Computer-aided Cancer Diagnosis and Grading via Sparse Directional Image Representations

Hadi Rezaeilouyeh
University of Denver

Follow this and additional works at: https://digitalcommons.du.edu/etd
Part of the Biomedical Devices and Instrumentation Commons, and the Other Engineering Commons

Recommended Citation
https://digitalcommons.du.edu/etd/1369

This Dissertation is brought to you for free and open access by the Graduate Studies at Digital Commons @ DU. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Digital Commons @ DU. For more information, please contact jennifer.cox@du.edu,dig-commons@du.edu.
Computer-aided Cancer Diagnosis and Grading via Sparse Directional Image Representations

A Dissertation
Presented to
the Faculty of the Daniel Felix Ritchie School of Engineering and Computer Science
University of Denver

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

by
Hadi Rezaeilouyeh
November 2017
Advisor: Dr. Mohammad H. Mahoor
Abstract

Prostate cancer and breast cancer are the second cause of death among cancers in males and females, respectively. If not diagnosed, prostate and breast cancers can spread and metastasize to other organs and bones and make it impossible for treatment. Hence, early diagnosis of cancer is vital for patient survival. Histopathological evaluation of the tissue is used for cancer diagnosis. The tissue is taken during biopsies and stained using hematoxylin and eosin (H&E) stain. Then a pathologist looks for abnormal changes in the tissue to diagnose and grade the cancer. This process can be time-consuming and subjective. A reliable and repetitive automatic cancer diagnosis method can greatly reduce the time while producing more reliable results. The scope of this dissertation is developing computer vision and machine learning algorithms for automatic cancer diagnosis and grading methods with accuracy acceptable by the expert pathologists.

Automatic image classification relies on feature representation methods. In this dissertation we developed methods utilizing sparse directional multiscale transforms—specifically shearlet transform—for medical image analysis. We particularly designed theses computer visions-based algorithms and methods to work with H&E images and MRI images. Traditional signal processing methods (e.g. Fourier transform, wavelet transform, etc.) are not suitable for detecting carcinoma cells due to their lack of directional sensitivity. However, shearlet transform has inherent directional
sensitivity and multiscale framework that enables it to detect different edges in the tissue images. We developed techniques for extracting holistic and local texture features from the histological and MRI images using histogram and co-occurrence of shearlet coefficients, respectively. Then we combined these features with the color and morphological features using multiple kernel learning (MKL) algorithm and employed support vector machines (SVM) with MKL to classify the medical images.

We further investigated the impact of deep neural networks in representing the medical images for cancer detection. The aforementioned engineered features have a few limitations. They lack generalizability due to being tailored to the specific texture and structure of the tissues. They are time-consuming and expensive and need prepossessing and sometimes it is difficult to extract discriminative features from the images. On the other hand, feature learning techniques use multiple processing layers and learn feature representations directly from the data. To address these issues, we first developed a deep neural network containing multiple layers of convolution, max-pooling, and fully connected layers, trained on the Red, Green, and Blue (RGB) images along with the magnitude and phase of shearlet coefficients. Then we developed a weighted decision fusion deep neural network that assigns weights on the output probabilities and update those weights via backpropagation. The final decision was a weighted sum of the decisions from the RGB, and the magnitude and the phase of shearlet networks. We used the trained networks for classification of benign and malignant H&E images and Gleason grading. Our experimental results show that our proposed methods based on feature engineering and feature learning outperform the state-of-the-art in terms of classification accuracy, sensitivity, specificity, F1 score, and area under the curve (AUC) and hence are promising computer-based methods for cancer diagnosis and grading using images.
Acknowledgements

This dissertation is dedicated to my parents for their support, dedication, and inspiration for me.

I would like to express great appreciation to my research advisor, Dr. Mohammad H. Mahoor, for his mentorship, support and patience. Dr. Mahoor helped me develop the required skills to successfully implement new ideas and develop methods during this research.

I would like to thank my dissertation committee members Drs. Ronald Delyser, Matthew Rutherford, and Kateri McRae for reviewing this dissertation and their valuable feedback.

I am grateful for Dr. Kourosh Jafari-Khouzani for sharing his Gleason grading dataset with me. This made it possible for me to evaluate my developed methods for automatic Gleason grading.

I appreciate Drs. Francisco G. La Rosa, Samuel Chang, and Priya N Werahera for sharing their prostate cancer data with us and for their valuable discussions. It made it feasible for me to evaluate my developed methods for automatic cancer diagnosis in histological and MR images.

I would like to thank my fellow graduate students Behzad Hasani, Ali Mollahosseini, and Pooran Negi for their help and camaraderie.

I acknowledge that this research was partially supported by grant IIP- 1230556 from the National Science Foundation.
# Table of Contents

Acknowledgements .......................... iv  
List of Tables .............................. vii  
List of Figures .............................. viii  
List of Symbols ............................ xi  

1 Introduction ................................ 1  

2 Literature review ......................... 8  
   2.1 Breast cancer diagnosis .................. 8  
   2.2 Prostate cancer diagnosis ............... 11  
   2.3 Gleason grading ........................ 13  

3 Sparse directional representation systems 16  
   3.1 Applied harmonic analysis ............... 16  
   3.2 Shearlet transform ........................ 21  

4 Proposed methods .......................... 28  
   4.1 Feature engineering ........................ 28  
       4.1.1 Histogram of Shearlet Coefficients (HSC) .... 28  
       4.1.2 Co-occurrence of Shearlet Coefficients (CSC) .... 30  
       4.1.3 Feature fusion using multiple kernel learning algorithm .... 31  
   4.2 Feature learning .......................... 35  
       4.2.1 Medical image classification via deep neural networks and shearlet transform .... 35  
       4.2.2 A weighted decision fusion framework for automatic Gleason grading using shearlet transform and deep learning .... 39  

5 Experiments and Results .................... 47  
   5.1 Based on feature engineering ............... 48  
       5.1.1 Breast cancer diagnosis using histogram of shearlet coefficients .... 48  
       5.1.2 Prostate cancer diagnosis and Gleason grading via histogram of shearlet coefficients .... 53
5.1.3 Automatic Gleason grading via shearlet transform and multiple kernel learning ................................. 55
5.1.4 Prostate cancer detection in Magnetic Resonance (MR) images 69
5.2 Based on feature learning ........................................... 74
5.2.1 Breast cancer diagnosis and prostate Gleason grading via deep learning .......................................... 74
5.2.2 Automatic Gleason grading via weighted decision fusion framework .............................................. 87

6 Conclusions ................................................................. 97
6.1 Conclusions ............................................................. 97
6.2 Future research ......................................................... 98
6.3 Publications and Patent .............................................. 99
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Confusion matrix</td>
<td>48</td>
</tr>
<tr>
<td>5.2</td>
<td>Average classification accuracy for breast cancer diagnosis using DSC and HSC methods</td>
<td>52</td>
</tr>
<tr>
<td>5.3</td>
<td>Comparing our breast cancer diagnosis classification results with state-of-the-art</td>
<td>53</td>
</tr>
<tr>
<td>5.4</td>
<td>Classification results for prostate cancer diagnosis</td>
<td>55</td>
</tr>
<tr>
<td>5.5</td>
<td>Gleason grading classification results using shearlet features</td>
<td>66</td>
</tr>
<tr>
<td>5.6</td>
<td>Gleason grading classification results using color channel histograms</td>
<td>67</td>
</tr>
<tr>
<td>5.7</td>
<td>Gleason grading classification results using morphological features</td>
<td>67</td>
</tr>
<tr>
<td>5.8</td>
<td>Gleason grading classification results using all of features</td>
<td>68</td>
</tr>
<tr>
<td>5.9</td>
<td>Gleason grading classification results comparing with state-of-the-art</td>
<td>69</td>
</tr>
<tr>
<td>5.10</td>
<td>Prostate cancer classification results for Mp-MRI</td>
<td>74</td>
</tr>
<tr>
<td>5.11</td>
<td>Classification results for breast cancer diagnosis using shearlet transform and deep learning (mean±std)</td>
<td>81</td>
</tr>
<tr>
<td>5.12</td>
<td>Classification results for Gleason grading using shearlet transform and deep learning (mean±std)</td>
<td>82</td>
</tr>
<tr>
<td>5.13</td>
<td>Classification results for Gleason grading using weighted decision fusion and deep learning (mean±std)</td>
<td>95</td>
</tr>
</tbody>
</table>
List of Figures

1.1 Estimated new cancer cases and deaths in males and females in 2017 [1]. 2
1.2 Epithelial structure of the prostate gland. Notice the difference between the basal and luminal cells in location, shape and size of the cells. ................................................................. 3
1.3 Benign and malignant prostate images. Notice the difference in the size of glands, basal cells, and luminal cells. ................................. 4
1.4 A flow diagram showing a CADx system that can provide insightful information for the pathologists in cancer diagnosis. 7

3.1 Medical images are dominated by anisotropic features. .......... 19
3.2 Curve approximation using (a) isotropic objects and (b) anisotropic objects. ................................................................. 20
3.3 Tiling of the frequency plane by (a) short time Fourier transform, (b) wavelet transform, (c) curvelet transform, and (d) shearlet transform. 22
3.4 Frequency support of shearlets for different scaling and shear values. 24
3.5 Shearlet function $\psi_{0,2,-1,0}$ in (a) time domain and (b) frequency domain. 25
3.6 Frequency domain implementation of the discrete shearlet transform [2]. The Laplacian pyramid is applied at each resolution level $j$ to decompose the image $f_{a}^{j-1}$ into a low pass filtered image $f_{a}^{j}$, a quarter of the size of $f_{a}^{j-1}$, and a high pass filtered image $f_{d}^{j}$. Then the directional filtering is applied to extract the shearlet coefficients with different directions. This process will repeat for the rest of the resolution levels. ................................................................. 26
3.7 Edges with different directions can be detected using shearlets. . 27

4.1 Our proposed histogram of shearlet coefficients method. Histograms are extracted from shearlet coefficients at each decomposition level and concatenated to compose the histogram of shearlet coefficients. . 29
4.2 Histograms of shearlet coefficients for (a) correctly classified pair of histological breast images and (b) misclassified pair. ........... 30
4.3 Kernel trick where we map the data to a different space to make it linearly separable. ................................................................. 33
4.4 Our CNN consisting of three layers of convolution and max-pooling followed by two fully connected layers. ......................... 37
4.5 Our deep neural network that employs CNNs to learn features from
the magnitude and phase of shearlet coefficients extracted from five
decomposition levels along with the RGB image. The learned features
go through a fully connected layer and are used for classification of
the medical images.

4.6 A simple two layers neural network.

4.7 Backpropagation in the output layers.

4.8 Deep neural network based on decision fusion from RGB, magnitude,
and phase of the shearlet coefficients.

5.1 Sample breast tissue images of the university of California Santa Bar-
bara bio-segmentation benchmark dataset [3].

5.2 DSC and HSC features extracted from a pair of benign and malignant
images.

5.3 Sample prostate tissue images from our dataset. Notice the changes
in texture from benign to malignant.

5.4 Block diagram of our proposed system for Gleason grading.

5.5 Gleason grading system [4].

5.6 Sample prostate tissue images.

5.7 Sample Gleason grade 2 to 5 images and their corresponding shearlet
coefficients. Notice that the shearlet coefficients can highlight the
structure of the cell and the random scattering of the epithelial cells.

5.8 Green, red, and blue color channel histograms for sample Gleason
grades 2 to 5. Notice the shift in the green color channel histogram
as the grade increases.

5.9 A sample Gleason grade 2 image and the segmented mask.

5.10 Our proposed method for Mp-MRI prostate cancer detection.

5.11 Sample Mp-MRI imaging sequence from our dataset. The red color
indicates the tumor boundaries. Notice the visual differences in the
tumor appearance from one Mp-MRI sequence to another.

5.12 Our proposed method for medical image analysis via shearlet trans-
form and deep learning.

5.13 All 104 augmented images for a sample breast tissue image.

5.14 Sample benign and malignant breast tissue images and their corre-
sponding magnitude and phase of shearlet coefficient.

5.15 Convolutional layer filters from three different layers for RGB and
magnitude of the shearlet coefficients.

5.16 Convolutional layer features from three different layers for RGB and
magnitude of the shearlet coefficients.

5.17 ROC curves for breast cancer diagnosis using feature engineering
(red) and our deep neural network (blue).

5.18 Confusion matrices for Gleason grading.

5.19 Our network’s graph based on weighted decision fusion from RGB,
magnitude, and phase of shearlet coefficients.
5.20 Gleason grade 2-5 samples of our dataset. ........................................ 89
5.21 Training accuracy vs. iterations. Notice how the accuracy increases
as the number of iterations increases. ...................................................... 89
5.22 Training loss vs. iterations. Notice how the network’s loss decreases
as the number of iterations increases. ...................................................... 90
5.23 Decision fusion weights for RGB, Mag, and Phase CNNs. ................. 91
5.24 Convolutional layer filters from three different layers for RGB, mag-
nitude, and phase of shearlets. ................................................................. 92
5.25 Pooling layer features for RGB, magnitude, and phase of the shearlet
coefficients. .............................................................................................. 93
5.26 ResNet architecture. ................................................................. 96
List of Symbols

- $a$: scale
- $s$: shear
- $t$: location
- $\psi$: shearlet function
- $\theta$: rotation
- $\phi$: mapping function
- $x$: input training sample
- $d$: kernel weight
- $C$: SVM hyperparameter
- $\zeta$: slack variable
- $b$: bias
- $y$: label
- $O$: objective function
- $\alpha$: Lagrange multiplier
- $E$: error
- $\eta$: learning rate
- $p$: softmax function
- $L$: loss
- $W$: probability weight vector
Chapter 1

Introduction

Cancer is a group of disease described by the abnormal cell growth that can spread and metastasize into the other organs [1]. It is caused by various factors, e.g. tobacco, diet, genetics, etc. Cancer can be treated using chemotherapy, radiation, surgery, etc. There were about 14.5 million Americans with a history of cancer alive in 2014 [1]. About 1,688,780 new cases of cancer are estimated in 2017. About 600,920 Americans are estimated to die from cancer in 2017. Cancer is the second most frequent cause of death in the US after cardiovascular disease.

Prostate cancer and breast cancer are the most frequently diagnosed cancers in men and women, respectively [1]. They rank second among the other cancers as the cause of death (Figure 1.1). Cancer if remains untreated can transfer into the other tissues and develop new tumors. Therefore, early diagnosis of cancer is vital for the patient’s survival and treatment planning.
Cancer diagnosis and grading generally includes pathologists going through the images of biopsy samples taken from patients, examining the images under a microscope and making decisions relying on their personal knowledge and experience. To better understand the histology of the prostate cancer the epithelial structure of the prostate is shown in Figure 1.2. Prostatic epithelium includes two layers of the cells. The inner cell layer consists of secretory cells and the outer layer is made up of basal cells [5]. The basal and secretary cells have different shapes and sizes. Usually there is no noticeable cytoplasm in the basal cells while the secretory cells have clearly visible cytoplasm.
Figure 1.2: Epithelial structure of the prostate gland. Notice the difference between the basal and luminal cells in location, shape and size of the cells.

To better illustrate the histology of the prostate when cancer happens two sample benign and cancer images are shown in Figure 1.3. In the benign image (Figure 1.3a), two layers of epithelial cells are clearly visible around the lumen inside the large glands. The outer layer basal cells separate the inner layer luminal cells from the stroma. The amount of cytoplasm (pink) is much larger in the luminal cells comparing to the basal cells. However when cancer happens (Figure 1.3b), the glands become much smaller, the basal cells are gone, and the cancerous secretory cells that now have enlarged nuclei invade the stroma.

Pathologists usually look for changes in the cell structure and the distribution of the cells through the tissue. This process can be subjective, time-consuming, and prone to variability due to the qualitative nature of the diagnosis by the pathologist. Therefore, to overcome this problem and enhance the reliability of cancer diagnosis and grading, it is very important to design and develop repetitive and reliable automatic cancer diagnosis techniques that work based on the quantitative measures. These techniques can help the pathologists make faster and more reliable diagnosis and ultimately save the patient’s life.
Figure 1.3: Benign and malignant prostate images. Notice the difference in the arrangement, size and shape of the glands, basal cells, and luminal cells.
Automatic cancer diagnosis techniques rely heavily on feature representation methods. A wide variety of feature representation methods have been developed in the literature such as Haar, HOG, SIFT, SURF, etc. Methods based on signal processing usually apply Fourier transform or wavelet transform on the images and extract features from the coefficients. However, the aforementioned transforms have limitations. Fourier transform does not include the spatial information of the images and wavelet transform does not have directional sensitivity. These motivated researchers to develop a new set of transforms called shearlet transform [6]. Shearlet transform is a sparse directional multiscale representation of signals. The directional sensitivity and the multiscale analysis of signals make the shearlet transform an excellent choice for histological image analysis where the structure and shape of the cells go through changes due to cancer. In this Research, we develop methods that utilize shearlet transform for automatic cancer diagnosis and grading of prostate and breast cancer.

Our main contributions in this dissertation are two-folded.

- We extract features from shearlet coefficients using the histogram representation of the shearlet coefficients and utilize the engineered features for medical image classification. We further improve our method by extracting more statistics from the shearlet coefficients using co-occurrence [7] of the shearlet coefficients. We also extract color and morphological features from the prostate images and combine the aforementioned features using multiple kernel learning (MKL) [8]. MKL eliminates the need for a feature selection technique that otherwise could be cumbersome due to the different nature of the features being combined.

- We develop a different framework by eliminating the need for feature engineering. To this end, we employ deep neural networks [9] equipped with the
phase and magnitude of shearlet coefficients [10] along with the RGB data. This helps us learn the most appropriate abstractions and features instead of using feature engineering and makes our proposed framework more general and less application-dependent. Our deep neural network is a convolutional neural network (CNN) [11] consisting of multiple layers of convolution, max pooling, and fully connected layers. We ultimately design an improved deep neural network method by assigning weights on the probabilities and updating those weights using backpropagation. This helps boost the contribution of each feature set separately, which leads to better classification accuracy.

The immediate objective of this dissertation is to develop and evaluate medical image analysis methods for automatic cancer diagnosis and grading as we aim to bring the image analysis and cancer pathology together. The ultimate objective is to develop a computer-aided diagnosis (CADx) system for the pathologists to enhance their performance in practice. A flow diagram of such CADx system is depicted in Figure 1.4 where the output of our CADx system along with the feedback from the radiologist would help the pathologist make a faster and more reliable diagnosis.

The remainder of this dissertation is organized as follows. Chapter 2 presents a literature review on the breast cancer diagnosis, prostate cancer diagnosis, and Gleason grading in the histological and magnetic resonance images. Chapter 3 reviews the evolution of harmonic analysis methods including Fourier transform, wavelet transform, and shearlet transform and compares them. Chapter 4 presents our proposed frameworks using both feature engineering and feature learning methods for cancer detection. Chapter 5 shows the experimental results of our proposed methods on several databases and compares our results with the state-of-the-art. Finally, Chapter 6 presents our conclusions and achievements and gives advice for future research on this topic.
Figure 1.4: A flow diagram showing a CADx system that can provide insightful information for the pathologists in cancer diagnosis.
Chapter 2

Literature review

Automatic cancer diagnosis includes three steps: preprocessing, feature extraction, and classification [12]. Preprocessing is used to remove the noise and improve the image quality. It might also include cell nuclei segmentation. Feature extraction is performed to extract the morphological, textural, or intensity-based features from the images. It can be performed in the cellular level or tissue level. The features extracted in this step are used in the classification step to diagnose the cancer or grade its severity. There is an extensive amount of research for automatic cancer diagnosis in the literature. There are several overviews and surveys covering this topic as well. In this dissertation, we focus on the breast cancer diagnosis, prostate cancer diagnosis, and Gleason grading tasks and categorize the research work based on the three automatic cancer diagnosis steps explained above. Furthermore, since we had the histopathology images of the breast tissues and MRI and histopathology images of the prostate tissue, we will focus on the research work with similar data.

2.1 Breast cancer diagnosis

There is a tremendous amount of research conducted on automatic breast cancer diagnosis since breast cancer is the most common form of cancer among women
and the laborious cancer diagnosis process can be greatly reduced using automatic image analysis methods. An overview of the methods proposed for breast cancer histopathology image analysis is presented in [13]. In this dissertation, we categorize these methods based on their feature extraction and classification methods since those are the main focus of this research.

Some of the automatic breast cancer diagnosis methods are based on feature extraction using signal processing methods [14–17]. Wan et al. [14] decompose the images into multiscale patches using a dual-tree complex wavelet transform. Then they extract statistical features from wavelet coefficients and use SVM for mitosis detection. Wang et al. [15] apply a top-bottom hat transform followed by wavelet decomposition and multiscale region growing to segment the cell nuclei. Then they extract the shape and textural features and use SVM for classification of breast cancer histopathology images. Mousa et al. [16] extract horizontal, vertical, and diagonal coefficients from wavelet decomposition, compute the energy of coefficients and use them for classification of benign and malignant breast tissue images. Kothari et al. [17] apply Fourier transform on the images and extract shape-based features. Then they use the extracted features for classification of histological renal tumor subtypes.

Other types of features have also been used for automatic breast cancer diagnosis [18–23]. Irshad et al. [18] extract co-occurrence features, run-length features, and scale-invariant feature transform (SIFT) features from the images and use them for mitosis detection. Ojansivu et al. [19] extract local binary pattern (LBP) and local phase quantization (LPQ) features from the images and use SVM for classification of breast cancer morphology. Ko et al. [20] apply color space transform, nuclei segmentation, and watershed processing to extract features from the images and use SVM for classification. Irshad et al. [21] apply Laplacian of Gaussian, thresholding,
and active contour models on the images to segment the nuclei. Then they extract morphological and statistical features from the cells and use decision tree for the classification and mitosis detection. Boucheron et al. [22] use watershed-based and marker-based segmentation of the cell nuclei to extract morphological features and use them for the classification of benign and malignant breast tissue images. Filipczuk et al. [23] present a hybrid k-means based segmentation method that first thresholds the image to highlight the background objects. Then they use k-means clustering to segment the cell nuclei from the background. Finally they extract morphological features from the nuclei and use them for the classification of benign and malignant tissue images.

Classification methods also play an important role in automatic breast cancer diagnosis [24–37]. Methods based on support vector machines have been extensively used for breast cancer diagnosis [24–26]. Some of the classical breast cancer diagnosis methods use fuzzy c-means [27] or a hybrid fuzzy-genetic methodology for classification [28, 29]. Paul et al. [30] use regenerative random forest equipped with the automatic feature selection for mitosis detection in the histopathological Breast Cancer Images. Couture et al. [31] use sparse coding to learn dictionary of features that represent the images. Then they use a logistic regression classifier for tumor histology. Neural networks have also been used as classifier for breast cancer diagnosis [32,33]. Due to the shortcomings of feature extraction methods, recently feature learning has gained a lot of attention. These methods have been able to outperform traditional methods. Cruz-Roa et al. [34] apply convolutional neural networks on the whole slide images to detect invasive ductal carcinoma. Zejmo et al. [35] use GoogLeNet and AlexNet CNNs equipped with SVM to classify the benign and malignant cytological specimens. Ciresan et al. [36] use deep CNNs to detect the mitosis in breast cancer histology images. They won the ICPR 2012 mitosis detec-
tion challenge. Wang et al. [37] combine engineered features–including morphology, texture, and color–with CNN to detect mitosis in breast cancer pathology images.

2.2 Prostate cancer diagnosis

Prostate cancer is the most common cancer among men in the United States. There has been tremendous amount of research conducted on automatic prostate cancer diagnosis using medical image analysis [38–54]. In this dissertation we focus on two imaging modalities that are commonly used: histology and magnetic resonance imaging (MRI). Lopez et al. [38] review the automatic prostate cancer diagnosis in digitized histopathology and Wang et al. [39] review the automatic prostate cancer diagnosis in multiparametric MRI. In the following we review the literature and categorize them based on their image modality and feature extraction and classification methods.

Some of the automatic prostate cancer diagnosis methods use histopathological images [40–43]. Salman et al. [40] propose a supervised classification scheme based on the joint intensity histograms of hematoxylin and eosin to delineate prostate cancer areas in the histological images. They compare their method with the intensity histograms, histogram of oriented gradients, and manual annotation by pathologists. Ali et al. [41] propose a CADx system based on wavelet packets and support vector machines. Gertych et al. [42] use joint histograms of local binary patterns and local variance as texture features and utilize support vector machine with random forest to classify the prostate tissues into stroma, benign, and malignant areas. Tabesh et al. [43] propose a multifeature method for prostate cancer diagnosis and Gleason grading. They aggregate the texture, color, and morphological features and choose the best set of features using sequential forward feature selection technique. Then they use Gaussian, K-nearest, and SVM for classification.
Magnetic resonance imaging (MRI) has also been widely used for prostate cancer detection [44–54]. Cobelli et al. [44] use apparent diffusion coefficient (ADC) values and ADC ratios for prostate cancer detection and grading. Puech et al. [45] present a CAD software called ”ProCAD” that analyses the dynamic contrast enhanced MRI data and diagnoses the cancer using a seeded region growing algorithm. Tamada et al. [46] review the importance of diffusion weighted MRI for prostate cancer diagnosis and grading. Multiparametric magnetic resonance imaging MP-MRI has been widely used for prostate cancer detection [43-49]. Giannini et al. [47] propose an automatic CAD system for prostate cancer detection on MP-MRI. First they create a malignancy probability map from all voxels of the prostate. Then they perform feature selection and classification on ADC and T2 weighted images. Niaf et al. [48] propose a method for prostate cancer diagnosis in MP-MRI based on pattern classification via kernel based learning of qualitative and quantitative labels when there is some uncertainty in the data. They compare their proposed SVM with classical SVM and fuzzy SVM. Cameron et al. [49] extract hybrid textural morphological features from the MP-MRI data and use them to delineate benign and malignant areas. Niaf et al. [50] propose a CAD system for prostate cancer diagnosis. They propose an SVM with feature selection and smoothness terms to delineate the benign and malignant areas. Khalvati et al. [51] propose a set of textural features for automatic prostate cancer detection in MP-MRI. They extract the first and second order statistical features from Gabor and Kirsch filters. Fehr et al. [52] extract the first and second order statistical texture features from the ADC and T2-w MRI and use them for automatic cancer diagnosis and Gleason grading. Vos et al. [53] propose an automatic CAD system for prostate cancer detection using Hessian blob detection and segmentation. Liao et al. [54] propose a deep learning approach based on the stacked independent subspace analysis to learn the most appropriate features from the MR images and perform segmentation.
2.3 Gleason grading

The Gleason grading system [4] is the main method for histological grading of the prostate cancer and helps with the prostate cancer prognosis and therapy. Similar to automatic breast and prostate cancer diagnosis, the Gleason grading literature mainly focus on the feature extraction and classification methods. Therefore, we categorize the literature based on their feature extraction and classification methods. Furthermore, the histological images are mainly used for automatic Gleason grading and we will only focus on these images.

The majority of the Gleason grading methods focus on the feature extraction and feature representation methods [55–73]. These methods use texture features [55–63], morphological features [64–71], or a combination of texture and morphological features [72,73]. Wavelet transform has been widely used for automatic Gleason grading [55–58]. Jafari et al. [55,56] propose a method based on the energy and entropy of multiwavelet transform coefficients. They use a k-NN classifier as their classification method. Lopez et al. [57] propose a combination of wavelet and fractal features and use SVM for classification. Almuntashri et al. [58] combine the features extracted from Haar wavelet transform and fractal analysis and use SVM for classification. Huang et al. [59] propose two feature extraction algorithms based on fractal analysis and perform classification using Bayesian, k-NN, and SVM classifiers. Khurd et al. [60] propose a method using the random forests to combine the filter responses into textons and use SVM for classification. Wang et al. [61] use the bag of words features extracted from the histological images along with SVM for automatic Gleason grading. Harder et al. [62] extract co-occurrence matrix features from the histological images of the prostate and use them for Gleason grading. Lin et al. [63] extract higher order statistical moments from the curvelet coefficients and use SVM for classification. Ali et al. [64] propose an active contour model combining
the shape priors with the boundary and regions based energy terms for cell nuclei segmentation. Then they use SVM for classification. Sparks et al. [65] use manifold learning to extract low dimensional manifold representations from high dimensional features and use SVM for classification. Naik et al. [66] propose a method for automatic segmentation of nuclei and Gleason grading using manifold learning applied on the morphological features. Khurd et al. [67] first compute networks from cells extracted from images and extract statistical features from network location. Then they use SVM for classification. Niazi et al. [68] extract luminal and architectural features from the images and use linear subspace for classification. Naik et al. [69] use Bayesian classifier along with level-set algorithm and template matching for nuclei segmentation and Gleason grading. Nguyen et al. [70] use graph cut for nuclei segmentation and spatial arrangement of the nuclei for Gleason grading. Ren et al. [71] use structure features and Delaunay triangulation to segment the individual cell nuclei and perform Gleason grading. Tabesh et al. [43, 72] extract multiple features including color channel histograms, fractal features, wavelet features, and morphological features and use Gaussian classifiers with a greedy search feature selection method for Gleason grading. Lopez et al. [73] extract multiple features including complex wavelet features, quaternion color ratios, and local binary pattern features and use SVM for classification.

Other automatic Gleason grading methods focus on the classification algorithms [74–77]. Farjam et al. [74] segment the glandular regions using texture features and K-means clustering and use the extracted features in a tree-structured algorithm for classification. Greenblatt et al. [75] extract quaternion wavelet transform and local binary patterns features from the histological images and propose a quaternion neural network along with SVM for classification. Kallen et al. [76] use a pretrained neural networks called ”OverFeat” that was trained on photographic images for
automatic Gleason grading. Jacobs et al. [77] use max-margin conditional random fields to segment the histological images into regions of benign, Gleason grade 3, and grade 4.

As mentioned in this chapter, wavelet transform has been widely used as the texture features for automatic cancer diagnosis and grading. However, wavelet lacks directional sensitivity and use isotropic scaling which makes it not suitable for detecting edges. On the other hand, shearlet transform [6] has inherent directional sensitivity and use parabolic scaling which makes it suitable for detecting curvilinear features which are the most dominant features in medical images. This motivated us to utilize shearlet transform for automatic cancer diagnosis and grading [78–82]. We extract histogram of shearlet coefficients from the histological images and use SVM for classification of benign and malignant breast tissue images [78]. We also use the histogram of shearlet coefficients for prostate cancer detection and Gleason grading in the histological images [79] and magnetic resonance images [80]. We propose a multi-feature automatic Gleason grading method based on shearlet transform and multiple kernel learning (MKL) [81]. We extract texture features from the histological images using co-occurrence of shearlet coefficients. Then we combine shearlet features with the color and morphological features utilizing MKL and perform classification. In a recent study [82], we explore the possibility of feature learning for microscopic medical image classification using deep neural networks. To this end, we apply shearlet transform on the histological images of breast and prostate tissue and extract the magnitude and phase of shearlet coefficients. Then we combine them with the RGB image data and use them as input to our convolutional neural network (CNN). Our CNN consists of multiple layers of convolution followed by max-pooling and fully connected layers. Our experiments show that our proposed method outperforms the traditional feature engineering methods.
Chapter 3

Sparse directional
representation systems

Shearlet is a framework for efficient representation of multidimensional data [83]. Most of its success comes from the fact that despite its predecessors (e.g. Fourier transform, wavelet transform, etc.), shearlet can detect edges and anisotropic features which dominate the multidimensional realm. In this chapter, first we briefly go over harmonic analysis and classical signal processing methods including Fourier and wavelet transforms. Then we present shearlet transform in details and compare its properties with the other representation systems.

3.1 Applied harmonic analysis

In music, the harmonics of a note with frequency $f$ are the integer multiples of that frequency, i.e. $2f$, $3f$, and so on. In mathematics, applied harmonic analysis focuses on the efficient representation and analysis of periodically recurrent data. It includes breaking the mathematical data into the sum of simpler components (analysis) and reconstructing the data from the expansion coefficients (synthesis).
One of the pioneer methods of harmonic analysis is the Fourier transform that was developed in the late 18th century. Fourier transform is one of the most powerful techniques for signal representation that resolves general functions into the sum (integral) of simple functions with some special properties [84]. The continuous Fourier transform of an integrable function \( f \) is defined as follows

\[
\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i x \xi} \, dx \quad (3.1.1)
\]

and the inverse Fourier transform is defined as follows

\[
f(x) = \int_{-\infty}^{\infty} \hat{f}(\xi) e^{2\pi i x \xi} \, d\xi \quad (3.1.2)
\]

Although Fourier transform is a powerful tool for representation of signals, it has a crucial disadvantage in finding the location of the peaks of the signal. Since local perturbation of a signal results in a change in all Fourier coefficients simultaneously, Fourier transform can only analyze the global structure of a signal [85]. Also the original Fourier transform only contains information of the signal in the frequency domain.

The shortcomings of Fourier transform lead to the birth of time-frequency methods, specifically wavelet transform in the early 1980s. Wavelet transform is a system of time-scale elements that can provide local information of the signal in different resolutions. A wavelet is a function \( \psi \in L^2(\mathbb{R}) \) that satisfies \( \int_{\mathbb{R}} | \hat{\psi}(\xi) |^2 / |\xi| \, d\xi = 1 \). The continuous wavelet system is defined as follows

\[
a^{-1/2} \psi(a^{-1}(x - t)) : a > 0, t \in \mathbb{R} \quad (3.1.3)
\]
Then the continuous wavelet transform of a function $f$ at location $t$ and scale $a$ is defined as follows

$$W\psi f(a,t) = a^{-1/2} \int_{\mathbb{R}} f(x)\psi^{*}(a^{-1}(x-t))dx, \quad a > 0, t \in \mathbb{R} \quad (3.1.4)$$

where $^*$ represents the operation of complex conjugate.

From the above formulas it is obvious that the wavelet transform can provide the location of the peaks of the signal in different scales. The discrete wavelet transform is achieved by sampling the continuous wavelet transform parameters $a$ and $t$ as $2^{-j}$ and $2^{-j}k$ respectively where $j, k \in \mathbb{Z}$. The discrete wavelet system is defined as follows

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^{j}x - k) : j, k \in \mathbb{Z} \quad (3.1.5)$$

Wavelet transform has many excellent properties. It is localized in both time (spatial) and frequency domain while the Fourier transform is localized in frequency domain only. Wavelet has better resolution in both time and frequency domain. Wavelet has less computational complexity due to the logarithmic division of frequency plane in contrast to the equally spaced frequency divisions in fast Fourier transform (FFT) [83]. However, there are some disadvantages with wavelet transform as well. The traditional wavelet transform—which is defined using the isotropic dilations— is unable to detect the geometry of set of singularities of multivariate functions and lacks directional sensitivity. This is while the natural and medical images are dominated by anisotropic features and structures. Figure 3.1 shows a sample breast tissue image and the anisotropic features in it. Neurophysiology also supports this claim, since it is customarily accepted that neurons are highly directional and react strongly to curvelike structures [86].

Another important property of wavelet transform is that wavelet provides sparse representation for isotropic features in the signal. The degree of sparsity is usually
measured as the decay rate of the best n-term approximation which is also called the asymptotic approximation rate [83]. The sparse approximation of a function is defined as follows [6]:

**Theorem:** Let $f$ be a function with continuous first and second derivatives $C^2$ away from piecewise $C^2$ curves, and let $f^T_N$ be the approximation to $f$ using the $N$ largest coefficients of the transform $T$. Then the $N$-term approximation error is defined by

$$\|f - f^T_N\|_2^2$$

(3.1.6)

where $\|\cdot\|_2$ is the $l - 2$ norm. The $N$-term approximation error for Fourier, wavelet, and optimal transform are as follows

- Fourier approximation error: $\|f - f^F_N\|_2^2 \leq C.N^{-1/2}, N \to \infty$
- Wavelet approximation error: $\|f - f^W_N\|_2^2 \leq C.N^{-1}, N \to \infty$
- Optimal approximation error: $\|f - f^O_N\|_2^2 \leq C.N^{-2}, N \to \infty$

Figure 3.1: Medical images are dominated by anisotropic features.
We can observe that the wavelet transform is sparser than the Fourier transform which means the wavelet coefficients sequence has very few non-zero entries. This has multiple benefits since we can detect important features by thresholding the wavelet coefficients with the largest absolute values which can lead to high compression rates for wavelets [87]. This is crucial for tasks such as feature extraction and classification which lead to the success of wavelets in such areas.

However, we can also observe that wavelet cannot reach the optimal approximation error. This is due to the fact that wavelets are defined using isotropic scaling which makes them inherently isotropic objects and therefore not optimal for approximating anisotropic objects which are the dominant features in natural images [86]. To overcome this shortcoming of traditional wavelets, several set of transforms were suggested including the directional wavelets [88], the complex wavelets [89], the ridgelets [90], the contourlets [91], the curvelets [92], and the shearlets [2, 6, 93]. These transforms are much better suited to detect anisotropic features since they are based on anisotropic objects and align with the curvilinear structure much better than wavelets as shown in Figure 3.2.
The first sparse directional representation system that achieved the optimal approximation error and was able to detect anisotropic features in the images was the curvelet transform [92]. The curvelets are based on three parameters: scale, translation, and direction. The curvelets enforce anisotropy by parabolic scaling i.e. \( \text{length}^2 \approx \text{width} \). For every scale \( j \geq 0 \) the parabolic scaling matrix is defined as \( A_j := \begin{pmatrix} 2^j & 0 \\ 0 & 2^{j/2} \end{pmatrix} \). This makes them suitable for approximating curvilinear features in the images. It was proved in [94] that analyzing elements with elongated and orientable supports are needed to achieve optimally sparse approximation of piecewise smooth multivariate functions. This idea was the main reason for constructing curvelets and shearlets.

A major drawback of the curvelets is the way the orientation is defined. The curvelets use the rotation matrix defined as \( R_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \) for \( j \in \mathbb{Z} \) and \( \theta \in \mathbb{T} \). Therefore, curvelets are based on rotation and a rotation destroys the discrete lattice structure [85]. Hence, the discrete curvelet transform cannot be directly implemented. This motivated the researchers to create shearlet systems which define the direction using directional matrix which makes the discrete implementation of shearlets feasible.

### 3.2 Shearlet transform

The shearlet transform was first introduced by Guo et al. [95] in 2006. Shearlet is the first directional sparse representation system that has all the excellent properties of curvelet, but is also more adapted to the digital realm due to using shear matrices for directions. This makes the digital implementation of shearlets to be consistent with the continuous domain. An important property of representation systems is their tiling of frequency plane which directly controls their ability to pro-
Figure 3.3: Tiling of the frequency plane by (a) short time Fourier transform, (b) wavelet transform, (c) curvelet transform, and (d) shearlet transform.

Figure 3.3 shows the tiling of the frequency plane by short time Fourier transform, wavelet transform, curvelet transform, and shearlet transform. We can observe that the shearlets and curvelets have elongated and orientable elements which is due to their use of parabolic scaling and directional sensitivity and in turn lead to optimal approximation error when detecting anisotropic features [94].

Shearlet uses the parabolic scaling matrix $A_a = \begin{pmatrix} a & 0 \\ 0 & \sqrt{a} \end{pmatrix}$ and shear matrix $S_s = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}$ where $a \geq 0$ and $s \in \mathbb{R}$, to generate elements at different scales and directions. A shearlet function $\psi \in L^2(\mathbb{R}^2)$ needs to satisfy the admissibility condition $\int_{\mathbb{R}^2} |\hat{\psi}(\xi_1, \xi_2)|^2 d\xi_1 d\xi_2 = 1$ for the shearlet transform to be invertible [96]. Then the continuous shearlet system generated by $\psi$ is defined as

$$a^{-3/4} \psi(A_a^{-1}S_s^{-1}(., - t)) \quad : \quad a > 0, s \in \mathbb{R}, t \in \mathbb{R}^2 \quad (3.2.1)$$

The continuous shearlet transform of some $f(x) \in L^2(\mathbb{R}^2)$ is defined as

$$S\!H_\psi f(a, s, t) = a^{-3/4} \int_{\mathbb{R}^2} f(x) \psi^*(A_a^{-1}S_s^{-1}(x - t)) dx, \quad : \quad a > 0, s \in \mathbb{R}, t \in \mathbb{R}^2 \quad (3.2.2)$$
The parameters involved in constructing shearlets provide necessary tools for sparse directional representation of data. If we let the scale parameter $a$ converge to zero, it produces functions that look like needle due to parabolic scaling which helps with approximating anisotropic features. Despite the rotation angle that was used by curvelets, the shear parameter $s$ helps detect directions by slope. This leads to nice digital domain implementations for shearlets. Finally, the location parameter $t$ helps find the location of the singularities accurately.

An important concept that helps us understand the tiling of frequency plane in representations systems is the concept of support. The support of a real valued function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is defined as $\text{supp } f = \{ x \in \mathbb{R}^n : f(x) \neq 0 \}$. Let the shearlet function $\psi \in L^2(\mathbb{R}^2)$ be defined as a tensor product by $\hat{\psi}(\xi) = \hat{\psi}_1(\xi_1) \hat{\psi}_2(\xi_2)$ where $\psi_1$ is a continuous wavelet such that $\hat{\psi}_1 \in C^\infty(\mathbb{R})$ and $\text{supp } \hat{\psi}_1 \subseteq [-2, -1/2] \cup [1/2, 2]$ and $\psi_2$ is a bump function such that $\hat{\psi}_2 \in C^\infty(\mathbb{R})$ and $\text{supp } \hat{\psi}_2 \subseteq [-1, 1]$. Then the support of each shearlet function $\hat{\psi}_{a,s,t}$ can be determined as $\text{supp } \hat{\psi}_{a,s,t} \subseteq \{ (\xi_1, \xi_2) : \xi_1 \in [-2/a, -1/2a] \cup [1/2a, 2/a], |\xi_2/\xi_1 - s| \leq \sqrt{a} \}$. Therefore, as shown in Figure 3.4, each continuous shearlet has a frequency support on a pair of trapezoids oriented along a line with slope $s$.

Next, we investigate the directionality of the shearlets. It was proven in [97] that the directions of the singularities in a function $f$ can be detected using the decay rate of the corresponding shearlet transform. The shearlet coefficients decay faster than any other polynomial as the scale parameter $a \rightarrow 0$ except when the translation parameter $t$ is located on the singularity and the shear parameter $s$ points to the direction perpendicular to the singularity. To better understand this suppose a distribution $f$ has a singularity in the $y$ direction located at the origin $(0, 0)$. We can observe that $f = \delta(x_2 - px_1), p \in \mathbb{R} \setminus \{0\}$. Then we can find the Fourier transform as $\hat{f} = \delta(\xi_2 + \frac{1}{p}\xi_1)$. We can observe that the line $\xi_2 = -\frac{1}{p}\xi_1$ is
perpendicular to the line $x_2 = px_1$ along which $f$ is placed. We can conclude that there is maximal overlap between the analyzing shearlet elements and $\hat{f}$ if $\hat{\psi}(a, s, t)$ is stretched along the line $\xi_2 = -\frac{1}{p}\xi_1$. Therefore, the shearlet coefficients with the correct $t$ and the shear parameter perpendicular to the direction of the singularity give strong responses to the singularity. This phenomena is shown in Figure 3.5 where $p = 1$. Figure 3.5b shows the shearlet function $\psi_{0,2,-1,0}$ in frequency domain. It is clear that this function is stretched in the $x$ direction. Figure 3.5a shows the shearlet function in time domain. We can observe that this function is stretched in $y$ direction which is perpendicular to the $x$ direction. We can have the same discussion for the translation parameter $t$ and prove that continuous shearlet transform can detect the location of the singularities.

The discrete shearlet transform can be obtained by sampling the continuous shearlet transform $SH_{\psi} f(a, s, t)$ on discrete set of parameters. Let’s define $M_{as} = S_{s}A_{a} = \begin{pmatrix} 1 & \sqrt{as} \\ 0 & \sqrt{a} \end{pmatrix}$. We will discretize this matrix as $M_{jl} = B^lA^j$ by choosing $a = 2^{-j}$ and $s = -l$ with $j, l \in \mathbb{Z}$ where $B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $A = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}$. Also let’s replace the
Figure 3.5: Shearlet function $\psi_{0.2,-1,0}$ in (a) time domain and (b) frequency domain.

translation parameter $t$ by a point in the discrete lattice $\mathbb{Z}^2$. Then we can obtain the discrete shearlet system as

$$\psi_{j,l,k} = 2^{3j/2} \psi(B^l A^j x - k) \quad (3.2.3)$$

where $\hat{\psi}(\xi) = \hat{\psi}(\xi_1, \xi_2) = \hat{\psi}_1(\xi_1) \hat{\psi}_2(\frac{\xi_2}{\xi_1})$ for $j, l \in \mathbb{Z}, k \in \mathbb{Z}^2$. This makes sure that the discrete shearlet transform has the compact support on a pair of trapezoids similar to the continuous shearlet transform [6].

There are a few numerical implementations of the discrete shearlet transform. In this dissertation we will focus on the method proposed by Easley et al. [2]. They propose a frequency domain implementation and a spatial domain implementation. Their Fourier-based implementation is based on the Laplacian pyramid filter and directional filtering using pseudo polar discrete Fourier transform. Figure 3.6 shows this procedure. Their spatial-domain-based implementation is based on the inverse Fourier transform of the band-limited window functions implemented in the frequency domain. We refer our reader to [2] for more information on this.

A more recent implementation of the discrete shearlet transform was proposed by Hauser et al. [10]. They use fast Fourier transform for a frequency domain implementation of the shearlets. Their method is called fast finite shearlet transform.
Figure 3.6: Frequency domain implementation of the discrete shearlet transform [2]. The Laplacian pyramid is applied at each resolution level $j$ to decompose the image $f_{a}^{j-1}$ into a low pass filtered image $f_{a}^{j}$, a quarter of the size of $f_{a}^{j-1}$, and a high pass filtered image $f_{d}^{j}$. Then the directional filtering is applied to extract the shearlet coefficients with different directions. This process will repeat for the rest of the resolution levels.

(FFST). Figure 3.7 shows the shearlet coefficients extracted from a sample H&E breast tissue image using FFST for different directions. We use their implementation to extract the magnitude and phase from complex shearlet coefficient.

To summarize this chapter, here are some of the main highlights of the shearlet transform:

- Shearlets are well localized since they are compactly supported in frequency domain and have fast decay rate in spatial domain.

- Shearlets use parabolic scaling that makes them suitable for curve approximation and anisotropic feature detection.

- Shearlets have high directional sensitivity due to the design of shearing parameters.
• Shearlets are well localized in spatial domain due to the design of translation parameter.

• Shearlets provide optimally sparse representations.

• Shearlets have fast numeric implementations.

• Shearlets provide a unified treatment of the continuum and digital domains.

Figure 3.7: Edges with different directions can be detected using shearlets.
Chapter 4

Proposed methods

In the previous chapter we showed that the shearlet transform is suitable for detecting edges and anisotropic features in images due to its directional sensitivity and multi-scale representation framework. In this chapter, we present our proposed methods to utilize the shearlet transform for computer aided cancer diagnosis and grading. We start with the feature engineering of the shearlet coefficients and advance to the feature learning methods.

4.1 Feature engineering

4.1.1 Histogram of Shearlet Coefficients (HSC)

The purpose of the feature extraction methods is to find the most discriminative features in the images which then will be used for the classification. To this end, we will start with extracting statistics (histogram) from the shearlet coefficients. A simple, yet effective method for representing the data is to use the histogram method. A histogram shows the frequency distribution of the data. As we explained in the previous chapter, the shearlet transform can find the location and direction of the singularities in the images. Therefore, the histogram of magnitude of shearlet
coefficients can be an efficient representative of the edges and other anisotropic features in the images [98]. Also since the shearlet transform produces a lot of coefficients, using the discrete shearlet coefficients directly can be cumbersome and cause system memory problems. Furthermore, shearlets are sparse and most of the coefficients have values close to zero. Therefore, using the histograms would be more efficient. To this end, we propose the histogram of shearlet coefficients (HSC) method. We apply the discrete shearlet transform with certain number of scales and direction on the image. Then at each decomposition (scale) level, we find the histogram of shearlet coefficients with a fixed number of bins. Finally, we concatenate these histograms from different decomposition levels into a single histogram similar to [98] and use that to represent the image. Figure 4.1 shows our proposed HSC method.
4.1.2 Co-occurrence of Shearlet Coefficients (CSC)

The histogram method is a simple and effective method to represent the images. However, there is a major drawback with the histogram method. The histogram does not include any spatial information on the local structure in the images. This is due to the fact that when calculating the histograms, we only count the frequency of occurrence of values falling into each bin. Hence, we do not take into account the spatial relationship of the pixels in the images. To further illustrate the shortcomings of the histogram method, a pair of histograms is shown in Figure 4.2. Figure 4.2a shows the histogram of shearlet coefficients for a correctly classified pair of benign and malignant breast tissue images. Figure 4.2b shows the histograms for a failed case. In this case, the peaks and the shape of the histograms are very similar which indicates we need a more sophisticated method to represent shearlet features. This was a motivation to extract statistical texture features from the shearlets in this section and also perform feature learning later.

![Histograms of shearlet coefficients for (a) correctly classified pair of histological breast images and (b) misclassified pair.](image)

Figure 4.2: Histograms of shearlet coefficients for (a) correctly classified pair of histological breast images and (b) misclassified pair.

Moreover, texture analysis methods characterize the regions in the images based on their texture contents. They quantify texture qualities (e.g. rough, bumpy,
smooth, or silky) as a function of the spatial arrangements of the image pixels intensities. This is of special interest to us since in medical image analysis the texture of the image plays an important role in feature representation. To this end, we propose extracting the co-occurrence matrix [2] features from the discrete shearlet coefficients. A co-occurrence matrix of an image shows the distribution of co-occurring two pixels at an offset. The offset is a vector that represents the position difference between the pixels in a certain direction.

To find the co-occurrence of shearlet coefficients, first we apply the discrete shearlet transform on the image and extract the coefficients. Then we follow the process described in [7] to find the co-occurrence matrix. Finally, we extract 20 statistics from the co-occurrence matrix. These statistics include correlation, autocorrelation, energy, cluster prominence, entropy, contrast, maximum probability, cluster shade, sum of squared variance, dissimilarity, homogeneity, sum of average, sum of entropy, difference of entropy, sum of variance, difference of variance, inverse difference, information measure of correlation, and inverse difference momentum.

4.1.3 Feature fusion using multiple kernel learning algorithm

In chapter 2 we mentioned that feature extraction is an important step in automatic cancer diagnosis. There are various types of features that can be extracted from the images including texture, morphological features, intensity-based features, etc. Sometimes it is necessary to extract more than one type of feature from the image to have a successful classification. A good example of such applications is the automatic Gleason grading which relies on the color, morphological, and texture features and will be discussed in the next chapter. In this case, we need to combine these various features. One method is to use feature selection techniques to choose the most discriminative features by minimizing the classification error. These meth-
ods are often sensitive to the optimization parameters and need extensive fine-tuning of the parameters and sometimes don’t coverage at all. Also, depending on the optimization method, they might get stuck in a local minimum and never reach the global minimum. Therefore, the classification results would not be optimal. Also from the classification point of view, it is important to find the best hyperparameters and kernels for the input features. This can be challenging in the case of SVM where there are multiple choices for the parameters and kernels. Specially since these features might be coming from different sources, they will need different set of parameters and kernels [99]. To overcome these issues, we propose utilizing multiple kernel learning (MKL) algorithm for fusing different types of features. In MKL, a linear combination of some fixed kernels is used to create a kernel model. Then it will automatically find the best set of hyperparameters and kernels from the poll of possible choices provided. This will also eliminate the need of feature selection techniques. Here we utilize a state-of-the-art multiple kernel learning algorithm called SimpleMKL [8].

To better understand the MKL algorithm, we start with the kernel tricks in SVM which is the most common base learner for MKL. SVM is a linear discriminative classifier for binary classification problems. Given input data with labels, SVM finds the linear discriminant hyperplane to categorize the data. However, if the data is not linearly separable, SVM will not be able to find a hyperplane to separate the categories in the data. It is possible to transform the data into a higher dimensional feature space where it becomes linearly separable. This mapping function is called a kernel. Depending on the number of dimensions of the original feature space, we usually need a high dimensional kernel. This would need a lot of computational power. Luckily, when solving the SVM we only need to calculate the inner product of the transformed points. This is where the kernel trick is useful. The kernel trick
Figure 4.3: Kernel trick where we map the data to a different space to make it linearly separable.

is used to define the similarity measure in the transformed space in terms of the original space without calculating the transformation function. Therefore, we can choose a kernel that transforms the data into a high dimensional space where the data is linearly separable and yet it is simple to calculate the similarity measure by finding the inner product of the transformed points. The kernel trick process is illustrated in Figure 4.3.

Some of the most common kernels used in the literature are linear kernel $K_{\text{Lin}}(x_i,x_j) = x_i.x_j$, Gaussian kernel $K_{\text{Gaus}}(x_i,x_j) = e^{-\|x_i-x_j\|^2/2S^2}$, $S \in \mathbb{R}^{++}$, and polynomial kernel $K_{\text{Poly}}(x_i,x_j) = (x_i.x_j + 1)^q$, $q \in \mathbb{N}$. These kernels are each suitable for a specific type of data. Therefore, choosing the best type of kernel and its parameters that is suitable for our data is very important for training. The MKL approach combines multiple kernels and parameters. This is of specific importance for us since medical images often highlight various changes in the tissue and need different feature representation method. Each of these features might need different kernels to adequately represent them. Therefore, combining the kernels is an excellent approach to combine these information coming from different sources [100].

There are various multiple kernel learning methods in the literature. In this dissertation we utilize the SimpleMKL method developed by Rakotomamonjy et al. [8]. They define the kernel as a linear combination of multiple kernels and solve the
SVM optimization problem using gradient descent as follows

$$\begin{align*}
\text{minimize} & \quad J(d) = \frac{1}{2} \sum_{m=1}^{p} \frac{1}{d_m} \|w_m\|^2 + C \sum_{i=1}^{N} \zeta_i \\
\text{with respect to} & \quad w_m \in \mathbb{R}^{S_m}, \zeta \in \mathbb{R}^N, b \in \mathbb{R}, d \in \mathbb{R}^p
\end{align*}$$

subject to
\begin{align*}
y_i \left( \sum_{m=1}^{p} W_m^T \phi_m(x_i^m) + b \right) & \geq 1 - \zeta_i \quad \forall i, \sum_{m=1}^{p} d_m = 1
\end{align*}

where $x_i$ is the input training sample, $W$ is the weight vector, $d_m$ is the kernel weight, $C$ is a trade-off parameter between the model simplicity and classification error, $\zeta$ is the vector of slack variables which measure the degree of misclassification of the data, $b$ is the bias term of the separating hyperplane, $\phi$ is the mapping function, and $y_i$ is the label of input training sample.

Rakotomamonjy et al. [8] solve the above optimization problem by first solving the canonical SVM optimization problem with given $d$, and then using a reduced gradient descent algorithm to update $d$ based on the following gradient:

$$\frac{\partial J(d)}{\partial d_m} = -\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j K_m(x_i, x_j) \quad \forall m$$

To further extend the above binary classification method to multiclass classification, they define a combined objective function as $O(d) = \sum_{p \in P} O_p(d)$ where $P$ is the set of all pairs of binary classifiers and $O_p(d)$ is the binary SVM objective value associated with the pair $p$. Then the gradient can be calculated as:

$$\frac{\partial O(d)}{\partial d_m} = -\frac{1}{2} \sum_{p \in P} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_{i,p} \alpha_{j,p} y_i y_j K_m(x_i, x_j) \quad \forall m$$

where $\alpha_{i,p}$ is the Lagrange multiplier of the $i-th$ sample in the $p-th$ class. We will use this one-against-all approach for our automatic Gleason grading experiments.
4.2 Feature learning

Engineered features have some shortcomings. They usually need preprocessing and the classification results depend on the accuracy of preprocessing. Furthermore, some of the feature engineering methods have inherent limitations. A good example is the wavelet transform which does not have directional sensitivity. Moreover, feature engineering methods are application-specific and cannot be generalized to different applications. On the other hand, feature learning techniques do not need any preprocessing and can be transferred to different applications since they are data-driven [34]. This motivated us to further improve our proposed method by employing deep neural networks equipped with the magnitude and phase of shearlet coefficients along with the histological images for medical image classification.

4.2.1 Medical image classification via deep neural networks and shearlet transform

We design a deep neural network that utilizes the phase and the magnitude of shearlet coefficients along with the RGB data as the inputs. We specifically extract the phase of shearlet coefficients since most of the information of the signal is contained in the phase [101] and also the features extracted from the phase are invariant to the image contrast and noise [102]. However, since the phase of a signal is non-trivial it is difficult to design phase features. This was another motivation to learn the features instead of feature engineering. We use the magnitude of shearlet coefficients since as we previously showed, the magnitude is a direct representative of the edges in the signal and the higher the magnitude, the higher the chance of an edge existing in that location [93]. The magnitude and phase of shearlets are accompanied with the imagery data (RGB) since we are dealing with the histological H&E images and the color information is very important for the correct diagnosis. Finally, we deploy deep neural networks as an evolution step to extract the most
discriminative abstracts directly from the aforementioned features and employ them for the automatic cancer diagnosis and grading. In the following first we present the feature extraction from the shearlet coefficients and then our deep neural network.

For a shearlet coefficient extracted at scale $a$, shear $s$, and location $t$ denoted by $c(a, s, t) = u + iv$, we find the magnitude and phase as follows

$$Mag(a, s, t) = |c(a, s, t)| = \sqrt{u^2 + v^2}$$

$$Phase(a, s, t) = \angle c(a, s, t) = \tan^{-1} \frac{v}{u}$$

(4.2.1)

After extracting the magnitude and phase of shearlets at each decomposition level, we combine them with the RGB images and feed them to our neural network.

Traditional machine learning techniques (e.g. SVM, KNN, etc.) are not fully capable of analyzing natural images in their raw form since they need appropriate feature representation techniques to extract the most discriminative features from the images for a successful classification [103]. On the other hand, deep learning techniques which consist of multiple layers of processing units are capable of learning the most appropriate features from the raw images directly [2]. Deep learning is a branch of machine learning that extracts high level abstractions from the data similar to how our brain works [103]. Deep neural networks are neural networks with many layers between the input and the output. These methods usually outperform the classical machine learning methods that use the feature engineering concept. Convolutional neural networks (CNN) are biologically-inspired feed-forward neural networks that include successive layers of convolution and pooling followed by a fully connected layer [11]. The architecture of our proposed CNN is illustrated in Figure 4.4. The building blocks of our CNN are described as follows:

- Convolutional layer: Applies a 2-D convolution on the input using Gaussian
Figure 4.4: Our CNN consisting of three layers of convolution and max-pooling followed by two fully connected layers.

filters and sends the output through a non-linear activation function such as rectified linear unit (ReLU).

- Max-pooling layer: Combines similar features together by finding the maximum of local patches and therefore acts as a feature dimension reduction layer.

- Fully connected layer: Acts as an inner product operation and connects all the neurons from the previous layer to all the neurons in the next layer.

Our proposed deep neural network is illustrated in Figure 4.5. The inputs to our neural networks are magnitude and phase of shearlet coefficients extracted from each decomposition level, along with their associated RGB image. Then we send them through separate CNNs individually. The purpose of separating the features extracted from different decomposition levels of shearlets is to let the CNN learn features extracted from different scales and directions. Then the output of CNNs are combined together using a fully connected layer which in turn sends the learned features to the classification layer which is a softmax classifier.
Figure 4.5: Our deep neural network that employs CNNs to learn features from the magnitude and phase of shearlet coefficients extracted from five decomposition levels along with the RGB image. The learned features go through a fully connected layer and are used for classification of the medical images.
4.2.2 A weighted decision fusion framework for automatic Gleason grading using shearlet transform and deep learning

We previously showed in Figure 3.7 that the sum of all shearlet coefficients extracted from an image is a good approximation of the image. Different shearlet decomposition levels correspond to different scales of the features in the image. Lower decomposition levels correspond to the coarser features while higher decomposition levels represent the finer details in the image. We also showed in Figure 1.2 that the color information of the tissue is important for a correct diagnosis. That was the motivation behind including the RGB images in our neural network in the previous subsection. In our neural network depicted in Figure 4.5 we find the classification accuracy based on the concatenation of the outputs of the different CNNs. However, it would be beneficial if we investigate the impact of each feature set (RGB, magnitude, and phase of shearlets) separately since they represent different type of features, e.g. RGB features represent the color features while the shearlet coefficients represent the texture of the tissue. Therefore, in this subsection we propose to assign weights on each feature set decision and learn those weights through backpropagation. Then we calculate the classification accuracy using a weighted sum of the decisions from the CNNs. This helps us understand the contribution of each feature set separately and adjust the network accordingly. Following is the detailed formulation of this approach.

Backpropagation is a method used for training the neural networks [104]. It optimizes the weights by minimizing the classification error and includes two phases: forward pass and backward pass. To better understand how the learning process works let’s start with a simple two layers neural network as shown in Figure 4.6.
During a forward pass, we calculate the network prediction given the inputs, weights, and biases. To this end, first we find the total net input to each hidden layer neuron and pass it through an activation function e.g. logistic function to find the output as follows:

$$net_{h1} = w_1 \times i_1 + w_2 \times i_2 + b_1 \times 1 \tag{4.2.2}$$

$$out_{h1} = \frac{1}{1 + e^{-net_{h1}}} \tag{4.2.3}$$

and we repeat the same process for $h_2$. We also need to find the net input and output of neurons in the output layer of our neural network as follows:

$$net_{o1} = w_5 \times out_{h1} + w_6 \times out_{h2} + b_2 \times 1 \tag{4.2.4}$$

$$out_{o1} = \frac{1}{1 + e^{-net_{o1}}} \tag{4.2.5}$$
and we repeat the same process for \( o_2 \). Finally we can find the total error using sum of the squared error as follows:

\[
E_{\text{total}} = \sum \frac{1}{2} (\text{target} - \text{output})^2
\]  

(4.2.6)

We find this error for \( o_1 \) and \( o_2 \) and the total error of the output is calculated as the sum of these errors:

\[
E_{\text{total}} = E_{o_1} + E_{o_2}
\]  

(4.2.7)

During a backward pass, we calculate the partial derivative of the total error with respect to each weight. This is then used to update the weights to produce output values that are closer to the target values by minimizing the error. To calculate the derivative of the total error with respect to the output layers, we use the chain rule as follows:

\[
\frac{\partial E_{\text{total}}}{\partial w_5} = \frac{\partial E_{\text{total}}}{\partial o_1} \times \frac{\partial o_1}{\partial \text{net}_{o_1}} \times \frac{\partial \text{net}_{o_1}}{\partial w_5}
\]  

(4.2.8)

This process is depicted in Figure 4.7. The first term in the right hand side of the equation 4.2.8 can be calculated as follows:

\[
E_{\text{total}} = \frac{1}{2} (\text{target}_{o_1} - \text{out}_{o_1})^2 + \frac{1}{2} (\text{target}_{o_2} - \text{out}_{o_2})^2
\]  

(4.2.9)

\[
\frac{\partial E_{\text{total}}}{\partial o_1} = -(\text{target}_{o_1} - \text{out}_{o_1}) = \text{out}_{o_1} - \text{target}_{o_1}
\]  

(4.2.10)

The second term in the right hand side of the equation 4.2.8 is the partial derivative of the logistic function and can be calculated as follows:

\[
\text{out}_{o_1} = \frac{1}{1 + e^{-\text{net}_{o_1}}}
\]  

(4.2.11)

\[
\frac{\partial \text{out}_{o_1}}{\partial \text{net}_{o_1}} = \text{out}_{o_1} \times (1 - \text{out}_{o_1})
\]  

(4.2.12)
Figure 4.7: Backpropagation in the output layers.

The last term of the equation 4.2.8 can be calculated as follows:

\[ \text{net}_{o1} = w_5 \times \text{out}_{h1} + w_6 \times \text{out}_{h2} + b_2 \times 1 \quad (4.2.13) \]

\[ \frac{\partial \text{net}_{o1}}{\partial w_5} = \text{out}_{h1} \quad (4.2.14) \]

Combining them all together we can rewrite the equation 4.2.8 as follows:

\[ \frac{\partial E_{\text{total}}}{\partial w_5} = (\text{out}_{o1} - \text{target}_{o1}) \times \text{out}_{o1} (1 - \text{out}_{o1}) \times \text{out}_{h1} \quad (4.2.15) \]

Finally to update the \( w_5 \) weight, we multiply the above gradient with a learning rate \( \eta \) and subtract that from the current value of \( w_5 \) as follows:

\[ w_5^+ = w_5 - \eta \times \frac{\partial E_{\text{total}}}{\partial w_5} \quad (4.2.16) \]

Then we repeat this process to update the other output layer weights \( w_6, w_7, \) and \( w_8 \).
For the hidden layers weights, we use a similar approach and find the gradients as follows:

\[
\frac{\partial E_{\text{total}}}{\partial w_1} = \frac{\partial E_{\text{total}}}{\partial \text{out} h_1} \times \frac{\partial \text{out} h_1}{\partial \text{net} h_1} \times \frac{\partial \text{net} h_1}{\partial w_1} \quad (4.2.17)
\]

\[
\frac{\partial E_{\text{total}}}{\partial \text{out} h_1} = \frac{\partial E_{o1}}{\partial \text{out} h_1} + \frac{\partial E_{o2}}{\partial \text{out} h_1} \quad (4.2.18)
\]

\[
\frac{\partial E_{o1}}{\partial \text{out} h_1} = \frac{\partial E_{o1}}{\partial \text{out} o1} \times \frac{\partial \text{out} o1}{\partial \text{net} o1} \quad (4.2.19)
\]

\[
\frac{\partial E_{o1}}{\partial \text{net} o1} = \frac{\partial E_{o1}}{\partial \text{out} o1} \times \frac{\partial \text{out} o1}{\partial \text{net} o1} \quad (4.2.20)
\]

\[
\frac{\partial \text{net} o1}{\partial \text{out} h_1} = w_5 \quad (4.2.21)
\]

Then we follow the same process for \( \frac{\partial E_{o2}}{\partial \text{out} h_1} \). We can find \( \frac{\partial \text{out} h_1}{\partial \text{net} h_1} \) as follows:

\[
\text{out} h_1 = \frac{1}{1 + e^{-\text{net} h_1}} \quad (4.2.22)
\]

\[
\frac{\partial \text{out} h_1}{\partial \text{net} h_1} = \text{out} h_1(1 - \text{out} h_1) \quad (4.2.23)
\]

Which leads to finding \( \frac{\partial E_{\text{total}}}{\partial w_3} \). Now we can update the weight \( w_1 \) as follows:

\[
w_1^+ = w_1 - \eta \times \frac{\partial E_{\text{total}}}{\partial w_1} \quad (4.2.24)
\]

We repeat the same process for \( w_2, w_3, \) and \( w_4 \) and update them all. This completes one run of the backpropagation. We repeat this process many times to achieve the desired minimum error which leads to the best classification accuracy.

Now let’s take a look at our proposed weighted decision fusion approach. For a regular softmax with loss we have:

\[
p_j = \frac{e^{o_j}}{\sum_k e^{o_k}} \quad (4.2.25)
\]
\[ L = -\sum_j y_j \log p_j \]  
(4.2.26)

\[ \frac{\partial L}{\partial o_i} = p_i - y_i \]  
(4.2.27)

where \( p_j \) is the softmax function and \( L \) is the loss. For a softmax with weighted probabilities we have:

\[ p_{jw} = W \times p_j = \frac{W \times e^{o_j}}{\sum_k e^{o_k}} \]  
(4.2.28)

where \( W \) is the probability weight vector. Then the loss can be written as:

\[ L = -\sum_j y_j \log p_{jw} \]
\[ = -\sum_j y_j \log (W \times p_j) \]
\[ = -\sum_j y_j (\log W + \log p_j) \]
\[ = -\sum_j y_j \log W - \sum_j y_j \log p_j \]
(4.2.29)

Then the derivative of the loss with respect to the weights can be calculated as:

\[ \frac{\partial L}{\partial W} = -\frac{1}{W} \sum_j y_j \]
(4.2.30)

and since the sum of the probabilities out of softmax is 1, i.e. \( \sum_j y_j = 1 \), we have:

\[ \frac{\partial L}{\partial W} = -\frac{1}{W} \]
(4.2.31)

Therefore, we can update the probability weights using the following formula:

\[ W^+ = W - \eta \times \frac{\partial L}{\partial W} = W + \eta \times \frac{1}{W} \]
(4.2.32)
Since we have three networks, we can update the weights as follows:

\[
W_{\text{RGB}}^+ = W_{\text{RGB}} + \eta \times \frac{1}{W_{\text{RGB}} + W_{\text{Mag}} + W_{\text{Phase}}} \tag{4.2.33}
\]

\[
W_{\text{Mag}}^+ = W_{\text{Mag}} + \eta \times \frac{1}{W_{\text{RGB}} + W_{\text{Mag}} + W_{\text{Phase}}} \tag{4.2.34}
\]

\[
W_{\text{Phase}}^+ = W_{\text{Phase}} + \eta \times \frac{1}{W_{\text{RGB}} + W_{\text{Mag}} + W_{\text{Phase}}} \tag{4.2.35}
\]

Since we are working with the probabilities, we need to have the following criteria:

\[
W_{\text{RGB}} + W_{\text{Mag}} + W_{\text{Phase}} \approx 1 \tag{4.2.36}
\]

To apply this constraint we can normalize the weights as follows:

\[
W'_{\text{RGB}} = \frac{W_{\text{RGB}}}{W_{\text{RGB}} + W_{\text{Mag}} + W_{\text{Phase}}} \tag{4.2.37}
\]

\[
W'_{\text{Mag}} = \frac{W_{\text{Mag}}}{W_{\text{RGB}} + W_{\text{Mag}} + W_{\text{Phase}}} \tag{4.2.38}
\]

\[
W'_{\text{Phase}} = \frac{W_{\text{Phase}}}{W_{\text{RGB}} + W_{\text{Mag}} + W_{\text{Phase}}} \tag{4.2.39}
\]

The associated network is depicted in Figure 4.8.
Figure 4.8: Deep neural network based on decision fusion from RGB, magnitude, and phase of the shearlet coefficients.
Chapter 5

Experiments and Results

In this chapter we present the experimental results of the breast and prostate cancer diagnosis and Gleason grading using our proposed methods. We have divided the experiments into two main categories: based on feature engineering, and based on feature learning. For feature engineering, we evaluate our shearlet-based feature representation methods for the breast and prostate cancer diagnosis and Gleason grading of the histological H&E images. We also present our experimental results for the prostate cancer diagnosis in the magnetic resonance images. For feature learning, we evaluate both of our deep neural network methods for the breast cancer diagnosis and Gleason grading of the H&E images. For each experiment, we describe the dataset and parameter settings and compare our results with state-of-the-art. We use various metrics to measure the classification performance of our methods including area under the curve, classification accuracy, sensitivity, specificity, and F1 score. We describe these metrics in the following.

Table 5.1 visualizes the confusion matrix. Confusion matrix is a table that is used to measure the performance of a classification given the true and predicted labels. It includes True Positive (TP), False positive (FP), True Negative (TN), and False
Negative (FN). TP represents the number of instances that were correctly classified as positive. TN is the number of instances that were correctly classified as negative. FP and FN are the number of instances that were incorrectly classified as being positive and negative, respectively. Sensitivity is calculated as $Sensitivity = \frac{TP}{TP + FN}$ and represents the classifier’s ability to correctly detect the patients that have cancer. Specificity is calculated as $Specificity = \frac{TN}{TN + FP}$ and measures the classifier’s ability to correctly detect patients who don’t have cancer. Accuracy is calculated as $Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$ and represents the proportion of total number of correct predictions. F1 score is calculated as $F1 = \frac{2TP}{2TP + FP + FN}$ and is a measure of classifier’s accuracy. Area Under the Curve (AUC) is another classification measure that is calculated as the area under the Receiver Operating Characteristic (ROC) curve and ranges from 0 to 1 with 1 being a perfect classification. An important property of the AUC measure is that it is independent of the test population which makes it suitable for cases where we have unbalanced data.

Table 5.1: Confusion matrix

<table>
<thead>
<tr>
<th>Predicted</th>
<th>True</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TP</td>
</tr>
<tr>
<td></td>
<td>FP</td>
</tr>
</tbody>
</table>

5.1 Based on feature engineering

5.1.1 Breast cancer diagnosis using histogram of shearlet coefficients

This experiment was a pilot study, designed as a proof of concept to show that shearlet transform can be used for medical image analysis. In this subsection, we evaluate the performance of our proposed histogram of shearlet coefficients method.
for the breast cancer diagnosis of the histological images. For this experiment we used the university of California Santa Barbara bio-segmentation benchmark dataset [3]. This publicly-available dataset includes 58 H&E images of the breast tissue consisting of 32 benign and 26 malignant images. A pair of benign and malignant images is presented in Figure 5.1. The images were of high quality and we did not need to do any preprocessing since this is a standard dataset for the breast cancer diagnosis and segmentation.

![Sample breast tissue images](image)

Figure 5.1: Sample breast tissue images of the university of California Santa Barbara bio-segmentation benchmark dataset [3].

To apply the shearlet transform on the images we used the MATLAB [105] toolbox provided by Easley et al. [2]. We converted the images to gray scale since the toolbox would only work on 2-D images and we were not interested in the color information for texture representation of the images. We used 5 decomposition levels of shearlets. Then we extracted two types of features from the images. First we used the discrete shearlet coefficients (DSC) directly. We used 5 decomposition levels with 1, 2, 4, 8, 16 directional filters in each level, respectively. We used $32 \times 32$, $16 \times 16$, and $8 \times 8$ spatial filter size at each decomposition level and reported the results for the filter size that returned the best classification accuracy. Therefore, for a $512 \times 512$
image, we had a feature size of $512 \times 512 \times (1 + 2 + 4 + 8 + 16) = 262,144 \times 31$. To extract the histogram of shearlet coefficients (HSC) from the images, we followed the procedure explained in Section 4.1.1 of chapter 4. We used a fixed number of 59 bins for the histograms. We extracted histograms for each filter at each decomposition level. Therefore, for each image, we extracted a $1 \times 59 \times (1+2+4+8+16) = 1 \times 1829$ feature vector. Figure 5.2 illustrates the DSC and HSC features extracted from the pair of images in Figure 5.1. The blue and red colors represent the features extracted from the malignant and benign images, respectively. Note the visual differences between the benign and malignant features in both the DSC and histograms.
After extracting the DSC and HSC features, we fed them into a classifier for breast cancer diagnosis. We used hold-out cross-validation with half of the data for training and the other half for testing. The samples were randomly drawn from
the dataset. This procedure was repeated 50 times and we reported the average classification results. For classification, we used the MATLAB toolbox for support vector machines [106]. We tried different SVM kernels such as Gaussian, polynomial, and linear with different kernel parameters. We obtained the best results using the linear kernel SVM and reported the results. The average classification accuracy for both of the HSC and DSC methods is reported in Table 5.2. We can observe that both of the HSC and DSC methods achieve good results. This is due to the fact that the larger shearlet coefficients correspond to the stronger edges in the images. Also notice that by increasing the number of decomposition levels, the DSC accuracy slightly increases. This can be due to the fact that higher decomposition levels correspond to the finer details in the images which are not detectable at lower levels.

Table 5.2: Average classification accuracy for breast cancer diagnosis using DSC and HSC methods

<table>
<thead>
<tr>
<th>Number of decomposition levels</th>
<th>HSC</th>
<th>DSC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.73</td>
<td>0.75</td>
</tr>
<tr>
<td>2</td>
<td>0.68</td>
<td>0.73</td>
</tr>
<tr>
<td>3</td>
<td>0.69</td>
<td>0.74</td>
</tr>
<tr>
<td>4</td>
<td>0.65</td>
<td>0.74</td>
</tr>
<tr>
<td>5</td>
<td>0.65</td>
<td>0.75</td>
</tr>
</tbody>
</table>

We also compared our results with state-of-the-art [22,107,108]. The classification results are presented in Table 5.3. These methods are all based on segmentation of the cell nuclei. We are achieving comparable results without going through the cumbersome task of cell nuclei segmentation. It is usually difficult and requires
deep understanding of the biological content of the images. Furthermore, their classification results highly depend on the segmentation accuracy and any error in segmentation adversely affects the classification accuracy. These methods use the wavelet transform which has inherent limitations due to the lack of directional sensitivity. On the other hand, our method is based on the shearlet transform which is excellent for detecting the anisotropic features in the images.

Table 5.3: Comparing our breast cancer diagnosis classification results with state-of-the-art

<table>
<thead>
<tr>
<th>Method</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>F1 score</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSC</td>
<td>0.87</td>
<td>0.77</td>
<td>0.85</td>
<td>0.75</td>
</tr>
<tr>
<td>HSC</td>
<td>0.81</td>
<td>0.76</td>
<td>0.81</td>
<td>0.73</td>
</tr>
<tr>
<td>Weyn et al. [107]</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.76</td>
</tr>
<tr>
<td>Boucheron et al. [22]</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.74</td>
</tr>
<tr>
<td>Van deWouwer et al. [108]</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.65</td>
</tr>
</tbody>
</table>

5.1.2 Prostate cancer diagnosis and Gleason grading via histogram of shearlet coefficients

In previous section, we proved the concept of application of the shearlet transform for the breast cancer diagnosis. In this section, we evaluate the histogram of shearlet coefficients method for the histological prostate cancer diagnosis and Gleason grading. We have divided our experiments to two parts: prostate cancer diagnosis and Gleason grading.

For prostate cancer diagnosis, we acquired 10 H&E images of the prostate gland from the university of Colorado school of medicine. Our data includes 5 benign and 5 malignant images. A pair of benign and malignant prostate tissue image is presented
in Figure 5.3. Notice the change of texture from benign to malignant. Shearlet has
directional sensitivity which makes it an excellent choice to represent the changes in
small contours of carcinoma cells. We utilized the histogram of shearlet coefficients
method with a fixed number of 60 bins to extract features from the images. For
classification, we used a SVM classifier with a linear kernel. We used half of the
data for training and the other half for testing. The samples were drawn randomly.
We repeated this process 50 times and reported the average classification accuracy.
We also extracted features using the histogram of oriented gradients (HOG) method
[109] and compare the classification results with HSC. The classification accuracy,
sensitivity, specificity, and F1 score of these experiments are reported in Table 5.4.
We can observe that our HSC method achieves perfect classification results and
outperforms the HOG method. This is due to directional sensitivity and multiscale
decomposition of the shearlets.

Figure 5.3: Sample prostate tissue images from our dataset. Notice the changes in texture
from benign to malignant.
Table 5.4: Classification results for prostate cancer diagnosis

<table>
<thead>
<tr>
<th>Method</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>F1 score</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSC</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>HOG</td>
<td>1</td>
<td>0.33</td>
<td>0.4</td>
<td>0.6</td>
</tr>
</tbody>
</table>

For the Gleason grading pilot study, we acquired 13 Gleason grade 4 and 5 grade 3 images from the university of Colorado school of medicine. Then we extracted the histogram of shearlet coefficients with 60 bins from the images. We used a SVM with linear kernel for classification. We used leave one out (LOO) cross-validation for training and testing. We compared our classification accuracy with the HOG method. We were able to achieve 0.89 classification accuracy using the HSC method while the accuracy for HOG method was 0.70. We were able to show the superiority of our HSC method over state-of-the-art. In the next section, we will perform more comprehensive experiments for automatic Gleason grading.

5.1.3 Automatic Gleason grading via shearlet transform and multiple kernel learning

5.1.3.1 Methodology

In previous section we showed that the histogram of shearlet coefficients can be used for automatic Gleason grading. In this section, we evaluate our multifeature Gleason grading approach using the shearlet transform and multiple kernel learning. For this purpose, we extract texture features via co-occurrence of the shearlet coefficients and combine them with the color and morphological features using multiple kernel learning and perform classification. Our proposed method is illustrated in Figure 5.4.
Figure 5.4: Block diagram of our proposed system for Gleason grading.
The Gleason grading system [4] is the standard test for grading of the prostate cancer as shown in Figure 5.5. It classifies the prostate cancer from grade 1 to grade 5. The higher Gleason grades represent the higher malignancy levels. The Gleason score is calculated as the sum of the two most dominant grades in the tissue and varies from 2 to 10. Patients with a score of 2 to 4 have a higher chance of survival while patients with the score of 8 to 10 will most probably die of cancer [110].

To better illustrate the changes that the tissues go through when the prostate cancer advances to higher grades, we present samples images in Figure 5.6. Figure 5.6a shows a normal prostate tissue which consists of gland units surrounded by stroma (pink). These gland units consist of lumen (white) surrounded by the epithelial cells (blue). Figure 5.6b shows a cancerous Gleason grade 2 tissue where the
epithelial cells randomly duplicate and disturb the normal structure of the glands. Furthermore, the malignant cells show irregular morphology which can be seen in the cell nuclei. They have larger nuclei and lack sufficient cytoplasm. When the cancer advances further to Gleason grade 5, most of the gland formation and stroma is gone and the cells become poorly differentiated. These phenomena can be observed in Figure 5.6c. This motivated us to extract the color, texture, and morphological features from the images for automatic Gleason grading.

For texture feature representation, we propose the statistics extracted from the co-occurrence of the shearlet coefficients (CSC) as explained in Chapter 4 Section 4.1.2. Since we extract 20 statistics from the co-occurrence matrix, we will have a texture feature vector of size $1 \times 20$ for each image. To better illustrate the effectiveness of the shearlet transform for histological texture representation, we show the third decomposition level shearlet coefficients of the sample Gleason Grades 2 to 5 images in Figure 5.7.
Figure 5.6: Sample prostate tissue images.
Figure 5.7: Sample Gleason grade 2 to 5 images and their corresponding shearlet coefficients. Notice that the shearlet coefficients can highlight the structure of the cell and the random scattering of the epithelial cells.
For color features, we calculate the histogram of the color channels from the RGB, HSV, CIELAB, CIELUV, and YCbCr color spaces. The reason to use the color features is due to the fact that the histological images are stained with H&E staining technology which highlights different structures of the cell. As cancer advances, the structure of the cell goes through major changes which affect the color distribution of the cell. Figure 5.8 shows the red, green, and blue color channel histograms of the sample Gleason grade 2 to 5 images. We can observe that the histogram of green channel moves towards the lower green channel intensity values as the Gleason grade increases. Our observation is in accordance with the results from [43]. However, a similar conclusion cannot be derived from the red or blue color channels. This is due to the fact that the RGB color space has redundant color information and does not match with the human perception of the color. By converting to the other color spaces, we will have more meaningful color information which is close to the human perception of the color. This motivated us to convert the images to the other color spaces and calculate the histograms. Overall, we use 5 color spaces, each has 3 components. We find the color channel histogram using 30 bins which returned the best preliminary results. Therefore, for each image, we calculate a color feature vector of size $1 \times 5 \times 3 \times 30 = 1 \times 450$. 
For the morphological feature representation, we extract the cell nuclei from the images, calculate the cell nuclei area, and use that as a feature. To this end, we propose using the mean shift clustering algorithm [111] for the color approximation and then thresholding in the HSV color space to segment the cell nuclei similar to [112]. We previously explained the color changes of the tissue as cancer happens and advances. However, we cannot directly apply thresholding on the RGB colors to separate the cell nuclei from the rest of the tissue since there are many color features in the image. To reduce the color feature space, we use the mean shift
algorithm. For this purpose, we use a window around each data point and calculate the mean inside it. Then we shift the window to the location of the mean value and repeat the process until it converges. This window will move towards the more congested areas (modes) of the data which helps us find the most important parts of the data. After applying the mean shift on the image which reduces the distinct colors in the image, we convert the image to HSV color space and apply a threshold on the hue value. This is to separate the cell nuclei (blue hue) from the rest of the tissue (pink hue). Therefore, we create a mask image with pixel values of 1 where the hue value of the mean shifted image in HSV space is between 0.70 and 0.88, and pixel values of 0 otherwise. Finally, the morphological feature vector is the number of white pixels in the mask image. Figure 5.9 shows a Gleason grade 2 image with the corresponding segmented nuclei mask image.

(a) Sample Gleason grade 2 image
(b) Segmented mask

Figure 5.9: A sample Gleason grade 2 image and the segmented mask.
5.1.3.2 Results

In this section, we report the classification results of each feature set separately as well as all the feature combined using the MKL algorithm and compare our results with state-of-the-art. For these experiments, we used the Gleason grading images utilized by Jafari et al. [56]. This dataset contained 100 images of Gleason grade 2 to 5 which were labeled by the expert pathologists. The images were of different sizes. Therefore, when extracting features from the images, we divided the features by the image size to have fair comparison. We used the color images for feature extraction except for the shearlet features where we converted the images to gray scale since shearlets did not need the color information. After extracting the features, we used the principal component analysis (PCA) algorithm for feature dimension reduction and to emphasize the variance and highlight the strong patterns in the data which then helps with the classification. To this end, we used the first few eigenvectors that captured at least 90% of the total variance in the data. For training and testing the algorithms, out of 100 images we randomly chose 60 images for the training, 10 images for the validation (e.g. SVM parameters), and 30 images for the testing. We repeated this process 50 times and reported the average classification results. We used the sensitivity, specificity, F1 score, and accuracy to measure the performance of the classification. For classification, we used the SVM classifier with polynomial and Gaussian kernel functions as follows:

\[
K_{Poly}(x_i, x_j) = (x_i, x_j + 1)^q, q \in \mathbb{N}
\]

\[
K_{Gaus}(x_i, x_j) = e^{-\|x_i - x_j\|^2/2S^2}, S \in \mathbb{R}^+\]

(5.1.1)

We used the polynomial kernels with \( q \in \{1, 2, 3\} \) and Gaussian kernels with \( S \in \{0.5, 1, 2, 5, 7, 10, 15, 20, 100, 1000\} \). For MKL, we used the following formula
to combine the features:

\[
K_{i,j} = \sum_{m=1}^{13} \left[ d_m k_m(x_i, x_j) + d_{m+13} k_{m+13}(y_i, y_j) + d_{m+26} k_{m+26}(z_i, z_j) \right] \quad (5.1.2)
\]

Where for \(i-th\) and \(j-th\) samples, \((x_i, x_j)\) is the color feature pair, \((y_i, y_j)\) is the shearlet feature pair, and \((z_i, z_j)\) is the morphological feature pair, \(k_m(., .)\) is one of the 13 kernels described above, and \(d = (d_1, d_2, ..., d_{39})^T\) is the kernel weight vector that will be optimized using the MKL algorithm. We also included the baseline kernel methods [14] as follows:

\[
K_{\text{Average}_{i,j}} = \frac{1}{13} \sum_{m=1}^{13} k_m(x_i, x_j)
\]

\[
K_{\text{Product}_{i,j}} = \left( \prod_{m=1}^{13} k_m(x_i, x_j) \right)^{\frac{1}{13}} \quad (5.1.3)
\]

The classification results using the shearlet coefficients features are presented in Table 5.5. To be able to apply the discrete shearlet transform on the images, first we made the images square size using the bicubic interpolation. Then we used the shearlet toolbox provided by Easley et al. [2] to apply the shearlet transform on the images using 2, 3, 4, and 5 decomposition levels. Finally, we extracted both of the histogram of shearlet coefficients (HSC) with 60 bins and the statistics from the co-occurrence of shearlet coefficients (CSC) as explained in the previous section. We can observe that the histogram of shearlet coefficients does not perform well for this test. As we explained before, this is due to the limitations of the histograms to find the spatial relations between the pixels and the texture of the images. On the other hand, co-occurrence of shearlet coefficients with 5 decomposition levels returns the best results due to the fact that the higher decomposition levels can detect finer details in the image which is suitable for higher Gleason grade images. Overall, we were able to achieve good classification accuracy of 84% using the shearlet features.
Table 5.5: Gleason grading classification results using shearlet features

<table>
<thead>
<tr>
<th>Feature</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>F1 score</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSC</td>
<td>0.59</td>
<td>0.86</td>
<td>0.56</td>
<td>0.58</td>
</tr>
<tr>
<td>CSC, 2 levels</td>
<td>0.61</td>
<td>0.82</td>
<td>0.59</td>
<td>0.59</td>
</tr>
<tr>
<td>CSC, 3 levels</td>
<td>0.68</td>
<td>0.89</td>
<td>0.66</td>
<td>0.67</td>
</tr>
<tr>
<td>CSC, 4 levels</td>
<td>0.76</td>
<td>0.92</td>
<td>0.74</td>
<td>0.78</td>
</tr>
<tr>
<td>CSC, 5 levels</td>
<td>0.84</td>
<td>0.94</td>
<td>0.81</td>
<td>0.84</td>
</tr>
</tbody>
</table>

The classification results using the color features are presented in Table 5.6. We can observe that the green channel histogram is showing promising results as expected. However, the red and blue channel histograms do not return promising classification results. Moreover, we achieved good classification results using the HSV color channels. Furthermore, we were able to improve the classification accuracy of the single color channels by 8% after combining with the other color channels. This further justified our choice of including the other color spaces. Overall, we were able to achieve 90% classification accuracy using the combined color features.

The classification results using morphological features are presented in Table 5.7. To apply the mean shift algorithm on the images we used the MATLAB toolbox provided by Comaniciu et al. [111]. We used the spatial resolution $h_s = 2$, range resolution $h_r = 6.5$ and minimum region area $S = 20$. Then we applied thresholding on the HSV image and calculated the cell nuclei area in the mask. We were able to achieve 90% classification accuracy using the morphological features.

The classification results using each feature separately and combined are presented in Table 5.8. We noticed that each feature set needed a different SVM kernel. Linear, polynomial, and Gaussian kernels returned the best results for color, morphological, and texture features, respectively. This further justified our choice in utilizing MKL for feature combination. Individually, the color channel histograms and morpholog-
Table 5.6: Gleason grading classification results using color channel histograms

<table>
<thead>
<tr>
<th>Color channel</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>F1 score</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td>0.73</td>
<td>0.90</td>
<td>0.70</td>
<td>0.74</td>
</tr>
<tr>
<td>Green</td>
<td>0.84</td>
<td>0.94</td>
<td>0.82</td>
<td>0.82</td>
</tr>
<tr>
<td>Blue</td>
<td>0.78</td>
<td>0.92</td>
<td>0.74</td>
<td>0.75</td>
</tr>
<tr>
<td>Y</td>
<td>0.80</td>
<td>0.93</td>
<td>0.78</td>
<td>0.77</td>
</tr>
<tr>
<td>Cb</td>
<td>0.73</td>
<td>0.91</td>
<td>0.71</td>
<td>0.74</td>
</tr>
<tr>
<td>Cr</td>
<td>0.75</td>
<td>0.89</td>
<td>0.72</td>
<td>0.75</td>
</tr>
<tr>
<td>Hue</td>
<td>0.82</td>
<td>0.94</td>
<td>0.80</td>
<td>0.80</td>
</tr>
<tr>
<td>Saturation</td>
<td>0.78</td>
<td>0.92</td>
<td>0.75</td>
<td>0.78</td>
</tr>
<tr>
<td>Value</td>
<td>0.74</td>
<td>0.91</td>
<td>0.72</td>
<td>0.72</td>
</tr>
<tr>
<td>(L)</td>
<td>0.76</td>
<td>0.94</td>
<td>0.74</td>
<td>0.77</td>
</tr>
<tr>
<td>A</td>
<td>0.72</td>
<td>0.94</td>
<td>0.69</td>
<td>0.74</td>
</tr>
<tr>
<td>B</td>
<td>0.76</td>
<td>0.94</td>
<td>0.74</td>
<td>0.75</td>
</tr>
<tr>
<td>(L')</td>
<td>0.73</td>
<td>0.90</td>
<td>0.75</td>
<td>0.76</td>
</tr>
<tr>
<td>U</td>
<td>0.74</td>
<td>0.94</td>
<td>0.72</td>
<td>0.75</td>
</tr>
<tr>
<td>V</td>
<td>0.72</td>
<td>0.94</td>
<td>0.70</td>
<td>0.74</td>
</tr>
<tr>
<td>Combined</td>
<td>0.90</td>
<td>0.96</td>
<td>0.89</td>
<td>0.90</td>
</tr>
</tbody>
</table>

Table 5.7: Gleason grading classification results using morphological features

<table>
<thead>
<tr>
<th>Feature</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>F1 score</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell nuclei area</td>
<td>0.89</td>
<td>0.97</td>
<td>0.88</td>
<td>0.90</td>
</tr>
</tbody>
</table>

Color and morphological features returned the best classification accuracy. However, when we combined
color and morphological features with the shearlet features, we were able to achieve
higher classification accuracy due to the importance of texture features for Gleason
grading and also shearlet’s ability to capture anisotropic texture features in the tis-
sue. We passed each feature set through the PCA for feature dimension reduction.
The new feature vectors for color channel histograms, morphological features, and
shearlet features are of size \(100 \times 10\), \(100 \times 1\), and \(100 \times 8\), respectively. For feature
combination, we considered two scenarios: using the single kernel SVM and using
the multiple kernel SVM. For the single kernel SVM, we concatenated all of the
features into a single feature vector and used the SVM with polynomial and Gaus-
sian kernels for the classification. For the MKL, we used the MATLAB toolbox for the SimpleMKL provided by Rakotomamonjy et al. [8]. We also implemented the baseline methods average and product kernels and included them in the results. We can observe from the results in Table 5.8 that when combining the features using the single kernel SVM, we are not gaining much in the classification results which illustrates we need a more sophisticated method to combine the features. However, we are able to boost the classification accuracy when using the MKL. MKL outperforms the baseline kernel methods (averaging and product kernel) as well.

<table>
<thead>
<tr>
<th>Method</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>F1 score</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Co-occurrence of shearlets</td>
<td>0.84</td>
<td>0.94</td>
<td>0.82</td>
<td>0.84</td>
</tr>
<tr>
<td>Color channel histograms</td>
<td>0.90</td>
<td>0.96</td>
<td>0.89</td>
<td>0.90</td>
</tr>
<tr>
<td>Morphological features</td>
<td>0.89</td>
<td>0.97</td>
<td>0.88</td>
<td>0.90</td>
</tr>
<tr>
<td>Single polynomial kernel</td>
<td>0.91</td>
<td>0.97</td>
<td>0.97</td>
<td>0.91</td>
</tr>
<tr>
<td>Single Gaussian kernel</td>
<td>0.80</td>
<td>0.93</td>
<td>0.79</td>
<td>0.78</td>
</tr>
<tr>
<td>Averaging kernel</td>
<td>0.89</td>
<td>0.96</td>
<td>0.87</td>
<td>0.89</td>
</tr>
<tr>
<td>Product kernel</td>
<td>0.69</td>
<td>0.90</td>
<td>0.66</td>
<td>0.68</td>
</tr>
<tr>
<td>MKL</td>
<td>0.93</td>
<td>0.98</td>
<td>0.92</td>
<td>0.94</td>
</tr>
</tbody>
</table>

The results of comparing our method with state-of-the-art are presented in Table 5.9. Comparing with the multiwavelet method proposed by Jafari et al. [56], we achieved higher classification results while not using any feature selection techniques or weights on the features. They used the simulated annealing algorithm, which is a slow optimization method with a high chance of getting trapped in the local minimum. Instead, we used the shearlet transform as a robust texture analysis tool and also the MKL algorithm as a feature fusion/classification technique. Comparing with the Gabor filter [113] and HOG [109], we were able to outperform these methods, thanks to the directional sensitivity of the shearlets. Compared to the histogram of shearlet coefficients [78,98], our proposed co-occurrence of shearlet
coefficients achieved better classification results. This is due to the fact that the histograms cannot highlight the texture and the spatial relationships of the singularities in the image. Compared to Zhou et al. [114], they use the same ideas of using the shearlets for texture analysis. However, we were the first group that proposed this method. Furthermore, we extract more statistics from the shearlet coefficients and combine them with the color and morphological features using the MKL which makes our method more robust and perform better.

Table 5.9: Gleason grading classification results comparing with state-of-the-art

<table>
<thead>
<tr>
<th>Method</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>F1 score</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our method</td>
<td>0.93</td>
<td>0.98</td>
<td>0.92</td>
<td>0.94</td>
</tr>
<tr>
<td>Jafari et al. [56]</td>
<td>0.69</td>
<td>0.89</td>
<td>0.66</td>
<td>0.69</td>
</tr>
<tr>
<td>Rezaeilouyeh et al. [78]</td>
<td>0.59</td>
<td>0.86</td>
<td>0.56</td>
<td>0.58</td>
</tr>
<tr>
<td>Schwartz et al. [98]</td>
<td>0.62</td>
<td>0.88</td>
<td>0.57</td>
<td>0.62</td>
</tr>
<tr>
<td>Zhou et al. [114]</td>
<td>0.69</td>
<td>0.88</td>
<td>0.67</td>
<td>0.69</td>
</tr>
<tr>
<td>Gabor [113]</td>
<td>0.52</td>
<td>0.71</td>
<td>0.55</td>
<td>0.50</td>
</tr>
<tr>
<td>HOG [109]</td>
<td>0.45</td>
<td>0.82</td>
<td>0.42</td>
<td>0.47</td>
</tr>
</tbody>
</table>

5.1.4 Prostate cancer detection in Magnetic Resonance (MR) images

The main diagnostic tools for the prostate cancer are serum concentration of prostate specific antigen (PSA), digital rectal exam (DRE), and transrectal ultrasound (TRUS) guided biopsies. However, these methods are inaccurate. When using DRE, deep or small cancers are missed and when using PSA, false positives are common and a 15% false negative rate is reported [115]. Magnetic resonance imaging (MRI) as a noninvasive imaging method has shown superior capabilities due to its improved ability to visualize and localize the prostate gland compared to the TRUS [115]. There are several types of MRI, each evaluating a different anatomical
property. These tests can help determine the size and the location of the prostate tumors, see if the cancer has spread to the other areas, and potentially determine the cancer’s aggressiveness [116]. T2-weighted imaging (T2W) as an anatomic imaging technique can help identify the prostate cancer based on the pathological changes within the prostate. Anatomic imaging is also an excellent technique for evaluating the spread of the cancer outside the gland. However, the image contrast is not very specific. As a result, it is important to obtain functional imaging measurements to most accurately identify the cancer within the prostate. Diffusion Weighted Imaging (DWI) can be used to assess the prostate cancer presence, spatial extent and aggressiveness. DWI is sensitive to the motion of the water molecules in the tissue at microscopic spatial resolution which makes it a suitable choice for the prostate cancer detection. Apparent diffusion coefficient (ADC) map of the diffusion weighted images can be used for risk stratification since it correlates with the histopathological grade of the disease. Dynamic Contrast Enhanced (DCE) MRI is performed by injecting a Gadolinium-based MR contrast agent into the patient and measuring the changes in the uptake and washout of the contrast agent, which is increased in the prostate cancer. DCE can also be used to assess the prostate cancer presence, spatial extent and aggressiveness [116]. These MRI methods have great potential for the prostate cancer diagnosis, however they each have some shortcomings as well. It is difficult to detect the cancer in the transitional and central zones of the prostate when using T2W. Medium prostate cancer grades (e.g. grade 3) and intermediate volume ($\leq 1cc$) tumors may remain undetected when using the DCE. The prostate cancer detection accuracy may depend on the size and the location of the tumor (anterior vs. posterior) when using the ADC. On the other hand, multiparametric MRI (Mp-MRI) integrates several tests (mentioned above) to give the physician a more complete picture of the patient’s condition. The combined results can let the physician know about the severity of the disease, and inform a treatment plan for
the patient [116]. Interpretation of Mp-MRI data is labor intensive, expensive, and highly operator-dependent. To assist the radiologist to correctly interpret and diagnose prostate cancer on Mp-MRI, we propose a computer-aided diagnostic (CADx) method in this dissertation. To this end, we apply the shearlet transform on the Mp-MRI region of interest (ROI) and extract the histogram of shearlet coefficients from the ROIs. Then we use the SVM to classify the ROIs as benign or malignant. Our proposed method is illustrated in Figure 5.10.

![Figure 5.10: Our proposed method for Mp-MRI prostate cancer detection.](image)

To evaluate the performance of our histogram of the shearlet coefficients method for the prostate cancer diagnosis, we acquired the Mp-MRI data from 4 patients at the university of Colorado hospital. Our Mp-MRI data included the T2W, ADC, and DCE imaging sequences. The prostate border and tumor boundaries were
manually delineated using our radiologist with the help of the histopathological maps provided by our pathologist. Figure 5.11 shows some sample Mp-MRI sequences from our dataset. The red color represents the tumor boundary. Notice the visual differences in the tumor appearance from one MRI sequence to another. 10 benign and 10 malignant ROIs were extracted from each Mp-MRI imaging sequence for each patient. The ROIs were upsampled to $320 \times 240$ to be able to apply the shearlet transform on them. We used 4 decomposition levels of the shearlets and extracted the histogram of shearlets using 60 bins and concatenated them and used them as the feature vector for the classification. Then we used the SVM with linear kernel to perform the classification. We divide our data to two halves and use half of data for training and the other half for testing. We repeated this process 50 times and reported the average classification sensitivity, specificity, and accuracy. We also compared our method with the Gabor [113] filters and the HOG method [109]. The classification results are presented in Table 5.10. We can interpret the results from different perspectives. From the intra-patient point of view, we can conclude that the ADC and DCE sequences work better than T2W. However, patient #4 has great classification results using T2W which emphasizes on the necessity of including the different MRI sequences for a better diagnosis. From the inter-patient point of view if we combine the ROIs from all patients and perform classification, we can conclude that the DCE has the best performance. Furthermore, comparing to the HOG and Gabor methods, our method achieves much better classification results. This is due to the fact that the shearlet transform has inherent directional sensitivity, which makes it suitable for characterizing the small contours of cancer cells.
Figure 5.11: Sample Mp-MRI imaging sequence from our dataset. The red color indicates the tumor boundaries. Notice the visual differences in the tumor appearance from one Mp-MRI sequence to another.
Table 5.10: Prostate cancer classification results for Mp-MRI

<table>
<thead>
<tr>
<th>patient #</th>
<th>Parameter</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ADC</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>DCE</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>T2W</td>
<td>0.60</td>
<td>1</td>
<td>0.80</td>
</tr>
<tr>
<td>2</td>
<td>ADC</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>DCE</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>T2W</td>
<td>0.80</td>
<td>0.60</td>
<td>0.70</td>
</tr>
<tr>
<td>3</td>
<td>ADC</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>DCE</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>T2W</td>
<td>1</td>
<td>0.60</td>
<td>0.80</td>
</tr>
<tr>
<td>4</td>
<td>ADC</td>
<td>0.80</td>
<td>1</td>
<td>0.90</td>
</tr>
<tr>
<td>4</td>
<td>DCE</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>T2W</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Combined</td>
<td>ADC</td>
<td>0.89</td>
<td>1</td>
<td>0.97</td>
</tr>
<tr>
<td>Combined</td>
<td>DCE</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Combined</td>
<td>T2W</td>
<td>0.92</td>
<td>0.83</td>
<td>0.94</td>
</tr>
<tr>
<td>HOG</td>
<td>ADC</td>
<td>0.90</td>
<td>0.35</td>
<td>0.63</td>
</tr>
<tr>
<td>HOG</td>
<td>DCE</td>
<td>0.70</td>
<td>0.80</td>
<td>0.75</td>
</tr>
<tr>
<td>HOG</td>
<td>T2W</td>
<td>0.80</td>
<td>0.90</td>
<td>0.85</td>
</tr>
<tr>
<td>Gabor</td>
<td>ADC</td>
<td>0.70</td>
<td>1</td>
<td>0.85</td>
</tr>
<tr>
<td>Gabor</td>
<td>DCE</td>
<td>0.60</td>
<td>0.75</td>
<td>0.68</td>
</tr>
<tr>
<td>Gabor</td>
<td>T2W</td>
<td>0.65</td>
<td>0.40</td>
<td>0.53</td>
</tr>
</tbody>
</table>

5.2 Based on feature learning

5.2.1 Breast cancer diagnosis and prostate Gleason grading via deep learning

In this subsection, we evaluate our deep learning framework explained in Chapter 4 Section 4.2.1. To this end, we acquire the histological images of the breast cancer and prostate Gleason grading. Then we apply the fast finite shearlet transform (FFST) on the images and extract the magnitude and phase of the shearlet coefficients at different directions and scales. Then we feed the shearlet features along with the RGB data to our deep neural network for feature learning and classification. Our deep neural network is a convolutional neural network that consists of
several layers of convolution and max pooling followed by fully connected layers. A block diagram of our approach is depicted in Figure 5.12. Here the procedure is divided into two parts. First we train our CNN using the training data and then test our approach using the trained CNN and test data.

We evaluated our method for the breast cancer diagnosis and prostate Gleason grading. For the breast cancer diagnosis, we used the University of California, Santa Barbara Biosegmentation Benchmark dataset [3] which contained 58 images of benign and malignant tissues. For Gleason grading, we used the the prostate Gleason grading dataset used by Jafari et al. [56]. This dataset contained 100 images of Gleason grades 2 to 5. We needed large amount of the data to train our deep neural network. Therefore, we augmented both datasets. To this end, we performed the mirroring, patches, rotation, and scaling of the images. We used three mirroring scenarios (horizontal, vertical, and horizontal and vertical). Then we rotated each image counter clockwise 10 times with a rotation randomly chosen between 10 and 90 degrees. For scaling, we resized each image by a factor of 2. We extracted patches from the top left, top right, bottom left, bottom right, and center of the image, each half the size of the original image. We also combined the above operations to further augment the datasets. Overall we were able to create 104 images out of each original image. Therefore, after data augmentation, we had 6,032 breast tissue images and 10,400 Gleason grading images. All the images were resized to 128 × 128 for normalization purposes. Figure 5.13 shows the data augmentation process for a sample breast tissue image.
Figure 5.12: Our proposed method for medical image analysis via shearlet transform and deep learning.
To extract the shearlet coefficients from the images, first we apply the fast finite shearlet transform (FFST) on each image using the MATLAB toolbox provided by Hauser et al. [10]. We used five decomposition levels with 1, 8, 8, 16, and 16 directions in each level, respectively. Therefore, we had a total of $1 + 8 + 8 + 16 + 16 = 49$ subbands of the shearlets. Then we extracted the magnitude and phase of the shearlet coefficients from each subband. These shearlet features along with the RGB images were the input to our deep neural network. Figure 5.14 shows a sample pair of benign and malignant breast tissue images along with the magnitude and phase of the shearlet coefficients extracted from them. We can observe the changes in the texture of the image as the tissue transforms from benign to malignant. These changes are reflected in the magnitude of the shearlet coefficients since the
magnitude is a direct representative of the edges in the image. However, the phase of the shearlet coefficients is non-trivial which justifies our choice of feature learning instead of feature engineering.

We implemented our convolutional neural network in Caffe [117]. Our CNN consisted of 3 layers of convolution and max-pooling. For the convolutional layers, we used 64 Gaussian filters of size $5 \times 5$ with a standard deviation of 0.0001 and bias of zero. The step between each filter application was 2 pixels. For the activation function we used a ReLU function. For the max-pooling layer, we applied it on the local patch of units inside a $3 \times 3$ region of the input feature map with a 2 pixels step between the pooling regions. We used the LRN layers to normalize the local input regions and the fully connected layers to concatenate the outputs of the CNNs. We used the stochastic gradient descent algorithm with the momentum of 0.9 and the weight decay of 0.05 in all experiments. We used mini-batches of 32 samples. We used the initial learning rate of 0.001 for our models. We also used dropout layers with a threshold of 0.7 to prevent the overfitting. We found these values based on the performance of the validation set. Same parameters were used for both breast cancer diagnosis and Gleason grading. For sampling the data for the training and test, we performed the fivefold cross-validation. For this purpose, we divided our non-augmented original dataset into 5 sets and used 4 sets for the training and 1 set for the test. This process was repeated 5 times and the average classification results were reported. The augmented images were only used for the training. We only used the original images for the test. Therefore, each image had been used for either the training or the test.

We designed different scenarios for each classification task. First we used only the RGB images as the input to the CNN. Then we combined the RGB with the magnitude of the shearlets. Finally, we combined the RGB images with the mag-
Figure 5.14: Sample benign and malignant breast tissue images and their corresponding magnitude and phase of the shearlet coefficients.
nitude and phase of the shearlets. Separating the input in such manner helped us investigate the contribution of each feature set separately and combined. The classification results for the breast cancer diagnosis and Gleason grading are presented in Tables 5.11 and 5.12, respectively. We can observe that we are able to boost the average classification accuracy of the breast cancer diagnosis and Gleason grading by 15% and 12%, respectively, after adding the magnitude and phase of the shearlet coefficients to the RGB images. We also compared our results with state-of-the-art methods based on feature engineering. It is clear that our methods outperform state-of-the-art.

To inspect the CNN filters we visualize the convolutional filters in Figure 5.15. Notice how the convolutional filters for the RGB and shearlets change from layer 1 to layer 3. We also visualize the features in Figure 5.16. This figure shows the RGB and shearlet features as they go through each convolution layer. Notice how these features evolve as they advance through the network.

To further investigate the performance of our deep neural network method we plot the receiver operating characteristic (ROC) curves for the breast cancer diagnosis using our best DNN method (RGB+ Magnitude+ Phase of the shearlet coefficient) and the best state-of-the-art method based on the feature engineering (Boucheron et al. [22]) in Figure 5.17. An ROC curve shows the true positive rate against the false positive rate using different thresholds. Based on the ROC curves and the area under the curve (AUC) values we can conclude that our deep neural network method outperforms the best feature engineering method. Furthermore, the confusion matrix for the Gleason grading experiments is presented in Figure 5.18. Here we compare our deep neural network method with the best feature engineering method (Jafari et al. [56]). We can observe that our method can perfectly classify the Gleason grade 2, 3, and 4 images. The only misclassified cases happen
Table 5.11: Classification results for breast cancer diagnosis using shearlet transform and deep learning (mean±std).

<table>
<thead>
<tr>
<th>Method</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>F1 score</th>
<th>AUC</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>RGB</td>
<td>0.91±0.08</td>
<td>0.59±0.09</td>
<td>0.76±0.05</td>
<td>0.68±0.02</td>
<td>0.71±0.02</td>
</tr>
<tr>
<td>RGB + Magnitude of shearlets</td>
<td>1</td>
<td>0.62±0.10</td>
<td>0.84±0.03</td>
<td>0.78±0.01</td>
<td>0.84±0.01</td>
</tr>
<tr>
<td>RGB + Magnitude + Phase of shearlets</td>
<td>1</td>
<td>0.72±0.10</td>
<td>0.89±0.03</td>
<td>0.82±0.01</td>
<td>0.86±0.03</td>
</tr>
<tr>
<td>Boucheron et al. [22]</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.74</td>
</tr>
<tr>
<td>Rezaeilouyeh et al. [78]</td>
<td>0.93±0.09</td>
<td>0.60±0.10</td>
<td>0.79±0.06</td>
<td>0.74±0.02</td>
<td>0.74±0.09</td>
</tr>
</tbody>
</table>
Table 5.12: Classification results for Gleason grading using shearlet transform and deep learning (mean±std).

<table>
<thead>
<tr>
<th>Method</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>F1 score</th>
<th>AUC</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>RGB</td>
<td>0.80±0.02</td>
<td>0.91±0.01</td>
<td>0.71±0.01</td>
<td>0.72±0.02</td>
<td>0.76±0.06</td>
</tr>
<tr>
<td>RGB + Magnitude of shearlets</td>
<td>0.84±0.01</td>
<td>0.91±0.02</td>
<td>0.81±0.03</td>
<td>0.79±0.02</td>
<td>0.84±0.04</td>
</tr>
<tr>
<td>RGB + Magnitude + Phase of shearlets</td>
<td>0.89±0.01</td>
<td>0.94±0.01</td>
<td>0.85±0.02</td>
<td>0.84±0.01</td>
<td>0.88±0.05</td>
</tr>
<tr>
<td>Jafari et al. [56]</td>
<td>0.82±0.01</td>
<td>0.91±0.02</td>
<td>0.73±0.02</td>
<td>0.78±0.02</td>
<td>0.83±0.09</td>
</tr>
<tr>
<td>Rezaeilouyeh et al. [79]</td>
<td>0.78±0.03</td>
<td>0.91±0.01</td>
<td>0.69±0.03</td>
<td>0.74±0.01</td>
<td>0.78±0.11</td>
</tr>
<tr>
<td>Wavelet packet [118]</td>
<td>0.82±0.02</td>
<td>0.92±0.01</td>
<td>0.73±0.01</td>
<td>0.74±0.02</td>
<td>0.78±0.07</td>
</tr>
<tr>
<td>Co-occurrence matrix [7]</td>
<td>0.81±0.01</td>
<td>0.92±0.01</td>
<td>0.72±0.02</td>
<td>0.73±0.02</td>
<td>0.77±0.09</td>
</tr>
</tbody>
</table>
Figure 5.15: Convolutional layer filters from three different layers for RGB and magnitude of the shearlet coefficients.
Figure 5.16: Convolutional layer features from three different layers for RGB and magnitude of the shearlet coefficients.
in the grade 5 which is the most difficult task in the Gleason grading [43]. However, the best feature engineering method has some misclassified cases in the grades 3, 4, and 5 and shows less accuracy comparing to our method.

Figure 5.17: ROC curves for breast cancer diagnosis using feature engineering (red) and our deep neural network (blue).
Figure 5.18: Confusion matrices for Gleason grading.

(a) Confusion matrix for Gleason grading using feature engineering.

(b) Confusion matrix for Gleason grading using our deep neural network method.
5.2.2 Automatic Gleason grading via weighted decision fusion framework

In this subsection, we evaluate our weighted decision fusion deep learning framework explained in Chapter 4 Section 4.2.2. To this end, we use the same histological images of the prostate Gleason grading as in the previous subsection and we extract the data the same way. We use the same CNN but we design a different network this time. We feed the RGB, magnitude, and phase data to separate CNNs. Then we calculate the probabilities and assign weights on them and update those weights through backpropagation. Our network's graph is depicted in Figure 5.19.

For this experiment we evaluated our method against our Gleason grading dataset. This dataset contained 100 images of the prostate tissue with Gleason grades 2 to 5. Figure 5.20 shows some samples of this dataset. We augmented this dataset to 10400 images as explained in the previous subsection. Then we applied the shearlet transform on the images and extracted the magnitude and phase of the shearlet coefficients. We fed these shearlet data along with the images to our deep neural network.

We implemented our deep neural network in the Tensorflow [119]. We used the same CNN structure and parameters as the previous subsection. The training accuracy is depicted in Figure 5.21. We were able to achieve perfect training accuracy after 5500 iterations. The network loss is shown in Figure 5.22. The loss was minimized to 0.0025 after 5500 iterations.

Figure 5.23 shows the decision fusion weights during the training. We can observe that the weights are being updated during the training which justifies our choice of assigning different weights for the RGB, magnitude, and phase of the shearlet coefficients CNN outputs. All three weights were initialized at 1/3 and changed
Figure 5.19: Our network's graph based on weighted decision fusion from RGB, magnitude, and phase of shearlet coefficients.
Figure 5.20: Gleason grade 2-5 samples of our dataset.

Figure 5.21: Training accuracy vs. iterations. Notice how the accuracy increases as the number of iterations increases.
Figure 5.22: Training loss vs. iterations. Notice how the network's loss decreases as the number of iterations increases.

during the training. We tried different initialization for the weights and the initialization value of 1/3 returned the best results. The final values of the decision fusion weights for the RGB, magnitude, and phase of the shearlets were 0.35, 0.33, and 0.32, respectively.

To illustrate the effectiveness of CNN filters we visualize the convolutional filters for the RGB, magnitude, and the phase of shearlets randomly selected from 3 layers of CNN in Figure 5.24. Notice how the convolutional filters represent the texture of the prostate tissue image. The RGB filters are more visually informative than the magnitude and phase of shearlets. The neuron representing the magnitude filters from the first CNN is off, hence the dark image. We also visualize the features in Figure 5.25. This figure shows the RGB, magnitude, and the phase of shearlet features as they go through each pooling layer. Notice how these features represent the prostate tissue.
(a) Decision fusion weight for RGB CNN.

(b) Decision fusion weight for magnitude of shearlets CNN.

(c) Decision fusion weight for phase of shearlets CNN.

Figure 5.23: Decision fusion weights for RGB, Mag, and Phase CNNs.
Figure 5.24: Convolutional layer filters from three different layers for RGB, magnitude, and phase of shearlets.
For testing the algorithm, we utilized the same classification scheme as the previous subsection. Table 5.13 shows our classification results. We compared our weighted fusion method with our previous regular CNN method, deep Residual
Network (ResNet) [120], state-of-the-art feature engineering methods, and the majority voting. We were able to improve the classification accuracy to 0.92 using our weighted decision fusion method. Comparing to our previous regular CNN method, we achieved higher classification accuracy by assigning weights on the decisions. We implemented ResNet with 14 layers which included 6 residual blocks as depicted in Figure 5.26. Comparing to ResNet [120], we achieved higher classification accuracy which justifies our choice in network architecture. Comparing to majority voting, we achieved higher classification accuracy which justifies our choice of learning the fusion weights instead of simple majority voting of the labels. Our method also outperformed the state-of-the-art feature engineering methods.
Table 5.13: Classification results for Gleason grading using weighted decision fusion and deep learning (mean±std).

<table>
<thead>
<tr>
<th>Method</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>F1 score</th>
<th>AUC</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weighted decision fusion</td>
<td>0.91±0.02</td>
<td>0.95±0.01</td>
<td>0.89±0.01</td>
<td>0.88±0.01</td>
<td>0.92±0.03</td>
</tr>
<tr>
<td>ResNet [120]</td>
<td>0.87±0.01</td>
<td>0.91±0.01</td>
<td>0.83±0.02</td>
<td>0.83±0.01</td>
<td>0.87±0.03</td>
</tr>
<tr>
<td>Majority voting</td>
<td>0.88±0.02</td>
<td>0.92±0.01</td>
<td>0.84±0.02</td>
<td>0.83±0.01</td>
<td>0.87±0.03</td>
</tr>
<tr>
<td>Regular CNN</td>
<td>0.89±0.01</td>
<td>0.94±0.01</td>
<td>0.85±0.02</td>
<td>0.84±0.01</td>
<td>0.88±0.05</td>
</tr>
<tr>
<td>Jafari et al. [56]</td>
<td>0.82±0.01</td>
<td>0.91±0.02</td>
<td>0.73±0.02</td>
<td>0.78±0.02</td>
<td>0.83±0.09</td>
</tr>
<tr>
<td>Rezaeilouyeh et al. [79]</td>
<td>0.78±0.03</td>
<td>0.91±0.01</td>
<td>0.69±0.03</td>
<td>0.74±0.01</td>
<td>0.78±0.11</td>
</tr>
<tr>
<td>Wavelet packet [118]</td>
<td>0.82±0.02</td>
<td>0.92±0.01</td>
<td>0.73±0.01</td>
<td>0.74±0.02</td>
<td>0.78±0.07</td>
</tr>
<tr>
<td>Co-occurrence matrix [7]</td>
<td>0.81±0.01</td>
<td>0.92±0.01</td>
<td>0.72±0.02</td>
<td>0.73±0.02</td>
<td>0.77±0.09</td>
</tr>
</tbody>
</table>
Figure 5.26: ResNet architecture.
Chapter 6

Conclusions

6.1 Conclusions

In this dissertation we presented our developed feature representation and learning methods for the breast and prostate cancer diagnosis and Gleason grading. We compared our methods with state-of-the-art and showed the superiority of our methods in terms of the classification measures. Our main conclusions are as follows.

We employed the shearlet transform as our main analysis tool due to its directional sensitivity and multiscale framework which makes the shearlet an excellent choice to detect the anisotropic features in the small carcinoma cells. We designed three methods to extract the features from the shearlet coefficients: the discrete shearlet coefficients (DSC), the histogram of shearlet coefficients (HSC), and the statistics extracted from the co-occurrence of the shearlet coefficients (CSC). We further improved our CSC method by combining the CSC features with the color and morphological features. We employed the multiple kernel learning (MKL) algorithm to combine the aforementioned features and perform classification via SVM. We evaluated our proposed methods for the histological breast cancer diagnosis, prostate cancer diagnosis, and Gleason grading. We also evaluated our HSC method for the
MRI prostate cancer diagnosis. We compared our results with state-of-the-art and showed that our methods outperform the other methods.

Due to the limitations of the feature engineering methods, we further investigated the impact of the deep neural networks in representing the histological images for the cancer detection. First, we designed a framework for automatic cancer diagnosis and grading via deep learning and shearlet transform. To this end, we extracted the magnitude and phase of the shearlet coefficients and used them along with the RGB images as the inputs to our deep neural network. Our deep neural network was a convolutional neural network consisting of multiple layers of convolution and pooling followed by the fully connected layers. Then, we designed a weighted decision fusion neural network to empower the contributions of different feature sets. We performed extensive experiments using both frameworks for different classification tasks to show the generalizability and superiority of our deep learning techniques compared with the state-of-the-art.

Despite the limitations, the results of our studies are very promising. It is possible that with further development and evaluations, a CADx system for the breast and prostate cancer detection and grading can be developed and deployed in the clinical environments.

6.2 Future research

This dissertation is very comprehensive in terms of the different aspects and methods of automatic cancer diagnosis since we developed both computer vision and machine learning techniques to address this issue. We further improved our machine learning methods via deep neural network which is the state-of-the-art in machine learning nowadays. Therefore, for future research, I would recommend exploring
the possibility of combining different image modalities for a more comprehensive diagnosis. A good example would be registering the histological slide and the MRI volume for tumor assessment.

Another possible area of research to explore would be the application of our proposed methods for the tumor localization in the histological images. We showed in this dissertation that the shearlet transform is suitable for detecting carcinoma cells due to its inherent directional sensitivity and multiscale framework that enables it to detect different edges in the tissue images. Therefore, it would be interesting to see how our deep neural network equipped with the shearlets performs for pixel-based localization of the tumors in H&E images.

6.3 Publications and Patent

Our publications are listed below in a chronological order:


- H. Rezaeilouyeh, M. H. Mahoor, J. J. Zhang, F. G. La Rosa, S. Chang, and P. N. Werahera, "Diagnosis of prostatic carcinoma on multiparametric magnetic resonance imaging using shearlet transform," in Engineering in Medicine and


Considering the extensive amount of research performed during my PhD research and our promising results, the university of Denver tech transfer office decided to file a provisional patent on our medical image analysis methods. The patent title and filing number are as follows:


We have filed the Nonprovisional Patent Application for this patent. This patent is based on our proposed multifeature medical image analysis method via shearlet transform and MKL and also our deep neural network methods for medical image analysis.
Bibliography


[44] F. De Cobelli, S. Ravelli, A. Esposito, F. Giganti, A. Gallina, F. Montorsi, and A. Del Maschio, “Apparent diffusion coefficient value and ratio as non-


[80] H. Rezaeilouyeh, M. H. Mahoor, J. J. Zhang, F. G. La Rosa, S. Chang, and P. N. Werahera, “Diagnosis of prostatic carcinoma on multiparametric mag-


