Radon Projections as Image Descriptors for Content-Based Retrieval of Medical Images

by

Aditya Sriram

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

Waterloo, Ontario, Canada, 2018

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the 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.

Statement of Contribution

This thesis includes two original papers that have been previously published/submitted for publication in peer reviewed journals, as follows:

Thesis Chapter	Publication Title	Publication Status	Contributions
3	Learning Autoencoded Radon Projections [100]	Published at IEEE Symposium Series on Computational Intelligence in 2017	Wrote the entire manuscript, coded the proposed framework, conducted experiments, and reflected on edits provided by co-authors.
4	Forming Local Intersections of Projections for Retrieval of Histopathology Images [101]	Submitted to European Conference on Computer Vision in 2018	Wrote the entire manuscript, coded and designed the proposed hand-crafted descriptor, conducted experiments, and reflected on edits provided by co-authors.

I certify that I have obtained a written permission from the copyright owners to include the above published materials in my thesis. I certify that the above material describes work completed during my registration as a graduate student at the University of Waterloo.

Abstract

Clinical analysis and medical diagnosis of diverse diseases adopt medical imaging techniques to empower specialists to perform their tasks by visualizing internal body organs and tissues for classifying and treating diseases at an early stage. Content-Based Image Retrieval (CBIR) systems are a set of computer vision techniques to retrieve similar images from a large database based on proper image representations. Particularly in radiology and histopathology, CBIR is a promising approach to effectively screen, understand, and retrieve images with similar level of semantic descriptions from a database of previously diagnosed cases to provide physicians with reliable assistance for diagnosis, treatment planning and research.

Over the past decade, the development of CBIR systems in medical imaging has expedited due to the increase in digitized modalities, an increase in computational efficiency (e.g., availability of GPUs), and progress in algorithm development in computer vision and artificial intelligence. Hence, medical specialists may use CBIR prototypes to query similar cases from a large image database based solely on the image content (and no text). Understanding the semantics of an image requires an expressive descriptor that has the ability to capture and to represent unique and invariant features of an image. Radon transform, one of the oldest techniques widely used in medical imaging, can capture the shape of organs in form of a one-dimensional histogram by projecting parallel rays through a two-dimensional object of concern at a specific angle. In this work, the Radon transform is re-designed to (i) extract features and (ii) generate a descriptor for content-based retrieval of medical images. Radon transform is applied to feed a deep neural network instead of raw images in order to improve the generalization of the network. Specifically, the framework is composed of providing Radon projections of an image to a deep autoencoder, from which the deepest layer is isolated and fed into a multi-layer perceptron for classification. This approach enables the network to (a) train much faster as the Radon projections are computationally inexpensive compared to raw input images, and (b) perform more accurately as Radon projections can make more pronounced and salient features to the network compared to raw images. This framework is validated on a publicly available radiography data set called "Image Retrieval in Medical Applications" (IRMA), consisting of 12,677 train and 1,733 test images, for which an classification accuracy of $\approx 82\%$ is achieved, outperforming all autoencoder strategies reported on the IRMA dataset. The classification accuracy is calculated by dividing the total IRMA error, a calculation outlined by the authors of the data set, with the total number of test images.

Finally, a compact handcrafted image descriptor based on Radon transform was designed in this work that is called "Forming Local Intersections of Projections" (FLIP).

The FLIP descriptor has been designed, through numerous experiments, for representing histopathology images. The FLIP descriptor is based on Radon transform wherein parallel projections are applied in a local 3×3 neighborhoods with 2 pixel overlap of gray-level images (staining of histopathology images is ignored). Using four equidistant projection directions in each window, the characteristics of the neighborhood is quantified by taking an element-wise minimum between each adjacent projection in each window. Thereafter, the FLIP histogram (descriptor) for each image is constructed. A multi-resolution FLIP (mFLIP) scheme is also proposed which is observed to outperform many state-of-the-art methods, among others deep features, when applied on the histopathology data set KIMIA Path24. Experiments show a total classification accuracy of $\approx 72\%$ using SVM classification, which surpasses the current benchmark of $\approx 66\%$ on the KIMIA Path24 data set.

Acknowledgements

Dedication

I would like to dedicate this work to my mother, father, and brother (Meera Sriram, Iyer Sriram, and Abhishek Sriram, respectively) who have paved the path before me upon whose shoulder I stand. Thank you for your love, care and unconditional support.

A special thanks to my wife, Akanksha, without whose never-failing support and encouragement has led me to believe that she might be a product of my imagination.

As for my source of inspiration and guidance, I would like to dedicate this thesis to my wonderful supervisors Dr. Hamid Tizhoosh and Dr. Shahryar Rahnamayan. My endeavor in research began with Dr. Rahnamayan back in 2015, and it has been my honor and pleasure to have worked with you thus far. Thank you, Dr. Rahnamayan, for teaching me the value of education.

To Dr. Tizhoosh, whose passion for teaching set a new standard for anyone involved in education, training and development, or any endeavor in which one human being seeks to support the growth and development for another. I will forever be thankful for our guidance, unconditional support, and trust. Dr. Tizhoosh is the funniest and most knowledgeable advisors who genuinely cares for his students. I hope that I could be as lively, enthusiastic, and energetic as Dr. Tizhoosh and that someday be able to command an audience as well as he can.

Finally, a special thanks to my manager at IBM, Deepen Manek, for supporting me during the developments of my work.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AE Autoencoder.

CBIR Content-Based Image Retrieval.

CNN Convolution Neural Network.

FLIP Forming Local Intersection of Projections.

HOG Histogram of Oriented Gradient.

IRMA Image Retrieval in Medical Applications.

KNN K-Nearest Neighbor.

LBP Local Binary Patterns.

mFLIP Multi-resolution Forming Local Intersection of Projections.

MLP Multi-Layer Perceptron.

SVM Support Vector Machine.

WSI Whole Slide Imaging.

Chapter 1

Introduction

The production of digital images has increased significantly over the years, both in terms of resolution and volume. Particularly in the medical imaging domain, a large number of digitized images are stored but are not computationally utilized to assist clinicians. To address this problem, Content-Based Image Retrieval (CBIR) is a computational tool to extract low-level features representative of the image for accurate and efficient retrieval purposes. These CBIR systems are capable of operating at a large scale; however, the accuracy of its retrieval depends heavily on the extracted features from images. The motivation of research in this field is to learn a medical data set such that CBIR systems can be adopted to automatically index images by extracting salient features using a sophisticated computer vision or artificial intelligence (AI) algorithm. These salient features should hold dominant characteristics of an image that are informant and non-redundant, such as: corners, edges, blobs, color, etc. Hence, CBIR systems use these generated features and aim is to retrieve an image closer to that of the query image using intelligence algorithms for both feature extraction and search.

An accurate CBIR system depends on the feature representation as well as on the similarity measure to find similar images. Although machine learning techniques have existed for decades, the recent deep learning algorithms have topped metric standards in a variety of applications. Recently, deep learning has become a popular approach for various applications. In reality, deep learning algorithm is a re-branding of neural networks from the 1950s, with multiple processing layers (called *hidden layers*) interconnecting neurons as several layers are employed. The claim of deep learning algorithms is to rely on raw data such that each hidden layer learns an "abstraction" of data (that is learning a feature set). The concatenation of the hidden layers enables the network to generalize the image based on all extracted features within every hidden layer. Although this approach is well suited for

certain applications, deep learning algorithms struggle when tackling high-dimensional data [81]. For instance, digital pathology is a major contributor in the medical imaging domain; the images of which can range from $50,000 \times 50,000$ to $100,000 \times 100,000$ pixels and higher. A common approach is to use many small sub-images after down-sampling; however, this causes loss of information making it all the more difficult to generalize properly. Hence, there is a need to develop compact feature extraction algorithm that can globally represent an image. This descriptor can be used to not only provide more manageable inputs for a network to analysis, but also significantly reducing resources and computational time.

Radon transform, introduced in 1917 [85], is a mathematically derived method that uses parallel projections for reconstructing scenes and objects. In the medical domain, Radon is used for tomography for creating images from projections taken at a cross-sectional plane of the patient. In terms of image processing, Radon transform converts a 2-dimensional image along a direction to a 1-dimensional signal, capturing the cumulation of pixel intensities along the direction, depicted in Fig. 1.1. Since the inverse Radon can reconstruct an image from a 1D Radon signal, the intuition is that Radon transform preserves features in the form of a 1D signal which are compact embeddings of an image. An advantage of Radon transform over deep learning is its familiarity within medical specialists, as it is used in various medical systems; hence, Radon is well accepted and understood in the medical domain. This work attempts to use Radon projections as image descriptors for two purposes: i) to study the effects of Radon when used in-place of raw images as an input parameter to a deep learning algorithm, and ii) to design a global descriptor using Radon projections and compare its performance and accuracy metrics to deep learning algorithms and other established global descriptors. Both these approaches will be evaluated on publicly available data sets. This thesis is hopeful to convey that Radon projections can contribute to construction of powerful image descriptors for medical imaging. To support this claim, Radon projections are used as a preliminary feature, an input parameter, for a proposed deep learning framework - validated on an publicly available radiography data set, called Image Retrieval in Medical Applications (IRMA) [1]. Also, a handcrafted global descriptor called Forming Local Intersection of Projections (FLIP) is designed and validated on the KIMIA Path24 [9] data set consisting of large histopathology images.

The first contribution of this thesis is a supervised framework which is composed of an Autoencoder (AE) for dimensionality reduction [41], and a Multi-Layer Perceptron (MLP) for classifying the AE features. This framework is developed to study if Radon transform can be used in place of raw images as input parameters. The purpose of using Radon transform is to study its effects as a feature extractor and to examine if there is a better alternative to raw images as inputs to a learning algorithm. This study aims to show, against popular belief, that deep networks are in-fact dependent on the quality of the in-
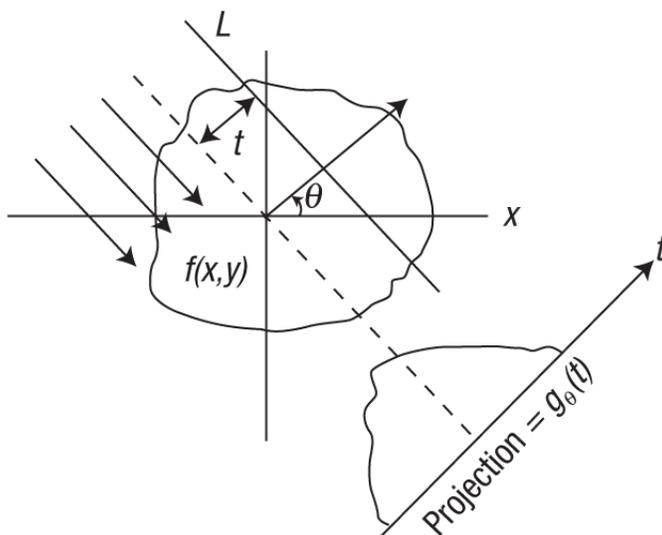
Figure 1.1: Computing Radon transform of an image [86].

put parameter and cannot just rely on raw data. There have been various studies that outlines the benefit of learning Radon features as compared to raw images when using deep learning algorithms, such as: [51], [66], and [103]. In addition, there have also been reported studies that confirm the rationale of using Radon features over Histogram of Oriented Gradient (HOG) features, such as: [11] and [59]. To validate this study, equi-distant Radon projections are applied to a raw image - these features are vectorized and provided as an input parameter to the proposed framework, evaluated on the IRMA (radiography) data set. For the proposed framework, the initial experimentation suggests that Radon transform outperforms both raw images as well as HOG features as input features. Similarly, in this study, Radon features are observed to be faster at computation and better in generalization, achieving a classification accuracy of 82% [100], outperforming the previous Autoencoder benchmark of 80% [103]. The accuracy measure is computed by dividing the IRMA error by the total number of test images.

Recently, handcrafted descriptors have been merely in the shadow due to the dominance of deep learning algorithms. In terms of literature contribution, it is evident that there is limited attention to classical computer vision techniques as most research emphasizes on applying deep learning for image classification and recognition in computer vision. Contrary to deep learning hype, descriptors have historical credentials when incorporating it into a CBIR system. Among many advantages, descriptors do not necessarily need to be "trained"; they can be computed on raw images to extract a representation of an image

and stored for quick retrieval. To that, the second novelty of this thesis is a handcrafted global descriptor called FLIP based on Radon transform. At a high-level, the FLIP is uses local 3×3 neighborhoods within which four equi-distant Radon projections are obtained. Thereafter, the intersections between the adjacent projections are calculated to obtain compound projections. Once calculated for all local neighborhoods, the projections are rescaled and counted to form a histogram. In addition, an advanced version of FLIP, called Multi-resolution Forming Local Intersection of Projections (mFLIP) is also proposed.

Since deep learning architectures have become dominant in computer vision, the intent for the FLIP and mFLIP is to compete with established deep architectures in contextually understanding histopathology scans. The purpose is to illustrate that hand-crafted descriptors can be a contender against deep learning architectures, offering minimum running cost and resources. At a high-level, FLIP and mFLIP is based on Radon transform captured within a small 3×3 neighborhood. Within each neighborhood, a relationship between the adjacent projections are computed based on logical "AND", and finally by rescaling all the projections, a histogram is obtained which is a compact representation of the entire image. The rationale to develop FLIP and mFLIP is to capture the textures that is translation-invariant, which is representative of the image yet light to store. In order to evaluate the descriptiveness of FLIP, the KIMIA Path24 [9] data set is adopted which contains anonymized histopathological patches extracted from 24 Whole Slide Imaging (WSI). In total, there are 27,000 training and 1,325 test labelled patches of size 1000×1000 pixels. The proposed descriptor surpasses deep learning approaches reported on the KIMIA Path24 data set. In fact, the mFLIP algorithm achieves a high overall accuracy (η_{total}) of $\approx 72\%$ when classifying using Support Vector Machine (SVM), and $\approx 60\%$ when retrieving using χ^2 distance. In comparison, the current benchmark is the hand-crafted ELP [106] approach which obtains an accuracy of $\approx 66\%$ with SVM classification. On the other hand, the best deep learning approach is TL-Inception-v3, which achieves a classification accuracy of $\approx 57\%$.

Chapter 2

Background on Image Descriptors

CBIR enables automatic indexing of extracted low-level image features; these salient features include edges, corners, color, and texture [67] [42]. In the medical domain, CBIR may be used for navigating, browsing, and querying-by-example such that similar images are obtained through similarity calculations between these extracted low-level image descriptors [29] [72]. One of the rationals for extracting features from images is, of course, dimensionality reduction, e.g., projecting high-dimensional points from images onto low-dimensional feature points, such that processing and retrieving images are computationally quicker, reducing overhead [49]. Conventionally, there are two types of features (also known as *descriptors*), namely local and global features.

"Local features" are used for recognizing and identifying an object, e.g., the ability to recognize a person in the scene/frame of concern [71]. Compact vector representations extracted from local neighborhoods within an image are used to form descriptors which contain distinctive patterns and structures such as edges, corners, and points. These features are identified due to immediate change in color, texture, or intensity across its surrounding image patches. A few well-known local descriptors include Scale-Invariant Feature Transform (SIFT) [70], Speeded Up Robust Features (SURF) [16], and Harris-Stephen corner detector [40]. For local features, distinction from its surrounding weighs more than the representation of feature itself [114]. Particularly, some medical images are highly textured and hold salient features for analyzing more than color and intensity values [36]. In general, local features are not widely used in medical domain as these images possess homogeneity properties which are ambiguous because of the imprecise nature of the images being captured [94]. To that, this literature review puts emphasis on global descriptors instead.

"Global descriptors", on the other hand, are features that are extracted to represent the entirety of the image, hence used for detecting and classifying an object in the scene/frame of concern [69]. Widely used in the medical domain, global descriptors are composed of locally extracted features that are generally concatenated together to form an image representation. Global descriptors generally excel for contour representations, and as shape descriptors and texture features. Fig. 2.1 provides a pictorial representation of global and local descriptors. Some notable examples of global hand-crafted features include: HOG [30], Local Binary Patterns (LBP) [77], Gabor filters [47], and Dual Tree Complex Wavelet Transform (DT-CWT) [89].
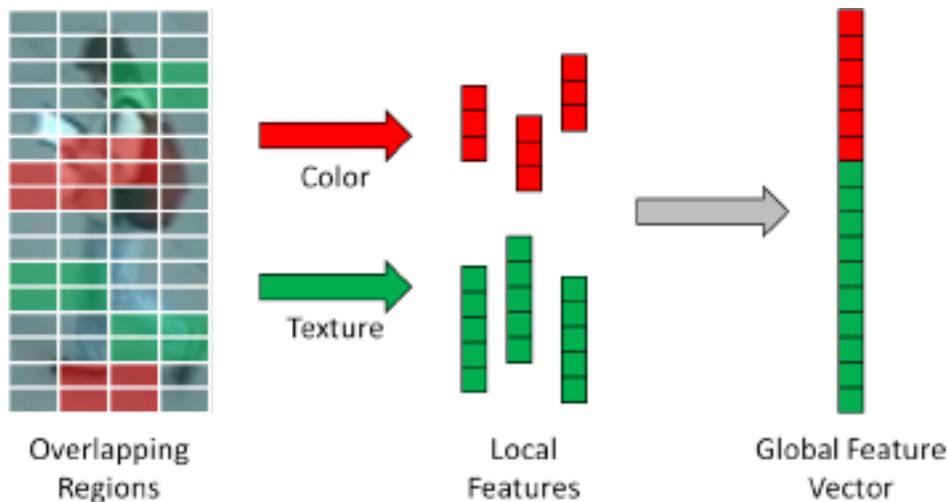
Figure 2.1: Difference between global and local descriptors [87].

Extracted features represent the image as a point in the high-dimensional feature space, upon which a similarity metric can be applied to compare images. Hence, the similarity between the descriptors quantifies the image content enabling the image search and retrieval.

In this work, the background literature will mainly focus on global descriptors as the proposed FLIP and mFLIP descriptors are both global descriptors and will be validated against other global descriptors. This section will provide a comprehensive literature review on global descriptors, automated feature extractors using deep learning algorithms, and more importantly Radon projections as the base for both FLIP and mFLIP.

2.1 Global Features

Images are a set pixel values representing color or gray-level intensity. In general, descriptor algorithms use these pixel values to calculate and to form feature vectors for classification and retrieval purposes. Renowned global descriptors, such as HOG and LBP, are algorithms that extract salient features and formulate them into a concise vector that represent the entirety of the image. Simply put, for global features, the entire object is represented by a single feature vector, as opposed to local features which are comprised of many feature vectors representing smaller parts of the object. In medical imaging, global features are extensively used because these algorithms are expected to capture the texture and spatial relationships of an image more adequately. This subsection will cover global feature extractors that are widely used in the computer vision and/or medical imaging domain, namely HOG and LBP.

2.1.1 Histogram of Oriented Gradients

HOG is one of the popular and successful global descriptors in image processing that debuted as a powerful algorithm for "person detection" and later for person tracking [122]. Developed by Dalal and Triggs [30] at the CVPR conference in 2005, HOG is a robust global descriptor which produces distinct features based on occurrences of edge orientations in local windows which are then formed into a histogram that represents the image at its entirety. Although HOG is not rotational invariant, the descriptor captures high frequency of gradient orientation making it invariant to geometric and photometric transformations. At a high-level, to compute the HOG descriptor, let's say an 8×8 cell is obtained within the image for which these cells are organized into overlapping blocks. Within the 8×8 cell, a total of 64 gradient vectors are computed and placed into a 9-bin histogram. Since the histogram ranges from 0 to 180 degrees, a 9-bin histogram means 20 degrees per bin. All histograms are then concatenated for the entire image to form the HOG descriptor, which is depicted in Fig. 2.2.

There have been further developments in applying HOG to applications other than pedestrains, cars, buses, animals, and bicycles. One such approach was to apply HOG descriptors for detecting humans in videos [31]. Later on in 2006, Zhu et al. in [124] presented a formulation of HOG in combination with cascading classifiers using AdaBoost algorithm which was reported to be 70% faster than the original algorithm.

A common strategy with global descriptors is to bind them with a powerful classification algorithm such as SVM. Although SVM was officially introduced by Boser et al. in 1992
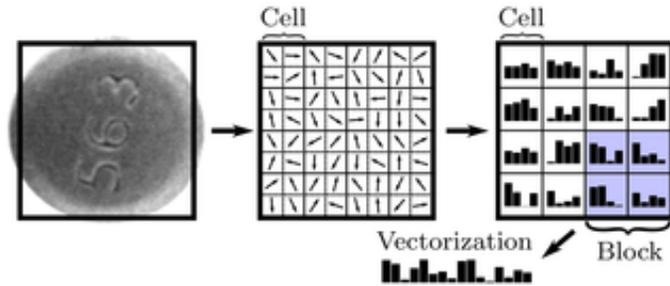
Figure 2.2: Computing HOG for an image [21].

[19], the classification algorithm was originally developed by Vapnik and Chervonenkis as a linear classifier, using the maximum-margin hyperplane algorithm, which dates back to 1963. In 1992, SVM was reformulated to analyze and categorize data that are both linear and non-linear in nature, achieved by what is called the "kernel trick". At a high-level, the kernel trick is to implicitly map data points into high-dimensional feature spaces to explore new patterns of data that is not achievable at the current input dimension [44]. The coupling of HOG and SVM is among the most popular approaches for classifying dense descriptors.

In [27], Christobel used HOG and SVM to detect and classify mammogram images. Particularly, the mammogram images were pre-processed using CLAHE (Contrast Limited Adaptive Histogram Equalization) method and the images were dilated to enhance the image quality. The HOG descriptor is computed on the enhanced images which achieved very good results for cancer detection by reducing false positive rates. The aforementioned approach yields an accuracy, sensitivity, and specificity of 90%, 86.36%, and 5.55%, respectively. Barbu in [15] adopted HOG and a non-linear SVM with quadratic kernel function, for cell localization. Specifically, the authors use 3×3 cell blocks of 6×6 pixel cells with 9 bins to obtain a HOG descriptor with 81 coefficients. To validate this approach, the authors have constructed an in-house dataset which contains various types of cells being composed of 29 positive and 29 negative cases. The approach achieves a 90% detection rate with recall and precision values of 0.9 each. Agarwal et al. in [5] apply HOG to automatically recognize and classify standard views of an echocardiogram, captured through a 2D ultrasound probe. Using the HOG along with SVM, the authors were able to classify standard views namely the Parasternal Long Axis (PLAX) and the Short Axis (SAX) B-mode echocardiograms. It is observed that HOG can effectively discriminate feature encodes by computing edges/gradients in the images, which yields a 98% accuracy when tested on a fairly large database of 703 echocardiogram images. In 2014, Pomponiu et al. [82] employed the HOG descriptor along with SVM for classifying and filtering the

8

mass and normal tissue regions. After testing on 1881 regions of interest (ROIs), the authors conclude that extensive simulations are to be conducted to illustrate the capacity of HOG descriptor to improve the performances of mass detection systems. More recently in 2015, Adetiba and Olugbara in [4] proposed an artificial neural network (ANN) ensemble utilizing HOG features and SVM for predicting lung cancer. The authors validated their method on the openly available dataset "Integrated Genomic Database of Non-Small Cell Lung Carcinoma" (IGDB.NSCLC), for which the ANN ensemble and HOG achieved an accuracy of 95.90% and mean square error of 0.0159. Song et al. in [95] proposed a descriptor based on Multi-coordinate HOG (MCHOG), validated on the publicly available Interstital Lung Disease (ILD) multi-class dataset, representing more than 150 disorders of the lung parenchyma. For all of the five tissue categories, the proposed algorithm achieved over an 80% true positive rate. Xu et al. in [121] exploited a multi-scale HOG for angle closure classification. The authors demonstrate that HOG features with higher dimensions outperform low dimensional clinical features in terms of angle closure classification accuracy. When validated on a clinical dataset of 2048 images, the proposed method achieves a AUC value (area under the curve) of 0.835±0.068, accuracy of 75.8%±6.4%, and a specificity of 85%, outperforming other classification approaches based on clinical features. Pons et al. in [83] adopted HOG configured with 10×10 cell to detect lesions in breast ultrasound images. Validated on 326 images from different patients (54 malignant lesions, 109 benign lesions, and 163 healthy breasts), HOG obtained a sensitivity of 86% with 0.28 false-positive detection per image when classifying breast lesions. When detecting malignant lesions, the HOG descriptor obtains an Az value (the area under the ROC curve) of 0.93 and a sensitivity of 78% at a 1.15 false-positive detections per image. Recently in 2017, Abdel-Nasser et al. [3] evaluated the performance of five texture methods with the proposed CAD system: gray level co-occurrence matrix features, local binary patterns, phase congruency-based local binary pattern, HOG and pattern lacunarity spectrum. In particular, the authors use a 3×3 cell size, 4×4 cells for the block size, and a 9-bit histogram yielding a feature vector that is 144 in length. Validated on a clinical database of 31 malignant and 28 benign breast ultrasound image sequences, from Cambridge University, HOG is observed outperform all other descriptors experimented by the authors. In particular, HOG achieved the highest AUC of 0.989 when classifying benign versus malignant tumors. In comparison, LBP yields the second-best AUC of 0.950. Finally, Shastria et al. in [105] introduced a descriptor called Discrimination Potentiality Histogram of Oriented Texture (DP-HOT) which is based on HOG and Gabor filter to classify mammogram patches as malignant/benign. When evaluated on the openly available IRMA dataset, the authors claim to have achieved more than 92% accuracy.

2.1.2 Local Binary Patterns

LBP were introduced by Ojala et al. [77] as an efficient multi-resolution descriptor for texture classification that is scale and translational invariant. However, there have been developments on rotational-invariant LBP [61], which is particularly useful for when the medical images captured at different orientations. However, since there are regulations for capturing images, in reality, rotational-invariant extractors are obscure. Designed as a particular case of texture spectrum model [75] [76], LBP is a powerful image descriptor that is used, among others, for facial recognition and expression classification [6]. Fig. 2.3 provides a pictorial illustration on computing LBP for a small neighborhood.
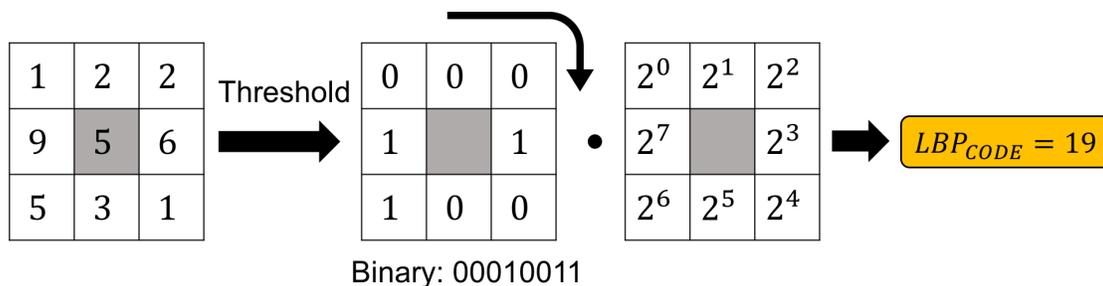
Figure 2.3: Computing LBP on an image neighborhood [20].

Image descriptors quantify image characteristics such as shape, color, texture, edges, or corners [56, 93, 102]. A good example are LBP [77] as a descriptor for texture classification which is, among others, rotation invariant [80]. Designed as a particular case of texture spectrum model [75, 76], LBP is a powerful image descriptor that has certainly set high accuracy standards in the medical domain, including for CT images [96, 97], x-ray [55], and digital pathology [9].

Liu et al. [68] utilized LBP histogram based on PCA as the local descriptor to encode both the retinal image and its edge map. The authors use a 2-class SVM classifier to identify the presence of normal macula and each of three types of macular pathologies, namely, macular edema, macular hole, and age-related macular degeneration, in the OCT slice centered at the fovea. The authors conducted experiments on a dataset of 326 OCT scans from 136 subjects. The results show that the proposed method was very effective (all AUC > 0.93).

Qureshi et al. [84] proposed a combined approach for meningioma subtype classification using subband texture (macro) features and micro-texture features. These are captured using the Adaptive Wavelet Packet Transform (ADWPT) and LBP, respectively. SVM

is adopted to classify the extracted features. Although LBP features did not provide higher overall classification accuracies than ADWPT, however it provided higher accuracy for a meningioma subtype that is difficult to classify otherwise. It has been shown that the ADWPT performs better than LBP in most of the feature selection and classifier combinations. The LBP features provide 59% to 73% classification accuracy on their own. Although the combined feature set does not improve the overall classification accuracy, it improves the classification accuracy of meningiotheliamatous, which is a difficult subtype to classify correctly, by 7%.

In [116], authors extract the center-symmetric LBP from image and compute co-occurrence of pixel pairs in local pattern map using gray-level co-occurrence matrix, which is assumed to be more robust than frequency of patterns (histogram). Dubey et al. [35] propose a new feature descriptor named local diagonal extrema pattern (LDEP) for CT image retrieval. By using the first order local diagonal derivatives, the values and indexes of the local diagonal extremes are obtained, and the descriptor is generated on the basis of the indexes and the relationship between the center pixel and local diagonal extremes. Greenspan et al. present a localized statistical framework based on Gaussian mixture modelling and Kullback-Leibler (KL) matching for medical image retrieval. It combines a continuous, probabilistic and region-based image representation scheme, along with an information-theoretic image matching measure to categorize and retrieve x-ray images [38]. Camlica et al. used two methods to retrieve x-ray images: an AE to reduce the size of the image for feature extraction [22], and a SVM classifier trained with LBP features derived from saliency based image regions [23].

It is worth noting that most current medical CBIR systems are limited to the images with specific modality, organ, or diagnostic studies, and usually cannot be applied directly to other medical applications [60]. Moreover, although CBIR can be applied to several medical cases, the semantic gap between the extracted image features and the doctors' interpretation of the images limits its application. This is because generic features of images may not be suitable for similarity matching for medical images [8]. Liao et al. [63] used dominant local binary patterns (DLBP) and a symmetric Gabor filter responses which is observed to achieve the highest accuracy in some of the texture databases, which include: Brodatz, Outex, Meastex, and many others. In particular, the DLBP captures the most frequent patterns to provide a texture description; whereas, the Gabor filter captures texture at a global level which complements the DLBP during retrieval.

There have been numerous approaches that validate the use of LBP in the medical domain. Nanni et al. [74] presents a descriptor called Elongated Quinary Pattern (EQP) which is inspired from LBP, which basis its analysis on an elliptic neighborhood and adopts quinary encoding for evaluating the local gray-scale difference. The authors evaluate this

11

approach on three bio-imaging data sets: the Infant COPE database of neonatal facial images, the 2D-HeLa dataset of fluorescence microscope images, the pap smear data sets of smear cells images. The EQP along with SVM classification is observed to outperform the traditional LBP approach. The authors claim that concatenating features to obtain a longer descriptor yields a better result as it exploits information from different neighborhood for when classifying. Sorensen et al. [97] set the benchmark in emphysema dataset using LBP along with K-Nearest Neighbor (KNN) for classification. Emphysema is a CT data set composed of specific regions of interest (ROI) images of lung, that is categorized into 3 classes. Specifically, a 95.2% classification accuracy was achieved with a pulmonary function test (PFT) achieving correlation coefficients of up to $|r| = 0.79$ in 39 subjects. Ko et al. [55] used local wavelet-centre symmetric LBP (WCS-LBP) along with random forest to classify x-ray images (2007 IRMA data set). The authors then use maximum posterior probability to retrieve the best result for every test image. Experimental results on the 2007 IRMA data set yield an average precision and recall performance of 93.10% and 89.43%, respectively. Liu et al. [68] use LBP to capture texture and shape of retinal OCT images along with their edge maps. The authors use 2-class Support Vector Machine classifiers to form a multi-scale spatial pyramid. The validation this approach is to identify the presence of normal macula and each of three types of macular pathologies, namely, macular edema, macular hole, and age-related macular degeneration, in the OCT slice centered at the fovea. The data set is composed of 326 OCT scans from 136 subjects, yielding a result of all AUC > 0.93. Sorensen et al. [96] used LBP along with KNN to classify different texture patterns in lung computed tomography. The proposed method is evaluated on a set of 168 regions of interest comprising normal tissue and different emphysema patterns, for which it achieves a classification accuracy of 95.2%. Unay et al. [111] designed an LBP as a texture operator for MRI brain image analysis. Since the LBP is rotational invariant, the LBP effectively applied to MR bias field yielding a robust result for even 40% intensity variations. In addition, the data set adopted contains images rotated up to 60deg in both clockwise and counterclockwise directions using three different interpolation methods. The author concluded that the results were reasonable, making LBP a promising texture descriptor for various MR brain image analysis applications. Unay et al. then built upon the previous idea by presenting a quicker search and retrieval approach for brain MR images [112] and [113]. This time, the authors used two complementary intensity invariant features, local binary patterns and Kanade-Lucas-Tomasi feature points, and compared them against a baseline method. Experiments results show that LBP with spatial context yields the highest accuracy and performs quicker than Kanade-Lucas-Tomasi feature points – which is observed to considerably degrade the performance. Kim et al. [53] designed an LBP for allow for a simple and efficient feature extraction based classification of x-ray images. For a fast and accurate classification task, Random Forests is adopted which is a decision

tree based ensemble classifier. The concatenation of LBP and Radon forest if observed to improve accuracy when compared to other approaches in the data set; in particular, the training and testing speed is noticeably better for the aforementioned approach.

2.2 Deep Features

Generally, features that are extracted by deep learning algorithms can be composed of both local and global features, for this reason we believe deep features deserves a section of their own. Deep features are mainly extracted by Convolution Neural Network (CNN) and AE, wherein features are extracted through a learning process [123], [118], and [18]. In essence, deep learning algorithms are designed to generalize a data set by learning relationships between the input parameters and automatically extracting salient features, such as corners, edges, and blobs. These features are observed to be globally expressive local features and symmetries; hence these features are locally extracted to compose an overall global representation of an image. Fig. 2.4 provides a pictorial representation of deep features.
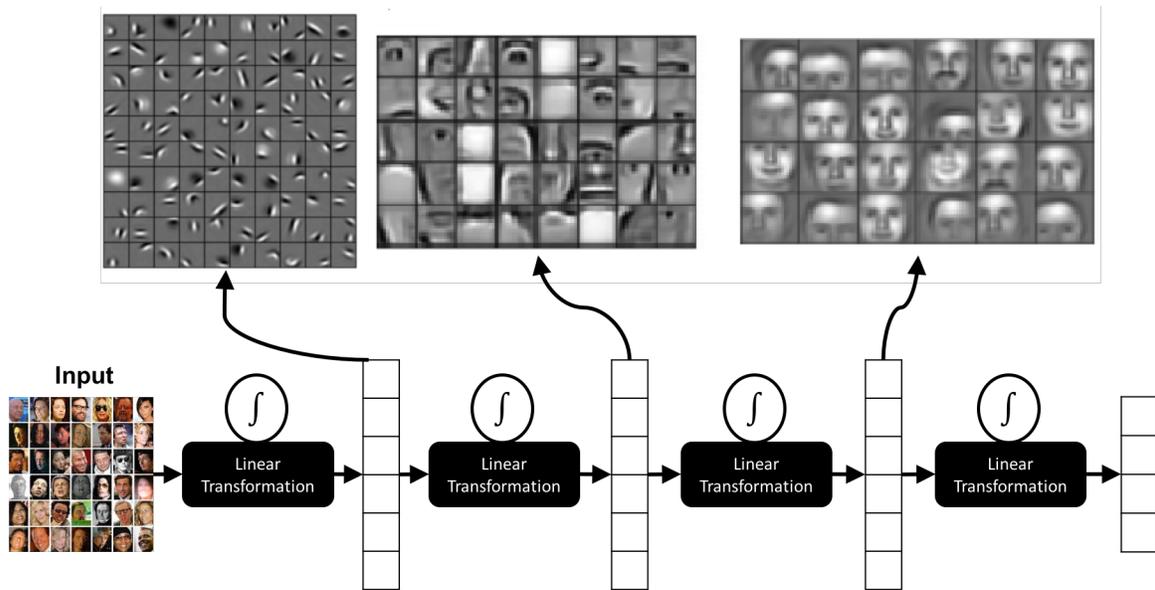
Figure 2.4: Extracting deep features from face images.

Popular deep learning algorithms such as CNN [12] and AE [41] have been adopted to generalize the data set to provide state-of-the-art performance for feature extraction

and retrieval problems. In particular, these networks posses activation functions for each layer which are local features describing a particular image region. These local features are aggregated to provide a global descriptor that represents the image at its entirety. An advantage of use deep learning strategies is their ability for transfer-learning, i.e., to fine-tune CNN models (supervised) pre-trained by natural image data set and to apply on medical images [92]; this means that the network should be trained only once for a particular image category or recognition task. Specifically in the medical imaging domain, deep learning algorithms are widely used for feature extraction and retrieval; there have been various surveys on this topic [39, 62, 64]. Similar to LBP, deep descriptors have set many competitive accuracy standards, specifically in the histopathology domain [9,98,120].

Recent developments in machine learning have opened up new opportunities for image retrieval. Among others, AE, introduced in 2006 by Hinton and Salakhutdinov [41], are designed to compress high-dimensional data by removing redundancies and preserving salient features. Deng [34] presented a notable survey on AE, covering the mathematical aspects and a general overview of the types of learning and applications for which AE can excel. Goyal and Benjamin [37], on the other hand, describe the history of deep learning and describe various image data sets for which the deep learning architectures can be exploited.

Initially, AE were trained on the MNIST dataset [41] for extracting features and retrieval [57]. In 2015, Camlica et al. [22] presented an AE to eliminate image regions that possess low encoding errors by validating its retrieval accuracy against both LBP and SVM. Tested against the IRMA dataset, roughly 50% of the image area was removed which increased the retrieval speed by 27% with less than 1% decrease in the accuracy. Sze-To et al. [103] developed a Radon Autoencoder Barcode (RABC) to hash images into binary codes using the IRMA dataset. Their method obtained the lowest total error of ≈ 344 using 512-bit codes for retrieval. In 2016, Tizhoosh et al. [108], introduced Autoencoded Radon Barcode (ARBC) to autoencode Radon projections using mini-batch stochastic gradient descent by binarizing the outputs from each hidden layer during training, and to produce a *barcode* per layer. A comparison with other methods (e.g., RBC, SURF, and BRISK) suggested that ARBC can achieve the lowest IRMA error of ≈ 392.

Motivated to exploit the potentials of AE for medical image retrieval, in this work a novel approach will be proposed which classifies autoencoded Radon projections using MLP to calculate the class probability for each query image. Radon transform is widely used in medical applications as means for reconstructing images using equidistant parallel rays penetrating an object from many directions [88] [33]. Specifically, an AE is adopted for dimensionality reduction. The autoencoded features (compressed image/histogram/projections) are then classified by an MLP to assign the image to a class. Radon projections are ex-

14

amined as features based on some recent success; Shujin and Tizhoosh [125] combined Radon projections and SVM for CBIR tasks. As well, Tizhoosh et al. [109] presented MinMax Radon barcodes which are observed to retrieve images 15% faster compared to "local thresholding" [78].

Specifically in the medical imaging domain, deep learning algorithms are used for feature extraction and retrieval; there have been various surveys on this topic, which include: [39], [64], and [62].

Recently, complex deep learning algorithms have been proven as imperative tools for big data images, yielding high accuracy and efficiency for diagnosing histopathological images [65]. Xu et al. [120] adopted deep learning for automatic extraction of histopathology images and multiple instance learning (MIL) for classifying the features. The authors observed that the automatic feature learning outperforms the manually extracted features. The best trained network using MIL on course labels achieves an accuracy of 96.30% using deep learning features. The histopathology images were extracted from 132 patients, each image of which is of dimension 10000×10000 pixels. The authors sampled 200×200 pixels with an overlap step size of 100 pixels, yielding a total of 9801 patches for each image. In 2015, Kumar et al. [58] adopted an AE, as a deep feature extractor, to classify lung nodules as either malignant or benign. The validation of the was evaluated using the National Cancer Institute (NCI) Lung Image Database Consortium (LIDC) data set which is comprised of 4303 instances containing 4323 nodules. After a 10-fold cross-validation, the network achieved an overall accuracy of 75.01% with a sensitivity of 83.35% and false positive of 0.39/patient. Bar et al. [13] developed a deep CNN to identify different types of pathologies in chest x-ray images. The algorithm was validated on a data set which consisted of 93 images, achieving an area under curve (AUC) of 0.93 for Right Pleural Effusion pathology detection, 0.89 for Enlarged heart detection and 0.79 for when classifying between malignant and benign chest radiographs. Shin et al. [92] adopted a transfer learning approach by obtaining a pre-trained ImageNet (via fine-tuning), with five-fold cross-validation, to detect thoraco-abdominal lymph node (LN) detection and interstitial lung disease (ILD) classification of the ImageNet data set. The authors achieved their best result on the mediastinal LN detection by obtaining a 85% sensitivity at 3 false positive per patient. The authors conclude that deep pre-trained CNN can be used for CAD systems for medical imaging tasks. Bar et al. [14] adopted a pre-trained deep CNN architecture, trained on ImageNet which is a non-medical data set, to detect and classify chest radiograph pathologies. The CNN with GIST features was fine-tuned and validated on on a 433 image data set, obtaining an area under curve (AUC) of 0.87-0.94 for different pathologies. Ciresan et al. [28] designed a use deep max-pooling CNN to detect and classify mitosis in breast histology images. The network utilizes raw RGB squared images, each images of

which contain mitosis which are detected by applying the classifier on a sliding window. The network is evaluated on the MITOS data set, comprised of 50 images which correspond to 50 high-power fields of 5 different biopsy slides stained; in total, this data set contains 300 mitosis. The proposed architecture won the ICPR 2012 mitosis detection competition, achieving a precision, recall, F1-score of 0.88, 0.70, 0.782, respectively. More recently, in 2016, Janowczyk and Madabhushi [48] developed a single architecture, which is used to detect and classify digital pathology images. In particular, the structured network achieves an F-score of 0.83 from 12,000 nuclei segmentations, F-score of 0.84 across 1735 regions for epithelium segmentation, F-score of 0.83 from 795 tubule segmentations, F-score of 0.90 for detecting 3064 lymphocytes, F-score of 0.53 for detecting 550 mitotic events, F-score of 0.7648 on 50,000 images of invasive ductal carcinoma, and a classification accuracy of 0.97 across 374 images containing lymphoma. Recently in 2016, Wang et al. [119] developed a deep learning algorithm to detect and classify metastatic breast cancers from whole slide images containing sentinel lymph node biopsies - the network of which won the International Symposium on Biomedical Imaging (ISBI) grand challenge. The proposed network achieved an AUC of 0.925 for WSI classification and a score of 0.7051 for the tumor localization task. When combining the deep learning knowledge with a pathologist, the AUC increases to 0.995, reducing human error by approximately 85%. In 2016, Spanhol et al. in [99] utilized a CNN architecture, namely AlexNet, to classify between malignant and benign tissue by adopting the BreaKHis dataset. The authors tried various techniques to train the CNN, including: extracting random patches from originally dimensioned images and sliding window technique with 0.5 overlap to obtain patches from each train image. The authors of [99] achieved a score of 87% when using a CNN with different fusion rules (see [54] for details). For each CNN, which are trained with different inputs (i.e., size and number of patches), the fusion rules are applied to the fully-connected layer which contains the softmax activation. In particular, three rules are applied: sum, product, and max - of which the "max" rule is observed to provide the best result when retrieving at patient level. This patient score is, however, outlined for a specific fold, and is not reflective of the entire data set. Bayramoglu et al. in [17] proposed a method to classify histopathology images using BreaKHis data set; however, the proposed CNN is independent from magnification level. The authors created magnification independent models, which are supposedly scalable, and the model is fine-tuned using new samples. In particular, the authors designed a multi-task CNN that is capable of predicting both the magnification level and if the query image is benign or malignant. The patient score for the proposed framework is 83.25% which is very close to PFTAS approach, which obtained a patient score of 83.33%.

2.3 Radon Transform as Feature Extractor

Radon transform provides a graphical representation of an object, also referred to as sinogram, through a number of sine-waves that have different amplitudes and phases. These waves are observed to hold features that represent the texture and shape of the objection. Radon projections have been widely adopted in the medical imaging including x-ray imaging, computed tomography (CT) scans, and a variety of other radiological imaging techniques.

There have been many attempts to use Radon transform as means for generating rich descriptors without prior knowledge of the object [88]. Hoang and Tabbone [43] proposed a descriptor that employs both Radon with Fourier and a modification of the Mellin transform (RFM) to extract features, without normalization, that are invariant to rotation, scaling, and translation. The authors validate the effectiveness of this algorithm on noisy, occlusive, and deformed data sets - the result and performance of which is reported to be reasonable compared to commonly used pattern descriptors. Nacereddine et al. [73] adopted Radon transform to develop a new descriptor called "Phi-signature" which is observed to be invariant to geometric transformations. In particular, the authors validate the effectiveness of the multi-level representation of the descriptor using Phi-signature and R-signature, and compare it to a generic Fourier descriptor (GFD). The proposed RT-based descriptor is observed to increase the discrimination capability, thereby performing faster and more reliably than GFD when tested against graphical symbols. Chen and Chen [26] presented a feature-based invariant descriptor called Radon Composite Features (RCFs) designed particularly for planar shapes. The essence of the algorithm is to transform binary shapes into a 1-dimensional feature vector. In particular, the feature shapes are extracted from the Radon transform using statistical and spectral analysis which is observed to overcome the invariance encountered by conventional geometric transformations. Experimental results indicate that the RCF provides better discrimination between shapes when compared to several state-of-the-art techniques in literature. Jadhav and Holambe [45] presented a pattern recognition technique for face recognition based on the combination of Radon and discrete cosine transforms (DCT). The Radon transform is observed to enhance the low frequency components, imperative for extracting salient information from facial structures; whereas, the DCT compresses the feature yielding a low-dimensional feature vector. Thereafter, the same authors [46] proposed a pattern recognition algorithm using Radon and wavelet transform, validated on facial features. The combination of Radon and wavelet transform provides a descriptor that is invariant to the changes in facial expression and illumination - this is accomplished by computing the Radon projection in different angles enabling the algorithm to capture features of the face at various directions. The

wavelet transform, on the other hand, complements the features from the Radon space by providing multi-resolution features of the facial images. The proposed algorithm is validated on FERET, ORL, Yale, and YaleB databases outlining the superiority of the proposed method with some of the existing algorithms.

The first attempt to quantify features onto a histogram is presented by Tabbone et al. [104]. Wherein, the authors propose a histogram composed of Radon transformations called HRT, designed to be invariant to geometrical transformations. The authors validate on an openly available shapes, trademarks and structural database for which the black and white images, when computed using HRT descriptor, provides a shape that is a re-partition on the angular length. The authors mention that retrieval using Euclidean distances, to compare the histogram shape, is debatable. To facilitate a better retrieval using Radon transform, Dara et al. [32] presented a novel method of content-based image retrieval based on 3-D generalized Radon transform. The authors obtain descriptors using Radial Integration Transform (RIT) and Spherical Integration Transform (SIT) which provides the shape characteristics of the data. These descriptors were used for retrieval purposes, for which the author claims a satisfactory result in terms of both precision versus recall and time taken for retrieval.

Particularly in the medical imaging domain, substantial work has been put-forward by researches to develop a Radon descriptor that has the ability to generalize any medical modality. Inspired by the barcodes embedded in many products we encounter on a daily basis, a novel Radon barcode for medical image retrieval system was proposed in 2015 [107]. The Radon barcode is a binary vector generated based on Radon Transform with selected projection angles and projection binarization operation that can tag/annotate a medical image or its regions of interest. Hence, large image archives can be efficiently searched to find matches via Hamming distance, e.g., KNN search, or hashing-based methods, e.g., locality sensitive hashing.

In 2016, Shujin and Tizhoosh [125] combined Radon projections and SVM for CBIR. Using the 2009 IRMA data set, a total error of 294.83 was obtained for 1,634 images. Babaie et al. [10] propose a CBIR approach for large data sets using a single projection based on Radon barcodes (SP-RBC). This approach supposedly captures the required global features of an image with a purpose to expedite weak learning approaches. The proposed method is tested on IRMA 2009 data set and it is observed to substantial decrease in the error rate when comparing against other non-learning methods. Although these methods are effective, they do not compare well against leading methods such as LBP or deep features. In early 2018, Tizhoosh et al. [106] introduced a descriptor called "Encoded Local Projections" (ELP), which generates a Radon-based histogram for classification. In particular, the authors extract local neighborhoods of an image from which an "anchor

projection" is obtained by calculating the angle at which the amplitude is maximized. Using the anchor projection, three equi-distant projection is calculated from which the projections are re-scaled using minmax approach, and counted to form a histogram. This approach is validated on three public datasets (IRMA, KIMIA Path24, and CT Emphysema), all of which provides competitive results. For the KIMIA Path24, the SVM and ELP combination achieved a patch-to-scan accuracy of ≈83%, and a whole-scan accuracy of ≈80%.

Radon transform is also used to generate descriptive features when used as an input parameter for deep learning algorithms, which is often proved to yield higher accuracy when retrieving like-images. Sze-To et al. [103] designed an AE with Radon barcodes as inputs to hash images and convert them into binary codes for retrieval. When validating this approach on the IRMA data set, the Autoencoded Radon Barcode (ARBC) is observed to retrieve 9.27 times faster. Later, Tizhoosh et al. [108] improved the ARBC by using mini-batch stochastic gradient descent along with binarizing the weights from each hidden layer during training - this approach is observed to produce a barcode per-layer. The revised ARBC achieves a 392.09 IRMA error.

Although there have been many works on Radon features for image processing, this thesis explores the possibilities of utilizing Radon projections as image descriptors. In particular, there are two novelties that manifests the properties of Radon. For the first novelty, Radon features extracted from raw images and trained using an Autoencoder and MLP for classification. Thereafter a comparative test is performed against learning raw image and learning HOG features extracted from raw image. The idea for using Radon features for learning is influenced by [103] and [108]. Although there have been works on using Autoencoder for learning Radon projections, the novelty presented in Chapter 3 is to use the compressed Radon features from Autoencoder, and pass it onto MLP for classifying the features. For the second novelty, a global hand-crafted descriptor called FLIP is designed which uses equi-distant Radon projections extracted from small neighborhoods, of size $3 times 3$, to form a compact histogram - which is scale and translation invariant. Influenced by [106] to use local Radon projections as descriptors, the purpose is to design a descriptor that "counts" the projections into a histogram, which is inspired by [75]. A multi-resolution version of FLIP called mFLIP is designed which, motivated by [77]. As expected, the experimentation results suggests that mFLIP performs better than FLIP, which is further discussed in Section 5. This suggests that Radon projections are viable features that provides a feasible approach for CBIR.

Chapter 3

Learning Autoencoded Radon Projections

This chapter is based on the paper "Learning Autoencoded Radon Projections" authored by Sriram et al. [100], which was published in IEEE Symposium Series on Computational Intelligence, 2017.

3.1 Motivation

An AE is an unsupervised bottle-neck network designed to learn encodings of a dataset by reducing the dimensionality, thereby preserving compact features in its most compressive layer [41]. Depicted in Fig. 3.1, AE only expect a flattened vector as an input which is then compressed to a user-defined ratio - this compressed layer holds latent embeddings that is features of the input image. Thereafter, the AE decompresses the latent embedding layer back to its original dimension, which provides an approximate representation of the input image, also known as learned representation. Hence, to confirm the validity of the compressed layer, one would have to look at the output layer to visually confirm if the image displayed is similar to that of the input image.

It is the claim for machine learning algorithms to "understand" an image only using its raw pixel intensity values [41]. However, since AE are dimensionality reduction algorithm, the most compressive layer is only useful and representative if the input is descriptive for learning. AE being an unsupervised network negates the associated labels, this network is designed to build structure in the data without supervision. However, in the medical field,
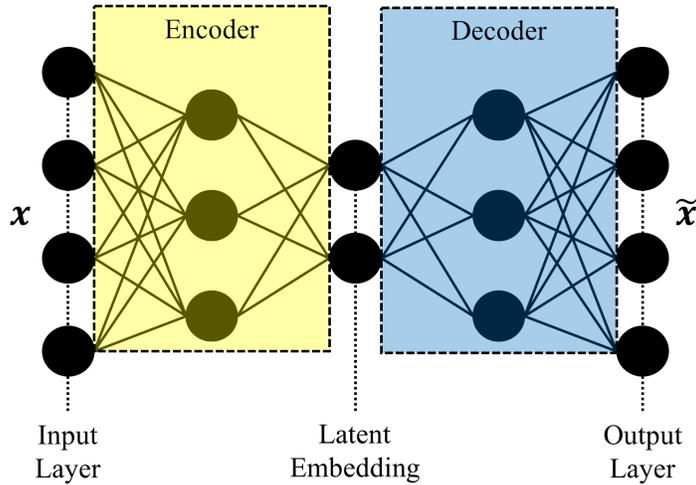
Figure 3.1: Example of an AE which illustrates that the input and the output is similar confirming that the deepest layer holds salient features of the input image.

labels are of utmost importance, as they are annotated by medical specialists. In order for AE to bridge the relationship between the latent embeddings and the associated labels, a MLP is adopted. Illustrated in Fig. 3.2, MLP accepts data along with coressponding labels as input parameters and learns to build a relationship in order to generalize the data set. In particular, the MLP learns to classify the data providing a class probability output between 0 and 1 for a pre-defined number of classes (the last layer being softmax) for each image that passes through the network. The motive of this chapter is to propose a multi-phase framework that utilizes the compression behavior of an AE which is then sequentially trained with a MLP for classification. To that, the proposed framework incorporates the AE for compressing the images, the latent embedding of which is then passed onto the MLP along with its class labels as input parameters. The output of the MLP is a class probability based on the latent embeddings.

Over the years, AE have been widely adopted for encoding medical images, from which the most compressive layer is preserved as a feature vector representing the input image. In this thesis, a new framework is proposed for retrieving medical images by classifying Radon projections. In particular, this framework is a supervised learning approach which is composed of two networks concatenated together, namely: AE for dimensionality reduction followed by an MLP for classification. For each image provided to the framework, equi-distant Radon projection is extracted from the raw image forming a vector that is representative of the image - illustrated in Fig. 3.3. Thereafter, these Radon projections
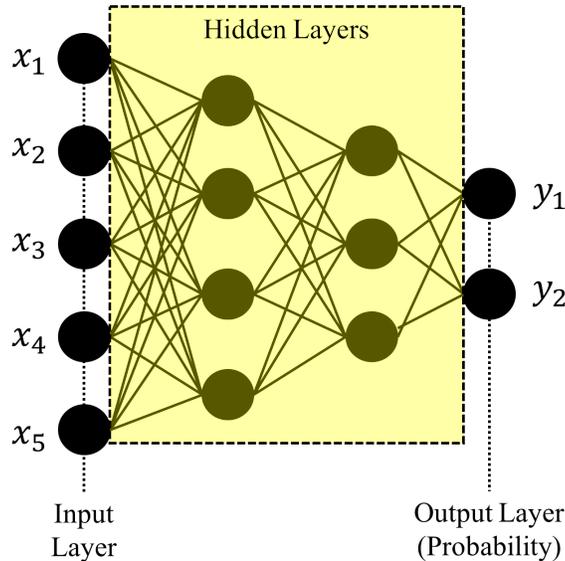
Figure 3.2: Example of an Multi-layer perceptron wherein the input is the raw data along with the class labels, for which the output is a class probability for a pre-defined number of classes ranging between 0 and 1 (1 being the best match).

are provided as an input to an AE from which the deepest layer containing the feature vector is preserved. The feature vector from the AE is passed onto an MLP which classifies the feature vector thereby building a relationship between the compressed layer and the associated class. The integration of MLP is observed to promote a rather shallow learning architecture which makes the training faster.

This chapter also compares the learning behavior of an AE and multi-layer perceptron for when the provided input is the raw image, Radon projections of an image, and HOG extracted from an image. The proposed framework is benchmarked on the openly-available IRMA dataset containing 14, 410 radiograph images distributed across 57 different classes. Through experiments, it is apparent that Radon features are better inputs for this proposed framework, as they yield a higher accuracy and quicker computation. Experiments showed an IRMA error of 313 (equivalent to $\approx 82\%$ accuracy) when providing the Radon projections as an input parameter. This outperforms other works, such as [103], [91], and [108], on retrieval from IRMA dataset using AE. The proposed method is implemented using *Keras* and *Theano*, and the networks were trained on Nvidia's Tesla K-80 GPU provided by SHARCNET [2].

(a) Radon projection at 0°

(b) Radon projection at 45°

(c) Radon projection at 90°

(d) Radon projection at 135°
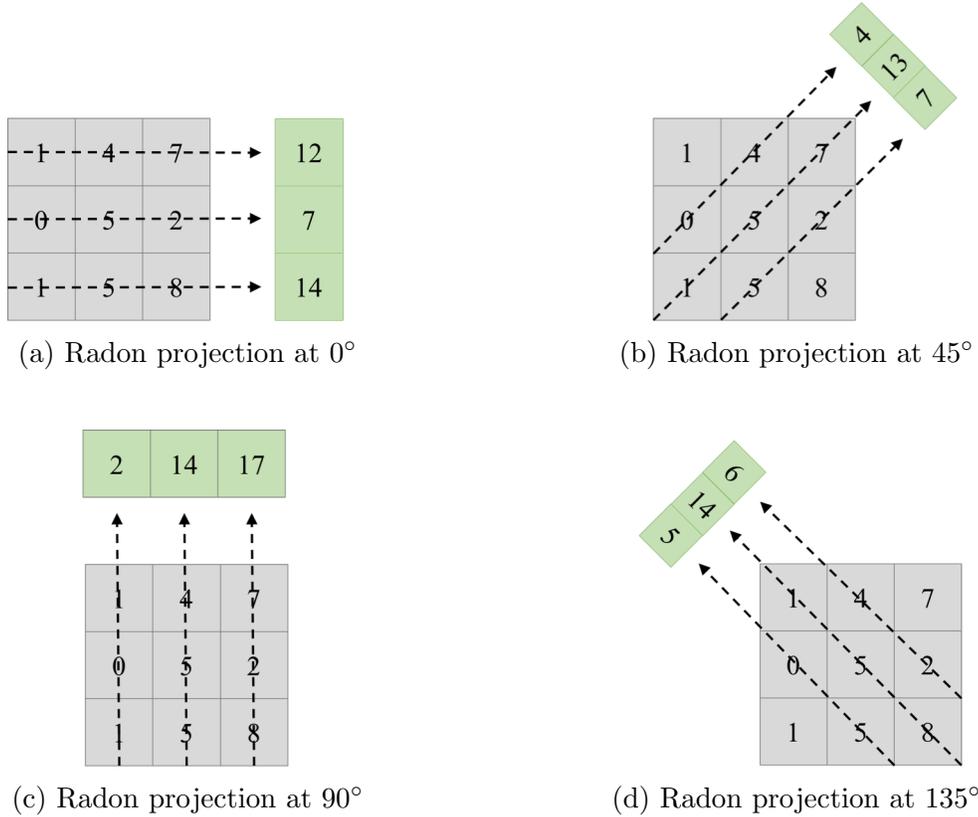
Figure 3.3: Computing equi-distant Radon projection of an image at 0°, 45°, 90°, and 135°. In this case, an image of size 3×3 is adopted, each cell of which represents a pixel intensity value.

3.2   Training an Autoencoder

As for training the networks, the proposed sequential framework is comprised of three components: (i) AE for dimensionality reduction, (ii) multi-layer perceptron classification, and (iii) probabilistic retrieval. To begin, Radon projections are extracted from each image and vectorized to form an input to the AE. It is observed that obtaining Radon projections from images not only decreases the computational time, but it is also observed to remove redundancies when compressing the Radon features, preserving only salient features for classification. As equi-distant Radon projections are capturing the intensity of pixel value across four angles (0°, 45°, 90°, and 135°), the obtained Radon signal is observed to be unique to the image, thereby removing unnecessary inter-pixel redundancies.

For the classification component, an MLP is adopted to establish a feature-to-category relationship, wherein features are obtained from the compressed Radon projections in the deepest autoencoded layer. Fig. 3.4 provides a schematic representation of the proposed framework. In order to validate the framework, the networks were trained using 1) raw pixels, 2) HOG features, and 3) Radon projections. The experimental results of which suggest Radon projections to be the most descriptive inputs for the proposed framework.
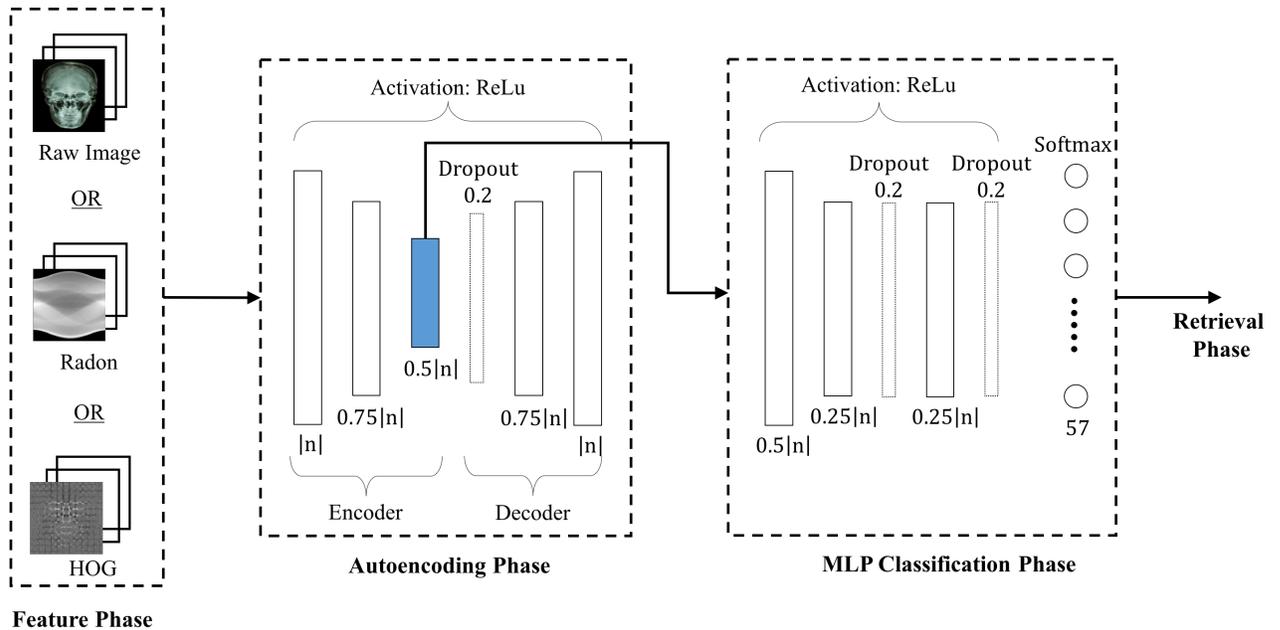
Figure 3.4: Classification of radiography images into 57 classes by MLP using the deepest layer of AE. The input parameter is a choise of (i) raw pixels, (ii) Radon projections, or (iii) histogram of oriented gradients [100].

As depicted in Fig. 3.4, the proposed framework is divided into three phases: (i) extracting features from raw image, (ii) training the AE, and (iii) training MLP classifier using AE's deepest layer as input along with the class labels. For sake of comparison, the framework is trained for three types of features: Radon projection, HOG, and raw images (pixel intensity values) which are extracted and pre-processed prior to training the AE. Illustrated in Table 3.1, the pre-processing steps depend on the type of input feature provided to the network. For raw images, the pixel intensities are normalized between 0 and 1, and padded to maintain a square size - as AE are restricted to have the same input dimension for all images when training. Radon projections, on the other hand, are normalized by dividing each projection by the maximum value, and then standardized,

Input Feature	Pre-processing Step
Input x-ray Image	Normalize $[0, 1]$, Pad to square, Re-size to $N \times N$
Radon projections	Standardization $\mu = 0$, $\sigma = 1$
HOG features	As is
Raw features	Flattened pre-processed image

Table 3.1: Pre-processing applied to the inputs prior to training the AE [100].

with $\mu = 0$, $\sigma = 1$. Finally, HOG features are provided to the network unprocessed.

Regardless of the feature extraction method, all features are vectorized and provided to the AE for compression. To obtain an optimally trained network, early stopping with an epoch patience of 3 along with 10-fold cross validation is configured for both AE and MLP networks. The final performance measure is the average of the values computed in the loop of the 10-folds – which ensures the creation of an optimal model when the number of samples is very small. The final classification probabilities are produced by the best configuration.

There are several motivations behind the multi-phase training architecture for the proposed method. The latent embeddings from the compressed autoencoded layer is observed to preserve salient attributes that are free from redundancies; this holds true for when the input is descriptive for generalization. The concise and dense latent embeddings from the AE enhances the convergence and classification accuracy of MLP. A faster convergence also contributes to a rather shallow AE that is easier to train. In essence, a sequential training can be thought of as pre-training steps, enabling further fine-tuning for a better classification. As for testing the proposed framework, each query is passed through both trained AE and the MLP to obtain a final class assignment which is used later for similarity matching during the retrieval process.

3.3   Retrieval Strategy

To increase the retrieval accuracy, the images within the top-five probable classes are selected from the trained database. Along with the top-five class probabilities, an equidistant normalized Radon projections are obtained from the query image. For retrieval, the Radon projection for the query image is compared with the top-five class probability from the MLP using the KNN method. In this case, KNN yields the index of the most

```
procedure GETFEATURES(I, F_s):
I_p ← double(I) (Normalize to [0, 1]);
D_max ← max(I_p.shape) (Find max dimension of the image);
I_p ← Pad(I_p, D_max) (Convert I_p to D_max × D_max size, aligning to center of larger dimen-
sion);
I_p ← Resize(I_p, (N, N)) (resizes I_p to N × N);

if F_S is HOG then
    I_f ← ReShape2Vector(HOG(I_q, n_hog)) (reshapes 2D to 1D) ;
else if F_s is Radon then
    I_f ← Normalize(R(I_q, θ), axis = 1);
    I_f ← ReShape2Vector(I_f);
else
    I_f ← ReShape2Vector(I_q) ;
end
Return I_f ;

end procedure
```

Algorithm 1: Pre-processing the image I and its features I_f [100].

similar image, as illustrated in Fig. 3.5.

The Radon projection for each query image along with the top-five class probability for when the query image is passed to the framework is weighted together for retrieval. In particular, the Euclidean distance between the projections is computed to retrieve the best matched image. Generally, the retrieval phase is structured such that the MLP is employed for global search (class assignment), and the input feature (i.e., Radon projection) is computed for local search (to retrieve the most similar image). Analogously, the same procedure is applied to HOG and raw images, when they are provided as input features to the designed framework (see Algorithm 1). The overall architecture is illustrated in Fig. 3.5 and Algorithm 2. In essence, a total of 57 probabilistic outputs are ranked from which the top 5 are used in a KNN search to find the best match for the query image. After various empirical experimentation, it is important to note that KNN with L_2 distance provided the best accuracy.

The aforementioned framework is only compared against other AE approaches reported on the glsirma data set. The IRMA accuracy measures for the proposed framework is outlined in Table 5.2, and discussed further in Chapter 5.1.1.

It is the claim that deep learning algorithms are a flexible function that can learn to
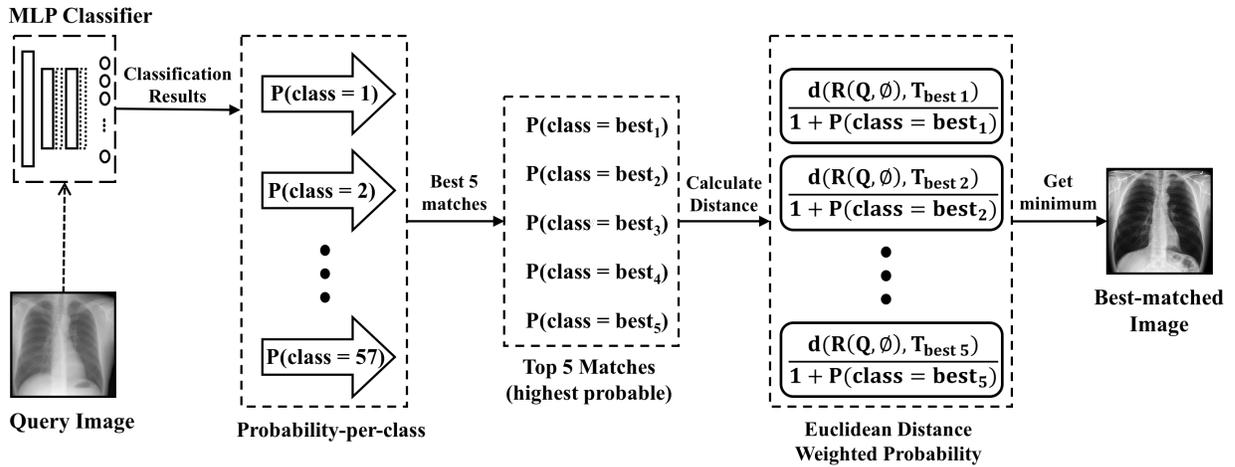
Figure 3.5: For each image, the features (such as Radon, HOG, or raw pixels) are extracted, compressed by the Autoencoder, and passed to MLP for classification. The classification is sorted to determine the best class. Thereafter, the image is retrieved using KNN search within the top 5 predicted classes, wherein the summation of the probabilities is 1 [100].

fit raw data. However, this chapter provides evidence that deep learning algorithms can perform better when provided with features as input parameters. This strategy promotes the deep structures to get a better understanding of the data which further helps in feature extraction and classification. In the coming chapter, a hand-crafted radon descriptor is proposed to advance the importance of radon as a feature extractor.

I_q: the query Image, AE: the trained autoencoder, MLP: the trained MLP, F_S: defined input features (i.e., Radon, HOG, raw images)

procedure RETRIEVE(I_q, AE, MLP, F_s):
$I_f^q \leftarrow$ GetFeatures(I_q, F_s);
$I_{\text{enc}} \leftarrow$ Autoencoder(I_f);
$P_c \leftarrow$ MLP(I_{enc}) (get class probabilities);
$\text{Top}_{idx} \leftarrow$ argsort(P_c, $'decreasing'$)[: 5] (get indexes of top five classes);
$D = []$ (Initialize empty array);
$\text{Idx} = []$;
$m = 0$ (number of candidate images);
for i in Top_{idx} **do**
    for j in $|C_i|$ *(for all images in class C_i)* **do**
        $I_f^{i,j} \leftarrow$ GetFeatures(I_i^j, F_s) (get features for the candidate image)
        $d, \text{idx} \leftarrow$ KNN($I_f^{i,j}, I_f^q$) (perform knn search)
        $m = m + 1$
        $D[m] \leftarrow \frac{d}{1+P_c[i]}$ (normalize w.r.t. to prediction probability)
        $\text{Idx}[m] \leftarrow idx$
    end
end
$Best\_Match = \text{Idx}[\text{argmin}(D)]$;
return $Best\_Match$

Algorithm 2: Retrieving the best match for the query I_q [100].

Chapter 4

Forming Local Intersections of Projections

The materials of this chapter is currently under review at the European Conference on Computer Vision, titled "Forming Local Intersections of Projections (FLIP) for Retrieval of Histopathology Images" authored by Sriram et al. [101].

4.1 Motivation

WSI in the field of pathology is a digital imaging solution which refers to scanning a specimen smeared on a glass-slide thereby digitizing the slide. Recently, content-based image retrieval has gained traction in pathology for diagnosing, analyzing and retrieving similar images contextually. However, since digital pathology scans are large in size, it becomes difficult for established deep learning algorithms to process. A commonly used strategy for deep learning algorithms is to down-sample the images in order to process such large images; this technique, however, causes loss in quality of data. Hence, the motivation is to develop a global feature extractor that can concisely representing a large pathology scan into a compact yet dense vector for quick retrieval.

In this section, a novel global image descriptor called FLIP is proposed for representing large histopathology images. The proposed descriptor is based on Radon transform which computes parallel projections in small local neighborhoods of gray-level images to form a histogram for retrieval. In particular, four equi-distant projections are computed for each

local neighborhood of 3×3 at angles 0°, 45°, 90°, and 135°. After which for each neighborhood, the intersection (a numerical equivalence of logical AND) is calculated between all adjacent projections - an entropy strategy inspired from information theory. Thereafter, each neighborhood is re-scaled between 0 to 128 from which a histogram is constructed by counting the intensity instance of each neighborhood to form a representation for the entire image. Finally, from the obtained histogram, the first bin is neglected (to remove redundancy) to form the FLIP histogram with a 127 bins.

In addition, this section also introduces an extended version of FLIP called mFLIP. In essence, the FLIP descriptor for 1000×1000, 750×750, 500×500, and 250×250 images are concatenated to form the mFLIP descriptor. It is observed that the mFLIP descriptor outperforms all feature sets on the KIMIA Path24 data set, which consists of approximately 27,000 histopathology patches of size 1000×1000 belonging to 24 classes. Experimental results show that the mFLIP descriptor achieves an overall accuracy (η_{total}) of $\approx 72\%$ when combined with SVM classification, and $\approx 60\%$ when retrieving using χ^2 distance. Not only does mFLIP set the current benchmark on the KIMIA Path24 data set, it also surpasses all deep learning algorithms reported in the aforementioned data set.

4.2   Localizing projection features

Examining Radon projections to form a global descriptor is an active area of research which is observed to work well for pathology images [106]. In this section, the proposed handcrafted image descriptor, FLIP, is discussed in greater detail on its influence when representing histopathology images. It is important to note that the FLIP descriptor is observed to outperform various established deep feature algorithms, including pre-trained networks as well as other powerful global descriptors like LBP and HOG. The performance of the FLIP algorithm is verified on the KIMIA Path24 data set. For illustrating the variety of textures in this data set, Fig. 4.1 depicts nine randomly selected samples from the KIMIA Path24 data set.

4.2.1   Projection Features

Introduced by J. Radon in 1917 [85], Radon transform is widely used in the medical domain as a way of reconstructing a function from the values of its projections in $\mathbb{R}^n$ space [115]. Although the intent here is not reconstruction, for the purposes of the FLIP descriptor, the Radon transform is adopted to compute a profile that is a projection of the image
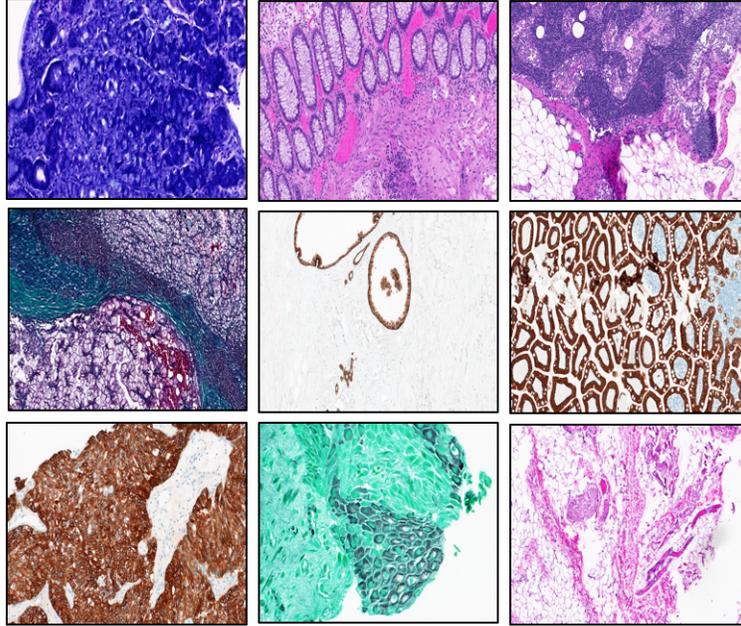
Figure 4.1: Sample thumbnails of scans depicting variety of cellular textures present in the KIMIA Path24 dataset. The images represent approximately 20× magnification, which is a portion of the WSI.

structures. When examining a function $f(x, y)$, one can project $f(x, y)$ along a number of projection angles; these projections are the sum (or integral) of $f(x, y)$ values along lines constituted by each angle θ. Hence, the projection creates a new image $R(\rho, \theta)$ which is a *sinogram* with $\rho = x\cos\theta + y\sin\theta$. Therefore, when using the Dirac delta function $\delta(\cdot)$, the Radon transform of a two-dimensional image $f(x, y)$ can be defined as its line integral along a line inclined at an angle θ and at a distance ρ from the origin [90]:

$$R(\rho, \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\delta(x\cos\theta + y\sin\theta - \rho)dxdy \qquad (4.1)$$

where $\delta(\cdot)$ is the Dirac Delta function which is non-zero only, wherein $-\infty < \rho < \infty, 0 \leq \theta < \pi$.

When designing the FLIP descriptor, a total of four equi-distant Radon projections were computed for each 3×3 neighborhood for each image. The Radon projections for each neighborhood is hard-coded instead of using a Radon transformation library, as depicted in Fig. 4.2. Since we are only interested in computing four equidistant projections per-

neighborhood, we add the pixel intensity across the projection angles of 0°, 45°, 90°, and 135°.
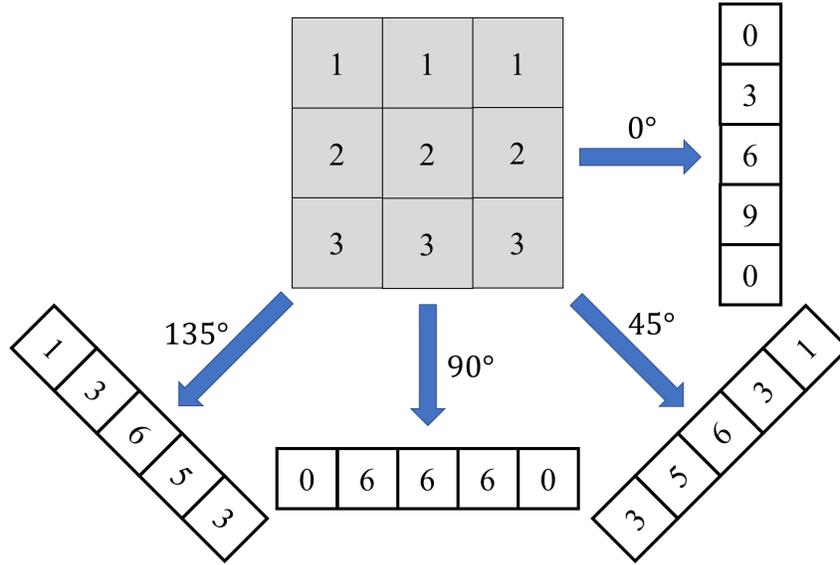
Figure 4.2: Computing hand-crafted Radon projection for a local 3×3 neighborhood.

4.2.2   Forming Local Intersections of Projections Histogram

FLIP is a compact histogram based on some general knowledge and intuitive expectations. Given the scan $\mathbf{S}$, we are interested in describing the grayscale image $\mathbf{I}(\subset \mathbf{S})$ via a short descriptor, or in this case a histogram, $\mathbf{h}$ using Radon projections $R(\rho, \theta)$ to transform the intensities $f(x, y)$ of $\mathbf{I}$. We can process all local neighbourhoods $\mathbf{W}_{ij} \subset \mathbf{I}$. For each neighbourhood $\mathbf{W}_{ij}$, we capture n_P projections with $0 < n_P \ll 180$: $\mathbf{p}_{ij}^1, \mathbf{p}_{ij}^2, \ldots, \mathbf{p}_{ij}^{n_P}$. It is observed that some individual projections from different (and dissimilar) images to be quite similar. Hence, the intersection of adjacent projections is calculated to quantify the spatial correlations of a given neighborhood pattern. The intersection of projections can be thought of as the logical "AND" operator, capturing a unique characteristic between the projections. In literature, "AND" logical operators compute the relation between two observations, wherein only if both the relations are true will the outcome be true. Intuitively, if the relations between two or more observation is true, then the combination of observations brings about a unique relationship which must hold true. To explain this predicament, if we are to identify a person in public who is tall or with brown hair, or

wearing a gray sweater or carrying a brown bag - this will be quite difficult to identify, because several people may hold these features. However, if a person is to be described as "a tall man AND brown hair AND wearing a gray sweater AND carrying a brown bag", then this does in fact narrow down the search for quickly identifying the person because we have formed compound features of this individual using the logical "AND" . Hence a combination of the features when combined together provides a unique characteristic which is expected to be discriminant. Therefore, when computing the intersection of adjacent Radon projections, we receive n_P intersection vectors $\mathbf{v}_{\min}^k$ as:

$$\mathbf{V}_{k,m} = \min\left(\mathbf{p}_{ij}^k, \mathbf{p}_{ij}^{(k+1)\%n_P}\right),\tag{4.2}$$

with $k = 1, 2, \ldots, n_P$ and $m = 1, 2, \ldots, n_W$ where n_W is the total number of local neighborhoods of the image $\mathbf{I}$. Hence, we will have $n_P \times n_W$ intersections of local Radon projections. These projections have different values which are also subject to intensity fluctuations. Thereafter, the four intersection of projections are re-scaled between 0 to 128 for all projection values across all neighborhoods of the image.

$$\bar{\mathbf{V}}_{k,m} = \left\lceil L \times \frac{\mathbf{V}_{k,m} - p_{\min}}{p_{\max} - p_{\min}} \right\rceil.\tag{4.3}$$

Now, the re-scaled Radon projections values are counted $\bar{\mathbf{V}} \in 1, 2, \ldots, L, \forall m = 1, 2, \ldots, n_P \times n_W$ to form the FLIP histogram $\mathbf{h}$.

Fig. 4.3 shows the main steps for creating the FLIP histogram. Capturing $n_P = 4$ equi-distance projections at $0°$, $45°$, $90°$, and $135°$. Algorithm 3 provides the pseudo-code for calculating the FLIP descriptor.

For comparison, Fig. 4.4 compares the FLIP histogram with the pixel intensity histogram (ranging from 0 to 255) for the given sample images. The image histogram is a count of all the pixel intensities which forms a vector of length 255. On the contrary, the FLIP histogram is only 127 in length, and captures the radon projections computed for every neighborhood within the image.

4.3   Multi-resolutional FLIP

A multi-resolutional approach is motivated by physics and biological vision which is now widely adopted in computer vision, image analysis, and signal processing algorithms. The intuition behind multi-resolution representation is to accommodate for structural changes

33
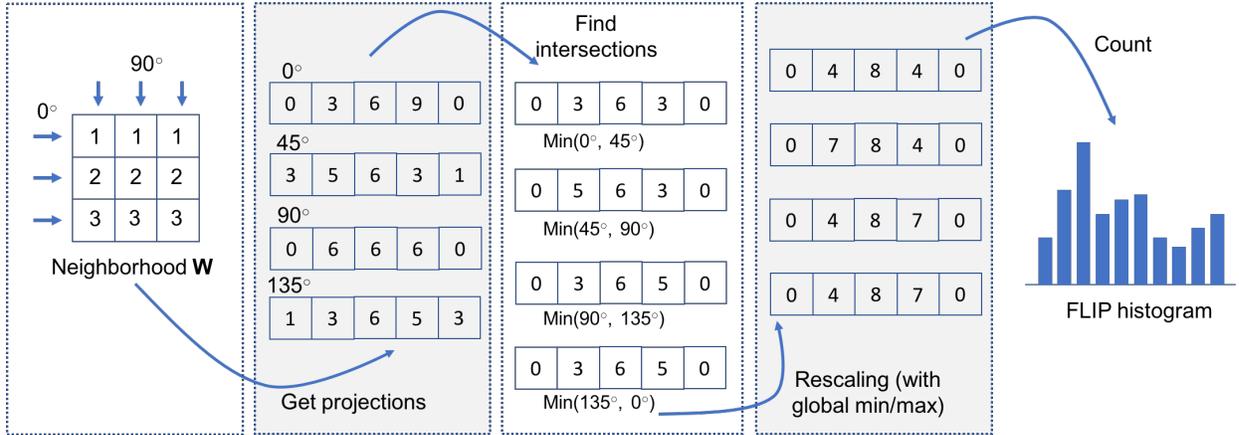
Figure 4.3: A simplified overview of the FLIP histogram extraction. A certain number of projections, here $n = 4$, is computed for each neighbourhood, from which the intersection of all adjacent projections is computed. After re-scaling all projections of all neighbourhoods in the image based on global min/max projection values, the FLIP histogram can be assembled.

that one observes in real-world objects [24]. The mFLIP approach is built upon the traditional FLIP descriptor. In the case of KIMIA Path24 dataset, we compute the FLIP for four image dimensions (inclusive of original and resized resolutions) - 1000×1000, 750×750, 500×500, and 250×250. After obtaining a 127 length FLIP histogram for each of the resolution, we concatenate the histograms in descending order of resolution to form a mFLIP histogram length of 508. It is the mFLIP histogram that sets the benchmark in the KIMIA Path24 data set, beating deep learning features as well as traditional computer vision techniques. Specifically, a mFLIP histogram of 508 in length achieves a total accuracy of 72% when using generalized histogram intersection SVM kernel classifier. Fig. 4.5 provides a pictorial representation to compute the mFLIP for a given image.

4.3.1 Indexing and Testing

The KIMIA Path 24 is a multi-class openly available histopathology data set consisting of approximately 27,000 train images and 1,325 test images of size 1000×1000, along with their associated classes. As described in [9], homogeneity and other techniques are implemented to on all 24 scans to only focus on the specimen instead of the background (noised portion). Once the specimen is obtained, the scans are divided into a regular grid delivering approximately 27,000 patches of size 1000×1000. For each patch the FLIP
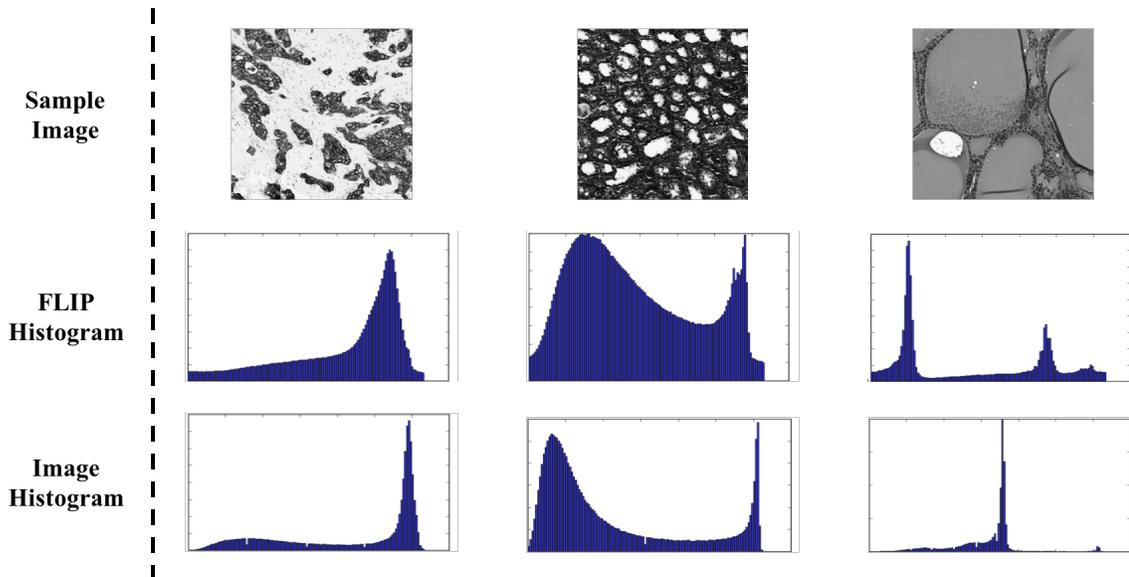
Figure 4.4: FLIP vs. Pixel intensity histogram for three randomly sampled patches.

histogram is obtained and saved in a database for retrieval. Once saved, the 27,000 train images are used for training using SVM classifier. Subsequently, for every FLIP histogram from the test set, the closest indexed FLIP histogram is obtained with its associated class for validity. Hence, the main task of the retrieval is to find the best match for a FLIP-computed patch. Fig. 4.6 presents a graphical overview of the overall approach to extract and test the FLIP histograms.

Although SVM obtained the best result for classification, the motive for retrieval is to also observe the performance of the FLIP when a learning algorithm is not applicable. To that, two matching strategies were incorporated: i) using a distance measure to quantify the (dis)similarity between the mFLIP histogram of the query patch and the mFLIP histogram of every image in the database, or ii) to apply SVM learning to assign a class to the query patch. For the distance-based retrieval, several strategies were used including: χ^2, histogram intersection, Pearson coefficient, cosine similarity, and L_1 and L_2 metrics. In terms of distance-based retrieval, χ^2 obtains the best result which reduces the accuracy approximately 10% when compared to SVM, these results are reported as well. The best result and the current benchmark for the KIMIA Path24 data set is achieved when combining the mFLIP global descriptor along with the generalized histogram intersection kernel SVM for classification.
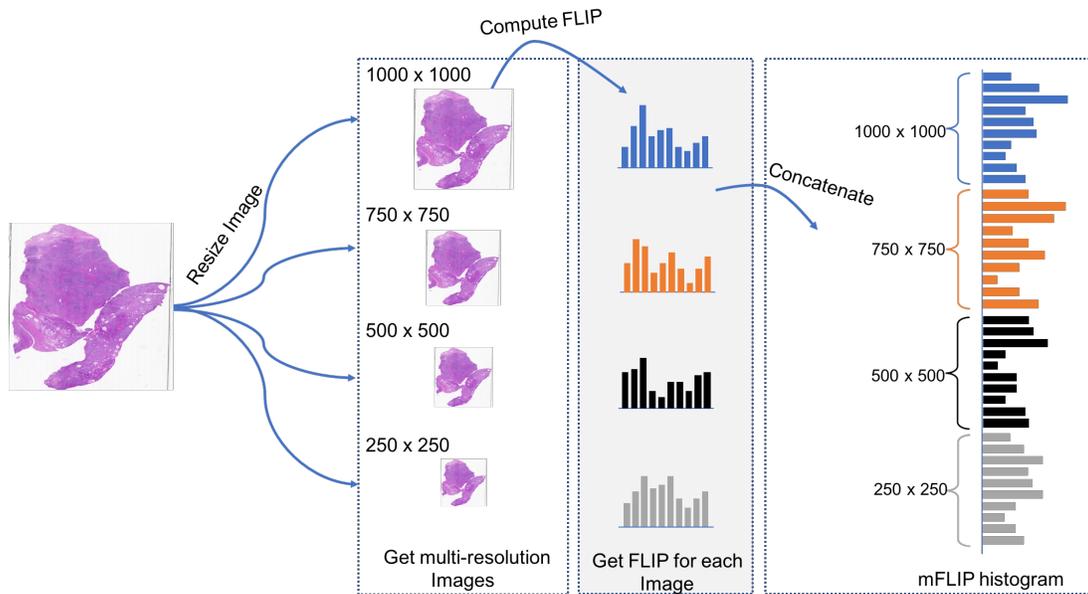
Figure 4.5: Multi-resolution Forming Local Intersections of Projections.
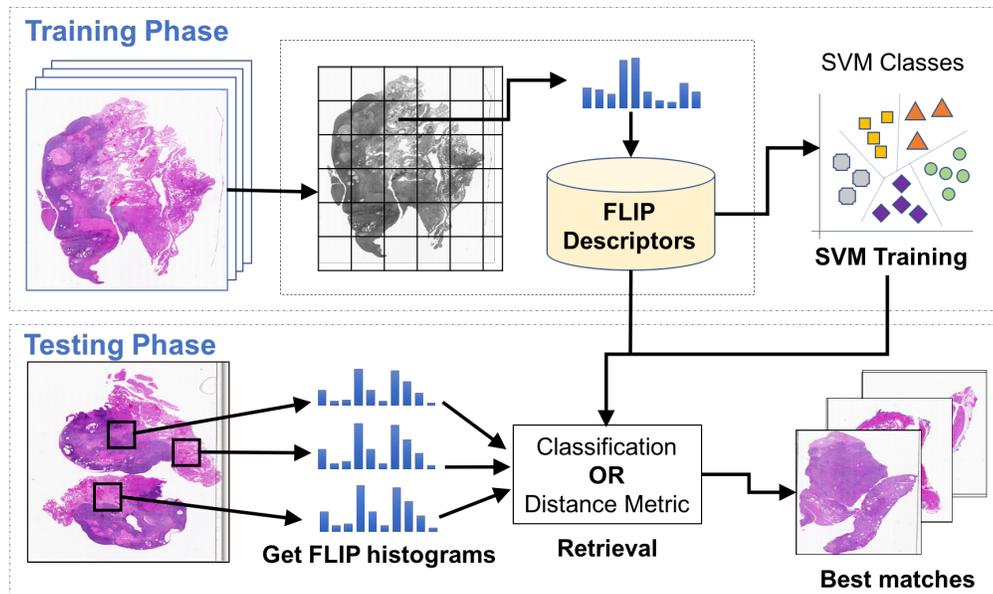
Figure 4.6: Provides a thorough overview of the extracting and testing phase for computing and evaluating the FLIP global descriptors.

Input   : An image **I** as part of a whole scan **S**: $\mathbf{I} \subset \mathbf{S}$
Output: The FLIP histogram **h**
Set neighbourhood size and overlap;
$L \leftarrow 128$ (histogram default length);
$\mathbf{h} \leftarrow \emptyset$;
$\mathbf{F} \leftarrow \emptyset$;
$\mathbf{I}_g \leftarrow$ Convert image **I** to gray-scale;

foreach *window* $\mathbf{W}_i$ *in image* $\mathbf{I}_g$ **do**
    $R_{(0,45,90,135)} \leftarrow$ RadonTransform($\mathbf{W}_i$);
    $R_{\min_1} \leftarrow \min(R_0, R_{45})$;
    $R_{\min_2} \leftarrow \min(R_{45}, R_{90})$;
    $R_{\min_3} \leftarrow \min(R_{90}, R_{135})$;
    $R_{\min_4} \leftarrow \min(R_{135}, R_0)$;
    $R_{\min} \leftarrow$ concatenate $(R_{\min_1}, R_{\min_2}, R_{\min_3}, R_{\min_4})$;
    $\mathbf{F} \leftarrow$ AppendRow($R_{\min}$);
end
$f_{\min}, f_{\max} \leftarrow$ FindMinMax($\mathbf{F}$);
$\mathbf{F} \leftarrow$ reScale($\mathbf{F}, f_{\min}, f_{\max}, L$);
$\mathbf{F} \leftarrow \mathbf{F}[\mathbf{1} : \mathbf{128}]$ (127 length histogram);
for $i = 1$ *to* $\mathbf{F}_{rows}$ **do**
    for $j = 1$ *to* $\mathbf{F}_{cols}$ **do**
        $\mathbf{h}(\mathbf{F}(i,j)) \leftarrow \mathbf{h}(\mathbf{F}(i,j)) + 1$;
    end
end
Return **h**;

Algorithm 3: FLIP – Forming Local Intersections of Projections.

Chapter 5

Experiments and Analysis

This chapter provides a detailed analysis and experimental results for the aforementioned Autoencoded Radon projection framework as well as the novel FLIP and mFLIP global image descriptors applied on their respective benchmark data sets.

5.1 Evaluation of Autoencoded Radon Projections

The following subsection emphasizes on the performance of the proposed autoencoded Radon projections framework on the IRMA data set. In addition to the experimental results, this section provides a overview of the IRMA data set as well as retrieval strategies.

5.1.1 IRMA Radiography Data-set

IRMA is a publicly available radiograph data set collected at the Department of Diagnostic Radiology at the RWTH Aachen University, in Germany [1]. This data set is composed of $12,677$ training and $1,733$ testing images of anonymized patients of different ages and genders. The radiographs are of various resolutions captured at different angular positions and represent multiple body parts (such as: lung, leg, arms, skull, etc.) [110]. The IRMA data set is quite challenging for machine learning algorithms to generalize due to noise, variational backgrounds, rotations, artifacts and nonuniform category distribution, as in the data set is dominated by chest x-rays. In terms or evaluation, each image in this data set has a unique "IRMA code" which belongs to one of the 57 categories (2005 version). The IRMA code computation is provided along with the data set which is intended to provide

confirmation on the validity of the proposed framework, and to compare against literature for benchmarking. In essence, when comparing the query image against the retrieved image, the IRMA code is computed which provides an error between 0 and 1 (0 being the better match). In reality, the IRMA code is not available; in this case, it is designed and numerically computed by medical professionals solely for benchmarking purposes. Hence, the total IRMA error is calculated for all 1,733 test images for which the best match and its associated class is obtained from the training set. The summation of the IRMA code for all 1,733 query images quantifies the performance of the proposed framework, which is then compared to literature. The presented framework, that is autoencoded Radon projections and classifying it using an MLP, sets the benchmark on the IRMA data set when comparing against other AE techniques reported on this data set.

IRMA Challenges

Each medical imaging data set has its challenges which requires exploring different strategies to overcome. The IRMA data set is a particularly difficult data set for machine intelligent algorithms, especially for deep learning approaches. The obvious challenges of the data set include: *(i)* limited samples, *(ii)* nonuniform categorical distribution between the 57 categories, and *(iii)* diverse sources of variable or irrelevant information in the images.

Limited Samples: The IRMA is a rather small data set, with 12,677 and 1,733 training and testing images, respectively. The limited sample of images poses as a threat to deep Learning algorithm, disabling the network from generalizing. This is because, for a network to generalize, a large set of data is to be provided such that feature extraction is improved. However, since the provided data set is not large, an AE is constrained to have a relatively shallow network to avoid overfitting the data. In order to reflect on this problem, data augmentation is implemented – which creates synthetic data of an existing data by flipping the image along its axes and rotating them to obtain multiple combinations of the same image. For completeness, the AE and MLP were also validated on augmented data set, however, the testing accuracy is observed to have decreased. For this reason, the proposed method did not include augmented data results.

Categorical Imbalance: IRMA is a non-uniform data set that is biased towards the first class; the majority of the images in the training set are lung radiographs, as depicted in Fig. 5.1. An imbalance data set causes the network to "memorize" certain images which disallows for generalizing the data set. In this case, it is often observed that deep learning algorithms tend to "over-learn" lung x-rays which causes mis-classification during the retrieval stage.
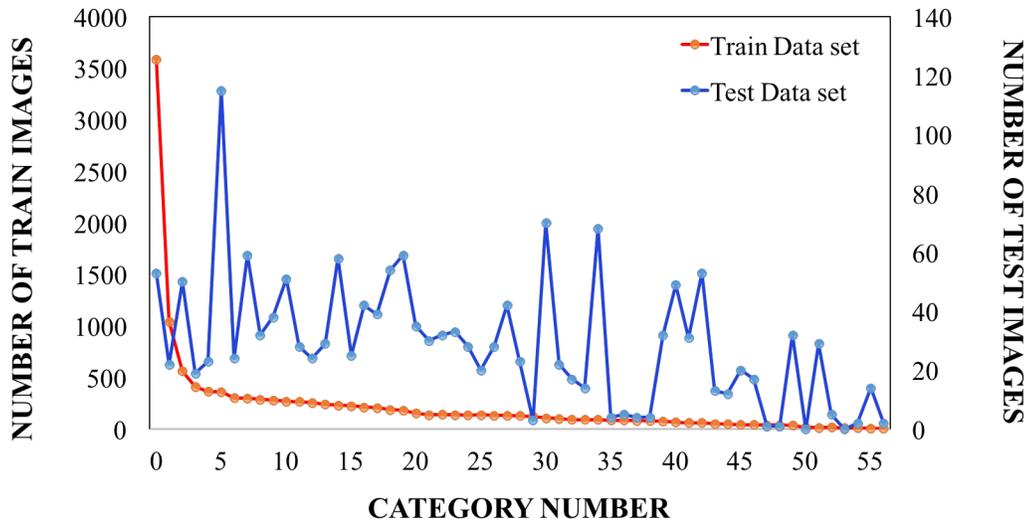
Figure 5.1: Non-uniform categorical distribution of training vs testing data in IRMA data set across 57 classes.

Diverse sources of variable or irrelevant information: To add further complexity for training, the IRMA data set contains numerous challenges which confuses the network during training: variable background color or lack of, artefacts, burnt-in annotations, landmarks and rotations, to mention some. Although these noises impact the performance of the network, they were included in the training and testing phase in order to be comparable against literature. As depicted in Fig. 5.2, there exists various types of problems within the images in the IRMA data set.

Experimental Results

For validating the proposed supervised sequential framework, three series of experiments were performed. These experiments include evaluating the framework using three feature inputs to the AE, i.e., (*i*) Radon projections, (*ii*) HOG features, and (*iii*) raw images. Regardless of the input parameter to the framework, the image size and the feature dependent parameter(s) are altered. The feature dependent parameters includes the number of equi-distant projections for Radon transform, the number of orientations and the bin size for HOG features, and the compression (down-sampling) ratio for raw images. Evidently, the retrieving and the training strategy needs to be of the same type of input feature.

The performance of the proposed framework is validated by calculating the accumulated
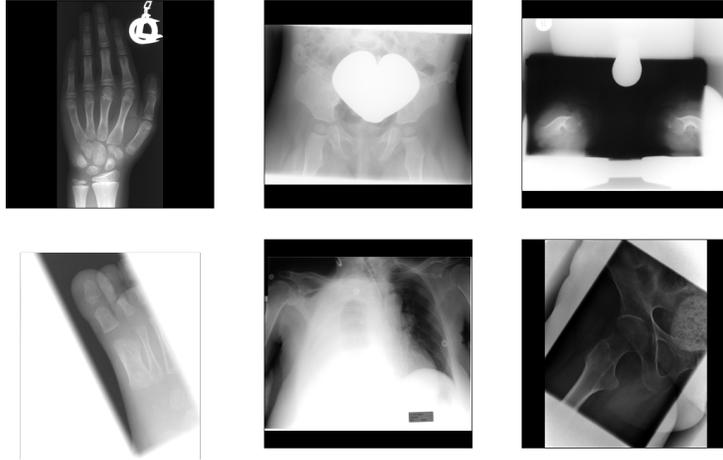
Figure 5.2: Challenging aspects of the IRMA data set: some images have burnt-in annotations, artifacts and landmarks. Different orientations and background intensities also pose challenges [100].

IRMA code for every proper and improper classification. Hence, to obtain the lowest IRMA error, namely when using Radon projections, 57 indexed KD-trees are generated whereas the ith KD-Tree is indexed using 16 equidistant 1D vectorized and normalized Radon projections from pre-processed 256×256 down-sampled training images which belong to the i-the category. The closest match for the provided query image is retrieved based on the minimum distance when comparing its vector representation against all vectors in the training set. Particularly, 16 equidistant projections are chosen based on results presented in [108]. As well, a 256×256 size was chosen empirically to unify the image dimensions to simplify the retrieval task.

For the first experiment, the Radon projections are calculated from pre-processed training images and normalized on a per-projection basis. The normalized Radon projections are then concatenated to form a one-dimensional vector which is provided to the framework. The combination of Radon features and the framework yielded an IRMA error of 313, which is the best result on the IRMA dataset using an AE. For the second series of experiments, HOG descriptors were provided as input to the framework, which obtained an IRMA error of $\approx$ 326 for 8 orientations. For the final experiment, the raw images were down-sampled and provided as input to the framework, resulting in an IRMA error of $\approx$ 349 for 50% compression that is 64×64 sized input images. Table5.4 shows the experimental results - wherein the dark gray cells are the best results and the light gray cells: are the best value for a given input dimension.

41

Table 5.1: Comparative results: For different image sizes (first column), we measured the IRMA error and network accuracies (second column). For Radon projections (third column) four scenarios were tested with 8, 10, 16 and 20 equidistant projections. For HOG features (fourth column), four different gradient histograms were calcutes with 4, 8, 10 and 16 directions. Raw images (last column) were compressed at 25% and 50% [100].

Input size	Error/Accuracy	Radon				HOG				Raw Image	
		8	10	16	20	4	8	10	16	25%	50%
32×32	IRMA Error	345	341	335	340	410	408	408	405	351	354
64×64	IRMA Error	332	332	330	328	329	326	328	331	349	353
128×128	IRMA Error	325	323	323	320	332	329	335	335	349	350
256×256	IRMA Error	317	316	313	314	336	333	341	350	351	356

In conclusion, the proposed supervised sequential framework, which incorporated data compression as well as classification is observed to be competent for medical imaging applications. The lowest IRMA error of 313 was achieved by training a classifier using 16 equidistant normalized Radon projections from pre-processed training images with a 10-fold cross validation, early stopping, and dropout regularizations included. It is evident that Radon projections are descriptive input parameters to the framework outperforming both the HOG feature and raw pixel values. In addition, Radon projections were observed to consume less memory and to be faster ($\approx$ 25GB of RAM for 2.5 hours) when compared to HOG ($\approx$ 30GB of RAM for 3 hours) and raw images ($\approx$ 35GB of RAM for 4.5 hours). Additionally, weighting the retrieval distances using the MLP classification probability is observed to consistently yield better results compared to simply using L-2 distances. Fig. 5.3 shows the effect of weighting of similarity using classification probabilities. The images from left to right are sorted in increasing order of Radon projection similarity, wherein the highlighted image shows the best match after probability weighting.

It is evident that the test accuracy is not necessarily a good estimation of the IRMA error. This is perhaps due to the very high categorical imbalance among training/test data. However, there persists a relationship between the IRMA error and the test accuracy – an IRMA error of 313 is achieved for the highest test accuracy of 57%.

For the sake of completeness, the IRMA training images were also augmented to obtain 70,000 images. This strategy is implemented to confirm if increasing the number of images can assist the learning algorithms to perform better. However, despite the increase in number of training images, the IRMA error was similar to the best achieved.

After several experiments on various other learning parameters, the best activation function was observed to be *ReLU* (rectified linear unit) when compared to *Sigmoid*, *tanh*, and *rmsprop* functions. Also, a dropout rate of 0.2 was established across all models after empirically experimenting on the data set. Table 5.2 shows the relative ranking of

Query Image | Top 5 Similar Cases
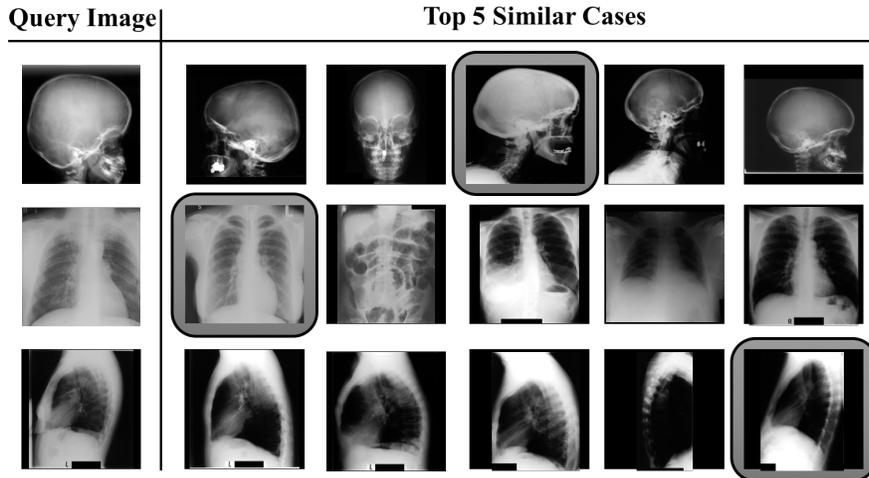
Figure 5.3: Depiction of the "best match" (highlighted) strategy when using Radon projections. Certainly every retrieved image has an associated IRMA code which can be compared with the query's IRMA code to calculate the retrieval error [100].

the proposed approach against existing methods in the literature. Beside the IRMA error ($E_{\text{tot}} \in [0, 1733]$), an accuracy estimate is provided as well: $A = 1 - E_{\text{tot}}/1733$. Even though the proposed method does not yield the least overall error when compared to literature, it is, however, the lowest IRMA error when using an AE for dimensionality reduction. Table 5.2 provides the benchmark IRMA error for completeness, however, the purpose of our framework is to compare only against AE approaches.

Method	Error	Accuracy(%)
Camlica et al. [23]	146.55	91.54%
Avni et al. [7]	169.50	90.22%
Proposed Framework	**313.17**	**81.93%**
Autoencoder – Sze-To et. al [103]	344.08	80.14%
Autoencoder – Sharma et. al [91]	376.00	78.30%
Autoencoder – Tizhoosh et. al [108]	392.09	77.37%

Table 5.2: IRMA results: The lower section compares methods using autoencoders. The upper section reports the best results by non-neural approaches [100].

43

5.2 Evaluation of FLIP Descriptor

This section examines the performance of the FLIP and the mFLIP algorithms, and their evaluation on the openly available KIMIA Path24 histopathology data set. In addition to the discussing the metrics for retrieval and accuracy calculations, this section also discusses the structure and some challenges when training and testing on the KIMIA Path24 data set.

5.2.1 KIMIA Path24 Data-set

KIMIA Path24, introduced by Babaie et al. [9], is a publicly available histopathology data set comprised of 24 WSI depicting various tissue patterns and body parts. Through homogeneity and other statistical metrics, all scans are broken down into 1,325 and 27,055 test and training patches, respectively. These patches were captured from each of the 24 scans to ensure that only specimens are obtained, ignoring the background and other irrelevant parts. Each of the captured patches are of size 1000×1000 are extracted (0.5mm×0.5mm). Specifically, the scans are captured using *TissueScope LE 1.01* in the bright field using a 0.75 NA lens. Fig. 4.1 provides a pictorial representation of randomly selected thumbnails from the KIMIA Path24 data set.

Since FLIP and mFLIP are relatively compact global descriptors, the entire dimension, that is 1000×1000 image, could be utilized and processed (no downsampling necessary). In particular, using NVIDIA Cuda for parallel processing, it takes roughly 20 seconds to obtain the mFLIP for the originally sized image. For deep learning algorithms it becomes time consuming and resource heavy for when the input image is high dimensional. In fact, for each of the deep learning algorithms, the images were down-sampled to 250×250 in order to train the data set. For completeness, the mFLIP and the FLIP were also computed on the down-sampled 250×250 images which resulted in 2∼3% loss in accuracy. Hence, the reported results are for gray-scaled 1000×1000 images as they yield better results.

Accuracy Measurement

A total of $n_{tot} = 1,325$ test patches P_s^j are obtained which belong to one of the 24 classes available $\Gamma_s = \{P_s^i | s \in S, i = 1, 2, \ldots, n_{\gamma_s}\}$ with $s = 0, 1, 2, \ldots, 23$ [9]. In order to compare our retrieval against literature, the retrieval calculation outlined in [9] is adopted. Hence,

for a retrieved image R for any experiment, the patch-to-scan accuracy η_p can be given as:

$$\eta_p = \frac{\sum_{s \in S} |R \cap \Gamma_s|}{n_{tot}} \tag{5.1}$$

In addition, the whole-scan accuracy η_W is calculated to be:

$$\eta_W = \frac{1}{24} \sum_{s \in S} \frac{|R \cap \Gamma_s|}{n_{\gamma_s}} \tag{5.2}$$

Conclusively, the total accuracy η_{total} is obtained which is comprised of both patch-to-scan and whole-scan accuracies: $\eta_{total} = \eta_p \times \eta_W$.

Experimentation on Deep Learning

Since deep learning algorithms have caught much attention in any facet of data prediction and analysis, a performance comparison of mFLIP and FLIP against current established deep networks on the KIMIA Path24 data set is inevitable. The claim behind deep learning is to learn data representations automatically by "training" a network to generalize a data set. Some claim that deep learning network is a very flexible function that we recently learned to fit; others claim that these networks are a smooth lookup table. Regardless, the main idea behind deep learning algorithms is to extract one-level of abstraction per layer, learn these abstractions and fit it to data to generalize. In this subsection, four different deep learning structures were evaluated and compared against the proposed mFLIP descriptor. These deep learning algorithms include: (i) Pre-trained VGG16 classified on linear-SVM, (ii) Fine-tuned VGG16, (iii) Pre-trained Inception V3 classified using linear-SVM, and (iv) Fine-tuned Inception V3. Both the VGG16 and the Inception V3 are convolutional neural networks that are established networks with optimized weights trained thoroughly on a particular data set. The idea here is to leverage a well-established network and optimize its weights to fit the KIMIA Path24 data set.

Pre-Trained CNN as a Feature Extractor: For the first series of experiment, the pre-trained VGG16 and Inception V3 networks are adopted for feature extraction; these networks along with its optimized weights are built into the Keras library in Python. In particular, the fully-connected layers from both the networks are extracted to obtain the feature vector. This feature vector is provided to a linear SVM for classification. As for the linear-SVM classifier, built-in Python packages scikit-learn and LIBSVM were utilized [79] [25]. In addition, Python libraries NumPy [117] and SciPy [50] were leveraged to manipulate and store the data.

Fine-tuned CNN as a Classifier: For the second series of experiment, the established VGG16 and Inception V3 networks are fine-tuned as classifiers; these networks along with their optimized weights are built into the Keras library in Python. For the VGG16 network, the fully-connected layer is removed from the convolutional layers and provided to a bottleneck to extract features through the convolutional layers. These features are initially used to pre-train the new fully-connected MLP layers, hence forcing the network to initialize and optimize its weights. Thereafter, the new fully connected MLP is of a single 256-vector length with ReLU activation function, followed by a softmax layer for obtaining class probabilities. Finally, the new fully-connected model is attached back onto the VGG16 convolutional layers and trained on each convolutional block, except the last block, in order to receive the adjusted classification weights.

Similarly for Inception V3 network, the original fully-connected layer is replaced with a single 1024 dense ReLU layer followed by a softmax layer for classification. Furthermore, the new fully-connected layer is trained on a bottleneck structure to extract features which is then attached back onto the original convolutional layers for fine-tuning the final two inception blocks. Table 5.4 provides the accuracy measures for aforementioned four pre-trained deep learning experiments. It is thereby evident that even the state-of-the-art deep learning algorithm is not a match for the mFLIP global descriptor.

Experimentation with mFLIP descriptor

Multiple experiments were performed with different configurations in form of $\text{mFLIP}_{(L,w,\Delta),D}$ where $L = |\mathbf{h}|$ is the histogram length, w is the window size, Δ is the pixel stride (overlap), and D is the distance measure or classification scheme. Specifically, we experimented with L being 127 and 508 (after removing the first bin), w of 3×3 and 10×10, and Δ of no overlap and full-overlap. The best configuration is empirically chosen to be of $\text{FLIP}_{(127,3,3),svm}$, and $\text{mFLIP}_{(508,3,3),svm}$.

Table 5.3 provides the accuracy measures for various FLIP and mFLIP combinations. When the FLIP is configured with utilizing the original dimensions, with a neighborhood size of 3×3 and $\Delta = 3$ pixel stride and a histogram length of $L = 127$, the best retrieval accuracy (η_{total}) of 55.24% is achieved using SVM classifier (generalized histogram intersection kernel). On the other hand, we obtain a 46.98% accuracy when using histogram intersection distance metric for retrieving - the lowest distance is the best match when comparing histograms. On the other hand, the benchmark score for the KIMIA Path24 is achieved by mFLIP - utilizing four concatenated image dimensions of 1000×1000, 750×750, 500×500, 250×250, with a neighborhood size of 3×3 and $\Delta = 3$ pixel stride and a histogram length of $L = 508$ (each dimension of which gets a FLIP descriptor of 127

concatenated together). The best retrieval accuracy (η_{total}) of 72.42% is achieved using the generalized histogram intersection kernel SVM classifier and a 59.93% accuracy when using χ^2 distance metric.

	F1-Score	Precision	Recall	η_p	η_W	η_{total}
mFLIP$_{(508,3,3),\ \chi^2}$	76.78	77.55	77.28	77.28	77.55	59.93
mFLIP$_{(508,3,3),\ histInt}$	74.24	75.38	74.87	74.87	75.38	56.44
mFLIP$_{(508,3,3),\ svm}$	**84.54**	**85.52**	**84.68**	**84.68**	**85.52**	**72.42**
FLIP$_{(127,3,3),\ \chi^2}$	67.28	68.27	67.62	67.62	68.27	46.16
FLIP$_{(127,3,3),\ histInt}$	67.80	69.03	68.07	68.07	69.03	46.98
FLIP$_{(127,3,3),\ svm}$	73.88	74.54	74.11	74.11	74.54	55.24

Table 5.3: mFLIP and FLIP results for different retrieval strategies (χ^2, histogram intersection, and svm) for a histogram length of $L = 127$, generated using neighborhood size of with no-overlap ($\Delta = 3$). Best results are highlighted in bold.

Table 5.4 provides a comparison of mFLIP against deep learning methods reported in literature. The fine-tuned Inception V3 delivers $\eta_{total} = 56.98$ which is slightly higher than the FLIP accuracy. However, all deep learning approaches are considerably lower when compared to the current benchmark, mFLIP$_{(508,3,3)}$ which achieves a $\eta_p = 85.53$, $\eta_p = 84.68$, and $\eta_{total} = 72.42$. These results are quite encouraging that a handcrafted global descriptor can surpass deep features that are the result of numerous hours/days of training.

Method	η_W	η_p	η_{total}
mFLIP$_{(508,3,3),svm}$	**85.53**	**84.68**	**72.42**
ELP [106]	82.70	79.90	66.07
LBP$^u_{(24,2),svm}$	77.80	73.30	57.02
TL-Inception-v3 [52]	76.10	74.87	56.98
FLIP$_{(127,3,3),svm}$	74.54	74.11	55.24
FE-Inception-v3 [52]	71.24	70.94	50.54
FE-VGG16 [52]	64.96	65.21	42.36
TL-VGG16 [52]	66.23	63.85	42.29
CNN [9]	64.75	64.98	41.80

Table 5.4: FLIP and mFLIP results for the KIMIA Path24 dataset. FLIP uses different retrieval strategies (χ^2, comb, and svm) for a 127 histogram length. Best results for η_p, η_W and η_{total} are highlighted in bold. The results for methods marked are bolded.

Chapter 6

Conclusion

This thesis investigated the uses of Radon transform for deep learning and computer vision techniques to complement CBIR systems. In particular, two contributions were put forward in this thesis. The first contribution is a comparative study of Radon transform used in place of raw images and HOG descriptor as an input for the proposed sequential framework; that is, AE for dimensionality reduction and a multi-layer perceptron for classifying the compressed Radon embedding. This approach is validated on the IRMA data set, which consists of approximately 14,000 radiography images of various body-parts; and the retrieval accuracy of which is computed on the provided IRMA code published by the authors of the data set [1]. The second novelty is a handcrafted global descriptor based on Radon projections called FLIP, as well as an multi-resolution version of the FLIP is discussed. Both the FLIP and mFLIP descriptors compete directly with deep learning algorithms and other established global descriptors. These descriptors are validated on the KIMIA Path24 data set [9], which is a histopathology data set comprised of approximately 29,000 digital scans of size 1000×1000 obtained from 24 classes.

At first, Radon transform is explored as an input feature for the proposed supervised framework, that is AE and MLP trained sequentially for retrieval of medical images. Three features are employed for comparison, namely: Radon projections, HOG features, and raw pixels. In essence, each of these features are provided to an AE which compresses the data to form a latent embedding representative of the image. Thereafter, these embeddings along with its labels are provided to the MLP for classification. When evaluating the proposed framework on the IRMA dataset, it is evident that the dimensionality reduction capabilities of AE tends to remove the redundancies of the input vector; which helps to expedite the training process and avoid over-fitting of data. The lowest IRMA error of 313 is obtained when providing Radon projections as input to the framework. Not only

48

do Radon projections provide the best accuracy, but they also accelerate the training process as Radon features are compact yet descriptive. The aforementioned IRMA error is the lowest IRMA error achieved when using AE for dimensionality reduction and feature extraction.

In this thesis, a new global feature descriptor called FLIP was also introduced along with its extended multi-resolution version, mFLIP - which currently sets the benchmark for KIMIA Path24 data set. The mFLIP descriptor outperforms any deep learning and handcrafted computer vision algorithms reported on the aforementioned data set. It appears that mFLIP is particularly good for histopathology images as the proposed algorithm is observed to capture the texture of each image through the means of Radon projections, and quantify these projections onto a condensed and concise histogram. In addition, the process of localizing and capturing the Radon projections for each neighborhood allows for quick computation and does not require learning or expensive GPUs/CPUs for extracting features from images. Moreover, since both mFLIP and FLIP histograms are descriptive yet compact representation of the image, saving these histograms in place of images immensely expedites the retrieval and classification process. In particular, the mFLIP descriptor of length 508 achieves an η_{total} of $\approx 60\%$ accuracy using χ^2 distance for retrieval, and $\approx 72\%$ when using SVM classification using generalized histogram intersection kernel. This is quite encouraging as mFLIP does not require training and down-sampling, and it generates much more compact descriptors with less storage demand.

Future Work: While this thesis demonstrated the potential of radon projections as image descriptors for CBIR, there are many opportunities to extend the scope of this thesis. The following presents future direction.

Learning Autoencoded Radon Projections

- Train Autoencoder and MLP sequentially, and fine-tune the compressed layers.

- Use SVM in-place of MLP, perhaps it may yield a better accuracy

- Perhaps a greedy layer-wise approach for training an autoencoder

Forming Local Intersections of Projections

- Use filters for calculating local Radon projections

- Create a multi-magnification FLIP that depends on the magnification of the pathology image

- Use curvelet transform for a more descriptive and faster generalization instead of local Radon transform.

References

[1] IRMA (image retrieval in medical applications). https://ganymed.imib.rwth-aachen.de/irma/index_en.php. Accessed: 2017-03-02.

[2] SHARCNET (the shared hierarchical academic research computing network). http://www.sharcnet.ca. Accessed: 2017-29-01.

[3] Mohamed Abdel-Nasser, Jaime Melendez, Antonio Moreno, Osama A Omer, and Domenec Puig. Breast tumor classification in ultrasound images using texture analysis and super-resolution methods. *Engineering Applications of Artificial Intelligence*, 59:84–92, 2017.

[4] Emmanuel Adetiba and Oludayo O Olugbara. Lung cancer prediction using neural network ensemble with histogram of oriented gradient genomic features. *The Scientific World Journal*, 2015.

[5] Dhruv Agarwal, KS Shriram, and Navneeth Subramanian. Automatic view classification of echocardiograms using histogram of oriented gradients. In *IEEE 10th International Symposium on Biomedical Imaging (ISBI)*, pages 1368–1371. IEEE, 2013.

[6] Timo Ahonen, Abdenour Hadid, and Matti Pietikäinen. Face recognition with local binary patterns. *European Conference on Computer Vision*, pages 469–481, 2004.

[7] Uri Avni, Jacob Goldberger, and Hayit Greenspan. Addressing the imageclef 2009 challenge using a patch-based visual words representation. In *The Cross-Language Evaluation Forum (CLEF)*, 2009.

[8] Paulo Mazzoncini de Azevedo-Marques and Rangaraj Mandayam Rangayyan. Content-based retrieval of medical images: landmarking, indexing, and relevance feedback. *Synthesis Lectures on Biomedical Engineering*, 8(1):1–143, 2013.

[9] Morteza Babaie, Shivam Kalra, Aditya Sriram, Christopher Mitcheltree, Shujin Zhu, Amin Khatami, Shahryar Rahnamayan, and Hamid R Tizhoosh. Classification and retrieval of digital pathology scans: A new dataset. In *Computer Vision for Microscopy Image Analysis (CVMI) Workshop*, 2017.

[10] Morteza Babaie, Hamid R Tizhoosh, Shujin Zhu, and ME Shiri. Retrieving similar x-ray images from big image data using radon barcodes with single projections. *arXiv preprint arXiv:1701.00449*, 2017.

[11] Morteza Babaie, HR Tizhoosh, Amin Khatami, and ME Shiri. Local radon descriptors for image search. *arXiv preprint arXiv:1710.04097*, 2017.

[12] Artem Babenko and Victor Lempitsky. Aggregating local deep features for image retrieval. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1269–1277, 2015.

[13] Yaniv Bar, Idit Diamant, Lior Wolf, and Hayit Greenspan. Deep learning with non-medical training used for chest pathology identification. In *Proceedings of Society of Photographic Instrumentation Engineers (SPIE)*, volume 9414, page 94140V, 2015.

[14] Yaniv Bar, Idit Diamant, Lior Wolf, Sivan Lieberman, Eli Konen, and Hayit Greenspan. Chest pathology detection using deep learning with non-medical training. In *IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pages 294–297. IEEE, 2015.

[15] Tudor Barbu. Svm-based human cell detection technique using histograms of oriented gradients. *17th WSEAS International Conference on Applied Mathematics (AMATH)*, 4:11, 2012.

[16] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. *European Conference on Computer Vision*, pages 404–417, 2006.

[17] Neslihan Bayramoglu, Juho Kannala, and Janne Heikkilä. Deep learning for magnification independent breast cancer histopathology image classification. In *23rd International Conference on Pattern Recognition (ICPR)*, pages 2440–2445. IEEE, 2016.

[18] Yoshua Bengio, Aaron C Courville, and Pascal Vincent. Unsupervised feature learning and deep learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 1, 2013.

[19] Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the 5th Annual Workshop on Computational Learning Theory*, pages 144–152. ACM, 1992.

[20] Sheryl Brahnam, Lakhmi C Jain, Loris Nanni, Alessandra Lumini, et al. *Local binary patterns: new variants and applications*. Springer, 2014.

[21] Blaž Bratanič, Franjo Pernuš, Boštjan Likar, and Dejan Tomaževič. Real-time rotation estimation using histograms of oriented gradients. *Public Library of Science (PloS) one*, 9(3):e92137, 2014.

[22] Zehra Camlica, Hamid R Tizhoosh, and Farzad Khalvati. Autoencoding the retrieval relevance of medical images. In *International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pages 550–555. IEEE, 2015.

[23] Zehra Camlica, HR Tizhoosh, and Farzad Khalvati. Medical image classification via svm using lbp features from saliency-based folded data. *arXiv preprint arXiv:1509.04619*, 2015.

[24] Chi-Ho Chan, Josef Kittler, and Kieron Messer. Multi-scale local binary pattern histograms for face recognition. *Advances in Biometrics*, pages 809–818, 2007.

[25] Chih-Chung Chang and Chih-Jen Lin. Libsvm: a library for support vector machines. *ACM transactions on Intelligent Systems and Technology (TIST)*, 2(3):27, 2011.

[26] Yun Wen Chen and Yan Qiu Chen. Invariant description and retrieval of planar shapes using radon composite features. *IEEE Transactions on Signal Processing*, 56(10):4762–4771, 2008.

[27] J Shiny Christobel and C Vinoth Kumar. Feature based breast cancer detection using mammographic images. *Indian Journal of Scientific Research(IJSR)*, pages 257–262, 2017.

[28] Dan C Cireşan, Alessandro Giusti, Luca M Gambardella, and Jürgen Schmidhuber. Mitosis detection in breast cancer histology images with deep neural networks. In *International Conference on Medical Image Computing and Computer-assisted Intervention*, pages 411–418. Springer, 2013.

[29] Ricardo da Silva Torres and Alexandre X Falcao. Content-based image retrieval: theory and applications. *Journal of Theoretical and Applied Informatics (RITA)*, 13(2):161–185, 2006.

[30] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 886–893. IEEE, 2005.

[31] Navneet Dalal, Bill Triggs, and Cordelia Schmid. Human detection using oriented histograms of flow and appearance. In *European Conference on Computer Vision*, pages 428–441. Springer, 2006.

[32] Petros Daras, Dimitrios Zarpalas, Dimitrios Tzovaras, and Michael G Strintzis. Efficient 3-d model search and retrieval using generalized 3-d radon transforms. *IEEE Transactions on Multimedia*, 8(1):101–114, 2006.

[33] Stanley R Deans. *The Radon transform and some of its applications.* Courier Corporation, 2007.

[34] Li Deng. A tutorial survey of architectures, algorithms, and applications for deep learning. *Transactions on Signal and Information Processing*, 2014.

[35] Shiv Ram Dubey, Satish Kumar Singh, and Rajat Kumar Singh. Local diagonal extrema pattern: a new and efficient feature descriptor for ct image retrieval. *IEEE Signal Processing Letters*, 22(9):1215–1219, 2015.

[36] Tristan Glatard, Johan Montagnat, and Isabelle E Magnin. Texture based medical image indexing and retrieval: application to cardiac imaging. In *Proceedings of International Workshop on Multimedia Information Retrieval*, pages 135–142, 2004.

[37] Soren Goyal and Paul Benjamin. Object recognition using deep neural networks: A survey. *arXiv:1412.3684*, 2014.

[38] Hayit Greenspan and Adi T Pinhas. Medical image categorization and retrieval for pacs using the gmm-kl framework. *IEEE Transactions on Information Technology in Biomedicine*, 11(2):190–202, 2007.

[39] Hayit Greenspan, Bram van Ginneken, and Ronald M Summers. Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. *IEEE Transactions on Medical Imaging*, 35(5):1153–1159, 2016.

[40] Chris Harris and Mike Stephens. A combined corner and edge detector. In *Alvey Vision Conference*, volume 15, pages 147–151. Manchester, UK, 1988.

[41] Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 2006.

[42] PS Hiremath and Jagadeesh Pujari. Content based image retrieval using color, texture and shape features. In *International Conference on Advanced Computing and Communications (ADCOM)*, pages 780–784. IEEE, 2007.

[43] Thai V Hoang and Salvatore Tabbone. Invariant pattern recognition using the rfm descriptor. *Pattern Recognition*, 45(1):271–284, 2012.

[44] Thomas Hofmann, Bernhard Schölkopf, and Alexander J Smola. Kernel methods in machine learning. *The Annals of Statistics*, pages 1171–1220, 2008.

[45] Dattatray V Jadhav and Raghunath S Holambe. Radon and discrete cosine transforms based feature extraction and dimensionality reduction approach for face recognition. *Signal Processing*, 88(10):2604–2609, 2008.

[46] Dattatray V Jadhav and Raghunath S Holambe. Feature extraction using radon and wavelet transforms with application to face recognition. *Neurocomputing*, 72(7):1951–1959, 2009.

[47] Anil K Jain and Farshid Farrokhnia. Unsupervised texture segmentation using gabor filters. *Pattern Recognition*, 24(12):1167–1186, 1991.

[48] Andrew Janowczyk and Anant Madabhushi. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *Journal of Pathology Informatics*, 7, 2016.

[49] Hervé Jégou, Matthijs Douze, Cordelia Schmid, and Patrick Pérez. Aggregating local descriptors into a compact image representation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3304–3311. IEEE, 2010.

[50] Eric Jones, Travis Oliphant, and Pearu Peterson. {SciPy}: open source scientific tools for {Python}. https://www.scipy.org/, 2014.

[51] Amin Khatami, Morteza Babaie, Abbas Khosravi, HR Tizhoosh, Syed Moshfeq Salaken, and Saeid Nahavandi. A deep-structural medical image classification for a radon-based image retrieval. In *IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)*, pages 1–4, 2017.

[52] Brady Kieffer, Morteza Babaie, Shivam Kalra, and HR Tizhoosh. Convolutional neural networks for histopathology image classification: Training vs. using pre-trained networks. *arXiv preprint arXiv:1710.05726*, 2017.

[53] Seong-Hoon Kim, Ji-Hyun Lee, Byoungchul Ko, and Jae-Yeal Nam. X-ray image classification using random forests with local binary patterns. In *2010 International Conference on Machine Learning and Cybernetics (ICMLC)*, volume 6, pages 3190–3194. IEEE, 2010.

[54] Josef Kittler, Mohamad Hatef, Robert PW Duin, and Jiri Matas. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3):226–239, 1998.

[55] Byoung Chul Ko, Seong Hoon Kim, and Jae-Yeal Nam. X-ray image classification using random forests with local wavelet-based cs-local binary patterns. *Journal of Digital Imaging*, 24(6):1141–1151, 2011.

[56] ByoungChul Ko, Soo Yeong Kwak, and Hyeran Byun. Svm-based salient region (s) extraction method for image retrieval. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 2, pages 977–980. IEEE, 2004.

[57] Alex Krizhevsky and Geoffrey E Hinton. Using very deep autoencoders for content-based image retrieval. In *European Symposium on Artificial Neural Network*, 2011.

[58] Devinder Kumar, Alexander Wong, and David A Clausi. Lung nodule classification using deep features in ct images. In *12th Conference on Computer and Robot Vision (CRV)*, pages 133–138. IEEE, 2015.

[59] Soorya S Kumar and Jiji CV. Histogram of radon projections: A new descriptor for object detection. *National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics*, 2015.

[60] Thomas M Lehmann, MO Gold, Christian Thies, Benedikt Fischer, Klaus Spitzer, Daniel Keysers, Hermann Ney, Michael Kohnen, Henning Schubert, and Berthold B Wein. Content-based image retrieval in medical applications. *Methods of Information in Medicine*, 43(4):354–361, 2004.

[61] Zhi Li, Guizhong Liu, Yang Yang, and Junyong You. Scale-and rotation-invariant local binary pattern using scale-adaptive texton and subuniform-based circular shift. *IEEE Transactions on Image Processing*, 21(4):2130–2140, 2012.

[62] Znaonui Liang, Gang Zhang, Jimmy Xiangji Huang, and Qmming Vivian Hu. Deep learning for healthcare decision making with emrs. In *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 556–559. IEEE, 2014.

[63] Shu Liao, Max WK Law, and Albert CS Chung. Dominant local binary patterns for texture classification. *IEEE Transactions on Image Processing*, 18(5):1107–1118, 2009.

[64] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen AWM van der Laak, Bram van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *arXiv preprint arXiv:1702.05747*, 2017.

[65] Geert Litjens, Clara I Sánchez, Nadya Timofeeva, Meyke Hermsen, Iris Nagtegaal, Iringo Kovacs, Christina Hulsbergen-Van De Kaa, Peter Bult, Bram Van Ginneken, and Jeroen Van Der Laak. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Scientific Reports*, 6:26286, 2016.

[66] Xinran Liu, Hamid R Tizhoosh, and Jonathan Kofman. Generating binary tags for fast medical image retrieval based on convolutional nets and radon transform. In *International Joint Conference on Neural Networks (IJCNN)*, pages 2872–2878, 2016.

[67] Ying Liu, Dengsheng Zhang, Guojun Lu, and Wei-Ying Ma. A survey of content-based image retrieval with high-level semantics. *Pattern Recognition*, 40(1):262–282, 2007.

[68] Yu-Ying Liu, Mei Chen, Hiroshi Ishikawa, Gadi Wollstein, Joel S Schuman, and James M Rehg. Automated macular pathology diagnosis in retinal oct images using multi-scale spatial pyramid and local binary patterns in texture and shape encoding. *Medical Image Analysis*, 15(5):748–759, 2011.

[69] Jakub Lokoč, David Novák, Michal Batko, and Tomáš Skopal. Visual image search: feature signatures or/and global descriptors. In *International Conference on Similarity Search and Applications*, pages 177–191. Springer, 2012.

[70] David G Lowe. Object recognition from local scale-invariant features. In *The Proceedings of 7th IEEE International Conference on Computer Vision*, volume 2, pages 1150–1157. IEEE, 1999.

[71] Luis Montesano and Manuel Lopes. Learning grasping affordances from local visual descriptors. In *IEEE 8th International Conference on Development and Learning*, pages 1–6. IEEE, 2009.

[72] Henning Müller, Nicolas Michoux, David Bandon, and Antoine Geissbuhler. A review of content-based image retrieval systems in medical applicationsclinical benefits and future directions. *International Journal of Medical Informatics*, 73(1):1–23, 2004.

[73] Nafaa Nacereddine, Salavatore Tabbone, Djemel Ziou, and Latifa Hamami. Shape-based image retrieval using a new descriptor based on the radon and wavelet transforms. In *20th International Conference on Pattern Recognition (ICPR)*, pages 1997–2000. IEEE, 2010.

[74] Loris Nanni, Alessandra Lumini, and Sheryl Brahnam. Local binary patterns variants as texture descriptors for medical image analysis. *Artificial intelligence in Medicine*, 49(2):117–125, 2010.

[75] Timo Ojala, Matti Pietikainen, and David Harwood. Performance evaluation of texture measures with classification based on kullback discrimination of distributions. In *International Conference on Pattern Recognition*, volume 1, pages 582–585. IEEE, 1994.

[76] Timo Ojala, Matti Pietikäinen, and David Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29(1):51–59, 1996.

[77] Timo Ojala, Matti Pietikainen, and Topi Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002.

[78] Nikhil R Pal and Sankar K Pal. A review on image segmentation techniques. *Pattern Recognition*, 26(9):1277–1294, 1993.

[79] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12(Oct):2825–2830, 2011.

[80] Matti Pietikäinen, Abdenour Hadid, Guoying Zhao, and Timo Ahonen. Local binary patterns for still images. In *Computer vision using local binary patterns*, pages 13–47. Springer, 2011.

[81] Tomaso Poggio, Hrushikesh Mhaskar, Lorenzo Rosasco, Brando Miranda, and Qianli Liao. Why and when can deep-but not shallow-networks avoid the curse of dimensionality: A review. *International Journal of Automation and Computing*, 14(5):503–519, 2017.

[82] Victor Pomponiu, Harishwaran Hariharan, Bin Zheng, and David Gur. Improving breast mass detection using histogram of oriented gradients. In *Medical Imaging 2014: Computer-Aided Diagnosis*, volume 9035, page 90351R. International Society for Optics and Photonics, 2014.

[83] Gerard Pons, Robert Martí, Sergi Ganau, Melcior Sentís, and Joan Martí. Computerized detection of breast lesions using deformable part models in ultrasound images. *Ultrasound in Medicine & Biology*, 40(9):2252–2264, 2014.

[84] Hammad Qureshi, Olcay Sertel, Nasir Rajpoot, Roland Wilson, and Metin Gurcan. Adaptive discriminant wavelet packet transform and local binary patterns for meningioma subtype classification. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 196–204, 2008.

[85] Johann Radon. On the determination of functions from their integral values along certain manifolds. *IEEE Transactions on Medical Imaging*, 5(4):170–176, 1986.

[86] Hitesh Rajput, Tanmoy Som, and Soumitra Kar. Using radon transform to recognize skewed images of vehicular license plates. *IEEE Computer Society*, 49(1):59–65, 2016.

[87] Peter M Roth, Martin Hirzer, Martin Koestinger, Csaba Beleznai, and Horst Bischof. Mahalanobis distance learning for person re-identification. In *Person Re-Identification*, pages 247–267. Springer, 2014.

[88] Jorge LC Sanz, Eric B Hinkle, and Anil Jain. *Radon and projection transform-based computer vision: algorithms, a pipeline architecture, and industrial applications*, volume 16. Springer Science & Business Media, 2013.

[89] Ivan W Selesnick, Richard G Baraniuk, and Nick C Kingsbury. The dual-tree complex wavelet transform. *IEEE Signal Processing Magazine*, 22(6):123–151, 2005.

[90] Jin S Seo, Jaap Haitsma, Ton Kalker, and Chang D Yoo. A robust image fingerprinting system using the radon transform. *Signal Processing: Image Communication*, 19(4):325–339, 2004.

[91] S Sharma, I Umar, L Ospina, D Wong, and HR Tizhoosh. Stacked autoencoders for medical image search. In *International Symposium on Visual Computing*, pages 45–54. Springer, 2016.

[92] Hoo-Chang Shin, Holger R Roth, Mingchen Gao, Le Lu, Ziyue Xu, Isabella Nogues, Jianhua Yao, Daniel Mollura, and Ronald M Summers. Deep convolutional neural

networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging*, 35(5):1285–1298, 2016.

[93] Mei-Ling Shyu, Shu-Ching Chen, Min Chen, Chengcui Zhang, and Kanoksri Sarinnapakorn. Image database retrieval utilizing affinity relationships. In *Proceedings of the 1st ACM International Workshop on Multimedia databases*, pages 78–85. ACM, 2003.

[94] John R Smith and Shih-Fu Chang. Automated binary texture feature sets for image retrieval. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 4, pages 2239–2242. IEEE, 1996.

[95] Yang Song, Weidong Cai, Yun Zhou, and David Dagan Feng. Feature-based image patch approximation for lung tissue classification. *IEEE Transactions on Medical Imaging*, 32(4):797–808, 2013.

[96] Lauge Sørensen, Saher Shaker, and Marleen De Bruijne. Texture classification in lung ct using local binary patterns. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 934–941, 2008.

[97] Lauge Sorensen, Saher B Shaker, and Marleen De Bruijne. Quantitative analysis of pulmonary emphysema using local binary patterns. *IEEE Transactions on Medical Imaging*, 29(2):559–569, 2010.

[98] Fabio Alexandre Spanhol, Luiz S Oliveira, Caroline Petitjean, and Laurent Heutte. Breast cancer histopathological image classification using convolutional neural networks. In *International Joint Conference on Neural Networks (IJCNN)*, pages 2560–2567. IEEE, 2016.

[99] Fabio Alexandre Spanhol, Luiz S Oliveira, Caroline Petitjean, and Laurent Heutte. Breast cancer histopathological image classification using convolutional neural networks. In *2016 International Joint Conference on Neural Networks (IJCNN)*, pages 2560–2567. IEEE, 2016.

[100] Aditya Sriram, Shivam Kalra, HR Tizhoosh, and Shahryar Rahnamayan. Learning autoencoded radon projections. *arXiv preprint arXiv:1710.01247*, 2017.

[101] Aditya Sriram, Hamid R Tizhoosh, Shivam Kalra, Brady Kieffery, and Shahryar Rahnamayan. Forming local intersections of projections for retrieval of histopathology images. *Submitted to 15th European Conference on Computer Vision (ECCV)*, 2018.

[102] Yongqing Sun and Shinji Ozawa. A hierarchical approach for region-based image retrieval. In *IEEE International Conference on Systems, Man and Cybernetics*, volume 1, pages 1117–1124. IEEE, 2004.

[103] Antonio Sze-To, Hamid R Tizhoosh, and Andrew KC Wong. Binary codes for tagging x-ray images via deep de-noising autoencoders. *arXiv preprint arXiv:1604.07060*, 2016.

[104] Salvatore Tabbone, Oriol Ramos Terrades, and Sabine Barrat. Histogram of radon transform. a useful descriptor for shape retrieval. In *19th International Conference on Pattern Recognition (ICPR)*, 2008.

[105] Deepti Tamrakar and Kapil Ahuja. Density-wise two stage mammogram classification using texture exploiting descriptors. *arXiv preprint arXiv:1701.04010*, 2017.

[106] Hamid Tizhoosh and Morteza Babaie. Representing medical images with encoded local projections. *IEEE Transactions on Biomedical Engineering*, 2018.

[107] Hamid R Tizhoosh. Barcode annotations for medical image retrieval: A preliminary investigation. In *IEEE International Conference on Image Processing (ICIP)*, pages 818–822. IEEE, 2015.

[108] Hamid R Tizhoosh, Christopher Mitcheltree, Shujin Zhu, and Shamak Dutta. Barcodes for medical image retrieval using autoencoded radon transform. *arXiv preprint arXiv:1609.05112*, 2016.

[109] HR Tizhoosh, Shujin Zhu, Hanson Lo, Varun Chaudhari, and Tahmid Mehdi. Min-max radon barcodes for medical image retrieval. In *International Symposium on Visual Computing*, pages 617–627. Springer, 2016.

[110] Tatiana Tommasi, Barbara Caputo, Petra Welter, Mark Oliver Güld, and Thomas M Deserno. Overview of the clef 2009 medical image annotation track. In *Workshop of the Cross-Language Evaluation Forum for European Languages*, pages 85–93. Springer, 2009.

[111] Devrim Unay, Ahmet Ekin, Mujdat Cetin, Radu Jasinschi, and Aytul Ercil. Robustness of local binary patterns in brain mr image analysis. In *International Conference of Engineering in Medicine and Biology Society*, pages 2098–2101. IEEE, 2007.

[112] Devrim Unay, Ahmet Ekin, and Radu Jasinschi. Medical image search and retrieval using local binary patterns and klt feature points. In *International Conference on Image Processing*, pages 997–1000. IEEE, 2008.

[113] Devrim Unay, Ahmet Ekin, and Radu S Jasinschi. Local structure-based region-of-interest retrieval in brain mr images. *IEEE Transactions on Information technology in Biomedicine*, 14(4):897–903, 2010.

[114] Joost Van De Weijer and Cordelia Schmid. Coloring local feature extraction. *European Conference on Computer Vision*, pages 334–348, 2006.

[115] Surender Varma, Biplab Banerjee, et al. Representation of desert sand dunes by surface orientations using radon transform. In *International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 2733–2736. IEEE, 2012.

[116] Manisha Verma and Balasubramanian Raman. Center symmetric local binary co-occurrence pattern for texture, face and bio-medical image retrieval. *Journal of Visual Communication and Image Representation*, 32:224–236, 2015.

[117] Stéfan van der Walt, S Chris Colbert, and Gael Varoquaux. The numpy array: a structure for efficient numerical computation. *Computing in Science & Engineering*, 13(2):22–30, 2011.

[118] Ji Wan, Dayong Wang, Steven Chu Hong Hoi, Pengcheng Wu, Jianke Zhu, Yongdong Zhang, and Jintao Li. Deep learning for content-based image retrieval: A comprehensive study. In *Proceedings of 22nd International Conference on Multimedia*, pages 157–166. ACM, 2014.

[119] Dayong Wang, Aditya Khosla, Rishab Gargeya, Humayun Irshad, and Andrew H Beck. Deep learning for identifying metastatic breast cancer. *arXiv preprint arXiv:1606.05718*, 2016.

[120] Yan Xu, Tao Mo, Qiwei Feng, Peilin Zhong, Maode Lai, I Eric, and Chao Chang. Deep learning of feature representation with multiple instance learning for medical image analysis. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1626–1630. IEEE, 2014.

[121] Yanwu Xu, Jiang Liu, Ngan Meng Tan, Beng Hai Lee, Damon Wing Kee Wong, Mani Baskaran, Shamira A Perera, and Tin Aung. Anterior chamber angle classification using multiscale histograms of oriented gradients for glaucoma subtype identification. In *International Conference of Engineering in Medicine and Biology Society (EMBC)*, pages 3167–3170. IEEE, 2012.

[122] Hanxuan Yang, Ling Shao, Feng Zheng, Liang Wang, and Zhan Song. Recent advances and trends in visual tracking: A review. *Neurocomputing*, 74(18):3823–3831, 2011.

[123] Dong Yu and Li Deng. Deep learning and its applications to signal and information processing. *IEEE Signal Processing Magazine*, 28(1):145–154, 2011.

[124] Qiang Zhu, Mei-Chen Yeh, Kwang-Ting Cheng, and Shai Avidan. Fast human detection using a cascade of histograms of oriented gradients. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 1491–1498. IEEE, 2006.

[125] Shujin Zhu and HR Tizhoosh. Radon features and barcodes for medical image retrieval via svm. *arXiv preprint arXiv:1604.04675*, 2016.