A Deep Learning Approach to Classify Colon Diseases

By
Mist. Mariya Hossan Shayle
213-15-4303

FINAL YEAR DESIGN PROJECT REPORT

This Report Presented in Partial Fulfillment of the Requirements for the
Degree of Bachelor of Science in Computer Science and Engineering

Supervised by

Mr. Amir Sohel
Assistant Professor

Department of Computer Science and Engineering
Daffodil International University

Co-Supervised by

Mr. Md. Shahriar Shakil
Lecturer

Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
Dhaka, Bangladesh

September 16, 2025

APPROVAL

This Project titled **A Deep Learning Approach to Classify Colon Diseases**, submitted by Mist. Mariya Hossan Shayle, ID No: **213-15-4303** to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on **16 September, 2025**.

BOARD OF EXAMINERS

Dr. S. M Aminul Haque
Professor and Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Chairman

Dr. Md. Alamgir Kabir
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Mr. Md. Assaduzzaman
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Dr. Md. Zulfiker Mahmud
Professor
Department of Computer Science and Engineering
Jagannath University

External Examiner

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Mr. Amir Sohel, Assistant Professor,** Department of Computer Science and Engineering, Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

Mr. Amir Sohel
Assistant Professor
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised by:

Mr. Md. Shahriar Shakil
Lecturer
Department of Computer Science and Engineering
Daffodil International University

Submitted by:

Mariya

Mist. Mariya Hossan Shayle
Student ID: 213-15-4303
Department of Computer Science and Engineering
Daffodil International University

ACKNOWLEDGEMENTS

This work would not have been possible without the support and contributions of many individuals over the past two semesters. We are deeply grateful to everyone who has assisted us in one way or another.

First, we express our heartfelt thanks and gratefulness to the almighty for His divine blessing making it possible for us to complete the **Final Year Design Project(FYDP)** successfully.

We are grateful and wish our profound indebtedness to **Mr. Amir Sohel, Assistant Professor,** Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh. Deep knowledge and keen interest of our supervisor in the field of deep learning to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

We would like to express our heartfelt gratitude to the Head of the Department of Computer Science and Engineering, for his kind help in finishing our project and also to other faculty members and the staff of the Department of Computer Science and Engineering, Daffodil International University.

We would like to thank our entire course-mates at Daffodil International University, who took part in this discussion while completing the coursework.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

ABSTRACT

Early and proper diagnosis of Colon diseases is crucial, as it is essential to successful clinical treatment and patient outcomes. InceptionResNet V2, Traditional convolutional neural network architectures (Xception and ConvNeXt-Tiny) can theoretically be used to label medical imaging. However, they are expensive, uninterpretable, or fail to perform fine-grained discriminative image recognition tasks on complex endoscopic images. We utilize CareNet, a simple yet highly discriminative deep learning model, to address these challenges specifically in colon disease classification. CareNet applies both EfficientNet-B0 (as a baseline, founded on average pooling worldwide, max pooling worldwide, channel-gating and refining convolutional pooling, and an attention-pooling mechanism. This design is more computationally efficient, experienceable in the context of global features, and sensitive to local features. This analysis of the colon endoscopy dataset on a benchmark has identified that CareNet performs well, achieving 99.22% classification accuracy on the colon endoscopy dataset, compared to state-of-the-art models on the same dataset. Moreover, the results of cross-validation prove its strength and ability to generalize to other folds of data. The CareNet, which proposes a solution for clinical decision-support in diagnosing colon disease balance using real-world data, is more precise, effective, and comprehensible compared to existing studies.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

This chapter discusses the applicability of diagnosing colorectal disease, with special reference to the possibility that endoscopic image analysis can be improved using deep learning.

1.1 Introduction

The colon is a long, tube-like structure with a length of approximately 150 cm, and it features taeniae coli, haustra, and appendices epiploicae. These characteristics make it appear segmented and distinguishable from the small intestine [1]. Colon diseases comprise a wide range of gastrointestinal disorders that have severe health effects on the population across the globe. Several other disorders are known to have precisely been associated with the colon such as inflammatory diseases which include ulcerative colitis, structural defects such as colon polyps, and other related gastrointestinal disorders that can affect their quality of life [2]. The severity of colon disease indicates gut inflammation at a specific point in time, and the total, long-term disease burden is affected by various clinical and quality-of-life factors [3]. According to the study of Global Burden of Disease in 2019,, digestive diseases are prevalent worldwide, resulting in 2.86 billion cases, 8 million deaths, and 277 million disability-adjusted life years (DALYs) recorded across 204 countries and territories [4]. Recent studies in Bangladesh emphasize the growing prevalence of gastrointestinal disorders in different populations. For instance, a cross-sectional survey of 1,019 undergraduates found that 38.24% were affected by functional gastrointestinal disorders, with constipation as the most prevalent subtype, and many lifestyle and psychological factors were found to have a significant relationship with the condition [5]. Similarly, in rural settings, a door-to-door survey of 3,351 adults revealed that 33.1% of the adults experienced esophageal symptoms (mainly heartburn), and female gender, low income, and psychological distress were identified as risk factors [6]. Among the students of the Dhaka-based private university, the irritable bowel syndrome prevalence was found to be 31.63% and its association was significantly associated with anxiety, stress, diet, and lifestyle [7]. In addition, a massive surveillance study in the coastal area of Bangladesh has shown that 7% of the participants expressed gastrointestinal symptoms, especially diarrhea, which was more prevalent in women, older people, low-income earners, and those who lived near the sea [8].

The initial clinical assessment of colon diseases traditionally starts with a detailed clinical history, physical examination, and a panel of non-invasive laboratory tests, most often complete blood count (CBC) and C-reactive protein (CRP). Doctors diagnose by direct visualization of the mucosa of the colon by colonoscopy or endoscopy, which offers the opportunity to obtain tissue to confirm the diagnosis by histopathology, as well as providing ancillary structural evaluation through the use of imaging techniques like computed tomography (CT) scans or barium enema [9]. With rising technologies, deep learning is also leaving an impact on the medical field. Much research has been conducted on these technologies to improve disease diagnosis and treatment. For example, Nadim et al. (2024) [10] used image resizing, contrast enhancement, gamma correction, and data augmentation to improve model generalization. Conversely, a hybrid architecture combining re-trained deep learning models with Kernel Extreme Learning Machine (KELM) and Mutation Boosted Dwarf Mongoose Optimization Algorithm (MB-DMOA) optimization, introduced by Gowthamy and Ramesh (2024) [11], is proposed as a suitable way to diagnose lung and colon cancer based on histopathology pictures correctly. In the same way, Nadimi et al. (2025) [12] used the CAM-based ways, as well as CartoonX and Pixel Rate Distortion Explanations, to make the model more open to interpretation, hoping to achieve a higher level of trust, robustness, and fidelity of the classification process by the use of real rendering poly data.

In light of these advances, we employed various preprocessing techniques, including adjusting brightness and contrast, Contrast Limited Adaptive Histogram Equalization (CLAHE), image equalization, geometrical rotation, blurring with a Gaussian filter, and adding noise with a Gaussian filter, to enhance the generalization ability of our structure. The outputs of the quantitative assessment of the processed images were the Structural Similarity Index (SSIM), Peak Signal-to-Noise Ratio (PSNR), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). Our hybrid neural network design is an efficient network featuring a convolutional block attention (CBAM) module, which is utilized to detect colon diseases in endoscopic images actively. To evaluate the predictive quality and computer efficacy, we compared the computer model with various pre-trained architectures using three metrics: dependability, rationale of damage, and mean training times in each epoch. We then tested the given model using it as a mere data analysis tool with cross-validation and Scott-Knott statistics. Moreover, the model was combined with explainable artificial intelligence (XAI) methods to enhance the interpretability, plausibility, and clinical credibility of the predictions made by the model.

1.2 Motivation

Manual interpretation of endoscopic images is a tedious yet specialized process. There is also a significant possibility of inter-observer variability, and consequently, the risk of interference with the validity of the diagnosis and its implications for the patient. Designing an automatic system that currently achieves visual separation of similar diseases, with a prudent yet uncomplicated architecture that can overcome noise and natural variations in medical imaging images, is among the foremost computational targets. Attention-based models or deep learning are likely to enable a higher likelihood of better diagnostic throughput and accuracy, along with the opportunity for customized data augmentation pipelines.

Not only will the resolution of this diagnostic problem have a positive impact on the field of medical image analysis, but it will also lead to the development of diagnostic instruments that can be scaled and have a high level of reliability, enabling clinicians to make more informed decisions based on them. In addition, it will result in a greater quality of knowledge in the expertise of state-of-the-art convolutional neural network designs and domain-specific augmentation techniques, which will be pivotal in the innovation of artificial intelligence to aid healthcare.

1.3 Objectives

- Develop an automated deep learning model capable of classifying endoscopy images and identifying diseases such as ulcerative colitis, polyps, and esophagitis.
- Develop and apply class-specific data augmentation to the image sets to enhance model robustness, albeit at the expense of diminished diagnostic features.
- Make the dataset fit for training, validation, and testing, ensuring proper representation of each class.
- Identify and resolve issues that cause the network to be slow and difficult to apply in various situations.
- Evaluate the performance of the proposed model rigorously using standardized validation and test sets, thereby ensuring its generalizability and clinical relevance.

1.4 Methodology

We employed a multi-class endoscopy image dataset comprising three disease classes (ulcerative colitis, polyps, and esophagitis) and one regular class in this study. Class-specific augmentations in the Albumentations library were used to augment the data, making it more robust, and to upsample. In the case of classes of non-ulcerative lesions, we used uniform enhancement, which includes brightness and contrast adjustments, adaptive enhancement, contractive rotation, blurring, denoising, and resizing. Conversely, the samples of ulcerative colitis were heterogeneity-enriched with regulated extrapolation of the pathological phenotype.

Any endoscopic image was scaled to 224 x 224 and divided into a 70:15:15 ratio for training, validation, and test sets. On the preprocessing side, ImageDataGenerator was employed to generate the image transformations required by EfficientNetB0.

Based on this, the CBAM (Convolutional Block Attention Module) is suggested as a means of accentuating both the spatial and channel-wise discriminative features. A convolutional neural network, named CareNet, is then developed on top of the EfficientNet backbone. Mixed-precision training had a lower latency and generated less memory overhead. Besides, it enhances the strength in optimization and the capability of generalization. To achieve this, label smoothing was employed, and AdamW with firm and cosine decay learning rate schedules were implemented. The process of training the model involved employing a mechanism of early stopping and model checkpointing to determine the correct weights, which were used to prevent overfitting.

Moreover, the test set was applied systematically to evaluate the performance of the fully trained CareNet model's classification. Unbiased performer and advanced augmentation and preprocessing: We have demonstrated that augmentations achieved through both custom preprocessing and data augmentation, as well as the augmentation of an attention-based architecture, provide benefits for colon disease classification. This demonstrates the model's robustness and relevance to the clinical setting in which it is applied.

1.5 Project Outcome

The possible outcomes of this project are:

- A high-performance classification model capable of identifying various types of colon diseases based on endoscopy images.
- A more powerful generalization is through the discretion of class-specific augmentation and attention mechanisms, resulting in remarkable output with wide imaging variations.
- Better explainability by use of explainable AI (XAI) techniques, e.g., Grad-CAM, Score-CAM, LIME, and SHAP, which provide visual explanations to the decisions of the model.
- A benchmark system that can serve as a generic one for other classification problems of medical images with similar trouble.
- This may be a potential clinical support tool that assists gastroenterologists in making faster and more precise image diagnoses, thereby limiting subjectivity in image interpretation.

1.6 Organization of the Report

This report is divided into six chapters:

I. Chapter 1 yields the background of the research, including its motivation, aims, the chosen methodology, and a general outline of the research.
II. Chapter 2 gives the background study. The literature review acknowledges the current research and trends in wireless capsule endoscopy, and then proceeds to analyze comparable applications, followed by a gap analysis that identifies the research need.
III. Chapter 3 outlines the research methodology, including requirements analysis, system architecture, functional and non-functional specifications, data flow diagrams, user interface design, and the project plan and task assignment.
IV. Chapter 4 records the implementation and results of the system, including the research scenario, test process, assessment requirements, system performance, comparative analysis, and discussion of the findings.
V. Chapter 5 addresses engineering standards and design issues, which encompass compliance with applicable software, hardware, and communications standards, as well as societal, environmental, ethical, and sustainability considerations.
VI. Lastly, the work is summarized in Chapter 6, including key findings, limitations, and future research directions.

Chapter 2

Background

This chapter provides the background, motivation, and an overview of existing studies on the topic, highlighting the need for automated identification of colon diseases using deep learning. It also conducts a literature review, highlighting existing problems and defining the research gap that the proposed CareNet model attempts to fill.

2.1 Introduction

The potential of deep learning combined with a hybrid computational environment has brought swift changes in colon disease detection. Various reports have suggested convolutional neural networks (CNNs) and hybrid models to improve the choice of diagnosis based on histopathological and colonoscopy image data. Nevertheless, most methods suffer from weaknesses related to their inability to explain findings, computation-intensity, poor feature selection, small diversity in the dataset, or uncertain effectiveness of tests being developed. In order to better understand these challenges, one may refer to a comparative analysis of recent literature, pointing out the areas where particular approaches have specific gaps and what the following literature has provided in the form of innovative ideas. Besides following the development of model architecture and preprocessing approaches, this review also points out major areas in which models showed potential to be advanced in accuracy, interpretability, and clinical applicability.

2.2 Literature Review

Several studies have examined CNN and hybrid lung and colon cancer diagnosis methods with the LC25000 data. Hasan et al. (2024) [13]. introduced LW-MS-CCN, integrated with XAI, achieving an accuracy of 99.20%, but it still has overfitting, and clinical generalisability is possible. Similarly, Ochoa-Ornelas et al. (2024) [14]. presented a hybrid MobileNetV2-EfficientNetB3 that uses MEGWO-LCCHC as optimization and achieved 94.8% accuracy, while Al-Mamun Provath et al. (2023) [17]. presented a comparison between standard CNNs and their proposed model, and their proposed model reached an accuracy of 97.0%. Hadiyoso et al. (2023) [19]. further optimized the classification using VGG16-CNN with CLAHE, which provided 98.96 % These studies are very accurate, but they typically are heavy on LC25000, scaleability, computational requirements, and evaluation metrics.

Other authors worked on the issue of feature selection and other datasets. For example, Mohamed et al. (2023)[16]. used CNN models on 10,000 colon images with GOA-based feature reduction, achieving several models above 98% accuracy; however, models were limited in complexity. Karthikeyan et al. (2024) [15]. created a CNN using the ranking algorithms on a set of 334 patient images and saw relatively higher scores than ANN and BPNN (91% accuracy); however, feature selection made the procedure more complex. In addition to histopathology, CT colonography data were used by Kumar et al. (2023) [18].; they achieved an accuracy of 94.165% using GoogLeNet; however, their model required large amounts of data for generalization.

The clinically oriented publications were not so promising. Schiele et al. (2021) [20]. used InceptionResNet V2 to predict the presence of colon cancer in 291 patients. The model they trained (Big-CoMet) achieved 80% accuracy, and the limits of their model were only constrained by the dataset (a single sample per patient). Wang et al. (2021) [21]. achieved more than 98% accuracy through GAP-based VGGNet and ResNet models on colonoscopy images for polyp detection. However, they still cannot hold out against the occlusion phenomenon. By comparison, using the much more narrow in scope and small by comparison Warwick-QU dataset (165 images), Sarwinda et al. (2021) [22]. only achieved 85-88% accuracy with ResNet18 and ResNet50, which proves that larger and more diverse datasets matter. Table 2.1 depicts the summary of literature reviews.

Table 2.1: Summary of Literature Reviews.

Author (s)	**Year**	**Title**	**Methodology**	**Key Findings**
Hasan et al.	2024	An End-to-End Lightweight Multi-Scale CNN for the Classification of Lung and Colon Cancer with XAI Integration	Quantitative Analysis	High-performing CNN models have the potential to be useful in clinical diagnosis without being able to generalize on uncontrolled data.
Ochoa-Ornelas et al.	2024	A Hybrid Deep Learning and Machine Learning Approach with Mobile-EfficientNet and Grey Wolf Optimizer for Lung and Colon Cancer Histopathology Classification	Quantitative Analysis	Hybrid optimization framework is more accurate in diagnosing than the traditional ML models, at the cost of computation speed.
Karthikeyan et al.	2024	Colorectal cancer detection based on convolutional neural networks (CNN) and ranking algorithm	Quantitative Analysis	Ranking-integrated CNN outperforms ANN/BPNN techniques in detecting colorectal cancer to a huge extent.
Al-Mamun Provath et al.	2023	Classification of Lung and Colon Cancer Using Deep Learning Method	Quantitative Analysis	Grasshopper optimization is efficient at enhancing the performance of a CNN in various architectures in the case of binary classification tasks.
Mohamed et al.	2023	Colon Disease Diagnosis with Convolutional Neural Network and Grasshopper Optimization Algorithm	Quantitative Analysis	New CNN architecture achieves better results than known models (VGG16, ResNet50) in the multi-classification of histopathology.
Kumar et al.	2023	Comparative Assessment of Colon Cancer Classification Using Diverse Deep Learning Approaches	Quantitative Analysis	Transfer learning (AlexNet, GoogLeNet) yields better results in detection of colon cancer based on CT.
Hadiyoso et al.	2023	Diagnosis of lung and colon cancer based on clinical pathology images using	Quantitative Analysis	The preprocessing of histogram equalization and VGG16 give an almost perfect solution to the

		convolutional neural network and CLAHE framework		classification of the histopathological sample.
Schiele et al.	2021	Deep Learning Prediction of Metastasis in Locally Advanced Colon Cancer Using Binary Histologic Tumor Images	Case Study	Training of Single-section InceptionResNetV2 gives clinically viable metastasis prediction in advanced adenocarcinoma.
Wang et al.	2021	An improved deep learning approach and its applications on colonic polyp images detection	Case Study	The high-accuracy Pediatric colonoscopy images can be detected in pooling modification of global average.
Sarwinda et al.	2021	Deep Learning in Image Classification using Residual Network (ResNet) Variants for Detection of Colorectal Cancer	Case Study	ResNet50 has a better performance than ResNet18 when detecting colorectal cancer but has worrying sensitivity variables.

2.3 Gap Analysis

We conducted a gap analysis of the existing literature to position the proposed CareNet model better. This discussion highlights the strengths and weaknesses of previous studies, as they must be accurate, robust, efficient, and have a diverse dataset that can be interpreted and have clinical utility. The table 2.2 depicts the gap analysis.

Table 2.2: Gap analysis of existing research and comparison with our proposed model

Features	**Hasan etal. (2024)**	**Ochoa Orne Asetal (2024)**	**Al-Mamun Provath etal.(2023)**	**Hadiyoso et al. (2023)**	**Mohamed et al. (2023)**	**Karthikeyan et al. (2024)**	**Kumar et al. (2023)**	**Schiele et al. (2021)**	**Wang Et al. (2021)**	**Sarwinda Et al. (2021)**	**Proposed CareNet (2025)**
High Classification accuracy	yes	Moderate	yes	yes	yes	Moderate	Moderate	low	yes	low	yes
Robust against overfitting	No	No	No	No	No	No	No	No	Partial	No	yes
Light architecture	No	No	No	No	No	No	No	No	No	No	yes
Computational efficiency	No	No	No	No	No	No	No	No	No	No	yes
Large & diverse dataset usage	No	No	Limited	Yes (LC25000)	yes	No (334 images)	Yes	No	Moderate	Very limited(165 images)	yes
Feature selection integrated	No	No	No	No	yes (GOA)	yes	No	No	No	No	Yes (attention + pooling)

XAI support	Partial	No	No	No	No	No	No	No	No	No	Yes
Clinical applicability	Limited	Limited	Limited	Limited	Limited	Limited	Limited	Limited	Limited	Very Limited	yes

2.4 Summary

In short, although previous studies to detect colon diseases demonstrated high accuracy, most of them had limitations like their possible overfitting, excessive computational complexity, small or insufficiently diverse datasets, and feature selection issues, the lack of adequate evaluation indicators, and the lack of interpretability of the results. We addressed these limitations by combining our lightweight EfficientNetB0-CBAM hybrid network with state-of-the-art preprocessing and validation methods, offering efficiency, reliability, and clinical practicability.

Chapter 3

Research Methodology

This section presents the systematic process followed in designing, training and testing the planned CareNet model. It outlines the arrangement of the datasets, preprocessing methods, model design, training processes, and assessment methods, and presents a clear roadmap of the experiment framework.

3.1 Methodology

The proposed framework for colon disease diagnosis is divided into four stages: preprocessing, model development, evaluation and clinical validation. Preprocessing included brightness-contrast adjustment, CLAHE, histogram equalization, rotation, and Gaussian blurring with noise up to resizing, with data partitioned to training, validation and testing datasets (70:15:15). The model CareNet based on EfficientNet-B0 with adding attention layers and custom classifiers was evaluated with the measures of accuracy, precision, recall and F1-score while explainable AI was incorporated to interpretability. Benchmarking experiments against Xception, InceptionResNetV2 and ConvNextTiny demonstrated CareNet's outstanding performance and clinical potential.

3.1.1 Overview

There are four stages in the automated colon disease diagnosis process: data collection and preprocessing, model structure and code development, performance and simulation, and clinical evidence. Image quality was enhanced and variability was counted using the following preprocessing schemes: brightness-contrast optimization, CLAHE, histogram equalization, rotation, noisy Gaussian blurring, and resizing. The endoscopic dataset was divided into a ratio of 70:15:15 to ensure strict evaluation and prevent overfitting, resulting in training, validation, and test sets.

The CareNet design proposed is an integration of EfficientNet-B0, efficient feature extraction, and spatial and channel-wise attention to identify clinically interesting patterns. These enhanced characteristics can then be mapped to disease categories by a custom classifier. Accuracy, precision, recall, and F1-score metrics were used to evaluate the CareNet methods, and explainable artificial intelligence techniques were used to achieve a thorough insight into the predictions. Compared to other benchmark models, like Xception, InceptionResNetV2 and ConvNeXt-Tiny, CareNet showed improved performance. Lastly, clinical validation was performed to demonstrate that the framework not only presented good experimental results but also showed reliability and scalability in a natural diagnostic context.

3.1.2 Proposed Methodology

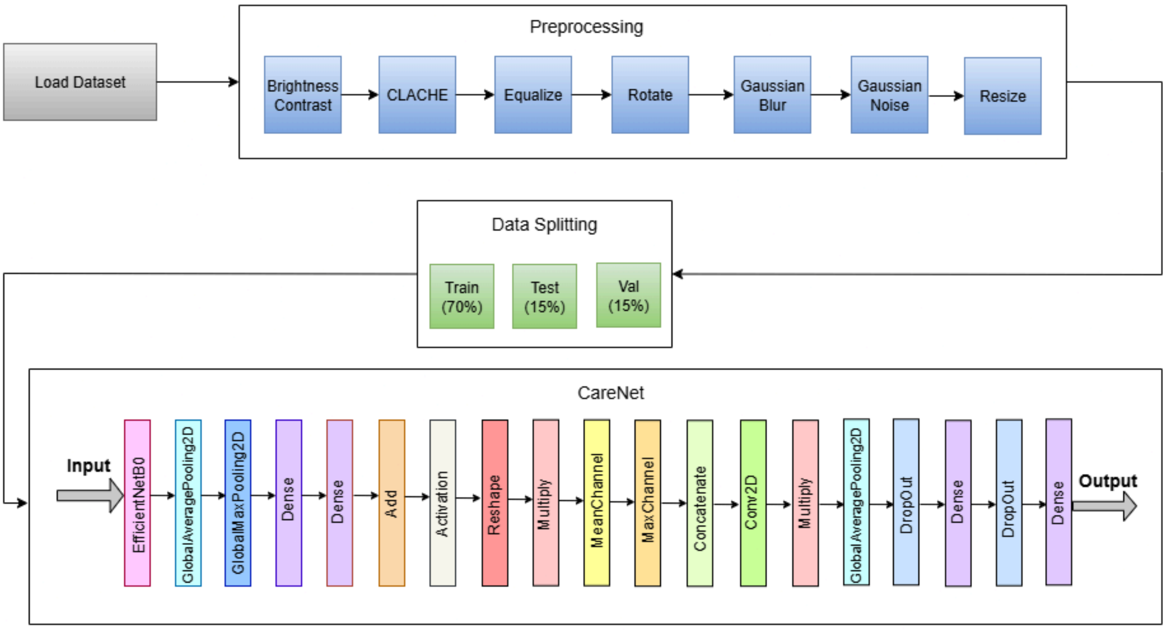


Fig 3.1: Proposed Methodology of CareNet

Figure 3.1 illustrates the proposed methodology of CareNet. The suggested colon disease classification scheme combines state-of-the-art image preprocessing, data partitioning systems, and a custom deep learning architecture to improve diagnostic accuracy. The pipeline starts with a thorough preprocessing phase, where the raw endoscopic data is enhanced with brightness-contrast adjustment, Contrast-Limited Adaptive Histogram Equalization (CLAHE), equalization, rotation, Gaussian blurring, Gaussian noise injection, and resizing. Such operations enhance the visual quality and robustness to illumination, directional changes, and imaging noise changes.

After preprocessing the images, the dataset is split into training, validation, and test sets separately in the ratio of 70:15:15 to support efficient model training, hyperparameter optimization, and objective evaluation.

CareNet, a hybrid architecture incorporating EfficientNetB0 as the backbone in hierarchical feature extraction, realizes the classification stage. EfficientNetB0 offers computational efficiency, and the extracted features are optimized with the help of a convolutional attention module. Specifically, Mean-Channel and Max-Channel operations attend to channel- and spatial-level saliency, allowing the network to focus on disease-relevant areas. Such attention-boosted features are then processed by convolutional, pooling, and dense layers, with dropout regularisation helping to prevent over-fitting. Lastly, a softmax-activated and fully connected dense layer provides the probability of the four diagnostic categories in a single layer.

In general, the offered methodology integrates state-of-the-art convolutional modelling, attention, and large-scale preprocessing to ensure robustness and high accuracy in classifying colon disease image.

3.2 Detailed Methodology and Design

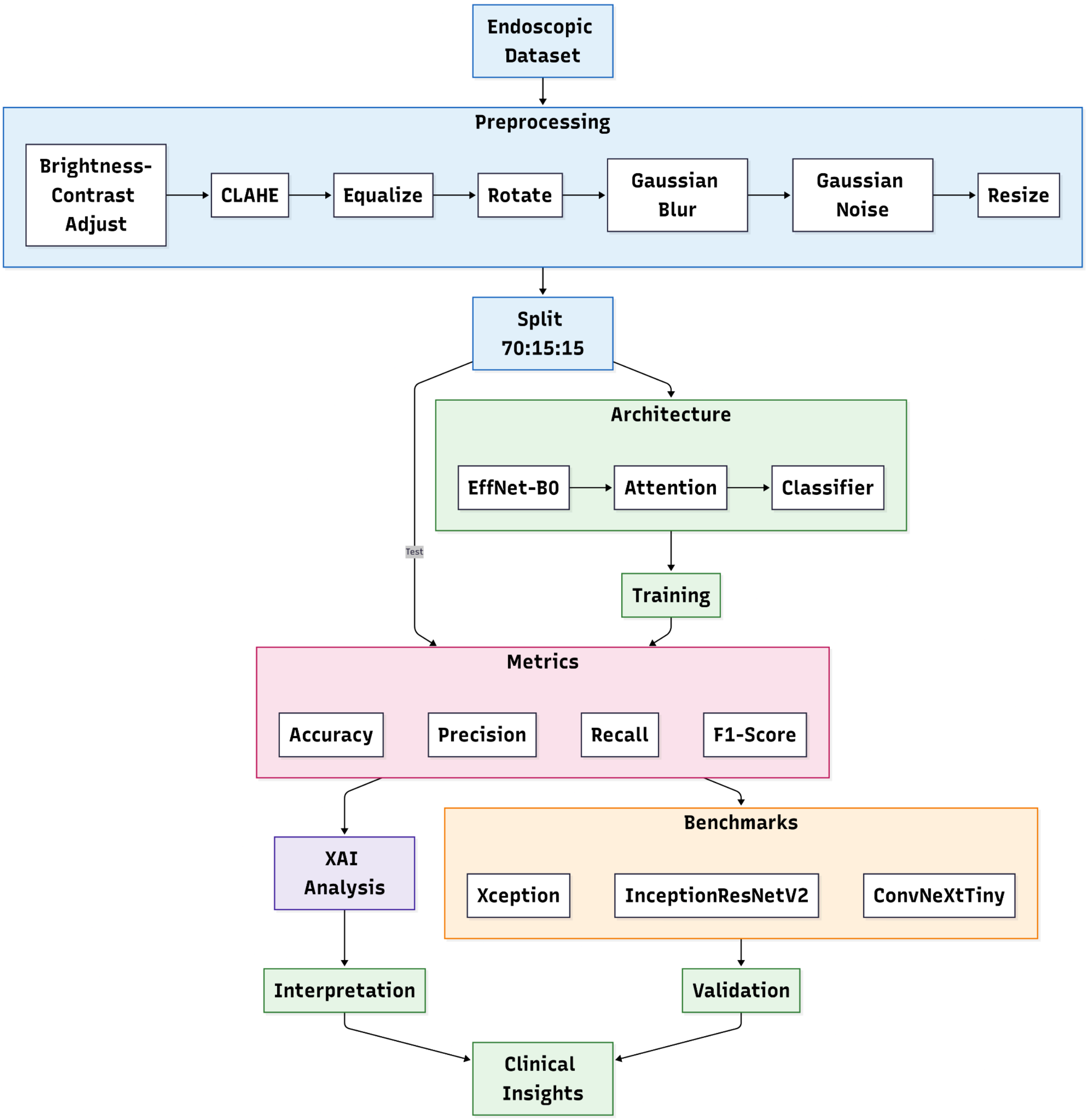


Fig 3.2: Detailed Methodology of CareNet

Figure 3.2 shows the systematic automatic colon disease diagnosis workflow based on endoscopic images. The proposed paradigm is divided into four stages: data acquisition and preprocessing, model architecture and development, performance evaluation, and clinical validation. Preprocessing included brightness-contrast adjustment, Contrast-Limited Adaptive Histogram Equalisation (CLAHE), rotation, equalization, Gaussian blurring combined with noise addition, and further resizing input dimensions. The dataset was partitioned into the training, validation, and test sets, applying a ratio of 70:15:15 to provide a stringent evaluation.

The CareNet architecture implements an EfficientNet-B0 backbone for feature extraction, spatial and channel-wise attention layers to enhance the saliency of clinically relevant regions, and a custom-designed classifier for which disease classification is completed. Supervised training methods were used for the model learning, and accuracy, precision, recall, and F1 score were the endpoints for model evaluation. Explainable AI visualizations and opacities were provided as a means of interpretability. Competitive benchmarking against the widely known backbone models, such as Xception, InceptionResNetV2, and ConvNextTiny, proved the performance of our approach.

Lastly, the workflow is geared towards clinical translation, as demonstrated by statistically proven results on independent datasets, the interpretation of model outputs into action-oriented diagnostics, and testing on real-world data to assess reliability. These automated methods for gastrointestinal disease classification provide an end-to-end solution that meets the requirements of rigorous methods, yielding strong, translational, and interpretable results.

3.3 Project Plan

The implementation of the planned CareNet structure is carried out according to a plan to maintain systematized implementation and replenishment. The project is subdivided into different phases that address a major object of the working process such as the work with the dataset preparation, up to clinical validation.

Phase 1 Data Acquisition and Preprocessing.

- Endoscopic image datasets of many different disease classes and normal cases are collected.
- Use of preprocessing techniques such as adjustments in brightness and contrast, CLAHE, histogram equalization, Gaussian blurring and noise addition, rotation and resizing.
- A 70:15:15 ratio division of the endoscopic dataset into training, validation, and test sets to enable equal consideration in appraisal.

Phase 2 - Model Development

- The setting up of EfficientNet-B0 to use as the feature extractor.
- Application of individually-designed pooling and attention-based modules.
- Final CareNet architecture with colon disease-specific classifier layer is built.

Phase 3 - Training and optimization.

- AdamW optimizer, cosine decay learning rate schedule and label smoothing model training.
- Mixed-precision training to cut down computation time.
- Early stopping, dropout, and model checkpointing as techniques to help prevent overfitting.

Phase 4 - Assessment and benchmarking.

- Accuracy // Precision // Recall // F1-score as main measures to measure the performance.
- Comparison with Benchmark Models such as InceptionResNetV2, Xception and ConvNeXt-Tiny.
- Multiple fold statistical validation is done to determine the generalizability and the strength.

Phase 5 - Explainability and Clinical Validation.

- Explanation AI (XAI) visualization using explainable AI (XAI) like Grad-CAM, Grad-CAM++, Score-CAM, SHAP and LIME.
- R2 values analysed in order to understand the model.
- Which is the ultimate test validation on independent test sets.

Phase 6 - Reporting and Documentation.

- Self-organization of experimental data, figures, its analysis and publication in academic volumes is included.
- Translating clinical insight and advice from frontline practice and practicalities.

3.4 Task Allocation

Table 3.1 depicts the timeline of the principal activities in each period of the project, from week 12 to week 46.

Table 3.1: Task allocation with estimated and actual work period

Tasks	Weeks																	
	12	14	16	18	20	22	24	26	28	30	32	34	36	38	40	42	44	46
Data Acquisition and Preprocessing																		
Model Development																		
Training and Optimization																		
Evaluation and Benchmarking																		
Explainability and Clinical Validation																		

Reporting and Documentation																		

Estimated Work Period	
Actual Work Period	

3.5 Summary

In this section, the approach taken in developing the proposed CareNet framework on colon disease classification was elaborated. The data were preprocessed and augmented to ensure high-quality and balanced input data. CareNet architecture designs have been developed, incorporating efficient learners by combining EfficientNet-B0 with attention and pooling to achieve effective feature learning. Evaluation used standard metrics and explainable AI to confirm accuracy, reliability and interpretability. They also described a planned project with a specific timetable and several distinct phases that had to be followed, which would lead to an ordered implementation and process completion. In general, the methodology contains a rigorous, transparent, and clinically relevant data-to-deployment pipeline.

Chapter 4

Implementation and Results

This chapter describes the practical implementation of the CareNet project, including the development environment, preprocessing pipeline implementation, and model architecture and results.

4.1 Environment Setup

A reproducible and managed computational environment would be of key importance to the development, as well as training, of deep learning models. The following setup has been done for the implementation and experimenting of the CareNet system.

A. **Hardware Configuration:**
    a. CPU: Intel(R) Core-i5(TM) 11 Rockestron Processor
    b. RAM: 16 GB DDR4.
    c. Storage: 1TB NVMe SSD for High Speed data loading/model checkpointing
B. **Programming library and these Framework:**
    a. Operating System: Windows 11 pro
    b. Development Languagepython 3.8.10
    c. Deep Learning Framework - Tensorflow 2.9.1 + Keras API range of model-building, training and evaluation
C. **Key Python Libraries:**
    a. OpenCV 4.5.4: Ensuring Calculations for image preprocessing (CLAHE, filtering, resizing)
    b. NumPy 1.21.5 & Pandas 1.4.4 For efficient numeric operations and manipulating data
    c. scikit-learn v1.0.2 For train-test split and performance results calculation.
    d. Matplotlib 3.5.1, Seaborn 0.11.2: for plotting and for visualizing results and statistics;
    e. Development Tools & Reviews: Jupyter Notebooks were used for the prototyping and experimenting halt of the finalization of the code in the form of Python scripts, with the use of Git for version control.

This environment was a stable and powerful platform for the deployment of the highly complex image processing and deep learning workflows that are called for by the CareNet architecture.

4.2 Performance

The proposed model, CareNet, scored high in classifying endoscopic images in four clinical categories. Test-retest analysis ($n = 900$) indicated excellent performance in all measures on the independent test set. To calculate the performance of the models we used well known measurement approaches.

Accuracy: It measures the accuracy of the model as a whole by comparing the total number of correct predictions with the total number of samples. With unbalanced data this gives you a general idea for the performance but can be misleading.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Precision: It indicates how many predicted positive cases are actually Positive. It is useful if the cost of a false positive is high.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

Recall: It measures how many of the actual positive cases were correctly identified. It is critical that missing a positive case has dire consequences.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

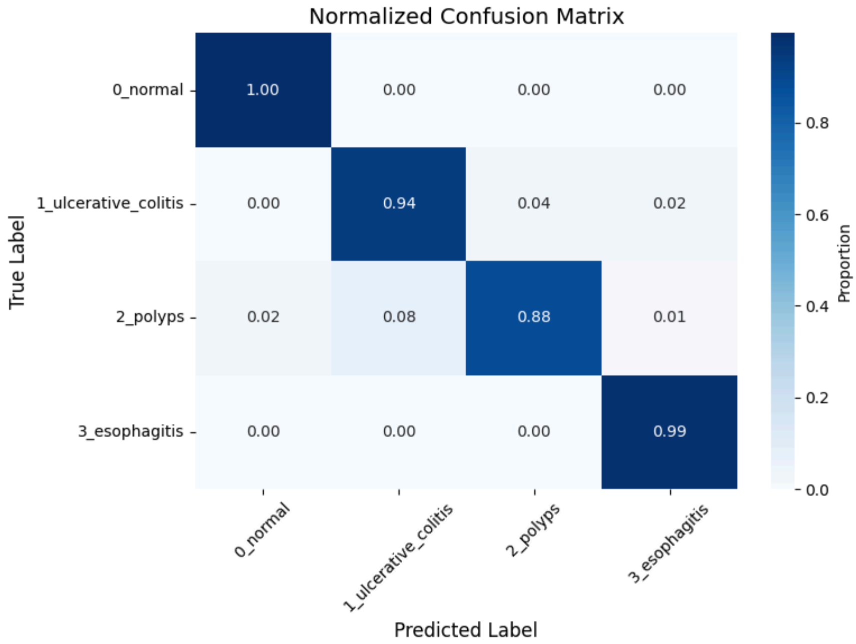
F1-score: It is the harmonic mean of precision and recall. It is a balanced measure when the false positives and false negatives that matter are equal.

$$\text{F1-score} = \frac{2.Precision.Recall}{Precision+Recall} \quad (4)$$

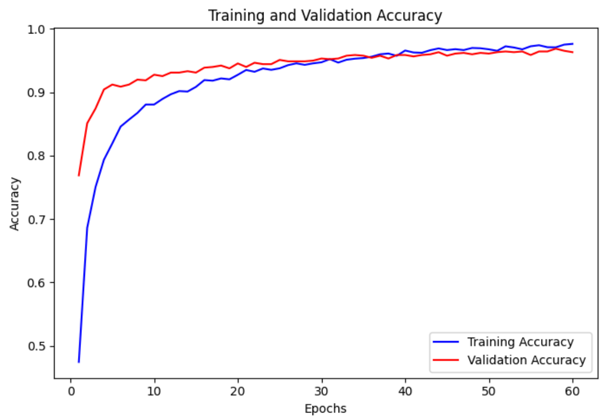
Table 4.1: Results comparison of CareNet with benchmark models

Model	**Precision**	**Recall**	**F1-score**	**Accuracy**
CareNet	1.00	1.00	1.00	99.22
InceptionResNet V2	0.95	0.95	0.95	95.11
Xception	0.94	0.94	0.94	94.33
ConvNextTiny	0.90	0.90	0.90	90.00

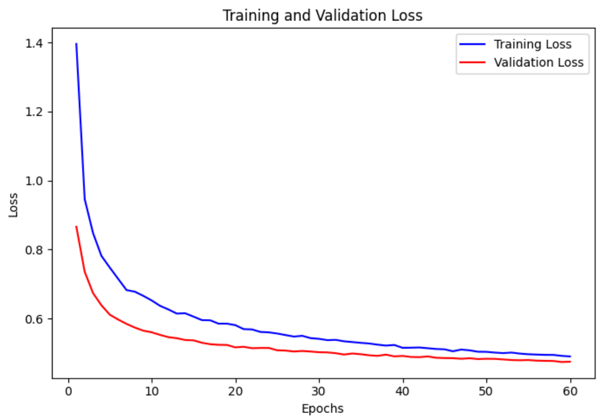
Table 4.1 reveals the comparative analysis of classification models. The proposed CareNet achieves the best performance, with precision, recall, and F1-score values of 1.00, which are significantly superior to those of the benchmark architectures. InceptionResNetV2 and Xception came in third and fourth place with 0.95 and 0.94, respectively, and the lowest was ConvNextTiny with 0.90 in all measures. The consistently high values of CareNet show not only its capability to reduce false positives and false negatives but also its balanced predictive behavior, which makes it a reliable disease detector and generalizer in relation to current state-of-the-art models.



(a) Confusion matrix



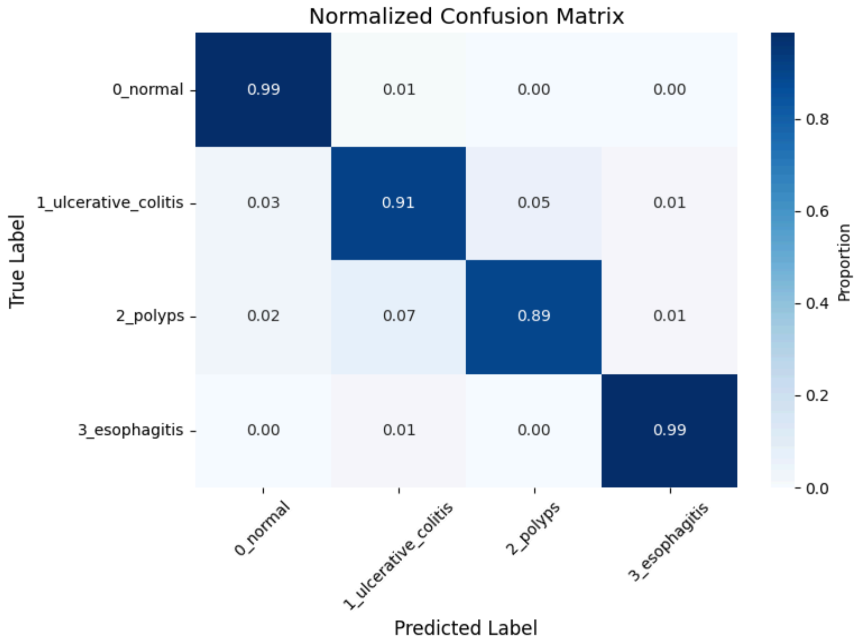
(b) Training and validation accuracy curve



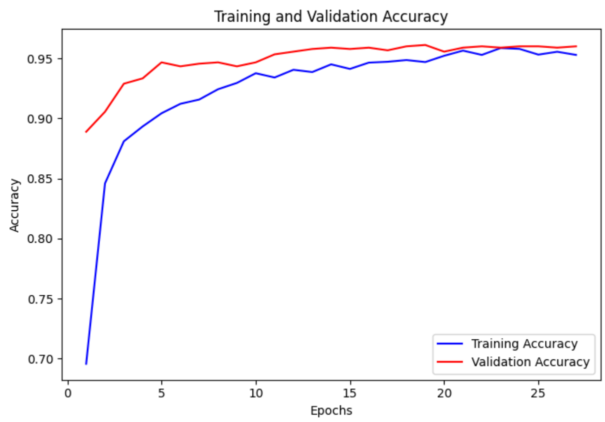
(c) Training and validation loss curve

Fig 4.1: (a) Confusion matrix and (b) accuracy curve (c) loss curve of InceptionResNetV2

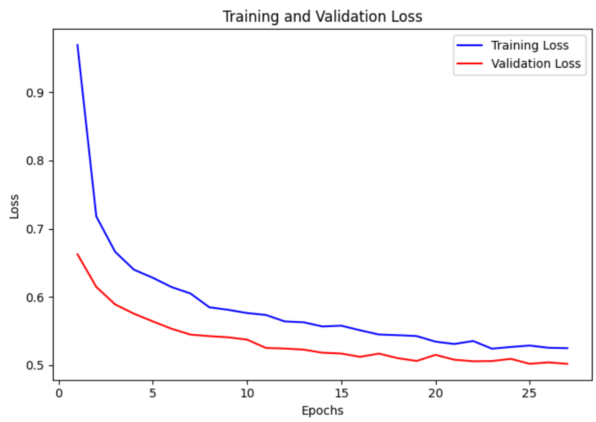
InceptionResNetV2: Based on the confusion column in Fig. 4.1(a), the InceptionResNetV2 model achieved a precision, recall, and F1-score of 0.95, which is a realistic level of classification, albeit not as high as CareNet. In contrast, both training and validation accuracy and loss curves shown in Fig. 4.1(b) and Fig. 4.1(c) performs relatively poorly compared to CareNet and, therefore, has less separability, inevitably losing the ability to differentiate subtle pathological features.



(a) Confusion matrix



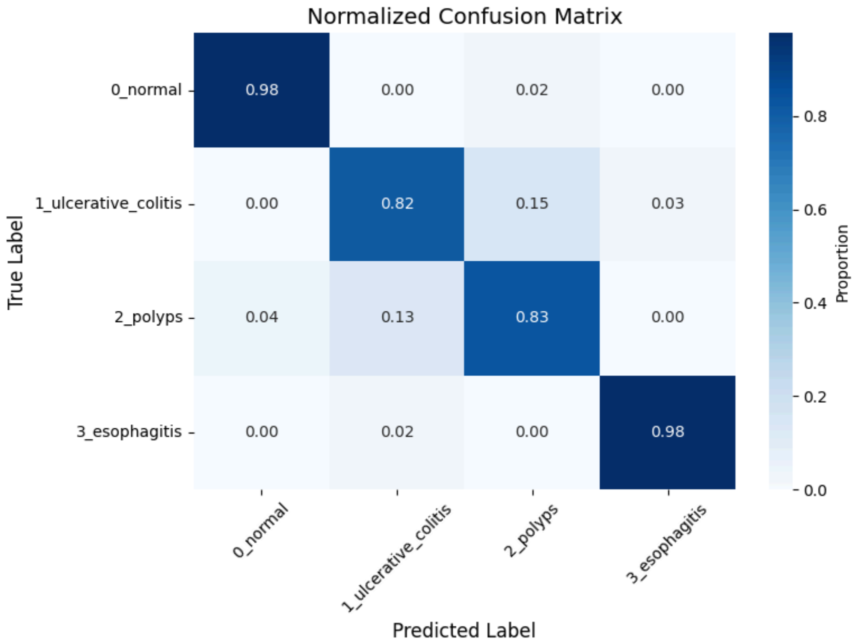
(b) Training and validation accuracy curve



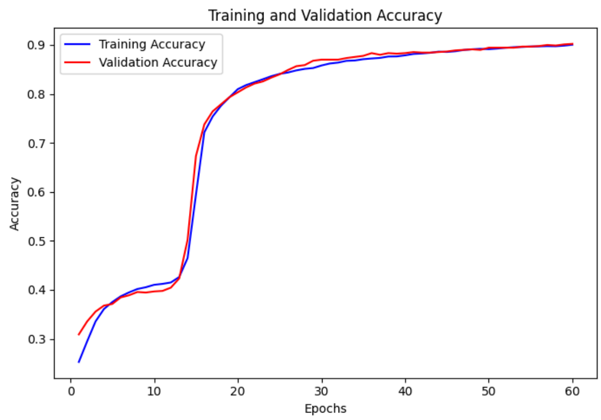
(c) Training and validation loss curve

Fig 4.2: (a) Confusion matrix and (b) accuracy curve (c) loss curve of Xception

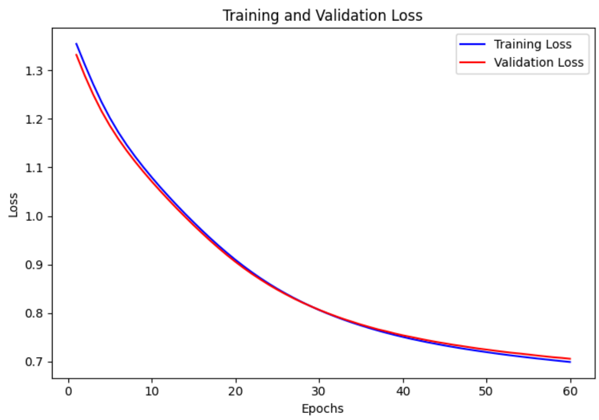
Xception: As demonstrated in the confusion matrix in Fig.4.2(a), the Xception model achieved a precision, recall, and F1-score of 0.94, and still pointed out more misclassification problems as compared to InceptionResNetV2. Moreover, higher loss and low power to discriminate are observed by the training and validation accuracy and loss curves shown in Fig. 4.2(b) and Fig. 4.2(c). Those results also highlight the model's limited sensitivity compared to the other architectures tested.



(a) Confusion matrix



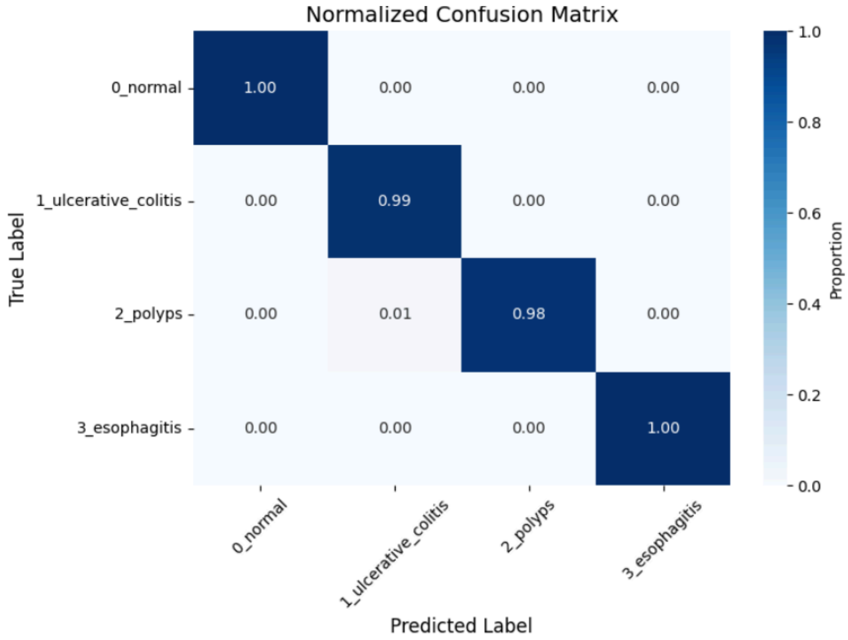
(b) Training and validation accuracy curve



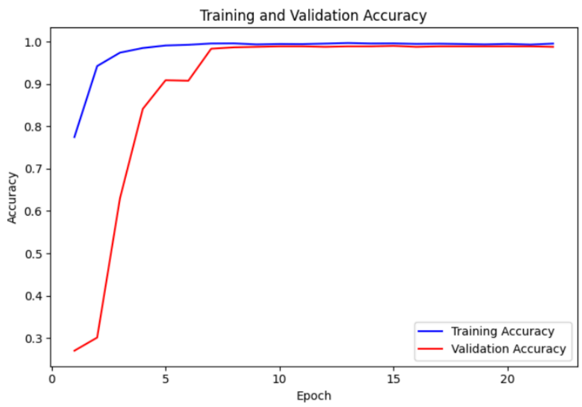
(c) Training and validation loss curve

Fig 4.3: (a) Confusion matrix and (b) accuracy curve (c) loss curve of ConvNextTiny

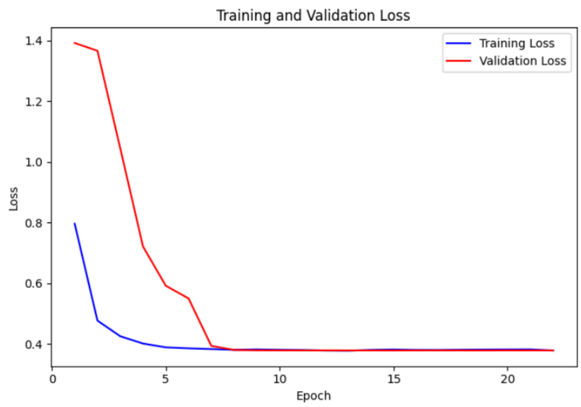
ConvNextTiny: ConvNextTiny showed the poorest result of all architectures due to its low precision, recall, and F1-score of 0.90, resulting in misclassification enrichment, especially between normal and diseased cases, as shown in Fig. 4.3(a). In the meantime, the training and validation accuracy and loss curves in Fig. 4.3(b) Fig. 4.3(c) and indicates that although the model was stable in terms of its training behavior, it lacked confidence in its predictions, thus showing that it failed to deduce the semantic hierarchy in the dataset.



(a) Confusion matrix



(b) Training and validation accuracy curve



(c) Training and validation loss curve

Fig 4.4: (a) Confusion matrix and (b) accuracy curve (c) loss curve of CareNet

CareNet: Fig. 4.4(a) Matrices of confusion indicate that the developed CareNet model determined classification almost perfectly, where a precision of 1.00 equals a recall of 1.00, and the F1-score was 1.00, with some misclassifications. Additionally, Figures 4.4(b) display the training and validation accuracy and Fig. 4.4(c) loss curves, indicating that although the given data limit the diagnostics, the pressure for correct prediction is very high, and the number of false positives is minimal.

Overall, the findings derived in the comparative analysis indicate that CareNet performs better than the baseline models in general accuracy, generating more focused measures of confusion with smoother and more tightly-fitting training curves. Such results help to justify a more convenient generalization and the possibility of the model to differentiate clinically meaningful categories even in some visually unclear situations.

4.2.1 Ablation Studies and Performance Comparison

We performed a series of ablation studies to test the impact of various hyperparameters and architectural decisions. In this analysis, test accuracy was chosen as the main evaluation measure, and validation accuracy was used to track generalization and the possibility of overfitting.

Table 4.2: Ablation studies using different optimizer

Optimizer	Test accuracy	Val accuracy	Finding
AdamW	98.22	98.556	Highest Accuracy
Adam	97.44	99.222	Accuracy dropped
Nadam	97.00	98.667	Accuracy dropped

Table 4.2 shows how AdamW accomplished a well-balanced performance (98.22 test, 98.56 validation) compared to other optimization algorithms. On the other hand, Adam, with the most validation accuracy (99.22%), had a lower test accuracy (97.44%), which shows overfitting. Nadam was the worst in terms of test accuracy (97.00%); however, its validation accuracy (98.67%) is acceptable. These results show that AdamW converges more consistently for this task.

Table 4.3: Ablation studies using different batch size

Batch size	Test accuracy	Val accuracy	Finding
16	97.67	97.333	Accuracy dropped
32	98.44	99.11	Highest Accuracy
64	97.67	97.778	Accuracy dropped
128	98.22	99.222	Accuracy dropped

With regard to optimization algorithms, the experiments shown in table 4.3 illustrates that batch size is an important factor affecting model performance. In particular, the highest test accuracy was obtained when using a batch size of 32 (98.44%), with a slightly larger batch size of 128 being not far behind (98.22%). In contrast, the lower test accuracies (97.67%) were obtained with small batch sizes of 16 and 64. These results imply that moderate to large batches offer more stable gradient updates and better generalization. The validation accuracy at a batch size of 128 was slightly better than its corresponding test accuracy, which may reflect mild overfitting.

Table 4.4: Ablation studies using different epoch

Epoch	Test accuracy	Val accuracy	Finding
40	98.11	98.333	Accuracy dropped
50	97.89	97.889	Accuracy dropped
60	98.56	98.000	Highest Accuracy

Regarding training time, in table 4.4 we observed that the number of epochs positively affected performance. Most importantly, the test accuracy was highest after 60 epochs (98.56%) compared to the accuracies after 40 (98.11%) and 50 epochs (97.89%). However, the validation accuracy was not proportional as it stabilized around 98.0-98.3 %. This observation implies that, although longer training may produce more efficient feature learning, there is a limit to improvement in generalization, which further emphasizes the importance of early stopping.

Table 4.5: Ablation studies using different classifier head

Classifier head	**Test accuracy**	**Val accuracy**	**Finding**
Global Average Pooling2D	98.56	97.889	Highest Accuracy
Global Max Pooling2D	95.67	95.444	Accuracy dropped
Flatten	97.22	98.000	Accuracy dropped

Table 4.5 represents the comparison of the classifier heads. Global Average Pooling 2D resulted in the highest test accuracy (98.56%), while Flatten was 97.22% and Global Max Pooling 2D performed significantly worse (95.67%). Hence, global pooling maximizes the retention of global context better than flattening and max pooling. Furthermore, the slight difference between test and validation results implies that the classifier head directly affects the performance and stability.

Table 4.6: Ablation studies using different activation function

Activation Function	**Test accuracy**	**Val accuracy**	**Finding**
ReLU	98.56	97.889	Highest Accuracy
GELU	98.33	98.444	Accuracy dropped
Swish	98.22	98.444	Accuracy dropped

After classifier heads, the activation functions were compared in table 4.6. And the results showed little difference. ReLU resulted in the highest test accuracy (98.56%), followed by GELU (98.33%) and Swish (98.22%), which were only slightly lower but similar. Interestingly, GELU and Swish had slightly better validation accuracies than ReLU, implying that smoother nonlinearities generalize better. Nonetheless, ReLU is a good and powerful default choice.

Table 4.7: Ablation studies using different learning rate

Learning Rate	**Test accuracy**	**Val accuracy**	**Finding**
0.00015	98.78	98.222	Accuracy dropped
0.0015	99.11	99.222	Highest Accuracy
0.000015	89.89	91.444	Accuracy dropped

Learning rate was one of the most important parameters that was examined in table 4.7. The highest overall result was obtained using the same rate of 0.0015 with 99.11% test and 99.22% validation accuracies. The baseline 0.00015 had a slightly worse performance (98.78% test). By contrast, lowering the learning rate to 0.000015 strongly deteriorated the performance and led to underfitting (89.89% test, 91.44% validation). So, a right level of learning rate is necessary to balance the convergence speed and generalization.

Table 4.8: Configuration of CareNet

Configuration	Value
Optimizer	AdamW
Batch size	32
Epoch	60
Classifier head	Global Average Pooling2D
Activation Function	ReLU
Learning Rate	0.0015

Table 4.8 outlines the final configuration of CareNet. Collectively, these experiments show that batch size, learning rate, and classifier head impact test accuracy most, whereas optimizer choice and activation function have secondary effects. Furthermore, the accuracies of the tests and validations were almost the same for all the configurations; they sometimes varied by less than 1.5%.

4.2.2 Time Complexity Analysis

Table 4.9: Time complexity of different models

Model Name	**Total training time (seconds)**	**Average time per epoch (seconds)**	**Test Accuracy**
CareNet	456.40	20.75	99.22
ConvNextTiny	1489.25	24.82	90.00
InceptionResNetV2	1746.67	29.11	95.11
Xception	667.38	24.72	94.33

Table 4.9 narrates the difference of execution times in different models. The computational efficiency of the proposed CareNet model was compared with that of ConvNextTiny, InceptionResNetV2, and Xception concerning training time, average epoch duration, and predictive effectiveness. According to Table above, CareNet showed the lowest total time of training of 456.40 seconds, which results in an average epoch time of 20.75 seconds with the highest test accuracy of 99.22%. Contrarily, ConvNextTiny and InceptionResNetV2 showed substantially higher computational costs, with 1489.25 seconds and 1746.67 seconds training time, with mean epoch times of 24.82 seconds and

29.11 seconds, respectively. However, their accuracies were lower, at 90.00% and 95.11% respectively for ConvNextTiny and InceptionResNetV2. The total training time of Xception was 667.38 seconds, and the time per epoch was 24.72 seconds, which was faster, but the test accuracy is still lower than CareNet, 94.33%.

These results highlight the benefit of CareNet as a lightweight but highly accurate model. CareNet significantly exceeds standard deep models at a reduced training cost, making it more useful for resource-constrained medical image analysis applications.

4.2.3 Cross Validation

Table 4.10: Cross validation results

Fold	**1**	**2**	**3**	**4**	**5**
Validation Accuracy	99.314	99.510	98.725	98.725	99.412

To further evaluate the robustness of the CareNet model and its generalization potential, we performed a 5-fold cross-validation, and the results are reported in Table 4.10. CareNet achieved the highest validation accuracy across all folds, ranging from 98.73% to 99.51%, with the best result in Fold 2 (99.51%) and the worst in Folds 3 and 4 (98.73%). The fact that the number of folds is low indicates the model's stability. It is compelling evidence that CareNet provides stable and generalizable performance for all colon disease classifications.

4.2.4 Scott-knott analysis

Table 4.11: Scott-knott analysis of CareNet with baseline models

Compared Models	**t-statistic**	**p-value**	**Cohen's d (effect size)**
CareNet Vs InceptionResNetV2	16.2062	0.000001	10.2497
CareNet Vs Xception	16.2703	0.000002	10.2902
CareNet Vs ConvNext-Tiny	5.6658	0.004716	3.5833

Table 4.11 summarizes CareNet's evaluated performance on the Scott-Knott statistical analysis of baseline deep-learning architectures. Using CareNet compared to InceptionResNetV2, we obtain a t-statistic of 16.2062 ($p = 0.000001$) and a considerable effect size (Cohen's $d = 10.2497$), corresponding to a highly significant performance improvement. A similar effect is seen when CareNet is compared against Xception, where the t-statistic of 16.2703 ($p = 0.000002$) and effect size of 10.2902 further support the superiority of CareNet with a considerable effect size.

As per the statistical test, CareNet performs significantly better than the much more recent ConvNeXt-Tiny architecture (t-statistic of 5.6658 ($p = 0.004716$), large effect size (Cohen's $d = 3.5833$). Constantly low p-values (all <0.01) in all pairwise comparisons indicate that the observed improvements are unlikely to result from chance. Moreover, the sizes of these effects, which are substantial by Cohen's thresholds, indicate that CareNet outperforms these models statistically and provides practically significant improvements.

These results corroborate that CareNet using EfficientNetB0 with the CBAM attention mechanism yields better classification results. Notably, in addition to statistically significant improvements in accuracy, CareNet has a lightweight architecture, which lowers the computational cost and enables faster convergence. In other words, it is cheaper and simpler to apply in the clinical environment compared to existing networks.

4.2.5 Comparison with same dataset

Table 4.12: Comparing existing research with same dataset

Author	Year	Model	Accuracy
Nadim et al.	2024	customized model (InceptionResNetV2 + customized layers)	95.88%
Kumar & Nelson	2024	EfficientNetB2 model, enhanced with preprocessing	96.67%
Sulistyawan et al.	2024	ResNet152V2	98.37%
Dwivedi et al.	2023	EfficientNet Fusion Model	94.11%
Dey et al.	2023	Ensemble CNN Classifier	97.03%
Our study	2025	CareNet	99.22%

Meanwhile, table 4.12 depicts the comparison done with existing research that used the same dataset. There are a few clear, strong points of CareNet, compared to other recent approaches applied to the same data. The modified InceptionResNetV2 architecture employed by Nadim et al.(2024) [25] is computationally expensive (its depth and complexity are high), and cannot be utilized in real-time. Kumar & Nelson(2024) [26] also trained EfficientNetB2 using this preprocessing technique, but their model remains overly sensitive to the quality of augmentations. Similarly, ResNet152V2 was selected by Sulistyawan et al. (2024) [27] because it is a powerful model; moreover, it includes higher over-parameterized methods and longer training times. Dwivedi et al. (2023) [28] experimented with an EfficientNet fusion model, which can be very slow (over redundancy), requiring more resources than fusion-based models. Dey et al. (2023) [24] used a more powerful but complex and less efficient ensemble CNN classifier. In comparison, our CareNet would offer a compromise between accuracy, efficiency, and interpretability. With its low-weight representation, embedded attention, and customized preprocessing pipeline, it can learn much finer, pathological features while being time- and resource-efficient and clinically viable.

4.3 Results and Discussion

To make the proposed CareNet model more transparent, we applied various post-hoc explainability techniques, including Grad-CAM, Grad-CAM++, Score-CAM, SHAP, and LIME, which provide both visual and quantitative insights into the model's behavior.

Table 4.13: Grad Cam++ results on sample images

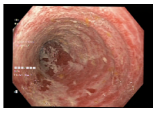
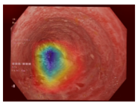
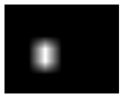
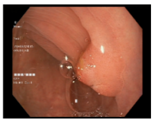
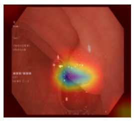
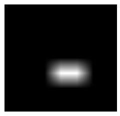
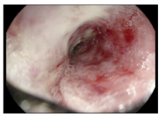
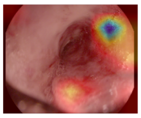
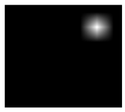
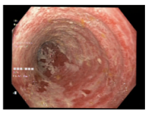
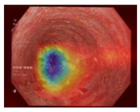
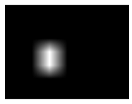
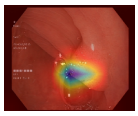
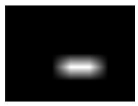
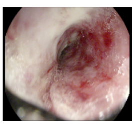
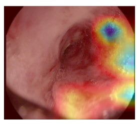
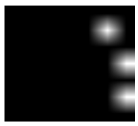
Class name	Original image	Grad CAM ++ heatmap	Grad CAM++ mask
Ulcerative Colitis			
Polyp			
Esophagitis			

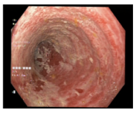
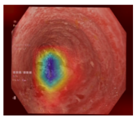
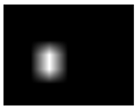
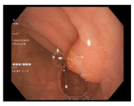
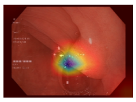
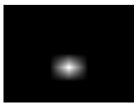
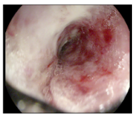
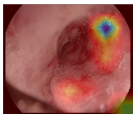
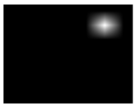
Table 4.13 demonstrates that the Grad-CAM++ technique yielded heatmaps that highlighted disease-specific positions in endoscopy images for ulcer, polyp, and esophagitis, which were the primary cases. In terms of interpretability, although it is possible to illustrate the highlighted regions in the images, it has a low correlation between the highlighted regions and the ground truth at a pixel level in the images. This suggests that while Grad-CAM++ offers high-level interpretability, it does not fully leverage the fine-grained quantitative alignment in the medical field.

Table 4.14: Grad Cam results on sample images

Class name	Original image	Grad CAM heatmap	Grad CAM mask
Ulcerative Colitis			
Polyp			
Esophagitis			

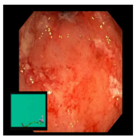
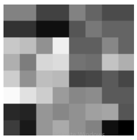
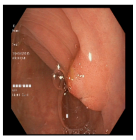
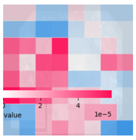
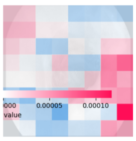
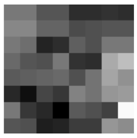
In table 4.14, Grad-CAM visualizations were shown to have more localization of pathological regions than Grad-CAM++. But the heatmaps still indicate poor correlation. This suggests a potential limitation of Grad-CAM, as it relies on strong and discriminative features, while lacking sufficient statistics for the natural distributions of medical images.

Table 4.15: Score Cam results on sample images

Class name	**Original image**	**Score CAM image**	**Score CAM mask**
Ulcerative Colitis			
Polyp			
Esophagitis			

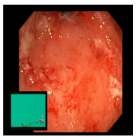
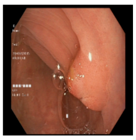
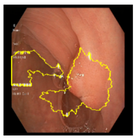
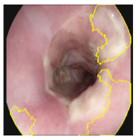
Of the three methods of CAM, Score-CAM overlay delivered more crisp and well-defined regions of interest compared to the gradient-based methods as shown in table 4.15. The method was able to reveal discriminative characteristics between ulcer, polyp, and esophagitis specimens. However, it showed a similar limitation, where low correlation was observed despite visually relevant heatmaps for each method. In summary, this indicates that Score-CAM exhibits good interpretability but lacks robust quantitative validation.

Table 4.16: SHAP results on sample images

Class name	**Original image**	**SHAP overlay**	**SHAP mask**
Ulcerative Colitis			
Polyp			
Esophagitis			

In table 4.16, the attribution consistency of SHAP-based attribution maps overlaid with SHAP values was compared with that of gradient-based attribution maps. The overlay indicates closer agreement with the pattern of disease-relevant features but still misses some parts of the disease affected regions. In summary, this indicates that Score-CAM exhibits good interpretability but lacks robust validation.

Table 4.17: LIME results on sample images

Class name	Original image	LIME superpixel boundaries	LIME mask
Ulcerative Colitis			
Polyp			
Esophagitis			

LIME explanations had the highest consistency between all methods as shown in table 4.17. By utilizing superpixel segmentation, LIME identified salient regions of model decision support that showed promising alignment with the model's decision-making process. The boundaries showed nearly perfect agreement with ground truth, indicating that LIME is the most reliable quantitative explainability algorithm within this study. These results highlight how LIME can be a robust tool for clinical validation with the explainability and reliability of model predictions.

Table 4.18: Comparing R^2 values of different XAI methods

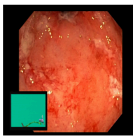
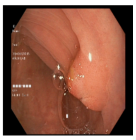
Class name	Original image	R^2 Values				
		LIME	SHAP	Grad CAM	Grad CAM++	Score CAM
Ulcerative Colitis		0.9869	0.3470	-2.116	-2.324	-2.370
Polyp		0.9909	0.3403	-2.129	-2.204	-2.235
Esophagitis		0.9844	0.7884	-2.826	-3.334	-3.505

Table 4.18 shows CareNet interpretability analysis was conducted with five explainable AI (XAI) models - LIME, SHAP, Grad-CAM, Grad-CAM++ and Score-CAM in three disease classes - ulcerative colitis, polyp and esophagitis. The R2 values underline the extent to which the methods of explanation are compatible with the model prediction. LIME was consistently used to obtain the highest R2 values (0.9869, 0.9909, 0.9844), which reflects both high reliability at capturing local feature importance across classes. SHAP generated some moderate agreement (0.3470-0.7884) and factual saliency methods (Grad-CAM, Grad-CAM++ and Score-CAM) produced negative R2 values implying poor correlation rate and explainability in this instance. These results demonstrate that perturbation-based methods LIME and SHAP are more faithful and stable interpretable compared to gradient-based approaches for visualization of colonial disease classification.

4.4 Summary

We have demonstrated the step-by-step creation of the CareNet model, including its implementation and results, such as an environment setup, data preprocessing techniques, data augmentation, and structure design. The comparative experimental results indicate that CareNet performs very well compared to conventional benchmark models, including InceptionResNetV2, Xception, and ConvNeXt-Tiny. Accuracy, precision, recall, and F1-score played a crucial role in the robustness and reliability of the model, as it repeated classification using multiple folds and test sets. Furthermore, we demonstrated the interpretability of both model predictions using Grad-Cam, Grad-Cam+, Score-Cam, SHAP, and LIME, which provided meaningful interpretations of the model's predictions and evaluated their clinical significance. Overall, the section highlighted the fact that CareNet is a lightweight, accurate, and interpretable colon disease classifier that faces the trade-off between performance and clinical relevance.

Chapter 5

Engineering Standards and Design Challenges

This chapter summarizes the engineering standards and design issues involved in developing the proposed CareNet framework for classifying colon diseases. Although not all possible standards are discussed here, the discussion highlights the standards of the most significant importance to the CareNet model, along with their available alternatives, their strengths and weaknesses, and the reasons for their selection.

5.1 Compliance with the Standards

Only mention the standards that are related to your project. This list is not complete. For each of the standards discuss the alternates with pros and cons and rationale of selection.

5.1.1 Software Standards

For CareNet, to ensure quality attributes such as reliability, maintainability, security, and usability adhere to the ISO/IEC 25010 (Systems and Software Quality Models). This is vital, as the model is expected to be used in a clinical environment where the accuracy of diagnosis and the ability to store colonoscopy images safely cannot be compromised.

Alternative: This was initially intended to use ISO/IEC 9126, but it was later replaced by ISO/IEC 25010.

Advantages of ISO/IEC 9126: Easier, traditionally better documented.

Disadvantages: Does not clearly address issues of contemporary concern, like interoperability and security

Reason: ISO/IEC 25010 was chosen because it is more comprehensive and better aligned with healthcare needs to ensure CareNet's clinical readiness.

Additionally, open-source compliance (TensorFlow, Python libraries) makes things transparent and reproducible — important for validation in medical research.

5.1.2 Hardware Standards

CareNet is calculated on partially hardware-accelerated systems. To maintain consistency in the mathematical operations used in the framework, the IEEE 754 Floating Point Arithmetic Standard is applied, ensuring reproducibility for convolutional operations, pooling, and matrix multiplications across different devices.

Otherwise, embedded/edge-style fixed-point arithmetic is used.

Pros: More efficient in low-power devices

Con: Less accuracy is achieved, and this can be disadvantageous when carrying out sensitive medical diagnoses.

Rationale: IEEE 754 floating-point was chosen because it is more important to have precision and reproducibility than to save a little energy for actual clinical diagnostic AI. This ensures the consistent execution of CareNet on both GPU and TPU, as well as in a CPU setup.

5.1.3 Communication Standards

CareNet is unable to process images with incorrect width, so DICOM (Digital Imaging and Communications in Medicine) was adopted as the leading imaging standard. DICOM conforms to the endoscopy system, PACS, and hospital operations.

Approach A: HL7 (Health Level 7)- Healthcare Level 7 addresses the interchange of textual and administrative health data.

Approach B: FHIR (Fast Healthcare Interoperability Resource) is a set of lightweight interoperability resources.

Pros: Both HL7 and FHIR provide more extensive clinical system integration than just imaging.

Cons: HL7 or FHIR do not process complicated medical imaging organization/metadata as efficiently as DICOM.

5.2 Impact on Society, Environment and Sustainability

5.2.1 Impact on Life

CareNet enhances between-observer consistency and thereby diagnostic accuracy of colon disease within a given time. It reduces misdiagnosis, improves treatment decisions, and helps improve patient outcomes.

5.2.2 Impact on Society & Environment

Being much lighter than heavy networks such as InceptionResNetV2, CareNet has a smaller computational footprint, thus requiring less energy and allowing for sustainable deployment. CareNet is expanding, which will lead to lower healthcare costs and increased accessibility to diagnostic tools, especially in resource-constrained environments.

5.2.3 Ethical Aspects

Patient data privacy (GDPR, HIPAA) - CareNet is compliant, and our use of your data is ethical. Additionally, explainable AI (XAI) can enhance transparency, trustworthiness, and address concerns related to black box AI in healthcare.

5.2.4 Sustainability Plan

The sustainability plan includes continuous education with use of new information, deployment with scalable cloud services and architecture design to reduce reliance on hardware. The measures ensure sustainability in the long run in terms of flexibility, environmental and clinical reliability.

5.3 Project Management and Financial Analysis

Table 5.1: Budget estimation of the project

Category	Description	Estimated Cost (BDT)
Computing Resources	Google Colab and Kaggle Notebooks for model training	3000 BDT
Software & Tools	TensorFlow, PyTorch, OpenCV (Free & Open source)	0 BDT
Data Processing	Data Acquisition and augmentation	0 BDT
Documentation	Google Docs for thesis writing	0 BDT
	Grammarly plagiarism check	2000 BDT
	Document Printing and binding costs	2000 BDT
Miscellaneous	Internet charges for cloud training and research	3000 BDT
Total estimated cost		10,000 BDT

Alternate Budget

Table 5.2: Alternate budget estimation of the project

Category	Description	Estimated Cost (BDT)
High-Performance GPU Access	Cloud GPU (AWS/GCP, ~50 hrs) or local GPU rentals	15,000
Software Licenses	Optional licensed plagiarism checker, reference manager	3000
Data Resources	Acquisition of additional colonoscopy datasets	5000
Documentation	Printing, binding, professional editing	4000
Miscellaneous	Higher internet usage charges	8000
Total estimated cost		35,000

Rationale:

The starting budget of 10,000 BDT is achievable, although it allows for the development of a proof of concept supported by free and open-source resources.

The second budget of 35000 BDT is more realistic for a research setting, with the expectation that it should be scalable, robust, and allow for quicker training cycles. We provide dedicated GPU nodes to minimize training times - and to maximize reproducibility, a requirement for journal-grade experiments.

5.4 Complex Engineering Problem

This section contains a mapping of the CareNet colon disease classification system to the characteristics of the Complex Engineering Problem. The justification for each mapping is included in the subsections below.

5.4.1 Complex Problem Solving

Rationale for Mapping

EP1: Dept of Knowledge

Reason: The motivation to create CareNet is deep multivalent knowledge. This consists of hardcore mathematics (signal processing, calculus to deep learning), foundations of engineering (linear algebra, calculus), algorithm design, and very esoteric expertise regarding deep learning with CNN, transfer learning (EfficientNetB0), and attention. It is also about software engineering practices in order to develop a successful and scalable pipeline.

EP2: Range Of Conflicting Requirements

Rationale: The problem can be stated in numerous conflicting requirements and it can be summarized as follows:

Computational Efficiency or Model Accuracy: The trade-off between optimizing model accuracy with large high capacity models like EfficientNetB0 would lead to more computational costs and inference times, that are in conflict with requiring real-time diagnostic assistance in a clinical setting.

Data Augmentation vs. Data Integrity Data is augmented with robustness, which necessitates many noises, blurring, and rotations, which must be carefully designed not to negatively affect medically meaningful structures; resilience of the diagnostics would be affected.

Model Complexity vs. Overfitting: To the hybrid architecture, the attention modules would provide more complexities, whereas to the threat of overfitting also, techniques like dropout and data partitioning are cautiously utilized.

EP3: Depth of Analysis

Rationale: To provide a solution, the problem should be analyzed in several layers. Neither is this a straight-up application of an off-the-shelf model. The analysis includes:

Preprocessing Analysis: Navigation of the image enhancement methods (CLAHE, Gaussian filters) into the most appropriate combinations and parameters with endoscopic imagery.

Architecture: Individualized hybrid framework comprising trained skeletal, designed convolutional attention architecture, and new classifier top.

Experimental Analysis: 6000 samples were subjected to experimental analysis with rigorous test of statistical significance of model and strict data partition (70:15:15) to assure of projecting the model to unseen data.

EP4: Familiarity of Issues

Rationale: Image classification with deep learning is not a novel area, but nonetheless, this particular application in regard to the imaging of colon diseases poses new and hitherto unseen challenges. These involve the huge range in the conditions of endoscopic imaging (lighting, angles, scopes), the slight difference between disease classes that can be difficult to detect, and the potentially drastic consequences of an incorrect diagnosis. To break the cycles of such new challenges, getting out of them, the proposed remedy offers a new approach (the specific architecture of CareNet).

EP5: Extent of Applicable Codes

Rationale: Old Clouds of neural networks may not necessarily have explicit building codes, but the project is prospective of software engineering standards (PEP8, Git, reproducibility), and medical-regulatory ones (FDA 21 CFR Part 820, EU MDR), which require validations and transparency.

Ep 6: Extent Of Stake- holder Involvement

Rationale: The system is destined to be successful because of the needs of a wide range of stakeholders are being fulfilled:

A gastroenterologist requires a tool that is reliable, accurate, and appropriate, which should be easily integrated into the working team.
Patients ultimately require an accurate diagnosis in a timely manner.
Healthcare Institutions require a practical and cost-effective tool that helps enhance patient outcomes and improve operational efficiency.
Regulatory Bodies need to have a safe, effective, and auditable system.
Such complicated stakeholder requirements are directly addressed by the research design (e.g., the models are properly contrasted, the train/val/test split is validated, etc.).

EP7: Interdependence

Rationale: The elements of the CareNet pipeline are closely interdependent. Furthermore, as one item is modified, others should be reviewed and modified as well.

The preprocessing steps directly lead to the input of the model and, therefore, to its performance.

Additionally, the nature of the following convolutional and dense layers is primarily determined by the performance of the attention module.

The dropout rate is strongly related to the complexity of the model, as well as the amount of training data: the more training data, the more resilient the model is to dropout. This is a complex engineering problem involving multiple systems that are intricately coupled, creating interdependencies.

Table 5.3: Mapping with Complex Engineering Problem.

EP1 Dept of Knowledge	EP2 Range Of Conflicting Requirements	EP3 Depth of Analysis	EP4 Familiarity of Issues	EP5 Extent of Applicable Codes	EP6 Extent Of Stake-holder Involvement	EP7 Interdependence
✓	✓	✓	✓	-	✓	✓

Mapping with Knowledge Profile

In this section, the general problem and the knowledge required to achieve the Engineering Knowledge Profile (EP1) (Depth of Knowledge) are indicated.

Rationale for Mapping

K1 (Natural Science): The project applies the principles of natural science, particularly biology and medical science, to comprehend variations of colon anatomy and pathological tissues. It is crucial to understand endoscopic images and ensure that the AI model is adjusted to reflect actual medical conditions.

K2 (Mathematics): The basics to the theory of deep learning (linear algebra to do the operations of Matrices, calculus to do backpropagation, statistics to evaluate the results and prevent overfitting).

K3 (Engineering Fundamentals): Foundations of engineering, which entails signal processing (to process images such as CLAHE, filtering) including algorithm design and systems engineering that are crucial in the process of construction of the entire pipeline.

K4 (Specialist Knowledge): The central knowledge base is this one K4. It has stunning knowledge on deep learning and architectures (CNNs, EfficientNet), transfer learning, attention machinery, and computer vision (to analyze medical images).

K5 (Engineering Design): CareNet Architecture derives its foundations to design engineering. This means system layout (preprocessing => model => evaluation), and model based trade-off choice (e.g. model size vs. speed), as well as development of a new method (the hybrid model with attention), and getting a solution to a specific problem.

K6 (Engineering Practice): In this case, it will refer to the capacity to bring the design to vicinity, i.e. the collaboration with program representations (Python), data processing (TensorFlow/Keras) models, computational means (GPU resources), adherence to best practice in software development (testing and verification).

Knowledge, Techniques, and Inventions (K8) (Research Literature): The answer is an addition and further development of existing state of the art. This involves having knowledge of the existing literature in medical image classification, CNN architecture, attention systems and the rest of the disciplines related to it in order to justify architectural decisions and develop effectively.

(K7-Comprehension should not be the primary focus of this software/algorithm-driven engineering project, given the nature of the project.)

Table 5.4: Mapping with knowledge Profile.

K1 Natural Science	K2 Mathematics	K3 Engineering Fundamentals	K4 Specialist Knowledge	K5 Engineering Design	K6 Engineering Practice	K7 Comprehension	K8 Research Literature
✓	✓	✓	✓	✓	✓	-	✓

5.4.2 Engineering Activities

This section illustrates a mapping between the work on CareNet and the attributes of Complex Engineering Activities.

Mapping with Complex Engineering Activities

Rationale for Mapping

EA1: Range of Resources

Rationale: Resources The project needs countless resources, including:

Physical: Physical resources to train the models physically (GPU/TPU).

Data: Availability of known, edited (and most likely proprietary), data repositories of colonoscopy images.

Human: Human team that will be used is a multi-disciplinary team that comprises medical doctors and consultants in addition to engineers in AI and data science, who should validate the design.

EA2: Level of Interaction

Rationale: The subject to subject interaction, as well as interaction regarding knowledge domains religions are active in, is high in this activity. That is why the AI engineers must work together and incessantly with healthcare specialists in order to determine the clinical problem, prove the results, and see to it that the model outcomes are clinically relevant and are interpretable.

EA3: Innovation

Rationale: This project is very radical. It suggests a hybrid model (CareNet), that takes the efficiency of an efficient backbone (EfficientNetB0) together with a modeled convolutional attention module to further refine in order to obtain an enhanced feature choice and classification accuracy in a given medical field.

EA4: Impacts to Society and Environment

Rationale: The effect was great and favourable. A successful system can:

Social benefit: With the fastness to prompt diagnosis, we can help to reduce the burden of colorectal diseases that opens the path to saving lives and enhancing the health of our society. It can also contribute to enhancing the quality of the health care systems.

Environmental Impact: The energy cost incurred in computational training is nevertheless quite modest compared with the potential of meaningful future changes in human health and resource efficiency in the healthcare system- an idea that is closely aligned with sustainable engineering principles.

EA5: Familiarity

Rationale: Partly familiar as deep learning is a mature field, and the issue of clinical adaptation and novelty of architecture makes the problem less commonly known.

Table 5.5: Mapping with Complex Engineering Activities.

EA1 Range of re-sources	EA2 Level of Interaction	EA3 Innovation	EA4 Consequences for society and environment	EA5 Familiarity
✓	✓	✓	✓	-

5.5 Summary

CareNet development is associated with the features of complex problem-solving, knowledge profiles, and complex engineering activities, as deep mathematical, engineering, and specialist knowledge is required, and conflicting multiple demands are balanced (i.e., accuracy vs. efficiency, augmentation vs. diagnostic integrity). It combines multi-layer analysis in preprocessing, architecture, experimentation, and deep learning; however, it is a mature science because clinical adaptation and new architectures render the problem familiar to some degree. Software standards drive it and aim to anticipate the form of compliance requirements set by medical regulatory authorities, while meeting the diverse needs of stakeholders, including clinicians, patients, institutions, and regulators. One can find high-performance computing, curated medical datasets, interdisciplinary collaboration, and innovation, all the way up to the societal benefits and interdependence of all components of the system as a whole, which characterize the activity as an actual complex mathematical engineering task.

Chapter 6

Conclusion

Based on the insights and discoveries from the CareNet project, this chapter provides a comprehensive overview of the project, as presented in this report. It identifies the inventory of limitations or gaps in the current work and provides a detailed plan for further work in the upcoming phase.

6.1 Summary

In this paper, we present the design and early development of CareNet, a new hybrid deep learning architecture for classification of colon diseases from endoscopic images. The proposed approach consists of a complete pipeline design, starting by including a large block for image pretreatments (noisy processing, CLAHE processing, noise injection) to further improve robustness. A properly split data set (70% training data, 15% validation and 15% testing data) was used with the aim of carrying out a strict assessment of the model. The primary innovation of this work is in the CareNet model itself, which consists of an EfficientNetB0 backbone for efficient feature extractions along with an attention module adapted specifically for convolutional attention. This function (based on Mean-Channel/Max-Channel operations) is suggested to evaluate the regions of the images that are of interest for the diagnosis which will ensure a higher accuracy for the network and therefore more interpretability. In one example, a non-linear full dropout model with a softmax classification function has been developed to solve the very difficult problem of automated medical diagnosis.

6.2 Limitation

The current stage of the project has certain limitations such as:

- **Size and Scope:** The size and variety of the provided dataset also limit the scope, practically precluding the opportunity for development and validation. To show good generalizability, larger and more detailed data would be required to mark images in various medical institutions and endoscopic instrumentation.
- **Phase of prototype:** This is a prototype used as a working model at the time. Nevertheless, it has not been proven in a clinical environment and has not been integrated into a real clinical workflow. Its results on the validation set are still encouraging, but it is not too early to use it in a clinical setting.
- **Explainability:** Deep learning models are black boxes, even when some attention mechanism is applied to make them more explainable. The aspect of its decision-making process also requires closer analysis of the model in case it wants to gain the trust of medical professionals.

6.3 Future Work

The following step in the next stage of the Final Year Design Project will be to polish the existing limitations and work towards a deployable state of the project, and this will entail the following steps:

- Data collection and maintenance: One of the key activities will be the ability to access more datasets and a greater diversity of Bangladeshi patients, as well as the clinical associate annotation of their data to increase model resilience and generalization.
- Clinical Validation: It will be subjected to in-depth validation on a held-out clinical testing set. Comparison with the current procedure, as well as performance against methods, will be made, preferably also against clinical availability services provided by practicing gastroenterologists, to measure the clinical relevance of performance.
- Deployment Prototyping: A highly scalable model will be developed, along with a simple software interface (e.g., a web dashboard) to visualize a potential clinical workflow and gather initial feedback on the model from medical stakeholders.

References

1. https://books.google.com.bd/books?hl=en&lr=&id=4bYnYFvGolsC&oi=fnd&pg=PP1&dq=colon+diseases+and+disorders&ots=lNPCZ0zXDY&sig=2y4aXRVgelU3gIY_46QcE8m1C6M&redir_esc=y#v=onepage&q&f=false

2. Mendonça, C. M., Correa NETO, I. J. F., ROLIM, A. D. S., & Robles, L. (2022). Inflammatory bowel diseases: characteristics, evolution, and quality of life. *ABCD. Arquivos Brasileiros de Cirurgia Digestiva (São Paulo)*, *35*, e1653.

3. Swaminathan, A., Day, A. S., Sparrow, M. P., Peyrin‐Biroulet, L., Siegel, C. A., & Gearry, R. B. (2024). Measuring disease severity in inflammatory bowel disease–Beyond treat to target. Alimentary pharmacology & therapeutics, 60(9), 1176-1199.

4. Wang, Y., Huang, Y., Chase, R. C., Li, T., Ramai, D., Li, S., ... & Pang, M. (2023). Global burden of digestive diseases: a systematic analysis of the global burden of diseases study, 1990 to 2019. *Gastroenterology*, *165*(3), 773-783.

5. Roy, S., Eva, F. N., Dev, D., Roy, S., Tipu, S. K., Chowdhury, S., ... & Hawlader, M. D. H. (2024). Prevalence and predictors of functional gastrointestinal disorder among the undergraduate students of Bangladesh. *PloS one*, *19*(12), e0315687.

6. Rahman, M. M., Ghoshal, U. C., Kibria, M. G., Sultana, N., Ahmed, F., Rowshon, A. H. M., & Hasan, M. (2020). Prevalence, risk factors, and healthcare‐seeking among subjects with esophageal symptoms: A community‐based study in a rural Bangladeshi population. *JGH Open*, *4*(6), 1167-1175.

7. Hasan, A. N., Asaduzzaman, M., Mahfuj, M. N., Islam, R., Bhuyan, M. M. K., Basak, P., & Sharif, A. B. (2025). Prevalence of irritable bowel syndrome and its association to mental health among the private university students of Dhaka, Bangladesh. *PLOS Global Public Health*, *5*(5), e0004670.

8. Fardousi, A., Novak, M., Sagar, S., Chowdhury, S., Haque, R., Rahman, H., ... & Hanifi, M. A. (2024). Prevalence of gastrointestinal symptoms is higher close to the sea: evidence from coastal Bangladesh. *medRxiv*, 2024-10.

9. Hong, S. M., & Baek, D. H. (2024). Diagnostic Procedures for Inflammatory Bowel Disease: Laboratory, Endoscopy, Pathology, Imaging, and Beyond. *Diagnostics*, *14*(13), 1384. https://doi.org/10.3390/diagnostics14131384

10. Nadim, M. I., Setu, S. A., Sohel, A., Zubaer, T., Rifat, R. I., & Rahman, M. M. (2024, December). A Hybrid Deep Learning Approach to Classify Colon Cancer from Colonoscopy Image. In *2024 13th International Conference on Electrical and Computer Engineering (ICECE)* (pp. 264-269). IEEE.

11. Gowthamy, J., & Ramesh, S. (2024). A novel hybrid model for lung and colon cancer detection using pre-trained deep learning and KELM. *Expert Systems with Applications*, *252*, 124114.

12. Nadimi, E. S., Braun, J. M., Schelde-Olesen, B., Khare, S., Gogineni, V. C., Blanes-Vidal, V., & Baatrup, G. (2025). Towards full integration of explainable artificial intelligence in colon capsule endoscopy's pathway. *Scientific Reports*, *15*(1), 5960.

13. Hasan, M. A., Haque, F., Sabuj, S. R., Sarker, H., Goni, M. O. F., Rahman, F., & Rashid, M. M. (2024). An End-to-End Lightweight Multi-Scale CNN for the Classification of Lung and Colon Cancer with XAI Integration. *Technologies*, *12*(4), 56. https://doi.org/10.3390/technologies12040056

14. Ochoa-Ornelas, R., Gudiño-Ochoa, A., & García-Rodríguez, J. A. (2024). A hybrid deep learning and machine learning approach with mobile-EfficientNet and grey wolf optimizer for lung and colon cancer histopathology classification. *Cancers*, *16*(22), 3791.

15. Karthikeyan, A., Jothilakshmi, S., & Suthir, S. (2024). Colorectal cancer detection based on convolutional neural networks (CNN) and ranking algorithm. *Measurement: Sensors*, *31*, 100976.

16. Mohamed, A. A. A., Hançerlioğullari, A., Rahebi, J., Ray, M. K., & Roy, S. (2023). Colon disease diagnosis with convolutional neural network and grasshopper optimization algorithm. *Diagnostics*, *13*(10), 1728.

17. Al-Mamun Provath, M., Deb, K., & Jo, K. H. (2023, February). Classification of lung and colon cancer using deep learning method. In *International Workshop on Frontiers of Computer Vision* (pp. 56-70). Singapore: Springer Nature Singapore.

18. Kumar, V. R. P., Arulselvi, M., & Sastry, K. B. S. (2023). Comparative assessment of colon cancer classification using diverse deep learning approaches. *Journal of Data Science and Intelligent Systems*, *1*(2), 128-135.

19. Hadiyoso, S., Aulia, S., & Irawati, I. D. (2023). Diagnosis of lung and colon cancer based on clinical pathology images using convolutional neural network and CLAHE framework. *Int. J. Appl. Sci. Eng*, *20*(1), 1-7.

20. Schiele, S., Arndt, T. T., Martin, B., Miller, S., Bauer, S., Banner, B. M., ... & Müller, G. (2021). Deep learning prediction of metastasis in locally advanced colon cancer using binary histologic tumor images. *Cancers*, *13*(9), 2074.

21. Wang, W., Tian, J., Zhang, C., Luo, Y., Wang, X., & Li, J. (2020). An improved deep learning approach and its applications on colonic polyp images detection. *BMC Medical Imaging*, *20*, 1-14.

22. Sarwinda, D., Paradisa, R. H., Bustamam, A., & Anggia, P. (2021). Deep learning in image classification using residual network (ResNet) variants for detection of colorectal cancer. *Procedia Computer Science*, *179*, 423-431.

23. https://www.kaggle.com/datasets/francismon/curated-colon-dataset-for-deep-learning

24. Dey, R. K., Rana, M. E., & Hameed, V. A. (2023, January). Analysing wireless capsule endoscopy images using deep learning frameworks to classify different GI tract diseases. In *2023 17th International Conference on Ubiquitous Information Management and Communication (IMCOM)* (pp. 1-7). IEEE.

25. Nadim, M. I., Setu, S. A., Sohel, A., Zubaer, T., Rifat, R. I., & Rahman, M. M. (2024, December). A Hybrid Deep Learning Approach to Classify Colon Cancer from Colonoscopy Image. In *2024 13th International Conference on Electrical and Computer Engineering (ICECE)* (pp. 264-269). IEEE.

26. Kumar, A., & Nelson, L. (2024, December). Automated Classification of Gastrointestinal Conditions from Capsule Endoscopy Images Using EfficientNetB2. In *2024 International Conference on Innovation and Novelty in Engineering and Technology (INNOVA)* (Vol. 1, pp. 1-6). IEEE.

27. Sulistyawan, J. S., Nugroho, K. S., Isnan, M., & Pardamean, B. (2024, November). Comparative Analysis of CNN Models for Multiclass Detection on Colon Disease. In *2024 9th International Conference on Information Technology and Digital Applications (ICITDA)* (pp. 01-06). IEEE.

28. Dwivedi, A. K., Srivastava, G., & Pradhan, N. (2023, March). NFF: A Novel Nested Feature Fusion Method for Efficient and Early Detection of Colorectal Carcinoma. In *Proceedings of Fourth International Conference on Computer and Communication Technologies: IC3T 2022* (pp. 297-309). Singapore: Springer Nature Singapore.

213-15-4303

ORIGINALITY REPORT

20%	16%	13%	13%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	5%
2	www.mdpi.com Internet Source	2%
3	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%
4	Submitted to United International University Student Paper	1%
5	Alistair Michael Baretto, Raj Sunil Salvi, Shreyas Rajesh Labhsetwar, Piyush Arvind Kolte, Veerasai Subramaniam Venkatesh. "Analysis of Dimensional Influence of Convolutional Neural Networks for Histopathological Cancer Classification", 2021 4th Biennial International Conference on Nascent Technologies in Engineering (ICNTE), 2021 Publication	<1%
6	www.nature.com Internet Source	<1%
7	Thangaprakash Sengodan, Sanjay Misra, M Murugappan. "Advances in Electrical and Computer Technologies", CRC Press, 2025 Publication	<1%
8	Submitted to RDI Distance Learning Student Paper	<1%