Graduate Program in Applied Computing

Applied Computing

Academic Master

Felipe André Zeiser

DeepBatch: A Hybrid Deep Learning Model for an Interpretable Diagnosis of Breast Cancer in Whole-Slide Images

São Leopoldo, 2020

Felipe André Zeiser

DEEPBATCH: A HYBRID DEEP LEARNING MODEL FOR AN INTERPRETABLE DIAGNOSIS OF BREAST CANCER IN WHOLE-SLIDE IMAGES

Master Thesis presented as a partial requirement to obtain the Master's degree from the Applied Computing Graduate Program of the Universidade do Vale do Rio dos Sinos — UNISINOS

Advisor:
Prof. Cristiano André da Costa, PhD

Co-advisor:
Prof. Gabriel de Oliveira Ramos, PhD

São Leopoldo
2020

Z47d Zeiser, Felipe André.
DeepBatch: a hybrid deep learning model for an interpretable diagnosis of breast cancer in whole-slide images / Felipe André Zeiser. – 2020.
135 f. : il. ; 30 cm.

Dissertação (mestrado) – Universidade do Vale do Rio dos Sinos, Programa de Pós-Graduação em Computação Aplicada, 2020.
"Advisor: Prof. Cristiano André da Costa
Co-advisor: Prof. Gabriel de Oliveira Ramos."

1. Breast cancer. 2. Histopathological images. 3. Deep learning. 4. Interpretable diagnosis. 5. Convolutional neural network. I. Título.

CDU 004

Dados Internacionais de Catalogação na Publicação (CIP)
(Bibliotecária: Amanda Schuster – CRB 10/2517)

(Esta folha serve somente para guardar o lugar da verdadeira folha de aprovação, que é obtida após a defesa do trabalho. Este item é obrigatório, exceto no caso de TCCs.)

To my family.

"I think, therefore I am."
— RENE DESCARTES

ACKNOWLEDGEMENTS

This work was only possible due to the collaboration of several people in different aspects of the research. I would like to thank my advisor Prof. Cristiano André da Costa, who accepted me under his supervision as a master's student, sharing his knowledge and time in conducting the work. I would also like to thank my co-advisor, Prof. Gabriel de Oliveira Ramos, for helping me in this work with feedbacks always very useful. I thank the pathologist and Prof. Adriana Vial Roehe for the advice in the development of my research. I would like to thank Cristiano, Gabriel, and Adriana for their patience during this period with several reviews, meetings, and for being a reference of excellence in research for me. Thank you very much.

Special thanks to my family for their emotional and financial support throughout my life. To my father, Darci, and my mother, Melita, for being my inspiration for life. To my brother for the friendship and the moments of fun. I would like to thank my cousin Elixandro Weber and his family for their friendship, support, and company during my stay in Ivoti - RS.

I would also like to thank Prof. Tiago Zonta and Prof. Nuno Cavalheiro Marques for his support during my graduation. You made a fundamental role in building my knowledge, and that contributed to the opportunity to develop this work. I would like to thank the project colleagues who helped me in the research development, in particular, Henrique C. Bohn and Ismael Santos. I also want to thank the excellent professors of Unisinos PPGCA (Postgraduate Program in Applied Computing). Likewise, I would like to thank Unisinos' secretaries and staff for their help and patience during the master's degree.

I would also like to express my thanks to my laboratory colleagues, Fabiano, Guilherme, João, Luis, and Régis, for the moments of fun, coffee, and suggestions for work and life. Without you, that laboratory is not the same. I would like to thank all colleagues at uHospital for their collaboration with the project. I also want to thank all my friends who somehow contributed to the work and my life, thank you very much.

Finally, this work would not be possible without the financial support of several funding agencies. I would like to thank CNPq and CAPES for my master's scholarship. FAPESC for its recognition in the Prêmio Stemmer de Inovação Catarinense. NVIDIA for donating a GPU. UNOESC for providing the infrastructure to carry out the research experiments.

ABSTRACT

CONTEXT: Advances in studies at the molecular and genetic sequencing have brought significant progress in understanding the behavior, treatment, and prognosis of breast cancer. However, the diagnosis of breast cancer in the early stages is still essential for successful treatment. Currently, the gold standard for breast cancer diagnosis, treatment, and management is a histological analysis of a suspected section. Histopathology consists of analyzing the characteristics of the lesions through sections of tissue stained with Hematoxylin and Eosin. However, pathologists are currently subjected to high workloads, mainly due to the fundamental role played by histological analysis in the treatment of the patient. In this context, applications that can reduce the time of histological analysis, provide a second opinion, or even point out suspicious locations as a screening tool can help the pathologist. **OBJECTIVE:** We envision two main challenges: the first, how to identify cancerous regions in Whole Slide Imaging (WSI) using Deep Learning (DL) with an accuracy comparable to the pathologist's annotations, considered the gold standard in the literature. And second, how a DL-based model can provide an interpretable diagnosis. The scientific contribution consists in proposing a model based on Convolutional Neural Networks (CNN) to provide a refined and multi-class segmentation of WSI in breast cancer. **METHODOLOGY:** The methodology consists of proposing and developing a model called DeepBatch. The DeepBatch is divided into four modules: Preprocessing, ROI Detection, ROI Sampling, and Cell Segmentation. These modules are organized to decode the information learned using CNNs in interpretable predictions for pathologists. The Preprocessing module is responsible for removing background and noise from WSI. At ROI Detection, we use the U-Net convolutional architecture to identify suspicious regions in low magnification WSI. Suspected areas identified are mapped from low magnifications by ROI Sampling to $40\times$ magnifications. Cell Segmentation then segments high-magnification areas. Segmentation is performed using a ResNet50/U-Net. To validate the DeepBatch, we use datasets from different sources that can be used together or separately in each module, depending on the module's objective. **RESULTS:** The evaluations performed demonstrate the feasibility of the model. We assessed the impact of four-color spaces (RGB, HSV, YCrCb, and LAB) for multi-class segmentation of breast cancer WSI. We used 205 WSI of breast cancer for training, validation, and testing. For the detection of suspicious regions by ROI Detection we obtained a IoU of 93.43%, accuracy of 91.27%, sensitivity of 90.77%, specificity of 94.03%, F1-Score of 84.17%, and an AUC of 0.93. For the refined segmentation of WSI by the Cell Segmentation module we obtained a IoU of 88.23%, accuracy of 96.10%, sensitivity of 71.83%, specificity of 96.19%, F1 -Score of 82.94%, and an AUC of 0.86. **CONCLUSION:** As a contribution, DeepBatch provides refined segmentation of breast cancer WSIs using a cascade of CNNs. This segmentation helps the interpretation of the diagnosis by the pathologist, accurately presenting the regions considered during the inference of WSI. The results indicate the possibility of using the model as a second reading system.

Keywords: Breast Cancer. Histopathological Images. Deep Learning. Interpretable Diagnosis. Convolutional Neural Network. Whole-Slide Image. Image Segmentation.

RESUMO

CONTEXTO: Os avanços em estudos à nível molecular e de sequenciamento genético trouxeram significativos progressos no compreendimento do comportamento, tratamento e prognóstico do câncer de mama. No entanto, o diagnóstico do câncer de mama em estágios iniciais ainda é essencial para o sucesso no tratamento. Atualmente, o padrão ouro para o diagnóstico, tratamento e gerenciamento do câncer de mama é a análise histológica de uma seção suspeita. A histopatologia consiste na análise das características das lesões através de cortes de tecido corados com Hematoxilina e Eosina. Contudo, atualmente os patologistas são submetidos a altas cargas de trabalho, muito devido ao papel fundamental desempenhado pela análise histológica no tratamento do paciente. Neste contexto, aplicações que possam reduzir o tempo da análise histológica, fornecer uma segunda opinião, ou ainda apontar locais suspeitos como ferramenta de triagem podem auxiliar o patologista. **OBJETIVO:** Desse modo, vislumbramos dois desafios principais: o primeiro, como identificar regiões cancerígenas em *Whole Slide Imaging* (WSI) utilizando *Deep Learning* (DL) com uma precisão comparável as anotações do patologista, consideradas como padrão ouro na literatura. E, segundo, como um modelo baseado em DL pode fornecer um diagnóstico interpretável. A contribuição científica consiste em propor um modelo baseado em *Convolutional Neural Networks* (CNN) para fornecer uma segmentação refinada e multi-classe de WSI do câncer de mama. **METODOLOGIA:** A metodologia consiste em propor e desenvolver um modelo chamado de DeepBatch. O DeepBatch é dividido em quatro módulos: Preprocessing, ROI Detection, ROI Sampling e Cell Segmentation. Estes módulos são organizados de maneira a decodificar as informações aprendidas utilizando CNNs em previsões interpretáveis para os patologistas. O módulo de *Preprocessing* é responsável pela remoção do fundo e de ruído das WSI. No *ROI Detection* utilizamos a arquitetura convolucional U-Net para identificar regiões suspeitas em WSI de baixa ampliação. As regiões suspeitas identificadas são mapeadas de baixas ampliações pelo *ROI Sampling* para ampliações de $40\times$. As regiões em alta ampliação são então segmentadas pela *Cell Segmentation*. A segmentação é realizada utilizando uma ResNet50/U-Net. Para a validação do DeepBatch, utilizamos datasets de diversas fontes que podem ser utilizados em conjunto ou separadamente em cada módulo, conforme o objetivo do módulo. **RESULTADOS:** As avaliações realizadas demonstram a viabilidade do modelo. Avaliamos o impacto de quatro espaços de cores (RGB, HSV, YCrCb e LAB) para a segmentação multi-classe de WSI do câncer de mama. Utilizamos para o treinamento, validação e teste 205 WSI do câncer de mama. Para a detecção de regiões suspeitas pelo *ROI Detection* obtivemos uma IoU de 93.43%, acurácia de 91.27%, sensibilidade de 90.77%, especificidade de 94.03%, F1-Score de 84.17% e uma AUC de 0.93. Para a segmentação refinada das WSI pelo módulo de *Cell Segmentation* obtivemos uma IoU de 88.23%, acurácia de 96.10%, sensibilidade de 71.83%, especificidade de 96.19%, F1-Score de 82.94% e uma AUC de 0.86. **CONCLUSÃO:** Como contribuição, o DeepBatch fornece uma segmentação refinada das WSI do câncer de mama utilizando uma cascata de CNNs. Esta segmentação facilita a interpretação do diagnóstico pelo patologista, fornecendo com precisão as regiões consideradas durante a inferência da WSI. Os resultados indicam a possibilidade de utilização do modelo como um sistema de segunda leitura.

Palavras-chave: Câncer de Mama. Imagem Histopatológica. Aprendizado Profundo. Diagnóstico Interpretável. Redes Neurais Convolucionais. Whole-Slide Image. Segmentação de Imagem.

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

2D	Two-Dimensional
3D	Three-Dimensional
AI	Artificial Intelligence
ANN	Artificial Neural Network
AUC	Area Under the ROC curve
BACH	The Grand Challenge on Breast Cancer Histology Images
CAD	Computer-Aided Detection
CADx	Computer-Aided Diagnosis
CNN	Convolutional Neural Networks
DIP	Digital Image Processing
DL	Deep Learning
DNA	Deoxyribonucleic Acid
FN	False Negative
FP	False Positive
HSV	Hue, Saturation e Value
H&E	Hematoxylin and Eosin
INCA	Instituto Nacional do Câncer
KNN	k-Nearest Neighbors
LDA	Linear Discriminant Analysis
LSTM	Long Short-Term Memory
MIL	Multiple Instance Learning
ML	Machine Learning
MLP	Multilayer Perceptron
PCA	Principal Component Analysis
RCNN	Recurrent Convolutional Neural Network
RF	Random Forest
RGB	Red, Green e Blue
RNA	Ribonucleic Acid
ROC	Receiver Operating Characteristic
ROI	Region of Interest
SLR	Systematic Literature Review
SOM	Self-Organizing Map
SSA	Stacked Sparse Autoencoder

SVD	Singular Value Decomposition
SVM	Support Vector Machines
TCGA	The Cancer Genome Atlas Program
TN	True Negative
TP	True Positive
TUPAC16	Tumor Proliferation Assessment Challenge 16
WHO	World Health Organization
WSI	Whole Slide Imaging

CONTENTS

1 INTRODUCTION

The last few decades have been of significant advances in the health area, especially in molecular studies and Deoxyribonucleic Acid (DNA) sequencing, allowing a deeper understanding of the causes of cancer initiation and progression. However, cancer prevention and treatment remains a worldwide public health problem. Estimates from World Health Organization (2019a) (WHO) show this problem, pointing to 9.6 million deaths due to cancer in 2018. Also, according to the organization, 20 to 50% of these deaths could be avoided with early detection and adequate treatment.

In the female population, breast cancer is the one with the highest incidence and mortality rates in the world (BRAY et al., 2018; World Health Organization, 2019a; DEBELEE et al., 2020). In Brazil, breast cancer incidence rates are even more worrying when compared to global averages. For a group of 100,000 women, approximately 63 have breast cancer, while the global average is around 46 (World Health Organization, 2019a). Estimates by the Instituto Nacional do Câncer (INCA) pointed to almost 60 thousand new cases and more than 15 thousand deaths for the year 2018 in Brazil due to breast cancer (BRASIL, 2019).

Although hereditary factors are related to the development of the disease (about 5 to 10%), the main risk factors are related to menstruation (early or late menopause), reproduction (over the age of 30), hormonal intake (contraceptives or hormone replacement), overweight, alcohol consumption, and smoking (BRAY et al., 2018; World Cancer Research Fund, 2019; BRASIL, 2019).

Early detection is essential for combating the disease and for a better prognosis. Clinical breast examination or imaging tests, such as mammography, ultrasound, or magnetic resonance, are non-invasive methods used to perform breast cancer screening (BRASIL, 2019; World Health Organization, 2019a). However, the gold standard for confirming the diagnosis, treatment, and management of breast cancer is given by the histological analysis of a section of the suspected area by a pathologist (BEJNORDI et al., 2017b; World Health Organization, 2019b; YAN et al., 2019; BENHAMMOU et al., 2020).

To perform the microscopic analysis, the suspect tissue undergoes a process of fixation and staining on glass slides. The slides are usually stained with Hematoxylin and Eosin (H&E) (ROY et al., 2019). Hematoxylin is responsible for staining the DNA of the cell nucleus, Ribonucleic Acid (RNA) of the cytoplasm, and cartilage, producing a bluish color in the structures (ROY et al., 2019). While eosin stains cytoplasm structures and pink-colored collagen (YAN et al., 2019). The staining process is necessary because most of the cells and extracellular material are colorless (MESCHER, 2018).

However, some characteristics in histological analysis cause intra and inter-observer variations in diagnosis (MITTAL; SARASWAT, 2019). These variations occur mainly due to the subtle differences between the cancer cells of the types of breast cancer. Besides, the chemical properties of the reagents, differences in the protocols used, and the skills of

pathologists cause variations in the analysis of histopathological slides (KHAN; SIRINUKUNWATTANA; RAJPOOT, 2015; DEBELEE et al., 2020).

Recently, full-slide scanners have become a useful tool for pathologists. With digitalization, case sharing can take place without the slides being physically sent and providing new possibilities for a case evaluation. The digitization of the slides, known in the literature as Whole Slide Imaging (WSI) (SPANHOL et al., 2016a), spread interest in research for the development of computer systems to assist pathologists. For the analysis of histopathological images of breast cancer, Computer-Aided Detection (CAD) or Computer-Aided Diagnosis (CADx) systems have already proved useful, for example, for the detection of mitosis (WANG et al., 2014; MAROOF et al., 2020), cell classification (PAN et al., 2018), and nuclei segmentation (BAKER et al., 2018).

Therefore, the main focus of this Master Thesis is to provide the pathologist with a second opinion for the analysis of histopathological slides of breast cancer in an interpretable way, to improve the diagnosis. To achieve this goal, we propose a model based on a cascade of Convolutional Neural Networks (CNN) that work together to provide a diagnostic suggestion. The model obtains global information from WSI at lower magnification, optimizing the use of computational resources. Based on this global information, a refined analysis is performed at higher magnification. We base the extraction of knowledge on the pathologists' workflow, which locates relevant regions based on a global view, to perform a local and detailed analysis at higher magnification.

1.1 Motivation

The current moment in computing is a reflection of a series of developments and advances, mainly in processing and storage capacities. These advances have enabled the development and improvement of digital image processing techniques and Artificial Intelligence (AI). Using these techniques, some CAD systems applied to medical or biomedical imaging can achieve results close to human capacity (VETA et al., 2014). Especially in recent years, intelligent systems to aid the diagnosis, prognosis, classification or segmentation of pathologies have become the focus of several lines of research (SPANHOL et al., 2016a; CRUZ-ROA et al., 2018; ARESTA et al., 2019).

The need for detection in the early stages of breast cancer supports the need for an increasingly accurate and agile diagnosis (BRASIL, 2019; World Health Organization, 2019a). However, the pathologist is subjected to high workloads in public health care (MAJEED et al., 2018). Allied to this, the various restrictions regarding the heterogeneity of cancer cells, different staining protocols, and chemical properties of the reagents are all factors that contribute to increasing the analysis time, the level of concentration, and the dependence on pathologists' experience and expertise (SPANHOL et al., 2017). For most of the studies in the literature, the intelligent analysis of histopathological images of breast cancer is directed only

to analyze sections of the WSI (SPANHOL et al., 2016a; SPANHOL et al., 2017; GANDOMKAR; BRENNAN; MELLO-THOMS, 2018a), is based on small datasets (ROY et al., 2019), or just classify the slides (BEJNORDI et al., 2017b). Therefore, this Master Thesis presents a methodology for the intelligent analysis of breast cancer WSI. The model explores WSI analysis based on the pathologist's slide analysis routine and provides the pathologist with interpretable information used during the inference to the diagnosis.

1.2 Research Question

This work seeks to answer the following research question:

Is an intelligent model based on deep learning capable of detecting and providing an interpretable diagnosis of breast cancer to pathologists in Whole-Slide Imaging with precision comparable to the golden standard?

This work seeks to answer which characteristics are necessary so that a CAD can be optimized to the interests of the pathologist in assisting the WSI analysis of breast cancer. In this way, we conducted a Systematic Literature Review (SLR) to identify the state of the art, challenges, and open opportunities in assisting the diagnosis in histopathological images of breast cancer.

In this master's thesis, we are interested in analyzing how CNNs can learn the characteristics necessary to detect breast cancer in WSI. As a way to facilitate the learning of CNNs, preprocessing techniques can be used to remove any unwanted noise from the WSI, such as pen markings. We selected CNNs as techniques to compose the methodology of this Master Thesis based on the results analyzed in the SLR.

Finally, as a way of providing an interpretable diagnosis, CNN will be responsible for generating WSI global and local segmentation. The segmentation will present the affected area to the pathologist. The inferences produced by the CNNs will be compared with the golden standard utilizing performance metrics identified in the SLR. In this Master Thesis, we consider the annotations labeled by pathologists and provided by public datasets as the golden standard.

1.3 Goals

The general goal of this work is to propose an intelligent and interpretable architecture to assist in the detection of breast cancer WSI. And so, this CAD system offers a means to reduce intra and inter-observer variations and allow a faster and more accurate diagnosis. To achieve this objective, we have defined some specific goals:

- Conduct initial research to understand the breast cancer WSI analysis process;

- Conduct a SLR to obtain the state of the art of our object of study, identify possible gaps, challenges, technologies used, and available public datasets;
- Propose an intelligent and hybrid architecture to fill the identified gaps;
- Implement the proposed architecture;
- Validate the architecture based on the validation metrics identified in the SLR using images from public datasets.

1.4 Scientific Contributions

In this Master Thesis, we investigate an intelligent model based on Deep Learning (DL) for the detection of breast cancer in WSI. With the model, we intend to provide the pathologist with information regarding the inference made by the techniques, allowing the interpretation of the diagnostic suggestion. In this way, the main contributions of this work can be described below.

- **State of the art mapping.** In order to identify the state-of-the-art for intelligent analysis of histopathological images and the challenges that involve this area, we reviewed the literature. We selected articles published in the last decade in conferences and journals. As a way to conduct the study analysis methodology, we used the SLR method (KITCHENHAM; CHARTERS, 2007; ROEHRS et al., 2017; COVER et al., 2018; MONTENEGRO; COSTA; RIGHI, 2019; GUBERT; COSTA; RIGHI, 2019; ARJI et al., 2019). In this context, we present an extensive literature review in intelligent histopathological image analysis of breast cancer. In particular, we propose an updated taxonomy emphasizing aspects of the proposed methods, as divided into four main topics: preprocessing techniques, processing techniques, post-processing techniques, and evaluation metrics.
- **Novel convolutional architecture for WSI segmentation.** Based on the challenges and open gaps identified in the SLR, we defined the direction of our research. In this way, the main scientific contribution of this dissertation is the proposal of a novel convolutional model for the segmentation of WSI at the pixel level. The model provides segmentation using a cascade of CNNs organized in two levels. In the first, global level, our method analyzes WSI at low magnification and with a larger receptive field. In the second level, our method examines WSI at high magnification, focusing on details and cellular characteristics.
- **Interpretable inferences.** The proposed model is based on the analysis carried out by the pathologist, analyzing global characteristics, and then focusing on local cellular functions. The pixel-level classification allows the pathologist to be provided in an interpretable and detailed way with the cellular aspects that influenced the diagnosis,

both globally and locally. Providing the locations of the regions with precision can assist and accelerate the process of analysis and identification of neoplastic areas in WSI of breast cancer. Finally, we provide a set of heat maps that can help in decision making.

1.5 Text Organization

The remainder of the study is divided into seven chapters. Chapter 2 presents the concepts related to the present work, introducing technologies and algorithms that are used for the development of the proposed system. Then, in Chapter 3, the related works are discussed to present the state-of-the-art in WSI intelligent analysis of breast cancer. Chapter 4 presents the proposed model and design decisions. Chapter 5 presents the validation methodology and metrics for the model. Chapter 6 presents the results, evaluation, and discussions for the proposed model. Finally, Chapter 7 presents the final considerations regarding the findings and future work.

2 BACKGROUND

This chapter presents the essential concepts used in this Master Thesis that guide the development of intelligent architecture to assist in the detection and classification of breast cancer WSI. Characterization of fundamental properties and concepts regarding cancer and histopathology are detailed in Section 2.1. In Section 2.2, we explain some useful digital image preprocessing techniques to assist in the process of learning intelligent algorithms. Meanwhile, in Section 2.3, the theory related to artificial intelligence relevant to this work is discussed. Finally, Section 2.4 presents neural network architectures applied to semantic segmentation.

2.1 Pathology and Breast Cancer

Pathology can be defined as the study of structural and functional changes that manifest themselves as diseases in the body (RUBIN et al., 2015). Given the diversity of the human body, the pathology is subdivided into several specific areas, such as microbiology, hematology, biochemistry, genetics, and cyto-histology (HERRINGTON, 2014). Within histopathology, cell and tissue analysis, both structural and behavioral, plays a fundamental role in the diagnosis and staging of cancer, as well as generating future prognosis (YAN et al., 2019). Within this point, in the next subsections, we present concepts regarding histopathological analysis and breast cancer.

2.1.1 Cancer

The first evidence of cancer, also called a malignant neoplasm or malignant tumor, goes back to periods of prehistory, with mention of the disease in Egyptian, Indian, and Babylonian writings (RUBIN et al., 2015). The evasion of rural life towards an industrial and urban style and behavior exposes man, increasingly, to the growing list of factors that stimulate the emergence of neoplasms (HOFF et al., 2012). A neoplasm can be characterized as an abnormal mass of tissue with uncontrolled growth and proliferation (HERRINGTON, 2014; KUMAR; ABBAS; ASTER, 2018). The capacity for self-renewal, induction of inflammation and tissue invasion, resistance to antiproliferative factors, avoidance of cell death and immunity mechanisms, induction of angiogenesis (formation of new blood vessels), metabolic changes for adaptation contribute to the development of carcinogenesis (HANAHAN; WEINBERG, 2011).

The disease originates from genetic disorders, that is, mutations in human genes that regulate multiplication and differentiation (how much normal cells look like neoplastic cells), providing advantages to altered cells and allowing uncontrolled and autonomous proliferation (HERRINGTON, 2014; BRASILEIRO FILHO, 2016; KUMAR; ABBAS; ASTER, 2018).

The characteristics of such cells will probably reflect the characteristics of the tumors in question, even with the cell population of the same tumor having different aspects, that is, the tumor cells may have different characteristics (HOFF et al., 2012).

There is also a relevant classification that differentiates, due to pathophysiological characteristics, two types of neoplasms: malignant and benign. Malignant neoplasms, or cancer, are tumors capable of rapidly expanding, invading adjacent or distant tissues (metastases). Benign neoplasms are characterized by a well-defined tissue mass, with slower growth and without invasion in adjacent tissues. They are less lethal than the malignant ones, but they can act by compressing neighboring organs or vasculatures, thus evolving into complications (HOFF et al., 2012; DEVITA; LAWRENCE; ROSENBERG, 2018). The classification is confirmed only with a detailed analysis of the suspect tissue. This analysis is known as histopathology.

2.1.2 Histopathology and Cancer

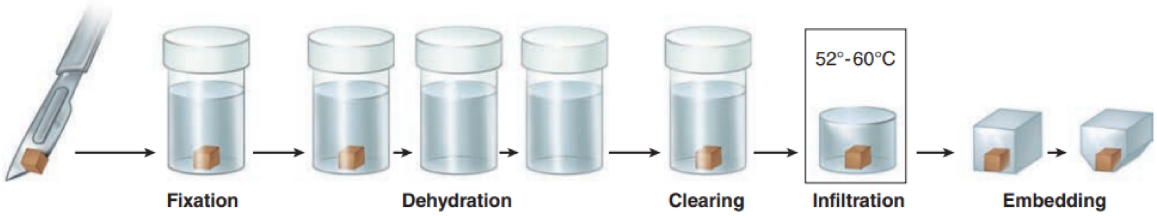
Although findings in clinical and radiological exams (X-ray, magnetic resonance and computed tomography) contribute to the diagnosis and staging of different types of cancer, the gold standard is the macro and microscopic pathological anatomical analysis. In the microscopic examination, the analysis is performed at the cellular level (histopathology) (HOFF et al., 2012; BEJNORDI et al., 2017b; YAN et al., 2019).

Histopathology consists of studying samples of diseased tissue (biopsy) in a microscopic inspection (ROY et al., 2019). With the histopathological exam, it is also possible to obtain indicators for the staging (study to assess the location and extent of the tumor used to choose the ideal treatment) and prognosis (evaluation of possible development or future responses) of cancer (HERRINGTON, 2014).

Obtaining tissue for analysis is done through a biopsy, which can be performed in three ways: excisional when the entire region of the lesion is removed for analysis; incisional, the surgeon removes part of the tissue from the lesion; and by needle, being a less invasive method where light instruments are used with the aid of radiological methods to remove a small but sufficient tissue sample from the lesion (HERRINGTON, 2014). After the acquisition of the tissue, it is prepared according to the steps illustrated in Figure 1.

After the tissue preparation is completed, the resulting sections are placed on glass slides to undergo a process of coloring the structures and examination by electron microscopy (HERRINGTON, 2014). The most frequently used staining method is the combination of H&E (MESCHER, 2018). H&E staining the structures by relating chemical affinities of the dye to simple cell structures. Eosin, a pink dye, due to its acidic and negatively charged (anionic) character, stains basic cellular structures. As an example of structures stained with Eosin, we can mention the cell cytoplasm and mitochondria (KUMAR; ABBAS; ASTER, 2016). Hematoxylin, a bluish-colored dye, has a basic character and positive charge (cationic),

Figure 1 – Process of tissue preparation for histological study. The process steps are briefly: fixation, the tissues are placed in chemical solutions that cross-link proteins and inactivate enzymes; Dehydration, removal of water from tissues using alcoholic solutions; Clearing, process to lighten the tissue; Embedding, the tissue is placed in a mold of melted paraffin to harden; Trimming, the result is then trimmed to section the tussue.



Source: Mescher (2018)

stains acidic cell structures called basophils (ROY et al., 2019). As an example of a basophilic structure, we can mention the cell nucleus itself, which is composed of its acidic groups. The staining process is necessary for most cells and extracellular material to be colorless when observed via microscopy (MESCHER, 2018). An example of a histopathological slide image of breast tissue stained with H&E is shown in Figure 2.

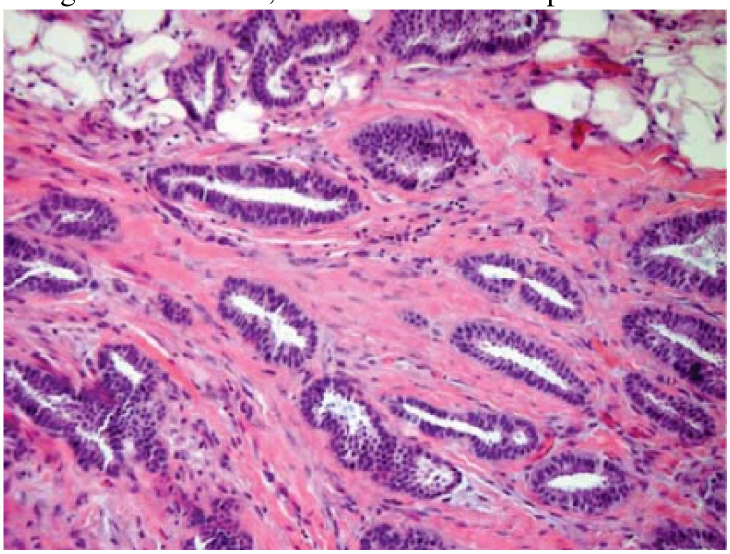
The diversity of colors produced by the interaction of an acid and basic dyes on the slide makes it possible to identify and evaluate even functional issues of the cell. The wide range of tones produced, the intensity of the dye in the medium, and the texture enhanced by the pigment that appears on the slides exposes important cellular information to the medical environment (CHAN, 2014), capable of communicating and differentiating, even, normal or pathological cellular processes (MITTAL; SARASWAT, 2019).

For the diagnosis of neoplasms, and the distinction between benign and malignant, in histopathological slides, the following characteristics are briefly considered: the degree of cell differentiation, mitotic activity, growth pattern, invasion of adjacent tissue, and presence of metastases (HERRINGTON, 2014; RUBIN et al., 2015). Still, when studying and analyzing sections of stained tissue, the pathologist needs to stick to artifacts produced during the stages of the tissue preparation process. Because this process can distort tissues, cause wrinkles, dyes may not correctly dye specific cellular structures, in addition to being aware of interpreting a three-dimensional (3D) structure in a two-dimensional (2D) space (MESCHER, 2018).

2.1.3 Breast Anatomy

The anatomy of the breast is directly related to the behavior of breast cancer (JESINGER, 2014; KUMAR; ABBAS; ASTER, 2016). Therefore, it is essential to understand the structures that make up the mammary gland. The breasts are superficial glands that protrude from the anterior wall of the human chest. They are formed by glandular tissue, which provides nutrition to newborns, fibrous and adipose tissue forming the stroma that supports and shapes the breast

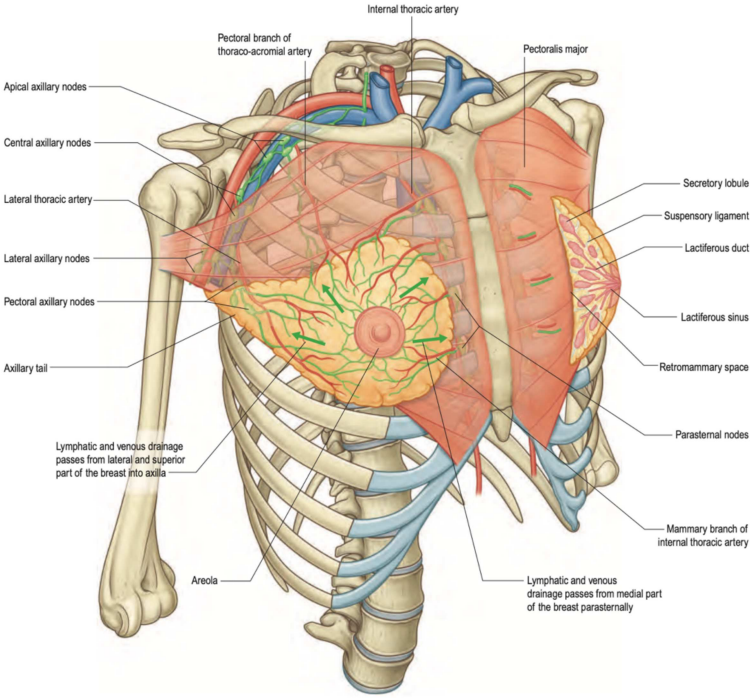
Figure 2 – Example of breast tissue stained with H&E. The pink tones represent structures of the cytoplasm and collagen. Meanwhile, bluish tones are components of the tissue core.



Source: Pawlina e Ross (2018)

(MOORE; DALLEY; AGUR, 2013). Amid these tissues, there is rich vascularization for tissue nutrition, lymphatic ducts, and nerves. The breasts are located in what we call the subcutaneous mesh, anterior to the pectoralis major and minor muscles (STANDRING, 2015). A detailed representation of the breast anatomy is illustrated in Figure 3.

Figure 3 – The relations of the breast.



Source: Adapted from Standring (2015)

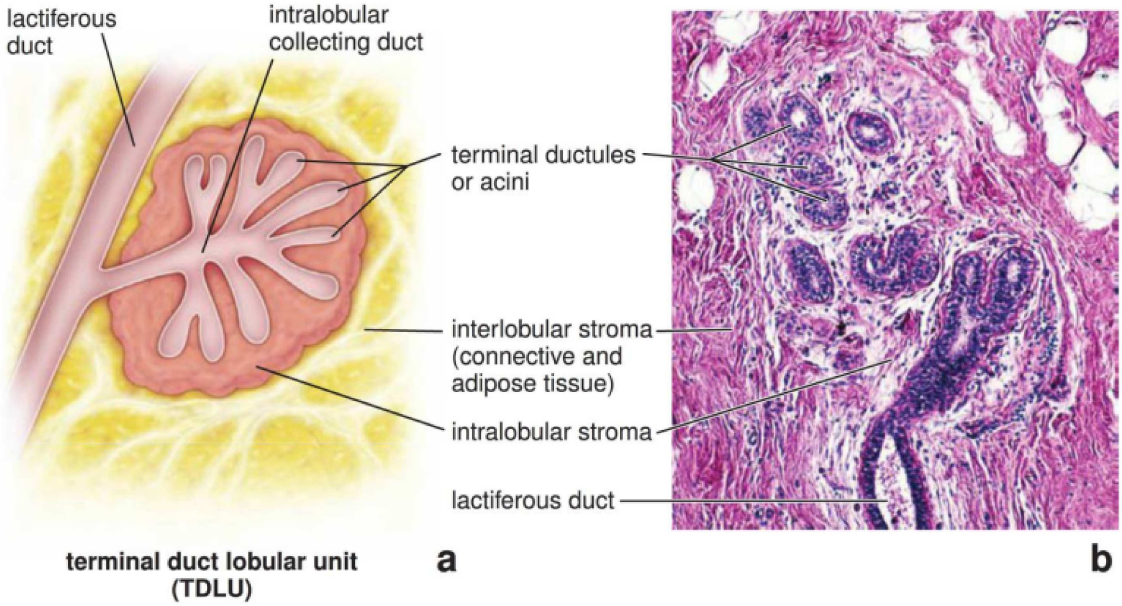
The breast is accommodated in a space that extends from the lateral margin of the external, long bone in the medial portion of the chest, up to the midline of the ribs II to VI. A pectoral

fascia forms this bed over the pectoralis major and serratus anterior muscle. Between the breast and the space of the fascia (mentioned above) is the retro mammary space, which gives a certain degree of movement to the breast. A part of the mammary gland can be found in the axillary region (MOORE; DALLEY; AGUR, 2013). The visualization of the locality and involvement with the superficial chest can be visualized in Figure 3.

2.1.4 Breast Cytology and Histology

The mammary gland is composed of 15 to 20 lobules of tubuloalveolar glands, located in the subcutaneous tissue. These glands differ during the woman's development from apocrine sweat glands. The lobes radiate from the nipple, also called the mammary papilla, located at the center of the areola, and divide into numerous lobular units of the terminal duct (PAWLINA; ROSS, 2018). Each of these terminal units of the mammary gland is a set of small secretory alveoli or terminal ducts surrounded by intralobular stroma. It should be noted that each lobe is composed of at least three essential parts (PAWLINA; ROSS, 2018): terminal ducts, which are part of the inactive gland and, when activated, differentiates into alveoli producing breast milk; intralobular collecting duct, which transports glandular secretions from the alveoli into the galactophore duct in the nipple; and intralobular stroma, which is nothing more than the tissue that is attached to the alveoli and is sensitive to maternal hormones. The components are illustrated in Figure 4.

Figure 4 – **A**, the lobe is composed of multiple acins (responsible for milk production). **B**, the lobular duct is on the left, coming out of the lobe.



Source: Pawlina e Ross (2018)

2.1.5 Breast Cancer

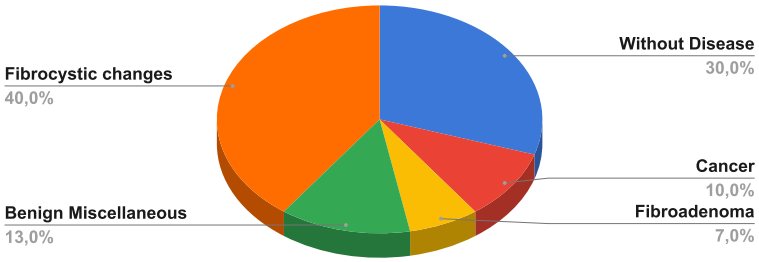
The development of breast tumors is one of the most common in humans. In the female population, they reach up to a quarter of the total number of cancers in the world (World Health Organization, 2019a). It is noteworthy that, even though the number of confirmed tumors is 100 times higher in women, men can also be affected by the disease (MAKKI, 2015).

Although most breast cancers start in cells close to the ducts or lobes (American Cancer Society, 2019), it can manifest in any of the cells of the mammary gland (MAKKI, 2015). This characteristic makes neoplasms highly heterogeneous, both at the molecular and clinical levels (POLYAK, 2007; CIANFROCCA; GRADISHAR, 2009; ELIYATKIN et al., 2015). The WHO defines classifications for the types of breast cancer, but there are doubts about whether this classification is biologically significant (LAKHANI; CANCER; ORGANIZATION, 2012; ELIYATKIN et al., 2015). These doubts are justified by the wide variation in morphological characteristics, immunohistochemical profiles, and pathological subtypes, which result, in most cases, in a single outcome for each patient (MAKKI, 2015). In the next subsections, we discuss and present some characteristics that differentiate benign and malignant tumors, as well as some representative subdivisions.

2.1.5.1 Benign Breast Lesions

Benign lesions of the breast correspond to more than 50% (Figure 5) of the cases of histopathological analyzes of breast cancer (KUMAR; ABBAS; ASTER, 2016). Benign lesions that are not associated with an increased risk of malignancy include duct ectasia (dilation or enlargement of the tubular structure), simple fibroadenoma, fibrosis, mastitis, mild hyperplasia, and cysts (DABBS, 2017).

Figure 5 – Histopathological findings in a series of women seeking breast lump evaluation.

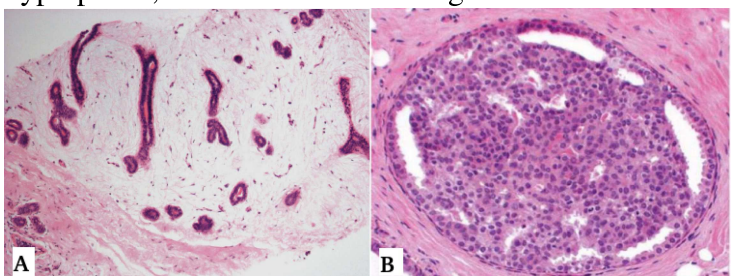


Source: Adapted from Kumar, Abbas e Aster (2016)

Cysts and fibrosis are the most common non-proliferative lesions, characterized by increased stroma and dilation of the ducts (KUMAR; ABBAS; ASTER, 2016). Another common histopathological finding is fibroadenoma (Figure 6 - A), common in women of the

third age, with stromal proliferation associated with a variable number of tubular structures (HERRINGTON, 2014). In hyperplasia, on the other hand, there is an increase in the number of cells that fill the duct (Figure 6 - B), in a normal duct this filling is a double layer of cells (KUMAR; ABBAS; ASTER, 2013).

Figure 6 – **A**, Fibroadenoma: biopsy with a well-defined mass of connective tissue and tubule-like structures. **B**, Hyperplasia, duct filled with heterogeneous cells.



Source: Adapted from Herrington (2014)

In ductal ectasia, dilation of the ducts occurs, with chronic inflammation, usually without significant symptoms and characteristics that can develop into a malignant neoplasm (HERRINGTON, 2014). Fibroadenoma is not usually associated with the development of malignant neoplasms. However, in more severe cases, with the presence of cysts or calcification, the chance is slightly higher (DABBS, 2017). Phylloid tumors and papillomas are benign lesions that can also be associated with the development of breast cancer (RUBIN et al., 2015). Phylloid tumors, in the vast majority, are benign lesions, have a well-defined abnormal mass, and are composed of stromal cells and covered by epithelium (HERRINGTON, 2014). Meanwhile, papilloma is more frequent in premenopausal women, with heterogeneous cell proliferation in the ducts, and is generally associated with nipple secretion (KUMAR; ABBAS; ASTER, 2013). In this work, we will use the definition of benign disease for all changes found in histopathological images that are not considered carcinogenic changes.

2.1.5.2 Malignant Breast Lesions

Breast carcinoma is the most common malignancy among women (KUMAR; ABBAS; ASTER, 2016). Of these, almost 95% are originated and limited to the breast duct and lobular system (MAKKI, 2015). When they are of epithelial origin, and there has not yet been an invasion of adjacent tissues, they are referred to in the literature as carcinomas *in situ* (MOINFAR, 2007). Data refer to that, of the patients who did not receive the appropriate treatment, from 20% to 30% subsequently developed invasive carcinoma (RUBIN et al., 2015). When the tumor escapes from the epithelial tissue and compromises other tissues becoming invasive, it can be classified into two types: those that arise from the ducts, ductal carcinoma, and those that arise from the lobes, lobular carcinoma. However, both types of tumors originate in the same segment of the mammary gland, called the lobular unit of the terminal

duct, as seen in Figure 4 (MAKKI, 2015). The carcinoma develops in the mammary stroma and infiltrates the basement membrane. In these cases, there is the possibility that invasions to the surrounding tissues reach the lymph nodes or blood vessels and create metastases (KUMAR; ABBAS; ASTER, 2016). The following are the main classifications defined by WHO and some relevant characteristics for carcinomas:

- **Ductal Carcinoma *in situ***: this type of tumor is characterized by its proliferation in the epithelial cells limited to the ducts or lobes. It is considered the precursor carcinoma of invasive pathologies (MAKKI, 2015). Corresponds to about 25% of breast cancers in screened populations (KUMAR; ABBAS; ASTER, 2016);

- **Lobular Carcinoma *in situ***: it is also an abnormal proliferation of cells within ducts and lobes, that is, intralobular, but without altering the spaces involved during the spread of the neoplasia (KUMAR; ABBAS; ASTER, 2013). It happens with the proliferation of small and heterogeneous cells with weak intercellular cohesion. This type of carcinoma occurs between 25% to 30% (around 1% per year) of patients who are monitored for more than 20 years (MAKKI, 2015);

- **Invasive Ductal Carcinoma**: of invasive breast cancers, 50% to 70% fall into this category. In Invasive Ductal Carcinoma, malignant ductal proliferation occurs with an invasion of the mammary stroma. It is a heterogeneous group of tumors classified according to the architectural structure of the cells, with a series of variations in morphology, location, and immunohistochemistry (MAKKI, 2015). Invasive Ductal Carcinoma can be further divided into Invasive Ductal Carcinoma of non-specific type, Tubular Carcinoma, Mucinous Carcinoma, Papillary Carcinoma, Adenoid Cystic Carcinoma, Metaplastic Carcinoma, Micropapillary Carcinoma, Lipid-rich Carcinoma, Secretory Carcinoma, or as Cell Carcinoma Acinar;

- **Invasive Lobular Carcinoma**: it consists of cells morphologically similar to Lobular Carcinoma *in situ*, and almost 70% of cases are derived from it. In addition, the cells invade the stroma and organize themselves aligned in cord forms (KUMAR; ABBAS; ASTER, 2016). The cell group is usually rounded, small, uniform, and without cohesion. It has a standard growth characteristic with a single infiltration for the breast stroma (MAKKI, 2015). Invasive Lobular Carcinoma can be further divided into Pleomorphic Lobular Carcinoma, Histiocytoid Carcinoma, Signet Ring Cell Carcinoma, or Tubulo-lobular Carcinoma;

Currently, with all technological advances that involve the health area, the deepening of molecular knowledge and DNA sequencing has allowed to develop content and understanding of the initiation and progression of diseases, including breast cancer. Thus, this new classification appears based on the expression of proteins and genes (GOBBI, 2012). The main

breast cancer subtypes, according to this classification, are luminal A, luminal B, protein HER2, and triple-negative (MAKKI, 2015).

Finally, the correct classification of the lesion according to histological characteristics is essential for defining the appropriate treatment and influencing the outcome and response to treatment (MAKKI, 2015). Therefore, a detailed analysis of the suspected tissue sections by the pathologist is of fundamental importance. Image processing techniques are already used in the pathologist's routine to assist in the detection of suspicious regions, and present in software such as ImageJ (RASBAND et al., 1997). In this regard, we present in the next section essential concepts for image processing and its applications in WSI.

2.1.5.3 Prognostic

The tumor stage directly influences the patient's prognosis (RUBIN et al., 2015). The staging of a malignant neoplasm is a classification of how far the tumor has spread to determine the appropriate treatment (KUMAR; ABBAS; ASTER, 2016). The system used for more than 50 years for the staging of malignant neoplasms is TNM, based on the anatomical extension of the primary tumor (T), lymph node characteristics (N), and the presence or absence of distant metastasis (M) (KUMAR; ABBAS; ASTER, 2013). In the last revision of the system, there was a substantial modification, in addition to the anatomical characteristics of the lesion, the system now integrates biomarkers as the second level of prognostic modifiers (GIULIANO et al., 2017).

Another essential prognostic classification is histological grading. The system currently used is known as the Nottingham Classification System. It evaluates the percentage of tubular differentiation, nuclear pleomorphism, and the mitotic index (GALEA et al., 1992). From this percentage, the lesion is classified in one of the following degrees: G1- well differentiated; G2 - moderately differentiated, and G3 - weakly differentiated (RUBIN et al., 2015).

2.2 Digital Image Processing

Digital Image Processing (DIP) is used to assist or automate the solution of several problems, such as remote sensing, automated inspection of manufactured products, analysis of medical images, and security systems (JäHNE, 2005). DIP, unlike human vision, can cover almost the entire electromagnetic spectrum (GONZALEZ; WOODS, 2018). Therefore, DIP can be defined as a set of computational techniques to capture, represent, manipulate, and extract information from digital images (JäHNE, 2005; PEDRINI; SCHWARTZ, 2008).

2.2.1 Digital Image

The computational representation of a digital image can be performed in several ways, the most common being the spatial representation (JäHNE, 2005). The spatial representation is composed of a finite set of $M \times N$ elements, where M and N correspond, respectively, to the height and width of the image (GONZALEZ; WOODS, 2018). Each of these elements, also called pixels, constitutes a discretization of a rectangular region of the scene or component to be digitally characterized (BURGER; BURGE, 2009).

A monochrome digital image $f(x, y)$, where x and y are spatial coordinates, and f corresponds to the intensity or level of gray in the image, can be represented as a 2D matrix (JäHNE, 2005; GONZALEZ; WOODS, 2018). For a color digital image, in general, a 3D representation is required (BURGER; BURGE, 2009). The most used standardized models for the 3D representation of digital color images are:

- **RGB** *(Red, Green and Blue)*: each channel represents one of the primary colors, and the sum of the intensities for each channel forms the final color (GONZALEZ; WOODS, 2018);
- **HSV** *(Hue, Saturation and Value)*: the first channel represents the base color of the spectrum, the second channel indicates the hue of the final color (the lower the value, the closer to a shade of gray the color will be), and the last channel defines the brightness of the color (BURGER; BURGE, 2009) ;
- **CIELAB**: it is represented by three channels, L, which corresponds to intensity, A colored component that varies from green to magenta, and B colored part that changes from blue to yellow (GONZALEZ; WOODS, 2018).

2.2.2 Morphological Image Processing

Morphological image processing (or just morphology) is the branch of digital image processing that focuses on the study of geometric structures present in the image (MARQUES FILHO; VIEIRA NETO, 1999). The morphological operations are applied to the images with the structuring elements, which are small sets or sub-images used to search for the information of interest (GONZALEZ; WOODS, 2018).

Dilation is a morphological process responsible for increasing the structures that make up an image (BURGER; BURGE, 2009). The operation can be defined mathematically as:

$$A \oplus B = \{z | [(\widehat{B})_z \cap A] \subseteq A\} \tag{2.1}$$

The Equation 2.1 indicates that the expansion of A by B is the set of all the displacements of z, so that $\widehat{B}$ and A overlap at least one instead (GONZALEZ; WOODS, 2018). Where A is

the image, and B is the structuring element.

While the expansion expands the elements that make up an image, erosion does the reverse process, reducing the size of the objects.

$$A \ominus B = \{z|(\widehat{B})_z \subseteq A\} \tag{2.2}$$

The Equation 2.2 indicates that the erosion of A by B is the set of all points z, so that B, translated by z, is contained in A (GONZALEZ; WOODS, 2018). Where, again, A is the image, and B is the structuring element.

2.2.3 Image Segmentation

Segmentation is defined as the subdivision of an image into two or more significant regions (GONZALEZ; WOODS, 2018). The division is carried out at the pixel level, so those pixels with the same label are connected concerning some visual or semantic property (GHOSH et al., 2019). With the segmented image, it is possible to identify the boundaries between the objects and the shape of the objects (JäHNE, 2005).

The thresholding technique is equivalent to classifying each pixel of the image in background and object based on a predefined threshold value (BURGER; BURGE, 2013). It is a useful and straightforward technique to perform an image binarization or to remove noise (GONZALEZ; WOODS, 2018).

The Equation 2.3 defines mathematically how the thresholding of an image occurs. Therefore, given an input image, a threshold value T is specified, this threshold is a number referring to the intensity of color or shade of gray present in the image. After setting the threshold, iterates over all the pixels of the image to check their values and divide them into two classes: the first class is made up of pixels with values less than or equal to T that assume the value 0 (background); the second class is composed of values higher than T that assume the value 1 (object). Therefore, the segmented image can be obtained by $g(x, y)$ (BURGER; BURGE, 2009):

$$g(x,y) = \begin{cases} 1, \text{ if } f(x,y) > T \\ 0, \text{ otherwise} \end{cases} \tag{2.3}$$

The definition of the threshold T is an essential factor in the segmentation of the image, being responsible for defining which part of the image (background or object) the pixel belongs. A method to obtain a global threshold, to separate an object from the background, was proposed by Otsu (1979), wherefrom an image in shades of gray, ideal value for the limit is determined. The proposed algorithm iterates through all the pixels of the image looking for a value that minimizes the sum of the variance between classes, separating the image in background and object, that is, trying to make the pixel variance in each of the categories to be minimal (BURGER; BURGE, 2013; GONZALEZ; WOODS, 2018).

2.2.4 Adaptive thresholding

It is essential to consider that segmentation with the thresholding technique only achieves satisfactory results in images in which there is a distinction between the colors of the background and the objects (SONKA; HLAVAC; BOYLE, 2015). A solution for images with variation in lighting, backgrounds, or objects of uneven color tones can be the use of adaptive thresholding. Adaptive thresholding uses variable limits to perform segmentation, considering local characteristics of the image (BURGER; BURGE, 2013).

One of the most basic ways to perform adaptive thresholding is to split the original image into several sub-images and apply the Otsu (1979) method to each of these sub-images (GONZALEZ; WOODS, 2018). Another approach is to calculate the threshold for each (x, y) point based on the characteristics of your neighborhood (BURGER; BURGE, 2013). A common way to perform this processing is taking into account the standard deviation and the average of the pixels in the vicinity of each point in the image so that the segmented image can be calculated as:

$$g(x,y) = \begin{cases} 1, \text{ if } f(x,y) > a\sigma_{xy} + bm_{xy} \\ 0, \text{ otherwise} \end{cases} \tag{2.4}$$

where $f(x, y)$ is the input image, a and b are positive constants, σ and m correspond, respectively, to the standard deviation and the average of the pixels contained in the neighborhood S_{xy} (GONZALEZ; WOODS, 2018).

However, the simple use of image processing algorithms does not guarantee the identification of patterns in images is due to the variability of factors that influence them, such as luminosity, depth, and even the location of objects in the image (HAN et al., 2017; PAN et al., 2018; FENG; ZHANG; YI, 2018; MITTAL; SARASWAT, 2019). Thus, the use of techniques that can provide the computer with artificial learning capacity is necessary to achieve satisfactory results.

2.3 Artificial Intelligence

"Can machines think?" is a discussion that started in the 50s, so it is considered as one of the first articles to discuss AI (TURING, 1950). After seven decades, AI has transformed human interactions significantly, influencing which meals we eat, which routes we take to work, which people we relate to, which destinations we visit, among others (ALPAYDIN, 2016; MÜLLER; GUIDO et al., 2016). However, for AI applications on a global scale to be possible, advances in various areas of computing were necessary, allowing the existence of mobile and accessible computing with significant capacity for processing, storing, and transmitting information in real-time (ALPAYDIN, 2016).

AI is a broad area, studying systems that can automate intellectual tasks performed by

humans (COPPIN, 2010). In the early days of AI, the automation of tasks was implemented, defined rules manually by programmers, based on human knowledge of the process. This process was also known as symbolic AI (MÜLLER; GUIDO et al., 2016). However, this approach was not efficient for solving complex problems, such as image classification and speech recognition (CHOLLET, 2018). The two main issues of rule-based AI: small changes may require rewriting the entire system and the need for in-depth knowledge of the problem to be automated (MÜLLER; GUIDO et al., 2016).

In this context, AI systems need to acquire their knowledge and be able to adapt to changes (ALPAYDIN, 2016). It is from this problem that a new AI subfield, Machine Learning (ML), emerges. Unlike symbolic AI, ML seeks to obtain a set of rules that solve the problem. For this, ML algorithms are trained to identify a function $f : X \rightarrow Y$ that maps the input X (data) to an output Y (predictions for the problem) (BEKKERMAN; BILENKO; LANGFORD, 2011; CHOLLET, 2018). Therefore, ML can be defined as the process of gathering a data set and developing an algorithm that generates a statistical model based on this data set to solve a practical problem (BURKOV, 2019).

Within ML, algorithms are often classified into four categories: **supervised learning**, **unsupervised learning**, **semisupervised learning**, and **reinforcement learning** (GÉRON, 2017). In supervised learning, the data used in training needs to be previously labeled by an expert. Based on this data, the algorithms are trained, mapping an input X to an output Y, producing a model (BURKOV, 2019). After training, the model can be saved and used to predict results based on new data (GOODFELLOW; BENGIO; COURVILLE, 2017). Unsupervised learning uses data where only the entry X is known, requiring no data labeled (MÜLLER; GUIDO et al., 2016). The algorithms seek to find similarities, patterns, or irregularities to group data, detect anomalies, or reduce the dimensionality of the X entry (GÉRON, 2017). As the data labeling process is usually expensive and time-consuming, some techniques can use partially labeled data. These algorithms are classified as semisupervised learning. When using semisupervised learning techniques, it is expected that the unlabeled data can somehow help the model to find a better solution (BURKOV, 2019). Finally, in reinforcement learning, the learning system, called an agent, can observe the environment, selecting and executing actions, and receiving rewards in exchange (GÉRON, 2017). Reinforcement learning techniques can be characterized as an agent who learns a policy from reward signs, interacting with his environment. The agent aims to find an ideal policy that maximizes your cumulative reward (SUTTON; BARTO, 2018).

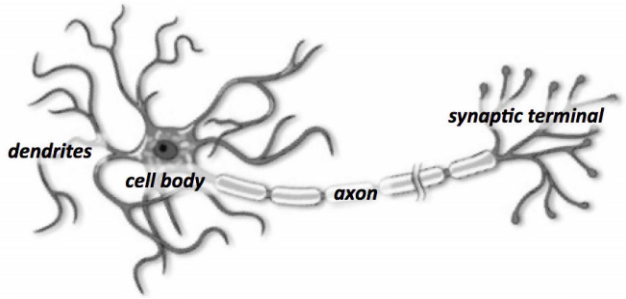
Within ML techniques, we can highlight a subset of methods that have been gaining prominence in research in recent years called DL. DL is based on the neural connections that exist in the human brain (GOODFELLOW; BENGIO; COURVILLE, 2017). The connections are simulated using dozens and sometimes hundreds of levels of mathematical functions providing a different interpretation of the data it receives (CHOLLET, 2018). With regard to this focus, we discuss in the next subsections relevant details of the Artificial Neural Network

(ANN). This theoretical basis is necessary to understand the methods discussed in the next sections.

2.3.1 Artificial Neural Networks

The human brain has hundreds of billions of interconnected neurons that process information from the human body or the external environment (BUDUMA; LOCASCIO, 2017). A biological neuron is composed of four main elements: dendrites, axons, cell body, and synaptic terminals (RUSSELL; NORVIG, 2010). In Figure 7, a biological neuron is illustrated, and the dendrites are the receptors for information and stimuli. Each of these inputs has a connecting force, which can be weakened or reinforced according to its use. Through the process of weakening or strengthening connections, the learning process occurs (BUDUMA; LOCASCIO, 2017). When the information reaches the cell body, it is added and propagated along the axon reaching the synaptic terminals that are connected to other neurons (HAYKIN, 2009).

Figure 7 – A representation of the structure of a biological neuron.



Source: Buduma e Locascio (2017)

With the neurobiological analogy as a source of inspiration, research to create an ANN has been conducted since the beginning of AI (HAYKIN, 2009; RUSSELL; NORVIG, 2010; BUDUMA; LOCASCIO, 2017). In the next subsections, we cover notions that form the basis of ANNs and the learning process with neural networks.

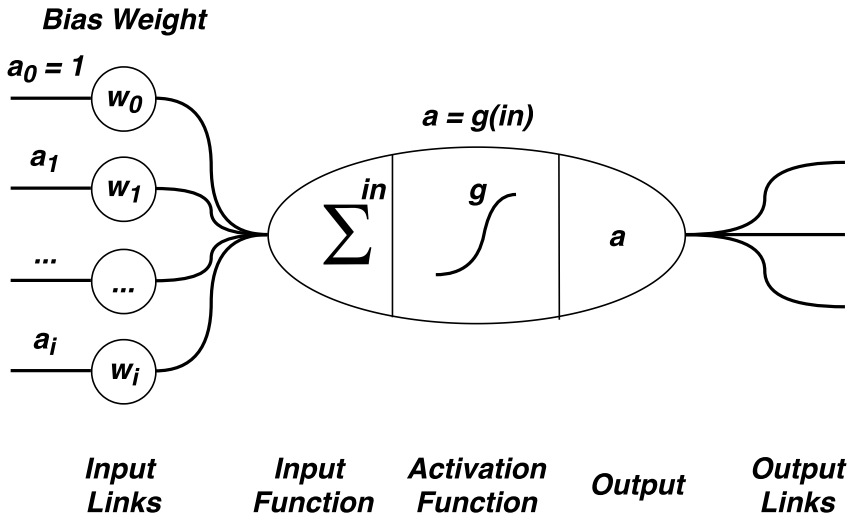
2.3.1.1 Artificial Neuron

McCulloch e Pitts (1943) were the first to study and propose a mathematical model of an artificial neuron. The proposed model was composed of one or more binary inputs and a binary output. With this model, any logical proposition door could be built as desired (GÉRON, 2017). However, this model was still very primitive (RUSSELL; NORVIG, 2010). An evolution was proposed by Rosenblatt (1958), allowing the input to the neuron to be any numerical value and the output being calculated from a threshold function, this model became known as *Perceptron*.

As in the biological neuron, the artificial neuron (Figure 8) is made up of some a_i entries,

each multiplied by a specific w_i weight that corresponds to the strength of the connection (BUDUMA; LOCASCIO, 2017). In addition, each neuron usually has a constant as a fictitious input from the network, known as bias, and with a value normally equal to 1 (RUSSELL; NORVIG, 2010). The weighted inputs are added together, and the result is applied to a g function, also known as an activation function, to produce the output of the neuron (HAYKIN, 2009).

Figure 8 – A simple mathematical model for a neuron.



Source: Adapted from Russell e Norvig (2010)

2.3.1.2 Activation function

The input of a neuron is a linear transformation ($a_i * w_i$). However, complex problems, such as image classification, are non-linear problems. Thus, the objective with the use of the activation functions is to introduce nonlinearity to the neuron (GOODFELLOW; BENGIO; COURVILLE, 2017). Therefore, an activation function $\alpha(v)$ defines the output of the neuron v and must be differentiable and capable of mapping any entry from $-\infty$ to $+\infty$ to a specific interval, for example, between $\{0.1\}$ or $\{-1.1\}$ (HAYKIN, 2009).

- **Step or threshold**: the output is equal to 1 if the value obtained by the weighted sum of the inputs is greater than or equal to zero. Otherwise, it will be -1, is defined by:

$$\alpha(v) = \begin{cases} 1 \text{ se } v \leq 0 \\ -1 \text{ se } v < 0 \end{cases} \tag{2.5}$$

  It is important to note that the step function is not differentiable (Figure 9 (B)); therefore its use in ANN is not indicated (HAYKIN, 2009).

- **Sigmoid**: the sigmoid function graphically has an "S" shape (Figure 9 (A)). We can also see that when the weighted sum of the inputs is small, the output will be close to 0. When

the weighted sum is large, the output will be close to 1. The sigmoid function is defined by:

$$\alpha(v) = \frac{1}{1 + e^{v}} \tag{2.6}$$

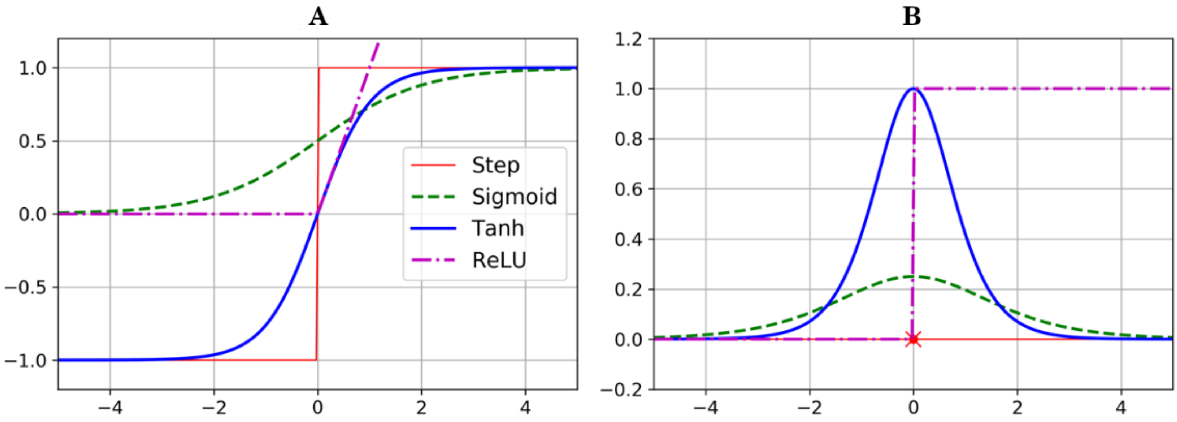
- **Tanh**: tanh graphically behaves similarly to sigmoid (Figure 9 (A)). However, the range is from -1 to 1. Tanh is defined by:

$$\alpha(v) = tanh(v) \tag{2.7}$$

- **ReLU**: the ReLU activation function is often used in computer vision tasks (BUDUMA; LOCASCIO, 2017). Its main advantage over sigmoidal is the fact that it does not suffer from gradient saturation, and its computation is faster. The ReLU returns the maximum between 0 and the value of the weighted sum of the inputs in a neuron, is defined as:

$$\alpha(v) = max(0, v) \tag{2.8}$$

Figure 9 – Activation functions and their derivatives. (A) Activation functions. (B) Derivatives of the activation functions.



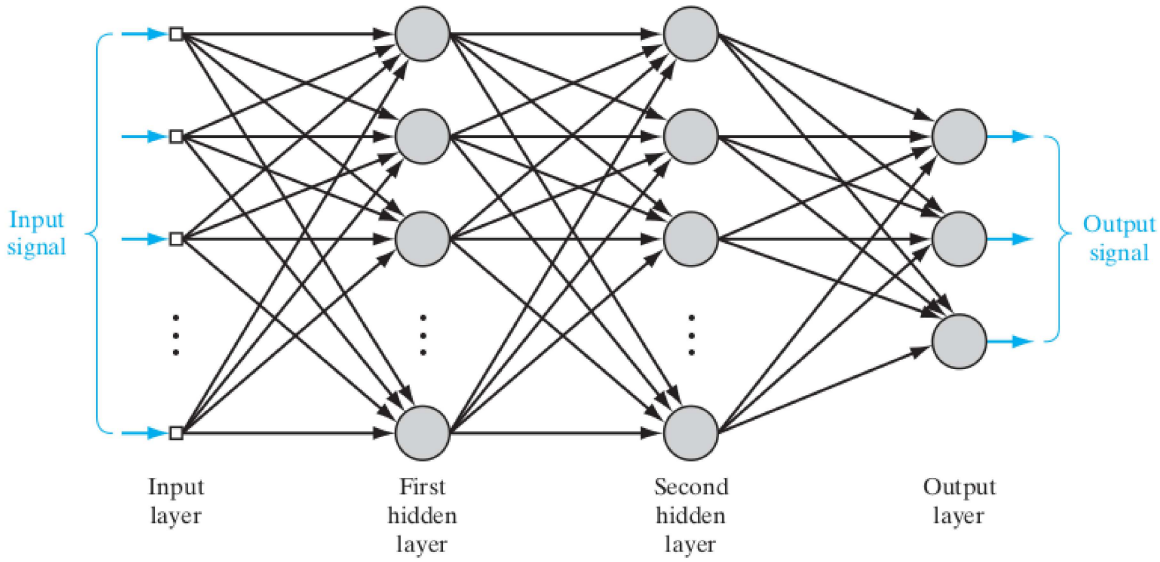
Source: Géron (2017)

2.3.1.3 Multilayer Perceptron

Some of Perceptron's limitations, for example, the unique OR classification problem (XOR) (MINSKY; PAPERT, 1969), can be eliminated by organizing a set of neurons in layers. The organization of neurons in layers is known as Multilayer Perceptron (MLP) (RUSSELL; NORVIG, 2010). An MLP consists of an input layer, one or more hidden layers, and an output layer. Each output signal from a neuron is passed on to all neurons in the adjacent layer; this process is known as propagation. The output signal of a layer is given by processing all output

signals from the previous layer (COPPIN, 2010), as can be seen in Figure 10. The input layer usually just passes the data without modifying it. Most computing takes place in the hidden layers. The output layer converts the activations of the hidden layer into an output, as a classification (BURKOV, 2019).

Figure 10 – Representation of a Multilayer Perceptron.



Source: Adapted from Haykin (2009)

Mathematically an MLP can be expressed by the following formula:

$$y^i = \sigma(\sum_{i=0}^{n} W^i y^{i-1} + b^i) \tag{2.9}$$

where y^i is output from the layer i, W^i corresponds to the weights of the layer, y^{i-1} the outputs from the previous layer, b^i to the bias of the current layer and σ the activation function (HAYKIN, 2009).

The weight update process, also known as learning, is an iterative method. In a neural network, learning occurs by propagating the inputs on the network and computing an error in the output layer. This error is measured using a loss function (or cost function). The measured error is then backpropagated from the output neurons to all other errors using the back-propagation algorithm. In the next subsections, we present the essential concepts of these two fundamental elements in the neural network learning process.

2.3.1.4 Loss functions

The purpose of the loss functions is to measure the difference between the estimated and true values for a given data set (BURKOV, 2019). The lower the value obtained after calculating the loss function, the better the model behaves (CHOLLET, 2018). However, it is necessary to define the loss function carefully; for example, it is worse to classify a non-spam message as

spam, than to classify spam as non-spam (RUSSELL; NORVIG, 2010). Thus, it is necessary to understand how each loss function is calculated and choose the one that is consistent with our problem (CHOLLET, 2018).

In Table 1 we can see some loss functions currently used. Binary Cross-Entropy measures the performance between the classification of the model with a probability between 0 and 1. The closer the output is to 0, the closer the model's prediction to the expected value. However, Binary Cross-Entropy can only be used for binary classification problems. For multiclass classification problems, we must use the Categorical Crossentropy, which calculates a separate loss for each observed class and makes a sum of the result to infer the error of the model (GOODFELLOW; BENGIO; COURVILLE, 2017).

For regression problems, we can use Mean Absolute Error or Mean Squared Error. Mean Absolute Error can be defined as the absolute sum of the differences between the target and predicted variables. Mean Squared Error is the mean of the difference squared between the forecasts and the actual observations (GOODFELLOW; BENGIO; COURVILLE, 2017).

Table 1 – Some loss functions.

Name	**Goal**	**Equation**
Binary Crossentropy	Classification	$-(y \log(p) + (1-y) \log(1-p))$
Categorical Crossentropy	Classification	$-\sum_{M}^{c=1} y_{o,c} \log(p_{o,c})$
Mean Absolute Error	Regression	$\frac{\sum_{i=1}^{n} \lvert y_i - y_i^p \rvert}{n}$
Mean Squared Error	Regression	$\frac{\sum_{i=1}^{n} (y_i - y_i^p)^2}{n}$

Source: Elaborated by the author.

2.3.1.5 Gradient Descent and Back-propagation

To obtain the set of weights that minimize the value obtained from the loss function, the back-propagation method is used. While the x input signals propagate from the input layer to each of the hidden layers producing an output y', the back-propagation takes the opposite path (RUSSELL; NORVIG, 2010). From the output y' and with the output labeled y, the difference in the forecast is calculated with a loss function. The gradient of the loss function is then computed (GOODFELLOW; BENGIO; COURVILLE, 2017). The gradient of a real and continuous function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is the column vector of the partial derivatives of the function at any point.

The gradient obtained for the loss function is then propagated from the output layer to the hidden cameras to calculate the error in each neuron. The hidden layer error is obtained by calculating those derived by the chain rule (GOODFELLOW; BENGIO; COURVILLE, 2017). The neuron weights are then updated by subtracting a proportion of the gradient from the current value of the weights. The calculation of this proportion is defined by multiplying the gradient by the learning rate (COPPIN, 2010). The learning rate is a hyperparameter used to control the

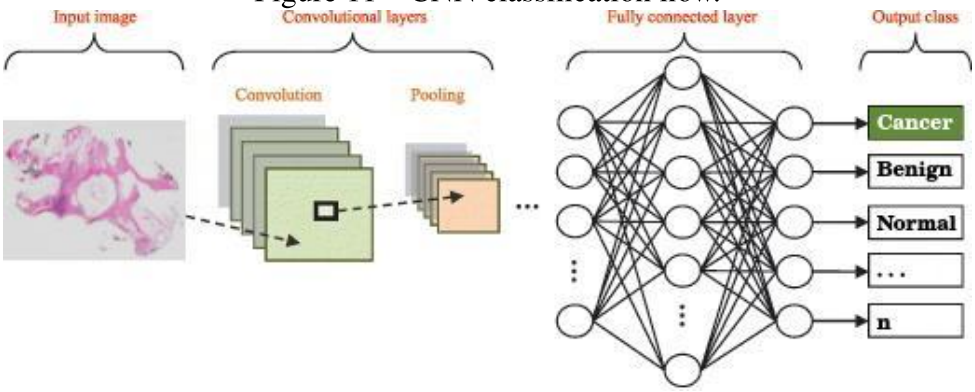
magnitude of the update, which influences the convergence of the algorithm. The definition of this parameter must be done carefully. If the value is too high, the ANN can fluctuate, and if it is too low, the training can be slow and end up stuck in local minimums (HAYKIN, 2009).

2.3.2 Convolutional Neural Networks

Human vision is incredibly advanced, being able to detect objects and patterns in our visual field in fractions of a second (BUDUMA; LOCASCIO, 2017). However, for computer systems, this is still a complex task, due to the variability of objects, dependence on the point of view and the high variability of classes (CIRESAN et al., 2011). Traditionally, the identification of patterns in images was carried out in two stages: resource extraction and classification. The resources described the images numerically and served as input for the classifiers. The main problem with this approach was the dependence on accuracy in identifying patterns with the resource extraction stage (LECUN et al., 1998).

In this context, the use of DL models, such as CNN, proves to be efficient in reducing the dependence on the resource extraction stage (RAWAT; WANG, 2017). Like MLP, CNN was inspired by the visual recognition process of our brain (HAYKIN, 2009). A CNN architecture typically has three layers: convolution, subsampling, and completely connected (KANG; WANG, 2014). Figure 11 illustrates a typical CNN architecture for image classification. The convolutional and subsampling layers are directly inspired by the notions about the functioning of cells in visual neuroscience (HUBEL; WIESEL, 1962).

Figure 11 – CNN classification flow.



Source: Adapted from Rawat e Wang (2017)

In the next subsections, we present the essential components of a CNN, along with some key ideas of the CNN.

2.3.2.1 Convolutional Layers

A convolution is a special type of linear operation in two functions of an argument with a real value that expresses how the other modifies the shape of one (GOODFELLOW; BENGIO;

COURVILLE, 2017). The $s(t)$ convolution of two functions $x(t)$ and $w(a)$ is an integral that expresses the overlap of these two functions, shifting $w(a)$ over $x(t)$, being defined as:

$$s(t) = (x * w)(t) = \int_{-\infty}^{+\infty} x(a)w(t-a)da \tag{2.10}$$

The function $s(t)$ can be seen as the function $x(t)$ weighted by $w(a)$. The convolution operation presented in Equation 2.10 assumes that we can obtain data at any time. However, normally in computer systems, time is discrete, and data is received or collected at regular intervals (GOODFELLOW; BENGIO; COURVILLE, 2017). Therefore, the operation of convolution in discrete data is defined by:

$$s[t] = (x * w)[t] = \sum_{a=-\infty}^{\infty} x[a]w[t-a] \tag{2.11}$$

In the context of CNN, we can adapt the discrete convolution process by taking the function $x[t]$ as the input image I, the function $w[t]$ as the filter F and the result $s[t]$ as S feature map, the operation being defined by:

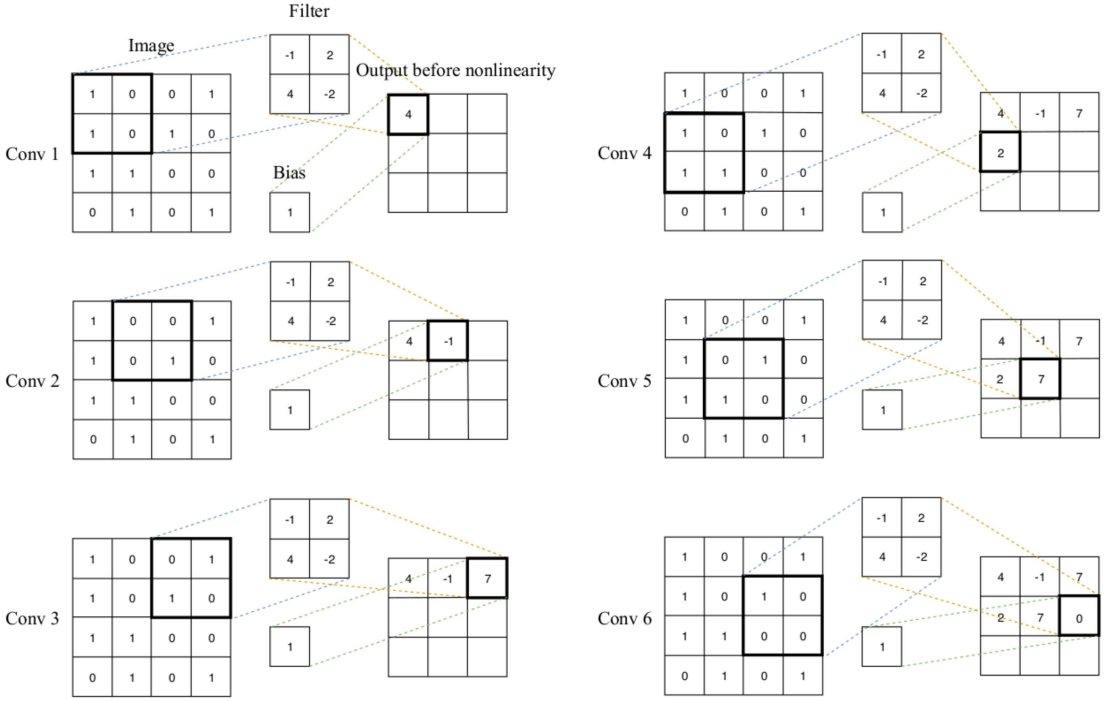
$$S[i,j] = (I * F)[i,j] = \sum_{n}\sum_{m} I[m,n]F[i-m, j-n] \tag{2.12}$$

Making an analogy with MLP, in CNN, the trainable weights are a set of combined filters, forming a convolutional layer (CHOLLET, 2018). However, while in an MLP, a neuron is connected to all neurons of the next layer in the CNN, a neuron affects only a subset of the neuron of the next layer (GOODFELLOW; BENGIO; COURVILLE, 2017). This change in the weight connection provides CNNs with two advantages: sparse connections and weight sharing (LECUN; BENGIO; HINTON, 2015). In images, patterns are often repeated in different positions. In this way, with sparse connections, we can identify small sets of patterns, such as lines and curves in the initial layers, and more complex patterns such as animals, people, and cars in the last layers (ZEILER; FERGUS, 2014). The identification of the patterns is performed using small filters (usually much smaller than the input images) that are slid over the entire image (GOODFELLOW; BENGIO; COURVILLE, 2017). This process is illustrated in Figure 12. Therefore, each filter is responsible for detecting a type of feature in the image, thus generating several feature maps throughout the processing of an image (LECUN; BENGIO; HINTON, 2015).

A convolutional layer has some hyperparameters that need to be controlled, namely:

- **Filter size**: defines the size of the convolution filter, in Figure 12 the size is $2x2$ pixels;
- **Number of filters**: defines the number of filters in the current layer, in the representation of Figure 12 we have only one filter;
- **Padding**: in Figure 12, we can see that the size of the obtained feature map is smaller

Figure 12 – A filter convolving across an image.



Source: Burkov (2019)

than the size of the input image. Therefore, in some situations, it is interesting to fill the edges of the input image with zeros or based on some statistical measures so that the size of the feature map is equal to the input image;

- **Stride**: defines the size of the step that the filters perform on convolution, in Figure 12, for example, the step is 1 pixel.

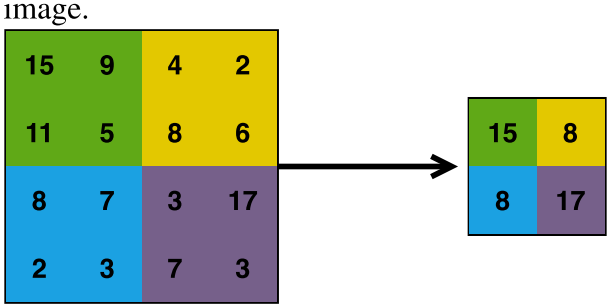
2.3.2.2 Pooling Layers

The objective with the use of pooling functions is to reduce the spatial resolution of feature maps (LECUN et al., 1998; LECUN; BENGIO; HINTON, 2015). Like the convolutional layers, the pooling layers have hyperparameters to be defined, namely: size, padding, and stride. However, the pooling layers do not have trainable parameters (GÉRON, 2017). A common pooling function is max pooling, which returns the maximum value for a region (GOODFELLOW; BENGIO; COURVILLE, 2017). Figure 13 illustrates the max-pooling process.

2.3.2.3 Fully Connected Layers

Finally, in the last layer of convolutional architectures, fully connected layers are used. In these layers, the behavior of neurons is connected to all activations of the previous layer, as

Figure 13 – $2x2$ pixel max-pooling process with a 2 pixel stride applied to a $4x4$ pixel image, resulting in a $2x2$ pixel image.



Source: Adapted from Rawat e Wang (2017)

in MLP (GÉRON, 2017; GOODFELLOW; BENGIO; COURVILLE, 2017; RAWAT; WANG, 2017).

2.3.3 Overfitting and Underfitting

For an ML model to be considered efficient, it must also perform well on data not used in the training process. This capability is called generalization (GOODFELLOW; BENGIO; COURVILLE, 2017). Usually, when we start the process of training an ML model, we divide our data set into three parts: training, validation, and testing. In the case of ANN, the training set is used in the weight optimization process. With the validation set, we can monitor the model's behavior, and thus select the best set of weights. Finally, the test set is used to measure the performance of the model to solve the task (GÉRON, 2017).

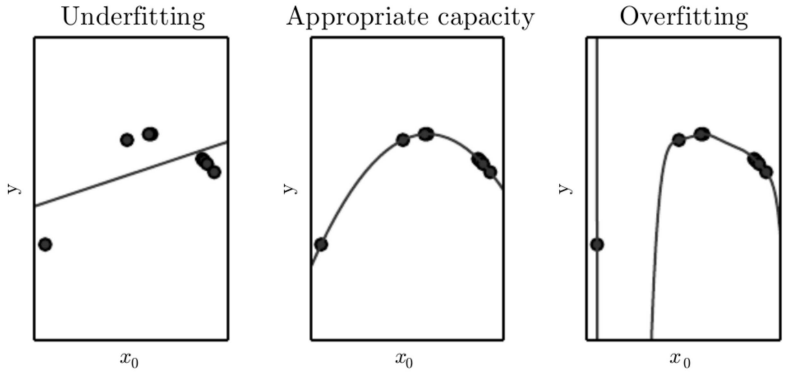
The monitoring of the model's behavior regarding the metrics obtained for the training and validation sets is essential to identify two challenges in ML: underfitting and overfitting (Figure 14). Underfitting occurs when the model is unable to obtain a low error for the task being trained in the training set. Overfitting, on the other hand, occurs when the gap between training error and validation is very large (GOODFELLOW; BENGIO; COURVILLE, 2017).

Some of the techniques that help to avoid overfitting or underfitting on CNNs are:

- **Dropout**: Dropout is a recent method for regularizing ML models (GOODFELLOW; BENGIO; COURVILLE, 2017). The main idea of a dropout layer is to disable a percentage of neurons during training randomly. In this way, each time a slightly different neural network is used (SRIVASTAVA et al., 2014). With neurons deactivated randomly, the models are not dependent on a neuron or connection, thus significantly reducing overfitting. In addition, dropout layer processing is computationally cheap (GOODFELLOW; BENGIO; COURVILLE, 2017).

- **Early Stopping**: the Early Stopping technique stops training before the model overfits. The interruption is performed based on a stop criterion on the loss function rates in the training and validation set. For example, if the loss function has not changed for some

Figure 14 – (Left) A linear function fit to the data suffers from underfitting — it cannot capture the curvature that is present in the data. (Center) A quadratic function fit to the data generalizes well to unseen points. It does not suffer from a significant amount of overfitting or underfitting. (Right) A polynomial of degree 9 fit to the data suffers from overfitting.



Source: Goodfellow, Bengio e Courville (2017)

time, it means that the algorithm has locked in a local minimum, and the additional adjustment of the weights is unnecessary and leads to a waste of computational processes and the excessive adjustment of the model (GOODFELLOW; BENGIO; COURVILLE, 2017).

- **Batch Normalization**: it is a layer that allows the network to converge more quickly (IOFFE; SZEGEDY, 2015). The process consists of adding a data normalization step between the layers of the network. Since H is a set of feature maps of the layer to be normalized, we have:

$$H' = \frac{H - \mu}{\sigma} \tag{2.13}$$

where μ is the average and σ is the standard deviation of the feature maps of the layer to be normalized. H' are normalized feature maps that are processed in the same way as if the network were processing H without normalization (GOODFELLOW; BENGIO; COURVILLE, 2017).

- **Transfer Learning**: when we start training an ANN, the initialization of the weights is usually done at random. For a CNN model, for example, to adjust these weights in order to obtain a generalization of the problem, thousands or millions of labeled examples are needed. However, getting these labeled and public data sets is not always easy, as in health data. One solution is to use the transfer learning technique. The technique consists of using trained weights for a A problem in the initialization of the network weights to solve the B problem. During the training stage, you can adjust the weights of all layers or even freeze some, avoiding their update (GOODFELLOW; BENGIO; COURVILLE, 2017). However, it is essential to note that this technique is useful in data sets that share

common characteristics (SHORTEN; KHOSHGOFTAAR, 2019).

- **Data Augmentation**: in contrast to the previously mentioned techniques that work on model characteristics, the data augmentation addresses the training data set (SHORTEN; KHOSHGOFTAAR, 2019). The process consists of generating synthetic data based on the original set. The operations generally applied to images are rotation, translation, horizontal and vertical flip, subsampling, distortions, among others. However, some attention is needed to avoid generating noise in the training dataset. For example, in the character recognition task, the letters 'b' and 'd' will have the same format if the image flips, so they are not operations to be considered to increase the data set (GOODFELLOW; BENGIO; COURVILLE, 2017). Another more recent technique for increasing data is the use of Generative Adversarial Networks (WEI et al., 2019).

2.4 Convolutional Neural Networks for Semantic Segmentation

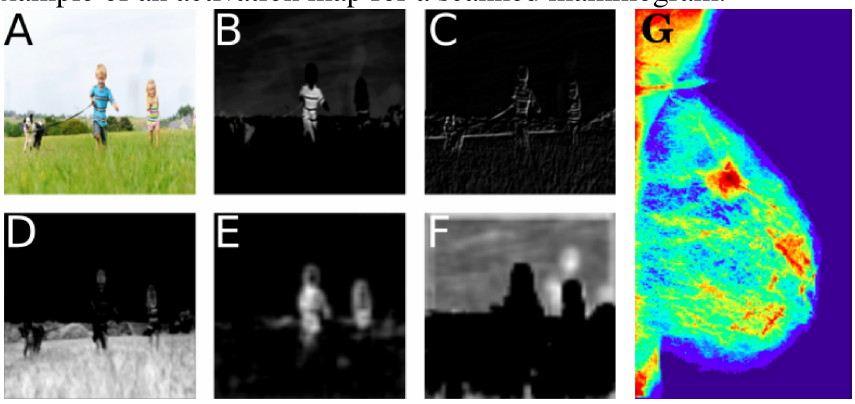
Semantic segmentation is a problem derived from image segmentation (GARCIA-GARCIA et al., 2017). In semantic segmentation, each pixel is classified into one of the predefined sets of classes, so those pixels belonging to the same class belong to a unique semantic entity in the image (GHOSH et al., 2019). Semantic segmentation is considered one of the most challenging problems in computer vision (GUO et al., 2018).

The oldest methods performed the semantic segmentation process using resource extraction, taking into account, for example, contours and borders (YU et al., 2018). However, these methods are dependent on the researcher's knowledge of the problem (SHEN; WU; SUK, 2017). With the development of DL, it was observed that after CNN training, the convolutional nuclei produced activation maps that referred to segmentation masks (Figure 15) (ZEILER; FERGUS, 2014). Most modern semantic segmentation methods use this property of CNNs to generate, in some way, the segmentation masks (GHOSH et al., 2019). In this section, we discuss the main techniques that use convolutional cameras to perform semantic segmentation.

2.4.1 Fully Convolutional Network

Based on the results obtained in the image classification Long, Shelhamer e Darrell (2015), they adapted classification networks to carry out the classification pixel by pixel. This work was one of the first to adopt a convolutional methodology. The methodology consisted of removing the fully connected layers from the AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012), VGG 16 (SIMONYAN; ZISSERMAN, 2014), and GoogLeNet (SZEGEDY et al., 2015) architectures by replacing them with a $1x1$ convolutional layer. This layer is commonly called deconvolution because it creates an output larger than the input (LONG; SHELHAMER; DARRELL, 2015). In addition, the authors added jump connections in the network to combine high-level resources with specific representations obtained in the last

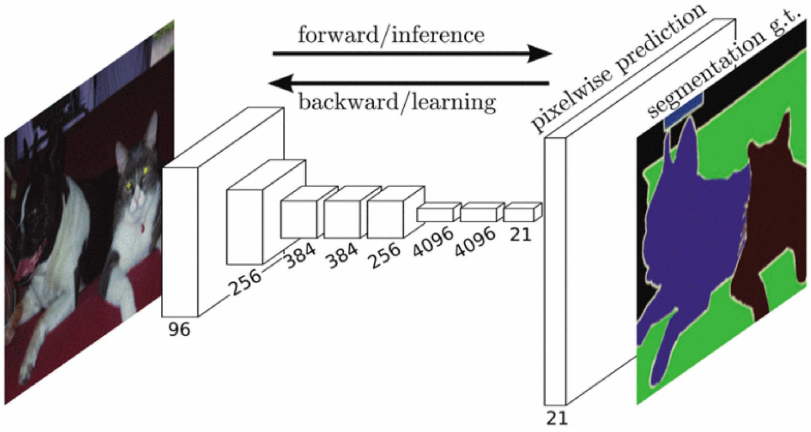
Figure 15 – (A) Input image. (B and C) Two activation maps of the first layers of a CNN, indicating the identification of lines, contours, and the child's shirt. (D, E, and F) Activation maps of the last layers, showing the identification of complex objects such as grass, people, and the sky. (G) Example of an activation map for a scanned mammogram.



Source: Adapted from Ghosh et al. (2019)

layers of the network. Figure 16 illustrates the architecture of the Fully Convolutional Network that achieved the best results for the VOC2012 test suite.

Figure 16 – Fully Convolutional Network

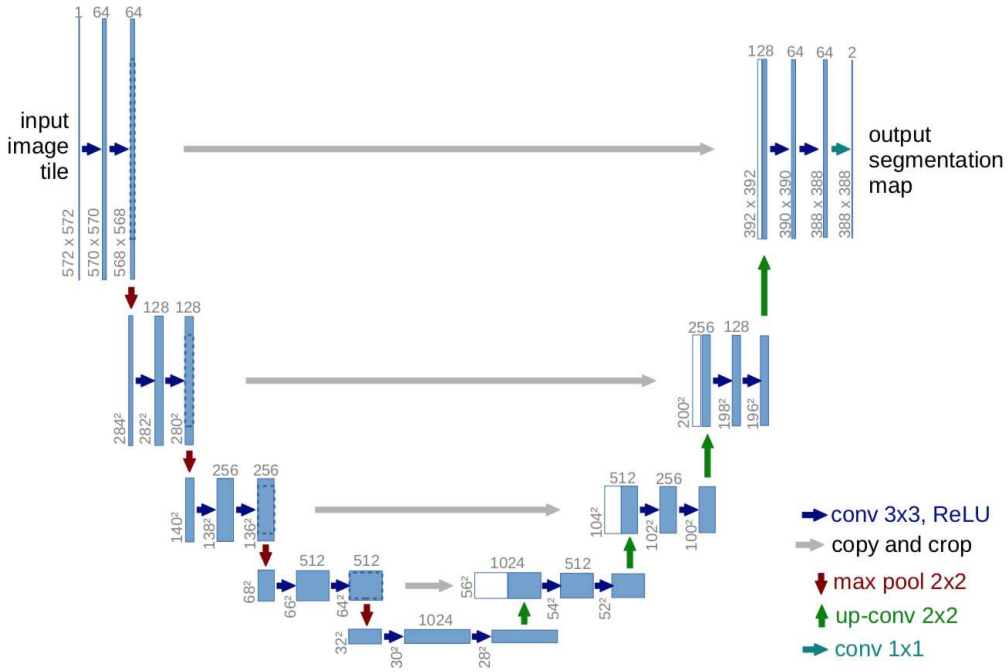


Source: Long, Shelhamer e Darrell (2015)

2.4.2 U-Net

The U-Net was designed especially for the segmentation of medical images (RONNEBERGER; FISCHER; BROX, 2015). Its name is due to the symmetrical shape of the model (Figure 17), which can be divided into two parts: downsampling and upsampling. Downsampling extracts features with two 3x3 convolutions and gradually reduces the spatial size of the images using the max-pooling operation. While upsampling reconstructs the images using the deconvolution and concatenation process with the characteristics extracted in the downsampling process. In the end, a 1x1 convolutional layer is used to map each resource map to the number of classes in the data set.

Figure 17 – U-Net

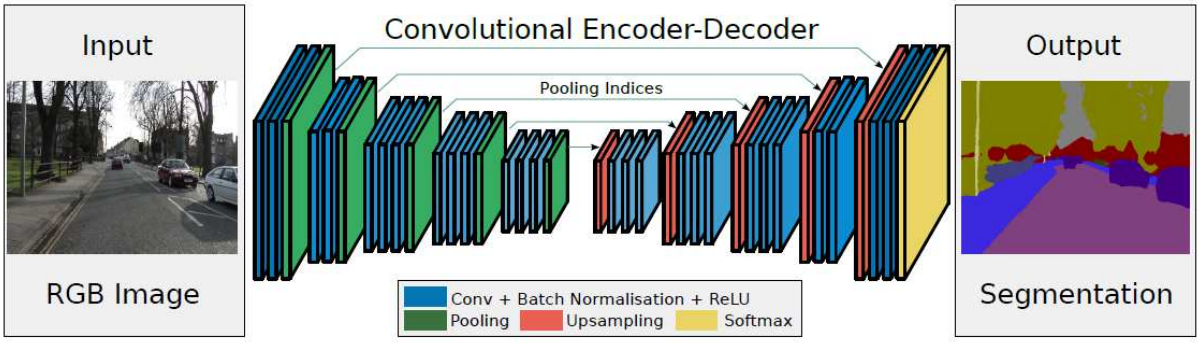


Source: Ronneberger, Fischer e Brox (2015)

2.4.3 SegNet

The SegNet network architecture is similar to U-Net; it also has a network division in downsampling and upsampling (BADRINARAYANAN; KENDALL; CIPOLLA, 2017). The difference is in the number of convolutional layers and the addition of batch normalization after each convolutional layer. SegNet's downsampling stage is topologically identical to the convolutional layers of VGG16 (SIMONYAN; ZISSERMAN, 2014). The fully convolutional layers of the VGG16 network were removed, and replaced by deconvolutional layers, corresponding to the convolutional layers of the downsampling stage. An illustration of the SegNet architecture is shown in Figure 18.

Figure 18 – SegNet

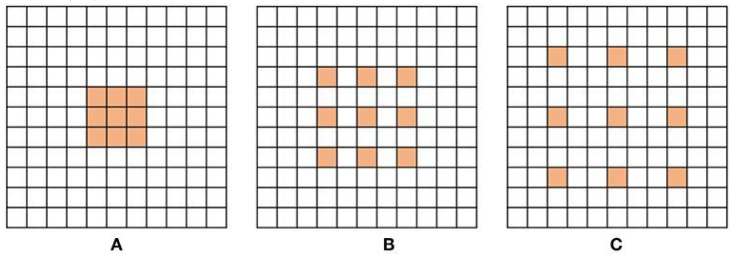


Source: Badrinarayanan, Kendall e Cipolla (2017)

2.4.4 Dilated Residual Network

The global information of the images in the last layers of the convolutional architectures that use pooling is inferior. Pooling layers are useful for eliminating redundant information and speeding up training. However, this causes the loss of some global details (JEONG et al., 2019). The extended convolution helps to capture global information in a larger region without reducing the resolution (YU; KOLTUN; FUNKHOUSER, 2017). This process is due to the way in which the convolution process occurs, which starts to be performed in a spaced manner according to some specified expansion rate; an example is shown in Figure 19.

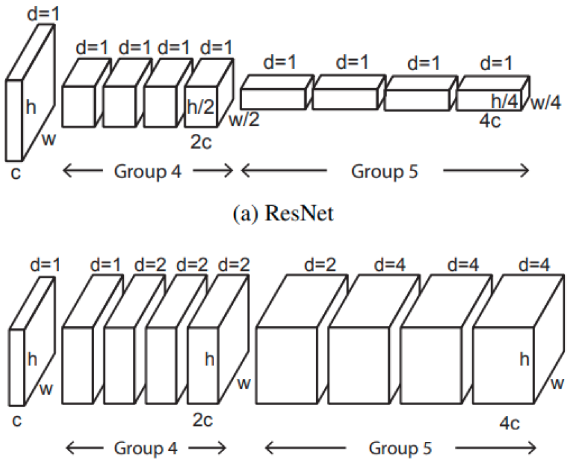
Figure 19 – Examples of dilated convolution filter with 3x3 size. (A) Dilation factor = 1, (B) Dilation factor = 2 and (C) Dilation factor = 3.



Source: Jeong et al. (2019)

The Dilated Residual Network (DRN) is an architecture that uses the dilated convolution (YU; KOLTUN; FUNKHOUSER, 2017). In this network, the authors adapted the ResNet architecture (HE et al., 2016), replacing certain convolutional layers with dilated convolutional layers. Figure 20 illustrates the process of changing part of ResNet to DRN.

Figure 20 – Adaptation of the ResNet architecture to DRN. Where d is the expansion factor, h and w are the height and width of the maps, and c is the number of characteristic maps.



Source: Jeong et al. (2019)

2.5 Final Remarks

This chapter presented the main concepts involved in the proposal for an intelligent breast cancer WSI analysis system, characterizing the development of breast cancer and how it is identified by histological analysis. The chapter also includes techniques considered state of the art for digital image processing and pattern identification in images.

CAD systems meet the need to provide means of early detection of breast cancer (VETA et al., 2014). In this sense, for intelligent analysis of WSI, some characteristics make the process challenging. Unlike most natural images, WSI are images with billions and sometimes trillions of pixels. These pixels transmit essential information for the detection of breast cancer. And the same challenges faced by the pathologist in the analysis of the slides are transferred to CAD systems. Thus, the use of PDI in conjunction with AI techniques is necessary to achieve results comparable to state of the art in the detection of breast cancer in WSI. In the next chapter, we analyze, using an SLR, the works considered state of the art to investigate the current methods, challenges, and research opportunities that guided this Master Thesis.

3 RELATED WORK

This chapter aims to present studies and work from different authors, related to our proposed thematic and model. This SLR is divided into three main parts, methodology, results and discussion and challenges, and future directions. The methodology describes the procedures adopted for the selection of studies. In the discussion, we present an analysis of the selected studies, proposing a taxonomy and identifying datasets, pre-processing methods, and intelligent techniques used in the processing of breast cancer WSI. Finally, we identified challenges and open gaps in the literature, indicating future research directions.

3.1 Methodology

In this section, we describe the protocol adopted for performing the SLR, presenting the research questions, the search and selection strategy, quality criteria, and how we obtain the information to answer our questions. The main advantage of an SLR is the identification, evaluation, and interpretation of works with methodology and defined results for a given research question without the bias of a traditional literature review (KITCHENHAM; CHARTERS, 2007).

3.1.1 Study design

This study aims to map relevant work and identify the gaps in the intelligent analysis of breast cancer histopathological images for future research. The SLR steps are defined below, based on methodologies used in previous work (ROEHRS et al., 2017; COVER et al., 2018; MONTENEGRO; COSTA; RIGHI, 2019; GUBERT; COSTA; RIGHI, 2019; ARJI et al., 2019).

- Research questions: introduce the research questions investigated;
- Search strategy: outline the strategy and libraries explored to collect data;
- Article selection: explain the criteria for selecting the studies;
- Quality assessment: describe the quality assessment applied to the selected studies;
- Data extraction: compare the selected studies and research questions.

3.1.2 Research Questions

Defining research questions is the fundamental part of a SLR (KITCHENHAM; CHARTERS, 2007; COVER et al., 2018) because, based on these questions, we get evidence to define, apply, and gain knowledge in the intelligent analysis of breast cancer images. Thus,

we organized our questions into two groups: General Question (GQ) and Specific Question (SQ). The Table 2 lists all the research questions investigated.

Table 2 – Research questions.

Group and identifier	Issue
	General questions (GQ)
GQ1	What is the state of the art in intelligent methods to aid in the diagnosis of breast cancer, based on histopathological images?
GQ2	What is the taxonomy for intelligent detection of breast cancer in histopathological images?
GQ3	What are the main challenges in this area?
	Specific questions (SQ)
SQ1	What public datasets are used?
SQ2	What are the preprocessing techniques used?
SQ3	What machine learning techniques are used for breast cancer detection?
SQ4	What are the metrics for the validation of the work?

Source: Elaborated by the author.

3.1.3 Search strategy

To identify the answers to the previously proposed questions, we need to select a set of relevant studies. So the next step is to define keywords and search scope (ROEHRS et al., 2017; COVER et al., 2018). The keywords were determined based on the authors' experience, reading articles of the area of interest, and final validation by an experienced pathologist. We then grouped the keywords by boolean operators correlating synonyms, acronyms, and words inserted in the same context (KITCHENHAM; CHARTERS, 2007). Figure 21 presents the search string used in the selected bases.

Figure 21 – Search string used for database queries.

Search String

```
("breast cancer" OR "breast carcinoma") ("histopathology"
OR "histopathological") ("medical image" OR "biomedical
image" ) ("detect" OR "classify" OR "segment") ("machine
learning" OR "deep learning" OR "artificial intelligence")
```

Source: Elaborated by the author.

To search for articles, we selected the following databases: ACM Digital Library, IEEE Xplore Library, PubMed, Google Scholar, ScienceDirect, Scopus, SpringerLink, and Web of

Science. We based our choice on the significant sample indexing of relevant full-text journals and conference proceedings in the health field involving the detection of breast cancer in histopathological images. Given the specifications of each base, we adapt our string to the rules provided by them.

3.1.4 Article selection

After the base searches, we applied some exclusion criteria to keep only the most representative articles for our search problem. First, we performed an impurity removal, removing duplicate papers, and non-English studies. In the next steps, we performed the analysis of the titles and abstracts of the articles, removing works that did not meet our criteria. The criteria are based on our search string and are defined as follows:

- Exclusion criteria 1: articles that are not explicitly applied to histopathological images of breast cancer;

- Exclusion criteria 2: articles that do not use intelligent methods to perform image analysis.

After this step, we removed theses, books, and articles shorter than four pages, focusing only on scientific articles from journals and conferences. The final process consisted of a complete analysis of the works, selecting only the papers that presented architectural proposals and methods matching our search string.

3.1.5 Quality assessment

For a SLR, the quality of the corpus is essential. Quality criteria assist in the selection of relevant works (KITCHENHAM; CHARTERS, 2007; COVER et al., 2018). In our work, we defined some indexes to measure the quality of the selected studies. The h5-index is a way of quantifying the relevance of journals and conferences in the last five years, based on citations of articles published in the journal (MONTENEGRO; COSTA; RIGHI, 2019). We defined a cutoff for the studies in this SLR, considering only articles published in journals and conferences with h5-index greater than or equal to 10. We also use the SJR indexes for journals and CORE for conferences. However, the CORE index does not index some conferences. Thus, we chose to keep the studies as long as they reach the minimum score defined in the h5-index. Maintaining these studies is justified by the relevance of the works for the discussion of research questions. In addition, some specific conferences are relatively recent and have not yet been indexed by the index.

3.1.6 Data extraction

In this section, we describe (Table 3) the correlation between our questions and the sections in which we expect to find relevant information to answer them.

Table 3 – Quality Assessment of article structure and related questions.

Section	Description	Research Questions
Open Content		
	Title	GQ1, GQ2, SQ1
	Abstract	All questions
	Keywords	GQ1, SQ1
Article Content		
	Introduction	All questions
	Method	All questions
	Results	All questions
	Discussion	All questions
	Conclusion	All questions

Source: Elaborated by the author.

3.2 Results and discussion

In this section, we present the results and discussions of SLR, based on the research questions defined above.

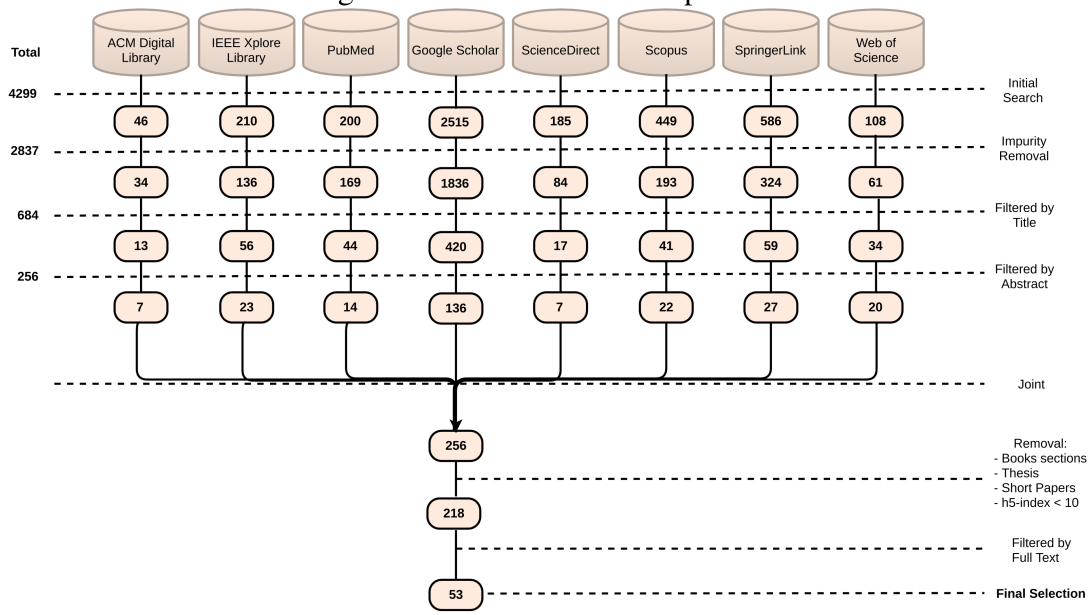
3.2.1 Article selection

We use the methodology mentioned above in the process of selecting articles from each database to compose our corpus. We also limit our selection by year of publication, selecting only studies from the last ten years and excluding patents, citations, and articles published in journals or conferences with h5-index less than 10.

The process of selecting corpus articles is detailed in Figure 22. Our search in the databases listed above and according to the proposed string returned a total of 4,299 articles. After the first filtering, removing impurities remained 2,837 (65.99 %) articles. In the second stage, the selection was based on the title of the articles, removing studies that were not related to intelligent breast cancer analysis techniques, remaining 684 (24.10 %) articles. Thus, we performed the complete reading of the abstract, with a total of 256 (37.42 %) studies remaining after this stage. The next step was the removal of sections of books, thesis, and articles with less than four pages, leaving 218 (85.15 %) studies. Finally, we performed a complete analysis of the articles, taking into account our previously defined criteria in the Section 3.1.4 and Section 3.1.6, reaching a final corpus of 53 (24.31 %) articles. We divided

our corpus into primary studies published in Journals, Table 4, and in Conferences, Table 5, sorted in descending order by year of publication.

Figure 22 – Article selection process.



Source: Elaborated by the author.

Table 4: Final corpus of articles published in journals.

References	Publisher	h5	SJR
(GUO et al., 2019)	Nature	151	1,414
(VO; NGUYEN; LEE, 2019)	Elsevier	94	1,62
(XIE et al., 2019)	Frontiers	59	1,888
(LI; WU; WU, 2019)	IEEE	57	0,609
(MURTAZA et al., 2019)	Springer	45	0,335
(DAS; DUTTA, 2019)	Elsevier	39	0,57
(MITTAL; SARASWAT, 2019)	Elsevier	32	1,278
(ALOM et al., 2019)	Springer	29	0,666
(ROY et al., 2019)	Elsevier	29	0,805
(GHAZVINIAN ZANJANI et al., 2019)	SPIE	16	0,6
(Sabeena Beevi; NAIR; BINDU, 2019)	Elsevier	16	0,438
(ALQUDAH; ALQUDAH, 2019)	Taylor & Francis	15	0,207
(CRUZ-ROA et al., 2018)	PLOS	180	1,1
(OLIVEIRA et al., 2018)	Elsevier	92	1,19

Continued on next page

Table 4 – continued from previous page

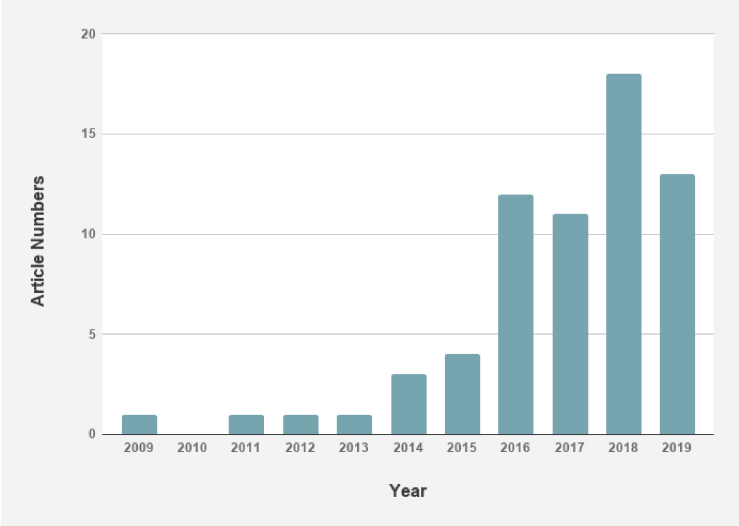
References	Publisher	h5	SJR
(GECER et al., 2018)	Elsevier	74	1,363
(ZHENG et al., 2018)	IEEE	66	2,188
(LI et al., 2018)	Elsevier	57	2,452
(BARDOU; ZHANG; AHMAD, 2018)	IEEE	57	0,609
(FONDÓN et al., 2018)	Elsevier	39	0,57
(GANDOMKAR; BRENNAN; MELLO-THOMS, 2018b)	Elsevier	32	1,025
(SAHA; CHAKRABORTY; RACOCEANU, 2018)	Elsevier	29	0,805
(FENG; ZHANG; YI, 2018)	Springer	28	0,625
(PAN et al., 2018)	Elsevier	27	0,402
(HAN et al., 2017)	Nature	151	1,414
(ZHENG et al., 2017)	Elsevier	74	1,363
(WAN et al., 2017b)	Elsevier	71	0,996
(PAN et al., 2017)	Elsevier	71	0,996
(WAN et al., 2017a)	Elsevier	71	0,996
(WAHAB; KHAN; LEE, 2017)	Elsevier	39	0,57
(BEJNORDI et al., 2017b)	SPIE	16	0,6
(KRAWCZYK et al., 2016)	Elsevier	77	1,216
(XU et al., 2016)	IEEE	66	2,188
(ZHANG et al., 2016)	IEEE	56	1,122
(CHAN; TUSZYNSKI, 2016)	The Royal Soc.	32	1,131
(BALAZSI et al., 2016)	SPIE	16	0,6
(KORKMAZ; KORKMAZ; POYRAZ, 2016)	Springer	11	0,61
(TASHK et al., 2015)	Elsevier	67	0,873
(ZHANG et al., 2015)	IEEE	66	2,188
(KHAN; SIRINUKUNWATTANA; RAJPOOT, 2015)	IEEE	56	1,122
(WANG et al., 2014)	SPIE	16	0,6
(LOUKAS et al., 2013)	Hindawi	31	0,414
(ISSAC NIWAS et al., 2012)	Springer	45	0,565
(DUNDAR et al., 2011)	IEEE	69	1,256

Table 5: Final corpus of articles published in conferences.

References	Publisher	h5	CORE
(YAN et al., 2019)	IEEE	16	-
(GUPTA; BHAVSAR, 2018)	IEEE	47	-
(BAKER et al., 2018)	IEEE	12	C
(GANDOMKAR; BRENNAN; MELLO-THOMS, 2018a)	SPIE	10	-
(SPANHOL et al., 2017)	IEEE	23	-
(SPANHOL et al., 2016a)	IEEE	32	A
(BAYRAMOGLU; KANNALA; HEIKKILA, 2016)	IEEE	28	B
(LI; PLATANIOTIS, 2015)	SPIE	13	-
(CRUZ-ROA et al., 2014)	SPIE	13	-
(JELEN; FEVENS; KRZYZAK, 2009)	SPIE	14	-

In Figure 23, we present an evolution of the publications of the corpus in the last decade. Analyzing the distribution over time, we can see that the amount of primary studies is practically all located in the previous five years, indicating the relevance of research in this area today.

Figure 23 – Publication chronology. The numbers above years indicate the number of articles published.



Source: Elaborated by the author.

This distribution is related to higher processing, storage, and, above all, data transmission capacities that impacted the interest of several research groups in improving the intelligent algorithms (GHAZVINIAN ZANJANI et al., 2019). In the context of intelligent analysis of

WSI, these advances have allowed the analysis and manipulation of images even by personal computers. In addition, the realization of challenges and the availability of public datasets contribute to spread interest in the problem and to improve the results in the automatic analysis of WSI for breast cancer.

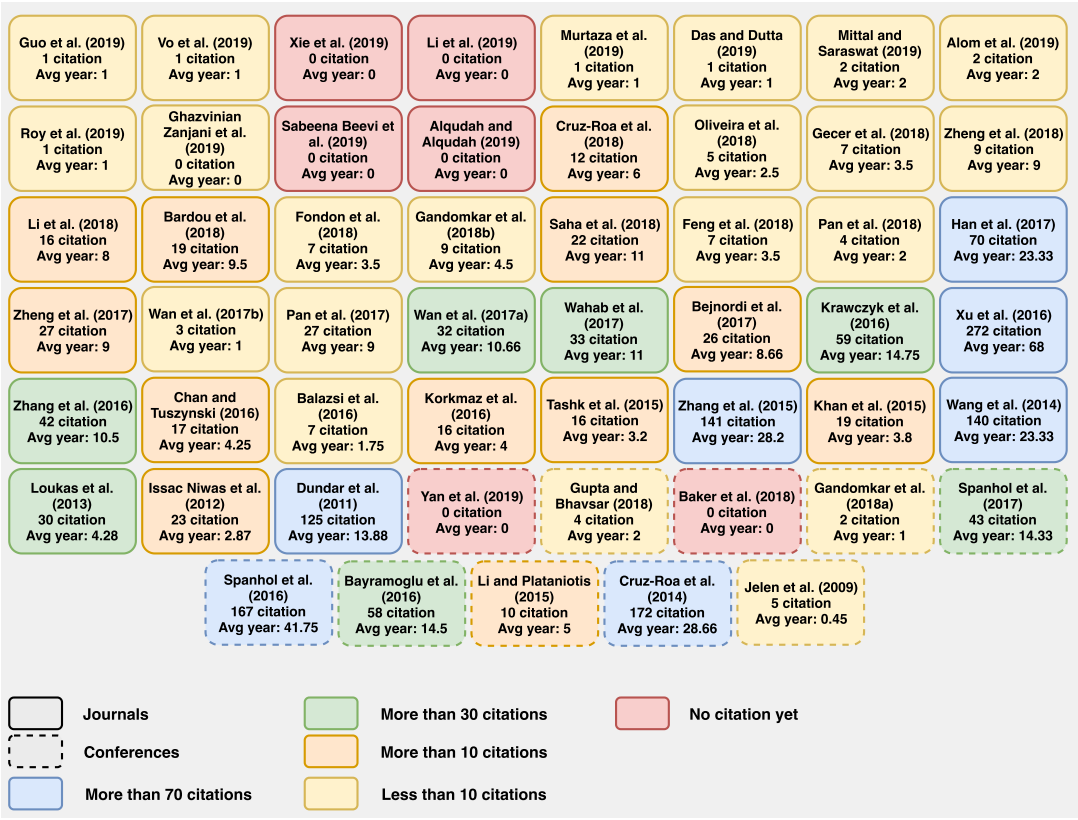
3.2.2 Data Extraction and Answers to the Research Questions

Based on the 53 articles selected during SLR, we discussed the main challenges, models, techniques, and objectives of the studies to answer research questions in the context of intelligent histopathological analysis of breast cancer imaging.

GQ1 What is the state of the art in intelligent methods to aid in the diagnosis of breast cancer, based on histopathological images?

As a way to answer the question GQ1, we present in Figure 24 the number of citations of the corpus, based on the values given by Google Scholar. Also, an average of citations per year is presented. With this analysis, it is possible to measure the relevance of each study to the current literature, and consequently identify the most significant methods to conduct future work.

Figure 24 – State of the Art articles considered in this survey with number of citations and average per year. Articles are organized in order of publication and classification by h5-index. Information about Journals articles is represented by solid lines, while dashed lines represent details from conferences.



Source: Elaborated by the author.

According to Figure 24, studies without citations or with few citations are concentrated in the last two years. It is also possible to notice that the articles with the highest citation rates and average citation per year are located in the last five years.

In the studies analyzed, we can highlight the use of deep learning techniques, such as CNN. Even with the scarcity of available data, CNN can achieve the best results in the analyzed literature for binary classification, multi-class classification of patches (regions of interest extracted from WSI or directly from histopathological slides), or the segmentation of WSI (SPANHOL et al., 2016a; HAN et al., 2017; GUO et al., 2019). Another highlight is the use of CNN in conjunction with the representation of image resources (morphology, color, and texture features) to reduce the cost of computational processing (WANG et al., 2014).

CNN can also be used as extractors of characteristics for the classification of breast cancer lesions (ZHENG et al., 2017). Alternatively, feature extraction techniques were used to classify intraductal lesions (DUNDAR et al., 2011). In the specific questions, we present in greater detail all the methods selected for the corpus of the article.

GQ2 What is the taxonomy for intelligent detection of breast cancer in histopathological images?

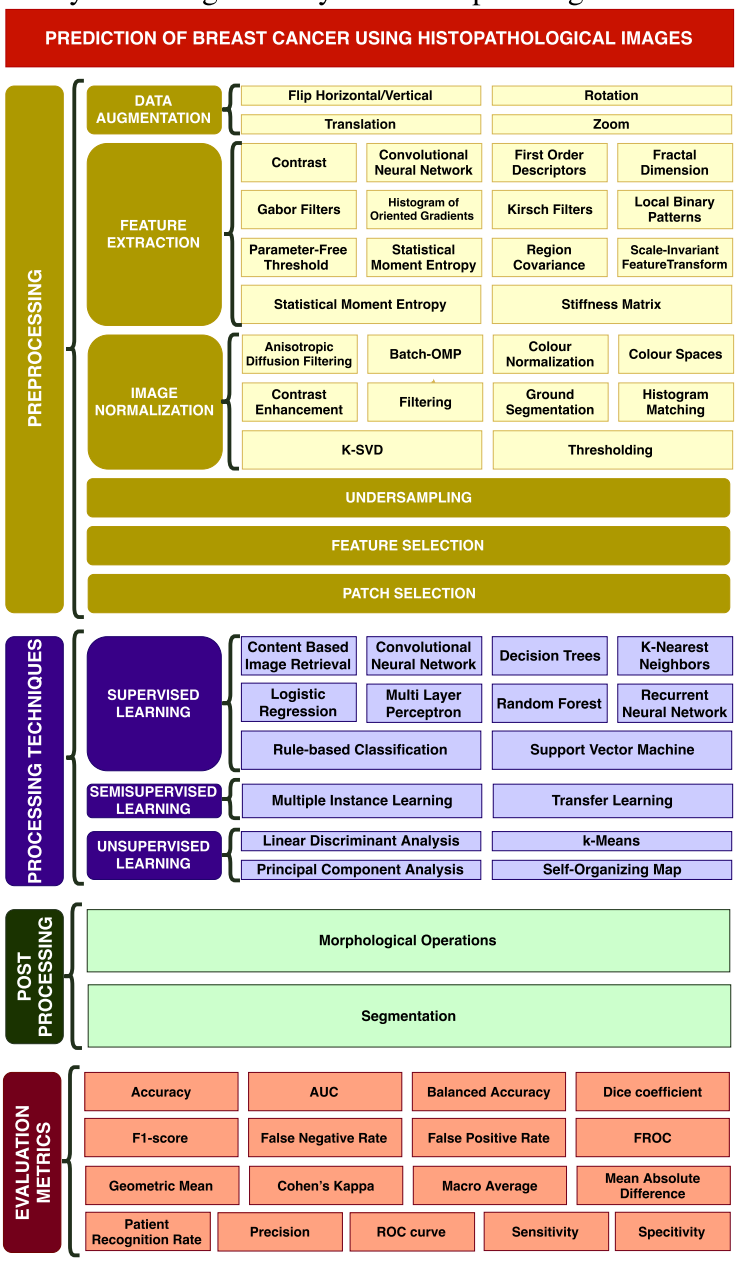
With the analysis of the corpus studies, we defined a taxonomy as a way to classify the intelligent analysis of histopathological images of breast cancer. This taxonomy is illustrated in Figure 25, is organized into classes that are based on the technologies, methods, and problems identified in the studies. Articles generally combine multiple classes, so the classes are not mutually exclusive. In the process of analyzing the studies, we defined four main categories: (1) Preprocessing techniques, (2) Processing techniques, (3) Post-processing techniques, and (4) Evaluation metrics. State of the art in the field of intelligent analysis of histopathological images has been grouped into the classes corresponding to the steps generally performed in a CAD system (Preprocessing techniques, processing techniques, post-processing techniques) and the evaluation metrics used in the analyzed studies (PEDRO; MACHADO-LIMA; NUNES, 2019).

In the first class, we present the preprocessing techniques that seek to assist in the learning process. Preprocessing techniques for analyzing histopathological images are applied to remove background images (GUO et al., 2019), color normalization (KHAN; SIRINUKUNWATTANA; RAJPOOT, 2015; LI; PLATANIOTIS, 2015; GANDOMKAR; BRENNAN; MELLO-THOMS, 2018a) or to enlarge the dataset (KRAWCZYK et al., 2016; GUO et al., 2019).

In the second class, we define the types of learning techniques that are used for breast cancer histopathological image analysis. Using intelligent techniques, jobs can perform classification (BARDOU; ZHANG; AHMAD, 2018), segmentation (PAN et al., 2017), or detection tasks (GECER et al., 2018), and the data during the training step can be labeled or not.

In the third class are defined techniques that seek to improve the quality of results after learning processes based on histopathological images. The goal with the use of post-processing techniques is mainly noise removal to improve outcomes. For segmentation tasks, for example,

Figure 25 – Taxonomy of intelligent analysis of histopathological data from breast cancer.



Source: Elaborated by the author.

we can remove isolated pixels using the morphological cleaning and opening operation (PAN et al., 2017). Finally, the fourth class presents the metrics used to evaluate intelligent methods.

After each class, the taxonomy presents the corpus attributes related to each central concept. This breakdown is presented later in this article when we answer the specific questions.

GQ3 What are the main challenges in this area?

To answer GQ3, we list and identify the challenges found according to the corpus studies. The problems were then summarized and grouped into two groups, as shown in Table 6. Based on this grouping, we selected some articles and discussed the challenges related to intelligent analysis of breast cancer histopathological images.

Table 6 – Challenges regarding the intelligent analysis of breast cancer histopathological images.

Field	Challenges	Articles
Computer Science	Clinical Acceptance	(BALAZSI et al., 2016)
	Different Protocols of Analysis and Stained	(LOUKAS et al., 2013; KORKMAZ; KORKMAZ; POYRAZ, 2016; LI; PLATANIOTIS, 2015; KRAWCZYK et al., 2016; FENG; ZHANG; YI, 2018; YAN et al., 2019; VO; NGUYEN; LEE, 2019; XIE et al., 2019; ROY et al., 2019)
	Few Labeled Images	(XU et al., 2016; KRAWCZYK et al., 2016; SPANHOL et al., 2017; BEJNORDI et al., 2017b; FENG; ZHANG; YI, 2018; OLIVEIRA et al., 2018; GUO et al., 2019; VO; NGUYEN; LEE, 2019)
	Ignorance of Patient's Medical History	(ISSAC NIWAS et al., 2012)
	Images Size	(CRUZ-ROA et al., 2014; SPANHOL et al., 2016a; ZHENG et al., 2017; WAN et al., 2017b; PAN et al., 2018; GUO et al., 2019; YAN et al., 2019; GHAZVINIAN ZANJANI et al., 2019)
	Image Processing Time	(TASHK et al., 2015; MITTAL; SARASWAT, 2019)
Pathology	Different Magnification Factors	(ZHANG et al., 2016; BAYRAMOGLU; KANNALA; HEIKKILA, 2016; GANDOMKAR; BRENNAN; MELLO-THOMS, 2018a; CRUZ-ROA et al., 2018; ALOM et al., 2019)
	Disagreement in Diagnosis	(DUNDAR et al., 2011; CHAN; TUSZYNSKI, 2016; HAN et al., 2017; WAN et al., 2017b; GANDOMKAR; BRENNAN; MELLO-THOMS, 2018a; GANDOMKAR; BRENNAN; MELLO-THOMS, 2018b; BARDOU; ZHANG; AHMAD, 2018; MURTAZA et al., 2019)
	Tissue Overlap and Structure Similarity	(JELEN; FEVENS; KRZYZAK, 2009; WANG et al., 2014; ZHANG et al., 2015; KHAN; SIRINUKUNWATTANA; RAJPOOT, 2015; WAHAB; KHAN; LEE, 2017; HAN et al., 2017; PAN et al., 2017; FENG; ZHANG; YI, 2018; GECER et al., 2018; LI et al., 2018; BAKER et al., 2018; PAN et al., 2018; SAHA; CHAKRABORTY; RACOCEANU, 2018; ZHENG et al., 2018; FONDÓN et al., 2018; GUPTA; BHAVSAR, 2018; MITTAL; SARASWAT, 2019; Sabeena Beevi; NAIR; BINDU, 2019; VO; NGUYEN; LEE, 2019; LI; WU; WU, 2019; XIE et al., 2019; DAS; DUTTA, 2019; ALQUDAH; ALQUDAH, 2019)

Source: Elaborated by the author.

The processes of preparation, staining, and digitization of tissue slides carried out by pathologists may suffer from inadequate fixations. These problems can generate changes in tissue morphology and induce incorrect tissue morphometry by analyzing images (ROY et al., 2019). Another challenge are sections that have not been adequately stained (these may be over- or under-stained) and assembled, this may result in regions out of focus or missing parts (LI et al., 2018). In addition, slight variations in coloring conditions can lead to changes in the appearance of the tissue (YAN et al., 2019). Also, the subtle differences between multiple classes and high coherence of cancer cells ultimately lead to heterogeneity in the color distribution in images (HAN et al., 2017; MITTAL; SARASWAT, 2019).

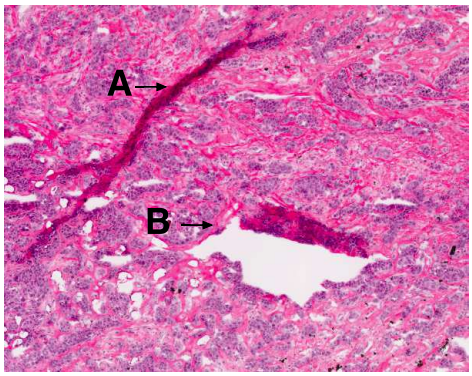
Due to the above resources, there is a shortage of pathologists for the analysis of slides in H&E, mainly due to the difficulty of inheriting or innovating the necessary experience and expertise (HAN et al., 2017). Due to the lack of even primary care, the process of labeling WSI or sections thereof is an expensive and time-consuming process, reflecting the scarcity of public data on the available histopathological images (KRAWCZYK et al., 2016; GECER et al., 2018; GUO et al., 2019; YAN et al., 2019). According to (FENG; ZHANG; YI, 2018), published breast cancer data sets are generally involved on a small scale and without expert notes. Another problem for the creation of several sets of biomedical images is the impossibility of using crowdsourcing methods, as in natural images, precisely due to experience and expertise to interpret them (YAN et al., 2019). In addition, it is common for the labeling process to suffer inter and intraobserver variations (MITTAL; SARASWAT, 2019).

Technical and computational problems are common when discussing intelligent analysis techniques. In the histopathological images, the main problem in the computational and technical scope is the size of the WSI. A WSI, for example, cannot be used directly on a neural network because it contains billions of pixels, in the range of $100Kx100K$ pixels (GECER et al., 2018; GUO et al., 2019; YAN et al., 2019). Associated with this, another limitation is the different nucleus shapes, sophisticated tissue appearance, tissue overlap, and texture difference of the tissues that make up the image (Figure 26) (HAN et al., 2017; PAN et al., 2018; FENG; ZHANG; YI, 2018; MITTAL; SARASWAT, 2019). These subtle differences in images are directly aligned with the number of breast cancer types (FENG; ZHANG; YI, 2018). These characteristics, coupled with the limited publicly available data and the differences between the occurrence of cancer types, cause a class imbalance in histopathological images of breast cancer, making it a challenge to apply Machine Learning (KRAWCZYK et al., 2016).

Motivated by the factors mentioned, selecting the magnification factor for intelligent analysis is another challenge (ZHANG et al., 2016). The study of (GANDOMKAR; BRENNAN; MELLO-THOMS, 2018b) made a comparison for the classification at first for benign/malignant and after for eight histological types of breast cancer. The best result for binary classification was the $40\times$ magnification factor, while for histological types of breast cancer, the best result was obtained at $200\times$ magnification.

Finally, there is a need for studies that may cover the patient's medical history (ISSAC

Figure 26 – Example of H&E stained tissue section. A) tissue overlap; B) Image artifact due to tissue rupture.



Source: Elaborated by the author.

NIWAS et al., 2012). Interesting research along these lines is the fusion of mammographic data with histopathological images (KORKMAZ; KORKMAZ; POYRAZ, 2016). Allied to this, the studies proposed so far do not verify the real clinical acceptability of the methods and how they fit the pathologist's routine.

SQ1 What public datasets are used?

The purpose of SQ1 is to identify publicly available datasets and their characteristics used by articles reviewed by SLR. In Table 7 we list all public datasets with their main features.

BreakHis is a set of images collected from histopathological slides of partial mastectomy or excisional biopsy from 82 patients, with manual selection of significant regions by pathologists (SPANHOL et al., 2016b). Already Bioimaging 2015 is proposed for use in the challenge of a congress. After the event, the data were made available for use and comparison with the results obtained in the contest. The collection was similar to BreakHis, with images labeled and selected by two pathologists (ARAUJO et al., 2017). Camelyon16 and Camelyon17 have also emerged for use in a challenge, but unlike the previous ones, they provide WSI, thus allowing complete tissue analysis. In addition, they provide the boundaries of the abnormalities present in each WSI and the case diagnosis (BEJNORDI et al., 2017a).

UCSB Breast Cancer Cell is a smaller dataset of benign and malignant breast cancer regions with the ground truth of the nuclei present in the patches (GELASCA et al., 2008). Mitos-Atypia-14 is the extension of the Mitos-ICPR12 dataset, providing information on the presence of mitosis and degree of nuclear atypia in H&E stained patches of five patients labeled by two senior pathologists and three junior pathologists (ROUX et al., 2014).

Table 7 – Public data sets of histopathological images of breast cancer identified in the review.

Dataset	Field of Study	Image type	Number of images	Magnifications	Patients	Image size (pixels)	Annotations	Image format
BREAKHIS	adenosis, fibroadenoma, phyllodes tumor, tubular adenona, carcinoma, lobular carcinoma, mucinous carcinoma and papillary carcinoma.	Patches extracted from WSI	7,909	40, 100, 200, 400	82	700X460	Patch-level notes	PNG
BIOIMAGING 2015	normal, benign,in situ carcinoma and invasive carcinoma	Sections of slides	269	200	-	2040X1536	Image-level notes	TIFF
CAMELYON16	Normal and metastases	WSI of lymph node sections	399	1, 10, 40	399	-	Pixel level annotation	TIFF
MITOS-ATYPIA-14	Nuclear Atypias and Mitosis	Sections of slides	1420	20, 40	-	1539X1376 and 1663X1485	Mitosis annotation or atypia at image level	TIFF
ICIAR 2018	normal, benign, in situ carcinoma, invasive carcinoma	WSI of breast cancer and patches extracted from WSI	400	-	-	-	Pixel level and image level annotations	SVS
CAMELYON17	Normal and metastases	WSI of lymph node sections	1000	-	-	-	Pixel level annotation	TIFF
TCGA	ductal and lobular neoplasms, epithelial neoplasms, complex epithelial neoplasms, adenomas and adenocarcinomas, cystic and mucinous neoplasms, squamous cell neoplasms, fibroepithelial neoplasms, adnexal and skin appendage, basal cell neoplasms,mature b-cell lymphomas	Various	-	-	3.816	-	Genomic, epigenomic, transcriptomic, and proteomic data	Various formats
UCSB Breast Cancer Cell	normal and malignant	No	58	-	-	896X768 768X512	Cell segmentation	TIFF
TUPAC16	Mitosis	WSI of breast cancer	821	-	-	$\approx$ 50000X50000	Proliferation score based on mitosis counting, molecular proliferation score, regions of interest for mitosis, and annotation of some regions of mitosis	SVS

Source: Elaborated by the author.

ICIAR 2018 is an extension of the Bioimaging 2015 dataset, with the addition of WSI providing the delimitation of lesions in each WSI in four classes: normal, benign, in situ carcinoma, invasive carcinoma (ARESTA et al., 2019). The Cancer Genome Atlas Program (TCGA) is a regularly updated project coordinated by the National Cancer Institute, currently with information on 33 different types of tumors from over 11,000 patients. The TCGA provides the patient-level diagnosis, and WSI of the biopsy performed and, in some cases, the patient outcome (National Cancer Institute, 2019). Tumor Proliferation Assessment Challenge 16 (TUPAC16) is the result of WSI analysis from TCGA by a group of 15 pathologists. The dataset provides information on mitotic and proliferation scores based on WSI gene expression (VETA et al., 2019).

SQ2 What are the preprocessing techniques used?

Another result of SLR is the identification of the methods used for image preprocessing. Supporting this issue, we discuss some challenges raised by selected articles related to the objective of question SQ3, focusing on macro challenges presented on Table 8.

Table 8 – Main preprocessing techniques applied to intelligent analysis of breast cancer histopathological images.

Technique	**Goal**
Data Augmentation	Technique for generating synthetic dataset-based images to increase and balance dataset
Feature Extraction	It aims to represent an image based on significant characteristics for the problem
Feature Selection	Techniques applied to get the most relevant information in the learning process
Image Normalization	Set of techniques applied to noise reduction, contrast and reduce the difference between color of histopathological image of breast cancer
Patch Selection	Selection of sub-images of significant regions in images to increase the number of training examples, or decrease the computational cost during learning
Undersampling	Deleting images from the learning process to balance the dataset

Source: Elaborated by the author.

In histopathological images, the tissue sample does not completely cover the image, unlike natural images, remaining a non-woven background in the image (KHAN; SIRINUKUNWATTANA; RAJPOOT, 2015; ZHENG et al., 2017). Thus, removing these regions from the image is done in some works to speed up the training process or improve the feature extraction process (GUO et al., 2019).

In addition, color variability because of the use of different scanners and cameras, and the staining processing carried out at the different laboratories is a problem for feature extraction and training of intelligent techniques (KHAN; SIRINUKUNWATTANA; RAJPOOT, 2015; LI;

PLATANIOTIS, 2015; GANDOMKAR; BRENNAN; MELLO-THOMS, 2018b). Thus using the geodetic Singular Value Decomposition (SVD) method (YAN et al., 2019; DAS; DUTTA, 2019), color deconvolution (WAN et al., 2017a; Sabeena Beevi; NAIR; BINDU, 2019), methods based on color space transformation (LI; WU; WU, 2019), contrast enhancement (BAKER et al., 2018) and mean normalization and standard deviation (PAN et al., 2018) help to standardize images and facilitate the training of machine learning techniques.

The lack of labeled and reliable publicly available data is another challenging process for using supervised learning techniques (KRAWCZYK et al., 2016; GUO et al., 2019; YAN et al., 2019). Furthermore, (KRAWCZYK et al., 2016) also points to the problem of class imbalance on public bases. The lack of data is challenging for supervised learning algorithms because for homogeneous and unbiased learning, and there is a need for the set to have the same amount of examples for each class. When artificial neural networks are used, this lack of data is usually resolved by applying data augmentation or transfer learning.

For histopathological images, it is common to use data augmentation processes such as mirror, flip, rotation, translation, color jittering, and cut (BEJNORDI et al., 2017b; PAN et al., 2018; GHAZVINIAN ZANJANI et al., 2019; DAS; DUTTA, 2019; VO; NGUYEN; LEE, 2019; ALOM et al., 2019; YAN et al., 2019; GUO et al., 2019; ROY et al., 2019). By rotating and flipping images, we can simulate a real lab environment because pathologists do not adopt a fixed orientation to perform image analysis (YAN et al., 2019). However, it is essential to use this technique only in the training set. This care, when applying data augmentation, is necessary to avoid that the results obtained are not too optimistic considering the variability of the H&E images.

Additionally, the subsampling technique can be used to reduce the intra-class difference of sets. It is a less popular technique, which is to exclude specific images from higher sampled classes (KRAWCZYK et al., 2016).

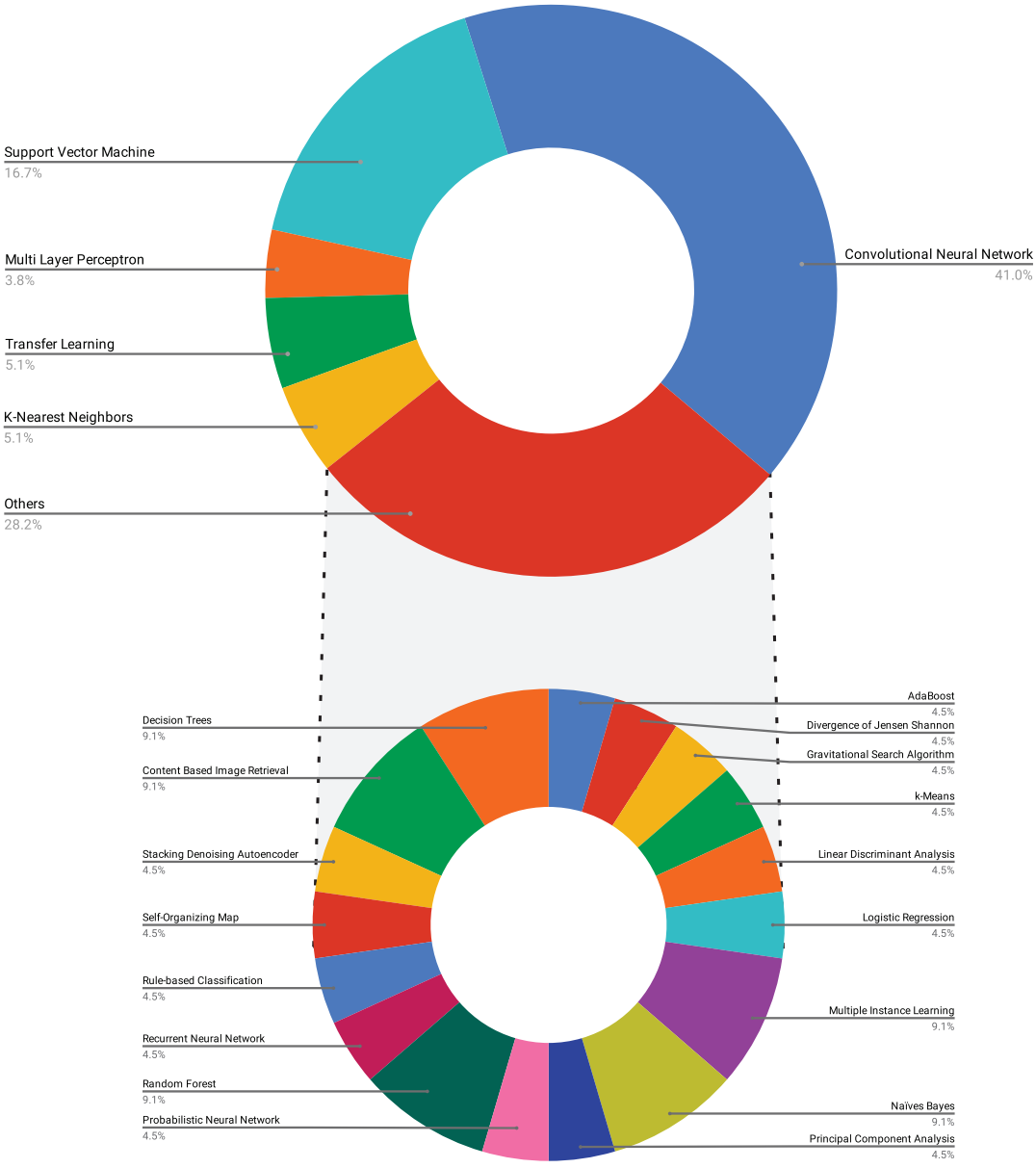
SQ3 What machine learning techniques are used for breast cancer detection?

We analyzed all articles in the corpus to identify machine learning techniques used within the histopathological images of breast cancer. The techniques and their respective percentages of use are shown in Figure 27. There were situations where the same article used more than one technique; in these situations, we accounted for all the techniques used by the method proposed in the article. Based on the classification presented in the taxonomy, we will discuss how the techniques were employed in three topics: supervised, semi-supervised, and unsupervised learning.

Supervised Learning:

Based on the work analyzed, we can highlight, especially in recent years, results obtained with CNN. Adapted CNN models often arise from problems with classification, detection and segmentation of natural images such as Res-Net-v2 (GANDOMKAR; BRENNAN; MELLO-THOMS, 2018a), AlexNet (MURTAZA et al., 2019; SPANHOL et al., 2016a), VGGNet (Sabeena Beevi; NAIR; BINDU, 2019) , Faster R-CNN (LI et al., 2018), and

Figure 27 – Techniques used for the analysis of histopathological images and their percentages of use in our corpus.



Source: Elaborated by the author.

GoogleNet (HAN et al., 2017). Already in (ALOM et al., 2019) is proposed the fusion of several networks (Inception-v4, ResNet, and Recurrent Convolutional Neural Network (RCNN)) for the classification of histopathological images. This technique has as its main advantage of better performance and the use of the same number of parameters compared to an RCNN or traditional Residual Network.

The classification of various magnifications of histopathological images is also studied in the literature (BAYRAMOGLU; KANNALA; HEIKKILA, 2016). The proposed method is based on a single CNN for all amplifications. As pointed out by the authors, the model proved stable in the classification for any amplification available BreakHis. However, the size of the WSI precludes direct use on CNN. Sampling methods of regions of interest are applied in the literature to process image patches on CNN (BEJNORDI et al., 2017b; CRUZ-ROA et al., 2018; GECER et al., 2018). Moreover, the use of CNN may face a problem of clinical acceptance, as the technique is considered a black box in certain situations the pathologist may be resistant to lack of evidence for the diagnosis (BALAZSI et al., 2016).

Due to the high variability of histopathological images, classification in the most diverse histopathological types of breast cancer is a challenging process, as already pointed out in GQ3. The use of stacked CNN (MURTAZA et al., 2019) is a proposed solution using the AlexNet network for initial classification into benign and malignant. Then two CNN networks (one for benign and one for malignant class) are used for classification into histopathological types.

The use of method sets has also been applied to intelligent methods for analyzing histopathological images, such as using the pre-trained VGGNet feature extraction network (MITTAL; SARASWAT, 2019). For this, the authors remove the softmax layer from the network, obtaining a resource vector with 4096 characteristics. Finally, they make a comparison between four classifiers: Support Vector Machines (SVM), Linear Discriminant Analysis (LDA), Random Forest (RF) and k-Nearest Neighbors (kNN). In this paper, the authors point to RF as the best classifier. In (YAN et al., 2019), the authors used the pre-trained Inception-v3 network for feature extraction and a Long Short-Term Memory (LSTM) network for classification into four classes (normal, benign, in situ carcinoma, and invasive carcinoma). Similar works using CNN as feature extractors are also proposed by (WAN et al., 2017b), (GUPTA; BHAVSAR, 2018), (MURTAZA et al., 2019), (VO; NGUYEN; LEE, 2019), and (SPANHOL et al., 2017).

In (BARDOU; ZHANG; AHMAD, 2018) and (CRUZ-ROA et al., 2014), a comparison is made between CNN and feature extraction with traditional classifiers. In both studies, CNN achieves the best results. However, handcrafted resources can also be used in conjunction with CNN to improve classifier accuracy (WANG et al., 2014).

Feature extraction can also be done manually based on calculations such as region covariance (KHAN; SIRINUKUNWATTANA; RAJPOOT, 2015), local binary pattern (TASHK et al., 2015; BALAZSI et al., 2016; ALQUDAH; ALQUDAH, 2019), entropy (TASHK et al., 2015; JELEN; FEVENS; KRZYZAK, 2009), stiffness matrix (TASHK et al.,

2015), Log-Gabor Wavelet (ISSAC NIWAS et al., 2012; BALAZSI et al., 2016), fractal dimension (CHAN; TUSZYNSKI, 2016), among others. In the analyzed corpus, the classification of images based on descriptor representations is also performed by kNN (KHAN; SIRINUKUNWATTANA; RAJPOOT, 2015; TASHK et al., 2015), SVM (TASHK et al., 2015; LOUKAS et al., 2013; FONDÓN et al., 2018; SAHA; CHAKRABORTY; RACOCEANU, 2018; ALQUDAH; ALQUDAH, 2019), PNN (LOUKAS et al., 2013), and Decision Trees (BAKER et al., 2018). However, using feature extraction-based techniques requires advanced knowledge of the problem to obtain essential features that fit a real environment (SPANHOL et al., 2016a).

(LI; PLATANIOTIS, 2015) makes a comparison between selecting the appropriate color space for image classification. The authors compare the classification results with an SVM based on features extracted from the RGB, HSV, CIE L * a * b *, and stain-dependent H and E decomposition model spaces. The results indicate that statistically, the utilization of the H and E decomposition model is better.

Another similar line is the use of Content-Based Image Retrieval to indicate the probability of malignancy in a region (ZHENG et al., 2018) or for the diagnosis of preinvasive cancer (ZHANG et al., 2015). This technique consists of mapping significant descriptors based on the histopathological images of breast cancer with the known diagnosis, for later classification of unknown cases based on the comparison of the characteristics of the images (ZHENG et al., 2018).

Semisupervised Learning:

In the pathological classification process, it is common to obtain several samples of the tissues under study. However, labeling these is an expensive process. A way to solve this problem, (OLIVEIRA et al., 2018) proposes the use of Multiple Instance Learning (MIL). According to the author, this technique consists of grouping sets of images or patches and assigning a unique label to each set. In the case of binary classification, a set of images is considered positive when it has at least one positive instance. And, negative, only if all images are negative in the collection.

In (DUNDAR et al., 2011), MIL is used for the classification of intraductal breast lesions (usual ductal hyperplasia , atypical ductal hyperplasia and ductal carcinoma in situ). In work, the author performs the extraction of resources based on qualitative characteristics used by the pathologists (Cell size, cell shape, and nucleoli) and morphological (perimeter and mean of gray level intensity), with an accuracy of 87.9% in the cases analyzed in the set of tests.

Unsupervised Learning:

K-means is used to undersampled non-mitotic regions in breast cancer histopathological images (WAHAB; KHAN; LEE, 2017). Principal Component Analysis (PCA) can be used as a feature selection algorithm, reducing the computational load on processing less significant instances (Sabeena Beevi; NAIR; BINDU, 2019). Meanwhile, Self-Organizing Map (SOM) and PCA techniques are used in the process of classifying malignancy based on characteristics

extracted from segmented nuclei (JELEN; FEVENS; KRZYZAK, 2009).

In general, combining multiple features yields better results compared to jobs that use unique features (ZHANG et al., 2016). Interesting approaches in this line are the use of Stacked Sparse Autoencoder (SSA) as feature extractors for both WSI classification and core detection (XU et al., 2016; ZHANG et al., 2016). A SSA is known as an encoder/decoder neural network, using backpropagation to produce an output value as close to the input value (XU et al., 2016). Thus, by using the information provided by the intermediate layers, we can, in theory, obtain a set of resources that represent the information contained in the image (ZHANG et al., 2015).

SQ4 What are the metrics for the validation of the work?

With the fourth specific question (SQ4), we intend to raise the work evaluation metrics. When an intelligent analysis method is proposed, a test set is usually set aside to be classified and benchmarked to study performance. For this purpose, different validation metrics are used. In Figure 28 we present the metrics used in the corpus, a brief description of each, the task in which it can be applied, and the frequency of use. The metrics are not exclusive, and more than one may be used in the corpus.

Accuracy, the main and most widely used metric in the corpus, measures the percentage of correctly classified samples (VO; NGUYEN; LEE, 2019). Nevertheless, using accuracy alone can underestimate the efficiency of the method because in cases of unbalanced data (there are more samples for one class than another), the approach can efficiently classify only one category. Thus, metrics such as specificity and sensitivity, which measure the ability to classify correctly, respectively, the normal and abnormal classes, are useful to complement the evaluation of the model (XIE et al., 2019). Another way to check how the method behaves with unbalanced classes is by using F1-Score. F1-Score evaluates the model with a measurement between 0 and 1, with value 1 being the perfect model (FENG; ZHANG; YI, 2018).

Meanwhile, with the precision, we can get the proportion of abnormal cases sorted correctly in the test suite (FENG; ZHANG; YI, 2018). With the Receiver Operating Characteristic (ROC) curve, we can graphically evaluate the models by plotting sensitivity versus 1 - specificity on various thresholds (BEJNORDI et al., 2017b; ALOM et al., 2019). And from the ROC curve, we get another interesting metric, the Area Under the ROC curve (AUC), ranging from 0 to 1. The AUC measures the area to the right of the ROC curve, the closer to 1, the better the model (KRAWCZYK et al., 2016; YAN et al., 2019).

However, even with the use of metrics, it is not easy to make a direct comparison of the studies. This problem is due to the lack of information present in the studies on how the evaluations were made, the lack of public datasets and methods that are difficult to replicate. Thus, competitions such as ICIAR 2018, BioImaging2015, and CAMELYON can be used as appropriate references, as it keeps model evaluation servers evaluating methods on a standard data set.

Figure 28 – Performance evaluation methods applied in selected studies.

Metric	Description	Applicable Task	Equation	Quantity
Accuracy	Consisting of the ability to correctly classify the classes.	Cancer classification, nuclear atypia classification, cancer multi-classification, nucleus segmentation, and mitosis detection	$\frac{TP+TN}{TP+TN+FP+FN}$	
AUC	Provides a scalar measure of a classifier's performance and has been used in unbalanced domains.	Cancer classification, cancer multi-classification, tumour segmentation, and mitosis detection	$1+\left(\frac{TP}{TP+FN}\right)-\left(\frac{FP}{FP+TN}\right)$	
Balanced Accuracy	A balanced assessment is achieved even in cases where the classifier is capable of classifying only one class.	Cancer detection	$\frac{Sensitivity+Specitivity}{2}$	
Dice coefficient	It is a metric used to measure similarity between two datasets.	Nucleus segmentation, and tumor segmentation	$\frac{2\lvert X\cap Y\rvert}{\lvert X\rvert+\lvert Y\rvert}$	
F1-score	It is a harmonic measure between precision and sensitivity.	Cancer classification, nucleus classification, cancer multi-classification, nucleus segmentation and mitosis detection	$\frac{2*TP}{2*TP+FN+FP}$	
False Negative Rate	With this metric, it is possible to identify the amount of wrongly labeled positive.	Cancer classification	$\frac{FN}{FN+TP}$	
False Positive Rate	With this metric, it is possible to identify the amount of negative wrongly labeled.	Cancer classification, and nucleus detection	$\frac{FP}{FP+TN}$	
Free-response Receiver Operating Characteristic curve	This is the True Positive Rate plot versus the number of false positives per image.	Tumor detection	-	
Geometric Mean	It considers a balancing between the accuracy over the instances of the minority and majority classes at the same time .	Cancer classification	$\sqrt{\left(\frac{TP}{TP+FN}\right)*\left(\frac{TN}{FP+TN}\right)}$	
Cohen's Kappa	Cohen's Kappa statistic measures interrater reliability	Cancer classification, and multi-classification	$\frac{P_o-P_e}{1-P_e}$	
Macro Average	It is used to evaluate multiclass rating performance. The F1-score of all n categories is computed to obtain the global average.	Cancer classification	$\frac{1}{N}\sum_{i=1}^{n}F_i$	
Mean Absolute Difference	It can evaluate the average difference between each rating point against the labels.	Mitosis detection	$\frac{1}{N}\sum_{1}^{n}[min\lVert S_i^c-S_i^g\rVert_2]$	
Patient Recognition Rate	It is the correct rating rate at the patient level.	Cancer classification, and multi-classification	$\frac{\sum_{p=1}^{N_p}\frac{N^p}{N_{im}^p}}{N_p}$	
Precision	This metric represents the fraction of positive instances classified correctly.	Cancer classification, nucleus classification, cancer multi-classification, nucleus segmentation, cell detection and mitosis detection	$\frac{TP}{TP+FP}$	
Receiver operating characteristic curve	To obtain the ROC curve we must plot the True Positive Rate versus the False Positive Rate for different rating thresholds.	Cancer classification, cancer multi-classification, cell detection and mitosis detection	-	
Sensitivity	With this metric we can measure the amount of actual positives that are correctly labeled.	Cancer classification, nucleus classification, cancer multi-classification, nucleus segmentation, cell detection and mitosis detection	$\frac{TP}{TP+FN}$	
Specitivity	With this metric we can measure the amount of actual negatives that are correctly labeled.	Cancer classification, nucleus classification, cancer multi-classification, nucleus segmentation, cell detection and mitosis detection	$\frac{TN}{TN+FN}$	

Caption

Quantity of Articles

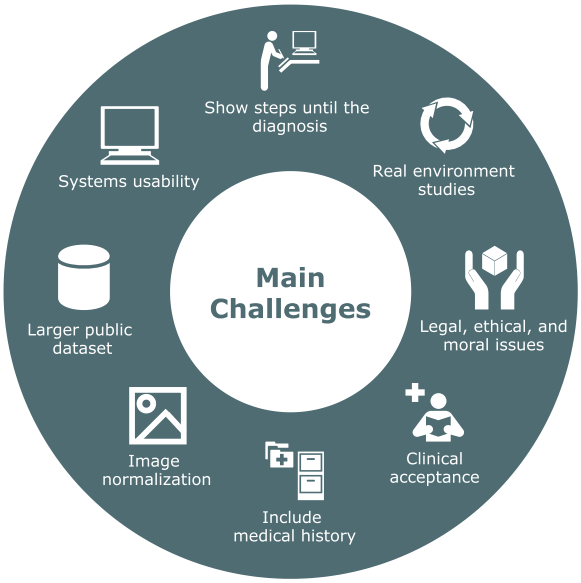
<2 <4 <8 <16 <32 <64

Source: Elaborated by the author.

3.3 Challenges and Research Opportunities

Based on the analysis of the selected studies, we identified several research challenges and opportunities. In Figure 29, we group the main challenges encountered during the SLR. It is essential to highlight that the gaps found are concentrated mainly around the pathologist. An open gap is the need to study the clinical acceptability of CAD systems, and how they can be adapted to the routine analysis of breast cancer histopathological slides (BALAZSI et al., 2016). Providing some information that will lead the healthcare professional to understand how the model has achieved the outcome can, at a minimum, help overcome distrust in the use of machine learning applied to health. Another point that can increase clinical acceptability is benchmarking platforms simulating real laboratory environments.

Figure 29 – Main challenges related to intelligent analysis of breast cancer histopathological images.



Source: Elaborated by the author.

Regarding opportunities at the computational level, there is a need for studies using intelligent methods applied to a significant amount of data. The use of compiled data from several institutions allows studies on color normalization, reduction of artifacts, and differences in WSI resolution. In addition, we have identified research opportunities in the detection of significant regions in breast cancer WSI. This step is of fundamental importance for the processing of WSI in intelligent methods, such as CNN.

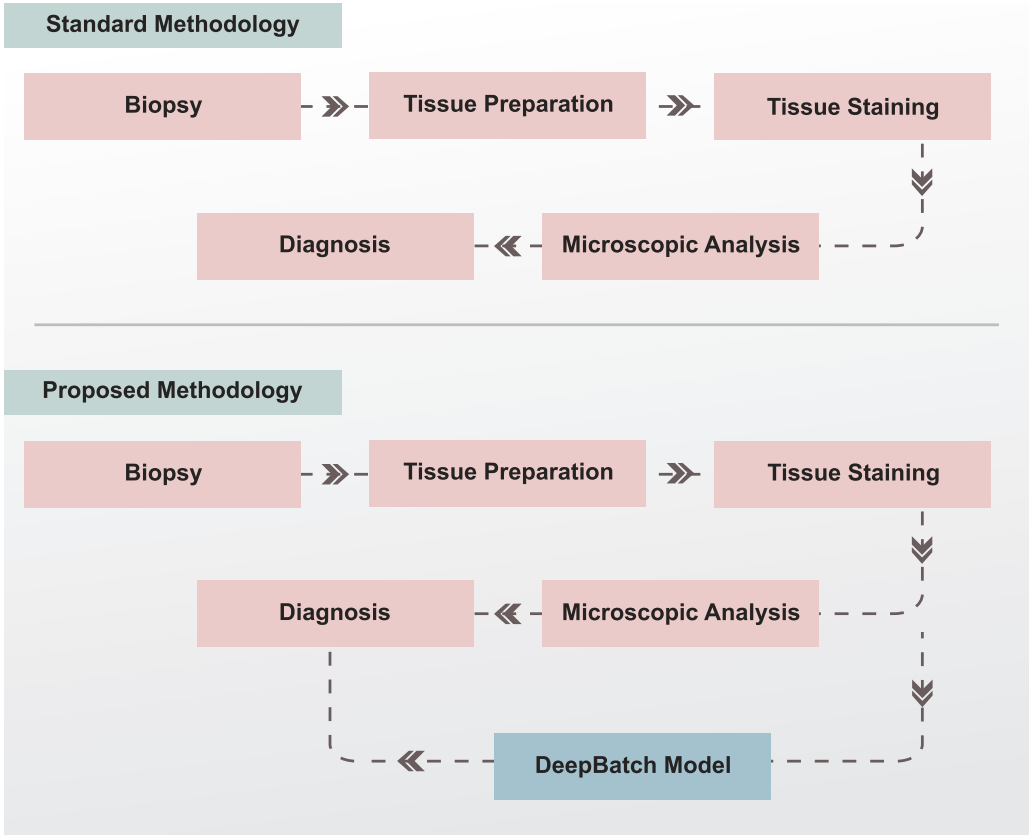
Studies that manage to reconcile a better interpretation, providing clear and understandable information and, at the same time, achieve results comparable to state of the art may offer less resistance for validation by regulatory bodies. Thus, the development of systems that can simulate the workflow adopted by the pathologist can be an opportunity for research. The adoption of an intelligent method similar to that of the pathologist may be favorable concerning the challenges and discussions on the ethical and moral issues of the use of intelligent methods.

In the related work presented, there are several open research issues. To close the scope of this Master Thesis, this work focuses on three opportunities found in the related work: (a) provide the diagnosis in a clear and interpretable way, through demarcation of the regions at high magnifications, refined segmentation at the cellular level and a diagnosis generated in natural language; (b) use of a WSI compilation from different sources; (c) bringing the flow of the intelligent method for the identification of breast cancer in WSI closer to the flow adopted by the pathologist. Concerning these problems, we can observe several solutions proposed in the literature (BALAZSI et al., 2016; SPANHOL et al., 2017; GANDOMKAR; BRENNAN; MELLO-THOMS, 2018b; VO; NGUYEN; LEE, 2019; GUO et al., 2019). However, we note that the use of a hybrid model using CNN is a promising approach. We can also highlight the lack of studies that address the three problems together. In this way, our scientific contribution aims to promote the use of CNN as a model that addresses the issues related to the identification of breast cancer in WSI by providing an interpretable diagnosis. In the next chapter, we present the technologies for the development of the model and detail the architecture of the DeepBatch model.

4 DEEPBATCH MODEL

As already noted, the patient's prognosis depends on the correct classification of the histopathological type and the stage of the disease (MAKKI, 2015). In this way, applications that are capable of being integrated into the pathologist's routine, and that can reduce the restrictions in histopathological analysis can contribute to a more accurate and reliable diagnosis. In summary, the standard methodology performed by the pathologist consists of acquiring the biopsy, preparing and staining the tissue, microscopic analysis, and diagnosis. The purpose of this Master Thesis is to add a step in parallel to the microscopic examination (Figure 30), with the DeepBatch model serving as a support system for medical diagnosis. The name DeepBatch was obtained by generating an acronym from the phrase "a DEEP learning model for Breast cAncer deTeCtion in Histopathological image".

Figure 30 – Proposal to integrate the DeepBatch model into the pathologist's routine.



Source: Elaborated by the author.

However, even for the most modern computer systems, automatic WSI analysis is a challenging task. There are four main reasons: (1) different analysis and staining protocols; (2) scarcity of pixel-labeled WSI; (3) size of the WSI, and (4) complex and heterogeneous biological structures. The DeepBatch model seeks to solve these restrictions with a hierarchical structure. This chapter details all the parts that make up the DeepBatch model, starting with an overview and explaining each of the architecture modules. Section 4.1

presents the design decisions for the realization of the proposed model. Section 4.2 describes an overview and the functionalities of each DeepBatch module.

4.1 Project Decisions

The modules that make up the DeepBatch were designed in a way that resembles the workflow adopted by the pathologist. The model recognizes suspicious WSI regions at a global level for further local analysis. The proposed strategy is modeled to provide information on the areas that influenced the diagnosis and so that the model can be easily adapted to the pathologist's routine.

Based on these conceptual definitions, we define the organization of the model at a technological level. The first decision is related to the size of the WSI, which has billions of pixels. Thus, due to the limitation of computational resources, we defined the use of WSI in a magnification of $2.5\times$ for the global detection of suspicious regions. As a way to speed up the training of the intelligent technique of the ROI Detection module, we apply image processing techniques to remove background and noise present in WSI. Remove the background and noise in images for processing on CNNs increases the likelihood of obtaining more accurate results (RAJNOHA; BURGET; POVODA, 2018).

In DeepBatch, for the global screening of suspicious regions, we use a CNN. The CNN architecture in this module will be responsible for producing a binary segmentation with the location of suspicious areas based on the pre-processed low-resolution WSI ($\approx 3000 \times 3000$). As this is an intermediate stage, FN must not occur. Therefore, higher PF rates are acceptable at this level.

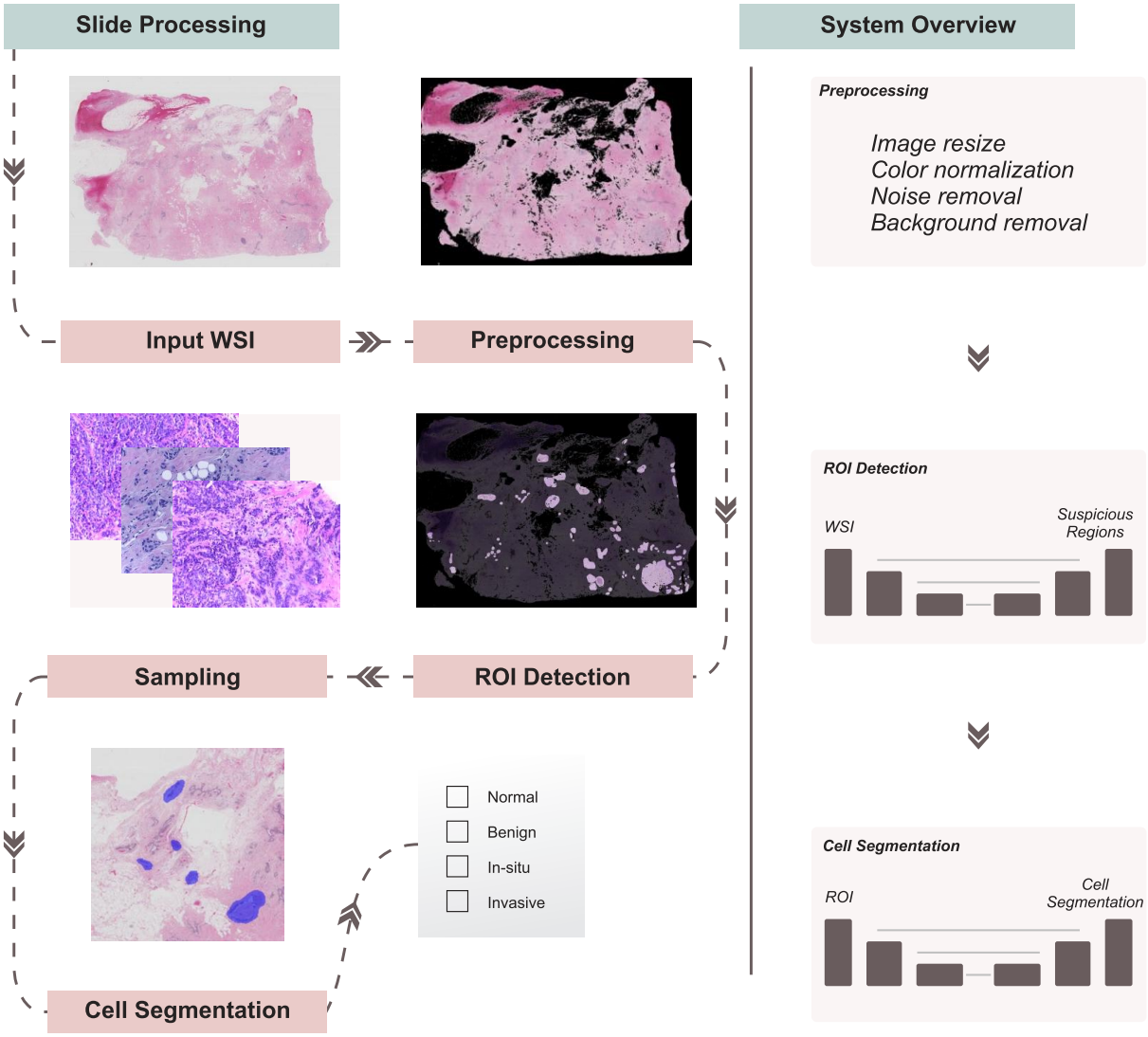
Based on the segmentation produced by the global analysis, we defined a sample of Regions of Interest (ROI) in WSI at a magnification of $40\times$. The size of the ROI will be fixed and limited by the computational capacity available. Another CNN will process the sampled ROI. This CNN will be responsible for producing a detailed multiclass segmentation for WSI. The decision to use CNNs is based on the results obtained by state of the art identified in the SLR.

4.2 Architecture

One of the conceptual bases of the DeepBatch model is to transmit to the pathologist a view of which regions and characteristics of the WSI influenced the diagnosis. The core of the DeepBatch is composed of several CNN arranged in cascade. These CNNs work together to optimize computing resources and provide refined segmentation of WSI. Segmentation provides pathologists with a view of which regions have influenced diagnosis at global and local levels. Thus, the DeepBatch seeks to allow a diagnosis that is understandable and interpretable by the pathologist. The DeepBatch inference process (Figure 31) is divided into four main modules: Preprocessing, ROI Detection, ROI Sampling, and Cell Segmentation. In Figure 31, we can see

two main components Slide Processing and System Overview. In the next subsections, each module of the DeepBatch will be presented, detailing its components and techniques involved.

Figure 31 – DeepBatch model. In Slide Processing, we show an example of WSI being analyzed by the DeepBatch model. In the System Overview, we demonstrate, in a conceptual way, the operation of the DeepBatch.



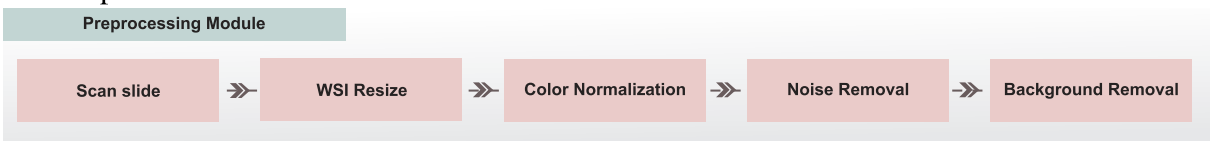
Source: Elaborated by the author.

4.2.1 Preprocessing Module

As highlighted in the previous sections, WSI are images with billions of pixels, variations in coloring, differences in tissue thickness, and a tissue composed of complex, heterogeneous biological structures. In this way, the Preprocessing Module in the DeepBatch is responsible for mitigating the problems faced in the automatic analysis of WSI. Figure 32 presents an overview of the steps that make up the Preprocessing Module.

The colors present in the WSI transmit information that plays an essential role in the

Figure 32 – Preprocessing Module steps. WSI goes through a process of preparation for use by the DeepBatch model.



Source: Elaborated by the author.

diagnosis. In computing, colors are represented by color spaces, which aim to represent the colors visible by humans. However, there are significant differences in the number of colors in the visible electromagnetic spectrum that each model is capable of representing (GONZALEZ; WOODS, 2018). Thus, in this step, we will evaluate the quantitative and qualitative performance of four-color spaces (RGB, HSV, YCbCr, and LAB) for the identification of regions with malignant potential.

To minimize the interference of color variation, it is common in the literature to use color normalization (KHAN; SIRINUKUNWATTANA; RAJPOOT, 2015; LI; PLATANIOTIS, 2015; GANDOMKAR; BRENNAN; MELLO-THOMS, 2018b). In this step, we chose to use the method proposed by Khan et al. (2014). Briefly, the algorithm consists of a non-linear mapping of an image to a target image using a representation derived from color deconvolution.

Due to hardware limitations, it is not possible to process WSI directly on a CNN. In this way, for this module, we use WSI in a magnification of 2.5×, reducing the complexity in this stage. In addition, as a way to speed up the WSI segmentation process, we have removed the white borders present in the WSI using the adaptive segmentation of Otsu (OTSU, 1979) to separate the bottom of the tissue.

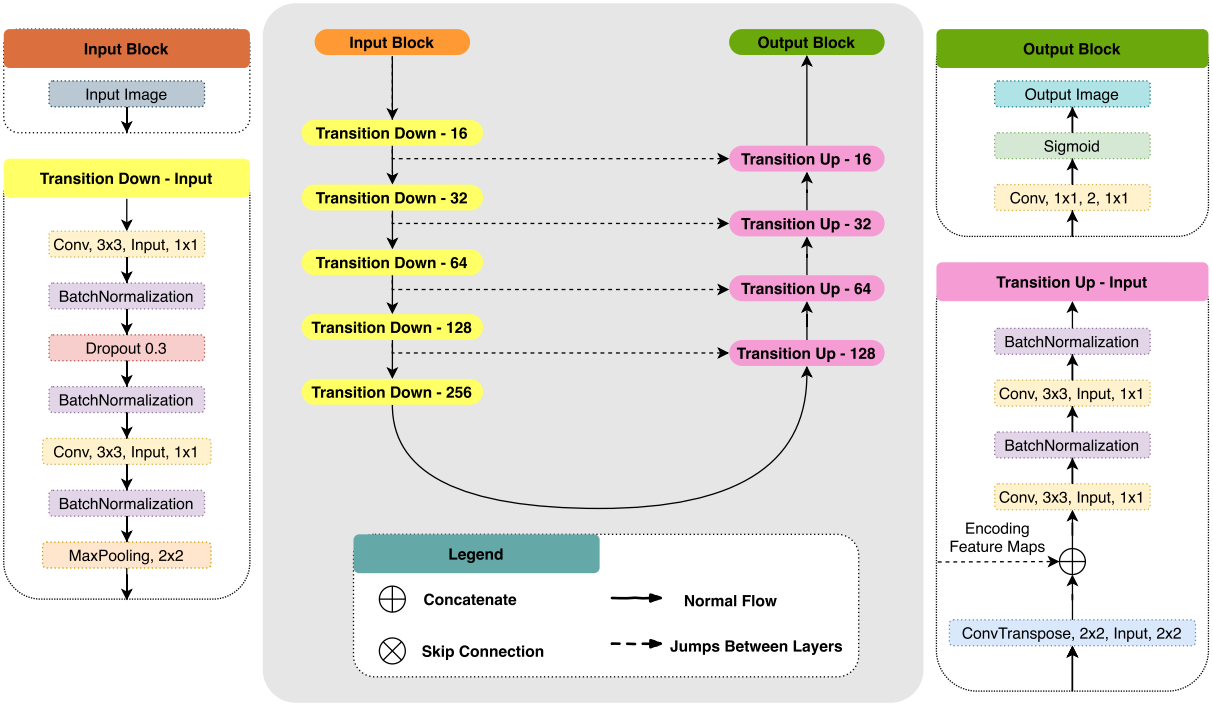
4.2.2 ROI Detection Module

The main function of the ROI Detection Module is to classify each pixel of the preprocessed WSI as an abnormal region or not. For this classification, DeepBatch proposes the use of CNN in U-Net format, which will provide a probability associated with each pixel. In Figure 33, we present the convolutional architecture used in this module.

Before CNN training, some steps can be used to assist in the training and inference of WSI (Figure 34). The first is related to the computational capacity of image processing. Even with magnification of 2.5×, WSI still has a relatively large resolution ($\approx 3000 \times 3000$ pixels). Thus, for processing at CNN, WSI can be divided into non-overlapping and computationally processable patches. We also propose the use of data augmentation, with the generation of synthetic data applying transformations such as rotations, zoom, horizontal and vertical inversions.

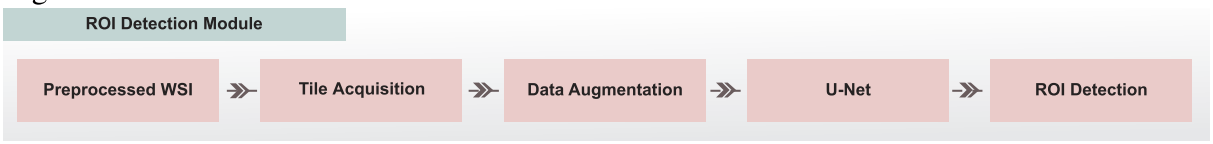
Finally, this module must be efficient in detecting suspicious regions in WSI. Thus, the occurrence of a low rate of False Negative (FN) at the expense of a high number of False

Figure 33 – Convolutional architecture for ROI Detection Module. The gray part of the figure represents the summarized flow of the network. On the left side, the detailed blocks of the downsampling of the images are represented. In the right part, the detailed blocks of the upsampling of the images are represented. Each convolutional layer is represented as follows: *kernel*, *feature maps* and *stride*.



Source: Elaborated by the author.

Figure 34 – ROI Detection Module steps. Preprocessed WSI undergoes tile sampling and data augmentation to be trained on a CNN.



Source: Elaborated by the author.

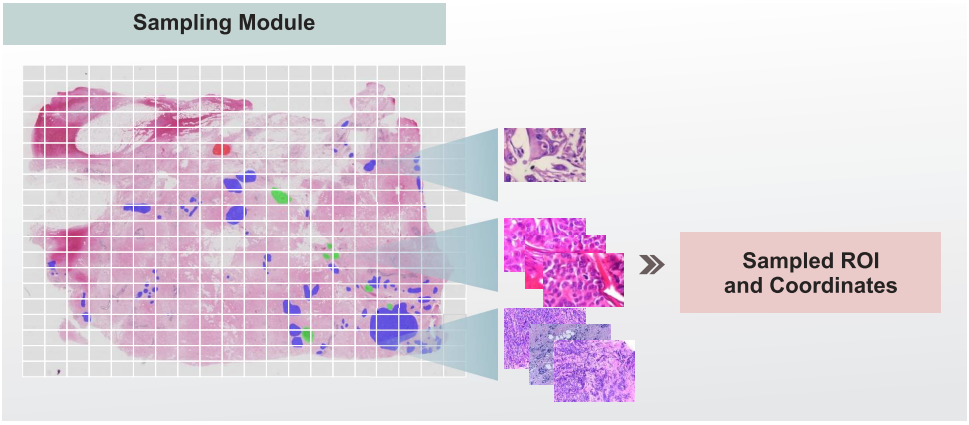
Positive (FP) is desirable. Therefore, a higher penalty will be applied to areas with the presence of tumors in the WSI erroneously classified during training.

4.2.3 Sampling Module

Based on the suspicious regions identified in the ROI Detection Module, the Sampling Module will select regions (called ROIs) of high pixel density at higher magnification of the WSI. To decide the region to be sampled, the average of the probabilities of the pixels of the current patch will be taken into account whether or not it is a neoplastic region. The mapping will correspond to the multiplication of the WSI reduction factor by the coordinates of the segmented region, obtaining the set of coordinates in high resolution. This set of coordinates

will be used to generate the training set for the Cell Segmentation Module. An illustration of this process can be seen in Figure 35.

Figure 35 – Sampling Module. Suspicious regions identified by the ROI Detection Module are mapped to high magnification regions.



Source: Elaborated by the author.

4.2.4 Cell Segmentation Module

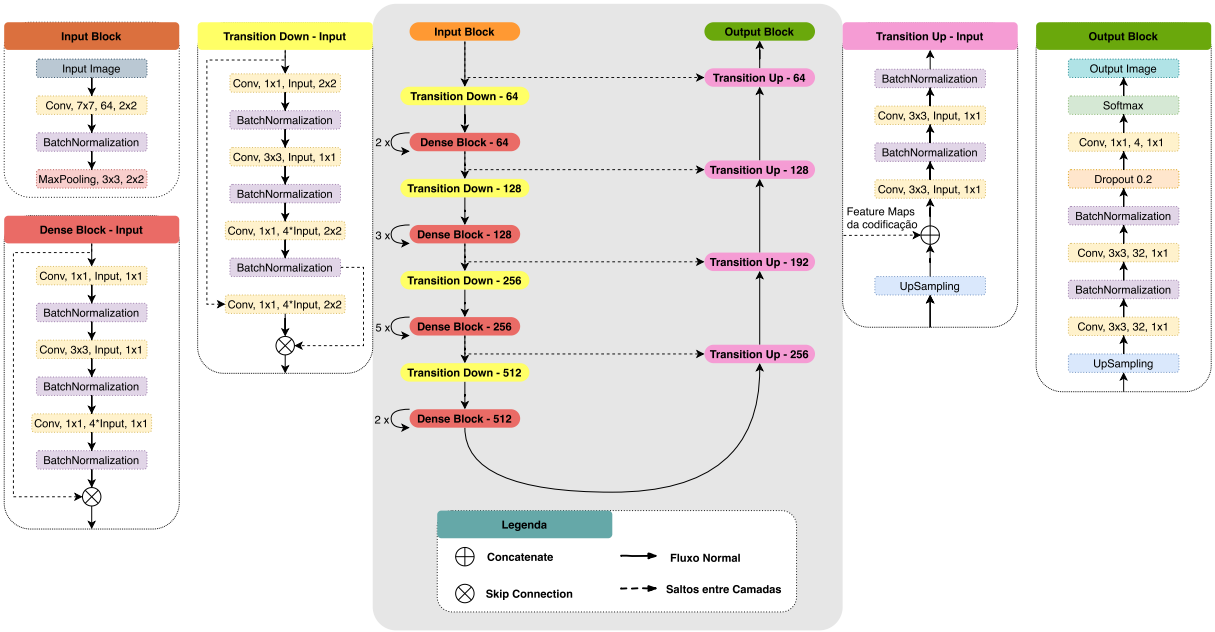
The ROIs sampled by the Sampling Module are used as input to the Cell Segmentation Module (Figure 36). In this module, a higher capacity for cellular complexity is required. Thus, to speed up the training of convolutional architecture, we made some modifications to the U-Net architecture (Figure 37). In the downsampling stage, we use a pre-trained ResNet50 (HE et al., 2016) for the ImageNet set. The small jumps between the convolutional layers of the ResNet50 architecture allow a faster convergence and less chance of gradient vanish (HE et al., 2016). In this module, we also use data augmentation techniques to increase the variability of images shown to the convolutional architecture.

Characteristics considered for the diagnosis of the patient, such as cell differentiation, mitotic activity, growth pattern, and invasion of adjacent tissue, are visible and considered in the inference for the diagnosis proposed by the DeepBatch model. The segmentation of regions at the cellular level provides pathologists with a view of the areas that influenced WSI analysis. This function aims to give the model greater credibility and increase the DeepBatch clinical acceptability.

Figure 36 – Cell Segmentation Module steps. High magnification images go through a tile sampling process and data augmentation to be segmented by a CNN.

Source: Elaborated by the author.

Figure 37 – Convolutional architecture for the Cell Segmentation Module. The gray part of the figure represents the summarized flow of the network. Next to each *Dense Block*, the number of times it will be repeated in the architecture is specified. On the left side, the detailed blocks of the downsampling of the images are represented. In the right part, the detailed blocks of the upsampling of the images are represented. Each convolutional layer is represented as follows: *kernel*, *feature maps* and *stride*.



Source: Elaborated by the author.

4.3 Final Remarks

Therefore, the DeepBatch model proposes a WSI analysis architecture in modules. The modules work in cascade to optimize the computational resources. In a first step, WSI is preprocessed by a set of techniques that will be detailed in Section 5.4.2. The ROI Detection Module will use the images resulting from this module for global analysis and identification of suspicious regions. The Sampling Module will map low magnification regions to high pixel density regions. Finally, the Cell Segmentation Module will use a pre-trained convolutional architecture to produce a refined multiclass segmentation and heat map for suspect regions. Based on the experiments detailed in the next chapter and with the results obtained and presented in Chapter 6, we can say that the DeepBatch model is capable of answering our

research question. Therefore, in the next chapter, we present in detail the techniques, procedures, and meta parameters used in the implementation of the DeepBatch model.

5 MATERIALS AND METHODS

This chapter is organized as follows. In the beginning, the data sets for evaluation, and the evaluation criteria are described. After that, implementation strategies and details are defined DeepBatch model.

5.1 Materials Description

Recently, pathology services in Europe initiated procedures for complete digitization of the slides analyzed (DIMITRIOU; ARANDJELOVIć; CAIE, 2019). This process meets the proposals for automated analysis of slides (SAHA; CHAKRABORTY, 2018). However, as demonstrated by the SLR, there is a shortage of publicly available WSI breast cancer datasets.

As a way to solve this gap, in this study, we use a compilation of datasets identified in the literature and publicly available. The selected datasets and their main characteristics are listed in Table 9. Details on obtaining and organizing the datasets are discussed in the next subsections.

5.1.1 BACH

The Grand Challenge on Breast Cancer Histology Images (BACH) consists of two sets called A and B, the first consisting of 400 images of WSI sections with four classes: normal, benign, carcinoma in situ, and invasive carcinoma. Set B consists of 40 high-resolution WSI containing all of the sampled tissue. The 40 WSI in set B is divided into two subsets, 30 for training and 10 for testing. Of the 30 WSI provided for training, only ten are labeled at the pixel level in four classes: normal, benign, carcinoma in situ, and invasive carcinoma. Thus, in this work, only 10 WSI are used, due to the need for labeling at the pixel level (ARESTA et al., 2019). In Figure 38, we present an example of WSI from the BACH dataset.

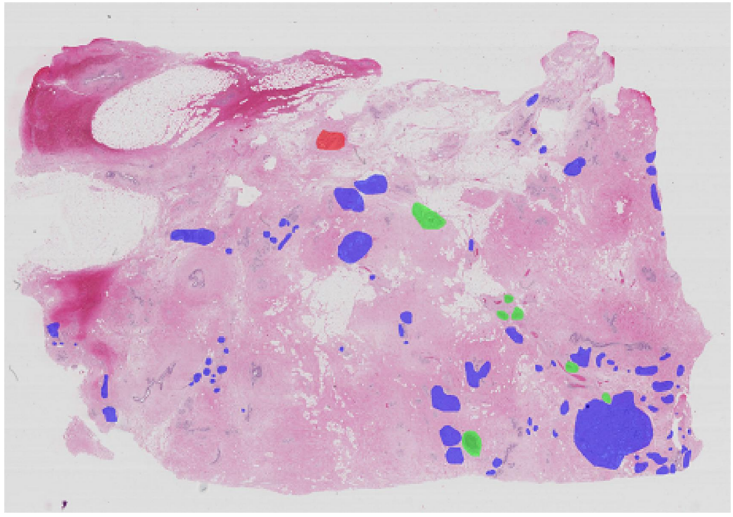
The base WSIs were acquired between 2013 and 2015 in patients from Portugal using a Leica SCN400 digital slide scanner and made available in SVS format. In addition, each SVS file provides ten magnification levels for the samples, with a minimum of 0.04$\times$ and a maximum of 40$\times$. As for the labels, an XML file was provided containing the coordinates for each region labeled (ARESTA et al., 2019).

Table 9 – Public datasets of histopathological images of breast cancer.

Dataset	WSI	Number of images	Lesions types	Patients	Annotation type
BACH (B)	Yes	40	normal, benign, in situ carcinoma, invasive carcinoma	40	pixel-wise
BreCaHAD	No	162	mitosis, apoptosis, tumor nuclei, non-tumor nuclei, tubule, and non-tubule	-	pixel-wise
BREAKHIS	No	7,909	adenosis, fibroadenoma, phyllodes tumor, tubular adenona, carcinoma, lobular carcinoma, mucinous carcinoma and papillary carcinoma.	40, 100, 200, 400	82
BIOIMAGING 2015	No	269	normal, benign,in situ carcinoma and invasive carcinoma	200	-
CAMELYON17	Yes	1000	Normal and metastases	-	-
HASHI	Yes	596	invasive carcinoma	596	pixel-wise
TCGA	Yes	-	ductal and lobular neoplasms, epithelial neoplasms, complex epithelial neoplasms, adenomas and adenocarcinomas, cystic and mucinous neoplasms, squamous cell neoplasms, fibroepithelial neoplasms, adnexal and skin appendage, basal cell neoplasms,mature b-cell lymphomas	3.816	patient level, clinical data, outcomes, genetic expressions, and others
UCSB Breast Cancer Cell	No	58	normal and malignant	-	pixel-wise

Source: Elaborated by the author.

Figure 38 – Example of a pixel-wise annotated WSI from the BACH dataset. Red is benign; Green is in situ; Blue is invasive.

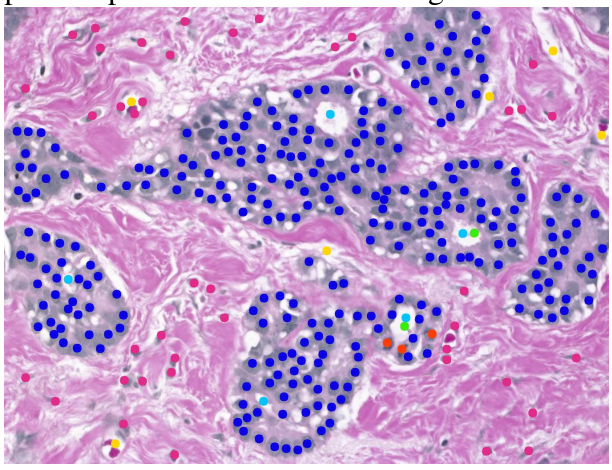


Source: Aresta et al. (2019)

5.1.2 BreCaHAD

BreCaHAD provides a set of 162 breast cancer histopathological images with pixel-level markings of mitosis, apoptosis, tumor nuclei, non-tumoral nuclei, tubule, and non-tubule. The images were obtained from several regions of histopathological slides. Thus, the dataset contains images with clear boundaries to little differentiated structures, and with a lack of typical features (AKSAC et al., 2019). In Figure 39, we present an example of a labeled image from the BreCaHAD dataset. The images provided by the dataset have a dimension of 1360x1024 pixels.

Figure 39 – Example of a pixel-wise annotated image from the BreCaHAD dataset.

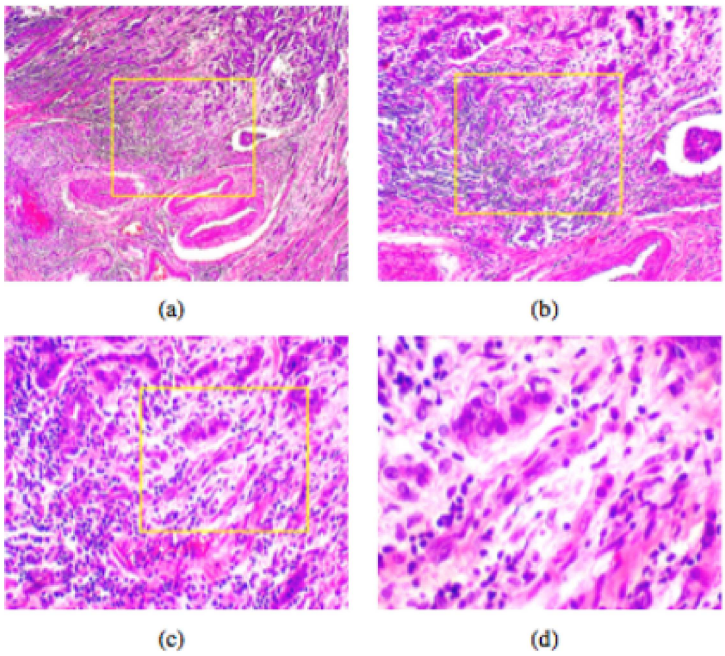


Source: Aksac et al. (2019)

5.1.3 BREAKHIS

BreakHis is a set of images collected from histopathological slides of partial mastectomy or excisional biopsy from 82 patients, with manual selection of significant regions by pathologists. The result was also confirmed with complementary tests, such as immunohistochemical analysis (SPANHOL et al., 2016a). In Figure 40, we present an example of the dataset.

Figure 40 – An example of a slide section of the BREAKHIS base seen in different magnification factors: (a) 40X, (b) 100X, (c) 200X, and (d) 400X.



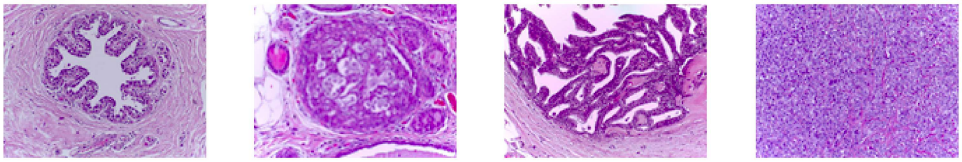
Source: Spanhol et al. (2016b)

The dataset provided by the authors is divided into benign and malignant. With four histopathological types for benign cases: adenosis, fibroadenoma, phyllodes tumor, and tubular adenosine; and four malignant tumors: carcinoma , lobular carcinoma, mucinous carcinoma and papillary carcinoma in dataset (SPANHOL et al., 2016a).

5.1.4 BIOIMAGING 2015

The set of Bioimaging images was made available as part of a challenge at the 4th Internation Symposium in Applied Bioimaging in 2015. The dataset contains 269 images (2048x1536 pixels) extracted from histopathological slides in a 200x magnification. The images were labeled by two pathologists experienced in four histopathological types: normal, benign, in situ carcinoma, and invasive carcinoma (ARAUJO et al., 2017). Figure 41 presents an example of the images provided by the BIOIMAGING database.

Figure 41 – Example of tissue sections from the BIOIMAGING 2015 dataset. A) normal tissue; b) benign tissue; c) carcinoma in situ; or d) invasive carcinoma.



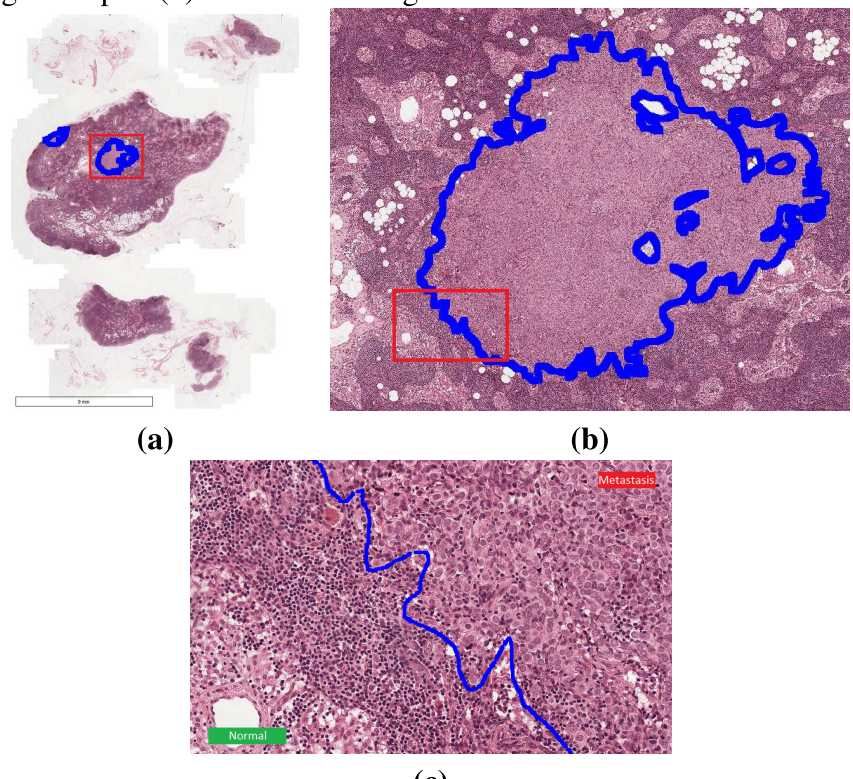
(a) (b) (c) (d)

Source: Araujo et al. (2017)

5.1.5 CAMELYON17

CAMELYON17 is the combination of images from the challenges of CAMELYON16 and CAMELYON17. It contains 1399 sentinel lymph node WSI collected in five medical centers: the Radboud University Medical Center, the Utrecht University Medical Center, the Rijnstate Hospital, the Canisius-Wilhelmina Hospital, and LabPON. Metastases are divided into three categories in this dataset: Macro (metastases greater than 2.0 mm), Micro (metastases greater than 0.2 mm and less than 2.0 mm), and ITC (single tumor or a cluster of tumors smaller than 0.2 mm or less than 200 cells) (LITJENS et al., 2018).

Figure 42 – Example of a metastatic region of the CAMELYON17 dataset. (a) WSI at 1× magnification; (b) red rectangular region of part (a) seen in a 10× magnification; (c) red rectangular region of part (b) seen at 40× magnification.



(a) (b)

(c)

Source: Litjens et al. (2018)

Of the 1399 WSI, 449 are labeled at the pixel level, with 399 referring to the challenge for the year 2016 and 50 for the year 2017. The WSI from CAMELYON16 was first analyzed by an experienced laboratory technician and a clinical Ph.D. student. Two experienced pathologists then inspected the notes. As for the CAMELYON17 WSI, an experienced pathologist labeled the WSI at the slide level, and five experienced observers delimited the metastases. Finally, two resident pathologists reviewed the demarcations (LITJENS et al., 2018).

5.1.6 HASHI

This dataset was made available in the work of Cruz-Roa et al. (2018). It consists of 596 WSI scanned by Aperio or Ventana scanners. All images have been reduced in size by a factor of 32:1. In Table 10, the origins, and the number of cases for each source are shown.

The cases provided by this dataset were labeled at the pixel level by three experienced pathologists. The annotation occurred in a 2x magnification using ImageScope v11.2 and Image Viewer v3.1.4.

Table 10 – HASHI dataset.

Site	**Cases**	**Scanner**
Hospital of the Univ. of Pennsylvania	239	Aperio
Univ. Hospitals Case Medical Center/ Case Western Reserve Univ.	110	Ventana
Cancer Institute of New Jersey	52	Aperio
The Cancer Genome Atlas	195	Aperio

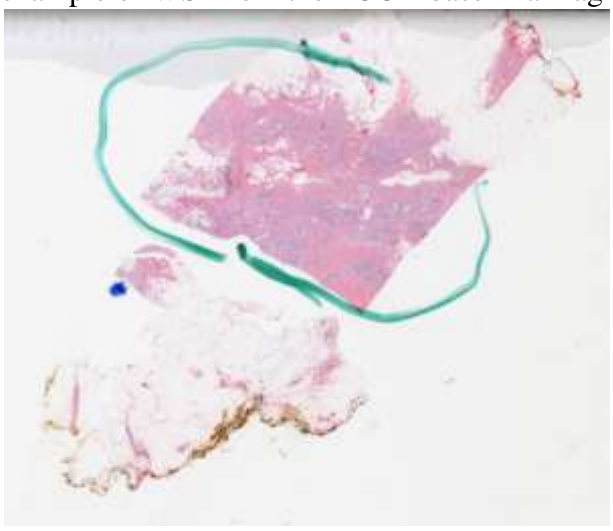
Source: Cruz-Roa et al. (2018)

5.1.7 TCGA

TCGA is a historic cancer genomics program overseen by the National Cancer Institute. It has studies and samples of 33 types of cancer and some tools for data processing, visualization, and other computational tools developed by researchers and collaborators of the program (National Cancer Institute, 2020). In Figure 43, we present an example of breast cancer WSI from the TCGA base.

In the case of breast cancer, the program provides 8,897 cases, providing demographic information, medical history of the patient, WSI, with patient-level labels, histopathological reports, among others (National Cancer Institute, 2020). However, as this dataset is a compilation of cases from different institutions, it can suffer from technical variations, and in the coloring process (NAYAK et al., 2013).

Figure 43 – An example of WSI from the TCGA base in a magnification of $1\times$.

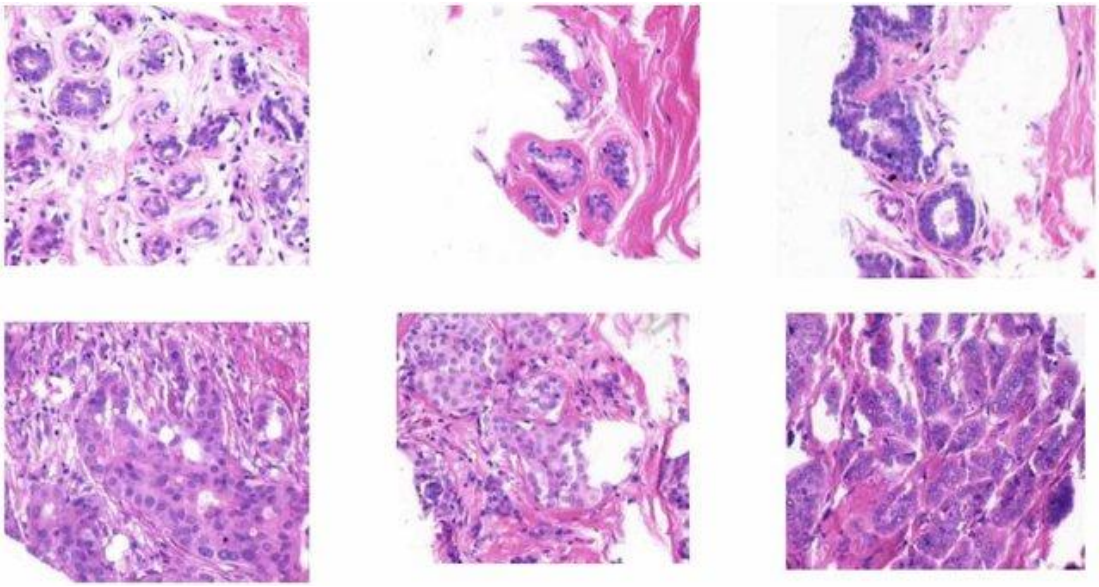


Source: National Cancer Institute (2020)

5.1.8 UCSB Breast Cancer Cell

The UCSB Breast Cancer Cell is a set of 58 images (896x768) with pixel-level annotations extracted from slides stained with H&E. The dataset was built to allow cell segmentation for benign and malignant classes (UCSB Center for Bio-Segmentation, 2020). In Figure 44, we present examples of images of WSI sections of the UCSB dataset.

Figure 44 – Sample images of the UCBS dataset. Top row - benign tumor images, and bottom row - malignant tumor images.



Source: UCSB Center for Bio-Segmentation (2020)

5.2 Metrics: Quantifying the Quality of Predictions

For the evaluation of the models, metrics were defined for the stages of the model at the pixel, image, and case level. The use of these metrics is in agreement with the most used metrics for each problem or learning objective identified in the SLR. The next subsections explore each of the defined metrics.

5.2.1 Metrics Based on the Confusion Matrix

In a binary classification problem, instances are mapped to two classes, for example, positive or negative. Formally, a classifier X will map the instances I to one of the two classes positive or negative $\{p, n\}$. Thus, given any binary classifier, there are four possible results (FAWCETT, 2006). These results are shown in Table 11.

Table 11 – Confusion Matrix

	Labeled		
Predicted		*Positive*	*Negative*
	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

Source: Adapted from Fawcett (2006)

From this matrix, we can calculate classification performance metrics that reflect the performance of the classifier (RUUSKA et al., 2018). Based on the analyzed literature, the most common metrics are:

- **Sensitivity**: model's ability to detect the presence of the disease, and is defined by the function:

$$sen = \frac{TP}{TP + FN} \tag{5.1}$$

- **Specificity**: the model's ability to detect the absence of the disease, and is defined by the function:

$$spec = \frac{TN}{TN + FP} \tag{5.2}$$

- **Accuracy**: model's ability to correctly classify each case, and is defined by the function:

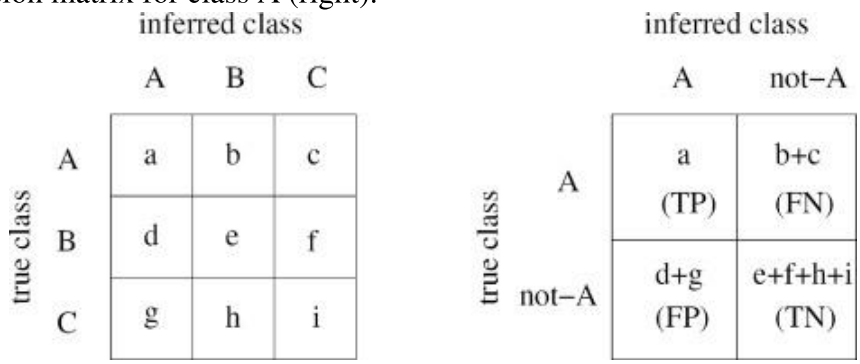
$$acu = \frac{TP + TN}{TP + TN + FP + FN} \tag{5.3}$$

- **F1-Score**: it is a harmonic measure between precision and sensitivity, and is defined by the function:

$$F1 - Score = \frac{2 * TP}{2 * TP + FN + FP} \tag{5.4}$$

In certain steps of the proposed model, the problem to be solved involves multi-classes. Thus, it is necessary to generalize the confusion matrix to use these metrics for the evaluation of the model. Formally, for a data set with C_k classes, an array of size $k \times k$ is defined. Each cell $[i, j]$ represents the observation frequency of the real class C_i and the inferred class C_j. This matrix can be represented in up to k binary confusion matrices, one for each class C_i (RUUSKA et al., 2018). A representation of this transformation can be seen in Figure 45.

Figure 45 – An example of a 3×3 confusion matrix for classes A–C (left) and the corresponding binary confusion matrix for class A (right).



Source: Ruuska et al. (2018).

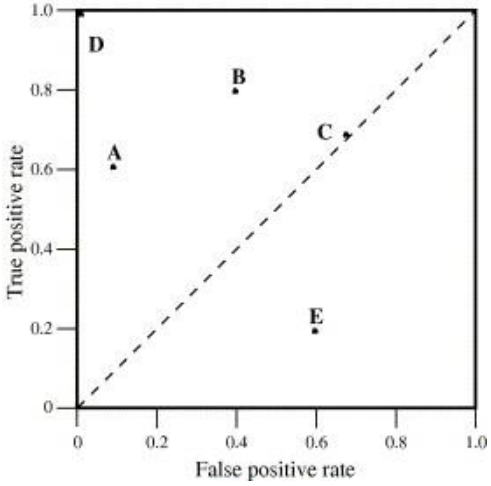
This transformation from a multi-class confusion matrix to the corresponding binary is an excellent tool for visualizing the model's behavior. However, in practice for model evaluation, this transformation is not necessary, is performed automatically by libraries such as *scikit-learn*.

5.2.2 ROC graphs

A graphical way to evaluate classifiers is to use the ROC, a two-dimensional graph in which the rate of true positives is plotted on the Y-axis, and the rate of false positives is plotted on the X-axis (PEDRO; MACHADO-LIMA; NUNES, 2019) axis. Figure 46 shows a ROC chart with five classifiers.

Some points on the graph of the ROC curve need special attention. For example, the $(0, 0)$ point indicates a strategy that always issues a negative rating; the $(1, 1)$ point indicates that the model always generates a positive rating; finally, the $(0, 1)$ point represents the perfect strategy, correctly classifying each instance (RUUSKA et al., 2018). In addition, when the classifier's performance is under the dotted line that divides the graph into two parts, as in Figure 46, its performance can be considered random.

Figure 46 – A basic ROC graph showing five discrete classifiers.



Source: Ruuska et al. (2018).

Connected the points created using different decision limits, we can obtain the AUC (CARTER et al., 2016). Which is used to gauge how well a model can differentiate classes (PEDRO; MACHADO-LIMA; NUNES, 2019). Also, the higher the AUC value, the better the model performance (RUUSKA et al., 2018).

5.2.3 Dice coefficient

Dice coefficient measures the similarity between two sets of data, based on the overlap of the two sets. This metric can be defined as:

$$DSC(A,B) = \frac{2\,|A \cap B|}{|A + B|} \tag{5.5}$$

where A and B are any two sets. Also, the closer to 1.0 the value obtained by the metric, the better the model is generalized.

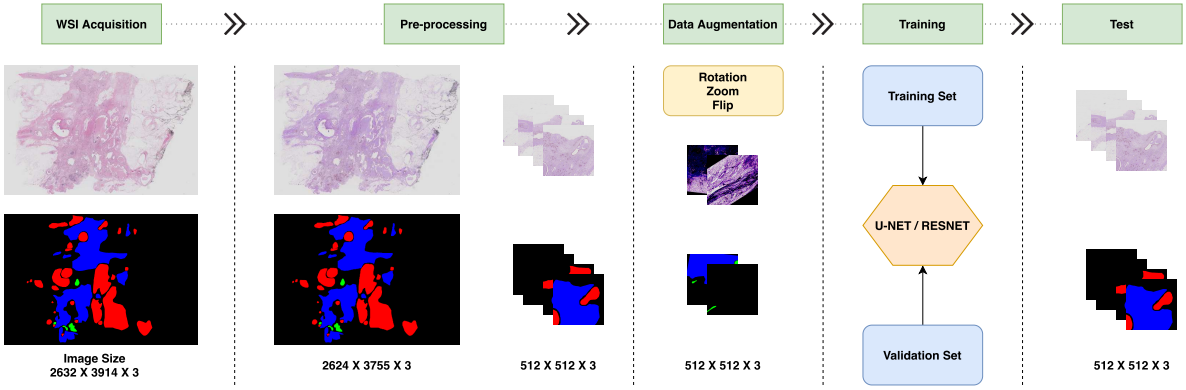
5.3 Comparative Analysis of the Influence of Color Spaces on WSI Segmentation

The colors present in the WSI transmit a large amount of information and play an essential role in the diagnosis. In computing, colors are represented by color spaces, which aim to represent the colors visible to humans. However, there are significant differences in the amount of colors in the visible electromagnetic spectrum that each model is capable of representing (GONZALEZ; WOODS, 2018). In this section, we present the methodology used to analyze the behavior of breast cancer WSI multi segmentation for four-color spaces (RGB, HSV, YCbCr, and LAB) using a ResNet50/U-Net.

The methodology (Figure 47) is divided into four stages: pre-processing, data augmentation, training, and testing. The pre-processing of the images consists of converting the color spaces

and normalizing the images. The data augmentation step describes the methods used to create synthetic images. In the training stage, CNN and the parameters used are defined. Finally, the test step presents the results and behavior of CNN for each color space.

Figure 47 – Diagram of the proposed methodology, based on a U-Net and Resnet model for the WSI segmentation of breast cancer.



Source: Elaborated by the author.

5.3.1 WSI Acquisition and Preprocessing

In this evaluation, we used the BACH (B) dataset with 10 pixel-labeled WSI in four classes: normal, benign, carcinoma in situ, and invasive carcinoma. Although the image base is relatively small, it is essential to highlight the size of these images with billions of pixels (ARESTA et al., 2019). In this way, transfer learning and data augmentation techniques are useful to assist the process of learning models with datasets with few images.

For pre-processing, we use WSI in a 2.5× magnification. The use of WSI in this magnification is due to the computational capacity available. As a way to speed up the WSI segmentation process, we removed the white borders present in the WSI using the adaptive segmentation of Otsu (1979) to separate the bottom of the tissue. After segmentation, the images are normalized by the mean and standard deviation of the overall pixel intensity (SPANHOL et al., 2016a).

The segmented images were divided into three sets, training (70%), validation (10%), and testing (20%). CNNs need a significant amount of images in the training process to obtain a generalization of the problem. Thus, each set went through the extraction of non-overlapping patches of 256×256 pixels.

5.3.2 Data augmentation

One technique that helps in the convergence and learning of CNN is the use of data augmentation. At this stage, we use the *ImageDataGenerator* class of the framework *Keras* applied to the training and validation sets. The class in question generates synthetic data in

real-time based on the images extracted from the WSI. The transformations applied to the patches in this step were: rotations, zoom, and horizontal and vertical inversions.

5.3.3 Convolutional Model

Figure 37 show the convolutional architecture used in this step. The encoder used for this step was pre-trained for the ImageNet image set. The learning rate used in the training was 0.0001, a *batch size* of 16 images, *padding same* for the convolutional and *maxpooling* layers and the *Adam* algorithm for optimization of the weights.

5.4 DeepBatch Experimental Setup

After analyzing the behavior of the color spaces in the WSI multi segmentation for breast cancer, we concentrated our efforts on the details of the DeepBatch model. The implementation of each of the modules of the DeepBatch model architecture was developed in the Python language. In the next subsections, we present this implementation in technical detail.

5.4.1 Datasets Distribution and Analysis

A well-prepared data set is a crucial factor in defining a good generalization for DL techniques. In this way, we use different sets of information from the dataset, following the objectives of the modules. In Table 12, we present the datasets and what information is used in each module. These datasets were joined into a single set of images and divided randomly into sets of training, testing, and validation.

Table 12 – Public datasets used in DeepBatch modules.

Module	**Dataset**	**WSI**	**Number of images**	**Annotation type**
ROI Detection	HASHI	No	195	pixel-wise
	BACH (B)	Yes	10	pixel-wise
	TCGA	Yes	195	-
Cell Segmentation	BACH (B)	Yes	10	pixel-wise

Source: Elaborated by the author.

After choosing the datasets, we analyze quantitative measurements of the WSI of the selected datasets. In Table 13, we present the maximum, minimum, and average values for height, width, and the number of pixels present in the WSI.

In Figure 48, we present the distribution of the images concerning height and width. Based

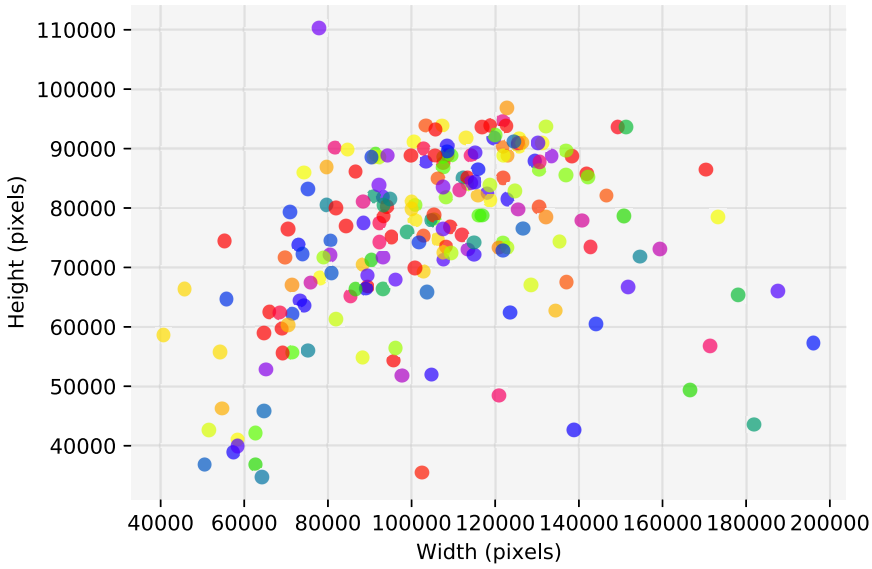
Table 13 – Whole-slide imaging statistics used in DeepBatch modules.

Measure	Size in pixels
Max width	196,111
Max height	110,336
Max size	14,733,475,680
Min width	40,704
Min height	34,700
Min size	1,861,134,657
Avg width	105,334
Avg height	75,386
Avg size	8,073,512,637

Source: Elaborated by the author.

on this distribution, we can see that most WSI are concentrated in the range of 60,000 to 100,000 pixels in height and 80,000 to 140,000 pixels in width.

Figure 48 – Distribution of WSI according to height and width in pixels.



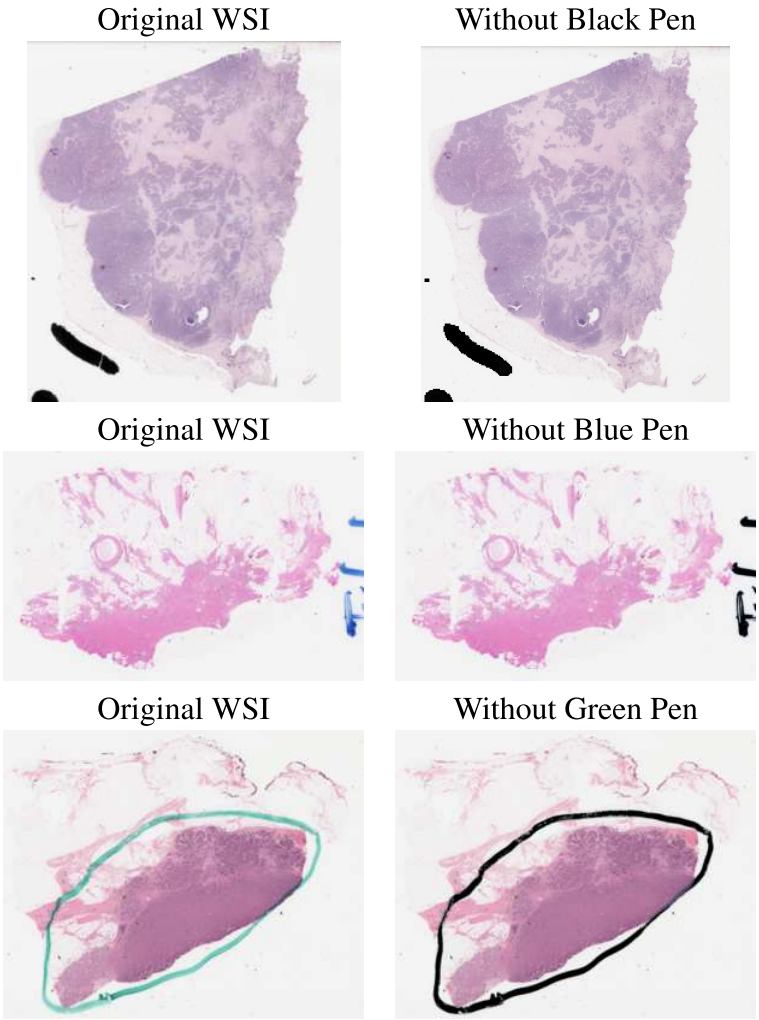
Source: Elaborated by the author.

5.4.2 Preprocessing Module

As already mentioned, this module is responsible for preparing WSI for the overall analysis of cases. Thus, the entry in this module is WSI in magnification of 40×. In these WSI, we apply a 32:1 downsampling, as in works identified in the literature (CRUZ-ROA et al., 2018; GUO et al., 2019). In some of these WSI, markings are found with pens in the colors: black, blue, and green. These demarcations sometimes cover regions noted as cancerous. This feature can cause noise on CNN. In this way, we apply a set of thresholds in each RGB channel to

remove the demarcations. In Figure 49, we present an example for each pen color. In Table 14, we present the thresholds used for each channel. The thresholds used were based on the work of Dusenberry e Hu (2019). For this, intervals were obtained for the pixel values of the demarcations in black, blue, and green. Based on these values, we defined a set of thresholds that suited the task of removing the demarcations without affecting areas of tissue.

Figure 49 – Examples of removing pen markings in WSI. On the left, black, blue, and green markings. On the right, the regions removed in black pixels.



Source: Elaborated by the author.

The next step was to remove the bottom of the WSI using the adaptive segmentation of Otsu (OTSU, 1979). Based on the coordinates found in the background removal, we cut the white edges present in the WSI. As a result, the percentage of tissue in the WSI increased from 32.18% to 40.02%. In Figure 50, we can see an original WSI with background removal and cutting the edges.

The masks generated during the pre-processing of the WSI were also applied to the base annotation images. This process aims to avoid noise on the network, as the segments provided are not perfectly delimited with the tissue. In Figure 51, we can see the difference between

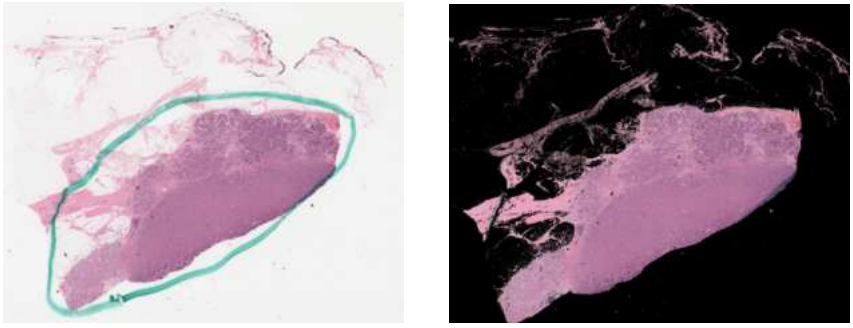
Table 14 – Set of thresholds used to remove pen demarcations in WSI. The columns R, G, and B represent the operator and the threshold applied to each of the pixels in the RGB color space.

Filter for	**R**	**G**	**B**
Black Pen	< 128	< 128	< 128
Blue Pen	> 60	> 120	< 190
	> 120	> 170	< 200
	> 175	> 210	< 230
	> 145	> 180	< 210
	> 37	> 95	< 160
	> 30	> 65	< 130
	> 130	> 155	< 180
	> 40	> 35	< 85
	> 30	> 20	< 65
	> 90	> 90	< 140
	> 60	> 60	< 120
	> 110	> 110	< 175
Green Pen	> 150	< 160	< 140
	> 70	< 110	< 110
	> 45	< 115	< 100
	> 30	< 75	< 60
	> 195	< 220	< 210
	> 225	< 230	< 225
	> 170	< 210	< 200
	> 20	< 30	< 20
	> 50	< 60	< 40
	> 30	< 50	< 35
	> 65	< 70	< 60
	> 100	< 110	< 105
	> 165	< 180	< 180
	> 140	< 140	< 150
	> 185	< 195	< 195

Source: Elaborated by the author.

Figure 50 – An example of WSI resized after going through the Preprocessing Module. In brackets, the size of each WSI.

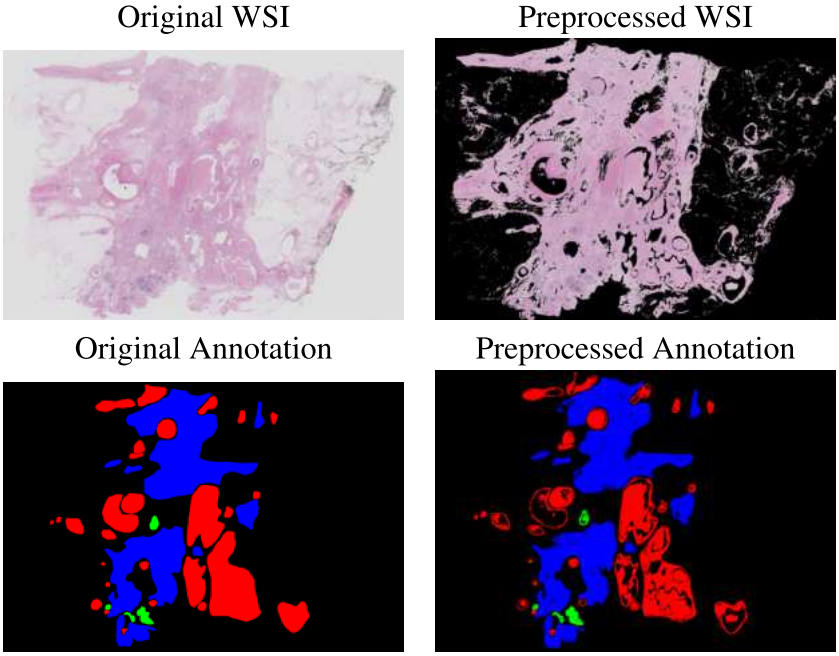
Resized WSI (2824×3927) Preprocessed WSI (2481×3597)



Source: Elaborated by the author.

using this step or not. In addition, it is essential to note that there are many non-tissue regions removed from the original annotation in this process.

Figure 51 – An example of an annotation after going through the Preprocessing Module.



Source: Elaborated by the author.

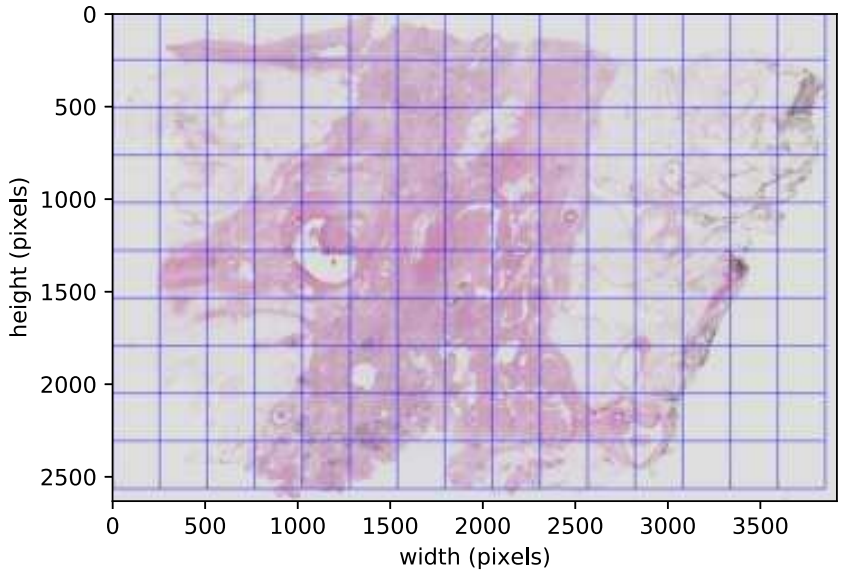
5.4.3 ROI Detection Module

The ROI Detection module is responsible for generating a global segmentation of the WSI. In this module, the images produced by the Preprocessing Module were divided into three sets: training (70%), validation (10%), and test (20%). However, even at low magnification, WSIs are still considered high-resolution images for direct processing on a CNN. Also, WSI has a large proportion of non-tissue information. So, as a way to solve these two problems for training and validation sets, we extracted from the 256×256 pixels tiles images. In Figure 52, an example of this sample of tiles is presented. Also, as a way to decrease the percentage of non-tissue regions, we only use tiles with more than 10 % of the pixels representing tissue.

This processing resulted in 26,016 tiles. For the training of convolutional architecture, we use the $ImageDataGenerator$ class from the $Keras$ framework. This class generates synthetic images on-the-fly based on the tiles. The transformations applied to the tiles were rotation, zoom, and inversions and vertical. For each training epoch, 40 thousand tiles were produced. It is important to note that a new set is generated at each time, which makes training more robust since a different collection of images is presented at each interaction.

The convolutional architecture used in this module was presented in Section 4.2.2 in Figure 33. The values for the meta parameters used for U-Net were:

Figure 52 – Tile sampling process for the ROI Detection module. Each rectangular region is a 256×256 pixels image generated and used in the training and validation steps as a basis for data augmentation.



Source: Elaborated by the author.

- **Learning rate:** 0.0001;
- **Batch size:** 16 images;
- **Padding:** same;
- **Optimizer:** Adam;
- **Loss function:** Binary cross-entropy.

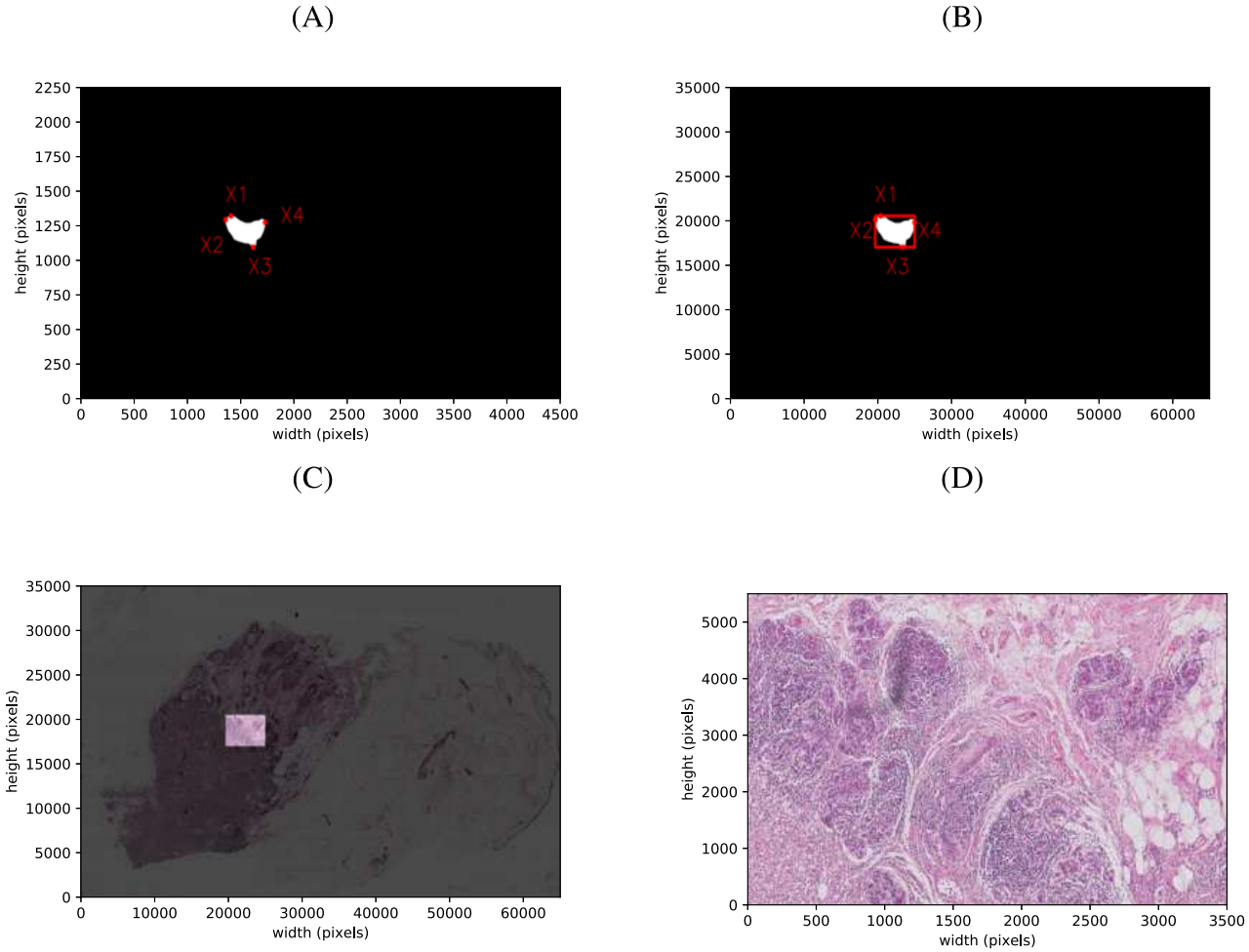
5.4.4 Sampling Module

This module is responsible for mapping the regions identified as suspicious by the ROI Detection Module from areas at low magnifications to high magnifications ($\approx 60K \times 40K$). This mapping occurs by collecting the major and minor coordinates of the region targeted for the x and y axes. In Figure 53, we present an example of this mapping.

5.4.5 Cell Segmentation Module

In this module, the inputs are regions in high resolution and of different formats. These regions were grouped by case into three sets of training (70%), validation (10%), and test (20%). The regions that make up the sets are made up of images of different formats. Thus, we adopted a process similar to that performed in the ROI Detection Module, extracting a set of 256×256

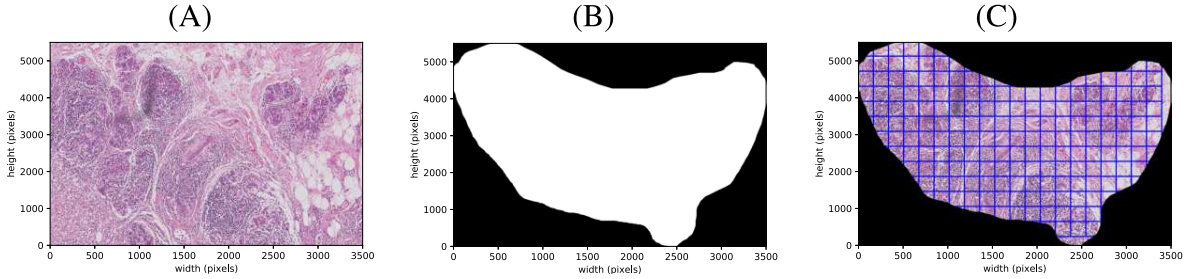
Figure 53 – Sampling process from low to high magnification regions. (A) Region segmented by the ROI Detection Module with points with greater and lesser coordinates for the y-axis (X1, X3) and x-axis (X2, X4). (B) The rectangular region in high magnification obtained based on the coordinates identified in step (A). (C) The rectangular region highlighted in WSI. (D) Region sampled at 40× magnification that will be used as an input to the Cell Segmentation Module.



Source: Elaborated by the author.

pixels from the segmented regions. In Figure 54, we present an example of generating tiles based on an area of a WSI at a magnification of 40×.

Figure 54 – Tile sampling process for the Cell Segmentation module. (A) Tissue section at 40× magnification. (B) Segmentation produced by the ROI Detection Module. (C) Segmented tissue with tiles of 256×256 pixels.



Source: Elaborated by the author.

As mentioned in Section 4.2.4, the purpose of this module is to generate a refined, multiclass segmentation for breast cancer WSI. Thus, since multiclass segmentation requires a more substantial data set to produce a significant result, we opted for the use of transfer learning and adapted a ResNet50 network for semantic segmentation. ResNet50 was used as an encoder, extracting resources that will be used by a decoder based on U-Net to produce a multiclass segmentation. The convolutional architecture used in this module was presented in Figure 37. The values for the meta parameters used for ResNet50/U-Net were:

- **Learning rate:** 0.00001;
- **Batch size:** 8 images;
- **Padding:** same;
- **Optimizer:** Adam;
- **Loss function:** (0.5 * Categorical cross-entropy) + 0.5 * (Dice-Index).

5.5 Final Remarks

This chapter presented the methodology for this Master Thesis. Potential datasets were introduced that could be used together or separately for each module, in accordance with the module's objective. We presented the methodology adopted for the pre-processing of WSI. We define the characteristics and meta parameters for implementing the ROI Detection module. Finally, we present the methodology for mapping low-magnification regions to high magnifications using the Sampling module, as well as defining the implementation of the Cell Segmentation module.

The proposed methodology aims to define the experiments for the validation of the DeepBatch model. We can highlight two relevant points consistent with our scientific contribution. The first is related to the analysis of WSI. We are allowing, through CNN, the complete analysis of each case. The second point, on the other hand, is how the diagnostic aid information is transmitted to the pathologist. We propose the use of semantic segmentation. Semantic segmentation can offer the pathologist two important information in decision making, the separation of tissue by classes and heat maps that can indicate border regions. Finally, this work is limited to the analysis of WSI stained with H&E breast cancer. In the next chapter, we present the results and discussions based on this methodology.

6 RESULTS AND DISCUSSIONS

This chapter presents the results and discussions for the experiments carried out to validate the DeepBatch. The methodology used to carry out these experiments is described in Chapter 5. The experiments described in this chapter were performed on a system that contains a 12 GB NVIDIA TITAN XP GPU, Intel Core i7-9700K CPU, and 32 GB of RAM. The chapter is divided into four sections. Section 6.1 presents the results and discussions for analyzing the influence of color spaces on WSI multi segmentation. In Section 6.2, we evaluate the performance of the ROI Detection module. Then, in Section 6.3, experiments for the Cell Segmentation Module are evaluated. Finally, in Section 6.4, we present a comparison of the results obtained with the related works identified in the SLR.

6.1 Comparative Analysis of the Influence of Color Spaces on WSI Segmentation

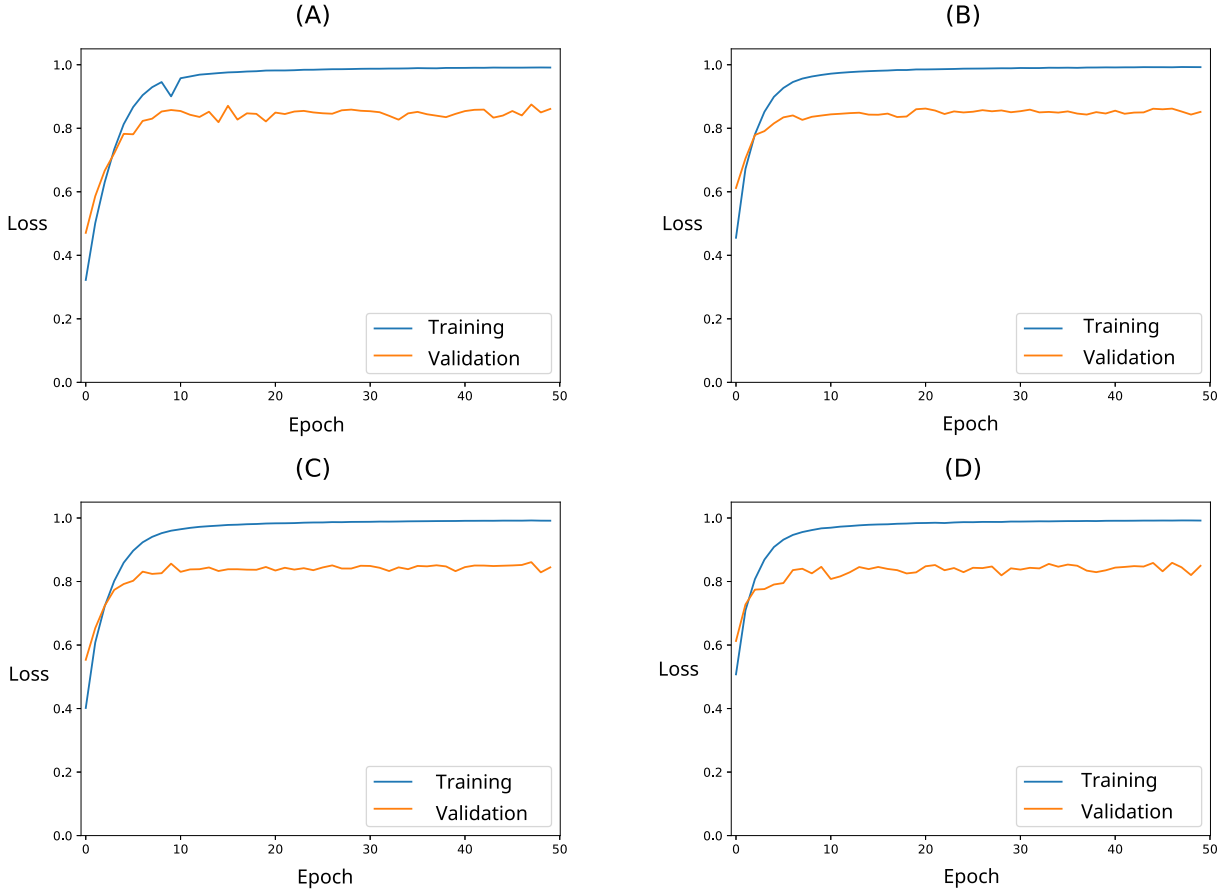
The preparation, coloring, and digitization processes of tissue slides are susceptible to inadequate fixations, which can generate changes in tissue morphology and induce incorrect tissue morphometry through image analysis (ROY et al., 2019). Another difficulty encountered are sections that have not been adequately stained (these may be over or under-stained) and assembled, this may result in regions out of focus, or missing parts (LI et al., 2018). Also, slight variations in coloring conditions can lead to changes in the appearance of the tissue (YAN et al., 2019). In this way, colors and textures can directly influence the behavior of classification, detection, or segmentation architectures. Computationally, colors are represented only by numbers. However, there are several formats to perform this representation, the most common being RGB, HSV, YCrCb, and LAB. In this section, we present and evaluate the results obtained using the methodology proposed in Section 5.3 the proposed convolutional architecture (Figure 37) for each of the selected color spaces (RGB, HSV, YCrCb, and LAB).

6.1.1 Training and Validation of CNNs

During the training of the CNNs for each color space, the metric used to measure the loss each epoch was the Dice Index. The convolutional architectures were trained for 50 epochs based on the training set and validated at the end of each epoch in the validation set to obtain the best set of weights. Figure 55 shows the indexes obtained for each epoch in the training and validation set.

All CNNs showed statistically similar behaviors for the training and validation set. For the four-color spaces tested, we can see relative stability in the growth of the Dice Index in the training set and small variations in the validation set. This behavior indicates that there was no overfitting during training.

Figure 55 – Loss graphs for training and validation data for each CNNs: (A) RGB; (B) HSV; (C) YCrCb; (D) LAB.



Source: Elaborated by the author.

6.1.2 CNNs Evaluation

The results for the CNNs evaluation stage are based on the set of tests for each color space. The choice of the best set of weights was made automatically based on the loss in each epoch for the validation set. Table 15 shows the values obtained for the metrics at the pixel level and the Dice Index in the test set.

Table 15 – Results for the test set for each color space: (A) RGB; (B) HSV; (C) YCrCb; (D) LAB. In bold, the best results. (*Accu. pixel* is the accuracy at the pixel level. The columns *Benign, In situ, Invasive, and Background* correspond to the percentage of pixels correctly classified for each of these classes).

CNN	Dice Index	Pixel accu.	Benign	In situ	Invasive	Background
A	0.6923	84.67%	78.22%	**82.63%**	76.64%	91.49%
B	**0.7085**	83.85%	**80.95%**	72.73%	**89.83%**	90.51%
C	0.6881	**84.85%**	71.56%	74.17%	81.84%	**93.70%**
D	0.6979	83.80%	73.45%	75.03%	80.81%	91.75%

Source: Elaborated by the author.

With the results obtained for the Dice Index metric, we can see, again, the statistical similarity between the indices. The largest variation between the Dice Index is 2.04%, between the YCrCb and the HSV. This similarity is reflected in the percentages obtained for accuracy at the pixel level, which presents a variation of at most 1.05% between color spaces.

However, when we analyze the percentages at the pixel level for each class, this similarity does not occur. Where, for example, for the classification of benign regions, the percentage variation is 9.39% between color spaces. As for the tissue affected by invasive carcinoma, there is a better classification for the convolutional architecture of the HSV space with a percentage of 89.83% accuracy. In the classification of areas with carcinoma in situ, the RGB space proved to be more effective for detecting pixels belonging to diseased regions.

In general, the convolutional architecture for the HSV color space showed the best quantitative results. However, it demonstrates some difficulty in detecting benign tissue. Another conclusion that we can have is that even presenting the best accuracy at the pixel level, the YCrCb convolutional architecture presented significantly lower results for the regions with cancer. This is a reflection of the high precision for the classification of pixels as background, which are proportionally larger areas in the images, reflecting a higher weight for the overall accuracy.

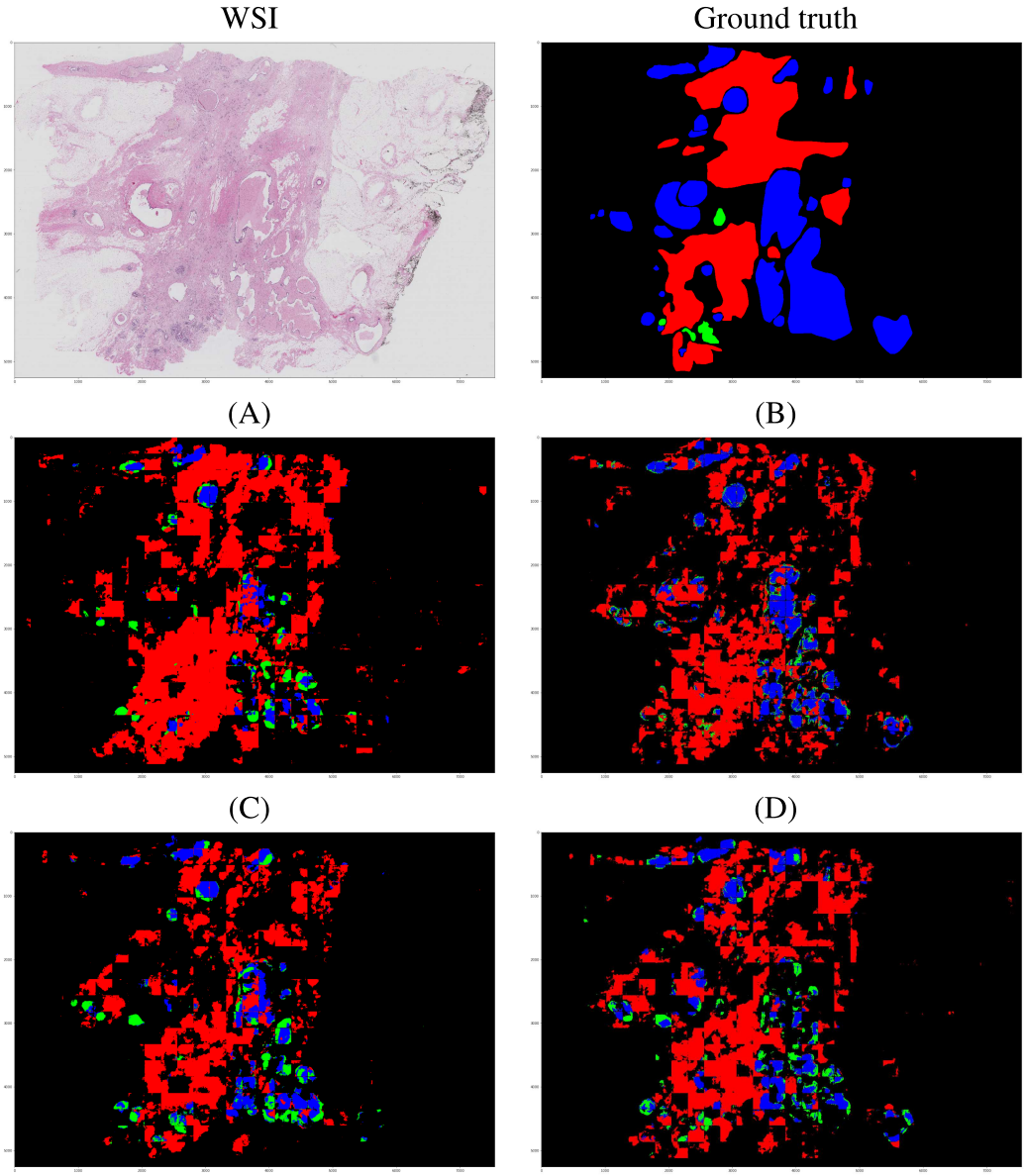
6.1.3 Qualitative Evaluation

In this section, we present the qualitative results for the multi-segmentation of a WSI (Figure 56) of the test set for each of the color spaces. In Figure 56, the black color indicates the background or healthy tissue. The blue color indicates the presence of tissue with benign characteristics. The green color indicates the presence of carcinoma in situ. The red color indicates the presence of invasive carcinoma.

Analyzing the qualitative performance, the color space that presents the best result is the convolutional architecture for the HSV color space (convolutional architecture (B)). The best representation, whether quantitative or qualitative, can be associated with the way that colors are represented in the HSV space, where one channel only represents the base color, and the other two channels by brightness and color saturation (GONZALEZ; WOODS, 2018). In this way, the convolutional architecture can be more robust for small variations, since the Hue channel presents color in its primary form, without alterations in brightness or saturation decreasing.

Specifically, for case A01, it is possible to notice that convolutional architectures A, C, and D detect regions with carcinoma in situ more efficiently. Meanwhile, convolutional architecture B has significantly higher accuracy for detecting benign areas and invasive carcinoma.

Figure 56 – BACH case A01 with the segmentation provided by the base and the results for the CNNs: (A) RGB; (B) HSV; (C) YCrCb; (D) LAB. In the annotations, the black color indicates healthy tissue or background, blue benign tissue, green in-situ carcinoma, red invasive carcinoma.



Source: Elaborated by the author.

6.2 ROI Detection Module

CNN-based architectures are considered, based on the SLR, the state of the art for classification, detection, or segmentation. CNNs, in comparison with resource extraction techniques, have as their main advantage the ability to learn the characteristics necessary for inference. However, to perform the training and inference of CNNs, significant computational processing is required. In this way, WSI processing, with billions of pixels, is not possible directly on a CNN. As a way to reduce the consumption of computational resources, we analyzed the performance of the convolutional architecture (Figure 33 in low-resolution WSI

applying the methodology presented in Section 5.4.3 to identify suspicious regions. In the next subsections, we present quantitatively and qualitatively the results obtained for the ROI Detection Module.

6.2.1 Training and Validation

The convolutional architecture for the ROI Detection Module was trained for 100 epochs in the selected datasets (Table 12). During the training, synthetic samples were generated based on tiles extracted from WSI. For each epoch, a total of $78,048$ synthetic tiles of 256×256 pixels were generated on-the-fly for architecture training. In the validation set, we do not apply data augmentation techniques, thus totaling $6,824$ tiles in the set. At the end of each epoch, we evaluate the performance of the architecture using the loss functions and a set of metrics. In Figure 57, we present the indexes obtained for each of the 100 epochs.

Following the behavior of the convolutional architecture for the values of the loss function at each time (Figure 57 (A)), we can see that CNN did not suffer from overfitting. For the metrics in the validation set (Figure 57 (C)), we can see significant variations each epoch, with a slight upward trend at the end of the 100 epochs. This behavior can reflect the learning process of architecture, refining its weights at each interaction with the training set.

6.2.2 Evaluation

In addition to measuring the behavior of loss function and metrics each epoch to assess the performance of a convolutional architecture, it is important to measure performance based on assessment metrics on an unused dataset during training and validation. In this step, we use a percentage of 20% of the selected WSI to evaluate the ROI Detection Module (Table 12). The weights used for the evaluation of the convolutional architecture refer to the epoch with the lowest loss for the validation set. It is important to note that the WSI used in the test stage did not go through the tile sampling stage. Thus, the entrance to convolutional architecture is a WSI with 32:1 downsampling. This is an interesting feature of networks composed only of convolutional layers, which allow the use of images of any dimension without affecting the inference. In Table 16, we present the percentages and indices obtained for the metrics in the test set. In Figure 58, we present the ROC curve for the ROI Detection Module.

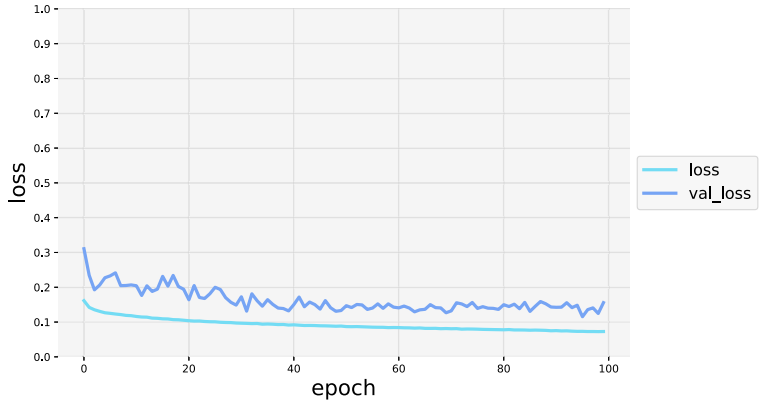
Table 16 – Results for the performance metrics in the test set for the ROI Detection Module.

IoU	**Accuracy**	**Sensitivity**	**Specificity**	**F1-Score**	**AUC**
93.43%	91.27%	90.77%	94.03%	84.17%	0.93

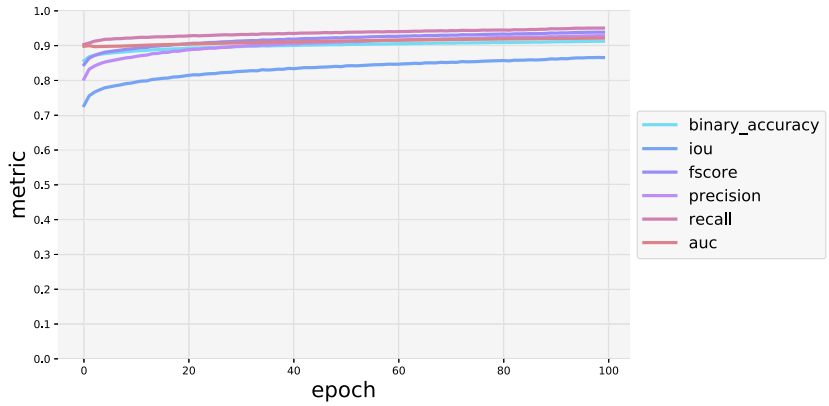
Source: Elaborated by the author.

Figure 57 – Graphs for loss and metrics for each epoch. (A) Loss over epochs; (B) Metrics over epochs for the training set; (C) Metrics over epochs for the validation set.

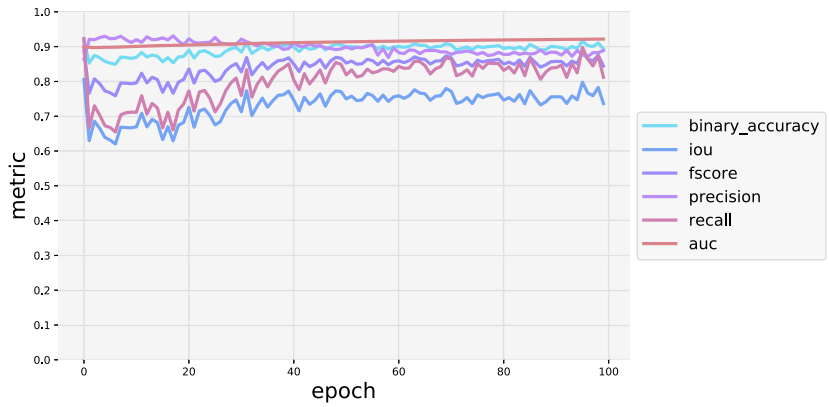
(A)



(B)

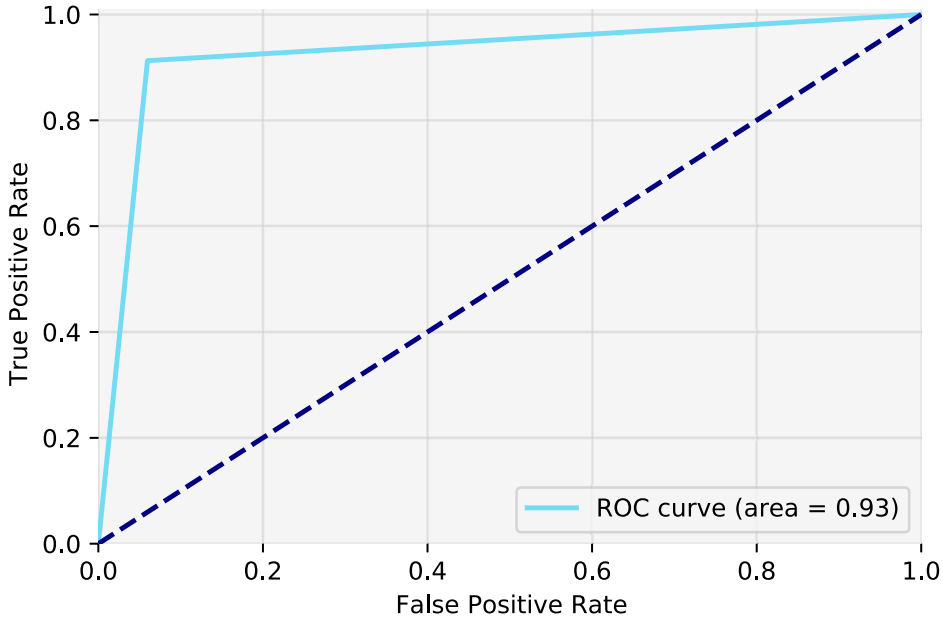


(C)



Source: Elaborated by the author.

Figure 58 – The performance of mass segmentation in terms of ROC curves in test set for the ROI Detection Module.



Source: Elaborated by the author.

6.2.3 Qualitative Evaluation

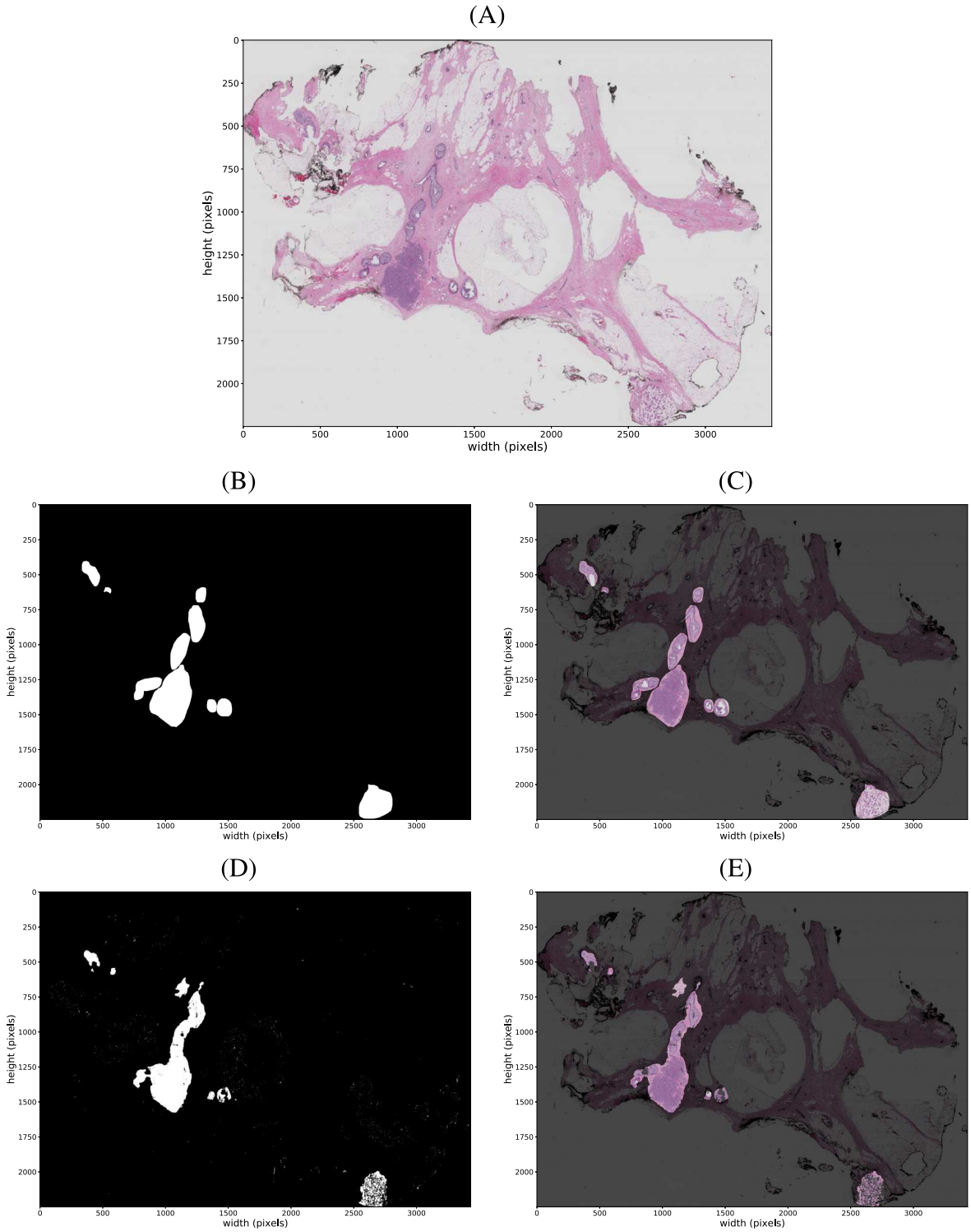
In this section, we present results for the segmentation of suspicious regions in WSI. In Figures 59 and 60, the input WSI of the DeepBatch, the annotation of the pathologist, and the annotation inferred by the computational architecture are presented. In the annotations, the color black indicates a healthy tissue or the background. In contrast, the white color indicates a suspicious region, which deserves attention in the next stages.

In Figure 59, we can see a relative similarity between the regions labeled by the pathologist and those inferred by the convolutional architecture. In comparison with the pathologist's annotations, we can see the ability of convolutional architecture to separate background pixels within suspect regions. Figure 60, on the other hand, presents an example with suspicious areas that can be considered FP compared to the pathologist's label. However, as this module aims to serve as a filter for the next phase, this behavior is desired and will be treated in Cell Segmentation Module.

6.3 Cell Segmentation Module

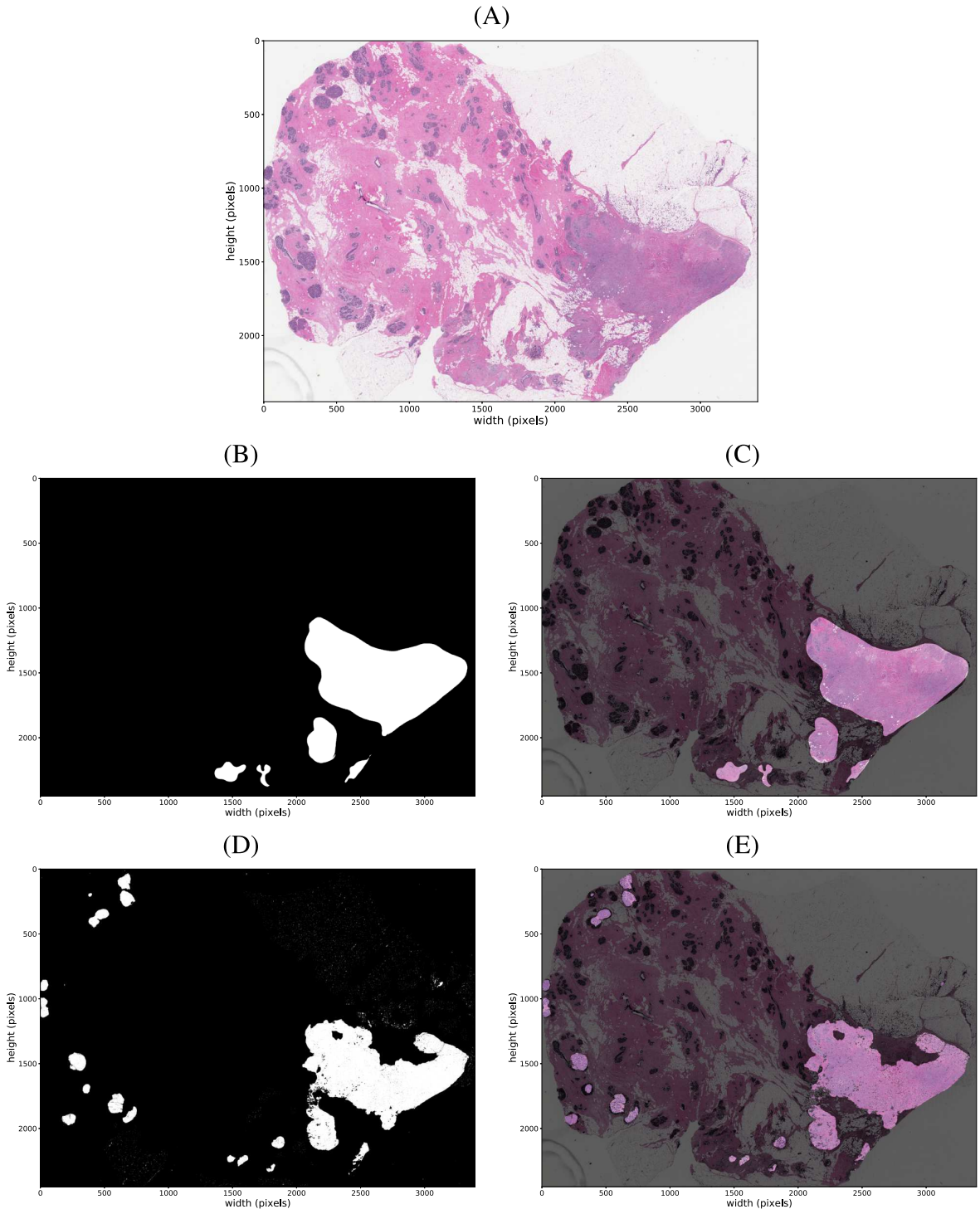
The size of the WSI reflects the amount of information present in these images. The interaction of H&E dyes makes it possible to identify and evaluate even cell functional issues (CHAN, 2014; MITTAL; SARASWAT, 2019). Thus, for the identification of neoplasic regions in a WSI, detailed analysis at high magnifications is necessary. Therefore, in this section, we present the performance of the convolutional architecture (Figure 37) in suspicious regions in

Figure 59 – BACH case A03. (A) WSI input for inference. (B) Pathologist annotation. (C) WSI with pathologist annotation highlighted. (D) Annotation produced by the ROI Detection Module. (E) WSI with inferred annotation highlighted. In the annotations, the black color indicates a healthy tissue or background. The white color indicates suspicious tissue.



Source: Elaborated by the author.

Figure 60 – TCGA case TCGA-A2-A0CZ. (A) WSI input for inference. (B) Pathologist annotation. (C) WSI with pathologist annotation highlighted. (D) Annotation produced by the ROI Detection Module. (E) WSI with inferred annotation highlighted. In the annotations, the black color indicates a healthy tissue or background. The white color indicates suspicious tissue.



Source: Elaborated by the author.

an magnification of $40\times$ applying the methodology presented in Section 5.4.5 for multi segmentation of WSI. The classes analyzed in this experiment are grouped into four groups: healthy tissue and background, benign tissue, carcinoma in situ, and invasive carcinoma. In the next subsections, we present quantitatively and qualitatively the results obtained for the Cell Segmentation Module.

6.3.1 Training and Validation

For the segmentation of regions in high magnification, the convolutional architecture (Figure 37) was trained for 50 epochs in the selected dataset (Table 12). The weights for this architecture were pre-trained for ImageNet set. The WSI were divided into tiles of 256×256 for the training and validation sets, according to the methodology detailed in Section 5.4.4. For the training set, we generated images on-the-fly, totaling $62,476$ images for each epoch. The validation set was made up of $27,275$ images. In Figure 61, the results for the loss function are presented, and the performance metrics for each epoch during training and validation.

The learning, according to the values for the loss in Figure 61 (A), in the convolutional architecture, occurred more accentuated up to epoch 10. After, there was a relative stabilization of the network with less significant variations. Significant variation also occurs for performance metrics, especially for the training set (Figure 61 (B)). As for the performance metrics in the validation set (Figure 61 (C)), we can see a significantly greater variation when compared to the training set.

6.3.2 Evaluation

The evaluation of the convolutional architecture for the Cell Segmentation Module occurred for the best set of weights obtained in the training phase based on the loss obtained in the validation set. At this stage, we use a 20% data set selected for the Cell Segmentation Module (Table 12). The selection of suspicious regions occurred according to the methodology presented in Section 5.4.4. The difference at this stage for the methodology was the way in which the images were sampled. The tiles were sampled from the WSI in sizes of 2048×2028, increasing CNN's visual field by 8 times for each inference. In Table 17, we present the performance metrics obtained for the test set. In Figure 62, we present the ROC curve for the Cell Segmentation Module.

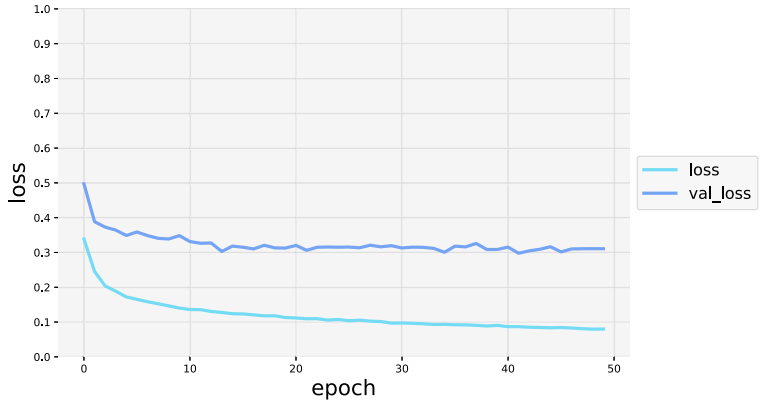
Table 17 – Results for the performance metrics in the test set for the Cell Segmentation Module.

IoU	**Accuracy**	**Sensitivity**	**Specificity**	**F1-Score**	**AUC**
88.23%	96.10%	71.83%	96.19%	82.94%	0.86

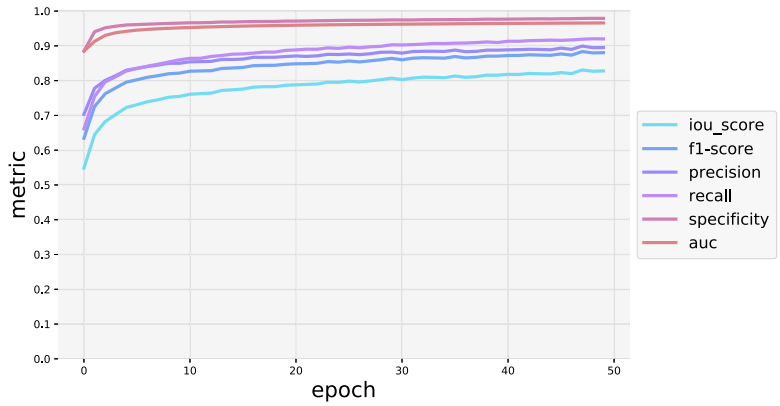
Source: Elaborated by the author.

Figure 61 – Graphs for loss and metrics for each epoch. (A) Loss over epochs; (B) Metrics over epochs for the training set; (C) Metrics over epochs for the validation set.

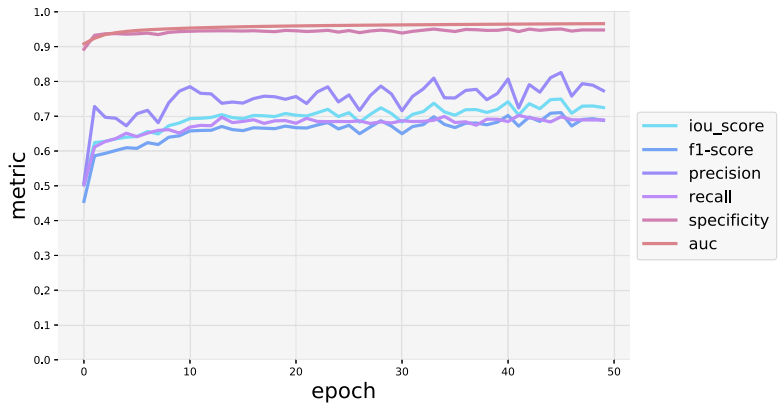
(A)



(B)

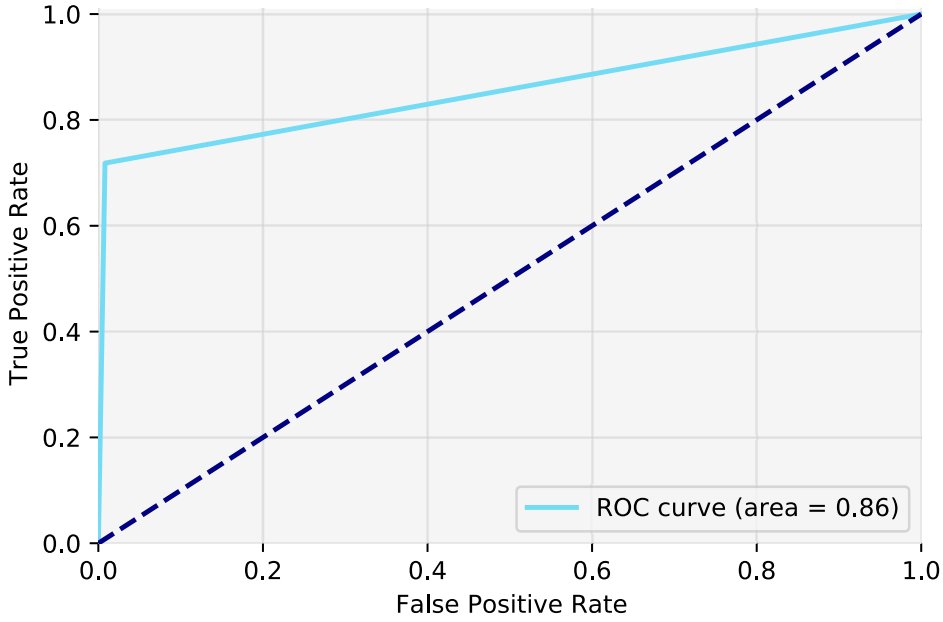


(C)



Source: Elaborated by the author.

Figure 62 – The performance of mass segmentation in terms of ROC curves in test set for the Cell Segmentation Module.



Source: Elaborated by the author.

6.3.3 Qualitative Evaluation

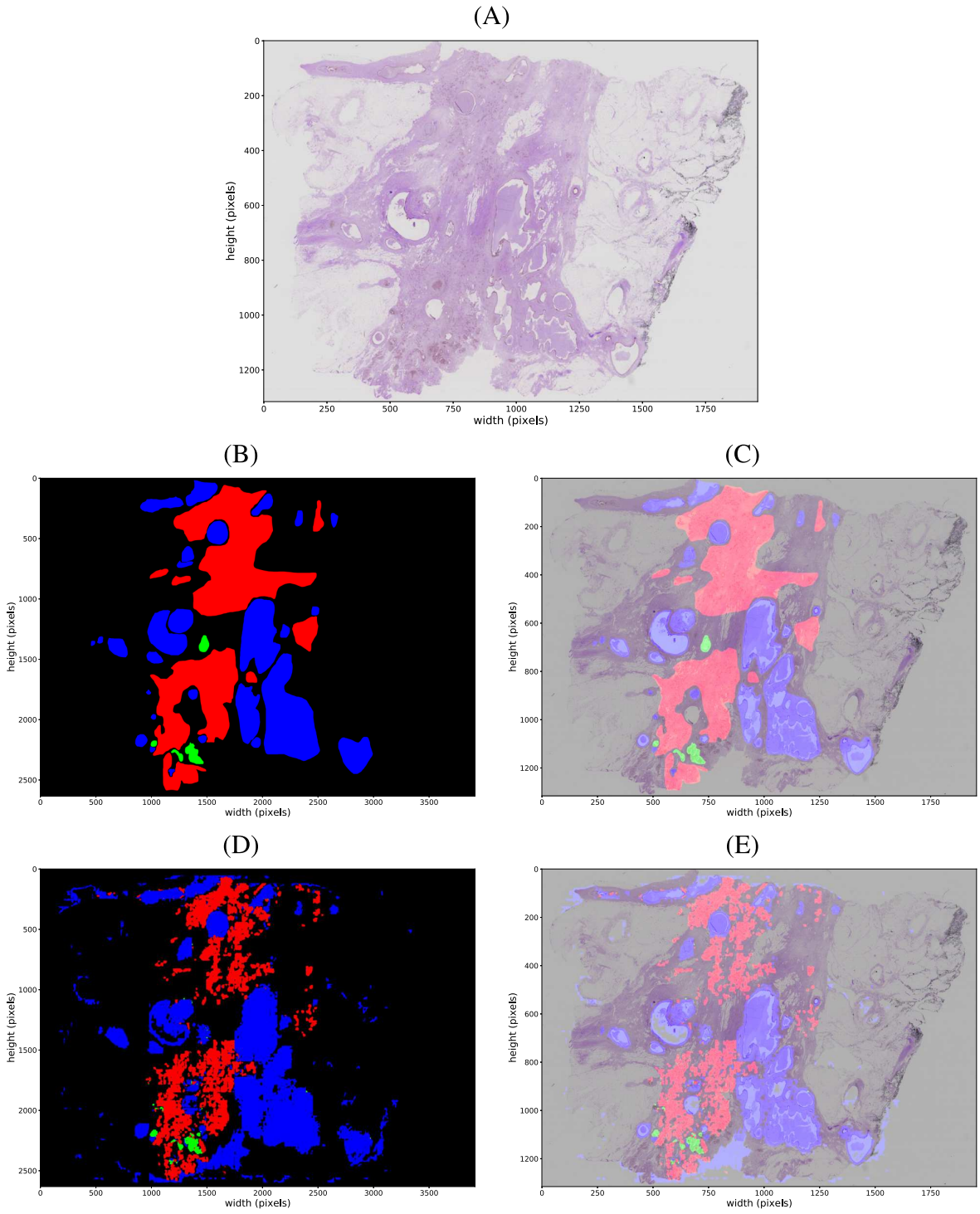
In this section, we present qualitative results for the multi-segmentation of WSI produced in the Cell Segmentation Module. Two examples for segmentation are presented in Figures 63 and 64. In the figures, the black color indicates the background or healthy tissue. The blue color indicates tissue with benign characteristics. The green color indicates the presence of carcinoma in situ. The red color indicates the presence of invasive carcinoma.

In Figure 63 and 64, we can observe a relative similarity with the labels considered gold standard. We can also highlight the ability of convolutional architecture to identify microscopic regions (if compared to the size of WSI), as is the case of regions with carcinoma-in-situ (green color) in Figure 63. However, unwanted behavior is presented by the convolutional architecture, especially in Figure 63, classifying background regions of the WSI as benign areas.

In Figure 63, in the center of the golden standard image, we can see a region with in-situ carcinoma (green color). The convolutional architecture demarcates this same region as being of benign characteristics. This misclassification can indicate two things: (a) the area may be mislabeled, and (b) it may be necessary to use more regions in-situ in the dataset for better generalization. A viable solution would be to pre-train the encoder as a patch classifier using other datasets. Then, use the weights to perform the segmentation of breast cancer WSI.

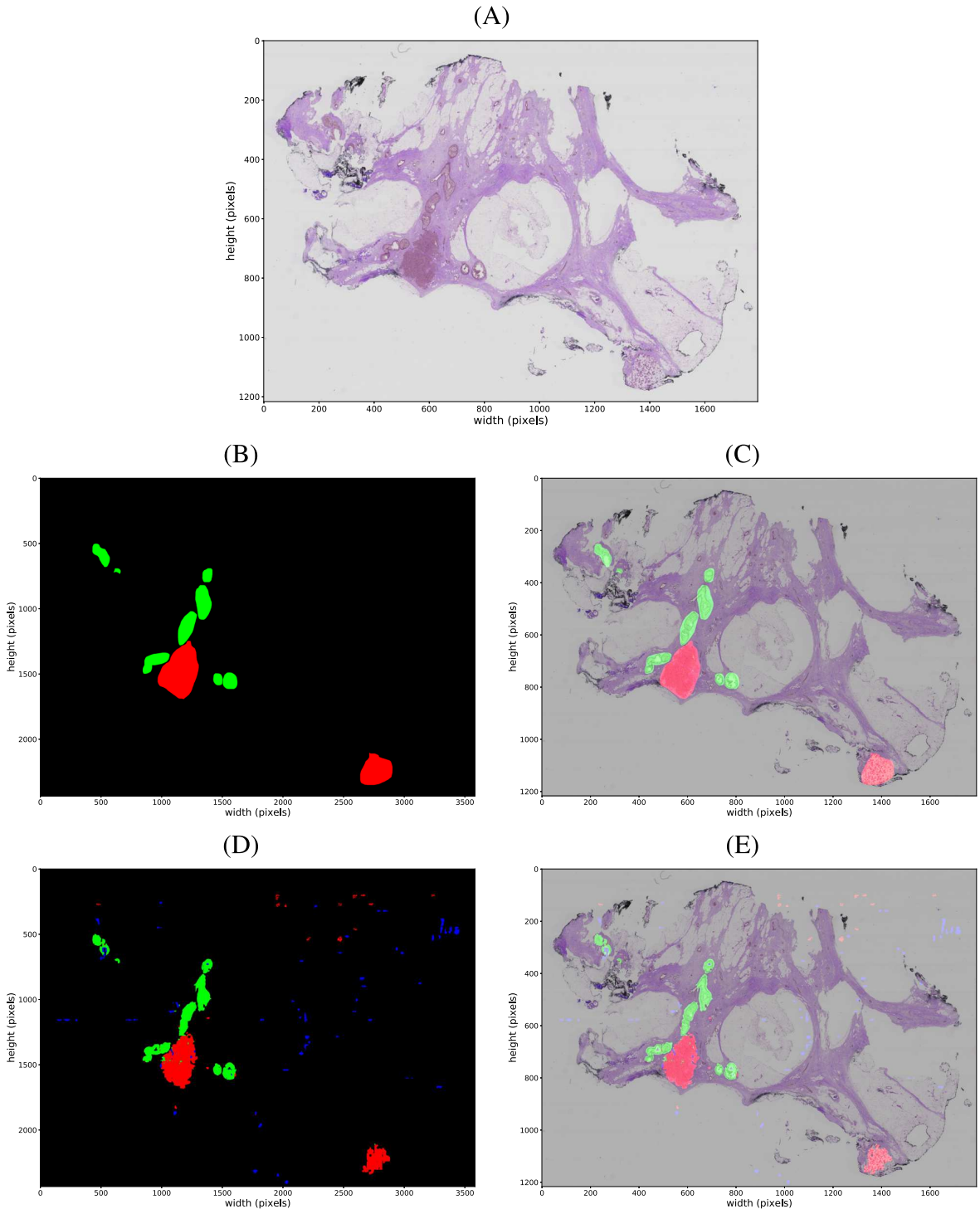
Another way to analyze the inferences produced by the convolutional architecture for the Cell Segmentation Module is through heat maps. In Figures 65 and 66, we present the respective heat maps for cases A01 and A03 of the BACH dataset. Heat maps can be useful in cases of

Figure 63 – BACH case A01. (A) WSI input for inference. (B) Pathologist annotation. (C) WSI with pathologist annotation highlighted. (D) Annotation produced by the Cell Segmentation Module. (E) WSI with inferred annotation highlighted. In the annotations, the black color indicates healthy tissue or background, blue benign tissue, green in-situ carcinoma, red invasive carcinoma.



Source: Elaborated by the author.

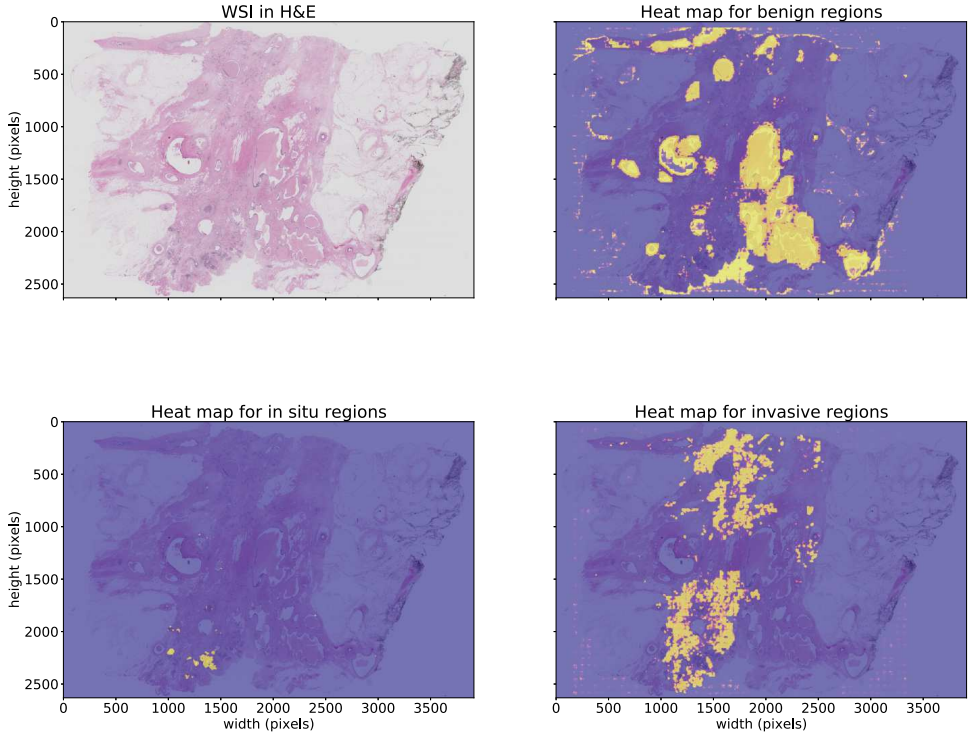
Figure 64 – BACH case A03. (A) WSI input for inference. (B) Pathologist annotation. (C) WSI with pathologist annotation highlighted. (D) Annotation produced by the Cell Segmentation Module. (E) WSI with inferred annotation highlighted. In the annotations, the black color indicates healthy tissue or background, blue benign tissue, green in-situ carcinoma, red invasive carcinoma.



Source: Elaborated by the author.

uncertainty. For example, situations can occur where a pixel group has very close probabilities for two or more classes. With separate heat maps for each class, a pathologist can analyze the region and confirm the diagnosis.

Figure 65 – Heat map for case A01 from BACH.



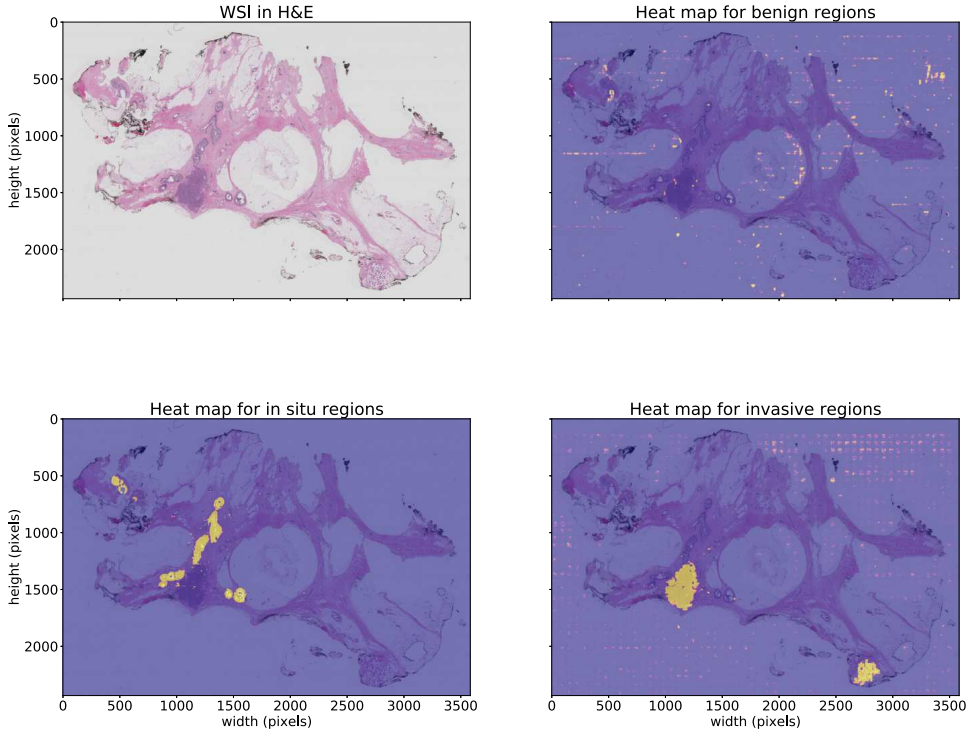
Source: Elaborated by the author.

An important finding in our research occurred during an analysis of the multi-class segmentation produced by the Cell Segmentation Module by a senior pathologist. This finding refers specifically to a region present in case A03 of the BACH dataset. At the time, it was pointed out the existence of an area with neoplastic characteristics at WSI. In the dataset, this region was labeled as healthy tissue. The same was repeated in the analysis provided by the Cell Segmentation Module. However, from the analysis of the heat maps locally, it was possible to observe a higher probability of the pixels in that region characterizing an invasive carcinoma. The area and the respective heat maps are shown in Figure 67.

The behavior shown in Figure 67 may indicate an essential contribution to this work. Even with the lack of annotation by the pathologist, the convolutional architecture was able to generate an activation for the suspect region. In addition to indicating the generalization capacity of CNN, it can serve as a tool to detect difficult regions or as support in border cases.

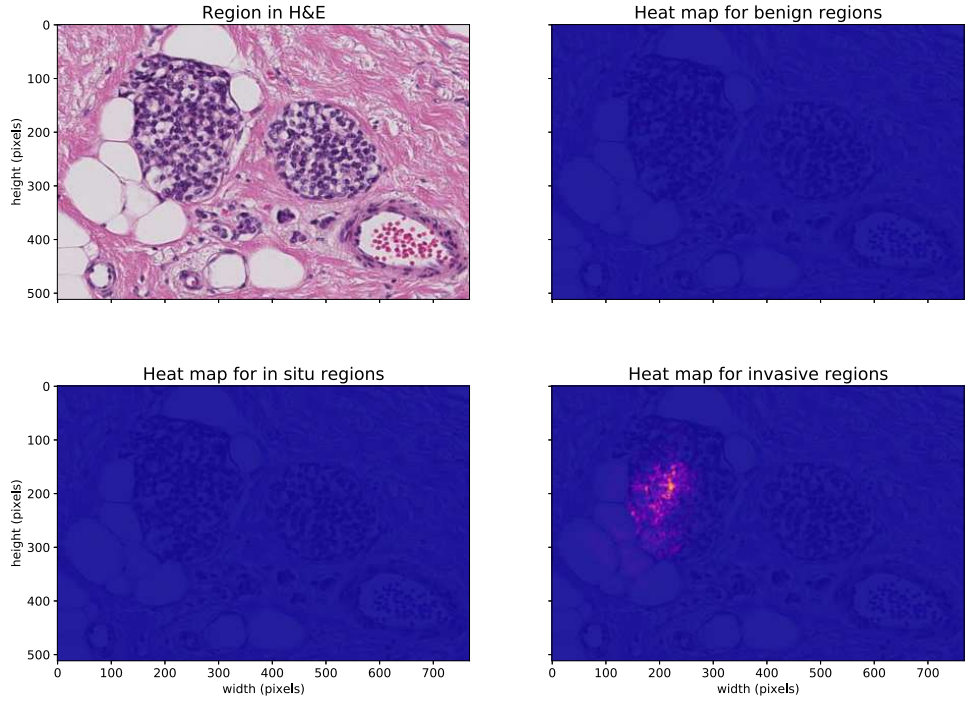
Here it is also important to highlight that the region of the heatmap that presented an activation for an invasive carcinoma corresponds to $\approx 100 \times 100$ pixels. That is a small region compared to an entire WSI. This same region in Figure 67 is presented here in a magnification of $20\times$. Meanwhile, Figures 64 and 66 are presented in $1\times$ magnification. Thus, the same region that in Figure 67 corresponds to $\approx 100 \times 100$ pixels, in Figures 64 and 66 corresponds

Figure 66 – Heat map for case A03 from BACH.



Source: Elaborated by the author.

Figure 67 – Heat map for a suspected region in the BACH case A03.



Source: Elaborated by the author.

to a region with $\approx 5 \times 5$ pixels. Finally, this region demonstrates, in a practical way, the challenge that is the histopathological analysis for a CNN.

6.4 Discussion

In this section, we discuss and compare the DeepBatch with representative works from the literature. A detailed discussion of the state of the art in the analysis of histopathological images of breast cancer is presented in Chapter 3.

The DeepBatch aims to serve as a tool to assist the pathologist during the analysis and diagnosis of breast cancer in WSI. The use of CNNs to support the pathologist's routine has been widely explored in the literature. Some works in the literature are based on data sets based on sections of the WSI (SPANHOL et al., 2016a; HAN et al., 2017; GANDOMKAR; BRENNAN; MELLO-THOMS, 2018b; MURTAZA et al., 2019; ROY et al., 2019). However, these approaches suffer from the loss of contextual information and make it challenging to apply in real environments. Still, we can find works that are dependent on the knowledge of the problem on the part of the researcher, as they need to extract resources manually (ZHENG et al., 2018; BEJNORDI et al., 2017b).

Strategies based on full analysis of WSI focus on binary classification or segmentation. As direct processing of WSI is not yet possible in CNNs, sampling methods are used to identify representative regions. The tiles can be chosen employing adaptive statistical sampling, such as by the quasi-Monte Carlo method and classified into invasive carcinoma or non-invasive areas (CRUZ-ROA et al., 2018). CNN-based approaches to tile sampling are proposed in the literature. Gecer et al. (2018) proposes a cascade of CNNs at different levels of magnification to identify regions of interest. However, the areas are based on view-ports obtained during the analysis of WSI by a group of pathologists. Thus, there is no guarantee that the regions observed during the study are consistent with suspicious areas and reflect a classification accuracy of 55%. In Guo et al. (2019), a classification of tiles is presented in an enlargement of $40\times$ using an InceptionV3 and subsequent segmentation using an FCN in healthy tissue and regions with the presence of metastases. This approach provides segmentation for WSI with a $1/8$ downsampling rate. The inference time for a WSI is approximately 11.5 minutes.

In contrast, our research aims to allow an inference of WSI based on characteristics automatically extracted by a cascade of CNNs, without the need for manual intervention. As for highlights, compared to the methodologies based on WSI sections (SPANHOL et al., 2016a; HAN et al., 2017; GANDOMKAR; BRENNAN; MELLO-THOMS, 2018b; MURTAZA et al., 2019; ROY et al., 2019), is the DeepBatch ability to analyze, during the inference process, contextual information from the entire WSI. The detection of suspicious regions is based on labels provided by pathologists and ensures higher sensitivity in the detection of areas for local analysis. The use of a CNN for global analysis of WSI significantly reduces the computational cost. And, it allows providing segmentation at the pixel level in four classes: background or healthy tissue, benign tissue, in-situ carcinoma, and invasive carcinoma. The inference time for each WSI is ≈ 2 minutes.

A direct quantitative comparison with related works is difficult. Due to the use of private

datasets, methodologies with different image cuts, selecting specific aspects of the datasets, removing some images, and with different classification or segmentation objectives. In this way, we organized in Table 18 a quantitative comparison with similar works.

Table 18 – Comparison of the DeepBatch model with the related works. *Acc* is the Accuracy, *Sen* is the Sensitivity, *Spe* is Specificity, *F1* is the F1-Score, and *AUC* is the area under the ROC curve.

Study	Goal	Dataset	IoU	Acc	Sen	Spe	F1	AUC
Guo et al. (2019)	Pixel-level segmentation in normal and metastases in WSI with a 1:8 downsampling	Camelyon16	80.69	-	-	-	-	**0.96**
Roy et al. (2019)	Classification of histopathology images into normal, benign, in situ and invasive carcinoma	BACH (A)	-	87.00	**90.00**	91.48	**89.85**	-
Gecer et al. (2018)	Classification of WSI into non-proliferative or proliferative changes, atypical ductal hyperplasia, ductal carcinoma in situ, and invasive carcinoma	Private	-	55.00	-	-	-	-
Cruz-Roa et al. (2018)	Pixel-level segmentation in normal and invasive carcinoma in WSI with a 1:32 downsampling	HASHI	76.00	-	87.00	92.00	-	0.90
Han et al. (2017)	Classification of histopathology images into adenosis, fibroadenoma, phyllodes tumor, tubular adenoma, ductal carcinoma, lobular carcinoma, mucinous carcinoma, and papillary carcinoma	BreakHis	-	94.70	-	-	-	-
Bejnordi et al. (2017b)	Classification of WSIs into normal/benign, ductal carcinoma in-situ, and invasive ductal carcinoma	Private	-	81.30	-	-	-	-
Spanhol et al. (2016a)	Classification of histopathology images into benign, and malignant	BreakHis	-	90.00	-	-	-	-
DeepBatch model	Pixel-level segmentation of WSI in four classes: background/normal, benign, carcinoma in situ, and invasive carcinoma.	BACH (B), HASHI, and TCGA	**88.23**	**96.10**	71.83	**96.19**	82.94	0.86

Source: Elaborated by the author.

In Table 18, we can observe the high heterogeneity of the datasets used. This behavior is reflected in the objectives of each study. For example, in the work of (GUO et al., 2019), an analysis of WSI obtained from the sentinel lymph node is made. These WSI have different characteristics from those of breast cancer. For example, the regions affected by cancer present themselves in a well-defined way and with less variability. This characteristic can be decisive to obtain better performance metrics when compared to our work, as in AUC.

The work of Roy et al. (2019), for example, aims to identify the same classes of our research using the BACH dataset. However, there is a point that deserves attention when comparing our

model to the method proposed by Roy et al. (2019). We need to be careful when comparing this method with the DeepBatch due to the used subset of the BACH. The authors use the set (A), which is composed of patches extracted from WSI with labels per image. Thus, it is not possible to perform segmentation at the pixel level. Therefore, better behavior is expected for the classification task.

When analyzing an image at the pixel level, it is possible to have inferior results to image classification tasks. This problem is linked to the characteristics of CNNs, and the way that annotation occurs. CNNs learn characteristics such as textures, colors, and formats by analyzing not only the current pixel to define which class this pixel belongs. However, when the annotation is not provided properly, respecting the limits of the regions or objects, for example, labeling a background region as invasive carcinoma, CNN can classify the pixel as background. In fact, CNN proved to be efficient in learning and generalization, but due to the characteristics of data annotation, performance metrics can point to inferior performances. Therefore, the statistically lower sensitivity achieved by the DeepBatch when compared to the golden standard may be related to the annotations of the BACH dataset. That as pointed out in Figures 51 and 67 have some inconsistencies.

Another point that deserves attention is the finding presented in Figure 67, which represents a significant contribution from our research and reinforces the need for pixel-level studies for breast cancer WSI. For example, comparing our model with methods in the literature that aims at classifying WSI or sections of WSI, it would not be possible to inform the pathologist about the possibility of atypical behavior in that region. Finally, it is essential to note that a model or methodology cannot be evaluated only by a performance metric. Thus, we can see in Table 18 that our model has indexes comparable to the state of the art and even statistically higher, as for the IoU. Therefore, based on quantitative and qualitative assessments, and in comparison with the state of the art, we can consider that the DeepBatch is capable of providing a refined segmentation comparable to the golden standard.

7 CONCLUSION

This study began by investigating the different aspects of intelligent analysis of histopathological images of breast cancer. To support our analysis, we define research questions by delimiting our analysis based on them. To determine the most advanced studies, we rank them by the number of citations and average citations per year. We define taxonomy as a way to categorize studies based on the techniques used and their objectives. Having established the taxonomy, we were able to understand relationships that are important to understanding intelligent analysis of breast cancer in histopathological images. We identify aspects of the challenges and concerns surrounding this area. Another contribution of the study is the identification of public data sets of histopathological breast cancer images used by the corpus. Regarding intelligent methods: We list the main method evaluation metrics and their definitions; and the preprocessing, training, and machine learning techniques applied. Finally, we identify the problems and point out directions for the development of future work.

The study of state of the art enabled the identification of challenges and open opportunities that guided the development of the research. In this Master Thesis, we present a hybrid model for the diagnosis of breast cancer in WSI. The DeepBatch provides the pathologist with an interpretable diagnostic suggestion with a refined segmentation of pixel-level regions in WSI at magnifications of 40 $\times$. Also, we provide a set of heat maps that can assist in diagnosis in border cases. A cascade of CNNs analyzes WSI to provide segmentation and heat maps. In the DeepBatch, a global WSI inference is used to identify ROI to be investigated locally. The local analysis of the WSI is performed in an magnification of 40 $\times$ with tiles of 2048×2048.

The analysis of histopathological images is commonly performed in WSI sections or providing classifications at high magnifications. The main objective of this Master Thesis is to show the feasibility of using DL to provide an interpretable and comparable diagnosis to the golden standard. In this regard, this Master Thesis contributes to the advancement of literature in two main ways.

The first contribution of this work is the proposal of a set of CNNs for segmentation at the pixel level of the WSI. The convolutional architectures have been trained with various datasets. This approach guarantees a better generalization in the identification of suspicious regions. The results obtained for the evaluation metrics in the test sets confirm the effectiveness of the CNNs, both for detecting suspicious areas and for refined multiclass segmentation. Also, the assessment metrics are comparable to state of the art identified in the SLR.

As a second contribution, we propose an interpretable WSI analysis model focused on the pathologist's workflow. The methodology used by DeepBatch for inferences allows us to provide precise segmentations. In this way, our method can improve the accuracy and consistency in the WSI evaluation of breast cancer. Consequently, impacting an improvement in prognosis for patients. Thus, with the approach proposed by the DeepBatch model, we believe that the pathologist can adopt a CAD system with higher acceptability in his routine.

Finally, we believe that our method can generalize and identify complex tissue patterns for other types of cancer. Thus, although this Master Thesis has considered the WSI analysis of breast cancer using DL, the methodology can be extended to different types of cancer.

7.1 Publications

As partial contributions throughout the research, several articles were produced for publication in journals and events. The following are published articles and articles that have been submitted for evaluation:

- Published articles:
  - Zeiser, F. A., da Costa, C. A., Zonta, T., Marques, N. M., Roehe, A. V., Moreno, M., da Rosa Righi, R. (2020). **Segmentation of Masses on Mammograms Using Data Augmentation and Deep Learning.** Journal of Digital Imaging, 1-11. doi:10.1007/s10278-020-00330-4
  - Zeiser, F. A., da Costa, C. A., Ramos, G. O., Roehe, A. V., Bohn, H. C., Santos, I., da Rosa Righi, R., (2020). **Análise Comparativa da Influência dos Espaços de Cores na Segmentação Multi-Classe de *Whole Slide Imaging* do Câncer de Mama Utilizando *Deep Learning***. XX Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS), 1-12.
- Published abstract:
  - Zeiser, F. A., Zonta, T., Costa, C. A., Marques, N. M. C., Roehe, A. V., Moreno, M. (2019). **Auxílio no Rastreamento de Massas em Mamografias Digitalizadas Utilizando a Rede Neural Convolucional U-Net.** Seminário De Iniciação Científica E Seminário Integrado De Ensino, Pesquisa E Extensão - UNOESC.
- Submitted article:
  - Zeiser, F. A., da Costa, C. A., Roehe, A. V., da Rosa Righi, R., Marques, N. M., (2020). **Breast Cancer Intelligent Analysis of Histopathological Data: A Systematic Review**. Computer Methods and Programs in Biomedicine, 1-41.

7.2 Future Work

The DeepBatch presents several opportunities for extension and future work. To improve the generalization of convolutional architectures, we can carry out training on ResNet-50 as a patch-classifier for WSI. With this pre-trained network, weights can be used in the coding part of the U-Net network. The use of pre-trained coders for similar problems can help to increase some performance metrics, such as for sensitivity in the Cell Segmentation module.

Regarding color spaces, we intend to study their use in a single convolutional architecture. This use can increase the samples submitted to CNN and help generalize the model. With this approach, we also hope to combine the benefits presented by each color space for the multi-class segmentation of WSI.

Multimodal architectures present an interesting study potential. Thus, we can concentrate research efforts on studying the impact of using the patient's clinical history, or genetic information, such as those present in the TCGA dataset. Similarly, the combination of histopathological analysis with immunohistochemical analysis can contribute significantly to accelerating the review of cases.

In order to provide a more interpretable diagnosis, we can combine semantic segmentation architectures with image descriptors using Recurrent Neural Networks. These descriptors can describe changes that are present and are considered by the pathologist in the analysis of WSI, such as mitosis or pleomorphism presence. We see potential in the integration of our methodology with datasets that have the pathologist's report for this task. Furthermore, the BreCaHAD dataset can also be used, as it has pixel-level labels for mitosis, apoptosis, tumor nuclei, and among others.

It is important to note that we believe that our model is capable of presenting similar results in the detection of other types of cancer in WSI. Also, we intend to study methods applied to images collected by satellites, which have similarities in size and present the need to identify small objects in the images. Still, we believe it is necessary to study how this model can adapt to the pathologist's routine. And so, look for ways and methods that can be adjusted to provide better acceptability. Finally, it is necessary to verify the impact of CAD systems, such as DeepBatch, on legal, moral, and ethical issues for the use of these methods in health care.

REFERENCES

AKSAC, Alper et al. Brecahad: a dataset for breast cancer histopathological annotation and diagnosis. **BMC research notes**, BioMed Central, v. 12, n. 1, p. 1–3, 2019.

ALOM, Md Zahangir et al. Breast Cancer Classification from Histopathological Images with Inception Recurrent Residual Convolutional Neural Network. **Journal of Digital Imaging**, feb 2019.

ALPAYDIN, Ethem. **Machine learning**: the new ai. 1. ed. Cambridge: MIT Press, 2016. 225 p. ISBN 9780262529518.

ALQUDAH, Ali Mohammad Amin; ALQUDAH, Ali Mohammad Amin. Sliding Window Based Support Vector Machine System for Classification of Breast Cancer Using Histopathological Microscopic Images. **IETE Journal of Research**, p. 1–9, mar 2019.

American Cancer Society. **What Is Breast Cancer?** 2019. Disponível em: <https://www.cancer.org/cancer/breast-cancer/about/what-is-breast-cancer.html>. Acesso em: 06 out. 2019.

ARAUJO, Teresa et al. Classification of breast cancer histology images using convolutional neural networks. **PLOS ONE**, Public Library of Science, v. 12, n. 6, p. 1–14, 06 2017.

ARESTA, Guilherme et al. Bach: Grand challenge on breast cancer histology images. **Medical Image Analysis**, v. 56, p. 122 – 139, 2019. ISSN 1361-8415.

ARJI, Goli et al. A systematic literature review and classification of knowledge discovery in traditional medicine. **Computer Methods and Programs in Biomedicine**, v. 168, p. 39 – 57, 2019. ISSN 0169-2607. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0169260718312859>.

BADRINARAYANAN, V.; KENDALL, A.; CIPOLLA, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 39, n. 12, p. 2481–2495, Dec 2017.

BAKER, Qanita Bani et al. Automated Detection of Benign and Malignant in Breast Histopathology Images. In: **15th International Conference on Computer Systems and Applications**. IEEE, 2018. p. 1–5. Disponível em: <https://ieeexplore.ieee.org/document/8612799/>.

BALAZSI, Matthew et al. Invasive ductal breast carcinoma detector that is robust to image magnification in whole digital slides. **Journal of Medical Imaging**, v. 3, n. 2, may 2016.

BARDOU, Dalal; ZHANG, Kun; AHMAD, Sayed Mohammad. Classification of Breast Cancer Based on Histology Images Using Convolutional Neural Networks. **IEEE Access**, v. 6, n. 6, p. 24680–24693, jun 2018.

BAYRAMOGLU, Neslihan; KANNALA, Juho; HEIKKILA, Janne. Deep learning for magnification independent breast cancer histopathology image classification. In: **23rd International Conference on Pattern Recognition**. [S.l.]: IEEE, 2016. p. 2440–2445.

BEJNORDI, Babak Ehteshami et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. **JAMA**, v. 318, n. 22, p. 2199–2210, 12 2017. ISSN 0098-7484.

BEJNORDI, Babak Ehteshami et al. Context-aware stacked convolutional neural networks for classification of breast carcinomas in whole-slide histopathology images. **Journal of Medical Imaging**, v. 4, n. 04, p. 1, dec 2017.

BEKKERMAN, Ron; BILENKO, Mikhail; LANGFORD, John. **Scaling up Machine Learning: Parallel and Distributed Approaches**. USA: Cambridge University Press, 2011. ISBN 0521192242.

BENHAMMOU, Yassir et al. Breakhis based breast cancer automatic diagnosis using deep learning: Taxonomy, survey and insights. **Neurocomputing**, v. 375, p. 9 – 24, 2020. ISSN 0925-2312. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0925231219313128>.

BRASIL, Ministério da Saúde; Instituto Nacional do Câncer José Alencar Gomes da Silva. **Câncer de mama**. Rio de Janeiro: Ministério da Saúde, 2019. Disponível em: <https://www.inca.gov.br/tipos-de-cancer/cancer-de-mama>. Acesso em: 25 set. 2019.

BRASILEIRO FILHO, Geraldo. **Bogliolo - Patologia**. 9. ed. Rio de Janeiro: GUANABARA, 2016. 2038 p.

BRAY, Freddie et al. Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. **CA: a cancer journal for clinicians**, Wiley Online Library, v. 68, n. 6, p. 394–424, 2018.

BUDUMA, Nikhil; LOCASCIO, Nicholas. **Fundamentals of deep learning: Designing next-generation machine intelligence algorithms**. [S.l.]: O'Reilly Media, Inc., 2017.

BURGER, W.; BURGE, M.J. **Principles of Digital Image Processing: Advanced Methods**. 1. ed. London: Springer, 2009. 273 p.

BURGER, Wilhelm; BURGE, Mark J. **Principles of Digital Image Processing: Advanced Methods**. [S.l.]: Springer Publishing Company, 2013. ISBN 1848829183.

BURKOV, Andriy. **The hundred-page machine learning book**. [S.l.]: Andriy Burkov Quebec City, Can., 2019.

CARTER, Jane V. et al. Roc-ing along: Evaluation and interpretation of receiver operating characteristic curves. **Surgery**, v. 159, n. 6, p. 1638 – 1645, 2016.

CHAN, Alan; TUSZYNSKI, Jack A. Automatic prediction of tumour malignancy in breast cancer with fractal dimension. **Royal Society Open Science**, v. 3, n. 12, dec 2016.

CHAN, John K. C. The wonderful colors of the hematoxylin–eosin stain in diagnostic surgical pathology. **International Journal of Surgical Pathology**, v. 22, n. 1, p. 12–32, 2014.

CHOLLET, Francois. **Deep Learning mit Python und Keras: Das Praxis-Handbuch vom Entwickler der Keras-Bibliothek**. [S.l.]: MITP-Verlags GmbH & Co. KG, 2018.

CIANFROCCA, Mary; GRADISHAR, William. New molecular classifications of breast cancer. **CA: A Cancer Journal for Clinicians**, v. 59, n. 5, p. 303–313, 2009.

CIRESAN, Dan Claudiu et al. Flexible, high performance convolutional neural networks for image classification. In: **Twenty-Second International Joint Conference on Artificial Intelligence**. [S.l.: s.n.], 2011.

COPPIN, Ben. **Inteligência artificial**. 1. ed. Rio de Janeiro: LTC, 2010. 636 p.

COVER, G.S. et al. Computational methods for corpus callosum segmentation on mri: A systematic literature review. **Computer Methods and Programs in Biomedicine**, v. 154, p. 25 – 35, 2018. ISSN 0169-2607.

CRUZ-ROA, Angel et al. Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks. **Medical Imaging 2014: Digital Pathology**, v. 9041, p. 904103, 2014.

____. High-throughput adaptive sampling for whole-slide histopathology image analysis (HASHI) via convolutional neural networks: Application to invasive breast cancer detection. **PLoS ONE**, v. 13, n. 5, p. e0196828, may 2018.

DABBS, David J. **BREAST PATHOLOGY**. 2. ed. Philadelphia: Elsevier, 2017.

DAS, Dev Kumar; DUTTA, Pranab Kumar. Efficient automated detection of mitotic cells from breast histological images using deep convolution neutral network with wavelet decomposed patches. **Computers in Biology and Medicine**, v. 104, n. May 2018, p. 29–42, jan 2019.

DEBELEE, Taye Girma et al. Survey of deep learning in breast cancer image analysis. **Evolving Systems**, Springer, v. 11, n. 1, p. 143–163, 2020.

DEVITA, Vincent T.; LAWRENCE, Theodore S.; ROSENBERG, Steven A. **DeVita, Hellman, and Rosenberg's cancer: Principles & practice of oncology**. 11. ed. [S.l.]: Wolters Kluwer Health Adis (ESP), 2018. 2432 p.

DIMITRIOU, Neofytos; ARANDJELOVIć, Ognjen; CAIE, Peter D. Deep learning for whole slide image analysis: An overview. **Frontiers in Medicine**, v. 6, p. 264, 2019.

DUNDAR, M. Murat et al. Computerized classification of intraductal breast lesions using histopathological images. **IEEE Transactions on Biomedical Engineering**, v. 58, n. 7, p. 1977–1984, jul 2011.

DUSENBERRY, Mike; HU, Fei. **Deep Learning for Breast Cancer Mitosis Detection**. 2019. Disponível em: <https://github.com/CODAIT/deep-histopath/raw/master/docs/tupac16-paper/paper.pdf>.

ELIYATKIN, Nuket et al. Molecular classification of breast carcinoma: From traditional, old-fashioned way to a new age, and a new way. **The journal of breast health**, v. 11, n. 2, p. 59–66, 2015.

FAWCETT, Tom. An introduction to roc analysis. **Pattern Recognition Letters**, v. 27, n. 8, p. 861 – 874, 2006. ISSN 0167-8655.

FENG, Yangqin; ZHANG, Lei; YI, Zhang. Breast cancer cell nuclei classification in histopathology images using deep neural networks. **International Journal of Computer Assisted Radiology and Surgery**, v. 13, n. 2, p. 179–191, feb 2018.

FONDÓN, Irene et al. Automatic classification of tissue malignancy for breast carcinoma diagnosis. **Computers in Biology and Medicine**, v. 96, p. 41–51, may 2018.

GALEA, Marcus H. et al. The nottingham prognostic index in primary breast cancer. **Breast Cancer Research and Treatment**, v. 22, n. 3, p. 207–219, Oct 1992.

GANDOMKAR, Ziba; BRENNAN, Patrick C.; MELLO-THOMS, Claudia. A framework for distinguishing benign from malignant breast histopathological images using deep residual networks. **14th International Workshop on Breast Imaging**, p. 54, 2018.

____. MuDeRN: Multi-category classification of breast histopathological image using deep residual networks. **Artificial Intelligence in Medicine**, v. 88, p. 14–24, jun 2018.

GARCIA-GARCIA, Alberto et al. **A Review on Deep Learning Techniques Applied to Semantic Segmentation**. 2017.

GECER, Baris et al. Detection and classification of cancer in whole slide breast histopathology images using deep convolutional networks. **Pattern Recognition**, v. 84, p. 345–356, dec 2018.

GELASCA, Elisa Drelie et al. Evaluation and benchmark for biological image segmentation. In: IEEE. **2008 15th IEEE International Conference on Image Processing**. [S.l.], 2008. p. 1816–1819.

GÉRON, Aurélien. **Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts**. [S.l.: s.n.], 2017.

GHAZVINIAN ZANJANI, Farhad et al. Impact of JPEG 2000 compression on deep convolutional neural networks for metastatic cancer detection in histopathological images. **Journal of Medical Imaging**, v. 6, n. 02, p. 1, 2019.

GHOSH, Swarnendu et al. Understanding deep learning techniques for image segmentation. **ACM Comput. Surv.**, Association for Computing Machinery, New York, NY, USA, v. 52, n. 4, ago. 2019.

GIULIANO, Armando E. et al. Breast cancer—major changes in the american joint committee on cancer eighth edition cancer staging manual. **CA: A Cancer Journal for Clinicians**, v. 67, n. 4, p. 290–303, 2017.

GOBBI, Helenice. Classificação dos tumores da mama: atualização baseada na nova classificação da Organização Mundial da Saúde de 2012. **Jornal Brasileiro de Patologia e Medicina Laboratorial**, scielo, v. 48, p. 463 – 474, 12 2012. ISSN 1676-2444.

GONZALEZ, Rafael C.; WOODS, Richard E. **Digital Image Processing**. 4. ed. New York: Pearson, 2018. 624 p.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep Learning**. 1. ed. Cambridge: MIT Press, 2017.

GUBERT, Luis Claudio; COSTA, Cristiano André da; RIGHI, Rodrigo da Rosa. Context awareness in healthcare: a systematic literature review. **Universal Access in the Information Society**, 2019.

GUO, Yanming et al. A review of semantic segmentation using deep neural networks. **International Journal of Multimedia Information Retrieval**, v. 7, n. 2, p. 87–93, Jun 2018.

GUO, Zichao et al. A Fast and Refined Cancer Regions Segmentation Framework in Whole-slide Breast Pathological Images. **Scientific Reports**, v. 9, n. 1, p. 882, dec 2019.

GUPTA, Vibha; BHAVSAR, Arnav. Sequential modeling of deep features for breast cancer histopathological image classification. **IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops**, v. 2018-June, p. 2335–2342, 2018.

HAN, Zhongyi et al. Breast Cancer Multi-classification from Histopathological Images with Structured Deep Learning Model. **Scientific Reports**, v. 7, n. 1, p. 4172, dec 2017.

HANAHAN, Douglas; WEINBERG, Robert A. Hallmarks of cancer: the next generation. **cell**, Elsevier, v. 144, n. 5, p. 646–674, 2011.

HAYKIN, Simon S. **Neural networks and learning machines**. Third. Upper Saddle River, NJ: Pearson Education, 2009.

HE, Kaiming et al. Deep residual learning for image recognition. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2016. p. 770–778.

HERRINGTON, C Simon. **Muir's Textbook of Pathology**. 15. ed. Boca Raton: Taylor & Francis Group, 2014. 589 p.

HOFF, Paulo Marcelo Gehm et al. **Tratado De Oncologia**. [S.l.]: EDITORA ATHENEU, 2012. ISBN 9788538803126.

HUBEL, David H; WIESEL, Torsten N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. **The Journal of physiology**, Wiley Online Library, v. 160, n. 1, p. 106–154, 1962.

IOFFE, Sergey; SZEGEDY, Christian. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: **Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37**. [S.l.]: JMLR.org, 2015. (ICML'15), p. 448–456.

ISSAC NIWAS, S. et al. An expert support system for breast cancer diagnosis using color wavelet features. **Journal of Medical Systems**, v. 36, n. 5, p. 3091–3102, oct 2012.

JELEN, Lukasz; FEVENS, Thomas; KRZYZAK, Adam. Influence of nuclei segmentation on breast cancer malignancy classification. **Medical Imaging 2009: Computer-Aided Diagnosis**, v. 7260, p. 726014, 2009.

JEONG, Yunhee et al. Dilated saliency u-net for white matter hyperintensities segmentation using irregularity age map. **Frontiers in Aging Neuroscience**, v. 11, p. 150, 2019.

JESINGER, Robert A. Breast anatomy for the interventionalist. **Techniques in Vascular and Interventional Radiology**, v. 17, n. 1, p. 3 – 9, 2014.

JäHNE, Bernd. **Digital image processing**. 6. ed. Berlin: Springer, 2005. 598 p.

KANG, Kai; WANG, Xiaogang. **Fully Convolutional Neural Networks for Crowd Segmentation**. 2014.

KHAN, Adnan Mujahid et al. A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution. **IEEE Transactions on Biomedical Engineering**, v. 61, n. 6, p. 1729–1738, 2014.

KHAN, Adnan Mujahid; SIRINUKUNWATTANA, Korsuk; RAJPOOT, Nasir. A Global Covariance Descriptor for Nuclear Atypia Scoring in Breast Histopathology Images. **IEEE Journal of Biomedical and Health Informatics**, v. 19, n. 5, p. 1637–1647, 2015.

KITCHENHAM, B.; CHARTERS, S. **Guidelines for performing Systematic Literature Reviews in Software Engineering**. 2007.

KORKMAZ, Sevcan Aytac; KORKMAZ, Mehmet Fatih; POYRAZ, Mustafa. Diagnosis of breast cancer in light microscopic and mammographic images textures using relative entropy via kernel estimation. **Medical and Biological Engineering and Computing**, v. 54, n. 4, p. 561–573, apr 2016.

KRAWCZYK, Bartosz et al. Evolutionary undersampling boosting for imbalanced classification of breast cancer malignancy. **Applied Soft Computing**, v. 38, p. 714–726, jan 2016.

KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey E. Imagenet classification with deep convolutional neural networks. In: **Advances in neural information processing systems**. [S.l.: s.n.], 2012. p. 1097–1105.

KUMAR, V.; ABBAS, Abul K.; ASTER, Jon C. **Robbins Patologia Basica**. 9. ed. São Paulo: Elsevier, 2013. 927 p.

____. **Robbins & Cotran Patologia — Bases Patológicas das Doenças**. 9. ed. São Paulo: Elsevier, 2016. 2696 p.

____. **Robbins Patologia Basica**. 10. ed. São Paulo: Elsevier, 2018. 952 p.

LAKHANI, S.R.; CANCER, International Agency for Research on; ORGANIZATION, World Health. **WHO Classification of Tumours of the Breast**. [S.l.]: International Agency for Research on Cancer, 2012. (IARC Who Classification of Tum).

LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. Deep learning. **Nature**, v. 521, n. 7553, p. 436–444, 2015. Disponível em: <https://doi.org/10.1038/nature14539>.

LECUN, Y. et al. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, v. 86, n. 11, p. 2278–2324, Nov 1998. ISSN 1558-2256.

LI, Chao et al. DeepMitosis: Mitosis detection via deep detection, verification and segmentation networks. **Medical Image Analysis**, v. 45, p. 121–133, apr 2018.

LI, Xingyu; PLATANIOTIS, Konstantinos N. Color model comparative analysis for breast cancer diagnosis using H and E stained images. **Medical Imaging 2015: Digital Pathology**, v. 9420, p. 94200L, 2015.

LI, Yuqian; WU, Junmin; WU, Qisong. Classification of Breast Cancer Histology Images Using Multi-Size and Discriminative Patches Based on Deep Learning. **IEEE Access**, v. 7, p. 1–1, 2019.

LITJENS, Geert et al. 1399 H&E-stained sentinel lymph node sections of breast cancer patients: the CAMELYON dataset. **GigaScience**, v. 7, n. 6, 05 2018.

LONG, J.; SHELHAMER, E.; DARRELL, T. Fully convolutional networks for semantic segmentation. In: **2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 2015. p. 3431–3440.

LOUKAS, C. et al. Breast Cancer Characterization Based on Image Classification of Tissue Sections Visualized under Low Magnification. **Computational and Mathematical Methods in Medicine**, v. 2013, p. 1–7, 2013.

MAJEED, Hassaan et al. Label-free quantitative evaluation of breast tissue using spatial light interference microscopy (slim). **Scientific reports**, Nature Publishing Group, v. 8, n. 1, p. 1–9, 2018.

MAKKI, Jaafar. Diversity of breast carcinoma: Histological subtypes and clinical relevance. **Clinical Medicine Insights: Pathology**, v. 8, p. CPath.S31563, 2015.

MAROOF, Noorulain et al. Mitosis detection in breast cancer histopathology images using hybrid feature space. **Photodiagnosis and Photodynamic Therapy**, v. 31, p. 101885, 2020. ISSN 1572-1000. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1572100020302398>.

MARQUES FILHO, Ogê; VIEIRA NETO, Hugo. **Processamento Digital de Imagens**. 1. ed. Rio de Janeiro: Brasport, 1999. 331 p.

MCCULLOCH, Warren S.; PITTS, Walter. A logical calculus of the ideas immanent in nervous activity. **The Bulletin of Mathematical Biophysics**, v. 5, n. 4, p. 115–133, 1943.

MESCHER, Anthony L. **Junqueira's Basic Histology: Text and Atlas, Thirteenth Edition: Text and Atlas, Thirteenth Edition**. 15. ed. New York: McGraw-Hill Education, 2018. 523 p.

MINSKY, Marvin; PAPERT, Seymour. Perceptrons - an introduction to computational geometry. In: . [S.l.: s.n.], 1969.

MITTAL, Himanshu; SARASWAT, Mukesh. An automatic nuclei segmentation method using intelligent gravitational search algorithm based superpixel clustering. **Swarm and Evolutionary Computation**, v. 45, p. 15–32, 2019.

MOINFAR, F. **Essentials of Diagnostic Breast Pathology: A Practical Approach**. 1. ed. Berlin: Springer Berlin Heidelberg, 2007. 483 p.

MONTENEGRO, Joao Luis Zeni; COSTA, Cristiano André da; RIGHI, Rodrigo da Rosa. Survey of conversational agents in health. **Expert Systems with Applications**, v. 129, p. 56 – 67, 2019.

MOORE, Keith L; DALLEY, Arthur F; AGUR, Anne MR. **Clinically oriented anatomy**. [S.l.]: Lippincott Williams & Wilkins, 2013.

MÜLLER, Andreas C; GUIDO, Sarah et al. **Introduction to machine learning with Python: a guide for data scientists**. [S.l.]: O'Reilly Media, Inc., 2016.

MURTAZA, Ghulam et al. Breast Cancer Multi-classification through Deep Neural Network and Hierarchical Classification Approach. **Multimedia Tools and Applications**, apr 2019.

National Cancer Institute. **Outcomes & Impact of The Cancer Genome Atlas**. 2019. Disponível em: <https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga/history>.

____. **The Cancer Genome Atlas Program**. 2020. Disponível em: <https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga>. Acesso em: 17 april 2020.

NAYAK, N. et al. Classification of tumor histopathology via sparse feature learning. In: **2013 IEEE 10th International Symposium on Biomedical Imaging**. [S.l.: s.n.], 2013. p. 410–413.

OLIVEIRA, Luiz Eduardo et al. Multiple instance learning for histopathological breast cancer image classification. **Expert Systems with Applications**, v. 117, p. 103–111, mar 2018.

OTSU, Nobuyuki. A threshold selection method from gray-level histograms. **IEEE Transactions on Systems, Man, and Cybernetics**, v. 9, n. 1, p. 62–66, 1979.

PAN, Xipeng et al. Accurate segmentation of nuclei in pathological images via sparse reconstruction and deep convolutional networks. **Neurocomputing**, v. 229, p. 88–99, mar 2017.

____. Cell detection in pathology and microscopy images with multi-scale fully convolutional neural networks. **World Wide Web**, v. 21, n. 6, p. 1721–1743, nov 2018.

PAWLINA, Wojciech; ROSS, Michael H. **Histology: a text and atlas: with correlated cell and molecular biology**. [S.l.]: Lippincott Williams & Wilkins, 2018.

PEDRINI, Hélio; SCHWARTZ, William Robson. **Análise de imagens digitais: princípios, algoritmos e aplicações**. [S.l.]: Thomson Learning, 2008.

PEDRO, Ricardo Wandré Dias; MACHADO-LIMA, Ariane; NUNES, Fátima L.S. Is mass classification in mammograms a solved problem? - a critical review over the last 20 years. **Expert Systems with Applications**, v. 119, p. 90 – 103, 2019.

POLYAK, Kornelia. Breast cancer: origins and evolution. **The Journal of Clinical Investigation**, The American Society for Clinical Investigation, v. 117, p. 3155–3163, 11 2007.

RAJNOHA, M.; BURGET, R.; POVODA, L. Image background noise impact on convolutional neural network training. In: **2018 10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)**. [S.l.: s.n.], 2018. p. 1–4.

RASBAND, Wayne S et al. **ImageJ**. [S.l.]: Bethesda, MD, 1997.

RAWAT, Waseem; WANG, Zenghui. Deep convolutional neural networks for image classification: A comprehensive review. **Neural Computation**, v. 29, n. 9, p. 2352–2449, 2017. Disponível em: <https://doi.org/10.1162/neco_a_00990>.

ROEHRS, Alex et al. Personal health records: A systematic literature review. **Journal of Medical Internet Research**, v. 19, n. 1, p. e13, jan 2017.

RONNEBERGER, Olaf; FISCHER, Philipp; BROX, Thomas. U-net: Convolutional networks for biomedical image segmentation. In: **Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015**. Cham: Springer International Publishing, 2015. p. 234–241.

ROSENBLATT, Frank. The perceptron: A probabilistic model for information storage and organization in the brain. **Psychological Review**, v. 65, n. 6, p. 386–408, dez. 1958.

ROUX, L et al. **MITOS-ATYPIA-14**. 2014. Disponível em: <https://mitos-atypia-14.grand-challenge.org/Dataset/>.

ROY, Kaushiki et al. Patch-based system for Classification of Breast Histology images using deep learning. **Computerized Medical Imaging and Graphics**, v. 71, p. 90–103, jan 2019.

RUBIN, R. et al. **Rubin's Pathology: Clinicopathologic Foundations of Medicine**. 7. ed. Philadelphia: Wolters Kluwer, 2015. 1618 p.

RUSSELL, S.; NORVIG, P. **Artificial Intelligence: A Modern Approach**. Third. Upper Saddle River, NJ: Prentice Hall, 2010.

RUUSKA, Salla et al. Evaluation of the confusion matrix method in the validation of an automated system for measuring feeding behaviour of cattle. **Behavioural Processes**, v. 148, p. 56 – 62, 2018. ISSN 0376-6357.

Sabeena Beevi, K.; NAIR, Madhu S.; BINDU, G. R. Automatic mitosis detection in breast histopathology images using Convolutional Neural Network based deep transfer learning. **Biocybernetics and Biomedical Engineering**, v. 39, n. 1, p. 214–223, 2019.

SAHA, M.; CHAKRABORTY, C. Her2net: A deep framework for semantic segmentation and classification of cell membranes and nuclei in breast cancer evaluation. **IEEE Transactions on Image Processing**, v. 27, n. 5, p. 2189–2200, 2018.

SAHA, Monjoy; CHAKRABORTY, Chandan; RACOCEANU, Daniel. Efficient deep learning model for mitosis detection using breast histopathology images. **Computerized Medical Imaging and Graphics**, v. 64, p. 29–40, mar 2018.

SHEN, Dinggang; WU, Guorong; SUK, Heung-Il. Deep learning in medical image analysis. **Annual Review of Biomedical Engineering**, v. 19, n. 1, p. 221–248, 2017.

SHORTEN, Connor; KHOSHGOFTAAR, Taghi M. A survey on image data augmentation for deep learning. **Journal of Big Data**, v. 6, n. 1, p. 60, Jul 2019. Disponível em: <https://doi.org/10.1186/s40537-019-0197-0>.

SIMONYAN, Karen; ZISSERMAN, Andrew. Very deep convolutional networks for large-scale image recognition. **arXiv preprint arXiv:1409.1556**, 2014.

SONKA, Milan; HLAVAC, Vaclav; BOYLE, Roger. **Image Processing, Analysis, and Machine Vision**. 4. ed. Stanford: Cengage Learning, 2015. 930 p. ISBN 1-133-59360-7.

SPANHOL, Fabio A. et al. Deep features for breast cancer histopathological image classification. **2017 IEEE International Conference on Systems, Man, and Cybernetics**, v. 2017-Janua, p. 1868–1873, 2017.

SPANHOL, Fabio Alexandre et al. Breast cancer histopathological image classification using Convolutional Neural Networks. In: **Proceedings of the International Joint Conference on Neural Networks**. [S.l.]: Institute of Electrical and Electronics Engineers Inc., 2016. v. 2016-Octob, p. 2560–2567.

SPANHOL, F. A. et al. A dataset for breast cancer histopathological image classification. **IEEE Transactions on Biomedical Engineering**, v. 63, n. 7, p. 1455–1462, July 2016.

SRIVASTAVA, Nitish et al. Dropout: a simple way to prevent neural networks from overfitting. **The journal of machine learning research**, JMLR. org, v. 15, n. 1, p. 1929–1958, 2014.

STANDRING, Susan. **Gray's anatomy e-book: the anatomical basis of clinical practice**. [S.l.]: Elsevier Health Sciences, 2015.

SUTTON, Richard S; BARTO, Andrew G. **Reinforcement learning: An introduction**. [S.l.]: MIT press, 2018.

SZEGEDY, Christian et al. Going deeper with convolutions. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2015. p. 1–9.

TASHK, Ashkan et al. Automatic detection of breast cancer mitotic cells based on the combination of textural, statistical and innovative mathematical features. **Applied Mathematical Modelling**, v. 39, n. 20, p. 6165–6182, oct 2015.

TURING, Alan M. Computing machinery and intelligence. **Mind**, v. 59, n. 236, p. 433, 1950.

UCSB Center for Bio-Segmentation. **Bio-Segmentation**. 2020. Disponível em: <https://bioimage.ucsb.edu/research/bio-segmentation>. Acesso em: 17 april 2020.

VETA, Mitko et al. Predicting breast tumor proliferation from whole-slide images: The tupac16 challenge. **Medical Image Analysis**, Elsevier BV, v. 54, p. 111–121, May 2019.

VETA, M. et al. Breast cancer histopathology image analysis: A review. **IEEE Transactions on Biomedical Engineering**, v. 61, n. 5, p. 1400–1411, May 2014.

VO, Duc My; NGUYEN, Ngoc Quang; LEE, Sang Woong. Classification of breast cancer histology images using incremental boosting convolution networks. **Information Sciences**, v. 482, p. 123–138, may 2019.

WAHAB, Noorul; KHAN, Asifullah; LEE, Yeon Soo. Two-phase deep convolutional neural network for reducing class skewness in histopathological images based breast cancer detection. **Computers in Biology and Medicine**, v. 85, p. 86–97, jun 2017.

WAN, Tao et al. Automated grading of breast cancer histopathology using cascaded ensemble with combination of multi-level image features. **Neurocomputing**, v. 229, p. 34–44, mar 2017.

____. Automated mitosis detection in histopathology based on non-gaussian modeling of complex wavelet coefficients. **Neurocomputing**, v. 237, p. 291–303, may 2017.

WANG, Haibo et al. Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features. **Journal of Medical Imaging**, v. 1, n. 3, p. 034003, oct 2014.

WEI, Jerry et al. **Generative Image Translation for Data Augmentation in Colorectal Histopathology Images**. 2019.

World Cancer Research Fund. **Breast cancer statistics**. 2019. Disponível em: <https://www.wcrf.org/dietandcancer/cancer-trends/breast-cancer-statistics>. Acesso em: 25 set. 2019.

World Health Organization. **Cancer**. 2019a. Disponível em: <https://www.who.int/health-topics/cancer#tab=tab_1>. Acesso em: 25 set. 2019.

____. **Cancer - Diagnosis and Treatment**. 2019b. Disponível em: <https://www.who.int/cancer/treatment/en/>. Acesso em: 25 set. 2019.

XIE, Juanying et al. Deep Learning Based Analysis of Histopathological Images of Breast Cancer. **Frontiers in Genetics**, v. 10, p. 80, feb 2019.

XU, Jun et al. Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images. **IEEE Transactions on Medical Imaging**, v. 35, n. 1, p. 119–130, jan 2016.

YAN, Rui et al. A Hybrid Convolutional and Recurrent Deep Neural Network for Breast Cancer Pathological Image Classification. **IEEE International Conference on Bioinformatics and Biomedicine**, p. 957–962, 2019.

YU, Fisher; KOLTUN, Vladlen; FUNKHOUSER, Thomas. Dilated residual networks. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2017. p. 472–480.

YU, Hongshan et al. Methods and datasets on semantic segmentation: A review. **Neurocomputing**, v. 304, p. 82 – 103, 2018. ISSN 0925-2312.

ZEILER, Matthew D.; FERGUS, Rob. Visualizing and understanding convolutional networks. In: FLEET, David et al. (Ed.). **Computer Vision – ECCV 2014**. Cham: Springer International Publishing, 2014. p. 818–833.

ZHANG, Xiaofan et al. Fusing Heterogeneous Features From Stacked Sparse Autoencoder for Histopathological Image Analysis. **IEEE Journal of Biomedical and Health Informatics**, v. 20, n. 5, p. 1377–1383, 2016.

____. Towards large-scale histopathological image analysis: Hashing-based image retrieval. **IEEE Transactions on Medical Imaging**, v. 34, n. 2, p. 496–506, 2015.

ZHENG, Yushan et al. Feature extraction from histopathological images based on nucleus-guided convolutional neural network for breast lesion classification. **Pattern Recognition**, v. 71, p. 14–25, nov 2017.

____. Histopathological Whole Slide Image Analysis Using Context-Based CBIR. **IEEE Transactions on Medical Imaging**, v. 37, n. 7, p. 1641–1652, 2018.