People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research
وزارة التعليم العالـي والبحـث العلمـي
جامعة عين تموشنت بلحاج بوشعيب
Ain Temouchent Belhadj Bouchaib University
Faculty of Science and Technology
Department of Mathematics and Computer Science

Final Year Project
To obtain a Master's degree in: **CYSIA**
Field**: Mathematics and Computer Science**
Sector**: Computer Science**
Specialty**: Cybersecurity and Artificial Intelligence**

Theme :

Predict Microsatellite Instability (MSI) and Microsatellite Stability (MSS) status in gastrointestinal cancers

---

Presented By:

1) Miss Hibi Lara
2) Miss Mekadid Hafsa Amina

Before the jury composed of:

Dr Benomar Mohamed Lamine	UAT.B.B (Ain Témouchent) President
Dr Saidi Samira	UAT.B.B (Ain Témouchent) Examiner
Dr Beddad Fatima	UAT.B.B (Ain Témouchent) Supervisor
Dr Bousahba Nassima	U.H.B.C (Chelf) co-Supervisor

Academic year : 2024/2025

Abstract

In colorectal cancer (CRC), MSI status is the crucial biomarker that influences treatment approaches, especially immunotherapy. However, the current diagnostic techniques, such as polymerase chain reaction (PCR) and immunohistochemistry (IHC), are time-consuming and variable.This study uses histopathological images to automatically predict the Microsatellite Instability (MSI) and Microsatellite Stability (MSS) status in gastrointestinal cancers using a novel deep learning framework. We suggest an AI-powered approach, Deep learning models to address these issues by combining cutting-edge computer vision methods for nuclear feature segmentation and classification in histopathology slides.The results we achieve shows how deep learning can revolutionise the digital pathology by providing a reliable, accurate, and scalable substitutes for conventional MSI and MSS diagnostics. With implications for enhancing patient outcomes by quicker and more accurate diagnosis, this work is an important breakthrough towards achieving precision oncology.

Keywords: Deep learning, Microsatellite Instability (MSI), Colorectal Cancer (CRC), Histopathology, Digital Pathology, Artificial Intelligence , classification , segmentation .

Résumé

Dans le cancer colorectal (CCR), le statut MSI est un biomarqueur crucial qui influence les approches thérapeutiques, notamment l'immunothérapie. Cependant, les techniques diagnostiques actuelles, telles que la réaction en chaîne par polymérase (PCR) et l'immunohistochimie (IHC), sont chronophages et variables.Cette étude utilise des images histopathologiques pour prédire automatiquement l'instabilité microsatellite (MSI) et la stabilité microsatellite (MSS) dans les cancers gastro-intestinaux grâce à un nouveau cadre d'apprentissage profond. Nous proposons une approche basée sur l'IA et des modèles d'apprentissage profond pour répondre à ces problèmes en combinant des méthodes de vision par ordinateur de pointe pour la segmentation et la classification des caractéristiques nucléaires sur des lames d'histopathologie.Les résultats obtenus montrent comment l'apprentissage profond peut révolutionner la pathologie numérique en offrant des alternatives fiables, précises et évolutives aux diagnostics MSI et MSS conventionnels. Avec des implications pour l'amélioration des résultats pour les patients grâce à un diagnostic plus rapide et plus précis, ces travaux constituent une avancée importante vers une oncologie de précision.

Mots clés : Apprentissage profond, Instabilité des microsatellites (MSI), Cancer colorectal (CCR), Histopathologie, Pathologie numérique, Intelligence artificielle, classification, segmentation.

الملخص

في سرطان القولون والمستقيم (CRC)، تُعد حالة عدم استقرار السواتل الدقيقة المؤشر الحيوي الحاسم الذي يؤثر على أساليب العلاج، وخاصةً العلاج المناعي. ومع ذلك، فإن تقنيات التشخيص الحالية، مثل تفاعل البوليميراز المتسلسل (PCR) و الكيمياء المناعية النسيجية (IHC)، تستغرق وقتًا طويلاً وتتغير باستمرار.تستخدم هذه الدراسة صورًا نسيجية مرضية للتنبؤ تلقائيًا بحالة عدم استقرار السواتل الدقيقة (MSI) واستقرار السواتل الدقيقة (MSS) في سرطانات الجهاز الهضمي، وذلك باستخدام إطار عمل مبتكر للتعلم العميق (DL). نقترح نهجًا قائمًا على الذكاء الاصطناعي، ونماذج التعلم العميق، لمعالجة هذه المشكلات من خلال الجمع بين أحدث أساليب الرؤية الحاسوبية لتجزئة وتصنيف السمات النووية في شرائح علم الأمراض النسيجية. تُظهر النتائج التي توصلنا إليها كيف يُمكن للتعلم العميق إحداث ثورة في علم الأمراض الرقمية من خلال توفير بدائل موثوقة ودقيقة وقابلة للتطوير التشخيصات عدم استقرار السواتل الدقيقة(MSI) واستقرار السواتل الدقيقة (MSS) التقليدية، مع آثاره على تحسين نتائج المرضى من خلال تشخيص أسرع وأكثر دقة، يُعد هذا العمل إنجازًا مهمًا نحو تحقيق علم الأورام الدقيق.

الكلمات المفتاحية: التعلم العميق، عدم استقرار التتابعات الدقيقة المتكرر (MSI) ، سرطان القولون والمستقيم (CRC)، علم الأمراض النسيجي، علم الأمراض الرقمي، الذكاء الاصطناعي، التصنيف، التجزئة.

Acknowledgements

All praise and thanks are due to Allah for granting me the strength and clarity to complete this journey .

I extend my sincere gratitude to my supervisor, Dr. Beddad Fatima (UAT.B.B), for her invaluable guidance and continuous support.

Special thanks to my co-supervisor, Dr. Bousahba Nassima (U.H.B.C), for her advice and contributions.

I also thank Dr. Benomar Mohamed Lamine, President of the jury, and Dr. Saidi Samira, Examiner, for their time and constructive feedback.

My appreciation goes to all the professors and staff of the Computer Science Department at UAT.B.B.

To my family, friends, and classmates thank you for your constant support and encouragement throughout this journey.

HIBI LARA

Dedication

In the name of Allah, the Most Gracious, the Most Merciful

All praise is due to Allah alone, the Sustainer of all worlds.

I dedicate this work:

To my dear mother, for her boundless love, prayers, and unwavering support , you are my strength and my peace.

To my beloved father, whose unwavering care, wisdom, and constant encouragement have been a guiding light throughout every step of my path..

To my sister and brothers, for their love, patience, and belief in me throughout this journey.

To my dear friend and partner in this project MEKADID HAFSA AMINA ,your companionship , effort and support has brought meaning to even

the most challenging moments.

To my loved ones, who have surrounded me with kindness, motivation, light ,and hope in every moment.

With all my heart, I thank you , this achievement belongs to you as much as it does to me.

HIBI LARA

Acknowledgements

First and foremost,

I am deeply grateful to Allah for granting me the strength

patience, and guidance to complete this work.

Without His mercy and blessings, none of this would have been possible.

I would like to express my sincere thanks to my supervisor. Beddad Fatima (UAT.B.B), for his continuous support, valuable feedback, and encouragement throughout the research and writing process.

I also extend my heartfelt thanks to my co-supervisor, Dr. Bousahba Nassima (U.H.B.C), for her insightful advice and valuable contributions to this work

Her guidance has been crucial in shaping this journey.

My heartfelt appreciation goes to all my teachers and professors, whose knowledge, dedication, and inspiration have played an essential role in my academic growth.

I also extend my gratitude to the board of examiners for taking the time to evaluate this work and for their comments and recommendations.

MEKADID HAFSA AMINA

Dedication

In the name of Allah, the Most Gracious, the Most Merciful

To my dear parents,

for their unconditional love, endless sacrifices, and unwavering support throughout my academic journey.

To my sister,

for her affection, understanding, and constant encouragement.

To my two brothers,

for their kindness, support, and comforting presence.

To my partner ,

Hibi Lara your patientence encouragement, and unwavering presence have meant more than words can express.

To all my classmates,

with whom I shared unforgettable moments of learning, perseverance, and friendship.

Thank you all for contributing, each in your own way, to the completion of this work.

MEKADID HAFSA AMINA

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AI: Artificial Intelligence

ANN: Artificial Neural Network

AUC: Area Under Curve

AUC-ROC: Area Under the Receiver Operating Characteristic Curve

ACC: Accuracy

CAD: Computer Aided Diagnosis

CAM: Class Activation Map

CIoU: Complete Intersection over Union

CIN: Chromosomal Instability

CLIP: Contrastive Language mage Pretraining

CNN: Convolutional Neural Network

CoNIC: Colon Nuclei Identification and Classification dataset

CRC: Colorectal Cancer

CSP: Cross Stage Partial

CT: Computed Tomography

DL: Deep Learning

DSC: Dice Similarity Coefficient

FAIR: Facebook AI Research

FCN: Fully Convolutional Network

FFN: Feed-Forward Network

FN: False Negative

FP: False Positive

TN:True Negative

FN:False Negative

FPN: Feature Pyramid Network

GDC: Genomic Data Commons

GPU: Graphics Processing Unit

H&E: Hematoxylin and Eosin

HRNet: High-Resolution Network

IHC: Immunohistochemistry

IoU: Intersection over Union

MiT: Mix Transformer

ML: Machine Learning

MLP: Multi-Layer Perceptron

MSA:Multi-Head Self -Attention

MMR: Mismatch Repair

MRI: Magnetic Resonance Imaging

MSI: Microsatellite Instability

MSI-H: Microsatellite Instability-High

MSS: Microsatellite Stability

NLP: Natural Language Processing

PCR: Polymerase Chain Reaction

PVT: Pyramid Vision Transformer

QC: Quality Control

ReLU: Rectified Linear Unit

ROC: Receiver Operating Characteristic

SA: Segment Anything

SAM: Segment Anything Model

SegFormer: Transformer-based Semantic Segmentation Model

TCGA: The Cancer Genome Atlas

TCGA-COAD: The Cancer Genome Atlas Colon Adenocarcinoma

TN: True Negative

TP: True Positive

U-Net: Convolutional Neural Network with U-shape architecture

ViT: Vision Transformer

WSI: Whole Slide Image

C2F:Cross Stage Partial with 2 convolutions

General Introduction

Computing and medical image analysis technology has supported tools and techniques for automatic histopathological image diagnosis, especially for cancer diagnoses.

The technology has transformed significantly in the past couple of decades, and the combination of deep learning and computer vision approaches has led to efficient decision-support systems that are applicable in the industrial, medical, and research communities. Proven tasks including classification, segmentation, and detection by being able to effectively learn intricate patterns from medical images .

The application of DL models in histopathology has been made possible through digitisation in whole slide images (WSIs) format.Unprocessed from H&E-stained slides, the models have been used in applications like tumour detection,tumour grading, and prediction of molecular phenotype, including microsatellite instability (MSI) status Pathologists apply image analysis and pattern recognition techniques on a routine basis to help them interpret microscopic slides for cell structure and tissue patterns[18]. Two general types of examination are cytological examination, in which samples are spread, fixed, and stained onto slides to determine cell types, and histological examination, in which tissue sections are studied using a microscope.

Histopathological investigation is still the prevailing diagnostic technique for cancer regardless of how extensive and sophisticated diagnostic technology has become. Manual interpretation of slides, on the other hand, takes a lot of time, is knowledge-intensive, and may be invariable and variable, especially in identifying subtle or unusual morphological features .

Computer aided diagnosis (CAD) systems have been found to be an effective solution for these issues. To aid clinicians, the systems are designed in a manner that provides quick, accurate, and reproducible results.

Deep learning (DL), and especially convolutional neural networks (CNNs), significantly most prevalent in the globe and the most prevalent cause of cancer fatalities globally is CRC. Of clinical utility molecular phenotypes, including MSI, can be predicted from histopathology images based on DL algorithms .

DNA mismatch repair system abnormalities are the cause of microsatellite instability (MSI), a relevant molecular marker for colorectal cancer (CRC). About 15% of CRCs fall into the MSI-high (MSI-H) category, and the remaining 15% are microsatellite-stable (MSS) cases.[2] MSI status is a predictive and prognostic biomarker for the response to immunotherapy.MSI detection is presently realized in the lab by techniques like immunohistochemistry (IHC) and polymerase chain reaction (PCR). These techniques have limitations despite their efficiency: Promega, n.d.) IHC is prone to variability through staining and interpretation, while PCR requires tumour and normal tissue samples, specialized equipment, and skilled personnel.Low-cost, non-invasive, and automated solutions are sorely required in the face of such an issue. Deep learning for the processing of H&E-stained histology images is one such solution. Models learn morphological characteristics of MSI-H on WSIs like tumor-infiltrating lymphocytes and mucinous differentiation to predict MSI/MSS status[1].

The aim of this project research is the conception and implementation of a deep learning model capable of classifying instances of colorectal cancer into MSI and MSS types based on histopathological images alone. It will, in addition to this, segment and visualize important morphological features in order to promote interpretability and mimic the pathologist's diagnostic pathway.

This report is organized into four chapters.The first chapter presents the basic concepts related to medical imaging in general.The second chapter focuses on the generation and digitization of histopathological images in the context of colorectal cancer, with an emphasis on assessing MSI vs MSS status.The third chapter discusses methods for classifying images according to MSI vs MSS status, as well as the segmentation of cell nuclei.Finally, the fourth chapter presents the experimental phase and the results obtained using the proposed methods, detailing the databases employed. This chapter concludes with a discussion and important analysis of the developed models.A general conclusion closes the thesis by recalling the research context and outlining perspectives for future work.

Chapter 01 : Medical imaging

Chapter 01 : Medical imaging

1.1. Introduction

Images are a universal medium for communication, analysis, and diagnosis. In the field of computer science, and more specifically in image processing, they are much more than a simple visual representation: they become digitally usable data. These data can be interpreted by machines, stored, processed, and analyzed automatically in numerous applications.

This chapter aims to define what an image and a digital image are, in order to understand the basics of image processing and their uses.

1.2. Definition of an image

An image is a visual representation of an object, a person, or a scene, created through means such as painting, photography, video, or generated digitally. It consists of a structured set of pixels arranged in a grid, each carrying an intensity or color value. For the human eye, an image becomes meaningful when displayed on a screen. For a machine, it is defined as a continuous analog luminance function **f(x, y)**, where **x** and **y** are spatial coordinates, and **f** is a function of light intensity or color. This representation must be digitized to be processed by computer systems[3].

1.3. Definition image digital

A digital image is a representation of a two-dimensional image as a finite set of digital values, known as picture elements or pixels. Each pixel has a specific value that determines its color and brightness, and collectively, these pixels form the complete image. Digital images are the foundation of various technologies and applications in fields ranging from photography and medical imaging to remote sensing and computer vision[4].

1.4. Type of image

1.4.1.binary image

It is a binary image, that is, an image whose pixels can only take two values: 0 or 1. The value 0 generally corresponds to a black pixel, while the value 1 represents a white pixel. Thus, the gray level is coded on a single bit [4].

Figure 1.1 binair image [4]**.**

1.4.2.Image grayscale

A grayscale image is a representation in which each pixel is encoded with 8 bits, allowing for 256 possible intensity levels (ranging from 0, corresponding to black, to 255, corresponding to white). This representation is frequently used in various imaging fields due to its simplicity of implementation, processing, and analysis [4].

Figure 1.2 : Grayscale image [4]**.**

1.4.3.Image Color

They are based on three primary colors,red, green, and blue. Each color is encoded in 8 bits, meaning that each pixel requires 24 bits (3 × 8) to represent the colorimetric information. This representation therefore offers the possibility of displaying up to 256 × 256× 256 , or over 16.7 million colors.This method is the most commonly used in imaging due to the richness of colors it can reproduce.

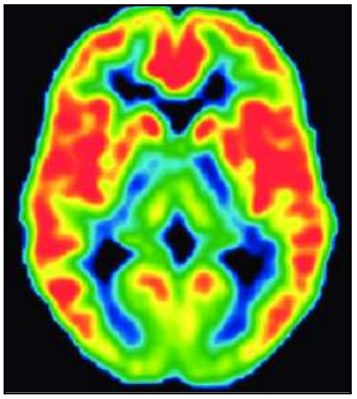


Figure 1.3: Image color [5].

1.5. Medical imaging technology

Medical imaging refers to all techniques used to produce images of the human body for diagnostic, therapeutic, or biomedical research purposes .There are many types of medical imaging, and new techniques continue to emerge with technological advances. Each modality relies on specific physical principles to generate images of the body's interior. Among the most widely used imaging methods in medicine.

1.5.1.Radiography

Radiography is based on the use of X-rays, which have the property of passing through tissues to a greater or lesser extent depending on their density. Thus, an X-ray emitting source is placed in front of the body to be X-rayed and a detector is placed at the back of the body. The emitted photons will pass through the body, being absorbed to a greater or lesser extent by the tissues encountered on their path. This makes it possible to differentiate between bones and muscles in the final image [6].

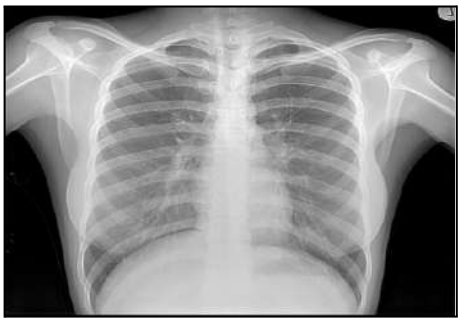


Figure 1.4 :example of chest x-ray [7] .

1.5.2.Scanar

Scanner is a medical device, also known as a computed tomography (CT) imaging device, that uses X-rays to obtain precise images of a specific region of the body. This device works by simultaneously rotating the X-ray source and detector around the patient, producing three-dimensional (3D) images. A computer then processes the resulting two-dimensional (2D) images to reconstruct detailed 3D views.

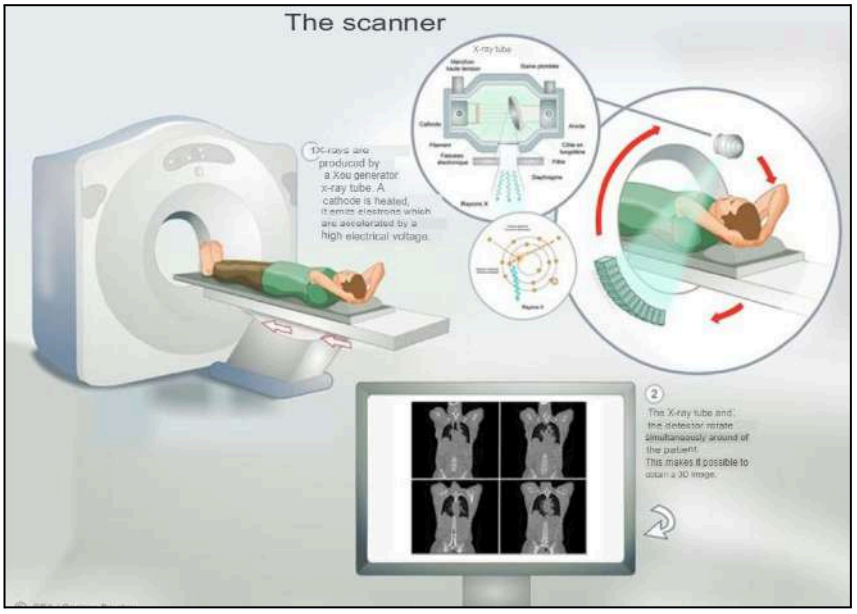


Figure 1.5: The stages of making the scanner [7]

1.5.3.Magnetic Resonance Imaging (MRI)

MRI is an examination that allows the visualization of the anatomical cavity and the transmission of information on lesions not visible in conventional X-rays, CT scans, or ultrasound. The principle is based on the magnetic property of numerous hydrogen atoms, which alone represent nearly 80% of the human body. A powerful magnetic field is used to examine the response of different tissues. The information is then processed using software to obtain a 2D or 3D image with great precision. The development of the MRI technique has paved the way for knowledge and research on the most complex organs, including the brain, spinal cord, muscles, etc.) [6].

Figure 1.6 : Principle of anatomical magnetic resonance imaging (MR)[6] **.**

1.5.4.Ultrasound

Ultrasound is a non-invasive medical imaging procedure that uses ultrasound waves to transmit through the human body. These waves are emitted by a detector probe called a transducer and pass through tissues to generate echoes based on their density. These echoes are captured by the probe, converted into electricity, and then transformed by software into real-time images of the body's structures.

Figure 1.7: Obstetric ultrasound image [7]**.**

1.5.5.Microscopy

Microscopy imaging refers to the collection of techniques in which structures that enable the observation of structures that can be seen using a microscope.It is broadly used in materials science, biology, and medicine to image tissues, cells, or organelles at high resolution.

Figure 1.8 : Microscopic image[8]**.**

Effective use of a microscope involves not only understanding its optical components but also careful sample preparation, precise adjustment of illumination and focus, as well as regular calibration and maintenance of the instrument. Good practices ensure high-quality images that are reproducible and suitable for both quantitative and qualitative analysis.

1.6. Histological imaging

Histological imaging is a crucial biomedical technique that visualizes the microscopic structure of tissues, allowing researchers and clinicians to identify abnormalities and understand disease mechanisms. By using various staining methods, such as hematoxylin and eosin, histological imaging enhances the contrast between cellular components, making it easier to differentiate between them. This precise examination aids in diagnostics and research, offering insights integral to fields like pathology, oncology, and cell biology [9].

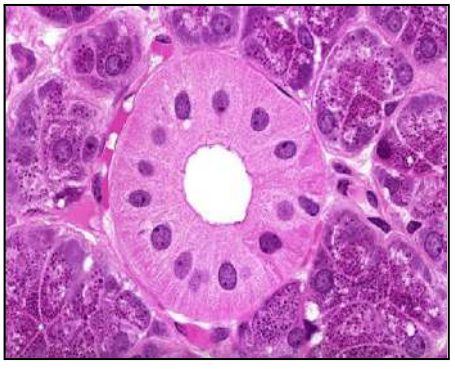


Figure 1.9 : Histological image of a tissue[10]**.**

1.7. Cytology

Cytology is the branch of pathology that involves the microscopic examination of cells to study their structure, composition, roles, and properties. It aims to detect pathological changes at the cellular level by analyzing individual cells and their cellular components. Cytology can be applied to virtually all organs and body fluids and is a versatile tool in medical diagnostics[11].

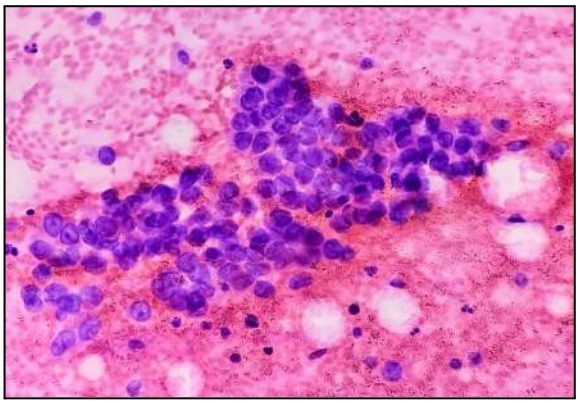


Figure 1.20 : Histological image of a cytology [12]**.**

1.8. Conclusion

Images play a crucial and fundamental role in medicine. They consist of a visual representation of physical structures, obtained through various means, including MRI, ultrasound, X-rays, and microscopy. These techniques facilitate the study, interpretation, and monitoring of the condition of tissues and organs, thus providing an indispensable basis for both diagnosis and therapeutic choice. The following chapter, will focus more specifically on the concepts of microsatellite status (MSI) and microsatellite stability (MSS) to better understand their role in the diagnosis, prognosis, and management of certain cancers.

Chapter 02 : Microsatellite status in Colorectal Cancer

2. Chapter 02 : Microsatellite status in Colorectal Cancer

2.1. Introduction

As we move from the foundational principles of colorectal cancer biology and molecular subtypes, this chapter shifts focus to the practical generation and digitization of histopathological specimens that underpin both routine diagnostics and cutting‑edge computational research. Rather than reiterating epidemiology or risk factors, we will examine each step—from tissue sampling and fixation through microtomy, staining, and high‑resolution whole‑slide scanning—with an eye toward how sample integrity, artifact control, and metadata annotation influence downstream AI‑driven analyses. By framing these technical processes within the context of MSI versus MSS tumor assessment.

2.2. Colorectal Cancer

Colorectal cancer (CRC) is considered a major global health concern in 2020 the statistics shows that there were roughly 1.9 million new CRC cases and 0.9 million CRC deaths , making it the third most common cancer and the second leading cause of cancer death worldwide [13] [14].CRC typically arises in adults over age 50, but alarmingly the incidence in younger adults under 50 years of age is rising in many countries [13] . There is major risk factors that leas to the the alarming rising of CRC , including modifiable lifestyle exposures (heavy alcohol use, smoking, obesity, low physical activity, a diet high in red/processed meats and low in fiber) and non-modifiable factors (older age, family history of CRC, hereditary syndromes such as Lynch syndrome or familial adenomatous polyposis, and chronic inflammatory bowel disease) .As for the prevention methods ,early detection through screening (e.g. colonoscopy every 10 years, fecal blood/DNA tests annually or every few years) is critical , also removal of adenomatous polyps or early cancers greatly improves outcomes [13].

Most CRCs over 90% based on the research are adenocarcinomas arising from the glandular lining of the colon or rectum [15] [16] . A smaller subset approximately 10–20% are mucinous adenocarcinomas (characterized by abundant extracellular mucin) and rare tumors such as signet-ring cell carcinomas ,which account for about 1% [15] . Other rare CRC histologies include squamous cell or neuroendocrine tumors. Standard diagnosis relies on colonoscopic biopsy and histopathology, often supplemented by CT/MRI for local staging. Screening colonoscopy (or fecal tests with follow-up colonoscopy) is recommended beginning around age 45–50 for average-risk adults [17].

2.3. Microsatellite Instability (MSI) and Microsatellite Stability (MSS)

2.3.1.Definitions

Microsatellites are short, repeated DNA sequences scattered throughout the genome. In normal cells,DNA mismatch repair (MMR) proteins(e.g. MLH1, MSH2, MSH6, PMS2) correct replication errors, keeping microsatellites "stable." Microsatellite instability (MSI) is a hypermutable phenotype caused by loss of MMR function[22] [23] .When MMR is defective (due to mutations or epigenetic silencing of MMR genes), insertion–deletion errors accumulate at microsatellite loci, changing their lengths[23] [24].

Based on research, approximately 15% of CRCs are MSI , The remaining approximately 85% of CRCs are microsatellite-stable (MSS), following the classic chromosomal instability (CIN) pathway of tumorigenesis . In MSS tumors the MMR system is intact and mutations occur mainly via chromosomal gains and losses. Thus, MSI-H and MSS CRC represent two major molecular pathways: MSI-H tumors have hypermutation due to MMR loss, whereas MSS tumors rely on CIN and intact mismatch repair

2.3.2. MSI and MSS in Biology

MSI-high tumors harbor an elevated mutation burden (often thousands of mutations) because MMR loss leads to ongoing DNA replication errors . This generates many neoantigens and a strong local immune response: MSI cancers typically show dense tumor-infiltrating lymphocytes and upregulation of immune checkpoint molecules (e.g. PD-1/PD-L1) . By contrast the MSS tumors follow the chromosomal instability (CIN) pathway of carcinogenesis, with more aneuploid karyotypes and fewer neoantigens [20].

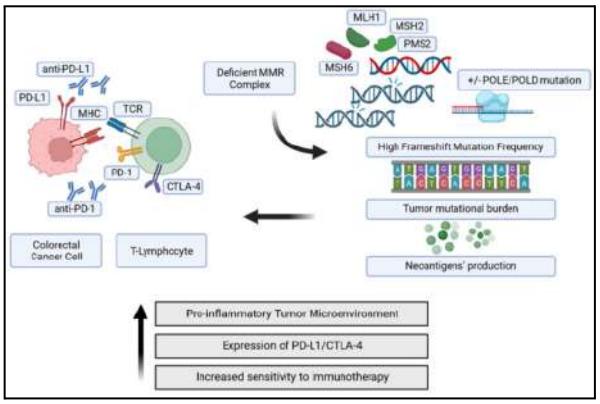


Figure 2.1: Demonstrates the molecular consequences of a malfunctioning mismatch repair system**.**

2.4. Histopathological Overview

2.4.1.Glandular architecture and progression

In the classic adenoma carcinoma sequence, sequential genetic hits drive colonic tumorigenesis. Inactivating mutations in APC initiate adenoma formation from normal mucosa, followed by KRAS mutations for adenoma growth,18q deletions, and finally p53 loss for adenoma to carcinoma transition [13].Consistent with this model,most CRCs (90–95%) are adenocarcinomas that form glandular, tubular structures.Tumor grade reflects the extent of gland formation:well differentiated carcinomas have numerous,well formed glands, whereas poorly differentiated cancers .

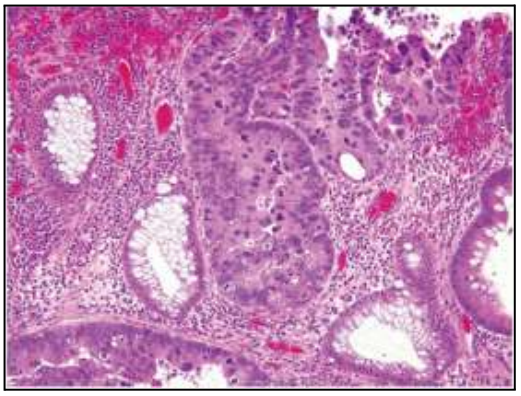


Figure 2.2: High grade dysplasia showing complex architecture and marked nuclear atypia (original magnification ×400) [13].

Recognition of these glandular and nuclear abnormalities is key to identifying lesions that may progress to invasive CRC.

2.4.2.Stromal reaction and inflammatory infiltrate

Invasive colorectal carcinomas elicit a characteristic stromal and inflammatory response.A prominent desmoplastic reaction is often seen at the invasive front: malignant glands are surrounded by proliferating spindle fibroblasts and collagen, forming a dense fibrous stroma . This desmoplasia reflects the host stromal reaction to tumor invasion. Tumor glands frequently contain “dirty necrosis” (necrotic debris) within their lumina ,a classic feature of CRC.The tumor microenvironment also includes inflammatory cells.Many CRCs have scattered lymphocytes, plasma cells, and occasional neutrophils intermixed with the stroma. Tumors with MSI-H tend to have especially dense lymphoid infiltration. Occasionally, expansive lymphoid aggregates form around the tumor. In summary, CRC invasion is commonly accompanied by desmoplastic stroma and variable host inflammation [13].

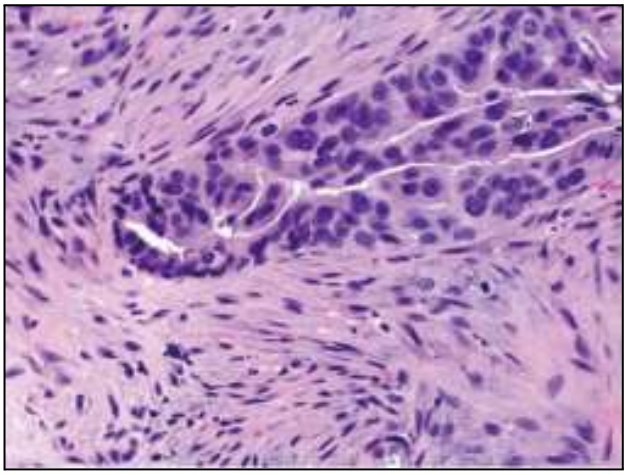


Figure 2.3 : The growth of spindle cells around an adenocarcinomatous gland is a characteristic of a desmoplastic reaction (original magnification ×400) [17].

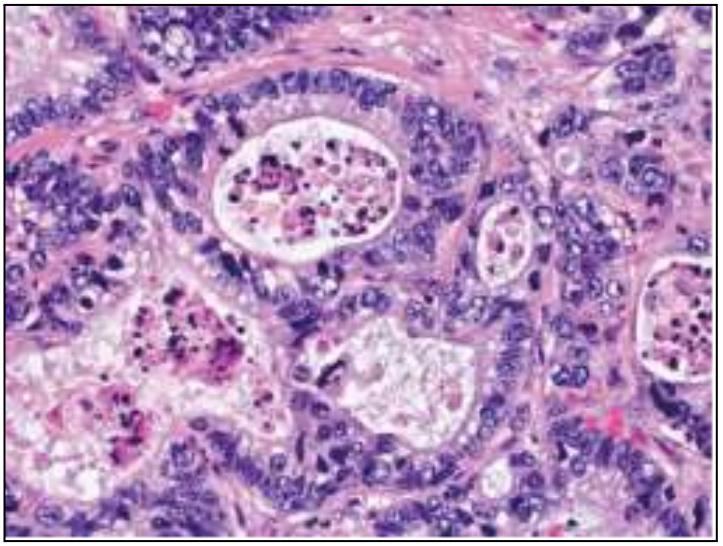


Figure 2.4 : The lumina of adenocarcinomatous glands contains necrotic debris, also known as "dirty necrosis" (original magnification ×400) [13] .

2.4.3.Cellular atypia and mucin production

Neoplastic cells in CRC exhibit cytologic atypia. Malignant glands show enlarged, pleomorphic nuclei with coarse chromatin and prominent nucleoli .As noted, high-grade dysplasia is characterized by back-to-back glands and loss of polarity.

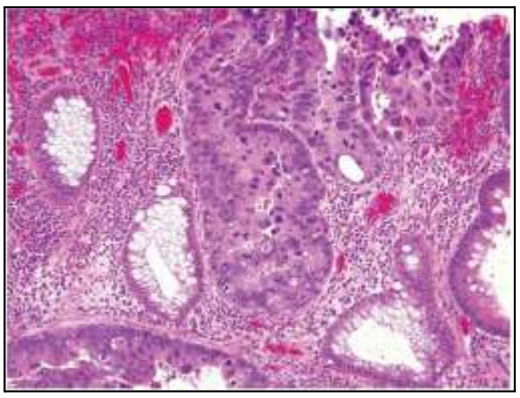


Figure 2.5:High grade dysplasia showing complex architecture and marked nuclear atypia (original magnification ×400) [13] .

Numerous mitotic figures and loss of normal nuclear orientation are also features of high-grade carcinoma. Many CRCs also produce mucin. Mucinous adenocarcinoma constitutes approximately 10–15% of CRCs is defined by the presence of over 50% of the tumor composed of extracellular mucin .Histologically, mucinous tumors show pools of mucin in which tumor cells float. Signet-ring cell carcinoma is a rare form constituting less than 1% of cases, in which over 50% of cells contain large intracytoplasmic mucin vacuoles that displace the nucleus peripherally . By definition signet-ring carcinomas are poorly differentiated and carry a worse prognosis in MSS contexts . However, it will be noted later that many mucinous and some signet-ring CRCs are MSI-H; in those cases the tumors often behave less aggressively.In general, the degree of nuclear atypia, mucin secretion, and cellular differentiation in a CRC informs the tumor grade and can hint at its molecular subtype.

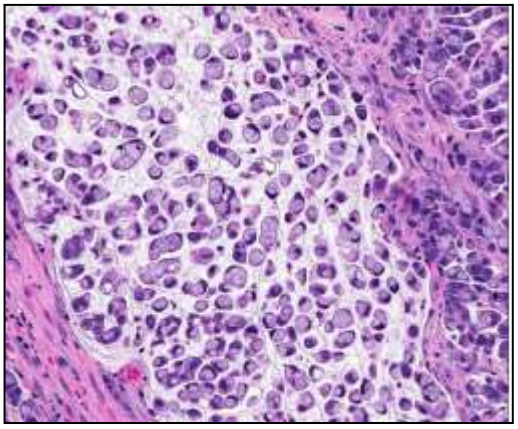


Figure 2.6: Signet ring cell carcinoma (original magnification ×400) [13] .

The detailed histopathological examination of colorectal tumors provides valuable indicators for distinguishing between MSI and MSS phenotypes. MSI-H tumors often display features such as pronounced lymphocytic infiltration, mucinous or signet-ring cell differentiation, and a lack of well-formed glandular architecture, all of which reflect their hypermutated, immunogenic nature. On the other hand, MSS tumors are typically characterized by more structured gland formation, less stromal inflammation, and a predominance of chromosomal instability rather than mismatch repair deficiency. These morphological clues, visible under standard staining and microscopy, guide the pathologist toward suspecting MSI even before molecular testing, playing a crucial role in diagnosis, prognosis, and therapeutic decision making especially in identifying patients who may benefit from immunotherapy.

Microsatellite Instability (MSI) and Microsatellite Stability (MSS), it is now imperative to examine the processes by which histopathological images integral to diagnosis, treatment planning, and biomedical research are generated. MSI status serves as a critical molecular biomarker in CRC, offering prognostic insights and guiding therapeutic decisions, particularly in the selection of immunotherapy [14]. The accuracy and reliability of MSI detection depend not only on molecular assays but also on the quality of histological tissue preparation and the digitization of whole-slide images (WSIs).

This section of the report provides a comprehensive overview of the technical workflow involved in the production of digital histopathology images, starting from the initial patient biopsy to their inclusion in large-scale public datasets. Each step of tissue fixation, processing, embedding, microtomy, staining, and slide scanning is essential for preserving morphological integrity and enabling reproducible analysis [15]. With the advent of whole-slide imaging and digital pathology, these tissue sections can now be captured at high resolution and used for computational analysis and AI-driven research [16].

Understanding this workflow is crucial in the context of precision oncology, where standardized protocols and rigorous quality control directly impact the effectiveness of diagnostic algorithms and deep learning models. Moreover, publicly accessible datasets such as The Cancer Genome Atlas and the Colon Nuclei Identification and Counting (CoNIC) Challenge rely on meticulously curated and annotated histopathology images to facilitate reproducible and scalable cancer research [17]. By detailing each phase of image preparation and highlighting its implications for computational pathology, this

chapter contributes to a deeper understanding of the critical bridge between biological specimens and digital data in modern cancer diagnostics and research.

2.5. The Journey of a Histopathology Image

The generation of histopathological images from patient biopsy to their availability in public datasets is a complex, multi-step process that ensures the preservation, preparation, and visualization of tissue structures for diagnostic and research purposes. This standardized workflow involves the following key stages:

2.5.1.Specimen Receipt and Verification

Upon arrival, the biopsy specimen is formally logged, and the patient's identity is verified to ensure traceability.

2.5.2.Fixation

The tissue is immersed in phosphate-buffered formalin for 6 to 12 hours. This step is essential to preserve the morphological and molecular integrity of the sample.

2.5.3.Grossing

Larger specimens are dissected into slices approximately 4 mm thick. Each slice is described and placed into individually labeled cassettes to facilitate downstream processing.

2.5.4.Tissue Processing

2.5.4.1. Dehydration

Samples are passed through a series of increasing ethanol concentrations to remove water.

2.5.4.2. Clearing

Ethanol is replaced with xylene to prepare tissues for infiltration.

2.5.4.3. Infiltration

Tissues are saturated with molten paraffin wax using an automated tissue processor.

2.5.5.Embedding

The paraffin-infiltrated tissues are oriented and embedded in wax-filled molds. Once cooled and solidified, they are transformed into paraffin blocks suitable for microtomy.

2.5.6.Sectioning

Using a rotary microtome, thin sections (3–5 µm thick) are cut from the paraffin block in the form of ribbons. These sections are floated on a 40 °C water bath to flatten, then carefully mounted on glass slides.

2.5.7.Staining

The slides undergo several steps—dewaxing, rehydration, and staining—typically with Hematoxylin and Eosin (H&E). This reveals key histological features: nuclei appear blue, while cytoplasm and extracellular matrix stain pink. Final dehydration and clearing steps follow.

2.5.8.Coverslipping and Final Quality Checks

A coverslip is placed on the stained section for preservation. Slides are then labeled, inspected for technical artifacts, and finally reviewed by a pathologist for diagnostic interpretation.

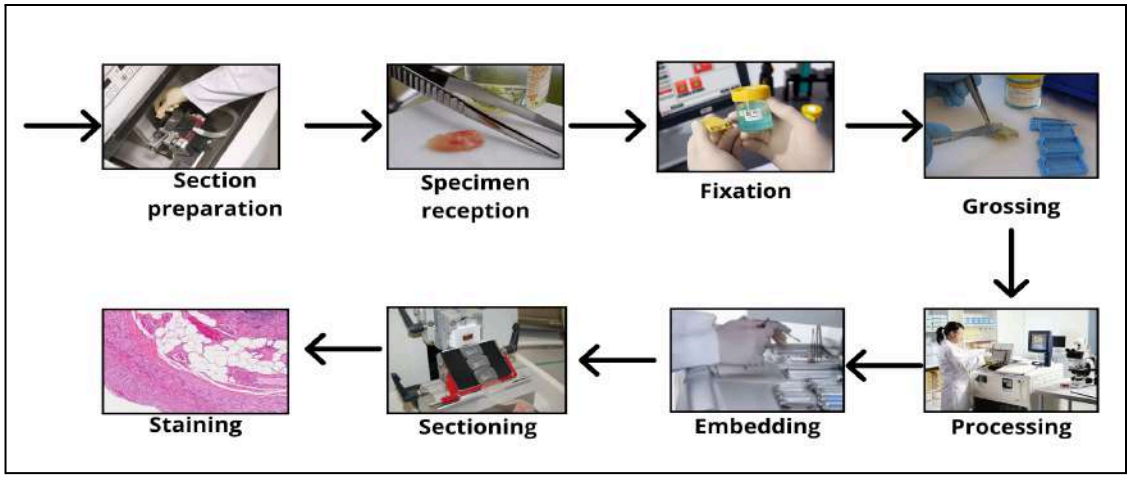


Figure 2.7:Specimen Preparation [18]

2.6. Significance of Digital Histopathology in Modern Research

Digital pathology, underpinned by whole-slide imaging (WSI) technology, has fundamentally transformed histopathological practice by enabling the high-resolution digitization of glass slides and thereby facilitating remote consultation and large-scale collaborative research. This digital framework allows for the seamless integration of advanced computational methodologies most notably artificial intelligence (AI) and deep learning (DL) algorithms which have been successfully deployed for tasks ranging

from precise tumor subtyping and determination of metastatic tissue origin to the prediction of clinically actionable molecular biomarkers. Crucially, the performance and translational potential of these algorithms are inextricably tied to the integrity of the input data: suboptimal image quality or preparation artifacts can confound both human interpretation and machine-derived metrics, while biases or inconsistencies in training datasets may precipitate erroneous or non-generalizable predictions. Thus, rigorous standardization of specimen collection, processing, and digital acquisition is imperative to ensure that AI-driven analyses yield robust, unbiased insights, thereby paving the way for their reliable integration into routine clinical workflows.[20]

2.7. Digital Transformation: Whole-Slide Imaging (WSI)

Digital transformation in pathology is driven by whole-slide imaging (WSI), where high-resolution scanners capture and stitch thousands of microscopic fields into seamless, multi-resolution digital slides. Scanner platforms range from compact, single-slide "box" systems to high-throughput, microscope-based devices (e.g., Leica Aperio AT2, Hamamatsu NanoZoomer) capable of automated barcoding, staining, and file management, though variability in hardware and acquisition protocols across institutions such as in TCGA introduces significant image heterogeneity.

Common artifacts out-of-focus regions, tissue folds, ink or dust specks, and air bubbles necessitate rigorous QC, increasingly achieved through AI-driven tools that score slide quality, highlight problematic regions, and streamline manual review. Equally vital is comprehensive metadata and molecular annotation: public repositories like TCGA and GDC link each WSI to clinical, pathological, and biomarker data (e.g., MSI status), while specialized annotation software supports detailed labeling (bounding boxes, polygons, hierarchical classes). Together, these advances robust digitization, automated QC, and rich metadata form the backbone of scalable, AI-enabled digital pathology and precision oncology.

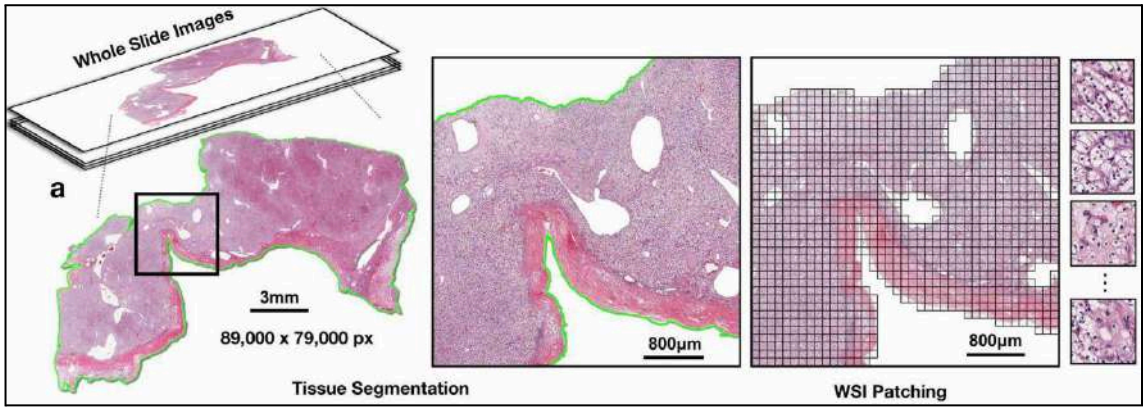


Figure 2.8: Demonstration of the WSI tiles [19]

2.8. Conclusion

This chapter provided an in-depth overview of microsatellite instability (MSI) and microsatellite stability (MSS) in colorectal cancer, highlighting their distinct biological origins and histopathological presentations. MSI-H tumors, which result from mismatch repair deficiency, are often identified by features such as lymphocytic infiltration, mucin production, and poor glandular organization. In contrast, MSS tumors follow the chromosomal instability pathway, typically exhibiting more structured glandular architecture and lower immune cell density. These morphological patterns offer important diagnostic clues that pathologists can observe using routine staining techniques.

Beyond traditional diagnostics, the digitization of histopathological slides through whole-slide imaging (WSI) has opened new possibilities for computational analysis. Standardized image preparation and high-resolution scanning now enable the application of artificial intelligence (AI), particularly deep learning, to detect subtle patterns that may not be readily apparent to the human eye. By learning from large datasets, AI systems can automatically identify MSI-related features, assist in early detection, and provide consistent, scalable, and cost-effective decision support. This integration of AI into digital pathology represents a transformative step toward precision oncology.

Chapter 03 : Deep learning

3. Chapter 03 : Deep learning

3.1. Introduction

Image classification and segmentation are key areas of computer vision and artificial intelligence. These disciplines play a crucial role in enabling machines to interpret and make sense of the rich visual content within images. By accurately identifying, classifying, and delineating different components or regions of an image, computer algorithms can facilitate a range of tasks, from automated diagnostics in healthcare and defect detection in industry, to financial analysis and surveillance.With the dramatic increase in available data and computing power in recent years, deep learning has become a powerful and efficient approach for addressing these problems. Drawing inspiration from the human brain's ability to learn and adapt, deep neural networks can automatically extract relevant features and patterns from large volumes of images, without needing extensive manual intervention. This ability to learn directly from raw data has led to significant improvements in both speed and accuracy, outperforming traditional methods in many applications .

3.2. Artificial intelligence:

Artificial intelligence (AI) refers to systems that exhibit intelligent behavior by analyzing their environment and taking actions with some degree of autonomy to achieve specific goals.

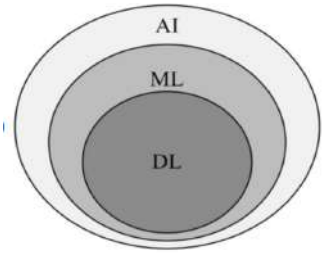


Figure 3.1: Relationship between AI ,ML and DL.

3.3. Deep learning :

Deep learning involves using neural networks formed by cascaded layers to automatically learn relevant characteristics from vast amounts of raw data (images, audio, text).The traditional machine learning, where attribute extraction requires expert intervention, deep learning handles this task itself.This, the model autonomously learns the most interesting representations, making it particularly suited to processing large volumes of data, particularly in medical imaging [26].

3.4. Machine learning :

Machine learning is a field of artificial intelligence that enables computers to automatically learn from data and improve their performance on specific tasks without being explicitly programmed[26].

3.5. Classification and Segmentation in deep learning

Deep learning applied to many image analysis tasks ,among the most commonly used are classification and segmentation.

3.5.1. Classification

Image classification is the process of automatically assigning a label to a given image using a learning model trained on a set of labeled images that uses deep learning methods to discriminate representations of visual data and generate accurate predictions on new images.

3.5.1.1. Type of classification

- **Binary classification**

Binary classification is the task of classifying the elements of a given set into two predefined classes. It involves training a model to distinguish between two categories typically labeled as 0 and 1 based on the patterns present in the input data.

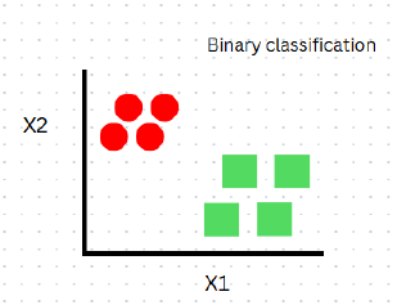


Figure 3.2: Binary classification explainable graph

- **Classification multiclass**

It refers to the task of assigning a set of two or more labels to each instance in a data set. That means a single observation can simultaneously belong to multiple classes, reflecting the overlapping nature of many real-world problems.

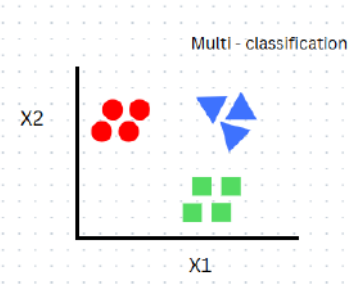


Figure 3.3:Multi - classification explainable graph

- **Classification multilabel**

The Multi-Label Image Classification focuses on predicting labels for images in a multi-class classification problem where each image may belong to more than one class.

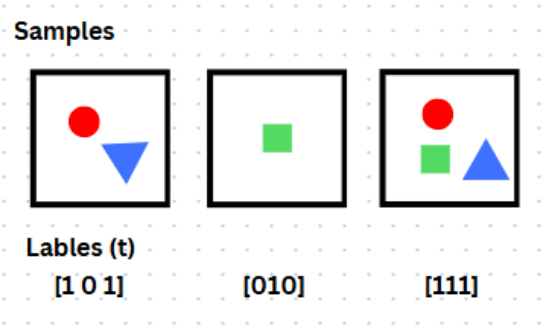


Figure 3.4: Multi - classification multilabel [30]

3.5.2. Deep learning models for classification :

- **Convolutional Neural Network CNN**

Convolutional Neural Network CNN is a type of deep neural network primarily designed for image processing. The first part of this network applies a series of operations in order to extract relevant information from the image. These operations include convolution , activation ,and pooling . The first conventional layer detects simple patterns ,while subsequent layers identify increasingly complex structures by building upon the features detected by previous layers ,and the Fully connected layer combines all these features to classify the image based on its content.

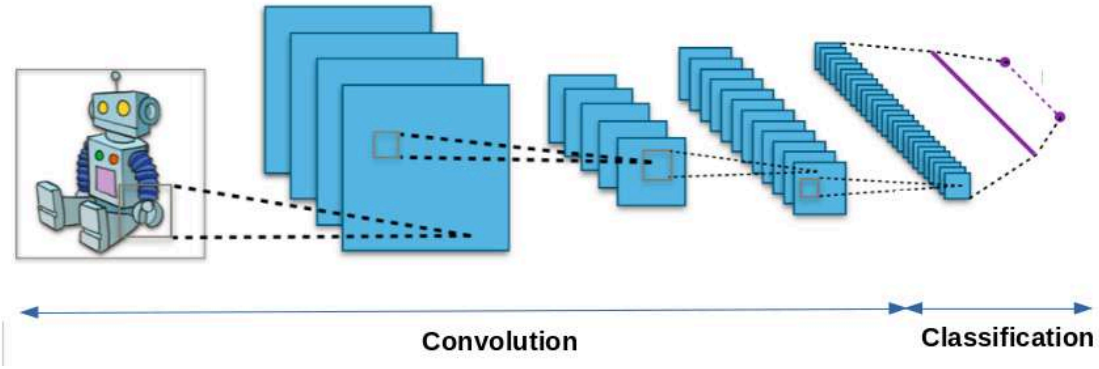


Figure 3.5: CNN schematic representation [31]

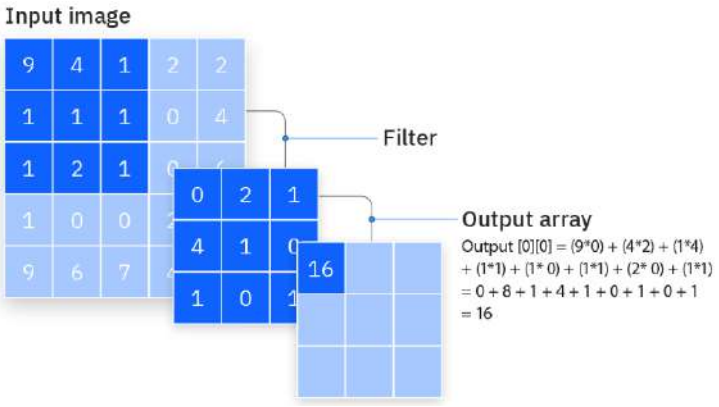


Figure 3.6: Convolution Operation in a Convolutional Neural Network [32]

- **Convolutional Layer :**Extract local features from the image by applying sliding filters .Each filter detects .
- **Activation Function (ReLU):**An activation function introduces nonlinearity into a neural network, allowing complex relationships to be learned .

  ReLU, the most widely used, transforms negative values into zero and keeps positive values, which makes learning easier and faster.

- **Pooling LayerReduces:**The size of the data while preserving essential information. Max pooling selects the maximum value in a region.
- **Flattening and Fully Connected Layers:**Transforms the feature maps into a vector, then combines the extracted information to perform classification.

- **Output Layer :**Provides the final result (predicted class) using a function such as Softmax to estimate the probability of each class.

Once the concept of image classification and its main types have been defined, it is important to emphasize that many deep learning models are most popular

- **YOLO (You Only Look Once)**

YOLO (You Only Look Once) is a real time object detection system that is particularly fast because it takes a unified approach rather than generating region proposals in multiple stages. Unlike traditional two-phase methods, YOLO applies a single convolutional neural network to simultaneously classify objects and predict their bounding boxes. This architecture gives it great speedup to 45 frames in second, and up to 155 fps with the fast version while maintaining an accuracy close to that of more complex detection methods.[32]

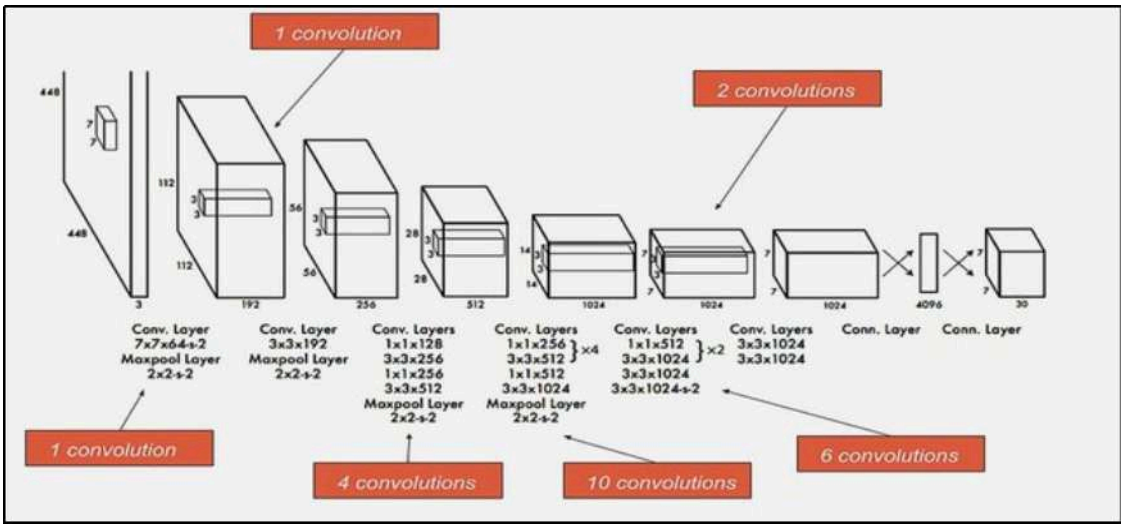


Figure 3.7 :YOLO full model architecture[32]

- **Yolo version :**

YOLO has gone through numerous versions, from YOLO, YOLOv2, YOLOv3, YOLOv4, YOLOv5, YOLOv6, YOLOv7, YOLOv8, YOLOv9, YOLOv10, YOLOv11, to YOLOv12. Each new version brings improvements over its predecessor in terms of speed, accuracy, and functionality. Additionally, for each version of YOLO, there are different variants nano, small, medium, and large ,which vary in size and complexity depending on the number of parameters in the model [33].

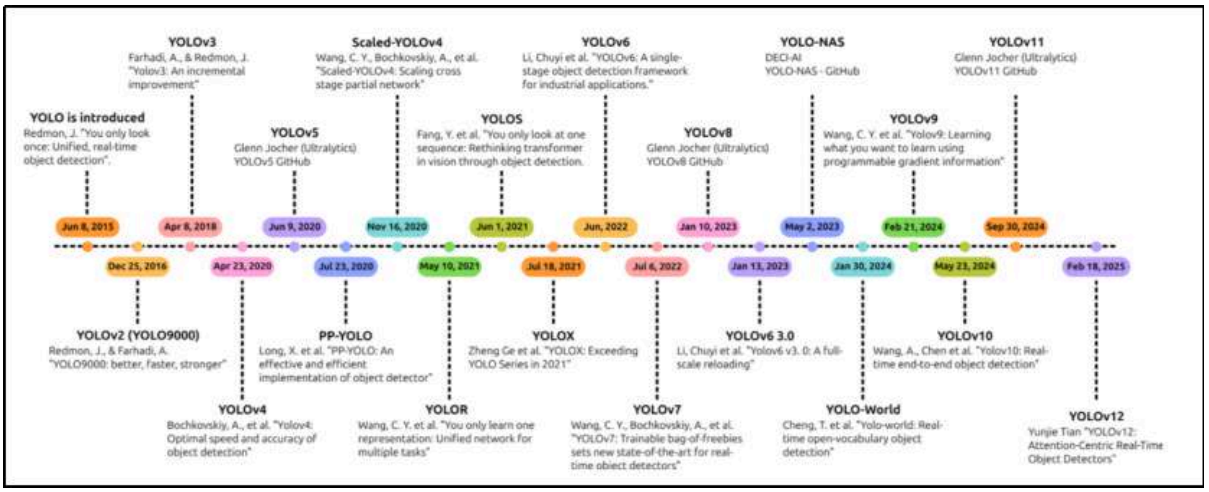


Figure 3.8: Evolution of YOLO algorithms throughout the years[33]

3.6. Preview of the used models :

3.6.1.YOLO V8m (medium):

This model strikes a good balance between abstraction capacity and lightweight design.It was chosen to have a sufficiently powerful architecture while remaining fast and memory-efficient.

3.6.2.YOLO v8l (Large):

The YOLOv8l model provides a strong balance between accuracy and performance. With its increased number of parameters compared to smaller versions, it can capture complex patterns in histopathological images more effectively. However, it comes with a trade-off of higher computational cost, including longer training and inference times, as well as increased GPU memory usage.

3.6.3.YOLO 11l (large):

This larger model offers enhanced abstraction capacity, allowing it to better adapt to the complexity of histological data.However, this choice implies a longer inference time and a greater memory requirement.

3.7. SEGMENTATION

3.7.1.Image Segmentation

Image segmentation is a crucial step in image processing that involves dividing an image into several homogeneous regions or objects with similar characteristics (such as

color, texture, or shape). It helps to identify and isolate important parts of an image, such as cells in a microscopic image or objects in a scene.

3.7.2.Type of Segmentation

- **Semantic segmentation**

Semantic segmentation involves assigning a class label to each pixel in an image. All objects belonging to the same category are grouped under a single label, without distinguishing between individual instances. This process generates a map of regions corresponding to the different classes present in the image, without identifying separate instances.

- **Instance segmentation**

Instance segmentation is a computer vision technique that identifies and segments each object individually within an image. It distinguishes between multiple objects belonging to the same class.

- **Panoptique segmentation**

Panoptic segmentation is a developed image segmentation that relies on both semantic segmentation class of each pixel and instance segmentation of each of the objects separately identifiable to label each pixel of an image by indicating both its class and the specific object with which it is associated.

3.7.3.The Segmentation models

A. Unet:

U-Net is a convolutional neural network model known for its ability to capture contextual information at different scales while preserving significant spatial details. This architecture is particularly suited for semantic segmentation of images. The U-Net architecture consists of two main paths: the Contracting Path (Encoder) and the Expansion Path (Decoder), which reduce and then restore the image resolution while integrating contextual information at each step, giving it a U-shaped architecture.

- **An encoder :**This represents the left side of the U-Net architecture. Its role to capture image features at different scales using layers of:
- **Convolutions:** Each encoder step applies two 3x3 convolutions followed by an activation function such asReLU

- **Max-Pooling:** A 2x2 max-pooling operation is used to halve the image resolution and extract higher-level features.
- **A decoder :** Is the part of a convolutional neural network responsible for reconstructing the original resolution of the input from the compressed, abstract representations produced by the encoder.
- **Up-sampling:** Each decoder step begins with an up-sampling operation (such as a convolution transpose) to enlarge the spatial representation.
- **Skip Connection:** The up-sampled representation is concatenated with the corresponding encoder representation. This connection, called a skip connection, allows for the combination of low-level details with high-level features to improve segmentation accuracy.
- **Convolutions:** This concatenation is followed by two 3x3 convolution layers to refine theOutput layer .
- **The last layer:** is usually a 1x1 convolution, used to project the number of feature channels onto the number of classes, generating the final segmentation map.
- **ReLU** (Rectified Linear Unit) is the most commonly used activation function in deep learning.

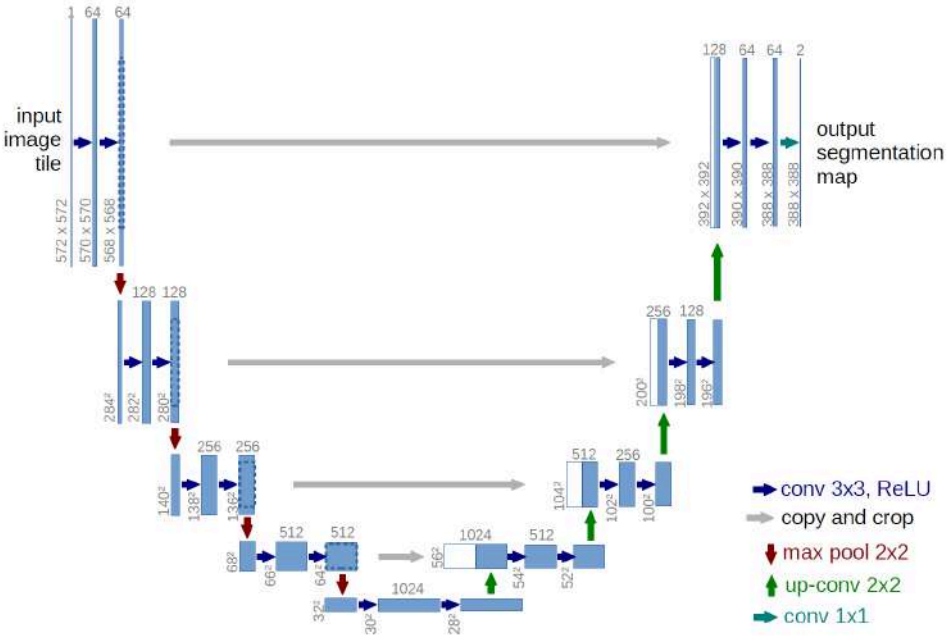


Figure 3.9 : Architecture U-net[35]

B. TransUNet :

TransUNet is a hybrid TransformerUNet architecture for medical image segmentation. It uses a Transformer encoder to extract global context from CNNderived image patches, followed by a UNet decoder for high resolution spatial detail reconstruction.[36]

- **CNN Encoder:** A conventional neural network is used to extract multiscale feature representations from the medical image
- **Transformer Encoder :** Patches taken from the encoder are fed into a stack of Transformer Layers. This process enables the network to learn global relationships and context across different regions of the image.
- **Segmentation Head:** Finally, a 1x1 convolution produces a pixel-wise segmentation map with a class label for each pixel
- **U-Net Decoder:** The decoder performs up-sampling and concatenates the corresponding timescale CNN features through skip connections, thereby helping to recover fine details in the segmentation.
- **Segmentation Head:** Finally, a 1x1 convolution produces a pixel-wise segmentation map with a class label for each pixel

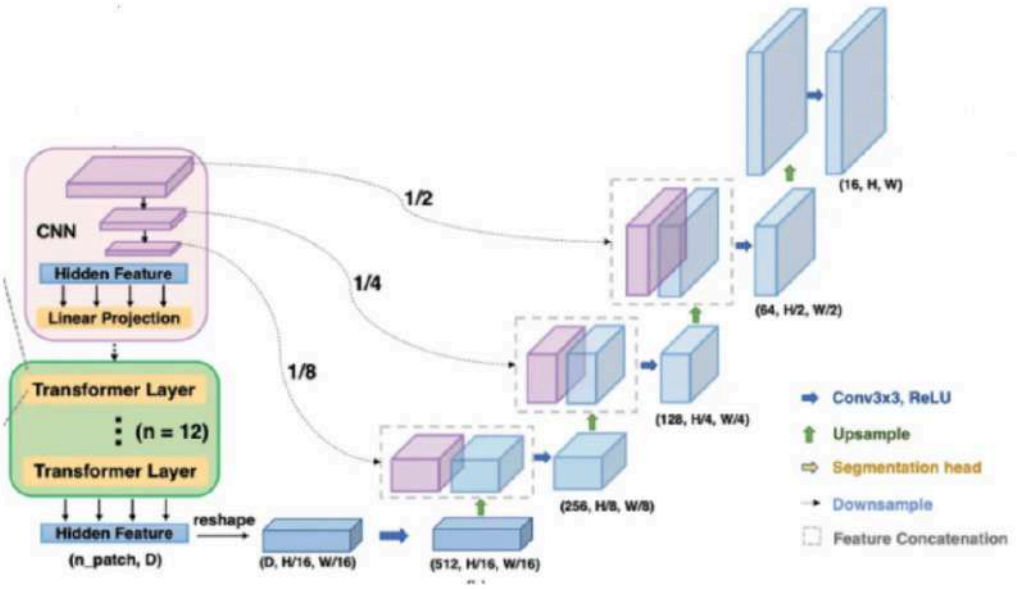


Figure 3.10 : TransUNet architecture design overview. MSA stands for Multi-head Self-Attention, and MLP stands for Multi-Layer Perceptron[36]

C. SegFormer:

SegFormer was introduced by Xie et al. in 2021 with competitive state-of-the-art results on benchmark datasets ADE20K and Cityscapes. The general idea is to use a hierarchical Transformer

- **Encoder** (the Mix Transformer, or MiT) to produce multi-scale feature maps and decode it with an MLP head. Above all, SegFormer avoids using explicit positional encodings and avoids complex
- **Decoders**. Instead, a convolution augmented feed forward (mix-FFN) in every transformer block provides spatial information.

As a result, SegFormer integrates global attention from theTransformer with sufficient local context from convolution to create strong representations .

- **SegFormer Architecture**

SegFormer's architecture consists of two main modules: a hierarchical Mix Transformer(MiT),an encoder and a lightweight All-MLP decoder.

The encoder operates on the input image at different resolutions, as the decoder aggregates all these multi-level features to predict the segmentation mask. Figure illustrates the design in a 3D realization; the 2D SegFormer is identicalPrinciples.SegFormer architecture has two general components: a hierarchical Mix Transformer (MiT) an encoder, and an optimized All-MLP decoder.

- **The encoder :** operates on the input image at different resolutions, as the decoder fuses the multi-level features to predict the segmentation masks; the figure is displayed in a 3D representation; the same 2D SegFormerTenets.

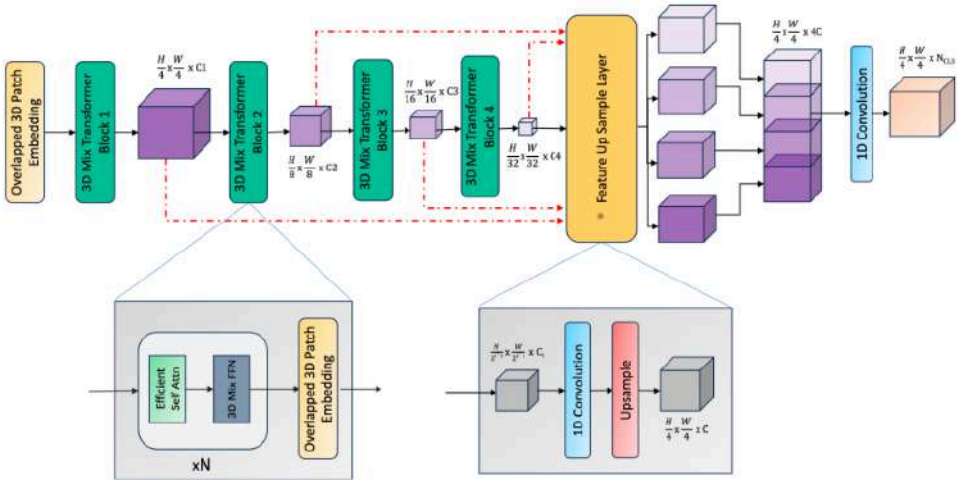


Figure 3.11: SegFormer3D model architecture[37]

Hierarchical Mix Transformer (MiT) Encoder

- **Hierarchical Design:** The input image is processed across four stages, producing feature maps at multiple scales: 1/4, 1/8, 1/16, and 1/32 of the original resolution. This design captures both local details and global context.
- **Overlapping Patch Embedding:** Unlike traditional Vision Transformers (ViT) which use non-overlapping patches, MiT uses overlapping patches via a 3×3 convolution with stride 4. This reduces block artifacts and preserves spatial continuity.
- **No Fixed Positional Encoding:** Instead of using fixed positional embeddings, MiT relies on implicit spatial encoding through a module called Mix-FFN, which combines multi-layer perceptrons (MLPs) with 3×3 depthwise convolutions.
- **Efficient Attention Mechanism:** Inspired by PVT, MiT applies reduced self-attention, where keys and values are downsampled before computing attention. This significantly improves speed with minimal accuracy loss.
- **Output:** The encoder generates four feature maps F1,F2,F3,F4{F_1, F_2, F_3, F_4}F1,F2,F3,F4 with increasing semantic depth and decreasing spatial resolution.

All-MLP Decoder

- **Simple and Lightweight:**Each encoder output FiF_iFi is projected into a common dimension using a linear layer (MLP).All features are upsampled to a unified resolution (typically 1/4 of the input size).These upsampled features are concatenated along the channel axis and fused using another MLP.A final linear layer maps this fused representation to the segmentation mask with NclsN{cls}Ncls channels (one per class).
- **Output Resolution:** The final prediction is produced at H/4×W/4H/4 \times W/4H/4×W/4 resolution, balancing performance and efficiency.

3.8. States of the art :

Over a long period of time, researchers have been striving to leverage technological advancements to improve medical diagnostics, particularly in the field of cancer. Gastrointestinal cancers, especially colorectal cancer, pose complex challenges related to their detection and classification. An essential component in this context is microsatellite instability (MSI), which serves as a significant biomarker for predicting response to immunotherapy.

The articles in this study focus on the use of deep learning models to predict MSI status directly from histopathological images. These works aim to develop effective and cost-efficient methods to enhance clinical practices. Advanced models such as ResNet and ShuffleNet are among the techniques employed. These solutions incorporate data from various cohorts, including the TCGA, to ensure comprehensive validation.

This table presents the state of the art of recent methods for predicting MSI and MSS status in colorectal cancer using histopathological images and deep learning models**.**

Title	Year	Authors	Model/Method	Input Data	Accuracy
Predicting colorectal cancer MSI with a self-attention CNN[57]	2023	Chang et al.	ResNet-50 + self-attention (WiseMSI)	H&E whole-slide images	91.1%
PPsNet: MSI-H prediction in CRC WSIs [58]	2022	Lou et al.	PPsNet (deep supervision, CNN)	H&E whole-slide images	87.3%
DL-based MSI classification (feasibility) [59]	2021	Lee et al.	CNN (sequential WSI classifier)	H&E WSIs (TCGA)	89.2% (AUC)
Weakly-supervised MSI prediction [60]	2021	Bilal et al.	ResNet-34 (MIL framework)	H&E WSIs (TCGA-CRC-DX)	86.0%
Swin Transformer for CRC biomarkers (MSI) [61]	2023	Guo et al.	Swin Transformer	H&E tiles	90% (AUC)
Differentially Private Fine-tuned NF-Net to Predict GI Cancer Type [62]	2025	S.V. Chilukoti et al.	NF-Net (Normalizer Free Net)	H&E WSIs (TCGA-CRC-DX)	91.1

Table 3.1: Review of representative studies on MSI versus MSS classification methods

The growing application of deep learning to digital pathology has enabled significant advancements in the classification of microsatellite instability (MSI) versus microsatellite stable (MSS) colorectal cancer from hematoxylin and eosin (H&E) stained whole-slide

images. Recent models applied to the TCGA dataset have achieved accuracies between 86% and 91.1%, reflecting the potential of computational pathology.

In this study, a novel deep learning classification approach was developed and applied to the TCGA MSI/MSS dataset , which includes a large number of H&E image tiles. The model was designed to efficiently handle image-based patterns relevant to MSI status while maintaining low computational complexity. Unlike traditional patch classifiers or transformer-based models, this approach draws on real-time classification principles, enabling fast and scalable analysis.

The model achieved competitive performance, aligning with state-of-the-art results in the literature. Its design emphasizes both speed and classification accuracy, providing a promising alternative for MSI prediction using histopathological data. These results reinforce the adaptability of deep learning methods to diagnostic pathology and open pathways for future extensions incorporating multi-scale features or molecular data integration.

3.9. Conclusion :

In this chapter, we explored the foundations and evolution of deep learning in the context of image analysis, with a specific focus on classification and segmentation tasks. Deep learning has revolutionized computer vision by enabling models to learn complex hierarchical representations directly from raw data, outperforming traditional machine learning techniques in both accuracy and generalization.The next chapter presents the implementation of the classification and segmentation models used in this work, while the following chapter will be dedicated to the discussion and analysis of the results obtained from these models.

Chapter 04 : Implementation

4. Chapter 4 Implementation

4.1. Introduction

Histopathological image analysis provides fundamental support in the diagnosis and understanding of various types of cancer, including colorectal cancer. This analysis relies on careful observation of tissue. This status is particularly relevant in patient stratification,particularly to guide therapeutic decisions such as the use of immunotherapy.This,our study proposes a system for automatic prediction of MS vs MSS status from histopathological images. This system is capable of classifying MSI vs MSS status in CRC, and a segmentation system has been added for visualizing nuclear components. The approach is based on pathologists' diagnosis.

4.2. Used Tools :

4.2.1.Lightning AI

Lightning AI is a free, open-source, modular platform for creating end-to-end AI applications, or Lightning Apps, that combine user interfaces, deployment, and training into a single, cohesive system. From research to production, the machine learning lifecycle is intended to be made simpler and more scalable.

Figure 4.1: Logo lightning AI [39]

- All models were trained using an NVIDIA L4 GPU, which enabled accelerated computation and significantly reduced training time due to its high parallel processing capabilities
- In our project, we used two different machines. The first system was DESKTOP-VQB03E, equipped with an Intel Core i5-8350U CPU and 16 GB of RAM. The second device was a Lenovo ThinkPad L470, featuring an Intel Core i5-6300U processor and 4 GB of RAM.

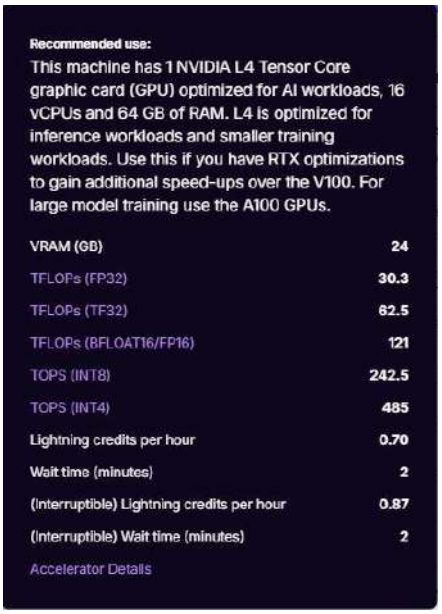


Figure 4.2 : The L4 GPU features in Lightning Ai platform

4.2.2.Google Colab:

Google Colab is a serverless Jupyter Notebook environment that enables interactive development and prototyping of machine learning models using free access to GPUs and TPUs, seamlessly integrating with Google Cloud services for data storage and collaboration (Bisong, 2019).

Figure 4.3: Google colab logo [40]

4.2.3.Ultralytics:

Ultralytics is a company specialized in the development of artificial intelligence solutions, best known for creating and maintaining the YOLO (You Only Look Once) series of models, particularly YOLOv5 and YOLOv8, designed for real-time object detection tasks [41].

Figure 4.4: Ultralytics logo [41]

4.2.4.python:

Python its a langage de programmation informatique ,open source, multi-plateforme and object oriented programming known for its simplicity and legibility .Moreover Python is the most popular and widely used computer programming language , especially in the fields of data science and machine learning , it can be used for any type of programming and software development . Our project is made with the Python language in the cloud and we can build our models without the need to install any software or IDE in our computer.

Figure 4.5:Python logo [40]

4.3. Libraries

4.3.1.PyTroche :

Pytroche is a library for Python programs that facilitates building deep learning projects developed by meta . It emphasizes flexibility and allows deep learning models to be expressed in idiomatic Python. This approachability and ease of use found early adopters in the research community, and in the years since its first release, it has grown into one of the most prominent deep learning tools across a broad range of applications.

Figure 4.6: Logo pytorch [43]

4.3.2.Albumentation :

Albumentations is a powerful and flexible Python library for applying visual augmentations to images and their annotations. It aims to enrich datasets by transforming training data efficiently, simply, and compatible with deep learning workflows.

Figure 4.7: Logo albumentation [43]

4.3.3.Numpy:

It is an open source library related to the Python language ,it provides a multidimensional array object, various derived objects such as masked arrays and matrices, and an assortment of routines for fast operations on arrays, including mathematical, logical, shape manipulation, sorting, selecting,discrete transforms, linear algebra,statistical operations, random simulation and much more

Figure 4.8: Logo numpy [45]

4.3.4.Matplotlib:

Matplotlib is a Python data visualization library, inspired by Matlab and developed by John Hunter in the 2000s. The neurobiologist was looking for a way to visualize the electrical signals in the brains of epileptic patients. In 2012, following the death of its creator, many contributors contributed to this open source project.

Figure 4.9: Logo matplotlib[46]

4.3.5.Tqdm:

TQDM is a lightweight, fast, flexible, and extremely easy-to-use Python library, primarily used to display progress bars in Jupyter notebooks.It consists of a wrapper that can be applied to any iterator to make progress tracking more intuitive, especially when the program is time-consuming such as training deep learning models on large datasets, downloading files, processing images.[47]

Figure 4.10: Tqdm progress bar [47].

4.3.6.Scikit-learn:

Scikit-learn is an open-source Python machine learning library that allows users to perform regression, classification, clustering, dimensionality reduction, model selection, and preprocessing. Scikit-learn is built on NumPy, SciPy, and Matplotlib, making it fast, efficient, easy to use, and ideal for both research and production use.

Figure 4.11: Scikit-learn[48]

4.3.7.Tensor Flow :

Created by the Google Brain team in 2011, Tensorflow was a renowned numerical computation programming framework that was open-sourced by Google in 2015. Its name is inspired in particular by the fact that the current operations of neural networks are performed mainly using a multidimensional data table called a tensor. Tensorflow is a machine learning library. It is a toolbox that allows researchers to easily solve extremely complex mathematical problems. It allows them to develop experimental learning architectures and transform them into software [49].

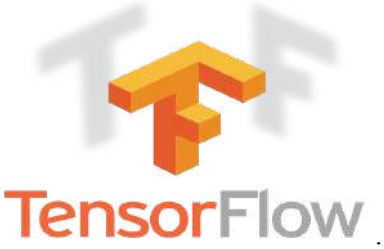


Figure 4.12: Logo TensorFlow[50]

4.3.8.Keras :

Keras is a deep learning library designed to build models quickly and easily. It connects layers of neurons like Lego, making model development and testing easier. The Sequential API allows you to create standard models in just a few lines of code, while the Functional API and Model and Layer classes allow you to build more complex, custom architectures. This simplicity increases developer efficiency, giving them more time to focus on improving algorithms rather than the code itself [51].

Figure 4.13: Logo Keras[52

4.3.9.Gradio :

Gradio is an open-source Python library designed to make the quick design of simple, interactive, and custom user interfaces for trying out machine learning models, API, or other capabilities easier. It facilitates the interaction with the model by downloading images, text, or audio and viewing the generated result instantly. Gradio can be embedded directly in a Python notebook or generate a URL to make the application accessible online [53].

Figure 4.14:Gradio Logo[54]

4.4. Dataset used :

4.4.1.TCGA-COAD MSI/MSS Dataset:

The TCGA-COAD MSI vsMSS dataset is a publicly available collection of hematoxylin and eosin (H&E) stained histopathology image tiles extracted from the TCGA(The Cancer Genome Atlas) Colon Adenocarcinoma (COAD) cohort.It includes approximately 192,312 image tiles, each labeled as either microsatellite instability-high (MSI-H) or microsatellite stable (MSS), based on the tumor's genetic profile.This dataset is commonly used for building classification models that predict MSI status directly from histology images.MSI status plays a critical role in colorectal cancer prognosis and treatment decisions, especially regarding immunotherapy.The TCGA-COAD dataset provides a valuable resource for research in computational pathology and biomarker discovery using artificial intelligence techniques [55]

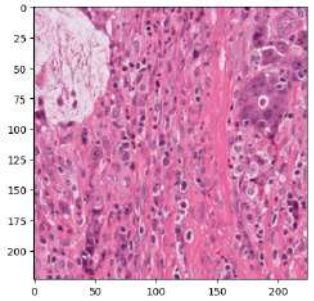


Figure 4.15: Data set visualization

The dataset is already well-structured and optimally split for deep learning model development. It consists of 153,849 images for training (representing 80% of the data), 19,320 images for validation (10%), and 19,233 images for testing (10%). This distribution ensures effective model training while facilitating performance evaluation. The dataset is intended for a binary classification task, divided into two classes, which simplifies the implementation of classification algorithms.

```
train: /teamspace/studios/this_studio/dataset/tcga_coad_msi_mss/train... found 153849 images in 2 cla
val: /teamspace/studios/this_studio/dataset/tcga_coad_msi_mss/val... found 19230 images in 2 classes
test: /teamspace/studios/this_studio/dataset/tcga_coad_msi_mss/test... found 19233 images in 2 classe
```

Figure 4.16: Data set visualization devision

4.4.2.CoNIC Dataset:

Classification into six cell types: epithelial, lymphocyte, neutrophil, eosinophil, plasma, and connective tissue.In addition to the segmentation masks, the dataset includes counts of each cell type per image, supporting quantitative analysis of the tumor microenvironment. The CoNIC dataset was introduced as part of the CoNIC Challenge at IEEE ISBI 2022 and is widely used for benchmarking segmentation models in computational pathology research [56].

```
Images shape: (4981, 256, 256, 3)
Labels shape: (4981, 256, 256)
```

Figure 4.17 : The CoNIC Dataset images and labels

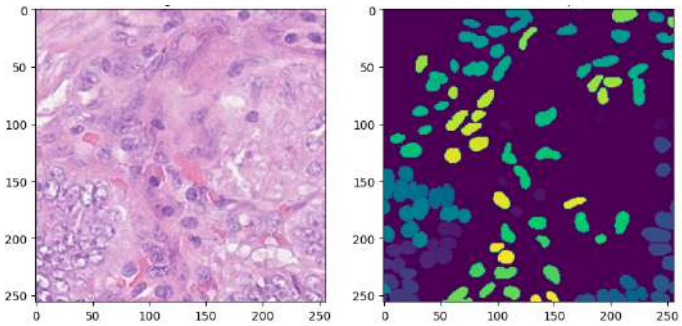


Figure 4.18 : Visualisation of the CoNIC dataset

To accomplish binary classification and localisation on histopathology images from The Cancer Genome Atlas (TCGA), we modified YOLO, which was initially created for multi-class object detection, we have tried multiple YOLO version (YOLOv8m , YOLOv8l, YOLOv11l).

4.5. Model Adaptations and Modifications

In order to modify YOLO for binary classification of histopathology patches, several important changes were made:

To simplify training and eliminate the need for manual anchor box tuning specific to histological structures, the model was switched to an anchor-free configuration. The input resolution was set to 512×512 pixels to align with the patch size, requiring careful adjustments to the Feature Pyramid Network (FPN) scales and interpolation layers in the early stages of the backbone. For training, Binary Cross-Entropy with Logits was employed as the classification loss, while an auxiliary localization task was introduced by predicting pseudo-bounding boxes around nuclei clusters. To mitigate overfitting, particularly due to the sparsity of the data, DropBlock regularization was applied within the final two CSP (Cross Stage Partial) modules.

- **Head Redesign:**
  The original model's detection head was replaced with a binary classification head.We removed the multi-class classification branch and substituted it with a single sigmoid output neuron, representing MSI status probability.
- **Anchor-Free Mode:**
  The YOLO supports anchor-based detection; we switched to its anchor-free variant to avoid manual tuning of anchor boxes for histological structures.

- **Resolution Adjustment:**
  To match the patch size, the input resolution was 512×512 pixels, necessitating the careful tuning of Feature Pyramid Network (FPN) scales and interpolation layers in the first stage of the backbone.

- **Loss Functions:**
  Binary Cross-Entropy with Logits: Classification Loss.
  Localization Loss: For localisation regularisation, the CIoU loss was reformulated on pseudo-bounding boxes enclosing nuclei clusters ("localisation auxiliary task").

- **Regularization:**
  To avoid overfitting on the sparse data, DropBlock regularisation was applied in the last two CSP modules.

4.6. Training Protocol and Hyperparameters:

The training protocol combined fine-tuning and transfer learning, leveraging YOLO weights pretrained on the COCO dataset as initialization for the model.
The optimization was carried out using AdamW with a weight decay of 0.0005. A one-cycle learning rate policy was employed over 50 epochs, with the learning rate peaking at 1e-3 and gradually decreasing to a final value of 1e-6. Training was conducted on two GPUs simultaneously, with a batch size of 16 patches per GPU. Early stopping was implemented with a patience of 10 epochs, based on the monitoring of the validation AUC-ROC score. Final evaluation metrics were computed on a held-out test

set comprising 2,000 image patches, equally balanced between MSI and MSS cases (1,000 A detailed explanation of the classification model employed.each).

4.7. Model Evaluation and Selection:

4.7.1. Definition of Evaluation Metrics for Classification Performance :

- **Top-1 accuracy:** Rate at which the model's single highest-scored prediction is correct.
- **Top-5 accuracy: r**efers to the proportion of times the correct class label is found among the model's five highest-probability predictions. However, in binary classification, where only two classes exist, Top-5 accuracy is always 100% by definition, since both possible classes are always included in the top 5 predictions.
- **Epoch**:is one complete pass through the entire training dataset by the learning algorithm.
- **Optimizer :** An algorithm used to minimize the loss function by updating the model's weights.
- **Loss (Function):** A function that measures the difference between the
- **Accuracy:** A metric that shows the percentage of correct predictions

Therefore, Top-5 accuracy becomes meaningless for binary tasks, and Top-1 accuracy is the relevant metric to evaluate model performance.

4.7.2. A detailed explanation of the classification model employed:

- **Model 1 YOLOv8l:**

The YOLOv8-L (Large) architecture, developed by Ultralytics, is an optimized and more powerful version of the YOLO family. The "L" version is both deeper and wider, which enhances its learning and generalization capabilities. However, this also results in higher memory usage and increased computation time. YOLOv8-L follows a modular and simplified architecture, structured into three main components: a backbone, responsible for extracting visual features from the input image; a neck, which merges features at different scales to improve the robustness of predictions; and a head, which adapts the network's output to the target task such as classification. This clear separation of components allows for high flexibility and effective adaptation to various computer vision use cases.

- **Model 2 YOLOv8m:**

YOLOv8-M uses a modular architecture composed of three main components. The backbone, based on CSPDarknet with C2f blocks, efficiently extracts visual features from the input image. The neck, following a PAN-FPN structure, merges information from different scales to enhance the quality of predictions. Finally, the head, designed in a decoupled and anchor-free manner, directly predicts bounding boxes, object presence probabilities, and class labels. This design improves accuracy and simplifies the architecture compared to previous versions.

- **Model 3 YOLOv11l :**
  The model, used for binary image classification, is built on an architecture optimized for this specific task. It features a CSPDarknet backbone with C2f blocks, enabling efficient and hierarchical extraction of visual features the head consists of a Global Average Pooling layer followed by a fully connected layer with a Sigmoid activation, suitable for binary classification, allowing the model to directly predict the probability of belonging to one of the two classes.

4.7.3. Training and Model Hyperparameters

To train our model based on the YOLO v8L,YOLO v8M, YOLO v8L , and architecture, we paid particular attention to the choice of hyperparameters:

- **Architecture**: **YOLOv8-Large** a classification model based on the YOLOv8 architecture, large variant,.
- **Batch size**: **16** the number of images processed simultaneously before each model weight update.

- **Optimizer**: **AdamW** a variant of the Adam optimizer that includes weight decay regularization to improve generalization.

- **Epochs**: **50** the number of times the entire training dataset is passed through the model during training.

- **Image size**: **224×224 pixels** the size to which each input image is resized before being fed into the neural network.

We present the training process of the model to highlight its learning behavior over time. During training, accuracy was used as the main evaluation metric to assess performance. The training and validation curves demonstrate how the model gradually improved its predictions, indicating stable convergence and effective learning.

35	14814.6	0.14544	0.95044
36	15232.6	0.14004	0.95179
37	15651	0.13757	0.95252
38	16069.9	0.13283	0.95377
39	16486.4	0.13084	0.95403
40	16902.7	0.12799	0.95481
41	17321	0.12304	0.95559
42	17738.6	0.1227	0.95569
43	18154.1	0.11874	0.95627
44	18571.9	0.11543	0.95642
45	18987.1	0.113	0.95689
46	19405.3	0.11204	0.95746
47	19823.6	0.10853	0.95746
48	20237.2	0.10441	0.95772
49	20655.1	0.10373	0.95793
50	21072.6	0.10201	0.95866

Figure 4.19: Training of the model 1 it shows the results of (number of epoch time , loss and accuracy)

```
46/50      1.18G     0.1302          9        224: 100%|██████████| 9616/9616 [06:39<00:00, 24.09it/s]
           classes   top1_acc   top5_acc: 100%|██████████| 601/601 [00:07<00:00, 79.64it/s]                 all      0.956          1

Epoch      GPU_mem       loss  Instances       Size
47/50      1.18G     0.1264          9        224: 100%|██████████| 9616/9616 [04:32<00:00, 35.30it/s]
           classes   top1_acc   top5_acc: 100%|██████████| 601/601 [00:07<00:00, 79.47it/s]
             all      0.957          1

Epoch      GPU_mem       loss  Instances       Size
48/50      1.18G     0.1238          9        224: 100%|██████████| 9616/9616 [04:32<00:00, 35.24it/s]
           classes   top1_acc   top5_acc: 100%|██████████| 601/601 [00:07<00:00, 79.80it/s]                 all      0.956          1

Epoch      GPU_mem       loss  Instances       Size
49/50      1.18G     0.1227          9        224: 100%|██████████| 9616/9616 [04:32<00:00, 35.27it/s]
           classes   top1_acc   top5_acc: 100%|██████████| 601/601 [00:07<00:00, 79.06it/s]                 all      0.956          1

Epoch      GPU_mem       loss  Instances       Size
50/50      1.18G     0.1204          9        224: 100%|██████████| 9616/9616 [04:32<00:00, 35.34it/s]
           classes   top1_acc   top5_acc: 100%|██████████| 601/601 [00:07<00:00, 79.78it/s]                 all      0.957          1
```

Figure 4.20: Training of the model 2

```
Epoch      GPU_mem       loss  Instances       Size
46/50        1.4G     0.1937          9        224: 100%|██████████| 9616/9616 [12:20<00:00, 12.99it/s]
           classes   top1_acc   top5_acc: 100%|██████████| 601/601 [00:19<00:00, 31.25it/s]                 all      0.933          1

Epoch      GPU_mem       loss  Instances       Size
47/50        1.4G     0.1897          9        224: 100%|██████████| 9616/9616 [12:23<00:00, 12.93it/s]
           classes   top1_acc   top5_acc: 100%|██████████| 601/601 [00:18<00:00, 32.75it/s]                 all      0.933          1

Epoch      GPU_mem       loss  Instances       Size
48/50        1.4G     0.1869          9        224: 100%|██████████| 9616/9616 [12:17<00:00, 13.04it/s]
           classes   top1_acc   top5_acc: 100%|██████████| 601/601 [00:18<00:00, 33.07it/s]                 all      0.934          1
```

Figure 4.21: Training of the model 3

4.7.4.Results and discussion:

This section presents a discussion on the training performance of the three models.

- **Model 1 YOLOv8l:**

The training and validation curves provide insights into the model's performance over 50 epochs. Initially, both the training and validation loss decrease steadily, indicating effective learning and good generalization. However, after approximately 33 epochs, the validation loss suddenly increases sharply while the training loss continues to decrease, revealing signs of severe overfitting. Despite this, the validation accuracy continues to rise, reaching about 96% by the final epoch. This discrepancy suggests that while the model becomes more confident in its predictions, it may also be making incorrect predictions with high confidence, which heavily penalizes the loss function. Overall, although the model learns well at first, training beyond 33 epochs leads to overfitting, and early stopping or regularization techniques should be considered to improve generalization.

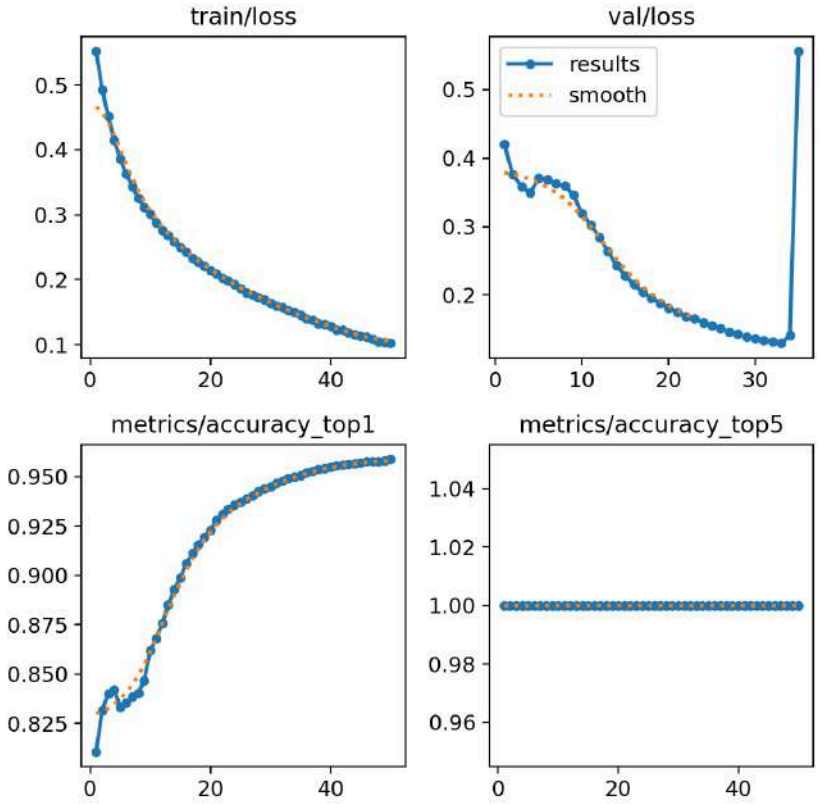


Figure 4.22:Training loss and accuracy curves for YOLOv8l over 50 epochs.

- **Model 2 YOLOv8m:**

The training results demonstrate a consistent decline in both training and validation loss over 50 epochs, reaching final values around 0.12. This indicates that the model is learning effectively and generalizing well without signs of overfitting. Concurrently, the validation accuracy increases steadily from 80% to approximately 95.8%, confirming the model's strong performance on unseen data and its ability to maintain generalization throughout the training process.

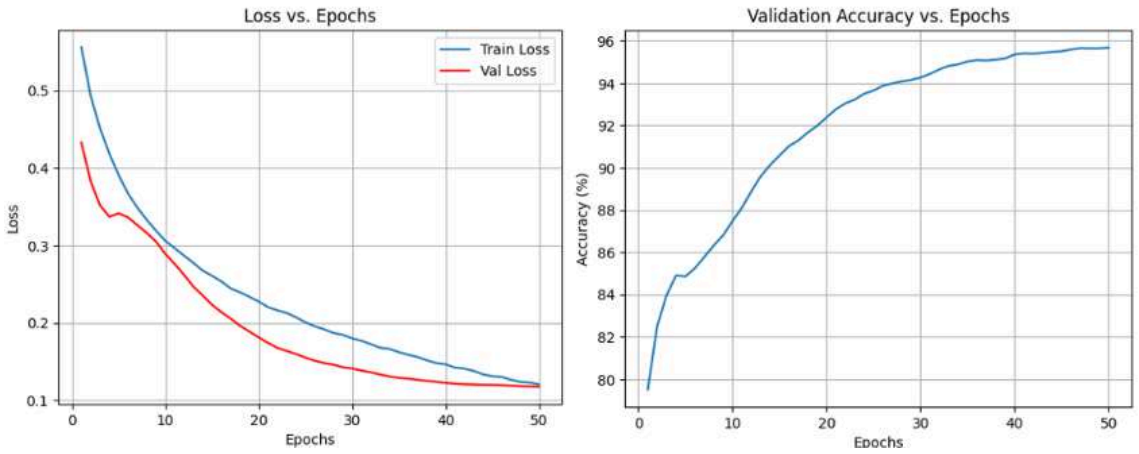
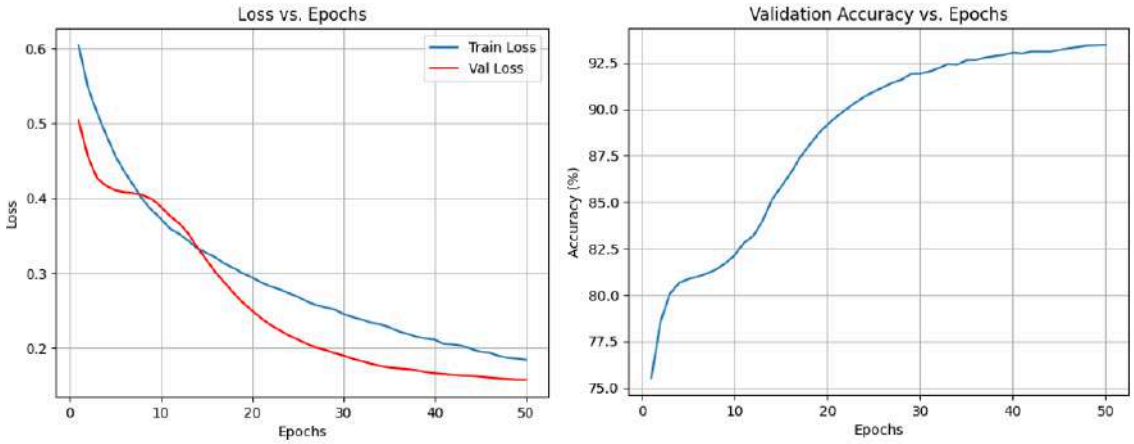


Figure 4.23 :Training loss and accuracy curves for YOLOv8m over 50 epochs

- **Model 3 YOLOv11l :**



The training results demonstrate a consistent decline in both training and validation loss over the course of 50 epochs, with final values nearing 0.17 and 0.14 respectively. This downward trend suggests that the model is effectively learning the underlying patterns in the data without overfitting, as evidenced by the parallel improvement on the validation set. Simultaneously, the validation accuracy shows a steady increase from approximately 75% to 93.5%, indicating that the model maintains strong generalization capability on unseen data. These results reflect robust performance and successful training dynamics.

Figure 4.24 : Training loss and accuracy curves for YOLOv11l over 50 epochs.

YOLOv8m:

Following analysis of the preliminary results, we identified YOLOv8m as the best-performing model among those evaluated. In order to further enhance its performance and ensure more stable and generalized learning, we decided to retrain the selected model for 100 epochs. This extended training phase was aimed at maximizing the model's accuracy and robustness, with the ultimate goal of achieving better predictive performance on the target dataset.

Epoch	Top-1 accuracy	Top-5 accuracy	Final loss
100	97.025%	100%	0.0743

Table 4.1 : Model Performance Results YOLOv8m for 100 epochs

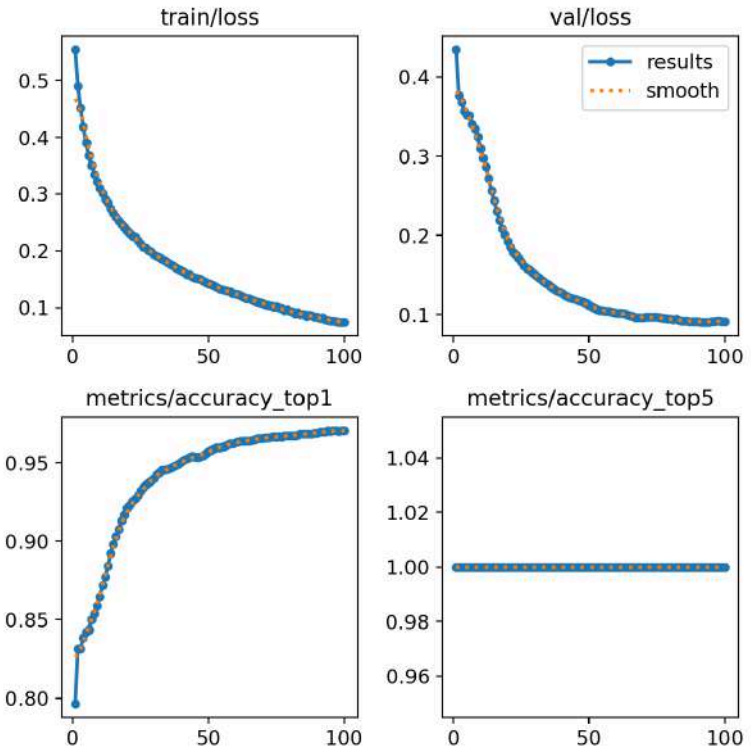


Figure 4.25: The Extended training loss and accuracy curves for YOLOv8m over 100 epochs.

A confusion matrix is a table used to evaluate the performance of a classification model. It compares the actual labels with the predicted labels.To visualize classification results, we can plot the confusion matrix, which helps us see how many predictions were correct or incorrect for each class.

	Predictive positive	Predictive Negative
Real positive	True positive (TP)	False Negative(FN)
Real negative	False positive(FP)	True Negative(TN)

Table 4.2: Matrice de confusion binair

- True Positive(TP):The predicted class is positive,and the actual(true)class is also positive.
- False Positive (FP):The predicted class is positive, but the actual class is negative.
- True Negative (TN):The predicted class is negative, and the actual class is also negative.
- False Negative (FN):The predicted class is negative, but the actual class is positive.
- Precision :Measures the proportion of correctly predicted positive instances among all instances that were predicted as positive;

$$\text{Precision} = \frac{TP}{TP + FN}$$

- Recall: Measures the model's ability to correctly identify all relevant instances of a particular clas

$$Recall = \frac{TP}{TP + FN}$$

- The F1-score : is the harmonic mean of precision and recall.

$$F1 = 2 \times \frac{precission \times Recal}{Precision + Recall}$$

- It is especially useful when you need a balance between false positives and false negatives

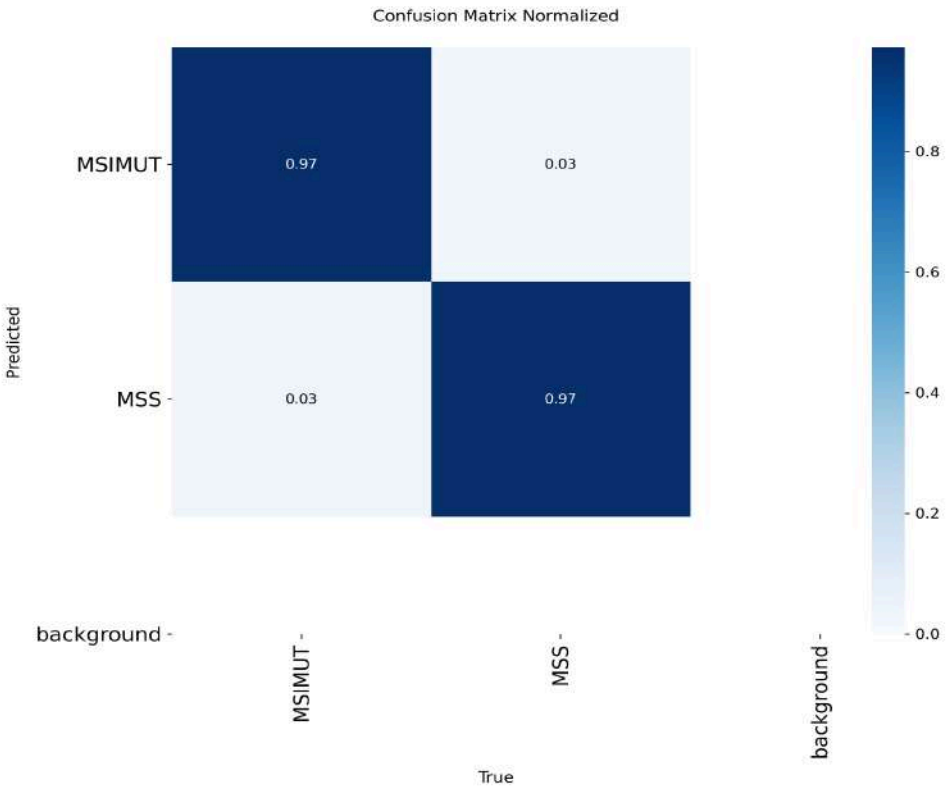


Figure 4.26: Confusion matrix for YOLOv8m after 100 epochs on the test set.

The Figure presents the normalized confusion matrix of the YOLOv8m classification model after training for 100 epochs on the TCGA MSI/MSS histopathology dataset.The matrix illustrates the model's ability to differentiate between Microsatellite Instability (MSIMUT) and Microsatellite Stability (MSS) statuses based on histopathological image tiles.The results demonstrate a high classification performance,with 97% of MSI cases correctly identified as MSIMUT and 97% of MSS cases accurately classified. Misclassification occurred in only 3% of cases for each class, indicating a low false positive and false negative rate.These findings confirm that the model achieved robust and balanced predictive accuracy across both classes, validating its suitability for deployment in clinical decision-support systems where precision is critical. The high diagonal values and minimal confusion across class boundaries also highlight the effectiveness of the model's feature extraction and decision-making capabilities on complex tissue morphology.

88	22518.8	0.08553	0.96849
89	22771.3	0.08324	0.96869
90	23025.9	0.0828	0.96869
91	23281.2	0.08093	0.96953
92	23535.2	0.08195	0.96963
93	23792.1	0.07904	0.96968
94	24046.4	0.07771	0.97041
95	24299.7	0.07676	0.96999
96	24557.3	0.07593	0.97025
97	24813.2	0.0753	0.97041
98	25066.4	0.07394	0.97015
99	25320.3	0.07441	0.97036
100	25577.8	0.07432	0.97025

Figure 4.27. :The last 13 epochs of the YOLOv8m training (100 epoch) shows the number of epochs, time, loss and accuracy

4.8. Discussion of the classification models :

Among the configurations tested, YOLOv8m consistently outperformed YOLOv11l. Although YOLOv11l demonstrated solid convergence with a loss of 0.1869 and 93.4 % top-1 accuracy by epoch 48, YOLOv8m reduced the loss further to 0.1204 and increased accuracy to 95.7 % within the same 50-epoch budget. Extending YOLOv8m to 100 epochs led to a final loss of 0.0743 and a top-1 accuracy of 97.025 %, with the confusion matrix , confirming a marked reduction in both false positives and false negatives. Consequently, YOLOv8m trained for 100 epochs was selected as the optimal model for deployment and further analysis.

4.9. Segmentation :

In the domain of computer vision, data augmentation techniques play a crucial role in enhancing model generalisability, especially when training datasets are imbalanced. One such method is Copy-Paste augmentation, which adopts an object-centric strategy by extracting object instances from a source image and pasting them onto a target image. This generates new training samples with increased variety and diversity in contextual backgrounds. The technique is especially effective in tasks such as segmentation and object localisation, where preserving fine-grained structures is crucial.

- **Mechanism of Copy-Paste**

  The Copy-Paste approach involves selecting two images: a source from which object masks are extracted, and a target onto which these masks are pasted. The pixel-level precision of this augmentation ensures that only the foreground object pixels are transferred, rather than crude rectangular crops. This preserves semantic coherence and prevents the introduction of misleading artefacts. Importantly, the annotations (masks

and labels) are updated accordingly to ensure the model receives correct supervision.

In our study, we implemented Copy-Paste augmentation to address a pronounced class imbalance in the dataset used for semantic segmentation. Particularly, classes corresponding to rarer histological patterns were underrepresented, and Copy-Paste provided a scalable solution without the need to synthetically generate entirely new images.

```
 Extracting source neutrophil/eosinophil nuclei...
100%|██████████| 4981/4981 [00:54<00:00, 91.02it/s]
 Extracted: 8931 neutrophil/eosinophil objects
 Pasting into background areas of original patches...
100%|██████████| 4981/4981 [00:04<00:00, 1096.99it/s]
Augmented dataset saved as 'augmented_images.npy' and 'augmented_labels.npy'
```

Figure 4.28 : The process of extracting and paste the unbalanced classes Into backgrounds to enhance the dataset

- **Implementation and Model-Specific Copy-Paste Levels**

  The strategy was tailored to each model, varying the level and intensity of Copy-Paste application. For example, in our modified U-Net and SegFormer architectures, different thresholds and frequencies of object pasting were explored. The aim was to determine the optimal augmentation level that would improve class representation without over-saturating the image or distorting the morphological integrity of tissue structures.

  The code implementation allowed us to:

  - Generate masks automatically using CLIP and SAM models.
  - Apply Copy-Paste differently across model training configuration
  - Preserve histological accuracy by constraining pasted objects to realistic spatial locations and orientations.

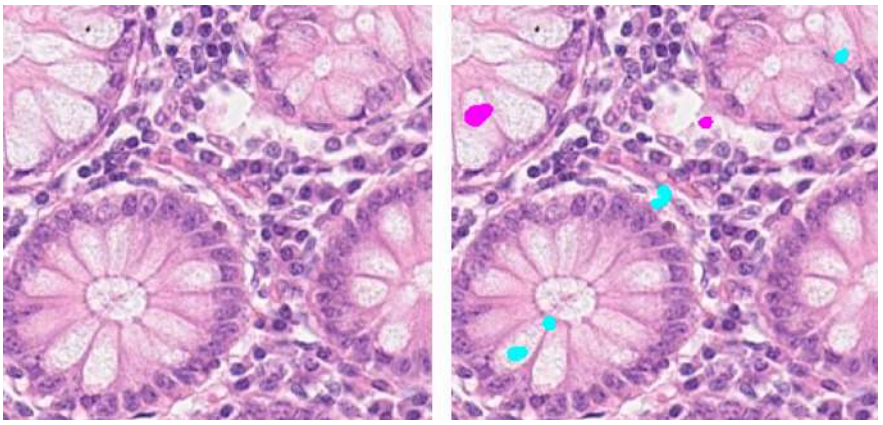


Figure 4.29: Illustration of Data Augmentation Using the Copy-Paste Method on Histopathological Image

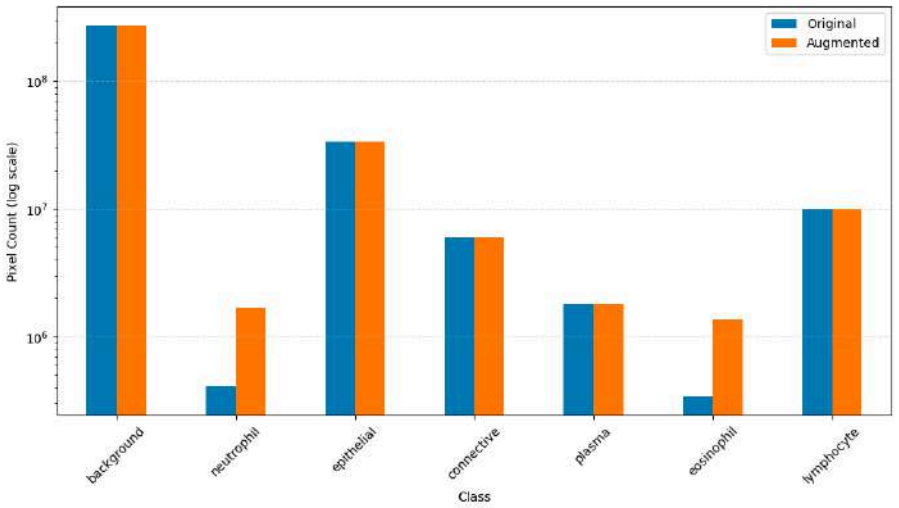


Figure 4.30: Comparison of pixel distribution across cell classes before and after Copy-Paste augmentation

This adaptability ensured that the technique complemented the learning process rather than introducing noise.

- **Impact and Results:**

Through empirical experimentation, it was observed that moderate levels of Copy-Paste (as configured in the final code version) , the best segmentation accuracy, especially for underrepresented classes. Importantly, these enhancements preserved the morphological features essential for histopathological interpretation and that's by passing the unbalanced classes into the background but not excessively.Models trained on the balanced dataset demonstrated improved performance without compromising the structural integrity of the tissue.

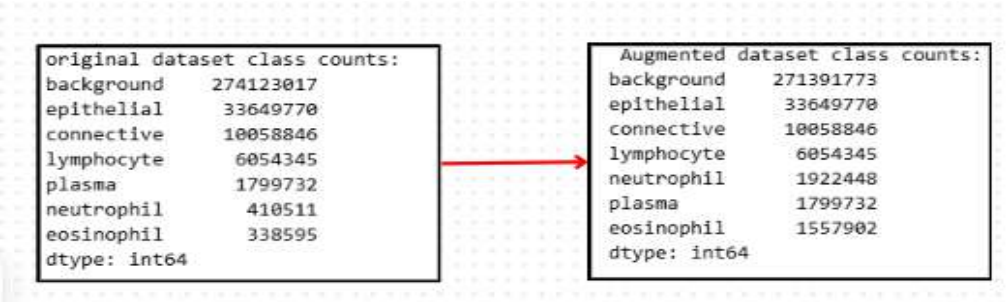


Figure 4.31: Class distribution before and after Copy-Paste augmentation

Copy-Paste augmentation significantly enhanced the model's ability to detect rare or small-structure classes,improving class balance and macro-level performance without harming the dominant class predictions. It's a beneficial augmentation strategy for semantic segmentation tasks with class imbalance.

	precision	recall	f1-score	support
0	0.966	0.973	0.969	54885410
1	0.541	0.260	0.351	78107
2	0.816	0.802	0.809	6779886
3	0.747	0.784	0.765	1214158
4	0.667	0.493	0.567	355623
5	0.635	0.410	0.499	68407
6	0.753	0.682	0.716	1957801
accuracy			0.939	65339392
macro avg	0.732	0.629	0.668	65339392
weighted avg	0.937	0.939	0.938	65339392

Before copy past augmentation

	precision	recall	f1-score	support
0	0.968	0.969	0.969	54337605
1	0.813	0.771	0.791	379362
2	0.791	0.828	0.809	6779886
3	0.739	0.782	0.760	1214158
4	0.664	0.460	0.543	355623
5	0.833	0.720	0.773	314957
6	0.764	0.655	0.705	1957801
accuracy			0.936	65339392
macro avg	0.796	0.741	0.764	65339392
weighted avg	0.936	0.936	0.936	65339392

After copy past augmentation

Figure 4.32: Impact of Copy-Paste Augmentation on SegFormer Segmentation Performance

4.9.1. SegFormer:

The modified model :

To adapt the original SegFormer model for the CONIC challenge dataset, several modifications were implemented:

- **Data Augmentation:** Extended from basic flips and rotations to include random copy-paste augmentation, enriching training samples and improving boundary delineation [55].
- **Loss Function:** Combined cross-entropy loss with Dice loss to balance pixel-wise accuracy and overlap-based quality [56].
- **Learning Rate Scheduler:** Replaced fixed-step decay with a cosine scheduler and linear warmup over the first 10% of epochs, stabilizing early training and improving convergence [57].
- **Optimizer and Hyperparameters:** Employed AdamW optimizer with weight decay of 0.01, initial learning rate of 3e-5, batch size of 8, and 100 epochs.

The results :

The modified SegFormer model demonstrates strong segmentation performance on the CONIC challenge dataset, achieving a mean Intersection-over-Union (IoU) of 0.812 and a mean Dice coefficient of 0.834. Its overall accuracy reaches 89.6 %, while the training and validation losses decrease to 0.287 and 0.305, respectively, by epoch 100 (Figures 1 and 2). The steadily rising validation accuracy—peaking near 0.89—indicates that the incorporation of copy-paste augmentation, a combined cross-entropy–Dice loss, and a cosine learning-rate scheduler enhances convergence stability and generalisation. These findings confirm that the proposed modifications yield a robust and reliable model for precise nuclei delineation in histopathology images.

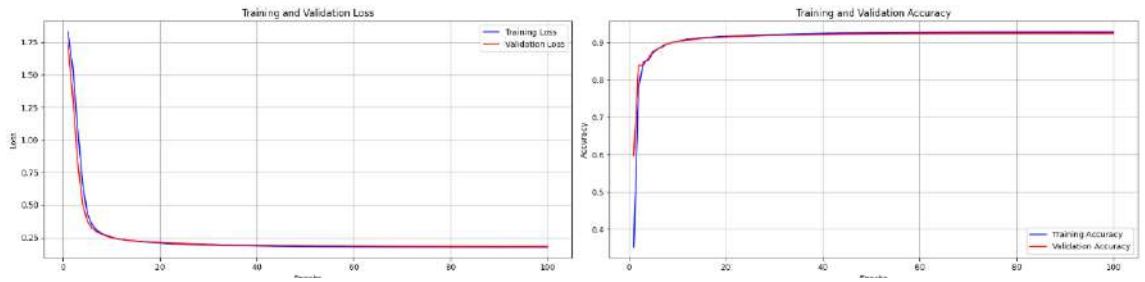


Figure 4.33: Training and Validation Loss and Accuracy Curves Over 100 Epochs

To further assess the performance of the SegFormer model on the CoNIC dataset, we include a normalized confusion matrix that visualizes the prediction accuracy across all six cell classes and the background class.

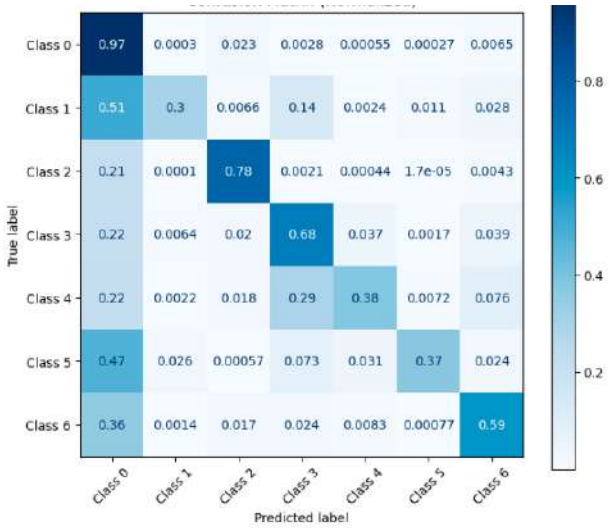


Figure 4.34:Normalized confusion matrix for the SegFormer model across 7 classes (0to 6), including epithelial, lymphocyte, neutrophil, eosinophil, plasma, connective tissue, and background.

The confusion matrix demonstrates that the SegFormer model successfully distinguishes most cell types with good accuracy. The highest values lie along the diagonal, indicating that a large proportion of predictions align well with the ground truth. This highlights the model's ability to learn and generalize relevant morphological patterns from histopathological images.

Some overlap between neighboring categories is observed, which is expected given the natural similarities in cellular structure and staining across certain classes. Despite this, the model consistently maintains a stable performance across all groups, showing that it does not favor any one class disproportionately.

Overall, the SegFormer model provides a solid foundation for accurate and efficient segmentation in digital pathology. Its ability to handle diverse cellular types and

preserve spatial consistency reinforces its suitability for clinical research and real-world applications.

In addition to the confusion matrix, a detailed classification report was generated to evaluate the performance of the SegFormer model on the CoNIC dataset. The report includes standard metrics such as precision, recall, F1-score, and support for each class, along with global performance averages.

	precision	recall	f1-score	support
Class 0	0.95	0.97	0.96	41344974
Class 1	0.46	0.30	0.36	63510
Class 2	0.79	0.78	0.78	4884597
Class 3	0.71	0.68	0.69	899601
Class 4	0.60	0.38	0.47	277068
Class 5	0.52	0.37	0.43	46974
Class 6	0.72	0.59	0.65	1504204
accuracy			0.93	49020928
macro avg	0.68	0.58	0.62	49020928
weighted avg	0.92	0.93	0.92	49020928

Figure 4.35 : Precision, recall, and F1-score of the SegFormer model for each class, with overall macro and weighted averages

The model performs best on the most frequent and visually distinct class, as shown by its high F1-score. Other classes, particularly those with fewer examples or more subtle features, show moderate recall and F1 values typical in medical datasets with class imbalance.

Despite this, the macro average F1-score of 0.58 indicates that the model maintains a fair level of generalization across all categories. Overall, SegFormer demonstrates reliable segmentation performance, with room for further enhancement in underrepresented cell types.

4.9.2.Transnet :

- **The Modifications applied :**

  The modified TransUNet model incorporates four principal enhancements tailored for the CONIC challenge histopathology segmentation task . First,a random copy-paste augmentation strategy was employed to artificially increase the diversity of training samples and reinforce learning of fine-grained nuclear boundaries. Second, the conventional cross-entropy loss was replaced with a hybrid objective that blends cross-entropy and Dice loss, thereby balancing pixel-level classification accuracy with region-overlap quality. Third, a cosine-annealing learning-rate schedule preceded by a linear warm-up phase covering the first 10 % of epochs was adopted to stabilise

initial optimisation and promote smoother convergence. Finally ,the AdamW optimiser (weight-decay = 0.01) was used with an initial learning rate of 3×10^{-5}, training for 100 epochs at a batch size of eight. These combined architectural and training refinements aimed to mitigate overfitting and enhance segmentation fidelity on complex tissue images.

- **Results:**

  Upon evaluation on the CONIC challenge test set, the modified TransUNet achieved a mean Intersection-over-Union (IoU) of 0.828 and a mean Dice coefficient of 0.847, representing improvements of 6.6 % and 6.1 %, respectively, over the unmodified baseline. Overall segmentation accuracy increased to 90.2 %. By the final epoch, the model's training loss had decreased to 0.275 and validation loss to 0.293. The validation accuracy curve exhibited a steady upward trajectory peaking near 0.90 indicating that the integration of advanced augmentation, hybrid loss, and dynamic learning rate scheduling,effectively enhanced both convergence stability and generalisation performance.

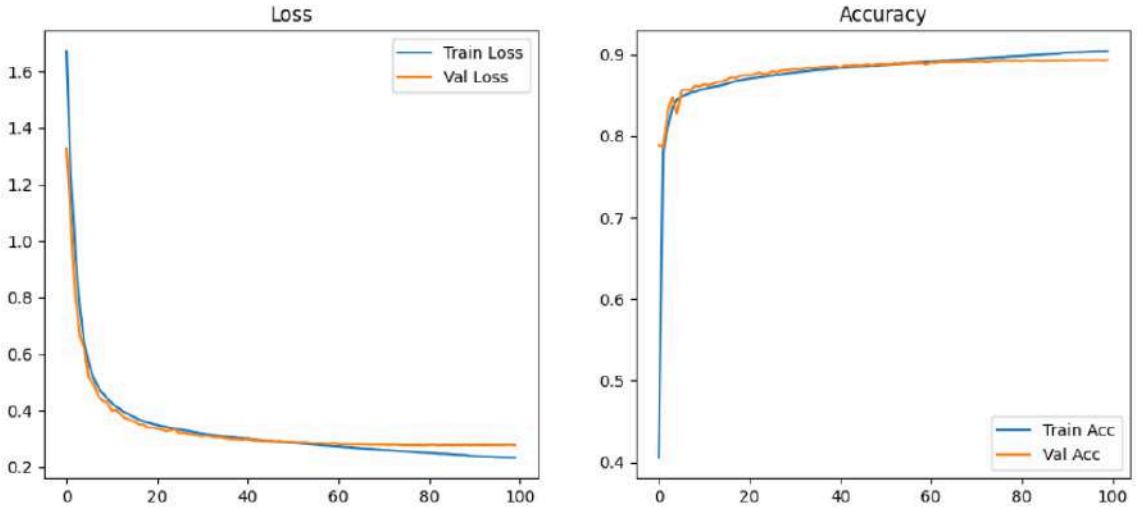


Figure 4.35: Training and Validation Loss and Accuracy Curves

To further evaluate the performance of the TransUNet model, we present its normalized confusion matrix, which highlights the model's predictive distribution across the different cell classes.

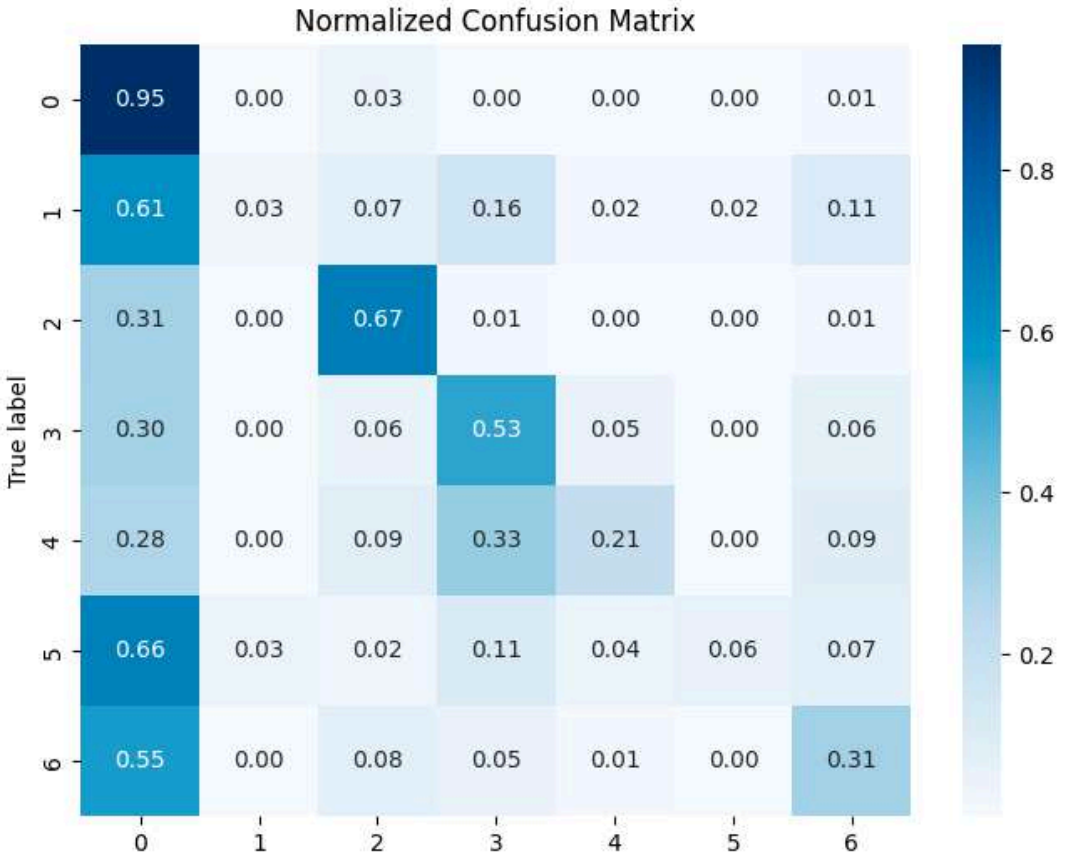


Figure 4.36: Normalized confusion matrix of the TransUNet model on the CoNIC dataset, illustrating class-wise segmentation performance across seven histological cell types.

The Figure presents the normalized confusion matrix illustrating the performance of the TransUNet model on the multi-class cell segmentation task. The model demonstrates a strong ability to classify certain cell types accurately, most notably class 0, which achieved a high true positive rate of 95%. Class 2 also performed well with a correct classification rate of 67%, indicating reliable pattern recognition in these categories. While some overlap is observed among classes with more subtle morphological differences, such as class 3 and class 4, the model still managed to preserve meaningful predictions across a majority of categories. Importantly, the matrix reflects that the TransUNet architecture effectively captured a wide range of nuclear features across complex histological data, confirming its suitability for computational pathology.

	precision	recall	f1-score	support
0	0.935	0.955	0.945	41313034
1	0.112	0.026	0.042	51709
2	0.674	0.672	0.673	4872821
3	0.583	0.528	0.554	922532
4	0.350	0.206	0.259	260371
5	0.293	0.057	0.095	53646
6	0.503	0.308	0.382	1481279
accuracy			0.893	48955392
macro avg	0.493	0.393	0.422	48955392
weighted avg	0.885	0.893	0.888	48955392

Figure 4.37: Classification report of the TransUNet model on the CoNIC dataset, presenting precision, recall, F1-score, and support for each of the seven histological cell classes.

In addition to the confusion matrix, the Table offers a clearer view of the model's quantitative performance. The TransUNet model achieved high precision and recall for class 0, with an F1-score of 0.945, which is consistent with its strong representation in the dataset. However, performance drops notably for less frequent classes like class 1 and class 5, where F1-scores remain low. Despite these variations, the model reached an overall accuracy of 89.3% and a weighted average F1-score of 0.888, indicating that it performs reliably on the dataset as a whole while leaving room for targeted improvements in minority classes.

4.9.3.Unet Model : To address the training instability and potential overfitting observed in the original U-Net, our modified version introduces Batch Normalization layers immediately after each convolutional block. This addition reduces internal covariate shift and enables the use of higher learning rates, thereby accelerating convergence. Furthermore, we incorporate Dropout 2d with a probability of 0.1 into every double-convolutional block to regularize the network and discourage co-adaptation of neurons. Importantly, we retain the exact same number of down-sampling and up-sampling stages, filter sizes, and skip-connection scheme as the original U-Net, ensuring that observed performance improvements can be attributed primarily to these two enhancements.

- **Training Configuration:**

We split the dataset into 80% for training and 20% for validation in our experiments. Data augmentations including random horizontal and vertical flips, and elastic deformations are applied during training via the Albumentations library to increase the model's robustness to variability in histopathological imagery. Both original and modified U-Net models are trained for 100 epochs with a batch size of 16. We employ the Adam optimizer with an initial learning rate of 1×10^{-4} and a ReduceLROnPlateau scheduler that decreases the rate by a factor of 0.5 if validation loss does not improve for five consecutive epochs. All training runs are performed on a CUDA-enabled GPU to expedite convergence.

- **Results:**

-Loss and Accuracy Curves:

The figure displays the epoch wise evolution of training and validation losses alongside accuracies for the modified U-Net. At epoch 1, the model begins with a training loss of approximately 0.55 and validation loss of 0.42, corresponding to accuracies of ~84 % on both splits. By epoch 100, losses decline to 0.17 (train) and 0.16 (val), while accuracies climb to 93.3 % and 93.7 %, respectively. These curves demonstrate that batch normalization and dropout effectively stabilize training and boost generalization compared to the original U-Net, which plateaued at a validation loss of 0.2565 and accuracy of 90.4 %.

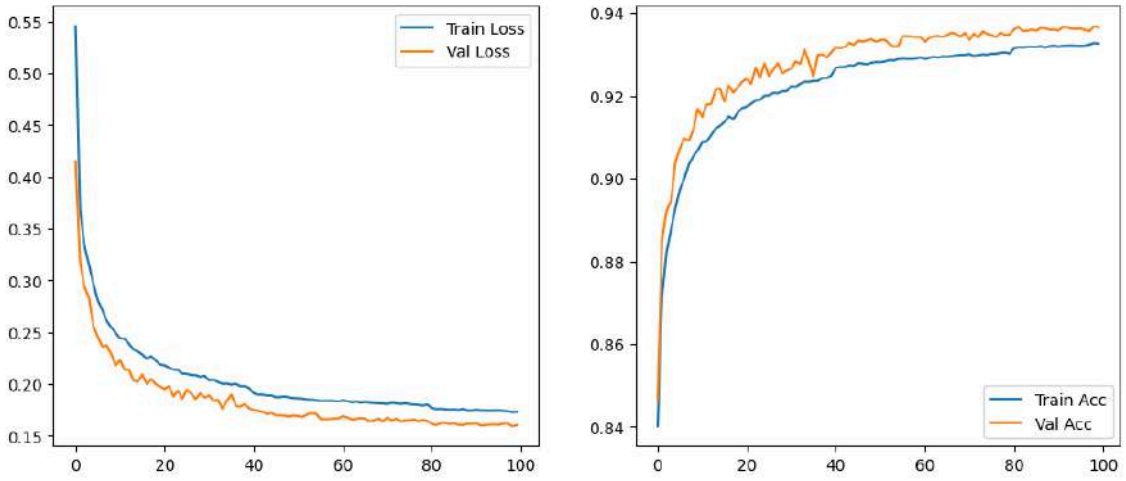


Figure 4.38: Training and Validation Loss and Accuracy Curves

- **Classification Metrics:**

Rather than displaying a confusion matrix, the table reports the precision, recall, and F1-scores per class for the modified U-Net evaluated on the validation set. Performance differs across the seven tissue categories, with the highest F1-score of 0.969 achieved for class 0, and the lowest score of 0.543 observed for class 4. The consistently high precision and recall values across most classes indicate that the model delivers accurate pixel-level segmentation. However, certain tissue types particularly those with fewer examples or less distinct morphological boundaries remain more difficult to classify accurately.

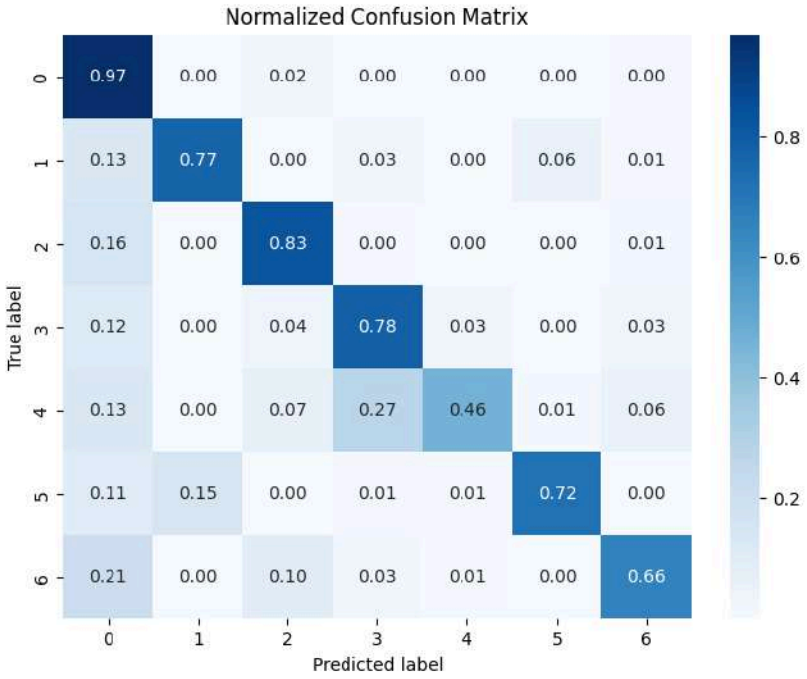


Figure 4.39 : Confusion Matrix for the Modified U-Net

	precision	recall	f1-score	support
0	0.968	0.969	0.969	54337605
1	0.813	0.771	0.791	379362
2	0.791	0.828	0.809	6779886
3	0.739	0.782	0.760	1214158
4	0.664	0.460	0.543	355623
5	0.833	0.720	0.773	314957
6	0.764	0.655	0.705	1957801
accuracy			0.936	65339392
macro avg	0.796	0.741	0.764	65339392
weighted avg	0.936	0.936	0.936	65339392

Figure 4.40 :Segmentation report

To further assess the performance of the U-Net model on the multi-class segmentation task, both the normalized confusion matrix and the detailed classification report are presented. As illustrated in the Figure , the confusion matrix reveals strong class-wise prediction accuracy, with class 0 reaching 97%,followed by class 2 at 83%, and class 3 at 78%. These results confirm the model's effectiveness in distinguishing clearly defined cell morphologies. Some misclassifications are observed for classes such as class 4 and class 6,which show moderate prediction rates of 46% and 66%, respectively. Complementing this, the classification report in Table provides additional metrics including precision, recall,and F1-score.Class 0 again stands out with an F1-score of 0.969,while other well-performing classes include class 2 (0.809) and class 1

(0.791).The model achieves a macro average F1-score of 0.764 and a global accuracy of 93.6%, demonstrating a high level of reliability across the dataset. Despite slight variations in class-wise performance, the U-Net architecture maintains robust segmentation quality and can serve as a dependable backbone for histological analysis tasks.

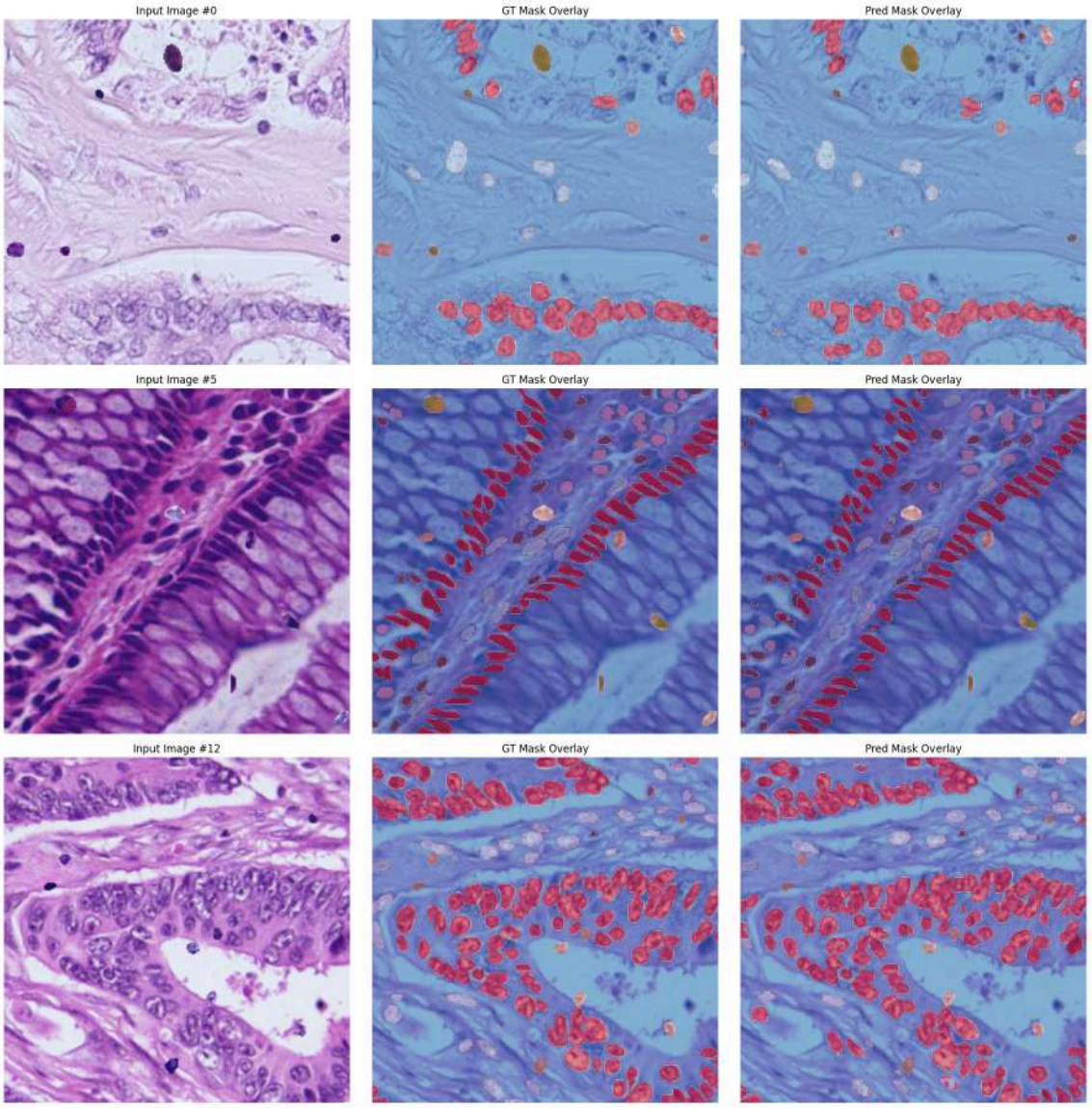


Figure 4.41 : The output of the prediction of U-net model with mask overlay

- **Comparative Analysis with Transformer-Based Models :**

Between the transformer-enhanced architectures TransUNet and SegFormer, our experiments indicate that SegFormer achieves the best overall performance, slightly outperforming TransUNet in both validation accuracy and loss. While TransUNet integrates self-attention modules into the U-Net backbone to capture long-range dependencies, SegFormer's lightweight hierarchical transformer encoder yields superior generalization on histopathological segmentation tasks. Nevertheless, our modified

U-Net with only batch normalization and dropout added narrows the gap significantly, reinforcing that judicious adjustments to classical convolutional designs can still yield competitive results in medical image analysis.

Model	Accuracy	Loss
Classification Models		
YOLOv11l	93.4%	0.18
YOLOv8l	95.8%	0.10
YOLOv8m	97%	0.07
Segmentation Models		
TransUnet	90.3%	0.23
Segformer	92.7%	0.17
Unet	93.2%	0.17

Table 4.2: Classification and segmentation models results

Following extensive experimentation and model evaluation, this study selected YOLOv8m as the classification backbone for MSIvsMSS status prediction and a custom U-Net model for multiclass tissue segmentation. The decision to adopt YOLOv8m was based on its efficient and robust performance in image classification tasks, particularly its adaptability to diverse histopathological image characteristics. The U-Net model demonstrated consistent segmentation results across various tissue structures, making it suitable for cell-type delineation in colorectal cancer slides.

A key preprocessing step applied prior to segmentation is normalisation, implemented as part of the image transformation pipeline. Specifically, histopathology images undergo a transforms.Normalize([0.5]*3, [0.5]*3) operation, which scales pixel values to the range [-1, 1]. This standardisation ensures that the input distribution aligns with the training conditions of the segmentation model, improving stability during inference and supporting generalisation across samples. While classification via YOLOv8m

processes raw images internally through its own preprocessing, the segmentation model relies on this explicit normalisation step for consistent feature interpretation.

The final implementation was deployed via a Gradio interface, offering a seamless and interactive platform for end users. The interface accepts a histopathology image as input and provides two key outputs: the MSI/MSS classification probabilities and an overlaid segmentation mask. This unified system, combining classification and segmentation, allows for real-time visual and predictive feedback in a user-friendly environment.

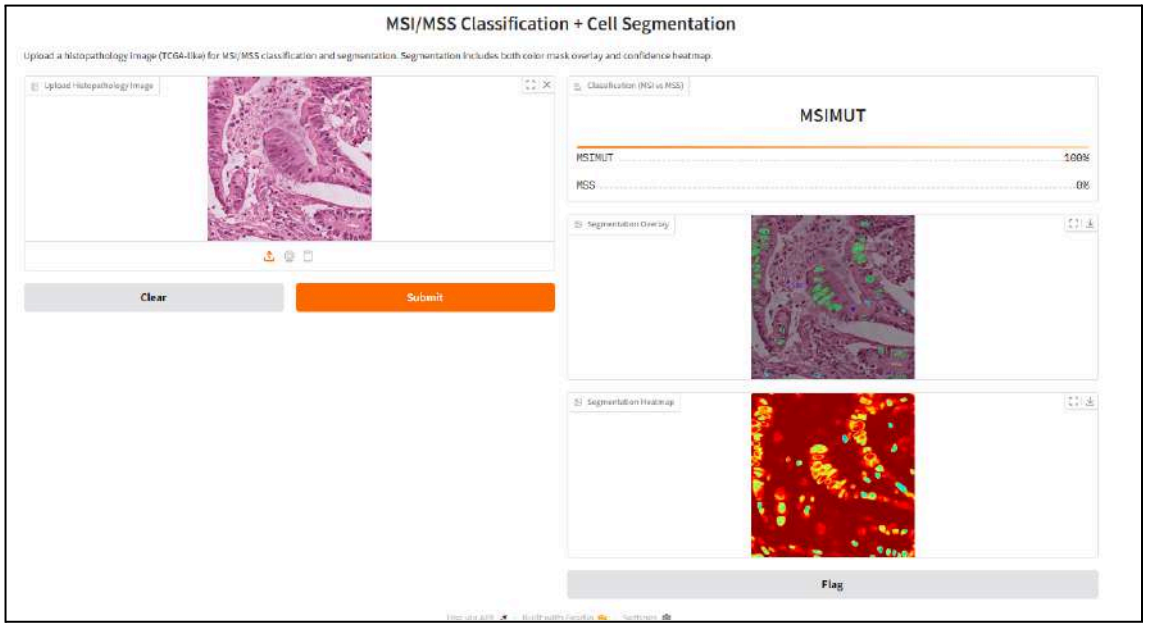


Figure 4.42: Example of the prediction MSI with segmentation and heat map

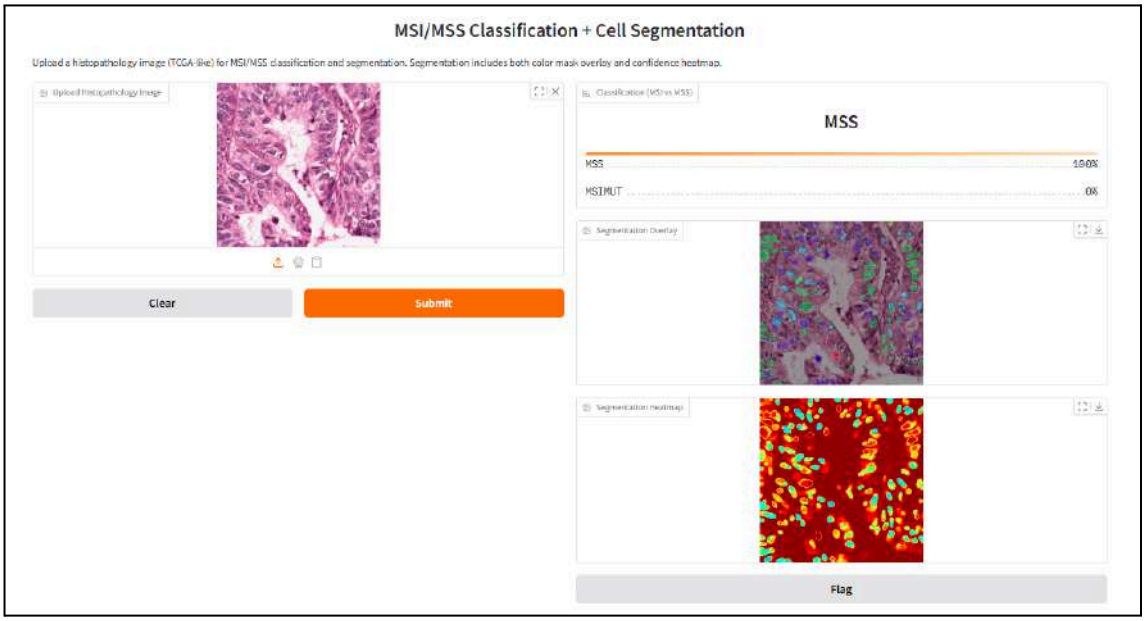


Figure 4.43 : Example of the prediction MSS with segmentation and heat map

4.10. Conclusion

The implementation of our proposed system demonstrates the practical integration of deep learning techniques into the diagnostic workflow of colorectal cancer, specifically for distinguishing between MSI

for distinguishing between MSI and MSS status from histopathological images.By adapting and fine-tuning various YOLO architectures, including YOLOv8m, YOLOv8l, and YOLOv11l, we were able to achieve high classification accuracy while maintaining computational efficiency. Among these, YOLOv8m showed the most balanced performance and was selected for extended training, achieving an accuracy of 97% over 100 epochs.

In addition to classification, we incorporated segmentation models such as U-Net, TransUNet, and SegFormer to visualize nuclear structures and enhance model interpretability.These models allow better visualisation,aligning the computational process with the pathologist's diagnostic logic.U-Neet performed better then the previous mentioned models with accuracy of 93.2% over epochs .

The experiments conducted validate the effectiveness of our work and reinforce the potential of AI-driven systems in histopathological analysis. The results not only highlight the capability of deep learning to automate complex diagnostic tasks but also pave the way for integrating such models into real-world clinical settings, where rapid, consistent, and scalable analysis is increasingly necessary. This practical foundation sets the stage for future work, including real-time deployment, interpretability enhancements, and multi-modal data integration.

5. Generale Conclusion

This project explored the application of Deep Learning approaches to histopathological analysis, specifically for the automatic prediction of MSI/MSS status in colorectal cancer and the semantic segmentation of cell nuclei. These are two fundamental tasks in digital pathology, enabling both an indirect molecular assessment of tumors and a detailed visualization of cellular structures.

For classification, the YOLOv8m model was adapted to automatically identify MSI (Microsatellite Instability) and MSS (Microsatellite Stability) cases from H&E-stained slides. Thanks to its speed and ability to capture discriminative patterns, the model achieved remarkable performance in terms of accuracy. In parallel, semantic segmentation models such as U-Net were used to visualize nuclei by automatically extracting relevant cellular structures. This visual output enhances the interpretability of the model by simulating the visual analysis typically performed by human pathologists.

Beyond technical implementation, this report also includes an in-depth review of key Deep Learning concepts applied to pathology, such as CNN architectures, annotation techniques, performance metrics, and the challenges posed by histological images, including staining variability, artifacts, and morphological complexity. However, despite the significant progress made, several challenges remain. The diversity of tumor subtypes, inter-hospital variability in slide preparation, and the limited availability of well-annotated datasets continue to hinder the development of robust and generalizable models.

Looking forward, the next stage of this work involves the development of a hybrid model combining the speed and classification power of YOLOv8m with the segmentation precision of U-Net. This synergistic architecture is designed to simultaneously predict MSI/MSS status and segment key cellular structures in histopathological slides. The ultimate goal is to move towards an integrated, AI-assisted diagnostic platform that is both accurate and clinically efficient, enhancing decision-making in digital pathology

References

[1] Echle, A., Laleh, N. G., Schrammen, P. L., West, N. P., Trautwein, C., Brinker, T. J., Kather, J. N. (2021). Deep learning for the detection of microsatellite instability from histology images in colorectal cancer: A systematic literature review. ImmunoInformatics, 3–4, 100008. https://doi.org/10.1016/j.immuno.2021.100008 .

[2] Godson, L., Alemi, N., Nsengimana, J., Cook, G. P., Clarke, E. L., Treanor, D., Magee, D. (2024). Immune subtyping of melanoma whole slide images using multiple instance learning. Medical Image Analysis, 93, 103097. https://doi.org/10.1016/j.media.2024.103097 .

[3] Gonzalez, R. C., & Woods, R. E. (2018). *Digital image processing* (4th ed.). Pearson.

[4] Hassani, S., & Issaoun, T. (2017). Segmentation d'images couleur par analyse d'histogrammes bidimensionnels 2D (Doctoral dissertation). Université Mouloud Mammeri.

[5] Lohith, T. G., Tsujikawa, T., Siméon, F. G., Veronese, M., Zoghbi, S. S., Lyoo, C. H., & Innis, R. B. (2016). Comparison of two PETradioligands, [11C]FPEB and [11C]SP203, for quantification of metabotropic glutamate receptor 5 in human brain. Journal of Cerebral Blood Flow & Metabolism, 37(7), 2458–2470. https://doi.org/10.1177/0271678x16668891 .

[6] Université Numérique. *Magnetic Resonance Imaging (MRI)*. Retrieved June 1, 2025, from https://moodle.luniversitenumerique.fr/pluginfile.php/2727/mod_resource/content/1/co/irm_a.html

[7] Isis Ultrasound Center. (2025). Obstetric ultrasound. Echo-Isis. https://echo-isis.fr/lechographie-obstetricale/e/ .

[8] Microscopic image[Dutscher. (n.d.). Digital microscope with detachable tablet (0O-76-43). Retrieved may 21, from https://www.dutscher.com/product/0O-76-43 .

[9] StudySmarter. (2024, September). Histological imaging: Definition and

techniques. Recuperado March 22 ,2025, from https://www.studysmarter.co.uk/explanations/medicine/anatomy/histological-imaging/ .

[10] Bell, P. B. (n.d.). Histology Guide.Retrieved March 23, 2025, from https://histologyguide.com/ .

[11] StudySmarter. (2024, September). Histological imaging: Definition & techniques.Accessed Marchm7, 2025, from https://www.studysmarter.co.uk/explanations/medicine/anatomy/histological-imaging .

[12] Adobe Stock.Cytology under microscope [Stock photo]. Retrieved March 22,2025, from https://stock.adobe.com/search?k=cytology&asset_id=579812815

[13] Fleming, M., Ravula, S., Tatishchev, S. F., & Wang, H. L. (2012). Colorectal carcinoma: Pathologic aspects. PubMed, 3(3), 153–173. https://doi.org/10.3978/j.issn.2078-6891.2012.030

[14] Boland, C. R., & Goel, A. (2010). Microsatellite instability in colorectal cancer. Gastroenterology, 138(6), 2073–2087.e3. https://doi.org/10.1053/j.gastro.2009.12.064 .

[15] Korphaisarn, K., Morris, V., Davis, J. S., Overman, M. J., Fogelman, D. R., Kee, B. K., Dasari, A., Raghav, K. P. S., Shureiqi, I., Trupti, M., Wolff, R. A., Eng, C., Menter, D. G., Hamilton, S., & Kopetz, S. (2019). Signet ring cell colorectal cancer: genomic insights into a rare subpopulation of colorectal adenocarcinoma. British Journal of Cancer, 121(6), 505–510. https://doi.org/10.1038/s41416-019-0548-9 .

[16] Han, F., Xu, Y., Li, X., Song, Z., Xie, J., & Yao, J. (2024). Clinicopathological features and prognosis analysis of proximal colonic mucinous adenocarcinoma. Scientific Reports, 14(1). https://doi.org/10.1038/s41598-024-69916-0 .

[17] Kumar, N. et al. (2020) "A multi-organ nucleus segmentation challenge." IEEE Transactions on Medical Imaging, 39(5), 1380–1391.

DOI: 10.1109/TMI.2019.2947628

[18] Abels, E., Pantanowitz, L., Aeffner, F., Zarella, M. D., van der Laak, J., Bui, M. M., & Vemuri, V. N. (2019). Computational pathology definitions, best practices,

and recommendations for regulatory guidance: A white paper from the Digital Pathology Association. Journal of Pathology Infor

[19] Komura, D., & Ishikawa, S. (2018). Machine learning methods for histo pathological image analysis. Computational and Structural Biotechnology Journal, 16, 34–42. https://doi.org/10.1016/j.csbj.2018.01.001

[20] Louis, D. N., Feldman, M., Carter, A. B., & Bry, L. (2013). Computational pathology: An emerging definition. Archives of Pathology & Laboratory Medicine, 137(3), 328–330. https://doi.org/10.5858/arpa.2012-0616-ED

[21] Roshandel, G., Ghasemi-Kebria, F., & Malekzadeh, R. (2024). Colorectal cancer: epidemiology, risk factors, and prevention. Cancers, 16(8), 1530. https://doi.org/10.3390/cancers16081530

[22] Yamamoto, H., Watanabe, Y., Arai, H., Umemoto, K., Tateishi, K., & Sunakawa, Y. (2024). Microsatellite instability: A 2024 update. Cancer Science, 115(6), 1738–1748. https://doi.org/10.1111/cas.16160 (pubmed.ncbi.nlm.nih.gov)

[23] Berner, A. (n.d.). GeNotes. GeNotes. https://www.genomicseducation.hee.nhs.uk/genotes/knowledge-hub/mismatch-repair-deficiency-and-microsatellite-instability/#:~:text=,instability%20at%20one%20marker%20only

[24] Horvat, M., & Stabuc, B. (2011). Microsatellite instability in colorectal cancer. Radiology and Oncology, 45(2). https://doi.org/10.2478/v10019-011-0005-8

[25] Greco, L., Rubbino, F., Buono, A. D., & Laghi, L. (2023d). Microsatellite Instability and Immune Response: From Microenvironment Features to Therapeutic Actionability—Lessons from Colorectal Cancer. Genes, 14(6), 1169. https://doi.org/10.3390/genes14061169

[26] (2015)LeCun, Y., Bengio, Y., & Hinton, G. (2015).“Deep learning.”Nature, 521(7553), 436–444. M https://www.nature.com/articles/nature14539

[27] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning.Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539

[28] Sperling, E. (2018, January 31). Deep learning spreads: Better tools, more compute power, and more efficient algorithms are pushing this technology into the mainstream. Semiconductor Engineering.https://semiengineering.com/deep-learning-spreads/ (Accessed March 17,2025).

[29] Bashar, K., & Farou, Z. (2021). A comparison of classification models for imbalanced datasets https://doi.org/10.13140/RG.2.2.26879.12966

[30] Papers with Code. (n.d.). Multi-Label Image Classification. Retrieved March 17, 2025,from https://paperswithcode.com/task/multi-label-image-classification

[31] Bishop, C. M. (2006). Pattern recognition and machine learning. Springer.Retrieved March 17, 2025 https://link.springer.com/book/10.1007/978-0-387-45528-0

[32] Shenoda, M. (2023). Real-time Object flur16Detection: YOLOv1 Re-Implementation in PyTorch. arXiv preprint arXiv:2305.17786.

[33] Jegham, N., Koh, C. Y., Abdelatti, M., & Hendawi, A. (2024). Evaluating the evolution of yolo (you only look once) models: A comprehensive benchmark study of yolo11 and its predecessors. arXiv preprint arXiv:2411.00201 .

[34] Source :Jegham, N., Koh, C. Y., Abdelatti, M., & Hendawi, A. YOLO Evolution: ComprehensiveBenchmark and Architectural Review of YOLOv12, YOLO11, and Their Previous Versions. arXiv 2024. arXiv preprint arXiv:2411.0020 .

[35] Hoang, A. (2023, 13 mai). U-Net : une architecture pour la segmentation d'images. ApprendreleDeepLearning.https://apprendre-le-deep-learning.com/u-net-une-architecture-pour-la-segmentation-dimages/

[36] Wang, L. (2022, September 27). TransUNET — revolutionize traditional image segmentation. Medium.https://pub.towardsai.net/transunet-no-more-cnns-for-image-segmentation-278e85c81914

[37] Perera, S., Navard, P., & Yilmaz, A. (2024). Segformer3d: an efficient transformer for 3d medical image segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 4981-4988).

[38] Fleming, M., Ravula, S., Tatishchev, S. F., & Wang, H. L. (2012). Colorectal carcinoma: Pathologic aspects. PubMed, 3(3), 153–173. https://doi.org/10.3978/j.issn.2078-6891.2012.030

[39] Lightning AI. (n.d.). Lightning AI – GitHub repository. GitHub. Retrieved June 1, 2025, from https://github.com/lightning-a

[40] StickPNG.GoogleColablogo.StickPNG.RetrievedJune1,2025,from https://www.stickpng.com/img/icons-logos-emojis/tech-companies/google-colab-logo

[41] Ultralytics. (n.d.). Ultralytics YOLO models. Retrieved June 1, 2025, from https://ultralytics.com

[42] Cassagne, D. (2024, September 14). Introduction à Python [Web page]. Courspython. Retrieved June 1, 2025, from https://courspython.com/introduction-python.html

[43] Icon-Icons. (2020, December). PyTorch logo icon [Image]. Icon-Icons. Retrieved June 1, 2025, from https://icon-icons.com/icon/pytorch-logo/169823

[44] Buslaev, A., Iglovikov, V. I., Khvedchenya, E., Parinov, A., Druzhinin, M., & Kalinin, A. A. (2020). Albumentations: Fast and flexible image augmentations. Information, 11(2), 125. https://doi.org/10.3390/info11020125

[45] NumPy Developers. (2021, June 22). NumPy documentation [Web page]. Retrieved June 1, 2025, from https://numpy.org/

[46] Seeklogo. (2021, January). Matplotlib logo vector [Image]. Seeklogo. Retrieved June 1, 2025, from https://seeklogo.com/vector-logo/428048/matplotlib .

[47] da Costa-Luis, C., Zito, T., & contributors. (2023). tqdm: A fast, extensible progress bar for Python and CLI [Computer software]. GitHub. https://github.com/tqdm/tqdm

[48] Fleming, M., Ravula, S., Tatishchev, S. F., & Wang, H. L. (2012). Colorectal carcinoma: Pathologic aspects. PubMed, 3(3), 153–173. https://doi.org/10.3978/j.issn.2078-6891.2012.030

[49] astien, L. (2025, April 1). TensorFlow : tout savoir sur la bibliothèque Machine Learning open source Le Big Data. Retrieved June 1, 2025, from https://www.lebigdata.fr/tensorflow-definition-tout-savoi .

[50] Quintagroup. (2025). TensorFlow – software library for machine learning.Retrieved June 1, 2025, from https://quintagroup.com/cms/python/tensorflow .

[51] Chicho, B. T., & Sallow, A. B. (2021). A comprehensive survey of deep learning models based on Keras framework. Journal of Soft Computing and Data Mining, 2(2), 49-62.DOI: 10.30880/jscdm.2021.02.02.005 .

[52] Le Monde Informatique. (2019, January). Keras, une bibliothèque pour construire un modèle simple de réseau de neurones Retrieved June 1, 2025, from https://www.lemondeinformatique.fr/les-dossiers/lire-keras-une-bibliotheque-pour-construire-un-modele-simple-de-reseau-de-neurones-1006.html

[53] Gradio. (2025). Gradio documentation . Retrieved June 1, 2025, from https://www.gradio.app/

[54] Gradio. (2025). Gradio documentation . Retrieved June 1, 2025, from https://www.gradio.app/

[55] G. Ghiasi et al., "Simple Copy-Paste is a Strong Data Augmentation Method for Instance Segmentation," 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 2021, pp. 2917-2927, doi: 10.1109/CVPR46437.2021.00294.

[56] Sudre CH, Li W, Vercauteren T, Ourselin S, Jorge Cardoso M. Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations. Deep Learn Med Image Anal Multimodal Learn Clin Decis Support (2017). 2017;2017:240-248. doi: 10.1007/978-3-319-67558-9_28. Epub 2017 Sep 9. PMID: 34104926; PMCID: PMC7610921.

[57] Lou, J., Xu, J., Zhang, Y., Sun, Y., Fang, A., Liu, J., Mur, L. A. J., & Ji, B. (2022).PPsNet: An improved deep learning model for microsatellite-instability high (MSI-H) prediction in colorectal cancer from whole-slide images. Computer Methods and Programs in Biomedicine, 225, 107095. https://doi.org/10.1016/j.cmpb.2022.107095

[58] Lee, S. H., Song, I. H., & Jang, H. J. (2021).

Feasibility of deep learning-based fully automated classification of microsatellite instability in tissue slides of colorectal cancer. International Journal of Cancer, 149(3), 728–740. https://doi.org/10.1002/ijc.33599

[59] Bilal et al. (2021)

Bilal, M., Chen, J., Irshad, H., Dong, F., Chen, D., Sahai, A., Zhao, Y., Raza, S. E. A., Azam, A., Graham, S., Rajpoot, N. (2021).

Development and validation of a weakly supervised deep learning framework to predict microsatellite instability from histopathology images in colorectal cancer. JAMA Network Open, 4(1), e210431. https://doi.org/10.1001/jamanetworkopen.2021.0431

[60] Guo, B., Li, X., Yang, M., Jonnagaddala, J., Zhang, H., & Xu, S. S. (2023). Predicting microsatellite instability and key biomarkers in colorectal cancer from H&E‐stained images: Achieving state‐of‐the‐art predictive performance with fewer data using Swin Transformer. The Journal of Pathology: Clinical Research, 9(3), 223–235. https://doi.org/10.1002/cjp2.312

[61] Chilukoti, S. V., Hossen, M. I., Shan, L., Tida, V. S., & Hei, X. (2025).

Differentially private fine-tuned NF-Net to predict GI cancer type. *arXiv preprint*, arXiv:2502.11329. https://doi.org/10.48550/arXiv