



**HAL**  
open science

# A joint discriminative-generative approach for tumour angiogenesis assessment in computational pathology

Oumeima Laifa

► **To cite this version:**

Oumeima Laifa. A joint discriminative-generative approach for tumour angiogenesis assessment in computational pathology. Bioengineering. Sorbonne Université, 2019. English. NNT : 2019SORUS230 . tel-02954133

**HAL Id: tel-02954133**

**<https://theses.hal.science/tel-02954133>**

Submitted on 30 Sep 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Agency for  
Science, Technology  
and Research



**SORBONNE  
UNIVERSITÉ**  
CRÉATEURS DE FUTURS  
DEPUIS 1257

# THÈSE DE DOCTORAT DE SORBONNE UNIVERSITÉ

Ecole doctorale  
*Informatique, Télécommunications et Electronique (Paris)*

## Titre de la thèse

**A joint Discriminative-Generative approach for tumour  
angiogenesis assessment in Computational Pathology**

Présentée par  
**oumeima LAIFA**

Pour obtenir le grade de  
DOCTEUR de SORBONNE UNIVERSITÉ

Soutenue le 02 septembre 2019

Jury composé de :

Prof. Yannick KERGOSIEN	Université de Cergy-Pontoise	<i>président de jury</i>
Thierry DELZESCAUX, PhD, HDR	CEA DVS 12BM MIRCen	<i>examineur</i>
Prof. Maria Gloria BUENO GARCIA	Universidad de Castilla-La Mancha	<i>rapporteur</i>
A/Prof. Vlad POPOVICI, HDR	Masaryk University	<i>rapporteur</i>
A/Prof. Nicolas LOMENIE, HDR	Université Paris Descartes	<i>examineur</i>
Hwee Kuan LEE, PhD	BII, A*STAR Singapore	<i>co-directeur de thèse</i>
Prof. Daniel RACOCEANU	Sorbonne Université	<i>directeur de thèse</i>



# Contents

Acknowledgement	1
Abbreviations	3
Thesis Abstract	5
Resumé de la Thèse	7
Introduction	9
<b>I Cancer</b>	<b>13</b>
I.1 Definition . . . . .	13
I.2 Breast Cancer . . . . .	15
I.2.1 Anatomy . . . . .	15
I.2.2 Grading . . . . .	16
I.3 Colorectal cancer . . . . .	18
I.3.1 Anatomy . . . . .	18
I.3.2 Grading . . . . .	19
I.4 Tumour micro-environment . . . . .	21
<b>II Tumour angiogenesis</b>	<b>25</b>
II.1 Introduction . . . . .	25
II.2 Blood vessels mechanism in tumor . . . . .	26
II.2.1 Different stages of tumour growth . . . . .	28
a) The a-vascular stage . . . . .	28

b)	The vascular stage . . . . .	28
c)	Metastatic invasion . . . . .	28
II.3	Tumour Angiogenesis activation . . . . .	28
II.4	Efforts to prevent tumour angiogenesis . . . . .	31
II.5	Motivation . . . . .	33
<b>III</b>	<b>Immunochemistry for preclinical tumor angiogenesis assessment</b>	<b>39</b>
III.1	Confocal fluorescence microscopy versus Conventional fluorescence microscopy . . . .	40
III.2	Virtual Fluorescence microscopy Axio 2D . . . . .	42
III.3	Database preparation . . . . .	43
III.3.1	Cell culture and tumor model . . . . .	43
III.3.2	Immunohistochemistry labeling (IHC) . . . . .	44
III.3.3	Automated multi-fluorescence scan imaging using VFM Axio 2D Zeiss . . . .	45
III.4	Markov Random Field for preclinical tumor angiogenesis assessment . . . . .	49
III.4.1	Preprocessing . . . . .	49
III.4.2	Nuclei density maps . . . . .	49
III.4.3	Apoptotic endothelial cell assessment . . . . .	51
a)	The prior probability . . . . .	55
b)	The conditional probability . . . . .	56
III.4.4	Validation process . . . . .	56
III.4.5	Results . . . . .	58
III.5	Conclusion . . . . .	60
<b>IV</b>	<b>Digital pathology for tumour angiogenesis assessment</b>	<b>63</b>
IV.1	Digital / Computational Pathology . . . . .	63
IV.2	Whole slide imaging . . . . .	65
IV.3	Histopathology . . . . .	65
IV.4	Hematoxyline and Eosine staining . . . . .	67
IV.5	Computer aided diagnosis and pattern recognition in Digital pathology . . . . .	68
IV.5.1	Marked Point Process . . . . .	69
a)	Definition . . . . .	69
b)	Algorithm . . . . .	69

c)	Marked Point Process (MPP) versus Markov Random Field (MRFs)	72
d)	Sampling and optimization strategies in MPP algorithm . . . . .	72
e)	Biomedical Imaging application in histopathology and microscopy data . . . . .	73
f)	Challenges of MPP framework . . . . .	74
IV.5.2	Neural Network and deep learning . . . . .	75
a)	Convolutional Neural Network (CNN) . . . . .	75
b)	Fully Convolutional Network (FCN) . . . . .	76
IV.6	Data description . . . . .	77
IV.6.1	Challenges . . . . .	79
IV.6.2	Model design . . . . .	84
a)	Design of the contrast predictor G . . . . .	89
b)	Design of the overlap predictor . . . . .	90
c)	Design of the loss function . . . . .	92
d)	Configuration of the neural network . . . . .	94
IV.7	Experiment . . . . .	94
IV.8	Results . . . . .	96
IV.9	Proof of concept of CNN-MPP model on synthetic data-set . . . . .	98
IV.9.1	Experiment . . . . .	98
a)	Results . . . . .	102
IV.10	CNN-MPP comparison with Active contours and FCN . . . . .	102
IV.10.1	Comparison of CNN-MPP model with pure learning driven model FCN . . .	102
IV.10.2	Comparison of CNN-MPP model with pure based-knowledge driven model Active contours (AC) . . . . .	103
IV.10.3	Discussion of the CNN-MPP algorithm . . . . .	107
a)	Limits . . . . .	107
b)	Generative Adversarial Networks (GANs) versus CNN-MPP Networks	107
IV.10.4	Conclusion . . . . .	112
<b>Conclusions</b>		<b>114</b>
<b>Bibliography</b>		<b>118</b>



# Acknowledgement

Starting a Ph.D. project is a very fascinating and challenging experience in the same time. Therefore I would like to thank my supervisors, Prof Daniel RACOCEANU and Prof Hwee Kuan LEE for their advice, support and guidance throughout the period.

My Ph.D project was one of the most exciting experience in my academic career, during which time I got the opportunity to discover two different countries: France and Singapore. With different perspectives and visions about the future of Computer-aided-diagnosis to improve the accuracy of therapies in cancer field. Although, drawing up a Ph.D project is a personal experience with a lot of challenges, it would not be possible without the participation and the support of many people. Therefore, I would highlight my gratitude to the two scientific structures that have been hosting me: Laboratoire d'Imagerie Biomedicale in Paris, France (LIB) and at the second Bioinformatics Institute A\*STAR (BII) in Singapore. In Imaging Informatics Division.

I would like to thank the LIB/ITD team under the supervision of Dr. Lori BRIDAL for helping me to understand and learn about histopathology, tissue staining and helping during the experimental phase of my project in Paris.

I would like to express my gratitude to Dr Lori Bridal for all the significant administrative efforts she made.

I would like to thank the CEF (Centre d'Explorations Fonctionnelles, Cordeliers' Research Center) for their technical support and help with animals care. This work was performed, in part, on a platform of France Life Imaging network partly funded by the grant "ANR-11-INBS-0006".

Furthermore, I would like to thank Computer Vision and Pattern Discovery group (CVPD) at BII who welcomed me in their team for the two last years of the project.

I would like to thank my flatmate and colleague Weiven TEE, for his care, support and sharing reflexions.

Special thanks go to master students that I have co-supervised and supervised in Paris during their internships, Remi Lemont and Myriem abdelhak for helping in accomplishing different tasks during my project and also to all my colleagues in France and Singapore: Connie Kou Khor Li, Mustafa Omit, Frahan Akram and Sreetama Basu for their help and support.

Last but not least, I would thank my friends and my colleagues Laura Marin in Peru and Bacem ben Cheik in USA for their care, support and attention through all the up and downs of this 3 years.

Mum, Dad, Sarra, Mootez thank you for your support across oceans.

# List of Abbreviations

ANN	Artificial Neural Network
CNN	Convolutional neural network
CAD	Computer Aided Diagnosis
CASP3	Caspase-3 protein
FDA	Food and Drug administration
FN	False Negative
FP	False Positive
FPR	False Positive Rate
IDC	INvasive Ductral Carcinoma
IF	ImmunoFluorescence
IHC	ImmunoHistoChemistry
ROI	Region of Intrest
TME	Tumour Microenviroment
TN	True Negative
TP	True positive
WSI	Whole Slide Imaging
AAm	Anti-angiogenic mice
BII	Bioinformatics Institute A*STAR
CEF	Centre d'Explorations Fonctionnelles
CEUS	Contrast Enhanced Ultrasound
CNN	Convolutional Neural Network
CT26	Colorectal Tumour Cells
DAPI	4', 6-diamidino-2-phénylindole
GMM	Gaussian Mixture Model
<i>H&amp;E</i>	Hematoxylin and Eosin
ITD	Imagerie et développement de nouvelles thérapies
LIB	Laboratoire d'Imagerie Biomedicale
MRF	Markov Random Field
MPP	Marked Point Processes
MRcc	Metastatic Renal Cell Carcinoma
OCT	Optimum Cutting Temperature
OS	Overall Survival Rate
Pm	Placebo mice
PFS	Progression Free Survival
QUS	Quantitative Ultrasound
ROIs	Regions of Interest
TIFF	Tagged Image File Format
VEGF	Vascular Endothelial Growth Factor
VEGF-R	Vascular Endothelial Growth Factor Receptor
WHO	World Health Organization
VFM	virtual fluorecence microscopy Axio 2D Zeiss

# Thesis Abstract

Angiogenesis is the process through which new blood vessels are formed from pre-existing ones. During angiogenesis, tumour cells secrete growth factors that activate the proliferation and migration of endothelial cells and stimulate over production of the vascular endothelial growth factor (VEGF). The fundamental role of vascular supply in tumour growth and anti-cancer therapies makes the evaluation of angiogenesis crucial in assessing the effect of anti-angiogenic therapies as a promising anti-cancer therapy.

In this study, we establish a quantitative and qualitative panel to evaluate tumour blood vessels structures on non-invasive fluorescence images and histopathological slide across the full tumour to identify architectural features and quantitative measurements that are often associated with prediction of therapeutic response.

We develop a Markov Random Field (MRFs) and Watershed framework to segment blood vessel structures and tumour micro-environment components to assess quantitatively the effect of the anti-angiogenic drug Pazopanib on the tumour vasculature and the tumour micro-environment interaction.

The anti-angiogenesis agent Pazopanib was showing a direct effect on tumour network vasculature via the endothelial cells crossing the whole tumour. Our results show a specific relationship between apoptotic neovascularization and nucleus density in murine tumor treated by Pazopanib.

Then, qualitative evaluation of tumour blood vessels structures is performed in whole slide images, known to be very heterogeneous. We develop a discriminative-generative neural network model based on both learning driven model convolutional neural network (CNN), and rule-based knowledge model Marked Point Process (MPP) to segment blood vessels in very heterogeneous

images using very few annotated data comparing to the state of the art.

We detail the intuition and the design behind the discriminative-generative model, and we analyze its similarity with Generative Adversarial Network (GAN). Finally, we evaluate the performance of the proposed model on histopathology slide and synthetic data. The limits of this promising framework as its perspectives are shown.

# Résumé de la thèse

L'angiogenèse est le processus par lequel de nouveaux vaisseaux sanguins se forment à partir du réseaux préexistant. Au cours de l'angiogenèse tumorale, les cellules tumorales sécrètent des facteurs de croissance qui activent la prolifération et la migration des cellules endothéliales et stimulent la surproduction du facteur de croissance endothélial vasculaire (VEGF). Le rôle fondamental de l'approvisionnement vasculaire dans la croissance tumorale et le développement des thérapies anticancéreuses rend l'évaluation de l'angiogenèse tumorale, cruciale dans l'évaluation de l'effet des thérapies anti-angiogéniques, en tant que thérapie anticancéreuse prometteuse.

Dans cette étude, nous établissons un panel quantitatif et qualitatif pour évaluer les structures des vaisseaux sanguins de la tumeur sur des images de fluorescence non invasives et des images histopathologique sur toute la surface tumorale afin d'identifier les caractéristiques architecturales et les mesures quantitatives souvent associées à la réponse thérapeutique ou prédictive de celle-ci.

Nous développons un pipeline formé de Markov Random Field (MRF) et Watershed pour segmenter les vaisseaux sanguins et les composants du micro-environnement tumoral afin d'évaluer quantitativement l'effet du médicament anti-angiogénique Pazopanib sur le système vasculaire tumoral et l'interaction avec le micro-environnement de la tumeur.

Le pazopanib, agent anti-angiogénèse, a montré un effet direct sur le système vasculaire du réseau tumoral via les cellules endothéliales. Nos résultats montrent une relation spécifique entre la néovascularisation apoptotique et la densité de noyau dans une tumeur murine traitée par Pazopanib.

Une évaluation qualitative des vaisseaux sanguins de la tumeur est réalisée dans la suite de l'étude. Nous avons développé un modèle de réseau de neurone discriminant-générateur basé sur un

modele d'apprentissage : réseau de neurones convolutionnels (CNN) et un modèle de connaissance basé sur des règles Marked Point Process (MPP) permettant de segmenter les vaisseaux sanguins sur des images très hétérogènes à l'aide de très peu de données annotées.

Nous détaillons l'intuition et la conception du modèle discriminatif-génératif, sa similarité avec les Réseaux antagonistes génératifs (GAN) et nous évaluons ses performances sur des données histopathologiques et synthétiques. Les limites et les perspectives de la méthode sont présentées à la fin de notre étude.

# Introduction

Cancer is a disease characterized by abnormal cell division (proliferation) leading to a functional disorder of the human organ generating the disease. Globally, nearly 1 in 9 deaths are due to cancer. In 2018, cancer was responsible for 9.6 million deaths around the world [1]. Most of the present anti-cancer therapies are based on anti-angiogenic approach. It has been shown that solid tumours are angiogenesis-dependent following the research of Folkman, as described by Ribatti et al. [2].

Angiogenesis is the process through which new blood vessels are formed from existing ones. During angiogenesis, tumour cells secrete growth factors that activate the proliferation and migration of endothelial cells and lead to capillary morphogenesis. Hypoxic tumour regions remain despite neovascularization, which continue to stimulate over production of the vascular endothelial growth factor, VEGF [3, 4]. Folkman hypothesized that tumours would wither and shrink by blocking this factor. Therefore, the fundamental role of vascular supply in tumour growth and anti-cancer therapies make the evaluation of angiogenesis crucial in assessing the effect of anti-angiogenic therapies.

Since many years, such therapies are designed to inhibit the vascular endothelial growth factor (VEGF), then, to increase the Progression Free Survival (PFS) and the Overall Survival rate (OS). For example, pazopanib is an oral angiogenesis inhibitor targeting vascular endothelial growth factor receptor (VEGF-R). In Metastatic Renal Cell Carcinoma (mRcc), single agent Pazopanib was shown to extend the PFS in Sternberg's work when inhibiting VEGF compared with placebo [5, 6]. However, VEGF inhibition is not effective in all cancers. In Gianni's work [5], anti-angiogenic therapy showed no improvement in PFS and OS in metastatic first line Breast cancer as reported in Vasudev and al. [5]. The same result is shown in Hecht's work for metastatic first line colorectal

cancer as reported in Mbeunkui review [7]. Therefore, we need a deeper understanding of how the neo-angiogenic vasculature can be effectively targeted in tumours. The assessment of anti-angiogenic therapy effect is required not only in VEGF production, but also in vasculogenesis. Therefore, we need to develop metrics able to assess the effect of drugs on blood vessels within tumour tissue. Since the amount of information in microscopy images slides represents a huge amount of data for any computer algorithm, we first need to segment the blood vessels in these images. On the other hand, in tumour microscopy images, blood vessel segmentation remains a major challenge for current bio-medical imaging applications due to the heterogeneity of the tumour's architecture and the spatial distribution of the surrounding cells in the micro-environment. Digital Pathology allows us to investigate the complexity of such environment.

The objective of this study is to develop a pipeline of image processing and machine vision techniques, which use digitized slides of human tissue and mice cells to analyze and detect blood vessels in tumour. The proposed algorithm should be configurable depending on the type of histopathology images used. Hence, the aim is to overcome some flaws in the existing algorithms and to propose a robust pipeline which pathologists can use to diagnose any type of cancer involving angiogenesis.

Quantitative evaluation of tumour blood vessels structures with histopathological precision across the full tumour cross section, may ultimately contribute to the identification of architectural features that are often associated with or predictive of therapeutic response.

Then, qualitative evaluation of tumour blood vessels structures is performed in whole slide images which are known to be very heterogeneous. Although several papers have showed that pure rule based knowledge driven model such as stochastic Marked Point Processes (MPP) or active contours (AC) are promising tools for segmentation in histopathology images, this model fails with very heterogeneous, complex and noisy images. On the other hand, pure learning data driven models such as Convolutional Neural Networks (CNN) have been the mainstream approach in the field of semantic segmentation since good performance was achieved for similar tasks in anticancer therapies. However, CNN needs a large number of example to achieve good performances which is not always the case in histopathology research and clinical workflow. We develop, a novel Neural Network architecture combining a pure rule-based knowledge-driven (MPP) and a pure learning data driven models (CNN) to overcome heterogeneity and labelling problems. We show that our model is not only performing well in segmentation of blood vessel in very heterogeneous images

with very few data (such as whole slide images), but can also be used as flexible pre-processing to 'clean up' histopathology images for any segmentation method based on pure rule-based knowledge.

Proof of concept was done on Active contours method to enhance the ability of using our method as a pre-processing stage on heterogeneous image allowing stochastic processes such as Active Contours to perform on very heterogeneous images such as histopathology images on *H&E* (Hematoxylin and Eosin stain).

This thesis is structured in four chapters (see figure .1). In chapter I, we give a definition of cancer disease and types of cancer in which we are interested during our research. We emphasize the importance of the microenvironment interaction as a target of a lot of anti-angiogenesis therapies.

Chapter II gives a comprehensive review on the main process behind the activation of tumour growth : tumour angiogenesis. In this chapter we will emphasize the importance of assessing this process and its implication in new anti-cancer therapies. We will highlight, the motivations and background of our study.

Chapter III presents a framework proposed to assess tumour angiogenesis in preclinical histopathology tissue with fluorescence imaging. First, we will describe in detail the experiment protocol developed. We will, subsequently, propose a dedicated algorithm based on probabilistic approach : Markov Random Field to segment endothelial cells, second we developed statistical method to assess quantitatively the effect of anti-angiogenic drug Pazopanib on endothelial cells and thus the other components of the microenvironment.

Chapter IV describes the assessment of tumour angiogenesis in clinical histopathology tissue. We propose a novel method combining Neural Network and Marked Point Process for segmentation of blood vessel in *H&E* staining. Essentially, we have developed a new architecture, capable of segmenting blood vessel on very noisy images with very few annotated data for overcome two main problems: from a histopathological point of view, we overcome the annotation of a large number of data by a pathologist. From a computer science point of view, we establish a method able to segment and 'clean up' very heterogeneous images, using a combination of pure learning method and pure rule based method. Then, we propose a proof of concept by, first, testing the algorithm on synthetic data and second, by comparing with previous methods. Finally, We test our concept as a pre-processing algorithm to enhance the performance of Active contours in histopathology images.

At the end, we present our overall conclusions, by discussing our methods limits and our perspectives.

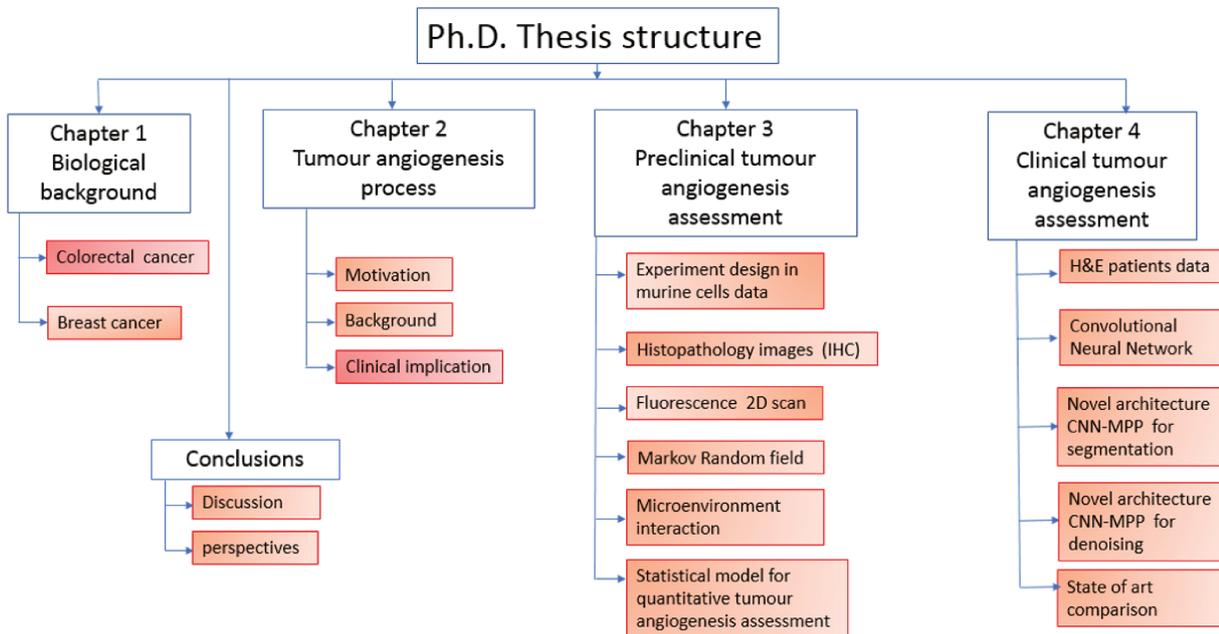


Fig. .1 *Thesis structure*

# Chapter I

## Cancer

### Abstract of the chapter

In this chapter, we, first give an overview about cancer mechanism. We detail, the importance of histopathology grading by Pathologists and its implication in patient prognosis and therapies. We focus on both breast cancer and colorectal cancer based on the alarming statistics reported recently by WHO. Finally, we emphasis the role of blood vessel as a major component of the tumour micro-environment matrix and its interaction with other tumour micro-environment components.

### I.1 Definition

The human body is made of  $3.72 \times 10^{13}$  of cells which grow by dividing process (mitosis) to form new cells, thereby replacing damaged (via apoptosis) or old ones during a human life[8]. This dividing process is activated and tightly controlled by genetic signal, coming from the genetic material of the human body: the genome.

The genome is constantly being modified by several internal factors related to metabolism and epigenetic regulation, or external factors such as environmental exogenous physical, chemical agents including ultraviolet (UV) light, ionizing radiation (IR), air pollutants and others. Besides,

biological exogenous factors such as chemotherapeutic drugs and lifestyle (tobacco, stress and etc.) can play a major role in this modification process [9].

Therefore, cells have evolved multiple DNA repair pathways to preserve the genome integrity when damage arises, as described in details in Iyama et. al work reviewing the DNA repair mechanism in actively dividing cells [10]. However, prolonged DNA damage can induce mutagenesis. Mutagenesis is a process by which the genome of an organism is changed by mutations [11]. Mutagenesis includes substitution base pairs, insertion/deletions and chromosomal rearrangements, by causing genomes instability, an essential step in unregulated cell growth and, therefore, the development of cancer. Indeed, mutations in some genes allow cells to bypass the tightly regulated checkpoint in mitosis (cell division), causing cells to grow uncontrollably [12]. We call these cells cancer/tumour cells, which are different from normal cells because of their invasive behaviour. For example, tumour cells are able to bypass the programmed apoptotic pathways (programmed cell death). The apoptosis is crucial in maintaining a proper cells population as in removing DNA damaged cells [13, 14, 15].

Benign tumours do not spread via metastasis, while other malignant tumours spread into surrounding tissues through the lymphatic and blood system and form secondary tumour in other organ [16]. This spreading behaviour is controlled by very complex processes termed Epithelial–mesenchymal transition (EMT) and angiogenesis which are 2 crucial tumor progressions. While EMT is necessary for tumour cells progression, angiogenesis is the important part of the vascular phase in tumour growth and metastasis. The main biomedical motivation behind our research is the assessment of the tumour angiogenesis process.

Essential, cancer can originate from most, if not all tissues that are actively undergoing mitosis. According to College of American Pathologists (CAP), there are more than 100 known types of cancer in very diverse locations such as Breast DCIS, Adrenal Gland, Trophoblastic Tumors, Uterine Sarcoma, Squamous Cell Carcinoma and many other types in different location in human body [17].

According to the World Health Organization (WHO), cancer is the second leading cause of death in the world after cardiovascular diseases. In 2018, 18.1 million new cases have been recorded, and cancer was responsible for 9.6 million deaths. Globally, nearly one in eight men and one in eleven women deaths are due to cancer [1].

Although this disease exists in all human populations, the International Agency for Research and Cancer (IARC) reports a geographical diversity in its occurrence. In 2018, 57.3% of global death caused by cancers occurred in Asia and 23.4% in Europe [1].

In Singapore, in 2018, 35 persons were diagnosed, everyday, with cancer [18]. The most common cancers for men is colorectal Cancer (17.2%) (1 from 6 cancer diagnosis), lung cancer (14.8%) and prostate cancer (13.00%). The most common cancers for women are breast cancer (29.1%) and colorectal cancer (14.8%).

In France, according to the National Institute of Cancer (Institut National du Cancer - INC), only for 2018, we estimate to 382000 the number of new incidences recorded and to 157400 the number of deaths. The most common cancer for men are lung cancer with 27 500 cases and colorectal Cancer with 21 500 cases. For women, breast cancer is ranked number 1 (29.1%) then colorectal cancer (12.00%) [19].

Given the alarming statistics in breast and colorectal cancers, our study is focusing on these ones. Therefore, we start by giving a brief introductory overview of breast and colon anatomy of the corresponding cancers.

## I.2 Breast Cancer

### I.2.1 Anatomy

The breast is composed of glandular ducts, lobules, connective tissue, and adipose tissue (fat). Lobules are the milk producing glands and they are connected to nipple by lactiferous ducts conducting milk. These three main structures of the breast are protected by fatty and connective tissues.

There are many types of breast cancer. Most breast cancers arise from glandular ducts and lobules [20]. Most breast cancers are carcinomas. Carcinomas are tumors that start in the epithelial cells that line organs and tissues throughout the body. An even more specific term is often used. For example, most breast cancers are a type of carcinoma called adenocarcinoma, which starts in cells that make up glands (glandular tissue). Breast adenocarcinomas start in the ducts (the milk ducts) - ductal carcinomas - or the lobules (milk-producing glands) - lobular carcinomas.

Invasive lobular carcinoma spreads from the lobules to the ducts, on the other hand, invasive ductal carcinoma spreads outside ducts to spread through the lymphatic system [21, 22]. In our research, the available dataset for breast cancer tumour angiogenesis assessment concerns breast carcinoma.

### 1.2.2 Grading

All breast cancer grades do not have the same aggressivity. The anatomopathological exam of a tumour sample is the one that allows to assess the exact type of cancer and define its grade.

Grading involves labeling of tissue regions into groups based on their morphologies. The specimen comes from biopsy, an invasive medical or surgical procedure performed by healthcare professional, and involving human cell tissue extraction to define the stage of cancer. Subsequently, the pathologists characterize the breast cancer based on specific features which are architectural and cytological features, nuclear grade and number of mitosis [23] to answer the following questions:

- How aggressive is the breast cancer: a low grade breast cancer tends to be less aggressive and the spreading potential is less, however a high grade breast cancer tends to be very aggressive with high chance of mitosis.
- How far the breast cancer has spread.
- How large is the tumour size.

All these elements are essential in patients prognosis and the chance recurrence. Therefore, grading is a crucial part in breast cancer diagnosis, as it guides the Oncologist to select the most appropriate anti-cancer therapeutic strategy for the patient. However, grading requires a labeling/annotation procedure as the Pathologist needs to extract precise features to properly evaluate it. Indeed, in order to define the grade of cancer, Pathologists need to annotate the histopathology images by hand to identify and characterize some features, crucial supporting their decision. The complexity of the anatomy and the interaction between its different components make this task very tedious, time consuming and sensitive to subjective human judgments. Although many grading systems have been used in many cancers to reduce the inter-variability between different pathologists decisions, those systems are still prone to human annotation, a subjective tedious task

that computer-aided diagnosis could support.

In accordance with hospital protocol, pathologists may use different grading systems. In breast cancer, the most known is the Nottingham Histologic Score system, or Elston-Ellis modification of the Scarff-Bloom-Richardson [24]. Following this grading system, pathologists have to assess the glandular differentiation, nuclei pleomorphism and mitotic index and give a score for each element from 1 to 3. The final grade is the sum of the three scores:

- In grade 1: tumour has a score of 3-5.
- In grade 2: tumour has score of 6-7.
- In grade 3: tumour has score of 8-9.

1. Glandular/Tubular Differentiation: Describes the capacity of the tumour to recreate a normal glands as shown in figure I.1 - Score 1: > 75% of tumour area are forming glandular structure.  
- Score 2: 10% to 75% of tumour area are forming glandular structure. - Score 3: < 10% of tumour area are forming glandular structure.
2. Nuclear pleomorphism: Describes shape, size and color variability of cell nuclei. - Score 1: little increase of epithelial cells nuclei comparing to normal ones, little variation in size, uniform chromatin - Score 2: moderate variability in size and shape, visible nucleoli. - Score 3: marked variation in size and shape with occasional very large forms.
3. Mitotic proliferation index: is a score indicating how much the tumour cells are dividing(mitosis). It depends on the field diameter of the microscope used by the pathologist when doing the biopsy. So the Pathologist counts how many mitotic cells are located in the most dense 10 high power adjacent fields. - Score 1: less than or equal to 7 mitosis events of mitotic cells - Score 2: 8-14 mitosis - Score 3: equal or greater then 15 mitosis

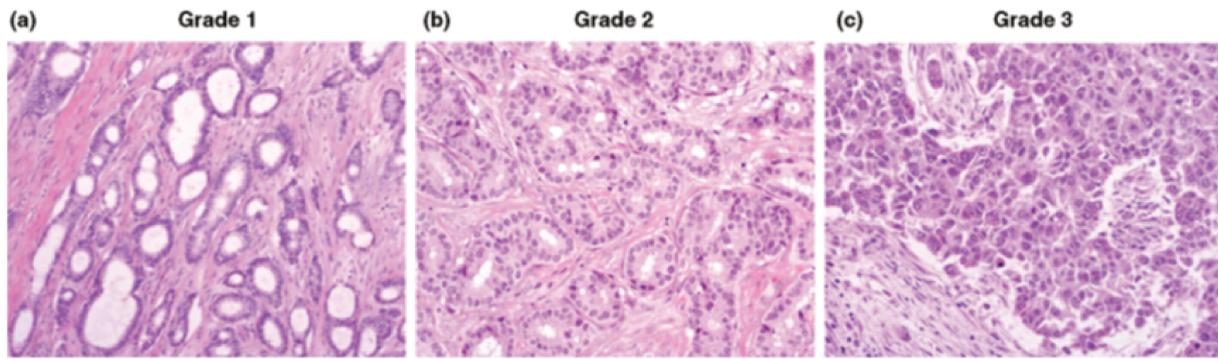


Fig. I.1 (a) A grade 1 of ductal carcinoma that demonstrates a tubule formation ( $> 75\%$ ), a mild degree of nuclear pleomorphism, and low mitotic count. (b) grade 2 (c) grade 3 ductal carcinoma with a marked degree of nuclear pleomorphism and frequent mitoses and no tubule formation ( $< 10\%$ )[25].

## I.3 Colorectal cancer

### I.3.1 Anatomy

Colon, also called large intestine, is a huge muscular tube of 6 feet (1.8m) for the excretion of solid waste. The colon plays a major part in the final digestion stage. It is divided into four parts: the ascending colon, the transverse colon, the descending colon and the sigmoid colon as shown in figure I.2. The function of the large intestine is removal of undigested food and waste after the nutrients are removed from it to be absorbed by the epithelial cells lining the internal wall of the colon.

Colon cancer also know as Colorectal cancer, is one of the most common causes of cancer-associated death. It is the second-leading cause of cancer deaths in both men and women [26]. Colorectal adenocarcinoma makes up to 95% of all colorectal cancer types. It develops in the lining of the colon or rectum typically in the inner lining and then spread to other layers. There are other less frequent types:

- Gastrointestinal carcinoid tumors: They account for 1% of all colorectal cancers.
- Primary colorectal lymphomas develops in the lymphatic system.
- Gastrointestinal stromal tumors develops from the interstitial cells of Cajal.
- Leiomyosarcomas which represents 0.1 % of all colorectal cases

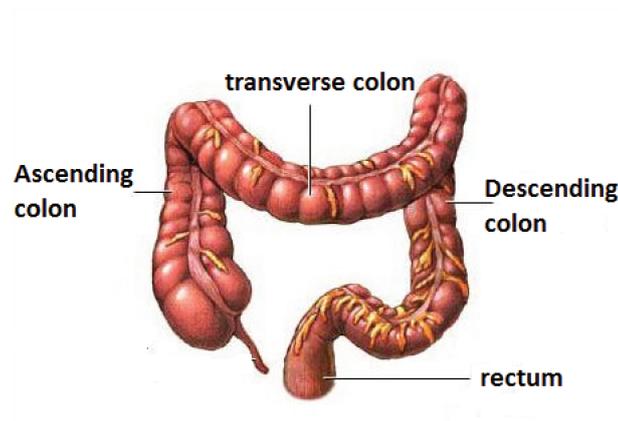


Fig. I.2 *The colon anatomy.*

In our research, we investigate Colorectal cancer on murine cells (no clinical dataset is available).

### **I.3.2 Grading**

Colorectal carcinoma grading is usually based on the degree of tumor differentiation. This grading assessment is in turn based on the cancer cells and the differences in the morphology and their behaviour comparing to normal cells.

Colorectal carcinoma is often described as low grade or high grade. Pathologists attribute a low grade to tumour that is very similar to normal cells in term of morphology. Low grade cancer cells tend to grow slowly and they rarely spread. However in high grade tumour, the cancer cells look more abnormal. Higher grade cancer cells tend to grow more quickly and are more likely to spread.

The Pathologist also gives a grade from 1 to 4 to the colorectal cancer. The lowest the number, the lowest the grade. Grades 1 to 3 apply only to adenocarcinomas whereas grade 4 is only used for squamous cell carcinomas.

Grading is a very important task in the prognosis of patients as shown in the table below:

Colorectal cancer stages	5 years age standardized Relative Survival (%)
1	98.9
2	84.6
3	71.0
4	10.3

Table I.1: 5 years Age-standardized Relative Survival for males. This ratio reflects the chances of surviving assuming that cancer is the only possible cause of death [27].

Colorectal cancer stages	5 years age standardized Relative Survival (%)
1	94.1
2	89.2
3	64.5
4	9.9

Table I.2: 5 years Age-standardized Relative Survival for females. This ratio reflects the chances of surviving assuming that cancer is the only possible cause of death

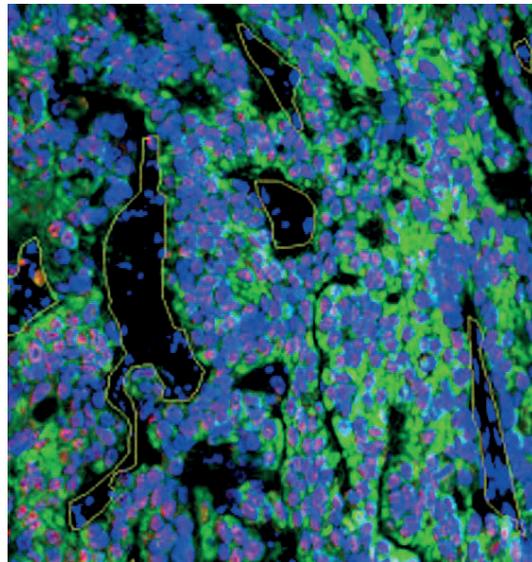


Fig. I.3 Example of colorectal adenocarcinoma in a fluorescence image of immunostained specimen. The blue stain shows the nuclei, the green represents the endothelial cells, pink color represents immune cells and the marked region of interest with green are the lumen of the colon gland. Our pre-clinical data is similar to this type of images as we are using a fluorescence imaging

## I.4 Tumour micro-environment

Tumors often become more aggressive in their behaviour and more "malignant" in their characteristics during their life history, although the time course may be quite variable from one patient to another in term of type of cancer and mechanisms of tumor progression.

In [28], Nowell stated the different factors playing major roles in tumour growth and progression. Through this study, Nowel emphasis two main factors:

- Genetic instability of tumor cells: there is evidence that most neoplastic cells (tumour cells) are more genetically unstable than comparable normal cells and this may be a major factor contributing to the phenomenon of tumour growth.
- Host Factors: One must also consider those factors in the host environment including the immune surveillance, specific tumor growth factors as tumour growth regulators and interaction with tumour microenvironment components.

Indeed, it becomes well known that tumour growth is determined not only by the genome but also by their interactions with the different components of the microenvironment which modulate the malignancy of tumor growth [29]. Tumour inhabits a cellular environment composed of different components such as immune cells, collagen, adipose tissue (fat), blood network, lymphatic network and many other cells types as shown in figure I.4. Together, these form the tumour microenvironment (TME).

The interaction between tumour and TME impacts the tumour growth and the progression of the disease. For this reason, important research has been carried out in order to improve knowledge about the tumour behaviour according to its interaction with the microenvironment components in order to treat tumour more effectively.

This knowledge forms the major thrust in the progress of cancer therapies in the last decade and in recent transitional oncology that has been giving popularity which aims to translate laboratory research into new anticancer therapies based on targeted anticancer therapy (TAT). Contrary to conventional surgery, chemotherapy, and radiation therapy, target anticancer therapy refers to systemic administration of drugs with particular mechanisms that specifically act on well-defined

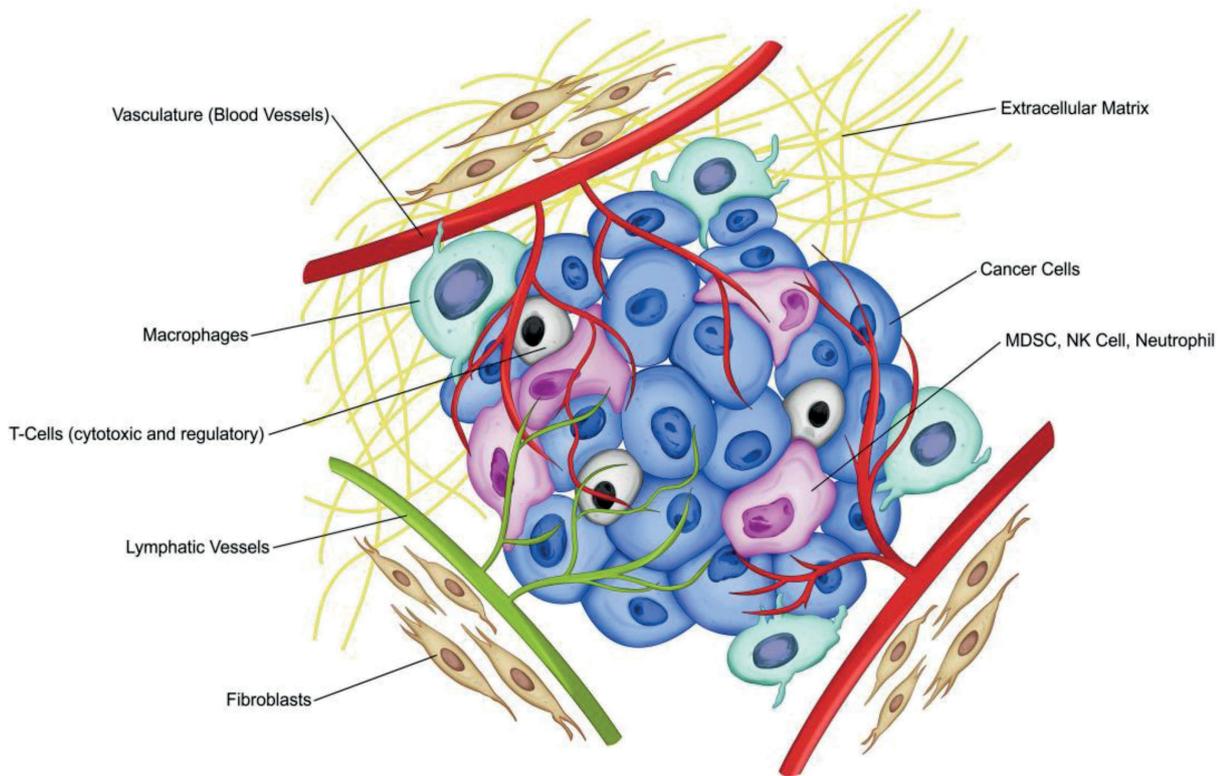


Fig. I.4 *Illustration of the complexity of the tumor micro-environment (TME) [30]*

targets or biologic pathways that, when activated or inactivated, may cause regression or destruction of the tumour growth.

In figure I.5 a brief review listing currently known targeted anticancer therapies summarized by Li et.al. [31]. Immunotherapy and anti-angiogenesis are two examples of TAT.

The tumour microenvironment affects the outcome of these two therapies substantially. In [32], Smyth et. al. emphasize that the tumour microenvironment can help determine which immune suppressive pathways become activated to restrain anti-tumour immunity to improve cancer immunotherapy which marks a turning point in cancer therapies in the last 10 years [33]. The inflammatory microenvironment is known to have a great impact on the tumour behaviour [34]. Significant correlation have been shown to exist between clinical outcome and immune cell presence, relative abundance as well as spacial proximity of immune cells to invasive cancer cells. Same to chemotherapy, Albini and et. al. [35] demonstrated that the tumour microenvironment can be used as a target to improve chemo-prevention and reduce side effects after chemo-therapies.

However, anti-angiogenesis TAT is controlled by the main component in the human body and

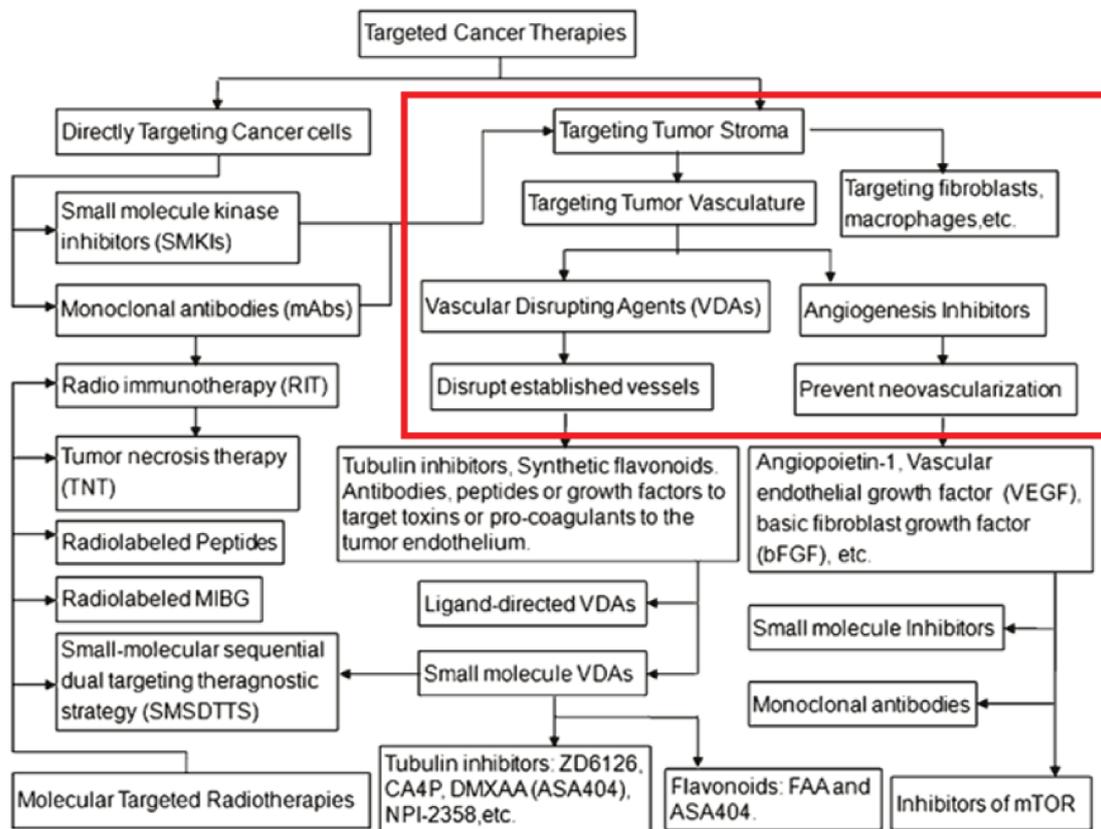


Fig. I.5 A scheme listing currently known targeted anticancer therapies (TATs). The red box emphasis the anti-angiogenic therapies in which we are interested[31].

a key component in the tumour microenvironment: blood vessel network. It has been known for more than a century that tumours have the ability to induce their own blood supply and that the tumour vasculature is actually more extensive to that of normal tissues. In fact, tumours induce their neovasculature by driving the angiogenesis process. Endothelial cells lining the internal wall of blood vessels are then playing a central role in the development of the tumour as they form tumor-associated blood vessels [36].

The angiogenesis process is going to be established in details in chapter II: tumour angiogenesis, where we emphasize the importance to assess this complex process in order to develop more accurate TAT therapies by focusing on Targeting Tumour Vasculature.



# Chapter II

## Tumour angiogenesis

### Abstract of the chapter

In this chapter, we detail the mechanism of tumour angiogenesis and its role in each phase of tumour growth. we describe the 'switch-on' process as the passage from the dormant to proliferative tumour. We present anti-angiogenic therapies as a promising anti-cancer therapy. Finally, we explain our motivations and background according to the angiogenesis assessment.

### II.1 Introduction

A tumour is a cluster of mutated cells that are no longer sensitive to the regulation of cell division. Therefore, each cell divides anarchically following the secretion of its own growth signals. During tumour growth, an interaction starts between the tumor and its micro-environment. Over the past decade, the discovery of tumor cells located near blood vessels has highlighted the existence of functional interactions between blood vessels and tumor cells.

## II.2 Blood vessels mechanism in tumor

Like a normal tissue, the growth of tumor cells relies not only on the supply of oxygen and nutrients but also the evacuation of metabolic waste and carbon dioxide. These needs are provided by the blood vessels as one of the major components of the microenvironment. Cells require oxygen and nutrients for their survival. Diffusion of oxygen and nutrients happens within 100-200 micrometres of the blood vessels. Blood vessels ensure that blood transport conveys necessary nutrients and eliminates cells' waste in all human organs. Cells are, thus, located 100-200 micrometres from blood vessels. If the cells aggregation goes beyond this diameter, new blood vessels are created from preexisting ones. This makes blood vessels a key element in the development of normal or tumour cells in multi-cell organs [37]. The process of creating new blood vessels from preexisting ones is called angiogenesis, vasculogenesis or neovascularization.

The wall of a blood vessel consists of 3 layers (figure II.1). These layers are well identified and are defined: first Tunica adventitia, second Tunica media and third Tunica intima [38]. The adventitia, which represents the outer layer, consists of connective tissue, fibroblasts, macrophages, nerve endings and vasa vasorum (for the supply of oxygen and nutrients). The media is the most central layer of the blood vessel and is the thickest layer. The innermost layer is the intima. The intima consists of an extra-cellular base of collagen, fibroblast and a monolayer of endothelial cells.

In addition to their exchange function, blood vessels actively contribute to tissue regeneration and growth processes. The growth of solid tumors beyond 1 to 2 mm is dependent on the development of a vascular network dedicated to them. Many studies [39, 40, 41, 42, 43] have shown that the interaction between blood vessels and tumor growth are due to endothelial cells. In fact, the entire process of angiogenesis in tumour has been attributed to the endothelial cells. In absence of endothelial cells, necrosis or apoptosis of the tumor cells has been observed. Thus, the growth of solid tumors is dependent on tumor neovascularization formed via angiogenesis process. We have found 3 stages in the tumour growth which are highly dependent on the blood-vessel network as shown in figure II.2: 1) a-vascular stage, 2) the vascular stage and 3) the metastatic invasion.

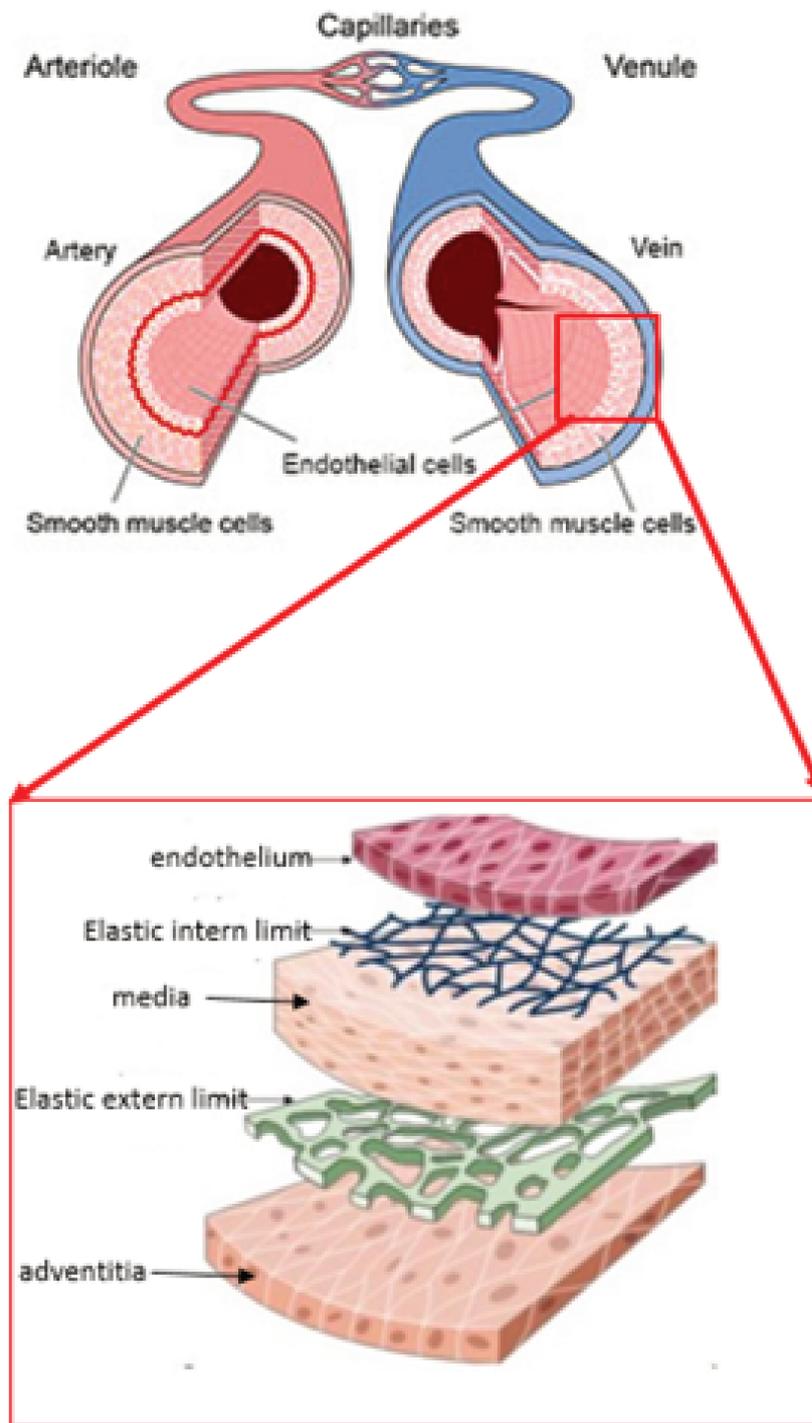


Fig. II.1 3D cut of different layers of the arterial wall, endothelium is the inner layer. It consists of endothelial cells responsible for tumor neovascularization.

### **II.2.1 Different stages of tumour growth**

#### **a) The a-vascular stage**

At this stage, the tumor only measures up to a few millimeters. It is therefore content with the contributions of nutrients and oxygen provided by the preexisting blood networks in its immediate environment. Proliferative cells that are located on the surface of the tumor will not lack oxygen and nutrients. They will, therefore, divide permanently. However, the proliferative cells that are located more deeply in the tumor and are called quiescent cells can no longer divide due to lack of oxygen and nutrients. From a certain size, the tumor becomes large. Proliferative cells are big consumers of oxygen. They do not leave enough oxygen and nutrients for the quiescent cells. If the tumor becomes larger, apoptotic areas appear in the heart of the tumor because of hypoxia. The limit of oxygen and nutrient supply at the a-vascular stage may limit the size of the tumor.

#### **b) The vascular stage**

When the tumor consumes almost all the nutrients and oxygen in its environment, it will tend to find other solutions to increase this intake in order to continue its growth. Therefore, the tumor will encourage the body to create new blood vessels that will directly irrigate the tumor by the angiogenesis process.

#### **c) Metastatic invasion**

Blocks of cells that are in contact with neoformed vessels can become detached from the tumor and migrate through the lymphatic or blood network to other regions of the body to attach. By settling on other regions, these clusters multiply in their turn anarchically and form a new tumor. These new tumors are called metastases.

## **II.3 Tumour Angiogenesis activation**

The passage of the tumor from a dormancy stage to an activation stage is conditioned by the activation of angiogenesis. The absence of this process causes inactivity of the tumor as illustrated

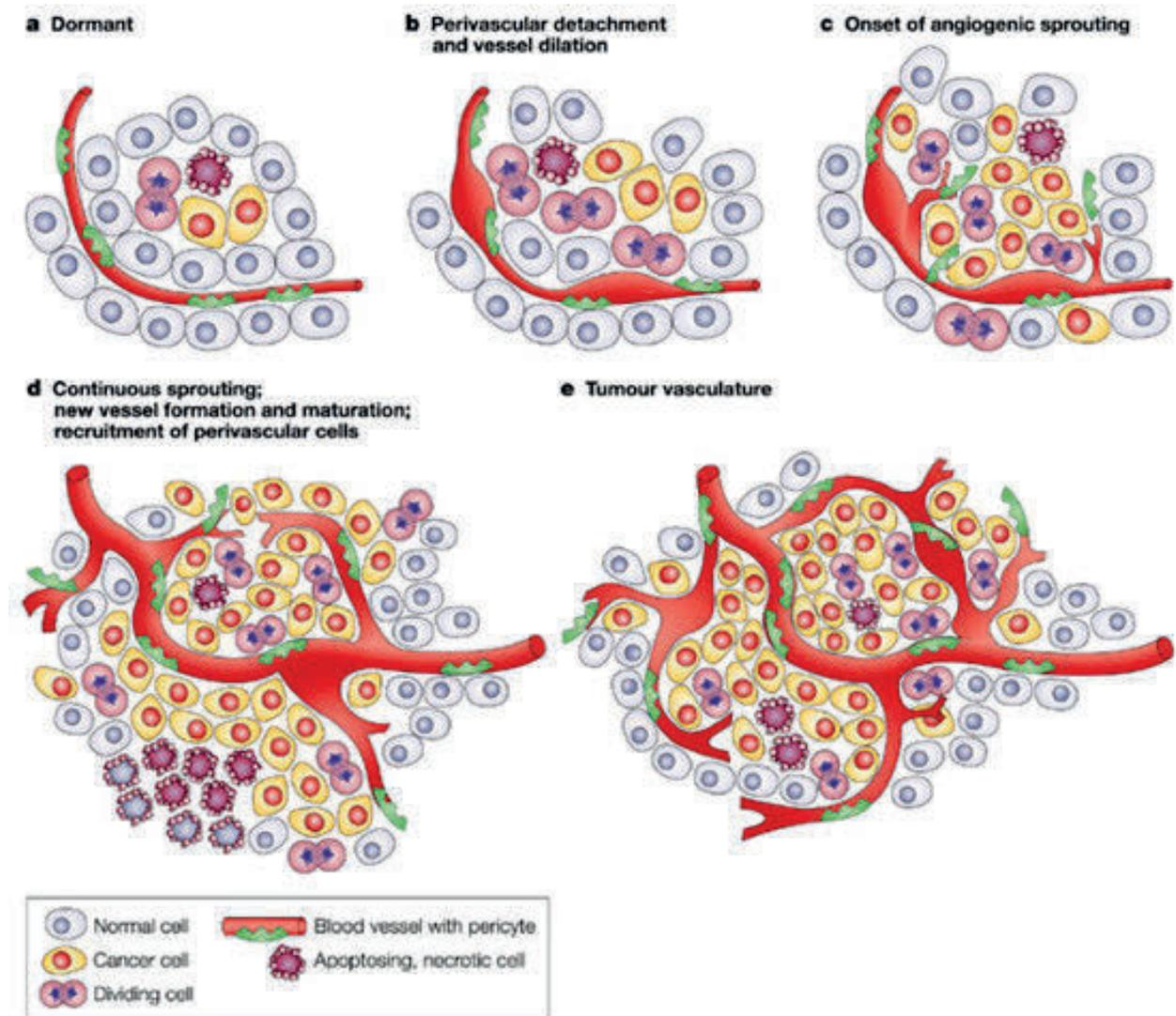


Fig. II.2 The stages of tumor angiogenesis [44].

in figure II.3. Therefore, the tumor remains benign. By definition angiogenesis is the formation of new blood vessels from preexisting vessels [38]. This process includes the follows behaviour:

- a) Continuous activation of the proliferation signals.
- b) Insensitivity to processes inhibiting cell growth.
- c) Resistance to cell death.
- d) Activation of the processes of invasion and formation of metastases.

Angiogenesis of tumors is regulated by local balance in the activities of pro-angiogenic and anti-angiogenic factors. Each angiogenic activating factor has its antiangiogenic inhibiting factors

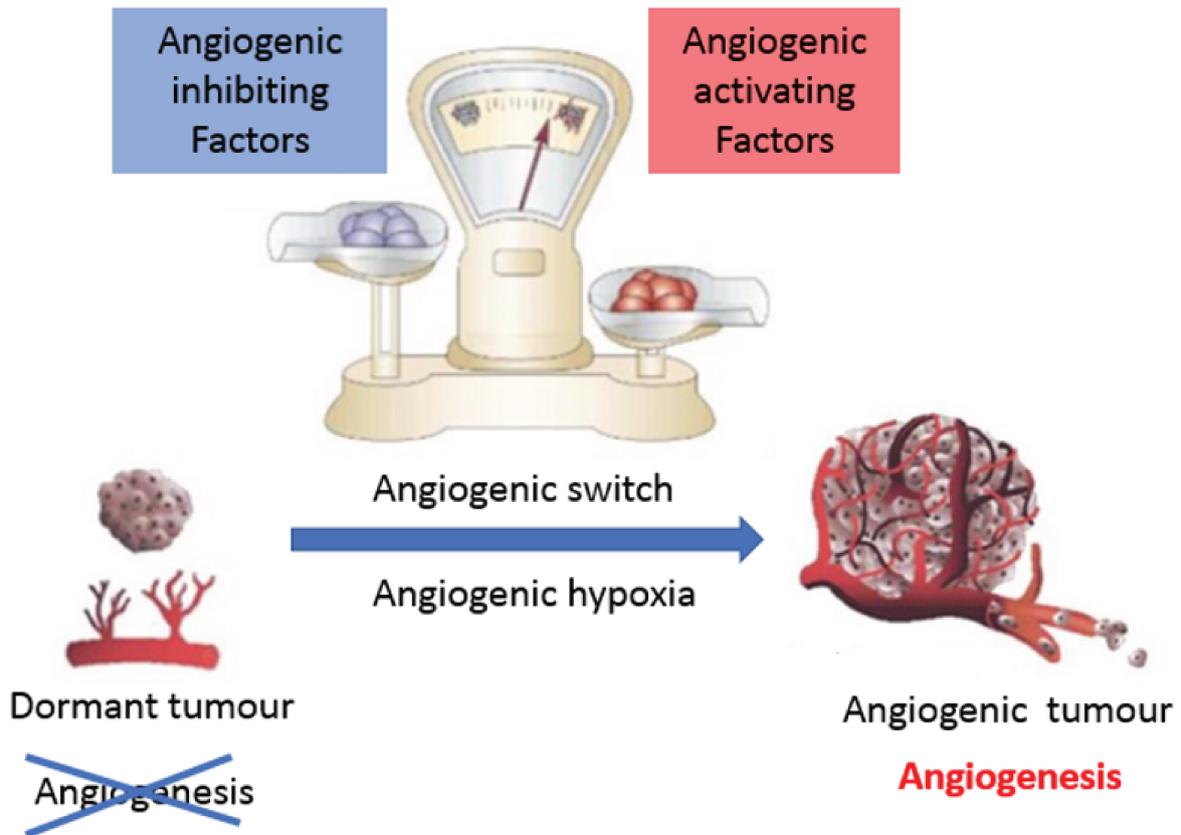


Fig. II.3 The passage from dormant tumour to angiogenic tumour and the angiogenesis switch concept.

as shown in table II.1. Indeed, when the tumour grows more than 1mm, hypoxic areas are more frequent, tumor suppressor genes are inactivated and pro-angiogenic proteins are activated. The angiogenic switch is the result of down regulation of anti-angiogenic and pro-angiogenic factors.

In most adult tissues, the balance between pro-angiogenic and anti-angiogenic signaling is kept resulting in the angiogenesis switch to be "off". In this situation, the secretion of pro-angiogenic factors is balanced by the production of anti-angiogenic molecules. This leads to a dormant, non-angiogenic tumor that is restricted to a microscopic size usually less than 0.5 to 1 mm in diameter [45]. When the angiogenesis switch is "on", the tumours often express increased levels of pro-angiogenic proteins [46]. Among these pro-angiogenic proteins, members of the vascular endothelial growth factor family (VEGF) have a predominant and crucial role in inducing angiogenesis [47]. After the primary vascular plexus is formed, more endothelial cells are generated, which can form new capillaries by sprouting or by splitting from the original VEGF stimulated the

Angiogenic activators factors	Antiangiogenic inhibitors factors
Vascular endothelial growth factor (VEGF)	Endostatin
Transforming growth factor- $\alpha/\beta$	Angiostatin
Platelet-derived endothelial cell growth factor	Interferon- $\alpha/\beta$
hypatocyte growth factor	Interleukin-I2
Tumour necrosis factor- $\alpha$	Platelet factor 4 fragment
Epidermal growth factor	Angiopoietin-2
Placental growth factor	Human macrophage
Tissue factor	Tissue inhibitor of metalloproteinase-1/2
Interleukin-6/8	Vascular endothelial growth inhibitor
Angiogenin	Vasostatin
Angiopoietin-1	Anti-thrombin III fragment
Cyclooxygenase-2	Osteopontin fragment

Table II.1: Endogenous angiogenic and antiangiogenic factors.

migration and proliferation of endothelial cells, formation of a lumen and functional maturation of new endothelium [37].

## II.4 Efforts to prevent tumour angiogenesis

During tumour angiogenesis, tumor cells secrete growth factors that activate the proliferation and migration of endothelial cells and allow capillary morphogenesis. As shown in figure II.4, the endothelial cells form a one-cell thick layer wall called endothelium that lines all blood vessels such as arteries, arterioles, venules, veins and capillaries. Hypoxic tumor regions remain despite this neovascularization. Hypoxic zones continue to stimulate overproduction of vascular endothelial growth factor, VEGF [45, 48, 4].

This mechanism has been noticed in vitro, but it is unclear which factors act in vivo. In vivo, one may be VEGF as it is an endothelial-specific growth factor. By definition, antiangiogenic agents block tumor angiogenesis by acting on VEGF or its receptors.

In 1971, Folkman proposed that since tumour growth and metastasis are angiogenesis depend, blocking angiogenesis could be a good strategy to stop tumour growth [49]. This dependence is the motivation behind targeting of angiogenic factors VEGF, with antiangiogenic therapy.

The relationship between anti-angiogenic and anti-tumor activity was first evaluated in studies with chemotherapy. Chemotherapy refers to the administration of intravenous drugs for cancer

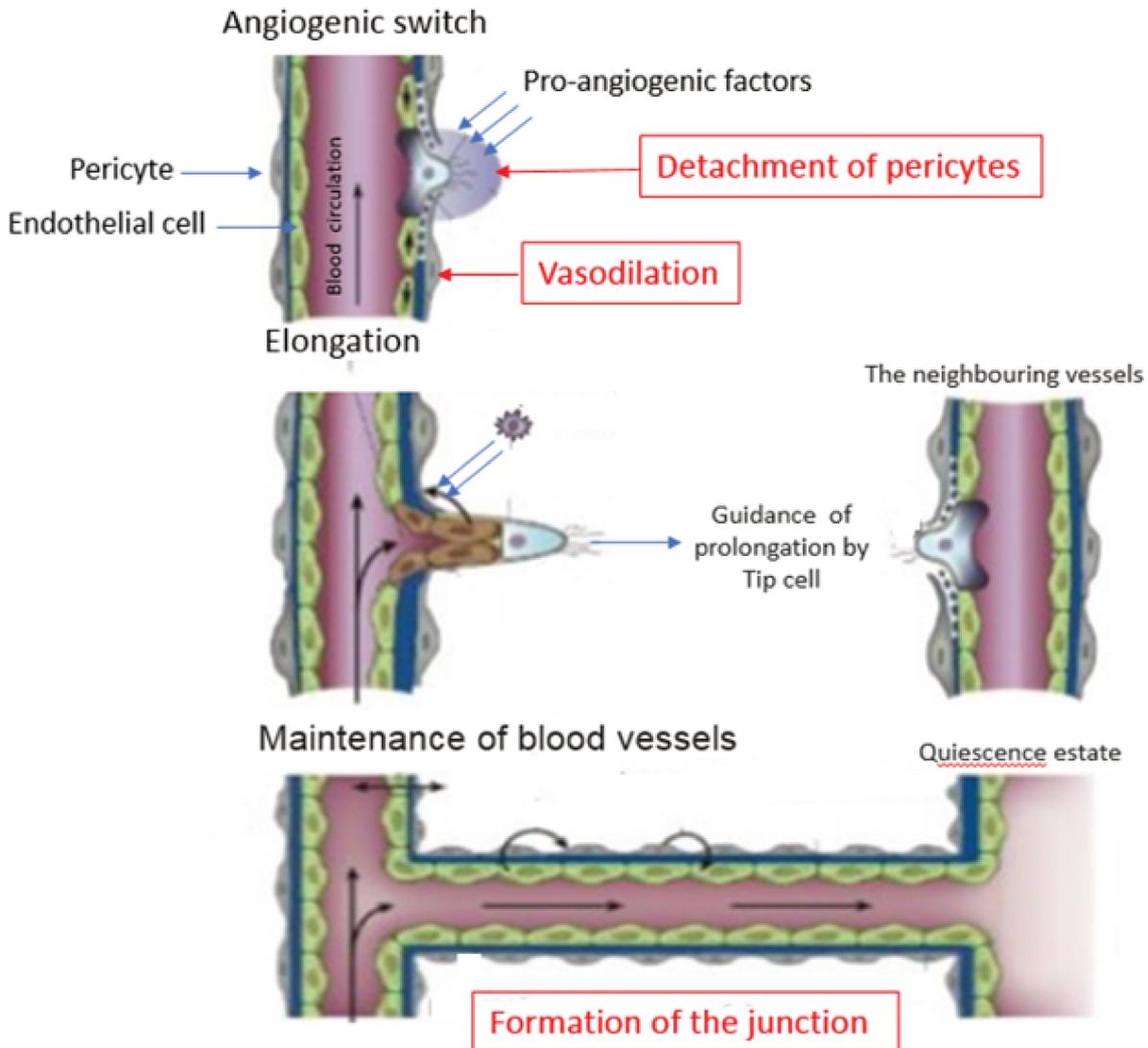


Fig. II.4 Roles of the VEGF family in angiogenesis switch.

treatment [50]. Conventional dosing schedules for chemotherapy, in which maximum tolerated doses are followed by off-therapy intervals to rescue bone marrow, allow recovery of endothelial cells. However, chemotherapy has shown many drawbacks such as tumor recurrence and increased risk of acquired drug resistance. Besides, as the drugs are inducted by the blood circulation, safe tissue can be damaged [51, 52]. Nowadays, anti-angiogenic therapies have taken over chemotherapies (figure II.5). Therefore, many clinical trials have been developed to block pro-angiogenic factors such as VEGF [53, 5, 7].

Two general approaches have been used in the development of antiangiogenic agents: inhibition of proangiogenic factors (eg. therapy with anti VEGF antibody) and therapy with endogenous



Fig. II.5 *Effect of the pazopanib anti-angiogenic treatment on the tumor.*

inhibitors of angiogenesis (eg. angiostatin, endostatin). Therefore, many of these anti-angiogenic agents have been approved in clinical trial to block VEGF. These agents react either by blocking VEGF production by a tumor cell, by blocking its receptor, or by neutralizing VEGF itself. They can be considered as “indirect” angiogenesis inhibitors, because they block a tumor cell product or its receptor.

Many tumour angiogenic inhibitor agents have been approved by FDA. Although they have the same mechanism, preclinical and clinical trails show that not all inhibitors are effective in all cancers. In this context, we cite Pazopanib agent which is an oral angiogenesis inhibitor targeting vascular endothelial growth factor receptor (VEGF-R) [5, 6]. In Metastatic Renal Cell Carcinoma (mRcc), single agent Pazopanib compared with placebo was shown to extend Progression Free Survival (PFS) in Sternberg and collaborator’s work [6]. However, Pazopanib VEGF inhibition is not effective in all cancers[5]. Therefore, we need more understanding on how the neo-angiogenic vasculature can be effectively targeted in tumors [7] in order to improve angiogenesis targeted therapies.

## II.5 Motivation

In the literature, several studies have been done by the national and international scientific communities for the qualitative evaluation of tumour angiogenesis to optimize patient prognosis and anti-angiogenic treatments [54, 55, 56].

As detailed previously, angiogenesis has a significant implication for the design of anti-angiogenic therapy. Therefore, understanding its mechanism, in preclinical and clinical trail should help

Generic name	FDA-Approved Indication
Bevacizumab	Colorectal, non-small-cell lung, and glioblastoma multiforme
Thalidomide	Myeloma
Lenalidomide	Myeloma (myelodysplastic syndrome (MDS))
Sorafenib	Renal cell and hepatocellular carcinoma
Sunitinib	Renal cell and gastrointestinal carcinoma
Temsirolimus	Renal cell carcinoma
Axitinib	Renal cell carcinoma
<b>Pazopanib</b>	Renal cell carcinoma, kidney cancer, and advanced soft tissue sarcoma
Cabozantinib	Thyroid cancer
Everolimus	Kidney cancer, advanced breast cancer, pancreatic neuroendocrine tumors (PNETs), and subependymal giant cell astrocytoma
Ramucirumab	Stomach cancer and gastroesophageal junction adenocarcinoma
Regorafenib	Colorectal cancer and gastrointestinal stromal tumor
Vandetanib	Thyroid cancer
Ziv-aflibercept	Colorectal cancer

Table II.2: FDA-approved inhibitors. These angiogenesis inhibitors are being used in conjunction with other anticancer chemo-therapeutics.

develop effective therapeutic anti-angiogenesis strategies. Assessment of anti-angiogenesis therapies should be done qualitatively and quantitatively to have a specific overview. Both assessments requires an accurate description of the behaviour of the new blood vessels formed in the tumour. For this purpose, we first need a segmentation tasks. Indeed, we need to segment the blood vessel first in order to characterize it. However, in contrast to normal vessels, tumour vasculature is highly disorganized, vessels are tortuous and dilated, with different diameters and excessive branching [57, 58]. These features make the segmentation of blood vessels in tumour images a very challenging tasks. Tumour blood vessels are a part of a very complex matrix termed tumor microenvironment (TME). In medical imaging, TME is well-known by its diverse morphological profiles, with complex and various shapes, especially in malignant tumour.

Besides, a complex disease such as cancer requires an microscopic examination of the tissue that pathologists cannot do by naked eye. The expanding field of digital pathology by adopting novel acquisition methods, especially non invasive ones, such as fluorescence imaging and innovative image analysis technologies such as image processing hence now opened up a new dimension to address our purpose.

Recently, computerized image analysis techniques are highly required since they can expand medical knowledge and provide accurate investigations, especially if these techniques allow to dis-

cover new aspects of the disease that cannot be investigated by naked eyes. However, the most novel computerized image analysis techniques such as neural network require histopathology annotation on the images. Indeed, nowadays image processing network remains a useful tool in pattern recognition and segmentation. However, this tool needs a large amount of annotated data to train our model to ensure a good accuracy. It is common knowledge that the more annotated data a neural network algorithm has access to, the more accurate it can be. This annotation task is a very time-consuming and tedious task for the pathologists.

We are dealing, thus, with the problems stated below:

- (a) Very heterogeneous and complex data.
- (b) Very complex shapes for blood vessels in histopathology images.
- (c) Annotating data is very time-consuming and tedious for Pathologists.
- (d) histopathology sections are an invasive exam so that we usually have a limited amount of data.

In this study, We define our challenges as follow:

1. Which Neural Network is able to segment blood vessels in very heterogeneous images using only very few annotated data?
2. Which imaging model to use for analyzing angiogenesis in preclinical specimen?
3. Which imaging model to use for analyzing tumour angiogenesis in clinical specimen?

Despite their very important role in tumour growth, blood vessels in histopathology stained images have not been widely studied due to their complexity and heterogeneity. To our knowledge, only one study has been dedicated to segmenting blood vessels in histopathology slices precisely in *H&E* images. This study was published in February 2018 by Faliu Yi and al [59]. Although this study presents an automatic method based on neural network to segment blood vessels, no quantitative assessment was performed. Moreover, the proposed method requires the pathologist to annotate a large amount of data (350) by hand which is a very tedious and time consuming task.

In our study, we came up with a full panel to assess tumour angiogenesis process in both pre-clinical and clinical specimen using automatic algorithms. Our panel includes both quantitative

and qualitative assessment tools using multimodal invasive and non invasive imaging.

Our research reflects the multidisciplinary environment in which cancer research has developed nowadays, and includes experiment design, image processing and statistical analysis. Our two main contributions are:

1. Developing a quantitative technique to apply and assess tumour blood vessels and their interaction with nuclei, apoptosis distribution across whole-slice histological sections of tumors, by using noninvasive imaging techniques: fluorescence imaging.
2. Proposing a novel method in neural network which combines pure data driven learning model and rule based knowledge driven model (Marked Point Process) as qualitative techniques to assess tumour angiogenesis in *H&E*. This novel method overcomes two major problems: segmentation of blood vessels in very heterogeneous and complex images and annotation, a very tedious task for the Pathologist.

Considering our main biological contribution, we will try to investigate if there is any direct interaction between anti-angiogenic therapies and tumour blood vessel formation. Indeed, all studies dedicated to tumour angiogenesis have been investigating the effect of anti-angiogenic agents on the pro-angiogenic protein (such as VEGF). To our knowledge, no study has been investigating the direct interaction between anti-angiogenic agents and tumour blood vessels formation.

Therefore, we will try to answer the following questions:

- (a) Do anti-angiogenic agents such that pazopanib have any direct effect on the endothelial cells besides the inhibition of the pro-angiogenic protein?
- (b) Do anti-angiogenic agents have any impact on the tumour cells?

We aim at developing a pipeline to assess angiogenesis with histopathological precision, to contribute to the identification of architectural features that are often associated with therapeutic responses. Blood vessel segmentation is the first important task to reach this objective.

To achieve this target, we structure the pipeline as detailed in figure [II.6](#).

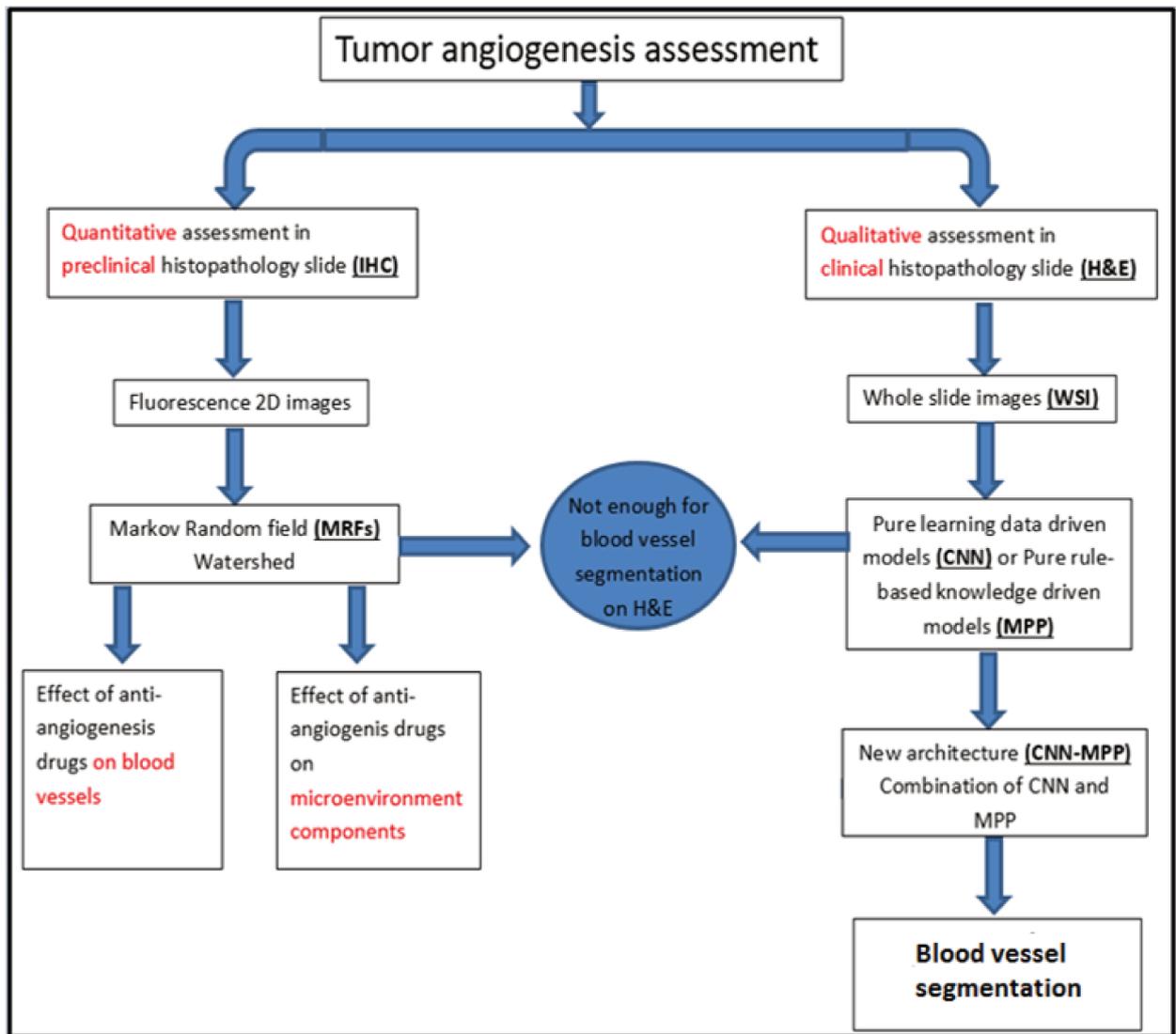


Fig. II.6 The pipeline of the thesis. In this figure, we detailed the developed quantitative and qualitative tools to assess the tumour angiogenesis.



# Chapter III

## immunochemistry for preclinical tumor angiogenesis assessment

### Abstract of the chapter

Axio-scan 2D fluorescence imaging is a non-invasive imaging technique allowing a faster and complete acquisition of immunochemistry staining. We develop an experimental protocol to collect pre-clinical data (mice), then we establish a automatic framework based on Markov Random field and Watershed, in order to segment endothelial cells, apoptosis areas and nuclei. In order to investigate the effect of the Pazopanib as anti-angiogenic drug on tumour vasculature and the tumour micro-environment, we develop a statistic framework to qualitatively assess the apoptotic percentage of endothelial cells among the overall apoptotic area.

### III.1 Confocal fluorescence microscopy versus Conventional fluorescence microscopy

During this study, our motivation is to assess experimentally if the anti-angiogenic drug Pazopanib has a direct effect on blood vasculature and therefore on other components of the tumour micro-environment. For that, we develop a bio-imaging framework including:

1. An experimental design on murine cells using fluorescence imaging,
2. A computer algorithm to segment endothelial cells (blood vessels),
3. Metrics for quantitative assessment of the drug effect on endothelial cells and nuclei.

The fluorescence microscopy imaging is based on acquiring fluorescence data from a sample, such as biological tissue, to facilitate diagnosis of the presence or absence of disease or other abnormality [60]. Several recent studies have shown the interest of fluorescent microscopy imaging in drugs investigation [61, 62]. This kind of imaging can be an important tool which allows better reproducibility than non-fluorescent immunostaining.

The fluorescence microscopy takes advantage of the phenomenon of fluorescence in addition to conventional observation by reflection or absorption of artificial visible light in a microscope. Immunofluorescent staining makes use of antibodies to locate and identify patterns of protein expression in cells by a fluorescent microscope. Therefore, primary antibody binds to antigen (figure III.1). Then antibody-antigen complex is bound by a secondary antibody conjugated to a fluorochrome. Upon absorption of high energy light, the fluorochrome emits light at its own characteristic wavelength (fluorescence) and thus allows detection of antigen-antibody complexes as described in figure III.2 The fluorescent information is filtered out according to the wavelength. Only specific structures at specific wavelength is detected by the filter and converted to pixel intensity information in the image.

Fluorescent anti-bodies (FPs) have become valuable tools for investigating biological processes in living cells. They represent suitable biochemical markers for a variety of biological applications, non-invasive monitoring of gene expression as well as for studying subcellular localization and dynamics of proteins [63].

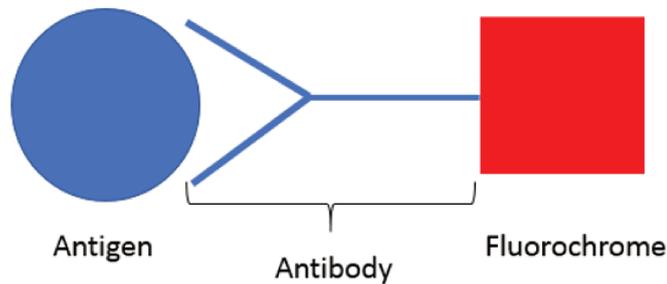


Fig. III.1 The fluorochrome is fixed on the antibody to form the whole complex with the antigen

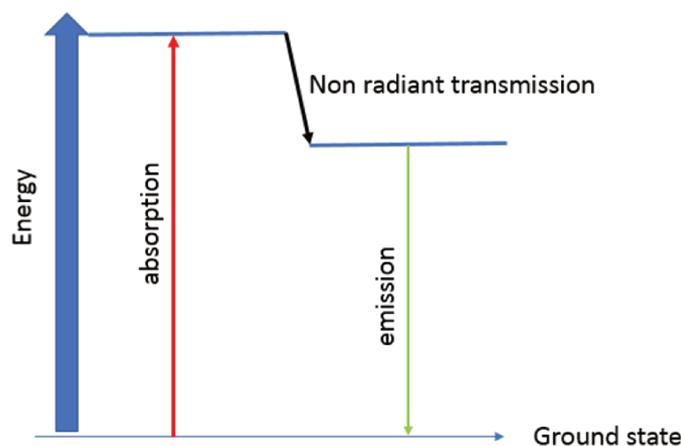


Fig. III.2 The emission-absorption concept behind the fluorescence phenomena in fluorescence imaging

Recently, confocal laser scanning microscopy (CLSM) or laser confocal scanning microscopy (LCSM) was widely used as a novel imaging technique to assess vascular endothelial growth factor (VEGF) for developing anti-angiogenic therapeutic studies, as it represents a powerful tool for generating high-resolution images and 3D reconstruction of specimen. These studies concerned gastrointestinal cancers for *in vivo* visualization [64], colorectal cancers [65] and immunohistochemical staining of cancer stem cell markers in hepatocellular carcinoma [66]. CLSM was used to overcome limitations of traditional fluorescence microscopes. Indeed, in the conventional fluorescence microscope, all parts of the specimen in the optical path are excited at the same time and the resulting fluorescence is detected by the microscope including a large unfocused background part. In contrast, confocal microscope uses point illumination and a pinhole to eliminate out-of-focus

<b>Confocal fluorescence microscopy</b>	<b>VFM Axio 2D Zeiss</b>
Makes use of antibodies	Makes use of antibodies
Locate and identify patterns of VEGF protein expression in cells	Locate and identify patterns of VEGF protein expression in cells
Few Cross Talk	Low possibility of Cross Talk
Very time consuming	Fast and reliable slide scanner
Low axial resolution	High resolution and precision
Partially automated systems	Fully automated and boxed systems

Table III.1: VFM Axio 2D Zeiss versus Confocal fluorescence Microscopy

signal - the name "confocal" stems from this configuration. As only light produced by fluorescence very close to the focal plane can be detected, the image's optical resolution - particularly in the sample depth direction -, is much better than that of wide-field microscopes. The CLSM achieves a controlled and highly limited depth of focus.

## III.2 Virtual Fluorescence microscopy Axio 2D

Although CLSM has been widely used in the last decade for developing drug trails for anti-cancer therapies, it has its own drawbacks such as low resolution in the z-axis acquisition and photo damage because of illuminating the tissue sample through entire z-axis, but the most important one is the speed. Indeed, the acquisition is very slow as it scans the specimen point-by-point. For that reason, to do one whole slide image acquisition, we need one full working day. If we try to make it faster, we are compromising the resolution of the image.

In order to take advantage of the fluorescence concept and overcome the speed problem, we use for our experiment, virtual fluorescence microscopy Axio 2D developed by Zeiss. This technology allows scanning many whole slide images at once with very interesting speed as it scans the specimen by field of view which is a fixed parameter by the user. Therefore, the acquisition time is limited without compromising the resolution of the image. We detailed in table III.1 the advantage of using the virtual fluorescence microscopy (VFM) Axio 2D Zeiss comparing to confocal fluorescence microscopy.

## III.3 Database preparation

### III.3.1 Cell culture and tumor model

CT26 is a murine colorectal carcinoma cell line from a Balb /c mouse which are an albino laboratory-bred strain of the house mice which is the most widely used in animal experiment. Preclinically, the murine CT26 colon carcinoma line has become a platform model for evaluating the potential of drug combinations with immune checkpoint inhibitor antibodies, for studying the mechanisms of tumor growth and metastasis and evaluating the effects of different drugs and therapies in animals.

To assess the pre-clinical tumour angiogenesis process, we established a protocol for colon carcinoma cell lines culture in "Animalerie Lab" in Sorbonne University in Paris, France.

This protocol was realised with the collaboration of ITD team (Développement d'imagerie et de thérapie ciblée pour le cancer) under Laboratoire d'imagerie biomédicale (LIB) supervision as part of a project with the Fondation pour la Recherche Médicale (FRM).

Murine colorectal tumor cells, CT26 (American Type Culture Collection, Virginia) were cultured in a specific medium. Two thousand tumor cells in 100  $\mu$ L of  $2 \times 10^{-6}$  cells/ml were then injected subcutaneously on the left flanks of 3 Balb /c mice to generate first generation tumors: G1.

16 days after cell implantation, we consider that tumors are large enough to be used to make a second generation of tumors: G2. Therefore, tumors were removed after euthanasia of the mice. Excised tumors were dissected into 20 – 40 $mm^3$  fragments which were then implanted in the left flanks of eight other Balb /c mice ( $n = 8$ ) via a small incision that was closed with one or two sutures. The date of fragment implantation is referred to as day zero (D0). We underline the need to make G2 tumors because they are more reproducible.

Thereafter, ultrasound imaging was performed every 2 to 3 days to evaluate the tumor size evolution. On Day 7, four mice, AA1, AA2, AA3 and AA4 were treated by gavage with anti-angiogenic drug, Pazopanib, repeated day by day (2 mg diluted in 100 microL VOTRIENT®). Mice labeled P1, P2, P3 and P4, received placebo by the same route and administration schedule. On Day 18, all mice were euthanized and tumors were removed carefully to conserve orientation relative to the anterior surface (skin side) and the longitudinal axis (We kept this information

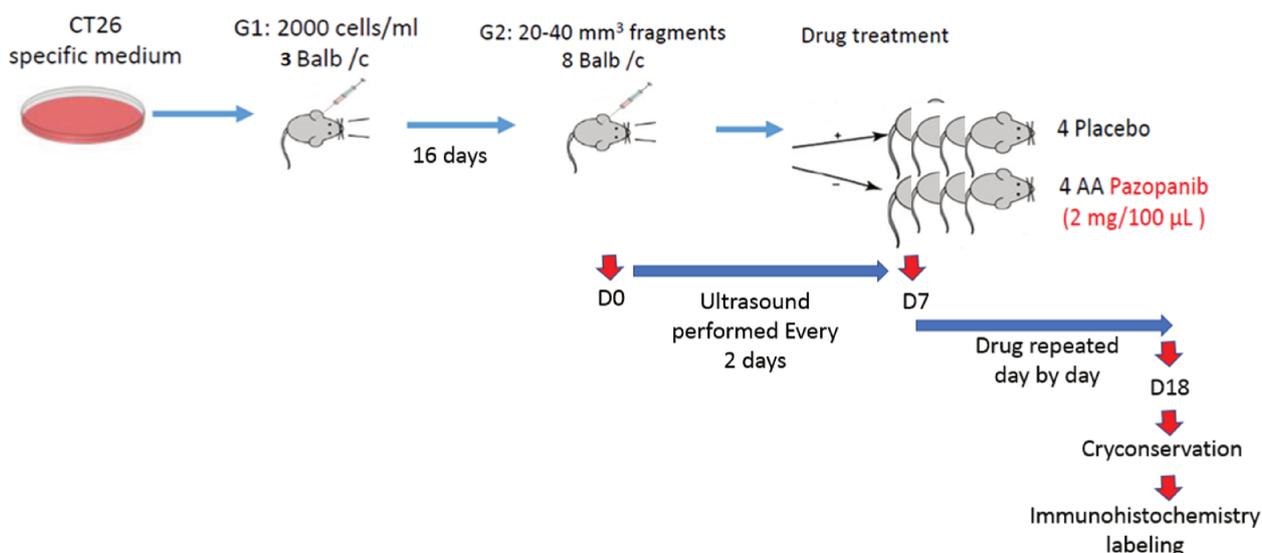


Fig. III.3 *Biological experiment design of IHC specimen for fluorescence scan. G1 refers to the first generation culture of colorectal tumour. G2 refers second generation culture of colorectal tumour. Drug treatment is anti-angiogenic Pazopanib. Cryopreservation refers to the process where whole cells or tissues are preserved at very low temperatures, in our case  $-80^{\circ}\text{C}$ .*

by putting a mark on the sample). Tumors were prepared for cryoconservation and embedded in Optimum Cutting Temperature gel (tissue-tek<sup>®</sup> OCT, VWR, France).

Tumors were stored in a freezer at  $-80^{\circ}\text{C}$  until labeling for immunohistochemistry.

All immunohistochemistry marking was performed on the same day for the entire sample set. Tumors were sectioned to obtain a central section along the longitudinal tumor axis following the Frozen section procedure with a microtome. Sections were 8 to 10 microns thick.

### III.3.2 Immunohistochemistry labeling (IHC)

Immunohistochemistry (IHC) is a method of locating proteins in the cells of a tissue section by detecting antigens using antibodies. IHC exploits the fact that an antibody specifically binds to antigens in biological tissues. This staining method allows to localize differentially expressed proteins in the tissue. In the table bellows, we detailed the bio-marker and its main activity. In our experiment, as we are interested to assess the apoptosis caused by the Pazopanib drug on endothelial cells and its effect on nuclei on the tumour surface we use DAB, CASP3 and CD31.

Triple immunohistochemistry marking was performed on each of the 8 slices (1 slice per mouse).

Biomarkers	Cellular event
KI-67	Assessment of the tumour proliferation activity
Phospho-histone-H3	Cell undergoing mitosis
<b>CD31</b>	Presence of endothelial cells in histological tissue
<b>CASP3 protein</b>	Apoptotic cells
<b>Diaminobenzidine (DAB)</b>	Detect fingerprints in blood
Feulgen	To identify chromosomal material or DNA
Schiff	To detect polysaccharides such as glycogen
Wright	Differentiation of blood cell types.
Toluidine blue	Tissues rich in DNA and RNA

Table III.2: IHC antibodies and their cellular activities

Triple labeling provides marking of the neovascularization, apoptosis and total nuclei across the tumor surface such that the relative position of structures can be evaluated. Isolectin B4-A488 was used to mark endothelial cell showing neovascularization area. Caspase 3 IgG antibodies followed by a second anti-IgG antibody TRITC were used to mark apoptosis. 4', 6-diamidino-2-phénylindole (DAPI) was used to mark nuclei.

### III.3.3 Automated multi-fluorescence scan imaging using VFM Axio 2D Zeiss

After labeling tumor slices, DP data was obtained using a 2D scanner, Zeiss Axioscan (Zeiss, France) as shown in figure III.4. It allows for faster acquisition more than any fluorescence microscopy. The scan is made field by field and assembled the high resolution fields into a whole slice image. The system can record data for up to 3 separate labels by filtering for the fluorophore of each marker. The RGB images obtained were recorded as .CZI files and visualized using the image processing software ZEN.

The utility of this device is the ability to go up to 4 immunostains in histochemical analysis. Indeed the acquisition of the images is done by filtering the different wavelengths reflected by the stained specimen, so that each wavelength corresponds to a structure in the tumour microenvironment, as shown in table III.3. At the device, there are 4 color filters (Dapi, GFP, Cy3, Cy5) that can retrieve the information needed, according to the modification of the acquisition parameters. Therefore, different structures of the tumor can be visualized separately. In our case we target the nuclei, the endothelial cells and the apoptotic area. The 3 different images obtained by the 2D scan are shown in figure III.5 As the scanner can filter up to 4 wavelength, we can see separately: the

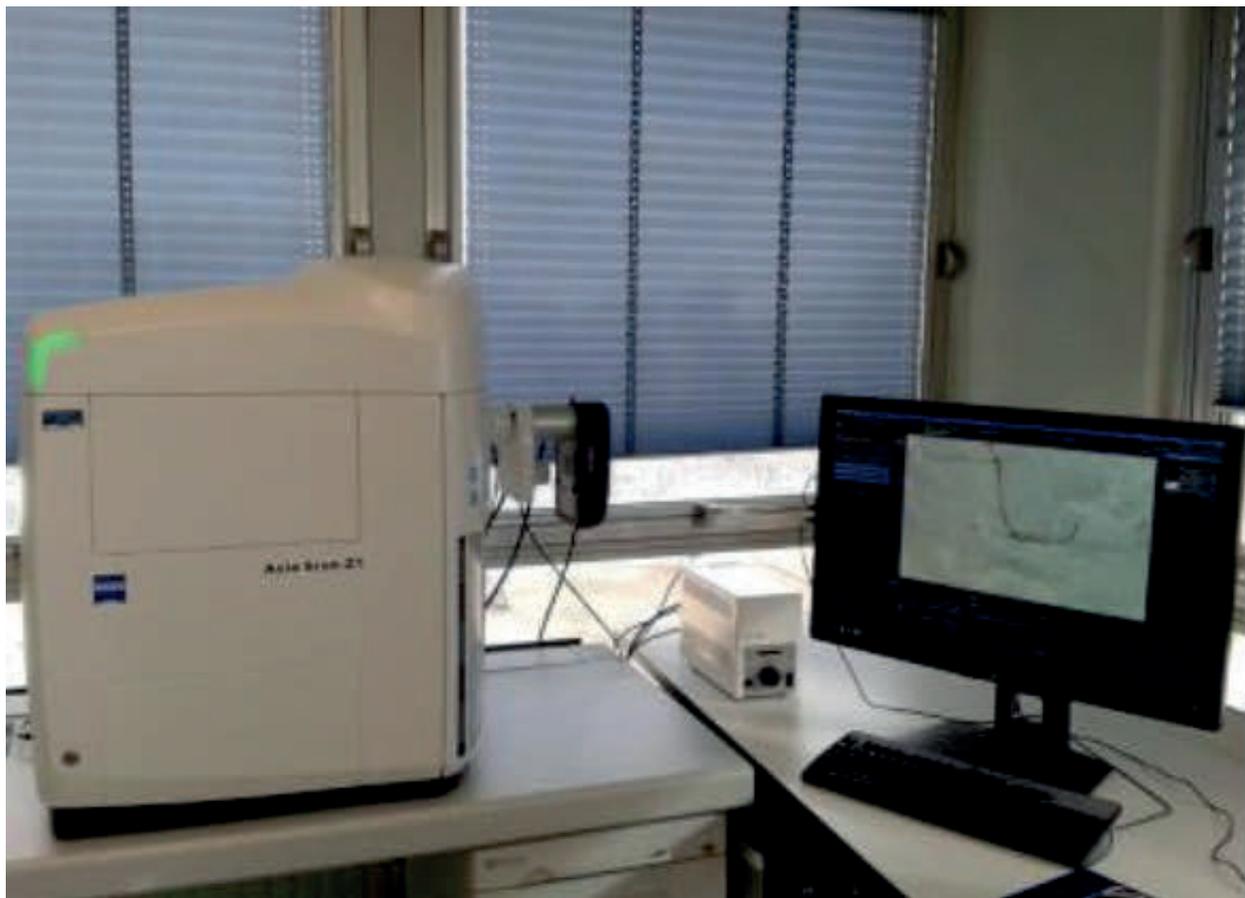


Fig. III.4 The Zeiss Axioscan blade scanner from the "Animalerie" at the Cordeliers site, Sorbonne University, Paris, France. The scanner is equipped with a workstation that allows you to view the images acquired by Zeiss.

nuclei in blue (figure III.5.a) as the DAPI marks the DNA tumour cells, DNA, and consequently nuclei, the endothelial cells of the blood vessels in green (figure III.5.b) as isolectine marks the endothelial cells and the apoptosis caused by the antiangiogenic drug Pazopanib as Trict marks the apoptotic areas in Red (figure III.5.c).

Filters	Excitation	Emission
DAPI	342-388	420-470
GFP	450-490	500-550
Cy3	530-555	570-640
Cy5	625-645	665-715

Table III.3: Absorption Spectra of Fluors Commonly Conjugated to Secondary Antibodies.

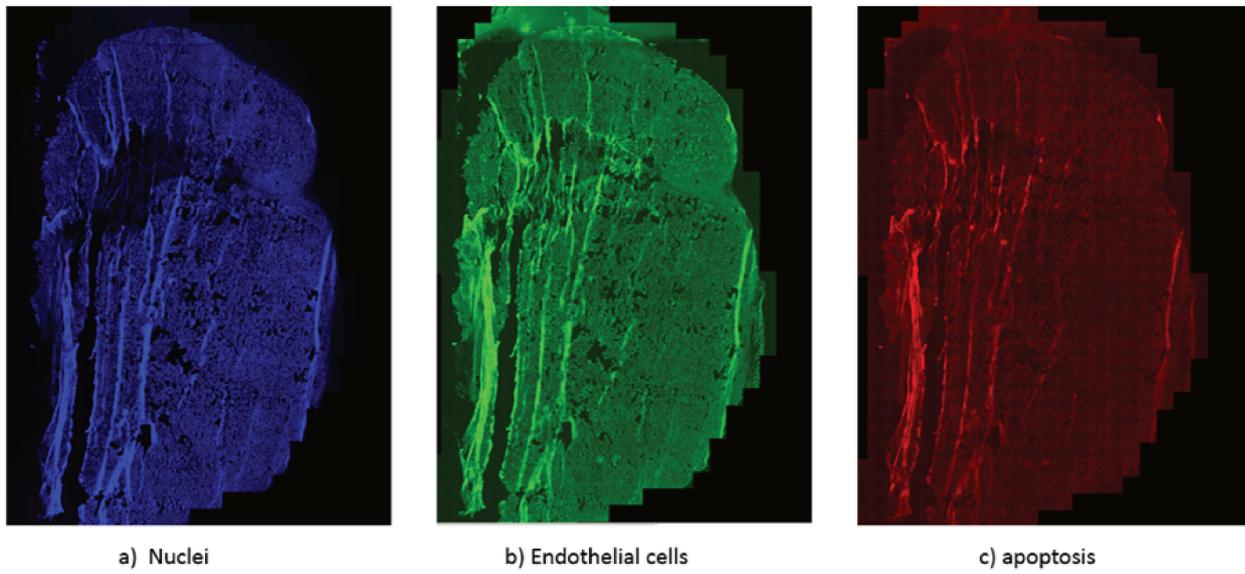


Fig. III.5 The images after reconstruction and pre-treatment of the AA1 anti-angiogenic mouse (a) the blue channel representing all the nuclei of the tumor surface (b) The green channel representing the endothelial cells (c) the red channel representing apoptotic areas on the tumor surface.

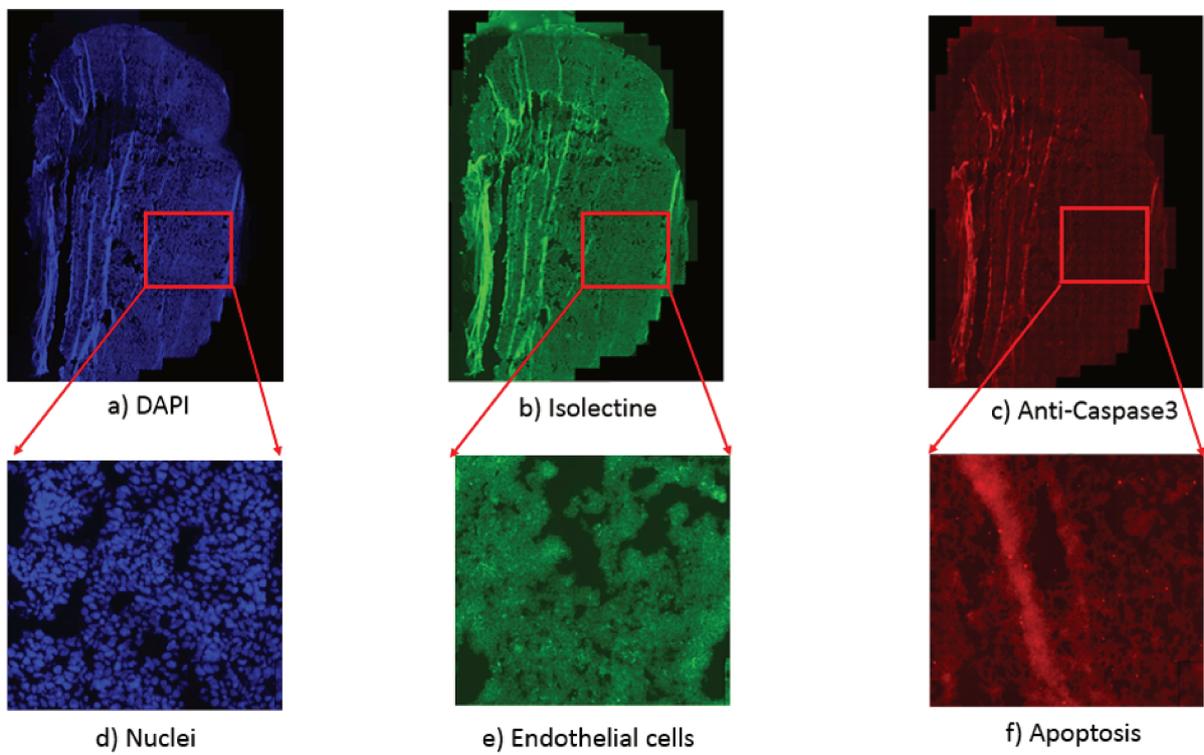


Fig. III.6 Patch extracted from the whole fluorescence images.

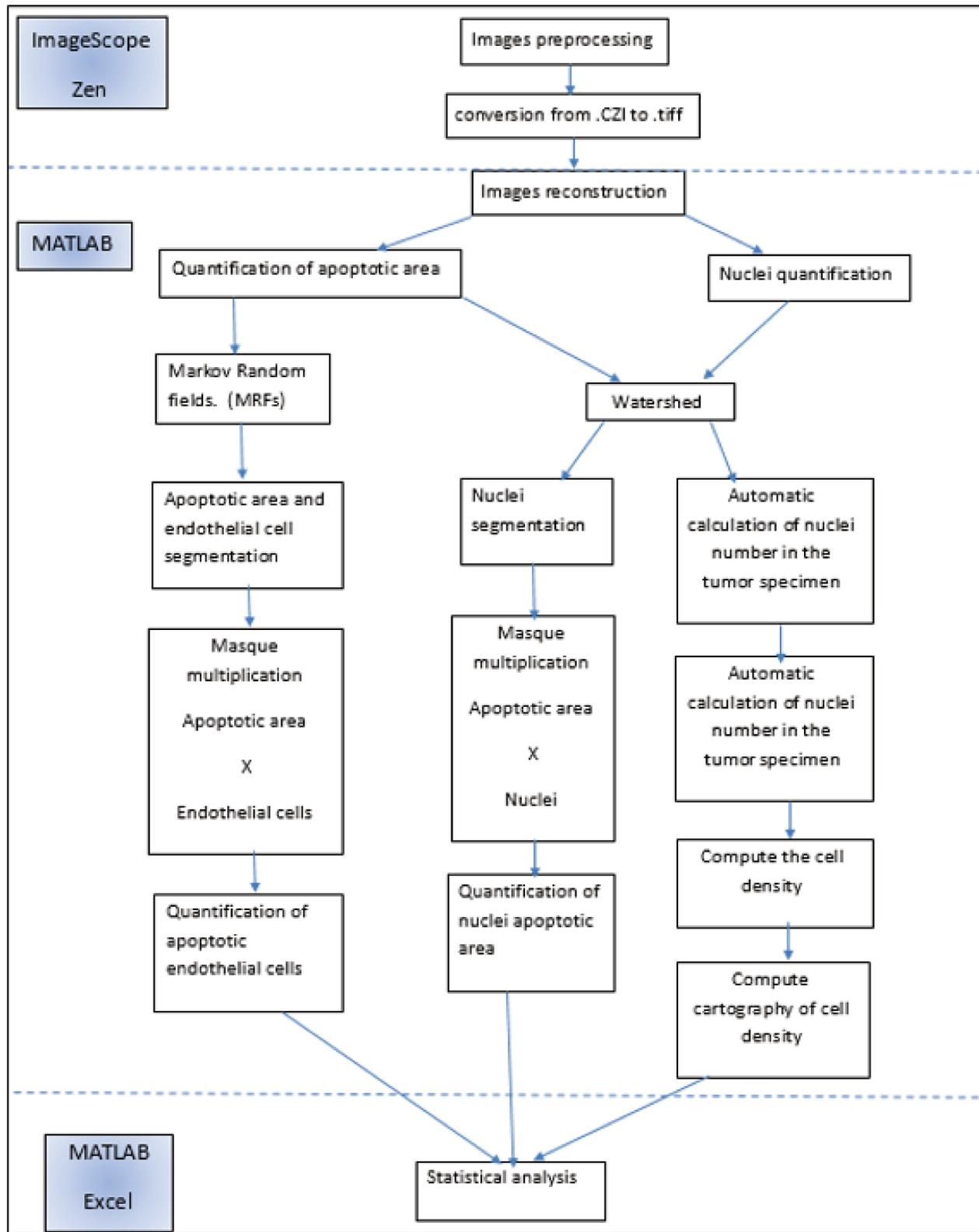


Fig. III.7 Flowchart for quantitative angiogenesis assessment in fluorescence imaging.

## III.4 Markov Random Field for preclinical tumor angiogenesis assessment

### III.4.1 Preprocessing

Quantification software was developed in house using Matlab (version 9.0) to evaluate the number and density of structures based on the fluorescent microscope data. However, CZI files are not a supported format in Matlab. Thus, files were exported in Tagged Image File Format (TIFF) to preserve the best image quality possible. Our algorithm starts by obtaining 3 binary masks of images for each of the 8 tumor sections. Regions of interests (ROIs) were extracted at 0.325  $\mu\text{m}/\text{pixel}$  resolution and stored in Tiles TIFF files with tiles size (1024\*1024 pixels). This tile size was fixed by our team biologist.

For each tumor's slice, segmentation was performed on each tile (1024\*1024) separately. To contribute to the assessment of Pazopanib anti-angiogenic effect on vascular and cellular structures, quantification of the spatial distribution of nuclei cell, apoptotic area and neovascularization apoptotic area (endothelial cell) was required.

We aim to get 3 binary masks of every tumor's slice:

1. nuclei segmentation,
2. apoptotic area segmentation,
3. apoptotic neovascularization.

For each tumor slice, trained biologist has manually annotated the ground truth on the tiles using Aperio Imagescope for final evaluation. One day of work was approximately needed for each slide to be annotated.

### III.4.2 Nuclei density maps

In the literature [67, 68, 69] one usually use a simple threshold method to separate the background from pixels belonging to an object of interest. In our case, each one of the  $n$  connected groups of pixels tagged as nuclei is analyzed, and then watershed segmentation was performed inside each component to split it into individual nuclei. Euclidean distance was used to define connection.

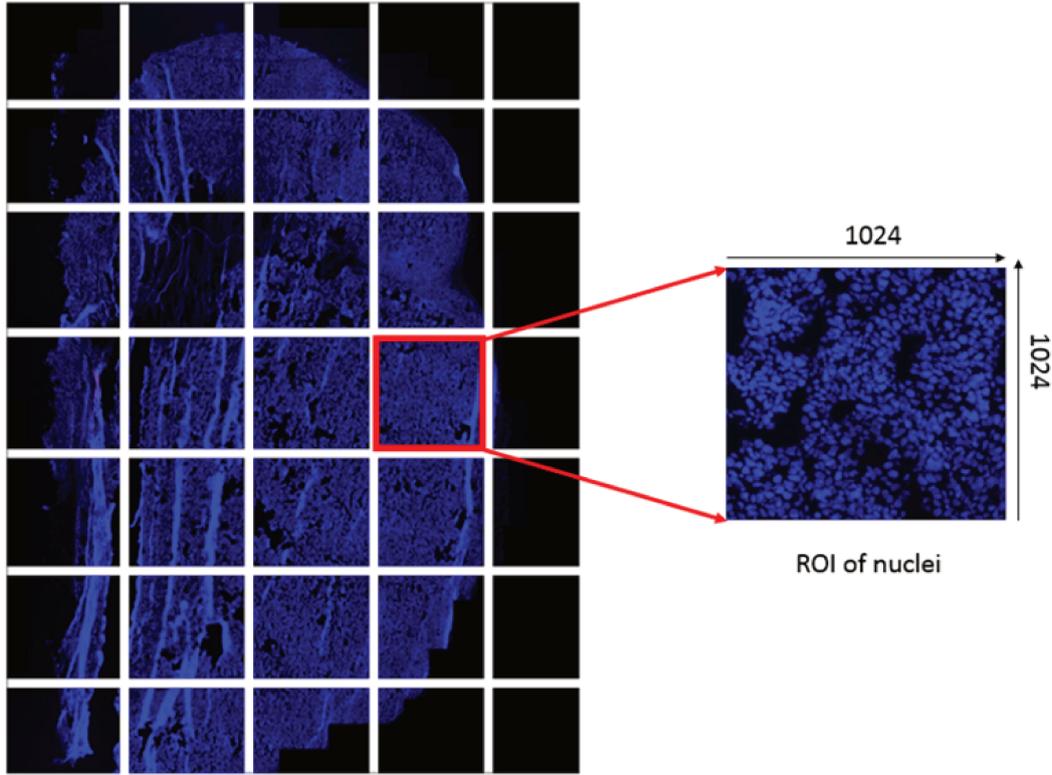


Fig. III.8 The nuclei cross the tumour surface. all the segmentation process was performed with a sliding window of 1024x1024 pixels

Watershed segmentation method was applied to the blue images describing the nuclei distribution. This method provided a binary mask of all nuclei in the tumor area as shown in figure III.13. These segmentation tools are derived from mathematical morphology [70]. Watershed segmentation allows a partition into multiple regions according to a topographic approach [71, 72].

Then, by calculating the number of nuclei on the whole mask in each (1024\*1024) pixel tile by sliding window, we obtained maps describing the local density (nuclei number per tile over the tile area) to assess the nuclei distribution. In fact, the local density per each tile zone on the image was quantified on the mask following equation III.4.2 and grouped into a matrix called the nuclei matrix.

$$ND_{tile} = \frac{NN_{tile}}{TA_{tile}}$$

Where  $ND_{tile}$  is the nuclei density per tile,  $NN_{tile}$  is nuclei number cross the tile area and

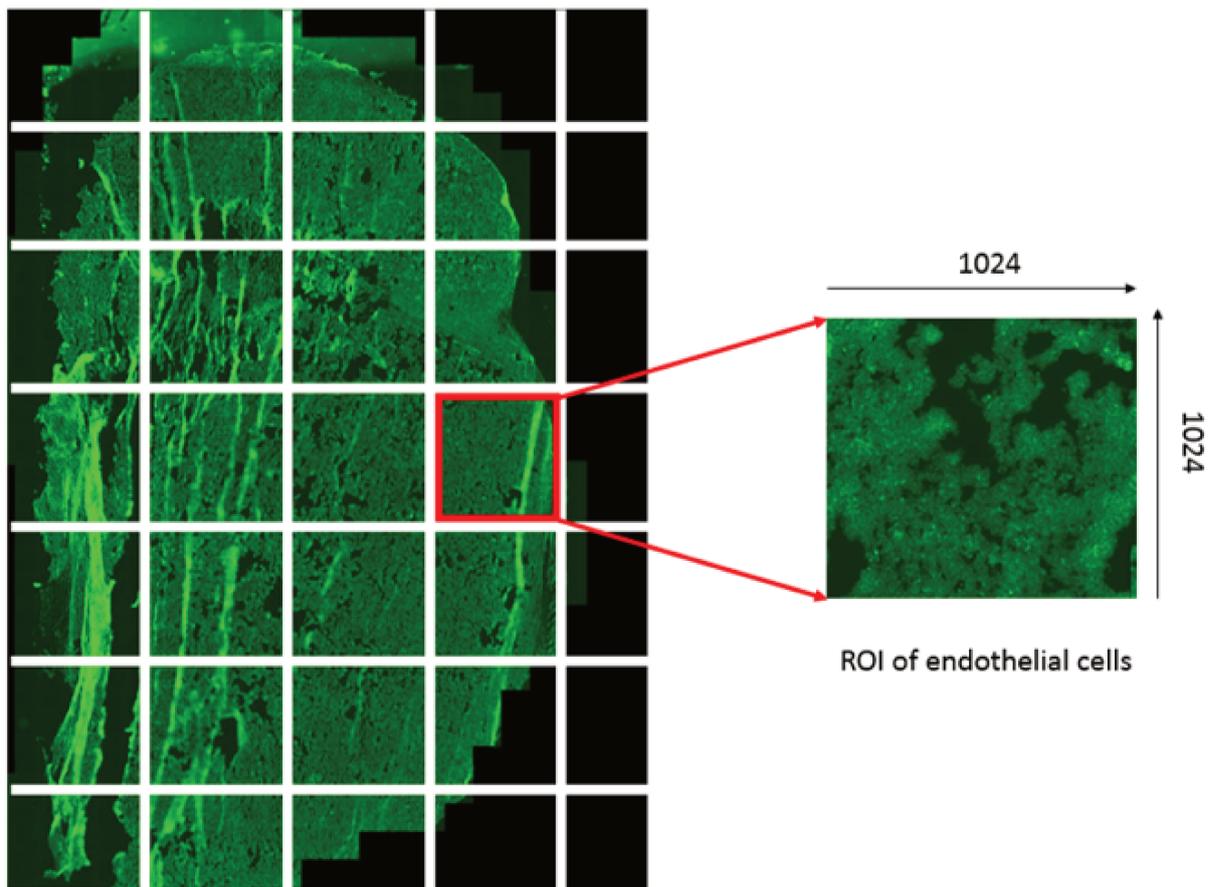


Fig. III.9 The endothelial cells cross the tumour surface. all the segmentation process was performed with a sliding window of 1024x1024 pixels

$TA_{tile}$  is the tile area. Based on the cell structure, maps of the density of nuclei were calculated across each whole slide fluorescence image.

### III.4.3 Apoptotic endothelial cell assessment

The structure of endothelial cells and apoptotic regions are not well defined as shown in figure III.8 and III.10 respectively. A global threshold was not sufficient to separate these regions from the background as shown in figure III.14 Especially, there are fluorophore artifacts shown in figure III.9 with dark green. Not all green area on the figure III.12 represent endothelial cells. We thus chose to apply a stochastic segmentation method to evaluate these features. We used Markov Random Field model segmentation (MRFs) based on pixel intensity on mono-channel: green one for endothelial cells and red one for apoptotic area.

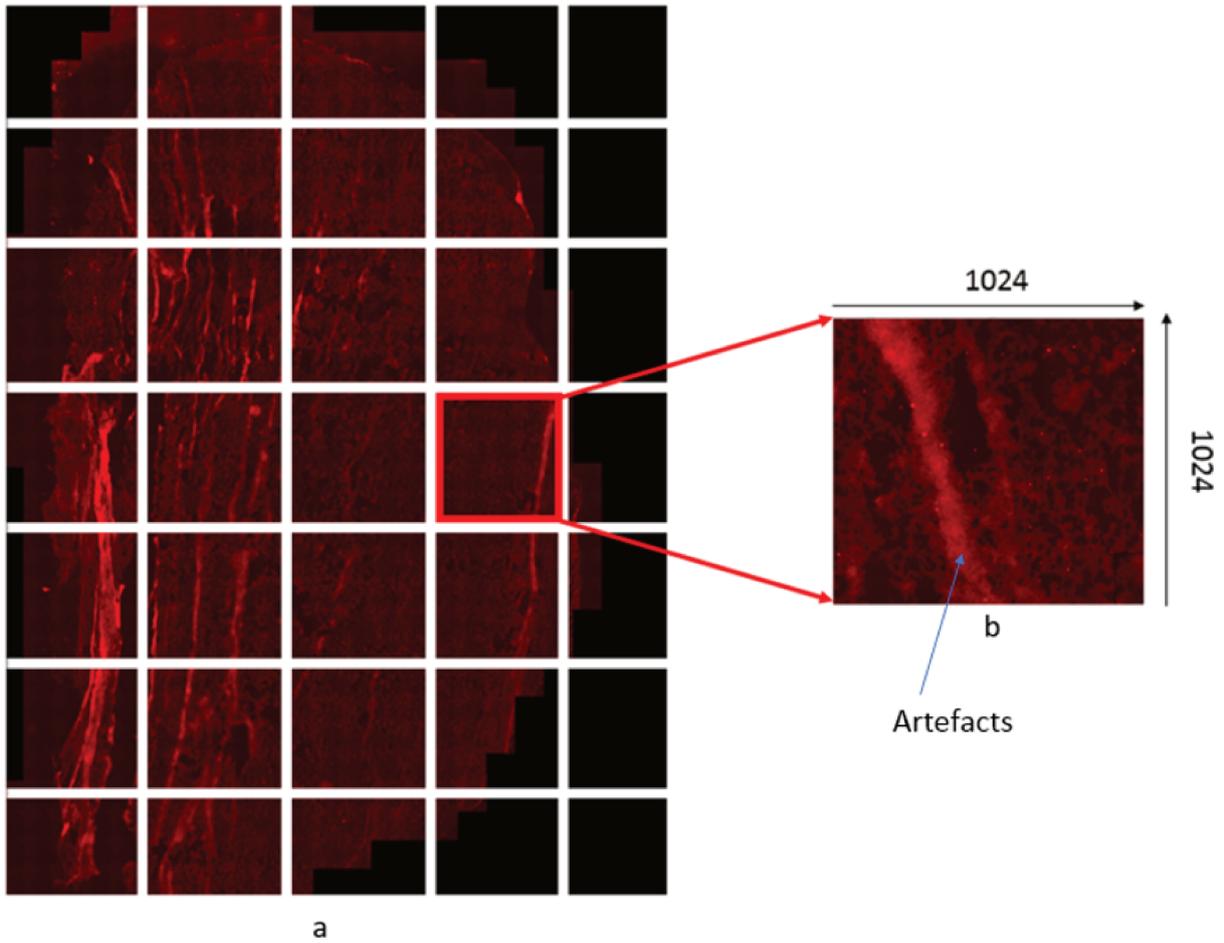


Fig. III.10 The apoptotic areas cross the tumour surface. all the segmentation process was performed with a sliding window of 1024x1024 pixels

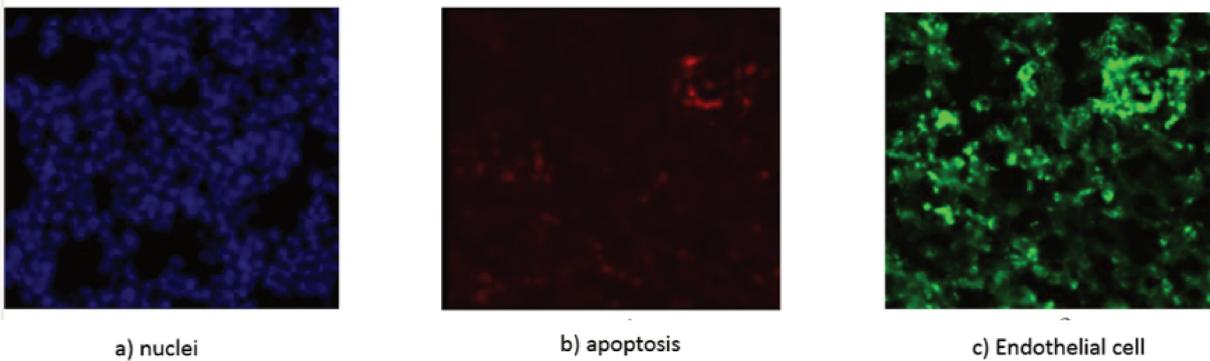


Fig. III.11 One field (1024\*1024 pixels) of the whole slice Multi-Fluorescence Scan imaging of tumor slice of P1 mouse. (a) DAPI channel representing nuclei in the field (b) Cyx channel representing apoptotic areas in the field (c) Cyx representing endothelial cells in the field (neovascularization) (Magnification: 215%)

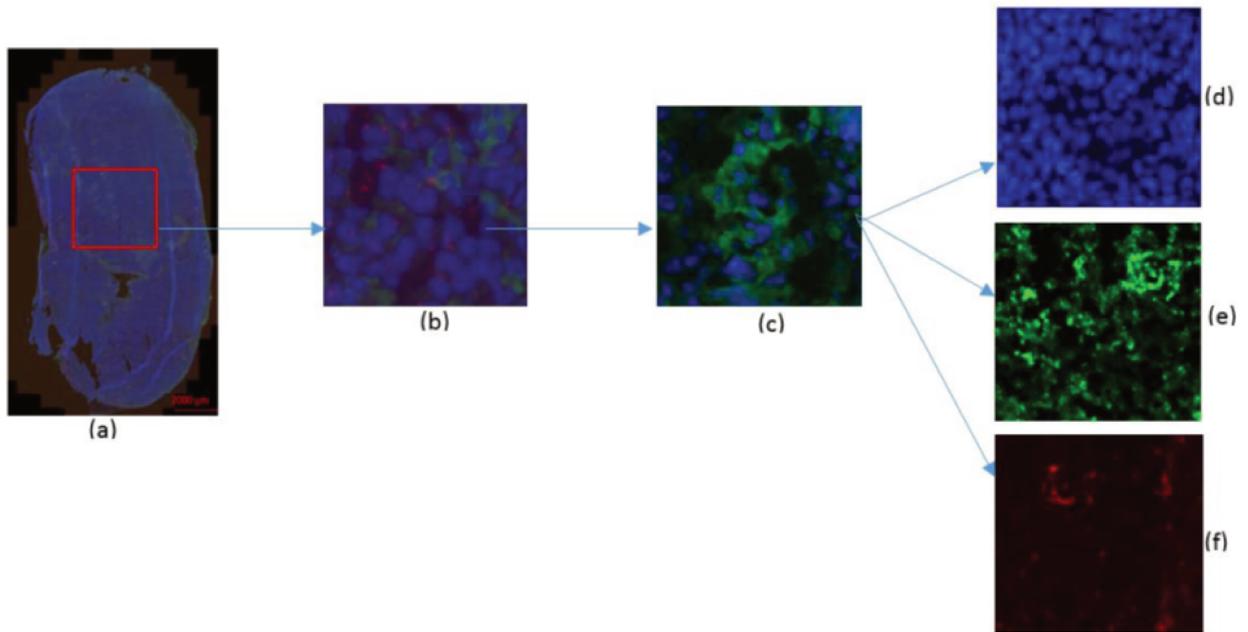
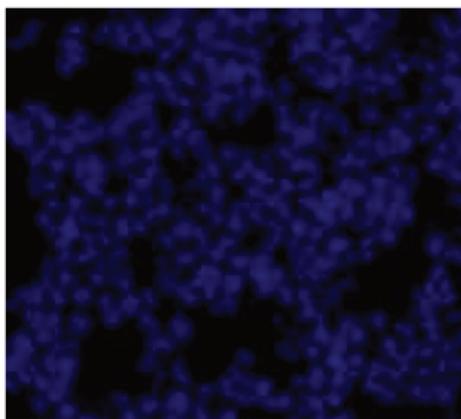
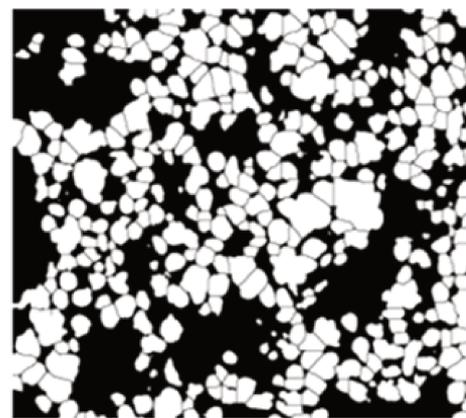


Fig. III.12 One field of the whole slice Multi-Fluorescence Scan imaging of tumour slice of P1 mouse, (a) Whole slice tumour of triple immunohistochemistry marking of P1 mouse (b) Region of interest triple immunohistochemistry marking of P1 mouse field



a) Nuclei in the input image



b) Watershed segmentation on the nuclei ROI

Fig. III.13 Nuclei segmentation on one ROI 1024x1024 using watershed algorithm

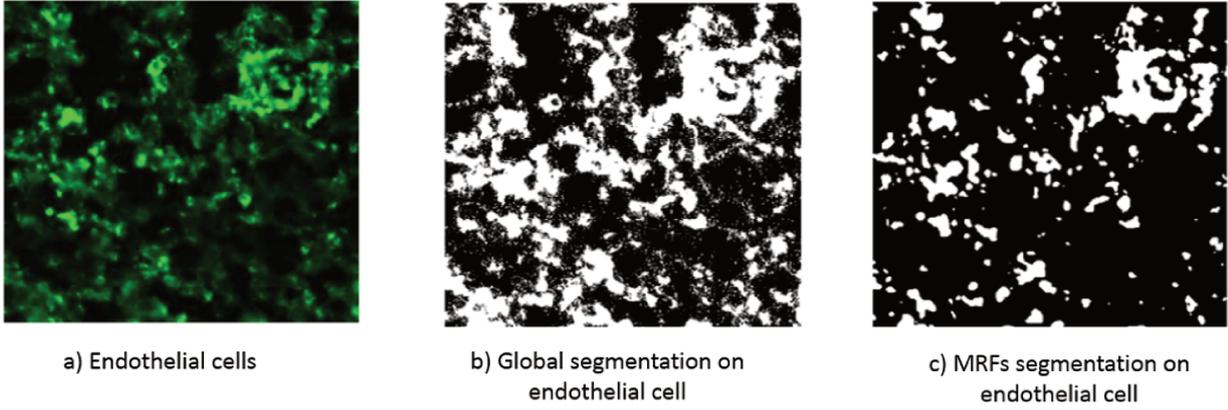


Fig. III.14 Comparison between the endothelial cells segmentation performed with a simple threshold and a Marked Random field threshold. On the left the 1024x1024 tile before segmentation. On the right the 1024x1024 tile after MFR segmentation. In the middle the 1024x1024 tile after Otsu threshold segmentation

We considered segmentation as a pixel labeling task to obtain an endothelial cell binary mask and a binary mask of apoptotic region throughout the tumor's cross-section. We aimed to classify pixels into two classes. For each tumor's slice, segmentation was performed on each tile (1024\*1024) separately. Our goal is to segment each tile on 2 classes: background and endothelial cell. Our approach starts with a Gaussian Mixture Model [73] (GMM) which segments the images on 3 classes: endothelial cell class, background and transition class pixel then a Markov Random Field (MRFs) [74] is applied to have the 2 classes. The third class represents uncertainty pixels between endothelial cell and artifacts. The main difficulty is that the model and its parameters are not known and need to be computed.

Our image  $Y$  is a set of intensity data  $Y = \{y_s\}$  where  $s$  is the location of the pixel  $y$  in the image and  $y_s$  its intensity.

Let us note  $X = \{x_s\}$  the segmented version of  $Y = \{y_s\}$  where  $x_s \in \{1, 2, \dots, K\}$  and  $K$  is the number of different regions.

In the Bayes theorem shown in equation III.4.3,  $P(Y)$  is a constant normalization factor. the conditional probability is then shown in equation III.4.3.

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$$

$$P(X|Y) \propto P(Y|X)P(X)$$

$P(X)$  represents prior probability of one region  $x_s$  and  $P(Y|X)$  is the conditional probability density function of  $y_s$  given the image region  $x_s$ .

### a) The prior probability

The prior probability is modeled by a Markov Random Field (MRFs). Local neighborhood between pixels is taken into account. For an  $s$  location of a pixel  $y$  the conditional probability satisfies the following equation:

$$p(X_x|X_r, s \notin r) = p(X_s|X_r, r \in N_s)$$

where  $N_s$  is a neighborhood of region  $s$  and  $r$  the neighbor region. Hammersley–Clifford theorem shows that the probability of  $X$  is given by a Gibbs distribution according to the formula a):

$$p(X) \propto \exp\left[-\sum_{c \in C} V_c(X)\right]$$

A subset  $c \subseteq Y$  is called a clique if every pair of pixels in this subset are neighbors and  $C$  denotes the set of cliques. For each clique  $c$  in the image, we can assign a value  $V_c(X)$  which is the clique potential of  $c$ . This clique potential is defined by equation a).

$$V_c(X) = \begin{cases} -\gamma, & \text{if } X_s = X_r \\ \gamma, & \text{otherwise.} \end{cases}$$

Neighboring pixels which have different classes will increase the energy, thus, lowering the prior probability. This should be more likely for neighboring pixels to have the same class.

**b) The conditional probability**

The conditional probability is also known as the likelihood. Here a Gaussian distribution  $P(Y|X)$  is defined by  $\hat{X} = (m_s, \sigma)$  the mean and standard deviation respectively. The aim is the estimate the unknown parameters that maximize the log-likelihood function shown by equation b).

$$X_{max} = \operatorname{argmax}_x \{P(x \in X | y \in Y)\}$$

To do so, we use Iterated Conditional Modes (ICM) algorithm. 40 iterations were enough for ICM to reach a stable segmentation. The ICM requires an initial segmentation. K-mean method was applied to initialize the ICM.

Apoptotic tumor area was calculated by multiplying the binary mask of total apoptotic cells by the binary mask of Nuclei. Endothelial apoptotic areas were calculated by multiplying the binary mask of endothelial cells by the mask of all apoptotic areas. Once all these results were obtained we generated our maps. Then, we calculated the percentage of endothelial apoptotic cells as follows:

$$A\% = \frac{ANA}{AA}$$

Where  $A\%$  is the percentage of endothelial apoptotic cell, ANA is apoptotic endothelial cell area and AA is the apoptotic area.

**III.4.4 Validation process**

The F-score[75] is used to measure the segmentation accuracy of masks obtained from our method: the apoptotic areas mask and the neovascularization mask. We consider a vascularization mask as true positive if it 50% area coincides with ground truth, otherwise it is considered false positive.

The difference between the number of ground truth objects and the number of true positives is the number of false negative. Equation III.4.4 defines F-score:

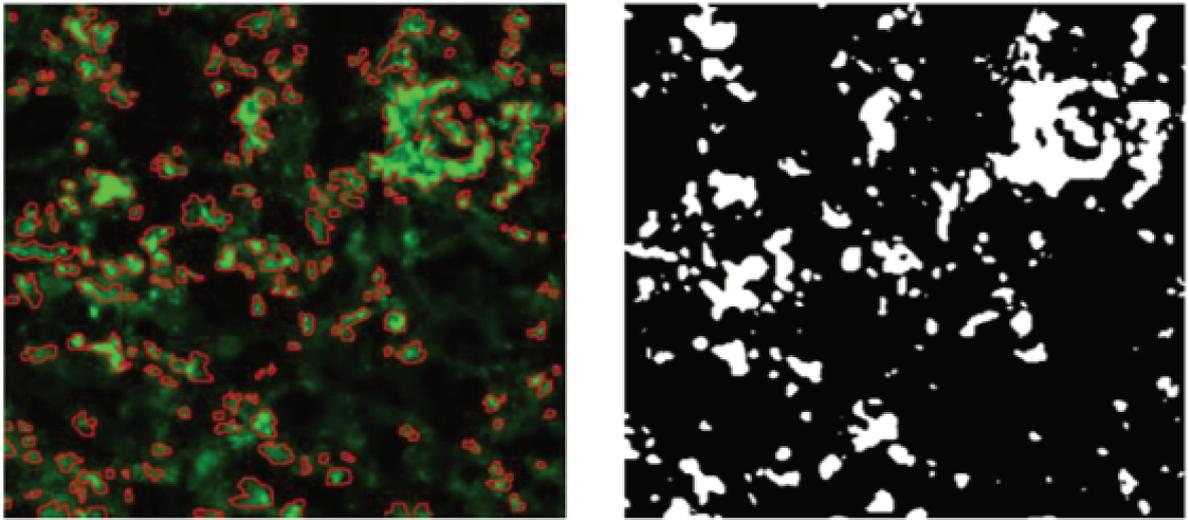
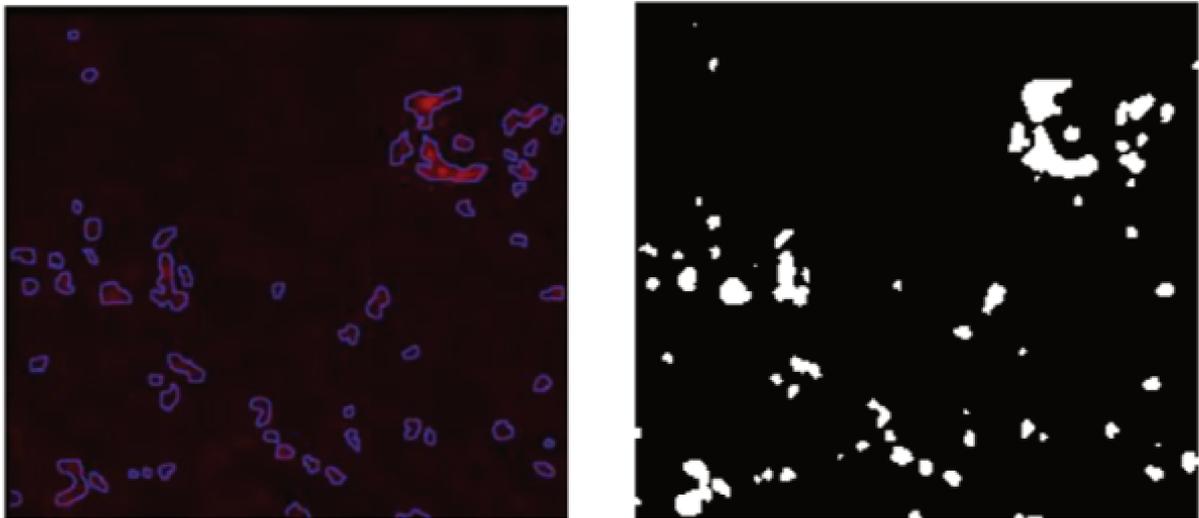


Fig. III.15 The comparison between the ground truth made by the biologist on the endothelial cells and the segmentation done by the MRFs segmentation



a) Ground truth of the apoptotic area

b) MRFs segmentation on the apoptotic area

Fig. III.16 The comparison between the ground truth made by the biologist on the apoptotic cells and the segmentation done by the MRFs segmentation

$$F_1score = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

TP, FP and FN denote respectively the number of true positives, false positives and false negatives from all tiles given by the database.

### III.4.5 Results

The percentage of apoptotic endothelial cells on the whole slice tumor relative to the total number of apoptotic cells exceeded 50% for all sections from the 4 mice treated with anti-angiogenic agent Pazopanib.

The highest percentage (up to 70%) was observed in tumor sections from AA1. The percentage of apoptotic endothelial cells in whole slices from mice receiving placebo did not exceed 40% as summarized in figure III.17.

The percentage of apoptotic cells on tumor area is mostly endothelial except for one aberrant value considering P2. Although, we still don't have a proof and this value needs to be more investigated, we believe that the fluors interaction can lead to some bias. Our finding supports Kadambi's work [76] showing that blocking vascular endothelial growth factor (VEGF) or its receptor (VEGFR2) leads to apoptosis of endothelial cells. The anti-angiogenesis agent Pazopanib was showing a direct effect on tumour network vasculature via the endothelial cells crossing the whole tumour.

Maps of the density of nuclei cells were calculated across each whole slide image. Overall, the density values were higher for placebo mice as compared to mice receiving Pazopanib III.18. The yellow color referring high nuclei density in the tumour sections was more frequently in tumour sections coming from placebo. Regional variability of the cell density can also be observed on these maps. The lowest density cells (blue color on the map) were detected in tumor slices presenting the highest apoptotic endothelial areas.

Because of the few numbers of our database (8 mice), we cannot conclude about a correlation between the apoptotic neovascularization and the nuclei density. Thus, we highlight the importance of working on more slices. However, our result shows a specific relationship between apoptotic

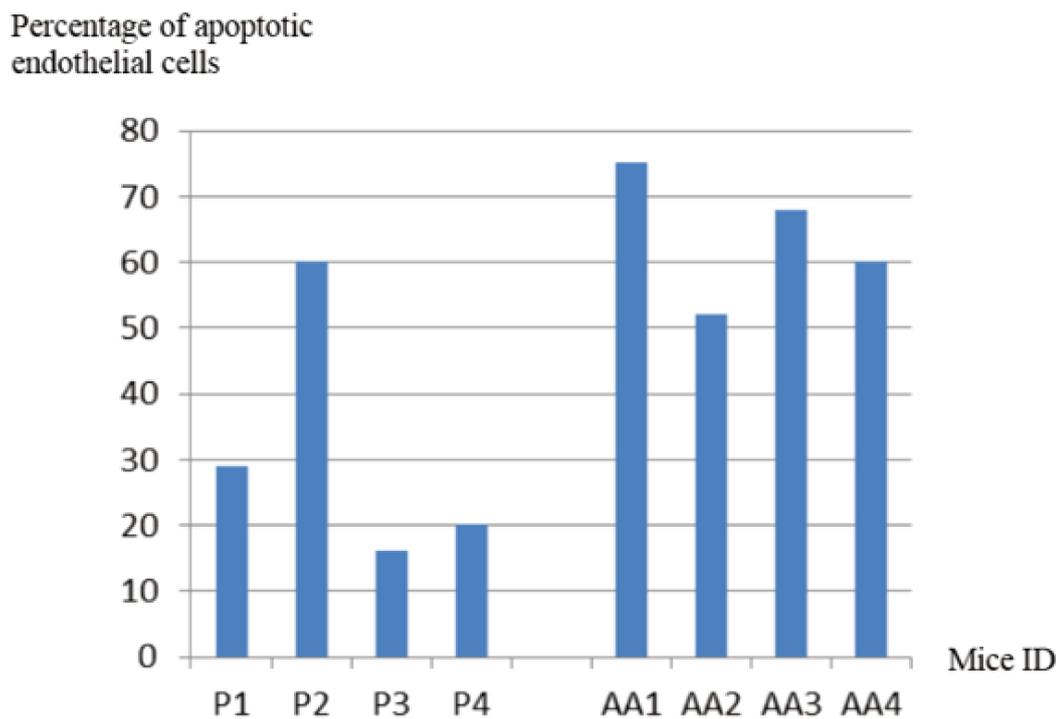


Fig. III.17 Percentage of apoptotic endothelial cells relative to the total apoptotic cells between placebo mice and mice treated by Pazopanib. P1, P2, P3 and P4 are mice receiving placebo. AA1, AA2, AA3 and AA4 are mice receiving Pazopanib.

neovascularization and nucleus density in murine tumor treated by Pazopanib anti-angiogenic which we need to investigate deeply for future work.

Results demonstrated that antiangiogenic therapy with Pazopanib generates apoptosis that is strongly associated with endothelial cells. Mapping of nuclei using a watershed segmentation method to separate different DAPI labeled cells on the tumor surface with an unsupervised method allowed us to establish homogeneous regions of cell density. Ultimately, this work will contribute to provide a high-resolution and full slide assessment of tumor modifications that can serve as reference measurements when validating non-invasive imaging techniques. Evaluation of our algorithm is shown in table III.4. The standard deviation was up to 0.003 which shows that our algorithm is stable.

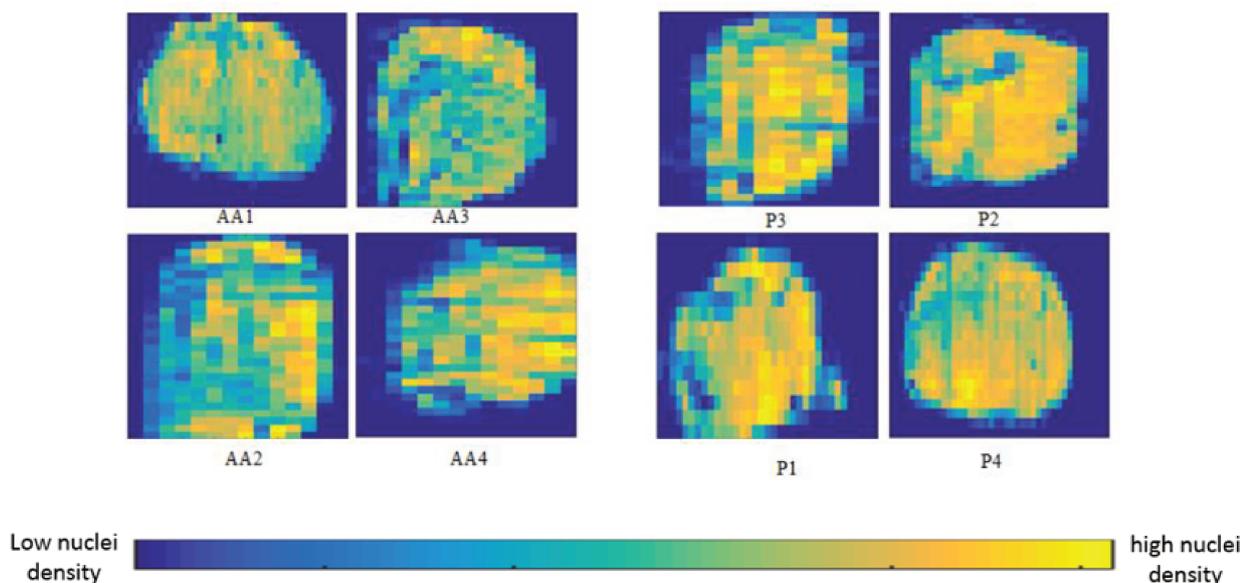


Fig. III.18 Maps of cell density are displayed for a whole slice segment of from each mouse. The maps labeled *P* are from mice receiving placebo and those labeled *AA* from mice receiving anti-angiogenic agent.

Mice ID	PRECISION	RECALL	ACCURACY	F-SCORE
P1	0.65 ±0.001	0.71 ±0.003	0.74 ±0.001	0.67±0.001
P2	0.71 ±0.003	0.72±0.001	0.69 ±0.003	0.71 ±0.002
P3	0.71 ±0.002	0.68 ±0.001	0.65 ±0.001	0.69 ±0.001
P4	0.69 ±0.001	0.72 ±0.002	0.73 ±0.002	0.70 ±0.002
AA1	0.68 ±0.002	0.69 ±0.003	0.71 ±0.001	0.69 ±0.001
AA2	0.71 ±0.001	0.72 ±0.001	0.70 ±0.001	0.71 ±0.001
AA3	0.66 ±0.001	0.71 ±0.001	0.69 ±0.002	0.67 ±0.001
AA4	0.77 ±0.001	0.65 ±0.003	0.65 ±0.002	0.70 ±0.003

Table III.4: Evaluation of the proposed algorithm.

### III.5 Conclusion

An experimental protocol as well as quantitative techniques were developed and applied to assess nuclei, apoptosis and endothelial cell distribution across whole-slice histological sections of tumors. Non-invasive imaging techniques were developed to probe the tumor environment during screening and therapy.

We first collected Multi-Fluorescence whole-slide images in a tumor model under anti-angiogenic Pazopanib therapy to develop and test techniques and quantitatively assess the cell density, apoptosis and endothelial cells. Watershed segmentation was applied to evaluate the number of cells

throughout the tumor sections. MRFs model was applied to resolve the endothelial cell and apoptotic areas segmentation problem. Stochastic image processing methods were used because of undefined structure of blood vessels and apoptotic cells in the image of our database.

During this first part of our research, we faced two challenges: First, because the Zeiss scanner does not provide information about the Z dimension, thick or folded regions of tumor sections were not fully interpretable. To minimize folding, sections were cut as thinly and carefully as possible. Secondly, tears in the sections occurred in some cases. The need to work with thin sections exacerbated the risk of tearing. The level of bias in quantification resulting from tears and folds remains to be assessed in future work but this bias is not anticipated to be very strong based on the relatively limited surface area involved with such artifacts in our sections.

Quantitative evaluation of tumor structures with histopathological precision across the full tumor cross section may ultimately contribute to the identification of architectural features that are often associated or predictive of therapeutic response. Data can also serve as valuable reference information as non-invasive imaging techniques. Finally, we emphasize the need to collect more data in order to conclude about a correlation between the apoptotic neovascularization caused by anti-angiogenic agent Pazopanib and the nuclei density and to improve the acquisition quality in order to reduce the fluors artifacts to minimize the bias in the measurements.

In order to complete the tumour angiogenesis assessment, we complement our study by a qualitative assessment using clinical data coming from a biopsy exam.



# Chapter IV

## Digital pathology for tumour angiogenesis assessment

### Abstract of the chapter

In this chapter, we design a Discriminative-Generative model to segment tumour blood vessels from high-scale heterogeneous (*H&E*) images. Our model is based on both data-driven model - Convolutional Neural Network (CNN) - and rule-based knowledge model - Marked Point Process (MPP). The main contribution of our algorithm is the ability to segment very heterogeneous images using a Neural network model based on very few annotated data, due to the knowledge coming from the MPP model (8 images of 512x512 pixels each). We describe the principle and the design of the model and we test it on histopathology slide and synthetic data.

### IV.1 Digital / Computational Pathology

Initial digital imaging framework in microscopy, Virtual Microscopy evolve, in the early 2000 (figure [IV.1](#)), into a technology framework termed "Digital Pathology" (DP), or more specifically, "Computational Pathology" (CP), including - a large panel digital technologies related to image

processing /analysis and to artificial intelligence. DP/CP represents the most important anatomic pathology revolution, with a strong impact in medical education, clinical medicine and medical research (see figure IV.2). This new technology, allows the acquisition, the management and the interpretation of digital information in histopathology [77, 78]. Besides, healthcare professionals can post, transfer and transmit digitized glass slide via computer networks [79]. Finally, the emergence of DP/CP and the high-resolution of histopathology images, allow studying meaningful correlations between these images and the OMICS (genomics, proteomics, transcriptomics, metabolomics) signature of relevant regions of interest (ROI) in the tissue, a bridge towards the Integrative Digital Pathology. A European Society of Digital and Integrative Pathology (ESDIP<sup>1</sup>) has recently been created to deal with these challenges, as to structure the sustainable support to European Congresses on Digital Pathology (ECDP<sup>2</sup>).

In Medical education, DP changed the teaching from analogical to numerical field as the pathological training was transformed from microscope-based to fully-digitized lessons where the number of trainers is no more limited by the number of microscope. Moreover, DP allows to standardize the teaching as the educational materials, including sets of digital images of microscopic specimens, are the same for all students and pathologists. This is extremely beneficial, especially in rare diseases, when the number of relevant biopsies is usually very limited [80].

For clinical medicine, DP makes the investigation of complex diseases (such as cancer) easier and accessible, as it offers the flexibility of manipulating workload allocations of cases. Indeed, DP allows fast cases transfer between different services and laboratories [81]. Besides, Interactions between healthcare professional becomes more fluid, as DP allows departments to proactively interact and share clinical repositories between each others. Indeed, as current medicine is evolving into a more and more dynamic multidisciplinary environment, pathologist needs, nowadays, to share diagnosis with other healthcare specialists such as clinicians, pathologists (second opinion is compulsory for some specific cancer cases for example), radiologists, biologists, surgeons, ... for a better assessment of complex diseases.

In medical research, anti-angiogenic trial requires a long term assessment on the outcome of patients. This assessment involves a large groups of patients interchanged between different labs

---

<sup>1</sup>European Society of Digital and Integrative Pathology: <https://digitalpathologysociety.org/>

<sup>2</sup>European Congress on Digital Pathology: <https://digitalpathologysociety.org/events/>

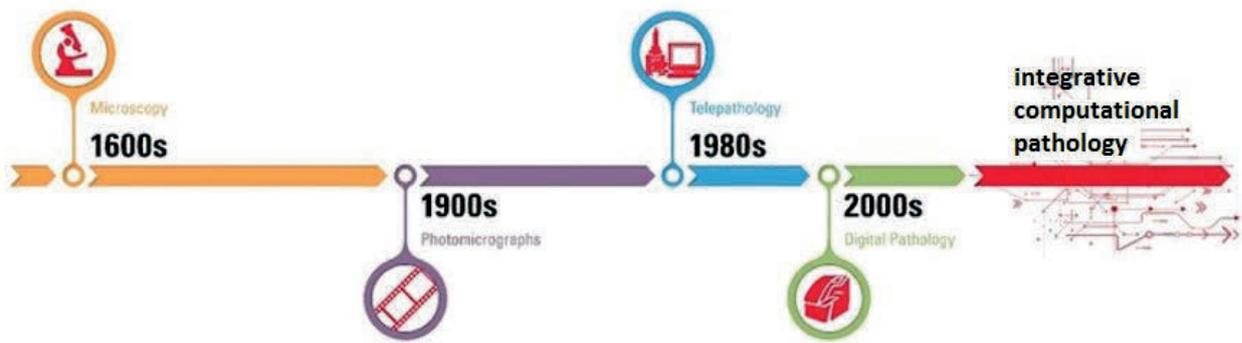


Fig. IV.1 *Illustration of histopathology over time. Emergence of Digital Pathology in early 2000s. Evolution towards Integrative Computational Pathology.*

and platforms. In this case, DP offers the possibility of standardizing quantitative measurement of immunohistochemical samples by applying standardized automated analysis of digital images. On the other hand, this kind of trails often needs pathologist's annotation which is practical when the specimen is digitized [82].

## IV.2 Whole slide imaging

The first Whole slide imaging (WSI) scanners were introduced in the late 1990s [83]. Prior to this date, digital images in histopathology was just limited to the static acquisition by using a microscope-mounted camera which limited research and clinical utility as it captured only specific regions on the stain[84]. WSI, as illustrated in figure IV.3, capture the different regions of the stain, with different resolutions, in a pyramidal pathway, with respect to supplement 145 (Whole Slide Microscopic Image IOD and SOP Classes) of the Digital Imaging and Communications in Medicine (DICOM).

## IV.3 Histopathology

Histopathology is defined as the examination of a tissue specimen under a microscope by a pathologist, in order to assess the manifestation of the disease [85]. In the case of observing physiological cancer symptoms, for example, the histopathological examination is prescribed for a patient, in order to allow the observation and determination of cancer malignancy after a tissue sample is removed

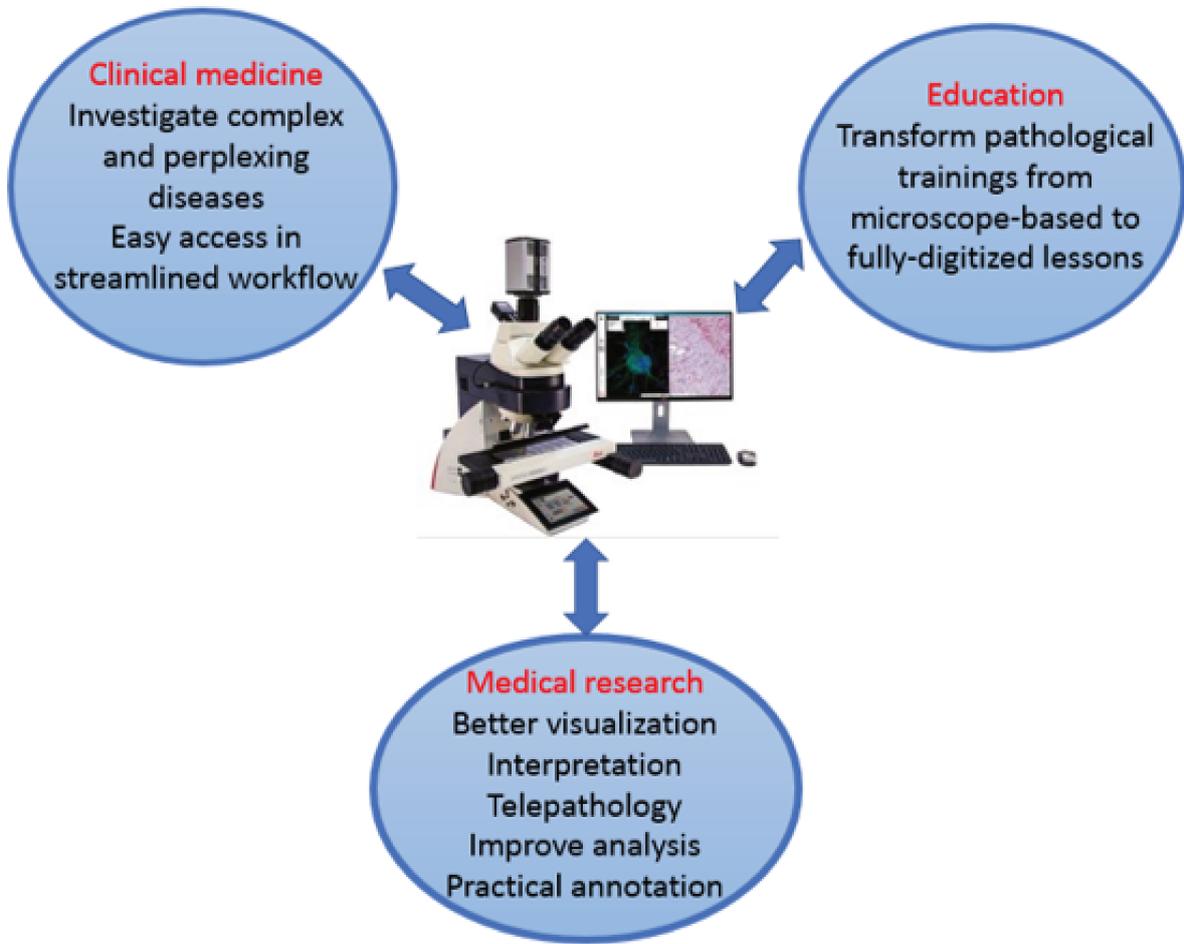


Fig. IV.2 Illustration of digital pathology impact in clinical medicine, education and medical research fields.



Fig. IV.3 Whole slide imaging (WSI) pathway. The histological specimen is scanned in the WSI work station which can be manipulated in a pyramidal way by defining the different resolutions.

from the suspected organ by biopsy or surgical resection. With the microtome, histopathology lab operators can create very thin specimen slices to be stained and observed under a microscope.

This specimen undergoes a preparation procedure after which a pathologist grades the malignancy of cancers by observing the architecture of the tissue, the distribution and the morphology of the cell, in very fine details. This procedure allows the pathologists to diagnosis the severity of the disease as it delivers rich visual content at the cellular level, featuring disease characteristics that are not observable in other medical imaging modalities. However, histopathology exam requires a very highly skilled expertise, as this exam is very important to decide about the treatment strategy and therefore the patient prognostic. Besides, this exam is very time consuming and tedious due to the biological complexity and heterogeneity of the human tissue. Histopathology is the golden standard to detect, interpret and diagnose the disease.

## IV.4 Hematoxyline and Eosine staining

In histopathology, staining refers to the fact of coloring a specimen to study the structure of its components which allows a tissue based diagnosis. The most common staining used in pathologist's daily routine is hematoxyline and eosine (*H&E*). In contrast to other staining method, besides its accessible cost, (*H&E*) staining allows an extensive tissue morphology study, together with the detection of particular cell types as of the tissue structure and the interaction between its components, which makes this staining termed as "routine staining" in histopathology [86].

The *H&E* stain uses a combination of two dyes: hematoxylin and eosin which stain different tissue elements. Hematoxylin reacts like a basic dye, therefore, it stains acidic or basophilic structures such as cell nuclei as it contains DNA and nucleoprotein, endoplasmic reticulum ribosomes as it contain RNA. On the other hand, eosin reacts like an acidic dye, therefore, it stains basic or acidophilic structures such are cytoplasm, cell membranes and extracellular fibers. Hematoxylin dye appears with purplish blue color in histopathology images however eosin dye is typically reddish or pink [87].

## IV.5 Computer aided diagnosis and pattern recognition in Digital pathology

Looking at the characteristic features of the disease using the biopsy remains the gold standard in diagnostics. Indeed, through this process, histopathologists try to find the complex patterns characterizing the disease as suggested by the grading systems. Due to the advent of computer vision and pattern recognition, active research tried to develop and improve automated systems to detect these patterns in order to help histopathologists with this challenging and time-consuming decision [88, 89, 90].

Analysis of histopathological images by pattern recognition poses a challenging computer vision and artificial intelligence problem due to different dye concentration, uneven cuts, image acquisition artifacts, dye artifacts and the complexity of the tissue shapes of the tumour microenvironment. Automated image processing tools need to overcome these challenges to complement the work done by the pathologist in a very accurate way, as this task directly concerns patients' prognosis.

Many metrics have been developed to characterize nuclei and glands in tumour specimens as many features reflect the presence of cancer condition. To our knowledge, no metric has been developed for features related to blood vessel assessment (detection, segmentation, quantification) in whole slide images (WSI).

The complexity of the morphology and the environment of the blood vessels, as the high-resolution of the WSI used in DP makes this task very difficult to handle in a robust way, automatically.

Deep neural networks (DNN) represent, nowadays, the most effective machine learning technology in biomedical domain, concerning areas of interest related to OMICS, BioImaging, Medical Imaging, Brain and Body Machine Interface, as the Public and Medical Health Management [91]. Within the DNN family, the Convolutional Neural Networks (CNNs) were inspired by the neurobiological model of the visual cortex, where the cells are sensitive to small regions of the visual field [92, 93, 94].

Recent successes in DP challenges (i.e. Camelyon<sup>3</sup>) highlight the relevance of mixing DL archi-

---

<sup>3</sup>Camelyon challenge - breast cancer metastases in whole-slide images of histological lymph node sections: <https://www.camelyon.org/>

tures with Bayesian models. In this context, in the next section, we give the background of the discriminative – Deep Neural Networks (DNN) – and generative – Marked Point Process (MPP) – pattern recognition approaches inspiring our blood vessels assessment pipeline. Indeed, the combination of these families of pattern recognition methods presents an interesting complementary advantage of probabilistic modelling of the blood vessels environment characteristics (MPP) and the efficiency of the Deep Learning (DL) methods in a precise, framed context.

### IV.5.1 Marked Point Process

#### a) Definition

In pattern recognition, the conditional probabilistic approach corresponds to estimating a variable  $X$  knowing an observable variable  $Y$ . We define an energy model in variables space or functions and optimize it such that the variable  $X$  represents the optimal guess. The energy model is usually composed of two terms: an internal term and an external term. The internal term is a regularization term which imposes conditions on  $X$  according to the desired structure. The external term, is the data fidelity term or the likelihood term. We aim at estimating the variable  $X$  with a defined cost function  $L$  where  $\hat{X}$  is the estimator as follows:

$$\hat{X} = \text{Argmin}[E_{intern}(X) + E_{extern}(X, Y)]$$

$E_{intern}$  represents the internal term and  $E_{extern}$  is the external term. Image segmentation can be simulated as probabilistic approaches, where we aim to assess a pixel  $i$  defined by its features (generally pixels) to a class. Many probabilistic approaches have been introduced in image segmentation and pattern recognition. One of the most widely used was Marked Point Process (MPP) which was the main inspiration behind our approach.

#### b) Algorithm

As MPP is an object recognition algorithm which has to decide if there is any object of specificity in a scene and determine its number and characteristic (shapes, size, location, spatial relationship),

it takes into consideration the geometry of the random set points (line, circle, ellipse, star etc.).

Let's  $\Omega = \{\omega_1, \omega_2, \omega_3, \dots, \omega_n\}$  be a set of random objects, and  $\Omega_i$  all the possible realization of  $\omega$ . A marked point  $\omega_i$  is an object in  $\Omega$  described by its position  $x_i$  and its mark  $m_i$ . The mark  $m_i$  represents the geometric definition of the object which can be a simple shape (line, circle, ellipse) or a more complicated one (star-shape, sphere, ellipsoid).

Let's denote  $\Omega$  a set of random marked points in  $2D$  dimension respecting a Gibbs distribution. Each Marked Point is defined by  $x_i$  position and  $m_i$  mark where  $i \in \{1, 2, 3, 4, 5, \dots\}$ . The object we aim to detect is a circle where  $x_i$  is the circle center and  $r_i$  is a circle radius. Thus, the mark of the object  $m_i = (x_i, r_i)$  is following a Gibbs distribution:

$$P(w) = \frac{1}{Z_\beta} \exp(-\beta U(w))$$

where  $U$  is the energy model and  $\beta = \frac{1}{T}$ , where  $T$  is the temperature of the energy model.

This temperature term defines the sensitivity of the probability of the different energy and  $z_\beta = \sum_{w \in W} \exp(-\beta U(w))$  is a normalization constant. Such a problem represents an optimization problem, in which we need to find  $w$  having the highest probability. This target is reached by minimizing the energy model  $U$ , defined as follows:

$$U = \gamma d \sum_i U_d(w_i) + \gamma p \sum_{i \neq j} U_p(w_i, w_j)$$

MPP algorithm is a rule based knowledge driven model based on the definition of a cost function. Indeed, the assumption behind MPP algorithm is to detect the target object on the image by predicting its prior shape within the minimization of a cost function. Thus, the algorithm sample the process in order to extract the configuration of objects minimizing the energy through an optimization method usually a dynamic of multiple births and deaths process or Monte Carlo optimizer (MCMC). This optimizer decides if the simulated prior object in the configuration is fitting the target object in the image, we keep it, otherwise we delete it and we test a new created (set of) object(s).

Eventually, this dynamic lead to an estimation of the total number of target object present on

the image. The density associated with the MPP is defined with respect to a Poisson measure. The cost function consists of a regularization term, which introduces constraints on the objects and their interactions and a second term of attachment to the data termed data fidelity, which makes possible the location of the object to be detected on the image. The cost function is defined as follows:

$$U(w) = \sum_i U_d(w_i) + \sum_{i \neq j} U_p(w_i, w_j)$$

Where  $U_p$  is the interaction term controlling the overlapping between detected object  $w_i$  and its neighbor  $w_j$  of the configuration  $w$  and  $U_d$  is the data fidelity term which measures the relevance between the configuration  $w$  and the image. The aim of the data fidelity is to evaluate that the object is correctly placed. Indeed, an object  $w_i$  correctly placed should give a low energy value for  $U_d$ . In previous works, the metric defining the data fidelity term was configured according to the characteristic of the structure of the object that ones aim to detect. For example, in [95], to detect nuclei in *H&E* slides, Avenel et al. evaluated the data fidelity term using to the Bhattacharya distance which quantifies the similarity between two probability distributions. Therefore it measures the amount of the relevance between the configuration  $w$  and the nuclei in the *H&E* image [96]. The authors make use of this metric because the equation used in the Bhattacharya distance is highly adaptable with the characteristic of the nuclei in the image. Indeed, the nuclei structure shows a good contrast between the inside of the nuclei and its borders. Therefore, to correctly detect the nuclei, the authors define an inside region which is an ellipse corresponding to the nuclei shape and a thick border (20 pixels) corresponding to the border of the nuclei. They call, respectively the two regions: inside and outside region and they evaluate data fidelity by computing the Bhattacharya distance metric given by the formula b):

$$db(D_{in}, D_{out}) = \frac{(\mu_{in} - \mu_{out})^2}{4\sqrt{\sigma_{in}^2 + \sigma_{out}^2}} - \frac{1}{2} \log \frac{2\sigma_{in}\sigma_{out}}{\sigma_{in}^2 + \sigma_{out}^2}$$

Where  $D_{in}$  and  $D_{out}$  are respectively the inside and outside region and  $\mu_{in}$ ,  $\mu_{out}$ ,  $\sigma_{in}$  and  $\sigma_{out}$  are respectively the mean and the standard deviation of the inside and outside defined region distribution. To compute the Bhattacharyya distance, good contrast between the outside region

$D_{out}$  and the inside region  $D_{in}$  of the object is highly recommended.

### c) **Marked Point Process (MPP) versus Markov Random Field (MRFs)**

In the last decade, research has been shown that optimization models are a useful way to resolve recognition and segmentation problem in images. The main advantage of MRFs models is that prior information can be imposed locally [74]. Although MRFs model tends to preserve well the arbitrary boundary information, it fails to consider the geometric property of the object contradictory to MPP model which incorporates the geometric constraint. However, the main drawback of MPP is that this algorithm is not able to describe the boundary information if shape does not fit a geometric model exactly which is not the case of the blood vessels in histopathologist slides.

### d) **Sampling and optimization strategies in MPP algorithm**

A numerical image can be considered as Markov chain as it satisfies the two conditions:

- Positivity: as all image pixels have a positive value.
- Neighbourhood: image acquisition creates a superposition in neighbours pixels values creating thus a spacial dependency that can be modelled by clic in Markov Chain.

Therefore, two methods have been reported in the state of the art to optimize the energy model in MPP algorithm.

**d).1 Markov Chain Monte Carlo (MCMC)** is a well-known method for obtaining information about distributions, especially for estimating posterior distribution in Bayesian interface [97]. It allows one to characterize a distribution without knowing all of it's mathematical properties by randomly taking samples out of it. A particular strength of MCMC is that it can be used to draw samples from distributions even when all that is known about distribution is how to calculate the density for different samples. This property makes it so useful as an optimizer in MPP framework [98]. MCMC methods have been around for almost as long as Monte Carlo techniques, even though their impact on Statistics has not been realized until the early 1990s.

**d).2 Birth and death process** is a stochastic process. This process is a homogeneous, irreducible and continuous time Markov process. In MPP, state represents a current size of a population and the transition is limited to a birth or death. This type of counting process is a simulated annealing algorithm which allows to find the optimal solution  $w_i$  to a given function  $f(w_i)$ .

Assume  $n$  is a population number at  $t=i$ . Two possibilities may occur at  $t=i+1$ :

- Birth step: the population increase +1
- Death step: the population decrease -1

We calculate the probability of the two transitions as follows:

$$Pr\{N_{i+1} = n + 1 | N_i = n\} = \frac{\lambda(n)}{\lambda(n) + \mu(n)}$$

$$Pr\{N_{i+1} = n - 1 | N_i = n\} = \frac{\mu(n)}{\lambda(n) + \mu(n)}$$

The birth and death transition are independent from each other. The state called birth is when the variables population of a set  $\Omega$  increase by one so that the process goes from state  $n$  to state  $n + 1$ . However State called death is when the variables population in the same set  $\Omega$  decrease by one. So the process goes from state  $n$  to state  $n - 1$ . Although the two optimization methods stated above showed good results in segmentation's using MPP algorithm, they remain a heavy process with a big computational time.

### e) Biomedical Imaging application in histopathology and microscopy data

The strength and versatility of the MPP framework is demonstrated by the wide range of image analysis tasks and application domains in which it has been incorporated. Comprehensive coverage of early works and foundations of the methods can be found in the books [99, 100]. The MPP framework can be disassociated for simplicity of tracing its evolution and applications into the simulation mechanism and the energy function design components. The adaptation of the energy functional is necessary for its application to various tasks depending on the patterns of interest and data modality. The simulation and optimization of the energy function are strongly supported by a mathematical framework which too is an active area of research. Majority of MPP applications

begin with approximating the objects to be detected with a data energy modeling that should be adaptable to the data type. In the following sections we will review several applications from Biomedical Imaging application domain in histopathology and microscopy data.

Early biomedical models include coronary tree extraction [101]. Vasculature networks modeling is a frequently encountered problem presenting a tree hierarchy relations among repeated units of the model for which MPP framework offers an attractive unsupervised solution [102]. Other similar applications are found in quantification of neuron morphology [103, 104, 105]. The MPP framework is versatile for efficient adaptation to a varied template detection problems (automatic dendrite spines detection, vascular system detection) applied to images from multiple modalities, ranging from x-ray tomographies to 2-Photon microscopy images [106, 102].

Automatic dendrite spines detection from x-ray tomography volumes [106] it has been adopted in unsupervised vascular and neural network modeling [102].

#### f) Challenges of MPP framework

The two main struggles in adapting the MPP framework for a new application in spite of its versatility is in the estimation of its several categories of parameters and then the long time to convergence for the MCMC simulation. Besides, the amount of various parameters that should be fixed empirically to make the algorithm stable and ensure the convergence. The various parameters of an MPP model can be organized into 3 categories: Parameters of objects corresponding to the descriptive marks of the chosen objects are often application and data specific. Second, parameters of Energy Model, are calibrated – in the literature – according to the model and the application data. Finally, the parameters of Simulation which are the most critical parameters and the sensitivity of the model to these parameters are very noticeable. Unless the birth intensity is initialized optimally, there is little possibility of reaching close to the global optimum. It is generally set as an over estimation of the expected number of objects. The critical parameters require initialization close to the optimum. Otherwise, the simulation has no means of converging near the global optimum. A general framework for a fully automatic approach for parameter estimation is, nevertheless, an open and challenging problem.

### IV.5.2 Neural Network and deep learning

Deep learning (DL) is an ensemble of learning based algorithms. It allows a data representations based on an intensity values of pixels in images [107]. In medical imaging, this family of algorithm has dramatically improved the state-of-the art, especially in histopathology field [108, 109, 110, 111]. DL provides various algorithms for supervised and non supervised feature learning [112, 113]. These efficient algorithms replace handcrafted features, as they provide an accurate feature extraction tool.

A Deep Neural Network (DNN) is composed of different layers connected to each other, each layer being composed of different nodes. Each node is having its own weights and bias and their interactions provide more extraction and composition of the image features. Therefore, in a DNN, the number of layers, the size of each layer and the interaction between layers enables having different type of features extraction.

With the phenomenal increase of the use of Deep learning for medical image analysis and quantification, the research and the engineering have been more focused on developing new architecture to learn the desired representations for a given task. in this area, many DNN architectures have been used in medical imaging such that:

- Convolutional Neural Network (CNN) [114]
- Recurrent Neural Network (RNN) [115]
- Deep belief Neural Networks (DBN) [116]

#### a) Convolutional Neural Network (CNN)

The Convolutional Neural Network (CNN) is one of the widely used Neural Network. It takes advantage of the fact that the input is an image which the pixels go through an ensemble of non-linear computation to exact features according to a certain task. Many architectures have been reported in state of art, the used architecture being related to the task we aim to realize. For example, a simple CNN for binary classification can have the following architecture: an ensemble of CONV layers followed by a RELU layer followed by POOL layer and finally a Fully Connected layer (FC).

CONV layers are composed of different nodes. Each node has its weights and biases. The CONV layer will compute the output of nodes that are connected to local regions in the input. Assume that the input image has the shape of  $512 \times 512 \times 3$  as an RGB image, and the kernel size is  $1 \times 1$  pixels, thus the CONV layer computes the output of the dot product between the input image and the kernel. If we decide to use 12 kernels (filters) and a stride = 1, the output of the CONV layer is  $512 \times 512 \times 12$ .

The RELU layer will apply an element-wise activation function, such as the  $\max(0, x)$  threshold. The unchanged volume will have dimensions  $512 \times 512 \times 12$ .

The POOL layer is used to perform a down-sampling operation along the spatial dimension, the output volume will be  $128 \times 128 \times 12$ .

FC result in a volume of  $1 \times 1 \times 2$  (2 refers to the number of classes, as our example is a binary classification, the number of classes is equal to two). The output of this layer is a class score, for each class the FC computes the prediction score referring the among of each class according to the number of our dataset.

The aim of this architecture is to transform each pixel in the input image into a final class score. The weights and biases are going to be trained with an optimization algorithm which could be Adam or stochastic gradient descend (SGD), through a back-propagation process.

In the CNN the cost function will control the optimization process as it compute the difference between the CNN prediction and the ground truth given during the training process.

## b) Fully Convolutional Network (FCN)

The fully convolutional network (FCN) or UNET is an NN made only by CONV layers as it is specific for image segmentation. The output of the neural network is a segmented image with the same size as the input image. This architecture shows good accuracy in segmentation task for nuclei and gland in histopathology images on *H&E* whole slide images. FCN can take different forms (with or without shortcut, number of layers in the down-sampling and up-sampling) but it is usually split into two parts: first part in down-sampling when we go from simple to very deep semantic presentation of the input image and we call usually this part of the neural network: the

# Network Architecture

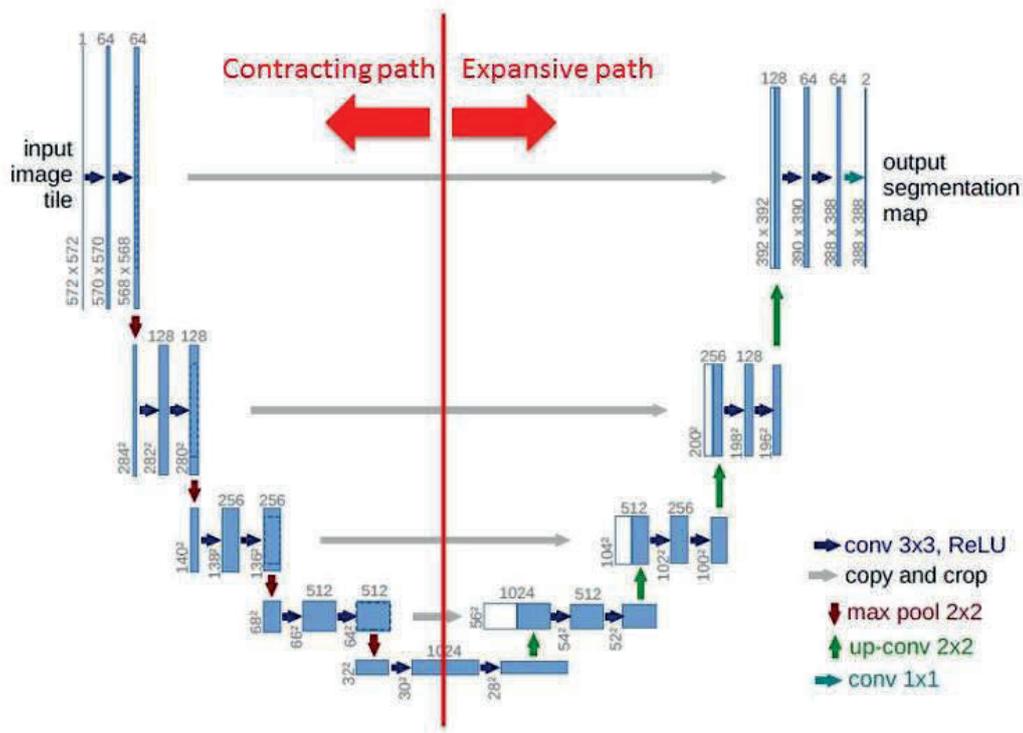


Fig. IV.4 The UNET architecture representing the two paths: the contracting path and the expansive path. The contracting path is to go from a simple pixels representation to a deep semantic representation. The expansive path is to end-up with a segmented image having the same dimension as the input image

contracting path, and a second part of up-sampling to end with a segmented image with the same size as the input image and we call this part of the neural network the expansive path. A shortcut is computed between the down layer and corresponding up layer to improve the back-propagation and resolve the vanishing gradient problem when the neural network is very deep. Figure IV.4 demonstrate the UNET architecture as an example of FCN.

## IV.6 Data description

Our data has been collected from Switzerland as a part of our collaboration with the Swiss Institute of Bioinformatics in Lausanne. The *H&E* data used is a subset of the data from Moor et al. [117]. Overall, there were two standard pathology haematoxylin-eosin-stained slides with breast tissue

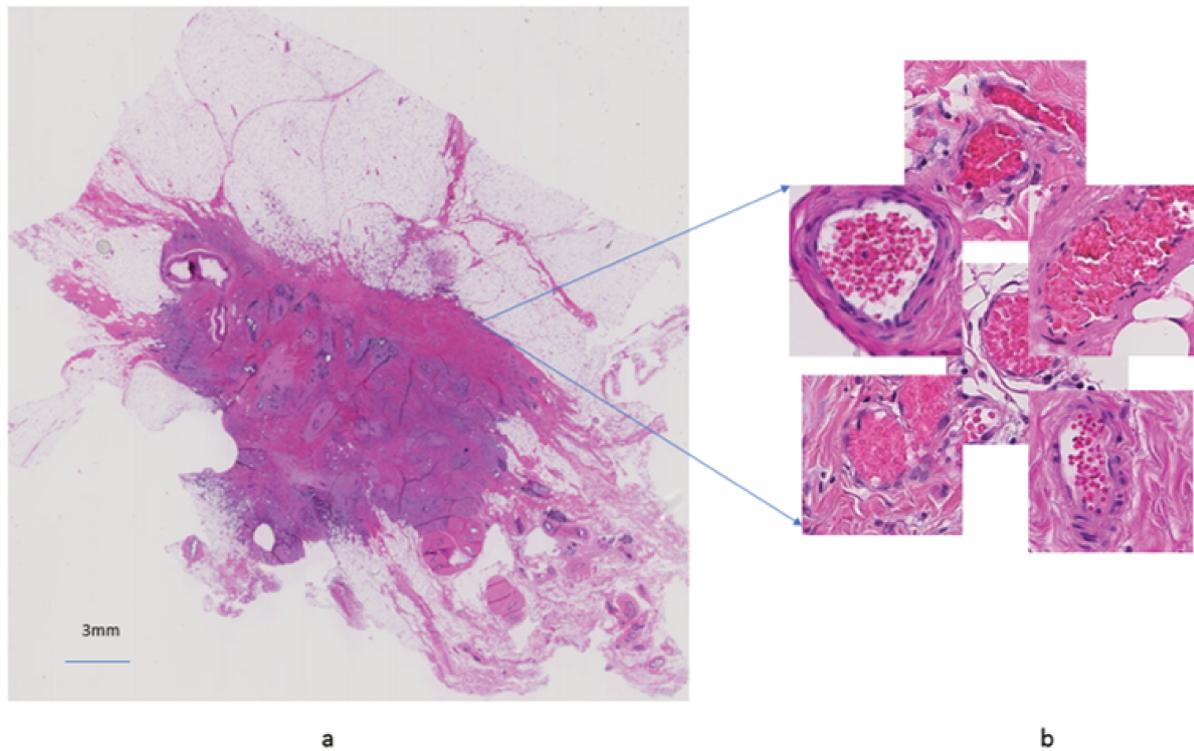


Fig. IV.5 Example of tissue with breast carcinoma grade 2: (a) whole slide imaging 150,000 x 100,000 pixels (b) cropped region of blood vessels 512x 512 with a resolution value of 0.5.

sections, containing a tumoural component from different cases. Images were obtained by whole-slide scanning of the pathology slides at 40x magnification, resulting in RGB images of about 150,000 x 100,000 pixels. The population of this data undercame a case-control. Patients with early breast cancer treated at the Department of Medical Oncology, Inselspital Bern, Switzerland.

All patient received adjuvant endocrine therapy. After biopsy, 10 microns sections were cut from cancer tissue of each patient. From each whole slide image, we crop 512x512 RGB region of interest (ROI) of blood vessels, as shown in figure IV.5.

We noticed that the blood vessels represent very different and complicated shapes which can be circles, ellipses or for more complex cases none defined shapes. besides the density of the blood cells, the internal part of the blood vessel can be different from a ROI to another (high, low or medium density) as shown in figure IV.6. Therefore, the implementation of an automatic framework to segment the blood vessel on this dataset is a very challenging task.

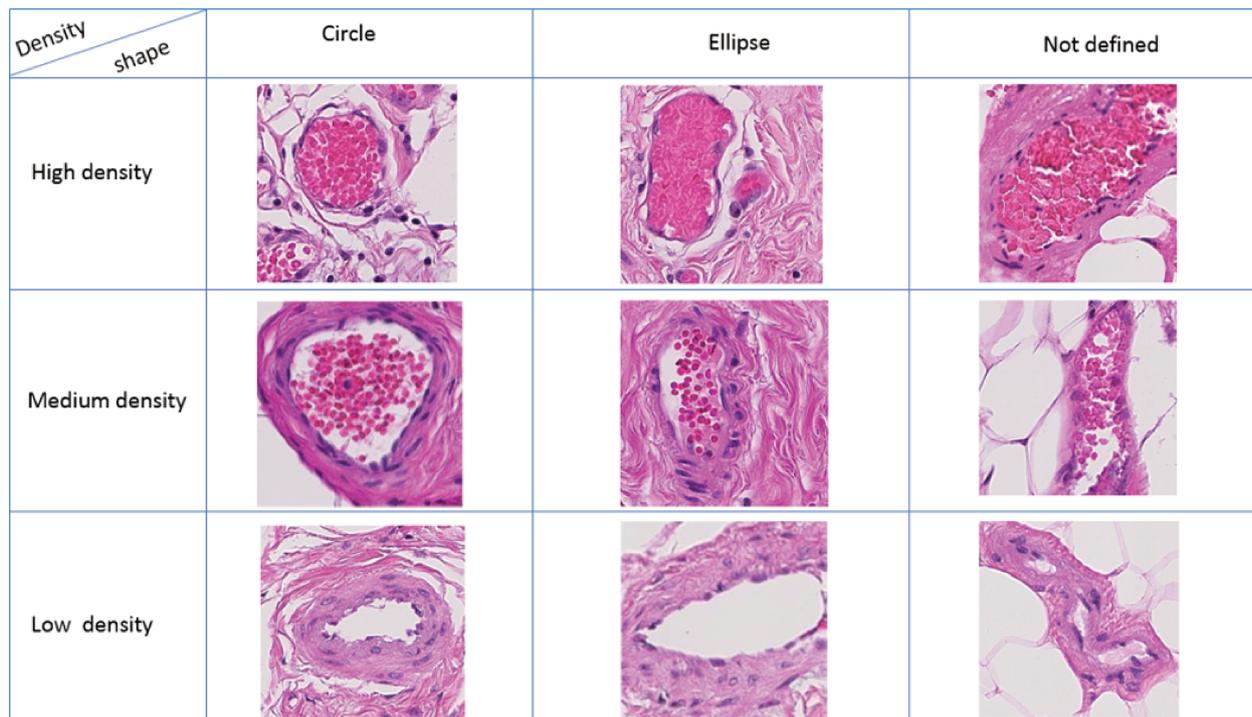


Fig. IV.6 The blood vessels have different shapes with different red blood cells density.

### IV.6.1 Challenges

The baseline developed in chapter III allowed us to assess quantitatively the tumour angiogenesis process on preclinical specimen of fluorescence images. However, for complex data including more heterogeneous information, such as the data detailed in figure IV.6, our initial algorithm is not accurate-enough. Indeed, our previous baseline is a pure rule-based segmentation algorithm which anticipates relatively uniform and distinguish colors or textures within the whole image. This assumption leads to over-segmentation, as shown in figure IV.7, as the rule defined is not adoptable to more challenging data such as *H&E*.

Although, the existence of non-uniform staining of the specimen surface leading to different intensity and color can be technically corrected with a pre-processing algorithm such as color normalization [118], The image heterogeneity remains a challenging problem, reducing the accuracy of the object segmentation. Indeed, our *H&E* data is very heterogeneous, including information not only from blood vessels but also from nuclei (in purple), adipocyte or empty regions (in white) and collagen (in pink). The pixels intensity distribution is very variable from a ROI to an other, as detailed in IV.8 The texture of the blood vessel is similar to collagen area. Therefore, blood vessel

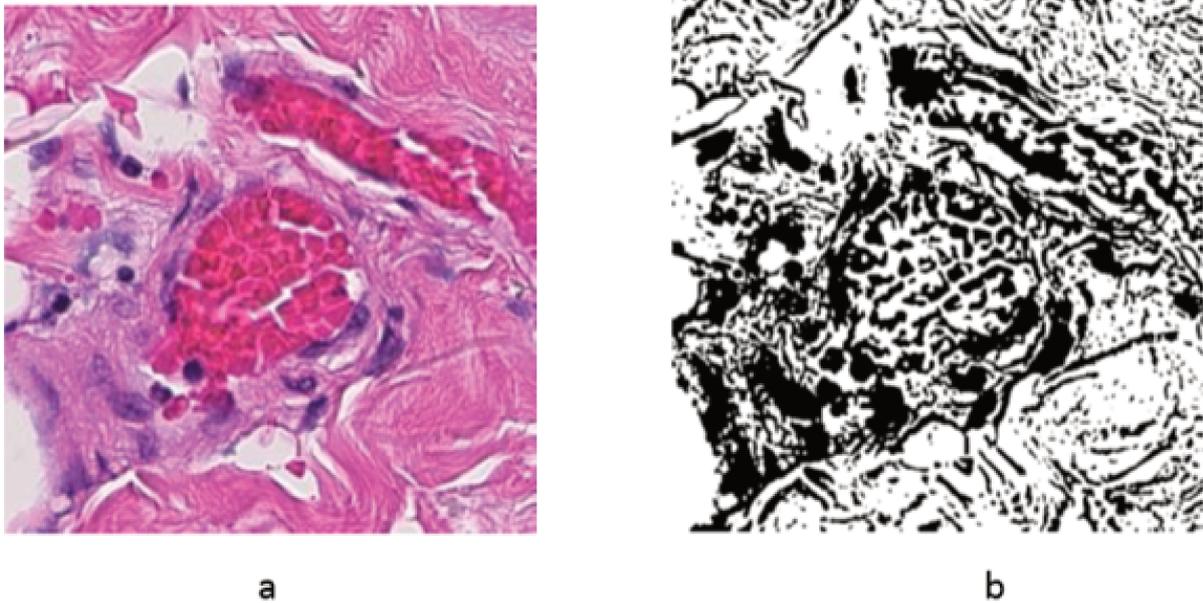


Fig. IV.7 MRFs on H&E whole slide images. (a) input 512x512 RGB images of the blood vessel (b) the result of the segmentation using MRFs baseline

segmentation is a very challenging task; especially from heterogeneous and complex microenvironment, as it's usually the case in histology.

For the other components of the microenvironment such as nuclei, many others discriminative rule-based segmentation algorithms such as marker-controlled watershed segmentation [119],

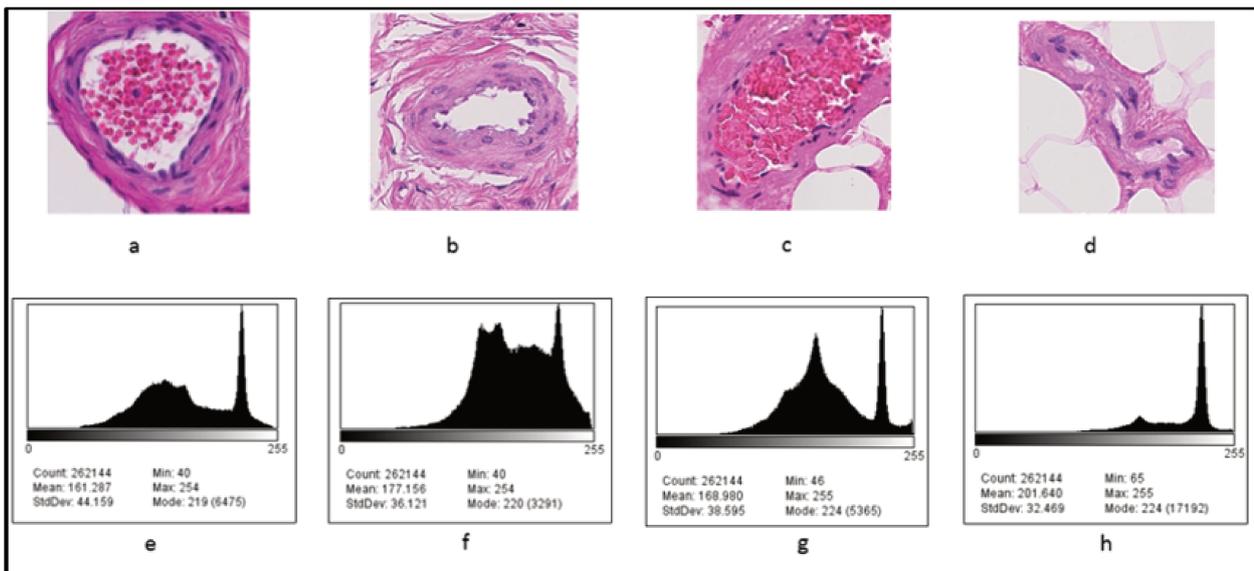


Fig. IV.8 MRFs on HE whole slide images. (a) input 512x512 RGB images of the blood vessel (b) the result of the segmentation using MRFs baseline

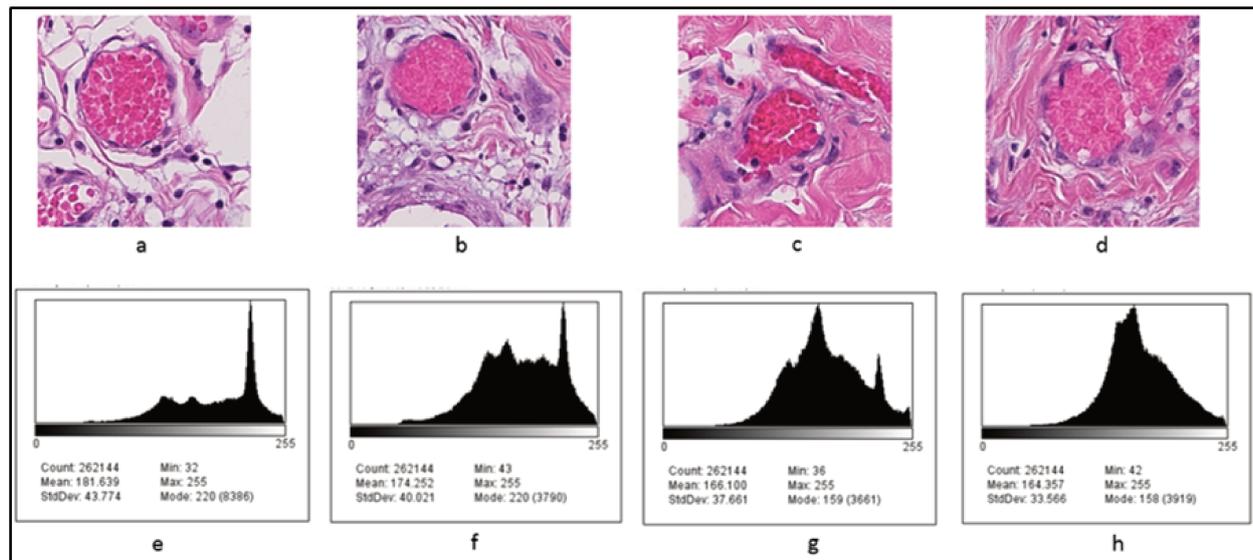


Fig. IV.9 (a),(b),(c) and (d) represent 4 cropped regions containing blood vessels from the whole slide images of breast carcinoma of the same patient. Each crop is 512x512 pixels RGB image. The blood vessel that we aim to segment is the big pink circle in the image. (e), (f), (g) and (h) represent respectively its intensity distribution

Ref	Organ	year	Stain	Method	Accuracy
[122]	skin	2017	dermoscopy	K-means clustering	96.5
[123]	liver	2015	dermoscopy	Morphological reconstruction	0.94
[124]	lung	2015	dermoscopy	Automatic Immunostained Vessel Image Filtering Algorithm	Not reported
[125]	colon	2015	IHC	Clustering	Not reported

Table IV.1: A summary of state-of-art of blood vessel segmentation using pure rule-based knowledge. No study has been reported according to blood vessel segmentation using pure rule-based algorithm in *H&E* stain.

Contour-based “minimum-model” [120] and MPP [121] have been widely used. However, To our knowledge, no rule-based segmentation model has been developed for blood vessel segmentation in *H&E* slides. The table IV.1 gives a state-of-art of the rule-based method used to assess angiogenesis but those studies are limited to dermoscopy stain or IHC stain. No *H&E* has been reported. This can be due to the complexity of the data which make the definition of the rule in the model a very challenging task.

Recently, several papers have showed that energy based rule model such as stochastic Marked Point Processes (MPP) are a promising tool for segmentation in digital pathology (see section III.5.1 for details of state of the art) In Kulikova and Avenel’s work, MPP model was used to guide the decision of pathologists in the histological grading of breast carcinoma by detecting automatically

the morphology of cell nuclei in *H&E* whole slides [121, 95]. The well-defined nuclei borders with the high contrast between the center and the border region of the nuclei make the use of Bhattacharyya distance to control the data fidelity term, a very successful metric on which the data fidelity energy loss is based.

In [126], Basu developed a pipeline using MPP to extract connected neuronal networks by fitting special configuration of marked objects to the centreline of the neurite branches. Although, Basu used an MPP model to achieve the segmentation, she emphasized the need to define a new energy model as the Bhattacharyya metric failed in the 3D segmentation task especially when the image is heterogeneous without well-defined border.

To our knowledge, no previous work has been published to test the MPP model in blood vessel segmentation in digital pathology. In the works mentioned above, the shape is usually an ellipse/ellipsoid as the nuclei has the similar appearance to ellipses. Unfortunately, this condition is missing in our dataset. Although we are using the same type of data as in Avenel and Kulikova works (*H&E* slides), the Bhattacharyya distance is not adaptable to our segmentation task because of the specific structure of the blood vessel. Indeed, as described in figure IV.6, in addition to the heterogeneity problem of the *H&E*, blood vessels have very complexes shapes. The border of the blood vessel can be composed of collagen layers including nuclei (figure IV.9.c), a mix of nuclei, blood cells and collagen tissue (figure IV.9.d) or, moreover, a mix of fat cells and nuclei such as figure IV.9.a. The structure, thus, is not homogeneous in the border of blood vessels to make use of Avenel's and Kulikova's methodology, in our case.

In order to deal with this problem, we need to adapt the Bhattacharyya metric to our dataset and design a new energy model with precise descriptors in the MPP model, so that we can get an accurate segmentation of the blood vessel. Therefore, in reason of the complexity of the task of blood vessels segmentation and of the rules setting in the model for very noisy and heterogeneous data, pure learning data driven models are not the best solution to our task.

On the other hand, Neural network (NN) remains a useful tool in pattern recognition and segmentation showing high accuracy in very complex tasks for nuclei segmentation. However, to our knowledge, the only study using NN for blood vessel segmentation in *H&E* was published in 2018 by F. Yi et al. [59] as detailed in table IV.2

Ref	Organ	year	Stain	Method	Accuracy %
[127]	skin	2017	dermoscopy	FCRN	93.1
[128]	skin	2017	dermoscopy	FCRN	96.3
[129]	skin	2017	dermoscopy	U-net	71.5
[59]	breast cancer	2018	<i>H&amp;E</i>	FCN	75

Table IV.2: A summary of state-of-art of blood vessel segmentation methods using pure discriminative (learning-based) algorithm

Although this pure data driven model showed an accuracy of 0.77 (Jaccard index), it needed 350 images at size of 384x384 pixels and 35 images at size 1600x1600 pixels, a big amount of data - usually difficult to accede.

It is common knowledge that the more data a NN algorithm has access for training, the more accurate it is. However, data is usually restricted in histopathology due to the invasive procedure that include a human specimen, the specific protocols which need to be accepted by the patients, as to the important amount of work necessary for a proper protocol design, as to the data and labels preparation. Indeed, to reach an interesting performance, DNN needs a large amount of annotated data - a very tedious and time consuming task for Pathologists. Indeed, to properly train a DNN model, Pathologists need to annotate by hand all the images of the data-set to compare the ground truth with the NN predictions. This task is affected by the human judgment which can be very variable from one Pathologist to another. For the reasons stated above, a pure learning data driven model such DNN is not enough the resolve our task. Nevertheless, its ability to deal with very complex tasks may be very useful in our case.

Concerning the segmentation of very complex shapes in very heterogeneous large scale and high resolution images, both Discriminative (i.e. learning-based / data-driven) and Generative (i.e. rule-based / knowledge-driven) models have, separately, their own drawbacks. However, combining both of them in a end to end model is a promising idea, able to reduce each ones' limits by. Indeed, as MPP model has the advantage of modeling (using conditional probabilities, rules etc.) the available knowledge (and therefore not using labeled data) and DNN has the advantage of resolving very complex tasks in a data-driven approach, we aim at designing an end-to-end model to overcome the problems detailed above, by taking advantages of both approaches (see figure IV.10).

We design a joint Discriminative-Generative approach trainable end-to-end that we call CNN-MPP model where the relationship between the CNN and MPP is complementary. CNN helps

Models	Model drawbacks	Model benefits
Pure learning data driven models (U-net, RCNN...)	Need <b>a lot</b> of labeled data	<b>Effective</b> for detection on very complex images. <b>Effective</b> for complex tasks.
Pure Rule based knowledge driven models (MPP, Active contours)	<b>Not effective</b> for detection on very complex images. <b>Not effective</b> for complex tasks.	Need <b>no</b> labeled data

Fig. IV.10 Combining data-driven and knowledge-driven models to design the CNN-MPP model.

the MPP energy model to resolve complex task and MPP helps the CNN with the Bhattacharyya energy such as CNN doesn't need a large amount of training data to perform good accuracy. Indeed, the input image goes through the CNN feed-forward, then on the CNN output image we compute the Bhattacharyya energy of the MPP. We define a loss function such that it maximize the Bhattacharyya value of the MPP. The pixels values of the CNN output image, then, keep changing through the optimization process of the loss function via the back-propagation, until we get the blood vessel segmentation, as detailed in figure IV.11.

According to the designed framework, we define our challenges as follows:

- What is the optimal energy model to limit the number of data for the CNN?
- How can we back-propagate through the MPP energy layers ?
- How to design a proposed algorithm which is very flexible to a large amount of class of histological images with few data?

## IV.6.2 Model design

Theoretically, the Bhattacharyya distance is defined as follows:

$$db(D_{in}, D_{out}) = \frac{(\mu_{in} - \mu_{out})^2}{4\sqrt{\sigma_{in}^2 + \sigma_{out}^2}} - \frac{1}{2} \log \frac{2\sigma_{in}\sigma_{out}}{\sigma_{in}^2 + \sigma_{out}^2}$$

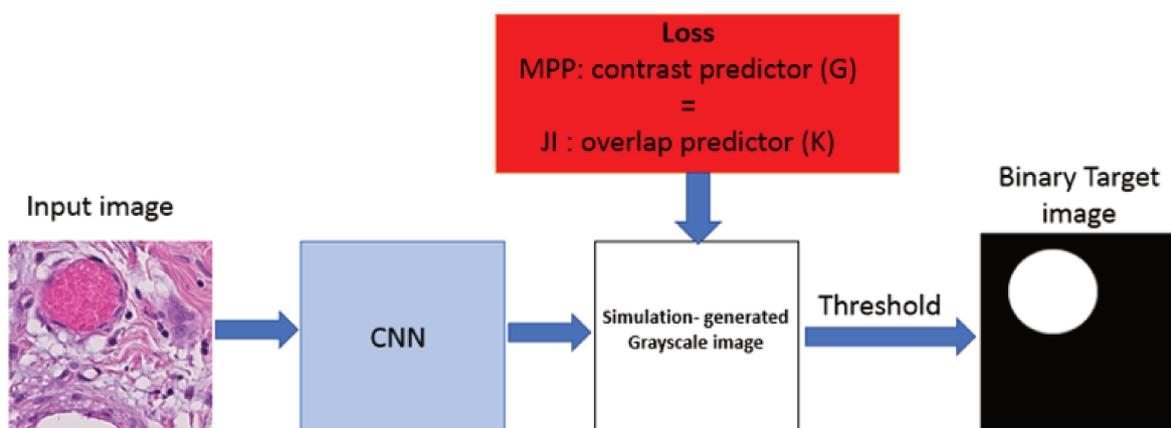


Fig. IV.11 *CNN-MPP model is a joint Discriminative-Generative approach based on both learning/data-driven model CNN and rule-based/knowledge-driven model MPP.*

According to the first term of the equation, which maximize the contrast between the inside and the outside region, good detection of the object leads to big value of  $db$  and detected boundaries of the blood vessel and bad detection of the object leads to small value of  $db$  and no detected boundaries of the blood vessel as detailed in figure IV.12 where we described the different Bhattacharyya values scenarios and its consequences on the blood vessel detection.

This condition is fulfilled only when the image is homogeneous with defined contour structure. As demonstrated by simulation in figure IV.13, heterogeneous areas in the images will give misleading detection of the boundaries of the blood vessels. In figure IV.13, the Bhattacharyya value is high when we have the correct detection and low when we have bad detection. However, in our simulation a third case raises: The Bhattacharyya value can be high in bad detection too. Indeed, having a good contrast between the inside and outside region does not guaranty a good overlap between the object and the blood vessel we want to segment. Therefore, designing a model based only on an energy model taking into consideration only the contrast predictor calculated with the Bhattacharyya distance doesn't guaranty a good detection of the blood vessel.

We need to design the model such that The Bhattacharyya value is high only in the good detection. This leads to design a model such that high Bhattacharyya value is only in the good overlap. A pure contrast predictor based on Bhattacharyya value is not suitable for very heterogeneous

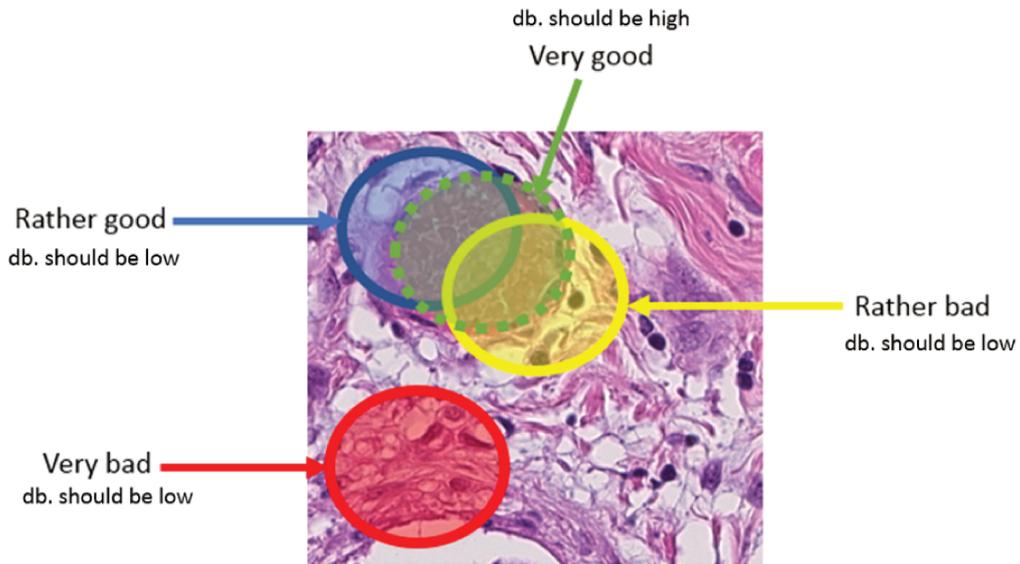
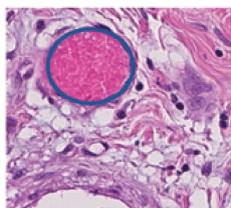


Fig. IV.12 Different detection scenarios of the blood vessels and the corresponding Bhattacharya value ( $db$ ): according to the  $db$  equation  $db$  value should be high only in the correct location otherwise  $db$  should be low

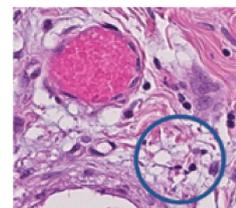
**Very heterogeneous and noisy images**



Good detection, High Bhattacharya value



Bad detection, High Bhattacharya value



Bad detection, low Bhattacharya value

**Bhattacharya energy is not suitable in very heterogeneous images**

Fig. IV.13 Different simulations assessing the behaviour of the Bhattacharya value on heterogeneous images

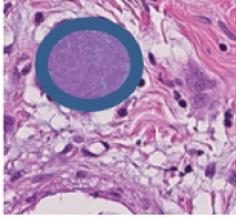
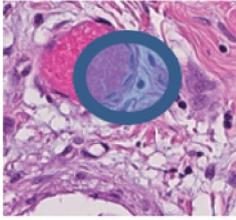
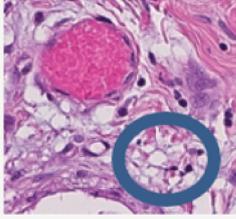
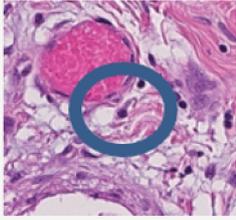
	
High db = 0.98 High JI = 1	High db = 0.76 low JI = 0.5
	
Low db = 0.01 Low JI = 0	Low db = 0.23 Low JI = 0.18

Fig. IV.14 The relation between the Jaccard index (JI) and the Bhattacharya value (db).

images. Therefore, we need to combine it to the overlap predictor, as detailed in figure IV.14.

As shown in figure IV.15, our idea is to combine a designed overlap predictor to a defined contrast predictor such as. When the overlap is correct, the contrast is good, corresponding to a good location of the blood vessel. In the other scenario, corresponding to an incorrect overlap, the

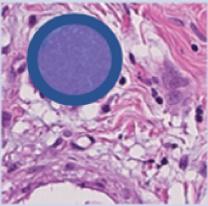
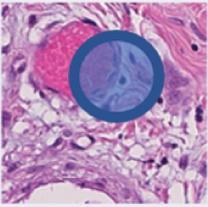
illustration	detection	The contrast predictor (G)	The overlap predictor (K)
	Good detection	Low	low
	Bad detection	High	High

Fig. IV.15 The relation between the contrast predictor (G) and the overlap predictor (K) in the CNN-MPP model.

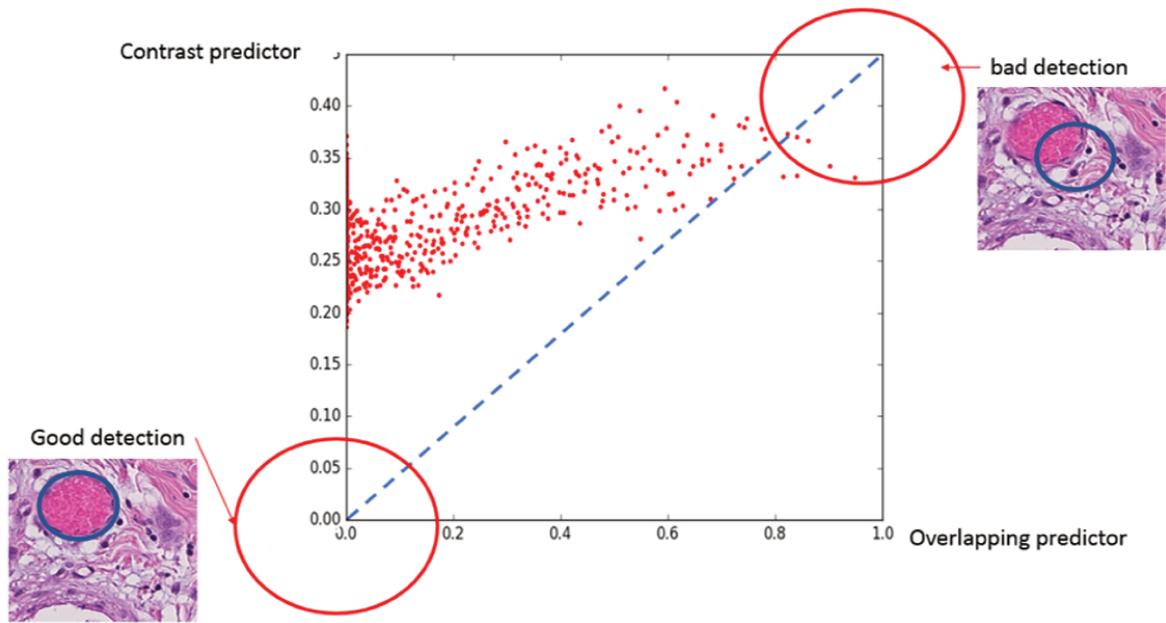


Fig. IV.16 The regression on simulations: the red dots represent the different values of  $db$  calculated on the  $512 \times 512$  image; the blue line represents the linear regression on the red dots, i.e. the perfect regression line, as an output of the CNN-MPP algorithm.

contrast is low and the blood vessel is not correctly located.

As detailed in the graph IV.16, our problem can be solved by a regression model, formalizing the relationship between the overlap and the contrast predictors. The blue line corresponds to the perfect regression in the model.

The CNN allow us to define the appropriate contrast predictor by comparing it to the overlap predictor controlled by the loss function of the model. The way we design the loss function, in our CNN, allow us to combine both predictors to get the correct location of the blood vessel. Therefore, we design a neural network with few labels thanks to the rule-based knowledge, using the Bhattacharyya similarity.

To start, we implement our model focusing on the first challenging case of blood vessel shown in figure IV.6 with circle shape and high density of blood cells. Our model design is detailed in 4 parts:

- a) Design of the contrast predictor
- b) Design of overlap predictor

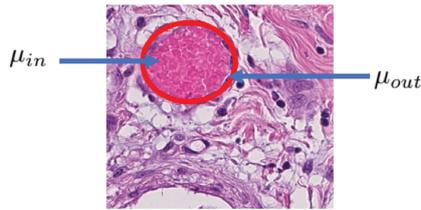


Fig. IV.17 The design of the contrast predictor  $db$

- c) Design of the loss function
- d) Configuration of the neural network

Next, we describe how to design the two predictors  $K$  and  $G$  as the two components of the loss function of the model.

#### a) Design of the contrast predictor $G$

We set the contrast predictor such that:

- The Bhattacharyya value is high in good localization.
- The Bhattacharyya value is small in bad localization.

To satisfy these two rules, we define  $\mu_{in}$ , the mean of the pixels intensities inside the circle  $C(x, y)$  and  $\mu_{out}$ , the mean of the pixels intensities in a defined boundaries of  $C$  (20 pixels) called respectively inside and outside region, as shown in figure IV.17. This design satisfies the first term of the Bhattacharyya value  $(\mu_{in} - \mu_{out})^2$ .

We generate different couples of masks called inside and outside masks, where the inside mask is a circle  $C_i(x_i, y_i)$  and outside mask is the corresponding boundary of the circle  $C_i$ . The different circles have different center positions  $(x_i, y_i)$  and fixed  $r^*$ .  $r^*$  corresponds to the radius ground truth of the blood vessels and  $x_i = x^* + \epsilon$  and  $y_i = y^* + \epsilon$  where  $\epsilon$  corresponds to errors around the correct center of the blood vessels  $(x^*, y^*)$ .

For each couple of mask with different  $(x_i, y_i)$ , we compute the Bhattacharyya distance  $db$  between the Neural Network output  $U(x_i, y_i)$  and the different couple of masks, as shown in figure IV.18.

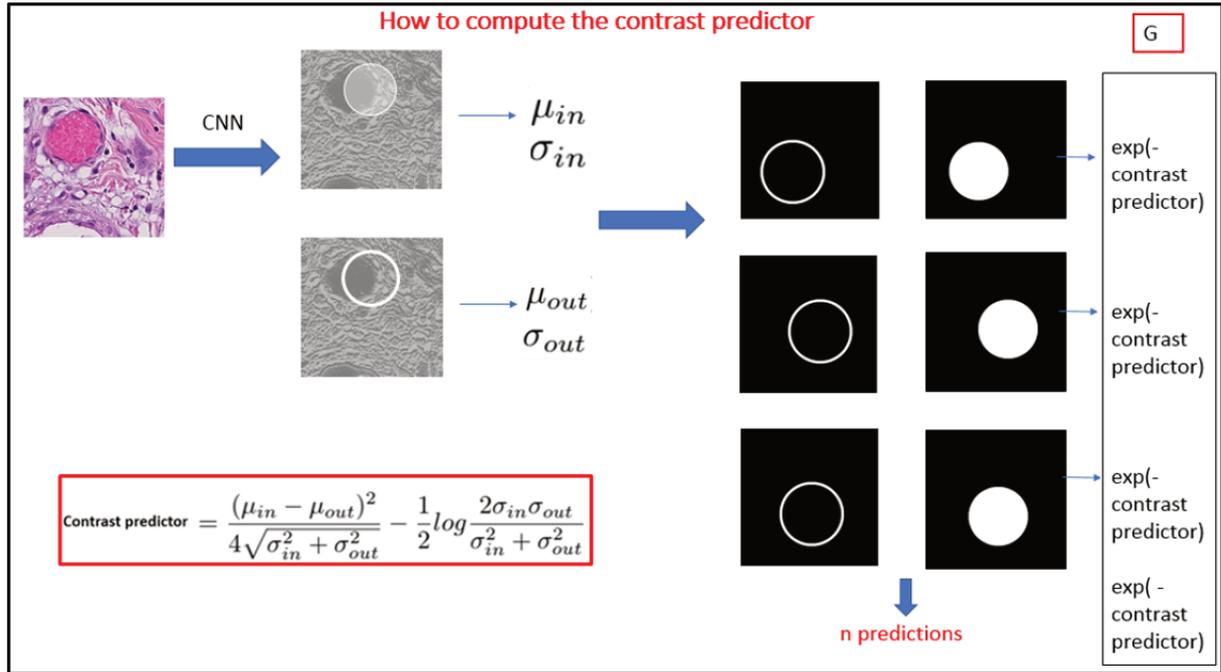


Fig. IV.18 The design of the contrast predictor db.  $\mu_{in}$  and  $\sigma_{in}$  are computed between the inside binary mask and the neural network output. However,  $\mu_{out}$  and  $\sigma_{out}$  are computed between the outside masks and the neural network output.  $G$  is a transformation function of db.

For each couple of inside-outside masks, we have one db value calculated by the Neural Network during the training.

$$db \in [0, \infty]$$

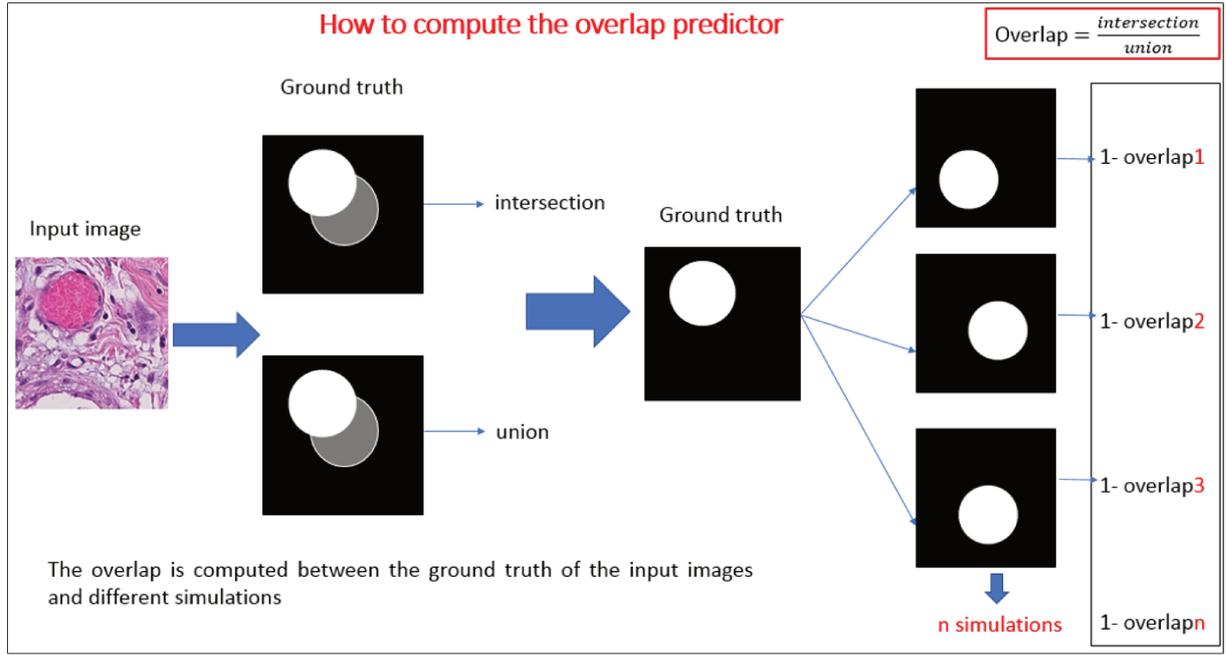
## b) Design of the overlap predictor

To design the overlap predictor, we use the Jaccard index similarity (JI).

The JI, also known as the Jaccard similarity coefficient, is a statistic measurement used for comparing the similarity of sample sets. It is defined as the size of the pixel's intersection divided by the size of the pixel's union of the sample sets [130].

We compute the JI between the Ground truth  $GT(x^*, y^*)$  of the input image and the same inside mask generated, to calculate the contrast predictor db.

$I(x, y)$  is the input image and  $GT(x^*, y^*)$  is the ground truth drawn by hand (the mask of

Fig. IV.19 The design of the overlapping predictor  $K$ 

the perfect CNN output). Suppose that  $S_1(x_1, y_1)$ ,  $S_2(x_2, y_2)$  and  $S_3(x_3, y_3)$  are 3 automatically-generated random inside-masks. Therefore, the JI for each mask is computed between the binary GT of the input image and each inside binary mask, separately, as follows:

$$JI_i(GT, S_i) = \frac{|GT \cap S_i|}{|GT \cup S_i|}$$

where  $i = 1, 2, 3$

The JI value is interpreted as follows:

- If  $JI \simeq 1$ , the intersection is equal to the union and we detect correctly the blood vessel: this represents the ideal case.
- If  $JI \simeq 0$ , there is no intersection, so we don't detect the blood vessel: this represents the worst case.

**b).1 Automatic masks generation for CNN-MPP model** We aim at training our model with masks generated automatically. Therefore, we generate a 1700 grid pairs of masks (inside and outside). The corresponding framework is as follows:

1. On 512x512 matrix, we create circles  $C_i(x_i, y_i, r^*)$ , as we are interested for now to the first challenging case where the blood vessel is approximated/simulated by a circle.

- $i \in (1, 2, 3, \dots, N)$
- $(x_i, y_i)$  coordinates of the center of the circle  $C_i$
- $r^*$  the radius of the circle - the same for all circles.

2. The distance between  $C_i(x_i, y_i, r^*)$  and  $C_{i+1}(x_{i+1}, y_{i+1}, r^*)$  is equal to a step size  $s=8$  pixels.

The choice of  $C_i(x_i, y_i)$  affects the linear regression between the overlap and the contrast predictors. Indeed:

1. If the distance between the generated circles centers  $C_i(x_i, y_i) \in C$  and the  $GT(x^*, y^*)$  is big, most JI values are going to be equal to zero.
2. If the distance between the generated circles centers  $C_i(x_i, y_i) \in C$  and the  $GT(x^*, y^*)$  is very small, most JI value are going to be equal to 1.
3. If the number of all the possible distances  $(x_i, y_i)$  is roughly the same, we have a better regression curves, as the model trains on all possible distances, generating a more accurate output image.

To avoid the two first scenarios, We compute all JI values for all the generated inside-masks and we pick up for each interval from  $[0,1]$  roughly the equivalent numbers of masks (45 masks in total for each GT).

### c) Design of the loss function

Let  $I(x, y)$  be the RGB input image of a size 512x512. Let  $U(x_i, y_i)$  be the output image of the Neural Network, of a size 512x512 pixels. Let db be the Bhattacharyya value of the MPP model.

The conditions to be respected in our Loss function are the following:

1. db is calculated on  $U_i(x_i, y_i)$  image;
2. db is represented by the coordinates of the circle center  $(x_i, y_i)$ , the radius  $r^*$  and the output of the Neural Network  $U_i(x_i, y_i)$ ;

3. let  $x^*, y^*, r^*$  be the correct location and radius, draw by hand, of the blood vessel (ground truth);
4. db should have the highest value in the ground truth location  $x^*, y^*, r^*$ ;
5. Around  $x^*, y^*, r^*$ , the deviation ( $\epsilon$ ) makes db value smaller in the neighborhood of the correct location. Therefore, db should be high only in case of correct location, otherwise db is small;

The contrast predictor db should be high in correct location and small in the neighborhood of correct location (and therefore, the overlap predictor). We thus define the loss function of the neural network such that it minimizes the difference between the overlap predictor and the contrast predictor. Thus, when the overlap is good, the contrast is correct, too.

Let's  $G(db)$  be a transformation function of the contrast predictor db and  $K(JI)$  be the transformation function of the overlap predictor. While training the neural Network model, we minimize the loss function through the epochs. Therefore, G and K should be defined such that both quantities are in the same range.

By minimizing the loss function, we should make the Bhattacharyya value the highest possible in the good location and the lowest possible otherwise. K values do not depend on the neural network training. The loss function of our neural network will thus be:

$$Loss = \sum_{i=1}^N (G(db_i) - K(JI_i))^2$$

where N is the number of possible simulated masks.

Then,  $G(db_i) = \exp(-db_i)$  and  $K(JI_i) = 1 - JI_i$

Where i is the index of different possible masks with different location  $(x_i, y_i)$ .

By minimizing  $G(db_i)$  through neural network training epochs, we maximize the Bhattacharyya value to get the good location of the blood vessel. However,  $K(JI_i)$  is a fix quantity as it is calculated between the random possible masks and the ground truth of the input image. As K does not depend on the neural network, through the optimization of the loss function, minimizing the difference between K and G is going to make G low in good location (as K is low in good location)

and high otherwise.

$$G(db_i) \in [0, 1] K(JI_i) \in [0, 1]$$

While optimizing the neural network, and since  $K$  does not depend on the neural network output, the lost function becomes:

$$\begin{aligned} Loss &= \sum_{i=1}^n (G(db_i(U)) - K(JI_i))^2 \\ \frac{\partial Loss}{\partial w} &= \sum_{i=1}^n \frac{\partial (G(db_i(U)) - K(JI_i))^2}{\partial w} \\ &= 2 \sum_{i=1}^n (G(db_i(U)) - K(JI_i)) \frac{\partial (G(db_i(U)) - K_i)}{\partial w} \\ &= 2 \sum_{i=1}^n (G(db_i(U)) - K(JI_i)) \frac{\partial G(db_i)}{w} \end{aligned}$$

$$\frac{\partial Loss}{\partial b} = 2 \sum_{i=1}^n (G(db_i(U)) - K(JI_i)) \frac{\partial G(db_i)}{b}$$

where  $w$  are the weights and  $b$  are the bias of the neural network. Figure IV.20 represents the whole framework described above.

#### d) Configuration of the neural network

In literature, the only work reported according to segmentation of blood vessels in *H&E* slides was using (FCN), therefore, we first tried to use this same architecture in our CNN-MPP model.

## IV.7 Experiment

To train our model, we use only 8 RGB images with 512x512 pixels representing the 3 categories of high density blood cells: surrounded by nuclei, surrounded by adipocyte and surrounded by collagen

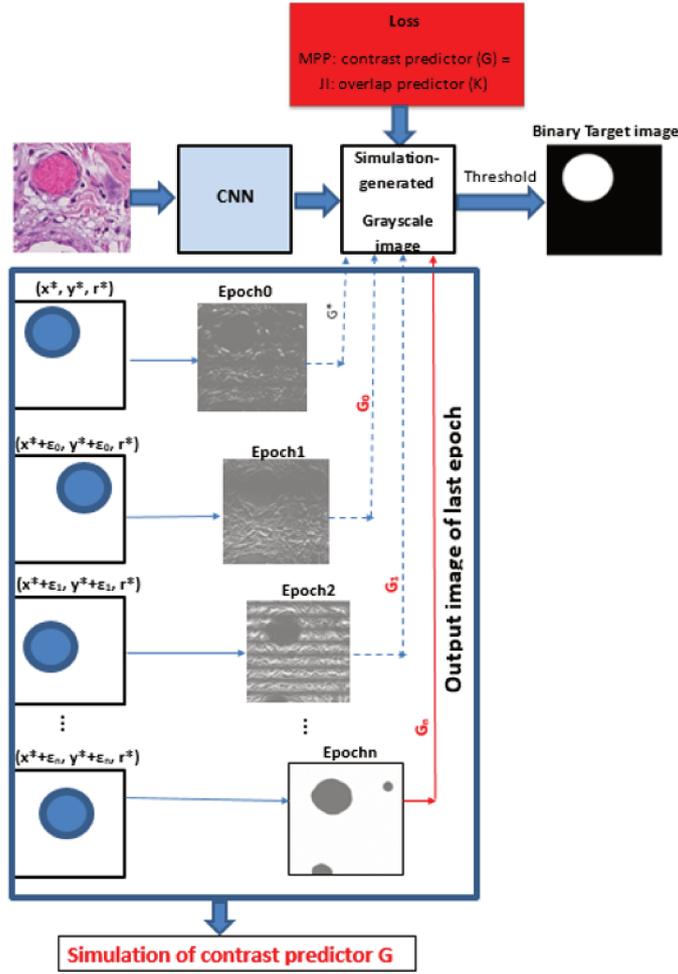


Fig. IV.20 The input image  $I(x, y)$  Hematoxylin and Eosin staining of tumour section showing a blood vessel.  $x^*y^*$  are the correct coordinates of the blood vessel center given by the ground truth,  $r^*$  is the correct radius of the blood vessel,  $U$  is the neural network output on which the Bhattacharyya distance  $db$  is computed.  $\epsilon_0, \epsilon_1, \dots, \epsilon_n$  are different errors of the wrong location  $(x, y)$  and  $r_*$ .

and nuclei as shown in figure IV.21. We process a data augmentation (translation, rotation) and we split our data: 4 images for training, 2 images for validation and 2 images for testing. For each image, we compute 45 inside/outside masks.

We use a Fully Convolutional Neural Network (FCN) model with 6 down-sampling blocs of CONV-MAXPOOLING-RELU layers and 6 up-sampling blocks of CONV-UP-Sampling layers. A shortcut is performed between the down- and up-sampling layers. We use Adam optimizer and a batch size of 2 images for training. The learning rate is 0.001 and we train our algorithm during 500 epochs.

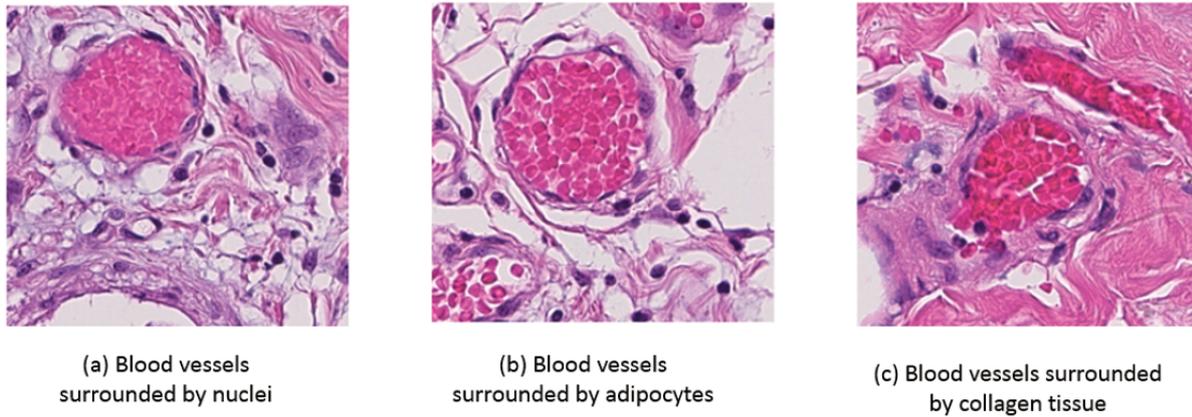


Fig. IV.21 The category of data type used for CNN-MPP model.

## IV.8 Results

Through the training, the loss function decreases from 0.37 to attend a value of 0.22 after 500 epochs, as shown in figure IV.22. The red curve represents the CNN-MPP loss function during the training and the blue curve shows the CNN-MPP loss function during the validation process. As the two curves are almost superimposed, we can say that our model doesn't over-fit the training data. We compare different optimizer (Adam, SGD) and we end by using the Adam optimizer.

We remind that our objective is to train the model such that the contrast predictor  $G$  is equal to the overlap predictor  $K$ . Figure IV.23 shows that the regression curve between the contrast predictor  $G$  and the overlap predictor  $K$ , tends to fit the straight line shape during the training (figure IV.23) and the testing (figure IV.24). The blood vessels surrounded by collagen and nuclei (shown by yellow triangles in figure IV.23), was the hardest to train by our model. Indeed, as the contrast predictor depends on the boundaries regions of the blood vessels too (outside region), if this area is less homogeneous (including information from different components of the microenvironment in this case collagen and nuclei), the contrast cannot be very high between the inside and the outside.

The final image given by the CNN-MPP algorithm is a grayscale image on which we apply a global threshold to obtain the binary output image as described in figure IV.26.

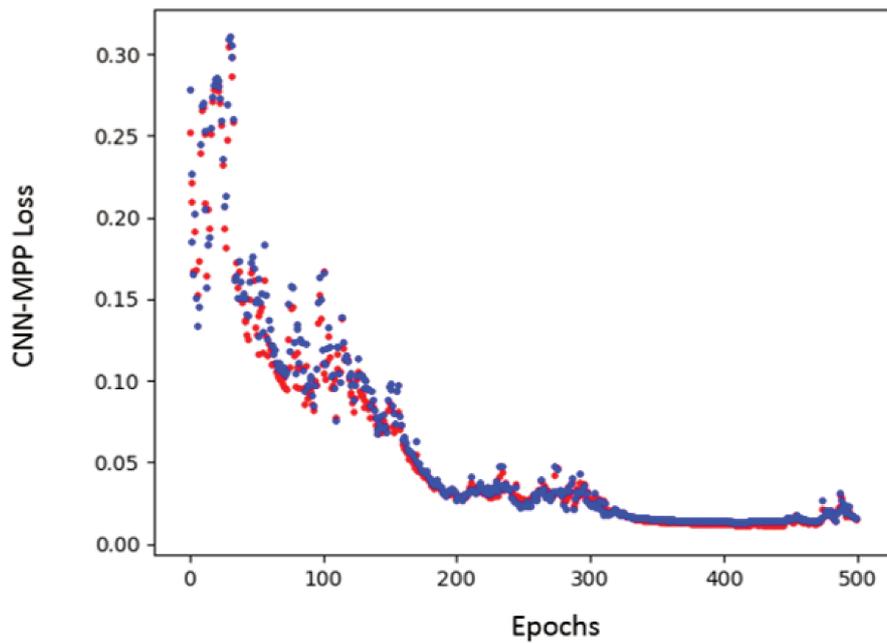


Fig. IV.22 The CNN-MPP loss function during the training and the validation of the CNN-MPP model. The blue curve shows the loss function during validation and the red curve shows the loss function during the training.

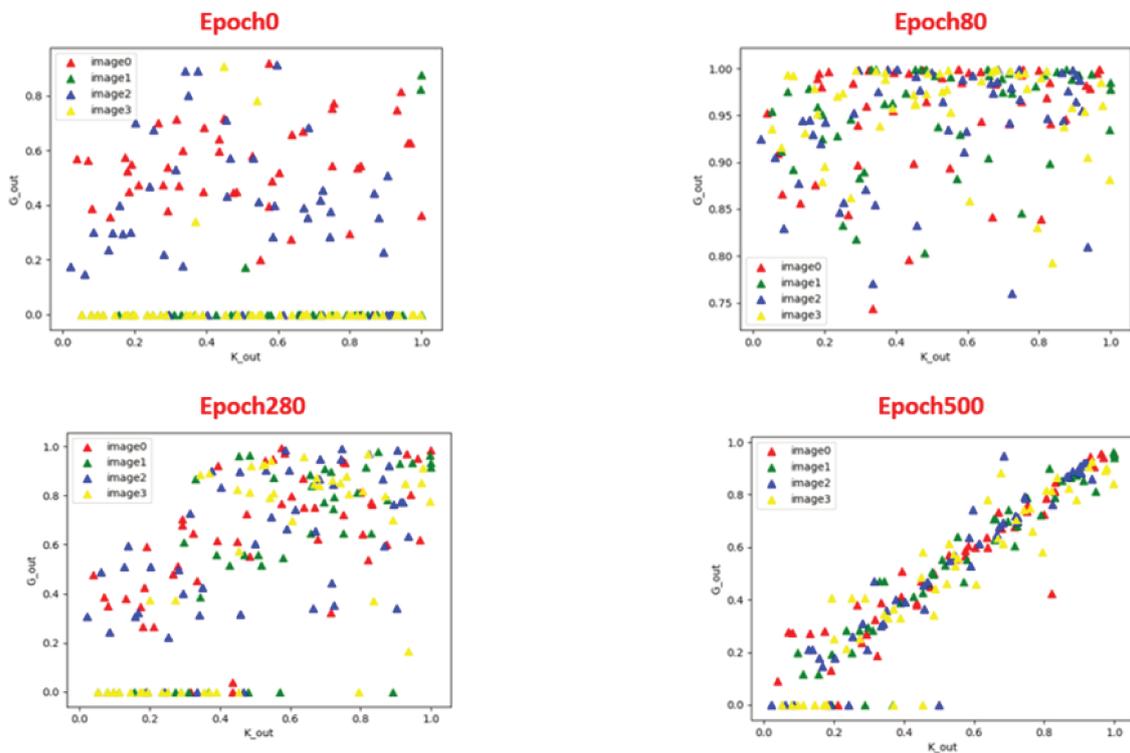


Fig. IV.23 The overlap predictor versus contrast predictor through 500 epochs for 4 training data. Each image is of 512x512 pixels and each color refers to the contrast predictor versus the overlap predictor for one image.

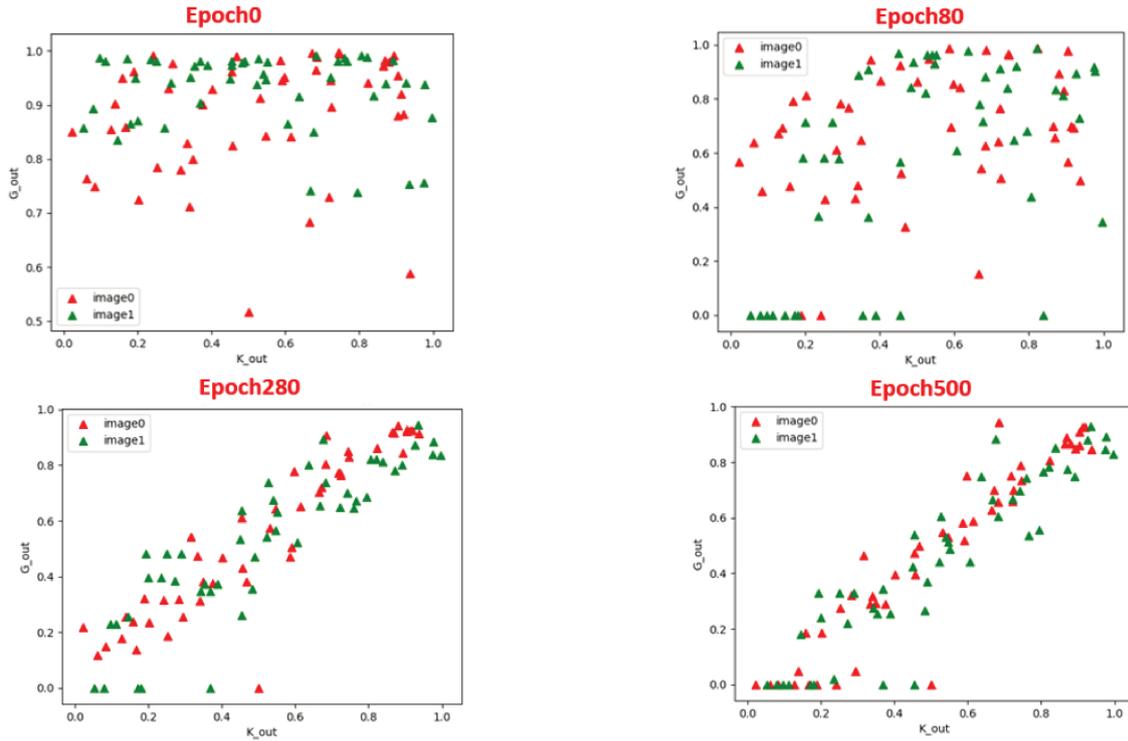


Fig. IV.24 The overlap predictor versus contrast predictor through 500 epochs for 2 testing images. Each image is of 512x512 pixels and each color refers to the contrast predictor versus the overlap predictor for one testing image.

## IV.9 Proof of concept of CNN-MPP model on synthetic data-set

We believe that the number of data used to develop our algorithm is not enough to conclude about its performances. Therefore, we gave a proof of concept by testing our algorithm on an augmented database, generated automatically (synthetic data).

### IV.9.1 Experiment

Histopathology images of blood vessels are very heterogeneous as they include information of the surrounding components of the microenvironment such as collagen, fibroblasts and nuclei. We generate, thus, synthetic data which are similar to histopathology images of blood vessels. 30 RGB images of 512x512 pixels were automatically generated. For each image, we call foreground - the circle which is simulated to the blood vessel and the background - the region outside the circle.

For this purpose, we used the next protocol :

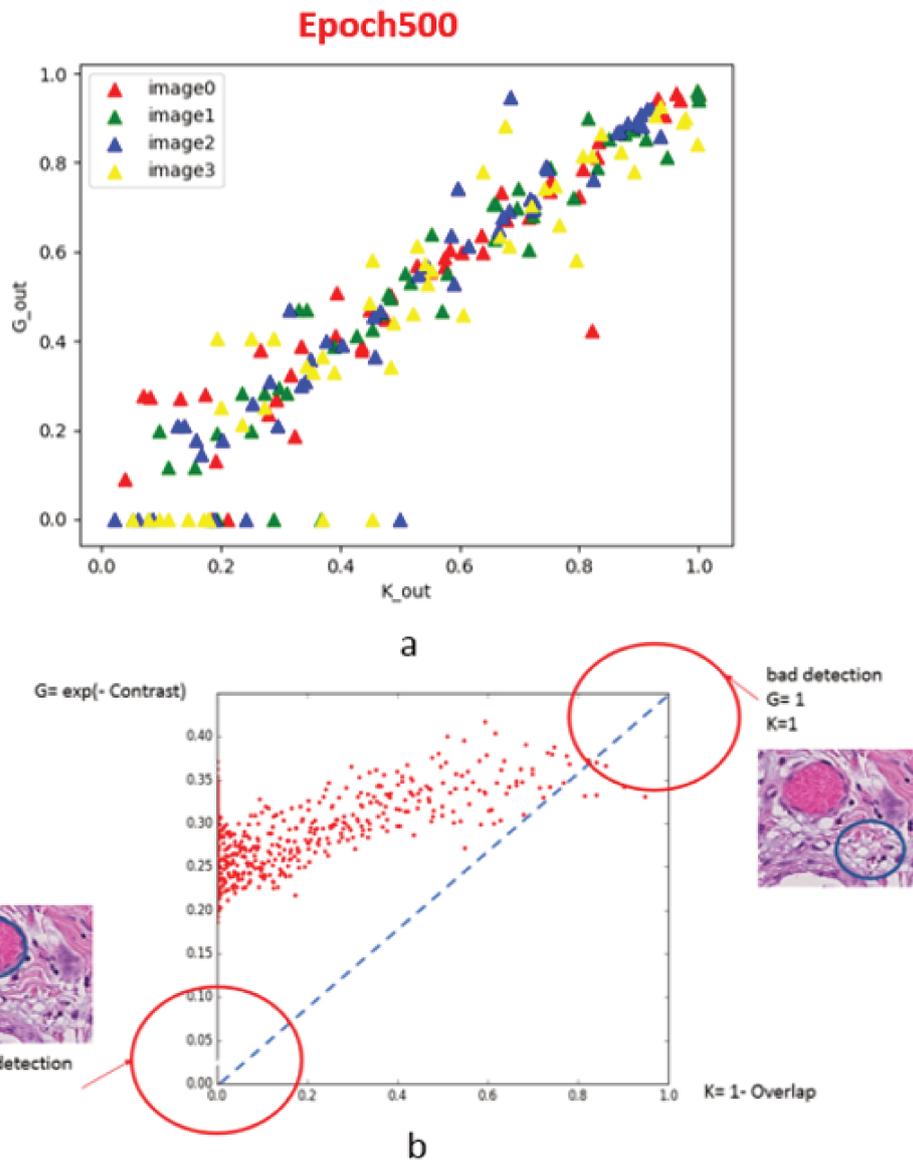


Fig. IV.25 The final regression given by CNN-MPP after 500 epochs on H&E data compared to the theoretic graph b.

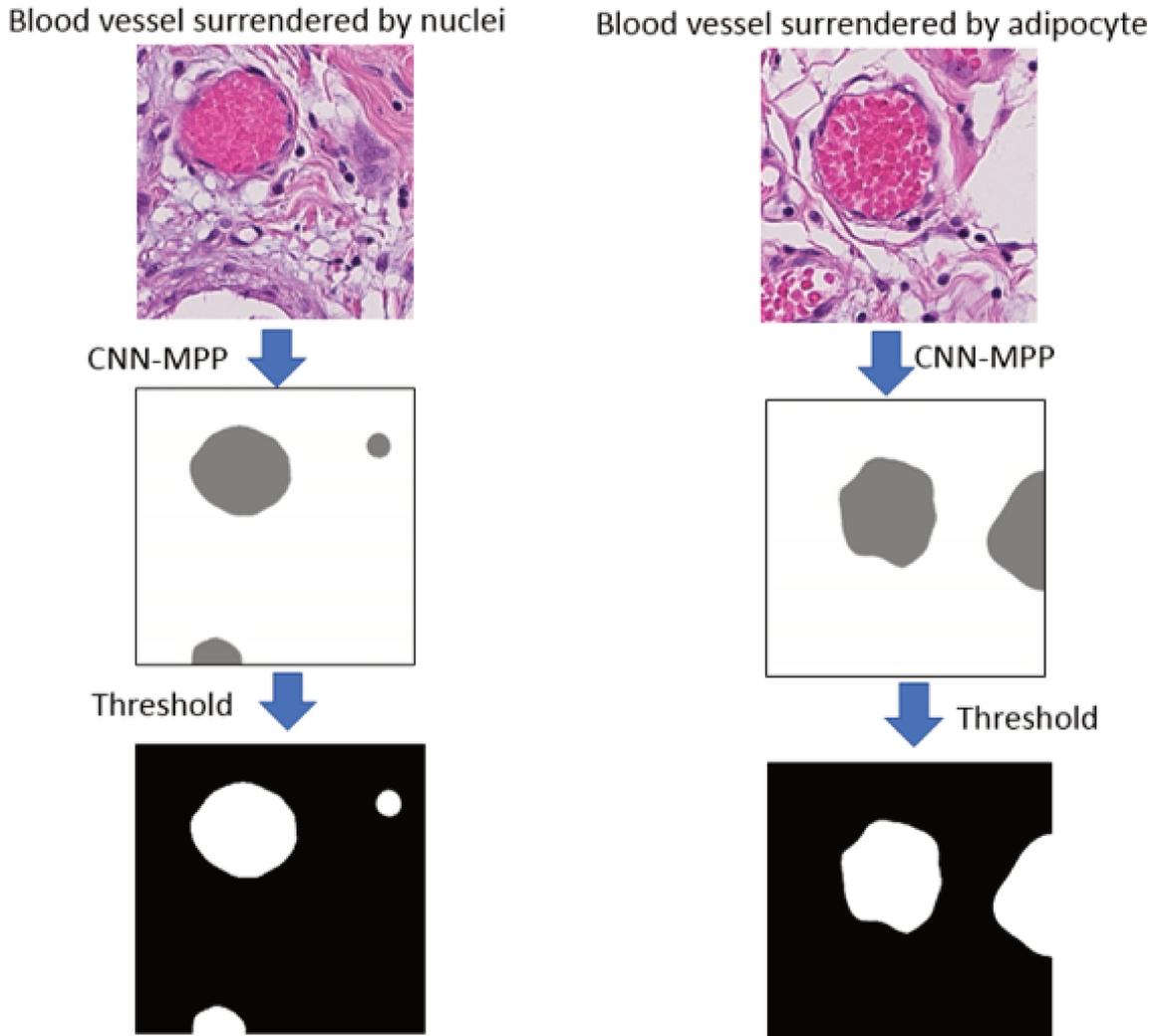


Fig. IV.26 The segmentation result on the testing data.

- First, we use the online free data-set: the Describable Textures Data-set (DTD) to download two sets of images of different color and texture with different sizes: background and foreground.
- Second, we generate automatically 30 binary masks (size 512x512 pixels) of circles with different radius  $r_i \in (50, 150)$  and different locations  $(x_i, y_i) \in [20, 340]$  as shown in figure IV.27
- Then, we multiply randomly each mask of the 30 binary masks with one image from the foreground set to obtain different circles with different colors and textures.
- Finally, we add the different circles foreground obtained previously to the different backgrounds to get the final images shown in figure IV.28 We split the 30 synthetic images : 10 data for training, 10 data for validation and 10 data for testing.

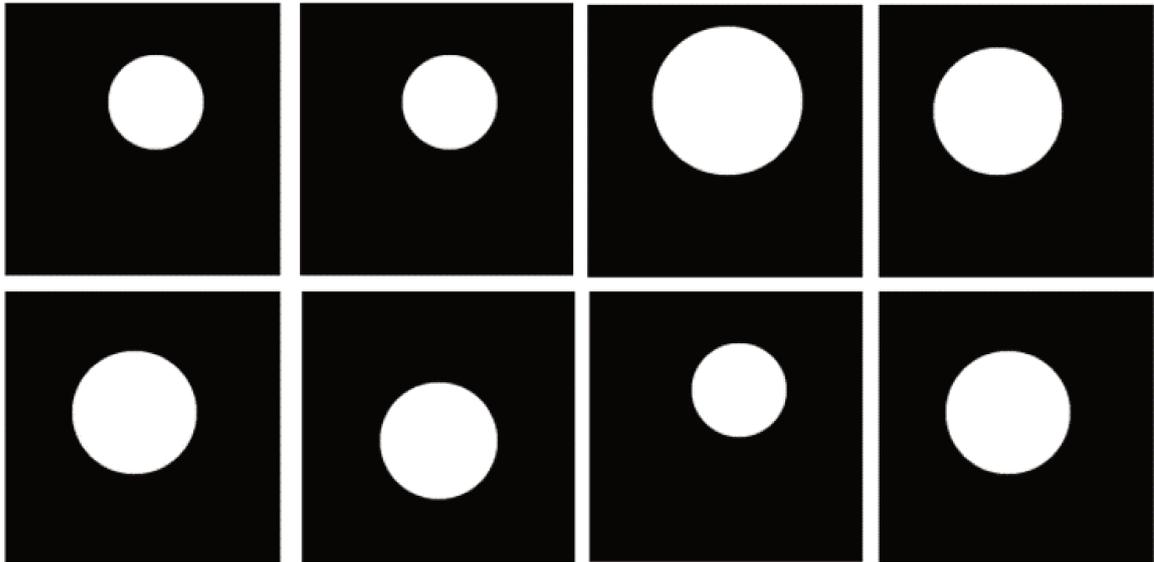


Fig. IV.27 Example of the generated GT data used to create the synthetic data. Each ROI represents  $512 \times 512$  binary images with different  $(x_i, y_i, r_i)$

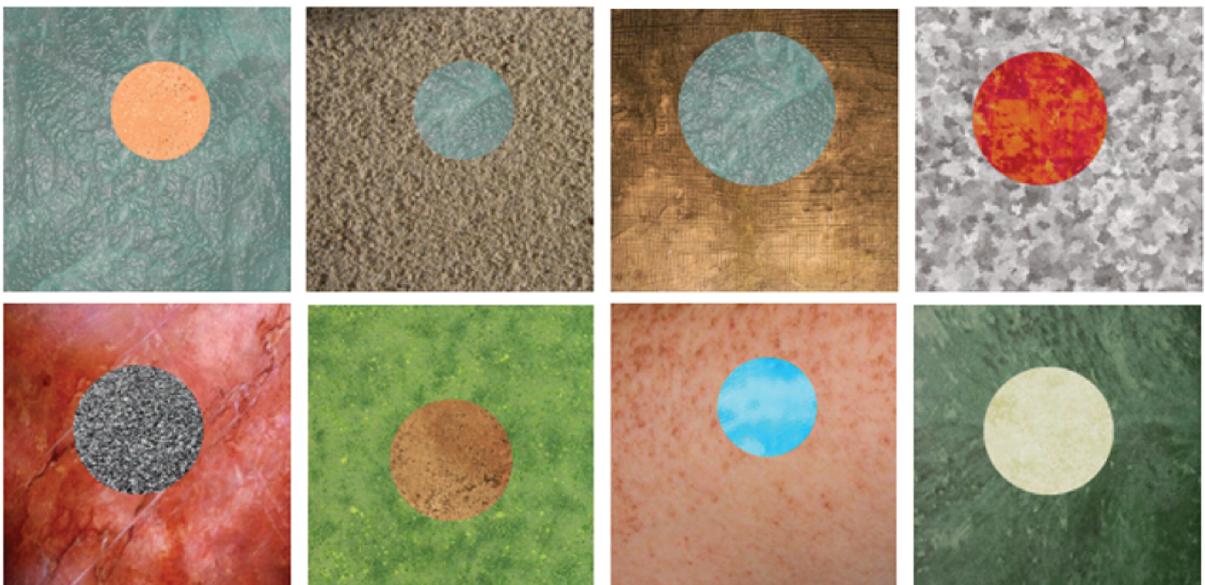


Fig. IV.28 Example of the generated synthetic data used to create the synthetic data. Each ROI represents  $512 \times 512$  RGB images with different  $(x_i, y_i, r_i)$

The parameters according to this experiment are the following:

- The kernel size for all the CONV layer is 3x3 pixels with padding operation.
- We use RELU as activation function.
- For the POOL operation, we use 2D Maxpooling with 2x2 pixels stride.
- Each bloc of CONV/POOL layer is followed by a batch normalization with a fixed momentum=0.99 and epsilon=0.001.
- The number of channels in first layer is 8 channels then it is multiplied by 2 for each CONV layer.
- Batch size is equal to 5.
- We use Adam as optimizer

#### a) Results

Although the training data was limited to 10, the loss function decreases progressively from 0.27 to reach zero after 1000 epochs (red curve for training in figure IV.29).

During the 1000 epochs, the relationship between the contrast predictor and the overlap predictor tends to fit the straight-line IV.30 during the training phase and during the testing phase (figure IV.31).

## IV.10 CNN-MPP comparison with Active contours and FCN

### IV.10.1 Comparison of CNN-MPP model with pure learning driven model FCN

We compare the CNN-MPP algorithm with the only study reported in the literature using FCN. Although the FCN model demonstrated a good accuracy using 350 images for training and 80 images for testing, the FCN model over-fit when we try to train it on less data (8 data). Figure IV.32 shows the difference on the performance between FCN model and CNN-MPP model. The loss function decreases from 0.8 to reach zero after 500 epochs for both method for the training. However, for the FCN the validation, loss did not decrease comparing to the CNN-MPP validation loss.

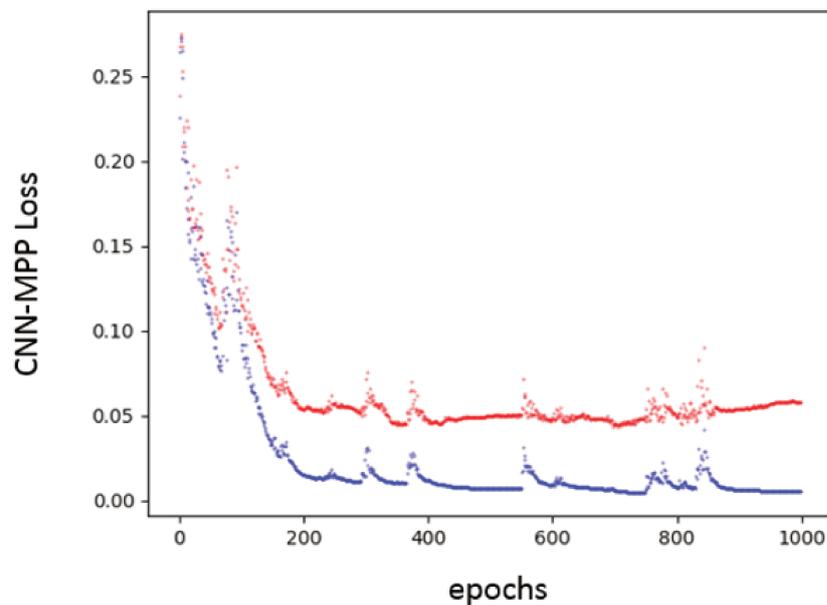


Fig. IV.29 The loss function of the CNN-MPP model for the synthetic data-set. The red curve represents the validation loss. The blue color shows the training loss

#### IV.10.2 Comparison of CNN-MPP model with pure based-knowledge driven model Active contours (AC)

Active contours (AC) was proposed by Kas and al. in 1988 [131] and it is one of the widely used algorithm in medical image segmentation [132, 133]. AC is a pure rule based knowledge driven model which the main idea is to formulate an energy functional by using image statistics, curvature and gradient information and minimize it through iterations. Same as MPP, Active contours are based on the definition of an energy function that we try to minimize based on precise rules. This rule is based on the fact of minimizing the energy functional such that the level-set curve towards the desired object boundary. To date, various active model have been used for medical image segmentation.

Our main goal according to the CNN-MPP model was to simplify the very heterogeneous image into more meaningful or easier image to analyze that any other rule based method can use for segmentation by developing a regression model on both contrast and overlap information. However, the use of CNN as a learning driven model allowed us to get the final segmentation that

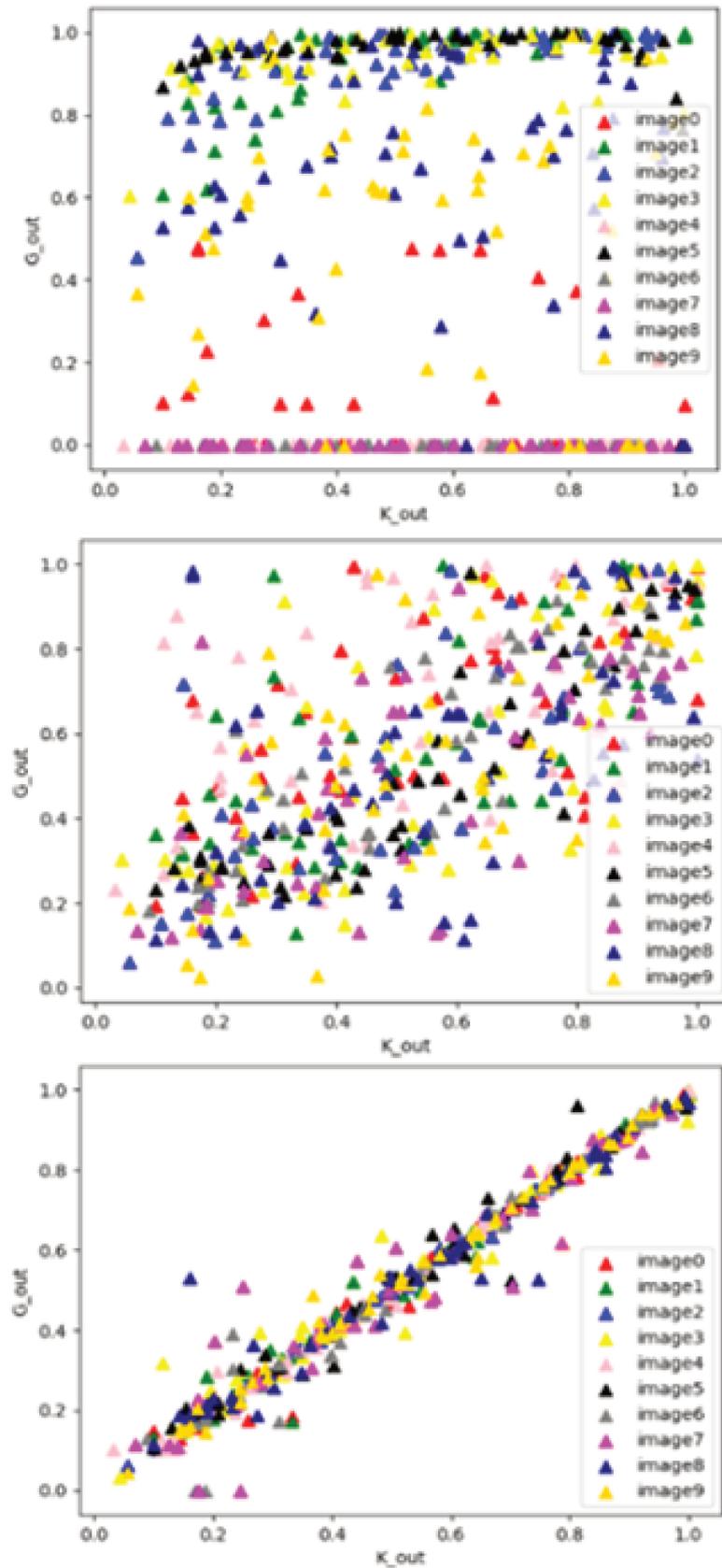


Fig. IV.30 The relationship between the contrast predictor and the overlap predictor through 500 epochs during the training of the algorithm CNN-MPP on 10 synthetic images.

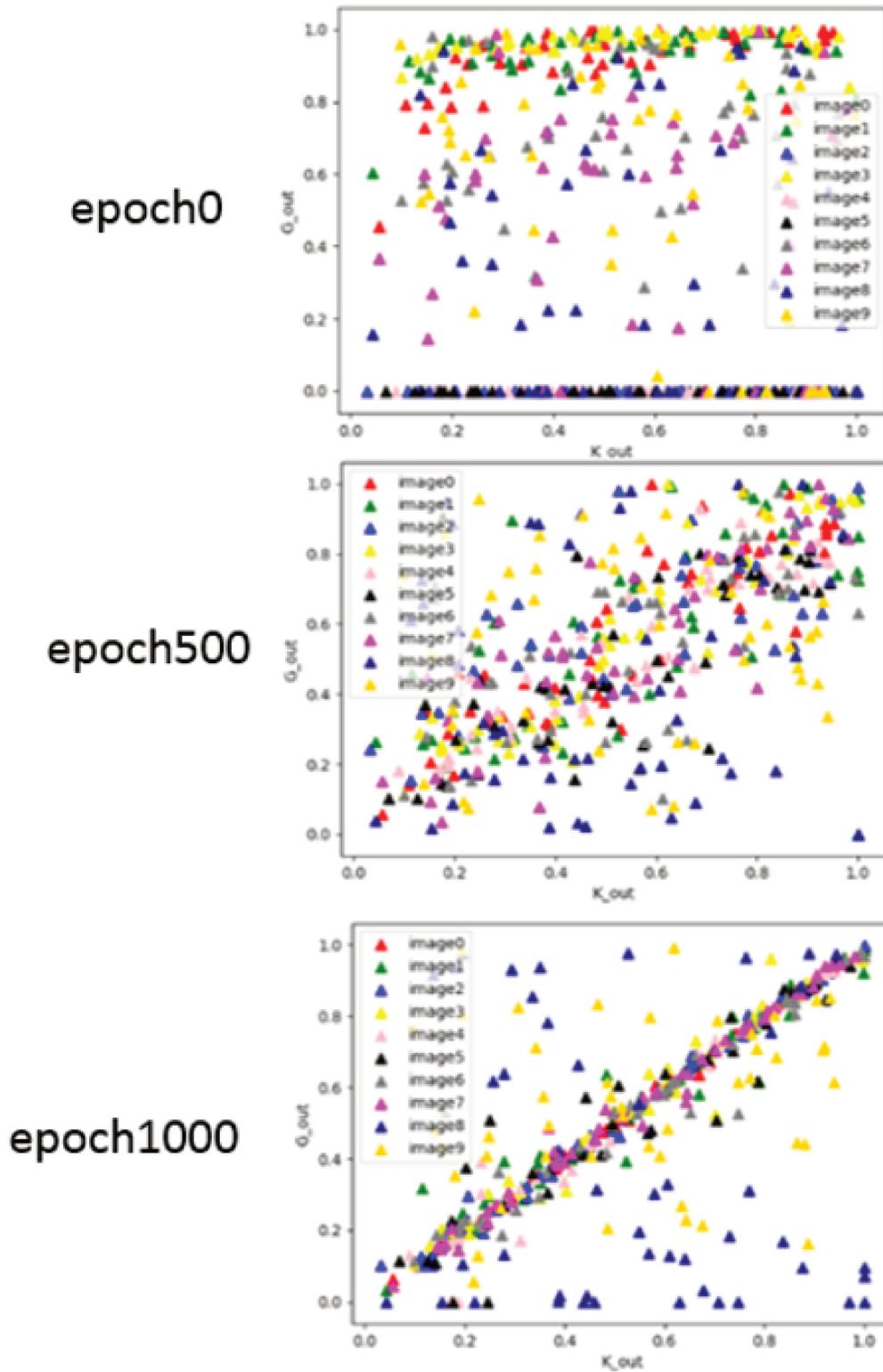


Fig. IV.31 The relationship between the contrast predictor and the overlap predictor through 500 epochs during the test on 10 synthetic images.

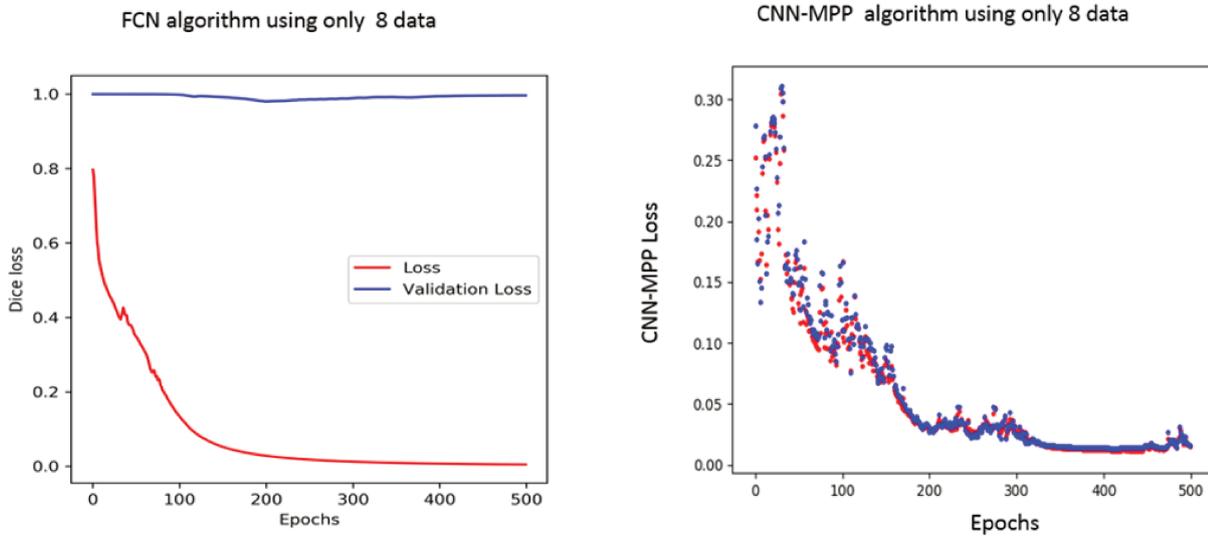


Fig. IV.32 Comparison of the Loss function using a FCN model and CNN-MPP model on *H&E* data. The red curve shows the training while the blue curve shows the validation loss.

we compare to AC. As demonstrated in figure IV.33, using a pure based knowledge algorithm such as AC leads to an over-segmentation on very heterogeneous data such as *H&E* which is not the case with a mixed algorithm such as CNN-MPP model. The segmented output image using CNN-MPP model is shown in figure IV.34.

Images	Jaccard	DICE	Precision	sensitivity	Accuracy
Nuclei AC	0.14	0.24	0.17	0.44	0.60
Adipocyte AC	0.20	0.33	0.25	0.48	0.66
Nuclei and collagen AC	0.10	0.18	0.11	0.46	0.64
Nuclei CNN-MPP	0.55	0.71	0.81	0.63	0.93
adipocyte CNN-MPP	0.48	0.65	0.67	0.63	0.88
nuclei and collagen CNN-MPP	0.21	0.34	0.22	0.80	0.73

Table IV.3: Comparison between the performance of AC algorithm on *H&E* images and performance of CNN-MPP on *H&E*, for different type of blood vessels.

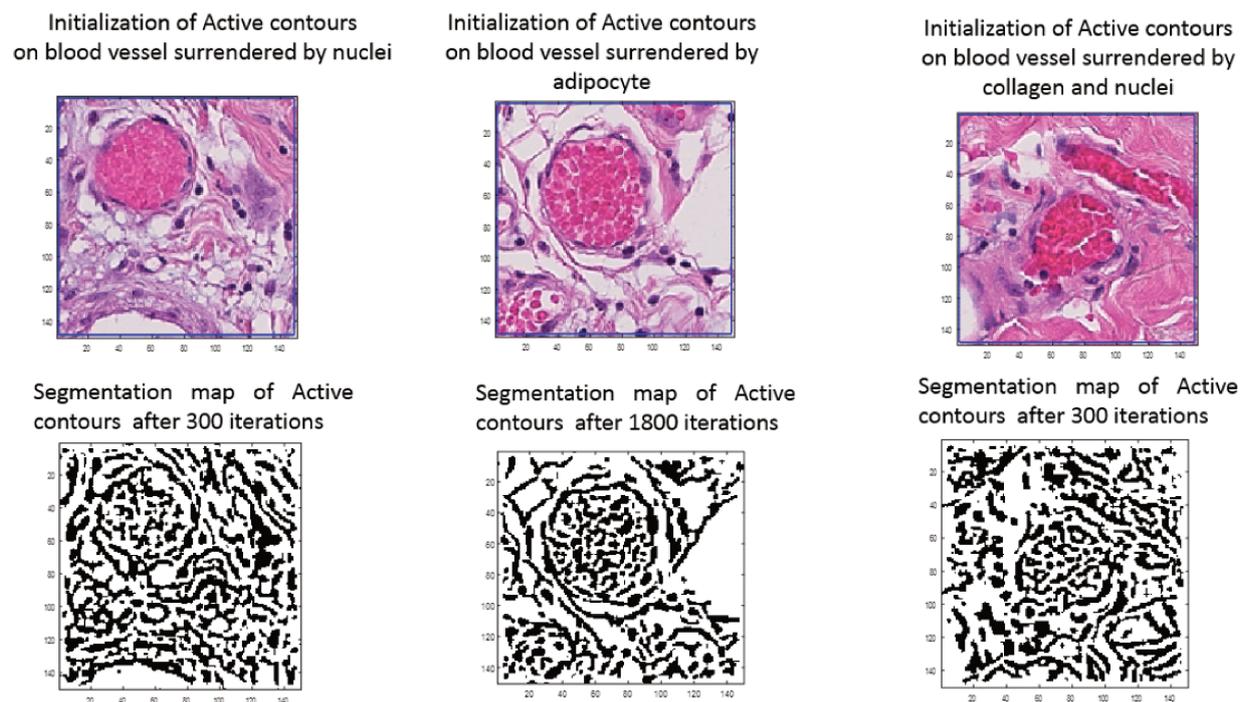


Fig. IV.33 Performance of a pure knowledge based method: active contours tested on our heterogeneous data.

### IV.10.3 Discussion of the CNN-MPP algorithm

#### a) Limits

Although, we show that our algorithm is able to learn from very few amount of data - thanks to the rule fixed in the loss function by the Battacharya distance of the MPP algorithm -, it has its limited number of data not to go below. Figure IV.35 shows that, the CNN-MPP algorithm is enable to train if we use less then 4 images. in figureIV.36 the algorithm was enable to perform for the testing phase and the loss function did not decrease as detailed in figure IV.37 This result needs more investigations to give explanation about the limits of the algorithm.

#### b) Generative Adversarial Networks (GANs) versus CNN-MPP Networks

Generative Adversarial Networks (GAN) are a neural network architecture composed of two main parts : a discriminator neural network and a generator neural network (figure IV.38. The generator neural network generates samples from a distribution (random variables) and the discriminator neural network decides if the generated data are real or fake. Therefore, the discriminator is a classifier which

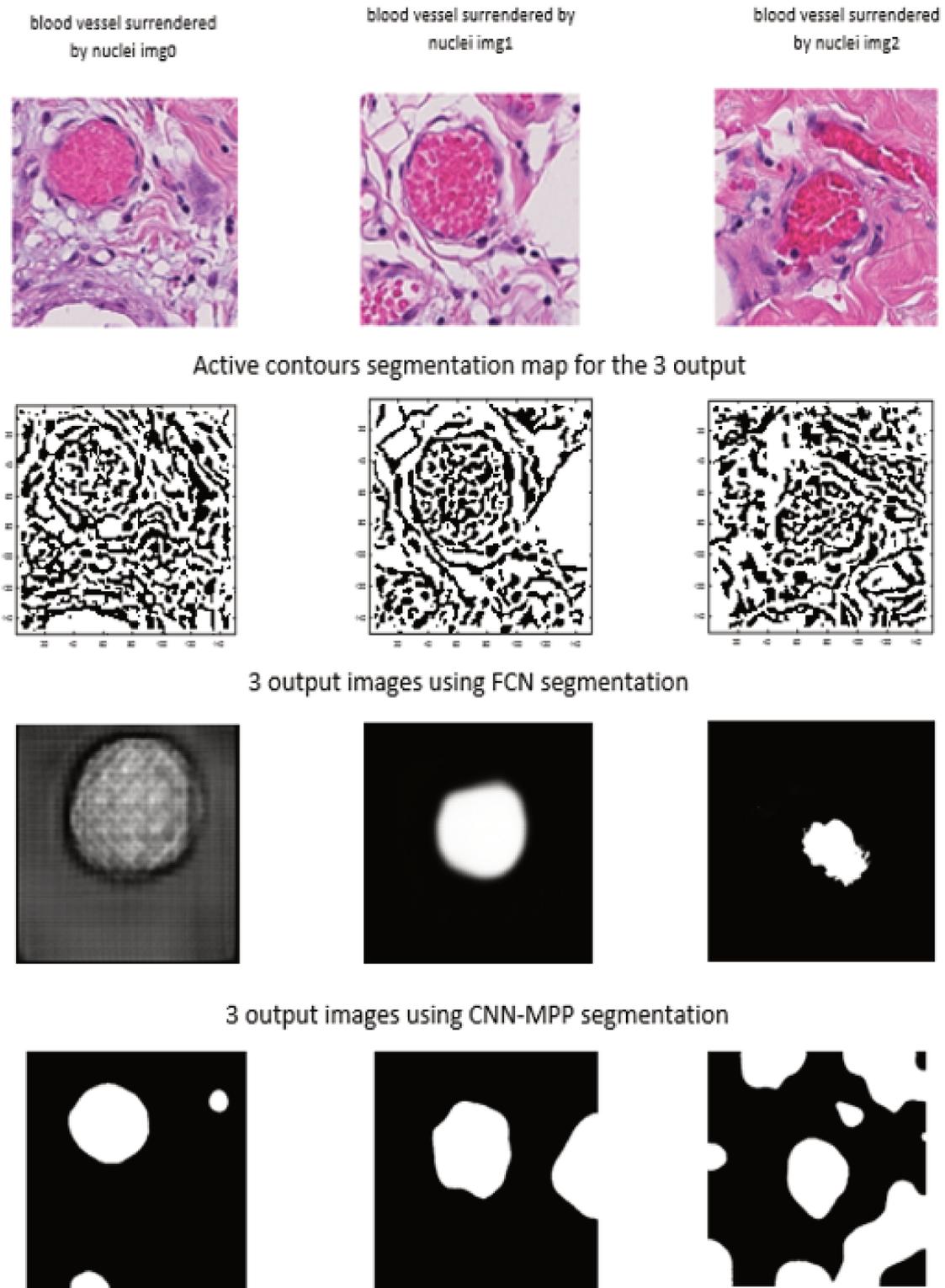


Fig. IV.34 Comparison between Performance of a pure knowledge based method: active contours and FCN tested on our heterogeneous data versus the performance of the CNN-MPP algorithm

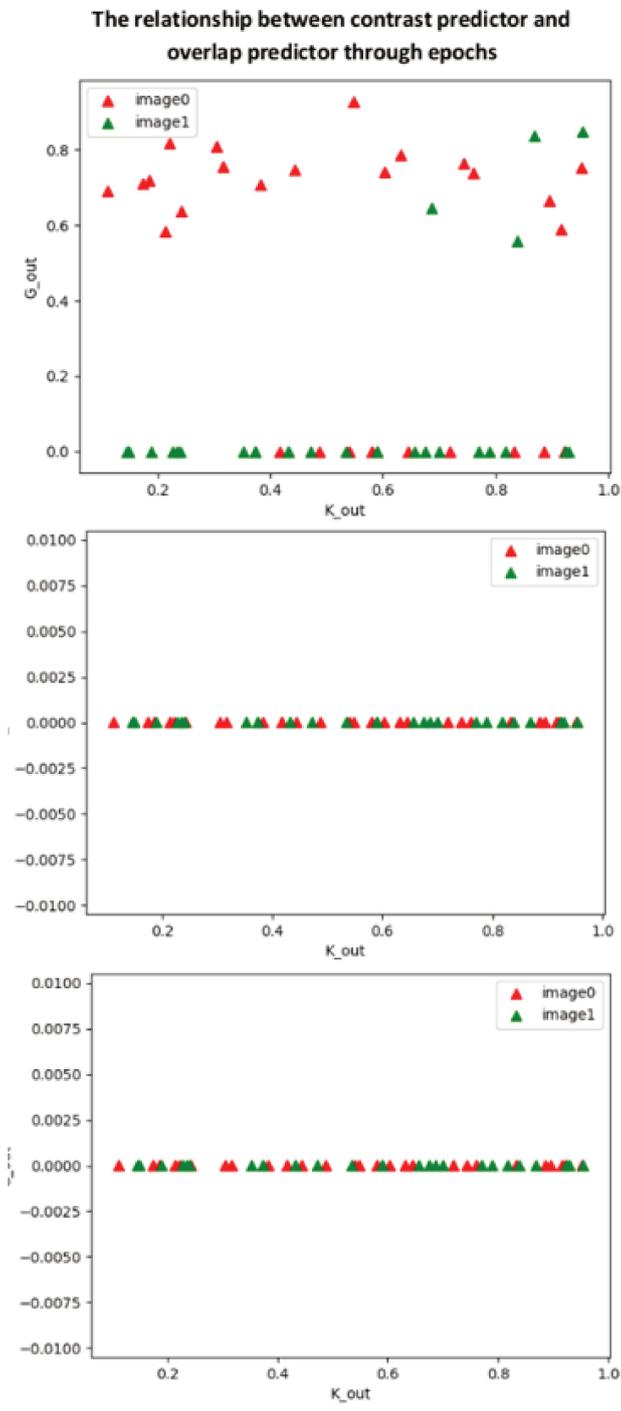


Fig. IV.35 The algorithm is not performing the regression on the training data for less than 4 images.

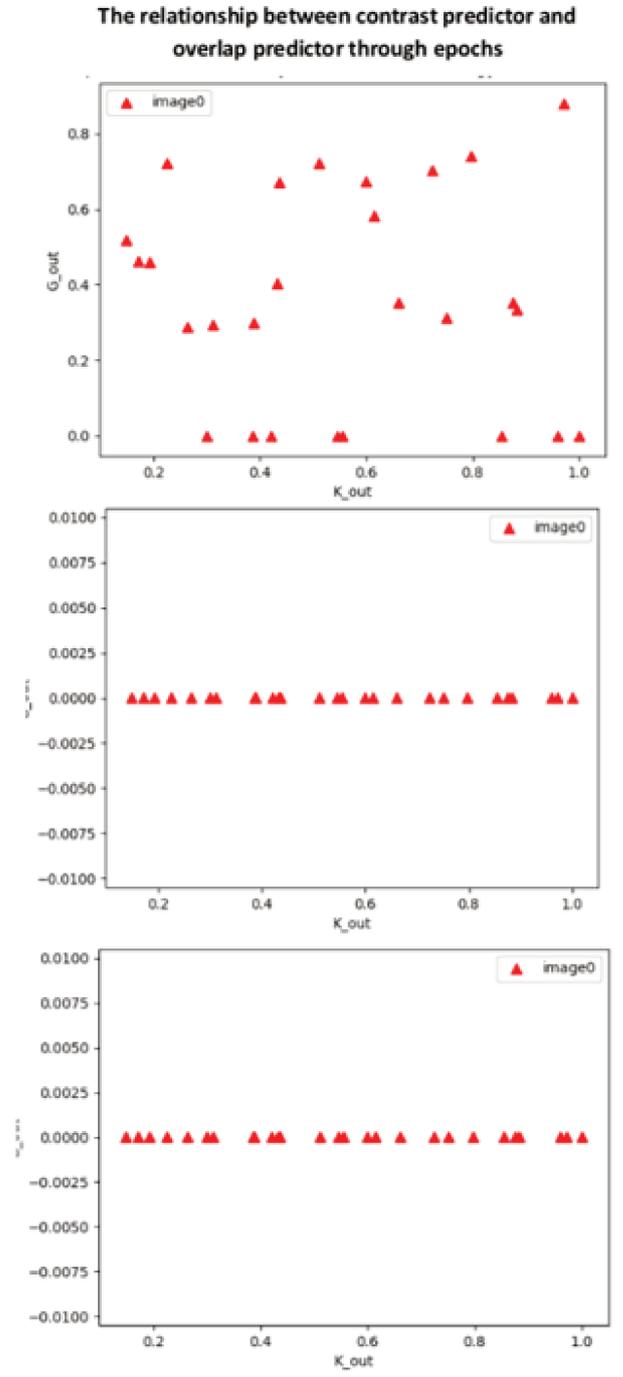


Fig. IV.36 The algorithm is not performing the regression on the testing data for less than 2 images.

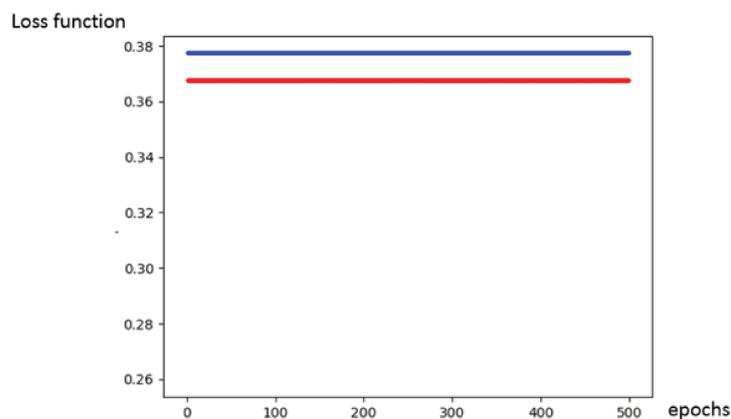


Fig. IV.37 Training loss and validation loss for only 2 training and one testing images. The blue line shows the validation loss and the red line shows the training loss. The algorithm cannot train in less than 4 images for training.

decide about the belonging of each generated data to a class (fake or real). To illustrate that, we give the email example, given all the words in the email, the discriminator decides if it is a spam or not. This type of neural network can learn to mimic any data distribution [91].

In this section, we discuss the similarity between our algorithm and the GANs. Indeed, both algorithm have the two main parts: discriminator and generator (figure IV.39). However, each part is ensured by a different structure in the two algorithms. For the GAN, both generator and discriminator's parts are composed of a Neural Network with two different loss functions which is not the case of the CNN-MPP algorithm.

In the CNN-MPP algorithm, the generator part is composed of the different simulations of MPP created from a prior knowledge of the shape of the structure we want to segment (the circle shape of the blood vessel) and the CNN architecture which is going to generate the different image through each epochs on which we evaluate the contrast predictor. On the other hand, the Loss function evaluates each contrast predictor generated by comparing it to the overlap predictor to decide for each generated simulation if the contrast predictor is equal to the overlap predictor and doing a feed-back control on the simulations through the epochs to make for each generated simulation the value of G equal to the value of K. Therefore, the discriminator in CNN-MPP algorithm is a regression function to assess each generated simulation.

## Training GANs

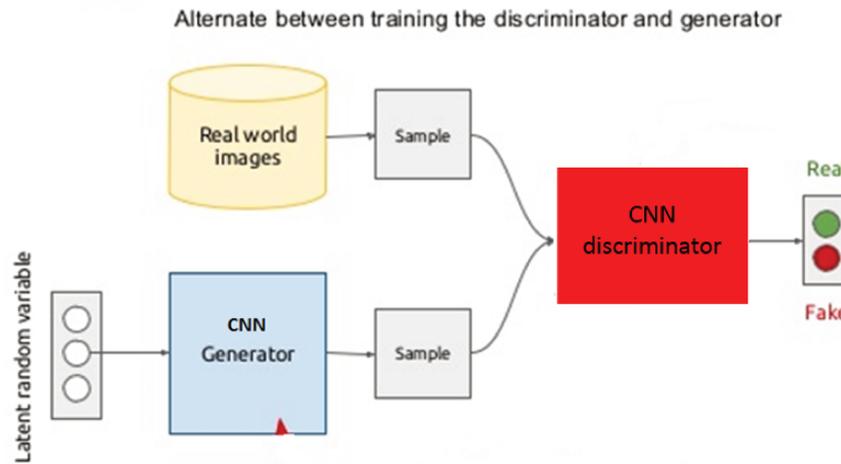


Fig. IV.38 The GAN framework.

### IV.10.4 Conclusion

In this chapter, we proposed a new algorithm to segment blood vessel on very heterogeneous data using a combination of discriminative-generative approach based on both learning driven model and based-knowledge model. Our main goal was to realise a regression on a number of generated simulations to match the contrast predictor to a overlap predictor. At this stage of our research, we tried to explain the concept, design the idea and test it on 2 dataset. However, we believe that more investigations should be done in order to test the algorithm on more histopathology data to conclude about its performance and its limits. Besides, our study is limited to the blood circle shape, thus, in future work, we aim to extend our algorithm for more complex shapes. Finally, as CNN-MPP algorithm includes a Neural Network part, we believe, that testing more architecture and an additional tuning of the hyperparameters is crucial to improve the algorithm and making it more stable. Therefore, in futur work, we need to answer the following questions:

- Which is the best Neural Network architecture to use in our model to make it stable?
- Is our model accurate in more IHC data or others *H&E* data?
- How to make CNN-MPP model useful for more complex shapes?

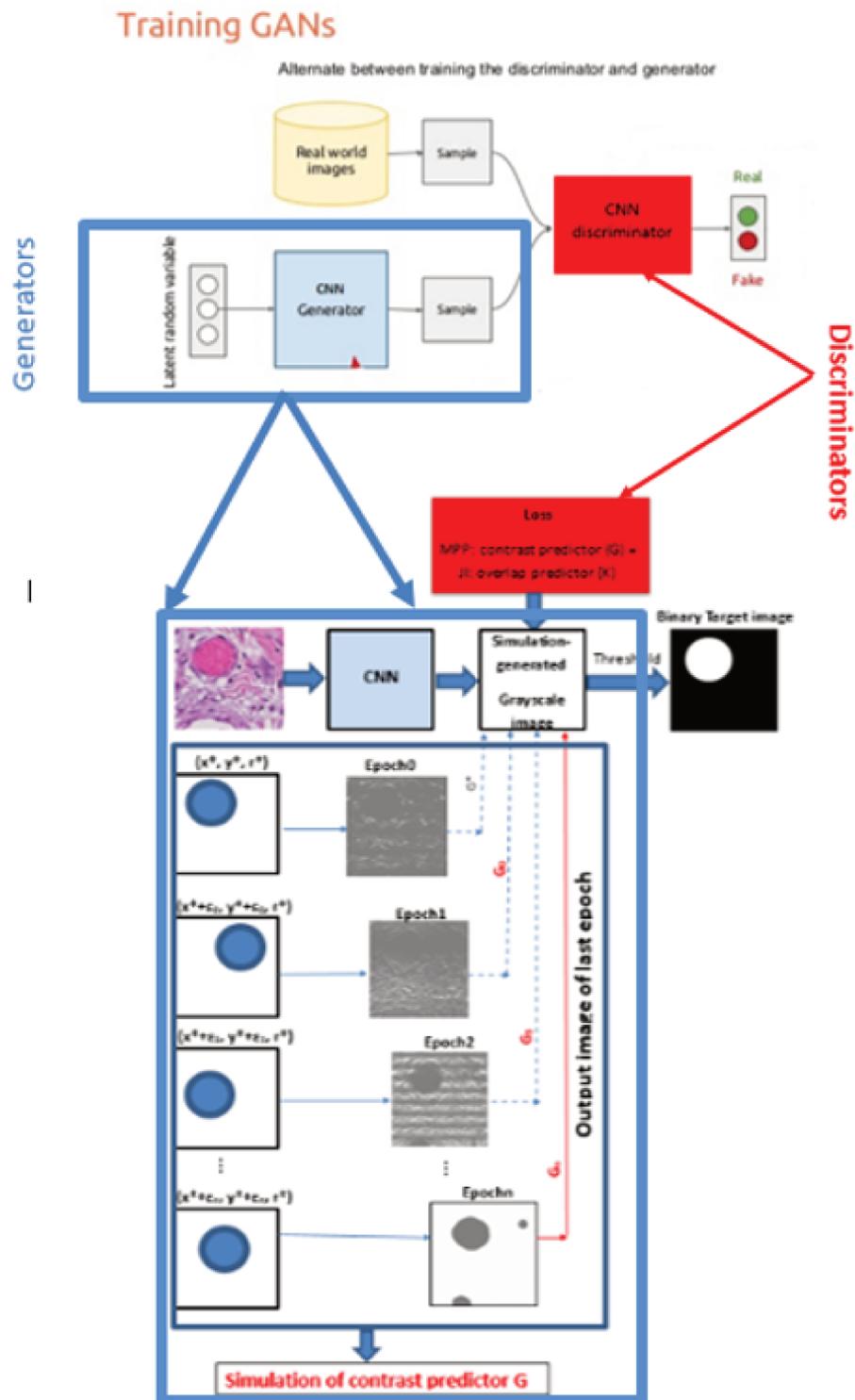


Fig. IV.39 Comparison between the GAN framework and CNN-MPP framework.



# Conclusions

We report in this manuscript the different investigations done to answer our objective which is the tumour angiogenesis assessment in breast and colorectal cancer.

Our study started with a bibliography investigation to highlight the importance of a type of transitional therapy – the anti-angiogenesis treatment – and its place as a promising anti-cancer strategy. We introduced the importance of integrative computational pathology as a very useful tool to answer histology's requests according to investigation diagnosis and treatment of cancer disease based on very heterogeneous, noisy and complex histopathology data.

Our research emphasizes the need to develop new methods to assess qualitative and quantitative tumour angiogenesis parameters for the following reasons : first anti-angiogenesis therapies was shown to improve prognosis and was shown to extend Progression Free Survival (PFS) for patients with cancer disease such as metastatic Renal Cell Carcinoma (mRcc); second, anti-angiogenesis therapies need more investigations on the tumour network and its behaviour during treatment, for effective personalized therapies. Third, those investigations are based on blood vessel network which is very complex, random and hard to assess by naked eyes by histologists. Finally, the state-of-the-art methods are either not accurate – because of the complexity of histology data – or they need a lot of labeled data – which is a very time consuming and tedious task (practically impossible to handle at a massive scale) for histopathologists.

Our work responds to the need of a multidisciplinary environment, in which cancer research is growing, nowadays. Our contribution is twofold: medical – by assessing the drug effect and giving statistic assessment – and computational – by developing new methods to help doctors in their usual framework. We believe that both sides are important to improve the communication between doctors and scientists, as to offer improved diagnosis, prognosis and care conditions to patients.

Our assessment is composed by two main parts : a first study is using pre-clinical data, where we quantitatively assess the tumour angiogenesis and a second study, using clinical data, where we qualitatively assess the tumour angiogenesis.

The first study concerns the mouse model and fluorescence images of tumour. We developed, an experimental protocol to collect pre-clinical data using a fast and non invasive whole slice image acquisition based on fluorescence. Segmentation tasks based on MRF and watershed techniques were performed to compute the statistical analysis using an information combination coming from three fluorescent stains : nuclei, endothelial cells (blood vessels) and apoptotic cells (the drug effect). Our proposed protocol aimed to quantify the ratio of apoptotic cells corresponding to endothelial cells in order to evaluate the pazopanib effect as an anti-angiogenesis drug. The results demonstrated a significant effect of the drug on the tumour blood vessels network. We confirm thus our first hypothesis emphasizing that anti-angiogenic agent pazopanib is not only a VEGF inhibitor but it has also a direct effect on the tumour blood vessel network which can be an important element to improve and guide anti-angiogenesis therapies in future work. Besides, we showed that the drug has an effect on the microenvironment components by reducing the cell density on the tumour surface.

For individual patients, optimization of dose and schedule in anti-angiogenic therapies remains a challenge. Specific phenotypic changes to be quantified during anti-angiogenic therapy are needed. Biopsies can provide us this information; however, they are difficult to obtain. New imaging technology such as fluorescence imaging can serve as reference measurements as non-invasive and reproducible imaging techniques.

The second study is based on human sample from biopsy sections. We developed an original neural network to assess qualitatively the tumour angiogenesis in very heterogeneous and complex images. Our method joint discriminative and generative pattern recognition approaches. Our algorithm shows a proof of concept to solve the data labelling problem in histopathology framework, as it uses only few data for training to resolve very complex tasks. Besides, our algorithm is useful for pure based knowledge driven model: the active contour, as a denoising pre-processing to ensure better segmentation accuracy. To our knowledge, our work is the second in term of blood vessels segmentation in histopathology images and the first work developing a new neural network architecture based on both learning data driven and rule based knowledge driven using only limited

data number. At this stage of the research, we aimed to develop a new idea and showing a proof of concept. Although, the accuracy is debating the state of art, we believe that we need to test our method on more data to conclude to the accuracy of this new architecture.

Our two studies are complementary as they correspond to pre-clinical and clinical cases using both non invasive image acquisition (fluorescence scan) and invasive techniques (biopsy).



# Bibliography

- [1] F. BRAY, J. FERLAY, I. SOERJOMATARAM, R. L. SIEGEL, L. A. TORRE & A. JEMAL; «Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries»; *CA: a cancer journal for clinicians* **68**, p. 394–424 (2018).
- [2] D. RIBATTI; «Judah Folkman, a pioneer in the study of angiogenesis»; *Angiogenesis* **11**, p. 3–10 (2008).
- [3] A. M. BYRNE, D. J. BOUCHIER-HAYES & J. H. HARMEY; «Angiogenic and cell survival functions of vascular endothelial growth factor (VEGF)»; *Journal of cellular and molecular medicine* **9**, p. 777–794 (2005).
- [4] Y. SINGH; «Tumor angiogenesis: clinical implications»; *Nepal Journal of Neuroscience* **1**, p. 61–63 (2004).
- [5] N. S. VASUDEV & A. R. REYNOLDS; «Anti-angiogenic therapy for cancer: current progress, unresolved questions and future directions»; *Angiogenesis* **17**, p. 471–494 (2014).
- [6] C. N. STERNBERG, I. D. DAVIS, J. MARDIAK, C. SZCZYLIK, J. WAGSTAFF, P. SALMAN, A. KAVINA, M. CHEN, L. PANDITE, R. E. HAWKINS *et al.*; «Pazopanib in locally advanced or metastatic renal cell carcinoma: results of a randomized phase III trial»; *Journal of Clinical Oncology* (2010).
- [7] F. MBEUNKUI & D. J. JOHANN; «Cancer and the tumor microenvironment: a review of an essential relationship»; *Cancer chemotherapy and pharmacology* **63**, p. 571–582 (2009).
- [8] E. BIANCONI, A. PIOVESAN, F. FACCHIN, A. BERAUDI, R. CASADEI, F. FRABETTI, L. VITALE, M. C. PELLERI, S. TASSANI, F. PIVA *et al.*; «An estimation of the number of cells in

the human body»; *Annals of human biology* **40**, p. 463–471 (2013).

- [9] D. TRICHOPOULOS, F. P. LI & D. J. HUNTER; «What causes cancer?»; *Scientific American* **275**, p. 80–87 (1996).
- [10] T. IYAMA & D. M. WILSON III; «DNA repair mechanisms in dividing and non-dividing cells»; *DNA repair* **12**, p. 620–636 (2013).
- [11] V. CHOPRA; «Mutagenesis: Investigating the process and processing the outcome for crop improvement»; *CURRENT SCIENCE-BANGALORE*- **89**, p. 353 (2005).
- [12] E. L. GOODE, C. M. ULRICH & J. D. POTTER; «Polymorphisms in DNA repair genes and associations with cancer risk»; *Cancer Epidemiology and Prevention Biomarkers* **11**, p. 1513–1530 (2002).
- [13] T. REYA, S. J. MORRISON, M. F. CLARKE & I. L. WEISSMAN; «Stem cells, cancer, and cancer stem cells»; *nature* **414**, p. 105 (2001).
- [14] C. E. BRONNER, S. M. BAKER, P. T. MORRISON, G. WARREN, L. G. SMITH, M. K. LESCOE, M. KANE, C. EARABINO, J. LIPFORD, A. LINDBLOM *et al.*; «Mutation in the DNA mismatch repair gene homologue hMLH 1 is associated with hereditary non-polyposis colon cancer»; *Nature* **368**, p. 258 (1994).
- [15] M. AARNIO, R. SANKILA, E. PUKKALA, R. SALOVAARA, L. A. AALTONEN, A. DE LA CHAPELLE, P. PELTOMÄKI, J.-P. MECKLIN & H. J. JÄRVINEN; «Cancer risk in mutation carriers of DNA-mismatch-repair genes»; *International journal of cancer* **81**, p. 214–218 (1999).
- [16] J. A. JOYCE & J. W. POLLARD; «Microenvironmental regulation of metastasis»; *Nature reviews cancer* **9**, p. 239 (2009).
- [17] R. L. SIEGEL, K. D. MILLER & A. JEMAL; «Cancer statistics, 2017»; *CA: a cancer journal for clinicians* **67**, p. 7–30 (2017).
- [18] P. CORREA; «Human gastric carcinogenesis: a multistep and multifactorial process—first American Cancer Society award lecture on cancer epidemiology and prevention»; *Cancer research* **52**, p. 6735–6740 (1992).

- [19] S. BOUÉE, P. GROSCLAUDE, A. ALFONSI, V. FLORENTIN, F. CLAVEL-CHAPELON & F. FAGNANI; «Projection de l'incidence du cancer du sein en 2018 en France»; *Bulletin du Cancer* **97**, p. 293–299 (2010).
- [20] S. PANDYA & R. G. MOORE; «Breast development and anatomy»; *Clinical obstetrics and gynecology* **54**, p. 91–95 (2011).
- [21] B. WEIGELT, J. L. PETERSE & L. J. VAN'T VEER; «Breast cancer metastasis: markers and models»; *Nature reviews cancer* **5**, p. 591 (2005).
- [22] L. HUTCHINSON; «Breast cancer: challenges, controversies, breakthroughs»; (2010).
- [23] E. ELSTON & I. ELLIS; «Method for grading breast cancer.»; *Journal of clinical pathology* **46**, p. 189 (1993).
- [24] H. BLOOM & W. RICHARDSON; «Histological grading and prognosis in breast cancer: a study of 1409 cases of which 359 have been followed for 15 years»; *British journal of cancer* **11**, p. 359 (1957).
- [25] E. A. RAKHA, J. S. REIS-FILHO, F. BAEHNER, D. J. DABBS, T. DECKER, V. EUSEBI, S. B. FOX, S. ICHIHARA, J. JACQUEMIER, S. R. LAKHANI *et al.*; «Breast cancer prognostic classification in the molecular era: the role of histological grade»; *Breast Cancer Research* **12**, p. 207 (2010).
- [26] L. RICCI-VITIANI, D. G. LOMBARDI, E. PILOZZI, M. BIFFONI, M. TODARO, C. PESCHLE & R. DE MARIA; «Identification and expansion of human colon-cancer-initiating cells»; *Nature* **445**, p. 111 (2007).
- [27] B. S. OOI, Y. H. HO, K. W. EU & F. SEOW CHOEN; «Primary colorectal signet-ring cell carcinoma in Singapore»; *ANZ journal of surgery* **71**, p. 703–706 (2001).
- [28] P. C. NOWELL; «Mechanisms of tumor progression»; *Cancer Research* **46**, p. 2203–2207 (1986).
- [29] M. WANG, J. ZHAO, L. ZHANG, F. WEI, Y. LIAN, Y. WU, Z. GONG, S. ZHANG, J. ZHOU, K. CAO *et al.*; «Role of tumor microenvironment in tumorigenesis»; *Journal of Cancer* **8**, p. 761 (2017).

- [30] K.-A. NORTON, C. GONG, S. JAMALIAN & A. S. POPEL; «Multiscale agent-based and hybrid modeling of the tumor immune microenvironment»; *Processes* **7**, p. 37 (2019).
- [31] J. LI, F. CHEN, M. M. CONA, Y. FENG, U. HIMMELREICH, R. OYEN, A. VERBRUGGEN & Y. NI; «A review on various targeted anticancer therapies»; *Targeted oncology* **7**, p. 69–85 (2012).
- [32] M. J. SMYTH, S. F. NGIOW, A. RIBAS & M. W. TENG; «Combination cancer immunotherapies tailored to the tumour microenvironment»; *Nature reviews Clinical oncology* **13**, p. 143 (2016).
- [33] J. COUZIN-FRANKEL; «Cancer immunotherapy»; (2013).
- [34] J. K. ADAM, B. ODHAV & K. D. BHOOLA; «Immune responses in cancer»; *Pharmacology & therapeutics* **99**, p. 113–132 (2003).
- [35] A. ALBINI & M. B. SPORN; «The tumour microenvironment as a target for chemoprevention»; *Nature Reviews Cancer* **7**, p. 139 (2007).
- [36] S. CHOUAIB, C. KIEDA, H. BENLALAM, M. Z. NOMAN, F. MAMI-CHOUAIB & C. RUEGG; «Endothelial cells as key determinants of the tumor microenvironment: interaction with tumor cells, extracellular matrix and immune killer cells»; *Critical Reviews<sup>TM</sup> in Immunology* **30** (2010).
- [37] W. RISAU; «Mechanisms of angiogenesis»; *nature* **386**, p. 671 (1997).
- [38] S. DAL-ROS; *Dysfonction endothéliale et pathologies cardiovasculaires: rôle du stress oxydant et effets protecteurs des polyphénols végétaux*; Thèse de doctorat; Strasbourg (2009).
- [39] F. W. ORR, H. H. WANG, R. M. LAFRENIE, S. SCHERBARTH & D. M. NANCE; «Interactions between cancer cells and the endothelium in metastasis»; *The Journal of pathology* **190**, p. 310–329 (2000).
- [40] J. RAK, J. FILMUS & R. KERBEL; «Reciprocal paracrine interactions between tumour cells and endothelial cells: the ‘angiogenesis progression’ hypothesis»; *European Journal of Cancer* **32**, p. 2438–2450 (1996).

- [41] M. DELLIAN, C. ABELS, G. KUHNLE & A. GOETZ; «Effects of photodynamic therapy on leucocyte-endothelium interaction: differences between normal and tumour tissue»; *British journal of cancer* **72**, p. 1125 (1995).
- [42] O. OUDAR; «Spheroids: relation between tumour and endothelial cells»; *Critical reviews in oncology/hematology* **36**, p. 99–106 (2000).
- [43] P. P. NAWROTH & D. M. STERN; «Modulation of endothelial cell hemostatic properties by tumor necrosis factor.»; *Journal of Experimental Medicine* **163**, p. 740–745 (1986).
- [44] G. BERGERS & L. E. BENJAMIN; «Angiogenesis: tumorigenesis and the angiogenic switch»; *Nature reviews cancer* **3**, p. 401 (2003).
- [45] P. CARMELIET & R. K. JAIN; «Angiogenesis in cancer and other diseases»; *nature* **407**, p. 249 (2000).
- [46] M. SHIBUYA; «Vascular endothelial growth factor (VEGF) and its receptor (VEGFR) signaling in angiogenesis: a crucial target for anti-and pro-angiogenic therapies»; *Genes & cancer* **2**, p. 1097–1105 (2011).
- [47] G. D. YANCOPOULOS & F. W. ALT; «Developmentally controlled and tissue-specific expression of unrearranged VH gene segments»; *Cell* **40**, p. 271–281 (1985).
- [48] J. H. HARMEY & D. BOUCHIER-HAYES; «Vascular endothelial growth factor (VEGF), a survival factor for tumour cells: Implications for anti-angiogenic therapy»; *Bioessays* **24**, p. 280–283 (2002).
- [49] J. FOLKMAN; «Role of angiogenesis in tumor growth and metastasis»; *Semin Oncol* **29**, p. 8–15 (2002).
- [50] O. CREECH JR, E. KREMENTZ, R. F. RYAN & J. N. WINBLAD; «Chemotherapy of cancer: regional perfusion utilizing an extracorporeal circuit»; *Annals of surgery* **148**, p. 616 (1958).
- [51] A. STEIN, W. VOIGT & K. JORDAN; «Chemotherapy-induced diarrhea: pathophysiology, frequency and guideline-based management»; *Therapeutic advances in medical oncology* **2**, p. 51–63 (2010).

- [52] M. L. SLEVIN, L. STUBBS, H. J. PLANT, P. WILSON, W. M. GREGORY, P. J. ARMES & S. M. DOWNER; «Attitudes to chemotherapy: comparing views of patients with cancer with those of doctors, nurses, and general public.»; *Bmj* **300**, p. 1458–1460 (1990).
- [53] G. C. JAYSON, R. KERBEL, L. M. ELLIS & A. L. HARRIS; «Antiangiogenic therapy in oncology: current status and future directions»; *The Lancet* **388**, p. 518–529 (2016).
- [54] C. A. STATON, S. M. STRIBBLING, S. TAZZYMAN, R. HUGHES, N. J. BROWN & C. E. LEWIS; «Current methods for assaying angiogenesis in vitro and in vivo»; *International journal of experimental pathology* **85**, p. 233–248 (2004).
- [55] I. TADEO, G. BUENO, A. P. BERBEGALL, M. M. FERNÁNDEZ-CARROBLES, V. CASTEL, M. GARCÍA-ROJO, S. NAVARRO & R. NOGUERA; «Vascular patterns provide therapeutic targets in aggressive neuroblastic tumors»; *Oncotarget* **7**, p. 19 935 (2016).
- [56] M. FERNANDEZ-CARROBLES, I. TADEO, G. BUENO, R. NOGUERA, O. DÉNIZ, J. SALIDO & M. GARCÍA-ROJO; «TMA vessel segmentation based on color and morphological features: application to angiogenesis research»; *The Scientific World Journal* **2013** (2013).
- [57] D. M. LONG; «Capillary ultrastructure in human metastatic brain tumors»; *Journal of neurosurgery* **51**, p. 53–58 (1979).
- [58] E. BULLITT, S. AYLWARD, K. SMITH, S. MUKHERJI, M. JIROUTEK & K. MULLER; «Symbolic description of intracerebral vessels segmented from magnetic resonance angiograms and evaluation by comparison with X-ray angiograms»; *Medical Image Analysis* **5**, p. 157–169 (2001).
- [59] F. YI, L. YANG, S. WANG, L. GUO, C. HUANG, Y. XIE & G. XIAO; «Microvessel prediction in H&E Stained Pathology Images using fully convolutional neural networks»; *BMC bioinformatics* **19**, p. 64 (2018).
- [60] P. TREPAGNIER; «Asynchronous fluorescence scan»; (2003); uS Patent 6,639,668.
- [61] A. ALLALOU & C. WÄHLBY; «BlobFinder, a tool for fluorescence microscopy image cytometry»; *Computer methods and programs in biomedicine* **94**, p. 58–65 (2009).

- [62] X. CHEN, X. ZHOU & S. T. WONG; «Automated segmentation, classification, and tracking of cancer cell nuclei in time-lapse microscopy»; IEEE Transactions on Biomedical Engineering **53**, p. 762–766 (2006).
- [63] D. PARRELO, C. MUSTIN, D. BRIE, S. MIRON & P. BILLARD; «Multicolor whole-cell bacterial sensing using a synchronous fluorescence spectroscopy-based approach»; PloS one **10**, p. e0122848 (2015).
- [64] S. FOERSCH, R. KIESSLICH, M. J. WALDNER, P. DELANEY, P. R. GALLE, M. F. NEURATH & M. GOETZ; «Molecular imaging of VEGF in gastrointestinal cancer in vivo using confocal laser endomicroscopy»; Gut **59**, p. 1046–1055 (2010).
- [65] M. GOETZ, A. ZIEBART, S. FOERSCH, M. VIETH, M. J. WALDNER, P. DELANEY, P. R. GALLE, M. F. NEURATH & R. KIESSLICH; «In vivo molecular imaging of colorectal cancer with confocal endomicroscopy by targeting epidermal growth factor receptor»; Gastroenterology **138**, p. 435–446 (2010).
- [66] S. LINGALA, Y.-Y. CUI, X. CHEN, B. H. RUEBNER, X.-F. QIAN, M. A. ZERN & J. WU; «Immunohistochemical staining of cancer stem cell markers in hepatocellular carcinoma»; Experimental and molecular pathology **89**, p. 27–35 (2010).
- [67] A. E. CARPENTER, T. R. JONES, M. R. LAMPRECHT, C. CLARKE, I. H. KANG, O. FRIMAN, D. A. GUERTIN, J. H. CHANG, R. A. LINDQUIST, J. MOFFAT *et al.*; «CellProfiler: image analysis software for identifying and quantifying cell phenotypes»; Genome biology **7**, p. R100 (2006).
- [68] S. S. AL-AMRI, N. V. KALYANKAR *et al.*; «Image segmentation by using threshold techniques»; arXiv preprint arXiv:1005.4020 (2010).
- [69] E. CUEVAS, H. SOSSA *et al.*; «A comparison of nature inspired algorithms for multi-threshold image segmentation»; Expert Systems with Applications **40**, p. 1213–1219 (2013).
- [70] J. SERRA & P. SOILLE; *Mathematical morphology and its applications to image processing*; tome 2 (Springer Science & Business Media) (2012).
- [71] D. WANG; «A multiscale gradient algorithm for image segmentation using watersheds»; Pattern recognition **30**, p. 2043–2052 (1997).

- [72] S. BEUCHER & F. MEYER; «The morphological approach to segmentation: the watershed transformation»; *Mathematical morphology in image processing* **34**, p. 433–481 (1993).
- [73] H. GREENSPAN, A. RUF & J. GOLDBERGER; «Constrained Gaussian mixture model framework for automatic segmentation of MR brain images»; *IEEE transactions on medical imaging* **25**, p. 1233–1245 (2006).
- [74] Z. KATO & T.-C. PONG; «A Markov random field image segmentation model for color textured images»; *Image and Vision Computing* **24**, p. 1103–1114 (2006).
- [75] F. M. LORD & M. R. NOVICK; *Statistical theories of mental test scores* (IAP) (2008).
- [76] A. KADAMBI, C. M. CARREIRA, C.-O. YUN, T. P. PADERA, D. E. DOLMANS, P. CARMELIET, D. FUKUMURA & R. K. JAIN; «Vascular endothelial growth factor (VEGF)-C differentially affects tumor vascular function and leukocyte recruitment: role of VEGF-receptor 2 and host VEGF-A»; *Cancer research* **61**, p. 2404–2408 (2001).
- [77] J. GRIFFIN & D. TREANOR; «Digital pathology in clinical use: where are we now and what is holding us back?»; *Histopathology* **70**, p. 134–145 (2017).
- [78] L. PANTANOWITZ; «Digital images and the future of digital pathology»; *Journal of pathology informatics* **1** (2010).
- [79] M. PROCHOREC-SOBIESZEK; «Future perspectives of digital pathology»; *Nowotwory. Journal of Oncology* **66**, p. 277–284 (2016).
- [80] M. CHANDRA *et al.*; «Digital pathology slides in medical education»; *Indian Journal of Dermatopathology and Diagnostic Dermatology* **1**, p. 17 (2014).
- [81] D. N. LOUIS, H. W. VIRGIN IV & S. L. ASA; «“Next-generation” pathology and laboratory medicine»; (2011).
- [82] M. G. ROJO, V. PUNYS, J. SŁODKOWSKA, T. SCHRADER, C. DANIEL & B. BLOBEL; «Digital pathology in Europe: coordinating patient care and research efforts»; *Stud Health Technol Inform* **150**, p. 997–1001 (2009).

- [83] S. PARK, A. V. PARWANI, R. D. ALLER, L. BANACH, M. J. BECICH, S. BORKENFELD, A. B. CARTER, B. A. FRIEDMAN, M. G. ROJO, A. GEORGIU *et al.*; «The history of pathology informatics: A global perspective»; *Journal of pathology informatics* **4** (2013).
- [84] N. FARAHANI, A. V. PARWANI & L. PANTANOWITZ; «Whole slide imaging in pathology: advantages, limitations, and emerging perspectives»; *Pathol Lab Med Int* **7**, p. 23–33 (2015).
- [85] W. F. LEVER *et al.*; «Histopathology of the Skin.»; *Histopathology of the skin.* (1949).
- [86] A. C. RUIFROK, D. A. JOHNSTON *et al.*; «Quantification of histochemical staining by color deconvolution»; *Analytical and quantitative cytology and histology* **23**, p. 291–299 (2001).
- [87] A. H. FISCHER, K. A. JACOBSON, J. ROSE & R. ZELLER; «Hematoxylin and eosin staining of tissue and cell sections»; *Cold Spring Harbor Protocols* **2008**, p. pdb-prot4986 (2008).
- [88] D. C. CIREŞAN, A. GIUSTI, L. M. GAMBARDELLA & J. SCHMIDHUBER; «Mitosis detection in breast cancer histology images with deep neural networks»; dans «International Conference on Medical Image Computing and Computer-assisted Intervention», p. 411–418 (Springer) (2013).
- [89] O. RONNEBERGER, P. FISCHER & T. BROX; «U-net: Convolutional networks for biomedical image segmentation»; dans «International Conference on Medical image computing and computer-assisted intervention», p. 234–241 (Springer) (2015).
- [90] J. XU, L. XIANG, Q. LIU, H. GILMORE, J. WU, J. TANG & A. MADABHUSHI; «Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images»; *IEEE transactions on medical imaging* **35**, p. 119–130 (2016).
- [91] R. ZEMOURI, N. ZERHOUNI & D. RACOCEANU; «Deep Learning in the Biomedical Applications: Recent and Future Status»; *Applied Sciences* **9**, p. 1526 (2019).
- [92] B. L. B. Y. H. P. LECUN, Y.; «Gradient-based learning applied to document recognition»; *Proc. IEEE* p. 2278–2324 (1998).
- [93] M. U. S. J. CIRESAN, D.; «Multi-column deep neural networks for image classification»; *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition* p. 3642–3649 (2012).

- [94] M. K. M. Y. K. Y. MATSUGU, M.; «Subject independent facial expression recognition with robust face detection using a convolutional neural network»; *Neural Netw.* **16**, p. 555–559 (2003).
- [95] C. AVENEL & M. S. KULIKOVA; «Marked point processes with simple and complex shape objects for cell nuclei extraction from breast cancer H&E images»; dans «Medical Imaging 2013: Digital Pathology», , tome 8676p. 86760Z (International Society for Optics and Photonics) (2013).
- [96] E. CHOI & C. LEE; «Feature extraction based on the Bhattacharyya distance»; *Pattern Recognition* **36**, p. 1703–1709 (2003).
- [97] D. GAMERMAN & H. F. LOPES; *Markov chain Monte Carlo: stochastic simulation for Bayesian inference* (Chapman and Hall/CRC) (2006).
- [98] D. VAN RAVENZWAALJ, P. CASSEY & S. D. BROWN; «A simple introduction to Markov Chain Monte–Carlo sampling»; *Psychonomic bulletin & review* **25**, p. 143–154 (2018).
- [99] D. J. DALEY & D. VERE-JONES; *An introduction to the theory of point processes: volume II: general theory and structure* (Springer Science & Business Media) (2007).
- [100] A. BADDELEY & M. V. LIESHOUT; «Stochastic geometry models in high-level vision»; *Journal of Applied Statistics* **20**, p. 231–256 (1993).
- [101] C. LACOSTE, X. DESCOMBES & J. ZERUBIA; «Road network extraction in remote sensing by a Markov object process»; dans «Proceedings 2003 International Conference on Image Processing (Cat. No. 03CH37429)», , tome 3p. III–1017 (IEEE) (2003).
- [102] X. DESCOMBES, F. PLOURABOUÉ, A. EL BOUSTANI, C. FONTA, G. LE DUC, R. SERDUC & T. WEITKAMP; «Vascular network segmentation: An unsupervised approach»; dans «2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)», p. 1248–1251 (IEEE) (2012).
- [103] S. BASU, M. KULIKOVA, E. ZHIZHINA, W. T. OOI & D. RACOCEANU; «A stochastic model for automatic extraction of 3d neuronal morphology»; dans «International Conference on Medical Image Computing and Computer-Assisted Intervention», p. 396–403 (Springer) (2013).

- [104] S. BASU, W. T. OOI & D. RACOCEANU; «Improved marked point process priors for single neurite tracing»; dans «2014 International Workshop on Pattern Recognition in Neuroimaging», p. 1–4 (IEEE) (2014).
- [105] S. BASU & D. RACOCEANU; «Reconstructing neuronal morphology from microscopy stacks using fast marching»; dans «2014 IEEE international conference on image processing (ICIP)», p. 3597–3601 (IEEE) (2014).
- [106] X. DESCOMBES, G. MALANDAIN, C. FONTA, L. NEGYESSY & R. MOSKO; «Automatic dendrite spines detection from X-ray tomography volumes»; dans «2013 IEEE 10th International Symposium on Biomedical Imaging», p. 436–439 (IEEE) (2013).
- [107] Y. LECUN, Y. BENGIO & G. HINTON; «Deep learning»; *nature* **521**, p. 436 (2015).
- [108] M. E. VANDENBERGHE, M. L. SCOTT, P. W. SCORER, M. SÖDERBERG, D. BALCERZAK & C. BARKER; «Relevance of deep learning to facilitate the diagnosis of HER2 status in breast cancer»; *Scientific reports* **7**, p. 45 938 (2017).
- [109] W. SUN, B. ZHENG & W. QIAN; «Computer aided lung cancer diagnosis with deep learning algorithms»; dans «Medical imaging 2016: computer-aided diagnosis», , tome 9785p. 97850Z (International Society for Optics and Photonics) (2016).
- [110] K. PREUER, R. P. LEWIS, S. HOCHREITER, A. BENDER, K. C. BULUSU & G. KLAMBAUER; «DeepSynergy: predicting anti-cancer drug synergy with Deep Learning»; *Bioinformatics* **34**, p. 1538–1546 (2017).
- [111] A. KADURIN, A. ALIPER, A. KAZENNOV, P. MAMOSHINA, Q. VANHAELEN, K. KHRABROV & A. ZHAVORONKOV; «The cornucopia of meaningful leads: Applying deep adversarial autoencoders for new molecule development in oncology»; *Oncotarget* **8**, p. 10 883 (2017).
- [112] Y. GUO, Y. LIU, A. OERLEMANS, S. LAO, S. WU & M. S. LEW; «Deep learning for visual understanding: A review»; *Neurocomputing* **187**, p. 27–48 (2016).
- [113] A. RADFORD, L. METZ & S. CHINTALA; «Unsupervised representation learning with deep convolutional generative adversarial networks»; arXiv preprint arXiv:1511.06434 (2015).

- [114] F. MILLETARI, N. NAVAB & S.-A. AHMADI; «V-net: Fully convolutional neural networks for volumetric medical image segmentation»; dans «2016 Fourth International Conference on 3D Vision (3DV)», p. 565–571 (IEEE) (2016).
- [115] T. MIKOLOV, M. KARAFIÁT, L. BURGET, J. ČERNOCKÝ & S. KHUDANPUR; «Recurrent neural network based language model»; dans «Eleventh annual conference of the international speech communication association», (2010).
- [116] H. LEE, R. GROSSE, R. RANGANATH & A. Y. NG; «Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations»; dans «Proceedings of the 26th annual international conference on machine learning», p. 609–616 (ACM) (2009).
- [117] A. MOOR, C. GUEVARA, H. ALTERMATT, R. WARTH, R. JAGGI & S. AEBI; «PRO\_10—A new tissue-based prognostic multigene marker in patients with early estrogen receptor-positive breast cancer»; *Pathobiology* **78**, p. 140–148 (2011).
- [118] G. D. FINLAYSON & G. Y. TIAN; «Color normalization for color object recognition»; *International Journal of Pattern Recognition and Artificial Intelligence* **13**, p. 1271–1285 (1999).
- [119] M. VETA, P. J. VAN DIEST, R. KORNEGOOR, A. HUISMAN, M. A. VIERGEVER & J. P. PLUIM; «Automatic nuclei segmentation in H&E stained breast cancer histopathology images»; *PloS one* **8**, p. e70221 (2013).
- [120] S. WIENERT, D. HEIM, K. SAEGER, A. STENZINGER, M. BEIL, P. HUFNAGL, M. DIETEL, C. DENKERT & F. KLAUSCHEN; «Detection and segmentation of cell nuclei in virtual microscopy images: a minimum-model approach»; *Scientific reports* **2**, p. 503 (2012).
- [121] M. KULIKOVA, A. VEILLARD, L. ROUX & D. RACOCEANU; «Nuclei extraction from histopathological images using a marked point process approach»; dans «Medical Imaging 2012: Image Processing», , tome 8314p. 831428 (International Society for Optics and Photonics) (2012).
- [122] P. KHARAZMI, M. I. ALJASSER, H. LUI, Z. J. WANG & T. K. LEE; «Automated detection and segmentation of vascular structures of skin lesions seen in dermoscopy, with an application to basal cell carcinoma classification»; *IEEE journal of biomedical and health informatics* **21**, p. 1675–1684 (2017).

- [123] Y. LIANG, F. WANG, D. TREANOR, D. MAGEE, G. TEODORO, Y. ZHU & J. KONG; «Liver whole slide image analysis for 3D vessel reconstruction»; dans «2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)», p. 182–185 (IEEE) (2015).
- [124] C.-H. TSOU, Y.-C. LU, A. YUAN, Y.-C. CHANG & C.-M. CHEN; «A heuristic framework for image filtering and segmentation: application to blood vessel immunohistochemistry»; *Analytical Cellular Pathology* **2015** (2015).
- [125] J. N. KATHER, A. MARX, C. C. REYES-ALDASORO, L. R. SCHAD, F. G. ZÖLLNER & C.-A. WEIS; «Continuous representation of tumor microvessel density and detection of angiogenic hotspots in histological whole-slide images»; *Oncotarget* **6**, p. 19 163 (2015).
- [126] S. BASU, W. T. OOI & D. RACOCEANU; «Neurite tracing with object process»; *IEEE transactions on medical imaging* **35**, p. 1443–1451 (2016).
- [127] L. YU, H. CHEN, Q. DOU, J. QIN & P.-A. HENG; «Automated melanoma recognition in dermoscopy images via very deep residual networks»; *IEEE transactions on medical imaging* **36**, p. 994–1004 (2017).
- [128] Y. YUAN, M. CHAO & Y.-C. LO; «Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance»; *IEEE transactions on medical imaging* **36**, p. 1876–1886 (2017).
- [129] N. C. CODELLA, Q.-B. NGUYEN, S. PANKANTI, D. GUTMAN, B. HELBA, A. HALPERN & J. R. SMITH; «Deep learning ensembles for melanoma recognition in dermoscopy images»; *IBM Journal of Research and Development* **61**, p. 5–1 (2017).
- [130] R. REAL & J. M. VARGAS; «The probabilistic basis of Jaccard’s index of similarity»; *Systematic biology* **45**, p. 380–385 (1996).
- [131] M. KASS, A. WITKIN & D. TERZOPOULOS; «Snakes: Active contour models»; *International journal of computer vision* **1**, p. 321–331 (1988).
- [132] A. YEZZI, S. KICHENASSAMY, A. KUMAR, P. OLVER & A. TANNENBAUM; «A geometric snake model for segmentation of medical imagery»; *IEEE Transactions on medical imaging* **16**, p. 199–209 (1997).

- [133] L. HE, Z. PENG, B. EVERDING, X. WANG, C. Y. HAN, K. L. WEISS & W. G. WEE; «A comparative study of deformable contour methods on medical image segmentation»; *Image and vision computing* **26**, p. 141–163 (2008).