UIVERSIDADE FEDERAL DO PARANÁ

FABIO ALEXANDRE SPANHOL

AUTOMATIC BREAST CANCER CLASSIFICATION FROM HISTOPATHOLOGICAL IMAGES: A HYBRID APPROACH

CURITIBA
2018

FABIO ALEXANDRE SPANHOL

AUTOMATIC BREAST CANCER CLASSIFICATION FROM HISTOPATHOLOGICAL IMAGES: A HYBRID APPROACH

Tese apresentada como requisito parcial à obtenção do título de Doutor em Ciência da Computação, Programa de Pós-Graduação em Informática, setor de Ciências Exatas, Universidade Federal do Paraná (UFPR).

Área de concentração: *Ciência da Computação*.

Orientador: Prof. Dr. Luiz Eduardo Soares de Oliveira.

CURITIBA

2018

Catalogação na Fonte: Sistema de Bibliotecas, UFPR
Biblioteca de Ciência e Tecnologia

S735a Spanhol, Fabio Alexandre
Automatic breast cancer classification from histopathological images: a hybrid approach / Fabio Alexandre Spanhol. – Curitiba, 2018.

Tese - Universidade Federal do Paraná, Setor de Ciências Exatas, Programa de Pós-Graduação em Informática, 2018.

Orientador: Luiz Eduardo Soares de Oliveira .

1. Mamas – Câncer. 2. Histopatologia. 3. Redes neurais (Computação). 4. Convoluções (Matemática). 5. Processamento de imagens. I. Universidade Federal do Paraná. II. Oliveira, Luiz Eduardo Soares de. III. Título.

CDD: 616.99449

Bibliotecário: Elias Barbosa da Silva CRB-9/1894

MINISTÉRIO DA EDUCAÇÃO
SETOR SETOR DE CIENCIAS EXATAS
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO INFORMÁTICA

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em INFORMÁTICA da Universidade Federal do Paraná foram convocados para realizar a arguição da tese de Doutorado de **FABIO ALEXANDRE SPANHOL** intitulada: **Automatic Breast Cancer Classification From Histopathological Images: A Hybrid Approach**, após terem inquirido o aluno e realizado a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de doutor está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

Curitiba, 10 de Agosto de 2018.

LUIZ EDUARDO SOARES DE OLIVEIRA
Presidente da Banca Examinadora

PAULO RODRIGO CAVALIN
Avaliador Externo

DAVID MENOTTI GOMES
Avaliador Interno

LUCAS FERRARI DE OLIVEIRA
Avaliador Interno

ALCEU DE SOUZA BRITTO JR
Avaliador Externo

MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO PARANÁ
Setor de Ciências Exatas
Coordenação de Pós-Graduação em Informática

What we call chaos is just patterns we haven't recognized. What we call random is just patterns we can't decipher.

Chuck Palahniuk

This thesis is dedicated to the memory of my father Alcino Spanhol, who was so early taken from us due to cancer. His life teachings about honesty, perseverance, and integrity will be with me for all my days.

Acknowledgments

First and foremost I want to thank my advisor Professor Luiz Eduardo Soares de Oliveira for the continuous support, even in the harshest times. It has been an honor for me to be your Ph.D. student. I could not have imagined having a better advisor for my Ph.D study. He has shown me, by his stimulating example, what a good researcher should be. And here I have to mention Jefferson Gustavo Martins, my personal friend and co-worker at UTFPR-Toledo. After exploring the current research of different potential Ph.D. supervisors, it was through his recommendation that I decided to contact Professor Luiz Eduardo to be my advisor.

Besides my advisor, I would like to thank the rest of my thesis committee members: Professor Dr. Alceu de Souza Britto Jr (PUCPR), Dr. Paulo Cavalin (IBM Research-Brazil), Professor Dr. Lucas Ferrari de Oliveira (UFPR), and Professor Dr. David Menotti (UFPR), for their valuable insightful comments and suggestions which incented me to widen my research from various perspectives.

I am also especially grateful for the time spent at the LITIS lab (*Université de Rouen/France*) under supervision of Professors Dr. Laurent Heutte and Dr. Caroline Petitjean, who gave access to the laboratory and research facilities. I would like to express my sincere gratitude for all the support, welcoming and kindness that my son Marcus and I received in Rouen. The professional relationship and the cultural experience were priceless gains. Thank you for the opportunity.

During my stay in Rouen I met an extremely friendly Ph.D. candidate and real nice person named Hongliu Cao, aka Léo. I am completely sure that Léo is the Chinese most French ever. Indeed, my time at LITS was made enjoyable in large part due to Léo's friendship. I have appreciated the camaraderie and local expertise. In fact, Léo also introduced me to two other foreign Ph.D. candidates: Wassim Swaileh (from Yemen) and Roger Trullo (from Colombia). Although we only spent a short time together, drinking some beers and sharing a *shisha* session, it was an amazing experience to have met you guys. I am grateful for the time spent with my Brazilian colleague Rafael Will Macêdo de Araújo, also a Ph.D. candidate (IME-USP) at the LITIS lab. My trips to the Leclerc in St. Étienne du Rouvray have never been the same after Rafael's company.

I would like to thank my Ph.D. colleagues at UFPR, Paulo Ricardo Lisboa de Almeida, André Gustavo Hochuli (Taz), Jeovane Alves (Jeová), and Ivan Pires for the friendship and camaraderie. There were countless lunches at the university restaurant and equal times in the cafeteria, even when we were surrounded by pigeons and bees. I thank for all the fun we have had in the last four years.

I gratefully acknowledge the funding sources that made my Ph.D. work possible, including my university, UTFPR-Toledo. I was funded by the CAPES foundation for my 4 years. My work was also supported by the colleagues of the TSI group at UTFPR-Toledo.

I would like to thank the valuable collaboration of P&D Laboratory-Brazil, for providing the slide collection of histopathological studies of breast tissue employed to build the BreaKHis database. In particular, I would like to acknowledge the pathologists Fábio Negretti of P&D, for their valuable feedback throughout the image acquisition process. Also I am grateful to Carlos Eduardo Pokes, medical student, for his great commitment to the digitization process.

Moreover, I am particularly indebted to my dear friends Edson Tavares de Camargo and his wife Heloísa. Hosting me in their house in Curitiba, for almost one year, they made me feel at home.

My endless gratitude to my parents, Alcino Spanhol (*in memoriam*) and Deisy do Carmo Bahú Spanhol, who raised me with perseverance and dignity, despite all the poverty and adversities. Both of you will always be my moral inspiration.

Lastly, I would like to thank my core family for all their immense love, patience and encouragement: my wife Claudiana Soerensen (Clau) and my son Marcus Vinícius Soerensen Spanhol (Vini). You have always supported me in all my pursuits and inspire me to be better. I love you!

Fabio Alexandre Spanhol
June 2018

RESUMO

O câncer é um grave problema de saúde pública mundial, apresentando altas taxas de mortalidade e sobrecarga dos sistemas de saúde públicos e privados. Especialmente na população feminina, o câncer de mama surge como o segundo mais incidente e um dos tipos mais letais quando não devidamente diagnosticado e tratado. Apesar do progresso significativo alcançado pelas tecnologias de diagnóstico por imagem, a biópsia é a única maneira de diagnosticar com confiança se o câncer está realmente presente. O diagnóstico final do câncer de mama, incluindo graduação e estadiamento, ainda continua sendo feito por patologistas aplicando inspeção visual de imagens histológicas sob o microscópio. A análise histopatológica é uma tarefa altamente especializada, demorada, extremamente dependente da experiência dos patologistas e diretamente influenciada por fatores tais como fadiga e diminuição da atenção. Avanços recentes em técnicas de processamento de imagem e de aprendizado de máquina permitem construir sistemas CAD (Computer-Aided Diagnosis) que podem ajudar os patologistas a serem mais produtivos, objetivos e consistentes no processo de diagnóstico. Infelizmente, há também uma falta de bases de dados de imagem histológicas rotuladas abrangentes e públicas destinadas à pesquisa em sistemas CAD. Bases de dados rotuladas são cruciais para desenvolver e validar sistemas de aprendizado de máquina. Além disso, o desempenho da maioria dos sistemas de classificação convencionais depende da representação de dados apropriada e grande parte dos esforços são dedicados a engenharia de características, um processo difícil e demorado que usa o conhecimento prévio do especialista no domínio para criar características úteis. Contribuindo para eliminar esta lacuna, o presente trabalho apresenta um novo conjunto de dados publicamente disponível chamado BreaKHis, o qual contém imagens histopatológicas de tumores mamários. Este trabalho também apresenta uma abordagem alternativa para classificar estas imagens desafiadoras evitando qualquer segmentação explícita. Tal abordagem explora descritores texturais manuais e representação automática, particularmente empregando Redes Neurais Convolucionais, bem como o paradigma Aprendizado de Instâncias Múltiplas. Os resultados experimentais obtidos demonstraram a viabilidade desta proposta, dando indicações para melhorar tal modelo.

Palavras-chave: *reconhecimento de padrões, câncer de mama, imagem histopatológica, aprendizagem profunda, rede neural convolucional, aprendizado de instâncias múltiplas, processamento de imagens, aprendizado de máquina, diagnóstico auxiliado por computador.*

ABSTRACT

Cancer is a serious public health problem worldwide, presenting high mortality rates and overloading both public and private health systems. Especially in the female population, breast cancer emerges as the second most incident and one of the most lethal types when not properly diagnosed and treated. In spite of the significant progress reached through the diagnostic imaging technologies, biopsy is the only way to diagnose with confidence if cancer is really present. Final breast cancer diagnosis, including grading and staging, still continues to be done by pathologists applying visual inspection of histological images under the microscope. Histopathological analysis is a highly specialized and time-consuming task, extremely dependent on the experience of the pathologists and directly influenced by factors such as fatigue and decrease of attention. Recent advances in image processing and machine learning techniques, which allow CAD (Computer-Aided Diagnosis) systems to be built, which in turn can assist pathologists to be more productive, objective and consistent in the diagnosis process. Unfortunately, there is also a lack of comprehensive and public annotated histological image databases intended for research in CAD systems. Annotated databases are crucial for developing and validating machine learning systems. Moreover, the performance of most conventional classification systems relies on appropriate data representation and much of the efforts are dedicated to feature engineering, a difficult and time-consuming process that uses prior expert domain knowledge of the data to create useful features. By contributing to eliminate this gap, the present work introduces a new publicly available image dataset named BreaKHis, which contains histopathological images of breast tumors. This work also presents an alternative approach to classify these challenging images, avoiding any explicit segmentation. Such approach explores hand-crafted textural descriptors and automatic representation, particularly using Convolutional Neural Networks as well as the paradigm Multiple Instance Learning. The obtained experimental results have demonstrated the feasibility of this proposal, giving directions for improvement in such model.

Keywords: pattern recognition, breast cancer, histopathological image, deep learning, convolutional neural network, multiple instance learning, image processing, machine learning, computer-aided diagnosis.

List of Figures

List of Tables

List of Acronyms

k-NN	k-Nearest Neighbor
1-NN	1-Nearest Neighbor
ACR	American College of Radiology
ANOVA	Analysis of Variance
APR	Axis-Parallel Hyper Rectangle
AR	Association Rule
AUC	Area Under the ROC Curve
BC	breast cancer
BI-RADS	Breast Imaging-Reporting Data System
BlVC	Berkeley Vision and Learning Center
BR	Bloom-Richardson
CAD	Computer-Aided Detection
CADx	Computer-Aided Diagnosis
CC	Conventional Cytology
CCD	Charge-Coupled Device
CLBP	Completed Local Binary Pattern
CNB	Core Needle Biopsy
CNN	Convolutional Neural Network
CSC	Convolutional Sparse Coding
CSDCNN	Class structure-based Deep Convolutional Neural Network
CT	Contourlet Transform
DBT	Digital Breast Tomosynthesis
DCIS	Ductal Carcinoma *In Situ*
DD	Diverse Density
DDSM	Digital Database for Screening Mammography
DeCAF	Deep Convolutional Activation Feature
DII	Digital Infrared Imaging
DT	Decision Trees
EIT	Electrical impedance imaging
EM	Expectation Maximization
EM-DD	Expectation Maximization-Diverse Density
ER	Estrogen Receptor

ETD	Extralobular Terminal Duct
FC	Fully Connected
FDA	Food and Drug Administration
FDCT	Fast Discrete Curvelet Transform
FFDM	Full-Field Digital Mammogram
FN	False Negative
FNA	Fine Needle Aspiration
FNAC	Fine Needle Aspiration Citology
FNR	False Negative Rate
FOV	Field-of-View
FP	False Positive
FPR	False Positive Rate
FV	Fisher Vectors
GA	Genetic Algorithm
GBM	Glioblastoma Multiforme
GLCM	Gray-Level Co-occurrence Matrix
GONN	Genetically Optimized Neural Network
HE	Hematoxylin-Eosin
IARC	International Agency for Research on Cancer
IDC	Invasive Ductal Carcinoma
IHC	Immunohistochemistry
ILC	Invasive Lobular Carcinoma
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
INCA	*National Institute of Cancer José Alencar Gomes da Silva*
IRT	Infrared Thermograph
ITD	Intralobular Terminal Duct
JScSPM	Joint Sparse coding based SPM
KIRC	Clear Cell Kidney Carcinoma
LBC	Liquid-Based Cytology
LBP	Local Binary Pattern
LBP-TOP	Local Binary Pattern from Three Orthogonal Planes
LCIS	Lobular Carcinoma *In Situ*
LDA	Linear Discriminant Analysis
LPQ	Local Phase Quantization
LPQ-TOP	Local Phase Quantization from Three Orthogonal Planes
LS-SVM	Least Square Support Vector Machine

mBR	Modified Bloom-Richardson
MIAS	Mammographic Image Analysis Society
MIBB	Minimally Invasive Breast Biopsy
MIL	Multiple Instance Learning
MILCNN	Multiple Instance Learning Convolutional Neural Network
MLP	Multi-Layer Perceptron
MRI	Magnetic Resonance Imaging
MS	*Health Ministry*
NN	Neural Network
NOS	Not Otherwise Specified
NST	No Special Type
OCLBP	Opponent Color Local Binary Pattern
ORB	Oriented FAST and Rotated BRIEF
PAP	Papanicolaou Cytology
PEM	Positron Emission Mammography
PFTAS	Parameter-Free TAS
PgR	Progesterone Receptor
PNG	Portable Network Graphics
PNN	Probabilistic Neural Network
QDA	Quadratic Discriminant Analysis
RBF	Radial Basis Function
RBM	Restricted Boltzmann Machine
ReLU	Rectified Linear Unit
RF	Random Forest
RGB	Red-Green-Blue
ROC	Receiver Operating Characteristic
ROI	Region of Interest
RS	Rough Set
SBB	Stereotactic Breast Biopsy
SCG	Scaled Conjugate Gradient
ScSPM	Sparce coding based linear SPM
SGD	Stochastic Gradient Descent
SIFT	Scale Invariant Feature Transform
SIL	Single Instance Learning
SOB	Surgical (open) Biopsy
SPM	Spatial Pyramid Matching
SSVM	Sparse Support Vector Machine
STFT	Short-Term Fourier Transform
SURF	Speed-Up Robust Features
SUS	*Sistema Único de Saúde*
SVM	Support Vector Machine

t-SNE	t-Distributed Stochastic Neighbor Embedding
TAS	Threshold Adjacency Statistics
TCGA	Cancer Genome Atlas
TDLU	Terminal Ductal Lobular Unit
TN	True Negative
TP	True Positive
UCI	University of California at Irvine
UCSB-BB	UCSB Biosegmentation Benchmark
VABB	Vacuum Assisted Breast Biopsy
VLAD	Vector of Locally Aggregated Descriptors
WBCD	Wisconsin Breast Cancer Dataset
WCRF	World Cancer Research Fund
WHO	World Health Organization
WSI	Whole-Slide Imaging

SUMMARY

Chapter 1

Introduction

Despite remarkable current advances in diagnosis and treatment, cancer continues constituting a massive public health problem around the world. Factors such as population aging and adoption of unhealthy habits [84] — before restricted to industrialized countries — are now contributing to expand the incidence of this disease. There was a 20% increase in the last decade according to World Cancer Research Fund (WCRF) [286] and 27 million new cases of this disease are expected to occur until 2030 [37]. Considering all its types, cancer is the second most common cause of deaths recorded in developed countries and it is recently replacing cardiac diseases as the main cause of death in several western nations [123, 160]. There were 8.2 million deaths in 2012 [93], according to International Agency for Research on Cancer (IARC), which is part of the World Health Organization (WHO). Besides, WHO's projections estimate 17 million deaths due to cancer by 2030, and developing countries will be the ones mostly affected [93].

In Brazil, cancer is an extremely worrying health issue as well. *National Institute of Cancer José Alencar Gomes da Silva* (INCA)/ *Health Ministry* (MS) estimates 596 thousand new cases of cancer in 2016 [137], being skin cancer the most prevalent (182 thousand cases), followed by prostate cancer (68 thousand), breast cancer (57 thousand), intestine cancer (33 thousand), and lung cancer (27 thousand). Among all types of cancer, excluding skin cancer, breast cancer (BC) is the second most common among women. And the mortality rate of BC is very high when compared to the other types of cancer. According to IARC, while the cancer mortality rate increased by 8% in 2012, the mortality rate of BC was 14% in the same period [93]. Furthermore,

compared to other countries, Brazil shows higher diagnostic rates for breast and prostate cancer. BC, especially, presents 59 cases per 100 thousand inhabitants (while the worldwide average is about 43 per 100 thousand).

Even though cancer has a hight incidence in more developed regions, mortality rate is relatively higher in developing countries due to difficulties in achieving early diagnosis as well as other restrictions faced by poor women in accessing technical advances against this disease. Quoting Dr. Christopher Wild, director of IARC, "an urgent need in cancer control today is to develop effective and affordable approaches to the early detection, diagnosis, and treatment of BC among women living in less developed countries" [138, p.2].

Detection and diagnosis of BC can be achieved through non invasive methods and biopsy. Non invasive methods are basically imaging procedures: diagnostic mammograms (x-ray), Magnetic Resonance Imaging (MRI) of the breast, breast ultrasound (sonography) and thermography. Although the use of imaging procedures for cancer screening is widespread, biopsy is the only way to tell with confidence if cancer is really present. Among biopsy techniques, procedures such as Fine Needle Aspiration (FNA), Core Needle Biopsy (CNB), Vacuum Assisted Breast Biopsy (VABB) and Surgical (open) Biopsy (SOB) stand out. Biopsy procedures collect samples of cells or tissue. These samples must be fixed across a glass microscope slide for subsequent staining and microscopic examination. Then, histopathologists use the benefits of a wide range of colored dyes to obtain pretty useful information about the lesions and the tissue compositions. Diagnosis from a histopathology image remains the "gold standard" in diagnosing almost all types of cancer, including BC [238].

In this context, considering the cancer impact in public health, especially BC for the female population, as well as the urgency to provide supporting tools for pathologists, an automatic system is proposed to classify BC using digitized images from histological slides.

1.1 Motivation

Even in face of recent advances in the comprehension of the molecular biology of BC progression and the discovery of new related molecular markers, the histopathological analysis remains the most widely used method for BC diagnosis [164]. Regardless of the significant progress reached by diagnostic imaging technologies, the final BC diagnosis, including grading and staging, still continues being done by pathologists applying visual inspection of histological samples under a microscope. As part of the making diagnosis process, one fundamental strategy of the pathologist — when looking at a histologic slide — is based on pattern recognition [221]. Pattern recognition is the realization that the histologic picture conforms to a previously learned picture of the disease. In fact, current pathological diagnosis is based on the subjective opinion of pathologists (experts). This reality can be observed in many countries, Brazil included. Histopathological analysis is a highly specialized and time consuming task, extremely dependent on the experience of the pathologists and directly influenced by factors such as fatigue and a decrease in attention.

Recently, digitized tissue histopathology has become feasible to the application of computerized image analysis and machine learning techniques. Factors such as large increases in available computational power, cheaper storage devices and considerable improvements in image analysis algorithms have disseminated Computer-Aided Detection (CAD)/Computer-Aided Diagnosis (CADx) systems in the daily routine of pathology laboratories. These systems have emerged for disease detection, diagnosis, and prognosis prediction to complement the opinion of the human expert [113], the pathologist. Recent advances in image processing and machine learning techniques allow the building of CAD/CADx systems that can assist pathologists to be more productive, objective and consistent in the diagnosis. As pointed by Gurcan *et al.* [113], there is a pressing need for CAD/CADx systems to relieve the workload of pathologists by filtering obviously benign areas, so that the experts can focus on the more difficult-to-diagnose cases [75].

1.2 Challenges

Classification of histopathology images into distinct histopathology patterns, corresponding to non-cancerous or cancerous condition of the analyzed tissue, is often the primordial goal of the image analysis module, which is part of the computer-aided diagnosis systems for cancer. Thus, the main challenge of this work is the development of a reliable classification system capable of dealing with the inherent complexity of histopathological images. In order to accomplish this task, some relevant issues should be properly addressed:

- **Lack of comprehensive and public histological image databases**. By reviewing the literature, we noticed that most of the works on BC histopathology image analysis are carried out on small datasets, which, most of the time, are not shared with the scientific community.

- **Feature extraction**. Features derived from segmented nuclei and glands from histological tissue images are usually a prerequisite for the extraction of higher-level information regarding the state of the disease. However, segmenting this kind of image is not a trivial task. Besides, problems in the segmentation step generally compromise the overall performance of the system.

- **Variation in staining procedure**. Whether for visual analysis or computerized analysis, the tissue must be prepared, and staining is essential in order to highlight important cellular structures used in diagnosis. Pathologists routinely use this color information in their practice, and some works in the literature have also included color information in digital histopathology image analysis [253]. However, such systems have to deal with the appearance variability of stained histological sections, which is the result of inconsistent preparation of intra and inter-lab.

- **Confounder tissue patterns**. Automated classification of histopathology involves identification of multiple classes, including benign, cancerous, and confounder categories. The confounder tissue classes can often mimic and share attributes with both the diseased

and normal tissue classes, and can be particularly difficult to identify, both manually and through automated classifiers [83].

1.3 Hypothesis

Considering the aforementioned aspects, the present work establishes the following basic hypothesis: "**is it possible to create an automatic method to aid BC diagnosis using digitized image database from histological slides, avoiding explicit segmentation of these images?**". The primary purpose consists in offering an automatic system to aid pathologists in the BC classification task, resulting in faster, more objective and consistent diagnosis, thus minimizing the inter- and intra-expert variability.

1.4 Goals

The primary goal of this research is to propose an automatic system to aid pathologists in BC diagnosis. This automatic system must distinguish, in a microscopic level, breast lesions as malignant or benign, using digitized images from histopathological slides as input. Suspect lesions must be classified with high accuracy rates, thus minimizing the occurrence of false-negative predictions. Furthermore, the proposed system should be usable in real conditions, observed in current clinical routines. Recent works dealing with BC classification are focused on Whole-Slide Imaging (WSI) [82, 104, 151, 295, 296]. However, the broad adoption of this and other forms of digital pathology are still facing obstacles such as the high cost of implementing and operating the technology, insufficient productivity for high-volume clinical routines, intrinsic technology-related concerns, unsolved regulatory issues, as well as "cultural resistance" from the pathologists [89]. Therefore, we propose an image acquisition protocol based on conventional photomicroscopy systems, which is widely available and used by pathologists.

So as to accomplish the general goal, the following secondary goals must be satisfied:

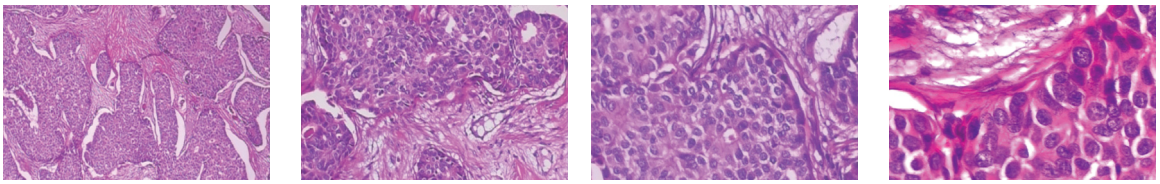
- To create and publicly share a database of images in high resolution — labeled by pathologists — from digitized histopathological slides of breast tumors;
- To analyze, test and/or propose frameworks to extract discriminant features suitable for BC classification;
- To assess classification strategies capable of recognizing similar patterns in unknown images, through the use of discriminative features extracted from image database;
- To validate the algorithms and the developed prototypes using tests from the image database, evaluating the performance in terms of false-acceptance and false-rejection.

1.5 Contributions

This work aims to offer some original scientific-technological contributions. Therefore, among the scientific-technological contributions, the following are highlighted:

- **Planning, collection, documenting and publishing of a histopathological image database related to breast tumors**. Unfortunately, there is a lack of comprehensive and public histopathological image databases intended for research in the CAD/CADx systems. In a recent review, Veta *et al.* [277] point out that the main obstacle in the development of new histopathology image analysis methods is the lack of large publicly available annotated datasets. Annotated databases are also crucial for development and validation of machine learning systems. In order to deal with the scarcity of public databases of histopathological images, this work introduces a novel BC histopathological images database called BreaKHis, which is composed of microscopic images from breast tumors surgical biopsy slides, stained with Hematoxylin-Eosin (HE). Images are divided into benign and malignant tumors, which have been collected at four different magnification factors: 40×, 100×, 200×, and 400×. A representative image of each magnification factor can be seen in Figure 1.1. This database has been built in collaboration with the P&D

Laboratory [1] – Pathological Anatomy and Cytopathology, Paraná, Brazil. The BreaKHis database is freely available from now on upon request for research purposes[2]. We will describe the dataset in more details in Appendix B. Until now, the dataset counts with more than 1070 registered users around the world. This distribution can be seen in the Figure 1.2. Published in [256].



(a) (b) (c) (d)

Figure 1.1: Image samples from the BreaKHis database. Distinct areas, belonging to the same slide of breast malignant tumor (stained with HE), seen in different magnification factors: (a) 40×, (b) 100×, (c) 200×, and (d) 400×.
Source: The author (2015).

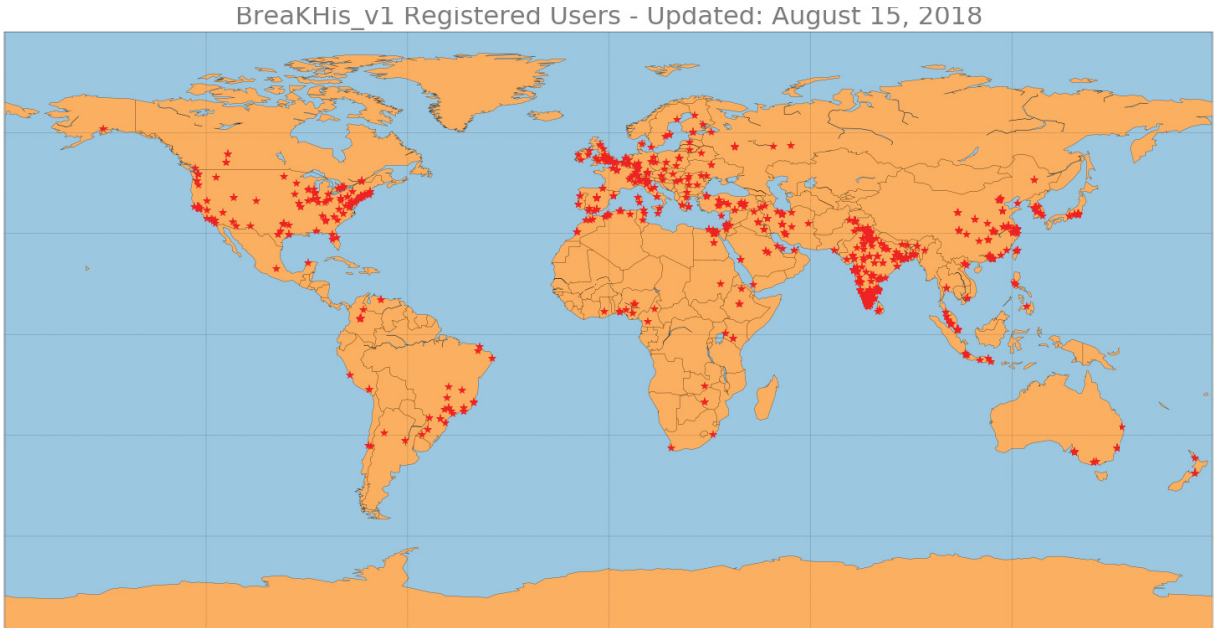


Figure 1.2: BreaKHis registered users worldwide. Each red star represents the location of a user.
Source: The author (2018).

[1] http://www.prevencaoediagnose.com.br/
[2] http://web.inf.ufpr.br/vri/breast-cancer-database

- **Development of an efficient and robust classification system for BC classification**. This hybrid system must perform automatic image analysis in HE breast tissue digitized images, collect quantitative information, identify pathological patterns of BC and provide correct differentiation between types of breast lesions. The system can work as a "second opinion" for medical experts, supporting mainly inexperienced pathologists, reducing workload and making the final diagnosis more objective. In addition, by extrapolating this initial application, the classifier system may, eventually, be applied to other histopathological classification problems in which there is a similarity with the images analyzed here. Published in [256, 257, 259].

- **Assessment of transfer learning approach applied to a real-world histopathological image dataset aiming to improve classification performance**. In another perspective, a deep learning approach combined with transfer learning should be evaluated in order to model high-level abstractions in the database. Published in [258].

- **Assessment of Multiple Instance Learning (MIL) approach applied to a real-world histopathological image dataset aiming to deal with inherent ambiguously labeled data of this dataset**. Submitted to the journal *Expert Systems With Applications (ESWA)* [262].

1.6 Document Outline

This document is composed of six chapters and two appendices.

In the current chapter, the relevance of researches to aid BC diagnosis was evidenced. BC, especially for the female population, is still associated to high incidence and mortality rates, making this disease a public health concern in many countries. The initial chapter also outlined BC diagnosis through histopathologic analysis, a task executed by human experts, but very time-consuming and prone to misinterpretation. Motivation to work with this problem, main goals and contributions of the research were presented as well.

In Chapter 2 we offer a pretty basic background to the reader, in aspects related to the terminology and procedures of pathology applied in BC diagnosis. This knowledge is essential for a proper understanding of the research, of our approach and of the achieved results. Aiming to provide the reader with a proper comprehension of the computational techniques used in this study, we also introduce the descriptors and classifiers selected in this Chapter, as well as the approaches based on deep learning and MIL. However, we have assumed that the reader is already familiar with the general theory related to image processing, machine learning and pattern recognition. Thus, this document does not show a more detailed revision about these topics. Additionally, the following references are recommended: [35, 108] for the subject of image processing, [31, 85, 273] for pattern recognition and [118, 243] for machine learning.

A review of the state-of-the-art for pattern recognition in BC classification is given in Chapter 3. Some main approaches in this field are detailed as well. In particular, the results achieved by different methods applied to our dataset BreaKHis are also detailed.

Chapter 4 contains our proposed alternative model along with the methodology that will be used to validate this model. We also present a brief description of our public dataset BreaKHis in this chapter.

Chapter 5 shows the experiments. Firstly, an experimental setup and protocol to evaluate the conventional image classification system applied on BreaKHis dataset is detailed. Chapter 5 also contains the experiments executed using an alternative deep learning approach with CNN and MIL. These results are discussed to establish a baseline to move forward with the research. The final considerations are presented in Chapter 6.

Finally, Appendix A presents a more detailed revision about BC and Appendix B describes the BreaKHis database in a comprehensive way.

Chapter 2

Fundamental Background

Firstly, for a proper understanding of the present research, this chapter provides some fundamental background knowledge about breast cancer (BC). A much more detailed revision can be found in Appendix A. After that, we briefly introduce the descriptors and classifiers used in the baseline experiments. Finally, this chapter also covers a short revision of the approaches based on deep learning as well as multiple instance learning.

2.1 Breast Cancer

Cancer is a common designation for several diseases characterized by the abnormal and accelerated replication of the cells [160]. This disordered cellular division is named *neoplasia* and produces a tissue mass called *neoplasm* or tumor. A tumor is not necessarly a cancer. Considering their potential aggressivity, tumors can be classified into benign or malignant. Normally, cells in benign tumors don't spread to other parts of the body and these tumors do not present a risk for the patient's life [238]. On the other hand, cells in malignant tumors can invade surrounding tissue, enter the circulatory system and/or lymphatic system, spread themselves to other parts of the body (*metastasis*) [161] and may eventually lead to the patient's death [238]. Cancer is a synonym for malignant tumor [238].

Generally, when a malignant tumor originates in connective tissues such as muscle, tendon, fat, and cartilage, it is named *sarcoma*. When a malignant tumor arises in organs or gland tissue, such as breast, cervix, liver, lung, etc., it receives the denomination of *carcinoma*. BC is a type of carcinoma (with various subtypes) [126], i.e., a malignant neoplasm that arises in the breast tissue. However, BC is not a single disease, but rather composed of several distinct subtypes associated with different origins, evolutions and possible therapeutic interventions [226]. Using histopathological analysis, BC can be classified into different tumor types and characterized in their specificities. Such classification is essential because it determines different prognoses, surgery planning and distinct therapies for each type.

Detection and diagnosis of BC can be achieved by imaging procedures, such as diagnostic mammograms (x-rays), magnetic resonance imaging, ultrasound (sonography), and thermography [144]. Imaging for cancer screening has been investigated for more than four decades [260]. However, biopsy is the only way to diagnose with confidence if cancer is really present. Among biopsy techniques, the most common are Fine Needle Aspiration (FNA), Core Needle Biopsy (CNB), Vacuum Assisted Breast Biopsy (VABB) and Surgical (open) Biopsy (SOB) [42]. The procedure consists in collecting samples of cells or tissues, which are fixed across a glass microscope slide (histological slide) for subsequent staining and microscopic examination. The slides are stained with various dyes in order to make the tissue components visible. The combination of two dyes – hematoxylin (deep blue-purple color) and eosin, known as HE – is one of the main stain protocols used in pathology and is essential for recognizing various tissue types and morphologic changes, which are key aspects of the contemporary cancer diagnosis [96]. Thus, diagnosis from a histopathology image is the “gold standard” in diagnosing almost all types of cancer, including BC [164, 238]. Staging and grading provide a general idea on how quickly a cancer may grow and which treatments may work best. The final BC diagnosis, including grading and staging, is done by pathologists applying visual inspection of histological samples under the microscope. Pathologists identify abnormalities by changes in the size, shape or arrangement of cells.

2.2 Feature Extraction

Humans have innate skills to process and recognize complex patterns, e.g., images. However, even experts are normally unable to explain exactly how they take certain decisions. By trying to incorporate the information used by a human expert in his/her analysis, feature sets, extracted directly from images, are generated. Our approach for classification is based on statistical model, where each pattern must be represented as d features (measurements) and is considered a point in a d-dimensional space. Ideally, features should be meaningful. In the data representation aspect, the main goal is to determine those features that mapped to pattern vectors occupy compact and disjoint regions in a d-dimensional feature space [140]. Regarding the classification aspect, or the decision making, the objective is to establish decision boundaries in the features space which separated patterns belonging to distinct classes. Thus, in the context of this work, features are numerical descriptors extracted directly from histopathological images. These sorts of images are naturally complex patterns, showing very high density of biological structures, color variations due to the staining technique, distortions, diverse illumination, artifacts and other noises from the acquisition device.

The goal is to summarize the appearance of the entire image in a single vector descriptor. Mainly, global features were applied, i.e., a function of the whole image. These features are quantifications describing how gray levels are distributed over the image pixels, i.e., an attempt to capture the image texture. Although there is no precise definition of "texture" concept, an image can be characterized in terms of its visual appearance as: coarse, fine, smooth, homogeneous, soft, gritty, etc. These properties are textural coefficients that describe the image and they are obtained by exploiting space relations underlying the gray-level distribution. Figure 2.1 presents real textures found in histopathological images (HE staining).

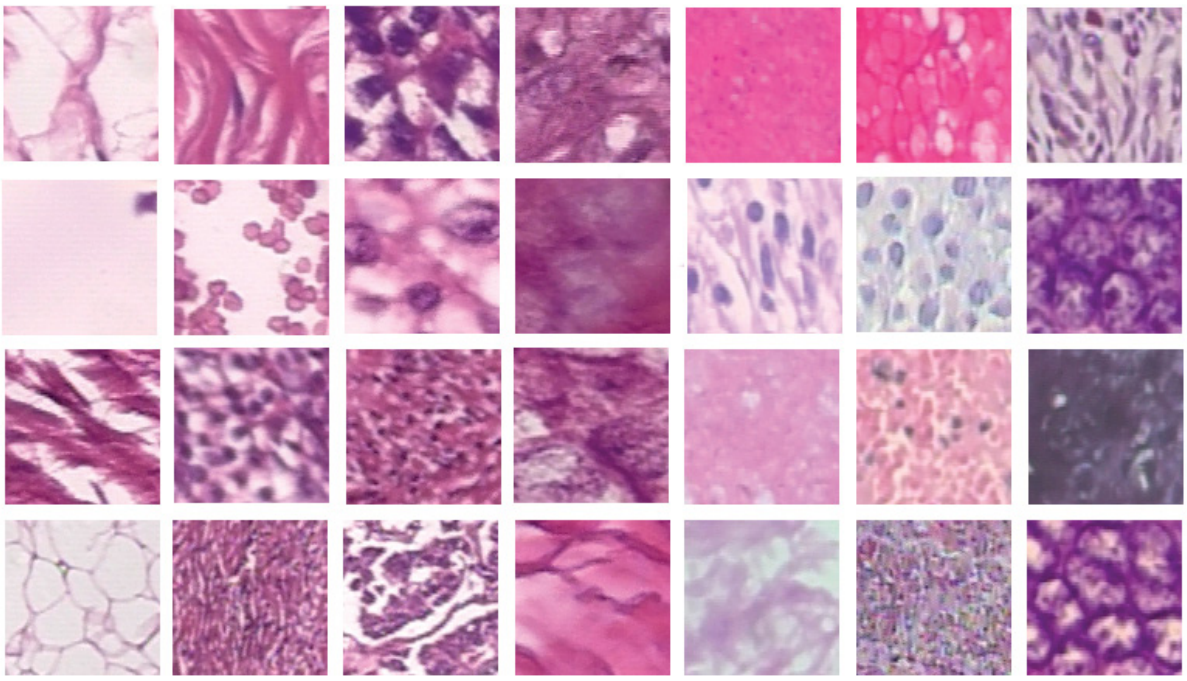


Figure 2.1: Examples of real textures present in histopathological images (HE staining).
Source: The author (2015).

The idea of using texture to describe the histopathological images is to have a simple, but powerful representation, so that we could establish some basis for further comparison. Besides, using texture representation allows us to avoid the explicit segmentation and extraction of structure properties such as cell nuclei size and shape, tubular formation, etc. However, these histological images present intrinsic heterogeneity, i.e., the tissue is a complex structure not a homogeneous texture. In Figure 2.2, three different tissues from the same ductal carcinoma image at 40× are highlighted. It is possible to note pretty distinguished textures.

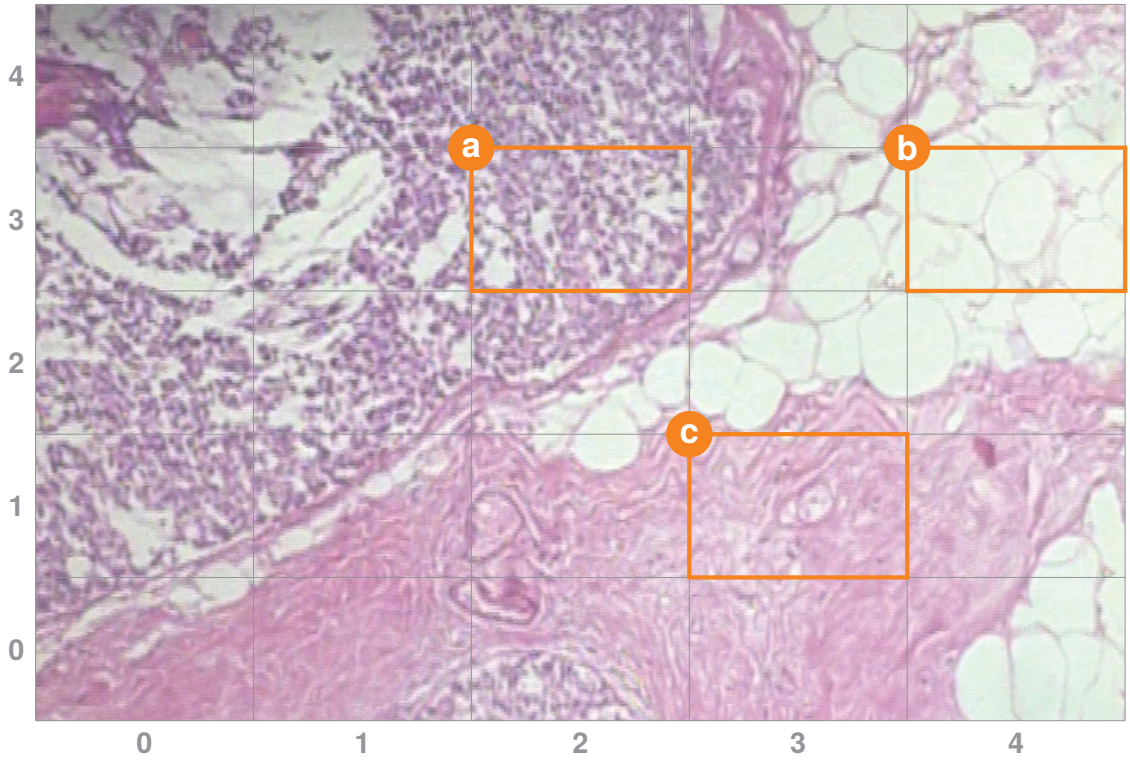


Figure 2.2: Image example from ductal carcinoma (HE staining) at 40×, showing distinct tissues: (a) cell nuclei, (b) fat tissue, and (c) collagen-rich stroma.
Source: The author (2015).

2.2.1 Descriptors

In this section, we briefly describe the six different operators used to extract texture descriptors.

2.2.1.1 GLCM

Gray-Level Co-occurrence Matrix (GLCM) is proposed by Haralick [117] in 1973, and it is widely used to characterize texture images. In our experiments, four adjacency directions 0^o, 45^o, 90^o, 135^o are used to compute the GLCM, and 8 gray levels. On the GLCM, 13 Haralick parameters are computed [117] shown in Table 2.1: angular second moment, contrast, correlation, sum of squares, variance, inverse difference moment, sum average, sum variance, sum entropy, entropy, difference variance, difference entropy, information measures of correlation 1, and

information measures of correlation 2. These individual features are computed using the GLCM (Gray-Level Co-occurrence Matrix) G, a symmetric matrix with dimension $N_g \times N_g$, where N_g is the number of gray levels in a particular image. Adjacency matrix G stores in position (i, j) the number of times that a pixel takes the value i next to a pixel with value j. Thus, each element (i, j) is considered the probability of a pixel of value i to be adjacent to a pixel of value j. Furthermore, given different ways to define operator $nextto$, it is possible to obtain slightly distinct variations of features. Finally, we obtain a 13-dimensional feature vector by averaging feature vectors obtained in the four directions.

Table 2.1: The thirteen co-occurrence features proposed by Haralick.

Feature	Description
f_1	*angular second moment*
f_2	*contrast*
f_3	*correlation*
f_4	*sum of squares: variance*
f_5	*inverse difference moment*
f_6	*sum average*
f_7	*sum variance*
f_8	*sum entropy*
f_9	*entropy*
f_{10}	*difference variance*
f_{11}	*difference entropy*
f_{12}	*information measures of correlation 1*
f_{13}	*information measures of correlation 2*

Source: [117].

2.2.1.2 LBP

Original Local Binary Pattern (LBP), proposed by Ojala *et al.* [207], in 1996 labels the pixels of an image by thresholding a 3×3 neighborhood of each pixel with the center value. Then, considering the results as a binary number and the 256-bin histogram of the LBP labels computed over a region, they used this LBP as a texture descriptor. The LPB operator $\text{LBP}_{P,R}$ produces 2^P different binary patterns that can be formed by the P pixels in the neighbor set on a circle of radius R. However, certain bins contain more information than others, and as a result, it is possible to use only a subset of the 2^P LBPs. These fundamental patterns are known as uniform patterns.

In 2002, LBP variants were proposed in [208]. LBP^{ri} and LBP^{riu2} have the same $\text{LBP}_{P,R}$ definition, but they only have 36 and 10 patterns, respectively. LBP^{ri} accumulates all binary patterns in only one bin, which keeps the same minimum decimal value $LBP^{ri}_{P,R}$, when their P bits are rotated. LBP^{riu2} combines the definitions of LBP^{u2} and LBP^{ri}. Thus, it uses only the uniform binary patterns and accumulates those that keep the same minimum decimal value $LBP^{riu2}_{P,R}$ in only one bin when their P bits are rotated.

As mentioned above, the LBP operator [208] consists in computing the distribution of binary patterns in the circular neighborhood of each pixel. These patterns are obtained by thresholding neighboring pixels compared to the central pixel. The neighborhood is characterized by a radius R and a number of neighbors P. The LBP code for a neighborhood is obtained by combining the thresholded neighborhood with powers of two and summing up the result. In our experiments, we chose a standard value of $P = 8$ neighbors, providing a 10-dimensional feature vector. It has been observed that a vast majority of patterns have a number of spatial transitions less than or equal to 2. Patterns with number of transitions superior to 2 are clustered into one single bin.

2.2.1.3 CLBP

One of the latest variants of LBP is Completed Local Binary Pattern (CLBP) [110]. It provides a complete modelling of the LBP, which is based on three components extracted from a local region: center pixel, sign and magnitude. The center pixel is simply coded by a binary code after global thresholding, with the threshold set as the average gray level of the whole image. For the two other components, a neighborhood of radius R and the number of neighbors P is considered, similarly to LBP. The difference signs and the difference magnitudes are then computed and coded by a specific operator into a binary format so that they can be readily combined to form the final CLBP histogram [110]. Note that the operator coding the sign component corresponds to the original LBP operator. Finally, the complete descriptor is composed of the center gray level (C), and of these two components derived from the local difference, sign (S) and magnitude (M).

Similar to LBP, for a given local region, CLBP calculates the difference between the center pixel g_c and each of its P circularly and evenly spaced neighbors g_p located in a circle of radius R, defined as $d_p = g_p - g_c$. However, here the difference d_p is decomposed into two components: sign, which is defined as $s_p = sign(d_p)$, and magnitude, which is defined as $m_p = |d_p|$.

The operator related to the component C , named $CLBP\_C$, is defined in Equation 2.1, where c_I corresponds to the average gray level of the whole image

$$CLBP\_C_{P,R} = t(g_c, c_I) \tag{2.1}$$

where

$$t(x, c_I) = \begin{cases} 1 & \text{if } x \geq c_I \\ 0 & \text{otherwise} \end{cases}$$

While the operator defined to code the component S ($CLBP\_S$) corresponds to the original LBP operator, the Equation 2.2 presents the operator to code the component M ($CLBP\_M$). In this case, the threshold c is originally defined as the mean value of m_p from the whole image.

$$CLBP\_M_{P,R} = \sum_{p=0}^{p-1} t(m_p, c)2^p \tag{2.2}$$

where

$$t(x, c) = \begin{cases} 1 & \text{if } x \geq c \\ 0 & \text{otherwise} \end{cases}$$

More details about the rotation invariant version of $CLBP\_M$ and different schemes to combine the histograms of codes provided by each operator $CLBP\_C$, $CLBP\_S$, and $CLBP\_M$ can be found in [110]. We have assessed different configurations suggested in [110] and the best

results observed in our experiments have been obtained with the combination of all components using a 3D joint histogram, while the best values for the parameters P and R, are 24 and 5, respectively, yielding a 1352-dimensional feature vector.

2.2.1.4 LPQ

Local Phase Quantization (LPQ) is a descriptor for texture classification robust to image blurring and also invariant to uniform illumination changes, according to its creators [121]. The descriptor was proposed by Ojansivu and Heikkilä in 2008 [209] and it is based on the Fourier phase spectrum. It utilizes the local phase information extracted by using the 2D DFT or, more precisely, a Short-Term Fourier Transform (STFT) computed over a rectangular $M \times M$ neighborhood N_x at each pixel position x of the image $f(x)$. The quantized coefficients are represented as integer values between 0-255, using binary coding described in [209]. These binary codes are generated and accumulated in a 256-bin histogram, similar to the LBP method [208]. The accumulated values in the histogram are used as the LPQ 256-dimensional feature vector.

In our experiments, a variant of LPQ, named LPQ-TOP [213], produced better results. Similar to the Local Binary Pattern from Three Orthogonal Planes (LBP-TOP) [297] for LBP, a variant of the LPQ, namely Local Phase Quantization from Three Orthogonal Planes (LPQ-TOP) has been proposed [213]. Actually, LPQ-TOP applies the original LPQ on the XY, XZ and YZ plans of dynamic images and concatenates the LPQ histograms, for a total of 768 elements. As we have static images, we only used the 256 elements for the XY plan. The main difference is that LPQ and LPQ-TOP use different default values for their parameters.

2.2.1.5 SURF and ORB

Speed-Up Robust Features (SURF) is a scale-rotation-invariant detector and descriptor proposed by Bay *et al.* [23, 24], following the same principles of his predecessor Scale Invariant Feature Transform (SIFT) [176]. But, SURF strikes a good balance between accuracy and speed.

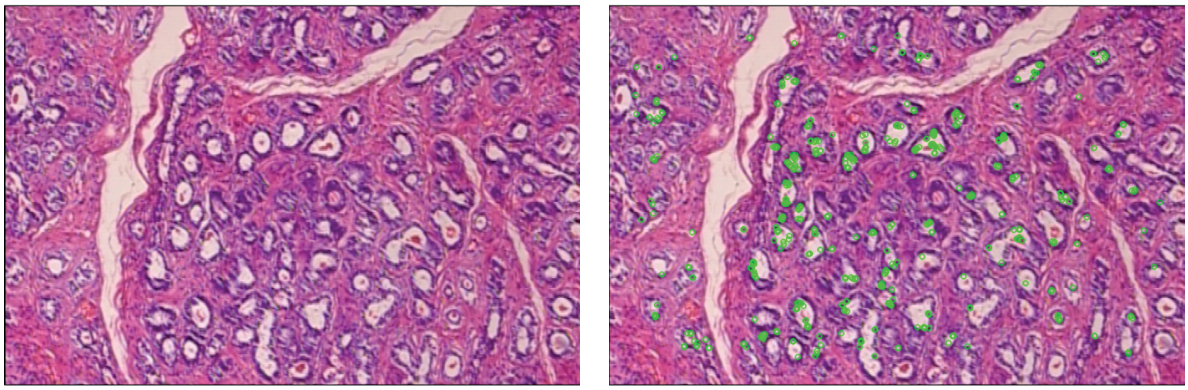
Local features are fast computed at automatically determined keypoints. A technique called multi-resolution transforms the input image into coordinates. There are two steps in the SURF algorithm: detection of interest points and description of the detected points. Each point is represented by the tuple $(y, x, \sigma, score, Laplacian, \theta, D_0, ..., D_{63})$, where y,x is the position; σ is the scale; θ is the angle of the orientation; $score$; $Laplacian$ is the sign of the detector; and D_i is the descriptor, a 64-dimensional array.

But, both SIFT and SURF require license fees for the usage of their original algorithm. Thus, Oriented FAST and Rotated BRIEF (ORB) [239] has been proposed as a free alternative to the traditional SIFT [176] and SURF [23] keypoint detectors in terms of computational cost and matching performance. It is designed to be rotation invariant and resistant to noise. Their authors claim that ORB is at two orders of magnitude faster than SIFT. The rationale behind the keypoint detectors is to find interesting points from an image, at which, statistics of local gradient directions of image intensities are accumulated to give a description of the local image structures in a local neighborhood around each interest point, with the intention that this descriptor should be used for matching corresponding interest points between different images.

ORB is based on the well-known FAST keypoint detector [237] and the BRIEF keypoint descriptor [49]. ORB works as follows: first it employs FAST to find keypoints, then applies the Harris corner measure to find top N points among them. Since FAST features do not have an orientation component, an efficiently computed orientation is added. This orientation compensation mechanism makes it rotation invariant. While BRIEF uses randomly chosen sampling pairs, ORB learns the optimal sampling pairs. In order to compute the orientation that is not provided by the FAST algorithm, ORB computes the intensity-weighted centroid of the patch with the located corner at the center. The direction of the vector from this corner point to centroid supplies the orientation. In order to improve the BRIEF descriptors against rotation, ORB moves them according to the orientation of the keypoints. For any feature set of n binary tests at location (x_i, y_i), it defines a $2 \times n$ matrix S, which contains the coordinates of these pixels. Then using the orientation of patch, θ, its rotation matrix is found and rotates S to get the rotated version S_θ. Then, ORB discretizes the angle to increments of $2\pi/30$ and constructs a lookup

table of precomputed BRIEF patterns. As long as the keypoint orientation θ is consistent across views, the correct set of points S_θ will be used to compute its descriptor.

In this work we have used the OpenCV implementation [38] with the default parameters, which returns a 32-dimensional vector for each keypoint. Best results have been achieved using 500 keypoints, considering a balance between runtime and improvement of recognition rate. At the end, the image is represented by a single 32-dimensional vector that contains the average of all keypoints. Keypoint detectors such as ORB, SIFT, and SURF, have been proposed initially for object matching/tracking. In these cases, the keypoints are used as features to perform the matching. When using these keypoint detectors as a texture descriptor, it is common practice to average them out in order to produce a single feature vector. Figure 2.3 exemplifies the keypoints detected by ORB in a benign tumor image. In this case, the keypoints highlight small terminal duct units that proliferate in a tubular adenoma.



(a) Original image of a tubular adenoma (benign tumor) HE stain at 40× magnification.

(b) ORB-detected keypoints.

Figure 2.3: Example of keypoints detected by ORB in a benign tumor image.
Source: The author (2015).

2.2.1.6 TAS and PFTAS

The Threshold Adjacency Statistics (TAS) method was presented by Hamilton *et al.* in 2007 [115], for cell phenotype image classification. The method first applies a threshold to the image so as to create a binary image. In order to define the threshold, first, the average intensity, μ, of those pixels with intensity of at least 30, is calculated. This value was defined empirically by

the authors in [115], once they observed that intensities below 30 are in the general background. Then, the experimental image is thresholded to the range $\mu - 30$ to $\mu + 30$. From each white pixel of this experimental image, nine adjacency statistics are computed according to Figure 2.4. The first threshold statistic is then the number of white pixels with no white neighbors; the second one is the number with one white neighbor, and so forth, up to the maximum of eight. Finally, the 9-dimensional feature vector is normalized by dividing each by the total number of white pixels in the threshold image. This process is repeated for two other sets of threshold: $\mu - 30$ to 255 and μ to 255. The final feature vector is the concatenation of three 9-dimensional vectors, giving a total of 27 statistics.

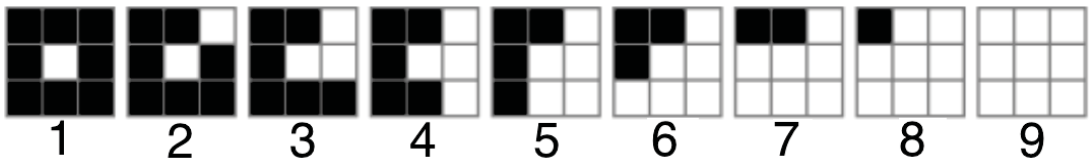
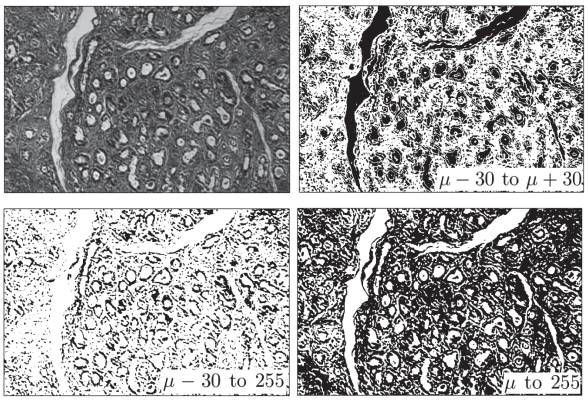


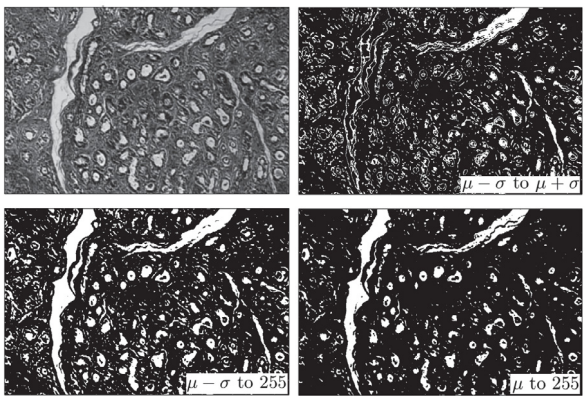
Figure 2.4: Nine threshold adjacency patterns in TAS. Each white pixel is evaluated regarding its neighbors.
Source: Adapted from [115].

The use of a fixed value to threshold the image makes the algorithm somehow problem-dependent. To overcome this difficulty, Coelho *et al.* [69] proposed using Otsu algorithm [211] to replace the fixed parameter. This method is known as Parameter-Free TAS (PFTAS). To illustrate the difference between TAS and PFTAS, Figure 2.5 shows the three resulting images of both methods. Considering that BC images in BreaKHis database share some similarities with those sorts of images for which TAS was originally thought and regarded to these descriptors, empirical observations demonstrated that recognition rates are better for color images when compared to grayscale images. Thus, we have used the PFTAS [69], which is the parameter-free version of TAS, in all three RBG channels. Its principle is to accumulate the pixels in the histogram bins, according to their number of white neighbors, in multiple-threshold binarized images. The original image is binarized using 3 different threshold ranges: $[\mu - \sigma, \mu + \sigma]$, $[\mu - \sigma, 255]$, and $[\mu, 255]$, where μ is an Otsu's defined threshold, and σ is the standard deviation of the above threshold pixels of the image. For each binarized image, a normalized histogram of pixels having i (i ranging from 0 to 8) white pixels as neighbors is computed. All 3 histograms

are concatenated to form a 27-dimensional feature vector for each one of three RGB channels, yielding an 81-dimensional feature vector. Finally, the computed image binarization 81-vector and its bitwise negated version (using bitwise not operator ~) are concatenated, resulting in a 162-dimensional feature vector. Figure 2.6 shows intermediate thresholds applied on the original RGB color image of a ductal carcinoma from BreaKHis dataset.



(a) TAS. From left to right, top to bottom: original image, binarized images using threshold ranges $[\mu - 30, \mu + 30]$, $[\mu - 30, 255]$, and $[\mu, 255]$.



(b) PFTAS. From left to right, top to bottom: original image converted to grayscale, binarized images using threshold ranges $[\mu - \sigma, \mu + \sigma]$, $[\mu - \sigma, 255]$, and $[\mu, 255]$.

Figure 2.5: Thresholds resulting from TAS and PFTAS, both applied on the same image converted to grayscale.
Source: The author (2015).

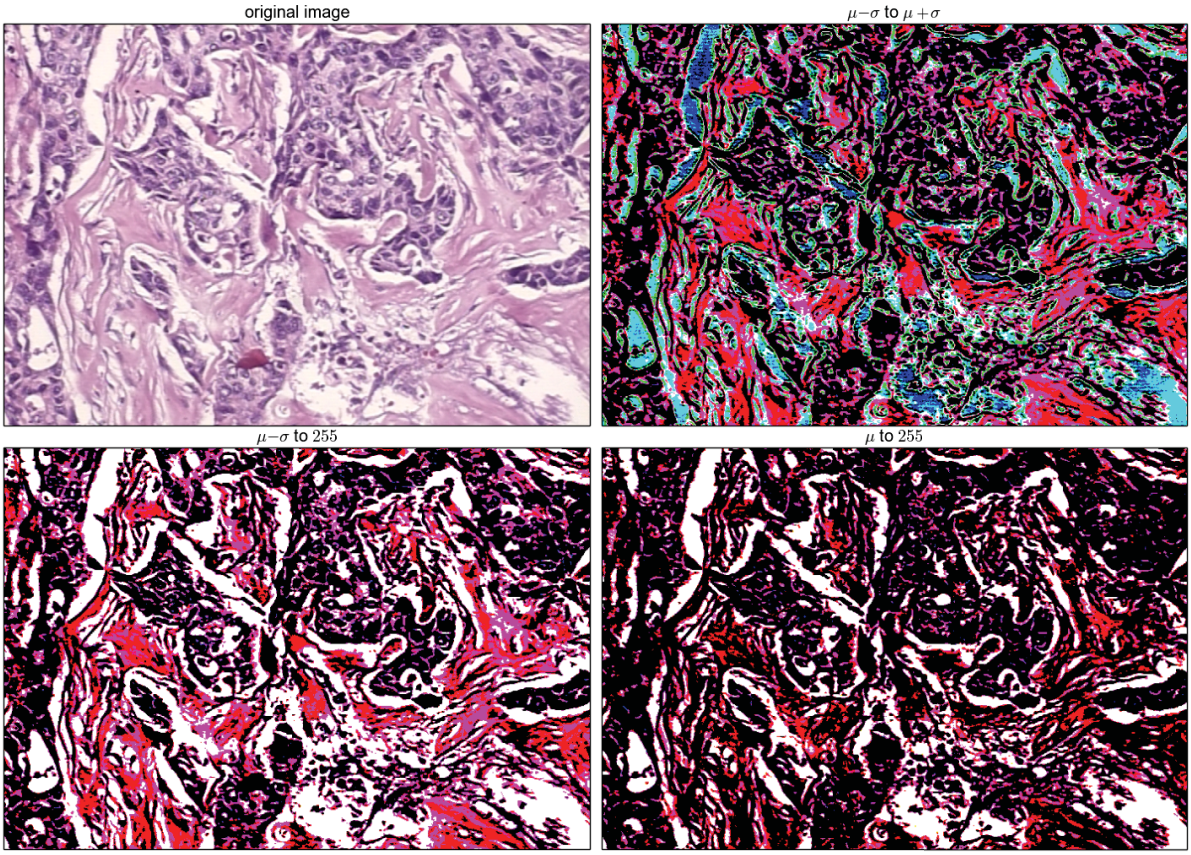


Figure 2.6: PFTAS thresholding on an image of a malignant tumor. From left to right, top to bottom: original image, binarized images using threshold ranges $[\mu - \sigma, \mu + \sigma]$, $[\mu - \sigma, 255]$, and $[\mu, 255]$.
Source: The author (2014).

2.3 Classifiers

Different classifiers were used to assess the aforementioned feature sets (Section 2.2). These classifiers will be presented in the following sections.

2.3.1 k-NN

A k-Nearest Neighbor (k-NN) is a type of instance-based learning that stores all available training data and classifies the testing samples based on a similarity measure (e.g., Euclidean

distance) [85]. The algorithm k-NN assumes that the data is a feature space, more specifically, the data points are represented in a metric space. The data can be scalars or multidimensional arrays. It does not use the training data points to do any generalization. Therefore, there is no explicit training phase and k-NN makes the decision based on the entire training data set (or a subset of them). However, the testing phase is very costly, in terms of time and memory. The arbitrary parameter k determines how many neighbors (points closer in terms of distance metrics) will be considered in the classification phase. When the number of classes is 2, k usually is an odd number. In particular, the 1-NN is often used to assess the discriminating power of the features.

2.3.2 QDA

Quadratic Discriminant Analysis (QDA) is closely related to Linear Discriminant Analysis (LDA), where it is assumed that the measurements of each class are normally distributed. LDA was formulated by Ronald A. Fisher in 1936 [97]. Although LDA is a linear transformation technique more commonly applied for dimensionality reduction, e.g., in the pre-processing phase in the pattern recognition and machine learning applications, it has a practical use as classifier too. LDA computes the directions ("linear discriminants") that will represent the axes that maximize the separation between multiple classes. The main goal is to project a feature space (a dataset n-dimensional samples) onto a smaller subspace k (where $k \leq n - 1$) while maintaining the class-discriminatory information. However, unlike LDA, in QDA there is no assumption that the covariance of each of the classes is identical [273]. So, QDA is a general discriminant function with a quadratic decision boundary, which can be used to classify datasets with two or more classes.

2.3.3 SVM

Support Vector Machine (SVM), a very popular classification algorithm based on statistical learning, builds a hyperplane in a high-dimensional space, which can be used for

classification and regression. Differently from other linear discriminant functions, it provides the optimal hyperplane that separates two classes of data with as wide a margin as possible [70]. This leads to good generalization accuracy on unseen data, and supports specialized optimization methods that allow SVM to learn from a large amount of data. We use a SVM with RBF kernel. Two parameters are required to train the RBF kernel, γ the coefficient of the exponent, and C the penalty term of the error. A grid search was performed to choose those values of γ and C.

2.3.4 Random Forest

Random Forest (RF) is an ensemble approach that combines Decision Trees (DT) predictors. It is a hybrid of the bagging algorithm and the random subspace method, and uses DT as the base classifier. The principle behind ensemble methods is that a group of weak learners (in this case the decision trees) can come together to form a strong learner [46]. One of the advantages of the RF is that it is quite fast and able to deal with unbalanced data. RF is a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest. The generalization error for forests converges to a limit as the number of trees in the forest becomes large [46]. Each tree is constructed from a bootstrap sample of the original dataset. An important point is that the trees are not subjected to pruning after construction, enabling them to be partially overfitted to their own data sample. In order to further diversify the classifiers, at each branch in the tree, the decision of which feature to split is restricted to a random subset of size n, from the full feature set. The random subset is chosen anew for each branching point. n is suggested to be $log_2(N + 1)$, where N is the size of the whole feature set.

A DT is a tree-structured classification model, which is easy to understand, even by non-expert users, and can be efficiently induced from data. The induction of decision trees is one of the oldest and most popular techniques for learning discriminatory models, which has been developed independently in the statistical [45] and machine learning [132, 229] communities. DT learning uses a decision tree as a predictive model which maps observations about an item for conclusions about the item's target value. Tree models where the target variable can take a

finite set of values are called classification trees. In these tree structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels. Decision trees where the target variable can take continuous values (typically real numbers) are called regression trees.

2.4 Deep Learning and CNNs

The performance of most conventional classification systems relies on appropriate data representation and much of the effort is dedicated to *feature engineering*, a difficult and time-consuming process that uses prior expert domain knowledge of the data to create useful features [29]. In this sense, the descriptors presented in Section 2.2.1 are hand-crafted features by feature engineering. On the other hand, *feature learning* can directly learn data representations from raw input data (training set) and detect data-driven features, i.e., learning representations and transformations of the raw data to facilitate the extraction of useful information from it [28, 167]. In contrast to feature engineering, feature learning can extract and organize the discriminative information from the data, not requiring the design of feature extractors by a domain expert.

Among the representation learning techniques, deep learning can yield more non-linear and more abstract representations, using an architecture formed by the composition of multiple levels of representation [27, 28]. Essentially, the deep learning approach proposes learning models with multilayer representations of the data with multiple levels of abstraction [167]. Each layer can be seen as a "distributed representation" of the inputs, where representation is a vector of units to extract information about specific features. Each unit in a layer is a separate feature of input and these units are not mutually exclusive. Considering a hierarchical view of this architecture, higher layers (levels) are built on top of lower layers. The lowest level is related to the original raw input and the subsequent higher levels represent more abstract concepts, ranging from lower abstract representation to higher abstract representations [27]. The overall composition gives a deep, layered model, in which each layer encodes progress from the low-level input details, such as image pixels and audio waveform, to high-level output concepts, such as

the identification of a face or a spoken word. In summary, deep learning allows to build complex concepts by using the combination stacked of simpler concepts [29].

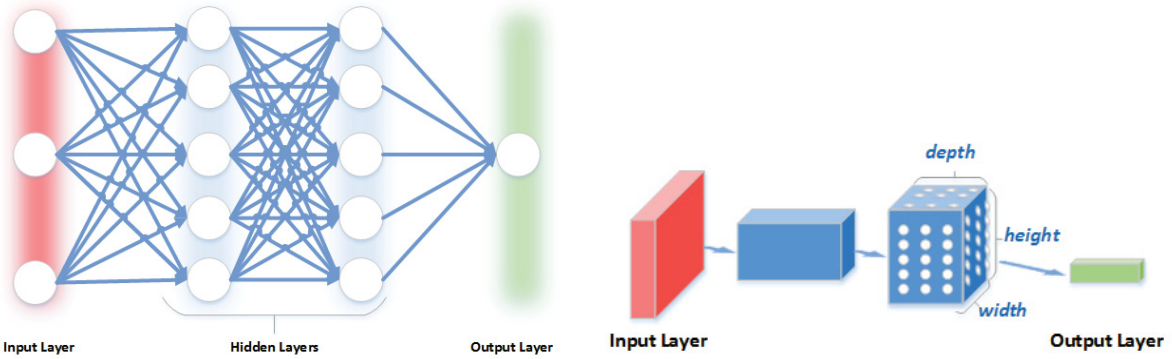
A deep net is trained by feeding it the input and letting it compute layer-by-layer in order to generate output for comparison with the correct answer. After computing the error at the output, the error flows backwards through the net by back-propagation. At each step backwards, the model parameters are tuned in a direction that tries to reduce the error. This process sweeps over the data, thus improving the model as it goes. Typically, training is an iterative process that involves multiple passes of the input data until the model converges.

In this context, Convolutional Neural Networks (CNNs) are a particular type of deep, feedforward network which have gained attention from the research community and industry, achieving empirical success in tasks such as speech recognition, signal processing, object recognition, natural language processing and transfer learning. In particular, CNNs have achieved success in image classification problems, including medical image analysis [67, 71, 227, 289]. CNN are biologically-inspired variants of Multi-Layer Perceptrons (MLPs), very similar to regular Neural Networks (NNs): they are made up of neurons that have learnable weights and biases. Each neuron receives some inputs, performs a dot product and optionally follows it with a non-linearity. The whole network still expresses a single differentiable score function: from the input signal (e.g., raw image pixels) on one end to class scores at the other end. And they still have a loss function (e.g. SVM/Softmax) on the last (fully-connected) layer. In summary, a CNN consists of multiple trainable stages stacked on top of each other, followed by a supervised classifier, and sets of arrays named *feature maps* represent both input and output of each stage [168]. Input can be signals such as image, audio, and video. For example, considering color images, at the input each feature map is a 2D array storing a color channel of the input image. The output consists of a set of arrays, where each feature map represents a particular feature extracted at locations of the associated input.

2.4.1 CNN Architecture

The simplest example of a deep learning model is the feedforward deep network or MLP [29]. It consists of a collection of neurons connected in an acyclic graph organized into a succession of neuron layers. A neuron computes a function on inputs from the preceding layer and passes the result (neuron's activation) on to outputs in the succeeding layer. Within each layer, all neurons compute the same function, but individual neurons may have distinct sets of inputs and outputs and may assign different weights to their inputs. The MLP receive as input a single vector in its input layer. Then, this input is transformed thorough a sequence of intermediary layers (hidden layers). In regular NN each hidden layer is fully-connected, i.e., each neuron compounding the layer is fully connected to all the neurons from the preceding layer, but neurons within a layer are not connected. The last fully-connected layer, named output layer, represents the class scores considering classification tasks. An example of this architecture is shown in Figure 2.7(a).

However, an otherwise arrangement used in CNNs is 3D volumes of neurons. The layers of a CNN have neurons arranged in 3 dimensions: width, height, and depth. Every CNN layer transforms an input 3D volume to a correspondent output 3D volume with some differentiable function. For instance, considering as input patches of dimensions 32×32 pixels extracted from the original Red-Green-Blue (RGB) image available in BreaKHis dataset (see Chapter 4, Section 4.2). For each patch, the expected output is the probability of it belonging to the benign or malignant class. Thus, the input is a volume of activations $32 \times 32 \times 3$ ($width \times height \times depth$) that holds the image, where the width (32) and the height (32) are the image dimensions, and the depth (3) encodes the three channels (Red, Green, and Blue format). The final output layer would have dimensions $1 \times 1 \times 2$, because the CNN will reduce the original image into a single vector of class scores. This general processing path can be seen in Figure 2.7(b): the depicted net transforms (intermediate blue volumes) the original image (input red volume) layer-by-layer from the original pixel values to the final class scores (output green volume).



(a) A regular 3-layer Neural Network with two fully connected hidden layers.

(b) CNN highlighting the 3-dimensional (width, height, depth) arrangement of the neurons in one of the layers.

Figure 2.7: Distribution of neurons in a regular Neural Network compared to the typical arrangement found in CNNs.
Source: The author (2015).

2.4.1.1 CNN Layers

There are three main types of layers used to build CNN architectures: *convolutional layer*, *pooling layer*, and *fully-connected layer*. Normally, a full CNN architecture is obtained by stacking several of these layers. An example of typical CNN architecture with two feature stages is shown in Figure 2.8.

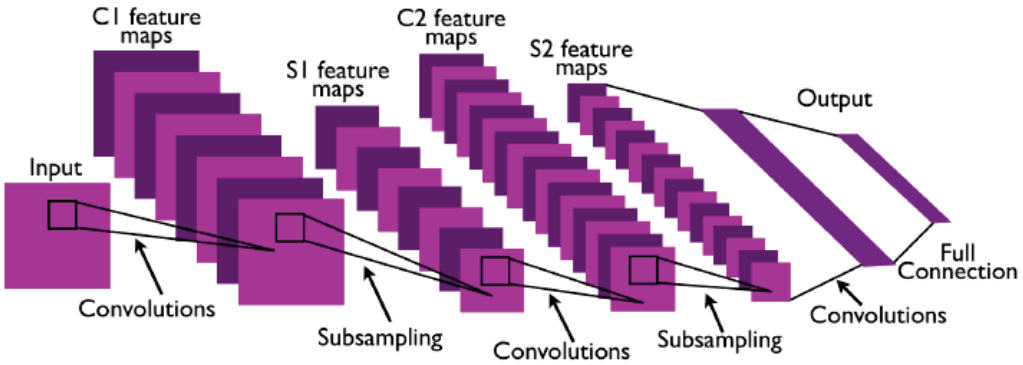


Figure 2.8: An example of a typical CNN architecture with two feature stages.
Source: Extracted from [168].

In a CNN, the key computation is the convolution of a feature detector with an input signal. Convolutional layer computes the output of neurons connected to local regions in the input, each one computing a dot product between their weights and the region they are connected to in the input volume. The set of weights which is convolved with the input is called **filter** or **kernel**. Every filter is spatially small (width and height), but extends through the full depth of

the input volume. For inputs such as images, typical filters are small areas (e. g., 3×3, 5×5, or 8×8) and each neuron is only connected to this area in the previous layer. The weights are shared across neurons, leading the filters to learn frequent patterns that occur in any part of the image. The distance between the applications of filters are called **stride**. If the stride hyperparameter is smaller than the filter size, the convolution is applied in overlapping windows.

Convolution with a collection of filters, like the learned filters (also named feature maps or activation maps) in Figure 2.9, improves the representation: at the first layer of a CNN the features go from individual pixels to simple primitives, like horizontal and vertical lines, circles, and patches of color. We can see that the model learns the filters for horizontal and vertical edges, and also learns the filters that resemble Gabor filters (edge detectors) [36, 98]. In contrast to conventional single-channel image processing filters, these CNN filters are computed across all of the input channels. Due to its translation-invariant property, convolutional filters yield a high response wherever a feature is detected.

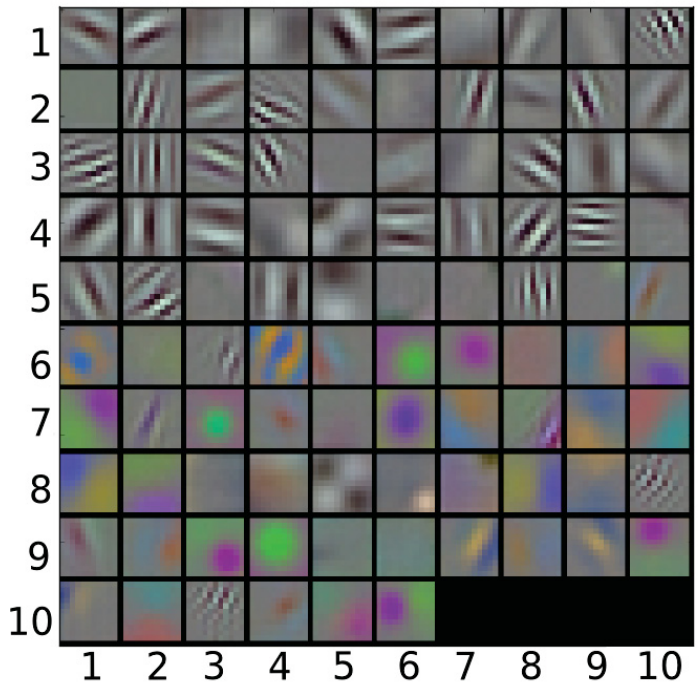


Figure 2.9: The first layer of learned convolutional filters in CaffeNet. This Caffe [141] reference model is based on AlexNet [158]. These filters are tuned to edges of different orientations, frequency, phase and colors. The filter outputs expand the dimensionality of the visual representation from the three color channels of the image to these 96 primitives. Deeper layers further enrich the representation.
Source: The author (2015).

When different features can be learned in different spatial locations the weights sharing scheme of the convolutional layer is relaxed and this layer is then referred to as *locally-connected layer*.

The insertion of a pooling (subsampling) layer between two successive convolutional layers is common. The main objective of this practice is to progressively reduce the spatial size of the representation. Thus, reducing the number of parameters and computations required by the network helps the overfitting control. The pooling layer downsamples the volume spatially in each depth slice of the input volume independently. Thus, the pool operator resizes the input along the width and the height, discarding the activations. In practice, the *max pooling* function, which applies a window function to the input patch, and computes the maximum in that neighborhood, has shown better results [248]. Furthermore, the pooling units can perform other functions like *L2-norm pooling* or *average pooling*.

In a fully-connected layer, neurons have full connections to all activations in the previous layer and their activations can be computed using a matrix multiplication followed by a bias offset. This type of layer is standard in a regular NN. The last fully-connected layer holds the net output, such as probability distributions over classes [90, 158].

2.5 Multiple Instance Learning

Supervised learning is a subfield of machine learning, where a predictive function is inferred from a set of labeled training examples, in order to map each input instance to its output label. In a conventional setting, the training dataset consists of instances equipped with their corresponding labels. While instances are relatively easy to obtain, the expensive data-labeling process with human-based ground-truth descriptions remains the major bottleneck in order to obtain large-scale datasets. This issue gives rise to a novel paradigm in machine learning, with the so-called weakly supervised learning, namely when having a partially-labeled training dataset [302].

Multiple Instance Learning (MIL) provides an elegant framework to deal with weakly supervised learning, modeling incomplete knowledge of labels in the training set. In comparison with strong (i.e., fully-labeled) supervised learning, where every training instance is assigned a discrete or real-valued label, the rationale of the MIL paradigm is that instances are naturally grouped in labeled bags, without the need for all the instances of each bag to have individual labels. In the binary classification case, a bag is labeled positive if it has at least one positive instance; on the other hand, a bag is labeled negative if all of its instances are negative [100]. With such training data grouped in labeled bags, MIL algorithms seek to classify either unseen bags (i.e., bag-level classification) or unseen instances (i.e., instance-level classification). Figure 2.10 shows an example of a problem naturally formulated as a MIL application: image classification task based on its visual content. Considering that the target class is "beach", for example, the positive images are those displaying a beach, while the negative images will be those displaying any other type of visual content. In Figure 2.10, the images in the top row are positive (beach images), while the images in the bottom row are negative (non-beach images). We can notice that there are regions of the image that are related with the target class (the regions that belong to the sand and sea), whereas there are regions that are not specifically related to it (e.g., the sky, umbrella, chairs, mountain, palm trees, etc.).

Basically, in order to obtain a beach image (positive bag), the simultaneous occurrence of sea and sand regions (positive instances) is necessary, while the rest of the regions (negative instances) are not necessary in order to make the image positive, even though such regions can be present as well. On the other hand, a non-beach image (negative bag) contains only other regions (negative instances) or even an insufficient number of regions, e.g., a desert containing sand is not a beach. As a result, each image can be described as a bag, i.e., an unordered set $\mathcal{X} = \{\vec{x}_1, \ldots, \vec{x}_N\}$, where the elements $\vec{x}_i$ are instances in a d-dimensional *instance space*, $\vec{x}_i \in \mathbb{R}^d$, and N can vary across the bags. Some MIL assumptions related to bags and its instances are presented in the following section.

While the multiple instance paradigm arose in many domains prior to the 1990's, MIL was first described explicitly and studied by Dietterich *et al.* in 1997 [78]. The original

Figure 2.10: Each image is a labeled bag $\mathcal{X}_i$ and the red ellipses represent instances $\vec{x}_j$. Beach images (positive bags) are depicted in the top row and non-beach images (negative bags) are in the bottom row. Notice that some instances are in both positive and negative bags.
Source: The author (2017).

motivation in MIL was the drug activity prediction, where experts provide activity labels for bags of molecules, labeling each individual molecule being costly and hard to set up. MIL is central in many applications in various domains, such as in bioinformatics, text processing, computer vision and image processing, to name a few [122]. Indeed, in many applications, ground-truth labeling is expensive in general and instances can be often grouped in bags, where each bag has a set of partially-labeled instances. An example is facial recognition, where several images of the same person, taken from different angles, can be considered as instances in a bag (the bag being the person) [122]. Note that only the MIL paradigm can apprehend these types of situations. Of particular interest is the image-based pathology classification for medical decision making, since it is relatively easy and part of the clinical protocol to take many images of some organs or tissues (physiology) under study; on the other hand, labeling each image is a time-consuming process dominated by human effort. MIL has many applications in medical imaging, as shown in a recent review [228].

2.5.1 MIL Methods

Under the standard MIL assumption, positive bags contain at least one positive instance, while negative bags contain only negative instances.

We denote by L_B the label of a bag B, defined as a set of instances, each one described by its feature vector: $B = \{b_1, b_2, ..., b_N\}$. We denote by l_k the label of each instance b_k. We can now define the label of a bag, following the standard MIL assumption:

$$L_B = \begin{cases} +1 & \text{if } \exists l_k : l_k = +1; \\ -1 & \text{if } \forall l_k : l_k = -1. \end{cases} \quad (1) \quad (2.3)$$

There are other – more relaxed – assumptions, such as a bag is labeled positive when it contains a sufficient number of positive instances; we refer the reader to [100] for further reading.

MIL methods are usually divided into two groups, depending on how they exploit the information in the data [10]. The first group consists of methods that consider the discriminative information at the instance level.

Learning algorithms do not focus at the larger scale of a bag, but at the local scale of instances. An advantage of these methods is that they are able to classify instances, when needed. However, they require that instances have a precise label, a requirement not all MIL problems meet. The instance level methods include APR, DD, SVM based approaches. The second group consists of the methods that consider the discriminative information to be at the bag level. These methods are usually more accurate, since they can model the distribution of each class and the relations between classes [53]. However, they cannot classify single instances, but only bags. An example of such methods is Citation-kNN [280]. For a review on MIL methods, we refer the reader to [10, 53, 122].

In the following subsections, we briefly describe the well-established MIL methods that have been implemented and applied to the BreaKHis dataset.

2.5.1.1 Axis-Parallel Hyper Rectangle (APR)

The MIL paradigm was first introduced in the seminal work of Dietterich *et al.* [78], motivated mainly by an application in biochemistry. The goal was to predict whether a molecule would be binding to a given receptor or not. Each molecule, which can be considered as a bag, can take many different spatial conformations, namely the instances. The methodology to solve the MIL problem is to design a hyper rectangle (called Axis-Parallel Hyper Rectangle (APR)) in the feature space aimed at containing at least one positive instance from each positive bag while excluding all the instances from negative bags. A molecule is classified as positive (resp. negative) if one (resp. none) of its instances belongs inside the APR.

2.5.1.2 Diverse Density (DD) and Its Variants

Diverse Density (DD) [183] is closely related to the idea of the APR. The DD defines a function over the feature space, such that it is high at points that are both close to instances from positive bags, and far away from instances which are in negative bags. The DD algorithm attempts to find the local maxima of this function (called the positive instance targets or prototypes) by maximizing diverse density (i.e. conditional likelihood) over the instance space, using gradient ascent with multiple starting points. Usually the search is repeated using the instances from every positive bag as the starting points. One of DD's shortcomes is that it is sensitive to noisy negative instance.

The DD approach has given rise to many variants. The most known is the Expectation Maximization-Diverse Density (EM-DD) method [294]. In this variant, the DD measure is maximized iteratively with the Expectation Maximization (EM) algorithm.

A probabilistic derivation of DD is given below. We denote positive bags as B_i^+, and the jth instance in that bag as B_{ij}^+. Let's suppose each instance can be represented by a feature vector (or a point in the feature space), and we use B_{ijk}^+ to denote the value of the kth feature of instance B_{ij}^+. Likewise, B_i^- denotes a negative bag and B_{ij}^- is the jth instance in that bag. The true

concept is a single point t defined by maximizing the diverse density defined as $DD(t) = P(t \mid B_1^+, ..., B_n^+, B_1^-, ..., B_m^-)$ over the feature space. Using Bayes rule and assuming uniform prior over the concept location, this is equivalent to maximizing the following likelihood:

$$\arg\max_x P(B_1^+, ..., B_n^+, B_1^-, ..., B_m^- \mid t)$$

By making additional assumption that the bags are conditionally independent given the concept point t, this decomposes to:

$$\arg\max_x \prod_i P(B_i^+ \mid t) \prod_i P(B_i^- \mid t)$$

Using Bayes rule once more with the uniform prior assumption, this is equivalent to:

$$\arg\max_x \prod_i P(t \mid B_i^+) \prod_i P(t \mid B_i^-)$$

which gives a general definition of the Diverse Density. Given the fact that boolean label (say, 1 and 0) of a bag is the result of "logical-OR" of the labels of its instances, $P(t \mid B_i)$ is instantiated using the noise-or model:

$$P(t \mid B_i^+) = 1 - \prod_j 1 - P(B_{ij}^+)$$

$$P(t \mid B_i^-) = 1 - \prod_j 1 - P(B_{ij}^-)$$

Finally, $P(t \mid B_{ij}^+)$ (or $P(t \mid B_{ij}^-)$) is estimated (though not necessarily) by a Gaussian-like distribution. Without close-form solution to the above maximization problem, gradient ascent method is used to search the feature space for the concept point with (local) maximum DD. Usually the search is repeated using the instances from every positive bag as the starting points.

2.5.1.3 Citation-kNN

The Citation-kNN, an adaptation of the k-NN algorithm, is the first non-parametric approach [280]. The principle is to first apply the k-NN algorithm to bags, where the distance

between bags is measured with the minimum Hausdorff distance. The latter is defined as the shortest distance between any two instances from each bag:

$$\text{Dist}(A, B) = \min_{a_i \in A} \min_{b_j \in B} ||a_i - b_j||$$

for any two bags A and B, where a_i and b_j are instances from each bag. This distance is used by a k-NN to classify a new bag, as the regular k-NN approach. The citation-kNN method adds a final step that makes the process more robust: in addition to the neighbors of the bag to be classified, other bags, called *citers*, are also considered in the classification rule.

2.5.1.4 mi-SVM and MI-SVM

Two alternative generalizations of the maximum margin idea used in SVM classification have been proposed in [12]. On one hand, the mi-SVM is based on the instance-level paradigm. Since the instance labels are not known, they are treated as hidden variables subjected to constraints defined by their bag labels. The mi-SVM method attempts to recover the instance labels and, at the same time, find the optimal discriminant function.

On the other hand, the bag-level paradigm is adopted by the MI-SVM. Its goal is to maximize the bag margin, defined between the positive instances of the positive bags, and the negative instances of the negative bags. In this setting, the bag is not represented by all of its instances, but only by the "extreme" ones, in the same sense as *support vectors* in conventional SVM. Moreover, mi-SVM and MI-SVM inherit also the kernel trick, thus allowing to use linear, polynomial and RBF kernels.

2.5.1.5 Non-parametric MIL

This recent technique is designed as a modified version of the k-NN classifier [276]. The non-parametric MIL approach employs a new formulation based on distances to k-nearest neighbors. The idea is to parse the MIL feature space with a Parzen window technique, using

different sized regions. Conversely to the majority vote used in k-NN, the vote contributions are the kernelized distances in the feature space. Non-parametric MIL has shown enhanced robustness to labeling noise on various datasets as well as competitive results compared to the state-of-the-art.

2.5.1.6 MILCNN

Deep learning networks have been overwhelming machine learning, pattern recognition and computer vision fields for a few years. MIL is no exception to this rule [127, 152, 217, 263, 281, 300]. Sun *et al.* proposed a Multiple Instance Learning Convolutional Neural Network (MILCNN) in [263]. This framework was initially designed for the data augmentation problem: in object detection, labels are not always preserved when the images are split for data augmentation. All the images produced through the data augmentation process share the same label. Thus, the proposed method considers data augmentation generated images as a bag, by combining a CNN with a specific MIL loss function derived with respect to the bag.

MILCNN is a combination of deep residual network and multiple instance learning loss layer. It is like a typical CNN containing convolutional layers, batch normalization layers, residual network layers, rectified linear unit, fully-connected layers, pooling layers and multiple learning loss function.

Chapter 3

State-of-the-Art Review

This chapter reviews the literature related to pattern recognition applied to automatic breast cancer (BC) classification, specially computer-aided detection/diagnosis (CAD/CADx) using histopathological image processing. The present work is focused on digitized images from the breast biopsy, in particular, surgical (open) biopsy (see Appendix 2, Section A.3.2.3). Therefore, this state-of-the-art analysis will be limited to those works reporting classification experiments on image databases digitized from the breast biopsy.

However, it is difficult to establish a "ground truth" in order to compare the different methods for BC classification, particularly with histopathology image analysis methods. Despite the significant increase of research in the field of histopathology focusing on automatic cancer diagnosis and classification, many of the works available in literature relate results of experiments performed using the own researchers' private datasets, distinct evaluation methods and performance metrics [139]. In short, there is a scarcity of annotated image datasets publicly available, as well as, unified benchmarks. In this sense, the recent availability of the BreaKHis dataset [256] constitutes a relevant unified benchmark to compare distinct approaches. Thus, some works presenting results of BC classification in unavailable datasets are briefly discussed in the Section 3.2.1.1 and a more detailed review focusing on BreaKHis dataset related works is given in Section 3.1.

3.1 Categories of Breast Cancer Classification Works

Considering the feature representation, automatic BC classification related works can be grouped into two main categories. In the first group, those works that automatically process images in order to acquire the feature descriptors are included. In this case, image sources can be mammograms, breast ultrasound, magnetic resonance breast imaging, breast thermograms and direct digitization of biopsies in microscope slides. The second group is compounded by studies using previously extracted feature descriptors, such as the well known Wisconsin Breast Cancer Dataset (WBCD). Figure 3.1 summarizes this categorization.

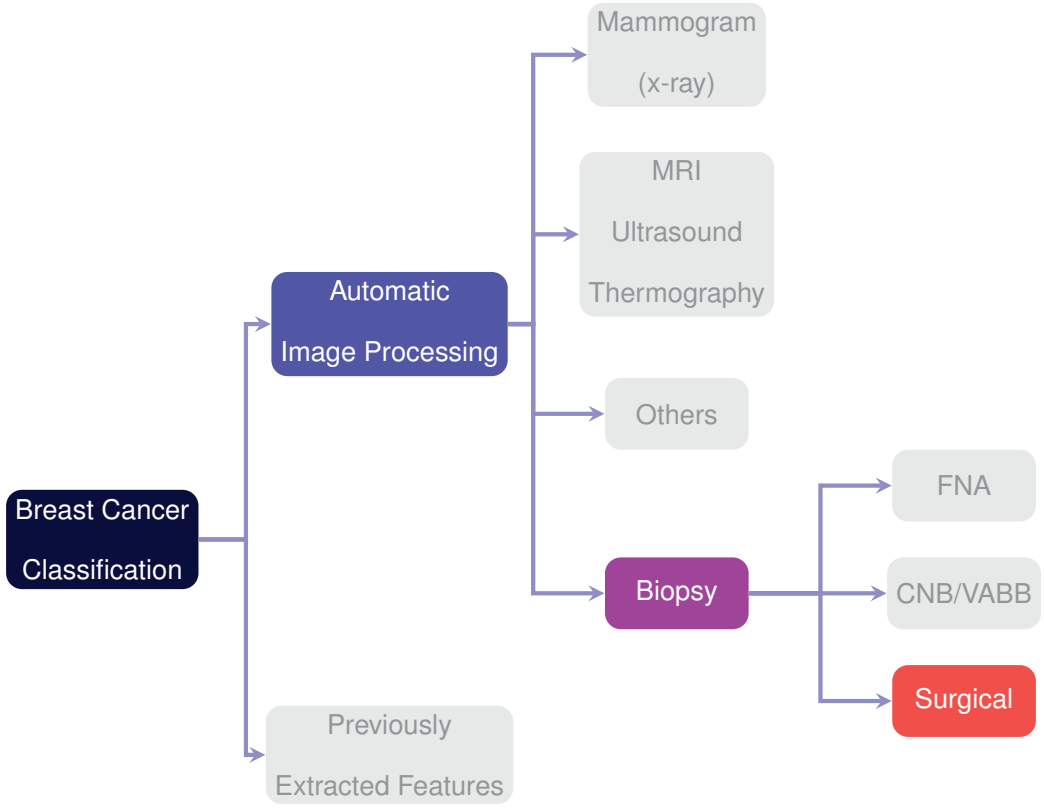


Figure 3.1: Categories of related works on breast cancer classification.
Source: The author (2015).

3.2 Research Using Automatic Image Processing

The automatic imaging processing for cancer diagnosis has been explored as a topic of research since the 1970s [260]. However, this subject continues to be the central point of recent

researches. In this section, we review the latest studies related to automatic image processing for BC diagnosis.

3.2.1 Images From Biopsy Sources

After preparing the tissue, mounting it on a glass slide, and staining it to enhance the cellular structures, microscopes are used to acquire digital histocytologic images. In the most commonly available configuration, named photomicroscopy, the magnified view of the sample in light microscope is observed by the pathologist, who selects the area of interest and digitally captures this area using a Charge-Coupled Device (CCD) camera attached to the microscope. Thereafter, the resulting digital images are ready for input into CAD/CAD-x systems. This image source is used in works like [82, 104, 151, 295, 296]. Notwithstanding its wide availability and relatively low cost, the conventional photomicroscopy approach presents some limitations such as eventual suboptimal image quality and the existence of only a few selected images (photographs) per slide, causing the impossibility to see a high-resolution overview of the whole slide and navigate in distinct magnifications (full objective lens zooming functionality) [284].

Avoiding limitations imposed by the photomicroscopy method, the Whole-Slide Imaging (WSI) has more recently become popular in large centers. WSI consists of the creation of a single, high magnification digital image of an entire microscopic slide, rather than one small microscopic field. Current WSI systems are composed of illumination systems, microscope optical components, and a precise focusing system to capture the images [105]. In summary, an automated microscope scans an entire slide at one or more resolutions, combines consecutive small images together into a single large image — normally some gigabytes [1] in size, — thus generating a "*virtual slide*". A virtual slide is a comprehensive digital rendering of the whole slide, visible at resolutions of less than $0.5\mu m$, using interactive software on a computer screen [107]. This software emulates the experience of using a light microscope, thus allowing the

[1] Unlike conventional digital image files containing a single image view at a single resolution, WSI files are usually formatted as multiresolution pyramids that contain multiple images comprising of multiple magnifications, normally being a vendor proprietary file format based on TIFF package of multiple JPEG 2000 images. The images of the same virtual slide can be assembled in various ways, depending on the particular scanner being used: tiling, line scanning, dual sensor scanning, dynamic focusing, or array scanning [284].

pathologist to freely navigate through histological sections in a complete range of standard magnifications [105].

3.2.1.1 Unavailable Datasets

The BC biopsy image dataset from the Israel Institute of Technology (Technion) was used in [293, 295, 296]. Technion BC image dataset [41] is public and composed of 361 760×570 JPEG images acquired from breast tissue biopsy slides stained with HE at 40× magnification. During the acquisition process, a digital CCD camera attached to a microscope was used. Then, a pathologist classified the samples as being normal, carcinoma *in situ* and invasive carcinoma, from which 119 samples were normal tissue, 102 as carcinoma *in situ*, and 140 as invasive ductal or lobular carcinoma.

Unfortunately, there is a notorious scarcity of public histopathological image database. At this moment, the location of the database used by Zhang *et al.* [293, 295, 296], as referenced in literature [2], is not available anymore. Moreover, other researchers use their own databases, not publicly available [82, 95, 104, 150, 151].

Zhang *et al.* [295] proposed a new cascade Random Subspace ensembles scheme with reject option for microscopic biopsy image classification. In their work, the combination of three different feature descriptors aim to fully describe image biopsy, capturing local texture (LBP), global texture (GLCM) and shape information (Curvelet Transform). The system achieved a classification accuracy of 99.25% (with a rejection rate of 1.94%) for the BC biopsy image dataset from the Israel Institute of Technology (Technion). In a new paper [296], Zhang *et al.* introduce a classification scheme based on a one-class kernel principle component analysis (KPCA) model ensemble for classification of medical images.

The analysis of histopathological images for automatic segmentation and identification of nuclei was explored by Kowal *et al.*[151]. This research used 500 breast FNA biopsy images collected from 50 patients (cases) of the Regional Hospital in Zielona Góra, Poland. Biopsy

[2] `ftp://ftp.cs.technion.ac.il/pub/projects/medic-image`

samples were digitized using a CCD video camera mounted on top of a microscope and the images were stored in 8 bit RBG BMP format files with 704×578 pixels. There are 10 images per case in the dataset, which is distributed into 25 malignant and 25 benign cases. Four distinct clustering algorithms k-means, fuzzy c-means, competitive learning Neural Networks and Gaussian mixture models were evaluated for clustering in the color space. Forty two morphological, topological and textural features were extracted from the segmented nuclei. The system accuracy rate for correctly classifying the cases as malignant or benign was between 96-100%.

Filipczuk *et al.* [95] conducted a study at the same Polish Regional Hospital, using a database with 737 images of cytological material collected from FNA biopsies of 67 patients. This breast cytological material was digitalized into virtual slides and a pathologist manually selected 11 distinct areas which were converted to 8-bit RGB TIFF files of 1538×828 pixels. The final dataset contains 25 benign (275 images) and 42 malignant (462 images) cases. Four different classifiers were applied in 25 features describing the nuclei. Their proposed method has achieved an effectiveness rate of 98.51% on classifying samples as benign or malignant.

George *et al.*[104] presented a fully automated method for cell nuclei identification and segmentation in breast cytological images. Extracted features through this method feed a system which classifies image samples as malignant or benign breast tumor. The dataset of this study was obtained in cooperation with the Early Cancer Detection Unit-Obstetrics and Gynecology Department, at the Ain Shams University Hospital, Egypt. Fine Needle Aspiration Citology (FNAC) samples where collected from breast lumps and stained with May-Grünwald-Giemsa[3] stain or Diff[4]. There are 92 RGB JPEG format microscopic images of 2560×1920, including 45 images of benign, and 47 of malignant tumors.

In order to classify breast cancer tissues, in a recent work, Bruno *et al.* [48] have presented a texture method based on the association among Fast Discrete Curvelet Transform

[3]This is the commonly used staining method of blood and others smears. It is used for morphological inspection and differential counting of blood cells. It combines the effect of acidic eosin, alkaline methylene blue, and azure. As a result, methylene blue stains the acidic components of the cell in blue tones, eosin stains the alkaline components of the cell in orange-red, and azure stains the basic cellular components in red and purple.

[4] Diff-Quik staining is used for the rapid differential staining of a variety of hematological and other smears to yield qualitative results similar to the Wright-Giemsa stain technique

(FDCT), Local Binary Pattern (LBP) and feature selection by Analysis of Variance (ANOVA). They evaluated the proposed method using images from four breast cancer datasets. Three of them containing digital/film mammography images and one containing Hematoxylin-Eosin (HE) histopathological images acquired by a CCD camera. The fourth evaluated dataset is the UCSB Biosegmentation Benchmark (UCSB-BB) [103], which contains 58 Red-Green-Blue (RGB) TIFF color images with 896×768 pixels. Half of these images are malignant cases and the other half benign. In this study, the authors converted the images into grey level images.

Table 3.1 summarizes these aforementioned classification methods which use histopathological images from biopsy sources.

Moreover, other relevant aspects of the breast cancer diagnosis have been explored. Quantitative image analysis applied on HE tissue images has been done in the last decade and some works have focused on BC grading [15, 22, 82, 148, 189, 224, 270]. Doyle *et al.* [82] introduced an image analysis methodology to automatically distinguish between (1) BC and non-cancer images and (2) between low and high grades – Bloom-Richardson (BR) grading system – of BC from digitized histopathology. Basavanhally *et al.* [22] used a multi-Field-of-View (FOV) classifier to successfully discriminate low, medium, and high Modified Bloom-Richardson (mBR) grade and identify specific image features at different FOV sizes that are important for distinguishing mBR grade in HE stained ER+ BC histology slides.

Human experts (pathologists) analyze and describe the slide images in terms of objects (such as cells, nuclei), the size relationship between one object and another, the spatial relationship between components like adjacency to tissues, the arrangement of glands and so on. For instance, Loukas *et al.* [175] employed 100× magnification as applied in routine clinical practice for deriving a diagnostic grade based on the architecture of the tissue section. Besides that, some works have addressed the characterization of the tumors in WSI. Approaches include representation of spatial relations of biological entity and decomposition of tumors in distinguishable patches. Loménie and Racoceanu [174] presented a formalization of spatial relation modeling techniques, convenient for structural representations of large image data by means of interest point sets and their morphological analysis. These authors have applied the proposed framework to process

large histopathological images capturing the spatial configuration of biological entities, which is important in the prognosis process. Even being a contribution for generic applications, this work is focused on automatic BC grading from WSI. The spatial distribution of distinct biological entities such as nuclei, tubular formations and *lumina* is analyzed. In addition, a simplified architectural/structural representation is presented to deal with large size of WSI (about eight gigabytes of data). The goal of Nayak *et al.* [202] is to characterize tumors decompounding WSI of histology section in patches such as stroma, normal, tumor, transition into necrosis and necrotic, etc. Their method, based on a variation of the Restricted Boltzmann Machine (RBM), performs an automated feature learning from unlabeled images for classifying distinct morphometric regions within a WSI. A dataset of small image blocks of 1000×1000 was used, which has been extracted from Glioblastoma Multiforme (GBM) and Clear Cell Kidney Carcinoma (KIRC) from the Cancer Genome Atlas (TCGA) [5] archive. Evaluation of 1,400 and 2,500 samples of GBM and KIRC achieved a performance of 84% and 81%, respectively. The learned model was projected on each WSI (of size 20000×20000 pixels or larger) for characterizing and visualizing tumor architecture. Zhou *et al.* [301] also conducted experiments using the same TCGA public database. The authors proposed a multispectral feature learning model that automatically learns a set of convolution filter banks from separate spectra to efficiently discover the constitutive tissue morphometric signatures, based on Convolutional Sparse Coding (CSC). The learned feature representations were aggregated through the Spatial Pyramid Matching (SPM) framework and finally classified using a linear SVM.

Despite the colors being relevant for diagnosis by histological images, usually in preprocessing for pattern recognition purpose these images are converted to grayscale. Thus, some methods try to avoid discard the important color information. Recently, Shia *et al.* [253] introduced a method for classification of color medical images, using a hepatocellular carcinoma histological image dataset to evaluate the new method. This method, named Joint Sparse coding based SPM (JScSPM), unlike other popular techniques such as Sparce coding based linear SPM (ScSPM) and its variants, extract the color information in each color channel and the correlation between channels.

[5] http://cancergenome.nih.gov/

Table 3.1: Comparison among classification methods using biopsy images (unavailable datasets).

Authorship [ref](year)	Dataset	Biopsy	Acquisition	# Images	Preprocessing	Features	Results	
Doyle *et al.* [82] (2008)	nonpublic	?	$CCD + M$	48	?	– Grey Level – Haralick – Gabor filter	SVM	Ac = 95.80
Zhang [293] (2011)	public [41]	FNA	$CCD + M$	361	– image enhancement and conversion	– LBP – Curvelet Transform	Serial Fusion of Random Subspace Ensemble	Ac = 97.00, ReR = 0.80
Kowal *et al.* [151](2013)	nonpublic	FNA	$CCD + M$	500	– image enhancement and cropping – cell nuclei segmentation	– morphometric – GLCM – topology	k-NN Naive Bayes Decision Tree	Ac = 93.11 Ac = 90.22 Ac = 87.56
Filipczuk *et al.* [95](2013)	nonpublic	FNA	WSI	737	– selection of 11 areas of interest – cell nuclei segmentation	– shape-based – GLCM – GLRLM – topology	k-NN Naive Bayes Decision Tree SVM	EfR = 98.51, MCC = 0.97 EfR = 98.51, MCC = 0.97 EfR = 97.01, MCC = 0.94 EfR = 98.51, MCC = 0.97

$CCD + M$: CCD digital camera attached to light microscope • WSI: Whole-Slide Imaging • Ac: Accuracy (%) • ErR: Error Rate (%) • ReR: Rejection Rate (%) • EfR: Effectiveness Rate (%) • MCC: Matthews Correlation Coefficient • Se: Sensivity (%) • Sp: Specificity (%) • ?: Non-reported

Table 3.1: (continued)

Authorship [ref](year)	Dataset	Biopsy	Acquisition	# Images	Preprocessing	Features	Results	
Zhang *et al.* [295](2013)	public [41]	FNA	$CCD + M$	361	– image enhancement and conversion	– CLBP – GLCM – Curvelet Transform	Random Subspace Ensemble	Ac = 99.25, ReR = 1.94
George *et al.* [104](2014)	nonpublic	FNA	$CCD + M$	92	– image resizing, enhancement, and conversion – cell nuclei detection – false findings eliminations – cell nuclei segmentation	– 10 shape-based – 2 textural	SVM MLP LVQ PNN	Se = 88.48, Sp = 90.99 Se = 80.66, Sp = 76.28 Se = 87.95, Sp = 85.69 Se = 96.32, Sp = 94.57
Ko *et al.* [150](2014)	nonpublic	FNA	WSI	100	– image enhancement and cropping – cell detection and segmentation – cell nuclei detection and segmentation	– 9 shape-based – 12 graph-based – 12 Haralick's textural	SVM with LDA	Ac = 90.60
Zhang *et al.* [296](2014)	public [41]	FNA	$CCD + M$	361	– image enhancement and conversion	– CLBP – GLCM – Curvelet Transform	One-class KPCA	Ac = 92.28, ErR = 7.72

$CCD + M$: CCD digital camera attached to light microscope • WSI: Whole-Slide Imaging • Ac: Accuracy (%) • ErR: Error Rate (%) • ReR: Rejection Rate (%) • EfR: Effectiveness Rate (%) • MCC: Matthews Correlation Coefficient • Se: Sensivity (%) • Sp: Specificity (%) • ?: Non-reported

Table 3.1: (continued)

Authorship [ref](year)	Dataset	Biopsy	Acquisition	# Images	Preprocessing	Features	Results	
Bruno *et al.* [48](2016)	public [103]	?	$CCD + M$	58	– image conversion	– Curvelet Transform – LBP	Polinomyal Decision Tree Random Forest SVM	Ac = 100.00 Ac = 67.00 Ac = 72.00 Ac = 72.00

$CCD + M$: CCD digital camera attached to light microscope • WSI: Whole-Slide Imaging • Ac: Accuracy (%) • ErR: Error Rate (%) • ReR: Rejection Rate (%) • EfR: Effectiveness Rate (%) • MCC: Matthews Correlation Coefficient • Se: Sensivity (%) • Sp: Specificity (%) • ?: Non-reported

Source: The author (2016).

3.2.1.2 BreaKHis Dataset for Benchmarking

Since the publication of the BreaKHis dataset [256], a growing number of researchers are using this public available dataset to evaluate their proposed methods for breast cancer classification such as [5, 25, 55, 59, 60, 73, 79, 111, 112, 116, 145, 194–196, 241, 255, 257, 258, 282, 299]. Most of the BreaKHis related papers are generally based on automated feature learning and deep learning techniques to improve the performance on this dataset [5, 25, 55, 60, 73, 116, 194–196, 255, 257, 258, 282, 299].

Initally, in the benchmark approach [256], Spanhol *et al.* presented an evaluation of different combinations of six different visual feature descriptors along with different classifiers. Computing the average over five trials, the authors report accuracies ranging from 80% to 85%, which may vary depending on the image magnification factor. In the proposed experimental protocol, the BreaKHis dataset has been randomly divided into a training set (70%) and a testing set (30%). Patients used for building the training set were not used for the testing set. Such protocol, including the 5 training-test splits and the respective results, is considered the comparative baseline. Note that the distribution of the samples for each trial is publicly available and allows a fair comparison of methods. However, several published works don't strictly follow this protocol making their results not directly comparable with the baseline.

The first deep learning approach using the BreaKHis was originally published by Spanhol *et al.* [257]. In this work, the authors presented the results achieved using a CNN trained directly from the original histopathology images. Given that CNNs generally require large datasets, they make use of the patches trick, which consists of extracting sub-images at both training and test phases. During training, the idea is to increase the training set by means of extracting a large number of patches at randomly-defined positions. Then during test, patches are extracted from a grid, and after classifying each patch, their classification results are combined. The authors show that, with this approach, an increase of about 4–6 % can be observed in the baseline accuracy [256].

Bayramoglu *et al.* [25] proposed a method which uses CNNs to classify the same BC histopathology images, but independent of the magnifications factors. In their work, two different CNN architectures are used: a single task CNN is used to predict malignancy, and a multi-task CNN is used to predict both malignancy and image magnification level simultaneously. In general, their experimental results are similar to the previous baseline results obtained from hand-crafted features [256]. However, the authors claim that the multi-task system does not depends on specific magnification factors and could leverage the training by using data from distinct imaging devices.

The use of DeCAF features on the BreaKHis dataset was investigated by Spanhol *et al.* [258]. The experimental evaluation shown that these features can be a viable alternative to fast development of high-accuracy BC recognition systems, generally achieving better results than traditional hand-crafted textural descriptors [256] and, in some cases outperforming task-specific CNNs [257].

A novel regularization technique for CNNs was presented by Akbar *et al.* [5]. The key concept is the transitioning between convolutional layers and Fully Connected (FC) layers. The modified transition model learns filters of varying sizes and then collapses them via average pooling. According to the authors, this model achieved an overall test accuracy of 82.7% on the BreaKHis dataset. However, the employed experimental protocol is not clear, especially how the distinct magnification factors were considered. The authors themselves admit that the result is not directly comparable.

Song *et al.* [255] report the results of their transfer learning-based method which uses a VGG-VD model [254] pretrained on ImageNet in order to extract features. The extracted local features are encoded by Fisher Vectors (FV) and the resulting descriptors are transformed by applying an adaptation layer. The final image classification is performed using a linear-kernel Support Vector Machine (SVM).

Another deep learning multi-classification model was proposed by Han *et al.* [116]. The so called Class structure-based Deep Convolutional Neural Network (CSDCNN) is an end-to-end

recognition method that adopts a hierarchical representation with feature space constraints which maximize the Euclidean distance of inter-class labels. In order to evaluate the performance of the method, the original BreaKHis dataset and the same database augmented on the training set for oversampling imbalanced classes it were used. The achieved performance, for both image level and patient level analysis is enhanced by using the BreaKHis dataset augmented. In the binary classification task, the average accuracy was 96.07% and 96.25% for image level and patient level, respectively. The authors claim that a great advantage of the CSDCNN is that it deals with WSI. However, all the conducted experiments are carried out using just the published version of the BreaKHis dataset, notably composed of images acquired by the conventional photomicroscopy method which is widely available and relatively low cost.

Dimensionality reduction strategies for CNN-based features were investigated by Cascianelli *et al.* [55]. The main purpose of that work was to establish the optimal trade-off between accuracy and dimensionality.

A CNN-based solution is also presented by Das *et al.* [73]. They employ transfer learning of a pre-trained CNN for specific breast histopathology features and majority voting using random multi-views at multi-magnification. Again, the authors describe their method as being a "histopathology whole slide classification approach", but all the conducted experiments are carried out using just the published version of the BreaKHis dataset, notably composed of images acquired by the conventional photomicroscopy method.

Aiming to address the two-class BC classification considering class and the sub-class labels as prior knowledge, Wei *et al.* [282] proposed a method based on CNN, named BiCNN. Two datasets are used: the original and an augmented version. Both datasets were divided into a training set, a validation and a test set, corresponding to 50%, 25%, and 25% of the images, respectively. The reported average recognition rate for patient level and image level was 97%.

An ensemble model containing three custom Convolutional Neural Network (CNN) classifiers using transfer learning was proposed by Zhy *et al.* [299]. The custom model is inspired by VGGNet [254] architecture and the Max rule is used to ensemble the individual

classifiers. Similarly to other published works, in this work the authors have used a patch-based approach which randomly sample five sub-images of size 224 × 224 pixels. Moreover, a data augmentation is also done performing zooming, horizontal and vertical flips. After ensembling and cross-validation, the achieved percentage accuracies were 93.3 ± 2.3, 94.6 ± 2.2, 94.8 ± 3.2, and 88.4 ± 4.1 for the magnifications 40×, 100×, 200×, and 400×, respectively.

Aiming to explore the usefulness of the local features and the frequency domain information associated with the hierarchical property of the CNNs, Nahid and Yinan [196] proposed a CNN-based model along with residual blocks. The model named CNN-CH, which achieved the best results, utilizes features with Contourlet Transform (CT) coefficients, histogram information and LBP. The experiments were performed on each magnification factor separately, achieving accuracies of 94.40%, 95.93%, 97.19%, and 96.00%, for 40×, 100×, 200×, and 400×, respectively.

More recently, Motlagh *et al.* [194] have examined different frameworks of Inceptions models [266] and ResNet [119] architectures on cancer digital images, including BreaKHis. Aiming to prepare the data for further feature extraction and classification, the authors have applied color map selection and data augmentation as pre-processing steps. Then, fine-tuned pre-trained deep CNNs were used to classify malignant and benign cases, as well as, perform the sub-types detection. Approximately 85% of available images were randomly selected to create the training set, and the remaining images were used for testing. This work achieved 94.8% and 96.4% classifying benign and malignant sub-types, respectively. Considering the classes benign and malignant, the accuracy of the binary classification task was 98.7%.

Relying on a more traditional approach, Gupta and Bhavsar [112] compare their results directly with [25], which also considers a magnification independent model, but using a deep learning approach. They proposed an integrated multi-scale model which uses joint color-texture features and a classifier ensemble to deal with the large variability appearance present in the breast histopathology images. The final color-texture descriptor used is the concatenation of six different representations: normalized color space representation, multilayer coordinate clusters representation, Gabor features on Gaussian color model, Gabor chromatic features, complex

wavelet features and chromatic features, and Opponent Color Local Binary Pattern (OCLBP). A majority voting among a pool of heterogeneous classifiers combines the individual outputs to classify the testing samples. In general, this method improves the recognition rate on patient-level achieved by the previous deep learning method by 4% [25] and yields a lesser variance in scores. In another published work, Gupta and Bhavsar [111] analyze performance of the same proposed framework, also comparing with the CNN results achieved by Spanhol *et al.* [257]. A Receiver Operating Characteristic (ROC) analysis and the related Area Under the ROC Curve (AUC) computation on patient-level scores is presented, suggesting better results achieved by the integrated model.

Chan and Tuszynski [59] presented an attempt to classify breast tumors based on malignancy using fractal dimension as the descriptor and SVM as the classifier. The authors have employed their own experimental protocol, making efforts to classify the tumor subtypes present in the BreaKHis dataset, not achieving competitive results. Moreover, considering the classification into benign and malignant slides at 40 ×, despite the F1 score of 97.9%, the reported accuracy is only 55.6%.

Kahya *et al.* [145] proposed an adaptative sparse support vector aiming to perform the selection of relevant features for describing histopathological images, removing noisy features and improving the classification performance. Their method is based on the Sparse Support Vector Machine (SSVM) with L_1-norm with Wilcoxon rank sum test to weight each feature. In their experimental setting using 10 folds cross validation, the proposed adaptative SSVM achieved improved performance in all magnification factors, being the average classification accuracy 94.97 (40×), 93.62 (100×), 94.54 (200×), and 94.42 (400×).

Finally, in a recent paper, Dimitropoulos *et al.* [79] used the encoding of histological images as Vector of Locally Aggregated Descriptors (VLAD) representations on the Grassmann manifold to perform automated grading of invasive BC. In that work, each histological image is represented as a set of multidimensional spatially-evolving signals mapped and encoded on the Grassmann manifold, a non-Euclidean space. This methodology tries to avoid the modeling of the domain expert (pathologist) knowledge in order to detect histologic primitives, aiming to model

directly from raw images without segmentation. The efficiency of the proposed method in the grading of invasive BC is assessed using a dataset [6] created by the authors, which consists of 300 breast cancer annotated images of grades 1–3 collected from 21 patients of the General Hospital of Thessaloniki, Greece. On the other hand, considering only the classification of benign and malignant cases available in the BreaKHis dataset, the proposed method presents an improvement of 6.53% in average when compared to the CNN-based deep learning approach [257].

Table 3.2: Comparison among classification methods using BreaKHis dataset.

Authorship (year) [ref]	Protocol	Approach	Zoom	Patient Accuracy (%)
Spanhol *et al.*(2016) [256]	Yes	Conventional	40×	83.9 ± 4.1
			100×	82.1 ± 4.9
			200×	85.1 ± 3.1
			400×	82.3 ± 3.8
Spanhol *et al.*(2016) [257]*	Yes	Deep Learning	40×	90.0 ± 6.7
			100×	88.4 ± 4.8
			200×	84.6 ± 4.2
			400×	86.1 ± 6.2
Bayramoglu *et al.*(2016) [25]	Yes	Deep Learning	40×	83.0 ± 3.0
			100×	83.1 ± 3.5
			200×	84.6 ± 2.7
			400×	82.1 ± 2.4
Spanhol *et al.*(2017) [258]	Yes	Deep Learning	40×	84.0 ± 6.9
			100×	83.9 ± 5.9
			200×	86.3 ± 3.5
			400×	82.1 ± 2.4
Akbar *et al.*(2017) [5]	?	Deep Learning	40×	82.7 ± ?
			100×	? ± ?
			200×	? ± ?
			400×	? ± ?
Song *et al.*(2017) [255]	Yes	Deep Learning	40×	90.0 ± 3.2
			100×	88.9 ± 5.0

[6]Available on `http://doi.org/10.5281/zenodo.834910`

Table 3.2: (continued)

Authorship (year) [ref]	Protocol	Approach	Zoom	Patient Accuracy (%)
			200×	86.9 ± 5.2
			400×	86.3 ± 7.0
Han *et al.*(2017) [116]	No	Deep Learning	40×	94.5 ± 2.1
			100×	93.2 ± 1.4
			200×	94.7 ± 3.6
			400×	93.5 ± 2.7
Cascianelli *et al.*(2017) [55]	No	Deep Learning	40×	87.0 ± ?
			100×	85.2 ± ?
			200×	85.0 ± ?
			400×	81.3 ± ?
Das *et al.*(2017) [73]	?	Deep Learning	40×	94.8 ± ?
			100×	94.4 ± ?
			200×	94.6 ± ?
			400×	93.5 ± ?
Wei *et al.*(2017) [282]	No	Deep Learning	40×	97.0 ± ?
			100×	97.2 ± ?
			200×	97.9 ± ?
			400×	97.5 ± ?
Zhy *et al.*(2017) [299]	No	Deep Learning	40×	93.3 ± 2.3
			100×	94.6 ± 2.2
			200×	94.8 ± 3.2
			400×	88.4 ± 4.1
Nahid and Yinan (2018) [196]	No	Deep Learning	40×	94.4 ± ?
			100×	95.9 ± ?
			200×	97.2 ± ?
			400×	96.0 ± ?
Motlagh *et al.*(2018) [194]	No	Deep Learning	40×	98.7 ± ?
			100×	98.7 ± ?
			200×	98.7 ± ?
			400×	98.7 ± ?

Table 3.2: (continued)

Authorship (year) [ref]	Protocol	Approach	Zoom	Patient Accuracy (%)
Gupta and Bhavsar (2017) [112]	No	Conventional	40×	87.2 ± 3.7
			100×	88.2 ± 3.3
			200×	88.89 ± 2.5
			400×	85.82 ± 3.8
Chan and Tuszynski(2016) [59]	No	Conventional	40×	55.6 ± ?
			100×	? ± ?
			200×	? ± ?
			400×	? ± ?
Kahya *et al.*(2017) [145]	No	Conventional	40×	94.9 ± ?
			100×	93.6 ± ?
			200×	94.5 ± ?
			400×	94.4 ± ?
Dimitropoulos *et al.*(2017) [79]	Yes	Conventional	40×	91.8 ± ?
			100×	92.2 ± ?
			200×	91.6 ± ?
			400×	90.5 ± ?

Symbol * indicates the combination of multiple classifiers • Symbol ? indicates unpublished/unknown.

Source: The author (2018).

3.2.2 Images From Non-Invasive Sources

Currently mammography is the standard screening tool for BC, detecting anatomical changes and structural distinction between normal and abnormal (eventually tumoral) breast tissue. This procedure uses low energy x-ray to produce images on high contrast and high resolution. Being the mammography a routinary procedure, the CAD/CADx systems present in mammographs already incorporate many research results of the last decades. The list of works related to machine learning and image processing in mammograms is pretty extensive.

Recently published studies on this diagnostic image field present a wide range of techniques, covering topics such as blood vessel segmentation [193], analysis of bilateral mammographic asymmetry [57, 94, 232], automatic breast density estimation and classification [63, 99, 128, 173, 250, 275], segmentation of region of interest, detection and classification of mammographic microcalcifications and masses [1, 3, 11, 19, 56, 64, 76, 109, 130, 133, 134, 159, 171, 172, 186, 197, 223, 242, 269, 292, 298], content-based image retrieval [143, 283], architectural distortion recognition [32, 215], geometrical alignment of mammogram sequences (mammogram registration) [88, 125, 179, 182, 234], image enhancement in post acquisition process [181, 216, 264], new coordinate system based on breast anatomy [39], and synthesis of 2D mammograms from Digital Breast Tomosynthesis (DBT) [249].

Unlike the notorious lack of basic public data for research with histopathological images of BC, there are some mammographic databases publicly available to the research community. The two most easily accessed and therefore most commonly used are the Mammographic Image Analysis Society (MIAS) database [261, 272] and the Digital Database for Screening Mammography (DDSM) [74, 120, 235].

Thermogram is a relatively cheap, non-ionizing and free-contact procedure used as a screening tool for BC detection. The approaches to improve this modality of early BC diagnosis have gaining space among some researchers of image processing and machine learning [13, 91, 101, 153–157, 165, 166, 192, 245].

3.3 Research Using Previously Extracted Features

Relatively few works in literature have concentrated on BC classification through previously extracted features using their own private databases. For instance, a study conducted by Huang *et al.* [131], focused on a hybrid SVM-based strategy with feature selection, searched for risk factors in BC. A total of 80 samples from breast tissue were collected in the Chung-Shan Medical University Hospital. The classification between BC and fibroadenoma achieved 86% of

averaged accuracy in a 5-fold cross validation. The obtained results also correlate the presence of 5 DNA viruses — HSV-1, EBV, CMV, HPV, and HHV-8 — with breast tumors.

However, several papers, such as [2, 4, 16, 30, 62, 86, 87, 104, 146, 147, 212, 222, 225, 240, 252, 296], use WBCD variations to assess different classification methods to aid BC diagnosis. Normally, high successful classification rates are reported. WBCD is a dataset of feature descriptors extracted from digitized BC biopsy samples. It is publicly available on the machine learning repository from the University of California at Irvine (UCI)[18]. The samples were acquired from FNA biopsies of breast mass. Each one of the 569 instances from the base has a descriptor formed by 30 real-valued numeric fields. These values describe the properties of cell nuclei present in the image and tell if the instance is a malignant or a benign case. Properties include attributes such as clump thickness, uniformity of cell size and shape, marginal adhesion, single epithelial cell size, bare nuclei, bland chromatin, normal nucleoli and mitoses. Table 3.3 summarizes the methods of several works – using WBCD database – with their respective performances.

Generally, in terms of applications involving medical diagnosis, it is expected that the automated system achieves high predictive accuracy and it has be based on a set of significant classification rules that can be validated by human experts. A BC diagnosis system, presented by Peña-Reyes and Sipper [222], combined fuzzy systems with Genetic Algorithm (GA), achieving a 97.36% accuracy. This system used simple rules to facilitate human interpretation about how the result was obtained. The final diagnosis was associated to confidence measures.

Aiming to optimize the set of rules extracted from trained neural networks, Setiono performed a preprocessing of neural network input data [252]. The preprocessing used by the author in WBCD is twofold. The first step consists in just eliminating – before training – samples with missing attribute values. Then, only the most relevant attributes are selected for classification, thus reducing the computation time. The resulting rule set is more concise and accurate than the previous methods reported in the literature, achieving a 98.10% accuracy rate.

The selection of relevant features to optimize accuracy and minimize classification time was explored in other works as well. Neural networks combined with Association Rule (AR) was the approach presented by Karabatak and Ince in [147]. The purpose of using AR is the dimensional reduction of feature space in order to make the diagnosis system faster. The traditional neural network model was compared to the proposed model of the neural network conjugated with AR. The correct classification rate of the new system was 95.6%.

A Least Square Support Vector Machine (LS-SVM) classifier was used by Polat and Güneş [225], reaching a 98.53% accuracy in a 10-fold cross validation. Another SVM method combined with feature selection was applied by Akay [4] over a training-test of distinct partitions extracted from WBCD. Their results show the highest accuracy at 99.51% using five features (80-20% training-test partition).

SVM technique was also used by Chen *et al.* in BC classification, but combined with Rough Set (RS) [62]. The RS reduction algorithm is used as a tool to select features, removing redundancies and further improving accuracy of SVM classifier, which achieved 99.41%.

A hybrid classifier method for BC, based on fuzzy-artificial immune system and k-Nearest Neighbor (k-NN) algorithm, was proposed by Şahan *et al.* in [240]. Indeed, in computer-aided diagnosis systems, short computing time and high accuracy are key-factors. The time consumed by algorithms k-NN can make them impracticable. Thus, the first step of the hybrid system introduced in [240] explores the large clustering capability of such systems to significantly reduce the data inputted into k-NN algorithm. An accuracy of 99.14% via 10-fold cross validation was reported.

Three classifier algorithms based on common architectures of neural networks, Multi-Layer Perceptron (MLP), Radial Basis Function (RBF), and Probabilistic Neural Network (PNN), were tested by Azar and El-Said for BC diagnosis [16]. Although MLP is one of the most widely used neural network architectures for biomedical applications, the best performance was achieved by a PNN classifier with 100 and 97.66% of accuracy rate in training and test phases, respectively. And MLP, using Scaled Conjugate Gradient (SCG) learning algorithm was ranked

second, reaching 97.80% and 96.34% classification accuracy for training and validation phases. Finally, RBF using 23 hidden neurons occupied the last position for accuracy rate at 96.05%. In the same problem, Padmavathi had already compared MLP and RBF architectures, besides logistic regression [212]. This previous study shows better result for the RBF model (97.00%), whereas MLP reached 93.64%, requiring a larger execution time when compared to RBF.

Elouedi *et al.* [87] tried to improve the classification of malignant instances by using the k-Means algorithm for clustering and the decision tree for classification. Their results demonstrate a gain in accuracy rate (from 94.56% to 95.14%) by splitting the original malignant instances into two clusters for posterior classification using a C4.5 classifier.

Bhardwaj and Tiwari [30] used Genetically Optimized Neural Network (GONN) in the WBCD database, introducing new crossover and mutation operators. Their approach showed results outperforming previous works in the literature, reaching a classification accuracy of 98.24%, 99.63% and 100% for 50-50, 60-40, 70-30 training-testing partition respectively and 100% for 10-fold cross validation.

Table 3.3: Accuracy rates achieved by different classifiers of the literature – WBCD Dataset.

Authorship (year) [ref]		Method	Accuracy (%)
Peña-Reyes and Sipper (1999) [222]	⊕	Fuzzy GA	97.36
Setiono (2000) [252]		Feed Forward Neural Network	98.10
Abonyi and Szeifert (2003) [2]		Supervised Fuzzy Clustering	95.57
Polat and Güneş (2007) [225]		LS-SVM	98.53
Şahan *et al.* (2007) [240]	⊕	Fuzzy-Artificial Immune System k-NN	99.14
Karabatak and Ince (2009) [147]	⊕	Neural Networks Association Rules	95.60
Akay (2009) [4]		SVM with Feature Selection	99.51
Padmavathi (2011) [212]	⊗	RBF MLP Logistic Regression	97.00 91.30 73.70
Chen *et al.* (2011) [62]	⊕	SVM Rough Set	99.41
Azar and El-Said (2012) [16]	⊗	PNN MLP RBF	97.66 96.34 96.05
Eleyan (2012) [86]	⊗	Naive Bayes Neural Networks k-NN SVM	95.60 96.09 96.36 96.56
Zhang *et al.* (2014) [296]		One-class KPCA	97.28
Elouedi *et al.* (2014) [87]	⊕	Decision Trees k-Means	95.14
Bhardwaj and Tiwari (2015) [30]		GONN	100.00
Karabatak (2015) [146]	⊗	Naive Bayes Weighted-Naive Bayes	96.17 98.54

Symbol ⊗ indicates comparison between methods and ⊕ stands for combination of methods.

Source: The author (2015).

Chapter 4

Methodology

This chapter presents the proposed alternative model and the methodological approach used to carry out the experiments in this research. In order to assess the validation of the proposed hybrid model, the following tasks will be performed:

- To build a histopathological image database composed of microscopic breast tumor images labeled by pathologists;
- To extract different hand-crafted texture descriptors from each image present in the database;
- To build a baseline supervised classification system based on traditional classifier algorithms and extracted texture descriptors;
- To evaluate the baseline supervised classification system, by computing the accuracy as a metric to verify how the model performs on testing unseen data;
- To investigate different Convolutional Neural Network (CNN) architectures and propose a suitable one to deal with histopathological image;
- To investigate the performance of state-of-the-art Multiple Instance Learning (MIL) methods for classifing histopathological breast cancer images.

The remainder of this chapter is organized as follows: Firstly, a general view of the proposed hybrid approach is given. Then, in Section 4.2, BreaKHis, a histopathological image database of breast tumors which was used in the experiments, is introduced. After that, the methodological strategies used to carry out the experiments, considering the traditional representation, the automatic representation using CNN, and the MIL approach, are detailed in sections 4.3, 4.4 and 4.5, respectively. Section 4.6 presents the evaluation criteria, highlighting the selected performance metrics. Finally, in Section 4.7 we present the software tools used for building the baseline system, as well perform the experiments based on the deep learning approach with CNN, and under the MIL paradigm.

4.1 The Proposed Alternative Approach

The main goal of this work is to build a classification system of breast cancer (BC) histopathological images which uses an alternative holistic approach, i.e., avoiding the explicit segmentation of the images. However, histopathological images are naturally challenging for the classification systems. This can be explained due to the inherent heterogeneity of these images because the tissue is composed of complex, non-homogeneous and very dense biological structures. Regarding to BC detection, some tissue patterns can occur in the normal tissue, benign lesions and malignant lesions. Besides, the appearance of the stained slides varies among different laboratories or even in the same lab, mainly in terms of section thickness and coloring. Figure 4.1 shows an example of color variation in histopathological images acquired using the same imaging condition, but from two different slides, both prepared in the same laboratory. As pointed by Li and Plataniotis [170], the problem of color variation in histopathological images is caused by inconsistent staining procedures (timing, concentrations, different manufacturers, reagents, etc.) and nonstandard imaging condition. Thus, these properties of the histopathological images might make the classification a hard task for pathologists or automatic systems. But robust histopathology image analysis systems must mitigate this variability in image appearance [277].

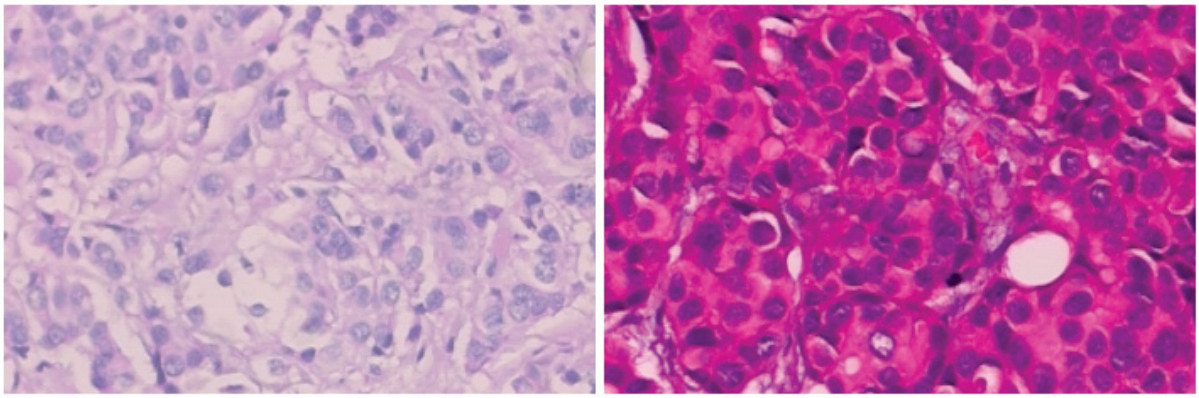


Figure 4.1: Example of color variation in histopathological images (HE staining).
Source: The author (2015).

We propose to investigate different representation approaches, considering both the Single Instance Learning (SIL) and MIL paradigms, to deal with these challenges. Under the SIL paradigm traditional supervised learning, using hand-crafted or even automatic representation, requires a training dataset that consists of various inputs with corresponding class labels. In many applications, however, it is hard or even impossible to accurately and consistently assign labels to the inputs of the training dataset, thus resulting in partial or ambiguously labeled data. MIL arises as a feasible alternative to deal with this issue by grouping instances into bags, without the need to label all the instances. In our case, MIL allows for the modeling of real situations such as digital histopathology images, when there are several images for one patient. Figure 4.2 presents a simplified block diagram representing the proposed model and its main aspects.

Firstly, we have created a large dataset of breast cancer histopathology images named BreaKHis, which is available to the scientific community [256]. All the experiments were performed using the BreaKHis dataset, aiming to make this real-life, challenging dataset a useful benchmark for research in breast cancer computer-aided diagnosis. We also have built a baseline system exploring the use of classic textural representation. This system, based on hand-crafted features, is covered in the Section 4.3.

Then, considering the limitations of the hand-crafted features, we propose the usage of an automatic learning representation to extract discriminant features from the histopathological

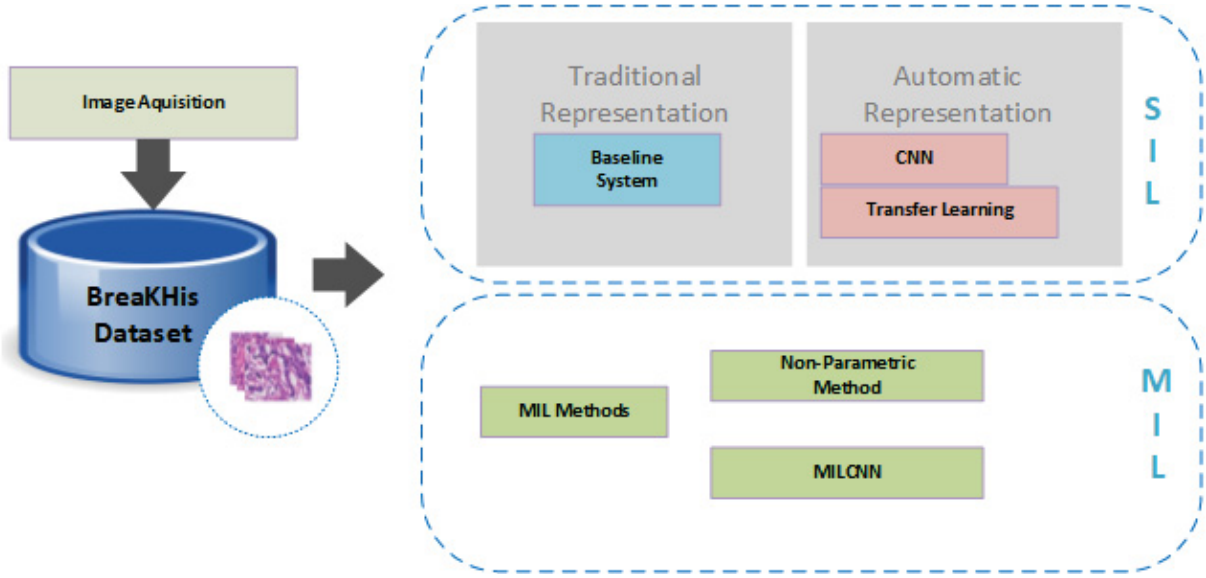


Figure 4.2: Simplified block diagram representing the proposed alternative approach.
Source: The author (2016).

images. In our approach, the automatic representation relies on supervised deep learning techniques, specially CNN and transfer learning. Such automatic representation is detailed in Section 4.4.

Additionally, we should cope with certain intrinsic aspects of the BreaKHis dataset, presented in the Section 4.2. To do so, we are employing a patch-discard strategy (Section 4.4.1.4) and finally a MIL formulation, detailed in Section 4.5.

4.2 The Histopathological Image Database BreaKHis

As discussed in Section 3.2.1, there is an evident scarcity of public annotated histopathological image database. At same time, the algorithms present in the histopathology image analysis systems need to be trained on selected image fields with well documented histopathological characterization. In a recent review, Veta *et al.* [277] point out that the lack of large publicly available annotated datasets would be the main obstacle in the development of new histopathology image analysis methods.

Contributing to mitigate this gap, we introduced in [256] a database called BreaKHis composed of 7,909 breast histopathological images acquired from 82 patients. The samples were generated from breast tissue biopsy slides, stained with Hematoxylin-Eosin (HE). Tumoral areas were identified and labelled by means of visual analysis of tissue sections under a microscope, performed by anatomo-pathologists of the P&D Lab. Final diagnosis of each case was produced by experienced pathologists and confirmed by complementary exams, such as the Immunohistochemistry (IHC) analysis.

These microscopic images of breast tumor tissues are divided into benign and malignant tumors. Both breast tumors, benign and malignant, can be sorted into different types based on the way the tumoral cells look under the microscope. At this time, there are four histological distinct types of breast benign tumors: adenosis (A), fibroadenoma (F), phyllodes tumor (PT), and tubular adenona (TA); and four malignant tumors (BC): carcinoma (DC), lobular carcinoma (LC), mucinous carcinoma (MC) and papillary carcinoma (PC) in dataset. Table 4.1 summarizes the image distribution.

Table 4.1: Images distribution by magnification factor and main class.

Magnification	Benign	Malignant	Total
40×	625	1,370	1,995
100×	644	1,437	2,081
200×	623	1,390	2,013
400×	588	1,232	1,820
Total	2,480	5,429	7,909
# Patients	24	58	82

Source: The author (2015).

A more complete description of the BreakHis dataset is shown in Appendix B, detailing the distinct biopsy methods used for collecting the samples, the acquisition process and the types of tumors present in this database.

4.2.1 A Priori Information

In this section, we point out some relevant *a priori* information about the BreaKHis dataset.

Patient-level Annotation: As previously mentioned, BreaKHis dataset is composed of 7,909 breast histopathological images acquired from 82 patients, one slide per patient, being 58 cases malignant and 24 benign. A patient (slide) was labeled by the pathologists as malignant because one or few areas within the slide presented cancerous patterns, i.e., it matched the criteria of malignancy. If the complete area of the lesion does not match any criteria of malignancy it is considered a benign tumor. Each labeled patient can be regarded as a collection of several images captured using four magnification factors: 40×, 100×, 200×, and 400×. Despite the fact that the collection (patient) has a label, the annotation is not provided to individual instances (images), but each instance inherits the label of the collection where it lies.

Imbalanced Classes: Currently, the two classes of BreaKHis dataset are not equality represented, i.e., the amount of examples of malignant and benign cases is substantially unbalanced. Considering the image distribution by class in the BreaKHis, there are 2480 benign images and 5429 malignant images. In order to get some intuition about the class distribution, we have performed a visual inspection on the BreaKHis images. Considering the high dimensional data, for visualization purpose, we employed the popular t-Distributed Stochastic Neighbor Embedding (t-SNE) technique, developed by Maaten and Hinton [180]. This nonlinear dimensionality reduction is well suited for visualization of high-dimensional datasets and has been successfully applied on large real-world datasets. The resulting t-SNE projection applied in each magnification factor is shown in Figure 4.3. In this scatter plot of BreaKHis dataset, the blue color represents benign instances and the malignant instances are presented in magenta.

Small Regions of Interest: Normally, in the clinical routine, only some small regions within the histopathological images are effectively used by pathologists to characterize the lesion into cancer (malignant) or non-cancer (benign). Regarding the BreaKHis images, such regions are not segmented or labeled, i.e., the exact location is unknown.

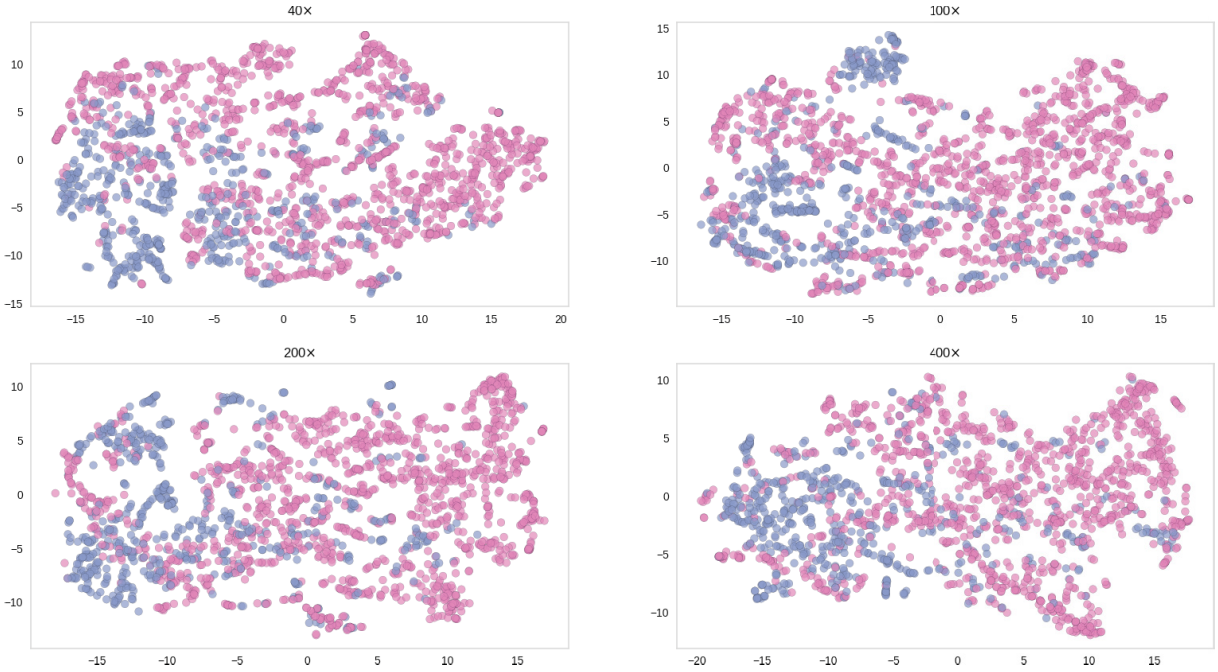
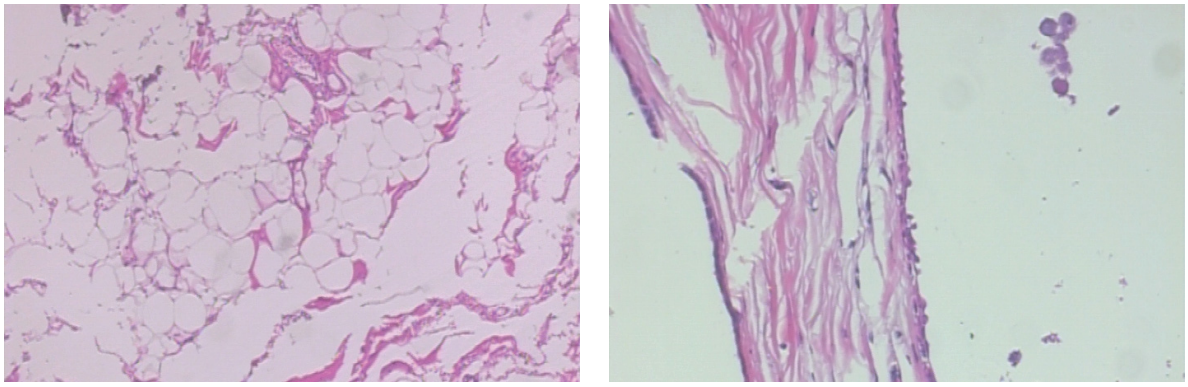


Figure 4.3: Visualization of the BreaKHis dataset in 2D space by t-SNE. Each point represents an image. The benign instances are shown in blue and the malignant in magenta. $Perplexity = 40$. **Source:** The author (2016).

Irrelevant Images: Some BreaKHis images do not contain enough visual information to classify such image as a malignant or benign tumor. Figure 4.4 depicts two examples of irrelevant images: (a) an image acquired from a malignant ductal carcinoma, at 40× magnification factor, and (b) an image from a benign adenosis at 100×. Both isolated images cannot be classified with confidence, since they present only common tissue or even almost all background.

4.3 Baseline System Using Hand-crafted Features

Aiming to evaluate the discrimination power of traditional representations, we built a conventional classification system using supervised learning and some of the well known classic texture descriptors found in the surveyed literature. Such a global approach to extract features tries to avoid the explicit segmentation of the histopathological images, a non-trivial and error-prone task. The representations used as textural descriptors were summarized in Chapter 2, Section 2.2.



(a) Irrelevant image, at 40×, extracted from a malignant slide.

(b) Irrelevant image, at 100×, extracted from a benign slide.

Figure 4.4: Examples of images irrelevant to classification.
Source: The author (2017).

We also intend to investigate the possible complementarity among the four magnification factors of the images available on the dataset, by emulating the analysis made by the pathologist, who starts the observation using smaller zoom levels and gradually goes to larger zooms, seeking details on tissue components. An intuitive approach suggests the use of diverse distinct representations for each magnification factor. Also, it is necessary to establish a strategy to combine such representations into a single descriptor capable of improving the recognition rate.

Usually, this feature combination problem is treated by using two basic methods: combining the features before the classification (early fusion or feature level fusion) or combining the output of the classifiers built upon different descriptors (late fusion or decision-level fusion). Considering the first method, a popular technique is concatenating all the features to construct a combined bigger feature vector [287]. Then, this new and normally high dimensional feature space is fed to a classifier. However, such practice can not guarantee the optimal performance [72], since it suffers from imbalanced input features and produces high dimensional descriptors [287]. Regarding the decision-level fusion method, the individual decisions of the classifiers may be combined by voting and averaging or used as input to a final classifier. This fusion method, with respect to our model, will be explored in Section 4.4.1.6.

4.3.1 Traditional Representation

Considering the traditional representation, six different operators were used to extract texture descriptors and they were presented briefly in Chapter 2. These include the most commonly textural descriptors found in the literature, such as Local Binary Patterns (LBP) [207, 208], Completed LBP (CLBP) [110], Local Phase Quantization (LPQ) [209], Grey-Level Co-occurrence Matrix (GLCM) [117], Threshold Adjacency Statistics [69, 115], and one keypoint descriptor, named Oriented FAST and Rotated BRIEF (ORB) [239]. Keypoint descriptors are most often used for object recognition. However, the literature shows that this kind of descriptor can produce interesting results for texture classification on microscopic images [185].

The operators that will be employed to extract texture descriptors are shown in Table 4.2. We will use such representations to train the classifiers presented in Chapter 2, Section 2.3.

Table 4.2: Summary of operators.

Acronym	Operator [reference]	Number of Features
$CLBP\_S/M/C$	Completed LBP: combination of operators $CLBP\_S$ (Sign), $CLBP\_M$ (Magnitude) and $CLBP\_C$ (Center) [110]	1,352
$GLCM$	Gray-Level Co-ocurrence Matrix [117]	13
LBP	Local Binary Pattern [207, 208]	10
LPQ	Local Phase Quantization [209]	256
ORB	Oriented FAST and Rotated BRIEF [239]	32
$PFTAS$	Parameter Free-TAS [69, 115]	162

Source: The author (2015).

4.3.2 Oracle

To verify the level of complementarity among the classifiers used in this study, we will compute the accuracy of the oracle, which is the upper limit in terms of performance of the pool of classifiers. The oracle is an abstract model defined in [162], which always selects the classifier that predicted the correct label, for a given query sample, if such a classifier exists. In other words, it represents the ideal classifier selection scheme. The motivation of reporting the oracle is to show the difference between the performance of a real-life classification system and the

abstract model of the oracle. Better results in oracle can indicate that room for improvement exists with a high potential to increase the accuracy. Considering this scenario, the classification system performance may be improved, for example, by using improved descriptors or designing a strategy to combine appropriate classifiers.

4.4 Automatic Representation Using Deep Learning

As stated in the introduction, the visual classification of histopathological images poses as a challenging task. This type of image presents high variability (inter- and intra-class) as well as complex geometrical structures, due to high morphological diversity and quite complex textures. Thus, the traditional approach to extract features suitable to classify histopathological images requires considerable modeling efforts guided by knowledge domain experts. Generally, the final representation is pretty customized, specific for each problem and hardly applicable in other contexts [278].

To overcome this limitation we will explore the automatic representation based on deep learning techniques, which are expected to extract the features directly from the input images, thus avoiding hand-crafted features [28]. Since the key-concept of deep learning is to discover multiple representation levels in order to extract more abstract semantics directly from the input data [167], we believe that the automatic representation would be a fruitful diversity source to train the classifiers.

In this sense, we selected CNNs, which have achieved success in image classification problems. Two main strategies will be considered:

1. Classifying our histopathological images using classical architectures, such as *LeNet* [169] and *AlexNet* [158], trained from scratch;

2. Using networks trained on different datasets to act as feature extractors in our dataset and potentially reuse the representation learned (transfer learning) [291].

Considering the first alternative, the aforementioned architectures demand many thousands of instances to properly train the network. Unfortunately, there is a lack of enough training images in the BreaKHis dataset for such purpose, a common fact on similar medical applications. To address this limitation we will use *data augumentation* methods aiming to artificially produce additional training instances by applying transformations to the original image [61, 158]. In summary, our data augmentation technique is based on downsampling the original image and extract several *patches* of this resized image. Then these patches will be used to train the CNNs. In order to maintain a compromise between satisfactory results and a feasible computational time, we propose to evaluate different approaches to extract the patches. Initially, methods such as random extraction and sliding window will be considered. A possible drawback in patch extraction approach is how to select the representative patches. Since the histopathological images are highly heterogeneous, some patches can present little textural information as we can see in the Figure 4.5. To overcome this problem we should define a metric based on measures of variability in order to evaluate each patch, by selecting it to compose the training set or discarding it as non-meaningful. In the classification phase, the test image will also be resized and the patches extracted. A similar deep learning method, but applied on texture classification of forest species, was presented by Hafemann *et al.* [114]. In this study, the CNN outperformed most of the earlier published results on a macroscopic image dataset [219] and another one composed of microscopic images [184].

The second alternative also deals with the limited amount of training examples, but transferring the representations learned from one large-scale dataset to other related domains. Training CNNs entirely from scratch would be unfeasible in some contexts, due to the nature of this type of network demanding large amounts of training examples and the contrasting lack of training data of some image datasets. Currently, a common tactic adopted by researchers is to use networks pre-trained on large datasets in order to produce representations which can be applied to classification tasks on small datasets. CNNs pre-trained on large-scale datasets can act as general-purpose feature extractors [251] and the representations learned are potentially transferable to several other related domains (*transfer learning*) [210]. To deal with the problem of scarcity of large medical image databases, Bar *et al.* [20] proposed using a deep learning approach

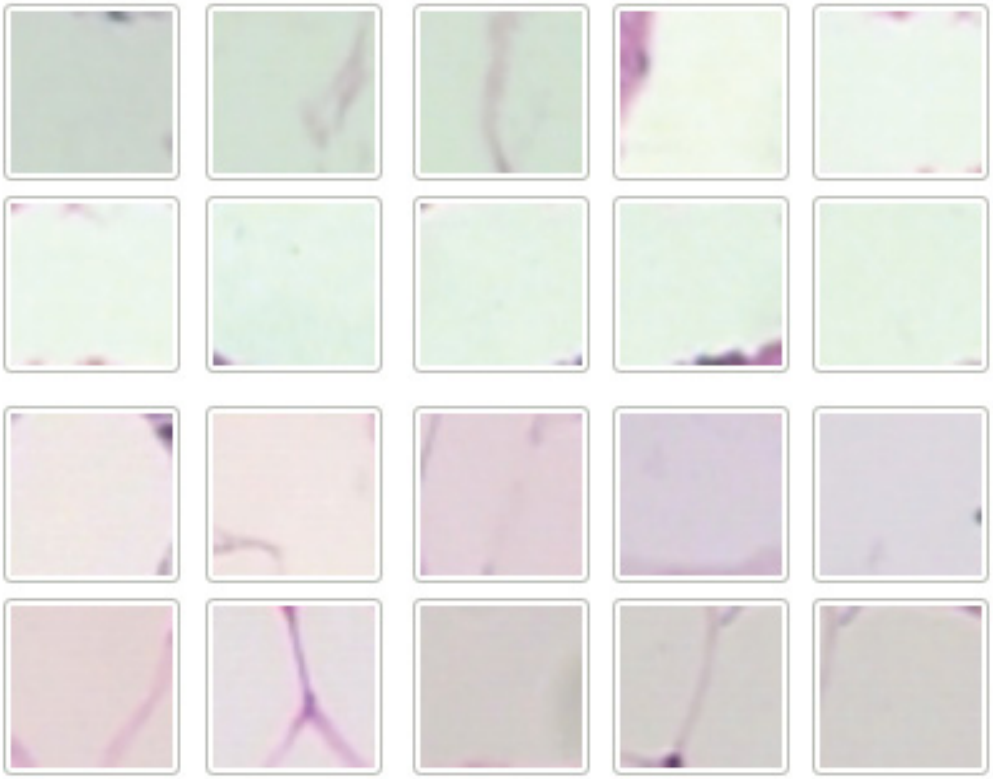


Figure 4.5: Examples of patches extracted from BreaKHis images. In the two upper rows there are patches from benign images. Malignant examples are in the two bottom rows.
Source: The author (2015).

based on non-medical learning and, for this purpose, they have trained a CNN with a well-known large scale non-medical image database. In a recent publication, Xie *et al.* [288] presented a transfer learning approach to extract large-scale socioeconomic indicators from high-resolution satellite imagery, using nighttime light intensities as a data-rich proxy for economic activity. Therefore, the features extracted by a CNN in a particular problem domain, known as *deep features*, can be transferred to different but related domains. If there is sufficient similarity between the domains, the representation learned by the CNN on a given problem can be effective in others as well. Moreover, the results of a recent approach by Simonyan and Zisserman [254] demonstrate that very deep models generalize well to different datasets and the extracted deep features achieve excellent performance even when used as input of other classifiers. Thus, focusing on transfer learning techniques, we expect to evaluate some CNNs pre-trained for image classification challenges like ImageNet Large Scale Visual Recognition Challenge (ILSVRC)[1].

[1] http://image-net.org/challenges/LSVRC/

Specifically, we will start with CNNs models such as *GoogLeNet* [265], *AlexNet* [158], and *VGG* [61, 254].

4.4.1 CNN Approach

A traditional approach to extract appropriate features for classification tasks in pathological images requires considerable efforts and effective expert domain knowledge, frequently leading to highly customized solutions, specific for each problem and hardly applicable in other contexts [278]. On the other hand, as a particular deep learning technique, CNNs have achieved success in image classification problems, exploring the possibility of learning features directly from input data, thus avoiding hand-crafted features [28]. In summary, a CNN consists of multiple trainable stages stacked on top of each other, followed by a supervised classifier. Learning a CNN that is convenient for classification tasks critically depends on the expertise of parameters. In the remaining of this section the method that will be used to classify the images from the BreaKHis dataset using a proper CNN is described.

4.4.1.1 CNN Architecture

In order to classify images from BreaKHis dataset, we will evaluate some previously existing deep neural network architectures. We will start with LeNet [169], a CNN known to work well on digit classification tasks. Then, we will choose a more complex model, specially designed to classify color images. Particularly, we expect to achieve competitive results using a variant based on the AlexNet [158]. The original AlexNet was proposed by Alex Krizhevsky to accurately classify images from the CIFAR-10 [2] dataset. This architecture is composed of multiple layers of convolution, pooling, Rectified Linear Unit (ReLU) nonlinearities, and local contrast normalization with a linear classifier on top of it all, as shown in Figure 4.6.

[2]The CIFAR-10 (Canadian Institute for Advanced Research) dataset consists of 60000 32×32 color images (50,000 for training, 10,000 for testing) in 10 mutually exclusive classes ('truck', 'automobile', 'plane, 'ship', 'cat', 'dog, 'bird', etc.), with 6,000 images per class. Source: `http://www.cs.toronto.edu/~kriz/cifar.html`

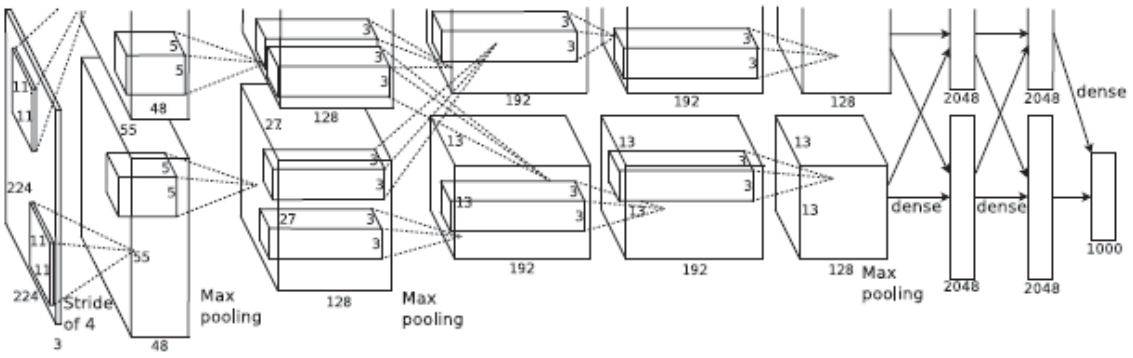


Figure 4.6: AlexNet CNN architecture.
Source: Extracted from [158].

Table 4.3 summarizes the parameters of the CNN layers, where CONV+POOL$_{max}$ stands for Convolutional Layer followed by Max-pooling layer, CONV+POOL$_{avg}$, Convolutional Layer followed by Average-pooling layer, and FC by fully-connected layer. In the end, the CNN architecture that provided the best results in our experiments contains the following layers and parameters:

- *input/data layer*: this layer loads input and produces output used to feed convolutional layers. Some transformations such as mean-subtraction (described below) and feature-scaling can be applied. In our case, inputs are images and the parameters are defining the image dimension (32×32 pixels) and number of channels – 3 for Red-Green-Blue (RGB). As a convenience for classification models, the input layer can also receive the label associated with the image;

- *convolutional layers*: a convolutional layer convolves the input image with a set of learnable filters, each producing one feature map in the output image. There are three convolutional layers in this model. The receptive fields (kernels) are of size 5×5, the zero-padding is set to 2 and the stride is set to 1. Each of the first two convolutional layers learns 32 filters and they are initialized from a Gaussian distribution with a standard deviation of 0.0001 and 0.01, respectively. The last layer learns 64 filters and it is initialized from a Gaussian distribution with a standard deviation of 0.0001.

- *pooling layers*: these layers are responsible for downsampling the spatial dimension of the input. There is one pooling-layer after each convolutional layer. All of them are set to use a

3×3 receptive field (spatial extent) with a stride of 2. The first pooling layer uses the most common max operation over the receptive field and the other two perform average pooling;

- *ReLU layers*: despite the ReLU activation function actually being a non-linear element-wise operator, we will treat it, for convenience, explicitly as a layer. There are three ReLU layers in this model. Given an input value x, the ReLU layer computes the neuron's output $f(x)$ as x if $x > 0$ and $(\alpha * x)$ if $x <= 0$. The parameter α specifies whether to leak the negative part by multiplying it with the slope value (0.01 or so) rather than setting it to 0. The default value of α is 0. So, when this parameter is not set, it is equivalent to the standard ReLU function $f(x) = max(0, x)$ [158], in other words, the activation is simply thresholded at zero;

- *inner-product layers* or *fully connected layers*: these type of layers treat the input as a simple vector and produce an output in the form of a single vector. There are two inner-product layers in this model. The second one, a fully-connected output layer with softmax activation, depends on the number of classes in the classification problem, i.e., 2 output filters for our binary classification problem.

Table 4.3: Summary of CNN layers.

	Layers				
	1	2	3	4	5
Type	$CONV + POOL_{max}$	$CONV + POOL_{avg}$	$CONV + POOL_{avg}$	FC	FC
Channels	32	32	64	64	2
Filter Size	5×5	5×5	5×5	–	–
Convolution Stride	1×1	1×1	1×1	–	–
Pooling Size	3×3	3×3	3×3	–	–
Pooling Stride	2×2	2×2	2×2	–	–
Padding Size	2×2	2×2	2×2	–	–

■ $CONV + POOL_{max}$: Convolutional Layer followed by Max-pooling layer
■ $CONV + POOL_{avg}$: Convolutional Layer followed by Average-pooling layer ■ FC: Fully-connected layer

Source: The author (2015).

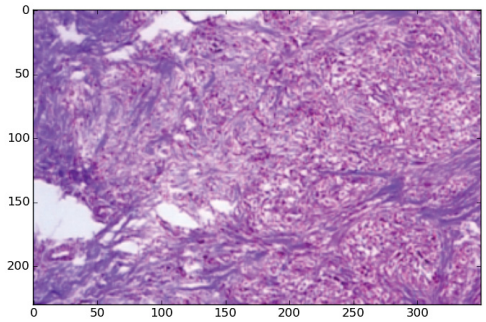
4.4.1.2 Training Strategies Using Small Patches

The proposed method aims at dealing with the high-resolution of the images generally used for histopathological BC classification. As pointed out in [114], adapting the existing deep neural network models for larger images can result in more complex architectures, with larger sets of parameters (more and larger layers), which can substantially increase the complexity of the model. As a consequence, the time that is necessary to fine-tune and train the parameters of the architecture can become very long. In order to deal with this problem, the proposed method is based on the extraction of random patches for training, and the combination of these patches for recognition.

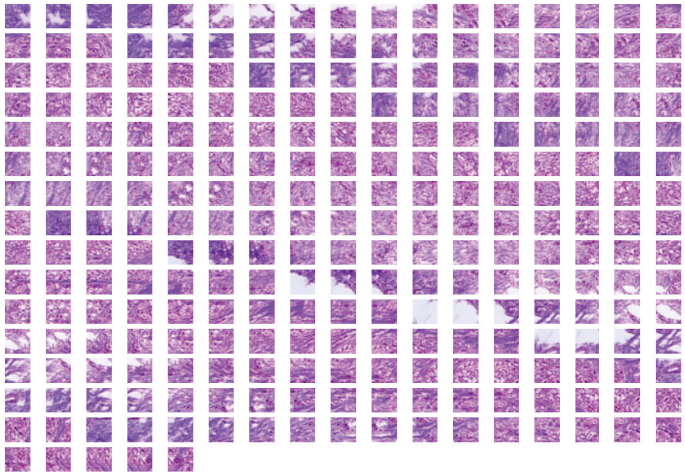
In order to learn the parameters of the CNN described in Section 4.4.1.1, only small patches of the images are used for training. The main idea is to extract patches with sizes that are close to those of the CIFAR dataset from the high resolution images . Since we are dealing with textures, the main premise is that these patches can contain enough information for training a model, provided an appropriate set of patches is extracted from each image.

Based on the results reported by Hafemann *et al.* in [114], where the best results were achieved by reducing the dimensionality of the images, in this work the original 700×460 images will be reduced to 350×230. Afterwards, we will extract patches using two different strategies. In the first one, we will use a sliding window with a 50% overlap, while in the second case the patches will be extracted randomly with no overlap control between patches. Also, based on the results reported in [114], we will assess two different image patch sizes (32×32 and 64×64). Figure 4.7 shows a resized image, as well as the respective 32×32 image patches.

In practice, this method brings translation-invariance to the model and acts as regularization, preventing the model from overfitting the training set. The sliding window strategy, allowing 50% of overlap between patches of 32×32 and 64×64, results in 260 and 54 patches by image, respectively. On the other hand, considering the random extraction strategy, for both patch sizes, we have fixed an arbitrary number of 1000 patches to be extracted from each input image. Table 4.4 summarizes the patch images strategies which we are going to evaluate in our work.



(a) Example of a BreaKHis image, breast malignant tumor stained with HE staining and acquired at 40× magnification, resized (50% of the original).



(b) Patches of dimension 32×32 extracted from resized image

Figure 4.7: Patches extracted from the original image after dimension reduction.
Source: The author (2015).

Table 4.4: Summary of patch image generation strategies

#	Patch Size	Strategy	Number of Patches
1	32×32	Sliding Window	260
2	64×64	Sliding Window	54
3	32×32	Random	1000
4	64×64	Random	1000

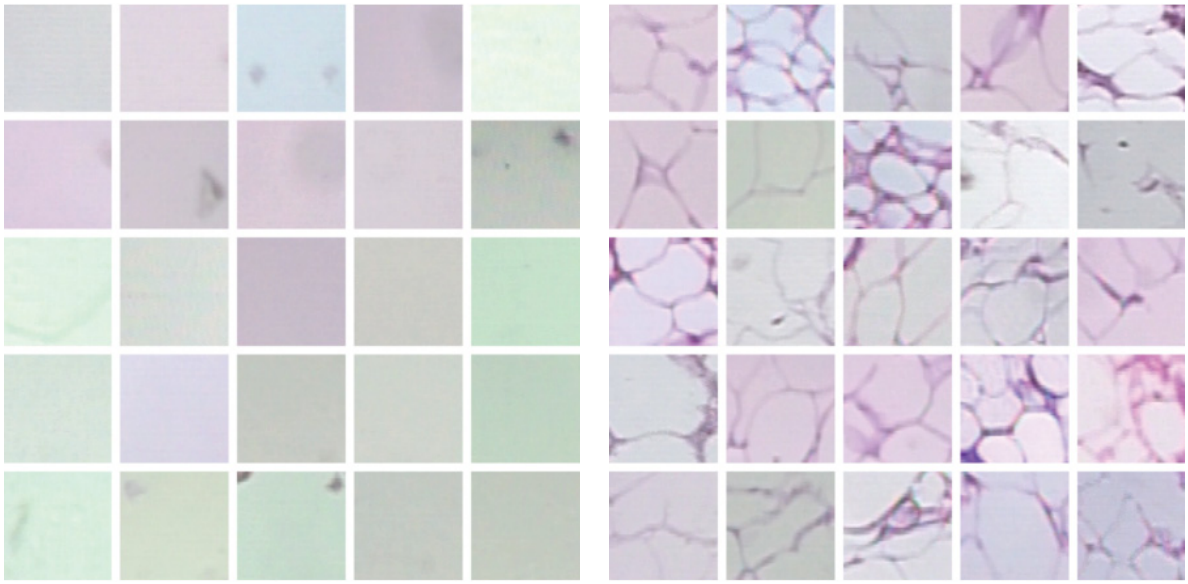
The training protocol adopted here is the purely supervised type, frequent in practical systems for speech and image recognition. As usual in the supervised mode, the Stochastic Gradient Descent (SGD) method [34], with backpropagation to compute gradients and a mini-batch size of 1, will be used to update the network's parameters, starting with a learning rate of 10^{-6}, in conjunction with a momentum term of 0.9 and a weight decay of 4^{-5}. The CNN will be trained for 80,000 iterations on a NVIDIA® Tesla® K40m GPU [205] using the Caffe framework [141].

4.4.1.3 Limitation of the Patch-based Strategy

Avoiding the necessity of performing an explicit segmentation of the images, the CNN method is based on the extraction of several small random patches (sub-images) for training,

and the combination of these patches for recognition. Considering this automatic representation approach, some promising results were achieved as presented in Chapter 5.

However, it was noted that the patterns of some patches are quite similar, independently if the patch comes from a malignant or a benign image. Examples of these patterns can be seen in Figure 4.8. This intersection can be partially explained by the natural presence of certain types of tissue (such as adipose tissue, collagen-rich stroma, connective tissue etc.) in most of the breast histopathological samples.



(a) Patches from background areas. (b) Patches from adipose tissue.

Figure 4.8: Examples of recurrent patterns.
Source: The author (2017).

Moreover, a significant number of extracted patches present almost plain texture, which normally corresponds to tissue spaces representing areas where liquid is formed, the central cavity of a tubular or other hollow structure or even areas without tissue (background). An example of such type of patch presenting homogeneous texture can be seen in Figure 4.9.

Definitely, selected patches containing only such common tissue pattern aren't adequate to distinguish between benign and malignant tumors. In this vein, we also assess an alternative strategy named patch filtering, which tries to discard irrelevant patches and provide only representative examples to train the CNN.

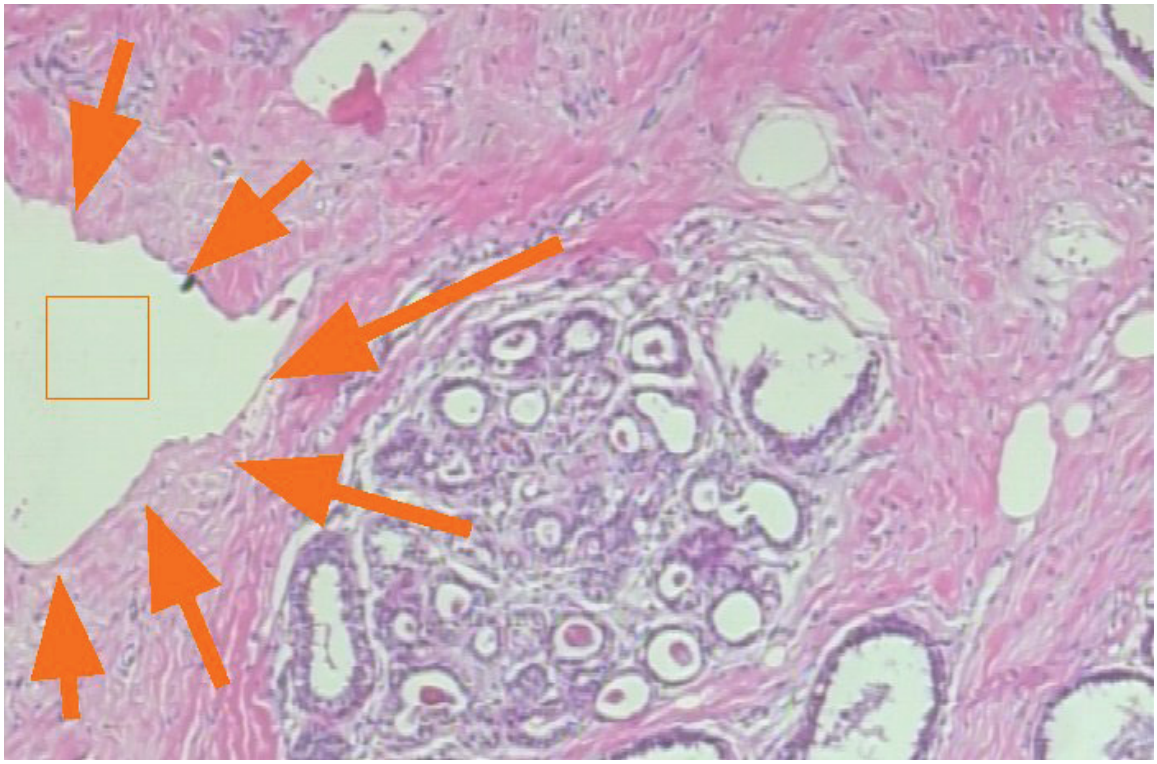


Figure 4.9: Example of selected patch inside a flat texture area of the image.
Source: The author (2017).

4.4.1.4 Patch Filtering

As previously discussed, in order to cope with non-discriminant patches — irrelevant to the classification, we propose to discard those patches presenting common patterns to both classes (see Figure 4.10). A possible approach is to use a clustering technique in the patches, aiming to separate them into three groups: pure malignant, pure benign and "mixed". Thus, only the "pure" groups will be considered as CNN's input for training. It is expected that using more discriminative sub-images improves the recognition rate.

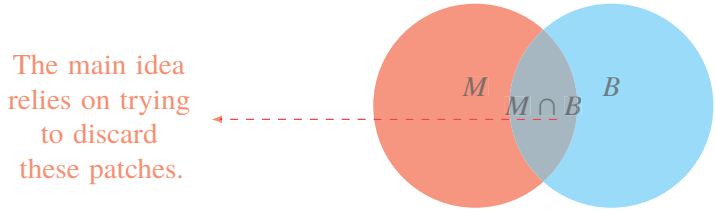


Figure 4.10: Malignant and benign class overlapping.
Source: The author (2017).

This method aims to evaluate each cluster, discarding those containing mixed patterns, and the general definition is as follows.

By using the previously extracted patches, we apply a pre-trained CNN to extract Deep Convolutional Activation Feature (DeCAF) features [17, 66, 81, 233] from the *ip1* layer of the selected network. This layer is an InnerProduct layer (usually referred to as the fully connected layer) that treats the input as a simple vector and produces an output in the form of a single vector. Considering this model, we have a 64-dimensional feature vector v_i per patch, saved in $\mathcal{V}$.

Once the representation from the *ip1* layer is extracted, we use a simple k-means clustering algorithm to create k clusters on the input dataset. A partitional clustering is performed, i.e., a division of the set of data instances (feature vector) into non-overlapping subsets (clusters). We expect to divide the patches into distinct groups, based on properties of coherence and similarity automatically learned by the CNN.

Thus, using the feature vectors in $\mathcal{V}$, we create the clusters C and evaluate the "purity" rate p^C for each cluster, discarding the clusters presenting purity less than a fixed threshold.

To do so, let's consider k the total number of clusters, $C = \{C_1, \ldots, C_k\}$ the set of clusters, $\mathcal{L} = \{L_1, \ldots, L_m\}$ the set of m distinct classes and $L_b = \{P_{1_b}, \ldots, P_{N_b}\}$ the set of labeled extracted patches, where $N_P = \sum_{j=1}^{m} \left|L_j\right| = \sum_{b=1}^{m} N_b$.

By using these assumptions, we compute the "purity" rate p^{C_q} as:

$$p^{C_q} = max\left(\frac{|L_1|}{N_P}, \ldots, \frac{|L_m|}{N_P}\right) \quad \text{such that, } \forall b, L_b \in C_q \tag{4.1}$$

where $1 \leq q \leq k$, and $L_b \in C_q$ represents the patches from L_b in the cluster C_q. Given a threshold λ, we discard the cluster C_q if $p^{C_q} < \lambda$.

Finally, for each remaining cluster, we assign the cluster's label to the patches in this respective cluster. These patches, expected to present more discriminative patterns, will be used as input to train the CNN.

4.4.1.5 Classification

For the recognition, patch results are combined for the whole image. Since the models are trained on the patches of the images, we require a strategy to divide the original test images into patches, run them through the model and then combine the results. The optimal result could be achieved by extracting all possible patches from the images, but this is too intensive computationally. Instead, we chose to extract the grid patches of the images, that is, the set of all non-overlapping patches, which in practice demonstrated reasonable balance between classification performance and computational cost.

Running the model, each patch outputs the probability of each possible class, given the patch image. In order to combine the results of all the patches of a given test image, we are also going to test different fusion rules, such as the well-known *Sum*, *Product* and *Max* rules [149].

4.4.1.6 Fusion

Initially, we expect to explore the limits of each representation approach as an individual model of classification. However, a single-classifier model might not be sufficient to achieve the required accuracy. The combination of multiple classifiers can provide advantages over the traditional monolithic approach to classifier design. Thus, we propose to build a pool of classifiers, using the different representations, and also design a strategy to combine such classifiers. It is expected that this will take advantage of the potential diversity arising from classifiers that use different representation approaches, hand-crafted and automatic. Besides focusing on the fusion of classifiers using the classical techniques proposed in [149], we expect to evaluate the dynamic selection of the classifiers approach [47]. The final goal is to build a hybrid model which explores the potentialities of the different representations.

Considering the automatic representation with CNNs, after classifying each patch individually, an appropriate fusion rule should be used to make the final decision about the original image. Initially, in order to combine the results of the patches extracted from the input image, we will apply fusion rules similar to the ones proposed in [149]. This is possible because the output produced by the CNN is interpreted as the probability of the patch belonging to each of the considered classes. However, other rules should be evaluated to propose the most appropriate. Besides, the fusion of the features extracted using CNNs trained from scratch on the dataset with *deep features*, extracted using CNNs trained on large-scale datasets, could be effective to improve the recognition rate.

4.5 MIL Paradigm

A relatively new learning paradigm called MIL allows for the training of a classifier from partially or ambiguously labeled data. It is the case of the BreaKHis dataset. In fact, the relevance of MIL for this type of application and dataset is naturally described in two different ways:

- **Each image is considered as a bag while its patches (or subimages) are the instances**. In the field of natural scene images, this first approach is related to region-based image categorization, where each instance encodes color, textural or spatial features related to that specific region [122]. In our binary setting, the image would be labeled "positive" (pathological) if it has at least one malignant patch; conversely, an image would be labeled benign if it does not have any portion labeled malignant. This multiple instance formalism is natural, since only a subset of the patches is labeled by the experts, making it possible that entire images might be healthy where the patient is diagnosed with a tumor. This is not the case in the strategy used so far, i.e., a single instance classification setting with instances inheriting the label of their image. Note that the underlying assumption is that image or bag label stems from the patient diagnostic. However, this might not be the case, since entire images might be healthy where the patient is diagnosed with a tumor.

- **Each patient is a bag, with the instances being its associated images or patches**. This makes full sense as the diagnosis (i.e., the label) is established only at the patient level. Furthermore, a patient diagnosed with a malignant tumor can still have some of his images described as tumor-free, i.e., healthy, as just said; and a healthy patient has inevitably all of his images healthy. These facts match the MIL assumption.

We consider the state-of-the-art of MIL methods. In particular we've investigated the seminal Axis-Parallel Hyper Rectangle (APR) [78], and the algorithms based on Diverse Density (DD) [183, 294], k-NN (Citation-kNN) [280] and Support Vector Machine (SVM) [12], as well as a recently-proposed non-parametric algorithm [276], and a deep learning approach revisiting CNN for MIL (MILCNN) [263]. As a second contribution, we've studied how MIL results compare to our single instance classification results. Of course in this case we suppose that instances inherit labels from the bags. We have examined whether it is preferable to cast this problem into a single instance one, or if MIL does indeed bring an added value, both at the image and the patient levels [6].

4.6 Evaluation Metrics

In binary classification problems two possible classes (positive and negative) are considered in the prediction. Thus, there are four possible results for the classification of a test instance, as shown in Table 4.5, grouped in correct predictions: True Positive (TP) or True Negative (TN), and incorrect predictions: False Positive (FP), or False Negative (FN).

Since we are dealing with a binary classification problem (benign vs malignant), as evaluation metrics, we have reported recognition rates and confusion matrices [92]. In this context, let B be a test set with N_B breast tumor images. If the classification system rejects N_{rej}, classifies correctly N_{rec}, and misclassifies the remaining N_{err}, then the rejection rate $RejRate_{img}$,

Table 4.5: Four possible results from two-classes prediction.

		True Class	
		Positive	Negative
Predicted	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

Source: The author (2015).

the recognition rate $RecRate_{img}$, and the error rate $ErrRate_{img}$ for images correctly classified are stated in Equations 4.2, 4.3, and 4.4, respectively.

$$RejRate_{img} = \frac{N_{rej}}{N_B} \times 100 \tag{4.2}$$

$$RecRate_{img} = \frac{N_{rec}}{N_B} \times 100 \tag{4.3}$$

$$ErrRate_{img} = \frac{N_{err}}{N_B} \times 100 \tag{4.4}$$

Using these definitions we also compute the reliability rate $RelRate_{img}$, as defined in equation 4.5.

$$RelRate_{img} = \frac{RecRate_{img}}{RecRate_{img} + ErrRate_{img}} \times 100 \tag{4.5}$$

Besides, the errors committed by the system can be analyzed in terms of False Positive Rate (FPR) and False Negative Rate (FNR). The FPR is the ratio (percentage) of negative

instances predicted as positive. The FNR measures the likelihood that a positive instance is incorrectly classified as negative.

$$FPR = \frac{FP}{FP + TN} \tag{4.6}$$

$$FNR = \frac{FN}{FN + TP} \tag{4.7}$$

However, considering that the final interest is to classify patients rather than images, we have also reported the recognition rate for patients correctly classified. Let $I^{P,M}$ be a test set of N_P breast tumor images of patient P for magnification M. If the classification system classifies correctly N_{rec}, then

$$PatientScore = \frac{N_{rec}}{N_P} \tag{4.8}$$

$$RecRate_{pat} = \frac{\sum PatientScore}{TotalPatients} \tag{4.9}$$

Finally, a common assumption in the use of recognition rate as an evaluation metric is that the class distribution among examples is constant and relatively balanced. This is not the case in this work. To better analyze the results for unbalanced problems, we have also drawn Receiver Operating Characteristic (ROC) curves, which are attractive because they are insensitive to changes in the class distribution. If the proportion of positive to negative instances changes in a test set, the ROC curves will not change [92]. So, along with the recognition rates, we present the confusion matrices and ROC curves with Area Under the ROC Curve (AUC). Using normalized units, AUC is the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative example [92]. This metric ranges between 0 and 1, and it is useful when comparing the global behavior of different classifiers.

4.7 Software Libraries

In order to extract the texture descriptors, build the baseline classification system and perform the respective experiments, our implementation uses the python language associated with specific libraries for image processing and machine learning. Thus, the following software tools were used:

- Mahotas [68], a computer vision and image processing library for python programming language;
- Milk, a machine learning toolkit for python, created by Luis Pedro Coelho, Mahotas's author;
- Scikit-learning [220], an open source machine learning library for python;
- Scikit-image [279], an open and free image processing library for python;
- OpenCV (Open Source Computer Vision) [38], an open source multi-platform computer vision library originally developed by Intel Corporation.

Considering the deep learning approach, we've also used Caffe [141], a deep learning framework BSD-licensed C++ library with python/numpy bindings for training and deploying general-purpose CNNs and other deep models. Caffe is developed and maintained by the Berkeley Vision and Learning Center (BlVC). This framework uses Nvidia CUDA for GPU computation, but the same models can be run in CPU or GPU mode.

Finally, for all MIL methods except the non-parametric and the MILCNN, we have used the implementation of the Jun Yang's MIL Library[3] with MATLAB 2017a. The non-parametric MIL algorithm was obtained from the author's website[4]. For the implementation of MILCNN in Python, Keras and Theano were used [65].

[3] CMU MIL toolbox: `http://www.cs.cmu.edu/~juny/MILL/`
[4] `https://github.com/ragavvenkatesan/np-mil`

Chapter 5

Experiments

This chapter focuses on a series of experiments performed using BreaKHis dataset [256]. Initially, the adopted experimental protocol is defined. After that, the results obtained by a conventional supervised classification system using hand-crafted features are presented and discussed. This conventional baseline system constitutes the basis for comparison with other approaches. Next, a summary of the results achieved by applying the automatic representation approach is presented, as well as a comparison with the best conventional system results. Additionally, some fusion strategies of the classifiers are presented in order to assess possible complementarity among the representations aiming to improve the performance. Finally, we investigate the Multiple Instance Learning (MIL) framework applied on the BreaKHis dataset, providing an extensive comparative analysis of the results achieved by MIL methods.

5.1 Experimental Protocol

We have adopted the experimental protocol published in [256]. Experiments conducted in this work are based on a subgroup of images randomly selected from the original image dataset described in Section 4.2.

The BreaKHis dataset has been randomly divided into a training (70%) and a testing (30%) set. To make sure the classifier generalizes to unseen patients, the system ensures that patients used to build the training set are not used for the testing set. Aiming to avoid the bias of using just "easier" instances (or more "difficult" ones) in each set of training and test, the images were distributed into five distinct trials (#1 − #5). Following the standard labeling convention used in medical studies, the label `positive` (resp. `negative`) refers to malignant (resp. benign) images. All trials include positive samples (malignant tumors) and negative samples (benign tumors). The positive group includes ductal carcinoma (most frequent), lobular carcinoma, mucinous carcinoma and papillary carcinoma. The negative group includes, in an undistinguished way, benign tumors like fibroadenoma, adenosis, tubular adenoma, and phyllodes tumor. This image distribution across the five trials is presented in Table 5.1. The results presented in this Chapter are the average of five trials. This protocol was applied independently to each of the four available magnifications.

Considering the Convolutional Neural Network (CNN) approach, aiming to handle the image high resolution (752×582) and augment data for training, the images were divided into small patches. Different sizes of patches were assessed and the size of 64×64 pixels has been shown to be particularly relevant for CNN-based classification. For training, 1000 patches have been randomly extracted from each input image. For test, we have chosen to extract the grid patches of the images, that is, the set of all non-overlapping patches, which in practice demonstrated reasonable balance between classification performance and computational cost.

Regarding the clustering analysis, a starting point to discard irrelevant patches, for each cluster, the most prevalent class of its elements is defined and the purity rate of the cluster regarding this class is computed. The strategy consists of eliminating those clusters containing a mixture of patches from distinct classes given a purity rate, i.e., only clusters presenting a purity rate greater than a threshold λ are kept. The formal notation was stated in Equation 4.1. Then, those remaining patches are used to create a new dataset $\mathcal{D}'$, split into training $\mathcal{D}'_T$ and testing $\mathcal{D}'_V$ sets. The set $\mathcal{D}'_T$ was used to train a CNN built to classify input patches into two classes: benign, and malignant. We have assessed the number of clusters k such that $k \in \{10, 20, 50\}$.

Analyzing the distribution of the patches into clusters, $k = 50$ it was chosen to keep low dense clusters. Also, different values of purity rate setting the threshold λ has been evaluated. It was noticed that assuming high purity rates, i. e. $\lambda \geq 0.99$, caused the total elimination of patches in certain folds. Thus, a $\lambda = 0.9$ was chosen.

Finally, for MIL experiments, the aforementioned patches form the instances, whereas bags will be considered at two levels: at the patient level, patches will be collected independently from all the patient's images; at the image level, patches will be originating from the image of interest. For this purpose, each patch is described with a 162-long feature vector of Parameter-Free TAS (PFTAS) features [69, 115]. These features have shown to be particularly relevant for the BreaKHis dataset, when assessed against many others such as Local Binary Pattern (LBP), Completed Local Binary Pattern (CLBP), Local Phase Quantization (LPQ), Gray-Level Co-occurrence Matrix (GLCM), as well as computer vision features, such as Oriented FAST and Rotated BRIEF (ORB) [256]. The hyper-parameters for each MIL method were optimized using grid search as follows.

For Axis-Parallel Hyper Rectangle (APR) [78]:

- Kernel Width: 0.999;
- Outside Probability: 0.023;
- GridNum: 25000.

For Diverse Density (DD) [183] and Expectation Maximization-Diverse Density (EM-DD) method [294]:

- Scaling: 1;
- Aggregate: average;
- Threshold: 0.5;
- No. of runs: 100 (DD), 500 (EM-DD);
- Iteration Tolerance (for EM-DD): 0.08.

For Citation-kNN [280]:

- Bag Distance Type: minimum;
- Instance Distance Type: Euclidean;
- Reference nodes considered: 5;
- CiterRank: 11.

For mi-SVM and MI-SVM [12]:

- Kernel: Linear, poly, RBF;
- KernelParam - NA/degree/gamma: (NA), 4, 0.32 (mi-SVM), (NA), 5, 0.17 (MI-SVM);
- CostFactor: 1/0.96/1 (mi-SVM), 1/1/1 (MI-SVM);
- NegativeWeight: 1/1/1;
- Threshold: 0.5/0.5/0.5.

For non-parametric MIL [276]:

- Averaged accuracy over 100 runs;
- Range of k for grid search: 50 (1–50) using elbow method;
- No. of Tsteps: 3000;
- Distance Method: Euclidean.

For MILCNN [263]:

- the structure is the same as the Multiple Instance Learning Convolutional Neural Network (MILCNN) for CIFAR10/100.

Table 5.1: Summary of images in datasets.

		Images in datasets – grouped by magnification and trial									
		Magnification									
		40×					100×				
		#1	#2	#3	#4	#5	#1	#2	#3	#4	#5
Benign	Training dataset	343	387	355	351	348	360	398	357	356	368
	Testing dataset	236	192	224	228	231	240	202	243	244	232
		200×					400×				
		#1	#2	#3	#4	#5	#1	#2	#3	#4	#5
	Training dataset	341	390	339	347	349	313	375	319	317	324
	Testing dataset	234	185	236	228	226	226	164	220	222	215
		40×					100×				
		#1	#2	#3	#4	#5	#1	#2	#3	#4	#5
Malignant	Training dataset	808	899	926	901	855	858	969	999	982	898
	Testing dataset	514	423	396	421	467	534	423	393	410	494
		200×					400×				
		#1	#2	#3	#4	#5	#1	#2	#3	#4	#5
	Training dataset	833	938	919	926	876	758	814	840	827	760
	Testing dataset	511	406	425	418	468	432	376	350	363	430
		40×					100×				
		#1	#2	#3	#4	#5	#1	#2	#3	#4	#5
Total	**Training dataset**	1,151	1,286	1,281	1,252	1,203	1,218	1,367	1,356	1,338	1,266
	Testing dataset	750	615	620	649	698	774	625	636	654	726
		200×					400×				
		#1	#2	#3	#4	#5	#1	#2	#3	#4	#5
	Training dataset	1,174	1,328	1,258	1,273	1,225	1,071	1,189	1,159	1,144	1,084
	Testing dataset	745	591	661	646	694	658	540	570	585	645

Source: The author (2015).

5.2 Baseline: Results of the Conventional Approach

Considering the conventional system, Table 5.2 reports the performance of all classifiers and descriptors we have assessed, at image level and patient level respectively. Since the medical decision is patientwise, we focus on the discussion at the patient level, and not at the image level.

We propose a two-level analysis of Table 5.2. Let us first focus on the influence of the magnification factors, by comparing columns (bold results). Interestingly, the magnification factors do not seem to have the same level of information. In particular, the first level (40×) exhibits the best results over CLBP, LBP and ORB. This slight tendency that 40× may be the most informative magnification factor is in accordance with the pathologist behavior, which starts by examining at factor 40 and switches to the next level, until he established his diagnosis. Note however that the 200× magnification factor also shows high potential, with the best results

over GLCM and PFTAS, higher than those obtained with the 40× level. The complementarity of the magnification factors may fruitfully be investigated in the future, for example, through a coarse-to-fine analysis.

The other level of analysis concerns the feature vector comparison (best results are underlined in Table 5.2). All feature vectors exhibit stable and close results. These results are little influenced by the classifiers: for each magnification factor and for each feature vector, the recognition rates of the four classifiers are in a range of less than 4%. However, considering the CLBP descriptor, the results fall out of this interval, e.g., recognition rates range from 41.5% to 75.6% for factor 40×. In particular, note that the results obtained by CLBP with Quadratic Discriminant Analysis (QDA) are far below the other mean recognition rates. Indeed, QDA is based on the estimation of covariance matrices: in order to make a proper estimation of these matrices, a large amount of samples is required, which should be all the greater given that CLBP is high dimensional (1,352).

As previously reported in Table 5.2, over all the feature vectors, the PFTAS performs best. Since the best overall performance (recognition rate of 85.1% for factor 200×) is achieved by the Support Vector Machine (SVM) trained with PFTAS descriptors, we focus on the SVM/PFTAS association and further analyze their performance, by drawing the associated ROC curve (Figure 5.1) and reporting the confusion matrices in Table 5.3, which confirms that 200 seems to be the most discriminant magnification factor. As we can see, most of the confusion occurs when a benign tumor is classified as malignant (high false positive rate). This may be partially explained, as pointed out by Kowal *et al.* [151], by the fact that one of the benign tumor present in the dataset (fibroadenoma) shares similar properties with a malignant tumor. In order to verify this hypothesis, we have performed an error analysis on the SVM/PFTAS results, which is summarized in Table 5.4. This analysis shows that independently of the magnification factor, about 30% of the errors committed by the classifier are due to benign tumors fibroadenoma classified as malignant class. One example of this misclassification is presented in Figure 5.2, where (a) shows a benign tumor classified as a malignant tumor and (b) presents a real malignant tumor. From these images we can see how complex the classification problem is.

Table 5.2: Mean recognition rates and standard deviations (at image level and patient level) of the classifiers trained with different descriptors. In bold the best results over the magnification factors are highlighted. For each magnification factor, underlined, with a gray background, the 5 best results are shown over the feature vectors and classifiers.

Accuracy at	Descriptor	Classifier	Magnification Factors			
			40×	100×	200×	400×
Patient Level	CLBP	1-NN	**73.6 ± 2.5**	71.0 ± 2.8	69.4 ± 1.5	70.1 ± 1.3
		QDA	39.4 ± 13.5	**51.7 ± 17.3**	50.3 ± 16.0	49.4 ± 15.5
		RF	**74.5 ± 0.7**	72.5 ± 3.8	70.0 ± 2.4	72.3 ± 2.1
		SVM	**77.4 ± 3.8**	76.4 ± 4.5	70.2 ± 3.6	72.8 ± 4.9
	GLCM	1-NN	74.7 ± 1.0	76.8 ± 2.1	**83.4 ± 3.3**	81.7 ± 3.3
		QDA	67.0 ± 6.0	74.2 ± 3.5	**78.6 ± 1.7**	77.0 ± 2.3
		RF	73.6 ± 1.5	76.0 ± 1.9	**82.4 ± 2.3**	79.8 ± 2.5
		SVM	74.0 ± 1.3	78.6 ± 2.6	**81.9 ± 4.9**	81.1 ± 3.2
	LBP	1-NN	**75.6 ± 2.4**	73.0 ± 2.4	72.9 ± 2.3	71.2 ± 3.6
		QDA	69.7 ± 3.8	69.7 ± 4.2	68.8 ± 4.7	**72.3 ± 4.6**
		RF	**74.0 ± 2.9**	73.1 ± 1.9	70.1 ± 2.5	70.7 ± 4.3
		SVM	**74.2 ± 5.0**	73.2 ± 3.5	71.3 ± 4.0	73.1 ± 5.7
	LPQ	1-NN	72.8 ± 4.9	71.1 ± 6.4	**74.3 ± 6.3**	71.4 ± 5.2
		QDA	**70.4 ± 1.1**	69.3 ± 4.2	67.2 ± 1.9	68.3 ± 1.8
		RF	**73.8 ± 5.0**	72.3 ± 5.5	73.4 ± 5.9	71.1 ± 3.8
		SVM	**73.7 ± 5.5**	72.8 ± 5.0	73.0 ± 6.6	**73.7 ± 5.7**
	ORB	1-NN	**71.6 ± 2.0**	69.3 ± 2.0	69.6 ± 3.0	66.1 ± 3.5
		QDA	**74.4 ± 1.7**	66.5 ± 3.2	63.5 ± 2.7	63.5 ± 2.2
		RF	**72.3 ± 1.8**	69.3 ± 1.0	68.6 ± 1.7	67.6 ± 1.2
		SVM	**71.9 ± 2.3**	69.4 ± 0.4	68.7 ± 0.8	67.3 ± 3.1
	PFTAS	1-NN	80.9 ± 2.0	80.7 ± 2.4	**81.5 ± 2.7**	79.4 ± 3.9
		QDA	83.8 ± 4.1	82.1 ± 4.9	**84.2 ± 4.1**	82.0 ± 5.9
		RF	81.8 ± 2.0	81.3 ± 2.8	**83.5 ± 2.3**	81.0 ± 3.8
		SVM	81.6 ± 3.0	79.9 ± 5.4	**85.1 ± 3.1**	82.3 ± 3.8
Image Level	CLBP	1-NN	**69.7 ± 3.1**	66.5 ± 2.7	66.8 ± 2.5	66.7 ± 3.3
		QDA	41.5 ± 11.8	**51.6 ± 14.0**	50.6 ± 14.2	49.1 ± 12.6
		RF	**69.8 ± 2.1**	**69.8 ± 4.6**	67.1 ± 2.9	68.1 ± 3.2
		SVM	**75.6 ± 4.9**	74.0 ± 5.9	68.1 ± 4.3	69.3 ± 5.7
	GLCM	1-NN	74.5 ± 1.8	74.8 ± 3.5	**82.0 ± 2.2**	78.6 ± 3.4
		QDA	66.9 ± 5.2	72.2 ± 2.0	**77.7 ± 2.4**	75.2 ± 2.0
		RF	72.0 ± 3.1	73.7 ± 2.1	**81.5 ± 1.7**	77.2 ± 2.8
		SVM	74.6 ± 2.1	76.5 ± 3.3	**80.9 ± 3.4**	78.2 ± 3.2
	LBP	1-NN	**70.3 ± 3.2**	68.2 ± 2.8	67.9 ± 2.6	66.9 ± 5.2
		QDA	66.4 ± 4.2	65.6 ± 4.7	65.3 ± 4.4	**67.9 ± 5.2**
		RF	**68.3 ± 3.2**	67.9 ± 3.7	67.0 ± 2.8	65.5 ± 4.6
		SVM	**68.9 ± 4.4**	68.3 ± 4.6	67.5 ± 4.0	67.0 ± 6.2
	LPQ	1-NN	69.8 ± 5.1	66.8 ± 6.1	**70.3 ± 5.8**	66.9 ± 5.2
		QDA	**67.3 ± 2.3**	66.3 ± 5.1	65.4 ± 3.2	64.4 ± 3.8
		RF	**69.0 ± 5.3**	67.6 ± 5.7	**69.0 ± 5.1**	67.8 ± 5.4
		SVM	**70.6 ± 6.7**	68.9 ± 5.9	69.5 ± 6.3	69.1 ± 7.3
	ORB	1-NN	**69.7 ± 2.1**	65.5 ± 3.5	67.0 ± 2.8	62.9 ± 2.1
		QDA	**72.8 ± 1.8**	63.0 ± 3.3	61.8 ± 2.4	61.8 ± 2.1
		RF	**70.1 ± 2.3**	65.4 ± 2.8	66.2 ± 2.1	63.8 ± 1.0
		SVM	**69.9 ± 2.3**	65.7 ± 2.4	66.0 ± 1.7	62.9 ± 2.0
	PFTAS	1-NN	79.1 ± 2.1	77.8 ± 3.0	**79.6 ± 1.9**	77.6 ± 4.0
		QDA	82.8 ± 3.6	80.7 ± 4.9	**83.3 ± 3.0**	80.5 ± 5.6
		RF	80.2 ± 1.9	80.4 ± 3.8	**82.4 ± 2.3**	80.0 ± 4.5
		SVM	79.9 ± 3.7	77.1 ± 5.5	**84.2 ± 1.6**	81.2 ± 3.6

Source: The author (2015).

Table 5.3: Confusion matrices produced by the SVM classifier trained with the PFTAS descriptor. B: benign, M: malignant.

	40×		100×		200×		400×	
	B	M	B	M	B	M	B	M
B	0.51	0.49	0.38	0.62	0.69	0.31	0.62	0.38
M	0.06	0.94	0.03	0.97	0.08	0.92	0.08	0.92

Source: The author (2015).

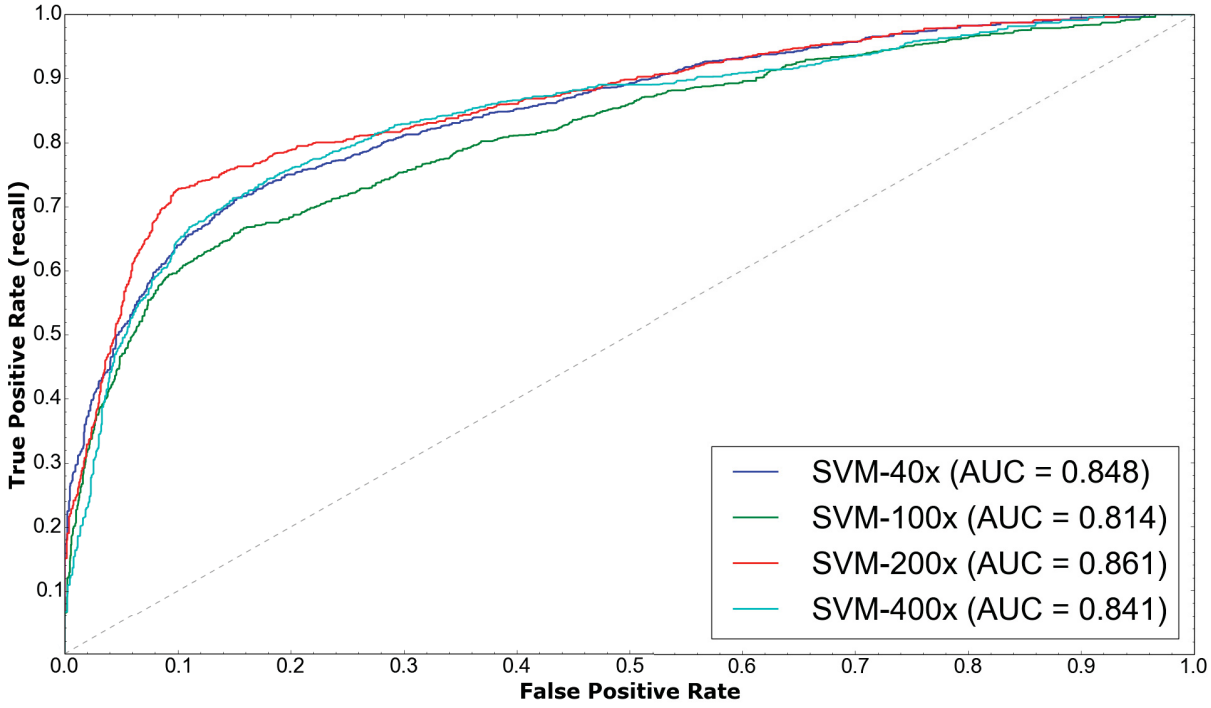


Figure 5.1: ROC curves for the confusion matrices presented in Table 5.3.
Source: The author (2015).

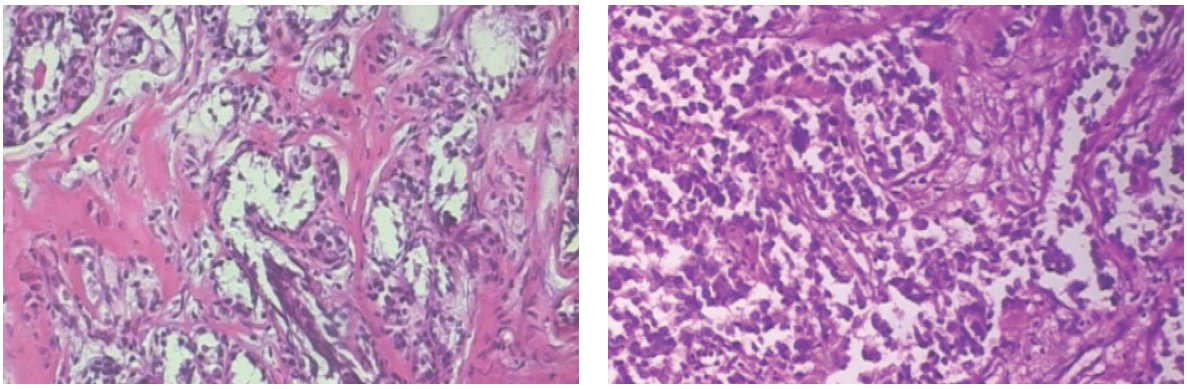
5.3 Results of the Automatic Representation Approach

As expected, on the histopathological images assessed, LeNet classification performance were considerably inferior to our previous results reported in [256], achieving about 72% of accuracy. So we focus on the other architecture. Among a few tested, the model which presented the best performance was a variant based on the AlexNet [158]. The training of these CNN models, described in 4.4.1, took about 40 minutes for the sliding window strategy and 3 hours for the random patch strategy, which contains a much bigger training set. Table 5.5 reports the accuracy of the CNNs at both patient and image levels, as defined in Equations 4.9 and

Table 5.4: Error distribution (%) of the SVM trained with PFTAS over subclasses. A large amount of false positive comes from fibroadenoma (benign) mistaken for malignant tumor.

Class	Subclass	Magnification Factors			
		40×	100×	200×	400×
Benign	Adenosis	15.7	21.7	9.7	10.3
	Fibroadenoma	**28.5**	**31.8**	**29.5**	**30.2**
	Phyllodes tumor	13.6	18.6	10.1	14.4
	Tubular Adenoma	23.1	19.5	15.6	16.5
Malignant	Ductal	11.6	2.8	13.9	8.7
	Lobular	0.0	0.0	0.2	3.2
	Mucinous	2.8	5.1	13.9	10.1
	Papillary	4.7	0.5	7.1	6.6

Source: The author (2015).



(a) Original image of benign tumor (HE stained) classified as a malignant tumor.

(b) Image of a real malignant tumor (HE stained).

Figure 5.2: Example of misclassification between benign and malignant breast tumors.
Source: The author (2015).

4.3. From these experimental results, we can see that the CNN outperforms almost all of the classifiers/descriptors used in the conventional system (Figure 5.3).

From Table 5.5 we may notice that training the CNN with a large number of 64×64 image patches extracted randomly from the image (strategy #4) seems to be a suitable strategy for low magnification factors such as 40× and 100×. In the case of the 40× magnification factor, the CNN was able to achieve an accuracy of about 5% better than the best result reported in Table 5.2. For higher magnification factors, though, training the CNN with a large number of image patches brings no benefit. In those cases, all strategies achieve similar results, which are also comparable to the ones reported in Table 5.2.

Table 5.5: Mean recognition rates and standard deviations (patient and image levels) of the CNN trained with the strategies presented in Table 4.4. The best results are in bold, and on a gray background.

Accuracy at	Strategy	Magnification Factors			
		40×	100×	200×	400×
Patient Level	1	80.5 ± 1.6	81.0 ± 3.0	**85.3 ± 3.8**	81.0 ± 1.5
	2	81.0 ± 1.9	82.8 ± 2.8	83.7 ± 2.8	81.1 ± 3.2
	3	81.7 ± 2.9	83.5 ± 5.0	82.9 ± 3.6	81.4 ± 5.1
	4	**88.6 ± 5.6**	**84.5 ± 2.4**	83.3 ± 3.4	**81.7 ± 4.9**
Image Level	1	79.9 ± 2.6	80.8 ± 3.7	**84.0 ± 3.2**	80.7 ± 1.8
	2	80.6 ± 2.1	81.0 ± 3.0	82.7 ± 1.9	**80.8 ± 3.1**
	3	81.8 ± 3.3	82.3 ± 4.9	82.4 ± 2.8	80.3 ± 4.0
	4	**89.6 ± 6.5**	**85.0 ± 4.8**	82.8 ± 2.1	80.2 ± 3.4

Source: The author (2015).

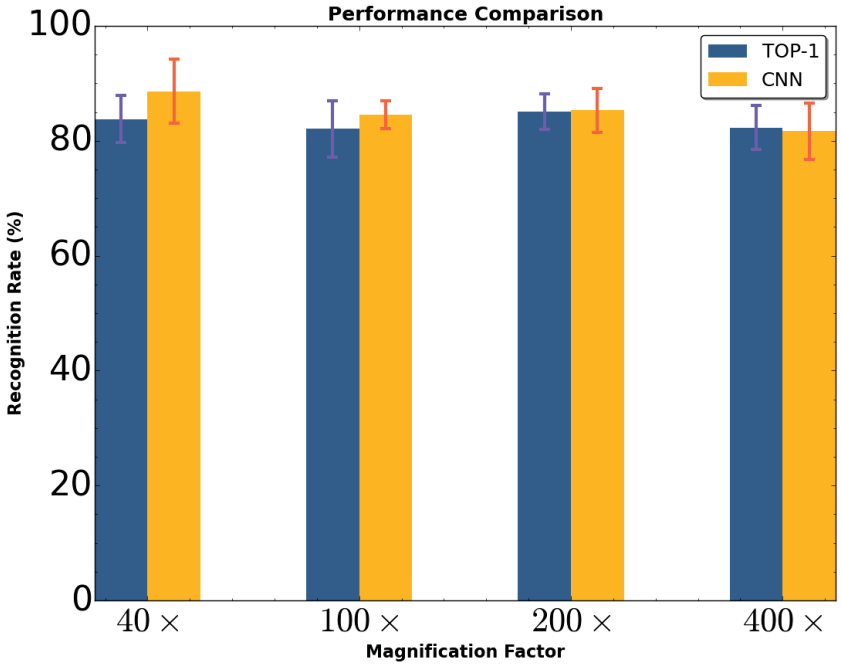


Figure 5.3: Recognition rates achieved by top-1 classifiers trained using conventional feature extraction compared to the CNN approach.
Source: The author (2015).

Considering the misclassifications, a similar behavior to that noticed in the conventional approach, has been observed in the automatic representation approach as well. Table 5.6 presents the confusion matrices produced by the network, and again there is a high proportion of benign tumors erroneously classified as a malignant type (high false positive rate).

Table 5.6: Confusion matrices produced by the CNN. B: benign, M: malignant.

	40×		100×		200×		400×	
	B	M	B	M	B	M	B	M
B	0.63	0.37	0.54	0.46	0.61	0.39	0.53	0.47
M	0.08	0.92	0.08	0.92	0.06	0.94	0.05	0.95

Source: The author (2015).

5.4 Combination

Initially, in order to verify the level of complementarity among the classifiers used in conventional system, we have computed the accuracy of the oracle, which is the upper limit in terms of performance of the pool of classifiers. Using this abstract fusion model, Table 5.7 shows the upper limit of the classifiers and the representations adopted in this work.

As we can see, despite the intrinsic complexity of the problem, the performance of the oracle is very high. Considering a single architecture of classifier trained with six different representations, the upper limit of the system achieves an average of 93%, except for the QDA classifier that reached 99.5% for the subset of 40× magnification images. When combining all the architectures and representations (24 experts) the upper limit increases to an average up to 99%. Note that for the 40× magnification, all test images could be correctly classified by at least one of the classifiers in the pool. However, getting a high oracle result does not imply anything about how feasible it is to get those results by combining the classifiers, but how feasible it is to get those results by dynamically selecting the classifier. Dynamic selection of the classifier is an alternative to the traditional combination of classifiers that has been successfully applied to complex pattern recognition problems [47].

Table 5.8 presents the hypothetical confusion matrices for the oracle. As we can see, the proposed pool of classifiers is able to solve most of the confusion. The challenge now lies in defining a winner strategy to combine or select the classifiers, given an input image.

Considering the automatic representation approach, since each network was trained with different inputs (i.e., size and number of patches), each classifier builds its own representation,

Table 5.7: Summary of accuracy of the oracle (%). The first 4 lines show the oracle for each classifier using six different representations. The last line reports the oracle considering all the 24 classifiers reported in Table 5.2.

Classifier	Magnification factor			
	40×	100×	200×	400×
1-NN	93.6	91.8	91.0	90.8
QDA	99.5	96.3	97.2	96.3
RF	92.3	89.5	90.6	90.3
SVM	93.1	90.8	93.2	92.6
All classifiers	100	98.7	98.8	98.8

Source: The author (2015).

Table 5.8: Hypothetical confusion matrices for the oracle. B: benign, M: malignant.

	40×		100×		200×		400×	
	B	M	B	M	B	M	B	M
B	1.00	0.00	0.96	0.04	0.96	0.04	0.97	0.03
M	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00

Source: The author (2015).

which gives us the perspective of improving such results through the combination of classifiers. As stated before, the CNNs have a final fully-connected layer with softmax activation that allows us to interpret the outputs of the networks as an estimation of the posterior probabilities. Therefore, different combination rules may be applied. In this work, we report the results obtained when combining the four patch image generation strategies, using the well-known *Sum*, *Product* and *Max* rules (see [149] for details).

Regarding the performance at image level, Table 5.9 shows that all combination rules produce very similar results and that none of them surpass the individual results reported in Table 5.5. On the other hand, the combination brings interesting improvements for all magnification factors (except the 200×) at patient level. The most noticeable result is for the 100× magnification factor, where the improvement is of about 4% and 6% when compared to the best CNN, and the best result presented in Table 5.2, respectively. In these cases, the *Max* rule outperforms the *Sum* and *Product* rules.

Table 5.9: Combination of CNNs using different fusion rules (at Patient and Image Levels). The best results are in bold, and on a gray background.

Accuracy at	Fusion Rule	Magnification Factors			
		40×	100×	200×	400×
Patient Level	*Sum*	88.4 ± 7.6	88.4 ± 4.8	83.8 ± 2.8	85.3 ±5.6
	Product	89.2 ± 7.4	88.4 ± 4.8	83.8 ± 2.8	85.3 ±5.6
	Max	**90.0 ± 6.7**	**88.4 ± 4.8**	**84.6 ± 4.2**	**86.1** ±6.2
Image Level	*Sum*	85.4 ± 5.2	83.3 ± 4.3	**83.1 ± 1.9**	**80.8 ± 3.0**
	Product	85.5 ± 5.3	83.4 ± 4.3	83.0 ± 1.8	**80.8 ± 3.0**
	Max	**85.6 ± 4.8**	**83.5 ± 3.9**	82.7 ± 1.7	80.7 ± 2.9

Source: The author (2017).

5.5 Patch-Filtering

In Table 5.10, since the results achieved by the approach based on hand-crafted descriptors [256] are a baseline, we compare our pre-training CNN patch filter method with the task-specific CNN method [257] and a method based on the Deep Convolutional Activation Feature (DeCAF) features [258]. We can observe that, despite some punctual gains, using the filtering-patches strategy did not bring general significant improvements in the CNN-based performance.

5.6 Results of the MIL Approach

We provide results for two different settings, as aforementioned. In the first setting, each patient is considered as a bag, which is labeled with its diagnosis, and the instances are the patches extracted from the images. As can be seen in Table 4.1, in average 25 images are available for each patient, at each magnification factor. Since around 100 patches are extracted per image in the testing, each bag (or patient) contains around 2500 instances. In the second setting, we consider each image as a bag; in this case, the instances are the patches, and a bag contains approximately 100 instances.

Table 5.10: Patch-filtering results. The best results are in bold, and on a gray background are the cases at which higher results are achieved using the patch-filtering method, compared to those presented in [256]. Results from the combination of multiple classifiers are marked with *.

%	Approach	Magnification factor			
		40×	100×	200×	400×
Patient	Hand-crafted visual descriptors[256]	83.8 ± 4.1	82.1 ± 4.9	85.1 ± 3.1	82.3 ± 3.8
	CNN trained from scratch[257]	88.6 ± 5.6	84.5 ± 2.4	85.3 ± 3.8	81.7 ± 4.9
	Combination of CNNs trained from scratch[257]*	**90.0 ± 6.7**	**88.4 ± 4.8**	84.6 ± 4.2	**86.1 ± 6.2**
	DeCAF features[258]	84.0 ± 6.9	83.9 ± 5.9	86.3 ± 3.5	82.1 ± 2.4
	Patch-Filtering Method	86.4 ± 5.7	83.6 ± 5.8	**92.1 ± 7.3**	85.0 ± 4.7
Image	Hand-crafted visual descriptors[256]	82.8 ± 3.6	80.7 ± 4.9	84.2 ± 1.6	81.2 ± 3.6
	CNN trained from scratch[257]	**89.6 ± 6.5**	**85.0 ± 4.8**	84.0 ± 3.2	80.8 ± 3.1
	Combination of CNNs trained from scratch[257]*	85.6 ± 4.8	83.5 ± 3.9	83.1 ± 1.9	80.8 ± 3.0
	[258]	84.6 ± 2.9	84.8 ± 4.2	84.2 ± 1.7	81.6 ± 3.7
	Patch-Filtering Method	85.3 ± 3.3	82.5 ± 3.0	**87.8 ± 4.9**	**82.1 ± 3.4**

Source: The author (2017).

The results are presented in Table 5.11 and Figures 5.4 and 5.5. Despite the large standard deviation values, we can attempt to observe some trends. As expected, DD-based approaches and APR yielded the poorest results which led us to think that positive instances are not clustered in a single area of the feature space. For SVM-based approaches, MI-SVM leads to enhanced results, which shows that a bag level paradigm is better suited to the data. At last, best classification rates are reported with the non-parametric MIL approach, especially for low magnification factors, and at the patient level.

Table 5.11: Summary of accuracy of MIL approach (%) at respective levels. Best results columnwise are in bold, and on a gray background.

	Patient as bag				Image as bag			
	40×	100×	200×	400×	40×	100×	200×	400×
Iterated-discrim APR [78]	73.8 ± 3.8	66.5 ± 4.1	84.2 ± 4.9	68.0 ± 5.6	70.4 ± 2.4	65.1 ± 5.0	81.3 ± 5.5	67.3 ± 4.9
DD [183]	70.5 ± 6.1	64.5 ± 4.3	68.3 ± 3.6	71.2 ± 3.3	71.2 ± 5.9	66.1 ± 5.4	66.7 ± 2.9	70.8 ± 3.8
EM-DD [294]	78.3 ± 5.6	80.6 ± 5.2	77.1 ± 6.3	78.7 ± 5.7	73.1 ± 5.4	76.4 ± 4.8	78.2 ± 5.2	76.2 ± 5.6
Citation-kNN [280]	73.7 ± 4.6	72.8 ± 5.4	75.7 ± 3.1	77.2 ± 3.6	73.1 ± 4.3	73.0 ± 5.7	71.3 ± 3.5	78.7 ± 3.1
mi-SVM Linear [12]	79.5 ± 4.3	83.4 ± 4.6	83.6 ± 4.7	81.0 ± 5.2	72.6 ± 4.4	80.6 ± 3.7	80.1 ± 4.9	78.2 ± 5.3
mi-SVM poly [12]	75.2 ± 6.1	79.8 ± 4.8	76.5 ± 3.9	68.5 ± 5.1	75.6 ± 5.7	78.7 ± 4.0	75.2 ± 5.6	69.2 ± 4.8
mi-SVM RBF [12]	77.8 ± 1.6	75.4 ± 1.5	73.8 ± 2.3	72.9 ± 3.4	77.9 ± 2.2	77.3 ± 2.1	74.6 ± 2.9	71.4 ± 3.9
MI-SVM Linear [12]	85.6 ± 5.6	82.1 ± 5.9	84.6 ± 4.8	80.9 ± 4.9	79.5 ± 4.1	78.2 ± 4.4	80.8 ± 4.7	78.9 ± 5.1
MI-SVM poly [12]	84.8 ± 2.7	82.5 ± 4.6	83.9 ± 4.2	81.3 ± 4.2	86.2 ± 2.8	82.8 ± 4.8	**81.7 ± 4.4**	**82.7 ± 3.8**
MI-SVM RBF [12]	79.0 ± 2.1	71.9 ± 2.9	76.2 ± 1.9	73.0 ± 3.5	78.3 ± 3.2	72.2 ± 3.0	76.8 ± 1.6	71.9 ± 2.4
Non-parametric [276]	**92.1 ± 5.9**	**89.1 ± 5.2**	**87.2 ± 4.3**	82.7 ± 3.0	**87.8 ± 5.6**	**85.6 ± 4.3**	80.8 ± 2.8	**82.9 ± 4.1**
MILCNN [263]	86.9 ± 5.4	85.7 ± 4.8	85.9 ± 3.9	**83.4 ± 5.3**	86.1 ± 4.2	83.8 ± 3.1	80.2 ± 2.6	80.6 ± 4.6

Source: The author (2018).

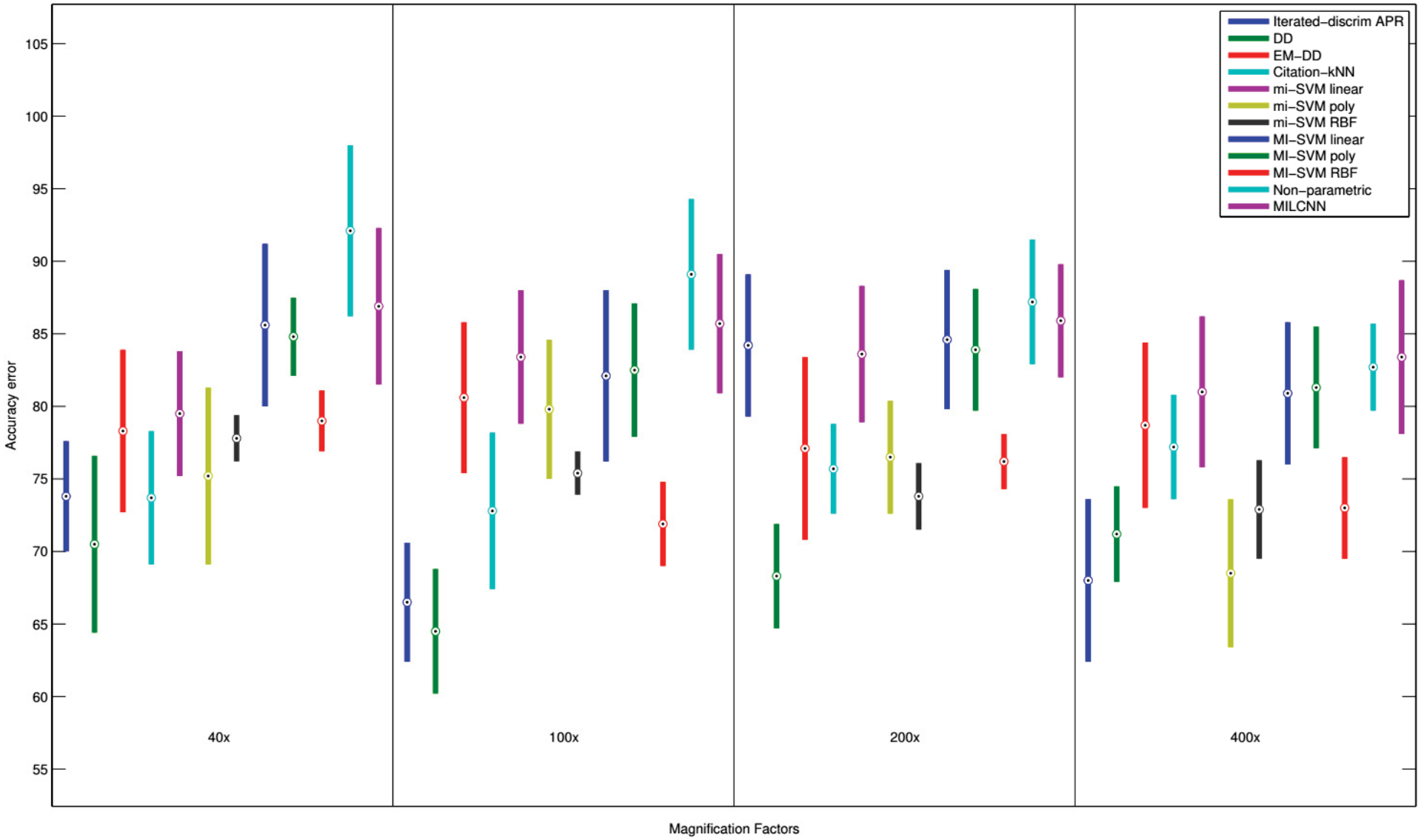


Figure 5.4: Accuracy results of MIL benchmark with patient as bag, corresponding to the left side of Table 5.11. Best viewed in color.
Source: The author (2018).

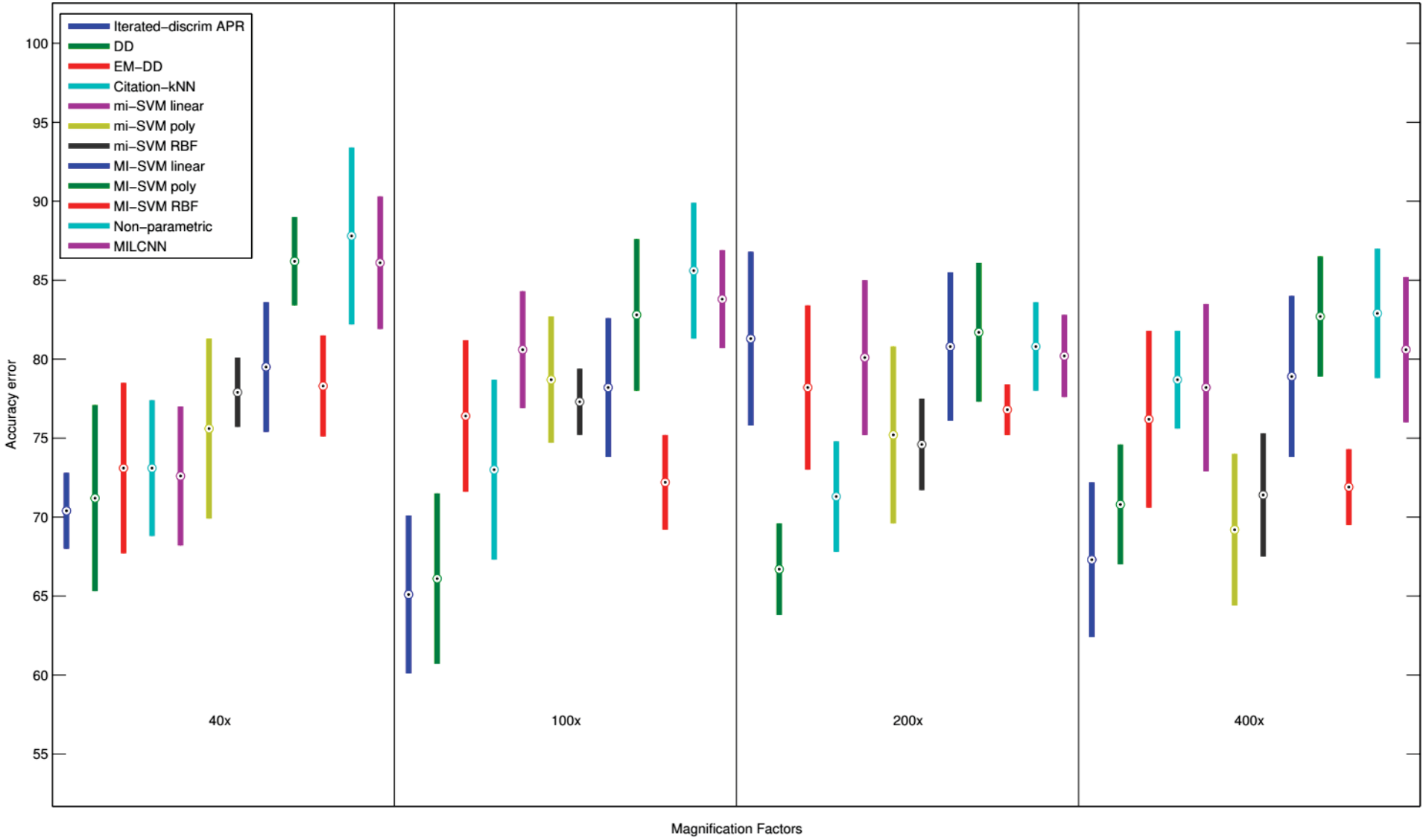


Figure 5.5: Accuracy results of MIL benchmark with image as bag corresponding to the right side of Table 5.11. Best viewed in color.
Source: The author (2018).

5.6.1 MIL vs SIL

In this section, we compare the best MIL result (non-parametric MIL) with a Single Instance Learning (SIL) setting using the same protocol (same trials, same training and test set distribution). In Table 5.12, we report results, using state-of-the-art classifiers, namely 1-Nearest Neighbor (1-NN),QDA, Random Forest (RF), and SVM, partly obtained from previous experiments [256]. Hyperparameters of these classifiers were tuned using grid search and only the best results were retained. These classifiers take as input the PFTAS feature vector describing each image. For the CNN approach, we used AlexNet [158, 257]. Decisions are taken on each patch and are fused together using the *Max* Fusion Rule.

Unsurprisingly, the CNN performs better than other machine learning models trained with hand-crafted textual descriptors (in accordance with [116]; however, their results are not comparable since they do not use the same folds), see Figure 5.6 and Figure 5.7. We observe that the non-parametric MIL brings interesting improvements for all magnification factors (except the 400×) at patient level, while remaining comparable to the CNN results. This suggests that instances, namely patches, may provide only partial, complementary information for the image or the patient level [6], and that a bag-based analysis is valuable for the analysis of histopathology images.

Table 5.12: Comparison of MIL (non-parametric) vs single instance (SIL) classification at respective levels. Best results columnwise are in bold, and on a gray background.

		Patient as bag (MIL) or level (SIL)				Image as bag (MIL) or level (SIL)			
		40×	100×	200×	400×	40×	100×	200×	400×
MIL	Non-parametric	**92.1 ± 5.9**	**89.1 ± 5.2**	**87.2 ± 4.3**	82.7 ± 3.0	**87.8 ± 5.6**	**85.6 ± 4.3**	80.8 ± 2.8	**82.9 ± 4.1**
SIL	CNN	90.0 ± 6.7	88.4 ± 4.8	84.6 ± 4.2	**86.1 ± 6.2**	85.6 ± 4.8	83.5 ± 3.9	**83.1 ± 1.9**	80.8 ± 3.0
	1-NN	80.9 ± 2.0	80.7 ± 2.4	81.5 ± 2.7	79.4 ± 3.9	79.1 ± 2.1	77.8 ± 3.0	79.6 ± 1.9	77.6 ± 4.0
	QDA	83.8 ± 4.1	82.1 ± 4.9	84.2 ± 4.1	82.0 ± 5.9	82.8 ± 3.6	80.7 ± 4.9	**83.3 ± 3.0**	80.5 ± 5.6
	RF	81.8 ± 2.0	81.3 ± 2.8	83.5 ± 2.3	81.0 ± 3.8	80.2 ± 1.9	80.4 ± 3.8	82.4 ± 2.3	80.0 ± 4.5
	SVM	81.6 ± 3.0	79.9 ± 5.4	85.1 ± 3.1	82.3 ± 3.8	79.9 ± 3.7	77.1 ± 5.5	84.2 ± 1.6	81.2 ± 3.6

Source: The author (2018).

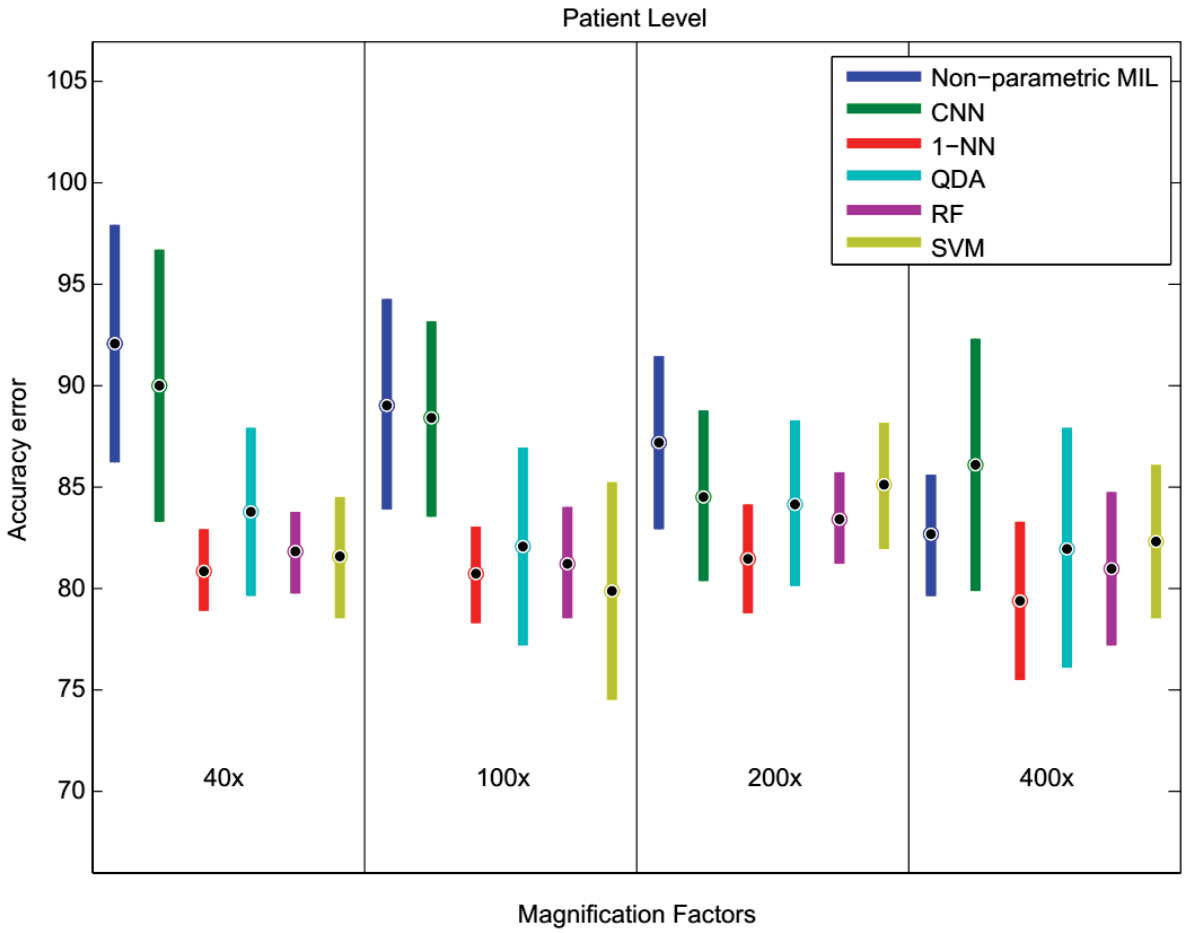


Figure 5.6: Accuracy results of MIL compared to SIL at patient level, corresponding to the left side of Table 5.12. Best viewed in color.
Source: The author (2018).

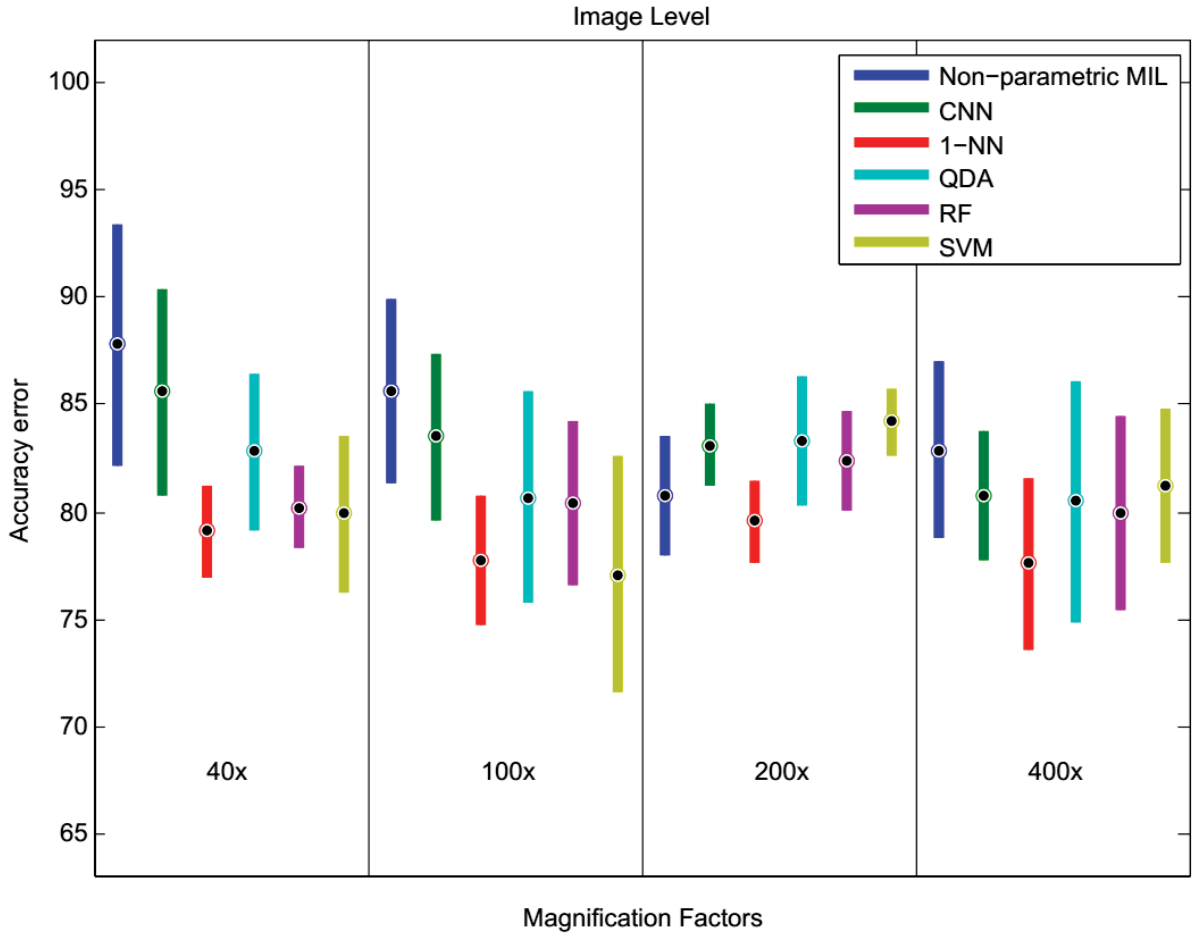


Figure 5.7: Accuracy results of MIL compared to SIL at image level corresponding to the right side of Table 5.12. Best viewed in color.
Source: The author (2018).

Chapter 6

Final Considerations

By building a baseline classification system for histopathological images, we have performed extensive experiments with the BreakHis dataset. This system uses six hand-crafted texture descriptors and four classic classifiers. By adopting an appropriate experimental protocol and pipeline processing for two-class classification of benign vs. malignant images, we have obtained promising results. However, there is room for improvement and additional challenges remain.

Furthermore, a deep learning approach was used to classify BreakHis images following the same experimental protocol. We have shown that we could use an existing Convolutional Neural Network (CNN) architecture and adapt it to the classification of breast cancer (BC) histopathological images. We have also proposed several strategies for training the CNN architecture, based on the extraction of patches obtained randomly or by a sliding window mechanism, that allows to deal with the high-resolution of these textured images without changing the CNN architecture designed for low-resolution images. Our experimental results, obtained on the BreaKHis dataset, showed improved accuracy achieved by CNN when compared to traditional machine learning models trained on the same dataset but with state-of-the-art texture descriptors. Future work can explore different CNN architectures and the optimization of the hyperparameters. Also, strategies to select representative patches in order to improve the accuracy can be explored.

Multiple Instance Learning (MIL) has been proved successful empirically for learning problems with label ambiguity where existing supervised or semi-supervised learning approaches are insufficient. In fact, MIL provides a classification framework that is particularly adapted to computer-aided diagnosis based on histopathological image analysis. In the case of the BreaKHis dataset, several hundreds images are available per patient. The patient can thus be considered as a bag, which is labeled with its diagnosis. Our MIL benchmark shows that the recently proposed non-parametric MIL is particularly efficient for the tasks of patient and image classification. Patient classification rates can reach up to 92.1% for the 40× magnification factor, a level never reached by conventional classification frameworks, which enhances the fact that instances are complementary and can be fruitfully considered in a MIL framework. In summary, non-parametric not only makes intuitive sense, but can also be a powerful tool for most general cases and magnification factors with histopathological images. MIL can thus leverage digital histopathological image classification and analysis to improve computer-aided diagnosis.

6.1 Future Work

This section presents some potentially promising directions for future works.

Creating a strategy for multiclass classification. A much more challenging goal would be to create a strategy for multiclass classification. We have performed experiments considering binary classification of instances, *i.e.*, classify instances (images individually or group of images belonging to a patient) into benign or malignant class. However, as stated in Chapter 2 Section 2.1, the histopathology classifies breast tumors into distinct types, which determines their suitable clinical treatment. Thus, a valid contribution would be to concentrate research efforts in order to propose an automatic method to classify, correctly and with confidence, a given test instance into one of the four malignant types or the four benign types.

Determining the tumor grading. Currently, pathologists determine the tumor grading by assessing the spatial organization of the tissue (e.g., distribution of cancer cells, nuclei morphological properties, interaction with the stroma, etc.). These parameters are evaluated in

small sample regions of the microscopic slide given a score considering some "scoring system" such as Nottingham Histologic Score System[1]. Aiming to complement the final decision, which determines the prognosis and the proper clinical intervention, respective hormone receptor status by Immunohistochemistry (IHC) is also analyzed in IHC-stained sections. Moreover, for binary classification of breast tumor (benign vs malignant) only the epithelial regions of the tissue are relevant, whereas for tumor grading and quantification of IHC, just tumoral tissue should be analyzed [277]. So, automatic BC diagnosis systems based on histopathological images should perform the detection and segmentation of tissue and its components.

Exploring different CNN architectures. We are currently engaged in experimenting other deep learning frameworks. With the acceleration of proposals in this area, no doubt that some more efficient networks will be proposed in the near future. Moreover, the oracle results also show that a single-classifier might not be enough, and that designing a strategy to combine or select the classifiers given an input image should help increase the accuracy.

Investigating MIL for histopathological image segmentation. We also want to investigate MIL for histopathological image segmentation. MIL can indeed be an adequate framework to find the location of the malignant region position in histopathological images [142, 152, 218, 290]. In digital histopathology, it is mostly applied when we have labels for an image (bag of pixels) and not the pixels themselves, where we are interested in learning a model that can classify the pixels. For instance, if the input is a dataset of images taken from a tumor or from benign specimens, but the desired output is the segmentation of the cancerous regions of an image, one can use MIL. Since manual labeling is too long, MIL can help in pixel labeling and clustering, and can serve as a feedback to the pathologist. In this case, the image is considered as a bag and the pixels as instances.

[1]Nottingham (or Elston-Ellis) modification of the Scarff-Bloom-Richardson grading system.

Appendix A

Fundamentals of Breast Cancer

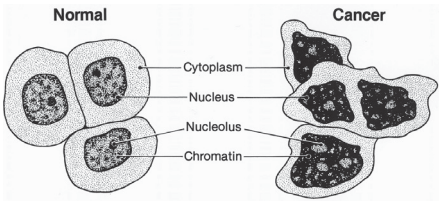
This text reviews three great subjects: cancer, pathology, and breast cancer, which is more detailed than the other two previous topics.

A.1 Cancer

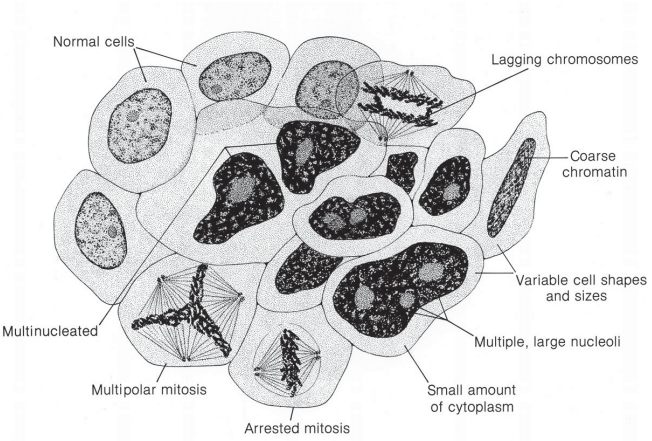
The cellular growing in living organisms can occur through ordered processes (hyperplasia, dysplasia and metaplasia) or disordered processes (neoplasia). Cancer [1] is a common designation for more than one hundred diseases which share the characteristic of accelerated and disordered cellular replication. As an abnormal cellular division process, the neoplasia [190] produces an accumulation of abnormal cells, a tissue mass ordinarily called tumor (or neoplasm). Oncology is the medical specialty focused on the study of tumors [160]. Figure A.1(a) shows a simplified cellular structural organization in normal and cancerous cells. Normal cells are characterized by large cytoplasm, single nucleus, single nucleolus and fine chromatin. Cancer cells, in turn, show the opposite: large size of the nucleus compared to the total cell size, small cytoplasm, multiple nuclei, multiple and large nucleoli, and coarse chromatin. In Figure A.1(b) cancer cells are illustrated undergoing cell division, showing abnormal mitosis.

Considering their biological behavior, tumors can be classified as benign or malignant. When the neoplasm presents a relatively slow and localized growth, in general, it does not represent a risk to the patient's life, and it is named benign tumor [238]. But, in spite of this denomination, eventually benign tumors may cause severe diseases [160], even the patient's death [238]. Other types of neoplasm are characterized by aggressive behavior, showing very fast growth, invading adjacent tissues and organs (local destructive capacity) and, eventually,

[1] The word cancer is the Latin translation of the Greek word *karkinos* which means grab. This designation was originally used because cancers are often very irregularly shaped, and, like a crab, they "grab on and don't let go."



(a) Structure of normal and cancer cells.



(b) Illustration of normal and cancer cells side-by-side. Cell division process on cancerous cells are leading to nuclear abnormalities.

Figure A.1: Major cellular characteristic structures in normal and cancer cells.
Source: Creator Pat Kenny (Illustrator) [203]. Modified by the author (2014).

even migrating to non-contiguous and distant sites (metastasis[2])[161]. This aggressive type of tumor is named malignant and cancer is a synonym.

There are distinct cancer types, using their own denominations which correspond to different types of body cells. Whether cancer begins in epithelial tissues, such as skin or mucosae, it is named *carcinoma*, and, if the epithelium origin is the secretory glandular type, such as found in the lung, stomach, liver, pancreas, etc., the tumor is called *adenocarcinoma* [190]. *Sarcoma* is the denomination for cancer originated from connective or other nonepithelial tissues such as bones, muscles or cartilages. Breast cancer is named carcinoma (with various subtypes) [126] and is detailed in Section A.3.

In order to feed the accelerated growth of its cancerous cells, the tumor itself releases substances to keep blood vessels opened and it also induces the formation (angiogenesis-inducing molecules) of brand new vessels, in a process known as *neoangiogenesis* or tumoral *angiogenesis*. Acquisition of capacity of angiogenesis by cancer cells is considered the most critical step in tumor growth and metastasis [231]. Figure A.2 shows a simplified schematic view of a tumor-associated angiogenesis.

[2]Metastasis is tumoral growing in non-contiguous (distant) site and independent of the original tumor.

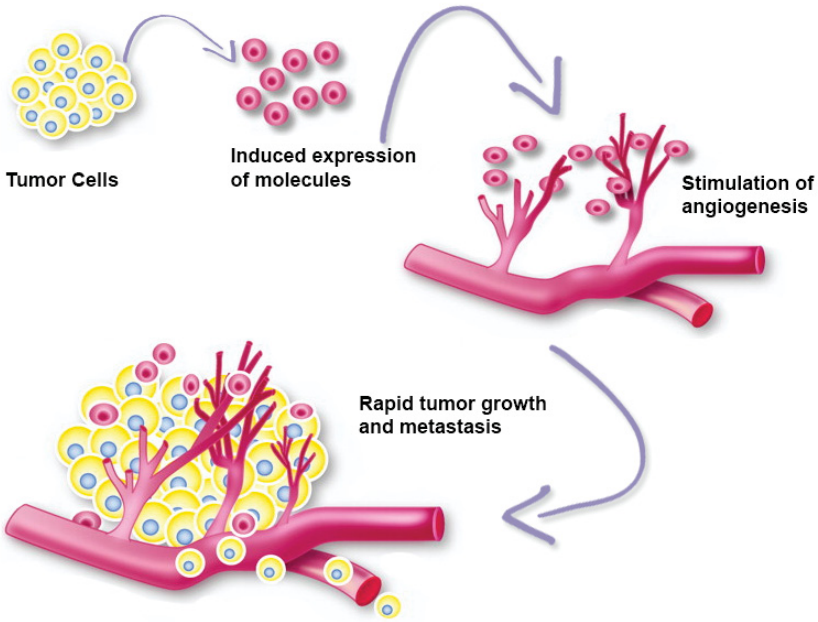


Figure A.2: Schematic presentation of tumor-induced angiogenesis.
Source: Based on [231]. Modified by the author (2014).

A.2 Pathology

Congregating basic science and clinical practice, pathology is focused on the study of structural and functional changes in cells, tissues and organs caused by diseases, being a rational systematization to support the diagnosis, prognosis and treatment of such illnesses [161]. Considering sample types, pathology can be divided into two main branches: (a) histopathology and (b) cytopathology. In Brazil, any cancer treatment will only initiate after malignancy confirmation given by a cyto- or histopathologic report [190].

A.2.1 Histopathology

Histopathology is the study of illness indicatives using by microscopic inspection of tissue samples prepared and fixated onto glass slides. These samples came from puncture biopsies[3] or surgical excisions.

In almost all multicellular animals (humans included), tissues are groups of specialized cells which perform the same function, associating themselves in different proportions to constitute organs and systems of these animal bodies [188]. Also, integrating tissues – providing structural and biochemical support to the surrounding cells – there are intercellular substances called extracellular matrix (interstitial matrix and basement membrane).

[3]Biopsies are cell or tissue samples removed during surgery or routine medical procedures.

A standard procedure in histopathology is the visual analysis of tissue sections under the light microscope. However, tissues are normally too thick for light to pass through them and they must be sliced to obtain appropriate thin translucent sections for the microscopic examination. Thus the sample tissue must be prepared for the histological study. The whole procedure, from fixation to observing a tissue in a light microscope, may take from half to two-and-half days [188]. The main goal is to preserve the original tissue structure and the molecular composition. The preparation includes the following steps: (a) fixation, (b) dehydration, (c) clearing, (d) infiltration, (e) embedding and (f) trimming. Figure A.3 shows the basic steps used in tissue preparation for light microscopy. First (a), solutions of chemicals conserve the proteins and inhibit the action of degradative enzymes. Then, all the tissue water is removed by the use of increasingly concentrated alcohol (70% – 100% ethanol) solutions (b), the alcohol is removed (c), the tissue is immersed in melted paraffin (d) and placed in a small mold in order to harden (e). Finally, the resulting paraffin block is trimmed to expose the tissue (f) and a microtome is used for sectioning the block. At the microtomy process, the block is sliced by the steel blade into extremely thin sections (generally at $1 - 10\mu m$[4] thickness). The paraffin slices are placed on the glass slides and allowed to adhere, deparaffinized, and stained for the light microscope study.

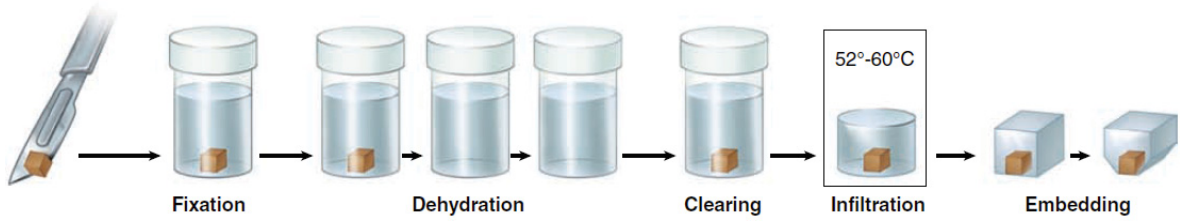


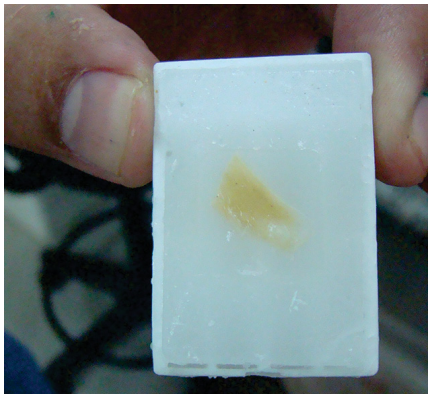
Figure A.3: Schematic presentation of sectioning fixed and paraffin-embedded tissue.
Source: Based on [188]. Modified by the author (2014).

Histological sections became transparent after the microtomy, that is, they present cellular structures that are not clearly differentiated and appear colorless when viewed under a microscope (Figure A.4(c)). Thus, to allow the visualization of cellular structures and other tissue components under the light microscope dyes are applied, as can be seen in Figure A.5(a). Various types of dyes are available to stain the tissue samples. The histopathologists get the benefits of a wide range of colored dyes in order to obtain useful information about the lesions and the tissue compositions. Despite its advantages, the staining process comes up with quite complex variations in staining concentrations and correlations, tissue fixation types, and fixation time periods.

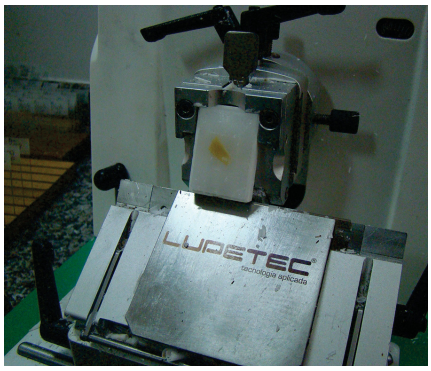
The dyes stain cell structures including the cytoplasm, nucleus, organelles, and extracellular components. This detailed view of the tissue allows the pathologists to make disease diagnosis based on the organization (or disorganization) of the cells and also shows any abnormalities or particular indicators in the actual cells (such as nuclear changes typically seen in cancer) [96]. In histology, the action of most dyes is based on the interaction between its acid or basic elements with positive or negative charged substances of tissue.

Due to its low cost and efficiency, the Hematoxylin-Eosin (HE) staining is used "routinely" with all tissue specimens to reveal the underlying tissue structures and conditions [96].

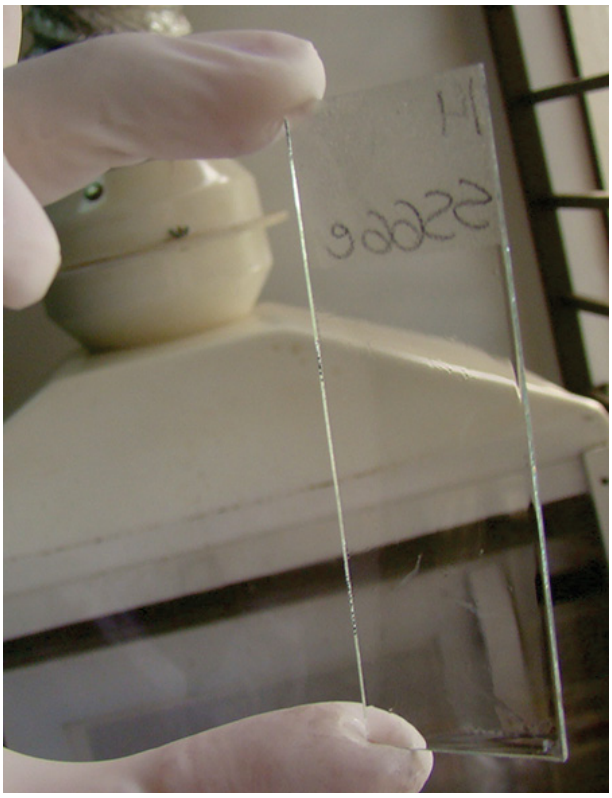
[4]One micrometer ($1\mu m$) is equivalent to $1/1000$ of a millimeter (mm) or 10^{-6} meters.



(a) Hardened block containing tissue and paraffin.



(b) Microtome with trimmed tissue specimen mounted in the paraffin block holder. The block is sliced by the steel blade into extremely thin sections (generally at $1 - 10\mu m$ thickness).

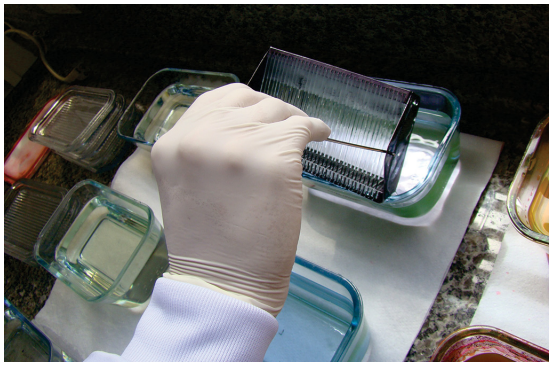


(c) After microtomy, the tissue section is placed on the glass slide and deparaffinized. Note that the section is completely colorless and in order to be studied microscopically, it must be stained.

Figure A.4: Preparation of tissue samples.
Source: Courtesy of the P&D Lab. Modified by the author (2014).

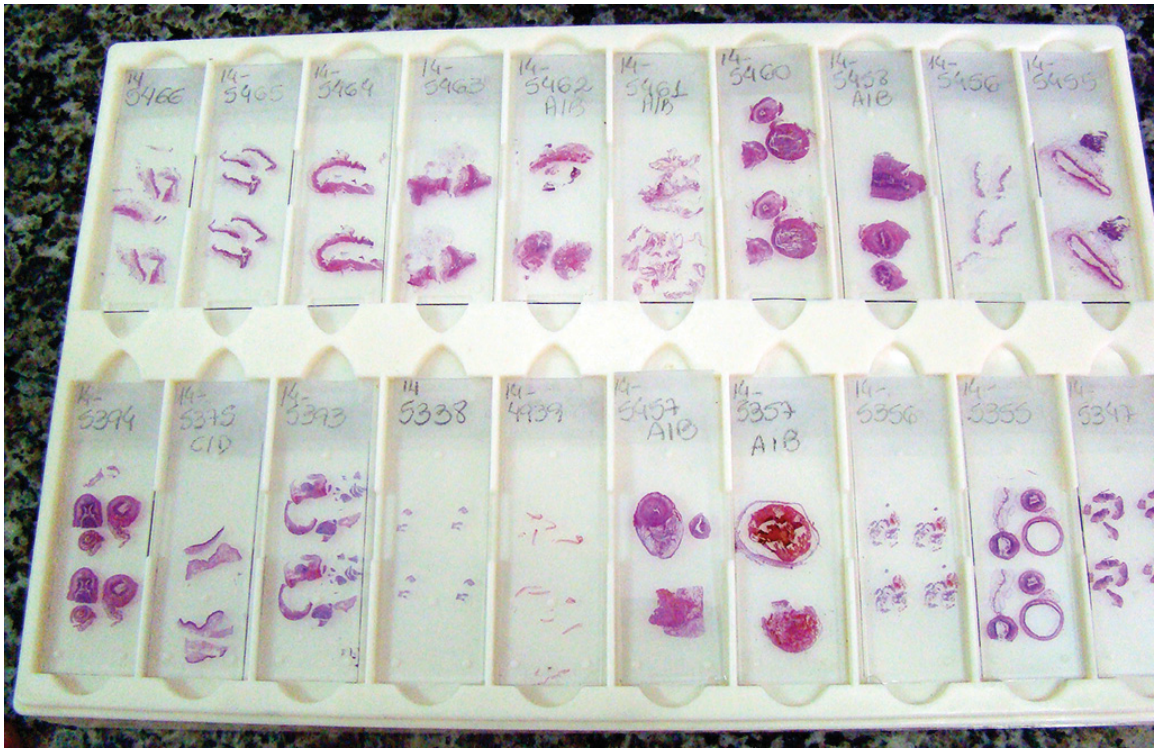
Hematoxylin is a compound extracted from the logwood tree and reacts like a basic dye with a purplish blue color. It stains acidic — or basophilic — structure, including the cell nucleous (which contains DNA and nucleoprotein), and organelles that contain RNA, such as ribosomes and the rough endoplasmic reticulum. On the other hand, eosin is an acidic dye that is typically dark red or pink. It stains basic — or acidophilic — structures, which include the cytoplasm, cell walls, and extracellular fibres. Tissues stained with HE present cytoplasm stained pink-orange and nuclei stained darkly, either blue or purple. Figure A.6 shows an example of breast tumor (carcinoma) with cellular structures highlighted by HE.

Although HE staining method is capable of evidentiate tissue structural features, it can't reveal all of the cellular components. Eventually, other staining techniques, such as different dyes (toluidine blue, orange gelb, methylene blue, methylene green, aniline blue WS) or the association of dyes, are used to highlight distinct cellular materials and in specific pathologic studies.



(a) Slides being immersed in staining solutions.

(b) Tissue sections properly stained with HE.



(c) Different slides ready for the pathologist visual inspection.

Figure A.5: HE staining process.
Source: Courtesy of the P&D Lab. Modified by the author (2014).

A.2.2 Cytopatholy

Cytopathology is focused on study and diagnosis of diseases at the cellular level, analysing the structure, the functionality and the chemistry of cells. Cytologic exams are extremely useful in malignant neoplasia diagnosis and their precursor lesions, as well as detecting the presence of infectious and parasitic agents [40].

Normally, specimen to cytology tests are collected from patients using minimally-invasive biopsy methods, such as smears, scrapes, puncture, centrifugation of liquids and others. A well known example is the screening test for detecting cancer of the neck of the womb (cervix) named *oncotic colpocytology* or popularly Papanicolaou Cytology (PAP). In this procedure, a health professional takes a sample of cells from the patient's cervix with a small soft brush. When

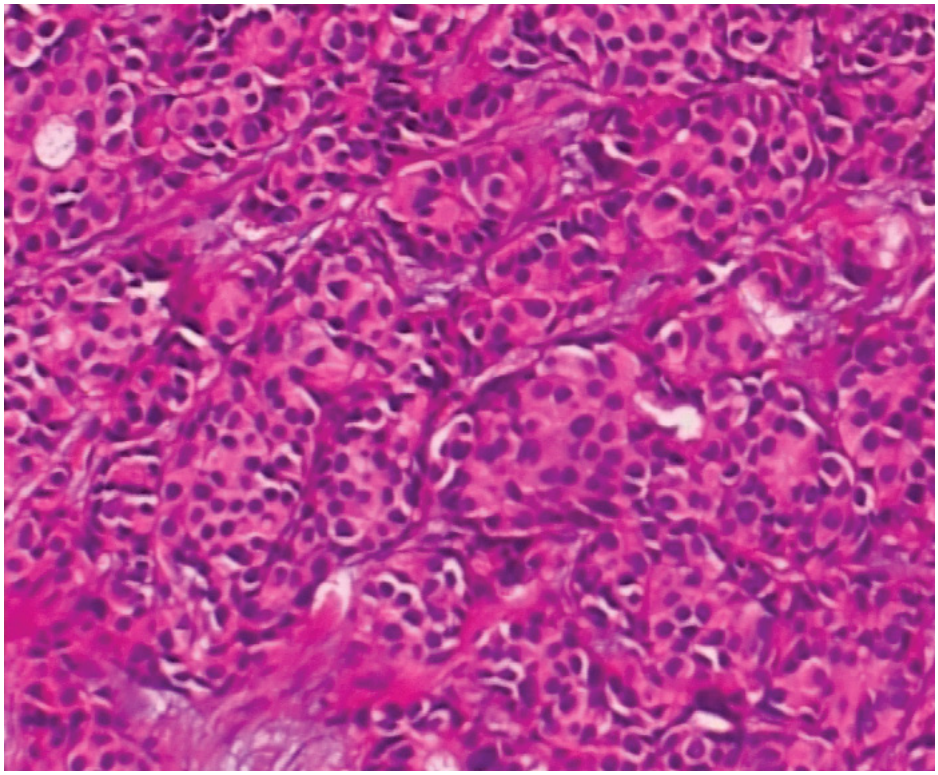
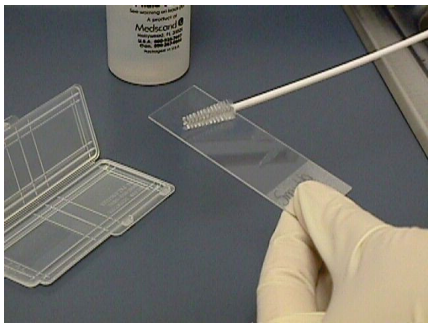


Figure A.6: Detail of HE section of a ductal carcinoma (at 100× magnification). It is possible see a pinkish red color identifying the cytoplasm and the nuclei highlighted in a darker tone of blue.
Source: The author (2014).

applying the Conventional Cytology (CC) method, the cells collected from the outer layer of the cervix are put directly in slide glass (Figure A.7(a)). Differently, in the Liquid-Based Cytology (LBC) technique, cells are put into an appropriate pot with preservative liquid (Figure A.7(b)). In many countries CC was replaced by LBC. However, in Brazil due to a decision of the *Sistema Único de Saúde* (SUS) aiming to reduce costs, CC is still widely used.

In CC, slides are delivered almost ready to pathological laboratory which only performs the fixation and staining process. Cells are heterogeneous and presents a malformed distribution on the slide (Figure A.9, highlight B) and this frequent condition of the sample is responsible for the occurrence of high false-negative rates, inconclusive results and, consequently, requisition of new samples.

Developed in the mid 1990s, LBC began as a collection technique for the cervix cytopathologic test [7]. The initial motivation was to offer a technique feasible for the automatic processing of collected material, in order to produce uniform, thin-layered slides (Figure A.9, highlight A), and minimize the presence of obscuring artifacts, such as overlapping cells. Thus, the shift from CC to LBC has occurred due to improvements in sample quality, reproducibility, sensitivity, and specificity [106], as well as the ability to perform automatic analysis by a computer-based imager [136]. In the first step of the LBC method, collected cells are immersed in preservative fluid stored on a proper bottle, shown in Figure A.8(a). Thereafter, the bottle is processed in a equipment, as shown in Figure A.8(b). This automatic process separates the cells from the liquid, removing the remnant blood, mucus and other impurities [106]. At the end, a



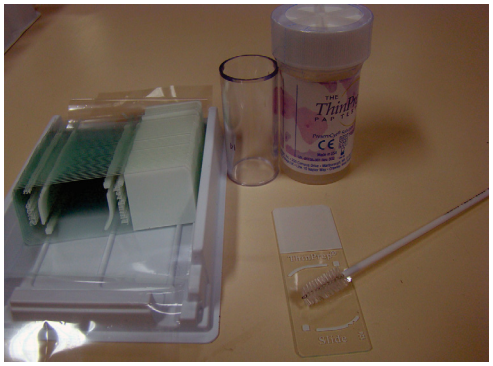
(a) Sample collection by conventional cytology. Only a small portion of the sample taken from the patient is transferred to the slide, and this method of material deposition onto a glass slide generates superimposed cells in irregular layers.

(b) Collection of sample using the LBC method for a cervix cancer screening test. The material is stored in Thinprep® bottle (sold by North American company Cytyc) which contains a preservative liquid. Afterwards, the cells immersed in liquid are processed using an automatic equipment.

Figure A.7: Cytology specimen collection.
Source: Courtesy of the P&D Lab. Modified by the author (2014).

thin cell layer is deposited directly onto the slide, forming a perfect circular area (Figure A.9, highlight A). Similar to CC, the slide is fixed and stained.

Figure A.9 shows the distinct appearance of microscope slides prepared by the conventional method (highlight A) and LBC (highlight B).



(a) Cytyc LBC Thinprep ® test package. Contains: collection bottle, filter, cervical brush, spatula, and proper microscope slide.

(b) Microscope slide being placed for processing in a Thinprep® 2000 processor. The procedure uses a liquid-based filtration process for slide preparation, whereby the sample is dispersed, randomized, filtered, and a representative sample is transferred to the slide. The automatic process guarantees the thin and uniform distribution of cells.

Figure A.8: LBC process.
Source: Courtesy of the P&D Lab. Modified by the author (2014).

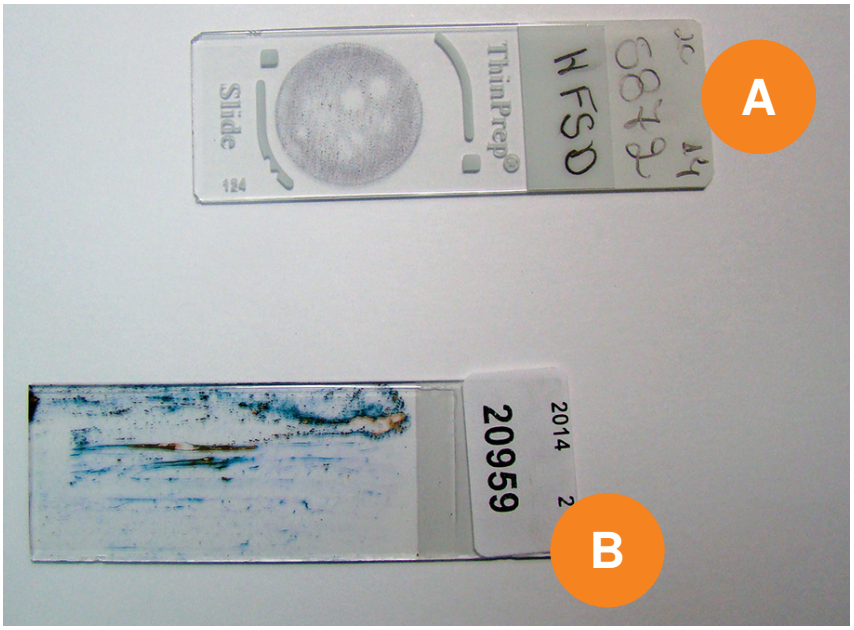


Figure A.9: Microscope slide preparation using LBC method compared to the conventional method. In (A) the microscope slide is prepared using LBC and in (B) using the CC method. In CC, the cell distribution is irregular due to the friction of the spatula/brush against microscope slide surface. On the other hand, the LBC presents a better cellular distribution (specimen collected at the centre circular area), thus facilitating diagnosis.
Source: Courtesy of the P&D Lab. Modified by the author (2014).

A.3 Breast Cancer

Like other types of cancer, breast cancer is the abnormal, fast and unordered proliferation of cells, in this case, from mammary tissue [247]. Individual genetic mutations — caused by several factors — are responsible for this disease that can begin in different breast regions. Thus, the female breast anatomy is briefly presented in Section A.3.1. Section A.3.2 is a brief discussion about diagnosis of breast cancer. Section A.3.3 brings a basic knowledge about breast cancer characterization, including types and stages. Finally, Section A.3.4 is focused on breast cancer in the male population.

A.3.1 Female Breast

The female mammary gland is an organ that has a dynamic behavior: demonstrates morphologic alteration throughout the reproductive life cycle (menstruation, pregnancy, lactation, menopause, etc.) and the age of women [84, 236]. Contrasting to other glands, the breast is functional during the lactation period [123].

In general, in mammal animals, including the human species, the mammary gland is constituted by lobes, lobules, milk ducts, connective tissue, fat, blood vessels and lymphatic vessels. Figure A.10 presents the main anatomical structures of the female breast (in cross section).

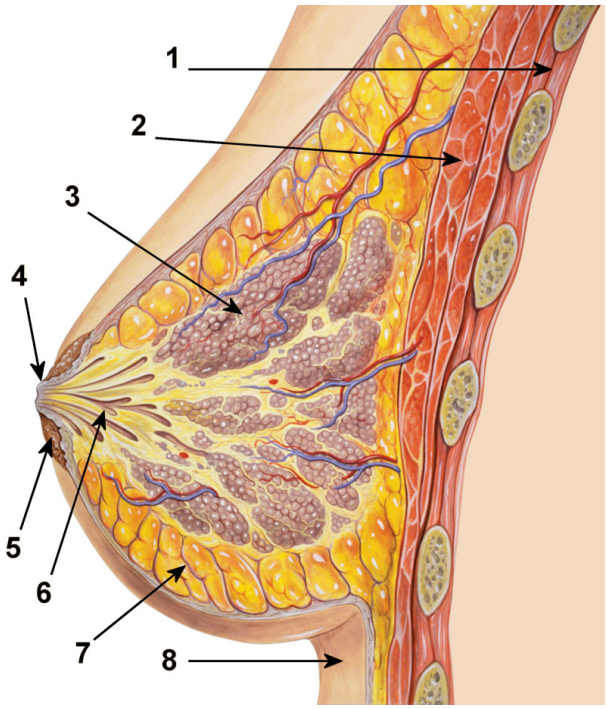


Figure A.10: Normal anatomy of a female mammary gland, in a cross section scheme. (1) Chest wall. (2) Pectoralis muscles. (3) Lobules. (4) Nipple. (5) Areola. (6) Milk duct. (7) Fatty tissue. (8) Skin.
Source: Patrick J. Lynch, medical illustrator; C. Carl Jaffe, MD, cardiologist (2007) [178].

Forming the normal female breast there are between 15 to 20 independent lobes or segments, separated by fibrous tissue, radially distributed from the nipple. The lobe is a well-defined part of an organ (the brain, the breast, the kidney, the liver, the lung, etc.), delimited by sulci, fissures, connective tissues or other anatomic structures. A lobe is visible without a microscope. Each breast lobe is composed of many tiny lobules (Figure A.10-3), at the end of which there are sacs (alveoli) where milk is produced in response to hormonal signals. Lobules are connected to the nipple (Figure A.10-4) through thin tubes (diameter of $1 - 2mm$) which are the milk ducts (Figure A.10-6). Ducts carry milk from the alveoli toward the dark area of the skin in the center of the breast (areola). From the areola (Figure A.10-5), the ducts join together into larger ducts (up to $4mm$) ending at the nipple.

The fibrofatty tissue, fibrous connective and adipose tissue (Figure A.10-7), forming the major components of the breast tissue, filling spaces between lobules and ducts [84]. The generic name *stroma* is given to this tissue type. Connective tissues and ligaments provide support to the breast and give it its shape and volume. Younger women might have denser and less fatty breast tissue, different from older women. The breast itself has no muscle tissue. However, muscles (Figure A.10-2) lie underneath the breasts, separating them from the chest wall (Figure A.10-1).

The basic micro-anatomic structure unit of mammary gland is called Terminal Ductal Lobular Unit (TDLU). The TDLU consists of the Extralobular Terminal Duct (ETD), the Intralobular Terminal Duct (ITD) and the lobule, the functional unit of the breast. In each lobule there is a cluster of round saccules called ductules, which differentiate in the secretory units or acini [5] during lactation. It is in TDLU region, where most benign and malignant tumors originate [84]. Figure A.11 presents a diagrammatic representation of TDLU.

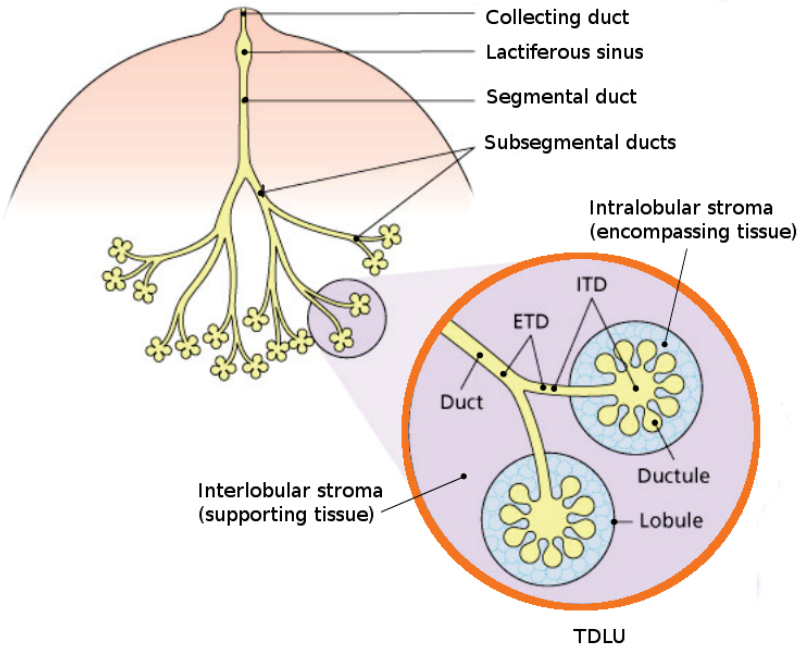
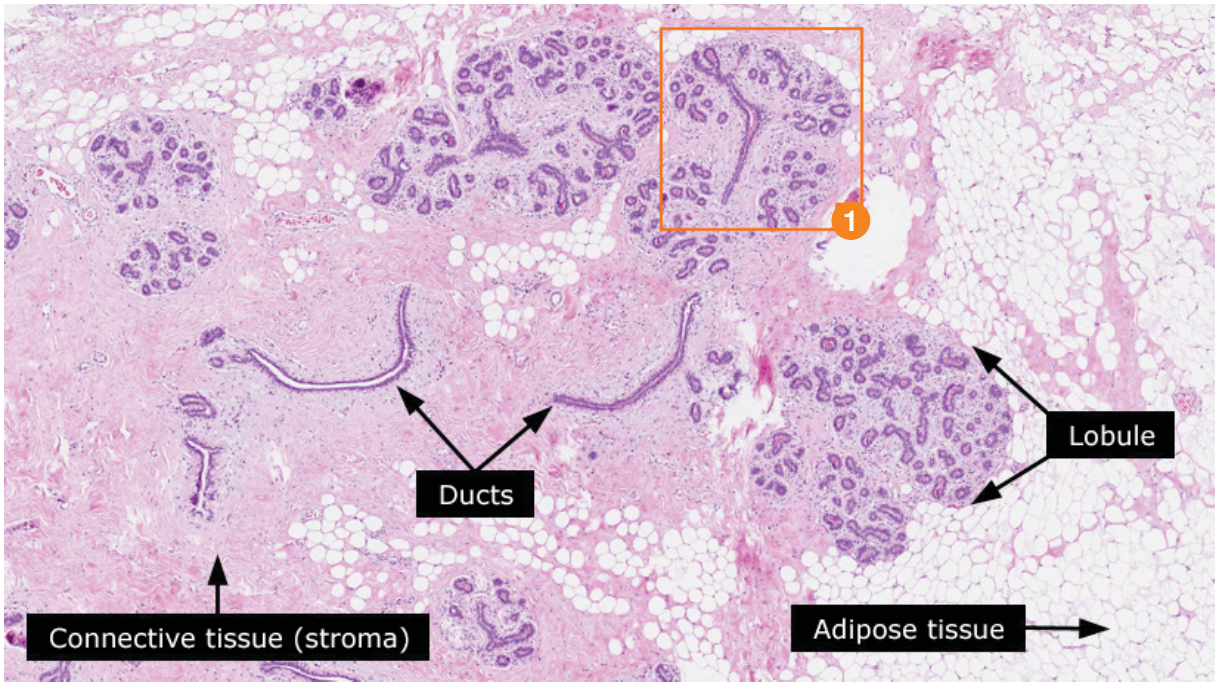
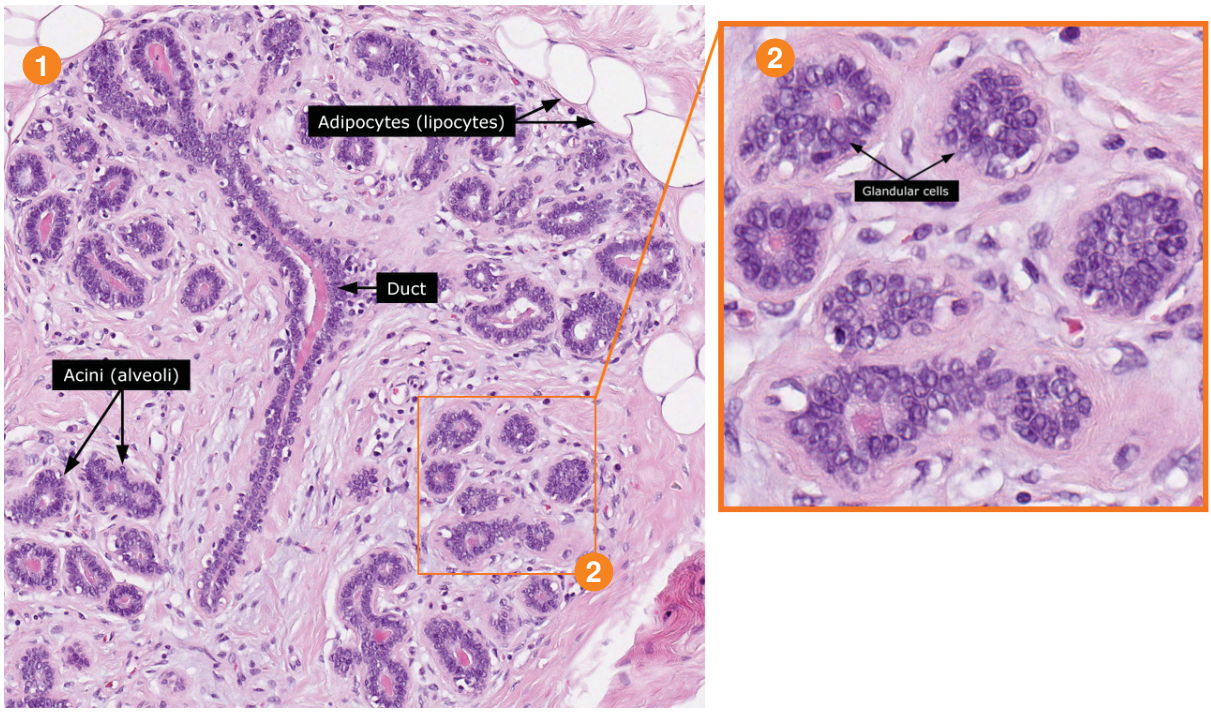


Figure A.11: Diagram of the non-lactating breast showing the arrangement of the ducts and lobules. The TDLU and its main components are highlighted.
Source: Based on [84, 124]. Modified by the author (2014).

[5] Acinus is a denomination for any cluster of cells that resemble a many-lobed "berry", such as raspberry. Acinus is Latin for berry.



(a) Main breast components: the subcutaneous adipose tissue and the functional glandular tissue that comprises both parenchyma and stroma.

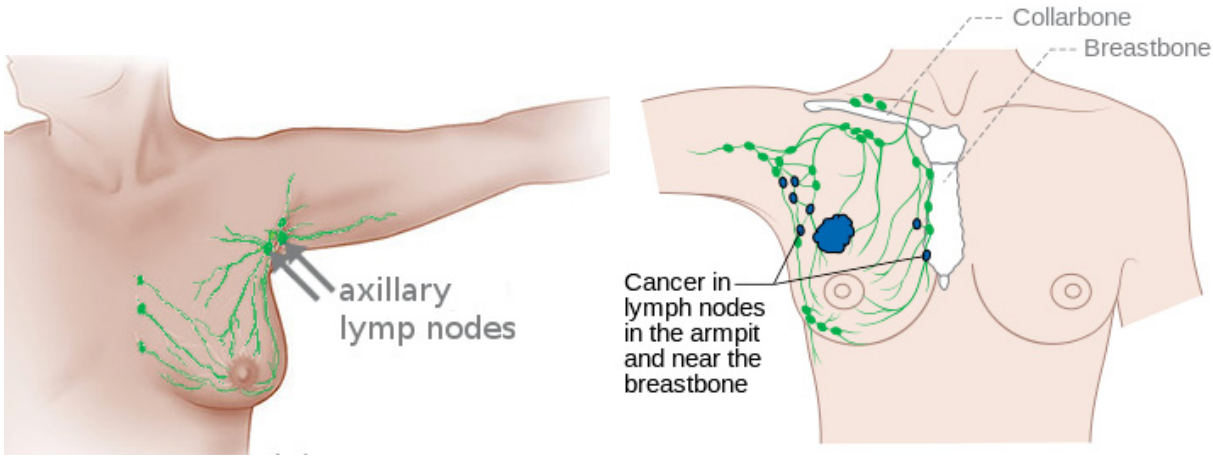


(b) Lobule detail 1 from the Figure above. In magnification 2, cells compounding the glandular tissue.

Figure A.12: Microscopic view of a normal breast. Histologic section HE stained.
Source: Based on [271]. Modified by the author (2014).

In the breast lymphatic vessels are still observed (Figure A.13), which transport a colorless fluid, rich in defense cells, named lymph. Distributed thru the lymphatic system there are small structures called lymph nodes or lymphatic nodes. These bean shaped structures store lymphocytes [161]. There are between five and six hundred lymph nodes in the human body,

many of which are grouped in clusters in different regions. A great part of the breast lymphatic vessels conduct towards the lymph nodes located in the axilla (axillary lymph nodes), highlighted in Figure A.13(a). If cancer cells reach these lymph nodes (Figure A.13(b)), the probability that the disease has spread to other organs rises considerably.



(a) Lymphatic vessels of female breast and axillary lymph nodes. There are about 30–50 lymph nodes in the axilla.

(b) Breast cancer invading lymph nodes.

Figure A.13: Lymphatic vessels of a female breast and axillary lymph nodes.
Source: Based on [50, 201]. Modified by the author(2014).

A.3.2 Breast Cancer Diagnosis

The main exams for the initial diagnosis of the breast cancer are imaging tests (see Section A.3.2.1). However, the final diagnosis, including malignancy, grading and tumor staging, can only be estabilished through the biopsy (see Section A.3.2.3) of the suspected area, which is analyzed by pathologists using anatomopathological exams (Section A.3.2.2).

A.3.2.1 Imaging Exams

The main imaging exams applied in breast cancer detection involve technologies such as diagnostic mammograms (x-ray), Magnetic Resonance Imaging (MRI), breast ultrasound (sonography), and thermography [144]. In Table A.1, the Food and Drug Administration (FDA)[6] approved technologies for breast cancer detection are listed. The main methods will be briefly presented in the following paragraphs.

Mammography is the current standard test for breast cancer screening[7], as well as for diagnostic investigation of findings at the physical (manual) examination [124, 268]. In fact,

[6]USFDA is a federal agency of the United States Department of Health and Human Services. This agency is responsible for protecting the public health by assuring the safety, efficacy and security of human and veterinary drugs, biological products, medical devices, the nation's food supply, cosmetics, and products that emit radiation.

[7]Screening refers to tests and exams used to find a disease like cancer in people asymptomatic and apparently healthy, in an attempt to achieve an earlier diagnosis.

Table A.1: Approved FDA imaging technologies for breast cancer diagnosis.

Technology
Conventional Mammography (film recorded)
Full-Field Digital Mammogram (FFDM)
Sonography (ultrasound)
MRI
Thermography (thermal imaging)
Positron Emission Mammography (PEM)
Electrical impedance imaging (EIT)

Source: Based on [77].

mammographic screening has moved the spectrum of breast pathology away from mostly large tumors, easily visualized and easily palpable, toward ever smaller and frequently noninvasive tumors [267].

Mammography is performed by an equipment which uses low-energy X-rays to produce high-resolution radiological images of the breast internal structure [246]. The mammographic images are a reflection of the breast anatomy and its occasional alteration by the pathological processes. Such images are evaluated by radiologists[8] who search for pathologic findings, e.g., lumps, microcalcifications, density variation in mammary tissue, asymmetry between left and right breasts, nipple deformities and other disease indications. Mammography has become widely and routinely used in breast cancer screening. The screening procedure is recommended every year for asymptomatic women above age 40 or younger when in the high risk category (previous biopsy showing cancer or high-risk lesions, strong positive family history, previous exposure to high doses of chest radiation, underlying genetic abnormalities, etc.).

The mammography image must present a high-contrast resolution, low noise and follow strict protocols, allowing for proper differentiation between normal tissue and potential abnormalities. In Figure A.14 a mammogram highlighting a non-palpable tumor is shown. For a long time the conventional mammography was an analogical using screen-film record, but it has been replaced by the digital detectors, called digital mammography [21] or FFDM. Overall, the digital mammography has significant benefits over conventional mammography, including decoupling of image acquisition and storage [58]. The appearance of the processed film in the analog technique is directly related to and dependent on the exposure and the processing values used. However, the digital image can be adjusted after exposure so as to optimize viewing [268].

Working as support tools for the human expert, Computer-Aided Detection (CAD) systems are designed to automatically highlight suspicious areas in the image and to produce preliminary interpretations that help the radiologist in the final diagnosis [124]. In the early 1990s, the digital mammography became widely used to automatically detect breast lesion areas which were indicative of cancer. Therefore, asymmetry tests between the left and right breasts, and executed pattern recognition algorithms are performed to evaluate breast tissue texture (microcalcifications, dense structures, lesions) [187]. The CAD system of digital mammography

[8]Radiologists are graduated medical professionals who specialize in diagnosing and treating diseases and injuries using medical imaging techniques, such as x-rays, computed tomography (CT), magnetic resonance imaging (MRI), nuclear medicine, positron emission tomography (PET) and ultrasound [285].

operates as a second reading of the mammogram and can automatically identify possible tumoral areas requiring special attention from the radiologist [124].

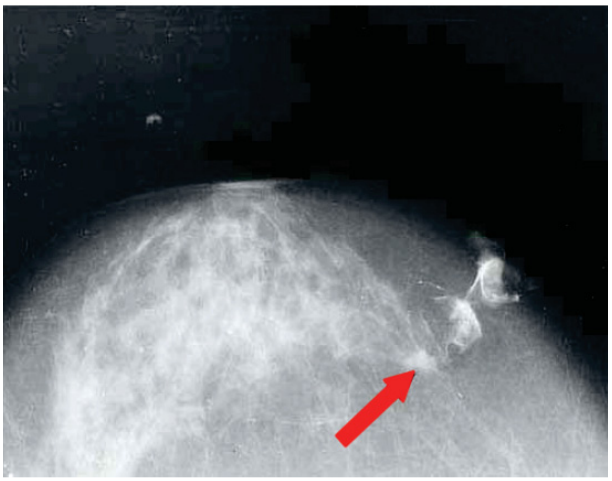


Figure A.14: Mammogram showing non-palpable tumor (arrow) near right border of mammary parenchyma.
Source: Based on [126]. Modified by the author (2014).

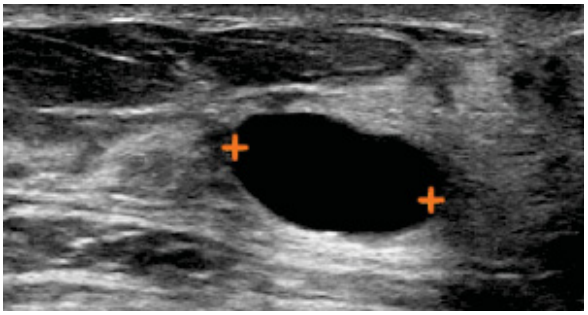
In order to standardize and unify the mammography report in the early 1990s, several US medical committees, headed by the American College of Radiology (ACR) established a classification system named Breast Imaging-Reporting Data System (BI-RADS)® [230]. This system, besides standardizing the report, it standardize the terminology and establishes categories, and it also suggests treatment conducts [200]. A summary of the BI-RADS® system is presented in Table A.2. In 2003, a new BI-RADS® edition also included breast ultrasound and breast MRI.

Table A.2: Classification BI-RADS®.

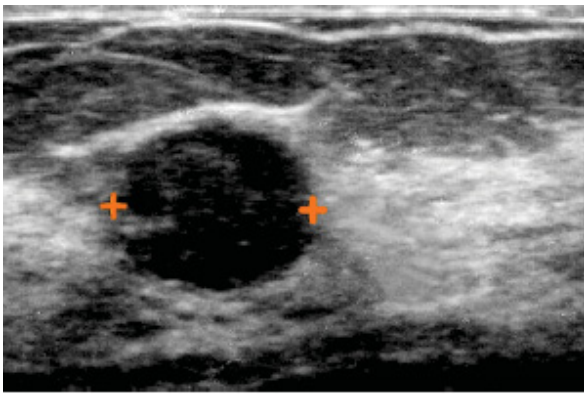
Category	Evaluation	Recommendation
0	Inconclusive	Additional imaging evaluation and/or comparison to prior mammograms is needed.
1	Normal findings (negative)	Back to screening (usual interval for age and risk).
2	Benign finding	Back to screening (usual interval for age and risk).
3	Probably benign finding	Short-term follow-up (usually at 6 months). Additional views and ultrasound examination are required to confirm this recommendation.
4	4A – Low suspicion 4B – Intermediate suspicion 4C – High suspicion	Histopathologic assesment
5	Highly suggestive of malignancy	Histopathologic assessment and probably appropriate treatment for malignancy required
6	Known biopsy-proven malignancy	Appropriate action has to be taken. Mammograms can be used to verify how the cancer is responding to treatment.

Source: Based on [124, 200].

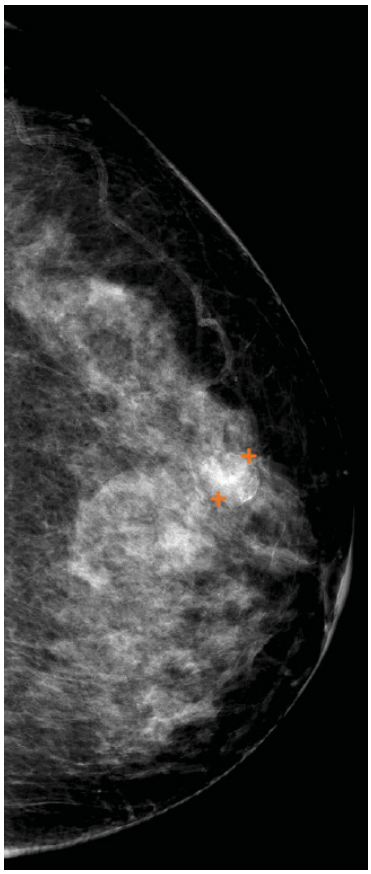
After mammography, sonography stands out as the second most important breast imaging exam [9]. Sonography — or Ultrasonography — uses emission of high-frequency sound waves (often in the range 10^3 to 10^6 Hz) in direct contact with the breast skin, where gel is applied in order to facilitate the propagation of the ultrasonic beams. Due to the acoustic impedance property, materials of different densities reflect sound waves in a distinct way, allowing that such reflection (echo) be measured for the diagnostic image formation. It is indicated for lesions that cannot be seen on mammography, especially when the breast is very dense [21] or to determine the presence of liquid inside in the nodules, i.e., distinguish between cystic and solid lesion. Moreover, since it produces real-time images, ultrasound is useful to guide the exact position to accomplish puncture, collect material for biopsies or establish pre-operative surgical markings [124], correctly including lesions unnoticeable to the touch. Figure A.15 presents examples of sonographic findings and the corresponding interpretation.



(a) Simple cyst can be definitively diagnosed sonographically: oval shape, thin wall, absence of internal echoes, and far wall enhancement of the soundbeam.



(b) This ultrasound image requires further assessment: well-circumscribed round hypoechoic lesion.



(c) Mammography correlating sonographic finding in A.15(b): demonstrates that the lesion corresponds to a partly calcified cyst with thickened contents.

Figure A.15: Clinically suspicious sonographic findings.
Source: Based on [124]. Modified by the author (2014).

[9] While mammography cannot match ultrasound in terms of early diagnostic precision, it is still an indispensable tool because microcalcifications are visible only on mammograms [124].

MRI is an imaging test, complementary to mammography and sonography, which apply electromagnetic waves and radio-frequency pulses (instead X-rays) to produce the images. Routinely for cancer detection and presurgical evaluation, this test is done after an intravenous application in the patient of a gadolinium-based contrast agent [102, 124]. The contrast agent allows the enhancement of the visualization of certain structures, and highlights neovascularization in the case of mammary tumors, as can be seen in Figure A.16. MRI test also allows detailed evaluation of nodes, because it reaches deeper regions of the breast tissue. It is extensively used to monitor the integrity of breast implants and recommended for screening in high-risk patients, such as those with a confirmed or suspected family history, with known genetic predisposition to breast cancer or who have already been affected by the disease.

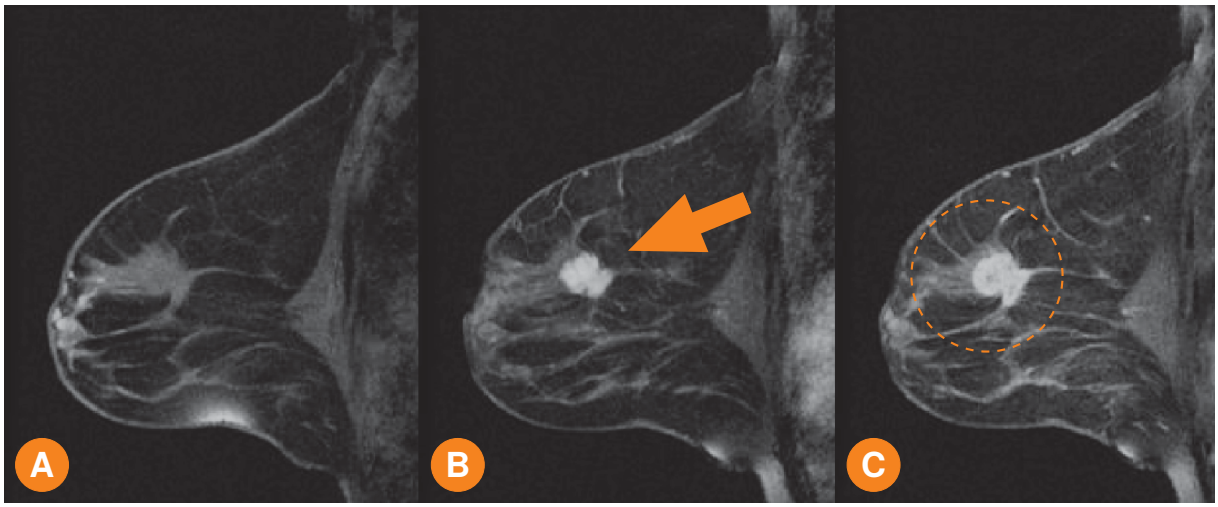
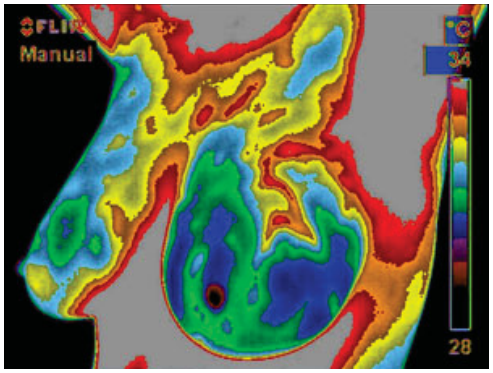


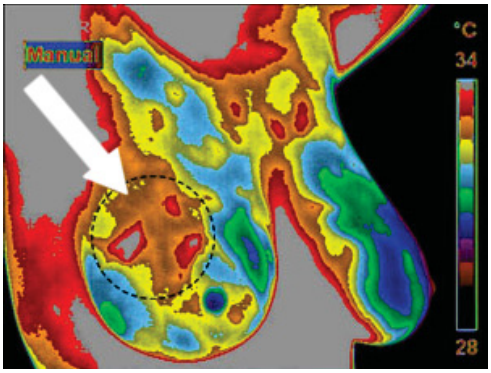
Figure A.16: Sequence of breast MRI images. In (A) breast image pre-contrast application. In (B, C) breast image post applied contrast. Note in (B) the highlighting of the tumor and in (C) the breast tissue neovascularization inducted by the tumor is also highlighted.
Source: Based on [102]. Modified by the author (2014).

The medical thermography, in particular Infrared Thermograph (IRT), is a rapid, passive and non-invasive method which has been successfully used to diagnose disorders such as breast cancer, diabetes, neuropathy and peripheral vascular issues [163]. Since 1982 thermography is approved by the FDA as a complementary exam to mammography for breast cancer diagnosis [77]. However, the FDA itself emphatically alerts that thermography is not a substitute for mammography and should not be used by itself for breast cancer screening or diagnosis[10]. Inflamed tissues, precancerous or surrounding a tumor, show higher vascular and chemical activity when compared to normal tissue [54]. This extra activity generates a greater emission of infrared radiation that can be detected [14]. Breast thermography or Digital Infrared Imaging (DII) is a procedure to map the thermal emission of the skin over the area of the breasts, in order to show thermal asymmetries that are indicative of the presence of cysts, infections, breast cancer or other diseases [245]. While other tests such as mammography and ultrasound detect anatomical changes already established in breast, thermography has the advantage of being a functional examination, which studies metabolic and vascular abnormalities of the breast (Figure A.17).

[10] http://www.fda.gov/MedicalDevices/Safety/AlertsandNotices/ucm257259.htm



(a) Left breast presenting thermal pattern compatible with normal vascularization.



(b) Right breast showing apparent temperature rise and vascularization (angiogenesis). The lump in the upper outer quadrant of the breast is highlighted. The patient's mamography was negative. Biopsy after DII confirmed the malignancy of the lesion.

Figure A.17: Typical breast thermography image.
Source: Based on [8]. Modified by the author (2014).

A.3.2.2 Anatomopathological Exams

Regarding breast cancer, if the physical examination (touch) detects palpable lumps or imaging exams find suspicious tissue areas, anatomopathological exams are required. Pathological exam analyses cellular and tissue microscopic alterations present in samples collected from biopsies or surgeries. The pathologist responsible for pathological examination can make the correlation with clinical and imaging tests. Generally the pathological diagnosis is considered definitive, but it may be inconclusive, due to limiting factors such as insufficient material collected or even if the collected sample is unrepresentative of the suspicious lesion. Section A.3.2.3 presents the main biopsy methods for breast cancer investigation.

A.3.2.3 Breast Biopsies

Once screening tests, such as mammography or breast ultrasound, have found suspicious changes, it is recommended to biopsy the region. Breast biopsy is the removal of a small amount of breast tissue for pathologic evaluation to determine whether it is cancerous or non-cancerous (benign). Particularly in breast lesion investigations, the main biopsy techniques can be grouped into needle (or percutaneous puncture) techniques and surgical techniques (Table A.3). In needle biopsy group there are techniques such as Fine Needle Aspiration (FNA), Core Needle Biopsy (CNB) and mammotomy. Surgical biopsy techniques can be done by incision or excision. The more appropriate biopsy method depends on several factors, such as, how suspicious the lesion is; the size, shape, and location of the lesion; the number of abnormalities present; the patient's medical history, etc. Most often, a needle biopsy is done first and then, if needed, a surgical biopsy is done. Figure A.18 shows a schematic representation of the four main breast biopsy techniques. FNA, CNB, and VABB are percutaneous methods, i.e., a needle is inserted through the skin. Surgical biopsy yields the largest breast tissue sample of all the breast biopsy methods, and the accuracy of a diagnosis is better, making it the "gold standard" of breast biopsy methods.

However, the surgical biopsy method is much more invasive than percutaneous methods, it requires stitches and can leave a scar.

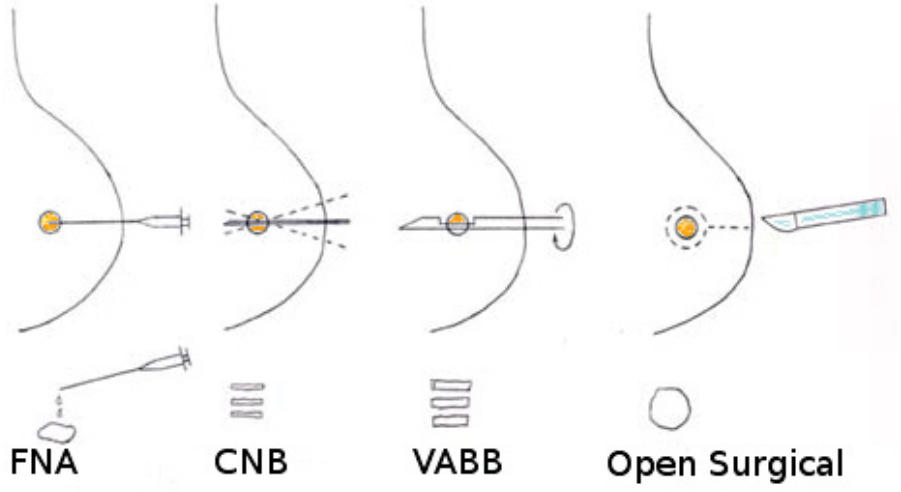


Figure A.18: Schematic representation of different breast biopsy types.
Source: The author (2014).

Table A.3: Types of breast biopsies.

Intervention	Procedure
Percutaneous Puncture (Needle)	FNA CNB VABB, Mammotome Biopsy (Mammotomy)
Surgical	Incisional Excisional

Source: Based on [42].

FNA was introduced in 1930 and became popular in the 70s [274]. In this technique a very thin needle ($20 - 21G$ [11]) is used. In general, the needle used during FNA is smaller than a needle that is normally used to collect blood samples. A 10 or $20ml$ syringe attached to the needle allows to aspirate fluids and clusters of cells from the puncture site [80, 206]. It is a fast and low cost procedure usually requiring no anesthesia. Several needle insertions are usually needed to guarantee that an adequate tissue sample is taken. The collected material is deposited on slides (samples are smeared on a microscope slide) for a cytological study. But, effectiveness of FNA is totally operator-dependent and the use of this procedure has gradually decreased because of the controversial rates of specimen inadequacy and suboptimal accuracy of diagnoses in inexperienced hands [80].

The CNB, on the other hand, uses a hollow core needle (16−, 14−, or 11−gauge needle), coupled to a special pistol, to take larger tissue cylinder samples of a breast mass, as well as nearby healthy breast tissue [206]. This CNB needle also has a special cutting edge. The procedure

[11]Considering needles, the measurement gauge (G) refers to the caliber, determining their diameter. The lower the G value, greater the diameter. Thus, $20G$ and $21G$ are equivalent to the external diameter of $0.9081mm$ and $0.8192mm$, respectively.

requires local anesthetic (lidocaine), because to facilitate the insertion of the needle a small incision in the skin is made with a bistoury. Are removed several fragments of a few millimeters which are packaged in bottles containing formaldehyde. If the lump is non-palpable (cannot be felt), CNB is performed under image-guidance using either Stereotactic Breast Biopsy (SBB)[12] or ultrasound. In this procedure, a device using radiographies obtained in different angles, guides the needle, positioning it in the correct area of concern.

Mammotomy, also called Vacuum Assisted Breast Biopsy (VABB) or Minimally Invasive Breast Biopsy (MIBB), is a minimally invasive image-guided technique established in 1995 and mainly used to assess non-palpable breast lesions detected on mammography [129, 204]. It uses a probe ($8 - 11G$) with a rotary cutting head and a cannula connected to a vacuum system to aspirate the removed breast tissue [206]. Unlike other puncture techniques, during the VABB procedure the biopsy probe is inserted only once into the breast through a small skin nick, and allows to extract larger fragments of lesioned breast areas [42, 177]. Although mammotomy is intended for diagnosis, in benign cases, such as fibroadenomas, the procedure can completely remove the tumor [204]. However, if malignancy of the lesion is confirmed, surgical procedures will be needed in order to remove larger amounts of breast tissue [177].

Figure A.19 presents a visual comparison between different needle sizes used in procedures VABB, CBN and FNA, respectively.

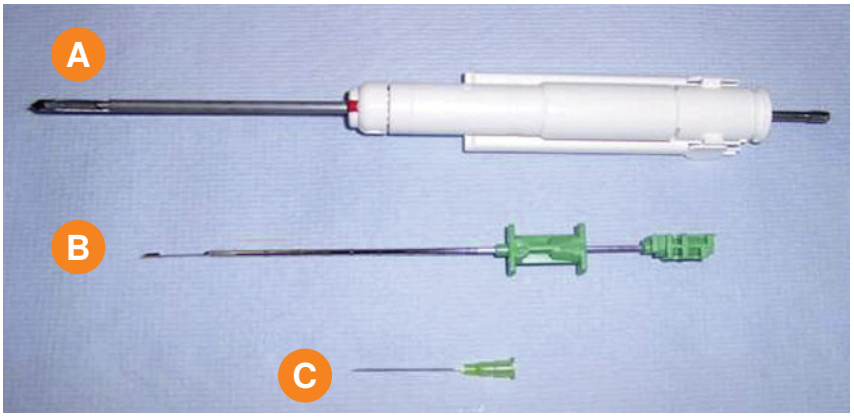


Figure A.19: Visual comparison of needles: (A) VABB equipment of $7G$ size, (B) CNB of $14G$, and (C) FNA of $21G$.
Source: Based on [206]. Modified by the author (2014).

Surgery biopsy is a procedure performed in the operating room by a surgeon who totally removes the lesion (including margins containing healthy tissue) or a representative part of it. The patient receives local anesthetic or a combination of intravenous sedation with local anesthetic. During an excisional surgical biopsy, the surgeon will attempt to completely remove the lesioned area, often along with a surrounding margin of normal breast tissue. This removal of the breast tumor ("lump"), and some of the normal tissue that surrounds it, is also called lumpectomy. Incisional breast biopsy is usually recommended for large lesions and the surgeon removes only part of the breast lesion. Thus, in suspicious nodules the incision is made directly over the nodule, avoiding excessive handling so as not to increase dissemination risk. Currently, considering all

[12]Stereotaxy is a surgical method for locating points within the patients body using an external, three-dimensional frame of reference usually based on the Cartesian coordinate system.

the breast biopsy methods, open surgical method presents the most reliable diagnosis with higher accuracy rates. However, compared to the percutaneous biopsy procedures, surgical biopsy requires a longer period of recovery, can leave significant scars and eventually changes the size and shape of the breast.

Additionally, the biopsy sample is tested for the presence of Estrogen Receptor (ER) and Progesterone Receptor (PgR) using a method called IHC. Female breast cancers containing these receptors are more likely to respond positively to hormonal therapy. In women, estrogen and estrogen hormones can stimulate the growth of some breast cancer cells. Thus, hormone therapy for breast cancer tries to lower the levels of estrogen and progesterone in the body and block their effects. When breast cancer cells have ER, the cancer is called ER-positive (ER+) breast cancer and PgR-positive (PgR+) when it has PgR. Tumor cells may contain none, one, or both of these hormone receptors [9]. About $75 - 80\%$ of breast cancers are ER- and/or PgR-positive and low-grade cancers are even more likely to be ER- and/or PgR-positive.

In some cases, no abnormalities were found in the breast, but the existence of increased lymph nodes were found in the armpit region. Likewise, biopsies can be made, such as FNA and CNB, or even surgical excision of the node. The material is sent for pathological evaluation.

A.3.3 Characterization of Breast Cancer

As seen in Section A.1, not all tumors are cancerous. Non cancerous tumors are called benign. Despite the fact that benign tumors can cause serious health problems, they cannot invade surrounding tissues and spread to other parts of the body (metastasize). For example, fibrocystic change is a benign lesion (non-cancerous condition) extremely common in which women develop cysts (accumulated packets of fluid), fibrosis (scarring of connective tissue), lumpiness, areas of thickening, tenderness, or breast pain [160]. Thus, these tumors are seldom life threatening.

On the other hand, cancer is a synonym for malignant tumors that penetrate and destroy healthy body tissues. Cells from a malignant tumor may also spread beyond the original site to other parts of the body (metastases). Therefore, untreated cancers can cause severe illnesses and death. A malignant tumor that has developed from cells in the breast is referred to as breast cancer.

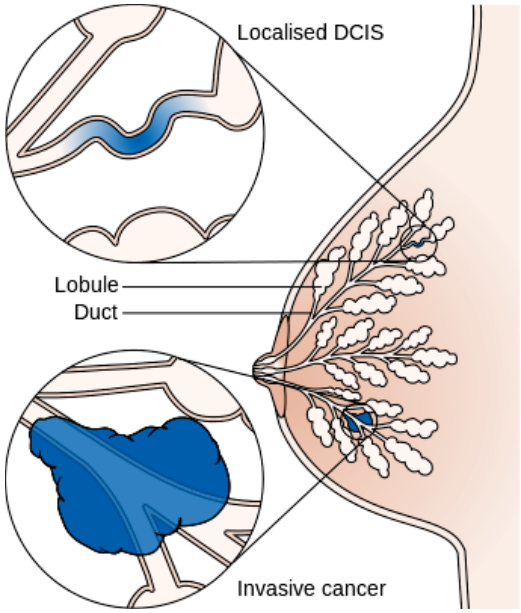
Breast cancer is not a single disease, but is rather composed of several distinct subtypes associated with different origins, evolution and possible therapeutic interventions [226]. By using histopathological analysis, breast cancer can be classified into different tumor types and characterized in their specificities. Such classification is essential because it determines different prognoses, surgery planning and distinct therapies for each type.

As seen in Section A.3.2.1, image screening procedures are indicated for the detection of breast cancer at early stages, even with no apparent symptoms. Among these tests, the most important is the mammogram, being associated in some cases with breast ultrasound or breast MRI. If a breast abnormality is detected with the physical exam, mammography or other additional breast imaging tests, the patient must be referred for a breast biopsy (see Section A.3.2.3). While physical breast exam, mammography and other breast imaging methods can help detect a breast abnormality, biopsy followed by pathological (microscopic) analysis is the

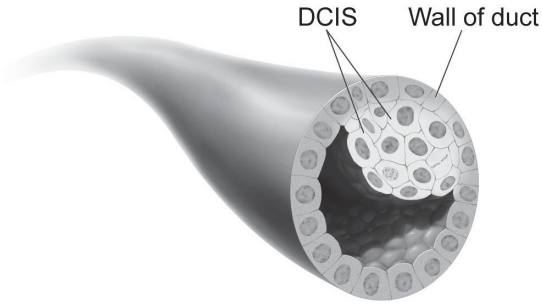
only reliable way to determine whether a lesion is benign or malignant (cancer), its type, grading and progression [113]. Further subclassification is based on morphology, immunohistochemistry and molecular profiling [238].

A.3.3.1 Breast Cancer *In Situ*, Invasive and Metastatic

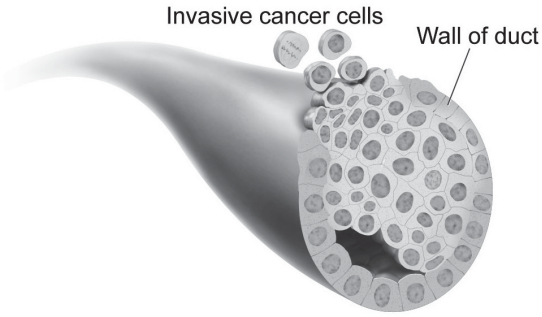
Whenever the tumor is confined to a certain structure it is called *in situ*, and typically does not have the property of developing metastasis [161]. When breast cancer spreads from its origin to another part of the body, the new tumor is formed by the same kind of abnormal cells from the original tumor, and the new tumor continues to receive the name of the original tumor [198]. If breast cancer becomes *metastatic* and spreads to the lung, for example, cancer cells in the lung are actually breast cancer cells. Thus, it is still breast cancer and not lung cancer. They would be treated the same way, too. Figure A.20(b) shows a duct containing breast carcinoma *in situ*, i.e., cancerous cells are restricted to the duct walls. Ductal Carcinoma *In Situ* (DCIS) is the most common form of non-invasive breast cancer. About 1 in 5 new breast cancer cases are DCIS [9]. Another form, Lobular Carcinoma *In Situ* (LCIS) or lobular neoplasia, is considered a non-cancerous condition, caused by abnormal cells inside lobules. It is often present in both breasts and the cells are all contained within the inner lining of the breast lobules. It tends to be diagnosed as a result of a biopsy performed on the breast for some other reason, because LCIS does not cause symptoms and usually does not show up on a mammogram. However, LCIS increases the risk of getting invasive breast cancer in either breast in the future. When cancer starts to reach neighboring regions it is called invasive or infiltrating [238]. Even small invasive carcinomas can potentially generate metastasis [161], but a cancer can be invasive without being metastatic. In Figure A.20(c) a carcinoma extrapolating the mammary duct to invading adjacent tissues is shown.



(a) Diagram showing Ductal Carcinoma *In Situ* (DCIS) and Invasive Ductal Carcinoma (IDS).



(b) A breast duct presenting DCIS. Abnormal cells have not spread outside of duct, thus other mammary tissues were not compromised.



(c) A breast duct presenting IDC.

Figure A.20: Ductal carcinoma *in situ* vs infiltrating ductal carcinoma.
Source: Based on [33, 51]. Modified by the author (2014).

A.3.3.2 Breast Cancer Types

Medical experts group breast cancers into two broad categories: (1) No Special Type (NST) and (2) Special Type. The distinction between the two categories follows the characterization of the cells making up the tumor. In NST — or Not Otherwise Specified (NOS) – tumor cells are undifferentiated under microscope. When tumor cells have features that classifies them as a particular type of cancer, it is classified as special type. Ductal carcinoma, the most frequent cancer which starts in the lining cells of ducts, is a NST tumor [274]. Figure A.21 shows the abnormal cells in duct. On the other hand, malignant tumors that are classed as special type include lobular carcinoma and some rare types of breast cancer (medullary, mucinous, papillary and others). Rare types of breast cancer represent less than 2% of all breast cancer cases. As additional complexity factor in classifying tumor types, a single breast tumor can be a combination of various breast cancer types or even a mixture of invasive and *in situ* cancer [9]. Besides, in some more rare types of breast cancer, the cancer cells may not form a defined tumor [9].

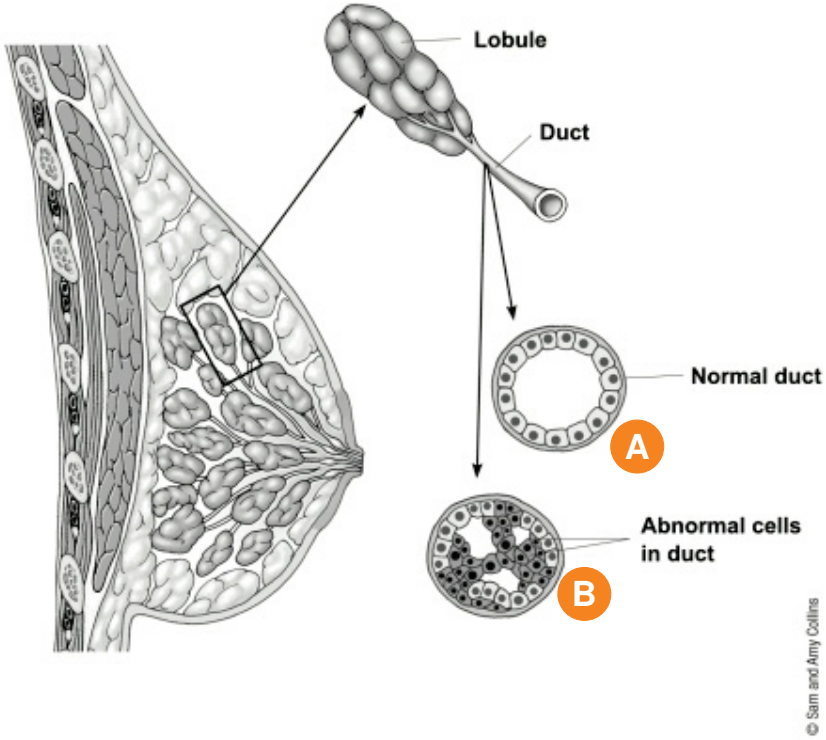


Figure A.21: Female breast highlighting a lobule and a duct. Section of duct presents two possibilities: (A) normal appearance and (B) abnormal cells growing inside a duct (DCIS).
Source: Based on [9]. Modified by the author (2014).

Table A.4 presents a summary of the usual denominations of the breast cancer types and their synonyms. The most common breast cancers are briefly presented in the following paragraphs.

Table A.4: Breast cancer types. The type often determines how the cancer behaves and what treatments are the most effective.

Denomination	Synonymy
Adenoid Cystic Carcinoma	Cylindroma, Adenocystic Carcinoma
Apocrine Tumor	–
Cribriform Carcinoma	–
Ductal Carcinoma *In Situ* (DCIS)	Intraductal Carcinomal
Inflammatory Breast Cancer	–
Invasive Ductal Carcinoma (IDC)	Infiltrating Ductal Carcinoma, Ductal Adenocarcinoma
Invasive Lobular Carcinoma (ILC)	–
Lobular Carcinoma *In Situ* (LCIS)	–
Medullary Carcinoma	–
Metaplastic Carcinoma	–
Mucinous Carcinoma	Colloid Carcinoma
Paget's Disease of the Nipple	–
Papillary Carcinoma	–
Squamous Cells Carcinoma	–
Tubular Carcinoma	–

Source: Based on [9, 52, 135, 161].

Depending on the morphology and the presumed site of origin, the vast majority of breast cancer occurrences in women is broadly classified as Invasive Ductal Carcinoma (IDC) and Invasive Lobular Carcinoma (ILC). Although it can affect women at any age, breast cancer is rare before the age of 25, except for family cases, and 77 % of the cases occur in women over 50 [161]. Thus, the IDC and ILC are more common in older women [43, 52], although the ILC can be found in more advanced ages. The ILC does not always form a well-defined node but a thickening of the breast tissue, it can not easily be detected on mammograms and, therefore, it is typically larger than other types when it is diagnosed [52].

IDC is the most common invasive breast cancer, representing more than 80 percent of all cases [43]. In a macroscopic view, IDC is characterized by a hard lump with irregular borders. On a mammogram, IDC usually looks like a mass with spikes radiating from the edges; sometimes it appears as a smooth-edged lump or as calcifications in the tumoral area.

Lobular Carcinoma is considered the second most commonly diagnosed form of breast cancer (about 10–15%) [43]. This type of cancer is more difficult to detect on imaging procedures, due to the way it grows with spreading branches. Also, ILC can show the formation of numerous small hard nodules mimicking sclerosing adenosis [191].

Tubular Carcinoma is a very-well differentiated infiltrating carcinoma with an excellent prognosis and its name comes from the tubular-angulated appearance of the cells [191]. This type of tumor is usually small and low graded. Normally, tubular carcinoma is accompanied by areas of DCIS.

Papillary breast cancer is usually a small tumor and it is a very rare type of invasive ductal breast cancer that accounts for less than 1% of all breast cancers. The name papillary comes from finger-like projections, or *papules*, found in lesion. In this carcinoma, cancerous cells are disposed in a pattern remembering bracken's shape [214].

Mucinous (also named Colloid) Carcinoma is one of the three subtypes of mucin-producing carcinomas of the breast [191]. A unusually large amount of mucous (mucin) is observed in cells that are part of this type of tumor, giving its name. It is a well-circumscribed tumor characterized by a low proliferative rate [244].

A.3.3.3 Breast Cancer Grading and Staging

Currently, pathologists determine the tumor grading by assessing the spatial organization of the tissue (e.g., distribution of cancer cells, nuclei morphological properties, interaction with the stroma, etc.). These parameters are evaluated in small sample regions of the microscopic slide in order to give a score considering some "scoring system" such as Nottingham Histologic Score System[13]. Complementing the final decision, for prognosis and clinical intervention, respective hormone receptor status by IHC is also analyzed in the IHC-stained sections.

Breast cancer survival is strongly influenced by tumor stage. Staging is the process of finding out how widespread cancer is when it is diagnosed, i.e., the presence of local and/or distant spread. Localized, the disease (Stage I) comprises approximately 60% of the cases, while

[13]Nottingham (or Elston-Ellis) modification of the Scarff-Bloom-Richardson grading system.

in about 5% cancer has spread to distant organs such as liver and bones (Stage IV). Approximately 35% are Stage II or III, indicating tumor spread to regional lymph nodes.

A.3.4 Male Breast Cancer

In general, testosterone hormone causes involution of male mammary gland. A normal male breast is primarily composed of ductal structures within collagenized stroma, with no or rare lobular elements compared to the female breast [40, 161]. Although rare, men can also be affected with breast cancer, accounting for less than 1% of all cases of the disease [44, 199, 244]. Risk factors, pathology, and prognosis of male breast cancer treatments are quite similar to those observed in the female population [244], however breast cancer in men is often diagnosed at later stages [199, 244]. All histological subtypes (*see* Section A.3.3.2) may occur, but the invasive papillary carcinoma constitutes the most frequent rare-type encountered in males and ILC is very unusual [161, 199, 214].

The treatments for breast cancer in men are similar to those in women [238], with the exception of surgical options. Given the breast volume and tumor location, the standard procedure for men is to have a mastectomy surgery, rather than a lumpectomy. Male breast cancers are usually hormone receptor positive tumors and hormonal therapy is also a common part of the systemic treatment. Genetic testing should be considered for men who develop breast cancer. The survival rates and prognoses for men are not as good as for women. Men have a 25 percent higher mortality rate than women. As mentioned previously, this is believed to be due in part to men getting diagnosed at a later stage of the disease.

Appendix B

BreaKHis — Dataset Description

This text is an expanded version of the BreaKHis description presented in [256]. The BreaKHis dataset contains digitized microscopic images of biopsies from benign and malignant breast tumors. Images were collected through a clinical study from January 2014 to December 2014. All patients referred to the P&D Laboratory, Brazil, during this period of time, with a clinical indication of BC were invited to participate in the study. The institutional review board approved the study and all patients gave written informed consents. All the data were anonymized.

Samples were generated from breast tissue biopsy slides, stained with HE (Figure B.1(b)). Most of the samples were collected by Surgical (open) Biopsy (SOB) and few slides were collected using CNB method. Tumoral areas were identified and labelled by anatomopathologists of the P&D Lab. The preparation procedure used in this work is the standard paraffin process, which is widely used in clinical routine. The main goal is to preserve the original tissue structure and molecular composition, allowing to observe it in a light microscope. The complete preparation procedure includes steps such as fixation, dehydration, clearing, infiltration, embedding, and trimming [188]. In order be mounted on slides, sections of 3 μm (1 μm = 1×10^{-6}m) were cut using a microtome. After staining, the sections were covered with a glass coverslip. Then the anatomopathologists identify the tumoral areas in each slide, by visual analysis of tissue sections under a microscope. Final diagnosis of each case was produced by experienced pathologists and confirmed by complementary exams such as the IHC analysis.

To date, the database is composed of 9,353 microscopic images of breast tumor tissues divided into benign and malignant tumors. Both breast tumors, benign and malignant, can be sorted into different types based on the way the tumoral cells look under the microscope. At this time, there are four histological distinct types of breast benign tumors: adenosis (A), fibroadenoma (F), phyllodes tumor (PT), and tubular adenona (TA); and four malignant tumors (breast cancer): carcinoma (DC), lobular carcinoma (LC), mucinous carcinoma (MC) and papillary carcinoma (PC) in the dataset. Table B.1 shows these types with respective acronyms.

Table B.1: Types of breast tumors present in the BreaKHis database.

Benign		Malignant	
Acronym	Denomination	Acronym	Denomination
A	Adenosis	DC	Ductal Carcinoma
F	Fibroadenoma	LC	Lobular Carcinoma
PT	Phyllodes Tumor	MC	Mucinous Carcinoma
TA	Tubular Adenoma	PC	Papillary Carcinoma

Source: The author (2014).

B.1 Image Acquisition

An Olympus BX-50 system microscope with a relay lens with magnification of 3.3× coupled to a Samsung digital color camera SCC-131AN, both shown in Figure B.1(a), was used to obtain digitized images from the breast tissue slides. This camera uses a 1/3" Sony Super-HAD™(Hole-Accumulation Diode) IT (Interline Transfer) CCD (Charge-Coupled Device) with pixel size 6.5μm ×6.25 μm and total pixel number 752×582. Images were acquired in a 3-channel RGB (Red-Green-Blue) TrueColor (24-bit color depth, 8 bits per color channel) color space using magnifying factors of 40×, 100×, 200× and 400×, corresponding to objective lens 4×, 10×, 20×, and 40× with ocular len 10×. The magnification power of the microscope is the product of the magnifications of all the lenses in the system [26], e.g., 10× ocular lens and 40× objective lens gives microscope magnification $10 * 40 = 400\times$. Increasing the magnification factor reduces the field of view and reveals more details. Field of view is the diameter of the circle of view seen when looking through the microscope. In Figure B.2 four digitized images can be seen — with the four magnification factors (a) 40×, (b) 100×, (c) 200×, and (d) 400× — acquired from a single slide of breast tissue containing a malignant tumor. The highlighted rectangle (manually added for illustrative purposes only) is the area of interest selected by the pathologist to be detailed in the next higher magnification factor. Pathologists usually use distinct magnifications in order to get an overview of the slide (scanning lens, e. g., objective 4×), view the whole or large portions of specimen (low power lens, e.g., 10× and 20× objective lens), and see the small, detailed parts of specimen (high power lens, e.g., 40×).

The camera is set for automatic exposure and focusing is done manually on the microscope looking at the digital image on the computer screen. Table B.2 shows the effective pixel size in microns for each magnifying factor and the objective lens we have used. The pixel size is the physical pixel size of the camera (6.5μm), divided by the relay lens magnification (3.3) and the objective lens.

The original image contains black borders on both the left and right sides, and text annotations in the upper left corner. In order to remove these undesired areas, the resulting images were cropped and saved in 3-channel Red-Green-Blue (RGB), with an 8-bit depth in each channel, using Portable Network Graphics (PNG) format with no compression, with a dimension of 700×460 pixels. Resulting images are raw images without normalization or color standardization.

Regarding the selection of image fields, the pathologists tried to maximize the number of images containing a single type of tumor. This means that most of the images contain a single

Table B.2: Magnification and digital resolution of the system.

Visual Magnification	Objective Lens	NA (Numerical Aperture)	Effective Pixel Size (μm)
40×	4×	0.10	0.49
100×	10×	0.25	0.20
200×	20×	0.40	0.10
400×	40×	0.65	0.05

Source: The author (2015).

type of tumor, but some of them also include transitional tissue (normal-pathologic). The process of acquiring images of different magnifications works as follows: First the pathologist identifies the tumor and sets a Region of Interest (ROI). To cover the whole ROI, several images are captured using the lowest magnification, i.e., 40×. The pathologist preferably selects images with a single type of tumor (majority of the cases), but some of the images also include transitional tissue, e.g., normal-pathological. In average, a total of 24 images per patient is captured from each slide using the lowest magnification. Then, the magnification is manually increased to 100× and a similar number of images is captured inside the initial ROI. This process is repeated for 200× and 400× magnifications, respectively. Resulting images were examined by a pathologist to guarantee that they were adequate for visual inspection and diagnosis. A final visual (i.e., manual) inspection discards out-of-focus images. Focusing is done manually on the microscope looking at the digital image on the computer screen.

B.2 Dataset Structure

The dataset BreaKHis is divided into two main groups: benign tumors and malignant tumors. Histologically benign is a term referring to a lesion that does not match any criteria of malignancy – e.g., marked cellular atypia, mitosis, disruption of basement membranes, metastasize, etc. Normally, benign tumors are relatively "innocent", present slow growing and remain localized. Malignant tumor is a synonym for cancer: the lesion can invade and destroy adjacent structures (locally invasive) and spread to distant sites (metastasize), causing death. Carcinoma is a denomination for those types of cancer originated from epithelial cells, i.e., that start in cells which make up the tissues lining the inner or outer surfaces of the body, such as skin or the tissue lining organs like the liver or kidneys. Breast cancer is a type of carcinoma — more precisely, adenocarcinoma — which begins in glandular tissue. The distribution of benign and malignant main groups is shown in Table B.3.

Both benign and malignant breast tumors can be sorted into different types based on the way the tumoral cells look under the microscope. Various types/subtypes of breast tumors can have different prognoses and treatment implications. At this time, there are four types of histological distinct types of benign breast tumors and four types of malignant breast tumors (breast cancer) in the BreaKHis dataset. Table B.4 shows these types with the respective acronyms. Figures B.3 and B.4 show typical patterns of the four benign tumors and the four malignant tumors, respectively.

(a) Olympus BX-50 microscope and Samsung digital camera SCC-131A, equipment used to acquire the images from the microscope slides.

(b) Microscope slides (HE stain). The three above are samples collected using CNB procedure and the three below were collected by SOB method.

Figure B.1: Microscope, CCD Digital Camera and Microscope Slides (HE stain).
Source: Courtesy of the P&D Lab. Modified by the author (2014).

Currently, most samples present in dataset were collected by the SOB method. Thus, numbers relative to benign samples, collected by the SOB method are shown in Table B.5 and the SOB malignant samples in Table B.6. SOB is also called partial mastectomy or excisional biopsy. This type of procedure, compared to any methods of needle biopsy, removes a larger size of tissue sample and is done in a hospital with general anesthetic.

Also, in the dataset there are some samples collected by CNB. In this procedure, a core needle, or hollow core needle, can be used to get small tissue samples from a breast lump, as well as nearby healthy breast tissue. CNB is more accurate than fine needle aspiration and less invasive compared to the SOB method. On the other hand, there is always the possibility that

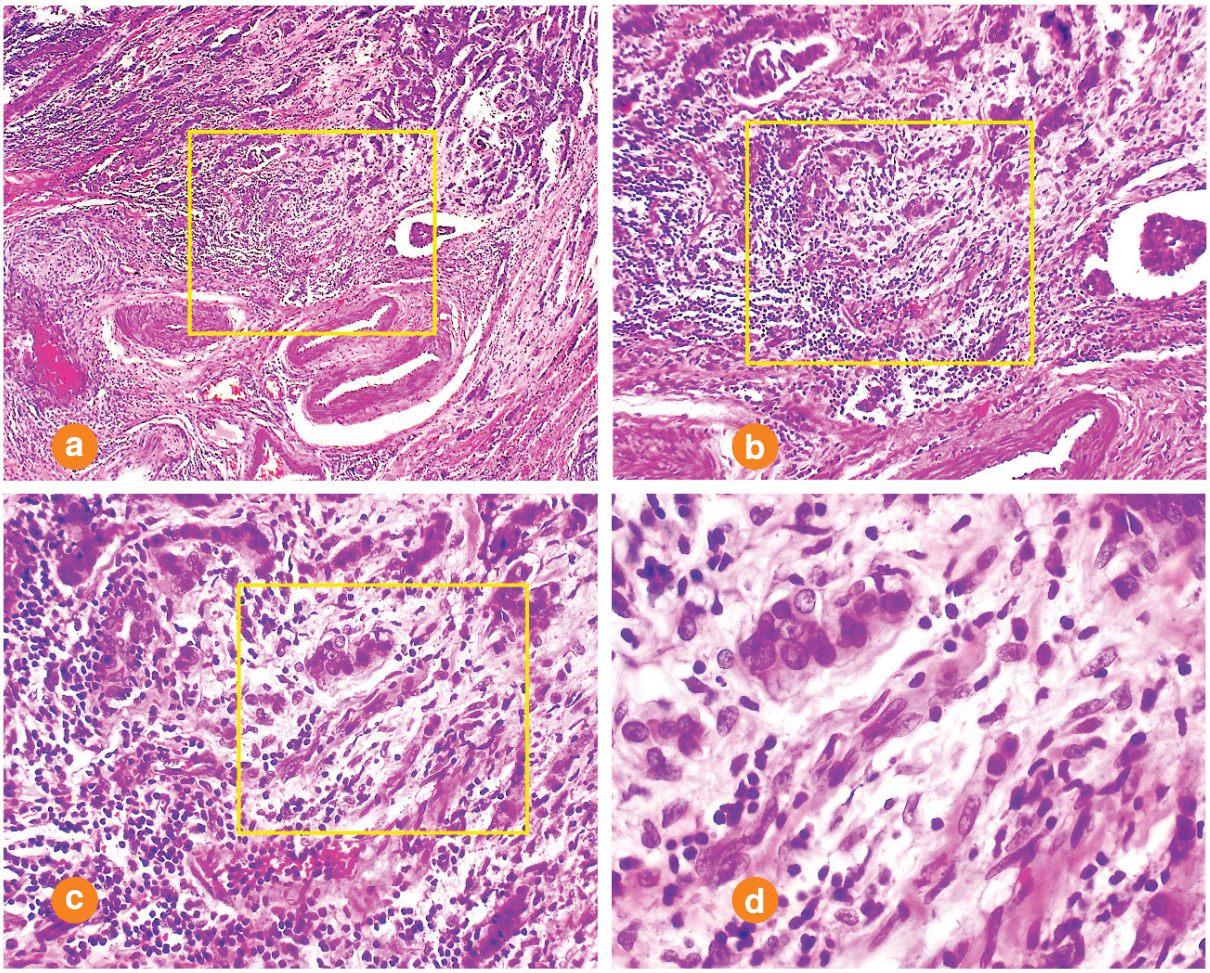


Figure B.2: A slide of malignant breast tumor (stained with HE) seen in different magnification factors: (a) 40×, (b) 100×, (c) 200×, and (d) 400×. Highlighted rectangle (manually added for illustrative purposes only) is the area of interest selected by the pathologist to be detailed in the next higher magnification factor.
Source: The author (2014).

Table B.3: Images distribution by magnification factor and main group.

Magnification	Benign	Malignant	Total
40×	652	1,649	2,376
100×	674	1,749	2,485
200×	652	1,648	2,355
400×	614	1,471	2,137
Total	2,592	6,517	**9,109**
# slides	26	71	**97**

Source: The author (2015).

the needle may miss a malignant area. Tables B.7 and B.8 present the distribution of CNB for benign and malignant groups, respectively.

Table B.4: Types of breast tumors present in database.

Benign		Malignant	
Acronym	Denomination	Acronym	Denomination
A	Adenosis	DC	Ductal Carcinoma
F	Fibroadenoma	LC	Lobular Carcinoma
PT	Phyllodes Tumor	MC	Mucinous Carcinoma
TA	Tubular Adenoma	PC	Papillary Carcinoma

Source: The author (2015).

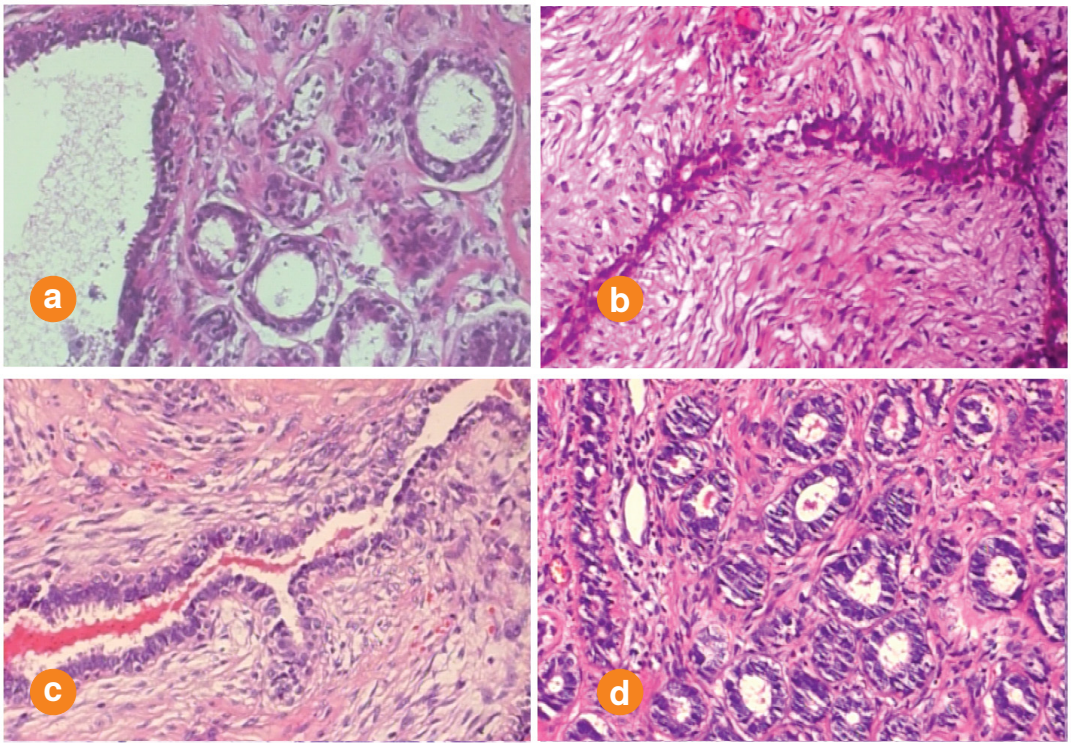


Figure B.3: Images from the dataset showing typical patterns of four benign breast tumors (stained with HE) as viewed under a microscope at a 100× magnification factor: (a) adenosis, (b) fibroadenoma, (c) phyllodes tumor, and (d) tubular adenoma.
Source: The author (2014).

In the BreaKHis dataset, each image filename stores information about the image itself: method of biopsy procedure, tumor class, tumor type, slide identification, and magnification factor. For example, `SOB_B_TA-14-16184-100-001.png` is the image 001, at magnification factor 100×, of a benign tumor of the tubular adenoma type, original from the slide 14-16184, which was collected by the SOB procedure. More formally, the format of image file name is given in Listing B.1, following a variant of the BNF (Backus-Naur Form) notation. Besides, the images are organized in folders and Figure B.5 shows a hierarchical view of the dataset.

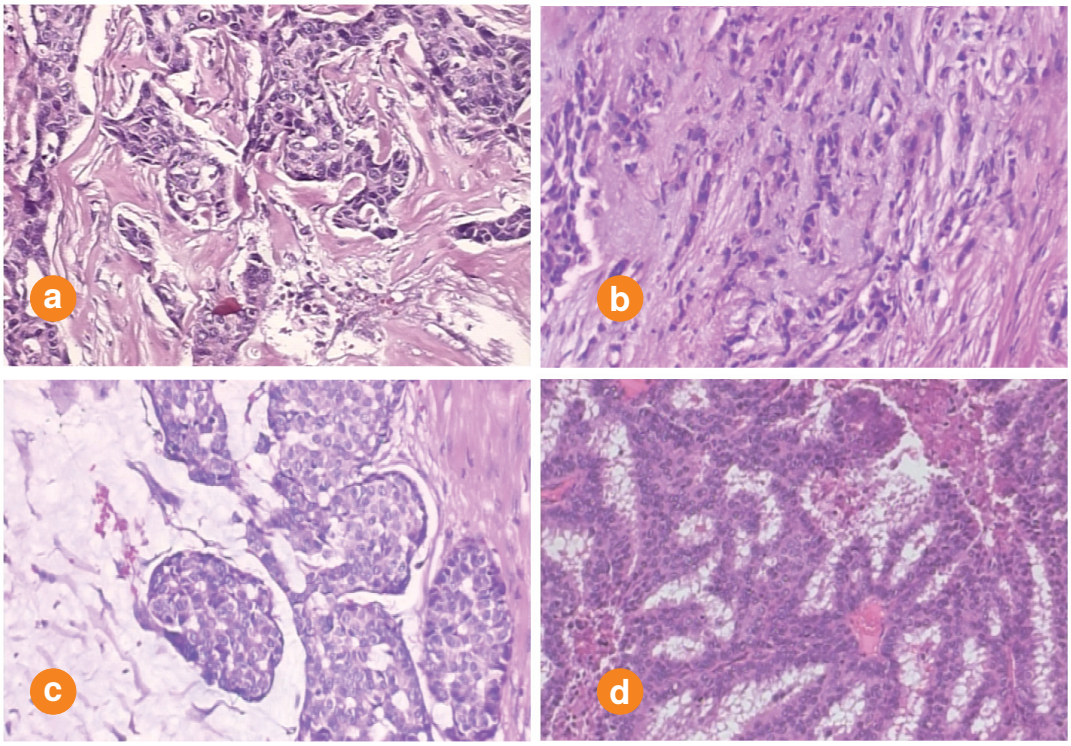


Figure B.4: Images from the dataset showing typical patterns of four malignant breast tumors (stained with HE) as viewed under a microscope at a 100× magnification factor: (a) ductal carcinoma, (b) lobular carcinoma, (c) mucinous carcinoma, and (d) papillary carcinoma.
Source: The author (2014).

Table B.5: Images distribution by magnification factor and benign subtypes – SOB procedure.

Magnification	Adenosis	Fibroadenoma	Tubular Adenoma	Phyllodes Tumor	Total
40×	114	253	122	109	598
100×	113	260	120	121	614
200×	111	264	111	108	594
400×	106	237	104	115	562
Total	444	1014	457	453	**2,368**
# slides	4	10	7	3	**24**

Source: The author (2015).

Table B.6: Images distribution by magnification factor and malignant subtypes – SOB procedure.

Magnification	Ductal	Lobular	Mucinous	Papillary	Total
40×	864	156	205	145	1,370
100×	903	170	222	142	1,437
200×	896	163	196	135	1,390
400×	788	137	169	138	1,232
Total	3,451	626	792	560	**5,429**
# slides	38	5	9	6	**58**

Source: The author (2015).

Table B.7: Images distribution by magnification factor and benign subtypes – CNB procedure.

Magnification	Tubular Adenoma	Total
40×	54	54
100×	60	60
200×	58	58
400×	52	52
Total	224	**224**
# slides	2	**2**

Source: The author (2015).

Table B.8: Images distribution by magnification factor and malignant subtypes – CNB Procedure.

Magnification	Ductal	Lobular	Mucinous	Total
40×	240	23	16	279
100×	269	29	14	312
200×	218	22	18	258
400×	199	25	15	239
Total	926	99	63	**1088**
# slides	11	1	1	**13**

Source: The author (2015).

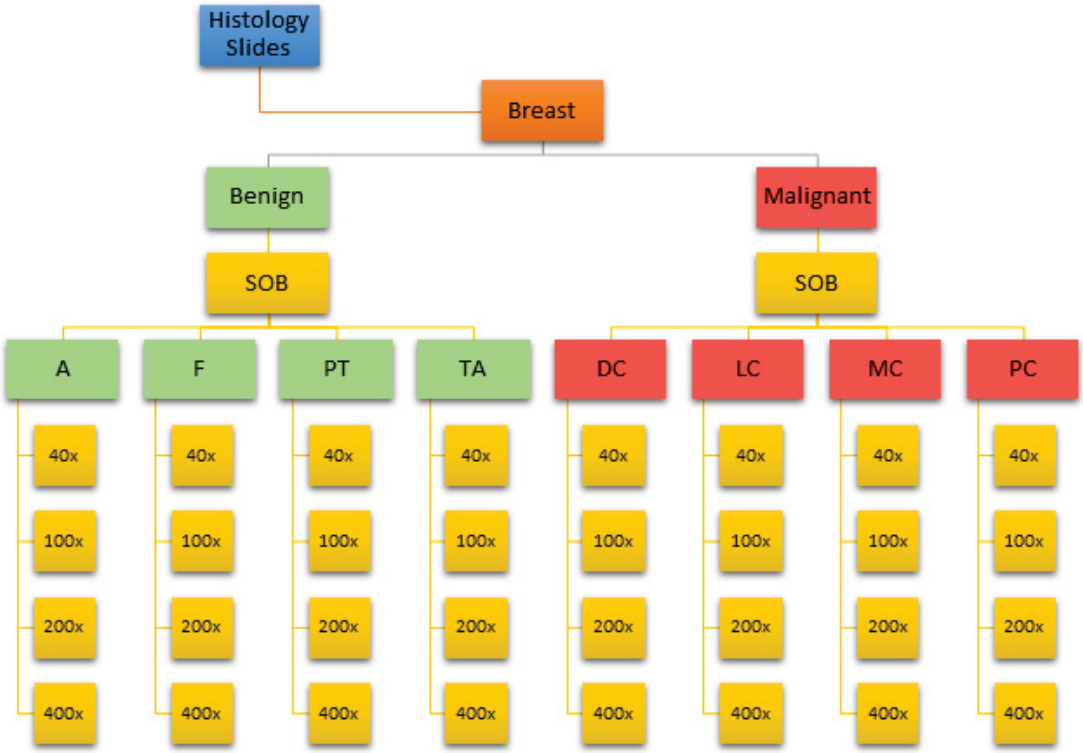


Figure B.5: Distribution of the images in the dataset.
Source: The author (2014).

```
<BIOPSY_PROCEDURE> "_" <TUMOR_CLASS> "_"
  <TUMOR_TYPE> "-" <YEAR> "-" <SLIDE_ID> "-"
  <MAG> "-" <SEQ>

<BIOPSY_PROCEDURE> ::= "CNB" | "SOB"
<TUMOR_CLASS> ::= "M" | "B"
<TUMOR_TYPE> ::= <BENIGN_TYPE> | <MALIGNANT_TYPE>
<BENIGN_TYPE> ::= "A" | "F" | "PT" | "TA"
<MALIGNANT_TYPE> ::= "DC" | "LC"| "MC"| "PC"
<YEAR> ::= <DIGIT><DIGIT>
<SLIDE_ID> ::= <INTEGER><SECTION>
<SEQ> ::= <INTEGER>
<MAG> ::= "40" | "100" | "200" | "400"
<INTEGER> ::= <INTEGER><DIGIT> | <DIGIT>
<SECTION> ::= <SECTION><LETTER> | <LETTER> | ""
<DIGIT> ::= "0" | ... | "9"
<LETTER> ::= "A" | ... | "Z"
```

Listing B.1: Format of image file name represented with a variant of BNF (Backus-Naur Form) notation. Terminal symbols are surrounded by quotes. An ellipsis symbol indicates a range of terminal values. An empty production (ϵ) is represented by "".

References

[1] Q. Abbas, M. E. Celebi, I. F. García. "Breast mass segmentation using region-based and edge-based methods in a 4-stage multiscale system". In: *Biomedical Signal Processing and Control* 8 (2 2013), pp. 204–214. DOI: 10.1016/j.bspc.2012.08.003.

[2] J. Abonyi, F. Szeifert. "Supervised fuzzy clustering for the identification of fuzzy classifiers". In: *Pattern Recognition Letters* 24.14 (2003), pp. 2195–2207.

[3] P. Agrawal, M. V. Vatsa, R. Singh. "Saliency based mass detection from screening mammograms". In: *Signal Processing* 99 (1 2014), pp. 29–47. DOI: 10.1016/j.sigpro.2013.12.010.

[4] M. F. Akay. "Support vector machines combined with feature selection for breast cancer diagnosis". In: *Expert Systems with Applications* 36.2 (2009), pp. 3240–3247.

[5] S. Akbar, M. Peikari, S. Salama, S. Nofech-Mozes, A. Martel. "Transitioning between Convolutional and Fully Connected Layers in Neural Networks". In: *Proceedings of the 3rd workshop on Deep Learning in Medical Image Analysis (DLMIA)*. (Quebec, Canada). Vol. 10553. Springer, Cham, Sept. 2017. DOI: 10.1007/978-3-319-67558-9.

[6] E. Alpaydin, V. Cheplygina, M. Loog, D. M. J. Tax. "Single- vs. multiple-instance classification". In: *Pattern Recognition* 48.9 (2015), pp. 2831–2838. DOI: 10.1016/j.patcog.2015.04.006.

[7] V. A. F. Alves, M. Bibbo, F. C. L. Schmitt, F. Milanezi, A. Longatto Filho. "Comparison of manual and automated methods of liquid-based cytology: a morphologic study". In: *Acta Cytologica* 2.48 (2004), pp. 187–193.

[8] W. C. Amalu, ed. *Case Studies*. Nov. 2012. URL: http://www.breastthermography.com/case_studies.htm (visited on 07/2014).

[9] American Cancer Society (ACS). *Breast Cancer Detailed Guide*. 2014. URL: http://www.cancer.org/acs/groups/cid/documents/webcontent/003090-pdf.pdf (visited on 09/2014).

[10] J. Amores. "Multiple instance classification: Review, taxonomy and comparative study". In: *Artificial Intelligence* 201 (2013), pp. 81–105. DOI: 10.1016/j.artint.2013.06.003.

[11] I. I. Andreadis, G. M. Spyrou, K. S. Nikita. "A CADx Scheme for Mammography Empowered With Topological Information From Clustered Microcalcifications' Atlases".

In: *IEEE Journal of Biomedical and Health Informatics* 19 (1 2015), pp. 166–173. DOI: 10.1109/JBHI.2014.2334491.

[12] S. Andrews, I. Tsochantaridis, T. Hofmann. "Support Vector Machines for Multiple-instance Learning". In: *Proceedings of the 2001 Conference on Advances in Neural Information Processing Systems 15*. (Vancouver, Canada). MIT Press, Dec. 2002, pp. 577–584.

[13] M. C. Araújo, R. C. F. Lima, R. M. C. R. d. Souza. "Interval symbolic feature extraction for thermography breast cancer detection". In: *Expert Systems with Applications* 41 (15 2014), pp. 6728–6737. DOI: 10.1016/j.eswa.2014.04.027.

[14] N. Arora, D. Martins, D. Ruggerio, E. Tousimis, A. J. Swistel, M. P. Osborne, R. M. Simmons. "Effectiveness of a noninvasive digital infrared thermal imaging system in the detection of breast cancer". In: *The American Journal of Surgery* 196.4 (2008), pp. 523–526.

[15] D. E. Axelrod, N. A. Miller, H. L. Lickley, J. Qian, W. A. Christens-Barry, Y. Yuan, Y. Fu, J.-A. W. Chapman. "Effect of quantitative nuclear image features on recurrence of ductal carcinoma in situ (DCIS) of the breast". In: *Cancer Informatics* 6 (2008), pp. 99–109. DOI: 10.4137/CIN.S401. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3994718/.

[16] A. T. Azar, S. A. El-Said. "Probabilistic neural network for breast cancer classification". In: *Neural Computing & Applications* 23.6 (2012), pp. 1737–1751.

[17] A. Babenko, A. Slesarev, A. Chigorin, V. Lempitsky. "Neural Codes for Image Retrieval". In: *Proceedings of the 13th European Conference of Computer Vision (ECCV)*. (Zurich, Switzerland). Springer International Publishing, Sept. 2014, pp. 584–599. DOI: 10.1007/978-3-319-10590-1_38.

[18] K. Bache, M. Lichman. *UCI Machine Learning Repository*. URL: http://archive.ics.uci.edu/ml (visited on 06/2014).

[19] A. M. Bagci, A. E. Cetin. "Detection of microcalcifications in mammograms using local maxima and adaptive wavelet transform analysis". In: *Electronics Letters* 38 (22 2002), pp. 1311–1313. DOI: 10.1049/el:20020907.

[20] Y. B. Bar, I. Diamant, L. Wolf, H. Greenspan. "Deep learning with non-medical training used for chest pathology identification". In: *Proceedings of SPIE Medical Imaging 2015: Computer-Aided Diagnosis*. (Orlando, USA). Ed. by L. M. Hadjiiski, G. D. Tourassi. Vol. 9414. Society of Photo-Optical Instrumentation Engineers (SPIE), Mar. 2015. ISBN: 9781628415049. DOI: 10.1117/12.2083124.

[21] V. Barth. *Diagnosis of breast diseases: integrating the findings of clinical presentation, mammography, and ultrasound*. New York: Tyeme Verlag, 2011. ISBN: 978-3-13-143831-7.

[22] A. Basavanhally, S. Ganesan, M. Feldman, N. Shih, C. Mies, J. Tomaszewski, A. Madabhushi. "Multi-field-of-view framework for distinguishing tumor grade in ER+

breast cancer from entire histopathology slides". In: *IEEE Transactions On Biomedical Engineering* 60 (8 2013), pp. 2089–2099. DOI: 10.1109/TBME.2013.2245129.

[23] H. Bay, T. Tuytelaars, L. van Gool. "SURF: Speeded Up Robust Features". In: *Proceedings of the 9th European Conference on Computer Vision (ECCV)*. (Graz, Austria). Ed. by A. Leonardis, H. Bischof, A. Pinz. Springer, May 2006, pp. 404–417. DOI: 10.1007/11744023_32.

[24] H. Bay, A. Ess, T. Tuytelaars, L. Van Gool. "Speeded-Up Robust Features (SURF)". In: *Computer Vision and Image Understanding* 110.3 (June 2008), pp. 346–359. ISSN: 1077-3142. DOI: 10.1016/j.cviu.2007.09.014. URL: http://dx.doi.org/10.1016/j.cviu.2007.09.014.

[25] N. Bayramoglu, J. Kannala, J. Heikkilä. "Deep Learning for Magnification Independent Breast Cancer Histopathology Image Classification". In: *23rd International Conference on Pattern Recognition*. (Cancun, Mexico). Vol. 1. IEEE, Dec. 2016, pp. 2440–2445. DOI: 10.1109/ICPR.2016.7900002.

[26] S. Bell, K. Morris. *An introduction to microscopy*. Boca Raton: CRC Press, 2010. ISBN: 978-1-4200-8450-4.

[27] Y. Bengio, A. Courville. "Deep Learning of Representations". In: *Handbook on Neural Information Processing*. Vol. 49. Springer: Berlin Heidelberg, 2013. ISBN: 978-3-642-36656-7. DOI: 10.1007/978-3-642-36657-4.

[28] Y. Bengio, A. Courville, P. Vincent. "Representation Learning: A Review and New Perspectives". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35 (8 2013), pp. 1798–1828. DOI: 10.1109/TPAMI.2013.50.

[29] Y. Bengio, I. J. Goodfellow, A. Courville. "Deep Learning". Book in preparation for MIT Press. 2015. URL: http://www.iro.umontreal.ca/~bengioy/dlbook.

[30] A. Bhardwaj, A. Tiwari. "Breast cancer diagnosis using Genetically Optimized Neural Network model". In: *Expert Systems with Applications* 42 (10 2015), pp. 4611–4620. DOI: 10.1016/j.eswa.2015.01.065.

[31] C. M. Bishop. *Pattern Recognition and Machine Learning*. 2nd ed. New York: Springer, 2006.

[32] S. K. Biswas, D. P. Mukherjee. "Recognizing Architectural Distortion in Mammogram: A Multiscale Texture Modeling Approach with GMM". In: *IEEE Transactions on Biomedical Engineering* 58 (7 2011), pp. 2023–2030. DOI: 10.1109/TBME.2011.2128870.

[33] D. Bliss. *NCI Visuals Online*. 2007. URL: https://visualsonline.cancer.gov/details.cfm?imageid=4353 (visited on 07/2014).

[34] L. Bottou. "Stochastic Gradient Tricks". In: *Neural Networks, Tricks of the Trade, Reloaded*. Ed. by G. Montavon, G. B. Orr, K.-R. Müller. Lecture Notes in Computer Science (LNCS 7700). Springer, 2012, pp. 430–445. URL: http://leon.bottou.org/papers/bottou-tricks-2012.

[35] A. Bovik. *The Essential Guide to Image Processing*. Burlington: Elsevier, 2009.

[36] C. Bovik, M. Clark, W. S. Geisler. "Multichannel Texture Analysis Using Localized Spatial Filters". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12 (1 1990), pp. 55–73. DOI: 10.1007/BF00204594.

[37] P. Boyle, B. Levin, eds. *World Cancer Report 2008*. Lyon: IARC, 2008. ISBN: 978-9-283-20423-7. URL: http://www.iarc.fr/en/publications/pdfs-online/wcr/2008/wcr_2008.pdf.

[38] G. Bradski. "The OpenCV Library". In: *Dr. Dobb's Journal of Software Tools* 25.11 (2000), pp. 120–125. URL: http://www.drdobbs.com/open-source/the-opencv-library/184404319.

[39] S. S. Brandt, G. Karemore, N. Karssemeijer, M. Nielsen. "An Anatomically Oriented Breast Coordinate System for Mammogram Analysis". In: *IEEE Transactions on Medical Imaging* 30 (10 2011), pp. 1841–1851. DOI: 10.1109/TMI.2011.2155082.

[40] G. Brasileiro Filho, ed. *Bogliolo Patologia*. 7th ed. Rio de Janeiro: Guanabara Koogan, 2006.

[41] *Breast Cancer Dataset*. URL: ftp://ftp.cs.technion.ac.il/pub/projects/medic-image (visited on 02/2015).

[42] Breastcancer.org. *Biopsy*. 2012. URL: http://www.breastcancer.org/symptoms/testing/types/biopsy (visited on 06/2014).

[43] Breastcancer.org. *Types of Breast Cancer*. 2013. URL: http://www.breastcancer.org/symptoms/types (visited on 06/2014).

[44] Breastcancer.org. *Male Breast Cancer*. 2014. URL: http://www.breastcancer.org/symptoms/types/male_bc (visited on 06/2014).

[45] L. Breiman, J. H. Friedman, R. Olshen, C. Stone. *Classification and Regression Trees*. Pacific Grove: Wadsworth & Brooks, 1984.

[46] L. Breiman. "Random Forests". In: *Machine Learning* 45 (1 2001), pp. 5–32. DOI: 10.1023/A:1010933404324. URL: http://dx.doi.org/10.1023/A%3A1010933404324.

[47] A. A. Britto Jr, R. Sabourin, L. E. S. Oliveira. "Dynamic Selection of Classifiers - A Comprehensive Review". In: *Pattern Recognition* 47 (11 2014), pp. 3665–3680. DOI: 10.1016/j.patcog.2014.05.003.

[48] D. O. T. Bruno, M. Z. do Nascimento, R. P. Ramos, V. R. Batista, L. A. Neves, A. S. Martins. "LBP operators on curvelet coefficients as an algorithm to describe texture in breast cancer tissues". In: *Expert Systems with Applications* 55 (15 2016), pp. 329–340. DOI: 10.1016/j.eswa.2016.02.019.

[49] M. Calonder, V. Lepetit, C. Strecha, P. Fua. "BRIEF:Binary robust independent elementary features". In: *Proceedings of the 11th European Conference on Computer Vision (ECCV)*. (Crete, Greece). Ed. by K. Daniilidis, P. Maragos, N. Paragios. Springer Berlin Heidelberg,

Sept. 2010, pp. 778–792. ISBN: 978-3-642-15560-4. DOI: 10.1007/978-3-642-15561-1_56.

[50] Cancer Research UK (CRUK). *Diagram 3 of 3 showing stage 3C breast cancer CRUK 401*. 2014. URL: http://commons.wikimedia.org/wiki/File:Diagram_3_of_3_showing_stage_3C_breast_cancer_CRUK_401.svg#/media/File:Diagram_3_of_3_showing_stage_3C_breast_cancer_CRUK_401.svg (visited on 06/2014).

[51] Cancer Research UK (CRUK). *Diagram showing ductal carcinoma in situ (DCIS) CRUK 115*. 2014. URL: http://commons.wikimedia.org/wiki/File:Diagram_showing_ductal_carcinoma_in_situ_(DCIS)_CRUK_115.svg#/media/File:Diagram_showing_ductal_carcinoma_in_situ_(DCIS)_CRUK_115.svg (visited on 06/2014).

[52] Cancer Research UK (CRUK). *Types of breast cancer*. 2014. URL: http://www.cancerresearchuk.org/cancer-help/type/breast-cancer/about/types/invasive-lobular-breast-cancer (visited on 06/2014).

[53] M.-A. Carbonneau, V. Cheplygina, E. Granger, G. Gagnon. "Multiple Instance Learning: A Survey of Problem Characteristics and Applications". In: *Pattern Recognition* 77 (2017), pp. 329–353. DOI: 10.1016/j.patcog.2017.10.009.

[54] P. Carmeliet, R. K. Jain. "Angiogenesis in cancer and other diseases". In: *Nature* 407 (2000), pp. 249–257.

[55] S. Cascianelli, R. Bello-Cerezo, F. Bianconi, M. L. Fravolini, M. Belal, B. Palumbo, J. N. Kather. "Dimensionality Reduction Strategies for CNN-Based Classification of Histopathological Images". In: *Proceedings of the 2017 International Conference on Intelligent Interactive Multimedia Systems and Services (KES-IIMSS)*. (Vilamoura, Portugal). Vol. 76. Springer, Cham, June 2017, pp. 21–30. DOI: 10.1007/978-3-319-59480-4_3.

[56] D. Cascio, F. Fauci, R. Magro, G. Raso, R. Bellotti, F. De Carlo, S. Tangaro, G. De Nunzio, M. Quarta, G. Forni, A. Lauria, M. E. Fantacci, A. Retico, G. L. Masala, P. Oliva, S. Bagnasco, S. Cheran, E. L. Torres. "Mammogram Segmentation by Contour Searching and Mass Lesions Classification With Neural Network". In: *IEEE Transactions on Nuclear Science* 53 (5 2006), pp. 2827–2833. DOI: 10.1109/TNS.2006.878003.

[57] P. Casti, A. Mencattini, M. Salmeri, R. M. Rangayyan. "Analysis of Structural Similarity in Mammograms for Detection of Bilateral Asymmetry". In: *IEEE Transactions on Medical Imaging* 34 (2 2015), pp. 662–671. DOI: 10.1109/TMI.2014.2365436.

[58] *Car practice guidelines and technical standards for breast imaging and intervention*. Tech. rep. Ottawa: Canadian Association of Radiologists (CAR), 2013. URL: http://www.car.ca/uploads/standards%20guidelines/20131024_en_breast_imaging_practice_guidelines.pdf (visited on 06/2014).

[59] A. Chan, J. A. Tuszynskim. "Automatic prediction of tumour malignancy in breast cancer with fractal dimension". In: *Royal Society Open Science* 3 (12 2016), pp. 1–10. DOI: 10.1098/rsos.160558.

[60] J. Chang, J. Yu, T. Han, H.-j. Chang, E. Park. "A Method for Classifying Medical Images using Transfer Learning: A Pilot Study on Histopathology of Breast Cancer". In: *Proceedings of the 2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom)*. (Dalian, China). IEEE, Oct. 2017. DOI: 10.1109/HealthCom.2017.8210843.

[61] k. Chatfield, K. Simonyan, A. Vedaldi, A. Zisserman. "Return of the Devil in the Details: Delving Deep into Convolutional Nets". In: *CoRR* abs/1405.3531 (2014). URL: http://arxiv.org/abs/1405.3531.

[62] H.-L. Chen, B. Yang, J. Liu, D.-Y. Liu. "A support vector machine classifier with rough set-based feature selection for breast cancer diagnosis". In: *Expert Systems with Applications* 38.7 (2011), pp. 9014–9022.

[63] Z. Chen, A. Oliver, E. Denton, R. Zwiggelaar. "Automated Mammographic Risk Classification Based on Breast Density Estimation". In: *Pattern Recognition and Image Analysis*. (Madeira, Portugal). Ed. by J. M. Sanches, L. Micó, J. S. Cardoso. Vol. 7887. Lecture Notes in Computer Science. Springer-Verlag Berlin Heidelberg, June 2013, pp. 237–244. ISBN: 978-3-642-38628-2. DOI: 10.1007/978-3-642-38628-2_28.

[64] Z. Chen, H. Strange, A. Oliver, E. R. E. Denton, C. Boggis, R. Zwiggelaar. "Topological Modeling and Classification of Mammographic Microcalcification Clusters". In: *IEEE Transactions on Biomedical Engineering* 62 (4 2015), pp. 1203–1214. DOI: 10.1109/TBME.2014.2385102.

[65] F. Chollet et al. *Keras*. https://keras.io. 2015.

[66] M. Cimpoi, S. Maji, I. Kokkinos, A. Vedaldi. "Deep Filter Banks for Texture Recognition, Description, and Segmentation". In: *International Journal of Computer Vision* 118.1 (2016), pp. 65–94. DOI: 10.1007/s11263-015-0872-3.

[67] D. C. Ciresan, A. Giusti, L. M. Gambardella, J. Schmidhuber. "Deep Neural Networks Segment Neuronal Membranes in Electron Microscopy Images". In: *Proceedings of 26th Annual Conference on Neural Information Processing Systems 2012 (NIPS)*. (Lake Tahoe, USA). Ed. by P. L. Bartlett, F. C. N. Pereira, C. J. C. Burges, L. Bottou, K. Q. Weinberger. Dec. 2012, pp. 2852–2860. URL: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.

[68] L. P. Coelho. "Mahotas: Open source software for scriptable computer vision". In: *Journal of Open Research Software* 1.3 (2013), pp. 131–155. DOI: http://dx.doi.org/10.5334/jors.ac.

[69] L. P. Coelho, A. Ahmed, A. Arnold, J. Kangas, A.-S. Sheikh, E. P. Xing, W. W. Cohen, R. F. Murphy. "Structured literature image finder: extracting information from text and images in biomedical literature". In: *Linking Literature, Information, and Knowledge for Biology*. Ed. by C. Blaschke, H. Shatkay. Vol. 6004. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2010, pp. 23–32. DOI: 10.1007/978-3-642-13131-8_4.

[70] C. Cortes, V. Vapnik. "Suport-Vector Networks". In: *Machine Learning* 20 (3 1995), pp. 273–297. DOI: 10.1023/A:1022627411411.

[71] A. Cruz-Roa, J. Arevalo Ovalle, A. Madabhushi, F. A. González Osorio. "A Deep Learning Architecture for Image Representation, Visual Interpretability and Automated Basal-Cell Carcinoma Cancer Detection". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*. Ed. by K. Mori, I. Sakuma, Y. Sato, C. Barillot, N. Navab. Vol. 8150. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2013, pp. 403–410. ISBN: 978-3-642-40762-8. DOI: 10.1007/978-3-642-40763-5_50.

[72] T. Damoulas, M. A. Girolam. "Combining feature spaces for classification". In: *Pattern Recognition* 42 (11 2009), pp. 2671–2683. DOI: 10.1016/j.patcog.2009.04.002.

[73] K. Das, S. P. K. Karriy Karri, A. G. Roy, J. Chatterjeey, D. Sheet. "Classifying Histopathology Whole-slides Using Fusion of Decisions from Deep Convolutional Network on a Collection of random Multi-views at Multi-magnification". In: *Proceedings of the 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI)*. (Melbourne, Australia). IEEE, Apr. 2017, pp. 1024–1027. DOI: 10.1109/ISBI.2017.7950690.

[74] *DDSM: Digital Database for Screening Mammography*. 2000. URL: http://marathon.csee.usf.edu/Mammography/Database.html (visited on 02/2015).

[75] C. Désir, C. Petitjean, L. Heutte, M. Salaun, L. Thiberville. "Classification of endomicroscopic images of the lung based on random subwindows and extra-trees". In: *IEEE Transactions on Biomedical Engineering* 59 (9 2012), pp. 2677–2683. DOI: 10.1109/TBME.2012.2204747.

[76] J. Dheeba, N. A. Singh, S. T. Selvi. "Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach". In: *Journal of Biomedical Informatics* 49 (2014), pp. 45–52. DOI: 10.1016/j.jbi.2014.01.010.

[77] M. Diakides, J. D. Bronzino, D. R. Peterson, eds. *Medical Infrared Imaging: Principles and Practices*. 3rd ed. Boca Raton: CRC Press, 2013.

[78] T. G. Dietterich, R. H. Lathrop, T. Lozano-Pérez. "Solving the multiple instance problem with axis-parallel rectangles". In: *Artificial Intelligence* 89.1 (1997), pp. 31–71. ISSN: 0004-3702.

[79] K. Dimitropoulos, P. Barmpoutis, C. Zioga, A. Kamas, K. Patsiaoura, N. Grammalidis. "Grading of invasive breast carcinoma through Grassmannian VLAD encoding". In: *PloS ONE* 12 (9 2017), pp. 1–18. DOI: 10.1371/journal.pone.0185110.

[80] H. A. Domanski, ed. *Atlas of fine needle aspiration cytology*. 1st ed. London: Springer-Verlag London, 2014. ISBN: 978-1-4471-2445-0.

[81] J. Donahue, J. Yangqing, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, T. Darrell. "DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition". In: *Proceedings of the 31th International Conference on Machine Learning (ICML)*. (Beijing, China). IMLS, June 2014, pp. 647–655.

[82] S. Doyle, S. Agner, A. Madabhushi, M. Feldman, J. Tomaszewski. "Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features". In: *Proceedings of the 5th IEEE International Symposium on Biomedical*

Imageging (ISBI): From Nano to Macro. (Paris, France). Vol. 61. IEEE, May 2008, pp. 496–499. DOI: 10.1109/ISBI.2008.4541041.

[83] S. Doyle, M. Feldman, N. Shih, J. Tomaszewski, A. Madabhushi. "Cascaded discrimination of normal, abnormal, and confounder classes in histopathology: Gleason grading of prostate cancer". In: *BMC Bioinformatics* 13 (282 2012). DOI: 10.1186/1471-2105-13-282. URL: http://www.biomedcentral.com/1471-2105/13/282.

[84] D. J. Dronkers, J. H. C. L. Hendriks, R. Holland, G. Rosenbusch, eds. *Practice of mammography: pathology – technique – interpretation – adjunct modalities*. New York: Tyeme Verlag, 2002.

[85] R. O. Duda, P. E. Hart, D. G. Stork. *Pattern Classification*. 2nd ed. New York: John Wiley & Sons, 2000.

[86] A. Eleyan. "Breast cancer classification using moments". In: *Proceedings of the 20th Signal Processing and Communications Applications Conference (SIU)*. (Muğla, Turkey). Ed. by A. Ö. Ercan. IEEE, Apr. 2012, pp. 1–4. DOI: 10.1109/SIU.2012.6204778.

[87] H. Elouedi, W. Meliani, Z. Elouedi, N. B. Amor. "A hybrid approach based on decision trees and clustering for breast cancer classification". In: *Proceedings of the 2014 6th International Conference of Soft Computing and Pattern Recognition (SoCPaR)*. (Tunis, Tunisia). IEEE, Aug. 2014, pp. 226–231. DOI: 10.1109/SOCPAR.2014.7008010.

[88] S. van Engeland, P. Snoeren, J. Hendriks, N. Karssemeijer. "A Comparison of Methods for Mammogram Registration". In: *IEEE Transactions on Medical Imaging* 22 (11 2003), pp. 1436–1444. DOI: 10.1109/TMI.2003.819273.

[89] A. J. Evans, E. A. Krupinski, R. S. Weinstein, L. Pantanowitz. "2014 American Telemedicine Association Clinical Guidelines for Telepathology: Another Important Step in Support of Increased Adoption of Telepathology for Patient Care". In: *Journal of Pathology Informatics* 6 (13 2015). DOI: 10.4103/2153-3539.153906.

[90] C. Farabet, C. Couprie, L. Najman, Y. LeCun. "Learning Hierarchical Features for Scene Labeling". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35 (8 2013), pp. 1915–1929. DOI: 10.1109/TPAMI.2012.231.

[91] O. Faust, U. R. Acharya, E. Y. K. Ng, T. J. Hong, W. Yu. "Application of infrared thermography in computer aided diagnosis". In: *Infrared Physics & Technology* 66 (2014), pp. 160–175. DOI: 10.1016/j.infrared.2014.06.001.

[92] T. Fawcett. "An introduction to ROC analysis". In: *Pattern Recognition Letters* 27 (8 2006), pp. 861–874. DOI: 10.1016/j.patrec.2005.10.010.

[93] J. Ferlay, I. Soerjomataram, M. Ervik, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D. Parkin, D. Forman, F. Bray. *GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11*. Tech. rep. 11. Lyon: International Agency for Research on Cancer, Jan. 2013. URL: http://globocan.iarc.fr (visited on 06/2014).

[94] R. J. Ferrari, R. M. Rangayyan, J. Desautels, A. F. Frère. "Analysis of asymmetry in mammograms via directional filtering with Gabor wavelets". In: *IEEE Transactions on Medical Imaging* 20 (9 2001), pp. 953–964. DOI: 10.1109/42.952732.

[95] P. Filipczuk, T. Fevens, A. Krzyżak, R. Monczak. "Computer-aided breast cancer diagnosis based on the analysis of cytological images of fine needle biopsies". In: *IEEE Transactions on Medical Imaging* 32.12 (2013), pp. 2169–2178. DOI: 10.1109/TMI.2013.2275151.

[96] A. H. Fischer, K. A. Jacobson, J. Rose, R. Zeller. "Hematoxylin and eosin staining of tissue and cell sections". In: *Cold Spring Harbor Protocols* (2008). DOI: 10.1101/pdb.prot4986.

[97] R. A. Fisher. "The use of multiple measurements in taxonomic problems". In: *Annals of Eugenics* 7 (2 1936), pp. 179–188. DOI: 10.1111/j.1469-1809.1936.tb02137.x. URL: http://onlinelibrary.wiley.com/doi/10.1111/j.1469-1809.1936.tb02137.x/pdf.

[98] I. Fogel, D. Sagi. "Gabor Filters as Texture Discriminator". In: *Biological Cybernetics* 61 (2 1989), pp. 103–113. DOI: 10.1007/BF00204594.

[99] P. Fonseca, J. Mendoza, J. Wainer, J. Ferrer, J. Pinto, J. Guerrero, B. Castaneda. "Automatic breast density classification using a convolutional neural network architecture search procedure". In: *Proceedings of SPIE Medical Imaging 2015: Computer-Aided Diagnosis*. (Orlando, USA). Ed. by L. M. Hadjiiski, G. D. Tourassi. Vol. 9414. Society of Photo-Optical Instrumentation Engineers (SPIE), Mar. 2015. ISBN: 9781628415049. DOI: 10.1117/12.2081576.

[100] J. Foulds, E. Frank. "A review of multi-instance learning assumptions". In: *The Knowledge Engineering Review* 25.1 (2010), pp. 1–25.

[101] S. V. Francis, M. Sasikala. "Automatic detection of abnormal breast thermograms using asymmetry analysis of texture features". In: *Journal of Medical Engineering & Technology* 37 (1 2013), pp. 17–21. DOI: 10.3109/03091902.2012.728674.

[102] I. Gardiner, U. Kuusk, B. B. Forster, A. Spielmann. "Breast magnetic resonance imaging". In: *BC Medical Journal* 47.10 (2005), pp. 543–548.

[103] E. D. Gelasca, B. Obara, D. Fedorov, K. Kvilekval, B. Manjunath. "A biosegmentation benchmark for evaluation of bioimage analysis methods". In: *BMC Bioinformatics* 10 (368 2009). DOI: 10.1186/1471-2105-10-368.

[104] Y. M. George, H. H. Zayed, M. I. Roushdy, B. M. Elbagoury. "Remote computer-aided breast cancer detection and diagnosis system Based on cytological images". In: *IEEE Systems Journal* 8.3 (2014), pp. 949–964. DOI: 10.1109/JSYST.2013.2279415.

[105] F. Ghaznavi, A. Evans, A. Madabhushi, M. Feldman. "Digital imaging in pathology:whole-slide imaging and beyond". In: *Annual Review of Pathology: Mechanisms of Disease* 8 (2013), pp. 331–359. DOI: 10.1146/annurev-pathol-011811-120902.

[106] R. K. Gibb, M. G. Martens. "The impact of liquid-based cytology in decreasing the incidence of cervical cancer". In: *Reviews in Obstetrics & Gynecology* 4 (2011), s2–s11.

[107] J. R. Gilbertson, J. Ho, L. Anthony, D. M. Jukic, Y. Yagi, A. V. Parwani. "Primary histologic diagnosis using automated whole slide imaging: a validation study". In: *BMC Clinical Pathology* 6 (4 2006). DOI: 10.1186/1472-6890-6-4. URL: http://www.biomedcentral.com/1472-6890/6/4.

[108] R. C. Gonzales, R. E. Woods. *Digital Image Processing*. 3rd ed. Upper Saddle River: Prentice Hall, 2007.

[109] J. Grim, P. Somol, M. Haindl, J. Daneš. "Computer-Aided Evaluation of Screening Mammograms Based on Local Texture Models". In: *IEEE Transactions on Image Processing* 18 (4 2009), pp. 765–773. DOI: 10.1109/TIP.2008.2011168.

[110] Z. Guo, L. Zhang, D. Zhang. "A completed modeling of local binary pattern operator for texture classification". In: *IEEE Transactions on Image Processing* 19 (6 2010), pp. 1657–1663. DOI: 10.1109/TIP.2010.2044957.

[111] V. Gupta, A. Bhavsar. "An Integrated Multi-scale Model for Breast Cancer Histopathological Image Classification with Joint Colour-Texture Features". In: *Proceedings of the 17th International Conference on Computer Analysis of Images and Patterns (CAIP)*. (Ystad, Sweden). Vol. 10425. Springer, Cham, Aug. 2017, pp. 354–366. DOI: 10.1007/978-3-319-64698-5_30.

[112] V. Gupta, A. Bhavsar. "Breast Cancer Histopathological Image Classification: Is Magnification Important?" In: *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. (Honolulu, USA). IEEE, July 2017, pp. 769–776. DOI: 10.1109/CVPRW.2017.107.

[113] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, B. Yener. "Histopathological Image Analysis: A Review". In: *IEEE Reviews in Biomedical Engineering* 2 (2009), pp. 147–171.

[114] L. G. Hafemann, L. E. S. Oliveira, P. Cavalin. "Forest Species Recognition Using Deep Convolutional Neural Networks". In: *Proceedings of the 22nd International Conference on Pattern Recognition (ICPR)*. (Stockholm, Sweden). IEEE, Aug. 2014, pp. 1109–1107. DOI: 10.1109/ICPR.2014.199.

[115] N. A. Hamilton, R. S. Pantelic, K. Hanson, R. D. Teasdale. "Fast automated cell phenotype image classification". In: *BMC Bioinformatics* 8 (110 2007). DOI: 10.1186/1471-2105-8-110. URL: http://www.biomedcentral.com/1471-2105/8/110.

[116] Z. Han, B. Wei, Y. Zheng, Y. Yin, S. Li Kejian Li. "Breast Cancer Multi-classification from Histopathological Images with Structured Deep Learning Model". In: *Scientific Reports* 7 (1 2017), pp. 1455–1462. DOI: 10.1038/s41598-017-04075-z.

[117] R. M. Haralick, I. Dinstein, K. Shanmugam. "Textural features for image classification". In: *IEEE Transactions On Systems Man And Cybernetics* 3.6 (1973), pp. 610–621. URL: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4309314.

[118] S. Haykin. *Neural Networks and Learning Machines*. 3rd ed. Upper Saddle River: Prentice Hall, 2009.

[119] K. He, X. Zhang, S. Ren, J. Sun. “Deep Residual Learning for Image Recognition”. In: *Proceedings of the 29th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. (Las Vegas, USA). IEEE, June 2016, pp. 770–778. DOI: 10.1109/CVPR.2016.90.

[120] M. Heath, K. Bowyer, D. Kopans, R. Moore, P. Kegelmeyer Jr. “The Digital Database for Screening Mammography”. In: *Proceedings of the 5th International Workshop on Digital Mammography (IWDM)*. (Toronto, Canada). Ed. by M. J. Yaffe. Medical Physics, June 2000, pp. 212–218.

[121] J. Heikkilä, V. Ojansivu. “Methods for local phase quantization in blur-insensitive image analysis”. In: *Proceedings of the 2009 International Workshop on Local and Non-Local Approximation in Image Processing (LNLA)*. (Tuusula, Finland). IEEE, Aug. 2009, pp. 104–111. DOI: 10.1109/LNLA.2009.5278397.

[122] F. Herrera, S. Ventura, R. Bello, C. Cornelis, A. Zafra, D. Sánchez-Tarragó, S. Vluymans. “Multiple instance learning”. In: *Multiple Instance Learning: Foundations and Algorithms*. Springer, 2016, pp. 17–33.

[123] C. S. Herrington, ed. *Muir’s Textbook of Pathology*. 5th ed. Boca Raton: CRC Press, 2014.

[124] S. H. Heywang-Koebrunner, I. Schreer, S. Barter. *Diagnostic breast imaging: mammography, sonography, magnetic resonance imaging, and interventional procedures*. 3rd ed. New York: Tyeme Verlag, 2014. ISBN: 978-3-13-102893-8.

[125] J. H. Hipwell, C. Tanner, W. R. Crum, J. A. Schnabel, D. J. Hawkes. “A New Validation Method for X-ray Mammogram Registration Algorithms Using a Projection Model of Breast X-ray Compression”. In: *IEEE Transactions on Medical Imaging* 26 (9 2007), pp. 1190–1200. DOI: 10.1109/TMI.2007.903569.

[126] S. A. Hoda, E. Brogi, F. C. Koerner, P. P. Rosen, eds. *Rosen’s Breast Pathology*. 4th ed. Philadelphia: Lippincott Williams & Wilkins, 2014.

[127] J. Hoffman, D. Wang, F. Yu, T. Darrell. “FCNs in the Wild: Pixel-level Adversarial and Constraint-based Adaptation”. In: *CoRR* abs/1612.02649 (2016).

[128] B.-W. Hong, B.-S. Sohn. “Segmentation of Regions of Interest in Mammograms in a Topographic Approach”. In: *IEEE Transactions on Information Technology in Biomedicine* 14 (1 2010), pp. 129–139. DOI: 10.1109/TITB.2009.2033269.

[129] L. Hoorntje, P. Peeters, W. Mali, I. B. Rinkesemail. “Vacuum-assisted Breast Biopsy: a critical review”. In: *European Journal of Cancer* 39.12 (2003), pp. 1676–1683.

[130] K. Hu, X. Gao, F. Li. “Detection of Suspicious Lesions by Adaptive Thresholding Based on Multiresolution Analysis in Mammograms”. In: *IEEE Transactions on Instrumentation and Measurement* 60 (2 2011), pp. 462–472. DOI: 10.1109/TIM.2010.2051060.

[131] C.-L. Huang, H.-C. Liao, M.-C. Chen. “Prediction model building and feature selection with support vector machines in breast cancer diagnosis”. In: *Expert Systems with Applications* 34 (2008), pp. 578–587.

[132] E. B. Hunt, J. Marin, P. J. Stone. *Experiments in induction*. New York: Academic Press, 1966.

[133] R. Hupse, N. Karssemeijer. "Use of Normal Tissue Context in Computer-Aided Detection of Masses in Mammograms". In: *IEEE Transactions on Medical Imaging* 28 (12 2009), pp. 2033–2041. DOI: 10.1109/TMI.2009.2028611.

[134] J. E. Iglesias, N. Karssemeijer. "Robust Initial Detection of Landmarks in Film-Screen Mammograms Using Multiple FFDM Atlases". In: *IEEE Transactions on Medical Imaging* 28 (11 2011), pp. 1815–1824. DOI: 10.1109/TMI.2009.2025036.

[135] Instituto Nacional de Câncer José de Alencar Gomes da Silva (Inca), Ministério da Saúde (MS), Coordenação de Prevenção e Vigilância (CONPREV). *Diagnóstico Histopatológico e Citopatológico das Lesões da Mama*. Tech. rep. Rio de Janeiro: Instituto Nacional de Câncer José de Alencar Gomes da Silva (Inca), 2002.

[136] Instituto Nacional do Câncer (INCA), Ministério da Saúde (MS). *Diretrizes brasileiras para o rastreamento do câncer do colo do útero*. Tech. rep. Rio de Janeiro: Instituto Nacional do Câncer (INCA), 2011. URL: http://bvsms.saude.gov.br/bvs/publicacoes/inca/rastreamento_cancer_colo_utero.pdf (visited on 07/2014).

[137] Instituto Nacional de Câncer José de Alencar Gomes da Silva (Inca), Ministério da Saúde (MS), Secretaria de Atenção à Saúde (SAS). *Estimativa 2016 – Incidência de Câncer no Brasil*. Tech. rep. Brasília: Instituto Nacional de Câncer José de Alencar Gomes da Silva (Inca), 2015. URL: http://www.inca.gov.br/wcm/dncc/2015/estimativa-2016.asp (visited on 12/2015).

[138] International Agency for Research on Cancer (IARC). *Latest world cancer statistics Global cancer burden rises to 14.1 million new cases in 2012: Marked increase in breast cancers must be addressed*. 2013. URL: http://www.iarc.fr/en/media-centre/pr/2013/pdfs/pr223_E.pdf (visited on 06/2014).

[139] H. Irshad, A. Veillard, L. Roux, D. Racoceanu. "Methods for Nuclei Detection, Segmentation, and Classification in Digital Histopathology: A Review – Current Status and Future Potential". In: *IEEE Reviews in Biomedical Engineering* 7 (2014), pp. 97–114. DOI: 10.1109/RBME.2013.2295804.

[140] A. K. Jain, R. P. W. Duin, J. Mao. "Statistical Pattern Recognition: A Review". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (1 2000), pp. 4–37.

[141] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. B. Girshick, S. Guadarrama, T. Darrell. "Caffe: Convolutional Architecture for Fast Feature Embedding". In: *CoRR* abs/1408.5093 (2014). URL: http://arxiv.org/abs/1408.5093.

[142] Z. Jia, X. Huang, E. I.-C. Chang, Y. Xu. "Constrained Deep Weak Supervision for Histopathology Image Segmentation". In: *IEEE Transactions on Medical Imaging* 36.11 (2017), pp. 2376–2388. DOI: 10.1109/TMI.2017.2724070.

[143] M. Jiang, S. Zhang, H. Li, D. N. Metaxas. “Computer-Aided Diagnosis of Mammographic Masses Using Scalable Image Retrieval”. In: *IEEE Transactions on Biomedical Engineering* 62 (2 2015), pp. 783–792. DOI: 10.1109/TBME.2014.2365494.

[144] J. E. Joy, E. E. Penhoet, D. B. Petitti, eds. *Saving women’s lives: strategies for improving breast cancer detection and diagnosis*. Washington, D.C.: National Academies Press, 2005. ISBN: 978-0-309-09213-5.

[145] M. A. Kahya, W. Al-Hayani, Z. Y. Algamal. “Classification of Breast Cancer Histopathology Images based on Adaptive Sparse Support Vector Machine”. In: *Journal of Applied Mathematics & Bioinformatics* 7 (1 2017), pp. 49–69.

[146] M. Karabatak. “A new classifier for breast cancer detection based on Naïve Bayesian”. In: *Measurment* 72 (2015), pp. 32–36. DOI: 10.1016/j.measurement.2015.04.028.

[147] M. Karabatak, M. C. Ince. “An expert system for detection of breast cancer based on association rules and neural network”. In: *Expert Systems with Applications* 36 (2 2009), pp. 3465–3469.

[148] B. Karaçali, A. Tözeren. “Automated detection of regions of interest for tissue microarray experiments: an image texture analysis”. In: *BMC Medical Imaging* 7 (2 2007). DOI: 10.1186/1471-2342-7-2. URL: http://www.biomedcentral.com/1471-2342/7/2.

[149] J. Kittler, M. Hatef, R. P. W. Duin, J. Matas. “On Combining Classifiers”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20 (3 1998), pp. 226–239. DOI: 10.1109/34.667881.

[150] C.-C. Ko, C.-Y. Tsai, C.-H. Lin, K.-S. Liao. “A computer-aided diagnosis system of breast intraductal lesion using histopathological images”. In: *Proceedings of the 29th International Conference on Image and Vision Computing New Zealand (IVCNZ)*. (Hamilton, New Zeland). ACM, Nov. 2014, pp. 212–217. ISBN: 978-1-4503-3184-5. DOI: 10.1145/2683405.2683430.

[151] M. Kowal, P. Filipczuk, A. Obuchowicz, J. Korbicz, R. Monczak. “Computer-aided diagnosis of breast cancer based on fine needle biopsy microscopic images”. In: *Computers in Biology and Medicine* 43.10 (2013), pp. 1563–1572.

[152] O. Z. Kraus, J. L. Ba, B. J. Frey. “Classifying and segmenting microscopy images with deep multiple instance learning”. In: *Bioinformatics* 32.12 (2016), pp. i52–i59. DOI: 10.1093/bioinformatics/btw252.

[153] B. Krawczyk, G. Schaefer. “A pruned ensemble classifier for effective breast thermogram analysis”. In: *Proceedings of the 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. (Osaka, Japan). IEEE, July 2013, pp. 7120–7123. DOI: 10.1109/EMBC.2013.6611199.

[154] B. Krawczyk, G. Schaefer. “An Analysis of Properties of Malignant Cases for Imbalanced Breast Thermogram Feature Classification”. In: *Proceedings of the 2013 2nd IAPR Asian Conference on Pattern Recognition (ACPR)*. (Naha, Japan). IEEE, Nov. 2013, pp. 305–309. ISBN: 978-1-4577-2176-2. DOI: 10.1109/ACPR.2013.45.

[155] B. Krawczyk, G. Schaefer. “Breast thermogram analysis using classifier ensembles and image symmetry features”. In: *IEEE Systems Journal* 8 (3 2014), pp. 921–928. DOI: 10.1109/JSYST.2013.2283135.

[156] B. Krawczyk, G. Schaefer, M. Woźniak. “Breast thermogram analysis using a cost-sensitive multiple classifier system”. In: *Proceedings of the IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*. (Hong Kong, China). Vol. 1. IEEE, Jan. 2012, pp. 507–510. ISBN: 978-1-4577-2176-2. DOI: 10.1109/BHI.2012.6211629.

[157] B. Krawczyk, G. Schaefer, M. Woźniak. “Combining one-class classifiers for imbalanced classification of breast thermogram features”. In: *Proceedings of the 2013 IEEE Fourth International Workshop on Computational Intelligence in Medical Imaging (CIMI)*. (Singapore). IEEE, Apr. 2013, pp. 36–41. ISBN: 978-1-4673-5919-1. DOI: 10.1109/CIMI.2013.6583855.

[158] A. Krizhevsky, I. Sutskever, G. E. Hinton. “ImageNet Classification with Deep Convolutional Neural Networks”. In: *Proceedings of 26th Annual Conference on Neural Information Processing Systems 2012 (NIPS)*. (Lake Tahoe, USA). Ed. by P. L. Bartlett, F. C. N. Pereira, C. J. C. Burges, L. Bottou, K. Q. Weinberger. Dec. 2012, pp. 1106–1114. URL: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.

[159] V. A. Krylov, J. D. Nelson. “Stochastic Extraction of Elongated Curvilinear Structures With Applications”. In: *IEEE Transactions on Image Processing* 23 (12 2014), pp. 5360–5373. DOI: 10.1109/TIP.2014.2363612.

[160] V. Kumar, A. K. Abbas, J. C. Aster, eds. *Robbins Basic Pathology*. 9th ed. Philadelphia: Elsevier, 2013.

[161] V. Kumar, A. K. Abbas, N. Fausto, eds. *Robbins & Cotran: Patologia - Bases Patológicas das Doenças*. 7th ed. Rio de Janeiro: Elsevier, 2004.

[162] L. I. Kuncheva. “A theoretical study on six classifier fusion strategies”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24 (2 2002), pp. 281–286. DOI: 10.1109/34.982906.

[163] B. B. Lahiri, S. Bagavathiappan, T. Jayakumar, J. Philip. “Medical applications of infrared thermography: A review”. In: *Infrared Physics & Technology* 55.4 (2012), pp. 221–235.

[164] S. Lakhani, E. I.O., S. Schnitt, P. Tan, M. van de Vijver. *WHO classification of tumours of the breast*. 4th ed. Lyon: WHO Press, 2012. ISBN: 978-9-2832-2433-4.

[165] G. Lalli, N. Manikandaprabu, D. Kalamani, C. N. Marimuthu. “A development of knowledge-based inferences system for detection of breast cancer on thermogram images”. In: *Proceedings of the 2014 International Conference on Computer Communication and Informatics (ICCCI)*. (Coimbatore, India). Vol. 1. IEEE, Jan. 2014, pp. 1–6. ISBN: 978-1-4799-2353-3. DOI: 10.1109/ICCCI.2014.6921743.

[166] N. Lanisa, N. S. Cheok, L. K. Wee. “Color morphology and segmentation of the breast thermography image”. In: *Proceedings of the 2014 IEEE Conference on Biomedical*

Engineering and Sciences (IECBES). (Kuala Lumpur, Malaysia). IEEE, Dec. 2014, pp. 772–775. DOI: 10.1109/IECBES.2014.7047614.

[167] Y. LeCun, Y. Bengio, G. Hinton. "Deep learning". In: *Nature* 521 (7553 2015), pp. 436–444. DOI: 10.1038/nature14539.

[168] Y. LeCun, K. Kavukcuoglu, C. Farabet. "Convolutional networks and applications in vision". In: *Proceedings of 2010 IEEE International Symposium on Circuits and Systems (ISCAS)*. (Paris, France). Springer-Verlag, June 2010, pp. 253–256. DOI: 10.1109/ISCAS.2010.5537907.

[169] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner. "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86 (11 1998), pp. 2278–2324. DOI: 10.1109/5.726791.

[170] X. Li, K. N. Plataniotis. "A Complete Color Normalization Approach to Histopathology Images Using Color Cues Computed From Saturation-Weighted Statistics". In: *IEEE Transactions on Biomedical Engineering* 62 (7 2015), pp. 1862–1873. DOI: 10.1109/TBME.2015.2405791.

[171] X. Liu, J. Tang. "Mass Classification in Mammograms Using Selected Geometry and Texture Features, and a New SVM-Based Feature Selection Method". In: *IEEE Systems Journal* 8 (3 2014), pp. 910–920. DOI: 10.1109/JSYST.2013.2286539.

[172] X. Liu, Z. Zeng. "A new automatic mass detection method for breast cancer with false positive reduction". In: *Neurocomputing* 152 (25 2015), pp. 388–402. DOI: 10.1016/j.neucom.2014.10.040.

[173] R. Llobet, M. Pollán, J. Antón, J. Miranda-García, M. Casals, I. Martínez, F. Ruiz-Perales, B. Pérez-Gómez, D. Salas-Trejo, J.-C. Pérez-Cortés. "Semi-automated and fully automated mammographic density measurement and breast cancer risk prediction". In: *Computer Methods and Programs in Biomedicine* 116 (2 2014), pp. 105–115. DOI: 10.1016/j.cmpb.2014.01.021.

[174] N. Loménie, D. Racoceanu. "Point set morphological filtering and semantic spatial configuration modeling: Application to microscopic image and bio-structure analysis". In: *Pattern Recognition* 45 (8 2012), pp. 2894–2911. DOI: 10.1016/j.patcog.2012.01.021.

[175] C. Loukas, S. Kostopoulos, A. Tanoglidi, D. Glotsos, C. Sfikas, D. Cavouras. "Breast cancer characterization based on image classification of tissue sections visualized under low magnification". In: *Computational and Mathematical Methods in Medicine* 2013 (2013). DOI: 10.1155/2013/829461.

[176] D. G. Lowe. "Object recognition from local scale-invariant features". In: *Proceedings of the Seventh IEEE International Conference on Computer Vision (ICCV)*. (Kerkyra, Greece). Vol. 2. IEEE, Sept. 1999, pp. 1150–1157. ISBN: 0-7695-0164-8. DOI: 10.1109/ICCV.1999.790410.

[177] C. Y. Lui, H. S. Lam. "Review of Ultrasound-guided Vacuum-assisted Breast Biopsy: Techniques and Applications". In: *Jornal of Medical Ultrasound* 18.1 (2010), pp. 1–10.

[178] P. J. Lynch, C. C. Jaffe. *File:Breast anatomy normal scheme.png, Center for Advanced Instructional Media*. 2007. URL: http://commons.wikimedia.org/wiki/File:Breast_anatomy_normal_scheme.png (visited on 06/2014).

[179] F. Ma, L. Yu, M. Bajger, M. J. Bottema. "Incorporation of fuzzy spatial relation in temporal mammogram registration". In: *Fuzzy Sets and Systems* 49 (2015). DOI: 10.1016/j.fss.2015.03.021.

[180] L. v. d. Maaten, G. Hinton. "Visualizing Data using t-SNE". In: *Journal of Machine Learning Research (JMLR)* 9 (2008), pp. 2579–2605.

[181] I. K. Maitra, S. Nag, S. K. Bandyopadhyay. "Technique for preprocessing of digital mammogram". In: *Computer Methods and Programs in Biomedicine* 107 (2 2012), pp. 175–188. DOI: 10.1016/j.cmpb.2011.05.007.

[182] K. Marias, C. Behrenbruch, S. Parbhoo, A. Seifalian, M. Brady. "A Registration Framework for the Comparison of Mammogram Sequences". In: *IEEE Transactions on Medical Imaging* 26.4 (2005), pp. 782–790. DOI: 10.1109/TMI.2005.848374.

[183] O. Maron, T. Lozano-Pérez. "A Framework for Multiple-instance Learning". In: *Proceedings of the 1997 Conference on Advances in Neural Information Processing Systems 10*. (Vancouver, Canada). MIT Press, Dec. 1997, pp. 570–576.

[184] J. G. Martins, L. E. Oliveira, S. Nisgoski, R. Sabourin. "A database for automatic classification of forest species". In: *Machine Vision and Applications* 24 (3 2013), pp. 567–578. DOI: 10.1007/s00138-012-0417-5.

[185] J. G. Martins, L. E. S. Oliveira, A. Britto Jr, R. Sabourin. "Forest Species Recognition based on Dynamic Classifier Selection and Dissimilarity Feature Vector Representation". In: *Machine Vision and Applications* 26 (2 2015), pp. 279–293. DOI: 10.1007/s00138-015-0659-0.

[186] K. J. McLoughlin, P. J. Bones, N. Karssemeijer. "Noise equalization for detection of microcalcification clusters in direct digital mammogram images". In: *IEEE Transactions on Medical Imaging* 23 (3 2004), pp. 313–320. DOI: 10.1109/TMI.2004.824240.

[187] A. J. Mendez, P. G. Tahoces, M. J. Lado, M. Souto, J. J. Vidal. "Computer-aided Diagnosis: Automatic Detection of Malignant Masses in Digitized Mammograms". In: *Med Phys.* 25 (1998), pp. 957–964.

[188] A. L. Mescher. *Junqueira's basic histology: text and atlas*. New York: McGraw-Hill Lange, 2013. ISBN: 978-0-07-178033-9.

[189] N. A. Miller, J.-A. W. Chapman, J. Qian, W. A. Christens-Barry, Y. Fu, Y. Yuan, H. L. A. Lickley, D. E. Axelrod. "Heterogeneity between ducts of the same nuclear grade involved by duct carcinoma in situ (DCIS) of the breast". In: *Cancer Informatics* 9 (2010), pp. 209–216. DOI: 10.4137/CIN.S5505. URL: http://www.la-press.com/heterogeneity-between-ducts-of-the-same-nuclear-grade-involved-by-article-a2250.

[190] Ministério da Saúde (MS), Secretaria de Atenção à Saúde (SAS), Departamento de Regulação, Avaliação e Controle (DRAC), Coordenação Geral de Sistemas de Informação (CGSI). *Manual de Bases Técnicas da Oncologia – SIA/SUS: Sistema de Informações Ambulatoriais*. Tech. rep. Brasília: Ministério da Saúde (MS) et al., 2013. URL: http://www1.inca.gov.br/inca/Arquivos/manual_oncologia_15_edicao_06_09_2013.pdf (visited on 06/2014).

[191] F. Moinfar. *Essentials of diagnostic breast pathology: a practical approach*. 1st ed. Heidelberg: Springer-Verlag Berlin Heidelberg, 2007. ISBN: 978-3-540-45117-4.

[192] M. R. K. Mookiah, U. R. Acharya, E. Y. K. Ng. "Data mining technique for breast cancer detection in thermograms using hybrid feature extraction strategy". In: *Quantitative InfraRed Thermography Journal* 9 (2 2012), pp. 151–165. DOI: 10.1080/17686733.2012.738788.

[193] J.-J. Mordang, N. Karssemeijer. "Vessel segmentation in screening mammograms". In: *Proceedings of SPIE Medical Imaging 2015: Computer-Aided Diagnosis*. (Orlando, USA). Ed. by L. M. Hadjiiski, G. D. Tourassi. Vol. 9414. Society of Photo-Optical Instrumentation Engineers (SPIE), Mar. 2015. ISBN: 9781628415049. DOI: 10.1117/12.2081804.

[194] N. H. Motlagh, M. Jannesary, H. Aboulkheyr, P. Khosravi, O. Elemento, M. Totonchi, I. Hajirasouliha. "Breast Cancer Histopathological Image Classification: A Deep Learning Approach". In: *bioRxiv* (2018). DOI: 10.1101/242818.

[195] A.-A. Nahid, M. A. Mehrabi, K. Yinan. "Histopathological Breast-Cancer Image Classification by Deep Neural Network Techniques Guided by Local Clustering". In: *BioMed Research International* (2018), pp. 1–24. DOI: 10.1155/2738.

[196] A.-A. Nahid, K. Yinan. "Histopathological Breast-Image Classification Using Local and Frequency Domains by Convolutional Neural Network". In: *Information* 19 (9 2018), pp. 1–26. DOI: 10.3390/info9010019.

[197] R. Nakayama, Y. Uchiyama, K. Yamamoto, R. Watanabe, K. Namba. "Computer-aided diagnosis scheme using a filter bank for detection of microcalcification clusters in mammograms". In: *IEEE Transactions on Biomedical Engineering* 53 (2 2006), pp. 273–283. DOI: 10.1109/TBME.2005.862536.

[198] National Cancer Institute (NCI). *What You Need to Know About ™Breast Cancer*. 2012. URL: http://www.cancer.gov/cancertopics/wyntk/breast/page3 (visited on 06/2014).

[199] National Cancer Institute (NCI). *Male Breast Cancer Treatment (PDQ ®)*. 2014. URL: http://www.cancer.gov/cancertopics/pdq/treatment/malebreast/patient (visited on 06/2014).

[200] National Cancer Institute (NCI). *Mammograms Fact Sheet*. 2014. URL: http://www.cancer.gov/cancertopics/factsheet/detection/mammograms (visited on 06/2014).

[201] National Cancer Institute (NCI). *SEER Cancer Statistics Factsheets: Breast Cancer*. 2014. URL: http://seer.cancer.gov/statfacts/html/breast.html (visited on 06/2014).

[202] N. Nayak, H. Chang, A. Borowsky, P. Spellman, B. Parvin. "Classification of tumor histopathology via sparse feature learning". In: *Proceedings of the International Symposium on Biomedical Imaging (ISBI): From Nano to Macro*. (San Francisco, USA). IEEE, Apr. 2013. DOI: 10.1109/ISBI.2013.6556499. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3850768/.

[203] *NCI Visuals Online*. 2001. URL: https://visualsonline.cancer.gov (visited on 07/2014).

[204] N. I. for Health, C. E. (NICE). *Image-guided vacuum-assisted excision biopsy of benign breast lesions*. 2006. URL: http://www.nice.org.uk/guidance/ipg156 (visited on 06/2014).

[205] NVIDIA Corporation. *NVIDIA Tesla Product Literature*. 2015. URL: http://www.nvidia.com/object/tesla_product_literature.html (visited on 12/2015).

[206] E. O'Flynn, A. Wilson, M. Michell. "Image-guided Breast Biopsy: state-of-the-art". In: *Clinical Radiology* 65.4 (2010), pp. 259–270.

[207] T. Ojala, M. Pietikäinen, D. Harwood. "A comparative study of texture measures with classification based on featured distribution". In: *Pattern Recognition* 1 (29 1996), pp. 51–59.

[208] T. Ojala, M. Pietikäinen, T. Mäenpää. "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns". In: *IEEE Transactions On Pattern Analysis And Machine Intelligence* 24 (7 2002), pp. 971–987. DOI: 10.1109/TPAMI.2002.1017623.

[209] V. Ojansivu, J. Heikkilä. "Blur Insensitive Texture Classification Using Local Phase Quantization". In: *Proceedings of the 3rd International Conference on Image and Signal Processing (ICISP)*. (Cherbourg-Octeville, France). Ed. by A. Elmoataz, O. Lezoray, F. Nouboud, D. Mammass. Vol. 5099. Springer Berlin Heidelberg, July 2008, pp. 236–243. DOI: 10.1007/978-3-540-69905-7_27.

[210] M. Oquab, L. Bottou, I. Laptev, J. Sivic. "Learning and Transferring Mid-level Image Representations Using Convolutional Neural Networks". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. (Columbus, USA). IEEE, June 2014, pp. 1717–1724. DOI: 10.1109/CVPR.2014.222.

[211] N. Otsu. "A threshold selection method from gray-level histograms". In: *IEEE Transactions on Systems, Man and Cybernetics* 9 (1 1979), pp. 62–66. DOI: 10.1109/TSMC.1979.4310076.

[212] P. Padmavathy. "A comparative study on breast cancer prediction using RBF and MLP". In: *International Journal of Scientific & Engineering Research* 2.1 (2011), pp. 1–5.

[213] J. Päivärinta, E. Rahtu, J. Heikkilä. "Volume local phase quantization for blur-insensitive dynamic texture classification". In: *Proceedings of the 17th Scandinavian Conference on Image Analysis (SCIA)*. (Ystad, Sweden). Ed. by A. Heyden, F. Kahl. Springer Berlin Heidelberg, May 2011, pp. 360–369. ISBN: 978-3-642-21227-7. DOI: 10.1007/978-3-642-21227-7_34.

[214] S. K. Pal, S. K. Lau, L. Kruper, U. Nwoye, C. Garberoglio, R. K. Gupta, B. Paz, V. Lalit, E. Guzman, A. Artinyan, G. Somlo. "Papillary carcinoma of the breast: an overview". In: *Breast Cancer Research and Treatment* 122 (3 2010), pp. 637–645. DOI: 10.1007/s10549-010-0961-5.

[215] G. Palma, I. Bloch, S. Muller. "Detection of masses and architectural distortions in digital breast tomosynthesis images using fuzzy and a contrario approaches". In: *Pattern Recognition* 47 (7 2014), pp. 2467–2480. DOI: 10.1016/j.patcog.2014.01.009.

[216] K. Panetta, Y. Zhou, S. Agaian, H. Jia. "Nonlinear Unsharp Masking for Mammogram Enhancement". In: *IEEE Transactions on Information Technology in Biomedicine* 15 (6 2011), pp. 918–928. DOI: 10.1109/TITB.2011.2164259.

[217] D. Pathak, E. Shelhamer, J. Long, T. Darrell. "Fully Convolutional Multi-Class Multiple Instance Learning". In: *3rd International Conference on Learning Representations (ICLR2015)*. (San Diego, USA). May 2015.

[218] D. Pathak, E. Shelhamer, J. Long, T. Darrell. "Fully Convolutional Multi-Class Multiple Instance Learning". In: *CoRR* abs/1412.7144 (2015).

[219] P. L. d. P. Paula, L. E. S. Oliveira, A. d. S. Britto Jr., R. Sabourin. "Forest Species Recognition Using Color-Based Features". In: *Proceedings of the 22nd International Conference on Pattern Recognition (ICPR)*. (Istanbul, Turkey). IEEE, Aug. 2010, pp. 4178–4181. DOI: 10.1109/ICPR.2010.1015.

[220] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[221] G. P. Pena, J. d. S. Andrade-Filho. "How does a pathologist make a diagnosis?" In: *Archives of Pathology & Laboratory Medicine* 133 (1 2009), pp. 124–132. DOI: 10.1043/1543-2165-133.1.124. URL: http://www.archivesofpathology.org/doi/abs/10.1043/1543-2165-133.1.124.

[222] C. A. Peña-Reyes, M. Sipper. "A fuzzy-genetic approach to breast cancer diagnosis". In: *Artificial Intelligence in Medicine* 17.2 (1999), pp. 131–155.

[223] D. C. Pereira, R. P. Ramos, M. Z. d. Nascimento. "Segmentation and detection of breast cancer in mammograms combining wavelet analysis and genetic algorithm". In: *Computer Methods and Programs in Biomedicine* 114 (1 2014), pp. 88–101. DOI: 10.1016/j.cmpb.2014.01.014.

[224] S. Petushi, F. U. Garcia, M. M. Haber, C. Katsinis, A. Tozeren. “Large-scale computations on histology images reveal grade-differentiating parameters for breast cancer”. In: *BMC Medical Imaging* 6 (14 2006). DOI: 10.1186/1471-2342-6-14.

[225] K. P. Polat, S. Güneş. “Breast cancer diagnosis using least square support vector machine”. In: *Digital Signal Processing* 17.4 (2007), pp. 694–701.

[226] K. Polyak. “Breast cancer: origins and evolution”. In: *The Journal of Clinical Investigation* 117 (11 2007), pp. 3155–3163. DOI: 10.1172/JCI3329.

[227] A. Prasoon, K. Petersen, C. Igel, F. Lauze, E. Dam, M. Nielsen. “Deep Feature Learning for Knee Cartilage Segmentation Using a Triplanar Convolutional Neural Network”. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*. Ed. by K. Mori, I. Sakuma, Y. Sato, C. Barillot, N. Navab. Vol. 8150. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2013, pp. 246–253. ISBN: 978-3-642-40762-8. DOI: 10.1007/978-3-642-40763-5_31.

[228] G. Quellec, G. Cazuguel, B. Cochener, M. Lamard. “Multiple-instance learning for medical image and video analysis”. In: *IEEE Reviews in Biomedical Engineering (RBME)* 10 (2017), pp. 213–234. DOI: 10.1109/RBME.2017.2651164.

[229] J. R. Quinlan. “Induction of decision trees”. In: *Machine Learning* 1 (1986), pp. 81–106.

[230] A. C. of Radiology (ACR). *About ACR BI-RADS® Atlas*. 2014. URL: http://www.acr.org/Quality-Safety/Resources/BIRADS/About-BIRADS (visited on 06/2014).

[231] N. Rahimi. “The ubiquitin-proteasome system meets angiogenesis”. In: *Molecular Cancer Therapeutics* 11 (3 2012), pp. 538–548. DOI: 10.1158/1535-7163.MCT-11-0555. URL: http://mct.aacrjournals.org/content/11/3/538.full.

[232] R. M. Rangayyan, R. J. Ferrari, A. F. Frère. “Analysis of bilateral asymmetry in mammograms using directional, morphological, and density features”. In: *Journal of Electronic Imaging* 16 (1 2007). DOI: 10.1117/1.2712461.

[233] A. S. Razavian, H. Azizpour, J. Sullivan, S. Carlsson. “CNN Features Off-the-Shelf: An Astounding Baseline for Recognition”. In: *2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. (Columbus, USA). June 2014, pp. 512–519. DOI: 10.1109/CVPRW.2014.131.

[234] F. J. P. Richard, P. R. Bakić, A. D. A. Maidment. “Mammogram Registration: A Phantom-Based Evaluation of Compressed Breast Thickness Variation Effects”. In: *IEEE Transactions on Medical Imaging* 25 (2 2006), pp. 188–197. DOI: 10.1109/TMI.2005.862204.

[235] C. Rose, D. Turi, A. Williams, K. Wolstencroft, C. Taylor. “Web services for the DDSM and digital mammography research”. In: *Proceedings of the 8th International Workshop on Digital Mammography (IWDM)*. (Manchester, UK). Ed. by S. M. Astley, M. Brady, C. Rose, R. Zwiggelaar. Vol. 4046. Springer-Verlag Berlin Heidelberg, June 2006, pp. 376–383. ISBN: 978-3-540-35627-1. DOI: 10.1007/11783237_51.

[236] M. H. Ross, W. Pawlina. *Histology: a text and atlas: with correlated cell and molecular biology*. 6th ed. Philadelphia: Lippincott Williams & Wilkins, 2011.

[237] E. Rosten, R. Porter, T. Drummond. "Faster and better: A machine learning approach to corner detection". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32 (1 2010), pp. 105–119. DOI: 10.1109/TPAMI.2008.275.

[238] R. Rubin, D. S. Strayer, E. Rubin, eds. *Rubin's Pathology Clinicopathologic Foundations of Medicine*. 6th ed. Philadelphia: Lippincott Williams & Wilkins, 2012.

[239] E. Rublee, V. Rabaud, K. Konolige, G. R. Bradski. "ORB: An efficient alternative to SIFT or SURF". In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. (Barcelona, Spain). IEEE, Nov. 2011, pp. 2564–2571. ISBN: 978-1-4577-1101-5. DOI: 10.1109/ICCV.2011.6126544.

[240] S. Şahan, K. P. Polat, H. Kodaz, S. Güneş. "A new hybrid method based on fuzzy-artificial immune system and k-nn algorithm for breast cancer diagnosis". In: *Computers in Biology and Medicine* 37.3 (2007), pp. 415–423.

[241] A. A. Samah, M. F. A. Fauzi, S. Mansor. "Classification of Benign and Malignant Tumors in Histopathology Images". In: *Proceedings of the 2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*. (Kuching, Malaysia). IEEE, Sept. 2017, pp. 102–106. DOI: 10.1109/ICSIPA.2017.8120587.

[242] M. Sameti, R. K. Ward, J. Morgan-Parkes, B. Palcic. "Image Feature Extraction in the Last Screening Mammograms Prior to Detection of Breast Cancer". In: *IEEE Transactions on Selected Topics in Signal Processing* 3 (1 2009), pp. 46–52. DOI: 10.1109/JSTSP.2008.2011163.

[243] C. Sammut, G. I. Webb, eds. *Encyclopedia of Machine Learning*. New York: Springer, 2011.

[244] M. E. Sanders, J. F. Simpson. *Breast pathology*. Ed. by D. E. Elder. Vol. 6. New York: DemosMedical, 2014. ISBN: 978-1-936287-68-0.

[245] K. I. Satoto, O. D. Nurhayati, R. Isnanto. "Pattern Recognition to Detect Breast Cancer Thermogram Images Based on Fuzzy Inference System Method". In: *Internatinal Journal of Computer Science and Technology* 2.3 (2011), pp. 484–487.

[246] Sociedade Brasileira de Mastologia (SBM). *Tipos e Técnicas de Mamografia*. 2011. URL: http://www.sbmastologia.com.br/cancer-de-mama/rastreamento-diagnostico-cancer-de-mama/tipos-e-tecnicas-de-mamografia-14.htm (visited on 06/2014).

[247] Sociedade Brasileira de Mastologia (SBM). *Câncer de Mama*. 2014. URL: http://www.sbmastologia.com.br/index/index.php/entendendo (visited on 06/2014).

[248] D. Scherer, A. Müller, S. Behnke. "Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition". In: *Proceedings of the 20th International Conference*

on Artificial Neural Networks: Part III. (Thessaloniki, Greece). Ed. by K. Diamantaras, W. D. Duch, L. S. Iliadis. Springer-Verlag, Sept. 2010, pp. 92–101.

[249] G. van Schie, R. Mann, M. Imhof-Tas, N. Karssemeijer. "Generating Synthetic Mammograms From Reconstructed Tomosynthesis Volumes". In: *IEEE Transactions on Medical Imaging* 32 (12 2013), pp. 2322–2331. DOI: 10.1109/TMI.2013.2281738.

[250] S. E. Selvan, C. C. Xavier, N. Karssemeijer, J. Sequeira, R. A. Cherian, B. Y. Dhala. "Parameter Estimation in Stochastic Mammogram Model by Heuristic Optimization Techniques". In: *IEEE Transactions on Information Technology in Biomedicine* 10 (4 2006), pp. 685–695. DOI: 10.1109/TITB.2006.874197.

[251] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, Y. LeCun. "OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks". In: *Proceedings of the International Conference on Learning Representations (ICLR 2014)*. (Banff, Canada). CBLS, Apr. 2014. ISBN: 78-952-10-5699-4. URL: http://arxiv.org/abs/1312.6229.

[252] R. Setiono. "Generating concise and accurate classification rules for breast cancer diagnosis". In: *Artificial Intelligence in Medicine* 18.3 (2000), pp. 205–2019.

[253] J. Shi, Y. Li, J. Zhu, H. Sun, Y. Cai. "Joint sparse coding based spatial pyramid matching for classification of color medical image". In: *Computerized Medical Imaging and Graphics* 41 (2015), pp. 61–66. DOI: doi:10.1016/j.compmedimag.2014.06.002.

[254] K. Simonyan, A. Zisserman. "Very deep convolutional networks for large-scale image recognition". In: *Proceedings of the 3rd International Conference on Learning Representations (ICLR2015)*. (San Diego, USA). CBLS, May 2015. URL: http://arxiv.org/abs/1409.1556.

[255] Y. Song, J. J. Zou, H. Chang, W. Cai. "Adapting Fisher Vectors for Histopathology Image Classification". In: *Proceedings of the 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI)*. (Melbourne, Australia). IEEE, Apr. 2017, pp. 600–603. DOI: 10.1109/ISBI.2017.7950592.

[256] F. Spanhol, L. S. Oliveira, C. Petitjean, L. Heutte. "A Dataset for Breast Cancer Histopathological Image Classification". In: *IEEE Transactions on Biomedical Engineering (TBME)* 63 (7 2016), pp. 1455–1462. DOI: 10.1109/TBME.2015.2496264.

[257] F. Spanhol, L. S. Oliveira, C. Petitjean, L. Heutte. "Breast cancer histopathological image classification using Convolutional Neural Networks". In: *2016 International Joint Conference on Neural Networks (IJCNN)*. (Vancouver, Canada). IEEE, July 2016, pp. 2560–2567. DOI: 10.1109/IJCNN.2016.7727519.

[258] F. Spanhol, P. Cavalin, L. S. Oliveira, C. Petitjean, L. Heutte. "Deep Features for Breast Cancer Histopathological Image Classification". In: *Proceedings of the 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. (Banff, Canada). IEEE, Oct. 2017, pp. 1868–1873. DOI: 10.1109/SMC.2017.8122889.

[259] F. A. Spanhol, F. Negretti, C. E. Pokes, L. S. Oliveira, M. C. Cavéquia. "Software para Auxiliar no Diagnóstico Histopatológico de Neoplasias de Mama: Resultados Prelim-

inares". Portuguese. In: *XVIII Congresso Brasileiro de Mastologia (CBM)*. (Curitiba, Brazil). SBM, June 2015.

[260] B. Stenkvist, S. Westman-Naeser, J. Holmquist, B. Nordin, E. Bengtsson, J. Vegelius, O. Eriksson, C. H. Fox. "Computerized Nuclear Morphometry as an Objective Method for Characterizing Human Cancer Cell Populations". In: *Cancer Research* 38.12 (1978), pp. 4688–4697.

[261] J. Suckling, J. Parker, D. R. Dance, S. Astley, I. Hutt, C. R. M. Boggis, I. Ricketts, E. Stamatakis, N. Cerneaz, S.-L. Kok, P. Taylor, D. Betal, J. Savage. "The Mammographic Image Analysis Society Digital Mammogram Database". In: *Exerpta Medica* 1069 (1994), pp. 375–378.

[262] P. J. Sudharshan, C. Petitjean, F. Spanhol, L. Oliveira, L. Heutte, P. Honeine. "Multiple Instance Learning for Histopathological Breast Cancer Image Classification". In: *Expert Systems With Applications (ESWA)* (2018). In press.

[263] M. Sun, T. X. Han, M.-C. Liu, A. Khodayari-Rostamabad. "Multiple Instance Learning Convolutional Neural Networks for object recognition". In: *23rd International Conference on Pattern Recognition (ICPR)*. (Cancun, Mexico). Vol. 1. IEEE, Dec. 2016, pp. 3270–3275. DOI: 10.1109/ICPR.2016.7900139.

[264] M. Sundaram, K. Ramar, N. Arumugam, G. Prabin. "Histogram Modified Local Contrast Enhancement for mammogram images". In: *Applied Soft Computing* 11 (8 2011), pp. 5809–5816. DOI: 10.1016/j.asoc.2011.05.003.

[265] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich. "Going deeper with convolutions". In: *Proceedings of the 28rd IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. (Boston, USA). IEEE, June 2015, pp. 1–9. DOI: 10.1109/CVPR.2015.7298594.

[266] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna. "Rethinking the Inception Architecture for Computer Vision". In: *Proceedings of the 29th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. (Las Vegas, USA). IEEE, June 2016, pp. 2818–2826. DOI: 10.1109/CVPR.2016.308.

[267] L. Tabár, T. Tot, P. B. Dean. *breast cancer: the art and science of early detection with mammography – perception, interpretation, histopathologic correlation*. New York: Tyeme Verlag, 2004.

[268] A. G. Taghian, B. L. Smith, J. K. Erban, eds. *Breast cancer: a multidisciplinary approach to diagnosis and management*. New York: DemosMedical, 2010. ISBN: 978–1–933864–44–0.

[269] S.-C. Tai, Z.-S. Chen, W.-T. Tsai. "An Automatic Mass Detection System in Mammograms Based on Complex Texture Features". In: *IEEE Journal of Biomedical and Health Informatics* 18 (2 2014), pp. 618–627. DOI: 10.1109/JBHI.2013.2279097.

[270] A. Tashk, M. S. Helfroush, H. Danyali, M. Akbarzadeh1. "A Novel CAD System for Mitosis detection Using Histopathology Slide Images". In: *Journal of Medical Signals and Sensors* 4 (2 2014), pp. 139–149. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3994718/.

[271] The Human Protein Atlas. *The Human Protein Atlas - Dictionary - Normal tissue - Breast*. 2008. URL: http://www.proteinatlas.org/learn/dictionary/normal/breast+1 (visited on 08/2014).

[272] *The mini-MIAS database of mammograms*. 1994. URL: http://peipa.essex.ac.uk/info/mias.html (visited on 02/2015).

[273] S. Theodoridis, K. Koutroumbas. *Pattern Recognition*. 4th ed. Academic Press, 2009. ISBN: 978-1-59749-272-0.

[274] G. Tse, P. H. Tan, F. Schmitt. *Fine needle aspiration cytology of the breast: atlas of cyto-histologic correlates*. 1st ed. Heidelberg: Springer-Verlag Berlin Heidelberg, 2013. ISBN: 978-3-642-34999-7.

[275] N. Vállez, G. Bueno, O. Déniz, J. Dorado, J. A. Seoane, A. Pazos, C. Pastor. "Breast density classification to reduce false positives in CADe systems". In: *Computer Methods and Programs in Biomedicine* 113 (2 2014), pp. 569–584. DOI: 10.1016/j.cmpb.2013.10.004.

[276] R. Venkatesan, P. S. Chandakkar, B. Li. "Simpler Non-Parametric Methods Provide as Good or Better Results to Multiple-Instance Learning". In: *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*. (Santiago, Chile). IEEE, Dec. 2015, pp. 2605–2613. DOI: 10.1109/ICCV.2015.299.

[277] M. Veta, J. P. W. Pluim, P. J. v. Diest, M. A. Viergever. "Breast cancer histopathology image analysis: a review". In: *IEEE Transactions On Biomedical Engineering* 61 (5 2014), pp. 1400–1411. DOI: 10.1109/TBME.2014.2303852.

[278] T. H. Vu, H. S. Mousavi, V. Monga, U. A. Rao, G. Rao. "DFDL: Discriminative feature-oriented dictionary learning for histopathological image classification". In: *Proceedings of the IEEE 12th International Symposium on Biomedical Imaging (ISBI)*. (New York, USA). IEEE, Apr. 2015, pp. 990–994. DOI: 10.1109/ISBI.2015.7164037.

[279] S. van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, T. Yu, the scikit-image contributors. "scikit-image: image processing in Python". In: *PeerJ* 2 (June 2014), e453. ISSN: 2167-8359. DOI: 10.7717/peerj.453. URL: http://dx.doi.org/10.7717/peerj.453.

[280] J. Wang, J.-D. Zucker. "Solving the Multiple-Instance Problem: A Lazy Learning Approach". In: *Proceedings of the Seventeenth International Conference on Machine Learning (ICML)*. (Stanford, USA). Morgan Kaufmann, July 2000, pp. 1119–1126.

[281] X. Wang, Y. Yan, P. Tang, X. Bai, W. Liu. "Revisiting multiple instance neural networks". In: *Pattern Recognition* 74 (2018), pp. 15–24. DOI: 10.1016/j.patcog.2017.08.026.

[282] B. W. Wei, Z. Han, X. He, Y. Yin. "Deep learning model based breast cancer histopathological image classification". In: *Proceedings of the 2017 IEEE 2nd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*. (Chengdu, China). IEEE, Apr. 2017, pp. 348–353. DOI: 10.1109/ICCCBDA.2017.7951937.

[283] L. Wei, Y. Yang, M. N. Wernick, R. M. Nishikawa. "Learning of Perceptual Similarity From Expert Readers for Mammogram Retrieval". In: *IEEE Journal of Selected Topics in Signal Processing* 3 (1 2009), pp. 53–61. DOI: 10.1109/JSTSP.2008.2011159.

[284] R. S. Weinstein, A. R. Graham, L. C. Richter, G. P. Barker, E. A. Krupinski, A. M. Lopez, K. A. Erpsa, A. K. Bhattacharyya, Y. Yagi, J. R. Gilbertson. "Overview of telepathology, virtual microscopy, and whole slide imaging: prospects for the future". In: *Human Pathology* 40 (8 2009), pp. 1057–1069. DOI: 10.1016/j.humpath.2009.04.006.

[285] *What is a radiologist?* 2014. URL: http://www.acr.org/Quality-Safety/Radiology-Safety/Patient-Resources/About-Radiology (visited on 08/2014).

[286] World Cancer Research Found International (WCRF). *Worldwide Cancer Statistics*. Jan. 2013. URL: http://www.wcrf.org/cancer_statistics/world_cancer_statistics.php (visited on 06/2014).

[287] H. Wu, L. He. "Combining visual and textual features for medical image modality classification with ℓ_p–norm multiple kernel learning". In: *Neurocomputing* 147 (11 2015), pp. 387–394. DOI: 10.1016/j.neucom.2014.06.046.

[288] M. X. Xie, N. Jean, M. Burke, D. Lobell, S. Ermon. "Transfer Learning from Deep Features for Remote Sensing and Poverty Mapping". In: *Proceedings of Thirtieth AAAI Conference on Artificial Intelligence (AAAAI-16)*. (Phoenix, USA). AAAI, Feb. 2016. URL: http://arxiv.org/abs/1510.00098.

[289] F. Xing, Y. Xie, L. Yang. "An automatic learning-based framework for robust nucleus segmentation". In: *IEEE Transactions on Biomedical Imaging* (2015). DOI: 10.1109/TMI.2015.2481436.

[290] Y. Xu, J.-Y. Zhu, I Eric, C. Chang, M. Lai, Z. Tu. "Weakly Supervised Histopathology Cancer Image Segmentation and Classification". In: *Medical Image Analysis* 18.3 (2014), pp. 591–604. DOI: 10.1016/j.media.2014.01.010.

[291] J. Yosinski, J. Clune, Y. Bengio, H. Lipson. "How transferable are features in deep neural networks?" In: *Proceedings of 26th Annual Conference on Neural Information Processing Systems 2014 (NIPS)*. (Montréal, Canada). Ed. by Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, K. Weinberger. Dec. 2014, pp. 3320–3328. URL: http://papers.nips.cc/paper/5347-how-transferable-are-features-in-deep-neural-networks.pdf.

[292] S. Yu, L. Guan. "A CAD system for the automatic detection of clustered microcalcifications in digitized mammogram films". In: *IEEE Transactions on Medical Imaging* 19 (2 2000), pp. 115–126. DOI: 10.1109/42.836371.

[293] B. Zhang. "Breast cancer diagnosis from biopsy images by serial fusion of random subspace ensembles". In: *Proceedings of the International Conference on Biomedical Engineering and Informatics (BMEI)*. (Shanghai, China). Vol. 1. IEEE, Oct. 2011, pp. 180–186. DOI: 10.1109/BMEI.2011.6098229.

[294] Q. Zhang, S. A. Goldman. “EM-DD: An Improved Multiple-Instance Learning Technique”. In: *Proceedings of the 2001 Conference on Advances in Neural Information Processing Systems 14*. (Vancouver, Canada). MIT Press, Dec. 2001, pp. 1073–1080.

[295] Y. Zhang, B. Zhang, F. Coenen, W. Lu. “Breast cancer diagnosis from biopsy images with highly reliable random subspace classifier ensembles”. In: *Machine Vision and Applications* 24.7 (2013), pp. 1405–1420. DOI: 10.1007/s00138-012-0459-8.

[296] Y. Zhang, B. Zhang, F. Coenen, J. Xiau, W. Lu. “One-class kernel subspace ensemble for medical image classification”. In: *EURASIP Journal on Advances in Signal Processing* 2014.17 (2014), pp. 1–13. DOI: 10.1186/1687-6180-2014-17.

[297] G. Zhao, M. Pietikäinen. “Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29 (6 2007), pp. 915–928. DOI: 10.1109/TPAMI.2007.1110.

[298] L. Zheng, A. K. Chan. “An artificial intelligent algorithm for tumor detection in screening mammogram”. In: *IEEE Transactions on Medical Imaging* 20 (7 2001), pp. 559–567. DOI: 10.1109/42.932741.

[299] W. Zhi, H. W. F. Yueng, Z. Chen, S. M. Zandavi, Z. Lu, Y. Y. Chung. “Using Transfer Learning with Convolutional Neural Networks to Diagnose Breast Cancer from Histopathological Images”. In: *Proceedings of the 24th International Conference on Neural Information Processing (ICONIP)*. (Guangzhou, China). Springer, Cham, Nov. 2017, pp. 669–676. DOI: 10.1007/978-3-319-70093-9_71.

[300] L. Zhou, Y. Zhao, J. Yang, Q. Yu, X. Xu. “Deep multiple instance learning for automatic detection of diabetic retinopathy in retinal images”. In: *IET Image Processing* 12 (4 2018), pp. 563–571. DOI: 10.1049/iet-ipr.2017.0636.

[301] Y. Zhou, H. Chang, K. Barner, P. Spellman, B. Parvin. “Classification of Histology Sections via Multispectral Convolutional Sparse Coding”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. (Columbus, USA). June 2014, pp. 3081–3088. DOI: 10.1109/CVPR.2014.394.

[302] Z.-H. Z. Zhou. “A brief introduction to weakly supervised learning”. In: *National Science Review* 5 (2017), pp. 44–53. DOI: 10.1093/nsr/nwx106.