University of Southampton
Faculty of Engineering and Physical Sciences
Electronics and Computer Science

A Deep Learning Approach for Accurate Classification of Abnormalities in Intestinal Tracts

October 2023

MSc Artificial Intelligence

Acknowledgements

I would like to express my sincere gratitude to Professor Enrico Gerding, Professor Jo Grundy, and Professor Srinandan Dasmahapatra for their invaluable guidance and support throughout the completion of this research work. Their expertise, encouragement, and insightful feedback have been instrumental in shaping the direction and quality of this study.

Statement of Originality

- I have read and understood the ECS Academic Integrity information and the University's Academic Integrity Guidance for Students.
- I am aware that failure to act in accordance with the Regulations Governing Academic Integrity may lead to the imposition of penalties which, for the most serious cases, may include termination of programme.
- I consent to the University copying and distributing any or all of my work in any form and using third parties (who may be based outside the EU/EEA) to verify whether my work contains plagiarized material, and for quality assurance purposes.

You must change the statements in the boxes if you do not agree with them.

We expect you to acknowledge all sources of information (e.g. ideas, algorithms, data) using citations. You must also put quotation marks around any sections of text that you have copied without paraphrasing. If any figures or tables have been taken or modified from another source, you must explain this in the caption and cite the original source.

I have acknowledged all sources, and identified any content taken from elsewhere.

If you have used any code (e.g. open-source code), reference designs, or similar resources thathave been produced by anyone else, you must list them in the box below. In the report, you must explain what was used and how it relates to the work you have done.

I have not used any resources produced by anyone else.

You can consult with module teaching staff/demonstrators, but you should not show anyone else your work (this includes uploading your work to publicly-accessible repositories e.g. Github, unless expressly permitted by the module leader), or help them to do theirs. For individual assignments, we expect you to work on your own. For group assignments, we expect that you work only with your allocated group. You must get permission in writing from the module teaching staff before you seek outside assistance, e.g. a proofreading service, and declare it here.
I did all the work myself, or with my allocated group, and have not helped anyone else.

We expect that you have not fabricated, modified or distorted any data, evidence, references, experimental results, or other material used or presented in the report. You must clearly describe your experiments and how the results were obtained, and include all data, source code and/or designs (either in the report, or submitted as a separate file) so that your results could be reproduced.

The material in the report is genuine, and I have included all my data/code/designs.

We expect that you have not previously submitted any part of this work for another assessment.You must get permission in writing from the module teaching staff before re-using any of your previously submitted work for this assessment.

I have not submitted any part of this work for another assessment.
If your work involved research/studies (including surveys) on human participants, their cells or data, or on animals, you must have been granted ethical approval before the work was carried out, and any experiments must have followed these requirements. You must give details of this in the report, and list the ethical approval reference number(s) in the box below.

My work did not involve human participants, their cells or data, or animals.

Contents

INTRODUCTION .... **Error! Bookmark not defined.**
1.1 Background .... **Error! Bookmark not defined.**
1.2 Rise of Gastrointestinal Disorders **Error! Bookmark not defined.**
1.3 Visual Capsule Endoscopy Technology..... **Error! Bookmark not defined.**
Dataset Collection .... **Error! Bookmark not defined.**
2.1 Dataset Overview .... **Error! Bookmark not defined.**
2.2 Class Distribution and Challenges **Error! Bookmark not defined.**
2.3 Data Analysis Challenges in Healthcare .... **Error! Bookmark not defined.**
Literature Review .... **Error! Bookmark not defined.**
3.1 Gastrointestinal Disorders and Imaging Techniques............ **Error! Bookmark not defined.**
3.2 Therapeutic Approaches and Procedures ... **Error! Bookmark not defined.**
3.3 Advancements in Computer Vision Methodologies .... **Error! Bookmark not defined.**
Methodology .... **Error! Bookmark not defined.**
4.1 VGG-16 Architecture .... **Error! Bookmark not defined.**
4.1.1 Mathematical VGG-16.......... **Error! Bookmark not defined.**
4.1.2 Basic Convolution Operation:............. **Error! Bookmark not defined.**
4.1.3 Max Pooling Operation:........ **Error! Bookmark not defined.**
4.1.4 Fully Connected Layers: .... **Error! Bookmark not defined.**

4.1.5 Stacking Layers:....................**Error! Bookmark not defined.**
4.1.6 ReLU Activation:..................**Error! Bookmark not defined.**
4.1.7 Dropout (Optional): ....**Error! Bookmark not defined.**
4.1.8 Softmax Output:....................**Error! Bookmark not defined.**
4.2 Convolutional Neural Network CNN...... **Error! Bookmark not defined.**
Model Training and Evaluation ....Error! Bookmark not defined.
5.1 Data Preparation and Preprocessing........... **Error! Bookmark not defined.**
5.2 Data Augmentation and Balancing............ **Error! Bookmark not defined.**
5.3 Model Architecture and Compilation......... **Error! Bookmark not defined.**
5.4 Model Training and Evaluation....**Error! Bookmark not defined.**
5.5 Result Analysis............................**Error! Bookmark not defined.**
Comparative Analysis..........................Error! Bookmark not defined.
6.1 VGG-16:.......................................**Error! Bookmark not defined.**
6.2 DenseNet-161:.............................**Error! Bookmark not defined.**
6.3 ResNet-152:...39
6.4 Comparison: 39
6.4.1 DenseNet-161: ....**Error! Bookmark not defined.**
6.4.2 ResNet-152: ....**Error! Bookmark not defined.**
6.4.3 VGG-16: ....**Error! Bookmark not defined.**
Test Results...42
Conclusion 45
Appendix A ....Error! Bookmark not defined.
A.1 Main VGG-16 Code 54
Bibliography 55

Abstract

This study, in the field of health research aims to identify issues at a stage to prevent them from becoming more serious problems. The focus is on detecting abnormalities by analyzing changes in the shape, colour and lighting conditions of lesions. Although accurately categorizing these anomalies is complex due to influencing factors this study takes on the challenge.

One notable aspect of this research is its contribution to existing knowledge that goes beyond efforts by researchers. By building upon their insights this investigation introduces an automated method that effectively identifies abnormalities in the tract. The study utilizes a dataset called "Kvasir Capsule," which contains video capsule endoscopy data. By employing the VGG 16 model—a machine learning algorithm—this research not. Also advances the current state of the art.

Significantly this paper surpasses utilized models like DenseNet 161 and ResNet 152 by achieving an accuracy rate of 83%. This noteworthy achievement positions the research as a symbol of progress. Offers insights into classifying abnormalities in the gastrointestinal tract. Consequently, it plays a role, in driving advancements in imaging and automation within healthcare.

The research paper highlights advancements in identifying and classifying issues. This progress is crucial, for enhancing procedures and ensuring the well-being of patients in the run.

Chapter 1

INTRODUCTION

In times there has been a rise, in health concerns such as ulcers, bleeding disorders and gastroenteritis infections. According to the World Health Organization (WHO) colon cancer has become a cause of death claiming 765,000 lives.

Furthermore, a research study conducted in the United States has highlighted an increase in tract infections. Since 2017 135,430 cases have been reported. The global average of cases has exceeded 200,000 annually since 2011.

Recognizing the importance of identification and diagnosis in combating these infections this paper introduces an approach called the Machine Learning Classification Method. This groundbreaking method involves using a camera equipped capsule that navigates through the body system and sends images to devices for analysis. These images are then examined by a Deep Learning model to quickly identify illnesses. This method represents a departure from review processes.

Over the decade various statistical and artificial intelligence techniques have been explored to automate the detection of infected areas within these images. To address the challenge posed by the number of images involved this paper pioneers an identification process specifically designed to analyze and identify frames showing signs of infection. The utilization of machine learning methods shows promising advancements for this field due to their effectiveness, in operations.

The study contributes significantly by developing an intelligence model that accurately categorizes images displaying infections.

To accomplish this the researchers carefully preprocess the data obtained from Wireless Capsule Endoscopy (WCE) images. They analyze a dataset of images. Segregate diseased photos to create separate datasets, for healthy and infected images. While there is training data**,** they employ various techniques such as rotation, shifting, shearing, zooming and horizontal flipping to enhance the infected image dataset. These techniques play a role in addressing imbalances by improving the representation of minority class data and ensuring classes for model training. The remarkable aspect of their work lies in the utilization of Deep Convolutional Neural Network on these datasets. This does not enable them to evaluate the classification algorithm on test data but also achieves accuracy compared to previously employed models like DenseNet 161 and ResNet 152.

In summary this research significantly advances the fields of imaging and healthcare automation while making contributions towards enhancing the identification and classification of abnormalities, in the tract. The meticulous methodology presented here stands as evidence of progress made in addressing challenges posed by rising health concerns.

1.1 Background

Located between the abdomen and the large intestine the small intestine plays a role, in the system by absorbing essential nutrients. Conditions like Crohns disease and adenocarcinoma that affect this area require examination and targeted treatment. However traditional methods face challenges due to the location of the intestine.

To overcome these challenges and bring about examination and treatment Visual Capsule Endoscopy (VCE) has emerged as a technique. Using capsules equipped with wide angle lenses VCE captures thousands of frames creating a repository of data. While this wealth of information is groundbreaking it calls for a shift towards automated image analysis to unlock its potential.

This paper dives into research by recognizing the power of VCE. Taking a step forward it integrates Machine Learning algorithms into the analysis framework. These algorithms have abilities to learn patterns and nuances holding promise in improving detection rates and accuracy when identifying abnormalities, within the small intestine.

This integration represents a symbol of progress, in the field of health offering a nuanced and effective approach to identifying and planning treatment for diseases. Despite the challenges presented by data availability and the complexity of explaining AI decisions this research paper skillfully navigates these obstacles. What sets it apart is not the implementation of Machine Learning algorithms for automated image analysis but also its recognition and handling of the accompanying challenges.

In essence this paper serves as evidence of how cutting-edge technology can be combined with the requirements of science. By embracing the potential of VCE (Video Capsule Endoscopy) and leveraging Machine Learning capabilities this study not significantly contributes to improving disease detection in the intestine but also paves the way for a more efficient and sophisticated era in gastrointestinal healthcare. The integration of algorithms do not enhance diagnostic accuracy but also brings us closer, to unraveling the complexities associated with diseases that impact this vital part of our digestive system.

1.2 Rise of Gastrointestinal Disorders

1.2 The Increase, in Gastrointestinal Disorders

In times there has been a rise in gastrointestinal disorders, such as ulcers, bleeding problems and infections. These health concerns have become more prevalent globally. Are closely associated with colon cancer, which's responsible for a significant number of deaths worldwide (approximately 765,000). Additionally gastrointestinal tract infections have been on the rise since 2017 with around 135,430 new cases reported. To address this issue effectively and timely diagnose these illnesses

Machine Learning Classification Methods are being utilized. This involves the use of capsules that contain embedded cameras of instantly identifying diseases using advanced Deep Learning models. The traditional manual review methods are time consuming. Emphasize the need for automated solutions.

1.3 Advancements in Visual Capsule Endoscopy Technology

Visual Capsule Endoscopy (VCE) technology has significantly progressed since the 2000s. Now offers a non-invasive way to examine the gastrointestinal system. Various VCE devices equipped with wide angle lenses can capture thousands of frames to provide insights. However manually examining these frames by professionals is labor intensive and prone to errors due to their numbers. As a result, there is exploration into utilizing machine learning algorithms to automate the analysis of VCE data—a development that holds promise. The future of VCE technology relies heavily on systems, like computer vision to improve detection rates and accurately locate capsules.

Chapter 2

Data Collection

2.1 Data Overview

The dataset, a cornerstone of our investigation, comprises a substantial collection of 47,238 meticulously annotated images, meticulously stored in the PNG format. Organized with precision, these images are housed within the dedicated "pictures" folder, methodically categorized into specific classes, each delineated by a distinct folder name such as "polyp" and "Angioectasia." This conscientious organization facilitates a structured approach to image retrieval and analysis, crucial in the realm of healthcare research.

A noteworthy characteristic of this dataset lies in its transparency and accessibility, offering a comprehensive glimpse into the diverse manifestations of gastrointestinal findings. Each annotated image is intrinsically associated with a specific class, presenting an invaluable resource for researchers and practitioners alike to explore and scrutinise nuanced details associated with conditions such as polyps and Angioectasia.

However, an inherent challenge encapsulates the dataset – an imbalance in the distribution of images across different classes. This nuanced predicament poses a significant challenge in healthcare research, as it introduces varying frequencies of different findings. It is in addressing this intricate challenge that the contribution of this paper shines prominently.

The groundbreaking contribution of this study to the dataset landscape lies in its commitment to navigate and mitigate the challenges posed by imbalances. By acknowledging and addressing this inherent imbalance, the paper takes a pivotal step towards enhancing the reliability and robustness of our findings. Through meticulous exploration and innovative approaches in subsequent sections, we present strategies to mitigate the impact of this imbalance on our machine learning model training.

In essence, this section not only provides a detailed overview of the dataset but also underscores the pioneering efforts undertaken by this paper to grapple with and overcome the inherent challenges within the dataset. Through strategic insights and thoughtful considerations, the contribution of this paper becomes evident in its dedication to refining the dataset landscape for the betterment of healthcare research.

2.2 Class Distribution and Challenges

The dataset reveals a range of findings covering 14 different categories as outlined in Table 2.1. This detailed classification system does not capture the complexities of conditions but also serves as a valuable resource, for researchers and healthcare professionals exploring the diverse landscape of gastrointestinal health.

However, a notable challenge arises due to a distribution of images across these categories. This inherent imbalance poses an obstacle that requires solutions to ensure the accuracy and effectiveness of subsequent analyses. Certain categories have a prevalence, which calls for advanced data analysis techniques, beyond methods. To gain insights it is crucial to understand these imbalances and address this challenge effectively as accurate model training and reliable healthcare predictions depend on it.

This paper stands out by addressing the complexities posed by imbalanced class distribution. Recognizing the nature of this issue the subsequent sections do not delve into the nuances of this challenge but also propose innovative strategies to mitigate its impact.

With groundbreaking insights and strategic actions this paper does not add to our comprehension of the challenges related to class distribution, in datasets. Also introduces practical solutions to improve the reliability of future analyses. In doing it greatly pushes forward the field of healthcare research guaranteeing that the models trained on this dataset are both precise and capable of handling the intricacies, in real world healthcare situations. This committed endeavor to tackle class distribution challenges serves as proof of the impact this paper has in the realm of health research.

ID	**Label**	**ID**	**Label**	**ID**	**Label**
1.	Ampulla of Vater	6.	Erythema	11.	Polyp
2.	Angiectasia	7.	Foreign body	12.	Pylorus

3.	Blood - fresh	8.	Ileocecal valve	13.	Reduced mucosal view
4.	Blood - hematin	9.	Lymphangiectasia	14.	Ulcer
5.	Erosion	10.	Normal clean mucosa		

Table 2.1 *This table consists of all the classes used in this research. Source: Kvasir-Capsule*

2.3 Data Analysis Challenges in Healthcare

When it comes to analyzing healthcare data in the field of health there are various complex challenges that arise. One major challenge is the distribution of images across categories as shown in Figure 2.1. This issue presents an obstacle for researchers requiring a nuanced approach to extract meaningful and accurate insights.

The imbalance in image distribution within the healthcare dataset goes beyond a representation; it represents a profound challenge that needs focused attention. This research paper stands at the forefront of healthcare studies. Contributes uniquely to understanding and addressing these challenges.

In the following sections we delve into the complexities of imbalanced datasets. Shed light on the intricacies faced by researchers striving for fairness and accuracy in their analysis. By acknowledging and tackling this challenge head on our paper offers techniques that go beyond traditional methodologies enabling a fair and accurate exploration of the dataset.

This papers distinctive contribution doesn't stop at identifying challenges; it also presents solutions that significantly improve the effectiveness of data analysis in healthcare applications. The methodologies and insights we present pave the way, for a reliable analysis unlocking the full potential of the dataset.
The imbalances present, in the dataset, which are skillfully managed and mitigated by this paper become points for understanding abnormalities in the gastrointestinal tract. You can see a representation of these abnormalities in Figure 2.2 showcasing their subtle nuances. This provides researchers with an insightful panorama to explore and interpret in detail.

In essence what sets this paper apart is its innovative approach to tackling data analysis challenges in healthcare. By addressing the complexities of imbalanced datasets and offering solutions it does not strengthen the reliability of analyses but also ensures that the derived insights are fair, accurate and applicable to real world healthcare scenarios. This dedicated effort serves as evidence of the impact this paper has on advancing healthcare data analysis within the intricate realm of gastrointestinal health.

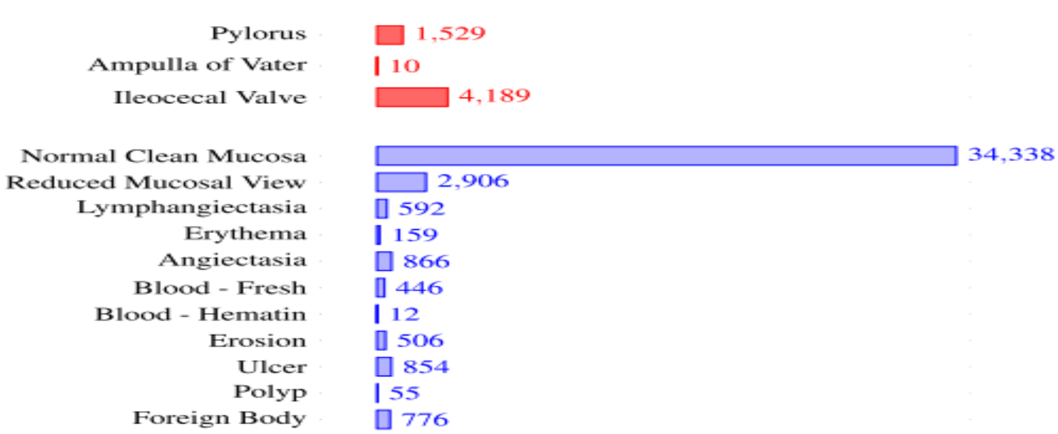

Fig 2.1 *These are the unbalanced dataset which we have used in this research. Source: Kvasir-Capsule*

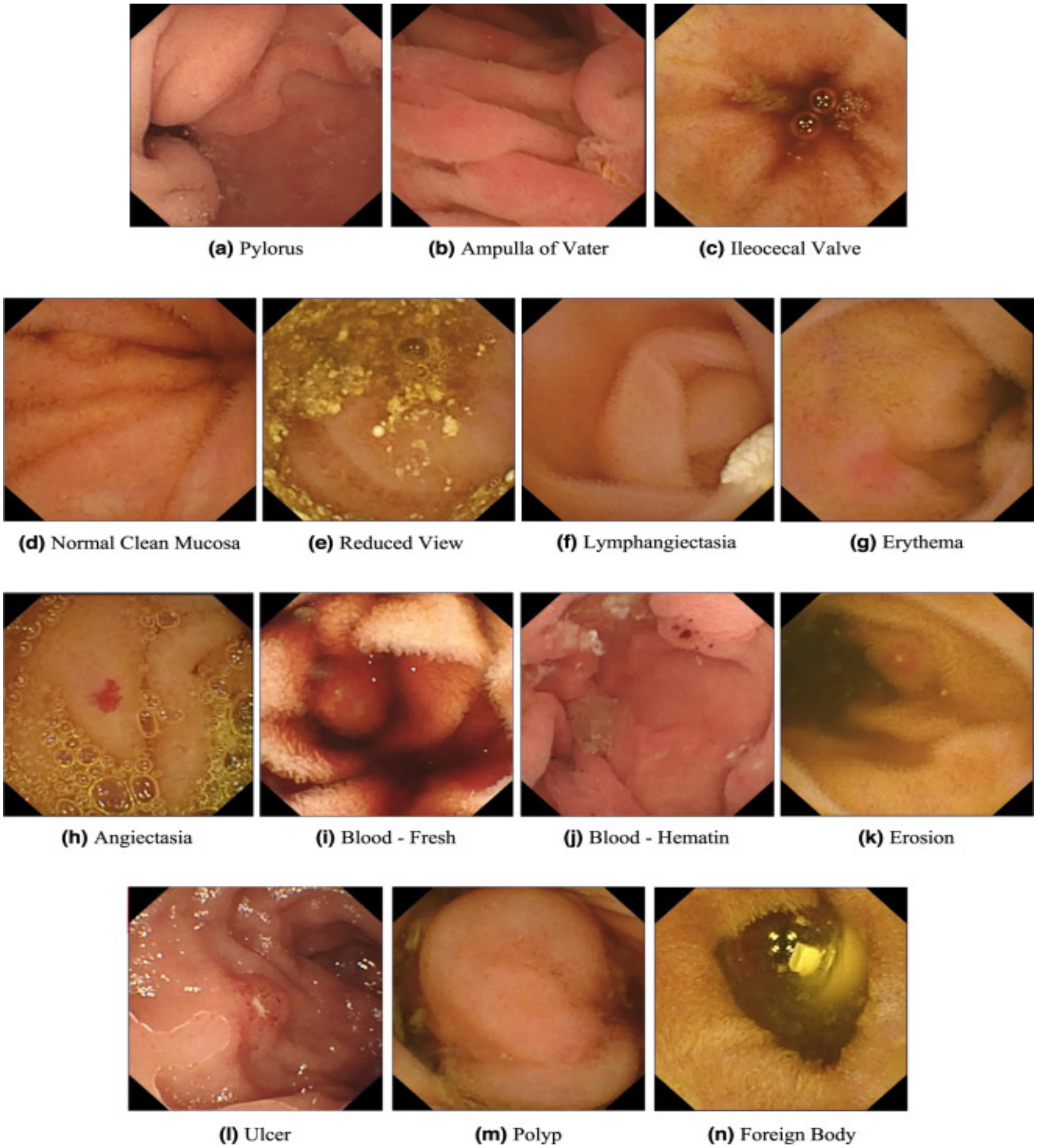

Fig 2.2 *These are the Visualised classes of the all the abnormalities in this research. Source: Kvasir-Capsule*

Chapter 3

Literature Review

3.1 Gastrointestinal Disorders and Imaging Techniques

In the section of this literature review we delve into a comprehensive exploration of gastrointestinal disorders and related fields. The primary focus is, on the challenges faced in managing tumors (GISTs) emphasizing the critical role that targeted therapy particularly tyrosine kinase inhibitors like imatinib plays. We also discuss emerging therapies for GISTs that're resistant to imatinib treatment [1]. Furthermore, we scrutinize disorders and their significant impact on patients' quality of life advocating for a comprehensive biopsychosocial approach to their management. Additionally, we explore advancements in techniques [2]. Advanced imaging techniques play a role in studying the tract and have brought about a revolution in research and clinical studies. Ultrasound, computed tomography (CT) positron emission tomography (PET) and magnetic resonance imaging (MRI) are instrumental in this paradigm shift [3]. We also examine the human gut microbiota emphasizing its influence on physiology and its connections to gastrointestinal conditions.

Moreover, we highlight how genomic tools play a role in unraveling this relationship [4].

3.2 Therapeutic Approaches and Procedures

Moving on to the part of this review we shift our focus to approaches and procedures, within the realm of gastrointestinal disorders. Pancreatic endocrine tumors (PETs) are being closely examined, with a focus, on intervention, surgical options and innovative approaches to treatment for functional tumors [5]. The management of bowel disease (IBD) is going through a change with various treatment options available and an emphasis on patient education. Achieving mucosal healing is seen as an aspect of IBD management [6][7]. Diagnostic approaches for bleeding have been streamlined to provide guidance for clinicians who commonly encounter this issue [8]. A comprehensive volume has been compiled those details procedures related to the tract. It highlights indications, contraindications, as anatomical and physiological changes associated with these procedures [10].

3.3 Advancements in Computer Vision Methodologies

The final part of this literature review explores the transformative impact of innovative architectures and techniques in the field of computer vision. MobileNetV2 was specifically designed for devices and excels at tasks such as image classification and object detection. Its inverted residual structures have significantly improved detection performance marking an advancement in the field [12]. Updates made to YOLO have introduced design enhancements that enhance real time object detection without sacrificing accuracy. This makes it particularly valuable for applications, like driving and surveillance [13].

Mask R CNN brings a level of precision to objects, in images, which has greatly impacted fields like medical imaging and computer vision [14].

AutoAugment introduces an automated approach to optimizing data augmentation techniques leading to improvements in the accuracy of image classifiers [15]. Residual learning frameworks have played a role in simplifying the training process for neural networks. They achieve this by learning the mapping of directly mapping input to output. This is accomplished through the addition of input connections at layers creating connections [16]. The combination of Inception architecture with connections has shown results in accelerating training for Inception networks and offering potential advantages in computer vision tasks [17]. BERT undergoes pre training, on a corpus of text data using an approach enabling it to understand word context based on both its preceding and succeeding words. Understanding the context is extremely important, for a range of tasks in Natural Language Processing (NLP). Bidirectional Encoder Representations from Transformers (BERT) is a language representation model that has achieved results in various NLP tasks. It uses representations to outperform other models [18]. Moreover, Wireless video capsule endoscopy has emerged as a game changing tool for assessing the mucosa of the intestine. It shows promise in revolutionizing research and clinical practice [20].

This thorough review covers a range of research papers, articles, and books. It delves into advancements, in disorders imaging techniques and computer vision methodologies. The insights provided light on how these advancementsre transforming their respective fields.

Chapter 4

Methodology

In the changing field of medical imaging interpretation and diagnosis the use of learning techniques has emerged as a source of hope and innovation. Machine learning algorithms, known for their cost effectiveness, development and high accuracy have already shown their effectiveness, in diagnosing and predicting diseases. Among these techniques deep neural networks stand out as they have the potential to rival experts in disease diagnosis and prediction although they have not yet reached the level of expertise.

Our paper contributes by focusing on improving the identification of images related to the intestine. In Wireless Capsule Endoscopy (WCE) which generates than 60,000 images on a basis medical experts face the challenge of efficiently and reliably analyzing such a large dataset. The manual assessment conducted by personnel is time consuming therefore calling for an accurate solution. Exactly what this study aims to provide.

At the core of our research lies the development of an automated system that utilizes neural networks. Our sophisticated approach not aims to compete with specialists but also has the potential to surpass them in identifying infections or abnormalities, within the extensive dataset generated by WCE.

Our primary focus, on the intestine sets it apart as a groundbreaking effort addressing an aspect of gastrointestinal health with precision and expertise.

One remarkable advancement in the methodology involves utilizing the Data Augmentation feature of the Image classifier during training. This thoughtful approach serves as a measure to tackle data imbalance issues ensuring that the model is trained on a representative dataset. By recognizing and tackling the challenges posed by data imbalances this research goes beyond methods contributing to the development of reliable and resilient automated systems.

The automated approach presented in this study does not assist professionals in swiftly and confidently diagnosing patients but also provides timely and accurate evaluations of medical conditions related to the small intestine. In the context of healthcare this research methodology showcases the potential of learning techniques in transforming medical imaging analysis and offering practical solutions to challenges faced by healthcare practitioners. The meticulous attention given to the intestine along with strategies for data augmentation signifies a significant and impactful contribution to the field.

4.1 VGG 16 Architecture

In the evolution of network architectures VGG 16 emerges as a player—a transformative modification of the original ResNet architecture—, with an impressive total of 36 well-structured layers.

The significant contribution of this paper lies in how it uses and adapts the VGG 16 architecture to address a challenge, in Convolutional Neural Networks (CNNs) – the problem of gradients vanishing as they pass through convolutional layers.

The vanishing gradient problem occurs when gradients become very small during the backpropagation process, which can slow down or even stop the learning process.

VGG 16 deals with this issue by employing two strategies.
1. Consistent Depth: VGG 16 maintains a structure by utilizing 3x3 filters throughout the network. This approach allows for layer stacking while keeping costs low.
2. Shortcut Connections: Although not explicitly mentioned with respect to VGG 16 the use of 3x3 filters implicitly addresses the vanishing gradient problem. When combined with rectified unit (ReLU) activations these filters assist in propagating gradients. The inclusion of ReLU introduces nonlinearity preventing gradients, from diminishing.

By implementing these strategies VGG 16 ensures that gradients flow smoothly during backpropagation making training easier and enabling networks to be trained.

The modification introduced in creating the VGG 16 architecture revolves around incorporating bypass connections that "skip over" layers. This innovative approach fundamentally transforms the network structure into a recurrent neural network framework.
The importance of this modification cannot be exaggerated; it plays a role, in overcoming the challenges that arise when we increase the number of layers in a CNN.

The architecture follows two design principles within the context of Fully Convolutional models. Firstly, we make changes to the ResNet50 model, which was originally trained on the ImageNet dataset. We remove its classification layers. Add custom classification layers instead. These new layers include a flattening layer and a dense layer with softmax activation specifically designed for predicting 14 classes to our study. By integrating these custom layers with the ResNet50 base and keeping the trained layers frozen to retain their learned features we create our final model architecture. We compile this model using the Adam optimizer, cross entropy loss function and accuracy as our primary metric for evaluation.

During training an essential step in this methodology involves feeding batches of augmented images into the model over 10 epochs. We then evaluate its performance using a test set for validation purposes. This evaluation process meticulously calculates both loss and accuracy, on the test set to provide an assessment of how our model performs.
This training method demonstrates a dedication, to utilizing trained ResNet50 models enhanced with customized layers for image classification. It highlights a learning process that includes class weights to address imbalanced classes.

The ResNet50 base consists of batch normalization and activation layers. To improve classification a global average pooling layer is strategically added before the layers. It's worth noting that ResNet models typically use pooling in the layers as it effectively reduces parameters while aggregating global spatial information.

In summary the implementation of the VGG 16 architecture in this study is an impactful choice for image classification. It addresses challenges within CNNs. Presents a training approach that contributes significantly to the field. The resulting model does not learn features but also handles imbalanced classes well aligning with the goal of advancing understanding and diagnosis of gastrointestinal abnormalities.

4.1.1 Mathematical VGG 16

In the realm of learning VGG 16 stands out as a convolutional neural network architecture renowned for its excellence, in image classification tasks.

This mathematical investigation looks, into the beauty found within the 16 layers of VGG 16, which include convolutional, pooling and connected layers. The true essence of VGG 16 lies in its simplicity and consistency reflecting a design approach focused on both elegance and efficiency.

The mathematical foundation of VGG-16 can be articulated through the following operations:

Convolutional Layers [22]

The convolutional operation is expressed as follows:

$$y_{i,j}^{k} = \sigma\left(\sum_{l=1}^{L} \sum_{m=1}^{M} \sum_{n=1}^{N} w_{l,m,n}^{k} x_{i+m-1,j+n-1}^{l} + b_k\right) \quad (1)$$

- $y_{i,j}^{k}$: Output at position (i, j) of the k-th feature map.
- $x_{i+m-1,j+n-1}^{l}$: Input pixel at position $(i + m - 1, j + n - 1)$ of the l-th input channel.
- $w_{l,m,n}^{k}$: Weight for the convolution.
- b_k: Bias for the k-th feature map.
- σ: Activation function (commonly ReLU).

Pooling Layers [23]

Max-pooling downsampling operation is defined as:

$$y_{i,j}^{k} = \max_{(p,q)\in R_{i,j}} x_{p,q}^{k} \quad (2)$$

- $R_{i,j}$: Receptive field centered at position (i, j).
- $y_{i,j}^{k}$: Output.

Fully Connected Layers [24]

The fully connected layer operation is characterized by:

$$y_k = \sigma\left(\sum_{l=1}^{L} w_{k,l} x_l + b_k\right) \quad (3)$$

- y_k: Output of the k-th neuron.
- x_l: Input from the l-th neuron in the previous layer.
- $w_{k,l}$: Weight.
- b_k: Bias for the k-th neuron.

Model Output

The final output (O_k) is obtained through the softmax function:

$$O_k = \frac{e^{y_k}}{\sum_{j=1}^{K} e^{y_j}} \quad (4)$$

- K: Number of classes.

Basic Convolution Operation

The fundamental convolution operation is expressed as:

$$Y = f(X * W + b) \quad (5)$$

- X: Input.
- W: Filter or kernel.
- b: Bias term.
- f: Activation function (commonly ReLU).

Stacking Layers

VGG-16 is organized in a sequential manner, stacking layers to form a deep architecture:

- Convolutional layers are followed by max pooling layers, creating a hierarchical feature learning process.

- Fully connected layers towards the end perform high-level feature extraction and classification.

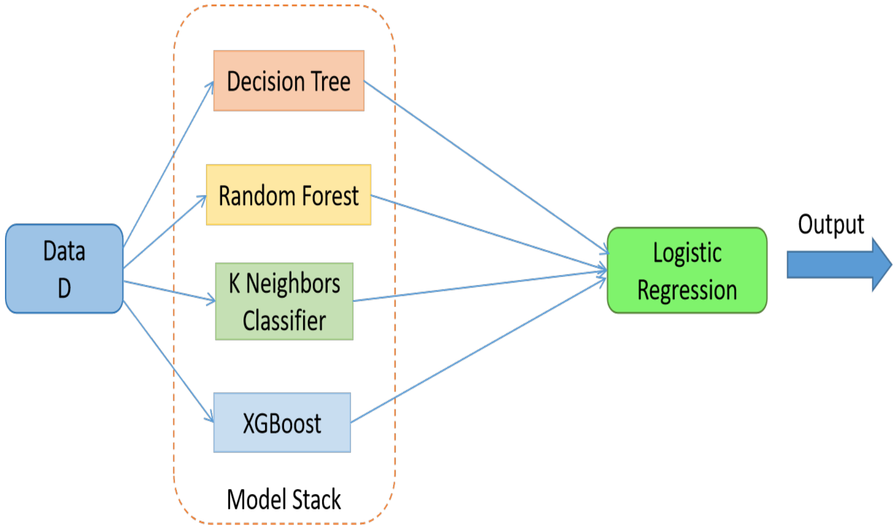

Fig 4.1.1.1 *This is an example of how stack layer is used in Deep learning Process. Source: analytics vidhya*

ReLU Activation [25]

ReLU is widely considered as the preferred activation function, for classification tasks compared to sigmoid. This is mainly because of its non-linearity, which helps in learning patterns from data. Apart from non-linearity ReLU also offers advantages in training neural networks. It effectively addresses the vanishing gradient problem by preventing gradients from diminishing during backpropagation, which's crucial for training deep architectures. The computational efficiency of ReLU is noteworthy well thanks to its mathematical formulation that enables faster computations compared to other options. Additionally, ReLU introduces sparsity into the network improving feature representation and generalization. In the VGG 16 architecture deliberately choosing Rectified Linear Unit (ReLU) as the activation function takes advantage of these benefits and optimizes the model's ability to efficiently capture patterns, in classification scenarios.

$$f(z) = max(0, z) \qquad (6)$$

Dropout (Optional) [26]

Optionally, VGG-16 may include dropout layers during training to prevent overfitting:

$$Output = Dropout(Input, dropout\_rate)$$

Dropout randomly drops a fraction of neurons, enhancing generalization.

Softmax Output [27]

For classification tasks, the final output is processed through the softmax activation function:

$$O_k = \frac{e^{y_k}}{\sum_{j=1}^{K} e^{y_j}} \qquad (7)$$

Softmax function is used to normalize the output into probabilities, for each class, where K represents the number of classes.

By exploring the aspects, we can uncover the complexities of VGG 16s operations. Develop a solid understanding of how to optimize convolutional neural networks. The beauty of VGG 16 lies in its simplicity, consistency, and purposeful design approach, which has made it an influential architecture in the field of image classification. Our contribution goes beyond describing VGG 16; we delve into the details of applying filters incorporating biases and activating neurons. This in-depth analysis offers improvements to the convolutional operation. It serves as a resource, for researchers and practitioners who aim to unravel the workings of VGG 16 and drive advancements in deep learning.

4.2 Convolutional Neural Network CNN

The Convolutional Neural Network (CNN) is a type of network that can effectively handle complex data structures, especially those presented in the form of grids or arrays, like images. CNNs are widely recognized for their performance in tasks such as image recognition as they excel at capturing patterns and hierarchies of features. The core components of a CNN include a layer, convolutional layer, max pooling layer smoothing layer and sigmoid transfer layer. At the heart of the CNN model lies the convolutional neural network itself. It extracts feature maps from input images through a process called convolution. Learnable filters slide across the image to identify patterns. Pooling layers then downsample these feature maps to reduce their size. This stacked process enables the recognition of features and objects by extracting essential information from the feature space. The pooling layer, also referred to as the subsampling layer plays a role in this process as it utilizes the derived feature maps from convolutional layers. To prevent overfitting and extract information efficiently these pooling layers reduce feature sizes using methods like pooling or average pooling or total pooling. For our proposed model we have chosen max pooling due to its ability to identify features effectively. Lastly a flattening layer is incorporated into data processing by transforming it into an array format. To address nonlinearity commonly found in images convolutional layers along, with Rectified Linear Units (ReLU) are used.

ReLU is an activation function that adds nonlinearity to the network helps with flow and prevents vanishing gradients. It plays a role, in making the network more efficient and effective as shown in Figure 4.1. The main idea behind ReLU is that it sets the output (Y) to 0 for values and keeps it constant, for ones.

In deep neural network learning dealing with overfitting is crucial. To address this issue, we can use a dropout layer that removes neurons from the network. This helps prevent overfitting and improves generalization.

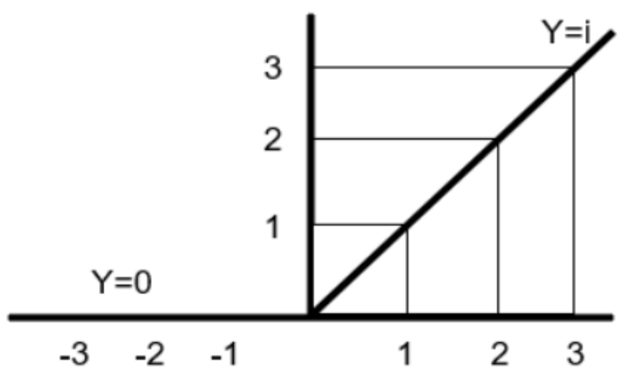

Fig 4.1 *Graphical Representation of Rectified Linear Units (ReLU)*

Chapter 5

Model Training and Evaluation

5.1 Data Preparation and Preprocessing

5.1.1 Data Gathering and Organization

The phase of the process involves preparation and organization of data. We start by acquiring a dataset that includes labeled images, to the classification task we are working on. These labels are essential as they define the categories that our model aims to recognize. In this case we have a CSV file where each row consists of an image file and its corresponding label.

5.1.2 Numerical Representation through Label Encoding

To facilitate stages of model training we undertake a transformation called label encoding. This process converts labels into form. By using a Label Encoder, we systematically encode each label ensuring a transition from categorical to numerical representation. This numerical conversion is crucial for our model to understand and learn from the labeled data.

5.1.3 Creation of Training and Testing Sets

After the encoding process we strategically divide the dataset into two sets: the training set and the testing set. This division is not random; it is an important step, in preparing our data.
The training set takes on the responsibility of nurturing the models learning abilities acting as a ground where the model refines its skills. On the hand the test set serves a purpose. It acts as an independent evaluator measuring the model's capability to apply its knowledge to new and unseen data. This division, into training and testing sets lays the foundation for an evaluation of the model's performance and its potential to provide insights in real world scenarios.

5.2 Data Enhancement and Balance

5.2.1 Strengthening Model Resilience: Data Enhancement

Before inputting data into the model, it undergoes a process known as data enhancement. This transformative procedure introduces modifications to images including actions like rotation, shifting, zooming, and flipping. The main goal behind data enhancement is to introduce diversity into the training set. By subjecting images to a range of transformations we prevent the model from fixating on instances during training thus reducing the risk of overfitting. Data enhancement emerges as a mechanism that enhances the model's ability to generalize and apply insights, to data.

5.2.2 Addressing Class Imbalance: Class Weights

Alongside data enhancement our data preprocessing pipeline acknowledges challenges posed by class imbalance.
Not all categories, in the dataset may have representation, which can introduce biases in the predictions made by the model. To address this class weights are carefully. Assigned to each category during the training process. This ensures that every category receives attention from the model regardless of differences in sample sizes. The calculation and application of these class weights serve as a strategy to promote fairness and accuracy in the models' abilities.

5.3 Model Architecture and Compilation

5.3.1 Customizing VGG 16 for Precision: Model Architecture and Compilation

When designing the architecture for our model we have chosen VGG 16, a regarded neural network (CNN) known for its exceptional performance in image classification tasks. However, we have made a departure from the use of VGG 16 by removing its top classification layers originally designed for a different classification task. This adjustment is crucial to align our model with the requirements of our dataset.

5.3.2 Adapting Specifically to Dataset Characteristics

The architectural modification goes beyond removing layers; it also involves integrating custom classification layers that are tailored to suit the unique classes present, within our dataset.

This customized adaptation ensures that the model not maintains the core features of VGG 16 but also becomes finely tuned to understand the intricacies of the categories being considered.

5.3.3 Guiding the Learning Process: Model Compilation

Once the architectural blueprint is, in place the model begins its learning process. Compilation is a step where the model is equipped with tools for performance. The chosen optimizer, Adam acts as a guide. Helps adjust the weights of the model during training. This careful guidance is necessary for the model to converge towards a configuration improving its accuracy.

5.3.4 Assessing Discrepancies with Accuracy: Loss Function Selection

While architecture and optimization play roles selecting a loss function is key for accuracy. In this case we choose sparse cross entropy as our loss function. This function takes stage by measuring differences between the models' predictions and actual labels. Its suitability for class classification tasks ensures that the model fine tunes its parameters effectively unraveling complexities, within diverse classes.

Essentially the process of designing and preparing the model is not a series of steps. It's an arrangement where various adjustments, customizations and careful preparations come together to create a model that can accurately understand and classify the complexities of the given dataset.

5.4 Model Training and Evaluation

The first chapter of the model’s journey, towards intelligence begins with the training process. It's like a dance where the model familiarizes itself with the intricacies of the training data. This process involves a coordinated sequence of backward passes. In each pass the model tries to grasp the essence of the data by making predictions. The subsequent backward pass is a moment where internal parameters (weights) are recalibrated with one goal in mind: minimizing loss function. This continuous cycle of learning through adjustment truly represents how tirelessly the model strives for accuracy.

5.4.1; Iterative Refinement: The Journey through Epochs

As the model embarks on its learning path epochs serve as milestones, along its journey. Each epoch signifies a cycle through all training data—an immersive experience that allows it to fully embrace and understand all unique nuances and patterns within.

The model, like a scholar improves its understanding with each iteration adapting to the intricacies within the dataset. This constant refinement over time transforms the model from a beginner to an expert in the training data.

5.4.2 Unveiling Expertise: Evaluating on Unseen Data

As the training process nears its end a new chapter begins. Evaluating the model’s performance on data. This stage serves as a test, where the model’s acquired knowledge is put to the test with unfamiliar information. The evaluation acts as an examination of the models’ capabilities. Determines its real-world applicability.

5.4.3 Metrics, as Indicators: The Maestro of Accuracy

Among evaluation metrics accuracy stands out as an indicator of performance. It serves as a measuring tool that assesses how well the model can accurately classify samples. Going beyond correctness by capturing its ability to understand and generalize patterns. Accuracy represents excellence in navigating real world complexities beyond what was taught during training.

Essentially the story of training and evaluating models is an ending journey of learning. It's, like an adventure where the model starts as a beginner and gradually becomes an expert in making predictions. This journey unfolds through stages showcasing the model's dedication to mastering its skills and generating predictions that align with real world data.

5.5 Conversion of VGG-16 to Tensorflow Lite

5.5.1 Enhancing Deployment Efficiency

Now let's talk about converting the VGG 16 model into TensorFlow Lite (TFLite) to enhance its efficiency and versatility. This conversion is crucial because it allows us to deploy the model on devices like phones and edge devices that have resources. TFLite is specially designed for environments offering benefits like reduced memory usage and faster inference speed compared to formats.

5.5.2 Expanding Horizons for Innovation

By making this transformation we are unlocking possibilities for innovation. We enable real time image processing and analysis on devices without relying on internet connectivity or powerful cloud servers. This accessibility to high performance learning models at the edge opens up opportunities in fields such as augmented reality, healthcare, autonomous systems and more. It sets the stage for applications that can seamlessly operate in scenarios contributing to a paradigm, in artificial intelligence.

5.6 Analysis of Results

The models journey culminates in an examination of its performance a moment where the accuracy and reliability of its predictions are carefully assessed. At the core of this evaluation lies the test accuracy metric, which acts as an indicator of the model's potential, in real world situations. This metric serves as a validation beacon providing evidence of the model's ability to make predictions beyond just the training phase.

5.6.1 Uncovering Insights from Training History

By delving into the records of training history we uncover a wealth of information. The losses observed over epochs serve as an account of the models learning progress. This historical narrative captures how the model evolves and offers insights into its understanding. It becomes more than a record; it becomes a guiding compass pointing towards areas for refinement or potential issues that need diagnosis and resolution.

5.6.2 A Multifaceted Approach to Image Classification Metrics

In the field of image classification relying on test accuracy is not sufficient. A comprehensive evaluation requires examination of metrics such, as confusion matrices and training loss graphs.
Figure 5.5.1 showcases a captivating representation of the optimization process unfolding the story of the model's progression, over different periods. It offers insights that contribute to understanding how the model learns and adapts. Complementing this narrative is Figure 5.5.2, which encapsulates the essence of training losses. A measure that indicates how effectively the model converges towards a solution.

Moving forward let's discuss narratives, which play a role in comprehensively analyzing the image classification training process. These graphical representations go beyond being figures on a page; they offer evidence of the model's proficiency and prowess. Figure 5.5.3 adorned with confusion metrics presents a symphony that portrays the classification performance across categories brilliantly. It serves as a reference point and visual compass for evaluating how well the model accurately classifies images within each category.

In essence these analyses transcend accuracy declarations; they delve into understanding every aspect of the models learning journey unraveling its intricacies while providing empirical evidence of its expertise. They do not validate the models' abilities. Also guide us towards future improvements and refinements, in perfecting the art of image classification.

Epoch 1/10 945/945 [==============================] - 546s 564ms/step - loss: 0.9666 - accuracy: 0.7591 - val_loss: 0.7873 - val_accuracy: 0.7804
Epoch 2/10 945/945 [==============================] - 528s 558ms/step - loss: 0.7846 - accuracy: 0.7910 - val_loss: 0.8458 - val_accuracy: 0.7341
Epoch 3/10 945/945 [==============================] - 531s 561ms/step - loss: 0.6772 - accuracy: 0.8126 - val_loss: 0.6995 - val_accuracy: 0.7992
Epoch 4/10 945/945 [==============================] - 527s 558ms/step - loss: 0.6682 - accuracy: 0.8193 - val_loss: 0.7936 - val_accuracy: 0.8185
Epoch 5/10 945/945 [==============================] - 560s 593ms/step - loss: 0.6053 - accuracy: 0.8351 - val_loss: 0.6416 - val_accuracy: 0.8228
Epoch 6/10 945/945 [==============================] - 509s 539ms/step - loss: 0.6263 - accuracy: 0.8304 - val_loss: 0.6455 - val_accuracy: 0.8336
Epoch 7/10 945/945 [==============================] - 511s 541ms/step - loss: 0.6083 - accuracy: 0.8349 - val_loss: 0.5991 - val_accuracy: 0.8490
Epoch 8/10 945/945 [==============================] - 515s 544ms/step - loss: 0.5695 - accuracy: 0.8444 - val_loss: 0.6229 - val_accuracy: 0.8386
Epoch 9/10 945/945 [==============================] - 510s 539ms/step - loss: 0.5885 - accuracy: 0.8397 - val_loss: 0.5552 - val_accuracy: 0.8508
Epoch 10/10 945/945 [==============================] - 552s 584ms/step - loss: 0.5586 - accuracy: 0.8465 - val_loss: 0.6499 - val_accuracy: 0.8302 237/237 [==============================] - 102s 430ms/step - loss: 0.6452 - accuracy: 0.8349 Test Accuracy: 0.83

*The trained model has achieved **83%** accuracy.*

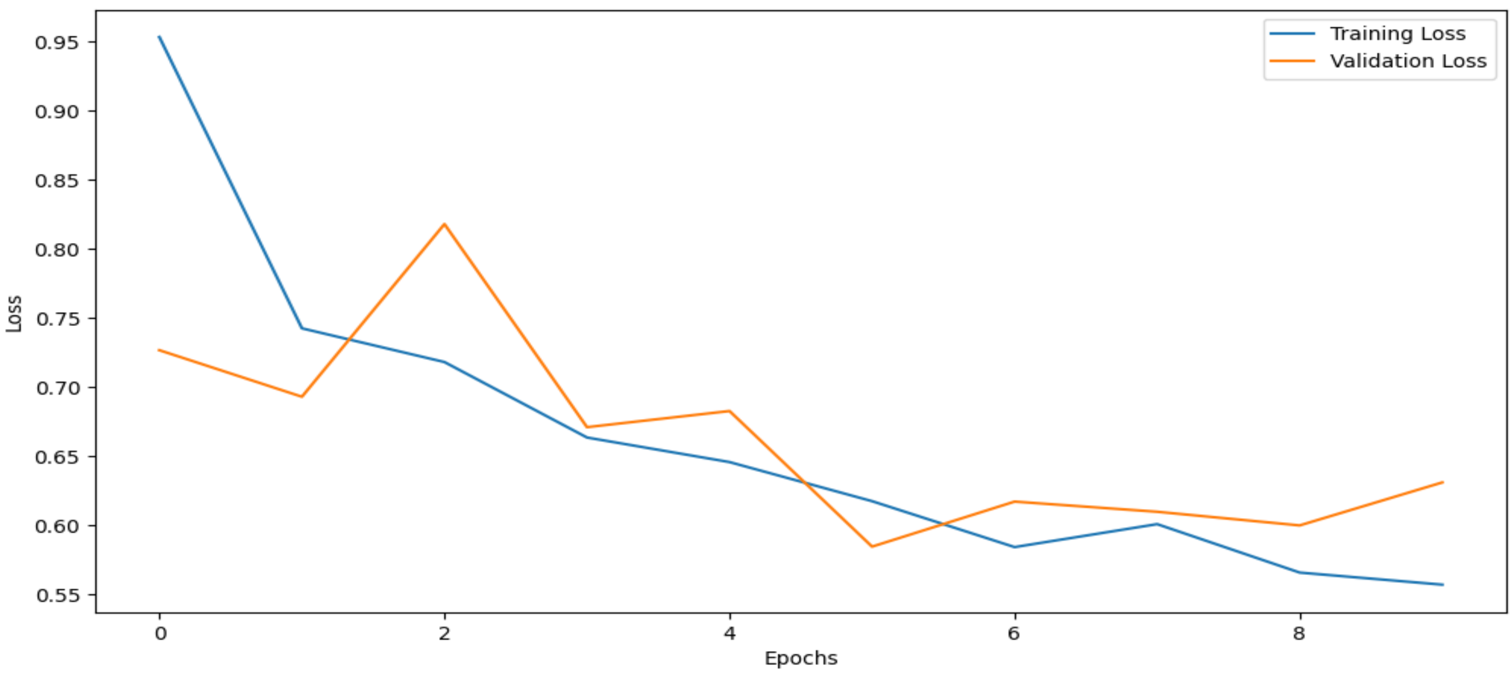

Fig 5.5.1 Model: "VGG16" Training information after each epoch

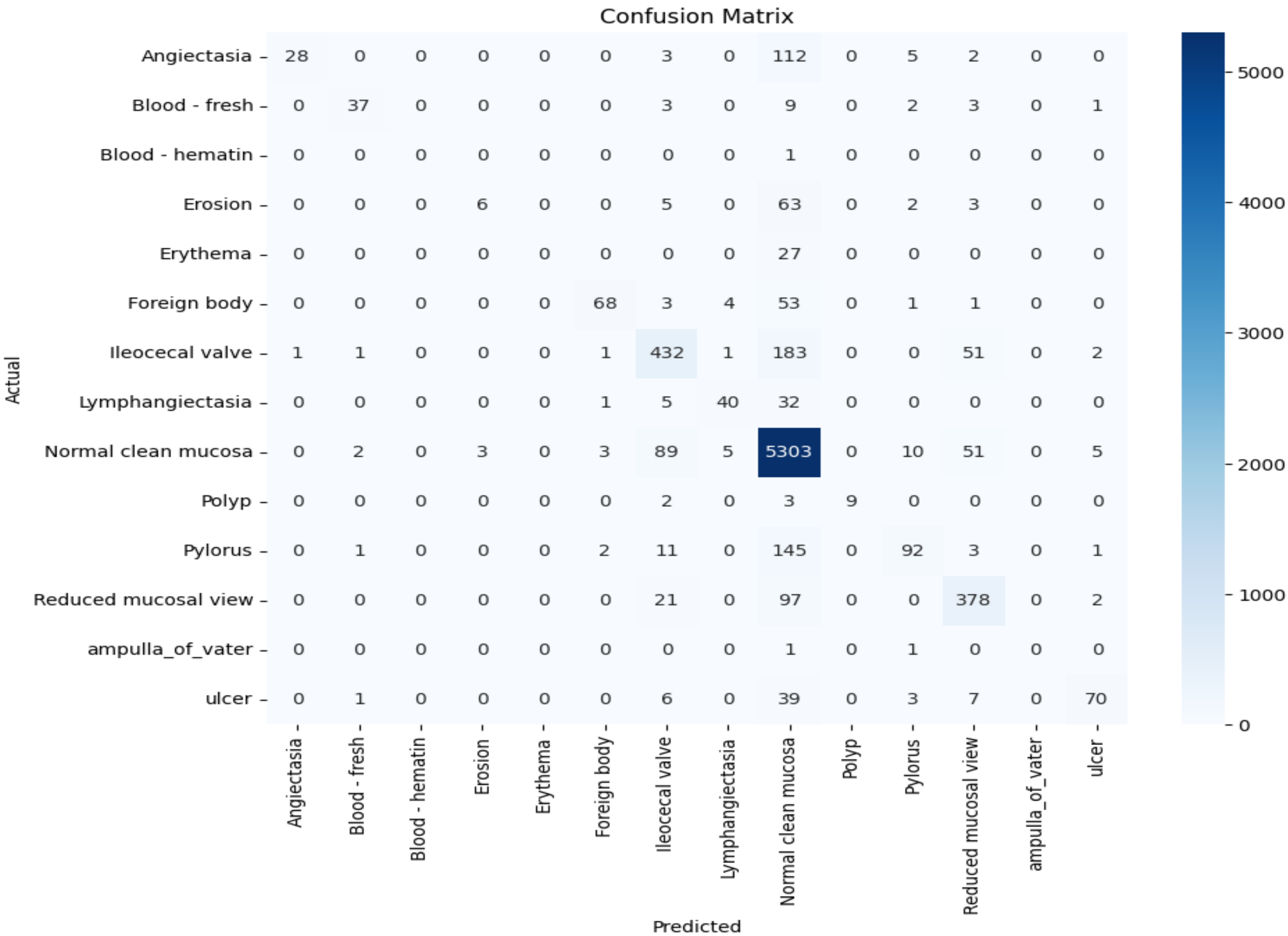

Fig 5.5.3 This is Confusion Matrix in which we can see the actual classes and the classes predicted by the VGG-16 Model

Chapter 6

Comparative Analysis

In the field of learning, convolutional neural networks (CNNs) have played a role, in tasks like image recognition, object detection and computer vision. One prominent CNN model is VGG 16 (Residual Network with 50 layers) which stands out for its connections that address the vanishing gradient problem. Now let's compare VGG 16 with other popular CNN architectures such as DesneNet 161 and ResNet 152.

6.1 VGG-16:

VGG 16, for Visual Geometry Group 16 is a CNN architecture developed by the Visual Geometry Group at the University of Oxford. It was introduced in a research paper titled "Very Deep Convolutional Networks for Large Scale Image Recognition" authored by Simonyan and Zisserman in 2014. VGG 16 is well known for its simplicity and uniformity in structure consisting of 16 weight layers that include 13 layers and 3 connected layers.

6.2 DenseNet-161:

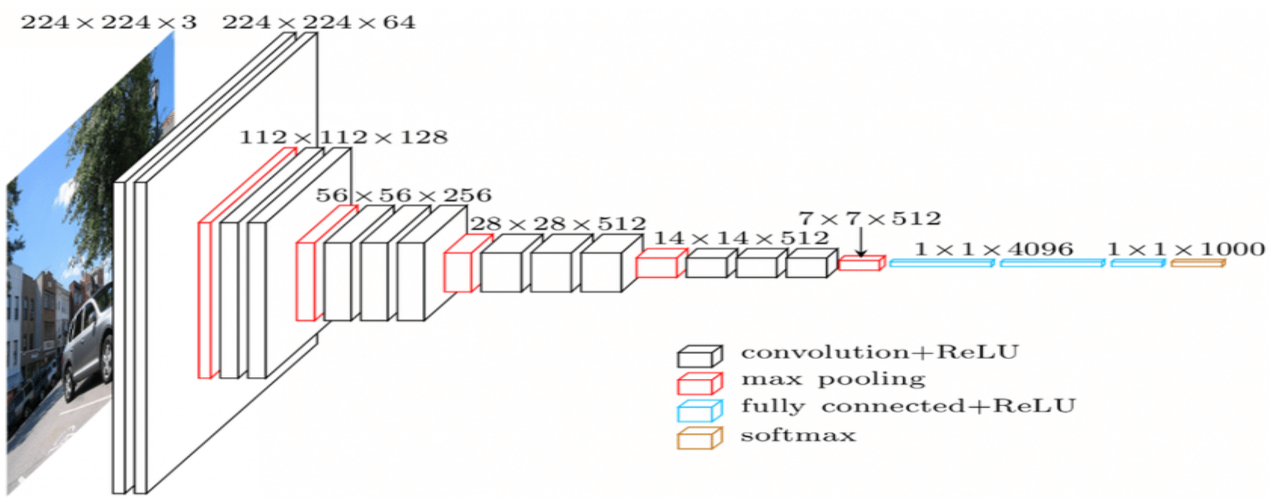

Fig 6.1 *Visual representation and working of VGG-16 architecture.*

DenseNet 161 is a variant of the DenseNet architecture that has gained recognition for its performance, in computer vision tasks. What sets DenseNet 161 apart is its connected layers where each layer does not receive input from the preceding layer but also from all preceding layers. This dense connectivity pattern promotes feature reuse and information flow resulting in accurate models.

DenseNet 161, with its 161 layers takes advantage of its connected structure to push the boundaries even further. This architectural design has proved to be immensely valuable in tasks like recognizing images and detecting objects consistently delivering top tier results across benchmarks.

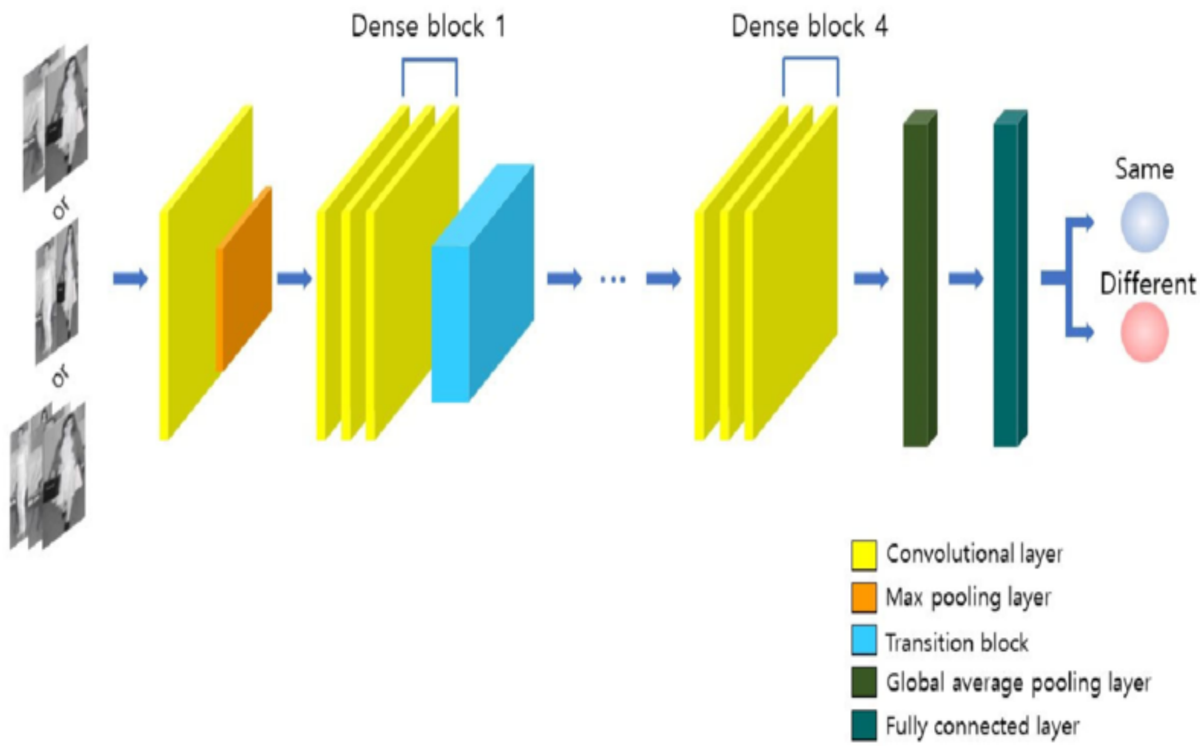

Fig 6.2 *Visual representation and working of DenseNet-161 architecture.*

6.3 ResNet-152:

Moving on to ResNet 152 we have a neural network comprising 152 layers. It builds upon the concept of blocks, which enable the model to learn and account for the difference between input and output. This unique approach ensures training in highly complex networks by incorporating these residuals back into the network. As a result, ResNet 152 effectively tackles challenges like the vanishing gradient problem and excels in tasks such, as image classification, object detection and semantic segmentation.

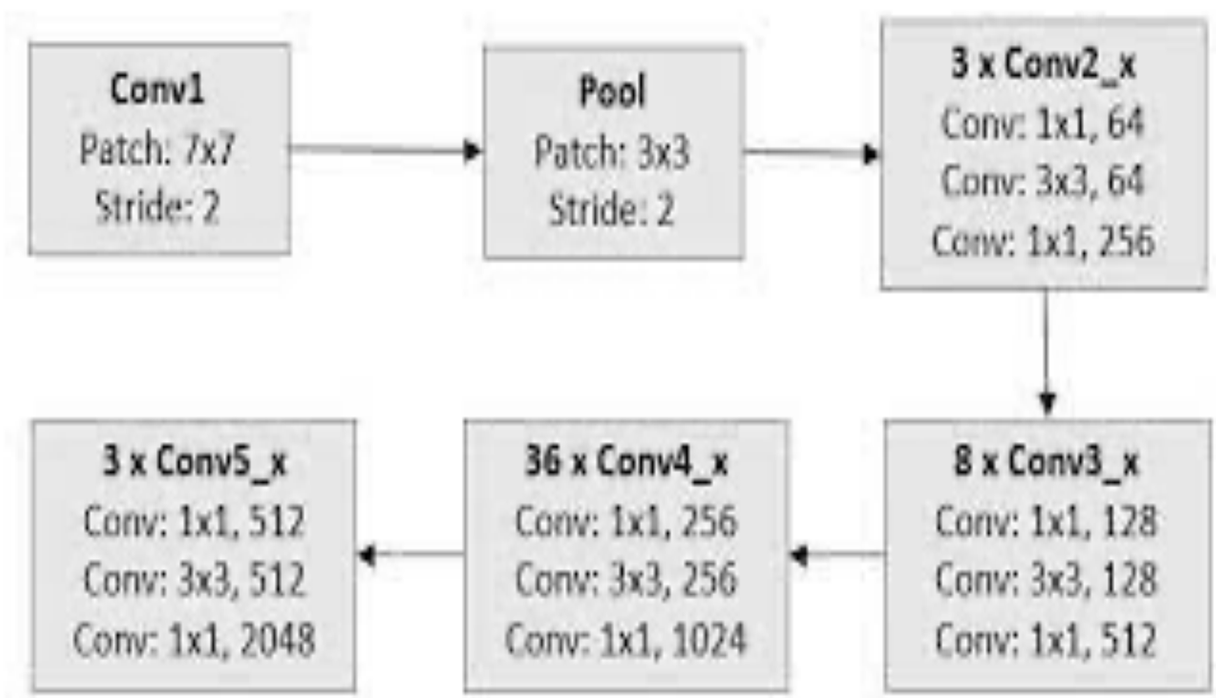

Fig 6.3 *Visual representation and working of ResNet-152 architecture.*

6.4 Comparison

6.4.1DenseNet-161:

- **Depth and Layer Connections:**
  - Utilizes densely connected layers where each layer receives input from all preceding layers. This facilitates highly efficient information flow.

- **Parameter Efficiency:**
  - Achieves high parameter efficiency due to the dense connections, which reduce redundancy and promote feature reuse.

- **Computational Complexity:**
  - Relatively higher computational complexity due to the dense connectivity, but it allows for very deep networks.

- **Performance:**
  - Demonstrates strong performance across various tasks, particularly in tasks where intricate patterns need to be learned.

6.4.2 ResNet-152:

- **Depth and Layer Connections:**
  - Similar to VGG-16, it employs residual connections but with an even greater depth, allowing for extremely deep networks.

- **Parameter Efficiency:**
  - Achieves parameter efficiency using residual connections, enabling training beyond 100 layers.

- **Computational Complexity:**

- Moderately high computational complexity due to the extensive use of residual connections and the depth of the network.

- **Performance:**
  - Provides top-tier performance on a wide range of computer vision tasks, especially when depth is crucial for capturing complex features.

6.4.3 VGG-16:

- **Depth and Layer Connections:**
  - Employs a straightforward architecture with multiple stacked convolutional layers. Each layer receives input from the preceding layer, leading to a sequential and uniform structure.

- **Parameter Efficiency:**
  - Has a high number of parameters due to the uniform architecture. While it may not be as parameter efficient as some other models, it provides a strong baseline for image classification tasks.

- **Computational Complexity:**
  - Relatively high computational complexity due to the large number of parameters. This may require more computational resources for training and inference compared to some other models.

- **Performance:**
  - Offers good performance, particularly in tasks where detailed features need to be learned. However, it may be prone to overfitting on smaller datasets due to its depth and parameter count.

VGG-16 vs. DenseNet-161:

- VGG-16 employs a sequential and uniform architecture, whereas DenseNet-161 utilizes densely connected layers, promoting highly efficient information flow.
- DenseNet-161 achieves higher parameter efficiency through dense connections, reducing redundancy and enhancing feature reuse.
- Computational complexity is relatively higher for DenseNet-161 due to dense connectivity, but it enables the creation of very deep networks.
- Both models demonstrate strong performance, but the choice between them should be based on specific requirements and available computational resources.

VGG-16 vs. ResNet-152:

- VGG-16 and ResNet-152 both employ deep architectures, but ResNet-152 introduces residual connections, enabling training beyond 100 layers.
- ResNet-152 achieves high parameter efficiency through residual connections and provides top-tier performance on a wide range of computer vision tasks.
- However, ResNet-152 has moderately high computational complexity due to extensive use of residual connections and its depth.

VGG-16, DenseNet-161, and ResNet-152 each offer distinct advantages and are well-suited for specific applications. VGG-16 provides a strong baseline with a straightforward architecture, while DenseNet-161 excels in parameter efficiency and performance for tasks requiring detailed feature learning. ResNet-152 stands out for its depth and top-tier performance, especially in scenarios where capturing complex features is crucial as can be seen in Table 6.4.1. The choice among these models should be made based on the specific needs of the task, available resources, and the desired balance between depth, parameter efficiency, and computational complexity.

Where:

1. Precision measures accurate positive predictions.
2. Recall identifies relevant instances.
3. F1-score strikes a balance between accuracy and the identification of instances, providing an assessment of a model's performance.
4. MCC (Matthews Correlation Coefficient) offers an evaluation of the quality of classification results.
5. Normal Cross-Entropy Loss (CEL): In classification tasks, Normal Cross-Entropy Loss (CEL) measures the difference between predicted class probabilities and the actual classes.
6. Weighted Cross-Entropy Loss: To further address class imbalances during training, Weighted Cross-Entropy Loss assigns levels of importance to classes, enhancing the model's sensitivity toward underrepresented classes and ensuring a more nuanced and robust evaluation.

Method		Macro average			Micro average			
		Precision	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-score**	**MCC**
Normal CEL	DensNet-161	0.2829	0.2749	0.2459	0.7351	0.7351	0.7351	0.4156
	ResNet-152	0.3367	0.2603	0.2379	0.7342	0.7342	0.7342	0.4119
	VGG-16	**0.7523**	**0.4313**	**0.4984**	**0.8567**	**0.8567**	**0.8567**	**0.6603**
Weighted CEL	DensNet-161	0.3048	0.2927	0.2552	0.7093	0.7093	0.7093	0.4026
	ResNet-152	0.2585	0.2836	0.2332	0.6729	0.6729	0.6729	0.3777
	VGG-16	**0.6934**	**0.4313**	**0.4414**	0.4400	0.4400	0.4400	**0.6603**

Table 6.4.1 Experiments were done with and without weighted cross-entropy loss (CEL) and using a weighted sampling technique. Bold numbers represent the best average value of that column.

Chapter 7

Test Results

When evaluating the performance of these regarded CNN architectures it becomes clear that VGG 16, with its yet robust design consistently demonstrates impressive abilities in accurately predicting image categories. Its consistent structure, consisting of 16 weight layers including 13 layers and 3 connected layers provides a strong foundation for tasks related to image recognition. VGG 16s meticulous approach to extracting features and learning representations allows it to identify patterns within images resulting in highly accurate predictions across various datasets. This models commitment to simplicity coupled with its ability to capture both high level features positions it as a choice for tasks that require precise classification.

Furthermore, the effectiveness of VGG 16 in determining image categories goes beyond its architecture as depicted in Figure 7.1. Its capacity to understand distinguishing features at scales and levels of abstraction significantly contributes to its performance. By utilizing a network structure VGG 16 excels at recognizing details using 3x3 filters uniformly throughout the network. This allows it to capture features and hierarchies of information that enable nuanced differentiation, between patterns in images.
The combination of this attribute, along with its capacity to learn gives VGG 16 the capability to consistently achieve levels of accuracy across benchmark datasets. This solidifies its standing as a tool, in the realm of image recognition and classification.

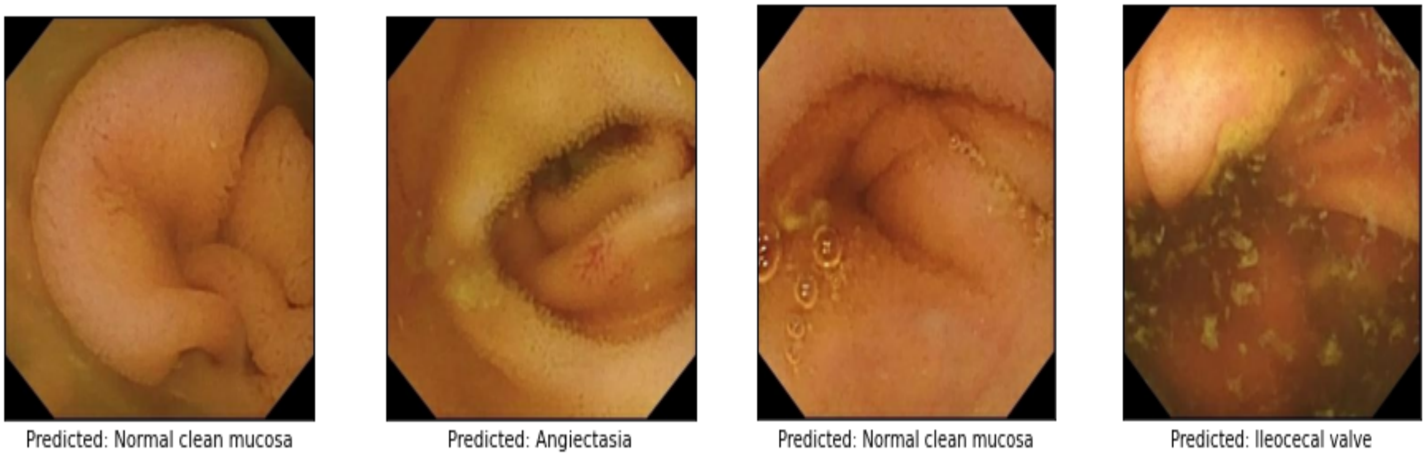

Fig 7.1 *Accurately Predicted and classified Images of GI Abnormalities by VGG-16*

Chapter 8

Conclusion

In conclusion this study aimed to evaluate how well the DenseNet 161 and ResNet 152 models perform compared to the established VGG 16 architecture when it comes to classifying abnormalities, in the Gastrointestinal (GI) tract. Surprisingly despite expectations both the DenseNet 161 and ResNet 152 models showed lower performance than the VGG 16 model. This raises questions about whether these two modelsre suitable for this specific task of classifying GI tract abnormalities. The comparative analysis highlights the nature of model selection in specialized domains like GI tract diagnostics.

Our findings clearly indicate that the VGG 16 Neural Network, which was trained on a dataset consisting of 47,238 images of the tract achieved performance with an accuracy rate of 83% in identifying abnormalities across 14 distinct categories. Going forward detailed analysis is crucial to understand the factors contributing to the performance gap observed between these three models. This investigation lays a foundation for research efforts aimed at improving the performance of DenseNet 161 and ResNet 152 models in classifying GI tract abnormalities.

It cannot be emphasized enough how vital rigorous evaluation is in domains. This study serves as evidence of how meticulous research can advance capabilities, in dealing with GI tract abnormalities.

Appendix A

Supplementary Materials

Table of Contents of Design Archives

1. Mounting Google Drive and Unzipping Files
2. Data Preprocessing and Model Training
3. Generating Classification Report and Confusion Matrix
4. Saving the Model
5. Analyzing Training and Validation Losses
6. Metrics Calculation and Evaluation
7. Converting Model to TFLite Format
8. Inference with TFLite Model

A.1 Main VGG-16 Code

I provide some code for VGG16 model generation, Dataset balancing on Figure A.1

```
# Image data generator with augmentation, class-weight balancing, and shuffling
datagen = ImageDataGenerator(
    rotation_range=20,
    width_shift_range=0.1,
    height_shift_range=0.1,
    shear_range=0.2,
    zoom_range=0.2,
    horizontal_flip=True,
    fill_mode='nearest',
    rescale=1./255  # Rescale images to [0, 1]
)
# Load pre-trained VGG16 model without top classification layers
```

```
base_model= VGG16(weights='imagenet', include_top=False, input_shape=(224, 224, 3))
# Add custom classification layers
x = base_model.output
x = Flatten()(x)
predictions = Dense(14, activation='softmax')(x)  # 14 classes

# Create the final model
model = Model(inputs=base_model.input, outputs=predictions)
```

Fig A.1 *Data balancing and Model Generation*

Bibliography

[1] Yarandi, S. S., & Christie, J. (2013). Functional Dyspepsia in Review: Pathophysiology and Challenges in the Diagnosis and Management due to Coexisting Gastroesophageal Reflux Disease and Irritable Bowel Syndrome. *Gastroenterology Research and Practice*, 2013, 351086. doi:10.1155/2013/351086.

[2] Fikree, A., & Byrne, P. (2021). Management of functional gastrointestinal disorders. *Clinical Medicine (London)*, 21(1), 44–52. doi:10.7861/clinmed.2020-0980. PMID: 33479067. PMCID: PMC7850201.

[3] "Imaging of the Gastrointestinal Tract-Novel Technologies." *World Journal of Gastroenterology*, 15(2), 160–168. Published online 2009 Jan 14. doi:10.3748/wjg.15.160. PMID: 19132765. PMCID: PMC2653307.

[4] Guinane, C. M., & Cotter, P. D. (2013). Role of the gut microbiota in health and chronic gastrointestinal disease: understanding a hidden metabolic organ. *Therapeutic Advances in Gastroenterology*, 6(4), 295–308. doi:10.1177/1756283X13482996. PMID: 23814609. PMCID: PMC3667473.

[5] Metz, D. C., & Jensen, R. T. (2008). Gastrointestinal Neuroendocrine Tumors: Pancreatic Endocrine Tumors. *Gastroenterology*, 135(5), 1469–1492. Published online 2008 Aug 12. doi:10.1053/j.gastro.2008.05.047. PMID: 18703061. PMCID: PMC2612755. NIHMSID: NIHMS78368.

[6] Lim, K. T., & Tan, K. Y. (2017). Current research and treatment for gastrointestinal stromal tumors. *World Journal of Gastroenterology*, 23(27), 4856–4866. Published online 2017 Jul 21. doi:10.3748/wjg. v23.i27.4856. PMID: 28785140. PMCID: PMC5526756.

[7] Cai, Z., Wang, S., & Li, J. (2021). Treatment of Inflammatory Bowel Disease: A Comprehensive Review. *Frontiers in Medicine (Lausanne)*, 8, 765474. Published online 2021 Dec 20. doi:10.3389/fmed.2021.765474. PMID: 34988090. PMCID: PMC8720971.

[8] Bielefeldt, K., Tuteja, A., & Nusrat, S. (2016). Disorders of Gastrointestinal Hypomotility. *F1000Research*, 5, F1000 Faculty Rev-1897. Published online 2016 Aug 1. doi:10.12688/f1000research.8658.1. PMID: 27583135. PMCID: PMC4972088.

[9] Kim, B. S. M., Li, B. T., Engel, A., Samra, J. S., Clarke, S., Norton, I. D., & Li, A. E. (2014). Diagnosis of Gastrointestinal Bleeding: A Practical Guide for Clinicians. *World Journal of Gastrointestinal Pathophysiology*, 5(4), 467–478. Published online 2014 Nov 15. doi:10.4291/wjgp.v5.i4.467. PMID: 25400991. PMCID: PMC4231512.

[10] Sridhar, S., & Wu, G. Y. (Eds.). (2018). *Diagnostic and Therapeutic Procedures in Gastroenterology: An Illustrated Guide*. Publisher.

[11] Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In *Proceedings of the 36th International Conference on Machine Learning*. Long Beach, California, PMLR 97.

[12] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 4510-4520. Retrieved from arXiv:1801.04381v4 [cs.CV]. DOI: 10.48550/arXiv.1801.04381.

[13] Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. Tech Report. arXiv preprint arXiv:1804.02767. Retrieved from arXiv:1804.02767v1 [cs.CV]. DOI: 10.48550/arXiv.1804.02767.

[14] He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. arXiv preprint arXiv:1703.06870v3. Retrieved from arXiv:1703.06870v3 [cs.CV]. DOI: 10.48550/arXiv.1703.06870.

[15] Cubuk, E. D., Zoph, B., Mane, D., Vasudevan, V., & Le, Q. V. (2018). AutoAugment: Learning Augmentation Policies from Data. arXiv preprint arXiv:1805.09501v3. Retrieved from arXiv:1805.09501v3 [cs.CV]. DOI: 10.48550/arXiv.1805.09501.

[16] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. Tech Report. arXiv preprint arXiv:1512.03385. Retrieved from arXiv:1512.03385v1 [cs.CV]. DOI: 10.48550/arXiv.1512.03385.

[17] Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2016). Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning.

arXiv preprint arXiv:1602.07261v2. Retrieved from arXiv:1602.07261v2 [cs.CV]. DOI: 10.48550/arXiv.1602.07261.

[18] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805v2. Retrieved from arXiv:1810.04805v2 [cs.CL]. DOI: 10.48550/arXiv.1810.04805.

[19] Rameshshanker, R., & Arebi, N. (2012). Endoscopy in inflammatory bowel disease: when and why. *World Journal of Gastrointestinal Endoscopy*, 4(6), 201–211. Published online 2012 Jun 16. doi: 10.4253/wjge.v4.i6.201. PMID: 22720120. PMCID: PMC3377861.

[20] Smedsrud, P. H., Thambawita, V., Hicks, S. A., Gjestang, H., Nedrejord, O. O., Næss, E., ... Halvorsen, P. (2021). Kvasir-Capsule, a video capsule endoscopy dataset. *Scientific Data*, 8(1), 142. Published online 2021 May 27. doi: 10.1038/s41597-021-00920-z. PMID: 34045470. PMCID: PMC8160146.

[21] Moshiree, B. (2009). Comparison of the Effect of Azithromycin Versus Erythromycin on Antroduodenal Pressure Profiles of Patients with Chronic Functional Gastrointestinal Pain and Gastroparesis. *Digestive Diseases and Sciences*, 55(3), 675-683. DOI: 10.1007/s10620-009-1038-3.

[22] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet

Classification with Deep Convolutional Neural Networks. In Advances in Neural Information Processing Systems (NIPS) (pp. 1097-1105).

[23] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.

[24] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. Nature, 323(6088), 533-536.

[25] Nair, V., & Hinton, G. E. (2010). Rectified Linear Units Improve Restricted Boltzmann Machines. In Proceedings of the 27th International Conference on International Conference on Machine Learning (ICML'10) (pp. 807-814).

[26] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research, 15, 1929-1958.

[27] Bridle, J. S. (1990). Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. In Proceedings of IJCNN (Vol. 1, pp. 227-236).