CONTENT BASED MEDICAL IMAGE RETRIEVAL

by

Neda Barzegar Marvasti

B.S., Electrical and Electronics Engineering, Tehran Azad University, 2006

M.S., Electrical and Electronics Engineering, Malek Ashtar University, 2010

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

Graduate Program in Electrical and Electronics Engineering

Boğaziçi University

2017

CONTENT BASED MEDICAL IMAGE RETRIEVAL

APPROVED BY:

Assoc. Prof. Burak Acar (Thesis Supervisor)	.....
Prof. Murat Saraçlar	.....
Assoc. Prof. Arzucan Özgür	.....
Assoc. Prof. Engin Erzin	.....
Assoc. Prof. Emine Şule Yazıcı	.....

DATE OF APPROVAL: 03.11.2017

ACKNOWLEDGEMENTS

First and foremost, I would like to sincerely thank my Ph.D. advisor Assoc. Prof. Burak Acar, who has patiently and with his immense knowledge led me through my entire research. I truly appreciate all his contributions of time, opinions and thoughts, and the financial support to make my Ph.D. study an exciting and productive experience.

I would like to thank my jury committee members Prof. Murat Saraçlar, Assoc. Prof. Arzucan Özgür, Assoc. Prof. Engin Erzin , and Assoc. Prof. Emine Şule Yazıcı for their insightful comments, which have remarkably increased the quality of my Ph.D. dissertation.

I appreciate Dr. Erdem Yoruk's thoughtful and innovative comments, suggestions and discussions throughout my research. I also thank Dr. Suzan Üskudarlı, Assoc. Prof. Taylan Cemgil, and Dr. Ceyhun Burak Akgül for their useful comments.

I want to thank all my fellow labmates at the VAVlab, specially Demet Yüksel and Akdülkadir Yazıcı for creating a warm and friendly environment at our workplace.

I would like to show my deepest gratitude to my family: my mother, who I truly wish were here among us, my beloved father and my lovely sisters Nasrin and Nooshin for their eternal love, continuous support, and encouragement in my whole life.

Last but not least, I thank my beloved husband, Reza, who has been a constant source of support and encouragement during the challenges of graduate school and our life. I am truly grateful for having you in my life!

ABSTRACT

CONTENT BASED MEDICAL IMAGE RETRIEVAL

Fast technological developments of different medical imaging and data collection techniques increase the expectation of more accurate interpretations and diagnoses of radiologists. However, to carefully analyze the resulting big medical data, reliable and fast systems are needed. Content-based medical image retrieval (CBMIR) is a valuable technique to assist radiologists by identifying similar images in a large archive. However, due to the huge semantic gap between low-level image features and high-level semantic features, the challenge of retrieving similar images utilizing the high-level user specified semantic labels, which are closer to the users understandings and interpretations, has attracted great interest from various researches. In this dissertation, an iterative search and retrieval scheme to identify similar images from a database of 3-dimensional liver computed tomography (CT) images is proposed via utilizing the combination of lesion and liver related semantic features and patients' metadata. At each retrieval iteration, the lesion related concepts are annotated in a specific order through a proposed computer aided medical image annotation (CMIA) scheme. The proposed radiologist-in-the-loop semi-automatic CMIA is based on a Bayesian tree structured model, linked to RadLex, to exploit the inter-dependencies between concepts to update the full annotation process and to guide the radiologist to input the most critical information at each iteration. Results show the effectiveness of this model-based interactive annotation scheme compared to the domain-blind models, as well as its advantage in the performance of the retrieval system, where a few number of manual annotations can significantly boost the retrieval accuracy. Moreover, better retrieval performance is achieved by incorporating a small contribution of the non-lesion data.

ÖZET

İÇERİK TABANLI TIBBİ GÖRÜNTÜ BULGETİR

Farklı tıbbi görüntüleme ve veri toplama teknikleriyle ilgili teknolojik gelişmeler radyologların daha doğru yorum ve teşhis beklentilerini arttırmaktadır. Bununla birlikte, ortaya çıkan büyük tıbbi verilerin dikkatli bir şekilde analiz edilmesi için güvenilir ve hızlı sistemler gereklidir. İçerik tabanlı tıbbi görüntü bulgetir (CBMIR) benzer bir görüntüyü büyük bir arşivde bularak radyologlara yardımcı olan değerli bir tekniktir. Bununla birlikte, düşük düzeyli imge özellikleri ve üst düzey semantik özellikler arasındaki büyük anlamsal boşluk nedeniyle, kullanıcıların anlayışlarına ve yorumlarına daha yakın olan, üst düzey kullanıcı tarafından belirlenen semantik etiketler kullanarak benzer görüntülerin bulgetir sistemi gerçekleştirilmesi araştırmacılar tarafından büyük bır ilgi görmüştür. Bu tezde, lezyon ve karaciğer ile ilgili semantik özellikler ve hastaların meta-verileri kombinasyonunu kullanarak 3-boyutlu karaciğer CT görüntülerinin bir veritabanından benzer görüntüleri tanımlamak için tekrarlayan bir arama ve bulgetir sistemi önerilmiştir. Her bulgetir iterasyonunda, lezyonla ilgili özellikler önerilen bir bilgisayar destekli tıbbi görüntü anotasyon (CMIA) şemasıyla spesifik bir sırayla anotasyon edilmektedirler. Önerilen döngü-deki-radyolog yarı otomatik CMIA, RadLex ile bağlantılı Bayes ağaç yapılı modeline dayanıyor ve kavramlar arasındaki karşılıklı bağımlılıkları kullanarak radyologun en kritik bilgiyi girmesine rehberlik etmektedir. Sonuçlar önerilen anotasyon modelin daha etkili ve verimli olduğunu göstermektedir. Bu üstünlük bulgetir performans açısından da görülmektedir. Bir kaç manuel anotasyon bulgetir doğruluk oranını belirgin bir şekilde artırabilir. Ayrıca metadata'nın küçük bir katkısı ile daha iyi arama ve bulgetir performansı elde edilir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

$E(\mathbf{r})$	Liver boundary-map generation
$\mathbf{F_I}$	Vector of low level image features
$\mathbf{F_M}$	Vector of metadata
$\mathbf{F_S}$	Vector of semantic features
$\mathcal{G}_D$	The domain-aware Bayesian network designed based on domain knowledge
$\hat{\mathcal{G}}_D$	The modified domain-aware Bayesian network designed based on domain knowledge using the ms-EM algorithm
$\mathcal{G}_{MWST}$	The maximum weight spanning tree Bayesian network
$\mathcal{G}_N$	The *naïve* Bayesian network
$\hat{\mathcal{G}}_N$	The modified *naïve* Bayesian network using the mS-EM algorithm
$I(\mathbf{r})$	Initial CT image
$I_0(\mathbf{r})$	Preprocessed CT image
$I_b(\mathbf{r})$	Liver edge-map image
$I_{P_{bg}}$	Background probability image
$I_p(\mathbf{r})$	Liver Profile image
$I_{P_{fg}}$	Foreground probability image
k	Iteration index
$P_s(\mathbf{r})$	Poisson solution of image
$\mathbf{r}$	Position vector
s	Vesselness scale
$S_0(\mathbf{r})$	Initial liver segmentation mask
$S_E(\mathbf{r})$	Lesion mask
$S_L(\mathbf{r})$	Liver mask
$S_V(\mathbf{r})$	Vessel mask
X_C	Concept variables
X_L	Latent variables
X_S	SVM observations

Y_L	Annotations of the second radiologist
Θ	Parameter set of the BN
μ	Mean
ρ	Similarity value
σ	Standard deviation

LIST OF ACRONYMS/ABBREVIATIONS

2D	Two Dimensional
3D	Three Dimensional
AL	Active Learning
BIC	Bayesian Information Criterion
BN	Bayesian Network
BP	Belief Propagation
CBIR	Content Based Image Retrieval
CBMIR	Content Based Medical Image Retrieval
CES	Clinical Experience Sharing
CIA	Computer-Aided Image Annotation
CMF	Conventional Multi-Scale Selection Filter
CMIA	Computer-Aided Medical Image Annotation
CPT	Conditional Probability Table
CT	Computed Tomography
DAG	Directed Acyclic Graph
EM	Expectation Maximization
EMD	Earth Mover's Distance
GMM	Gaussian Mixture Model
KNN	K-Nearest Neighbor
MD	Medical Doctor
MI	Mutual Information
MRI	Magnetic Resonance Imaging
mS-EM	Modified Structural Expectation Maximization
MTL	Multi-Task Learning
MWST	Maximum Weight Spanning Tree
ONLIRA	Ontology of Liver in Radiology
RBF	Radial Basis Function
ROI	Regions of Interest

SEM	Structural Expectation Maximization
SVM	Support Vector Machine
VE	Variable Elimination
VST	Visual Semantic Terms

1. INTRODUCTION

Due to the fast technological developments in the medical imaging within the last decades, there is a huge increase in the amount of medical data per study, which increase the expectation of high interpretation and diagnosis accuracy of the radiologists. However on the other hand, manual examination of large amount of medical images has become infeasible, as it is highly time consuming, error-prone, subjective, and costly. The performance and accuracy of a radiologist is highly dependent on his/her experience and training, and many other factors may degrade the manual examination process including fatigue and distraction. Moreover, radiologists' interpretation and diagnosis accuracy vary from person to person, or even from time to time. For ambiguous and rare cases, radiologist often asks for the second opinion or searches for previous similar cases, or rely on the reputable references in the literature. However, managing and searching through these large data and repositories is increasingly complex and tedious, therefore utilizing the computer-based assistance is an inevitable solution to administer the procedure while maintaining the interpretative accuracy. To this end, many machine learning approaches have been proposed in the literature [2], which offer some promising diagnostic techniques, enabling radiologists to examine various aspects of the cases with a higher diagnostic accuracy .

The challenges of big data analysis in medicine can be appropriately addressed through the content-based medical image retrieval (CBMIR) systems, which aim to retrieve similar medical images from a large database by integrating computer-based assistance into the medical image interpretation. The application of CBMIR system falls in the realm of clinical experience sharing (CES) [3], in which similar medical images can be obtained for both diagnostic aids and medical educational purposes. Conventional CBIR techniques focus only on the image content and are more suited for multimedia application where the content is defined at a coarser level. However, CBMIR focuses on subtle differences that have diagnostic value. This can be achieved by utilizing not only the image content but also the medical case information. The image content is represented with a comprehensive set of low level computer generated

features, while medical case information is represented by an expert's high-level interpretations (semantic concepts), non-image metadata (patient's demographics, clinical history, lab test results, physical examination, and etc.), as well as the knowledge models. This is because of the fact that in the diagnostics practice, radiologists always exploit the broader patient-specific knowledge rather than only concentrating on the image itself. All in all, implementing a medical image retrieval system requires higher levels of endeavor for the data processing than the conventional image search and retrieval systems [4].

The overall objective of this dissertation is to implement a CBMIR system to retrieve the 3D computed tomography (CT) liver images containing lesions similar to a query lesion that can be utilized for comparative diagnosis and medical education via sharing the collective experience over the medical community. Liver has been chosen as the application domain mainly because of its importance and also relative ease of data collection. The main contributions of this dissertation are (i) introducing a domain-aware Bayesian network (BN) based lesion model and (ii) proposing a computer-aided medical image annotation (CMIA) via a novel radiologist-in-the-loop scheme, and (iii) assessing the efficiency of the proposed CMIA approach in CBMIR system. The semantic concepts are generated using the ontology of liver for radiology (ONLIRA) [5], which are related to the Radlex terms. The proposed scheme considers the relation between semantic concepts in their annotations and models these relations using the tree-based BN structure. The search and retrieval part is based on the CMIA output as well as medical case information supplies.

This dissertation is organized as follows: Chapter 2 reviews the methods in CBIR and their transitions to the state of the art methods in CBMIR. Chapter 3 discusses the methods employed to realize a CBMIR. Experiments and results are presented in Chapter 4 and discussed in Chapter 5. Finally, Chapter 6 concludes the dissertation.

2. STATE OF THE ART

2.1. Content Based Image Retrieval (CBIR)

Over the last decades, the availability of big data in visual and multimedia information have led to several projects and developments in the area of content search technologies offering far more beyond the simple text-based search. This is a challenging task due to the usual vagueness and controversy of the visual interpretation of the content of images. Hence, teaching the image interpretation to machines requires cumbersome and complex algorithms. The years of 1994-2000 can be thought of as the initial phase of research and development on the content based image retrieval (CBIR) [6].

CBIR, which is widely used in image and multimedia applications including automatic photo tagging, visual search on the web and visual shopping, is a system that helps users to find similar image contents in a database of variety of different multimedia and images, which saves a tremendous amount of browsing time [4]. Anything ranging from image similarity function to robust image annotation engines falls under the purview of CBIR [7].

A generic CBIR system has two main components: image descriptors and the similarity measures. Image descriptors are image features that reveal the visual content that exist in image pixels/voxels by encoding their numerical representation. This visual information includes *color* and *texture* information, *location* of objects in the image or with respect to each other, and *shape* properties [8]. The image similarity measures mathematically analyze image descriptors across different images to compute their distances [8]. The shorter the distance, the more similar they are. Ideally, these measures should perform similar to human interpretation and similarity criteria. For a given query image, the system provides related images in the database, while the most similar ones are ranked ahead in the queue. The choice of the similarity metric depends on the image descriptor type/representation. For instance, for high dimensional feature

vectors, the conventional metric is the Euclidean distance [9], in which each image is represented with a vector containing its all features. For histogram-based features with different number of bins, the Earth mover's distance (EMD) has been proposed [10], and for image features that are represented by graphs, the graph matching theories have been introduced [11]. Another approach is to use pre-trained classifiers to retrieve the semantic meaning of an image from its visual image representation, which is a task dependent approach and requires the set of fixed labels. Then the similarity between the labels is used as the similarity metric [4].

2.2. Content Based Medical Image Retrieval (CBMIR)

Content based medical image retrieval refers to the retrieval of medical images, e.g. X-ray, ultrasound, magnetic resonance imaging (MRI), and CT scans, with similar medical content. The first content based medical image retrieval (CBMIR) approaches were direct application of CBIR systems designed for multimedia domain for retrieving medical images of the same modality, body part or orientation [12, 13]. These systems did not focus on diagnostic image content. Despite CBIR's success in the multimedia, such CBMIR systems has had small impact on radiology [4], which is potentially due to the different definitions of the content in multimedia and the medicine. In CBIR, the subtle differences between images are often ignored and macro differences are considered, i.e. two images are similar, if they belong to the same group, such as landscape images or cars. However in CBMIR, subtle differences play a crucial role as they indeed are the key points of clinical diagnoses, which must not be ignored. Hence, the main issue in CBMIR is defined as assessing the similarities or dissimilarities between images of the same modality, same organ and same disease, but with different pathological observations [8].

Since images with similar diagnosis may show different visual patterns or anatomical structures, it is fundamental to design descriptive and discriminative feature descriptors to retrieve only images with similar diagnosis. Many CBMIR applications exert the improvement and specialization of low-level image features to overcome this issue by capitalizing the subtle differences between medical images. Generally, features

are extracted from regions of interest (ROI), which is defined manually or segmented automatically, and the main focus is on the content of that specific area including its texture or the shape [14]. Ballerini *et al.* have utilized color covariance-based features and texture features as low level image features to retrieve similar skin lesions. The combination of Bhattacharyya and Euclidean distance metrics are computed to measure the similarity by Ballerini *et al.* [15]. Quellec *et al.* have proposed medical image retrieval based on the wavelet-based image signatures and similarity distance [16]. This method has been tested on diabetic retinopathy follow up and also screening mammography. In [17], Wei *et al.* evaluate the digital mammograms retrieval performance using different shape and margin features of the lesion masses and propose a hierarchical similarity measurement scheme, which replicates the radiologist evaluation procedure. Murala *et al.* propose local mesh patterns for biomedical image retrieval application to capture the underlying relationship among the surrounding neighbors for a given referenced pixel [18]. Tsochatzidis *et al.* propose a new texture descriptor to capture the margin and core specific mammographic mass properties [19].

Despite all the effort for increasing the discrimination power of low level image features, individually utilizing these features results in poor retrieval performance, which is due to the semantic gap between these features and the radiologist's high-level interpretation. The semantic gap is the lack of conformity between the extracted information from a specific visual data and its expert interpretation, which leads to the issue of user-dependent interpretations [20]. Visual image features are generally low-level compared to high-level concepts of images and are not descriptive enough to capture the semantic concept that users are interested in. Moreover, radiologists always interpret an image considering clinical disease-specific metadata as well, which consists of information such as medical history, laboratory results, and physical examinations. Hence, the domain knowledge needs to be integrated into the specialized query engine. Semantic annotations describe the high-level visual content (same as lesion shape, contrast pattern, and composition), which are related to radiologists' high level understanding and interpretations, and therefore can improve the performance of the retrieval process via facilitating the expert to search for the specific terms that he/she is interested in [21].

Recently, research has shifted from designing sophisticated low-level visual feature descriptors to bridge the semantic gap by enriching the semantic features (concepts) and incorporating them into the search and retrieval system. The generic method is to describe the image with a bag of visual concepts, which may be defined manually or automatically learned. Ma *et al.* have proposed the combination of visual and semantic features in a context-sensitive graph-based similarity measure for the retrieval of CT images of lung patients [22]. Napel *et al.* performed automatic 3D CT liver lesion search and retrieval, where similarities are computed based on both low-level and high-level manually annotated semantic features using the Manhattan distance [23]. Depeursinge *et al.* have performed case-based lung image search and retrieval utilizing both texture-based visual features and text-based clinical data in the similarity search [24].

Due to the effective enrichment of semantic features with visual concepts, the retrieval can be performed by applying techniques similar to the commonly used text-related methods in many successful information retrieval (IR) systems [25]. Rahman *et al.* [26] propose a query specific similarity search and retrieval scheme, which adapts itself on-line based on the query type and also via user feedback. In their scheme, SVM has been used as the classifier for computing labels of the image. Kumar *et al.* [27] have proposed a retrieval system of the multi-modal images based on tumor location, in which a graph represents the anatomical relations between tumors and other organs and their proximities. Kurt *et al.* [21] proposed a retrieval system based on semantic features, in which the Radlex ontological structure is used to compute the similarity metric. Zhang *et al.* [28] proposed a retrieval system based on the bag of visual words, where the less meaningful words are pruned using the probabilistic latent semantic analysis (pLSA). Rahman *et al.* describe images by color and texture features extracted from local patches in the images. They proposed a SVM based classification method to drive visual descriptors representing the frequency of image features while preserving their local spatial relations. Euclidean distance is applied as the similarity metric [29]. All of these works have vividly demonstrated the importance of the richness and accuracy of semantic features in the performance of the medical image retrieval.

2.3. Computer-Aided Medical Image Annotation (CMIA)

Computer-aided image annotation (CIA) is the process of automatically assigning visual words to digital images [30–32]. The main idea behind the CIA techniques is to automatically learn the mapping between semantic concepts and low-level image features from a large number of training image samples, and then apply the learned model to label new images.

Incorporating the concept of CIA into the medical domain leads to computer-aided medical image annotation (CMIA) methods, which utilize common lexicons and can be described as the task of assigning high-level labels (e.g. RadLex terms) to semantic concepts using low-level image features. They can be used as image tags in medical image/record search/retrieval systems to generate structured radiology reports, which is the primary barrier in exploiting data analytic tools to their full potential in medical informatics.

The pioneering examples of CMIA focused on annotating medical images with very high level concepts such as the imaged body parts and the modality used [33]. These systems were indeed, to a great extent, extension of applications of general purpose automatic image annotation methods [34]. Organizers of the ImageCLEF workshop series established their first medical image annotation challenge in 2005 with a similar goal, which was later developed to semantic annotations of medical images in 2014 [13, 35]. CMIA methods can be broadly categorized as individual and joint methods. The former approaches individually estimate the (annotation) value for each (semantic) concept using low-level image features, whereas the latter approaches exploit the inter-relations of (semantic) concepts as well.

The predominant individual CMIA approaches are multi-class classifiers such as SVM [36–38], bag of visual words [39], logistic regression [40], CBIR algorithms [41], and artificial neural networks [42]. Dimitrovski *et al.* used a combination of textual and visual features via two different SVM classifiers to individually annotate each image's modality [36], whereas Alahmer *et al.* used an SVM classifier to annotate the

lesion malignancy (benign/malignant) [37]. Depeursinge *et al.* suggested to use visual semantic terms (VST) of liver lesions that are learned from the image using linear combinations of high-order steerable Riesz wavelets using SVMs for each VST [38]. They also automatically derived a data-driven ontology of VSTs, which is later used for the retrieval of similar images [21]. Mougiakakou *et al.* trained five neural networks (NNs) for a 4-class classification of lesion types [42]. Kumar *et al.* employed a CBIR based on low-level image features and inferred the annotation of the test image from the annotations of the retrieved images [41].

Few approaches that pursue to exploit the inter-dependencies can be grouped as hierarchical classifiers and multi-task learning (MTL). In the former case, a hierarchy of concepts is built and individual classifiers are trained for each concept using the images related to that specific concept. During testing, hierarchical one-concept-at-a-time classification is performed starting from the most general concept (highest in hierarchy), which however is potentially prone to error propagation [43]. Mueen *et al.* used a hierarchical representation of IRMA codes to annotate X-ray images, where separate SVM classifiers were trained for each level of the hierarchy [44]. Ko *et al.* trained random forest classifiers to annotate X-ray images at lower levels of a hierarchical representation of IRMA codes, whereas inter-relations between IRMA codes are exploited for annotations at the higher levels [45]. Dimitrovski *et al.* used a single multi-concept classifier that implicitly models the hierarchy between concepts. Their approach employs a single classifier based on ensembles of predictive clustering trees for hierarchical multi-label classification to annotate medical X-ray images with IRMA codes [46]. Zare *et al.* pursued a Bayesian approach and proposed the probabilistic Latent Semantic Analysis (pLSA) scheme to annotate medical images [47], in which a set of latent concepts/topics are assumed in the generative process of visual and textual words from every image. pLSA estimates the posterior probability of each semantic feature and uses an SVM classifier for final annotation. On the other hand, the MTL-based approaches simultaneously learn several joint tasks (i.e. semantic annotations) by using shared hidden (regression) layers [48]. While they are primarily applied in the non-medical domain [49–52], recently, Chen *et al.* proposed MTL approach in semantic rating of lung nodules in CT images [53].

Active learning is another viable strategy for medical image annotation. It is commonly used in computer vision, image analysis, and pattern recognition. Here, the goal is to recover the ground-truth via sequential user feedback or some form of actively chosen elementary observations, which, similar to a "20 questions game", remove the uncertainty about the semantics as fast as possible and in accordance with the accumulated evidence. Such methods are primarily applied outside of the medical domain, both in fully automated and human-in-the-loop systems, such as road tracking [54], complex scene interpretation [55], fine-grained image classification [56], and volumetric image segmentation [57].

In this dissertation, a CBMIR system is proposed via a radiologist-in-the-loop CMIA scheme [58] that utilizes a Bayesian tree structured domain-aware model.In this dissertation, a CBMIR system is proposed via a radiologist-in-the-loop CMIA scheme [58] that utilizes a Bayesian tree structured domain-aware model. This model represents the inter-dependencies of the concepts. The system uses individual concept-specific classifiers trained on low-level image features to get noisy individual concept observations. The domain-aware model is used to solve for the optimal annotation for *lesion related* concepts jointly, as well as to guide the active learning component of the proposed scheme. In brief, a domain network model is employed in an interactive annotation/active learning scheme that takes classifier responses as initial noisy individual observations and sequentially requests manual input from the user for the most informative concept among the pending ones, to refine the automatic annotation. Hence, the proposed system combines the power of individual SVM classifiers with a statistical model that encodes concept dependencies, thereby, iteratively refining the complete multi-concept annotation of the radiological image. Unlike the Bayesian schemes, which model the semantic terms and their relations as they appear in medical literature [59], a bottom-up approach is pursued and used the semantic terms expressed in a domain ontology (ONLIRA) [5] linked to RadLex terms.

3. METHODS

3.1. Overview

In this dissertation, a semantic-based CBMIR system is proposed via a radiologist-in-the-loop CMIA scheme that utilizes a Bayesian tree structured domain-aware model [58]. As illustrated in Figure 3.1, the query image ($I(\mathbf{r})$), $\mathbf{r} \in \Re^3$, is pre-processed ($I_0(\mathbf{r})$), segmentation masks are generated ($S_L(\mathbf{r}), S_V(\mathbf{r}), S_E(\mathbf{r})$) and low-level image features are extracted ($\mathbf{F}_I$). The individual support vector machine (SVM) classifiers produce the initial image annotations ($\mathbf{X}_S$), which are modified ($\mathbf{X}_C$) using a BN model that represents their interrelations. The last state is an iterative scheme to improve the annotation and retrieval task through providing manual annotations ($\mathbf{X}_E$) as evidence.

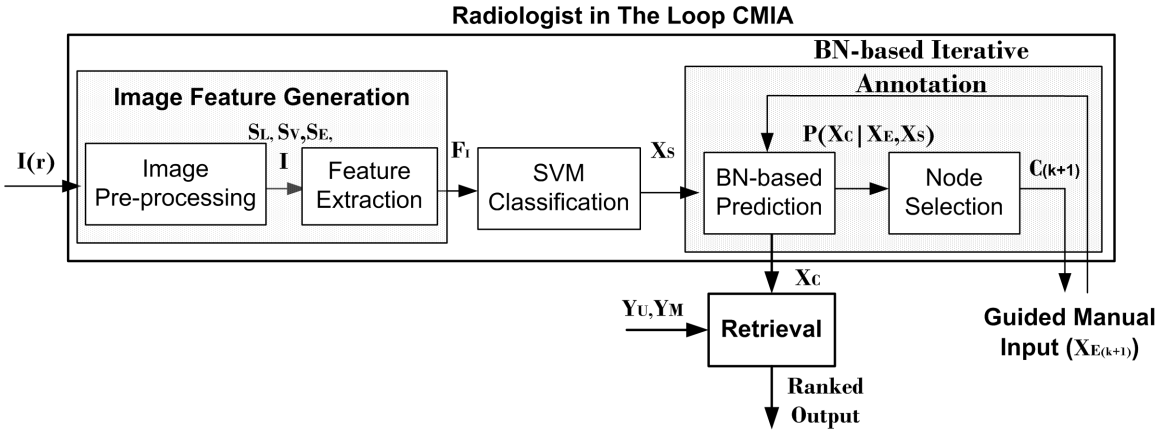


Figure 3.1. Iterative search and retrieval scheme.

3.2. Image Pre-Processing

The input 3D abdominal CT image is manually cropped to isolate the volume of interest, which is liver. Then, the cropped volume is isotropically re-sampled and processed by the edge enhancing diffusion filter for noise suppression to get the image I_0 [60]. As shown in Figure 3.2, the prepared image I_0 is given to the segmentation block which is composed of three steps: (i) Liver segmentation, (ii) Vessel segmentation,

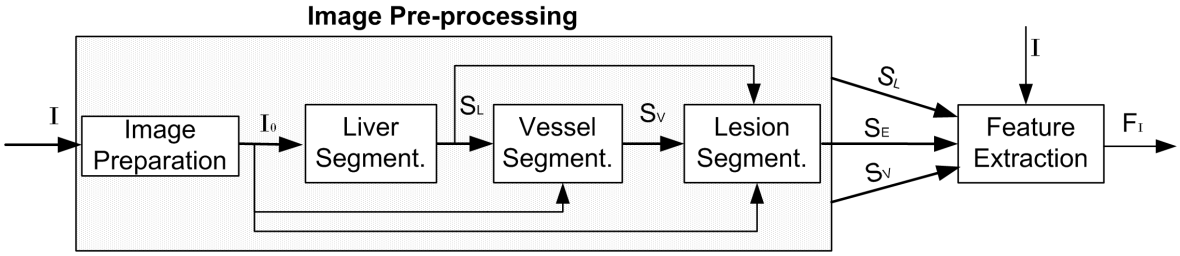


Figure 3.2. Block diagram of image feature generation scheme.

and (iii) Lesions segmentation.

3.2.1. Liver segmentation

The 3D liver segmentation method proposed by S. Çimen is used with a modified MinCut/MaxFlow graph-cut stage [61, 62]. The full pipeline is reviewed for the sake of completeness. The method uses the Poisson equation based shape representation and independently trains k-NN classifiers on 1D across-the-boundary CT profiles to generate liver-boundariness maps [63]. As shown in Figure 3.3, the liver segmentation pipeline consists of three phases: (i) Initial liver segmentation ($S_L^0(\mathbf{r})$) , (ii) Liver-boundariness map generation ($I_b(\mathbf{r})$), and (iii) Graph-cut based fine segmentation ($S_L(\mathbf{r})$).

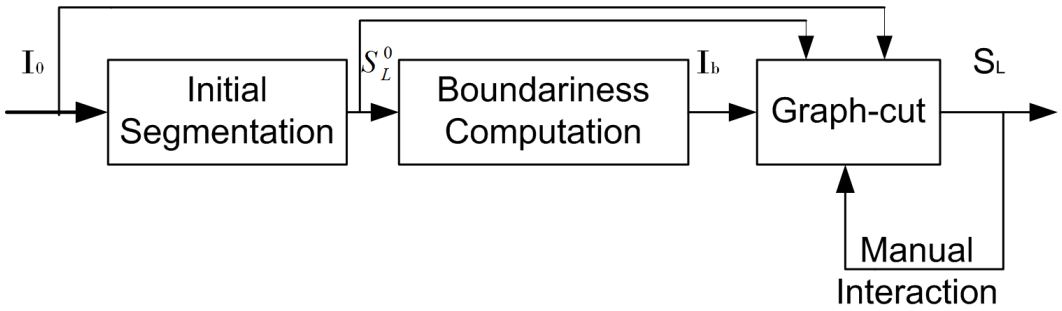


Figure 3.3. Block diagram of the liver segmentation method.

- Initial liver segmentation: Initial liver segmentation ($S_L^0(\mathbf{r})$) is achieved by an adaptive thresholding of the CT volume. The intensity distribution of the prepared image ($I_0(r)$) voxel value is modeled with a 5-component Gaussian mixture model (GMM) [64]. The component with the least variance among the two top ranked components sorted in a descending order with respect to their peak values

is identified as the liver parenchyma. Its mean and variance are set as the *liver mean* (μ_L) and the *liver variance* (σ_L^2). The liver parenchyma mask is defined as

$$S_{ini}(\mathbf{r}) = \begin{cases} 1 & t_l < I_0(\mathbf{r}) < t_u \\ 0 & \text{o.w.} \end{cases} \tag{3.1}$$

$$t_l = \mu_L - (1.5 \times \sigma_L \times g(\mathbf{r})) \tag{3.2}$$

$$t_u = \mu_L + (3.0 \times \sigma_L \times g(\mathbf{r})) \tag{3.3}$$

where t_u and t_l are the upper and lower adaptive thresholds and $g(\mathbf{r})$ is a gradient-based edge map which is 0 at the edges and 1 in non-edge areas. $S_{ini}(\mathbf{r})$ is further post-processed using a median filter with a kernel of size $7 \times 7 \times 7$, followed by morphological erosion with a 3D symmetric structure element of radius equal to two voxels. The largest connected component is retained as the initial liver segmentation mask $S_L^0(\mathbf{r})$.

- Liver-boundariness map computation: Using the initial segmentation $S_L^0(\mathbf{r})$, a compact representation of the shape of the liver is obtained following the Poisson equation based method proposed for 2D shapes by Gorelick *et al.* [65]. The Poisson equation is solved with the boundary conditions defined as

$$\begin{cases} P_s(\mathbf{r}|S_L^0(\mathbf{r}) == 1) = 1 & \text{, inner boundary} \\ P_s(\mathbf{r}||\mathbf{r} - \mathbf{o}| > K) = 1 & \text{, outer boundary} \end{cases} \tag{3.4}$$

where $\mathbf{o}$ is the center of mass of $S_L^0(\mathbf{r})$ and K is a sufficiently large (essentially larger than the bounding box) constant radius used to define the outer boundary as a spherical surface at infinity. Thus, $\nabla P_s(\mathbf{r})$ would be oriented with the liver surface normal to a great extent. $\nabla P_s(\mathbf{r})$ directions are used to get the 1D voxel profiles for liver-boundaryness map computation. The 1D voxel profiles are extracted at each point $\mathbf{r}$ to generate the the boundary profile image $\mathbf{I}_p(\mathbf{r})$ as:

$$i_p(\mathbf{r},k) = \begin{cases} I_0(\mathbf{r} + (k \times \frac{\nabla P_s(\mathbf{r})}{|\nabla P_s(\mathbf{r})|})) & , \quad |\nabla P_s(\mathbf{r})| \neq 0 \\ 0 & , \quad \text{o.w.} \end{cases} \tag{3.5}$$

$$\mathbf{I}_p(\mathbf{r}) = [i_p(\mathbf{r},-N)\ i_p(\mathbf{r},-N+1)\ \cdots\ i_p(\mathbf{r},N)]$$

where $i_p(\mathbf{r},k)$ is the vector field of sampled image profiles ($N = 4$). The profile lines at each point of the liver boundary in training images are used to train a KNN (K-nearest neighbor) classifier. MICCAI 2007 Liver Segmentation Grand Challenge (SLIVER07) training dataset is used for training [66]. Hence, the liver boundariness map $I_b(\mathbf{r})$ for the given image is computed as the probability of each voxel being on the liver boundary using this trained classifier.

- Graph-cut segmentation: The MinCut/MaxFlow graph-cut algorithm has been employed for the final segmentation [62]. Image is represented by a graph $G =< \nu, \xi >$, where nodes (ν) correspond to the voxels in the image and edges (ξ) are the weighted undirected edges between the voxels themselves (n-links) or the voxels and abstract foreground/background nodes (s/t-links), representing the pairwise (dis-)similarities. The algorithm optimally divides the graph into two disjoint graphs corresponding to the foreground and background, whereas the removed edges correspond to the minimum cut $C \subset \xi$. The weights of n-links $B_{i,j}$ between nodes i and j are defined as

$$B_{i,j} = \begin{cases} \frac{-ln(I_b(\mathbf{r_i}))}{|\mathbf{v}_{i,j}|} & , cos(\theta_{i,j}) \geq cos(\theta_{j,i}) \\ \frac{-ln(I_b(\mathbf{r_j}))}{|\mathbf{v}_{i,j}|} & , cos(\theta_{i,j}) < cos(\theta_{j,i}) \end{cases} \tag{3.6}$$

where $\mathbf{v}_{i,j} = \mathbf{r_i} - \mathbf{r_j}$ and $\theta_{i,j}$ is the angle between $\mathbf{v}_{i,j}$ and $\nabla P_s(\mathbf{r_i})$. The initial segmentation $S_L^0(\mathbf{r})$ is used as the foreground hard constraint to solve the graph-cut. Note that there is no background hard constraint.

Edge weights ($R_{i,f}$, $R_{b,i}$) between node at ($\mathbf{r_i}$) and the terminal nodes (f: foreground, b: background) are computed using a unimodal Gaussian distribution model. The foreground probability image ($I_{P_{fg}}$) is a unimodal Gaussian distribution with $\mu = mean(I_0(S_L^0(\mathbf{r}) == 1))$ and $\sigma^2 = var(I_0(S_L^0(\mathbf{r}) == 1))$. Then the

weights are computed as:

$$R_{i,f} = \begin{cases} K & , S_L^0(\mathbf{r_i}) = 1 \\ -ln(I_{P_{fg}}(\mathbf{r_i})) & , o.w. \end{cases} \tag{3.7}$$

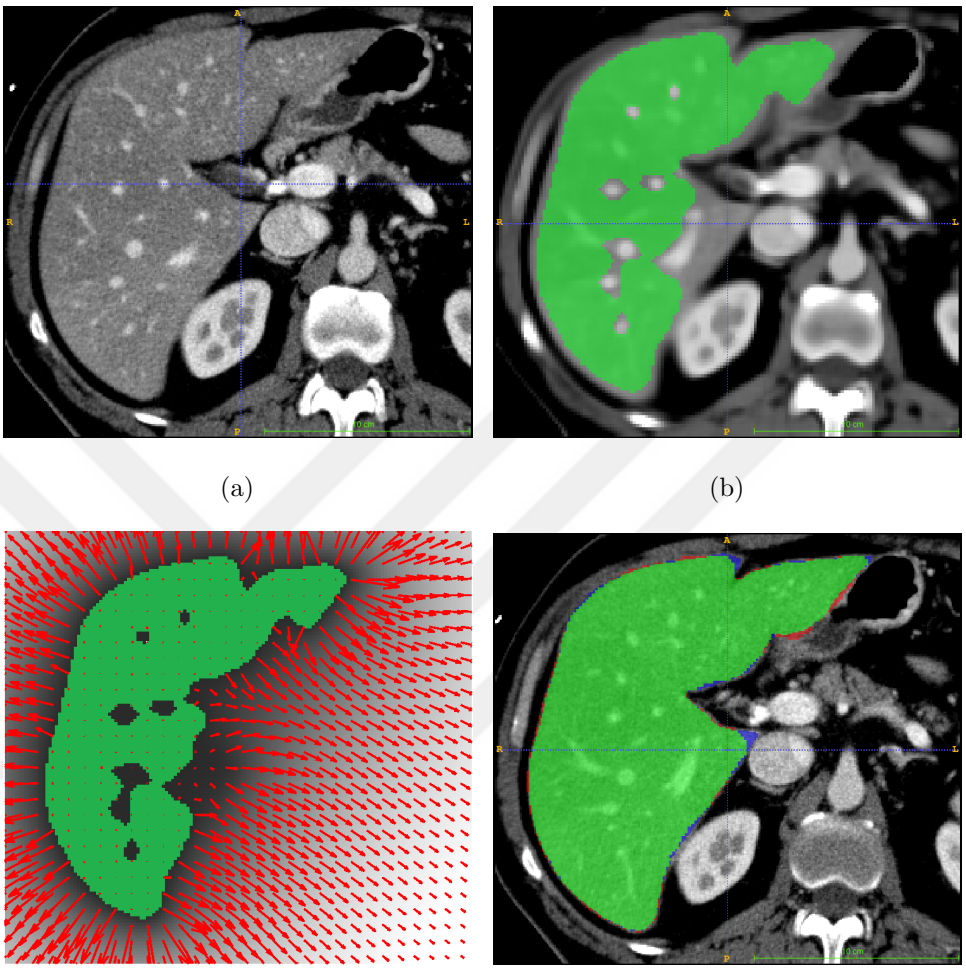
$$R_{i,b} = \begin{cases} 0 & , S_L^0(\mathbf{r_i}) = 1 \\ -ln(1 - I_{P_{fg}}(\mathbf{r_i})) & , o.w. \end{cases} \tag{3.8}$$

where K is a sufficiently-large constant that prohibits the corresponding edge to be in the cut C. This iterative process is manually checked and the foreground/background images are edited if needed. The graph-cut is applied iteratively until the final segmentation is satisfactory. The binary foreground mask is further processed by means of connected component analysis and the largest component is retained as the final liver segmentation mask, $S_L(\mathbf{r})$.

Figure 3.4 illustrates the visual segmentation results tested in MICCAI 2007 Liver Segmentation Grand Challenge (SLIVER07) dataset [66]. Figures 3.4(a) and 3.4(b) illustrate the original image and the initial segmentation mask applied on the diffusion filtered image, respectively. Figure 3.4(c) shows a 2D view of the Poisson shape representation (gray-level image) of the initial segmentation mask (green) and the Poisson gradient direction (red lines) at each voxel. Figure 3.4(d) shows the final segmentation mask together with the ground truth mask. The false negative, true positive and false positive voxels are shown in blue, green and red respectively.

3.2.2. Vessel segmentation

The multi-scale vesselness based vessel segmentation method proposed by Frangi *et al.* is performed inside the liver with a modified MinCut/MaxFlow graph-cut algorithm and a proposed image guided morphological filtering [62,67,68]. The full pipeline is depicted in Figure 3.5.



(a) (b)

(c) (d)

Figure 3.4. Liver segmentation: (a) Original image, (b) initial liver mask (green), (c) profile lines (red) and (d) final liver mask on true segmentation mask, (green) correctly segmented, (blue) false negative, and (red) false positive.

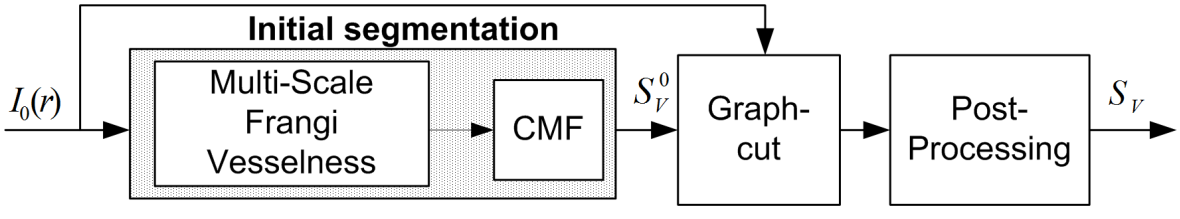


Figure 3.5. Block diagram of the vessel segmentation method.

- Initial vessel segmentation: The multi-scale vesselness approach captures the vessels via utilizing the eigen analysis of the Hessian of the processed image $I_0(\mathbf{r})$ at different scales (s). The Hessian of $I_0(\mathbf{r})$ is the $2nd$ derivative of it which can be defined as its convolution with the $3D$ Hessian of a Gaussian kernel. The Gaussian kernel is defined as $G(\mathbf{r}) = \frac{1}{(\sqrt{2\pi s^2})^3} \exp \frac{-|\mathbf{r}|^2}{2s^2}$. The eigenvalue of the k_{th} normalized eigen vector of the Hessian of $I_0(\mathbf{r})$ is defined as $\lambda_k(\mathbf{r}, s)$, where $|\lambda_1| \leq |\lambda_2| \leq |\lambda_3|$. Frangi *et al.* defined the vesselness image at every scale (s) as [67]

$$V_o(s) = \left\{ \begin{array}{lll} 0 & , & \lambda_2 > 0 \quad \text{or} \quad \lambda_3 > 0 \\ (1 - \exp(\frac{-R_A^2}{2\alpha^2})) \exp(\frac{-R_B^2}{2\beta^2})(1 - \exp(\frac{-S^2}{2\gamma^2})) & , & \text{o.w.} \end{array} \right\} \tag{3.9}$$

$$R_A = \frac{|\lambda_2|}{|\lambda_3|} \tag{3.10}$$

$$R_B = \frac{|\lambda_1|}{\sqrt{|\lambda_2 \lambda_3|}} \tag{3.11}$$

$$S = \sqrt{\sum_{j=1:3} \lambda_j^2} \tag{3.12}$$

where R_A is the parameter to distinguish plate-like and tubular-like structures with a value close to 0 for plate-like and 1 for tubular-like structures, R_B is the parameter to define blob-like structure, 1 for blob-like and 0 for other structures and S measures the second order structureness. α, β and γ are three parameters that control the sensitivity of the tubular detection filter. Small values of α and β (less than 0.3) can not discriminate between plate and tubular structures, whereas large values of them (larger than 0.7) produce more disconnected vessels because vessel junctions look more like plates rather than tubes. γ is set as a certain percentage of the maximum intensity value of the image. Note that, low γ values produce too much noise. We set $\alpha = 0.3, \beta = 0.7$ and $\gamma = 0.5 * max(I_0)$. For each voxel the vesselness measure (3.9) is analyzed at different scales (s) and the scale that maximizes vesselness (s_0) is selected and the corresponding vesselness value is saved in the vesselness map $I_v(\mathbf{r})$. The initial vessel mask $S_V^0(\mathbf{r})$ is computed

by applying threshold of 0.03 on $I_v(\mathbf{r})$.

- Graph-cut based modification: The Min-cut/Max-flow graph-cut algorithm is used to generate the vessel segmentation mask [62]. The initial vessel mask $S_V^0(\mathbf{r})$ is used as the foreground hard constraint. The background hard constraint is set using the eigen values related to the vesselness map image which is the background hard constraint is one where $\lambda_2(\mathbf{r}, s_0) \geq 0$ or $\lambda_3(\mathbf{r}, s_0) \geq 0$. This is based on the fact that the vessels are bright structures, hence the dark tubular structures can not be considered as vessels. Note that, in the non-contrasted CT images, this condition should be reversed. As for the edge weights, the weights of the n-links ($B_{i,j}$) are defined as:

$$B_{\mathbf{i,j}} \propto \exp\left(\frac{-(I_0(\mathbf{r}_i) - I_0(\mathbf{r}_j))^2}{2\sigma^2}\right)\left(\frac{1}{dist(\mathbf{r}_i, \mathbf{r}_j)}\right). \tag{3.13}$$

The weights of the terminal links are assigned by the vesselness map $I_v(\mathbf{r})$ as

$$R_{i,f} = \begin{cases} K & , S_V^0(\mathbf{r}_i) == 1 \\ 0 & , (\lambda_2(\mathbf{r}_i, s_0) \geq 0) \quad \text{or} \quad (\lambda_3(\mathbf{r}_i, s_0) \geq 0) \\ -ln(I_v(\mathbf{r}_i)) & , o.w. \end{cases} \tag{3.14}$$

$$R_{i,b} = \begin{cases} 0 & , S_V^0(\mathbf{r}_i) == 1 \\ K & , (\lambda_2(\mathbf{r}_i, s_0) \geq 0) \quad \text{or} \quad (\lambda_3(\mathbf{r}_i, s_0) \geq 0) \\ -ln(1 - I_v(\mathbf{r}_i)) & , o.w. \end{cases} \tag{3.15}$$

where $R_{i,f}$ and $R_{b,i}$ are the weights of the t-links between nodes at $\mathbf{r}_i$ and the terminal nodes (f: foreground, b: background).

- Post-processing: Finally, the connected component analysis with threshold of $100 \times r_{i_x} \times r_{i_y} \times r_{i_z} mm^3$ followed by an image guided morphological dilation with linear structures is carried out. In order to perform this task, 13 linear structures in 13 directions are defined inside a unit sphere as illustrated in Figure 3.6 in which $\phi \in \{0, 45, 90, 135\}$ and $\theta \in \{0, 45, 90, 135\}$. At every voxel of the vessel mask the linear structure which is aligned with the eigen vector of $\lambda_1(\mathbf{r}, s_0)$ at that voxel is chosen to dilate the vessel mask at that point. The resulted vessel

mask is denoted as S_V.

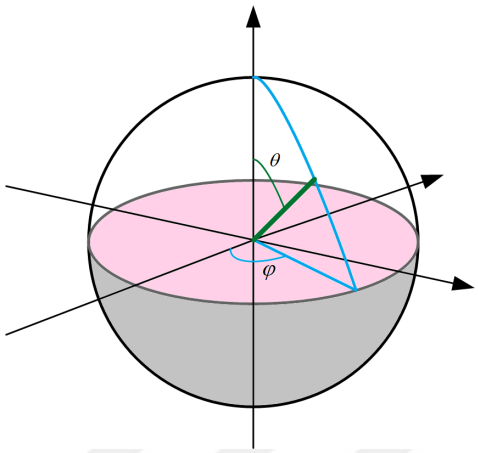


Figure 3.6. An illustration of linear structure generation inside the unit sphere, in which $\phi \in \{0, 45, 90, 135\}$ and $\theta \in \{0, 45, 90, 135\}$.

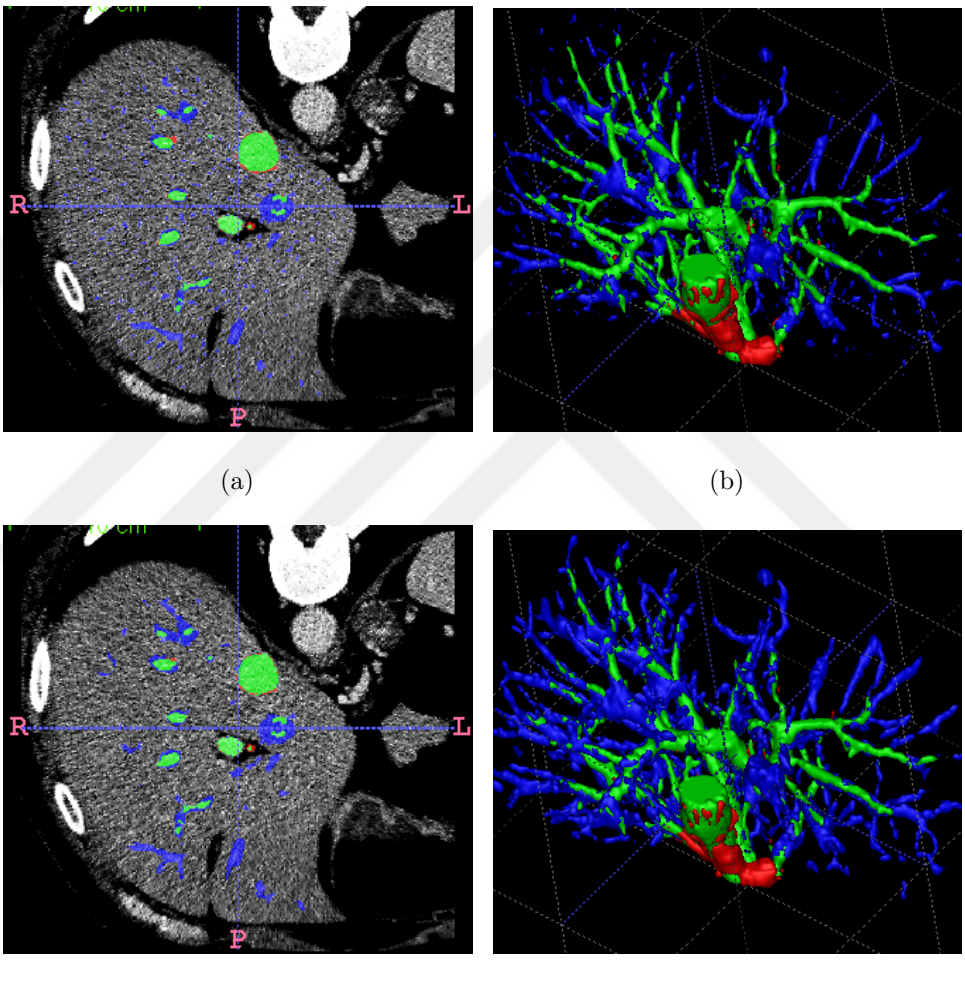
Figures 3.7(a), 3.7(b) and Figures 3.7(c), 3.7(d) show the visual segmentation results tested on 3D contrast enhanced CT liver images of IRCAD database [69] before and after post-processing step, respectively. False negative, true positive and false positive voxels are shown in red, green and blue respectively.

3.2.3. Lesion Segmentation

The lesion segmentation is carried on by applying the MinCut/MaxFlow graph-cut algorithm [62] by using the liver parenchyma ($S_L^0(\mathbf{r})$) and vessel masks ($S_V(\mathbf{r})$) as background hard constraint. Edge weights of the t-links are computed using the background probability image ($I_{P_{bg}}$) which is a unimodal Gaussian distribution with $\mu_{bg} = mean(I_0(S_L^0(\mathbf{r}) == 1))$ and $\sigma_{bg}^2 = var(I_0(S_L^0(\mathbf{r}) == 1))$. The weights are computed as:

$$R_{i,f} = \begin{cases} 0 & , (S_L^0(\mathbf{r}_i) = 1) \quad \text{or} \quad (S_V(\mathbf{r}_i) = 1) \\ -ln(1 - I_{P_{bg}}(\mathbf{r}_i)) & , o.w. \end{cases} \tag{3.16}$$

$$R_{i,b} = \begin{cases} K & , (S_L^0(\mathbf{r}_i) = 1) \quad \text{or} \quad (S_V(\mathbf{r}_i) = 1) \\ -ln(I_{P_{bg}}(\mathbf{r}_i)) & , o.w. \end{cases} \tag{3.17}$$



(a) (b)

(c) (d)

Figure 3.7. Vessel segmentation results before (a,c), and after (b,d) linear morphological post-processing. True-positive mask (green), false-negative (red), and false-positive (blue).

where K is a sufficiently-large constant that prohibits the corresponding edge to be in the cut C. The weights of the n-link $B_{i,j}$, between two node i and j, is defined as

$$B_{i,j} = \exp\left(\frac{-(I_{P_E}(\mathbf{r}_i) - I_{P_E}(\mathbf{r}_j))^2}{2\sigma_1^2})\right) \tag{3.18}$$

where $\sigma_1^2 = 1$, but its value slightly changed (manually) based on the segmentation results during the experiments. I_{P_E} is the lesionness probability image defined as

$$I_{P_E}(\mathbf{r}_i) = exp(\frac{-(I_0(\mathbf{r}_i) - \mu_E)^2}{2\sigma_E^2}) \tag{3.19}$$

where $\mu_E = mean(I_0(S_L^0(\mathbf{r}) == 1))$ and $\sigma_E^2 \ll var(I_0(S_L^0(\mathbf{r}) == 1))$. The idea behind the lessionness probability definition is that lesion can be hyperdense (brighter) or hypodense (darker) compared to the liver density. The only problematic case is when a lesion is peripherally calcified, which results in hyperdensity voxels around a hypodense lesion. Therefore, it is important to prevent the graph-cut from breaking these links between the inner and outer region of such lesions. Using Equation (3.18), if voxel $\mathbf{r}_i$ is related to a lesion and voxel $\mathbf{r}_j$ is related to the liver, $I_{P_E}(\mathbf{r}_i)$ is small, while $I_{P_E}(\mathbf{r}_j)$ is large resulting in a very low $B_{\mathbf{i,j}}$, which is preferable. However, if voxel $\mathbf{r}_i$ is related to a dark-lesion and $\mathbf{r}_j$ is related to a bright-lesion, $I_{P_E}(\mathbf{r}_i)$ is small and $I_{P_E}(\mathbf{r}_j)$ is also small making $B_{\mathbf{i,j}}$ large, which is also preferable. All segmentation masks are manually verified and corrected to avoid errors due to the image preprocessing step.

3.3. Data Generation

A liver lesion CT image is taken to be a medical case representing three components: (i) computer generated image descriptors (low level image features), (ii) doctors' radiological annotations (semantic features), and (iii) non-image metadata. The low level image features are (visual) features that describe liver visual characteristics, which are extracted from 3D liver CT images as described in Section 3.3. The semantic features used in this thesis are provided by an expert radiologist by completing a form designed in-house by using the ONLIRA ontology [5]. They describe images in the

higher level of semantic understanding as described in Section 3.3.2. The non-image metadata is non-radiological information about patient's demographics, medical history, laboratory results and the clinical observations, which are explained in Section 3.3.3.

3.3.1. Image Features

The image features/descriptors capitalize the visual cues contained in an image. They can be considered as two main groups based on their scope of application. The *global* image descriptors which summarize the general visual properties of the liver, vessels and set of lesions, and the *individual* lesion descriptors, which reflect finer levels of visual information related to individual lesions. The global image descriptors include:

- Liver volume, mean and variance
- Vessel volume and ratio of vessels to liver voxels
- Number of lesions, volume of smallest lesion, volume of largest lesion, Ratio of lesions to liver voxels, mean and variance of the whole lesions, and the lesions' histogram-based features including skewness, kurtosis, uniformity, smoothness, abcssia, entropy, threshold, and the low frequency coefficients of histogram's three level Haar wavelet transform.

The individual lesion descriptors are divided in to three types as shape, location, and texture features.

- Shape descriptors are geometric-based cues which are directly from the lesion mask and consist of volume, area, sphericity, convexity, compactness, solidity, maximum-extent, aspect-ratio and fourier descriptors.
- Location descriptors are extracted using all segmentation masks and describe the location of a lesion with respect to the vessels or inside the liver. These features contains lesion proximity to the nearest vessel, touch area ratio between lesion and vessels, vessel to lesion ratio inside the lesion convex hull and the anatomical location of lesion inside the liver.

- Texture descriptors include gray-scale, boundary and texture-based features which are derived from image intensity values.
  (i) *Gray-scale descriptors* are derived directly from raw pixel intensities and include mean and variance of lesion tissue, and the lesion's intensity histogram-based features containing skewness, kurtosis, uniformity, smoothness, abcssia, entropy, threshold, and the Harr wavelet's coefficients derived.
  (ii) *Boundary descriptors* include the lesion's boundary sharpness and lesion's boundary contrast.
  (iii) *Texture-based features* are photometric features that encode spatial organization of pixel values and include six Haralick's features, five Tamura features, the Gabor filter based features, and the invariant Hu moments.

58 low-level image features (a 320-dimensional feature vector) are extracted for every lesion image which are listed in Table 3.1 in which both global and individual features are combined to categorize them under *shape*, *location* and *texture*. Note that this grouping is used in the annotation step. For more details on these features, readers are referred to Appendix A.

3.3.2. Semantic Features

Radiologists describe their observations of CT scans in free-text radiology reports. Unfortunately these reports, written in the daily language, are not easy to process automatically. Hence, it is beneficial to record these observations in a structured manner. An ontology is a conceptualization of a domain in a structured way, which enables description of concepts and their properties, as well as relationships between the concepts. The ONLIRA ontology is employed as the multi-concept ontological domain representation for liver, developed based on the radiologists' manual annotation process [5]. It models the liver CT annotations with an emphasis on its properties and the relations between liver, hepatic veins and liver lesions. The ONLIRA ontology is built upon the RadLex lexicon with the extensions in attributes and additions of ontological relations between these attributes [70]. Readers are referred to Appendix B for full description of ONLIRA. Figure 3.8 shows a visual representation of ONLIRA

Table 3.1. Low-level Image Features.

	subGroup	**Features**
Shape & Location	Liver	Volume
	Vessel	Volume, Ratio in liver, Vessel to lesion ratio, Vessel to total ratio, Proximity to vessel, Touch area ratio
	Lesion Set	Number of lesions, Smallest lesion volume, Largest lesion volume, Ratio in liver
	Single Lesion	Volume, Area, Max extent , Aspect ratio, Sphericity, Compactness, Convexity, Solidity, Anatomical location, FourierDescriptors
Texture	Liver	Mean, Variance
	Lesion Set	Mean, Variance, Skewness, Kurtosis, Energy, Smoothness, Bin of histogram peak, Entropy, Threshold, Haar wavelet coeff.
	Single Lesions	Mean, Variance, Skewness, Kurtosis, Energy, Smoothness, Bin of histogram peak, Entropy, Threshold, Haar wavelet coeff., HuMoments, Gabor features, Haralick entropy, Haralick angular $2nd$ moment, Haralick interia, Haralick inverse difference moment, Haralick cluster shade, Haralick cluster prominence, Tamura coarseness, Tamura contrast, Tamura directionality-x, Tamura directionality-y, Tamura directionality-z, Boundary sharpness, Boundary contrast

with its 40 concepts, 12 relations and 36 properties. Utilizing ONLIRA, 72 semantic

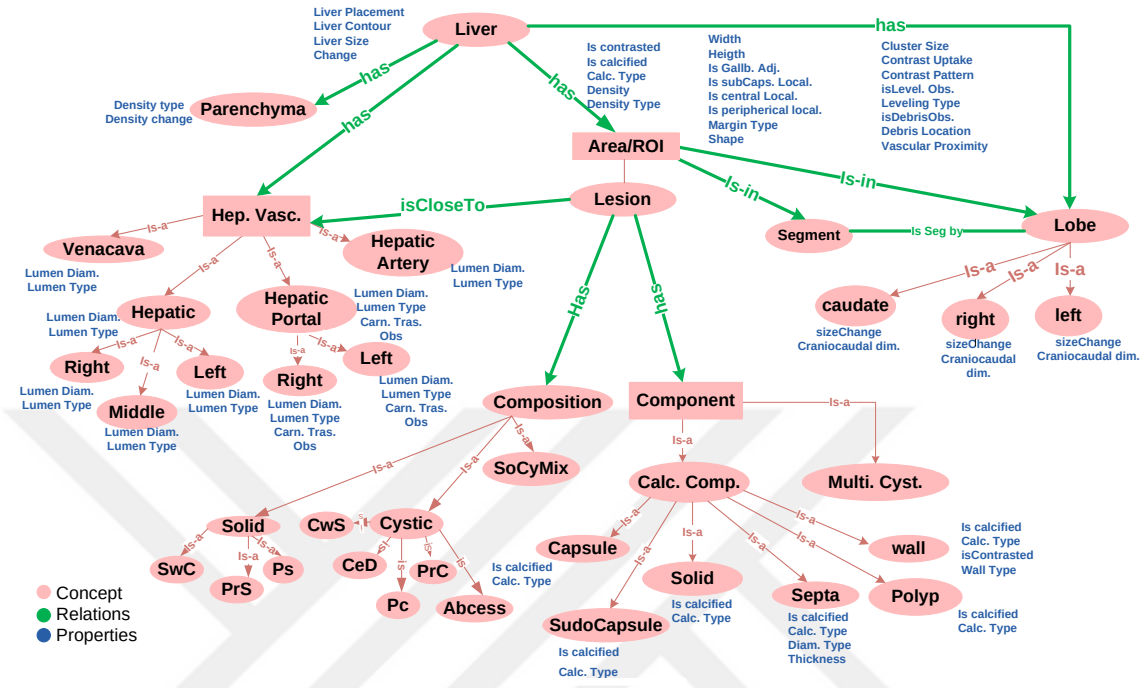


Figure 3.8. The ONLIRA ontology.

features clinically characterizing the liver (Table 3.2), the hepatic vascular (Table 3.3), and the liver lesions (Tables 3.4, 3.5 and 3.6) are defined.

The semantic features are collected via an in-house web application, called CaReRa-WEB [1]. Figure 3.9 shows a screenshot of the system. CaReRa-WEB allows the user to browse a database of cases to brows through CT images, mask ROIs (lesions) and fill in an annotation form about the masked lesion. Every sematic feature corresponds to a question in the form.

3.3.3. MetaData

The non-image metadata represents the non-image information on the liver case. The data is composed of both quantitative and qualitative properties and the description of metadata is enhanced with standardized terminologies, when possible. The defined non-image metadata include

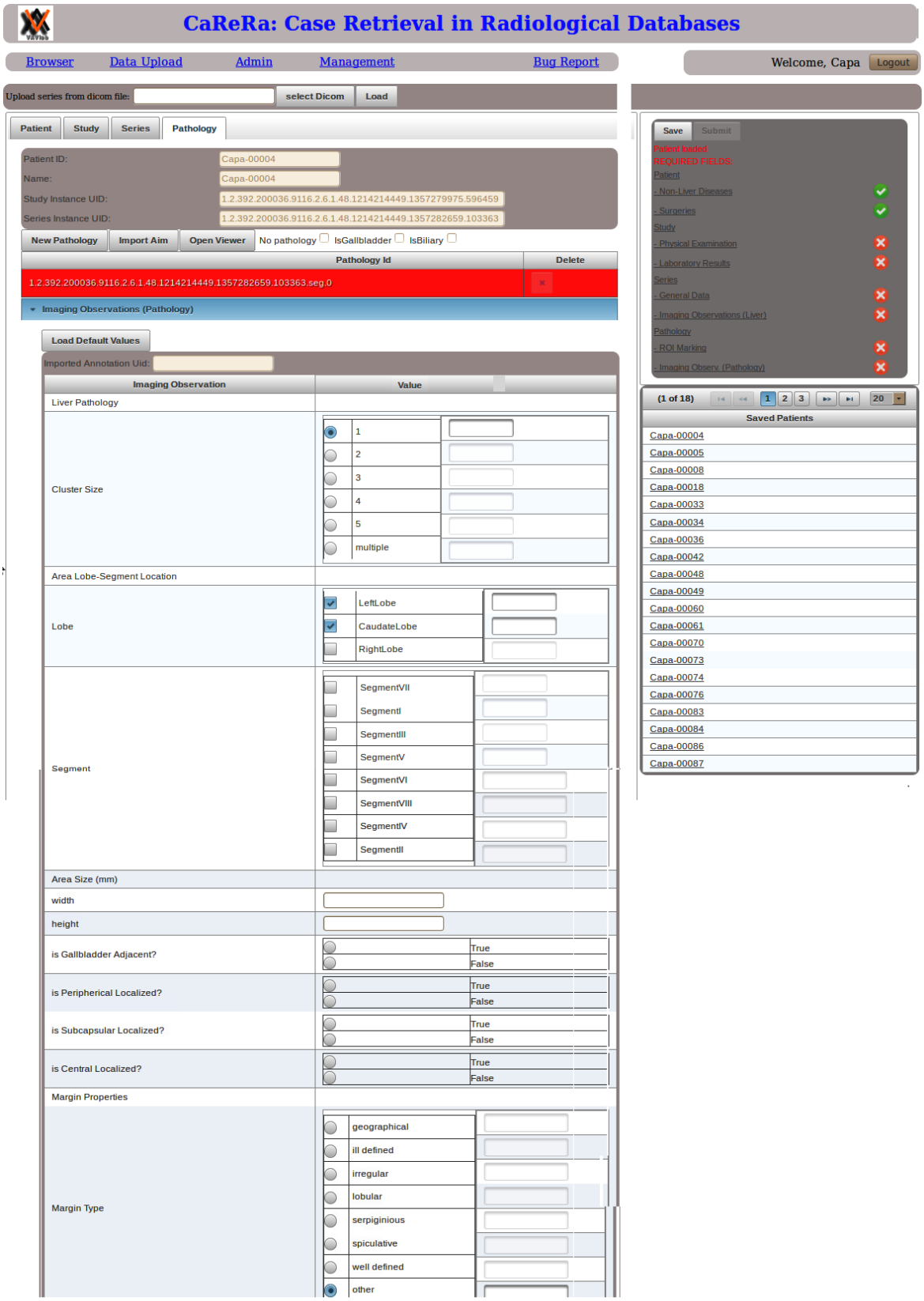


Figure 3.9. The screenshot of the CaReRa-WEB system [1].

Table 3.2. List of semantic features of liver derived from ONLIRA.

Group	**Feature**	**Range**
Liver	Liver placement	Downward_displacement, Normal_placement , Leftward_displacement, Upward_displacement, Other
	Liver contour	Irregular, Lobulated, Nodular, Regular, Other
	Liver size change	Decreased, Increased, Normal, Other
	Cranioc. dim.	The amount change in size of liver craniocaudal dimension (mm)
	Density type	Heterogeneous, Homogeneous, Other
	Density change	Decreased, Increased, Normal, Other
Right Lobe	Cranioc. dim.(mm)	The amount change in size of right lobe(mm)
	Size change	Decreased, Increased, Normal, Other
Left Lobe	Cranioc. dim.(mm)	The amount change in size of left lobe(mm)
	Size change	Decreased, Increased, Normal, Other
Caud. Lobe	Cranioc. dim.(mm)	The amount change in size of caudate lobe(mm)
	Size change	Decreased, Increased, Normal, Other

Table 3.3. List of semantic features of vessels derived from ONLIRA.

Group	**Feature**	**Range**
Hepatic artery	Lumen diam.	Decreased, Increased, Normal, Other
	Lumen type	Obliterated, Open, Partially_obliterated, Other
Hepatic portal vein	Lumen diam.	Decreased, Increased, Normal, Other
	Lumen type	Obliterated, Open, Partially_obliterated, Other
	Cavernous transformed	True,False,NA
Left portal vein	Lumen diam.	Decreased, Increased, Normal, Other
	Lumen type	Obliterated, Open, Partially_obliterated, Other
	Cavernous transformed?	True,False,NA
Right portal vein	Lumen diam.	Decreased, Increased, Normal, Other
	Lumen type	Obliterated, Open, Partially_obliterated, Other
	Cavernous transformed?	True,False,NA
Hepatic vein	Lumen diam.	Decreased, Increased, Normal, Other
	Lumen type	Obliterated, Open, Partially_obliterated, Other
Left hepatic vein	Lumen diam.	Decreased, Increased, Normal, Other
	Lumen type	Obliterated, Open, Partially_obliterated, Other
Middle hepatic vein	Lumen diam.	Decreased, Increased, Normal, Other
	Lumen type	Obliterated, Open, Partially_obliterated, Other
Right hepatic vein	Lumen diam.	Decreased, Increased, Normal, Other
	Lumen type	Obliterated, Open, Partially_obliterated, Other

Table 3.4. List of semantic features of lesions derived from ONLIRA (I).

Feature	**Range**
Cluster_size	1, 2, 3, 4, 5, Multiple
Contrast_uptake	NA, Dense, Heterogeneous, Homogeneous, Minimal, Moderate, Other
Contrast_pattern	Early_uptake_then_wash_out, Fix_contrast_in_late_phase, NA, Central, Heterogeneous, Homogeneous, Peripheric, Peripheric_nodular, Spokes_wheel, Undecided, Other
Composition	SolidCycsticMix, Solid, SolidWithCystic, Abcess, PureSolid, PredominantCystic, PredominantSolid, Cystic, PureCystic, CysticWithSolidComponent, CysticWithDebris
Leveling_observed?	True,False
Leveling_type	NA, Fluid_fluid, Fluid_gas, Fluid_solid, Gas_solid, Other
Debris_observed?	True,False,NA
Debris_location	NA, Floating_inside, Located_on_dependent_ position,Other
is_close_to_vein	NA, HepaticArtery, HepaticPortalVein, RightPortalVein, LeftPortalVein, HepaticVein, RightHepaticVein, MiddleHepaticVein,LeftHepaticVein, VenacavaInferior, AnteriorBranchOfRightPortalVein, PosteriorBranchOfRightPortalVein,Other
Vein_proximity	NA, Adjacent, Adjunct_to_contact, Bended, Circumscribed, Invaded, Other
Lobe	LeftLobe, CaudateLobe, RightLobe
Segment	SegmentI, SegmentII, SegmentIII, SegmentIV, SegmentV, SegmentVI, SegmentVII, SegmentVIII

Table 3.5. List of semantic features of lesions derived from ONLIRA (II).

Feature	**Range**
Width	a number in mm which represents width of the lesion
Height	a number in mm which represents heigth of the lesion
Gallbladder._adj.?	True,False
Peripheral_loc.?	True,False
Subcapsular_loc.?	True,False
Central_loc.	True,False
Margin_type	Geographical, Ill_defined, Irregular, Lobular, Serpiginous, Speculative, Well_defined, Other
Shape	Band, Fusiform, Irregular, Linear, Nodular, Ovoid, Round, Serpiginious, Other
Is contrasted(Area)	True,False,NA
Is calcified? (Area)	True,False,NA
Calc._type (Area)	NA, Coarse, Focal, Millimetric-fine, Punctuate, scattered, other
Density	NA, Hyperdense, Hypodense, Isodense, Other
Density_type	NA, Heterogeneous, Homogeneous, Other
Is_cont.?(wall)	True,False,NA
Wall_type	NA, Thick, Thin, Other

Table 3.6. List of semantic features of lesions derived from ONLIRA (III).

Feature	**Range**
is_Calcified? (Capsule)	True,False,NA
Calc._type (Capsule)	NA, Coarse, Focal, Millimetric-fine, Punctuate, Scattered, Other
Is_calc.? (polyp)	True,False,NA
Calc._type (polyp)	NA, Coarse, Focal, Millimetric-fine, Punctuate, Scattered, Other
Is_calc.? (pCapsule)	True,False,NA
Calc._type (pCapsule)	NA, Coarse, Focal, Millimetric-fine, Punctuate, is Scattered, Other
Is_calc.? (septa)	True,False,NA
Calc._type (septa)	NA, Coarse, Focal, Millimetric-fine, Punctuate, Scattered, Other
Diam._type (septa)	NA, Complete, Incomplete, Other
Thickness (septa)	NA, Thick, Thin, Other
Is_calc.? (component)	True,False,NA
Calc._type (component)	NA, Coarse, Focal, Millimetric-fine, Punctuate, Scattered, Other
Is_calc.? (wall)	True,False,NA
Calc._type (wall)	NA, Coarse, Focal, Millimetric-fine, Punctuate, Scattered, Other

- Demographics: Age and gender
- Diseases and Diagnosis are described using ICD-10 codes [71] and contain
  (i) Non-liver diseases (code / free text)
  (ii) Genetic diseases (code / free text)
  (iii) Liver pre-diagnosis (code / free text)
  (iv) Liver final diagnosis (code / free text)
- Drugs are presented by ATC codes [72] and includes
  (i) Regular drugs (code / free text)
  (ii) Non-regular drugs (code / free text)
- Surgeries (code / free text) are provided via Turkish Ministry of Health codes [73].
- Complaints: Kept as free text for future reference.
- Laboratory results

(i) AFP	(vi) Ammonia	(xi) GGT	(xvi) tBilirubin
(ii) Albumin	(vii) Amylase	(xii) HDL	(xvii) PT
(iii) ALP	(viii) APTT	(xiii) LDH	
(iv) AAT	(ix) AST	(xiv) iBillirubin	
(v) ALT	(x) CBC	(xv) dBilirubin	

- Physical examinations:

(i) Blood pressure	(iv) Oedema
(ii) Jaundice	(v) Ascites
(iii) Pulse	

3.4. Computer-Aided Medical Image Annotation (CMIA)

Computer-aided medical image annotation (CMIA) is a system which assign visual concepts to the medical images using their low level features. These visual concepts are more similar to the radiologists' perception and interpretation. CMIA is used as an initial step of the content based medical image retrieval (CBMIR) system which aims at finding similar lesion images in the dataset of 3D liver CT images.

The proposed domain-aware radiologist-in-the-loop semi-automatic CMIA approach utilizes a Bayesian tree structured model (designed for *lesion-related* concepts) to represent the inter-dependencies of the concepts while taking advantage of noisy observations of individual classifiers trained on low-level image features. The domain network model also guides the interactive annotation/active learning scheme to sequentially request manual input from the user to iteratively perform the full annotation.

3.4.1. Overview of the Proposed CMIA

This section describes the proposed CMIA approach for the annotation of lesion CT images [58]. Let us define the value of intended semantic features (generated from ONLIRA) to be annotated, as *semantic annotation*. Each *semantic annotation* is represented by a random variable called the *annotation label* which is taken from a valid set called *value ranges*. For instance *Wall_Type* is a semantic feature and *Thick, Thin and NA* are the possible *semantic annotations* which are represented by *annotation labels* $1, 2$ and 3 as its *value ranges*.

Three sets of random variables X_S, X_C and X_L are defined, standing for SVM observations of semantic annotations, the semantic annotations themselves and the latent variables, respectively. As shown in Figure 3.10, the image (I) is processed and a set of computer generated image features are computed (F_I) as described in Section 3.3. The noisy observation of annotation labels (X_S) are computed via the pre-trained individual SVM classifiers. In the next step, a Bayesian network model is used to infer the semantic annotations X_C using X_S as observation/evidence. As the final step, values of X_C are modified iteratively by the manual input value of the most important semantic feature in each iteration. X_S, X_C and X_L are represented in a bottom-up order in our tree-structured hierarchical models.

30 liver lesion related semantic features in the current model are used and represented as a set of *concept nodes* indexed from the set $C = \{1, \ldots, K = 30\}$. These are derived from the 41 lesion related semantic features expressed in ONLIRA (Section 3.3.2), in the following manner: (i) Semantic features that are constant across

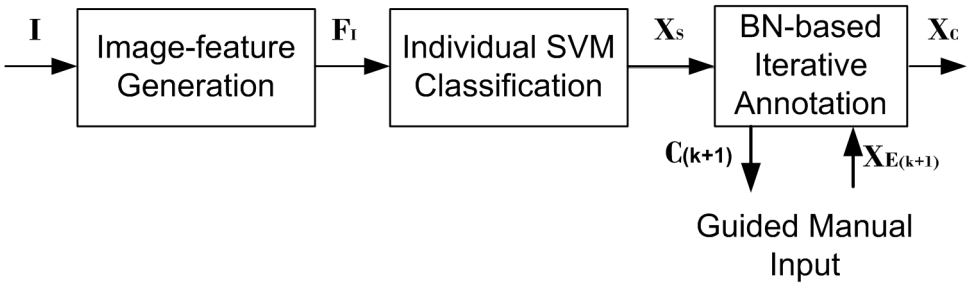


Figure 3.10. CMIA block diagram.

the dataset (*Debris_Location* and *Is_Debris_Observed*) are excluded. (ii) Information duplicating semantic features (*Segment* and *is_Close_To_Vein*) are also excluded, (iii) *Vascular_Proximity*, which can take multiple annotations is divided into separate semantic features with a binary value range, (iv) similar semantic features are merged by adding *False* to their possible semantic annotations (e.g. *Is_Contrasted* is merged with *Contrast_Uptake* and *Contrast_Pattern*, *Is_Calcified* is merged with *Calcification_Type*, and *Is_Leveling_Observed* is merged with *Leveling_Type*). Tables 3.7 and 3.8 list these semantic features categorized as *Location, Shape* and *Texture*.

3.4.2. SVM Observations of Semantic Features

Having a training set of annotation labels along with their corresponding image features, one can train an individual classifier for each annotation label separately. Kernel-SVM is a supervised binary classifier that constructs a hyperplane to separate data into distinct classes by maximizing the margin between them via performing a linear discrimination in a higher-dimensional space [74–76]. Data is transformed to higher-dimensional space via a kernel function (linear, Gaussian radial basis function (RBF), sigmoid, polynomial, and etc.) [77]. Kernel-SVM performs the optimization by simultaneously minimizing the classification error and maximizing the geometric margin between the classes. The points located on the margin of the separating hyperplane are called the support vectors and they specify the decision function.

The original SVM is designed for binary classification problems. However, one approach to convert the binary SVM to a multi-class classifier is *one-versus-one* (1_vs._1) approach also referred to as the pairwise coupling [78], in which binary SVM classi-

Table 3.7. Liver lesion location and shape semantic features along with their possible semantic annotations and value range.

Group	**Feature Name**	**Abbr.**	**Semantic annotations (value range)**
Location	Lobe	Lobe	Right(1), Caudate(2), Left(3), Right-Left(4)
	is_Gallbladder_Adjacent?	GalAdj	False(1), True(2)
	is_Peripherical_Localized?	PerLoc	False(1), True(2)
	is_Subcapsular_Localized?	SubLoc	False(1), True(2)
	is_Central_Localized?	CentLoc	False(1), True(2)
	Vascular_Proximity_Adjacent	Adj	False(1), True(2)
	Vasc_Prox._AdjToContact	Adt	False(1), True(2)
	Vasc_Prox._Bended	Ben	False(1), True(2)
	Vasc_Prox._Invaded	Inv	False(1), True(2)
	Vasc_Prox._NA	NA	False(1), True(2)
Shape	Cluster_Size	Cluster	1,2,3,4,5,multiple(6)
	Margin_Type	MarT	Illdefined(1),Irregular(2), Lobular(3),Other(4),Welldefined(5)
	Area_Shape	ASh	Irregular(1),Nodular(2),Ovoid(3) , Round(4),Serpiginious(5)
	Wall_Type	WaT	Thick(1),Thin(2),N/A(3)
	Septa_Diameter_Type	SepDia	Complete(1),Incomplete(2), N/A(3)
	Septa_Thickness	SepW	Thick(1),Thin(2),N/A(3)

Table 3.8. liver lesion texture semantic features along with their possible semantic annotations and value range.

Group	**Feature Name**	**Abbr.**	**Semantic annotations (value range)**
Texture	Contrast_Uptake	ContUp	Hetero.(1),Homoge.(2) ,Minimal(3), Moderate(4), False(5),N/A(6)
	Contrast_Pattern	ContPat	Hetero.(1),Homoge.(2), Periphe.(3) ,Perip_nodular(4),Undecided(5) ,False(6), N/A(7)
	Lesion_Composition	Comp	SolidCys.Mix(1),SolidCys.(2), PureSolid(3), PureCys.(4) PredominantSolid(5),Abcess(6)
	Leveling_Type	Level	Fluid gas(1), N/A(2)
	Density	Dens	Hyper.(1), Hypo.(2), Iso.(3), Other(4)
	Density_Type	ADens	Hetero.(1), Homoge.(2), Other(3)
	is_Wall_Contrasted?	WaCont	False(1), True(2), N/A(3)
	Area_Calc._Type	CaA	Coarse(1), Focal(2), False(3)
	Capsule_Calc._Type	CaCa	False(1), N/A(2)
	Polyp_Calc_Type	CaPo	False(1), N/A(2)
	Pseudocapsule_Calc._Type	CaPs	False(1), N/A(2)
	Septa_Calc_Type	CaSep	Coarse(1), Punctate(2) , False(3), N/A(4)
	Solid_Comp_Calc._Type	CaSo	False(1), N/A(2)
	Wall_Calc_Type	CaWa	Coarse(1), Punctate(2), False(3), N/A(4)

fiers are trained for all possible pairs, and in the test phase the final decision is made through the majority voting.

As discussed in Section 3.3, the size of value range for every semantic feature is different, i.e. some have binary range, while the others may take multiple values. Hence, based on the size of their range, binary or multi-label SVM classifiers with RBF kernels were individually trained for each label with its value range used as the class labels. Matlab^{TM}'s built-in machine learning toolbox was used to train the SVM classifiers. Since classification using all features is computationally costly and may cause over-fitting problem, each classifier was trained on a manually selected subset of 320-D features. Manual feature selection was based on the feature categories listed in Tables 3.7, 3.8 and Table 3.1. For every semantic feature from location or shape group, the location and shape related low level features were used. Whereas, for texture related semantic features, the low level features categorized as texture descriptors were utilized. During the experiments for every training set the low level features that were constant across the dataset were excluded.

The trained SVM classifiers are used to set the values for X_S where $S = \{37, \ldots, 66\}$. They serve as evidence in the subsequent Bayesian inferencing.

3.4.3. Domain-Aware BN Model

Bayesian networks are directed acyclic graphs (DAGs) that encode the cause-effect and conditional independence relationships among variables in a probabilistic reasoning system [79]. Nodes and edges represent the variables and their relations, respectively. Directed links exist between dependent nodes, and every node may be a discrete variable with different number of possible values (value range) or a continuous value with a defined distribution. In discrete domain (which is the case in this dissertation), every parent-child relation is defined by a conditional probability table (CPT) representing the probability of each value that a node can take given the value of its parents. The set of CPTs of a BN constitute the network parameters.

The joint probability distribution of all random variables in a BN representation is given as

$$P(x_1, ..., x_N) = \prod_{i=1}^{N} P(x_i|pa(x_i)) \tag{3.20}$$

where $pa(x_i)$ stands for the parent set of the node (the random variable) x_i. In a fully observed dataset, computing the network parameters is a matter of counting co-occurrences, whereas in the case of the existence of hidden nodes or missing values, the parameters should be estimated from the data using an expectation maximization (EM) or a similar scheme [80].

The EM scheme is an iterative approach, which is formulated as the maximization of the marginal likelihood of $P(D|\Theta)$, with respect to Θ, where D stands for 'observable' variables and Θ represents the network parameters. Representing the hidden/missing variables with h, the EM algorithm iteratively solves the maximization problem by initially fixing the Θ and optimizing $P(h|D, \Theta)$ in the E-step, followed by fixing $P(h|D, \Theta)$ and optimizing $P(D, h|\Theta)$ with respect to Θ in the M-step. The algorithm iterates until convergence. This method, which its algorithm is described in Figure 3.11, follows the idea of Bayesian estimation for the complete data in order to compute Θ which is the matter of counting. In other words, in each EM iteration, the network parameters (CPTs) are updated via probabilistic counting (i. e. if there is a missing variable, its probability resulted from the previous maximization step is used) [81].

The *domain-aware* BN model, denoted by $\mathcal{G}_D$, is manually constructed based on the ONLIRA ontology [5], by introducing *hidden/latent nodes* corresponding to the categorical groupings in Tables 3.7 and 3.8. $\mathcal{G}_D$ has a hierarchical structure with a set of six hidden nodes, indexed from the set $L = \{31, \ldots, 36\}$, representing the lesion at the top of the hierarchy, its shape, texture (with calcification as its sub-category), and location (with proximity as its sub-category) positioned further down as shown in Figure 3.12. The number of states of each hidden node is empirically set to three.

$k = 0$

Choose an initial setting for θ^0

while θ not converged **do**

  k := k + 1;

  for $n = 1$ to N **do**

    $q_k^n(h^n|D^n) = P(h^n|D^n, \theta^{k-1})$

  end for

  $\theta^k = argmax_\theta \sum_{n=1}^{N} < \log P(D^n, h^n|\theta) >_{q_k^n(h^n|D^n)}$

end while

return θ^k

Figure 3.11. Expectation maximization algorithm.

Larger number of states for hidden nodes not only increase the computational load, but also weaken the generalisability of the model, whereas small number of states hinder the modeling capacity of hidden nodes (latent variables). The direct links may connect two hidden nodes, a hidden node and a concept node or a concept node and its observation node.

Thus, model parameters Θ are conditional probabilities $P(X_i = \xi|X_{pa(i)} = \rho)$ for each child-parent pair $(X_i, X_{pa(i)})$ in the tree, and for each of their possible state configurations (ξ, ρ). Since there is no missing data in (X_c, X_S) pairs, the related parameters are simply computed by counting the co-occurrences over the training dataset. Whereas, the EM algorithm is used for maximum likelihood parameter estimation of the latent/hidden variables X_L.

As illustrated in Figure 3.12, there is one observation node (X_S) per concept node (X_C) and the SVM classifier outputs are used to set their values. X_S and X_C are both observable in the available data, but as will be explained later, for a test example, X_C are incrementally observed during the iterative annotation process. In all of the proposed and compared models, the joint distribution $(P(X_S, X_C, X_L) = P(X_L)P(X_C|X_L)P(X_S|X_C))$ is structured as a directed tree, where X_S are children of X_C in the one-to-one correspondence, and X_L are introduced as parents to various

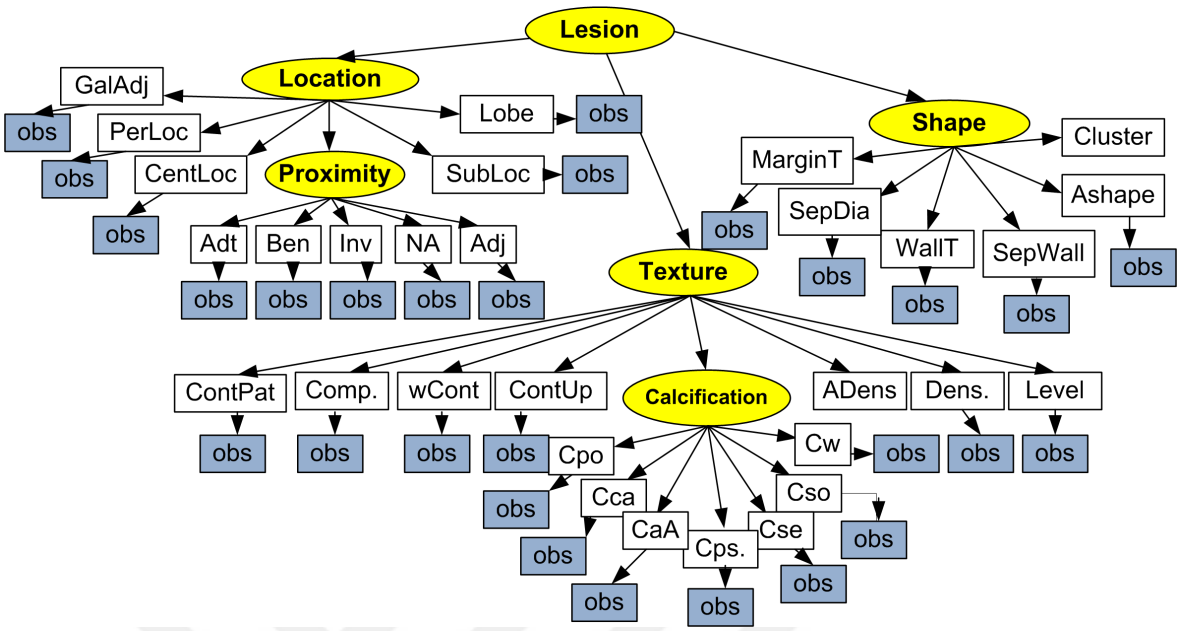


Figure 3.12. Domain-aware BN model($\mathcal{G}_D$): hidden (X_L), concept (X_C) and observation (X_S) nodes are shown in yellow, white, and blue, respectively.

subsets of X_C to encode their dependencies. In the following, X_S are not denoted explicitly for the sake of clarity.

3.4.4. Model Learning

In addition to the previously defined fixed structure, a modified version of the structural expectation maximization algorithm (SEM) is used to learn the network model from the dataset [82]. While the SEM algorithm uses a fixed number of hidden nodes, the modified structural expectation maximization (mS-EM) algorithm (See Figure 3.13) allows to change the number of hidden nodes, while searching for a structure, similar to Elidan *et al.* [83]. The mS-EM is initialized with a network model and it iteratively updates the model until there is no more improvement.

The network model learning is a data-centric optimization stage applied over a given initial network model to maximize $P(D|\mathcal{G})$, where D is the observed data and $\mathcal{G}$ is the network model. The general idea behind model learning is to search the space of all possible networks by maximizing the model's capacity of explaining the observed (training) data while simultaneously avoiding over-fitting. As the search space which

is generally very large, these algorithms employ heuristics to speed-up what otherwise would be a greedy search with respect to a generic cost function. Bayesian information criterion (BIC) score is used as the objective function for model learning because it is not only sensitive to the model's power to explain observed data but also penalizes the model complexity [84]. The BIC score is defined as

$$\mathrm{BIC}(\mathcal{G}|D) = \log_2 P(D|\Theta_{\mathcal{G}}, \mathcal{G}) - \frac{size(\mathcal{G})}{2} \log_2(N) \tag{3.21}$$

where N is the number of training instances, $\Theta_{\mathcal{G}}$ is an estimate of the maximum likelihood parameters for the structure $\mathcal{G}$ and $size(\mathcal{G})$ is the total number of independent parameter in the network which is defined as

$$size(\mathcal{G}) = \sum_i \left((dim(X_i) - 1) \prod_{j=1}^{M} dim(X_{pa_j(i)}) \right) \tag{3.22}$$

where $dim(.)$ stands for the cardinality of the corresponding variable's range set. In Equation (3.21), the left term takes care of the fit to the training data, while the right term penalizes the number of independent parameters space(complexity).

Given a model $\mathcal{G}_{cur}$ at the current iteration of mS-EM with the set of hidden nodes L, let U_l be the set of all triplets in $\mathcal{G}_{cur}$ that have a common hidden parent (Triplets in U may have different parents). The goal is to find the parent with high entropy (i.e. high uncertainty), and low mutual information (MI) with one of its triplets (i.e. this specific parent possesses low information on these children), to insert a new hidden node between the parent and its triplets to describe those children better. This can be formulated as selecting the hidden node with $\frac{H(l)}{max^*(H(l))} > \delta$ where $l \in L$ and δ is empirically set to 0.2 as lower values result in a more complex wider node search space, and higher values output a very narrow parent sets. $\max^*(H(X))$ is the maximum achievable entropy for X, which is limited by the cardinality of the X's range set. The triplets selection criterion can be defined as $\frac{H(u_{lk}|X_l)}{max^*(H(u_{lk}))} > \Gamma$, where $u_{lk} \in U$ is the k^{th} triplet of hidden node l and Γ is the threshold value which $\Gamma \in [0, 1]$.

$\mathcal{G}_{cur} \leftarrow \mathcal{G}_{input}$: Initialization
repeat
  $\Theta_{cur} \leftarrow \text{EM}(\mathcal{G}_{cur}, D)$: Parameter learning
  $\mathbf{U} \leftarrow sort(\mathcal{G}_{cur}, \Theta_{cur})$: Sort groups of children
  for all $u_{lk} \in U$ **do**
    $\mathcal{G} \leftarrow Insert(u_{lk}, \mathcal{G}_{cur})$: Insert hidden node
    $\mathcal{G} \leftarrow Rewire(\mathcal{G})$
    if $BIC(\mathcal{G}) > BIC(\mathcal{G}_{cur})$ **then**
      return $\mathcal{G}_{cur} = \mathcal{G}$
    end if
  end for
until length(U)=0
return $\mathcal{G}_{cur}$

Figure 3.13. The mS-EM model learning algorithm.

To find the triplet with low MI with their parent, let us consider a threshold value ϵ as for their MI

$$\begin{aligned} &MI(u_{lk}; X_l) < \epsilon \\ &H(u_{lk}) - H(u_{lk}|X_l) < \epsilon \\ &H(u_{lk}) - \epsilon < H(u_{lk}|X_l). \end{aligned} \tag{3.23}$$

Therefore, by replacing $H(u_{lk}) - \epsilon$ with $max^*(H(u_{lk})) \times \Gamma$, the triplet selection criteria can be written as

$$\Gamma < \frac{H(u_{lk}|X_l)}{max^*(H(u_{lk}))}. \tag{3.24}$$

Obviously, smaller Γ leads to more triplet candidates causing over-fitting, whereas larger Γ leads to less candidates rendering the search inefficient. In this work, $\Gamma = 0.5$ has been empirically selected.

After limiting the search space by applying the mentioned thresholds, elements of U are sorted in the descending order in a nested fashion, first with respect to the triplets' parent's entropies $H(X_l)$, and then with respect to $\frac{H(u_{lk}|X_l)}{max^*(H(u_{lk}))}$. The top triplet in U is chosen to add a new hidden node as its parent. Step 6 in Algorithm 3.13 corresponds to this hidden node insertion, forming a candidate network model $\mathcal{G}$. Figure 3.14(a) illustrates an example network, where in Figure 3.14(b) the top triplet in U are colored in red, and in Figure 3.14(c) a new hidden parent is introduced for them. The triplet selection aims to find the smallest possible seed that can attract other nodes as its siblings in the rewiring phase.

The rewiring phase at Step 7 in Algorithm 3.13 checks all *parent switching* moves on the candidate network model, where a *parent switching* move is defined as changing the parent of a node from one hidden node to another. The algorithm implements the move that increases the BIC score of the candidate network the most. The hidden nodes with one or no children are deleted. Figure 3.14(d) shows an example of this step, where the attracted node under the new hidden parent is colored in blue. The candidate network $\mathcal{G}$ is accepted as the new $\mathcal{G}_{cur}$ if its BIC score is higher than the previous model's, otherwise the next triplet in U is considered. The algorithm terminates, when none of the triplets in U qualifies for hidden node insertion.

3.4.5. Iterative Annotation Scheme

As one of the contributions of this work, a novel radiologist-in-the-loop scheme for iteratively annotating the semantic features (estimating the state of the concept node) in a given lesion CT image is proposed. More specifically, this system lets the user sequentially provide manual annotations in an optimal order, while simultaneously updating the full annotation for other semantic features by estimating the states of not-yet-annotated concepts conditional to the accumulated evidence. At each step, the decision of which concept to annotate is made based on the rate of increase in the certainty of pending estimations, such that the total amount of user effort (manual annotation) remains minimal until an overall certainty threshold is achieved.

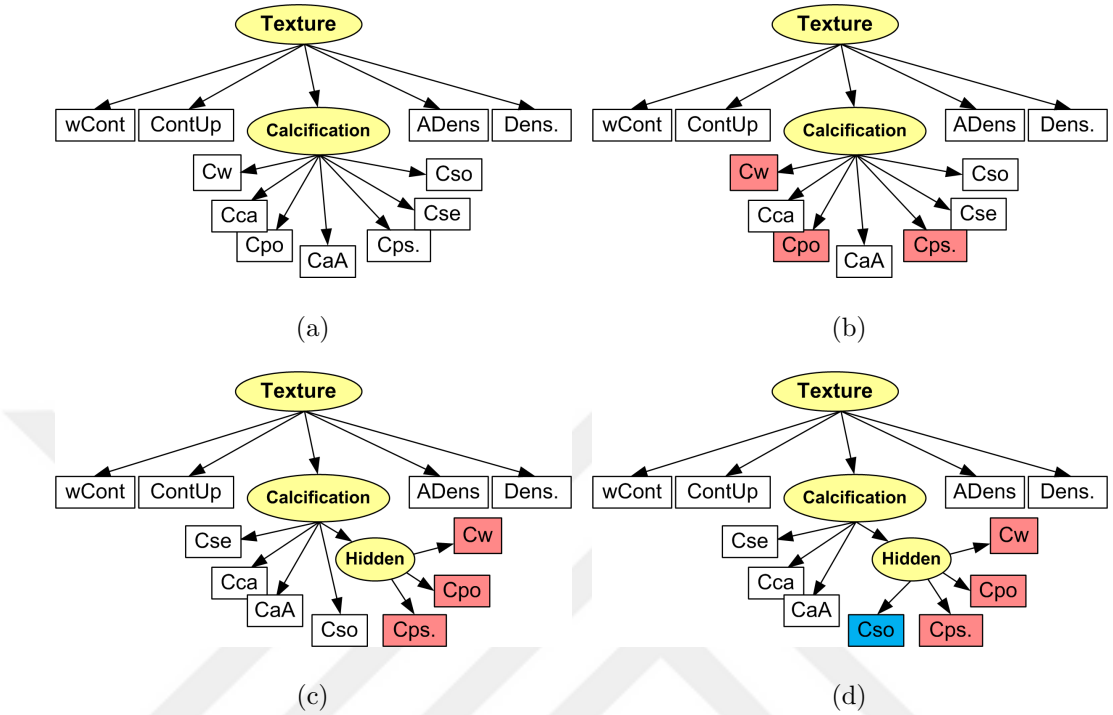


Figure 3.14. An example of model learning algorithm. (a) Initial graph, (b) a triplet (red nodes) is chosen for hidden node insertion, (c) a new hidden parent is added, and (d) the rewiring result.

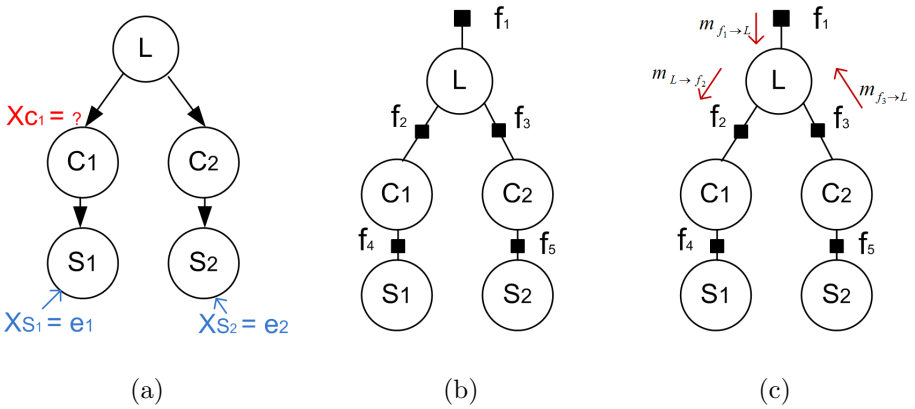


Figure 3.15. A toy example to demonstrate the inference in BNs, (a) the original BN, (b) the factor graph representation, and (c) the message passing on the factor graph.

Marginal inference is the computation of the distribution of the subset of variables given other/no variable. Equation (3.25) represents the example inference in Figure 3.15(a) for node C_1 given nodes S_1 and S_2.

$$\begin{aligned} P(C_1|S_1 = e_1, S_2 = e2) &= \sum_{L,C2} P(C_1, C_2, L|S_1 = e_1, S_2 = e_2) \\ &= \sum_{L,C2} P(L)P(C_1|L)P(C_2|L)P(S_1 = e_1|C_1)P(S_2 = e_2|C_2) \end{aligned} \tag{3.25}$$

To perform inference in BNs, different Bayesian approximate/exact inference methods have been proposed [85–88]. Message passing is the key element in an effective inference method, in which the graph information is summarized by local edge information. Exact inference methods, variable elimination (VE) [87] and belief propagation (BP) [85]), are directly applicable to a tree-structured network such as the ones used in this research.

VE is a method in which the summation in Equation (3.25) are distributed over the network step by step, resulting in elimination of a node as in Equation (3.26) below

$$\begin{aligned} P(C_1|S_1 = e_1, S_2 = e_2) &= \sum_{L,C2} P(L)P(C_1|L)P(C_2|L)P(S_1 = e_1|C_1)P(S_2 = e_2|C_2) \\ &= P(S_1 = e_1|C_1) \sum_{L} P(L)P(C_1|L) \sum_{C2} P(S_2 = e_2|C_2)P(C_2|L) \\ &= P(S_1 = e_1|C_1) \sum_{L} P(L)P(C_1|L)\tau(L, S_2 = e_2) \\ &= P(S_1 = e_1|C_1)\tau(C_1, S_2 = e_2). \end{aligned} \tag{3.26}$$

Note that, VE on trees is equivalent to message passing along tree branches [89]. The elimination of variables can be interpreted as the information transfer to a neighboring node on the graph. VE can always start from any leaf node and continue

the elimination through the root for tree structured networks, which guarantees any marginalization computation.

In contrast, BP, also known as sum-product, is a message passing algorithm in which messages are iteratively sent in both direction between neighbor variables speeding up the inferencing. Figure 3.15(b) illustrates the factor graph generated from the BN in Figure 3.15(a), while Figure 3.15(c) shows messages sent under the BP scheme. If the factor graph is a tree, BP exactly and efficiently computes the marginal of all nodes.

The the factor graph represented in the toy example (Figure 3.15(b)) are defined as

$$\begin{gathered} f_1(L) = P(L) \\ f_2(L, C_1) = P(C_1|L) \ , \ f_3(L, C_2) = P(C_2|L) \\ f_4(S_1, C_1) = P(S_1|C_1) \ , \ f_5(S_2, C_1) = P(S_2|C_2). \end{gathered} \tag{3.27}$$

A message is an unnormalized belief vector of a variable node of length of the cardinality of the variable node's range. Two types of messages (m) are defined One message from a factor node f to a variable node x is defined as the marginalization of the product of the factor and the incoming messages over other variables inside the factor, except the current node

$$m_{f \to x}(x) = \sum (x_1, ..., x_M) f(x, x_1, ..., x_M) \prod_{i=1}^{M} m_{x_i \to f}(x_i). \tag{3.28}$$

In the example graph (Figure 3.15(b)), the message from factor node (f_3) to the variable node (L) ($m_{f_3 \to L}(L)$) is

$$m_{f_3 \to L}(L) = \sum_{C_2} f_3(L, C_2) \prod m_{C_2 \to f_3}(C_2). \tag{3.29}$$

Another message from a variable node x to a factor node f is defined as the product of messages coming to that factor

$$m_{x_i \to f}(x_i) = \prod_{f_j \in f_N} m_{f_j \to x_i}(x_i) \tag{3.30}$$

where f_N is the set of factor connected to node x_j except the factor f. In the example graph (Figure 3.15(b)), message from the variable node L to the factor node $f_1(L)$ ($m_{L \to f_1}(L)$) is

$$m_{L \to f_2}(L) = m_{f_3 \to L}(L) \times m_{f_1 \to L}(L). \tag{3.31}$$

After sufficient number of iterations, the messages converge to a consensus state that gives the marginal probabilities of each variable node as the product of incoming messages

$$P(L) = m_{f_1 \to L}(L) \times m_{f_2 \to L}(L) \times m_{f_3 \to L}(L). \tag{3.32}$$

The process starts by picking an arbitrary node as the root node and further define the leaf nodes which are the starting point of the message passing. Every node that receives all incoming messages can start to send out a message during the iteration. The initialization of the leaf nodes depends on their types, if a leaf node is a variable node, its message is set to an all-one vector, whereas if it is a factor node, its message is the factor value itself. Note that if a variable node is observed, its message is an all-zero vector except the observed state which is set to 1. For instance in the given toy example, if the leaf node is considered as f_1, $m_{f_1 \to L}(L) = f_1$, while if S_1 becomes the leaf node, its initial message is $m_{S_1 \to f_4}(S_1) = 1$.

The proposed iterative annotation method is illustrated in Figure 3.16 and the corresponding algorithm is depicted in Figure 3.17. This strategy can be mathematically formulated as a sequential entropy minimization and falls within the domain of ac-

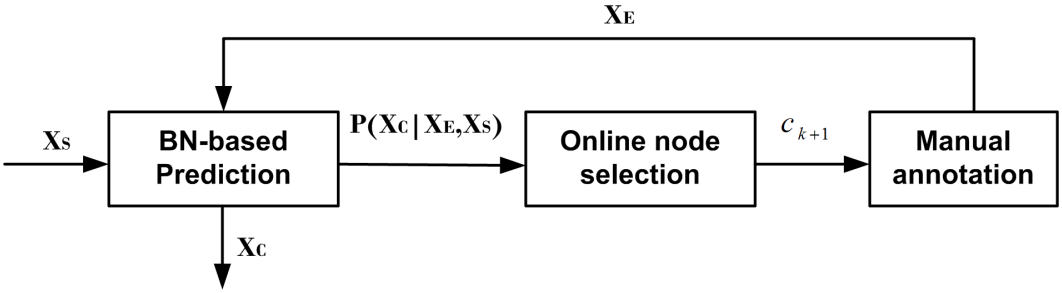


Figure 3.16. Iterative annotation scheme.

tive learning. In particular, let X_C be a K-dimensional random vector of concept nodes, and let X_{E_k} denote the k-dimensional random vector of evidence concepts realized as x_{E_k} up to the kth iteration of the annotation process, with $E_k = \{c_1, c_2, ..., c_k\} \subset C$ representing the set of concept states that have been manually annotated. The decision for the next manual annotation is made by

$$c_{k+1} = \arg \min_{c \in C \setminus E_k} H(X_{C \setminus (E_k \cup \{c\})} | X_c, x_{E_k}) \tag{3.33}$$

which returns the concept that when added to the current evidence set, decrease the overall entropy the most over the remaining variables. By chain rule, one can write this pending entropy as

$$H(X_{C \setminus (E_k \cup \{c\})} | X_c, x_{E_k}) = H(X_{C \setminus E_k} | x_{E_k}) - H(X_c | x_{E_k}). \tag{3.34}$$

Since the first term on the right hand side does not involve c, Equation (3.33) can be rewritten as

$$c_{k+1} = \arg \max_{c \in C \setminus E_k} H(X_c | x_{E_k}) \tag{3.35}$$

namely selecting the concept, whose state has the largest marginal entropy given the evidence so far. The marginal entropy is given by

$$H(X_c | x_{E_k}) = \sum_{x_c} P(X_c = x_c | x_{E_k}) \log P(X_c = x_c | x_{E_k}) \tag{3.36}$$

where the marginal distributions $P(X_c|x_{E_k})$ can be inferred using BP on the Bayesian tree over X_C. Similarly, using those inferred marginal, the states for the remaining concepts can be estimated by

$$\hat{x}_c = \arg\max_{x_c} P(X_c = x_c|x_{E_k}). \tag{3.37}$$

As more evidence accumulates, this process is expected to yield progressively better estimates.

Initialize the set $E_0 = \emptyset$ of annotated concepts
for $k = 0 : K - 1$ **do**
  for all $c \in C \setminus E_k$ **do**
    Compute state estimate $\hat{x}_c$ using (3.37)
    Compute entropy $H(X_c|x_{E_k})$
  end for
  Compute accuracy $Acc(k)$ using (4.1)
  Find $c_{k+1} = \arg\max_{c \in C \setminus E_k} H(X_c|x_{E_k})$
  Get annotation $x_{c_{k+1}}$ from user
  Update $E_{k+1} = E_k \cup \{c_{k+1}\}$
end for

Figure 3.17. Iterative annotation via entropy minimization

3.5. Semantic Based Search and Retrieval

A medical image search and retrieval system must be capable of (a) representing the image content, (b) providing methods to build subjective queries, and (c) assessing the similarity between the query and entries in the database. The scope of this dissertation is limited to the liver lesion (pathologies) CT images. As depicted in Figure 3.18, the proposed iterative search and retrieval scheme utilizes the CMIA, which is previously described, for optimal interactive and iterative refinement of search and retrieval. Specifically, the query is iteratively and interactively updated in the way that

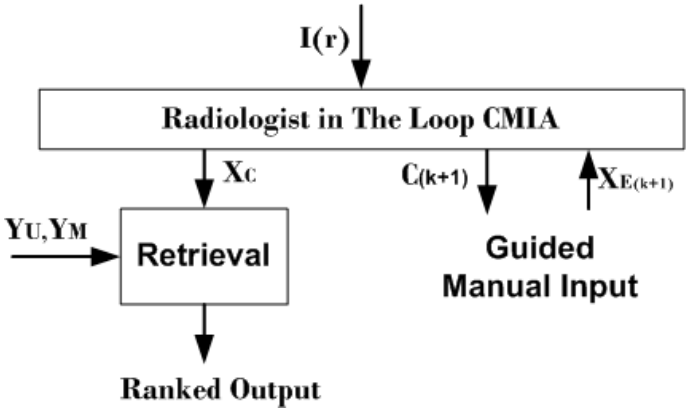


Figure 3.18. Block diagram of the semantic based search and retrieval method.

would improve the semantic representation of the query the most.

As described earlier, automatic annotation based on the low-level visual image features is not reliable and is error prone. Therefore, the CMIA is employed and assessed as an efficient way of improving the semantic query representation and consequently the search and retrieval performance. Liver lesion CT images are represented with a set of multi-label discrete features that are composed of the semantic image descriptors (concepts in CMIA), non-lesion image related semantic features and non-image metadata on the patient such as demographics, laboratory results, physical examinations, past and current disease and drugs.

Lesion-related semantic features were generated as described in Section 3.4.1. The non-lesion image related semantic features, which are extracted from ONLIRA as described in Section 3.3.2 and listed in Table 3.2 and Table 3.3, were processed to decrease the length of their feature vector. Originally there were 31 non-lesion semantic features, from which *Lumen_Diameter_[vessel-name]* and *Lumen_Type_[vessel-name]* were merged representing a new feature named *Lumen_[vessel-name]*. Moreover, features that were constant across the dataset were excluded from the final feature set which were *has_Liver_Placement*, *Lumen_HepaticPortalVein*, *Lumen_HepaticVein*, and *is_CavernousTransformation_Observed*. Final non-lesion image related semantic features includes 13 features named Y_U, $\{U \in 1, ..., 13\}$ and are listed in Table 3.9.

Table 3.9. List of processed liver and vessel semantic features used in the experiments along with their value range.

Feature Name	**Range**
'hasLiverContour'	Nodular(1), Regular(2)
'hasLiverDensityChange'	Decreased(1), Normal(2)
'hasDensity'	Heterogeneous(1),Homogeneous(2)
'Liver_hasSizeChange'	Decreased(1),Increased(2),Normal(3)
'RightLobe_hasSizeChange'	Decreased(1),Increased(2),Normal(3)
'LeftLobe_hasSizeChange'	Decreased(1),Increased(2),Normal(3)
'CaudateLobe_hasSizeChange'	Decreased(1),Increased(2),Normal(3)
'Lumen_HepaticArtery'	Decreased-Obliterated(1), Normal-Open(2)
'Lumen_LeftPortalVein'	Decreased-Obliterated(1), Decreased-Other(2) ,Normal-Open(3)
'Lumen_RightPortalVein'	Decreased-Obliterated(1), Decreased-Open(2), Decreased-Other(3), Normal-Open(4)
'Lumen_LeftHepaticVein'	Normal-Open(1), Other-Obliterated(2)
'Lumen_MiddleHepaticVein'	Decreased-Obliterated(1), Normal-Open(2)
'Lumen_RightHepaticVein'	Decreased-Obliterated(1), Normal-Open(2), Other-Other(3)

Non-image metadata collected in Section 3.3.3 were also processed and discretized by (i) removing the features that were constant/unavailable across the dataset, and (ii) discretizing the lab results and physical examinations using their normal value range listed in Table 3.10 to three possible values (low (1), normal (2) and high (3)). The non-image metada is represented by Y_M, where $\{M \in 14, ..., 35\}$, and are listed in Table 3.11.

Table 3.10. Normal range of laboratory results and physical examinations.

Test	**Unit**	**Normal range**	**Test**	**Unit**	**Normal range**
Albumin	g/dL	3.5-4.8	t-Bilirubin	mg/dL	0.2-1.3
ALP	U/L	male:53-128 female:42-98	LDH	U/L	140-280
ALT	U/L	7-56	GGT	U/L	8-78
AST	U/L	male:8-40 female:6-34	Pt	sec	11-15
d-Bilirubin	mg/dL	0-0.3	APTT	sec	18-28
i-Bilirubin	mg/dL	0.2-0.7	Afp	ng/mL	0-10
Odema	-	T,F,NA	Jaundice	-	T,F
Ascites	-	T,F,NA	Blood	Presure	120/80 - 140/90

Let us define the similarity between a query lesion i and each lesion in the dataset j as

$$d_{all}(i,j) = \alpha.d_{Lesion}(i,j) + (1-\alpha).d_{non-Lesion}(i,j) \tag{3.38}$$

where d_{Lesion} and $d_{non-Lesion}$ are the similarity distance based on the lesion and non-lesion (image and non-image) feature, respectively. The α is the contribution factor of each of these two similarity distances in the final similarity distance. The d_{Lesion} is the similarity distance between lesions, based on the lesion related semantic annotations

Table 3.11. The discretized metadata generated via current existing disease in the database, or the normal range annotation labels.

Feature Name	**Range Set**
DiseasesPastOther	NA(1), Drug abuse(2), Mix germ(3), Colon(4), Rectum(5), Pancreas(6) ,Coroner(7), Lung(8), Hernia(9), Atrial(10), Vertigo(11), Neuroendocrine(12), Pulmonary(13)Fusiform(14), Prostate(15), Teratoma(16), Sigmoid(17)
DiseasesPast	NA(1), I10(2), K85.9(3), E11.9(4), Q50.5(5)
DiseasesCurrentOther	NA(1), Metastasis(2), Cholangiocarcinoma(3), Cystic(4), Biliary(5),Mass(6), Donor(7), Gallbladder(8), Hemangioma(9), HBV(10), Liver failure(11)
DiseasesCurrent	NA(1), C22.0(2), K75.9(3), Q50.5(4), D18.00(5), I82.0(6)
Gender	Female(1), Male(2), Unknown(3)
Age	$[< 31(1), 31 - 44(2), 45 - 60(3), 61 - 75(4), > 75(5)]$
Lab Results(Table 3.10)	Low(1),Normal(2),High(3)

(X_C) defined as

$$d_{Lesion}(i,j) = \frac{1}{L}\sum_{l=1}^{L}\left(1 - \frac{\sum_{n=1}^{N}\delta(X_n^i, X_n^j).w_{nl}}{\sum_{n=1}^{N} w_{nl}}\right) \tag{3.39}$$

where $\delta(a,b)$ is the Dirac delta function.

$$\delta(a,b) = \begin{cases} 1 & a = b, \\ 0 & a \neq b. \end{cases} \tag{3.40}$$

$N = 30$ is the length of the lesion related semantic annotations, X_n^i is the n_{th} lesion annotation label of the query image i in the current CMIA iteration , w_{nl} is the weight presenting the importance of the co-occurrence of the n_{th} annotation between the query and the retrieved lesions, and L is the number of defined weight vectors. As described later in Section 4.3, in this dissertation $L = 4$. The $d_{non-Lesion}$ is the similarity, based on non-lesion data ($[Y_U, Y_M]$) defined as

$$d_{non-Lesion}(i,j) = 1 - \frac{\sum_{n=1}^{35}\delta(Y_n^i, Y_n^j)}{35}. \tag{3.41}$$

For a query lesion, the similarity distance d_{all} is computed with respect to all lesions in the dataset and sorted in the ascending order. The lower the distance, the more similar it is to the query lesion. As described earlier, at each CMIA iteration the annotated lesion features along with non-lesion image features and non-image metadata are used to compute the similarity distance between query lesion and every image in the database. Based on the computed similarity distance, lesions in the database, are sorted from the most to the least similar to the query lesion.

4. EXPERIMENTS

4.1. Data

Dataset contains 135 3D abdominal CT scans from 123 patients (58 men, 65 women, mean age of 54 years, and age range of $15 - 85$ years). The data includes eight types of liver lesion diagnoses: metastasis (32), hemangioma (4), hepatocellular carcinoma (30), focal nodular hyperplasia (1), abscess (inflammation) (1), cyst (19), colangiocarcinoma (3), and unknown cause (45). Each image was manually annotated by a board-certified radiologist via an in-house on-line web application, which required the radiologist to fill in an annotation form composed of lesion and non-lesion image related semantic features described in Section 3.3.2. The non-image metadata (Section 3.3.3) was manually input by the radiologist using printed radiology reports.

The abdominal CT scans were segmented using the interactive 3-level hierarchical segmentation, and low-level image features were computed for each liver along with the vessels and lesions inside as described in Section 3.3. The image preparation and segmentation mask of a sample liver image are illustrated in Figure 4.1. All lesion and non-lesion semantic features and non-image metadata are processed and the corresponding labels (X_C, Y_U, Y_M) are produced as described in Section 3.4.1 and Section 3.5. Table 4.1 illustrate the lesion related semantic features, while Table 4.2 shows the non-lesion features defined for the mentioned sample case.

The dataset was collected at İstanbul University, İstanbul Medical School, Department of Radiology under the approval of İstanbul University clinical Research Ethics Committee (approval no 372, file no 2010/206-14). The data was anonymized on site following HIPAA guidelines within the CaReRa project. [1]

[1]This work was in part supported by TÜBİTAK ARDEB Grant # 110E264 and by Boğaziçi University Research Funds Grant # 5324

(a) (b) (c)

Figure 4.1. Sample case: (a) The original image ($I(\mathbf{r})$), (b) the preprocessed image ($I_0(\mathbf{r})$), and (c) the liver S_L (red), vessel S_V (green) and lesion S_E (blue) segmentation masks overlayed on the original image.

4.2. Experiments of Computer Aided Medical Image Annotation (CMIA)

As the training step, the lesion-related semantic features X_C along with their corresponding manually selected low-level image features F_I are used to train the individual binary/multi-label SVM classifiers as described in Section 3.4.2. Simultaneously, the lesion-related semantic features are utilized to learn the network models and their parameters as described in Section 3.4.3.

As discussed in Section 3.4.3, CMIA was used to generate lesion annotation labels. For a test lesion CT image, image features are extracted and input to pre-trained SVM classifiers to generate the values of the observation nodes (X_S). At every CMIA iteration, the semantic annotation of lesion semantic features are predicted via applying Bayesian inference in the Bayesian network model using the observation values and user's manual input as evidence. The semantic feature (concept) *Cluster* represents the number of segments/parts of a lesion type. It is directly observed from segmentation output, hence not included in automatic annotation process. Therefore, 29 out of 30 concept nodes are subjected to CMIA.

Table 4.1. Sample case: lesion related semantic features and their semantic annotations along with their annotation label.

Abbr.	**Annotations(value)**	**Abbr.**	**Annotations(value)**
Lobe	Right-Left(4)	ContUp	False(4)
GalAdj	False(1)	ContPat	False(6)
PerLoc	True(2)	Comp	PureSolid(3)
SubLoc	False(1)	Level	N/A(2)
CentLoc	False(1)	Dens	Hypo.(2)
Adj	False(1)	ADens	Homoge.(2)
Adt	False(1)	WaCont	False(1)
Ben	True(2)	CaA	False(3)
Inv	False(1)	CaCa	False(1)
NA	False(1)	CaPo	False(1)
Cluster	1	SepW	Thin(2)
MarT	Welldefined(5)	CaPs	False(1)
ASh	Ovoid(3)	CaSep	Punctate(2)
WaT	Thin(2)	CaSo	False(1)
SepDia	Complete(1)	CaWa	False(3)

Performance is assessed via prediction accuracy versus the iteration number, which is also equal to the number of manual annotations k (see Section 3.4.5), and reported as $Acc(k)$ given in Equation 4.1. During the experiments, the manual annotations were provided by setting the concept states to their *correct* values. Prediction accuracy is defined as the percentage of correctly annotated concept nodes.

$$Acc(k) = \frac{k + \sum_{c \in C \setminus E_k} \delta(\hat{x}_c, x_c)}{K} \tag{4.1}$$

where $\delta(a, b)$ is the Dirac delta function. The sooner the accuracy $Acc(k)$ approaches 1, the better the model is. Note that, for $k = 0$ the estimations are made without any

manual annotations whereas for $k = K$, the user has already manually provided all the annotations for the concept variables making $Acc(K) = 1$.

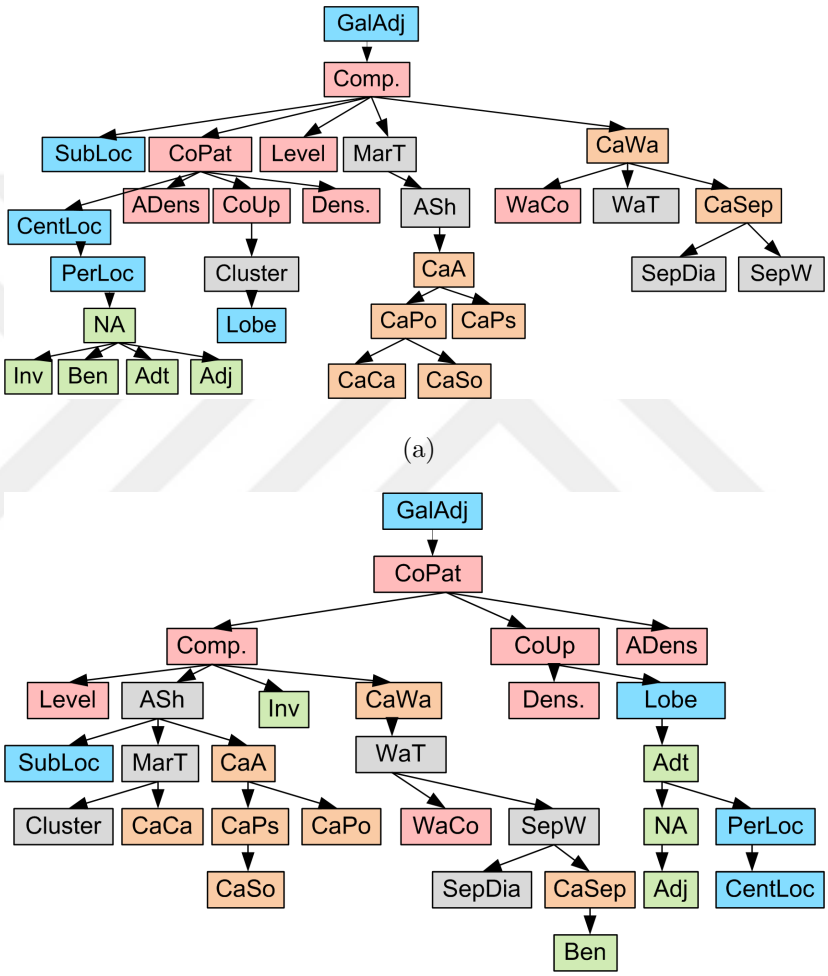
A two-fold cross validation study was performed, for which the dataset is manually divided into two by spreading the cases in a way that the samples of each and every lesion semantic label is uniformly distributed in both two sets as much as possible. Each data subset is used for training and testing in separate runs.

The $Acc(k)$ curves computed for each test case in both runs of the 2-fold cross-validation are averaged over all test results. The required number of manual annotations (k^*) to reach $Acc^* = 90\%, 95\%$ and 99% for each test case, i.e. $k^* = \underset{k}{\text{argmin}}(Acc(k) \geq Acc^*)$, are also computed and their means over all lesions are reported in following experiments.

4.2.1. The Importance of Hidden (Latent) Nodes

The added value of latent variables in the model is demonstrated by comparing $Acc(k)$ curves for $\mathcal{G}_{MWST}$ and $\mathcal{G}_N$, which are respectively, a maximum weight spanning tree (MWST) constructed over X_C only (no hidden node) [90], and a *naïve* Bayes model with a single 3-state hidden root. Figure 4.2 demonstrate the $\mathcal{G}_{MWST}$ models learnt in each fold of the cross validation.

Figure 4.3 shows the average $Acc(k)$ curves of 2-fold cross validation experiments using these two models. The results show the significant improvement in annotation performance due to the addition of a single hidden node to the network model specifically for $k < 17$ (number of manual annotations). The average number of required manual annotations to achieve certain accuracy levels are summarized in Figure 4.4 where the difference in the percentage of required manual annotations, between $\mathcal{G}_{MWST}$ and $\mathcal{G}_N$, are 10.68%, 8.58% and 6.97% for accuracy levels of 90%, 95% and 99%, respectively. It clearly demonstrates the importance of latent variable addition to the network from the radiologist point of view.



(a)

(b)

Figure 4.2. The purely data driven, domain-blind $\mathcal{G}_{MWST}$ models, constructed automatically using (a) the first group and (b) the second group of the dataset.

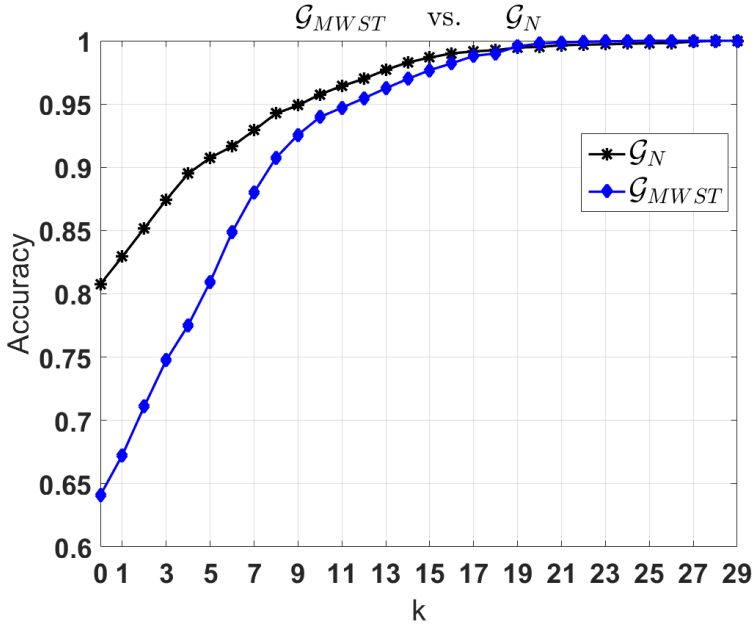


Figure 4.3. The annotation performance, $Acc(k)$ curves, versus number of manual annotations k, for $\mathcal{G}_{MWST}$ and $\mathcal{G}_N$ over 2-fold cross-validation experiments.

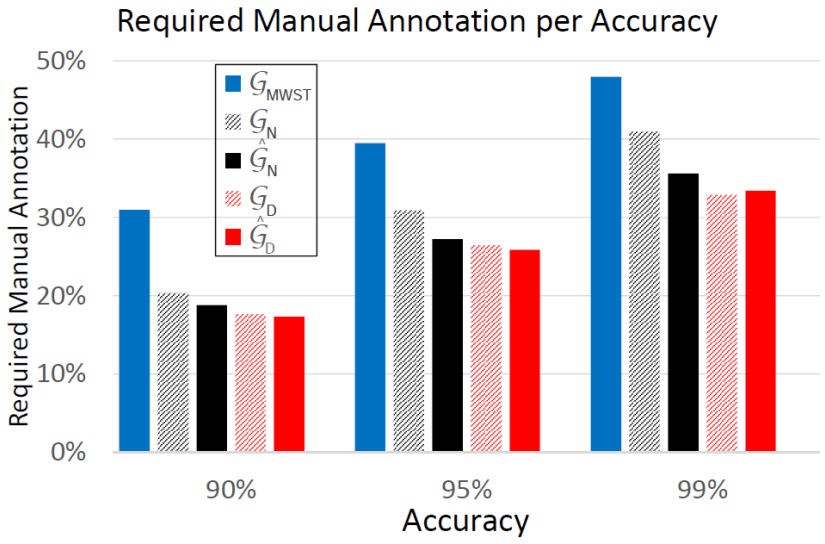


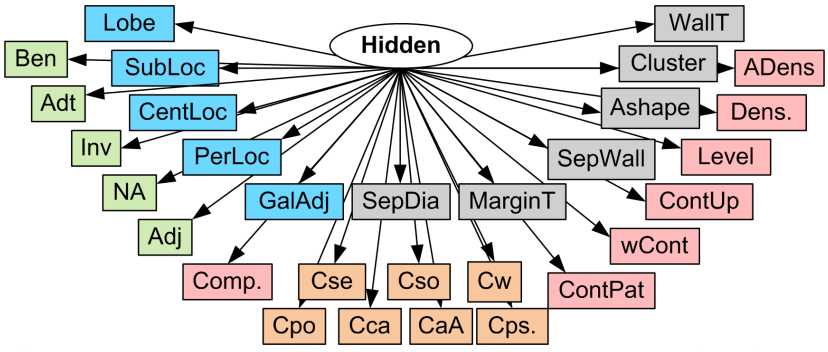
Figure 4.4. The mean percentage of required manual annotations to reach 90%, 95%, and 99% accuracy for $\mathcal{G}_{MWST}$, $\mathcal{G}_N$, $\hat{\mathcal{G}}_N$, $\mathcal{G}_D$, $\hat{\mathcal{G}}_D$ models.

4.2.2. Domain-Aware vs. Domain-Blind Models

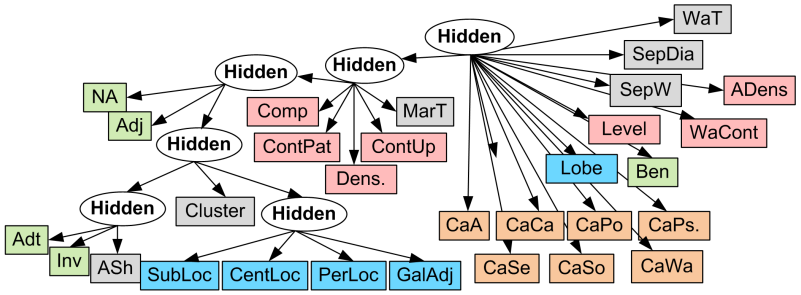
The effect of utilizing the prior domain knowledge in construction of the Bayesian network models is assessed by comparing domain-blind and domain-aware models before and after the model learning described in Section 3.4.4. Figures 4.5 and 4.6 illustrate the domain-blind and domain-aware models before and after model learning, where the concept nodes (each of which is connected to an observation node that is not illustrated in the figure) are depicted with rectangles and the hidden nodes are depicted with circles. Figure 4.5(a) shows the domain-blind *naïve* Bayes model $\mathcal{G}_N$, whereas Figures 4.5(b) and 4.5(c) depict the domain-blind model $\hat{\mathcal{G}}_N$ learned by mS-EM algorithm from the initial $\mathcal{G}_N$ over the first and second group of data respectively. Figure4.6(a) illustrate the domain-aware network model constructed manually by exploiting the domain knowledge, $\mathcal{G}_D$, while Figures 4.6(b) and 4.6(c) show the domain-aware network model $\hat{\mathcal{G}}_D$ learned by mS-EM algorithm started from the initial $\mathcal{G}_D$ over the first and second group of data respectively.

Specifically, comparisons between the domain-blind $\mathcal{G}_N$ and $\hat{\mathcal{G}}_N$ and domain-aware models $\mathcal{G}_D$ and $\hat{\mathcal{G}}_D$ were carried out. $\hat{\mathcal{G}}_N$ is the model learned from each fold of data through mS-EM algorithm initialized from *naïve* model $\mathcal{G}_N$. $\hat{\mathcal{G}}_D$ is the model learned from each fold of data through mS-EM algorithm initialized from domain-aware model $\mathcal{G}_D$.

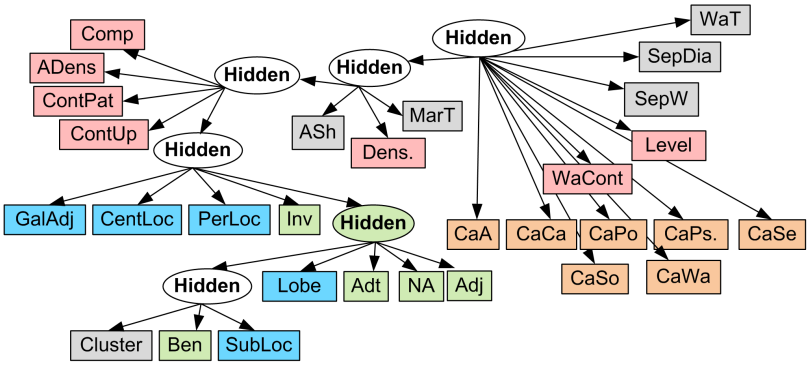
Figure 4.7 shows the corresponding $Acc(k)$ curves. As illustrated in Figure 4.4, a comparison of $\hat{\mathcal{G}}_N$ and $\hat{\mathcal{G}}_D$ reveals an improvement of 1.51%, 1.32% and 2.20% for 90%, 95% and 99% accuracy, respectively. Despite the significant improvement from $\mathcal{G}_N$ to $\hat{\mathcal{G}}_N$, the domain aware model's performance is not surpassed. The comparison of $\hat{\mathcal{G}}_N$ and $\mathcal{G}_D$, showing 1.15%, 0.77% and 2.73% improvement for 90%, 95% and 99% accuracy, confirms this observation. The results indicate that the model learning is capable of improving the performance of the domain-blind model more than that of the domain-aware model. However an important observation is that the latter always outperforms the former. Slight decrease in the performance of $\mathcal{G}_D$ at 99% accuracy is an indication of the importance of prior domain knowledge in contrast to domain-blind



(a)

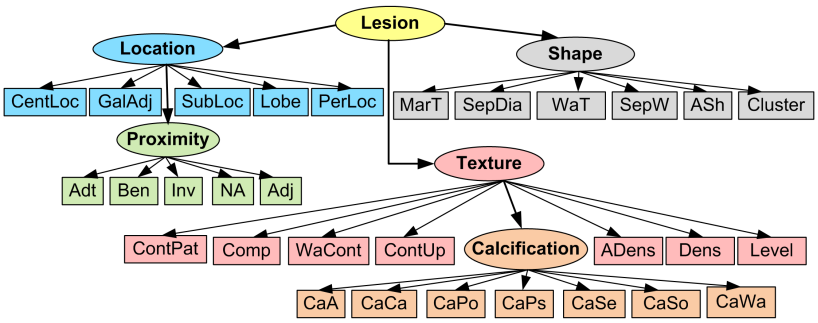


(b)

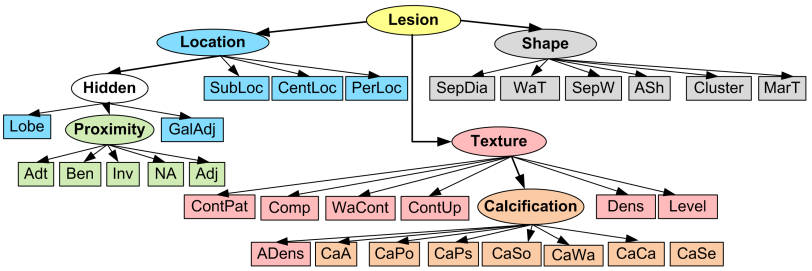


(c)

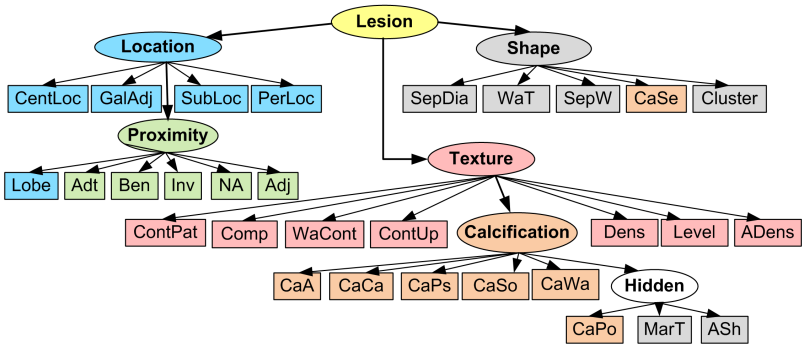
Figure 4.5. Model learning performance on domain-blind network. (a) $\mathcal{G}_N$, (b) $\hat{\mathcal{G}}_N$ on $1st$ fold and (c) $\hat{\mathcal{G}}_N$ on $2nd$ fold of data. The concept and hidden nodes are depicted with rectangles and circles respectively.



(a)



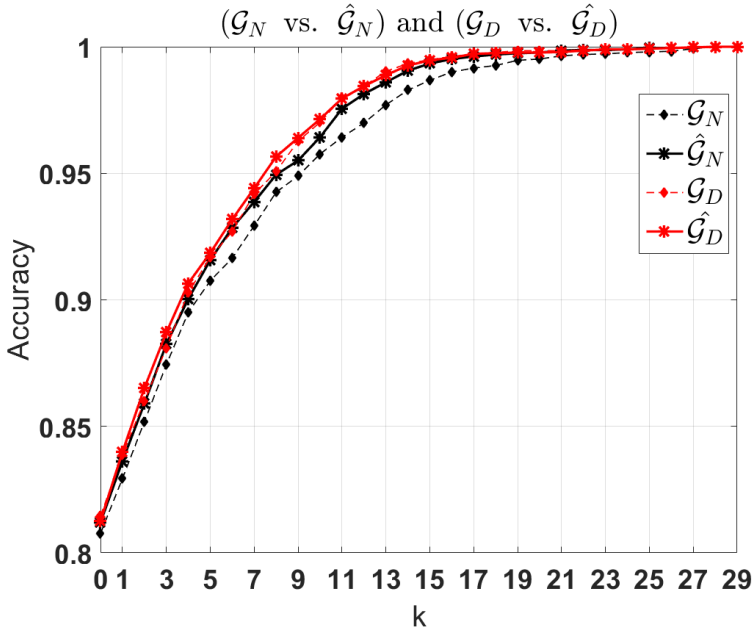
(b)



(c)

Figure 4.6. Model learning performance on domain-aware network. (a) $\mathcal{G}_D$, (b) $\hat{\mathcal{G}}_D$ on $1st$ fold and (c) $\hat{\mathcal{G}}_D$ on $2nd$ fold of data. The concept and hidden nodes are depicted with rectangles and circles respectively.

data-centric model construction approaches.



(a)

Figure 4.7. The $Acc(k)$ curves for $\mathcal{G}_N$, $\hat{\mathcal{G}}_N$, $\mathcal{G}_D$, $\hat{\mathcal{G}}_D$.

4.2.3. The Online Iterative Annotation

The use of the proposed model for dynamically selecting the next concept node (*online*) to be manually annotated (i.e. the next evidence to be provided manually to the system), as described in Section 3.4.5, in comparison to the *random* selection, and selection according to a prefixed order (*offline*) was evaluated. In the random selection case, there is no preference across the concept nodes and one is selected at random at each round of iteration of the annotation scheme for manual input. In the offline case, the concept nodes were pre-ordered according to their entropies computed on the training data. The proposed *online* selection scheme however falls within the *active learning* paradigm, where the network model is actively employed to select the most important node in an online manner, given the evidence (manual inputs) provided so far. All experiments were conducted using $\hat{\mathcal{G}}_D$ (i.e. the model learned from data using mS-EM algorithm starting from initial domain-aware model $\mathcal{G}_D$), and Figure 4.8 shows the corresponding $Acc(k)$ curves. The difference between random and offline/online selections is clearly obvious for all iterations and all accuracy levels which are above

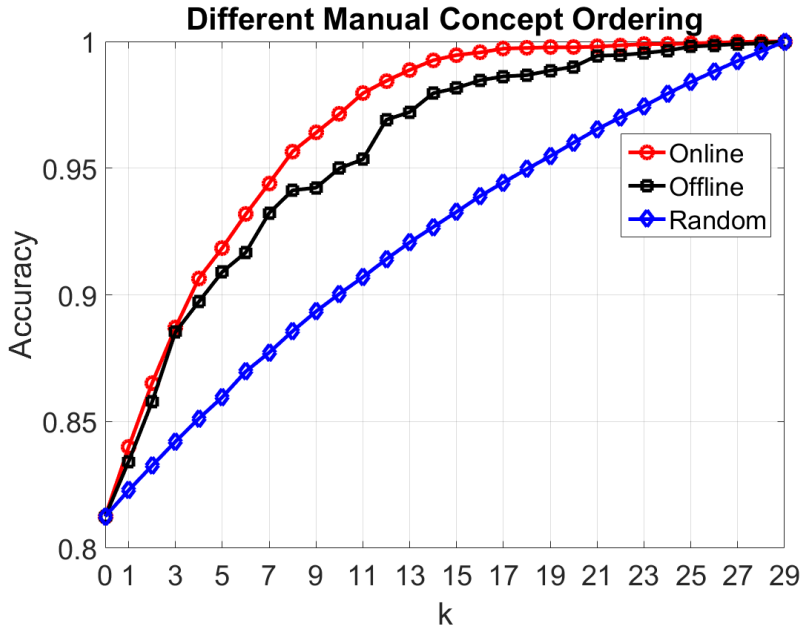


Figure 4.8. The iterative annotation accuracy $Acc(k)$ curves with random, offline (pre-ordered), and active (online) manual node selection schemes versus number of manual annotations k.

80%, whereas the difference between online and offline selection is significant after $k = 4$ iteration for accuracy levels higher than 90%. The mean number of concept nodes to be annotated manually at each iteration to reach 90%/95%/99% accuracy with $\hat{\mathcal{G}}_D$ model is 9.34/17.01/25.76 for random selection, 6.24/9.11/13.32 for offline selection and 5.01/7.50/9.68 for online selection. The results clearly illustrate the superiority of the online selection for the annotation performance in comparison to the offline and random selection schemes. From another point of view, these results show that 7.50/9.11/17.01 number of manual annotations are required on the average for 95% prediction accuracy with online/offline/random node selection, respectively. This corresponds to a 27.3% and 32.8% improvement from random (no-guidance) node selection to offline and online node selections, respectively.

4.3. Retrieval Experiments

As described in Section 3.5, similar lesion search and retrieval is applied given a query lesion image. The similarity ground truth $\rho \in (1, 3)$ is provided by a different radiologist for 10 versus 50 lesion cases, where $\rho = 1$ stands for *non-similar* and $\rho = 3$

represents *very similar* lesions. The similarity values are entered manually for each of the 10 test lesions in comparison to every 50 lesions in the database generating in 500 similarity values. The ground truth includes the annotation labels of four semantic features (without any knowledge of previous annotation labels) as *Contrast*, *Composition*, *Periphery* and *Density*, which are referred to as $\widetilde{X}_l$, $l \in \{l, ..., L = 4\}$ as listed in Table 4.3.

However, as described earlier, image annotation is an subjective task, hence for some lesions the annotation labels $\widetilde{X}_L$ may be different from the corresponding annotation labels X_C. The gold standard similarities are based on subjective annotations $\widetilde{X}_L$. Hence, the feature weighting for every semantic annotation in X_C with respect to every semantic annotation in $\widetilde{X}_L$ is applied to compensate the weighting in Equation (3.38). These importance weights W are learned computing the F-score of annotations in X_C for each one of the subjective annotations in $\widetilde{X}_l$ [91]. The normalized weight matrix is defined as $W = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1L} \\ w_{21} & w_{22} & \dots & w_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ w_{N1} & w_{N2} & \dots & w_{NL} \end{bmatrix}$ where w_{nl} is the weight assigned to the n_{th} annotation in X_C with respect to the l_{th} annotation in $\widetilde{X}_l$, where $N = 30$ and $L = 4$.

The model $\hat{\mathcal{G}}_D$ was used in the retrieval experiment which was learned by applying the mS-EM algorithm (Section 3.4.4) on the whole dataset except the 10 test cases initialized with $\mathcal{G}_D$ (Section 3.4.3). The iterative retrieval scheme as described in Section 3.5 is performed for 10 test cases, for which the ground truth similarity to 50 cases are known. At each iteration, user is asked to input the most critical concept value and after performing the annotation for the rest of unknown concepts, the retrieval is performed for every query lesion. All of 50 lesion cases are ranked in the ascending order based on their similarity to query.

The retrieval performance at each iteration k and each retrieved position p is computed via normalized discounted cumulative gain ($NDCG$) evaluation metric [92].

The $NDCG$ is a standard technique used to measure the effectiveness of information retrieval algorithms, when the ground truth is available. It is performed based on the position of the retrieved images (lesions images) in the rank list. Higher $NDCG$ in a position indicates that more similar lesions than dissimilar ones to the query were appeared in front of the list than those appearing later up to this position. $NDCG$ is defined as the ratio of discounted cumulative gain (DCG) over ideal discounted cumulative gain ($IDCG$) as follows

$$NDCG(k,p) = \frac{DCG(k,p)}{IDCG(k,p)} \tag{4.2}$$

where

$$DCG(k,p) = \left\{ \begin{array}{ll} \rho_p, & p = 1 \\ DCG(k,p-1) + \frac{\rho_p}{log_2(p)}, & \text{o.w.} \end{array} \right\} \tag{4.3}$$

in which ρ_p is the similarity value of the $p-th$ retrieved lesion, and $IDCG$ is defined as the DCG when all cases were retrieved in the correct (ideal) order, which is

$$IDCG(k,p) = max(DCG(k,p)). \tag{4.4}$$

Note that, $NDCG(p) = 1$ implies the perfect retrieval of p images. Finally, the averaged retrieval performance over all 10 test cases is computed at each iteration k and each position p as

$$aNDCG(k,p) = \frac{1}{10}\sum_{q=1}^{10}(NDCG_q(k,p)) \tag{4.5}$$

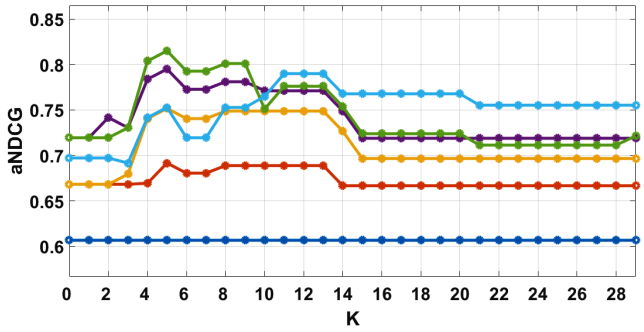
where $NDCG_q$ is the NDCG values of the test lesion q.

As stated earlier, a lesion case is described by two types of feature vectors: lesion labels/annotation X_C and non-lesion features Y_N; $N = \{1, ..., 35\}$ including liver and vessel labels Y_U (listed in Table 3.9) as well as the metadata Y_M (listed in Table

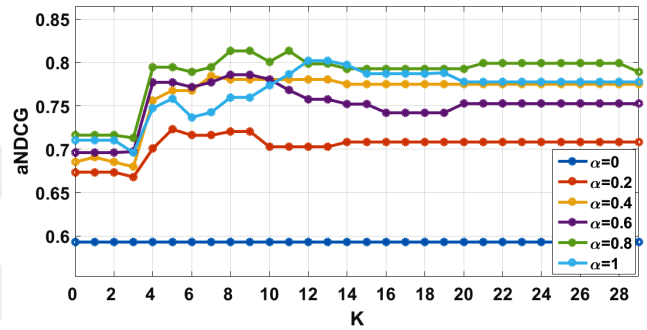
3.11). The retrieval is performed by combining these features to compute the similarly distance d_{all} given in Equation (3.38). The effect of these features in the retrieval performance is assessed by varying α from 0 to 1, where $\alpha = 0$ and $\alpha = 1$ stand for only non-lesion features and only lesion features, respectively.

Figure 4.9 illustrates the effect of α on the retrieval performance at positions $p = 1, 2, 3, 5$. As can be seen, results of non-lesion data augmentation in the retrieval renders a better performance despite the fact that the relevancy gold standard is only based on the texture related semantic features. Moreover, 20% contribution of non-lesion data ($\alpha = 0.8$) in the similarity measure results in the best performance. This observation is due to the discrepancy between the subjective and defined annotations, which is complemented by non-lesion feature. Results also show that after 4 manual annotations, the retrieval performance reach more than 76% at each position. For few number of manual inputs $k < 4$, since the annotations are not accurate enough, the non-lesion features may not be able to fill the missing information and therefore fail to perform better. Note that, the desirable case is to have fewer annotations and yet higher retrieval performance in the beginning of the retrieval positions. This can be accessed by 4 manual annotations and the contribution factor of $\alpha = 0.8$. Figure 4.10 depicts the $aNDCG(k, p)$ surface for $\alpha = 0.8$ and $\alpha = 1$, which shows the improvement of retrieval performance by increasing the number of manual annotation k at every position p.

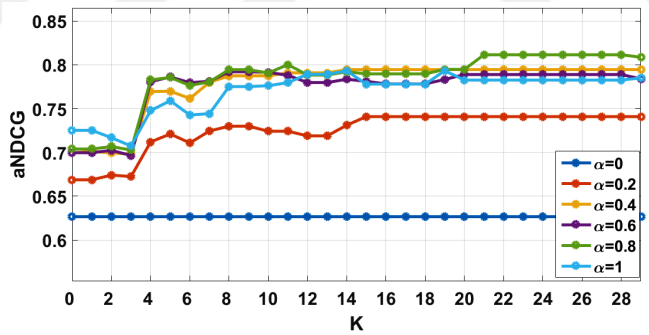
In order to visually compare the results, one high and one low performing retrieval results are illustrated in Figures 4.11 and 4.12, respectively, where each row represents the retrieved images at the current iteration k. In other words, the first row means that no manual input is provided, while the last row includes five manual annotations. Each column is the position of the retrieved images. Note that, green color stands for $\rho(i, j) > 1.8$, yellow represents $1.4 < \rho(i, j) < 1.8$, and $\rho(i, j) < 1.4$ is depicted in red. These intervals are chosen by $\frac{max(\rho)-min(\rho)}{3}$, which is the interval of provided similarity gold standard divided by three. As can be seen, in both high and low performing retrieval results, after four manual annotations more similar images are retrieved, which confirms the previous discussions.



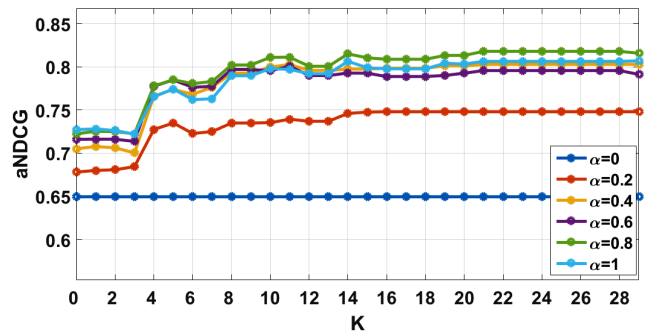
(a)



(b)

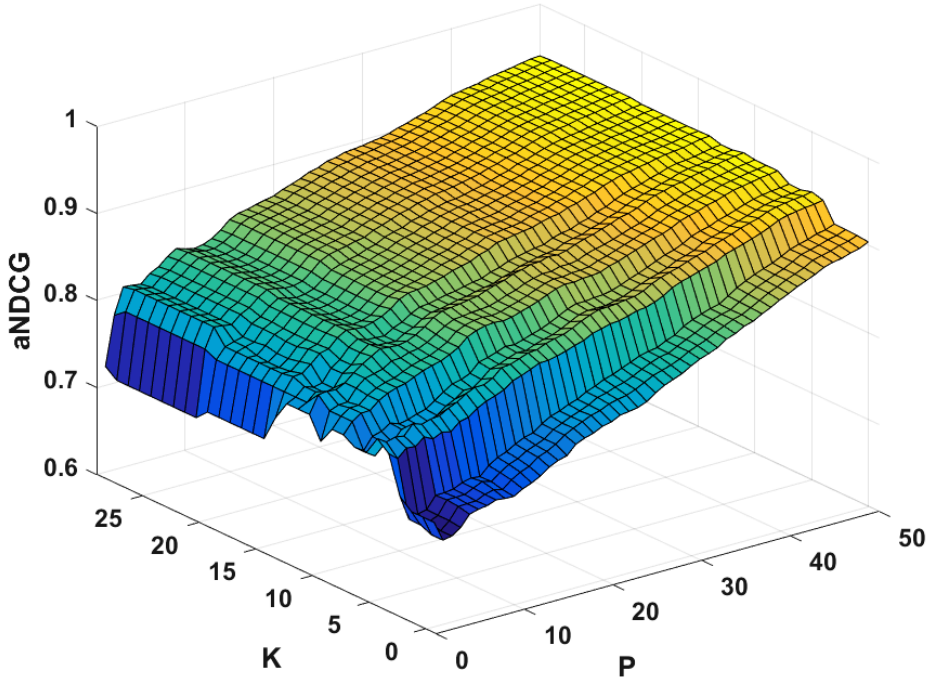


(c)

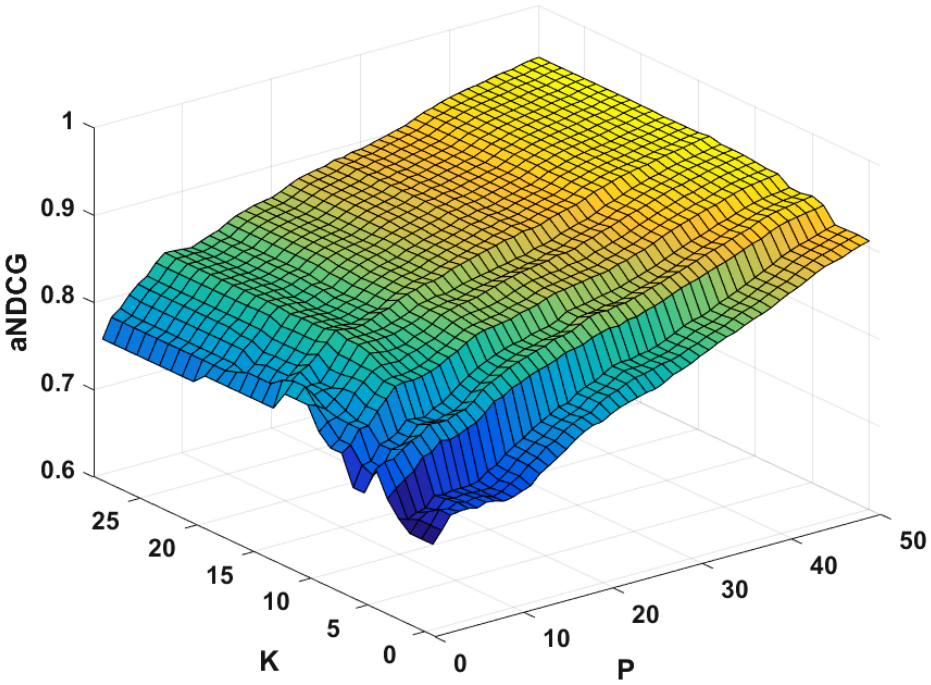


(d)

Figure 4.9. The $aNDCG(k, p)$ at retrieved position p and iteration (manual annotation) k, where the similarity distance is computed as a function of α. (a) $p = 1$, (b) $p = 2$, (c) $p = 3$, and (d) $p = 5$.

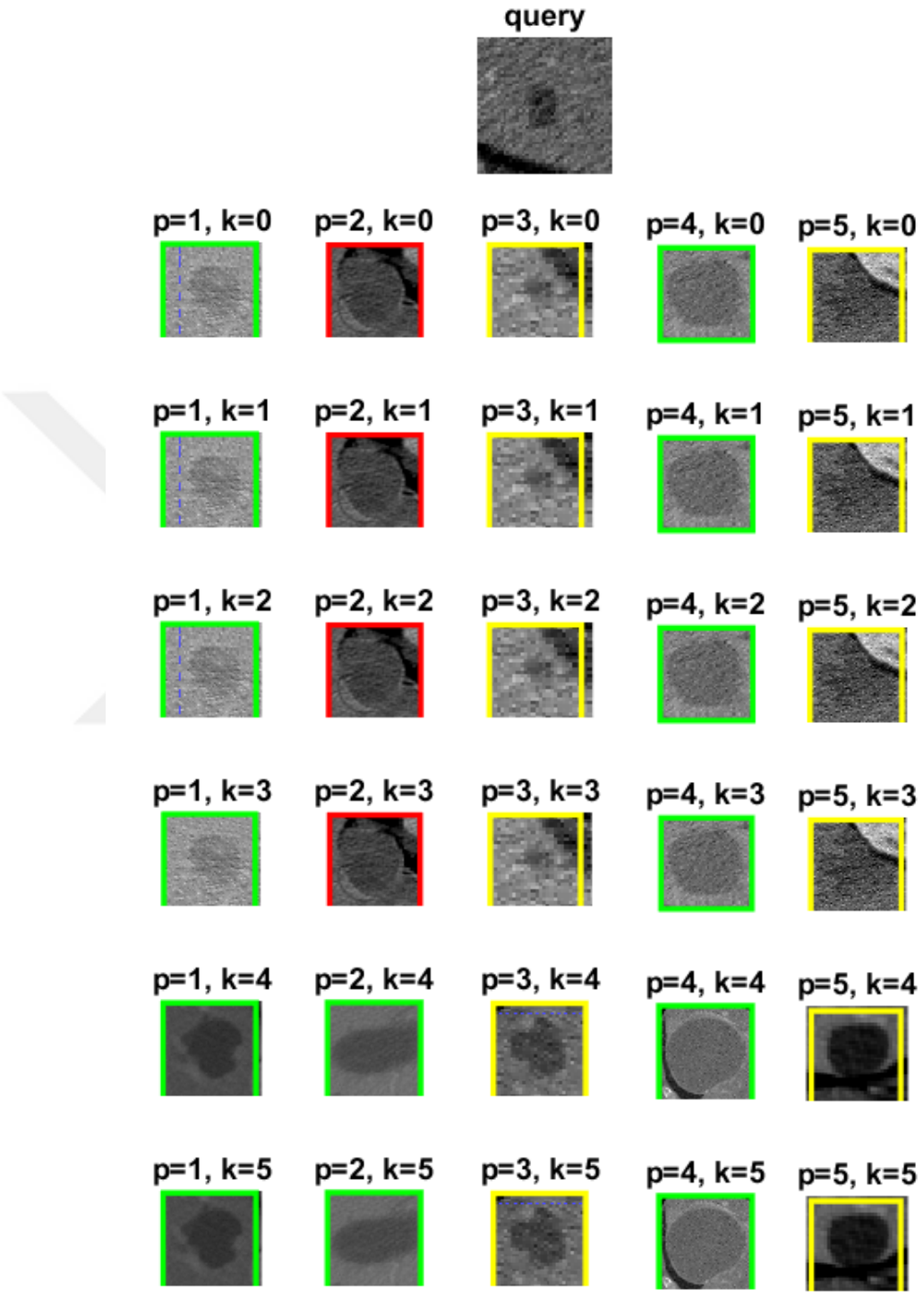


(a)



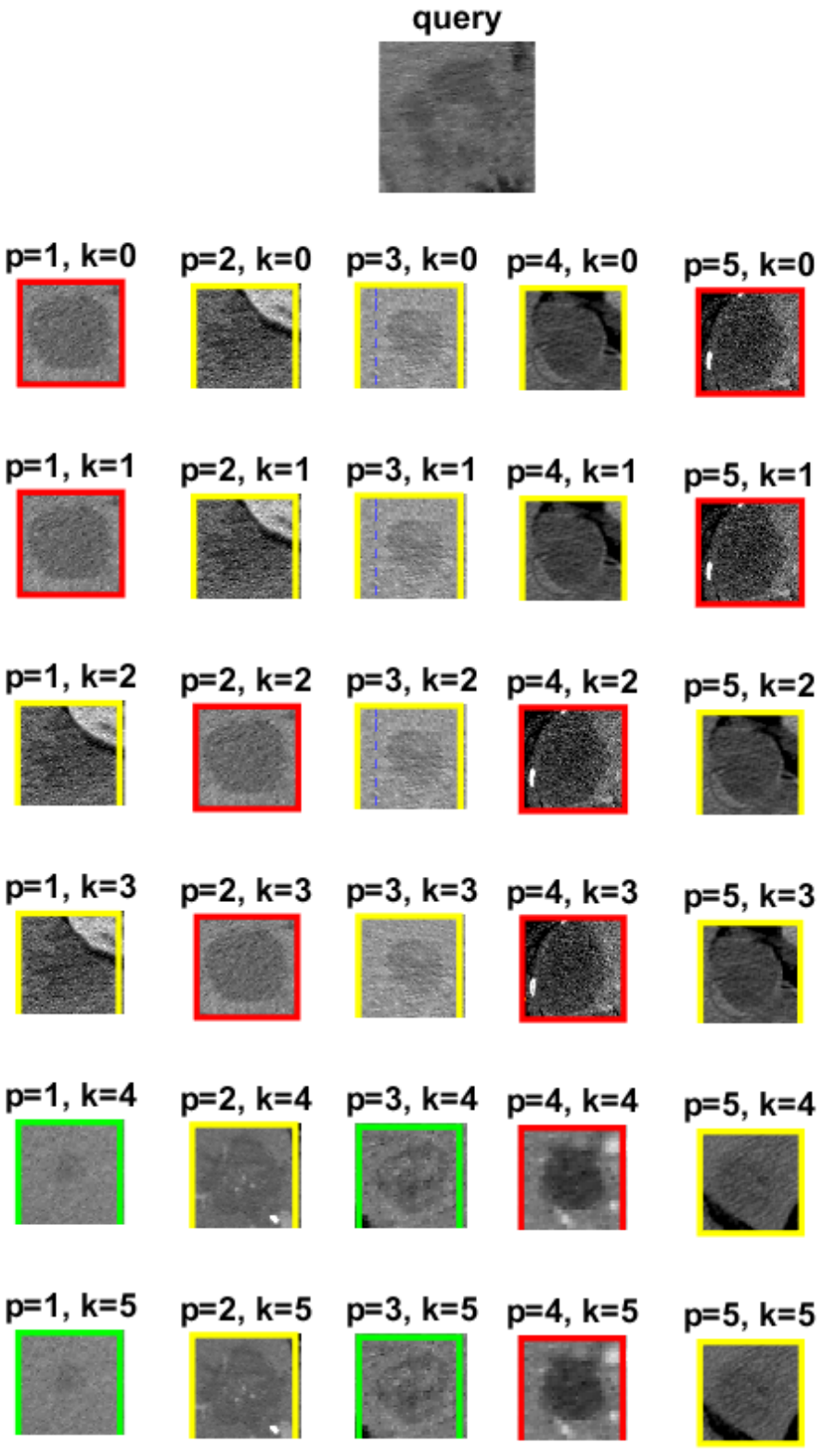
(b)

Figure 4.10. The $aNDCG(k, p)$ at retrieved position p and iteration (manual annotation) k for a fixed α value. (a) $\alpha = 0.8$ and (b) $\alpha = 1$.



(a)

Figure 4.11. One of the best performing retrieval results. Each row contains the first five retrieved images (p) at each iteration (k).



(a)

Figure 4.12. One of the lowest retrieval results. Each row contains the first five retrieved images (p) at each iteration (k).

Table 4.2. Sample case: non-lesion data and their corresponding labels.

Name	**Annotations**	**Name**	**Annotations**
LiverContour	Regular(2)	Albumin	2.9(1)
LiverDensityChange	Normal(2)	ALP	260(3)
hasDensity	Homogeneous(2)	ALT	71(3)
Liver_SizeChange	Normal(3)	AST	23(2)
RightLobe_SizeChange	Normal(3)	d-Bilirubin	3.09(3)
LeftLobe_hasSizeChange	Normal(3)	i-Bilirubin	1.17(3)
CaudateLobe_SizeChange	Normal(3)	t-Bilirubin	4.26(3)
Lumen_HepaticArtery	Normal-Open(2)	Pt	14(2)
Lumen_LeftPortalVein	Normal-Open(3)	LDH	258(2)
Lumen_RightPortalVein	Normal-Open(4)	GGT	109(3)
Lumen_LeftHepaticVein	Normal-Open(1)	APTT	33.2(3)
Lumen_MiddleHepaticVein	Normal-Open(2)	Afp	3(2)
Lumen_RightHepaticVein	Normal-Open(2)	Jaundice	T(2)
DiseasesPastOther	pulmonary (10)	Odema	F(1)
DiseasesPast	K85.9(3)	Ascites	F(1)
DiseasesCurrentOther	Liver-cyst(4)	Gender	M(2)
DiseasesCurrent	NA(1)	Age	73(4)
BloodPresure	120/80(2)		

Table 4.3. Subjective annotations used for manual similarity assessment.

Name	Lable	Range
Contrast	$\widetilde{X}_1$	Homogeneous, Heterogeneous
Composition	$\widetilde{X}_2$	Solid, Cystic,Mix
Periphery	$\widetilde{X}_3$	Smooth,Infiltrated
Density	$\widetilde{X}_4$	Hyper-dense,Hypo-dense,Iso-dense

5. DISCUSSION

Every content based medical image retrieval system contains three main components: *(a)* image content representation, *(b)* subjective query generation, and *(c)* assessing the similarity between the query and the cases in the database. In this work to represent the image content, 58 features were extracted from the image. These features represented the shape, location and texture of the liver, vessels and lesions. Texture features were extracted from the intensity image, whereas the shape and location features were extracted from the generated segmentation masks. Although the methods used to generate the segmentation masks were modified to capture the corresponding components the best, all segmentation masks were manually checked and edited. Any error due to the segmentation or feature extraction can propagate to the query generation part and result in low retrieval performance.

In this dissertation, subjective queries are the visual related semantic feature vectors. A radiologist-in-the-loop computer-aided medical image annotation was proposed to generate the semantic annotations. The semantic annotations for each lesion image were generated from low-level image features. As the initial annotation step, SVM method was utilized where individual multi-label SVM classifiers were used to assign the initial semantic annotations to each semantic feature. These initial semantic annotations were used as observation nodes in the BN-based model to generate the final annotations. This BN model captures the causal relation between the semantic features and generates the full annotation considering the inter-relations for every lesion. The proposed domain-aware model $\mathcal{G}_D$ was manually generated from the ONLIRA ontology which represents the relations between lesion, vessels and liver, as well as the properties of them. Moreover, to increase the annotation accuracy an online learning method has been proposed to iteratively identify the most important feature to be manually annotated.

The proposed model $\mathcal{G}_D$ was generated by introducing 6 hidden nodes to group the related semantic features. The effect of hidden nodes in the annotation perfor-

mance was examined by comparing the annotation accuracy of domain-blind models $\mathcal{G}_{MWST}$ and $\mathcal{G}_N$. $\mathcal{G}_{MWST}$ was directly learned from the data and $\mathcal{G}_N$ is the *naïve* model generated manually by connecting all nodes to the hidden node. As described in Section 4.2.1, the annotation result shows the improvement in the performance only by introducing one hidden node to the model. This improvement is due to the fact that hidden nodes can capture the underlying structure from data, while decreasing the complexity of the network by reducing the number of its parameters. For instance, the number of parameters of $\mathcal{G}_{MWST}$ and $\mathcal{G}_N$ are 312 and 203 respectively.

In order to examine the annotation capability of the proposed domain aware network, the proposed mS-EM model learning was used to learn the model from the data itself. As the results summarized in Figure 4.7 show, the structure learning is capable of learning relatively complex structures with high CMIA performance, such as $\hat{\mathcal{G}}_N$, starting from a very simple *naïve* model. The improvement due to the structure learning/refinement is more evident between $\mathcal{G}_N$ and $\hat{\mathcal{G}}_N$, than between $\mathcal{G}_D$ and $\hat{\mathcal{G}}_D$, which is due to the difference in the number of latent variables of the former. The size of model parameters are 233, 233, and 239 and the number of hidden nodes are 6, 6, and 7 for $\hat{\mathcal{G}}_N$, $\mathcal{G}_D$, and $\hat{\mathcal{G}}_D$, respectively. Yet, $\hat{\mathcal{G}}_N$ does not outperform $\mathcal{G}_D$ or $\hat{\mathcal{G}}_D$. This result suggests that the domain knowledge has the potential of replacing costly and domain-blind structure learning methods that are also prone to over fitting to data in the case of incomplete and sparse training datasets. Hence, it can be concluded that the domain-aware network construction can by-pass the challenges of sparse training data and/or incomplete domain representation, whereas in the presence of rich data the proposed structure learning approach is capable of constructing a network with similar CMIA performance.

The difference between domain-aware and domain-blind models after model learning is less compared to the difference before applying the model learning. This confirms (a) the importance of latent variables, (b) the ability of proposed method to learn from data. Moreover, incorporating the prior domain knowledge has complementary benefits compared to data-centric domain-blind learning approaches. A closer look at the domain-blind model after model learning ($\hat{\mathcal{G}}_N$) shows that the mS-EM algorithm groups

texture, calcification and proximity nodes to a great extent but distributes shape and location nodes more. The behaviour of mS-EM, or similar data-centric structure learning algorithms, is driven by the data, hence in case of sparse data, they are prone to overfitting.

Increasing the cardinality of hidden nodes up to some level can increase the annotation performance of the model which is shown in Figure 5.1. These results show the performance of $\mathcal{G}_N$ while different cardinalities are assigned to the hidden node $(3, 7, 11, 18, 21)$ in comparison to the performance of the original $\mathcal{G}_D$. As can be seen, by increasing the number of states up to 7 the annotation performance improves and by further increasing the number of states it stars to decrease which is due to the overfitting problem. Although the performance of $\mathcal{G}_N$ improves by increasing the number of states, it can not outperform the performance of $\mathcal{G}_D$. Note that, by increasing the number of states of the hidden node the complexity of the model increases which can be seen from the number of parameters which are 475, 611, 747, 1223, and 1495 for 7, 11, 18, and 21 number of states, respectively. Thus, to keep the complexity of the models low, all the hidden nodes were assigned to have 3 states.

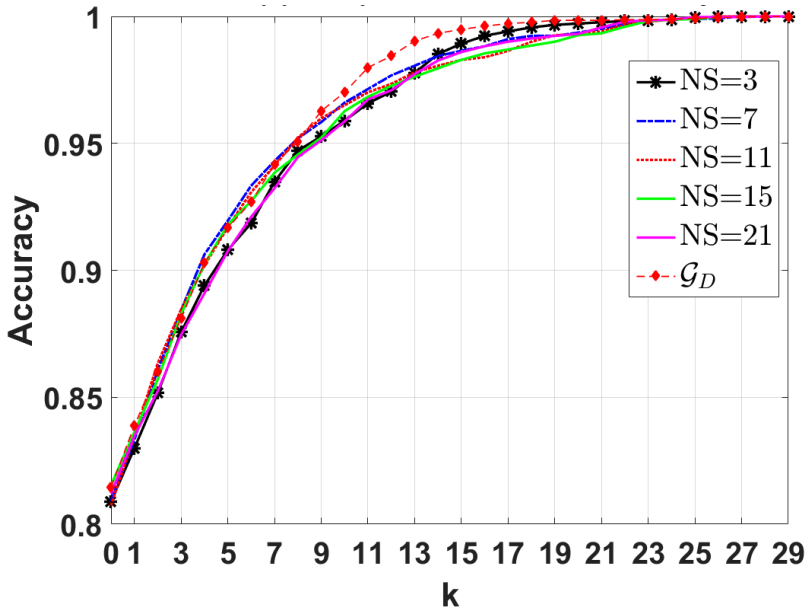
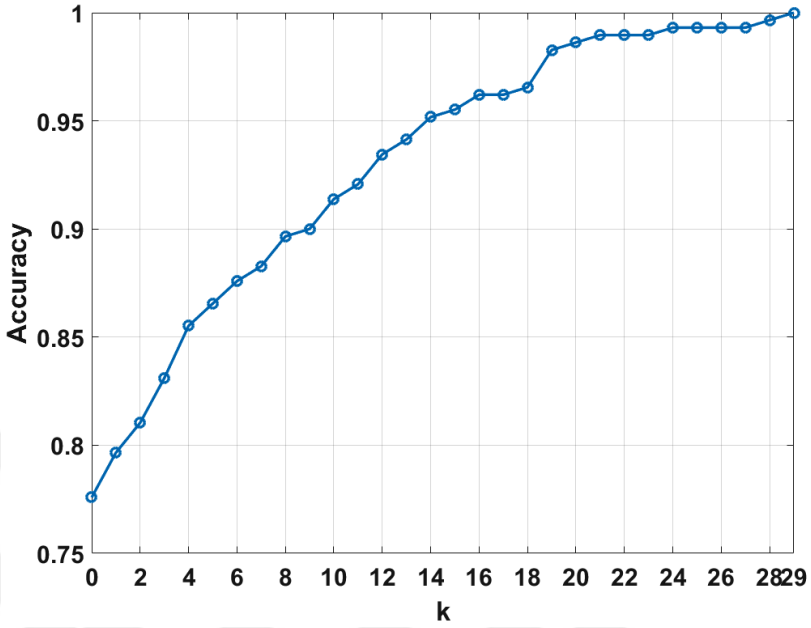


Figure 5.1. The effect of cardinality of hidden node in $\mathcal{G}_N$ over 2-fold cross-validation experiments.

Online annotation leads the manual input by identifying the most important node to be annotated at each iteration. Thus, it is more personalized, i.e. the selection order is more based on the user interest, whereas the offline method is population based, which means that the order selection is based on the dataset. This observation justifies the annotation performance superiority of the online selection method. The results clearly exhibit the benefits of case-specific online selection of the concept node for the next manual annotation, when the number of manual annotations are more than three. However, in the cases, where fewer annotations are needed (i.e. up to 3), the difference between the performance of offline and online node selection is subtle, which results in exploiting the low computational complexity of the offline method. As an intuitive case in which k is very large ($27 < k < 29$), no matter what the selection method is, the annotation performance is high due to the presence of nearly all observations. Applicational point of view, if the limitation on the number of manual input k is of the main concern, online method should be perused, whereas for less annotation runtime, offline method is of interest.

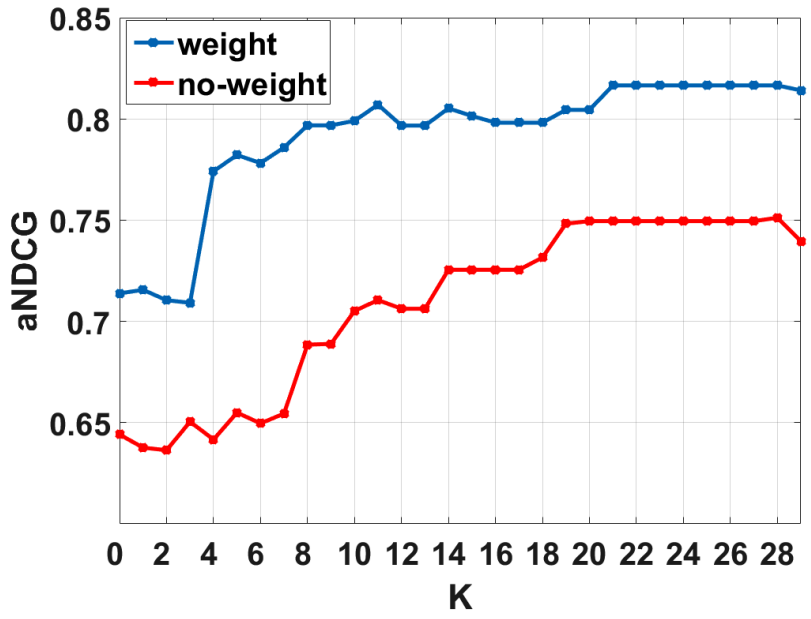
High performance of the proposed CMIA approach is expected to have a direct implications in the content based medical image retrieval system, which can be seen by considering the results illustrated in Figure 4.10. As a general trend, it is observed that by increasing the number of manual annotations, the retrieval performance improves due to the more accurate semantic features available. More important, by only feeding $k = 4$ manual annotations, retrieval performance reaches more than 77% in accuracy, which is similar to that of the annotation performance being more than 85% after 4 manual inputs as shown in Figure 5.2.

As described earlier, the gold standard similarity was provided by a different radiologist which was based on the given subjective semantic annotations $\widetilde{X}_L$. Whereas the computer-aided automatic annotation was carried out by considering the system defined semantic annotations X_C. Although this discrepancy poses challenges for search and retrieval, it also simulates a realistic experimental setup that includes the subjectivity of the similarity. To capture the criteria of the subjective gold standard, different weights were assigned to each feature. Figure 5.3 illustrates the importance



(a)

Figure 5.2. Annotation accuracy for the 10 query lesion images using the $\hat{\mathcal{G}}_D$ model trained on the rest of 125 images.



(a)

Figure 5.3. The effect of weight assignment for aNDCG at the 4_{th} position (p = 4).

of the weight assignment in the retrieval performance.

One of the main component of any retrieval system is the similarity metric. In this study, the co-occurrences between the feature vectors is used as the metric, however, as the future work, the same semantic relations considered in the CMIA step can be utilized to compute the similarity between the lesions. More clearly, currently two labels are not considered similar if they are not exactly the same. This strict similarity criterion can however be further loosen by defining a wider degree of similarity definitions. For instance, as for the lesion composition, "Solid" and "Cystic" are completely different, whereas "Solid" and "SolidCysticMix", from the pathological point of view, might share some similar information, which can be captured through a more complex similarity metrics via integration of the information from ONLIRA/Redlex ontologies in the similarity metric computation.

6. CONCLUSION

In this study, an interactive search and retrieval system is proposed via utilizing a flexible user-interactive computer-aided annotation for semantic query preparation to collect similar lesions from a dataset of 3D abdominal CT liver images using the restricted vocabularies generated based on ONLIRA ontology.

The results for a radiologist-in-the-loop computer aided medical image annotation system that uses a domain-aware Bayesian model are presented, which illustrate the importance of the latent variables, as well as the domain-knowledge in designing the model. To the best of my knowledge, the proposed system is the first CMIA method that (a) incorporates the inter-relations between annotation concepts, (b) selects the optimal node in an online manner for sequential manual inputs which falls in the domain of active learning. The use of the prior domain knowledge in model construction has also been evaluated and results support the idea of domain-aware model construction as an efficient and effective approach whose benefits would mostly be seen in case of sparse training data.

The results for the iterative search and retrieval system demonstrate the notable effect of the proposed radiologist-in-the-loop CMIA system to accelerate the retrieval performance. Moreover, the effectiveness of incorporating non-lesion feature to compensate the error of annotations for limited user interactions has been shown.

Further research would include the incorporation of the ONLIRA/Radlex ontology in the similarity assessment of the search and retrieval system.

REFERENCES

1. VAVlab, "CaReRa-WEB, Case Retrieval in Radiological Databases", `https://vavlab.ee.boun.edu.tr:5904/CareraWeb2/`, accessed at November 2017.

2. Wang, S. and R. M. Summers, "Machine Learning and Radiology", *Medical image analysis*, Vol. 16, No. 5, pp. 933–951, July 2012.

3. Barzegar Marvasti, N., C. B. Akgül, B. Acar, N. Kökciyan, S. Üsküdarlı, P. Yolum, R. Türkay and B. Bakır, "Clinical Experience Sharing by Similar Case Retrieval", *Proceedings of ACM International Workshop on Multimedia Indexing and Information Retrieval for Healthcare*, pp. 67–74, October 2013.

4. Akgül, C. B., D. L. Rubin, S. Napel, C. F. Beaulieu, H. Greenspan and B. Acar, "Content-Based Image Retrieval in Radiology: Current Status and Future Directions", *Journal of Digital Imaging*, Vol. 24, No. 2, pp. 208–222, April 2011.

5. Kökciyan, N., R. Türkay, S. Üsküdarlı, P. Yolum, B. Bakır and B. Acar, "Semantic Description of Liver CT Images: An Ontological Approach", *IEEE Journal of Biomedical and Health Informatics*, Vol. 18, No. 4, pp. 1363–1369, July 2014.

6. Smeulders, A. W., M. Worring, S. Santini, A. Gupta and R. Jain, "Content-Based Image Retrieval at the End of the Early Years", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 12, pp. 1349–1380, December 2000.

7. Datta, R., D. Joshi, J. Li and J. Z. Wang, "Image Retrieval: Ideas, Influences, and Trends of the New Age", *ACM Computing Surveys*, Vol. 40, No. 2, p. 5, April 2008.

8. Müller, H., N. Michoux, D. Bandon and A. Geissbuhler, "A Review of Content-Based Image Retrieval Systems in Medical Applications-Clinical Benefits and Fu-

ture Directions", *International Journal of Medical Informatics*, Vol. 73, No. 1, pp. 1 – 23, September 2004.

9. Jain, A. K. and A. Vailaya, "Image Retrieval Using Color and Shape", *Pattern Recognition*, Vol. 29, No. 8, pp. 1233–1244, August 1996.

10. Rubner, Y., C. Tomasi and L. J. Guibas, "The Earth Mover's Distance as A Metric for Image Retrieval", *International Journal of Computer Vision*, Vol. 40, No. 2, pp. 99–121, November 2000.

11. Bunke, H., C. Irniger and M. Neuhaus, "Graph Matching- Challenges and Potential Solutions", *Proceedings of International Conference on Image Analysis and Processing*, pp. 1–10, September 2005.

12. Keysers, D., H. Ney, B. B. Wein and T. M. Lehmann, "Statistical Framework for Model-Based Image Retrieval in Medical Applications", *Journal of Electronic Imaging*, Vol. 12, No. 1, pp. 59–68, January 2003.

13. Deselaers, T., H. Mller, P. Clough, H. Ney and T. M. Lehmann, "The CLEF 2005 Automatic Medical Image Annotation Task", *International Journal of Computer Vision*, Vol. 71, pp. 51–58, August 2007.

14. Xu, J., J. Faruque, C. F. Beaulieu, D. Rubin and S. Napel, "A Comprehensive Descriptor of Shape: Method and Application to Content-based Retrieval of Similar Appearing Lesions in Medical Images", *Journal of Digital Imaging*, Vol. 25, No. 1, pp. 121–128, February 2012.

15. Ballerini, L., X. Li, R. B. Fisher and J. Rees, "A Query-by-Example Content-Based Image Retrieval System of Non-Melanoma Skin Lesions", *Proceedings of MICCAI International Workshop on Medical Content-Based Retrieval for Clinical Decision Support*, pp. 31–38, Springer, March 2009.

16. Quellec, G., M. Lamard, G. Cazuguel, B. Cochener and C. Roux, "Wavelet Opti-

mization for Content-based Image Retrievalin Medical Databases", *Medical Image Analysis*, Vol. 14, No. 2, pp. 227–241, April 2010.

17. Wei, C.-H., S. Y. Chen and X. Liu, "Mammogram Retrieval on Similar Mass Lesions", *Computer Methods and Programs in Biomedicine*, Vol. 106, No. 3, pp. 234–248, June 2012.

18. Murala, S. and Q. J. Wu, "Local Mesh Patterns Versus Local Binary Patterns: Biomedical Image Indexing and Retrieval", *IEEE Journal of Biomedical and Health Informatics*, Vol. 18, No. 3, pp. 929–938, May 2014.

19. Tsochatzidis, L., K. Zagoris, N. Arikidis, A. Karahaliou, L. Costaridou and I. Pratikakis, "Computer-Aided Diagnosis of Mammographic Masses Based on A Supervised Content-based Image Retrieval Approach", *Pattern Recognition*, November 2017.

20. Shih, T. K., *Distributed Multimedia Databases*, IGI Global, Hershey, PA, USA, 2002.

21. Kurtz, C., A. Depeursinge, S. Napel, C. F. Beaulieu and D. L. Rubin, "On Combining Image-Based and Ontological Semantic Dissimilarities for Medical Image Retrieval Applications", *Medical Image Analysis*, Vol. 18, No. 7, pp. 1082–1100, October 2014.

22. Ma, L., X. Liu, Y. Gao, Y. Zhao, X. Zhao and C. Zhou, "A New Method of Content Based Medical Image Retrieval and Its Applications to CT Imaging Sign Retrieval", *Journal of Biomedical Informatics*, Vol. 66, pp. 148–158, February 2017.

23. Napel, S. A., C. F. Beaulieu, C. Rodriguez, J. Cui, J. Xu, A. Gupta, D. Korenblum, H. Greenspan, Y. Ma and D. L. Rubin, "Automated Retrieval of CT Images of Liver Lesions on The Basis of Image Similarity: Method and Preliminary Results", *Radiology*, Vol. 256, No. 1, pp. 243–252, May 2010.

24. Depeursinge, A., A. Vargas, F. Gaillard, A. Platon, A. Geissbuhler, P.-A. Poletti and H. Müller, "Case-Based Lung Image Categorization and Retrieval for Interstitial Lung Diseases: Clinical Workflows", *International Journal of Computer Assisted Radiology and Surgery*, Vol. 7, No. 1, pp. 97–110, January 2012.

25. Rahman, M. M., B. C. Desai and P. Bhattacharya, "Medical Image Retrieval with Probabilistic Multi-Class Support Vector Machine Classifiers and Adaptive Similarity Fusion", *Computerized Medical Imaging and Graphics*, Vol. 32, No. 2, pp. 95–108, March 2008.

26. Rahman, M. M., S. K. Antani and G. R. Thoma, "A Learning-Based Similarity Fusion and Filtering Approach for Biomedical Image Retrieval Using SVM Classification and Relevance Feedback", *IEEE Transactions on Information Technology in Biomedicine*, Vol. 15, No. 4, pp. 640–646, July 2011.

27. Kumar, A., J. Kim, L. Wen, M. Fulham and D. Feng, "A Graph-Based Approach for the Retrieval of Multi-modality Medical Images", *Medical Image Analysis*, Vol. 18, No. 2, pp. 330–342, February 2014.

28. Zhang, F., Y. Song, W. Cai, A. G. Hauptmann, S. Liu, S. Pujol, R. Kikinis, M. J. Fulham, D. D. Feng and M. Chen, "Dictionary Pruning with Visual Word Significance for Medical Image Retrieval", *Neurocomputing*, Vol. 177, pp. 75–88, February 2016.

29. Rahman, M. M., S. K. Antani and G. R. Thoma, "A Medical Image Retrieval Framework in Correlation Enhanced Visual Concept Feature Space", *Proceedings of IEEE International Symposium on Computer-Based Medical Systems*, pp. 1–4, IEEE, August 2009.

30. Tang, H. L., R. Hanka and H. H. S. Ip, "Histological Image Retrieval Based on Semantic Content Analysis", *IEEE Transactions on Information Technology in Biomedicine*, Vol. 7, No. 1, pp. 26–36, March 2003.

31. Dimitrova, N., "Context and Memory in Multimedia Content Analysis", *IEEE MultiMedia*, Vol. 11, No. 3, pp. 7–11, July 2004.

32. Goh, K. S., E. Y. Chang and B. Li, "Using One-Class and Two-Class Svms for Multiclass Image Annotation", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 17, No. 10, pp. 1333–1346, October 2005.

33. Lehmann, T., H. Schubert, D. Keysers, M. Kohnen and B. B. Wein, "The IRMA Code For Unique Classification of Medical Images", *Proceedings of SPIE*, Vol. 5033, pp. 440–451, May 2003.

34. Zhang, D., M. M. Islam and G. Lu, "A Review on Automatic Image Annotation Techniques", *Pattern Recognition*, Vol. 45, No. 1, pp. 346–362, January 2012.

35. Marvasti, N. B., N. Kökciyan, R. Türkay, A. Yazici, P. Yolum, S. Üsküdarli and B. Acar, "ImageCLEF Liver CT Image Annotation Task 2014", *Proceedings of CLEF 2014 Evaluation Labs and Workshops: Online Working Notes*, Vol. 1180, pp. 329–340, September 2014.

36. Dimitrovski, I., D. Kocev, I. Kitanovski, S. Loskovska and S. Džeroski, "Improved Medical Image Modality Classification Using a Combination of Visual and Textual Features", *Computerized Medical Imaging and Graphics*, Vol. 39, pp. 14–26, January 2015.

37. Alahmer, H. and A. Ahmed, "Computer-aided Classification of Liver Lesions From CT Images Based on Multiple ROI", *Procedia Computer Science*, Vol. 90, pp. 80–86, December 2016.

38. Depeursinge, A., C. Kurtz, C. Beaulieu, S. Napel and D. Rubin, "Predicting Visual Semantic Descriptive Terms From Radiological Image Data: Preliminary Results With Liver Lesions in CT", *IEEE Transactions on Medical Imaging*, Vol. 33, No. 8, pp. 1669–1676, August 2014.

39. Yang, W., Z. Lu, M. Yu, M. Huang, Q. Feng and W. Chen, "Content-based Retrieval of Focal Liver Lesions Using Bag-of-Visual-words Representations of Single and Multiphase Contrast-Enhanced CT Images", *Journal of Digital Imaging*, Vol. 25, No. 6, pp. 708–719, December 2012.

40. Gimenez, F., J. Xu, Y. Liu, T. T. Liu, C. F. Beaulieu, D. L. Rubin and S. Napel, "Automatic Annotation of Radiological Observations in Liver CT Images", *Proceedings of Symposium of American Medical Informatics Association*, pp. 257–63, November 2012.

41. Kumar, A., S. Dyer, J. Kim, C. Li, P. H. Leong, M. Fulham and D. Feng, "Adapting Content-based Image Retrieval Techniques For the Semantic Annotation of Medical Images", *Computerized Medical Imaging and Graphics*, Vol. 49, pp. 37 – 45, April 2016.

42. Mougiakakou, S. G., I. Valavanis, K. Nikita, A. Nikita and D. Kelekis, "Characterization of CT Liver Lesions Based on Texture Features and a Multiple Neural Network Classification Scheme", *Proceedings of International Conference of the IEEE Engineering in Medicine and Biology Society*, Vol. 2, pp. 1287–1290, October 2003.

43. Silla, J., N. Carlos and A. A. Freitas, "A Survey of Hierarchical Classification Across Different Application Domains", *Data Mining and Knowledge Discovery*, Vol. 22, No. 1-2, pp. 31–72, January 2011.

44. Mueen, A., R. Zainuddin and M. S. Baba, "Automatic Multilevel Medical Image Annotation and Retrieval", *Journal of Digital Imaging*, Vol. 21, No. 3, pp. 290–295, September 2008.

45. Ko, B. C., J. Lee and J.-Y. Nam, "Automatic Medical Image Annotation and Keyword-Based Image Retrieval Using Relevance Feedback", *Journal of Digital Imaging*, Vol. 25, No. 4, pp. 454–465, August 2012.

46. Dimitrovski, I., D. Kocev, S. Loskovska and S. Džeroski, "Hierarchical Annotation of Medical Images", *Pattern Recognition*, Vol. 44, No. 10, pp. 2436–2449, March 2011.

47. Zare, M. R., A. Mueen and W. C. Seng, "Automatic Classification of Medical X-Ray Images Using a Bag of Visual Words", *IET Computer Vision*, Vol. 7, No. 2, pp. 105–114, April 2013.

48. Caruana, R., "Multitask learning", *Learning to Learn*, pp. 95–133, Springer, 1998.

49. Torralba, A., K. P. Murphy and W. T. Freeman, "Sharing Features: Efficient Boosting Procedures for Multiclass Object Detection", *Proceedings of IEEE Computer Vision and Pattern Recognition*, Vol. 2, pp. 762–769, June 2004.

50. Fan, J., Y. Gao and H. Luo, "Integrating Concept Ontology and Multitask Learning to Achieve More Effective Classifier Training for Multilevel Image Annotation", *IEEE Transactions on Image Processing*, Vol. 17, No. 3, pp. 407–426, February 2008.

51. Fan, J., H. Luo, Y. Gao and R. Jain, "Incorporating Concept Ontology for Hierarchical Video Classification, Annotation, and Visualization", *IEEE Transactions on Multimedia*, Vol. 9, No. 5, pp. 939–957, August 2007.

52. Ladicky, L., P. H. S. Torr and A. Zisserman, "Human Pose Estimation Using a Joint Pixel-Wise and Part-Wise Formulation", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3578–3585, June 2013.

53. Chen, S., J. Qin, X. Ji, B. Lei, T. Wang, D. Ni and J.-Z. Cheng, "Automatic Scoring of Multiple Semantic Attributes With Multi-Task Feature Leverage: A Study on Pulmonary Nodules in CT Images", *IEEE Transactions on Medical Imaging*, Vol. 36, No. 3, pp. 802–814, March 2017.

54. Geman, D. and B. Jedynak, "An Active Testing Model for Tracking Roads in

Satellite Images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 18, pp. 1–14, January 1996.

55. Jahangiri, E., E. Yörük, R. Vidal, L. Younes and D. Geman, "Information Pursuit: A Bayesian Framework for Sequential Scene Parsing", *arXiv preprint arXiv:1701.02343*, January 2017.

56. Branson, S., C. Wah, F. Schroff, B. Babenko, P. Welinder, P. Perona and S. Belongie, "Visual Recognition with Humans in the Loop", K. Daniilidis, P. Maragos and N. Paragios (Editors), *Proceedings of European Conference on Computer Vision*, Vol. 6314 of *Lecture Notes in Computer Science*, pp. 438–451, Springer, September 2010.

57. Konyushkova, K., R. Sznitman and P. Fua, "Introducing Geometry in Active Learning for Image Segmentation", *Proceedings of IEEE International Conference on Computer Vision*, pp. 2974–2982, August 2015.

58. B. Marvasti, N., E. Yörük and B. Acar, "Computer-Aided Medical Image Annotation: Preliminary Results with Liver Lesions in CT", *IEEE Journal of Biomedical and Health Informatics*, In press, November 2017.

59. Antal, P., G. Fannes, D. Timmerman, Y. Moreau and B. D. Moor, "Using Literature and Data to Learn Bayesian Networks as Clinical Models of Ovarian Tumors", *Artificial Intelligence in Medicine*, Vol. 30, No. 3, pp. 257–281, March 2004.

60. Weickert, J., *Anisotropic Diffusion in Image Processing*, Vol. 1, Teubner Stuttgart, Copenhagen, Denmark, 1998.

61. Cimen, S., *Liver Segmentation in 3D CT Data*, Master's Thesis, Bogazici University, 2011.

62. Boykov, Y. and G. Funka-Lea, "Graph Cuts and Efficient N-D Image Segmenta-

tion", *International Journal of Computer Vision*, Vol. 70, pp. 109–131, November 2006.

63. Wimmer, A., G. Soza and J. Hornegger, "A Generic Probabilistic Active Shape Model for Organ Segmentation", *Proceedings of Conference on Medical Image Computing and Computer Assisted Interventions*, pp. 26–33, March 2009.

64. Furukawa, D., A. Shimizu and H. Kobatake, "Automatic Liver Segmentation Based on Maximum a Posterior Probability Estimation and Level Set Method", *Proceedings of MICCAI Workshop on 3D Segmentation in the Clinic: A Grand Challenge*, pp. 117–124, March 2007.

65. Gorelick, L., M. Galun, E. Sharon and R. Basri, "Shape Representation and Classification Using the Poisson Equation", *Proceedings of IEEE Computer Vision and Pattern Recognition*, pp. 61–67, December 2004.

66. Heimann, T. *et al.*, "Comparison and Evaluation of Methods for Liver Segmentation from CT Datasets", *IEEE Transactions on Medical Imaging*, Vol. 28, No. 8, pp. 1251–1265, February 2009.

67. Frangi, R. F., W. J. Niessen, K. L. Vincken and M. A. Viergever, "Multiscale Vessel Enhancement Filtering", *Proceeding of International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 130–137, Springer-Verlag, October 1998.

68. Marvasti, N. B. and B. Acar, "Assessment of MRF Based Joint Scale Selection and Segmentation For 3D Liver Vessel Segmentation Task", *Proceedings of Conference on Signal Processing and Communications Applications*, pp. 1–4, April 2013.

69. IRCAD France, "IRCAD Training Center", `http://www.ircad.fr/softwares/3Dircadb/3Dircadb1/index.php?lng=en`, accessed at August 2017.

70. Kundu, S., M. Itkin, D. Gervais, V. Krishnamurthy, M. Wallace, J. Cardella,

D. Rubin and C. Langlotz, "The IR Radlex Project: An Interventional Radiology Lexicon-A Collaborative Project of the Radiological Society of North America and the Society of Interventional Radiology", *Journal of Vascular and Interventional Radiology*, Vol. 20, pp. 433–435, April 2009.

71. World Health Organization, "ICD10 Codes", `http://www.who.int/classifications/icd/en/`, accessed at August 2017.

72. WHO Collaborating Centre for Drug Statistics Methodology, "ATC Codes", `http://www.whocc.no/atc_ddd_index/`, accessed at August 2017.

73. Turkish Ministry of Health , "USBS Codes", `https://skrs3.sagliknet.saglik.gov.tr/`, accessed at August 2017.

74. Vapnik, V. N., *Statistical Learning Theory*, Vol. 1, Wiley, New York, USA, 1998.

75. Hsu, C.-W. and C.-J. Lin, "A Comparison of Methods For Multiclass Support Vector Machines", *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 13, No. 2, pp. 415–425, March 2002.

76. Alpaydin, E., *Introduction to Machine Learning*, The MIT Press, 2014.

77. Campbell, C., "Kernel Methods: A Survey of Current Techniques", *Neurocomputing*, Vol. 48, No. 1–4, pp. 63 – 84, October 2002.

78. Wu, T.-F., C.-J. Lin and R. C. Weng, "Probability Estimates for Multi-Class Classification by Pairwise Coupling", *Journal of Machine Learning Research*, Vol. 5, pp. 975–1005, August 2004.

79. Koller, D. and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*, MIT press, 2009.

80. McLachlan, G. and T. Krishnan, *The EM Algorithm and Extensions*, Wiley, Hoboken, NJ, USA, 2nd edn., 2008.

81. Barber, D., *Bayesian Reasoning and Machine Learning*, Cambridge University Press, 2012.

82. Friedman, N., "Learning Belief Networks in The Presence of Missing Values and Hidden Variables", *Proceedings of International Conference on Machine Learning*, Vol. 97, pp. 125–133, July 1997.

83. Elidan, G. and N. Friedman, "Learning Hidden Variable Networks: The Information Bottleneck Approach", *Journal of Machine Learning Research*, Vol. 6, pp. 81–127, December 2005.

84. Schwarz, G., "Estimating the Dimension of A Model", *Annals of Statistics*, Vol. 6, No. 2, pp. 461–464, March 1978.

85. Pearl, J., "Fusion, Propagation, and Structuring in Belief Networks", *Artificial Intelligence*, Vol. 29, No. 3, pp. 241–288, September 1986.

86. Spiegelhalter, D., A. Thomas, N. Best and W. Gilks, "BUGS 0.5: Bayesian Inference Using Gibbs Sampling Manual (version ii)", *MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK*, pp. 1–59, 1996.

87. Zhang, N. L. and D. Poole, "Exploiting Causal Independence in Bayesian Network Inference", *Journal of Artificial Intelligence Research*, Vol. 5, pp. 301–328, December 1996.

88. Minka, T. P., "Expectation Propagation for Approximate Bayesian Inference", *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*, pp. 362–369, Morgan Kaufmann Publishers Inc., August 2001.

89. Murphy, K. P., *Machine Learning: A Probabilistic Perspective*, The MIT Press, 2012.

90. Chow, C. and C. Liu, "Approximating Discrete Probability Distributions with

Dependence Trees", *IEEE Transactions on Information Theory*, Vol. 14, No. 3, pp. 462–467, May 1968.

91. Chen, Y.-W. and C.-J. Lin, *Combining SVMs with Various Feature Selection Strategies*, Springer, Berlin, Heidelberg, 2006.

92. Järvelin, K. and J. Kekäläinen, "Cumulated Gain-Based Evaluation of IR Techniques", *ACM Transactions on Information Systems*, Vol. 20, No. 4, pp. 422–446, October 2002.

93. Strela, V., P. N. Heller, G. Strang, P. Topiwala and C. Heil, "The Application of Multiwavelet Filterbanks to Image Processing", *IEEE Transactions on Image Processing*, Vol. 8, No. 4, pp. 548–563, April 1999.

94. Schmitt, R. and P. Fritz, "OP5-A 3D-Fourier-Descriptor Approach to Compress and Classify 3D Imaging Data", *Proceedings OPTO 2009 & IRS^2 2009*, pp. 133–138, May 2009.

95. Xu, J., S. Napel, H. Greenspan, C. F. Beaulieu, N. Agrawal and D. Rubin, "Quantifying the Margin Sharpness of Lesions on Radiological Images for Content-based Image Retrieval", *Medical Physics*, Vol. 39, p. 5405, September 2012.

96. Tesar, L., D. Smutek, A. Shimizu and H. Kobatake, "3D Extension of Haralick Texture Features for Medical Image Analysis", *Proceedings of International Conference on Signal Processing, Pattern Recognition, and Applications*, pp. 350–355, February 2007.

97. Majtner, T. and D. Svoboda, "Extension of tamura texture features for 3d fluorescence microscopy", *Proceedings of International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission*, pp. 301–307, IEEE, October 2012.

98. Gabor, D., "Theory of Communication Part 1: The analysis of information",

Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering, Vol. 93, No. 26, pp. 429–441, November 1946.

99. Heeger, D. J., "Model for the Extraction of Image Flow", *Journal of the Optical Society of America A*, Vol. 4, No. 8, pp. 1455–1471, August 1987.

100. Lo, C.-H. and H.-S. Don, "3D Moment Forms: Their Construction and Application to Object Identification and Positioning", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 11, No. 10, pp. 1053–1064, October 1989.

101. Haralick, R. M., K. Shanmugam *et al.*, "Textural Features for Image Classification", *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 3, No. 6, pp. 610–621, November 1973.

102. Tamura, H., S. Mori and T. Yamawaki, "Textural Features Corresponding to Visual Perception", *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 8, No. 6, pp. 460–473, June 1978.

103. Hu, M.-K., "Visual Pattern Recognition by Moment Invariants", *IRE Transactions on Information Theory*, Vol. 8, No. 2, pp. 179–187, August 1962.

APPENDIX A: IMAGE FEATURES

Low level Image descriptors/features are categorized into two groups, *i* the global image descriptors and *ii* the individual image descriptors. The global image descriptors cover the basic and liver-wide global statistical properties which are deduced directly from the CT volumes and the associated segmentation masks. The individual descriptors are generated for each single lesion in the liver separately. They are grouped in three types as geometric, location and texture features. Following the same notation in this dissertations where $\mathbf{r}$ is a position vector, S_L^0 is the liver parenchyma mask, S_L is the liver mask, S_V is the vessel mask, and S_E is the lesion mask the image features can be computed as follow.

A.1. Global image descriptors

- Liver related Features
  - (i) Liver volume is defined as $\sum_r(S_L(\mathbf{r})) \times r_x \times r_y \times r_z$, where $r_x \times r_y \times r_z$ is the voxel size in mm^3.
  - (ii) Liver's mean (μ_L^0) and variance (σ_{L0}^2) are computed using voxels inside the liver parenchyma as $\mu_L^0 = \frac{\sum_{\mathbf{r}} I(S_L^0(\mathbf{r})==1)}{\#(S_L^0(\mathbf{r})==1)}$ and $\sigma_{L0}^2 = \frac{\sum_{\mathbf{r}}(I(S_L^0(\mathbf{r})==1)-\mu_L^0)^2}{\#(S_L^0(\mathbf{r})==1)}$.
- Vessels related Features
  - (i) Vessel volume is defined as $\sum_r(S_V(\mathbf{r})) \times r_x \times r_y \times r_z$.
  - (ii) Ratio in liver is the percentage of vessel voxels inside the liver mask which is defined as $\frac{\sum_r(S_V(\mathbf{r})==1)}{\sum_r(S_L(\mathbf{r})==1)}$.
- Lesion Global Features

  Let us define the lesions intensity image at lesion voxels as $I_E(\mathbf{r}) = I(S_E(\mathbf{r}) == 1))$ and K as the number of bins for the histogram of $I_E(\mathbf{r})$ then X_i and P_i are the value and probability of the i_{th} histogram bin, respectively. The Lesion global features are defined as follows
  - (i) Ratio in liver is the percentage of lesion voxels inside the liver which is defined as the $\frac{\sum_r(S_E(\mathbf{r})==1)}{\sum_r(S_L(\mathbf{r})==1)}$.

(ii) Smallest and largest Lesions' volume are the volumes of the smallest and the largest component in the lesion mask, respectively.

(iii) Lesions' mean(μ_E) and variance(σ_E^2) are computed using voxels inside the lesion mask and defined as $\mu_E = \sum_i X_i P_i$ and $\sigma_E^2 = \sum_i P_i (X_i - \mu_E)^2$.

(iv) Number of lesions is the number of individual objects inside the lesion mask.

(v) Smoothness is computed as $S = 1 - \frac{1}{1+\sigma_E^2}$.

(vi) Skewness(asymmetry) is computed as $A = \frac{\sum_i P_i (X_i - \mu_E)^3}{\sigma_E^3}$.

(vii) Kurtosis is a measure of whether an image's intensity distribution is peaked or flat relative to the normal distribution, which is computed as $K = \frac{\sum_i P_i (X_i - \mu_E)^4}{\sigma_E^4}$.

(viii) Energy (uniformity) is $U = \sum_i P_i^2$.

(ix) Entropy is defined as $E = -\sum_i P_i * log(P_i)$.

(x) Threshold is the portion of voxels inside the $I_E(\mathbf{r})$ that have higher intensity than liver mean μ_L, to find portion of hyperdense lesion voxels.

(xi) Bin of histogram peak is the bin index in which histogram of $I_E(\mathbf{r})$ is maximum.

(xii) Haar-wavelet coefficients are the low-frequency coefficients of the histogram's three-level Haar-wavelet transform [93]

A.2. Individual Lesion Features

These features are calculated for each single object (lesion) in the lesion mask which are grouped into three categories: Geometric features, Location features, and the Texture features. Considering each lesion mask as $S_{E_k}(\mathbf{r})$, definitions are as follows

- Geometric Descriptors

  The geometric/shape descriptors are calculated directly from the lesion mask of each individual lesion.

  (i) Volume is defined as $V_{E_k} = \sum_r (S_{E_k}(\mathbf{r}) == 1) \times r_x \times r_y \times r_y$,

  (ii) Area is defined as the number of voxels on the surface of the lesion, where the surface voxels can be computed by morphological dilation as $Surface = S_{E_k}(\mathbf{r}) - Dilate(S_{E_k}(\mathbf{r}) == 1)$. Area is $A_{E_k} = \sum_r (Surface(\mathbf{r})) \times r_x \times r_y \times r_y$.

(iii) Sphericity is defined as the ratio of the lesion volume to a the ratio of its volume as a sphere (V_0) which is computed as $Sphericity = \frac{V_{E_k}}{V_0}$ where $V_0 = \frac{4\pi d_{max}^3}{3}$ and d_{max} is the maximum distance of the lesions surface to its center.

(iv) Convexity is computed as $Convexity = \frac{A_{convex_k}}{A_{E_k}}$ where A_{convex_k} is the area of the convex hull of the lesion.

(v) Compactness (shape factor) is a numerical quantity representing the compactness degree of a shape and defined as $Compactness = \frac{\sqrt{A_{E_k}^3}}{V_{E_k}}$.

(vi) Solidity measures the similarity of the lesions's area with it's convex area and is computed as $Solidity = \frac{A_{E_k}}{A_{convex_k}}$.

(vii) Maximum extentis the maximum diameter of the lesion d_{max} which is the maximum distance of the lesions surface to its center.

(viii) Aspect-ratio is the ratio of the largest diameter to the smallest diameter orthogonal to it $Aspect\_ratio = \frac{d_{min}}{d_{max}}$.

(ix) Fourier Descriptors of a shape are defined as the 2D-Fourier transform of the sampled-shape vector [94]. This vector is constructed by sampling the shape ($Surface$) of the lesion in the polar space and saving the distance to lesion center d_{mn}, where the degree intervals are 30 resulting in $m = 7$ elevation (vertical degree) and $n = 13$ azimuth (horizontal) rotation points. Note that, the absolute value of the Fourier descriptor is used to avoid the change in the value due to the rotations. To make these values scale free, all values are normalized by the first Fourier value. Hence, regardless of the lesion location in the liver, the Fourier descriptors remain the same.

- Location Descriptors

  The location descriptors refer to the location of the lesion with respect to the liver and vessels, and computed from the segmentation masks.

  (i) Proximity to Vessels is defined as the shortest distance between the lesion's surface and the nearest vessel to it as mm^3.

  (ii) Touch area ratio is the portion of the lesion's surface mask that touches a vessel.

(iii) Anatomical location is used to locate every lesion inside the liver. In order to perform this task, five critical points in the liver are assigned by an expert (3 points on portal vein, and 2 point on the hepatic vein) and for every lesion, the distance of the lesion' center to these points are computed and saved in a vector.

(iv) Vessel to lesion ratio measures the amount of vessel around the lesion. To this end, the convex hull of the lesion dilated by 3 voxels and the ratio of vessels to the lesions in the this volume is computed.

(v) Vessel to total ratio is the ratio of vessel to all voxels in this volume.

- Texture Descriptors

  The texture descriptors describe the gray-scale features, the boundary related features and the features encoding the texture of the image.

  (i) Boundary Descriptors The boundary descriptors capture the contrast and sharpness across the lesion boundary as proposed by Schmitt *et al.* [95]. One dimensional gray-scale profiles are extracted along the normal direction of the lesion surface for ± 2 voxels . The contrast is the signed difference between the two ends of a profile line. The sharpness refers to the rate of intensity change across the profile. These two parameters are calculated via fitting a sigmoid function to the profile lines and computed along the lesions boundary. Two $30 - bin$ histograms representing these two parameters over the whole lesion surface are generated from all profile lines.

  (ii) Gray-scale features which are derived directly from raw pixel intensities of individual lesions include Mean, variance, skewness, kurtosis, energy, smoothness, bin index of histogram peak, entropy, threshold, and haar wavelet coeff., which are described in details in global feature section.

  (iii) Texture Descriptors encode spatial organization of voxel values in an image regions. Four texture descriptors contain Haralick's features [96], Tamura features [97], Gabor filter descriptors [98, 99], and 3-D Hu moments [100] have been employed.

  Six Haralick's features are derived from co-occurrence matrices (C) in 6 distance ($1, 2, 4, 6$ and 8 in 13) and 13 (45 degree gap) directions [96, 101].

For each distance, the final feature value is computed by averaging over all angular directions, which results in total of 30 texture characteristic, 6 characteristics for each voxel distance. The definition of these features are

$$\text{Uniformity of Energy} = \sum_{ij} c_{ij}^2 \tag{A.1}$$

$$\text{Entropy} = -\sum_{ij} c_{ij} ln(c_{ij}) \tag{A.2}$$

$$\text{local homogeneity} = \sum_{ij} \frac{c_{ij}}{1 + (i-j)^2} \tag{A.3}$$

$$\text{contrast} = \sum_{ij} c_{ij} |i-j| \tag{A.4}$$

$$\text{Cluster Prominence} = \sum_{ij} [(i-\mu) + (j-\mu)]^4 c_{ij} \tag{A.5}$$

$$\text{Cluster Shade} = \sum_{ij} [(i-\mu) + (j-\mu)]^3 c_{ij} \tag{A.6}$$

Tamura features were first introduced by Tamura [102] and later were extended to 3D features by Majtner *et al.* [97]. They correspond to human visual perception, namely 'coarseness', 'contrast', 'directionality', 'line-likeness', 'regularity', and 'roughness'. In this thesis, the first three features have been considered as they better discriminate the human visual perception.

Coarseness is the most fundamental texture feature, which finds the largest size at which a texture exists.

$$A_L(x,y,z) = \frac{1}{2^{3L}} \sum_{i=x-m}^{x+m-1} (\sum_{j=y-m}^{y+m-1} (\sum_{k=z-m}^{z+m-1} f(i,j,k))) \tag{A.7}$$

where $m = 2^{L-1}$ and L is the size of the neighborhood, which is set to $0 \leq L \leq 4$. Then, the differences in 3 direction are calculated

$$\begin{aligned} E_{L,x} &= |A_L(x+2^{L-1},y,z) - A_L(x-2^{L-1},y,z)| \\ E_{L,y} &= |A_L(x,y+2^{L-1},z) - A_L(x,y-2^{L-1},z)| \\ E_{L,z} &= |A_L(x,y,z+2^{L-1}) - A_L(x,y,z-2^{L-1})| \end{aligned} \tag{A.8}$$

and finally at each point, the size leading to the highest difference value is selected

$$S(x,y,z) = \text{argmax}_{L=0,\ldots,4;d=(x,y,z)}(E_{L,d}(x,y,z)). \tag{A.9}$$

The resulting histogram of the coarseness image (S) is saved as the Tamura coarseness feature with 16 bins.

Contrast of the image is influenced by dynamic range of gray-levels in the image, polarization of the distribution of the black and white, sharpness of edges, and period of repeating patterns standing for the picture quality.

The image contrast is calculated as $F_{con} = \frac{\sigma}{\sqrt[4]{\alpha_4}}$, where $\alpha_4 = \frac{\mu_4}{\sigma^4}$ and $\mu_4 = \frac{1}{N_0 N_1 N_2} \sum_{i=1}^{N_0} \sum_{j=1}^{N_1} \sum_{k=1}^{N_2} (X(i,j,k) - \mu_{E_k})^4$ is the fourth moment around the mean. F_{con} is measured in a window neighborhood of every voxel. The choice of window size is critical, and here, the window size is set to the size of the image divided by 10. However, if resulting value is less than 5, the window size is set to 3. The output histogram of the contrast image is saved as the Tamura contrast feature with 16 bins.

Directionality calculates the direction of the texture along three axes. Three different kernels to compute derivation in three direction along three axis are

generated. Then, the derivatives in x-axis($\triangle x$), y-axis($\triangle y$) and z-axis($\triangle z$) are computed through the convolution with these three kernels. Finally, directionalities θ_1, θ_2 and θ_3 at every voxel are computed as:

$$\begin{aligned} \theta_1 &= tan^{-1}\frac{\triangle y}{\triangle x} + \frac{\pi}{2} \\ \theta_2 &= tan^{-1}\frac{\triangle z}{\triangle y} + \frac{\pi}{2} \\ \theta_3 &= tan^{-1}\frac{\triangle x}{\triangle z} + \frac{\pi}{2} \end{aligned} \tag{A.10}$$

The resulting histograms of these three directionality images are saved as the Tamura directionality features (histogram range is between 0 to 180 with 16 bins).

The Gabor filter, originally introduced by Gabor [98] and later extend to 3D space [99], is defined as the multiplication of sinusoid waves with a Gaussian window. Gabor features are computed by filtering the image with a bank of orientation and scale sensitive Gabor filters and computing their mean and standard deviation of the output in the frequency domain. In order to extract features, it is required to create a database of Gabor kernels in different scales and different orientations. Then, by convolving the image with each one of these kernels, a Gabor image is created. Finally, a histogram is constructed from these Gabor images. In this dissertation, Gabor filtering has been applied on the original CT images in 4 scales and 13 direction. The size of the kernel is chosen as a quarter of original CT image and each kernel is convolved with the original CT image (not just the lesion). The lesion part is extracted from the result to create the corresponding histogram and 52 features are generated.

Invariant Hu moments, first introduced in [103], are the set of moments, which are invariant under translation, scale, and rotation. Based on the extension of these features to 3D space proposed by Lo *et al.* [100], the 3D

invariant moment of order $p+q+r$ is defined as

$$M_{pqr} = \sum_x \sum_y \sum_z x^p y^q z^r I(x,y,z) \tag{A.11}$$

where (x,y,z) is an arbitrary pixel on the surface and $I(x,y,z)$ is the image. The central moment of order $p+q+r$ is defined as

$$\mu_{pqr} = \sum_x \sum_y \sum_z (x-x_0)^p (y-y_0)^q (z-z_0)^r I(x,y,z) \tag{A.12}$$

where $x_0 = \frac{M_{100}}{M_{000}}$, $y_0 = \frac{M_{010}}{M_{000}}$, and $z_0 = \frac{M_{001}}{M_{000}}$. Then, the scaling normalized moment η_{pqr} is computed as

$$\eta_{pqr} = \frac{\mu_{pqr}}{M_{000}^{\frac{p+q+r+3}{3}}}. \tag{A.13}$$

In order to reduce the computation complexity, the 2-order 3D invariants are defined as

$$\begin{aligned} J_1 &= \eta_{200} + \eta_{020} + \eta_{002} \\ J_2 &= \eta_{200}\eta_{020} + \eta_{200}\eta_{002} + \eta_{020}\eta_{002} - \eta_{101}^2 - \eta_{110}^2 - \eta_{011}^2 \\ J_3 &= \eta_{200}\eta_{020}\eta_{002} - \eta_{002}\eta_{110}^2 - \eta_{020}\eta_{101}^2 - \eta_{200}\eta_{011}^2 + 2\eta_{110}\eta_{101}\eta_{011} \end{aligned} \tag{A.14}$$

APPENDIX B: ONTOLOGY OF LIVER FOR RADIOLOGY (ONLIRA)

ONLIRA is an ontology for liver proposed by Kokciyan *et al.* [5], which includes

- Essential concepts: liver, lesion, and hepatic vascularity.
- Individual properties of these concepts, such as the size or density of a liver.
- Relationships between the concepts. For instance, between a liver concept and a lobe concept, one can specify a has_Lobe relation to show that a liver contains lobes.

B.1. Concepts

There are 3 concepts and 40 sub_concepts defined in ONLIRA as

(i) Liver: Allows to describe various properties of the liver in ONLIRA.

(ii) Lesion: There may be abnormal areas of the liver that a radiologist wishes to identify, which can be expressed through lesion concept. ONLIRA defines Lesion as a subclass of an Area.

(iii) Hepatic vascular: allows to describe non-ordinary characteristics of the hepatic vascular system and it contains eight vasculature as "Hepatic artery", "Hepatic portal vein", "Left portal vein", "Right portal vein", "Hepatic vein", "Left hepatic vein", "Middle hepatic vein", "Right hepatic vein".

B.2. Relations

There are 12 relations (object properties) defined in ONLIRA, which shows the relations between the concepts. These relations and the questions generated from them are as follows

(i) isLocatedInLobe

Description: The lobe in which the lesion is located?

Question: Lobe?

Domain: "Area"

Range: Lobe

(ii) isLocatedInSegment

Description: The Segment in which the lesion is located?

Question: Segment?

Domain: "Area"

Range: Segment

(iii) hasLesionComposition

Description: The composition of the liver?

Question: Lesion composition?

Domain: "Lesion"

Range: Lesion composition (Solid, SolidCycsticMix, SolidWithCystic, PureSolid, PredominantSolid, PureCystic, PredominantCystic, CysticWithSolidComponent, CysticWithDebris, and Abcess)

(iv) isCloseToVein

Description: To which vessels the lesion is close?

Question: is Close to Vein?

Domain: "Lesion"

Range: Hepatic Vascularity ("Hepatic artery", "Hepatic portal vein", "Left portal vein", "Right portal vein", "Hepatic vein", "Left hepatic vein", "Middle hepatic vein", "Right hepatic vein", "Posterior branch of right portal vein", "Anterior branch of right portal vein")

B.3. Properties

There are 32 properties defined in ONLIRA, which show the data properties assigned to each concept listed as follows. Note that the questions in CaReRaWEB are also included.

(i) hasLiverPlacement

Question: Liver placement in the body?

Description: The abnormality of liver location in the body.

Domain: "Liver"

Range: "Downward_displacement", "Leftward_displacement", "Normal_placement", "Upward_displacement", "Other"

(ii) hasLiverContour

Question: Liver surface?

Description: The property of liver surface.

Domain: "Liver"

Range:"Irregular" , "Lobulated" , "Nodular", "Regular" , "Other"

(iii) hasSizeChange

Question: Liver size change?/ Right lobe size change?/ Left lobe size change?/ Caudate lobe size change?

Description: To report any change in the size of the liver or any of its lobe. This property is used to generate four questions in the form related to each lobe of the liver and liver itself.

Domain: "Liver or Lobe"

Range: "Decreased" , "Increased" , "Normal" , "Other"

(iv) Craniocaudal Dimension(mm)

Description: the craniocaudal size of the liver or any of its lobe in mm. This property is used to generate four questions in the form related to each lobe of the liver and liver itself.

Question: Liver craniocaudal dimension(mm)/ Right lobe craniocaudal dimension(mm)/ Left lobe craniocaudal dimension(mm)/ Caudate lobe craniocaudal dimension(mm)

Domain: "Liver or Lobe"

Range: "sizeMM"

(v) hasDensity

Description: Density type of lesion or liver. This property is used to generate two questions in the form, related to lesion and the liver.

Question: Liver density type?/ Lesion density type?

Domain: "Area or Liver or Parenchyma"

Range: "Heterogeneous" , "Homogeneous" , "Other"

(vi) hasLiverDensityChange

Question: Liver density change?

Description: The change in the density of liver parenchyma.

Domain: "Liver"

Range: "Decreased" , "Increased" , "Normal" , "Other"

(vii) hasLumenDiameter

Question: Hepatic artery lumen diameter?, Hepatic portal vein lumen diameter?, Left portal vein lumen diameter?, Right portal vein lumen diameter?, Hepatic vein lumen diameter?, Left hepatic vein lumen diameter?, Middle hepatic vein lumen diameter?, Right hepatic vein lumen diameter?

Description: The change in the lumen diameter of each of the Hepatic vascular. This property generate 8 semantic features. Domain: "Hepatic Vascular"

Range: "Decreased", "Increased", "Normal", "Other"

(viii) hasLumenType

Question: Hepatic artery lumen type?, Hepatic portal vein lumen type?, Left portal vein lumen type?, Right portal vein lumen type?, Hepatic vein lumen type?, Left hepatic vein lumen type?, Middle hepatic vein lumen type?, Right hepatic vein lumen type?

Description: The lumen type of each of the Hepatic vasculature. This property generates eight questions.

Domain: "Hepatic Vascularity"

Range: "Obliterated" , "Open" , "Partially_obliterated" , "other"

(ix) isCavernousTransformationObserved

Description: Is any branch of Hepatic portal vein transformed cavernously? This generate three semantic features. Questions: Is cavernous transformation observed? (Hepatic portal vein)/ Is cavernous transformation observed? (Left portal vein)/ Is cavernous transformation observed? (Right portal vein)

Domain: "Hepatic portal vein"

Range: boolean

(x) hasLesionQuantity

Description: Number of the lesions from current lesion_type in the liver.

Question: Lesion Cluster Size?

Domain: "Lesion"

Range: 1, 2, 3, 4, 5, "multiple"

(xi) hasAreaLengthFirst

Description: Width of the lesion in 2D view which is the slice in which lesion has the largest of its size.

Question: Area width?

Domain: "Area"

Range: sizeMM

(xii) hasAreaLengthSecond

Description: Height of the lesion in 2D view which is the slice in which lesion has the largest of its size.

Question: Area height?

Domain: "Area"

Range: sizeMM

(xiii) isGallbladderAdjacent

Description: Is lesion near to Gallbladder?

Question: Is gallbladder adjacent?

Domain: "Area"

Range: boolean

(xiv) isPeriphericalLocalized

Description: Is lesion located in the periphery of liver?

Question: Is peripheral localized?

Domain: "Area"

Range: boolean

(xv) isSubcapsularLocalized

Description: Is the lesion located in the sub_capsule of the liver?

Question: Is lesion subcapsular localized?

Domain: "Area"

Range: boolean

(xvi) isCentralLocalized

Description: Is lesion located in the center of the liver?

Question: is Central Localized?

Domain: "Area"

Range: Boolean

(xvii) hasAreaMarginType

Description: The type of the lesion's margin?

Question: Margin Type?

Domain: "Area"

Range: "Geographical", "Ill_defined", "Irregular", "Lobular", "Serpiginous", "Speculative", "Well_defined", "Other"

(xviii) hasAreaShape

Description: Shape of the lesion?

Question: Shape?

Domain: "Area"

Range: "Band" , "Fusiform" , "Irregular" , "Linear" , "Nodular" , "Ovoid" , "Round" , "Serpiginous", "Other"

(xix) isContrasted

Description: Is wall/lesion contrasted? This property generates two questions in the form.

Question: is area contrasted?/ is wall contrasted?

Domain: "Area or wall"

Range: Boolean , NA

(xx) hasLesionContrastUptake

Description: Contrast level of the lesion after injecting the contrast media.

Question: Lesion contrast uptake?

Domain: "Lesion"

Range: "Dense", "Heterogeneous", "Homogeneous", "Minimal", "Moderate", "Other"

(xxi) hasLesionContrastPattern

Description: Contrast pattern of the lesion after injecting the contrast media.

Question: Lesion contrast pattern?

Domain: "Lesion"

Range: "Central", "Early uptake then wash out", "Fixing contrast in late phase", "Heterogeneous", "Homogeneous", "Peripheral", "Peripheral nodular", "Spokes wheel", "Undecided", "NA", "Other"

(xxii) isCalcified

Description: Is lesion or any calcifiable component calcified? This property assigns to seven questions.

Question: is Calcified? (Area)/ is Calcified? (Capsule)/ is Calcified? (Polyp)/ is Calcified? (Pseudocapsule)/ is Calcified? (Septa)/ is Calcified? (Solid Component)/ is Calcified? (Wall)

Domain: "Area or Calcifiable component"

Range: Boolean

(xxiii) hasCalcification

Description: The calcification type of lesion or any calcifiable component. This property assigns to 8 semantic features.

Question: Calcification type? (Area)/ Calcification type? (Capsule)/ Calcification type? (Polyp)/ Calcification type? (Pseudocapsule)/ Calcification type? (Septa)/ Calcification type? (Solid Component)/ Calcification type? (Wall)

Domain: "Area or Calcifiable component"

Range: "Coarse", "Focal", "Millimetric-fine", "Punctate", "Scattered", "Other", "NA"

(xxiv) hasAreaDensity

Description: The lesion density compare to liver parenchyma.

Question: Density?

Domain: "Area"

Range: "Hyperdense", "Hypodense", "Isodense", "Other"

(xxv) hasSeptaDiameter

Description: Septum's diameter type

Question: Septa Diameter Type?

Domain: "Septa"

Range: "Complete", "Incomplete", "Other"

(xxvi) hasSeptaWidth

Description: The thickness of the septum diameter.

Question: Septa Thickness?

Domain: "Septa"

Range: "Thick", "Thin", "Other"

(xxvii) isLevelingObserved

Description: Is texture leveling observed inside the lesion?

Question: is Leveling observed?

Domain: "Lesion"

Range: Boolean

(xxviii) hasLesionLeveling

Description: Levels of the lesion's texture

Question: Lesion Leveling Type?

Domain: "Lesion"

Range: "Fluid_fluid", "Fluid_gas", "Fluid_solid", "Gas_solid", "Other"

(xxix) isDebrisObserved

Description: Is debris observed in the lesion?

Question: is Debris observed?

Domain: "Lesion"

Range: Boolean, NA

(xxx) hasDebrisLocation

Description: Where is the debris located?

Question: Debris location?

Domain: "Lesion"

Range: "Floating_inside", "Located_on_dependent_position", "Other", NA

(xxxi) hasWallType

Description: Type of the wall of the lesion.

Question: Wall type?

Domain: "Wall"

Range: "Thick", "Thin", "Other"

(xxxii) hasLesionVeinProximity

Description: The proximity of the vessels to the current lesion.

Question: Vascular proximity?

Domain: "Lesion"

Range: "Adjacent", "Adjunct_to_contact", "Bended", "Circumscribed", "Invaded", "Other", "NA"