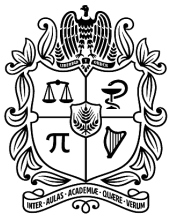


Universidad Nacional de Colombia
School of Engineering

Efficient interaction with large medical imaging databases

Germán Corredor Prada

Universidad Nacional de Colombia
School of Engineering, Electrical and Electronic Engineering Department
Bogota, Colombia
2018

Efficient interaction with large medical imaging databases

Germán Corredor Prada

In fulfillment of the requirements for the degree of:
Ph.D. in Engineering - Electrical Engineering

Advisor:
Eduardo Romero Castro, Ph.D.

Research Field:
Signal processing and artificial vision
Research Group:
Computer Imaging and Medical Applications Laboratory - CIM@LAB

Universidad Nacional de Colombia
School of Engineering, Electrical and Electronic Engineering Department
Bogota, Colombia
2018

A mis padres y mi hermano

Acknowledgements

First, I want to thank to Universidad Nacional de Colombia that has been my second home for more than 12 years. Thanks for all the experiences, opportunities, and life lessons!
I also want to express my gratitude to my thesis advisor, professor Eduardo Romero, who has trusted and supported me along this process. Thank you for your friendship, time, patience, dedication, and teachings. Everything has meaningfully contributed to my personal and professional growth. A special thanks to professor Marcela Iregui with whom I had the pleasure of working before starting my PhD program and who helped me to get in touch with professor Romero.
Thanks to the people of CIM@LAB group who were an important part of this process. I would like to give special thanks to my partners and friends Josué, Lina, Pablo, Cristian, Diana, Raúl, Luisa, David Romo, David Trujillo, Carlos, Fabio, and Fabián. I also want to thank to María Teresa for her kindness and willingness to help me every time.
I am indebted with the personnel of Departament of Pathology of Universidad Nacional de Colombia who provided advising, supplied study cases, and helped with experimentation. A special thanks to Dr. Viviana Arias who was always available and willing to work with us.
Thanks to the CCIPD and all the work team in Cleveland (USA). A special thanks to professor Anant Madabhushi for his support, academic guidance, and the opportunity to work in Cleveland; this was an invaluable personal and professional experience for me. My special gratitude to my friends Cheng, Prateek, Rakesh, Jon, Pranjal, Lin, Nate, and Xiangxue.
I also want to thank to all my friends who were present along this process, specially to Camilo, Diana García, Duván, Jaime, Borrego, Diana Gamba, Camila, Eljach, Daniel, Lina Bonilla, and Adriana. A special thanks to Marcela who was a very important part of this process; thank you for your affection, patience, and company.
To my girlfriend, Paula, thanks for your love, support, understanding, and teachings. Your comments and explanations have taught me a lot and have deeply contributed to improve my thesis. Thanks for motivating me to be a better person every day!
I am deeply grateful with my family. Thanks to my parents for their love, support, and teaching me how to be a better person and to give the best of myself. Thanks to my brother for his unconditional trust and support.
Finally, thanks to all people who, direct or indirectly, accompanied me along this process and were willing to help me when I needed it.

Abstract

Everyday, a wide quantity of hospitals and medical centers around the world are producing large amounts of imaging content to support clinical decisions, medical research, and education. With the current trend towards Evidence-based medicine, there is an increasing need of strategies that allow pathologists to properly interact with the valuable information such imaging repositories host and extract relevant content for supporting decision making. Unfortunately, current systems are very limited at providing access to content and extracting information from it because of different semantic and computational challenges. This thesis presents a whole pipeline, comprising 3 building blocks, that aims to to improve the way pathologists and systems interact. The first building block consists in an adaptable strategy oriented to ease the access and visualization of histopathology imaging content. The second block explores the extraction of relevant information from such imaging content by exploiting low- and mid-level information obtained from from morphology and architecture of cell nuclei. The third block aims to integrate high-level information from the expert in the process of identifying relevant information in the imaging content. This final block not only attempts to deal with the semantic gap but also to present an alternative to manual annotation, a time consuming and prone-to-error task. Different experiments were carried out and demonstrated that the introduced pipeline not only allows pathologist to navigate and visualize images but also to extract diagnostic and prognostic information that potentially could support clinical decisions.

Keywords: Histopathology, Digital pathology, Pathological marker

Resumen

Diariamente, gran cantidad de hospitales y centros médicos de todo el mundo producen grandes cantidades de imágenes diagnósticas para respaldar decisiones clínicas y apoyar labores de investigación y educación. Con la tendencia actual hacia la medicina basada en evidencia, existe una creciente necesidad de estrategias que permitan a los médicos patólogos interactuar adecuadamente con la información que albergan dichos repositorios de imágenes y extraer contenido relevante que pueda ser empleado para respaldar la toma de decisiones. Desafortunadamente, los sistemas actuales son muy limitados en cuanto al acceso y extracción de contenido de las imágenes debido a diferentes desafíos semánticos y computacionales. Esta tesis presenta un marco de trabajo completo para patología, el cual se compone de 3 bloques y tiene como objetivo mejorar la forma en que interactúan los patólogos y los sistemas. El primer bloque de construcción consiste en una estrategia adaptable orientada a facilitar el acceso y la visualización del contenido de imágenes histopatológicas. El segundo bloque explora la extracción de información relevante de las imágenes mediante la explotación de información de características visuales y estructurales de la morfología y la arquitectura de los núcleos celulares. El tercer bloque apunta a integrar información de alto nivel del experto en el proceso de identificación de información relevante en las imágenes. Este bloque final no solo intenta lidiar con la brecha semántica, sino que también presenta una alternativa a la anotación manual, una tarea que demanda mucho tiempo y es propensa a errores. Se llevaron a cabo diferentes experimentos que demostraron que el marco de trabajo presentado no solo permite que el patólogo navegue y visualice imágenes, sino que también extraiga información de diagnóstico y pronóstico que potencialmente podría respaldar decisiones clínicas.

Palabras claves: Histopatología, Patología digital, Marcadores patológicos

Contents

1 Introduction

Evidence-based medicine

Medicine practice has meaningfully changed during the last decades. Some years ago, unsystematic observations from clinical experience were considered as a valid way of building knowledge about patient prognosis, the value of diagnostic tests, and the efficacy of treatment [55]. Likewise, a combination of thorough traditional medical training and common sense was sufficient to allow a physician to evaluate new tests and treatments [55].
However, starting in the late 1960s, different authors published works evidencing weaknesses in medical decision making at the level of both individual patients and populations. These works highlighted problems such as lack of controlled trials, errors in clinical reasoning, gaps in evidence, high variations in how physicians practiced, large number of inappropriate procedures performed by physicians, among others [117, 172, 46, 66].
In response to these problems, in the last years, a new paradigm for medical practice has gained popularity: Evidence-base medicine (EBM). EBM is defined as "the conscientious, explicit, judicious, and reasonable use of current best evidence in making decisions about the care of individual patients"[111]. EBM states that clinical decisions should be based on evidence rather than the beliefs or intuition of experts. It also promotes to change the training of practitioners to ensure that the patient care actually delivered does meet standards[46].
Most physicians around the world are aiming for practice of EBM. Now, proper EMB practice requires accurate and reproducible diagnoses as well as the possibility of establishing prognostics and predictive factors, understanding prognostic factors as those related to the natural disease evolution and predictive factors as those related to some kind of intervention. Providing such set of diagnostic, prognostic, and predictive factors is the labor of pathology.

Pathology

Pathology is, literally translated, the *study of disease* and involves the analysis of structural and functional changes at the levels of cells, tissues, and organs affected by a disease [166]. It employs different molecular, microbiological, immunologic, and morphologic techniques to understand the signs and symptoms of patients, thereby offering a rational basis for clinical attention and treatment. Thus, pathology has been considered as a bridge between basic sciences and clinical medicine, and it provides the **scientific foundation** for all of medi-

cine [91] (See Figure **1-1**). Effective healthcare is supported by information that pathology services provide to clinicians; unfortunately, this fact is often not appreciated by the general public or even healthcare professionals [122].

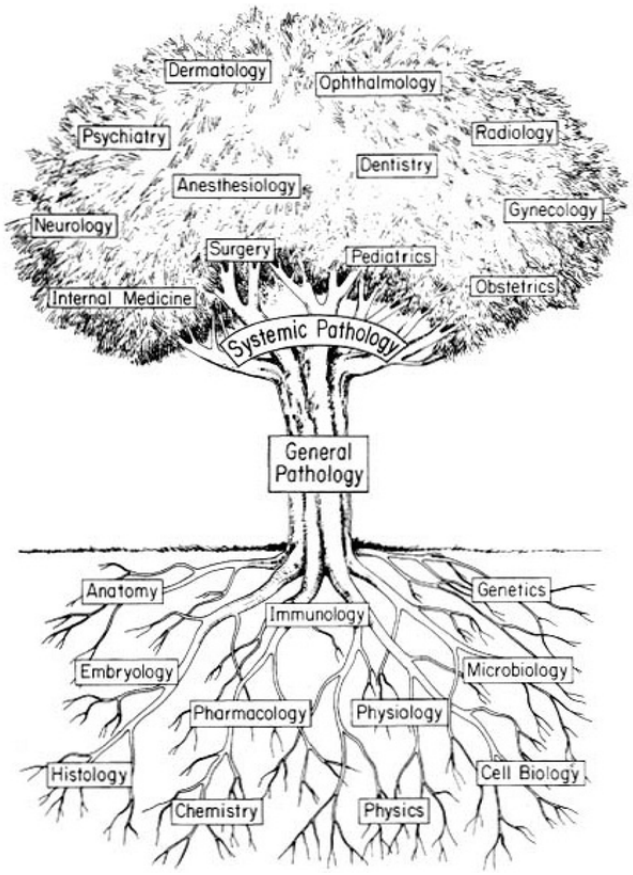


Figure 1-1: The Tree of Medicine. The trunk is General Pathology, which draws from all the basic sciences, and divides into the many branches of Special Pathology; each one of these supports a specialized field of Medicine, the crown of the tree. Source: [108].

Pathology focuses on the following four aspects of the disease process [166, 91]:

- **Etiology.** This term refers to origin of disease, including the underlying causes and modifying factors. Factors that cause disease are usually grouped into two classes: genetic and acquired; however, it is known that some diseases may be caused by a combination of inherited genetic susceptibility and various environmental triggers.
- **Pathogenesis.** It refers to the mechanisms involved in the origin of disease. It describes how etiologic factors trigger cellular and molecular changes that give rise to the specific functional and structural abnormalities that characterize a disease.

- **Morphologic changes.** They are related to structural alterations in cells or tissues. Such alterations could be either characteristic of a disease or diagnostic of an etiologic process. Traditionally, morphology has been used in diagnostic pathology to determine the nature of the disease and follow its progression.

- **Functional derangements and clinical manifestations.** This aspect refers to the functional consequences of morphologic changes. The result of genetic, biochemical, and structural changes in cells and tissues are functional abnormalities, which lead to the symptoms and signs of disease, as well as its clinical course and outcome.

Histopathology

Histopathology is a sub-discipline of pathology that comprises the study of diseased tissue. Histopathologists look at tissues and cells removed from patients undergoing a tissue biopsy or a surgical procedure. The process is summarized as follows. First, they examine the tissue at naked eye looking for abnormalities and then select pieces to examine in more detail. These small pieces are treated with chemicals so that very thin slices can be cut. Aiming to identify the structures present in the cut slices, they are stained using different techniques [122, 157]. Most diagnoses are based in assessment of hematoxylin and eosin (H&E) stained tissues. Hematoxylin stains nucleic acids in blue tones while Eosin colors acidophilus structures (e.g., cytoplasm) in varying degrees of pink [52] (See Figure **1-2**). Finally, the stained slices are examined by histopathologists under a microscope.

Histopathology specimens have a vital importance in patient care [8]. Using information from them, pathologist can establish a diagnosis and provide the physician who is treating the patient with information for decision making and treatment panning [8, 157].

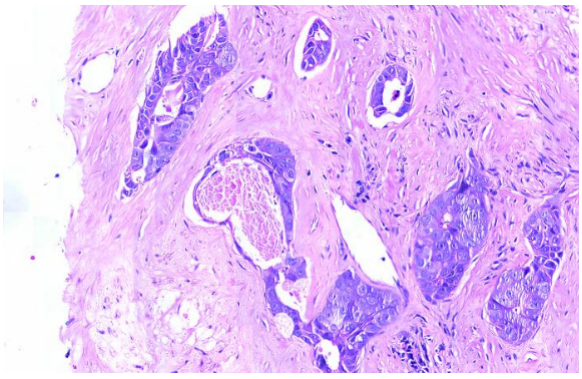


Figure 1-2: Histologic specimen of a breast tissue stained with hematoxylin (blues) and eosin (pinks).

Pathology challenges

As scientific foundation of medicine, pathologists carry a great responsibility, their decisions have crucial care repercussions. An incorrect diagnosis could result in a patient to undergo unnecessary surgical treatment and therapy, or have a delay in beginning treatment, which could be fatal [175]. However, this is not an easy task since pathology faces several challenges, as described below.

- High workloads. Demand for pathology services is raising, which represents an increased number of biopsies to assess and reports to fill out [81]. This has incidence in different aspects. First, turn around times might increase, so high-risk cases could be delayed. In addition, reliability of a rendered diagnosis might be affected.

- Cumbersome and time consuming tasks. Part of a pathologist's duties include time consuming tasks such as manual counting of the number of cells existent in a sample tissue. For example, during a myelography procedure, pathologists have to manually count the number of blood cells of different types to either diagnose or stage the disease in patients with hematological diseases (leukemia, anemia, lymphomas, etc.). Although pathologists employ auxiliary tools, e.g., cell counter devices (See Figure **1-3**), the whole task is cumbersome and time-consuming.

- Subjectivity. Recognizing, analyzing and understanding changes at the level of the tissue architecture or individual cells is a subjective process [122], so the inter-observer variability is a constant challenge. Usually, pathologists must face highly complex cases which make difficult to reach a consensus.

- Communication and inter consulting barriers. Usually, some cases must be sent from a laboratory to a distant location so other experts may examine the case and provide a second diagnostic opinion. This is not only a time consuming task but also there is a risk cases are lost or damaged while being transported.

- Lack of dynamic educational strategies. Education in pathology meaningfully relies on consulting books, which are expensive, easily outdated, and have limited interaction possibilities. Pathology is a dynamic field and, as such, it needs to provide trainees with proper tools that improve the learning process and allow them to face the complexity of the field.

Digital pathology

Digital pathology consists in the analysis of histopathological samples by means of computer-based tools (See Figure **1-4**). The first step involves digitization of tissue glass slides, i.e.,

Figure 1-3: A blood cell counter, used to manually count white blood cells. Source: http://www.naugramedical.com/

the histopathology specimens. This process is carried out by scanning devices, which take a set of consecutive photographies of the glass slide and then stitch them out by registration algorithms.
Images are captured at a high resolution level (usually, 40×) to provide detailed visual information of the tissue. Due to the large resolution of the resulting images, the file sizes are on the order of gigabytes. Finally, images are stored in a server so they can be accessed. Stored images can be visualized by means of software tools specially developed for navigation of tissue virtual slides also known as whole slide images [56].
Digital slides provide important benefits over glass slides. First, digital slides are not fragile and do not get affected by passage of time. In addition, digital samples can be easily shared between experts located in different physical locations to do remote pathology consultation without the need of physically shipping slides around [106, 150]. Furthermore, access to large digital repositories of tissue slides is a huge potential educational resource for medical students and pathology residents who could be trained to identify and recognize pathology cases in a dynamic fashion [106].
An additional benefit of digital pathology is that, by means of computer vision and other algorithms, quantitative data could be extracted from such images to provide pathologist with information to support their diagnostic and prognostic tasks.

Challenges of digital pathology

Everyday, a large quantity of medical images are being produced in hospitals and medical centers for diagnosis and therapy; unfortunately, the valuable information of this data repositories is not being exploited [118]. The development of systems able to to efficiently manage, access, and share such data could be highly valuable for clinicians, especially with the current trend toward evidence-based practice of medicine [27]. Nonetheless, current systems are far from this objective since their interaction mechanisms are very limited, i.e., they do not allow users to easily and dynamically visualize the content and extracting information from it. Some reasons for that are described hereafter.

- Quantity of data. As previously stated, histopathology specimens are digitized at a high magnification to provide the pathologist with the required information for diag-

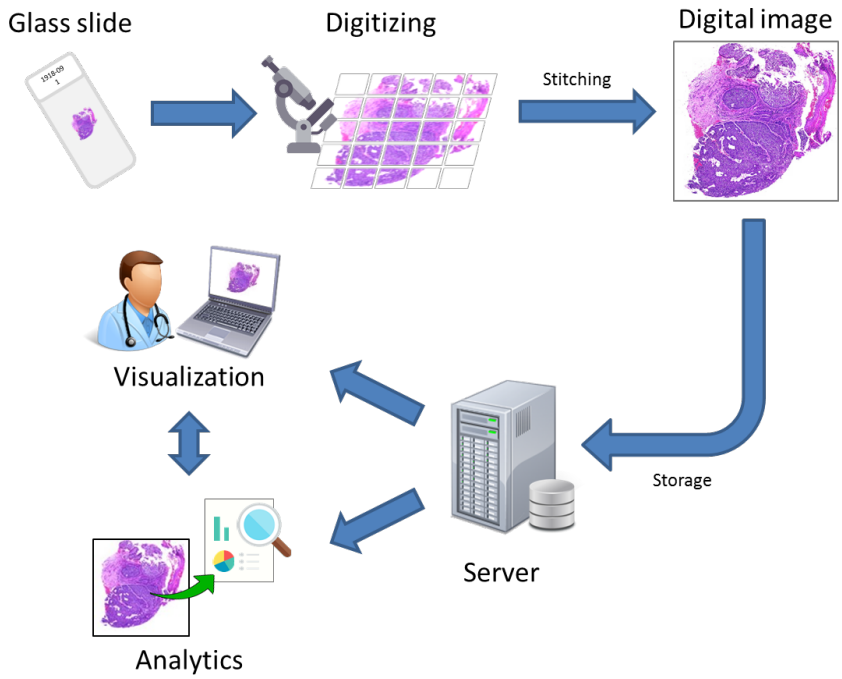


Figure 1-4: General workflow of digital pathology. First, a glass slide containing a tissue sample is digitized by means of a specialized scanner. Usually, such scanners take different consecutive pictures of the sample and then use registration algorithms to obtain a single final digital image. Such an image is stored into a server, so it can be later accessed for either visualization or automatic analysis purposes.

nostic; for this reason, these images are very large (on the order of gigabytes). Proper interaction with such large content represents a very challenging task in terms of storage, memory, processing, and streaming. Efficient and effective approaches to allow pathologists to visualize and interact with content should be developed.

- Complexity and variability of biological structures. Histopathological tissues exhibits a wide range of structures with different levels of complexity, so accurate identification of them is a challenging task. Different automatic approaches have attempted to model some structures based on low-level features such as color, shape, or texture; unfortunately, visual similarity is not enough. First, some structures reveal a high intra variability, i.e., a same structure may appear in very visually different manners. Likewise, two very similar structures could have a complete different biological meaning. In both cases, the context play a very important role to guide a pathologist towards an accurate diagnosis. Effective automatic strategies must inquire not only visual appearance of basic structures such as nuclei or glands, but also their context and architecture.

- Limited annotated data. Modern machine learning algorithms have made a good job at recognizing pattern in histopathology images; however, most of them need proper quantities of annotated data to work. Despite the large quantity of generated data everyday at hospitals and medical centers, most of them are not annotated. Data annotation is an expensive and time-consuming labor that, in addition, should only be performed by expert people to assure the quality of the content. Unfortunately, experts hardly have time for this kind of tasks, especially because the amount of content of such image databases is beyond their manual indexing capability [101, 27, 180].

Thesis outline

As previously stated, current medical systems do not provide proper interaction mechanisms thereby limiting a good practice of EBM. For this reason, in this thesis we provide a set strategies that attempt to improve the way pathologists and systems interact. In particular, we aim to answer two research questions:

- How to efficiently access to histopathology content?
- How to extract relevant information from this content?

The remaining chapters of the thesis are organized as follows:

- **Chapter 2: *Efficient access to digital histopathology images***

  The large size of images obtained from histopathology specimens represents a computational challenge for visualization and interaction. In this chapter, a low-computational-cost strategy to ease visualization of histopathology images is introduced. The work presented in this chapter has been published in a journal and different conferences:

  - Germán Corredor, Eduardo Romero, and Marcela Iregui. *An adaptable navigation strategy for Virtual Microscopy from mobile platforms.* Journal of Biomedical Informatics, vol 54, pp 39-49, 2015
  - Germán Corredor, Marcela Iregui, Viviana Arias, and Eduardo Romero. *Flexible architecture for streaming and visualization of large virtual microscopy images.* Proc. of the MICCAI - Workshop on Medical Computer Vision 2013, Nagoya - Japan, 2013
  - Germán Corredor, Marcela Iregui, Viviana Arias, and Eduardo Romero. *Concurrent access to a virtual microscope using a web service oriented architecture.* Proc. of the IX International Seminar on Medical Information Processing and Analysis, Mexico City - Mexico, 2013

- **Chapter 3: *Exploiting information from nuclei and their context for histopathological analytics***

  Considering that *the cell is the fundamental unit of pathology* [91, 166], in this chapter, we present different strategies for exploiting information from cell nuclei in conjunction with their context. This information is then employed for identification of pathological structures and inferring prognostic information. Specifically, two study cases are presented; the former is focused on breast cancer cases and the latter on lung cancer cases. The work presented in this chapter has been accepted for publication in a journal and was published different conferences:

  - Germán Corredor, Xiangxue Wang, Yu Zhou, Cheng Lu, Pingfu fu, Konstantinos Syrigos, David Rimm, Michael Yang, Eduardo Romero, Kurt Schalper, Vamsidhar Velcheti, and Anant Madabhushi. *Spatial architecture and arrangement of tumor-infiltrating lymphocytes for predicting likelihood of recurrence in early-stage lung cancer*. Accepted for publication in Clinical Cancer Research.
  - Xiangxue Wang, Germán Corredor, Eduardo Romero, Andrew Janowczyk, Yu Zhou, Michael Yang, Vamsidhar Velcheti, and Anant Madabhushi. *Computerized Density Estimation of Tumor-Infiltrating Lymphocyte in H&E TMAs Predicts Recurrence in Early Stage Non-Small Cell Lung Cancer*. Proc. of the USCAP 106th annual meeting, San Antonio - USA, 2017
  - Paula Toro, Germán Corredor, Xiangxue Wang, Viviana Arias, Vamsidhar Velcheti, Anant Madabhushi, and Eduardo Romero. *Quantifying expert diagnosis variability when grading tumor-Infiltrating lymphocytes*. Proc. of the 13th International Symposium on Medical Information Processing and Analysis, San Andres - Colombia, 2017
  - Juan García-Arteaga, Germán Corredor, Xiangxue Wang, Vamsidhar Velcheti, Anant Madabhushi, and Eduardo Romero. *A lymphocyte spatial distribution graph based method for automated classification of recurrence risk on lung cancer images*. Proc. of the 13th International Symposium on Medical Information Processing and Analysis, San Andres - Colombia, 2017
  - Germán Corredor, Xiangxue Wang, Cheng Lu, Vamsidhar Velcheti, Eduardo Romero, and Anant Madabhushi. *A Watershed and Feature based approach for automated detection of lymphocytes on lung cancer images*. Proc. of SPIE Medical Imaging 2018, Houston - USA, 2018
  - Germán Corredor, Cristian Barrera, Paula Toro, Ricardo Moncayo, Hannah Gilmore, Anant Madabhushi, and Eduardo Romero. *Detection and grading of ductal carcinoma in situ by using structural features*. 14th European Congress on Digital Pathology and the 5th Nordic Symposium on Digital Pathology, Helsinki - Finland, 2018

- Germán Corredor, Cristian Barrera, Xiangxue Wang, Anant Madabhushi, and Eduardo Romero. *Phenotyping Tumor Infiltrating Lymphocytes on H&E Tissue Images: Predicting Recurrence in Lung Cancer*. Submitted to SPIE Medical Imaging 2019, San Diego - USA, 2019

- **Chapter 4: *Using interactions from experts for histopathological analytics***

  In this chapter, we present an approach that employs implicit relevance feedback from experts to identify regions of interest in histopathology images. By means of this strategy, high-level information is obtained from experts while they perform routine tasks, i.e., they do not need to manually annotate images. Specifically, a study case focused on skin cancer cases is presented. The work presented in this chapter has been published in a journal and different conferences:

  - Germán Corredor, Jon Whitney, Viviana Arias, Anant Madabhushi, and Eduardo Romero. *Training a cell level classifier for detecting basal cell carcinoma by combining human visual attention maps with low level handcrafted features*. Journal of Medical Imaging, vol 4(2), 2017
  - Germán Corredor and Eduardo Romero. *Learning Histopathological Regions of Interest by fusing bottom-up and top-down information*. Proc. of the International Conference on Image Processing (ICIP) 2015, Quebec city - Canada, 2015
  - Daniel Santiago, Germán Corredor, and Eduardo Romero. *A sparse representation of the pathologist's interaction with whole slide images to improve the assigned relevance of regions of interest*. Proc. of the 13th International Symposium on Medical Information Processing and Analysis, San Andres - Colombia, 2017

- **Chapter 5: *Conclusions*** The final chapter in this work presents some conclusions and discuss the potential impact of the interaction strategies presented in this work. Some future research directions are also presented.

2 Efficient access to digital histopathology images

2.1. Introduction

Virtual Microscopy (VM) may be thought of as a collection of techniques that facilitate a set of Whole Slide Images (WSIs) can be examined from any place and at any time. Typically, a histopathological specimen is digitized at the higher possible magnification to provide the pathologist with the required information for diagnostic, research, training or educational tasks [34]. During the last decade, the dynamic interpretation of WSIs has been integrated with many pathology activities such as teaching, research, digital archiving, teleconsultation, and quality assurance testing [2]. Different works have studied the viability and agreement of diagnoses by using WSIs, reporting promising results [51, 2, 112]. Recently, medical schools in the United States have introduced digital pathology courses and virtual slide laboratories, promoting a generation of pathology trainers who may prefer digital pathology imaging over the traditional hands-on light microscopy [171]. A large variety of technical solutions supported these studies, e.g., Aperio ImageScope [51], home systems such as U-DPS [2], DMetrix Digital Eyepiece [171], or WebScope [112].

Several technical and logistical barriers, however, have delayed WSI becomes a widely accepted pathology modality [124]. A proper management of the number of files generated by a WSI demands large memory, processing, and storage resources since the size of a WSI is typically on the order of gigabytes. Furthermore, since there is not a common image format for virtual slides, a large number of proprietary or vendor-specific formats has been constantly modified as long as new scanners have been introduced [161]. Standardization not only allows an user to perform certain functions in an optimal way, but it also offers quality guarantees, interoperability, independency from vendors and equipments, access to new technologies, and possibilities to scale applications according to new requirements. A wider VM use will require full integration with laboratory information systems, seamless connectivity over broadband networks, efficient workstations, cost-effective storage solutions, and standards-based informatics transactions for integrating information with WSI [171, 124, 80]. Lately, image quality improvements, smaller scan times, and image-viewing browsers have converted digital pathology into an actual opportunity [124]. Overall, actual clinical scenarios require access to these files from any location, reason by which mobile devices might be considered as the support nodes of a VM network. However, such devices are still very

resource limited [120] and, yet communication channels have remarkably improved, network bandwidths are frequently insufficient.

This problem has been addressed using a variety of approaches, the most common consisting in constructing pyramidal data structures that deal with different image scales that are stored as independent files [3]. For a requested Region of Interest (RoI) to be displayed, a complex combination of pyramidal files must be composed and this is usually computationally expensive. Some pyramidal approaches (e.g., HD View, Zoomify, Gigapan, and Google Earth) have been evaluated from a VM standpoint, reporting pleasant interaction experiences when navigating a single WSI from a conventional computer [3]. Nevertheless, these approaches might be very limited when displaying a WSI from a low resource mobile device since in such a case, applications should deal with variable storage requirements, low compatibility, high processing demand, and poor adaptation to different displays. Likewise, limited devices may have trouble managing a large number of files since their cache space may be easily overflowed. Aperio [9], a commercially available software allows an user to pan and zoom in and out virtual slides, but this system is computationally very demanding and requires a powerful infrastructure. Similar aproaches are OpenSlide [63], NYU Virtual Microscope [72], and Deep Zoom (formerly called Seadragon) [114], among them, Openslide is an open source library devised to display WSIs and is compatible with different image formats. The NYU Virtual Microscope uses the Google Maps API and Deep Zoom is part of the Microsoft Silverlight platform, a proprietary software with a very limited mobile version. These last three applications are based on a pyramidal structure and share the limitations aforementioned for mobile devices. A different approach was proposed by Hadwiger et al. who introduced a multi-resolution virtual memory that performs dynamic updates and deals with missing data [70]. This system is not based on any standard, uses the lossy JPEG version and was devised to display data at a full resolution, a bottleneck in limited devices.

VM demands highly flexible, efficient, manufacturer independent, and standard-based tools [161, 137]. An alternative to the artificial pyramidal approach is the JPEG2000 standard, founded on the concept of making available any piece of required information, i.e., a particular spatial region at any desired quality and magnification. The standard appears to be flexible enough as to address the issue of streaming and visualizing demanding content in mobile devices [139], such as WSIs. This standard was smartly conceived to be granular, i.e., an image can be decomposed and compressed in small independent parts (grains) of information at different levels of magnification, several degrees of quality and independent spatial representation, facilitating a separated access and process of specific regions of the image, while also supporting large file sizes and a larger dynamic range of the pixel representation [156]. In addition, by the JPIP (JPEG2000 Interactive Protocol) standard, the client may demand specific RoIs from the server, instead of remotely accessing the whole JPEG2000 content [155]. Nevertheless, the JPEG2000 standard complexity may make it very expensive in computational terms [161] and therefore unrealistic at supporting a VM network. Basically, data allocation can be an actual burden of the navigation while the decoding process

may be on the order of 2-5 seconds, even when decoding a small VS of 9000 × 12000 pixels. There exist some applications using different JPEG2000 implementations, all of them decompressing data at the server side and leaving to the client a purely passive role at receiving the raw decoded information to be displayed, for instance IIPImage [128], Djatoka [33], JVSMicroscope [161], and Web Microscope [103], being the latter a reference in certain academic and clinical institutions. This strategy throws away the JPEG2000 high compression rates since only uncompressed data are transmitted and ignores the potential processing improvement at the client side. Other works [75, 60, 59] have explored the JPEG2000 as an interaction tool for VM by modifying the decoder implementation, retrieving and decompressing specific portions of the codestream; unfortunately, this tightly-coupled solution could be hardly extended to different platforms. Finally, Rosenbaum et al. proposed to send only the RoI encoded information and to complete the missing codestream (untransmitted) at the client side with a pre-defined template [139], but then the decompression times result equivalent because of the size of data.

2.2. A strategy to efficiently access and interact with histopathological images

2.2.1. An adaptable strategy for interaction with histopathological images from mobile devices

This section introduces an adaptable and low computational cost VM framework that exploits the JPEG2000 potentiality at both the server and client sides. The possibility of meeting any requirement, i.e., any spatial region at any size, with a desired magnification and quality, makes this proposal adaptable to new scenarios, in particular to the training and educational VM. In that case, a group of pathologists or students, might simultaneously access the same WSI and therefore saturate the network. The main contributions of this work are:

- Unlike most existent solutions, this strategy has been devised to maximally exploit the processing resources at both the client and the server so the server sends compressed data and the client decompresses data, even under very limited computational capacity.

- A smart decoding strategy addressed to construct any RoI by setting the requested region to an image which can then be decompressed by any standard decoder.

- A flexible and scalable data management strategy that efficiently retrieves JPEG2000 compressed data at the server side, independently of the image size, by a coupled designed meta level index file.

- A loosely-coupled architecture, web service oriented, providing functionalities that support interoperable and standard interaction over the network. This highly adaptable architecture adjusts the content to the user requirements, the device capacity and the network bandwidth, while it offers a progressive lossless visualization.

Materials and Methods

JPEG2000 Overview

JPEG2000 is a highly flexible image coding standard that optimizes interaction with compressed data [147, 153]. A key feature of this standard is that it encodes multiple resolution levels and quality layers. Resolution is related with the number of pixels that are needed to ensure that, at a particular image size, the displayed information is maximum. In contrast, the quality is a function of the number of bits that are used to represent a pixel. Resolution flexibility implies that an image can be retrieved at a low resolution (a small version of the image) and can be enlarged (by a factor of two) by adding the missing data and only these data [121]. The quality is connected with the concept of progressive user interaction and consists in displaying a very basic version of the image, with few details, that are progressively added as long as the user demands more information, until reaching a full lossless visualization, if needed [121].

The JPEG2000 norm is based on the Discrete Wavelet Transform (DWT) and the Embedded Block Coding with Optimal Truncation (EBCOT), both endowing the data representation with high granularity [154] (see Figure **2-1**). The DWT decomposes the input image into frequency subbands, producing a natural multi-resolution decomposition, with basically two wavelets: the Daubechies 9-7 for lossy compression, and the reversible Daubechies 5-3 for lossless compression [76]. The DWT image is divided into tiles that allow random access to spatial regions with different frequential information. The EBCOT compresses the image into small blocks (code-blocks) that encode the DWT coefficients of each subband. Each of the codeblocks, composed of a set of bit-planes, is ordered by levels of relevancy known as the quality layers, each containing a part of the whole information. Finally, the packet, the basic JPEG2000 information unit, is responsible for storing compressed data at a particular resolution level, a single spatial region and a unique quality level. The standard allows a progressive reconstruction of the original image by dynamically adding missing packets, improving the visualization of the image until a perfect reconstruction is obtained [156].

Smart decoding strategy

The JPEG2000 decompression process is expensive because of the decoding and the inverse transforming processes, a fact that has limited the JPEG2000 application in VM. A partial remedy to this bottleneck has consisted in assigning the processing responsability to a server which decodes the codestream and sends the resultant raw data [128, 33, 161, 103]. The client acts as a simple information receptor and the potential client resources are never used,

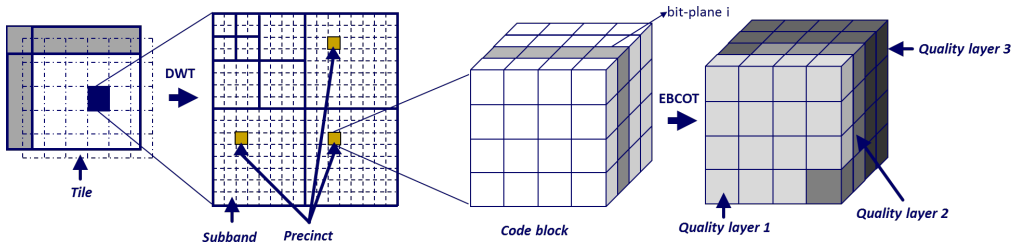


Figure 2-1: JPEG2000 data partition. The image is divided into smaller rectangular regions known as tiles. The DWT is applied independently within each tile component, yielding the respective subband tree structure. Each subband of each tile is further partitioned into code-blocks, which are then independently coded. After the EBCOT encoding, optimal truncation points for each quality layer are identified. Finally, each resolution of each tile component is grouped into precincts that represent specific spatial regions of the image.

overloading the communication channel by transporting uncompressed data. Furthermore, the transmission of raw data necessarily reduces the possiblity of storing relevant information at the client side and thus the potentiality of implementing effective cache policies that may reduce the network traffic. The option of decoding at the client side has been introduced either by decompressing the whole image, a real problem with the WSI sizes, or by adapting the decoder implementation to decompress specific packets [60, 59]. The main drawback of this last solution is the inevitable dependence on the decoder implementation, or the problem of managing the dynamic organization of data, which in some cases has been approximated by completing the requested codestream with zeros [139], but then the decompressing times result to be equivalent to those obtained with the whole image. In summary, most of the existent VM applications have ended up by using JPEG2000 as a simple compression format, without exploiting its flexibility at representing the data.

Unlike previous approaches, the proposed strategy effectively integrates the client to the processing by generating, for each requested RoI, a new small JPEG2000 coded image that meets the desired RoI, i.e, same dimensions, resolution levels and quality layers. This is achieved by modifying the image main header and assembling that header with the packets associated with the RoI. In this way, an efficient decompression is accomplished by processing exactly the required image portion, at any desired quality and magnification. A resultant side advantage of this strategy is the independence of the implementation, i.e., any decoder can be used.

Flexible data management

In an actual navigation scenario, the retrieval of specific data from a JPEG2000 file demands an intensive search within the codestream to localize the desired packets, whose location is coded in structures known as tag-trees [76]. Any individual query requires these structures to be decoded [76], a process that may take about 2 seconds for a single packet of a large

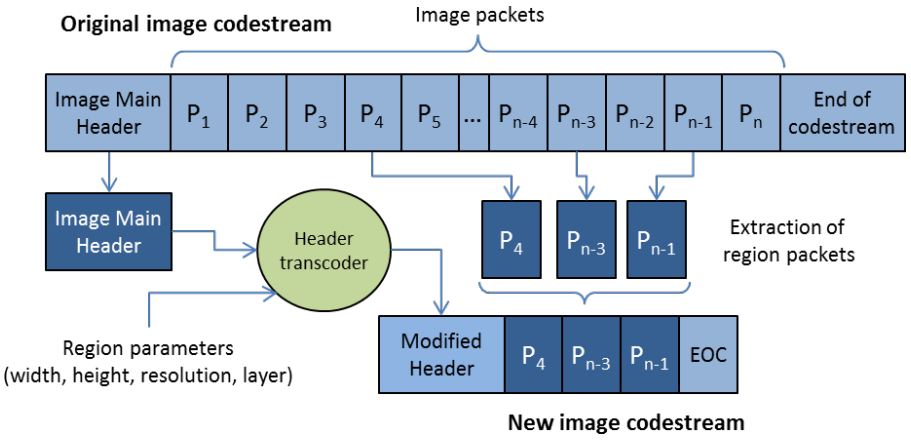


Figure 2-2: Illustration of the smart decoding strategy, which is performed in three steps: (1) Extraction of the image main header and packets that represent a region r. (2) Modification of the image header by adjusting the main data using the region r features (width, height, resolution level, quality level). (3) Assembling the modified main header and the extracted packets to generate a new smaller codestream that represents the requested RoI.

WSI. This problem has been overcome by using index files [154]. The herein used index files are based on the JPIP standard specification [77] and they are simple text files that provide an organized structure of the general image data at two different levels (see Figure **2-3**), the global image information (width, height, progression order, number of components, number of quality layers, number of decomposition levels, etc) and the particular local configuration at the level of packets (quality layer, component, resolution, precinct number and byte ranges), thereby facilitating any application to identify and to extract bytes directly from the JPEG2000 files and therefore to meet complex user requirements.

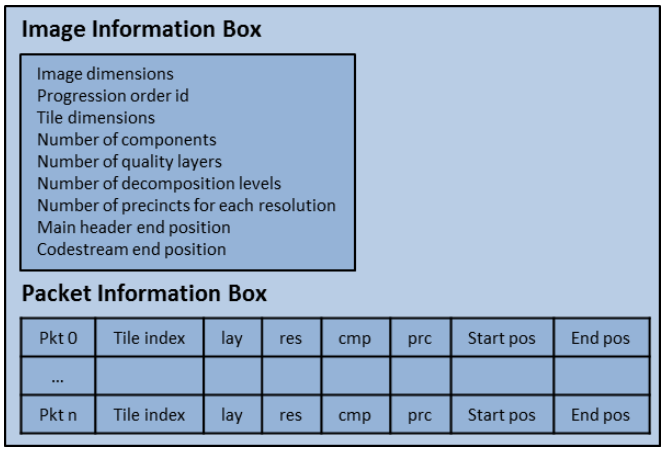


Figure 2-3: Structure of the adapted index file.

In general, when navigating an image, a user requests Windows of Interest (WoI) that are identified by their image location, size (weight and height), resolution and quality. The system must therefore use a parser that maps the user request to the specific packets in the compressed image or codestream, after the image information box in the index file (See Figure **2-3**). Using the WoI coordinates, the parser computes the identifiers of those precincts[1] associated to the spatial query. This information, together with with the specific requested resolution and layer, is used to calculate the corresponding packet IDs. Once these IDs are found, a search in the index file determines the initial and final locations in the codestream of the specific bytes, i.e., the position of this packet in the compressed file. Finally, these bytes are extracted directly from the compressed file.
While these index files are very important to accelerate the time required to locate a packet within the codestream, they require an associated efficient access technique. Depending on the WSI size, the index files may result as large as a WSI and the search process may become as slow as to become an actual navigation bottleneck. For this reason, indexation was herein optimally managed by designing a hash-based structure composed of multiple small index files that may be selectively loaded to meet a client request. Each file stores data from a given set of packets. When a packet information is required, its identification number is used to determine the index file containing the necessary data. This strategy provides scalability and proper performance regardless the image size, since just one index file must be accessed and loaded in memory.

Architecture overview
The proposed strategy exploits and extends the benefits of both the JPEG2000 and JPIP standards, adapting the content to the device capacity and user needs (see Figure **2-4**). Basically, any navigation request may be assembled and resources may be optimized if a flexible architecture is capable of implementing the standard granularity [141, 139]. In particular, the level of quality is a variable of the user needs and hence the navigation might be speeded up by setting a maximum quality. Likewise, navigation may also be accelerated if the decoding policies, at both the server and client sides, are completely adaptable to the bandwidth. Any architectural approach must then be flexible enough as to cope with all these different scenarios.
The proposed architecture is loosely coupled and allows integration of different caching and prefetching models, which speed up the browsing performance [41, 59]. Furthermore, the transmitted content can be adapted to the screen size, supporting several devices with different capacities (not only mobile ones). This architecture consists of three loosely coupled layers, described hereafter. Figure **2-5** illustrates the information flow through the different modules of the proposed architecture.

Storage layer: This layer is the repository of the JPEG2000 compressed images and their

[1]Precincts are a JPEG2000 image spatial partition

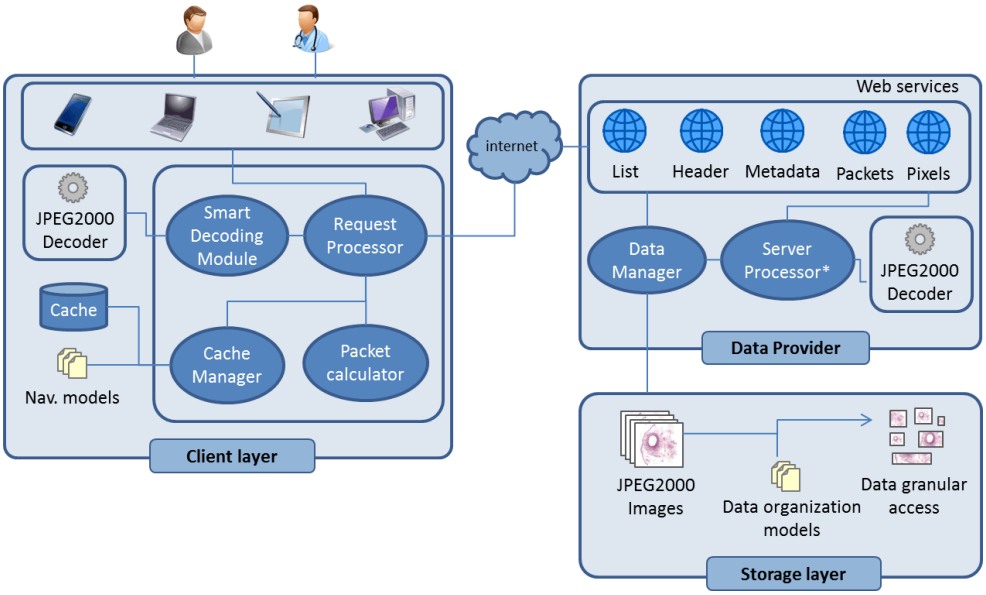


Figure 2-4: Top-level runtime view of the architecture.

respective index files, which facilitate access to such image files. When a new compressed image is stored, its index file is constructed using the information of the main header and the packet markers in the codestream. The proposed approach is independent of any database engine since data are stored as files. This layer is platform independent, i.e., it can be run from any operating system (Windows or UNIX based) and the required storage space depends exclusively on the image sizes. The application that generates the index files was developed using the Java SE platform, which is also platform independent.

Data provider Layer: This layer provides web service interfaces for a client accesses to data in the storage layer. The web services were developed using the Java EE platform and run over any Java Enabled Application Server. Such web services are interoperable and may be consumed by any client application. Four main services are available: *List* sends a list of the available images. *Header* receives an image name and returns the image main header. *Metadata* takes as input an image name and sends specific information, namely dimensions, progression order, number of precincts, number of components, number of quality layers, number of resolution levels, among others. Finally, for the *Packets* service, given an image name and a list of packet IDs, it sends the bitstream of each packet. The Data Manager module is an intermediate between the web services and the storage layer.
Alternatively, the *Pixels* service takes an image name and a WoI request (coordinates, resolution and layer), and returns the pixels of that window. This service may be suitable when decoding can not be performed at the client side, for example devices with very low capacities or for web-based applications, which generally have no support for the JPEG2000 standard. This service connects to a Server Processor module which is responsible for packet calcu-

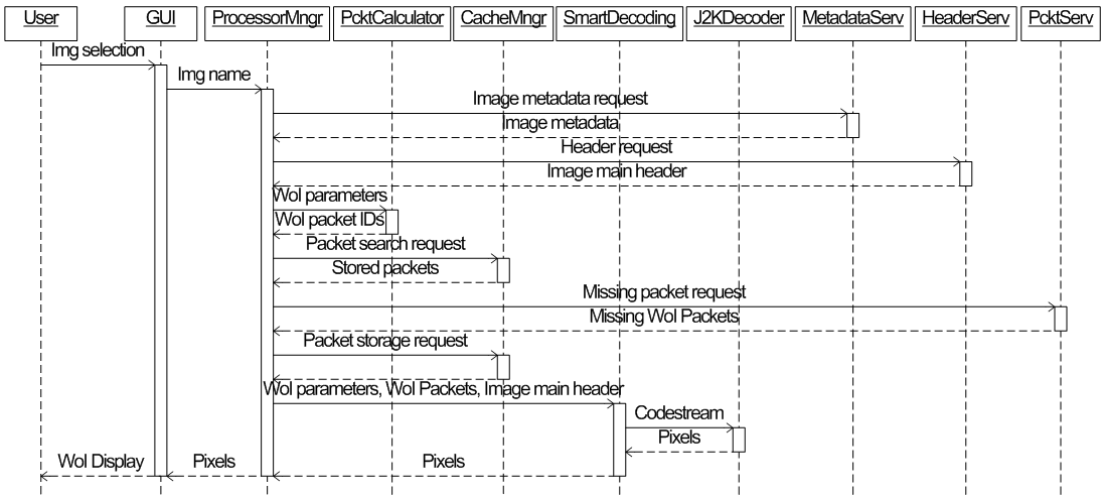


Figure 2-5: Information flow through the different architecture modules. When a user selects an image, a message is sent to both the Metadata and Header services, requesting these data. Then, an initial low resolution spatial request is generated for a user visualizes and interacts with a region of the image. For doing so, the packets of such region (WoI) are calculated and requested to the Packet service if they are not available in the cache memory. Then, the packets, the region parameters and the image main header are used to build a codestream of such requested region. This compliant codestream may be uncompressed with any standard JPEG2000 (J2K) decoder. When a user interacts with the interface (panning, zooming in/out or refining quality), the Processor Manager receives the requested region parameters and the process is repeated, starting from the packet calculation until WoI display.

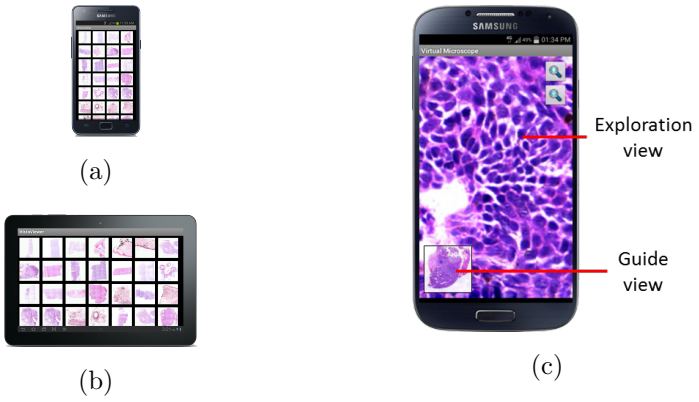
lation and generates a compliant codestream to be uncompressed using a JPEG2000 decoder.

Client Layer: It is composed of several modules for a user may visualize and interact with the WSI. A first module is a graphic user interface (GUI) with panning, zooming-in/out and quality operations. A second module is the cache manager that administrates the memory where previously requested data may be stored. This module removes old/unused data when this is full and takes advantage of spatial, resolution and quality scalability. The size of this cache memory is configurable according to the device capacity and adaptable to different models, for example, the Least Recently Used or the Least Frequently Used. The fact that the data representation is so granular facilitates the design of more complex cache policies that are constructed, after the principle of storing only what is relevant and will be used in the future. A third module is the Request Processor that demands the required data either to the cache manager (if they were previously requested) or to the corresponding web service. This module uses the Packet Calculator to identify the packets that are requested.

In the smart decoding module, the retrieved data from the codestream are mapped to a JPEG2000 image which then can be decompressed using any standard decoder. Finally, raw data (pixels) are displayed by the GUI.

GUI design

From any mobile device the first view is a dynamic list of the available WSIs. The user then selects that one to be examined and may easily switch between different WSIs of the dataset, as illustrated in Figures 2.6(a) and 2.6(b). Each thumbnail image is associated to a list of metadata, containing clinical information, that is pop out when long pressing the thumbnail image. If the pathologist picks a WSI, the navigation starts by displaying two views, a guide window that displays a low magnification version of the WSI and serves for the expert to be oriented within the WSI, and an exploration window showing a RoI (Figure 2.6(c)). At each WSI, the expert may pan, zoom in or out and refine the quality by using some gestures and interface elements.



(a) (b) (c)

Figure 2-6: System GUI. Subfigures (a) and (b) present visualization of the available WSIs in a smartphone and a tablet, respectively. Subfigure (c) presents the GUI for visualization of a WSI. It shows the exploration and guide views. The latter displays a thumbnail that helps to track the explored region within the WSI

Experimentation

Dataset

Experiments were performed with a dataset consisting of twenty skin biopsies of different patients, stained with Hematoxylin-Eosin. Most of these cases are skin basal cell carcinomas, two were lentigines, a melanocytic nevus, and an irritated seborrheic keratosis. The diagnosis difficulty was considered as moderate by our expert dermato-pathologist and, in general, they took less than a minute to perform a diagnosis. The cases were provided by the Pathology Department of Universidad Nacional de Colombia. The cases were collected between 2009 and 2014 and were randomly selected from a set of 98 patients. All the cases were anonymized.

The slides were digitized at 40x using a tri-ocular CARL ZEISS Axiostar plus microscope coupled to a DXM1200 Nikon color digital camera, controlled by a custom motorized scanner. The WSIs were JPEG2000 compressed using 10 quality layers, 4 decomposition levels (5 resolutions), the lossless filter (W5x3) and precinct sizes of 64×64 for the first resolution level and dyadic increasing sizes for the other levels. The WSI resolutions vary from 104 to 340 mega pixels, and their sizes range between 630 MB and 972 MB. The proposed approach was compared with two baseline approaches: the lossy and lossless versions of the JPEG standard, constucting two pyramidal structures with different scale (spatial scalability) and quality (Signal-to-noise ratio scalability) versions of the original image, based on the Google Earth API documentation [64]. Images were JPEG and JPEG-lossless (JPEG-LS) compressed using 10 quality layers, 5 resolution levels, and tiles of 64×64 for the first resolution level and dyadic increasing sizes for the other levels. Finally, the experiments were run using a 1 Mbps network. Table 2.2.1 shows a quantitative comparison of some representative WSIs in terms of formats, resolution, file size and number of files.

Image	1	2	3	4	5
Diagnosis	Nodular basal cell carcinoma	Nodular basal cell carcinoma	Nodular basal cell carcinoma	Solar lentigo	Melanocytic nevus
Resolution	15360x14336	22528x14336	11264x13824	36864x9216	17408x11776
JPEG2000 size	120.9 MB	176.5 MB	153.2 MB	198.5 MB	120.7 MB
JPEG pyramid size	502 MB	758 MB	421 MB	480 MB	499 MB
JPEG-LS pyramid size	1280 MB	1860 MB	1400 MB	1890 MB	1140 MB
Raw size	630 MB	924 MB	768 MB	972 MB	586 MB
# of JPEG2000 files	3	3	3	3	3
# of JPEG & JPEG-LS files	10500	15400	12800	16200	9350

Table 2-1: Comparison of some WSIs of the dataset in terms of formats, resolution, file size and number of files. The number of JPEG2000 files includes index files.

Architecture deployment

The server application was developed using the Java EE 1.5 platform and was deployed on the GlassFish Application Server. The storage and data provider layers run on a computer with 4 GB RAM memory and 2.4 GHz quad core processor. The client application for a single user navigation was implemented in the Android platform. The tests were performed using the Samsung GT-i9100 (Galaxy S2) device, with operating system Android 4.1.2, 800×480 display size, 1024 MB RAM memory and 1.2 GHz dual-core processor. Decoding was carried

out using the Kakadu library [152] version 2.2.3 and the JasPer library [1] version 1.900.1, under the Java Native Interface.
Simultaneous access experiments were run on a desktop client device that randomly executed recorded requests of actual expert navigations. The client application was developed in the Java platform and tests were performed on a computer with 4GB RAM memory and 2.71 GHz dual core processor, using The Kakadu library version 2.2.3 as the decoder.

Evaluation

Provided that limited processing power, memory and bandwidth are the most critical constraints of any mobile device, evaluation is addressed to measure an efficient trade off between the navigation time and the percentage of used resources. The evaluation of the presented strategy was carried out by addressing the following quality attributes:

- Efficacy: The system complies the user requirements and the diagnostic task is not altered by the system.
- Efficiency: The tasks are performed using the system resources appropiately (low response times and low memory consumption).
- Concurrency: The system can simultaneously attend a given number of users without affecting the response times.

According to the previously mentioned considerations, the experiments assessed: the global perception of the user during the diagnosis and its accuracy, size of the image representations, memory consumption, and response times in transmission and decoding. These last were measured for a single and multiple users. For doing that, two sets of experiments were executed, a mobile client device was firstly used to evaluate the system performance during a single user navigation, and then, a desktop client device was used to evaluate concurrent access.
Given that efficiency does not depend on the image content, but on its size and diagnostic information, the single user navigation experiments were carried out using one WSI, the one that spanned the largest navigation times, i.e., the largest number of requests.

Experiment 1: Navigations of Pathologists
An initial test was made with two pathologists with at least ten years of professional experience. They were requested to diagnose the 20 WSIs, using a custom GUI design with a list of the available WSI. Overall, pathologists are not familiar with computer, so for avoiding any navigation bias because of an inappropriate use of the GUI, before the first navigation, each pathologist was instructed about this interface with a test image. Each of the navigation operations were then previously assessed by them, the different zooms, the resizing

operations, the spatial jumps and the quality improvement. When they picked a particular WSI, the application displayed two windows: the exploration and guide views. The guide view resulted very useful since it facilitated a preliminar diagnosis that drove the navigation. This view may be hidden, if needed, to increase the exploration area. Pathologists performed different exploration paths, identifying relevant regions and diagnosing them, by using different gestures, namely taps, drags, pinches and stretches.

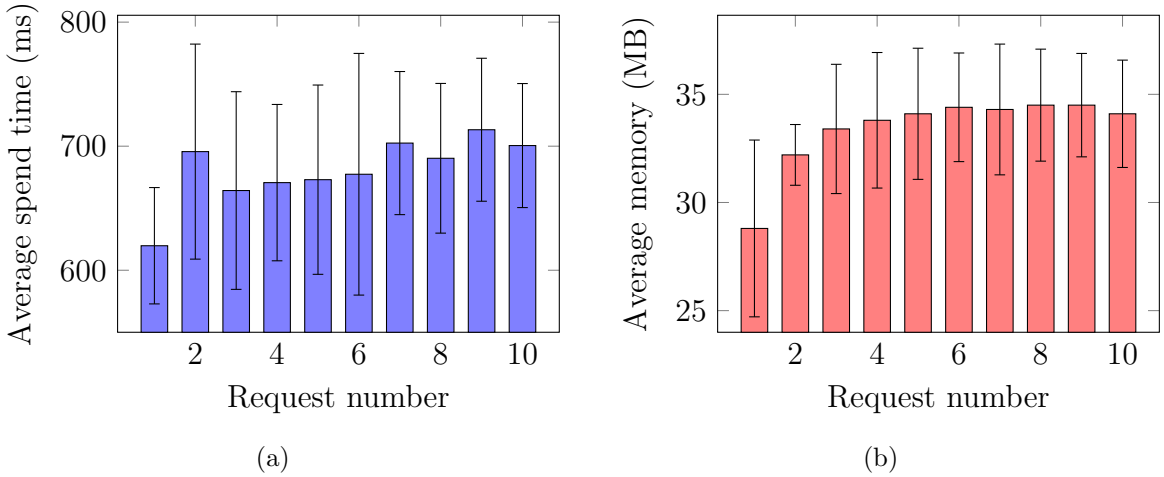


Figure 2-7: Performance of the system for the 10 first navigation steps of the pathologists during a diagnosis task on the whole dataset. Subfigure (a) presents the response times and Subfigure (b) presents the memory usage

At the end of the navigations, the pathologists reported a pleasant interaction experience, they agreed for all the cases and they concluded that the application might then be suitable to perform diagnosis tasks. In average, the experts spent about 1 minutes per WSI and 3.6 seconds per examined region. During the first 10 navigation steps, the average response time remained below the 800 milliseconds (Figure 2.7(a)) and the system memory consumption remained below the 38 MB, only a 3.7 % of the testing mobile device capacity (Figure 2.7(b)). Finally, expert's requests were recorded to perform the concurrency tests and to compare the proposed model performance with the baseline (JPEG and JPEG-LS). In such tests, variables such as size of the respresentations, transmission, decoding and memory usage were measured. Results are shown in the following sections.

Experiment 2: Size of the representation

The WSIs were JPEG2000 encoded and their average file size, including indexes, was 151.9 MB. Likewise, after construction of the JPEG and JPEG-LS pyramidal structures, an average of 12850 files were generated per image, resulting from splitting the image into different blocks, each coded in 5 resolutions and 10 quality layers. The JPEG pyramidal construction had a final average size of 532 MB, while the JPEG-LS structure, a final average size of 1.514

GB (Figure **2-8**). Intergroup comparison under a one-way analysis of variance (ANOVA) showed significant differences ($p < 0{,}05$) and pair-wise post hoc test (Bonferroni) indicating that these differences could be attributed to differences between JPEG2000 and the other formats. It should be strengthen out that while JPEG2000 data organization allows progressive reconstruction by size, resolution and quality refinement as long as more data are received, the pyramidal representation for JPEG and JPEG-LS requires storage and transmission of redundant data.

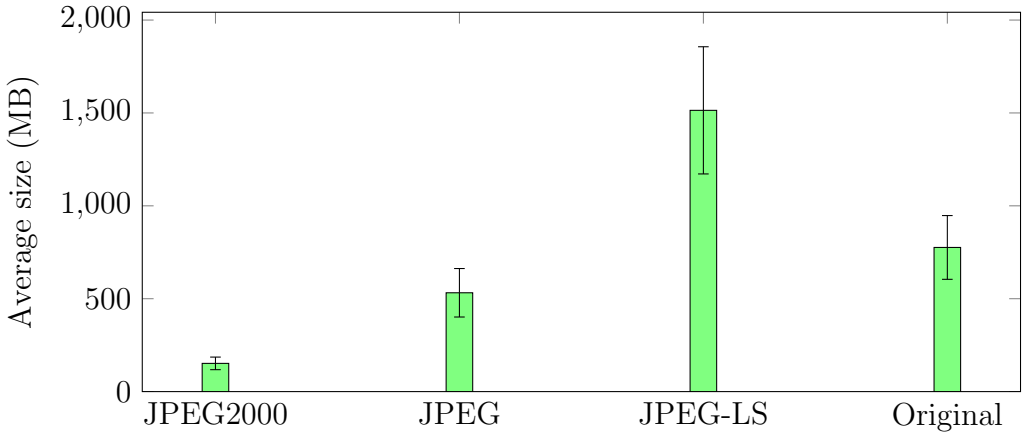


Figure 2-8: Size comparison of the image representations. JPEG2000, including index files, presents the lowest size, contrasted to the JPEG and JPEG-LS pyramidal structures.

Experiment 3: Transmission efficiency

Latency was estimated by means of the data size and the bandwidth, measuring the transmitted bytes per request. Figure 2.10(a) shows these results for the WSI with the largest number of requests. The ANOVA test showed no significant differences among the means of the three formats ($p < 0{,}05$) and the pair-wise post hoc test (Bonferroni) indicated that these differences may be attributed to differences between JPEG2000 and the other formats. When quality refinements and magnifications were required (requests 9-14), it can be seen that JPEG-LS is highly demanding.

Experiment 4: Decoding Performance

Figure 2.10(b) presents the decoding time for each of the requested regions during a navigation. In this case, the ANOVA indicates significant differences ($p < 0{,}05$) and the post hoc Bonferroni test showed insignificant difference between JPEG y JPEG-LS. Results show that these times were larger for the JPEG2000 approach. Nevertheless, it is important to keep in mind that this experiment in paticular was performed using the JasPer library [1]. In constrast, the decoding times for the previous navigation path using Kakadu library [152] were

about one third, a figure that could even decrease more with the Intel IPP based JPEG2000 library (×10 more rapid) [73] or the last Kakadu release (version 7.3.3) that includes a "speed pack"that speeds up to 40 % or 50 % the decoding process [152]. Hence, the use of more efficient or hardware-based decoders might improve the decoding times.

Experiment 5: Effect of the WSI size

The effect of the WSI size on the speed and efficiency of the system was assessed by including 3 new WSI images with compressed sizes of 3.89 GB, 2.6 GB and 1.3 GB and whose raw sizes were 12.5 GB, 9 GB and 4 GB, respectively. This experiment consisted in measuring the response times of requesting the image dimensions and the main header from the index file by the respective web services. Likewise, three different requests were also measured, namely, a 256×256 spatial, panning and zoom-in queries. Results, shown in Figure 2.2.1, demonstrate that in spite of the large size differences in the set of images, the response times are quite similar, not exceeding the 120 milliseconds, i.e., the WSI size has a moderate impact on the system performance.

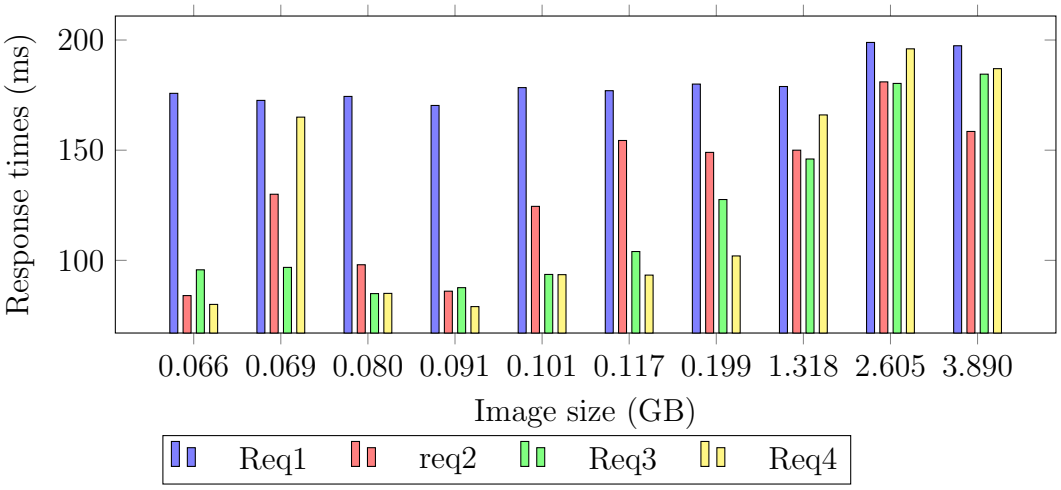


Figure 2-9: Effect of the WSI size on the performance of the system. The response times of four requests were measured for 10 different images: 1) Image dimensions and main header, 2) a 256×256 region, 3) a panning query, and 4) a zoom-in query.

Experiment 6: Memory usage

The figure 2.10(c) presents the use of memory for the three assessed strategies, recording the quantity of used memory in Megabytes each second during the navigation. The ANOVA indicated significant differences ($p = 0{,}002$) while the post hoc Bonferroni test showed no significant difference between JPEG2000 and the other formats. These results show that both strategies, JPEG and JPEG-LS, present a higher memory consumption that rapidly increases during the navigation. As long as the navigation lasts, more complex variable requests must be met, namely overlapped regions, magnifications or quality refinements, making the

device memory to fill quickly. It is worthy to mention that during experimentation, consecutive execution of different navigation protocols using the JPEG and JPEG-LS approaches caused memory overflows.

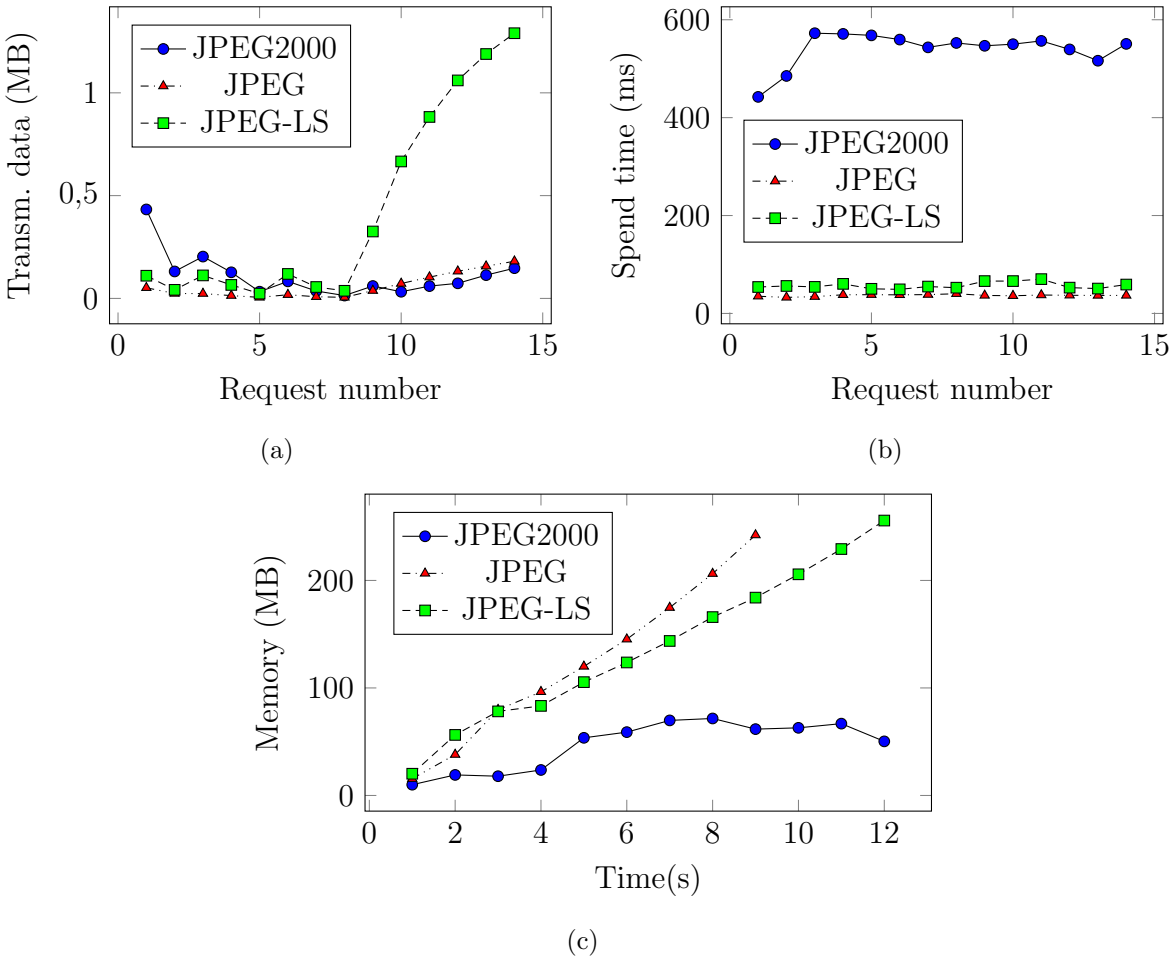
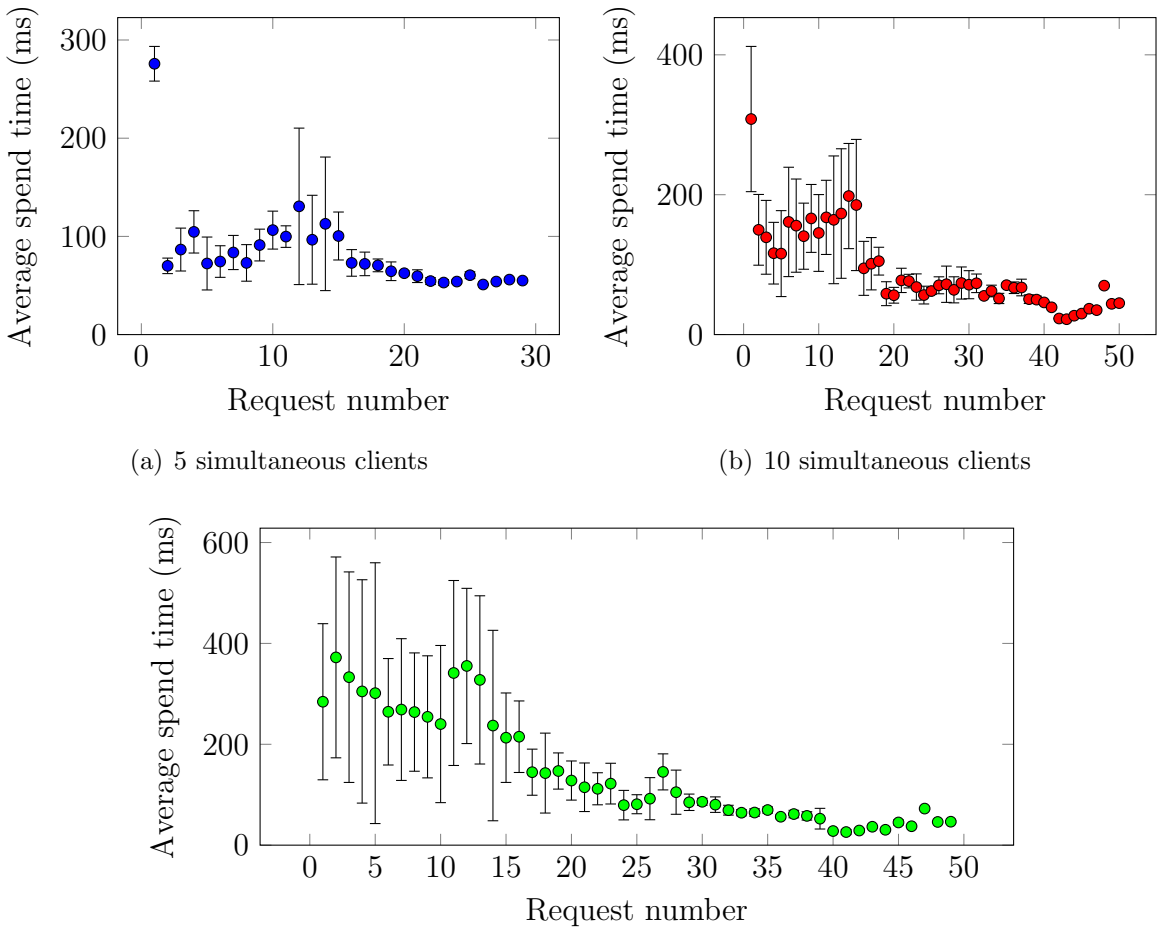


Figure 2-10: Performance of the system measured during the browsing of a selected WSI. a) Quantity of transmitted data for each request, b) Decoding time for each request, c) use of memory during the execution time.

Experiment 7: Simultaneous access performance

The access times were measured for several concurrent clients, using a client desktop device that randomly executed the diagnostic paths previously recorded. Each of these observation paths is composed of a set of requests, for which the application recorded the time spanned between a particular request and the data display.

Figure **2-11** presents the results for interaction of multiple concurrent users. Subfigure (a) presents the results with five simultaneous clients requesting data to the server, not exceeding the 280 milliseconds. In subfigure (b), ten simultaneous clients did not go beyond the $300\,ms$.



(a) 5 simultaneous clients

(b) 10 simultaneous clients

(c) 20 simultaneous clients

Figure 2-11: Average response times measured for a set of concurrent clients requesting data corresponding to different image navigation paths.

Finally, Subfigure (c) shows the results of twenty simultaneous clients remaining below the 600 milliseconds. For the first experiment (5 simultaneous clients), significant differences were not demonstrated under an ANOVA analysis ($p > 0{,}05$). However, the ANOVA for the other two experiments showed significant differences ($p < 0{,}05$). In general terms, the system presented response times under half a second even if a basic server was used (4 GB RAM memory and 2.4 GHz quad core processor) and the system was overloaded with a large number of concurrent users. Of course, much more powerful servers may attend a larger number of users.

Discussion

This section has presented a system fully integrated with an actual VM workflow that easily adapts to the device capacity and the expert interaction needs. This is achieved by means

of three key elements. A smart decoding strategy, a flexible data management and an architecture devised to adapt the navigation to the device resources and the network bandwidth, while it maintains the granular data representation of the JPEG2000 standard.

Dynamic interaction with WSIs is a very complex and challenging task due to their large sizes, on the order of Gigabytes [34]. Currently, the most common visualization state-of-the-art strategy consists in constructing a pyramidal structure composed of different magnifications of the same WSI. Each of these WSI enlargements is then split into small tiles that facilitate access to specific information and each tile is stored in a separate file, while each level of the pyramid (magnification) is stored in a separate folder [3]. This artificial granularity enables applications to fetch only the required tiles, instead of downloading the entire image. However such pyramids present important limitations and disadvantages, i.e., construction, management and uploading of thousand of files that result heavy, wasteful and cumbersome to manage[3]. Furthermore, neither the client cache management nor the communication channel are efficiently administered, for instance operations such as zoom and quality refinement end up by transmitting redundant data and by rapidly overflowing the system memory, as herein demonstrated. In addition, some applications that use these approaches are computationally very demanding, requiring a powerful and expensive infrastructure. During some experiments of the present investigation, the pathologists assessed the v10 stand alone version of Aperio[9] and found out that such application allows interaction with three WSI, but ten of them blocked a standard computer (2.8 GHz quad-core processor and 5 GB RAM memory). Therefore, at least with this version, it would be complicated to serve a large number of concurrent pathologists. Other disadvantage of these pyramidal approaches is related to the particular image formats, most of them use lossy compression formats such as JPEG, an actual issue in medical applications, i.e., Physicians hardly have accepted a compression rate of 2:1 [75]. Other lossless formats, such as BMP or JPEG-LS, are not suitable because of their high storage and transmission costs, as illustrated by the set of experiments herein presented. Unlike these approaches, the system herein introduced takes advantage of the multiresolution nature of the JPEG2000 standard and easily deals with many requests using a single structure. Although JPEG2000 has many advantages regarding data management, the decoding time is longer than other standards and the actual access to the content can be slow [75]. Furthermore, the J2K decoder implementations are partially granular, complicating a selective RoI decoding at the client side. Different approaches have attempted to improve the degree of granularity. On the one hand, some works [128, 33, 161] have developed solutions with a lazy client, that is to say, the client just receives the decoded data to be displayed, but these approaches may increase the server loading when attending several clients. On the other hand, some approaches [75, 60, 59] have opted for tightly-coupled decoder adaptations that enable decoding of specific RoIs, but these solutions can hardly evolve. Other approach [139] aimed to optimally use the channel bandwidth by sending the specific packets of a particular request, while the untransmitted codestream positions were completed with predefined data, but at the price of demanding important processing resour-

ces at the client side when processing large datastreams. In contrast, the proposed system achieves an adaptable and selective decompression at the client side, exploits the potential of the available devices, allows the smart use of advanced cache models and lightens the server loading. Regarding mobile applications, some works have explored interaction with WSIs, concluding that a successful interpretation is feasible from such devices [149, 131]. Nevertheless, some of these applications are just viewers that connect to VM pyramidal systems such as Zoomify or Aperio, inheriting some of the previously mentioned limitations. A mobile tele-radiology imaging system using JPEG2000 was proposed by Kim et al.[85], using low resolution and lossy versions of tomography images. Nonetheless, the authors conclude that the size of mobile devices was not functional because the magnification and details needed for diagnosis require zooming, RoI selection and high quality, characteristics that were not therein offered. On the contrary, the proposed approach achieves a selective lossless RoI visualization on the mobile device and accomplishes random spatial access of RoIs at any magnification and quality.

Different strategies have been proposed to improve interaction with large images. Some have explored JPEG2000 as an interaction tool, speeding up the navigation by implementing cache and prefetching techniques [74, 75, 59]. In this work, a simple and loosely coupled cache technique avoided redundant transmission. Interaction can also be enhanced by selectively compressing the more relevant areas with lossless quality and the rest of the image [11] with some loses. Different works have adapted and implemented wavelet-based [107, 58] and JPEG2000 approaches [11, 151] to ease the access to specific image areas. In the proposed approach the entire WSI was assumed to be relevant and a free lossless navigation was so possible. In this case, the granular RoI was dynamically constructed in real time, i.e., while the user was interacting with the WSI and not during compression. However, some studies have shown that a pathologist need not explore the entire slide, but instead she/he focuses her/his analysis on a few number of visual fields [68, 69]. Recognition of such relevant regions may be a potential source of knowledge for diagnostic tasks, medical training and reduction of computational and transmission costs. Integration of the proposed technique with RoI coding approaches might also be very useful.

Yet the discussed elements strengthen out the proposed approach, the presented system shows some limitations. The main issue is related to the fact that few web applications offer support for the JPEG2000 format, and then external decoders are required. This weakness was herein mitigated by uncoupling the system from the particular decoder and constructing general plug-ins or interfaces that communicate with any decoding standard implementation. Another important limitation is the time for decoding, larger for the proposed approach when compared with the state-of-the-art JPEG pyramidal approach, but still appropriate in terms of a seamless navigation experience. In fact, the authors conducted a short usability evaluation, consisting in 4 questions for the two pathologist that assessed the system: From 1 to 5, what was their opinion about the user friendliness, relevance, response times and functionality. Results showed scores of 4.8, 4.8, 4 and 3.8 for each of the respective items.

Pathologists pointed out that this prototype was very friendly and very useful for academic and clinic environments. Regarding the User Interface and functionality, this was considered as basic since it just displays a RoI at a time, and some extra functions should be added, for instance, interaction with the thumbnail, diagonal panning and zoom in based on a selected point. Likewise, they agree about the response times were appropriated but some work is still required to improve the transition between frames. Finally, yet only two pathologists were part of the evaluation, these experiments conclude a proper performance of the proposed architecture. Future work includes to enhance the GUI design and to perform a deeper usability test and to release stable prototype of the system.

2.2.2. A web-based interactive tool for education and research

Exploration of histopathological slides is a potential source of knowledge and new medical training and research paradigms. Commonly, such slides are explored using an optical microscope to analyze the structure and architecture of the tissue and to provide a diagnosis[110]. However, such physical slides have some issues: they are fragile, hard to transport, deteriorate with time, and may easily be lost. Furthermore, they have limited access because just a single person can examine a sample at any time [123]. Traditionally, pathology education has involved using optical microscopes in laboratory sessions. However, this model has experienced important changes due to curricular reforms that have reduced availability and time of laboratory practices, lack of space, and tools and tendencies towards cooperative work. New pedagogical approaches aim to interpretation and identification of structures in histology samples, more than to acquisition of hand skills such as microscope manipulation [160].

Recently, Digital Pathology has earned interest in the academic community since it solves some of the conventional microscopy issues: virtual information is easily accessible and always available, provides simultaneous access to the same sample, and is not deteriorate with time. In addition, working with digital images permits to annotate, set regions of interest, apply image processing techniques such as segmentation, perform pattern recognition, and use information for diagnosis support [110].

Digital Pathology has a great potential for training, so multiple educative institutions have included it in their curricula, complementing their histopathology courses, with positive results [109, 31, 134, 40, 171]. Inclusion of digital pathology in education has important benefits. First, it reduces costs for institutions since they do not have to invest in buying and maintenance of optical microscopes and physical slides [31, 40, 123]. Second, it offers flexibility to access content at any moment and place, not only during laboratory practices [90, 110]. Third, informatics tools give learners freedom and fast access to explore and study different slides, thereby potentially improving their diagnostic skills [86]. Fourth, virtual microscopy systems provide access to remote users to observe in real time certain slides [32]. They promote a collaborative environment in which learners and experts visualize and

discuss some samples [110, 160]. Fifth, professors can annotate the slides to present them to the class with the possibility of hiding data to evaluate students. These systems can be integrated to other academic systems to extend their functionality, for example, management of tests. Finally, such slides could be incorporated into clinical databases [31, 160].

Despite its strengths, implementation of digital pathology in education has important limitations that have hindered a wide adoption of it [32]. A remarkable issue is the size of virtual slides, that are on the order of gigabytes [123], which demands smart strategies to dynamically and efficiently access and visualize images. Other problem is related to the costs of infrastructure and resource support required to deploy a training system [123, 32]. Lack of standardization is another inconvenience since there is not a common image format for virtual slides, a large number of proprietary or vendor-specific formats has been constantly modified as long as new scanners have been introduced [161]. Finally, some educative systems have been developed as non-interoperable or stand-alone applications, for example, ReportTutor[47], which limits their accessibility and wide usage and does not profit the potential mobile devices offer. Current web and mobile technologies present a great opportunity to develop rich and interactive applications that facilitate access to content from anywhere and at any moment.

This section introduces an accessible, flexible, and low-computational-cost system for digital pathology training and research. The proposed approach exploits the potential of web technologies, integrating an interactive interface, mobile responsive, with a set of modules devised to perform specialized training and research tasks in the pathology field. It takes advantage of the JPEG2000 standard to efficiently deploy WSI at a desired quality and magnification. It also includes an interactive annotation tool for WSIs, to create content for learning, collaborative, and discussion purposes. The system also was devised to serve as a tracking tool, in which navigations over WSIs are registered. In this way, all the registered information can be easily used to feedback the learning and teaching processes, build reference databases, study navigation patterns, analyze intra-observer variability, among others.

System overview

Access and visualization strategy

Given the large size of WSIs, the most common visualization strategy consists in constructing a pyramidal structure composed of different image files, generally in the JPEG format. This pyramid contains magnifications of the same WSI, each split into small tiles to ease random spatial and resolution access [123]. However, as previously stated, this approach presents important limitations regarding management, quality, flexibility and performance [3, 36]. On the other hand, the JPEG2000 standard appears to be a suitable format for WSIs [75] since it offers some features such as support to large file sizes, high compression rates, lossless compression, and granular representation, i.e., a spatial region of an image can be retrieved at different levels of magnification and degrees of quality [153].

In Section 2.2.1, a strategy to efficiently access and display JPEG2000-coded WSIs without requiring a powerful infrastructure was presented. In this section, an adaptation to this model was used to allow users to dynamically interact with WSIs from a web browser (Figure **2-12**). For this purpose, two main applications were deployed.

The first application is a provider of web services; basically, it receives a specific request and sends back the corresponding response. These requests include image metadata, clinical information, and a specific image region at a particular resolution. In the latter case, when the region parameters are received (coordinates and resolution level), an image is generated following the methodology described in Section 2.2.1. In summary, the corresponding JPEG2000 packets needed to construct such a region are computed and extracted. Next, the packets and the corresponding image header are used to build a codestream, which is uncompressed using a standard JPEG2000 decoder. Finally, resulting pixels are sent to the requester application.

The second application is a visualizer and was built based on the OpenSeadragon software[163]. It works as an interface between the user and the provider of services. When a user wants to navigate a particular image, the visualizer creates an empty image pyramid composed of different tiles and scales. The tiles of the pyramid are filled dynamically when the user interacts with the image. For example, if the user moves towards a particular image region to visualize it, the visualizer computes the coordinates and resolution of such a region and sends the corresponding parameters to the provider of services. Then, the provider sends back the corresponding image pixels and they are received by the visualizer, which displays them into the screen.

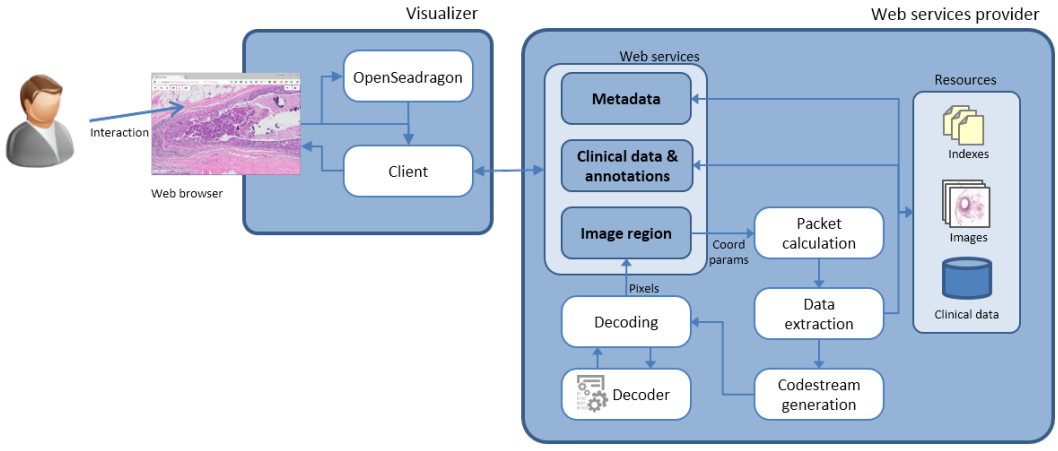
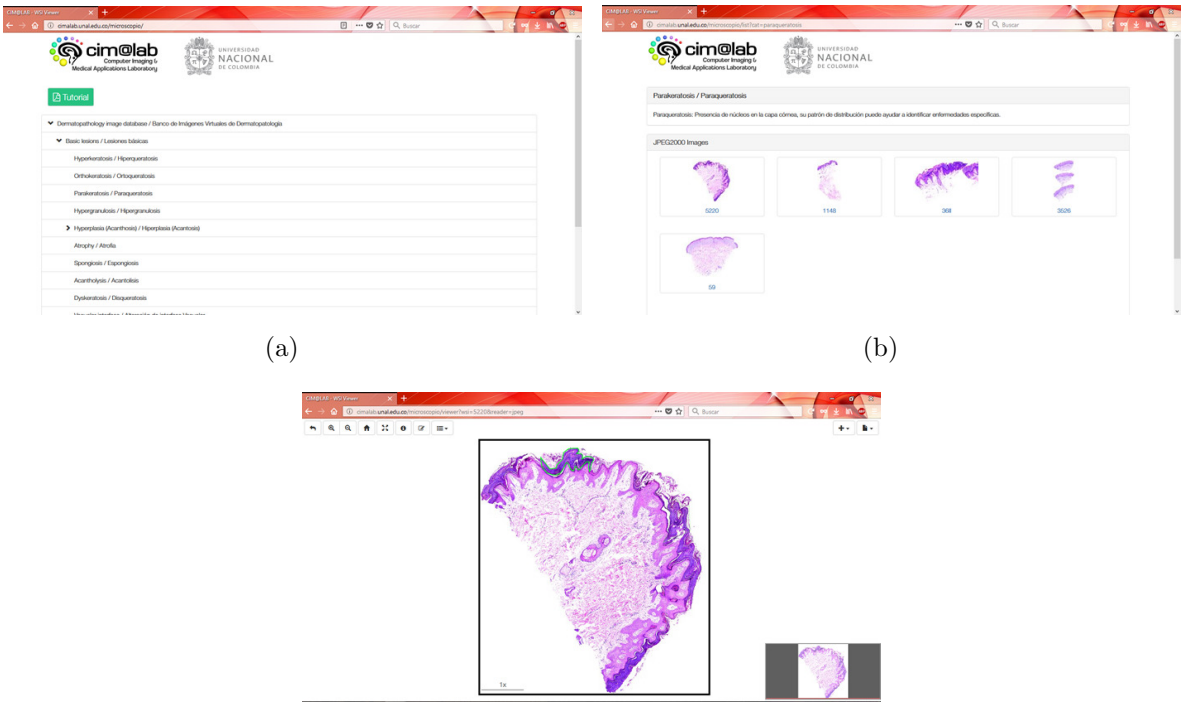


Figure 2-12: Architecture of the web-based interactive tool for visualization of WSIs.

Graphical user interface

Figure **2-13** shows some screen-shots of the user interface. It comprises three main screens. The first screen shows a tree-view of categories and subcategories. When a user selects an

item, the second screen is displayed showing the list of images (and corresponding thumbnail) within the selected category or subcategory. Finally, when user selects one image on the list, the third screen is deployed presenting the image itself. In such a third screen, user can navigate thorough the image by panning or zooming in specific areas for detailed examination.



(a) (b)

(c)

Figure 2-13: Main screens of the graphical user interface of the tool for visualization of WSIs. Subfigure (a) shows a tree-view of categories and subcategories. Subfigure (b) presents the list of images existent into a certain category. Subfigure (c) displays the interface for navigation of the WSI.

The graphical user interface of the system was implemented using the HTML 5, CSS 3, and JavaScript technologies, which ease the development of rich Internet applications. It was designed to be easy to use and interactive. This application runs at any modern browser and does not require to use or to install external plug-ins to work properly, so users can easily access to it from anywhere. In addition, it has responsive design, i.e., the content is adapted according to the user device, thereby enabling mobile access.

Navigation tracking

The presented system was devised to track the navigation movements of the different users while using the application. Specifically, it registers data derived from user interaction such as movements (panning and zoom) and elapsed time. Likewise, it registers points of interest

during navigations that are explicitly set by users by double-clicking. This information can be used for different purposes. For example, educators can visualize the way learners are exploring a WSI to detect problems during diagnostic tasks and to take corrective actions. Similarly, this can serve for students to see the way an expert navigates a slide and learn from that. This information is also very helpful to study navigation patters and identification of Regions of Interest (RoIs), which favors different activities; for example, the development of strategies such as caching or prefetching[2], the design smart tutoring systems which provide feedback that can help improve learner's diagnostic skills [47, 123], the indexation of histopathological databases, the use of selective compression strategies[3].
The system includes an interactive visualization tool that reproduces the registered navigation paths. It sequentially displays each visualized window step-by-step. Additionally, it also displays the marked points of interest on the image during navigation as showed in Figure **2-14**.

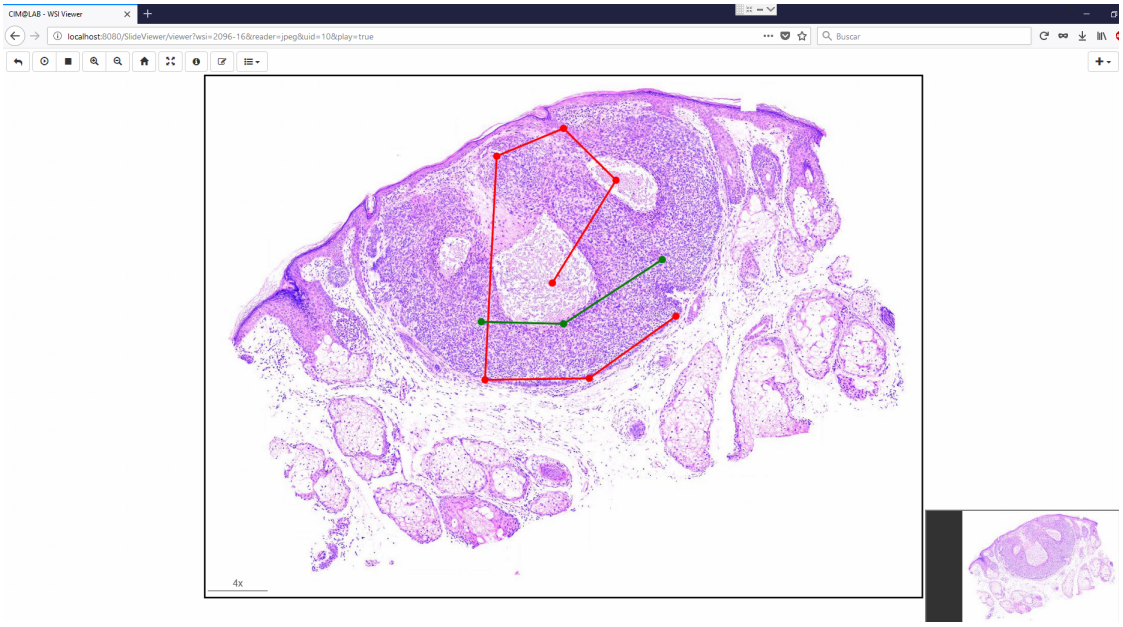


Figure 2-14: View of the navigation tracking tool. When users navigate the WSIs, they can indicate points of diagnostic interest by double-cliking on the image. The tracking tool enables the visualization of both the movements performed by a user and the points of interest he/she marked.

Clinical data and annotation editor

The system includes a panel showing the clinical data of the WSI (clinical record, diagnosis,

[2]Cache is a memory space where information previously processed might be used in the future. Prefetching consists in anticipating the user requirements to make data available before the user requests them
[3]A set of RoIs are explicitly defined to be lossless compressed while irrelevant data is lossy compressed

notes, etc.). Furthermore, it has an annotation editor that allows experts to delineate specific areas of the image with certain pathological meaning and including them some explanatory text. The annotations can be downloaded in JSON format for analytics and information extraction purposes. Figure **2-15** shows an example of an image annotation.

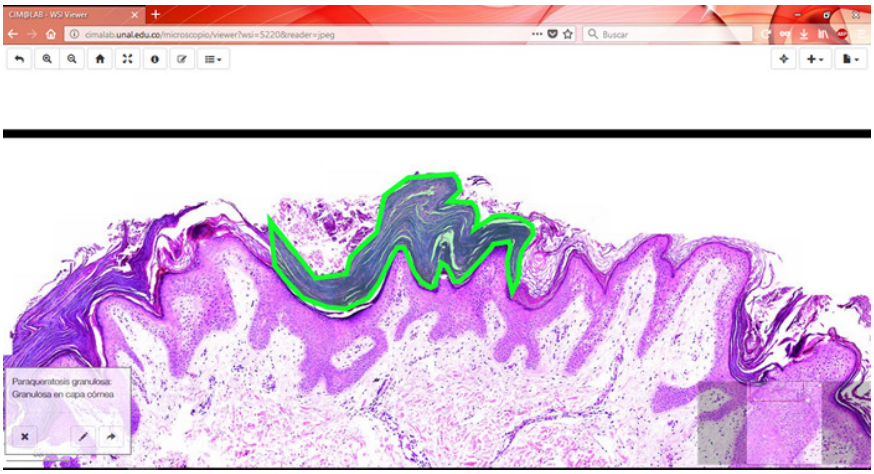


Figure 2-15: Example of an annotation on a WSI. The green polygon delineates a histopathological concept. The bottom-left square shows the corresponding text describing the concept.

2.3. Products

Journal papers

- Germán Corredor, Eduardo Romero, and Marcela Iregui. *An adaptable navigation strategy for Virtual Microscopy from mobile platforms*. Journal of Biomedical Informatics, vol 54, pp 39-49, 2015

Conference papers

- Germán Corredor, Marcela Iregui, Viviana Arias, and Eduardo Romero. *Flexible architecture for streaming and visualization of large virtual microscopy images*. Proc. of the MICCAI - Workshop on Medical Computer Vision 2013, Nagoya - Japan, 2013

- Germán Corredor, Marcela Iregui, Viviana Arias, and Eduardo Romero. *Concurrent access to a virtual microscope using a web service oriented architecture*. Proc. of the IX International Seminar on Medical Information Processing and Analysis, Mexico City - Mexico, 2013

Indirect products

- **Undergraduate thesis:** Darwin Díaz. Advisors: Germán Corredor, Eduardo Romero, and Angel Cruz-Roa. *Plataforma web de telepatología para la navegación eficiente de láminas virtuales de histopatología como apoyo a la enseñanza, investigación y trabajo colaborativo en cáncer*. Universidad de los Llanos, Villavicencio - Colombia, 2018
- **Colombian journal:** Sylvia Hernández, Germán Corredor, and Marcela Iregui. *Modelo de interacción para la navegación en imágenes panorámicas o de gran tamaño en Dispositivos Móviles de pantalla táctil*. Scientia et Technica, vol 19(3), pp. 290-298, 2014
- Sylvia Hernández, Germán Corredor, and Marcela Iregui. *Interaction Model for Large-Image Navigation in Mobile Devices*. Proc. of the 5th Latin American Conference on Networked and Electronic Media, Manizales - Colombia, 2013
- Paula Toro, Germán Corredor, Eduardo Romero, and Viviana Arias. *Presentation of an annotated whole-Slide image database of Dermatopathology for Postgraduate Training in Pathology and Dermatology*. Proc. of the 5th Digital Pathology Congress, London - UK, 2017
- Paula Toro, Germán Corredor, Eduardo Romero, and Viviana Arias. *Web application for pathology training using a low cost platform and annotated whole slide images*. Accepted for presentation at the 14th European Congress on Digital Pathology and the 5th Nordic Symposium on Digital Pathology, Helsinki - Finland, 2018

Software

- Web application for exploration of histopathologic slides.
  URL: http://cimalab.unal.edu.co/microscopio/

3 Histopathological analytics by learning from imaging data

3.1. Introduction

Large quantity of medical images are being produced daily in hospitals and medical centers for diagnosis and therapy [118, 27, 42]. In consequence, health care organizations need systems and technologies that enable health professionals to efficiently manage, access, and share such data [181]. These systems should allow experts to find content into those large repositories and use it to assist them in different manners, for example, to support their decisions by providing quantitative information or to make comparison between similar images in case of differential diagnosis [27, 99].
In most of the current medical imaging systems, access and retrieval of content is achieved by means of text, i.e., images are annotated by keywords or descriptive text and organized by topical or semantic hierarchies in traditional Database management systems. When a user wants to access to content, he/she inputs keywords that are compared to the text marked in the images and results are retrieved based on some similarity criteria [27, 132]. However, as the databases grow, the traditional keywords based methods to retrieve a particular image becomes inefficient and suffers from the following limitations:

1. Lot of manual labor. Manual annotation is an expensive and time-consuming procedure that is also highly prone to errors. Furthermore, it should only be performed by expert people to guarantee the quality of data, which is completely unrealistic since the whole content volume of very large image databases is beyond the manual indexing capability of human experts [39, 101, 27, 180, 132, 169].

2. Inaccuracy of the annotation due to the human subjectivity. A single image could represent different things to different persons and it could be hard to describe the diversity and ambiguity of image contents [101, 169].

3. The keywords increase linguistic barrier to share image data globally [168].

According to Müller et al. [118], the goal of medical information systems is "To deliver the needed information at the right time, the right place, to the right persons in order to improve the quality and efficiency of care processes". Nonetheless, the simple text-based retrieval is

very far from such a goal. Consequently, there is a strong need for efficient and effective strategies to interact with relevant content from such image repositories.
In response to such problematic, a strategy called Content-based Image Retrieval (CBIR) has been proposed. CIBR systems have been widely used to search into large image databases (medical, satellite, artistic, among others) and their objective is to aid users to retrieve relevant and objective information [130]. In general, CBIR enables image indexing by visual features and image retrieval by similarity of such features [102]. These characteristics include, but are not limited to, color, shape, borders, and texture. Several techniques have been used to extract visual image features. For example, extraction of color has been performed by color histograms, covariance matrices, coherence vectors, among others [27, 101]. Textures have been extracted using strategies such as co-ocurrence matrices, Tamura features, Wavelet transform coefficients, and Gabor filters [118, 101, 27]. Finally, some shape descriptors commonly used are Fourier descriptors, Turning functions, Beam angle statistics, Zernike moments, Generalized complex moments, Morphological descriptors, etc [27].
Figure **3-1** illustrates the operation of a typical CBIR system. Generally, CBIR systems operate in two phases. The first is an off-line step in which information is pre-processed and the indexes are created. In this stage, the visual contents of each image in the database is extracted, a set of characteristic features (a multidimensional feature vector) computed using a feature extraction process. This feature vector is finally stored in a metadata repository. The second phase is presented when a user performs a search. In this stage, the user inputs a query, commonly an example image, which is converted into a feature vector using an online feature extraction process. Then, the similarity between the feature vector of the usersś query and the feature metadata items is calculated and ranked. Finally, retrieval is performed by applying an indexing scheme that can be used to support fast retrieval and to make the system scalable to large image databases [27].

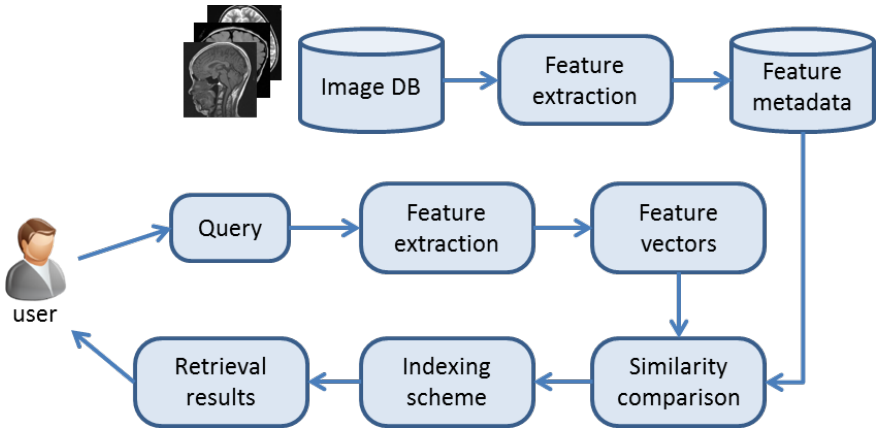
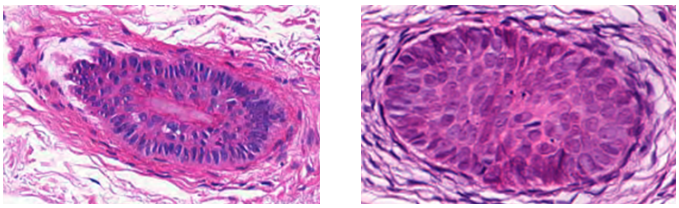


Figure 3-1: Diagram of a typical CBIR system (adapted from [27]).

While research in this field has been productive in recent years, there are different challenges to address to improve the performance of CBIR systems and make them a powerful tool to

support different professional tasks. One of these challenges is the curse of dimensionality, which refers to the problem caused by the exponential increase in volume associated with adding extra dimensions to Euclidean space [83]. In addition, CIBR systems must deal with the semantic gap, defined as the difficulty of describing high-level content by means of low-level features [101]. Although visual features could aid to find similar cases, similar visual features may not imply similar diagnoses or symptoms. Figure **3-2** illustrates a situation in which two images are visually similar (shape, texture, and color), but they have a very different meaning. While image (a) is a hair follicle, a normal structure, image (b) is a nodule of basal cell carcinoma, a cancerous structure. Although an expert can easily identify and differentiate such structures, this is not an easy task for a computer program that only uses information coming from visual primitives such as color, texture, orientation, or shape.



(a) (b)

Figure 3-2: Two visually similar structures obtained from skin samples. Image (a) is a hair follicle, a normal structure. Image (b) is a basal cell carcinoma nodule, a cancerous structure.

In this chapter, we present different strategies to automatically identify pathological structures and infer prognostic data from H&E stained images. These approaches aim to provide pathologists with objective and quantitative information that can support their decision making. The proposed strategies start by reducing the dimension of data, so instead of using the information of all the image pixels, the analysis is focused on image cell nuclei. This decision has two main reasons:

1. The cell is the basic unit of pathology. Modern pathology is the study of cellular abnormalities, i.e., pathologists understand diseases in the context of normal cellular structure and function [91].

2. Due to the high dimensionality of medical images, extraction of relevant content could be an intractable computational problem. For this reason, strategies for dimensionality reduction should be applied in such a way that there is an equilibrium between amount and utility of data.

Considering that pathologists do not pay attention to cells individually, but they examine their context and their relationships with other cells, the hereafter presented strategies exploit local and contextual information of nuclei.

In the subsequent subsections, two study cases are described. The former illustrates an approach that uses nuclear information for automatic identification and grading of ductal carcinoma in situ. The latter presents a strategy in which visual and topological features of lymphocyte nuclei are exploited to predict patient prognosis in cases of non-small cell lung cancer.

3.2. Study case: Automatic detection and grading of Ductal Carcinoma In Situ

Breast cancer comprises several kinds of lesions with different severity grades. From such lesions, ductal carcinoma in situ (DCIS) is the most common non-invasive breast cancer type. In this case, tumor cells are still located in the origin tissue (the milk ducts) and have not spread into any surrounding tissue[21]. Although DCIS is not life-threatening, it is synonymous of a high risk of developing invasive carcinoma and these patients may require additional surveillance, prevention, or treatment to reduce their risks. Early detection results then crucial in these cases[48]. Unfortunately, detection of DCIS is challenging since this cancer type is observed as a set of lesions with highly variable morphology, biomarker expression, genomic profile, and natural progression[16] (See Figure **3-3**). Usually, DCIS is categorized into three grades: low, moderate, and high. Low-grade lesions contain cancerous cells that look very similar to normal or atypical ductal hyperplasic cells. Moderate-grade lesions contain cancerous cells slightly different to normal cells. Finally, high-grade DCIS is characterized by well-differentiated and fast growing cancerous cells [21]. Previous studies have revealed low levels of agreement among experts when analyzing DCIS lesions[48, 16], a definite issue in the clinical practice. Misclassification of breast lesions may lead to over/under treatments of lesions identified during breast screening[48]. In this context, automatic measures may contribute for discrimination between breast lesions.

This section presents an automatic strategy that classifies microscopic Field of Views (FoVs) extracted from breast histology images into two classes: DCIS and non-DCIS. Furthermore, this strategy also classifies DCIS lesions into 3 grades. The presented approach characterizes each FoV using different nuclear features and their context. Each nucleus is threefold characterized by its own morphological properties (size, shape, color, texture, etc.), by its neighbor nuclei features within a determined radius, and a distance to other image nuclei. Unlike other state-of-the-art methods, any feature in this approach exploits nuclei relative information, i.e., each nucleus is not only characterized by its own information but also by how that nucleus feature is with respect to the surrounding nuclei. This method shows a good classification performance while it shows fast training times and needs no large annotated datasets.

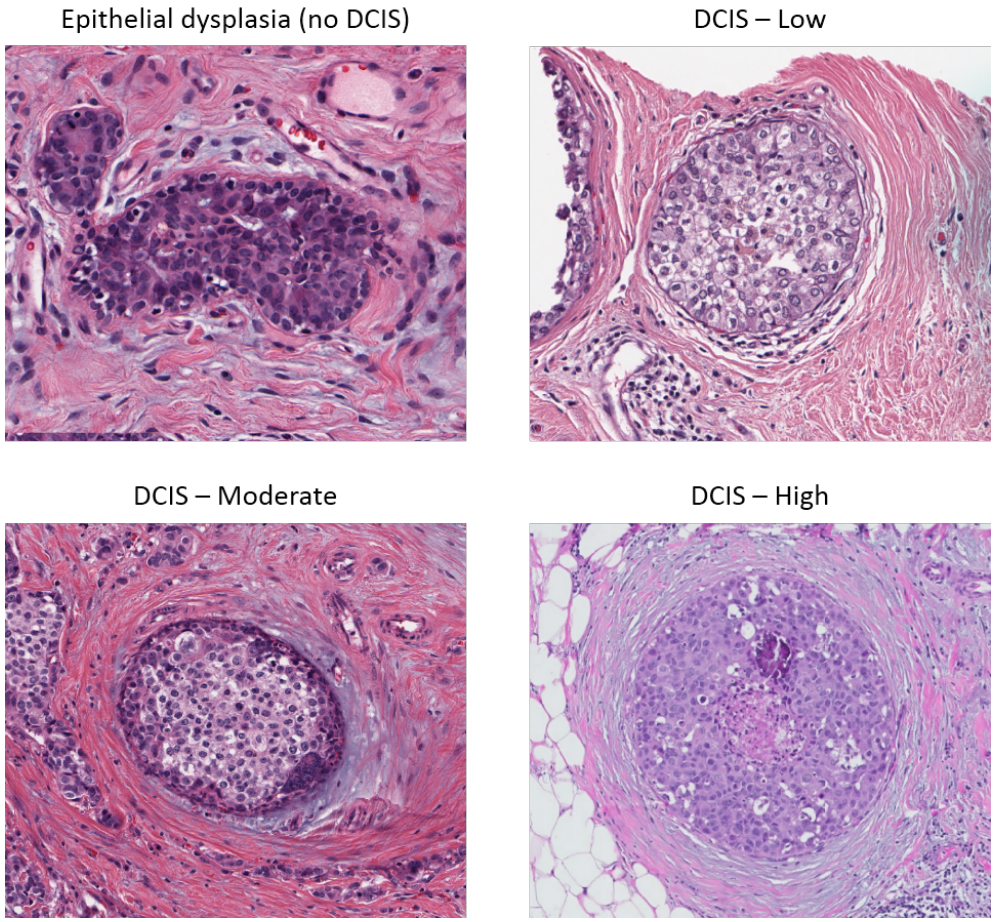


Figure 3-3: Illustration of three different breast lesions. The first corresponds to an epithelial dysplasia (non-cancerous) while the remaining 3 correspond to different DCIS grades.

3.2.1. Extraction of features

The pathology semiology is based on identifying abnormalities in terms of color, shapes, sizes, textures, and spatial arrangement of present structures at different scales[91, 164]. Inspired by this observation, the underlying idea behind the present work is that after nuclei are automatically identified from breast tissue images, different nuclei-based features are used to characterize such images.

Nuclei segmentation

Automatic nuclei segmentation is performed by a watershed-based algorithm[165]. This method applies a set of mathematical operations at different scales to identify candidate nuclei in H&E stained images. This method was selected by its visual efficacy, simplicity, speed, and absence of adjustable parameters.

Local features and regional features

Previous works[48] have shown that nuclear morphological features are useful to characterize DCIS. For this reason, after nuclei are segmented, a set of low-level features are extracted from them, including characteristics of shape (Zernike moments, ratio between axes, etc.), texture (Haralick, entropy, etc.), and color (mean intensity, mean red, etc.). This set of local features is used to characterize each nucleus.
Besides nuclear local features, pathologists also examine the nuclei context/neighborhood looking for particular patterns. Different approaches have used graph-based techniques to characterize nuclei neighborhoods[14, 6, 94]; however, these features only take into account the spatial distribution and ignore the variability of other features. For this reason, for each image nucleus, a set of circles with incremental radii of k=dL*10,dL*20,dL*30 pixels were placed at the nucleus center (dL=20 pixels, the average diameter of all the detected nuclei). Different radii were used aiming to model a multi-scale approach. Finally, a set of regional features are computed within each circle and used to characterize the nucleus. These features aim to measure the neighborhood density and variations in color, shape, and texture.
Once nuclei are characterized, each feature (local and regional) is represented by a histogram to characterize the FoV. For this purpose, the dynamic feature range is set between the maximum and minimum values found along the whole set of FoVs. The dynamic range is divided into ten intervals and the bins of the histogram are constructed as the number of occurrences within each of the particular intervals. The final histogram is normalized thereby obtaining a probability distribution function.

Cellularity features

Since cancer is characterized by an uncontrolled proliferation of cells, features related to the number of nuclei and their grouping grade were also included. This grouping index was computed as follows: a fully-connected graph is built using nuclei as nodes and the inverse of the Euclidean distance between nuclei is set as edge weights. The grouping factor for each node is computed as the sum of the weights of every edge connected to such node[36]. A high/low value means that such particular nucleus is close/far to other nuclei. Finally, the FoV was also characterized by different statistical measures of such grouping grade (mean, median, mode, etc.)

3.2.2. Experimentation

Dataset

A group of 1102 FoVs (1024x1024 pixels) were extracted from a set of H&E breast histology samples from 28 different patients. The cases were acquired from Indiana Hospital and scanned into WSIs at University Hospitals in Cleveland - Ohio using Aperio and Philips scanners. The FoVs were automatically extracted from a set of manual annotations carried

out by an expert pathologist. The final set comprised 400 non-DCIS, 106 low-grade, 251 moderate-grade, and 345 high-grade FoVs.

Experimental setup

Two experiments were carried out. The first experiment attempted to classify between DCIS and non-DCIS. A 10-fold cross validation scheme was used. At each iteration, 70 % of the whole set of FoVs was used to train a Gradient Boosted Regression Trees classifier[15] and 30 % was used to test the trained model. Finally, the measured performances along the 10 iterations were averaged. The second experiment aimed to distinguish between the different DCIS grades; this experiment followed a methodology similar to the first experiment, but in this case only the images labeled by the pathologist as DCIS were used. The presented method was compared with two approaches: The former uses only morphological features and the latter combines morphological and graph-based (Voronoi, Delaunay, etc.) features[14].

Results

Figure **3-4** illustrates the average receiver operator characteristic (ROC) curves of the predictions using the three different approaches, and Table **3-1** presents the corresponding accuracies and f-measures. Results show that the presented approach outperforms the baseline approaches in all the tested scenarios.

		Morph.	**Morph. + Graphs**	**Ours**
DCIS vs non-DCIS	Acc.	0.87 +/- 0.01	0.89 +/- 0.01	0.95 +/- 0.01
	F-meas.	0.83 +/- 0.01	0.85 +/- 0.01	0.93 +/- 0.01
Low vs Moderate, High	Acc.	0.87 +/- 0.02	0.88 +/- 0.02	0.91 +/- 0.02
	F-meas.	0.41 +/- 0.10	0.45 +/- 0.06	0.58 +/- 0.12
Moderate vs Low, High	Acc.	0.68 +/- 0.02	0.76 +/- 0.02	0.88 +/- 0.02
	F-meas.	0.49 +/- 0.03	0.63 +/- 0.04	0.82 +/- 0.03
High vs Low, Moderate	Acc.	0.65 +/- 0.03	0.78 +/- 0.01	0.94 +/- 0.01
	F-meas.	0.65 +/- 0.03	0.77 +/- 0.01	0.94 +/- 0.01

Table 3-1: Accuracies and F-measures measured for the tested approaches. First column presents the results of an approach that classifies FoVs based just on morphological features. Second column shows the results of a strategy that uses the morphological and graph-based features reported in[14]. The third column presents the results of the introduced approach that uses local, regional, and cellularity features.

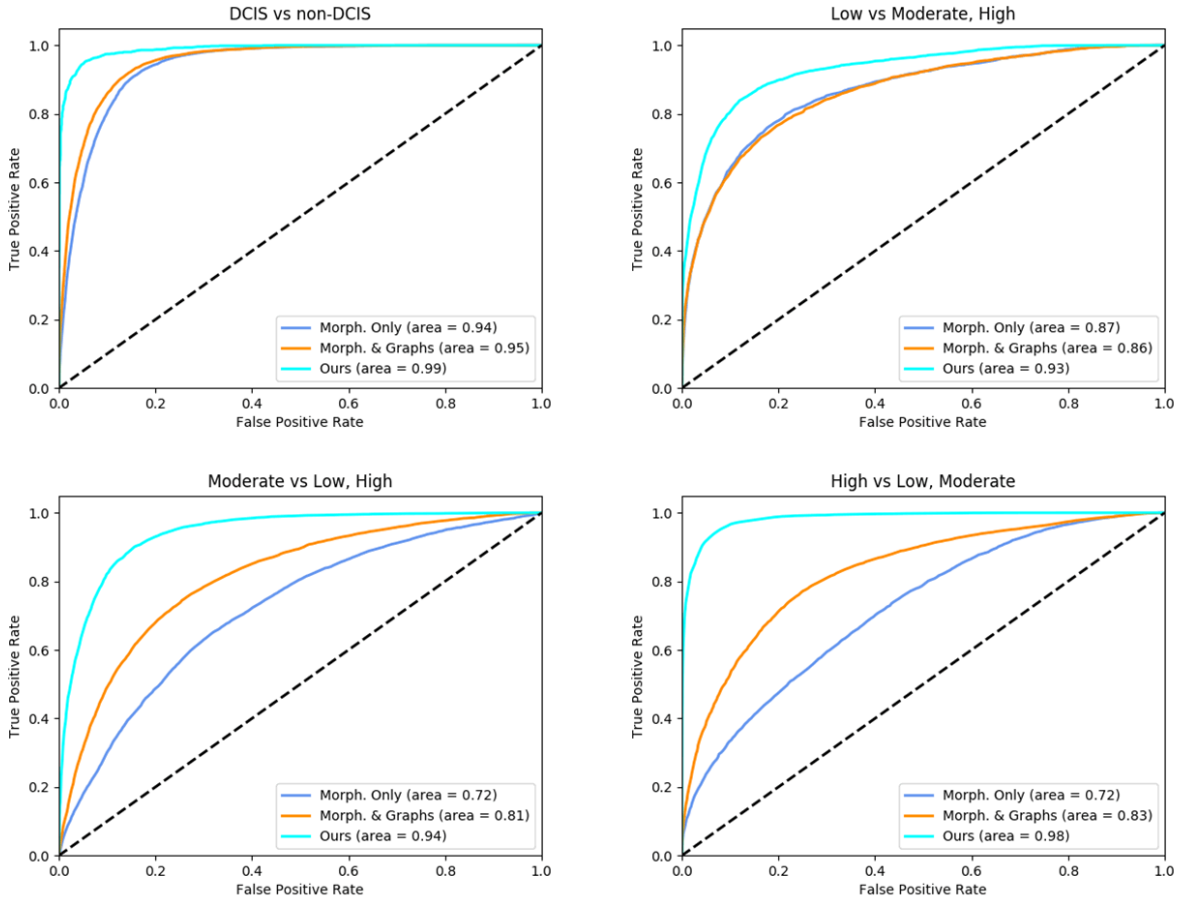


Figure 3-4: Average ROC curves for correctly identifying DCIS and DCIS grades in the test set of FOVs using the introduced approach and the comparative strategies (Morphological only features and morphological + graph-based features). To generate an adequate precision of the ROC curves, 100 repeats of 10-fold cross-validation were run.

3.3. Study case: Association between Tumor-Infiltrating Lymphocytes and prognosis in patients with lung cancer

Early-stage (stages I and II) non-small cell lung cancer (NSCLC) [57, 22] is typically treated with complete surgical resection of the tumor. However, even after the entire resection of the tumor, 30-55 % of patients develop disease recurrence within the first 5 years of surgery [162]. The ability to identify patients at highest risk of recurrence could help to identify those cases who may gain maximum benefit with adjuvant chemotherapy following surgery. Lung cancer histopathology is characterized by a complex interplay of cancer nuclei, immu-

ne cells (lymphocytes and plasma cells), fibroblasts, and pericytes/endothelial cells. Recent evidence suggests the interaction of cancer nuclei with immune cells is associated with likelihood of disease progression, and also influences tumor development, invasion, metastasis, and patient outcome [57, 22]. Recently, several independent studies [143, 43, 20, 22, 100] have shown association between patient survival, disease prognosis, and treatment response with an increased density of Tumor-Infiltrating Lymphocytes (TILs) in NSCLC and other cancer types. Additionally, there is substantial recent evidence that suggests that the density of TILs is associated with the chemotherapy response for a variety of different cancer types [24, 167, 87, 82, 100].

Unfortunately, despite the reported evidence [143, 43, 20, 22] showing a high correlation between density of TILs and prognosis of NSCLC, substantial inter-reader variability at estimating TIL density has meant that TIL count is not routinely employed in the clinic as a prognostic marker of outcome for NSCLC. Brambilla et al. [20], for instance, determined that inter-reader agreement between two pathologists was at best moderate (Kappa = 0.59). While attempts have been made to formalize guidelines for TIL grading in the context of breast cancer [140], these efforts have been lagging in the context of lung cancer.

Over the last few years there has been substantial interest in developing automated nuclear segmentation and detection algorithms for identifying and quantifying the extent of TILs from routine Hematoxylin and Eosin (H&E) pathology images [7, 136, 49]. Detection and segmentation of TILs have been carried out using different strategies, such as extracting visual features [7], applying morphological operations [13, 50], or using Deep Learning models [78]. Similarly, different works have employed different methods to automatically approach the density of TILs. In [7], authors determine whether lymphocyte density is high or low depending on whether the number of lymphocytes is higher or lower than the average of lymphocytes for the study population. Other approaches [136, 49] have used specialized software suites that compute a density estimation from automatic detection of lymphocytes.

Likewise, beyond just TIL density, there has also been recent interest in looking at spatial patterns of TILs with and their relation with the disease outcome and prognosis. Multiplexed quantitative fluorescence (QIF) and immunohistochemistry (IHC) based methods have been employed for objectively identifying TIL subtypes and some attempts have tried to correlate the spatial arragement and density of these TIL subtypes with the disease NSCLC outcome [143, 12, 100]. For instance, Schalper et al. [143] found out that increased levels of CD3 and CD8 TILs were associated with improved disease outcome. Similarly, Barua et al. [12] showed that spatial interplay between tumor and regulatory T cells was associated with overall survival in NSCLC. Furthermore, Liu et al. [100] demonstrated that presence of CD8+ and FOXP3+ TILs is correlated to response to platinum-based neoadjuvant chemotherapy in advanced NSCLC.

Interestingly, recent evidence suggests computer extracted spatial patterns and morphologic attributes of TILs from routine H&E images also appear to be associated with prognostic outcome of disease. Basavanhally et al. [13] explored the use of graph network algorithms to

spatially characterize the arrangement of machine identified TILs in HER2+ breast cancer H&E images in order to predict the TIL grade (i.e. high or low). Yu et al. [179] and Luo et al. [104] extracted a number of quantitative morphological nuclei and surrounding cytoplasm features from H&E tissue images of early-stage NSCLC patients (e.g., area, shape, intensity, texture, density, etc.) for predicting survival. In addition, Wang et al. [170] found that nuclear architecture features added prognostic value to nuclear shape and texture features for predicting early versus late recurrence in early-stage NSCLC. In [79], Saltz et al. used a deep learning model to identify patches of TILs in images, which were clustered using different similarity metrics. From such clusters, several indices were obtained and their correlation with patient survival was studied in different cancers types.

Given the evidence demonstrating the importance of TILs for prognostics tasks, in this section, we study the correlation between density and spatial arrangement of TILs and patient outcome, specifically disease recurrence in early-stage NSCLC. First, we introduce a simple and effective automatic approach for detecting lymphocytes on H&E images. Once we have identified lymphocytes, we proceed to extract information from their density and spatial arrangement, and then we analyze the correlation of such information with the patient prognosis. Finally, we explore a strategy that groups TILs based on their contextual information and then studied the association of the identified TIL families with prognosis.

3.3.1. Identification of lymphocytes on histopathological images

As previously stated, automatic detection and quantification of lymphocyte infiltration could potentially develop image based prognostic tools [140]. Thanks to the advances in computer vision and image analysis, a TIL biomarker for cancer prognosis is a real possibility [7]. However, lymphocyte segmentation in Hematoxylin and Eosin (H&E) stained histopathology images is not an easy task because of the similar appearance between lymphocyte nuclei and other structures (See Figure **3-5**). Additional challenges include biological variability, histological artifacts, and high prevalence of overlapping objects [13]. An automated lymphocyte detection algorithm has to be able to deal with these challenges so that an objective and precise infiltration grade can be measured.

Different complex approaches have attempted to address the issue of detecting lymphocytes by means of visual features. Basavanhally et al. [13] presented a scheme that starts by combining a region growing algorithm with a maximum *a posteriori* estimation and Markov random fields to identify lymphocytes. In another approach [92], a first set of candidates were obtained by applying different operations such as sigmoid contrast enhancement, conditional hole filling, adaptive active contour, and extraction of Haralick features. These candidates were then classified as lymphocytes or non-lymphocytes using the scale-invariant feature transform (SIFT) algorithm. In a different approach [7], the authors used the Visilog software, an image analysis suite, to perform an automatic detection of lymphocytes based on a three-step process: detection based on color thresholding, boundary definition with a watershed

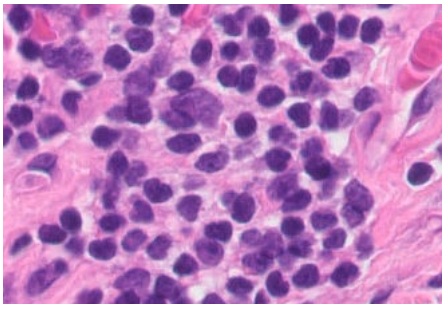


Figure 3-5: A patch extracted from an H&E stained lung cancer image. Figure illustrates the prevalence of overlapping objects and similar appearance between lymphocyte nuclei and other structures.

approach, and size based filtering of the detected elements.

In recent years, Deep Learning (DL) methods have become popular since they have shown outstanding performance in different computer vision and pattern recognition tasks [138]. DL architectures are constituted by multiple layers, containing linear and non-linear data transformations, aiming to construct a reduced and meaningful representation. In the digital pathology domain, DL models have been used for different tasks such as mitosis identification and localization of regions of interest in histological images [138]. However, DL models are highly demanding of powerful hardware infrastructures and large sets of annotated data for training, requirements hard to meet in real medical scenarios.

This subsection presents a framework that takes advantage of some discriminating visual features of lymphocytes in order to identify them in patches extracted from lung cancer Whole Slide Images (WSIs). It starts by detecting nuclei by a watershed-based nuclei segmentation algorithm. A main goal of this work was to define a simple but discriminating set of visual features to discriminate lymphocytes from other structures; for this reason, some shape, texture, and color features were carefully selected and extracted from each segmented nucleus. Finally, a Support Vector Machine (SVM) classifies each nucleus as either lymphocyte or non-lymphocyte. An SVM was used because it showed a higher performance compared with other classifiers, namely random forest, linear discriminant analysis, and quadratic discriminant analysis. An important advantage of this approach, compared to other state-of-the-art methods, is its simplicity since it is very easy to use and implement; besides, it presents a good performance, fast training times, and accurate results. This strategy is compared with a second approach constructed upon a Deep Learning model that was trained to identify lymphocytes. This model receives as input, patches from WSIs and outputs probability maps showing pixels that are likely lymphocytes; then, a threshold is applied to such map and a watershed-based algorithm segments and splits the connected/overlaid cases to set the final candidates.

Materials and methods

Dataset

13 Fields of View (FoV) of 1000×1000 pixels at $\times 40$ were extracted from a set of H&E WSIs from 5 patients diagnosed with lung cancer. These FoV were chosen as containing both lymphocytes and tumoral cells. The WSIs were provided by the Pathology Department of University Hospitals in Cleveland-Ohio, USA. FoVs with lymphocytes were manually annotated, validated, and approved by a pathologist and used to train the different models. A total of 3420 structures were annotated, being 2352 lymphocytes while the remaining 1068 corresponded to other structures (e.g., cancerous nuclei). The study cohort was limited to 13 FoVs owing to the arduous work required to annotate the images and the time constraints of the pathologist who validated annotations. This classification task is challenging since lymphocytes and cancerous cells are very alike and annotations have to be checked out several times.

Feature selection

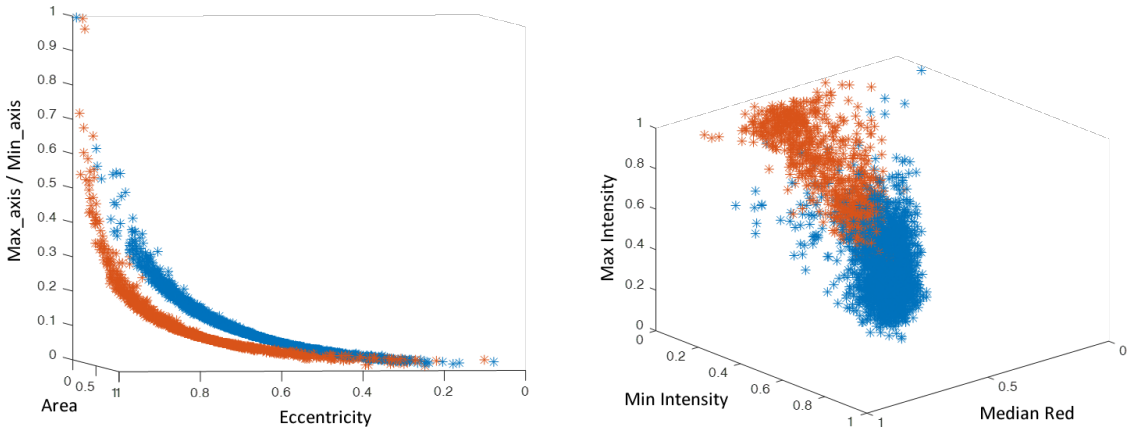
In H&E images, lymphocyte nuclei are generally distinguished from other cell nuclei by their smaller size, more circular shape, and a darker homogeneous staining [92, 13], thereby suggesting that visual features may provide a good clue to differentiate lymphocytes from other structures. For a pathologist, the color is one of the most discriminating characteristics of lymphocytes. This is why a main feature was the median of the red r, g, b channel of the segmented nucleus as well as the minimum and maximum of the luminance channel. Additionally, taking into account that lymphocyte texture is more homogeneous than other nuclei since this cell has a differentiated nucleolus, the entropy was also selected. Finally, provided that shape and size of lymphocytes are usually different from other structures, area, eccentricity, and major and minor axes ratios were also chosen. Figure **3-6** presents some feature spaces illustrating the discriminating properties of the selected features. The features selected corresponded to representative color and shape characteristics.

Model training

First, a normalization process, a color-based technique commonly used to compensate the different color variations on histopathology images, is applied to input images. Then, the previously described features are extracted from the annotated nuclei. Finally, a Support Vector Machine (SVM) with a linear kernel was trained to classify each structure as a lymphocyte or non-lymphocyte. Training takes about 8.4 seconds on a laptop computer with 8 GB RAM and an Intel Core i7 3.1 GHz processor. The training accuracy, i.e., accuracy of the model using the same training set for validation, of this approach was 95.99 %.

Model testing

Since manual annotation of nuclei is a time-consuming and unrealistic task for sets of WSI,



(a) (b)

Figure 3-6: Feature spaces of some selected visual characteristics. Blue dots represent lymphocytes and orange non-lymphocytes. 3D plots show how discriminating are the selected features.

a nuclei segmentation algorithm first detects nuclei candidates and then the trained SVM model classified such nuclei as either lymphocytes or non-lymphocytes.
Nuclei segmentation is a very challenging task and is out of the scope of the present article; for this reason, a watershed-based method [165, 176] was used by its documented advantages, namely simplicity, speed, and absence of adjustable parameters. This method starts by un-mixing the color images followed by morphological operations. A marker-controlled watershed segmentation is then applied at multiple scales and with different markers; next, a post-processing for rejection of false regions is performed; finally, results from multiple scales are merged.
Once nuclei are segmented, the input images are color-normalized and the previously mentioned visual features are extracted. Based on such extracted features, the SVM model defines whether a nucleus is lymphocyte or not. Nuclei segmentation takes about 4.8 minutes and feature extraction about 9.4 seconds on the previously mentioned laptop computer.

Comparative Strategy

It is well-known that DL models are able to automatically determine relevant features from a set of samples. For this reason, to train a model, a set of pixels is randomly selected from the manually annotated nuclei and lymphocytes. A set of patches (size 32×32), centered on the selected pixels, is extracted from the images. A single patch might then contain none, one, or multiple nuclei. Finally, each patch is labeled as positive if its center pixel corresponds to a lymphocyte and negative otherwise. In order to increase the number of training samples, different transformations (e.g., rotation and mirroring) were applied to the patches. Finally, a set of 213436 patches is obtained (144046 patches correspond to the positive class).
The annotated patches are then used to train a Deep Neural Network classifier. The architec-

ture selected for this task is the well-known CIFAR-10 AlexNet network schema, which was previously used for automatic detection of lymphocytes [78]. This architecture is composed of three identical blocks (See Figure **3-7**). Each block is composed of a Convolution Neural Layer (CNL), a Rectifier Linear Unit, and a Maximum Pool operator. Each CNL corresponds to a set of convolutional filters learned from the lymphocyte and non-lymphocyte classes. Finally, two fully connected layers yield the probability of representing the membership of each nucleus to the lymphocyte class. The network parameters (size of kernels, number of feature maps, loss function, etc.) were the same previously reported in the literature [78]. Training of this model takes about 3 hours using a server with 32 GB RAM, 8 2.5 GHz CPUs, and 3 GPUs: a GeForce GTX 660, a GeForce GTX Titan, and a Tesla K20c.

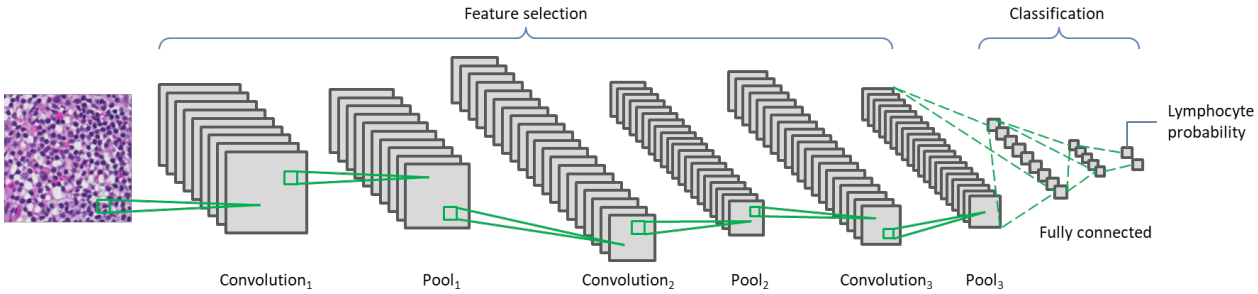
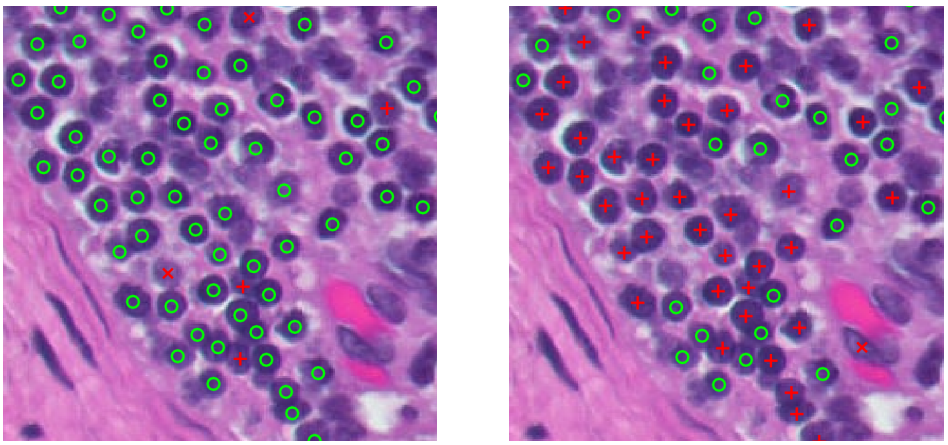


Figure 3-7: Deep learning architecture used to classify nuclei as either lymphocytes or non-lymphocytes. The illustration is based on the Figure presented by Romo et al. [138].

During testing, patches extracted from WSIs are used as input of the DL trained model. During this process, the membership of each patch to the lymphocyte class is determined. Finally, neural network outputs a probability map highlighting the pixels that likely correspond to lymphocytes. Different image processing operations could be applied to such maps to determine the final candidates. In this case, a simple threshold (128) was applied; finally, a watershed-based algorithm was used to segment and split the connected/overlaid cases.

Experimental results

Validation of the presented approaches was carried out using the previously described dataset (See Section 3.3.3). Automatic lymphocyte detection was graded by the number of lymphocyte centroids that correctly overlapped, judged as correct when centroids were within one nuclear radius. Figure **3-8** shows some visual results. These results show that DL approach is missing an important quantity of lymphocytes; in contrast, the presented framework detects most of the lymphocytes, even the overlapped/adjacent structures.
Quantitative results, presented in Table **3-2**, show that the Deep Learning approach has a very high precision (95.29), but a very low recall (39.60). In contrast, the presented framework shows good precision and recall metrics (89.12 and 83.57, respectively). Furthermore, the F-score for the presented approach outperforms by about 30 % the DL approach.



(a) (b)

Figure 3-8: Visual results of the lymphocyte detection process. Green circles represent lymphocyte hits, red plus signs represent missed lymphocytes, and red crosses represent false positives and false negatives. Subfigure (a) shows results using the presented simplified framework and Subfigure (b) presents results of the DL approach.

	Precision	**Recall**	**F-score**	**NAL**	**Acc. NAL**
Presented approach	89.12	83.57	86.31	621	64.5
Deep learning	95.29	39.60	55.95	2018	28.2

Table 3-2: Performance metrics for identifying lymphocytes on patches of lung cancer images. NAL stands for number of detected lymphocytes absent in manual annotations and Acc. NAL is the accuracy of the prediction for 10 % of NAL.

During testing, both approaches (DL and visual features) detected lymphocytes that were absent from the manual annotations (See Table **3-2**, column Non annotated lymp.). For this reason, 10 % of these lymphocytes were randomly selected and manually labeled by a pathologist to test the accuracy of such prediction. After this process, 28.2 % of the lymphocytes detected by the DL model were accurate and 64.5 % of the detections of the presented framework were true positives.

Our results show that the balance between precision and recall for the DL model is poor, being outperformed for the simple approach based on visual features (∼30 % higher F-score). While DL models have shown interesting results in different computer vision tasks, they have important limitations, for example, their need of large amounts of annotated data; more data means more annotations. Another limitation is related to training times, that are quite longer than using the presented simplified approach. This is an important constraint because it meaningfully limits model updating, i.e., adding new training data that could improve its

classification performance. It is worth noting that architecture of the used DL model has been widely used in many other tasks showing successful results, yet in the present case is very limited. A factor influencing the net performance is definitely the number of samples required for training and this is probably the case here. Although automatic differentiation between lymphocytes and cancerous cells is challenging, combination of an accurate nuclei segmentation strategy (e.g. watershed) and a set of simple but discriminating features could be good enough. General results suggest this approach provides good classification metrics and does not require high computational capabilities, long training times, nor thousands of annotated data.

Although results using the introduced simple approach were good, it is worth remembering the model was trained using a dataset with a few number of patients, so the question about how well this method generalize to other cases is still open. The model was designed taking into account two important aspects: 1) the visual features selected for identification of lymphocytes were based on the visual information pathologists use in their daily routine to differentiate lymphocytes from other structures, 2) lymphocytes have a low biological variability, meaning their appearance is similar regardless the tissue sample. For this reason, it is expected the model will have a good performance on new images; however, more experimentation is needed to validate this assertion. New experimentation could include either adding more training data from different patients or using data augmentation strategies.

3.3.2. Extracting topological and density information from Tumor-Infiltrating Lymphocytes for prognosis

Recent evidence has demonstrated that TIL density and spatial arrangement of cells appear to be prognostic of disease outcome in NSCLC. For this reason, in this section we present and evaluate computer extracted spatial TIL (SpaTIL) features relating to 1) the spatial architecture of TIL clusters, 2) co-localization of clusters of both TILs and cancer nuclei, and 3) variation in density of TIL clusters across the tissue slide image.

The association between disease recurrence and the SpaTIL features was explored on a total of n=301 patients with stages I and II of non-small cell lung cancer. Additionally, we also compared the SpaTIL features in terms of their ability to predict recurrence in these patients against the manually estimated degree of TILs by two thoracic pathologists. Finally, the SpaTIL features were also compared against an automated estimation of TIL density via a computerized algorithm for identifying TILs in pathology images.

Materials and methods

Datasets

Tissue microarrays (TMAs) H&E-stained samples from three independent and well-characterized NSCLC cohorts in stage I and II were included in this study, representing a total of n=301

patients. The three cohorts were represented by D1 (n=70), D2 (n=119), and D3 (n=112). A fourth dataset, named D4 (n=112), was also included, containing tissue punches corresponding to the same patients in D3 but extracted from different regions of the tumor. The corresponding clinico-pathological information from patients in D1-D4 was collected from clinical records and pathology reports. Patients from cohort D1 and D2 contained the formalin-fixed paraffin-embedded (FFPE) tumor samples from previously reported retrospective collections of NSCLC patients [143] and 0.6-mm cores from each tumor were arrayed in the form of TMA. Another 116 patients provided two punches from the same tumor consisted of D3 and D4, also in the form of TMA. Standard TMA preparation protocol was described in [28]. Datasets D1 and D2 were scanned and digitized at 20x magnification. The dataset D3 and D4 were scanned and digitized at 20x using a Ventana iScan HT Scanner (serial: BI15N7205). Finally, a 1500 pixel x 1500 pixel image at 20x magnification was extracted to represent unique patient from each dataset. The first cohort (D1) was employed for feature discovery and model training. This dataset included samples from 350 patients and was collected independently at Sotiria General Hospital and Patras University General Hospital between 1991 and 2001. Cohorts D2 and D3 were used for independently validating the trained classifier. D2 comprised samples from 202 patients and was collected at Yale Pathology between 1988 and 2003. D3 comprised tissue images from 189 patient samples and was collected at Cleveland Clinic between 2004 and 2014. D3 and D4 were used to quantitatively assess the ability of the approach to deal with intra-tumoral heterogeneity. Figure **3-9** illustrates the inclusion and exclusion criteria for patient selection for this study.

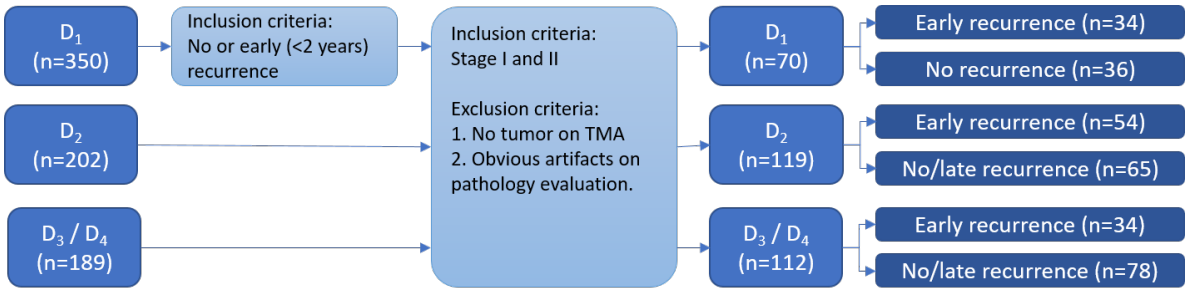


Figure 3-9: Patient selection workflow for the datasets included in this study.

Automatic characterization of TILs

Identification of Lymphocytes

The first step was to identify the spatial location of the TILs on the digitized H&E images. A watershed-based algorithm [165] was used for automatically detecting nuclei on H&E images. This method applies a set of mathematical operations (fast radial symmetry transform and regional minima) at different scales to identify nuclei candidates. This method was selected

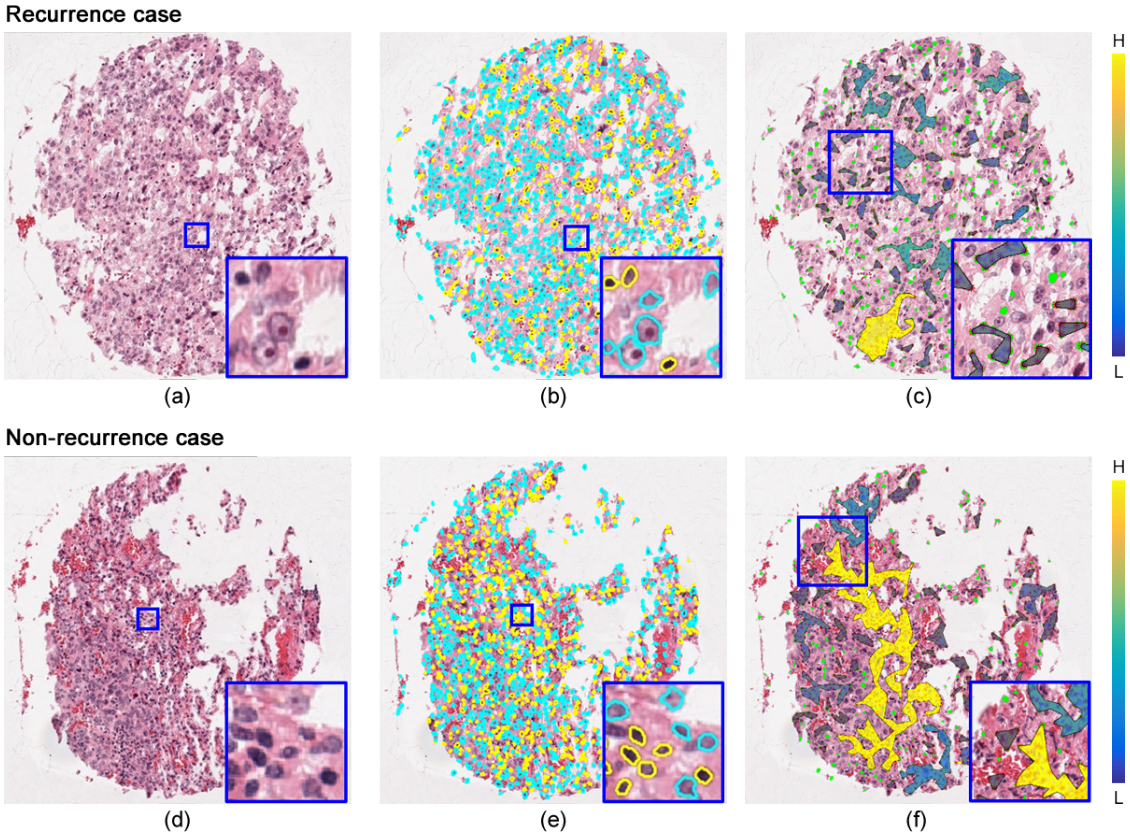


Figure 3-10: Representative TMA tissue spots of recurrent (top row) and non-recurrent (bottom row) early-stage NSCLC cases. The first column (a, d) shows the original H&E-stained images. Identification of TILs (yellow) and non-TILs (cyan) is presented in the second column (b, e). The third column (c, f) illustrates the qualitative representation of one of the SpaTIL features overlaid on the original images, specifically, the variation in the density of lymphocyte clusters. The color bars represent the density measurement (H stands for highly dense clusters while L stands for low-density or sparse clusters). Non-recurrence cases are characterized by the presence of more high-density clusters while recurrence cases were comprised of a larger number of low-density clusters.

by its visual efficacy and documented advantages, namely simplicity, speed, and lack of adjustable parameters. Once nuclei were detected, the method presented in Subsection 3.3.1 (also presented in [37]) was used to identify lymphocytes. This approach takes advantage of the fact that TILs tend to be smaller compared to cancerous nuclei, they also tend to be more rounded and with a darker homogeneous staining. Once all the nuclei candidates have been identified by the watershed approach described above, a machine classifier using 7 features related to texture, shape, and color attributes of the segmented nuclei is used to identify individual nuclei as being either lymphocytes or non-lymphocytes (See Figure **3-10**-b).

Spatial TIL Graph Construction

A graph is a mathematical structure composed of finite sets of objects (nodes) that capture global and local relationships via pairwise connections (edges) between the nodes. Graphs have been previously used to characterize nuclear architecture in histopathologic images due to their ability for representing spatial information such as neighborhood relationships and spatial arrangement of nuclei [13, 6, 5]. In order to evaluate a spatial network of TILs and to extract the corresponding SpaTIL features, we first identify sets of clusters of proximally adjoining TILs and non-TILs respectively. We first represented each of the individual TILs and cancer nuclei by their centroids, which in turn represents the nodes of a graph. Using the approach described in [6, 5], each node is connected to others based on the Euclidean distance, a weighting function favoring the connectivity between proximal nodes. After this process, multiple disconnected subgraphs or clusters of TILs are generated. This process was also performed with all the non-TILs (See Figure **3-10**-c).

SpaTIL features

Two sets of features were extracted. The first comprises 20 features related to spatial arrangement of TILs are extracted from the TIL cluster graphs. These features include first-order statistics (e.g. mean, mode, median) of the following: number of lymphocytes within the clusters, ratio between the area of the lymphocyte clusters and area of the TMA spot, ratio between the numbers of TILs within the cluster and the cluster area, among others. The second set includes 65 features describing the relationship between lymphocyte and non-lymphocyte clusters were extracted for each image. They include, for instance, the ratio between the density (ratio between the number of nuclei within the cluster and the cluster area) of a non-lymphocyte cluster and the density of its closest lymphocyte cluster, the intersecting areas of the lymphocyte clusters and non-lymphocyte clusters, a value indicating if the nearest neighbor of a lymphocyte cluster is either a lymphocyte or a non-lymphocyte cluster.

Feature selection

The Minimum Redundancy Maximum Relevance (mRMR) feature selection method [126]

identified the SpaTIL features most correlated with recurrence in the discovery set D1, while also eliminated redundant features. Table **3-5** presents the top SpaTIL features identified via this approach.

Comparative strategies

Inter-reader variability in TIL estimation by human readers

Two expert pathologists experienced in grading TILs were asked to determine the infiltration grade for each of the images in D1 and D2 via visual evaluation. A simple developed in-house custom web application was used by the readers which in turn enabled them to assign an infiltration score to each image. Infiltration options were defined based on findings reported in [143, 20], as 0) no-infiltration (virtual absence of TILs), 1) low (1 %-33 %), 2) moderate (34 %-66 %), and 3) high (67 %-100 %). The agreement among pathologists during the TIL-grading task was measured, and for this purpose two indices were computed: Pearson's correlation [18] and Cohen's Kappa coefficient [29]. The Kappa index is a widely used measure to determine the agreement among a set of experts making categorical judgments, considering agreement may occur by chance.

Computer based estimation of TIL-density

We also extracted TIL-density-based (DenTIL) features and compared the prognostic performance of these DenTIL features against the SpaTIL features. The DenTIL features included ratio between the number of lymphocytes and the TMA spot area, ratio between the sum of lymphocyte areas and area the TMA spot, ratio between the number of lymphocytes and the number of non-lymphocytes, and a grouping value indicating how close are lymphocytes to each other (computed as the sum of the inverse distances between lymphocytes).

Statistical analysis

Classification

A Quadratic Discriminant Analysis (QDA) classifier was trained using the top SpaTIL features (QS) identified from D1 to separate patient cases into two classes: recurrence and non-recurrence. We chose this classifier owing to the fact that it has no hyper parameters to tune, and is able to learn quadratic boundaries and therefore this is more flexible than linear classifiers. Similarly, another QDA classifier was trained using DenTIL features (QD) on the training set D1. The performance of the QDA classifiers QS and QD in distinguishing between early stage NSCLC patients who did and did not have recurrence was evaluated on the independent validation sets D2, D3, and D4. For each image in the test sets, QS and QD assigned a probability of recurrence. Classifier performance was evaluated via the area under the receiver operating characteristic (ROC) curve (AUC). The recurrence and non-recurrence labels predicted by QS and QD were compared with the ground truth labels (true

patient outcomes) to determine classifier accuracy and AUC. The AUC obtained for QS on D3 and D4 were quantitatively compared to evaluate the effect of spatial tissue sampling on the classifier performance.

Statistical and Survival Analysis

Recurrence-free survival (RFS) was defined as the time interval between the date of diagnosis and the date of death or recurrence (whichever happened first). Patients who were still alive without recurrence at the last reported date were labeled as censored. The Kaplan-Meier survival analysis was used to examine the difference of RFS between different patient groups categorized by the classifier output. As previously stated, D1 was used to train a classification model which was applied on datasets D2 and D3. The difference of survival in each group predicted by the classifier was assessed by log-rank test. Multivariable Cox regression was employed to examine the predicting ability of the QS and QD classifiers when controlling the effects of clinical and pathological parameters, namely gender, age, tumor stage, T-stage, and N-stage. P-values were two-sided assessed, and all values under 0.05 were considered as statistically significant. A Kaplan-Meier analysis was also carried out based on the TIL density estimation by the human readers. For this purpose, patients were split into two groups: High-TIL and Low-TIL. Since pathologists graded each case from 0 to 3, case patients with TIL categories of 0 to 2 were considered as being part of the low-TIL group and those with scores of 3 were grouped in as part of the high-TIL tumor category. This strategy was based on the approach described in [143, 20].

Experimental Results

Clinico-pathologic Features of the Patient Cohorts

The median follow-up for patients was 40.91 months, 45.33 months, and 70.92 months for D1, D2, and D3, respectively. By the end of the study/follow-up, 34 out of 70 patients (48.6 %) in D1, 54 of 119 (45.4 %) in D2, and 34 out of 116 (30.4 %) in D3 had developed recurrence. A summary of clinical and pathological features for the 3 cohorts is presented in Table 1, and a summary discriminated by dataset is presented in Table **3-3**.

Experiment 1: Prognostic ability of SpaTIL in early stage NSCLC

Figures **3-11**-c, **3-12**-c, and **3-13**-a illustrates the ROC curves and corresponding AUCs for the SpaTIL classifier (Qs) for predicting recurrence in NSCLC on D1, D2, D3, and D4. Concordance indexes (CI) of the binary classifier for the 4 datasets were 0.7002, 0.7248, 0.7008, and 0.7063, respectively.

Significance of clinical and pathologic variables with patients' survival time in the test sets was evaluated by log-rank test . Multivariate survival analysis controlling the effect of major pathological and clinical variables is presented in Table **3-4**. Patients identified by the Qs classifier as having poor prognosis had statistically significantly worse disease-specific sur-

Variable	Sub-variables	N (%)
Number of patients		301
Average age		64.3 +/- 10.5
Sex	Male	176 (58.5)
	Female	125 (41.5)
N-Pathological	0	206 (68.4)
	1	95 (31.6)
T-Pathological	1	136 (45.2)
	2	165 (54.8)
Stage	I/IA/IB	220 (73.1)
	II/IIA/IIB	81 (26.9)
Recurrence	Non-recurrence	135 (44.9)
	Recurrence	166 (55.1)
Tumor types	Adenocarcinoma	135 (44.9)
	Squamous Cell Carcinoma	89 (29.6)
	Others	77 (25.6)

Table 3-3: Summary of clinical and pathological features for the patients in D1, D2, and D3.

vival. The calculated hazard ratio was 3.08 (95 % confidence interval, 2.1 - 4.5, p=7.3e-5), meaning that patients with recurrence were approximately 3 times more likely to develop disease recurrence and die from it.

Figures **3-11**-d, **3-11**-d, and **3-11**-b illustrate the Kaplan-Meier plots corresponding to the SpaTIL features for D1, D2, and D3, respectively. Qs was found to be prognostic for D1, D2, and D3 (p-values < 0.02).

Experiment 2: Comparison of Human and Machine based assessment of TIL Density for Predicting Recurrence in Early stage NSCLC

Pearson's correlation and Cohen's Kappa index were computed to measure the agreement among pathologists. Subfigures **3-11**-a and **3-12**-a show the computed values for the two pathologists for D1 and D2, respectively. The overall computed Kappa (considering both analyzed datasets) was 0.5028. When computed independently for each dataset, Kappa indices were 0.3781 and 0.6216 for D1 and D2, respectively. On the other hand, the correlation coefficients were 0.5704 for D1 and 0.7902 for D2.

Subfigures **3-11**-b and **3-12**-b illustrate the Kaplan-Meier plots for both pathologists on D1 and D2, respectively. For reader 1, no statistically significant correlation between TIL estimation and outcome was observed for D1 (p=0.1373) nor D2 (p=0.2584). Conversely, for reader 2, a significant statistical correlation was observed for set D1 (p=0.0082), but not for

set D2 (p=0.0712).

Subfigures **3-11**-d, **3-12**-d, **3-13**-b, and **3-13**-e illustrate the Kaplan-Meier plots corresponding to the QS classifier on D1, D2, D3, and D4, respectively. The QS classifier was found to have a statistically significant statistical correlation between the classifier and patient outcome for D1, D2, D3, and D4 (p-values = 0.0034, 0.0005, 0.0014, and 0.0152, respectively).

Subfigures **3-11**-e, **3-12**-e, **3-13**-c, and **3-13**-f illustrate the Kaplan-Meier plots corresponding to the QD classifier on D1, D2, D3, and D4, respectively. The QD classifier was found to have a statistically significant statistical correlation between the classifier and patient outcome for D1 and D2 (p-values = 0.0082 and 0.0003, respectively); the predictions were not statistically significant for D3(p=0.3602) and D4 (p=0.3638), respectively.

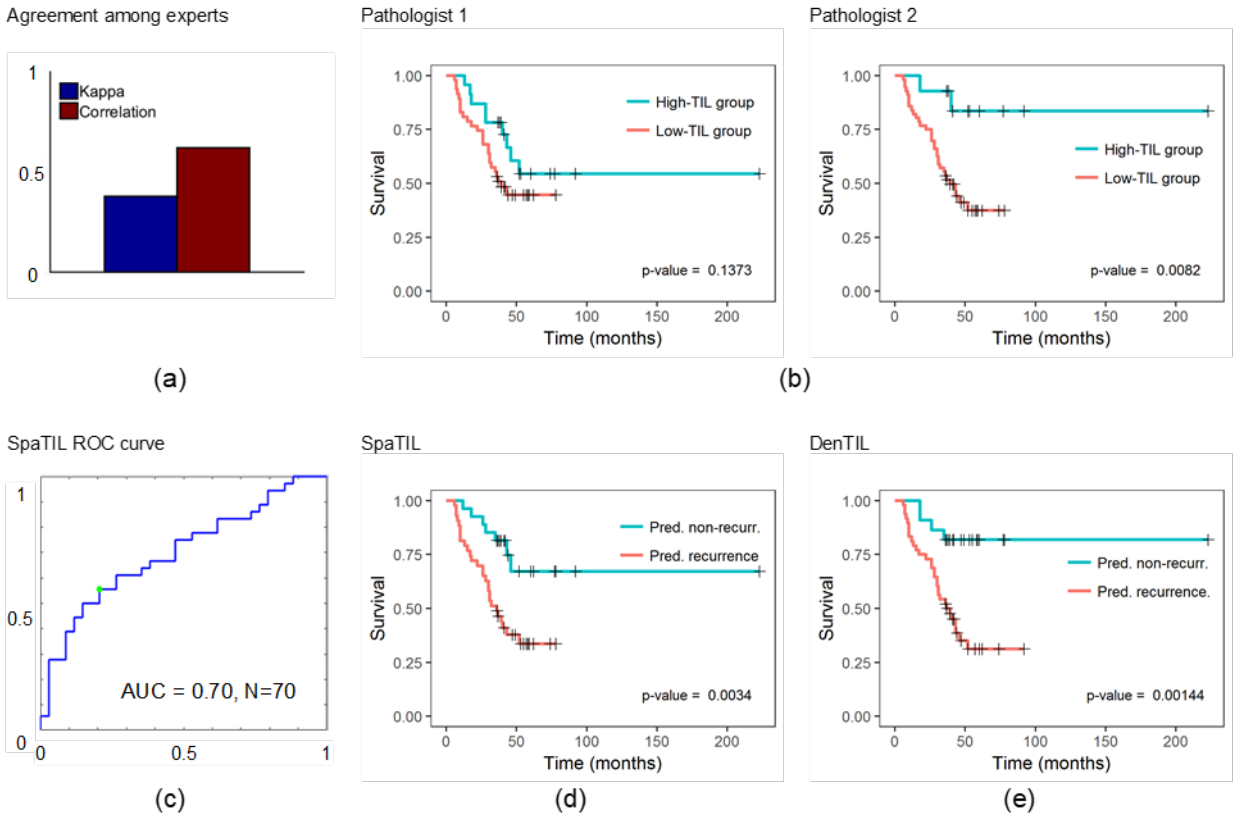


Figure 3-11: Prognostic prediction results for human readers, QD, and QS for D1. a) Bar chart illustrating the Kappa index and correlation coefficient computed for the agreement among readers 1 and 2, b) Kaplan-Meier curves for readers 1 and 2, c) ROC curve and corresponding AUC for QS, d) Kaplan-Meier plot for QS classifier, e) Kaplan-Meier plot for QD classifier.

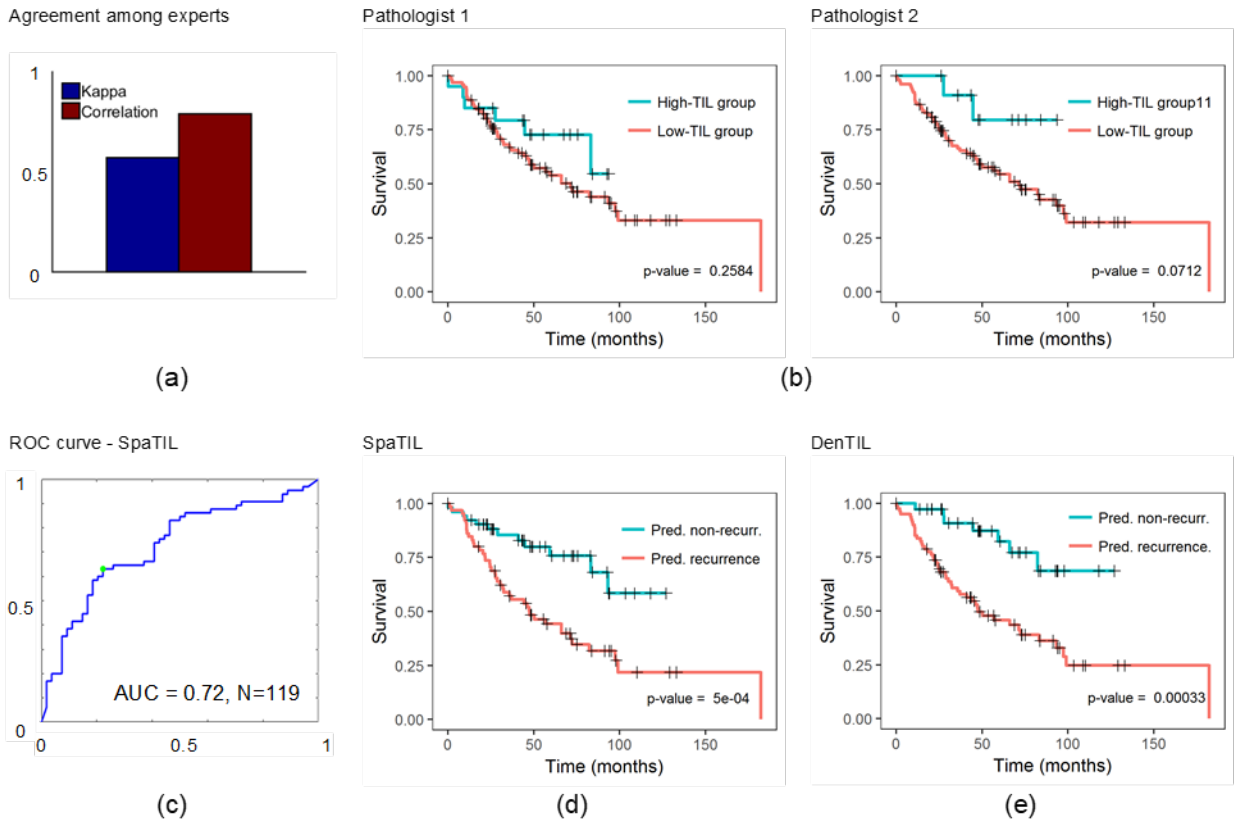


Figure 3-12: Prognostic prediction results for human readers, QD, and QS for D2. a) Bar chart illustrating the Kappa index and correlation coefficient computed for the agreement among readers 1 and 2, b) Kaplan-Meier curves for readers 1 and 2, c) ROC curve and corresponding AUC for QS, d) Kaplan-Meier plot for QS classifier, e) Kaplan-Meier plot for QD classifier.

Discussion

NSCLC is one of the most common lung cancers and presents high recurrence rates. Previous works have reported that 30-50 % of patients develop recurrence and die of their disease despite curative resection [162]. Identification of patients which disease is likely to recur can help clinicians for decision making, treatment planning, and guide the administration of adjuvant therapies. Traditionally, the prognosis of NSCLC has been evaluated according to TNM staging system; unfortunately, growing evidence has shown that such a system is not accurately enough [57, 162]. Therefore, alternative biomarkers are required as well as strategies that ease stratification of patients based on the risk of recurrence. Several studies have demonstrated that the presence and density of TILs are strongly correlated with the clinical response in patients with different kinds of cancer, namely lung, breast, ovarian, pancreatic, colorectal, skin, among others [7, 136, 140, 57, 142, 20, 22]. In the case of NSCLC, different

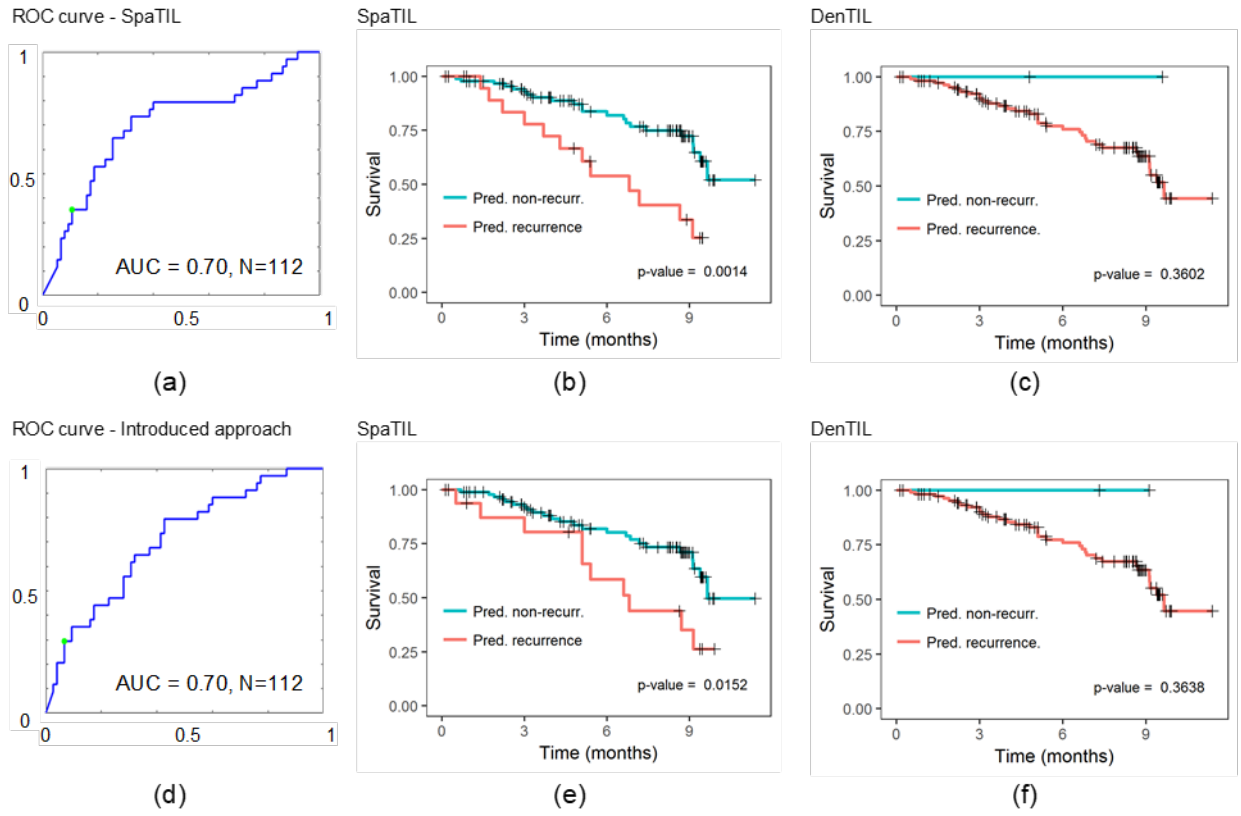


Figure 3-13: Prognostic prediction results for QD and QS for D3 (top row) and D4 (bottom row). First column (a, d) shows the ROC curve and corresponding AUC for QS, second column (b, e) presents the Kaplan-Meier plots for QS classifier, and third column (c, f) illustrates the Kaplan-Meier plot for QD classifier.

works have demonstrated a high association between the presence of TILs and patient outcome. For instance, in [143], Schalper et al. showed that the presence of certain subpopulations of TILs were associated with recurrence in early-stage NSCLC. Similarly, Liu et al. [100] found that presence of specific TIL subtypes is predictive of response to platinum-based neoadjuvant chemotherapy in advanced NSCLC patients. In such works, however, pathologists estimate by eye the quantity of lymphocytes, a time consuming and prone to error task. Different studies have shown that human based estimation of TILs presents poor reproducibility and has, at best, moderate inter-agreement [20, 7]. Different works have employed computer based approaches for automatic estimation of TILs [7, 136, 49]. Although these strategies mitigate the subjectivity and improve reproducibility, different studies support that the mere quantification of TILs might not be enough for predicting prognosis. Previous works have shown information from different cells within the tumor area has a prognostic value. For example, in [179] and [104], authors found association between different morpho-

Variable	**Hazard Ratio (95 % Confidence Interval)**	**p-value**
Gender Male vs. Female	1.1836 (0.8494-1.6492)	0.3193
T-stage T1 vs. T2	0.9768 (0.6834-1.3961)	0.8975
N-stage N0 vs. N1	0.9128 (0.5929-1.4051)	0.6784
Stage Stage I vs. Stage	1.0107 (0.6478-1.5771)	0.9625
Tumor subtypes ADCs vs. SCC vs. Others	0.9225 (0.7178-1.1856)	0.5287
SpaTIL recurrence vs. non-recurrence	3.0791 (2.1034-4.5075)	**7.29e-05**
Two-sided p<0.05 (in bold) was considered as statistically significant.		

Table 3-4: Multivariable Cox analysis of independent prognostic ability of SpaTIL controlling for major clinical parameter on the test set.

logical nuclei and surrounding cytoplasm features and survival in patients with early-stage NSCLC. In other study [170], researchers showed that spatial architecture of cells is highly correlated to recurrence in early-stage NSCLC. Other studies have also revealed that location of immune cells with respect to cancer cells might be of biological relevance [143]. All this evidence suggests that tumor microenvironment and interplay between cancerous cells and lymphocytes plays a determinant role in the disease progression and patient prognosis. In this work, we presented a set of features based on the spatial architecture of TILs (SpaTIL), devised to capture the TIL local density, variances of the density, architecture, and co-localization of TIL and cancerous nuclei. All nuclei in the images were first identified by a watershed-based algorithm. A feature based classifier distinguished these nuclei as either lymphocytes or non-lymphocytes (mostly, cancer nuclei). Then, based on the Euclidean distance between nuclei, different cell cluster graphs of lymphocytes and non-lymphocytes are built. From such clusters, different topological and density measurements are extracted, including local density of clusters, area of clusters, the intersected area between clusters of lymphocytes and non-lymphocytes, characterization of the neighborhood of clusters, graph-based metrics (betweenness and closeness centrality), among others. Two specific questions were addressed in this study. First, could SpaTIL features independently predict recurrence in early stage NSCLC. Secondly, would the SpaTIL features offer more prognostic capability compared to TIL density alone based off 1) manual assessment by expert human readers and 2) and a computerized automated TIL density estimator. The SpaTIL features were

Rank	Name	Description
F1	AvgLCNLC	Mean number of lymphocyte clusters surrounding a non-lymphocyte cluster
F2	MoDLC	Mode of the density of all lymphocyte clusters.
F3	StdANLC	Standard deviation of the area of non-lymphocyte clusters.
F4	AvgANLC	Mean of the area of non-lymphocyte clusters.
F5	Me2N_NLCLC	Considering the two nearest neighbor clusters to a non-lymphocyte cluster, median number of lymphocyte clusters.
F6	Avg2N_NLCLC	Considering the two nearest neighbor clusters to a non-lymphocyte cluster, mean number of lymphocyte clusters.
F7	Me3N_NLCLC	Considering the three nearest neighbor clusters to a non-lymphocyte cluster, median number of lymphocyte clusters.
F8	Avg3N_NLCLC	Considering the three nearest neighbor clusters to a non-lymphocyte cluster, mean number of lymphocyte clusters.
F 9	Mo4N_NLCLC	Considering the four nearest neighbor clusters to a non-lymphocyte cluster, mode of the number of lymphocyte clusters.
F 10	AvgRBNLCLC	Mean of the ratio of betweenness centralities of non-lymphocyte clusters to their closest lymphocyte cluster.
F 11	AvgRCNLCLC	Mean of the ratio of the closeness centralities of non-lymphocyte clusters to their closest lymphocyte cluster.

Table 3-5: List of the most discriminating spaTIL features for predicting recurrence in NSCLC. In this study, density is defined as the ratio of the number of cells within the cluster to the cluster pixel area, betweenness centrality is a measure of centrality in a graph based on shortest paths, and closeness centrality is a measure of centrality calculated as the sum of the length of the shortest paths between the node and all other nodes in the graph.

evaluated by training a Quadratic Discriminant Analysis classifier to predict recurrence on 3 different cohorts: D2 (n=119), D3 (n=112), and D4 (n=112); with D3 and D4 containing cases of the same patients but extracted from different tumor regions for assessing the incidence of tumor heterogeneity. The SpaTIL classifier was modeled using a cohort comprising 70 patients (D1). On two TMAs with 119 and 112 patients, the SpaTIL classifier yielded CIs

of 0.72, and 0.70, respectively. A Kaplan-Meier analysis showed a strong association between the predictions of the SpaTIL classifier and recurrence for D2 (p=0.0005), D3 (p=0.001), and D4 (p=0.01). Likewise, other factors were part of the analysis and the multivariate analysis for the automatic strategy, controlling T and N classification, showed results also strongly associated with recurrence; the calculated hazard ratio was 3.08 (95 % confidence interval, 2.1 - 4.5, p=7.3e-5). We also compared the prognostic performance of TIL estimation carried out by two human readers. A Kaplan-Meier analysis was conducted for each pathologist; results showed that no significant statistical correlation was found between Pathologist 1 and prognosis for any dataset (p>0.05) while there was a significant statistical correlation between TIL grade estimation of Pathologist 2 and patient outcome for D1 (p=0.006). In addition, the agreement among expert pathologists was studied for D1 and D2. The computed Kappa value was 0.5028, which is lower than the values previously reported for NSCLC (K=0.59) [20] and breast pathology (K=0.72) [26].This moderate agreement might be due to the fact that TIL grading in lung pathology lacks a well-defined set of guidelines, so each pathologist might either focus on different areas of the tissue during examination (e.g., epithelium or stroma) or count different cells within the infiltration (e.g., plasma cells). In contrast, the results obtained by using SpaTIL features appear to suggest that the spatial arrangement of TILs and cancer nuclei were strongly associated with recurrence in early stage NSCLC (p<0.05). Two related approaches to the work presented here have been published in the literature. In [84], aiming to deal with tumor heterogeneity, Khan et al. started by extracting several TMA cores for each patient in the dataset. Then, they measured the infiltration for each core as the ratio of lymphocytes to all cells. Finally, they computed an infiltration score for each patient based on the skewness of the measured infiltration values. Such a score was found to be associated with poor prognosis in breast cancer, specifically HER2 (p=0.018). In the present work, SpaTIL features were also tested in tumor heterogeneity conditions. D3 and D4 cohorts contain samples of the same patients but extracted from different tumor areas, and results support the idea of such features are prognostic in spite of tumor heterogeneity ($p \leq 0.01$). Similar to Khan et al. [84], in this article the ratio of lymphocytes to all cells in a TMA core was computed and used in conjunction with other density-based features (DenTIL) to assess their prognostic value. In this case, however, results were statistically significant only for D2 (p=0.0003) but not for D3 nor D4 ($p \geq 0.36$). In [79], Saltz et at. used a deep learning model to identify patches from whole slide images (WSIs) containing TILs. Next, similar patches were clustered by means of affinity propagation. Finally, from such patch clusters, authors computed different measures and indices, namely Ball and Hall, Banfield and Raftery, C, determinant ratio, among others. Five associations between cluster index and patient survival were found to be significant for different tumor types: Ball-Hall for breast invasive carcinoma (p=0.007), C-index for lung adenocarcinoma (p=0.003), Banfield-Raftery for prostate adenocarcinoma (p=0.013), Determinant Ratio for prostate adenocarcinoma (p=0.012), and Banfield-Raftery for skin cutaneous melanoma (p=0.001). Similar to SpaTIL, the approach presented by Saltz et al. [79] supports the idea that spa-

tial arrangement of TILs has a prognostic value; however, SpaTIL features include not only information from TILs but also from cancer nuclei. Under the idea that spatial location of immune cells with respect to cancer cells might be of biological relevance [143], SpaTIL features attempt to characterize the interplay between clusters of TILs and non-TILs within the tumor region. Our work did however show some limitations. First, unlike the study of Saltz et al. [79], our approach was evaluated on tissue microarrays and not on WSIs. Despite representing small portions of the whole tumor, TMAs allow pathologists to rapidly evaluate large clinical cohorts. Previous works have stated that results using TMAs are concordant with other studies in the field using TIL-based biomarkers from WSIs [140, 143, 84]. Even though our approach did evaluate different punches of the tissue from different locations within the tumor and revealed that SpaTIL features were prognostic tumor heterogeneous conditions, this approach will clearly need to also be evaluated on whole slide images. Second, even though two independent validation cohorts were employed (one of them with TMA cores from different tumor locations), the number of samples in the study is relatively small in terms of the complexity of this problem. Future work will include a further analysis of the proposed approach on WSIs and in a larger independent test set. Third, the employed datasets were highly heterogeneous, with images obtained from very different places and scanners. Although the success of SpaTIL features despite such heterogeneity is a promising sign regarding their robustness, further investigation into sensitivity of scan parameter heterogeneity on identification of lymphocytes and non-lymphocytes is required. In summary, in this work, we presented an approach that exploits a set of density and spatial topological features related to arrangement of clusters of TIL and non-lymphocytes within the tumor area, which were found to be predictive of patient recurrence in NSCLC cases. With an additional larger, multi-site validation, this approach could potentially form the basis for an image based companion diagnostic test to identify which early-stage lung cancer patients are at increased risk for recurrence and hence candidates for adjuvant chemotherapy.

3.3.3. Identifying groups of lymphocytes and their incidence in prognosis

As previously stated, it is well-acknowledged that most tumors trigger an immune response modulated by TILs[140]. Different studies have shown that density of TILs is highly correlated with the disease progression, patient survival, and treatment response in different types of cancer [140, 143, 7]. Recent works, however, have suggested that studying immune infiltration might not be as valuable as evaluating the relative concentrations of the different TIL subtypes, each of which may have different biological roles in tumor control [143]. While some of these TIL populations might stimulate an anti-tumor response, others might promote cancer progression [140]. Unfortunately, the different TIL subtypes are difficult to spot by manual inspection in Hematoxylin and Eosin (H&E) stained samples.
Multiplexed quantitative fluorescence (QIF) or immunohistochemistry (IHC) have been used

to identify TIL subtypes including CD3, CD4, and CD8. In [143], Schalper et al. showed that the increased levels of CD3 and CD8 TILs were associated with improved outcome in non-small cell lung cancer (NSCLC). Interestingly, in [30], the spatial interplay between different families of intratumoral TIL cells was correlated with survival in pancreatic cancer. More recently, in [12], the authors showed that the spatial interplay between tumor and regulatory T cells was associated with survival in NSCLC. However, in all of these studies QIF or IHC was employed for TIL subtyping, non-trivial and tissue-destructive processes.

Even though the different TIL subtypes may not be visually discernible on routine H&E tissue slide images, there is no doubt that spatial architecture of the different TIL phenotypes could be more prognostic of disease outcome compared to TIL density alone. In this section we present an approach called Phenotyping of Tumor Infiltaring Lymphocytes (PhenoTIL) for identifying potential TIL cluster families and their spatial architecture from routine H&E images, potentially obviating the need for more expensive and tissue destructive approaches such as QIF and IHC. Quantitative measurements derived from the PhenoTIL approach were then applied to predicting recurrence in early-stage NSCLC and the approach compared strategies based off the density of TILs and using the extracted local and contextual TIL features with no clusterization.

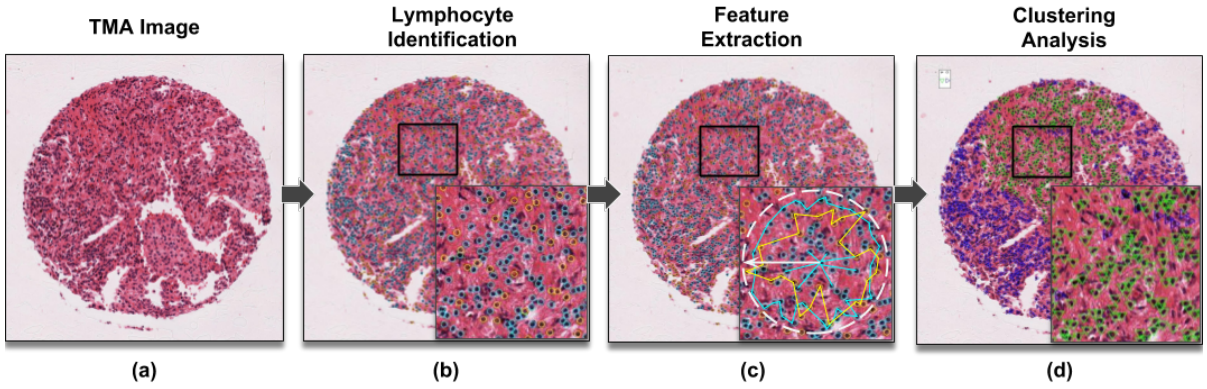


Figure 3-14: Illustration of PhenoTIL methodology for identifying TIL subtypes.

Methods

Identification and characterization of lymphocytes

A watershed-based algorithm [165] is firstly applied to segment nuclei on the image. We first segment all possible nuclei and then subsequently classify them either lymphocytes or non-lymphocytes (mainly, tumor cells) using different nuclei texture, shape, and color features [37] (See Figure **3-14**(b)). Subsequently, each lymphocyte is characterized by its own local morphological features and by a set of contextual features to describe the lymphocyte and its surroundings/neighborhood. For each lymphocyte, a set of circles with incremen-

tal radii[1] of $k = dL \times 10, dL \times 20, dL \times 30$ pixels were placed at the lymphocyte center; $dL = 20$ pixels being the average diameter of all the detected lymphocytes. Finally, a set of features are computed within each circle and used to characterize the lymphocyte (See Figure **3-14**(c)). These features aim to measure the grouping factor of lymphocytes, relative lymphocyte-cellular density, lymphocyte tissular density, variation of the median intensities of lymphocytes, relative lymphocyte-cellular interspersing, among others. The full set of features is described in Table **3-6**.

Table 3-6: Contextual Features

Sum of the inverse Euclidean distances between the lymphocyte on the center of the circle and the other lymphocytes in the circled area.
Ratio between the sum of the area of lymphocytes and the area of non-lymphocytes.
Ratio between the sum of the areas of lymphocytes and the eosinophilic tissue area.
Ratio between the number of lymphocytes and the number of non-lymphocytes.
Average of the squared differences between the median intensities of lymphocytes.
Ratio between the area of the convex hull containing all the lymphocytes and the convex hull containing non-lymphocytes.
Intersection between the convex hull containing all the lymphocytes and the convex hull containing all the non-lymphocytes.
Median of distances from each lymphocyte to closest non-lymphocyte neighbor.
Ratio between the median of the distance between all the lymphocytes and their closest lymphocyte neighbor and the median of the distance between all the non-lymphocytes and their closest non-lymphocyte neighbor.
Number of lymphocytes within a convex hull of non-lymphocytes.
Sum of distances between the centered lymphocyte and its non-lymphocytes neighbors.

Clustering analysis

[1]Different radii were used aiming to model a multi-scale approach

Gaussian mixture models are probabilistic models that assume all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters. In such cases, a Dirichlet Process as a prior distribution of the number of clusters, makes that the probability mass can be easily re-distributed (Figure **3-15**). DPGMMs can then automatically learn the number of clusters adjusted to the observed data, which make them suitable for unsupervised clustering scenarios.

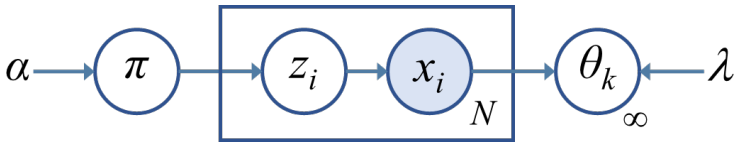


Figure 3-15: Illustration of a DPGMM. x_i are the observed data points, z_i is a set of labels assigning x_i to one of the k clusters. Cluster parameters are π (mixture proportions) and θ (cluster means and covariances) with associated uninformative priors (α and λ).

Since biological subtypes of lymphocytes are not visually distinguishable in H&E images and it is unknown the number of possible groups existent in a set of images, DPGMMs provide a convenient framework for TIL clustering.
Finding the lymphocyte clusters is summarized as follows. First, it is assumed that there exists an infinite set of latent groups, each described by a set of parameters (e.g., a Gaussian with mean μ and standard deviation σ). Then, each lymphocyte must be assigned to a cluster, similarly to the "Stick-Breaking process", as follows:

- There is a stick of length one.
- A random variable $\beta_1 \sim Beta(1, \alpha)$ is generated, a real number between 0 and 1 is set by the Beta distribution, with expected value $1/(1+\alpha)$. α is the concentration o scaling parameter, a positive real number that influences the dispersion of data points (A low α value will generate more tightly clustered data points while a high value will generate more clusters). The stick is then Break off at β_1 and w_1 is the stick length at the left.
- Now take the stick at the right, and generate $\beta_2 \sim Beta(1, \alpha)$. Break off the stick at β_2 and again, w_2 is the length of the stick to the left, i.e., $w_2 = (1-\beta_1)\beta_2$.

Image characterization
Once lymphocytes are assigned to a group, histopathologic images may be characterized using the clusters obtained via the DPGMM. For this purpose, a particular image is then represented by a histogram containing the number of TIL occurrences per cluster. These image histograms are normalized, resulting in the corresponding probability distribution functions.

Experimentation

Dataset

The dataset consisted of 178 H&E-stained tissue cores, corresponding to 178 different cases, from two independent and well-characterized collections of NSCLC represented in tissue microarrays (TMAs). TMAs were prepared using standard procedures and digitally scanned at $\times 20$. Each tissue core was digitized within an image with 1500×1500 pixels which was labeled as either *Recurrence* or *Non-recurrence*, information taken from the patient medical record.

Experiment 1: Assessing the performance of PhenoTIL

PhenoTIL was evaluated in terms of its ability to determine whether a NSCLC case may have recurrence or not. For this purpose, the whole dataset (See subsection 3.3.3) was randomly split into a learning ($n = 100$) and a test ($n = 78$) sets. The learning set was used to determine the clustering parameters and to train a Linear Discriminant Analysis (LDA) classifier model that separates cases into two classes: recurrence and non-recurrence.

Different α values were used for clustering, and the performance of each set of clusters was assessed. In addition, PhenoTIL was compared against two different approaches. The first is a model in which images are characterized by lymphocyte local and contextual features but with no clustering. In this case, features pertaining to lymphocyte mean, variance, skewness, and kurtosis are used to build the feature vector. A second prediction model was also constructed based off the density of TILs in the image, i.e., the ratio between the number of extant lymphocytes to the total tissue area.

Rows 1 and 2 of Figure **3-16** present some visual results corresponding to two non-recurrence cases while rows 3 and 4 illustrates two recurrence cases. These results show that some lymphocyte populations tend to appear more frequently on the recurrence cases while others are more common in non-recurrence cases.

Left panel of row 5 of Figure **3-16** presents the Receiver Operating Characteristic (ROC) curves corresponding to 1) a model using only TIL density, 2) a model using only local and contextual TIL features, and 3) the PhenoTIL model using different α values , which in turn resulted in 5, 8, and 10 clusters respectively. The corresponding AUCs were 0.58, 0.61, 0.80, 0.84, and 0.81, respectively. Results show that PhenoTIL outperformed the other models and a total of 8 clusters resulted in the highest area under the curve.

Experiment 2: Survival analysis using PhenoTIL

Right panel of row 5 of Figure **3-16** shows a Kaplan-Meier survival curve. We employed the previously optimally determined number of clusters (8) for locking down the PhenoTIL classifier for the survival analysis on the test set. PhenoTIL resulted in highly statistically significant separation between the patients who did and did not have recurrence following surgery.

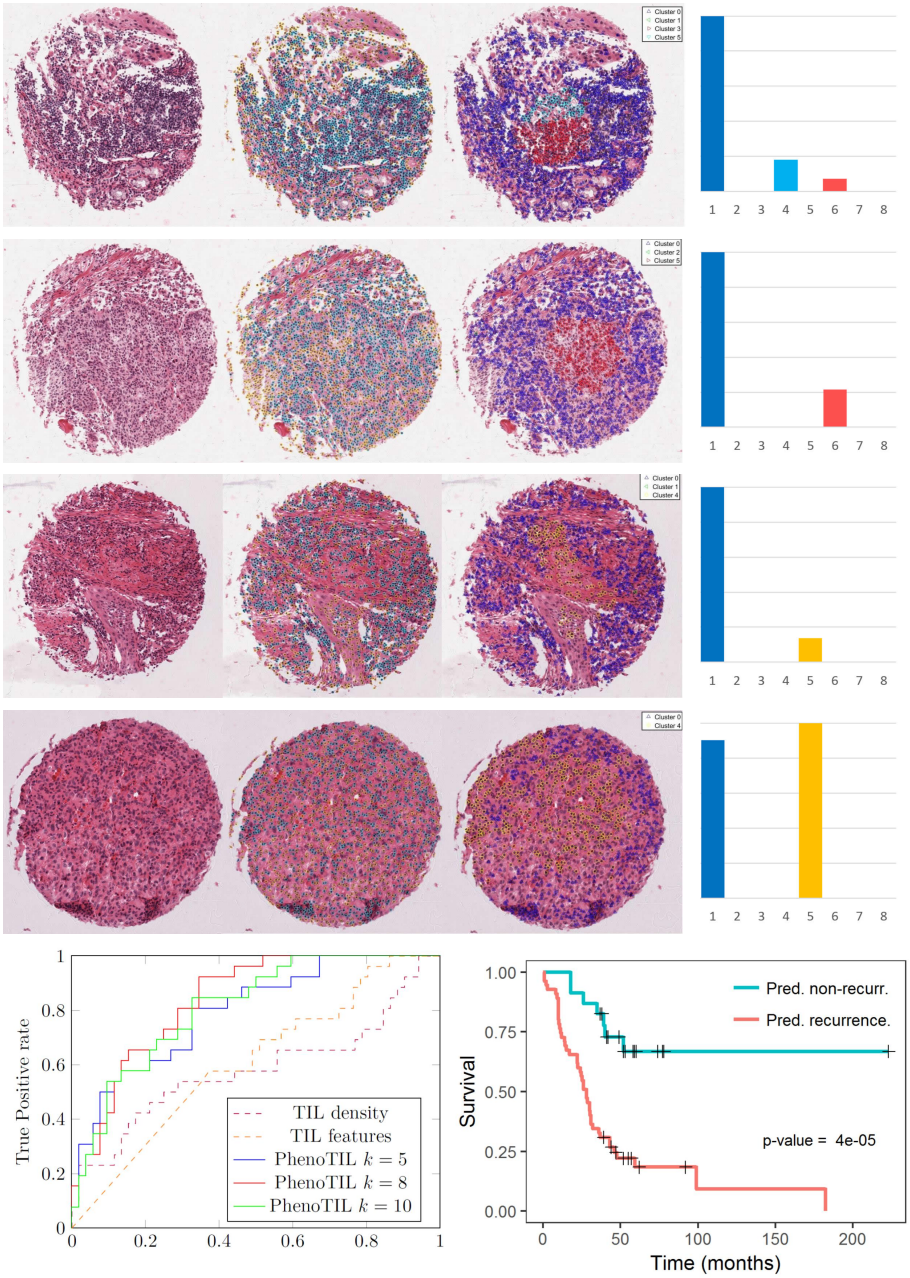


Figure 3-16: Summary of results.Rows 1 and 2 present two non-recurrence cases while rows 3 and 4 show two recurrence cases. 1st column: original image, 2nd column: TIL identification, 3rd column: TIL clustering (8 clusters), and 4th column: corresponding histogram derived from clusters. Row 5 illustrates the ROC curves of PhenoTIL and comparative strategies and the Kaplan-Meier curve of PhenoTIL using 8 clusters.

3.4. Concluding remarks

Usually, the common approach to extract relevant content from images is using low-level features, namely color, texture, orientation, or shape. Unfortunately, this strategy results highly limited in the medical field because of the high variability and complexity of biological structures. Structures very visually similar could have completely different pathological meanings. Aiming to deal with this gap between semantics and raw visual information, in this chapter we have presented different approaches that exploit the relationships between different constituent elements of an image. Specifically, we identified nuclei and studied their local and contextual information to obtain objective information that can be used for both rendering a diagnosis and establishing the patient prognosis.

In a first study case, nuclear information is used to automatically identify from a set of fields of view which of them contained DCIS and their corresponding grade. For this purpose, we characterized each image nucleus by its local morphological information (size, shape, color, texture, etc.), contextual information (neighborhood visual features), and grouping factor (distance to other nuclei). Previous studies have shown that detection of DCIS is a challenging task because of the variable morphology, biomarker expression, genomic profile, and natural progression of this type of lesion. Furthermore, such studies have revealed low levels of agreement between expert pathologists. Experimental results suggest the introduced approach could provide pathologists with objective and quantitative information, thereby easing decision-making and treatment planning.

In a second study case, we explored the prognostic potential of TILs for patients with NSCLC. Different studies have shown a a high correlation between the infiltration grade of lymphocytes and the patient prognosis (response to therapy, survival, recurrence, etc.); unfortunately, this information is not used in clinical practice because of a lack of standardized methodologies and objective strategies to quantify the infiltration. In this chapter we presented a first approach called SpaTIL that quantifies topological and density features from TILs on H&E-stained samples. Extraction of such features started by automatically detecting TILs and non-TILs (mostly cancerous nuclei) on histopathological images; then, TILs and non-TILs were grouped into small cell clusters based on the Euclidean distance. From each cell cluster, density and neighborhood characteristics were extracted. The prognostic ability of SpaTIL features was tested using 3 independent datasets; results showed a high correlation with with recurrence in early-stage NSCLC.

Recent works [143], however, have stated that lymphocytes should not be analyzed as a whole since there are different TIL families or sub-groups with different prognostic characteristics. For this reason, this chapter also presented a second approach called PhenoTIL that aims to perform a more granular analysis of lymphocytes. Unfortunately, TIL families are not easily distinguishable on H&E images, so non-standard, expensive, tissue-destructive, and time-consuming methods such as IHC must be applied to identify such sub-groups. Therefore, lymphocytes were characterized on H&E images by their local and contextual information.

Then, taking into account that we do not know how many sub-groups of lymphocytes we have, we used a soft-clustering approach aiming to naturally group lymphocytes, i.e., we did not explicitly state how many groups there were but the algorithm automatically learned the number adjusted to the observed data. Finally, we studied the relationship between the presence of certain groups and the patient prognosis. For this purpose, we used 2 of the datasets previously used to validate SpaTIL, but in this case we removed some noisy samples (deteriorated tissue, staining problems, etc.). Although we found, again, a high correlation with recurrence in NSCLC, more experimentation needs to be carried out to get a definitive conclusion.
Study and quantification of lymphocytic infiltration have a great potential. Predicting whether or not a patient will have recurrence might guide clinicians or better decision making and treatment planning. In addition, this would allow clinicians for identifying patients with higher risk disease who would benefit from adjuvant chemotherapy.

3.5. Products

Journal papers

- Germán Corredor, Xiangxue Wang, Yu Zhou, Cheng Lu, Pingfu fu, Konstantinos Syrigos, David Rimm, Kurt Schalper, Michael Yang, Eduardo Romero, Vamsidhar Velcheti, and Anant Madabhushi. *Spatial architecture and arrangement of tumor-infiltrating lymphocytes for predicting likelihood of recurrence in early-stage lung cancer*. Accepted for publication in Clinical Cancer Research.

Internships

- Research visitor at Center for Computational Imaging and Personalized Diagnostics. Case Western Reserve University, Cleveland - Ohio, U.S.A. (October 2015 - January 2016)

- Research visitor at Center for Computational Imaging and Personalized Diagnostics. Case Western Reserve University, Cleveland - Ohio, U.S.A. (May 2016 - August 2016)

- Research visitor at Center for Computational Imaging and Personalized Diagnostics. Case Western Reserve University, Cleveland - Ohio, U.S.A. (February 2018 - May 2018)

Conference papers

- Xiangxue Wang, Germán Corredor, Eduardo Romero, Andrew Janowczyk, Yu Zhou, Michael Yang, Vamsidhar Velcheti, and Anant Madabhushi. *Computerized Density Estimation of Tumor-Infiltrating Lymphocyte in H&E TMAs Predicts Recurrence in*

Early Stage Non-Small Cell Lung Cancer. Proc. of the USCAP 106th annual meeting, San Antonio - USA, 2017

- Paula Toro, Germán Corredor, Xiangxue Wang, Viviana Arias, Vamsidhar Velcheti, Anant Madabhushi, and Eduardo Romero. *Quantifying expert diagnosis variability when grading tumor-Infiltrating lymphocytes*. Proc. of the 13th International Symposium on Medical Information Processing and Analysis, San Andres - Colombia, 2017
- Juan García-Arteaga, Germán Corredor, Xiangxue Wang, Vamsidhar Velcheti, Anant Madabhushi, and Eduardo Romero. *A lymphocyte spatial distribution graph based method for automated classification of recurrence risk on lung cancer images*. Proc. of the 13th International Symposium on Medical Information Processing and Analysis, San Andres - Colombia, 2017
- Germán Corredor, Xiangxue Wang, Cheng Lu, Vamsidhar Velcheti, Eduardo Romero, and Anant Madabhushi. *A Watershed and Feature based approach for automated detection of lymphocytes on lung cancer images*. Proc. of SPIE Medical Imaging 2018, Houston - USA, 2018
- Germán Corredor, Cristian Barrera, Paula Toro, Ricardo Moncayo, Hannah Gilmore, Anant Madabhushi, and Eduardo Romero. *Detection and grading of ductal carcinoma in situ by using structural features*. 14th European Congress on Digital Pathology and the 5th Nordic Symposium on Digital Pathology, Helsinki - Finland, 2018
- Germán Corredor, Cristian Barrera, Xiangxue Wang, Anant Madabhushi, and Eduardo Romero. *Phenotyping Tumor Infiltrating Lymphocytes on H&E Tissue Images: Predicting Recurrence in Lung Cancer*. Proc. of SPIE Medical Imaging 2019, San Diego - USA, 2019

Indirect products

- Pablo Álvarez, Guatizalema Castro, Germán Corredor, and Eduardo Romero. *A statistical model for characterization of histopathology images*. Proc. of the X International Seminar on Medical Information Processing and Analysis, Cartagena - Colombia, 2014
- Pablo Álvarez, Germán Corredor, Juan García-Arteaga, Eduardo Romero. *A Low Dimensional Entropy Based Descriptor of Several Tissues in Skin Cancer Histopathology Samples*. Proc. of the 11th International Seminar on Medical Information Processing and Analysis, Cuenca - Ecuador, 2015
- David Romo-Bucheli, Germán Corredor, Juan García-Arteaga, Viviana Arias, and Eduardo Romero. *Nuclei Graph Local Features for Basal Cell Carcinoma Classification in Whole Slide Images*. Proc. of the 12th International Symposium on Medical Information Processing and Analysis, Tandil - Argentina, 2016

- Cristian Barrera, Germán Corredor, Sunny Alfonso, Andrés Mosquera, and Eduardo Romero. *An automatic segmentation of gland nuclei in gastric cancer based on local and contextual information*. Accepted for presentation in the SIPAIM â€“ SAMBA Biomedical Workshop, Granada - Spain, 2018

- Sunny Alfonso, Germán Corredor, Ricardo Moncayo, Paula Toro, and Eduardo Romero. *A method to detect glands in histological gastric cancer images*. Submitted to the 14th International Seminar on Medical Information Processing and Analysis, Mazatlan - Mexico, 2018

4 Histopathological analytics by learning from experts

4.1. Introduction

The development of computer-assisted histomorphometric tools for interrogating digital pathology slide images allows for the discovery of patterns and objective measurements associated with disease aggressiveness, grade, and patient outcome [106, 19, 105, 178]. Nevertheless, the size of a WSI, which typically runs in the order of several gigabytes, poses computational and diagnostic challenges [67]. Despite the large size of a sample, in general, pathologists require very little time to examine samples and reach a diagnosis [135]. They rapidly identify regions of potential diagnostic interest, i.e. areas that usually contain abnormal patterns that characterize a particular set of pathologies [148, 119]. Although every case is different, it is widely acknowledged that diseases present certain characteristic patterns and pathologists are educated to recognize them, even under very noisy conditions [25, 166].

There is a growing demand in pathology to use objective measurements that support treatment decisions, but few systems in use can meet this need [88]. The ability to identify the RoIs which contain pertinent diagnostic information could 1) reduce the time pathologists need to dedicate to each sample by focusing their attention on relevant areas [88] and 2) improve computational image analysis tools, thus enabling improved objective measurements of histological data [88, 116]. Different approaches [180, 132, 169] have attempted to identify relevant information on medical imaging databases using manual annotations, a very time consuming, laborious, and impractical approach when processing vast image collections. Other approaches have used automated RoI identification by handcrafted features (e.g. color, shape, and texture). For instance, Doyle et al. [44] applied a boosted Bayesian multi-resolution classifier along with texture features to identify areas of adenocarcinoma in prostate biopsy slide images. Low resolution RoIs were identified by a texture classifier and then mapped onto the next higher resolution to generate high resolution RoIs. Similarly, Peikari et al. [125] presented a texture analysis technique that automatically triages clinically important regions on histopathological images. Another method to automatically detect RoIs in WSIs involves the use of handcrafted graph features. Graphs are mathematical constructions composed of finite sets of connected objects (nodes) that capture global and local relationships via pairwise connections between its members (edges). These structures can be used to quantitatively represent spatially distributed information, such as neighborhood relationships,

and the spatial arrangement of structural tissue primitives (e.g. nuclei, lymphocytes, and glands)[145, 96, 13, 94]. Different graph-based features such as Delaunay triangles, Voronoi diagrams, and minimum spanning trees, have been extracted and combined with other visual features to effectively characterize histopathologic images [129, 13, 6, 5, 106].

While these approaches turn out to be simple and efficient, in practice they suffer from different drawbacks. One of these drawbacks is the semantic gap, a concept related with the impossibility of fully describing the global meaning of the image by using low-level image features [65, 101, 133]. The semantic gap might be minimized by introducing prior knowledge from the pathologist as part of the learning workflow. This prior knowledge could be used to inform low-level feature algorithms where in the images to initially "look". These algorithms could combine pathologists knowledge with features such as 1) low-level image cues such as color and texture of the epithelium and the stroma, and 2) mid-level structural motifs such as nuclear architectural patterns, collagen patterns, and gland morphology.

There are several methods for either predicting or identifying the interesting RoIs in a WSI based on the pathologist knowledge. For example, Peter et al. [127] described a method using a random forest algorithm to learn and identify potential RoIs within an image. These regions were then shown to an expert who interactively flagged the regions which were actually relevant. This in turn, allowed a continuous updating of the scoring function for different RoIs. This approach requires substantial manual interaction, which is time consuming and therefore difficult to implement in clinical practice. Alternatively, this high level expert domain knowledge might be captured by implicitly extracting information from the pathologist actions with the virtual microscope during a diagnosis task [89, 62, 135, 113]. One manner consists in passively recording the navigations of pathologists during a diagnostic task, a strategy perfectly suited for real clinical scenarios. In such an approach, the idea is to identify and characterize the most frequently visited areas, since very likely they contain relevant diagnosis information.

In this chapter, we present a strategy oriented to automatically detect regions of interest from histopathological images by including the pathologist into the learning workflow. Considering that manual annotations is an unrealistic task, in the hereafter introduced approach, we captured implicit information from the expert interactions. Specifically, we recorded information from their movements (panning and zooming) while performing a diagnostic task and use it for latter analysis.

We present a study case that consists in detection of regions of interest in basal cell carcinoma, a type of skin cancer. In a first approach (Section 4.2.1), the high-level information captured from experts is used in conjunction with a visual attention map of the sample attempting to identify the most relevant regions and caching them. Although this approach showed a good predictive ability, the learned models were image-specific, so they hardly could be generalized. For this reason, a second approach (Section 4.2.2) attempted to extend the first approach by combining nuclei information with the navigations of pathologists to identify which regions are more likely to be cancerous.

4.2. Study case: Detection of regions of interest in basal cell carcinoma

4.2.1. Integrating visual attention maps and interactions of experts

Digitized versions of tissue samples can be navigated, shared and remotely accessed under client-server architectures, facilitating inter-consultation and education scenarios [53]. However, a typical Whole Slide Image (WSI) size is of the order of gigabytes, a bottleneck when visualizing a pathological case. Identification of relevant Regions of Interest (RoIs) is crucial towards development of VM applications, a claim based on several statements. First, it favors strategies such as caching or prefetching[1] that speed up navigation [98, 41, 75, 61]. Second, this process may also serve to design smart educational systems that might guide the training strategies of pathology residents. Third, these strategies can be useful for indexing histopathological databases and to perform selective compression.

Some studies [159, 158, 45] have shown that when the visual process is task-driven, identification of RoIs involves a complex interaction between two complementary sources of information, the bottom-up flow, provided by information coming up from low level image features, and the top-down flow, provided by the expert who establishes relationships between an acquired knowledge and the tissue information [62]. In consequence, the relevant content of a WSI might be located by strategies that capture the navigation patterns defined by the pathologist expertise, acknowledging that there exist identifiable patterns from the pathologist interaction process [135]. Now, while the examination path is very likely connected with a particular clinical meaning, the underlying relationship is difficult to establish. The question addressed in this paper is then how to use the observational path defined by a particular navigation to determine which are the relevant RoIs, i.e., the relevance map of a WSI.

Different information search strategies have previously used graph structures. For instance, the Google PageRank algorithm [23, 93], behind the successful Google search engine, is a classical graph representation that has demonstrated how relevancy can be naturally integrated and exploited to improve interaction between users searching information and web pages [54]. Google's strategy however can not be straightforwardly extrapolated to the VM scenario since only a small number of experts is always available. This article presents a relevance map constructed from low level features that are integrated and modulated by the high level knowledge from actual pathologists' navigations. For doing so, a graph structure is constructed, being the graph nodes, the spatial regions of the WSI and the edges, the weights (relevance) of each of these regions. Initially defined by a Visual Attention Model (VAM), edges are modified by an evidence function that is generated using a Bayesian strategy that

[1] Cache is a memory space where information previously processed shall be used in the future, while prefetching consists in anticipating the user requirements to make data available before the user requests them

learns the expert preferences during actual navigations.

Model description

The Bottom-up and Top-down flows

The bottom-up information arranges low level captured features. In this work, the model proposed by Harel et al. [71], which integrates information from color, intensity and orientation, was used to calculate a coarse salience map: a 2D function whose intensity values stand for the relevance levels or the probability map. The pathologist knowledge is included by modifying the relevance level using information from the particular visited regions, the time spent at each and specific associated exploration activities, namely panning and zooming.

The Bayesian strategy

Navigations are considered as occurrences of a random variable whose probability is determined by the relevance distribution, the posterior distribution $P(\Theta|\mathcal{X}) = P(\mathcal{X}|\Theta)P(\Theta)$, where $\mathcal{X}$ are the particular regions visited during a navigation, Θ the distribution parameters, $P(\mathcal{X}|\Theta)$ the likelihood, the probability of having a particular navigation $\mathcal{X}$ given a distribution with parameters Θ and $P(\Theta)$ is the *a priori* probability of the parameters. Provided that there exist k probable regions that can be independently visited during a navigation, the success of visiting exactly a single region at a time for N independent trials is estimated as a multinomial distribution, with the same probability of success for each trial since statistical independence is assumed. The image is divided into M disjoint regions whose state of visits is stored at each of the navigation steps[2]. A navigation is then defined as a set of steps, each storing the location of the visited regions. The probability of visiting the region i, $P(x_i|\Theta)$, given the relevance distribution parameters is estimated assuming independence. The probability of visiting certain region r can be modeled as a multinomial probability:

$$P(x_1, \ldots, x_k|\Theta) = Multi(X|\Theta) = \prod_{i=1}^{k} p_i^{x_i} \tag{4-1}$$

where $x_1, x_2, ..., x_k$ are binary variables that indicate which region is visited, i.e., if the region visited is r then $x_r = 1$ and $\forall i \neq r$, $x_i = 0$. The parameter vector Θ is the set of probabilities $\{p_1, p_2, ..., p_k\}$, being p_i the probability of visiting the region i. A navigation is a path composed of N steps, being the step x^t a set of states $x_1^t, x_2^t, ..., x_k^t$ of the WSI spatial regions, with $\sum_{i=1}^{k} x_i^t = 1$, i.e., just a single region is visited at each step. The resultant probability is given by:

$$P(\mathcal{X}|\Theta) = \prod_{i=1}^{k} p_i^{N_i} \tag{4-2}$$

[2]at a particular navigation step, a state of visit is one if that region has been visited and null otherwise.

where $N_i = \sum_{t=1}^{N} x_i^t$ stands for the number of times the *i-th* region is visited during a navigation. This information is obtained by recording the expert pathologists in the vector $n = (N_1, N_2, ..., N_k)$.

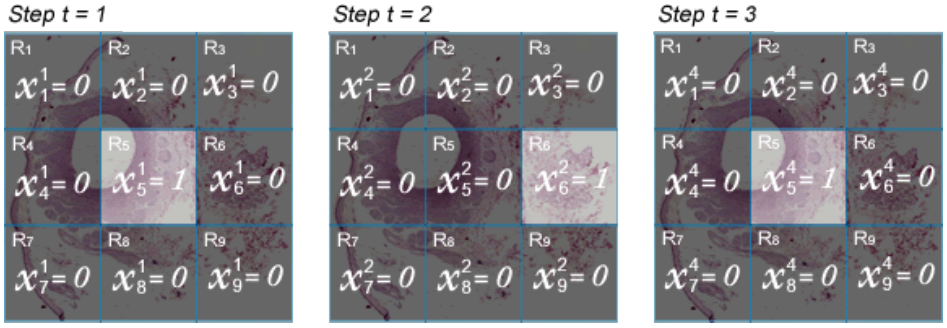


Figure 4-1: WSI navigation modeled as multinomial probability. The Figure illustrates a WSI divided into 9 regions, which is navigated in three steps. At each step, a single region is visited. In this example, at the end of the navigation, region 6 was visited just once, so $N_6 = \sum_{t=1}^{3} x_6^t = 1$, and region 5 was visited twice (step $t = 1$ and $t = 3$), then $N_5 = \sum_{t=1}^{3} x_5^t = 2$.

Under a Bayesian inference framework, the multinomial distribution parameters $\{p_1, \ldots, p_k\}$ are modeled as random variables that are described by a Dirichlet Distribution, i.e., the multinomial conjugate. Hence the prior distribution is:

$$P(\Theta|\alpha) = Dirichlet(\Theta|\alpha) = \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_k)} \prod_{i=1}^{k} p_i^{\alpha_i - 1} \tag{4-3}$$

where $\Theta = (p_1, p_2, ..., p_k)$, $\alpha = (\alpha_1, \alpha_2, ..., \alpha_k)$ and $\alpha_0 = \sum_{i=1}^{k} \alpha_i$. Given these prior and likelihood, the posterior, which also has a Dirichlet distribution is given by

$$P(\Theta|\mathcal{X}, \alpha) = \frac{\Gamma(\alpha_0 + N)}{\Gamma(\alpha_1 + N_1)\cdots\Gamma(\alpha_k + N_k)} \prod_{i=1}^{k} p_i^{\alpha_i + N_i - 1} \tag{4-4}$$

where $\alpha = (\alpha_1, \alpha_2, ..., \alpha_k)$ is a vector with the initial parameters of the prior distribution given by the saliency map and α_i is the initial visit probability assigned by the VAM to the *ith* region. Θ is estimated using a Bayesian parameter estimation:

$$\Theta_{Bayes} = E\left[\Theta|\mathcal{X}, \alpha\right] = \frac{\alpha_i + N_i}{\alpha_0 + N} \tag{4-5}$$

Two variations were introduced, a discrete relevance map that takes into account the number of times a region is visited, and a structured relevance map that models the WSI spatial regions as nodes and their relationships as edges. Hereafter, both models are further described.

The discrete relevance map

The relevance map probability is assigned to each region according to the number of times that region was previously visited, i.e., RoIs are those that are more frequently visited.
Taking into account equations (4-4) and (4-5), an estimated parameter for each region is

$$P(p_i|\mathcal{X},\alpha) = \frac{\Gamma(\alpha_0+N)}{\prod_{i=1}^{k}\Gamma(\alpha_i+N_i)}\left(\frac{\lambda\alpha_i+(1-\lambda)N_i}{\lambda\alpha_0+(1-\lambda)N}\right)^{\alpha_i+N_i-1} \tag{4-6}$$

where λ is a parameter that controls the relative importance of the prior with respect to the likelihood. The higher λ is, the more important the prior.
With a prior α and a navigation n_1, an estimate θ_1 is calculated, if a new navigation n_2 is available, a new estimate θ_2 is calculated using θ_1 as prior, and so on.

The structured relevance map
In the present context, searching particular RoIs amounts to figure out a preferential information flux through a net of nodes belonging to a fully-connected graph, being each node a particular histhological region and each edge a similarity (or dissimilarity) measure among regions. If an image is partitioned and its parts are somehow associated to a fully-connected graph, the interaction process between an user and the image is defined by a particular path of that graph.
The structured relevance map herein proposed performs a flexible WSI splitting that may represent any relevant region. This model defines a graph $G = (V, E)$, where V is the set of regions of the WSI with a relevance score associated and E is the set of edges representing the hierarchical relevance relationships among nodes. A first step is then to build a discrete saliency map and structure it as a graph. For doing so, the saliency map of an image is obtained from a bottom-up VAM. Then, a set of local minima are determined by a two-dimensional first-order derivative of the saliency map; each local minimum determines a saliency level which defines an image region with a similar relevance value (iso-saliency). Every iso-relevant region is represented by a node of the graph which connects other regions or nodes by relevance levels. The resulting graph corresponds to a tree that encodes the hierarchical relevance relationships of the image regions (See Subfigures **4-2**(a) and **4-2**(b)).
The idea is to enrich the tree structure described above with information from the pathologists' navigations. For achieving so, the path followed by an expert pathologist during a diagnostic exploration is recorded and the posterior probability $P(\mathrm{v}|\mathcal{X})$ is estimated given a navigation sample $\mathcal{X}$, where $v = v_1, v_2, \ldots, v_k$ is the probability of visiting the node i, modeled as a Dirichlet probability:

$$P(v_i|\mathcal{X},\alpha) = \frac{\Gamma(\alpha_0+N)}{\prod_{i=1}^{k}\Gamma(\alpha_i+N_i)}\left(\frac{\lambda\alpha_i)+(1-\lambda)N_i)}{\lambda\alpha_0+(1-\lambda)N}\right)^{\alpha_i+N_i-1} \tag{4-7}$$

where $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_k)$ is the set of initial parameters of the prior distribution given by the discrete saliency map.
During a navigation then, a visited region may overlap an image area corresponding to several nodes or just a small portion of the area associated to a single node, a more accurate

estimation of the probability distribution may be obtained considering not only the number of times an area is visited (as for the discrete relevance map), but also the intersected area. That is to say, the relevance of an area depends also on its Euclidean distance to an overlapped area, in this way, the more distant non-overlapped areas are, the lower the relevance value is. A smooth saliency region function redistributes the probability mass not only within the tree but also at the level of the node itself. This is modeled as:

$$\gamma = \frac{\Gamma(\alpha_0 + N)}{\prod_{i=1}^{k} \Gamma(\alpha_i + N_i)} \tag{4-8}$$

$$\delta = \frac{1}{\mathbf{D}_{(x,y)}(p,q)} \tag{4-9}$$

$$P(v_i(x,y))|\mathcal{X},\alpha) = \delta\gamma(\frac{\lambda\alpha_i(x,y)) + (1-\lambda)N_i(x,y))}{\lambda\alpha_0 + (1-\lambda)N})^{\alpha_i+N_i-1} \tag{4-10}$$

Where $\mathbf{D}_{(x,y)}(p,q)$ is the euclidean distance between the current pixel (p,q) and the nearest pixel (x,y) overlapped by the navigation path. Afterwards, a normalization is performed and the graph hierarchy is updated: the nodes corresponding to the more visited regions will have a higher relevance score. A new saliency map is built by the union of every region, as shown in Figure **4-2**. As soon as a new navigation is available, the probability is redistributed. Again, the λ parameter weights the probability maps.

Experimentation

Experimental setup

The evaluation dataset consisted of a total of twenty skin biopsies of patients diagnosed with different types of basal cell carcinoma, embedded in paraffin and stained with Hematoxylin-Eosin[3]. The WSI resolutions vary from 104 to 340 mega pixels, and their sizes range between 630 MB and 972 MB. Four pathologists, with at least ten years of professional experience, navigated these virtual slides and diagnose them using a custom virtual microscope (each pathologist was trained to use the virtual microscope with two test images). The twenty WSIs were randomly displayed. Each pathologist was asked to run over the WSIs up to a diagnosis was set. During examinations, the pathologist interactions (Visited WoIs, time spanned at each RoI, magnification level changes) were recorded for creating navigation heat maps. A navigation heat map shows how often a pixel was visited by a pathologist during a diagnostic task. Each image pixel has a visit counter, initially set to zero. When a pathologist is looking at a certain window, the all the pixels of such region increase their visit counter. Similarly, when a pathologist zooms in a region, the visit counter of the pixels belonging to the magnified region will increase. Thus, pixels within magnified regions will have higher visit counter values than pixels on other regions.

[3]The set of histological samples was provided by the Pathology Department of the Universidad Nacional de Colombia and is representative of what pathologists usually observe in this pathology

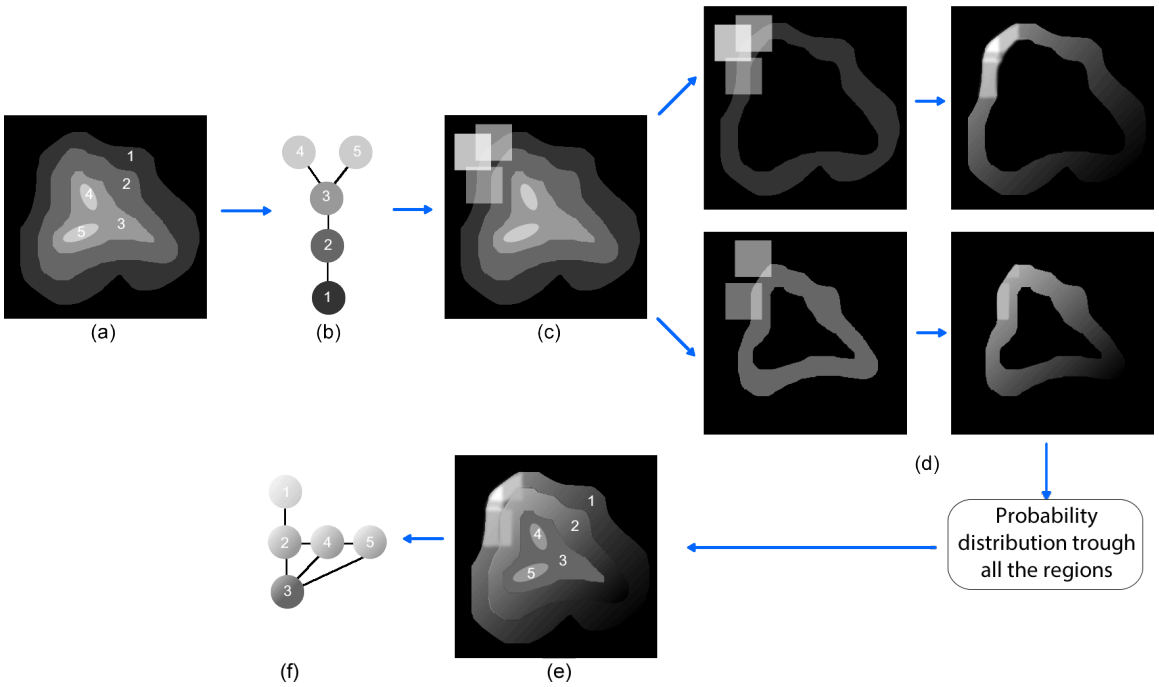


Figure 4-2: Redistribution of probabilities in the structured relevance map. (a) and (b) correspond to the saliency function and its corresponding graph. (c) A recorded navigation (gray-level squares) overlapping the discrete saliency prior. (d) Intra-region probability distribution by multinomial analysis and distance function. (e) and (f) final relevance map and its corresponding graph representation.

Evaluation and Results

The two presented strategies stored the most relevant regions and evaluation consisted in checking out whether the pixels requested during a navigation were or not available in that cache space (percentage of cache hits).

In the first experiment, the prediction capability of the discrete relevance map was evaluated using a fixed cache size, corresponding to the 5 % of the image size, and 9 different λ values that assessed the importance of the prior with respect to the evidence. First, an initial experiment using the saliency map calculated from the bottom-up VAM was performed. A first posterior probability $P(\Theta_1|\mathcal{X}_1)$ was calculated using a navigation and tested for a second navigation. Then, a second and a third posterior probability $P(\Theta_2|\mathcal{X}_2)$, $P(\Theta_3|\mathcal{X}_3)$ were calculated using the previous navigations and evaluated for a new one. Results in Figure **4-3** show that When using the VAM approach, the percentage of cache hits is constant for the different values of λ, about a 13 %. On the contrary, when the first posterior probability is calculated, the percentage of cache hits increases to 21.6 %, 8 % higher than the VAM, which may amount to a delay of about 12,5 s for an image of 458 megapixels and 24 bits-per-pixel being transmitted on a network with a speed of 3.5 Megabits/second. If a second navigation is involved and a second posterior probability is calculated, the percentage of

cache hits reaches a 22.4 % while for a third posterior probability, this figure is 22.8 %. These results demonstrate that integration of new knowledge increases the predictive capability of the model, especially for a λ value of 0.7, while larger values of this λ make that performance decreases, likely due to the lack of more evidence.

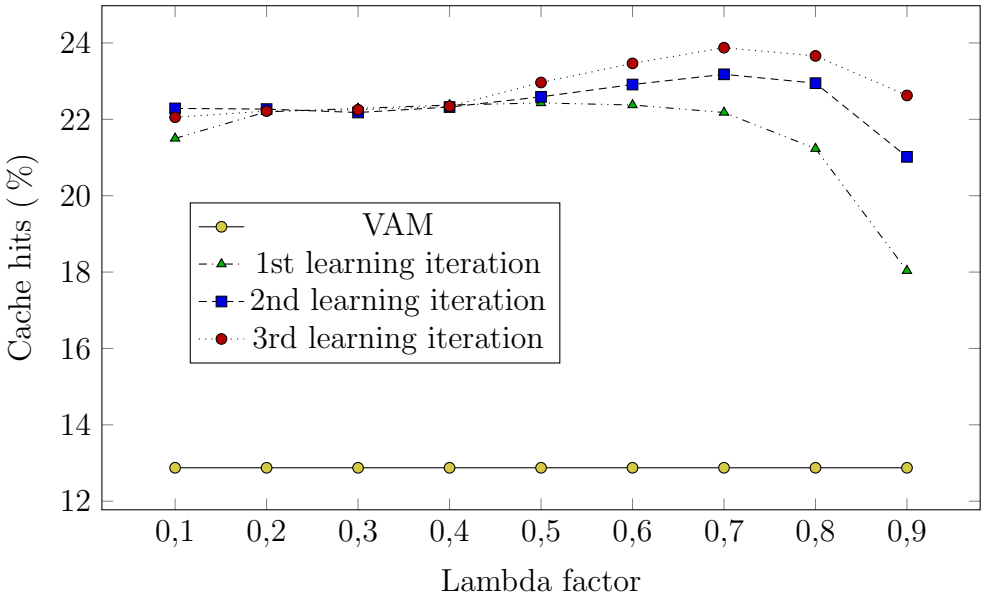


Figure 4-3: Prediction ability of the discrete relevance map using a cache size of the 5 % of the image size. Results using the VAM and after one ($P(\Theta_1|\mathcal{X}_1)$), two ($P(\Theta_2|\mathcal{X}_2)$) and three ($P(\Theta_3|\mathcal{X}_3)$) learning iterations. Different λ values were used to model the importance of the prior with respect to the evidence.

In a second test, the two proposed Bayesian models were compared. A cross-validation leave-one-out scheme was implemented: For each image, two navigations were used to build and train a relevance map with the discrete and structured relevance maps, and one navigation was used to measure the percentage of cache hits. In this case, performance was measured as the difference of hits for different cache sizes between the two methods. Results, shown in Figure **4-4**, demonstrate that the structured relevance map has more cache hits than the discrete relevance map. For a cache size of 1 % of the image, the difference is about 1.8 millions of pixels, which might represent a delay of about 12 s for a 458 megapixels and 24 bits-per-pixel image being transmitted on a network with a speed of 3.5 Megabits/second. For a cache size of 20 %, the difference is about 14.5 millions of pixels, which may represent a more than 1,66 min for an image and a network with the previously described features.

4.2.2. Integrating nuclear data and interactions of experts

In Section 4.2.1, a Bayesian framework predicted regions an expert might visit based on the visual data and previous visits of other experts, attempting to improve the cache perfor-

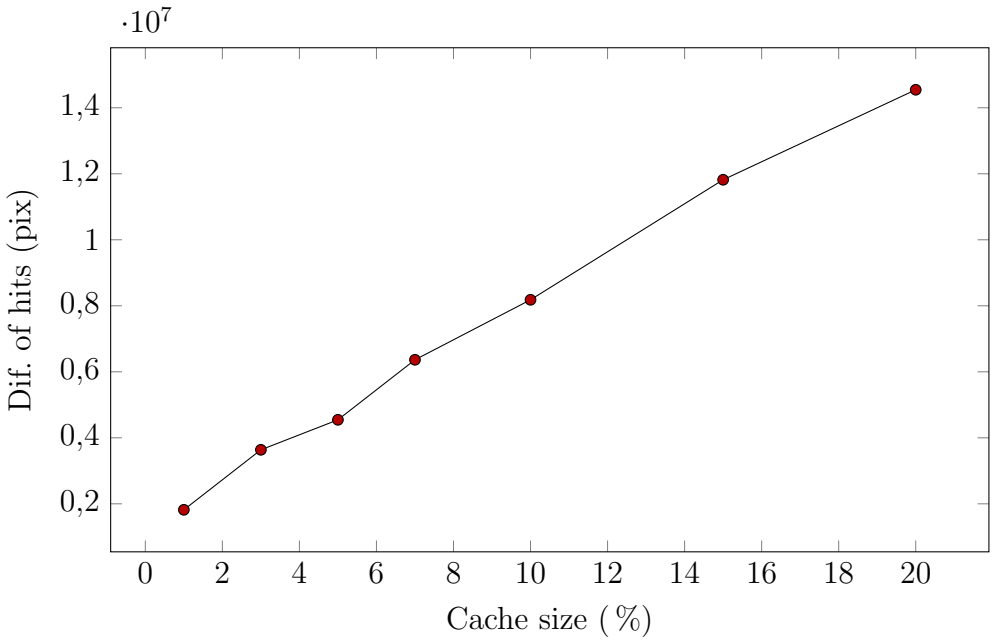


Figure 4-4: Evaluation of the discrete and structured relevance maps. Performance is measured as the difference of pixels hits for different cache sizes between the two methods. The cache size corresponds to a percentage of the WSI size.

mance for different navigation tasks. This approach begins by learning a set of candidate relevant regions using salient information coming from visual features such as color, shape, and orientation. These candidate regions are then pruned based on feedback information from the navigation trends of the pathologist. Unfortunately, this approach is limited because the learned models are specific for each image, so they hardly could be generalized.

For this reason, this section introduces a model that aims to integrate low-, mid- and high-level information to determine how likely a nucleus is cancerous. The model exploits the implicit knowledge extracted from actions performed by pathologists while navigating WSIs. Briefly, the method starts by segmenting the WSI nuclei and computing a set of nuclear visual features typically altered in regions with cancer, namely color, size, and spatial distribution [166]. Then, each nucleus is assigned a likelihood of being cancerous based off the number of times said nucleus is visited by a group of pathologists while performing a diagnostic task. Finally, the relative importance of each of the visual features is calculated by performing a multi-linear regression between the visual features associated with each nuclei and the corresponding likelihood of cancer presence. The primary contribution of this work is a simple and adaptable integration of the high-level expert knowledge (implicitly extracted) with handcrafted visual features in order to create a classifier for cancer detection on a patch-by-patch basis within WSIs. The underlying rationale here is that prior information, derived from pathologists' interactions with the WSI, improves the predictive performance of handcrafted features. The model is trained offline with a set of WSIs and then applied to

classification/prediction task in new images (previously unseen). A second contribution of this method is that it may provide an alternative approach to defining expert ground truth since this method enables identification of areas that represent a high level of interest. Figure **4-5** illustrates the workflow of the new approach to identify suspicious RoIs from whole slide images of basal cell carcinoma.

The presented method was evaluated in terms of its ability to determine if a set of patches or fields of view (FOVs), extracted from WSIs, contained cancerous regions. The validation of our approach was performed on a study case involving the detection of basal-cell carcinoma (BCC) from WSIs. BCC is the most common malignant skin cancer [38]. Although skin cancer is a threat to human life, fortunately it can be cured if it is detected and removed before it makes metastasis [144]. However, detecting BCC is challenging since it is difficult to distinguish the actual tumor from its surrounding noncancerous tissue [10]. If BCC remains undiagnosed, the tumor can grow, requiring more extensive resection that can result in visual or functional deformities. Currently, diagnosis relies on the experience and subjective judgment of each pathologist [10]. The development of objective measures for BCC detection could help improve clinical decision support approaches for disease diagnosis and could also potentially facilitate better strategies for treatment planning. The present paper shows how integration of navigations and low-level image features could help to identify cancerous regions in BCC WSIs.

Low-level model

Nuclear Segmentation

The low-level image features employed in this work are based off spatial nuclear architecture and arrangement. Consequently, the first challenge in extracting these measurements is carefully identifying these nuclei on digitized pathology images.

While a number of different nuclear segmentation methods for digitized pathology images have been previously presented [4, 50, 78, 177], we opted to go with the method of Wienert et al. [173]. This strategy, whose major advantage is its simplicity, involves the use of an adaptive tracing technique that detects contours which are filtered based off the strength of the image gradient. The convex-hull of each detected region is employed to separate out nuclei clusters. Nuclei classification is, subsequently, performed on these clusters in order to 1) distinguish between overlapping and conjoined nuclei and 2) distinguish nuclei from similar appearing primitives such as lymphocytes. According to the authors [173], with a validation set containing 7931 manually annotated cells from 36 images of different organs, this segmentation method presents a precision = 0.908 and recall = 0.859. After these results, one could expect that about a fifteen percent of the population might be missed; however, the fact that the segmentation process is applied to the complete WSI makes that in practice the probability distribution function of these features would change very little by the misdetected nuclei.

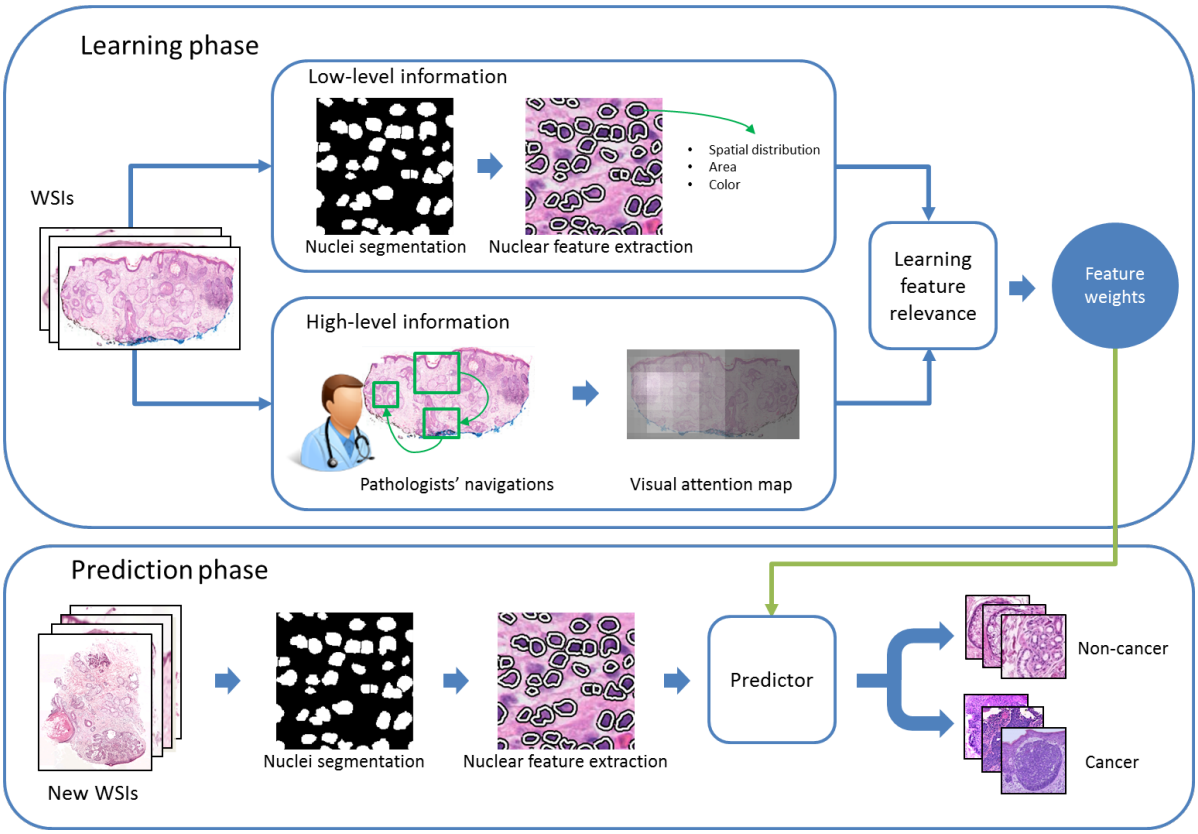


Figure 4-5: Illustration of the process of identifying cancer RoIs by fusing information from handcrafted features and interactions with the pathologist. In the learning phase, two types of information are extracted from a set of WSIs: low and high-level. Low-level features originate from the visual content of the image, specifically from nuclei arrangements and their sizes and colors. High-level information, such as magnification-weighted location focus, come from the visual attention of the experts. These two information sources are then integrated by a multilinear model that learns the relative importance (weights) of each handcrafted feature for predicting RoIs, i.e., cancer areas. In the prediction phase, low-level features are extracted from new WSIs. Using the learned weights, a nuclei cance likelihood is calculated to predict whether a specific patch is either cancerous or not.

Nuclear feature extraction

Experts frequently base their diagnosis of disease presence or absence on nuclear features [166]. Consequently, nuclei represent critical visual information for analysis in digital histopathology. For this reason, we attempted to determine the degree to which a nucleus might represent cancer. Identifying individual nuclei that represent cancer might in turn enable identifying regions involved with cancer. For this reason, each nucleus was assigned a likelihood of being cancerous or not based on its visual properties; hence, features that are characteristic of morphologic disruption in cancer were extracted, namely attributes pertaining to spatial distribution, size, and color of the nuclei.

Since cancer is characterized by rapid cell proliferation and the formation of cell clusters, our model was designed to assign a likelihood to each nucleus according to its grade of grouping. A high value for an individual nucleus means that it is close to other nuclei and has a high probability of being cancerous. The model is defined as follows: Once nuclei are segmented, the center of gravity of each individual cell is associated with a node in a graph which represents nuclei spatial relationships. The weights of the graph edges correspond to the inverse of the Euclidean distances between the connected nuclei. The likelihood of a node being cancerous is then computed by adding the corresponding weights of every edge connecting to that node. In this way, this feature captures the local spatial distribution of cells. Particular attention was paid to setting a threshold value that prevented large dense clusters from dominating the likelihood distribution. If the analysis is performed without such threshold, the probability mass of the likelihood is basically concentrated into very few foci, typically one or two. The idea behind this threshold is to achieve a likelihood distributed as homogeneously as possible. For this reason, this threshold was set to 0.02, a value that ensured a number of foci reached a number of Gaussians with similar probability mass. This value was established after an exhaustive search for the optimal threshold values on a subset of images from the training set. This model is defined by Equation

$$x_{1,i} = \sum_{j=1}^{n-1} \frac{1}{D_{i,j}} \times b_\theta, \tag{4-11}$$

where $x_{1,i}$ is the likelihood of the spatial distribution-based feature for the i-th nucleus, n the number of nuclei, and $D_{i,j}$ is the Euclidean distance between nuclei i and j, θ is the threshold value, and b_θ is a binary value that is 0 when $D_{i,j} > \theta$ and 1 otherwise. This model generates very high values if nuclei are close together or overlapping, a useful condition for the model detects clustered nuclei. It should be strengthen out that these values are calculated between centroids and therefore the minimum distance between two nuclei is at least the nucleus radius, in consequence this quantity will never diverge.

Cancer cells are also characterized by the proliferation of cells with large nuclei. For this reason, our method employs a model that clusters cells based on their area in order to highlight homogeneous neighborhoods. In this process, the area of each nucleus is computed

and modeled as a graph in which nodes are the nuclei and weights are the inverse of the differences between areas. Here, the likelihood of each nucleus being cancerous corresponds to the sum of the weights of the nodes connecting to it. High likelihood values typically are ascribed to cells with large areas surrounded by other similar cells.
It is also known that cancer cells and their environment (surrounding area) present slight differences in color when compared to normal cells. Taking this into account, the averaged YUV color values of each nucleus and its surrounding area are also extracted. The area surrounding the nucleus is defined as a circumference with radius $\times$1,5 the nucleus major axis length but excluding, of course, the nucleus itself and any other intersecting nuclei. This value aims to model a nucleus-cytoplasm relationship that usually is 1:4, but in the present investigation was set to 1:3 trying to take into account the shrinkage effect of the whole histological procedure (approximately 25 % according to [174]).
Finally, after applying these models and the subsequent feature extraction, a set of 8 feature values were obtained for each nucleus. These included spatial distribution, size, mean nuclear Y, U, and V values, and mean Y, U, and V values of the region surrounding each nucleus.

High-level model

We had a set of pathologists perform a series of diagnostics tasks on WSIs and captured the coordinates of the window of interest they visited and the type of action the expert performed (panning, zooming in or out). All of this information was recorded for each reader and used to generate the respective visual attention maps.

Visualization strategy and tracking system

Given the size of the WSIs, the most common visualization strategy consists in constructing a multi-resolution pyramid of the image, compressed in JPEG format; however, this approach presents important limitations regarding management, quality, flexibility and performance [3, 123, 36]. However, the JPEG2000 (J2K) standard, characterized by a natural multi-resolution decomposition, lossless compression, and random spatial access [75, 156], better suits the requirements for this type of interaction. In [36], the J2K model was adapted to improve the dynamic interaction between WSIs and a custom virtual microscope enabled to run on small devices like smartphones or tablets. The principle of this navigation is that the client requests RoIs instead of the entire image, thereby reducing the amount of information requested. When users navigate a particular WSI, they have as reference a window of interest that is moved along a navigation path, and whose coordinates and magnification are encoded as spatial coordinates. These coordinates define the J2K packets needed to construct the specific requested region. The packets are then extracted from the compressed file, combined with the region parameters and the image main header, decompressed using a J2K decoder, and used by the client to reconstruct the original RoI. Here, this strategy was adapted and used to enable pathologists to visualize different WSIs.

Furthermore, the visualization system was tuned to register the different navigation movements and actions performed by the pathologists during the diagnostic task. The tracking system records the time, location, and magnification activity along the entire diagnostic path. This information is subsequently stored for offline analysis.

The visual attention map

Visual attention maps are a summary of the most visited regions of an image by a set of experts during a diagnosis task. In this work, the visual attention map is built using the frequency with which each image pixel is visited; this approach assumes that regions most visited by a set of pathologists are those likely containing cancer. Information related to the time pathologists spend at examining a particular region was ignored by this model since it could be highly noisy. Generation of this map starts by setting to zero a visit counter for every pixel. Each time a pathologist visits certain region, the visit counter of each pixel belonging to such a region increases (See Figure **4-6** for illustration). Consequently, pixels from regions more frequently visited will have higher visit counter values. The final visual attention map is generated by summing all the visits from every pathologist, followed by a normalization process that sets 1 to the highest value and 0 to the lowest. From this map, we infer the likelihood of a cancerous nucleus depending on its location, i.e., a nucleus in a frequently visited area will have a higher likelihood value than a nucleus in a non-visited region.

Learning feature relevance

As previously stated, traditional approaches usually employ only low-level features for building machine learning classifiers for disease diagnosis. While these strategies have showed good results at discriminating between classes, they disregard the concepts behind the image and ignore the expert domain knowledge.

In this work we employ a mathematical formulation that enables the integration of higher-level expert knowledge with lower-level image derived features. The set of low-level features used to determine the degree to which a nucleus is cancerous can be thought of as the prior class conditional information, invoking the Bayesian formulation. Additionally, information about which nuclei are decisive in rendering a diagnosis can be implicitly extracted from the visual attention map, information that can be thought of as evidence or class prior. Thus the likelihood function of an individual nuclei was defined as a linear combination of its visual features as,

$$y = \sum_{k=0}^{f} w_k x_k, \tag{4-12}$$

with x_k the k-th feature vector, x_0=1 (the bias factor) and w_k the corresponding weight (the model parameters to be learned), f the number of different features, and y the resulting

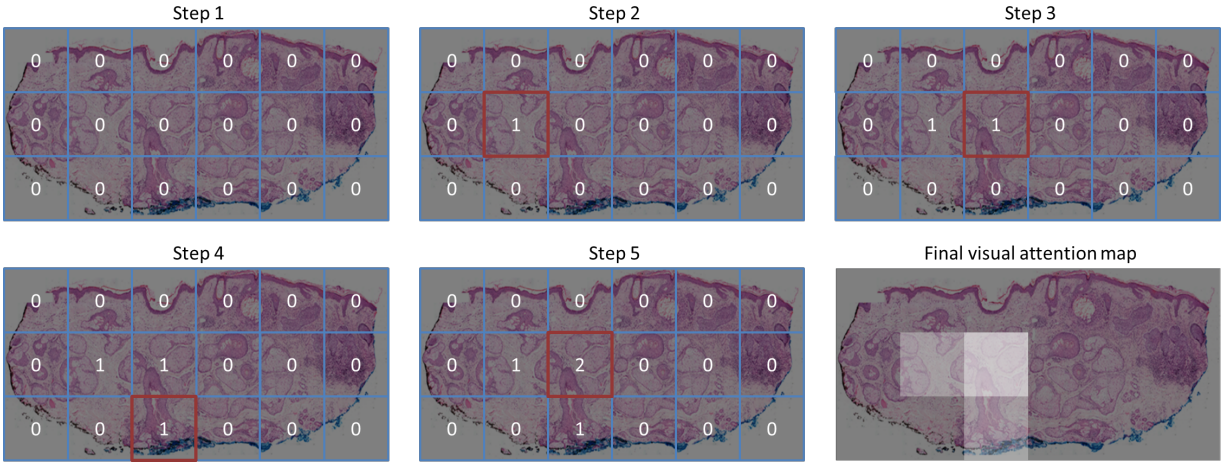


Figure 4-6: Generation of the WSI visual attention map. The visual attention map is generated at a pixel-level. Every pixel belonging to the navigation window is set to the relevance level, defined as the number of times said pixel was included within one of the observation windows. In this example, a pathologist exploring the WSI changed the observation window. Each time this happened, the number of visits was continuously updated for every pixel within the observation window. In step 1, all image pixels were initialized with a "visit counter."of 0. Steps 2-4 show visits to different image regions (red squares); in these cases, the visit counter of the pixels within the visualized regions increases. In step 5, a region is re-visited and the visit counters are incremented for every pixel within the observation window. Finally, a visual attention map is obtained from the total number of pixel visits. The more frequently a region is visited, the more likely it is cancerous (brighter).

likelihood. The bias term is usually introduced in multilinear regression models to avoid the regression line is forced to pass through the origin. In other words, if the bias is not used, when all of the feature regressors or predictors are zero, the predicted likelihood value should be zero, a meaningless statement in this context since nuclei relevance is always non null by other features than can have any influence on such likelihood value. So, a bias term independent of the regressors was added to allow the hyperplane described by the learned weights to naturally capture the statistical relations.

In order to integrate this information with the expert knowledge, the information extracted from the visual attention map was used as the objective function in this linear model. The weight or relative importance of each feature is then learned by performing a multi-linear regression using the least squares method [115].

Model validation

Validation of the presented approach was carried out using 24 patient studies as described below (See Subsection "Dataset"). Firstly, the feature weights were estimated using the model described in Section 4.2.2. This process was carried out using the 70 % of the cases (17 randomly selected) (See Figure **4-7**). Secondly, each of the images within the remaining 30 % (7 images) were split into smaller patches or fields of view (FOVs) with a size of 1024×1024 pixels. This FOV typically contains most of the relative primitives and surrounding structures needed to render a diagnosis by a pathologist. This FOV size was empirically determined by the pathologists. FOVs with no tissue were discarded. Next, each FOV was labeled by an expert pathologist as cancer or not-cancer. The number of FOVs is dependent on the WSI size, i.e., the larger the WSI the higher the number of FOVs. Once all the FOVs are collected, the number of samples within the cancer and noncancer classes were balanced by randomly removing a set of FOVs with the most represented class. 176 FOVs corresponding to 5 patients were used to train a classifier to predict the presence of cancer on a patch-by-patch basis while the remaining 98 FOVs, from 2 patients, were used for independent testing.

The first step to predict the cancer presence in a FOV was to calculate the likelihood of being cancerous for each of its nuclei. For this purpose, each nuclear feature was multiplied with its corresponding weight in the weighting vector (previously learned). Features that were positively correlated with cancer were assigned larger positive weights, features inversely correlated with cancer were given negative weights, and uncorrelated features were assigned weights of 0. Thus, the numerical result of multiplying the feature and weighting vectors was a direct indication of the likelihood that each nucleus was cancerous. Finally, the likelihood of a FOV being cancerous is calculated from the average likelihood of its individual constituent nuclei.

Baseline

The model described in this work (M_im) (defined by Equation 4-11) attempts to infer the relevance of visual features from implicit knowledge, extracted from interactions between a group of pathologists and WSIs. The comparative strategy or baseline model (M_ex) uses the very same visual features as M_im (See Subsection "Nuclear feature extraction"), but the estimation of weights is performed from explicit knowledge, i.e., nuclei manually labeled by an expert as cancerous or non-cancerous (See Figure **4-8**). The baseline feature weights are used to predict the likelihood of cancer on a per-nucleus basis, across the different WSIs. Finally, from these WSIs, a set of FOVs is extracted and a classifier is trained to predict the presence or absence of cancer at the FOV level.

Dataset

The dataset consisted of Hematoxylin-Eosin (H&E) slides from patients diagnosed with BCC.

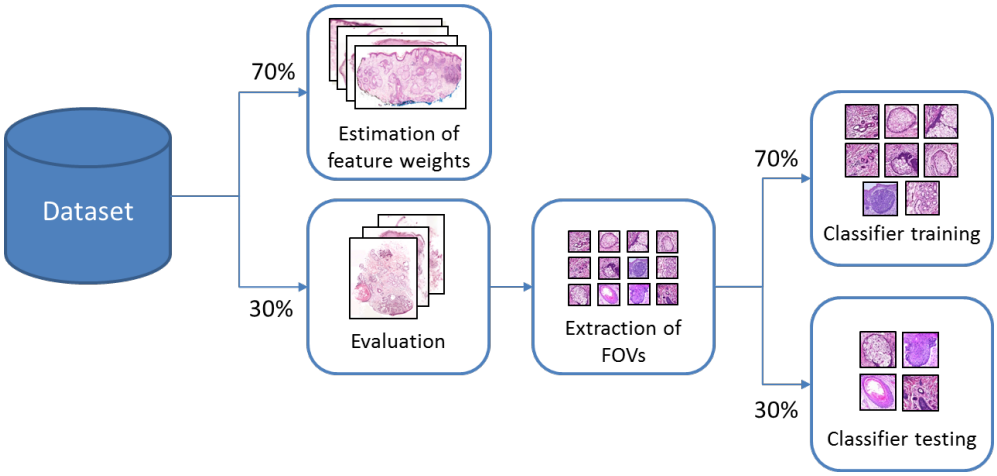


Figure 4-7: Dataset partition

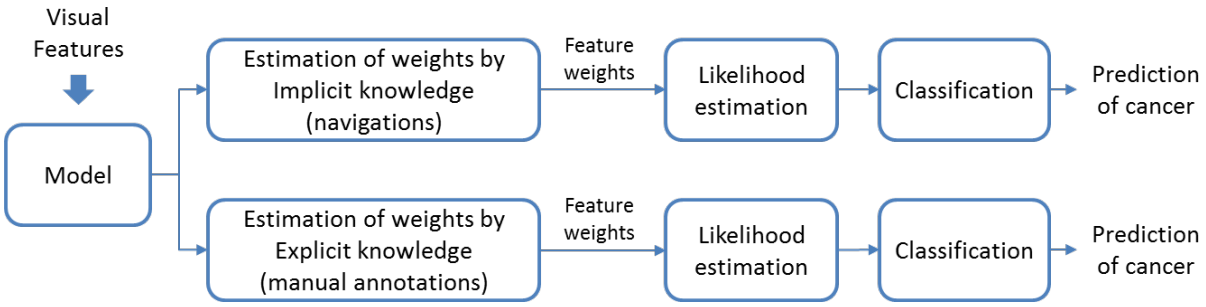


Figure 4-8: Validation methodology

These samples exhibit unique and mixed BCC types, namely superficial, nodular, micronodular, morpheaform, and trabecular. They were collected from the Pathology Department of Universidad Nacional de Colombia between 2009 and 2014 and were randomly selected from a set of 98 patients previously diagnosed with BCC and for whom slides were available. All the cases were anonymized. The study cohort was limited to 24 cases, each from a different patient, owing to the time constraints involving digitizing, manually annotating the samples, and recording the navigations of 4 different pathology readers. The slides were digitized at 40× using a tri-ocular CARL ZEISS Axiostar plus microscope coupled to a DXM1200 Nikon color digital camera, controlled by a custom motorized scanner.

Experimental Results

The visual attention map

Four pathologists with at least ten years of professional experience navigated 17 WSIs (70 % of all the cases) using a customized virtual microscope. Each pathologist was asked to exa-

mine each WSI until a diagnosis of disease presence or absence had been made. The set of WSIs used in the present investigation were determined to be of intermediate diagnostic difficulty, as defined by our dermato-pathologist. The test demanded only a general diagnosis and all four pathologists reached the same diagnosis in each case. Overall, the four pathologists explored mostly the regions with cancer and usually the difference in diagnostic times was due to the difficulty in finding the cancerous region. Once this region was located, pathologists frequently focused on local details that supported a refinement of the diagnosis. This navigation profile was not strongly impacted by the type of carcinoma in each slide; i.e. whether it was any one of the unique and mixed BCC types, namely superficial, nodular, micronodular, morpheaform, and trabecular.

Navigation patterns presented by pathologists during diagnosis tasks present a very high variability, associated to different factors such as the experience level or the complexity of the visual patterns to identify. Although patterns may be highly variable, previous works have shown that a group of experts exploring the same histological slide tend to visit similar locations [62]. Different pathologists navigate the same WSI assuming that the intersection of their navigations mitigate the noise generated by a single navigation. While four pathologists were part of the experiment, results of previous works suggest that including more pathologists has the potential of improving even more these results [35].

Learning feature relevance

As defined in Section 4.2.2, the likelihood a nucleus is cancerous is given by the linear combination of its features and the weights of each feature. In order to learn these feature weights, the least squares method was applied to approximate the likelihood of each nucleus by employing the visual attention map. This estimation process was performed using the *learning set*, i.e., 70 % of all the cases.

The least squares method was used to determine the feature weights; the average mean square error for the regression fit was 4.88 %. Interestingly, the spatial distribution-based feature has the highest weight among all the features. This makes intuitive sense given that cancer is characterized by high cellular proliferation and cluster formation. Figure **4-9** shows some visual results of images whose relevance values were predicted using the learned weights. Table **4-1** shows the weights learned for each of the image features considered in this study.

Model validation

From the *evaluation set*, i.e., 30 % of all the cases (7 patients), 274 FOVs were extracted. 176 FOVs (from 5 patients) were used to train an SVM to predict the presence of cancer on a patch-by-patch basis while the remaining 98 FOVs (from 2 patients) were used for independent testing.

The accuracy, area under the receiver operating characteristic curve (ROC), and model performance were compared against the predictions made by a model trained with explicit manual annotations.

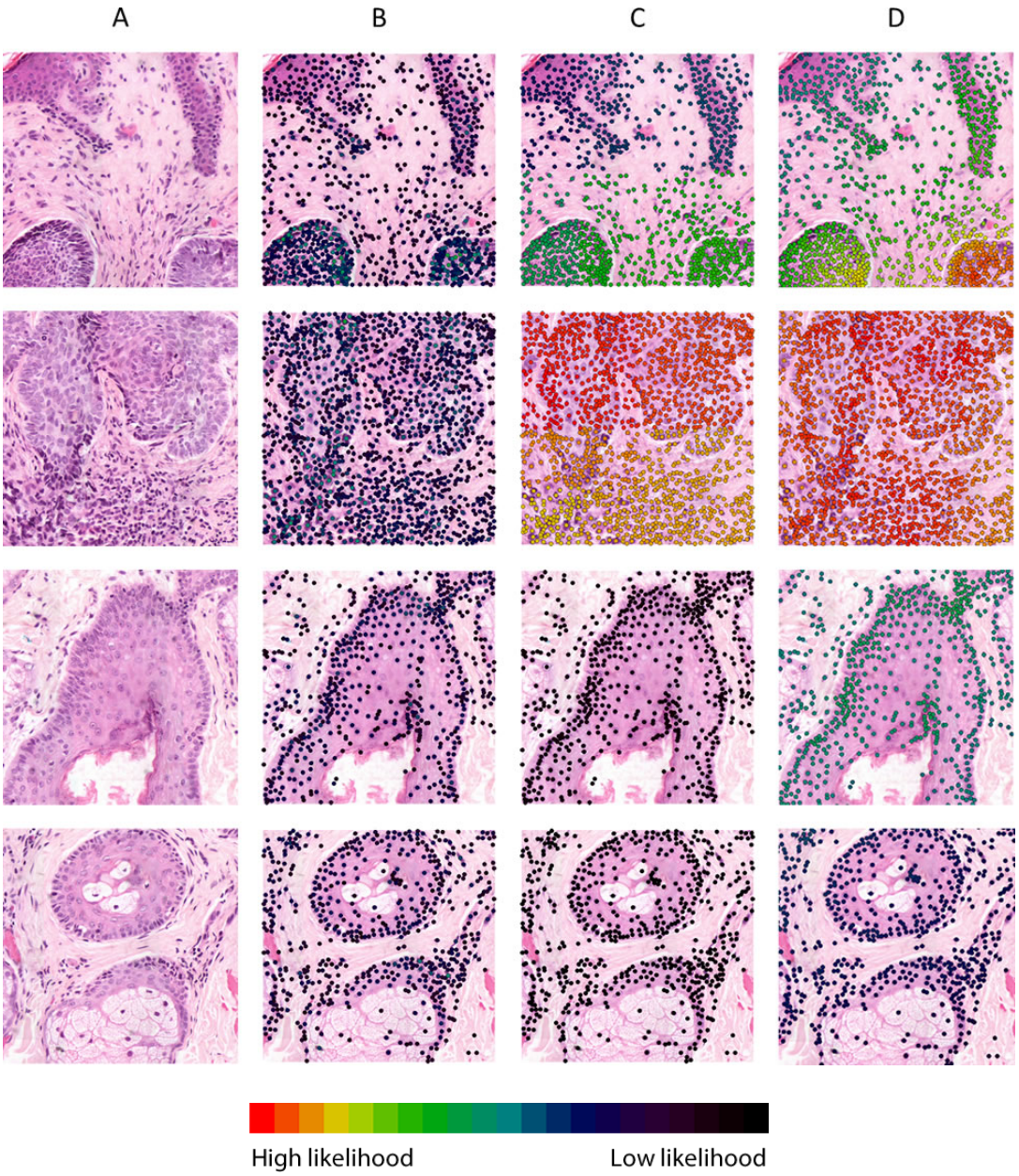


Figure 4-9: Examples of the predicted likelihood for representative fields of view. The color of each nucleus (dot) represents the relative likelihood of each nucleus being cancerous (a heat map color palette was used, in which blue and red represent low and high relevance, respectively). Column A presents the original image, column B shows the likelihood based on low level features, column C displays the likelihood inferred from navigations of pathologists, and column D illustrates the final likelihood calculated with the presented approach. First row is a visual field with two structures: a carcinoma nodule (bottom) and epidermis (top), second row corresponds to a cancerous area, and third row corresponds to a follicle, a normal/healthy structure.

Feature	Weight
Bias factor	0.3623
Spatial distribution	0.2272
Size	0.0231
Mean nuclear Y	0.1142
Mean nuclear U	-0.0354
Mean nuclear V	-0.0292
Mean Y of the region surrounding a nucleus	-0.0624
Mean U of the region surrounding a nucleus	-0.1355
Mean V of the region surrounding a nucleus	0.0146

Table 4-1: Learned weights for each visual feature.

Results, presented in Table 2, show that classification performance is slightly higher when using nuclear features weighted by visual attention information, in contrast to using a model trained with manual annotations.

	Acc.	Precision	Recall (Sensitivity)	F-measure	Specificity
Impl. Model	74.49	75.00	86.44	80.31	56.41
Baseline	73.47	77.97	77.97	77.97	66.67

Table 4-2: Performance metrics for identifying the presence of cancer in a set of FOVs.

Figure **4-10** shows the ROC curve corresponding to the SVM. The solid line corresponds to the ROC curve for implicit model (M_im), where the Area Under the Curve (AUC) is equal to 0.7771. In contrast, the dashed line represents the baseline (M_ex) ROC curve with an AUC=0.7750. Although differences are not meaningful, these results demonstrate that the presented approach, based on knowledge implicitly extracted from pathologists, can be used instead of the classical approach that require specific manual annotations.

Discussion

In this work, we introduced a new computational model that takes advantage of low, mid-, and high-level image information to predict the likelihood of cancer presence in WSIs. Low-level information was extracted from nuclear visual properties (spatial distribution, area and color), and high level information was extracted from visual attention maps, generated from pathologists interactions during diagnostic tasks. Our approach was able to identify RoIs within the whole slide images that seemed to be critical in predicting the presence of cancer; areas where the pathologists tend to focus on when making their diagnostic decision. As

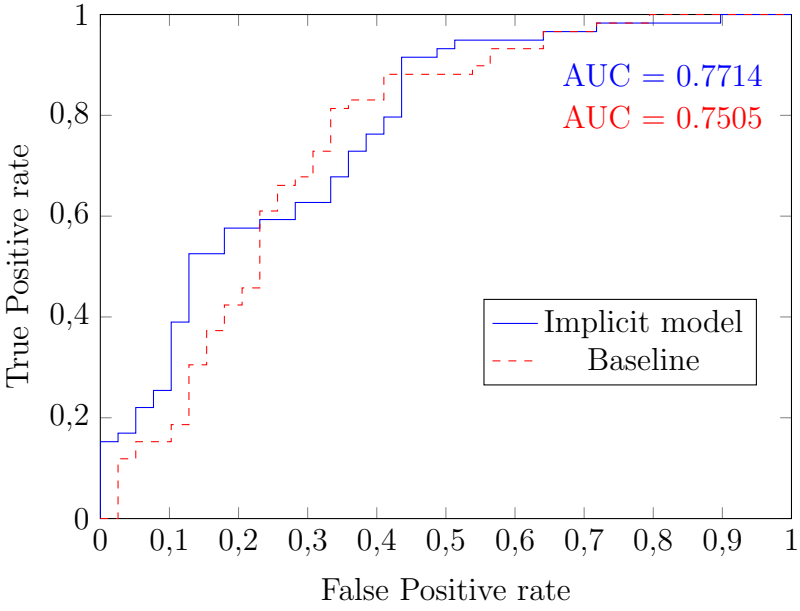


Figure 4-10: ROC curves for correctly identifying cancer in the test set of FOVs using the integrated and the baseline

previously stated, there is a wide body of evidence that abnormal tissue is characterized by nuclear patterns which can substantively inform the pathologists decision [135]. These patterns are interpreted by expert pathologists in ways that cannot easily be captured by simple models. As a first step, our experiments used nuclear features extracted from H&E slides of BCC samples to train a classifier to identify whether or not a given FOV contained cancer. This model was informed using an automated information gathering process that identified what RoIs in tissue slides a pathologist typically tends to focus on, and used that information to interpret the significance of nuclei features. While in this work we demonstrated the applicability of the new model in the diagnosis of BCC alone, this approach may be able to help incorporate expert knowledge into the interpretation of a wide variety of disease types.

The method was evaluated in terms of its ability to determine if a set of FOVs extracted from BCC WSIs contained cancerous regions. For this purpose the whole dataset was randomly split into a *learning* and a *evaluation* sets. The learning set was used to estimate the weights of a multi-linear regression (the least squares method) of visual features, aiming to approximate the visual attention maps generated by interactions of a pathologist. Subsequently, the WSIs in the experimentation set were split into 274 FOVs. 176 FOVs (from 5 patients) were used to train an SVM learning classifier to predict which FOVs were cancerous or noncancerous. The remaining 98 FOVs (from 2 patients) were employed as the test set.

The SVM yielded an accuracy of 74.49 % and an F-measure of 80.31 %, in turn representing an improvement of 1.02 % and 2.34 % over a baseline representation using a model trained

with manual annotations. The ROC curve of this SVM classifier yielded an AUC=0.7714 for the integrated model and AUC=0.7505 for the baseline.

Our approach aims to identify diagnostically relevant information from the WSI by explicitly modeling and integrating attributes pertaining to the interaction of an expert with a WSI. The method may potentially be prone to different types of noise. An example may be where the cancer region is too small, with many healthy cells surrounding a region of interest, e.g. micronodular carcinoma. The results presented herein suggest that even if different sub-types of skin cancer manifest on the slides, at the level of the cell, they may still be similar. The model aims to estimate the primary relationships in the feature space and uses these learned relationships to then predict the tissue class. In this particular case, since some data might not be the most çlass representative", linear regression was performed on about 830,000 different nuclei from 17 different WSIs. Our results suggest that the use of a large number of training exemplars might help to offset concerns regarding image noise.

Experimental results appear to suggest the following two issues. First, handcrafted nuclear features (spatial distribution, area, and color) were found to be independently discriminating of BCC and benign regions. Second, the presented approach, using implicit knowledge from pathologists, has a slightly better performance than the very same model but trained using manual annotations of individual nuclei. Consequently, this approach shows a high potential of being used in a real scenarios since experts are not asked to manually annotate the relevant structures, an obviously time consuming and prone to error task. Therefore, relevant information can be passively collected during pathologists' routine tasks (e.g. performing a diagnosis or teaching) for posterior analysis.

While visual saliency is a useful clue for focusing on relevant information, these maps tend to incorrectly suppress targets and pop-out distractors [97]. Past experience suggests that most models cannot replicate the ability of expert knowledge to describe entire populations of nuclei [135]. Different experiments [158, 45] have shown that when the visual process is task-driven, identification of RoIs requires complex interaction between two complementary sources of information. These include the *bottom-up flow* of information coming from low level image features such as color, edges, intensity or texture, and the *top-down flow* of expert knowledge, which establishes relationships between experience and a particular application [62]. In this work, visual information coming from nuclear handcrafted features was combined with high level information obtained from visual attention maps. These maps captured the focus and visual attention of expert pathologists during a focused cancer identification task, demonstrating that higher-level understanding can complement and enrich the prediction possible by low-level image features.

In a recent work [17], authors presented an approach for breast histopathology image segmentation in which an image is represented as a graph using an Euclidean spatio-colour-texture distance based similarity. This work utilized similar mathematical structures by introducing a graph-based model that assigns higher likelihood to clustered nodes and area-related homogeneous neighborhoods. Experimental results showed that these features were useful for

a tasks such as cancer detection. However, these results also indicate that this information is enriched by incorporating higher level information, pertaining to the regions a pathologist tends to focus on while rendering a diagnosis, and not necessarily the rationale behind the diagnosis itself.
Our work did however have its limitations. A major limitation of the presented approach was the size of the dataset. Since this work required the digitization and manual annotation of WSIs, and the capture of diagnostic navigation paths of 4 different pathologists, it was necessary to limit the number of images due to the limited time availability of our experts. Another limitation of this work is that we only focused on selective features of nuclear morphology and architecture while ignoring potential value that could be derived from sub-visual features such as from the stroma [146]. Interestingly, pathologists do not currently spend a great deal of time interrogating stromal architecture, but recent results suggest that tumor adjacent normal appearing regions might contain substantial prognostic information [95]. Future work will include validation of our approach on a larger independent test set, evaluation of our approach on other study cases apart from BCC, as well as the consideration of new visual and architectural nuclear features.

4.3. Products

Journal papers

- Germán Corredor, Jon Whitney, Viviana Arias, Anant Madabhushi, and Eduardo Romero. *Training a cell level classifier for detecting basal cell carcinoma by combining human visual attention maps with low level handcrafted features.* Journal of Medical Imaging, vol 4(2), 2017

Conference papers

- Germán Corredor and Eduardo Romero. *Learning Histopathological Regions of Interest by fusing bottom-up and top-down information.* Proc. of the International Conference on Image Processing (ICIP) 2015, Quebec city - Canada, 2015

- Daniel Santiago, Germán Corredor, and Eduardo Romero. *A sparse representation of the pathologist's interaction with whole slide images to improve the assigned relevance of regions of interest.* Proc. of the 13th International Symposium on Medical Information Processing and Analysis, San Andres - Colombia, 2017

Indirect products

- Lina Guzmán, Germán Corredor, and Eduardo Romero. *A Radiology Image Retrieval System based on User Preferences.* Proceedings of the 12th International Symposium

on Medical Information Processing and Analysis, Tandil - Argentina, 2016

5 Conclusions

In general, we consider this thesis has presented two breakthrough contributions:

- As far as we know, we introduced the first attempt to implicitly include the expert in the learning process involved in the automatic detection of regions of interest in histopahological images. Previous works [127] have included participation of the expert; however, they require an active manual interaction that is not similar to what pathologists do in daily routine. In contrast, in the presented approach, high-level knowledge is passively extracted from interactions (navigations) of pathologists with the WSI; then, it is used to determine the relative importance of low- and mid-level nuclear features for predicting the likelihood that a nucleus in a WSI is cancerous. Experimental results showed this approach has a comparable performance with an approach based on explicit manual annotations.

  Modern machine learning algorithms are doing a great job at identifying patterns in histopathology images; however, some of them (e.g., deep learning approaches) depend on large quantities of manually annotated data. Manual annotation is an unrealistic task because it must be performed by experts (to guarantee the quality of information), who hardly have time for such kind of activities. Looking for alternative methods for capturing high-level information is still an open question.

  We understand the presented approach is still noisy and requires a way more extensive experimentation; nonetheless, it has a lot of potential to be used in real clinical scenarios due to the speed and ease with which it can absorb higher-level information from routine tasks of physicians and implement that knowledge to generate objective diagnosis.

- We explored the prognostic potential of tumor-infiltrating lymphocytes for patients with non-small cell lung cancer. Although several works[7, 20, 24] have demonstrated a high correlation between the infiltration grade of lymphocytes and the patient prognosis (response to therapy, survival, recurrence, etc.), this information is not being used in clinical practice because of a lack of standardized methodologies and objective strategies to quantify the infiltration. In the presented work, we found out that spatial arrangement of nuclei lymphocytes and interplay between immune and cancerous cells show a high correlation with disease recurrence in early-stage non-small cell lung cancer. Predicting whether or not a patient will have recurrence might guide clinicians

for better decision making and treatment planning. In addition, this would allow clinicians for identifying patients with higher risk disease who would benefit from adjuvant chemotherapy. It is worth to say that, although this work was focused on lung cancer, lymphocytes have proven to have a prognostic value for different kinds of cancers, so application to other solid cancer types is another possible extension.

In addition to such mentioned contributions, the presented work also has explored other approaches with a high potential. For example, we presented an adaptable strategy oriented to ease the visualization of histopathology images. This approach effectively exploits and extends the granularity of the JPEG2000 standard and integrates it with different strategies to achieve a lossless, loosely-coupled, decoder, and platform independent implementation, adaptable to any interaction model. As a result of this work, a web-based tool was developed. It allows users to interact with the content in different ways: image visualization and navigation, querying clinical data and annotations, visualization of the diagnostic paths followed by expert pathologists, etc. Currently, this application is publicly available (http://cimalab.unal.edu.co/microscopio/) and contains more than 250 histology samples (about 40 % of them have annotations and clinical data), which have been included within the last months. Although the application is still in development, it has a great educative potential since it provides high-quality and easy-to-share material for physicians and learners. Similarly, in this work we also explored a strategy for detection and grading of ductal carcinoma in situ. Although DCIS is not life-threatening, people who suffer it are at a high risk of developing invasive carcinoma, so they may require additional surveillance. Unfortunately, detection and grading of DCIS is challenging because of the high variability of the lesions; a low agreement among experts have been reported [48]. In this work, we exploited information of nuclear features at different scales to detect and grade DCIS. This kind of approach might provide pathologists with objective and quantitative tools that facilitate decision making and treatment planning.

In summary, this thesis has presented a whole pipeline for pathology. This pipeline includes strategies to provide pathologists access to visual content and to extract diagnostic and prognostic information from such content. A very important point of this pipeline is that it includes the expert as a part of the learning process.

Although this approach has a great potential, still there are different things to explore and improve. Future work will include enhancing the web application for visualization of histopathology images by making it more robust and able to provide access to many more simultaneous users. We plan to expand its functionality, so it will include user management, more educative material, tools for evaluation, and an analytics tool displaying objective metrics and suggestions. In collaboration with the Pathology Department of Universidad Nacional de Colombia, we plan to add more cases including clinical data and annotations. Furthermore, we understand that the number of samples used in the different experiments along the thesis is relatively small in terms of the complexity of the addressed problem. For this reason, using the web application, we plan to record the interactions of many more

experts. With the new added cases and the high-level information obtained from annotations and recorded navigations, we could extend the experimentation and propose better predictive models.

Bibliography

[1] M.D. Adams and F. Kossentini. Jasper: A software-based jpeg-2000 codec implementation. http://www.ece.uvic.ca/ frodo/jasper/.

[2] Shaimaa Al-Janabi, André Huisman, Aryan Vink, Roos J. Leguit, G. Johan A. Offerhaus, Fiebo J.W. ten Kate, and Paul J. van Diest. Whole slide images for primary diagnostics of gastrointestinal tract pathology: a feasibility study. *Human Pathology*, 43(5):702 – 707, 2012.

[3] L Alfaro, E Poblet, P Català, A Navea, and M García Rojo. Compatability of virtual microscopy equipment: an analysis of different panoramic image software. *Revista Española de Patología*, 44(1):8–16, 2011.

[4] S. Ali and A. Madabhushi. An integrated region-, boundary-, shape-based active contour for multiple object overlap resolution in histological imagery. *IEEE Transactions on Medical Imaging*, 31(7):1448–1460, July 2012.

[5] Sahirzeeshan Ali, James Lewis, and Anant Madabhushi. *Spatially Aware Cell Cluster(SpACCl) Graphs: Predicting Outcome in Oropharyngeal p16+ Tumors*, pages 412–419. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.

[6] Sahirzeeshan Ali, Robert Veltri, Jonathan A Epstein, Christhunesa Christudass, and Anant Madabhushi. Cell cluster graph for prediction of biochemical recurrence in prostate cancer patients from tissue microarrays. In *SPIE Medical Imaging*, pages 86760H–86760H. International Society for Optics and Photonics, 2013.

[7] Marc-Antoine Allard, Jean Baptiste Bachet, Alain Beauchet, Catherine Julie, Robert Malafosse, Christophe Penna, Bernard Nordlinger, and Jean-François Emile. Linear quantification of lymphoid infiltration of the tumor margin: a reproducible method, developed with colorectal cancer tissues, for assessing a highly variable prognostic factor. *Diagnostic Pathology*, 7(1):1–11, 2012.

[8] D.C. Allen and R.I. Cameron. *Histopathology Specimens: Clinical, Pathological and Laboratory Aspects*. Springer International Publishing, 2017.

[9] Aperio Technologies Inc. Aperio image analysis. http://tmalab.jhmi.edu/aperiou/userguides/Image_Analysis_UG.pdf, 2009.

[10] Seong-Joon Baek, Aaron Park, Jin-Young Kim, Seung Yu Na, Yonggwan Won, and Jaebum Choo. *Detection of Basal Cell Carcinoma by Automatic Classification of Confocal Raman Spectra*, pages 402–411. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.

[11] J. Bartrina-Rapesta, J. Serra-Sagrist, and F. AulÃ-Llins. Jpeg2000 roi coding through component priority for digital mammography. *Computer Vision and Image Understanding*, 115(1):59–68, 2011.

[12] Souptik Barua, Penny Fang, Amrish Sharma, Junya Fujimoto, Ignacio Wistuba, Arvind U.K. Rao, and Steven H. Lin. Spatial interaction of tumor cells and regulatory t cells correlates with survival in non-small cell lung cancer. *Lung Cancer*, 117:73 – 79, 2018.

[13] A. N. Basavanhally, S. Ganesan, S. Agner, J. P. Monaco, M. D. Feldman, J. E. Tomaszewski, G. Bhanot, and A. Madabhushi. Computerized image-based detection and grading of lymphocytic infiltration in her2+ breast cancer histopathology. *IEEE Transactions on Biomedical Engineering*, 57(3):642–653, March 2010.

[14] Ajay Basavanhally, Shridar Ganesan, Michael Feldman, Natalie Shih, Carolyn Mies, John Tomaszewski, and Anant Madabhushi. Multi-field-of-view framework for distinguishing tumor grade in er+ breast cancer from entire histopathology slides. *IEEE transactions on bio-medical engineering*, 60:2089–99, Aug 2013.

[15] Carlos Becker, Roberto Rigamonti, Vincent Lepetit, and Pascal Fua. Supervised feature learning for curvilinear structure segmentation. In Kensaku Mori, Ichiro Sakuma, Yoshinobu Sato, Christian Barillot, and Nassir Navab, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*, pages 526–533, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[16] B. Ehteshami Bejnordi, M. Balkenhol, G. Litjens, R. Holland, P. Bult, N. Karssemeijer, and J. A. W. M. van der Laak. Automated detection of dcis in whole-slide h e stained breast histopathology images. *IEEE Transactions on Medical Imaging*, 35(9):2141–2150, Sept 2016.

[17] A.D. Belsare, M.M. Mushrif, M.A. Pangarkar, and N. Meshram. Breast histopathology image segmentation using spatio-colour-texture based graph partition method. *Journal of Microscopy*, 262(3):260–273, 2016.

[18] Jacob Benesty, Jingdong Chen, Yiteng Huang, and Israel Cohen. *Pearson Correlation Coefficient*, pages 1–4. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

[19] Rohit Bhargava and Anant Madabhushi. Emerging themes in image informatics and molecular analysis for digital pathology. *Annual Review of Biomedical Engineering*, 18(1):387–412, 2016.

[20] Elisabeth Brambilla, Gwénaël Le Teuff, Sophie Marguet, Sylvie Lantuejoul, Ariane Dunant, Stephen Graziano, Robert Pirker, Jean-Yves Douillard, Thierry Le Chevalier, Martin Filipits, Rafael Rosell, Robert Kratzke, Helmut Popper, Jean-Charles Soria, Frances A. Shepherd, Lesley Seymour, and Ming Sound Tsao. Prognostic effect of tumor lymphocytic infiltration in resectable non-small-cell lung cancer. *Journal of Clinical Oncology*, 34(11), 2016.

[21] Breastcancer.org. Ductal carcinoma in situ (dcis). http://www.breastcancer.org/symptoms/types/dcis/.

[22] Roy M. Bremnes, Tom Donnem, and Lill-Tove Busund. Importance of tumor infiltrating lymphocytes in non-small cell lung cancer? *Annals of Translational Medicine*, 4(7), 2016.

[23] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30(1-7):107–117, April 1998.

[24] Jason R. Brown, Hallie Wimberly, Donald R. Lannin, Christian Nixon, David L. Rimm, and Veerle Bossuyt. Multiplexed quantitative analysis of cd3, cd8, and cd20 predicts response to neoadjuvant chemotherapy in breast cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 20:5995–6005, Dec 2014.

[25] G. Bueno, M. Fernández-Carrobles, O. Deniz, and M. García-Rojo. New trends of emerging technologies in digital pathology. *Pathobiology*, 83:61–69, 2016.

[26] Laurence Buisseret, Christine Desmedt, Soizic Garaud, Marco Fornili, Xiaoxiao Wang, Gert Van den Eyden, Alexandre de Wind, Sebastien Duquenne, Anais Boisson, Celine Naveaux, Francoise Rothé, Sandrine Rorive, Christine Decaestecker, Denis Larsimont, Martine Piccart-Gebhart, Elia Biganzoli, Christos Sotiriou, and Karen Willard-Gallo. Reliability of tumor-infiltrating lymphocyte and tertiary lymphoid structure assessment in human breast cancer. *Modern Pathology*, 30:1204, June 2017.

[27] Tom Weidong Cai, Jinman Kim, and David Dagan Feng. *4 - Content-Based Medical Image Retrieval*, pages 83 – 113. Biomedical Engineering. Academic Press, Burlington, 2008.

[28] Robert L. Camp, Gina G. Chung, and David L. Rimm. Automated subcellular localization and quantification of protein expression in tissue microarrays. *Nature Medicine*, 8:1323–, October 2002.

[29] Jean Carletta. Assessing agreement on classification tasks: The kappa statistic. *Comput. Linguist.*, 22(2):249–254, June 1996.

[30] Julienne L. Carstens, Pedro Correa de Sampaio, Dalu Yang, Souptik Barua, Huamin Wang, Arvind Rao, James P. Allison, Valerie S. LeBleu, and Raghu Kalluri. Spatial computation of intratumoral t cells correlates with survival of patients with pancreatic cancer. *Nature Communications*, 8:15095–, April 2017.

[31] Yuk-Kwan Chen, Shue-Shang Hsue, Dai-Chung Lin, Wen-Chen Wang, Jing-Yi Chen, Cheng-Chung Lin, and Li-Min Lin. An application of virtual microscopy in the teaching of an oral and maxillofacial pathology laboratory course. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology*, 105(3):342–347, 2008.

[32] Brian K. Chiu, Kim Solez, and Consolato M. Sergi. Digital pathology for e-learning and digital education - a review. *Journal of Information Technology and Application in Education*, 3(4), 2014.

[33] Ryan Chute and Herbert Van de Sompel. Introducing djatoka. a reuse friendly, open source jpeg 2000 image server. *D-Lib Magazine*, 14(9), 2008.

[34] L.A.D. Cooper, A.B. Carter, A.B. Farris, Fusheng Wang, Jun Kong, D.A. Gutman, P. Widener, T.C. Pan, S.R. Cholleti, A. Sharma, T.M. Kurc, D.J. Brat, and J.H. Saltz. Digital pathology: Data-intensive frontier in medical imaging. *Proceedings of the IEEE*, 100(4):991–1003, April 2012.

[35] G. Corredor and E. Romero. Learning histopathological regions of interest by fusing bottom-up and top-down information. In *Image Processing (ICIP), 2015 IEEE International Conference on*, pages 3200–3204, 2015.

[36] Germán Corredor, Eduardo Romero, and Marcela Iregui. An adaptable navigation strategy for virtual microscopy from mobile platforms. *Journal of Biomedical Informatics*, 54:39 – 49, 2015.

[37] Germán Corredor, Xiangxue Wang, Cheng Lu, Vamsidhar Velcheti, Eduardo Romero, and Anant Madabhushi. A watershed and feature-based approach for automated detection of lymphocytes on lung cancer images. In *SPIE Medical Imaging*. International Society for Optics and Photonics, 2018.

[38] Angel Cruz-Roa, John Arevalo Ovalle, Anant Madabhushi, and Fabio González Osorio. *A Deep Learning Architecture for Image Representation, Visual Interpretability and Automated Basal-Cell Carcinoma Cancer Detection*, pages 403–410. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.

[39] E de Ves, J Domingo, G Ayala, and P Zuccarello. A novel Bayesian framework for relevance feedback in image content-based retrieval systems. *Pattern Recognition*, 39(9):1622–1632, 2006.

[40] Fred R. Dee. Virtual microscopy in pathology education. *Human Pathology*, 40:1112–1121, 2009.

[41] A. Descampe, C. De Vleeschouwer, M. Iregui, B. Macq, and F. Marques. Pre-fetching and caching strategies for remote and interactive browsing of jpeg2000 images. *IEEE Transactions on Image Processing*, 16(5):1339–1354, 2007.

[42] Ivica Dimitrovski, Dragi Kocev, Ivan Kitanovski, Suzana Loskovska, and Sašo Džeroski. Improved medical image modality classification using a combination of visual and textual features. *Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society*, 39:14–26, January 2015.

[43] Tom Donnem, Sigurd M. Hald, Erna-Elise Paulsen, Elin Richardsen, Samer Al-Saad, Thomas K. Kilvaer, Odd Terje Brustugun, Aslaug Helland, Marius Lund-Iversen, Mette Poehl, Karen Ege Olsen, Henrik J. Ditzel, Olfred Hansen, Khalid Al-Shibli, Yury Kiselev, Torkjel M. Sandanger, Sigve Andersen, Francesco Pezzella, Roy M. Bremnes, and Lill-Tove Busund. Stromal cd8+ t-cell density - a promising supplement to tnm staging in non-small cell lung cancer. *Clinical Cancer Research*, 21(11):2635–2643, 2015.

[44] S. Doyle, M. Feldman, J. Tomaszewski, and A. Madabhushi. A boosted bayesian multiresolution classifier for prostate cancer detection from digitized needle biopsies. *IEEE Transactions on Biomedical Engineering*, 59(5):1205–1218, May 2012.

[45] John Duncan and Glyn W. Humphreys. Visual search and stimulus similarity. *Psychological Review*, 96(3):433 – 458, 1989.

[46] David Eddy and John Billings. The quality of medical evidence: Implications for quality of care. *Health Affairs*, 7(1):19–32, 1988.

[47] GilanM. El Saadawi, Eugene Tseytlin, Elizabeth Legowski, Drazen Jukic, Melissa Castine, Jeffrey Fine, Robert Gormley, and RebeccaS. Crowley. A natural language intelligent tutoring system for training pathologists: implementation and evaluation. *Advances in Health Sciences Education*, 13(5):709–722, 2008.

[48] J.G. Elmore, G.M. Longton, PA Carney, and et al. Diagnostic concordance among pathologists interpreting breast biopsy specimens. *JAMA*, 313(11):1122–1132, 2015.

[49] Ann C. Eriksen, Johnnie B. Andersen, Martin Kristensson, RenÃ© dePont Christensen, Torben F. Hansen, Sanne KjÃ¦r-Frifeldt, and Flemming B. SÃ¸rensen. Computer-assisted stereology and automated image analysis for quantification of tumor infiltrating lymphocytes in colon cancer. *Diagnostic Pathology*, 12:65, August 2017.

[50] H. Fatakdawala, J. Xu, A. Basavanhally, G. Bhanot, S. Ganesan, M. Feldman, J. E. Tomaszewski, and A. Madabhushi. Expectation-maximization-driven geodesic active contour with overlap resolution (emagacor): Application to lymphocyte segmentation on breast cancer histopathology. *IEEE Transactions on Biomedical Engineering*, 57(7):1676–1689, July 2010.

[51] Jeffrey L. Fine, Dana M. Grzybicki, Russell Silowash, Jonhan Ho, John R. Gilbertson, Leslie Anthony, Robb Wilson, Anil V. Parwani, Sheldon I. Bastacky, Jonathan I. Epstein, and Drazen M. Jukic. Evaluation of whole slide image immunohistochemistry interpretation in challenging prostate needle biopsies. *Human Pathology*, 39(4):564 – 572, 2008.

[52] Andrew H. Fischer, Kenneth A. Jacobson, Jack Rose, and Rolf Zeller. Hematoxylin and eosin staining of tissue and cell sections. *Cold Spring Harbor Protocols*, 2008(5):pdb.prot4986, 2008.

[53] P. Fontelo, E. DiNino, K. Johansen, A. Khan, and M. Ackerman. Virtual microscopy: Potential applications in medical education and telemedicine in countries with developing economies. *Proceedings of the 38th Hawaii International Conference on System Sciences*, pages 1–7, 2005.

[54] Vitaly Friedman. Google pagerank: What do we know about it? *Smashing Magazine*, 2007.

[55] Guyatt G, Cairns J, Churchill D, and et al. Evidence-based medicine: A new approach to teaching the practice of medicine. *JAMA*, 268(17):2420–2425, 1992.

[56] Brandon D Gallas, Marios A Gavrielides, Catherine M Conway, Adam Ivansky, Tyler C Keay, Wei-Chung Cheng, Jason Hipp, and Stephen M Hewitt. Evaluation environment for digital and analog pathology: a platform for validation studies. *Journal of Medical Imaging*, 1(3):037501–, October 2014.

[57] Y. Geng, Y. Shao, W. He, W. Hu, Y. Xu, J. Chen, C. Wu, and J. Jiang. Prognostic role of tumor-infiltrating lymphocytes in lung cancer: a meta-analysis. *Cellular Physiology and Biochemistry*, 37(4):1560–1571, 2015.

[58] D. Gibson, M. Spann, and S.I. Woolley. A wavelet-based region of interest encoder for the compression of angiogram video sequences. *Information Technology in Biomedicine, IEEE Transactions on*, 8(2):103–113, June 2004.

[59] F. Gómez, D. Marín, and E. Romero. A soft-cache strategy for pathologist's navigation in virtual microscopy. *Microscopy Research and Technique*, 74(5):401–414, 2011.

[60] Francisco Gómez, Marcela Iregui, and Eduardo Romero. Virtual microscopy using jpeg2000. In *Computer Analysis of Images and Patterns*, volume 4673 of *Lecture Notes in Computer Science*, pages 181–188. Springer Berlin Heidelberg, 2007.

[61] Francisco Gómez and Eduardo Romero. A model for predicting pathologist's velocity profiles when navigating virtual slides. *Microscopy Research and Technique*, 73:85–98, 2009.

[62] Francisco A. Gómez. *Efficient navegation of medical mega-images*. PhD thesis, Universidad Nacional de Colombia, 2010.

[63] Adam. Goode, Benjamin. Gilbert, Jan. Harkes, Drazen. Jukic, and Mahadev. Satyanarayanan. OpenSlide: A vendor-neutral software foundation for digital pathology. *Journal of Pathology Informatics*, 4(1):27, 2013.

[64] Google Inc. Adding photooverlays (google earth api documentation). https://developers.google.com/kml/documentation/photos?hl=en.

[65] T Venu Gopal, P Prasanna Rani, and Kranthi Kumar.K. Relevance feedback: A proposed method for reducing semantic gap in content-based image retrieval using images with relevance level. *International Journal of Engineering Sciences Research*, 4(2):842–850, 2013.

[66] Trisha Greenhalgh. Effectiveness and efficiency: Random reflections on health services. *BMJ : British Medical Journal*, 328(7438):529–529, 2004.

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

[68] Ricardo Gutiérrez, Francisco Gómez, Lucía Roa-Peña, and Eduardo Romero. A supervised visual model for finding regions of interest in basal cell carcinoma images. *Diagnostic Pathology*, 6(1):26, 2011.

[69] Ricardo Gutierrez and Eduardo Romero. A visual model approach to extract regions of interest in microscopical images of basal cell carcinoma. *Diagnostic Pathology*, 8(Suppl 1):S36, 2013.

[70] M. Hadwiger, J. Beyer, Won-Ki Jeong, and H. Pfister. Interactive volume exploration of petascale microscopy data streams using a visualization-driven virtual memory approach. *Visualization and Computer Graphics, IEEE Transactions on*, 18(12):2285–2294, Dec 2012.

[71] Jonathan Harel, Christof Koch, and Pietro Perona. Graph-based visual saliency. In *Advances in Neural Information Processing Systems 19*, pages 545–552. MIT Press, 2007.

[72] Institute for Innovations in Medical Education. NYU Virtual Microscope. http://education.med.nyu.edu/virtualmicroscope.

[73] Intel Corp. Performance tools for software developers - application notes - intel ipp jpeg2000 and jasper in ksquirrel. http://software.intel.com/en-us/articles/performance-tools-for-software-developers-application-notes-intel-ipp-jpeg2000-and-jasper-in-ksquirrel.

[74] M. Iregui, P. Chevalier, and B. Macq. Optimal caching mechanisms for jpeg2000 communications. *EUSIPCO - European Signal Processing Conference*, 3:201–204, 2002.

[75] M. Iregui, F. Gómez, and E. Romero. Strategies for efficient virtual microscopy in pathological samples using jpeg2000. *Micron*, 38(7):700–713, 2007.

[76] ISO/IEC. 15444-1. information technology - jpeg2000 image coding system - part 1: Core coding system, 2000.

[77] ISO/IEC. 15444-9. information technology - jpeg2000 image coding system - part 9: Interactivity tools, apis and protocols, 2004.

[78] Andrew Janowczyk and Anant Madabhushi. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *Journal of Pathology Informatics*, 7(1):29, 2016.

[79] Joel Saltz, Rajarsi Gupta, Le Hou, Tahsin Kurc, Pankaj Singh, et al. Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images. *Cell Reports*, 23(1):181–193.e7, 2018.

[80] J.P. Johnson, E.A. Krupinski, M. Yan, H. Roehrig, A.R. Graham, and R.S. Weinstein. Using a visual discrimination model for the detection of compression artifacts in virtual pathology images. *Medical Imaging, IEEE Transactions on*, 30(2):306–314, Feb 2011.

[81] Hassan MH Kamel. Trends and challenges in pathology practice: Choices and necessities. *Sultan Qaboos University Medical Journal*, 11(1):38–44, November 2010.

[82] Shinichiro Kashiwagi, Yuka Asano, Wataru Goto, Koji Takada, Katsuyuki Takahashi, Satoru Noda, Tsutomu Takashima, Naoyoshi Onoda, Shuhei Tomita, Masahiko Ohsawa, Kosei Hirakawa, and Masaichi Ohira. Use of tumor-infiltrating lymphocytes (tils) to predict the treatment response to eribulin chemotherapy in breast cancer. *PLoS ONE*, 12(2):e0170634, January 2017.

[83] Eamonn Keogh and Abdullah Mueen. *Curse of Dimensionality*, pages 314–315. Springer US, Boston, MA, 2017.

[84] Adnan Mujahid Khan and Yinyin Yuan. Biopsy variability of lymphocytic infiltration in breast cancer subtypes and the immunoskew score. *Scientific Reports*, 6:36231–, November 2016.

[85] DK Kim, EY Kim, KH Yang, CK Lee, and SK Yoo. A mobile tele-radiology imaging system with jpeg2000 for an emergency care. *Journal of Digital Imaging*, 24(4):709–718, 2011.

[86] Min Hyung Kim, Youngwoon Park, Donghee Seo, Yu Jin Lim, Dong-Il Kim, Chul Woo Kim, and Woo Ho Kim. Virtual microscopy as a practical alternative to conventional microscopy in pathology education. *Basic Appl Pathol*, 1:46–48, 2008.

[87] Mariko Kochi, Takayuki Iwamoto, Naoki Niikura, Giampaolo Bianchini, Shinobu Masuda, Taeko Mizoo, Tomohiro Nogami, Tadahiko Shien, Takayuki Motoki, Naruto Taira, Yutaka Tokuda, Hiroyoshi Doihara, Junji Matsuoka, and Toshiyoshi Fujiwara. Tumour-infiltrating lymphocytes (tils)-related genomic signature predicts chemotherapy response in breast cancer. *Breast cancer research and treatment*, Sep 2017.

[88] Sonal Kothari, John Phan, Todd Stokes, and May Wang. Pathology imaging informatics for quantitative analysis of whole-slide images. *Journal of the American Medical Informatics Association*, 20:1099–1108, 2013.

[89] E. Krupinski, A. Tillack, L. Richter, J. Henderson, A. Bhattacharyya, K. Scott, A. Graham, M. Descour, J. Davis, and R. Weinstein. Eye-movement study and human performance using telepathology virtual slides. implications for medical education and differences with experience. *Human Pathology*, 37(12):1543 – 1556, 2006.

[90] Rakesh K Kumar, Gary M Velan, Sami O Korell, Madan Kandara, Fred R Dee, and Denis Wakefield. Virtual microscopy for learning and assessment in pathology. *The Journal of Pathology*, 204(5):613–618, 2004.

[91] V. Kumar, A.K. Abbas, and J.C. Aster. *Robbins & Cotran Pathologic Basis of Disease*. Robbins Pathology. Elsevier Health Sciences, 2014.

[92] Y. Kuo, C. Ko, and M. Lee. Lymphatic infiltration detection in breast cancer h&e image prior to lymphadenectomy. *Biomedical Engineering: Applications, Basis and Communications*, 26(04):1440007, 2014.

[93] Page Lawrence, Brin Sergey, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford University, 1998.

[94] George Lee, Rachel Sparks, Sahirzeeshan Ali, Natalie N. C. Shih, Michael D. Feldman, Elaine Spangler, Timothy Rebbeck, John E. Tomaszewski, and Anant Madabhushi. Co-occurring gland angularity in localized subgraphs: Predicting biochemical recurrence in intermediate-risk prostate cancer patients. *PKoS ONE*, 9(5), 2014.

[95] George Lee, Robert W. Veltri, Guangjing Zhu, Sahirzeeshan Ali, Jonathan I. Epstein, and Anant Madabhushi. Nuclear shape and architecture in benign fields predict biochemical recurrence in prostate cancer patients following radical prostatectomy: Preliminary findings. *European Urology Focus*, pages–, 2016.

[96] Geng Li, Murat Semerci, Bülent Yener, and Mohammed J Zaki. Effective graph classification based on topological and label attributes. *Statistical Analysis and Data Mining*, 5(4):265–283, 2012.

[97] J. Li, Y. Tian, and T. Huang. Visual saliency with statistical priors. *International Journal of Computer Vision*, 107(3):239–253, 2014.

[98] Tsai-Yen Li and Chih wei Chiang. Data management for visualizing large virtual environments. In *Proceedings of International Symposium on Multimedia Information Processing*, 1999.

[99] Yue Li and Chia-Hung Wei. Medical image retrieval: Multiple regression models for user's search target. In *Machine Learning and Cybernetics (ICMLC), 2011 International Conference on*, volume 4, pages 1436–1443, 2011.

[100] Hui Liu, Tiantuo Zhang, Jin Ye, Hongtao Li, Jing Huang, Xiaodong Li, Benquan Wu, Xubing Huang, and Jinghui Hou. Tumor-infiltrating lymphocytes predict response to chemotherapy in patients with advance non-small cell lung cancer. *Cancer Immunology, Immunotherapy*, 61(10):1849–1856, Oct 2012.

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

[102] L. Long, S. Antani, T. Deserno, and G. Thoma. Content-Based Image Retrieval in Medicine - Retrospective Assessment, State of the Art, and Future Directions. *International Journal of Healthcare Information Systems and Informatics*, 4(1):1–16, 2010.

[103] Mikael Lundin, Johan Lundin, Juho Konsti, and Jorma Isola. Web microscope. http://www.webmicroscope.net/.

[104] Xin Luo, Xiao Zang, Lin Yang, Junzhou Huang, Faming Liang, Jaime Rodriguez-Canales, Ignacio I Wistuba, Adi Gazdar, Yang Xie, and Guanghua Xiao. Comprehensive computational pathological image analysis predicts lung cancer prognosis. *Journal of Thoracic Oncology*, 12(3):501–509, 2017.

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

[106] Anant Madabhushi and George Lee. Image analysis and machine learning in digital pathology: Challenges and opportunities. *Medical Image Analysis*, 33:170 – 175, 2016. 20th anniversary of the Medical Image Analysis journal (MedIA).

[107] I. Maglogiannis, C. Doukas, G. Kormentzas, and Thomas Pliakas. Wavelet-based compression with roi coding support for mobile access to dicom images over heterogeneous radio networks. *Information Technology in Biomedicine, IEEE Transactions on*, 13(4):458–466, July 2009.

[108] Guido Majno and Isabelle Joris. *Cells, Tissues, and Disease: Principles of General Pathology*. Oxford University Press, 2004.

[109] Alberto M. Marchevsky, Anju Relan, and Susan Baillie. Self-instructional virtual pathology laboratories using web-based technology enhance medical school teaching of pathology. *Human Pathology*, 34(5):423–429, 2003.

[110] Diana Marín and Eduardo Romero. Virtual microscopy systems: Analysis and perspectives. *Biomédica*, 31(1):144–155, 2011.

[111] Izet Masic, Milan Miokovic, and Belma Muhamedagic. Evidence based medicine - new approaches and challenges. *Acta Informatica Medica*, 16(4):219–225, September 2008.

[112] Cesare Massone, H. Peter Soyer, Gian Piero Lozzi, Alessandro Di Stefani, Bernd Leinweber, Gerald Gabler, Masoud Asgari, Renata Boldrini, Leonardo Bugatti, Vincenzo Canzonieri, Gerardo Ferrara, Kazuo Kodama, Darius Mehregan, Franco Rongioletti, Shahbaz A. Janjua, Vahid Mashayekhi, Ismini Vassilaki, Bernhard Zelger, Borut Å½gavec, Lorenzo Cerroni, and Helmut Kerl. Feasibility and diagnostic agreement in teledermatopathology using a virtual slide system. *Human Pathology*, 38(4):546 – 554, 2007.

[113] Ezgi Mercan, Selim Aksoy, Linda G. Shapiro, Donald L. Weaver, Tad T. BrunyÃ©, and Joann G. Elmore. Localization of diagnostically relevant regions of interest in whole slide images: a comparative study. *Journal of digital imaging*, 29(4):496–506, August 2016.

[114] Microsoft Corporation. Deep zoom. http://www.microsoft.com/silverlight/deepzoom/.

[115] Steven Miller. *The Method of Least Squares*. Mathematics Department Brown University,, 2006.

[116] X. Moles-Lopez, O. Debeir, I. Salmon, and C. Decaestecker. *Whole slide imaging and analysis for biomarker evaluation in digital pathology*, pages 776–787. Formatex Research Center, 2014.

[117] Chassin MR, D H Kosecoff J FAU Solomon, R H Solomon DH FAU Brook, and Brook RH. How coronary angiography is used. clinical determinants of appropriateness. (0098-7484 (Linking)).

[118] Henning Müller, Nicolas Michoux, David Bandon, and Antoine Geissbuhler. A review of content-based image retrieval systems in medical applications-clinical benefits and future directions. *International journal of medical informatics*, 73(1):1–23, 2004.

[119] Dilip B Nagarkar, Ezgi Mercan, Donald L Weaver, Tad T Brunye, Patricia A Carney, Mara H Rendi, Andrew H Beck, Linda G Frederick, Paul D Shapiro, and Joann G Elmore. Region of interest identification and diagnostic agreement in breast pathology. *Modern Pathology*, 29:1004–1011, 2016.

[120] M.T. Nkosi and F. Mekuria. Cloud computing for enhanced mobile health applications. In *Cloud Computing Technology and Science (CloudCom), 2010 IEEE Second International Conference on*, pages 629–633, Nov 2010.

[121] Rita Noumeir and Jean François Pambrun. Using jpeg 2000 interactive protocol to stream a large image or a large image set. *J Digit Imaging*, 24(5):833–843, 2011.

[122] G. Orchard and B. Nation. *Histopathology*. Fundamentals of Biomedical Science. OUP Oxford, 2011.

[123] Liron Pantanowitz, Janusz Szymas, Yukako Yagi, and David Wilbur. Whole slide imaging for educational purposes. *Journal of Pathology Informatics*, 3(1):46, 2012.

[124] Liron Pantanowitz, Paul Valenstein, Andrew Evans, Keith Kaplan, John Pfeifer, David. Wilbur, Laura Collins, and Terence Colgan. Review of the current state of whole slide imaging in pathology. *Journal of Pathology Informatics*, 2(1):36, 2011.

[125] M. Peikari, M. J. Gangeh, J. Zubovits, G. Clarke, and A. L. Martel. Triaging diagnostically relevant regions from pathology whole slides of breast cancer: A texture based approach. *IEEE Transactions on Medical Imaging*, 35(1):307–315, Jan 2016.

[126] Hanchuan Peng, Fuhui Long, and C. Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance,, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1226–1238, Aug 2005.

[127] Loïc Peter, Diana Mateus, Pierre Chatelain, Noemi Schworm, Stefan Stangl, Gabriele Multhoff, and Nassir Navab. *Leveraging Random Forests for Interactive Exploration of Large Histological Images*, pages 1–8. Springer International Publishing, 2014.

[128] Denis Pitzalis, Ruven Pillay, and Christian Lahanier. A new concept in high resolution internet image browsing. In *Proceedings of the Conference on Electronic Publishing*, 2006.

[129] Judith Prewitt. Graphs and grammars for histology: An introduction. In *Proc Annu Symp Comput Appl Med Care*, pages 18–25, 1979.

[130] Ghanshyam Raghuwanshi, Nishchol Mishra, and Sanjeev Sharma. Content based Image Retrieval using Implicit and Explicit Feedback with Interactive Genetic Algorithm. *International Journal of Computer Applications*, 43(16):8–14, 2012.

[131] J Ramey, KM Fung, and LA Hassell. Use of mobile high-resolution device for remote frozen section evaluation of whole slide images. *J Pathol Inform*, 2(41), 2011.

[132] U Ravindran and T Shakila. Content based image retrieval for histology image collection using visual pattern mining. *International Journal of Scientific & Engineering Research*, 4(4), 2013.

[133] Branimir Reljin, Goran Zajić, Nikola Reljin, and Irini Reljin. Adaptive clustering of image database (acid) as an efficient tool for improving retrieval in a cbir system. *Series Studies in Health Technology and Informatics*, 179:172 – 183, 2012.

[134] P.H.J. Riegman, W.N.M. Dinjens, M.H.A. Oomen, W.F. Clotscher, R.J.J.R. Scholte, W. Sjoerdsma, A.R.A. Riegman, and J.W. Oosterhuis. Applications of virtual microscopy. In Sajeesh Kumar and BruceE. Dunn, editors, *Telepathology*, pages 53–62. Springer Berlin Heidelberg, 2009.

[135] Lucia Roa-Peña, Francisco Gómez, and Eduardo Romero. An experimental study of pathologist's navigation patterns in virtual microscopy. *Diagnostic Pathology*, 5(1):71, 2010.

[136] H. S. Robins, N. G. Ericson, J. Guenthoer, K. C. O'Briant, M. Tewari, C. W.Drescher, and J. H. Bielas. Digital quantification of tumor infiltrating lymphocytes. *Science translational medicine*, 5(214), 2013.

[137] Marcial García Rojo, Gloria Bueno GarcÃa, Carlos Peces Mateos, Jesús González García, and Manuel Carbajo Vicente. Critical comparison of 31 commercially available digital slide systems in pathology. *International Journal of Surgical Pathology*, 14(4):285–305, 2006.

[138] David Romo-Bucheli, Andrew Janowczyk, Hannah Gilmore, Eduardo Romero, and Anant Madabhushi. Automated tubule nuclei quantification and correlation with oncotype dx risk categories in er+ breast cancer whole slide images. *Scientific Reports*, 6(32706), 2016.

[139] R. Rosenbaum and H. Schumann. Jpeg2000-based viewer guidance for mobile image browsing. In *Multi-Media Modelling Conference Proceedings, 2006 12th International*, 2006.

[140] R. Salgado, C. Denkert, S. Demaria, N. Sirtaine, F. Klauschen, G. Pruneri, S. Wienert, G. Van den Eynden, F. L. Baehner, F. Penault-Llorca, E. A. Perez, E. A. Thompson, W. F. Symmans, A. L. Richardson, J. Brock, C. Criscitiello, H. Bailey, M. Ignatiadis, G. Floris, J. Sparano, Z. Kos, T. Nielsen, D. L. Rimm, K. H. Allison, J. S. Reis-Filho, S. Loibl, C. Sotiriou, G. Viale, S. Badve, S. Adams, K. Willard-Gallo, and S. Loi. The evaluation of tumor-infiltrating lymphocytes (tils) in breast cancer: Recommendations by an international tils working group 2014. *Annals of Oncology*, 26(2):259–271, 2015.

[141] C.M. Sarraf and R.M. Wakim. Improving jpeg 2000 images delivery over gprs mobile networks. In *Proceedings of the 6th WSEAS Int. Conf. on Electronics, Hardware, Wireless and Optical Communications*, 2007.

[142] Peter Savas, Roberto Salgado, Carsten Denkert, Christos Sotiriou, Phillip K. Darcy, Mark J. Smyth, and Sherene Loi. Clinical relevance of host immunity in breast cancer: from TILs to the clinic. *Nat Rev Clin Oncol*, 13(4):228–241, April 2016.

[143] Kurt A. Schalper, Jason Brown, Daniel Carvajal-Hausdorf, Joseph McLaughlin, Vamsidhar Velcheti, Konstantinos N. Syrigos, Roy S. Herbst, and David L. Rimm. Objective measurement and clinical significance of tils in non-small cell lung cancer. *Journal of the National Cancer Institute*, 107(3):1–9, 2015.

[144] Deepti Sharma and Swati Srivastava. Automatically detection of skin cancer by classification of neural network. *International Journal of Engineering and Technical Research*, 4, 2016.

[145] Harshita Sharma, Norman Zerbe, Sebastian Lohmann, Klaus Kayser, Olaf Hellwich, and Peter Hufnagl. A review of graph-based methods for image analysis in digital histopathology. *Diagnostic Pathology*, 1(61), 2015.

[146] M Sharma, A H Beck, J A Webster, I Espinosa, K Montgomery, S Varma, M van de Rijn, K C Jensen, and R B West. Analysis of stromal signatures in the tumor microenvironment of ductal carcinoma in situ. *Breast Cancer Res Treat*, 123(2):397–404, 2010 Sep 2010.

[147] A. Skodras, C. Christopoulos, and T. Ebrahimi. The jpeg 2000 still image compression standard. *Signal Processing Magazine, IEEE*, 18(5):36–58, Sep 2001.

[148] Sriram Somanchi and Daniel B. Neill. Discovering anomalous patterns in large digital pathology images. In *Proceedings of the 8th INFORMS Workshop on Data Mining and Health Informatics*, 2013.

[149] Rich Stoner. Wholeslide ios application. http://wholeslide.com/.

[150] Y. Sucaet and W. Waelput. *Digital Pathology*. SpringerBriefs in Computer Science. Springer International Publishing, 2014.

[151] Pablo G. Tahoces, J. Ramón Varela, María J. Lado, and Miguel Souto. Image compression: Maxshift roi encoding options in jpeg2000. *Comput. Vis. Image Underst.*, 109(2):139–145, February 2008.

[152] D. Taubman. Kakadu software. http://www.kakadusoftware.com/.

[153] D. Taubman. Remote browsing of jpeg2000 images. In *Proc. IEEE Int. Conf. Image Proc.*, volume 1, pages 229–232, 2002.

[154] D. Taubman and R. Prandolinib. Architecture, philosophy and performance of jpip: Internet protocol standard for jpeg2000. *Proceedings of the SPIE*, 5150:791–805, 2003.

[155] D. Taubman and R. Rosenbaum. Rate-distortion optimized interactive browsing of jpeg2000 images. In *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on*, volume 3, pages III–765–8 vol.2, Sept 2003.

[156] David Taubman and Michael Marcellin. *JPEG 2000: Image Compression Fundamentals, Standards and Practice*. Kluwer Academic Publishers, Norwell, MA, USA, 2001.

[157] The Royal College of Pathologists. Histology. https://www.rcpath.org/discover-pathology/i-want-a-career-studying/human-tissue/histopathology-careers.html.

[158] Anne Treisman and Stephen Gormican. Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, 95(1):15 – 48, 1988.

[159] Anne M. Treisman and Garry Gelade. A feature-integration theory of attention. *Cognitive Psychology*, 12(1):97 – 136, 1980.

[160] M. Triola and W. Holloway. Enhanced virtual microscopy for collaborative education. *BMC Medical Education*, 11(1):4, 2011.

[161] VilppuJ. Tuominen and Jorma Isola. The application of jpeg2000 in virtual microscopy. *Journal of Digital Imaging*, 22(3):250–258, 2009.

[162] Hidetaka Uramoto and Fumihiro Tanaka. Recurrence after surgery in patients with nsclc. *Translational Lung Cancer Research*, 3(4), 2014.

[163] Antoine Vandecreme, Ian Gilman, Mark Salsbery, and Chris Thatcher. Openseadragon, an open-source, web-based viewer for high-resolution zoomable images, implemented in pure javascript, for desktop and mobile. https://openseadragon.github.io/.

[164] M. Veta, J. P. W. Pluim, P. J. van Diest, and M. A. Viergever. Breast cancer histopathology image analysis: A review. *IEEE Transactions on Biomedical Engineering*, 61(5):1400–1411, May 2014.

[165] Mitko Veta, Paul Jand van Diest, Robert Kornegoor, André Huisman, Max A Viergever, and Josien P W Pluim. Automatic nuclei segmentation in h&e stained breast cancer histopathology images. *PLoS ONE*, 8(7), 2013.

[166] Abul K. Abbas Vinay Kumar and Jon C. Aster. *Robbins Basic Pathology*, chapter Cell Injury, Cell Death and Adaptations, pages 1–29. Saunders-Elsevier, 2013.

[167] Ke Wang, Jianjun Xu, Tao Zhang, and Dan Xue. Tumor-infiltrating lymphocytes in breast cancer predict the response to chemotherapy and survival outcome: A meta-analysis. *Oncotarget*, 7:44288–44298, Jul 2016.

[168] Xiang-Yang Wang, Yong-Wei Li, Hong-Ying Yang, and Jing-Wei Chen. An image retrieval scheme with relevance feedback using feature reconstruction and SVM reclassification. *Neurocomputing*, 127:214–230, March 2014.

[169] Xiang-Yang Wang, Hong-Ying Yang, Yong-Wei Li, Wei-Yi Li, and Jing-Wei Chen. A new SVM-based active feedback scheme for image retrieval. *Engineering Applications of Artificial Intelligence*, 37:43–53, January 2015.

[170] Xiangxue Wang, Andrew Janowczyk, Yu Zhou, Rajat Thawani, Pingfu Fu, Kurt Schalper, Vamsidhar Velcheti, and Anant Madabhushi. Prediction of recurrence in early stage non-small cell lung cancer using computer extracted nuclear features from digital h&e images. *Scientific Reports*, 7(1):13543, October 2017.

[171] Ronald S. Weinstein, Anna R. Graham, Lynne C. Richter, Gail P. Barker, Elizabeth A. Krupinski, Ana Maria Lopez, Kristine A. Erps, Achyut K. Bhattacharyya, Yukako Yagi, and John R. Gilbertson. Overview of telepathology, virtual microscopy, and whole slide imaging: prospects for the future. *Human Pathology*, 40(8):1057 – 1069, 2009.

[172] John Wennberg and Alan Gittelsohn. Small area variations in health care delivery. *Science*, 182(4117):1102–1108, 1973.

[173] S. Wienert, D. Heim, K. Saeger, A. Stenzinger, M. Beil, P. Hufnagl, M. Dietel, C. Denkert, and F. Klauschen. Detection and segmentation of cell nuclei in virtual microscopy images: a minimum-model approach. *Sci Rep*, 2, 2012.

[174] L. Winsor. *Tissue processing*. Churchill Livingstone, 1994.

[175] Winston Salem Journal. Error rate for cancer test results slim, but can it be devastating for patients. http://www.journalnow.com/news/local/error-rate-for-cancer-test-results-slim-but-can-it/article_d020348b-335d-526b-b77f-f194c8c765e1.html.

[176] H. Xu, C. Lu, R. Berendt, N. Jha, and M. Mandal. Automatic nuclei detection based on generalized laplacian of gaussian filters. *IEEE Journal of Biomedical and Health Informatics*, 21(3):826–837, 2016.

[177] H. Xu, C. Lu, and M. Mandal. An efficient technique for nuclei segmentation based on ellipse descriptor analysis and improved seed detection algorithm. *IEEE Journal of Biomedical and Health Informatics*, 18(5):1729–1741, Sept 2014.

[178] Yan Xu, Jun-Yan Zhu, Eric I-Chao Chang, Maode Lai, and Zhuowen Tu. Weakly supervised histopathology cancer image segmentation and classification. *Medical Image Analysis*, 18(3):591 – 604, 2014.

[179] Kun-Hsing Yu, Ce Zhang, Gerald J. Berry, Russ B. Altman, Christopher Ré, Daniel L. Rubin, and Michael Snyder. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nature Communications*, 7:12474, August 2016.

[180] Q. Zhang and E. Izquierdo. Histology image retrieval in optimized multifeature spaces. *IEEE Journal of Biomedical and Health Informatics*, 17(1):240–249, Jan 2013.

[181] Yi Zhuang, Nan Jiang, Zhiang Wu, Qing Li, Dickson K.W. Chiu, and Hua Hu. Efficient and robust large medical image retrieval in mobile cloud computing environment. *Information Sciences*, 263:60–86, April 2014.