PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

ESCUELA DE POSGRADO

MAESTRÍA EN PROCESAMIENTO DE SEÑALES E IMÁGENES DIGITALES

Deep Learning for Semantic Segmentation versus Classification in Computational Pathology: Application to mitosis analysis in Breast Cancer grading

Submitted by

Gabriel Alexandro Jiménez Garay

In partial fulfillment of the requirements for the Degree of

Master in Digital Signal and Image Processing

in the Graduate School of the Pontificia Universidad Católica del Perú.

Thesis supervised by:

Daniel Racoceanu

Examining commitee members:

Mauricio David Cerda Villablanca

Maria Gloria Bueno García

Lima, Perú

April 10, 2019

Abstract

Existing computational pathology approaches did not allow, yet, the emergence of effective/efficient computer-aided tools used as a second opinion for pathologists in the daily practice. Focusing on the case of computer-based qualification for breast cancer diagnosis, the present article proposes two deep learning architectures to efficiently and effectively detect and classify mitosis in a histopathological tissue sample.

The first method consisted of two parts, entailing a preprocessing of the digital histological image and a free-handcrafted-feature Convolutional Neural Network (CNN) used for binary classification. Results show that the methodology proposed can achieve 95% accuracy in testing with an F1-score of 94.35%, which is higher than the results from the literature using classical image processing techniques and also higher than the approaches using handcrafted features combined with CNNs. The second approach was an end-to-end methodology using semantic segmentation. Results showed that this algorithm can achieve an accuracy higher than 95% in testing and an average Dice index of 0.6 which is higher than the results from the literature using CNNs (0.9 F1-score). Additionally, due to the semantic properties of the deep learning approach, an end-to-end deep learning framework is viable to perform both tasks: detection and classification of mitosis.

The results showed the potential of deep learning in the analysis of Whole Slide Images (WSI) and its integration to computer-aided systems. The extension of this work to whole slide images is also addressed in the last two chapters; as well as, some computational key points that are useful when constructing a computer-aided-system inspired by the described technology.

To Hidelbrando, Marisol, Andrea, and Jessenia,
for their support, love and strength throughout this journey.

Acknowledgments

This thesis work closes another chapter of my life, and it would not have been possible without the participation and support of many people. I would like to devote a few lines to thank all of those who contributed in one way or another to the accomplishment of this goal.

Firstly, I would like to thank my advisor Prof. Daniel Racoceanu for his guidance, support, trust and for giving me the opportunity to discover and explore one the most interesting topics on biomedical image processing: deep learning. Additionally, I would like to thank my evaluation committee for their comments, and the time they took to review the results in this work.

Secondly, I would like to thank all of my fellow friends in the master program. We shared a lot of great and challenging moments. I would miss our interesting discussions and jokes, but no so much the long nights at the lab. I wish you all the best in all of your dreams.

Special thanks go to my family and my beloved Jessy, for their support through life, and for setting the example as persons, and professionals. Thanks mom for every cup of coffee at night, and the words of encouragement. Thanks dad for your advice, and for sharing your wisdom with me in the countless conversations we had. Thanks Andrea for encouraging me to be a better brother and professional everyday, you are an example to follow. Finally, thanks Jessy for enduring me in this long journey.

Contents

Contents

Chapter 1

Computational Pathology and Image Analysis in Breast Cancer

1.1 Breast Cancer: Overview of the problem

Reports describing global patterns of cancer incidence, mortality rates, and trends still present breast cancer as the most frequently diagnosed cancer and the leading cause of cancer-related death among women, worldwide [1, 2].

Recent estimations by the World Health Organization (WHO) position breast cancer in the top 5 killing types of cancer [3]. Only in the U.S, nearly 330 080 new cases of breast cancer (invasive and non-invasive) and 40 920 deaths in 2018 were reported in [4]. In addition, as of January 2018, there were more than 3.1 million women with a history of breast cancer [4].

Although statistics seem alarming, death rates have been decreasing since the early 2000s. In [1], A. Jemal, *et al.* reported a favorable mortality trend in the United States, the United Kingdom, Australia, and France due to a reduction in the use of menopausal hormone therapy, early detection through mammography and improved treatment. Unfortunately, this scenario was not the same in some Eastern European, Asian, Latin American, and African countries. The unfavorable mortality trend in several of these countries may have been exacerbated by poor survival because of lack of or limited access to early detection services and treatment, and disproportionately high prevalence of gene mutations (e.g. BRCA1 and BRCA2 mutations) in certain regions [1].

There seems to be no specific pattern for breast cancer rates and trends. Even if incidence rates follow an increasing or, in some regions, stable trend, differences in risk factors (e.g. reproductive patterns and obesity), access to early detection and timely treatment amongst countries will determine the best course of action when targeting this problematic in a specific population. In addition, histologic changes and hormonal influences on the breast should be taken into consideration in the different stages of women's lives.

1.2 Breast Cancer Histology

Remarkable technological advances in breast cancer imaging have position mammography as the non-invasive gold standard screening technique to detect breast cancer early. Nevertheless, histological examination of tissue specimens remains the cornerstone for the diagnosis and accurate evaluation of breast diseases [5].

According to the World Health Organization (WHO), there exist several classes for breast tumor classification at the cellular level [6]. The most common type of breast cancer, known as breast carcinoma, is the one that arises from the epithelial cells of the breast. The histological classification of breast carcinoma is also extremely important considering the significant implications of the subtypes in the prognosis and treatment of the disease. The three most frequently subtypes, representing nearly 75% of breast carcinomas, are: *ductal carcinoma in situ* (DCIS), *invasive ductal carcinomas* (IDC) and *invasive lobular carcinoma* (ILC) [7, 8].

1.2.1 Histopathology and tissue analysis procedure

Histopathology refers to the study of microscopic structures of diseased or abnormal tissues. Generally, this examination is prescribed after observing physiological symptoms (e.g. tumors) in a patient. The process of this study starts by removing tissue samples from the breast by means of biopsy or surgical resection. Depending on certain factors such as location, size, or tumor composition (e.g. solid tumor or cyst) the doctor will determine whether to perform a Fine Needle Aspiration (FNA) or a Core Needle Biopsy.

Once obtained, tissue samples are prepared for viewing under a microscope. To observe the tissue under a microscope, the sample is first sectioned into thin cross sections with a microtome. Then, fixation is used to stop the metabolic activities and stabilize the tissue to prevent decay and preserve the histological structure. The fixation is made either with a chemical fixative (e.g. formaldehyde) or frozen section procedure. The latter technique of preparation is faster and therefore principally used in the examination of tissue during surgery. However, the histological slides produced by the frozen section procedure are of lower quality [7, 9].

After fixation, the tissue sections are mounted on glass slides and stained with pigments (e.g. H&E, Saffron, or molecular biomarkers) to enhance the contrast and highlight specific cellular structures under the microscope. Hematoxylin-and-eosin (H&E) staining is used widely in a histopathological analysis. Hematoxylin stains the cell nuclei in a blue-purple shade, while eosin stains cytoplasm and stroma in various shades from reddish to a pinkish color, and collagen in a pale pink shade (see Figure 1).

Finally, pathologists observe and evaluate the architecture of the tissue, the distribution and the morphology of the cells in very fine details, and determines the level of the severity of the disease. Due to different biological conditions of the organism, this process is time-consuming and requires highly skilled expertise.

1.2.2 Breast cancer grading

Histopathology plays an important part in determining the treatment strategy for patients with breast cancer. This strategy should also rely on an objective evaluation of the aggressiveness of the tumor. One

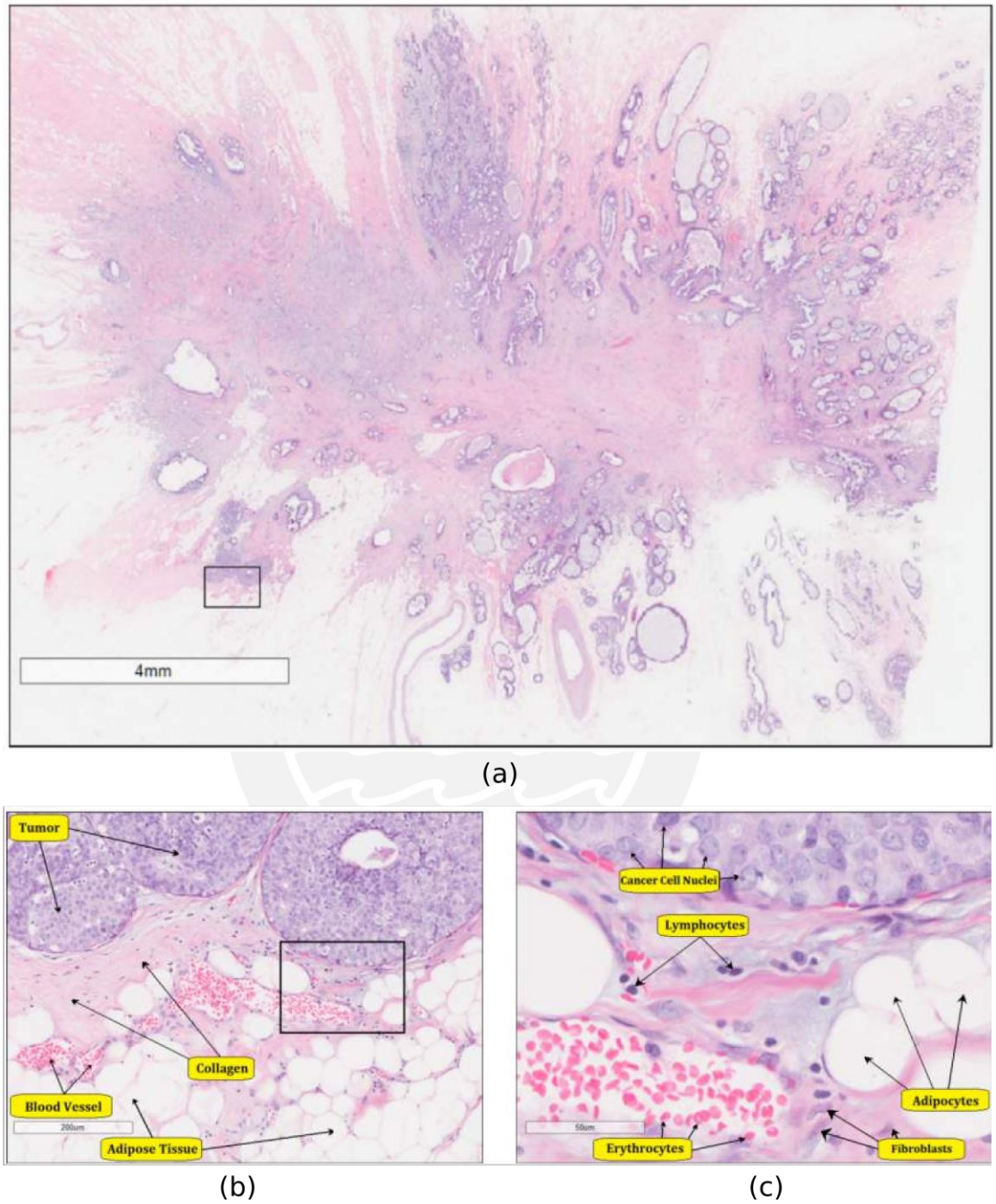


Figure 1: H&E staining effects digitized breast cancer tissue sample. (a) Complete digitized glass slide (0.5X magnification). (b) Different structures and its color properties in a region at 10X magnification. (c) Smaller structures and its color properties seen at 40X magnification. Images taken from [7].

of the most known and used systems is the Nottingham Histology Score, also refer as Scarff-Bloom-Richardson grading system [10]. The following descriptions summarize the criteria used to determine the Nottingham Histologic Score and the final grading of the tissue in more detail as seen in [11].

- **Glandular/Tubular Differentiation**

  Score 1: $\geq$ 75% of tumor area forming glandular/tubular structures.
  Score 2: 10% to 75% of tumor area forming glandular/tubular structures.

  Score 3: $\leq$ 10% of tumor area forming glandular/tubular structures.

- **Nuclear Pleomorphism**

  Score 1: Nuclei small with little increase in size in comparison with normal breast epithelial cells, regular outlines, uniform nuclear chromatin, little variation in size.

  Score 2: Cells larger than normal with open vesicular nuclei, visible nucleoli, and moderate variability in both size and shape.

  Score 3: Vesicular nuclei, often with prominent nucleoli, exhibiting marked variation in size and shape, occasionally with very large and bizarre forms.

- **Mitotic Count**

  Score 1: less than or equal to 7 mitoses per 10 high power fields.

  Score 2: $8 - 14$ mitoses per 10 high power fields.
  Score 3: equal to or greater than 15 mitoses per 10 high power fields.

- **Overall grade**

  Grade 1: scores of 3, 4, or 5.

  Grade 2: scores of 6 or 7.

  Grade 3: scores of 8 or 9.

The mitotic count score criteria is refer to a high power field (HPF) which is a parameter that varies according to the configuration of the microscope. On average, the diameter of a HPF is 0.55 mm [12].

1.3 Histopathological Image Analysis

1.3.1 Digital Pathology

Digital systems were introduced to the histopathological examination in order to deal with complex and huge amount of information obtained from tissue specimens. Digital images were originally generated by mounting a camera on the microscope. The static images captured only reflected a small region of the glass slide, and the reconstruction of the whole glass slide was not frequently attempted due to its complexity and time-consuming. However, precision in the development of mechanical systems has made possible the construction of devices such as whole slide digital scanners. The stored high-resolution images allow pathologists to view, manage, and analyzed the digitized tissue on a computer monitor in a similar manner as under an optical microscope.

Whole slide imaging (WSI) technology, also referred to as *virtual microscopy*, have proven to be useful in a wide variety of applications in pathology (e.g. image archiving, telepathology, image analysis). In essence, a WSI scanner operation principle consists in moving the glass slide a small distance

every time a picture is taken in order to capture the entire tissue sample. Every WSI scanner has six components: (a) a microscope with lens objectives, (b) light source (bright field and/or fluorescent), (c) robotics to load and move glass slides around, (d) one or more digital cameras for capture, (e) a computer, and (f) software to manipulate, manage, and view digital slides [13]. The hardware and software used for these six components will determine the key features to analyze when choosing a scanner. In [13], N. Farahani, *et al.* compared 11 WSI scanners from different manufacturers regarding imaging modality, slide capacity, scan speed, image magnification, image resolution, digital slide format, multilayer support, and special features their hardware and software may offer. This study showed that robotics and hardware used in a WSI scanner are currently state-of-the-art and almost standard in every device. Software, on the other hand, has some ground for further development.

1.3.2 Whole Slide Image reconstruction

The Digital Imaging and Communications in Medicine (DICOM) standard was adopted to store WSI digital slides into commercially available PACS (Picture Archiving and Communication System) and facilitate the transition to digital pathology in clinics and laboratories. Due to the WSI dimension and size, a new pyramidal approach for data organization and access was proposed by the DICOM Standards Committee in [14].

A typical digitalization of a 20mm$\times$15mm sample using a resolution of 0.25μm/pixel, also referred to as 40$\times$ magnification, will generate an image of approximately 80000$\times$ 60000 pixels. Considering a 24-bit color resolution, the digitized image size is about 15GB. Data size might even go one order of magnitude higher if the scanner is configured to a higher resolution (e.g. 80 $\times$ 100 $\times$), Z planes are used or additional spectral bands are also digitized. In any case, conventional storage and access to these images will demand excessive computational resources to be implemented into commercial systems. Figure 2 describes the conventional approach (i.e. *single frame* organization) which stores the data in rows that extend across the entire image.

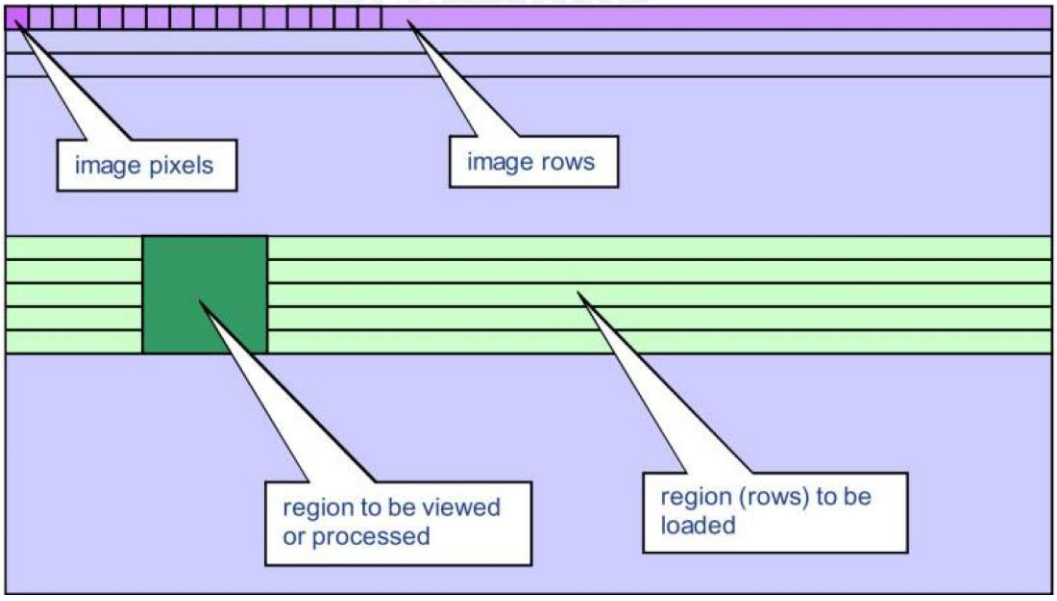


Figure 2: *Single frame* organization of Whole Slide Images. Image taken from [14].

Other types of organizations have also been studied. Figure 3 describes the storage of pixels in *tiles*, which decreases the computational time for visualization and manipulation of WSI by loading only the subset of pixels needed into memory. Although this approach allows faster access and rapid panning of the WSI, it fails when dealing with different magnifications of the images as it is the case in WSI scanners. Figure 4 depicts the issues with rapid zooming of WSI. Besides having to load a larger subset of pixels into memory, algorithms to perform the down-sampling of the image are time-consuming. At the limit, to render a low-resolution thumbnail of the entire image, all the data scanned must be accessed and down-sampled [14]. Stacking precomputed low-resolution versions of the original image was proposed in order to overcome the zooming problem. Figure 5 describes the *pyramidal* structure used to store different down-sampled versions of the original image. The bottom of the pyramid corresponds to the highest resolution and it goes up to the thumbnail (lowest resolution) image. For further efficiency, tiling and pyramidal methods are combined to facilitate rapid retrieval of arbitrary subregions of the image as well as access to different resolutions. In this combined approach, each image in the pyramid is stored as a series of tiles as shown in Figure 6.

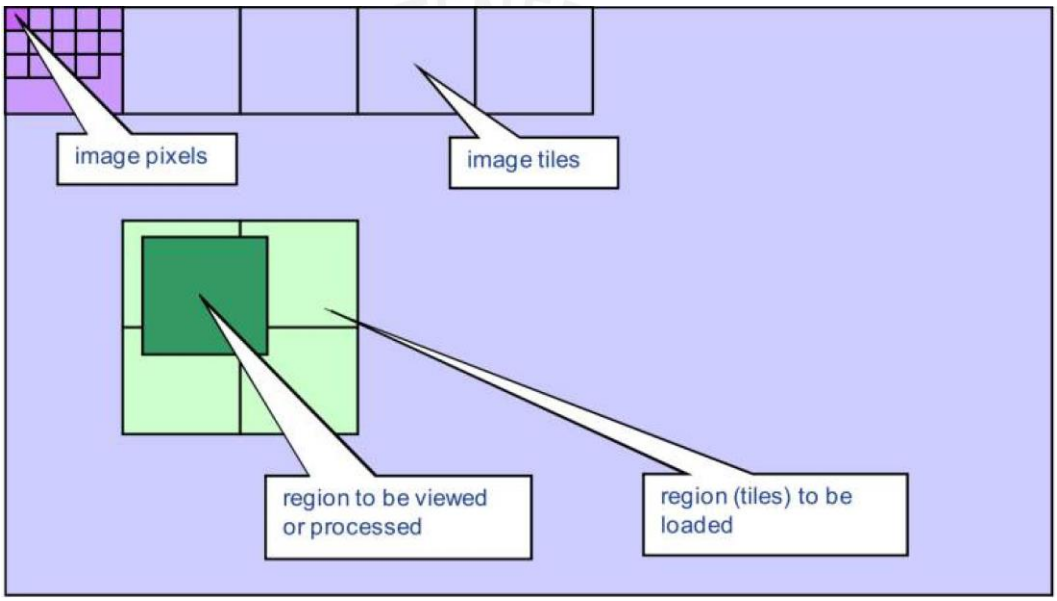


Figure 3: *Tiled image* organization of Whole Slide Images. Image taken from [14].

As mentioned in previous paragraphs, WSI can occupy several terabytes of memory due to the data structure. Depending on the application, lossless or lossy compression algorithms can be applied. Lossless compression typically yields a 3X–5X reduction in size; meanwhile, lossy compression techniques such as JPEG and JPEG2000 can achieve from 15X–20X up to 30X–50X reduction respectively [14]. Due to no standardization of WSI files' formats, scan manufacturers may also develop their own proprietary compression algorithms based on JPEG and JPEG2000 standards. Commercial WSI formats have a mean default compression value ranging from 13X to 27X. Although the size of WSI files is considerably reduced, efficient data storage was not the main issue when designing WSI formats for more than 10 years. In [15], H. Helin, *et al.* addressed this issue and proposed and optimization to the JPEG2000 format which yields up to 176X compression. Although no computational time has been reported in the aforementioned study, this breakthrough allows for efficient transmission of data through systems relying on Internet communication protocols.

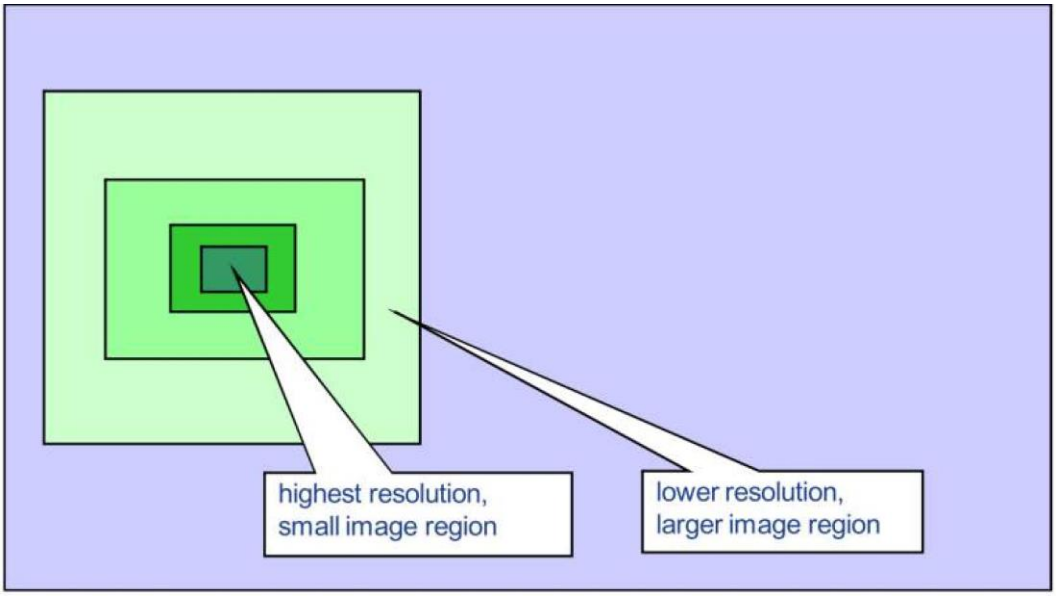


Figure 4: Rapid zooming issues using *tiled image* organization. Access to lower resolution versions imply large amount of data loaded into memory. Image taken from [14].

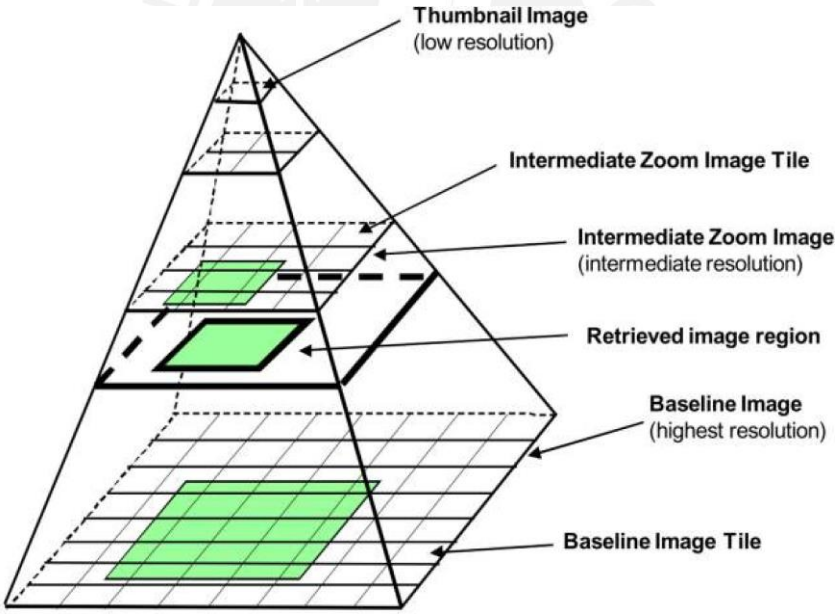


Figure 5: *Pyramidal* organization of Whole Slide Images. Image taken from [14].

1.3.3 Computational Pathology

Computational pathology is a term which refers to the integration of WSI technology and image analysis tools in order to perform tasks that were too cumbersome or even impossible to undertake manually. Image processing algorithms have evolved yielding enough precision to be considered in clinical applications. A few examples mentioned in [13] include morphological analysis to quantitatively measure

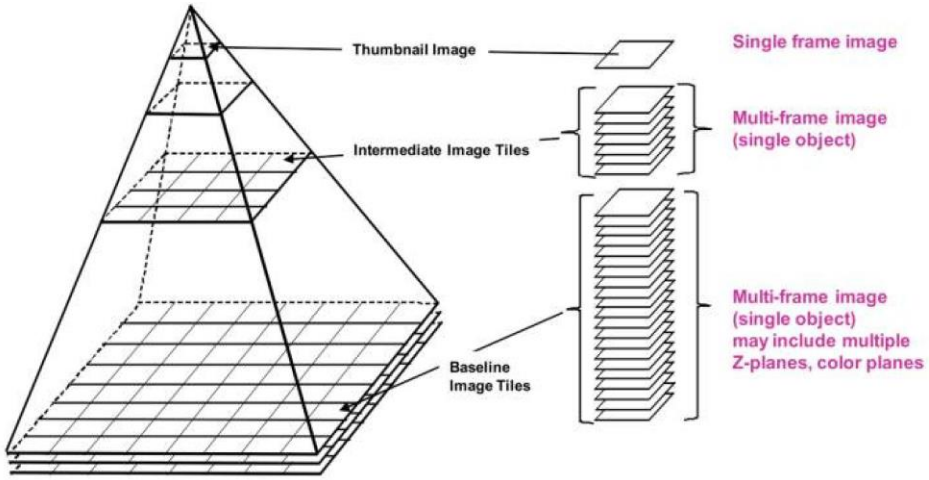


Figure 6: Integration of *tiled image* organization and *pyramidal* structure for Whole Slide Images. Image taken from [14].

histologic structures [16]; automated selection of regions of interest such as areas of most active proliferative rate [17]; and automated grading of tumors [18]. Moreover, educational activities have also benefited from the development of computational pathology. Virtual tutoring, online medical examinations, performance improvement programs, and even interactive *illustrations* in articles and books are being implemented thanks to this technology [13].

In order to validate a WSI scanner for clinical use, several tests are conducted following the guidelines developed by the College of American Pathologists (CAP). On average, reported discrepancies between digital slides and glass slides are in the range of 1% to 5%. However, even glass-to-glass slide comparative studies can yield discrepancies due to observer variability and increasing case difficulty [13].

Although several studies in the medical community have reported using WSI scanners to perform the analysis of tissue samples, pathologists remain reluctant to adopt this technology in their daily practice. Lack of training, limiting technology, shortcomings to scan all materials, cost of equipment and regulatory barriers have been reported as the principal issues [13]. In fact, it was until early in 2017 that the first WSI scanner was approved by the FDA and released to the market [19]. Nevertheless, WSI technology has the potential to enhance the practice of pathology by introducing new tools which help pathologists provide a more accurate diagnosis based on quantitative information.

1.4 Motivation and objectives

Besides its contribution to computer-aided diagnosis, computational pathology has opened up a new dimension in the study of complex diseases, by integrating innovative image analysis tools able to provide new features exposing different characteristics of the disease, impossible to analyze using classical microscopy techniques. In breast cancer, for instance, it allows pathologists to perform a comprehensive study of the microscopic structures in the tumor, and grade the tissue samples to yield an accurate diagnosis.

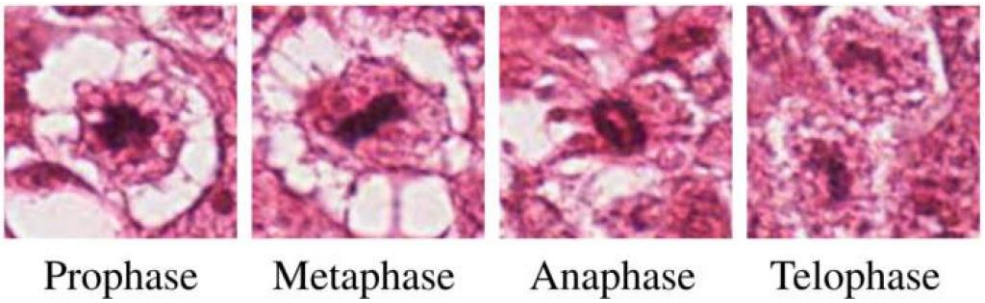


Figure 7: Mitosis stages seen in an WSI. Image taken from [20].

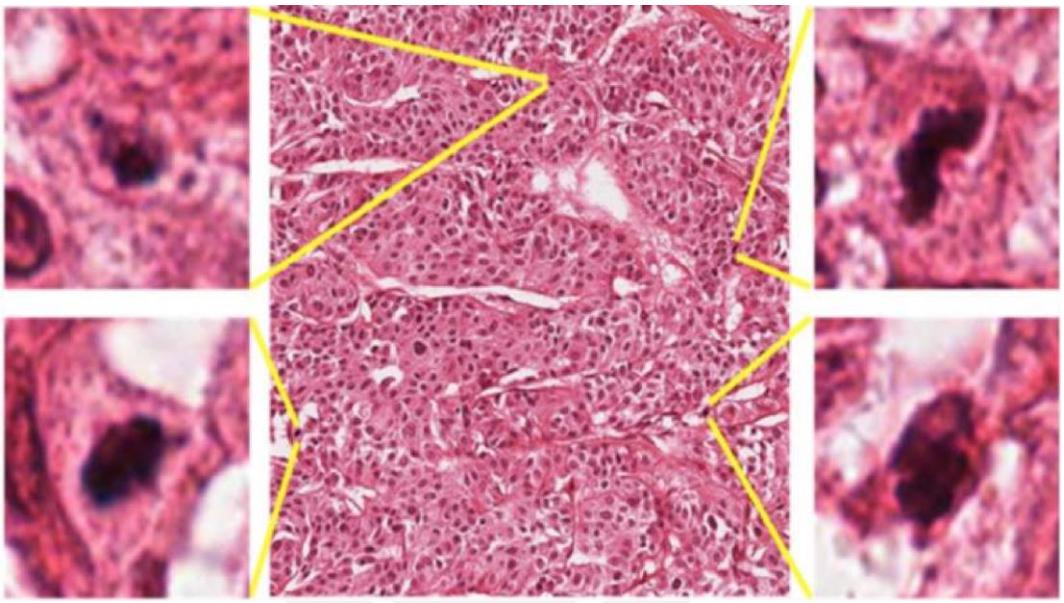


Figure 8: Different shapes and sizes of mitosis seen in a WSI.

As it was described in a previous section, grading of tissue samples is not only a key point for accurate diagnosis but also important in the selection of the patient's treatment. Pathologists frequently use the Nottingham Grading System to obtain an overall grade of the tissue being analyzed. From the three criteria evaluated by the NGS, mitotic counting is the most challenging tasks for doctors and image analysis algorithms.

Mitosis occupies a portion of the cell cycle. It has 6 phases (i.e. *interphase, prophase, prometaphase, metaphase, anaphase*, and *telophase*) followed by a final stage, in which the cell physically separates, named *cytokinesis* (see Figure 9). In the digitized version of a glass slide at enough magnification, we can visualize some of the stages of mitosis. Figure 7 depicts four different patterns frequently observed when evaluating a tissue. These differences in visual aspects make the automatic detection of mitosis a very complicated task.

As shown in Figure 8, there are some other characteristics which are also worth considering when developing a robust image analysis algorithm. In a WSI at 40X magnification we can find the following differences regarding mitosis:

- They do not appear with the same shape and size in the WSI.
- There exists considerable variations of intensity in the pixel surrounding and inside the nuclei. This is probably a direct effect of the staining process.
- Mitosis resemble other types of objects such as apoptosis, necrosis, dust particles, lymphocytes, among others.
- The number of mitosis in a HPF is usually lower than 10.

Regardless of the remarkable progress of digital pathology, the integration with image analysis techniques is still in its early beginnings. Therefore, the purpose of this thesis is to expand the research in computational pathology by introducing and comparing two deep learning approaches in order to automatically detect and classify mitosis in a whole slide image. In order to achieve this, several secondary objectives were also defined:

- Normalize images from different scanners in order to account for the differences in glass slide preparation, especially in the staining process.
- The methods studied should be robust and usable in clinical applications. Therefore, the methodology proposed should not rely on classical image processing techniques due to its variability and dependency on hand-crafted features. The introduction of image analysis algorithms used in natural images (e.g. cars, pedestrian, etc.) will definitely motivate the generation of applications in computational pathology using state-of-the-art image analysis techniques.

1.5 Overview of the document

This thesis is structured in five chapters.

- **Chapter 2:** reviews the algorithms used in recent research regarding mitosis detection and classification. It also introduces deep learning concepts of convolutional neural networks and semantic segmentation in order to fully understand the methodology proposed.
- **Chapter 3:** describes the methodology followed to detect and classify mitosis in frames taken from a WSI. The chapter describes all the preprocessing steps previous to the integration with two different deep learning approaches.
- **Chapter 4:** presents the results of the comparison between two deep learning approaches. It also discusses some key point regarding the implementation of the methodology, its robustness, and clinical applications. Finally,
- **Chapter 5:** presents the conclusions of the work and some perspectives for further research on the topic.

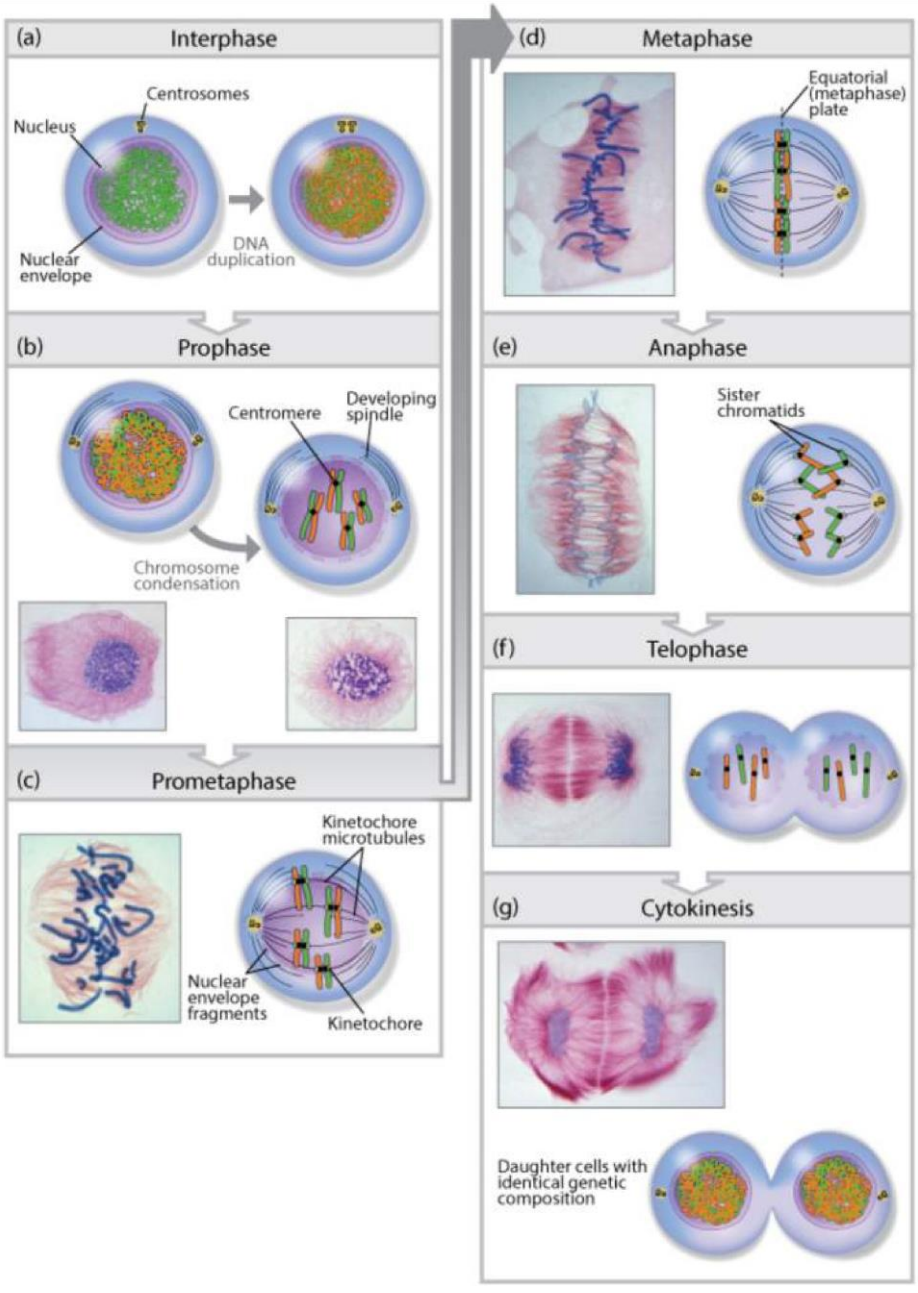


Figure 9: Mitosis stages. Image taken from [21].

Chapter 2

Review of Histopathological Image Analysis Methods

Microscopy has evolved remarkably over the years by incorporating imaging processing techniques. In the last decade, it has also benefited from the integration of artificial intelligence (AI) algorithms which have been shown to improve diagnostic accuracy and provide quantitative metrics useful for pathologists. In fact, in 2018, researchers at Google AI Healthcare, reported the integration of modern AI into a standard microscope to detect metastatic breast cancer in sentinel lymph nodes and prostate cancer in prostatectomy specimens [22]. Efforts made in this field are frequently driven by the need to overcome financial and workflow barriers encounter when using whole slide imaging scanners (e.g. prices of WSI scanners, IT infrastructure, operating personnel, among other). However, due to the advantages the latter technology poses, several researchers have been studying different alternatives to integrate AI and image processing algorithms with WSI. The present chapter will focus on the literature concerning the medical problematic detailed in Chapter 1: mitosis analysis in breast cancer; as well as those methods which can be applied to WSI analysis.

2.1 Detection and classification of cell nuclei in histological images

Operator-bias in cancer grading is undoubtedly one of the most important problems of cancer diagnosis and grading. In particular, nuclei analysis is the most cumbersome task for pathologists due its different properties and representations in a digital image. Regarding breast cancer and mitosis analysis, the image processing algorithms studied in the literature are categorized into two different groups: segmentation and classification.

Detection and segmentation of mitosis have been extensively studied in the literature. In [20], the authors suggested two different subcategories for segmentation algorithms: region based cell segmentation and boundary based cell segmentation. Among region-based approaches, X. Yang, *et al.* proposed a novel marker-controlled watershed algorithm which can effectively segment clustered cells with fewer over-segmentation [23]. A. Nedzved, *et al.* also applied morphological operations, and combined them with thinning algorithms to segment cells in histological images [24]. In order to improve robustness, different nuclear models using different morphological features were proposed and validated by G.

Lin, *et al.* in [25]. Moreover, in [20], authors proposed a segmentation controlled by the relative entropy between cells and background using opening and closing morphological operations. Similarly, A. Chowdhury, *et al.* applied entropy thresholding to detect and segment monocyte cells in order to track them using bipartite graph matching algorithms [26]. Contextual information from objects in an image was also reported as a methodology for detection and segmentation of cell nuclei. M. Seyedhosseini, *et al.* introduced a framework called multi-class multi-scale series contextual model, which uses contextual information from multiple objects and at different scales for learning discriminative models in a supervised setting [27]. Following the model of extracting information from several scales, Y. Al-Kofahi, *et al.* proposed an automatic segmentation algorithm using graph-based binarization and multi-scale Laplacian-of-Gaussian filtering [28]. On the other hand, regarding the second subcategory, level set methods are quite popular in the literature for boundary-based segmentation [29, 30, 31]. Active contour was also reported in [32] for automated segmentation of breast cancer nuclei. To sum up, most region-based methodologies assume a similarity in the size or region properties of different cells which is not the case for mitosis in breast cancer. Meanwhile, boundary-based algorithms are strongly dependent on the initial conditions and stopping criteria. Either way, all algorithms rely of prior knowledge for detection and segmentation which does not allow the development of fully automated systems.

Literature for cell nuclei classification is also as extended as the one reported for segmentation. O. Sertel, *et al.* proposed pixel level likelihood functions and component-wise two-step thresholding for mitosis counting in digitized images. They reach an average sensitivity of 81.1% and 12.2 false positive detections [33]. Unsupervised clustering algorithms were also studied by V. Roullier, *et al.* In [34], authors proposed a graph-based multi-resolution approach which yielded a good sensitivity ($\geq$70%) for Grade 1 and Grade 3 breast cancer in WSI. Similarly, B. Weyn, *et al.* proposed a k-nearest neighbor classification approach combined with wavelets features for multiscale image analysis of isolated nuclei in invasive breast cancer tissue. Authors achieved a recognition score of 76%; however, the presence of false negative cases restricted the immediate practical application [35]. Other machine learning approaches were also considered in the literature. H. Irshad, *et al.* evaluated different classifiers (decision tree, linear kernel SVM and non-linear kernel SVM) using color and texture features (e.g. blue ratio, Haralick, HMAX, and SIFT), and reached 76% in f-score [36]. Y. Tao, *et al.* also studied SVMs using 59 parameters combining geometric properties and intensity information. Results showed an 89.2% in accuracy using a specific subset of features [37]. An ensemble of cascade adaboosts [38] were also reported to yield significant improvement in mitosis classification. However, this method fails to demonstrate its robustness as region features used in the study may have significant variations among various datasets. Genetic algorithms for mitosis classification were proposed by R. Nateghi, *et al.* in [39]. The authors used genetic optimization algorithms to eliminate potential non-mitosis from the histological image. Then, texture features were computed from the remaining potential cell nuclei and classify using SVMs. Results were promising (78.47% f-score); however, they do not make a considerable improvement in the machine-learning category for mitosis classification. Taking a step further, the authors in [40] proposed the first complete grading system of breast cancer using histological images. In particular, for mitosis detection, they proposed two Gaussian models for classification using geometric and intensity features. The results obtained were promising; however, the system's scores tend to be slightly lower than pathologist's scores. With the development of artificial intelligence algorithms for image analysis, a few studies were conducted regarding the mitosis classification task. In [41], authors proposed a deep max-pooling convolutional neural network approach for mitosis classification which achieved 0.72 f-score. Recently, M. Saha, *et al.* improve the performance of convolutional neural networks (CNN) by adding additional information from hand-crafted features

[42]. Results showed a promising 0.9 f-score which entails the out-performance of deep-neural-network approaches over classical machine-learning ones. T. Araujo, *et al.* also contributed to the study of CNN by proposing a methodology which uses the features obtained from the convolutional layers as inputs of a support vector machine (SVM) classifier. This method achieved an 83.3% accuracy and 95.6% sensitivity [43]. Although these last results suggest better ways to address the mitosis challenges in histological imaging, the lack of available ground-truth data is still a major issue for transferring the developments in deep learning to the biomedical imaging domain. S. Albarqouni, *et al.* evaluated a method combining crowd-sourcing for data annotation and convolutional neural networks for classification. However, this proposal needs further research as the results reported do not exceed 0.8 in f-score [44]. To sum up, Table I summarizes and categorizes the methodologies exposed in the last paragraphs according to the image analysis algorithms employed.

Table I: Summary of the literature review.

Tasks	Categories	Methods proposed
Mitosis detection/segmentation	Region-based algorithms	Marker control watershed [23], Morphology [24, 25], Entropy thresholding [26], Contextual model [27], Graph binarization [28].
	Boundary-based algorithms	Level sets [29, 30, 31], Active contours [32].
Mitosis classification	Image processing and analysis	Pixel level likelihood [33], Unsupervised clustering [34], Wavelets [35], Textural features [36], Adaboosts [38].
	Machine learning	REMSS [20], SVM [37], Genetic algorithms [39], AggNet [44], Deep learning [45].

2.1.1 Challenges in mitosis detection and classification

Since the development of WSI scanners, many researchers have stepped in contributing with state-of-the-art methodologies to improve and integrate this technology in clinical procedures. In the academic field, a few challenges have been organized to get the attention of the scientific community and to address one of the most important issues concerning this technology: access to public datasets for validation of proposed algorithms. Although its importance, there have been only a few of these events organized in the past 6 years.

MITOS (Mitosis Detection in Breast Cancer) [46] was the first contest organized using WSI. It was presented at the International Conference on Pattern Recognition in 2012 (ICPR 2012). The dataset released was of relatively small size (5 WSI, 10 annotated HPFs per slide) and it did not account for the inter-subject variability in tissue appearance and staining. However, the methodologies presented helped expand the literature with promising results (0.782 f-score) [41]. To address the issues of the MITOS dataset, the *AMIDA (Assessment of Mitosis Detection Algorithms)* contest was released and

presented in the conference organized by the Medical Image Computing and Computer Assisted Intervention Society (MICCAI) in 2013. The dataset consisted of 23 subjects (12 for training and 11 for testing), with more than one thousand annotated mitotic figures by multiple observers. The top-performing method achieved a 0.611 f-score, and an error rate that is comparable to the inter-observer agreement among pathologists [47]. A year later, at ICPR 2014, the *MITOS-ATYPIA-2014* contest was released. It was organized as a follow-up and an extension of the MITOS contest. The major improvements were the additional nuclear atypia scoring and annotations provided by two senior pathologists and three junior pathologists [48]. In 2015, the INEB (Institute of Biomedical Engineering) organized a grand challenge for its International Symposium in Applied Bioimaging (Bioimaging 2015). The objective of the competition was to classify WSI of tissue samples into four different categories: normal tissue; benign lesion; carcinoma in situ; or invasive carcinoma. This implies that besides the analysis of nuclei, the methodologies proposed should also be able to retrieve information about the overall tissue organization. Results published in [43] showed the contribution of this challenge to the integration of deep learning algorithms into CAD systems for cancer diagnosis. Recently, in 2018, four different on-going challenges concerning nuclei analysis in WSI have been released. The *BACH (Breast Cancer Histology Image)* challenge was presented in the International Conference on Image Analysis and Recognition (ICIAR 2018) and it extended the Bioimaging 2015 challenge by incorporating a pixel-wise labeling task. Preliminary results showed an 87% accuracy for the nuclei classification [49]. The remaining challenges are organized by MICCAI. MoNuSeg [50], presented at MICCAI 2018, aims at segmenting nuclei from WSI of different patients and multiple organs. Meanwhile, in [51] the competition focuses on images extracted from a set of Glioblastoma and Lower Grade Glioma whole slide tissue images. Preliminary results in this challenge showed a 0.85 on the average of to dice coefficients. Finally, the challenge presented in [52] evaluates the performance of automated classification algorithms when information from two types of imaging data (i.e radiology images and pathology images) is used. Whole slides images correspond to two subtypes of lower grade glioma tumor cases: Oligodendroglioma and Astrocytoma. No preliminary results have been reported.

2.2 Deep learning background review

Lately, applications using deep learning along classical image processing techniques have increased. In the medical field, a deep learning approach is often preferred over classical techniques due to its robustness when dealing with different image modalities (e.g. MRI, CT, X-ray, WSI). Its use, relatively new in medical applications, entails the study of different methodologies such as Artificial Neural Networks (ANNs), Convolutional Neural Networks (CNNs), Generative Models, among others. The present thesis focuses on two specific groups of deep learning algorithms: Convolutional Neural Networks (CNNs), and Semantic Segmentation. A general description of each approach is provided as they will be compared in the following chapters.

2.2.1 Convolutional Neural Networks

Since the introduction of Convolutional Neural Networks by LeCun, *et al.* in [53] and further development in [54], CNNs have become the standard tool to analyze and process 2D signals. The design principles of CNNs are drawn from neuroscience following the study of neurophysiologists David Hubel and Torsten Wiesel. From a simplified point of view, CNNs resemble a part of the brain called V1, also

known as the primary visual cortex. In particular, a convolutional network layer is designed to capture three properties of V1: (a) it has a two-dimensional structure following the spatial map arrange of V1; (b) it has many detector units often characterized by linear functions mimicking simple cells found in V1; and (c) it has pooling units to mimic complex cells in V1 which are invariant to small shifts in the position of the feature [55]. These three properties are the basis for the three stages of a convolutional network layer which are shown in Figure 10.

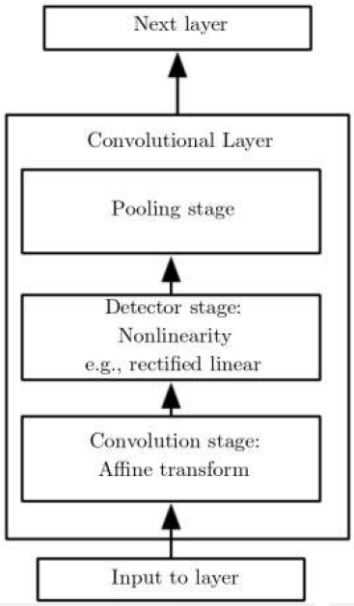


Figure 10: Three stages of a Convolutional Network layer. Image adapted from [55].

- *Convolution*, defined by Equation 1, represents the mathematical relationship between an input $x(t)$ and a kernel $w(t)$ to generate an output often referred as *feature map*. These feature maps, which can be generalized to a tensor of any dimension (see Equation 2), represent particular characteristics found in a signal which allow the characterization of the inputs making the classification possible. The convolution provides a means for working with inputs of variable sizes and it also leverages three key aspects of data organization: (a) *sparse interactions*, meaning most of the important features are located in and can be represented with a small amount of information (e.g. corner's pixels in an image); (b) *parameter sharing*, referring to the set of parameters found in every location (e.g. every pixel of an image); and, (c) *equivariant representations*, meaning the learning process will be invariant due to translations of the input signals. Therefore, compared to previous machine learning approaches, convolution represents a dramatical improvement on computation and statistical efficiency (e.g. less memory requirement and parameter sharing) [55].

$$s(t) = (x * w)(t) \tag{1}$$

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(m, n) K(i - m, j - m) \tag{2}$$

- *Activation function*, in an artificial neuron, computes the weighted sum of the feature map in order to determine the probability of the inputs representing a specific class. There are different

functions to select and they vary according to the application and the data being processed. The most frequently used function is the Rectified Linear Unit (ReLU), defined as $\max(0, x)$. It has the advantage of the non-linearity; high efficiency as nearly 50% of the neurons in a layer do not activate; and, low computational cost as it involves less complex mathematical operations as opposed to other functions such as sigmoid or tanh. This stage is sometimes called the *detector* stage [55].

- *Pooling functions* are used to compute aggregated statistics of the activated neurons at various locations in a signal. It allows further modifications of the outputs (i.e. deeper network) and also invariance due to translations. The most common function used in this stage is the *max pooling*.

In [56], Alex Krizhevsky, *et al.* proposed a deep convolutional neural network to classify 1.2 million images in the ImageNet dataset achieving a top-1 and top-5 error rate of 37.5% and 17%. This approach, currently known as the AlexNet, consists of five convolutional layers followed by max-pooling layers, three fully connected layers, and a final 1000-way softmax. The architecture used ReLU nonlinearity as activation functions and demonstrated that training error decreases slowly when using saturating neuron models such as the tanh. Figure 11 depicts the entire architecture and how it was implemented into two GPU in order to reduce computational training time.

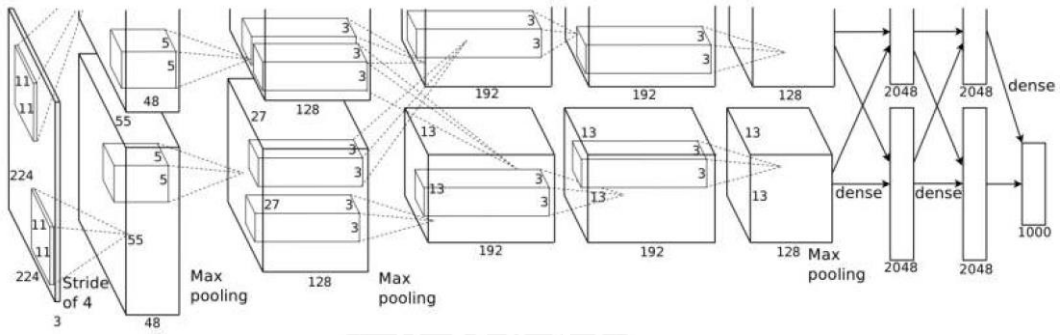


Figure 11: An illustration of the architecture of our CNN, explicitly showing the delineation of responsibilities between the two GPUs. One GPU runs the layer-parts at the top of the figure while the other runs the layer-parts at the bottom. The GPUs communicate only at certain layers. Image taken from [56].

Although in 2014, GoogLeNet [57] was introduced and reported certain improvements in the classification of the ImageNet dataset, the AlexNet remains a valid architecture to start developing image processing applications, especially when dealing with small datasets.

2.2.2 Semantic Segmentation

For certain applications in computer vision (e.g. surveillance, precision agriculture, self-driving cars, Geo-sensing, biomedical image processing) a better understanding of the image context is sometimes needed. Therefore, several methodologies regarding pixel-wise classification (i.e. outputs including localization) have been reported, starting from patch classification [45] (i.e. each pixel is given a class label based on its neighborhood), fully convolutional neural networks [58], up to more complex architectures such as the ones presented in [59] and [60].

For neural networks to work properly, a great amount of input data is necessary. Semantic neural networks and CNN are frequently trained using large datasets such as ImageNet, Pascal VOC or COCO datasets. These datasets correspond to natural images and do not contain relevant information to work in biomedical applications. However, in 2012, O. Ronneberger, *et al.* proposed a novel architecture build upon the fully convolutional network [61]. The U-net, depict in Figure 12, has 23 convolutional layers and none fully connected layers. Due to the very little training data available they used elastic transformations to augment the data which also allow the network to learn invariance to such deformations. Authors used stochastic gradient descend for training and ReLU functions as activations. They trained in two datasets achieving 92% and 77.5% intersection over the union (IOU), which was significantly higher than the state-of-the-art (83% and 46% respectively) [61]. This breakthrough in deep learning architectures clearly set the basis for semantic segmentation in biomedical applications.

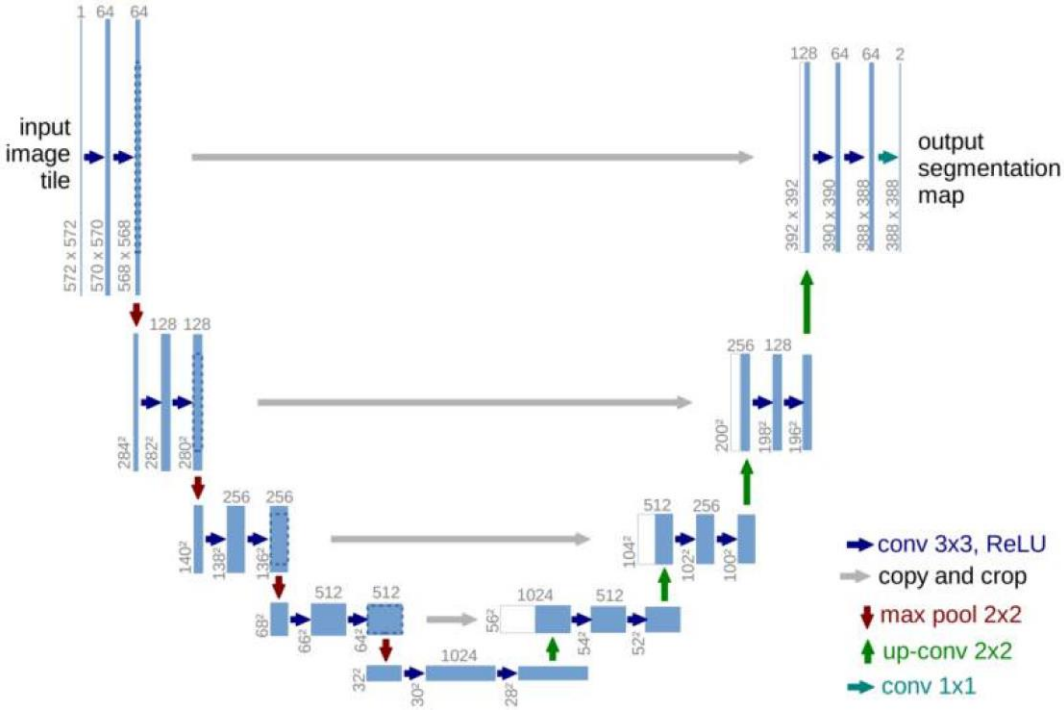


Figure 12: U-net architecture (example for 32x32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations. Image taken from [61].

Chapter 3

Mitosis detection in WSI using deep learning

Two different deep learning approaches were described in the previous chapter. We build upon these two methodologies (i.e. AlexNet and U-Net) a framework to compare both neural networks and define which is more suitable for biomedical applications, especially when dealing with Whole Slide Images. The present chapter will focus on the methodology followed in this project by describing the datasets, the preprocessing steps and the validation metrics for both deep learning approaches.

3.1 Dataset

Images used to evaluate the proposed methodologies were gathered from two public datasets corresponding to the ICPR-2012 contest [46, 62], and the MITOS-ATYPIA-2014 challenge [48]. Both datasets contain images corresponding to 10 High Power Fields (HPF) at 40x magnification, selected from a Whole Slide Image (WSI) by experienced pathologists. Two scanners were used to digitize the tissue samples: Aperio Scanscope XT with 1pixel = $0.2456\mu m$ resolution; and Hamamatsu Nanozoomer 2.0-HT with 1pixel = $0.2273\mu m$ resolution. The datasets combined add up to 2177 frames which are further processed to obtain the input patches for both neural networks. Table II summarizes the distribution of the data for both neural networks. To build the masks for the U-Net a manual segmentation of each mitosis and non-mitosis in the MITOS-ATYPIA 2014 dataset was necessary due to the lack of ground truth for region segmentation. In addition, data augmentation was only needed for the AlexNet as CNNs require large datasets as described before. This was implemented by rotating the patches in four different angles (0°, 45°, 90°, 180°), and incorporating data from a third dataset (AMIDA13) reported in [42]. Patch sizes were dependent on the network configuration, and, in the case of the AlexNet, the 71×71-pixel size patch corresponds to the largest mitosis found in both datasets. Finally, before extracting the patches, a color normalization was applied to every frame in the datasets.

Table II: Summary of the datasets used in the AlexNet and U-Net.

Neural Network	Patch size	Mitosis	Non-mitosis	Dataset
AlexNet	71×71	5850	7563	ICPR2012 & MITOS-ATYPIA2014 & AMIDA13
U-Net	128×128	327	0	ICPR2012
		946	3585	MITOS-ATYPIA2014

3.2 Color normalization

Hematoxylin-and-eosin (H&E) staining is widely used in a histopathological analysis. However, despite more advanced and automatized staining devices (if existing) this staining still remains highly dependent on staining providers, concentration, chemical reactivity, storage conditions, and timing. These factors, combined with the fact that light transmission depends on tissue thickness as on mechanical and optical properties of the scanners, generate one of the most common problems when using automated image processing algorithms: the color variability [7]. In [63], Hoffman et al., reported a comparison between several color normalization methods which conceptually resembles Reinhard's statistical methodology reported in [64]. Based on the latter, the RGB channels of each frame in the dataset were transformed into the *Lab* color space, in order to modify its color characteristics following Equation 1.

$$I_{n}^{(i)} = (I_{s}^{(i)} - \mu_{s}^{(i)})\frac{\sigma_{t}^{(i)}}{\sigma_{s}^{(i)}} + \mu_{t}^{(i)}, i \in \{L, a, b\} \quad (1)$$

$I_{n}^{(i)}$, $I_{s}^{(i)}$ correspond to the normalized and the source RGB frame transform to the *Lab* color space. In addition, $\sigma_{s,t}^{(i)}$ and $\sigma_{s,t}^{(i)}$ correspond to the standard deviation and mean value computed over all the pixels of the source's and target's *Lab* channel *i* respectively. The target image corresponds to the mean image of all the dataset. After normalization, the frame is returned to the RGB color space, where each channel showed more separable distribution when evaluating its histogram.

3.3 Blue Ratio computation

H&E staining presents certain characteristics which allow the automatic detection of potential mitosis inside the High Power Field (HPF). In fact, the Hematoxylin stains the cell nuclei in a blue-purple shade, while Eosin stains cytoplasm and stroma in various shades from reddish to a pinkish color, and collagen in a pale pink shade. Due to its particular blue color shade observed in the RGB image, it was reported in [7, 36, 42] that the blue ratio image highlights nuclei as it enhances the blue color layer following the transformation in Equation 2.

$$I_{BR} = \frac{255 \times B}{(1 + R + G)(1 + R + G + B)} \quad (2)$$

R, G, and *B* are the red, green and blue channels respectively. After the generation of the blue ratio image, a segmentation and morphological opening operation are applied. The objective is to retain the potential mitosis present in the HPF, for further analysis. The threshold and the radius of the opening

operation were obtained empirically by validating them in the ICRP2012 dataset. Testing was also performed using the MITOS-ATYPIA2014 dataset. Using the potential candidates selected from the blue ratio image, 71×71-pixel patches were extracted from the original HPF. These patches were augmented afterward by rotation to increase the input data for training and testing the AlexNet. Regarding the U-Net, the blue ratio image was added as a fourth layer (i.e. R, G, B, BR) to the HPF. Then, patches of 128×128 pixels, centered at the centroid of a mitotic/non-mitotic cell, were extracted from each frame and its corresponding labeled mask was also generated.

3.4 Deep learning architectures

Convolutional Neural Networks (CNN) have proven to outperform classic machine learning techniques regarding image processing applications. In Chapter 2, it was reported that statistics from textural features and color features are frequently used along with different classic classifiers and CNN. In our study, the performance of the CNN proposed in [56] (currently known as AlexNet) was evaluated. The input layer was modified in order to allow the CNN to work with color-normalized patches of 71×71 pixels. The softmax layer was also modified in order to obtain a binary classification rather than a 1000-class classification as originally proposed by the authors. No handcrafted features were added to the computation and the training was performed from scratch using stochastic gradient descent, a batch size of 128 and 100 epochs. Finally, all patches generated from the BR image where divided into three datasets in the following proportions: 70% for training, 20% for validation, and 10% for testing.

A second deep learning approach (so far unreported in the literature for this application) using the U-net, was also implemented. This 23-layer network was trained from scratch for 100 epochs using stochastic gradient descent, unitary batch size, Adam's optimizer, and a binary cross-entropy loss function. As it was detailed in Table II, the patches used for the U-Net were unbalanced; therefore, we chose 898 samples (701 non-mitotic patches and 197 mitotic patches) for testing and the remaining 2884 non-mitotic patches were randomly selected to match the 1076 mitotic patches. Additionally, these 2152 selected patches were again randomly divided into 90% for training and 10% for validation.

Finally, to increase the performance in terms of computational time, both neural networks were trained in a parallel platform using two Nvidia Quadro M2000 GPUs. Furthermore, the CNN was implemented using DIGITS interface from Nvidia and the Caffe framework. Meanwhile, the U-Net was implemented using Keras and Tensorflow as a backend framework for deep learning.

3.5 Validation metrics

In order to measure the performance of the proposed methodologies, five metrics were selected. Table III relates the metrics with its corresponding dataset and neural network.

- **Accuracy**: $\dfrac{TP + TN}{TP + TN + FP + FN}$
- **Sensitivity**: $\dfrac{TP}{TP + FN}$

- **Specificity**: $\frac{TN}{TN + FP}$
- **F1-score**: $2\frac{Precision \times Recall}{Precision + Recall}$
- **Dice index**: also known as the Sorensen-Dice coefficient is defined as $\frac{|\hat{Y} \cap Y|}{|\hat{Y}| + |Y|}$, where $\hat{Y}$ are the predicted labeled masks, and Y are the ground truth masks.

Table III: Metrics used to evaluate the performance of the algorithm in the testing dataset.

Neural Network	**Evaluation metric**	**Dataset used**
AlexNet	Accuracy	train/test
	Sensitivity	test
	Specificity	test
	F1-score	test
U-Net	Accuracy	train/test
	Dice index	train/test

Chapter 4

Results and discussion

4.1 Detection of potential mitosis

A critical step in the CNN approach using the AlexNet is the detection of potential mitosis in the HPF frames. Figure 13 resumes the results for the color normalization algorithm and blue ratio image computation. The threshold to generate the binary image from the BR image and the radius of the structure used for the opening operation were found to be $T_h = 30$, and $R_d = 1$ respectively. These parameters allow achieving a 100% detection of true positive mitosis in the validation dataset (ICPR2012). Additionally, when tested in the MITOS-ATYPIA2014 dataset, the true positive rate was approximately 0.99, meaning only 1 mitosis was not detected from the whole dataset. Regarding the semantic segmentation approach, the RBG input was normalized without any further preprocessing steps.

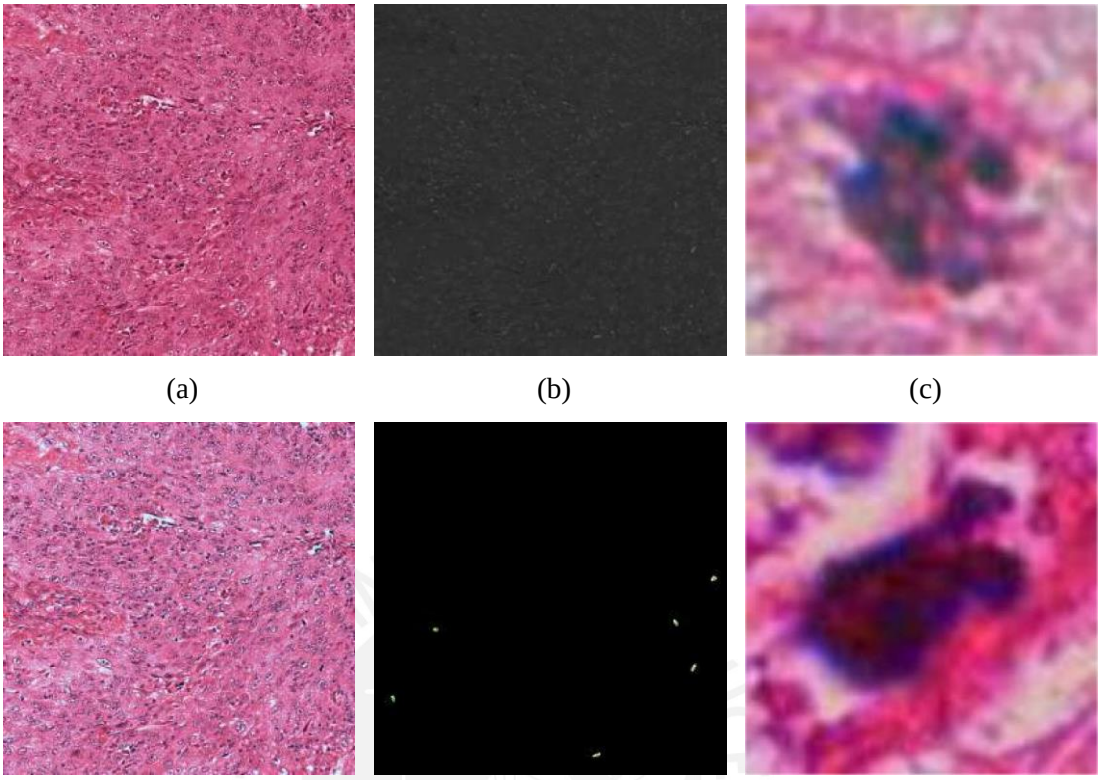
4.2 Convolutional Neural Network (AlexNet)

The network was trained for 100 epochs obtaining a 94.35% in accuracy using the validation-patch dataset. Regarding the loss value for the validation, it showed a descending behavior until epoch 40. At epoch 100, the loss value for the validation was only 0.1 higher than the one at epoch 40, which indicates that the CNN is still correctly modeling the validation dataset. Figure 14 resumes the results throughout the entire training and validation as well as the learning rate behavior.

The CNN was also evaluated using a testing-patch dataset which was never considered in training, nor in the validation process. Table IV shows the confusion matrix computed with the testing patch dataset. Additionally, Table V shows a resume of the metrics defined to evaluate and compare the performance of this CNN to other approaches present in the literature.

Table IV: AlexNet: Confusion matrix for the testing patch dataset.

	mitosis	**non-mitosis**	**accuracy/class**
mitosis	551	34	94.19%
non-mitosis	32	724	95.77%



(a) (b) (c)

(d) (e) (f)

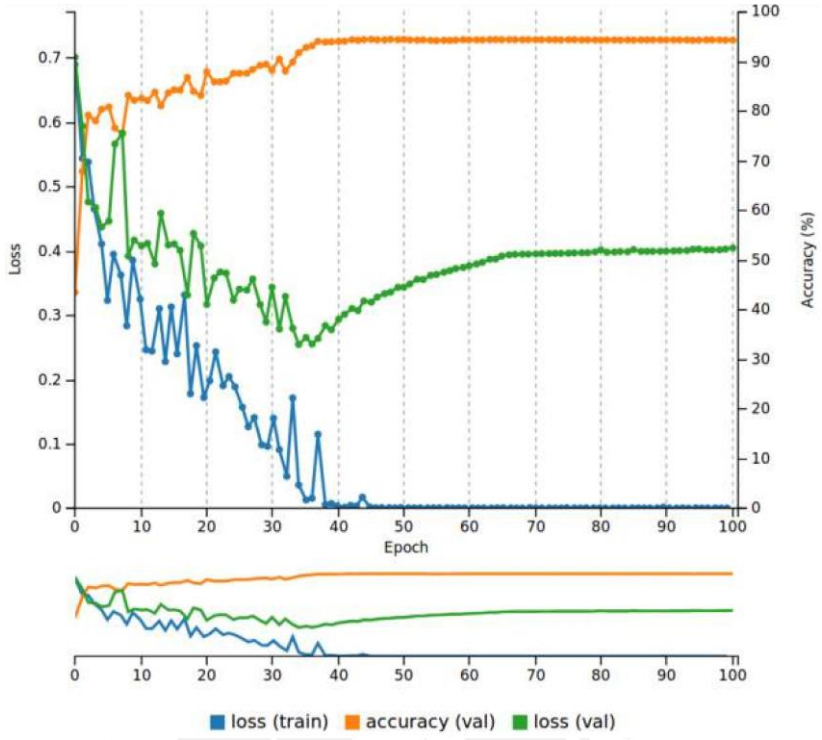
Figure 13: Results of the color normalization algorithm and blue ratio image generation for one frame: (a) Original HPF; (b) Blue ratio (BR) image; (c) Mitotic 71×71 patch generated with the thresholded version of the BR image; (d) Color normalized HPF; (e) Detection of mitosis centroids using the BR image to validate true positive rate; (f) Non-mitotic 71×71 patch generated with the thresholded version of the BR image.

Table V: AlexNet: Evaluation metrics in the testing patch dataset.

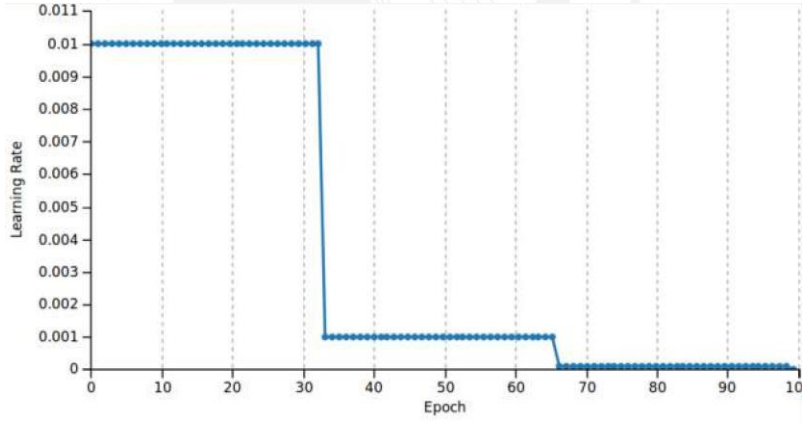
Accuracy	**Sensitivity**	**Specificity**	**F1-score**
95.08%	94.19%	95.77%	94.35%

4.3 Semantic segmentation (U-Net)

Table VI and Figure 15 resume the performance of the network after 100 epochs of training. The overall accuracy was 97.98% in the validation patch dataset, and 97.73% in the testing patch dataset. Regarding the convergence using the Adam optimizer, the loss value for both training and validation dataset showed a decreasing behavior. Early stopping of the network could be applied around epoch #24 due to the constant average loss value in the validation dataset. In addition, the Dice coefficient, obtained in both validation and testing datasets, indicates that nearly 60% of the predicted mask overlaps with the ground truth which translates into good network performance. In Figure 16 two RGB patches, ground truth masks, and predicted masks corresponding to mitosis and non-mitosis are shown. We found that almost every pixel corresponds to the ground truth classification and that borders are not well defined



(a)



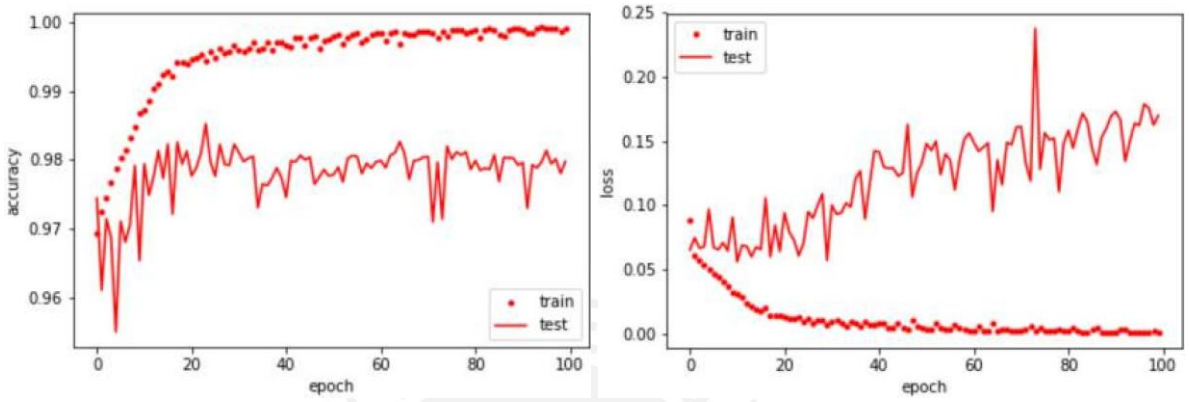
(b)

Figure 14: Results of the AlexNet training. (a) Training and validation loss values, and accuracy for the validation dataset (20% of the total patch dataset). (b) Learning rate behavior for 100 epochs.

in the predicted masks. Additionally, in certain cases, the classification is not 100% accurate and a low probability map (i.e predicted mask) is produced by the neural network as seen in Figure 17.

Table VI: U-Net: Evaluation metrics in training and testing.

Dataset	Train	Validation	Test
Dice index	0.9747	0.6117	0.5842
Accuracy	99.90%	97.98%	97.73%



(a) (b)

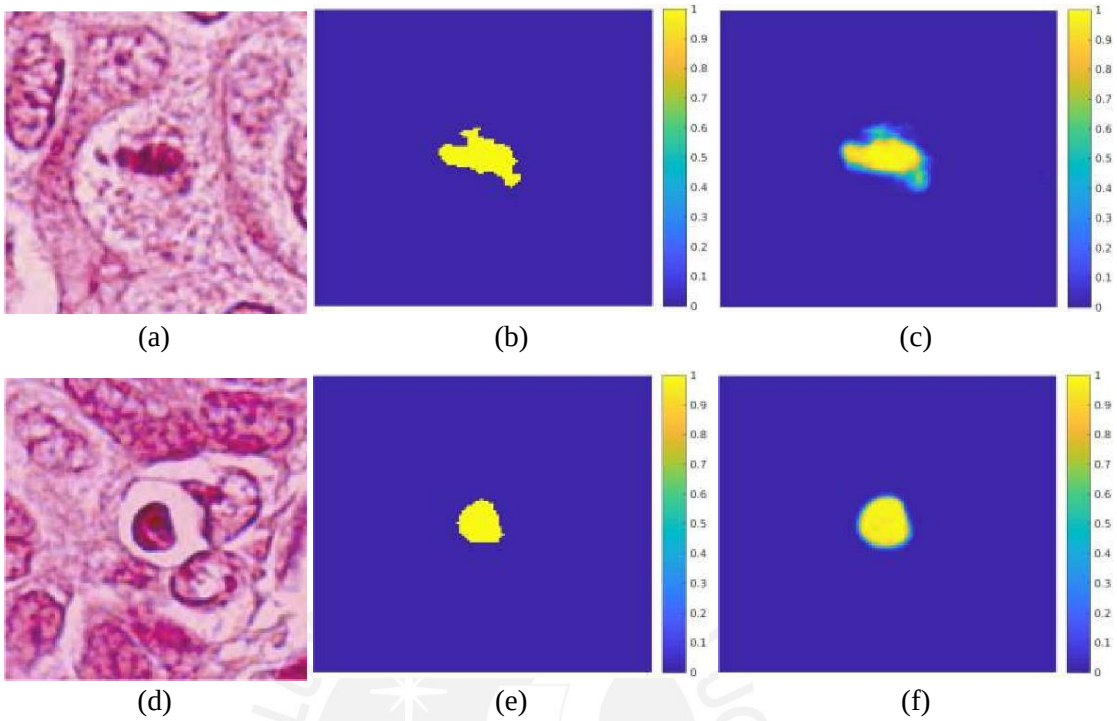
Figure 15: U-Net: Training and validation throughout 100 epochs. (a) Training and validation accuracy of each epoch. (b) Training and validation loss value of each epoch.

4.4 Discussion

Blue ratio has proven to be an efficient algorithm to detect potential mitosis with excellent accuracy (almost 0% false negative error). The analysis of the HPF containing the missed mitosis showed that the error presented when the WSI was scanned with the Hamamatsu scanner. This may suggest that the difference in resolution among scanners is likely to impact the quality of the algorithms used to preprocess the data. Additionally, commercial scanners perform a series of focusing actions during the digitalization of a tissue sample. These changes in the focus - due to fine mechanical slide scanners limitations - seem also likely to impact the quality of the whole slide image generated and, implicitly, of the analysis algorithms used for computation.

Although BR obtained high accuracy, it was computationally time-consuming due to the generation of false positive cases. To generate the nearly 13000 patches used in this work, approximately 2.5 hours were needed using a Core i7 processor computer with 24GB of RAM. To increase performance, both deep learning approaches were optimized to run in NVIDIA GPUs. The AlexNet, on a GeForce GTX980M, only took 20 minutes and 50 seconds to train, and 5 seconds to test nearly 1200 images. Meanwhile, the U-net, on two Quadro M2000, was trained in approximately 5 hours, and testing took 46 second per 128×128 pixels patch.

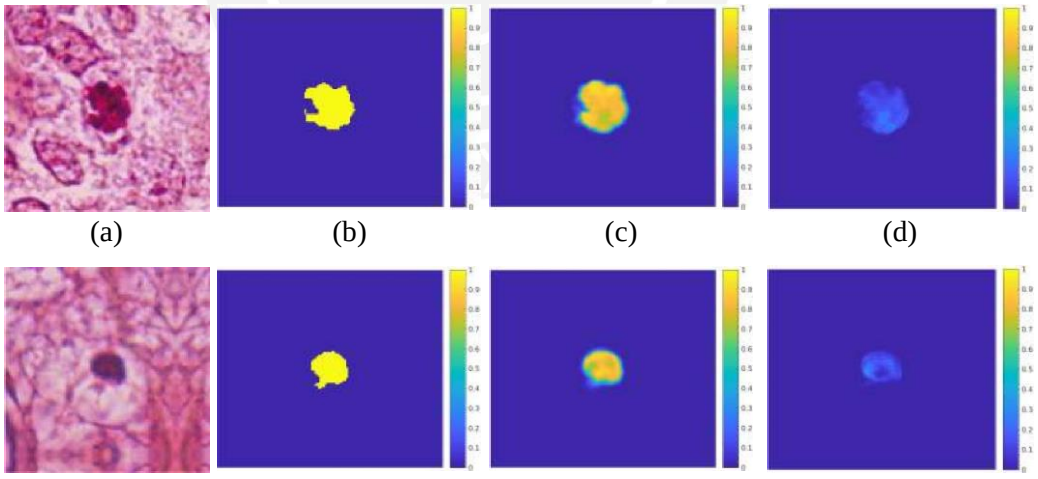
The second deep learning algorithm presented in this thesis, based on the U-net, is one approach closer to an end-to-end deep learning architecture. The complete elimination of handcrafted features (e.g. BR images) made the solution more robust to color and textural changes as observed in the slight increase of accuracy respect to AlexNet. Compared to the results in the literature, metrics obtained from both neural networks validate the fact that handcrafted features might introduce errors and subjectivity



(a) (b) (c)

(d) (e) (f)

Figure 16: Test results for the U-net. (a) HPF patch #37 with a mitosis. (b) GT mask of the mitosis. (c) Predicted mask for the mitosis class. (d) HPF patch #58 with a non-mitotic cell. (e) GT mask of the non-mitosis. (f) Predicted mask for the non-mitosis class.



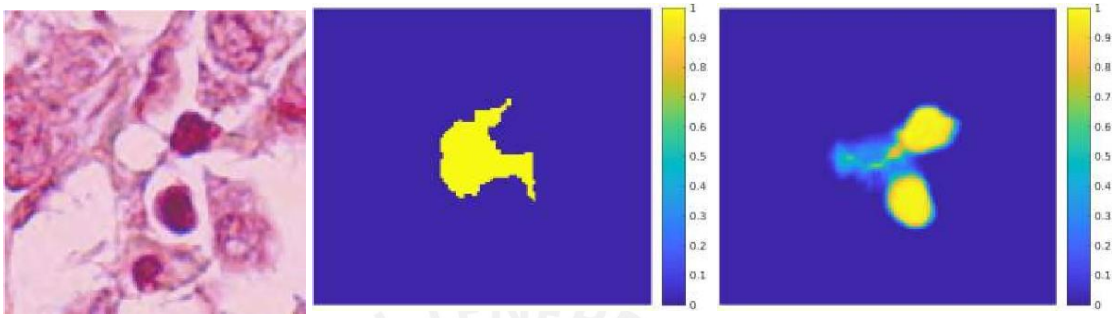
(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 17: U-net: Probability maps for both predicted classes in the testing-patch dataset (a) HPF patch #35 with a mitosis. (b) GT mask of the mitosis. (c) Predicted mask for the mitosis class. (d) Predicted mask for the non-mitosis class. (e) HPF patch #128 with a non-mitotic cell. (f) GT mask of the non-mitosis. (g) Predicted mask for the non-mitosis class. (h) Predicted mask for the mitosis class.

into the classification. To address this issue, we generated ground truth masks for the testing dataset using a Simple Linear Iterative Clustering (SLIC) algorithm to create super-pixels in an HPF frame

which correspond to mitotic or non-mitotic clusters of pixels. Figure 18 shows an example of a non-mitotic cell. The mask predicted using the SLIC algorithm involves several pink pixels from the patch which does not correspond the cell; meanwhile, the mask generated by the U-Net clearly identifies the region with predominant blue coloration representing the presence of a cell, in this case non-mitotic. Throughout the entire testing dataset, we found similar cases which validate the hypothesis that hand-crafted features introduce errors in the detection and classification of the behavior of cellular structures.



(a) (b) (c)

Figure 18: Evaluation of SLIC for mitosis selection. (a) HPF patch #75 with a non-mitosis. (b) Mask of the mitosis generated using SLIC. (c) Predicted mask for the mitosis class using U-Net.

The results in the present work outperform all classical machine learning approaches existing at the moment in the literature. However, for the U-net, further analysis is needed in order to improve border detection (an increase of Dice index) and increase the sensitivity up to 100% which is more beneficial in biomedical applications. This study presented two methods to analyze and classify a frame containing 10 HPF, usually observed by pathologists. Due to the structure of the networks and its performance, these algorithms can also be applied directly to the WSI. The direct application to WSI - which we will study in our future research - will improve the diagnosis of breast cancer as pathologists only evaluate small regions of the sample tissues.

Chapter 5

Conclusions and perspectives

In this thesis work, we have shown how two different deep learning approaches (i.e. AlexNet and U-Net) performs when dealing with the classification and detection of mitosis. The results suggest that the performance of classical image-processing methodologies and deep learning approaches combined with hand-crafted features can be noticeably improved (nearly 7% improvement in accuracy respect to the last work in [42]). Regarding the AlexNet we found that it is possible to adapt the neural network to work with small patches corresponding to single mitosis. Although this method has the disadvantage of using a pre-processing step (i.e. BR computation to detect potential mitosis), it outperforms the proposed algorithms in the literature.

On the other hand, due to the semantic advantage of the U-Net, we can directly detect (finding the *where*) and classify (finding the *what*) cellular structures in an end-to-end framework without the need of pre/post-processing steps. This suggests that the U-Net is more suitable and robust to process WSI allowing a stable and effective transition towards the full WSI analysis. Although the U-Net does not depend on the size of the input image, WSI analysis may turn out a greater challenge due to its size, different cellular patterns and structures, and artifacts introduce by scanners, staining process or other human factors. Therefore, future research to validate the use of deep learning architectures such as the U-Net in WSI is needed.

We have seen that the U-Net is an excellent tool to analyze histopathological images (validated in HPF so far). Therefore, we can extend the solution and apply it to different nuclei detection such as neoplastic nuclei, fibroblast, lymphocyte, adipose tissue (adipocytes), stroma, or blood vessels. These nuclei allow pathologists to have a better understanding of the tumor microenvironment which benefits the patient in terms of a targeted treatment to specific characteristics observed in the tumor. Additionally, we can also use the U-Net to analyze the spatial distribution of the tumor microenvironment to further understand tumor heterogeneity which might, in turn, provide some insight on why certain types of cancers are more resistant than others. Tumor heterogeneity can manifest as intra-tumor (meaning clones of cells responding differently to the same treatment), or inter-tumor (meaning, same kind of tumor behaving differently in different patients). In both cases, the semantic property of the U-Net allows to have a pixel-wise understanding of the tissue and therefore give the pathologists and oncologists useful information to diagnose and treat patients.

Finally, in this thesis, we used images of breast cancer tissues. However, this can be applied to all types of tissue as the deep learning architectures can be easily adapted to other inputs.

References

[1] Ahmedin Jemal, Melissa M. Center, Carol DeSantis, and Elizabeth M. Ward, "Global patterns of cancer incidence and mortality rates and trends," *Cancer Epidemiology Biomarkers and Prevention*, vol. 19, no. 8, pp. 1893–1907, 2010.

[2] Carol E. DeSantis, Freddie Bray, Jacques Ferlay, Joannie Lortet-Tieulent, Benjamin O. Anderson, and Ahmedin Jemal, "International Variation in Female Breast Cancer Incidence and Mortality Rates," *Cancer Epidemiology Biomarkers and Prevention*, vol. 24, no. 10, pp. 1495–1506, 2015.

[3] World Health Organization, "Cancer," http://www.who.int/news-room/facts-in-pictures/detail/cancer, 2018, Accessed: 2018-05-07.

[4] Breastcancer.org U.S., "Breast cancer statistics," http://www.breastcancer.org/symptoms/understand_bc/statistics, 2018, Accessed: 2018-05-07.

[5] R. Glynn Owens and Jennifer J. Ashcroft, "Breast Cancer Screening," *Journal of Psychosocial Oncology*, vol. 4, no. 4, pp. 15–26, 1987.

[6] Hans Peter Sinn and Hans Kreipe, "A brief overview of the WHO classification of breast tumors, 4th edition, focusing on issues and updates from the 3rd edition," *Breast Care*, vol. 8, no. 2, pp. 149–154, 2013.

[7] Bassem Ben Cheikh, *Morphologie mathématique sur les graphes pour la caractérisation de l' organisation spatiale des structures histologiques dans les images haut-contenu : application au microenvironnement tumoral dans le cancer du sein*, Ph.D. thesis, Universite´ Pierre et Marie Curie, 2017.

[8] Christie R. Eheman, Kate M. Shaw, Aliza Blythe Ryerson, Jacqueline W. Miller, Umed A. Ajani, and Mary C. White, "The changing incidence of in situ and invasive ductal and lobular breast carcinomas: United States, 1999-2004," *Cancer Epidemiology Biomarkers and Prevention*, vol. 18, no. 6, pp. 1763–1769, 2009.

[9] JR Jr. Wright, "The development of the frozen section technique, the evolution of surgical biopsy, and the origins of surgical pathology," *Bulletin of the History of Medicine*, vol. 59, pp. 295–326., 1985.

[10] H J G Bloom and W W Richardson, "Histological grading and prognosis of breast cancer," *British Journal of Cancer*, vol. 11, no. 3, pp. 359–77, 1957.

[11] Pedram Argani and Ashley Cimino-Mathews, "Grade of Breast Cancer," http://pathology.jhu.edu/breast/grade.php, 2015, Accessed: 2018-05-09.

[12] D. J. Meuten, F. M. Moore, and J. W. George, "Mitotic Count and the Field of View Area: Time to Standardize," *Veterinary Pathology*, vol. 53, no. 1, pp. 7–9, 2016.

[13] Liron Pantanowitz, Navid Farahani, and Anil Parwani, "Whole slide imaging in pathology: advantages, limitations, and emerging perspectives," *Pathology and Laboratory Medicine International*, vol. 7, pp. 23—-33, 2015.

References

[14] DICOM Standards Committee, "Digital Imaging and Communications in Medicine (DICOM): Supplement 145: Whole Slide Microscopic Image IOD and SOP Classes," Tech. Rep., 2010.

[15] Henrik Helin, Teemu Tolonen, Onni Ylinen, Petteri Tolonen, Juha N äpänkangas, and Jorma Isola, "Optimized JPEG 2000 Compression for Efficient Storage of Histopathological Whole-Slide Images," *Journal of pathology informatics*, vol. 9, pp. 20, 2018.

[16] Jun Kong, Lee A.D. Cooper, Fusheng Wang, Jingjing Gao, George Teodoro, Lisa Scarpace, Tom Mikkelsen, Matthew J. Schniederjan, Carlos S. Moreno, Joel H. Saltz, and Daniel J. Brat, "Machine-based morphologic analysis of glioblastoma using whole-slide pathology images uncovers clinically relevant molecular correlates," *PLoS ONE*, vol. 8, no. 11, pp. e81049, 2013.

[17] Hao Lu, Thomas G Papathomas, David van Zessen, Ivo Palli, Ronald R de Krijger, Peter J van der Spek, Winand N.M. Dinjens, and Andrew P Stubbs, "Automated Selection of Hotspots (ASH): enhanced automated segmentation and adaptive step finding for Ki67 hotspot detection in adrenal cortical cancer," *Diagnostic pathology*, vol. 9, pp. 216, nov 2014.

[18] Chien Ho, Fang-Cheng Yeh, AnilV Parwani, and Liron Pantanowitz, "Automated grading of renal cell carcinoma using whole slide imaging," *Journal of Pathology Informatics*, vol. 5, no. 1, pp. 23, 2014.

[19] ESMO, "FDA Allows Marketing of First Whole Slide Imaging System for Digital Pathology," *ESMO Oncology News*, pp. 7–9, 2017.

[20] Angshuman Paul and Dipti Prasad Mukherjee, "Mitosis Detection for Invasive Breast Cancer Grading in Histopathological Images," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 4041–4054, 2015.

[21] Clare O'Connor, "Cell division: Stages of mitosis," *Nature Education 1*, vol. 1, pp. 188, 2008.

[22] Po-Hsuan Chen, Krishna Gadepalli, Robert MacDonald, Yun Liu, Kunal Nagpal, Timo Kohlberger, Jason D. Hipp, and Martin C. Stumpe, "An Augmented Reality Microscope for Real-time Automated Detection of Cancer," Tech. Rep., Google AI Healthcare, Google Inc, Mountain View, CA, USA, 2018.

[23] Xiaodong Yang, Houqiang Li, and Xiaobo Zhou, "Nuclei segmentation using marker-controlled watershed, tracking using mean-shift, and Kalman filter in time-lapse microscopy," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 53, no. 11, pp. 2405–2414, 2006.

[24] Alexander Nedzved, Sergey Ablameyko, and Ioannis Pitas, "Morphological segmentation of histology cell images," in *Proceedings 15th International Conference on Pattern Recognition*, 2000, vol. 1, pp. 500–503.

[25] Gang Lin, Monica K. Chawla, Kathy Olson, Carol A. Barnes, John F. Guzowski, Christopher Bjornsson, William Shain, and Badrinath Roysam, "A multi-model approach to simultaneous segmentation and classification of heterogeneous populations of cell nuclei in 3D confocal microscope images," *Cytometry*, vol. 71, no. 9, pp. 724–736, 2007.

[26] Ananda S Chowdhury, Rohit Chatterjee, Mayukh Ghosh, and Nilanjan Ray, "Cell Tracking in Video Microscopy Using Bipartite Graph Matching," in *International Conference on Pattern Recognition*, 2010, pp. 2456–2459.

[27] Mojtaba Seyedhosseini and Tolga Tasdizen, "Multi-class multi-scale series contextual model for image segmentation," *IEEE Transactions on Image Processing*, vol. 22, no. 11, pp. 4486–4496, 2013.

[28] Yousef Al-Kofahi, Wiem Lassoued, William Lee, and Badrinath Roysam, "Improved automatic detection and segmentation of cell nuclei in histopathology images," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 4, pp. 841–852, 2010.

[29] Dipti Prasad Mukherjee, Nilanjan Ray, and Scott T Acton, "Level set analysis for leukocyte detection and tracking.," *IEEE transactions on image processing: a publication of the IEEE Signal Processing Society*, vol. 13, no. 4, pp. 562–572, 2004.

[30] Sumit K Nath, Kannappan Palaniappan, and Filiz Bunyak, "Cell segmentation using coupled level sets and graph-vertex coloring," *Medical Image Computing and Computer Assisted Intervention*, vol. 9, no. Pt 1, pp. 101–108, 2006.

[31] Oleh Dzyubachyk, Wiro Niessen, and Erik Meijering, "Advanced level-set based multiple-cell segmentation and tracking in time-lapse fluorescence microscopy images," in *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Proceedings, ISBI*, 2008, pp. 185–188.

[32] KM Lee, WN Street, and Kyoung-Mi Lee, "A fast and robust approach for automated segmentation of breast cancer nuclei," in *In Proceedings of the IASTED International Conference on Computer Graphics and Imaging*. 1999, pp. 42—-47, ACTA Press.

[33] Olcay Sertel, Umit V. Catalyurek, Hiroyuki Shimada, and Metin N. Gurcan, "Computer-aided prognosis of neuroblastoma: Detection of mitosis and karyorrhexis cells in digitized histological images," in *Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society: Engineering the Future of Biomedicine, EMBC 2009*, 2009, pp. 1433–1436.

[34] Vincent Roullier, Olivier L ézoray, Vinh Thong Ta, and Abderrahim Elmoataz, "Multi-resolution graph-based analysis of histopathological whole slide images: Application to mitotic cell extraction and visualization," *Computerized Medical Imaging and Graphics*, vol. 35, no. 7-8, pp. 603–615, 2011.

[35] Barbara Weyn, Gert Van De Wouwer, Andr éVan Daele, Paul Scheunders, Dirk Van Dyck, Eric Van Marck, and Willem Jacob, "Automated breast tumor diagnosis and grading based on wavelet chromatin texture description," *Cytometry*, vol. 33, no. 1, pp. 32–40, 1998.

[36] Humayun Irshad, Sepehr Jalali, Ludovic Roux, Daniel Racoceanu, GillesLe Naour, LimJoo Hwee, and Frédérique Capron, "Automated mitosis detection using texture, SIFT features and HMAX biologically inspired approach," *Journal of Pathology Informatics*, vol. 4, no. 2, pp. 12, 2013.

[37] Charles Y. Tao, Jonathan Hoyt, and Yan Feng, "A support vector machine classifier for recognizing mitotic subphases using high-content screening data," *Journal of Biomolecular Screening*, vol. 12, no. 4, pp. 490–496, 2007.

[38]F. Boray Tek, "Mitosis detection using generic features and an ensemble of cascade adaboosts," *Journal of Pathology Informatics*, vol. 4, no. 1, pp. 12, 2013.

[39] Ramin Nateghi, Habibollah Danyali, Mohammad Sadeghhelfroush, and Fattaneh Pourak Pour, "Automatic detection of mitosis cell in breast cancer histopathology images using genetic algorithm," in *2014 21st Iranian Conference on Biomedical Engineering, ICBME 2014*, 2011, pp. 1–6.

[40] Jean-Romain Dalle, Wee Kheng Leow, Daniel Racoceanu, Adina Eunice Tutac, and Thomas C. Putti, "Automatic breast cancer grading of histopathological images," in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2008, pp. 3052–3055.

[41] Dan C. Cireşan, Alessandro Giusti, Luca M. Gambardella, and J ürgen Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks," in *Lecture Notes in Computer Science*. 2013, vol. 8150 LNCS, pp. 411–418, Springer.

[42] Monjoy Saha, Chandan Chakraborty, and Daniel Racoceanu, "Efficient deep learning model for mitosis detection using breast histopathology images," *Computerized Medical Imaging and Graphics*, vol. 64, pp. 29–40, 2018.

[43] Teresa Araujo, Guilherme Aresta, Eduardo Castro, Jos éRouco, Paulo Aguiar, Catarina Eloy, António Polonia, and Aurélio Campilho, "Classification of breast cancer histology images using Convolutional Neural Networks," *PLoS One*, vol. 12, no. 6, pp. 1 – 14, 2017.

[44] Shadi Albarqouni, Christoph Baur, Felix Achilles, Vasileios Belagiannis, Stefanie Demirci, and Nassir Navab, "AggNet: Deep Learning From Crowds for Mitosis Detection in Breast Cancer Histology Images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1313–1321, 2016.

[45] Dan C. Ciresan, Alessandro Giusti, Luca M. Gambardella, and J ürgen Schmidhuber, "Deep Neural Networks Segment Neuronal Membranes in Electron Microscopy Images," in *2012 Conference on Neural Information Processing Systems*, 2012, pp. 2852—-2860.

[46] Ludovic Roux, Daniel Racoceanu, Nicolas Lom énie, Maria Kulikova, Humayun Irshad, Jacques Klossa, Frédérique Capron, Catherine Genestie, GillesLe Naour, and MetinN Gurcan, "Mitosis detection in breast cancer histological images An ICPR 2012 contest," *Journal of Pathology Informatics*, vol. 4, no. 1, pp. 8, 2013.

[47] Mitko Veta, Paul J. van Diest, Stefan M. Willems, Haibo Wang, Anant Madabhushi, Angel Cruz-Roa, Fabio Gonzalez, Anders B.L. Larsen, Jacob S. Vestergaard, Anders B. Dahl, Dan C. Cires¸an, Jürgen Schmidhuber, Alessandro Giusti, Luca M. Gambardella, F. Boray Tek, Thomas Walter, Ching Wei Wang, Satoshi Kondo, Bogdan J. Matuszewski, Frederic Precioso, Violet Snell, Josef Kittler, Teofilo E. de Campos, Adnan M. Khan, Nasir M. Rajpoot, Evdokia Arkoumani, Miangela M. Lacle, Max A. Viergever, and Josien P.W. Pluim, "Assessment of algorithms for mitosis detection in breast cancer histopathology images," *Medical Image Analysis*, vol. 20, no. 1, pp. 237–248, feb 2015.

[48] "MITOS - ATYPIA 14," `https://mitos-atypia-14.grand-challenge.org/`, 2014, Accessed: 2018-06-27.

[49] Teresa Ara újo, Guilherme Aresta, Catarina Eloy, António Polónia, and Paulo Aguiar, "Grand Challenge on Breast Cancer Histology Images," `https://iciar2018-challenge.grand-challenge.org/`,2018, Accessed: 2018-08-12.

References

[50] Neeraj Kumar, Verma Ruchika, Deepak Anand, and Amit Sethi, "MoNuSeg," 2018.

[51] Computational Precision Medicine Committee, "Digital Pathology: Segmentation of Nuclei in Images," http://miccai.cloudapp.net/competitions/83, 2018, Accessed: 2018-08-12.

[52] Computational Precision Medicine Committee, "Combined Radiology and Pathology Classification," http://miccai.cloudapp.net/competitions/82, 2018, Accessed: 2018-08-12.

[53] Y. Le Cun, O Matan, B Boser, J.S. Denker, D Henderson, R.E. Howard, W Hubbard, L.D. Jacket, and H.S. Baird, "Handwritten zip code recognition with multilayer networks," in *[1990] Proceedings. 10th International Conference on Pattern Recognition*, 1990, pp. 35–40.

[54] Yann LeCun, L éon Bottou, Yoshua Bengio, and Patrick Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2323, 1998.

[55] Ian Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep Learning*, MIT Press, 2016.

[56] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, USA, 2012, NIPS'12, pp. 1097–1105, Curran Associates Inc.

[57] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2015.

[58] Evan Shelhamer, Jonathan Long, and Trevor Darrell, "Fully Convolutional Networks for Semantic Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 640–651, apr 2017.

[59] Liang Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, apr 2018.

[60] Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross Girshick, "Mask R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, vol. 2017-Octob, pp. 2980–2988.

[61] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *International Conference on Medical Image Computing and Computer Assisted Intervention*, 2015, pp. 234–241.

[62] "Mitosis Detection in Breast Cancer Histological Images," http://ludo17.free.fr/mitos_2012/index.html, 2012, Accessed: 2018-06-27.

[63] Ryan A. Hoffman, Sonal Kothari, and May D. Wang, "Comparison of normalization algorithms for cross-batch color segmentation of histopathological images," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2014*, 2014, pp. 194–197.

References

[64] Erik Reinhard, Michael Ashikhmin, Bruce Gooch, and Peter Shirley, "Color transfer between images," *IEEE Computer Graphics and Applications*, vol. 21, no. 5, pp. 34–41, 2001.