of recent research. *IMIA Yearbook of Medical Informatics 2008*, 47(Suppl 1):128–144, 2008.

[89] Michiaki Mishima, Toyohiro Hirai, Harumi Itoh, Yasutaka Nakano, Hiroaki Sakai, Shigeo Muro, Koichi Nishimura, Yoshitaka Oku, Kazuo Chin, Motoharu Ohi, Takashi Nakamura, Jason H. T. Bates, Adriano M. Alencar, and Bela Suki. Complexity of terminal airspace geometry assessed by lung computed tomography in normal subjects and patients with chronic obstructive pulmonary disease. *Proceedings of the National Academy of Science of United States of America*, 96(16):8829–8834, 1999.

[90] Jeff Mitchell and Mirella Lapata. Vector-based models of semantic composition. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics*, pages 236–244, 2008.

[91] Jeff Mitchell and Mirella Lapata. Language models based on semantic composition. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 430–439, 2009.

164

[92] Fumihiro Mitsunobu, Kozo Ashida, Yasuhiro Hosaki, Hirofumi Tsugeno, Makoto Okamoto, Kazunori Nishida, Shingo Takata, Tadashi Yokoi, Michiaki Mishima, and Yoshiro Tanizaki. Complexity of terminal airspace geometry assessed by computed tomography in asthma. *American Journal of Respiratory and Critical Care Medicine*, 167(3):411–417, 2003.

[93] Kishore V. Mogatadakala, Kevin D. Donohue, andCatherine W. Piccoli, and Flemming Forsberg. Detection of breast lesion regions in ultrasound images using wavelets and order statistics. *Medical Physics*, 33(4):840–849, 2006.

[94] Florent Monay and Daniel Gatica-Perez. Plsa-based image auto-annotation: constraining the latent space. In *Proceedings of the 12th annual ACM international conference on Multimedia*, pages 348–351, 2004.

[95] Yasuhide Mori, Hironobu Takahashi, and Ryuichi Oka. Image-to-word transformation based on dividing and vector quantizing images with words. In *Proceedings of First International Workshop on Multimedia Intelligent Storage and Retrieval Management*, pages 405–409, 1999.

[96] Henning Muller and Thomas M. Deserno. *Content-Based Medical Image Retrieval*, pages 471–494. Springer-Verlag Berlin Heidelberg, 2011.

[97] Henning Muller, Nicolas Michoux, David Bandon, and Antoine Geissbuhler. A review of content-based image retrieval systems in medical appli-

cations: Clinical benefits and future directions. *International Journal of Medical Informatics*, 73:1–23, 2004.

[98] Andriy Myronenko and Xubo Song. Intensity-based image registration by minimizing residual complexity. *IEEE Transactions on Medical Imaging*, 29(11):1882–1891, 2010.

[99] Y. Nakano, H. Sakai, T. Hirai S. Muro, Y. Oku, K. Nishimura, and M. Mishima. Comparison of low attenuation areas on computed tomographic scans between inner and outer segments of the lung in patients with chronic obstructive pulmonary disease: incidence and contribution to lung function. *Thorax*, 54(5):384–389, 1999.

[100] Fei Peng, Kehong Yuan, Shu Feng, and Wufan Chen. Region feature extraction of brain ct image for classification. In *The 2nd International Conference on Bioinformatics and Biomedical Engineering*, pages 2495–2498, 2008.

[101] A. Ratnaparkhi. *Maximum Entropy Models for Natural Language Ambiguity Resolution*. PhD thesis, University of Pennsylvania, Philadelphia, PA, 1998.

[102] Paulo S. Rodrigues, Gilson A. Giraldi, Marcia Provenzano, Marcelo D. Faria, Ruey-Feng Chang, and Jasjit S. Suri. A new methodology based on q-entropy for breast lesion classification in 3-d ultrasound images. In *Proceedings of the 28th Annual International Conference on Engineering*

in Medicine and Biology Society (IEEE-EMBS 2006), pages 1048–1051, 2006.

[103] Yong Rui, Thomas S. Huang, and Shih-Fu Chang. Image retrieval: Past, present, and future. In *Proceedings of International Symposium on Multimedia Information Processing*, pages 1–23, 1997.

[104] Yong Rui, Thomas S. Huang, and Shih-Fu Chang. Image retrieval: Current techniques, promising directions, and open issues. *Journal of Visual Communication and Image Representation*, 10(1):39–62, 1999.

[105] N. Sager, Carol Friedman, M.S. Lyman, E.C. Chi, C. Macleod, S. Chen, and S. Johnson. The analysis and processing of clinical narrative. In *Proceedings of the Fifth Conference on Medical Informatics*, pages 1101–1105, 1986.

[106] Berkman Sahiner, Heang-Ping Chan, Marilyn A. Roubidoux, Lubomir M. Hadjiiski, Mark A. Helvie, Chintana Paramagul, Janet Bailey, Alexis V. Nees, and Caroline Blane. Malignant and benign breast masses on 3d us volumetric images: Effect of computer-aided diagnosis on radiologist accuracy. *Radiology*, 242(3):716–724, 2007.

[107] P.L. Schuyler, W.T. Hole, M.S. Tuttle, and D.D. Sherertz. The umls metathesaurus: representing different views of biomedical concepts. *Bulletin of the Medical Library Association*, 81(2):217–222, April 1993.

[108] P. Mohana Shankar, Vishruta A. Dumane, Catherine W. Piccoli, John M. Reid, Flemming Forsberg, and Barry B. Goldberg. Computer-aided classification of breast masses in ultrasonic b-scans using a multiparameter approach. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, 50(8):1002–1009, 2003.

[109] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, August 2000.

[110] Chi-Ren Shyu, Carla E. Brodley, Avinash C. Kak, Akio Kosaka, Alex M. Aisen, and Lynn S. Broderick. Assert: A physician-in-the-loop content-based retrieval system for hrct image databases. *Computer Vision and Image Understanding*, 74(1–2):111–132, 1999.

[111] Ingrid Sluimer, Arnold Schilham, Mathias Prokop, and Bram van Ginneken. Computer analysis of computed tomography scans of the lung: a survey. *IEEE Transactions on Medical Imaging*, 25(4):385–405, 2006.

[112] Arnold W.M. Smeulders, Marcel Worring, Amarnath Gupta, and Ramesh Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, December 2000.

[113] Jae H. Song, Santosh S. Venkatesh, Emily. F. Conant, Ted W. Cary, Peter H. Arger, and Chandra M. Sehgal. Artificial neural network to aid differen-

tiation of malignant and benign breast masses by ultrasound imaging. In *Proceedings of SPIE Medical Imaging 2005*, pages 148–152, 2005.

[114] Peter Spyns. Natural language processing in medicine: an overview. *Methods of Information in Medicine*, 35(4-5):285–301, December 1996.

[115] A T Stavros, D Thickman, C L Rapp, M A Dennis, S H Parker, and G A Sisney. Solid breast nodules: use of sonography to distinguish between benign and malignant lesions. *Radiology*, 196(1):123–134, 1995.

[116] Ricky K. Taira, Vijayaraghavan Bashyam, and Hooshang Kangarloo. A field theoretical approach to medical natural language processing. *IEEE Transactions on Information Technology in Biomedicine*, 11(4):364–373, July 2007.

[117] Ricky K. Taira, Stephen G. Soderland, and Rex M. Jakobovits. Automatic structuring of radiology free-text reports. *Radiographics*, 21(1):237–245, 2001.

[118] Jia-Wei Tian, Li-Tao Sun, Yan-Hui Guo, Ying-Tao Zhang, and Heng-Da Cheng. Computerized-aid diagnosis of breast mass using ultrasound image. *Medical Physics*, 34(8):3158–3164, 2007.

[119] Tatiana Tommasi, Barbara Caputo, Petra Welter, Mark Oliver Guld, and Thomas M. Deserno. Overview of the clef 2009 medical image annotation track. *CLEF working notes 2009*, 2009.

[120] Renuka Uppaluri, Theophano Mitsa, Milan Sonka, Eric A. Hoffman, and Geoffrey Mclennan. Quantification of pulmonary emphysema from lung computed tomography images. *American Journal of Respiratory and Critical Care Medicine*, 151(1):248–254, July 1997.

[121] Aditya Vailaya, Mario A. T. Figueiredo, Anil K. Jain, and Hong jiang Zhang. Image classification for content-based indexing. *IEEE Transactions on Image Processing*, 10:117–130, 2001.

[122] C. J. van Rijsbergen. *Information Retrieval*. Butterworths, London, 2nd edition, 1979.

[123] James Z. Wang and Jia Li. Learning-based linguistic indexing of pictures with 2-d mhmms. In *Proceedings of the tenth ACM international conference on Multimedia*, pages 436–445, 2002.

[124] Jie Wang, Shimiao Li, Chew Lim Tan, Boon Chuan Pang, C.C. Tchoyoson Lim, Cheng Kiang Lee, Qi Tian, and Zhuo Zhang. Registration based ct slice indexing for severe traumatic brain injury. 2011. submitted.

[125] Yong Wang. *Automatic Image Annotation and Categorization*. PhD thesis, Queen Mary, University of London, London, United Kindom, 2008.

[126] Changbo Yang, Ming Dong, and Farshad Fotouhi. Region based image annotation through multiple-instance learning. In *Proceedings of the thriteenth annual ACM international conference on Multimedia*, pages 435–438, 2005.

[127] Wei-Li Zhang and Xi-Zhao Wang. Feature extraction and classification for human brain ct images. In *Proceedings of the sixth International Conference on Machine Learning and Cybernetics*, pages 1155–1159, 2007.

[128] Pierre Zweigenbaum, Dina Demner-Fushman, Hong Yu, and Kevin B.Cohen. Frontiers of biomedical text mining: current progress. *Briefings in Bioinformatics*, 8(5):358–375, 2007.