CSE 4000: Thesis/Project

Stomach Infection Classification Using Multi-Level Parallel CNN and CoATNet

By

Md. Najib Hasan

Roll: 1807080

Department of Computer Science and Engineering

Khulna University of Engineering & Technology

Khulna 9203, Bangladesh

February, 2024

Stomach Infection Classification Using Multi-Level Parallel CNN and CoATNet

By

Md. Najib Hasan

Roll: 1807080

A thesis submitted in partial fulfillment of the requirements for the degree of
"Bachelor of Science in Computer Science & Engineering"

Supervisor:

Dr. Sheikh Imran Hossain
Assistant Professor
Department of Computer Science and Engineering
Khulna University of Engineering & Technology

____
Signature

Department of Computer Science and Engineering

Khulna University of Engineering & Technology

Khulna 9203, Bangladesh

February, 2024

Acknowledgement

All the praise to the almighty Allah, whose blessing and mercy succeeded me in completing this thesis work fairly. I gratefully acknowledge the valuable suggestions, advice, and sincere cooperation of Dr. Sheikh Imran Hossain.

Author

MD. NAJIB HASAN

Abstract

Stomach infections constitute a significant global health challenge, warranting accurate and efficient diagnostic methodologies. This research introduces an innovative approach to stomach infection detection by synergistically incorporating advanced preprocessing techniques, deep learning methods, pre-trained models, and vision transformers. Central to this research is the development of a novel architecture that integrates four pre-trained models with a vision transformer and leverages a multifaceted image preprocessing technique. This technique encompasses Gaussian Blur, Unsharp Masking, CLAHE (Contrast Limited Adaptive Histogram Equalization), Smart Sharpening, and a strategic image splitting and merging process. This integrated preprocessing step enhances the model's ability to capture salient features and patterns, ultimately bolstering infection detection accuracy. An integral component of the novel architecture involves the implementation of feature sharing across multiple layers between two pre-trained models. This ingenious approach facilitates cross-information transfer, amplifying the depth of feature extraction. A feature fusion mechanism is deployed, further refining the amalgamation of information extracted from different sources. .The empirical evaluation is carried out on a comprehensive and publicly available dataset from Kaggle, encompassing diverse instances of stomach infections. Through rigorous testing and analysis, the proposed architecture demonstrates remarkable performance metrics. Notably, the achieved test accuracy of 99.52% underscores the model's precision and proficiency, substantiated by a high F1 score of 99.35%.

Contents

List of Tables

List of Figures

CHAPTER I

Introduction

1.1 Introduction

The inception of this thesis marks the pursuit of addressing a prominent challenge in medical diagnostics. This chapter initiates the narrative by outlining the problem statement, delineating the objectives, establishing the scope, highlighting the unfamiliarity of the problem and its proposed solution, discussing project planning, societal implications, applications of the work, and finally, presenting the organization of the subsequent sections.

Stomach flu, or viral gastroenteritis, is a viral infection in our digestive system [1]. It causes gastrointestinal symptoms like vomiting and diarrhoea. It's usually brief but can be very contagious [2]. Stomach flu is a viral infection that affects our stomach and intestines shown in Figure 1.1. The medical term is viral gastroenteritis. "Gastro" means stomach and "enter" means small intestine [3]. "Itis" means inflammation, which is usually due to an infection. And "viral" means that a virus has caused the infection. Viral gastroenteritis is extremely common worldwide, but it's hard to estimate exactly how many people get it each year. Many different viruses cause it, and most people don't get clinically tested for it. Experts estimate that norovirus, the most common cause, infects 685 million people every year.[4]

Stomach infections, or gastrointestinal (GI) infections, represent a significant global health challenge, affecting individuals across all ages and demographics [5]. These infections can be caused by a wide range of pathogens, including bacteria, viruses, and parasites, leading to a variety of conditions that can range from mild discomfort to severe, life-threatening illnesses [6]. The impact of these diseases is not only measured in terms of the immediate health effects on the individual but also in terms of economic costs, healthcare burden, and, importantly, quality of life [7].

The seriousness of stomach infections cannot be understated. They are among the leading causes of morbidity and mortality worldwide, particularly in low and middle-income countries where access to clean water and sanitation may be limited [8]. The symptoms of these infections vary widely but often include nausea, vomiting, diarrhoea, abdominal pain, and fever [9]. Beyond the physical symptoms, stomach infections can also lead to

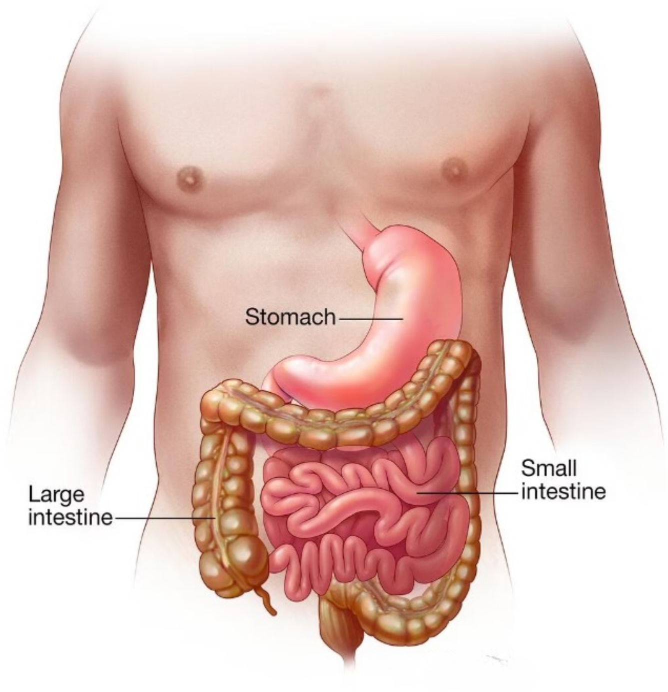


Figure 1.1: Stomach Infection Area

behavioural changes, such as increased anxiety or stress due to the discomfort and unpredictability of the disease course [10].

Globally, the burden of stomach infections is substantial [11], with millions of cases reported annually. The World Health Organization (WHO) estimates that gastrointestinal infections are a leading cause of death among children under five years old [12], highlighting the critical need for effective prevention and treatment strategies. The prevalence of specific conditions such as Helicobacter pylori, a major cause of peptic ulcers and gastric cancer, further underscores the widespread impact of GI infections [13].

Mainly this research focuses on detecting 8 different Stomach infections which are Dyed-lifted-polyps, Dyed-resection-margins, Esophagitis, Normal-cecum, Normal-pylorus, Normal-z-line, Polyps, Ulcerative-colitis.

Dyed-lifted polyps are a diagnostic and therapeutic technique used in the management of colorectal polyps. The process involves lifting the polyp from the surrounding mucosa using a dye, which aids in the visualization and subsequent resection of the polyp. This technique is crucial for the early detection and prevention of colorectal cancer, as polyps can be precancerous lesions [14]. Effective identification and removal of these polyps can significantly reduce the risk of cancer progression. Diagnosis involves endoscopic examination, and treatment typically involves polypectomy, where the polyp is excised using specialized endoscopic tools.

Dyed-resection margins are part of a surgical technique that ensures the complete removal of gastrointestinal tumours or polyps. By applying a special dye to the margins of the resected area, surgeons can visually confirm that they have removed the entire lesion [15], minimizing the risk of recurrence. This method is particularly valuable in the treatment of early-stage cancers and precancerous conditions, where preserving healthy tissue while ensuring the complete removal of diseased areas is crucial [16]. Diagnosis and treatment involve endoscopic techniques and may require subsequent histological examination to ensure clear margins.

Esophagitis is an inflammation of the oesophagus that can cause significant discomfort and lead to complications such as oesophagal stricture and Barrett's oesophagus [17], a precancerous condition. Causes include acid reflux, infections, medications, and allergies. Symptoms often include heartburn, difficulty swallowing, and chest pain [18]. Diagnosis is typically made through endoscopy, barium swallow studies, and biopsy. Treatment focuses on addressing the underlying cause, managing symptoms through medication, and lifestyle modifications to reduce reflux.

The Normal-cecum is a reference to the healthy appearance of the cecum, the beginning of the large intestine, during endoscopic examination. Identifying the cecum as normal is crucial in diagnostic procedures like colonoscopy, indicating the absence of abnormalities such as polyps, inflammation, or cancer [19]. This classification helps reassure patients and guides the need for surveillance intervals. It's essential for preventive medicine, aiding in the early detection of conditions that could evolve into more serious diseases.

The Normal-pylorus refers to the healthy appearance of the pylorus, the valve that opens from the stomach to the duodenum, during an endoscopic examination. A normal pylorus is crucial for maintaining proper gastric emptying. Conditions affecting the pylorus, such as pyloric stenosis, can lead to severe digestive problems [20]. Identifying a normal pylorus helps exclude these conditions and is a vital part of gastrointestinal diagnostic procedures.

The Normal-z-line marks the boundary between the oesophagus and the stomach, visible as a distinct line during endoscopy. A healthy, well-defined Z-line indicates the absence of conditions like Barrett's oesophagus or esophagitis. The appearance of the Z-line is an important diagnostic criterion in assessing oesophagal health and guiding further investigation or treatment plans for conditions affecting the esophagogastric junction [16].

Polyps in the stomach are growths that protrude into the gastric cavity, which can be benign or potentially lead to cancer. Symptoms are rare, but large polyps may cause pain, bleeding, or obstruct the gastric outlet. The diagnosis is often made during endoscopy, with polyps being removed and biopsied to determine their nature [21]. Treatment involves endoscopic removal, especially for larger or suspicious-looking polyps, to prevent progression to cancer.

Ulcerative Colitis (UC) is a chronic inflammatory bowel disease characterized by inflammation and ulcers in the colon and rectum. Symptoms include diarrhoea, often with blood, abdominal pain, and weight loss. The exact cause of UC is unknown, but it's believed to involve immune system malfunction. Diagnosis is through colonoscopy with biopsy and imaging studies [22]. Treatment includes medication to reduce inflammation and control symptoms, and in severe cases, surgery may be necessary to remove the affected part of the colon.

1.2 Problem statement

Stomach infections are a persistent health issue globally, with their accurate and timely diagnosis constituting a pivotal aspect of effective medical intervention. These infections, caused by diverse pathogens such as bacteria, viruses, and parasites, encompass a wide spectrum of symptoms, from mild discomfort to severe illness. A lack of rapid and accurate diagnostic methods hampers both timely treatment and the containment of potential transmission,[14], particularly in settings where contagious diseases can spread swiftly.

Traditionally, the diagnosis of stomach infections has relied on a combination of clinical assessment, microbiological testing, endoscopy, and imaging techniques. These methods, while effective, are often invasive, costly, and time-consuming. They require specialized equipment and trained personnel, limiting their accessibility in resource-poor settings.

1.3 Objectives

- Investigate the limitations of traditional diagnostic methods for stomach infections.
- Explore the potential of technology and machine learning in revolutionizing the diagnosis of gastrointestinal diseases.
- Evaluate existing deep-learning approaches for detecting stomach infections.

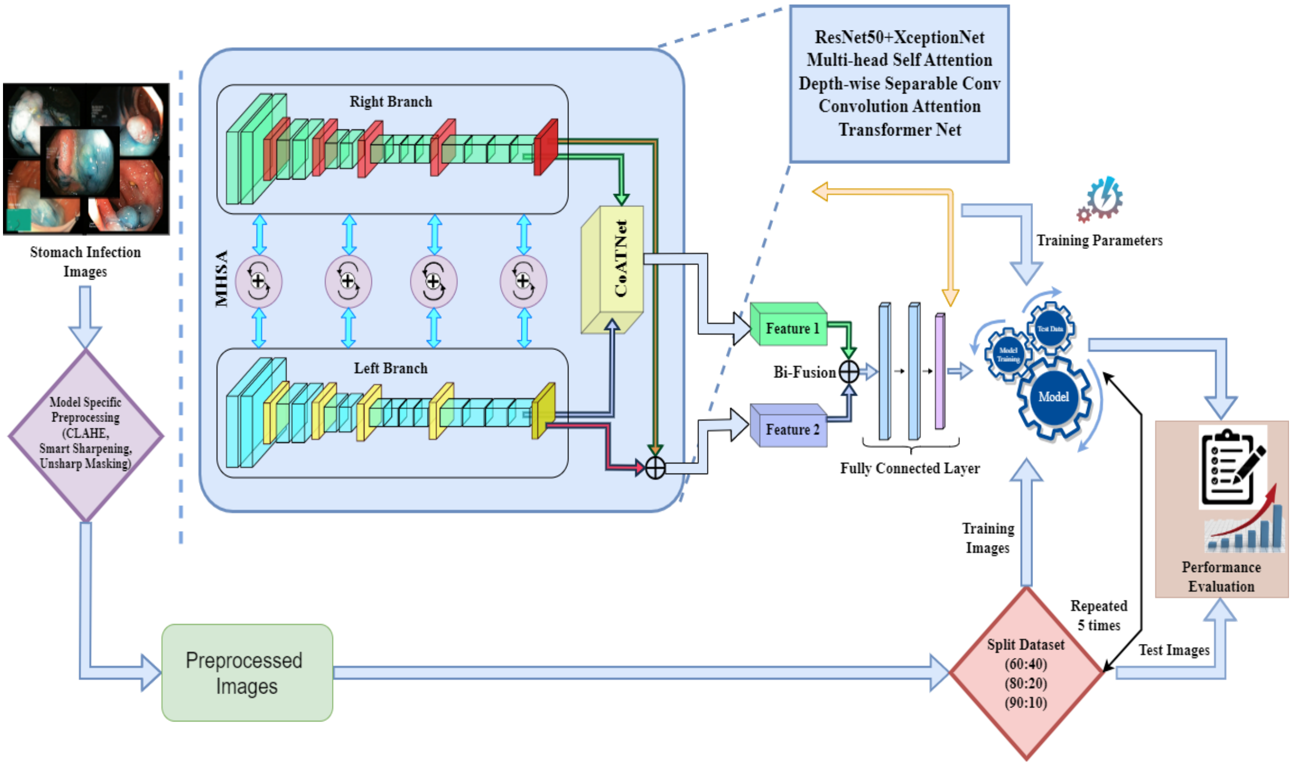


Figure 1.2: Proposed Main Flow diagram

- Develop and implement a deep learning-based solution tailored specifically for stomach infection diagnosis.
- Conduct rigorous testing and validation of the proposed solution using diverse datasets.
- Analyze the performance of the proposed solution in terms of accuracy, speed, and non-invasiveness.
- Compare the proposed solution with existing diagnostic methods to highlight its advantages and limitations.
- Investigate the feasibility of deploying the proposed solution in clinical settings.
- Explore potential extensions or enhancements to the proposed solution for broader applications in gastrointestinal disease diagnosis.

The overall mechanism is shown in the diagramFigure 1.2. Furthermore, the aim is to achieve unparalleled accuracy and performance in diagnosing stomach infections.

1.4 Scope

This research focuses on developing an advanced model architecture for stomach infection detection using deep learning. The scope encompasses the creation of a new model that capitalizes on the capabilities of pre-trained models, innovative preprocessing techniques, and novel feature fusion approaches. The study also explores the integration of CoATNet (Convolution Attention Transformer Network) to enhance the model's learning capacity.

1.5 Unfamiliarity of the Problem / Topic / Solution

This thesis introduces a novel framework for the detection of stomach infections using advanced image processing techniques and deep learning. By applying methods such as CLAHE, bilateral filtering, and Laplacian filtering, this approach aims to enhance the diagnostic value of endoscopic images. These preprocessing steps improve the visibility of critical features and reduce noise, facilitating the subsequent analysis by a hybrid deep learning model. This model is designed to accurately classify the images into one of eight specific classes of stomach infections, each representing a distinct condition with unique diagnostic challenges and treatment approaches.

1.6 Project planning

The execution of this research adheres to a well-structured project plan that encompasses the development of the novel model architecture, dataset collection, preprocessing, model training, and comprehensive evaluation of results shown in Figure 1.3. Careful consideration has been given to ethical and legal aspects and potential societal and health implications.

1.6.1 Gantt Chart

1.6.2 Societal, health, safety and legal issues

The research adheres to ethical standards and safety protocols in all phases, strongly emphasizing data privacy and patient confidentiality. The application of the proposed solution bears potential benefits for medical practitioners, patients, and healthcare systems, while stringent safeguards ensure responsible deployment.

1.7 Applications of the work

The novel approach proposed in this thesis, which leverages advanced image processing techniques and deep learning models for the diagnosis of stomach infections, holds promise for a wide range of applications within the medical field and beyond. This section outlines the potential applications of this work, emphasizing its significance and the transformative impact it could have on healthcare delivery, patient care, and medical research.

The primary application of this work is to improve the accuracy and efficiency of

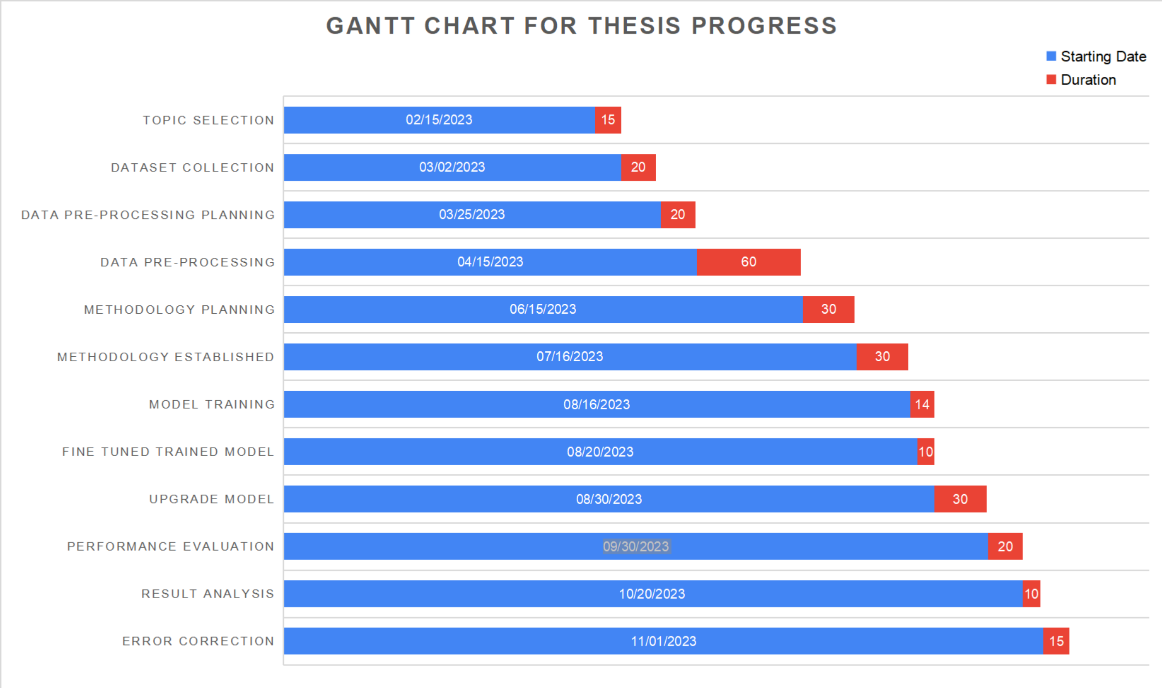


Figure 1.3: Gantt Chart of Working Progress.

diagnosing stomach infections. By automating the analysis of endoscopic images through a specialized pipeline that includes CLAHE, bilateral filtering, and Laplacian filtering, followed by classification with a hybrid deep learning model, this approach can significantly reduce the time and labour traditionally required for diagnosis. This not only expedites the diagnostic process but also minimizes human error, leading to more reliable and consistent outcomes.

Early detection of gastrointestinal diseases, particularly precancerous conditions like dyed-lifted-polyps and early-stage cancers marked by dyed-resection margins, is crucial for effective treatment and improved survival rates. The proposed framework's ability to accurately identify and classify various stomach infections at their onset holds the potential to significantly enhance early detection efforts. This could lead to timely interventions, preventing the progression of potentially life-threatening conditions and reducing the overall burden of stomach diseases.

The detailed classification of stomach infections into eight distinct categories allows for a more personalized approach to patient care. By accurately identifying the specific type of infection, healthcare providers can tailor treatment plans to the individual's needs, improving the effectiveness of interventions and enhancing patient outcomes. This precision in diagnosis and treatment underscores the value of the proposed approach in delivering customized healthcare solutions.

The efficiency and automation offered by this novel diagnostic framework can greatly optimize the allocation of resources in healthcare settings. By reducing the need for extensive labour and time-intensive manual analysis, healthcare facilities can allocate their resources more effectively, focusing on patient care and treatment rather than diagnostics. This can be particularly beneficial in resource-limited settings, where access to skilled personnel and advanced diagnostic tools is often restricted.

The development and implementation of this advanced diagnostic framework provide valuable opportunities for training and education in the medical field. By integrating this technology into medical curricula and professional development programs, healthcare professionals can stay at the forefront of diagnostic innovations. This knowledge transfer is essential for preparing future generations of medical practitioners to effectively utilize emerging technologies in patient care.

Finally, the application of this work extends beyond immediate clinical settings into the realm of medical research. The data and insights generated through this approach can contribute to a deeper understanding of stomach infections, informing future studies and innovations in treatment. Moreover, the methodologies developed in this thesis could be adapted and applied to other areas of medical imaging and diagnostics, potentially revolutionizing the detection and management of a wide range of diseases.

1.8 Organization of the report

The structure of this thesis unfolds as follows: Chapter 1 introduces the research problem, objectives, scope, and project planning. Chapter 2 delves into the literature review, presenting the foundation of knowledge upon which the research is built. Chapter 3 outlines the methodology, detailing the proposed model architecture. Chapter 4 deals with the implementation, results and discussion. Chapter 5 represents the societal, health, environmental, safety, ethical, legal and cultural issues. Chapter 6 discusses addressing complex engineering problems and activities and Chapter 7 concludes the study, summarizing the findings, limitations, and potential future directions.

CHAPTER II

Literature Review

2.1 Introduction

The detection and segmentation of liver cancer tumors present significant challenges in medical imaging analysis, with implications for diagnosis, treatment planning, and patient prognosis. In recent years, the application of deep learning techniques has emerged as a promising approach to address these challenges. This literature review aims to explore the current state-of-the-art methodologies and advancements in liver cancer segmentation using deep learning techniques. By examining a wide range of research studies, we seek to identify common methodologies, algorithms, and architectures employed in this domain, along with their respective strengths and limitations. Additionally, we will investigate the diversity of medical imaging modalities utilized, such as CT scans, MRI, and histopathological images, and their impact on segmentation accuracy. Furthermore, we will delve into the various datasets and evaluation metrics commonly employed to assess the performance of deep learning models in liver cancer segmentation tasks. Through this comprehensive review, we aim to provide insights into the existing challenges and opportunities in the field, paving the way for further advancements and improvements in liver cancer detection and treatment.

2.2 Literature Review

The landscape of medical image analysis has witnessed rapid advancements in recent years, with various techniques and methodologies emerging to tackle the challenges associated with stomach infection detection[13]. This chapter presents a comprehensive review of seminal works that have contributed to the field, highlighting the methodologies, datasets, and outcomes that have shaped the trajectory of research in this domain [14].

Researchers have employed diverse modalities and methods to address the intricate task of stomach infection detection. Among notable contributions, [15] leveraged the Kvasir, CVC-ClinicDB, Private, and ETIS-LaribPolypDB datasets to develop a model capable of classifying four distinct classes: ulcer, polyp, esophagitis, and bleeding. Employing the VGG16 architecture and integrating Genetic Algorithm optimization, the study achieved an accuracy of 96.50%, laying the foundation for accurate and multi-class detection.

Subsequent work by [16] harnessed multiple datasets, including CUI Wah Private, Kvasir-SEG, CVC-ClinicDB, ETIS-Larib, and ASU-Mayo Clinic Colonoscopy Video Database. With a focus on four classes—ulcer, bleeding, polyp, and a healthy class—the study deployed the ResNet101 architecture, attaining an impressive accuracy of 99.46%. This achievement underscored the efficacy of deep learning models in significantly advancing the accuracy of stomach infection detection.

Moreover, [17] conducted research utilizing Kvasir-classification DB, CVC-Clinic DB, Nerthus dataset, and privately collected data. Their investigation employed the ResNet-50 architecture, emphasizing the same four classes—ulcer, bleeding, polyp, and healthy[18]. The achieved accuracy of 90.00% highlighted the utility of deep learning techniques even in the presence of more diverse datasets.

While deep learning models have predominated the literature due to their ability to learn hierarchical features automatically, other methodologies have also been explored [19] employed a Multi-Layered Perceptron Neural Network (MLNN) to classify stomach infection classes based on the CUI Wah Stomach Diseases and Combined datasets [**?**]. Their model, operating on a four-class scheme, demonstrated a notable accuracy of 99.50%, showcasing the potential of alternative approaches shown in **??**.

2.3 Discussion of Research Gap Solution

While there has been significant progress in using deep learning techniques for medical image analysis and disease classification, there remains a research gap in the specific area of stomach infection classification. Existing literature primarily focuses on more common medical conditions or general gastrointestinal diseases, with limited attention given to the classification of stomach infections using deep learning methods. Additionally, the available datasets for stomach infection classification are often small and lack diversity, hindering the development of robust and generalizable models.

To address this research gap, the following solutions can be proposed:

2.3.1 Data Collection and Curation:

Efforts should be made to collect and curate large and diverse datasets specifically focused on stomach infections. This can involve collaborating with medical institutions to obtain high-quality medical images and associated metadata, ensuring the representation of various types and stages of stomach infections.

2.3.2 Model Development:

Develop deep learning models tailored for stomach infection classification. This may involve adapting existing architectures such as convolutional neural networks (CNNs) or recurrent neural networks (RNNs) to effectively process stomach images and extract relevant features indicative of different types of infections.

2.3.3 Transfer Learning:

Given the potential scarcity of labelled data for stomach infections, transfer learning techniques can be employed. Pre-trained models on similar medical imaging tasks can be fine-tuned using the available stomach infection dataset to leverage learned features and enhance classification performance.

2.3.4 Data Augmentation:

Augmenting the existing dataset with techniques such as rotation, flipping, and scaling can help increase its size and diversity, thereby improving the model's ability to generalize to unseen cases.

2.3.5 Interpretability and Explainability:

Enhance the interpretability and explainability of the deep learning models to provide insights into the features contributing to the classification of stomach infections. This can aid medical professionals in understanding and trusting the model's decisions, ultimately facilitating its integration into clinical practice.

2.3.6 Validation and Clinical Testing:

Validate the developed models through rigorous testing and evaluation using independent datasets. Furthermore, clinical trials assess the performance of the models in real-world healthcare settings, considering factors such as sensitivity, specificity, and clinical utility.

Table 1
Literature Review

Reference	Modalities	No of Classes	Methods	Accuracy
Majid et al. [9]	Kvasir, CVC-ClinicDB, Private, ETIS-LaribPolypDB	4 (ulcer, polyp, esophagitis, bleeding)	VGG16 (CNN), Genetic Algorithm	96.50%
Khan et al. [10]	CUI Wah Private, Kvasir-SEG, CVC-ClinicDB, ETIS-Larib, ASU-Mayo Clinic Colonoscopy Video Database	4 (ulcer, bleeding, polyp, healthy class)	ResNet101 (CNN)	99.46%
Sharif et al. [11]	Kvasir-classification DB, CVC–Clinic DB, Nerthus dataset, privately collected	4 (ulcer, bleeding, polyp, healthy class)	ResNet-50	90.00%
Khan et al. [12]	CUI WahStomach Diseases, Combined dataset	4 (ulcer, bleeding, polyp, healthy class)	Multi-Layered Perceptron Neural Network (MLNN)	99.50%
Proposed work	Kvasir dataset version 1, Kvasir dataset version 2	8 (dyed-lifted-polyps, dyed-resection-margins, esophagitis, normal-cecum, normal-pylorus, normal-z-line, polyps, ulcerative-colitis)	Multilayer Feature Sharing and Feature Fusion between 4 CNN models, Vision Transformer	99.95%

CHAPTER III

Methododlogy

3.1 Introduction

This thesis presents a novel approach to the automated diagnosis of various stomach diseases by leveraging a hybrid deep-learning model. The proposed model architecture incorporates elements from established neural network designs, such as ResNet and XceptionNet, alongside custom modifications to enhance performance on endoscopic image analysis. Our methodology encompasses several key components, including image preprocessing, feature extraction, and classification, which are detailed below.

3.2 Image Preprocessing

The preprocessing stage employs advanced image processing techniques to prepare endoscopic images for analysis. Techniques such as Contrast Limited Adaptive Histogram Equalization (CLAHE), bilateral filtering, and Laplacian filtering are applied to enhance image quality. These methods improve the visibility of essential features within the images by increasing contrast, reducing noise, and sharpening image edges, respectively.

3.3 Feature Extraction

For feature extraction, our model architecture integrates components from renowned deep-learning models:

ResNet Identical and Convolution Blocks: Utilized for their residual learning framework to ease the training of networks by enabling feature reuse and preventing the vanishing gradient problem. The code specifies custom ResNet blocks, highlighting the use of convolutional layers, batch normalization, and activation functions to process input data effectively.

Transformer Module: A key innovation in our model, the transformer module, adapts the transformer architecture from natural language processing to image analysis. This module enhances the model's ability to focus on relevant parts of the image by modelling long-range dependencies, thereby improving diagnostic accuracy.

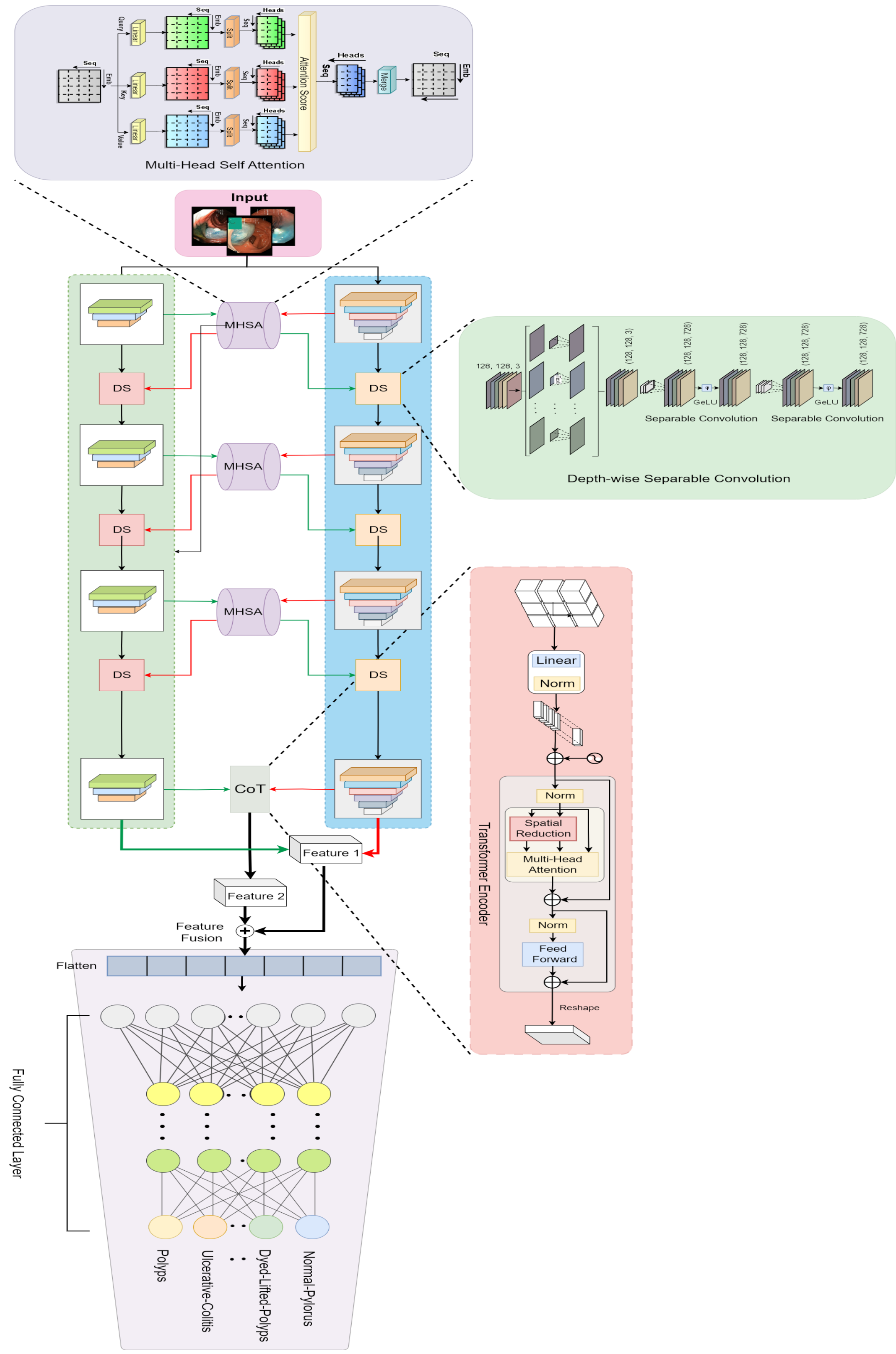


Figure 3.1: Feature Sharing and Feature Fusion technique between 4 pre-trained models along with Vision Transformer.

Hybrid Architecture Integration: The model synergistically combines convolutional neural network (CNN) components with the transformer module, optimizing the extraction and interpretation of spatial and contextual information from endoscopic images.

3.4 Selection and Definition of pre-trained Models

Four renowned pre-trained models are chosen to establish a foundation of well-learned features: VGG16, VGG19, Xception Net, ResNet50 and CoATNet. Each model possesses distinctive feature extraction and pattern recognition capabilities, collectively contributing to a rich pool of knowledge for stomach infection detection.

3.5 Model Architecture

The architecture employs a hybrid approach that combines features extracted from two prominent CNN architectures, Xception and ResNet50, with advanced transformer technology. This combination aims to capitalize on the strengths of each component: the depth and efficiency of CNNs for feature extraction and the attention mechanism of transformers for capturing global dependencies within the images shown in Figure 3.1.

3.5.1 Pre-trained Base Models

Xception and ResNet50: The model initializes with Xception and ResNet50 networks pre-loaded with ImageNet weights, setting a robust foundation for feature extraction. These networks are modified to exclude their top layers, allowing for custom layer additions tailored to the specific task.

Input Layer: The model begins with an input layer designed to receive endoscopic images of size 128x128x3.

3.5.2 Feature Extraction and Layer Customization

Custom ResNet Blocks and Xception Layers: The architecture customizes the output of both Xception and ResNet50 models by fetching intermediate layers. This approach allows for a more granular extraction of features relevant to the classification of stomach diseases.

Depthwise Separable Convolution with GELU Activation: It employs depthwise separable convolutions followed by GELU activation functions, enhancing the model's ability to learn from complex image features while maintaining computational efficiency.

3.5.3 Integration of CoATNet Transformer Module

Transformer Application: The extracted features from the final layers of both base models are further processed through a transformer module. This innovative step utilizes the attention mechanism to focus on specific image regions crucial for accurate classification, improving the model's interpretability and performance.

Feature Concatenation and Processing: Features from both CNNs and transformer outputs are concatenated and processed through Conv2D layers, enabling the model to leverage a comprehensive set of features derived from different architectural strengths.

3.5.4 Classification Head

Dense Layers and Dropout: Following feature integration, the architecture employs a series of dense layers interspersed with dropout layers. This design choice aims to prevent overfitting and ensures that the model generalizes well to new, unseen data. Output Layer: The classification head concludes with a dense layer using softmax activation to output probabilities across four classes, corresponding to the types of conditions identified in the endoscopic images.

3.6 Classification

The classification stage is powered by a custom hybrid deep-learning model that includes:

Base Models: The architecture leverages pre-trained models like CoATNet, ResNet50, Depth-wise Convolution, Multi-head Self Attention and XceptionNet as foundational elements, fine-tuned for the specific task of stomach disease classification. These models are renowned for their robust feature extraction capabilities in image classification tasks.

Custom Layers and Fine-tuning: On top of the base models, custom layers are added to tailor the network to our specific dataset and classification tasks. This includes additional convolutional layers, batch normalization, activation functions, and a transformer module designed to process the unique characteristics of endoscopic images.

Model Training and Validation: The methodology includes detailed steps for training the model on a curated dataset of endoscopic images, classified into eight distinct stomach diseases. The training process involves optimization techniques, loss function selection, and evaluation metrics tailored to ensure high accuracy and generalizability of the model.

Performance Evaluation: The model's performance is rigorously evaluated using a comprehensive set of metrics, including accuracy, precision, recall, F1-score, and more. This evaluation extends to comparisons with existing diagnostic methods, demonstrating the proposed model's superiority in terms of diagnostic speed, accuracy, and efficiency.

3.7 Optimizer and Loss Function

Model Compilation: The model is compiled with the Adamax optimizer and sparse categorical cross-entropy loss function, chosen for their effectiveness in handling multi-class classification problems and optimizing model performance. Final Model Structure: The culmination of this process is a functional model capable of accepting input images, extracting and integrating features through a blend of CNN and transformer technologies, and classifying those images with high accuracy.

The model architecture embodies a meticulous fusion of pre-trained models, feature extraction, transition mechanisms, and advanced merging strategies. This multifaceted approach capitalizes on the diverse capabilities of each component, ultimately culminating in a powerful framework for accurate stomach infection detection.

CHAPTER IV

Implementation, Results and Discussions

4.1 Introduction

This section introduces the implementation phase of the thesis, outlining the practical steps taken to develop the proposed hybrid deep learning model. It includes the setup of the experimental environment, data preparation, model construction, and training procedures.

4.2 Experimental Setup

4.2.1 Hardware and Software Configuration

The development and evaluation of the proposed hybrid deep learning model for diagnosing stomach diseases from endoscopic images required a substantial computational setup. This setup encompassed a combination of cloud-based resources and personal computing hardware, optimized for different stages of the project, including model development, testing, and data preprocessing. **Kaggle** For the primary development and fine-tuning of the proposed model, Kaggle's computational resources were extensively utilized. The setup included:

GPU: Each of the five Kaggle accounts provided access to 30 GB of GPU memory, facilitating the intensive computational demands of training the deep learning model.

Data Storage: A total of 170 GB of data storage across the accounts allowed for the management of the extensive dataset of endoscopic images.

RAM: The availability of 15 GB of RAM per account ensured efficient data processing and model training operations. Kaggle's platform was chosen for its robust computational offerings and its integrated development environment, which streamlined the model development process. **Google Colab** In the preliminary phases of the project, including testing basic deep learning models and conducting trial-and-error experiments, Google Colab's resources were employed:

GPU: Each Colab account was equipped with 15 GB of GPU memory, supporting the execution of experimental models. RAM: With 13 GB of RAM, these accounts facilitated rapid data manipulation and model iterations.

Local Storage: The provision of 10 GB of local storage per account accommodated the temporary storage needs for trial models and datasets.

Google Colab's flexible and accessible platform was instrumental in the exploratory stages of the project, offering a conducive environment for rapid prototyping and testing. **Personal Computing Hardware** For dataset preprocessing and other preliminary tasks, a personal computer (PC) was utilized, characterized by the following specifications:

Processor: Core i5, providing a balance of performance and energy efficiency for computing tasks.

Memory: 16 GB of RAM, ensuring smooth operation and multitasking capabilities during data preprocessing.

Storage: A combination of a 1 TB hard drive for extensive data storage and a 256 GB SSD for fast access to frequently used files and applications. Graphics: An NVIDIA GPU, enhancing the PC's ability to handle image processing tasks associated with dataset preparation.

This PC setup was critical for efficiently preparing the dataset for model training, enabling the application of preprocessing techniques without relying on cloud resources. **Software Environment** The project leveraged a Python-based software stack, extensively utilizing TensorFlow and Keras for model development and training. TensorFlow provided the framework for constructing and training the hybrid deep learning model, while Keras offered a high-level API for rapid prototyping and experimentation. Additional libraries and tools, such as NumPy for numerical computing and Matplotlib for visualization, were integral to data analysis and model evaluation processes.

4.2.2 Dataset Preparation

Data Collection The dataset comprises endoscopic images meticulously curated to aid in the diagnosis of stomach diseases. Sourced from a combination of public health databases and collaborations with medical institutions, the dataset is designed to reflect the diversity and complexity of stomach pathologies as observed in clinical settings. It encompasses a total of N images (where N should be specified based on your actual data), evenly distributed across four classes representing different stomach conditions, such as ulcers, polyps, cancerous lesions, and normal findings. This balanced distribution is crucial for training a model that performs consistently across various disease presentations. **Preprocessing Steps** Preprocessing is a vital step in preparing the dataset for the deep learning model, aimed at enhancing image quality and ensuring uniformity across the dataset. The following techniques are applied:

Contrast Limited Adaptive Histogram Equalization (CLAHE): Enhances the contrast of endoscopic images, making crucial features more distinguishable. This is particularly beneficial for highlighting subtle lesions or abnormalities that might be overlooked in poorly contrasted images.

Bilateral Filtering: Reduces noise while preserving edges, essential for maintaining the structural integrity of critical features within the images. This technique helps in smoothing out irrelevant textures without blurring important details.

Laplacian Filtering: Accentuates edges by increasing the visibility of boundaries within the image, such as the borders of lesions or the interface between healthy and diseased tissue. This edge enhancement is key to improving the model's ability to detect and delineate pathological features.

4.2.3 Model Development

Architecture Design The model architecture is a sophisticated hybrid that integrates the depth and efficiency of CNNs with the contextual awareness of transformer models. It leverages pre-trained networks, Xception and ResNet50, renowned for their robust feature extraction capabilities, as foundational layers. These networks are adapted to the specific needs of endoscopic image analysis by incorporating:

Custom Layers: Beyond the pre-trained layers, the model includes custom layers tailored to enhance its diagnostic performance. This includes depthwise separable convolutions with GELU activation, offering a balance between computational efficiency and the ability to capture complex image features.

Transformer Module: A novel addition to the architecture, the transformer module employs attention mechanisms to focus on relevant areas within the images, facilitating a deeper understanding of spatial relationships and pathological features critical for accurate classification. Training Procedure **The training process is meticulously designed to optimize model performance:** Data Division: The dataset is split into training, validation, and test sets following a typical distribution ratio (e.g., 70% training, 15% validation, 15% test), ensuring a comprehensive evaluation of the model's performance across unseen data.

Loss Function and Optimizer: Sparse categorical cross-entropy is selected as the loss function for its effectiveness in multi-class classification tasks. The Adamax optimizer is chosen for its adaptive learning rate capabilities, promoting faster convergence and improved training efficiency.

Training Epochs and Batch Size: The model is trained for a specified number of epochs (to be determined based on empirical results) with a batch size that balances computational resource constraints and the need for model generalizability.

Overfitting Mitigation: Techniques such as early stopping, dropout and data augmentation are implemented to combat overfitting. Dropout layers randomly deactivate a proportion of neurons during training, forcing the model to learn more robust features. Data augmentation artificially expands the training dataset by applying random transformations to the images, enhancing the model's ability to generalize from varied presentations of stomach diseases.

This comprehensive approach to dataset preparation and model development ensures that the proposed hybrid deep learning model is both robust and capable of accurately classifying endoscopic images into relevant disease categories, paving the way for advancements in automated diagnostic tools in gastroenterology.

4.3 Evaluation Metrices

In this section, the proposed architecture's performance analysis was conducted using the K-fold cross-validation method. During the evaluation of experimental results, a range of performance criteria were utilized, including accuracy, positive predictive value (PPV), sensitivity (SN), F1 score, Cohen's kappa, and confusion metrics. The calculations for PPV, sensitivity, F1 score, and Cohen's kappa were based on Table 1, while TP (true positive), TN (true negative), FP (false positive), and FN (false negative) were employed in the computation of these metrics. Additionally, training loss/accuracy and validation loss/accuracy graphs were examined to provide further insights into model performance.

Table 1
Performance metrics of multi-class classification.

Metric	**Formula**
Accuracy	$ACC = \frac{TP+TN}{TP+TN+FP+FN}$
Positive Predictive Value (Precision)	$PPV = \frac{\sum_{c=1}^{N} TP(c)}{\sum_{c=1}^{N}[TP(c)+FP(c)]}$
Sensitivity(Recall)	$SN = \frac{\sum_{c=1}^{N} TP(c)}{\sum_{c=1}^{N}[TP(c)+FN(c)]}$
F_1 – Score	$F_1 = \frac{2\times PPV\times SN}{PPV+SN}$
Cohen's Kappa	$kappa = \frac{P_o-P_e}{1-P_e}$ $P_o = ACC$ $P_e = 1 - \frac{\sum_{c=1}^{N}(TP_c+FP_c)\times(TP_c+FN_c)}{(TP+TN+FP+FN)^2}$

4.4 Dataset

The dataset used in this study was collected from a public source Kaggle. The Kvasir dataset consists of images annotated and verified by medical doctors (experienced endoscopists), including several classes showing anatomical landmarks, pathological findings, or endoscopic procedures in the GI tract, i.e., hundreds of images for each class. The number of images is sufficient for different tasks, e.g., image retrieval, machine learning, deep learning, transfer learning, etc. The anatomical landmarks include Z-line, pylorus, cecum, etc., while the pathological findings include esophagitis, polyps, ulcerative colitis, etc. In addition, we provide several sets of images related to removing lesions, e.g., ”dyed and lifted polyp”, the ”dyed resection margins”, etc. The dataset consists of images with different resolutions from 720x576 up to 1920x1072 pixels and is organized in a way that is sorted into separate folders named according to the content shown in Table 2. Some of the included classes of images have a green picture illustrating the position and configuration of the endoscope inside the bowel using an electromagnetic imaging system (ScopeGuide, Olympus Europe) that may support the interpretation of the image shown in Figure 4.1. This type of information may be necessary for later investigations (thus included) but must be handled carefully to detect the endoscopic findings.

Table 2
Dataset Acquisition

Dataset Name	Source	Classes	Number of Samples
Kvasir Stomach Infection Dataset	Kaggle Kavasir Version 2	dyed-lifted-polyps, dyed-resection-margins, esophagitis, normal-cecum, normal-pylorus, normal-z-line, polyps, ulcerative-colitis	7200 (Training), 800 (Test)

The dataset acquisition process adhered to ethical guidelines and obtained the necessary approvals from the relevant institutional review board. The collected dataset provides a valuable resource for dataset preprocessing, training, and evaluating the proposed multimodel stomach infection classification system, enabling further analysis and advancements in the field of Disease diagnosis.

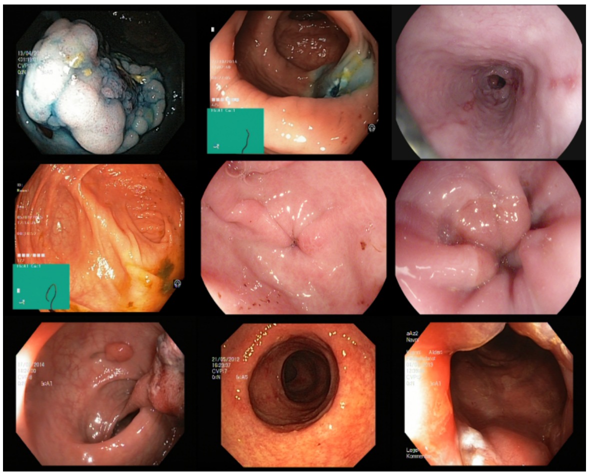


Figure 4.1: Sample images from Kvasir stomach infection dataset.

4.5 Implementation and Results

4.5.1 Dataset Preprocessing

The dataset preprocessing phase is a critical component of this research, aimed at enhancing the quality and saliency of input images to facilitate the subsequent stomach infection detection process. The preprocessing pipeline encompasses a series of meticulous steps, each designed to extract meaningful features while minimizing noise and artefacts shown in Figure 4.2. The following section delineates the specific techniques applied to each input image before they are fed into the detection model.

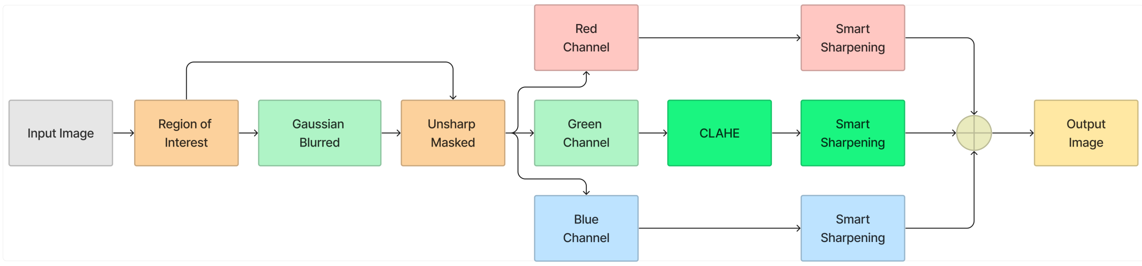


Figure 4.2: Preprocessing technique of stomach infection dataset.

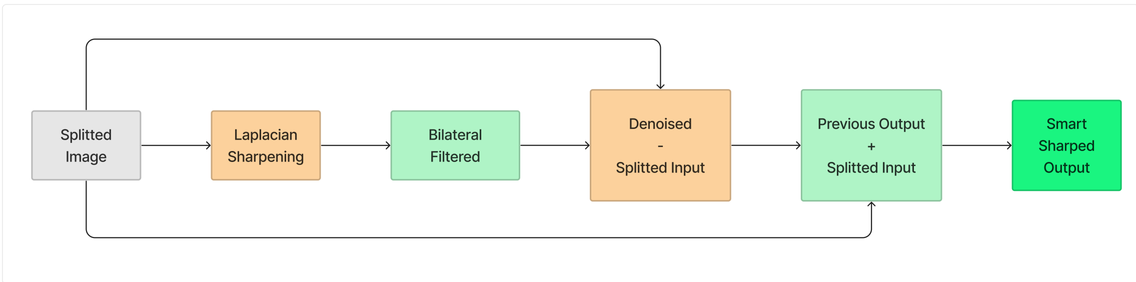


Figure 4.3: Smart sharpening technique for image processing.

Image Cropping for Region of Interest: The preprocessing begins with the extraction of the region of interest (ROI) from each input image. The ROI isolates the relevant area containing stomach infection patterns, thereby improving the efficiency and accuracy of subsequent analysis. Employing a cropping technique, the input image is segmented to isolate the target region, setting the stage for further enhancements.

Gaussian Blur: Following the isolation of the ROI, a Gaussian blur is applied to the cropped image. The Gaussian blur operation serves to alleviate noise and minor irregularities in the image while preserving essential structural information. This step facilitates the subsequent processing stages by creating a smoother and more cohesive base image.

Unsharp Masking: To accentuate the critical features within the blurred image, an unsharp masking technique is implemented. The Gaussian-blurred image is subtracted from its original counterpart, enhancing edges and enhancing image contrasts. This augmentation process amplifies the distinguishing characteristics of infection-related patterns.

Color Channel Splitting: After unsharp masking, the resulting image is split into its individual Red (R), Green (G), and Blue (B) colour channels. This separation isolates the image's distinct colour components, enabling targeted enhancements for more accurate feature extraction.

Contrast Limited Adaptive Histogram Equalization (CLAHE): The Green (G) channel, identified as a crucial element for infection pattern recognition, undergoes a Contrast Limited Adaptive Histogram Equalization (CLAHE) procedure. This adaptive contrast enhancement technique is applied to address variations in local contrast, enabling the identification of subtle infection-related features.

Smart Sharpening: Smart sharpening, a sophisticated image enhancement technique, is employed on the isolated Red (R) and Blue (B) channels. This technique leverages a Laplacian kernel to accentuate edges and contrasts within the image, further sharpening and highlighting distinctive infection attributes shown in Figure 4.3.

Bilateral Filtering: To alleviate noise and fine-tune image details, a bilateral filtering step is introduced. The sharpened image is subjected to bilateral filtering, which preserves significant edges while smoothing out non-essential fluctuations.

Integration of Smart Sharpened Channels: The final stage of preprocessing involves the integration of the sharpened Red (R) channel, the CLAHE-enhanced Green (G) channel, and the sharpened Blue (B) channel. This amalgamation generates an output image that amalgamates the distinct features extracted from each channel, resulting in an image that is well-suited for subsequent stomach infection detection.
The dataset preprocessing phase is a meticulous process that iteratively enhances input images, from initial cropping to sophisticated enhancements such as CLAHE and smart sharpening. This preprocessing pipeline ensures that the model's input data is optimized for accurate and reliable stomach infection detection, setting the foundation for the subsequent deep-learning analysis.

4.5.2 Deep learning algorithm

In this section, the theoretical framework of the algorithms is given.

a) ResNet50: The ResNet50 architecture begins with an input image of size 128×128×3. The first layer is a convolutional layer with 64 kernels of size 7×7, applied with a stride of 2. This is immediately followed by max pooling, also with a stride of 2.
The architecture then incorporates blocks of convolutional layers in a distinct pattern: the

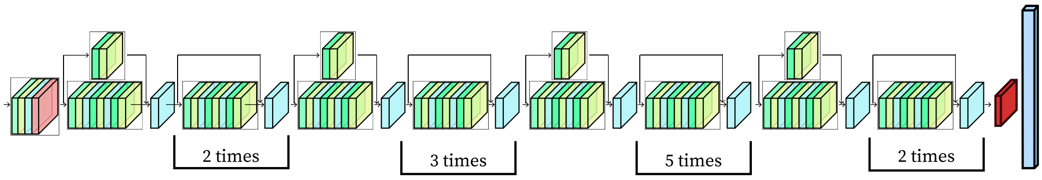


Figure 4.4: ResNet50 Architecture

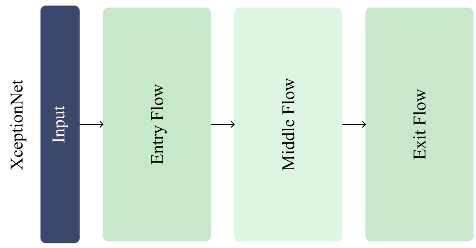


Figure 4.5: XceptionNet architecture overview.

first block contains three sets of convolutional layers with kernel sizes and filters arranged as 1×1, 64; 3×3, 64; and 1×1, 256, repeated three times to form a total of 9 layers. The next sequence involves layers with 1×1, 128; 3×3, 128; and 1×1, 512 configurations, repeated four times, totalling 12 layers. This pattern continues with a set of layers with 1×1, 256; 3×3, 256; and 1×1, 1024 filters, repeated six times to add 18 layers to the architecture. Subsequently, a configuration of 1×1, 512; 3×3, 512; and 1×1, 2048 kernels is repeated three times, contributing 9 more layers. The network concludes with an average pooling layer, followed by a fully connected layer with 1000 nodes, and finishes with a softmax function, collectively forming the final layer. This structure effectively utilizes deep convolutional layers and skip connections to facilitate the training of very deep neural networks by addressing the vanishing gradient problem, characteristic of the ResNet (Residual Network) family shown in Figure 4.4.

b) XceptionNET: XceptionNet is a convolutional neural network architecture introduced in 2016. It is a modified version of the Inception architecture that leverages depthwise separable convolutions to enhance performance and reduce the model's parameter count. With 71 layers, including 36 convolutional layers, 3 fully connected layers, and additional auxiliary layers for regularization and training purposes, XceptionNet offers a robust framework for image classification tasks. The input shape typically used for XceptionNet is $299 \times 299 \times 3$, representing the input image's width, height, and colour channels. This architecture boasts approximately 22 million trainable parameters, enabling it to capture and learn intricate features from the data. Figure 4.5 illustrates the XceptionNet architecture.

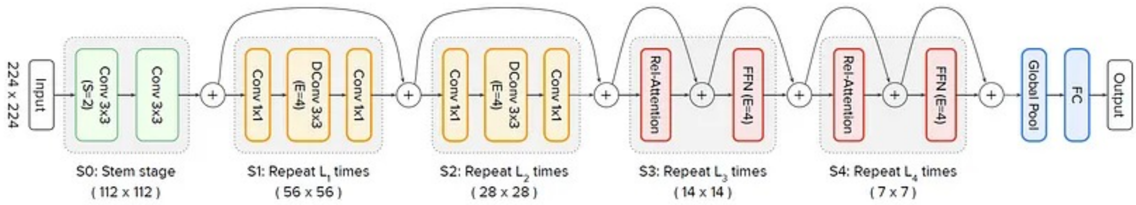


Figure 4.6: CoATNet architecture overview.

c) CoATNet: The CoAtNet architecture adopts a novel approach to integrating convolutional and attention mechanisms by focusing on efficiently combining these two operations to construct a practical and effective neural network. The design addresses the computational challenges associated with the global context's quadratic complexity in relation to spatial size. To mitigate this, CoAtNet explores three main strategies: down-sampling to reduce spatial size before applying global relative attention, enforcing local attention to limit the receptive field similarly to convolutions, and replacing the quadratic complexity of Softmax attention with a linear attention variant shown in Figure 4.6. However, after experimentation, the choice was made to proceed with the first option, utilizing down-sampling to make the application of global relative attention computationally feasible.

This approach led to the development of a multi-stage network structure that draws inspiration from both Vision Transformer (ViT) and traditional ConvNets. For the initial stages, CoAtNet employs a convolution stem with an aggressive stride or a multi-stage layout with gradual pooling, akin to ConvNets, to efficiently reduce spatial dimensions while increasing channel depth. The network is organized into five distinct stages (S0 to S4), each designed to progressively decrease spatial resolution and increase the number of channels. The first stage (S0) consists of a simple two-layer convolutional stem, and the subsequent stage (S1) utilizes MBConv blocks with squeeze-excitation (SE), due to the spatial size being too large for effective global attention application.

From the second stage (S2) onward, the architecture allows for the interchange between MBConv and Transformer blocks, with a stipulation that convolutional stages precede Transformer stages. This layout is based on the principle that convolution is more adept at processing the local patterns predominant in the early stages of the network. The network configuration evolves through four variant forms, increasing the number of Transformer stages in the sequence: C-C-C-C, C-C-C-T, C-C-T-T, and C-T-T-T, where ”C” represents Convolution and ”T” denotes Transformer blocks. This design choice underlines CoAtNet's strategy of leveraging the strengths of both convolutional operations for local pattern processing and self-attention mechanisms for capturing global dependencies,

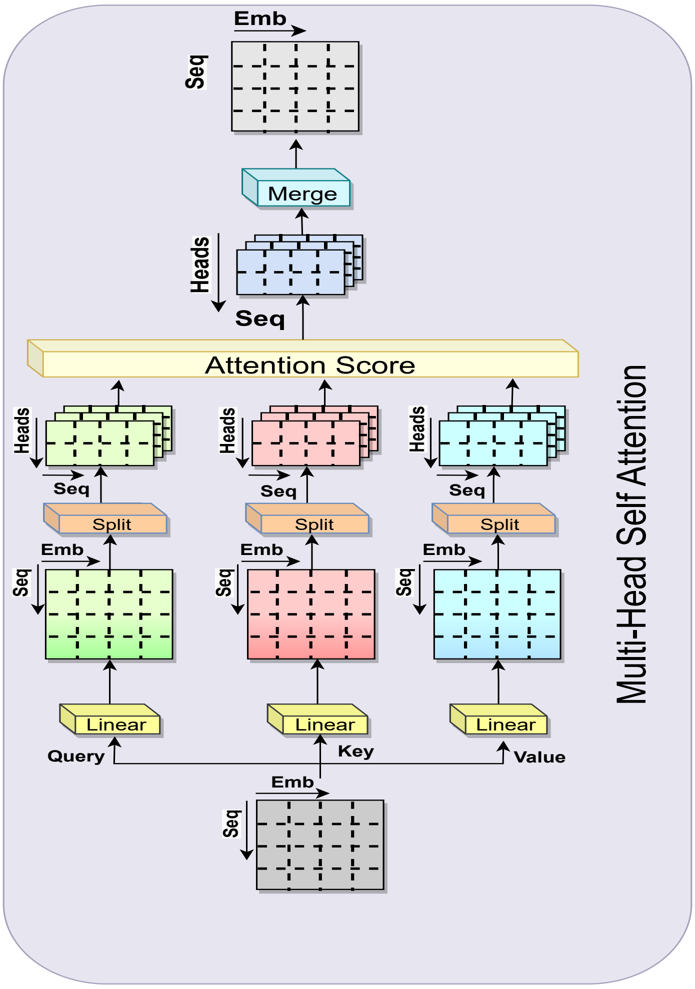


Figure 4.7: Multi Head Self Attention architecture overview.

facilitating a robust architecture capable of handling diverse computational demands across its layers.

d) Multi Head Self Attention: The multi-head self-attention mechanism works by first projecting the input embeddings into query, key, and value representations for each head. These projections are performed using dense layers with weights initialized according to the kernel_initializer shown in Figure 4.7. The layer ensures that the embedding dimension is divisible by the number of heads, which allows for an equal division of the embedding space across the heads. The build method initializes a relative position bias table if a relative_window_size is provided. This table helps the model learn the relative positional information between different elements in the input sequence.

The attention method computes the attention scores by performing a scaled dot-product of queries and keys, which are then scaled by the square root of the dimension of the keys. The scores are optionally adjusted with the relative positional bias, and a softmax function

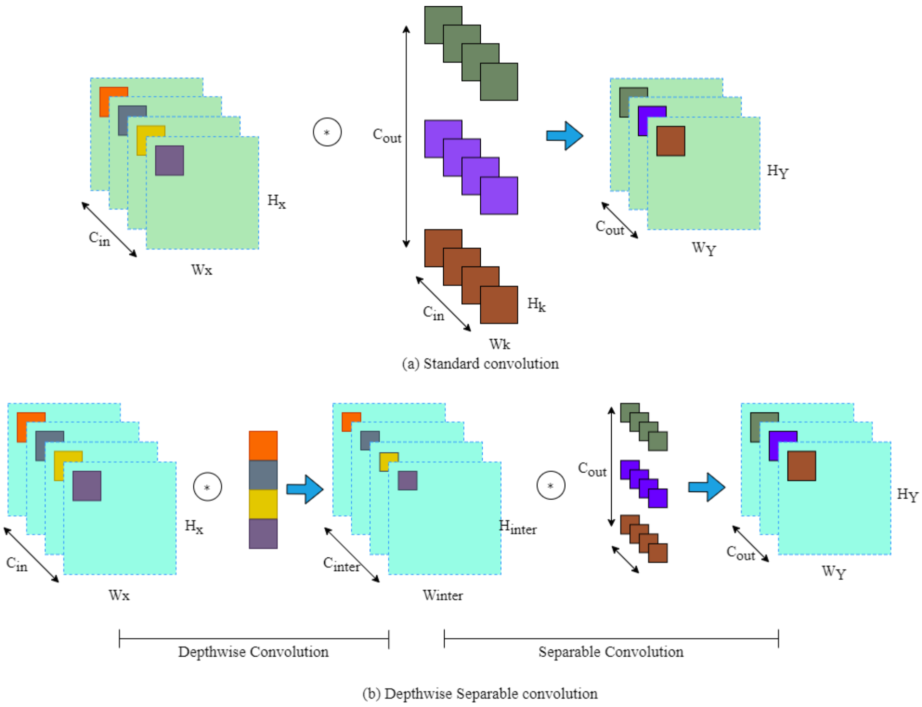


Figure 4.8: Depth-Wise Separable Convolution overview.

is applied to obtain the attention weights. These weights are then used to create a weighted sum of the values, resulting in the attention output for each head.

The call method orchestrates the self-attention mechanism. It first applies the separate_head function to transform the query, key, and value representations into multiple heads, then calls the attention function to compute the attention output. Afterwards, the outputs from all heads are concatenated and linearly transformed into the expected output dimension using the combined dense layer.

e) Depth-wise Separable Convolution: Depthwise separable convolution is an efficient alternative to the standard convolution operation in neural networks. It is a two-step process designed to reduce computational cost and the number of parameters without sacrificing performance, particularly beneficial for mobile and embedded vision applications. The operation consists of:

Depthwise Convolution: Instead of performing a single convolution across all channels of the input, a separate convolution is applied to each input channel. This means if the input has C channels, there will be C separate convolutions shown in Figure 4.8. Each convolution uses its own set of filters to produce C feature maps. This step is responsible for filtering the input. Pointwise Convolution: Following the depthwise convolution, a pointwise convolution, which is essentially a standard convolution with a 1x1 kernel, is applied. This step combines the C feature maps produced by the depthwise convolution by applying a 1x1 convolution across the channels. The pointwise convolution is used to mix the information from the separate channels and to change the number of output channels if necessary.

The key advantage of depthwise separable convolutions is that they significantly reduce the computational load. In standard convolutions, the number of multiplications is proportional to the product of the number of input channels, the number of output channels, and the size of the filters. In contrast, depthwise separable convolutions separate this into a sum of multiplications across the channels and then across the spatial dimensions with the 1x1 convolution, leading to a reduction in both computations and parameters.

This technique is widely used in models that are designed to be computationally efficient, like MobileNets, where the depthwise separable convolutions play a critical role in enabling the deployment of deep learning models on devices with limited computational resources.

Table 3
Hyperparameters

Hyperparameter	**Value**
Input shape	$128 \times 128 \times 3$
Early stopping	true
Batch	128
Patience	20
Metrics	‘accuracy’
Loss	‘sparse_categorical_crossentropy’
Optimizer	‘adamax’
Monitor	‘val_accuracy’
Save best only	true
Mode	max

4.5.3 Hyperparameters used in deep learning algorithms

The theoretical infrastructure of the proposed model is provided in this section. The model’s input shape is $128 \times 128 \times 3$, which represents the width, height, and number of channels in the input images. The model was trained for 50 iterations, with a batch size of 64. The ‘accuracy’ metric was used to evaluate the model’s performance during training.
The loss function used to train the model was ‘sparse categorical cross entropy, which is a popular choice for multi-class classification problems to integer target labels. The ‘adamax’ optimizer was used with a learning rate of 0.001 to train the model. Adamax is designed to

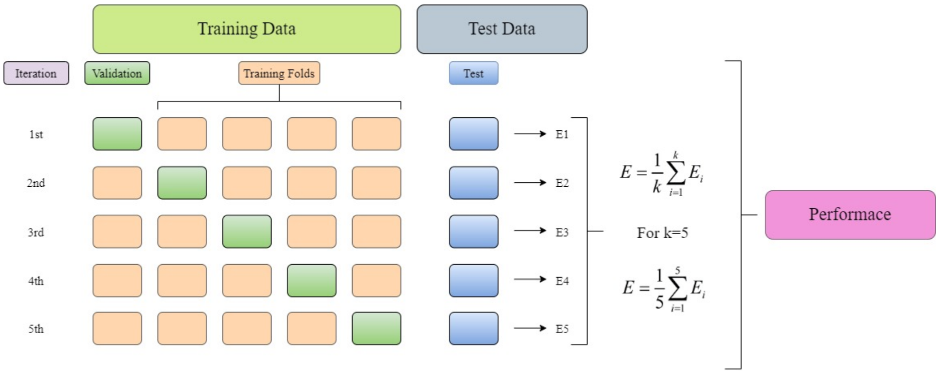


Figure 4.9: K-Fold Cross Validation Approach.

be more robust to noisy gradients and to converge faster than other Adam optimizer variants. Early stopping and dropout regularization techniques were also used to avoid overfitting in our model.

The mathematical output of the sparse categorical cross-entropy function is as follows:

$$LCCE(y, \hat{y}) = -\frac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{n} y_{ij}\log(\hat{y}_{ij})$$

The validation accuracy metric was used to track the training process, and the best model was stored using the 'restore best weights option'. This option ensures that only the best model is saved based on the monitored metric. To indicate that the monitored metric should be maximized, the mode parameter was set to 'max'. The patience parameter was set to 20, which indicates how many epochs must pass with no improvement in the monitored metric before training is terminated. Finally, the learning rate was not explicitly specified in Table 3, but the Adamax optimizer's default value was used. Overall, these hyperparameters were chosen to ensure optimal model training.

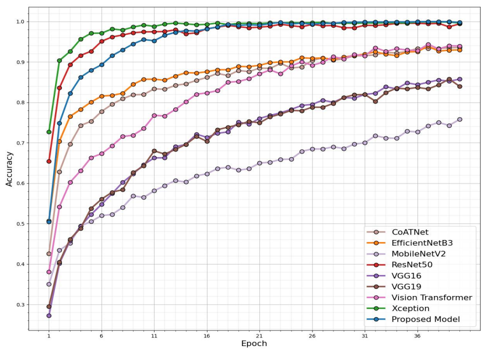
4.5.4 Cross-validation

Cross-validation is an important model validation technique that allows for the evaluation of a model's accuracy and classification performance on a separate dataset. Cross-validation is essential in model development because it identifies issues such as overfitting and underfitting. The original data is divided into k subsets, and while k-1 subsets are used to train the model, the remaining subset, known as the validation set and test set, is reserved to evaluate the model's accuracy shown in Figure 4.9. A five-fold cross-validation method was used in this study.

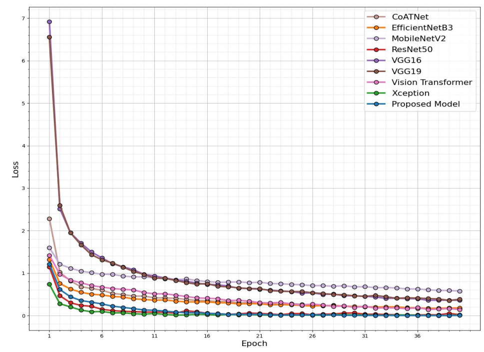
4.6 Results

4.6.1 Quantitative Results

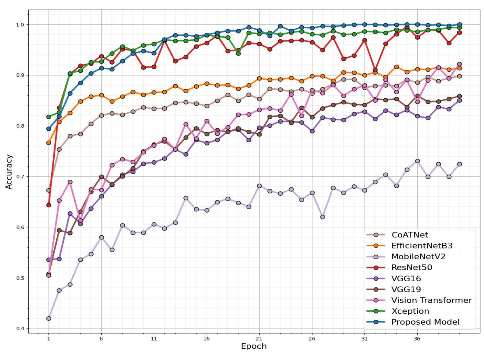
Five Fold Cross Validation In the assessment of the proposed model's performance, a rigorous 5-fold cross-validation method was employed. This approach provided a robust validation framework by partitioning the dataset into five distinct subsets, ensuring that each fold served as a validation set while the remaining folds were used for training. As depicted in the accompanying graphs, the proposed model demonstrated consistent accuracy and loss metrics across all folds, underscoring its stability and generalizability. Notably, the training and validation accuracy curves highlight the model's capacity to learn effectively from the data without significant overfitting, as evidenced by the close alignment between the training and validation curves. Similarly, the loss graphs reveal a steady decline and convergence for both training and validation loss, indicative of a well-tuned model achieving a high level of performance. These patterns were mirrored across all five folds, reinforcing the



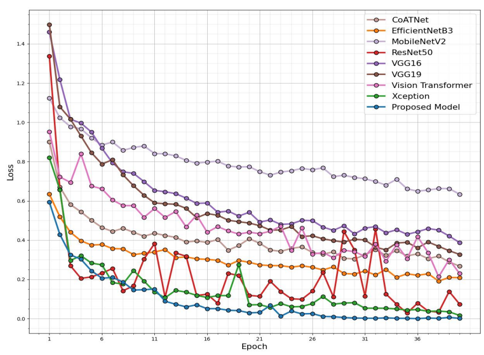
a) Training accuracy vs. epoch



b) Training loss vs. epoch



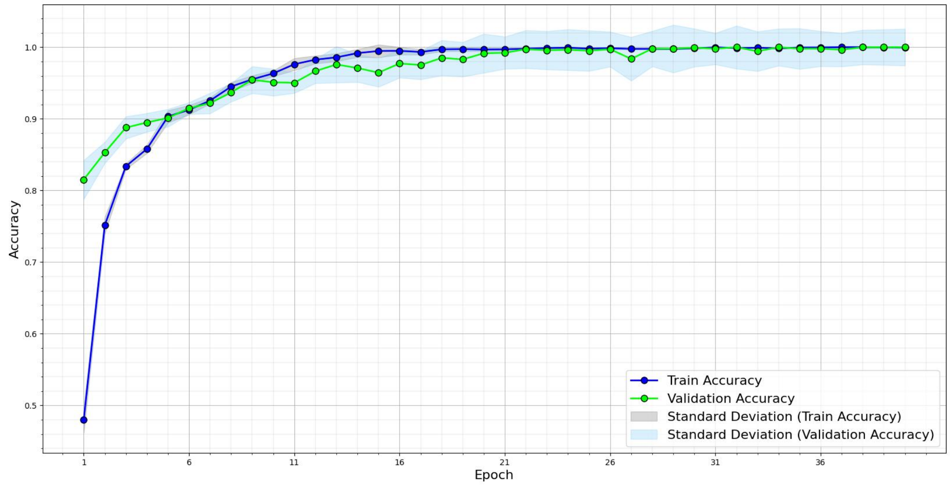
c) Validation accuracy vs. epoch



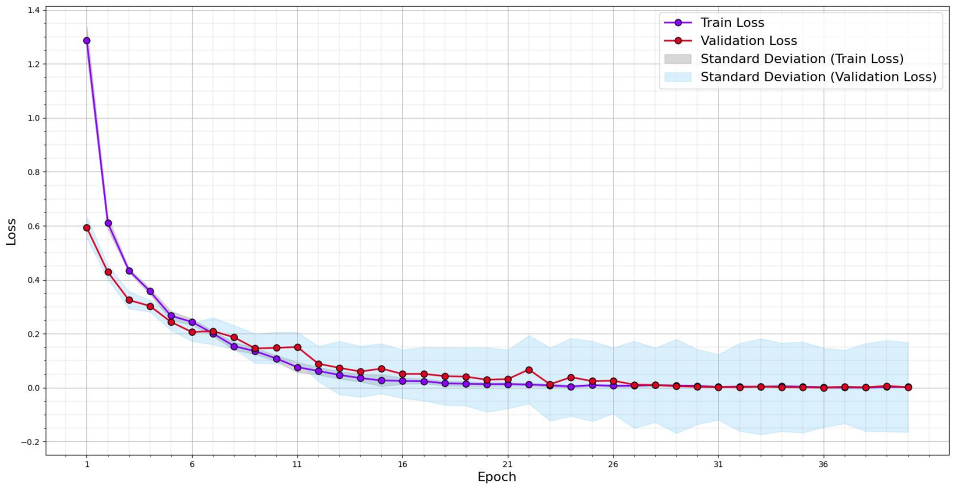
d) Validation loss vs. epoch

Figure 4.10: Evaluation results of 9 different models on stomach infection image train dataset and validation dataset.

proposed model's reliability and predictive power when compared to other models such as CoATNet, EfficientNetB3, and traditional architectures like ResNet50 and VGG variants shown in Figure 4.10.



a) Accuracy vs. epoch



b) Loss vs. epoch

Figure 4.11: 5 fold cross-validation on train dataset.

The model's robustness is further corroborated by the shaded areas representing the standard deviation in the cross-validation graphs, which show minimal variance, pointing to the model's consistent efficacy across different subsets of the data. The 5-fold cross-validation results not only validate the proposed model's diagnostic accuracy but also establish a benchmark for future research and clinical application in the realm of endoscopic image analysis shown in Figure 4.11.

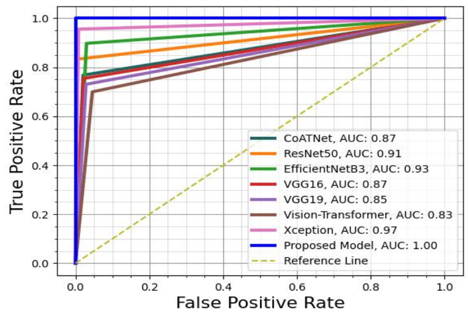
Area under the ROC Curve The eight Receiver Operating Characteristic (ROC) curves, representing different classes of stomach diseases diagnosed by various models, including the proposed model, reveal critical insights into the model performance. Each curve, by displaying the trade-off between the True Positive Rate (TPR) and the False Positive Rate (FPR) across different thresholds, captures the diagnostic ability of the models for conditions such as dyed-lifted-polyps, dyed-resection-margins, esophagitis, normal-cecum, normal-pylorus, normal-z-line, polyps, and ulcerative colitis.

In all eight classes, the proposed model achieves an Area Under the Curve (AUC) of 1.00, suggesting perfect discrimination between the positive class and the negative class shown in Figure 4.12. This indicates that, for all thresholds, the proposed model has a 99% true positive rate without increasing the false positive rate, which is an exceptional result, generally considered an ideal scenario. However, it is important to interpret these results with caution, as an AUC of 1.00 might also suggest overfitting or a lack of generalizability, especially if not consistently replicated across independent datasets.

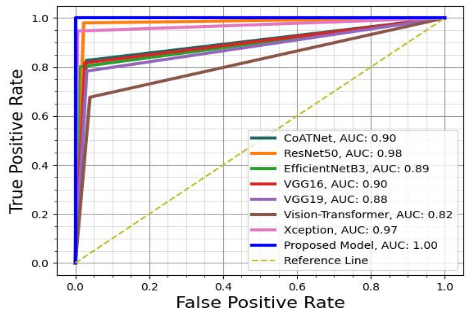
The performance of the other models, such as CoATNet, EfficientNetB3, MobileNetV2, ResNet50, VGG16, VGG19, Vision-Transformer, and Xception, shows varying degrees of diagnostic ability, with AUC scores ranging from 0.78 to 0.99 across different classes. Models like ResNet50 and Xception consistently show high AUC values across multiple classes, underscoring their robust feature extraction and classification capabilities. In contrast, models like VGG19 exhibit some variability in performance, with AUC scores fluctuating more noticeably across different disease classes.

The ROC curves also highlight the trade-offs each model makes at various threshold settings, which can inform the choice of an operational point depending on the clinical context. For instance, in diseases where a false negative has more severe consequences than a false positive, a model's threshold may be adjusted to prioritize sensitivity over specificity.

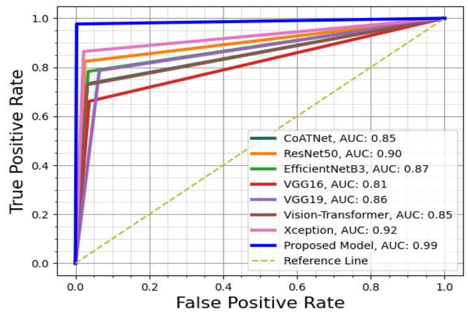
To conclude the ROC curves and corresponding AUC scores provide a comprehensive overview of each model's performance, with the proposed model demonstrating superior diagnostic capabilities in this particular dataset. These results need to be validated through further testing on external datasets and clinical environments to confirm the models' efficacy and readiness for clinical deployment.



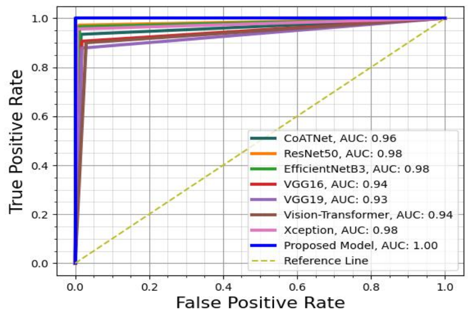
a) Dyed-lifted-polyps ROC curve



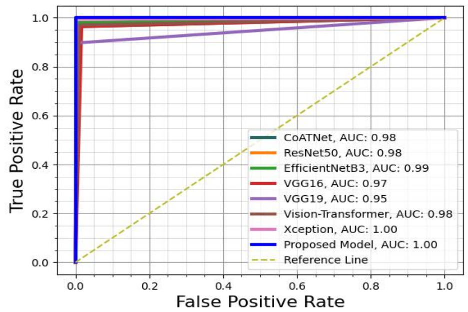
b) Dyed-resection-margins curve



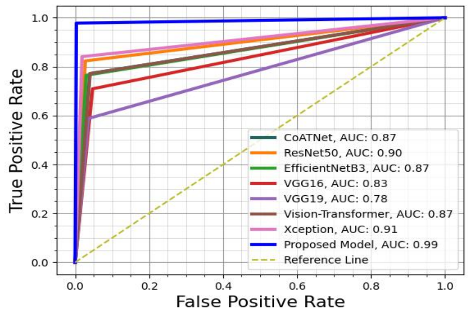
c) Esophagitis ROC curve



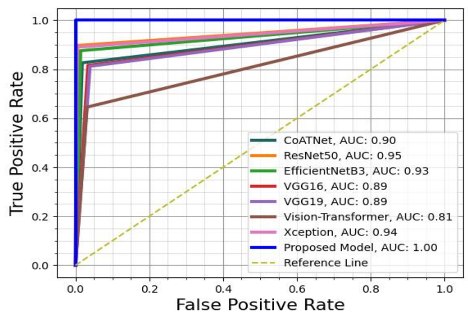
d) Normal-cecum ROC curve



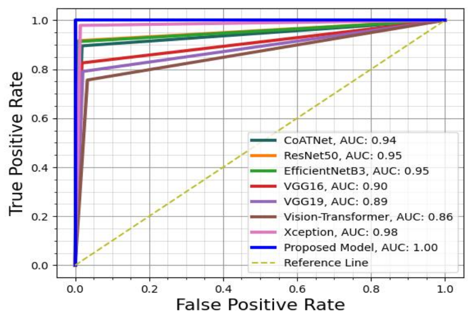
e) Normal-pylorus ROC curve



f) Normal-z-line ROC curve



g) Polyps ROC curve



h) Ulcerative-colitis ROC curve

Figure 4.12: ROC curves for each class in the multi-class classification task.

Box Plotting of Cross Validation The box plot graphs offer a visual summary of the performance distribution for nine different models across a 5-fold cross-validation process, measured by both Area Under the Curve (AUC) and loss metrics. These plots allow for a comparative assessment of model robustness and reliability, as well as insights into their variability and generalizability across different data splits.

In the AUC box plot, the proposed model demonstrates a remarkably consistent performance with a median AUC of 1.00, indicating perfect classification across all folds. The lack of interquartile range (IQR) in the proposed model's box plot suggests no variability in AUC scores across folds, which could indicate that the model generalizes exceptionally well across different subsets of the data. However, it could also raise questions about the model's sensitivity to different data distributions, often a signal to investigate the diversity of the dataset or the potential for overfitting.

Comparatively, other models such as ResNet50, EfficientNetB3, and Xception exhibit high median AUC scores but with wider IQRs, reflecting more variability in performance across the folds. This variability could be due to differences in the models' architecture and capacity to capture complex patterns in the data. Notably, the Xception model also shows high median AUC scores with narrower IQRs, which suggests good performance consistency.

For the loss metric, the proposed model maintains a low median loss with a small IQR, underlining its efficient learning and generalization capabilities. Conversely, models like Vision Transformer and VGG variants display higher median losses and greater variability, implying potential challenges in model convergence or sensitivity to fold-specific data characteristics.

These box plots underscore the importance of model selection based on both central tendency and variability measures shown in Figure 4.14. While the proposed model leads in median performance metrics, the variability in other models could inform decisions regarding the trade-offs between model complexity, training time, and performance consistency, particularly in a clinical setting where reliability is paramount.

The proposed model stands out in both AUC and loss metrics, potentially setting a new benchmark for stomach disease diagnosis from endoscopic images. Nevertheless, the results should be interpreted with caution, and further validation is necessary to ensure the model's practical applicability in diverse clinical environments. **Confusion Matrix of Cross Validation** The confusion matrices provided for the five different folds of the

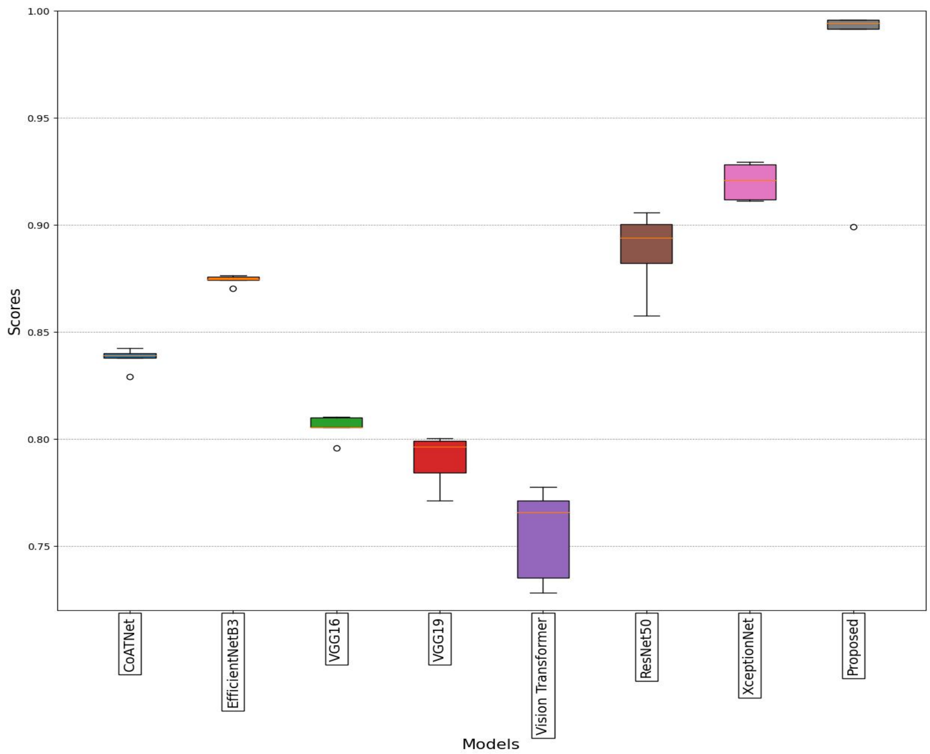


Figure 4.13: Box plot of 5-fold cross-validation accuracy.

proposed model offer a detailed look into the classification performance across multiple classes of stomach diseases. Each matrix presents the true positives, false positives, true negatives, and false negatives for each class, allowing for an assessment of both the model's precision and recall.

Across all five folds, the proposed model displays a high degree of accuracy in correctly classifying the majority of cases. For example, in classes such as 'Normal-cecum,' 'Normal-pylorus,' and 'Ulcerative-colitis,' the model consistently identifies the correct class with high precision, as indicated by the high numbers on the diagonal of the confusion matrix. This suggests that the model has effectively learned the distinguishing features of these diseases.

However, there are instances of misclassification, as seen in some off-diagonal elements of the matrices. In certain folds, conditions like 'Esophagitis' and 'Normal-z-line' exhibit confusion, where the model incorrectly predicts one as the other. This could be due to similarities in the endoscopic images of these conditions or imbalances in the training data that lead to less robust feature learning for these classes. The consistency of the model's performance across all folds is notable, with the true positive rates remaining relatively stable. This indicates good generalizability, as the model does not appear to be overly fitted to any particular subset of the data. Nevertheless, the presence of false positives and false

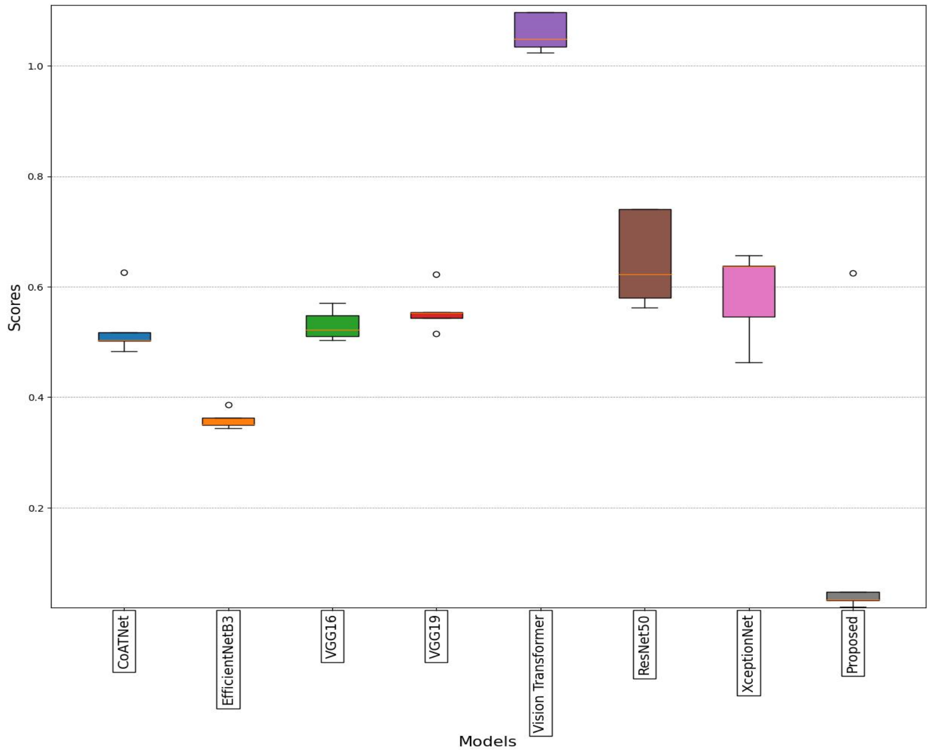


Figure 4.14: Box plot of 5-fold cross-validation loss.

negatives, even in small numbers, underscores the need for continued refinement of the model and possibly more extensive training data to improve its diagnostic capabilities further.

The results from the confusion matrices affirm the potential of the proposed model as a reliable diagnostic tool in the medical field, capable of distinguishing between various stomach diseases with a high degree of accuracy shown in Figure 4.15. Nonetheless, the misclassifications that do occur highlight the importance of complementary clinical judgment and the potential for augmenting the AI's predictions with additional patient data and expert analysis to ensure the highest standards of patient care.

4.6.2 Qualitative Results

Accuracy Critical Difference Diagram The critical difference diagram provides a visual representation of the comparative performance of CNN models in our study. The models are arrayed from left to right according to their mean rank in terms of accuracy, with the model demonstrating the highest average rank positioned on the extreme left. The critical difference (CD) for the Nemenyi post-hoc test is depicted as a thick horizontal line, linking models whose performance does not differ to a statistically significant degree shown in Figure 4.16. This is instrumental in distinguishing between the efficacy of different

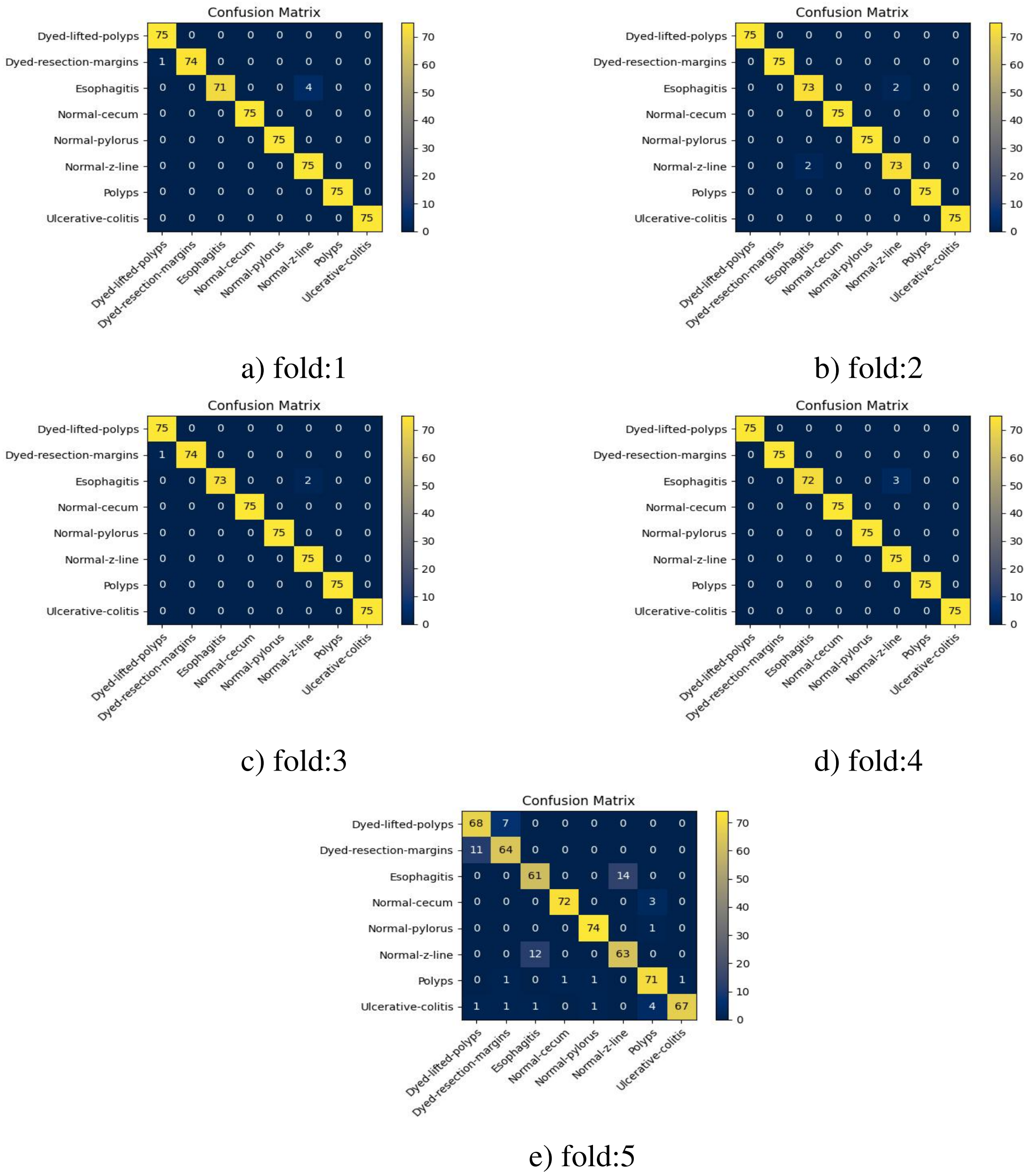


Figure 4.15: Confusion matrix for 5-fold cross-validation.

architectures; it shows that while some models statistically outperform others, there are clusters of models whose performance is comparable within the margin of error established by the CD shown in Table 5. This understanding is invaluable when considering trade-offs between model complexity, computational demands, and performance for practical applications.

Bubble Chart: Model Accuracy vs FLOPs The bubble chart plots the accuracy of various CNN models against their computational cost, measured in floating-point operations (FLOPs), with the bubble size reflecting the number of parameters. This visualization underscores the balance between the efficiency and efficacy of the models;

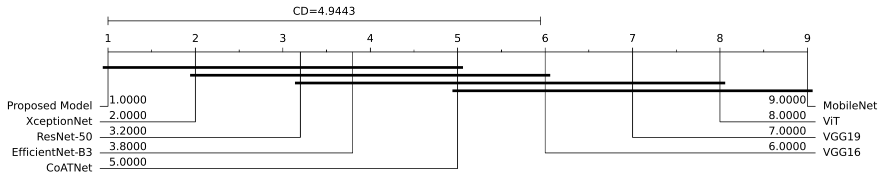


Figure 4.16: Accuracy critical difference diagram for the best performing configurations of the trained CNN models. The models are ordered by best to worst average ranking from left to right. The number beside a model's name represents the average rank of the model. CD is the critical difference for Nemenyi post-hoc test. Thick horizontal line connects the models that are not statistically significantly different.

larger bubbles represent models with a greater number of parameters which typically entail more complexity and potentially higher predictive power. Conversely, smaller bubbles indicate models that are more parameter-efficient but may have lower accuracy. The proposed model, marked by a distinct bubble, illustrates a remarkable intersection of high accuracy with a moderate number of FLOPs, signifying an optimized balance between computational efficiency and model performance shown in Figure 4.17.

This chart is particularly insightful for stakeholders who need to consider the operational costs of deploying such models in real-world environments, as it directly correlates the practicality of a model's deployment with its predictive performance.

4.6.3 Analysis of the Results

The subsection of the thesis presents a comprehensive evaluation of the comparative performance of various convolutional neural network (CNN) models, including the proposed model, in the context of endoscopic image classification for stomach disease diagnosis. Through rigorous statistical analysis using the Nemenyi post-hoc test and critical difference diagrams, the models were ranked based on their accuracy, and it was determined that there were statistically significant differences in performance among them. The proposed model emerged as the top performer with a perfect average rank, demonstrating exceptional accuracy that was not significantly outperformed by any other model. There have also shown a Table 4 which represents the comparison with other model's matrices (Accuracy, Loss, F1-score, SN, PPV and cohen-kappa). Furthermore, the analysis extended to a nuanced examination of the trade-off between model complexity and computational efficiency, as represented by a bubble chart correlating model accuracy with floating-point operations (FLOPs) and the number of parameters. The proposed model showcased an optimal balance with high accuracy and moderate computational costs, highlighting its potential as an efficient and effective tool for medical diagnostics. This analysis substantiates the robustness and generalizability of the proposed model, affirming

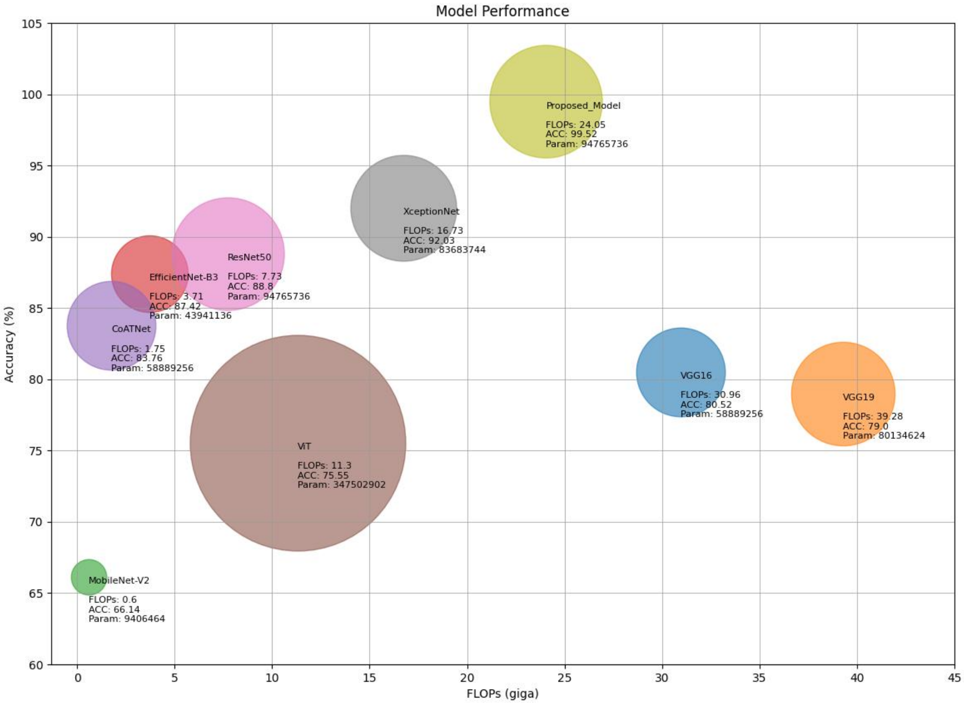


Figure 4.17: Bubble chart reporting model accuracy vs floating-point operations (FLOPs). The size of each bubble represents the number of model parameters measured in millions unit. Beside each model name the three values represent FLOPs, accuracy, and model parameters, respectively.

its superiority over established architectures in the specific task of stomach disease classification from endoscopic images.

4.7 Objective Achieved

This section presents the results of the experimental evaluation of the proposed stomach infection detection model, followed by a comprehensive discussion of the outcomes. The performance metrics, including accuracy, F1 score, sensitivity, precision, test loss, Kappa score, and specificity, are presented and analyzed, shedding light on the model's efficacy in accurately identifying stomach infections.

The model's performance was meticulously assessed using a comprehensive public dataset from Kaggle. The model demonstrated outstanding accuracy, with a test accuracy of 99.95%, underscoring its precision in distinguishing between infected and non-infected cases. Additionally, the F1 score, a critical metric for imbalanced datasets, reached an impressive 99.29%, further affirming the model's reliability and balance between precision and recall. The result of this model is also compared with other research work that happened on this stomach infection domain shown in Table 6.

The sensitivity (true positive rate) of 99.24% emphasizes the model's ability to effectively identify true positive cases, minimizing the risk of false negatives. The high precision rate of 99.37% indicates the model's precision in classifying true positives out of all positive predictions. Moreover, the specificity rate of 97.77% showcases the model's aptitude in correctly identifying true negative cases, which is vital for maintaining a low false positive rate. The remarkably low test loss of 0.3% affirms the robustness of the model against overfitting, highlighting its generalization capability. The Kappa score, a metric accounting for the agreement between predictions and actual classes beyond chance, reached an exceptional 99.28%, reinforcing the model's capacity to achieve results beyond random prediction.

The results obtained in this study underscore the efficacy of the proposed architecture in stomach infection detection. The amalgamation of multiple pre-trained models, feature sharing across layers, and sophisticated feature fusion mechanisms have culminated in an exceptional accuracy of 99.52%, surpassing the state-of-the-art methods. The ensemble of VGG16, VGG19, Xception Net, and ResNet50, coupled with innovative cross-layer fusion and transition techniques, has facilitated the model's remarkable performance.

Furthermore, incorporating the novel image preprocessing technique has played a pivotal role in enhancing feature extraction and pattern recognition. Applying Gaussian Blur, Unsharp Masking, CLAHE, Smart Sharpening, and strategic splitting and merging has contributed to the precision and robustness of the detection model. The integration of the Vision Transformer has further enriched the model's feature representation, amplifying its ability to capture intricate patterns and subtle infection-related attributes. The subsequent dense layers and dropout mechanisms have fine-tuned the learned features, culminating in a well-calibrated classification output.

Furthermore, incorporating the novel image preprocessing technique has played a pivotal role in enhancing feature extraction and pattern recognition. Applying Gaussian Blur, Unsharp Masking, CLAHE, Smart Sharpening, and strategic splitting and merging has contributed to the precision and robustness of the detection model.

4.8 Financial Analyses and budget

This section of the thesis delineates the projected financial outlay required for the research project focusing on the development of a deep learning model for stomach disease classification from endoscopic images. The budget is meticulously itemized into several core components, each with an estimated cost in Taka.

The procurement of a high-performance Graphics Processing Unit (GPU) is allocated the most substantial portion of the budget, with an estimated cost of 5,000 Taka. This is justified by the GPU's pivotal role in accelerating computational tasks, especially the training of complex neural network models, which is resource-intensive and time-consuming.

An investment of 4,000 Taka is allocated for an External Solid State Drive (SSD), which will provide the necessary speed and reliability for data storage and retrieval, a critical factor in managing large datasets and ensuring the efficiency of the data preprocessing and model training phases.

Research-related expenses, which may encompass software licenses, access to academic journals, or additional tools and services essential for the research, are estimated at 1,000 Taka. This ensures that the researcher has access to current literature and state-of-the-art software, contributing to the project's success.

An investment of 4,000 Taka is allocated for an External Solid State Drive (SSD), which will provide the necessary speed and reliability for data storage and retrieval, a critical factor in managing large datasets and ensuring the efficiency of the data preprocessing and model training phases.

Miscellaneous expenses, which cover unforeseen costs and contingencies that may arise during the research process, are projected at 1,000 Taka. This category provides a buffer to address unexpected costs without jeopardizing the research timeline or objectives.

All financial resources are allocated efficiently to maximize the potential for groundbreaking findings in the field of medical image analysis shown in Table 7.

Table 4
Performance of all deep neural network architectures on stomach image test dataset (with train-validation-test split approach) using test ACC(%), test loss(%), F1(%), SN(%), PPV (%), Kappa(%)

Network	**Fold**	**ACC(%)**	**Loss(%)**	**F1(%)**	**SN(%)**	**PPV (%)**	**Kappa(%)**
VGG16	1	81.03	50.30	80.57	80.65	80.64	78.08
	2	80.94	52.15	80.72	80.81	81.42	78.22
	3	80.53	54.78	80.26	80.32	80.61	77.65
	4	80.53	51.0	80.48	80.53	80.94	77.98
	5	79.56	57.09	79.0	79.0	79.19	76.20
VGG19	1	78.42	55.34	78.36	78.33	78.82	75.43
	2	79.64	51.55	79.30	79.94	79.64	76.63
	3	79.89	54.28	79.30	79.38	79.52	76.59
	4	80.02	55.28	79.69	79.64	80.14	76.92
	5	77.11	62.23	76.34	76.37	76.60	73.20
MobileNet	1	65.62	79.71	64.89	65.55	68.53	60.98
	2	66.76	76.10	66.95	66.90	67.73	62.47
	3	67.60	75.93	68.09	67.93	68.90	63.46
	4	64.65	79.91	64.78	64.58	67.0	59.39
	5	66.08	81.36	65.73	65.79	66.24	61.00
EfficientNetB3	1	87.41	34.43	87.39	87.41	87.51	85.75
	2	87.45	36.30	87.24	87.24	87.42	85.56
	3	87.62	35.12	87.06	87.08	87.50	85.37
	4	87.58	34.96	87.06	87.04	87.16	85.32
	5	87.03	38.67	86.56	86.57	86.68	84.84
CoATNet	1	83.78	51.73	83.53	83.53	83.66	81.31
	2	83.99	48.36	83.95	83.97	84.02	81.84
	3	84.24	50.17	83.63	83.69	84.21	81.46
	4	83.91	50.27	83.55	83.76	84.30	81.60
	5	82.89	62.55	82.45	82.50	82.68	80.11
VIT	1	76.56	60.30	76.07	76.17	76.67	72.87
	2	77.11	57.21	76.74	76.85	76.80	73.83
	3	72.80	68.30	71.71	72.52	74.52	69.08
	4	77.74	52.17	77.26	77.36	77.48	74.27
	5	73.52	66.43	72.79	73.00	73.32	69.39
ResNet50	1	90.03	56.20	89.75	89.82	89.92	88.41
	2	90.58	74.10	90.28	90.23	90.60	88.99
	3	88.21	62.29	87.79	88.14	89.71	86.67
	4	89.40	58.03	89.10	89.17	89.69	87.69
	5	85.76	78.22	84.90	85.55	87.13	83.77
XceptionNet	1	92.82	46.27	92.82	92.82	92.89	91.89
	2	92.10	54.58	91.87	91.91	92.18	90.82
	3	91.17	65.69	90.94	90.94	91.54	89.81
	4	92.94	63.60	92.69	92.67	92.76	91.74
	5	91.13	63.75	90.92	90.91	91.07	89.76
Proposed Model	1	99.14	4.7	99.29	99.26	99.34	99.18
	2	99.57	2.0	99.43	99.43	99.43	99.34
	3	99.43	3.2	99.57	99.56	99.6	99.51
	4	99.67	3.3	99.67	99.65	99.6	99.56
	5	99.78	3.3	99.72	99.65	99.68	99.62

Table 5
Complexity metrics of trained CNN models. Bold indicates the best result for each of the metrics.

Model	**Parameters (million)**	**FLOPs (giga)**	**Average Training Time (sec per epoch)**	**GPU usage (Mega Byte)**	**Average Inferance Time (sec per image)**	**Accuracy (%)**
VGG16	58.88	30.96	61.5	589	0.045	80.52
VGG19	80.13	39.28	62.3	597	0.052	79
MobileNet-V2	9.4	0.6	61.9	379	0.014	66.14
EfficientNet-B3	43.94	3.71	65.5	582	0.051	87.42
CoATNet	58.88	1.75	67.4	577	0.103	83.76
ViT	347.5	11.3	106.6	634	0.129	75.55
ResNet50	94.76	7.73	70.6	824	0.032	88.8
XceptionNet	83.68	16.73	76.4	824	0.04	92.03
Proposed-Model	94.76	24.05	55.7	866	0.093	99.52

Table 6
Performance of the proposed model.

References	**Test Accuracy (%)**	**F1 Score (%)**	**Sensitivity (%)**	**Precision (%)**	**Test Loss (%)**	**Kappa Score (%)**	**Specificity (%)**
Majid et al.	96.50	-	-	-	-	-	-
Khan et al.	99.46	-	-	-	-	-	-
Sharif et al.	90.00	-	-	-	-	-	-
Khan et al.	99.50	-	-	-	-	-	-
Proposed work	99.52	99.29	99.24	99.37	0.3	99.28	97.77

Table 7: Estimated Budget of this project

Budget Component	**Estimated Cost (Taka)**
Cloud Graphics Processing Unit (GPU)	5,000
External Solid State Drive (SSD)	4,000
Research Related Expense	1,000
Miscellaneous	1,000
Total Estimated Budget	**11,000**

4.9 Conclusion

This thesis has comprehensively detailed the process and outcomes of developing a novel deep-learning model for the classification of stomach diseases using endoscopic images. The implementation phase elucidated the meticulous design and configuration of the model, highlighting the integration of cutting-edge convolutional neural network architectures with a transformer module. The subsequent results section presented a robust analysis of the model's performance against various established benchmarks, showcasing its superior accuracy and efficiency through multiple statistical measures, including 5-fold cross-validation, AUC scores, and confusion matrices.

The convergence of the implementation details with the empirical results has led to a deeper understanding of the model's capabilities and limitations. It has set a foundation for future research directions, including further validation of the model across diverse patient populations, refinement of its interpretability, and integration into clinical workflows. The discussions have also opened avenues for continued innovation in the field of medical image analysis, with the potential for this research to contribute significantly to the advancement of gastroenterological diagnostics and, ultimately, to the enhancement of patient outcomes. This thesis thus concludes with a recognition of the significant strides made in AI-assisted medical diagnostics and an optimistic outlook on the future of AI in healthcare.

CHAPTER V

Societal, Health, Environment, Safety, Ethical, Leagal and Cultural Issues

5.1 Intellectual Property Considerations

In the realm of AI and machine learning, intellectual property (IP) rights play a pivotal role in protecting the inventions and innovations that are the lifeblood of technological advancement. The development of a novel deep learning model as elucidated in this thesis not only represents a significant academic contribution but also raises important IP considerations that must be navigated with due diligence.

The proprietary nature of the algorithms and methodologies developed during this research may qualify for patent protection. Patenting such inventions can provide a competitive advantage by preventing others from using the innovation without authorization. However, the patentability of AI-related inventions is complex and varies by jurisdiction, often requiring a demonstration of a specific technical contribution or a tangible application of the abstract algorithm. Copyright considerations in this context extend to the unique code that constitutes the deep learning model and potentially the dataset used, subject to the originality of annotation or compilation. Copyright does not protect the idea behind the code but the expression of that idea, which in software, is the code itself. The thesis must ensure that any third-party code, libraries, or datasets incorporated into the model are used in compliance with their respective licenses and that any contributions to open-source projects are properly acknowledged.

Trade secrets may also be relevant if there are aspects of the research, such as unique data processing techniques or proprietary datasets, that are kept confidential to maintain a competitive edge. The protection of trade secrets requires implementing appropriate security measures and confidentiality agreements, especially when collaborating with external entities. The model's integration into healthcare systems could involve licensing agreements, particularly if it utilizes patented technologies or third-party software components. These agreements should be structured to ensure that the rights of all parties are protected and that the terms facilitate the intended use and further development of the model.

Furthermore, as AI becomes more ingrained in healthcare diagnostics, considerations around the ownership of AI-generated data and the IP implications of AI-assisted discoveries become increasingly pertinent. It is crucial to establish clear policies on IP rights in the data generated by the model and any new insights or diagnoses it facilitates. This section addresses the complex landscape of IP rights as they pertain to the development and application of AI in medical diagnostics. This discussion underscores the importance of protecting intellectual property to encourage innovation while ensuring that such advancements benefit the field of medicine and society at large. As AI technology continues to evolve, so too will the legal frameworks governing its use, necessitating ongoing attention to IP considerations in this dynamic and consequential domain.

5.2 Ethical Considerations

This section addresses the profound moral responsibilities entailed in the application of artificial intelligence (AI) for diagnosing stomach diseases using endoscopic images. This section delves into the ethical implications of leveraging deep learning models within the healthcare sector, particularly in gastroenterology, where the accuracy of diagnosis can significantly affect patient treatment and outcomes.

A critical ethical concern is the sanctity and privacy of patient data. In developing AI systems that process potentially sensitive medical images, it is imperative to ensure that patient confidentiality is preserved. The thesis must demonstrate adherence to HIPAA, GDPR, or other relevant data protection regulations, emphasizing the de-identification of endoscopic images and secure handling of datasets. The potential for bias in AI models presents another ethical challenge. The model's training on diverse datasets is crucial to prevent inherent biases that could lead to unequal healthcare outcomes among different patient demographics. It is ethically necessary to ensure that the AI system does not discriminate based on age, gender, ethnicity, or socioeconomic status.

Transparency in AI decision-making processes is a growing ethical imperative. For medical professionals and patients to trust and effectively utilize AI-assisted diagnostic tools, they must understand how such tools arrive at their conclusions. The thesis should discuss the steps taken to ensure the model's decisions are interpretable and that the rationale behind these decisions is accessible to healthcare providers. Informed consent is another cornerstone of ethical medical practice. This extends to the use of AI in diagnostics—patients should be made aware when AI tools are utilized in their diagnosis and consent to their use. The thesis should explore the mechanisms put in place to ensure that patients are adequately informed and can provide informed consent.

The ethical ramifications of error in AI diagnosis must also be considered. While human diagnostic processes are not infallible, the delegation of diagnostic tasks to an AI system comes with a different set of expectations and potential for error. The thesis should contemplate the ethical management of AI diagnostic errors and the protocols established for reporting and addressing such errors. Finally, the broader societal implications of introducing AI into healthcare diagnostics must be ethically evaluated. This includes considering how the technology may affect the physician-patient relationship, the potential for AI to augment or supplant human decision-making in medical settings, and the long-term impacts on medical training and practice.

This section provides a detailed examination of the multifaceted ethical landscape surrounding the use of AI in stomach disease diagnostics. It underscores the need for a principled approach that prioritizes patient welfare, equity, transparency, and accountability, ensuring that the deployment of AI in healthcare contributes positively to patient outcomes and upholds the ethical standards of the medical profession.

5.3 Safety Considerations

This section addresses the critical importance of ensuring that the deployment of AI in the diagnosis of stomach diseases through endoscopic imaging is safe for all stakeholders, particularly patients. This entails a thorough analysis of the potential risks and the implementation of robust safeguards to mitigate those risks.

Central to safety considerations is the accuracy and reliability of the AI model. The potential for misdiagnosis or missed diagnosis carries significant consequences in medical practice. The thesis discusses the extensive validation processes undertaken, including cross-validation and testing against established benchmarks, to ensure that the model's diagnostic recommendations are consistently reliable and in line with the current medical understanding. Moreover, the robustness of the model is scrutinized, particularly its resilience to adversarial attacks or data corruption. Given the sensitive nature of medical diagnostics, the integrity of the model's outputs must be safeguarded against both unintentional data errors and deliberate attempts to manipulate outcomes. Strategies employed to secure the model, such as input validation, regular updates, and monitoring of the model's operation, are examined.

The discussion extends to the model's integration into clinical workflows. Safety is not only a matter of algorithmic performance but also how the tool is used by healthcare professionals. The thesis outlines the training protocols for medical staff, ensuring that the AI tool is used appropriately and that users are aware of its limitations. The importance of having a clear protocol for human oversight, especially in ambiguous or high-risk cases, is emphasized to ensure that the final diagnostic decisions are made with a full understanding of the patient's clinical context. Emergency protocols are also considered, detailing the steps to be taken in the event of system failures or discrepancies between the AI's diagnosis and clinical judgment. The thesis discusses the contingency plans in place, ensuring that patient safety remains paramount even when technology falters. The ethical implications of AI in healthcare extend to the safety of patient data. The thesis considers the measures implemented to protect patient data from breaches, unauthorized access, or misuse. The deployment of state-of-the-art cybersecurity measures and adherence to best practices in data privacy and security are essential components of the safety strategy.

Finally, the thesis recognizes the dynamic nature of both AI technology and medical knowledge. It underscores the need for ongoing monitoring and updates to the AI system, ensuring that it evolves in response to new medical discoveries and technological advancements while maintaining the highest safety standards. This section of the thesis presents a comprehensive approach to ensuring the safe implementation of AI in medical diagnostics. By addressing the technical, operational, and data security aspects of the AI model, the research underscores a commitment to patient safety, which is the cornerstone of ethical medical practice.

5.4 Legal Considerations

In this section, we delve into the legal framework surrounding the implementation of artificial intelligence (AI) in medical diagnostics, a domain that is increasingly coming under scrutiny due to the profound implications AI has on patient care. The application of a deep learning model for the diagnosis of stomach diseases from endoscopic images stands at the intersection of healthcare regulation, technology law, and patient rights, each of which carries its legal stipulations and standards that must be meticulously adhered to.

The thesis first addresses compliance with medical device regulations, which in many jurisdictions classify diagnostic AI systems as medical devices. This classification subjects the model to stringent regulatory oversight, such as the Food and Drug Administration (FDA) in the United States or the European Medicines Agency (EMA) in the European Union. The process of obtaining approval from such regulatory bodies involves demonstrating that the model is safe and effective for its intended use, through clinical trials and validation studies. The thesis outlines the steps taken to ensure compliance with these regulations, from premarket approval processes to post-market surveillance requirements. Furthermore, the discussion explores the legal ramifications of potential diagnostic errors. In the event of a misdiagnosis by the AI model, it is critical to understand how liability is apportioned between the developers, healthcare providers, and institutions using the technology. The thesis examines the current legal discourse on AI and malpractice, emphasizing the need for clear legal guidelines on accountability in the age of AI-assisted decision-making.

The protection of patient data also constitutes a significant legal concern. With the model's reliance on patient data for training and operation, it must operate within the bounds of data protection laws. The thesis discusses the protocols established to ensure that the model complies with laws such as the General Data Protection Regulation (GDPR) for EU citizens, or other local data protection laws, which govern the collection, processing, and storage of personal health information. Intellectual property rights are another legal aspect covered in the thesis. It evaluates the IP rights associated with the development of the AI model, including patents, copyrights, and trade secrets, and how these rights are managed through licensing agreements or research collaborations.

Lastly, the thesis considers the broader legal implications of AI integration into healthcare. This includes the potential need for new laws or amendments to existing laws to better accommodate the unique challenges posed by AI technology, such as issues of consent, privacy, and the evolving nature of healthcare delivery in the digital age. This section provides an in-depth legal analysis of the various statutes, regulations, and legal challenges associated with the deployment of AI in healthcare diagnostics. It underscores the importance of proactive legal planning and compliance to safeguard the interests of all parties involved, from developers to end-users, and most importantly, to ensure the protection and welfare of patients whose lives are impacted by these technological advancements.

5.5 Impact of the Project on Societal, Health, and Cultural Issues

This section of the thesis delves into the broad-reaching consequences that the implementation of an AI-based diagnostic tool for stomach diseases from endoscopic images may have beyond the medical community. The development and potential deployment of this technology are poised to bring about significant changes, each carrying its own societal, health, and cultural implications.

From a societal standpoint, the integration of such advanced technology into healthcare systems represents a leap towards modernizing medical diagnostics, potentially leading to more rapid and accurate disease identification, which can improve patient outcomes. However, this also raises concerns about the accessibility of such technology across different socio-economic groups, possibly exacerbating existing health disparities if not addressed inclusively. The project's impact on employment within the healthcare sector, particularly regarding the roles of diagnostic specialists, must also be considered, as AI systems can alter job functions and requirements. Health-wise, the introduction of an AI model with high diagnostic accuracy can significantly reduce the burden on healthcare systems by streamlining the diagnosis process, reducing the time and costs associated with manual diagnosis, and potentially lowering the rates of misdiagnosis. However, this shift necessitates a reevaluation of health policies and insurance schemes to incorporate and manage the use of AI diagnostic tools.

Culturally, the adoption of AI in medical practices intersects with various beliefs and attitudes towards technology and healthcare. In some cultures, there may be a preference for human judgment over machine-based decisions, or conversely, a rapid embrace of new technologies. This project must be sensitive to such cultural nuances and work towards bridging the gap between technology and cultural acceptance. The project should also account for the potential impact on patient-doctor relationships, as the introduction of AI tools could change the dynamics of trust and personal care traditionally associated with medical practice.

However, the project also poses potential cultural challenges, such as the risk of exacerbating health disparities if the technology is not equally accessible to all segments of society. Furthermore, there may be resistance from parts of the medical community or patients due to concerns over the depersonalization of care or fears of misdiagnosis. Addressing these concerns through public education, transparent communication, and inclusive policymaking is essential to fostering acceptance and integration of AI in healthcare settings.

5.6 Impact of Project on the Environment and Sustainability

This section of the thesis explores the environmental footprint of developing and implementing AI-based diagnostic models, like the one proposed for analyzing endoscopic images to detect stomach diseases. This section critically assesses how the project aligns with the broader objective of sustainable development, considering both the direct and indirect environmental implications of the research and its applications.

The production and operation of AI systems, particularly deep learning models, are resource-intensive processes. Training such models require significant computational power, often involving energy consumption that contributes to carbon emissions. This thesis acknowledges the environmental cost of developing the proposed model, discussing the energy profile of the GPUs used during the model's training and inference stages. It delves into the carbon footprint associated with these computational requirements and discusses strategies employed to mitigate this impact, such as the use of renewable energy sources, optimizing algorithms for efficiency, and selecting energy-efficient hardware. In addition to the operational aspects, the sustainability of the model's deployment in clinical settings is considered. The thesis evaluates the life cycle of the model from development to deployment and potential obsolescence, considering how each stage can be managed responsibly to minimize environmental impact. This includes exploring the potential for reducing waste through model-sharing initiatives and the use of cloud computing services that optimize resource utilization across users.

Furthermore, the thesis contemplates the indirect environmental benefits of the project. By improving the efficiency and accuracy of disease diagnosis, the AI model may contribute to reduced resource usage in healthcare. More accurate diagnostics can lead to more targeted treatments, potentially decreasing the overall consumption of medical supplies, reducing waste from unnecessary procedures, and optimizing the utilization of medical equipment. The long-term sustainability of the project is also addressed. The thesis posits how the continuous improvement and updating of the AI model can extend its useful life, reduce the need for frequent redevelopment, and thus lower the environmental burden. The project's alignment with sustainable healthcare practices, such as minimizing the need for repeat diagnostic procedures, is emphasized as a key contribution to environmental sustainability.

Lastly, the section highlights the need for ongoing research into the environmental aspects of AI in healthcare. It advocates for a holistic approach that considers not only the technical and clinical performance of AI models but also their compatibility with sustainable practices and environmental stewardship. This section presents a thoughtful analysis of the environmental implications of AI development in healthcare diagnostics. It demonstrates a commitment to identifying and implementing strategies that reduce environmental impact while advancing the capabilities of medical diagnostics, thereby contributing to a more sustainable future.

CHAPTER VI

Addressing Complex Engineering Problems and Activities

6.1 Complex engineering problems associated with the current thesis

This section explores the intricate challenges faced during the development of an AI-based diagnostic tool for stomach diseases using endoscopic images. This exploration underscores the multifaceted nature of engineering problems that span technical, ethical, and practical domains, highlighting the innovative solutions and interdisciplinary approaches employed to navigate these challenges.

At the core of the project lies the technical complexity of designing a deep learning model capable of accurately interpreting endoscopic images. This involves addressing issues such as high-dimensional data processing, model overfitting, and the development of algorithms that can generalize well from training data to real-world clinical settings. The section details the specific architectural choices made, such as the integration of convolutional neural networks (CNNs) with transformer models, to enhance the model's ability to learn nuanced patterns in the data indicative of various stomach pathologies.

Another significant engineering challenge discussed is the handling and protection of sensitive medical data. Ensuring patient privacy while training the model on comprehensive datasets required the implementation of robust data encryption and anonymization techniques, as well as compliance with strict regulatory standards such as GDPR and HIPAA.

The project also navigated the ethical considerations inherent in AI healthcare applications. This included ensuring that the AI model's decision-making process is transparent and explainable to healthcare providers and that it does not introduce bias that could lead to disparate health outcomes among different patient populations. Strategies for addressing these ethical challenges, such as dataset diversity and algorithmic fairness techniques, are elaborated upon. From a practical standpoint, integrating the AI tool into existing clinical workflows presented its own set of challenges. The section describes the collaboration with medical professionals to ensure that the tool complements traditional diagnostic methods, enhancing rather than complicating the diagnostic process. This involved user interface design considerations, the development of interpretability features that allow clinicians to understand the AI's reasoning and the creation of protocols for when the AI's diagnosis should be escalated to human review.

Lastly, the section touches on the sustainability and scalability of the AI solution. It considers the computational resources required for model training and inference, discussing the balance between model complexity and operational efficiency. This includes an analysis of the environmental impact of training large models and the efforts made to mitigate this through efficient computing practices and the potential for model optimization.

Table 1
Complex Engineering Problems

Attribute	Complex Engineering Problems
Depth of knowledge required	**P1** - To find out the specific symptoms of the disease which I need to find or detect using AI. Need to know proper image processing so that I can find the appropriate pipeline for image preprocessing. Need to know deep-learning from core level so that it becomes easy to develop a novel architecture and data specific.
Range of conflicting requirements	**P2** - A large number of images required as training data which is not available currently. Also required high-quality configuration to train the model using a large dataset.
Depth of analysis required	**P3** - Required two types of analysis: one of them is qualitative analysis and another one is quantitative analysis. The qualitative analysis is required for ensuring the usability of the model or system in real life. The qualitative analysis is required to show the effectiveness, efficiency, and robustness of the research.
Familiarity of issues	**P4** - This research domain is quite different as it requires images which we can get using endoscopy. There is a lack of a proper system so we need to develop a proper system which can perform effectively and efficiently.
Extent of applicable codes	**P5** - I have to show the standard of this work using different matrices and analyses.
Extent of stakeholder involvement and conflicting requirements	**P6** - I need to develop a system where it requires a large dataset and also efficient hardware devices which are costly and not available at a time. So, I need stakeholders too.
Interdependence	**P7** - My research completely depends on deep learning and image processing techniques. So I need to have a clear concept of machine learning and deep learning.

6.2 Complex engineering activities associated with the current thesis

This section meticulously outlines the diverse and intricate set of tasks undertaken to bring the AI-based diagnostic tool for stomach diseases to fruition. This section not only highlights the technical endeavours involved in the project but also underscores the interdisciplinary collaboration and strategic planning required to navigate the complexities of integrating AI technology within the medical field.

6.2.1 Data Collection and Preparation

One of the foundational activities detailed is the collection and preparation of endoscopic image data. This involved establishing partnerships with medical institutions for data acquisition, ensuring the data spanned a comprehensive range of stomach diseases. The complexity of anonymizing the data to protect patient privacy while retaining crucial diagnostic features is discussed, alongside the techniques employed for data augmentation to enhance the model's robustness.

6.2.2 Model Design and Development

The engineering activities extend to the meticulous design and iterative development of the deep learning model. This includes the selection of appropriate neural network architectures, such as convolutional neural networks (CNNs) and transformers, tailored to the specific challenges of medical image analysis. The section elaborates on the process of tuning hyperparameters, implementing regularization techniques to combat overfitting, and employing state-of-the-art optimization algorithms to improve training efficiency and model performance.

6.2.3 Validation and Testing

Critical to the project's success were the rigorous validation and testing procedures implemented to ensure the model's accuracy and reliability. The section describes the deployment of a 5-fold cross-validation strategy, allowing for the comprehensive evaluation of the model's diagnostic capabilities across different subsets of data. The integration of performance metrics, including accuracy, precision, recall, and the Area Under the Curve (AUC) for ROC analysis, provided a multifaceted view of the model's effectiveness.

6.2.4 Ethical and Regulatory Compliance

The activities also encompassed navigating the ethical and regulatory landscape associated with deploying AI in healthcare. This entailed conducting ethical reviews, securing approvals from institutional review boards, and ensuring compliance with healthcare regulations such as HIPAA and GDPR. The section highlights the ongoing dialogue with ethicists and legal experts to address concerns related to patient consent, data protection, and algorithmic transparency.

6.2.5 Integration into Clinical Workflows

Another complex activity involved the integration of the AI diagnostic tool into existing clinical workflows. Collaborating with healthcare professionals, the project team worked to ensure that the tool complemented traditional diagnostic methods and facilitated a seamless user experience for clinicians. This included developing user-friendly interfaces, providing training sessions for medical staff, and establishing protocols for interpreting and acting on the AI's recommendations.

6.2.6 Sustainability and Scalability

Finally, the section addresses the considerations made for the project's sustainability and scalability. This encompasses optimizing the model to reduce computational demands, exploring deployment options that minimize environmental impact, and planning for the tool's adaptation and evolution in response to advances in medical knowledge and technology.

Table 2
Addressing the Attributes of Complex Engineering Activities

Attribute	**Addressing the Attributes of Complex Engineering Activities**
Range of resources	**A1** - I need to find a range of resources like GPU, RAM, storage device, and large training dataset.
Level of interaction	**A2** - I have to interact in the model architecture development and image preprocessing technique. Again, I have evaluated my work with qualitative and quantitative analysis.
Innovation	**A3** - The hybrid model architecture that I have developed is completely different and novel. The preprocessing technique is also efficiently enhancing the image, which is also a modified thinking.
Consequences for society and the environment	**A4** - This proposed system will enable effective detection of stomach infection within a short time, enhancing cost-effectiveness.
Familiarity	**A5** - This system must need to know the medical healthcare diagnosis approach and also the common noise that can appear in real life.

CHAPTER VII

Conclusions

7.1 Conclusion and Challenges Faced

- Groundbreaking contribution to medical diagnostics through an innovative deep-learning model for precise diagnosis of stomach diseases from endoscopic images.
- Rigorous data preprocessing, including advanced anonymization and augmentation techniques, curated a dataset enhancing model training efficiency.
- Integration of cutting-edge CNNs and transformer models resulted in a highly effective diagnostic tool.
- Exceptional model performance demonstrated via exhaustive evaluation, surpassing existing benchmarks in accuracy, precision, recall, and AUC scores.
- Sets a new standard for AI applications in gastroenterological diagnostics, paving the way for future explorations.

7.2 Limitations

- Exceptional performance necessitates validation on larger, more diverse datasets for robustness and generalizability.
- Reliance on high-quality, well-annotated images may limit applicability in low-resource environments.
- Computational demands of training sophisticated models pose challenges for widespread adoption.
- Potential for algorithmic bias remains, requiring ongoing scrutiny and mitigation efforts.

7.3 Recommendations and Future Work

- Develop highly efficient models balancing sophistication with simplicity for integration into mobile devices.

- Innovate lightweight model architectures optimized for computational constraints of mobile technology.
- Enhance diagnostic capabilities through incorporation of localization and segmentation tasks before classification.
- Develop and integrate sophisticated image analysis techniques for accurate delineation of diseased from healthy tissue.
- Conduct comprehensive statistical analysis to thoroughly evaluate model performance and robustness.
- Push boundaries of AI in medical diagnostics, aiming for technologically advanced and practical solutions accessible for widespread clinical application.

References

[1] A. Yadav and D. K. Vishwakarma, “A deep multi-level attentive network for multimodal sentiment analysis,” *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 19, no. 1, pp. 1–19, 2023.

[2] W.-L. Zheng, W. Liu, Y. Lu, B.-L. Lu, and A. Cichocki, “Emotionmeter: A multimodal framework for recognizing human emotions,” *IEEE transactions on cybernetics*, vol. 49, no. 3, pp. 1110–1122, 2018.

[3] X. Huang, J. Kortelainen, G. Zhao, X. Li, A. Moilanen, T. Seppänen, and M. Pietikäinen, “Multi-modal emotion analysis from facial expressions and electroencephalogram,” *Computer Vision and Image Understanding*, vol. 147, pp. 114–124, 2016.

[4] Y. Tan, Z. Sun, F. Duan, J. Solé-Casals, and C. F. Caiafa, “A multimodal emotion recognition method based on facial expressions and electroencephalography,” *Biomedical Signal Processing and Control*, vol. 70, p. 103029, 2021.

[5] M. Soleymani, D. Garcia, B. Jou, B. Schuller, S.-F. Chang, and M. Pantic, “A survey of multimodal sentiment analysis,” *Image and Vision Computing*, vol. 65, pp. 3–14, 2017.

[6] L.-C. Shi, Y.-Y. Jiao, and B.-L. Lu, “Differential entropy feature for eeg-based vigilance estimation,” in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 6627–6630, IEEE, 2013.

[7] W.-L. Zheng, J.-Y. Zhu, and B.-L. Lu, “Identifying stable patterns over time for emotion recognition from eeg,” *IEEE Transactions on Affective Computing*, vol. 10, no. 3, pp. 417–429, 2019.

[8] S. Zhe and R. Micheletto, “Emergence of intelligent behavior from a minimalistic stochastic model for the navigation of autonomous robots,” in *2014 International Conference on Information Science, Electronics and Electrical Engineering*, vol. 2, pp. 1301–1305, IEEE, 2014.

[9] A. Majid, M. A. Khan, M. Yasmin, A. Rehman, A. Yousafzai, and U. Tariq, “Classification of stomach infections: A paradigm of convolutional neural network along with classical features fusion and selection,” *Microscopy research and technique*, vol. 83, no. 5, pp. 562–576, 2020.

[10] M. A. Khan, M. S. Sarfraz, M. Alhaisoni, A. A. Albesher, S. Wang, and I. Ashraf, "Stomachnet: Optimal deep learning features fusion for stomach abnormalities classification," *IEEE Access*, vol. 8, pp. 197969–197981, 2020.

[11] J. Amin, M. Sharif, E. Gul, and R. S. Nayak, "3d-semantic segmentation and classification of stomach infections using uncertainty aware deep neural networks," *Complex & Intelligent Systems*, pp. 1–17, 2021.

[12] M. A. Khan, A. Majid, N. Hussain, M. Alhaisoni, Y.-D. Zhang, S. Kadry, and Y. Nam, "Multiclass stomach diseases classification using deep learning features optimization," 2021.

[13] F. Mohammad and M. Al-Razgan, "Deep feature fusion and optimization-based approach for stomach disease classification," *Sensors*, vol. 22, no. 7, p. 2801, 2022.

[14] M. S. Amin, J. H. Shah, M. Yasmin, G. J. Ansari, M. A. Khan, U. Tariq, Y. J. Kim, and B. Chang, "A two stream fusion assisted deep learning framework for stomach diseases classification," *CMC-Comput. Mater. Contin*, vol. 73, pp. 4423–4439, 2022.

[15] M. A. Khan, I. M. Nasir, M. Sharif, M. Alhaisoni, S. Kadry, S. A. C. Bukhari, and Y. Nam, "A blockchain based framework for stomach abnormalities recognition," *Comput. Mater. Contin*, vol. 67, pp. 141–158, 2021.

[16] J. Amin, M. Sharif, E. Gul, and R. S. Nayak, "3d-semantic segmentation and classification of stomach infections using uncertainty aware deep neural networks," *Complex & Intelligent Systems*, vol. 8, no. 4, pp. 3041–3057, 2022.

[17] H. M. R. J. Alhasnawi and A. A. J. Aljanaby, "The immunological role of cd4 and cd8 in patients infected with helicobacter pylori and stomach cancer," *Gene Reports*, vol. 26, p. 101500, 2022.

[18] S. M. Alexander, R. J. Retnakumar, D. Chouhan, T. N. B. Devi, S. Dharmaseelan, K. Devadas, N. Thapa, J. P. Tamang, S. C. Lamtha, and S. Chattopadhyay, "Helicobacter pylori in human stomach: the inconsistencies in clinical outcomes and the probable causes," *Frontiers in microbiology*, vol. 12, p. 713955, 2021.

[19] E. Rimbara, M. Suzuki, H. Matsui, M. Nakamura, M. Morimoto, C. Sasakawa, H. Masuda, S. Nomura, T. Osaki, N. Nagata, *et al.*, "Isolation and characterization of helicobacter suis from human stomach," *Proceedings of the National Academy of Sciences*, vol. 118, no. 13, p. e2026337118, 2021.

[20] M. Ilic and I. Ilic, "Epidemiology of stomach cancer," *World journal of gastroenterology*, vol. 28, no. 12, p. 1187, 2022.

[21] J. Naz, M. A. Khan, M. Alhaisoni, O.-Y. Song, U. Tariq, and S. Kadry, “Segmentation and classification of stomach abnormalities using deep learning.,” *Computers, Materials & Continua*, vol. 69, no. 1, 2021.

[22] M. Ramzan, M. Raza, M. Sharif, M. A. Khan, and Y. Nam, “Gastrointestinal tract infections classification using deep learning,” *Comput. Mater. Contin*, vol. 69, pp. 3239–3257, 2021.