Escola Tècnica Superior d'Enginyeria ETSE-UV

Máster en Inteligencia Artificial Avanzada y Aplicada

Universitat de València

Trabajo Fin de Máster

Innovative Active Learning-Based Framework for Ulcerative Colitis Region Segmentation in Histological Images

Author:
Fernando García Torres

Tutors:
Valero Laparra
Rocío del Amor
Pablo Meseguer

VNIVERSITAT DE VALÈNCIA

Escola Tècnica Superior d'Enginyeria ETSE-UV

Trabajo Fin de Máster

Innovative Active Learning-Based Framework for Ulcerative Colitis Region Segmentation in Histological Images

Author:

Fernando García Torres

Tutors:

Valero Laparra
Rocío del Amor
Pablo Meseguer

Declaración de autoría:

Yo, Fernando García Torres, declaro la autoría del Trabajo Fin de Máster titulado "Innovative Active Learning-Based Framework for Ulcerative Colitis Region Segmentation in Histological Images" y que el citado trabajo no infringe las leyes en vigor sobre propiedad intelectual. El material no original que figura en este trabajo ha sido atribuido a sus legítimos autores.

Valencia, 8 de septiembre de 2023.

Fdo: Fernando García Torres

Resumen:

La incidencia de la colitis ulcerosa (UC), una enfermedad intestinal inflamatoria crónica, está experimentando un aumento mundial. En el tratamiento de la UC, es crucial reducir la inflamación intestinal, y la evaluación histológica desempeña un papel esencial, ya que la remisión histológica (RH) es el principal indicador de la enfermedad. Trabajos recientes en este campo han propuesto un novedoso índice denominado PHRI que cuantifica la actividad de la UC basándose en la presencia o ausencia de neutrófilos en diferentes compartimentos celulares: lámina propia, epitelio criptal, epitelio superficial y lumen criptal. La identificación de neutrófilos en las diferentes áreas de la imagen histológica para la gradación del índice PHRI utilizado para el diagnóstico de la UC es esencial, y el diagnóstico asistido por ordenador (CAD) se ha convertido en una valiosa herramienta. Debido a los avances en patología digital, los sistemas CAD se utilizan para digitalizar biopsias en *whole-slide images* (WSI) y permiten que sean compartidas entre centros clínicos y analizadas por varios patólogos. Este estudio introduce un algoritmo de *active learning* (AL), un enfoque innovador para aliviar la carga de trabajo de los patólogos mejorando la eficiencia en la identificación de regiones cruciales en estas imágenes para el diagnóstico de la colitis ulcerosa. Aprovechando una base de datos diversa de imágenes de diapositivas completas anotadas y un proceso de entrenamiento iterativo, el objetivo es optimizar el rendimiento del modelo de segmentación basado en un marco *encoder-decoder*.

Abstract:

The incidence of Ulcerative colitis (UC), a chronic inflammatory bowel disease, is experiencing a global increase. In the treatment of UC, it is crucial to reduce intestinal inflammation, and histological evaluation plays an essential role, as histological remission (HR) is the primary disease indicator. Recent works in the field have proposed a novel index called PHRI that quantifies UC activity based on the presence or absence of neutrophils in different cellular compartments: lamina propia, cryptal epithelium, surface epithelium and cryptal lumen. The identification of neutrophils in the different areas of the histological image for the grading of the PHRI index used for diagnosis of UC is essential, and computer-aided diagnosis (CAD) has become a valuable tool. Due to advances in digital pathology, CAD systems are used to digitalise biopsies into whole-slide images (WSI) and allow them to be shared between clinical centres and analysed by several pathologists. This study introduces an active learning (AL) algorithm, an innovative approach to alleviating the workload of pathologists by improving the efficiency in identifying crucial regions in these images for ulcerative colitis diagnosis. Leveraging a diverse database of annotated whole-slide images and an iterative training process, the aim is to optimize the performance of the encoder-decoder-based segmentation model.

Resum:

La incidència de la colitis ulcerosa (UC), una malaltia intestinal inflamatòria crònica, està experimentant un augment a nivell mundial. En el tractament de la UC, és crucial reduir la inflamació intestinal, i l'avaluació histològica juga un paper essencial, ja que la remissió histològica (RH) és el principal indicador de la malaltia. Treballs recents en aquest camp han proposat un nou índex anomenat PHRI que quantifica l'activitat de la UC basant-se en la presència o absència de neutròfils en diferents compartiments cel·lulars: làmina pròpia, epitelis criptal, epitelis superficial i llumen criptal. La identificació de neutròfils en les diferents àrees de la imatge histològica per a la graduació de l'índex PHRI utilitzat per al diagnòstic de la UC és essencial, i el diagnòstic assistit per ordinador (CAD) s'ha convertit en una eina valuosa. A causa dels avanços en patologia digital, els sistemes CAD s'utilitzen per digitalitzar biòpsies en *whole-slide images* (WSI) i permeten que siguen compartides entre centres clínics i analitzades per diversos patòlegs. Aquest estudi introdueix un algoritme d'*active learning* (AL), un enfocament innovador per alleujar la càrrega de treball dels patòlegs millorant l'eficiència en la identificació de les regions crucials en aquestes imatges per al diagnòstic de la colitis ulcerosa. Aprofitant una base de dades diversa d'imatges de diapositives completes anotades i un procés d'entrenament iteratiu, l'objectiu és optimitzar el rendiment del model de segmentació basat en un marc *encoder-decoder*.

Agradecimientos:

En primer lugar, quiero agradecer a mis tutores, por su orientación, paciencia y apoyo constante a lo largo de este proceso. Especialmente a Pablo y Rocío, ya que sus conocimientos y consejos fueron fundamentales para el desarrollo de esta investigación. Por último, a Valery, por la oportunidad, y a mis compañeros del CVBLab, a mis amigos y familia, por su respaldo y apoyo durante este desafío académico.

List of acronyms

AI	Artificial Intelligence
AL	Active Learning
CAD	Computer-Aided Diagnosis
CNN	Convolutional Neural Network
DL	Deep Learning
GT	Ground Truth
HR	Histological Remission
IBD	Inflammatory Bowel Disease
PHRI	PICaSSO Histologic Remission Index
UC	Ulcerative Colitis
WSI	Whole-Slide Image

Contents

List of Figures

List of Tables

Part I

Memory

Chapter 1

Introduction

1.1. Motivation and description of the problem

Ulcerative colitis (UC) is a chronic inflammatory bowel disease (IBD) characterized by inflammation and ulcers in the lining of the large intestine (colon) and rectum. The motivation for studying UC is rooted in its significant impact on the quality of life of affected individuals [5]. Understanding, diagnosing, and effectively treating UC are crucial for the successful management of the disease. UC has seen an increasing incidence worldwide [6]. It affects individuals of all ages but often manifests in young adulthood. Understanding its causes and mechanisms is essential to address this rising trend.

UC typically follows a relapsing-remitting cycle, with periods of active inflammation (flares) followed by remission. Managing these cycles effectively is critical to reducing symptoms and complications. UC also can vary in severity and extent and can be divided into categories depending on its extension, like proctitis, left-sided colitis or pancolitis [7]. The primary goal of UC treatment is to induce and maintain remission while improving the patient's quality of life. This involves:

1. Symptom control: Alleviating symptoms such as diarrhoea, abdominal pain, and rectal bleeding.

2. Inflammation reduction: Minimizing inflammation to promote colon lining healing.

3. Preventing relapses: Reducing the frequency and severity of disease flares.

4. Minimizing complications: Preventing or managing complications like colonic strictures, perforations, and colorectal cancer, which can arise from chronic inflammation.

5. Improving quality of life: Enhancing the patient's overall well-being, as UC can significantly impact daily activities.

Histological remission, beyond clinical remission, is an essential treatment target. It involves the complete healing of the mucosal lining of the colon, which can reduce the risk of relapse and complications [8]. Patients diagnosed with UC typically undergo clinical procedures like colonoscopies, during which endoscopists collect colon biopsies for pathologists to examine under a microscope, searching for disease-specific patterns.

To manage the disease, computer-aided diagnosis has become an indispensable tool in UC, complementing the expertise of pathologists and leading to more accurate, efficient, and personalized patient care. Digital pathology involves acquiring, managing, and interpreting pathology information using computer technology. In the context of UC, it can be used to assist pathologists in accurately diagnosing UC by providing digital slides for examination, and efficiently storing and retrieving large amounts of histological data. As digital pathology advances, biopsies are now transformed into whole-slide images (WSI), which are high-resolution, gigapixel images. These can be easily shared among medical centres and reviewed by various pathologists for collaborative study.

Deep learning, a subset of machine learning, is particularly valuable in the field of digital pathology for UC. Deep learning models can segment digital pathology images to identify inflammation, ulceration, and healthy tissue regions in colon biopsies. This enables a more detailed analysis of the disease's extent and severity. These models can assist pathologists by automatically detecting and quantifying histological features associated with UC, providing objective metrics for disease severity and healing. Nevertheless, the correct identification of tissues and structures by these segmentation algorithms requires exhaustive and time-consuming prior annotation work by specialist personnel, who must manually select the pixels belonging to one region or another.

In this context, this work presents an innovative algorithm based on Active Learning (AL), applied for the first time to the segmentation of regions of interest in the context of ulcerative colitis. This algorithm aims to significantly reduce the workload of pathologists by improving the efficiency of identifying these critical regions in histological images. Through a careful selection of data samples and an iterative training process, the algorithm allows for optimising the performance of the segmentation model, rooted in encoder-decoder techniques, which in turn results in a faster and more accurate diagnostic process for this disease. This framework strikes a balance between acquiring a diverse database of annotated WSIs while managing the annotation time cost for pathologists. The project relies on an extensive private database comprising WSIs from UC patients, along with pixel-level annotations for selected biopsies.

1.2. Project framework

This project falls under the research initiative 'Development of Artificial Intelligence using i-scan videos and digital histological images', led by the University of Birmingham. The project is also referred to as PICaSSO because of the development of a novel endoscopic scoring system for ulcerative colitis, known as the Paddington International virtual ChromoendoScopy ScOre (PICaSSO) [9].

The primary objective of the PICaSSO project is to create a comprehensive system for caring for patients with ulcerative colitis, which integrates endoscopic procedures and histological analysis. In pursuit of this aim, a groundbreaking scoring system (PICaSSO) was introduced to assess UC through endoscopy, demonstrating superior interobserver agreement compared to existing endoscopic scoring criteria [9]. They also created the PICaSSO Histologic Remission Index (PHRI), a simplified and neutrophil-based scoring system designed for tracking mucosal healing at the histological level [4]. Its main advantage, when contrasted with previous histologic remission indexes, lies in its simplicity and reliance on neutrophils, which diminishes the subjectivity typically linked with pathologists. This characteristic also renders the score well-suited for integrating deep

learning-based computer-aided diagnostic (CAD) systems.

The data utilised in this project were acquired through the study titled "A multicenter, international validation study of the i-scan endoscopic scoring system and a new histologic scoring system to define subtle mucosal inflammation in ulcerative colitis" (i-scan), which received approval from the West Midlands Research Ethics Committee (17/WM/0223).

The UPV, notably the Computer Vision and Behavior Analysis Lab (CVBLab), is actively involved in this project by designing and implementing automated algorithms for examining WSI (Whole Slide Images) from patients diagnosed with UC. These algorithms are rooted in deep learning methods. They are intended to categorise the acquired WSI as either indicative of histological remission or activity, following the novel, simplified, neutrophil-based scoring system known as PHRI.

1.3. Objetives

The primary objective of this TFM is to design and develop an innovative algorithm based on Active Learning, applied to the automatic segmentation of regions of interest of WSIs of patients with ulcerative colitis. To achieve the main objective, secondary objectives are detailed below:

1. Compile the histological image database, which comprises high-resolution WSIs from an international multicentre study. The management of these images poses challenges due to their considerable file size, ranging from 500 MB to 3 GB. The biopsies were acquired and digitised using similar methodologies at several centres and subsequently shared with the corresponding annotations provided by the pathologists.

2. Conduct a literature review to study the state-of-the-art (SoA) techniques of studies where active learning approaches have been employed for WSI segmentation with deep learning algorithms.

3. Perform pre-processing of the database. This process includes transforming the annotations into a label of the various areas of interest and removal of the image background.

4. Perform patient-level partitioning of the database. These partitions are conditioned by the availability of pixel-level annotation in certain WSIs.

5. Design and develop a segmentation algorithm based on deep learning for the prediction of regions of interest in WSI of patients diagnosed with UC.

6. Design and develop an active learning methodology to reduce the pathologists' workflow.

7. Analyse the minimum and efficient number of samples required to improve the performance of the region of interest segmentation network.

8. Propose future directions for the improvement of results obtained considering problems and constraints encountered during the project.

1.4. Document structure

In Chapter 2, ulcerative colitis is detailed in terms of epidemiology, scoring, and management. A brief review of the CAD systems implemented for UC diagnosis and treatment, which focus on endoscopy, is also presented alongside a novel index for histologic scoring in UC called PHRI. This chapter also introduces deep learning and active learning methods applied to histopathology image analysis.

In Chapter 3, the materials used in the project are described with a particular focus on the dataset of whole-slide images from patients with a UC diagnosis. The dataset contains pixel-level annotations of regions of interest with neutrophils for certain biopsies.

Chapter 4 first includes the methods for WSI preprocessing and a description of the data partitions carried out. It also contains the development of segmentation algorithms to predict the regions of interest in histologic images and the design and implementation of an AL and Human-in-the-loop framework to queue the sample selection.

The results and discussion for the different experiments are presented in Chapter 5. This section includes the external validation metrics in the test set and the results regarding calculating and visualising the uncertainty for the active learning framework implementation, as a metric for sample selection for retraining.

Finally, Chapter 6 is composed of the findings regarding the project and some notes about limitations and future lines of investigation in this field.

Chapter 2

Teoretical Framework

2.1. Ulcerative colitis

2.1.1. Definition, epidemiology and management

Ulcerative colitis is a chronic inflammatory bowel disease (IBD) that primarily affects the colon (large intestine) and the rectum. It is characterized by inflammation and the presence of ulcers (open sores or lesions) in the lining of the colon and rectum. This condition is one of the two main types of IBD, with the other being Crohn's disease [5]. A condition affecting solely the furthest part of the colon and the rectum is labelled as ulcerative proctitis. When the condition extends from the descending colon downwards, it is termed left-sided colitis. Disease that encompasses the entirety of the colon is known as pancolitis (Figure 2.1).

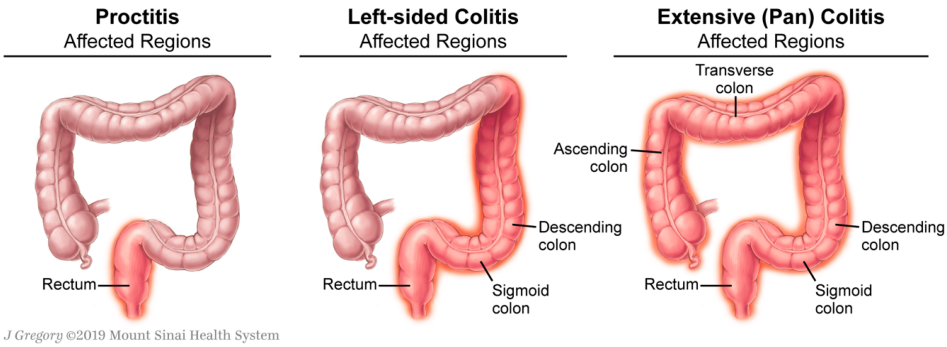


Figure 2.1: Ulcerative colitis disease extent [1].

The exact cause of UC is not fully understood but is believed to involve a combination of genetic, environmental, and immune system factors [5]. UC exhibits four distinct disease activity states: remission, mild, moderate, and severe [7]. The typical progression of UC involves intervals of remission alternating with episodes of acute exacerbation or disease flares, potentially necessitating an escalation of treatment, hospitalisation, and, in severe instances, surgical removal of the colon (colectomy). This is critical since a patient diagnosed with UC may not exhibit active inflammation at the endoscopic or histologic level. Therefore, it is imperative to distinguish between active UC and phases characterised by remitting inflammation or complete remission. The primary objective of treatment is to attain remission of the disease and avert complications such as infections,

the need for surgery, and the development of neoplastic growths, all while preserving the quality of life for patients [1].

The prevalence of UC is 24.3 per 100,000 person-years in Europe, while the rate in Asia and the Middle East is 6.3 per 100,000 person-years. In North America, the incidence is 19.2 per 100,000 person-years. These numbers reveal that the incidence and prevalence of IBD are rising globally and in various regions, underscoring its emergence as a worldwide health concern [6]. UC can occur at any stage of life, but it is often identified before age 30, and the condition seems to impact both men and women in equal proportions. Notably, around 20 per cent of individuals with UC have a close family member who also suffers from some form of inflammatory bowel disease (IBD) [5, 10]. The worldwide impact of ulcerative colitis is steadily increasing, leading to higher healthcare and societal expenditures. In the United States, it is estimated that the annual costs, encompassing both direct and indirect expenses associated with UC, range from $8.1 billion to $14.9 billion. In Europe, these costs are estimated to be between €12.5 billion and €29.1 billion [11]. These data show that the increasing disease incidence results in a rise in both direct and indirect costs. One potential solution to this issue is to harness the power of CAD systems, particularly emphasising using artificial intelligence (AI) methods that have demonstrated effectiveness in medical image processing. This approach can help manage the increasing incidence of disease and mitigate the associated rise in direct and indirect costs.

The diagnosis of ulcerative colitis cannot be definitively confirmed through a single diagnostic test. Instead, it is determined through a comprehensive evaluation of clinical symptoms, laboratory examinations, and assessments of endoscopic, histological, and radiological findings [12]. The management of UC patients has undergone significant advancements in recent years, mainly due to the development of innovative diagnostic techniques. Initially, management focused on inducing and sustaining clinical remission to prevent patient disability and colorectal cancer. Subsequently, endoscopy emerged as a pivotal protocol for disease assessment. Achieving endoscopy remission or mucosal healing became the gold standard for disease prognosis because it relies less on the highly subjective nature of clinical symptoms [13]. Recently, ulcerative colitis has included histologic remission as a treatment goal. In [8], it was found that UC patients in histological remission had a significantly lower relative risk of experiencing clinical relapse than patients with active histological inflammation.

Since management can be conducted at both the endoscopic and histologic levels, various tissue characteristics are affected as inflammatory activity increases. Among the endoscopic features, alterations in the vascular pattern and tissue bleeding are observed. Similarly, histological patterns seen in biopsies with active colitis include changes in crypt architecture, mucin depletion, and inflammatory infiltration by cells such as eosinophils and neutrophils [14]. In Figure 2.2, the connection between the endoscopic view and the histopathological examination is illustrated for both active (upper row) and remission (lower row) tissues.

Several factors, including the disease's severity, location, and progression, influence treatment decisions for ulcerative colitis. Proctitis, for instance, is typically managed with topical therapy involving 5-aminosalicylic acid (5-ASA) compounds. In cases of more extensive or severe disease, a combination of oral and local 5-ASA compounds, along with corticosteroids, may be employed to induce remission [12, 7]. Patients who do not respond to the initial treatment may necessitate hospitalisation. In such cases, intravenous steroids are commonly administered. If the condition remains unresponsive or

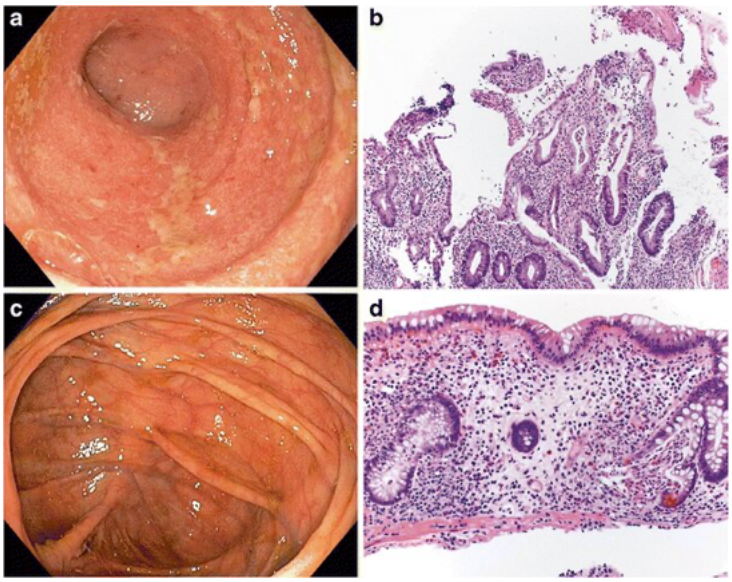


Figure 2.2: (a) Endoscopic appearance of severe ulcerative colitis. (b) Concordant biopsy with the endoscopic impression in *(a)*; severely active ulcerative colitis with erosion. (c) Endoscopic appearance of apparent normal mucosa in a patient with a history of ulcerative colitis. (d) A random biopsy from *(c)*: a histologic active disease with foci of neutrophilic cryptitis and crypt destruction [2].

refractory, additional therapeutic options are taken. These interventions are considered to address the more severe or treatment-resistant cases of ulcerative colitis. The most frequently conducted surgical procedure for patients with medically unresponsive UC, provided there are no complications such as perforation, is the restorative proctocolectomy (RPC) with ileal pouch-anal anastomosis (IPAA) [1].

2.1.2. Scoring on Ulcerative Colitis

In this context, we are introducing various indices for endoscopic and histologic assessment, with particular emphasis on a newly developed index that will serve as the central element in the deep learning algorithm.

- **Endoscopic scoring**: In UC, it is used to evaluate the extent and severity of inflammation in the colon and rectum. It helps clinicians determine the disease's activity level and guide treatment decisions. It may not capture microscopic inflammation, and inter-observer variability can influence scoring, where different endoscopists may assign slightly different scores to the same findings.

  - *Mayo score*: It is the most commonly used system. It grades disease activity on a scale from 0 to 3, with 0 indicating normal or inactive disease and 3 indicating severe disease with ulceration. This index was originally developed in 1987 and was initially employed in a clinical trial to assess the effectiveness of a novel therapy for UC [15].
  - *UCEIS*: The Ulcerative Colitis Endoscopic Index of Severity (UCEIS) was introduced in 2011, with a primary focus on assessing the endoscopic aspects of the disease, distinguishing it from the Mayo score. This scoring system considered as many as ten descriptors but ultimately selected three critical criteria to determine the final score. These criteria evaluate the vascular pattern, bleed-

ing, and erosions/ulcers, with grading based on the most severe lesion observed [16].

- **Histologic scoring**: It assesses the microscopic features of colonic tissue to determine the degree of inflammation, injury, and healing. It considers various histological features, including crypt architecture, mucin depletion, inflammatory infiltration and ulceration. It assigns a score based on the observed histological changes, allowing for a quantitative assessment of disease activity. Histological indices also present a particular component of subjectivity related to the scorer because specific patterns could have different interpretations, which may complicate its applicability to automatic CAD systems.

    - *Robarts Histologic Index* (RHI): The RHI assigns scores to features such as ulcerations, inflammatory infiltration, and neutrophils presence, with higher scores indicating more severe disease activity. The scores are then summed to provide an overall assessment of histologic disease activity. Each of these components of the score can be graded from 0 to 4 each [17].
    - *Nancy Histologic Index* (NHI): The Nancy Histologic Index assigns scores to histologic features, with higher scores indicating more severe disease activity. These histologic features include distortions of the crypt architecture, acute inflammatory cell presence, and ulceration, among others. These scores are then combined to provide an overall assessment of histologic disease activity. Unlike the RHI score, NHI does not grade these features between 0 and 4 [18].

It's important to note that both RHI and NHI take into account the presence of neutrophil infiltration within the mucosal tissue. Neutrophils are one of the three primary types of white blood cells, alongside basophils and eosinophils. Each cell type has its distinct nucleus and cytoplasmic characteristics, contributing to their unique appearances under microscopic examination.

This work primarily centres around a newly established histological scoring system for evaluating ulcerative colitis developed within the framework of the PICaSSO research initiative, in which the CVBLab participates. The PICaSSO Histologic Remission Index (PHRI) represents an innovative and simplified histologic index specifically designed to assess mucosal healing or disease activity in individuals diagnosed with UC [4].

PHRI assesses neutrophil infiltration in four distinct biopsy compartments: the lamina propria, surface epithelium, cryptal epithelium, and cryptal lumen. The lamina propria is the connective tissue that surrounds the crypts and typically contains varying numbers of plasma cells, neutrophils, and eosinophils, depending on the level of inflammation. The surface intestinal epithelium comprises a single layer of epithelial cells, forming a barrier between the tissue and the external environment. Lastly, the crypts are tube-like glands with a bordering cryptal epithelium that encircles the cryptal lumen. Visual examples for each of the four compartments from a WSI are presented in Figure 2.3.

The PHRI scoring system examines and analyses four distinct regions of interest within the biopsy sample. Within these four compartments, the degree of neutrophil infiltration, indicative of ulcerative colitis activity, is assessed. The PHRI grade for an individual biopsy is determined by adding the number of regions displaying significant histological findings, as outlined in Table 2.1.

This scoring system is notably recognised for its simplicity compared to other histological indices, such as NHI. PHRI assigns a grade to each biopsy on a five-point scale,

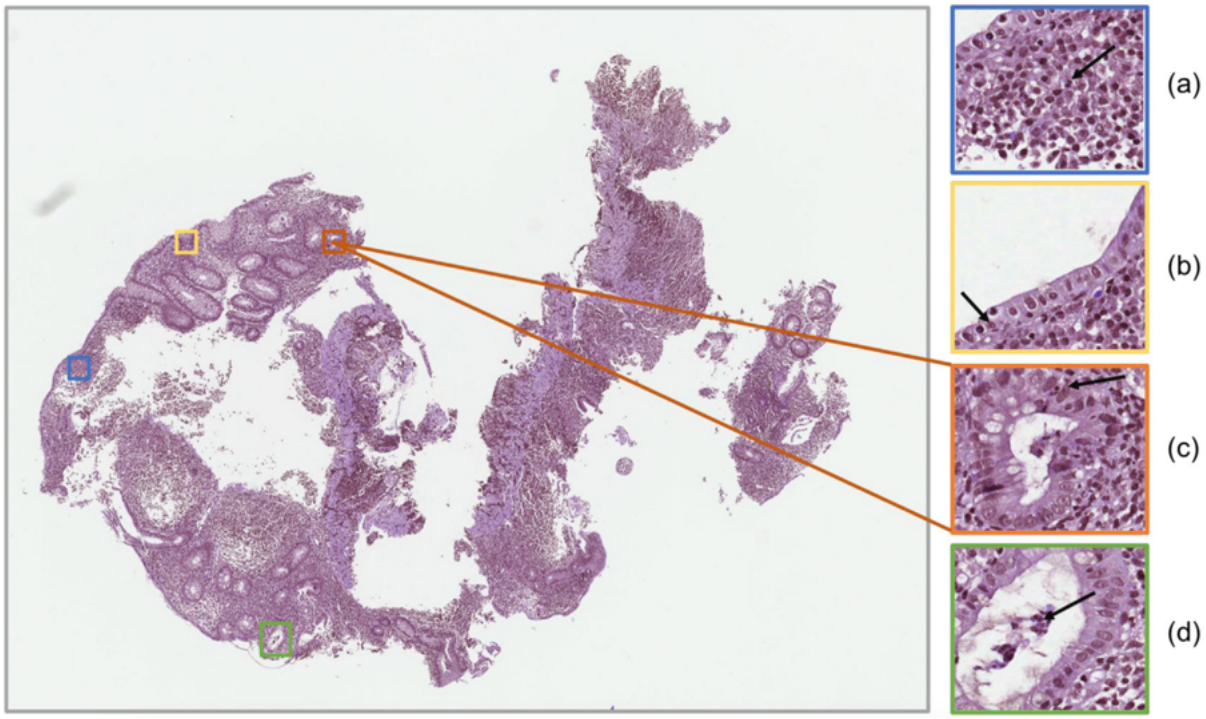


Figure 2.3: A Whole Slide Image of a biopsy with ulcerative colitis activity and its four structures of interest. The right column patches correspond to (a) lamina propria, (b) surface epithelium, (c) cryptal epithelium and (d) cryptal lumen. The black mark indicates the presence of a neutrophil [3].

ranging from 0 to 4. It's worth mentioning that the minimum score achievable on this index is zero, and is therefore considered to be in histological remission. In contrast, those with a PHRI score greater than zero are considered to have ongoing UC activity. The maximum score is four, indicating the presence of neutrophil infiltration in all compartments [4]. This threshold has effectively stratified patients into low and high-risk categories for adverse outcomes.

The primary objective in developing the PHRI was to create a straightforward grading system for ulcerative colitis to minimise the variability and subjectivity that can arise when pathologists examine tissue samples under a microscope. This scoring system has been aligned with endoscopic findings and has strongly correlated with clinical outcomes, such as RHI and NHI [4].

Table 2.1: PICaSSO Histologic Remission Index (PHRI) to predict histological remission [4].

Histologic finding	Score
Neutrophil infiltration in lamina propria	
Absent (No)	0
Present(Yes)	1
Neutrophil infiltration in epithelium	
Absent (No)	0
Present (Yes)	
- Surface epithelium	1
- Cryptal epithelium	1
- Crypt lumen	1
Total Score = sum of all above (maximum 4)	

Additionally, the inter-rater agreement among pathologists when using PHRI is excellent, surpassing the consensus achieved with other histological scoring systems. Simultaneously, the simplicity inherent in PHRI makes it well-suited for integration into artificial intelligence-based systems.

To classify the four areas of interest for PHRI calculation based on the presence or absence of neutrophils, as shown in Figure 2.3, experienced pathologists from the PICaSSO project performed pixel-level annotations indicating the four regions of interest. These annotations and the WSIs constitute the database with which the present work will be developed using deep learning and active learning techniques. With the workflow required by active learning, explained in the next section of the chapter, the aim is to improve the accuracy and performance of the segmentation algorithms as a preliminary step to a future calculation of the PHRI index.

2.1.3. CAD systems in Ulcerative Colitis

Computer-aided diagnosis has become a critical tool for clinicians, revolutionising medical decision-making. These systems, driven by advanced algorithms, artificial intelligence, and recent digital and medical technology advancements, substantially elevate diagnostic accuracy and efficiency [19]. They empower early disease detection and provide quantitative data for objective assessments. This synergy ultimately raises healthcare quality and augments patient outcomes. CAD's multidisciplinary approach encompasses multi-modal image processing, extensive data analysis, pattern recognition, and AI, enhancing its role in supporting clinical decision-making [20, 21]. Real-time clinical decision support, such as live video analysis during endoscopic procedures, exemplifies CAD's practical application. However, addressing model interpretability remains critical in implementing AI-based algorithms in medicine, as many deep learning methods function like black boxes, rendering their decision-making processes opaque [22].

As previously mentioned, the global rise in ulcerative colitis increases hospital workload. This trend underscores the growing importance of implementing CAD systems to support clinicians in managing this condition efficiently and effectively. Powered by advanced algorithms and artificial intelligence, CAD systems can help healthcare professionals diagnose and treat UC by providing valuable insights, enhancing diagnostic accuracy, and improving patient care [23]. Prior research in the field has achieved remarkable accuracy in tasks related to the classification of UC endoscopy images. These earlier AI systems primarily concentrated on classifying images and videos using established scoring systems such as Mayo and UCEIS. Remarkably, these AI systems exhibited results on par with those achieved by medical professionals, underlining the potential of AI in improving UC diagnosis and management [24].

Histopathology has received considerable attention in recent years, as it is the gold standard for cancer detection, which has an important clinical and societal impact. One of the most noteworthy advancements in this field is the digitalisation of histological tissue samples into WSIs. WSIs are high-resolution, gigapixel images of stained biopsy slides traditionally observed under microscopes in clinical settings. However, computational pathology has enabled the digitisation of these images. Although WSIs present a challenge for AI-based applications due to their immense complexity stemming from the vast number of pixels, they also offer the advantage of being analysable at various resolutions, depending on the desired magnification level [25].

While histological assessment plays a critical role in UC diagnosis and management, it's noteworthy that few studies have thus far concentrated on the analysis of whole-slide images to predict UC. In [26], the authors developed a deep-learning algorithm specifically designed to quantify the density of eosinophils within tissue samples from sigmoid colon biopsies. In this research, the assessment of disease severity was based on histological scoring systems, namely the Goebes and RHI scores. Authors in [3] used a Multiple Instance Learning (MIL) approach to predict UC activity (active or in remission) in WSI patches. In MIL tasks, the training dataset is structured into bags, with each bag comprising a collection of instances, and its goal is to teach a model to predict the bag label. They introduced a location constraint module that forces the feature framework to focus on the most significant patterns in the patches of each bag. However, they highlight the importance of improving neutrophil detection for a correct PHRI grading. Therefore, the primary goal of this work is to enhance the segmentation of the regions of interest where the presence or absence of neutrophils determines the area's activity. Improving this segmentation is deemed crucial, as it is expected to enhance the evaluation and grading of the PHRI index.

2.2. Active Learning

This section introduces the concept of active learning, a novel branch of deep learning focusing on the algorithms developed in this project.

2.2.1. Deep Learning

Deep learning is a subfield of artificial intelligence that focuses on training neural networks to perform complex tasks. Its success is attributed to its ability to handle large and diverse datasets, making it a powerful tool for solving complex problems and advancing AI research.

Deep learning has revolutionised various domains, including computer vision, which comprises object detection, semantic segmentation and image classification tasks. It is characterised by using deep neural networks with multiple layers and activation functions with trainable parameters, enabling the model to learn hierarchical features from data automatically. The training process in deep learning is marked by back-propagation, a technique that involves iteratively adjusting the model's weights based on the learning rate to minimise the network's error.

Ground truth is the variable we want to predict within the model. Deep learning-based algorithms can be classified into three types based on the availability of ground truth. They include supervised learning when ground truth is available for all the data, unsupervised learning when the ground truth is unknown, and semi-supervised knowledge when only a portion of the data contains ground truth annotations, while the remainder does not.

2.2.2. Definition of Active Learning

Active learning is a strategy employed in machine learning to minimise the quantity of labelled data required for a learning task. It involves selecting specific samples from an

unlabeled dataset, which are subsequently annotated by domain experts and integrated into the model's training process [27]. By utilising a well-designed sampling approach, AL can effectively reduce the overall volume of labelled data needed to train a model while enhancing its resilience to class imbalances. However, it's important to note that traditional AL methods do not inherently address the problem of noisy labels, which can still impact model performance.

In the field of computer vision, and in particular, in the field of semantic segmentation in the medical domain, medical image analysis is often confronted with the scarcity and cost of obtaining labelled data. In active learning frameworks, Human-in-the-loop is defined as a process in which users or specialists actively participate in the improvement of machine learning models by annotating data that is incorporated into the training set, gradually improving model performance while reducing annotation effort [28]. This approach is especially valuable for tasks requiring specialised knowledge, as it effectively combines human expertise with machine learning to train accurate models. Therefore, AL is invaluable for optimising the annotation process and improving the performance of segmentation models.

The application of AL for semantic segmentation in the medical domain follows a structured approach. Initially, an initial segmentation model is trained on a small labelled dataset as a starting point. AL algorithms then come into play, selecting regions or instances in unlabeled data where the model exhibits uncertainty or low confidence in its predictions [29]. In the context of semantic segmentation, this entails identifying image regions where the model struggles to differentiate between different classes or where boundaries are ambiguous. After selecting uncertain data samples, domain experts annotate these samples, and the resulting annotations are integrated into the training dataset (Human-in-the-loop). The model undergoes a fine-tuning process using this newly labelled data. This process becomes iterative; the model is re-evaluated on the labelled data, and uncertainty sampling is used again to pinpoint additional samples for annotation (Figure 2.4). This iterative process continues until the model reaches a satisfactory level of performance [30, 31, 32]. The iterative approach described accelerates model convergence, enhances segmentation accuracy, and is a valuable strategy for increasing the efficiency and effectiveness of medical image analysis systems.

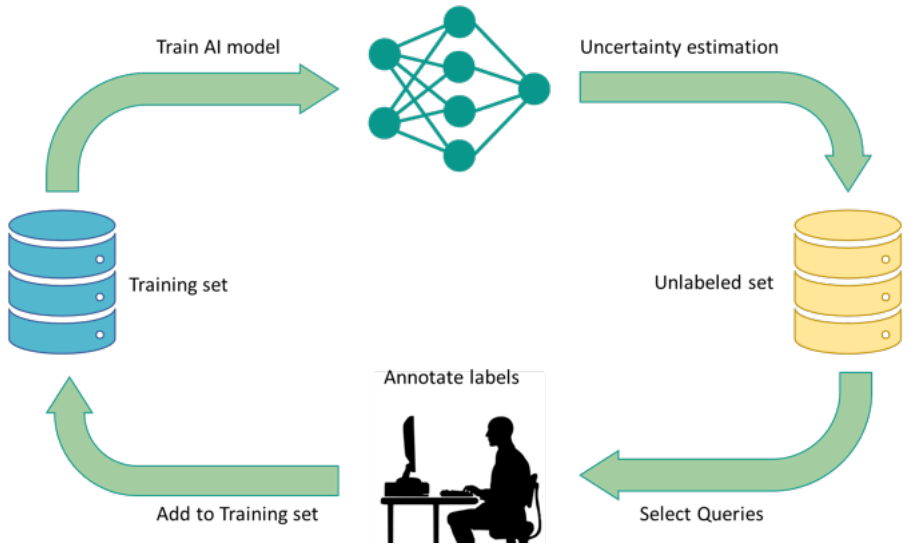


Figure 2.4: Active learning strategy in AI. Own elaboration.

2.2.3. AL in Histopathology

In recent years, Active Learning has been used in many applications of classification and segmentation of histopathology images [33]. Annotation of medical data is a time-consuming and labour-intensive task that requires significant dedication from skilled medical professionals. Active learning emerges as a valuable strategy in the medical field to alleviate this burden by reducing the need for extensive manual annotation efforts [34].

Most authors focus their research in histopathology images on finding the best strategy to extract the maximum information from the most informative unlabelled samples to be selected, queried for labels and then added to train the model further [29]. Several uncertainty measures have been researched to quantify a model's lack of confidence in its predictions for specific data points [35, 36].

If we focus on histopathology segmentation, several studies have been carried out with active learning approaches with good results, adjusting the number of annotations required. The authors from [37] developed a fully convolutional network (FCN) based segmentation model with an AL framework to select the most informative patches or regions from WSIs of each iteration. In this work, the authors generate a pixel-level uncertainty map of each area and then select the most informative unlabeled regions to annotate and append to the train set with satisfying model performance. In [38], the authors developed a patched-based segmentation assessment framework to evaluate the quality of segmentations in histopathology images with promising results. In [39], the authors propose an AL framework, which progressively integrates pixel-level annotations during training. It enabled a better CNN visualisation and interpretation of CNN predictions. As can be seen, many examples in the literature of AL frameworks applied to histopathology exist. Still, to the author's knowledge, there needs to be research focusing on using AL strategies for pixel-level segmentation of histological images of patients with ulcerative colitis.

Chapter 3

Materials

3.1. Database of Whole Slide Images

This section commences with a comprehensive database description, emphasising the acquisition protocol applied to capture the whole-slide images and pixel-level annotations. Additionally, it provides insight into the software and hardware infrastructure harnessed throughout the project, offering a global view of the project's data and computational resources.

3.1.1. Description of the database

The database of histological images employed in this project was sourced from a collaborative effort involving numerous medical centres worldwide, which conducted an international multicenter real-life prospective study on endoscopy and histology in UC [40]. During the respective endoscopic procedures, over 600 digitised biopsy samples were collected from patients diagnosed with UC.

These slides were initially utilised to validate the PHRI, a novel scoring system for assessing histological remission in UC. The clinical protocol involved the extraction of two biopsies from each patient, obtained from distinct sections of the colon (one from the sigmoid and another from the rectum) during the endoscopic procedure. It's worth noting that biopsies from the same patient were analysed independently, as active inflammation was not expected to be present in both samples simultaneously.

Following the surgical removal of colon tissue, samples must undergo a series of histologic procedures before being examined under a microscope by pathologists. These procedures encompass tissue fixation, which preserves cell morphology; tissue embedding to provide consistency; sectioning the tissue into thin slices using a microtome; and staining these slices, often with hematoxylin and eosin (HE) stain. Subsequently, the HE-stained tissue samples are transformed into digital WSIs using specialised scanning machines, a technology that has become prominent with the advancements in digital pathology.

Whole Slide Images are high-resolution, gigapixel-sized images obtained by digitising tissue samples, posing a unique challenge for deep learning systems due to their memory and computational resource demands. While all the biopsies in this study underwent a similar process of extraction and digitisation, a multi-centre database provides the opportunity to validate the model using images from various hospitals. This diversity

allows for assessing the model's performance across images exhibiting variations in staining colours or biopsy orientation, reflecting real-world clinical scenarios and enhancing the model's robustness and generalizability.

As previously highlighted in Table 2.1, PHRI is a histological remission index used to assess ulcerative colitis. Its primary focus is identifying neutrophils within four distinct tissue section regions: the lamina propria, cryptal epithelium, cryptal lumen, and surface epithelium. Pathologists meticulously analyse these compartments to score a single biopsy and determine the overall score by summing the number of regions exhibiting significant neutrophil presence. This scoring approach comprehensively assesses the biopsy's histological characteristics of ulcerative colitis activity.

In Figure 3.1, a large WSI exhibiting a high level of ulcerative colitis activity is showcased. The regions marked in the four colours are, in descending order, lamina propria (blue), cryptal epithelium (green), surface epithelium (cyan), and cryptal lumen (yellow), providing a comprehensive view of different areas of interest. All WSIs have been annotated with their four corresponding pixel-level regions, which serve as the ground truth (GT) for both model training and evaluation purposes.

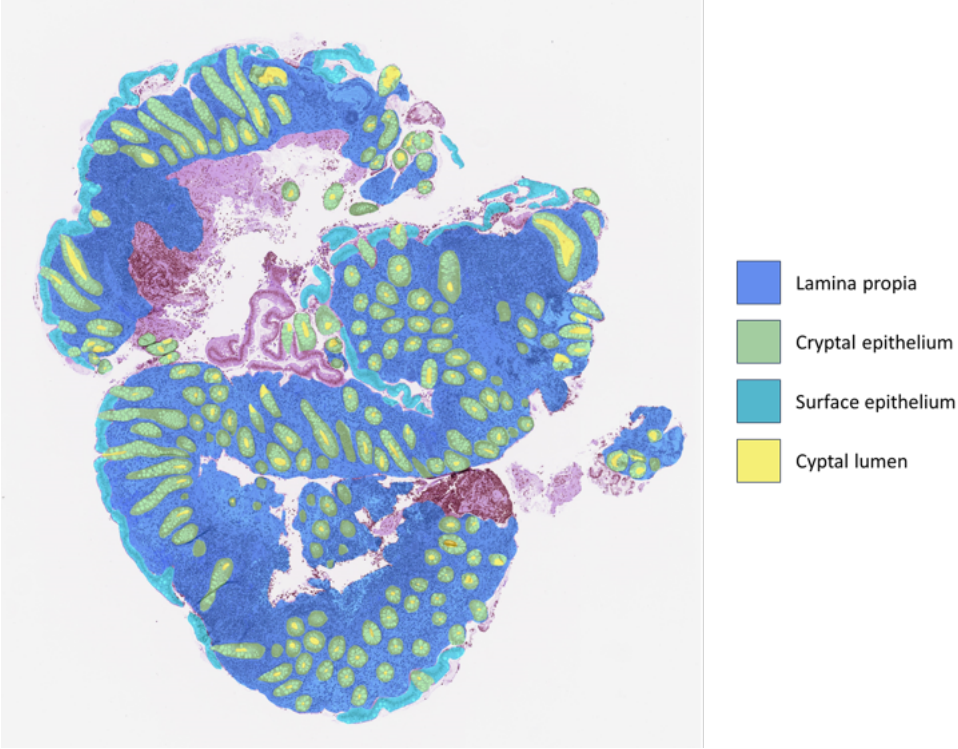


Figure 3.1: A Whole Slide Image of a biopsy with ulcerative colitis activity and its four structures of interest. In the right-hand column, each region is labelled according to the colour of the annotation. Own elaboration.

3.1.2. WSI annotations

Detailed image annotations hold immense significance in developing supervised deep-learning algorithms. However, obtaining such annotations in large datasets can be laborious and sometimes impractical. This challenge is particularly pronounced in medical imaging, where data availability is often constrained, and in digital histopathology, where the complexity of images can be exceedingly high.

The pathologists collaborating in the PICaSSO project were tasked with performing pixel-level annotations on only a small subset of WSIs from the entire database as a basis for addressing this challenge. In the beginning, they annotated 52 WSIs of the whole dataset. The aim of requesting these specific annotations was to have a first set of images to build the first segmentation algorithm for the regions of interest and to perform ablation experiments to optimise the model's hyperparameters. This first set of images and annotations, subdivided into train and test sets, form the basis from which the workflow of the active learning methodology used has been developed and which is explained in later sections.

These annotations used various colours to delineate the biopsy's four specific compartments of interest. This annotation process was facilitated by employing custom in-house software known as *PIXNORMOUS*, which had been tailored by engineers from the CVBLab to cater specifically to WSI annotations. This software gave the annotators a high degree of flexibility, enhancing their ability to carry out precise and detailed annotations. Finally, some annotation examples are presented in Figure 3.1.

3.2. Software

The deep learning algorithm was implemented using the *Python 3.7* programming environment. *Python* is a versatile, high-level, interpreted, interactive, and object-oriented programming language renowned for its wide-ranging applications. *Python*'s strength lies in its ability to seamlessly integrate modules and packages tailored for addressing specific tasks, making it an ideal choice for developing and executing complex machine learning algorithms.

The most used package for this work is *PyTorch*, a state-of-the-art open-source machine learning framework developed by *Facebook's AI Research lab (FAIR)*, which is playing a pivotal role in the field of artificial intelligence. Renowned for its dynamic computational graph construction and pythonic programming interface, *PyTorch* offers a versatile platform for scientists and researchers. Its dynamic nature enables on-the-fly model architecture modifications, facilitating experimentation and innovation. *PyTorch* is especially appealing for deep learning, thanks to its automatic differentiation capabilities, simplifying the complex task of gradient computation. With seamless GPU support and a thriving ecosystem of libraries, *PyTorch* empowers scientists to develop, train, and deploy sophisticated machine learning models, making it a fundamental tool in AI scientific research.

The training phase of the deep learning process, known for its significant computational demands, was executed on a remote server boasting high computational capabilities. To establish a seamless connection between the local working computer and the remote server, *VisualStudio Code* (Version 1.80.2) software was employed. This software enables the creation of an SSH (Secure SHell) connection. This secure and efficient communication protocol bridges the gap between the two systems, facilitating the execution of deep model training on the powerful server.

Another essential software utilised in this project is *MATLAB* (Version R2019b), a versatile application developed by *MathWorks*. *MATLAB* encompasses a multi-paradigm programming language and a robust numeric computing environment, making it an invaluable tool for tackling scientific and engineering challenges. In this project, *MATLAB* was harnessed for image processing tasks, underlining its utility in handling and analysing visual data.

3.3. Hardware

The project's core operations were executed on a computer equipped with an Intel Core i7-7700HQ processor, operating on a 64-bit Windows 10 system. This computer boasted 8GB of RAM, a 1TB external storage hard drive disk (HDD), and an NVIDIA GTX1050-4GB graphics card. However, a more robust resource for the intensive deep model training, the NVIDIA DGX A100 (now DGX), housed in the CVBLab, was employed. The DGX is one of the most potent hardware assets for artificial intelligence development, characterised by unparalleled computational density and performance, thanks to its 8 NVIDIA A100 GPUs, each equipped with 640GB of GPU memory.

Given the substantial size of WSI, ranging from 500MB to 2GB, a Synology DS918 NAS server with a substantial storage capacity of 32TB was utilized. The project workflow involved storing data and generating code on the local computer and then transferring it to the NAS server. An SSH connection was established between the NAS and the DGX system, enabling the execution of various scripts for training and inference using the deep models. This architecture allowed for the efficient handling of large WSI datasets and resource-intensive deep-learning tasks.

Chapter 4

Methodology

4.1. Overall methodological framework

To introduce this section, Figure 4.1 provides a flowchart encompassing all the methodology processes associated with developing algorithms for predicting ulcerative colitis in whole-slide images. Subsequent units will delve into the specific details of the methodology as outlined in the flowchart.

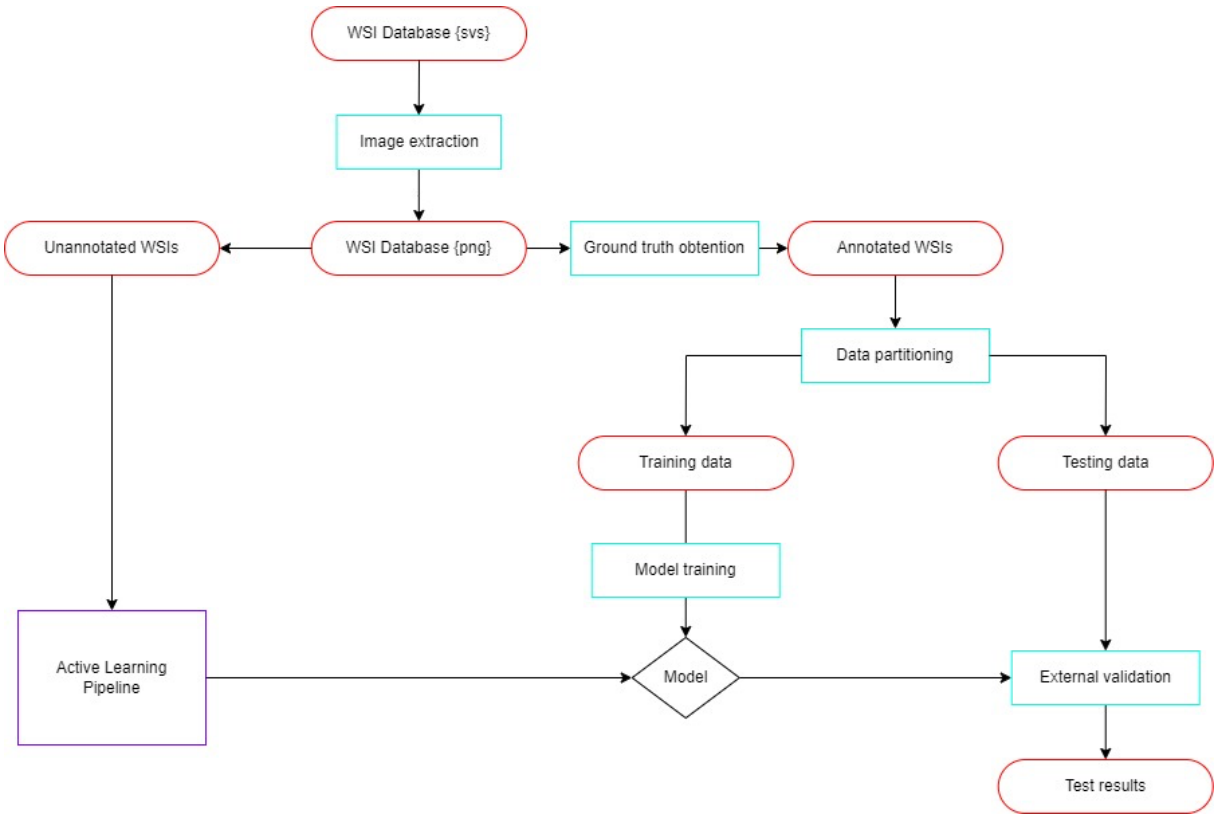


Figure 4.1: Flow chart of the proposed methodology. Own elaboration.

As explained in the previous section, the database to develop the project is a substantial dataset comprised of whole-slide images. WSI represents high-resolution, gigapixel images achieved by digitalising biopsy samples using specialised scanners. These WSI files are typically stored in .svs format, which requires specialised software like *Amperio Imagescope* for visualisation. To make these images accessible through standard software,

they must be converted into .png files. Some of these files come equipped with pixel-level annotations created by pathologists. These annotations are crucial as they dictate which images will be utilised for training and validating the machine learning models.

Afterwards, the model development includes all the programming stages for creating the deep-learning-based algorithm for segmenting the four regions of interest in UC histologic images. This algorithm based on encoder-decoder frameworks is trained using the selected biopsy images and their annotations.

Ultimately, our objective was to assess the efficacy and resilience of our model across a comprehensive database. To achieve this, we employed a collection of WSI pixel-level annotations for the algorithm's external validation. The evaluation of the test results primarily focuses on performance metrics commonly employed in segmentation problems and the efficiency of regions of interest segmentation.

Once this first model was evaluated, we used a batch of unlabeled WSIs to develop the Active Learning framework to enhance the model performance, which will be explained in subsequent sections.

4.2. WSI preprocessing

Whole-slide images represent digital biopsies of tissue sections, often containing redundant information due to multiple slices from the microtome being placed within the same crystal for subsequent digitisation using specialised scanners. Given the computational constraints associated with pixel-level segmentation in histopathology, selecting a representative slice for processing by deep models becomes essential, thus reducing image sizes.

Nevertheless, additional preprocessing steps, including image cropping and background elimination, are indispensable for mitigating computational costs and addressing data noise. In the database, all WSIs are downsampled to a 20x resolution, with the choice of downsampling magnification determining the final image size.

Typically, WSIs exhibit regions devoid of tissue surrounding the main content. These background pixels lack relevant information about the biopsy and should ideally be removed. To achieve this, the Otsu thresholding method, applied to the red channel of the RGB image, is employed to separate tissue from the background effectively.

4.3. Data partitioning

As mentioned earlier, the acquisition protocol involved obtaining two biopsies per patient. In light of this, patient-level partitions were implemented to construct the two essential datasets for machine learning: training and test sets. In patient-level partitioning, whole slide images of the same patient are grouped together in the same dataset, although they are analysed independently. The rationale behind this partitioning technique is to ensure that all images from a given patient reside within a single dataset. This approach avoids including images from the same patient in different sets, which could potentially introduce bias or inaccurately represent the model's capabilities due to recurring patterns or variations specific to individual patients. However, this issue only arises when dealing with fewer patients with only one biopsy, as multiple images from the same

patient are unavailable.

The division of the data into the subsets was based on the presence of pixel-level annotations in specific whole-slide images. Only WSIs that had associated pixel-level annotations to identify the four regions of interest were utilised for training the model. Out of 529 PICaSSO project WSIs, ground truth with the four regions of interest was obtained in only 52 WSIs (9,83%), manually annotated by the medical specialists for training and initial model evaluation. As explained in previous sections, this annotation task is exhaustive and time-consuming for pathologists, so the initial dataset available for the development of the segmentation model is limited. From the whole database, fifteen were selected as a WSI database without annotation for further use in the active learning framework to improve the performance of the segmentation model. The distribution of the data is shown in Table 4.1.

Table 4.1: Description of the WSI database

Database	Number of images
PICaSSO database	529
Annotated WSIs	52
Unannotated WSIs	15
Project Database	**67**

From the annotated WSI database, 33 images were utilised to develop the model and the remaining 19 histological images were designated as the test set to validate the model's performance externally (Table 4.2). Notably, this test set encompasses WSIs from medical centres that must be represented in the training data. This diversity allows for evaluating the model's robustness in a multicenter dataset where factors like colour staining may exhibit variations.

Table 4.2: Description of the database with pixel-level annotations.

Dataset	Number of centres	Number of patients	Number of images
Train	6	15	33
Test	5	6	19
Total	**7**	**21**	**52**

4.4. AL Framework for Segmentation of Regions of Interest in WSIs

An active learning approach aims to improve the first static segmentation model trained and evaluated with the first set of 52 WSIs annotated by the pathologists. Thirty-three of these images constitute the initial training set with which the static model is built, while the remaining 19 are kept out of the training as a test set for the whole process.

Once the first segmentation model has been built, the aim is to obtain the prediction of new images from a set of 15 unannotated images. Based on these predictions, the uncertainty of each of them will be estimated as a measure of the information they can contribute to the next model. From this set, the most informative samples will be selected, and pathologists will be required to correct the annotations predicted by the

model (human-in-the-loop process). Finally, the dynamic or iterable segmentation model will be retrained, and its performance will be evaluated with the test set. Figure 4.2 shows the framework for segmentation with Active Learning and Human-in-the-loop described.

Human-in-the-loop consists of the iterative interaction with the model outputs to guide the models to an optimum. In semantic segmentation problems, it can be implemented by guiding the authors to correct the output prediction of the models. Annotation time is considerably reduced as the annotators do not start the annotation from scratch.

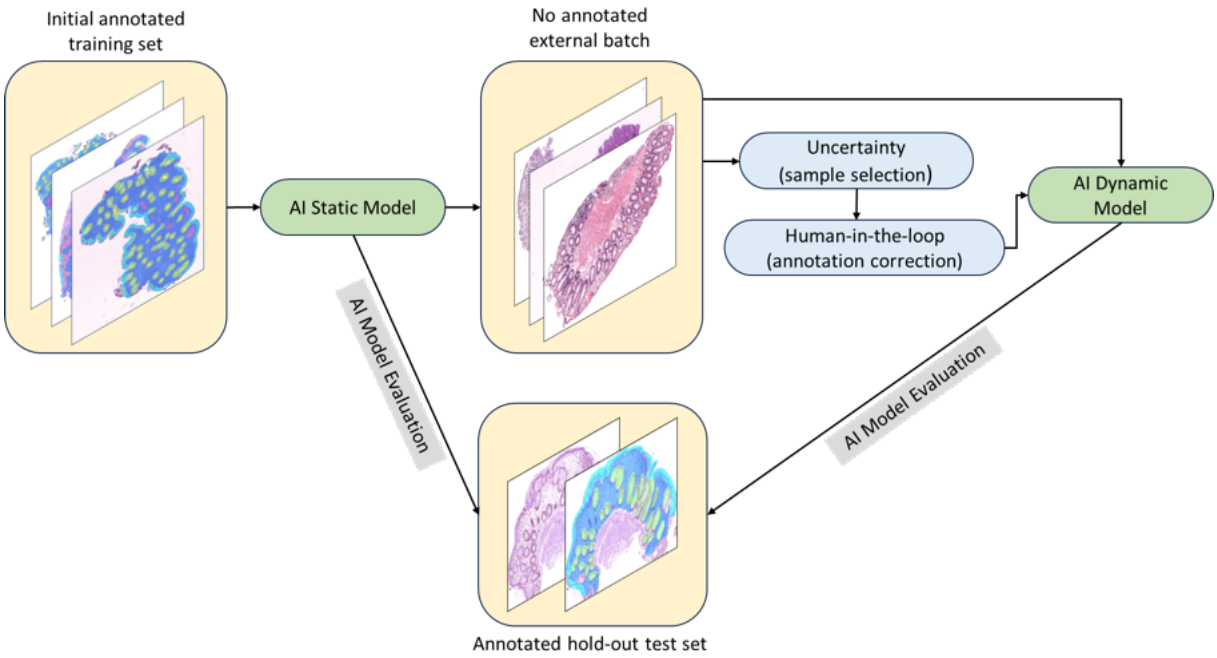


Figure 4.2: Framework for segmentation with Active Learning and Human-in-the-loop. Own elaboration.

4.4.1. U-Net for region segmentation

The proposed method consists of an end-to-end algorithm based on the implementation of U-net architecture for image segmentation from Ronnenberg et al. (2015), to perform whole-slide image segmentation [41]. The U-net formulation allows the detection of the different regions in WSI from patients with an ulcerative colitis diagnosis.

The graphical abstract of the U-Net is presented in Figure 4.3, and it is comprised of the encoder followed by the decoder. The encoder is a series of convolutional layers that gradually reduce spatial dimensions while increasing the number of feature channels. It extracts hierarchical features from the input image. The decoder is a symmetrical structure that upsamples the feature maps back to the original image resolution. It uses transposed convolutional layers to achieve this and gradually reduces the number of feature channels. One of the distinctive features of the U-Net is the skip connections (grey arrows in Figure 4.3) that directly connect corresponding layers between the encoder and decoder. These connections allow the network to preserve fine-grained details during upsampling. *ReLU* (Rectified Linear Unit) activations are used within the convolutional layers. In the final layer, a *Softmax* activation is used to obtain the most probable class out of the five in the images (four regions of interest and the background). The final layer is a convolutional layer that produces the segmentation mask with five classes.

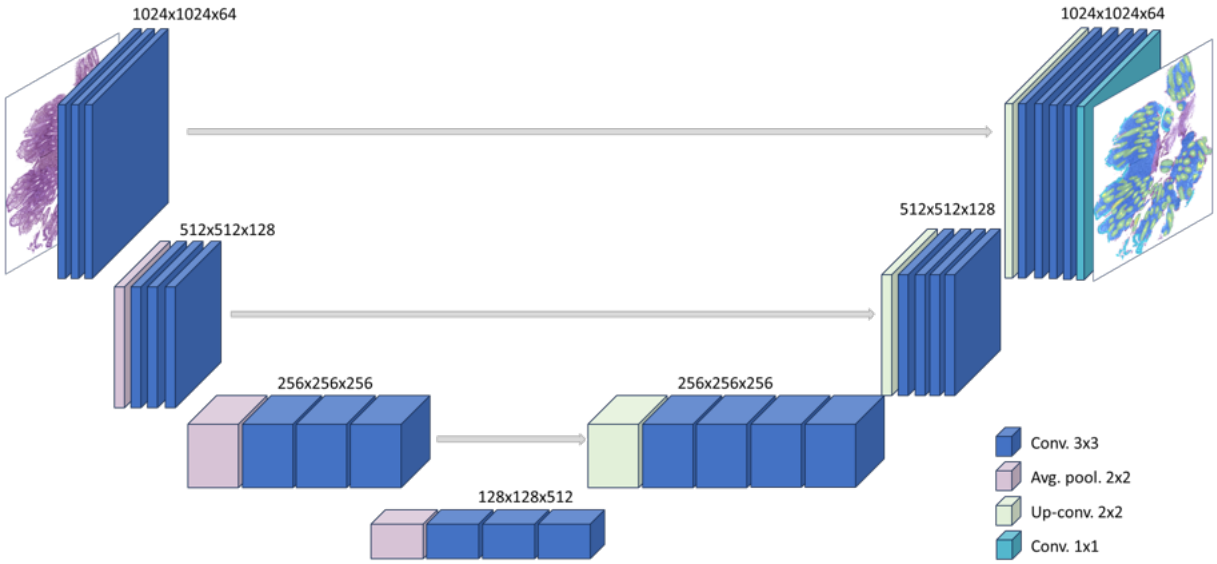


Figure 4.3: Graphical abstract of the proposed U-Net for WSI segmentation. Own elaboration.

4.4.2. Uncertainty calculation

Active learning involves selecting the most appropriate samples to retrain the algorithm optimally. For that purpose, we selected a subset of unannotated samples to evaluate the segmentation performance of the model. As a segmentation model, we trained a U-Net with residual connections between the encoder and decoder. Only a qualitative evaluation of the prediction can be made as the ground truth is unavailable for metrics calculation. We argue that selecting the most discriminative samples to retrain the algorithm will improve the model's performance.

In this context, uncertainty has been proposed in the literature as a measure of the informativeness of the samples [29]. One can assert that when a prediction carries higher uncertainty, a more significant opportunity exists to enhance our knowledge by incorporating the actual ground truth of that particular sample into the training set. This work also introduced three different measures for uncertainty estimation: least confident sampling, margin sampling, and Shannon Entropy:

- ***Least Confident Sampling***: It is calculated as one minus the maximum predicted probability. In other words, if the model assigns a high probability (close to 1) to one class and a low probability (close to 0) to all other classes, the uncertainty measure will be close to 0:

$$x_{LC} = \arg\max_x 1 - P_\theta(\hat{y}|x) \tag{4.1}$$

  where $\hat{y} = \arg\max_y P_\theta(y|x)$ and $P_\theta(y|x)$ is the estimated probability by the model that data instance x belongs to the correct class y. Instances with the highest uncertainty (lowest confidence) scores are selected for annotation. One limitation of least confident sampling is that it solely considers the information regarding the most likely label while disregarding information about the distribution of the remaining labels.

- ***Margin Sampling***: The difference between the highest predicted class probability and the second-highest predicted class probability is computed. A small margin

suggests that the model is uncertain about the correct class, while a large margin indicates higher confidence in the prediction:

$$x_M = \arg\min_x P_\theta(\hat{y_1}|x) - P_\theta(\hat{y_2}|x) \tag{4.2}$$

where $\hat{y_1}$ and $\hat{y_2}$ are the first and second most probable labels. This sampling method also does not consider the probabilities of all classes.

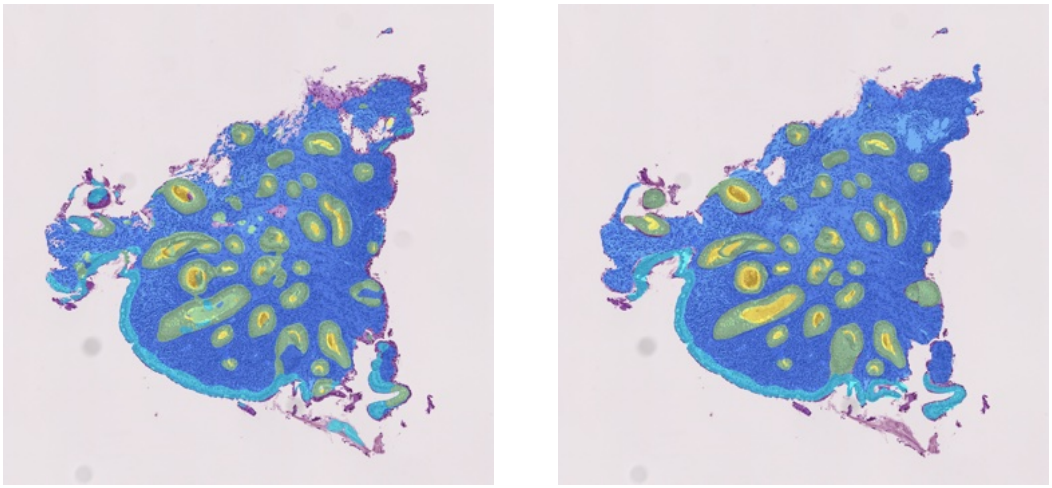
- ***Shanon Entropy***: Entropy serves as a metric to quantify the information needed for encoding a distribution. Consequently, it is commonly regarded as a means to assess uncertainty within machine learning tasks:

$$x_E = \arg\max_x -\sum_i P(y_i|x) \log P(\hat{y_i}|x) \tag{4.3}$$

where y_i covers all possible annotations. Entropy's versatility in extending to probabilistic multiclass annotations and accommodating more intricate structured data points has solidified its position as the preferred option for uncertainty-based query strategies.

4.4.3. Human-in-the-loop integration

Once the uncertainties of the images have been calculated, those with the highest level of information are obtained and selected for correction by the pathologists. For the calculation of the uncertainty metrics, the model obtains the prediction of the pixel-level classes corresponding to the regions of interest. This prediction constitutes the base annotation on which the pathologists had to correct those areas poorly predicted by the model within the so-called Human-in-the-loop process. For this purpose, the images selected with the highest uncertainty with their respective predicted annotation were uploaded to the *PIXNORMOUS* annotation software. Once the image and its segmentation had been uploaded, the pathologist modified those regions that had not been well predicted. Figure 4.4 shows an example of the Human-in-the-loop integration in the AL framework.



(a) Predicted segmentation (b) Corrected segmentation

Figure 4.4: Comparison between predicted and corrected segmentation after the Human-in-the-loop process. Own elaboration.

Firstly, Figure 4.4a reveals the predicted segmentation by the model. Once uploaded to *PIXNORMOUS* software, with this base annotation, the specialist corrects the wrong annotated regions according to its knowledge about the different tissue areas. Then, the corrected segmentation in Figure 4.4b is obtained, and prepared to be added in the next iteration of model retraining. In this way, the complete segmentations of the images with the highest uncertainty are obtained to provide the model with more information in the subsequent retraining process.

This protocol, which starts from the segmentation predicted by the model in that iteration, avoids pathologists having to perform the segmentation of the new images from scratch, having to correct only the regions where the network has performed worse. This significantly reduces the workflow and time spent by them on image annotation, mitigating the effort required for the improvement of the segmentation algorithm.

Chapter 5

Experiments and results

5.1. Evaluation metrics

To facilitate the comparison of various methods and the presentation of results in this chapter, we have chosen performance metrics commonly employed in segmentation problems. After obtaining the model's predictions and having access to the ground truth (GT) for each whole-slide image, the metrics we used for evaluating the segmentation performance of the different models are:

- **Categorical Dice** evaluates the similarity between two sets of segmented pixels: one produced by a segmentation model and another serving as the ground truth reference. The Categorical Dice metric is primarily used in pixel classification problems with discrete categories:

$$Dice = \frac{2 \cdot |X \cap Y|}{|X| + |Y|} \tag{5.1}$$

  where X represents the set of pixels segmented by the model (the prediction), Y represents the set of pixels from the ground truth, and $\cap$ represents the intersection of sets, i.e., the number of pixels that overlap between the prediction and the ground truth. The Dice coefficient varies between 0 and 1, where 0 indicates no overlap between the prediction and the ground truth, and 1 indicates a perfect match between the two sets of pixels.

- **Intersection over Union (IoU)**, also known as the Jaccard Index, is a common metric used in image segmentation tasks to evaluate the overlap between the predicted region (from a model) and the ground truth region. The following formula expresses the IoU:

$$IoU = \frac{|X \cap Y|}{|X \cup Y|} \tag{5.2}$$

  where $\cup$ represents the union of sets, the total number of pixels in either the prediction or the ground truth (or both). The IoU metric ranges from 0 to 1, where 0 indicates no overlap between the prediction and the ground truth, and 1 indicates a perfect match between the two regions. A higher IoU score indicates better agreement between the model's prediction and the ground truth.

5.2. Results and discussion

After training and validating the proposed methodology explained in section 4 in a subset of pixel-level annotated WSI, we retrained the model with the complete set of annotated images from the training and validation subsets, comprised of 33 WSIs. Once the final model was obtained, we performed an extensive external validation of the algorithm on a hold-out test set of 19 images. In this context, the Dice and IoU metrics were calculated for the initial model in the test set and were found to be 0.622 and 0.386 respectively.

Using the proposed framework, 15 new unlabelled images were selected from the WSIs database to undertake the iterative process by Active Learning. First, the uncertainty of each image was calculated according to the three methods explained in the section 4.4.2: least confident sampling, margin sampling and Shanon entropy. For the calculation of these metrics, the average uncertainty of each image was computed, considering only the probabilities of the pixels of the regions of interest, to normalise the calculation. Table 5.1 shows the uncertainty metrics of the 15 ordered from highest to lowest according to Shanon entropy.

Table 5.1: Uncertainty metrics for the complete AL batch. LC: Least confident sampling. M: Margin sampling. SE: Shanon Entropy.

Image name	LC	M	SE
01014 Rectum E_001	0.280	0.506	0.680
01014 Rectum E_002	0.260	0.534	0.625
03-06_Rectum_region_0	0.224	0.624	0.610
03-06_Rectum_region_3	0.210	0.645	0.578
12-03 Rectum_001	0.176	0.689	0.452
12-07 Rektum_002	0.164	0.716	0.435
04-06_RECTUM_002	0.161	0.722	0.430
04-12_SIGMOID_001	0.148	0.735	0.398
12-07 Rektum_001	0.144	0.746	0.386
04-12_SIGMOID_002	0.138	0.756	0.380
04-06_RECTUM_001	0.141	0.746	0.373
04-03_RECTUM_002	0.126	0.781	0.337
04-03_RECTUM_001	0.119	0.789	0.335
01010 Sigmoid D_001	0.121	0.780	0.325
12-05 Sigmoid_001	0.104	0.810	0.273

The sample with the highest uncertainty is shown below with its associated uncertainty maps according to the least confident sampling, Margin sampling and Shannon entropy methods, calculated at the pixel level (Figure 5.1). The uncertainty maps resolution corresponds to the output size of the image in the model (1024x1024), and therefore, their shape change compared with the raw WSI. It is worth noting that the margin sampling map was inverted in the form $1 - x_M$ because, as explained in section 4.4.2, in the specific case of margin sampling, the lowest value of the metric implies a higher uncertainty.

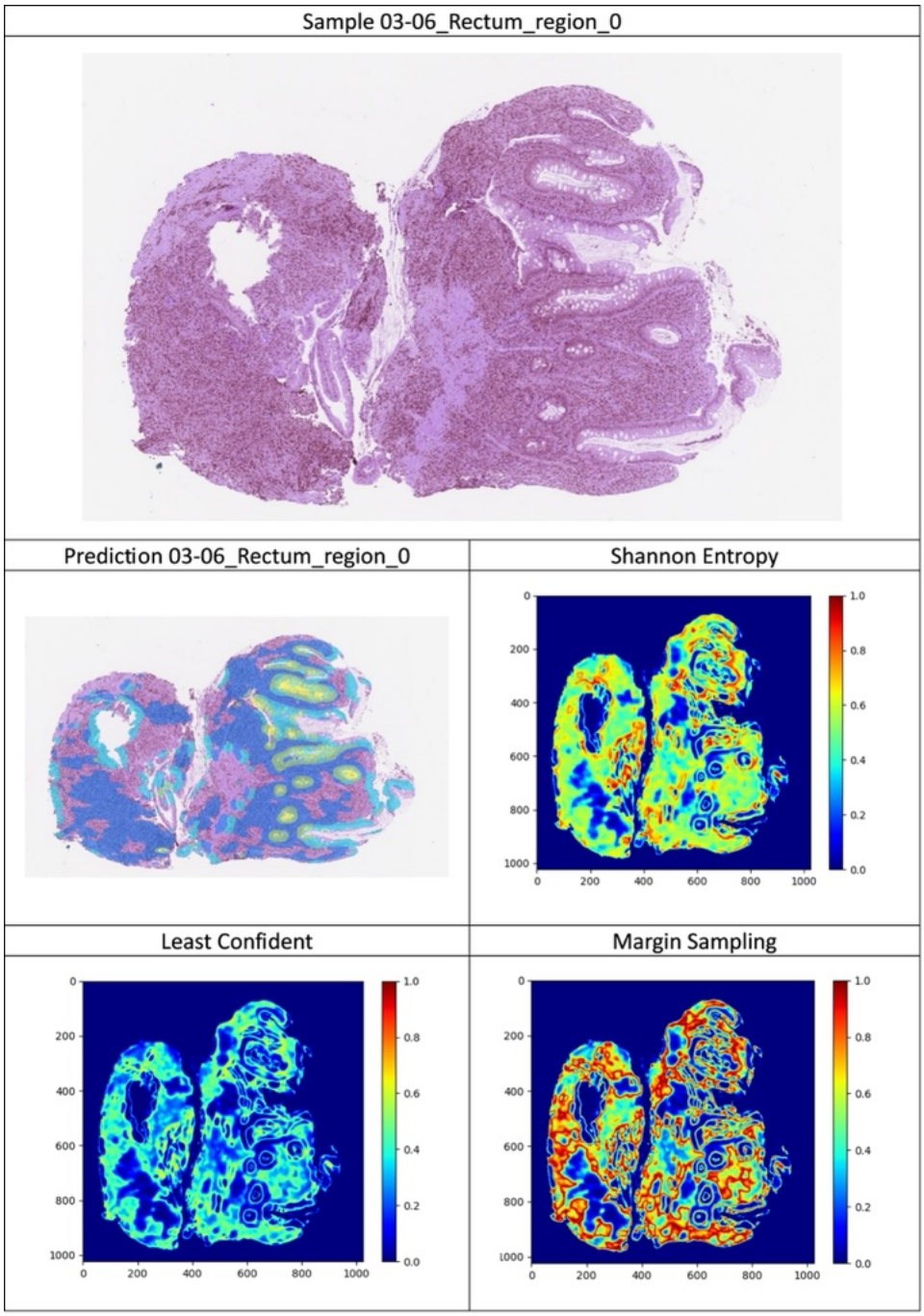


Figure 5.1: WSI with the highest level of uncertainty. Own elaboration.

Looking at the resulting uncertainty maps qualitatively and the prediction of the sample shown in Figure 5.1, the model seems to have difficulties distinguishing the region called lamina propia from other types of tissue that have no label. This could be because, as shown in Figure 3.1, the pathologists did not annotate all tissue zones, but only annotated the regions of interest for neutrophil detection (lamina propria, cryptal epithelium, cryptal lumen and surface epithelium) and segments of tissue were left unannotated, corresponding to the 'non-interest region'. As a consequence, the model confuses and predicts in uncertain samples such as the one in Figure 5.1 the non-interest region as lamina propria, given that in colour stain and cell structure they are very similar.

In contrast, Figure 5.2 shows the sample with the lowest level of uncertainty according to the Shanon entropy metric. On a qualitative level, it can be observed that most of the pixel uncertainty values are directly 0 or very close to that minimum uncertainty value. A higher level of uncertainty seems to be found at the edges of cryptal epithelium and cryptal lumen structures.

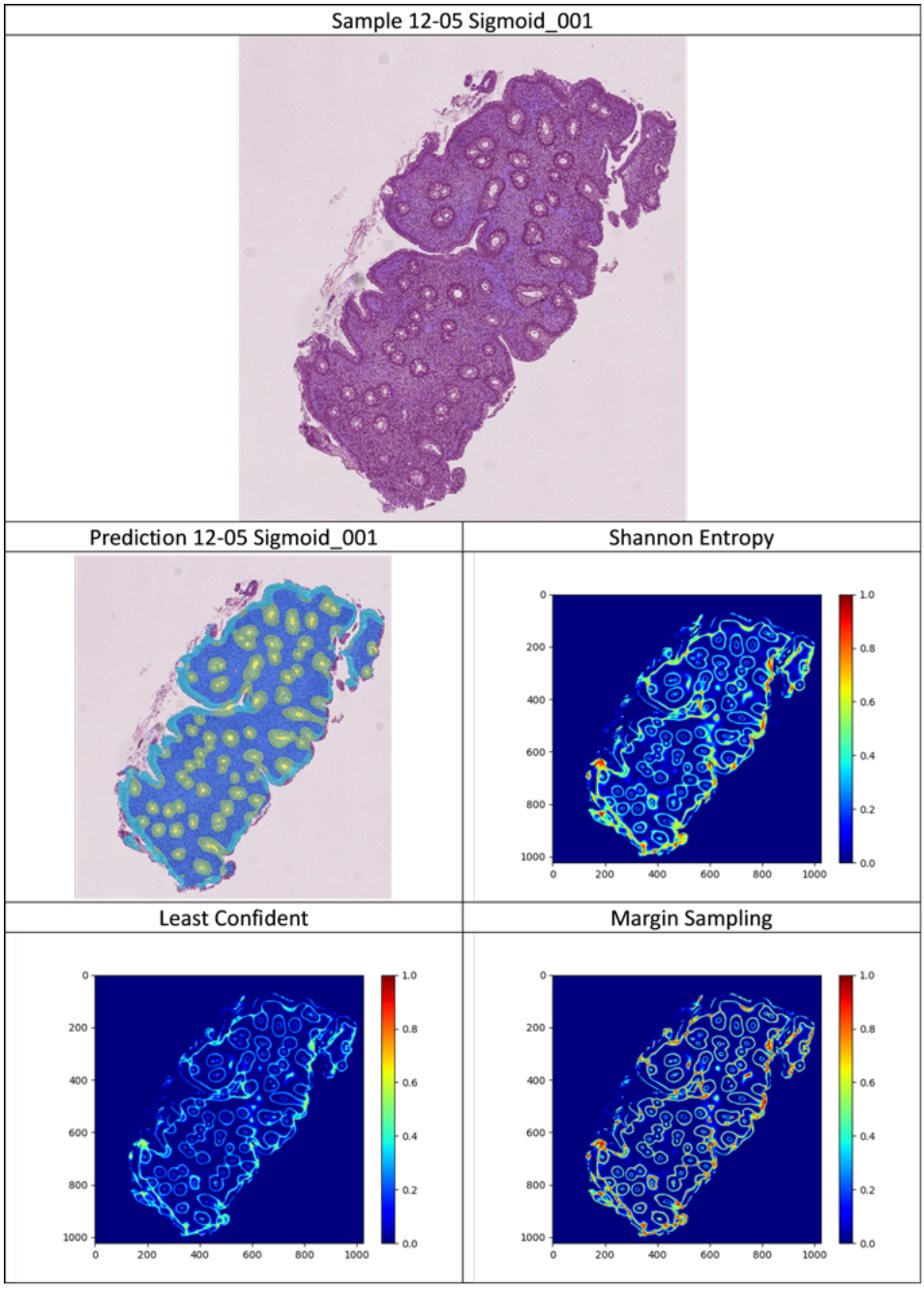


Figure 5.2: WSI with the lowest level of uncertainty. Own elaboration.

Considering the average of the calculated metrics per image and the formulation of each of them, we argue that the most appropriate selection criterion for this segmentation problem would be the Shanon entropy. The reason for choosing this metric was that, as explained in section 4.4.2, the Shanon entropy considers all classes' predictions. In contrast, the others only consider one or two probabilities. It is, therefore, the metric that provides the most information about the distribution of probabilities.

Afterwards, the model was re-trained first with the five samples with the highest uncertainty to determine the number of samples needed to find a significant improvement in network performance. For that purpose, with the preliminary model, we extracted the prediction of the five most uncertain samples from the unlabeled batch, and we asked the specialists to correct the annotations of these new five images (Human-in-the-loop). With this approach, the annotation time is substantially reduced as the annotators do not start the annotation from the beginning. The model was retrained jointly with the training set and the five new annotated images (AL w/ 5) in the next iteration step. Finally, we perform the external evaluation with the annotated hold-out test set to extract the test results of this iteration. As shown in Table 5.2, the new model trained by adding the five images with the highest uncertainty improves the Dice metric concerning the preliminary model, while the IoU metric decreases slightly.

Based on this result, there was no significant improvement, so the pathologists were asked to correct five more images to add the ten total images to the initial training set and retrain the model. Thus, a new iteration was performed within the framework of Active Learning and Human-in-the-loop, evaluating the performance of the retrained model with the test images at the end of the process. As a result, as can be seen in table 5.2, the test results of training with these ten additional images outperform the results of the previous models, increasing both Dice and IoU.

Finally, despite the apparent improvement obtained by training with ten additional corrected images, we requested specialists to correct the last five images to check if training with 15 images would result in an even more significant improvement than with 10. This way, we could assess the optimal number of images to correct in each iteration. Table 5.2 shows the results of the four models trained. As can be seen, the last model trained with the training set and adding 15 corrected annotations does not improve the previous model trained, adding the ten most uncertain samples. Therefore, we argue that the optimal number of annotation corrections required to improve the prediction of regions of interest in biopsies with UC through an active learning and human-in-the-loop approach corresponds to 10 WSIs.

Table 5.2: Test results for the preliminary model and the models retrained adding the five (AL w/ 5), the ten (AL w/ 10), and the 15 most uncertain samples (AL w/ 15).

WSI set	Number of WSIs	Dice	IoU
Training set	33	0.622	0.386
AL w/ 5	33 + 5	0.636	0.384
AL w/ 10	33 + 10	**0.651**	**0.415**
AL w/ 15	33 + 15	0.651	0.416

Chapter 6

Conclusions and future lines

6.1. Conclusions

The global rise in ulcerative colitis incidence presents a growing challenge for clinical centres, straining their resources and management. As demonstrated in this project, numerous scoring systems have been introduced to assess UC, encompassing both endoscopic and histological aspects. Recent clinical guidelines have endorsed histologic remission as the treatment objective, introducing a simplified, neutrophil-focused index known as PHRI for assessing histologic remission. To address this, the PICaSSO project's primary goal is to develop an automatic deep learning-based algorithm capable of predicting histologic remission in whole-slide images. For that purpose, this work aimed to improve the segmentation of regions of interest likely to contain neutrophils to further enhance the accuracy of the subsequent classification of the PHRI index.

The project benefited from an extensive database comprising whole-slide images from patients diagnosed with UC. This database was instrumental in validating the applicability of PHRI as an index with AI potential and in developing the current project. Notably, the dataset's standout feature is its multicenter and international origin. While this introduced challenges related to variability in acquisition techniques and colour staining, it also served as a valuable means to assess the model's robustness in predicting the diverse regions of interest.

The dataset included pixel-level annotations for selected whole slide images, and these annotations played a key role in the model development process. They provided critical information by indicating the precise locations of neutrophils and other key cells, which are essential for classifying UC according to the PHRI, as indicated in the table 2.1. Importantly, the patient-level partitions proposed in the study were influenced by the availability of annotations for each WSI, shaping the research approach accordingly.

Before algorithm development, an exhaustive literature review was conducted to explore novel techniques used in CAD systems for UC and the segmentation of whole-slide images. This investigation unveiled that encoder-decoder algorithms were the cutting-edge methods for image segmentation tasks, including WSIs. Furthermore, active learning methods were explored to enhance model performance by striking a balance between the need for annotations from pathologists, which incurs significant time investments. In addition to this, the review also delved into the primary uncertainty measures employed in active learning frameworks for segmentation tasks. The objective was to extract the most quantity of information from unlabeled images, a critical aspect of optimizing the

active learning process.

Since PHRI is essentially a neutrophil-based scoring system, the goal in developing the deep learning models was to improve the performance of an initial segmentation model to identify regions where neutrophils might be present. Pixel-level data, while valuable, can be challenging to acquire due to the time and effort required for annotation. Nevertheless, in this project, a strategic approach was adopted to maximize the utility of pixel-level annotations. This was achieved through the implementation of an active learning methodology, which enabled the models to benefit from this valuable information for enhanced performance.

The proposed active learning framework is based on the calculation of uncertainty maps of an unlabelled set of WSIs and the aggregation of the most uncertain samples to the training set, to obtain a more robust and accurate segmentation model. Before the addition, from the batch selected, we extracted their predictions and pathologists were asked to correct them. Then, the most uncertain WSIs with the corrected annotation were aggregated to the training set to retrain de segmentation model. This process was repeated in batches from three sizes (5, 10, 15) to find the optimum number of WSI annotations to be corrected to improve the model without increasing unnecessarily the implication of the pathologists.

In the last chapter, we aimed not only to present the results but also a further explain how was the model constructed and evaluated. The most significant conclusion about the results is that the model's ability to segment the four regions of interest of WSI improves when ten corrected annotations by pathologists are added to the training set. In conclusion, we reached the goal of the project which aimed to design an automatic deep learning-based system for segmenting histologic images before its PHRI classification framework. As no other AL approaches have used WSI for UC diagnosis, we will present for publication in clinical journals the algorithms and results obtained in this project.

6.2. Future lines

One of the main limitations of the present project is the labelling strategy for the various tissue segments, that determine the subsequent performance of the model. As previously explained, the model gets confused sometimes and predicts the non-relevant segments of tissue as a lamina propria, one of the interest areas, since they are very similar. This is because pathologists did not annotate all tissue areas, but only annotated the regions of interest for neutrophil detection, and tissue segments corresponding to the non-interest region were left unannotated. For this reason, in the masks obtained, where each of these areas is assigned a GT value between 1 and 4, the background and the non-interest region were annotated as GT = 0. Despite efforts to predict the region of non-interest, the results obtained remain unsatisfactory. Therefore, it will be imperative to explore new approaches in future research aimed at providing the neural network with the ability to distinguish between the region of non-interest and the lamina propria. Such advances would substantially improve the accuracy of the neural network, leading to a more reliable detection of the neutrophils present, thus contributing to a more accurate estimation of the PHRI.

Another fundamental limitation of the study is the low number of annotated images available for the initial training of the model. As mentioned above, the task of annotation requires subject matter expertise that is only available to pathologists and is a time-consuming task, being exhaustive in the annotations for better data quality. This is why, as this is an initial study of this type of iterative technique applied to UC segmentation, we had a small number of images with which to prepare the segmentation model. A larger number of segmented images will be needed in the future to improve performance and this is where the proposed active learning methodological framework can be of great importance.

The segmentation problem we have faced is, as discussed throughout the document, a step prior to full PHRI prediction for the characterisation of activity (in remission or active) of histological images. Grading the UC disease using the PHRI can significantly alleviate the pathologist's workload and aid in clinical decision-making. However, accurately predicting the clinical outcomes of UC patients based solely on the analysis of histological slides or endoscopic videos remains challenging. Clinical outcomes in this context cover various events that may occur during UC relapse phases. These events include surgical procedures, treatment changes, or hospitalization. It's worth noting that the latter two are typically associated with relapse episodes where UC symptoms worsen and clinical intervention becomes necessary.

Surgical intervention, often involving colectomy, is one such clinical outcome. Colectomy is a procedure used not only for UC but also for other colon diseases like cancer and diverticulitis. It entails the removal of the affected part of the bowel, where inflammation or cancer has rapidly progressed, to prevent its spread to other sections of the colon [1]. Consequently, it is imperative to differentiate between active UC and phases characterized by inflammation that is subsiding or complete remission. Notably, early diagnosis and improved treatments have contributed to a reduction in the incidence of colectomy in UC cases [42]. However, it's essential to recognize that both early and late postoperative complications associated with colectomy should not be underestimated [43].

The development and validation of PHRI as a simplified score for UC management, with applicability in AI-based systems, is part of a larger international project known as PICaSSO. The overarching goal of this project is to create an integrated system that

combines the analysis of endoscopic videos and histological evaluations to ultimately assess the clinical outcome of UC patients.

However, at this juncture, it appears more reasonable to combine both endoscopic and histological data. In the context of deep learning, the late fusion of multimodal information involves merging the extracted features from different data sources. By exploring feature fusion for endoscopic and histological data, the diagnostic capabilities of the outcome prediction algorithm are expected to improve. One challenge to address in feature fusion is the difference in feature sizes in the latent space. This difference arises from the use of distinct convolutional neural network (CNN) backbone architectures, resulting in different feature sizes. This discrepancy in feature dimensions will need to be managed effectively to integrate the information from both modalities successfully.

Bibliography

[1] Maia Kayal and Shailja Shah. Ulcerative colitis: Current and emerging treatment strategies. *Journal of Clinical Medicine*, 9(1), 2020.

[2] Rish Pai and Karel Geboes. Disease activity and mucosal healing in inflammatory bowel disease: a new role for histopathology? 472.

[3] Rocío Del Amor, Pablo Meseguer, Tommaso Lorenzo Parigi, Vincenzo Villanacci, Adrián Colomer, Laëtitia Launet, Alina Bazarova, Gian Eugenio Tontini, Raf Bisschops, Gert de Hertogh, et al. Constrained multiple instance learning for ulcerative colitis prediction using histological images. *Computer Methods and Programs in Biomedicine*, page 107012, 2022.

[4] Xianyong Gui, Alina Bazarova, Rocìo Del Amor, Michael Vieth, Gert de Hertogh, Vincenzo Villanacci, Davide Zardo, Tommaso Lorenzo Parigi, Elin Synnøve Røyset, Uday N Shivaji, et al. Picasso histologic remission index (phri) in ulcerative colitis: development of a novel simplified histological score for monitoring mucosal healing and predicting clinical outcomes and its applicability in an artificial intelligence system. *Gut*, 71(5):889–898, 2022.

[5] Kathleen A Head and Julie S Jurenka. Inflammatory bowel disease part i: Ulcerative colitis – pathophysiology and conventional and alternative treatment options. 8(3).

[6] Natalie A. Molodecky, Ing Shian Soon, Doreen M. Rabi, William A. Ghali, Mollie Ferris, Greg Chernoff, Eric I. Benchimol, Remo Panaccione, Subrata Ghosh, Herman W. Barkema, and Gilaad G. Kaplan. Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. *Gastroenterology*, 142(1):46–54.e42, 2012.

[7] Johannes Meier and Andreas Sturm. Current treatment of ulcerative colitis. 17(27):3204–3212.

[8] Sunhee Park, Tsion Abdi, Mark Gentry, and Loren Laine. Histological disease activity as a predictor of clinical relapse among patients with ulcerative colitis: systematic review and meta-analysis. *Official journal of the American College of Gastroenterology| ACG*, 111(12):1692–1701, 2016.

[9] Marietta Iacucci, Marco Daperno, Mark Lazarev, Razvan Arsenascu, Gian Eugenio Tontini, Oluseyi Akinola, Xianyong Sean Gui, Vincenzo Villanacci, Martin Goetz, Mark Lowerison, et al. Development and reliability of the new endoscopic virtual chromoendoscopy score: the picasso (paddington international virtual chromoendoscopy score) in ulcerative colitis. *Gastrointestinal endoscopy*, 86(6):1118–1127, 2017.

[10] Robert C. Langan, Patricia B. Gotsch, Michael A. Krafczyk, and David D. Skillinge. Ulcerative colitis: Diagnosis and treatment. 76(9):1323–1330.

[11] R. D. COHEN, A. P. YU, E. Q. WU, J. XIE, P. M. MULANI, and J. CHAO. Systematic review: the costs of ulcerative colitis in western countries. *Alimentary Pharmacology & Therapeutics*, 31(7):693–707, 2010.

[12] Torsten Kucharzik, Sibylle Koletzko, Klaus Kannengiesser, and Axel Dignass. Ulcerative colitis—diagnostic and therapeutic algorithms. 117(33):564–574.

[13] Britt Christensen and David T Rubin. Understanding endoscopic disease activity in ibd: how to incorporate it into practice. *Current gastroenterology reports*, 18(1):1–11, 2016.

[14] F Magro, C Langner, A Driessen, ARZU Ensari, K Geboes, GJ Mantzaris, V Villanacci, G Becheanu, P Borralho Nunes, G Cathomas, et al. European consensus on the histopathology of inflammatory bowel disease. *Journal of Crohn's and Colitis*, 7(10):827–851, 2013.

[15] Kenneth W Schroeder, William J Tremaine, and Duane M Ilstrup. Coated oral 5-aminosalicylic acid therapy for mildly to moderately active ulcerative colitis. *New England Journal of Medicine*, 317(26):1625–1629, 1987.

[16] Simon PL Travis, Dan Schnell, Piotr Krzeski, Maria T Abreu, Douglas G Altman, Jean-Frédéric Colombel, Brian G Feagan, Stephen B Hanauer, Marc Lémann, Gary R Lichtenstein, et al. Developing an instrument to assess the endoscopic severity of ulcerative colitis: the ulcerative colitis endoscopic index of severity (uceis). *Gut*, 61(4):535–542, 2012.

[17] Mahmoud H Mosli, Brian G Feagan, Guangyong Zou, William J Sandborn, Geert D'Haens, Reena Khanna, Lisa M Shackelton, Christopher W Walker, Sigrid Nelson, Margaret K Vandervoort, et al. Development and validation of a histological index for uc. *Gut*, 66(1):50–58, 2017.

[18] Aude Marchal-Bressenot, Julia Salleron, Camille Boulagnon-Rombi, Claire Bastien, Virginie Cahn, Guillaume Cadiot, Marie-Danièle Diebold, Silvio Danese, Walter Reinisch, Stefan Schreiber, et al. Development and validation of the nancy histological index for uc. *Gut*, 66(1):43–49, 2017.

[19] Heang-Ping Chan, Lubomir M. Hadjiiski, and Ravi K. Samala. Computer-aided diagnosis in the era of deep learning. *Medical Physics*, 47(5):e218–e227, 2020.

[20] Howard Lee and Yi-Ping Phoebe Chen. Image based computer aided diagnosis system for cancer detection. *Expert Systems with Applications*, 42(12):5356–5365, 2015.

[21] Guihua Chen and Jun Shen. Artificial Intelligence Enhances Studies on Inflammatory Bowel Disease. *Frontiers in Bioengineering and Biotechnology*, 9, 2021.

[22] Jianxing He, Sally L Baxter, Jie Xu, Jiming Xu, Xingtao Zhou, and Kang Zhang. The practical implementation of artificial intelligence technologies in medicine. *Nature medicine*, 25(1):30–36, 2019.

[23] Huangming Zhuang, Jixiang Zhang, and Fei Liao. A systematic review on application of deep learning in digestive system image processing. *The Visual Computer*, 39(6):2207–2222, June 2023.

[24] Yanyun Fan, Ruochen Mu, Hongzhi Xu, Chenxi Xie, Yinghao Zhang, Lupeng Liu, Lin Wang, Huaxiu Shi, Yiqun Hu, Jianlin Ren, Jing Qin, Liansheng Wang, and Shuntian Cai. Novel deep learning–based computer-aided diagnosis system for predicting inflammatory activity in ulcerative colitis. *Gastrointestinal Endoscopy*, 97(2):335–346, 2023.

[25] Muhammad Khalid Khan Niazi, Anil V Parwani, and Metin N Gurcan. Digital pathology and artificial intelligence. *The lancet oncology*, 20(5):e253–e261, 2019.

[26] Niels Vande Casteele, Jonathan A Leighton, Shabana F Pasha, Frank Cusimano, Aart Mookhoek, Catherine E Hagen, Christophe Rosty, Reetesh K Pai, and Rish K Pai. Utilizing deep learning to analyze whole slide images of colonic biopsies for associations between eosinophil density and clinicopathologic features in active ulcerative colitis. *Inflammatory Bowel Diseases*, 28(4):539–546, 2022.

[27] Wenyuan Li, Jiayun Li, Zichen Wang, Jennifer Polson, Anthony E. Sisk, Dipti P. Sajed, William Speier, and Corey W. Arnold. Pathal: An active learning framework for histopathology image analysis. *IEEE Transactions on Medical Imaging*, 41(5):1176–1187, 2022.

[28] Robert (Munro) Monarch. *Human-in-the-Loop Machine Learning: Active learning and annotation for human-centered AI*. Simon and Schuster, August 2021. Google-Books-ID: bNo2EAAAQBAJ.

[29] Samuel Budd, Emma C. Robinson, and Bernhard Kainz. A survey on active learning and human-in-the-loop deep learning for medical image analysis. *Medical Image Analysis*, 71:102062, July 2021.

[30] Jamshid Sourati, Ali Gholipour, Jennifer G. Dy, Sila Kurugol, and Simon K. Warfield. Active Deep Learning with Fisher Information for Patch-Wise Semantic Segmentation. In Danail Stoyanov, Zeike Taylor, Gustavo Carneiro, Tanveer Syeda-Mahmood, Anne Martel, Lena Maier-Hein, João Manuel R.S. Tavares, Andrew Bradley, João Paulo Papa, Vasileios Belagiannis, Jacinto C. Nascimento, Zhi Lu, Sailesh Conjeti, Mehdi Moradi, Hayit Greenspan, and Anant Madabhushi, editors, *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Lecture Notes in Computer Science, pages 83–91, Cham, 2018. Springer International Publishing.

[31] A. Smailagic, H. Noh, P. Costa, Devesh Walawalkar, Kartik Khandelwal, M. Mirshekari, Jonathon Fagert, A. Galdran, and Susu Xu. MedAL: Deep Active Learning Sampling Method for Medical Image Analysis. *ArXiv*, September 2018.

[32] Lin Yang, Yizhe Zhang, Jianxu Chen, Siyuan Zhang, and Danny Z. Chen. Suggestive Annotation: A Deep Active Learning Framework for Biomedical Image Segmentation. In Maxime Descoteaux, Lena Maier-Hein, Alfred Franz, Pierre Jannin, D. Louis Collins, and Simon Duchesne, editors, *Medical Image Computing and Computer Assisted Intervention MICCAI 2017*, Lecture Notes in Computer Science, pages 399–407, Cham, 2017. Springer International Publishing.

[33] Dwarikanath Mahapatra, Alexander Poellinger, Ling Shao, and Mauricio Reyes. Interpretability-Driven Sample Selection Using Self Supervised Learning for Disease Classification and Segmentation. *IEEE Transactions on Medical Imaging*, 40(10):2548–2562, October 2021. Conference Name: IEEE Transactions on Medical Imaging.

[34] André L S Meirelles, Tahsin Kurc, Jun Kong, Renato Ferreira, Joel Saltz, and George Teodoro. Effective and efficient active learning for deep learning-based tissue image analysis. *Bioinformatics*, 39(4):btad138, April 2023.

[35] Scott Doyle, James Monaco, Michael Feldman, John Tomaszewski, and Anant Madabhushi. An active learning based classification strategy for the minority class problem: application to histopathology annotation. *BMC Bioinformatics*, 12(1):424, October 2011.

[36] Dwarikanath Mahapatra, Behzad Bozorgtabar, Jean-Philippe Thiran, and Mauricio Reyes. Efficient Active Learning for Image Classification and Segmentation Using a Sample Selection and Conditional Generative Adversarial Network. In Alejandro F. Frangi, Julia A. Schnabel, Christos Davatzikos, Carlos Alberola-López, and Gabor Fichtinger, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, Lecture Notes in Computer Science, pages 580–588, Cham, 2018. Springer International Publishing.

[37] Xu Jin, Hong An, Jue Wang, Ke Wen, and Zheng Wu. Reducing the Annotation Cost of Whole Slide Histology Images using Active Learning. In *Proceedings of the 2021 3rd International Conference on Image Processing and Machine Vision*, IPMV '21, pages 47–52, New York, NY, USA, August 2021. Association for Computing Machinery.

[38] Si Wen, Tahsin M. Kurc, Le Hou, Joel H. Saltz, Rajarsi R. Gupta, Rebecca Batiste, Tianhao Zhao, Vu Nguyen, Dimitris Samaras, and Wei Zhu. Comparison of Different Classifiers with Active Learning to Support Quality Control in Nucleus Segmentation in Pathology Images. *AMIA Summits on Translational Science Proceedings*, 2018:227–236, May 2018.

[39] Soufiane Belharbi, Ismail Ben Ayed, Luke McCaffrey, and Eric Granger. Deep Active Learning for Joint Classification & Segmentation with Weak Annotator. In *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 3337–3346, Waikoloa, HI, USA, January 2021. IEEE.

[40] Marietta Iacucci, Samuel CL Smith, Alina Bazarova, Uday N Shivaji, Pradeep Bhandari, Rosanna Cannatelli, Marco Daperno, Jose Ferraz, Martin Goetz, Xianyong Gui, et al. An international multicenter real-life prospective study of electronic chromoendoscopy score picasso in ulcerative colitis. *Gastroenterology*, 160(5):1558–1569, 2021.

[41] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Lecture Notes in Computer Science, pages 234–241, Cham, 2015. Springer International Publishing.

[42] Laura E Targownik, Harminder Singh, Zoann Nugent, and Charles N Bernstein. The epidemiology of colectomy in ulcerative colitis: results from a population-based cohort. *Official journal of the American College of Gastroenterology| ACG*, 107(8):1228–1235, 2012.

[43] L Peyrin-Biroulet, A Germain, AS Patel, and JO Lindsay. Systematic review: outcomes and post-operative complications following colectomy for ulcerative colitis. *Alimentary pharmacology & therapeutics*, 44(8):807–816, 2016.

Part II

Budget

Chapter 7

Budget

7.1. Scope

This section is intended to conduct a financial assessment of the project's economic costs associated with the development of deep learning models for semantic segmentation in WSIs with UC.

7.2. Partial budgets

The project's overall budget has been subdivided into three distinct partial budgets: personnel costs, software costs, and hardware costs. This division enables us to provide a comprehensive breakdown and explanation of each of these individual budget components.

7.2.1. Personnel costs

This section aims to consider the human resources required to execute the project. The project's development has engaged the expertise of three researchers:

- D Valero Laparra Pérez-Muelas, Ps.D. assistant professor (UV).
- Dª Rocío del Amor del Amor, Ph.D. student (UPV - CVBLab).
- D Pablo Meseguer Esbri, Ph.D. student (UPV - CVBLab).
- D Fernando García Torres, master's student.

The student can be considered a junior biomedical engineer and has taken the lead in developing the majority of the project and writing this report. Since the master's thesis corresponds to 18 ECTS credits, the estimated time the student spent on project development is approximately 450 hours.

The work received guidance from an assistant professor at the 'Universitat de València', who oversaw the project and reviewed the final manuscript. Simultaneously, two PhD students from 'Universitat Politècnica de València' supervised the development of all project components. The breakdown of personnel costs is detailed in Table 7.1.

Table 7.1: Breakdowns of personnel costs

Description	Duration (h)	Unitary cost (€/h)	Total cost (€)
Assistant professor	15	30	450
Ph.D student	35	20	700
Ph.D student	50	20	1000
Student	450	12	5400
TOTAL			**€6650**

7.2.2. Hardware costs

The following section outlines the hardware costs, which encompass expenses associated with both the personal computer and the specialized hardware resources required for deep model training.

For the core activities of the project, such as programming and manuscript writing, a *Lenovo Legion Y520* laptop was used. This laptop is equipped with an *Intel Core i7®* processor and an *NVIDIA GTX 1050* graphics card. However, due to its limited computing power, an *NVIDIA DGX A100* system, owned by the CVBLab, was employed for training the neural networks. While the overall cost of the *DGX A100* system is approximately €200,000, only one of its eight GPUs was utilized for this project, resulting in a partial cost of €25,000. The breakdown of the hardware costs is presented in Table 7.2.

Table 7.2: Detail of hardware costs.

Description	Units (uds)	Unitary cost (€/uds)	Useful life (months)	Use time (months)	Total cost (€)
LENOVO LEGION Y520	1	999	84	8	95.14
NVIDIA DGX A100 (1 GPU)	1	25000	132	3	568.18
TOTAL					**€663.32**

7.2.3. Software costs

In the following section, the software costs are detailed, which encompass expenses related to licenses for computerized systems and programming environments.

It's worth noting that some free software resources were also utilized, such as *VisualStudio Code* software was utilized to establish a secure SSH connection between the student's computer and the DGX system for high-capacity GPU model training. Additionally, *Overleaf* for document writing and the *Pytorch* library for deep learning model development in *Python*. These resources are not included in the table.

Matlab (Mathworks ®) in its R2019b version was employed for various tasks, including image preprocessing tasks.

The breakdown of the software costs is presented in Table 7.3.

Table 7.3: Detail of software costs.

Description	Units (uds)	Unitary cost (€/uds)	Useful life (months)	Use time (months)	Total cost (€)
Matlab R2019b	1	800	12	6	400
TOTAL					**400€**

7.3. Total project cost

In conclusion, the economic evaluation of the project is summarized in Table 7.4. The total budget is calculated by summing the partial costs detailed in the preceding sections.

Table 7.4: Detail of the execution budget of the project.

Description	Cost (€)
Personnel costs	6650
Software budget	400
Hardware costs	663.32
Total budget	**€7,712.32**

Finally, the total cost of developing this Master's Thesis is determined by combining the material execution costs, general expenses, and industrial profit. The industrial profit is calculated as 6% of the general expenses, and the general expenses are augmented by a 13% percentage. Additionally, the applicable taxes for Spain, including the Value Added Tax (*Impuesto sobre el Valor Añadido* or IVA), which corresponds to 21% of the total cost, are included. The total amount is presented in Table 7.5.

Table 7.5: Detail of the total budget of the project.

Description	Cost (€)
Execution budget	7,712.32
General expenses	1,002.60
Industrial profit	462.80
SUM	**€9,177.72**
IVA (21%)	€1,927.32
TOTAL BUDGET	**€11,105.04**

Therefore, the total projected budget is **ELEVEN THOUSAND ONE HUNDRED AND AND FIVE EUROS AND FOUR CENTS**.