Image Area Reduction for Efficient Medical Image Retrieval

by

Zehra Camlica

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Systems Design Engineering

Waterloo, Ontario, Canada, 2015

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Content-based image retrieval (CBIR) has been one of the most active areas in medical image analysis in the last two decades because of the steadily increase in the number of digital images used. Efficient diagnosis and treatment planning can be supported by developing retrieval systems to provide high-quality healthcare. Extensive research has attempted to improve the image retrieval efficiency. The critical factors when searching in large databases are time and storage requirements. In general, although many methods have been suggested to increase accuracy, fast retrieval has been rather sporadically investigated. In this thesis, two different approaches are proposed to reduce both time and space requirements for medical image retrieval. The IRMA data set is used to validate the proposed methods. Both methods utilized Local Binary Pattern (LBP) histogram features which are extracted from 14,410 X-ray images of IRMA dataset. The first method is image folding that operates based on salient regions in an image. Saliency is determined by a context-aware saliency algorithm which includes folding the image. After the folding process, the reduced image area is used to extract multi-block and multi-scale LBP features and to classify these features by multi-class Support vector machine (SVM). The other method consists of classification and distance-based feature similarity. Images are firstly classified into general classes by utilizing LBP features. Subsequently, the retrieval is performed within the class to locate the most similar images. Between the retrieval and classification processes, LBP features are eliminated by employing the error histogram of a shallow (n/p/n) autoencoder to quantify the retrieval relevance of image blocks. If the region is relevant, the autoencoder gives large error for its decoding. Hence, via examining the autoencoder error of image blocks, irrelevant regions can be detected and eliminated. In order to calculate similarity within general classes, the distance between the LBP features of relevant regions is calculated. The results show that the retrieval time can be reduced,

and the storage requirements can be lowered without significant decrease in accuracy.

Acknowledgements

First of all, I would like to express my sincere gratitude to my advisors, Prof. Hamid R. Tizhoosh and Dr. Farzad Khalvati for supervising my studies and encouraging towards pursuing a graduate degree. Without their continuous support, patience and encouragement, the studies presented here could not possibly have been accomplished . I have learned a lot from them through our discussions.

I owe a debt of gratitude to my committee members Profs. Jonathan Kofman and John Zelek for reading my thesis. I appreciate the considerable amount of time and effort they have invested.

To my husband, Ahmet, any word could not express my feeling towards him. I am so fortunate to have such an invaluable life partner. Thank you for everything.

Thank you to Ayca Diriksoy, Ayse Erenay, Busra R. Eren , Derya Kara, Ezgi Yesiloz, Fatma Kiraz, Feyza G. Koroglu, Feyza Zeytinoglu, Gizem S.N. Dalgic, Humeyra Kiyak, Ilknur Umay, Nada Gohider, Nursefa Y. Zengin, Sinem Orbay and Zeynep Korkmaz for their endless friendship and support, and also thanks to my moral older sisters Betul Kara, Hulya Biyikli and Mahinur Fidan for inspirational advice that I will never ever forget throughout my life. Thank you all to my dearest friends.

I extend my deepest heartfelt thanks to my family Kadriye Aydin, Zohre Camlica, Ahmet Aydin, Yunus Camlica, Munteha Pac, Ayse Ozaydin, Osman M. Aydin, Ali S. Camlica and Abdullah Camlica for their faithful support and confidence in me. I would always be indebted to them.

Finally, I would like to thank my creator, God, for every time I had in my life with feeling as near me.

Funding for this research was provided by Turkish Government.

Dedication

To my husband.

Table of Contents

List of Tables

List of Figures

List of Acronyms

IRMA: Image Retrieval in Medical Application

LBP: Local Binary Pattern

SVM: Support Vector Machine

CBIR: Content Based Image Retrieval

WALRUS: WAveLet-based Retrieval of User-specified Scenes

PCA: Principle Component Analysis

LDA: Linear Discriminant Analysis

SPIRS: Spine Pathology and Image Retrieval Systems

RGB: Red Green Blue

CCV: Color Coherence Vector

FT: Fourier Transform

DCT: Discrete Cosine Transform

ASSERT: Automated Search Selection Engine with Retrieval Tools

PACS: Picture Archive and Communication System

TCIA: The Cancer Imaging Archive

NHANES: National Health and Nutrition Examination Surveys

ROI:Region Of Interest

EMD: Earth Movers Distance

MDL: Minimum Description Length

HMMS:Hidden Markov Models

Chapter 1

Introduction

With rapid growth of computer technologies, digitized information has gained tremendous interest for more than two decades. Since the steadily increasing amount of information has been made easily accessible through digital media, navigation and retrieving accurate and relevant information from big data has become one of the most important problems in information technology. Images are utilized in many application fields such as biomedicine, crime prevention, architecture, engineering, military, commerce, education and entertainment. Imaging provides important support in these areas. Especially in medicine, imaging constitutes a very fast and non-invasive method for diagnosis, treatment and monitoring of different disease. The digitization of images comes into prominence because of benefits in many areas. For instance, people can easily have a collection of hundreds or thousands of images stored on a personal computer. Most of the images may never be displayed again, but we frequently search for some of them in order to set them as a computer screen wallpaper, email them to a friend or print them. This kind of search may be performed manually, which is difficult in a larger collection of images. At this point, image retrieval

can be an effective choice to handle large collection of medical image data. In order to find and access relevant images in a medical database, efficient navigation and retrieval methods are indispensable. There are two kinds of medical image retrieval systems, namely text-based and content-based methods.

1.1 Text-based Image Retrieval

In text-based retrieval, images are labeled by taking advantage of keywords, classification codes or subject heading to search and retrieve images. Using these textual labels, retrieval can be easily and quickly applied. However, there is no understandably no general query pattern in text-based retrieval as different users query via different keywords for the same images. Another problem is that it requires sophisticated techniques and/or tremendous effort to annotate all images with descriptive metadata. As well, this method cannot prevent incorrect or missing results based on mistakenly labeled images. Finally, the spatial information in digital images can rarely be accurately and completely described in textual annotations. This may be impossible for many image categories.

1.2 Content-based Image Retrieval

In the Content-Based Image Retrieval (CBIR) technique, image contents are used to search and retrieve digital images. This technique is introduced to address the problems associated with text-based image retrieval. CBIR is generally a combination of techniques for retrieving relevant images based on automatically derived image features [1]. The main purpose of CBIR is to index and retrieve images efficiently to reduce the need for (automated and manual) annotations in the indexing process. Although CBIR systems differ in

details, the general concepts are rather similar. Extracting features of every image based on the image content and defining comparison rules for images are the main stages in CBIR. Image features represent the image used to measure similarity with other images in the database.

1.2.1 Feature extraction

Feature extraction is the first step in an image retrieval system. The main intention in extracting features is to create high-level descriptions from low-level data. Recent medical image retrieval systems are implemented with visual features, such as color, shape, texture, and other spatial characteristics. Visual features can be classified into low level, middle level and high level features. Almost all early systems were based on low level features that capture some image characteristics (color, shape). Recently, both mid-level (e.g. bagging approach) and high-level (semantics) image representations have emerged. General visual features are preferred in most CBIR systems, because they are independent of prior information and efficient in terms of computation [2]. The efficiency of a CBIR system relies on the quality of extracted features. If the features do not represent the image content effectively, similar images can hardly be retrieved. The features can be represented as symbolic and numeric data with vectors or graphs. These features can be extracted from the whole image (globally) or from regions of the image (locally).

Global features

Global features have had common usage as image descriptors since the beginning of CBIR. In [3], global features are extracted to obtain image semantics such as color histograms as color features, coarseness and direction as texture features and circularity and eccentricity

as shape features. These features are derived globally from the whole image. In the global representation of image, features cannot efficiently work for some semantically meaningful conditions such as scale and position. In other words, image retrieval using global features often gives ineffective results because of failure to capture semantic information within the image. For instance, global average of color in an image which has two cars with different color (yellow and red), is orange. This global representation does not describe exactly semantic meaning of the image. Another simple example of global features is a histogram. However, histograms do not contain information about location, shape and texture. Additionally, their results are sensitive to intensity, distortions and cropping [4]. There are some methods that combine two types of global features (texture and color) in order to retrieve images more precisely [5]. The retrieval for both features can be applied separately. However, the texture features are also extracted globally from the image and they cannot provide accurate description of the image in all cases. Although gist and scale invariant feature transform (sift) methods are commonly used global feature extraction methods in recent works [6], they need more complex function, computational time and feature storage areas comparing with low-level feature extraction methods.

Local features

Retrieval systems using local features have been developed to overcome limitations of global features. Local features represent images as collections of regions that consist of objects such as cars, eyes, flowers and humans. A key step in such an image retrieval system is a robust region detector algorithm [7]. The system takes an input image and separates regions by clustering pixels of this image that are similar. Each object in the image can be described by these local features [8]. The retrieval system should have the ability of identification of objects in the image in order to retrieve similar images in different location

or size. Some retrieval system examples which utilize local features are the Natra [9] and the Blobworld [10]. These two systems search based on regions of images. They require professional user guidance. For example, an user should provide segmented regions of the image and should select the types of features to compare. For this reason, the system result is highly influenced by the user. Another retrieval model is WALRUS [11], which is a robust model for scaling and translation of objects within an image. As a first step, an image is decomposed into local regions and then the similarity measure between two images can be defined by matching regions. The authors proposed a method that iteratively chooses the best pair of matching maximum area regions. R* three is cluster-based approach which is used to segment images into areas with the aim to increase efficiency of the system [12]. Moreover, in another image retrieval method [13], low-level features such as color, texture, and edge density are derived locally. A greedy region matching method is applied to reduce computational time. In this method, there is a threshold value in similarity detection. If the distance is more than threshold value, the region in the target image is automatically eliminated. While the aforementioned methods exemplify only comparison applied to a local region of an image and a local region of a target image, Integrated Region Matching (IRM) [14] compares similarity between a local area of an image and several local areas of target image. This measurement gives many relationships between images, and weighted summation of distances can define similarity. The Fuzzy Club [15] is another local feature-based retrieval method that combines color, texture and shape information. Firstly, theimage is divided by blocks and these local regions are reformed by unsupervised clustering. A color histogram is represented by fuzzy logic. Then color, texture and shape features are extracted. The second clustering is applied to features. Next, the distance and centroid of the regions' classes are described. Finally, the query only compares the candidate images that have regions in same classes. All these local-based methods are

strongly dependent on the segmentation results. In other words, if an error happens in segmentation or clustering part, the retrieval system can collapse.

1.2.2 Similarity measurement

After feature extraction, the similarity measurement begins. The similarity methods calculate the similarity of two images in terms of various features such as color, texture, shape etc. Then, search results are sorted with respect to their measured value of distance. The shorter distances correspond to higher similarity. There are many similarity measurement methods that have been used for image retrieval. Image features and descriptor selection and their representation methods directly affect the choice of similarity measurement method. Metrics on vector spaces (e.g., the Euclidean distance, the city-block-distance and the Minkowsky distance) are the most commonly implemented methods to calculate similarities due to their computational simplicity for linear searching.

Each visual feature set is usually stored in a high dimensional vector. In general, the dimension of feature vector space within similarity measurements is reduced for efficiency. For this purpose, principal component analysis (PCA) [16] or minimum description length (MDL) [17] can be applied. The concept of these reduction methods is to keep crucial information while reducing the size of the feature space. Some tree-based techniques, such as KD-trees [18] and R-trees [19], are also utilized for easy access to large feature vector space in CBIR systems. Furthermore, statistical methods can be trained to compare features. Some examples are neural networks [20], Bayesian networks [21] and Hidden Markov Models (HMMs) [22]. Different strategies have been developed to use various modalities in searching for CBIR, such as constrained hierarchies or classes, early fusion and fusion [23]. While the search process is restricted with some hierarchies or classes in

constrained methods, all images are searched in both early and late fusion methods. Using constrained method speeds up the retrieval process. Performing search within a local area (a certain class) or based on a hierarchical order takes less time than searching in the entire dataset. This method provides advantages especially for huge datasets [24]. However, if the class or hierarchy estimate fails, the system cannot work effectively.

1.3 Thesis Motivations

In order to discuss possible methodological developments and to benefit from them clinically, understanding the needs of CBIR in medical imaging is essential. The main purpose of a medical information system is to acquire information to provide efficient and high-quality healthcare. Retrieving similar images of the same modality from the same anatomic region can support the radiologist to diagnose the disease more confidently. In [25], experimental results prove clinical benefit of image retrieval systems. Visual features can be utilized with aim of improving diagnostics, teaching and research in medicine. In the medical educational domain, lecturers and students can use retrieval systems as an educational and training tool.

The most crucial problems in CBIR are related to metadata and semantic gap. Metadata are data types that consist of text information describing the visual information. Conceptually, if the content of an image is described by the text in metadata, the system could retrieve the images effectively using some text search algorithms. However, text-based systems are limited in practice. Moreover, the manual creation of meaningful metadata is a tedious task and impractical. On the other hand, automatic metadata systems are still unreliable, because images cannot be represented properly and descriptively. This representation problem of converting the low-level information into high-level infor-

mation of the image content is known as the semantic gap [26]. The semantic gap means 'the lack of matching between the retrieval information and real semantics' [27]. On the other hand, the computational load, when large image collections are managed, is most of the time not manageable or possible [23]. For these reasons, the design and development of efficient CBIR systems are still of interest. Speeding up the retrieval process with extending capability of database storage is clearly on demand in a medical setting.

1.4 Thesis Objectives and Contributions

The aim of this thesis is to propose a new CBIR system by reducing the overall retrieval time and storage requirements. The system first divides images into regions that correspond to blocks. Texture features are then extracted from each image block to which the proposed algorithms applied. The contributions of this work can be stated as follows:

1. Salient texture features are extracted from sub-block regions using local binary patterns. Instead of using a segmentation process, the salient region is detected by a context-aware algorithm and data is folded according to this salient region. In this manner, computational complexity and retrieval time are reduced.

2. Most systems compare the query image with every target image in the database to find the top matching images. The linear search is inefficient when the database is large. It is not necessary to search the entire database. In fact, it is possible to restrict a priori information regarding the general classes of images in the database in the feature space before a query is put forward. Only a part of the database needs to be searched, while a large portion of the database may be eliminated from the search. This certainly saves significant query processing time without compromising the retrieval precision.

3. To further increase the performance of the system, block (region) feature elimination

in retrieval is applied via using autoencoders. This thesis proposes a search algorithm that uses the same texture features from the divided blocks of the image to compute the distance between two images. This algorithm combines classification and the region-based searching algorithms that eliminate features using autoencoders. This way, the search area, the feature dimension, and hence the search time are reduced. Results illustrate that the system developed in this thesis significantly improves the overall retrieval efficiency compared to the existing systems.

1.5 Organization of the Thesis

The rest of the thesis is organized as follows. Chapter 2 summarizes some of the related work on CBIR and other primary research issues. In Chapter 3, an overview of the proposed CBIR systems, their principles, and techniques used for feature extraction, similarity measure, and indexing structures are described. The second implemented region-based retrieval systems with auto-encoder is also introduced in Chapter 3. Finally, Chapter 4 concludes with summary of the work and some suggestions for future work.

Chapter 2

Literature Review and Background

2.1 Literature Review

2.1.1 CBIR systems

CBIR systems can retrieve, manage and navigate huge visual data archives based on query images selected by the user or automatically. Recent medical image retrieval systems rely on visual features such as color, shape, texture etc. CBIR systems consist of two general processes, which are feature extraction and similarity measurement, in order to retrieve desired images from a huge visual database. IBM Almaden Research Center project [28] and Chabot project [29] are pioneer studies in CBIR area. Additionally, market-based CBIR systems have been developed such as VIRAGE [30] and NEC AMORE [31]. On the other hand, MIT Photobook [32], Columbia VisualSEEK and WebSEEK [33], UCSB NeTra [9] and Stanford WBIIS [34] are popular methods emerging from academia. Most of them have been using low-level of visual features (color or texture), while Blobworld

[10] and PicHunter [35] systems are utilizing higher level visual features (e.g., segmented part of image). Medical CBIR system samples are TELEMED [36], ASSERT [37], PACS [38] , IRMA [39], CervigramFinder [40] and SPIRS [41]. TELEMED is a visualization and segmentation tool in neurology, radiology and surgery. ASSERT (Automated Search Selection engine with Retrieval Tools) is a CBIR system for field of computed tomography of lung. PACS is Picture archive and Communication system which aims to integrate image modalities with other health information systems. IRMA (Image retrieval medical applications) is working on global medical image characteristics for retrieval process. The CervigramFinder works with cervicographic images and is created by National Cancer Institutes and National Library of Medicine. SPIRS (Spine Pathology and Image Retrieval Systems) is developed by National Library of Medicine in order to retrieve X-ray images from 17000 digitized spine radio graphs. Almost all recent CBIR systems have benefited from visual features of images.

Visual Features

Visual features ,which are fundamental for CBIR, can be classified into three levels. These are low-level features (primitive), middle-level features (logical) and high-level features (abstract). Early systems are based on low level features which represent characteristics of images (color or shape), but currently, middle and high-level image representations are grabbing attention. Middle-level features are obtained from particular parts of the image such as sub-image [42], important regions [43], and segmented and important details [44]. These features can be described with hierarchical [45] or graphical structure in order to improve retrieval results. Multiple learning and the bagging approach [46] represent images with visual words. Moreover, a review [47] describes the visual words which are related with dictionaries or the bag of visual words concept. High-level features are defined via semantic

design. Semantic design can be visual or textual. In the WordNet [48], link annotations are used for image concepts. Although color and texture features are defined as low-level features, they can be related to middle-level according to their extraction process. The methods utilizing color, texture and local features are investigated in following paragraphs because of their widespread usage in CBIR and also their relationship to the proposed algorithm.

Color is one of the most important low-level features of images. These features are measured based upon a particular color model or space. RGB, LUV, HSV and HMMD [49] are basic color models in the literature. Common color feature extracting methods are color histogram, color moments [50], color coherence vector [51] and color correlogram [52] etc. Moreover, dominant color descriptor, colour layout descriptor, colour structure descriptor and scalable colour descriptor are other color features which are standardized by MPEG-7 [53]. The simplest features among them are color moments that are used in many retrieval systems. The common color moments are mean, standard deviation and skewness. Since moments are very generalized features, they do not contain all the color information of an image. More specific color feature is color histogram which describes color distribution of an image [28]. In other words, color space is sampled into different bins and each color bin is defined based on frequency of pixels. Being independent from changes in resolution and rotations is a special advantage of the histogram method. Additionally, being simple to implement, efficient to compute and the low space requirement are some other advantages why histograms are common in CBIR [54]. Nevertheless, possible assignments for similar color intensities to different bins and the absence of any spatial information are the main disadvantages of using histograms [55]. To overcome these problems, partition-based histograms that contain spatial information by splitting the image into multiple partitions and calculating local histograms have also been developed [55]. Moreover, to solve the spatial

information problem in a histogram, the color coherence vectors (CCV) method has been proposed [55]. It investigates similar color regions in the image and counts the number of pixels in these regions. The method compares the number of pixels in the region with a threshold and classifies them as coherent or incoherent. For this reason, performance of CCV is usually better than a color histogram, but CCV is more complex. In this method, some spatial information is still missed. Determination of a threshold poses a potential problem as well. Another method is the color correlogram which is supposed to extract both color and spatial information from an image [52]. The method processes pixel position, intensity, probability of intensity and distance. A color correlogram is a gray level co-occurrence matrix (GLCM) with the color. A color correlogram is a combination of the 2D colors of any pixel pair and their spatial distance [52]. Therefore, complexity of the color correlogram is much more than complexity of the CCV and histogram because of high dimensionality and multiple matrix processing. Another histogram-based descriptor is the scalable color descriptor (SCD). The scalability issue distinguishes SCD from the regular histogram. Reducing the number of color bins by the Haar transform and removing some least significant bits from the quantized representations of the bin values are two methods for scalability. However, SCD has a similar problem as the conventional histogram which is missing spatial information. Color structure descriptor (CSD) also relies on a histogram [56]. The CSD is made while a structuring element is moving through the image. The window size is important for accuracy, and it is more complex than SCD. Dominant color descriptor (DCD) is also a type of histogram-based method [57]. A selected threshold applied to the bins and a dominant color bin is defined. This method is more accurate compared with the conventional histogram, but it loses efficiency in dissimilarity measurement. In summary, color moments are insufficient to represent the regions among all different kinds of color features. CCV, color correlogram and CSD perform well for image

representation, but they all have high computational complexity.

Texture is also one of the low-level image features that has common use in CBIR. While color features have been calculated by a pixel property, texture features are measured by using a group of pixels. In both image retrieval and semantic learning approaches, texture features are attractive to apply. In order to calculate texture features, various techniques have been proposed. These techniques can be grouped in two classes, spatial and spectral feature extraction. Pixel statistics and local pixel structures are measured as texture features in the spatial method. Texture extraction methods can be classified into three classes which are statistical, structural and model-based. The statistical texture method relies on low-level statistics of gray level images. Moments [58], Tamura texture features [59] and gray level co-occurrence matrix features [60] are regraded as statistical features in the spatial domain. Texture primitives and their placement rules [61] are used to extract structural features. Furthermore, Local Binary Patterns (LBPs) have been implemented originally to describe texture of the images [62]. LBP is a practical method to quantify color textures by utilizing patterns of local neighborhoods. LBP features have been used in various applications for texture classification [63]. LBP is widely considered as the state of the art texture descriptor because of low computational complexity and its invariance to changes in resolution. For model-based methods, random or generative models are applied in order to produce texture features. Markov random field [64], simultaneous auto regressive model [8] and fractal dimension [65] are some sample methods for texture feature extraction methods. Their computational cost increases due to including optimization part. Spectral features are calculated after converting images into the frequency domain. Fourier transform (FT) [66], discrete cosine transform [67], wavelet [68], and Gabor filters [53] are usual spectral feature extraction methods. FT and DCT give results rapidly without scale and rotation informations. Recently, the most applicable and accurate method is Curvelet

features [69] because it is able to successfully derive curvilinear properties [70]. The only problem for these methods are that they only can process a square area. In [71], a method which can transform an irregular texture region to a square texture region is proposed.

Shape features have been applied to retrieve images. Shape extraction can be classified into two groups which are contour-based and region-based methods. The boundary of the shape is of interest for contour-based methods, while the entire region is important for region-based methods. The contour-based method is more sensitive to noise with respect to area of interest. For this reason, the previous methods in image retrieval are generally region-based. Area, moment, circularity and eccentricity are common shape descriptors. In [72], the area method is performed. The work presented in [73] uses both eccentricity, elongation and area description methods. Eccentricity refers to the ratio of the major axis length to the minor axis length. The combination of many shape extraction methods has broad usage in order to design robust methods for image retrieval.

2.1.2 Similarity Metrics

Image distance measures are used to compare two images in CBIR in terms of features such as color, texture and shape. Then, search results are arranged with respect to their distance (similarity) to the query image. Intuitively, shorter distances refer to higher similarity. There are various kinds of similarity measurement methods that have been used for image retrieval. Image feature extraction, descriptor type and feature representation directly affect to distance metric selection. Metrics can be defined by vector spaces. Metrics on vector spaces such as Euclidean distance, city-block-distance and Minkowsky distance are commonly used similarity measurement methods, because their computational complexity is low. In addition, Earth Movers Distance (EMD) [74] introduces a solution by

computing an optimal agreement between the two multidimensional histograms. Another approach to calculate similarity is the graph-based technique, which is capable of representing both local color, texture, shape features and spatial distribution of region [75]. Graph matching refers to a special class of similarity measurement that is only applicable when the image content is represented by a graph [76]. A graph can illustrate both similarity and dissimilarity. In [77], the Blobworld approach [10], which is a type of graph-based descriptor, is applied. The similarity between the query and database images is calculated by matching the graph scheme [78]. Another method is classifier-based similarity measurement using the classification techniques of query by predetermined labels [79]. The statistical classifiers categorize new features by using high-level information derived from a training set of features with known labels [79].

2.1.3 Performance evaluation

Some evaluation criteria are applied to compare descriptors. These criteria can be classified into five groups in terms of feature extraction complexity, distance function complexity, storage requirements, effectiveness and validation environment. A feature extraction algorithm has to be fast, because the query extraction performance of descriptors directly affects system response time. Complexity analysis gives information about performance. The distance function of a descriptor takes too much time when a query is being processed in the CBIR system. For this reason, distance functions are expected to be fast during the search process. Indexing objects by a distance metric is also a crucial process. Storage requirement is really an important issue in evaluation. From this point of view, image compression techniques can help to find a way to reduce image features. For instance, dividing

into blocks with scaling-based or wavelet-based methods has been commonly used methods to encode image with requiring smaller space [80] [81] [82] and they can also be useful method for image retrieval feature reduction. Feature vectors representing image properties are stored by the image descriptor. The required storage space for feature vectors are directly affected by the number of images in the database. There are many dimensionality reduction techniques that can reduce storage size, e.g., PCA and LDA. The effectiveness of an image retrieval system is directly dependent on relevant results. In order to measure effectiveness in image retrieval systems, query and results should agree with each other. Hence, ground truth data can be used. Precision and Recall are the common choices to evaluate performance. Precision quantifies the proportion of relevant images to the retrieved set. Recall calculates the proportion of relevant retrieved images to all relevant images existing in the database. If the precision value is one, that means all retrieved images are relevant. If the recall value is one, all relevant images in the database are retrieved in the result set. When precision and recall values are both one, the system is considered as perfect. The validation environment includes testing image data to compare image descriptors. Several different evaluation measures and image databases are employed, making it difficult to perform a meta-analysis. Evaluation methods are rarely investigated in the literature. However, with increasing workshops where common benchmark datasets are used, the interest seem to be growing. There are some automatic evaluation systems, such as TREC-EVAL5, which were created by the CLEF organizers by considering top retrieval results of all submitted runs of the participating groups, applying Mean Average Precisions (MAP), Geometric Mean Average Precision (GMAP) and precision measurements [83].

2.1.4 Techniques in medical CBIR

In this section, various techniques currently used or proposed for medical image retrieval applications are described. Then, the general datasets and their evaluation techniques are briefly reviewed. Radiology serves as the main application field here.

Nowadays, the number of medical images is rapidly increasing. It is increasingly more complex to access and navigate these huge collections. The most important expenditures in clinical applications are constituted by imaging systems and their archiving systems [84]. Therefore, many methods from the computer vision and image processing fields have already been investigated for medical purposes. Patient identification, modality and study description [85], which are described in the DICOM standard [86], are performed to access data. In [22], web interfaces of medical image databases are defined. The Assert [87] system on the classification of high resolution computed tomographies of the lung [88] and the IRMA system for the classification of images into anatomical areas, modalities, and view points are rare samples to break away from past limited sources [89]. Recently, with usage of multi-modality devices which have the ability to register two modalities during the same imaging session, such as PET-CT [90] and PET-MR [91] scanners, performances have improved. Several studies have reported the potential clinical benefits of CBIR in clinical applications. For example, ASSERT [87] has gained improvement in the accuracy of diagnosis made by physicians [92]. Another CBIR experiment has been applied on liver CT providing real-time decision support [93]. An interface of a basic CBIR system retrieves images according to their calculated similarity score. The user can control the results and mark them as relevant or non-relevant in order to change the parameters for more effective results.

Features

The first medical information systems were initiated with providing only textual information about patients [89]. Recently, because of the rapid increase in amount of medical information and the increased accessibility to equipment for medical imaging, managing and storing of multimedia information have become relevant. Textual information has also been used to implement several studies in ImageCLEF medical on the same data [94]. Chu et al. [95] proposed a method that uses text information as query input for all data. In [96], this method was improved by contraction and combining group hierarchy approaches. Another combination method utilized correlation between visual and text features [97]. The performance of the combined method was higher than for the individual ones. The result was verified by different retrieval algorithms [98]. The latter use the ImageCLEF medical dataset detailed with information, which makes textual retrieval easier than visual retrieval. In addition to this dataset, the dataset without annotations was used in [99] and text features, visual features and combinations of them were also compared. The research concluded that visual features show better performance than textual features in indexing and retrieval. In comparison between combined features and others, the results indicate that the combination of them is more efficient in indexing, but there is no significant change in retrieval when they are combined. Semantic textual information is used with different kinds of information on decision trees for unsupervised classification by Quellec et al. [100].

As mentioned before, visual features are generally implemented in medical CBIR systems. Shape feature extraction methods used in medical image similarity retrieval are generally characterized as being region-based or boundary-based [101]. The shape represents a complementary space to color and texture [102]. This shape is defined as the probability density of a two dimensional variable using orthogonal moments [103]. Color

is one of the widely used visual features that can provide information about the visual content of an image. Some studies give vague description about general texture and color or gray level features [104]. Histogram samples that describe color and gray level features are explained in [105]. Local and global gray level features have been processed in [106]. Moreover, in [77] and [107] experiments are based on statistical distributions of gray levels to classify medical images and [108] uses a brightness histogram. Because medical images do not contain color, brightness calculations may be needed. If pathology images are normalized, they can be shown in different colors [109]. Also, texture and shape features play an important role for CBIR in the medical domain. To detect texture features, various edge detection methods have been performed, such as Canny [110] and Sobel [111]. Shape detection with Fourier filter [112], moments [113], co-occurrence matrices [114], Gabor filters [106], wavelets [114] and Markov models [112] methods have been frequently implemented in CBIR to extract texture features. Density is an important factor in mammography to find nodules within the breast [115]. The essential process in feature detection is segmentation and it is applied in many studies by different methods. In [116], segmentation is implemented for pathological images. Moreover, some basic morphological features are processed and the pattern spectrum is defined in [117]. The use of eigenimages for the retrieval of medical images in analogy to Eigenfaces for face recognition, is proposed in [118]. These features can help to easily retrieve images with histogram intersection method. Similar to general CBIR, semantic features are proposed for visual similarity queries with medical images [119]. In [120], decision forest tree method has been applied on different medical images with sub-windowing technique. Also, randomized trees are utilized by extracting gray level features from medical images for retrieval, classification, interest point detection and segmentation purposes [120]. A project for automatically attaching semantic labels to images or regions is described in [105]. As a result of these studies, performance has

been increased by combining textual and visual features or content and context of images, respectively.

Datasets

Medical content-based image retrieval studies are generally utilized for private datasets [37]. Using large publicly available datasets with ground truth facilitates benchmarking of CBIR methods. Several publicly available dataset are listed in the following. The ImageCLEF in the medical domain between 2005 and 2007 provide over 66,000 medical images [121]. These images consist of radiology, pathology, endoscopic, and nuclear medicine images. For the medical task of the ImageCLEF in 2013, over 300,000 images containing MR, CT, PET, ultrasound, and multi modalities in one image from the National Library of Medicine can be classified and retrieved with various methods. Another publicly accessible dataset is the PEIR Digital Library [122], which includes pathology images for educational purposes with text descriptions. These text descriptions can be used as ground truth. Furthermore, a chain study was performed over a 30 year period to illustrate health trends in USA, named the National Health and Nutrition Examination Surveys (NHANES) [123]. Only part of the dataset is publicly available and it consists of spine, hand and knee X-ray images. The Cancer Imaging Archive (TCIA) [124] has large image collections, which are built up for different specific purposes, e.g., the Lung Imaging Database Consortium (LIDC) of chest CT and X-rays [125]. Lastly, the VISCERAL project [126] is a new initiative with the purpose of providing 10 TB medical image data both for research and validation. The VISCERAL dataset should have two annotation standards; a gold corpus (experts annotation) and a silver corpus (participants consensus annotation). As of 2014, the plan is to provide retrieval benchmarks with larger dataset with weak label.

Medical Image Retrieval for Radiology

While general descriptors (color, texture, histograms) are frequently encountered features in all CBIR systems, they are also mostly applied in radiology. Vector distance similarity analyses are preferred in [79]. Furthermore, high dimensional feature vector spaces or with statistical classifiers are implemented in [127]. On the other hand, mixed descriptors are created by a combination of general descriptors and annotations [128]. Annotations help to fill the semantic gap in CBIR systems. For instance in [128], the system relies on an annotation tool in order to narrow the search area in the database. The other descriptor type is a specialized descriptor which exploits the interrelations between feature sets based on domain knowledge. For example in [37], the system retrieves pathology bearing regions in lung CT images with a human in the loop approach. Some examples of medical similarity measurement methods are the feature-vector distance matching method [77], graph matching method [129] and statistical classification [79] .

2.2 Background

Before description of the proposed methods, some established methods are explained as background.

2.2.1 Local Binary Patterns (LBP)

LBP features are utilized for texture description [130]. LBP is a practical method to measure gray-scale invariant texture features of images in local neighborhoods. The basic LBP operator replaces pixel values with labels by binarizing 3×3 neighborhoods of each pixel by a center pixel threshold. Then these pixel labels are converted to decimal numbers.

The binarizing function F for the basic LBP can be formulated as follows ;

$$F(I(Z_o), I(Z_i)) = \begin{cases} 0, & \text{if } I(Z_i) \leq I(Z_0), \text{ and} \\ 1, & \text{otherwise,} \end{cases} \tag{2.1}$$

where Z_0 is the center pixel and Z_i represents its 8 neighbors. LBP can also be defined as binary gradient representation. LBP features have been successfully used in many applications such as texture classification [63], face recognition [131], fingerprint identification [132], and automated cell phenotype image classification [133]. Main advantages of LBP are easy computation and its ability to produce multiple features in same dimension. In [134] and [135], LBPs are used to characterize medical images, for instance, magnetic resonance and mammography images. The method in [135] is also trained with a support vector machine (SVM) in order to distinguish mammographic masses from normal parenchyma. LBP is widely considered as the state-of-the-art texture descriptor because of low computational complexity and its invariance to changes in resolution.

In the proposed method, LBP is applied to multi-block patches in image with different scales. In other words, different neighborhood sizes are applied. After labeling the image parts, the histogram features are extracted from the local region labels. The region can be rectangular, circular or triangular.

2.2.2 Support Vector Machines (SVM)

The support vector machine (SVM) is a machine-learning method to classify data. Image annotation and classification can be considered as a first step for speeding-up image retrieval in large databases. SVM is a popular machine learning algorithm to perform reliable

and generally fast classification [136]. SVM takes a set of training data that consists of inputs and desired outputs [137] and SVM maps the input vectors to a higher dimension feature space, because separation of classes and detection of pattern are easier in higher dimension space [138]. Kernel functions map the input data into a higher dimension feature space, because this data could be easily separated or better structured [138]. The main task of SVM is to create hyperplanes using support vectors in a higher dimension space [139]. A hyperplane separates the space into two half spaces. A good separation is achieved by hyperplanes that have the largest distance (margin) from the nearest data points. There are many kernel function in SVM which are used for mapping [140]. The various kernel functions can be grouped into two methods which are local and global. Some local kernel functions examples are RBF, KMOD and Inverse Multi-quadric. Linear, Polynomial, Sigmoid are global kernels.

Recently, SVM methods have been widely utilized for medical applications. For instance, linear SVM methods with quadratic optimization have been proposed for CT brain images [141]. Also, SVM has been combined with K-NN classifiers [142] and with boosting [143]. Since it is common to have to deal with high-dimensional feature space, most indexing methods cannot perform within a reasonable time.

In this thesis, SVM is used as a statistical multi-class classifier in order to predict query image label for medical image retrieval purposes.

2.2.3 Context-Aware Saliency

The context-aware saliency algorithm detects image regions that best represent the 'scene' [144]. It is a region detection algorithm based on four principles observed in the psychological literature: local low-level considerations, global considerations, visual organizational

rules, and high-level factors. Local low-level factors consist of contrast and color. Visual organization rules state that visual forms may possess one or several centres of gravity and high-level factors include priors on the salient object location and object detection. Further descriptions and implementation of this algorithm are available on the authors' website [1]. The basic principles of the saliency detection equation are focused on both isolated pixels and the surrounding path. This approach promises immediate content. As a first step, pixel color similarity is defined and then the position distance between patches are considered. A dissimilarity equation measures distance dissimilarity based upon both color and position:

$$D = \frac{d_{color}(p_i, q_i)}{1 + c \times d_{position(p_i,q_i)}} \tag{2.2}$$

$$S_i = 1 - exp(-\frac{1}{K}) \sum_{i=1}^{n} d(p_i, q_i), \tag{2.3}$$

where p_i and q_i are compared patches, d_{color} and $d_{position}$ correspond to distance in terms of color and position respectively, c is a constant, K is the number of possible comparisons, D is dissimilarity between comparing parts and S_i is total similarity.

Multi scale enhancement has been applied. The same saliency equation is applied in different scales and the average of them is taken. Lastly, in order to obtain immediate content of salient region, the threshold is set and weighting is based on Euclidean distance metric. Most of salient region detection methods have been designed with general images by using their color distinctiveness [145]. Some of them consider patch distinctiveness [146] [144]. Some studies claims "medical saliency" by identifying informative regions based on internal image statistics [147]. The saliency algorithm is utilized in this thesis to assign relevant regions in the image in order to apply a folding method.

[1] http://webee.technion.ac.il/labs/cgm/Computer-Graphics-Multimedia/Software/Saliency/Saliency.html

2.2.4 Autoencoders

Autoencoders are a specific architecture of neural networks consisting of encoding and decoding operations in order to reach minimum error. As an unsupervised learning technique, the autoencoder utilizes backpropagation networks without training [148]. Autoencoders, having a general $n/p/n$ architecture, encode n inputs into p positions, and then decode p positions back into n outputs. $p<n$ is general case in which autoencoder functions as a compressor to reduce dimensionality. An autoencoder is basically a shallow neural network with some level of error. The error can be reduced if the network is deepened, for instance make it an $n/m/p/m/n$ autoencoder with $n>m>p$ [149]. For any $A \in A$ and $B \in B$, the autoencoder transforms an input vector $x \in F^n$ into an output vector $A \circ B(x) \in F^n$. The corresponding autoencoder problem is to find $A \in A$ and $B \in B$ that minimizes the overall distortion function:

$$minE(A, B) = min_{A,B} \sum_{t=1}^{p} E(X_t) = min \sum_{t=1}^{p} \triangle(A \circ B(X_t), X_t), \tag{2.4}$$

where A is class of functions from decoding (from G^p to F^n), B is a class of functions from encoding (from F^n to G^p), X is set of training vectors and Y is a target vector, $\triangle$ is the dissimilarity or distortion measurement via Hamming or L_n distance functions [149]. Deep architectures provide reuse of features and they conduct more high level features. In other words, high level features can be constructed by autoencoders.

Recently, the denoising autoencoder has been developed to reconstruct data from one of its corrupted versions [150]. Deep architectures can learn many layers' features of image color. Afterwards, they can encode images as the binary codes [151]. Mammograms can be compressed by use of patches of images [152]. There are broad surveys on deep learning and autoencoders [153]. [154] attempts to propose a generalized autoencoder via manifold

learning and uses it for digit and face image manifolds. [149] presents a generic framework about kinds of autoencoders. To deal with complex databases, stacked autoencoders have been implemented in [155]. Furthermore, [156] proposes a feature ensemble learning method based on sparse autoencoders for image classification. As mentioned in the next section, this work proposes to train a shallow autoencoder and record the error histogram of each class to eliminate image blocks that are irrelevant for retrieval.

Chapter 3

Image Area Reduction via Folding and Autoencoding

3.1 Introduction

With acquisition of many images each day in medicine, the necessity of searching and navigation in databases has drastically increased. CBIR methods have been developed for more than two decades to provide high-quality care through accurate and efficient diagnosis and treatment planning. In medical imaging, CBIR can enormously contribute to more reliable diagnosis by classifying and retrieving similar images. Almost all medical retrieval methods have been proposed to increase the accuracy of retrieval, because medical images are much more sensitive to process comparing to other images. Accurate retrieval methods are important, but also retrieval time and storage capacity play crucial roles in practice because efficiency also states on accuracy, computational time and storage. For this reason, the main purpose of this thesis is to obtain a high classification score and to

generate effective retrieval results with less computational complexity and lower storage requirement. In order to save time and to obtain a high classification score, the fundamental goals are envisioned to reach by reducing the number of features by decreasing the image area subject to feature extraction (traditionally the entire image is used).

With this in mind, two approaches are developed in this thesis. The first method is based on folding less relevant image regions whereas some alignment and summation operations are necessary to keep the intensity values of the folded regions. The second approach is a retrieval system which eliminates irrelevant image areas completely and then extracts features from the remaining image area. The steps of both approaches and extracted features are described in the following sections in detail. The benchmark dataset used in this study and validation details are described in this chapter . Finally, all results are illustrated and discussed from different perspectives.

3.1.1 Problem Definition and Dataset

While there have been many studies to increase retrieval accuracy in medical imaging, retrieval time and acquired storage have been rarely investigated in the literature. However, these issues play an important role especially because of the dramatic increase of medical images produced each day. A medium-size hospital can produce 2 terabytes (TB) of data per 100,000 studies [25]. Most hospitals produce 200,000 studies per year on average. This means that each year the hospital should store 4 TB of image data and maintain them for several years. Generally each hospital has some regulations about keeping the medical history of their patients for approximately 7 years. According to these general numbers, each hospital should support to store at least 28 TB of image data [124]. Moreover, navigation and search in big datasets take a large amount of time. The main retrieval

problems in big datasets are run time and storage. In order to examine these problems, a big dataset which has at least 10,000 images with their ground-truth should be select. Thus, the IRMA 2009 dataset was chosen to investigate how retrieval time and storage needs can bee reduced.

The IRMA 2009 (IRMA[1]), which is a publicly available dataset, has 14,410 X-ray images that have been randomly collected from daily routine work at the Department of Diagnostic Radiology of the RWTH Aachen University[2]. X-ray images in the dataset have been taken from different anatomical regions of body, from different directions and from different biological systems. Downscaled images were collected from different ages, genders, view positions, and pathologies [157]. Each image is labeled with IRMA codes. These codes correspond to the ground-truth of images. 193 classes are defined according to 2008 IRMA codes. The IRMA code comprises four axes with three to four positions: 1) the technical code (T) (modality), 2) the directional code (D) (body orientations), 3) the anatomical code (A) (body region), and 4) the biological code (B) (the biological system examined). The complete IRMA code consists of 13 characters TTTT-DDD-AAA-BBB, with each character in $\{0, \ldots, 9; a, \ldots, z\}$. As many as 12,677 images are separated for training. The remaining 1,733 images are used as test data. Figure 3.1 shows sample images from the IRMA 2009 dataset. The first lines on the bottom of the figures state the IRMA 2008 code of the image, and the second line is the IRMA 2005 code of images.

In this project, the IRMA 2009 dataset has been used with specified 2008 IRMA labels (consisting of 193 classes) for retrieval and annotation purposes. Otherwise, same dataset is utilized with general 2005 IRMA labels (consisting of 57 classes) for classification purpose. 2005 IRMA labels are more general than 2008 IRMA labels because it has been made of 6

[1]http://irma-project.org/

[2]http://www.rad.rwth-aachen.de/

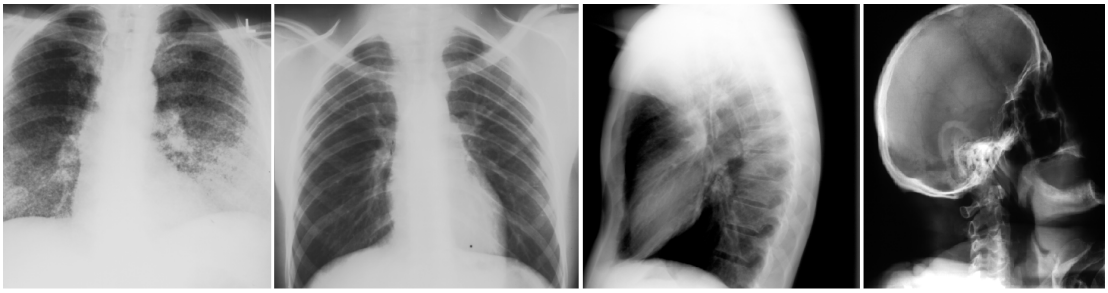

Figure 3.1: Sample images in IRMA 2009 dataset from different groups with IRMA 2008 and IRMA 2005 codes. The first line on the bottom of the images corresponds 2008 IRMA codes and the second line is 2005 IRMA codes.

characters from the top hierarchical classes, TT-D-AA-B. In the 2009 dataset, each image cannot have been coded according to the 2005 IRMA coding regularity. A total number of 12,631 images from a training set and 1,639 images from a testing set have 2005 IRMA codes.

3.2 First Area Reduction Approach: Image Folding

The first technique to implement is to reduce the image area via folding. A statistical classification method is developed with saliency-based folding. Classifier-based similarity measurement systems retrieve similar images by using classification of a query image according to a set of ground-truth labels. SVM is used as a statistical classifier. If the salient (significant) image regions are detected, then folding less significant (rectangular) regions is applied inwardly and their intensity values are added to the inward (salient) pixels. The folded image is again divided into fewer regions (3×3). Afterward, texture features are extracted from these blocks, because LBP from multi-blocks are histogram-based fea-

tures that contain texture and spatial information about images. In following sections, the details of this approach will be explained.

3.2.1 System Architecture

The first approach is comprised of preprocessing, off-line training and on-line usage (see Figure 3.2). During the preprocessing stage, saliency detection and image folding take place. In the off-line stage, SVM is trained on histograms of LBPs of not only folded images but also images which are not folded. To detect salient regions, the context-aware saliency algorithm [3] is used [158]. Saliency maps of images have been obtained to create a saliency template by averaging all saliency maps. The average of saliency maps is first calculated internally within each class, then the global average is taken across all classes. Image folding and calculation of the saliency template are detailed in following sections. After image folding, the folded image is divided into blocks. Then, the feature extraction process begins, where local texture features are extracted using the LBP method. LBP is implemented from two different scales to be applied on local sub-image areas. After that, histograms of these blocks are derived form LBP features. The LBP histograms of training sets of data are used to train SVM and support vector models stored. Then, multi-class SVM is applied as a classifier with a radial basis function in an offline stage. Both folded and not folded images from the IRMA training set have been utilized in order to compare the two methods based on efficiency and time. Lastly, learned SVM models have been tested with folded and not folded images from the testing dataset in the online stage. Alghorithm 3.2.1 and Figure 3.2 give a generic overview of the proposed system. The pre-processing of image data mainly consists of two steps. As a first step, a saliency

[3] http://webee.technion.ac.il/labs/cgm/Computer-Graphics-Multimedia/Software/Saliency/Saliency.html

Algorithm 1 Proposed Approach

1: ———— Pre-Processing————-

2: Read all images from Training set; I_i.

3: Calculate saliency template S^*

4: $I_i^F \leftarrow$ Apply folding on all images I_i,

5: Save S^* and all folded images I_i^F.

1: ———— Off-line Training ————

2: Read folded images I_i^F

3: Set number of classes N_C

4: Extract LBP features from folded images

5: Train SVM to learn model

6: Save SVM model

1: ———— On-line Classification ————

2: Read query image I_q

3: Read the Saliency template S^*

4: Read the SVM model

5: $I_s^q \leftarrow$Apply saliency template S^* on I^q

6: $I_F^q \leftarrow$ Apply folding F on I_s^q

7: Extract LBP features from I_F^q

8: Classify the query using SVM

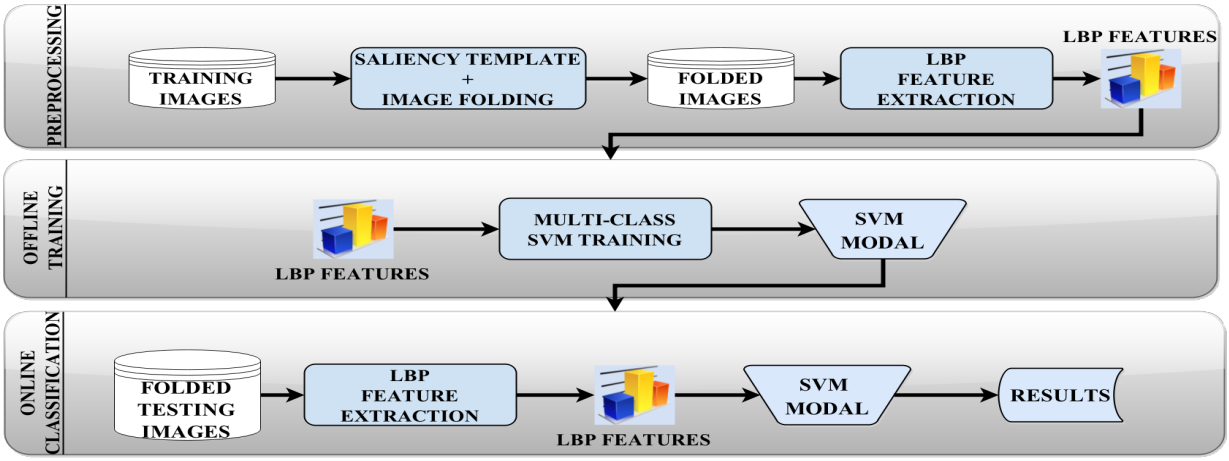

Figure 3.2: Three stages of image area reduction via folding.

template is created, and in the second step, image folding is formulated based on the saliency template as described below.

Saliency Map Template

The detection of salient regions of an image is crucial to extract effective information. This work proposes to create a saliency template by averaging all saliency maps which are detected by the aforementioned context-aware saliency algorithm. Saliency maps of all training images are generated and averaged to calculate a saliency map template (Algorithm 2). Figure 3.3 shows four saliency maps. The first three images are from different classes and the last one is the template which is created by averaging all saliency maps.

Algorithm 2 Pre-Processing Stage: Saliency Template S^*

1: $N_C \leftarrow$ number of classes; $i = 1$.
2: Initialize saliency template S^*
3: **while** $i < N_C$ **do**
4: Calculate the saliency map S_i for image I_i [158]
5: $S^* \leftarrow S^* + S_i$
6: $i \leftarrow i + 1$
7: **end while**
8: $S^* \leftarrow \frac{S^*}{N_C}$

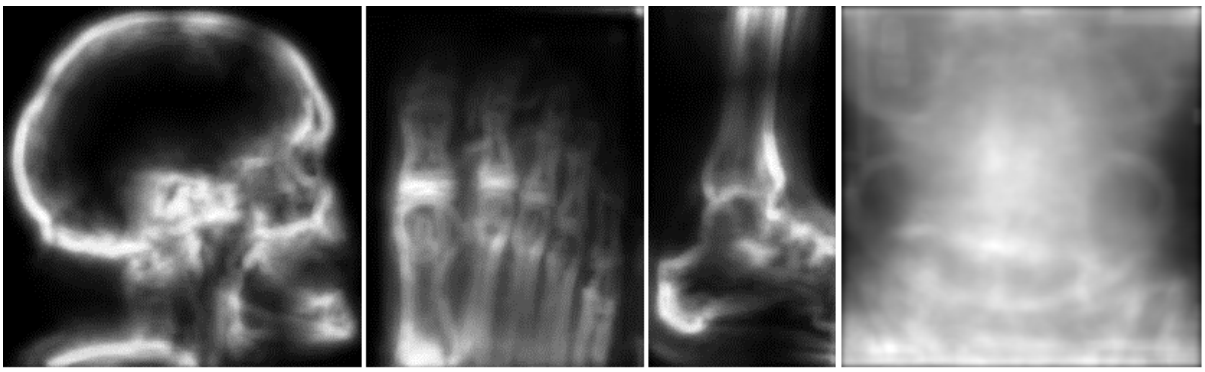

Figure 3.3: The first three images are samples of saliency maps, the last image is the template saliency map for all images

The salient, less salient and not salient areas are defined for training data by dividing images into N sub-blocks. Then, based on the saliency map, the folding is applied. The new images with reduced area can now be used for local pattern analysis. The average of saliency maps is first calculated internally within each class, then the average is taken across all classes.

Image Folding

Folding leads to the area reduction of an image. A is the rectangular region in image I ($A \subset I$) and I' is the folded image ($I' \subset I$). I' can be calculated through $I' = A + I \backslash A$ whereas the sign '\' denotes the set-theoretical subtraction. The main purpose of folding is to reduce the image area to decrease the dimensionality of features without loosing information. While area is reduced, pixel information is not deleted. Salient regions gain importance with the saliency template and non-salient regions become irrelevant without deleting. The folding is applied only row-wise and column-wise. This method can be extended to work block-wise as well. If block-wise folding or smaller region folding is applied, the area will be reduced, but the folding decision takes more time than the run time of row- and column-wise approaches. For this reason, the simple folding type, which is row- or column-wise, is used.

The proposed folding approach starts with dividing an image into equal parts. Each part is assigned a value according to a saliency template. Then, each part is folded with other possible parts to find the best combination of folding. The main step in folding corresponds to summation of pixel values in the saliency template covered by the candidate folding area. The image will be folded column-wise or row-wise. The folding stages are described in Algorithm 3 step by step. Folding starts with dividing the template saliency map into M parts (e.g., $N \times N$ with $N = 4$). All possible column folding cases are tried by aligning two columns and then taking the summation of all pixel values in the saliency template. The maximum pixel value of summed columns $s_{\max}^{c_i}$ are saved. Hence, for each folding case we have $s_{\max}^{\text{column}} = \sum_i s_{\max}^{c_i}$. All possible row folding cases are processed by aligning two rows and then taking the summation of all pixel values in the saliency

template. The maximum pixel value of summed rows $s_{\max}^{r_j}$ are saved. Hence, for each folding case we have $s_{\max}^{\text{row}} = \sum_j s_{\max}^{r_j}$. Then the method finds the folding F_{best} function that satisfies the following : $s = \min(s_{\max}^{\text{column}}, s_{\max}^{\text{row}})$. Finally, the folding F_{best} is applied to the image.

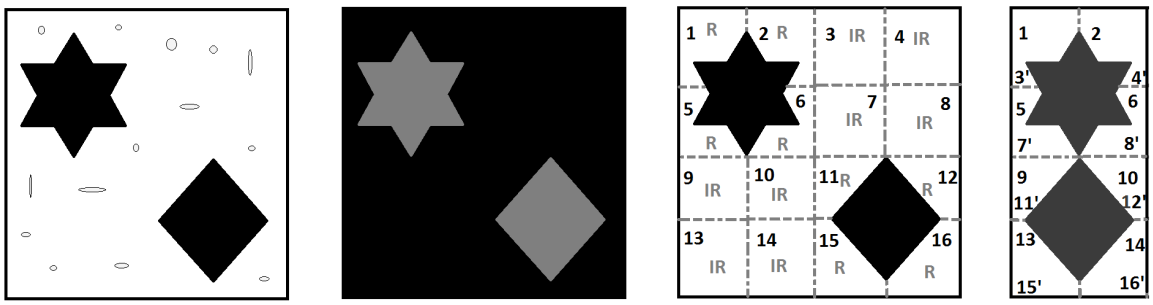

Figure 3.4: The input image (left) is processed to find a salient regions (second image from left). Subsequently, non-salient regions are marked as IR to be folded with relevant regions which are labeled with R. The blocks have been numbered and folded parts have been included with " ' " such as 15' means that 15 is folded. Column-wise folding on the example image is illustrated lastly (right).

After applying image folding, the image area is reduced by half. The ratio of salient to non-salient regions in the folded image is more than the ratio of salient to non-salient regions in unfolded regions. This shows that relevant regions are preserved in folding. As mentioned before, the folded image is divided into 9 (3×3) equal regions. LBP features are extracted from these regions by two scaling factors (1 and 2). Any number of blocks which is under 16 (4×4) may be selected, because image area is smaller than before. The rectangular block region is selected to be 3×3. Afterwards, LBP histogram features are trained with multi-class SVM [159] to classify images. SVM kernel type is set to Radial Basis Function. In the online part of the method, an image query is selected from the

Algorithm 3 Pre-Processing Stage: Image folding

Set number of blocks $M(NxN = 4x4)$

Read saliency template S^*

Read input image I

while not all combinations tested **do**

    Align two columns

    Take summation of all pixel values in S^*

    Keep $s_{\max}^{c_i}$(maximum value of summed columns)

    Update $s_{\max}^{\text{column}} \leftarrow \sum_i s_{\max}^{c_i}$

end while

while not all combinations tested **do**

    Align two rows

    Take summation of all pixel values in S^*

    Keep $s_{\max}^{r_j}$(maximum value of summed columns)

    Update $s_{\max}^{\text{row}} \leftarrow \sum_i s_{\max}^{r_j}$

end while

Find the folding F_{best} that satisfies $s = min(s_{\max}^{\text{column}}, s_{\max}^{\text{row}})$

Apply the folding F_{best} to I.

test dataset and LBP features are calculated from the saliency-based folded image as new images are encountered. SVM classification is performed with these features and compared to the SVM-based approach on the IRMA test dataset without folding.

3.2.2 Analysis of Results

SVM classification results on folded images have been compared with SVM classification results without folding in terms of accuracy and time. Classification accuracy is measured by a specific error score calculation method developed for the IRMA dataset. Accuracy measurement is crucial in order to show that the proposed method runs as effectively as existing methods. For this reason, ImageCLEF has defined an error score evaluation method in order to evaluate the performance of methods on IRMA datasets [157].The equation is

$$E = \sum_{i=1}^{n} \frac{1}{b_i} \frac{1}{i} \delta(I_i, \hat{I}_i), \tag{3.1}$$

where b_i is the number of possible labels at position i, and δ is the probability of decision (zero or one). δ is the decision error value in the equation. If the decision is correct, $\delta = 0$, if the decision is wrong, $\delta = 1$, where I_i is the correct code for an axis and $\hat{I}_i$ is the classified code for an axis. E is the error score. For every axis, the maximal possible error is computed and the errors are normalized between 0.25 and 0. If all positions in all axes are wrong, the error value is 1, or the reverse condition error value is 0.

Classification Error

All folded images in test datasets have been classified by multi-class SVM and the total error score is evaluated using the error score function. All images without folding in the

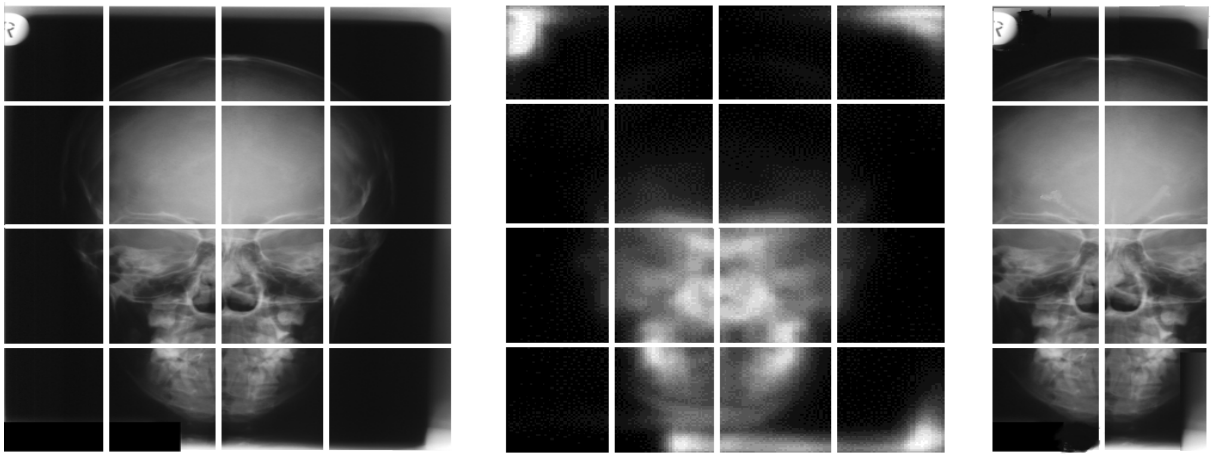

Figure 3.5: The input image (left) is processed to find salient regions using the saliency template (middle image). Column-wise folding is applied on the example image and selected column combinations are summed and normalized (right image).

test set have been also classified with multi-class SVM and the error score is measured by the same function. The error score for the proposed method of SVM image classification with multi-scale LBP on the saliency map-based folded image is 153.07. If images are not folded, the SVM classification error slightly decreases to 146.55. This slight decrease in error comes with a huge cost in computation, because the dimensions of features are twice the dimensions of the folded image. This means that the accuracy does not drop considerably while time and computational cost decrease. Saliency-based folding reduces the complexity without loosing important patterns in salient regions.

In order to test saliency template effects, folding is implemented without applying the saliency template on image. Without consideration of salient areas, folding has been tried in different directions. The error clearly and considerably increases. In addition, if we fold images in random ways, the accuracy significantly drops. Clearly, a saliency map plays a crucial role for the decision on how to fold an image. Figure 3.5 illustrates the folding

method application to a real image from the IRMA dataset. In the example, folding is applied column-wise according to their salient areas.

There are comparable studies in the literature. For instance, The IRMA dataset is used in ImageCLEF 2009 competition with 2008 IRMA code and basic LBP on 4×4 multi-blocks is applied by VPASabanci. The error score is reported as 261.2 [157]. In addition, the lowest error score in ImageCLEF 2009 with 2008 IRMA code is 169.5. The comparison of classifiers and SVM results are outlined in Table 3.1. According to this comparison table, the proposed approach provides higher accuracy than previous methods. Besides the error score, an example of the GUI interface in Figure 3.6 shows that classification works properly.

Memory and Time

Regarding memory and time, a decrease in area causes reduction in memory requirement and computational time. The image area is reduced by 50% with saliency-based folding. As the result of area reduction, the number of image blocks decreases too. Also the feature dimensionality reduces from 1888 to 1062 which is a 44% decrease. SVM needs 141.17 s training time and 92.51 s testing time without saliency-based folding. In contrast, with saliency-based folding, SVM only needs 60.36 s training time and 52.56 s testing time for all images. To neglect the overhead for the saliency calculations, and only by looking at the time required for testing (online execution), using the proposed approach accelerates the classification process by approximately 13%. Considering the challenges of big sets of image data in the medical field, this can be a tremendous improvement. As a result of the error score, dimension and time reductions, it is experimentally verified that the proposed method has succeeded in its goals. Moreover, Figure 3.6 illustrates the efficiency of the method in retrieving images correctly. In summary, a classification-

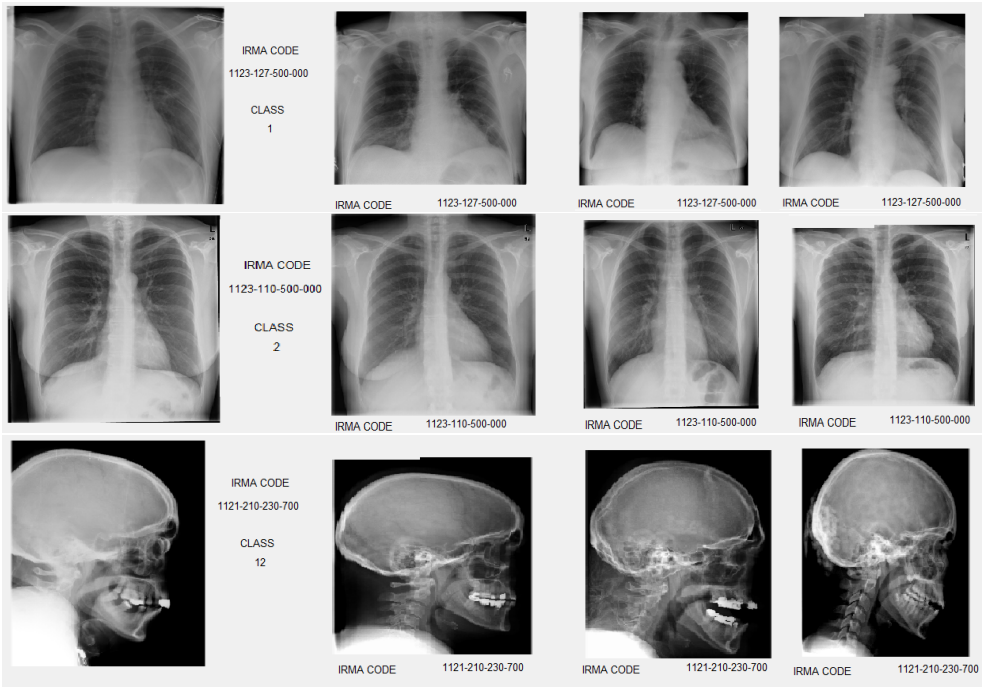

Figure 3.6: The images on the left are query images , the other three images on the right are three results from classified categories in the proposed study.

based CBIR system depends on good classification first to assign a query to the right image category within a short time. The time requirements become paramount when dealing with big data. The proposed medical image classification using the saliency-based folding method appears to be an effective method when support vector machines and local binary patterns are employed. Folding non-salient (non-relevant) parts of the image may lead to a slight increase of classification error. That may be expected since folding areas overlap with salient regions. Overlapping salient and not salient regions may result in slight distortion. However, the proposed approach does accelerate the online classification. This is an advantage that might be crucial for big image data (reduction from 53 ms per image

Method	Error score	t(ms)/image
The proposed 4×4 Multiscale LBP+SVM	146.55	53
The proposed 3×3 Multiscale LBP+SVM w. folding	153.07	30
TAU [157] (Patches+BoW+SVM)	169.5	–
VPASabanci [157] (Local+Block position)	261.2	–

Table 3.1: Image classification results (Error score and time TAU and VPASabanci results as reported in literature)

to 30 ms corresponding to 43% acceleration). Furthermore, storage requirements have been decreased from 1888 to 1062 feature vector length by the proposed folding method. With respect to time and memory reductions, the method gives better solution to image classification. The most critical part of the method is the decision of how to fold image blocks. Different approaches can be examined in future work to investigate the feasibility and the potential effect of folding blocks and not necessarily just folding rows and columns. Moreover, other methods can be applied by deleting non-salient blocks. We have examined this method by implementing an autoencoding approach.

3.3 Second Area Reduction Approach: Autoencoding Error

The previous system showed that non-salient regions cannot significantly affect the classification accuracy. Also, elimination of non-salient regions can speed up the method. In order to demonstrate ineffectiveness of unimportant regions on retrieval, another (different) approach is proposed in this section that reduces the image area as well. While the elim-

ination step is before the feature extraction process in the image folding method, in this approach, the elimination stage takes place after classification and before similarity measurements. The proposed method detects irrelevant image blocks in each medical image class by analyzing the error histogram of decoding errors when autoencoders are applied to each image block. Features from irrelevant blocks are eliminated based on error histograms of their classes. With this in mind, an $n/p/n$ autoencoder has been utilized while $p < n$. The hypothesis of this study is that the relevance of image blocks is directly proportional to error of an autoencoder when the hidden layer is smaller than input/output layer. There is assumption that images are widely free from noise.

3.3.1 System Architecture

In a specific retrieval task, all regions of an image may not be relevant to search similarity. For this reason, selected regions of interest in an image, ROI, can be efficiently processed by the retrieval system. In order to select relevant regions, an approach that attempts to eliminate small patches (blocks) of the image from the feature extraction process is developed. This region reduction should be based on some universal criteria to ensure a general approach. This approach consists of configuration, training, and testing retrieval parts. Algorithm 4 describes the main steps of the proposed approach. In order to implement a complete solution, LBP features and SVM are used to classify the images. In the configuration step, the number of image blocks k is determined. The percentage of area reduction and the number of autoencoder hidden layers are also decided by the user in this step. These variables affect the results as discussed in the results section. LBP features of all training images have been extracted by multi-block LBP descriptors with given block number k in the training step. Utilizing LBP histogram features and general ground-truth

classes (57 classes), SVM creates the support vector modal. This the SVM modal is used in the test section of this method to classify the query image. In addition to building SVM modal, the training section includes autoencoder error calculation. $n/p/n$ autoencoder errors are sorted by a histogram for all classes and image features have been eliminated with respect to the sorted error histogram (lower error indicating lower relevancy). The query image from the testing dataset is used to extract LBP features by using multi-blocking. After SVM classification, the LBP feature vector size is reduced via an error histogram elimination function. Lastly, similar images are retrieved by eliminated features derived from eliminated blocks by using similarity measurement.

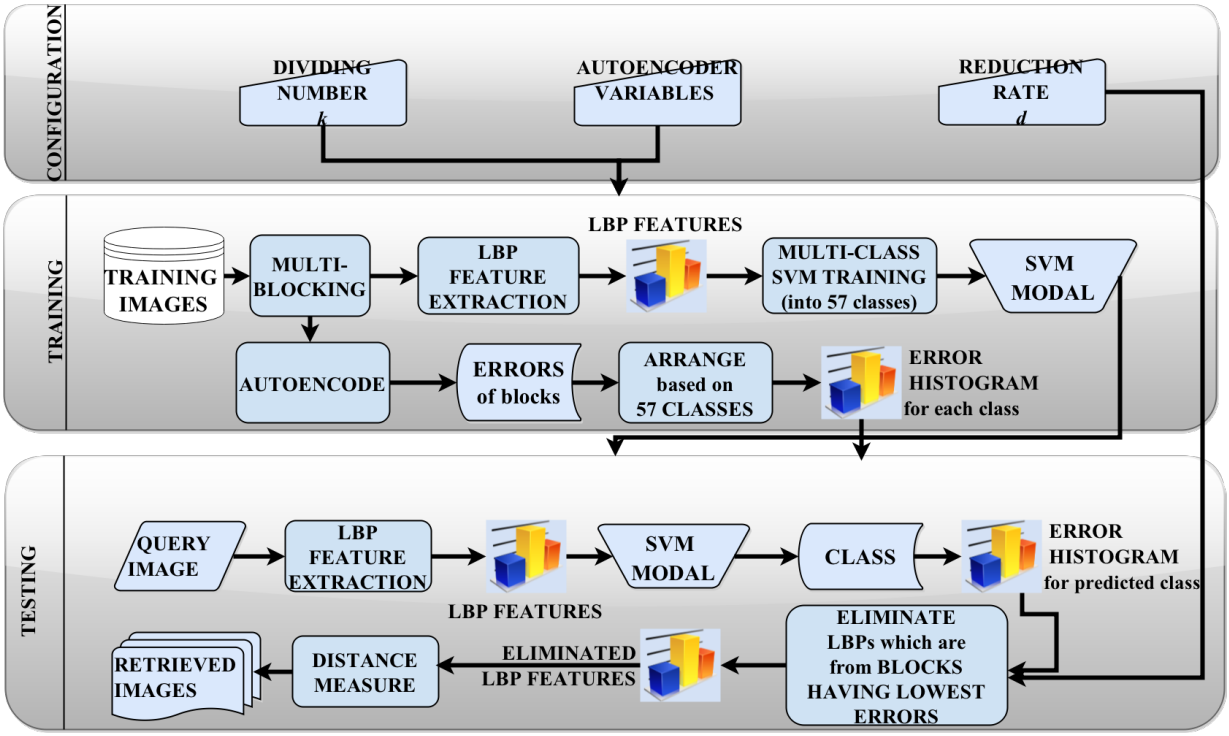

Figure 3.7: Three stages of the proposed image retrieval method .

LBP and SVM Settings

Local binary patterns have been extracted from 3×3 neighborhoods in each image block. The LBP descriptor calculates features from two scales (1 and 2). These binary patterns are converted from binary numbers to decimal numbers in order to calculate a histogram $\mathbf{h}_{\text{LBP}}$. These histograms were used for both classification and retrieval. The LBP histogram features from IRMA training dataset are utilized to train the multi-class SVM [159]. The radial basis function is set as the kernel for SVM. For each block in one scale, the LBP image descriptor creates 59 histogram features. For this reason, dimension of LBP features depends on the number of image blocks. Different blocks numbers have been tried within this method, such as 4×4, 5×5 and 6×6. The relationship between accuracy or time and block numbers can be derived from these results.

Shallow autoencoders reduction

As a neural network, autoencoders are explained in Section 2. An autoencoder with $p < n$ is used to reduce dimensionality. The $p < n$ autoencoder is basically a shallow neural network with some level of error. The purpose of this study is to design a shallow network to keep the decoding error high. Utilizing shallow networks, the relationship between a relevant region and error value can be examined. The errors in decoding image blocks are captured with a histogram matrix $\mathbf{H}$ for each class. These histograms will enable us to find retrieval-irrelevant image regions. Figure 3.8 illustrates the idea of relevance quantification via an autoencoding error histogram. The hypothesis of this work is that if the image block contains relevant information, the encoding error is expected to be high for a $n/p/n$ type of

autoencoder when $p < n$. In contrast, if uniform regions are encoded, low decoding error can be expected. Therefore, lowest decoding error states the least contribution to accurate retrieval. These blocks which have lowest decoding errors will be eliminated according to the desired area reduction rate (e.g., experts determined that 25% of the least relevant image area may be discarded). If an image I is divided into $k \times k$ blocks and the reduction rate is given as $d \in [0, 1)$, the task is to eliminate as many as $\lfloor d \times k \times k \rfloor$ blocks. The autoencoding errors are recorded for all blocks of a certain image class, in order to update the class histogram for every query. The SVM classifies the query image and assigns it to a certain general class (among 57 classes). After the classification, the most similar images to the query are retrieved using the reduced texture features. The proposed reduction of feature dimensionality using autoencoding error analysis occurs to improve the speed and space requirements of the intra-class retrieval task.

Similarity measurement

A distance metric calculates the similarity of two images in terms of their features. Then, search results are sorted with respect to their measured distance. The shorter distances correspond to higher similarity. Image features and descriptor selection and their representation methods directly affect the choice of distance metric. Metrics on vector spaces (e.g., The Euclidean distance, the city-block-distance and the Minkowsky distance) are the most commonly implemented methods due to their computational simplicity. In this work, the Pearson distance metric d is used:

$$d^2 = \sum_{i=1}^{n} \frac{(X_i - Y_i)^2}{X_i + Y_i} \tag{3.2}$$

$$p = \frac{\sum_{i=1}^{n}(X_i - \hat{X})(Y_i - \hat{Y})}{\sum_{i=1}^{n}\sqrt{(X_i - \hat{X})(Y_i - \hat{Y})}} \tag{3.3}$$

$$d = 1 - p \tag{3.4}$$

where X_i and Y_i are compared data positions, $\hat{X}$ and $\hat{Y}$ are the state mean of the data and d is the distance.

Accuracy Measurement for Retrieval

Retrieval evaluation is important in comparing results of similarity metrics with other methods. As common evaluation methods in medical image retrieval, precision and recall are calculated. Looking at the top m retrieved images, the number of correctly retrieved images (true positives T_p) and wrongly retrieved images (false positives F_p) can be used to calculate the precision $P_{\text{top m}}$ of the retrieval:

$$P_{\text{top m}} = \frac{T_p}{T_p + F_p}. \tag{3.5}$$

The average precision can be calculate with:

$$P_{\text{top m av}} = \frac{1}{x}\sum_{o=1}^{x} P_{\text{top m}} \tag{3.6}$$

where x is number of trials for the same number of results (same top$m$). Similar to precision, using the top m retrieved images, the number of correctly retrieved images (true positives T_p) and wrongly not-retrieved images (false negative F_n) can be used to calculate the recall $R_{\text{top m}}$ of the retrieval:

$$R_{\text{top m}} = \frac{T_p}{T_p + F_n} \tag{3.7}$$

$$R_{\text{top m av}} = \frac{1}{x} \sum_{o=1}^{x} R_{\text{top m}} \tag{3.8}$$

These measurements are applied 100 times within same classes for top 10, 20 and 30. After calculating all of them, the average of precisions and recalls are calculated in terms of their top retrieved image number m.

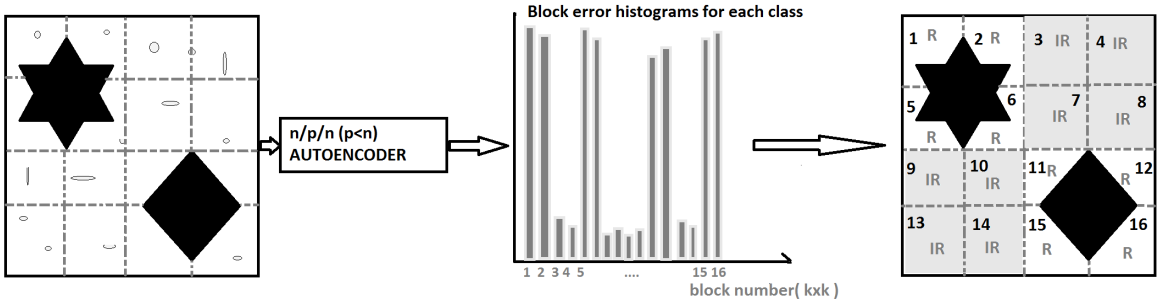

Figure 3.8: Schematic illustration of the proposed approach: Image blocks are autoencoded in an $n/p/n$ architecture. The encoding error of each block is recorded for each image class to create the error histogram. A desired reduction (e.g. $50\% = 8$ blocks) can be used to exclude a number of blocks (gray blocks) from feature extraction.

Algorithm 4 Proposed approach

——- **Configuration** ——-

Set k to divide the image into $k \times k$ blocks

Set the desired reduction rate $d \in [0, 1)$

Set $n/p/n$ for the autoencoder ($p < n$)

——- **Training** ——-

Get the number of training images $m_{\max}$

Initialize the feature matrix $\mathbf{F}$

Initialize the class vector $\mathbf{c}$

Initialize the error histogram $\mathbf{H}$ for all classes

for each $i \in \{1, 2, \ldots, m_{\max}\}$ **do**

  Read the training image $\mathbf{I}_i$ and its class c_i

  $\mathbf{c} \leftarrow c_i$

  for each $j \in \{1, 2, \ldots, k \times k\}$ **do**

    $\mathbf{B}_j \leftarrow \mathit{currentBlock}(\mathbf{I}_i)$

    $\mathbf{f} \leftarrow \mathit{extractLBPfeatures}(\mathbf{B}_j)$

    $\mathbf{F} \leftarrow \mathit{appendFeatures}(\mathbf{F}, \mathbf{f})$

    $\mathit{error} \leftarrow \mathit{autoEncode}(\mathbf{B}_j)$

    $\mathbf{H}(c_i, j) \leftarrow \mathbf{H}(c_i, j) + \mathit{error}$

  end for

end for

$[\mathbf{v}_1, \mathbf{v}_2, \ldots] \leftarrow \mathit{TrainSVM}(\mathbf{F}, \mathbf{c})$

Save support vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots$

Save the error histogram $\mathbf{H}$

——- **Testing** —— (See Algorithm 5)

Algorithm 5 Testing step in Proposed Approach

1: Read support vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots$

2: Read the error histograms $\mathbf{H}$

3: Read new image $\mathbf{I}_{\text{new}}$

4: **for each** $j \in \{1, 2, \ldots, k \times k\}$ **do**

5: $\quad \mathbf{B}_j \leftarrow \mathit{currentBlock}(\mathbf{I}_{\text{new}})$

6: $\quad \mathbf{f} \leftarrow \mathit{extractLBPfeatures}(\mathbf{B}_j)$

7: **end for**

8: $c_{\text{new}} \leftarrow \mathit{classifySVM}\ (\mathbf{f})$

9: $\mathbf{f}' \leftarrow \mathit{ignoreBlocks}(\mathbf{f}, \mathbf{H}, c_{\text{new}})$

10: $< I_1^*, I_2^*, I_3^*, \cdots > \leftarrow \mathit{calculateSimilarity}(\mathbf{F}, \mathbf{f}', \mathbf{H}, c_{\text{new}})$

11: Show retrieved images $< I_1^*, I_2^*, I_3^*, \cdots >$

3.3.2 Analysis of Results

The IRMA 2009 dataset is utilized for this second method for retrieval purpose. The classification error score measurement is explained in the first method previously mentioned. The classification error is calculated using the same Equation 3.1 defined for the ImageCLEF competition [157]. However, the error has been measured based on 2005 IRMA codes. According to 2005 IRMA codes, there are 57 classes. These classes are more general than 2008 classes, since 2008 IRMA codes classify same images into 193 categories. It means that 2008 IRMA codes are designed to show detailed classes. For this reason, while 2005 IRMA codes have been used to classify images into general classes, 2008 IRMA codes will be utilized for retrieval evaluation.

Results

First of all, the images are classified with multi-class SVM using LBP histogram features. We have tried to extract LBP features from the entire image (not blocks), but there is a large difference in classification accuracy ($\approx$ 50% decrease in accuracy). Analogous to global versus local thresholding, it appears that calculating LBP histograms for image blocks is more capable of capturing the spatial characteristics of the image compared to extracting only one LBP histogram for the entire image. As Table 3.2 illustrates, the LBP-SVM approach achieves the lowest error score comparing ImageCLEF results based on 2005 IRMA code, and hence the highest accuracy is reached while implementing 4×4 blocks (image divided into 16 regions).

Table 3.2: SVM accuracy with LBP features (59 dimensions for each block) for classification of 2005 IRMA image dataset containing 14,270 images constituting 57 image categories.

Method	Accuracy	Error score
The Proposed with 4×4 blocks	75.50%	116.77
The Proposed with 5×5 blocks	74.31%	121.34
The Proposed with 6×6 blocks	72.23%	133.90
TAU [157]	–	356
Idiap [157]	–	393

An example of this method implementation on figure is shown in Figure 3.9. Regardless of any classes, the autoencoders and error concept applications are illustrated in this figure. In the retrieval process, image blocks are autoencoded using the restricted Boltzmann machine (RBM) function. For fixed n inputs and outputs, different p values are tried.

However, the condition of $p < n$ is maintained, the error levels are not considerably affected from changes of p value. The number of iterations for the autoencoder is set to 5, and more iterations does not seem to change the results. As a similarity measurement of two vectors which represents two images, the Pearson distance metric is used. The cases with no reduction and different reduction levels are compared, such as 1/8, 1/4 and 1/2 corresponding to 12.5%, 25% and 50% image area reduction, respectively. The precision and recall are calculated to evaluate retrieval when the top 10, 20 and 30 images are in focus. Moreover, retrieval times are recorded for each case. Table 3.3 provides the averages of 100 runs in each classes for different settings. According to the Table 3.3, 6×6 takes longest time in each case. Precision and recall do not significantly change while block numbers are changing within the same reduction rate. In general, when the reduction rate increases, the precision and recall slightly decrease. Tables 3.4 and 3.5 summarize the rates of decreases in terms of precision, recall and time by evaluating the top 20 retrieval hits. It is obvious that for finer grid structures (6×6 blocks) the time savings of greater than 27% can be achieved where 50% of the image blocks have been ignored resulting in 50% reduction of the feature vector size. This becomes a significant result when it is observed that a slight decrease in accuracy (both precision and recall) less than 1% for the top 20 hits of the retrieval.

3.4 Summary and Analysis

Searching for similar images in large medical image archives is both necessary and challenging. The actual retrieval of similar images may need more computational resources through more costly with one-by-one comparisons. This becomes a serious task for medical imaging dealing with big image data. There are many methods to reduce the dimensionality

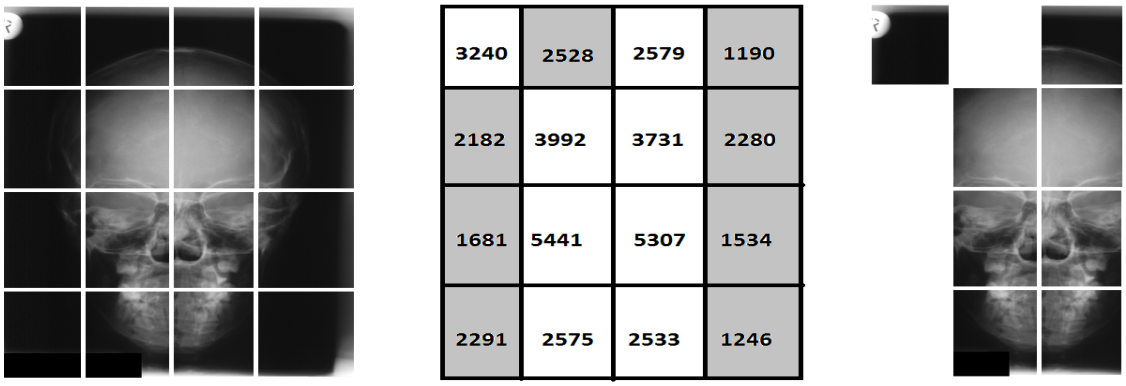

Figure 3.9: Illustration of the proposed approach: An example image from the IRMA dataset (left) is divided into 16 blocks and image blocks are autoencoded in an $n/p/n$ architecture. The error of each block is the recorded matrix (middle). A desired reduction (e.g. $50\% = 8$ blocks) is applied to exclude a number of blocks (gray blocks) from feature extraction. Gray blocks are eliminated because they have low error to generate the final results (right) which can be used to generate a shorter feature vector.

Table 3.3: The average precision P, recall R and time t are measured for top 10, top 20 and top 30 retrieval for randomly selected image from 57 classes and repeated 100 times. As distance metric, cross-correlation was used.

Blocks	Reduction	$P_{\text{Top 10}}$	$R_{\text{Top 10}}$	$P_{\text{Top 20}}$	$R_{\text{Top 20}}$	$P_{\text{Top 30}}$	$R_{\text{Top 30}}$	t(sec)
4×4	0	0.872	0.1807	0.868	0.2527	0.882	0.3214	0.01807
5×5	0	0.874	0.1783	0.873	0.2538	0.883	0.3222	0.02154
6×6	0	0.871	0.1792	0.877	0.2542	0.882	0.3220	0.02620
4×4	1/8	0.862	0.1761	0.866	0.2485	0.878	0.3199	0.01684
5×5	$\approx$1/8	0.869	0.1783	0.872	0.2531	0.880	0.3212	0.02005
6×6	$\approx$1/8	0.867	0.1786	0.873	0.2519	0.882	0.3221	0.02388
4×4	1/4	0.863	0.1762	0.862	0.2478	0.877	0.3205	0.01592
5×5	$\approx$1/4	0.870	0.1788	0.870	0.2497	0.879	0.3214	0.01833
6×6	1/4	0.869	0.1775	0.871	0.2513	0.879	0.3195	0.02182
4×4	1/2	0.856	0.1745	0.858	0.2458	0.865	0.3147	0.01396
5×5	$\approx$1/2	0.864	0.1762	0.862	0.2491	0.875	0.3185	0.01591
6×6	1/2	0.868	0.1790	0.870	0.2519	0.878	0.3212	0.01905

of image classification and retrieval tasks. In this chapter, retrieval methods were tested on 14,410 medical X-ray images using different approaches in order to reduce time, data storage needs and computational complexity. Motivated by the fact that in medical image analysis usually a certain region of interest (ROI) is of interest for user evaluation. Firstly, the proposed saliency-based folding demonstrated that relevant regions do not lose their importance, if they are summed with irrelevant regions. Summation is applied via column-wise or row-wise folding and relevant regions are defined by a context-aware saliency tem-

Table 3.4: Decrease in precision and recall and gain in time (for top 20 hits) compared to no reduction case where the LBP of all image blocks is calculated and used for retrieval comparisons.

	12.5% reduction			25% reduction			50% reduction		
Blocks	Precision	Recall	Time	Pre.	Rec.	Time	Pre.	Rec.	Time
4×4	0.20%	1.6%	6.8%	0.7%	1.9%	11.8%	1.2%	2.7%	22.7%
5×5	0.11%	0.2%	6.9%	0.3%	1.2%	14.9%	1.3%	1.8%	26.1%
6×6	0.40%	0.2%	8.8%	0.7%	1.1%	16.7%	0.8%	0.9%	27.2%

plate to find relevant regions in an image. This way, the image dimension is decreased, making feature extraction easier. In addition, classification runtime and computation complexity are reduced. Unlike folding, the autoencoding method eliminates irrelevant regions from the image. In other words, irrelevant regions are permanently deleted using autoencoder error histograms. Autoencoders have been investigated in the past with respect to their compression capabilities. In this work, a different approach for data reduction via autoencoder was designed. Feature elimination leads to reduction of both memory requirements and computational expense of the retrieval task. To decide which image blocks are rather irrelevant for the retrieval process, a $n/p/n$ autoencoder ($p < n$) was trained with the image blocks. The autoencoding errors are sorted in a histogram for all image classes. This histogram is then thresholded to exclude a certain percentage of the image area corresponding to low autoencoding error. Although experiments on the IRMA dataset illustrate that there is a slight decrease in precision and recall for the top 20 hits, the space requirements for the annotated feature vectors is reduced where simultaneously, the speed of the retrieval is increased.

Table 3.5: Decrease in precision and recall and gain in time (for top 10 hits) compared to no reduction case, where the LBP of all image blocks is calculated and used for retrieval comparisons.

	12.5% reduction			25% reduction			50% reduction		
Blocks	Precision	Recall	Time	Pre.	Rec.	Time	Pre.	Rec.	Time
4×4	1.1%	2.5%	6.8%	1.2%	2.5%	11.8%	1.8%	3.4%	22.7%
5×5	0.57%	0%	6.9%	0.46%	2.0%	14.9%	1.2%	1.2%	26.1%
6×6	0.46%	0.3%	8.8%	0.3%	0.9%	16.7%	0.34%	0.1%	27.2%

Chapter 4

Summary and Conclusions

Content-based Image Retrieval (CBIR) has become one of the most active research fields because of the development of digital technology and the rapid increase of imaging applications in medicine. An image is described by visual features, such as color, texture, shape, space etc. The critical issue in CBIR is how to construct effective methods to describe the image content, how to find images in a short time, and how to store them in a compact way to save storage space. The retrieval speed is an important issue in image retrieval systems as modern image archives usually have to store big image data generally meaning thousands of terabytes of data. If images are searched one by one, the computation usually becomes impractically long. However, if all images in a database are classified, similar images will be organized around one center or cluster. In order to improved such indexing structures as commonly used in recent CBIR systems, this thesis proposed time and storage reduction algorithms for both classification and retrieval of medical X-ray images. The IRMA 2009 medical dataset, which includes 14,410 X-ray images, was utilized to validate the performance of proposed algorithms.

The **first method** is based on statistical classification method. Classifier-based similarity measurement retrieves similar images by using a classifier according to ground-truth labels. SVM is used as a statistical classifier in this method. In the first step of this method, salient regions are detected with context-aware saliency algorithm [158] and a saliency template is created with the average of all saliency maps. Afterward, applying the saliency template to each image, relevant and irrelevant regions are identified. Then images are divided into several blocks. Using relevant regions, column- or row-wise folding is applied by summation of aligned image columns or rows. Features are extracted from the folded image to classify the query with SVM. The **second method** is a retrieval method consisting of classification and distance metrics. This method uses LBP features and SVM without folding. As class labels, this method used 2005 IRMA codes. While the first method classifies images into 193 classes, this code structure has 57 classes for the same images. After the classification, some blocks are deleted according to autoencoders decoding errors. Autoencoder neural networks were utilized to find relevant blocks by examining their decoding error. An $n/p/n$ autoencoders was applied when $p < n$. The main hypothesis behind this method is that if the autoencoder error gives large decoding error, this means that the input region contains complex or relevant structures. With respect to this hypothesis, an error histogram is generated for all classes. Subsequently, the irrelevant regions (blocks) that have the lowest error are eliminated. LBP features of remaining blocks can be compared with cross correlation (Pearson distance) similarity measurement to retrieve similar images within a given class. Classification accuracy was measured based on ImageCLEF error classification function. Retrieval results have been evaluated according to precision and recall measurements.

Comparing results of this SVM classification method on folded images with SVM classification without folding, the folding method speeds up the classification and reduces the

feature dimensions without significant decrease in accuracy. The second method results show that image retrieval time and storage needs decrease. Both methods aim to reduce run time and storage requirements for efficient medical image retrieval. Whether using the folding or autoencoding method, it seems that irrelevant regions do not contain effective information for retrieval. Finding these irrelevant regions can help to reduce feature dimensionality via image area reduction to reduce both computation time and storage needs. These two methods may be improved by using different texture or visual features instead of LBP. As future work, the folding method can be implemented by applying block-wise folding. If a global elimination function could be designed using only an autoencoder (without classification step), it would be a more effective method that cannot be affected by the performance of classification. Moreover, other feature dimension reduction methods, such as LDA and PCA, can be compared to these methods with respect to accuracy, run time and storage requirements.

References

[1] F. Long, H. Zhang, and D. D. Feng. Fundamentals of content-based image retrieval. In *Multimedia Information Retrieval*, 2002.

[2] Y.K. Chan and C.Y. Chen. Image retrieval system based on color complexity and color spatial features. *J. of Systems and Software*, 71(1):65–70, 2004.

[3] A. Gupta and R. Jain. Visual information retrieval. *Commun. ACM*, 40(5):70–79, May 1997.

[4] A.A.Salamah. Efficient content based image retrieval, 2010.

[5] D. Zhang. Improving image retrieval performance by using both color and texture features. *In Proc. of IEEE 3rd International Conference on Image and Graphics*, 18(20), December 2004.

[6] J. Liu, S. Zhang, W. Liu, X. Zhang, and D. N. Metaxas. Scalable mammogram retrieval using anchor graph hashing. In *Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on*, pages 898–901. IEEE, 2014.

[7] Y. Deng and B.S.Manjunath. Unsupervised segmentation of color-texture regions in images and video. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(8):800–810, August 2001.

[8] J. Z. Wang, J. Li, and G. Wiederhold. Simplicity: Semantics-sensitive integrated matching for picture libraries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:947–963, 2001.

[9] W. Y. Ma and B. S. Manjunath. Netra: A toolbox for navigating large image databases. In *Proceedings of the 1997 International Conference on Image Processing (ICIP '97) 3-Volume Set-Volume 1 - Volume 1*, ICIP '97, pages 568–, Washington, DC, USA, 1997. IEEE Computer Society.

[10] C. Carson, M. Thomas, S. Belongie, J. M. Hellerstein, and J. Malik. Blobworld: A system for region-based image indexing and retrieval. In *In Third International Conference on Visual Information Systems*, pages 509–516, 1999.

[11] N. Apostol, R. Rajeev, and S. Kyuseok. Walrus: A similarity retrieval algorithm for image databases. In *Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data*, SIGMOD '99, pages 395–406, New York, NY, USA, 1999.

[12] C. R. Venugopal and M. V. Sudhamani. Image retrieval from databases: an approach using region color and indexing technique. In Mislav Grgic, Naphtali Rishe, and Laurence Tianruo Yang, editors, *HPCNCS*, 2008.

[13] S. Nandagopalan, Dr. B. S. Adiga, and N. Deepak. A universal model for content-based image retrieval. 2(10):533 – 536, 2008.

[14] J. Li, J. Z. Wang, and G. Wiederhold. Irm: Integrated region matching for image retrieval. pages 147–156, 2000.

[15] R. Zhang and Z. Zhang. A clustering based approach to efficient image retrieval. In *14th IEEE International Conference on*, pages 339–346, 2002.

[16] U. Sinha and H. Kangarloo. Principal component analysis for content-based image retrieval. 22:1271–1289.

[17] C. Shyu Jennifer G. Dy L. S. Broderick C. E. Brodley, A. C. Kak and A. M. Aisen. Content-based retrieval from medical image databases: A synergy of human interaction, machine learning and computer vision. pages 760–767. AAAI / The MIT, 1999.

[18] J.S. Duncan C.C. Jaffe G.P. Robinson, H.D. Targare. Medical image collection indexing: shape-based retrieval using kd-trees. *Computerized Medical Imaging and Graphics*, 20(4):209–217, 1996.

[19] E. G. M. Petrakis and C. Faloutsos. Similarity Searching in Medical Image Databases. *IEEE Transactions on Knowledge and Data Engineering*, 9:435–447, 1997.

[20] R. Hanka and T. P. Harte. Curse of dimensionality: Classifying large multidimensional images with neural networks, 1997.

[21] Y. Liu and F. Dellaert. Classification driven medical image retrieval. In *Image Understandard Workshop. DARPA*, 1998.

[22] S. Berretti, A. Del Bimbo, and P. Pala. Content based retrieval of 3d cellular structures. In *Proc. of IEEE International Conference on Multimedia & Expo (ICME)*, Tokyo, Japan, August 2001.

[23] F. Hichem, C. Joshua, and A. C. Ben. Fusion of multi-modal features for efficient content-based image retrieval. In *FUZZ-IEEE*, pages 1992–1998. IEEE, 2008.

[24] A. MacFarlane and A. Tuson. Local search: A guide for the information retrieval practitioner. *Information Processing and Management*, 45(1):159 – 174, 2009.

[25] R. Thurmayr A. Horsch. How to identify and assess tasks and challenges of medical image processing. In *Proceedings of the Medical Informatics Europe Conference, St.Malo, France*, 2003.

[26] M. Worring A. W. M. Smeulders, A. Gupta S. Santini, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(12):1349–1380, December 2000.

[27] M. J. Fonseca and J. A. Jorge. Indexing high-dimensional data for content-based retrieval in large databases. In *DASFAA*, pages 267–274. IEEE Computer Society, 2003.

[28] W. Niblack J. Ashley Q. Huang B. Dom M. Gorkani J. Hafner D. Lee D. Petkovic M. Flickner, H. Sawhney and D. Steele. Query by image and video content: The qbic system. *Computer*, 28(9):23–32, September 1995.

[29] V. E. Ogle and M. Stonebraker. Chabot: Retrieval from a relational database of images, 1995.

[30] J.R. Bach, J.R. Fuller, G.Amarnath, H. Arun, H. Bradley, H. Rich, J. Ramesh, and S. Chiao-Fe. In *Storage and Retrieval for Image and Video Databases (SPIE)*, pages 76–87, 1996.

[31] Sougata Mukherjea, Kyoji Hirata, and Yoshinori Hara. Amore: A world wide web image retrieval engine. *World Wide Web*, 2(3):115–132, 1999.

[32] A. Pentland, R. W. Picard, and S. Sclaroff. Photobook: Content-based manipulation of image databases. *Int. J. Comput. Vision*, 18(3):233–254, June 1996.

[33] D. P. Huijsmans and N. Sebe. How to complete performance graphs in content-based image retrieval: add generality and normalize scope. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(2):245–251, 2005.

[34] J. Z. Wang, G. Wiederhold, O. Firschein, and S. X. Wei. Content-Based Image Indexing and Searching Using Daubechies' Wavelets. *Int. J. on Digital Libraries*, 1(4):311–328, 1997.

[35] I. J. Cox, M. L. Miller, S. M. Omohundro, and P.N. Yianilos. Pichunter: Bayesian relevance feedback for image retrieval. Vienna, Austria, 1996.

[36] R. Currell, C. Urquhart, P. Wainwright, and R. Lewis. Telemedicine versus face to face patient care: effects on professional practice and health care outcomes. 97(35):35+, August 2001.

[37] C. Shyu, C. Brodley, A. Kak, A. Kosaka, A. M. Aisen, and L. S. Broderick. Assert: A physician-in-the-loop content-based retrieval system for hrct image databases, 1999.

[38] C. Li, C. Wei, C. Wei, C. Li (corresponding, and Roland Wilson. A content-based approach to medical image database retrieval. In *in: Z. Ma (Ed.), Database Modeling for Industrial Data Management: Emerging Technologies and Applications, Idea Group Publishing*, page 291, 2005.

[39] P. Ghosh, S. Antani, L. R.Long , and G. R.Thoma. Review of medical image retrieval systems and future directions. In *CBMS*, pages 1–6. IEEE, 2011.

[40] Z. Xue, L. R. Long, S. Antani, J. Jeronimo, and G. R. Thoma. A web-accessible content-based cervicographic image retrieval system, 2008.

[41] W. Hsu, L. R. Long, and S. Antani. Spirs: A framework for content-based image retrieval from large biomedical databases. In Klaus A. Kuhn, James R. Warren, and Tze-Yun Leong, editors, *MedInfo*, volume 129 of *Studies in Health Technology and Informatics*, pages 188–192, 2007.

[42] M. Shyu, S. Chen , M. Chen , C. Zhang , and K. Sarinnapakorn . Image database retrieval utilizing affinity relationships. In *Proceedings of the 1st ACM International Workshop on Multimedia Databases*, MMDB '03, pages 78–85, New York, NY, USA, 2003. ACM.

[43] Y. Sun and S. Ozawa. Hirbir: A hierarchical approach to region-based image retrieval. *Multimedia Syst.*, 10(6):559–569, 2005.

[44] B. Ko, S. Y. Kwak, and H. Byun. Svm-based salient region(s) extraction method for image retrieval. In *ICPR (2)*, pages 977–980, 2004.

[45] J. Fan, Y. Gao, H. Luo, and R. Jain. Mining multilevel image semantics via hierarchical classification. *IEEE Transactions on Multimedia*, 10(2):167–187, 2008.

[46] D. A. Tran, S. R. Pamidimukkala, and K. Vu. Relevance-feedback image retrieval based on multiple-instance learning. In *Proceedings of IEEE/ACIS Conference on Computer and Information Science (ICIS 2008)*, Portland, Oregon, May 2008. IEEE Press.

[47] A. M. Marshall and S. Gunasekaran. Image retrieval-a review. In *International Journal of Engineering Research and Technology*, volume 3. ESRSA Publications, 2014.

[48] C. Fellbaum, editor. *WordNet: an electronic lexical database*. MIT Press, 1998.

[49] R. Schettini, G. CIOCCA, and S. Zuffi. A survey of methods for colour image indexing and retrieval in image databases. In *IN COLOR IMAGING SCIENCE: EXPLOITING DIGITAL*, pages 9–1. Media, John Wiley, 2001.

[50] G. A. Miller. Wordnet: A lexical database for english. *Commun. ACM*, 38(11):39–41, November 1995.

[51] G. Pass and R. Zabih. Histogram refinement for content-based image retrieval. In *Proceedings of the 3rd IEEE Workshop on Applications of Computer Vision (WACV '96)*, WACV '96, pages 96–, Washington, DC, USA, 1996. IEEE Computer Society.

[52] J. Huang, S.R. Kumar, M. Mitra, W.J. Zhu, and R. Zabih. Image indexing using color correlograms. In *Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97)*, CVPR '97, pages 762–, Washington, DC, USA, 1997.

[53] P. Salembier and T.Sikora. *Introduction to MPEG-7: Multimedia Content Description Interface*. John Wiley & Sons, Inc., New York, NY, USA, 2002.

[54] G. Gagaudakis and P.L. Rosin. Incorporating shape into histograms for CBIR. *Pattern Recognition*, 35(1):81–91, 2002.

[55] G. Pass, R. Zabih, and J. Miller. Comparing images using color coherence vectors. In *The Fourth ACM International Conference on Multimedia*, pages 65–73, 1996.

[56] K. Park, J. Jeong, and D. Lee. Olybia: Ontology-based automatic image annotation system using semantic inference rules. In Kotagiri Ramamohanarao, P. Radha Krishna, Mukesh K. Mohania, and Ekawit Nantajeewarawat, editors, *DASFAA*, volume 4443 of *Lecture Notes in Computer Science*, pages 485–496, 2007.

[57] Y. Liu, D. Zhang, and G. Lu. Region-based image retrieval with high-level semantics using decision tree learning. *Pattern Recogn.*, 41(8):2554–2570, August 2008.

[58] R. C. Gonzalez and R. E. Woods. *Digital Image Processing (3rd Edition)*. Upper Saddle River, NJ, USA, 2006.

[59] H. Tamura, S. Mori, and T. Yamawaki. Texture features corresponding to visual perception. *IEEE Transactions on Systems, Man and Cybernetics*, 8(6), 1978.

[60] S. B. Park, J.W. Lee, and S. Kim. Content-based image classification using a neural network. *Pattern Recognition Letters*, 25(3):287–300, 2004.

[61] A. K. Jain, R. P. W. Duin, and J. Mao. Statistical pattern recognition: A review. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(1):4–37, January 2000.

[62] T. Ojala, M. Pietikäinen, and T. Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(7):971–987, 2002.

[63] H. Müller, P. Clough, T. Deselaers, and B. Caputo. *ImageCLEF: Experimental Evaluation in Visual Information Retrieval*. 2010.

[64] F. Liu and R. W. Picard. Periodicity, directionality, and randomness: Wold features for image modeling and retrieval. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 18:722–733, 1996.

[65] B. B. Chaudhuri and N. Sarkar. Texture segmentation using fractal dimension. *IEEE Trans. Pattern Anal. Mach. Intell.*, 17(1):72–77, January 1995.

[66] N. Hervé and N. Boujemaa. Image annotation: which approach for realistic databases? In *CIVR '07: Proceedings of the 6th ACM international conference on Image and video retrieval*, pages 170–177, New York, NY, USA, 2007.

[67] Z. Lu and H. Burkhardt. A content-based image retrieval scheme in jpeg compressed domain, 2005.

[68] J. Fan, Y. Gao, H. Luo, and G. Xu. Automatic image annotation by using concept-sensitive salient objects for image content representation. In *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '04, pages 361–368, New York, NY, USA, 2004.

[69] I. J.Sumana, M .M. Islam, D. Zhang, and G. Lu. Content based image retrieval using curvelet transform. In *Multimedia Signal Processing, 2008 IEEE 10th Workshop on*, pages 11–16, Oct 2008.

[70] J. Starck, E.J. Candes, and D. L. Donoho. The curvelet transform for image denoising. *IEEE TRANSACTIONS ON IMAGE PROCESSING*, 11(6):670–684, 2002.

[71] M. M. Islam, D. Zhang, and G. Lu. Region based color image retrieval using curvelet transform. In Hongbin Zha, Rin-ichiro Taniguchi, and Stephen J. Maybank, editors, *Computer Vision - ACCV 2009, 9th Asian Conference on Computer Vision, Xi'an, China, September 23-27, 2009, Revised Selected Papers, Part II*, volume 5995 of *Lecture Notes in Computer Science*, pages 448–457, 2009.

[72] J. Jeon, V. Lavrenko, and R. Manmatha. Automatic image annotation and retrieval using cross-media relevance models. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval*, SIGIR '03, pages 119–126, New York, NY, USA, 2003.

[73] A. Yavlinsky, E. Schofield, and S. Rüger. Automated image annotation using global features and robust nonparametric density estimation. In *Proceedings of the 4th International Conference on Image and Video Retrieval*, CIVR'05, pages 507–517, Berlin, Heidelberg, 2005.

[74] Y. Rubner, C. Tomasi , and L. J. Guibas. The earth mover's distance as a metric for image retrieval. *Int. J. Comput. Vision*, 40(2):99–121, November 2000.

[75] A. Kumar, J. Kim, L. Wen, M. Fulham, and D. Feng. A graph-based approach for the retrieval of multi-modality medical images. *Medical Image Analysis*, 18(2):330 – 342, 2014.

[76] E.G.M. Petrakis, C. Faloutsos, and K.-I. Lin. Imagemap: an image indexing method based on spatial similarity. *Knowledge and Data Engineering, IEEE Transactions on*, 14(5):979–987, Sep 2002.

[77] D. Keysers, J. Dahmen, H. Ney, B. Berthold, and T. M. Lehmann. Statistical framework for model-based image retrieval in medical applications. *Journal of Electronic Imaging*, 12(1):59–68, 2003.

[78] H. Bunke, C. Irniger, and M. Neuhaus. Graph matching - challenges and potential solutions. In Fabio Roli and Sergio Vitulano, editors, *ICIAP*, volume 3617 of *Lecture Notes in Computer Science*, pages 1–10, 2005.

[79] M. Siadat, H. Soltanian-Zadeh, F. Fotouhi, and K. Elisevich. Content-based image database system for epilepsy. *Computer Methods and Programs in Biomedicine*, 79(3):209 – 226, 2005.

[80] P. Domingues, J. Silva, T. Ribeiro, N. M. M. Rodrigues, M. B. De Carvalho, and S. M. M. De Faria. Optimizing memory usage and accesses on cuda-based recurrent pattern

matching image compression. In *Computational Science and Its Applications–ICCSA 2014*, pages 560–575. Springer, 2014.

[81] K. R. Rao and P. Yip. *Discrete cosine transform: algorithms, advantages, applications*. Academic press, 2014.

[82] N. C. Francisco, N.M.M. Rodrigues., E. A. B. Da Silva, and S. M. M. De Faria. A generic post-deblocking filter for block based image compression algorithms. *Signal Processing: Image Communication*, 27(9):985–997, 2012.

[83] A. Wojnar and A. M. G. Pinheiro. Annotation of medical images using the surf descriptor. In *ISBI*, pages 130–133, 2012.

[84] T. M. Lehmann, M. O. Guld, C. Thies, B. Fischer, D. Keysers, M. Kohnen, H. Schubert, and B. B. Wein. Content-based image retrieval in medical applications for picture archiving and communication systems, 2003.

[85] B. Revet. *DICOM Cook Book for Implementations in Madalities*. Eindhoven, Netherlands, 1997.

[86] T Frankewitsch and U Prokosch. Navigation in medical internet image databases. *Medical informatics and the Internet in medicine*, 26(1):115, 2001.

[87] C. Shyu, C. Brodley, A. Kak, A. Kosaka, A. M. Aisen, and L. S. Broderick. Assert: A physician-in-the-loop content-based retrieval system for hrct image databases, 1999.

[88] T. Beyer D. W. Townsend. A combined pet/ct scanner: The path to true image fusion. *Br J Radiol*, 75(9), Jan 2002.

[89] H. Mller, N. Michoux, D. Bandon, and A Geissbuhler. A review of content-based image retrieval systems in medical applications – clinical benefits and future directions, 2004.

[90] B. K. Swann S. B. Siegel W. I. Jung M. S. Judenhofer, C. Catana and R.E. Nutt. Pet/mr images acquiredwith a compact mr compatible pet detector in a 7-t magnet. *Radiology*, 3(244), 2007.

[91] H. Winer-Muram C. E. Brodley A.C. Kak C. Pavlopoulou A.M. Aisen, L.S. Broderick. Automated storage and retrieval of thin section ct images to assist diagnosis: System description and preliminary assessment. *Radiology*, 1(228), 2003.

[92] C. Rodriguez J. Cui J. Xu A. Gupta S.A. Napel, C. F. Beaulieu. Automated retrieval of ct images of liver lesions on the basis of image similarity: Method and preliminary results. *Radiology*, 1(256), 2010.

[93] R. Haux A. Winter. A three-level graph-based model for the management of hospital information systems,. *Methods Information Med.*, (34), 1995.

[94] W.W. Chu, H. Chih-Cheng, A.F. Cardenas, and R.K. Taira. Knowledge-based image retrieval with spatial and temporal constructs. *Knowledge and Data Engineering, IEEE Transactions on*, 10(6):872–888, Nov 1998.

[95] M. M. Rahman, S. Antani, L. R. Long, D. Demner-Fushman, and G. R. Thoma. Multi-modal query expansion based on local analysis for medical image retrieval. In Barbara Caputo, Henning Mller, Tanveer Fathima Syeda-Mahmood, James S. Duncan, Fei Wang 0002, and Jayashree Kalpathy-Cramer, editors, *MCBR-CDS*, volume 5853 of *Lecture Notes in Computer Science*, pages 110–119, 2009.

[96] H. Mller, J. Kalpathy-Cramer, C. E. Kahn, and W. Hersh. Comparing the quality of accessing medical literature using content-based visual and textual information retrieval.

[97] A. Nvol, T. M. Deserno, S. J. Darmoni, M. O. Gld, and A.R. Aronson. Natural language processing versus content-based image analysis for medical document retrieval. *JASIST*, 60(1):123–134, 2009.

[98] Q. Gwnol, L. Mathieu, B. Lynda, C. Guy, R. Christian, and C. Batrice. Medical case retrieval from a committee of decision trees. *IEEE Transactions on Information Technology in Biomedicine*, 14(5):1227–1235, 2010.

[99] S. Antani, L. Rodney, L. George, and R. Thoma. A biomedical information system for combined content-based retrieval of spine x-ray images and associated text information. In *Proceedings of the Indian Conference on Computer Vision, Graphics, and Image Processing*, pages 16–18, 2002.

[100] Kwak DM. Content-based ultrasound image retrieval using a coarse to fine approach. In *Ann N Y Acad Sci*, 2002.

[101] D. Zhang and G. Lu. Shape based image retrieval using generic fourier descriptors. In *Signal Processing: Image Communication 17*, pages 825–848, 2002.

[102] M. Stricker and M. Orengo. Similarity of color images. pages 381–392, 1995.

[103] C. Faloutsos, R. Barber, M. Flickner, J. Hafner, W. Niblack, D. Petkovic, and W. Equitz. Efficient and effective querying by image content. *J. Intell. Inf. Syst.*, 3(3-4):231–262, July 1994.

[104] A. Mojsilovic and J. Gomes. Semantic based categorization, browsing and retrieval in medical image databases. In *Proceedings. International Conference on Image Processing*, pages III–145–III–148, 2002.

[105] A. N. Krishna and B. G. Prasad. Article: Automated image annotation for semantic indexing and retrieval of medical images. *International Journal of Computer Applications*, 55(3):26–33, October 2012. Full text available.

[106] J.-P. Valle A. Geissbuhler H. Mller, A. Rosset. Integrating content-based visual access methods into a medical case database. 2003.

[107] M.O. Gld, B. B. Wein, D. Keysers, C. Thies, M. Kohnen, H. Schubert, and T.M. Lehmann. A distributed architecture for content-based image retrieval in medical applications, 2002.

[108] W.E. Snyder H. Qi. Content-based image retrieval in pacs. *J. Digital Imag.*, 2(12):81–83, 1999.

[109] D. Meyer-Ebrecht T. Wringer, J. Stockhausen. Automatic coregistration, segmentation and classification for multimodal cytopathology. *Proceedings of the Medical Informatics Europe Conference*, 2003.

[110] K. Veropoulos. Image processing and neural computing used in the diagnosis of tuberculosis. *IET Conference Proceedings*, pages 8–8(1), January 1998.

[111] C. E. Brodley, A. C. Kak, C. Shyu, G. Jennifer, L. S. Broderick, and A. M. Aisen. Content-based retrieval from medical image databases: A synergy of human interaction, machine learning and computer vision. In Jim Hendler and Devika Subramanian, editors, *AAAI/IAAI*, pages 760–767, 1999.

[112] M. E. Mattie, L. H. Staib, E. Stratmann, H. D. Tagare, J. S. Duncan, and P. L. Miller. Research paper: Pathmaster: Content-based cell image retrieval using automated feature extraction. *JAMIA*, 7(4):404–415, 2000.

[113] A. Sameer, L. Rodney, Long G., and R. Thoma. A biomedical information system for combined content-based retrieval of spine x-ray images and associated text information. In *Proceedings of the Indian Conference on Computer Vision, Graphics, and Image Processing*, pages 16–18, 2002.

[114] O.-K. Yoon C.-H. Park J.-U. Won K.-H. Park D.-M. Kwak, B.-S. Kim. Content-based ultrasound image retrieval using a coarse to fine approach. *Annals New York Acad. Sci*, 2002.

[115] N. Kehtarnavaz S. Baeg. Classification of breast mass abnormalities using denseness and architectural distorsion. *Electronic Lett. Comput. Vis. Image Anal.*, pages 1–20, 2002.

[116] J. Z. Wang. Pathfinder: Region-based searching of pathology images using irm. *AMIA Symp*, 2000.

[117] P. Korn, N. Sidiropoulos, C. Faloutsos, E.Siegel, and Z. Protopapas. Fast and effective retrieval of medical tumor shapes. *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, 10(6):889–904, 1998.

[118] H. Kangarloo U. Sinha. Principal component analysis for content-based image retrieval. *RadioGraphics*, 22(5), 2002.

[119] L. H. Tang, R. Hanka, H. H. S. Ip, and R. Lam. Extraction of semantic features of histological images for content-based retrieval of images, 1999.

[120] A. Criminisi and J. Shotton. *Decision forests for computer vision and medical image analysis*. Springer Science & Business Media, 2013.

[121] W. R. Hersh, H. Mller, and J. Kalpathy-Cramer. The imageclefmed medical image retrieval task test collection. *J. Digital Imaging*, 22(6):648–655, 2009.

[122] K. Panizzi P. G. Anderson K. N. Jones, D. E. Woode. Peir digital library: On-line resources and authoring system. *Proceedings of the American Medical Informatics Association Symposium*, page 1075, 2001.

[123] R. L.Sameer, L. R. Long, S. K. Antani, and G. R. Thoma. Image informatics at a national research center. *Computerized Medical Imaging and Graphics*, 29:171–193, 2004.

[124] K. W. Clark, B. A. Vendt, K. Smith, J. B. Freymann, J. Kirby, P. Koppel, S. M. Moore, S. R. Phillips, D. R. Maffitt, M. Pringle, L. Tarbox, and F. Prior. The cancer imaging archive (tcia): Maintaining and operating a public information repository. *J. Digital Imaging*, 26(6):1045–1057, 2013.

[125] S. G. Armato III, G. McLennan, and D. Yankelevitz D. R. Aberle M. F. McNitt-Gray, C. R. Meyer. Lung image database consortium: Developing a resource for the medical imaging research community. In *Radiology*, volume 3, 2004.

[126] G. Langs, H. Mller, B. H. Menze, and A. Hanbury. Visceral: Towards large data in medical imaging - challenges and directions. In *MCBR-CDS MICCAI workshop*, volume 7723 of *Springer LNCS*, Nice, France, 2013.

[127] Md. M. Rahman, P. Bhattacharya, and B. C. Desai. A framework for medical image retrieval using machine learning and statistical similarity matching techniques with relevance feedback. *Trans. Info. Tech. Biomed.*, 11(1):58–69, January 2007.

[128] H. Mller, A. Geissbhler, J. Marty, C. Lovis, and P. Ruch. The use of medgift and easyir for imageclef 2005. In Carol Peters, Fredric C. Gey, Julio Gonzalo, Henning Mller, Gareth J. F. Jones, Michael Kluck, Bernardo Magnini, and Maarten de Rijke, editors, *CLEF*, volume 4022 of *Lecture Notes in Computer Science*, pages 724–732, 2005.

[129] T. Lin and H. Zha. Riemannian manifold learning. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(5):796–809, May 2008.

[130] C. B. Akgl, D.L. Rubin, S. Napel, C. F. Beaulieu, H. Greenspan, and B. Acar. Content-based image retrieval in radiology: Current status and future directions. *J. Digital Imaging*, pages 208–222, 2011.

[131] P. J. Phillips, H. Moon, P. Rauss, and S.A. Rizvi. The feret evaluation methodology for face-recognition algorithms. In *Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97)*, CVPR '97, pages 137–, 1997.

[132] L. Nanni and A. Lumini. Local binary patterns for a hybrid fingerprint matcher. *Pattern Recognition*, 41(11):3461 – 3466, 2008b.

[133] L. Nanni and A. Luminia. A reliable method for cell phenotype image classification. *Artificial Intelligence in Medicine*, 43(2):87–97, June 2008a.

[134] D. Unay and A. Ekin. Intensity versus texture for medical image search and retrival. In *ISBI*, pages 241–244. IEEE, 2008.

[135] L. Xavier, O. Arnau, J. Mart, and J. Freixenet. Dealing with false positive reduction in mammographic mass detection, 2007.

[136] C.J. C. Burges. A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov.*, 2(2):121–167, June 1998.

[137] WaooAkhilesh A. Patheja P.S. and Maurya J. P. An enhanced approach for content based image retrieval. *Research Journal of Recent Sciences*, 1:415–418, 2012.

[138] Z. K. Anuragjain S. Kumar. A review of content based image classification using machine learning approach. *International Journal of Advanced Computer Research*, 2:2277–7970, 2012.

[139] M. Arevalillo-Herráez, F. J. Ferri, and S. Moreno-Picot. Distance-based relevance feedback using a hybrid interactive genetic algorithm for image retrieval. *Appl. Soft Comput.*, 11(2):1782–1791, March 2011.

[140] K. S. Durgesh and LekhaBhambhu. Data classification using support vector machine. *Journal of Theoretical and Applied Information Technology*, 2009.

[141] J. Umamaheswari and Dr. G. Radhamani. Quadratic program optimization using support vector machine for ct brain image classification, 2012.

[142] C.M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., NJ, USA, 2006.

[143] H.J. Xing, J.G. Wu, and X.F. Chen. Modified adaboost based ocsvm ensemble for image retrieval. In *International Conference on Machine Learning and Cybernetics*, volume 3, pages 1048–1053, 2012.

[144] S.Goferman, L. Zelnik-Manor, and A. Tal. Context-aware saliency detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 34(10):1915–1926, October 2012.

[145] M. Cheng, G. Zhang, N.J. M., X. Huang, and Shi-Min Hu. Global contrast based salient region detection. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 409–416, June 2011.

[146] A. Tal R. Margolin and L. Zelnik-Manor. What makes a patch distinct? *The IEEE Conference on Computer Vision and Pattern Recognition*, 2013.

[147] A. Sharon and K. Pavel. Unsupervised detection of abnormalities in medical images using salient features, 2014.

[148] D.E. Rumelhart, G.E. Hinton, , and R.J. Williams. Learning internal representations by error propagation. *Parallel Distributed Processing*, 1, 1986.

[149] P. Baldi. Autoencoders, unsupervised learning, and deep architectures. *JMLR: Workshop and Conference Proceedings 27*, pages 37–50, 2012.

[150] P. Vincent, H. Larochelle, Y. Bengio, and P.A. Manzagol. Extracting and composing robust features with denoising autoencoders. In *International Conference on Machine Learning*, pages 1096–1103, 2008.

[151] Alex Krizhevsky and Geoffrey E. Hinton. Using very deep autoencoders for content-based image retrieval. In *ESANN*, 2011.

[152] C.C. Tan and C. Eswaran. Using autoencoders for mammogram compression. *J. Med. Syst.*, 35(1):49–58, 2011.

[153] Y. Bengio, A. Courville, and P. Vincent. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828, 2013.

[154] W. Wang, Y. Huang, Y. Wang, and L. Wang. Generalized autoencoder: A neural network framework for dimensionality reduction. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 496–503, 2014.

[155] H. Shin, M. R. Orton, D. J. Collins, S. J. Doran, and M. O. Leach. Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4d patient data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1930–1943, 2013.

[156] Y. Lu, L. Zhang, B. Wang, and J. Yang. Feature ensemble learning based on sparse autoencoders for image classification. In *International Joint Conference on Neural Networks*, pages 1739–1745, 2014.

[157] H. Müller, P. Clough, T. Deselaers, and B. Caputo. *Overview of the CLEF 2009 Medical Image Annotation Track*. 2010.

[158] S. Goferman, L. Zelnik-Manor, and A. Tal. Context-aware saliency detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 34(10):1915–1926, 2012.

[159] C. Chang and C. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at URL=http://www.csie.ntu.edu.tw/ cjlin/libsvm.