

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

---

School of Science  
Department of Physics and Astronomy  
Master Degree in Physics

# Artificial Intelligence Applications in Digital Onco-hematology

Supervisor:  
Prof. Gastone Castellani

Submitted by:  
Alessandro Lapi

Co-supervisor:  
Dr. Nico Curti

Anno Accademico 2022/2023

# Abstract

Artificial intelligence and digital pathology have revolutionized histopathology by converting images into high-resolution digital formats and introducing computational methods for image analysis. This thesis contributed to this context by aiming to stratify patients with various hematological conditions based on bone marrow fibrosis. The study involved the quantification of the percentage of bone marrow fibrosis on the hematopoietic tissue, used as indicator for characterizing different hematological disorders on a dataset of 1123 whole slide images of bone marrow biopsies. Hematopoietic tissue has been segmented through a color decomposition technique, while the segmentation of the fibrosis needed a more powerful segmentation model, which was trained on an active semi-supervised learning (ASSL) pipeline. Thanks to the ASSL approach it was possible to overcome the lack of annotated patches, resulting in more effective segmentation. Masks obtained from segmentation were used to calculate fibrosis and tissue areas. Final comparative analysis between control and disease revealed lower fibrosis percentages in the control group. With respect to control, Acute Myeloid Leukemia and Myelodysplastic Syndrome showed no significant differences, while Myeloproliferative Neoplasms exhibited significant differentiation, underscoring the role of bone marrow fibrosis in these conditions.

# Contents

<b>Introduction</b>	<b>1</b>
<b>1 Images</b>	<b>3</b>
1.1 Digital images . . . . .	4
1.1.1 Image acquisition and formation . . . . .	4
1.1.2 Image representations . . . . .	6
1.1.3 Color images . . . . .	8
1.1.4 Image formats . . . . .	10
1.2 Digital images in pathology . . . . .	12
1.2.1 Tissue visualization . . . . .	13
1.2.2 Format and image structure . . . . .	16
<b>2 Radiomics</b>	<b>18</b>
2.1 Segmentation . . . . .	21
2.1.1 Segmentation techniques . . . . .	22
2.2 Features extraction . . . . .	27
2.3 Analysis . . . . .	29
<b>3 Artificial Neural Networks</b>	<b>32</b>
3.1 Artificial Neural Networks . . . . .	32
3.1.1 Training and Testing . . . . .	35
3.1.2 Learning approaches . . . . .	36
3.2 Convolutional Neural Networks . . . . .	39
3.2.1 U-Net . . . . .	41

3.3	Vision Transformers . . . . .	43
3.3.1	Swin Transformer . . . . .	46
3.4	Swin-Unet . . . . .	47
<b>4</b>	<b>Materials and Methods</b>	<b>49</b>
4.1	Dataset . . . . .	49
4.1.1	Digital Pathology . . . . .	50
4.1.2	Color Decomposition Preprocessing . . . . .	52
4.2	Segmentation . . . . .	54
4.2.1	Tissue Segmentation . . . . .	54
4.2.2	Bone Marrow Fibrosis Segmentation . . . . .	55
4.3	Analysis . . . . .	60
<b>5</b>	<b>Results</b>	<b>61</b>
5.1	Preprocessing . . . . .	61
5.2	Segmentation . . . . .	61
5.3	Analysis . . . . .	63
5.3.1	Overview . . . . .	63
5.3.2	Comparative Analysis . . . . .	66
<b>6</b>	<b>Discussion</b>	<b>74</b>
<b>7</b>	<b>Conclusions</b>	<b>78</b>
	<b>Bibliography</b>	<b>80</b>

# Introduction

In recent years, the intersection of medical image analysis, radiomics, and deep learning has paved the way for transformative advancements in the field of pathology [2]. Digital pathology represents a paradigm shift in the way pathologists diagnose diseases by embracing the power of digitization. At its core, digital pathology involves the conversion of traditional glass pathology slides into high-resolution digital images, which can be viewed and analyzed on computer screens. Exploring the profound impact of cutting-edge technologies on the way we understand, diagnose and treat cancer, the integration of artificial intelligence algorithms with digital pathology not only promises to improve the efficiency and accuracy of cancer diagnosis, but also has the potential to revolutionize personalized treatment strategies. As the reliance on digital pathology continues to grow, the integration of sophisticated computational techniques becomes imperative for extracting meaningful insights from histological digital slides. This thesis aims to fit in this field, with a comprehensive study conducted on a dataset comprising 1123 WSIs of bone marrow biopsies, to quantify the Bone Marrow Fibrosis in order to use it as an indicator to stratify patients with several onco-hematological disorders.

The first chapter of this thesis explores the evolution of digital images in pathology, tracing the trajectory from traditional glass slides to high-resolution digital counterparts. Chapter 2 delves into radiomics, an emerging field that translates medical images into mineable data. The general pipeline of a radiomics workflow is dissected, from image acquisition and segmentation to feature extraction and analysis. With a solid understanding of radiomics, Chapter 3 navigates the landscape of Artificial Neural Networks within the context of computer vision. This chapter overviews the most widely used methods, such as convolutional neural networks (CNNs), and Vision Transformers

(ViTs), elucidating their applications in medical image analysis. The fourth chapter illustrates materials and methods used in this thesis. Special attention is given to the segmentation of bone marrow fibrosis, which involves the implementation of an active semi-supervised learning pipeline, addressing the problem of having no annotated data. Chapter 5 shows the results, discussed then in the following chapter.

# Chapter 1

## Images

An image is a visual representation of objects, scenes, or concepts. In contemporary society, images are everywhere and serve different functions across various fields. They are employed for communication, artistic expression, entertainment, marketing, and numerous other purposes.

Since the discovery of X-rays, images have also assumed a pivotal role in the field of medicine. The use of medical imaging has revolutionized healthcare, providing indispensable tools in various medical applications and diagnostics.

Images in medicine have had an even greater impact with the advent of digitization. The first computed tomography scan performed on a patient dates back to 1971. Since then, technological advancements in image acquisition, storage, transfer, and analysis have led to the generation of huge amounts of data. Nowadays, the use of automated systems has become necessary to produce accurate and fast results from images, supporting clinicians in nearly every aspect of their work, including diagnosis, treatment planning, and real-time image-guided interventions.[4]

Next paragraphs include the main aspects of digital and medical images needed for the comprehension of the following chapters.

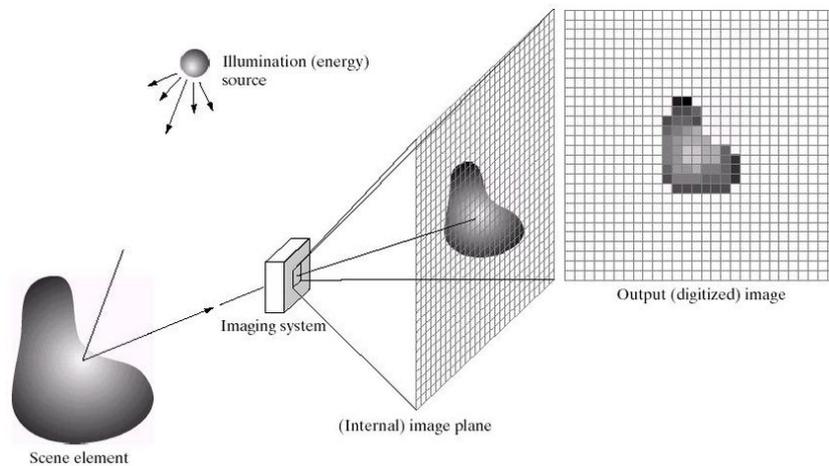


Figure 1.1: This scheme represents the process of acquisition and formation of images. [21]

## 1.1 Digital images

Digital images could be defined as a discrete numerical representation of an object. By a mathematical point of view, an image is a two-dimensional function,  $f(x, y)$ , where  $x$  and  $y$  are *spatial coordinates*, and the amplitude of  $f$  at any pair of coordinates  $(x, y)$  is the *intensity* of the image at that point. When the spatial coordinates  $x$  and  $y$ , and the intensity values of  $f$ , are all finite, i.e. discrete quantities, the resulting representation is termed a digital image [21].

In order to better understand the definition, may it is worth to spend few words about image acquisition and formation.

### 1.1.1 Image acquisition and formation

In the process of image acquisition, there are three elements involved:

- Energy source
- Scene element
- Imaging system

The energy source, also referred to “illumination”, at first glance makes us thinking about visible light. However it is a broader concept, as there are many alternatives like X-rays, infrared or less traditional sources, such as ultrasound or even a computer-generated illumination pattern.

Similarly, the scene elements could be familiar objects, but they can just as easily be molecules, buried rock formations, or a human brain. Depending on the nature of the source, illumination energy is reflected from, or transmitted through, objects. An example in the first category is light reflected from a surface, while, in the second category, we find X-rays passing through a patient’s body for the purpose of generating a diagnostic X-ray image.

The imaging system receive the energy reflected or transmitted by the object being sensed. The incoming energy is projected on a plane and transformed into a voltage by a combination of the input electrical power and sensor material that is responsive to the type of energy being detected.

Now a digitization of the output voltage will produce the digital image. The digitization process consists of a *spatial sampling* of the image plane and the *quantization* of the intensity associated to each position within the image.

Let  $g(s, t)$  be the continuous image function which would describe an analogical image. The spatial sampling consists of dividing the space of the continuous variables  $(s, t)$  into a finite number of discrete coordinates  $(x, y)$ . The image plane will be a grid, whose elements, the *picture elements*, are usually called **pixels**.

The intensity  $g(x, y)$  associated with each pixel still spans a continuous range of values. Quantization discretizes these values. The brightness of each pixel varies between two finite boundary values, typically associated with black and white. The interval between them is divided into a set of discrete numbers, called **grey-levels**. The digitization of the analogical image  $g(s, t)$  takes to the digital image  $f(x, y)$ , in which there can be found the characteristics mentioned in the previous definition of digital image.

In the process of image acquisition, digitization is closely related to image quality. The spatial quality of an image is influenced by the number of pixels it contains, and the process for determining the appropriate quantity follows the Nyquist-Shannon sampling theorem [49]. This theorem establishes a minimum requirement for the spatial sampling

rate, which should be at least half the size of the smallest detail in the image. Adhering to this principle helps prevent artifacts resulting from inadequate sampling. The **spatial resolution** of an image, which characterizes the smallest observable detail within it, is determined by both the number and dimensions of its pixels. The most common metric used to measure spatial resolution is *dpi*, i.e. dots per inch.

As for the spatial resolution, the intensity resolution can be defined as the smallest change in the intensity level that can be perceived. It is described by the **dynamic range**, a quantity which describes the ratio between the highest and lowest value of the image intensities displayed.

### 1.1.2 Image representations

The more common representation of images shows  $f(x, y)$  as it would be on a computer screen or in a photograph. Each point's intensity in the display is proportional to its  $f$  value. This kind of representation allows us to view results at a glance.

However, in digital image processing there are also other very useful representations, like matrices and histograms.

In **matrix representation** each element of the matrix corresponds to a pixel in the image, and the value of that element represents the gray-level of the pixel. So, let  $f(x, y)$  be an  $M \times N$  image. Then its matrix representation will be an  $M \times N$  numerical array:

$$f(x, y) = \begin{bmatrix} f(0, 0) & f(0, 1) & \dots & f(0, N - 1) \\ f(1, 0) & f(1, 1) & \dots & f(1, N - 1) \\ \vdots & \vdots & & \vdots \\ f(M - 1, 0) & f(M - 1, 1) & \dots & f(M - 1, N - 1) \end{bmatrix} \quad (1.1)$$

The matrix representation simplifies various image processing tasks, as you can apply mathematical operations and filters directly to the image matrix.

The dimensions of the matrix  $M, N$ , are clearly related to the spatial sampling referred to in the previous paragraph, which determines the size of the digital image. The same applies to the quantization of intensities. Generally the number  $L$ , levels of intensities, is chosen from powers of 2:

$$L = 2^k \quad (1.2)$$

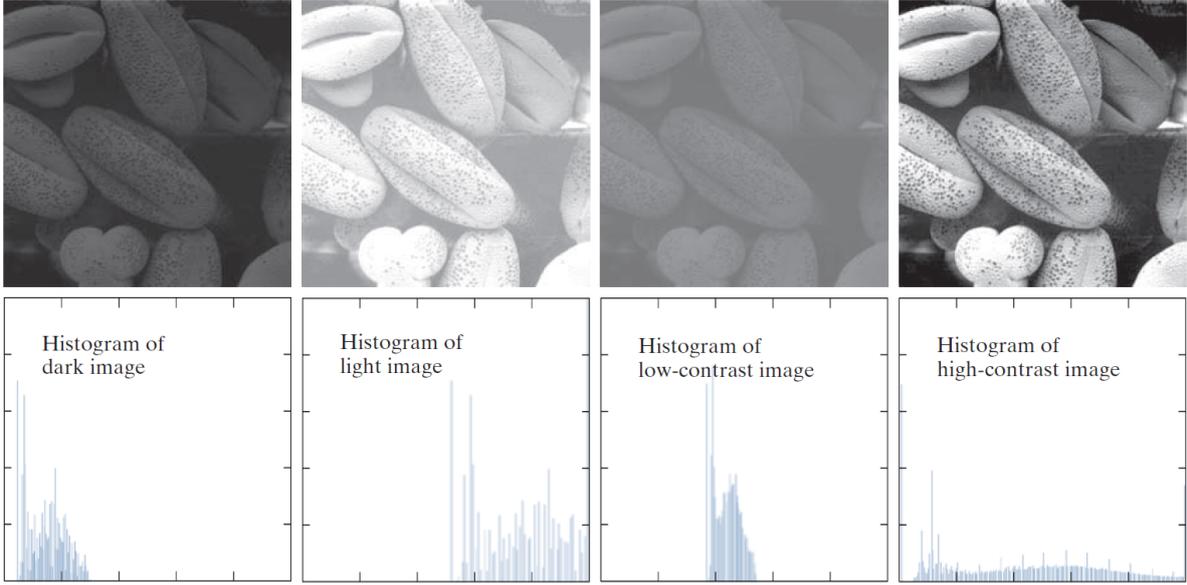


Figure 1.2: Different variants of the same image, each featuring its associated histogram below. [21]

so that the possible gray-levels of image pixels span the interval  $[0, 2^k - 1]$ . This brings us to the introduction of the **gray-level histogram**, a fundamental tool in image processing. By simply counting the occurrences of the values within the matrix we can easily construct the histogram. In fact, the histogram exhibits on the horizontal axis the  $L$  values mentioned above, while on the vertical axis, bars represent the number of pixels that have that intensity within the entire image. Thus, let  $n_k$  be the number of pixels with intensity  $l_k$ ,  $k = 0, \dots, L - 1$ , then the histogram is:

$$h(l_k) = n_k \quad (1.3)$$

We usually refer to the gray-level histogram what would actually be the *normalized histogram*:

$$p(l_k) = \frac{h(l_k)}{MN} = \frac{n_k}{MN} \quad (1.4)$$

with  $M$ ,  $N$  dimensions of the image. Here  $p$  indicates the probability of finding a pixel with gray-level  $l_k$  within the image.

Figure 1.2 provides valuable insights into the informations related to the main characteristics of images, contained within their histograms. Specifically, you can notice that dark images are characterized by a massive population of the lower intensity bins of the histogram. This observation aligns with our expectation, as in dark images, most pixels will possess lower gray-level values. For the same reason, in light images, the histogram shows a higher concentration in the bins associated with higher intensities. High-contrast or low-contrast image cases are also shown in Figure 1.2. As can be seen in histograms, the high-contrast images are characterized by a spread of pixel values over almost all available bins, while the low-contrast images show a populating of a few bins close to each other.

### 1.1.3 Color images

What has been described so far referred to grayscale images. To extend the discussion to color images, it may be useful to briefly recall some notions of color vision and color theory.

The **trichromacy theory**, initially proposed by Thomas Young [58] and later developed and confirmed by Hermann von Helmholtz [25], forms the cornerstone of our understanding of color vision. This theory elucidates how the human visual system perceives and differentiates a vast spectrum of colors.

According to the trichromacy theory, the human eye possesses three types of photoreceptor cells, commonly known as *cones*, each of which is sensitive to a specific range of wavelengths in the visible light spectrum.

Over the years, extensive experimental studies starting from Grassmann [23] and Maxwell [39] to modern research in molecular biology have provided substantial evidence in support of the trichromatic nature of human color vision.

The trichromacy theory, in conjunction with empirical studies, establishes that the human visual system can perceive any color by decomposing it into a combination of three primary colors. This insight has profound implications not only for our understanding of color vision but also for the development of **color models**.

Nowadays there are numerous color models. The most used ones in image processing are:

**RGB** : a combination of monochromatic primaries corresponding to red(R) = 700nm, green(G) = 546.1 nm and blue(B) = 435.8 nm. It is widely used in modern digital images.

**CMY** : secondary colors like cyan(C), magenta(M) and yellow(Y) are used as base. Assuming normalized values, the conversion from *CMY* to *RGB* is:

$$\begin{bmatrix} C \\ M \\ Y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

**HSV** : this model does not take in consideration pure colors. It uses hue(H), saturation(S) and value(V) to describe the possible combinations. Respectively, the three values represent the dominant color, its purity and its brightness on grey scale.

**Lab** : which combines *RGB* and *HSV* taking as variables the brightness(L), red-green contents(a) and yellow-blue contents(b).

Figure 1.3 shows the Cartesian representation of the *RGB* system. Each point in the *RGB* space is a combination of the primaries as can be seen by simply projecting the point along the axes. Assuming the values normalize to 1 we can model some common colors:

$$\begin{array}{lll} \text{red}=(1,0,0) & \text{green}=(0,1,0) & \text{blue}=(0,0,1) \\ \text{black}=(0,0,0) & \text{white}=(1,1,1) & \text{grey}=(0.5,0.5,0.5) \\ \text{cyan}=(0,1,1) & \text{magenta}=(1,0,1) & \text{yellow}=(1,1,0) \end{array}$$

as we can see the previous conversion from *CMY* to *RGB* is consistent with this definition.

As we can see in Figure 1.3, the 1D grayscale space is the diagonal connecting black to white of the cube in the RGB 3D space. Space dimensionality leads us to understand the extension of what was written before about image representation to color images, since

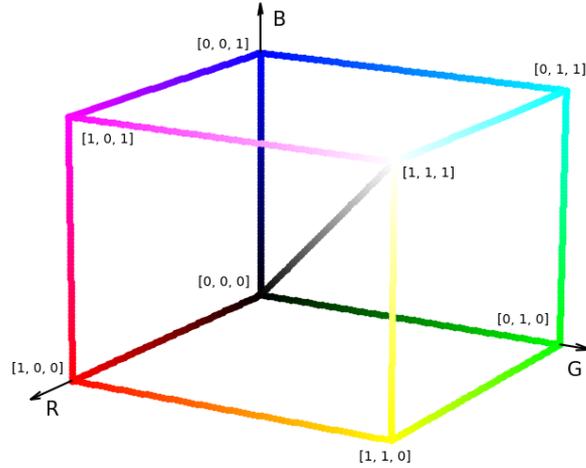


Figure 1.3: Cartesian representation of the RGB color model. To ensure the readability of the plot, only the colors on the edges and the gray-scale diagonal are shown.

the image matrix (1.1) becomes now a tensor  $M \times N \times 3$ , composed by each channel's own  $M \times N$  matrix. As well as the matrix, also the gray-level histogram is substituted by three histograms, one for each channel.

### 1.1.4 Image formats

*Image file formats* are the standard way in which digital image data is organized and stored within a file. A format specifies the structure of the data, how it is encoded, and whether compression is used [21]. These aspects are fundamental because they are strictly related with the usage of an image: the aim is always to find the optimal solution for having an image with the required quality that is as manageable as possible.

The first aspect to analyze is the image **file size**, which, expressed in bits ( $b$ ), is

$$b = M \times N \times k, \quad (1.5)$$

where  $M$  and  $N$  are the dimensions of the image, and  $k$  is the power of 2 that defines the range of intensities (see eq. (1.2)).

The type of images is usually defined through  $k$ , in fact we refer to  $k$ -bit images as a first characterization of them. Some examples are the following:

- Binary images: in this case  $k = 1$ . This is the simplest type of digital images. They use only two colors, typically black and white (or 0 and 1), to represent objects or shapes in an image. In image processing, binary images are usually used for tasks like image segmentation.
- Gray-scale images: as mentioned in the previous paragraphs, these images use shades of gray to represent brightness at each pixel. The most common gray-scale images are the *8-bits* images, i.e. the pixel values range from 0 (black) to 255 (white).
- Color Images: as already said in paragraph 1.1.3, according to the chosen model, they typically use a combination of three color channels. In this case the  $k$  bits are split among the channels; e. g. in *24-bits* images we have 8 bits for each channel.

The  $k$ -bit classification is suitable for images with discrete color or intensity. *Floating-point images* differ from them as they store intensity as floating-point numbers within a defined precision range, rather than integer values. They are often used for scientific or medical images to represent measurement values.

The other fundamental aspect in the choice of the image format is **compression**. It allows to efficiently store and transmit visual data. There are two primary approaches to image compression: *lossless* and *lossy techniques*.

Lossless compression techniques aim to reduce the size of an image file without sacrificing any image quality. These methods exploit redundancy and use various algorithms to represent the data more efficiently. When a losslessly compressed image is decompressed, it is an exact replica of the original, with no loss of information. This is crucial in applications where preserving every detail is necessary.

Lossy compression, on the other hand, achieves higher compression ratios by selectively discarding some image data that the human eye is less sensitive to. This can include removing fine details, reducing color precision, or optimizing the representation of image elements. While lossy compression significantly reduces file sizes, it comes at the cost of some loss in information.

The most common image formats and their properties are:

- JPEG (Joint Photographic Experts Group): it is commonly used for photographs, web images, digital cameras. JPEG uses a lossy compression algorithm that sacrifices some image quality to reduce file size.
- PNG (Portable Network Graphics): it is used for web graphics, logos, images with transparency. PNG uses lossless compression, preserving all image data without sacrificing quality. It typically results in larger file sizes compared to JPEG.
- BMP (Bitmap): it is mainly used for storing raw pixel data. BMP is often uncompressed, meaning it stores every pixel's color information without any compression. As a result, BMP files tend to be very large.
- TIFF (Tagged Image File Format): it is a versatile format that can support both lossless and lossy compression.
- DICOM (Digital Imaging and Communications in Medicine): it is a standard for managing medical images and associated data, ensuring compatibility and interoperability across various healthcare systems. It includes structured metadata, supports lossless and lossy compression, and prioritizes security for patient data, making it essential for medical imaging

## 1.2 Digital images in pathology

In pathology, biopsy analysis is traditionally considered the gold standard for cancer diagnosis. Tissue-based diagnostics have long been dependent on optical microscopy, specifically bright field microscopy. Digital pathology has revolutionized this field, guided by the introduction of the first **Whole Slide Imaging** (WSI) scanners in late '90. These scanners are basically robotic-controlled microscopes, equipped with highly specialized cameras featuring high-performance photosensors [45].

Before those scanners digital images in pathology were only static images, captured by simply placing a camera in front of the microscope. This technique has limitations,

mainly because it is possible to select only limited region of interest and photograph them [45].

WSI solves this problem by scanning entire glass slides, giving as output an image file that is a digitized reproduction of the glass slide, (complete with full objective lens zooming functionality) rather than a small microscopic field [45]. This capability, often referred to as "virtual microscopy" [18], enhances both the diagnostic and research processes, providing pathologists with a versatile tool for in-depth examination.

Digital pathology allows also the remote review of slides by different pathologists in real-time, a concept known as "telepathology" [17]. It facilitates collaboration and consultation among experts regardless of their physical locations, improving the quality and speed of diagnoses.

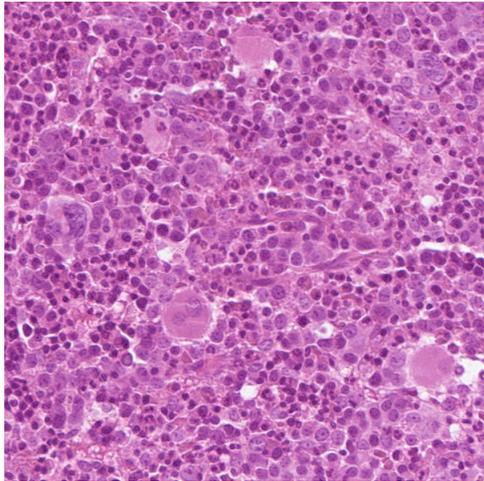
Furthermore, digital pathology opens the door to **radiomics** and **computer-aided diagnosis** (CAD) even in histopathology, through automated image analysis exploiting computer algorithms. These algorithms can extract quantitative biomarkers from WSIs, which can then be integrated with clinical data. This approach, termed "computational pathology" [2], holds the potential to enhance our understanding of diseases and improve patient care by providing valuable insights and predictive information.

The field of digital pathology is continually evolving, with a focus on improving image quality, standardization, data integration, and technological efficiency. These advancements are driving the adoption of WSI in clinical practice and research, offering new possibilities for pathology diagnosis and analysis [18].

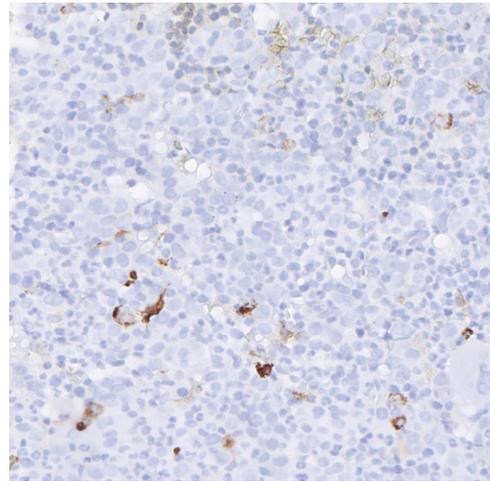
To establish the acceptance of a Whole Slide Imaging (WSI) system in clinical practice, a validation assessment is essential [16]. One crucial step in this process involves verifying that the diagnoses made using digital slides are consistent with those made using traditional glass slides. Recent studies have been conducted, yielding promising results in this regard [54].

### 1.2.1 Tissue visualization

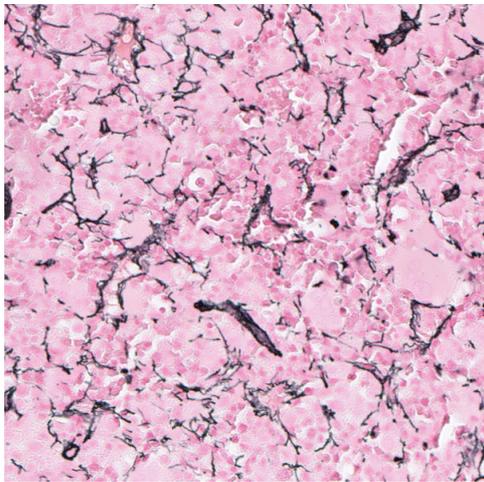
**Histological stainings** are techniques used in the field of pathology and histology to enhance the visibility and contrast of specific structures within biological tissues when viewed under a microscope. These stains involve the application of various chemical dyes



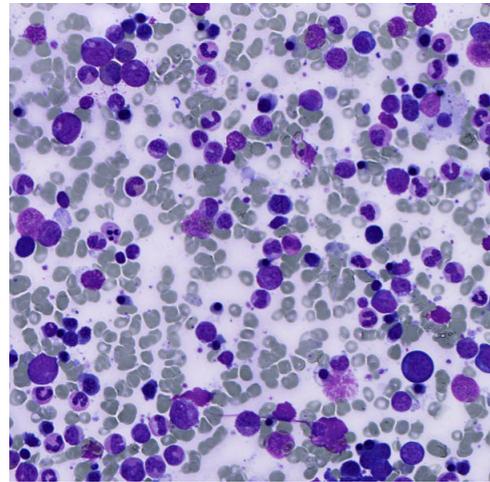
(a) Hematoxylin and Eosin (H&E)



(b) Immunohistochemical (IHC)



(c) Gomori trichrome stain



(d) May-Grünwald Giemsa (MGG)

Figure 1.4: Examples of different stains.

to tissue sections, allowing pathologists and researchers to distinguish different cell types and tissue components.

There are different types of histological stains, including:

- Hematoxylin and Eosin (H&E) [19]: H&E staining is one of the most common and widely used histological stains. Hematoxylin stains cell nuclei blue-purple, while eosin stains cytoplasm and other structures pink. This stain provides a broad overview of tissue architecture and aids in the identification of various cell types and tissue structures (Figure 1.4a).
- Immunohistochemical (IHC) [37]: IHC involves the use of antibodies that bind to specific proteins within tissue sections. This staining technique helps identify the presence and location of specific antigens or proteins within the tissue, allowing for the detection of diseases and the characterization of cellular markers (Figure 1.4b).
- Gomori [20]: Gomori trichrome staining is primarily used to examine muscle tissue and connective tissues in histological samples. It helps differentiate and highlight various structures within the tissue, making it particularly useful in diagnosing muscle diseases and assessing fibrosis (excessive collagen deposition) (Figure 1.4c).
- May-Grünwald Giemsa (MGG): it is a versatile histological and cytological staining technique used to visualize and differentiate various cellular components, including blood cells, bone marrow cells, and other tissue structures. It provides distinct coloration and contrast for different cell and tissue elements (Figure 1.4d).

Histological stains, which are commonly used in traditional glass slides, are also employed in WSIs. Stains play a significant role in computational pathology analysis. Despite existing guidelines for standard staining procedures, variations in staining between different images and even within the same image can still pose challenges to the effectiveness of quantitative analysis of digital slides. Consequently, automated algorithms are being developed to standardize colors in digital slides with the aim of enhancing the accuracy and consistency of these analyses [3].

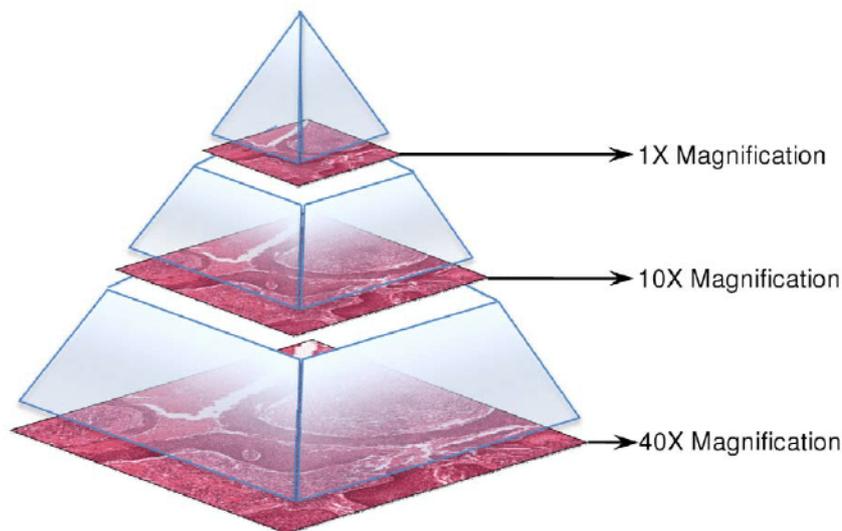


Figure 1.5: Illustration of the digital slide pyramid structure. Within a single file, a collection of images with different levels of magnification is stored. The pyramid base level represents the initial image captured by the scanning system. As one ascends the pyramid, each successive level is created by systematically reducing the resolution through downsampling techniques.

### 1.2.2 Format and image structure

Pathology is one of the last medical specialties to be digitized, largely because of the technical challenges associated with the utilization of histopathological images. The necessity for highly zoomable, extremely high-resolution, and consequently very large images, introduces complexities in various operations, including image capture, storage, viewing and annotation.

Unlike conventional digital images, which typically consist of a single view at a fixed resolution, WSI files are structured as multiresolution pyramids. These pyramids contain a multitude of images, each offering various magnifications (Figure 1.5).

The basic file format for WSIs is often the TIFF format. The key feature for which it is used extensively with WSIs is that it allows the storage of multiple images in a single file. The versatility of this format allows the use of both lossy and lossless compression techniques, however, to handle the immense size of these images, high quality lossy compression techniques based on JPEG2000 are often used. JPEG2000 is an advanced

version of JPEG, also lossy, but a bit more flexible.

However, not all vendors use these formats and compression techniques. One notable challenge in digital pathology is the absence of a universal file format standard. Every vendor tends to employ its proprietary file format. Lately, some efforts are being made to extend the DICOM format to WSIs [26], but until now, the responsibility for ensuring interoperability has been left to the manufacturers. As a result, it is not always possible to view WSI files from one vendor using another vendor's viewer.

In response to this interoperability gap, open-source initiatives like OpenSlide [22], Seeden Viewer, QPath, and Fiji Bioinformatics have emerged. These initiatives aim to fill that gap, but a universally accepted WSI file format standard has still to be fully developed.

# Chapter 2

## Radiomics

Cancer represents a global health challenge, standing as the second leading cause of death worldwide. The relentless growth of cancer cases places immense physical, emotional, and economic strain on individuals, families, communities, and healthcare systems [56]. The complex nature of cancer, characterized by high tumor heterogeneity and recurrence rates, highlights the urgent need for more effective diagnosis and treatment strategies.

In this context, medical imaging has proven to be an invaluable aid in the diagnosis, management, and monitoring of tumors. Nevertheless, the manual interpretation of medical images remains a subjective process, heavily reliant on the expertise and qualitative judgment of individual healthcare professionals. To enhance the accuracy and reliability of cancer diagnosis and treatment, there is a growing imperative to identify quantitative and consistent image biomarkers. **Radiomics**, an emerging methodology, offers a promising solution to this challenge, contributing to more objective and data-driven clinical decision-making. The name radiomics first appeared in 2012 [31] but its roots lie in earlier methods of quantitative image analysis.

Radiomics aims to extract quantitative and reproducible information from medical images, especially focusing on complex patterns that can be challenging for human observers to recognize or quantify visually [40].

Radiomic data are extracted in such a way that they can be used for data mining and for the development of descriptive and predictive models [51]. There are different types of studies, depending on the specific clinical needs they seek to address. *Diagnos-*

*tic* radiomics aims to distinguish between disease entities, helping in the classification of conditions, such as discerning benign from malignant tumors, or recognizing different subtypes. *Prognostic* radiomics is concerned with predicting patient outcomes, offering insights into disease progression, risk of recurrence, and survival rates. Radiomics for *treatment planning* is a tool for personalizing treatment approaches, particularly in radiation therapy, facilitating delineation of target volumes and optimizing treatment plans. *Delta radiomics* focuses on evaluating the efficacy of interventions by comparing pre- and post-treatment images and analyzing changes in radiomic characteristics. Each of these distinct types of radiomics studies addresses a wide range of clinical scenarios, providing insights and guidance that can be helpful to radiology and oncology professionals [33, 50, 63, 43].

The usual pipeline of radiomics can be divided in few steps: image acquisition and pre-processing, segmentation, feature extraction, analysis. Each step has its own methods and its own challenges.

- **Acquisition and pre-processing:** image acquisition, in medicine, is the process of capturing medical images using various imaging modalities such as computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), or histopathological Whole Slide Imaging (WSI). These techniques create detailed visual representations of the patient’s anatomy and pathology. However, during acquisition, factors like equipment settings and imaging protocols can introduce variability and imperfections into the images, which can influence radiomics features.

Standardized acquisition and shared pre-processing protocols are crucial to minimize these variations and ensure the quality of the acquired images, especially in terms of interoperability of different systems and reproducibility of different studies.

- **Segmentation:** it is the most critical step in the radiomic pipeline, essential for extracting quantitative information from medical images. It involves precisely identifying and isolating regions of interest (ROIs), such as tumors or organs, within

these images. The primary goal is to ensure accurate ROI definition, as the image segmentation heavily affects the radiomics features.

Segmentation methods are *manual*, *semi-automatic* or *fully automatic*. While automatic techniques are becoming more accurate, the gold standard is still the manual segmentation, since in medical images, the identification of the ROI could be a very complex task. However, humans are susceptible to errors and biases and the use of computer-aided methods can reduce them. Automation is often preferred to manual delineation, not only for reproducibility, but also for time efficiency, particularly with large datasets. Recently *Deep Learning* excels in medical image segmentation, showing great results across diverse medical imaging applications. Unlike conventional techniques dependent on handcrafted features, DL directly learns from data, capturing complex patterns that are difficult to express with explicit rules. On the other hand DL methods are usually specific to particular task and need a customized training.

- **Features extraction:** it is the heart of the radiomic pipeline, where the aim is to convert segmented regions of interest (ROIs) within medical images into a comprehensive set of quantitative features. These features, ranging from statistical properties to texture and shape characteristics, provide numerical descriptions of the ROIs. A critical challenge is posed by the lack of reproducibility and validation within the field of radiomic studies.

The extraction process in radiomics typically yields a vast number of features. Among these, certain features may prove irrelevant, while others might exhibit correlation, rendering some of them redundant. **Features selection** serves as a crucial optimization step in radiomics analysis, with the primary objective of ensuring that only pertinent information is incorporated into the process.

- **Analysis:** the focus shifts to exploiting the potential of radiomics data for various purposes. The analysis process is structured into statistical data analysis, model building and model validation. Exploratory analysis, a key component, involves unsupervised methods such as clustering and dimensionality reduction, improving understanding of the data. Model building can integrate radiomic data with

clinical or genomic information, using regression and classification. Model validation ensures model robustness and generalizability through cross-validation and independent test datasets to establish model reliability.

## 2.1 Segmentation

Let  $I$  be the entire spacial region of an image. **Image segmentation** is the process through which  $I$  is split into  $n$  sub-regions  $R_i$ , such that:

- a)  $I = \bigcup_{i=1}^n R_i$
- b)  $R_i$  is a connected set, for  $i = 0, 1, 2, \dots, n$
- c)  $R_i \cap R_j = \emptyset$  for all  $i$  and  $j$ ,  $i \neq j$
- d)  $P(R_i) = \text{TRUE}$  for all  $i$ , and  $P(R_i \cup R_j) = \text{FALSE}$  for any adjacent regions  $R_i, R_j$

Condition **a)** states the *completeness*: every pixel in the image must belong to one of the defined regions, ensuring that no pixel is left unaccounted for. **b)** focuses on *connectedness*, i.e. each region must be a connected set of pixels. The exact definition of connectedness may vary, but states that all pixels in a region are linked together in some way. Condition **c)** means that regions must be *disjoint*, i.e. distinct and non-overlapping: each pixel must be assigned to only one region. The last condition, **d)** involves the *region differentiation*: each pixel within a segmented region should satisfy specific properties represented by a logical predicate  $P$ . Adjacent regions, sharing a boundary, must exhibit distinguishable properties according to the predicate. For example, these properties may involve characteristics like intensity, color, or other attributes relevant to the segmentation task, ensuring that neighboring regions are distinguishable based on these properties.

Thus, in image segmentation, the process typically divides pixels into  $n$  subsets, usually referred to as **classes**. These classes represent distinct components within the image. However, the assignment of specific meaning or interpretation to these classes, known as **labeling**, often occurs after the segmentation process.

When there are only two classes, we talk about binary image segmentation task. In this case the labeling process can be relatively straightforward: the two classes usually correspond to the “object” and the “background”.

### 2.1.1 Segmentation techniques

Depending on specific requirements, image segmentation can be accomplished using various criteria for pixel assignment to different classes. Furthermore, the inherent variability within and across images makes it impractical to define a universal segmentation technique that fits all situations. As a result, numerous segmentation techniques exist, each following different approaches.

Some techniques are based on the principle of *similarity*, wherein segmentation involves classifying similar regions based on predefined criteria. Key algorithms falling into this category include thresholding and region-growing. Another common approach is based on the principle of *discontinuity*. Here, the objective is to identify abrupt changes among neighboring pixels in terms of intensity or other predefined criteria. Prominent algorithms in this category are edge detection methods. Additionally, there is a third category known as *special theory-based* segmentation. This category encompasses a variety of segmentation techniques and algorithms that draw from diverse fields of knowledge, such as wavelet transformation, morphology, fuzzy mathematics, genetic algorithms, and artificial intelligence. These approaches allow for specialized segmentation solutions tailored to specific image analysis tasks.

#### Similarity-based algorithms

Algorithms in this category aim to perform image segmentation by grouping pixels that share common characteristics. Further categorization can be made between *threshold-based algorithms* and *region-based algorithms*, such as region growing and split and merge. Here’s an overview of the main properties of these methods.

Image **thresholding** is popular in image segmentation applications due to its intuitive properties, ease of implementation, and computational efficiency. Let’s start with the simplest case: a high-contrast image  $f(x, y)$ , containing an object that stands out

against a relatively uniform background. In this scenario, one can easily imagine a histogram characterized by the presence of two fairly narrow and well-separated peaks. Intuitively, by selecting a threshold value ( $T$ ) between the two modes, we can obtain the following segmented image:

$$g(x, y) = \begin{cases} 1 & f(x, y) \geq T \\ 0 & f(x, y) < T \end{cases} \quad (2.1)$$

In this case, we refer to it as *global thresholding* since  $T$  remains constant across the entire image. To make things more complex, we can consider situations where the illumination is not uniform throughout the image. In such cases, it becomes necessary to divide the image into multiple regions to determine a threshold that varies according to the characteristics of the pixels in these regions. This is known as *local thresholding*, often referred to as *dynamic thresholding* when a threshold  $T(x, y)$  varies pixel by pixel.

Besides non-uniformity, another significant factor that complicates threshold-based segmentation is noise. Noise tends to widen the distributions in histograms and can lead to confusion between pixels belonging to objects and the background, due to the lack of spatial information of the histogram. For this reason, it is often necessary to apply smoothing filters to reduce noise before performing image segmentation.

The most popular image thresholding technique was presented by Otsu [44]. This method automatically determines an optimal threshold value for image segmentation. It works by finding a threshold that minimizes the intra-class variance (variance within the two classes of pixels: foreground and background) and maximizes the inter-class variance (variance between the two classes).

If an histogram exhibits three or more modes, *multiple thresholding* becomes necessary, involving the definition of two or more thresholds. However, when the number of thresholds exceeds two, it becomes very difficult, if not impossible, to determine them effectively for achieving good segmentation. In such instances, alternative methods, such as region-based approaches, are often much more effective.

**Region-based** segmentation algorithms are a class of image processing techniques that partition an image into distinct regions according to a certain criteria. These algorithms are used to group similar pixels together to form meaningful regions. The most

common examples are *region-growing*, *split and merge*.

**Region Growing** starts with one or more *seed* pixels, randomly selected or chosen according on criteria like intensity similarity. It then gradually expands the region by examining neighboring pixels. If a neighboring pixel meets the *similarity criterion*, it's added to the growing region. This process continues until a *stopping criterion* is met, such as reaching a predefined region size, a dissimilarity threshold, or the image boundaries. Region growing is particularly useful for objects with uniform regions and clear boundaries. It can adapt to the shape of objects and handle noise effectively. However, its sensitivity to seed point selection may lead to bad segmentations if the criteria are not well-defined.

**Split and Merge**, on the other hand, is an algorithm that starts with the entire image as one region and recursively splits and merges regions. It checks if a region meets a *splitting criterion* based on a local property like intensity variance or texture homogeneity. If the criterion is not met, the region is divided into smaller subregions. Conversely, if the *merging criterion* is satisfied, adjacent regions are combined into a single region. This approach allows the algorithm to adapt to non-uniform regions and various object sizes, producing more balanced regions with well-defined boundaries. However, setting appropriate splitting and merging criteria can be challenging, as the algorithm's performance heavily depends on them.

## Discontinuity-based algorithms

In images **edges** are defined as a collection of connected pixels located along the boundary that separates different regions. These edges typically occur at points of significant discontinuities in visual properties such as changes in gray-scale, color, texture, and other relevant attributes.

**Edge-based segmentation** involves the identification of image discontinuities, commonly referred to as *edge detection*. Subsequently, it aims to reconstruct these regions in a coherent manner by connecting the detected edges.

Since these methods involve the analysis of abrupt changes between pixels, they often contend with noise. To optimize algorithms of this kind and achieve effective segmentation, it is crucial to find the right balance between accuracy and noise immunity

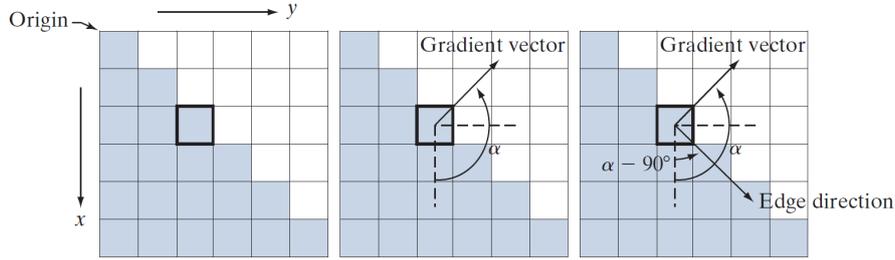


Figure 2.1: The gradient at a pixel provides information about edge strength and direction. The edge direction is always perpendicular to the gradient direction, helping identify edges in an image. [21]

[28].

The most common family of edge-based segmentation techniques uses the **gradient** operator. The gradient of an image  $f(x, y)$  at position  $(x, y)$  is defined as

$$\vec{\nabla} f(x, y) = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} = \begin{bmatrix} G_x \\ G_y \end{bmatrix} \quad (2.2)$$

while its amplitude and direction are simply given by

$$\nabla f(x, y) = (G_x^2 + G_y^2)^{1/2} \quad (2.3)$$

$$\alpha(x, y) = \tan^{-1} \left( \frac{G_x}{G_y} \right) \quad (2.4)$$

In Figure 2.1 is shown a graphical representation of the gradient, emphasizing the perpendicular direction of the edge relative to the gradient vector. This picture offers a valuable insight into how gradient analysis helps identify edge orientations.

The edge-detection exploiting the characteristics of the gradient is performed through these general steps:

1. Apply the derivative operator for edge detection
2. Measure edge strength using gradient amplitude
3. Keep edges with a magnitude above a specified threshold (eliminating weak edges).

4. Identify crack edges and decide to keep or reject them based on neighboring edges' confidence.
5. Repeat steps 3 and 4 with varying threshold values to identify closed boundaries and achieve image segmentation

This general framework forms the basis for several edge detection techniques, including the use of differential coefficients, the Laplacian of Gaussian, and the renowned Canny edge detector, introduced by J. Canny [7]. The core principles of Canny's approach are rooted in three key criteria: the pursuit of a low error rate; the achievement of precise edge localization; and the minimization of repeated responses. By incorporating an initial Gaussian filter to smooth the image, and selectively retaining points connected to strong edges, this algorithm excels in enhancing noise immunity and detecting real edges, the toughest challenges that put edge detection algorithms to the test.

### **Special theory-based algorithms**

Under this class of algorithms, following [28], there are many techniques, coming from different fields.

These algorithms can be broadly categorized into two main classes: supervised and unsupervised methods. Supervised methods, relying on annotated training data, are pivotal in tasks demanding precise object localization and classification, such as medical image analysis and autonomous driving. They encompass techniques like pixel-wise classification and semantic segmentation. On the other hand, unsupervised methods operate without the need for pre-labeled data, grouping pixels or regions based on inherent similarities. Clustering, is a fundamental concept in unsupervised learning. It seeks to identify natural groupings within datasets, allowing data to self-organize into clusters based on similarities. Clustering has widespread applications in areas like market analysis and social network analysis, making it a powerful tool for data exploration. In the context of image segmentation, clustering essentially involves classifying individual pixels into distinct attribute regions, enabling the extraction of valuable information from complex visual data. Notably, two commonly used clustering algorithms, K-means [36] and Fuzzy C-means [5], offer "hard" and "soft" segmentation, respectively, each with

unique characteristics suited for various data analysis and pattern recognition tasks.

Additionally, the landscape of image segmentation has been transformed by neural networks, specifically Convolutional Neural Networks and Vision Transformers. CNNs are adept at learning hierarchical features from images and excel in tasks like semantic segmentation. They are extensively employed in applications ranging from medical image analysis to object detection and scene understanding. Vision Transformers, a more recent innovation, harness self-attention mechanisms to capture long-range dependencies in images. Next chapter presents a deeper overview on Artificial Neural Networks and their use for image segmentation task.

## 2.2 Features extraction

The feature extraction phase is the central step in the radiomics workflow. Here, I will outline only the most commonly utilized features, while for a more comprehensive list, one may refer to the Image Biomarker Standardization Initiative (IBSI) [65] which aims to establish a standard in response to the critical challenges posed by the lack of reproducibility and validation within the field of radiomic studies.

Mayerhoefer et al. [40] classify radiomic features into four main categories, each offering unique insights into medical image analysis:

- **Statistical Features:** This category encompasses histogram-based and texture-based features. *First order statistics* features quantify the distribution of pixel intensity values within an image exploiting histogram's informations. Common metrics include mean, median, variance, skewness, and kurtosis of the intensity distribution. Texture analysis of images, first introduced in 1973 by Haralick et al. [24], exploits *second order statistics* to describe spatial patterns and variations in intensity, offering valuable information about tissue heterogeneity, an important biomarker in many tumors. The analysis involves matrices like the *gray-level co-occurrence matrix*, *gray-level run-length matrix*, and *gray-level size zone matrix*. These matrices quantify patterns, co-occurrence, and relationships between pixel intensities, aiding in the characterization of tissue heterogeneity and the identification of subtle textures within medical images.

- **Model-Based Features:** Model-based feature analysis in the context of radiomics involves using generative image models and stochastic models to interpret the texture of an image. In this approach, the parameters of the model are estimated and then employed for image analysis as radiomic features. However, in practice, a significant challenge arises from the computational complexity associated with estimating the parameters of stochastic models [38].
- **Transform-Based Features:** Transform-based features exploits mathematical transformations, such as Fourier or wavelet transforms, to extract relevant information from the image data. These features help uncover frequency domain patterns and variations within the images.
- **Shape-Based Features:** Shape-based features focus on the geometric properties of regions of interest in medical images. They are crucial for quantifying size, contour irregularities, and structural changes, making them especially relevant in tumor detection and characterization. Morphological features specifically describe the shape of a segmented region of interest and are entirely independent of the pixel values within that region. To compute morphological features, only a boolean mask of the segmented region is needed. In the category of morphological features, some key parameters are area, perimeter, compactness, and eccentricity. These parameters are fundamental for delineating structural information and identifying anomalies in anatomical shapes, making them especially valuable in contexts where changes in shape can serve as significant indicators of disease or abnormality.

Radiomics features can be also categorized on the basis of their relevance for the further analysis. *Relevant features* provide valuable information, while *irrelevant features* do not contribute to the algorithm’s process and are of no assistance. *Redundant features* offer no new insights to the learning algorithm, as their information can already be derived from other available data. This description of radiomics features is strictly related to **features selection**, an important step across feature extraction and analysis. This process is indispensable for several reasons. First, it helps in reducing the dimensionality of radiomics data, which can be extensive, making the analysis more efficient and guarding against the curse of dimensionality. Furthermore, selecting relevant features

improves the overall performance and interpretability of radiomics models, as it ensures that only the most informative data is used, minimizing noise and enhancing the quality of the results. Additionally, feature selection contributes to cost efficiency, as it allows researchers to prioritize the collection of the most critical features, saving valuable time and resources. Lastly, it facilitates the interpretability of the analysis outcomes, a particularly important factor in clinical settings. Various methods, such as *filter*, *wrapper* and *embedded techniques*, are employed for feature selection, and the choice of method depends on the specific context and dataset [60].

## 2.3 Analysis

Radiomics includes diagnostic, prognostic, treatment planning, and delta studies, each addressing distinct clinical needs like disease classification, outcome prediction, personalized treatment, and intervention efficacy assessment. The analysis phase in the radiomics pipeline is closely related to the objective of the study. Below I will outline some general techniques, which can be tailored to different specific cases.

Once the extracted features have been selected, usually the next step is a **data integration**, in which radiomic data related to the images are merged with clinical and genomic information. This approach combines information derived from medical images with the patient's clinical data, such as demographic data, medical history, medical records, and laboratory results. This integration allows the development of comprehensive models that consider the patient's overall health and history. In addition, the inclusion of genomic data provides insights into the genetic basis of disease and variability in individual patients. This approach to precision medicine, allows for the development of personalized treatment plans based on genetic markers that may influence disease progression or response to therapy. The synergy between radiomics, clinical data, and genomics promises to improve diagnosis, prognosis, and treatment planning, providing health care professionals with a more holistic view of the patient and enabling the development of personalized and optimized therapeutic strategies, early disease detection, and more informed clinical decisions.

**Explorative analysis** is an important phase that offers valuable insights into the

complex and high-dimensional radiomic data. It comprises both unsupervised techniques and dimensionality reduction methods. Unsupervised analysis includes *clustering*, which identifies natural groupings within the data. Clustering algorithms group similar data points based on their radiomic feature profiles, but also clinical and genomic data if included, allowing the identification of underlying patterns within the data. Clustering includes various methods, among which partitional clustering, hierarchical clustering, and density-based clustering are prominent. Partitional clustering involves dividing data into distinct non-overlapping subsets, with algorithms like K-means seeking to optimize specific criteria. Hierarchical clustering, on the other hand, builds a tree of clusters through recursive merging or splitting, creating a dendrogram that visually represents the cluster hierarchy. Density-based clustering, like by DBSCAN [15], identifies clusters based on data point density, making it suitable for datasets with irregular shapes and varying densities. Each clustering approach has its advantages and limitations, and the choice of algorithm depends on the dataset’s characteristics and the analytical goals. *Dimensionality reduction* techniques, such as t-Distributed Stochastic Neighbor Embedding [34] and Uniform Manifold Approximation and Projection [42], are employed to reduce high-dimensional data into lower dimensions, facilitating visualization and interpretation. This phase of analysis allows researchers to uncover hidden structures and relationships within the radiomic data, contributing to a deeper understanding of the medical images and their clinical implications.

**Model construction** depends on the specific clinical objectives. Regression models play role in predicting continuous outcomes, making them highly relevant for tasks such as estimating tumor volume, assessing changes in disease burden over time, or quantifying treatment response. On the other hand, classification models are particularly useful when the goal is to assign labels to different disease classes or subtypes. For example, they can be employed to differentiate between benign and malignant tumors, classify different cancer stages, or identify specific disease subtypes based on radiomic features. Classification models are essential for diagnostic radiomics, where accurate disease categorization is critical. With the advent of deep learning techniques, neural networks have found their place in radiomics. While deep learning has gained prominence, traditional machine learning models like Random Forest, Decision Trees, and Support Vector Ma-

chines persist in radiomics applications. Random Forest, known for its ensemble learning approach, is adept at handling both numerical and categorical data, making it suitable for feature selection and classification tasks, particularly in scenarios with a large number of features. Decision Trees, appreciated for their interpretability, can capture complex relationships in the data but may be sensitive to small variations. SVMs, with their effectiveness in high-dimensional spaces, find application in classifying and regressing radiomic data, particularly in situations where clear margins of separation between classes are crucial. Neural networks, particularly convolutional neural networks, excel in automatically extracting hierarchical features from images, making them valuable for tasks such as classification, segmentation, and feature extraction in radiomics. The choice between these models depends on factors like dataset size, problem complexity, interpretability needs, and computational resources, often leading to the adoption of hybrid approaches for optimal results in radiomics.

**Model validation** is fundamental step in the radiomics analysis process, playing a crucial role in ensuring the reliability and generalizability of the constructed models to real-world scenarios. Validation procedures assess the model's performance by evaluating its ability to make accurate predictions on data that it has not seen during training. This helps in estimating how well the model will perform on new, unseen datasets, thus avoiding overfitting to the specific characteristics of the training data. Common validation techniques in radiomics include cross-validation, where the dataset is split into subsets for training and testing iteratively, and external validation, where the model is tested on entirely independent datasets. Rigorous validation is essential to confirm the robustness of models, enhancing their utility in clinical applications and contributing to the credibility of findings in medical research.

# Chapter 3

## Artificial Neural Networks

Artificial intelligence (AI) focuses on creating systems that emulate intelligent behavior, using various approaches such as logic, search and probabilistic reasoning. Within AI, machine learning is a subset that involves training models to make decisions by adapting mathematical representations to observed data. Despite the broad scope of AI, machine learning, particularly deep learning, has mistakenly become synonymous with AI due to its explosive growth [46].

The adaptation of a deep network to data is the essence of deep learning, and currently these networks stand out as the most powerful and practical models of machine learning. Deep learning finds application in everyday scenarios, such as language translation, image recognition and voice interaction with digital assistants, demonstrating its widespread impact on modern technology.

### 3.1 Artificial Neural Networks

Artificial Neural Networks are computational models inspired by the structure and functioning of the human brain. They consist of interconnected nodes, or artificial neurons, organized into layers. Each connection between nodes has a weight associated with it, which is adjusted during the learning process.

Nodes are the fundamental units within a neural network. These artificial entities are inspired by the neurons found in the human brain and serve as the building blocks

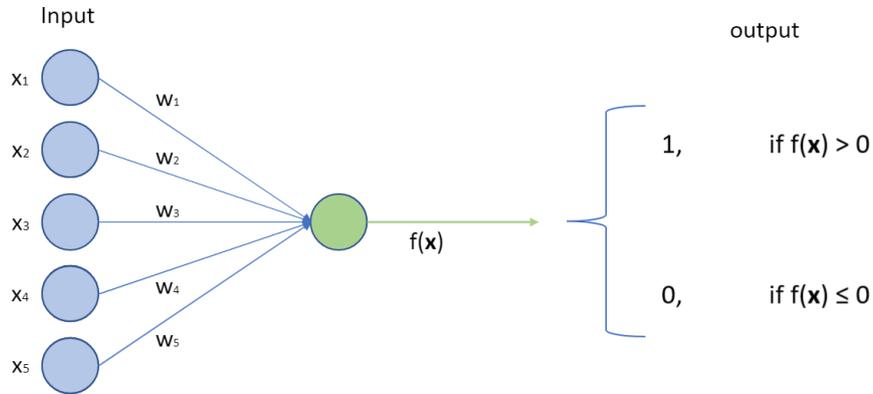


Figure 3.1: Representation of a simple neuron (green circle). The  $x_i$  elements are the inputs (blue circles) and the output is calculated based on the weighted sum of these inputs summed with a bias  $w_0$ .

of the network's architecture.

Each node is designed to process information, emulating the way biological neurons receive and transmit signals. The primary mechanism by which nodes operate is through an *activation function*. This function takes the weighted sum of inputs to the node and produces an output based on a predetermined rule.

In figure 3.1 it is showed the *simple perceptron* (SP), or simple neuron, i.e. the first artificial neuron, proposed in the late 1950s by Rosenblatt [48], inspired by the McCulloch and Pitts studies [41]. It works following the equation

$$f(\vec{x}) = h \left( \sum_{i=1}^n w_i x_i + w_0 \right), \quad (3.1)$$

where  $f(\vec{x})$  represents the output of the neural network for input vector  $\vec{x}$ ,  $h(\cdot)$  is the activation function, and  $\sum_{i=1}^n w_i x_i + w_0$  is the weighted sum of the input vector  $\vec{x}$  with weights  $w_i$  and bias  $w_0$ .

SP uses a Heaviside step activation function

$$H(t) = \begin{cases} 0 & \text{if } t < 0, \\ 1 & \text{if } t \geq 0, \end{cases} \quad (3.2)$$

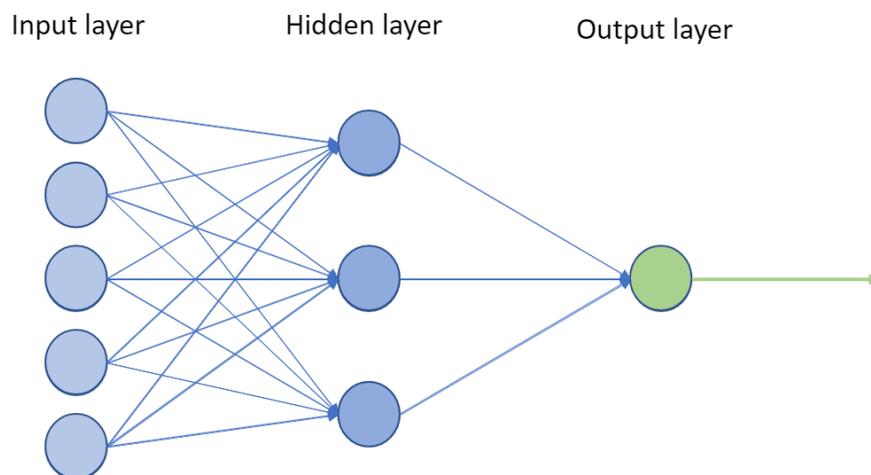


Figure 3.2: Representation of the multiple layer perceptron architecture. Notice the fully connected hidden layer between the inputs and the simple perceptron neuron.

which means the neuron is activated if the output is 1.

The simple neuron marked a significant step in the development of artificial neural networks, but it is capable of solving only linearly separable problems, with limitations in handling more complex, nonlinear patterns and binary classification tasks. This led to the evolution of the *Multi-Layer Perceptron* (MLP), a more sophisticated neural network architecture.

The key innovation of the MLP is the introduction of a hidden layer positioned between the input and output layers, as shown in figure 3.2. The addition of a hidden layer is particularly significant because it introduces the concept of non-linear activation functions. Three common non-linear activation functions employed in the hidden layer are the sigmoid function, hyperbolic tangent (tanh) function, and Rectified Linear Unit (ReLU).

The MLP is an example of *feedforward network*, i.e. a network architecture in which the  $j$ -th layer is fed only by the  $(j - 1)$ -th one. It is easy to understand the extension of the SLP to a multi-layer structure: each output of the first layer become an element of the input of the following.

The power of neural networks with a hidden layer is highlighted by the following

theorem:

**Theorem 1 *Universal Approximation Theorem***

*Let  $f : R^n \rightarrow R$  be a continuous function. There exists a feedforward neural network with a single hidden layer that can approximate  $f$  to arbitrary accuracy on any compact subset of  $R^n$ .*

This theoretical result, proved in 1989 by Cybenko for a sigmoid activation function [10], underscores the universality and versatility of neural networks, suggesting that, given a sufficiently large number of neurons in a single hidden layer, neural networks can learn and represent a wide range of complex relationships within data.

However, because of the computational cost involved in adding a large number of nodes which could be significant, it is often preferable to add hidden layers to the network, resulting in the transition from a *shallow neural network* (consisting of only one hidden layer) to a *deep neural network* (consisting of many hidden layers).

### 3.1.1 Training and Testing

The process of training a machine learning model involves randomly initializing parameters and iteratively adjusting them to minimize a loss function, which measures the difference between predicted and actual values. The simplest optimization algorithm is the *gradient descent method*, where the model “walks down” the loss function surface by taking steps in the direction of the steepest downhill slope. The training process continues until the gradient is flat and further improvement is not possible.

Following training, the model’s real-world performance is assessed through testing on a separate set of data. The evaluation includes computing the loss on this test data to determine how well the model generalizes to new, unseen information. The success of generalization is influenced by the representativeness and completeness of the training data. Underfitting occurs when a model is too simplistic to capture complex patterns, while overfitting happens when a model is excessively complex and memorizes training data, leading to poor performance on new data. Achieving a balance in model complexity is crucial to ensure it captures underlying patterns without falling into underfitting or overfitting [46].

### 3.1.2 Learning approaches

Within the domain of machine learning, different learning approaches have been developed to address different datasets and learning objectives. This section provides an overview of the most important and common learning paradigms: Supervised Learning, Unsupervised Learning and Reinforcement Learning [46]. Then, some paragraphs are dedicated to some interesting approaches which are growing within the deep learning framework, the Semi-Supervised, the Self-Supervised, and the Transfer Learning.

#### Supervised Learning

Supervised learning is a machine learning paradigm in which a model is trained on a labeled dataset consisting of input-output pairs. In this approach, each input in the dataset is associated with a corresponding correct output, providing the model with explicit guidance during training. The main goal for the model is to learn the underlying mapping and relationship between the corresponding inputs and outputs. This process involves the model iteratively adjusting its parameters to minimize the difference between its predictions and the actual output values. Common examples of supervised learning tasks are classification, in which the algorithm predicts the correct category or class for a given input, and regression, in which the goal is to predict a continuous numerical value. In classification tasks, the algorithm learns to assign inputs to predefined classes, while in regression tasks it predicts a numerical value associated with the input data.

#### Unsupervised Learning

Unsupervised learning is a branch of machine learning that deals with unlabeled data, avoiding explicit hints about predefined outcomes. In this paradigm, algorithms aim to autonomously discern patterns, relationships or inherent structures in the data. The absence of labeled data requires the algorithm to explore and identify meaningful representations or groupings on its own. Examples of unsupervised learning tasks are *clustering*, in which the algorithm groups similar data points, *dimensionality reduction*, which involves simplifying data by capturing its essential features, and *generative modeling*, in which the algorithm learns the underlying distribution of the data to generate new

similar instances. Unsupervised learning is particularly useful in scenarios where the intrinsic structure of the data is not well defined or when the discovery of hidden patterns is a primary goal.

## **Reinforcement Learning**

Reinforcement learning is a machine learning paradigm that relies on an agent learning to make decisions through iterative interactions with the environment. Unlike supervised learning, reinforcement learning does not rely on labeled datasets. Instead, the agent receives feedback from the environment in the form of rewards for favorable actions and penalties for unfavorable ones. The goal is for the agent to learn optimal strategies by navigating the environment and adapting its behavior based on the consequences of its actions. An illustrative example is training a computer program to play a game, in which the agent learns to achieve success (e.g. winning the game) by receiving positive reinforcement in the form of rewards and adapting its decisions to avoid negative consequences (e.g. losing), thus learning through a process of trial and error. Reinforcement learning is particularly effective in scenarios where explicit training data may not be available and the agent needs to learn from direct interactions with its environment.

## **Semi-Supervised Learning**

Semi-supervised learning represents a departure from traditional, fully supervised methodologies. While inductive models rely solely on labeled data for training, semi-supervised learning leverages both labeled and unlabeled data. This approach, often referred to as “transductive modeling”, enables the model to exploit patterns present in the unlabeled data, augmenting its ability to make informed decisions.

In semi-supervised learning, the model learns from the labeled data to establish the relationship between inputs and outputs. However, it also exploits the vast pool of unlabeled data to further refine its understanding and enhance performance. This is particularly advantageous when labeled data is limited or expensive to obtain.

Despite its benefits, semi-supervised learning comes with challenges. Unlike inductive models, semi-supervised models may need retraining when additional unlabeled data becomes available. This is because the model must continually adapt to evolving patterns

within the unlabeled data to maintain optimal performance.

## **Transfer Learning**

Transfer learning is a machine learning paradigm that leverages knowledge gained from solving one task to improve performance on a different but related task. In traditional machine learning, models are trained for a specific task with a fixed dataset, and the knowledge acquired during this process is typically not applicable to other tasks. Transfer learning, on the other hand, allows a model to generalize its learning across different domains, enabling it to adapt and perform well on new, unseen tasks.

The key idea behind transfer learning is to use a pre-trained model, often trained on a large dataset for a particular task, as a starting point for a new task. The knowledge acquired by the model in its initial training is transferred or fine-tuned to the new task with a smaller, task-specific dataset. This approach is particularly beneficial when labeled data for the new task is limited, as the model can capitalize on the extensive knowledge gained during its pre-training.

Transfer learning has proven effective in various domains, such as computer vision and natural language processing, where pre-trained models like convolutional neural networks (CNNs) for image recognition or language models for text understanding can be adapted for specific applications with minimal additional training. This technique accelerates model training, enhances performance, and promotes efficient use of resources in machine learning applications [30].

## **Self-Supervised Learning**

Self-Supervised learning is a machine learning paradigm in which a model learns from data without the need for explicit external labeling. In this approach, the training task is generated directly from the data, creating a pseudo-supervised learning scenario without relying on human annotations. The model is designed to perform its prediction generating its own supervisory signal. An illustrative example of self-supervised learning is the pre-training of a language model. In this context, the model might be instructed to predict missing words in a sentence, where portions of the text are deliberately removed, and the model learns to predict the omitted words based on the surrounding

context. Similarly, in the case of images, the model could be trained to predict the next sequence in an image, improving its ability to understand spatial and temporal relationships within visual data. *Contrastive learning* [62] is a self-supervised method that improves learning of representations by maximizing the agreement between similar instances and minimizing it for dissimilar ones. This is achieved through a contrastive loss, in which positive pairs (similar instances) are brought closer together and negative pairs (dissimilar instances) are moved farther apart. Self-supervised learning is valuable in scenarios where obtaining labeled data is difficult or expensive, allowing the model to derive meaningful representations independently from the data.

## 3.2 Convolutional Neural Networks

Convolutional neural networks (CNNs) have revolutionized the field of computer vision, offering a powerful tool for extracting complex features from complex visual data.

Unlike traditional neural networks, CNNs excel at processing grid-like data structures, such as images, through the application of three key mechanisms: *local receptive fields*, *weight sharing*, and *subsampling* [6].

Convolutional layers in neural networks consist of multiple planes known as feature or activation maps. Each map is computed by combining parameters within a local region of the input image, allowing it to detect specific patterns at different locations. This concept is related to local receptive fields. Weight sharing is implemented under the assumption that patterns appear in various image regions, using the same weights for the entire activation map. This approach mitigates the risk of overfitting by avoiding an exponential increase in the number of weights. Subsampling gradually reduces spatial dimensions without adding more weights, preserving informative features. Stacking multiple layers achieves a substantial reduction in spatial dimensions while progressively increasing feature planes, with convolutional layers handling local receptive fields and pooling layers managing subsampling (Figure 3.3).

These principles enable CNNs to automatically learn hierarchical representations of visual information, making them exceptionally effective in tasks such as image classification and object detection, where robust feature extraction and position invariance are

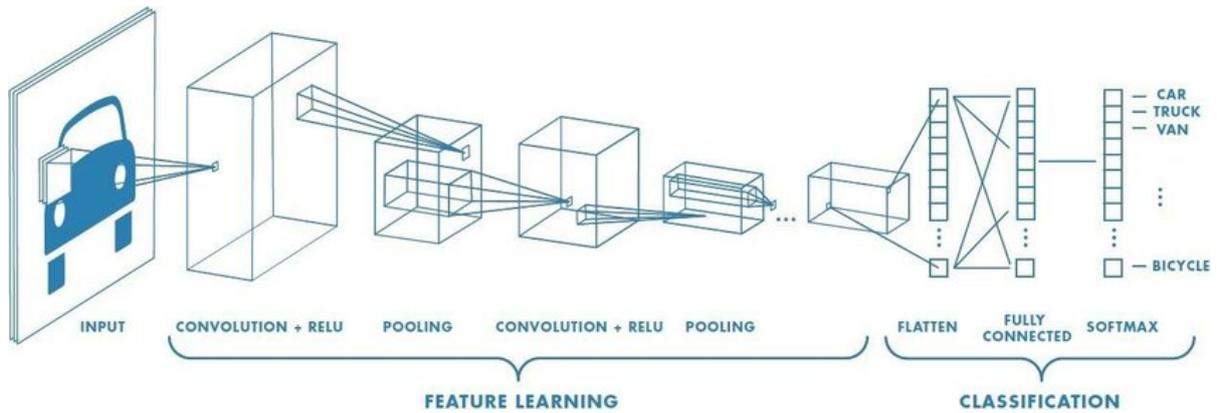


Figure 3.3: Structure of a convolutional neural network for classification.

requested [57].

### Convolutional Layers

The convolutional layer performs feature extraction through a combination of linear and nonlinear operations. This involves applying kernels to input tensors, creating feature maps that capture different features. The use of filters in the initial layers allows the network to detect fundamental features such as edges and colors, combining them progressively in the deeper layers to achieve more abstract representations. In particular, weight sharing ensures efficient detection of feature patterns at various positions in the image, promoting translation invariance and spatial hierarchy learning.

### Pooling Layers

Pooling layers downsample feature maps to introduce translation invariance and reduce the number of learnable parameters. Max pooling, the most popular type, extracts patches and outputs the maximum value, effectively downsampling in-plane dimensions. Global average pooling, an extreme downsampling method, computes the average of all elements in a feature map, reducing it to a 1x1 array while retaining depth. This operation is applied before fully connected layers, reducing parameters and enabling variable input sizes.

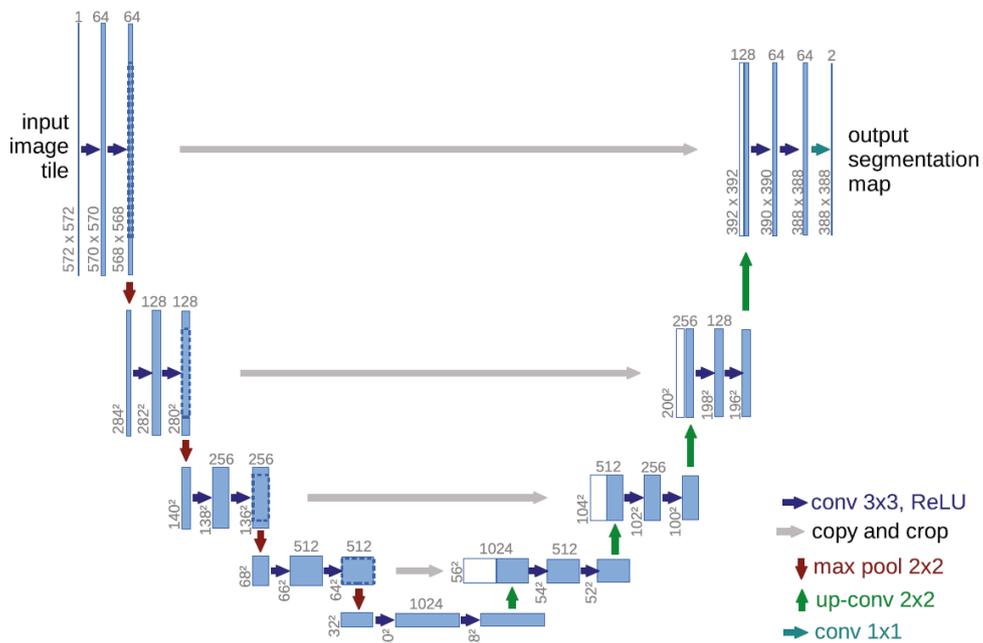


Figure 3.4: Architecture of the original U-Net. [47]

## Fully Connected Layers

Output feature maps from the last convolutional or pooling layer are flattened into a one-dimensional array and connected to fully connected (dense) layers. Each input is linked to every output through learnable weights. These layers map the features extracted by previous convolution and pooling layers to the network's final outputs, such as class probabilities in classification tasks.

### 3.2.1 U-Net

The U-Net architecture, introduced by Ronneberger et al. in 2015 [47], is a widely used network architecture, specially in medical image analysis. The original idea was taken as a starting point for the development of numerous variants, making it one of the most widely used methods for tasks such as image classification and segmentation [52]. This model excels in pixel-wise classification, making it particularly valuable for the task of image segmentation.

The U-Net architecture is a convolutional neural network (CNN) design, where the objective is to partition an image into distinct regions. Its name is derived from the U-shaped structure that characterizes its layout, featuring a contracting path (encoder), a central bottleneck, and an expansive path (decoder).

The contracting path is situated on the left side of the U-Net (Figure 3.4). Each blue box in this region represents a multi-channel feature map, with the number of channels indicated on top of the box. The x-y size of the feature map is specified at the lower left edge. Operations in this section typically involve successive convolutional layers, often followed by non-linear activation functions such as ReLU. Downscaling is achieved through processes like max pooling, which reduces spatial dimensions.

The central bottleneck acts as a bridge between the contracting and expansive paths. It comprises several convolutional layers aimed at capturing abstract features, crucial for accurate segmentation.

The right side of the U-Net constitutes the expansive path (Figure 3.4). Here, feature maps are upsampled to recover spatial dimensions lost during the contracting path. White boxes represent copied feature maps from the contracting path, forming skip connections. These connections enable the combination of high-level semantic information from the contracting path with detailed spatial information from the expansive path. The copied feature maps from the contracting path are concatenated with the upsampled feature maps in the expansive one.

Arrows in the diagram indicate the flow of information throughout the network. In the contracting path, information is progressively abstracted, and spatial dimensions are reduced. The bottleneck captures high-level features, while the expansive path restores information gradually, increasing spatial dimensions through up-convolutions. For classification tasks, a deeper output tensor is needed, with a depth corresponding to the number of classes to be identified.

U-Net's architecture is particularly effective for tasks like medical image segmentation. The combination of the contracting and expansive paths, along with skip connections, allows the network to exploit both high-level context and fine-grained details for accurate segmentation.

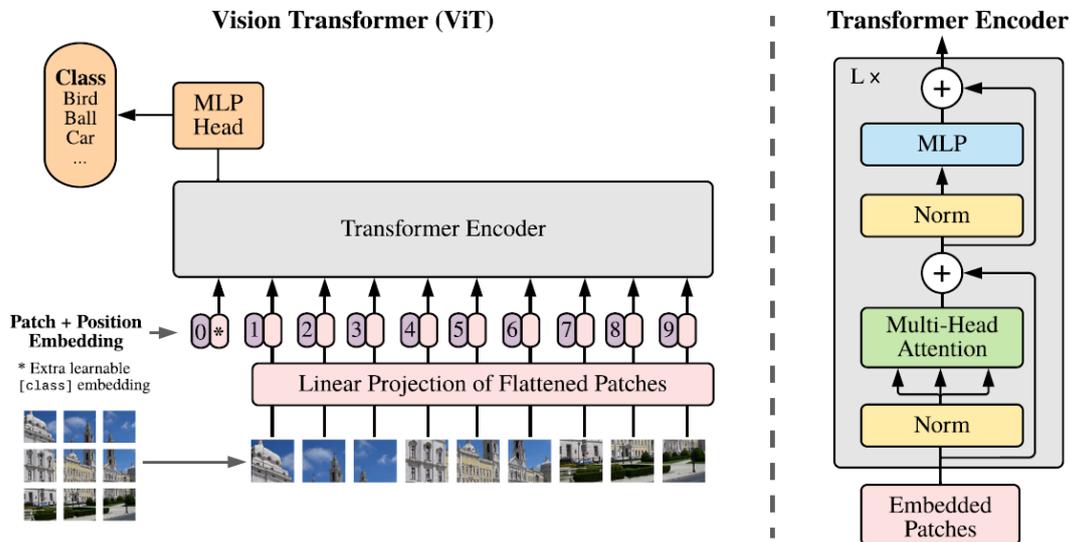


Figure 3.5: Architecture of a Vision Transformer (left), with a focus on the Transformer Encoder internal structure (right). [14]

### 3.3 Vision Transformers

Convolutional neural networks have dominated computer vision for the past decade, but their inherent bias toward local spatial structures limits their ability to model long-range dependencies [64]. Despite efforts to improve CNNs, these limitations persist, and researchers have explored alternatives. *Transformers*, introduced in 2017 by Vaswani et al. [55], which have proven effective in NLP, inspired a paradigm shift in computer vision through **Vision Transformers** (ViTs) [14].

The key elements they introduced are the *pre-training* strategy and the *self-attention mechanism*.

Transformers undergo two-stage training: initial pre-training on extensive non-annotated datasets using self-supervised learning and later fine-tuning on specific tasks. Contrastive learning ensures invariance to semantic changes. This approach enhances scalability and adaptability, enabling meaningful representation generation without costly manual annotations.

In figure 3.5 it is shown the architecture of a ViT. The main components are:

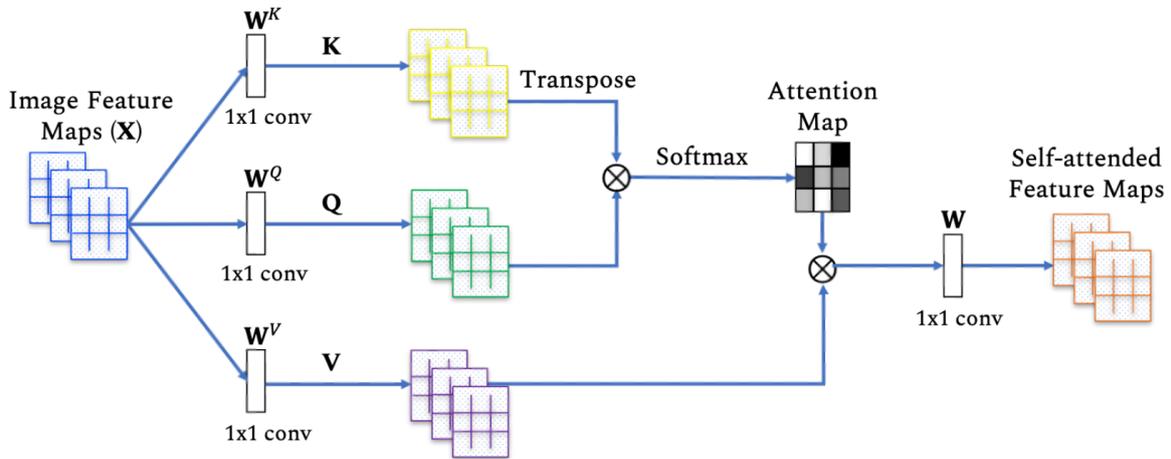


Figure 3.6: Example of a single self-attention head used in computer vision. [29]

- **Image Patching:** ViTs divide input images into non-overlapping patches, which are then reshaped into 1D vectors;
- **Embedding:** Reshaped patch vectors undergo linear embedding into a desired dimensional space using a learnable embedding matrix;
- **Positional Encoding:** Positional embeddings are added to incorporate spatial information, allowing the model to discern different patch positions within the image;
- **Transformer Layers:** Multiple Transformer layers process the combined patch and positional embeddings, exploiting self-attention mechanisms and feed-forward neural networks.

### Self-Attention in Transformers

Self-attention is a fundamental mechanism in the Transformer model, designed for structured prediction tasks by estimating the relevance of items within a sequence. It explicitly models interactions between all entities in a sequence, capturing global contextual information for each entity [29]. The main components of the attention mechanism are:

- **Query (Q):** it is a vector that represents the element of interest or the context being examined;

- **Keys (K):** they are vectors associated with the elements in the input sequence. They provide information about the context according to which the query will be evaluated;
- **Values (V):** they are vectors representing the content or information associated with each key. These values are combined on the basis of the attention weights to form the final output.
- **Compatibility Function:** it assesses the relationship between the query and each key. It produces a set of scores meaning how relevant each key is to the query.

To understand how self-attention works, let us consider a sequence of entities  $x_1, x_2, \dots, x_n$  represented by  $X \in R^{n \times d}$ , where  $d$  is the embedding dimension.

The self-attention layer utilizes learnable weight matrices  $W^Q$ ,  $W^K$ , and  $W^V$  to transform the input sequence into query ( $Q = XW^Q$ ), key ( $K = XW^K$ ), and value ( $V = XW^V$ ) representations. Then, for each entity in the sequence, self-attention computes the dot-product of the query with all the keys, normalized using the softmax operator. This results in each entity becoming a weighted sum of all entities in the sequence, where weights are determined by attention scores, given by the compatibility function. The output  $Z$  is then computed using the equation:

$$Z = \text{SoftMax} \left( \frac{QK^T}{\sqrt{d}} \right) V \quad (3.3)$$

To capture complex relationships among elements, **multi-head attention** is employed with multiple self-attention blocks (heads). Each block has its set of learnable weight matrices. The outputs of these attention blocks are concatenated and projected onto a weight matrix to obtain the final output [55].

Comparing self-attention with convolution, the first dynamically calculates filters instead of using static filters, resulting in higher flexibility to model both global and local features [64].

However, the quadratic complexity of self-attention hinders its applicability to longer sequences (high-resolution images). Some methods have been proposed to address this

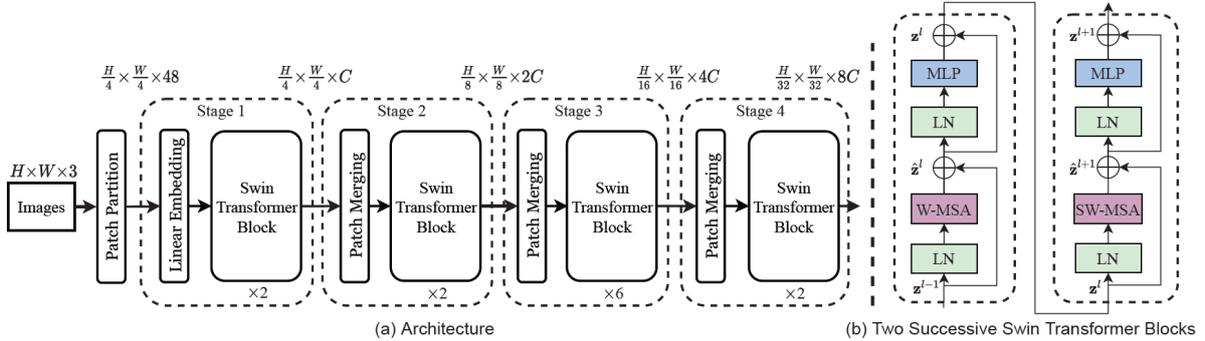


Figure 3.7: (a) The architecture of a Swin Transformer. (b) Two successive Swin Transformer Blocks. Notice the difference between W-MSA and SW-MSA multi-head self attention modules with regular and shifted windowing configurations, respectively. [32]

challenge. One of the most promising approaches in image segmentation and classification is represented by *multi-scale* family of ViT, which includes the Swin Transformer, described in the next paragraph.

### 3.3.1 Swin Transformer

Swin Transformer [32] has a multi-stage hierarchical architecture which computes attention within a local window, by partitioning the window into multiple subpatches.

To capture interactions between different windows (image locations), window partitioning is gradually shifted, along the hierarchy of the network, to capture overlapping regions.

At the heart of the Swin Transformer lies a unique architecture (Figure 3.7(a)), characterized by a patch-based tokenization, a hierarchical feature representation, and a shifted window-based self-attention mechanism. The initial stage involves splitting the input image into non-overlapping patches, treating each patch as a token, and applying a linear embedding layer to project raw pixel RGB values to a higher-dimensional space. The subsequent Swin Transformer Block maintains the number of tokens while introducing modified self-attention computations. Then, the hierarchical representation is achieved through patch merging layers that concatenate features of neighboring patches, progressively reducing the number of tokens. Each stage culminates in a Swin Trans-

former Block, transforming features while preserving resolution.

### **Swin Transformer Block**

The Swin Transformer Block (Figure 3.7(b)), redefines the traditional multi-head self-attention (MSA) module. It incorporates a window-based MSA module (W-MSA), followed by a two-layer Multi-Layer Perceptron (MLP) with GELU nonlinearity in between. Layer normalization (LN) is placed before each MSA module and MLP, and residual connections enhance the overall stability of the block.

The global self-attention mechanism in the standard Transformer architecture leads to quadratic complexity with respect to the number of tokens. To make self-attention more scalable for vision tasks, the proposed W-MSA with a fixed window size, results in linear complexity.

The introduction of shifted window partitioning in successive blocks is a crucial point to enhance modeling power. This shift creates overlapping regions between windows in consecutive blocks, promoting information flow and connecting neighboring non-overlapping windows of the previous layer.

## **3.4 Swin-UNet**

CNN-based methods, particularly U-Net variants, have demonstrated notable success in various applications. However, these approaches encounter challenges in meeting stringent accuracy requirements, primarily due to the intrinsic locality of convolution operations, which hinders the capture of explicit global and long-range information.

To address this limitation, Transformer-based methods leverage the self-attention mechanism, proving to be powerful tools. Nonetheless, their quadratic complexity poses implementation challenges when dealing with high-resolution image datasets.

The Swin Transformer, distinguished by its innovative architecture based on shifted windows, mitigates this complexity issue. Consequently, it has gained increasing traction in the vision domain, particularly in tasks like classification.

In the context of this work, the chosen methodology for the segmentation step involves employing Swin Transformer Blocks as the backbones of a U-Net architecture.

This strategic combination aims to harness the strengths of transformers in capturing global information while leveraging U-Net’s pixel-wise classification properties for effective segmentation.

This approach has been previously implemented in various works, such as the one by Cao et al. [8], yielding promising results. Further details regarding the implementation specifics for this work can be found in Chapter 4.

# Chapter 4

## Materials and Methods

The project aims to develop an image processing pipeline for the segmentation of Bone Marrow Fibrosis (BMF) components in histological images stained with GOMORI trichrome technique [20]. The primary goal is to generate a binary mask highlighting BMF components, enabling a quantitative measurement of their area relative to the total histological section. The investigation delves into the presence and distribution of fibrotic components. The project's quantitative approach seeks to develop a method for assessing the extent of BMF, offering valuable insights into the degree of fibrosis associated with different diseases.

### 4.1 Dataset

For this work, a dataset comprising 1123 bone marrow biopsy samples was curated by Humanitas Research Hospital in Milan, from June 2016 to December 2020. The digitization of the samples were performed using the Hamamatsu whole slide scanner and stored in the NDPI format in 2021. The NDPI file is a specialized format for whole slide images, characterized by a proprietary TIFF-like structure that encloses JPEG images along with metadata detailing their size and resolution. To facilitate the opening and processing of these images, the dataset was managed using OpenSlide software [22].

All slides were prepared using the Gomori trichrome stain, a histological technique mainly used for highlighting muscle and collagen fibers in tissue sections, particularly

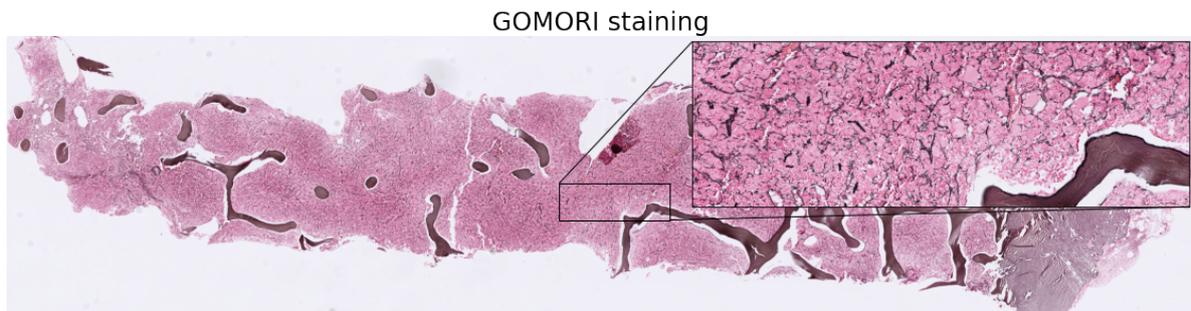


Figure 4.1: Whole Slide Image featuring a GOMORI trichrome staining.

in bone marrow biopsies to assess fibrosis. Figure 4.1 shows an example of a WSI stained with the GOMORI technique, characterized by two main colors: a pink-like color identifies the section soft tissue, while the brown-like color highlights the dense components. We can find the BMF components inside the pink-like areas, as a series of darker lines.

The patient cohort comprises 981 individuals, 548 males and 433 females. The age distribution within the cohort spans from 20 to 96 years. It is worth noting that some patients underwent multiple time point analyses, involving images related to subsequent biopsies. However, it is important to highlight that these additional data were not utilized for the specific type of study conducted.

This dataset encompasses a diverse spectrum of hematologic conditions, incorporating cases of Acute Myeloid Leukemia (AML), Myelodysplastic Syndrome (MDS), Myeloproliferative Neoplasms (MPN) and the combined Myelodysplastic Syndrome/Myeloproliferative Neoplasms (MDS/MPN). Additionally, a control group (CTR) has been included for comparative analysis.

#### 4.1.1 Digital Pathology

Bone marrow fibrosis is characterized by the increased deposition of reticulin fibers and in some cases collagen fibers within bone marrow. BMF is a key factor in diagnosing myelofibrosis, a type of MPN. In this condition, the bone marrow's abnormal deposition of reticulin and collagen fibrosis is thought to be caused by specific cells (myelofibrosis hematopoietic stem/progenitor cells). This process creates an unfavorable environment

that supports malignant over normal blood cell development [59].

A spectrum of benign conditions and malignant disorders is linked to a pathological increase in bone marrow stromal fibers. This category encompasses MDS, MPN, MDS/MPN and AML, each demonstrating an association with an abnormal rise in bone marrow fibrosis [11].

MDS, a heterogeneous group of disorders characterized by dysplasia and ineffective hematopoiesis, have only recently been associated with BMF. Numerous case reports and retrospective studies have contributed to the understanding of MDS with bone marrow fibrosis (MDS-F). It has been determined that MDS-F accounts for a notable percentage, ranging from 10% to 20%, of all MDS cases [27]. This recognition of MDS-F underscores the complexity of MDS as a disease entity and highlights the variability in its presentation. Understanding the incidence and implications of BMF in MDS is crucial for accurate diagnosis, appropriate treatment strategies, and prognostic considerations in this heterogeneous group of hematological disorders [12].

Acute myeloid leukemia is a complex hematologic malignancy marked by rapid myeloid blast proliferation in the bone marrow. AML is typically not associated with bone marrow fibrosis (BMF), even though instances of fibrosis can be observed. Despite the absence of specific international guidelines for AML with BMF, it emerges as an independent prognostic factor for overall survival (OS), underscoring its crucial role in predicting outcomes and treatment efficacy for newly diagnosed patients. The presence of BMF varies in AML across different age groups and risk categories, contributing to its heterogeneous nature [61].

MPNs represent a cluster of disorders marked by the excessive production of mature blood cells within the bone marrow, and closely associated with BMF. Understanding the variable presence of fibrosis in MPNs, especially in advanced forms like primary myelofibrosis, is crucial for accurate diagnosis and tailoring effective treatment strategies [59].

Myelodysplastic syndrome/myeloproliferative neoplasm overlap syndromes (MDS/MPN) are disorders that include features of both MDS and MPN. For this reason their diagnosis is proved challenging.

### 4.1.2 Color Decomposition Preprocessing

In histology, different stains are used to highlight specific components in tissues. The RGB color space, commonly used in digital images, may not accurately represent the original chemical colors introduced during staining. **Color decomposition** involves transforming the RGB color space into staining one, i.e. converting RGB channels into channels related to different components in histological section. Color decomposition supervised methods use a pre-determined staining matrix to convert images to the staining channels, requiring precise knowledge of the desired staining values. On the other hand unsupervised methods, like the Macenko algorithm [35], do not need prior knowledge and determine optimal color decomposition based on the input image.

In the dataset, we encountered some images with color or staining issues. To identify these problematic images, we curated a subset of 200 well-stained images. Utilizing the unsupervised Macenko’s color decomposition technique on low-resolution images, we computed the mean conversion matrix, from RGB to GOMORI stainings channels as

$$\begin{bmatrix} -2.9365 & 1.04966524 & 1.87777335 \\ 3.37487221 & -0.34343737 & -1.29851867 \\ -0.18034784 & 0.71461108 & -0.73782514 \end{bmatrix}$$

Subsequently, we applied this matrix to perform a supervised color decomposition on the entire dataset, using the following procedure.

Firstly, the image is transformed into optical density space. Optical density, a measure of light absorption, is utilized as an intermediate representation. This step is essential for standardizing the image’s representation, providing a consistent basis for subsequent transformations. The heart of the transformation lies in the color conversion matrix. This 3x3 matrix contains vectors corresponding to a transformation in stainings color space that projects the image onto the desired target color space. Following the conversion of the RGB colors into the GOMORI’s channels, we specifically selected the channel that highlights the soft tissue. Employing an Otsu threshold, we then generated a binary mask with the tissue as the foreground and all other components of the slide as the background.

This process allowed us to observe a distinct difference in the ratio of foreground to background between good-quality images and degraded ones. In Figure 4.2 are shown

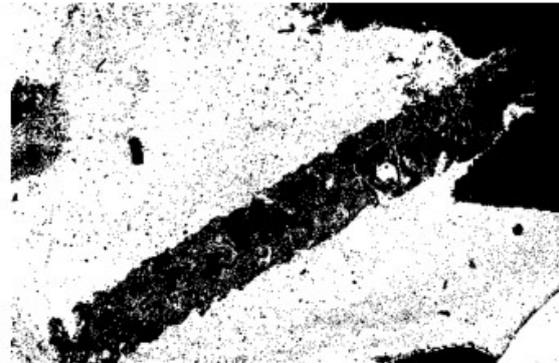
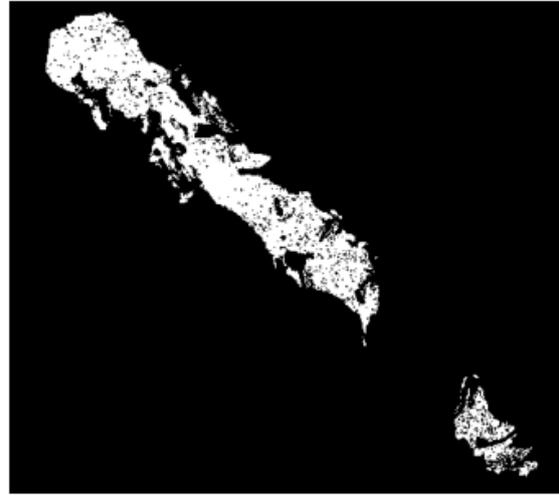
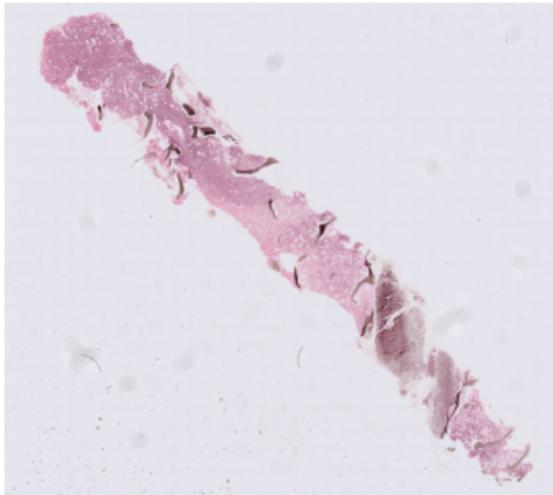


Figure 4.2: Results of the color decomposition preprocessing procedure on a good-quality WSI (top), and a poor-quality WSI (bottom). Notice that the binary mask is inverted in degraded slides.

Condition	AML	MDS	MDS/MPN	MPN	CTR	Total
N° patients	80	193	36	146	147	602
Males/Females	41/39	134/59	28/8	74/72	82/65	338/264
Age	21-88	29-92	56-89	21-89	20-90	20-92

Table 4.1: Dataset

the results of the procedure, in which can be noticed how poor-quality stained images lead to an inverted mask.

After this procedure we decided to flag inverted mask images by classifying them as degraded. Thus, the resulting dataset used for this study includes 602 patients (338 males and 264 females), whose age ranges from 20 to 92 years. The pathological conditions of the patients are resumed in Table 4.1.

## 4.2 Segmentation

The segmentation process involves both identifying the tissue within the slide and segmenting the bone marrow fibrosis.

### 4.2.1 Tissue Segmentation

For tissue identification, the same method was used as for the selection of images to be excluded. This involved implementing an Otsu threshold after a supervised color decomposition. The result is the isolation of soft tissue, excluding bone tissue and cells clear to staining (e.g. adipocytes).

The exclusion of adipocytes and bone tissue is of paramount importance since fibrosis is not observable in regions occupied by these elements in the images. As a consequence, this method of tissue identification serves to safeguard the accuracy of the fibrotic area to tissue ratio, preventing any inadvertent bias introduced by factors that should not impact the calculation.

## 4.2.2 Bone Marrow Fibrosis Segmentation

Regarding the segmentation of BMF, it was necessary a more powerful method. Fibrosis can be distinguished from tissue only at high resolution. Given the size of WSI making computationally very difficult to process them, deep learning models usually receive patches as input, i.e. portions of images of arbitrary size extracted from the original image [13]. Our choice for the patches size was  $W \times H = 256 \times 256$ . To avoid processing unnecessary data, the patches were extracted from the regions of the slides that correspond to the ROIs identified in the segmentation of the tissue. However, at such high resolutions, the images exhibit intricate patterns and much more variability, posing challenges in achieving effective results for the BMF segmentation.

The choice therefore fell on an **active semi-supervised learning** (ASSL) pipeline [9]. This training strategy address the problem of annotation availability, while minimizing the effort for clinicians. Starting with an unsupervised method as a **kickstarter** to achieve the first labels, we then trained a deep learning segmentation model on them.

The main phase of the ASSL approach consists of several rounds of validation. During each iteration of the active learning loop, model-generated segmentation predictions of unlabeled patches were presented to domain experts, specialized in bone marrow pathology. For each validation image, clinicians determined if the generated segmentation was accurate. The validation images (and corresponding segmentation masks) which satisfied the experts evaluation, were inserted into the training set as ground truth, for a next round of model training. A schematic representation of the proposed training strategy is shown in Figure 4.3.

With this type of training, one usually chooses as a stop criterion the achievement of a certain metric such as a certain percentage of valid images compared to those predicted within a round. In this case, a stop criterion given by a trade-off between time and accuracy was chosen by mutual agreement with the pathologists, having noted that the images produced were sufficiently valid after two validation rounds.

The ASSL strategy facilitated the efficient utilization of unlabeled data by progressively incorporating ground truth information through expert analysis, compensating for the absence of initial annotated data. This iterative approach facilitated the model's adaptation to the complexities of fibrosis segmentation within bone marrow biopsies, re-

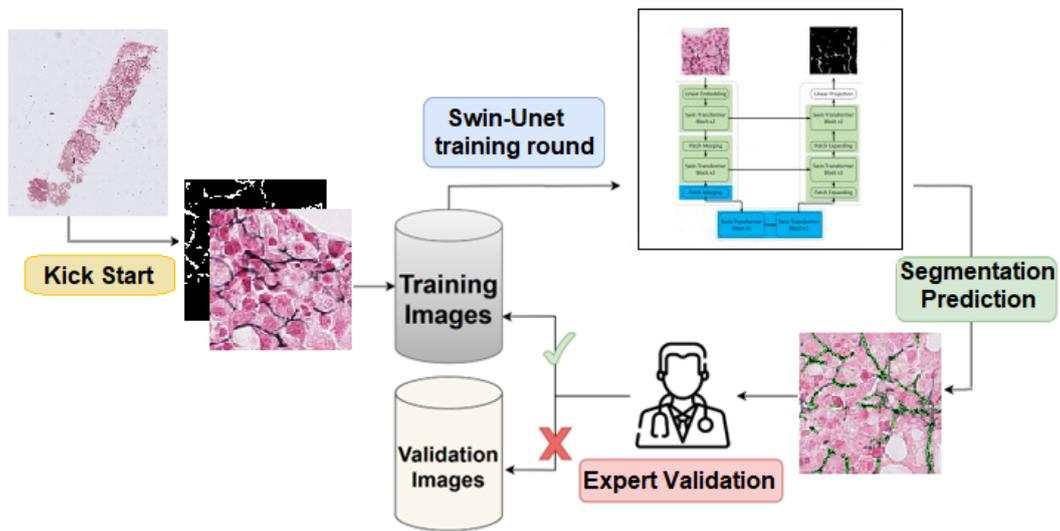


Figure 4.3: Active semi-supervised learning pipeline: after an unsupervised kick-start to produce the first labeled patches, the training continues through several rounds of segmentation prediction and expert validation.

sulting in a final model capable of accurately identifying and delineating fibrotic regions.

### **Color Deconvolution Kick Start**

As mentioned in the previous paragraph, giving the lack of annotations within the dataset, the ASSL pipeline needed an unsupervised method to start with the active learning strategy. Color deconvolution method, specifically Macenko’s approach, was chosen for the task.

The process begins by selecting a subset of patches from several images. There are two important aspects in this step: focus is on including only those patches that contains soft tissue to avoid useless computation, and taking patches from many WSIs for generalizability, allowing it to learn from a broader spectrum of cases. Macenko color decomposition is then applied to the selected subset of patches and a thresholding generates binary masks.

The combination of Macenko’s algorithm and thresholding leads to the creation of the first training set. This set includes both the original patches and corresponding training masks. Training masks are annotated images that indicate the BMF, helping the machine learning model learn and generalize patterns during the training process.

At high resolutions where fibrosis became visible, the inherent variability hindered effective segmentation. Moreover, the extraction of images in patches introduced certain inconsistencies, with some images not meeting optimal standards. Despite this, they hold value as an initial source of information regarding fibrosis. While not ideal for precise segmentation, these images offer a foundational understanding of fibrosis, serving as a preliminary insight.

### **Model**

The chosen segmentation model is based on a U-Net architecture [47], with a transformer-based model, the Swin-Transformer [32], as backbone. As described in Chapter 3, Swin-Unet comprises an encoder, bottleneck, decoder and skip connections. In Figure 4.4 is shown a scheme of the architecture. As already said the input images are the patches extracted from the whole slide images according to the ROI identified through the segmentation of the tissue.

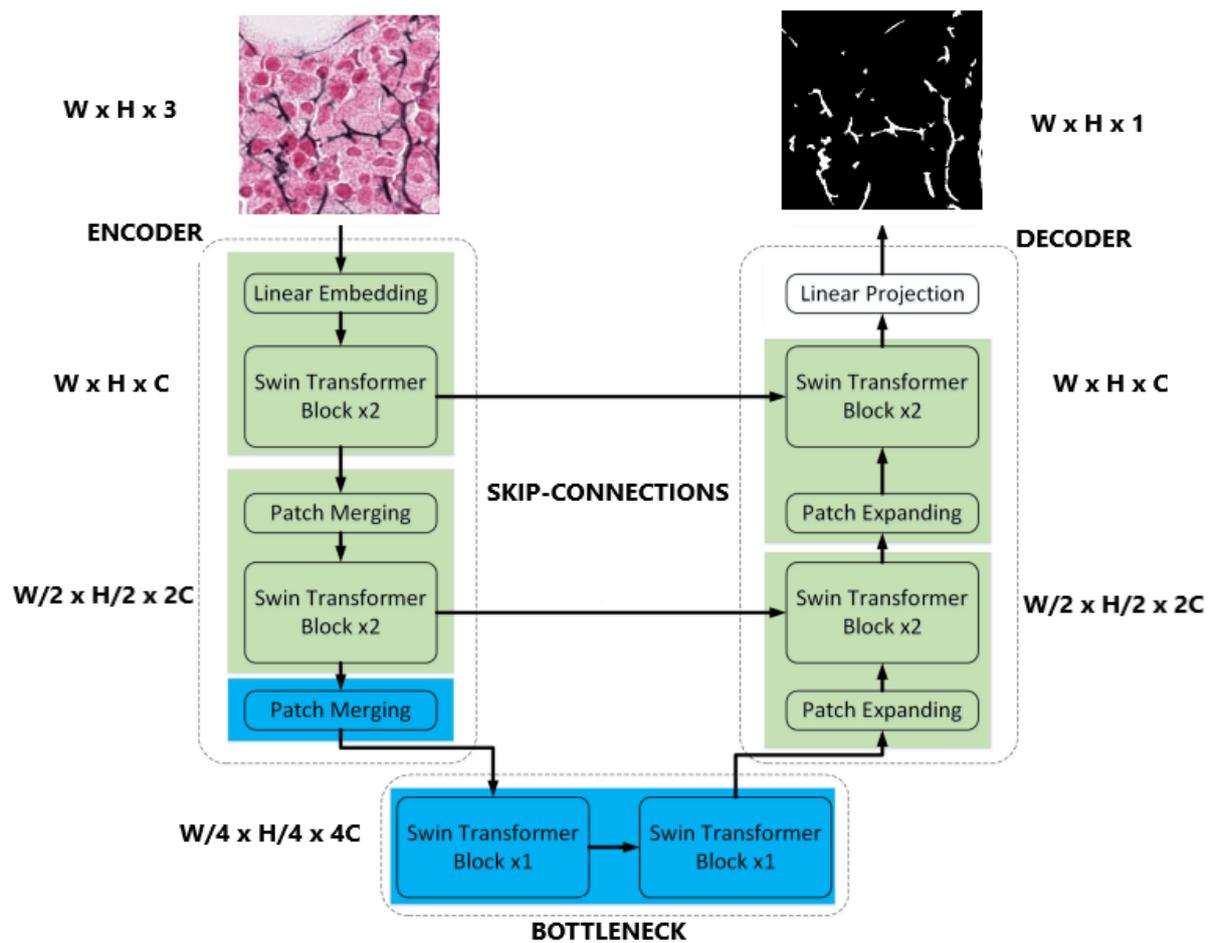


Figure 4.4: Architecture of the segmentation model employed in the ASSL pipeline.

Initially, a linear embedding layer is employed to project the feature dimension into an arbitrary dimension  $C$ , in this case set to 16. The resultant patch tokens go through layers of Swin Transformer Blocks and a patch merging, to generate hierarchical feature representations. Specifically, the patch merging layer is tasked with downsampling and increasing dimensionality, while the Swin Transformer block focuses on learning feature representations.

Taking inspiration from U-Net, we implemented a symmetric transformer-based decoder. This decoder features Swin Transformer Blocks and a patch expanding layers. Unlike the patch merging layer, a patch expanding layer is designed for up-sampling, reshaping adjacent feature maps into larger ones. Context features extracted are then fused with multiscale features from the encoder via skip connections, compensating for the loss of spatial information induced by downsampling. Finally, a linear projection layer is applied to these up-sampled features to produce pixel-level segmentation predictions.

The unique characteristics of the Swin Transformer model is the shifted windows mechanism implemented in the multi-head self attention blocks, which allow a significant reduction in terms of complexity. In this work the fundamental parameter of the window size was set to 8.

The implemented Swin-Unet segmentation model, pre-trained with ImageNet-1k, is built using the TensorFlow library [1]. The training setup involves a maximum of 100 epochs, utilizing the Adam optimizer known for its efficiency. A cosine decay learning rate schedule is employed, starting from  $1e-4$  and gradually increasing to  $1e-3$  over warm-up steps set at 10 times the training dataset’s cardinality.

During training, the Binary Intersection over Union (IoU) is monitored as the chosen metric for evaluating segmentation accuracy. A threshold of 0.5 is applied to the IoU calculation to determine correct pixel classification. The Binary Focal Crossentropy loss function is used to address class imbalance, with a modulating factor (gamma set to 1) focusing on hard-to-classify samples. Additionally, label smoothing with a factor of 0.1 is applied to prevent overconfidence in the model and encourage the learning of more robust features.

### 4.3 Analysis

The core of our analysis is based on the ratio of BMF to soft tissue, which serves as a quantitative metric to assess the distribution of fibrotic regions within the bone marrow [53]. These values were derived by quantifying the number of pixels constituting the foreground within the binary masks obtained through segmentation.

As a continuation of our investigation, the next chapter delves into the results derived from our analysis of the BMF segmentation and the hematopoietic tissue. Starting by observing the distributions of the data, particularly their stratification among distinct disease classes, we undertook a comparative analysis. Our approach initially involved examining data within two primary populations: controls and patients with onco-hematological diseases. Next, we extended our examination to perform a detailed analysis of the different hematologic diagnoses with each other.

# Chapter 5

## Results

### 5.1 Preprocessing

The preprocessing part involves data selection based on the quality of the staining. Since the dataset was digitized a few years after biopsy acquisition, a number of slides were not usable. As described in Chapter 4, the quality assessment was performed using a color decomposition technique featuring both the unsupervised Macenko algorithm and a supervised method. Poor-quality slides were characterized by an inverted tissue binary mask. In Figure 5.1 is shown the histogram of the ratio of the foreground area to the whole mask area, where two well separated groups can be easily recognized, with a threshold set to 0.5. A small set of images were found very close to the threshold (gray region in Figure 5.1), then they were evaluated manually.

### 5.2 Segmentation

The segmentation involved the identification of the soft tissue and the bone marrow fibrosis within the WSI of the final dataset resulting from the preprocessing step. Different techniques were used for the two tasks. The identification of the tissue was obtained with a color decomposition as we did in the preprocessing step. The segmentation of the bone marrow fibrosis (BMF) leverages an Active Semi-Supervised Learning (ASSL) strategy during the model training phase to accurately delineate fibrotic regions. In order to

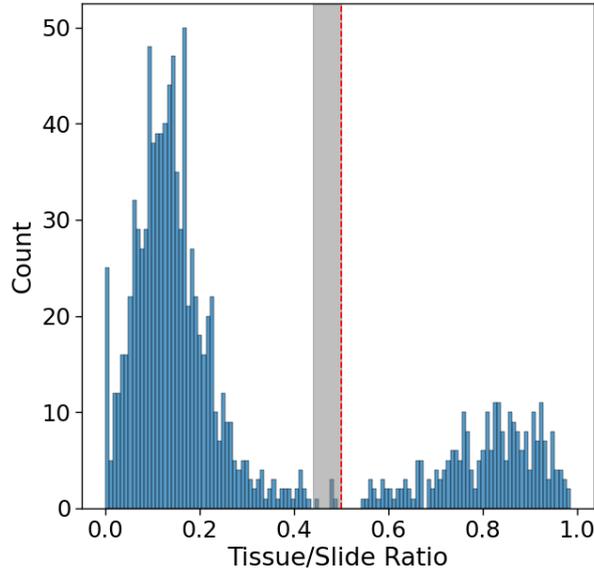


Figure 5.1: Histogram of the Tissue to Slide ratio used as a quality assessment criteria to flag poor-quality images presenting an inverted tissue binary mask, and resulting in Tissue/Slide Ratio larger than 0.5

face the lack of annotated patches, the color decomposition method have been used as a kickstarter to generate a set of 3427 annotated patches.

The predicted patches produced by the model trained with the kick-start, were subjected to experts analysis. Patches validated and accepted by the pathologists were included in the training set as ground truth for the successive ASSL training rounds. After two rounds of learning and validation, pathologists evaluated as ground truth 2611 patches, used as the the training set for the final segmentation model. In Figure 5.2 are shown the training loss and the IoU of the last round of training. Predictions after this round were considered sufficiently good by pathologists, then we decided, with a trade-off between time and accuracy, to stop the training procedure.

From the entire dataset were extracted more than 20 millions of patches, predicted with the model trained on 2611. Some issues have been found in smallest fibrotic components resulting as noisy meaningless label. However, they were solved in post-processing. Examples of the achieved BMF segmentation are shown in Figure 5.3 and 5.4.

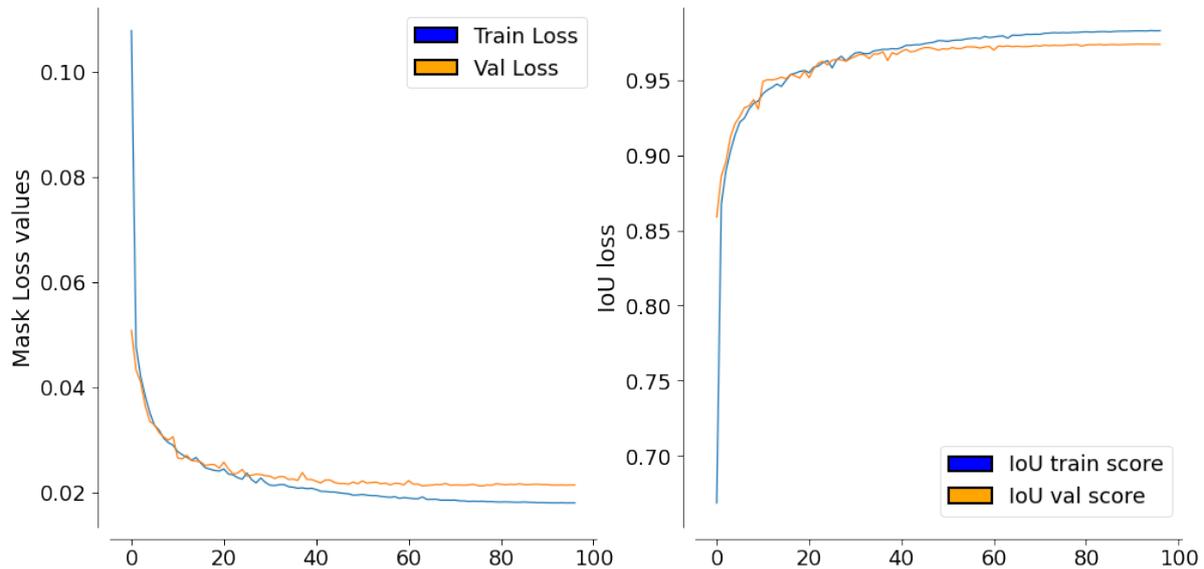


Figure 5.2: Loss and metrics related to the validation (orange) and training (blue) set as a function of the epochs of the last training round.

## 5.3 Analysis

In this section, we present a detailed analysis of the segmentation results, focusing on the ratio between Bone Marrow Fibrosis and soft tissue. This ratio serves as a quantitative measure to assess the distribution of fibrotic regions within the digital slides [53].

These values were derived by quantifying the number of pixels constituting the foreground within the binary masks obtained through segmentation of the hematopoietic tissue and the bone marrow fibrosis:

$$Ratio = \frac{Bone\ Marrow\ Fibrosis\ Area}{Hematopoietic\ Tissue\ Area}$$

### 5.3.1 Overview

To provide a comprehensive overview, we begin with the presentation of histograms depicting the distribution of the *Ratio* across the entire dataset, and related to each diagnosis. Complementing the histograms, Kernel Density Estimate (KDE) plots offer a smoothed representation of the data distribution. This initial exploration, shown in Figure 5.5, sets the foundation for subsequent analyses.

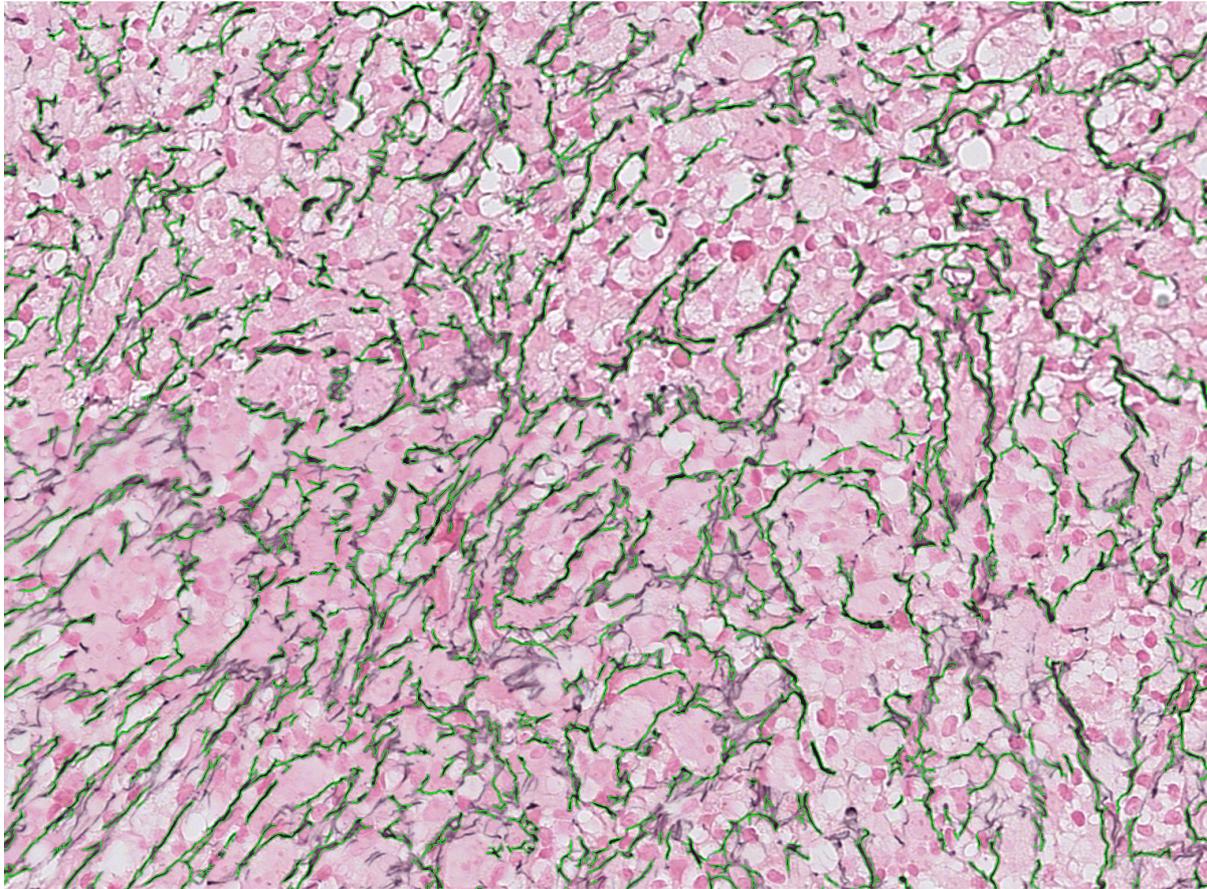


Figure 5.3: Example of segmentation of the BMF in a region of a WSI. Fibrotic components are delineated in green.

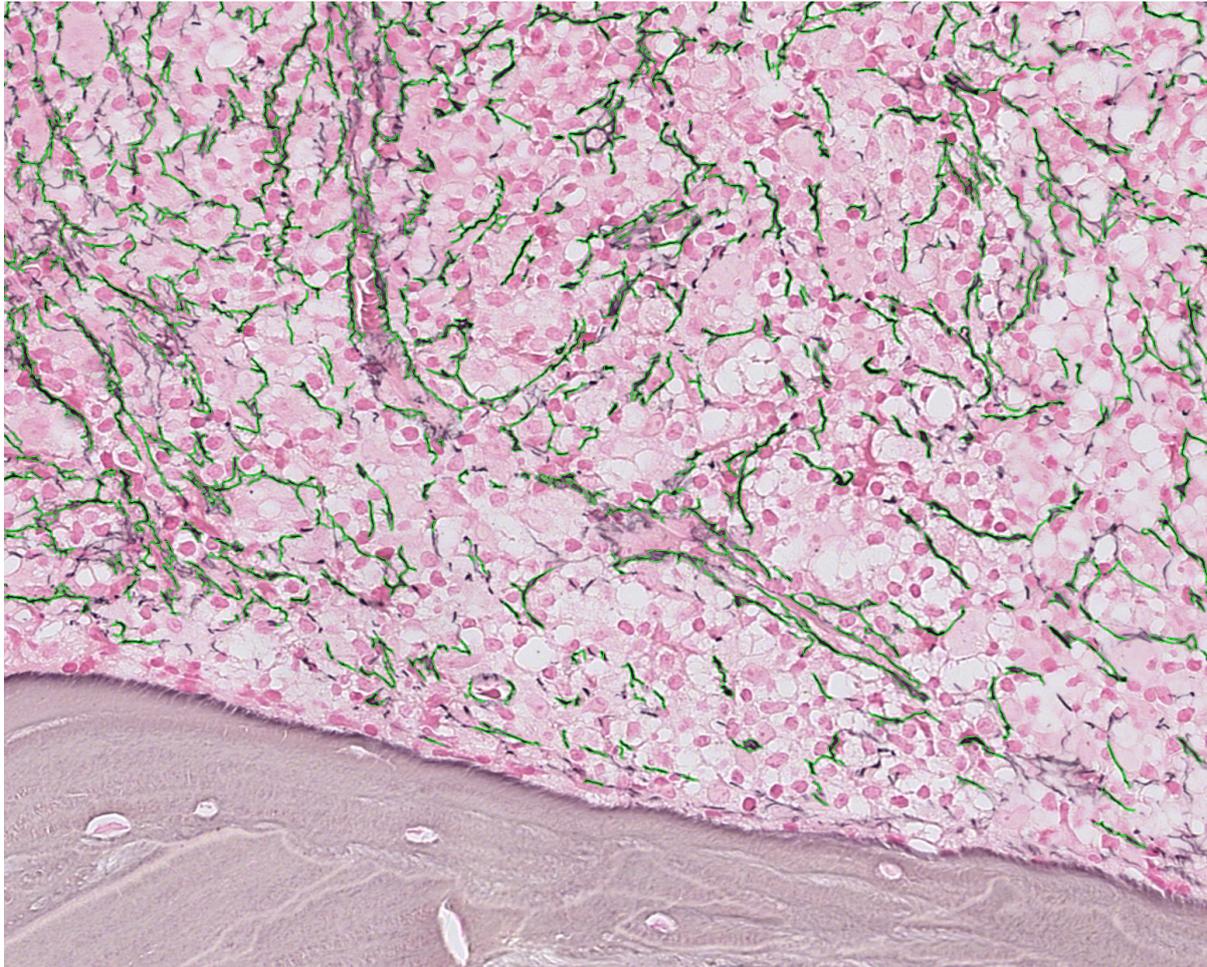


Figure 5.4: Example of segmentation of the BMF in a region of a WSI. Fibrotic components are delineated in green.

To discern potential patterns specific to different hematological conditions, we present KDE plots categorized by diagnosis (Figure 5.7). This exploration aims to uncover any distinctive trends or variations in the distribution of BMF ratios among various patient groups.

Since the distributions were overlapped, a better visualization of data can be given by the box plots (Figure 5.6). These visual tool enables a comparative analysis of the distribution, highlighting potential differences or similarities among diagnostic groups. Trying to emphasize a trend in data, we decided to applying to data the Otsu threshold method, which is commonly used in image processing, working as a one-dimensional discrete analogue of Fisher's Discriminant Analysis. A Multi-Otsu is applied to data in order to find an optimal grouping within the regions of the histogram. The Multi-Otsu method, applied to the data with a configuration for four classes, yielded thresholds of 0.03, 0.07, and 0.15, as illustrated in Figure 5.7 by the red vertical lines. The decision to opt for four classes was influenced by the limited number of samples in the MDS/MPN class, and the box plots suggested that it could potentially be treated similarly to the MPN class. Figure 5.8 illustrates the percentage of data associated with each diagnosis within the regions identified through the previously described method.

### 5.3.2 Comparative Analysis

In this section, we undertake a comparative analysis focusing on two distinct scenarios: CTR vs Disease and all diagnoses vs each other.

#### Control vs Disease

Firstly, in Figure 5.9, histograms and KDE plots depict the distribution of data for the CTR group and all other conditions treated collectively as a single class. While visual inspection suggests potential differences, it is essential to use statistical tests for rigorous evaluation of the observed distinctions. To proceed with the statistical evaluation, the data values are organized into a contingency table, as illustrated in Figure 5.10. By defining the condition of pathological grade of fibrosis in relation to the median, we aim to discern whether there is a statistically significant difference between the CTR and the combined set of other conditions.

The results of the Fisher Exact Test between CTR and other diagnoses are an odds ratio of 1.8 and a p-value of  $2 \times 10^{-3}$  suggesting a statistically significant result. Other statistical tests, including the Mann-Whitney U-test (Normalized U = 0.6, p-value =  $4 \times 10^{-5}$ ) and Kolmogorov-Smirnov test (Statistics = 0.19, p-value =  $4 \times 10^{-5}$ ), consistently reveal differences in terms of distribution between the CTR and other diagnoses.

### **All Diagnoses vs Each Other**

In Figure 5.6, is shown the boxplot for each diagnoses. Even in this case, to find statistically significant difference some statistical test were computed. Thus, for each pair, we generated contingency tables similar to the one for CTR vs Disease, shown in Figure 5.11.

In Figure 5.12 are shown the p-values and the the odds-ratio of the Fisher Exact Test for each pair of diagnosis. When conducting multiple tests, it becomes crucial to address the issue of inflated Type I error rates that can arise due to conducting numerous pairwise comparisons. To control the overall risk of making a Type I error, we can use the Bonferroni correction method, consisting on dividing the significance level by the number of tests. Even in this case Kolmogorov-Smirnov and Mann-Whitney tests were used and the results are shown in Figure 5.14 and 5.13.

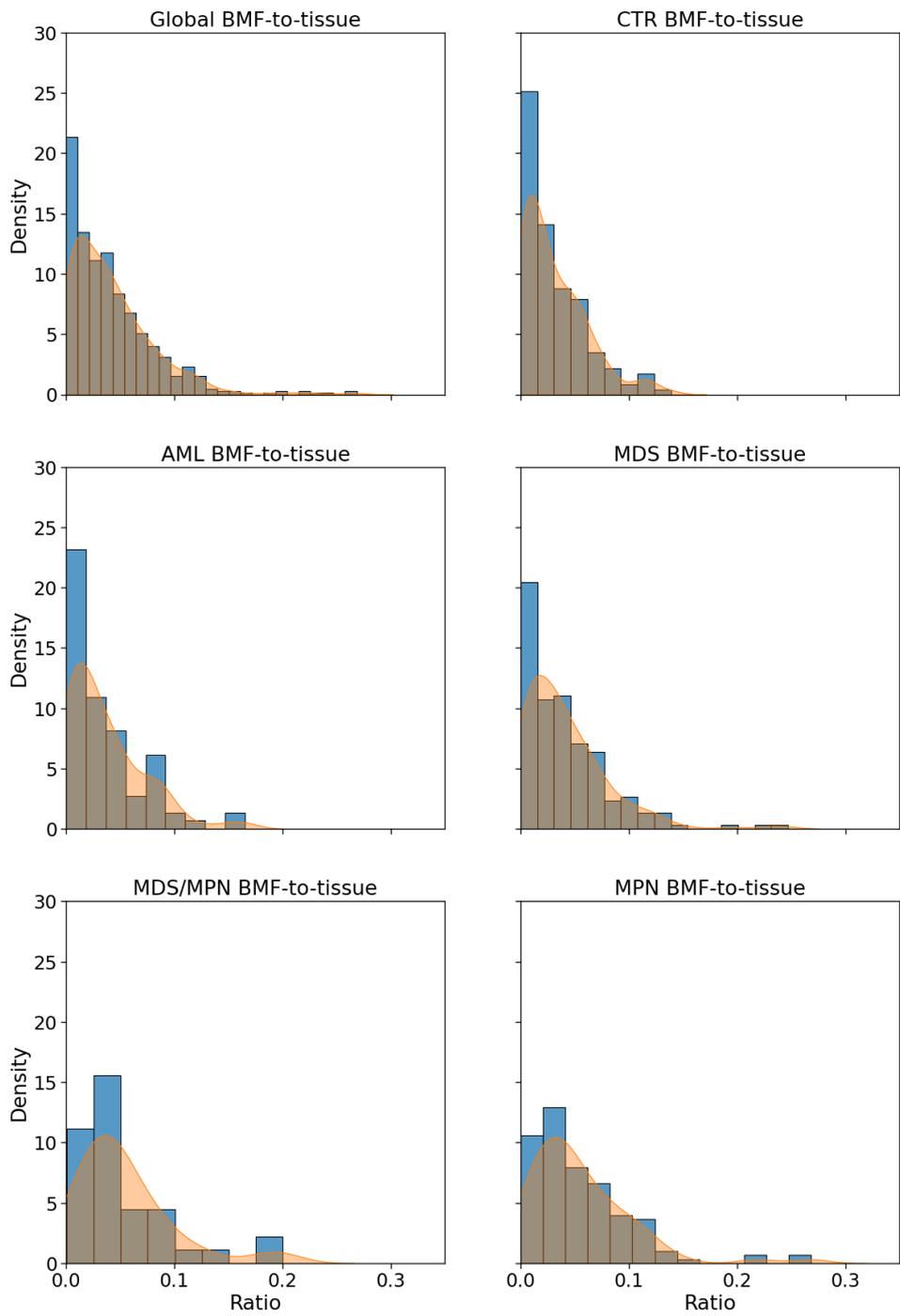


Figure 5.5: Normalized histograms (blue) and Kernel Density Estimation plots (orange), depicting the distributions of the Bone Marrow Fibrosis (BMF) to tissue ratio within the WSI.

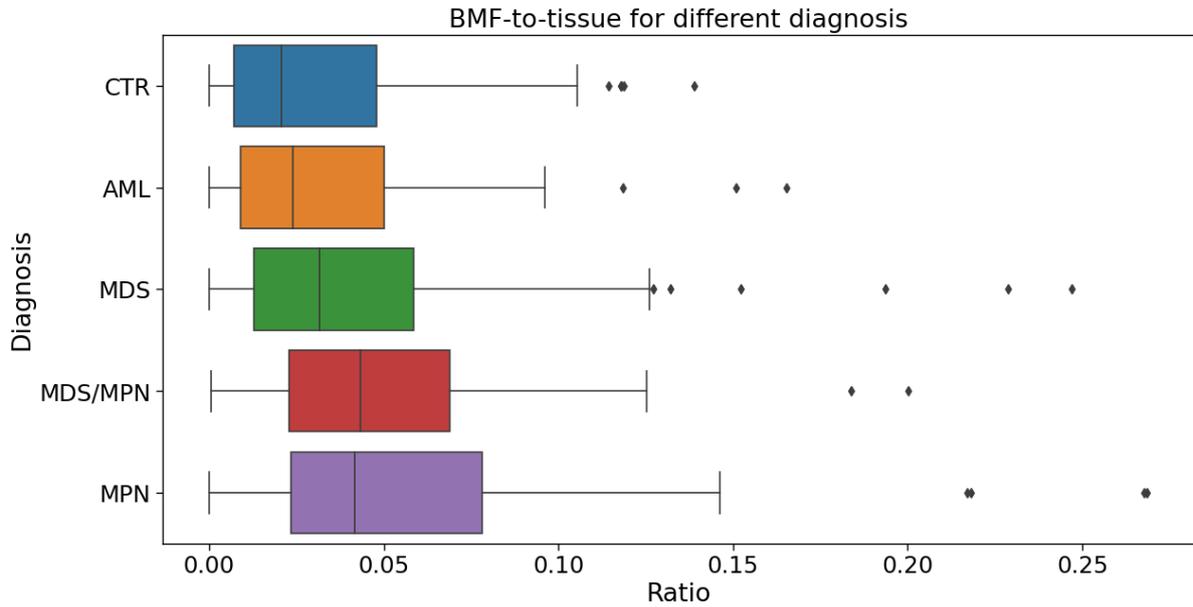


Figure 5.6: Box-plot with the Bone Marrow Fibrosis to tissue ratio for each diagnosis.

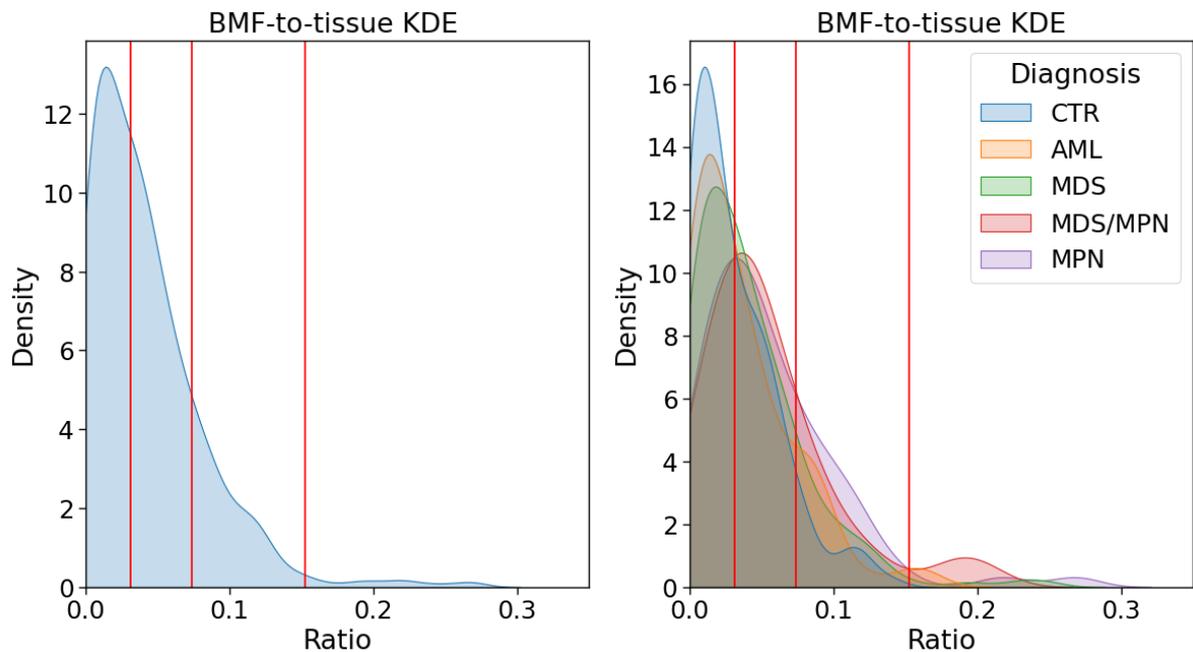


Figure 5.7: Kernel Density Estimation plots of the distribution of the BMF to soft tissue ratio. On the right there are the KDEs related to each diagnosis. The red lines identify the thresholds obtained with a Multi-Otsu method in order to find an optimal way to group data.

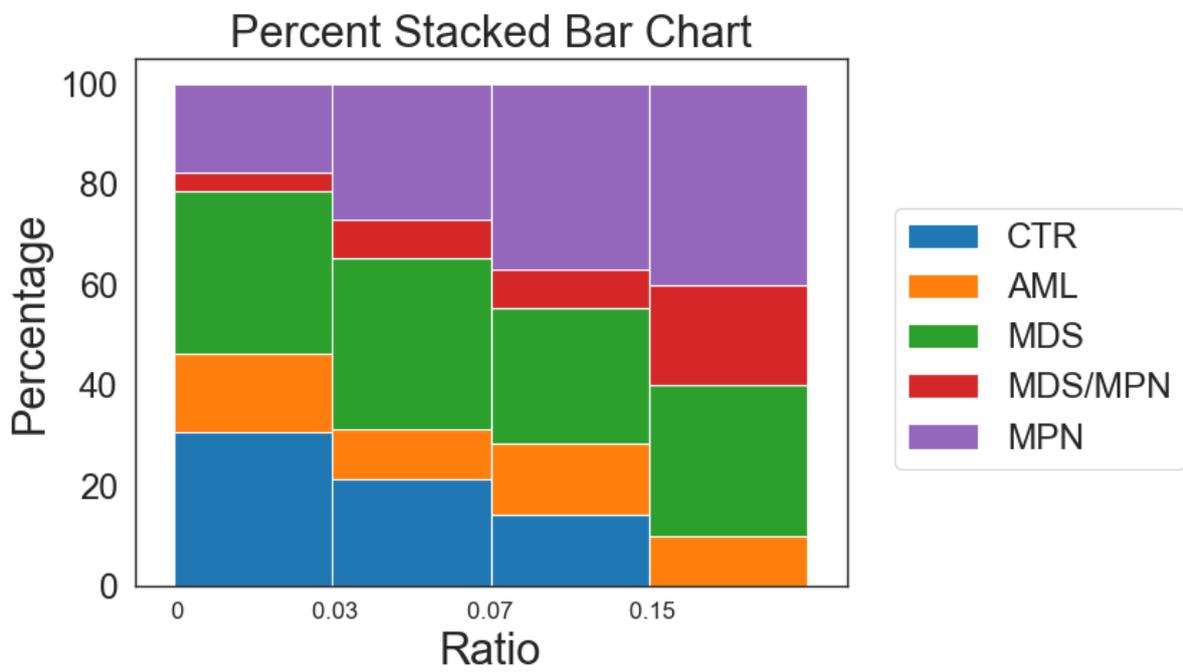


Figure 5.8: Percentage of data associated with each diagnosis within the regions identified through the Multi-Otsu method.

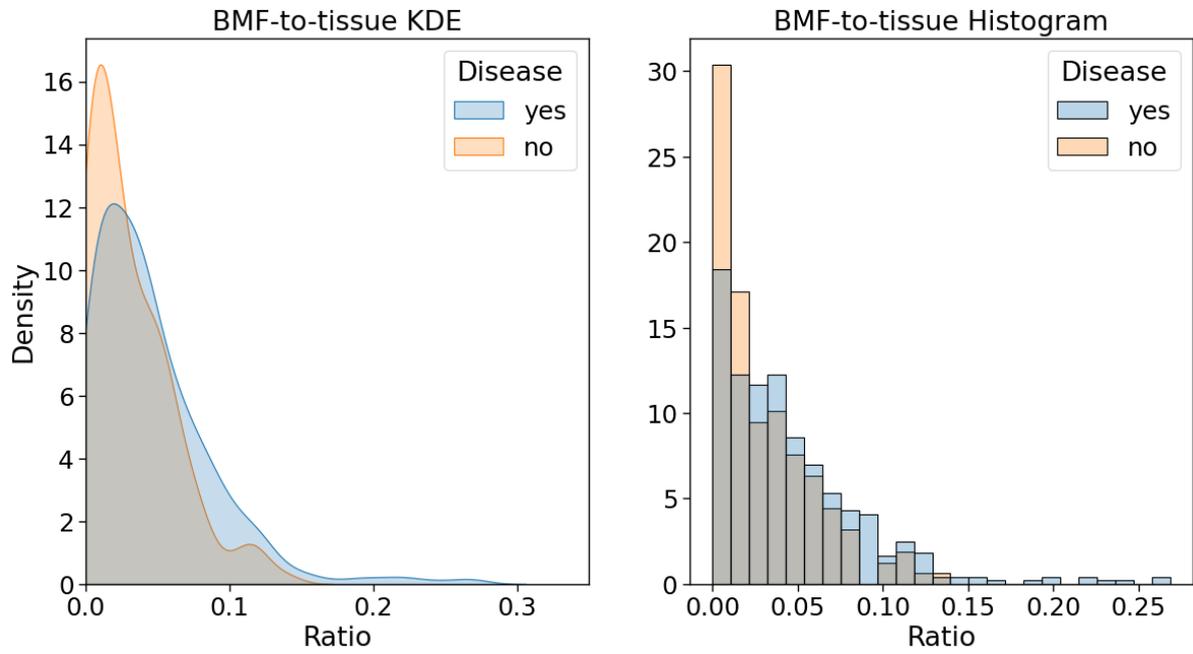


Figure 5.9: Kernel Density Estimation plots (left) and normalized histograms (right), of the control group and disease group.

Contingency table

Disease	no	90	57
	yes	211	244
		no	yes
		High Ratio (>median)	

Figure 5.10: Contingency table of the conditions of disease and high percentage of fibrosis with respect the the tissue. Threshold between high and low fibrosis is set by the median.



Figure 5.11: Contingency table of each couple of diagnosis and high percentage of fibrosis with respect to the hematopoietic tissue. Threshold between high and low fibrosis is set by the median.

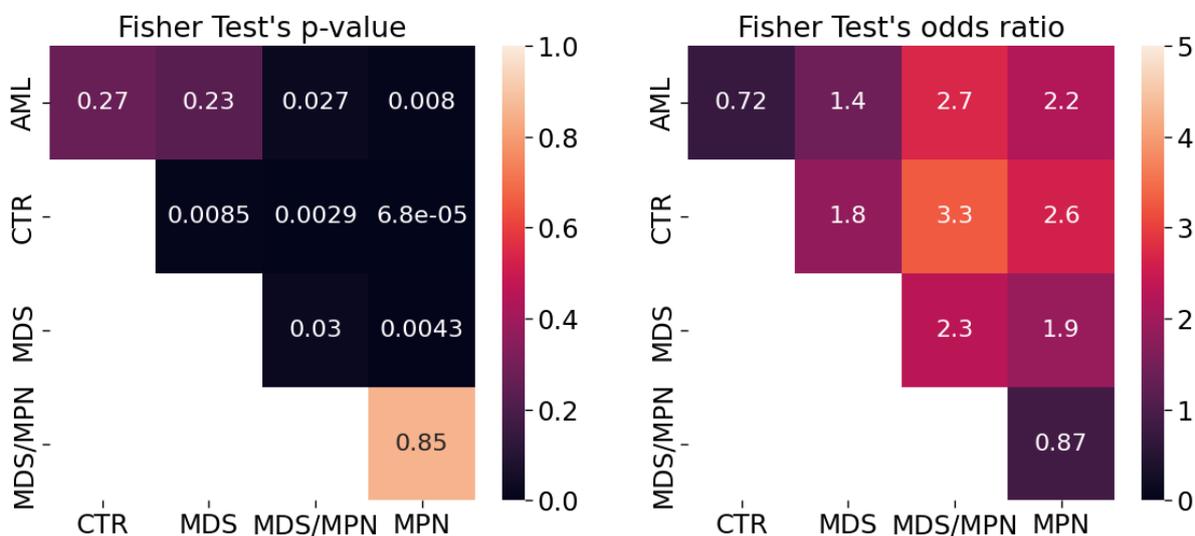


Figure 5.12: Results of the Fisher Exact test based on the contingency tables in Figure 5.11. Odds ratio for each pair of diagnosis are shown on the right, while on the left are reported the p-values.

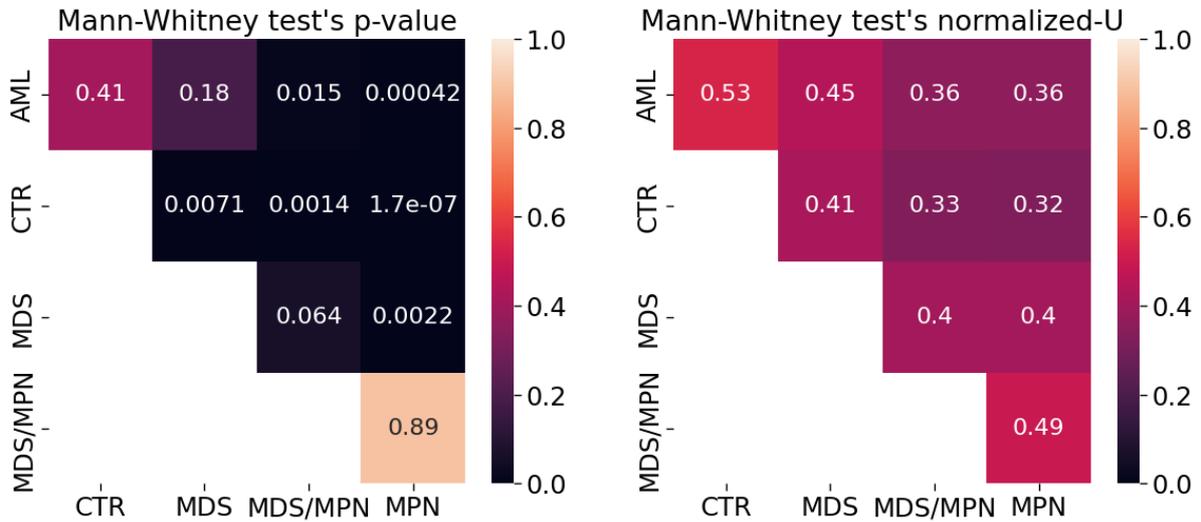


Figure 5.13: Results of the Mann-Whitney U-test on each pair of distributions. Normalized U statistics for each pair of diagnosis are shown on the right, while on the left are reported the p-values.

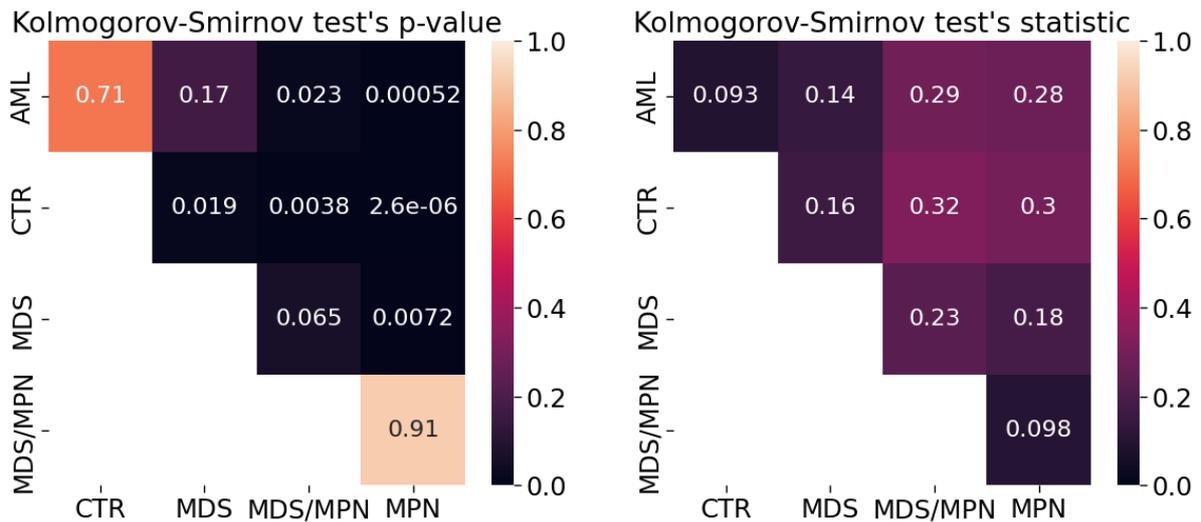


Figure 5.14: Results of the Kolmogorov-Smirnov test on each pair of distributions. Statistics for each pair of diagnosis are shown on the right, while on the left are reported the p-values.

# Chapter 6

## Discussion

This work was divided into two main components: the initial phase involves the curation of the dataset and subsequent segmentation of WSIs, while the latter part focuses on the analysis of the measured fibrotic component resulting from the segmentation.

The incorporation of staining decomposition in the dataset curation process stands out as a critical aspect of this study. By implementing conversion matrices, the isolation of tissue across all slides was successfully accomplished, and have been used as a selection methods to assess the quality of an digital slides to be included in the dataset. Moreover, for future investigations, there is potential to include WSIs that may have degraded over time or contain coloring errors. This entails exploring transformations between different color spaces, offering a prospective avenue to augment the dataset and incorporate additional valuable data.

The application of the staining decomposition technique extended beyond dataset curation to the identification of hematopoietic tissue, an important parameter used for assessing the BMF-to-tissue ratio. The methods employed in this study have demonstrated a high level of accuracy in segmentation, particularly in challenging regions containing bone and adipose tissue. Identifying tissue in these areas posed a greater difficulty, but the implemented techniques ensured precision. This precision is crucial, as it guarantees that the search for Bone Marrow Fibrosis (BMF) was conducted only where necessary, preventing computational waste. Additionally, this accuracy ensures the proper computation of the BMF-to-tissue ratio, contributing to the reliability and validity of the

study’s findings.

While the methodology achieved success in identifying tissue, it did not extend to the identification of fibrotic components using the same color decomposition technique. This choice stems from the fact that discerning fibrosis requires a high resolution. In this regard, the limitations of color decomposition in effectively capturing the heterogeneity and variability of high-resolution images ruled out its application for fibrotic component identification. This highlights the intricate nature of fibrotic patterns and emphasizes the necessity for alternative methodologies to tackle this specific task.

However, the adoption of color decomposition took a central role in the annotation of initial images, acting as a kick-start for implementing an active semi-supervised learning (ASSL) pipeline. Faced with the challenge of starting with the total absence of annotated data, this method facilitated the creation of a dataset of GOMORI-stained bone marrow biopsies. In particular, the validation of 2611 ground truth patches by medical professionals was achieved without the need for manual annotation. This approach significantly accelerated the otherwise time-consuming and critical phase of dataset annotation with a noteworthy speed-up.

An examination of the metrics from the latest training round reveals a great performance of the model. The absence of overfitting and the high accuracy in segmentation highlight the efficacy of the chosen model. However, it is proper to notice that, in discussions with domain experts, a conscious trade-off between the achieved accuracy and time constraints was assumed acceptable. This decision reflects the pragmatic considerations inherent in the research process, balancing the quest for precision with the practicalities of limited time resources.

The segmentation of both tissue and fibrosis has generated the masks necessary for quantifying fibrosis. A preliminary look at the data reveals that the distributions are quite similar, exhibiting a tail but with the majority of samples clustered near zero. The most interesting aspect, hinting at differences among various diagnoses, lies in samples with a more substantial fibrotic component. In some instances, these samples appear to exhibit a secondary peak. However, it’s essential to acknowledge that some of these data points might simply represent outliers, given the notably smaller quantity of data in that segment of the histogram.

Boxplots seem to highlight a discernible trend among the various data, emphasized by their arrangement. A similar outcome was achieved using a Multi-Otsu threshold algorithm. Optimal regions within the histogram were identified to analyze the trend of data pertaining to different diagnoses. The figure illustrates three distinct behaviors (Figure 5.8). The control group witnesses its percentage diminishing progressively compared to other diagnoses, eventually disappearing in the latter part of the histogram. Conversely, MPN and MDS/MPN observe their percentage steadily increasing towards segments of the histogram with higher BMF-to-tissue ratios. AML and MDS, on the other hand, appear to remain relatively constant.

This analysis suggests a nuanced relationship between the diagnosed conditions and the distribution of fibrotic components. The varying behaviors observed in the histogram and the trend highlighted by boxplots and threshold algorithms provide valuable insights into potential patterns associated with different diagnoses. However, the presence of outliers in the data should be carefully considered, and further statistical analyses have been necessary to confirm and quantify the observed trends.

The initial comparative analysis aimed to assess whether fibrosis could serve as an indicator to differentiate between the control group and those with a disease. The examination of the contingency table highlights that a high concentration of fibrosis, identified in this case by the median, is indeed an indicator (FT: odds ratio = 1.8, p-value =  $2 \times 10^{-3}$ ). The control group exhibits a relatively lower percentage of fibrosis compared to the group with diseases. The same outcome was confirmed through non-parametrical tests on distributions, showing a discernible differences between the two distributions (MW: normalized U = 0.61, p-value =  $4 \times 10^{-5}$ ; KS: statistics = 0.19, p-value =  $4 \times 10^{-5}$ ).

The focus has then shifted to exploring the differences between the control group and each individual diagnosis, followed by an examination of distinctions among the various diagnoses. Firstly, it is evident that the CTR cannot be differentiated from AML in any of the three tests presented (FT: odds ratio = 0.72, p-value = 0.27; MW: normalized U = 0.53, p-value = 0.41; KS: statistics = 0.09, p-value = 0.71). This observation aligns with existing literature, where BMF is not typically presented as a diagnostic indicator but rather as a prognostic factor for overall survival [61].

Moving on to the comparison with MDS, following adjustment for p-value, no signif-

icant differences were observed (FT: odds ratio = 1.8, p-value = 0.009; MW: normalized  $U = 0.41$ , p-value = 0,007; KS: statistics = 0.16, p-value = 0.019). Exploring subtypes within the disease could provide valuable insights. Literature reports indicate that approximately 10% to 20% of total MDS cases are classified as MDS with fibrosis (MDS-f), signifying a notable presence of fibrotic concentration in these cases [11].

The MDS/MPN overlap syndrome, inheriting characteristics from both MDS and MPN, is interesting not only for its distinction from the CTR but also for its comparison with diseases closely related to it. All tests highlight significant differentiation from the controls (FT: odds ratio = 3.3, p-value = 0.003; MW: normalized  $U = 0.33$ , p-value = 0,0014; KS: statistics = 0.32, p-value = 0.004). However, the comparison with MPN seems to suggest that the data essentially emanate from the same distribution (FT: odds ratio = 0.87, p-value = 0.85; MW: normalized  $U = 0.49$ , p-value = 0,89; KS: statistics = 0.10, p-value = 0.91). Additionally, the comparison with MDS does not reveal significant differences (FT: odds ratio = 2.3, p-value = 0.03; MW: normalized  $U = 0.4$ , p-value = 0.06; KS: statistics = 0.23, p-value = 0.07). It's crucial to note the limited number of samples for MDS/MPN (36), which may impact the statistical power of the analysis.

Finally, we turn to MPN, where the comparison with the control group exhibits a remarkably high level of significance (FT: odds ratio = 2.6, p-value =  $7 \times 10^{-5}$ ; MW: normalized  $U = 0.32$ , p-value =  $1.7 \times 10^{-7}$ ; KS: statistics = 0.3, p-value =  $3 \times 10^{-6}$ ). This outcome is not surprising, considering that BMF is a central pathological feature and the primary diagnostic criterion for myelofibrosis, a subtype of MPN.

# Chapter 7

## Conclusions

A comprehensive study was conducted on a dataset comprising 1123 WSIs of bone marrow biopsies stained with GOMORI, to quantify the BMF in order to use it as an indicator to stratify patients with several emathological disorders.

After the dataset curation involved employing a color decomposition technique, through which we excluded degraded WSIs, we quantified the amount of BMF within the slides. The process involved the identification of both the hematopoietic tissue and the bone marrow fibrosis, within the digital slides. Segmentation of tissue was obtained with the same color decomposition technique. The segmentation process proved highly effective, successfully excluding bone and adipose tissues.

However, the variety and complexity of high-resolution images posed more challenges in the identification of the fibrosis, necessitating the exploration of alternative techniques for this task. In addressing the segmentation of BMF, an ASSL pipeline was employed. Initiating with a color decomposition kickstart, this approach facilitated the creation of a dataset with annotated patches validated by experts as ground truth, eliminating the need for manual annotation. The absence of overfitting and the high accuracy in segmentation highlight the efficacy of the chosen model.

The obtained segmentation masks of tissue and fibrosis, were instrumental in evaluating the quantity of fibrosis, calculated as the area of BMF divided by the area of hematopoietic tissue. Comparative analyses were conducted between CTR and Disease groups, as well as across all diagnoses. The control group exhibits a relatively lower

percentage of fibrosis compared to the group with diseases. About the differences between each diagnosis based of the amount of fibrosis, we found results consistent with the literature. AML exhibited no significant differentiation from CTR as well as MDS. Contrastingly, a highly significant differentiation was observed in MPN, reinforcing BMF's pivotal role in myelofibrosis diagnosis.

To conclude, it is pertinent to consider avenues for improvement. Enhancements could be achieved by expanding the dataset, particularly by retrieving degraded images. Additionally, increasing rounds in the ASSL pipeline could refine segmentation accuracy. Furthermore, the incorporation of more clinical data, such as cytology or genomics, would provide a broader context for a more in-depth analysis.

# Bibliography

- [1] Martin Abadi et al. “TensorFlow: A System for Large-Scale Machine Learning”. In: *Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation*. OSDI’16. Savannah, GA, USA: USENIX Association, 2016, pp. 265–283. ISBN: 9781931971331.
- [2] Esther Abels et al. “Computational pathology definitions, best practices, and recommendations for regulatory guidance: a white paper from the Digital Pathology Association”. In: *J Pathol* 249.3, Epub 2019 Sep 3 (2019), pp. 286–294. DOI: 10.1002/path.5331.
- [3] Babak E Bejnordi et al. “Stain Specific Standardization of Whole-Slide Histopathological Images”. In: *IEEE Transactions on Medical Imaging* 35.2 (2016), pp. 404–415. DOI: 10.1109/TMI.2015.2476509.
- [4] E Bercovich and MC Javitt. “Medical Imaging: From Roentgen to the Digital Revolution, and Beyond”. In: *Rambam Maimonides Medical Journal* 9.4 (2018). Review, e0034. DOI: 10.5041/RMMJ.10355. URL: <https://doi.org/10.5041/RMMJ.10355>.
- [5] J.C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum Press, 1981. DOI: 10.1007/978-1-4757-0450-1. URL: <http://dx.doi.org/10.1007/978-1-4757-0450-1>.
- [6] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.
- [7] J. Canny. “A Computational Approach to Edge Detection”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-8.6 (Nov. 1986), pp. 679–698. DOI: 10.1109/TPAMI.1986.4767851.

- [8] Hu Cao et al. *Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation*. 2021. arXiv: 2105.05537 [eess.IV].
- [9] Nico Curti et al. “Effectiveness of Semi-Supervised Active Learning in Automated Wound Image Segmentation”. In: *International Journal of Molecular Sciences* 24.1 (2023). ISSN: 1422-0067. DOI: 10.3390/ijms24010706. URL: <https://www.mdpi.com/1422-0067/24/1/706>.
- [10] G. Cybenko. “Approximation by superpositions of a sigmoidal function”. In: *Math. Control Signal Systems* 2 (1989), pp. 303–314. DOI: 10.1007/BF02551274. URL: <https://doi.org/10.1007/BF02551274>.
- [11] Matteo G Della Porta and Luca Malcovati. “Myelodysplastic syndromes with bone marrow fibrosis”. In: *Haematologica* 96.2 (Feb. 2011), pp. 180–183. DOI: 10.3324/haematol.2010.039875. eprint: 21282718.
- [12] Matteo G Della Porta et al. “Clinical relevance of bone marrow fibrosis and CD34-positive cell clusters in primary myelodysplastic syndromes”. In: *Journal of Clinical Oncology* 27.5 (Feb. 2009), pp. 754–762. DOI: 10.1200/JCO.2008.18.2246. eprint: 19103730.
- [13] Nikolas Dimitriou, Ognjen Arandjelović, and Peter D. Caie. “Deep Learning for Whole Slide Image Analysis: An Overview”. In: *Frontiers in Medicine* 6 (2019), p. 264. DOI: 10.3389/fmed.2019.00264.
- [14] Alexey Dosovitskiy et al. “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale”. In: *CoRR* abs/2010.11929 (2020). arXiv: 2010.11929. URL: <https://arxiv.org/abs/2010.11929>.
- [15] Martin Ester et al. “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise”. In: *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*. KDD’96. Portland, Oregon: AAAI Press, 1996, pp. 226–231.
- [16] AJ Evans et al. “Validating Whole Slide Imaging Systems for Diagnostic Purposes in Pathology”. In: *Arch Pathol Lab Med* 146.4 (2022), pp. 440–450. DOI: 10.5858/arpa.2020-0723-CP.

- [17] Navid Farahani and Liron Pantanowitz. “Overview of Telepathology”. In: *Surg Pathol Clin* 8.2, Epub 2015 Apr 4 (2015), pp. 223–231. DOI: 10.1016/j.path.2015.02.018.
- [18] Navid Farahani, Anil Parwani, and Liron Pantanowitz. “Whole slide imaging in pathology: advantages, limitations, and emerging perspectives”. In: *Pathology and Laboratory Medicine International* 7 (2015), pp. 23–33. DOI: 10.2147/PLMI.S59826.
- [19] Andrew H Fischer et al. “Hematoxylin and Eosin Staining of Tissue and Cell Sections”. In: *CSH Protoc* 2008 (2008), pdb.prot4986. DOI: 10.1101/pdb.prot4986.
- [20] G Gomori. “GOMORI G. A rapid one-step trichrome stain”. In: *Am J Clin Pathol* 20.7 (July 1950), pp. 661–4. DOI: 10.1093/ajcp/20.7\_ts.661.
- [21] Rafael C. Gonzalez and Richard E. Woods. *Digital image processing*. Upper Saddle River, N.J.: Prentice Hall, 2008. ISBN: 9780131687288 013168728X 9780135052679 013505267X. URL: <http://www.amazon.com/Digital-Image-Processing-3rd-Edition/dp/013168728X>.
- [22] Adam Goode et al. “OpenSlide: A Vendor-Neutral Software Foundation for Digital Pathology”. In: *J Pathol Inform* 4 (2013), p. 27. URL: <http://www.jpathinformatics.org/text.asp?2013/4/1/27/119005>.
- [23] Hermann Grassmann et al. “On the theory of compound colors”. In: *Phil. Mag* 7 (1854), pp. 254–64.
- [24] R. M. Haralick, K. Shanmugam, and I. Dinstein. “Textural Features for Image Classification”. In: *IEEE Transactions on Systems, Man, and Cybernetics* SMC-3.6 (Nov. 1973), pp. 610–621. DOI: 10.1109/TSMC.1973.4309314.
- [25] Hermann von Helmholtz. *Handbuch der Physiologischen Optik*. Vol. 1. Leipzig: Voss, 1867.
- [26] Michael D. Herrmann et al. “Implementing the DICOM standard for digital pathology”. In: *Journal of Pathology Informatics* 9 (2018), p. 37. URL: <http://www.jpathinformatics.org/text.asp?2018/9/1/37/244882>.

- [27] Amit G Jain et al. “Myelodysplastic Syndromes with Bone Marrow Fibrosis: An Update”. In: *Annals of Laboratory Medicine* 42.3 (May 2022), pp. 299–305. DOI: 10.3343/al.m.2022.42.3.299. eprint: 34907099.
- [28] Wen-Xiong Kang, Qing-Qiang Yang, and Run-Peng Liang. “The Comparative Research on Image Segmentation Algorithms”. In: *2009 First International Workshop on Education Technology and Computer Science*. Vol. 2. 2009, pp. 703–707. DOI: 10.1109/ETCS.2009.417.
- [29] Salman H. Khan et al. “Transformers in Vision: A Survey”. In: *CoRR* abs/2101.01169 (2021). arXiv: 2101.01169. URL: <https://arxiv.org/abs/2101.01169>.
- [30] H. E. Kim et al. “Transfer learning for medical image classification: a literature review”. In: *BMC Medical Imaging* 22.1 (2022), p. 69. DOI: 10.1186/s12880-022-00793-7. URL: <https://doi.org/10.1186/s12880-022-00793-7>.
- [31] Philippe Lambin et al. “Radiomics: Extracting More Information from Medical Images Using Advanced Feature Analysis”. In: *Eur J Cancer* 48.4 (2012), pp. 441–446. DOI: 10.1016/j.ejca.2011.11.036. eprint: 22257792. URL: <https://pubmed.ncbi.nlm.nih.gov/22257792/>.
- [32] Ze Liu et al. “Swin Transformer: Hierarchical Vision Transformer using Shifted Windows”. In: *CoRR* abs/2103.14030 (2021). arXiv: 2103.14030. URL: <https://arxiv.org/abs/2103.14030>.
- [33] Zhenyu Liu et al. “The Applications of Radiomics in Precision Diagnosis and Treatment of Oncology: Opportunities and Challenges”. In: *Theranostics* 9.5 (2019), pp. 1303–1322. DOI: 10.7150/thno.30309. URL: <https://doi.org/10.7150/thno.30309>.
- [34] Laurens van der Maaten and Geoffrey Hinton. “Visualizing Data using t-SNE”. In: *Journal of Machine Learning Research* 9 (2008), pp. 2579–2605. URL: <http://www.jmlr.org/papers/v9/vandermaaten08a.html>.
- [35] Marc Macenko et al. “A method for normalizing histology slides for quantitative analysis”. In: *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. 2009, pp. 1107–1110. DOI: 10.1109/ISBI.2009.5193250.

- [36] J.B. MacQueen. “Some Methods for Classification and Analysis of Multivariate Observations”. In: *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*. Berkeley: University of California Press, 1967, pp. 281–297.
- [37] Shino Magaki et al. “An Introduction to the Performance of Immunohistochemistry”. In: *Methods Mol Biol* 1897 (2019), pp. 289–298. DOI: 10.1007/978-1-4939-8935-5\_25.
- [38] Andrzej Materka. “Texture analysis methodologies for magnetic resonance imaging”. In: *Dialogues Clin Neurosci* 6.2 (2004), pp. 243–250. DOI: 10.31887/DCNS.2004.6.2/amaterka. eprint: 22033841. URL: <https://pubmed.ncbi.nlm.nih.gov/22033841/>.
- [39] James Clerk Maxwell. “Theory of the perception of colors”. In: *Trans. R. Scottish Soc. Arts* 4 (1856), pp. 394–400.
- [40] Marius E. Mayerhoefer et al. “Introduction to Radiomics”. In: *J Nucl Med* 61.4 (2020), pp. 488–495. DOI: 10.2967/jnumed.118.222893. eprint: 32060219. URL: <https://pubmed.ncbi.nlm.nih.gov/32060219/>.
- [41] Warren S. McCulloch and Walter H. Pitts. “A Logical Calculus of the Ideas Immanent in Nervous Activity”. In: *Bulletin of Mathematical Biophysics* 5 (1943), pp. 115–133. DOI: 10.1007/BF02478259.
- [42] Leland McInnes, John Healy, and James Melville. *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*. 2020. arXiv: 1802.03426 [stat.ML].
- [43] Valerio Nardone et al. “Delta radiomics: a systematic review”. In: *Radiol Med* 126.12 (Dec. 2021), pp. 1571–1583. DOI: 10.1007/s11547-021-01436-7. eprint: <https://link.springer.com/article/10.1007/s11547-021-01436-7>. URL: <https://doi.org/10.1007/s11547-021-01436-7>.
- [44] Nobuyuki Otsu. “A Threshold Selection Method from Gray-Level Histograms”. In: *IEEE Transactions on Systems, Man, and Cybernetics* 9.1 (1979), pp. 62–66. DOI: 10.1109/TSMC.1979.4310076.

- [45] S Park, L Pantanowitz, and AV Parwani. “Digital imaging in pathology”. In: *Clin Lab Med* 32.4 (2012), pp. 557–584. DOI: 10.1016/j.c11.2012.07.006.
- [46] Simon J. D. Prince. *Understanding Deep Learning*. The MIT Press, Dec. 2023. ISBN: 9780262048644.
- [47] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-Net: Convolutional Networks for Biomedical Image Segmentation”. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Ed. by Nassir Navab et al. Vol. 9351. Lecture Notes in Computer Science. Springer, Cham, 2015. DOI: 10.1007/978-3-319-24574-4\_28.
- [48] Frank Rosenblatt. “The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain”. In: *Psychological Review* 65.6 (1958), pp. 386–408. DOI: 10.1037/h0042519.
- [49] C.E. Shannon. “Communication in the Presence of Noise”. In: *Proceedings of the IRE* 37.1 (1949), pp. 10–21. DOI: 10.1109/jrproc.1949.232969. URL: <https://doi.org/10.1109/jrproc.1949.232969>.
- [50] Rakesh Shiradkar et al. “Radiomics based targeted radiotherapy planning (RadTRaP): a computational framework for prostate cancer treatment planning with MRI”. In: *Radiat Oncol* 11.1 (2016), p. 148. DOI: 10.1186/s13014-016-0718-3. URL: <https://doi.org/10.1186/s13014-016-0718-3>.
- [51] Jonathan D. Shur et al. “Radiomics in Oncology: A Practical Guide”. In: *Radiographics* 41.6 (2021), pp. 1717–1732. DOI: 10.1148/rg.2021210037. eprint: 34597235. URL: <https://pubmed.ncbi.nlm.nih.gov/34597235/>.
- [52] Nahian Siddique et al. “U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications”. In: *IEEE Access* 9 (2021), pp. 82031–82057. DOI: 10.1109/ACCESS.2021.3086020.
- [53] Cheryl J Teman et al. “Quantification of fibrosis and osteosclerosis in myeloproliferative neoplasms: a computer-assisted image study”. In: *Leukemia Research* 34.7 (July 2010). Epub 2010 Jan 31, pp. 871–876. DOI: 10.1016/j.leukres.2010.01.005.

- [54] E. Torresani et al. “Diagnostic Concordance Between Traditional and Digital Workflows: A Study on 1427 Prostate Biopsies”. In: *Pathologica - Journal of the Italian Society of Anatomic Pathology and Diagnostic Cytopathology* 115.4 (Sept. 2023), pp. 221–226. DOI: 10.32074/1591-951X-896.
- [55] Ashish Vaswani et al. “Attention Is All You Need”. In: *CoRR* abs/1706.03762 (2017). arXiv: 1706.03762. URL: <http://arxiv.org/abs/1706.03762>.
- [56] World Health Organization. *Cancer*. Accessed: November 24, 2023. 2023. URL: [https://www.who.int/health-topics/cancer#tab=tab\\_1](https://www.who.int/health-topics/cancer#tab=tab_1).
- [57] Rikiya Yamashita et al. “Convolutional neural networks: an overview and application in radiology”. In: *Insights Imaging* 9.4 (2018), pp. 611–629. DOI: 10.1007/s13244-018-0639-9.
- [58] Thomas Young. “II. The Bakerian Lecture. On the theory of light and colours”. In: *Philosophical transactions of the Royal Society of London* 92 (1802), pp. 12–48.
- [59] Abdallah A Zahr et al. “Bone marrow fibrosis in myelofibrosis: pathogenesis, prognosis and targeted strategies”. In: *Haematologica* 101.6 (June 2016), pp. 660–671. DOI: 10.3324/haematol.2015.141283. eprint: 27252511.
- [60] Wenchao Zhang, Yu Guo, and Qiyu Jin. “Radiomics and Its Feature Selection: A Review”. In: *Symmetry* 15.10 (2023). ISSN: 2073-8994. DOI: 10.3390/sym15101834. URL: <https://www.mdpi.com/2073-8994/15/10/1834>.
- [61] Xia Zhang et al. “Significance of bone marrow fibrosis in acute myeloid leukemia for survival in the real-world”. In: *Frontiers in Oncology* 12 (Oct. 2022), p. 971082. DOI: 10.3389/fonc.2022.971082. eprint: 36276150.
- [62] Yuhao Zhang et al. *Contrastive Learning of Medical Visual Representations from Paired Images and Text*. 2021. URL: <https://openreview.net/forum?id=T4gXB0XoIUr>.
- [63] Baihui Zheng et al. “Radiomics score: a potential prognostic imaging feature for postoperative survival of solitary HCC patients”. In: *BMC Cancer* 18 (2018), p. 1148. DOI: 10.1186/s12885-018-5024-z. eprint: <https://bmccancer>.

biomedcentral.com/articles/10.1186/s12885-018-5024-z. URL: <https://doi.org/10.1186/s12885-018-5024-z>.

- [64] Hong-Yu Zhou et al. “ConvNets vs. Transformers: Whose Visual Representations are More Transferable?” In: *CoRR* abs/2108.05305 (2021). arXiv: 2108.05305. URL: <https://arxiv.org/abs/2108.05305>.
- [65] Alex Zwanenburg et al. “The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping”. In: *Radiology* 295.2 (2020). PMID: 32154773, pp. 328–338. DOI: 10.1148/radiol.2020191145. eprint: <https://doi.org/10.1148/radiol.2020191145>. URL: <https://doi.org/10.1148/radiol.2020191145>.