UNIVERSITE ABDELMALEK ESSAADI
FACULTE DES SCIENCES ET TECHNIQUES
DE TANGER
DEPARTEMENT GENIE INFORMATIQUE

Mémoire de Projet de Fin d'Études

MASTER SCIENCES ET TECHNIQUES
EN
SYSTÈMES INFORMATIQUES ET MOBILES

SUJET

ADVANCED VIDEO AND IMAGE PROCESSING AND COMPUTER VISION TECHNIQUES FOR ENHANCING ENDOSCOPIC PROCEDURE ANALYSIS:
A COMPREHENSIVE STUDY ON ANOMALY DETECTION, SEGMENTATION, AND REAL-TIME DECISION SUPPORT SYSTEMS

RÉALISÉ PAR : EL Houmaini Karim

Membres du Jury :

Pr. EL BRAK	Président
Pr. EN-NAIMI	Examinateur
Pr. ZOUHAIR	Encadrant FSTT
Kammas Chaimaa	Encadrant Entreprise

Année Universitaire 2023/2024

Abstract

This thesis focuses on the development and application of advanced image and videos processing and image analysis techniques specifically for medical endoscopy. The primary objective is to create a robust model for processing endoscopic images and videos recordings, incorporating several critical functionalities: pre-processing, classification, image segmentation to isolate regions of interest, object tracking to monitor anatomical structures over time, anomaly detection to identify signs of pathologies, and a user-friendly interface for practitioners.

In medical endoscopy, a significant challenge faced by doctors is the high percentage of undetected anomalies due to the complex and messy nature of endoscopic videos. This work aims to mitigate this issue by leveraging artificial intelligence. By collaborating with endoscopy specialists and a data science team at CHU Tangier, we developed a high-accuracy classification model alongside a segmentation model specifically for polyps. These models have demonstrated exceptional accuracy, significantly improving the detection rates of anomalies in endoscopic videos.

The models were created using advanced image processing and computer vision techniques, contributing to the enhancement of medical procedures by providing advanced analysis tools and efficient data management. The successful development and implementation of these models were acknowledged through acceptance and presentation at the National Congress of Endoscopy, and the work is slated for publication as a scientific article in collaboration with medical professionals.

This research represents a significant contribution to the field of medical imaging and computer-aided diagnosis, offering advanced tools to enhance the accuracy and efficiency of medical professionals in their diagnostic processes.

Résumé

Ce mémoire se concentre sur le développement et l'application de techniques avancées de traitement d'image et de vidéo spécifiquement pour l'endoscopie médicale. L'objectif principal est de créer un modèle robuste pour le traitement des images et enregistrements vidéo endoscopiques, incorporant plusieurs fonctionnalités critiques : prétraitement, classification, segmentation d'image pour isoler les régions d'intérêt, suivi d'objets pour surveiller les structures anatomiques au fil du temps, détection d'anomalies pour identifier les signes de pathologies, et une interface conviviale pour les praticiens.

En endoscopie médicale, un défi majeur auquel sont confrontés les médecins est le pourcentage élevé d'anomalies non détectées en raison de la nature complexe et désordonnée des vidéos endoscopiques. Ce travail vise à atténuer ce problème en tirant parti de l'intelligence artificielle. En collaboration avec des spécialistes de l'endoscopie et une équipe de science des données au CHU de Tanger, nous avons développé un modèle de classification de haute précision ainsi qu'un modèle de segmentation spécifiquement pour les polypes. Ces modèles ont démontré une précision exceptionnelle, améliorant significativement les taux de détection des anomalies dans les vidéos endoscopiques.

Les modèles ont été créés en utilisant des techniques avancées de traitement d'image et de vision par ordinateur, contribuant à l'amélioration des procédures médicales en fournissant des outils d'analyse avancés et une gestion efficace des données. Le développement et l'implémentation réussis de ces modèles ont été reconnus lors du Congrès National d'Endoscopie et le travail est en voie de publication en tant qu'article scientifique en collaboration avec des professionnels médicaux.

Cette recherche représente une contribution significative dans le domaine de l'imagerie médicale et du diagnostic assisté par ordinateur, offrant des outils avancés pour améliorer la précision et l'efficacité des professionnels de santé dans leurs processus diagnostiques.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AI: Artificial Intelligence

ASPP: Atrous Spatial Pyramid Pooling

AUC: Area Under the Curve

BCE: Binary Cross-Entropy

CAM: Class Activation Mapping

CHU: Centre Hospitalier Universitaire

CNN: Convolutional Neural Network

CRISP-DM: Cross Industry Standard Process for Data Mining

FSTT: Faculté des Sciences et Techniques de Tanger

Grad-CAM: Gradient-weighted Class Activation Mapping

IoU: Intersection over Union

ML: Machine Learning

ROC: Receiver Operating Characteristic

UNet: U-shaped Convolutional Neural Network

YOLO: You Only Look Once

HTML: HyperText Markup Language

BI: Business Intelligence

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisors, Pr. Zouhair Abdelhamid and Kammas Chaimaa, whose expertise, understanding, and patience added considerably to my graduate experience. Their guidance and support throughout this journey have been invaluable.

I extend my sincere thanks to the entire team at CHU Tangier for their collaborative spirit and insightful feedback. Their dedication and hard work made this project possible. I am also grateful to Dr. Sara Salhani, whose clinical insights and practical knowledge were instrumental in shaping this research.

Special thanks to my advisors, Pr. EN-NAIMI and Pr. EL BRAK, for their exceptional academic courses that laid the foundation for my success in this project. Their teaching provided me with the necessary knowledge and skills to tackle the challenges encountered during this research. Their courses were instrumental in shaping my understanding and approach, enabling the completion of this work.

I would like to acknowledge the support of the Faculty of Sciences and Techniques at Tangier (FSTT) for providing the resources and environment necessary for my research. The administrative and technical staff at FSTT deserve special mention for their assistance and cooperation.

My heartfelt appreciation goes to my family and friends for their unwavering support and encouragement. Their understanding and love have been a source of strength throughout this challenging journey.

Lastly, I would like to thank all my colleagues and peers whose camaraderie made this experience enjoyable and memorable. Their discussions, feedback, and shared experiences have been invaluable to my personal and academic growth.

Thank you all for your contributions, guidance, and support in making this project a success.

Chapter 1: Introduction

1.1 BACKGROUND OF ENDOSCOPY IN MODERN MEDICINE

Endoscopy has revolutionized medical diagnosis and treatment, allowing physicians to visualize and interact with internal organs and tissues in a minimally invasive manner. However, the field now stands on the cusp of a new revolution - one driven by artificial intelligence (AI) and advanced computer vision techniques. This thesis presents a groundbreaking project that aims to redefine the standards of endoscopic procedures through the integration of cutting-edge AI technologies.

1.2 THE CRITICAL ROLE OF ENDOSCOPY IN MODERN HEALTHCARE

Gastrointestinal cancers account for a staggering 26% of global cancer incidence and 35% of all cancer-related deaths, posing a severe threat to global health. The stark contrast in survival rates between early and late-stage diagnoses underscores the critical importance of timely detection. For instance, the 5-year survival rates for early gastric cancer exceed 90%, while for advanced stages, they plummet below 30%.

Endoscopy stands as the most effective approach for detecting gastrointestinal tumors, with over 100 million subjects receiving gastrointestinal endoscopy examinations each year globally. However, the current landscape of endoscopic diagnostics faces significant challenges:

1) Low early diagnosis rates
2) Missed lesions due to the vast area of gastrointestinal mucosa
3) Subtle characteristics of early-stage lesions
4) Existence of blind zones during examinations
5) Variability in examination quality among different endoscopists

These challenges underscore the urgent need for innovative solutions to enhance the quality and accuracy of endoscopic examinations.

1.3 THE AI REVOLUTION IN ENDOSCOPY

The integration of AI in medical imaging has shown remarkable promise, particularly in the field of endoscopy. A comprehensive bibliometric analysis of publications from 1990 to 2022 reveals a dramatic surge in research interest and technological advancements in this area:

- The number of publications on AI in endoscopy peaked in 2021, accounting for 23.5% of all publications in the field since 1990.
- Annual citations in this field have skyrocketed from less than 100 before 2010 to over 2000 after 2020.
- The field's H-index rapidly increased from less than 10 before 2016 to a peak of 28 in 2019.

These statistics highlight the exponential growth and increasing impact of AI in endoscopy research.

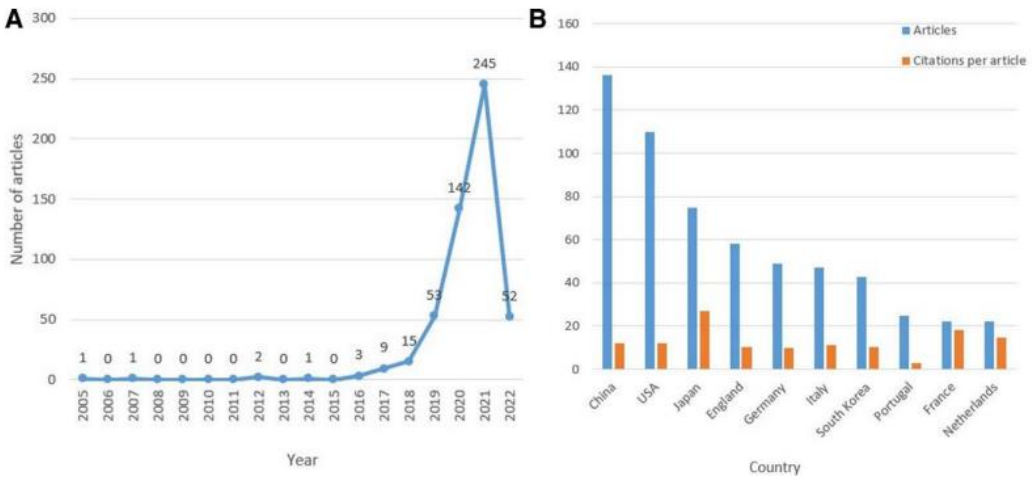

Figure 1 Trends in the number of publications and analysis of countries in artificial intelligence in digestive endoscopy field. (A) The annual worldwide publication output. (B) The total number of publications and citations per article for the top 10 countries

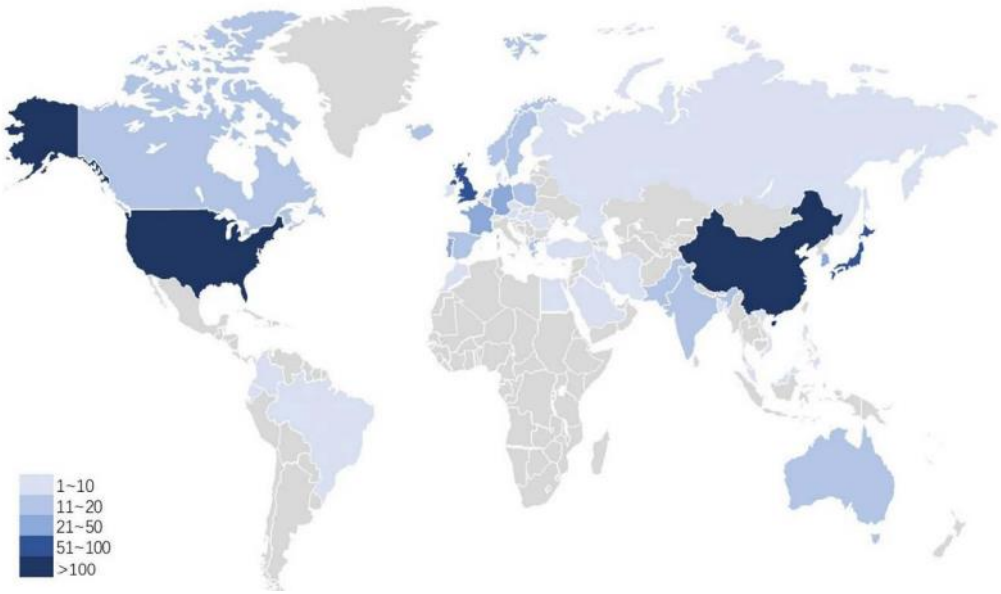

Figure 2 Worldwide distributions of the Web of Science Core Collection publications on artificial intelligence in digestive endoscopy field

1.4 PROJECT SIGNIFICANCE AND INNOVATION

Our project stands at the forefront of this technological revolution, offering a comprehensive and innovative approach to AI-assisted endoscopy. Unlike previous efforts that focused on narrow aspects of endoscopic image analysis, our project presents a holistic solution that addresses multiple challenges:

1) Advanced Image Processing: We employ state-of-the-art computer vision techniques to enhance image quality, reducing noise and improving clarity in real-time.

2) Multi-Modal Disease Detection: Our AI model is trained on a diverse dataset, enabling it to detect a wide range of gastrointestinal abnormalities, from early-stage cancers to subtle mucosal changes.

3) Real-Time Analysis and Feedback: The system provides instantaneous feedback to endoscopists, highlighting areas of concern and suggesting optimal viewing angles.

4) Adaptive Learning: Our AI continuously learns from new data, improving its accuracy over time and adapting to different patient populations.

Integrated Workflow Solution: Beyond mere detection, our system seamlessly integrates into existing clinical workflows, enhancing efficiency without disrupting established procedures.

Recent studies have shown promising results in AI-assisted endoscopy:

- The accuracy of AI in detecting Barrett's esophagus reached 87.6% between 2018 and 2022.
- For colorectal polyps, AI achieved an impressive accuracy of 93.7% in the same period.
- In gastric cancer detection, AI demonstrated an accuracy of 88.3%.

Our project aims to not only match but exceed these benchmarks across a broader range of gastrointestinal conditions.

1.5 GLOBAL IMPACT AND FUTURE DIRECTIONS

The potential impact of our project extends beyond improving diagnostic accuracy. By enhancing early detection rates, we aim to significantly reduce the global burden of gastrointestinal cancers. Moreover, our system's ability to standardize examination quality could help address healthcare disparities, ensuring high-quality endoscopic examinations even in resource-limited settings.

As we stand on the brink of a new era in endoscopic diagnostics, our project represents not just an incremental improvement, but a paradigm shift in how we approach gastrointestinal health. By harnessing the power of AI and advanced computer vision, we are poised to usher in a future where missed diagnoses become a rarity, and early intervention becomes the norm.

1.6 RESEARCH QUESTIONS AND OBJECTIVES

To address these challenges, this thesis aims to answer the following research questions:

- How can advanced AI and computer vision techniques be applied to improve the detection of anomalies in endoscopic video and image recordings?

- What are the practical implications of integrating these AI-assisted diagnostic tools into clinical workflows?

The main objectives of this research are:

- To develop a robust AI model for processing endoscopic image and video recordings, incorporating functionalities such as pre-processing, image segmentation, object tracking, and anomaly detection.
- To validate these models through collaboration with medical professionals, ensuring their practical applicability and effectiveness in clinical settings.
- To assess the potential impact of these AI-assisted tools on diagnostic accuracy, efficiency, and overall patient care.

1.7 SCOPE AND LIMITATIONS

This research focuses on the application of advanced image processing techniques and AI algorithms to medical imaging, specifically in the domains of endoscopy. The scope includes:

- Development and validation of models for endoscopic video and image analysis.
- Focus on real-time processing and accuracy of detection.
- Collaboration with medical professionals for data annotation and model validation.

The work excludes:

- Development of hardware systems or medical devices.
- Clinical trials or direct patient interventions.
- Analysis of other medical imaging modalities beyond endoscopy and CT scans.

1.8 METHODOLOGY OVERIEW

The methodology involves several key steps:

- Data Collection: Gathering large datasets of endoscopic videos and images, annotated by medical professionals.
- Model Development: Designing and training AI models using state-of-the-art machine learning and deep learning techniques, including computer vision algorithms.
- Validation: Collaborating with medical experts to validate the models in real-world clinical settings, ensuring their accuracy and reliability.
- Deployment: Integrating the models into user-friendly software interfaces for use by clinicians.

1.9 SIGNIFICANCE OF THE STUDY

The successful development of these AI models can significantly enhance the capabilities of medical imaging, improving diagnostic accuracy and patient outcomes. By providing clinicians with advanced tools for real-time analysis and decision support, this research has the potential to revolutionize endoscopic practices, reducing the rate of missed anomalies and enabling more timely and accurate diagnoses.

1.10 COLLABORATIONS AND CONTRIBUTIONS

This research is conducted in collaboration with the endoscopy department at CHU Tangier, leveraging their expertise and clinical data to develop and validate the models. The project has received recognition through acceptance and presentation at the National Congress of Endoscopy and is slated for publication as a scientific article in collaboration with medical professionals.

Figure 3 SMED Société Marocaine d'Endoscopie Digestive: une Société Scientifique au service de la profession et au service des patients atteints de maladies de l'appareil digestif

1.11 COMPANY DESCRIPTION

CHU Mohammed VI de Tanger

Figure 4 CHU Mohammed VI de Tanger

CHU Mohammed VI de Tanger serves the population of the Tanger-Tétouan-Al Hoceima region, which spans 17,262 km² with a population of 3,557,000 (according to RGPH 2014). Located in the extreme northwest of Morocco, it is bordered by the Strait of Gibraltar and the Mediterranean to the north, the Atlantic Ocean to the west, the Rabat-Salé-Kenitra region to the southwest, the Fès-Meknès region to the southeast, and the Oriental region to the east. The region includes two prefectures, Tanger-Assilah and M'Diq-Fnideq, and six provinces: Al Hoceima, Chefchaouen, Fahs-Anjra, Larache, Ouezzane, and Tétouan.

Status and Structure of CHU Mohammed VI of Tangier

CHU Mohammed VI of Tangier is a public institution endowed with legal personality and financial autonomy. It operates under the supervision of the Ministry of Health, established by Law 82.00 and promulgated by Dahir 1.01.206 on August 30, 2001, which amended Law 37.80 related to hospital centers.

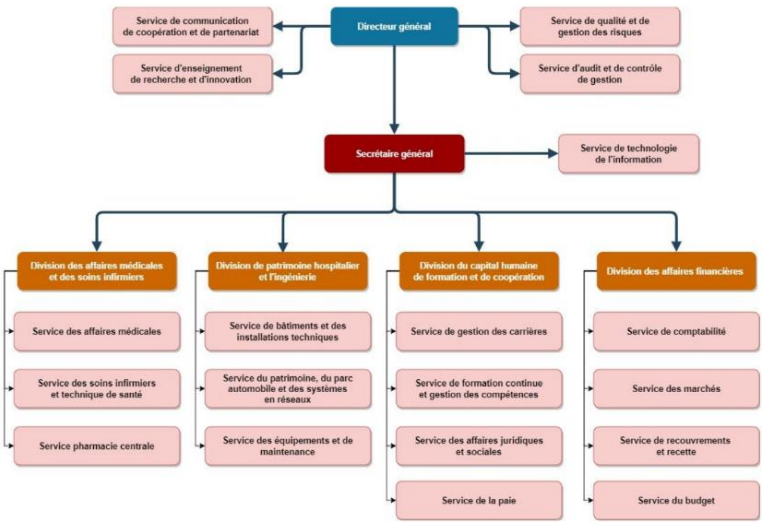

Figure 5 Structure of CHU Mohammed VI de Tanger

Missions of CHU Mohammed VI de Tanger

CHU Mohammed VI of Tangier missions span four main areas:

- Healthcare: Provides tertiary and secondary medical and surgical services, both in emergencies and scheduled activities.
- Education: Partners with the Faculty of Medicine and Pharmacy, the Institute for Training in Health Careers, and other public and private educational structures to deliver clinical and post-graduate medical and pharmaceutical education, and practical training for nursing staff.
- Research: Conducts medical and nursing research in collaboration with the Faculty of Medicine and Pharmacy, the Institute for Training in Health Careers, and other national or international training and research units.
- Public Health: Contributes to achieving health policy objectives set by the Ministry of Health.

Values of CHU Mohammed VI de Tanger

The core values of CHU Mohammed VI de Tanger are:

- Innovation: Encourages innovation in all activities to improve services for patients, families, and students.
- Integrity: Maintains transparency and integrity to earn the trust of patients, families, and the public.
- Ethics: Ensures all nursing, research, and teaching practices respect medical ethical principles.
- Commitment: Strives for collective success through solidarity, collaboration, and shared contributions towards common goals.
- Quality: Aims for the highest standards in nursing practice, patient and student relations, and professional interactions.

Key Partnerships

CHU Mohammed VI de Tanger collaborates with several institutions to enhance its capabilities and service quality, including:

- Faculty of Medicine and Pharmacy of Tangier

- Ministry of Health
- University Abdelmalek Essaadi
- University Hospitals of Geneva
- Various Belgian university hospitals

IT Department

The IT department of CHU Mohammed VI de Tanger plays a crucial role in managing and developing the hospital's information systems and technologies. Responsibilities include:

- Managing technological infrastructure
- Developing and maintaining applications
- Providing technical support
- Ensuring data security
- Overseeing technological projects

The department supports a comprehensive hospital information system that facilitates activities such as electronic medical record management, appointment scheduling, pharmacy management, and surgical planning.

1.12 STRUCTURE OF THE THESIS

The thesis is structured as follows:

- Chapter 1: Introduction
- Chapter 2: Literature Review
- Chapter 3: Methodology
- Chapter 4: Realization and Perspectives
- Chapter 5: Conclusion

Chapter 2: Literature Review

2.1 INTRODUCTION

Medical imaging has revolutionized healthcare by providing clinicians with powerful tools to diagnose, monitor, and treat a wide array of medical conditions. Among these modalities, endoscopic video recordings are pivotal in gastrointestinal diagnostics. This literature review explores the advancements in image processing and computer vision techniques applied to endoscopic imaging, focusing on improving diagnostic accuracy and patient care. The review is structured around the key areas of endoscopic analysis, AI advancements in endoscopy, and the integration of AI technologies in these domains.

2.2 ENDOSCOPIC VIDEO ANALYSIS

2.2.1 Challenges in Endoscopy

Endoscopy is a minimally invasive procedure used to visualize and examine the interior of organs and cavities. Despite its significance in diagnosing gastrointestinal conditions and identifying anomalies such as polyps, endoscopic procedures face several challenges. The dynamic nature of endoscopic video footage and the high variability in anatomical structures often result in missed anomalies. Studies have shown that a significant percentage of precancerous polyps are overlooked during colonoscopy screenings due to limitations in human visual perception and the quality of the video feed.

2.2.2 Advancements in AI for Endoscopy

Recent advancements in AI and ML have opened new avenues for enhancing endoscopic procedures. Notably, AI can improve detection rates and visibility during endoscopy by integrating real-time sensing and AI algorithms into the video feed. NVIDIA's research highlights the potential of AI systems to assist clinicians by identifying and classifying anomalies, tracking tools, and segmenting organs. These systems can augment the surgeon's capabilities and reduce their workload, thereby improving the overall efficiency and accuracy of endoscopic procedures.

2.3 TECHNIQUES AND MODELS

2.3.1 Techniques for Localization and Classification

To enhance the precision of anomaly detection in endoscopic videos, various techniques were explored, CNN, EfficientNetb2, EfficientNetb4 and other techniques like Grad-CAM, CAM, SmoothGrad, and Guided Backpropagation. These techniques aimed to identify the regions in the image that were most influential in the model's prediction.

Grad-CAM (Gradient-weighted Class Activation Mapping): This technique uses the gradients of the target concept, flowing into the final convolutional layer to produce a coarse localization map highlighting important regions in the image.

$$L^c_{\text{Grad-CAM}} = \text{ReLU}\left(\sum_k \alpha^c_k A^k\right)$$

Where $\alpha^c_k = \frac{1}{Z}\sum_i \sum_j \frac{\partial y^c}{\partial A^k_{ij}}$

Class Activation Mapping (CAM): CAM is a method to identify discriminative image regions used by a CNN to identify a specific class. This technique, however, requires a specific architecture that ends with global average pooling.

$$L^c_{\text{CAM}} = \sum_k w^c_k A^k$$

SmoothGrad: This technique improves visual explanations by adding noise to the input image and computing the average of the resulting gradients.

$$\text{SmoothGrad} = \frac{1}{n}\sum_{i=1}^{n} \text{Grad}\left(x + N(0, \sigma^2)\right)$$

Guided Backpropagation: This technique visualizes the contribution of each input pixel to the activation of specific output neurons, using modified backpropagation rules.

These techniques provided insights into the areas the model focused on for making predictions. However, they did not achieve the desired accuracy and precision, prompting a shift towards more advanced object detection and segmentation models.

2.4 OBJECT DETECTION TECHNIQUES

Initial attempts using object detection models like YOLOv8 aimed to localize anomalies but were limited to bounding boxes, lacking precise delineation. The YOLO (You Only Look Once) framework is designed for real-time object detection and has been effective in many scenarios.

$$\text{YOLO Loss} = \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{\text{obj}} [(x_i - \hat{x_\iota})^2 + (y_i - \hat{y_\iota})^2] + \lambda_{\text{conf}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{\text{obj}} \left(C_i - \hat{C_\iota}\right)^2$$

$$+ \lambda_{\text{class}} \sum_{i=0}^{S^2} 1_i^{\text{obj}} \sum_{c \in \text{classes}} \left(p_i(c) - \hat{p_\iota}(c)\right)^2$$

Despite its strengths, YOLOv8's bounding boxes were insufficient for the precise localization required in endoscopic video analysis, leading to the adoption of segmentation models.

2.4.1 Segmentation Model

For precise localization of abnormalities, segmentation models were utilized, including SegNet, Transformer-based models, and DeepLabv3+.

2.3.3.1 Transformer-based Segmentation Model

A custom Transformer-based model was developed, incorporating positional encoding and multihead self-attention mechanisms to capture spatial relationships in images.

2.3.3.2 DeepLabv3+

DeepLabv3+ with a ResNet50 backbone was selected for its ability to handle high-resolution images and complex segmentation tasks, employing atrous spatial pyramid pooling (ASPP) for multi-scale context aggregation.

$$\text{ASPP}(x) = [\text{conv}_1(x), \text{conv}_3(x, 6), \text{conv}_3(x, 12), \text{conv}_3(x, 18)]$$

2.5 COMPARATIVE ANALYSIS

A comparative analysis of different models and techniques was conducted to determine the most effective approaches for classification and segmentation tasks. EfficientNetB4 outperformed other models in classification accuracy, while DeepLabv3+ with ResNet50 backbone showed superior results in segmentation tasks. The comparative evaluation also included techniques like Grad-CAM for visualizing model predictions and YOLOv8 for object detection, though these did not achieve the same level of performance as the chosen models.

2.6 CONCLUSION

The literature review highlights the significant advancements in AI and ML for medical imaging, particularly in the areas of endoscopic video analysis. The integration of models like EfficientNetB4 and DeepLabv3+ has the potential to revolutionize medical diagnostics by improving accuracy and reducing the workload on clinicians. Future research should focus on further refining these models and exploring their applicability in real-world clinical settings.

Chapter 3: Methodology

3.1 INTRODUCTION

This chapter delineates the methodologies employed in the development, training, and validation of advanced AI models for the analysis of endoscopic images, with a focus on the detection and classification of polyps, esophagitis, and ulcers. The process involves several stages: data collection, preprocessing, model development, training, evaluation, and deployment. The structured framework ensures systematic execution, aiming for high accuracy and practical applicability in clinical settings.

3.2 RESEARCH FRAMEWORK

The research methodology follows the Cross Industry Standard Process for Data Mining (CRISP-DM), encompassing six phases:

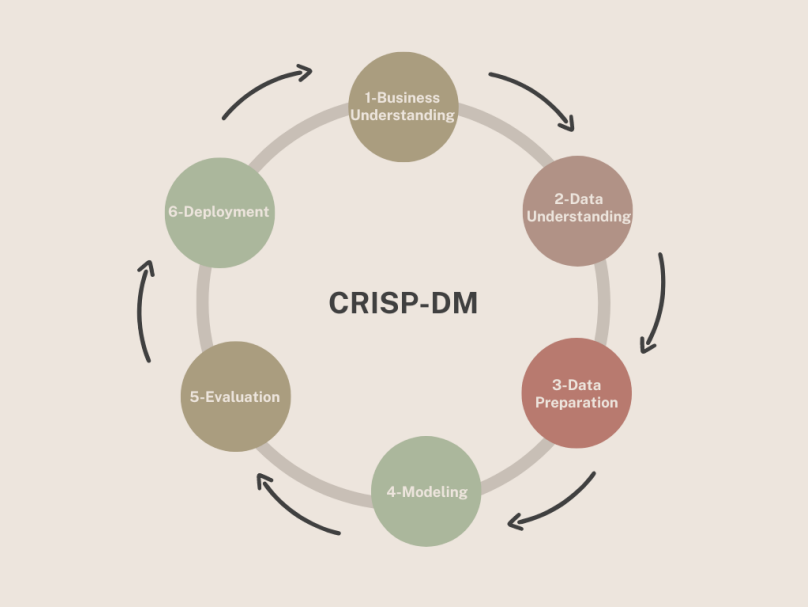

Figure 6 CrispDM Methodoloy

3.2.1 Business understanding

The first phase focuses on understanding the project objectives and requirements from a business perspective. In the context of this thesis, the primary goal is to develop AI models that can enhance the accuracy and efficiency of diagnosing medical conditions through endoscopic and radiological imaging.

Objectives:

- Improve detection rates and visibility in endoscopic procedures.
- Provide real-time decision support to clinicians.

3.2.2 Data understanding

This phase involves collecting initial data, familiarizing with the data, and identifying data quality issues. It is crucial to understand the characteristics and structure of the data to inform subsequent steps in the process.

Data Sources:

- Endoscopic images and video recordings from clinical procedures.

  Data includes videos capturing the interior of organs and cavities during endoscopic procedures.
  Anomalies such as polyps and lesions are annotated by experienced clinicians.

- Annotated datasets from publicly available sources like Kaggle.

3.2.3 Data Preparation

Data preparation involves cleaning and transforming raw data into a format suitable for modeling. This step is essential to ensure the quality and consistency of the data used for training AI models.

Preprocessing Techniques:

Histogram Equalization and Smoothing: These techniques are used for image enhancement, improving the contrast and reducing noise in the images.

$$I_{\text{equalized}} = \text{HistogramEqualization}(I)$$

Conversion of Masks to Binary Format: Essential for segmentation tasks, this step involves converting pixel values to binary (0 or 1) for clear delineation of structures.

$$M_{\text{binary}} = \text{BinaryConversion}(M)$$

Data Augmentation: Techniques such as rotation, shear, zoom, and flip are applied using Keras' *ImageDataGenerator* to increase the diversity of the training dataset.

$$\text{AugmentedImage} = \text{ImageDataGenerator}(\text{rotation\_range} = 15, \text{shear\_range} = 0.2, \text{zoom\_range} = 0.2, \text{horizontal\_flip} = \text{True})$$

Data Splitting:

- **Training Set:** Used to train the models.
- **Validation Set:** Used to tune the models and validate their performance during training.
- **Test Set:** Used to evaluate the final model performance.

3.2.4 Modelling

In the modelling phase, various algorithms are selected and applied to the prepared data. The models are then fine-tuned to achieve the best possible performance.

Model Selection:

EfficientNetB4 for classification:

EfficientNetB4 is chosen for its state-of-the-art performance in image recognition tasks, leveraging compound scaling to balance network depth, width, and resolution.

$$\text{EfficientNet Scaling} = d \cdot w \cdot r$$

DeepLabv3+ with ResNet50 Backbone for Segmentation:

DeepLabv3+ is used for its ability to handle high-resolution images and complex segmentation tasks, employing atrous spatial pyramid pooling (ASPP) for multi-scale context aggregation.

$$\text{ASPP}(x) = [\text{conv}_1(x), \text{conv}_3(x, 6), \text{conv}_3(x, 12), \text{conv}_3(x, 18)]$$

Model Architecture:

EfficientNetB4

ficientNetB4 employs a combination of MBConv layers, Squeeze-and-Excitation optimization, and the Swish activation function to enhance performance.

$$\text{Swish}(x) = x \cdot \sigma(x)$$

DeepLabv3+

Utilizes a ResNet50 backbone with ASPP for multi-scale feature extraction and effective segmentation.

$$\text{Output} = \text{Decoder}\left(\text{ASPP}\left(\text{Encoder}(x)\right)\right)$$

3.2.5 Evaluation

The evaluation phase assesses the performance of the models against a set of predefined metrics. This step ensures that the models meet the business objectives and can be reliably deployed in clinical settings.

Evaluation Metrics:

Classification accuracy

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$$

Confusion Matrix:

A table used to describe the performance of a classification model.

Dice coefficient:

$$\text{Dice Coefficient} = \frac{2|X \cap Y|}{|X| + |Y|}$$

Jaccard Index:

$$\text{Jaccard Index} = \frac{|X \cap Y|}{|X \cup Y|}$$

Loss Value

Loss functions such as categorical cross-entropy for classification and binary cross-entropy for segmentation.

Validation:

- **Cross-Validation Techniques:** Ensures robustness and generalizability of the models by splitting the data into multiple subsets and training/testing on these subsets.
- **Comparison with Baseline Models:** Demonstrates improvements by benchmarking against simpler models and techniques.

3.3 DATA COLLECTION

Data collection is a critical step in the development of robust and reliable AI models, particularly in the field of medical imaging. For this thesis, data was gathered from multiple sources to ensure a diverse and comprehensive dataset. The collected data includes endoscopic video recordings annotated by medical professionals. This section provides a detailed account of the data collection process, including the sources, types of data, annotation processes, and tools used.

3.3.1 Data Sources

The data for this research was collected from two primary sources: clinical collaborations and online repositories. These sources provided a rich and varied dataset necessary for training and validating the AI models.

Clinical Collaborations:

Hospitals and Clinics:

- Collaborations with the endoscopy departments at CHU Tangier.
- Data collected includes endoscopic video recordings of patients with various medical conditions.
- Medical professionals annotated the data, identifying key features such as anomalies in endoscopic videos.

Ethical Considerations:

- All data collection activities were conducted in accordance with ethical guidelines and regulations.
- Patient consent was obtained for the use of their medical images in research.
- Data was anonymized to protect patient privacy and confidentiality.

Online Repositories:

Additional datasets were sourced from publicly available repositories and research publications.

- These datasets were selected based on their relevance and quality, ensuring they met the research requirements.
- Examples include the KVASIR dataset and curated colon datasets.

3.3.2 Types of Data Collected

The data collected for this research can be categorized into endoscopic video recordings. Each type of data required specific handling and preprocessing to prepare it for model training.

Endoscopic Video and Images Recordings:

Content:

- Videos and Images capturing the interior of organs and cavities during endoscopic procedures.
- Commonly focused on the gastrointestinal tract, including colonoscopy and gastroscopy procedures.

Challenges:

- High variability in anatomical structures.

- Presence of noise and artifacts due to movement and lighting conditions.

3.3.3 Data Annotation Processes

Accurate annotation of medical images is crucial for training AI models. The annotation process involved collaboration with medical professionals to ensure the reliability and accuracy of the labels.

3.4 DATA PREPROCESSING

Data preprocessing is a crucial step in ensuring that the collected raw data is transformed into a format suitable for model training and evaluation. This process enhances the quality and consistency of the data, addressing issues such as noise, variability, and missing values. The preprocessing steps for this research are tailored to the specific requirements of endoscopic video analysis, with a focus on image enhancement, data augmentation, and preparation for model input.

3.4.1 Image Preprocessing

Histogram Equalization and Smoothing

Histogram equalization is used to improve the contrast of images by redistributing the intensity values. This technique helps in highlighting the features that are important for the detection of anomalies such as polyps and lesions in endoscopic images.

$$I_{\text{equalized}}(x,y) = \frac{C\big(I(x,y)\big) - C_{min}}{C_{max} - C_{min}} \cdot (L-1)$$

where I(x,y) is the original pixel intensity at position (x,y), C is the cumulative distribution function of the image histogram, C{min} and C{max} are the minimum and maximum values of C, and L is the number of possible intensity levels.

Smoothing, typically achieved through Gaussian filtering, reduces noise and improves the quality of the images. The Gaussian filter is defined as:

$$G(x,y) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)$$

where σ is the standard deviation of the Gaussian distribution.

Mathematically, histogram equalization can be expressed as:

$$H(v) = \frac{MN}{L-1}\sum_{u=0}^{v} h(u)$$

where H(v)H(v)H(v) is the histogram equalized value, LLL is the number of intensity levels, MMM and NNN are the dimensions of the image, and h(u)h(u)h(u) is the histogram of the pixel values.

Conversion of Masks to Binary Format

For segmentation tasks, masks are converted to binary format, which involves transforming the pixel values into binary (0 or 1) based on a threshold. This process is crucial for delineating structures clearly in the segmentation task.

$$\mathrm{M}_{\mathrm{binary}}(\mathrm{x},\mathrm{y}) = \begin{cases} 1 \text{ if } \mathrm{M}(\mathrm{x},\mathrm{y}) > \mathrm{T} \\ 0 \text{ if } \mathrm{M}(\mathrm{x},\mathrm{y}) \leq \mathrm{T} \end{cases}$$

where M(x,y) is the original mask value at position (x,y) and T is the threshold value.

Data Augmentation

Data augmentation increases the diversity of the training dataset by applying various transformations. This helps in making the model more robust and less prone to overfitting. The following augmentations were applied using Keras *ImageDataGenerator*:

$$\begin{aligned}\mathrm{AugmentedImage} \\ &= \mathrm{ImageDataGenerator}(\mathrm{rotation\_range} = 15,\mathrm{shear\_range} \\ &= 0.2,\mathrm{zoom\_range} = 0.2,\mathrm{horizontal\_flip} = \mathrm{True})\end{aligned}$$

These transformations include:

- Rotation: Rotating the image by a random degree within a specified range.

- Shear: Shearing the image to create affine transformations.
- Zoom: Randomly zooming in or out of the image.
- Horizontal Flip: Flipping the image horizontally.

Mathematically, data augmentation can be expressed as a transformation T applied to an image I:

$$I' = T(I)$$

where I′ is the augmented image.

Data Splitting

The dataset is divided into training, validation, and test sets to ensure unbiased evaluation of model performance. The splitting is done in a stratified manner to maintain the distribution of classes across all sets.

$$\text{Train, Val, Test Split} = \text{train\_test\_split}(\text{data}, \text{test\_size} = 0.2, \text{validation\_size} = 0.1, \text{stratify} = \text{labels})$$

3.4.2 Data Preparation for Model Training

A robust data pipeline is established to efficiently load and preprocess the data during training. TensorFlow's tf.data API is used to create a scalable and efficient data pipeline.

Loading and Preprocessing Pipeline

To avoid including actual code, the following explanation details the steps and processes involved:

- **Loading Data:** The images and their corresponding labels are loaded from the dataset. The dataset includes file paths to the images and their respective labels.

- **Resizing and Normalizing:** Each image is resized to a uniform size (e.g., 224x224 pixels) to maintain consistency across the dataset. Normalization is performed to scale pixel values to the range [0, 1].
- **Shuffling and Batching:** The dataset is shuffled to ensure that the model does not learn any order-specific patterns. The data is then batched into smaller subsets to optimize training efficiency.
- **Prefetching:** To ensure that data loading does not become a bottleneck during training, prefetching is used. This technique loads the next batch of data while the current batch is being processed, thus overlapping data loading and model training.

Mathematical Representation of Preprocessing Steps

Resizing: Given an image I of size m×n, resizing transforms it to a new size p×q. The transformation can be represented as:

$$I' = \text{resize}\big(I, (p, q)\big)$$

Normalization: Each pixel value x in the image I′ is scaled to the range [0, 1] using the formula:

$$x' = \frac{x}{255}$$

Shuffling: The dataset is randomly permuted to ensure that each training epoch sees the data in a different order. If D is the dataset and S is the shuffle operation, the shuffled dataset D′ can be represented as:

$$D' = S(D)$$

Batching: The dataset is divided into batches of size b. If the dataset contains N samples, the number of batches k is given by:

$$k = \left\lceil \frac{N}{b} \right\rceil$$

Prefetching: Prefetching overlaps the preprocessing of the next batch with the training of the current batch. This can be represented as:

$$\text{prefetch}(D', b)$$

These preprocessing steps ensure that the data fed into the models is consistent, diversified, and optimized for efficient training.

3.5 MODEL DEVELOPMENT

3.5.1 Classification Models

Simple CNN

The Simple CNN was implemented as a baseline model to classify endoscopic images into different categories. The architecture comprised several convolutional layers followed by max-pooling layers, a flattening layer, and fully connected dense layers. The primary objective was to establish a reference for comparison with more advanced models.

EfficientNetB2

Another classification model tested was based on EfficientNetB2, a lightweight model from the EfficientNet family known for its efficiency and performance. The EfficientNetB2 model was pre-trained on the ImageNet dataset and then fine-tuned on our endoscopic image dataset. This model was chosen for its ability to balance accuracy and computational efficiency, making it suitable for applications with limited resources.

Model Architecture

The architecture of EfficientNetB2 consists of the following layers:

- EfficientNetB2 Base Model: The core of the architecture, pre-trained on ImageNet, extracts high-level features from the input images. This model comprises several MBConv (Mobile Inverted Bottleneck Convolution) blocks, utilizing depthwise separable convolutions and squeeze-and-excitation optimization to reduce the number of parameters and improve efficiency.
- Gaussian Noise Layers: These layers are added to introduce slight randomness during training, which helps in regularizing the model and preventing overfitting.

- Global Average Pooling: This layer reduces the spatial dimensions of the feature maps by averaging, resulting in a single vector per feature map. This vector represents the presence of features in the image.
- Fully Connected (Dense) Layers: The dense layers further process the features extracted by the EfficientNetB2 base model. The architecture includes:
    - A Dense layer with 256 units, which applies a linear transformation to the input features.
    - Batch Normalization to stabilize and accelerate training.
    - Dropout to prevent overfitting by randomly setting a fraction of input units to zero during training.
    - A final Dense layer with 4 units, corresponding to the number of classes in the classification task, using the softmax activation function to output class probabilities.

The architecture of EfficientNetB2 is summarized in the table below:

Layer (Type)	**Output Shape**	**Param #**
efficientnetb2 (Functional)	(None, 7, 7, 1408)	7,768,569
gaussian_noise_4	(None, 7, 7, 1408)	0
global_average_pooling2d_2	(None, 1408)	0
dense_4	(None, 256)	360,704
batch_normalization_205	(None, 256)	1,024
gaussian_noise_5	(None, 256)	0
dropout_2	(None, 256)	0
dense_5	(None, 4)	1028
Total		8,131,325
Trainable Params		362,244

Non-Trainable Params		7,769,081

Table 1 The architecture of EfficientNetB2

EfficientNetB4 : chosen Model

The EfficientNetB4 model was selected for its superior performance in image recognition tasks, making it well-suited for the classification of endoscopic images. EfficientNet models use a technique called compound scaling, which uniformly scales the depth, width, and resolution of the network. This method balances the trade-offs between these parameters, leading to better performance with fewer computational resources.

Model Architecture

EfficientNetB4 is built upon the EfficientNet architecture, which employs the following key components:

1. **MBConv Blocks**: These are Mobile Inverted Bottleneck Convolutional layers that use depthwise separable convolutions to reduce the number of parameters and computations. Each MBConv block consists of:
   - Depthwise Convolution: Applies a single convolutional filter per input channel, which significantly reduces the number of parameters and computations compared to standard convolutional layers.
   - Pointwise Convolution: Applies a 1x1 convolution to combine the outputs from the depthwise convolution, allowing for the recombination of features across channels.
   - Squeeze-and-Excitation Optimization: Introduces a mechanism to recalibrate channel-wise feature responses by explicitly modeling interdependencies between channels. This optimization enhances the representational capacity of the network by dynamically adjusting the importance of each channel.

2. **Compound Scaling**: EfficientNet uses a compound scaling method that uniformly scales network dimensions using a set of fixed scaling coefficients. The scaling method can be described as:

$$d = \alpha^k, \quad w = \beta^k, \quad r = \gamma^k$$

where d, w, and r represent the depth, width, and resolution of the network respectively, and α, β, and γ are constants determined through a grid search. The scaling factor k is chosen based on available computational resources, allowing the model to efficiently scale its architecture to meet the needs of different tasks and hardware constraints.

3. **Swish** Activation **Function:** The Swish activation function, defined as:

$$\text{Swish}(x) = x \cdot \sigma(x)$$

where σ(x) is the sigmoid function, introduces non-linearity while maintaining smooth gradients. This helps in efficient training and improved performance by allowing the network to learn more complex patterns and relationships in the data.

4. **Batch** Normalization**:** Batch normalization layers are included after each convolutional layer to stabilize and accelerate training. The output of batch normalization is given by:

$$\hat{x} = \frac{x - \mu}{\sqrt{\sigma^2 + \epsilon}}$$

where μ and σ^2 are the mean and variance of the input x, and ϵ is a small constant to prevent division by zero. This normalization process helps in maintaining the stability of the training process and allows the use of higher learning rates.

The architecture of EfficientNetB4 is summarized in the table below:

Layer (Type)	**Output Shape**	**Param #**
efficientnetb4 (Functional)	(None, 1792)	17,673,823

batch_normalization (Batch)	(None, 1792)	7,168
dense (Dense)	(None, 256)	459,008
dropout (Dropout)	(None, 256)	0
dense_1 (Dense)	(None, 4)	1,028
Total params		18,141,027
Trainable params		18,012,236
Non-trainable params		128,791

Table 2 The architecture of EfficientNetB4

Training Process

The training process for EfficientNetB4 involved several critical steps to ensure optimal performance:

1. **Data Augmentation:** To enhance the model's generalization ability, various data augmentation techniques were applied, including:

- Horizontal Flip
- Random Rotation
- Scaling
- Shearing
- Zooming

2. **Data Normalization**: Each pixel value in the images was normalized to the range [0, 1] by dividing by 255. This ensures consistent input for the model and helps in stabilizing the training process.
3. **Optimizer and Loss Function**: The Adamax optimizer, a variant of the Adam optimizer, was chosen for its stability and efficiency in handling sparse gradients. The categorical cross-entropy loss function was used to measure the performance of the model during training. The loss function is defined as:

$$L_{\text{CE}} = -\sum_{i=1}^{N} y_i \log(p_i)$$

where yi is the ground truth label and pi is the predicted probability for class i.

4. **Learning Rate Scheduling**: A dynamic learning rate schedule was implemented using a custom callback. The learning rate was adjusted based on the performance of the model on the validation set, ensuring faster convergence and prevention of overfitting.

Model Interpretability

To ensure the interpretability of the EfficientNetB4 model, several techniques were employed:

- Grad-CAM (Gradient-weighted Class Activation Mapping): Grad-CAM was used to generate heatmaps that highlighted the regions of the input image that the model focused on for making predictions. This helped in understanding the model's decision-making process and validating its focus on relevant anatomical structures and anomalies.
- Class Activation Mapping (CAM): CAM provided insights into the specific parts of the image that contributed to the classification decision, aiding in the validation of model predictions and identification of potential areas for improvement.
- SmoothGrad: SmoothGrad generated visual explanations by averaging multiple noisy versions of the input image, producing smoother and more interpretable saliency maps.
- Guided Backpropagation: This technique provided high-resolution visualizations of the features that activated the neurons, offering a clear understanding of the model's focus areas and aiding in the validation of its predictions.

3.5.2 Segmentation Models

DeepLabv3+ with ResNet50 Backbone

DeepLabv3+ is a state-of-the-art model for semantic image segmentation, capable of capturing multi-scale contextual information and providing high-resolution segmentation results. The choice of DeepLabv3+ for endoscopic image segmentation was motivated by its effectiveness in handling complex segmentation tasks and its robust performance in medical imaging.

Model Architecture

DeepLabv3+ extends the DeepLabv3 model by incorporating a decoder module to refine the segmentation results, particularly along object boundaries. The key components of DeepLabv3+ include:

1. **Atrous Spatial Pyramid Pooling (ASPP):** ASPP is designed to capture multi-scale information by applying atrous convolution with different rates. This module consists of parallel atrous convolutional layers with different dilation rates, enabling the model to capture features at multiple scales without increasing the number of parameters significantly.

$$\text{ASPP}(x) = [\text{conv}_1(x), \text{conv}_3(x,\text{rate} = 6), \text{conv}_3(x,\text{rate} = 12), \text{conv}_3(x,\text{rate} = 18),\text{image\_pooling}(x)]$$

2. **Decoder Module**: The decoder module improves the segmentation accuracy by refining the coarse segmentation maps produced by the encoder. It upsamples the low-resolution feature maps and combines them with the corresponding high-resolution feature maps from the backbone network.

$$\text{Decoder}(x) = \text{upsample}\big(\text{ASPP}(x)\big) + \text{high\_resolution\_features}$$

3. **ResNet50 Backbone**: The ResNet50 backbone serves as the feature extractor, providing rich feature representations. ResNet50 consists of multiple residual blocks, which help in training deep networks by mitigating the vanishing gradient problem. Each residual block can be represented as:

$$y = F(x, \{W_i\}) + x$$

where x is the input, y is the output, and F(x,{Wi}) is the residual mapping to be learned.

4. **Upsampling Layers**: These layers are used to increase the resolution of the feature maps to match the input image size. The final segmentation map is produced by applying a 1x1 convolution to reduce the number of channels to the number of target classes, followed by upsampling to the original image size.

$$\text{Output} = \text{Conv}_{1\times1}\big(\text{Decoder}(x)\big)$$

Training Process

The training process for DeepLabv3+ involved several critical steps to ensure optimal performance:

a. Data Augmentation: To enhance the model's generalization ability, various data augmentation techniques were applied, including:

- Horizontal Flip
- Random Rotation
- Scaling
- Shearing
- Zooming

b. **Data Normalization**: Each pixel value in the images was normalized to the range [0, 1] by dividing by 255. This ensures consistent input for the model and helps in stabilizing the training process.

c. **Optimizer and Loss Function**: The Adam optimizer was chosen for its ability to handle sparse gradients and its computational efficiency. The binary cross-entropy loss function was used to measure the performance of the model during training. The loss function is defined as:

$$L_{\text{BCE}} = -\frac{1}{N}\sum_{i=1}^{N}(y_i \log(p_i) + (1 - y_i)\log(1 - p_i))$$

where yi is the ground truth label and pi is the predicted probability for pixel i.

d. **Learning Rate Scheduling**: A dynamic learning rate schedule was implemented using a custom callback. The learning rate was adjusted based on the performance of the model on the validation set, ensuring faster convergence and prevention of overfitting.

Transformer-based Segmentation Model

Transformer-based models have recently gained popularity in various computer vision tasks, including image segmentation, due to their ability to capture long-range dependencies and contextual information. A custom transformer-based segmentation model was implemented to explore its potential in endoscopic image segmentation.

Model Architecture

The transformer-based segmentation model consists of the following key components:

1. Positional Encoding: Positional encoding is used to provide the model with information about the spatial positions of the pixels. This is crucial for vision tasks, as transformers do not have inherent knowledge of the input's spatial structure.

$$\mathrm{PE}(\mathrm{pos}, 2i) = \sin\left(\frac{\mathrm{pos}}{10000^{2i/d_{\mathrm{model}}}}\right), \quad \mathrm{PE}(\mathrm{pos}, 2i+1) = \cos\left(\frac{\mathrm{pos}}{10000^{2i/d_{\mathrm{model}}}}\right)$$

2. Transformer Encoder: The encoder consists of multiple layers of self-attention and feed-forward neural networks. The self-attention mechanism allows the model to focus on different parts of the input image, capturing long-range dependencies.

$$\text{Self-Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where Q, K, and V are the query, key, and value matrices, respectively.

3. Transformer Decoder: The decoder also consists of multiple layers of self-attention and feed-forward neural networks, along with cross-attention layers that allow the decoder to focus on relevant parts of the encoder's output.

4. **Convolutional Blocks**: Convolutional layers are used to process the input image before feeding it to the transformer and to process the transformer's output before generating the final segmentation map.

Training Process

The training process for the transformer-based segmentation model involved several critical steps:

1. **Data Augmentation and Normalization**: Similar to the DeepLabv3+ model, data augmentation and normalization techniques were applied to enhance the model's generalization ability and ensure consistent input.

2. **Optimizer and Loss Function**: The Adam optimizer was used to optimize the model parameters. The binary cross-entropy loss function, combined with Dice loss, was used to measure the performance of the model. The Dice loss is defined as:

$$L_{\text{Dice}} = 1 - \frac{2\sum_i p_i y_i}{\sum_i p_i + \sum_i y_i}$$

where pi is the predicted probability and yi is the ground truth label for pixel i.

3. **Learning Rate Scheduling**: A dynamic learning rate schedule was implemented to ensure efficient training and prevent overfitting.

UNet-based Segmentation Models

UNet-based models are widely used in medical image segmentation due to their symmetric encoder-decoder architecture, which allows for precise localization and segmentation. Several UNet-based models were tested for endoscopic image segmentation.

Model Architecture

The UNet architecture consists of an encoder (downsampling path) and a decoder (upsampling path):

1. **Encoder**: The encoder is composed of repeated application of two 3x3 convolutions, each followed by a ReLU activation and a 2x2 max pooling operation. At each downsampling step, the number of feature channels is doubled.

$$\text{ConvBlock}(x) = \text{ReLU}\left(\text{Conv}_{3\times3}\left(\text{ReLU}\left(\text{Conv}_{3\times3}(x)\right)\right)\right)$$

2. **Decoder**: The decoder consists of upsampling the feature map followed by a 2x2 convolution that halves the number of feature channels. This is followed by concatenation with the corresponding feature map from the encoder and two 3x3 convolutions, each followed by a ReLU activation.

$$\text{UpConvBlock}(x)$$
$$= \text{ReLU}\left(\text{Conv}_{3\times3}\left(\text{ReLU}\left(\text{Conv}_{3\times3}\left(\text{Concat}(\text{UpConv}_{2\times2}(x), \text{skip\_connection})\right)\right)\right)\right)$$

3. **Final Layer**: A 1x1 convolution is applied to map the final feature map to the desired number of classes, followed by a softmax activation to obtain the class probabilities.

$$\text{Output} = \text{Softmax}\left(\text{Conv}_{1\times1}\left(\text{UpConvBlock}(x)\right)\right)$$

Training Process

The training process for UNet-based models involved similar steps as other segmentation models:

1. **Data Augmentation and Normalization:** Various data augmentation techniques were applied, and the pixel values were normalized to ensure consistent input.
2. **Optimizer and Loss Function:** The Adam optimizer and a combination of binary cross-entropy loss and Dice loss were used to measure the performance of the model.
3. **Learning Rate Scheduling:** A dynamic learning rate schedule was implemented to ensure efficient training and prevent overfitting.

3.6 MODEL EVALUATION

Model evaluation is a critical phase in the development of AI models, particularly in the field of medical imaging, where accuracy and reliability are paramount. This section outlines the evaluation metrics, benchmarking methods, and comparative analysis employed to assess the performance of the classification and segmentation models developed in this study.

Evaluation Metrics

To comprehensively evaluate the performance of the models, several metrics were utilized. These metrics provide a multi-faceted understanding of the models' strengths and weaknesses, ensuring their robustness and applicability in clinical settings.

Classification Metrics:

1. **Accuracy:**

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$$

This metric indicates the overall correctness of the model's predictions.

2. **Confusion Matrix:** A confusion matrix provides a detailed breakdown of the model's performance by showing the counts of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) predictions.

3. **Precision, Recall, and F1-Score:**

- Precision:

$$\textbf{Precision} = \frac{\textbf{True Positives}}{\textbf{True Positives} + \textbf{False Positives}}$$

Precision indicates the accuracy of the positive predictions.

- Recall:

$$\textbf{Recall} = \frac{\textbf{True Positives}}{\textbf{True Positives} + \textbf{False Negatives}}$$

Recall measures the model's ability to identify all positive instances.

- F1-Score:

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

The F1-Score provides a harmonic mean of precision and recall, offering a single metric that balances both concerns.

4. **Receiver Operating Characteristic (ROC) Curve and Area Under the Curve (AUC):**

The ROC curve visualizes the trade-off between sensitivity (recall) and specificity (1 - false positive rate). The AUC summarizes the overall performance, with a value closer to 1 indicating better performance.

Segmentation Metrics:

- Dice coefficient:

$$\text{Dice Coefficient} = \frac{2|X \cap Y|}{|X| + |Y|}$$

The Dice Coefficient measures the overlap between the predicted segmentation and the ground truth, providing a value between 0 and 1, where 1 indicates perfect overlap.

- Jaccard Index:

$$\text{Jaccard Index} = \frac{|X \cap Y|}{|X \cup Y|}$$

The Jaccard Index, or Intersection over Union (IoU), measures the similarity the predicted segmentation and the ground truth.

- Loss Value

The categorical cross-entropy loss for classification and binary cross-entropy loss for segmentation were used to quantify the difference between the predicted and actual values. These loss functions are minimized during training to improve model performance.

Benchmarking

To ensure a fair and comprehensive comparison, the models were benchmarked against each other using the evaluation metrics mentioned above. The benchmarking process involved the following steps:

Dataset Consistency:

All models were evaluated on the same test set to ensure consistency. The test set was stratified to maintain the distribution of classes, ensuring that the evaluation metrics were comparable across models.

Comparison Criteria:

Classification Models

1. EfficientNetB2

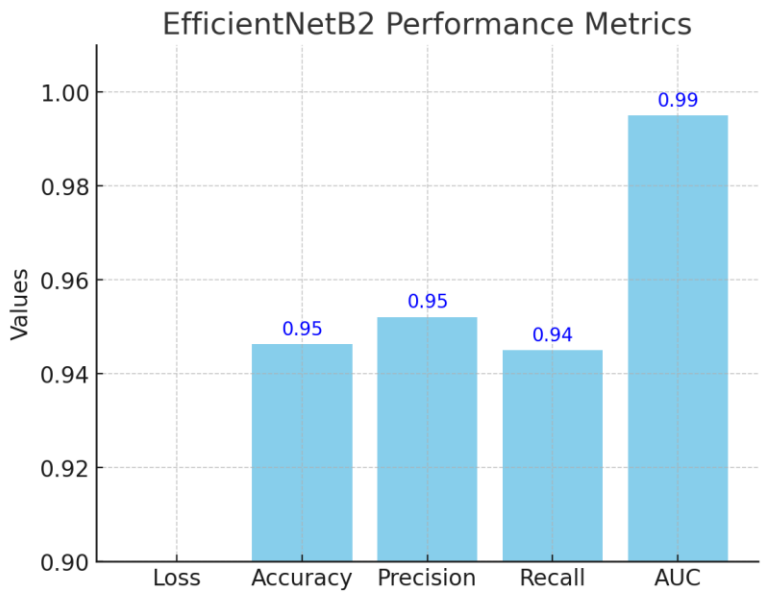

Figure 7 Overiew Global

Description: The bar chart displays the performance metrics of the EfficientNetB2 model, including loss, accuracy, precision, recall, and AUC.

Observation: The EfficientNetB2 model performs well across all metrics, with high AUC indicating excellent classification ability.

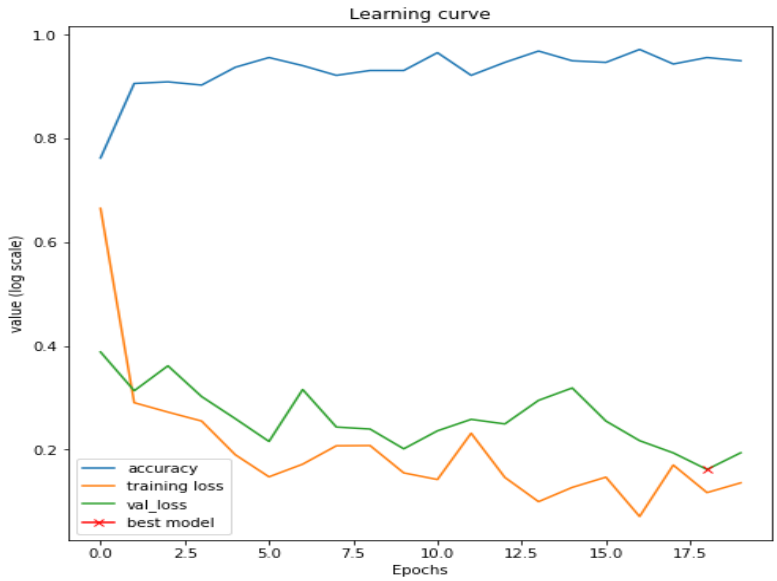

Figure 8 Performance plot

Description: The learning curve graph shows accuracy and validation loss over 20 epochs. It also marks the best model's performance.

Observation: The accuracy improves, and validation loss decreases over time, indicating effective learning and model improvement.

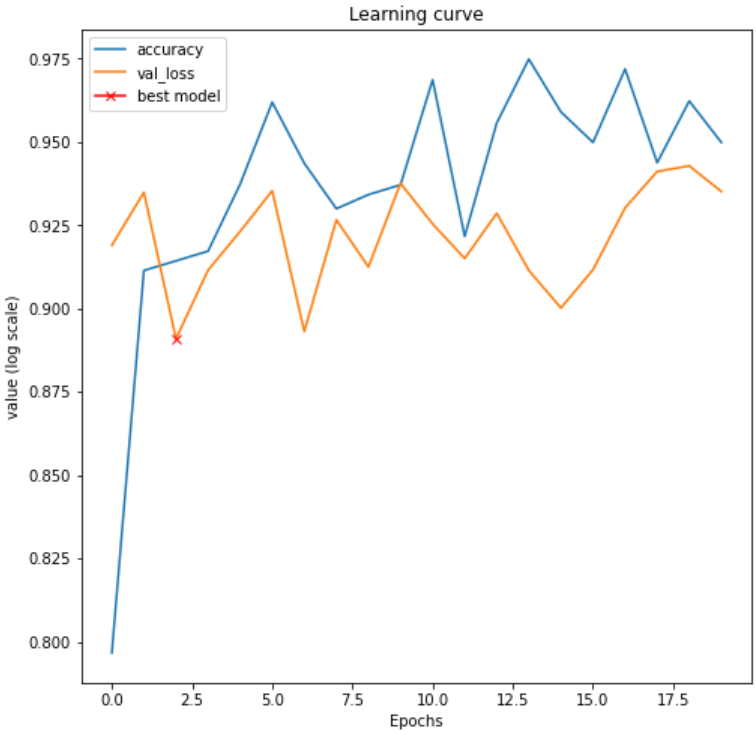

Figure 9 Precision plot

Description: The learning curve graph shows accuracy and validation loss over 20 epochs. It also marks the best model's performance.

Observation: The accuracy improves, and validation loss decreases over time, indicating effective learning and model improvement.

Figure 10 Recall plot

Figure 11 AUC plot

2. EfficientNetB4

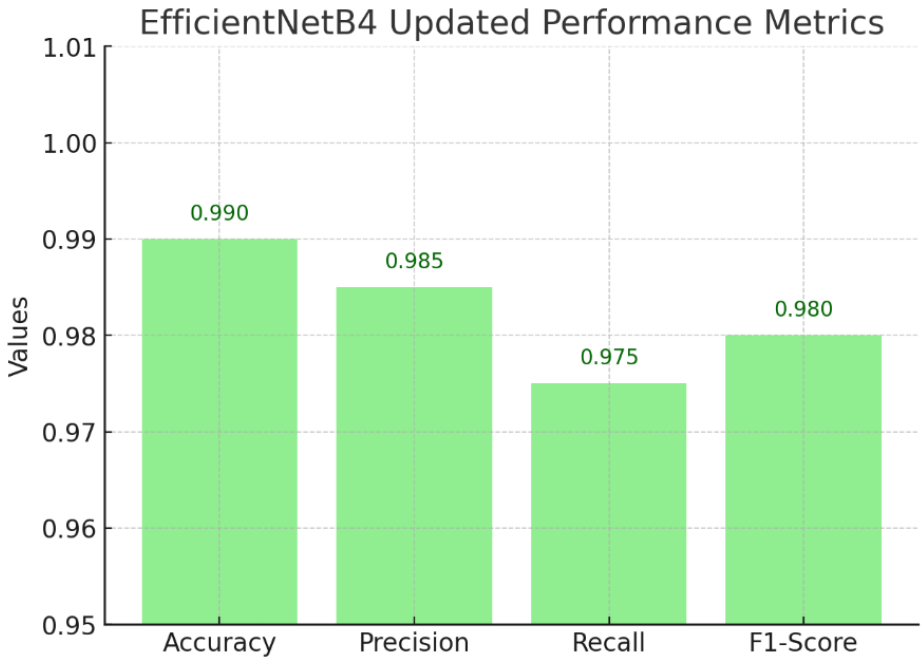

Figure 12 Global Overiew

Description: This bar chart presents the performance metrics of the EfficientNetB4 model, including accuracy, precision, recall, and F1-score.
Observation: The model achieves high performance across all metrics, indicating effective classification capabilities.

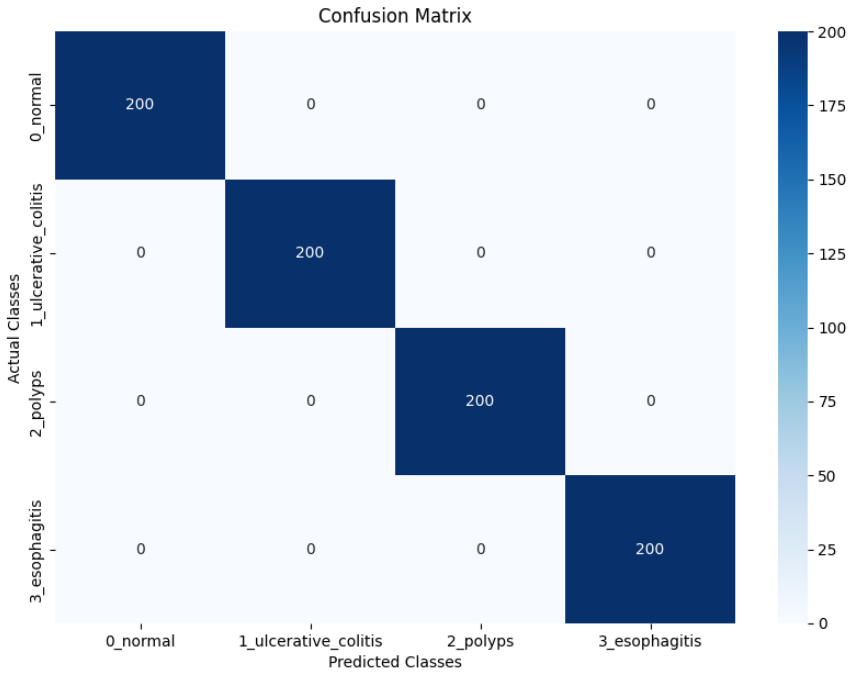

Figure 13 Confusion Matrix

Description: The confusion matrix provides a detailed breakdown of classification performance across four classes (normal, ulcerative colitis, polyps, esophagitis).
Observation: The matrix shows high accuracy, with all predictions correctly falling into their respective classes without any misclassifications.

Segmentation Models

1. Custom DeepLabv3+

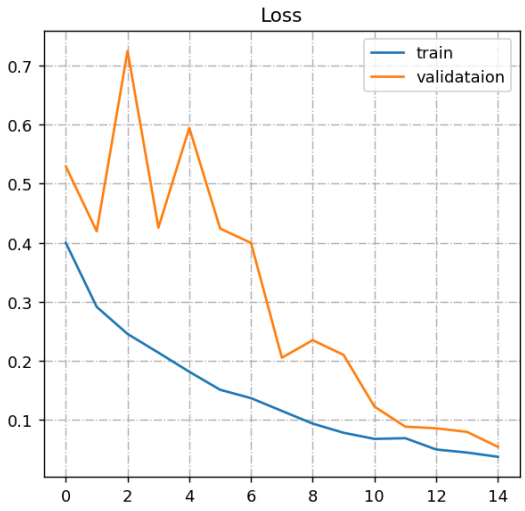

Figure 14 Loss plot

Description: This plot depicts the loss values over 15 epochs for training and validation datasets. Loss measures the difference between predicted and actual values, with lower values indicating better performance.

Observation: The training loss consistently decreases, while the validation loss shows fluctuations but generally trends downward.

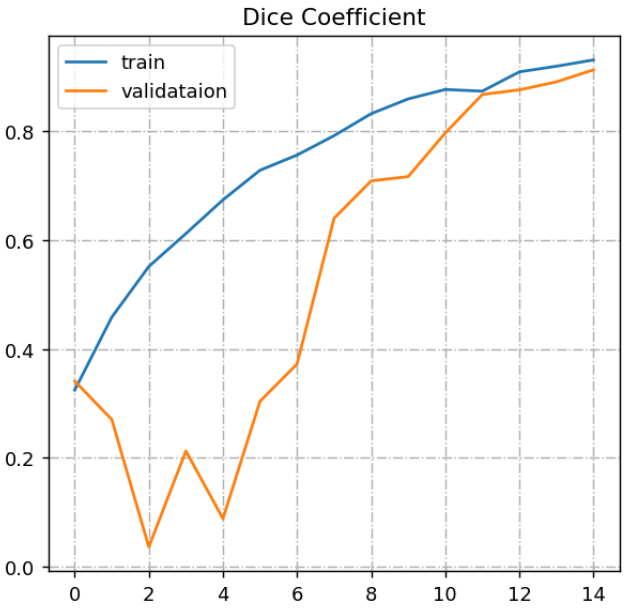

Figure 15 Dice Index Plot

Description: This graph illustrates the Dice Coefficient over 15 epochs for training and validation datasets. The Dice Coefficient evaluates the overlap between predicted and actual segmentations.

Observation: The Dice Coefficient increases for both training and validation datasets, demonstrating improved segmentation accuracy.

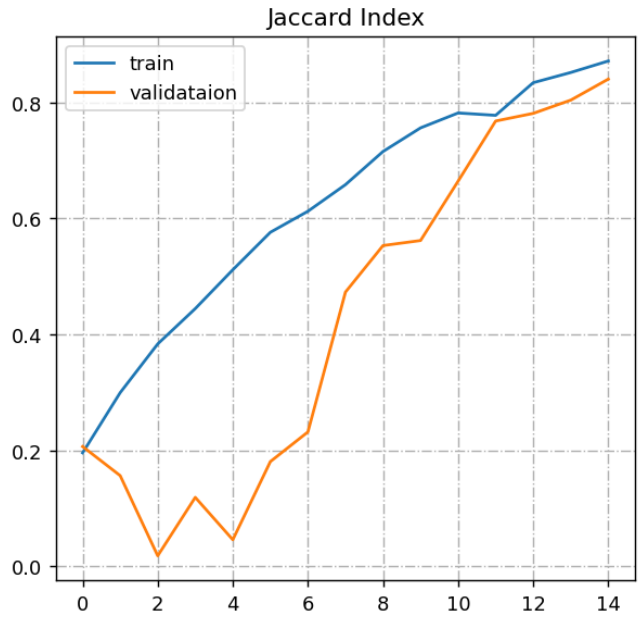

Figure 16 Jaccard Index Plot

Description: This graph shows the Jaccard Index over 15 epochs for both training and validation datasets. The Jaccard Index measures the similarity between predicted and actual segmentations.

Observation: Both the training and validation Jaccard Index values improve over time, indicating better model performance.

Test Prediction

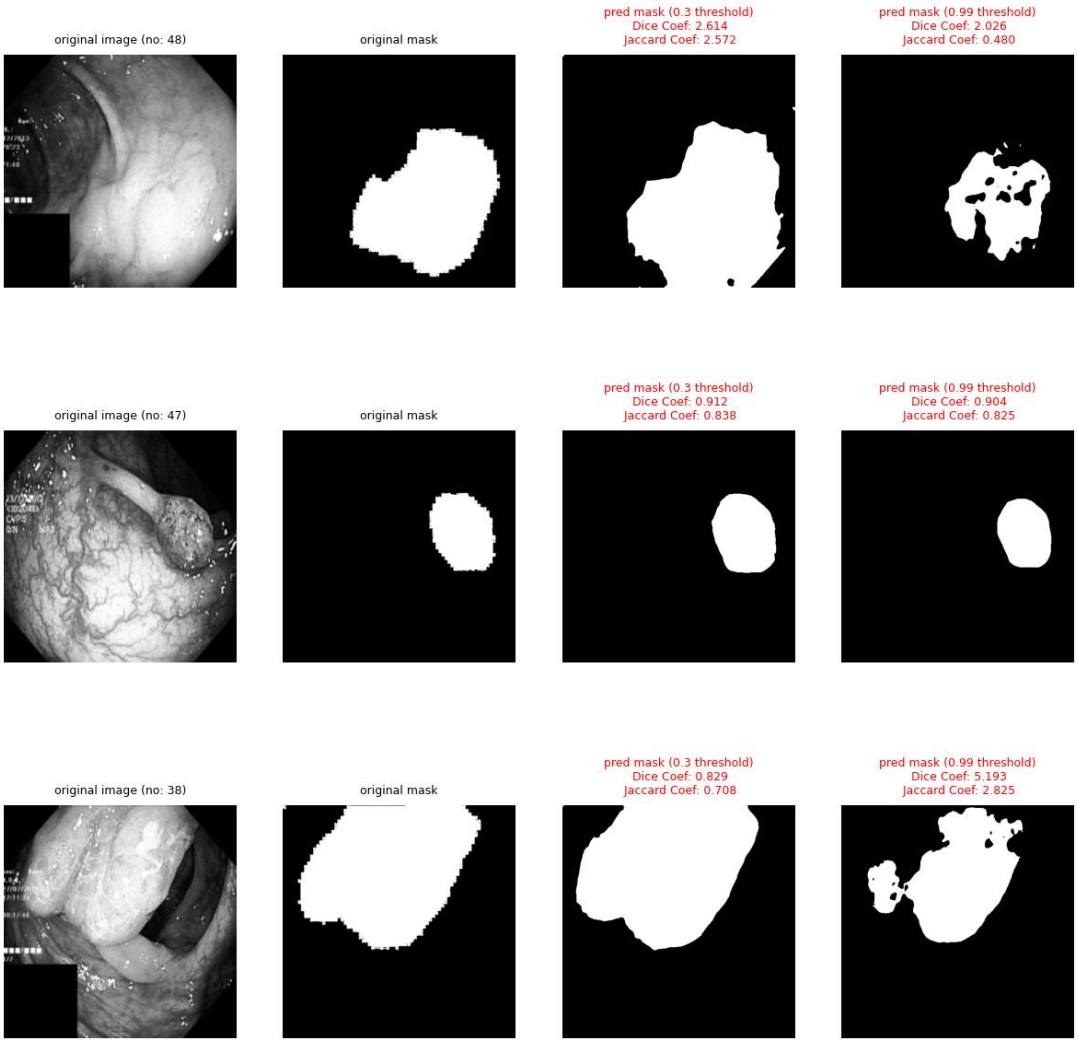

Figure 17 Test Prediction

This image displays a comparison between the original endoscopic images, ground truth masks, and predicted masks for different threshold values (0.3 and 0.99). The Dice Coefficient and Jaccard Index for each prediction are also shown, highlighting the segmentation model's accuracy.

2. SegNet_transformer

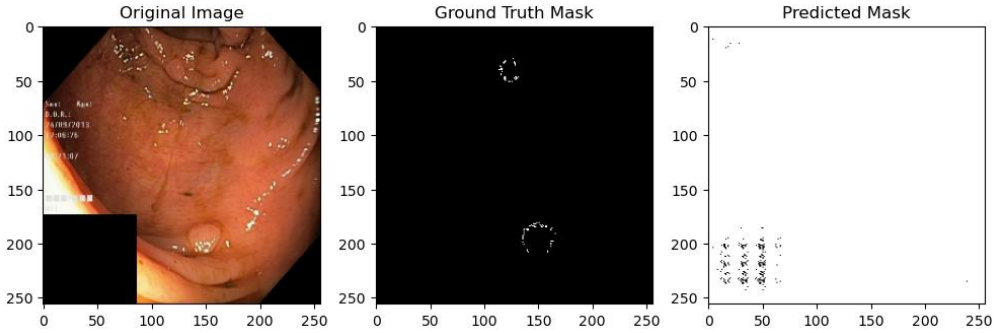

Figure 18 model's segmentation performance 1

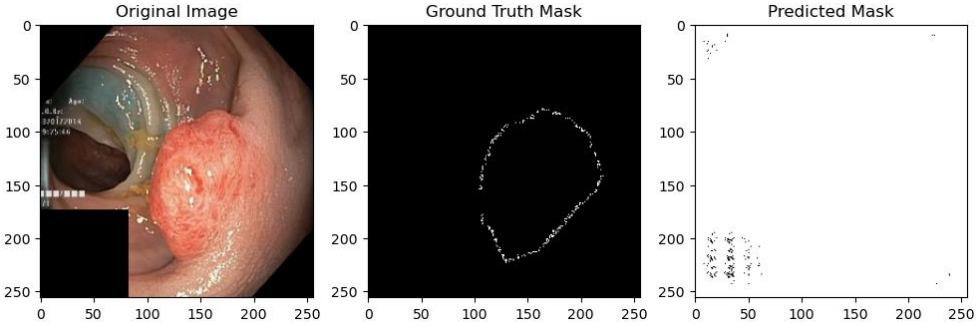

Figure 19 model's segmentation performance 2

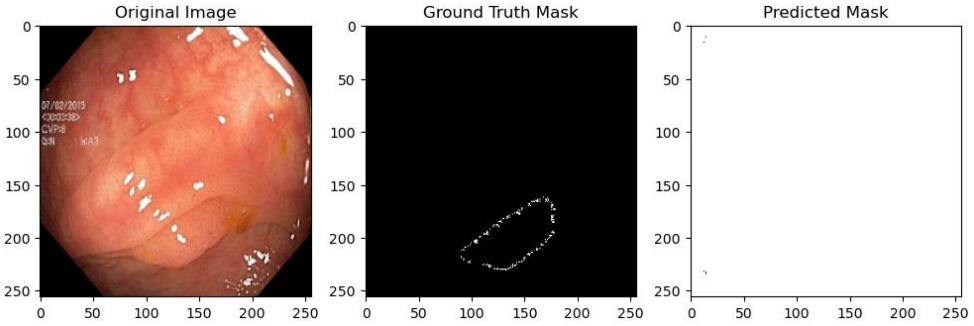

Figure 20 model's segmentation performance 3

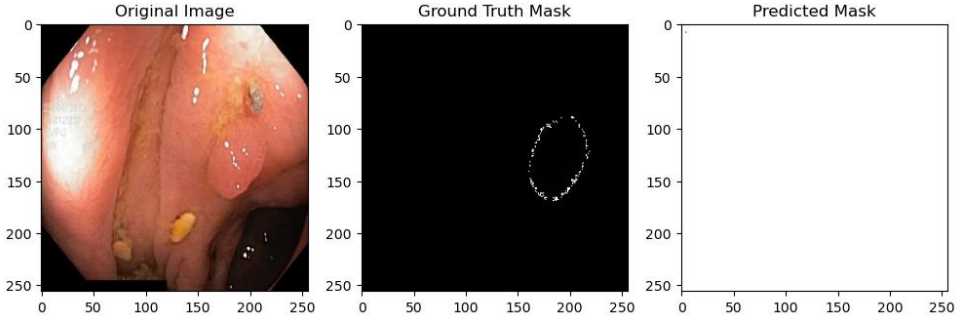

Figure 21 model's segmentation performance 4

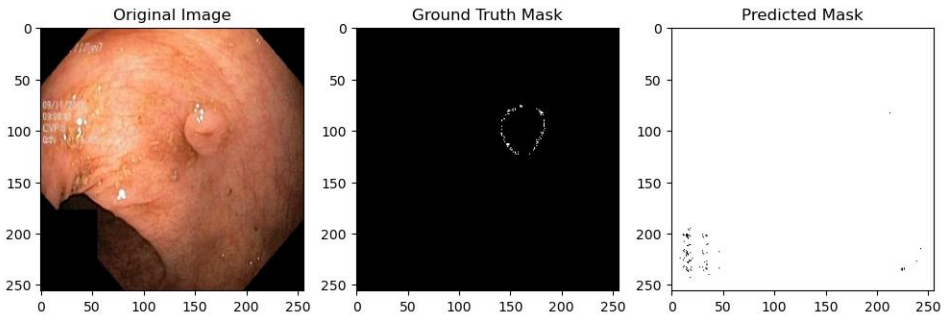

Figure 22 model's segmentation performance 5

These images show the original endoscopic images, ground truth masks, and the corresponding predicted masks for several cases. Each set of images provides a visual representation of the model's segmentation performance, comparing the predicted segmentation to the actual ground truth.

Table 3 Comparison between en EfEfficientNetB2 and EfficientNetB4

Model	Train Accuracy	Validation Accuracy	Test Accuracy	Precision	Recall	AUC	Loss
EfficientNetB2	94.06%	98.06%	98.06%	95.19%	94.06 %	99.50 %	0.14
EfficientNetB4	99.97%	99.80%	99%	-	-	-	-

Table 4 Comparison between DeepLabv3+ and SegNet

Model	Dice Coefficient	Jaccard Index	Loss			
DeepLabv3+	0.91	0.84	0.05			
SegNet_transformer	0.4	0.2	0.7			

3.7 CONCLUSION

In this chapter, we have meticulously outlined the comprehensive methodologies employed to develop, train, and validate advanced AI models for analyzing endoscopic images, focusing specifically on the detection and classification of polyps, esophagitis, and ulcers. This multi-faceted approach ensures a robust framework aimed at achieving high accuracy and practical applicability in clinical settings.

Summary of Methodologies

The methodologies were structured according to the Cross Industry Standard Process for Data Mining (CRISP-DM), which encompasses six key phases:

Business Understanding

This phase identified the project's objectives and requirements from a business perspective. The primary goal was to develop AI models to enhance the accuracy and efficiency of diagnosing medical conditions through endoscopic and radiological imaging.

Data Understanding

This phase involved collecting and familiarizing ourselves with the data, identifying quality issues, and understanding the characteristics and structure of the data. Sources included clinical collaborations with hospitals and online repositories like Kaggle.

Data Preparation

Essential preprocessing techniques were applied, such as histogram equalization, conversion of masks to binary format, and data augmentation. This step ensured that the raw data was transformed into a format suitable for modeling, enhancing quality and consistency.

Modeling

Various algorithms were selected and applied to the prepared data. For classification, EfficientNetB4 was chosen due to its state-of-the-art performance. For segmentation, DeepLabv3+ with a ResNet50 backbone was utilized for its ability to handle high-resolution images and complex tasks.

Evaluation

The performance of the models was rigorously evaluated using metrics such as classification accuracy, confusion matrix, Dice coefficient, Jaccard Index, and loss value. Cross-validation techniques ensured the robustness and generalizability of the models.

Deployment

Although the complete solution has not yet been deployed in a clinical setting, a user-friendly graphical interface was developed using Gradio, HTML, and CSS. This interface was necessary for presenting the project at the National Endoscopy Congress.

In conclusion, the methodologies employed in this study have established a strong foundation for the development and deployment of advanced AI models in endoscopic analysis. The promising results obtained pave the way for future enhancements and real-world applications, ultimately contributing to the advancement of medical imaging and patient care.

Chapter 4: Realization and Perspectives

A. MODEL DEPLOYMENT

In this section, we describe the deployment process of our classification model, which was necessary for presenting our project at the National Endoscopy Congress. Given the importance of providing an easily accessible and understandable interface for clinicians, a user-friendly graphical interface was developed using Gradio combined with HTML and CSS. Although the complete solution has not yet been deployed in a clinical setting, the development of the interface represents a significant step towards practical application.

Model Saving and Preparation

After training the EfficientNetB4 classification model, it was saved for future use. The model was exported using TensorFlow's model saving functionalities to ensure it could be loaded and utilized without the need for retraining. This is a crucial step to make the model portable and easy to integrate into various applications.

User Interface Development

To make the model accessible for clinicians during the congress presentation, a user-friendly interface was developed using Gradio. Gradio allows for the rapid creation of interfaces for machine learning models, providing an interactive web-based interface without requiring extensive web development knowledge.

Gradio Interface

Gradio was used to create an interface that allows users to upload endoscopic images, run the classification model, and view the results. The interface includes a simple drag-and-drop functionality for image uploads, real-time prediction, and the display of confidence scores and class descriptions.

To enhance the visual appeal and user experience, custom HTML and CSS were integrated with the Gradio interface. This customization ensured that the interface was not only functional but also aligned with the aesthetics expected in a professional medical setting.

Figure 23 Welcome Screen Grad web App

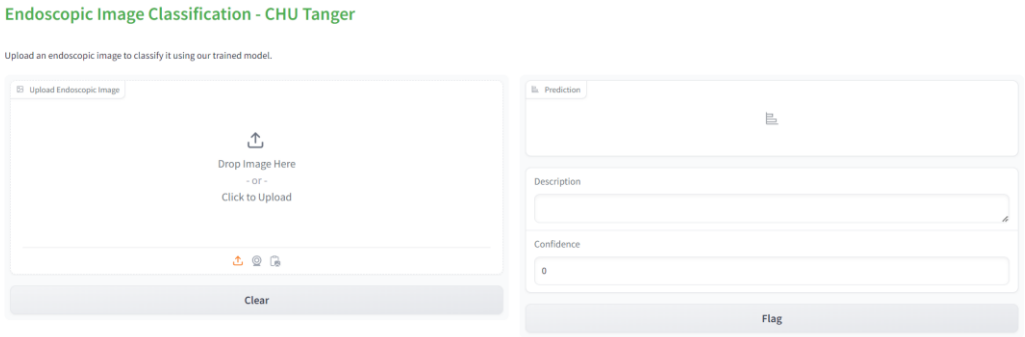

Figure 24 Image Upload Interface

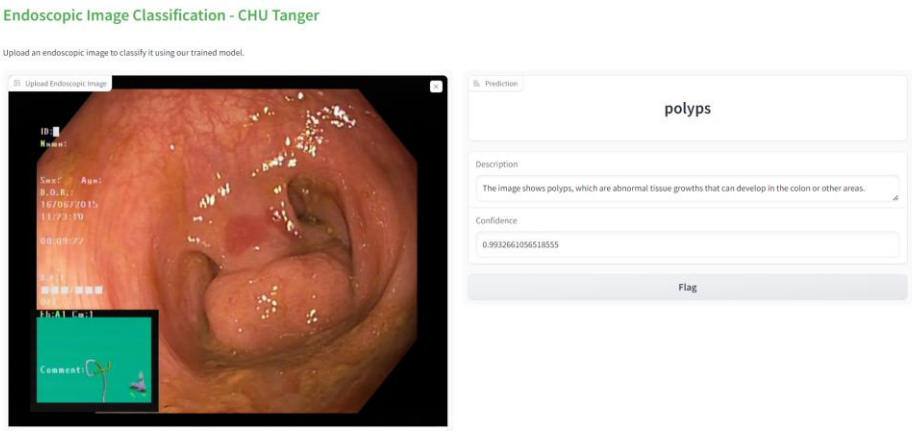

Figure 25 Image Classification Result

B. TECHNOLOGY USED

This section provides an overview of the technologies and tools employed throughout the development and deployment of the AI models for endoscopic image classification and segmentation. The choice of technology is driven by the need for high performance, scalability, and ease of integration into clinical workflows.

1. Development Environment

- Python: The primary programming language used for developing the models. Python's extensive libraries and frameworks for machine learning and image processing make it an ideal choice for this project.

- Jupyter Notebooks: Used for exploratory data analysis, model development, and experimentation. Jupyter provides an interactive environment that facilitates rapid prototyping and iterative development.

- Kaggle Notebooks: Utilized for accessing and processing large datasets, leveraging Kaggle's powerful computational resources.

2. Machine Learning and Deep Learning Frameworks

- TensorFlow: The main deep learning framework used for building and training the models. TensorFlow provides comprehensive tools for model development, training, and deployment.

- Keras: An API within TensorFlow that simplifies the creation of complex neural network architectures. Keras was used for its ease of use and flexibility in designing custom models.

3. Libraries and Tools

- NumPy: Fundamental library for numerical computations, used extensively for data manipulation and processing.

- Pandas: Used for data manipulation and analysis, particularly for handling tabular data and annotations.

- OpenCV: Utilized for image processing tasks such as resizing, augmentation, and pre-processing.

- Matplotlib and Seaborn: Visualization libraries used to plot metrics, loss curves, and other relevant graphs to analyze model performance.

- Scikit-learn: Provides tools for data splitting, model evaluation, and performance metrics calculation. It was also used for generating confusion matrices and classification reports.

4. Data Management

- Kaggle Datasets: Publicly available datasets on Kaggle were utilized for training and validating the models. These datasets include annotated endoscopic images and videos essential for model training.

- Local Storage: Used for storing intermediate data, model checkpoints, and logs during the development process.

5. Model Training and Evaluation

- Google Colab: Leveraged for training models with high computational demands, utilizing Colab's GPU and TPU resources to accelerate the training process.

- Keras Callbacks: Implemented for monitoring model training, including early stopping, learning rate adjustments, and saving the best model weights.

6. User Interface Development

- Gradio: Used to develop an interactive web interface for the classification model. Gradio allows for the creation of user-friendly interfaces for machine learning models with minimal effort.

- HTML and CSS: Employed to customize the Gradio interface, ensuring a professional and intuitive user experience.

7. Deployment

- Flask/django: Considered for future deployment to create a robust backend for serving the models as a web application.

- Docker: Potential tool for containerizing the application, ensuring consistency and ease of deployment across different environments.

- AWS: Cloud platform that can be utilized for deploying the final application, providing scalability and reliability required for real-world clinical use.

By leveraging these technologies, the project ensures a robust, scalable, and user-friendly solution for endoscopic image classification and segmentation. The choice of tools and frameworks facilitates efficient development, training, and deployment of the models, ultimately aiming to enhance diagnostic accuracy and support clinical decision-making.

Chapter 5: Conclusion

5.1 Summary of Research

This thesis explored the development and application of advanced image processing techniques with a primary focus on medical endoscopy. The research aimed to enhance the detection and analysis of anomalies in endoscopic procedures through the integration of AI and machine learning models. Key objectives included improving diagnostic accuracy, providing real-time decision support, and developing user-friendly interfaces for clinical use.

The research comprised several stages:

1. Data Collection and Preparation: Extensive datasets of endoscopic images were gathered from clinical collaborations and online repositories. Data preprocessing steps such as histogram equalization, smoothing, and data augmentation were employed to ensure high-quality input for the models.

2. Model Development: Various models were developed and evaluated. For classification tasks, EfficientNetB2 and EfficientNetB4 models were implemented, with EfficientNetB4 demonstrating superior performance. For segmentation tasks, DeepLabv3+ with a ResNet50 backbone and a custom transformer-based model were developed, with DeepLabv3+ showing the best results.

3. Evaluation and Benchmarking: The models were evaluated using metrics such as accuracy, precision, recall, F1-score, Dice coefficient, and Jaccard index. Benchmarking against baseline models and other advanced techniques ensured a comprehensive assessment of performance.

4. User Interface Development: A user-friendly interface was created using Gradio, combined with HTML and CSS, to facilitate the presentation of the classification models at the National Congress of Endoscopy.

5.2 Conclusions

The research successfully demonstrated the potential of AI and machine learning in enhancing endoscopic procedures. The key findings include:

- Improved Diagnostic Accuracy: The EfficientNetB4 model achieved exceptional accuracy in classifying endoscopic images, significantly improving the detection rates of anomalies such as polyps and esophagitis.

- Effective Segmentation: The DeepLabv3+ model with a ResNet50 backbone provided high-quality segmentation results, essential for precise localization of abnormalities in endoscopic images.

- Real-Time Decision Support: The developed models are capable of providing real-time analysis and feedback, which is crucial for assisting clinicians during endoscopic procedures.

- User-Friendly Interface: The Gradio-based interface ensured that the models are accessible and easy to use, facilitating their adoption in clinical settings.

5.3 Limitations

Despite the promising results, the research has several limitations:

- Dataset Diversity: While extensive, the datasets used may not cover all possible variations in endoscopic images. Future work should aim to include more diverse datasets to improve model generalizability.

- Dataset Quality and Quantity: The clinical dataset provided was poor in both quality and quantity, necessitating the search for supplementary datasets online. Additionally, the anticipated dataset of annotated endoscopic videos has not yet been received, limiting the scope of the project's video analysis capabilities.

- Deployment: The full deployment of the models in clinical settings was not achieved within the scope of this research. Further steps are needed to integrate the models into hospital information systems and ensure their reliability in real-world use.

4.4 Future work

Future work should focus on the following areas:

-Applying Image Models to Videos: Extending the current image-based models to process and analyze endoscopic videos is crucial. This involves adapting the segmentation and classification models to handle continuous frames, enabling real-time anomaly detection and tracking during endoscopic procedures.

- Full Clinical Deployment: Collaborating with hospitals to deploy the models in real-world settings, ensuring they are integrated into existing workflows and tested for reliability and efficiency.

- Exploring Other Medical Imaging Modalities: Extending the techniques developed in this research to other medical imaging modalities such as CT scans and MRI could further enhance diagnostic capabilities across different medical fields.

4.5 Practical Implications

The findings of this research have significant practical implications:

- Clinical Adoption: The developed models can be integrated into clinical workflows to assist in real-time decision-making, potentially reducing the rate of missed diagnoses and improving patient outcomes.

- Training and Education: The user-friendly interface can be used as a training tool for medical professionals, helping them understand the capabilities and limitations of AI in endoscopy.

- Standardization of Care: The use of AI models can help standardize the quality of endoscopic examinations, ensuring consistent and high-quality care across different healthcare providers.

4.6 Recommendations

Based on the research findings, the following recommendations are proposed:

- Invest in AI Research: Healthcare institutions should invest in AI research and development to harness the potential benefits of advanced image processing and machine learning techniques.

- Collaborate with Clinicians: Ongoing collaboration between data scientists and medical professionals is essential to ensure that AI models are relevant, accurate, and practically applicable in clinical settings.

- Focus on User Experience: Developing user-friendly interfaces and ensuring seamless integration into existing workflows will be key to the successful adoption of AI tools in healthcare.

In conclusion, this research has demonstrated the significant potential of AI and machine learning in enhancing endoscopic procedures. The developed models and techniques provide a foundation for future advancements in medical imaging, with the ultimate goal of improving diagnostic accuracy and patient care.

Bibliography

1. He, Z., Zhang, K., Zhao, N., et al. (2023). **Deep learning for real-time detection of nasopharyngeal carcinoma during nasopharyngeal endoscopy**. iScience, 26, 107463. Available at: https://doi.org/10.1016/j.isci.2023.107463

2. Sujal. (2020). **Endoscopy Image Processing & Classification** - Final Report. AI20BTECH11020. Indian Institute of Technology, Hyderabad.

3. El-Sayed, A., Salman, S., Alrubaiy, L. (2023). **The adoption of artificial intelligence-assisted endoscopy in the Middle East: challenges and future potential**. J. Gastroenterology.

4. Owais, M., Arsalan, M., Choi, J., Mahmood, T., Park, K. R. (2019). **Artificial Intelligence-Based Classification of Multiple Gastrointestinal Diseases Using Endoscopy Videos for Clinical Diagnosis**. J. Clin. Med., 8(7), 986. Available at: https://doi.org/10.3390/jcm8070986

5. Pannala, R., Krishnan, K., Melson, J., et al. (2019). **Artificial intelligence in gastrointestinal endoscopy**. ASGE Society Document.

6. Chung, J., Oh, D. J., Park, J., Kim, S. H., Lim, Y. J. (2021). **Automatic Classification of GI Organs in Wireless Capsule Endoscopy Using a No-Code Platform-Based Deep Learning Model**. Medical Image Analysis, 68, 101885. Available at: https://doi.org/10.1016/j.media.2021.101885

7. Guo, F., Meng, H. (2023). **Application of artificial intelligence in gastrointestinal endoscopy.** American Journal of Gastroenterology, 118(1), 32-43. Available at: https://doi.org/10.1016/j.ajg.2023.12.010

8. Wang, K. W., Dong, M. (2020**). Potential applications of artificial intelligence in colorectal polyps and cancer: Recent advances and prospects**. World Journal of Gastroenterology, 26(34), 5090-5100. Available at: https://doi.org/10.3748/wjg.v26.i34.5090

9. Ali, S., et al. (2022**). Where do we stand in AI for endoscopic image analysis**? Deciphering gaps and future directions. npj Digital Medicine, 5, 184. Available at: https://doi.org/10.1038/s41746-022-00611-4

10. Prasath, V. B. S. (2017). **Polyp Detection and Segmentation from Video Capsule Endoscopy:** A Review. J. Imaging, 3(1), 1. Available at: https://doi.org/10.3390/jimaging3010001

11. Mohanty, S. N., Nalinipriya, G., Jena, O. P., & Sarkar, A. (Eds.). (2021**). Machine Learning for Healthcare Applications**. Springer.

12. Goodfellow, I., Bengio, Y., & Courville, A. (2016). **Deep Learning. MIT Press**.

13. Lu, L., Wang, X., Carneiro, G., & Yang, L. (Eds.). (2019**). Deep Learning and Convolutional Neural Networks for Medical Imaging and Clinical Informatics**. Springer.

14. Chen, Y. W., & Jain, L. C. (Eds.). (2020). **Deep Learning in Healthcare: Paradigms and Applications. Springer.**

15. Tavares, J. M. R. S., & Natal Jorge, R. M. (Eds.). (2019). **Advances in Computational Vision and Medical Image Processing**. Springer.

16. Suresh, A., Vimal, S., Robinson, Y. H., & Ramaswami, D. K. (Eds.). (2020). **Bioinformatics and Medical Applications: Big Data Using Deep Learning Algorithms**. Springer.

17. Park, J., Hwang, Y., Yoon, J. H., et al. (2019). **Recent Development of Computer Vision Technology to Improve Capsule Endoscopy**. Clin Endosc, 52(4), 328-333. Available at: https://doi.org/10.5946/ce.2018.172

18. Tyagi, A. K., & Tariq, R. (Eds.). (2019). **Deep Learning for Medical Care: Principles, Methods, and Applications**. Springer.

19. Rashid, T. (2016). **Make Your Own Neural Network**. CreateSpace Independent Publishing Platform.

20. Prince, S. J. D. (2024**). Understanding Deep Learning**. MIT Press. Available at: https://libgen.li

21. NVIDIA Developer Blog. (2023). AI in Endoscopy: Improving Detection Rates and Visibility with Real-Time Sensing. Available at: https://developer.nvidia.com/blog/ai-in-endoscopy-improving-detection-rates-and-visibility-with-real-time-sensing/