CNN classification of immune cells in tumor sections

André da Veiga Ferro Forjaz Pereira

Thesis to obtain the Master of Science Degree in

Electrical and Computer Engineering

Supervisors:  Prof. Denis Wirtz
             Prof. Jorge Salvador Marques

Examination Committee
Chairperson: Prof. Dr. João Fernando Cardoso Silva Sequeira
Supervisor: Prof. Denis Wirtz
Members of the Committee: Prof. João Miguel Raposo Sanches

Outubro 2020
Declaration

I declare that this document is an original work of my own authorship and that it fulfills all the requirements of the Code of Conduct and Good Practices of the Universidade de Lisboa.
Acknowledgments

I would like to pay my special regards to my advisor, Prof. Denis Wirtz, for granting me the opportunity to develop this project and introducing me to several new fields of study. Also, I would like to recognize the essential guidance that he provided me through the study.

Second, I also want to thank Prof. Pei-Hsun Wu for his tirelessness in presenting me to new concepts, for his full time availability and for his persistent assistance.

In addition, I would like to acknowledge the important support given by Wirtz Lab and the Institute of NanoBioTechnology, at the Johns Hopkins University, for their unlimited support and funding.

I would also like to express my sincere gratitude to all members of Wirtz lab. In particular, to Kyu Han and Ashley Kiemten for the stimulating discussions and many helpful suggestions.

Fourth, I would like to thank my friends and everyone else that contributed to my project for their support.

Finally, I would like to thank my family for supporting me through this project. In special, I would like to thank my uncles and my parents for providing me with the means to make all of this happen.
Abstract

Modern deep learning approaches have the potential to analyze tumors for improved diagnostic and prognostic of patients. However, the accuracy and robustness of current deep learning models rely too heavily on experts to annotate tumor sections to establish a ground truth. This is particularly important for the identification of immune cells inside tumors, which play a central role in the onset and progression of solid tumors such as breast cancer. This thesis presents a solution for the problem of reliably identify T cell infiltration in tumors without an expert’s input, while using tumor sections stained with hematoxylin and eosin (H&E).

For our workflow, tumor sections are first stained with antibodies recognizing CD8+ T cells, FOXP3+ regulatory T cells, pan-cytokeratin+ cancer cells and CD31+ endothelial cells that constitute blood vessels. After imaging and washing of the tissue section, the same section is stained with H&E and imaged once more. Second, immunofluorescence (IF) and H&E images are accurately aligned via Fast-Fourier-transformation (FFT)-based image registration. Third, fast two-dimensional peak-finder method is used to accurately detect and transfer information from the IF to the H&E images. Lastly, a convolutional neural network (CNN) is trained using these annotated H&E images. The resulting trained CNN can classify cancer cells, T cells, and endothelial cells in breast cancer tissues using standard H&E images with a minimum accuracy of 94%, 85%, and 72%, respectively.

A first contribution of this thesis consists in the development of a robust method that can rapidly process and align large image sets of tumor tissue microarrays (TMAs) and accurately generate annotated H&E images using IF images as ground truth. A second contribution consists in using a trained and validated CNN to classify several cell subtypes on standard H&E stained tissue sections, without requiring IF images or expert’s annotations.

Keywords:
Convolutional Neural Networks, Biomedical Image Registration, Cell Detection, Breast Cancer, Immunofluorescent staining.
Resumo

As técnicas de Deep Learning desenvolvidas ultimamente têm vindo a demonstrar um enorme potencial relativamente à análise de tumores, podendo vir a melhorar os diagnósticos e os prognósticos. Todavia, atualmente, a precisão e a robustez destes modelos de Deep Learning dependem demasiado de anotações realizadas por peritos em tecidos contendo cancro, para estabelecer uma ground truth. Isto é particularlymente importante na identificação de linfócitos dentro de tumores, que representa um fator importante na progressão dos cancros. Esta dissertação apresenta uma solução para este problema, possibilitando a indentificação de linfócitos T em tumores, sem necessidade de intervenção de especialistas e utilizando secções coloradas com hematoxylin and eosin (H&E).

O trabalho desenvolvido começa por colorar secções de tumor com anticorpos que reconhecem linfócitos T CD8+, linfócitos T reguladores FOXP3+, células cancerígenas pan-citokeratin+ e células endoteliais CD31+ que constituem vasos sanguíneos. Após a digitalização e lavagem da secção de tecido, a mesma secção é tingida com H&E e digitalizada uma vez mais. Em segundo, as imagens immunofluorescence (IF) e H&E são alinhadas com precisão através de um método de transformação Fast-Fourier (FFT). Terceiro, um método rápido de localização de pico bidimensional é usado para detectar e transferir com precisão as informações das imagens IF para as imagens H&E. Por último, uma rede neural convolucional (CNN) é treinada usando essas imagens H&E anotadas. A CNN treinada resultante pode classificar células cancerígenas, linfócitos T e células endoteliais em tecidos cancerígenas da mama usando imagens H&E com uma precisão mínima de 94%, 85% e 72%, respectivamente.

Uma primeira contribuição desta tese consiste no desenvolvimento de um método robusto que possibilite processar e alinhar rapidamente grandes conjuntos de imagens de microarranjos de tecido tumoral (TMAs) e gerar imagens H&E anotadas com precisão, usando imagens IF como ground truth. Uma segunda contribuição consiste em usar uma CNN treinada e validada para classificar vários subtipos de células em secções de tecido tingidas com H&E, sem a necessidade de imagens IF ou anotações de especialistas.

Palavras-chave: Redes Neuronais Convolucionais, Alinhamento de imagens Biomédicas, Deteção de células, Cancro da mama, Coloração Immunofluorescente.
Contents

Declaration ................................................................. iii
Acknowledgments ......................................................... v
Abstract ................................................................. vii
Resumo .......................................................................... ix
List of Tables .................................................................... xv
List of Figures .................................................................... xvii
List of Abbreviations ..................................................... xix

1 Introduction ............................................................... 1
  1.1 Motivation ............................................................... 1
  1.2 Pathological and Histological Image Analysis ................. 2
    1.2.1 Standard Tissue Analysis ................................. 2
    1.2.2 Staining Techniques ........................................ 3
      1.2.2.1 Hematoxylin and Eosin Staining .................. 3
      1.2.2.2 Immunofluorescent staining ...................... 3
    1.2.3 Treatment Methods .......................................... 4
    1.2.4 Tumor Microenvironment ................................. 4
    1.2.5 Tissue Microarray ........................................... 5
  1.3 Integration of Deep Learning in Biomedicine .................. 6
    1.3.1 Historical Background ....................................... 6
    1.3.2 Convolutional Neural Networks as a solution for Automatic Annotation 6
  1.4 Objectives and Contributions .................................... 6
  1.5 Thesis Outline ....................................................... 7

2 Theoretical Background ............................................... 9
  2.1 Image Registration ................................................. 9
    2.1.1 Homologous Points ......................................... 9
    2.1.2 Transformation Functions ................................. 10
      2.1.2.1 Translation ............................................. 10
      2.1.2.2 Rotation ................................................ 11
      2.1.2.3 Scaling .................................................. 11
2.1.3 Similarity Transformation .................................................. 12
2.1.4 Parameter Estimation ......................................................... 12
2.1.5 Registration Method .......................................................... 13
2.1.5.1 Fast Fourier Transformation-based image registration ............... 13
2.2 Image Analysis ................................................................. 17
2.2.1 K-means Clustering .......................................................... 17
2.2.2 Peak color intensity detection .............................................. 18
2.3 Convolutional Neural Network ................................................ 19
2.3.1 Historical Background ....................................................... 19
2.3.2 Architecture Overview ...................................................... 19
2.3.2.1 Convolutional Layers .................................................... 20
2.3.2.2 Non Linear Layers ........................................................ 20
2.3.2.3 Pooling Layers ............................................................ 21
2.3.2.4 Fully Connected Layers ................................................ 21
2.3.3 Network Training ............................................................. 22
2.3.3.1 Loss Function ............................................................... 22
2.3.3.2 Hyperparameters ........................................................ 22
2.3.3.3 Optimization Algorithm ............................................... 23
2.3.3.4 Network Troubleshooting ............................................. 24
2.3.4 Reference Convolutional Neural Network Architectures ............... 24
2.3.4.1 MobileNet-V2 ............................................................. 25
3 Requirements for Workflow’s Architecture .................................. 27
3.1 Application in Biomedical Research ......................................... 27
3.2 Immunofluorescent Images as Ground Truth ............................... 28
3.3 Ensure Image Registration on input images .................................. 28
3.4 Dataset generation ............................................................... 28
3.5 Cell classification using Deep Learning tools ............................... 29
3.6 Implementation Goals .......................................................... 29
4 State of the Art .................................................................. 31
4.1 Existing annotation-independent architectures ............................. 31
4.2 Comparison and Research Gaps .............................................. 34
5 Methodology .................................................................... 37
5.1 Development Environment .................................................... 37
5.2 Overall Workflow ............................................................... 37
5.3 Selection of Dataset ............................................................. 39
5.4 Data Pre-Processing ............................................................. 39
5.4.1 Image Registration between IF and H&E stained images ............... 39
5.4.2 Cell detection in Immunofluorescent channels ............................................. 41
5.4.3 Convolutional Neural Network training ......................................................... 43
  5.4.3.1 Selection of the network’s architecture .................................................. 44
  5.4.3.2 Hyperparameters optimization ............................................................... 44
  5.4.3.3 Improvement of the network performance .............................................. 45
  5.4.3.4 Prevention of training overfitting ............................................................ 46
  5.4.3.5 TMA core classification using Trained Convolutional Neural Network ... 46
6 Results and Discussion ................................................................................. 47
  6.1 Image Registration ................................................................................. 47
    6.1.1 Evaluation Metrics ............................................................................. 47
    6.1.2 Experimental Results ........................................................................ 49
  6.2 Cell detection in Immunofluorescent Channels ............................................. 51
    6.2.1 Experimental Results ........................................................................ 51
  6.3 Convolutional Neural Network ................................................................. 54
    6.3.1 Evaluation metrics ............................................................................. 54
    6.3.2 Experimental Results ........................................................................ 54
      6.3.2.1 Training Results .......................................................................... 54
      6.3.2.2 Testing Results .......................................................................... 56
  6.4 Requirements Fulfilment .......................................................................... 60
7 Conclusions and Future Work ................................................................. 63
  7.1 Concept of an automatic annotation workflow ............................................ 64
  7.2 Assessment of the automatic annotation workflow ...................................... 64
  7.3 Future Work ............................................................................................ 65
Bibliography ................................................................................................. 67
List of Tables

2.1 Procedure for Image Registration Method Algorithm ............................................. 16
2.2 Common hyperparameters of Convolutional Neural Networks. .............................. 23
2.3 Common troubleshooting problems of Convolutional Neural Networks (based on [46]). . 24
2.4 MobileNet-V2 Architecture [47]. ........................................................................... 26

3.1 Requirements for an automatic annotation architecture using Immunofluorescence stained images. ............................................................... 30

4.1 Classification table for existing automatic annotation architectures. ........................ 34

6.1 Tissue and staining condition of the TMA cores. ..................................................... 52
6.2 Training and Validation datasets generated from the H&E stained TMA cores, using IF technique as ground truth. ............................................. 52
6.3 Testing dataset generated from the H&E stained TMA cores, using IF technique as ground truth. ............................................................... 53
6.4 Optimal training results and respective hyperparameters. ....................................... 56
List of Figures

1.1 Pathologists Tissue Analysis Workflow. ........................................ 3
1.2 Tissue staining Comparison between Hematoxylin and Eosin staining and Immunofluorescent staining. ........................................ 4
1.3 TMA tissue staining Comparison between Hematoxylin and Eosin staining and Immunofluorescent staining. ........................................ 5
2.1 Graphical output representation of a FFT when applied to the different image modalities . 15
2.2 K-means Clustering technique applied to a single TMA H&E stained core. ................. 17
2.3 Effect of the PSF on the resulting image of the given object. .............................. 18
2.4 An illustrative representation of a CNN architecture for classification. ................. 20
2.5 Representation of a Convolution operation on a RGB image [40] ............................ 21
2.6 Representation of a Pooling operation [41]. ............................................ 21
2.7 MobileNet-V1 network main improvements. ............................................. 25
2.8 MobileNet-V2 network main improvements [51]. ............................................. 26
5.1 Overall Workflow’s Architecture .......................................................... 38
5.2 Image registration method. ..................................................................... 41
5.3 Cell detection method including the color intensities of the surroundings. .......... 42
5.4 IF signal intensities threshold method applied to cell detection. ......................... 42
5.5 Overall representation of the cell detection method. ..................................... 43
5.6 CNN training pipeline.(based on figure from MATLAB’s website) ..................... 43
5.7 Comparison between pre-trained Networks. .............................................. 44
6.1 Image Registration subdivided into 36 sections ......................................... 49
6.2 Image Registration results in pixels .......................................................... 50
6.3 Image Registration results in microns ...................................................... 51
6.4 Heat map regarding the CNN’s first training ............................................. 55
6.5 Heat map regarding the CNN’s second training with initial frozen layers .......... 56
6.6 CNN predicted labels on test core 047 (Patient 24) ..................................... 57
6.7 Confusion matrix of test core 047 ............................................................ 57
6.8 Comparison between the CNN correct and incorrect predicted labels on independent test core #047 (Patient 24) ................................................ 58
6.9 CNN predicted labels on test core 208 (Patient 104) ........................................ 58
6.10 Confusion matrix of test core 208 ................................................................. 59
6.11 Comparison between the CNN correct and incorrect predicted labels on independent test core #208 (Patient 104) ................................................................. 60
List of Abbreviations

ADAM  Adaptive Moment Estimation. 23, 45

AI  Artificial Intelligence. 27, 63

CD8  CD8+ cytotoxic T-cells. 5, 39, 41, 42, 52, 53

CNN  Convolutional Neural Network. xvii, xviii, 6, 7, 9, 19–22, 24, 28–30, 32, 34, 35, 40, 43–46, 52, 54–58, 60, 61, 64, 65

DAPI  4’,6’-Diamino-2-Fenil-Indol. 37, 39–43, 48, 52

DFT  Discrete Fourier Transform. 13, 14

DL  Deep Learning. 6, 22, 28, 29, 31, 34, 44, 46, 60, 61, 63–66

DOF  Degrees of Freedom. 12

FFT  Fast Fourier Transform. xvii, 13–15, 48

FOXP3  Forkhead box P3 CD4+ regulatory T-cells. 5, 41, 42, 52, 53

GPU  Graphics Processing Unit. 19, 44

H&E  Hematoxylin and Eosin. xv, xvii, 2–7, 12, 15, 17, 18, 28–30, 32, 34, 35, 37, 39–43, 46–53, 60, 61, 63, 64

IF  Immunofluorescence. xv, xvii, 2, 3, 6, 7, 15, 28–32, 34, 37, 39–43, 46–49, 51–53, 56, 60, 61, 63–65

IHC  Immunohistochemistry. 2, 31, 32, 34, 35, 64

ILSVRC  ImageNet Large Scale Visual Recognition Challenge. 19, 24, 26

ML  Machine Learning. 5, 6

NCC  Normalized Cross-Correlation. 48

panCK  Pan-Cytokeratin. 41, 42, 52, 53
**PSF**  Point Spread function. 18

**ReLU**  Rectified Linear Unit. 20, 21

**SGD**  Stochastic Gradient Descent. 23

**SSD**  Summed Squared Difference. 48

**Th1**  CD4+ T-helper 1. 5

**Th2**  CD4+ T-helper 2. 5

**TILs**  Tumor Infiltrating Lymphocytes. 4

**TMA**  Tissue Microarray. xv, xvii, 5, 6, 9, 15, 17, 27, 28, 30, 34, 37, 39, 40, 43, 46, 47, 50–53, 56, 60, 63–65

**TME**  Tumor Microenvironment. 4, 5

**WSI**  Whole-Slide Imaging. 32, 34, 60
Chapter 1

Introduction

This chapter is accountable for presenting the importance of tissue annotations in the prognostic assessment and introduce important biomedical and computational concepts. Further in this chapter, the objectives established and contributions of this thesis are stated.

1.1 Motivation

According to the US National Cancer Institute [1], breast cancer is the most common type of cancer (276,4680 cases in 2020) and the third leading cause of cancer death (42,170 deaths in 2020) in American women. According to the American Cancer Society [2], 1 in 8 American women will be diagnosed with breast cancer in their lifetime, while 1 in 39 will die from breast cancer [3]. In Portugal, breast cancer represents the most common type of cancer in women. According to the Instituto de Medicina Molecular, 1 in 11 Portuguese women will develop breast cancer in their lifetime [4]. Breast cancer is also the leading cause of death in Portuguese women under 70 years old [5].

Currently, a breast-cancer patient's diagnostic represents a highly demanding task and relies heavily on an expert pathologist's assessment. This assessment includes staging and receptor status of the tumor, as well as scoring and annotations of tumor tissues, which can be sometimes inaccurate and inconsistent, with large variations from pathologist to pathologist [6]. Furthermore, in biomedical research, annotations of tumor tissues by experts are usually time-consuming and cannot be produced at scale, often leading to biased datasets [6, 7].

In recent years, advances in technology have allowed the development of improved computational processing power and storage. As a result, this improvement has had a direct impact on the current state of the art, by producing faster, more reliable and ameliorated algorithms in biomedical applications, in particular in oncology. In particular, deep learning (DL) algorithms have played a crucial role in this improvement, featuring multiple advantages for biological investigations [8]. More specifically, the use of Convolutional Neural Networks (CNNs) to analyze large image datasets, containing several levels of abstraction, has been a recurrent approach.

Although recent advances in technology have allowed scientists to surpass several barriers, the
solutions for this problem are still very limited. Over the last few years, an increased interest in a more reliable ground truth method has been noticed. However, regarding the ground truth selection, the state of the art displays an homogeneous pattern, being most of the research developed mainly using Immunohistochemistry (IHC) staining technique (only capable of labeling one marker per tissue section [9]) and lacking a robust workflow capable of classifying cell subtypes.

1.2 Pathological and Histological Image Analysis

In this section, basic concepts are addressed regarding the tissue analysis made by pathologists, tissue staining techniques, the tumor environment importance for treatment purposes and, lastly, tissue imaging formats.

1.2.1 Standard Tissue Analysis

When an abnormal form or mass is detected through screening imaging, often through a mammography test or Magnetic Resonance Imaging (MRI), pathologists are required to make an assessment whether the mass is benign or malignant. Hence, in order to diagnose if a patient has breast cancer or not, several steps must be taken as shown in figure 1.1.

The first step involves removing a sample of the abnormal mass (biopsy), followed by the specimen preparation. Once this step is complete, the specimen is stained with Hematoxylin and Eosin (H&E) and Pathologists determine if the sample is benign or malignant (cancerous) via visual inspection. In case a patient is diagnosed with breast cancer, further tests need to be made. To decide the patient's prognostic and to better predict the clinical outcome of breast cancer patients, the abnormal tissue sample is also tested using biomarkers. More precisely, the sample is stained using Immunohistochemistry (IHC) and/or IF staining techniques in order to test it for the estrogen receptor (ER), progesterone receptor (PR) and for HER-2 status of the patient [10]. In conclusion, by using fluorescent biomarkers to target cell subtypes, a more precise assessment about the patient's diagnosis, prognosis and clinical treatment can be made.
1.2.2 Staining Techniques

1.2.2.1 Hematoxylin and Eosin Staining

As aforementioned, the H&E staining method plays a crucial role in determining if a tissue sample is benign or malignant. This method is a cheap and ubiquitous staining method used worldwide in cancer diagnostication. Additionally, the image modality of the H&E staining method is included on the light microscopy field, which collects the light diffracted from the object and generate a digital representation in the microscope sensor [11]. From the staining point of view, the Hematoxylin in the H&E is responsible for staining all the cell nuclei with a dark blue color, as shown in figure 1.2(a). While the Eosin in the H&E stains the extracellular matrix and cytoplasm with a pink color. Finally, this method has some limitations, which is reflected mostly in the inability of identifying cell subtypes.

1.2.2.2 Immunofluorescent staining

To detect specific antigens on biological samples, Immunofluorescent (IF) staining comes as one of the most used techniques in research. This method uses antibodies to target specific subtypes of cells, which contain their matching antigen. At the same time, a fluorescent dye is added to these antibodies, allowing a visual location of the respective targeted cells, as shown in figure 1.2(b). The fluorescent digital images are in a different modality than the H&E images, as the fluorescent antibodies emit a specific wavelength that is acquired by the Charge-Coupled Device (CCD) camera sensors [12]. Although it is a well-established tool, its staining process is not trivial as it relies on multiple factors, which can result in weak signaling and incompatibility between biomarkers [9].
1.2.3 Treatment Methods

Depending on the current breast cancer stage, different treatment methods can be applied. If detected in early stages, the patient can be submitted to surgery to remove the tumor, followed by additional chemotherapy or radiation therapy.

Chemotherapy consists in giving the patient a group of anti-cancer drugs to destroy cancer cells in the entire body. Moreover, in recent years, a new and highly promising cancer treatment started to gain relevance in the scientific community called Immunotherapy. This particular treatment uses substances made from living organisms as a cancer treatment solution. More precisely, it uses one person’s immune system to dispute cancer, either by stimulating or inserting substances into the immune system [13]. Although breast cancer was not often described as being an indicated tumor to be administered Immunotherapy, this position will most likely change in the near future, as more than 70 new trials of Immunotherapy drugs opened in 2018 alone, according to Esteva [13].

1.2.4 Tumor Microenvironment

Tumor Microenvironment (TME), representing the surrounding cells to the tumor mass, has been shown to play a critical role in tumor progression. Depending on the healthy state of the TME, an assessment can be made referring either if a tumor will progress into advanced malignancies or if it will provide support against further invasions [14]. Lastly, one of the long recognized constitutions of the TME is Tumor Infiltrating Lymphocytes (TILs).

Assessing (TILs) has been shown to play an essential role when it comes to predict the patient response to therapy (Chemotherapy and Immunotherapy) and survival [15]. TILs studies the presence
of immune cells inside the invasive tumor and in the surrounding of the tumor cells. This study is not only performed at quantitatively level, as it also studies the types of infiltrating lymphocytes (phenotype). More precisely, CD4+ T-helper 1 (Th1) cells and CD8+ cytotoxic T-cells (CD8) provide essential aid in facilitating the antigen presentation and in destroying the tumor cells, respectively. Whereas, cells of type 2 CD4+ T-helper cells (Th2) are often linked with encouraging proliferation of B lymphocytes and Forkhead box P3 (FOXP3) CD4+ regulatory T-cells associated to the promotion of anti-inflammatory immune responses, resulting in tumor growth [16].

In conclusion, by assessing the TME and immune cells present both inside and outside the invasive tumor, predictions about tumor growth can be provided and information regarding the patient response to therapy can be obtained.

1.2.5 Tissue Microarray

First reported by Kononen et al. in 1998 [17], the Tissue Microarray (TMA) technique consists in placing hundreds of tissue samples on a microscope slide, as it can be seen in figure 1.3. This method allows researchers to analyze a significant number of tissues in a single experiment, from diseases at different stages and from multiple patients [18].

When applied to the study of breast cancer, TMAs are usually the product of hundreds of tissue biopsies from the resected tumors. Then, these tissue samples can follow the same preparation and staining procedures of regular tissue samples, as previously shown in figure 1.1. Once the TMA is prepared, large datasets can be built in combination with the patients clinical data. As a result, the generation of large varieties of data sets, can allow the creation of new fields of studies, which would mainly focus on the interpretation on large quantities of data.

In recent years, one of the most promising and disruptive fields is Machine Learning (ML). However, one of the biggest ML constraints is the constant requirement for big databases. In Biomedicine, TMAs emerge as a strong candidate, capable of providing ML algorithms large quantities of data.

Figure 1.3: TMA tissue staining Comparison between Hematoxylin and Eosin staining and Immunofluorescent staining.
1.3 Integration of Deep Learning in Biomedicine

1.3.1 Historical Background

Machine learning (ML) technology has been performing an active role in modern society and is frequently described as automated methods for data analysis. For Murphy, these automated methods can detect patterns in the data, which then can be used to perform decisions under uncertainty as well as forecast future data [19].

Deep Learning (DL) is considered to be a subtype of Machine Learning and has its roots in the early 1990’s, when LeCun et al. established a convolutional network able to perform the task of classifying hand-written digits and face recognition. More precisely, DL consists of computational models composed by multiple processing layers, which apply non-linear transformations to the data and are able to learn from these representations. In Biomedicine, DL can be applied to study data with multiple levels of abstraction and to identify types of tissues or cells [20].

1.3.2 Convolutional Neural Networks as a solution for Automatic Annotation

As seen in section 1.2.1, the conventional diagnostic procedure relies on tissue samples annotations, which is carried by Pathologists. This task is highly demanding and time consuming and, as a consequence, too much pressure is imposed on Pathologists. Alternatively, Convolutional Neural Networks (CNN) offer the possibility of accurately segment or classify tissue sections. As a result, a better morphological study of the classified objects could be further developed.

Additionally, the generation of biomedical datasets for research purposes, can be one further addressed application in the biomedical scientific community. This biomedical datasets, would be consisted of labelled image cells and could be tuned to the researchers preferences. Additionally, this datasets could be produced in larger quantities, at a faster pace and without requiring an annotation-expert.

1.4 Objectives and Contributions

This thesis’ objective is to design and develop a workflow capable of automatically annotating H&E sections at single cell resolution, using IF images as ground truth. This workflow starts by first aiming at an effective method of performing image registration between IF images and H&E images of a whole TMA slide. The second aim is to automatically detect immune cells (CD8+ cytotoxic T cells and FOXP3+ regulatory T cells), cancer cells (pan-cytokeratin+) and blood vessels (CD31+ endothelial cells) in IF images and annotate H&E images by transferring the detected (x, y) coordinates of the cells of the IF images. The last aim is to train and validate a CNN model capable of automatically detect immune cells, cancer cells and blood vessels in H&E sections, without requiring the IF images input.

The main contribution of this thesis is the development of a robust tool, capable of processing TMA images, aligned them, generate labelled image data sets from a reliable ground and, lastly, annotate/score large quantities of data in very limited time. Another contribution of this dissertation, reflects
the capacity of using a trained CNN to classify cell subtypes on H&E stained tissue sections, without requiring the IF images input or expert-annotations’ input (e.g. Pathologists input). The last contribution of this work is the support it can give to Pathologists, when determining the diagnosis and the prognosis of a patient, as it can build a fast, reliable and larger assessment of the tumor tissue sections. Additionally, this work can also support Researchers, by providing labelled cell image data sets and enhance the development of several fields of study.

1.5 Thesis Outline

This thesis is structured in six chapters, the first chapter was intended to give the reader an introduction about important concepts and relevant background in breast cancer, staining techniques and the growing role of Deep Learning in Biomedicine. In chapter 2, theoretical concepts and principles related to this work are explained, more precisely, about image registration, image processing techniques (cell detection) and Convolutional Neural Networks. Chapter 3 gives an essential insight in the dissertation’s requirements. In Chapter 4, an overview of the state of the art is given, an important comparison is established between the works developed and possible gaps in current practices are indicated. Chapter 5 is characterized by presenting the overall workflow and then stepping into individual processes of the respective workflow developed. In this chapter, it is given an insight on the steps and decisions elaborated during development of the workflow. Chapter 6 or, more precisely, results and discussion’s chapter describes the evaluation metrics utilized in the dissertation’s results, followed by an in-depth discussion and analysis of the same. Lastly, in Chapter 7, most important achievements of this thesis are delineated, and potential future work is outlined.
Chapter 2

Theoretical Background

The objective of this chapter is to give background about the theoretical concepts and principles related to the work developed by this dissertation. In section 2.1, it is explained the concept of Image Registration and, more importantly, the image registration method performed in this thesis. Section 2.2 exhibits additional image processing techniques and various image analysis techniques used throughout the dissertation, such as cell detection method, K-means clustering and centroid location. Finally, in section 2.3, the fundamentals of a Convolutional Neural Networks’ (CNN) architecture are point out, some essential CNN parameters are described and an insight on the implemented network is given.

2.1 Image Registration

Image Registration is a very utilized tool in the engineering field and is, by definition, the process of transforming different images into one coordinate system. In Biomedicine, Image Registration is performed to overlay images designated to be examined [21]. This approach is particularly beneficial when images display dissimilar dimensions, generally due to tissue movements during staining processes and/or as a result of tissue imaging performed using different examination machines. In this dissertation, Image Registration plays a fundamental role in the development of reliable data bases. More precisely, this technique is used to perform an alignment between the Immunofluorescent stained TMA images and the Hematoxylin and Eosin TMA stained image.

2.1.1 Homologous Points

When performing an Image Registration process, a reference image is defined and a geometric transformation is estimated, resulting in the alignment of the source images (images to be registered) accordingly to the reference coordinate system.

In order to have a competent image registration performance, the selected images involved in the process should exhibit an adequate number of similar features among themselves. This allows to identify common feature points or Homologous points that are present in the reference image and in the source images. To find the parameters of a transformation model to align images, the homologous points have
to be determined. The method to detect this points can vary and is dependent on the modality of the images involved in the process [22].

### 2.1.2 Transformation Functions

Image Registration can be subdivided into distinct categories depending on the image transformation performed, which can abridge to *Rigid Body*, *Affine*, *Projective*, and *Nonrigid* transformations [23]. In fact, these transformations can be described by the degrees of freedom or, in other words, the number of parameters required to characterize a transformation.

As mentioned previously, image registration estimates a geometric transformation between the moving image and the reference image. Hence, in a two dimensional space, given a set of n points \( p_i = (x_i, y_i) \) present in the reference image and the corresponding set of n points \( P_i = (X_i, Y_i) \) in the source image, it is possible to estimate the transformation function \( f \) between these two images. As a result, the relation between the reference image and the moving images can be represented by its transformation components [24]:

\[
\begin{align*}
X_i \approx f_x(x_i, y_i) &= f_x(p_i), \\
Y_i \approx f_y(x_i, y_i) &= f_y(p_i),
\end{align*}
\]

(2.1)

This transformation function \( f \) can be defined in several ways, depending on the main objective and focus of each image registration technique. When choosing the appropriate class for the transformation function, it is essential to analyze the main differences between the reference image and the test image. In particular, to understand the accuracy amongst the reference points and the moving correspondent points, the spatial density and organization of the control points. Once assessed the degree of misalignment between the images involved, the required transformation functions are selected. Some of the most used transformations are translation, rotation and scaling functions.

#### 2.1.2.1 Translation

Translation is one of the most commonly performed transformation function in the image registration field. As the name indicates, it performs a translation on each point according to each individual axis, as it is shown in equation 2.2. The transformed point, originated by the translation transformation, is denoted \( p'_i = (x'_i, y'_i) \) and preserves distances between the points, angles and parallelism. If the image registration technique is faultless, this point would necessarily correspond to corresponding point \( P_i = (X_i, Y_i) \) in the reference image.

\[
\begin{align*}
x'_i &= x_i + t_x, \\
y'_i &= y_i + t_y,
\end{align*}
\]

(2.2)

The equations 2.2 can be rewritten using matrix notation as displayed in equation 2.3 [24], which is displayed in Cartesian coordinates and corresponds to a non linear transformation. However, when converted to homogeneous coordinates, it assumes linear properties, as shown in equation 2.4 [25].
\[ p' = T + p = \begin{bmatrix} t_x \\ t_y \end{bmatrix} + \begin{bmatrix} x \\ y \end{bmatrix} \]  
(2.3)

\[
\begin{bmatrix}
  x' \\
  y' \\
  1
\end{bmatrix} =
\begin{bmatrix}
  1 & 0 & t_x \\
  0 & 1 & t_y \\
  0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
  x \\
  y \\
  1
\end{bmatrix}
\]  
(2.4)

### 2.1.2.2 Rotation

When the moving image appears rotated with respect to the reference image, a rotation transformation is suggested. The rotation transformation, shown in the equation 2.5, is performed in relation to the origin of the image [23]. Furthermore, it is also preserves the angle between the lines.

\[
\begin{align*}
x'_i &= x_i \cdot \cos(\theta) - y_i \cdot \sin(\theta), \\
y'_i &= x_i \cdot \sin(\theta) + y_i \cdot \cos(\theta),
\end{align*}
\]  
(2.5)

This transformation can be written as follows, using matrix notation [24]:

\[
p' = R \cdot p = \begin{bmatrix}
  \cos(\theta) & -\sin(\theta) \\
  \sin(\theta) & \cos(\theta)
\end{bmatrix}
\begin{bmatrix}
  x \\
  y
\end{bmatrix}
\]  
(2.6)

\[
\begin{bmatrix}
  x' \\
  y' \\
  1
\end{bmatrix} =
\begin{bmatrix}
  \cos(\theta) & -\sin(\theta) & 0 \\
  \sin(\theta) & \cos(\theta) & 0 \\
  0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
  x \\
  y \\
  1
\end{bmatrix}
\]  
(2.7)

When the points in the image are subjected to translation and rotation transformations, the process is termed **Rigid body transformation** or **Euclidean transformation**.

### 2.1.2.3 Scaling

A scaling transformation is a linear transformation that scales the points according to each image's axis. In a two dimensional space, this coincides with scaling the points in the \(x\) and \(y\) axis, \(s_x\) and \(s_y\), respectively. The result of this operation is visible in the equation 2.8, in a two dimensional space.

\[
\begin{align*}
x'_i &= s_x \cdot x_i, \\
y'_i &= s_y \cdot y_i,
\end{align*}
\]  
(2.8)

From the previous equations, this transformation can be written using matrix notation:

\[
p' = S \cdot p = \begin{bmatrix}
  s_x & 0 \\
  0 & s_y
\end{bmatrix}
\begin{bmatrix}
  x \\
  y
\end{bmatrix}
\]  
(2.9)
2.1.3 Similarity Transformation

In this dissertation, the geometric transformation utilized is *Similarity Transformation*, which includes translation, rotation and scaling operations. The similarity transformation’s mathematical statement is represented in the equation 2.11 [24]. This transformation represents \( s_x \) and \( s_y \) as the same numerical value or singular scaling factor of \( s \).

\[
\begin{bmatrix}
    x' \\
    y' \\
    1
\end{bmatrix} = \begin{bmatrix}
    s_x & 0 & 0 \\
    0 & s_y & 0 \\
    0 & 0 & 1
\end{bmatrix} \begin{bmatrix}
    x \\
    y \\
    1
\end{bmatrix}
\]

\( (2.10) \)

Under the Similarity transformation, a combination of the previously mentioned transformation functions in sections 2.1.2.1, 2.1.2.2 and 2.1.2.3. In particular, this dissertation uses translational, rotational ans scaling transformations to overcome the misalignment between the moving Immunofluorescent images and the reference H&E image. The matrix representation of the similarity function results from the multiplication of these matrix transformation, as shown in equation 2.12. The same result can be obtained by representing the mathematical statements in equations 2.11 [24].

\[
p' = T \cdot R \cdot S \cdot p \Leftrightarrow \\
\begin{bmatrix}
    x' \\
    y' \\
    1
\end{bmatrix} = \begin{bmatrix}
    1 & 0 & t_x \\
    0 & 1 & t_y \\
    0 & 0 & 1
\end{bmatrix} \begin{bmatrix}
    \cos(\theta) & -\sin(\theta) & 0 \\
    \sin(\theta) & \cos(\theta) & 0 \\
    0 & 0 & 1
\end{bmatrix} \begin{bmatrix}
    s_x & 0 & 0 \\
    0 & s_y & 0 \\
    0 & 0 & 1
\end{bmatrix} \begin{bmatrix}
    x \\
    y \\
    1
\end{bmatrix}
\]

\( (2.12) \)

The main characteristic of similarity transformations relies on the preservation of the parallelism between parallel lines in the images, as well as the ratio of lengths remain invariant. By analyzing the matrix representation of a similarity transformation in equation 2.12, it is possible to extract the \( s_x \), \( \theta \), \( t_x \) and \( t_y \) independent parameters and, therefore, the number of degrees of freedom (DOF). More precisely, this means that the four DOF of this transformation are correlated and comprise the four previously mentioned independent parameters.

2.1.4 Parameter Estimation

To estimate the parameters of an image alignment process, a registration method needs to be selected. Registration Methods are subdivided into Feature-based registration methods and Area-based registration methods.

Feature-based registration methods are the most commonly used techniques. This method exploit
features of images in order to establish a alignment pattern between the respective images, followed by
the association of the corresponding points and estimation of the geometric transformation. However, in
some particular cases, these methods tend to show some poor performance accuracy. More precisely,
these methods show low performance when displaying the position and distribution of the feature points
[26].

Area-based registration approaches use the image pixel values or its frequency domain to establish
a pattern between the reference image and the source image. Contrary to the Feature-based methods,
the Area-based approach shows robustness to slight content modifications in the registration process.
On the other hand, this procedure has the disadvantage of being sensitive to noise in the image, as it
reduces the peak value [26].

2.1.5 Registration Method

In this thesis, the method selected was an Area-based approach. This decision was supported by the
capacity of the method in finding pattern between images with slightly different content. As mentioned
in section 1.4, this condition corresponded to one of the main difficulties for this dissertation’s image
registration process. From this category of Registration methods, it was established the Fast Fourier
Transformation (FFT)-based image registration method developed by B. Reddy et al. [27] in order to
better determine the similarity’s parameters.

2.1.5.1 Fast Fourier Transformation-based image registration

The Fourier Transformation is a widely used technique. When applied in the image processing field, it is
mostly used to transform images from the spatial domain into the frequency domain, in order to achieve
a more compact image representation.

Through the conversion of images to the frequency domain, image patterns are revealed in the
respective images, which could not be visualized in the spatial domain. When applied in the image
registration field, this characteristic allows a deeper study and understanding of the images, as it can
enhance the perception of various image features. As a result, images with different modalities or
obtained through different sensors can be aligned using the Discrete Fourier Transformation (DFT) as
the technique to estimate the images registration parameters.

Another characteristic when transforming the domain of two dimensional images, is it allows to de-
compose the images into sine and cosine components, each one consisted with different amplitudes
and phases. Hence, considering \( f(x, y) \), as a finite the two dimensional (NxM) image in the spatial
domain, and \( F(u, v) \), as the image representation in the frequency domain, it is possible to describe the
2D DFT in equation 2.13. Where \( u \) and \( v \) do represent spatial frequencies and \( |F(u, v)| \) represents the
magnitude spectrum. If we acknowledge \( F_I(u, v) \) as the imaginary part of the image in the frequency
domain and \( F_R(u, v) \) as its real part, the phase angle spectrum can be obtained by computing the
\( \arctan(F_I(u, v)/F_R(u, v)) \) and, consequently, represented similarly to figures 2.1(c) and 2.1(f). Finally,
the inverse Fourier transformation can be obtained through the equation 2.14.
\[ F(u, v) = \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} f(x, y) e^{-j2\pi(\frac{xu}{N} + \frac{vy}{M})}, \quad (2.13) \]

\[ f(x, y) = \frac{1}{NM} \sum_{u=0}^{N-1} \sum_{v=0}^{M-1} F(u, v) e^{-j2\pi(\frac{xu}{N} + \frac{vy}{M})}, \quad (2.14) \]

However, one of the major inconveniences of the DFT process relies on its very time consuming aspect. With this in mind, the Fast Fourier Transform (FFT) is an algorithm that has gain a significant relevance in the last years, by providing a more efficient approach when converting signals. This characteristic comes particularly evident for a large number of operation N, where the DFT would compute in \( \Theta(N^2) \) time and FFT would compute the same amount of operations in \( \Theta(N\log_2(N)) \). If we consider the time of one operation to be equivalent to one nanosecond and the amount of operations performed equal to ten million, this would result in 27.78 hours expended for the DFT and the identical scenario would be computed in 0.23 seconds for the FFT. In the image Registration field, this FFT feature comes very convenient when processing a large amount of images and with large dimensions.

The FFT computation method consists in dividing the original image vector into two halves, calculate the DFT in each half recursively and, in the end, merge the results. Its output consists in an array of complex numbers that can be regrouped into a two dimension representation, for a more straightforward analysis. These number can be centered according to the frequency category, by shifting the zero-frequency component to the center of the spectrum, as shown in images 2.1(b) and 2.1(e). This operation displays the high frequencies (provide detail to the image) in the periphery of the FFT image output, while the low frequencies (shapes and patterns of the image) are displayed in the center of the spectrum.

Reddy and Cattterji [27], developed an FFT-based image registration method without requiring any control points from both reference and source images. This algorithm presents an efficient technique to estimate the translation, rotation and scaling image registration parameters.
Figure 2.1: Graphical output representation of a FFT when applied to the different image modalities

The alignment method proposed by Reddy and Catterji [27] is described in Table 2.1. Using this method, the similarity’s parameters can be obtained using phase correlation. This process can be broken down into two main stages: the first step consists in angle and scale estimation, while the following proceeding consists in estimating the shifting (translation) between the reference image and the source images. For the angle and scale estimation, it is possible to obtain it using the equation 2.21, considering $\xi = \log(r)$ and $c = \log(s)$ and applying the phase correlation technique. When estimating the translation shift between images, the method consists in locating the coordinates of the main peak of the inverse Fourier transform. This coordinates represent the translation required to align the images, and can be obtained through the normalized cross-power spectrum matrix, using equation 2.18 [26].

Additionally, I note that the image input format was processed under a range of shades of gray, which can extend from black to white (Grayscale).
Table 2.1: Procedure for Image Registration Method Algorithm

**Input**
Reference image \((f_2)\) and Source Image \((f_1)\), with a displacement of \((x_0, y_0)\), rotation of \(\theta_0\) and a scale factor of \(s\) relatively to \(f_2\), as exhibit in the following equation:

\[
f_2(x, y) = f_1(s \cdot x \cdot \cos(\theta_0) + s \cdot y \cdot \sin(\theta_0) + x_0, -s \cdot x \cdot \sin(\theta_0) + s \cdot y \cdot \cos(\theta_0) + y_0)
\]  
(2.15)

**Step 1: The Fourier Shift theorem**
The theorem states that a translation on an image can be estimate by computing the following equation:

\[
f_2(x, y) = f_1(x + x_0, y + y_0)
\]  
(2.16)

And the same property is also applied in the Fourier Transform:

\[
F_2(u, v) = e^{-j2\pi(ux_0+vy_0)} * F_1(u, v)
\]  
(2.17)

**Step 2: The Normalized Cross-power spectrum**
From the shift theorem in equation 2.17, it is possible to isolate the phase difference between the images, by calculating the cross-power spectrum [27]:

\[
\frac{F_1(u,v)F^*_2(u,v)}{|F_1(u,v)||F_2(u,v)|} = e^{-j2\pi(ux_0+vy_0)}
\]  
(2.18)

The inverse Fourier Transform of this equation results in an impulse, which its maximum value is at \((x_0, y_0)\). This impulse displays the shift location and is zero in the rest of the function.

**Step 3: Conversion to polar coordinates**
From equation 2.17, the Fourier transform can be computed, with the axis converted to the polar representation [28]:

\[
F_2(r, \theta) = e^{-j2\pi(u x_0 + w_0 y_0)} \cdot \frac{1}{s^2} F_1(\frac{r}{s}, \theta - \theta_0)
\]  
(2.19)

\[
|F_2(r, \theta)| = \frac{1}{s^2} |F_1(\frac{r}{s}, \theta - \theta_0)|
\]  
(2.20)

**Step 4: Conversion to logarithmic scale**
With the axis represented in the logarithmic scale, scaling can be characterized as a translational displacement [28]:

\[
|F_2(log(r), \theta)| = \frac{1}{s^2} F_1(log(r) - log(s), \theta - \theta_0)
\]  
(2.21)

Neglecting the factor \(\frac{1}{s^2}\) in equation 2.21, and considering \(\xi = log(r)\) and \(c = log(s)\) [27, 28]:

\[
|F_2(\xi, \theta)| = |F_1(\xi - c, \theta - \theta_0)|
\]  
(2.22)

**Output**
From equations represented in 2.21 and 2.22, the rotation angle \(\theta_0\) and scale \(s\) can be computed. As a result, the shifting, caused by the images misalignment, can be obtained by using 2D phase correlation technique.
2.2 Image Analysis

In this section, several image analysis methods are presented with the purpose of aiding the implementation and processing of the algorithms utilized in this thesis workflow. The main image processing techniques exploited in this dissertation are K-means Clustering applied to H&E stain separation in biological images, TMA centroid segmentation using several image processing techniques, peak color intensity detection dedicated to nuclei detection in H&E images and, lastly, the color intensity threshold technique to select the intended cells subtypes, as well as remove detected noise from each image.

2.2.1 K-means Clustering

The K-means Clustering technique consists in clustering data into groups or clusters. Each cluster contains similar features amongst the data inside it, but different features from one cluster to another one. This clustering techniques can present multiple variations, depending on the cluster initialization and distance measuring technique of the nearest centroid [29]. In biomedicine, this tool is particularly beneficial when it comes to enhance specific types of cells from the image background.

In this dissertation, K-means Clustering is implemented to segment all the nuclei present on the H&E stained image [30], as shown in figure 2.2(a). The output consists on a two dimensional image containing only all the dark blue objects (nuclei color) or the equivalent to the Hematoxylin channel, as explained in section 1.2.2.1. From this output, displayed in figure 2.2(b), the background is removed and the image only displays the targeted cell nuclei. Hence, by applying the cell detection algorithm to the k-means clustering nuclei segmentation output, the cell detection performance is improved.

The K-means Clustering algorithm applied to image segmentation, consists in an iterative process to accredit K number of clusters to the image. This method assigns K cluster centroids and computes the euclidean distances of every pixel in relation to those centroids. Finally, all the pixels are designated according to its nearest centroid. Thereupon, new centroids are recalculated, the euclidean distances are computed in order to this new centroids and each pixels reassigned to the new closest centroid. This process is repeated until it the minimum euclidean distance that satisfies the tolerance [29].

(a) Single TMA core stained with H&E  
(b) K-means Clustering Hematoxylin channel output of the single TMA core stained with H&E

Figure 2.2: K-means Clustering technique applied to a single TMA H&E stained core.
2.2.2 Peak color intensity detection

To detect cells in H&E stained images, an algorithm was implemented to measure the intensity peaks existing in each tissue section core. To implement this algorithm, the Hematoxylin channel (all cell nuclei) of the H&E images was extracted, resulting in a one channel gray scale image, as displayed in figure 2.2(b). In this image, spots containing higher intensity values, correspond to cells existent in the tissue core. As a result, an Fast 2D peak finder algorithm [31] was implemented to detect all cell nuclei (represented as white dots) in Hematoxylin image.

To better understand the functioning of the cell detection method, it is important to have the correct understanding about how each peak is represented in general imaging techniques. More precisely, cells are often displayed as a smooth point spread function (PSF). This representation is often utilized to describe the light distribution of point objects in microscopy. In practise, by measuring the intensity values through a line on each bright center (corresponding to each cell), the bright center or maxima is often represented as a Gaussian Distribution of a particular dimension. In conclusion, the output image of the microscope, results from the convolution between each point source of light, in the specimen, with the PSF, as shown in figure 2.3.

![Figure 2.3: Effect of the PSF on the resulting image of the given object [32].](image)

The Fast 2D peak finder algorithm allows to process 2D images, to locate the local maxima. In order to accurately target each peak, and assuming the PSF behaviour aforementioned, it is essential the assumption of detected noise (such as "salt and pepper") during the imaging process. Therefore, a 2D median filter function is implemented, which eliminates "salt and pepper" noise and a 2D convolution function, which smooths the image, are implemented in order to obtain only one PSF local maximum for each cell.
2.3 Convolutional Neural Network

Deep learning is considered to be a subtype of Machine Learning. For LeCun, deep learning consists of computational models which are composed by multiple processing layers, which apply non-linear transformations to the data. From these layers, the model is able to learn from these representations of data with multiple levels of abstraction [33].

For Goodfellow et al.[34], Convolutional Neural Networks (CNNs) are a class of neural networks (deep learning algorithm) that perform a linear operation called convolution, followed by a non-linearity operation (activation function). Instead of multiplying a general matrix, convolutional neural networks perform convolutions, which require less parameters. Additionally, CNNs are specialized in processing data under the image format. The process in which CNNs perform, can be related to how the brain processes the information extracted from one image. In addition, CNNs also displays some similar biological influences, as they are invariant to translations and share the concept of receptive fields.

The property of invariance to translations states the ability to detect the same input, when translated to some other location in the image [35]. This characteristic has become particularly useful towards detecting certain features in the object, despite its position on the image [34].

Regarding the concept of receptive fields, it explains the method as each neuron in our brain responds to stimuli. In the biological visual system, this feature addresses the ability to respond to stimuli belonging to restricted regions of the visual field [36]. Similarly, the same role is observed on each CNN’s neuron method to process data.

2.3.1 Historical Background

Deep Learning has its roots in the early 1990’s, when LeCun et al. successfully implemented a convolutional network capable of classifying hand-written digits and face recognition. The main objective for this network was focused on simplifying the network as much as possible, by reducing the number of parameters present in the network [37].

In 2010, the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) revolutionized the image analysis field. ILSVRC made available a classification challenge of a 15 million labeled high-resolution image data set, organized into nearly 22,000 different categories [38]. By using improved efficient Graphics Processing Unit (GPU), together with this growing availability of larger labeled data sets, allowed CNNs to accomplish remarkable benchmarks.

2.3.2 Architecture Overview

As mentioned in the previous section, Convolutional Neural Networks are known for applying a series of transformation, on hidden layers, to the input image. This operations observed in the hidden layers, consists on adjustments on the neurons’ learnable weights, in each fully connected layer. All neurons, in each layers, are connected to all neurons on previous layers. Additionally, filters are implemented to extract particular features of the images before they are introduced into the fully connected layers. The
resulting output, of the last fully-connected layer, comprises all class scores.

Figure 2.4 displays a basic CNN architecture, as well as an illustrative representation of the data in each layer. In particular, it allows a more in-depth understanding of the CNN's architecture, a visual representation of the transformations occurring at different stages of the network and, lastly, the probabilistic output of each class.

[Diagram of a CNN architecture]

Figure 2.4: An illustrative representation of a CNN architecture for classification [39].

Generally, Convolutional Neural Networks consist in the implementation of repetitive Convolutional, Pooling and Fully Connected layers in sequence. Additionally, Convolutional layers are often followed by a non-linear layer, which in figure 2.4 is represented as a Rectified Linear Unit (ReLU).

2.3.2.1 Convolutional Layers

On a convolution layer, given an image input (eg. function I) and the kernel (eg. function K), its convolution is similar to equation 2.23. In this equation, I represents a two-dimensional image input and K represents a two-dimensional Kernel [34]. The Kernel is spatially smaller than an input image (height and width), but extends into the depth of all image channels [36]. The output resultant of this convolution is often called feature map.

\[
S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(m, n)K(i - m, j - n)
\]  

(2.23)

As seen in figure 2.5, a value is outputted from the dot product of the three-dimensional image input with the three-dimensional kernel filter. Furthermore, the same operation is repeated in all spatial locations in the image [35]. The displacement size of the kernel is called a stride [36]. By doing so, a feature map is created based on that respective filter.

2.3.2.2 Non Linear Layers

The convolution is a linear operation, conducted in the convolutional layers. To enhance the class of transformations carried out by each layer, non-linear activation functions are applied after convolutions [36]. Some of the most commonly used non-linear activation functions are Sigmoid, Tanh and ReLU.

Regarding the Rectified Linear Unit, its selection is often used due to its reliability and efficiency. ReLU layers threshold their input values at zero. This means that for values bigger than zero the value remains unchanged, but for values smaller than zero the value outputted is zero. As a result, ReLU layers don’t interfere with the size of the volume.
2.3.2.3 Pooling Layers

Usually, after a ReLU layer, a Pooling layer is introduced. The main objective of the pooling process is to reduce processing time, allowing the CNN to perform with reduced spatial size representations. Pooling layers reduce the resulting size of the volume [35].

One of the most commonly used Pooling layers algorithms is Max Pool. This method uses a filter to go through the image, as shown in image 2.6. As a result, the maximum value is returned, from the portion of image covered by the kernel [40]. Lastly, this operation allows the extraction of dominant features and maintain the efficiency of the model [40, 41]. Additionally, the pooling operation provides translation invariance, which enables the CNN to extract features, regardless of the spatial position of the feature [34].

2.3.2.4 Fully Connected Layers

The fully connected layers, located in the classification layers, are usually present in the end of the CNN, as shown in figure 2.4. Firstly, the converted input image is flattened into a column vector. Subsequently, the fully connected layers, establish a connection between neurons in the current layer and the neurons locate in previous (inside the Fully Connected Layers). Finally, the resulting output of this operation, allows the possibility of learning from a non-linear function and culminates in the class scores [35].

One of the most used algorithms for classification is the Softmax, which computes the probability distribution of the classes, given its input. In practise, the output generated by the algorithm goes from zero to one and is assigned to each CNN label trying to predict [40, 41]. The resulting sum of all probabilities, predicted for all the CNN labels, is always equal to one.
2.3.3 Network Training

Through training, the network’s weights can be regulated in order to achieve better prediction accuracies. This is done by minimizing a loss function that evaluates the performance of the network in the training set of data.

2.3.3.1 Loss Function

Convolutional Neural Networks are evaluated by the means of a loss function, which measures the inconsistency between the predictions ($\hat{y}$) and the actual results ($y$) \[?\]. In order to improve the CNN’s performance, a reduction on the cost function ($\mathcal{L}$) must be observed. Ultimately, a general mathematical representation of the cost function is addressed in equation 2.24 [34].

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^{N} L(y^{(i)}, \hat{y}^{(i)}) \quad (2.24)$$

Regarding $L(y^{i}, \hat{y}^{i})$, distinct expressions can be implemented, according to the type of learning being prioritized. The main loss functions subdivisions are between a regression problem or classification problem. In respect to regression losses, the most common measurement of performance is the Mean Squared Error (MSE). Regarding the classification losses, the most commonly used are Cross Entropy loss function and SVM loss function [34].

For multiclass classification, Cross Entropy is often the preferred approach. Cross Entropy or frequently named log loss, establishes a performance comparison amongst the probability distribution, outputted from the Softmax function, and its true class label [42]. This comparison is given by equation 2.25, where $N$ is the number samples, $K$ being the number of classification classes, $t_{ij}$ represents a binary variable (indicator variable) that is equal to 1 if $j$ is the true class for sample $i$ and, lastly, $\hat{y}_{ij}$ represents the $i$th output probability distribution of the network for the $i$th training example.

$$\mathcal{L} = - \sum_{i=1}^{N} \sum_{j=1}^{K} t_{ij} \ln(\hat{y}_{ij}) \quad (2.25)$$

2.3.3.2 Hyperparameters

When training a DL model, several hyperparameters can be tuned in order to maximize the performance of the network. In particular, some of the most utilized hyperparameters are: Learning Rate, Mini batch size, number of Epochs, Dropout and Convolutional Layers filters’ dimensions. In table 2.2, a succinct explanation regarding the most frequent utilized hyperparameters is given.
Table 2.2: Common hyperparameters of Convolutional Neural Networks.

Learning Rate
The Learning Rate is a parameter of the optimization algorithm that controls weight update in each iteration, and its value can be fixed or be scheduled decay. A high learning rate, leads to large updates, which often results in an oscillatory tendency. On the other hand, by applying a small learning rate, a slow and impractical update process would result [34].

Mini batch size
The Mini batch size consists on the number of training samples, at which the network requires to update its parameters. By updating the model's parameters with more than one sample, a less noise optimization process and a faster convergence are allowed.

Number of Epochs
The number of Epochs refers the number of repetitions that the network goes through the entire training data set.

Dropout
Dropout is an hyperparameter responsible for avoiding the dependency between the neurons, established during the model's training. With this method, a subset of the network units are discarded, resulting in an improved generalization capability of the network.

2.3.3.3 Optimization Algorithm
In order to minimize the loss function by finding the optimal network's weight values, an optimization algorithm is required. One well established optimization algorithm is the Stochastic Gradient Descent (SGD) [43], given by the formula 2.26. For Fridman [44], a gradient represents how much the output changes, when we adjust the input a small portion. With the implementation of this algorithm, a quantitative measure of the weights variance can be assessed, in comparison with error variation.

\[
\omega_{t+1} = \omega_t - \gamma \nabla L(\omega)
\] (2.26)

Nevertheless, SGD algorithm updates are influenced in accordance to every data points, resulting in noisy convergences and lacking concepts such as Mini Batch update or Momentum.

In recent years, new and more robust optimization algorithms have been developed. However, most of them are also a variations or are based on the SGD. One algorithm that has gain relevance is Adaptive Moment Estimation (ADAM) [45]. ADAM merges the definition of Momentum with Adaptive Learning Rates, in order to converge faster. On one hand, Momentum helps the network to progress faster, by using velocity to update the weights and indicate the minima direction. On the other, by using Adaptive Learning rates, it is possible to schedule progressive decreases in the learning rate, as it converges to the local minimum. Altogether, ADAM stores the exponentially decaying average of the previous gradients, \( m_t \), (similar to momentum) and the exponentially decaying average of the previous squared gradient, \( v_t \), by using \( \beta_1 \) and \( \beta_2 \) hyperparameters to control the decay rates, as show in equations 2.27.

\[
\begin{align*}
    m_t &= \beta_1 \cdot m_{t-1} - (1 - \beta_1) \cdot g_t \\
    v_t &= \beta_2 \cdot v_{t-1} - (1 - \beta_2) \cdot g_t^2
\end{align*}
\] (2.27)
From the previous obtained expressions, it is possible to achieve the mean of the previous gradients in equation 2.28, where $g_t$ gradient and $\epsilon$ small number to avoid division by zero. As a result, this equation is used to update the weights, as shown in equation 2.29.

$$
\Delta \omega_t = -\gamma \frac{\nabla_t}{\sqrt{\nabla_t^2} + \epsilon} \ast g_t 
$$ (2.28)

$$
\omega_{t+1} = \omega_t + \Delta \omega_t 
$$ (2.29)

2.3.3.4 Network Troubleshooting

During training, several adversities can prevent the network from improving its performance. In table 2.3, common troubleshooting problems are stated in order to avoid the network from providing unsatisfactory and unreliable predictions.

<table>
<thead>
<tr>
<th>Table 2.3: Common troubleshooting problems of Convolutional Neural Networks (based on [46]).</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Class Imbalance</strong></td>
</tr>
<tr>
<td>Training Data sets may contain different amounts of data points for each label. Unequal label distribution, results in higher probabilities of predicting certain classes. As a result, the prediction score obtained is often misleading and tends to have a low performance.</td>
</tr>
<tr>
<td><strong>Overfitting</strong></td>
</tr>
<tr>
<td>Represents the correct capacity that the network has in predicting the training data set, but performs at a substandard level on unseen data (validation and testing data sets).</td>
</tr>
<tr>
<td><strong>Data Normalization</strong></td>
</tr>
<tr>
<td>In the input layer, the network rescales the data set’s values to a range comprised between 0 and 1. This allows the network, to process the information more efficiently. However, an incorrect normalization method can modify the actual information in the data set.</td>
</tr>
<tr>
<td><strong>L2 Regularization</strong></td>
</tr>
<tr>
<td>Represents the weights decay factor, which is used after the gradients are computed in the optimization algorithm.</td>
</tr>
<tr>
<td><strong>Early Stopping</strong></td>
</tr>
<tr>
<td>The tendency of the model to overfit, increases as the duration of the training extends. Therefore, the performance of the model can be monitored using an independent data set (validation set) and the training stopped if the performance degrades.</td>
</tr>
</tbody>
</table>

2.3.4 Reference Convolutional Neural Network Architectures

Over the last years, various CNN architectures have been a reference in the research state of the art. In the classification subtype, models such as AlexNet, GoogLeNet and ResNet are examples of well established networks in contests like ILSVRC. Most of the recent conducted research regarding CNN architecture improvements, has incorporated hyperparameters optimization, network pruning methods and connectivity learning [47].

One CNN that has acquired some attentions is MobileNet-V1, which was developed by Google and optimized to run on mobile devices (lightweight structure) [48]. MobileNet-V1’s main objective consists
on replacing expensive convolutions, by implementing depthwise separable convolutions. In this new
method, a separate filter is utilized for each input channel, instead having a filter with the depth of the
input, as in the conventional convolution approach, as shown in figure 2.7(a). Lastly, a 1x1 regular
convolution is performed across all channels (Point-wise convolution). After each convolution, a reLU6
activation function is inserted to prevent the excessive increment of the activations. Consequently, less
operations are computed, resulting in a faster process.

2.3.4.1 MobileNet-V2

Sandler et. al, developed MobileNet-V2 architecture, by merging the previous MobileNet-V1 upgrades
and innovative concepts [47]. The new modifications in the architecture, consisted on introducing in-
verted bottleneck blocks and residual connections.

In figure 2.8(a), a detailed representation of the bottleneck is presented, following by its respective
implementation in MobileNet-V2 architecture in figure 2.8(b). In general, each bottleneck block contains
a 1x1 Expansion layer, responsible for decompressing the information, by increasing the number of
channels of the input data. The increment in the number of channels can be given by an expansion
factor of $t$ (default value is 6). Subsequently, a Depthwise convolution is accountable for filtering the
data, while maintaining the same size. Lastly, as opposing to the expansion layer, the projection layer
decreases the number of output channels. In addition, the projection layer does not include a reLU6
activation function, due to the low dimension data (compressed data) [50].

Figure 2.7: MobileNet-V1 network main improvements.
The second improvement introduced in MobileNet-V2, relies on the insertion of Residual Connections, which helps providing a gradient flow. This problem is often visualized in deep networks, where a lose of gradient is often spotted.

The overall MobileNet-V2 architecture displayed both in figure 2.8(b) and table 2.4. In the table, each line represents a module that can be repeated n times. Additionally, a stride s is attributed to all first layers of each sequence, the remaining layers follow a stride 1. In the table, c corresponds to number of output channels and all spatial convolutions use 3x3 kernels [50].

With this in mind, an total of 17 bottleneck blocks form the core of the network, which is followed by a 1x1 convolution , an average pooling layers and, lastly, a convolutional layers and a classification layer. By analysing the respective table, when the stride is equal to two, the dimensions of the input decrease by half. Another possible observation, could be regarding the number of channels, which increases slightly through the network.

In conclusion, the MobileNet-V2 uses 55 layers and the number of parameters can vary from 1.7M to 6.9M for input size between the range of 96x96 to 224x224. In comparison, the well established AlexNet contains for 61.0M parameters for 227x227 input size, resulting in a MobileNet-V2 nearly 9 times lighter. At last, MobileNet-V2 achieved a 72.0% Top1 and 91.060% Top5 validation accuracy on 2018 ILSVRC.

<table>
<thead>
<tr>
<th>Input</th>
<th>Operator</th>
<th>t</th>
<th>c</th>
<th>n</th>
<th>s</th>
</tr>
</thead>
<tbody>
<tr>
<td>224 x 224 x 3</td>
<td>Convolution 2D</td>
<td>-</td>
<td>32</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>112 x 112 x 32</td>
<td>bottleneck</td>
<td>1</td>
<td>16</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>112 x 112 x 16</td>
<td>bottleneck</td>
<td>6</td>
<td>24</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>56 x 56 x 24</td>
<td>bottleneck</td>
<td>6</td>
<td>32</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>28 x 28 x 32</td>
<td>bottleneck</td>
<td>6</td>
<td>64</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>14 x 14 x 64</td>
<td>bottleneck</td>
<td>6</td>
<td>96</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>14 x 14 x 96</td>
<td>bottleneck</td>
<td>6</td>
<td>160</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>7 x 7 x 160</td>
<td>bottleneck</td>
<td>6</td>
<td>320</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>7 x 7 x 320</td>
<td>Convolution 2D 1 x 1</td>
<td>-</td>
<td>1280</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>7 x 7 x 1280</td>
<td>Average Pooling 7 x 7</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>1 x 1 x 1280</td>
<td>Convolution 2D 1 x 1</td>
<td>-</td>
<td>k</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Chapter 3

Requirements for Workflow’s Architecture

In this Chapter, the requirements for this dissertation and its implementation goals are established. This chapter comprises a fundamental role in formulating the state of the art in chapter 4, as it allows a more straightforward comparison between the existing papers present in this field. As a result, research gaps can be better determined in conjunction with a more clear perspective over the contributions of this dissertation.

3.1 Application in Biomedical Research

As mentioned by Topol [6], huge quantities of high resolution and large dimensions of data are recorded. This immense amount of data produced has clearly exceed the human analyzing capacities. This characteristic comes particularly evident in the Pathology field, where a patient diagnostic is established by pathologists, from analysing entire tissue sample slides. By introducing new computational algorithms and AI tools, pathologists would be able to establish a more consistent diagnosis amongst them, improve annotation accuracy and speed [6]. As a result, a considerable demand on algorithms appears to develop in order to accompany the data generation cadence, as well as to ensure a more reliable healthcare system.

With this in mind, the first requirement \((R1)\) expresses the ability of analyzing images with large dimensions. This requirement is very relevant as it imposes itself to all of the stages of the workflow implemented. This requirement can be a particularly demanding constraint, more specifically, when processing the image registration step and its respective validation.

The second requirement \((R2)\) asserts that the workflow must be able to process tissue samples with the TMA format, as previously described in section 1.2.5. This requirement assures that the workflow implemented can analyze discontinuous tissue samples throughout the tissue slide. Additionally, this qualification ensures that the workflow is also able to decompose and process the TMA image into individual image cores, as represented in the zoomed window on figure 1.3(b).
3.2 Immunofluorescent Images as Ground Truth

Deep Learning algorithms rely heavily on large quantities of data and, more importantly, on the quality of this datasets. This corresponds to one of the biggest adversities while performing DL algorithms. In Biomedical research, when analysing tissue samples, the quality of the image annotations is often guaranteed by pathologists, which represent the ground truth. Hence, this can be often a considerable obstacle when performing research, as it is largely dependent on pathologists in order to build a reliable dataset. In the same line of thought, IF images have been proved to correctly identify numerous types of cells. In other words, it is possible to use IF images as an alternative to guarantee the ground truth, when building reliable cell image datasets [12].

The third requirement (R3) states that this thesis must possess a reliable pathologist-independent method to classify the cell types in H&E images. In particular, this dissertation must utilize IF staining to insure a trustworthy dataset for the CNN training.

3.3 Ensure Image Registration on input images

Since the obtained H&E and the IF stained images are in dissimilar image modalities, the digital format of the images can contain distinct sizes. Additionally, during staining process (Figure 1.2.2.2), the tissue sample is first stained using IF markers, followed by its respective digital scanning of the IF images. This step is preceded by the removal of the previous dye (washed) and addition of the H&E staining. During this process, the tissue samples (individual TMA cores) can shift positions between the first and the second staining of the procedure. Therefore, an image alignment is required to ensure the correct matching between the information extracted from the IF images and the corresponding cell in the H&E image.

With this in mind, the fourth requirement (R4) in this dissertation establishes the concern in performing an image registration technique between the input images (IF images and the reference H&E image). This condition is utter decisive, as if the images aren’t properly aligned, the following steps of the workflow will be directly affected and their performance will not correspond to their true performance.

3.4 Dataset generation

In order to produce trustworthy data from the aligned IF and H&E images, two commitments are required. The first condition, relies on the detection of each cell type in its corresponding IF channel. This detection can be influenced by multiple factors such detected noise in the image, different staining color intensities throughout the image and deformed/missing tissue in the samples. However, the cell detection must be accurate in order to assure that the images feed to the CNN are corresponding to the cells targeted and are correctly centered, in order to assure the reliability of the dataset. The second qualification is the capacity of spatially locating the IF detected cells in the respective TMA core image stained with H&E. From the exact location in the H&E core images, the workflow must be capable of cropping a tile window
centering the intended cell. Each tile image, containing the addressed cell, must be correctly labeled and containing the appropriate CNN input format to assure the compatibility between the dataset and the project's CNN.

As previously mentioned, the fifth requirement ($R_5$) conveys the necessity of an automatic cell detection in IF channels. In general, this requirement can be described as the cell detection preformed in a reliable staining technique capable of identifying subtypes of cells accurately.

The sixth requirement ($R_6$) relies on building a dataset capable of training a CNN. This dataset consists on H&E images labeled according to each cell type, previously detected in the IF images. It also requires a significant amount of samples to be generated, under a specific input format, which varies according to the CNN utilized in the project.

### 3.5 Cell classification using Deep Learning tools

In Biomedicine, the employment of DL tools has acquired a noticeable relevance in the last years, in particular, when used for cell classification. One of the main DL techniques approaches utilized are Convolutional Neural Networks, best described in section 2.3. This is due to its noticeable characteristic of being able to extract high level abstractions from large datasets [52].

In this dissertation, the final aim consists on the classification of cell subtypes in H&E stained images, which were identified on IF images. The last requirement ($R_7$) states that the capacity to training a CNN network capable of distinguishing cell subtypes. To ensure an appropriate performance and correct evaluation of this condition, its implementation must be assured. Also, the validation and testing datasets of this CNN must also be provided from independent H&E images, in order to correctly assess the viability of the project.

### 3.6 Implementation Goals

This section is responsible for compiling all of this dissertation requirements, which are displayed in table 3.1. From the analysis of the table, a strong interconnection between this requirements is identified. In particular, the poor performance of one could directly influence the accomplishments of the subsequent requirements. Also, by gathering the essential requirements for this thesis, a more specific and clear vision is delimited for this project.

In this section, by establishing the requirements needed for this thesis, a better comparison can be made between this dissertation and the existing research in this field. This comparison is developed in a more extended manner in the state of the art section (4).
Table 3.1: Requirements for an automatic annotation architecture using Immunofluorescence stained images.

<table>
<thead>
<tr>
<th>ID</th>
<th>Description</th>
</tr>
</thead>
</table>
| R1 | **Application to images with large dimensions**  
The method should be implemented in large images in order to extract the most information possible. |
| R2 | **Application to medical images with the TMA format**  
The input images should processed under the TMA format and, more specifically, at the individual core level. |
| R3 | **Reliable pathologist-independent method**  
The method should not require the input of a Pathologist in order to assure the ground truth of the annotations provided. |
| R4 | **Performs image registration on input images**  
Due to the fact that the input images are in different modalities as well as due to the shift of the individual TMA cores during the staining procedures. An image alignment must take place to ensure the correct information transfer from the IF images to the H&E images. |
| R5 | **Automatic cell detection in immunofluorescent (IF) images**  
An algorithm must be capable of extracting the cell locations on Immunofluorescent (IF) images. |
| R6 | **Data set generation from IF channels**  
Production of a dataset, consisting of labelled images and the correct CNN training input format. |
| R7 | **Automatic cell classification on H&E images**  
Implementation of a CNN architecture capable of classifying cell subtypes in H&E stained TMA cores. This step must also be followed by a robust testing process. |
Chapter 4

State of the Art

This chapter is responsible for stating the existent research in the annotation-independent field. In particular, by using DL tools to classify tissue samples according to a reliable ground truth. The scope of this assessment can be better produced, by comparing with the actual research with the previously established requirements for this dissertation, as shown in chapter 3.

4.1 Existing annotation-independent architectures

In Biomedicine, in order to ensure a reliable ground truth, when using DL tools to make cell classification or segmentation, the state of the art frequently reflects an homogeneous pattern. In fact, the use of available public repositories or the employment of pathologists to compose the annotated tissue samples are the most recurrent approaches. As a result, significant constraints can stand out when conducting research, as it doesn’t allow a personalized study, can exhibit inconsistent tissue labelling and it’s time consuming, when performed by different pathologists [6], and can occasionally be induced into data set bias [7]. Over the last years, the pursuit for an alternative practise has increased, and the study of new imaging techniques using antibodies, such as Immunohistochemistry (IHC) and Immunofluorescent (IF) staining techniques, was introduced and has become a more recurrent approach to establish a reference ground truth [53].

The selection of the immuno-staining technique must be assessed according to the profile of each project, as each dye displays different benefits and disadvantages. The IHC staining technique has become the most widely used technique and the ground truth reference in the majority of the papers, as it is best suited for large tissue samples [54]. However, the IF technique as unveiled a set of advantages that can be very helpful when assembling datasets. While, the IHC dye uses visible light to detect only 2 types of cells and the surrounding tissues (possible noise in the detection) [9], the IF staining technique can support 4 to 5 distinct types of cells [54, 55]. In recent studies, an IF technique was capable of reaching 6 to 8 multiplex assays [55]. As result, the capacity to stain multiple types of cells is a major advantage when building a data set, providing a wider diversity of subtypes of cells and enabling a deeper study of the tissue environment.
Although this dissertation uses IF images to build the data sets, part of the existing research, to establish alternative ground truths references, was developed using IHC stained images. Additionally, some steps and procedures accomplished on the study of IHC techniques are transversal to the study of IF images, allowing a direct comparison between the two imaging techniques.

Geijs developed a project using IHC stained images to guarantee a reliable dataset [56]. The methodology implemented ensured the integration of images containing large dimensions (R1), which were decomposed and processed in smaller image tiles. This smaller image tiles were extracted from traditional whole-slide imaging (WSI) tissue format (R2). The project's ground truth was assured by using IHC stained images, which were annotated by a non-expert with a biomedical background and revised by a pathologist-resident (R3). An image registration algorithm was not applied to the multiplex IHC images, as the scanning is executed using the same machine and under the same conditions (R4). From the annotated IHC stained tissues, images were cropped centering the “hotspot” locations. This step allows the IHC dataset to be consisted of images containing an extensive perspective of the interest annotated regions (R5, R6). Finally, this project implemented multiple experiments in order to test different segmentation CNN architectures (U-net and FCNN). The main ambition of this network design was to segment tumor regions in IHC images (R7).

IHC stained tissues slides were essential in Bulten et al. work [57], as a way of providing a reliable ground truth method. More precisely, they implemented two CNNs to automate tissue annotations. The first CNN was used to map regions on the WSI images stained with IHC dyes (R1, R2). Whereas, the second CNN was applied to to carry out effectively the annotations on H&E stained images slides. In order to train the first CNN, 3493 annotations were made by three trained non-experts on 25 WSI slides (with an average of 140 annotations per slide) (R3). After the IHC interest regions were mapped, the IHC sections were registered to the H&E images, using a nonlinear image registration technique called Normalised Gradient Fields (NGF). This image registration method measures the alignment of image gradients (R4). The outputted masks generated from the IHC CNN, which were previously aligned with the H&E images, are used to signalize the interest H&E regions and decomposed those regions into 1024x1024 patches (R5). The H&E image data set (R6) generated is used as an input to train a six-level-deep U-Net to perform H&E segmentation (R7).

Nadarajan et al.’s work has established the state of the art of automated class segmentation on H&E tissues, conducting a biomarker ground truth [58]. This work used 149 breast cancer TMA cores stained for Multiplex IF biomarkers, followed by H&E staining (R1, R2). The binary label maps were obtained using a GE imaging software, the nuclei segmentation was obtained using a wavelet blob detector and the cell membranes and cytoplasm were segmented using a Frangi vesselness filter (R3). To register the multiplex IF channels with H&E image, a virtual H&E image is generated from the DAPI nuclei IF image. Then, the virtual H&E is registered with the real H&E and the resulting transformation matrix is applied to all the multiplex IF images (R4). Once the images were registered, Nadarajan et al. followed two approaches: to perform semantic segmentation on each class label or to train a CNN classifier to distinguish the same target classes. To segment images semantically, the previously generated class label maps were used as the ground truth to train the U-net in H&E images (R5, R6). The implemented
model was evaluated and displayed a dice score of 0.61 for the background class, 0.64 for nuclei score, 0.47 for cytoplasm and 0.40 for membrane (R7). The second approach also used the class label maps extracted from the IF images to generate H&E patches for each category. To extract these patches, all map regions were analyzed to assess the strongest IF signals (R5, R6). Once the best IF were selected, a modified AlexNet was trained, using the H&E patches, to distinguish three target cells, achieving a maximum testing accuracy of 70%. In particular, the classifier was able to predict 94% of the nuclei patches accurately, 79% of the cytoplasm patches and 36% membrane patches (R7).

Chang et al. constructed a pipeline to conceive a cancer cell image database [59], by using IF images to distinguish normal cells from cancerous cells. This pipeline used WSI tissues, stained with H&E and with an IF pan-cytokeratin antibody to stain cancer cells (R1, R2). To segment the nuclei in the H&E images, a set of Gabor filters as well as K-means clustering were used to differentiate and crop all individual nuclei from the different tissues (R5). Chang et al. implemented a feature-based technique to register H&E with the IF images and verify whether each cropped image contained the IF signal (ground truth). To improve the registration performance, the H&E segmented nuclei masks were utilized to match the DAPI IF channel (IF nuclei channel). Once the similar patterns were determined, a M-estimator SAmple Consensus (MSAC) algorithm was implemented to detect the corresponding matching point pairs, the scale and angle was obtained using a geometric transformation between those matching points (R4). With the images registered, a large database was generated containing the H&E stained cancerous cells and normal cells (R6). To better select the CNN architecture, a smaller dataset was built using pathologists' annotations (R3). Then, the selected architecture, was modified and trained to distinguish cancer cells from normal cells. The smaller testing dataset showed a prediction accuracy of 99.5%, while the big dataset showed a prediction accuracy of 91.3% (R7).
4.2 Comparison and Research Gaps

The comparison between different state of the art architectures is summarized in Table 4.1. By graphically representing all requirements for each study, it allows a visual representation of the current research gaps as well as this dissertation's contributions.

Table 4.1: Classification table for existing automatic annotation architectures.

<table>
<thead>
<tr>
<th>Papers</th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
<th>R4</th>
<th>R5</th>
<th>R6</th>
<th>R7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tumor segmentation in fluorescent TNBC immunohistochemical multiplex images using deep learning [56]</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Epithelium segmentation using deep learning in H&amp;E stained prostate specimens with immunohistochemistry as reference standard [57]</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>o</td>
<td>+</td>
<td>o</td>
</tr>
<tr>
<td>Automated, multi-class ground-truth labeling of H&amp;E images for Deep Learning using Multiplexed fluorescence microscopy [58]</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>o</td>
<td>o</td>
<td>+</td>
<td>o</td>
</tr>
<tr>
<td>Deep Learning based Nucleus Classification in Pancreas Histological Images [59]</td>
<td>+</td>
<td>-</td>
<td>o</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>o</td>
</tr>
</tbody>
</table>

+ Fully Fulfilled, o Partly Fulfilled, - Not Fulfilled

The first requirement (R1)- the application to large dimension images- was fulfilled by all the architectures studied, as it was required to build a considerable image dataset size capable of training various DL tools.

The application to a TMA image format (R2), was only implemented by one project. When using the TMA raw data format to build datasets, it allows the investigators to utilize tissue samples from multiple patients, in different cancer stages, extracted in different years and under different conditions. Therefore, this gives the ability of build a robust dataset capable of training a DL algorithm. In contrast, the remaining tissue format utilized by the other projects, consisted of WSI format. From the WSI papers, only one paper described using WSI from multiple patients.

One of the main relevance of this dissertation is the potential of assembling a reliable image dataset, for the study of cell subtype analysis using DL tools, without requiring the intervention of pathologists or experts (R3). By analysing table 4.1, half of the studies did not fulfill this requirement, one utilized pathologists to build a dataset in order to choose their CNN architecture and only one project was capable of fully develop an independent pathologist pipeline.

The fourth requirement (R4), consisted on performing an image registration technique to assure the alignment between the H&E image and the respective ground truth. Once provided an appropriate image alignment, the correct information could be accurately extracted from the ground truth images. In the final analysis, half of the projects executed an image registration technique to their images, one performed image registration but not at the single cell level, and one did not perform the image alignment between the images.

To extract information from the ground truth images (IHC or IF), it is required a single level cell detection for each ground truth image (R5). This condition is particularly relevant in dense cell regions,
where an inaccurate cell detection or segmentation could result in an incorrect cell labeling. Deriving out of the comparison made in table 3.1, one study did not perform this requirement; half of the projects implemented segmentation algorithms to identify all cells, but only processed the overall image masks (no single cell analysis); and, lastly, only one project developed an algorithm to segment all cells and, thereupon, locate them at the single cell level. By providing a more accurate cell detection method, a more reliable dataset is generated.

One essential condition for this thesis subsisted on the qualification to generate trustworthy H&E labelled image data sets (R6). Provided by the analysis of table 3.1, it is possible to observe an homogeneous pattern amongst almost all of the state of the art studies, where the majority of the studies fulfilled this requirement. From this pattern, only one study generated its data set consisted of IHC images (ground truth data set).

The last requirement (R7) stated necessity to training a CNN architecture to classify cell subtypes on H&E stained images, followed by a robust testing process. Most of the state of the art consists in using segmentation CNNs to map cancer regions and normal healthy tissue (binary decision). This segmentation allows a general assessment and representation of the tissue sample being studied, but limits its single cell analysis. From all studies developed, only one implemented. Although, the classification CNN is fulfilled, its binary choice training does not allow a subtype cell classification, which would allow a more in-depth study of the tumor microenvironment.

From the analysis of table 4.1, it is possible to scrutinize the various approaches developed in this field and its corresponding results. From this endeavor to formulate an reliable ground truth workflow, it is possible to acquiesce the existence of one requirement that was not fully fulfilled: the application of a CNN architecture capable of classifying cell subtypes on H&E stained images (R7). Another conclusion derived from table 4.1, subsists on convincingly fulfill all of the presented requirements in one single pipeline. By doing so, it would allow the development of more robust and trustworthy image data sets, as well as a more competent subtype cell classification CNN on H&E stained images.
Chapter 5

Methodology

This chapter describes the overall workflow used to automatically classify individual T cells, endothelial cells (blood vessels) and cancer cells in H&E stained tissues. This chapter also provides details about the different methods developed and used in each step of the workflow.

5.1 Development Environment

All the programming and analysis was developed with MATLAB software. The toolboxes required to operate the developed code are Image Processing, Parallel Computing, Statistics and Machine Learning, and Deep Learning Toolboxes.

The TMA images stained with H&E were pre-analyzed using the ImageScope software, this allowed a quick and straight assessment of the global H&E staining applied to the TMA. The same principle was similarly applied to each of the channels of the TMA images stained with Immunofluorescent antibodies, in which it was utilized the Fiji ImageJ software.

5.2 Overall Workflow

The developed workflow can be broken down into four main steps or methods, as illustrated in figure 5.1. Briefly, the first method is Image Registration. It is developed and used to align differently translated, scaled and rotated images of tissues stained differently. Here, Image Registration rigorously aligns the reference H&E images with the IF images.

The second step consists in automatic cell detection via the “nuclei channel” of the Immunofluorescent images (all sections are stained with the nucleus stain DAPI). With this in mind, the (x, y) coordinates of all detected cells are extracted from the nuclei channel (DAPI channel). Lastly, this step also extracts the color intensity value of each detected cell in every IF channel, enabling us later to threshold these values to improve the cell detection algorithm.

The third step, uses previously aligned H&E and IF images, to transfer the IF detected cells to the H&E images. This step automatically annotates the H&E images using the information provided by the IF images.
Figure 5.1: Overall Workflow Architecture
The fourth and last step of the workflow selects a CNN architecture that can be trained using the automatically annotated H&E images obtained in the previous steps. After the CNN is trained, H&E images can be annotated using only this trained CNN, without requiring IF images.

5.3 Selection of Dataset

One tissue microarrays (or TMA), obtained from US Biomax Inc, was used as a dataset. The microarray contained 208 cores from 104 breast cancer patients. Each core also contained information about the patient, such as age; sex; pathological diagnosis; grade and stage of the tumor; and the status of the ER, PR and HER2 receptors on tumor cells via a score. The TMA was stained with antibodies targeting molecules specific to four different types of cells, including CD8 for cytotoxic T cells, FOXP3 for regulatory T cells, pan-cytokeratin for cancer cells and CD31 for endothelial cells (blood vessels). In addition, all cell nuclei were stained with IF DAPI marker. After scanning these multiplex IF images, the tissue sections were washed and stained with standard H&E. Hence, it is possible to extracted six different images or channels from each TMA.

5.4 Data Pre-Processing

Before the application of the aforementioned workflow, key steps were taken to ensure that the code ran efficiently. In H&E slides, Hematoxylin (H) stains the nuclei of the cells, while Eosin (E) stains the cytoplasm and the collagen that surrounds these cells. Since, the goal is then to register the nuclei channel of the H&E channel with the nuclei in the IF DAPI channel, it was required to separate the H&E image into two separate channels to retrieve the isolated nuclei channel (Hematoxylin). This procedure, called Color Deconvolution, separated the nuclei in the images from the background, as previously mentioned in section 2.2.1.

5.4.1 Image Registration between IF and H&E stained images

Image Registration is the first main step in this workflow, shown in figure 5.1. This step aligns the Immunofluorescent (IF) images with the H&E images. The registration process utilized, consisted in first applying two automatic registration functions to all TMA tissue cores. Then, the TMA was subdivided into individual tissue cores using image processing tools and, lastly, apply another automatic image registration at the individual core level.

The automatic registration method, applied to images with large dimensions, is a method that requires heavy computational operations. However, it has the advantage of applying a translational, scaling and rotation transformation (similarity transformation in section 2.1.3) to the H&E images, which aligns them with the IF images. The technique used to generate the transformation matrix was imregcorr function from MATLAB, which estimates a geometric transformation between the images to be registered using phase correlation, better described in section 2.1.5.1. This step was followed by applying an imwarp function to perform the automatic registration technique. Although, this image registration
technique allows to align two very different types of image, it could not correctly align images at the cell level. This issue was mainly caused by the staining and washing processes, which caused the sections to shift their positions and deformation of the tissues. This issue represents the main difficulty while performing the automatic registration method, as it is very difficult to find a pattern between the images to be registered. Therefore, to correctly adjust the alignment made by the automatic registration, subdivision of the TMA into individual cores was added. An independent tissue core automatic registration enabled a more precise registration, down to the cell level.

To be able to transform the TMA into individual cores, image processing techniques were introduced with the objective of computing the centroid locations of the all the cores present in both Hematoxylin and IF images. This image processing workflow was firstly initiated by applying an edge function to the TMA image with the intention of only extracting the edges of the objects. Posteriorly, an *imdilate* function was executed, in the same TMA, resulting in the dilation of the of the edges of the objects. This step along with the implementation of an *imfill* function to fill the dilated edges of the objects, helped giving individual segmented arrays structures. Lastly, to remove the introduced noise, an *imerode* function was applied. With the arrays completely segmented, it is possible to compute the centroid locations using *regionprops* function. Once all the centroid locations are obtained, it is possible to crop the arrays from the whole-tissue section into individual independent arrays.

Finally, an automatic image registration was applied to all the individual cores in order to improve the alignment at a cell level. This step was achieved by, once again, combining the *imregcorr* function to produce the transformation matrix and *imwarp* function to execute the image registration.

In all steps of the image registration process, the Hematoxylin channel - extracted from the H&E stained TMA image- was used as the reference image. Whereas, the Immunofluorescent stained images were utilized as the images to be registered (source image) in the method, as shown in the figure 5.2. More precisely, the computation of the transformation matrices, utilized during the image registration process, were determined between the DAPI IF channel, which contains the all the cell nuclei, and the Hematoxylin channel (obtained in section 5.4), which also shares the same nuclei information. This was a very important step, has it allowed to use similar information to improve image registration performance from very different types of image. The transformation matrices calculated, to align the DAPI IF channel with the Hematoxylin channel, were applied to the remaining three Immunofluorescent channels, which have the same image dimensions and were obtained using the same sensor as the DAPI IF channel.

I note that I made sure the source images were the IF images, instead of the H&E image. By doing so, I allowed four images to be transformed instead of only transforming one H&E image. As counter intuitive as this step may resemble, this step allowed me to later be able to transfer the coordinates of the detected cells to H&E images with higher magnification/resolution, without requiring additional geometric transformations. As a result, the increment in the registration process time was compensated with higher quality annotated images to be used as input for the CNN, as well as reduction in further coordinate processing’s in order to correctly match with the H&E images with higher resolutions.
5.4.2 Cell detection in Immunofluorescent channels

After the Immunofluorescent images were aligned with the H&E image, the next stage relied on the transitioning of the information from the IF images (corresponding to the ground truth) to the H&E image. This procedure replaces the manual annotations typically completed by pathologists. In addition, this method allows to perform annotations more rapidly and at larger scales. To extract information from the IF images, two fundamental steps were implemented: apply peak finder in the DAPI IF channel (DAPI stains cell nuclei in the IF images) and measure the color intensity values of the detected cells in all the others IF channels.

The first step, consisted in separating the IF images into individual channels (each channel targets one specific cell subtype) and apply a color intensity peak finder to the DAPI IF channel. The function utilized for that purpose was `Fast2Dpeakfinder` in MATLAB, which is responsible for finding all the local maxima in a given image [31]. As a result, this function outputs an array containing the x and y coordinates of all the nuclei (peaks) in the DAPI IF channel, described in section 2.2.2. Since every cell has a nucleus, this process completes the detection of all cells present in the tissue section, regardless of the type of cell.

The second step, consists in positioning all the DAPI detected cell locations on all the remaining IF channels (CD8$^+$, FOXP3$^+$, panCK$^+$ and CD31$^+$ channels).

Thereupon, the IF signal intensity values are extracted from each cell's coordinates. In order to determine if a cell is positive for a particular IF channel (part of a cell subtype group), two conditions must be verified.

The first condition, states that a strong IF signal must be detected in that IF channel, for that specific cell's location. However, IF signals can vary from IF channel and be expressed with distinct visible features, depending on the binding characteristics of the antibody used. For instance, the DAPI and FOXP3$^+$ cells display a strong signal in the nucleus. Hence, as shown in figure 5.3, the IF signal intensities must be obtained not only from the cell (x,y) coordinates, but also from the surroundings of the coordinates. On the other hand, the CD8$^+$ and the panCK$^+$ IF signals are notorious for their binding to the surface of the target cell. As a result, in order to detect the intended cell on those IF channels, an extensive region must be utilized to accurately measure and detect the color intensity on the surface of

Figure 5.2: Image registration method.
the target cell. Lastly, the dimensions of the square shaped pixel region, for each cell detection, must consider the radius of each cell subtype. In particular, panCK+ cells would require a larger detecting region (bigger radius) than the CD8+ cells, as shown in figure 5.3.

![Image](image1.png)

**Figure 5.3:** Cell detection method including the color intensities of the surroundings.

The second condition, for the same cell location, verifies the existence of a strong signal in the DAPI channel. By detecting a strong DAPI signal, it is ensured that the detected object is considered a cell.

Altogether, for each IF channel, one cell must be positive (show a strong IF signal) for that given IF channel, as well as for the IF DAPI channel. If the images displayed a good immuno-stained quality, this step could have been performed automatically through the implementation of a threshold value for each IF signal. However, this selection process was made manually, due to the poor quality of some immuno-stained images utilized. As a result, it was implemented a manual threshold process to target the correct cells from the ground truth channels, as shown in the graph present in figure 5.4. In this graph, a small population was selected, corresponding to the strong IF FOXP3+ signal and strong IF DAPI+ signal cells. Lastly, this process was followed by a validation on the respective IF channel and its respective H&E stained image. In this particular example, 48 FOXP3+ DAPI+ previously selected cells were confirmed through visual validation.

![Image](image2.png)

**Figure 5.4:** IF signal intensities threshold method applied to cell detection.

I note that two other populations can be observed in the same graph. The population on the bottom left of the graph, corresponds to detected noise, while the other population on the bottom right of the
graph (strong IF DAPI signal and low IF cell subtype signal) represents the remaining cells present in the tissue core (negative cells). Additionally, I also note the necessary conversion to logarithmic scale on both axis, with the intention of better analyze the intended population.

To summarize, all steps aforementioned were executed with the purpose of accurately enumerate all cells present in each TMA core, locate each cell in the tissue core, and determine precisely their cell subtype. Figure 5.5 shows the successful extraction of the cell location and cell subtype in the IF-stained TMA tissue cores (Figure 5.5, A and B), followed by the corresponding cell locations in the H&E stained TMA cores (Figure 5.5C). This step of the workflow fully annotates all of the TMA cores with all the cell subtypes being studied (Figure 5.5D).

![Figure 5.5: Overall representation of the cell detection method.](image)

### 5.4.3 Convolutional Neural Network training

Training of the CNN is condensed into three main steps (Figure 5.6): selection of the CNN architecture followed by the implementation of network modifications and an iterative process to better determine the CNN hyperparameters and prevent overfitting of the data. Below, we briefly address these three different steps.

![Figure 5.6: CNN training pipeline.](image)
5.4.3.1 Selection of the network’s architecture

The objective of this Convolutional Neural Network (CNN) is to classify three different cell subtypes in breast cancer tissues, including CD8\(^+\)/FOXP3\(^+\) immune cells, panCK\(^+\) breast cancer cells, and CD31\(^+\) endothelial cells. Due to its reliability when applied to biomedical assignments, the CNN selected to accomplish this task was MobileNet-v2 [47, 60]. This CNN architecture selection was also influenced by the comparison established in figure 5.7 [61]. From the analysis of this graph, MobileNet-v2 network is able to provide an adequate accuracy compared to its relative prediction time of the network’s training. Lastly, another advantage of this network relies on the lightweight size, providing a fast and reliable training for big datasets, as well as allowing it to be used on smaller Graphics Processing Units (GPUs) (requires less processing capacity).

![Figure 5.7: Comparison between pre-trained Networks [61].](image)

To operate with this DL network, limited changes were made in order to adapt it to the intended application. The first adaptation, consisted on changing the input image size of the network. By generating a labelled image data set with this dimensions, two conditions were met: the target cell was successfully contemplated and its surrounding environment was also captured (helps improving the prediction accuracy). This reduction in the input data size, resulted in the capability of developing bigger data sets and, consequently, elaborate a more robust and accurate network.

Another adjustment required for this workflow, consisted on adjusting the CNN’s output size. As mentioned in the state of the art (in chapter 4), MobileNet-v2 is a pre-trained CNN conceived to classify more than a million images from the ImageNet database. Its default output size is set to classify 1000 different labels. Since, for this particular case, it is only required to distinguish 3 subtypes of cells, a modification in the last layers was required. Hence, the last fully connected layer and the classification layer were adjusted to this number of object categories. In sum, both the input and output size of this network were modified, as it is possible to observe in figure 5.6.

5.4.3.2 Hyperparameters optimization

The CNN’s hyperparameters play a crucial role in its training performance and, consequently, in its testing validation. Therefore, a major part of the CNN’s improvement relies on the hyperparameter’s tuning.
Different approaches can be utilized to better select this values: using grid search (exhaustive searching through a manually specified subset of the hyperparameter space), random search (exhaustive enumeration of all combinations) or using algorithms to determine the hyperparameters (Bayesian optimization, Gradient-based optimization or Evolutionary optimization) [62].

In this dissertation, the method selected to optimize the CNNs hyperparameters for training was based on the grid search approach. In particular, it consists on running a series of tests, where different hyperparameters values were experimented. By using this method, all possible scenarios were visualized and the best possible scenario, for this data set and this network, was adopted. The values utilized for this stack of training tests, were chosen in order to include a wide range of values typically selected for this network model and for the optimizer selected.

As previously mentioned in the Theoretical Background subsection (2.3.3.2), the hyperparameters selected to permute in each training were: Initial Learning rate, Mini Batch size and L2 Regularization. Additionally, the tuning of other essential hyperparameters, such as the learning rate drop factor, learning rate drop period, early stopping and maximum epochs run, were performed separately.

Finally, the optimization algorithm elected for this network's training was ADAM [45], as mentioned in section 2.3.3.3. This rather recent technique combines the use of momentum along with adaptive learning rates in order to converge rapidly.

5.4.3.3 Improvement of the network performance

The stack of tests performed to more accurately regulate the network's hyperparameters, contained two essential reinforcements to improve the CNN's accuracy: data set augmentation and re-training the network with the first layers frozen.

Data set size is a fundamental in order to provide a trustworthy and robust trained network. As a result, data set augmentation comes often as one solution, which allows to generate more data from the already existing data set. In this dissertation, two distinct data augmentations were generated from the existing data set. The first data augmentation was implemented for the first network's training and consisted on the application of small random rotations to the data set images (negative and positive rotations of 20 degrees), image reflection and small image translation in the x an y axis (negative and positive shifts of 5 pixels). The second and last data augmentation was introduced for the CNN's re-training, with the same characteristics as the first data augmentation. However, the rotation angle was increased, resulting in outputted images containing random rotations from 0 to 360 degrees.

The second approach to increase the performance of the network, consisted in freezing the weights existing in the first network's layers. By doing so, the network would only modify the weights operating in the last layers, which are known for helping in the predictions of smaller and more detailed features of the classification. Whereas, the frozen initial layers, responsible for detecting the objects shapes, edges and colours, which are kept unchanged.

I note that, for the re-training of the network with the initial layers frozen, the initial learning rate suffered an substantial decrease in its value. This can be explained by its close location to the local minimum. If the initial learning rate was too large, the CNN would not converge to its minimum value.
With this in mind, a series of tests (grid search technique) were also included in order to determine the optimal re-training CNN’s hyperparameters.

5.4.3.4 Prevention of training overfitting

DL techniques demand significant quantities of data in order to provide a robust network. In relatively small data sets, overfitting is often a common obstacle, when training a CNN. Overfitting is often described as the CNN capability to accurately predict the training data set, but performing poorly when predicting unseen data sets. As a result, several measures can be introduced to prevent or to diminish this effect.

The first implemented measure, explained in the previous section 5.4.3.3, consists in performing data augmentation on each mini batch, in order to always be feeding different data during the network’s training.

Another implemented resolution was early stopping, which consists in the ability to stop the network’s training before the model surpasses the ability to generalize from the data given, resulting in an overfit model to the training data.

The third implementation was L2 regularization, allowing the CNN to include a parameter to penalize large weights. More precisely, the inclusion of a parameter consisted of the squared norm of the weight matrix and multiplying it by the regularization parameters, all weights will diminish evenly and the cost function will be minimized.

The last implementation to prevent overfitting, consisted in shuffling the data on each training epoch and, consequently, preventing the network from optimizing the model based on the same data.

5.4.3.5 TMA core classification using Trained Convolutional Neural Network

To classify an H&E stained TMA core, using the trained CNN, a cell detection algorithm must first detect all of the cells existing in the respective TMA core. Once that step is finished, each cell is targeted and an image tile is cropped for each cell. As a result, a small dataset is originated containing all cells in that tissue core and, under that tile format, the CNN is able to classify those cells (Figure 5.6 under dataset input). This method can be applied blindly to predict unseen images or it can be verified by using the IF as a ground truth and to confirm the correct label for the cells.

To generate the testing cores, two important steps were executed: the usage of IF images to ensure the correct labelling of the testing cells and the adoption of independent patient and independent tissue cores for the testing. By previously selecting these independent TMA tissue cores from the training data set, it is established that none of the cells tested were from the same tissue section of the training data set and, therefore, display a degree of difference from the training data set. This second measure, also guarantees that none of the patients tested were used during the CNN’s training. In sum, a correct labelling of the testing data set and a proper independent validation method was ensure by these measures.
Chapter 6

Results and Discussion

This section is accountable for presenting the experimental results obtained after implemented the methodology exhibited in chapter 5. Thus, the results section is branched accordingly to the three main methods implemented: image registration results in section 6.1; cell detection results in Immunofluorescent images are presented in section 6.2 and, lastly, the Convolutional Neural Network training and its respective testing results in section 6.3. At the sections 6.1.1 and 6.3.1, evaluation metrics are specified for each technique performed, in order to more accurately estimate the performance of the workflow established. Lastly, at section 6.4, it is given an overview analysis of the proposed model and its implementation results.

6.1 Image Registration

Image registration was performed between an IF stained TMA image and the same TMA stained with H&E. Each IF image of the TMA has a size of 7800x10300 pixels and each H&E image has a size of 10150x13500x3 pixels. The multiplying factor 3 corresponds to the RGB channels. This method was implemented on one breast cancer TMA, containing 208 individual cores, with a size of 7800x10300 in case of the 5x resolution IF images and a size of 10150x13500x3 in case of the 5x resolution H&E image. As described in section 5.4.1, the 5x resolution TMA was subdivided into 208 individual cores in order to overcome tissue movements and deformations, caused by both washing and staining procedures, as well as allowing us to reach a pixel level registration. From each initial 5x TMA image, 208 individual image cores were generated, which in combination with the five initial TMA images, resulted in a total output of 1040 individual cores. Lastly, each individual core has a dimension of 660x660 (neglecting once again the third dimension in case of the H&E image).

6.1.1 Evaluation Metrics

As Cao mentioned [63], when it comes to image registration there is not a delimited standard error evaluation. However, it is possible to provide an overall performance assessment of the image registration technique performed based on certain measures. Commonly, this can be accomplished by comput-
ing the normalized cross-correlation (NCC) and the summed squared difference (SSD) between the reference image and the registered image. However, IF images and H&E images lack on common information, since the reference image and the registered image present two distinct types of staining and dissimilar color intensity levels from one image to the other. By applying these measures to these types of images, the evaluation would not demonstrate accurately the performance of the image registration technique [64]. With this in mind, the most reliable error measure for such different types of image would be to compute the distance in pixels between corresponding points in both images [65].

The error measure selected was the distance between two points or the sum of the Euclidean distances (L2 Norm), as shown by the equation 6.1. This error is obtained from estimating the deviation of any point \( p = (x, y) \) in the reference image and the corresponding point \( p' = (x', y') \) in the registered image. By applying a transformation \( f \), it is possible to represent \( p' = f(p) \) as a function of \( p \). Note that if the registration was faultless, the points \( p \) and \( p' \) would necessarily be coincident and the registration error would be zero.

\[
\text{Error} = \sqrt{(p' - f(p))^2} \quad (6.1)
\]

To obtain the Euclidean distances, each individual core was divided into 36 equal sections with a dimension of 110x110, as represented in figure 6.1. From each Hematoxylin channel, extracted from the H&E section, and its corresponding DAPI IF section, an image registration transformation matrix was computed. The geometric transformation was attained by using the same FFT based image registration method, but only to compute the translational parameters. From the transformation matrix, as shown in equation 6.2, it is attainable to measure the resolution of that specific image tile. In other words, by computing the transformation matrix of a particular section, we can estimate how much that section would have to dislocate in the x axis (\( \delta x \)) and in the y axis (\( \delta y \)) in order to completely aligned with the corresponding section. With the translational matrix 6.2, it is possible to compute the shift of that section, which corresponds to the error measure for that particular section of the image, represented in equation 6.3. By repeating the same error assessment to all the image sections, a more precise measurement is guaranteed throughout the image core instead of only one assessment per tissue core. Once the Euclidean distances of all sections are computed, the error of the entire tissue core is measured by plotting the mean of all 36 sections of that respective core, as shown in equation 6.4. In the end, the same procedure is repeated for all the 207 remaining tissue cores.

\[
T_{\delta x, \delta y} = \begin{bmatrix} 1 & 0 & \delta x \\ 0 & 1 & \delta y \\ 0 & 0 & 1 \end{bmatrix} \quad (6.2)
\]

\[
Res_{\text{section}} = \sqrt{(\delta x^2 + \delta y^2)} \quad (6.3)
\]

\[
Res_{\text{blob}} = \frac{1}{36} \sum_{i=1}^{6} \sum_{j=1}^{6} \sqrt{(\delta x_{ij}^2 + \delta y_{ij}^2)} \quad (6.4)
\]
According to Cao [63], the estimation of the error parameter includes the computation of the mean error, maximum error, median error and the variance of the error in order to correctly assess the image registration technique.

In order to have a global understanding of the image registration technique performance, it is computed the mean error. Subsequently, the maximum error in image registration is essential to provide a measure regarding the maximum deviation that the method presents. While the mean error can present an overall decent value, there can be some regions where the image registration can perform very poorly. The maximum error gives a better insight regarding the reliability of the method. Lastly, the median error provides information regarding the more recurrent error values evident in the image registration process. While the error variance presents knowledge related to the deviation of one error in regard to the mean error. Probabilistically speaking, the error variance is an error estimation given by an error range, where a given error value is more likely to be.

In figure 6.1, an exemplification of the evaluation method is displayed, in which one of the 36 sections of the H&E stained core is marked as well as it corresponding DAPI IF stained core. From the pairing of these sections, an error measurement was outputted through the computation of their respective transformation matrix. By computing the average of all sections in one tissue core, it is possible to estimate an overall error value of the core.

I note that during the error assessment of the tissue sections in each core, the sections which exhibit essentially white space and/or weak signal were neglected. This sections, in both H&E section and its corresponding IF section, lack on mutual common information and, therefore, the resulting outputted transformation matrix would not reflect accurately the image registration performance.

Figure 6.1: Image Registration subdivided into 36 sections

6.1.2 Experimental Results

In this section, the results of the image registration method, presented in section 4.2.3, are evaluated using the aforementioned metrics.
In figure 6.2, a graph is displayed showing the average registration error, in pixels, of all the cores present in the TMA. Together with the graph, a table is exhibited showing some statistical metrics and by analyzing the same, some important values can be extracted. More precisely, the mean value of all the Euclidean distances obtained is 0.64 pixels and the standard deviation is 0.31 pixels. The most recurrent error value displayed in the graph is given by the median and presents a value of 0.50 pixels. Lastly, to better assess the reliability of the method it was also computed the maximum registration error, which was 2.10 pixels and with only one occurrence.

![Figure 6.2: Image Registration results in pixels](image)

<table>
<thead>
<tr>
<th>Number of values</th>
<th>Registration error (pixels)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>0.40</td>
</tr>
<tr>
<td>25th Percentile</td>
<td>0.50</td>
</tr>
<tr>
<td>Median</td>
<td>0.60</td>
</tr>
<tr>
<td>75th Percentile</td>
<td>0.70</td>
</tr>
<tr>
<td>Maximum</td>
<td>2.10</td>
</tr>
<tr>
<td>Range</td>
<td>1.50</td>
</tr>
<tr>
<td>Mean</td>
<td>0.636</td>
</tr>
<tr>
<td>Std. Deviation</td>
<td>0.313</td>
</tr>
<tr>
<td>Std. Error of Mean</td>
<td>0.0217</td>
</tr>
</tbody>
</table>

By analyzing the registration error in pixels, we are able to have a general perception of the technique’s performance. However, to better assess and understand the meaning of this error, comparison with the size of cells is more meaningful. With this in mind, it is evident the need to convert the pixel metrics to micrometers, which is attainable by multiplying the error values by the pixel size. Knowing that the pixel size of the H&E image (reference image) in 5x magnification corresponds to 1.998 microns, it is possible to plot the corresponding error measurements in microns, as shown in the graph 6.3. Then, the quantitative evaluation of this experiment is also displayed in the table in the figure 6.3.

The mean error computed for all the tissue cores is 1.26 microns, the maximum error is 4.17 microns, the minimum error is 0.60 microns, the median error is 1.00 microns and the standard deviation is 0.62 microns.

According to Bezuidenhout et al. [66], the smallest diameter of a lymphocyte can vary from 8 to 10 microns, i.e. between 6.35 and 7.93 times larger than the registration mean error.

To determine the dimensions of panCK⁺ cells and endothelial cells, manual measurement of the cells was required. We found that the size of panCK⁺ cells varied from 16 to 27 microns (with a mean of around 21 microns), i.e. between 12.70 and 21.42 times larger than the registration mean error.

Endothelial cells are elongated; their measured length of the long axis was between 12 and 21 microns (mean = 16.0 microns) and 6 and 10 microns for the short axis (mean = 7.3 microns), i.e. between 9.52 and 16.67 times larger than the registration mean error for the long axis and between 4.76 and 7.93 times larger than the registration mean error.
In summary, cells with dimensions inferior to the image registration displacement (1.26 microns), would result in incorrect cell labelling. Since cells with dimensions smaller to the image registration mean error were not observed, the misalignment caused by the image registration error did not affect the performance of our cell detection method.

### 6.2 Cell detection in Immunofluorescent Channels

As previously explained in section 5.4.2, the implemented method can be described by first introducing a cell detection algorithm in IF images to extract the channel’s labelled cells and assure a reliable ground truth. This step is followed by the application of a threshold method to more accurately target the exact cells, as shown in figure 5.4. Another reason to implement a threshold method is related with the intentional implemented cells’ over detection, which ensures all cells in the IF channels are detected. This condition enforces the inclusion of a threshold procedure to remove the surplus of predicted cells.

Once all cells are accurately detected, each cell’s coordinates is located in its respective 20x magnification H&E stained TMA tissue core image. Then, multiple image tiles are cropped containing each individual cell. More precisely, from the cell’s coordinates extracted from the IF channels, reliable labelled H&E stained image datasets can be produced.

#### 6.2.1 Experimental Results

The cell detection algorithm was applied to all IF channels containing a strong and reliable staining signal. This means that many cores could not be used due to poor staining or tissue condition. With this restrictions in mind, in table 6.1, it is displayed the precise amount of tissue cores utilized to generate the dataset.

I also note that, not only the testing cores selected represented a considerable percentage of the number of cores containing all staining channels with acceptable condition, but also take into consider-
ation the previous established constraint of confining independent patients and independent cores, not utilized for the CNN training.

Table 6.1: Tissue and staining condition of the TMA cores.

<table>
<thead>
<tr>
<th>Immunofluorescent staining</th>
<th>DAPI</th>
<th>CD8+</th>
<th>FOXP3+</th>
<th>panCK+</th>
<th>CD31+</th>
<th>H&amp;E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nr. of stained cores</td>
<td>208</td>
<td>208</td>
<td>208</td>
<td>208</td>
<td>208</td>
<td>208</td>
</tr>
<tr>
<td>Nr. of cores with proper staining</td>
<td>208</td>
<td>35</td>
<td>133</td>
<td>77</td>
<td>90</td>
<td>201</td>
</tr>
</tbody>
</table>

All channels in good condition

Testing Dataset

<table>
<thead>
<tr>
<th>Nr. of cores</th>
<th>16</th>
</tr>
</thead>
</table>

From implementation of the cell detection algorithm to the tissues in good condition, a training, a validation and a testing datasets were generated. Both training and validation datasets can be visualized in table 6.2, according to each cell subtypes. In this table, all cell subtypes are displayed with its respective quantities. Due to the disparity between cells quantities, the dataset was balanced by random undersampling the majority of the classes, resulting in a significant reduction of the dataset (from 13255 to 7998 cells/data points). As a result, during the CNN training, the probability prediction between cells subtypes was equal, allowing a more accurate validation of the method by preventing the classifier from over-classifying the majority of the classes [67]. Once all the data was balanced, the dataset was split into training dataset (70%) and validation dataset (30%).

Table 6.2: Training and Validation datasets generated from the H&E stained TMA cores, using IF technique as ground truth.

<table>
<thead>
<tr>
<th>Generated Dataset</th>
<th>CD8+</th>
<th>FOXP3+</th>
<th>panCK+</th>
<th>CD31+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nr. of cells</td>
<td>2467</td>
<td>3571</td>
<td>4551</td>
<td>2666</td>
</tr>
<tr>
<td>Total nr. of cells</td>
<td>6038</td>
<td>4551</td>
<td>2666</td>
<td></td>
</tr>
<tr>
<td>Total nr. of data points</td>
<td>13255</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Balanced Dataset

<table>
<thead>
<tr>
<th>CD8+/ FOXP3+</th>
<th>panCK+</th>
<th>CD31+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nr. of Dataset points</td>
<td>2666</td>
<td>2666</td>
</tr>
<tr>
<td>Total nr. of data points</td>
<td>7998</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Training Dataset</th>
<th>Validation Training Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total nr. of data points</td>
<td>5598</td>
</tr>
</tbody>
</table>

As previously shown in table 6.1, a portion of the channels exhibiting a good tissue and staining quality, were reserved for the CNN testing. The designated tissues, were obtained from independent patients and independent tissue cores, to maximize the dissimilarity between the training and the testing
For the two selected testing tissue cores, the cell detection method was applied and H&E stained tissue images were annotated using the IF channels. In table 6.3, the generated testing datasets are displayed according to each cell subtype. The selected testing cores, not only displayed an elevated number of cancerous cells, as they included the presence of both T cells and endothelial cells (blood vessels).

Table 6.3: Testing dataset generated from the H&E stained TMA cores, using IF technique as ground truth.

<table>
<thead>
<tr>
<th>Tissue core #47 (Patient 24)</th>
<th>Generated Testing Dataset</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CD8+</td>
<td>FOXP3+</td>
</tr>
<tr>
<td>Nr. of cells:</td>
<td>89</td>
<td>46</td>
</tr>
<tr>
<td>Total nr. of cells</td>
<td>135</td>
<td>1812</td>
</tr>
<tr>
<td>Nr. of data points</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tissue core #208 (Patient 104)</th>
<th>Generated Testing Dataset</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CD8+</td>
<td>FOXP3+</td>
</tr>
<tr>
<td>Nr. of cells:</td>
<td>250</td>
<td>30</td>
</tr>
<tr>
<td>Total nr. of cells</td>
<td>280</td>
<td>999</td>
</tr>
<tr>
<td>Nr. of data points</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Combination of the Testing Datasets</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CD8+</td>
<td>FOXP3+</td>
</tr>
<tr>
<td>Nr. of cells:</td>
<td>339</td>
</tr>
<tr>
<td>Total nr. of cells</td>
<td>415</td>
</tr>
<tr>
<td>Nr. of data points</td>
<td>3479</td>
</tr>
</tbody>
</table>
6.3 Convolutional Neural Network

6.3.1 Evaluation metrics

To properly evaluate the performance of a classification CNN, two steps must be met: split the data into distinct datasets and perform an accuracy estimation method.

The first step, consists on dividing the groups into: training dataset, validation dataset and testing datasets. The training and validation datasets are utilized during the CNN training, one is used to update the parameters of the CNN and the other to estimate how the CNN would behave with unseen data. However, once the training is complete, a real-world performance test must be executed. This step is ensured by using the trained network to classify a testing dataset. In order to ensure a reliable testing dataset, this group should consist on 10-20% of the overall dataset [46]. Additionally, the testing dataset should consist on unused patients and tissues during training, to better describe a real-world application.

The second step, consist in using a confusion matrix as an evaluation metric. For classification CNN, this method allows to measure individually the accuracy of each class [46]. In addition, this method allows to estimate the precision and the recall of each label.

6.3.2 Experimental Results

As previously explained in the methodology section 5.4.3, a stack of 48 tests were arranged in order to identify the optimal training hyperparameters for MobileNet-v2 network. Additionally, the network's input size was modified from its default image size of 224x224x3 to 96x96x3, in order to ensure the optimal image dimensions for the cell visualization, to include the cell's environment in each image of the dataset. This step was conducted in accordance with the network advised input size [47], in order to fully match the network's filters size.

I also note that no input normalization was applied to the input data, due to the existence of large white areas in classes, such as blood vessels. Therefore, normalization methods utilizing parameters similar to the mean, would lead to incorrect normalization values (closer to 1).

6.3.2.1 Training Results

A heat map is utilized to graphically represent the CNN's first series of trainings, allowing to visualize a large volume of results, with multiple variables involved, by using a color-coded system. This heat map allows the representation of all predicted validation accuracies obtained during the CNN's training (value displayed inside each rectangle) as a function of the different hyperparameters values (values displayed in the peripheries of the heat map).

Given this points, in the analysis of the heat map in figure 6.4 , it is possible to realize that the network's prediction accuracy tends to be lower for bigger L2 Regularization values. In respect to the Initial Learning rate, the best CNN predictions were obtained for between the range of 0.005 and 0.01. Regarding the Mini Batch size hyperparameter, the highest validation accuracy values were attained
for the highest Mini batch size values. Finally, the trained network with the best performance obtained, achieved a validation accuracy of 92.44%. The hyperparameters, utilized during this training, were Initial Learning Rate of 0.01, an L2 Regularization value of 0.0001 and a Mini Batch size of 32, as better described in table 6.4.

The resulting CNN, which achieved the best training results, was utilized in a second training process. As best described in section 5.4.3.3, the initial layers were frozen and the weights in the last layers were adjusted, using a smaller learning rate in order to achieve the minima. Additionally, distinct data augmentation was performed in order to feed different data to network’s training. Lastly, another stack of 48 tests were set to achieve the optimal training hyperparameters, corresponding to the highest validation accuracy.

From the heat map generated in figure 6.5, an analysis of the hyperparameters’ behaviour is illustrated and the most favorable validation accuracy result is specified. In this analysis of the heat map, it is possible to realize that the network’s prediction accuracy also tends to be lower for bigger L2 Regularization values. In respect to the Initial Learning rate, the best CNN predictions were obtained for values of 0.0001. Regarding the Mini Batch size hyperparameter, the highest validation accuracy values were attained for the highest Mini batch size values as well. Finally, the CNN selected to classify the testing tissue cores achieved a validation accuracy of 93.20%, with an initial learning rate of 0.0001, a Mini Batch size of 32 and a L2 regularization value of 0.0001, as shown in figure 6.5.

I note that two training tests achieved the highest accuracy value of 93.20%, only differing in the L2 Regularization values (one with 0.0001 and the other with 0.01). As a result, the selected trained network corresponded to the model which achieved the highest testing accuracies on the independent datasets (L2 Regularization value of 0.0001).

![Figure 6.4: Heat map regarding the CNN's first training](image)

<table>
<thead>
<tr>
<th>Initial Learning Rate</th>
<th>L2 Regularization</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>0.001</td>
<td>0.0001</td>
</tr>
<tr>
<td>0.01</td>
<td>0.0001</td>
</tr>
<tr>
<td>0.02</td>
<td>0.0001</td>
</tr>
<tr>
<td>0.0005</td>
<td>0.0001</td>
</tr>
<tr>
<td>0.001</td>
<td>0.0001</td>
</tr>
<tr>
<td>0.02</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

55
In table 6.4, a summary containing some of the most relevant training hyperparameters is displayed for both trainings.

Table 6.4: Optimal training results and respective hyperparameters.

<table>
<thead>
<tr>
<th></th>
<th>1st Training</th>
<th>2nd Training</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning rate</td>
<td>0.01</td>
<td>0.0001</td>
</tr>
<tr>
<td>Mini batch size</td>
<td>32</td>
<td>32</td>
</tr>
<tr>
<td>Learning rate decay</td>
<td>0.4</td>
<td>0.1</td>
</tr>
<tr>
<td>Early Stopping patience (epochs)</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>L2 Regularization</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>Maximum number of epochs</td>
<td>30</td>
<td>14</td>
</tr>
<tr>
<td>Solver</td>
<td>ADAM</td>
<td>ADAM</td>
</tr>
<tr>
<td>Training duration (min)</td>
<td>98</td>
<td>35</td>
</tr>
<tr>
<td>Validation Accuracy</td>
<td>92.44</td>
<td>93.20</td>
</tr>
</tbody>
</table>

### 6.3.2.2 Testing Results

As formerly discussed in section 5.4.3.5, two TMA tissue cores were reserved for testing purposes: tissue core #047 and tissue core #208.

In testing core #047 (from patient #24), the tissue was classified using the re-trained CNN. This tissue core, contained 2088 testing cells. In particular, 89 were CD8⁺ cells, 46 were FOXP3⁺ cells, 1812 panCK⁺ and 141 CD31⁺ cells, as shown in table 6.3. From the classification of the positive IF stained cells present in that core, the predicted output labels were displayed in figure 6.6.
The quality of the predictions was assessed using confusion matrix (Figure 6.7), allowing for a direct comparison between the true class labels and the CNN’s predictions. By analysing the confusion matrix, we can observe a precision of prediction reached almost 100% for panCK$^+$ cells, while T cells and CD31$^+$ have 61.6% and 78.3%, respectively. On the other hand, the recall for panCK$^+$ was 95.7%, for T cells was around 85% and, lastly, for CD31$^+$ was 72.2%. This means that the network correctly classified 95.7%, 85% and 72.2% of the cells, respectively. In sum, the testing accuracy obtained for this testing core was 93.6%.

Also, the misclassified labels (marked inside the red squares) and the correctly classified labels (marked inside the green squares), were targeted and plotted in the test tissue core #047, as displayed in figure 6.8(a). As a result, it allowed a visual analysis, regarding their specific location. Furthermore, the
same procedure was reproduced for the misclassified labels only, allowing a more accurate perception of the unsuccessful predictions, represented in figure 6.8(b). The inaccurately predicted panCK$^+$ cells were mostly incorrectly classified as T cells, as scrutinized in figure 6.7, due to similar small dark nucleus cells exhibited on the edges of the cancer agglomerates. In relation to the inaccurately predicted T cells, the network labelled both CD31$^+$ and panCK$^+$ with small nuclei. Finally, the incorrectly predicted CD31$^+$ cells, were mostly labelled as cancer cells due to its close location to cancer agglomerates.

Figure 6.8: Comparison between the CNN correct and incorrect predicted labels on independent test core #047 (Patient 24)

The second test, tissue core #208, contained a large number of glycogen-rich clear cell carcinoma regions. In figure 6.9, labels predicted by the CNN are exhibited, cancer patterns are displayed and other important microenvironment features are presented, such as blood vessels and T cells.

Figure 6.9: CNN predicted labels on independent test core #208 (Patient 104)
The dataset label’s information are better described in table 6.3, allowing a more quantitative insight over the respective tissue core. By generating a confusion matrix for this second tissue core, the same evaluation metric is applied. Through the analysis of the confusion matrix, both T cell and panCK+ labels exhibited large precision values of 94.7% and 97.2%, respectively. Oppositely, the CD31+ only displayed a 57.1% precision value. Regarding the recall values, the label T cell achieved a value of 88.9%, panCK+ cells obtained 93.4% and, ultimately, CD31+ reached 83%. Finally, the overall testing accuracy achieved, for this testing core, was 92.0%.

As formerly executed for testing tissue #047, the correct and incorrect predicted labels are displayed on figure 6.11(a), in order to visual assess the overall performance of the trained network on tissue core #208. In addition, a more in-depth analysis was performed on the misclassified labels, as shown in figure 6.11(b). When analysing the figure, a pattern of incorrectly predicted panCK+ cells (true class) can be spotted in the surroundings of the Glycogen-rich clear cell carcinoma regions (white cancer area). The elongated shape of this edge cancer cells, similarly resembled CD31+ cells (endothelial cells), resulting in most of the incorrect predictions. Subsequently, no clear pattern was identified for T cells misclassification. Ultimately, a plausible justification for CD31+ incorrect labelling could be related with the cell agglomerates observed on both by CD31+ and panCK+ cells.

Finally, when combining the results of both testing core #047 and #208, the overall prediction accuracy of the model was 92.96%.
6.4 Requirements Fulfilment

In chapter 3, a list of requirements was presented for this dissertation, followed by a comparison between the established requirements and the current state of the art (Chapter 4). As a result, a study regarding the existent practised methodologies was produced and potential research gaps were outlined.

Both the first and second requirement are interconnected. However, the first requirement- implementation to large biomedical images - acquiesce the use of other tissue formats, such as whole tissue slides (WSI), whether the second requirement specifically defines the TMA format as the indicated method to build reliable datasets (multiple patients, multiple cancer types and in different stages). The first and second requirements (R1 and R2), were fully accomplished in this dissertation, as it the implemented workflow allowed the use of TMA tissues to create robust DL algorithms. To fulfill this requirements, the strategy implemented had to: support the TMA input image format, allow the implementation of multiple image processing techniques at the whole image level and, subsequently, convert, store and operate at the individual TMA cores.

One of the main thesis requisites was to build a Pathologist-independent annotation workflow, capable of establishing a reliable ground truth (R3). This was attained by employing IF images to signal target cell subtypes and extract the information to their respective H&E images.

In order to accurately assign the extracted information from the IF images to the intended H&E images, a cell level image registration was required (R4). This condition was achieved, by first applying one image registration at the whole image level, breaking the TMA images into individual TMA cores and, lastly, re-applying another image registration, this time, at the individual core level.

The fifth requirement (R5) states that the information extraction required an cell detection algorithm. In this dissertation, this requirement was fulfilled, by implementing an intensity peak finder on the Hematoxylin image stain (nuclei channel extracted from the H&E images, using k-means technique) of all TMA
cores.

As previously determined and further accomplished, one of the essential goals was to produce trustworthy datasets, capable of supporting the implementation of DL algorithms. In order to meet this criterion, the information transferred had to be accurately matched between the origin IF images and the destine H&E images. Additionally, a correct selection of the image dimensions had to be executed, to correctly target the intended cells, as well as allowing the visualization of the cell’s surrounding environment.

The last requirement (R7), was fulfilled into two separate steps. First step, included the implementation of a grid research algorithm to run 96 tests. As a result, the optimal trained CNN was obtained and subject to another stack of 48 tests, with its initial layers frozen. The second criterion - robust testing process- was met by testing the resulting trained CNN, on independent tissue cores and from independent patients, which were unused during the training process.

Given these points, the workflow implemented was capable of achieving all the requisites pre-established for this dissertation. More precisely, it was demonstrated that expert-independent labelled images can be acquired using immuno-stained images, in order to target cell subtypes.
Chapter 7

Conclusions and Future Work

In recent years, computing power, storage and, in general, technologies have been developing at a faster pace than ever before. In the computational field, DL algorithms have proven to be capable of classifying or segmenting images with several degrees of complexity. In the Biomedical field, pathologists and researchers have been introducing new staining techniques, to track a wider variety of cell subtypes, new antibodies have been developed and, lastly, new treatments are being introduced in the healthcare system. However, the large quantities of data generated, have clearly exceeded the human analysis limits [6]. Additionally, in the Pathology field, the introduction of new computation algorithms and AI tools, would result in more consistent diagnosis between pathologists, improve annotation accuracy and speed (generate more consistent and unbias datasets) [6]. Finally, human’s dependence on technology has never been higher [6]. As a result, the need for new solutions for this problems, as well as the demand for new algorithms hasn’t been met. This dissertation, allows the combining of this two promising fields. More precisely, this thesis objectives the production of expert-independent annotated datasets, for further DL algorithms application.

To produce trustworthy expert-independent labeled datasets, a workflow architecture needed to be implemented and several requirements needed to be established. The analysis of this architecture can be decomposed into three major phases. First, the developed pipeline must be capable of processing medical TMA images (H&E and IF stained), subprocess the TMA into individual cores and store it. Secondly, the workflow must possess a method to precisely extract information from the ground truth images. Consequently, this step should be followed by the production of a dataset. Lastly, a DL algorithm must be developed, to validate the reliability of the datasets previously generated.

The following sections are accountable for describing the formerly mentioned proposition, as well as presenting an overall evaluation of its practical implementation. In conclusion, additional future implementations are suggested and future projects are outlined.
7.1 Concept of an automatic annotation workflow

As described in this dissertation, an automatic annotation workflow would benefit both researchers and pathologists. For researchers, it would provide investigation independence, as it would not require the necessity of annotation experts, would help producing larger quantities of data, at a faster speed, and would help prevent the generation of bias datasets. For pathologists, it would help producing more consistent annotation results and improve their performance. Under those circumstances, it would provide annotation support to pathologists, from a reliable ground truth source, in this highly demanding task of producing diagnostics.

In the state of the art, chapter 4, an overview over the existent implemented projects is given. From this analysis, research gaps are delineated and possible improvements are established. More precisely, most of the conducted research is developed using IHC staining technique, providing fewer cell subtypes representation than the multiplex IF staining technique. By implementing an workflow capable of providing more cell subtypes, larger datasets could be generated and a wider variety of studies would be allowed to perform (study of the tumor infiltrating lymphocytes and tissue core microenvironment). Additionally, most of the studies developed lack on a fulfilling all the established requirements, most of the times missing either image registration alignment or only partly fulfilling a robust cell detection method or CNN validation.

With this in mind, the developed workflow is capable of processing tissue sections under the TMA format and perform an image registration technique between the H&E and IF stained cores, at the single cell resolution. In addition, the implemented framework allows the precise extraction of information from the IF stained tissue cores (ground truth images). This step is achieved, through the implementation of an intensity peak finder, followed by employment of threshold method to prevent the admittance of detected noise in the process. The result of this process, consists on the generation of cropped H&E stained image dataset, containing its corresponding cell subtypes label. The last step of the workflow, comprises the implementation of a CNN to accurately classify cell subtypes in H&E stained tissue cores and, subsequently, confirm its applicability to DL tools.

In the final analysis, this thesis’ contributions can be reflected on the implementation of a robust workflow capable of executing several operations all in one. In particular, extracting information, process it, generate image datasets and, lastly, implement a DL algorithm to classify cell subtypes in H&E stained tissues. Another contribution of this thesis, consists in producing entire datasets independently, without requiring expert intervention.

7.2 Assessment of the automatic annotation workflow

In chapter 4, relative to the state of the art, an overall representation of the existing research was outlined and research gaps were delimited. In accordance, the implemented architecture proposed the fulfillment of some of these aforementioned gaps.

In conformity with the research gaps established, requirements are also delineated in chapter 3, in
order to more accurately measure the achievements of the workflow and facilitate the comparison with the existent research. In chapter 5, a detailed explanation of the designed workflow is provided and an insight over the difficulties faced is given. As a result, it was ensured all requirements were achieved, it was verified the robustness against introduced noise in both cell detection as in the quality of the images provided.

To more accurately measure the performance of the implemented workflow, the pipeline was evaluated in three different phases. The first evaluation, corresponding to the image registration results, verified the correct alignment between the images. The results obtained, displayed an average registration error of 1.26 microns. In comparison with the utilized cell subtypes, which its smallest dimensions can be down to 6 microns (4.76 times larger than the registration error), the image registration technique could accurately target the designated cells. This registration error value was obtained, by breaking each tissue core into 36 equal sections and compute the translational error (Euclidean distance) it would require to align each section. The second evaluation, related with the cell detection results, guarantees a precise information extraction, from the multiplex IF images (ground truth). This method accurately detected all of the positive signals in the IF images, by first implementing an intensity peak finder and, after, apply a threshold method to clean the detected noise, as shown in figure 5.4. This process, results in the generation of a training dataset containing 5598 cells, a validation dataset of 2400 cells and a testing dataset of 3479 cells. The third, and last, evaluation of the workflow, is responsible for the CNN's implementation results, to ensure the reliability of the generated datasets, as well as its applicability to cell subtypes classification, using multiplex IF staining technique as ground truth.

Finally, in comparison with the studies mentioned in the state of the art, it is possible to conclude that, the implementation of a workflow containing strong image processing techniques, such as image registration, cell detection and intensity threshold methods, can provide more robust, reliable and accurate datasets and, consequently, improve the performance of DL algorithms. Another possible conclusion regarding the evaluation process refers that, due to the lack of tissue and staining quality, the number of independent testing tissue cores (from independent patients) could have been larger. However, the testing dataset size consisted of 3479 cells, corresponding to 30.3% of all the data generated and to 62% of the training dataset size utilized. Although, the number of tissue cores was reduced, the large quantities of positive IF stained cells were guaranteed during the selection of the testing tissues cores. During this process, it was ensured the inclusion of large areas of cancerous cells.

7.3 Future Work

Provided that, the previously established requirements for this dissertation were fulfilled, improvements could be applied to the implemented workflow and new studies could be introduced. Regarding the improvements to the workflow, this analysis can be also decomposed into the three main parts of the pipeline.

For the TMA processing and image alignment, improvements can be added. In particular, patient information - driver mutations, stage of cancer, breast density, and status of cell receptors - could be
used to group tissue cores, containing similar cancer types and generating multiple datasets and provide analysis of several subtypes of cancer.

For the cell detection method, an automatic threshold method could be implemented in order to prevent the manual selection of all the desired intensity values. Although, the implemented version relies on a manual selection method, this was decided due to the poor quality of tissue and staining utilized. Another implementation, consists on adding cell segmentation to study cell’s morphology and further allow the addition of cell segmentation DL algorithms.

For the DL process, several implementations can be introduced. The first implementation, consists on adding a coder-encoder network to the already established network, to provide segmentation results. Another implementation for the DL step, would use algorithms to better select the optimal network’s hyperparameters, such as Bayesian optimization. The third, and last, improvements to the network are regarding preventing overfitting of the model towards the training dataset. In particular, an interesting approach would be to implement cross-validation (generate multiple mini train test and use them to tune the network), network pruning (diminish model complexity by removing layers of the network), develop bigger training datasets and introduce a batch normalization technique capable of fitting the different classes.

Regarding new research projects opportunities, once the network is trained, a more in-depth analysis could be provided to the tissue cores. From this analysis, more information could be extracted from this tissues, which could help understanding cell’s behaviour and support cancer study in general. More precisely, assessments regarding the study of tumor infiltrating lymphocytes, the tumor microenvironment and provide additional tissue core statistics, such as cell morphology, cell density and proximity of cancerous cells to blood vessels.

All suggestions provided, aim towards the implementation of algorithms capable of providing support to pathologists, to researchers and to help promote, not only the study of breast cancer, but also to promote the development of cancer studies in general. By producing reliable cancer datasets, a powerful tool could be provided to further accelerate the study of diverse cancer research projects.


