**AUTOMATIC DETECTION AND CLASSIFICATION OF GASTROINTESTINAL PATHOLOGICAL FINDINGS FROM ENDOSCOPIC IMAGES USING HYBRID RESNET50-CNN**

**SAMSON KINYANJUI KATHUGUNA**

**A Research Proposal Submitted in Partial Fulfillment for the Award of** the Degree of Masters of Science in Computer Science, in the School of Computer Science and Information Technology, Dedan Kimathi University of Technology

**NOVEMBER, 2024**

# DECLARATION

*This research proposal is my original work and has not been presented in any university/institution for a degree or for consideration of any certification*

**Signature………………… Date January 12, 2025**

**Student’s Name……………………………………………………………………**

**Student’s Registration No:C241-01-2311/2023**

**Supervisors’ declaration caption:**

*This research proposal has been submitted for examination with our/my approval as University Supervisor(s).*

**Signature……………………………………….Date.…………………………….**

**Prof/Dr……………………Juliet Moso………………………………...................**

Department………………………………………………………………………….

Institution……………………………………………………………………………

**Signature……………………………………….Date.……………………………**

**Prof/Dr…………Patrick Gikunda……………………………………………….**

Department……………………………………………………………………. Institution: ………………………………………………………………………..

# DEDICATION

I dedicate this work to Almighty God, whose grace, wisdom, and guidance have been my strength throughout this journey. Without His blessings, this accomplishment would not have been possible.

To my beloved family, thank you for your unwavering support, patience, and encouragement.

To my supervisors, whose mentorship and insightful guidance have enriched this project, I am profoundly grateful.

### ACKNOWLEDGEMENT

I am deeply grateful to all those who have supported and encouraged me throughout the course of this project.

First, I would like to express my heartfelt gratitude to my supervisors, whose expertise, patience, and valuable feedback have been instrumental in shaping this research. Your guidance and dedication to my academic and professional growth have been invaluable.

I would also like to thank my family for their endless love, encouragement, and unwavering support. Your faith in me has given me the strength to persevere through every challenge, and I am profoundly thankful for all you have done to make this journey possible.

A special thanks to the institutions and organizations that provided the resources necessary for this research. Your support has been vital to the successful completion of this work.

Lastly, I extend my sincere appreciation to my friends and colleagues, who have been sources of motivation and camaraderie. Thank you for your encouragement and assistance, which have greatly contributed to the completion of this project.

Contents

[DECLARATION 2](#_Toc182393586)

[DEDICATION 3](#_Toc182393587)

[ACKNOWLEDGEMENT 4](#_Toc182393588)

[LIST OF TABLES 6](#_Toc182393589)

[LIST OF FIGURES 7](#_Toc182393590)

[Abstract 8](#_Toc182393591)

[CHAPTER ONE INTRODUCTION 9](#_Toc182393592)

[1.1 Background of the study 9](#_Toc182393593)

[1.2 Problem Statement 10](#_Toc182393594)

[1.3 Purpose/Aims/General Objective 10](#_Toc182393595)

[1.3.1 Specific Objectives 10](#_Toc182393596)

[Model development 11](#_Toc182393597)

[1.4 Significance of the Study/Rationale 11](#_Toc182393598)

[1.5 Scope/Delimitation 11](#_Toc182393599)

[1.6 Research Questions 11](#_Toc182393600)

[CHAPTER TWO LITERATURE REVIEW 12](#_Toc182393601)

[CHAPTER THREE METHODOLOGY 16](#_Toc182393602)

[3.1 Research Design 16](#_Toc182393603)

[3.2 Location of the Study 16](#_Toc182393604)

[3.3 Target Population/Materials 16](#_Toc182393605)

[3.3.1 Kvasir Dataset 16](#_Toc182393606)

[3.3.2 Gastrovision Dataset 16](#_Toc182393607)

[3.4 Sample Size and Sampling Techniques 17](#_Toc182393608)

[3.5 Data Collection Techniques 17](#_Toc182393609)

[3.5.1 Data Augmentation 17](#_Toc182393610)

[3.6 Model Development 17](#_Toc182393611)

[3.6.1 Feature **Extraction** 18](#_Toc182393612)

[3.6.2 Classification Network 18](#_Toc182393613)

[3.6.3 Grad-CAM Integration 18](#_Toc182393614)

[3.7 Model Evaluation Metrics 18](#_Toc182393615)

[3.8 Data Analysis 19](#_Toc182393616)

[3.9 Logistical and Ethical Considerations 19](#_Toc182393617)

[3.9.1 Logistical Considerations 19](#_Toc182393618)

[3.9.2 Ethical Considerations 19](#_Toc182393619)

[References 20](#_Toc182393620)

[Appendices 25](#_Toc182393621)

[Appendix 1: Work Plan 25](#_Toc182393622)

# LIST OF TABLES

[Table 1 State of the art Algorithms used in same task 14](#_Toc182381331)

# LIST OF FIGURES

[Figure 1: Sample images from the dataset 15](#_Toc182207020)

[Figure 2:The proposed methodology 16](#_Toc182207021)

# Abstract

The gastrointestinal (GI) tract is highly susceptible to various diseases, including severe conditions such as colorectal and stomach cancers, which contribute significantly to global mortality rates. Traditional diagnostic methods, particularly manual analysis of endoscopic images by specialists, face inherent challenges due to variations in image quality, inconsistent lighting, and high operator dependency. These limitations increase the risk of oversight and diagnostic errors, emphasizing the need for automated solutions that can enhance accuracy and efficiency. This Project proposal aims to develop and evaluate a hybrid deep learning model leveraging ResNet50 for feature extraction combined with a custom convolutional neural network (CNN) for the automatic detection and classification of GI pathological findings. The study utilizes endoscopic images from the Kvasir and Gastrovision datasets, which include various classes such as Esophagitis, Ulcerative Colitis, Polyps, and Normal findings. The integration of data augmentation techniques addresses class imbalances within the datasets, ensuring robust model training and enhanced generalization. Among the main objectives of this research is to improve model interpretability by incorporating Explainable AI (XAI) techniques such as Gradient-weighted Class Activation Mapping (Grad-CAM). Grad-CAM enables visualization of the regions within an image that significantly influence the model’s predictions, which is particularly crucial in medical settings where trust in AI outputs is paramount. The use of Grad-CAM will support endoscopists in understanding the rationale behind model decisions, fostering better clinical integration. The hybrid model architecture comprises the ResNet50 backbone for high-level feature extraction, followed by a custom CNN with convolutional, pooling, and fully connected layers designed to optimize classification accuracy. Techniques such as Batch Normalization and Global Average Pooling (GAP) are employed to stabilize training and mitigate overfitting. Expected results from this research include a robust, high-accuracy classification model capable of distinguishing between different GI tract conditions with a performance superior to traditional and some state-of-the-art methods. The incorporation of Grad-CAM is anticipated to offer clinicians visual insights that align with their domain knowledge, bridging the gap between Machine learning models predictions and human interpretability. The findings of this study are expected to contribute significantly to the field by reducing the workload of endoscopists, improving diagnostic consistency, and enabling earlier intervention for patients with GI diseases.

# CHAPTER ONE INTRODUCTION

## 1.1 Background of the study

The gastrointestinal (GI) tract, comprising the mouth, throat (pharynx), esophagus, stomach, small intestine, colon, and rectum, is prone to various mucosal infections. These range from minor, easily treatable infections to life-threatening diseases such as cancers (stomach, gastrointestinal cancer, and colorectal cancer). Colorectal cancer (CRC) is a severe condition affecting the colon and rectum, typically presenting as abnormal mucosal lining (Saad et al., 2024). According to the International Agency for Research on Cancer, colorectal and stomach cancers are the 3rd and 5th most common cancers globally, ranking 2nd and 5th in mortality rates, respectively (Ferlay et al., 2024). Additionally, other digestive diseases like peptic ulcer, appendicitis, inflammatory bowel disease (a term encompassing Crohn's disease and ulcerative colitis), and duodenitis contribute significantly to global mortality (PAHO, 2021).

Endoscopy remains the primary method for diagnosing GI tract infections. It is a minimally invasive procedure that uses a flexible, thin, and elongated tube called an endoscope to visualize the internal organs of the patient. The endoscope is equipped with a camera and a light source to transmit images of the organs to a monitor, allowing for accurate diagnosis and treatment planning [4][5]. Depending on the specific aim of the operation, the equipment used, and the internal structures being examined, there are diverse types of endoscopy. However, despite its wide usage, the effectiveness of endoscopy varies significantly with operator performance, leading to a 20% polyp miss-rate in the colon. Analyzing endoscopic images is time-intensive and demands high focus, leading to potential diagnostic errors. Therefore, enhancing the accuracy of detecting and classifying GI tract pathological findings is critical for early diagnosis and prevention of GI diseases [6].

Recent advancements in machine learning, particularly deep learning, have shown promise in improving the diagnosis of GI tract diseases. Deep learning, a subset of machine learning, mimics the functions of the human brain. Sharma et al. (2023) developed the U-Net model to assist radio-oncologists in treating cancer more efficiently. This model used six different architectures (Inception V3, ResNet50, VGG19, DenseNet121, InceptionResNetV2, and EfficientNetB) as backbones for feature extraction used in segmentation.

Reed et al. (2022) investigated the use of convolutional neural networks (CNNs) for diagnosing and grading ulcerative colitis (UC). They utilized pre-trained CNN models (ResNet50, VGG19, and DenseNet121) initialized with ImageNet weights, and conducted grid search to select the best hyperparameters using 5-fold cross-validation. Their study formulated the problem as binary classification and further graded the images identified as UC using Mayo grades. The DenseNet121 architecture achieved the highest accuracy of 87.50% and an AUC of 0.90, outperforming the "no skill" model's 72.02% accuracy and 0.50 AUC.

Similarly, Rudreshi et al. (2023) explored transfer learning using the MobileNetV3 model, fine-tuning it with 3000 images from the Kvasir dataset to classify eight categories: anatomical landmarks, pathological findings, and medical procedures. The model achieved an 88% accuracy on the test set, demonstrating the potential of transfer learning in GI tract diagnosis.

Despite the advancements in detecting and classifying GI tract findings using deep learning methods, most of the current approaches have conducted general classification of anatomical landmarks, pathological findings (abnormalities in GI tract), and medical procedure images altogether as GI diseases, this may lead to diagnostic inaccuracies by overlooking subtle but clinically significant findings. While those approaches that work on GI diseases have limited diseases. Therefore, there is a need for cost-effective efficient and quicker model that can detect and classify a wide range of pathological findings effectively in the GI tract. Additionally, there is need for greater explainability in these models, as the lack of transparency raises concerns regarding the trustworthiness of their black-box decision-making processes. Lack of transparency can hinder endoscopists' ability to interpret the models' predictions and integrate them into their diagnostic workflows, ultimately affecting the acceptance and utility of these models in clinical practice.

## 1.2 Problem Statement

Current approaches to GI tract diagnosis using deep learning typically suffer from limited scope, as they focus on general classifications or a restricted set of diseases. This leads to reduced generalizability and potential diagnostic inaccuracies. Additionally, many existing models lack transparency, hindering their integration into clinical practice due to concerns over their “black-box” nature. Therefore, there is a critical need for an efficient, cost-effective model capable of detecting and classifying a wider range of GI pathological findings while providing clear interpretability. This research aims to bridge this gap with a custom CNN architecture using ResNet50 features and Grad-CAM visualization to enhance diagnostic interpretability and trust on deep learning models predictions.

## 1.3 Purpose/Aims/General Objective

The main aim of this project is to develop a machine learning model for detection and classification of pathologic findings in the GI tract using endoscopic images.

### 1.3.1 Specific Objectives

1. Carry out a comprehensive literature review on machine learning techniques used in the detection and classification of Gastrointestinal tract pathological findings from endoscopic images.
2. Data Preprocessing (Resizing and Normalization of Images).
3. Model performance evaluation using the following metrics: Recall, Precision, Specificity, Accuracy, and F1 score.
4. Investigate the interpretability of the machine learning model by employing techniques such as feature importance analysis and visualization of decision boundaries.

## 1.4 Significance of the Study/Rationale

The proposed research is significant as it addresses the limitations of current deep learning models used for detecting and classifying gastrointestinal (GI) pathological findings. Unlike existing models that often focus on only one or two specific diseases, this study’s hybrid ResNet50-CNN model can generalize across a range of GI conditions from the initial experiments the proposed technique shows an accuracy of (95.96%). The integration of Grad-CAM interpretability also provides clear insights into the decision-making process, fostering trust among clinicians and supporting diagnostic consistency. This work has the potential to improve diagnostic outcomes, reduce operator dependency, and enhance the efficiency of clinical workflows, leading to better patient care.

## 1.5 Scope/Delimitation

This study will focus on the detection and classification of pathological findings in GI tract using endoscopic images from the Kvasir dataset and Gastrovision and potentially other datasets that might be related. The project will employ a hybrid model combining deep learning models. The study is not limited to a specific geographical location, but will be based on datasets that may originate from various sources.

## 1.6 Research Questions

1. What machine learning techniques are commonly used for the detection and classification of gastrointestinal (GI) tract pathological findings from endoscopic images?
2. What are the best practices for data preprocessing in medical image analysis?
3. Which machine learning algorithms are most effective for the detection and classification of GI tract pathological findings?
4. What evaluation metrics are commonly used to assess the performance of machine learning models in medical imaging, particularly for GI tract diagnosis?
5. What explainable AI (XAI) techniques can be applied to improve the interpretability of machine learning models in medical imaging?

# CHAPTER TWO LITERATURE REVIEW

Gastrointestinal (GI) disease is considered one of the supreme common diseases that usually infect people, causing complicated health conditions (Du et al., 2019). Based on the degree of injury, GI can approximately split into the precancerous lesion, primary GI cancer and progressive GI cancer, and benign GI diseases (Sharif et al., 2019). Among benign GI diseases, gastritis, and bleedings which will not depreciate into cancers in short term. In contrast, precancerous GI injury could depreciate into primary GI cancer.

Annually almost 0.7 million patients are diagnosed with gastric cancer.Since 2017, 135,430 new GI diseases arose in America. A global survey indicated that since 2017, 765,000 deaths occurred due to stomach cancer, 525,000 deaths are due to colon cancer. The poorest situations can be detected in the developing countries (e.g., the Asian countries and the Middle East) (Ali et al., 2019; Khan et al., 2020a). Moreover, among people diseased with GI diseases, 20% of them are from China, 18% from Brazil, 12% from Russia, 20% of EU, and 21% of the US (Sharif et al., 2019). The early diagnosis of GI tract Abnormalities is essential to reduce medical complications, cost of treatment, and lower death rates.

The traditional clinical method for diagnosing gastrointestinal (GI) tract conditions is through the intestinal biopsy, where samples are analyzed by medical experts using microscopes to detect the presence of cancerous or abnormal cells (Ali et al., 2019). This method, while effective, is invasive and requires a high degree of proficiency.In contrast, endoscopic imaging offers a less invasive alternative for visualizing the GI tract (Kainuma et al., 2015). During an endoscopic procedure, doctors can recognize and diagnose gastric anomalies in their early stages, which can lead to timely treatment and significantly reduce medical complications, treatment costs, and mortality rates associated with GI cancers (Hamashima et al., 2015). However, the manual examination of endoscopic images is challenging. The procedure can generate up to 60,000 frames in a single session, most of which are redundant (Khan et al., 2020b). Only a few frames may contain abnormal lesions, making the manual inspection process time-consuming and prone to human error. Anomalous frames can easily be overlooked, leading to potential misdiagnosis (Aoki et al., 2019).

Despite the advantages of endoscopy, it presents several challenges Presence of Multi-Class Artefacts (visual anomalies or distortions that are not part of the actual anatomical structures being examined) which can hinder visual interpretation by endoscopists and affect the robustness of deep learning methods applied to GI organs, as they can be confused with tissue of interest. There is also difficulty in Identifying Subtle Abnormalities in the GI tract by not being able to clearly distinguish between normal and abnormal tissues, especially subtle precancerous changes, makes manual examination highly demanding and increases the risk of misdiagnosis. The need for automated schemes to assist in the analysis of endoscopic images is evident. Such systems can help detect possible malignancies by analyzing the entire set of frames produced during the endoscopic process, thereby reducing the burden on medical experts and improving diagnostic accuracy (Aoki et al., 2019).

Machine learning (ML) has become a pivotal tool in the field of medical imaging, particularly through the development of computer-aided diagnosis (CADx) systems. These systems utilize ML algorithms to automatically diagnose various diseases across different parts of the human body, such as the brain, breast, and lungs (Attallah, Sharkas & Gadelkarim, 2019, 2020; Ragab, Sharkas & Attallah, 2019; Attallah, Ragab & Sharkas, 2020). In the context of gastrointestinal (GI) diagnostics, CADx systems analyze endoscopic images to identify and diagnose GI diseases (Khan et al., 2020b). The integration of ML into medical imaging offers numerous advantages, including reduced examination times, early detection of lesions, decreased treatment costs, and improved diagnostic accuracy compared to manual examination.

Deep learning, a subset of machine learning, employs neural networks with multiple layers to analyze complex data patterns. In medical imaging, convolutional neural networks (CNNs) have been particularly effective. For instance (Tsuyoshi et al., 2018) developed a CNN-based CAD system using GoogLeNet architecture then they trained it with over 26,000 colonoscopy images from patients with ulcerative colitis (UC). This system was able to demonstrate high performance in identifying normal mucosa and mucosal healing states, with areas under the receiver operating characteristic curves (AUROC of 0.86 and 0.98), respectively. These high levels of accuracy underscore the potential of deep learning to support less experienced endoscopists and reduce interobserver variability in diagnosing inflammation severity in UC patients.

Transfer learning is another powerful technique in machine learning, where a model pre-trained on one task is but can be fine-tuned to work on other tasks, mainly related tasks. This approach is beneficial in medical imaging due to the limited availability of labeled data. Pre-trained models, such as those trained on large datasets like ImageNet, can be fine-tuned on smaller, specific medical datasets. For instance, a CAD system designed to predict persistent histologic inflammation in UC patients utilized a convolutional neural network trained on a substantial dataset of endoscopic images. The system achieved diagnostic sensitivity, specificity, and accuracy of 74%, 97%, and 91%, respectively, demonstrating its effectiveness in identifying persistent inflammation (Yasuharu et al., 2019). By leveraging pre-trained models and transfer learning, medical imaging systems can achieve robust performance even with constrained data resources, facilitating earlier and more accurate disease diagnosis.

Table 1 State of the art Algorithms used in same task

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Authors** | **Purpose** | **Classes** | **Methodology** | **Accuracy (%)** | **Limitations** |
| Pei et al (2017) | Bowel detection and assessment | 2 | LSTM and PCA | 88.8 | Limited classes.  Used only temporal features.  Low accuracy |
| Nguyen et al (2020) | Classifying images to normal and abnormal | 2 | DenseNet,VGG-16,and Inception | 70.7 | Classify images to either Normal or abnormal.  Low classification accuracy |
| Reed et al(2022) | diagnosing and grading  ulcerative colitis (UC) | 2 | pre-trained CNN  models (ResNet50,  VGG19, and  DenseNet121) | highest  accuracy of  87.50% | Limited number of classes |
| Rudreshi et al (2023) | Classification of  Endoscopic images | 3 | MobileNetV3 | 88% | . Lack of Specificity Low accuracy |

Table 1 shows, deep learning and Machine learning techniques has invoked tremendous progress in automated image analysis. Before that, image analysis was commonly performed using systems fully designed by human domain experts. For example, such image analysis system could consist of a statistical classifier that used handcrafted properties of an image (i.e., features) to perform a certain task. Features included low level image properties such as edges or corners, but also higher-level image properties such as the speculated border of a cancer (Bas H.M. et al, 2022). In deep learning, these features are learned by a neural network (in contrast to being handcrafted) to optimally give a result (or output) given an input. An example of a deep learning system could be the output ‘GI Cancer’ given the input of an image showing a GI-cancer.

Various methods have been employed in Explainable AI to enhance the interpretability of medical imaging models. Among these methods, visual explanation techniques are prevalent. Techniques like Grad-CAM (Gradient-weighted Class Activation Mapping) and its variants such as guided Grad-CAM have shown high validity in providing visual explanations for model decisions (Arun et al., 2021). These methods highlight regions in the image that significantly contribute to the model's predictions, thereby helping clinicians understand and trust the model's decision-making process. Additionally, techniques like backpropagation and its guided version have been used, although they might sometimes emphasize edges rather than providing clear explanations (Adebayo et al., 2018).

Another set of techniques focuses on textual explanations, where models generate descriptive text to elucidate their predictions. For instance, image captioning methods generate text that describes the image content, providing an additional layer of interpretability. However, rigorous studies assessing the validity of textual explanations in medical imaging are limited, indicating a need for further research in this area (Arun et al., 2021). Example-based explanations, such as using triplet networks and influence functions, provide examples similar to the input image to help understand the model's reasoning. While these methods offer intuitive explanations, their robustness and validity in medical imaging contexts require more comprehensive evaluation.

The review of existing literature highlights significant advancements in the use of machine learning, particularly deep learning, for the automatic detection and classification of gastrointestinal (GI) tract pathological findingsfrom endoscopic images. Traditional diagnostic methods, relying heavily on manual analysis, are limited by operator expertise and the time-intensive nature of reviewing large volumes of endoscopic frames. Recent studies have demonstrated the potential of convolutional neural networks (CNNs), transfer learning, and explainable AI (XAI) techniques like Grad-CAM to address these challenges. While models such as ResNet, DenseNet, and MobileNet have shown promise, they often focus on a limited set of conditions, reducing their generalizability across diverse GI pathologies. Moreover, the “black-box” nature of these models raises concerns about clinical integration due to a lack of transparency in decision-making. This research aims to bridge these gaps by proposing a hybrid ResNet50-CNN model with integrated Grad-CAM visualization, providing both high accuracy and enhanced interpretability to support clinical decision-making.

# CHAPTER THREE METHODOLOGY

## 3.1 Research Design

This study employs a quantitative experimental design, using supervised machine learning techniques to classify GI tract pathological findings from endoscopic images. A hybrid model combining ResNet50 for feature extraction with a custom convolutional neural network (CNN) for classification was chosen due to its success in handling image-based tasks in medical diagnostics. The design is structured to achieve accuracy and interpretability through the integration of methods like Gradient-weighted Class Activation Mapping (Grad-CAM), which provides visual insights into the model’s decision-making process.

## 3.2 Location of the Study

The study will be conducted in a controlled lab setting using pre-collected, open-access datasets from international research institutions. The datasets—Kvasir and Gastrovision—were selected for their quality and relevance, containing expertly annotated images essential for training an effective model for GI tract disease classification.

# 3.3 Target Population/Materials

The target population in this study consists of images that represent a broad spectrum of gastrointestinal conditions. The Kvasir and Gastrovision datasets provide images across classes such as Esophagitis, Polyps, Ulcerative Colitis, and Normal findings. This diversity within the datasets ensures that the model is exposed to a range of GI conditions, supporting the development of a model that generalizes well across various pathological findings.

### 3.3.1 Kvasir Dataset

Contains images of different GI conditions and medical procedures, including anatomical landmarks and pathological abnormalities. The dataset includes over 4000 images, annotated by medical experts, which enhances the reliability of the data for training purposes.

### 3.3.2 Gastrovision Dataset

This multi-center dataset provides a wide array of images covering GI tract abnormalities, normal findings, and procedural cases, collected from institutions in Norway and Sweden. These images offer valuable diversity in terms of anatomical and pathological variations, aiding model robustness.

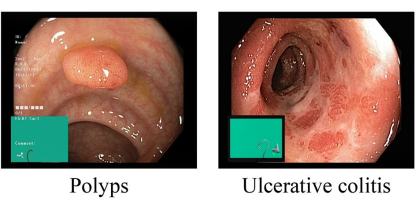


Figure 1: Sample images from the dataset

## 3.4 Sample Size and Sampling Techniques

The combined dataset includes approximately 8,000 images from Kvasir and a comparable number from Gastrovision, providing a large sample size for model training and evaluation. To ensure balanced learning, sampling methods will be applied to address class imbalance, particularly through data augmentation techniques such as flipping, rotation, and cropping. These augmentations will help mitigate the risk of overfitting by exposing the model to varied image transformations without altering the core content of each image.

## 3.5 Data Collection Techniques

The datasets, Kvasir and Gastrovision, are publicly available and have been pre-annotated by expert endoscopists, eliminating the need for additional data collection from human participants. The study will use all available classes to ensure the model's ability to generalize across a range of GI conditions. Each image will be resized and normalized as part of preprocessing to create uniform input dimensions for the model.

### 3.5.1 Data Augmentation

Augmentation techniques such as horizontal/vertical flips, random rotations, and brightness adjustments will be employed. These augmentations aim to simulate variability in real-world images, enhancing the model's generalization capabilities.

## 3.6 Model Development

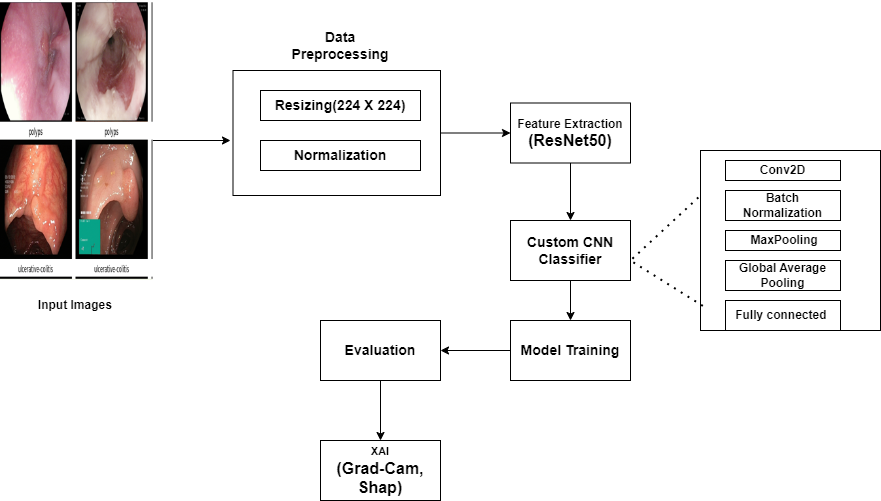


Figure 2:The proposed methodology

*Figure 2: The proposed methodology*

The model architecture comprises two main components: ResNet50 for feature extraction and a custom CNN for classification. The decision to use this hybrid model was based on ResNet50’s proven ability to capture intricate image features and the custom CNN’s flexibility in classifying these features into specific GI pathologies. The following steps outline the model's development process.

### 3.6.1 Feature **Extraction**

ResNet50, pre-trained on ImageNet, will be fine-tuned on the endoscopic image dataset. This pre-trained model provides a robust backbone for feature extraction, where its initial layers capture general image patterns, while later layers learn pathological findings specific features through fine-tuning.

### 3.6.2 Classification Network

The extracted features are passed into a custom CNN classifier. The custom **CNN** used for classification is built on top of the features extracted from the ResNet50 model. The model has an input layer that accepts feature maps of size 7x7x2048, corresponding to the output of the final convolutional block of ResNet50. This is followed by a convolutional layer with 32 filters, a 3x3 kernel size, ReLU activation, and 'same' padding, outputting a feature map of size 7x7x32. A Batch Normalization layer is applied next to stabilize the learning process, followed by a MaxPooling2D layer with a 2x2 pool size that reduces the spatial dimensions to 3x3x32. A second Conv2D layer with 64 filters, a 3x3 kernel size, and ReLU activation is then added, followed by another Batch Normalization layer and a second MaxPooling2D layer, reducing the output to 1x1x64. At this point, a Global Average Pooling (GAP) layer is used, which is used to condense the feature maps into a single vector of size 64, significantly reducing the number of parameters and mitigating overfitting

### 3.6.3 Grad-CAM Integration

To enhance interpretability, Grad-CAM is applied to generate heatmaps indicating regions in each image that contribute most significantly to the model’s classification. These visualizations provide insights for clinicians, increasing the model’s transparency and aiding in clinical decision-making.

## 3.7 Model Evaluation Metrics

The model's performance will be evaluated using the following metrics to ensure a comprehensive assessment of its diagnostic capabilities:

#### 3.7.1 Accuracy

The ratio of correctly classified images to the total number of images, providing a general measure of the model's performance.

#### 3.7.2 Precision

The proportion of true positive classifications out of all predicted positive cases, assessing the model's ability to avoid false positives.

#### 3.7.3 Recall (Sensitivity)

The proportion of true positive classifications out of all actual positive cases, evaluating the model’s capacity to detect GI pathological findings accurately.

#### 3.7.4 F1-Score

The harmonic mean of precision and recall, offering a balance between these two metrics, especially useful for imbalanced datasets.

#### 3.7.5 Specificity

The proportion of true negatives out of all actual negatives, measuring the model’s ability to avoid false negatives.

#### 3.7.6 AUC-ROC Curve

The Area Under the Receiver Operating Characteristic curve, illustrating the model’s diagnostic ability across all classification thresholds.

K-fold cross-validation (with 5 splits) will be employed to validate the model’s performance across multiple training and test splits. This technique ensures that the results are not biased by any specific subset of the dataset and provides an accurate reflection of the model’s robustness.

### 3.8 Data Analysis

Data analysis will focus on each evaluation metric for the model’s performance on various GI conditions. Descriptive statistics will summarize the model’s accuracy, precision, recall, F1-score, and specificity across folds. Visual analysis through Grad-CAM heatmaps will offer additional insights by highlighting the key image regions influencing the model’s decisions. Python libraries like TensorFlow, Keras, and Scikit-Learn will be used to implement and evaluate the model, while statistical libraries (e.g., Pandas, Matplotlib) will support data analysis and visualization.

## 3.9 Logistical and Ethical Considerations

### 3.9.1 Logistical Considerations

This study utilizes publicly available datasets, which means no additional data collection or patient involvement is required. The necessary computing resources for model training and testing will be provided through cloud-based or local high-performance computing facilities.

### Ethical Considerations

Since this study involves no direct human subjects and uses pre-annotated image datasets, it aligns with ethical guidelines for secondary data usage. However, maintaining data privacy and compliance with open-access licenses is essential. All datasets will be used solely for research, adhering to the terms of use specified by the data providers. Additionally, the research outputs will respect participant confidentiality, and no individual or patient-identifying information will be disclosed.

# References

Pogorelov, K., Randel, K. R., Griwodz, C., Eskeland, S. L., de Lange, T., Johansen, D., ... & Halvorsen, P. (2017, June). Kvasir: A multi-class image dataset for computer-aided gastrointestinal disease detection. In *Proceedings of the 8th ACM on Multimedia Systems Conference* (pp. 164-169).

Öztürk, Ş., & Özkaya, U. (2021). Residual LSTM layered CNN for classification of gastrointestinal tract diseases. *Journal of Biomedical Informatics*, *113*, 103638.

Saad, A. I., Maghraby, F. A., & Badawy, O. M. (2024). PolyDSS: computer-aided decision support system for multiclass polyp segmentation and classification using deep learning. *Neural Computing and Applications*,*36*(9), 5031-5057.

Ferlay J, Ervik M, Lam F, Laversanne M, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I, Bray F (2024). Global Cancer Observatory: Cancer Today (version 1.1). Lyon, France: International Agency for Research on Cancer. Available from: https://gco.iarc.who.int/today.

PAHO. The Burden of Digestive Diseases in the Region of the Americas, 2000–2019; Pan American Health Organization: Washington, DC, USA, 2021.

Sharma, N., Gupta, S., Koundal, D., Alyami, S., Alshahrani, H., Asiri, Y., & Shaikh, A. (2023). U-Net model with transfer learning model as a backbone for segmentation of gastrointestinal tract. *Bioengineering*, *10*(1), 119.

Pogorelov, K., Randel, K. R., Griwodz, C., Eskeland, S. L., de Lange, T., Johansen, D., ... & Halvorsen, P. (2017, June). Kvasir: A multi-class image dataset for computer-aided gastrointestinal disease detection. In *Proceedings of the 8th ACM on Multimedia Systems Conference* (pp. 164-169).

Gunasekaran H, Ramalakshmi K, Swaminathan DK, J A, Mazzara M. GIT-Net: An Ensemble Deep Learning-Based GI Tract Classification of Endoscopic Images. Bioengineering (Basel). 2023 Jul 5;10(7):809. doi: 10.3390/bioengineering10070809. PMID: 37508836; PMCID: PMC10376874.

Yaqoob MK, Ali SF, Bilal M, Hanif MS, Al-Saggaf UM. ResNet-Based Deep Features and Random Forest Classifier for Diabetic Retinopathy Detection. Sensors (Basel). 2021 Jun 4;21(11):3883. doi: 10.3390/s21113883. PMID: 34199873; PMCID: PMC8200077.

WHO, "Cancer Diseases " 3 March 2021.

World Health Organization, "Cancer cases over the world," 2022

Ali S., Ghatwary N., Braden B., Lamarque D., Bailey A., Reardon S., et al., "Endoscopy disease detection challenge 2020," arXiv preprint arXiv:2003.03376, 2020.

Kumar A., Singh S. B., Satapathy S. C., and Rout M., "MOSQUITO-NET: A deep learning based CADx system for malaria diagnosis along with model interpretation using GradCam and class activation maps," Expert Systems, p. e12695, 2021.

Zhang Z., Wang L., Zheng W., Yin L., Hu R., and Yang B., "Endoscope image mosaic based on pyramid ORB," Biomedical Signal Processing and Control, vol. 71, p. 103261, 2022.

Rahim, T., Hassan, S. A., & Shin, S. Y. (2021). A deep convolutional neural network for the detection of polyps in colonoscopy images. *Biomedical Signal Processing and Control*, *68*, 102654.

Luo, H., Xu, G., Li, C., He, L., Luo, L., Wang, Z., ... & Xu, R. H. (2019). Real-time artificial intelligence for detection of upper gastrointestinal cancer by endoscopy: a multicentre, case-control, diagnostic study. *The Lancet Oncology*, *20*(12), 1645-1654.

Klang, E., Barash, Y., Margalit, R. Y., Soffer, S., Shimon, O., Albshesh, A., ... & Kopylov, U. (2020). Deep learning algorithms for automated detection of Crohn’s disease ulcers by video capsule endoscopy. *Gastrointestinal endoscopy*, *91*(3), 606-613.

Ding, Z., Shi, H., Zhang, H., Meng, L., Fan, M., Han, C., ... & Hou, X. (2019). Gastroenterologist-level identification of small-bowel diseases and normal variants by capsule endoscopy using a deep-learning model. *Gastroenterology*, *157*(4), 1044-1054.

Stidham, R. W., Liu, W., Bishu, S., Rice, M. D., Higgins, P. D., Zhu, J., ... & Waljee, A. K. (2019). Performance of a deep learning model vs human reviewers in grading endoscopic disease severity of patients with ulcerative colitis. *JAMA network open*, *2*(5), e193963-e193963.

Alagappan, M., Brown, J. R. G., Mori, Y., & Berzin, T. M. (2018). Artificial intelligence in gastrointestinal endoscopy: The future is almost here. *World journal of gastrointestinal endoscopy*, *10*(10), 239.

Sharif, M., Attique Khan, M., Rashid, M., Yasmin, M., Afza, F., & Tanik, U. J. (2021). Deep CNN and geometric features-based gastrointestinal tract diseases detection and classification from wireless capsule endoscopy images. *Journal of Experimental & Theoretical Artificial Intelligence*, *33*(4), 577-599.

# Appendices

## Appendix 1: Work Plan

