QATAR UNIVERSITY

COLLEGE OF ENGINEERING

SEGMENTATION OF TUMOR REGIONS IN MICROSCOPIC IMAGES OF BREAST CANCER TISSUE

BY

DHOHA ABID

A Thesis Submitted to the Faculty of College of Engineering
in Partial Fulfillment of the Requirements
for the Degree of Master of Science

February 2016

Committee

The members of the Committee approve the thesis of **Dhoha Abid** defended on 18 January 2016.

<div align="right">

Dr. Nasir Mahmood Rajpoot
Thesis/Dissertation Supervisor

</div>

<div align="right">

Dr. Sumaya Al-Maadeed
Committee Member

</div>

<div align="right">

Prof. Sebti Foufou
Committee Member

</div>

<div align="right">

Prof. Chang-Tsun Li
Committee Member

</div>

Approved:

Dr. Rashid A. Qader M. Alammari, College of Engineering

Abstract

Nowadays, advances in the domain of digital pathology gave the means to replace the old optical microscopes by the Whole Slide Imaging (WSI) scanners. These scanners enable pathologists viewing conventional tissue slides on a computer monitor. Currently, several applications that aim to analyze human tissue are evolving remarkably. Segmentation of tumor regions in microscopic images of breast cancer tissue in one of the application that researchers are investigating extensively. Indeed, researchers are interested in such application not only because breast cancer is one of the pervasive cancers for human beings, but also segmentation is one of the basic and frequent tasks that pathologists have to perform in order to perform tissue analysis.

In this thesis, we addressed the task of segmentation of tumor regions in microscopic images of breast cancer tissue as a machine learning task. We developed different supervised and unsupervised learning frameworks. Our proposed frameworks encompass five steps: (1) pre-processing, (2) feature extraction, (3) feature reduction, (4) supervised and unsupervised learning, and (5) post-processing. We focused on the extraction of textural features, as well as utilization of supervised learning techniques. We investigated individually the MR8Fast, Gabor, and Phase Gradient features, as well as a combination of all these features. We investigated also different classifiers which are Naive Bayes, Artificial Neural Network, and Support Vector Machine, as well as a combination of the supervised learning results.

We conducted different experiments in order to compare the different proposed frameworks. Therefore, we developed different conclusions. The MR8Fast features are the most discriminating features compared to the Gabor and Phase Gradient that come in the second and third place respectively. Furthermore, the Naive Bayes classifier and the combination of classification results, that have been overlooked for the segmentation of tumor regions in microscopic images of breast cancer tissue, achieved better results compared to the Support Vector Machine classifier which has been extensively employed for this task. These promising conclusions promote the need for further work to investigate other textural features as well as other classifiers.

Contents

List of Figures

List of Tables

Acknowledgements

To my dear mother. . .

Chapter 1

Introduction

Breast cancer is the top cancer among women in developed and developing countries, and it represents 25% of the overall cancers [10]. Statistics indicate that this disease killed more than half a million women in 2012 where 58% of the total breast cancer deaths occur in developing countries compared to 42% that occur in developed countries. The percentage of deaths is less in developed countries because appropriate treatments for this disease are more accessible and available [13]. Breast cancer holders can survive and have a normal life again, if they get the right treatment at the beginning of their ailment [10].

Yet, there are many appropriate treatments for breast cancer, and every treatment depends on the stage of this disease [6]. A breast cancer holder has to receive the right treatments for her stage, otherwise she may die. To identify the breast cancer stage, pathologists - medical specialists who diagnose tissue samples - have to examine the breast cancer tissue. In the past, pathologists used to examine breast cancer tissue by visualizing it through the optical microscope slides [25]. While examining the breast cancer tissue, pathologists endeavor to identify the tumor regions in order to grade and score the severity of the disease [6]. This task was completely manual which makes it cumbersome and time consuming.

However, with the advances of the digital pathology [22, 25], the old optical microscopes are gradually replaced by the technology of Whole Slide Imaging (WSI) [22]. This technology does not only enable the digitization of tissue slides, but also automatic tissue analysis. The automatic tissue analysis include tasks such as identification of tumor regions and grading the cancer disease.

In this study, we consider identification of the tumor regions in breast cancer tissue. This task is also known as **segmentation of tumor regions in microscopic images of breast cancer tissue**. We detail the task of segmentation of tumor regions in the following section.

1.1 Segmentation of Tumor Regions

The segmentation of tumor regions in microscopic images consists of identifying and recognizing tumor regions from the non-tumor regions. To illustrate this task, we present in Figure 1.1 a microscopic image of breast cancer tissue with the corresponding ground truth data.
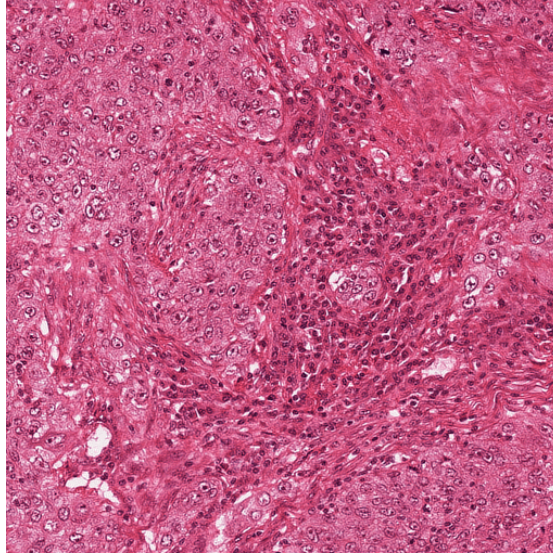
(A) A sample of microscopic image of breast cancer tissue

(B) Ground Truth -1-

(C) Ground Truth -2-

FIGURE 1.1: Microscopic image of breast cancer tissue with the corresponding ground truths data labeled by two different pathologits. The white labels refer to the tumor regions, while the black labels are assigned to the non-tumor regions.

Figure 1.1a presents a microscopic image of breast cancer tissue which is the input for the segmentation task, while Figures 1.1b and 1.1c present the ground truth data which what should be the output for this task.

The task of segmentation of tumor regions is one of the basics tasks in medical image analysis that doctors do often before analyzing the disease. This task has been applied

in past decade in several medical domains: colon cancer [16], pancreas cancer [32], and breast cancer [8, 9, 11, 18–20, 26, 27]. The solution for this task is not unique, and researchers are still investigating new techniques in order to achieve better results.

In this study, we consider the task of segmentation of tumor regions in microscopic images for breast cancer tissue. We address this task as a machine learning task as it has been done for the past decade [8, 9, 11, 18–20, 26, 27].

1.2 Motivation & Challenges

We are motivated to study the segmentation of tumor regions in microscopic images of breast cancer tissue for mainly the following reasons:

1. The manual segmentation of tumor regions in microscopic images of breast cancer tissue is labor intensive and time consuming. Automating this process would save the time of pathologists to do other tasks which cannot be performed automatically. On the other hand, saving the time of pathologists is also convenient for patients that would receive their analysis results earlier.

2. Although the task of segmentation of tumor regions in microscopic images of breast cancer tissue has been the interest of many researchers for the past decade [8, 9, 11, 18–20, 26, 27], this task is still subject of investigation for new techniques that achieve better segmentation results.

3. The inherited subjectivity of the task of segmentation of tumor regions in microscopic images of breast cancer tissue. Pathologists, often, judge differently the breast cancer tissue. What can be considered tumor regions by a particular pathologists, can be annotated as non-tumor regions by another pathologists. We illustrate this conflict in Figure 1.1 that present a microscopic image of breast cancer tissue annotated by two different pathologists. We can notice that the first ground truth, depicted in Figure1.1b, is different from the second ground truth presented in Figure 1.1c. This is because they are annotated by different doctors.

1.3 Public Dataset: MITOS 2012

We used the publicly available MITOS 2012 dataset[1]. This dataset was prepared by the team of professor Frédérique Capron at the pathology department of Pitié-Salpêtrière Hospital in Paris, France. The goal of preparing this dataset was for a competition that sought to better detect and grade mitotic cells automatically. However in this study, we

---

[1]http://ipal.cnrs.fr/ICPR2012

employ this dataset for the task of segmentation of tumor regions. We used this dataset because we could insure the ground truth for the segmentation of tumor regions. We elaborate more this dataset in Section 4.1.1 of Chapter 4.

1.4   Aim & Contributions

In this study, we addressed the tumor segmentation in microscopic images of breast cancer tissue as a machine learning task. We developed several frameworks that encompass the following steps: (1) Pre-processing, (2) Feature extraction, (3) Feature reduction, and (4) Learning process, and (5) Post-processing. We focus mainly on the feature extraction and the learning process step steps, and in this context our contributions can be summarized as follows:

1. We investigate individually different textural features: MR8Fast [5], Gabor [14], and Phase Gradient [24], as well as a combination of all these features. In addition, we rank these features from the most discriminating - that materialize better the differences between the tumor and non-tumor textures - to the less discriminating features.

2. We develop different supervised and unsupervised learning frameworks. We focus on the supervised learning frameworks that employ different classifiers. We investigate the Naive Bayes and Artificial Neural Network classifiers that have been overlooked for the task of segmentation of tumor regions in microscopic images for breast cancer tissues. We investigate also a combination of different classifiers results which would improve the results.

3. In order to reduce the number of training pixels, we develop the *Mean* approach. The *Mean* approach divides the image into equal patches. Then it computes the mean of the pixel intensities of every single patches in order to reduce the training data. We describe more this approach in Section 4.4.1 of Chapter 4.

1.5   Thesis Outline

The remainder of this thesis is organized as follows: we introduce in Chapter 2 a comprehensive background about image segmentation and breast cancer tissue, as well as the related work for segmentation of tumor regions in microscopic images of breast cancer tissue. We detail all the steps of the proposed frameworks in Chapter 3. We present the experimental evaluation results of the different proposed frameworks in Chapter 4. Finally, we conclude this study in Chapter 5.

Chapter 2

Background and Related Work

In this chapter, we present in the first section a comprehensive background on (1) image segmentation, (2) preparation of the histopathological images of breast cancer tissue, and (3) differences between tumor regions and non-tumor regions. In the second section of this chapter, we present the related work on segmentation of tumor regions in microscopic images of breast cancer tissue.

2.1 Background

2.1.1 Image Segmentation

Image segmentation is the task of dividing an image into different segments - a set of pixels - in order to have a different representation of the image [15]. Generally, image segmentation is performed according to some homogeneity criteria such as color, intensity, and texture. As well, image segmentation is the task that identifies and locates objects and boundaries within the image [15].

Image segmentation can be applied to different domains and applications [12]. These applications include: removing and filtering noise, identify objects, face recognition, finger print recognition, and medical image analysis. These different applications have inspired different image segmentation techniques over the past three decades [12]. The choice of the technique depends on the type of image. Indeed, the thresholding techniques that may work for MRI images [34], do not work for histopathological images that contain microscopic structures of human tissue. In the next sections, we elaborate more on the histopathological Breast cancer images.

2.1.2 Histopathological Breast Cancer Images

In this section, we describe *why* and *how* histopathological breast cancer images are acquired. Pathologists would resort to analyze histopathological breast cancer images because they are the most informative images among all other imaging tests [1]. Indeed, these images reveal crucial details about the breast tissue, as well as the cancerous cells. A sample of a histopathological breast cancer image is depicted in Figure 2.1. By

analyzing the structure of the cancerous cells in these images, pathologists can assert the presence and the grade of this disease [17].
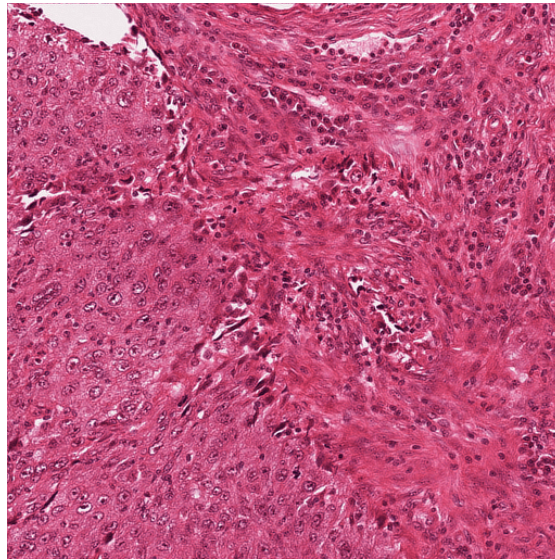
FIGURE 2.1: A sample of a histopathological breast cancer image

However, these histopathological images are very expensive as their preparation is very delicate and time consuming. For this reason, doctors would first perform other cheaper imaging tests such as mammograms. Then, only if they suspect breast cancer disease, they would proceed for the preparation and analysis of the histopathological images [1, 4].

Preparation of the histopathological breast cancer images starts with numbing the breast by a local anesthesia. So, the patient does not feel the pain, when the biopsy is performed. A needle is injected in the breast area and a small sample of the breast tissue is taken as shown in Figure 2.2. Then, the extracted tissue sample is taken to the laboratory in order to prepare the stained slides.
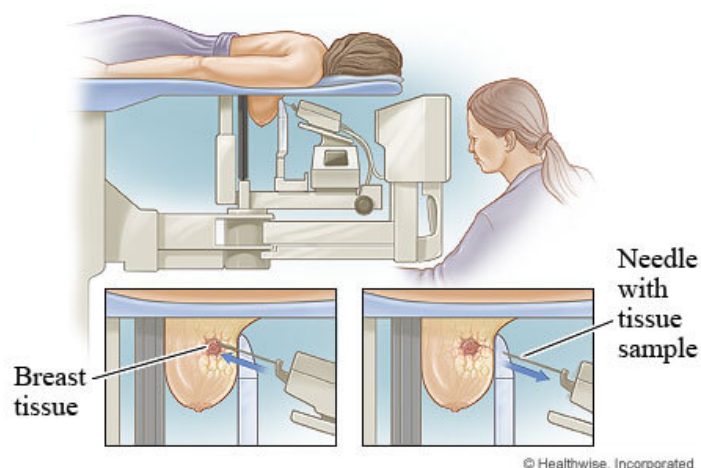
FIGURE 2.2: Stereotactic biopsy of the breast [4]

The preparation of the stained slides consists mainly of four steps [3]. First, the extracted tissue sample is divided into small cuts as shown in Figure 2.3a. Second, these small cuts are put into cassettes and embedded in the paraffin blocks as presented in figures 2.3b and 2.3c. Third, using these paraffin blocks, thin sections are extracted with a microtome as depicted in Figure 2.3d. In the fourth and last step, these thin sections are put onto glass slides to be stained by the Hematoxylin and Eosin *(H&E)* as shown in Figure 2.3e. The Hematoxylin and Eosin stains refer to the standard staining protocol: the Hematoxylin stains the nuclei with a dark purple color, while the Eosin stains other structures, such as cytoplasm and stroma, with a pink color. In Figure 2.3f, we present a sample of a H&E stained slide.

(A)   (B)   (C)

(D)   (E)   (F)
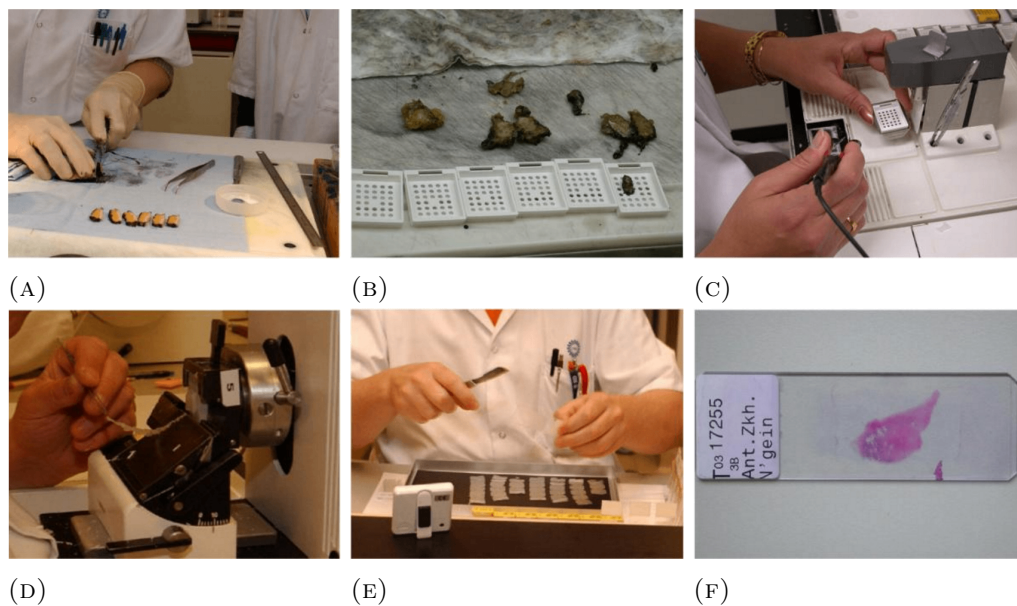
FIGURE 2.3: Preparation of H&E stained slide [3]

After preparing the stained slides, pathologists - experts in tissue analysis - start examining the breast cancer tissue. In the past and before the emergence of digital pathology, pathologists were used to examine these slides under optical microscopes [25]. These optical microscopes were not very practical because the slide tissue can be damaged quickly, and thus pathologists cannot keep record of the examined tissue. However, with the emergence of digital pathology, the old optical microscopes are gradually being replaced by to the Whole Slide Imaging (WSI) scanners [22]. These scanners enable pathologists to view the conventional stained slides on a computer monitor. These scanners do not only generate digital slides, at different resolutions, by scanning the conventional stained slides, but also help pathologists to analyze the images of the digital slides [22]. In Figure 2.4, we show one type of these scanners which is the Aperio scanner.

FIGURE 2.4: The Aperio ScanScope XT scanner [3]

2.1.3 Tumor & Non-tumor Regions

In this section, we elaborate on *how visually* to differentiate between tumor regions and non-tumor regions in the histopathological breast cancer images. The texture of both the tumor and non-tumor regions, as well as the structure of cancer and normal cells are the main criteria that pathologists consider in order to differentiate between both regions [7]. The texture of the tumor region is characterized by an invasive behavior of the cancer cells. These cancer cells have an uncontrollable growth as they begin to divide without stopping. For this reason, most of the time and unlike the normal cells, cancer cells are spread into the surrounding tissue [7].

As well, the cell structure is another criterion that pathologists examine to distinguish between the tumor and non-tumor regions. We illustrate in Table 2.1 the main differences between the structure of the cancer and normal cells [2].

TABLE 2.1: The structure of the normal and cancer cells

Structure	Normal cell	Cancer cell
Cytoplasm	Large	Small
Nucleus	Single	Multiple
Nucleolus	Single	Multiple and Large
Chromatin	Fine	Coarse

In Figure 2.5, we present a segmented histopathological breast cancer image. In this Figure, we highlight both the tumor and non-tumor regions. We can notice, in this Figure as well, a small white area that does not contain any cell. This area is the fat of the breast, and it is considered as non-tumor region.
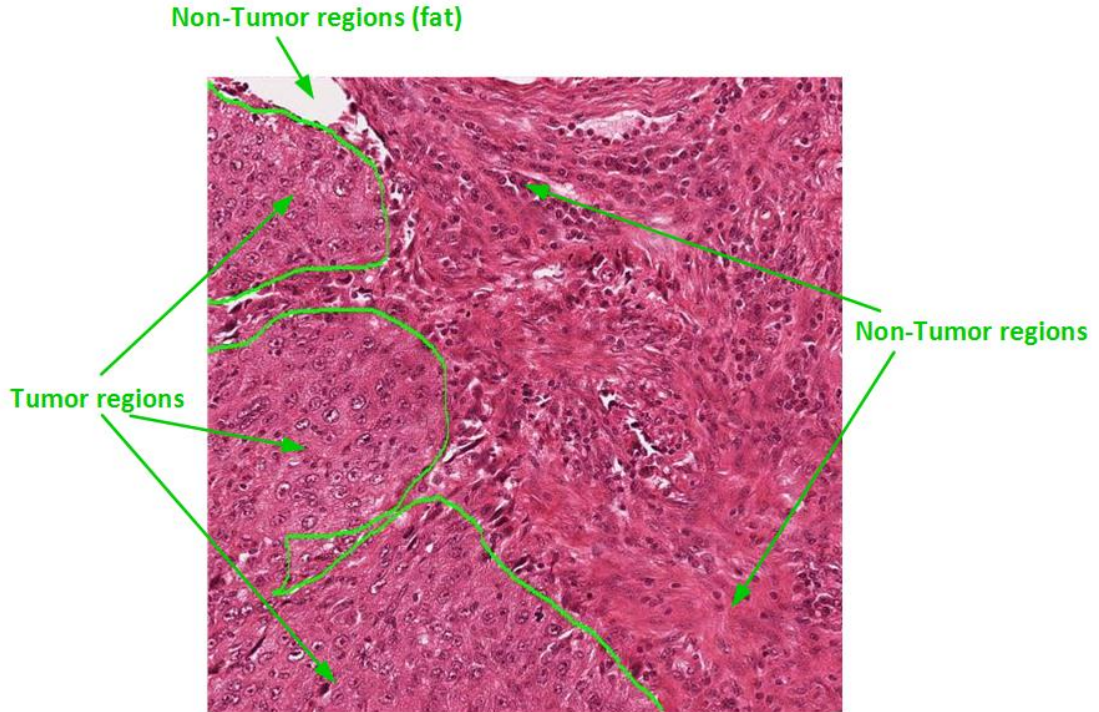
FIGURE 2.5: A sample of a histopathological breast cancer image segmented into tumor & non-tumor regions

2.2 Related Work: Segmentation of Tumor Regions

In this section, we present the related work to the segmentation of tumor regions in microscopic images of breast cancer tissue. In the past decade, researchers have dealt with this task by machine learning techniques [8, 9, 11, 18–20, 26, 27]. They have applied different learning techniques that span over: (1) the supervised learning [8, 9, 11, 26, 27], (2) the unsupervised learning [19, 20], and (3) the semi-supervised learning [18]. We describe some of these techniques in the following sections.

2.2.1 Supervised Learning

The supervised learning, also known as classification, can be defined as any machine learning algorithm that learns from labeled data. The input, or training data, consists of two elements: (1) data and (2) labels. The machine learning algorithm learns from this training data and outputs a trained model. This trained model can be applied to predict the labels of unseen data known also as testing data [31].

In order to address the segmentation of tumor regions in microscopic images of breast cancer tissue as a supervised learning problem, researchers have developed frameworks that consist mainly of the following steps: (1) pre-processing of the images, (2) feature extraction, (3) feature reduction/selection, (4) classification, and occasionally (5) post-processing of the classification results [8, 9, 11, 26, 27].

In the pre-processing step, different techniques were applied. Many contributions converted the RGB color space images to the CMY [11] and CIE Lab [26, 27] color space images. Qu et al. [27] applied contrast stretching to adjust image contrast, as well as smoothing and blurring techniques to remove image noise. After finishing the pre-processing step, the pre-processed images are used in the feature extraction step.

In the feature extraction step, researchers have investigated different types of feature. Many contributions applied textural features: the MR8 filter bank in [11], the Root Filter Bank RFB in [26], and the Haralick filter in [27]. Checkkouri et al. [11] applied topological and morphological features. Topological features consist of network features that connect the cancer cells in order to represent the tumor regions, whereas the morphological features are based on the size and shape of the nuclei of cancer cells. Akbar et al. [8, 9] applied spin intensity features that encode the distribution of intensity values within a circular regions. Qu et al. [27] applied the color intensity features of the CIE Lab color space images. The color intensity features cannot be considered as a robust feature for histopathological images. Indeed, the image color can change as the tissue stain can change from one laboratory to another. Whereas, textural features depend only on the structure and the patterns of the tumor regions. For this reason, we have chosen in this work to investigate mainly textural features such as Gabor, Phase Gradient and MR8Fast.

After finishing the feature extraction step, and if the number of extracted feature is large, feature reduction can be applied in order to decrease the number of features. Checkkouri et al. [11] applied the maximum relevance & minimum redundancy technique in order to select the most relevant features for the classification step.

In the classification step, Akbar et al. [8, 9] applied the Maximum Likelihood Predictor (MLP) classifier. Whereas, many other contributions relied on the Support Vector Machine (SVM) classifiers [11, 26, 27]. As we can notice a few number of classifiers were investigated. We consider in this study other classifiers such as Naive Bayes and Artificial Neural Network.

Occasionally, after finishing the learning step, a post-processing step is needed to improve the segmentation results. Generally, the post-processing step eliminate misclassified segment. Qu et al. [27] applied area threshold to remove the small connect tumor regions within the non-tumor regions. In our study, we apply a similar post-processing step in order to improve the results of the segmentation.

2.2.2 Unsupervised Learning

The unsupervised learning can be defined as any machine learning technique that learns from the data without labels[31]. Typically, there are three approaches of unsupervised learning: (1) clustering, (2) association rules, and (3) self-organizing map. In segmentation of tumor regions in microscopic images of breast cancer tissue, clustering technique has been mainly applied by Khan et al. [19, 20].

Khan et al. [19, 20] have developed the RanPEC and HyMaP frameworks that are composed of the following steps: (1) pre-processing of the images by removing the fat, and converting the RGB color space images to the CIE Lab color space, (2) extracting features from the pre-processed images, (3) reducing the number of features by applying ensemble of random projections, (4) applying K-Means clustering algorithm, and (5) post-processing of the clustered image.

In the RanPEC framework [19], Khan et al. have investigated mainly the feature extraction step. They used different features: Gabor features [14], Phase Gradient features [24], Orientation Pyramid features [35] , and Full Wavelet packet features [28]. The total number of used features is equal to 219 features which is relatively a large number. For this reason, Khan et al. [19] resorted to reduce the number of features by applying the ensemble of random projections technique. The number of reduced features is equal to 20 features.

As a limitation, this framework applied different features without unveiling the most discriminating feature among the four used features: Gabor [14], Phase Gradient [24], Orientation Pyramid [35] , and Full Wavelet packet [28]. For this reason and in a following contribution, Khan et al. [20] have developed the HyMaP framework. The HyMaP framework investigated the previous RanPEC framework while using only the Gabor [14] and Phase Gradient [24]. The investigation of only the Gabor and Phase Gradient has led to the following conclusion: the Gabor features discriminate better the hypo-cellular stroma - normal tissue not rich in cells -, while the Phase Gradient features discriminate better the hyper-cellular stroma - normal tissue rich in cells-.

Although HyMaP has narrowed the use of features to Gabor and Phase Gradient, we still need to know what are the most discriminating features. For this reason, in this study, we investigate each of the Gabor and Phase Gradient features apart. We investigate as well other features such as the MR8Fast features.

2.2.3 Semi-supervised Learning

The semi-supervised learning is a hybrid learning that combines the supervised and unsupervised learning. The semi-supervised learning uses both labeled and unlabeled data to perform the learning process [36].

Karaçali et al. [18] addressed the segmentation of tumor regions in microscopic images of breast cancer tissue as a semi-supervised learning process. For the pre-processing step, they converted the RGB color space image into the CIE lab color space image. As well, they excluded images with extensive fat from the learning process. For the feature extraction step, they applied pixel intensity features of the gray scale image. For the semi-supervised learning, K-Means algorithm is applied to cluster the blocks of the image into 3 clusters. Then, the log-likelihood ratio of these clusters is estimated. Then, the SVM classifier is applied to classify the unknown true likelihood ratio.

2.3 Summary

In this chapter, we presented a background for this study. This background encompasses information about image segmentation, histopathological breast cancer images, and the difference between tumor and non-tumor regions. As well, we presented the related work of the segmentation of tumor regions in microscopic images of breast cancer tissue. Different machine learning techniques were investigated: (1) Supervised learning, (2) Unsupervised learning, and (3) Semi-supervised learning. In this study, we propose different frameworks that are based mainly on supervised and unsupervised learning techniques. We detail these frameworks in the next Chapter.

Chapter 3

The Proposed Frameworks

In this chapter, we present our proposed frameworks that segment tumor regions in microscopic images of H&E stained Breast Cancer tissue. Our frameworks include the following steps: (1) Pre-processing of the images, (2) Feature extraction, (3) Features reduction, (4) The learning process that encompasses the supervised and unsupervised learning, and (5) Post-processing of the results. We illustrate all these steps in Figure 3.1. We describe in the following sections all the steps of the proposed frameworks, as well as our contributions that relate mainly to the learning process.

3.1 Pre-Processing

In this section, we present the pre-processing step that aims to improve the quality of the images, as well as remove unwanted distortions such as noise and stain excess. For this purpose, we apply different pre-processing techniques which are mainly: (1) Stain normalization, (2) Conversion to the gray color space, (3) Noise removal by smoothing. We describe these techniques in the following sections.

3.1.1 Stain Normalization

The stain normalization aims to correct the color inconsistency which results in the overall colors variations of the images. For the histopathological breast cancer images, the color inconsistency is due to many factors such as light, scanner type, and stain variations from one laboratory to another. We illustrate in Figure 3.2 the color variation in the dataset before and after stain normalization.
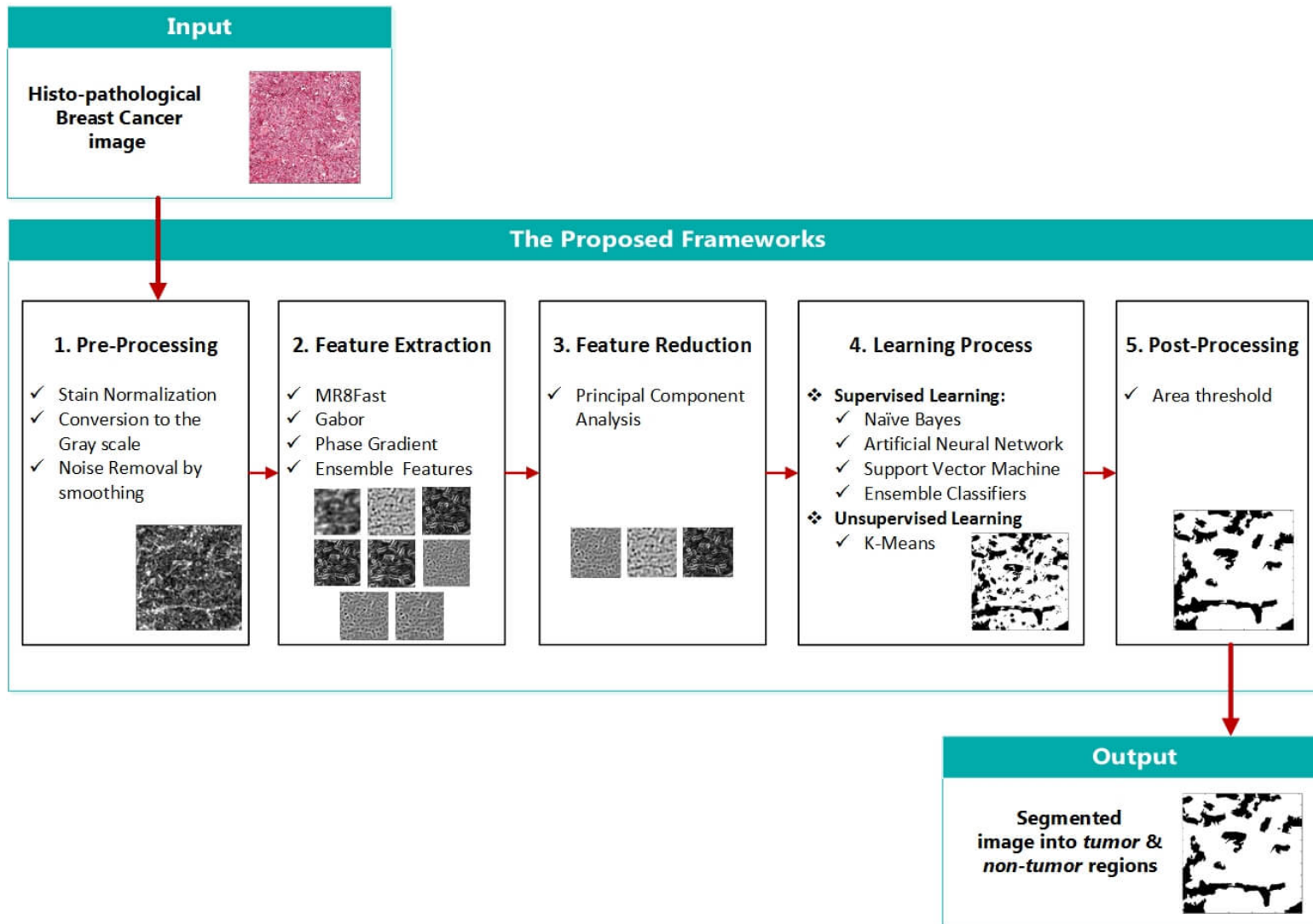
FIGURE 3.1: Block diagram of all the steps followed in the proposed frameworks
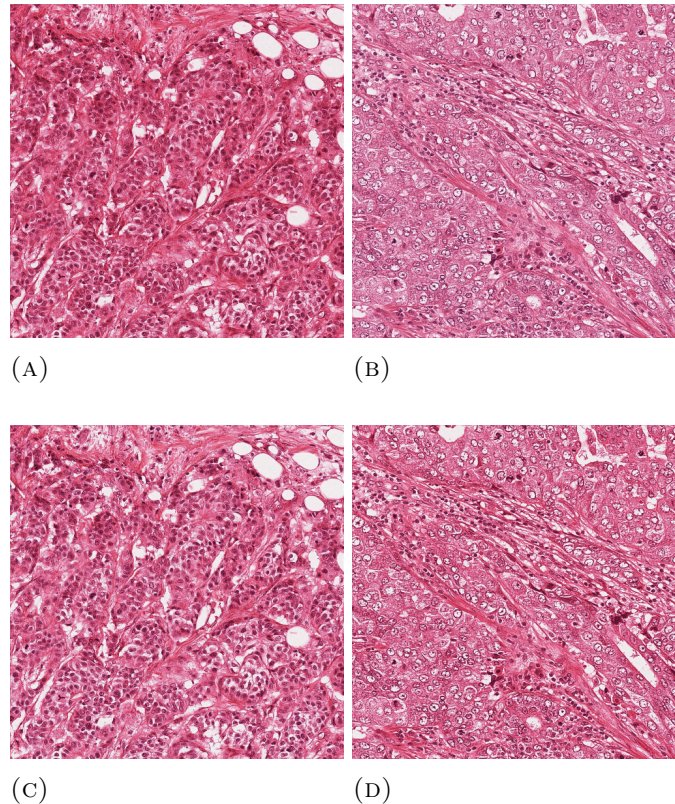
FIGURE 3.2: Illustration of the color variation in microscopic images of the H&E stained breast cancer tissue. Images (A) and (B) present color inconsistency before stain normalization, while images (C) and (D) present better color consistency after stain normalization.

There are several stain normalization techniques that can be found in the literature [21, 29]. We tried three popular techniques: (1) Histogram specification[1] , (2) Reinhard et al. [29], and (3) Macenko et al. [21]. We decided to apply the histogram specification technique because we observed that it enforces better the difference between the tumor and the non-tumor textures. We show in Figure 3.2c and 3.2d the images after stain normalization.

3.1.2 Conversion to the Gray Scale

In this step, we converted the microscopic images of H&E stained breast cancer tissue from the RGB to the CIE lab color space [2]. Then, we extract only the l channel space which is a gray scale version of these images that will be employed in all the following steps.

3.1.3 Noise Removal: Smoothing

Noise, in image processing, can be defined as the variation of brightness that does not match the real variation. Noise may arise during the image acquisition, such as sensor

---

[1]http://www2.warwick.ac.uk/fac/sci/dcs/research/combi/research/bic/software/sntoolbox/
[2]http://www.mathworks.com/help/images/ref/rgb2lab.html

noise, due to the poor illumination. In order to check noise, we selected randomly one image from the dataset, and we plotted its histogram as shown in Figure 3.3.
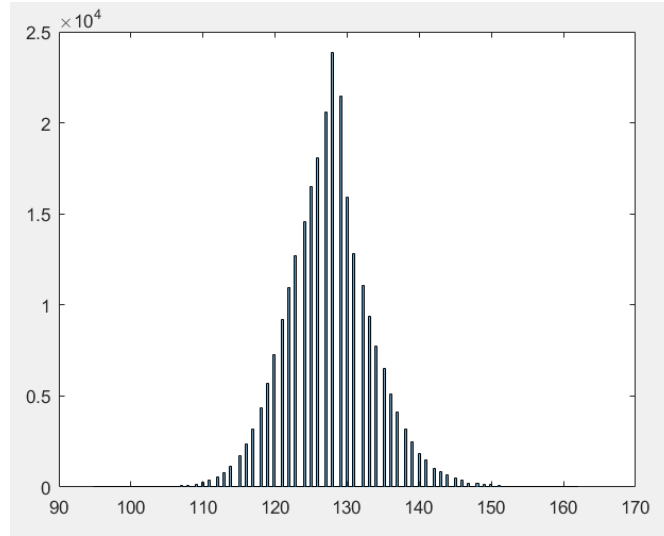
FIGURE 3.3: Histogram of a randomly selected image from the dataset

As shown in Figure 3.3, the histogram presents a normal distribution that can be correlated with a Gaussian filter. In this context, we employed a smoothing Gaussian filter to remove the Gaussian noise.

We present in Figure 3.4 a sample of a microscopic image of breast cancer tissue along with the corresponding pre-processed image. The pre-processed image will be used for the feature extraction step in order to extract textural features as detailed in the next section.
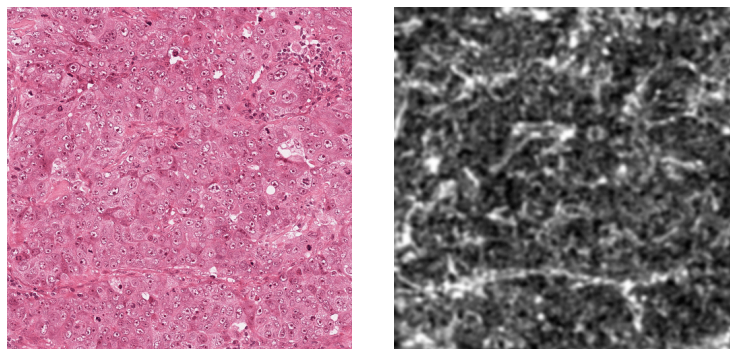
(A) A sample of microscopic image of breast cancer tissue

(B) The corresponding pre-processed image

FIGURE 3.4: A sample of microscopic image of breast cancer tissue and the corresponding pre-processed image that will be used for the feature extraction step

3.2 Feature Extraction: Textural Features

After pre-processing all the 50 images in the dataset, we proceed with extracting relevant features. In this context, features or attributes are specific properties or characteristics

extracted from the images in order to learn and recognize the tumor and non-tumor textures. In this study, we focus only on textural features. Textural features materialize the differences between the tumor and non-tumor textures. Our main reason for choosing textural features is their independence to the variation of pixels' intensities. We develop different frameworks to investigate each of the following textural features: MR8Fast [5], Gabor [14] and Phase Gradient [24], as well as a combination of all these features denoted as *Ensemble Features*. Our aim is to reveal the most discriminating feature among all these features. We describe all the investigated features in the following sections.

3.2.1 MR8Fast Features

MR8Fast features are extracted from the Maximum Response 8 (MR8) filter bank [5]. The Maximum Response (MR) filter bank is derived from a common Root Filter Set (RFS) bank that consists of 38 filters[3]. The RFS employs a Gaussian and a Laplacian of Gaussian filters with $sigma = 10$ pixels, an edge filter at 3 scales, and a bar filter at the same 3 scales. Both Gaussian and Laplacian are oriented filters that occur at 6 orientation at each scale. As MR8 measures the MR across orientations, the number of responses is reduced from 38 to 8; and this is why the MR8 filter consists of 38 filters with only 8 responses. That means the number of features extracted from the MR8Fast filters is 8 features for every single image in the dataset.

We present in Figure 3.5 8 filter responses of the MR8Fast filter bank of the image presented in Figure 3.4a. These generated responses are employed as textural features for the learning process steps.

(A)   (B)   (C)   (D)   (E)   (F)   (G)   (H)
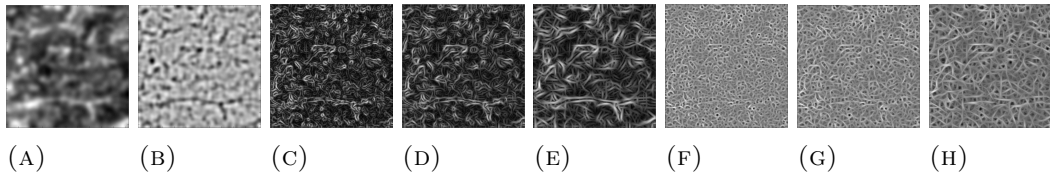
FIGURE 3.5: Sample of the 8 filter responses of the MR8Fast filter bank

3.2.2 Gabor Features

The Gabor features [14] are extracted from the Gabor filter bank [4]. The Gabor filter bank was firstly proposed by Dennis Gabor for edge detection. It is a linear filter that act similarly to the human visual system. It occurs at different orientations and scales to generate filter responses that are employed as features. In this study, we generated 84 Gabor filter responses which means 84 Gabor features for every single image in the dataset. As the number of Gabor features is relatively big, we reduced the 84 Gabor

---

[3]http://www.robots.ox.ac.uk/ṽgg/research/texclass/code/makeRFSfilters.m
[4]http://www.cs.rug.nl/ imaging/simplecell.html

features to only 15 features using the Principal Component Analysis (PCA) technique [33] described in Section 3.3.

We present in Figure 3.6 a sample of only 6 filter responses for Gabor filter bank out of a total of 84 filter responses. These filter responses correspond to the image presented in Figure 3.4a. The 84 Gabor filter responses are employed as textural features in the learning process steps.

(A)   (B)   (C)   (D)   (E)   (F)
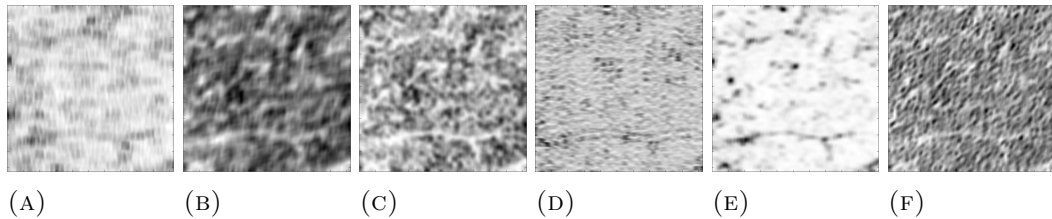
FIGURE 3.6: Sample of only 6 filter responses of the Gabor filter bank out of a total of 84 filter responses

3.2.3   Phase Gradient Features

The Phase Gradient features are extracted from the phase gradient filters [24]. The phase gradient filters employ the log-Gabor filters in order to decompose the image into the scales and orientations components. For each of these components, the local phase gradient is computed to give local frequency in a perpendicular direction to the radial direction of the log-Gabor domain. In this study, we extracted 48 Phase Gradient response filters which correspond to 48 features for every image in the dataset. As the number of features is relatively large, we reduce the 48 features to 10 using the PCA technique described in Section 3.3.

We present in Figure 3.7 a sample of only 5 filter responses for Phase Gradient filter bank out of a total of 48 filter responses. These filter responses correspond to the image presented in Figure 3.4a. The 48 Phase Gradient filter responses are employed as textural features in the learning process steps.

(A)   (B)   (C)   (D)   (E)
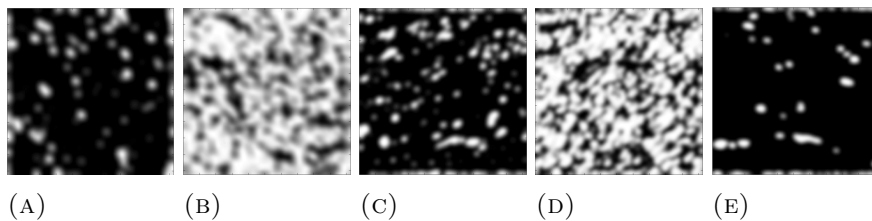
FIGURE 3.7: Sample of only 5 filter responses of the Phase Gradient filter bank out of a total of 48 filter responses

3.2.4 Ensemble Features

We combine all the previously detailed features MR8Fast, Gabor, and Phase Gradient in order to generate Ensemble Features. The Ensemble Features, referred as **Ens.F**, encompasses a total of 140 features (8 for MR8Fast +84 for Gabor +48 for Phase Gradient). A 140 features is relatively a big number that may include redundant and outlier features. In this context, we reduced the 140 features to only 20 features using the Principal Component Analysis technique [33] detailed in the following section.

3.3 Feature Reduction

In this section, we present the Principal Component Analysis (PCA) [33] that we used for feature reduction. The PCA technique is able to identify patterns in data presenting a large number of features, and point out their similarities and differences. This technique also provides simple steps that reduce the number of features without much loss of information. These steps are as follows: (1) subtract the mean from each features, (2) calculate the covariance matrix, (3) calculate the eigenvectors and eigenvalue of the covariance matrix, and (4) choosing components and forming a feature vector to generate the new dataset.

We present in Table 3.1 the number of each MR8Fast, Gabor, Phase Gradient, and Ensemble features before and after using the PCA reduction technique.

TABLE 3.1: Number of features before and after using the PCA reduction technique

Feature	Before reduction	After reduction
MR8Fast	8	8
Gabor	84	15
Phase Gradient	48	10
Ensemble Features	140	20

3.4 Learning Process

In this section, we present the learning process that segments histopathological breast cancer images into tumor and non-tumor regions. This process consists of a set of techniques that map the features generated in the previous sections to two different class labels. These two class labels correspond to the tumor and non-tumor regions. In this study, we aim to learn from every feature described in the previous sections using both supervised and unsupervised learning techniques. We detail these techniques in the next sections.

3.4.1 Supervised Learning

The supervised learning, known also as classification, can be defined as any machine learning algorithm that learns from labeled data. The input is the training data that consists of a set of histopathological breast cancer images associated with their ground truths. The output is a trained binary classification model that aims to segment new breast cancer images into tumor and non-tumor regions.

In this study, we endeavor to learn from every feature introduced in Section 3.2 using the following classifiers. We investigate the following classifiers: Naive Bayes (NB), Artificial Neural Network (ANN), and Ensemble Classifiers in order to segment histopathological breast cancer images. As well, we employ the Support Vector Machine (SVM) classifier that has been already used in the literature [8, 9, 11, 26, 27]. In the following sections, we illustrate briefly these classifiers.

Naive Bayes (NB)

The Naive Bayes (NB) classifier [23] is a probabilistic classifier based on the Bayes' theorem or the so-called Bayes' rule. The Bayes' theorem aims to compute the Posterior probability as described in Equation 3.1, where k is the class index, n is the number of features, and $\{x_1, ..., x_n\}$ is the set of features and known as data. This classifier is called *naive* because it assumes that all the features are conditionally independent given a class label. This assumption is referred as the *naive* assumption [23].

$$\underbrace{P(C_k|x_1, ..., x_n)}_{Posterior} = \frac{\overbrace{P(C_k)}^{ClassPrior} \overbrace{P(x_1, ..., x_n|C_k)}^{Likelihood}}{\underbrace{P(x_1, ..., x_n)}_{PredictorPrior}} \tag{3.1}$$

The Posterior $P(C_k|x_1, ..., x_n)$ is the probability of the Class C_k given the data $\{x_1, ..., x_n\}$. The Posterior probability is predicted from: (1) the Class Prior probability $P(C_k)$ which is the probability of the class before seeing the data, (2) the Likelihood $P(x_1, ..., x_n|C_k)$ which is the probability of the data given the class, and (3) the Predictor Prior $P(C_k)$ probability which is the probability of the data.

In the training phase, the class prior probability and the features' frequency tables are computed for every single class. In the testing phase, the features' frequency tables are used to compute the Likelihood and the Predictor Prior probabilities. The Likelihood and Predictor Prior probabilities are used with the Class Prior probability to compute the Posterior probability for every single class as shown in Equation 3.1. The class with the highest Posterior probability is the outcome of the NB classifier.

Artificial Neural Network (ANN)

The Artificial Neural Network (ANN) classifier is inspired from the biological neural networks of the human brain. In the human brain, the brain cells or neurons are connected to each other to have a pathway that transmits electrical signal. Similarly, the Artificial Neural Network consists of a set of neurons that are connected to each other. These neurons are organized into three layers: (1) *Input*, (2) *Hidden*, and (3) *Output*. The structure of the ANN classifier is presented in Figure 3.8.
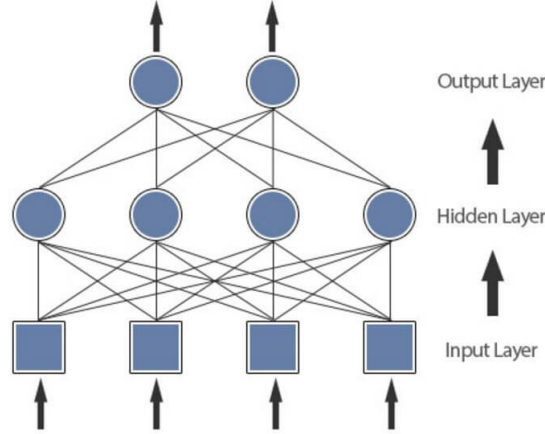
FIGURE 3.8: The structure of the Artificial Neural Network classifier

The neurons of the *Input* layer are rather the features' values that will be fed to the next *Hidden* layer. The Hidden layer can include more than one layer. Finally, the *Output* layer is composed of neurons equal to the number of predicted classes.

In the training phase, the neurons are initialized with random weights. Then, the network processes the data and compares the actual labels against predicted ones. The prediction errors are propagated back through the network, and the weights are adjusted. In the testing phase, the adjusted weights are used to classify the new data. It should be noted that in our framework we used an ANN of 4 layers.

Support Vector Machine (SVM)

In our framework, we used the Support Vector Machine (SVM) classifier that represents the features as points in the space. The SVM classifier aims to find the optimal hyperplane that separates the data features into separate classes. The hyperplane turns into a line or plane when the number of features is equal to 2 or 3 respectively. We describe the form of the linear SVM classifier in Equation 3.2.

$$f(\mathbf{x}_i) = \mathbf{w}^T x_i + b \tag{3.2}$$

Where $\mathbf{x}_i$ is the set of features with: $i \in \{1, ..., n\}$, and n is the number of features. The $\mathbf{w}$ is known by the weight vector, while the b is the bias.

In the training phase, the SVM classifier iterate over the data till the weight vector $\mathbf{w}$ is adjusted to better separate the data into the different classes. In the testing phase, the weight vector $\mathbf{w}$ is used to classify the data.

Ensemble Classifiers

We combine the segmentation results of the Naive Bayes, Artificial Neural Network, and Support Vector Machine classifiers in order to generate the Ensemble Classifiers results. The combination of the results is based on the majority voting which means every pixel in the image is classified as tumor only if at least two classifiers have classified that pixel as tumor, and vice versa. In this study, the Ensemble Classifiers is referred as **Ens.Clf**.

3.4.2 Unsupervised Learning

In addition to the supervised learning, we investigate as well the unsupervised learning. The unsupervised learning can be defined as any machine learning algorithm that learns from non-labeled data. In this study, we employ the k-Means clustering algorithm that has been already used in [19, 20] to segment histopathological breast cancer images into tumor and non-tumor regions.

The k-Means clustering algorithm takes as Input: (1) $k = 2$ the number of clusters that represents the tumor and non-tumor clusters, and (2) the set of n features where $n > k$. The Output is the set of the n features grouped into two separate clusters. We present the objective function of the K-Means algorithm in Equation 3.3.

$$J = \sum_{j=1}^{k} \sum_{i=1}^{N_j} \left\| \mathbf{x}_i^{(j)} - \mathbf{c}_j \right\|^2 \tag{3.3}$$

The K-Means algorithm aims to minimize the objective function - in this case a squared function - as follow: (1) every centroid or the cluster center is initialized randomly, and (2) the Euclidean distance between every single centroid $\mathbf{c}_j$ and each feature $\mathbf{x}_i$ is computed, where $i \in \{1, ..., N_j\}$ with N_j is the number of features associated to the cluster j, and $j \in \{1, ..., k\}$. Then (3) each feature vector is assigned to the closed centroid that has the minimal Euclidean distance from the feature vector. After that (4) the value of every centroid is computed again by averaging all the features $\mathbf{x}_i$ that assigned to that centroid. Finally (5) this process - from step (2) - is repeated again and again till the centroids' values stabilize.

The major problem with the k-Means algorithm is when the centroids are initialized with bad values. In this case, they may get stuck at local optima. Other than that, the k-Means algorithm and the unsupervised learning in general are considered relatively fast approaches for segmentation especially there is no need for the training step.

3.5 Post-processing

The post-processing step aims to improve the segmentation results of the learning step. We show in Figure 3.9b a sample of segmentation results before the post-processing step. We can see in this figure very small tumor areas within very large non-tumor areas, and vice-versa. These small areas are likely misclassified areas because classifying very small areas as tumor inside a very large connected non-tumor area, and vice-versa, would not be justified [27]. For this reason, we apply area threshold in order to remove the very small tumor areas within the large non-tumor regions, and vice-versa. We show a sample of the segmentation results after the post-processing step in Figure 3.9c.
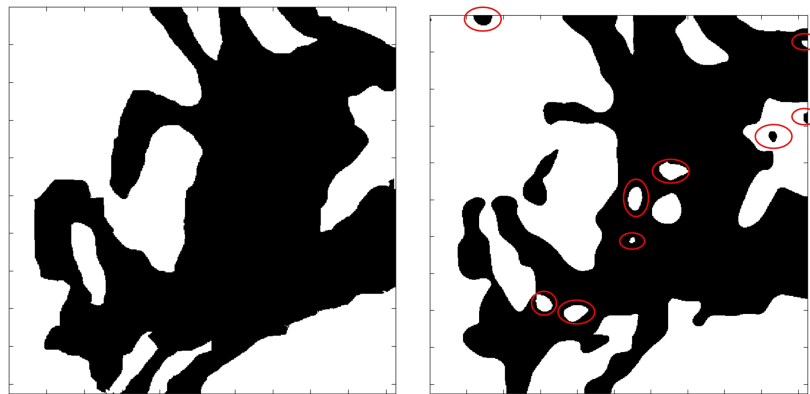
(A) A sample of a ground truth image

(B) Segmentation results before the post-processing step
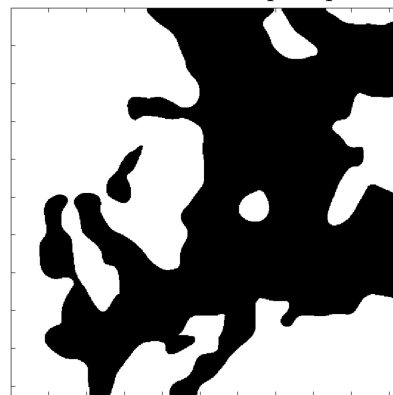
(C) Segmentation results after the post-processing step

FIGURE 3.9: Illustration of the post-processing step: area threshold. The white labels correspond to the tumor regions, while the black labels are assigned to the non-tumor regions.

As shown in Figure 3.9, the segmentation results after the post-processing has taken out the very small misclassified regions. The post-processed results are closer to the fused ground depicted in Figure 3.9a.

3.6   Summary

In this Chapter, we described the proposed frameworks that include the following steps: (1) pre-processing, (2) feature extraction, (3) feature reduction using PCA, (4) supervised and unsupervised learning, and (5) the post-processing. In this framework, our contributions are mainly in the supervised learning step. In this step, we used for the first time the Naive Bayes and the Artificial Neural Network classifiers, as well as Ensemble Classifiers in order to segment tumor regions in microscopic images of the H&E stained breast cancer tissue. Furthermore, our proposed frameworks learn separately from each of the MR8Fast, Gabor, and Phase Gradient features in order to disclose the most discriminating features. Every feature is employed with different supervised and unsupervised techniques to reveal as well the best combination of *features/learning techniques* that better segments tumor regions. We introduce the experimental results of the proposed frameworks in the next Chapter.

Chapter 4

Experimental Evaluation

In this chapter, we present in the first section the evaluation setup of this study. In the second section, we elaborate on the baseline system that we used as a reference for this work. In sections 4.3 and 4.4, we detail the different variants of our unsupervised and supervised frameworks, as well as compare these proposed frameworks with the baseline system. In the last section 4.5, we discuss and summarize the findings of this study.

4.1 Setup

In this section, we give an overview of the evaluation setup of this study. First, we describe on the dataset used in this work. Second, we present the evaluation metrics that we used to assess our frameworks.

4.1.1 Public Dataset: MITOS 2012

We used the MITOS 2012 dataset[1] that was prepared by the team of professor Frédérique Capron at the pathology department at Pitié-Salpêtrière Hospital in Paris, France. This dataset is publicly available, and was mainly designed for the MITOS 2012 competition. Participants in this competition were asked to develop robust algorithms to automatically detect and grade the mitotic cells. However, in this study, we employed the MITOS 2012 dataset to develop our frameworks for tumor segmentation.

This dataset encompasses microscopic images of Hematoxylin and Eosin (H & E) stained breast cancer tissue. These images were scanned at ×40 magnification by three different scanners: Aperio ScanScope XT slide scanner, Hamamatsu NanoZoomer 2.0-HT slide scanner, and 10 band multispectral microscope. Every image in this dataset has relatively small size of 512×512 pixels [30]. This dataset has a total of 50 H & E stained images. These images were taken from 5 different patients. Every patient contributed with 10 different images. We present an image sample from this dataset in Figure 4.1a .

---

[1]http://ipal.cnrs.fr/ICPR2012
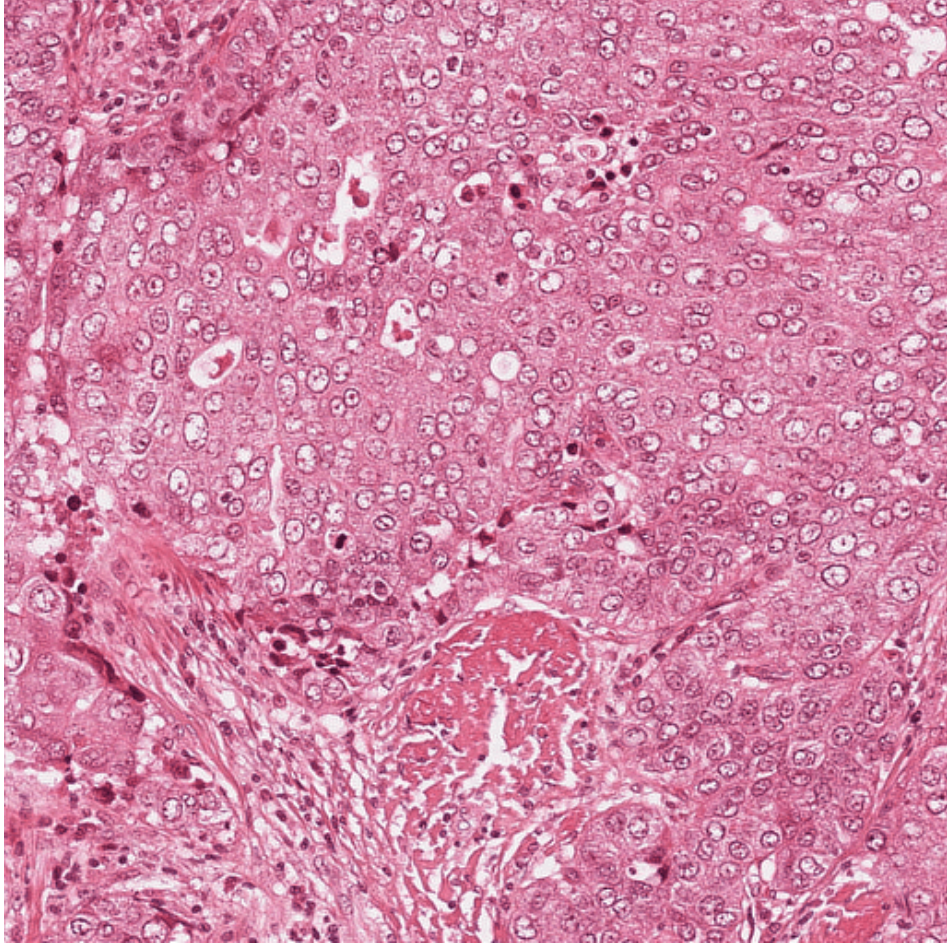
(A) A microscopic image of H&E stained breast cancer tissue

(B) First ground truth
- by Dr. Aisha -

(C) Second ground truth
- by Dr. Hesham -

(D) Third ground truth
- **Fused ground truth** -

FIGURE 4.1: A microscopic image of H&E stained breast cancer tissue and the corresponding ground truth data, the white labels are the tumor regions, while the black labels are the non-tumor regions

Ground Truth

In order to evaluate the results of our frameworks, we used three different sets of ground truth (GT) data. Two ground truth sets were manually and separately marked by two experienced pathologists Dr. Asha Rupani and Dr. Hesham El-Daly at Addenbrookes

Hospital, Cambridge, UK. Whereas, the third GT set is automatically generated from the two manually provided GT sets. This third GT, named *fused* GT, assigns tumor labels to the regions that were labeled as tumor by both pathologists and non-tumor labels to any other regions. We used the *fused* GT set in most of the evaluation results because this ground truth is considered more reliable, as well as the assessment and the comparison between the different framework variants become easier with one single GT set.

In Figure 4.1, we present an H & E stained Breast Cancer image, and the three associated GT images. Figure 4.1b and 4.1c illustrate the GT provided by the two experienced pathologists, whereas Figure 4.1d shows the corresponding *fused* GT.

4.1.2   Evaluation metrics

We use the following metrics to evaluate the results of our proposed frameworks: Accuracy, Precision, Recall, and F1-Score. We detail these evaluation metrics in the following sections. We present in Table 4.1 some useful notions for the next sections.

TABLE 4.1: A brief description of the terms True Positive, True Negative, False Negative, and False Positive.

Shortened Form	Description
True Positive (TP)	The set of pixels that belong to the tumor regions, and are correctly labeled as tumor pixels.
False Positive (FN)	The set of pixels that belong to the tumor regions, but are labeled as non-tumor pixels.
True Negative (TN)	The set of pixels that belong to the non-tumor regions, and are correctly labeled as non-tumor pixels.
False Positive (FP)	The set of pixels that belong to the non-tumor regions, but are labeled as tumor pixels.
n	Number of pixels in the testing data

Accuracy

Accuracy is defined as the proportion of the total number of pixels that are correctly classified over the total number of pixels in the entire dataset. A high accuracy implies that the majority of the pixels are classified correctly. In order to measure the accuracy, we use the Equation 4.1.

$$Accuracy = \frac{n_{TP} + n_{TN}}{n} \tag{4.1}$$

Where n_{TP} and n_{TN} denote the number of pixels that are TP and TN, respectively.

Precision

The precision is the proportion of the number of tumor pixels that are correctly classified over the total number of tumor pixels that are correctly and incorrectly classified as tumor regions. A higher precision means that the majority of the pixels, that are classified as tumor, are actually tumor pixels. In order to compute the precision, we use the Equation 4.2.

$$Precision = \frac{n_{TP}}{n_{TP} + n_{FP}} \tag{4.2}$$

Where n_{TP} and n_{FP} denote the number of pixels that are TP and FP respectively.

Recall

The recall, known also as sensitivity, is the proportion of the tumor pixels that are correctly classified over the total number of the actual tumor pixels. A higher recall means that the framework classified correctly the majority of the actual tumor pixels. In order to measure the recall, we used the Equation in 4.3.

$$Recall = \frac{n_{TP}}{n_{TP} + n_{FN}} \tag{4.3}$$

Where n_{TP} and n_{FN} denote the number of pixels that are TP and FN respectively.

F1-Score

The F1-Score combines the precision and recall metrics in one single number. It is known as well as the harmonic mean of the precision and recall. In order to compute the F1-Score, we use the Equation 4.4.

$$F1 - Score = 2.\frac{Precision.Recall}{Precision + Recall} \tag{4.4}$$

4.2 Baseline: The RanPEC Framework

As we detailed in Chapter 2, RanPEC [19] is an unsupervised learning framework that employs mainly the K-Means algorithm, and four textural features - Gabor [14], Phase Gradient [24], Orientation Pyramid [35] , and Full Wavelet packet [28] - in order to segment tumor regions in microscopic images of breast cancer tissue.

As described earlier in the section 4.1.1, our dataset encompasses images of five different patients, and three different sets of ground truth data. In this context, we report the evaluation results of the baseline framework for every single patient as well as for all the five patients [2]. In addition, we report the evaluation results against the three sets of ground truth. We present the evaluation results of the baseline framework in Table 4.2.

TABLE 4.2: The evaluation results of the baseline - The RanPEC Framework [19] -. Best results are boldfaced

Patient	Accuracy	Precision	Recall	F1-Score
Patient 1	0.83	0.85	0.84	0.85
Patient 2	0.70	0.93	0.70	0.80
Patient 3	0.71	0.96	0.71	0.82
Patient 4	0.83	0.85	0.88	0.86
Patient 5	0.79	0.92	0.81	0.86
All patients	0.77	0.91	0.78	0.84

(A) The evaluation results of the baseline against GT -1-

Patient	Accuracy	Precision	Recall	F1-Score
Patient 1	0.82	0.87	0.81	0.84
Patient 2	0.71	0.92	0.72	0.81
Patient 3	0.71	0.88	0.74	0.80
Patient 4	0.83	0.85	0.88	0.86
Patient 5	0.79	0.92	0.81	0.86
All patients	0.77	0.89	0.79	0.83

(B) The evaluation results of the baseline against GT -2-

Patient	Accuracy	Precision	Recall	F1-Score
Patient 1	0.83	0.83	0.86	0.84
Patient 2	0.72	0.91	0.73	0.81
Patient 3	0.71	0.88	0.74	0.80
Patient 4	0.83	0.83	0.91	0.87
Patient 5	0.80	0.91	0.83	0.86
All patients	**0.78**	**0.87**	**0.80**	**0.84**

Overall, the baseline system achieved relatively good results. For all patients, the accuracy is equal to 78%, and the F1-Score is equal to 84% as shown in Table 4.2. However, if we examine the evaluation results of every patient individually, we can notice that patients 2 and 3 have relatively low accuracy and recall values compared to the other patients. Indeed, the accuracy for patients 2 and 3 does not exceed 72% compared to more than 80% for the other patients. Figure 4.2 shows better the accuracy variation for all the patient in the baseline system, and confirm that patients 2 and 3 present a failure

---

[2]The evaluation results for all the patients is different from the results of the average of the five patients.

point of the baseline framework. In the following sections, we present our supervised frameworks that relatively overcome this issue.
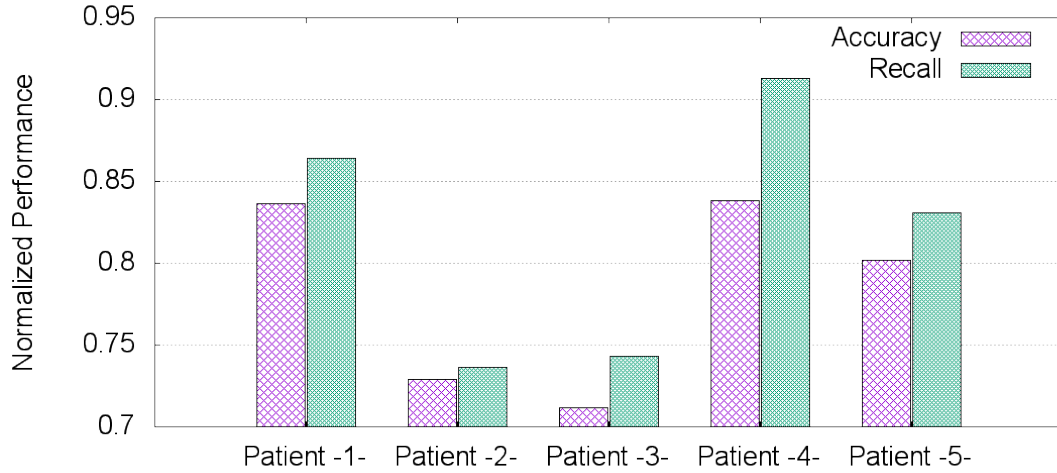
FIGURE 4.2: Accuracy and Recall values for the baseline - The RanPEC framework [19] -

4.3 The Unsupervised Frameworks

In this section, we present the evaluation results of our unsupervised frameworks. We investigate each of the following textural features individually: MR8Fast [5], Gabor [14], and Phase Gradient [24], as well as the combination of all these textural features. As well, we investigate the K-Means algorithm for the unsupervised learning process. We present and evaluate the results of every framework against the three different sets of ground truth. Lastly, we summarize, discuss, and compare the results of the proposed unsupervised frameworks with the baseline system.

4.3.1 Evaluating the MR8Fast Features

In this section, we present the evaluation results of the MR8Fast [5] features. We report the Accuracy, Recall, Precision and F1-Score for every patient and for all the five patients. As well, we report the results against the three different sets of ground truth data. The evaluation results of the MR8Fast features are depicted in the Table 4.3.

As shown in Table 4.3, the Accuracy and F1-Score are respectively equal to 76% and 82% for all the patients. These numbers are comparable to the baseline Accuracy and F1-Score which are equal to 78% and F1-Score 84%. Similarly to the baseline framework, the unsupervised MR8Fast framework has relatively low values of the Accuracy and Recall for the patients 2 and 3. The Accuracy and Recall values for these patients do not exceed the 69%.

TABLE 4.3: The evaluation results of the MR8Fast features. Best results are boldfaced.

Patient	Accuracy	Precision	Recall	F1-Score
Patient 1	0.87	0.89	0.87	0.88
Patient 2	0.61	0.91	0.61	0.73
Patient 3	0.67	0.96	0.67	0.79
Patient 4	0.82	0.86	0.85	0.86
Patient 5	0.80	0.92	0.82	0.87
All patients	0.75	0.91	0.75	0.82

(A) The evaluation results of the MR8Fast features against GT -1-

Patient	Accuracy	Precision	Recall	F1-Score
Patient 1	0.85	0.90	0.83	0.87
Patient 2	0.63	0.90	0.63	0.74
Patient 3	0.66	0.88	0.68	0.77
Patient 4	0.82	0.86	0.86	0.86
Patient 5	0.80	0.92	0.82	0.87
All patients	0.75	0.89	0.76	0.82

(B) The evaluation results of the MR8Fast features against GT -2-

Patient	Accuracy	Precision	Recall	F1-Score
Patient 1	0.87	0.86	0.89	0.88
Patient 2	0.63	0.88	0.63	0.74
Patient 3	0.67	0.88	0.69	0.77
Patient 4	0.83	0.84	0.88	0.86
Patient 5	0.81	0.91	0.84	0.87
All patients	**0.76**	**0.87**	**0.77**	**0.82**

(C) The evaluation results of the MR8Fast features against the fused GT

4.3.2 Evaluating the Gabor Features

In this section, we report the evaluation results of the Gabor [14] features. We present the Accuracy, Recall, Precision and F1-Score for every patient and for all the five patients. As well, we extend the results against the three different sets of ground truth data. The evaluation results of the Gabor features are introduced in the Table 4.4.

As we can notice from Table 4.4, the unsupervised Gabor framework has values of Accuracy and F1-Score respectively equal to 70% and 76% for all the patients. These number are lower than the baseline Accuracy and F1-Score which are equal to 78% and 84% respectively. Analogous to the baseline, the unsupervised Gabor framework has low values of the Accuracy and Recall for the patients 2 and 3. The Accuracy and Recall values for these patients do not exceed the 56%.

TABLE 4.4: The evaluation results of the Gabor features. Best Results are boldfaced

Patient	Accuracy	Precision	Recall	F1-Score
Patient 1	0.86	0.92	0.81	0.86
Patient 2	0.53	0.93	0.49	0.64
Patient 3	0.55	0.97	0.53	0.69
Patient 4	0.78	0.86	0.77	0.81
Patient 5	0.71	0.93	0.70	0.80
All patients	0.68	0.92	0.64	0.75

(A) The evaluation results of the Gabor features against GT -1-

Patient	Accuracy	Precision	Recall	F1-Score
Patient 1	0.83	0.93	0.77	0.84
Patient 2	0.55	0.92	0.50	0.65
Patient 3	0.58	0.89	0.55	0.68
Patient 4	0.78	0.86	0.77	0.82
Patient 5	0.71	0.92	0.70	0.79
All patients	0.69	0.91	0.65	0.75

(B) The evaluation results of the Gabor features against GT -2-

Patient	Accuracy	Precision	Recall	F1-Score
Patient 1	0.86	0.90	0.83	0.83
Patient 2	0.56	0.90	0.51	0.65
Patient 3	0.58	0.89	0.56	0.68
Patient 4	0.79	0.84	0.80	0.82
Patient 5	0.72	0.92	0.71	0.80
All patients	**0.70**	**0.89**	**0.66**	**0.76**

(C) The evaluation results of the Gabor features against the fused GT

4.3.3 Evaluating the Phase Gradient Features

In this section, we introduce the evaluation results of the Phase Gradient features [24]. We present the Accuracy, Recall, Precision and F1-Score for every patient as well as for all the five patients. In addition, we report the results against the three different sets of ground truth data. The evaluation results of the unsupervised Phase Gradient framework are introduced in Table 4.5.

As presented in Table 4.5, the unsupervised Phase Gradient framework has the lowest values compared to the unsupervised baseline, MR8Fast and Gabor frameworks. The Accuracy and F1-Score of the unsupervised Phase Gradient framework do not exceed respectively 66% and 71% for all the five patients compared to more than 70% and 76% for the other framework. Similarly to the baseline framework, the unsupervised Phase

TABLE 4.5: The evaluation results of the Gradient features. Best Results are boldfaced.

Patient	Accuracy	Precision	Recall	F1-Score
Patient 1	0.76	0.90	0.65	0.76
Patient 2	0.50	0.88	0.49	0.63
Patient 3	0.50	0.96	0.48	0.64
Patient 4	0.75	0.86	0.71	0.78
Patient 5	0.67	0.93	0.64	0.76
All patients	0.64	0.91	0.58	0.71

(A) The evaluation results of the Gradient features against GT -1-

Patient	Accuracy	Precision	Recall	F1-Score
Patient 1	0.74	0.91	0.62	0.74
Patient 2	0.52	0.86	0.50	0.63
Patient 3	0.52	0.87	0.49	0.62
Patient 4	0.75	0.86	0.72	0.78
Patient 5	0.67	0.93	0.64	0.76
All patients	0.64	0.89	0.59	0.71

(B) The evaluation results of the Gradient features against GT -2-

Patient	Accuracy	Precision	Recall	F1-Score
Patient 1	0.78	0.88	0.68	0.76
Patient 2	0.52	0.84	0.50	0.63
Patient 3	0.52	0.87	0.49	0.62
Patient 4	0.77	0.85	0.74	0.79
Patient 5	0.69	0.92	0.66	0.77
All patients	0.66	0.87	0.60	0.71

(C) The evaluation results of the Gradient features against the fused GT

Gradient framework has lower values of the Accuracy and Recall for the patients 2 and 3. The Accuracy and Recall values for these patients does not exceed the 52%.

4.3.4 Evaluating the Ensemble Features

In this section, we introduce the evaluation results of the combination of all the investigated features: MR8Fast [5], Gabor [19], and Phase Gradient [24]. We present the Accuracy, Recall, Precision and F1-Score for every patient and for all the five patients. As well, we report the results against the three different sets of ground truth data. The evaluation results of this framework are introduced in the Table 4.6.

As depicted in Table 4.6, the accuracy and F1-Score of the combination of all the features do not exceed 72% and 78% respectively for all the patients. These number are lower than the baseline and the unsupervised MR8Fast framework that have an accuracy and

TABLE 4.6: The evaluation results of combining MR8Fast, Gabor, and Phase Gradient features. Best results are boldfaced

Patient	Accuracy	Precision	Recall	F1-Score
Patient 1	0.86	0.94	0.80	0.86
Patient 2	0.56	0.91	0.54	0.68
Patient 3	0.56	0.97	0.54	0.69
Patient 4	0.81	0.90	0.79	0.84
Patient 5	0.73	0.94	0.72	0.81
All patients	0.70	0.93	0.66	0.77

(A) The evaluation results of combining MR8Fast, Gabor, and Phase Gradient features against GT -1-

Patient	Accuracy	Precision	Recall	F1-Score
Patient 1	0.83	0.94	0.76	0.84
Patient 2	0.58	0.90	0.56	0.69
Patient 3	0.58	0.89	0.56	0.69
Patient 4	0.81	0.90	0.79	0.84
Patient 5	0.73	0.94	0.72	0.81
All patients	0.71	0.91	0.66	0.77

(B) The evaluation results of combining MR8Fast, Gabor, and Phase Gradient features against GT -2-

Patient	Accuracy	Precision	Recall	F1-Score
Patient 1	0.87	0.92	0.82	0.87
Patient 2	0.58	0.88	0.56	0.69
Patient 3	0.59	0.89	0.56	0.69
Patient 4	0.83	0.88	0.82	0.85
Patient 5	0.75	0.93	0.74	0.82
All patients	**0.72**	**0.90**	**0.68**	**0.78**

(C) The evaluation results of combining MR8Fast, Gabor, and Phase Gradient features against the fused GT

F1-Score more than 76% and 82%. Again and similar to the baseline, this framework has lower values in the accuracy and recall for patients 2 and 3. The Accuracy and Recall values for these patients do not exceed the 59%.

4.3.5   Summary

In this section, we summarize and discuss the performance of our unsupervised frameworks while comparing it with RanPEC [19] the baseline system. Overall, the baseline system outperforms all our unsupervised frameworks with an Accuracy of 78% and F1-Score of 84%. Accordingly, we can sort our unsupervised frameworks as follow - the more

accurate first - (1) MR8Fast, (2) Ensemble of Features (Ens. Features), (3) Gabor, and
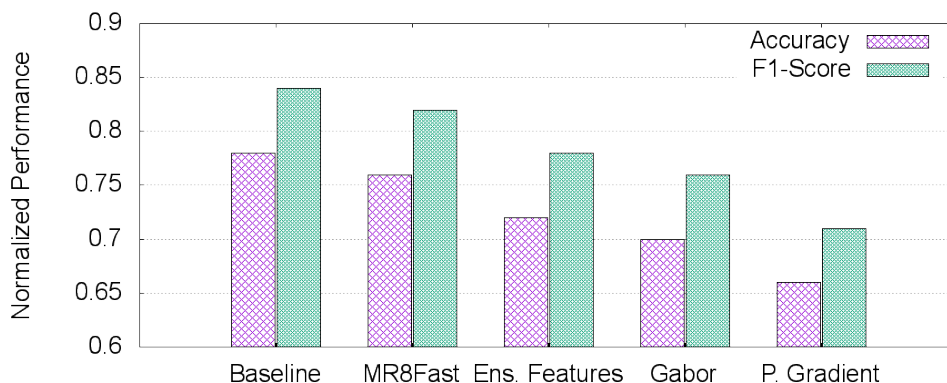(4) Phase Gradient (P.Gradient) as depicted in Figure 4.3

FIGURE 4.3: Comparison of the Accuracy and F1-Score values of the baseline framework and
all the unsupervised learning frameworks

Yet, our unsupervised MR8Fast framework, that employs only the MR8Fast features
[5], has a very close performance to the baseline. The Accuracy and F1-Score of this
framework are respectively 76% and 82% compared to 78% and 84% for the baseline
system that employs 4 different features - Gabor [14], Phase Gradient [24], Orientation
Pyramid [35] , and Full Wavelet packet [28] features -. Accordingly, we can conclude that
the MR8Fast features are highly discriminating features. In other words, the MR8Fast
features can materialize the differences of the tumor and non-tumor textures in the H &
E stained breast cancer images; and this is why the MR8Fast reported a better results
compared to the Gabor [13] and Phase Gradient [24] features.

On the other hand, the unsupervised Gabor framework outperformed the unsupervised
Phase Gradient framework by almost 5% in the Accuracy and F1-Score values. Accord-
ingly, we can sort the explored features as follow - the most discriminating first - (1)
MR8Fast, (2) Gabor, and (3) Phase Gradient.

Yet, as we combined all these features, the overall performance of the unsupervised
Ensemble Features framework outperformed only the unsupervised Gabor and Phase
Gradient frameworks, and did not outperform the unsupervised MR8Fast framework.
These results can be explained by the following: The total number of the MR8Fast
[5], Gabor [14], and Phase Gradient [24] features is equal to 140. The number of fea-
tures is relatively high and might include outlier and redundant features. Such outlier
and redundant features might have confused the K-Means algorithm to shape the right
clusters.

In order to understand the failure points of our frameworks, we examined the Accuracy
values recorded for every single patient. Figure 4.4 shows the Accuracy of the baseline
system and all our unsupervised frameworks for every single patient. As we can notice
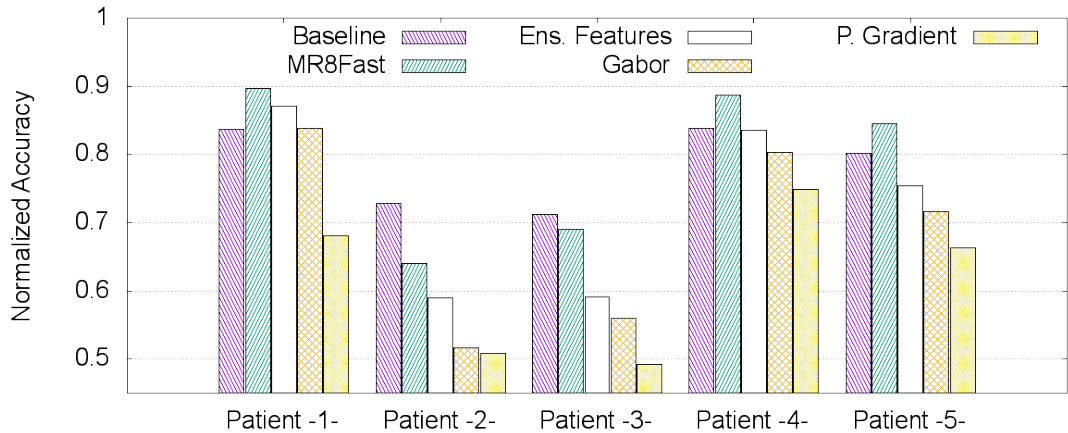
FIGURE 4.4: Comparison of the Accuracy values **per patient** for the baseline framework and all the unsupervised learning frameworks

from Figure 4.4, patients 2 and 3 have the lowest Accuracy values for all our unsupervised frameworks as well as the baseline system. The Accuracy values for patient 2 and 3 are less than 72% compared to more than 80% for the other patients. We endeavor to have better performance globally, as well as particularly for patients 2 and 3 using supervised learning. In the following sections, we present our supervised frameworks that outperform the Baseline system.

4.4 The Supervised Frameworks

In this section, we present the evaluation results of our supervised frameworks. We employ each of the MR8Fast [5], Gabor [14], Phase Gradient [24] features individually, as well as the combination of all these features. In order to reduce the number of features, we apply the Principal Component Analysis (PCA) technique. As well, we investigate the following classifiers Naive Bayes (NB), Artificial Neural Network (ANN), and Support Vector Machine (SVM), as well as the combination of all these classifiers (Ensemble classifiers). All these classifiers are trained and tested with the *fused* GT. In the following sections, we elaborate more on the training and testing data as well as the proposed supervised frameworks.

4.4.1 Training and Testing

The training data include only 50% of our dataset, whereas the remaining 50% is used for the testing. Our dataset consists of a total of 50 images. These images belong to 5 patients where every patient contributed with only 10 images. We select randomly 5 images from the 10 images of every patient to compose a training data of 25 images. For the testing data, we use the 25 remaining images. In other words the training data is different from the testing data.

Our dataset includes images of 512×512 pixels. That means the total number of pixels in one single images is equal to 262144 pixels. On the other hand, we have 25 images in the training data which means the total number of training pixels is equal to $262144 \times 25 \times numberoffeatures >> \mathbf{6553600}$. Obviously, the number of training pixels is extremely big which may not only slow down the learning process, but also confuse the classifier. For this reason, we reduce the number of training pixels. In this context, we consider two approaches: (1) the *Sample* approach, and (2) the *Mean* approach. We elaborate more these approaches in the following sections.

The Sample Approach

From the training images and after the feature extraction step of every entire image, we select randomly pixels from both the tumor and non-tumor regions. We select from every single image 3500 pixels from the tumor regions and 2000 pixels from the non-tumor regions. That means we reduced the training set to 2% from 6553600 pixels to 137500 while favoring the tumor regions - since we select a higher number of tumor pixels than non-tumor pixels -.

The Mean Approach

From the training data and after the feature extraction step, we divide every single response and the corresponding *fused* ground truth into 64 equal patches of size 64×64 pixels. We call the patches generated from the image responses *response patch*, and we call the patches generated from the *fused* ground truth *corresponding GT patch*. Then, we compute the mean for every *response patch* in order to reduce the response size from 512×512 to 64×64 pixels. These reduced responses are used for the training. Yet for the labels, we assign for every single *response patch* one single label from *the corresponding GT patch* in the following way: if the majority of pixels in the *corresponding GT patch* is tumor pixels, this *response patch* is assigned a tumor label, otherwise it is assigned non-tumor label. In this way, the training data is reduce from 6553600 pixels to 102400.

4.4.2 Evaluating the MR8Fast Features

In this section, we evaluate the MR8Fast [5] features by different supervised learning frameworks. We investigate the Naive Bayes (NB), Artificial Neural Network (ANN), and Support Vector Machine (SVM) classifiers, as well as the *Sample* and the *Mean* approaches that reduce the training data volume. In total, we develop 6 supervised frameworks for the MR8Fast features. We code our MR8Fast supervised frameworks as follow: **MR8Fast-CLASSIFIER**$_S$ for using the MR8Fast features with the appointed classifier by applying the *Sample* approach. Whereas **MR8Fast-CLASSIFIER**$_M$ for

using the MR8Fast features with the corresponding classifier by employing the *Mean* approach.
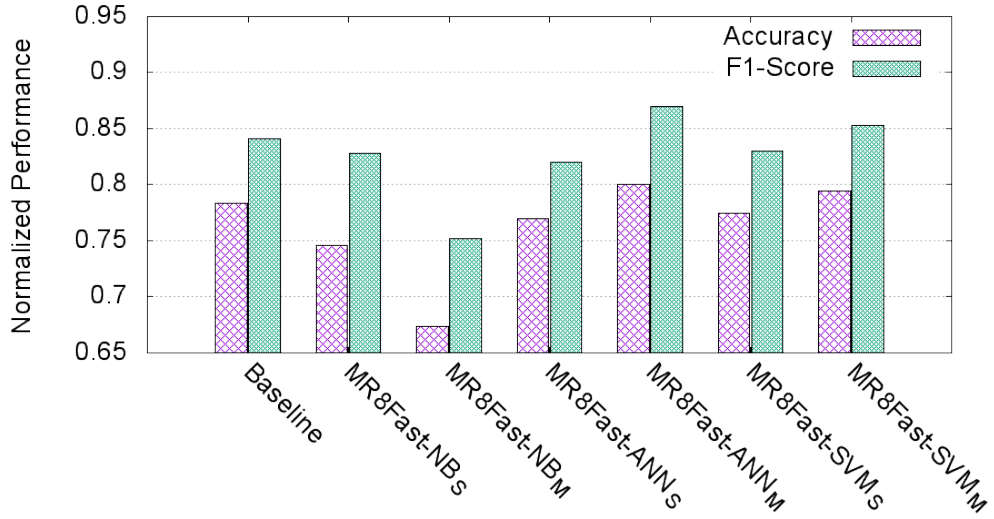

FIGURE 4.5: Comparison of the Accuracy and F1-Score values of the baseline and all the supervised MR8Fast frameworks.

We compare in Figure 4.5 the Accuracy and F1-Score values of the baseline framework and our MR8Fast supervised frameworks. We can notice that only the **MR8Fast-ANN**$_M$ and **MR8Fast-SVM**$_M$ frameworks outperform the baseline system by a 2%. On the

FIGURE 4.6: Comparison of the Accuracy values **per patient** of the baseline and the supervised MR8Fast frameworks based on the *Sample* approach

other hand, we examine the Accuracy of the MR8Fast supervised frameworks for every single patient as shown in Figure 4.6. We can notice that **MR8Fast-NB**$_S$, based on the Naive Bayes classifier, has the most balanced Accuracy values among the five patients: almost every patient has an Accuracy of more than 70%. However, the other frameworks **MR8Fast-ANN**$_S$ and **MR8Fast-SVM**$_S$, that are based on the Artificial Neural Network (ANN) and Support Vector Machine (SVM) respectively, have an accuracy of more

than 80% patients 1, 4, and 5, whereas for patients 2 and 3 the accuracy is less than 67%.

4.4.3 Evaluating the Gabor Features

In this section, we explore the Gabor [14] features by different supervised frameworks. We investigate the following classifiers: Naive Bayes (NB), Artificial Neural Network (ANN), and Support Vector Machine (SVM). As well we apply the *Sample* and the *Mean* approaches to reduce the volume of training data. Thereby and as depicted in Figure 4.7, we develop 6 supervised frameworks for the Gabor features. We name these frameworks as follow: **Gabor-CLASSIFIER**$_S$ for using only the Gabor features with the appointed classifier while reducing the training data volume by the *Sample* approach. Yet, the **Gabor-CLASSIFIER**$_M$ for using the Gabor features with the corresponding classifier with the *Mean* approach.

FIGURE 4.7: Comparison of the Accuracy and F1-Score values of the baseline and all the supervised Gabor frameworks

As shown in Figure 4.7, our **Gabor-ANN**$_M$ framework has a similar performance as the baseline system: Accuracy and F1-Score values equal to 78% and 84%. While the other supervised Gabor frameworks couldn't compete with the baseline system.

However, if we examine the Accuracy values for every single patient as shown in Figure 4.8, we can notice that the **Gabor-NB**$_S$, based on the Naive Bayes classifier, is the only framework that has balanced Accuracy values among the five patients: every patient has an Accuracy of more than 72%. Whereas the frameworks **Gabor-ANN**$_S$ and **Gabor-SVM**$_S$ report a gab between the accuracy values of every single patient that can range to more than 35%. Patients 2 and 3 report the lowest Accuracy values.
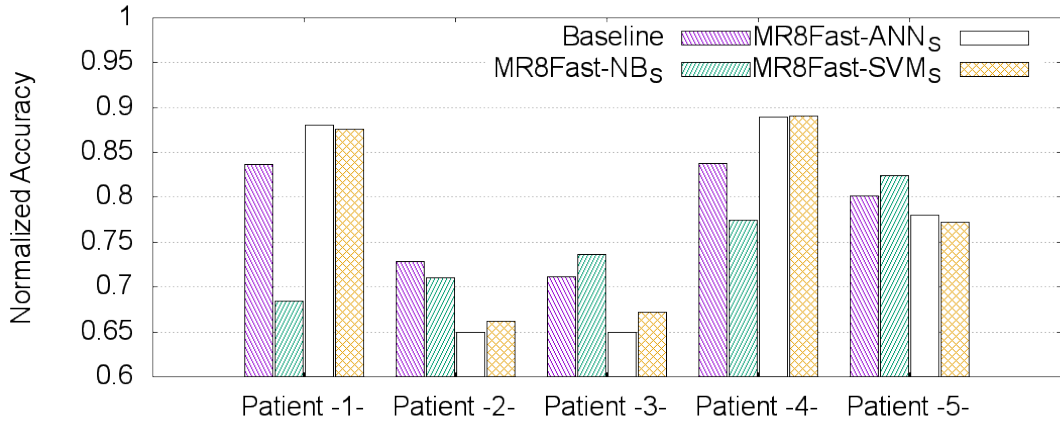
FIGURE 4.8: Comparison of the Accuracy values **per patient** of the baseline and the supervised Gabor frameworks based on the *Sample* approach

4.4.4 Evaluating the Phase Gradient Features

In this section, we evaluate the Phase Gradient [24] features while applying the following classifiers: Naive Bayes (NB), Artificial Neural Network (ANN), and Support Vector Machine (SVM). In order to reduce the training data volume, we employ the *Sample* and the *Mean* approaches that are clarified in section 4.4.1. We generated 6 supervised frameworks for the Phase Gradient features, and they are coded as follow: **P.Gradient-CLASSIFIER**$_S$ for using only the Phase Gradient features with corresponding classifier by applying the *Sample* approach. While **P.Gradient-CLASSIFIER**$_M$ refers to the frameworks that employ the Phase Gradient features that are based on the *Mean* approach. We present the evaluation results of supervised Phase Gradient frameworks in Figures 4.9 and 4.10.
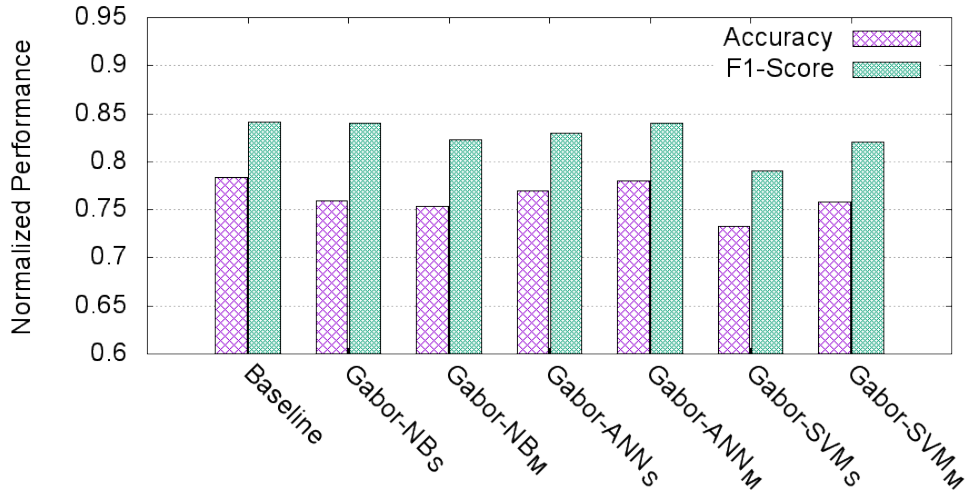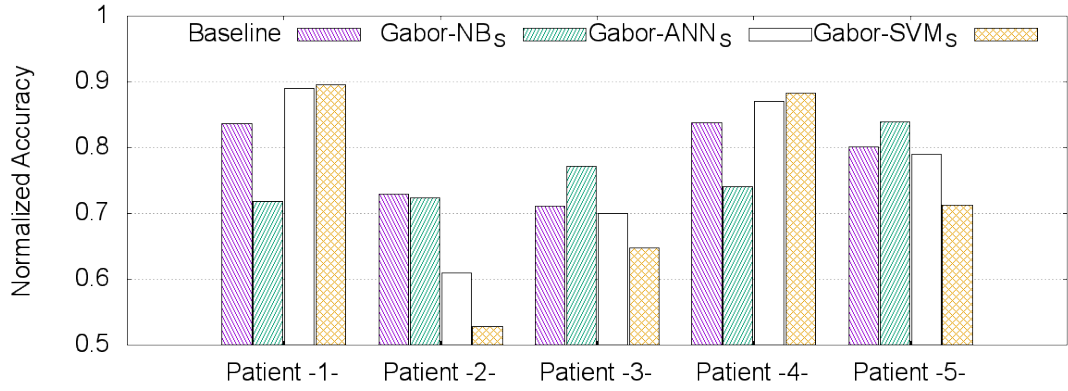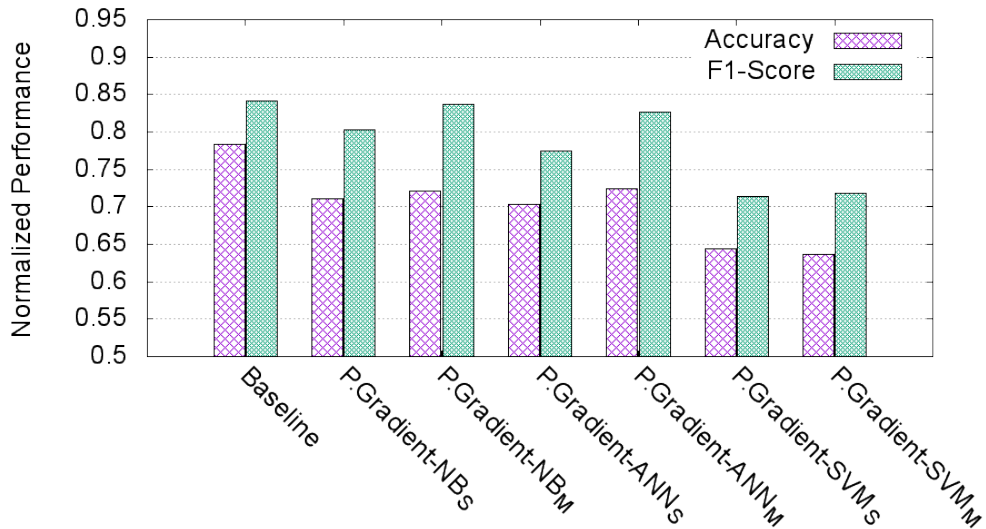

FIGURE 4.9: Comparison of the Accuracy and F1-Score values of the baseline and all the supervised Phase Gradient frameworks

As shown in Figure 4.9, our Phase Gradient supervised frameworks outperform the baseline system. These average results can be explained as the Phase Gradient features are not a very discriminating features for the tumor texture in the H & E stained breast cancer images.

We examine as well the Accuracy values if these Phase Gradient frameworks for every single patient as shown in Figure 4.10. We can notice again that patients 2 and 3 report the lowest performance compared to other patients. The Accuracy values for patients 2 and 3 are less than 52% compared to more than 70% for the other patients.
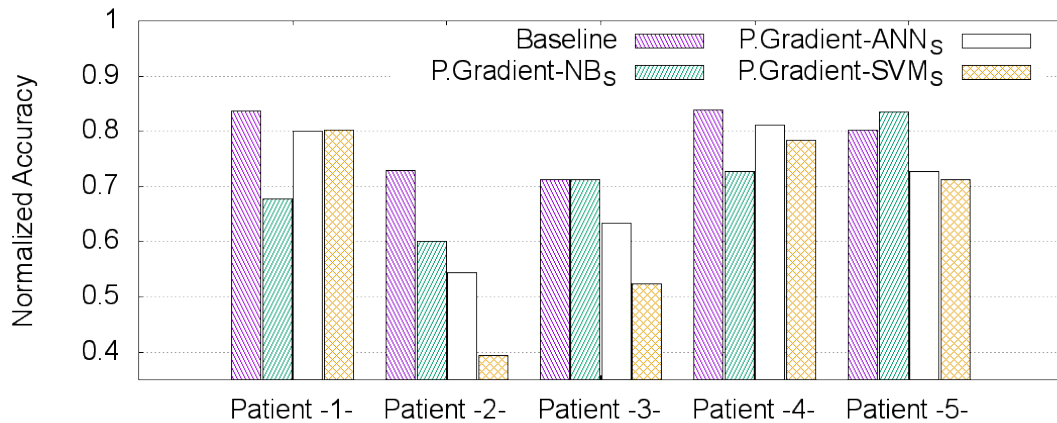

FIGURE 4.10: Comparison of the Accuracy values **per patient** of the baseline and the supervised Phase Gradient frameworks based on the *Sample* approach

4.4.5   Evaluating the Ensemble Features

In this section, we evaluate the combination of the MR8Fast [5], Gabor [14], and Phase Gradient [24] features by our supervised frameworks. We investigate the following classifiers Naive Bayes (NB), Artificial Neural Network (ANN), and Support Vector Machine (SVM), as well as the ensemble of these classifiers (Ens.Clf). In order to reduce the training data volume, we apply both the *Sample* and the *Mean* approaches - explained in the section4.4.1 -. In total, we generated 7 Ensemble-Features supervised frameworks, and they are coded as follow: **Ens.F-CLASSIFIER**$_S$ for the *Sample* based approach, and **Ens.F-CLASSIFIER**$_M$ for the *Mean* based approach. The framework **Ens.F-Ens.Clf** refers to the supervised Ensemble of features framework that employ the ensemble of classifiers.

We compare in Figure 4.11 the overall performance of our Ensemble-features (Ens.F) supervised frameworks. As shown in figure 4.11, our frameworks **Ens.F-Ens.Clf** and **Ens.F-NB**$_S$ outperform the Accuracy and F1-Score of the baseline system by more than**3%** and **2%** respectively.
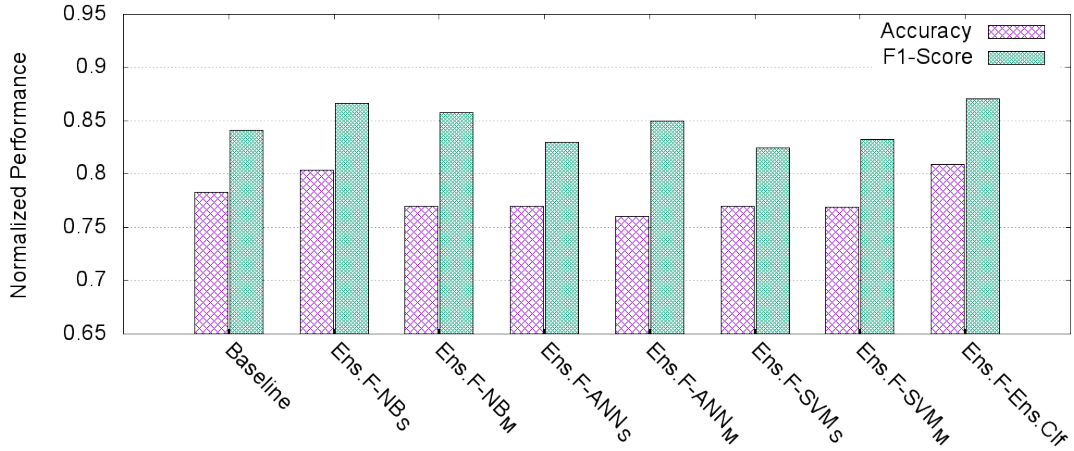
FIGURE 4.11: Comparison of the Accuracy and F1-Score values of the baseline and all the supervised Ensemble Features frameworks

In addition, we examine the Accuracy of both the **Ens.F-Ens.Clf** and **Ens.F-NB**$_S$ frameworks for every single patient as shown in Figure 4.12. We can notice that these frameworks do not only outperform the overall performance of the baseline system, but also these frameworks provide a balanced performance among the five patients. The overall Accuracy of these frameworks is around 81% while the lowest Accuracy per patient is more than 75%.
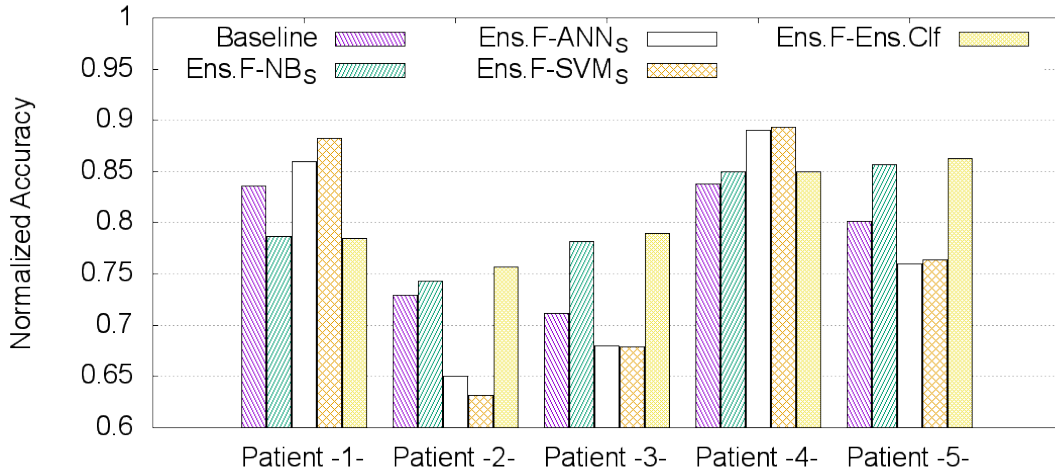
FIGURE 4.12: Comparison of the Accuracy values **per patient** of the baseline and the supervised Ensemble Features frameworks based on the *Sample* approach

4.5 Summary & Discussions

In this section, we summarize and discuss the several results of our proposed frameworks that segment tumor regions in H&E stained breast cancer images. In these frameworks we investigated textural features: MR8Fast [5], Gabor [14], and Phase Gradient [24],

as well as a combination of all these features. We explored each of these features individually. We noticed that the MR8Fast [5] frameworks, that employ only the MR8Fast features, have always better Accuracy values - up to 76% - compared to the Gabor [14] and Phase Gradient [24] frameworks - Accuracy values do not exceed 70% and 73% respectively. As well, we noticed that the Phase Gradient frameworks, that employ the Phase Gradient features, have always the lowest Accuracy values - up to 73%. Accordingly, the most discriminating features, for tumor and non-tumor texture in microscopic images of breast cancer tissue, are the MR8Fast features, followed by the Gabor features, followed by the Phase Gradient features that are considered to be the less discriminating features.

(A) An image of Patient -2- (B) GT -1- (C) GT -2-

(D) Baseline segmentation (E) Ens.F-NB$_S$ segmentation

FIGURE 4.13: A sample image of **patient -2-** (A) with the corresponding ground truth images (B and C), as well as the segmenation results of the baseline framework (D) and one of our best frameworks *Ens.F-NB$_S$* (E). The white labels correspond to the tumor regions, while the black labels are assigned to the non-tumor regions.

In this study, our developed frameworks include both unsupervised and supervised frameworks. Our unsupervised frameworks employ each of the MR8Fast, Gabor, and Phase Gradient features individually, as well as a combination of all these features. In addition, our unsupervised frameworks employ the K-Means clustering algorithm similarly to the Baseline framework. Our unsupervised frameworks were not able to outperform the Baseline framework. Our best unsupervised framework is the **MR8Fast** that has an Accuracy value equal to 76% compared to 78% for the Baseline framework. These

(A) An image of patient -3-   (B) GT -1-   (C) GT -2-

(D) Baseline segmentation   (E) Ens.F-NB$_S$ segmentation
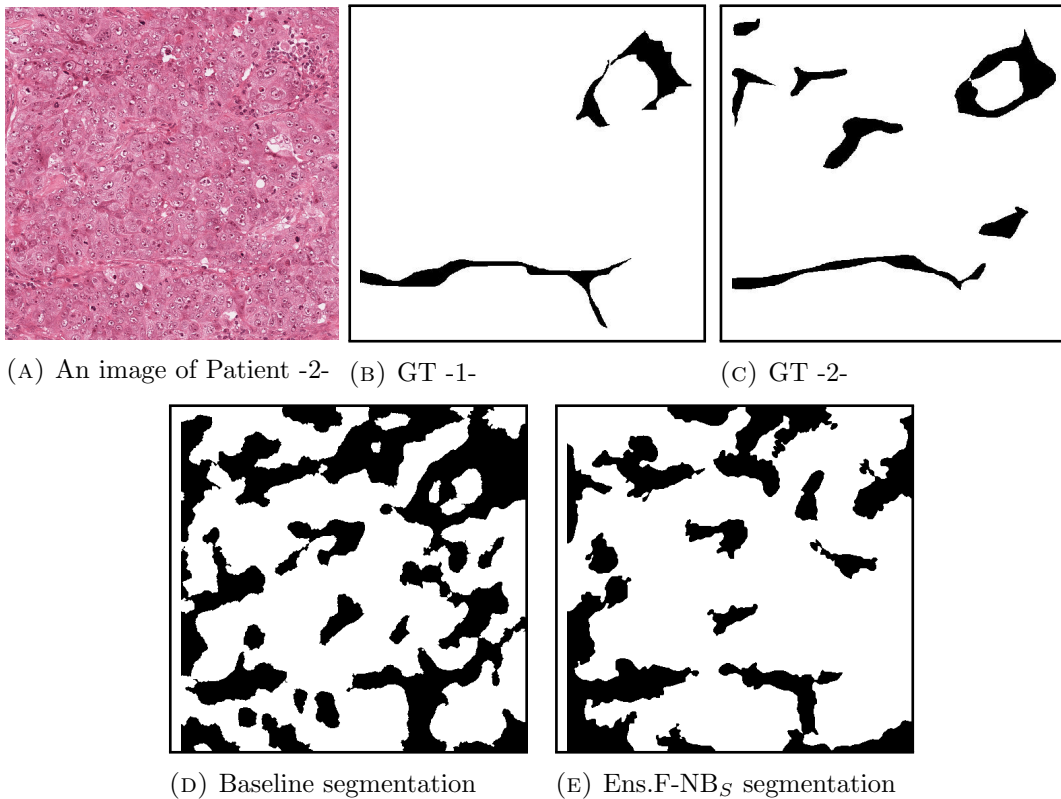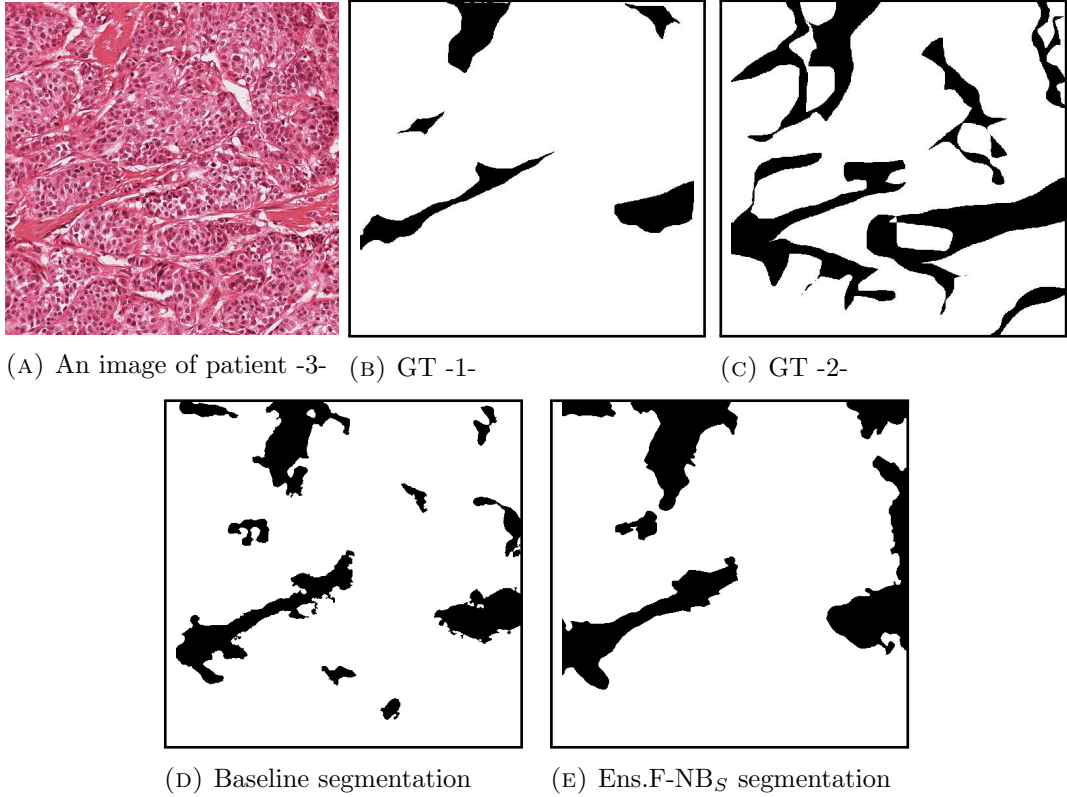
FIGURE 4.14: A sample image of **patient -3-** (A) with the corresponding ground truth images (B and C) ,and the segmenation results of the baseline framework (D) and one of our best frameworks *Ens.F-NB$_S$* (E). The white labels correspond to the tumor regions, while the black labels are assigned to the non-tumor regions.

results can be explained to the following reasons: (1) the baseline framework employs other features such as the full wavelet and orientation pyramids, and (2) the baseline framework uses several runs of K-means algorithm after random projections into 20 different random bases.

Yet, to understand better the results of our unsupervised framework and the baseline framework, we computed separately the Accuracy values for every single patient. The results for our unsupervised frameworks as well as the baseline framework reveal that patients 2 and 3 report the *lowest* Accuracy values - less than 72% - compared to the other patients - more than 85%. Accordingly, these two patients are the failure points for our unsupervised frameworks as well as the baseline system which can be accounted for the following reasons:

- the images of patients 2 and 3 are difficult cases to differentiate between the tumor and non-tumor texture. We present in Figures 4.13 and 4.14 samples of images from the dataset of patients 2 and 3. We can notice from these images that the texture of the tumor and non-tumor regions are very similar which make it difficult to distinguish between the tumor and non-tumor texture.

- the ground truth for these images can be inexact, especially, as we can notice in Figure 4.13 and 4.14, the ground truth images provided by both pathologists are different.

In order to have better segmentation results in overall as well as for patients 2 and 3, we considered different supervised frameworks. Our supervised frameworks include also the MR8Fast, Gabor, and Phase Gradient features, as well as a combination of all these features - called Ensemble Features. Our supervised frameworks employ different classifiers which are Naive Bayes, Artificial Neural Network, and Support Vector Machine, as well as a combination of all these classifiers - that we called Ensemble Classifiers. Our best supervised frameworks are the $\textbf{Ens.F-NB}_S$ and **Ens.F-Ens.Clf** frameworks that apply Ensemble Features together with the (1) Naive Bayes classifier and (2) Ensemble classifiers respectively. We compare in Figure 4.15 the performance of our best supervised frameworks $\textbf{Ens.F-NB}_S$ and **Ens.F-Ens.Clf**, as well as the Baseline framework.

As shown in Figure 4.15a, our supervised frameworks $\textbf{Ens.F-NB}_S$ and **Ens.F-Ens.Clf** outperform the Accuracy of the Baseline framework by 3% and 4% respectively. In addition, both frameworks have better performance for patients 2 and 3 compared to the Baseline framework. Indeed, the Accuracy for patients 2 and 3 are improved by at least by 2% and 7% respectively.
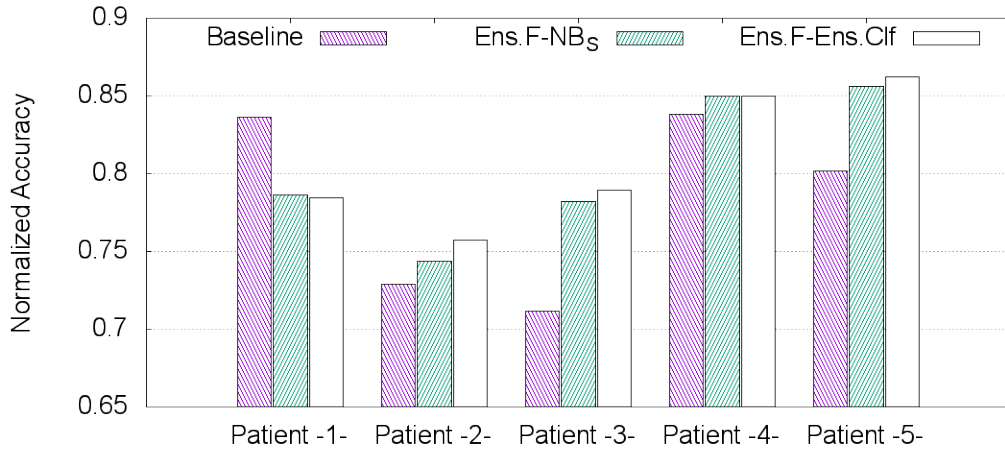
The results of our best supervised frameworks $\textbf{Ens.F-NB}_S$ and **Ens.F-Ens.Clf** that outperform the Baseline framework, can be explain by the good combination of Ensemble features with the Naive Bayes classifier and Ensemble classifiers. The $\textbf{Ens.F-NB}_S$ framework employs Ensemble features that combine different textural features - MR8Fast, Gabor, and Phase Gradient -, as well as the Naive Bayes classifier. The Naive Bayes is a probabilistic classifier that we employ in the task of segmentation of tumor regions in microscopic images of breast cancer tissue. This classifier achieved better results with an overall Accuracy 80% compared to only 76% for the Support Vector Machine classifier which is the only classifier used in the literature for such task.

On the other hand our supervised framework **Ens.F-Ens.Clf** that employs also Ensemble Features with Ensemble Classifiers outperform the Accuracy of the baseline framework and our $\textbf{Ens.F-NB}_S$ framework by 4% and 1% respectively. These results can be explained by the joint endeavor of each of the Naive Bayes, Artificial Neural Network, and Support Vector Machine classifiers. The Combination of all these classifiers is employed also for the task of segmentation of tumor regions in microscopic images of breast cancer tissue.

Although our supervised frameworks $\textbf{Ens.F-NB}_S$ and **Ens.F-Ens.Clf** improved the overall Accuracy, the Accuracy values for patients 2 and 3 are still the lowest compared
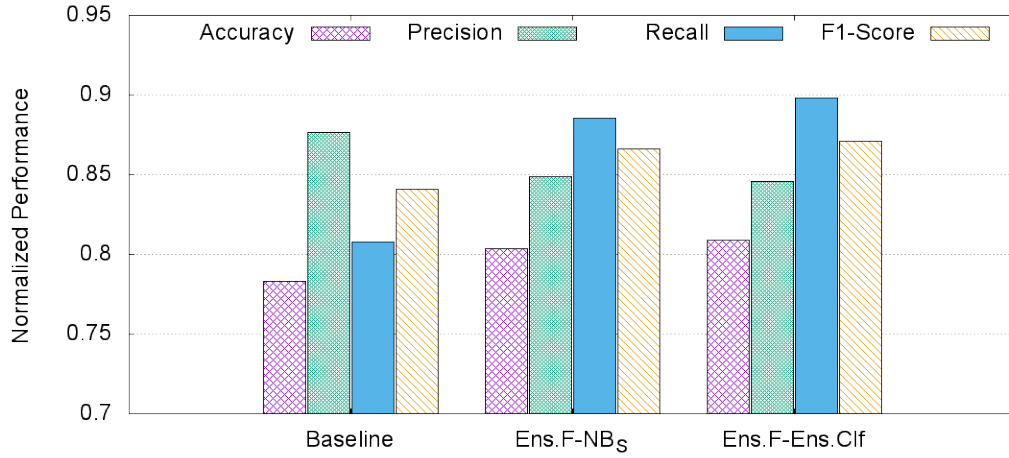
(A) Comparison of the Accuracy and F1-Score values of the baseline and all the supervised Ensemble Features frameworks

(B) Comparison of the Accuracy values **per patient** of the baseline and the supervised Ensemble Features frameworks based on the *Sample* approach

FIGURE 4.15: Comparative results of our **best** supervised frameworks that outperform the Baseline framework

to the other patients. These results can reinforce our previous assumption which claims that patients 2 and 3 represent difficult cases.

It should be noted that in order to compare our experimental results, we considered only the results that refer to the *Sample* approach for reducing the training data, and **not** the *Mean* approach - both detailed in Section 4.4.1. Indeed, the *Mean* approach is based only on the mean to reduce the pixels number of the training data which may not be fully representative. The mean may lead to loss of information of the reduced pixels as it does not consider outliers such as very big and very small values. For, this reasons, we relied more on the *Sample* approach, and in this context, we consider the standard deviation together with the mean to reduce the training data as a future work. We summarize and detail more the future work in the following Chapter.

Chapter 5

Conclusions and Future Work

In the first section of this Chapter, we summarize and conclude our work for the task of segmentation of tumor regions in microscopic images of H&E stained breast cancer tissue. In the second section, we give perspectives on the future work that can supplement our study.

5.1   Summary & Conclusions

In this study, we considered the segmentation of tumor regions in microscopic images of breast cancer tissue as a machine learning task. Accordingly we developed different frameworks that include the following steps: (1) pre-processing, (2) feature extraction, (3) feature reduction, (4) learning process, and (5) post-processing. We focused mainly on the feature extraction and learning process steps.

In the pre-processing step, we applied different techniques in the following order: (1) stain normalization in order to have better color consistency of the overall images, (2) conversion to the gray scale, and (3) noise removal by smoothing using Gaussian filter. The resultant pre-processed image is used for the feature extraction step.

In the feature extraction step, we investigated textural features that consist of the MR8Fast, Gabor, and Phase Gradient features, as well as a combination of all these features. These features are originally different filter responses that are generated from the corresponding filter banks. In this work, we concluded that the MR8Fast is the most discriminating features which means it materializes better the differences between the tumor and non-tumor textures. The Gabor and Phase Gradient features come in the second and third place respectively with Phase Gradient features are the less discriminating features.

As the number of generated features is large, we used the technique of Principal Component Analysis to reduce the number of features without too much information loss. Then, we applied the reduced features in the learning process in order to identify the tumor and non-tumor regions.

In the learning process step, we investigated both unsupervised and supervised learning frameworks. Our unsupervised frameworks employed only the K-Means clustering

algorithm, while our supervised frameworks applied different classifiers: Naive Bayes, Artificial Neural Network, and Support Vector Machine, as well as a combination of these classifiers results. The results of the learning process consist of segmentation results that require more enhancements. For this reason, we applied a post-processing step which includes the technique of area threshold in order to improve the segmentation results of the learning process.

Our experimental results reveal that the proposed unsupervised frameworks were not able to compete with the Baseline framework, whereas our best supervised frameworks **Ens.F-NB**$_S$ and **Ens.F-Ens.Clf** outperformed the overall Accuracy of the Baseline framework by 3% and 4% respectively. The **Ens.F-NB**$_S$ and **Ens.F-Ens.Clf** frameworks investigated the combination of the MR8Fast, Gabor, Phase Gradient features. The joint efforts of all these features materializes better the differences of the tumor and non-tumor textures. In our best supervised frameworks **Ens.F-NB**$_S$ and **Ens.F-Ens.Clf**, we investigated also the Naive Bayes classifier, and Ensemble classifiers - combination of classifier results - respectively. These two learning techniques outperformed the Support Vector Machine classifier which has been employed extensively in the literature for the task of segmentation of tumor regions in microscopic image of breast cancer tissue.

5.2 Future Work

Given the proposed frameworks and the subsequent conclusions, we developed several directions for future work that may supplement our study:

- In this study, we experienced the difficult cases of the images of patients 2 and 3 that recorded the lowest Accuracy values. These cases should be more investigated:

  1. Consider the performance of our best frameworks with a third ground truth provided by a third experienced pathologist.
  2. Consider other textural features that characterize better the differences of tumor and non-tumor textures in these particular patients.
  3. investigate other types of features such as network features, as well as other more recent learning strategies such as deep convolutional neural networks.

- Consider the proposed frameworks with another bigger dataset that includes other patients. As well, investigate the robustness of our supervised frameworks in supporting new tumor texture - knowing that the tumor texture may change from one patient to another due to the anatomical differences between patients.

- In this study, we developed the *Mean* approach to reduce the training dataset. However, we did not fully consider the results that refer to this approach because it is based only on mean. As a future work, we consider the standard deviation along with the mean to have more reliable results.

Bibliography

[1] Histo-pathological breast cancer images. http://www.cancer.org/cancer/breastcancer/detailedguide/breast-cancer-diagnosis. Accessed: 2015-10-17.

[2] Natinal cancer institute-what is cancer? http://www.cancer.gov/about-cancer/what-is-cancer. Accessed: 2015-11-17.

[3] Preparation of the H & E stained slide. http://amida13.isi.uu.nl. Accessed: 2015-10-17.

[4] Stereotactic biopsy of the breast. http://www.webmd.com/breast-cancer/stereotactic-biopsy-of-the-breast. Accessed: 2015-10-17.

[5] Texture classification. http://www.robots.ox.ac.uk/ṽgg/research/texclass/filters.html. Accessed: 2015-12-07.

[6] Understanding breast cancer. http://www.breastcancer.org/symptoms/understand_bc. Accessed: 2015-12-15.

[7] What is cancer? http://www.seattlecca.org/what-is-cancer.cfm.

[8] S. Akbar, T. Amaral, S. J. McKenna, A. Thompson, and L. Jordan. Tumour segmentation in breast tissue microarray images using spin-context. In *Proc. Medical Image Understanding and Analysis*, pages 25–30, 2012.

[9] S. Akbar, S. J. McKenna, T. Amaral, L. Jordan, and A. Thompson. Spin-context segmentation of breast tissue microarray images. *Ann BMVA*, 2013:1–11, 2013.

[10] P. Boyle and J. Ferlay. Cancer incidence and mortality in Europe, 2004. *Annals of oncology*, 16(3):481–488, 2005.

[11] A. Chekkoury, P. Khurd, J. Ni, C. Bahlmann, A. Kamen, A. Patel, L. Grady, M. Singh, M. Groher, N. Navab, et al. Automated malignancy detection in breast histopathological images. In *SPIE Medical Imaging*, pages 831515–831515. International Society for Optics and Photonics, 2012.

[12] R. Dass and S. Devi. Image segmentation techniques 1. *International Journal of Electronics & Comminucation Technology*, 3:66–70, 2012.

[13] J. Ferlay, H.-R. Shin, F. Bray, D. Forman, C. Mathers, and D. M. Parkin. Estimates of worldwide burden of cancer in 2008: Globocan 2008. *International journal of cancer*, 127(12):2893–2917, 2010.

[14] I. Fogel and D. Sagi. Gabor filters as texture discriminator. *Biological cybernetics*, 61(2):103–113, 1989.

[15] R. C. Gonzalez and R. E. Woods. Digital image processing. *ed: Prentice Hall Press, ISBN 0-201-18075-8*, 2002.

[16] C. Gunduz-Demir, M. Kandemir, A. B. Tosun, and C. Sokmensuer. Automatic segmentation of colon glands using object-graphs. *Medical image analysis*, 14(1):1–12, 2010.

[17] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener. Histopathological image analysis: A review. *Biomedical Engineering, IEEE Reviews in*, 2:147–171, 2009.

[18] B. Karaçali and A. Tözeren. Automated detection of regions of interest for tissue microarray experiments: an image texture analysis. *BMC Medical Imaging*, 7(1):2, 2007.

[19] A. M. Khan, H. El-Daly, and N. Rajpoot. RanPEC: Random projections with ensemble clustering for segmentation of tumor areas in breast histology images. In *Medical Image Understanding and Analysis*, pages 17–23, 2012.

[20] A. M. Khan, H. El-Daly, E. Simmons, and N. M. Rajpoot. HyMaP: A hybrid magnitude-phase approach to unsupervised segmentation of tumor areas in breast cancer histology images. *Journal of pathology informatics*, 4(Suppl), 2013.

[21] M. Macenko, M. Niethammer, J. Marron, D. Borland, J. T. Woosley, X. Guan, C. Schmitt, and N. E. Thomas. A method for normalizing histology slides for quantitative analysis. In *ISBI*, volume 9, pages 1107–1110, 2009.

[22] A. Madabhushi. Digital pathology image analysis: opportunities and challenges. *Imaging in Medicine*, 1(1):7–10, 2009.

[23] K. P. Murphy. Naive bayes classifiers. *University of British Columbia*, 2006.

[24] K. Murtaza, S. Khan, and N. Rajpoot. Villagefinder: Segmentation of nucleated villages in satellite imagery. In *Proceedings of the British Machine Vision Conference*, pages 83.1–83.11. BMVA Press, 2009. doi:10.5244/C.23.83.

[25] L. Pantanowitz et al. Digital images and the future of digital pathology. *Journal of pathology informatics*, 1(1):15, 2010.

[26] M. Peikari, M. Gangeh, J. Zubovits, G. Clarke, and A. Martel. Triaging diagnostically relevant regions from pathology whole slides of breast cancer: A texture based approach. *Medical Imaging, IEEE Transactions on*, pages 307–315, 2015.

[27] A. Qu, J. Chen, L. Wang, J. Yuan, F. Yang, Q. Xiang, N. Maskey, G. Yang, J. Liu, and Y. Li. Segmentation of hematoxylin-eosin stained breast cancer histopathological images based on pixel-wise svm classifier. *Science China Information Sciences*, pages 1–13, 2015.

[28] N. M. Rajpoot. Texture classification using discriminant wavelet packet subbands. *Circuits and Systems, 2002. MWSCAS-2002. The 2002 45th Midwest Symposium on*, 3:III–300–III–303 vol.3, 2002.

[29] E. Reinhard, M. Ashikhmin, B. Gooch, and P. Shirley. Color transfer between images. *IEEE Computer graphics and applications*, (5):34–41, 2001.

[30] L. Roux, D. Racoceanu, N. Loménie, M. Kulikova, H. Irshad, J. Klossa, F. Capron, C. Genestie, G. Le Naour, and M. N. Gurcan. Mitosis detection in breast cancer histological images an ICPR 2012 contest. *Journal of pathology informatics*, 4, 2013.

[31] C. Sammut and G. I. Webb. *Encyclopedia of machine learning.* Springer Science & Business Media, 2011.

[32] A. Shimizu, T. Kimoto, H. Kobatake, S. Nawano, and K. Shinozaki. Automated pancreas segmentation from three-dimensional contrast-enhanced computed tomography. *International journal of computer assisted radiology and surgery*, 5(1):85–98, 2010.

[33] L. I. Smith. A tutorial on principal components analysis. *Cornell University, USA*, 51:52, 2002.

[34] H. Suzuki and J.-i. Toriwaki. Automatic segmentation of head MRI images by knowledge guided thresholding. *Computerized medical imaging and graphics*, 15(4):233–240, 1991.

[35] R. Wilson and M. Spann. Finite prolate spheroidal sequences and their applications. ii. image feature description and segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 10(2):193–203, 1988.

[36] X. Zhu. Semi-supervised learning literature survey. 2005.